ЗАРУБЕЖНЫЕ СТАТИСТИЧЕСКИЕ ИССЛЕДОВАНИЯ

ЗСИ

# РАНГОВЫЕ КОРРЕЛЯЦИИ

•

М. КЕНДЭЛ

# RANK CORRELATION METHODS

MAURICE G. KENDALL, M. A., Sc. D

Fourth edition

GRIFFIN LONDON, 1970

# РАНГОВЫЕ КОРРЕЛЯЦИИ

**М. КЕНДЭЛ**

Перевод с английского, научное редактирование и предисловие **Е. М. Четыркина и Р. М. Энтова**

«СТАТИСТИКА» МОСКВА 1975

# ЗАРУБЕЖНЫЕ СТАТИСТИЧЕСКИЕ ИССЛЕДОВАНИЯ

(ТЕОРИЯ И МЕТОДЫ)

**З. С. И.**

*Шестая книга серии*

●

ВЫШЛИ ИЗ ПЕЧАТИ:

**Введение в теорию порядковых статистик**

**П. Массе. Критерии и методы оптимального определения капиталовложений**

**Г. Тейл. Экономические прогнозы и принятие решений**

**Г. Харман. Современный факторный анализ**

**Н. Дрейпер, Г. Смит. Прикладной регрессионный анализ**

ГОТОВЯТСЯ К ПЕЧАТИ:

**Э. Маленво. Статистические методы эконометрии. Вып. 1 и 2.**

**Л. Закс. Статистическое оценивание.**

●



К $\frac{10803^{1}—023}{008(01)—75}$ 127—75

1 Второй индекс 10805

## ПРЕДИСЛОВИЕ К РУССКОМУ ИЗДАНИЮ

В нашей стране неуклонно расширяется круг научных и практических работников, пользующихся в своих исследованиях математико-статистическими методами анализа. В связи с этим за последние годы был издан ряд переводных и оригинальных монографий, посвященных прикладному регрессионному анализу. Однако обычные методы регрессионного анализа оказываются недостаточными в тех случаях, когда признакам наблюдаемого явления не удается однозначно приписать те или иные абсолютные значения. В таких случаях исследователь часто пользуется методами ранговой корреляции; с помощью таких методов удается значительно расширить как возможности регрессионного анализа, так и круг решаемых задач.

В последнее время в прогнозировании и при решении ряда других задач стали широко применяться экспертные оценки. Методы ранговой корреляции в этой области являются едва ли не единственным путем обобщения экспертных оценок.

Элементарное изложение методов упорядочения и приемов ранговой корреляции можно найти в обычном курсе прикладной математической статистики. Однако в литературе до настоящего времени отсутствовала работа, которая была бы специально посвящена разностороннему теоретическому анализу методов ранговой корреляции. Предлагаемая книга призвана в какой-то мере восполнить этот пробел.

Имя автора—видного статистика Мориса Джорджа Кендэла—хорошо известно читателю. Он является одним из авторов книги «An introduction to the theory of statistics»[1] и трехтомного фундаментального труда «The Advanced Theory of Statistics»[2]. М. Кендэлу принадлежит, в частности, ряд новых идей и технических приемов в исследовании корреляции рангов; характерно, что один из рассматриваемых в книге коэффициентов ($r$) в мировой литературе обычно называют коэффициентом ранговой корреляции Кендэла.

Книга «Ранговые корреляции» впервые была издана в 1948 г., и с тех пор она пользуется неизменным вниманием читателей и неодно-

[1] Русский перевод: Юл Д. Э. и Кендэл М. Дж. Теория статистики. М., Госстатиздат, 1960.

[2] Первые два тома переведены на русский язык: Кендалл М., Стьюарт А. Теория распределений. М., «Наука», 1966 и Кендалл М., Стьюарт А. Статистические выводы и связи. М., «Наука», 1973.

кратно переиздавалась. Предлагаемый перевод выполнен с четвертого издания. Автор стремится всюду, где это представлялось возможным, приводить содержательную постановку задачи и снабжать изложение конкретными примерами. Большое внимание уделено условиям применения каждого из методов измерения ранговой корреляции. Особый интерес вызывает теоретический анализ проблемы так называемых связанных рангов. Следует отметить, правда, что при переходе к математическому обоснованию приводимых соотношений автор в ряде случаев опускает промежуточные выкладки и соображения. Вследствие этого требуется определенная самостоятельная работа, для того чтобы полностью разобраться в логике и последовательности формальных доказательств.

Студенты, изучающие курс математической статистики, а также практические работники обратят особое внимание на гл. 2, 3, 4, содержащие достаточно подробную информацию о практической технике расчета показателей ранговой корреляции. Читатели, заинтересованные в более строгом математическом обосновании приводимых соотношений, найдут соответствующий материал в гл. 5, 7, 10—12.

Переводчики стремились сохранить терминологию, встречавшуюся в ранее выполненных переводах. К числу исключений принадлежит термин tied ranks, который в книге Дж. Юла и М. Кендэла переведен как «объединенные ранги». Мы полагали, что более точно мысль автора воспроизводится в словах о «связи между рангами» (вместо «объединения рангов») и «связанных» (вместо «объединенных») рангах.

*Е. М. ЧЕТЫРКИН, Р. М. ЭНТОВ*

# *ГЛАВА 1.* ИЗМЕРЕНИЕ РАНГОВОЙ КОРРЕЛЯЦИИ

## Вводные замечания

**1.1.** Ряд объектов, расположенных в соответствии с некоторым признаком (в неодинаковой мере присущим этим объектам), называют *упорядоченным*. Сам процесс такого упорядочения называется *ранжированием*, а каждому члену ряда присваивается *ранг*.

**1.2.** Чаще всего ранги обозначаются порядковыми числительными $1, 2, \ldots, n$, где $n$ — количество объектов (хотя подобный выбор, вообще говоря, не столь существен). Таким образом, если какой-либо предмет или кто-либо после ранжирования будет занимать пятое место в ряду, его ранг равен 5. В дальнейшем этими числами часто будем оперировать так, как если бы они представляли собой количественные числительные, обычно фигурирующие в арифметике; будем складывать, вычитать и даже умножать эти числа. Поэтому следует разобраться в том, что означают подобные операции.

**1.3.** Предположим, например, что при ранжировании некоторого набора по признаку $A$ ранг объекта оказался равен 5, тогда как при ранжировании по другому признаку $B$ его ранг составил 8. Что выражает разница рангов, равная 3? Вычитание «пятого места» из «восьмого» не имеет смысла, и все-таки эта операция может иметь определенное содержание. Ведь когда мы говорим, что при упорядочении по признаку $A$ ранг объекта равен 5, это эквивалентно следующему утверждению: при упорядочении по $A$ четыре объекта оказались впереди, или, иными словами, данному объекту *предпочли* четыре других. Аналогично при ранжировании по признаку $B$ данному объекту предпочли семь других. Следовательно, при ранжировании по критерию $B$ количество предпочитаемых объектов на 3 превосходит число объектов, предпочитаемых при ранжировании по признаку $A$; цифра 3 в данном случае представляет собой не порядковое, а количественное числительное.

Стоит ли в самом начале курса обращать внимание на столь тонкие различия? Если читатель полагает, что это не имеет смысла, он может отложить рассмотрение указанных вопросов до тех пор, пока это не потребуется по ходу изложения. Однако он должен с самого начала осознать, что связанные с ранжированием вычислительные процедуры чаще всего основываются на измерении количества объектов, а не их порядковых номеров.

**1.4.** Можно назвать много различных способов упорядочения; упомянем лишь некоторые из них.

а. Задача может сводиться просто к упорядочению объектов по месту, которое они занимают в пространстве или во времени. Расположим, например, карты в колоде в некотором порядке, а затем перетасуем их. Новое расположение карт также характеризуется определенным порядком, ранжированием. Сравнив его со старым, можно увидеть, насколько тщательно были перетасованы карты. В этой задаче нас интересует только общее расположение карт в колоде, и мы не стремимся, скажем, упорядочить объекты в соответствии с «возрастанием» или «убыванием» того или иного присущего всем им признака.

б. Упорядочить объекты можно и по некоторому качеству, для которого не существует объективной абсолютной шкалы измерения. Мы можем, например, ранжировать образцы горных пород по твердости исходя из следующего простого критерия: $A$ тверже $B$, если $A$ оставляет царапину на $B$, когда они соприкасаются. Если $A$ оставляет царапину на $B$, а $B$ — на $C$, то $A$ будет оставлять царапину на $C$. Таким образом, прибегнув к ряду сопоставлений, мы сможем с достаточной точностью упорядочить рассматриваемые объекты (если только наш набор не включает такие два объекта, которые обладают одинаковой твердостью; (этот особый случай будет рассмотрен в гл. 3). Однако подобный способ не позволяет *измерить* абсолютную величину твердости горных пород. Мы всегда можем установить, что $A$ тверже $B$. Однако до тех пор, пока не построена та или иная шкала измерения абсолютных величин, мы не можем утверждать, что $A$, скажем, вдвое тверже $B$.

в. Упорядочение может проводиться в соответствии с измеряемой (или теоретически исчисляемой) величиной некоторого признака. Например, мы можем располагать людей в том или ином порядке в зависимости от их роста, а города — по численности населения. При этом не всегда требуется прибегать к самому процессу измерения: можно «на глаз» построить группу студентов по росту; однако в таких случаях критерий, по которому мы ранжируем, должен допускать возможность непосредственных сопоставлений.

г. Можно упорядочивать объекты по некоторому признаку, величину которого, по нашему мнению, в принципе можно измерить, но на практике (или даже теоретически) не удается прибегнуть к такому измерению в силу тех или иных причин. Например, мы можем упорядочить ряд лиц по их интеллектуальным способностям, полагая, что такое качество действительно существует и что можно разместить людей в том или ином порядке в соответствии с интенсивностью этого признака. В гл. 11 мы рассмотрим метод, который в некоторых случаях позволяет дать ответ на вопрос о том, правомерны ли подобные предпосылки. Этот случай отличается от ситуации, упоминавшейся в пункте «б», поскольку в данном примере содержательные соображения убеждают нас в том, что ранжирование возможно, тогда как в ситуации пункта «б» мы просто выдвигаем гипотезу относительно возможности подобных измерений.

**1.5.** Количественную характеристику, которая может менять свое значение при переходе от одного из элементов совокупности к другому, в теоретической статистике называют *случайной величиной*. Так, значения того или иного признака, которые можно измерить, представляют собой, разумеется в рамках соответствующей шкалы измерения, случайную величину. Такой набор мы всегда можем упорядочить, руководствуясь местом, которое занимает на шкале измерения каждый объект, после чего имеем право сказать, что значения *случайной величины* представлены соответствующими рангами. Следовательно, можно рассматривать процесс упорядочения как не совсем точный способ выражения порядковых отношений между элементами — не совсем точный потому, что он не позволяет нам судить о том, насколько близко друг к другу расположены на шкале измерения различные элементы рассматриваемой совокупности. Per contra* проигрывая в точности, процесс ранжирования выигрывает в общности подхода. Допустим, например, что мы «растянули» отрезок, характеризующий шкалу измерения; больше того, допустим, что мы с разной интенсивностью «растягивали» отдельные промежутки рассматриваемого отрезка. В любом случае порядок расположения элементов не изменится, или, выражаясь языком математики, такое упорядочение *инвариантно* относительно изменений масштаба шкалы.

**1.6.** Теория рангов впервые возникла как ответвление теории случайных процессов. На начальной стадии в рангах чаще всего видели просто удобный аппарат, благодаря которому удается обойтись без измерения абсолютной величины переменных и тем самым сэкономить время или усилия. Благодаря использованию рангов можно было избежать трудностей, связанных с построением объективной шкалы абсолютных значений. Позднее статистика рангов смогла завоевать признание благодаря своим собственным достоинствам. В начальных разделах книги наше внимание будет сосредоточено на самом процессе упорядочения, независимо от существования тех или иных шкал измерения абсолютных величин. Таким образом, предлагаемые методы обладают достаточно большой степенью общности. В гл. 9 и 10 будет рассмотрено соотношение между рангами и случайными величинами.

## Ранговая корреляция

**1.7.** Предположим, что группу учеников ранжировали в соответствии с их способностями, обнаруженными на уроках музыки и математики. Обозначим детей буквами от $A$ до $J$ и выпишем следующие две последовательности рангов:

| ученики | $A$ | $B$ | $C$ | $D$ | $E$ | $F$ | $G$ | $H$ | $I$ | $J$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| математика | 7 | 4 | 3 | 10 | 6 | 2 | 9 | 8 | 1 | 5 | (1.1) |
| музыка | 5 | 7 | 3 | 10 | 1 | 9 | 6 | 2 | 8 | 4 | |

* Per contra (лат.) — *здесь* однако. — *Прим. перев.*

Рассмотрим теперь вопрос, существует ли зависимость между музыкальными и математическими способностями. Даже беглого взгляда на приведенные данные достаточно, для того чтобы увидеть, что четкого соответствия между ними не существует. Однако некоторые ученики занимают одинаковое или почти одинаковое место в обоих рядах. Наличие (или отсутствие) связи между этими показателями станет более очевидным, если мы расположим элементы первого ряда в порядке возрастания (в последовательности натуральных чисел):

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ученики | $I$ | $F$ | $C$ | $B$ | $J$ | $E$ | $A$ | $H$ | $G$ | $D$ | |
| математика | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | (1.2) |
| музыка | 8 | 9 | 3 | 7 | 4 | 1 | 5 | 2 | 6 | 10 | |

Нужно определить степень соответствия между этими двумя последовательностями порядковых оценок, или, другими словами, измерить тесноту *ранговой корреляции*. Поэтому изложим методику построения соответствующего коэффициента корреляции, обозначив его буквой $\tau$ (тау).

1.8. Коэффициент корреляции должен обладать следующими тремя свойствами:

а) если между последовательностями порядковых оценок имеется полное соответствие, т. е. если каждый объект занимает одно и то же место в обоих рядах, то $\tau$ должен быть равен $+1$, что означает полную положительную корреляцию;

б) если налицо полная отрицательная зависимость, т. е. если в одной последовательности оценки расположены в обратном порядке по сравнению с другой, $\tau = -1$, что означает полную отрицательную корреляцию;

в) в остальных ситуациях $\tau$ лежит между предельными значениями; можно утверждать, что возрастание $\tau$ от $-1$ до $+1$ в некотором приемлемом для нас смысле характеризует увеличивающееся соответствие между двумя последовательностями порядковых оценок.

Первые два соображения просто вводят общепринятый масштаб измерения; как ни условен подобный масштаб, он в высшей степени полезен для практики.

1.9. В первой последовательности (1.1) выделим какую-нибудь пару рангов, например $A$ и $B$. Их значения 7 и 4 образуют обратный порядок величин (прямым порядком мы будем называть порядок натурального ряда 1,..., 10); парам, образующим обратный порядок, будем приписывать значения $-1$; парам, значения которых образуют прямой порядок, $+1$. Во второй последовательности ранги $A$ и $B$ равны соответственно 5 и 7; они образуют прямой порядок, и, следовательно, в этой последовательности паре $AB$ приписывается значение $+1$.

Перемножив значения, приписанные этим парам, в первой и во второй последовательностях, получим произведение, равное $-1$. Ясно, что для любой пары оно будет равно $+1$ в тех случаях, когда соответствующие ранги в обеих последовательностях расположены в одинаковом порядке, и $-1$, если эти ранги образуют различный порядок. Можно сказать, что мы приписываем значения $+1$ или $-1$ в зависимости от того, согласован или не согласован между собой порядок пары в обеих последовательностях.

Проделаем все эти вычисления для каждой пары, полученной из 10 элементов. В результате придется взять 45 пар, каждой из которых будут приписаны следующие значения (запишем соответствующие цифры для каждой пары, с тем чтобы читатель мог непосредственно проследить за всеми расчетами; однако, как будет показано ниже, на практике можно обойтись и без столь громоздких вычислений):

| Пара | Значение | Пара | Значение | Пара | Значение | Пара | Значение |
|---|---|---|---|---|---|---|---|
| *AB* | $-1$ | *BF* | $-1$ | *DE* | $+1$ | *FH* | $-1$ |
| *AC* | $+1$ | *BG* | $-1$ | *DF* | $+1$ | *FI* | $+1$ |
| *AD* | $+1$ | *BH* | $-1$ | *DG* | $+1$ | *FJ* | $-1$ |
| *AE* | $+1$ | *BI* | $-1$ | *DH* | $+1$ | *GH* | $+1$ |
| *AF* | $-1$ | *BJ* | $-1$ | *DI* | $+1$ | *GI* | $-1$ |
| *AG* | $+1$ | *CD* | $+1$ | *DJ* | $+1$ | *GJ* | $+1$ |
| *AH* | $-1$ | *CE* | $-1$ | *EF* | $-1$ | *HI* | $-1$ |
| *AI* | $-1$ | *CF* | $-1$ | *EG* | $+1$ | *HJ* | $-1$ |
| *AJ* | $+1$ | *CG* | $+1$ | *EH* | $+1$ | *IJ* | $-1$ |
| *BC* | $+1$ | *CH* | $-1$ | *EI* | $-1$ | | |
| *BD* | $+1$ | *CI* | $-1$ | *EJ* | $-1$ | | |
| *BE* | $-1$ | *CJ* | $+1$ | *FG* | $-1$ | | |

Сумма значений, равных $+1$ (назовем ее $P$), составляет 21, а сумма значений, равных $-1$ (обозначим ее $Q$), составляет $-24$. Сложив эти два числа, получим общую сумму приписанных значений $S$, равную $-3$.

Если бы во всех парах наблюдался одинаковый порядок, то каждое из 45 приписываемых им значений было бы положительным; следовательно, максимальное значение $S$ равно 45. Рассуждая аналогично, приходим к выводу, что минимальное значение $S$ должно составлять $-45$. Таким образом, значение $\tau$ равно:

$$\frac{\text{Действительная сумма приписываемых значений} = -3}{\text{Максимально возможная сумма этих значений} = 45} = -0{,}07.$$

Эта величина близка к нулю; отсюда следует, что корреляция между двумя последовательностями рангов очень мала. Нулевое значение $\tau$ может интерпретироваться как свидетельство независимости, которая лежит, так сказать, на полпути между полной положительной зависимостью и полной отрицательной зависимостью[1].

**1.10.** Рассмотрим теперь общий случай, когда имеются две последовательности рангов, каждая из которых содержит $n$ членов; количество пар, подлежащих сравнению, равно числу способов, с помощью которых можно выбрать два предмета из набора, содержащего $n$ предме-

---

[1] Этот критерий следует применять с достаточной осмотрительностью. Коэффициент $\tau$ может обращаться в нуль и в тех случаях, когда изучаемые соотношения связаны некоторой сложной зависимостью.

тов; эта величина равна $^1/_2\, n\,(n-1)$; иногда ее обозначают также $\binom{n}{2}$. Указанное число характеризует наибольшую возможную сумму приписанных значений; такая величина может быть достигнута лишь тогда, когда порядок рангов в обеих последовательностях полностью совпадает. Обозначив общее количество приписанных значений буквой $S$, введем следующее определение коэффициента корреляции:

$$\tau = \frac{S}{\frac{1}{2}\,n\,(n-1)}. \tag{1.3}$$

Пусть $P$ и $Q$ означают соответственно суммы приписываемых положительных и отрицательных значений $\left(\text{так что } P + Q = \frac{1}{2}\,n\,(n-1)\right)$; тогда можно записать эквивалентные формулы для вычисления

$$\tau = \frac{P-Q}{\frac{1}{2}\,n\,(n-1)} = \tag{1.4}$$

$$= 1 - \frac{2Q}{\frac{1}{2}\,n\,(n-1)} = \tag{1.5}$$

$$= \frac{2P}{\frac{1}{2}\,n\,(n-1)} - 1. \tag{1.6}$$

**1.11.** Для того чтобы найти величину $S$ (или, что равносильно, значения $P$ и $Q$), не требуется проделывать всю описанную выше вычислительную процедуру. Существуют и более простые методы. Наиболее легким из них, вероятно, является следующий.

а. Рассмотрим формулу (1.2). В тех случаях, когда одна последовательность рангов представляет собой натуральный ряд 1, 2, ..., $n$, оценки, приписываемые каждой паре значений этого ряда, положительны. Следовательно, значения $+1$, входящие слагаемыми в сумму $P$, будут приписываться только тем парам второго ряда, которые образуют прямой порядок. Требуется лишь пересчитать их. Вторая последовательность рангов имеет вид:

8 9 3 7 4 1 5 2 6 10.

Рассмотрим сначала пары, которые с первым элементом, т. е. с 8, образуют остальные элементы последовательности; мы видим, что справа от элемента 8 расположены два члена, которые превосходят 8 по величине. Следовательно, первые слагаемые суммы $P$, составляют $+2$. Рассмотрим теперь пары, которые образуют с элементом 9 остальные элементы (кроме пары 8 9, которая уже была учтена ранее); они увеличивают $P$ на $+1$. Аналогично, рассматривая пары, которые стоящие справа элементы образуют с цифрой 3, мы приходим к выводу, что следующее слагаемое суммы $P$ равно $+5$. Продолжая этот расчет находим:

$$P=2+1+5+1+3+4+2+2+1=21.$$

Следовательно, из (1.6) имеем:

$\tau = \frac{42}{45} - 1 = -0{,}07$, что совпадает с полученным ранее результатом.

б. Допустим, что очень сложно так организовать последовательности, чтобы одна из них была расположена в строгом порядке. Тогда можно поступить следующим образом. Запишем друг под другом рассмотренные выше последовательности, а над ними выпишем числа натурального ряда от 1 до 10:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| $A$ | 7 | 4 | 3 | 10 | 6 | 2 | 9 | 8 | 1 | 5 |
| $B$ | 5 | 7 | 3 | 10 | 1 | 9 | 6 | 2 | 8 | 4 |

Элементу 1 ряда $B$ соответствует элемент 6 в последовательности $A$. В натуральном ряду справа от 6 стоят 4 элемента. Включим в $P$ слагаемое, равное $+4$, и вычеркнем элемент 6 из процедуры ранжирования в соответствии с порядком натурального ряда. Далее, элементу 2 ряда $B$ соответствует элемент 8 ряда $A$, а в натуральном ряду справа от 8 стоят два члена. Включаем в $P$ еще одно слагаемое, равное $+2$, и вычеркиваем 8 из верхнего ряда. Закончив этот расчет, мы найдем общую сумму.

$$4+2+5+3+2+1+2+1+0=21,$$

которая совпадает с полученным выше значением $P$.

Желая убедиться в правильности подобной вычислительной процедуры, перепишем наши последовательности рангов таким образом, чтобы элементы последовательности $B$ совпадали с членами натурального ряда (в таком порядке мы рассматривали ранее элементы обоих рядов):

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 5 | 8 | 3 | 10 | 1 | 7 | 2 | 9 | 6 | 4 |
| $A$ | 6 | 8 | 3 | 5 | 7 | 9 | 4 | 1 | 2 | 10 |
| $B$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Легко видеть, что при использовании метода «б» слагаемые $P$ образуются согласно тому же закону, которым мы пользовались при сопоставлении рядов $B$ и $A$ по методу «а». Например, в ряду $A$ справа от 6 стоят 4 элемента, которые по величине превосходят 6, справа от 8 в ряду $A$ — два члена, которые по величине превосходят 8, и т. д.

**1.12.** Чтобы получить некоторое представление о тех значениях, которые может принимать $\tau$ в различных случаях, выпишем некоторые последовательности чисел от 1 до 10 и значения $\tau$, характеризующие степень соответствия этих последовательностей порядку чисел натурального ряда. Читателю рекомендуется самостоятельно проделать соответствующие расчеты, сверив результаты своих вычислений с приведенными в таблице значениями:

| Исходная последовательность чисел | | | | | | | | | | | Значения $\tau$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *a* | 4 | 7 | 2 | 10 | 3 | 6 | 8 | 1 | 5 | 9 | +0,11 |
| *b* | 1 | 6 | 2 | 7 | 3 | 8 | 4 | 9 | 5 | 10 | +0,56 |
| *c* | 7 | 10 | 4 | 1 | 6 | 8 | 9 | 5 | 2 | 3 | —0,24 |
| *d* | 6 | 5 | 4 | 7 | 3 | 8 | 2 | 9 | 10 | 1 | +0,02 |
| *e* | 10 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | +0,60 |
| *f* | 10 | 9 | 8 | 7 | 6 | 1 | 2 | 3 | 4 | 5 | —0,56 |

### $\tau$ как коэффициент неупорядоченности

**1.13.** Введенный нами коэффициент может служить количественной характеристикой общего соответствия между отдельными парами элементов (слово «соответствие» здесь означает соответствие порядков рассматриваемых элементов). Благодаря этому он может быть использован при согласовании между собой двух последовательностей. Чтобы лучше разобраться в том, что означает этот коэффициент, прибегнем к следующему способу рассуждений. Рассмотрим две последовательности, каждая из которых содержит числа от 1 до 7:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| *A* | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| *B* | 6 | 3 | 5 | 7 | 1 | 2 | 4 |

Мы можем перейти от *B* к *A*, последовательно меняя местами стоящие рядом числа. Например, в последовательности *B* будем перемещать влево число 1; тогда после четырех перестановок мы получим:

6 3 5 1 7 2 4
6 3 1 5 7 2 4
6 1 3 5 7 2 4
1 6 3 5 7 2 4

Затем будем перемещать влево число 2, для этого потребуются еще четыре перестановки:

1 2 6 3 5 7 4.

Поменяем местами числа 3 и 6:

1 2 3 6 5 7 4.

Далее проведем трехкратную перестановку числа 4:

1 2 3 4 6 5 7.

Наконец, поменяем местами числа 6 и 5. В результате нам удалось получить последовательность $A$, т. е. последовательность натурального ряда.

Вся процедура перехода от $B$ к $A$ потребовала 13 перестановок, и мы не могли бы получить тот же результат, применяя меньшее число взаимных перемещений. Можно было бы проделать больше перестановок, например дважды поменять местами числа 1 и 2, а затем начать описанную выше процедуру перемещения. Покажем, что всегда существует некое минимальное число перестановок, необходимое для перехода от одной последовательности к другой, содержащей то же количество элементов. Обозначим это число буквой $s$.

В следующей главе будут выведены две эквивалентные формулы:

$$s = Q$$

и

$$s = \frac{1}{2}\left[\frac{1}{2}n(n-1) - S\right], \tag{1.7}$$

которые устанавливают простое соотношение между числом перестановок $s$ и $Q$ — количеством отрицательных значений, или $S$ — общим количеством значений, приписанных парам объектов. В рассматриваемом примере $S = -5$, $n = 7$ и, следовательно, мы вновь можем убедиться в том, что

$$s = \frac{1}{2}(21 + 5) = 13.$$

Из (1.5) и (1.7) следует, что

$$\tau = 1 - \frac{2s}{\frac{1}{2}n(n-1)}. \tag{1.8}$$

Таким образом, $\tau$ представляет собой простую функцию минимального числа перестановок соседних элементов, необходимых для перехода от одного порядка элементов к другому; говоря короче, $\tau$ может служить количественной характеристикой, мерой неупорядоченности.

## Коэффициент Спирмэна ρ

1.14. Рассмотрим еще один коэффициент ранговой корреляции, обозначаемый ρ (ро). Его называют коэффициентом К. Спирмэна в честь автора, который впервые ввел такой показатель при исследованиях в области психологии. Вернемся теперь к нашим двум последовательностям рангов, состоящим из 10 элементов (см. (1.1)).

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Математика | 7 | 4 | 3 | 10 | 6 | 2 | 9 | 8 | 1 | 5 |
| Музыка | 5 | 7 | 3 | 10 | 1 | 9 | 6 | 2 | 8 | 4 |
| **Разности $d$** | 2 | —3 | 0 | 0 | 5 | —7 | 3 | 6 | —7 | 1 |
| **Разности $d^2$** | 4 | 9 | 0 | 0 | 25 | 49 | 9 | 36 | 49 | 1 |

Из ранга по математике мы вычли ранг по музыке и записали результат в строку, названную «разности $d$». Легко увидеть, что сумма этих разностей должна равняться нулю, поскольку речь идет о разности двух величин, каждая из которых представляет собой сумму числа от 1 до 10 (тем самым обеспечивается возможность арифметической проверки). Кроме того, мы выписали квадраты этих разностей. Обозначив их сумму через $S(d^2)$, определим коэффициент Спирмэна с помощью следующего соотношения:

$$\rho = 1 - \frac{6S(d^2)}{n^3 - n}. \tag{1.9}$$

Вычислим коэффициент Спирмэна для нашего примера:

$$\rho = 1 - \frac{6 \times 182}{990} = -0{,}103.$$

**1.15.** Пусть заданы две одинаковые последовательности. Тогда все разности $d$ равны нулю; из (1.9) следует, что $\rho = 1$. Предположим теперь, что элементы двух последовательностей расположены в обратном порядке. Покажем, что в этом случае $\rho = -1$.

Пусть $n$ — нечетное число, равное $2m + 1$. Без утраты общности одну из последовательностей можно расположить в натуральном порядке, тогда наши последовательности и соответствующие разности будут иметь следующий вид:

$$\begin{array}{lccccccc}
A & 1, & 2, \ldots, & m, & m+1, & m+2, \ldots, & 2m, & 2m+1 \\
B & 2m+1, & 2m, \ldots, & m+2, & m+1, & m, \ldots, & 2, & 1 \\
\hline
d & -2m, & -(2m-2), \ldots, & -2, & 0, & 2, \ldots, & 2m-2, & 2m
\end{array}. \tag{1.10}$$

Таким образом, сумма квадратов составляет:

$$\begin{aligned}
S(d^2) &= 8[m^2 + (m-1)^2 + \ldots + 2^2 + 1^2] = \\
&= 8m(m+1)(2m+1)/6 = \\
&= \frac{1}{3}(n-1)(n+1)(n) = \\
&= \frac{1}{3}(n^3 - n).
\end{aligned}$$

Подставив это значение в (1.9), можно найти численное значение:

$$\rho = 1 - \frac{6}{3} = -1.$$

Допустим теперь, что $n$ — четное число, скажем, $n = 2m$. Выпишем наши последовательности и соответствующие разности:

$$\begin{array}{lccccccc}
A & 1, & 2, \ldots, & m, & m+1, \ldots, & 2m-1, & m \\
B & 2m, & 2m-1, \ldots, & m+1, & m, \ldots, & 2, & 1 \\
\hline
d & -(2m-1), & -(2m-3), \ldots, & -1, & 1, \ldots, & 2m-3, & 2m-1
\end{array}. \tag{1.11}$$

Следовательно,

$$\begin{aligned} S(d^2) &= 2[(2m-1)^2+(2m-3)^2+\ldots+3^2+1^2] = \\ &= 2[(2m)^2+(2m-1)^2+(2m-2)^2+\ldots+3^2+2^2+1^2] - \\ &\quad -2[(2m)^2+(2m-2)^2+\ldots+2^2] = \\ &= 4m(2m+1)(4m+1)/6 - 8m(m+1)(2m+1)/6 = \\ &= \frac{1}{3}(n^3-n). \end{aligned}$$

Подставляя, как и в предшествующем случае, полученный результат в (1.9), мы найдем, что $\rho = -1$.

Таким образом, коэффициент $\rho$ может принимать значения $-1$ и $+1$. На стр. 32 будет показано, что коэффициент Спирмэна не может лежать вне этих границ и достигает экстремальных значений только при полной согласованности или полной рассогласованности между элементами двух последовательностей.

**1.16.** Читателю, знакомому с методикой расчета статистической оценки дисперсии и, в частности, с вычислением среднего квадратического отклонения, легче понять причины, побудившие нас возвести в *квадрат* разности рассматриваемых рангов, и лишь затем сложить их. Совершенно очевидно, что при построении коэффициента мы не можем воспользоваться суммой разностей, $S(d)$, поскольку она равна нулю. На это можно возразить, что просуммировав абсолютные величины разностей, можно было бы получить несколько более простой коэффициент; и к этому сводилось, между прочим, одно из первоначальных предложений Спирмэна. Но в таком случае возникли бы затруднения на последующих ступенях анализа, особенно усложнились бы проблемы выборочного обследования. Поэтому в дальнейшем изложении мы не будем пользоваться подобной методикой.

**1.17.** Оценка $Q$ в уравнении (1.5) представляет собой просто число расположенных в неодинаковом порядке пар, образованных элементами двух последовательностей. Любое такое нарушение порядка мы будем называть «инверсией»; тем самым $\tau$ оказывается линейной функцией от количества инверсий. Любопытно, что $\rho$ тоже можно считать коэффициентом инверсии, если только предположить, что каждая инверсия взвешена. В самом деле, предположим, что некоторая пара рангов $(i, j)$ образует инверсию (пусть $i < j$). Припишем этой инверсии оценку $(j - i)$. Тогда общую сумму приписанных значений можно представить так:

$$V = \frac{1}{2} S(d^2) \tag{1.12}$$

и, следовательно,

$$\rho = 1 - \frac{12V}{(n^3-n)}. \tag{1.13}$$

Этот результат будет доказан в следующей главе.

Рассмотрим пример. Обратимся вновь к двум последовательностям, приведенным в 1.13.

$$\begin{array}{cccccccc} A & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ B & 6 & 3 & 5 & 7 & 1 & 2 & 4 \end{array} \tag{1.14}$$

Выпишем те пары рангов, которые образуют инверсию:

| Пары рангов, образующие инверсию | Вес | Пары рангов, образующие инверсию | Вес |
|---|---|---|---|
| 6 3 | 3 | 5 1 | 4 |
| 6 5 | 1 | 5 2 | 3 |
| 6 1 | 5 | 5 4 | 1 |
| 6 2 | 4 | 7 1 | 6 |
| 6 4 | 2 | 7 2 | 5 |
| 3 1 | 2 | 7 4 | 3 |
| 3 2 | 1 | | 40 |

Как было показано в 1.13, общее число инверсий равно 13 и

$$\tau=-\frac{5}{21}=-0{,}24.$$

Просуммировав вес, приходящийся на каждую инверсию, получим число 40, следовательно,

$$\rho=1-\frac{480}{336}=-0{,}43.$$

Нетрудно подсчитать, что для последовательности (1.14) $S(d^2) = 80$, а затем, воспользовавшись формулой (1.12), проверить значение $V$.

## Сопряженные последовательности рангов

1.18. Следует отметить, что коэффициенты $\tau$ и $\rho$ обладают общим свойством. Рассмотрим сначала отношение между последовательностью $A$ и последовательностью $B$ $(1, \ldots, n)$, в которой элементы расположены в натуральном порядке, а затем соотношение между $A$ и такой последовательностью $B'$ $(n, \ldots, 1)$, в которой элементы расположены в порядке, обратном $B$. Тогда значения коэффициента $\tau$ окажутся равными по величине и противоположными по знаку. Это следует из самого определения $\tau$; ведь изменить порядок расположения элементов $B$ на противоположный — значит изменить знак каждого слагаемого, входящего в сумму $S$, и, следовательно, в конечном счете изменить знак $S$. Таким образом, справедливо следующее утверждение: если коэффициент корреляции рангов между последовательностями $A$ и $B$ (причем элементы в каждой из них не обязательно должны быть расположены в натуральном порядке) равен $\tau$, то корреляция рангов между $A$ и $B'$ измеряется величиной $-\tau$. Рассмотрим, например, такие две последовательности, содержащие по семь элементов:

$$\begin{array}{cccccccc} A & 4 & 1 & 6 & 7 & 5 & 2 & 3 \\ B & 7 & 6 & 3 & 1 & 4 & 5 & 2 \end{array} \tag{1.15}$$

Переставим эти пары таким образом, чтобы элементы последовательности $A$ были расположены в натуральном порядке:

$$\begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 6 & 5 & 2 & 7 & 4 & 3 & 1 \end{array}$$

Теперь легко найти значение $\tau$, оно равно — 11/21, или —0,52. Переставив элементы верхней последовательности в обратном порядке, мы получим следующие пары:

$$\begin{array}{cccccccc} C & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\ D & 6 & 5 & 2 & 7 & 4 & 3 & 1 \end{array}$$

В таком случае корреляция рангов $C$ и $D$ измеряется величиной +0,52. Расположим теперь пары таким образом, чтобы порядок элементов в последовательности $D$ совпадал с порядком $B$ (1.15).

В таком случае

$$\begin{array}{cccccccc} A' & 4 & 7 & 2 & 1 & 3 & 6 & 5 \\ B & 7 & 6 & 3 & 1 & 4 & 5 & 2 \end{array} \tag{1.16}$$

Последовательности рангов $A$ и $A'$ будем считать сопряженными. Рассматривая степень их корреляции с последовательностью $B$, мы получим одинаковые по величине, но противоположные по знаку коэффициенты $\tau$.

**1.19.** Аналогичные соотношения, но не столь простым способом, можно вывести для коэффициента $\rho$. Значения $\rho$ при сопоставлении последовательностей $A$, $B$ и при сопоставлении последовательностей $A'$, $B$ равны по величине и противоположны по знаку. В следующей главе эти результаты будут доказаны в общем виде. До тех пор читатель может самостоятельно убедиться, в частности, в том, что при сопоставлении между собой последовательностей (1.15) $\rho = -17/28$, а для (1.16) $\rho = +17/28$.

**1.20.** Таким образом, если мы пользуемся коэффициентами $\tau$ и $\rho$ для измерения корреляции рангов, то шкала допустимых значений в определенном смысле симметрична относительно нуля. Она ограничена числами +1 и —1, и любому положительному значению коэффициентов $\tau$ или $\rho$ соответствует равная ему по модулю и обратная по знаку величина. Отрицательное значение указанных коэффициентов характеризует корреляцию рангов в том случае, когда одна из исходных последовательностей расположена в обратном порядке, а другая сохраняет прежний порядок. Можно утверждать, что такие шкалы не имеют смещения.

**1.21.** Не следует полагать, что численные значения $\tau$ и $\rho$ будут одинаковы для любых двух последовательностей (если только не рассматриваются случаи полной согласованности и рассогласованности). Сопоставляя последовательности, приведенные в 1.12, с порядком натурального ряда, можно получить следующие значения коэффициентов:

| Последовательность | $\tau$ | $\rho$ |
|---|---|---|
| *a* | +0,11 | +0,14 |
| *b* | +0,56 | +0,64 |
| *c* | −0,24 | −0,37 |
| *d* | +0,02 | +0,03 |
| *e* | +0,60 | +0,45 |
| *f* | −0,56 | −0,76 |

На этом примере можно убедиться в том, что в практических задачах неизбежно приходится сталкиваться со следующей проблемой. Подобно паре термометров, один из которых измеряет температуру по Цельсию, а другой — по Фаренгейту, у наших коэффициентов разные масштабы, при этом они отличаются друг от друга не только шкалой измерения, но и тем, что при подсчете коэффициента $\rho$ инверсиям более отдаленных (по величине) друг от друга элементов приписываются бóльшие веса. На практике мы чаще всего сталкиваемся с такой ситуацией: если значения обоих коэффициентов не слишком близки к единице, то $\rho$ примерно на 50% превосходит $\tau$ по своей абсолютной величине.

## Неравенство Дэниелса

**1.22.** Если заданы последовательности ранговых оценок, можно установить неравенства, которые связывают между собой коэффициенты $\tau$ и $\rho$. Первое неравенство было выведено Дэниелсом:

$$-1 \leqslant \frac{3(n+2)}{n-2}\tau - \frac{2(n+1)}{n-2}\rho \leqslant 1, \tag{1.17}$$

где $n$ — количество членов в сопоставляемых последовательностях. При больших $n$ можно пользоваться приближенным соотношением:

$$-1 \leqslant 3\tau - 2\rho \leqslant 1. \tag{1.18}$$

Если $\tau$ больше нуля, то может достигаться устанавливаемый неравенством верхний предел, а нижний предел недостижим; если $\tau$ меньше нуля, может достигаться нижний предел, а верхний — недостижим. При $\tau = 0$ разность может принимать оба предельных значения. Следовательно, могут существовать и такие последовательности, у которых корреляция рангов характеризуется $\tau = 0$ и $\rho = 0{,}5$. Но эти случаи носят довольно специфический характер. Рассмотрим, например, такие последовательности:

$$\begin{array}{ccccccccc} A & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ B & 5 & 6 & 7 & 8 & 1 & 2 & 3 & 4 \end{array} \tag{1.19}$$

Найдем

$$\rho = -0{,}53; \quad \tau = -0{,}14.$$

## Неравенство Дарбина—Стюарта

1.23. Другие неравенства для $\rho$ и $\tau$ были выведены Дарбином и Стюартом: они позволяют, зная величину $\tau$, установить верхний и нижний пределы, которыми ограничены значения $\rho$. С помощью этих неравенств удается установить связь между $V$ (см. 1.12) и $Q$ для любых последовательностей:

$$V \geqslant \frac{2}{3} Q \left(1 + \frac{Q}{n}\right). \tag{1.20}$$

В некоторых случаях указанное выражение может превратиться в равенство. Используя соотношения (1.13) и (1.5), мы получаем

$$\rho \leqslant 1 - \frac{1-\tau}{2(n+1)} [(n-1)(1-\tau)+4], \quad \tau \geqslant 0, \tag{1.21}$$

т. е., зная величину $\tau$ ($\tau \geqslant 0$), мы можем отыскать верхний предел для значений $\rho$. Из выражения (1.17) находим нижний предел значения $\rho$:

$$\rho \geqslant \frac{3n\tau-(n-2)}{2(n+1)}, \quad \tau \geqslant 0. \tag{1.22}$$

При больших $n$ можно пользоваться приближенной формулой, имеющей вид:

$$\frac{3}{2}\tau - \frac{1}{2} \leqslant \rho \leqslant \frac{1}{2} + \tau - \frac{1}{2}\tau^2, \quad \tau \geqslant 0, \tag{1.23}$$

причем $\rho$ может принимать два предельных значения. Пусть, например, $\tau = 0$, тогда $-\frac{1}{2} \leqslant \rho \leqslant \frac{1}{2}$. Если $\tau = \frac{1}{2}$, $\frac{1}{4} \leqslant \rho \leqslant \frac{7}{8}$, при $\tau = 0{,}9$, $0{,}85 \leqslant \rho \leqslant 0{,}995$.

При $\tau < 0$

$$\frac{1}{2}\tau^2 + \tau - \frac{1}{2} \leqslant \rho \leqslant \frac{3}{2}\tau + \frac{1}{2}. \tag{1.24}$$

В следующей главе будут приведены доказательства этих результатов. Описанные неравенства позволяют лучше осознать уже высказывавшееся выше соображение: хотя коэффициенты $\rho$ и $\tau$ связаны между собой, эта связь не столь элементарна.

## Некоторые замечания

1.24. Вообще говоря, коэффициент $\rho$ легче рассчитать, чем $\tau$. И все же, как будет показано в последующих главах, в силу целого ряда практических и теоретических соображений больший интерес представляет коэффициент $\tau$, а не $\rho$. Важнейшие методы, приводимые в этой книге, также предполагают применение коэффициента $\tau$. Мы не будем на данной стадии изложения сравнивать между собой относительные достоинства этих коэффициентов, но все же здесь стоит упомянуть одно любопытное практическое соображение.

Иногда, после того как ранжирование уже проведено, могут появиться новые элементы и возникает необходимость в дальнейшем упорядочении всей последовательности. Аналогичная ситуация может сложиться также при следующих обстоятельствах: предположим, что мы выписываем ранговые оценки множества неупорядоченных объектов, различающихся между собой по величине или отмеченных неодинаковыми условными значками; при этом легко допустить ошибки, которые обнаружатся при завершении процесса ранжирования, — некоторые элементы последовательности окажутся неучтенными. Это потребует вычисления коэффициента $\rho$ *de novo** с использованием всей совокупности данных, в то время как коэффициент $\tau$ при добавлении к последовательности новых элементов не требует полного пересчета данных. Поясним это соображение с помощью числового примера.

*Пример 1.1.*

Нескольким фирмам были разосланы опросные листы, в которых содержалась просьба конфиденциально сообщить норму выплачиваемого дивиденда, которую компания предполагает огласить на ближайшем годовом собрании акционеров. Будем полагать, что все фирмы могут ответить на этот вопрос, однако не исключена следующая возможность: фирмы, предполагающие выплатить более высокие дивиденды, менее охотно станут отвечать на запрос, задерживая ответ, или вообще окажутся от заполнения опросного листа. Будем полагать, кроме того, что все нормы дивиденда различны. Подобные предположения, вероятно, не слишком реалистичны, однако они упростят построение числового примера.

К некоторому сроку от фирм будет получено определенное число ответов; теперь необходимо завершить наше обследование и сформулировать полученные результаты. Насколько правомерно полагать, что присланные ответы могут служить репрезентативной характеристикой всей совокупности адресатов? Есть ли какое-либо основание полагать, что дивиденды в фирмах, которые ответили раньше, имеют систематические отличия от дивидендов в фирмах, ответивших позже?

Допустим, что мы получили 15 ответов в следующем порядке:

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (*A*) Порядок получения ответа | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| (*B*) Норма дивиденда (в %) | 15 | 13 | 12 | 16 | 25 | 8 | 9 | 14 | 17 | 11 | 18 | 20 | 10 | 21 | 19 |
| (*C*) Ранг нормы дивиденда | 8 | 6 | 5 | 9 | 15 | 1 | 2 | 7 | 10 | 4 | 11 | 13 | 3 | 14 | 12 |

Если действительно существует зависимость между временем получения ответа и величиной дивиденда, такая зависимость должна проявиться в корреляции рангов этих величин. При этом упорядочение нормы дивиденда проводится в порядке ее возрастания (соответствующие порядковые номера приведены в последней строке таблицы). Рас-

---

* de novo (*лат.*) — заново. — *Прим. перев.*

смотрим корреляцию рангов между последовательностями $A$ и $C$; в таком случае $S = 25$.

$$\tau = \frac{25}{105} = +0{,}24.$$

Подсчитаем величину $S\,(d^2)$:

$$S\,(d^2) = 392,$$

$$\rho = 1 - \frac{392}{560} = +0{,}30.$$

Из приведенных расчетов следует, что между последовательностями $A$ и $C$ существует некоторая, хотя и слабая, положительная корреляция; в гл. 3 будет показано, как проверить существенность такой зависимости. Однако для нашего примера это не столь уж важно. Предположим, что после того, как коэффициенты $\tau$ и $\rho$ были уже вычислены, получены еще два ответа, в которых указаны нормы дивиденда 7 и 23%. Чтобы включить эти числа в уже упорядоченную последовательность, нужно изменить величины почти всех рангов в строке $C$. Кроме этого, надо заново рассчитать значения разностей $d$ и сумму $S\,(d^2)$. А если ответы будут поступать и в последующий период, то всю эту работу придется проделывать еще раз.

Однако можно без особого труда учесть дополнительное влияние на $S$, оказываемое прибавлением двух новых элементов, если прибегнуть к следующему ходу рассуждений. Новый элемент последовательности $B$, равный 7% и имеющий ранг 16 в последовательности $A$, может оказать влияние на корреляцию рангов лишь в силу того, что соответствующие ему числа вносят определенные изменения в порядок, образуемый остальными пятнадцатью парами. Норма дивиденда характеризуется наименьшей величиной, следовательно, новый элемент добавит в сумму $S$ слагаемое, равное $-15$. Аналогичным образом новый элемент, представленный нормой дивиденда в 23%, вносит в сумму $S$ слагаемое, равное 14. Следовательно, вновь полученная сумма $S$ будет на $14 - 15 = -1$ превосходить старую, т. е. составит 24. Теперь можно вычислить новое значение $\tau$:

$$\tau = \frac{24}{136} = +0{,}18.$$

Используя эту методику, можно рассчитать новые значения $\tau$, не прибегая на каждом этапе к повторному упорядочению всего массива данных.

Следует отметить, что в этом примере ранги, образующие последовательность $C$, определяются случайной величиной, указанной в ответах нормой дивиденда.

В то же время ранги, характеризующие очередность получения ответов, не «опираются» на какую-то особую переменную величину (хотя, если мы запасемся терпением и будем в каждом случае измерять продолжительность времени до момента поступления каждого последующего ответа, можно считать, что ранги, которые будут указаны в строке $A$, определяются упорядочением соответствующих промежутков времени).

**1.25.** В заключение рассмотрим три примера, иллюстрирующие возможности использования корреляции рангов.

*Пример 1.2*

Имеется 12 одинаковых по размеру дисков, окраска которых отличается тоном — от светло-голубого до темно-синего. С помощью колориметрического испытания можно получить объективную оценку интенсивности цвета. Для того чтобы оценить, как тонко модельер одежды различает цветовые оттенки, ему показывают все эти диски и предлагают расположить их в определенном порядке — по степени интенсивности цвета. При этом получают, скажем, следующие результаты:

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Порядок дисков, основанный на объективных оценках | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Порядок, основанный на оценках модельера | 1 | 4 | 7 | 2 | 3 | 5 | 8 | 12 | 10 | 6 | 11 | 9 |

С помощью корреляции рангов мы стремимся дать количественную характеристику способности модельера различать оттенки синего цвета.

Найдем сначала значение $P$:

$$11+8+5+8+7+6+4+0+1+2+0=52,$$

$$\tau=\frac{104}{66}-1=+0{,}58.$$

Налицо положительная корреляция рангов, степень соответствия довольно велика, но все же далека от полной эквивалентности. В гл. 3 мы покажем, как проверить существенность исчисленных коэффициентов.

В этом примере измерялась степень согласованности порядка, установленного на основе объективных оценок, с порядком, определенным на основе субъективного выбора. Модельер не сумел достичь полного успеха, что может объясняться его неумением различать малозаметные оттенки, либо отсутствием сосредоточенности, либо какими-то другими факторами; однако, какова бы ни была действительная причина, в любом случае мы можем проверить, насколько субъективные оценки модельера отличаются от заранее заданных объективных оценок.

*Пример 1.3*

Рассмотрим ситуацию, при которой некоторый ряд участниц конкурса красоты должен быть упорядочен тремя членами жюри. Их оценки распределились следующим образом:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Член жюри $A$: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Член жюри $B$: | 5 | 4 | 1 | 7 | 2 | 8 | 3 | 6 | 9 |
| Член жюри $C$: | 2 | 5 | 1 | 3 | 4 | 7 | 6 | 9 | 8 |

В отличие от условий предыдущего примера в этом случае не существует объективных оценок. Нас интересует вопрос, в какой степени различаются между собой мнения членов жюри; в этом случае уже не может возникнуть проблема отклонения субъективных оценок от некоторых стандартных объективных значений.

Выясним парную корреляционную зависимость между оценками членов жюри:

$$\tau\ (A \text{ и } B) = 0{,}33,$$

$$\tau\ (B \text{ и } C) = 0{,}44,$$

$$\tau\ (C \text{ и } A) = 0{,}67.$$

Отсюда следует, что мнения членов жюри $A$ и $C$ в большей степени совпадают между собой, чем суждения $A$ и $B$ или $B$ и $C$. Привлекает внимание сравнительно слабая согласованность оценок $A$ и $B$.

*Пример 1.4*

В табл. 1.1 приведены данные об обороте внешней торговли (импорт плюс экспорт) и численности населения некоторых государств в 1938 г. В соответствующих столбцах указаны оценки рангов, полученные в результате упорядочения стран по этим двум признакам.

Значение $P$ равно 72, следовательно, $\tau = +0{,}20$. По-видимому, это число довольно точно отражает общую зависимость между переменными. В целом для государства с бóльшим населением характерен бóльший объем внешнеторговых операций; однако Китай представляет собой исключение, поэтому утверждение о положительной корреляционной зависимости нельзя считать строгим.

Таблица 1.1. **Внешняя торговля и численность населения некоторых государств в 1938 г.**

| Государства | Торговля (импорт плюс экспорт; млрд. ф. ст.) | Ранг, определенный по объему торговли | Население (млн. человек) | Ранг, определенный по численности населения |
|---|---|---|---|---|
| Соединенное Королевство | 1,330 | 1 | 47,6 | 5 |
| США | 1,024 | 2 | 130,0 | 2 |
| Дания | 0,142 | 11 | 3,8 | 15 |
| Франция | 0,450 | 4 | 42,0 | 8 |
| Германия | 0,882 | 3 | 79,2 | 3 |
| Греция | 0,045 | 16 | 7,1 | 13 |
| Голландия | 0.276 | 7 | 8,7 | 11 |
| Италия | 0,232 | 8 | 43,4 | 7 |
| Япония | 0,309 | 5 | 72,8 | 4 |
| Норвегия | 0,098 | 13 | 2,9 | 16 |
| Испания | 0,052 | 15 | 25,6 | 9 |
| Швеция | 0,201 | 9 | 6,3 | 14 |
| Аргентина | 0,180 | 10 | 13,0 | 10 |
| Бельгия | 0,307 | 6 | 8,4 | 12 |
| Бразилия | 0,121 | 12 | 44,1 | 6 |
| Китай | 0,085 | 14 | 410,0 | 1 |

При анализе такого рода данных часто встречается ситуация, когда показатели очень сильно отличаются друг от друга, например население Норвегии составляет 2,9 млн. человек, а население Китая — 410 млн. Следует иметь в виду, что при значительном различии исходных величин присутствие одной-двух переменных, характеризующихся большими значениями, может существенно исказить общую картину, поскольку колебания этих переменных могут просто «перекрыть» колебания многих малых величин.

Ранжируя отдельные элементы, мы устанавливаем какой-то более приемлемый порядок, когда каждому государству отводится то или иное место в зависимости от размеров этой страны. Правильность подобного подхода зависит от предмета исследования; следует подчеркнуть, однако, что бывают ситуации, когда использование самих переменных величин, казалось бы, обеспечивает более точные результаты, и все же абсолютные величины могут в большей степени искажать картину, чем ранговые оценки, поскольку на самом деле такие абсолютные величины менее пригодны для описания зависимостей, которые мы стремимся измерить.

Для читателя, знакомого с описываемой в обычном курсе статистики методикой расчета смешанных корреляционных моментов, добавим, что коэффициент корреляции между приведенными в табл. 1.1 величинами—оборотом торговли и численностью населения — равен 0,006. Включение в расчет стран, располагающих огромным населением, таких, как Китай, приводит к тому, что средняя теснота корреляционной взаимосвязи между объемом торговли и населением практически сводится к нулю.

## Библиография

Методика расчетов коэффициентов $\tau$ при оценке корреляционных зависимостей между случайными величинами, характеризующимися нормальным распределением, описана в [33] и [24], см. гл. 9 и 10. Независимо от этих авторов коэффициент $\tau$ использовал Кендэл [48], видевший в нем только характеристику степени соответствия ранговых оценок. Способы применения коэффициента $\tau$ при измерении степени неупорядоченности описаны в [36], [26] и [65].

Методы использования коэффициента $\rho$ рассматриваются в работах [88 и 89], [75], [55].

Вывод неравенств, связывающих коэффициенты $\rho$ и $\tau$, можно найти в [11 и 12], а также [20].

## *ГЛАВА 2.* ВВЕДЕНИЕ В ОБЩУЮ ТЕОРИЮ РАНГОВОЙ КОРРЕЛЯЦИИ

**2.1.** В этой главе мы перейдем к изложению общей теории корреляции рангов и выведем некоторые результаты, которые в предыдущей главе излагались без доказательства. Если читатель интересуется лишь практическими приложениями и готов принять на веру приведенные выше результаты, он может просто опустить эту главу. Если же читатель знаком с теорией корреляции[1], ему, быть может, стоит просмотреть эту главу, чтобы узнать, как можно с помощью единой теории связать различные коэффициенты, используемые обычно в корреляционном анализе.

### Обобщенный коэффициент корреляции

**2.2.** Пусть дана совокупность, включающая $n$ объектов; при этом рассматриваются два свойства этих объектов: $x$ и $y$. Для того чтобы иметь возможность идентифицировать любой желаемый порядок объектов, перенумеруем их от 1 до $n$; в таком случае можно утверждать, что признак $x$ будет принимать значения $x_1, \ldots, x_n$, а признак $y$ — значения $y_1, \ldots, y_n$. Эти значения могут представлять собой абсолютные величины либо ранговые оценки.

Каждой паре элементов, скажем $i$ и $j$, мы будем приписывать $x$-оценку (назовем ее $a_{ij}$), которая обладает следующим свойством: $a_{ij} = -a_{ji}$. Аналогичным образом введем $y$-оценку, используя для этого символ $b_{ij}$, причем $b_{ij} = -b_{ji}$. Знак $\Sigma$ будет означать суммирование по всем значениям $i$ и $j$ от 1 до $n$. В этом случае обобщенный коэффициент корреляции $\Gamma$ можно определить следующим образом:

$$\Gamma = \frac{\Sigma a_{ij} b_{ij}}{\sqrt{\Sigma a_{ij}^2 \, \Sigma b_{ij}^2}}. \tag{2.1}$$

Будем полагать, что $a_{ij}$ принимает нулевое значение в тех случаях, когда $i = j$.

---

[1] Теория статистической корреляции излагается в книге G. Udny Yule and M. G. Kendall. An Introduction to the Theory of Statistics, 14th ed. (Русский перевод: Юл Д. Э., Кендэл М. Дж. Теория статистики. М., Госстатиздат, 1960. — *Прим. ред.*)

## $\tau$ как частный случай обобщенного коэффициента корреляции

2.3. Такое определение обобщенного коэффициента распространяется на индексы $\tau$, $\rho$ и коэффициент корреляции $r$, причем в каждом из этих случаев меняется лишь содержание оценок $a_{ij}$ и $b_{ij}$.

Предположим, например, что мы приписываем паре значений $x$ и $y$ оценку $+1$, если $p_j > p_i$ ($p_i$ представляет собой ранг $i$-го элемента, определенный по признаку $x$), и оценку $-1$, если $p_j < p_i$.
Тогда

$$\left.\begin{aligned} a_{ij} &= +1 \quad \text{при} \quad p_i < p_j \\ a_{ij} &= -1 \quad \text{при} \quad p_i > p_j \end{aligned}\right\}. \tag{2.2}$$

Аналогичные соотношения будут выражать и оценки $b$. В таком случае сумма $\Sigma a_{ij}b_{ij}$ окажется равной удвоенной сумме $S$ (удвоенной, поскольку при суммировании каждая пара будет встречаться дважды: один раз в сочетании $ij$, а другой—в сочетании $ji$). Далее, сумма $\Sigma a_{ij}^2$ оказывается равной числу слагаемых $a_{ij}$, т. е. она составляет $n$ $(n-1)$. То же относится и к сумме $\Sigma b_{ij}^2$. Подставив эти результаты в выражение (2.1), можно убедиться, что $\Gamma$ совпадает с коэффициентом $\tau$, который был определен в гл. 1.

## $\rho$ как частный случай обобщенного коэффициента корреляции

2.4. Вместо того чтобы приписать $a_{ij}$ значения $\pm 1$, положим

$$a_{ij} = p_j - p_i, \tag{2.3}$$

аналогичным образом определим и оценки $b_{ij}$:

$$b_{ij} = q_j - q_i, \tag{2.4}$$

где $q_i$ — ранг $i$-го элемента, определенный по признаку $y$. Величины $p_i$ и $q_i$ принимают значения от 1 до $n$, следовательно, суммы квадратов $\Sigma\,(p_j - p_i)^2$ и $\Sigma\,(q_j - q_i)^2$ равны между собой. Подставив эти суммы в выражение (2.1), мы получаем:

$$\Gamma = \frac{\Sigma\,(p_j - p_i)(q_j - q_i)}{\Sigma\,(p_j - p_i)^2}. \tag{2.5}$$

Рассмотрим числитель этого выражения:

$$\sum_{i,\,j=1}^{n} (p_j - p_i)(q_j - q_i) = \sum_{i=1}^{n}\sum_{j=1}^{n} p_i q_i + \sum_{i=1}^{n}\sum_{j=1}^{n} p_j q_j -$$
$$-\sum_{i=1}^{n}\sum_{j=1}^{n} (p_i q_j + p_j q_i) = 2n \sum_{i=1}^{n} p_i q_i - 2 \sum_{i=1}^{n} p_i \sum_{j=1}^{n} q_j =$$
$$= 2n \sum_{i=1}^{n} p_i q_i - \frac{1}{2} n^2 (n+1)^2, \tag{2.6}$$

так как $\Sigma p_i$ и $\Sigma q_i$ равны сумме первых $n$ элементов натурального ряда, т. е. составляют $^1/_2\, n\,(n+1)$.

Кроме того,

$$S(d^2)=\sum_{i=1}^{n}(p_i-q_i)^2=2\Sigma p_i^2-2\Sigma p_i q_i \tag{2.7}$$

и, следовательно, воспользовавшись выражением (2.6), можно написать:

$$\Sigma(p_j-p_i)(q_j-q_i)=2n\,\Sigma p_i^2-\frac{1}{2}n^2(n+1)^2+nS(d^2). \tag{2.8}$$

Но ведь $\Sigma p_i^2$ представляет собой сумму квадратов первых $n$ членов натурального ряда; эта сумма равна $\frac{1}{6}n(n+1)(2n+1)$. Подставив указанную величину в правую часть выражения (2.8), можно привести ее к следующему виду:

$$\frac{1}{6}n^2(n^2-1)-nS(d^2). \tag{2.9}$$

Далее замечаем, что

$$\begin{aligned}\Sigma(p_j-p_i)^2&=2n\,\Sigma p_i^2-2\,\Sigma p_i\,p_j=\\&=2n\,\Sigma p_i^2-2\,(\Sigma p_i)^2=\\&=\frac{1}{6}n^2(n^2-1).\end{aligned} \tag{2.10}$$

Подставив полученные значения (2.9) и (2.10) в выражение (2.5), мы находим, что

$$\Gamma=1-\frac{6S(d^2)}{n^3-n}. \tag{2.11}$$

Таким образом, Г в этом случае принимает значение коэффициента Спирмэна $\rho$.

## Коэффициент парной корреляции как частный случай обобщенного коэффициента корреляции

2.5. На сей раз предположим, что наши оценки основываются на значениях, которые в действительности принимали рассматриваемые переменные величины, и запишем

$$\left.\begin{aligned}a_{ij}&=x_j-x_i\\b_{ij}&=y_j-y_i\end{aligned}\right\}. \tag{2.12}$$

В таком случае

$$\frac{1}{2}\sum_{i,\,j}(x_j-x_i)(y_j-y_i)=n\sum_i x_i\,y_i-\sum x_i\,y_j. \tag{2.13}$$

$$\frac{1}{2}\sum_{i,\,j}(x_j-x_i)^2=n\sum x_i^2-(\sum x_i)^2. \tag{2.14}$$

Правая часть выражения (2.13) представляет собой не что иное, как ковариацию $x$ и $y$, умноженную на $n$, а правая часть выражения (2.14) — умноженную на $n$ дисперсию $x$[1].

Подставляя эти значения в выражение (2.1), получаем:

$$\Gamma = \frac{\operatorname{cov}(x, y)}{\sqrt{(\operatorname{var} x \operatorname{var} y)}}. \tag{2.15}$$

В этом случае Г представляет собой обычный коэффициент корреляции между $y$ и $x$.

**2.6.** Из предыдущего параграфа следует: если предположить, что ранги представляют собой случайные величины, то величину $\rho$ можно рассматривать как коэффициент корреляции между ними. Обратимся теперь к непосредственной проверке этого утверждения.

Выше уже было показано, что сумма первых $n$ членов натурального ряда может быть представлена следующим образом:

$$\Sigma p_i = \frac{1}{2} n(n+1)$$

и, следовательно, первый момент ( т. е. арифметическая средняя) будет составлять:

$$\mu_1' = \frac{1}{2}(n+1). \tag{2.16}$$

Аналогично находим:

$$\Sigma p_i^2 = \frac{1}{6} n(n+1)(2n+1)$$

и, таким образом, величина дисперсии может быть найдена из следующего соотношения:

$$\mu_2 = \frac{1}{n}\Sigma p_i^2 - \mu_1'^2 = $$

$$= \frac{1}{12}(n^2 - 1). \tag{2.17}$$

Пользуясь соотношениями (2.7) и (2.10), можно найти значение $^1/_2\, S(d^2)$:

$$\frac{1}{2} S(d^2) = \frac{1}{12} n(n^2 - 1) - n\left[\frac{1}{n}\Sigma p_i q_i - \frac{1}{n}\Sigma p_i \frac{1}{n}\Sigma q_i\right].$$

Первый смешанный момент (момент произведения) записан в правой части в квадратных скобках, он равен

$$\mu_{11} = \frac{1}{12}(n^2 - 1) - \frac{1}{2n} S(d^2).$$

---

[1] Соответствующие определения приведены в 3.10.

Исходя из этого, определяем коэффициент корреляции:

$$\frac{\mu_{11}}{\sqrt{[\mu_2(x)\,\mu_2(y)]}} = 1 - \frac{6S(d^2)}{n^3 - n} = \rho.$$

2.7. Таким образом, из соотношения (2.1), характеризующего обобщенный коэффициент корреляции Г, можно вывести как частные случаи коэффициенты $\tau$, $\rho$ и $r$, причем переход к тому или иному коэффициенту зависит от выбора оценок, с помощью которых мы измеряем различия между рассматриваемыми объектами. Наиболее просто определяется величина коэффициента $\tau$; при этом каждой паре значений рассматриваемых переменных приписывается оценка, равная единице, независимо от того, насколько удалены друг от друга ранжируемые элементы. Несколько более сложны оценки, используемые при расчете коэффициента $\rho$; при этом разностям между элементами приписываются веса, которые увеличиваются по мере отдаления элементов друг от друга (т. е. по мере того как возрастает число объектов, заключенных в промежутке между двумя элементами). Оценки, используемые при расчете коэффициента $r$, должны характеризовать абсолютную величину различий между значениями соответствующих переменных, измеряя эти разности с помощью какой-то объективной шкалы (если таковая существует). Какой из этих методов предпочесть? А может быть следует применить и другие пригодные методы? Ответ на эти вопросы зависит от конкретных условий.

2.8. В 1.8 отметили три свойства коэффициента корреляции рангов, которые существенно облегчают его использование. Легко видеть, что и коэффициент Г принимает значения от $+1$ до $-1$; это следует из неравенства Коши—Шварца:

$$(\Sigma ab)^2 \leqslant \Sigma a^2\, \Sigma b^2.$$

Покажем теперь, что при некоторых условиях коэффициент Г обладает также третьим свойством, упомянутым в 1.8. Действительно, пусть порядок двух соответствующих пар рангов в двух последовательностях не совпадает между собой. Поменяем местами элементы одной из пар (тем самым порядок будет восстановлен); в таком случае коэффициент Г увеличится, если только: 1) оценки $a_{ij}$ и $b_{ij}$ отличны от нуля; 2) величина этих оценок не уменьшается по мере удаления друг от друга элементов последовательности. Эти условия должны соблюдаться и при исчислении коэффициентов $\tau$, $\rho$ и коэффициента корреляции $r$.

Используем введенную ранее систему обозначений. При этом будем полагать, что ранговая оценка $r$-го элемента $p_r$ больше ранга $s$-го элемента $p_s$; допустим также, что $q_r$ представляют собой соответствующие ранги в другой последовательности. Символом $\Sigma''$ обозначим результат суммирования по всем $i$ и $j$, за исключением $r$ и $s$.

Теперь приступим к вычислениям:

$$\begin{aligned}\Sigma a_{ij} b_{ij} = &\Sigma'' a_{ij} b_{ij} + \Sigma'' a_{rj} b_{rj} + \\ &+ \Sigma'' a_{ir} b_{ir} + \Sigma'' a_{sj} b_{sj} + \\ &+ \Sigma'' a_{is} b_{is} + a_{rs} b_{rs} + a_{sr} b_{sr} = \\ &= \Sigma'' a_{ij} b_{ij} + 2\Sigma'' a_{rj} b_{rj} + \\ &+ 2\Sigma'' a_{sj} b_{sj} + 2a_{rs} b_{rs}. \qquad (2.18)\end{aligned}$$

Поменяем местами $p_r$ и $p_s$. В результате $\Sigma a_{ij} b_{ij}$ принимает следующий вид:

$$\Sigma a_{ij} b_{ij} = \Sigma'' a_{ij} b_{ij} + 2\Sigma'' a_{sj} b_{rj} + 2\Sigma'' a_{rj} b_{sj} - 2a_{rs} b_{rs}. \qquad (2.19)$$

Приращение суммы характеризуется положительной величиной; оно равно:

$$-2\Sigma'' (a_{sj} - a_{rj})(b_{sj} - b_{rj}) - 4a_{rs} b_{rs} = -2\Sigma (a_{sj} - a_{rj})(b_{sj} - b_{rj}). \qquad (2.20)$$

В таком случае если $p_j > p_r$ (и следовательно, $> p_s$), то $a_{sj} > 0$, $a_{rj} \geqslant 0$ и в силу условий, налагаемых на поведение оценок, $a_{sj} \geqslant \geqslant a_{rj}$. Если же $p_r > p_j \geqslant p_s$, то $a_{sj} \geqslant 0$, $a_{rj} \leqslant 0$ и, следовательно, $a_{sj} - a_{rj} = a_{sj} + a_{jr} > 0$. Когда $p_r > p_s > p_j$, $a_{sj} < 0$, $a_{rj} < 0$ и $a_{jr} \geqslant a_{js}$, так что $a_{sj} - a_{rj} \geqslant 0$. Таким образом, во всех случаях $a_{sj} - a_{rj} \geqslant 0$ и по меньшей мере в одном случае неравенство оказывается строгим, т. е. $a_{sj} - a_{rj} > 0$.

Аналогично можно показать, что $b_{sj} - b_{rj} \leqslant 0$ и по крайней мере в одном случае это неравенство оказывается строгим, т. е. $b_{sj} - -b_{rj} < 0$. Следовательно, сумма, записанная в правой части выражения (2.20), представляет собой отрицательную величину, поэтому приращение должно быть положительным, что и требовалось доказать.

**2.9.** К подобному подходу мы будем часто прибегать в последующих главах и особенно в тех случаях, когда будут рассматриваться проблемы выборочных показателей. А сейчас приведем доказательство следующих двух положений, содержащихся в гл. 1:

а) в 1.13 утверждалось, что минимальное число перестановок, необходимых для того, чтобы перейти от одной последовательности рангов к другой, определяется только величиной выражения $S$;

б) в 1.19 утверждалось, что значения $\rho$, полученные при коррелировании последовательности $A$ с последовательностью $B$ и с сопряженной с ней последовательностью $B'$, равны по величине и противоположны по знаку.

**2.10.** Вначале докажем второе утверждение. При этом достаточно рассмотреть случай, когда последовательность $A$, содержащая элементы $p_i$, сопоставляется с порядком натурального ряда $B = 1, \ldots, n$ и с порядком, обратным натуральному, $B' = n, \ldots, 1$. Исчисляемую при корреляции между $A$ и $B'$ величину $S$ $(d^2)$ мы назовем $S'$ $(d^2)$.

Запишем оба выражения:

$$S(d^2)=\Sigma(p_i-i)^2=\Sigma p_i^2+\Sigma i^2-2\Sigma ip_i=$$
$$=\frac{1}{3}n(n+1)(2n+1)-2\Sigma ip_i.$$
$$S'(d^2)=\Sigma[p_i-(n+1-i)]^2=$$
$$=\frac{1}{3}n(n+1)(2n+1)-2\Sigma(n+1-i)p_i.$$

Из этого следует, что

$$S(d^2)+S'(d^2)=\frac{2}{3}n(n+1)(2n+1)-2\Sigma(n+1)p_i=$$
$$=\frac{2}{3}n(n+1)(2n+1)-n(n+1)^2=$$
$$=\frac{1}{3}(n^3-n).$$

Таким образом,

$$\rho+\rho'=2-\frac{6}{n^3-n}[S(d^2)+S'(d^2)]=0,$$

что и требовалось доказать.

**2.11.** Теперь перейдем к доказательству того, что число перестановок $s$ можно определить следующим образом:

$$s=\frac{1}{4}n(n-1)-\frac{1}{2}S. \tag{2.21}$$

Для этого сначала покажем, что $s$ не превосходит величины, записанной в правой части выражения (2.21), а затем, что $S$ не может быть меньше ее. Из этих двух утверждений следует справедливость равенства (2.21).

Без потери общности можно предположить, что одна из последовательностей характеризуется натуральным порядком 1,2. ..., $n$; ведь нас интересуют только перестановки, и мы всегда можем перенумеровать элементы одной из последовательностей (скажем, второй) в натуральном порядке. Пусть элементу с номером $i$, содержащемуся во второй последовательности, соответствует элемент первой последовательности, имеющий ранг $p_i$; рассмотрим теперь процесс переупорядочения элементов первой последовательности.

Введем следующую функцию:

$$\left.\begin{aligned} m_{ij}&=1, \quad \text{если} \quad p_i>p_j \\ &=0, \quad \text{если} \quad p_i<p_j \end{aligned}\right\}. \tag{2.22}$$

Если ранг элемента равен $p_1$, этот объект можно переставить на первое место (в крайнее левое положение) с помощью $p_1 - 1$ перестановок. При этом элемент, имеющий ранг $p_2$, передвинется вправо на $m_{12}$ мест (позиций). Для того чтобы переместить этот элемент на второе место, потребуется $p_2 - 2 + m_{12}$ перестановок. Аналогично перемещение элемента, имеющего ранг $p_i$, на $i$-е место потребует

$$p_i - i + m_{1i} + m_{2i} + \ldots + m_{i-1,\, i}$$

перестановок. Сложив все эти величины, можно отыскать общее число перестановок:

$$\sum_{i=1}^{n} p_i - \sum_{i=1}^{n} i + \sum_{i<j} m_{ij} = \sum_{i<j} m_{ij}, \tag{2.23}$$

где знак $\sum\limits_{i<j}$ означает суммирование по всем $i < j$. Тогда величину можно определить из следующего выражения:

$$S = \sum_{i<j} (1 - 2m_{ij}), \tag{2.24}$$

поскольку каждый единичный вклад $m_{ij}$ входит в сумму со знаком плюс в том случае, если $p_i < p_j$, и со знаком минус, если $p_i > p_j$; причем в вычислениях участвуют только такие пары элементов, у которых $i < j$. Это делается для того, чтобы не учитывать дважды одни и те же пары. Таким образом, число перестановок, определяемое соотношением (2.23), составляет:

$$\frac{1}{2} \sum_{i<j} (1) - \frac{1}{2} S = \frac{1}{4} n(n-1) - \frac{1}{2} S,$$

потому что сумма единичных значений $m_{ij}$, удовлетворяющих условию $i < j$, равна половине общего числа членов, т. е. $\sum\limits_{i<j} (1) = {}^1/_2\, n\,(n-1)$. Следовательно, минимальное число перестановок $s$ определяется следующим неравенством:

$$s \leqslant \frac{1}{4} n(n-1) - \frac{1}{2} S. \tag{2.25}$$

Обозначим буквой $T$ общую последовательность перестановок, с помощью которых удается перейти от заданного порядка элементов к натуральному порядку. Разобьем $T$ на $n$ групп: $T_1, T_2, \ldots, T_n$, причем $T_1$ будет включать все перестановки элемента 1; в $T_2$ входят перестановки элемента 2 (но не входят перестановки элемента 1) и так далее, а $T_n$ — это пустое множество.

В каждой группе $T_i$ будем различать число перестановок, в результате которых данный элемент передвигается вправо; это число обозначим $A_i$, а число перестановок, в результате которых этот элемент перемещается влево, — $B$. Самое меньшее число перестановок, которое может содержаться в $T_1$, равно $P_1 - 1$. Но эти же перестановки сдвинут объект 2 на $m_{12}$ позиций вправо. В результате всех перемещений, входящих в состав $T_2$, элемент 2 передвинется на $A_2 - B_2$ позиций вправо, так что в совокупности он сместится на

$$m_{12} + A_2 - B_2$$

позиций; вместе с тем общее число перемещений элемента 2 должно составлять $2 - p_2$.

Следовательно,

$$B_2 - A_2 = p_2 - 2 + m_{12},$$
$$B_2 + A_2 \geqslant p_2 - 2 + m_{12}.$$

Аналогично для $i$-го элемента можно вывести следующее соотношение:

$$B_i + A_i \geqslant p_i - i + m_{1i} + \ldots + m_{i-1,\, i}.$$

Просуммируем эти неравенства по $i$ ($i = 1, \ldots, n - 1$). Принимая во внимание, что

$$p_n - n + m_{1n} + \ldots + m_{n-1,\, n} = 0,$$

можно вывести следующее соотношение:

$$s = \sum_{i=1}^{n-1} (A_i + B_i) \geqslant \Sigma p_i - \Sigma i + \sum_{i<j} m_{ij} \geqslant \frac{1}{4} n(n-1) - \frac{1}{2} S, \quad (2.26)$$

а объединив выражения (2.25) и (2.26), можно убедиться в справедливости равенства (2.21).

**2.12.** Теперь мы перейдем к доказательству соотношений (1.12) и (1.13), характеризующих $\rho$ как коэффициент взвешенных инверсий, и обобщим полученные результаты. Без потери общности можно предположить, что одна из последовательностей рангов характеризуется натуральным порядком $1, 2, \ldots, n$. Воспользуемся определением $m_{ij}$, заданным соотношением (2.22). В таком случае можно записать:

$$V = \sum_{i<j} m_{ij} (j - i). \quad (2.27)$$

Прибавим к этому выражению $\sum_{i>j} jm_{ij}$ и вычтем ту же величину; тогда

$$V = \sum_{i<j} jm_{ij} + \sum_{i>j} jm_{ij} - \sum_{i<j} im_{ij} - \sum_{i>j} jm_{ij} =$$
$$= \sum_{i,j} jm_{ij} - \sum_{i<j} i(m_{ij} + m_{ji}).$$

Символом $p_j$ будем обозначать ранг элемента первой последовательности, соответствующего $j$-му элементу второй последовательности; в этом случае, суммируя по $i$ слагаемые в первой сумме и по $j$ — слагаемые во второй сумме, можно вывести следующее соотношение

$$V = \sum_j j(n - p_j) - \sum_i i(n - i) = \sum_i i^2 - \sum_i ip_i. \quad (2.28)$$

Как и в выражении (2.7), величина $S$ $(d^2)$ будет определяться соотношением

$$S(d^2) = 2\Sigma i^2 - 2\Sigma ip_i$$

и, следовательно,

$$V=\frac{1}{2}S(d^2). \tag{2.29}$$

Но из этого непосредственно следует, что

$$\rho=1-\frac{12V}{n^3-n}. \tag{2.30}$$

**2.13.** В некоторых случаях пользуются другим способом исчисления $\rho$, хотя он менее удобен с точки зрения вычислений. Рассмотрим этот способ. Обозначим символами $a_{ij}$ и $b_{ij}$ единичные оценки, используемые при расчете обобщенного коэффициента корреляции Г. Введем также промежуточные суммы этих оценок:

$$a_i=\sum_{j=1}^{n}a_{ij},$$

$$b_i=\sum_{j=1}^{n}b_{ij}.$$

В таком случае

$$\rho=\frac{\sum\limits_{i=1}^{n}a_i b_i}{\frac{1}{3}(n^3-n)}. \tag{2.31}$$

Докажем это утверждение. Оценки $a_{ij}$ можно выразить через $m_{ij}$ с помощью следующего соотношения:

$$a_{ij}=1-2m_{ij},$$

поскольку $a_{ij}$ будет принимать значения $+1$, когда $p_i<p_j$ и $-1$, когда $p_i>p_j$. Вместе с тем ранг любого элемента последовательности можно выразить следующим образом:

$$p_i=\sum_{j=1}^{n}m_{ij}+1. \tag{2.32}$$

Применяя приведенные выше соотношения, можно записать:

$$p_i=-\frac{1}{2}\sum_j a_{ij}+\frac{1}{2}(n+1).$$

Пусть $\bar{p}$ означает среднее значение рангов $p_i$; оно равно $^1/_2\ (n+1)$, следовательно,

$$a_i=-2(p_i-\bar{p}). \tag{2.33}$$

Поскольку $\rho$ представляет собой коэффициент корреляции $p$ и $q$ (см. 2.4), мы можем теперь вывести соотношение (2.31)

$$\rho = \frac{\frac{1}{4}\sum_i a_i b_i}{\frac{1}{12}(n^3-n)} = \frac{\sum_i a_i b_i}{\frac{1}{3}(n^3-n)},$$

что и требовалось доказать.

**2.14.** Отсюда следует, что коэффициент $\rho$ можно представить в следующей форме:

$$\rho = \frac{3}{n^3-n}\Sigma a_{ij} b_{ik}, \tag{2.34}$$

причем суммирование проводится по всем $i$, $j$, $k$. С другой стороны, если провести суммирование при условии $j \neq k$, получим следующее выражение

$$\rho = \frac{3}{n^3-n}\left(\sum a_{ij} b_{ij} + \sum_{j\neq k} a_{ij} b_{ik}\right) =$$
$$= \frac{3}{(n+1)}\tau + \frac{3}{n^3-n}\sum_{j\neq k} a_{ij} b_{ik}. \tag{2.35}$$

Интересно рассмотреть это выражение, задавшись следующим вопросом: в какой мере согласован между собой порядок расположения пар, взятых из двух последовательностей. Если $a_{ij}$ и $b_{ij}$ характеризуются одинаковым знаком, имеет место согласованность первого типа (мы рассматриваем здесь только единичные оценки). Если $a_{ij}$ и $b_{ik}$ характеризуются одинаковым знаком, будем называть это согласованностью второго типа. Если же знаки этих оценок противоположны, имеет место *несогласованность*. В такой ситуации коэффициент $\tau$ просто характеризует удельный вес случаев согласованности первого типа при сравнении двух последовательностей рангов. Общее число возможных случаев согласованности второго типа составляет $n(n-1)(n-2)$, для того чтобы отличать эту величину от $k_1$ — числа случаев согласованности первого типа, равносильного $2P$ в уравнении (1.6), обозначим число случаев согласованности второго типа через $k_2$. Тогда

$$\Sigma a_{ij} b_{ij} = k_1 - [n(n-1) - k_1] = 2k_1 - n(n-1),$$
$$\Sigma a_{ij} b_{ik} = 2k_2 - n(n-1)(n-2).$$

Подставив полученные выражения в (2.35), находим:

$$\rho = \frac{6(k_1+k_2)}{n^3-n} - \frac{3(n-1)}{n+1}. \tag{2.36}$$

**Доказательство неравенства Дэниелса**

**2.15.** Поскольку соотношения типа $p_i > p_j > p_k > p_i$ невозможны, все три оценки $a_{ij}$, $a_{jk}$, $a_{ki}$ не могут быть одного знака, и, следовательно, их сумма должна составлять $\pm 1$. Аналогичное рассуждение можно провести и для $b$-оценок.

Положим:

$$(a_{ij}+a_{jk}+a_{ki})(b_{ij}+b_{jk}+b_{ki})=\varepsilon_{ijk}. \tag{2.37}$$

Теперь переименуем ранги $p_i$, $p_j$, $p_k$, $q_i$, $q_j$, $q_k$ в соответствии с их значениями в новые оценки $p_i'$, $p_j'$. Например, последовательность рангов 4 7 2 трансформируется в последовательность 2 3 1. В таком случае величина $\varepsilon_{ijk}$ будет равна $+1$ или $-1$, в зависимости от того, четна или нечетна перестановка оценок $p'$ относительно $q'$ (другими словами, от того, четным или нечетным окажется число перестановок соседних членов, требующееся для перехода от порядка одной последовательности к порядку другой). Если среди $i$, $j$ и $k$ какие-либо два числа оказываются равными между собой, величина $\varepsilon$ обращается в нуль.

Суммируя по всем значениям подписных индексов, находим:

$$3n\,\Sigma a_{ij}b_{ij}-6\Sigma a_{ij}b_{ik}=\Sigma\varepsilon_{ijk}. \tag{2.38}$$

Но ведь общее число возможных размещений элементов $i$, $j$ и $k$ по три составляет $n(n-1)(n-2)$, а $\Sigma\varepsilon_{ijk}$ представляет собой количество четных размещений. Введем отношение между этими величинами $U$:

$$U=\Sigma\varepsilon_{ijk}/n(n-1)(n-2). \tag{2.39}$$

Понятно, что $-1 \leqslant U \leqslant 1$; тогда из соотношения (2.38) следует:

$$-1\leqslant\frac{3(n+2)}{n-2}\tau-\frac{2(n+1)}{n-2}\rho\leqslant 1. \tag{2.40}$$

Переходя к пределу, мы получаем для больших $n$:

$$-1\leqslant 3\tau-2\rho\leqslant 1. \tag{2.41}$$

Верхний предел достижим при $\tau \geqslant 0$. Действительно, сопоставим между собой последовательность, в которой ранги $q$ расположены в натуральном порядке $1, \ldots, n$, и последовательность рангов $p$, в которой такой порядок регулярно чередуется с порядком, обратным натуральному (например, $m+1, m+2, \ldots, n, 1, 2, \ldots, m$). Тогда все величины $\varepsilon$ будут составлять $+1$, и отношение $U$ примет свое максимальное значение.

**Доказательство неравенства Дарбина—Стюарта**

**2.16.** Задача отыскания достижимого верхнего предела $\rho$ при заданном $\tau \geqslant 0$ эквивалентна задаче определения наименьшего значения $V$ при заданном $Q$. Рассмотрим для примера совокупность, содержащую числа от одного до шести, причем коэффициент $\tau$ составляет

0,6 и, следовательно, $Q = 3$. Если первая последовательность характеризуется натуральным порядком, $V$ принимает наименьшее значение в том случае, когда элементы второй последовательности расположены в следующем порядке: 214365, так как, для того чтобы перейти к новой последовательности, мы осуществляем три инверсии, 21, 43, 65, а вес каждой из этих инверсий принимает минимальное значение, равное единице.

Назовем *компактной* такую систему рангов, элементы которой образуют инверсию между собой и только между собой; ни одну компактную последовательность невозможно разбить на подпоследовательности того же типа. Все ранги, находящиеся слева от компактной системы, меньше, а стоящие справа — больше, чем любой из элементов компактной системы.

Пусть дана инверсия, вес которой равен 2, например $r+2$, $r$. Рассмотрим теперь совокупность, содержащую три числа: $r$, $r+1$ и $r+2$. Эти числа можно расположить тремя различными способами: $r+2$, $r+1$, $r$; $r+2$, $r$, $r+1$ и $r+1$, $r+2, r$. Если последовательность рангов минимальна, т. е. если исчисленная для этой последовательности величина $V$ характеризуется наименьшим значением (при заданном $Q$), то все группы из трех элементов, образующие инверсию с весом 2, должны иметь следующий порядок $r+2$, $r+1$, $r$ (исключение могут составить лишь группы из трех элементов, образующие единую компактную систему). Действительно, предположим, что кроме группы вида $r+2$, $r$, $r+1$ имеется еще отличная от нее группа $x+2$, $x$, $x+1$. Тогда, заменив их последовательностями $r+2$, $r+1$, $r$ и $x+1$, $x$, $x+2$, мы можем уменьшить величину $V$, оставив при этом $Q$ неизменным.

Группу из четырех элементов, содержащую инверсию рангов $r+3$ и $r$, т. е. инверсию с весом 3, можно построить двенадцатью различными способами. Единственный порядок, который не содержит инверсии элементов $r$, $r+1$ и $r+2$ с весом 2 и не совпадает с рассмотренными выше последовательностями типа $r+2$, $r$, $r+1$ или $r+1$, $r+2$, $r$, — это следующий порядок: $r+3$, $r+2$, $r+1$, $r$. Поэтому в минимальной последовательности рангов все группы, содержащие четыре элемента (за исключением, может быть, единой компактной системы), характеризуются порядком $r+3$, $r+2$, $r+1$, $r$.

Продолжая эти рассуждения, можно показать, что минимальная последовательность рангов должна состоять из групп вида $q+r$, $q+r-1, \ldots, r$; исключением может быть только единая компактная система, которую мы будем называть остаточной системой. Именно эта остаточная система и порождает больше всего осложнений в процессе доказательства нашей теоремы. В частности, такая остаточная система может включать всю рассматриваемую последовательность.

**2.17.** Рассмотрим сначала случай, когда остаточной системы не существует. Предположим, что имеется $\alpha_1$ групп вида $r+q_1, \ldots, r$; $\alpha_2$, групп вида $r+q_2, \ldots, r$ и т. д. Тогда

$$\Sigma\alpha_i q_i = n; \tag{2.42}$$

$$Q = \frac{1}{2}\Sigma\alpha_i q_i (q_i - 1) = \frac{1}{2}(\Sigma\alpha_i q_i^2 - n); \tag{2.43}$$

$$V = \frac{1}{6}\Sigma\alpha_i q_i (q_i^2 - 1) = \frac{1}{6}(\Sigma\alpha_i q_i^3 - n). \tag{2.44}$$

В силу неравенства Коши--Шварца

$$(\Sigma\alpha_i q_i)(\Sigma\alpha_i q_i^3) \geqslant (\Sigma\alpha_i q_i^2)^2$$

и воспользовавшись выражениями (2.42) — (2.44), можно записать следующее неравенство:

$$V \geqslant \frac{2}{3} Q\left(1 + \frac{Q}{n}\right). \tag{2.45}$$

В некоторых случаях, когда все $q_i$ равны между собой, приведенное выше выражение обращается в равенство.

**2.18.** Рассмотрим теперь случай, когда существует остаточная система $S$, содержащая $s$ рангов, порядок которой не совпадает с порядком последовательности $I = r+s, r+s-1, \ldots, r$. Предположим, что $S$ содержит $q$ инверсий, совокупный вес которых равен $v$, и что последовательность 1 содержит на $k$ инверсий больше, чем $S$. Тогда мы можем перейти от системы 1 к системе $S$ с помощью $k$ перестановок соседних элементов. Если мы сможем с помощью определенных перестановок разместить эти элементы в таком порядке, который обеспечивает наибольшее сокращение величины $v$, мы найдем требуемую последовательность, которая будет характеризоваться, скажем, $q$ инверсиями и минимальным значением $v$.

Допустим, что исходной последовательностью служит 1; тогда минимальное уменьшение величины $v$, которое можно обеспечить, проделав $k$ перестановок ($k \leqslant s-1$), равно $^1/_2\, k(k+1)$. Это уменьшение может быть достигнуто в результате серии перестановок, последовательно перемещающих наименьший ранг из крайнего правого положения влево. Однако, когда достигается крайняя левая позиция, оказывается нарушенным требование компактности системы $S$, поскольку наименьший ранг теперь сам образует компактную систему, состоящую из одного элемента.

Далее, всякий раз, когда, отправляясь от последовательности 1 как исходной, мы будем в любой последовательности проводить $k$ перестановок (причем $k \geqslant s-1$); полученная в результате этих перестановок система не может служить остаточной системой, образующей часть минимальной последовательности рангов. Дело в том, что, как отмечалось ранее, всегда можно с помощью $s-1$ перестановок продвинуть в крайнее левое положение наименьший ранг и тем самым нарушить условие компактности системы; в результате будет обеспечено то же количество инверсий $q$ при меньшей величине $v$.

Таким образом, в остаточной системе $S$, содержащей $s$ элементов и образующей часть минимальной последовательности ранговых оценок, только один ранг будет занимать место, отличающееся от места в системе 1, и количество перестановок, с помощью которых можно перейти от 1 к $S$, подчинено следующему условию $k \leqslant s - 2$.

**2.19.** Покажем теперь, что неравенство (2.45) может удовлетворяться только в тех случаях, когда система является остаточной. В действительности имеет место строгое неравенство

$$v > \frac{2}{3} q \left(1 + \frac{q}{s}\right) = \frac{2q}{3s}(q+s).$$

Это соотношение справедливо тогда, когда

$$\frac{1}{6} s(s^2-1) - \frac{1}{2} k(k+1) > \frac{2}{3} s\left[\frac{1}{2} s(s-1) - k\right]\left[\frac{1}{2} s(s-1) + s - k\right] >$$
$$> \frac{1}{6} s\,[s^2(s^2-1) - 4s^2k + 4k^2],$$

т. е.

$$k < \frac{s(4s-3)}{3s+4}. \qquad (2.46)$$

Однако, как показано выше, $k \leqslant s - 2$, поэтому соотношение (2.46) будет заведомо справедливо, если

$$s - 2 < \frac{s(4s-3)}{3s+4},$$

т. е. если $s^2 - s + 8 > 0$. Но последнее неравенство удовлетворяется при любых действительных значениях $s$. Таким образом, выражение (2.45) обращается в строгое неравенство только для остаточной системы.

Допустим, что $Q'$ представляет собой общее число инверсий в той части последовательности, которая остается после исключения остаточной системы, а $V'$ — соответствующую взвешенную сумму.

В таком случае, по определению компактной системы,

$$Q = Q' + q,$$
$$V = V' + v.$$

Кроме того, из соотношения (2.45) следует:

$$V' \geqslant \frac{2}{3}\left(\frac{Q'^2}{n-s} + Q'\right)$$

и

$$v > \frac{2}{3}\left(\frac{q^2}{s} + q\right).$$

Покажем, что в общем случае справедливо соотношение (2.45), т. е.

$$V'+v \geqslant \frac{2}{3}\left[\frac{(Q'+q)^2}{n}+Q'+q\right].$$

Это неравенство будет выполняться в том случае, когда

$$\frac{Q'^2}{n-s}+\frac{q^2}{s} \geqslant \frac{(Q'+q)^2}{n},$$

т. е.

$$n\,[sQ'^2+(n-s)\,q^2] \geqslant s\,(n-s)\,(Q'+q)^2,$$

что, в свою очередь, справедливо тогда, когда

$$s^2 Q'^2+(n-s)^2 q^2-2s\,(n-s)\,Q'q \geqslant 0,$$

т. е. $[sQ'-(n-s)\,q]^2 \geqslant 0$.
Убедившись, что последнее неравенство верно, мы тем самым доказываем справедливость соотношения (2.45). Это неравенство может обратиться в равенство только в том случае, когда отсутствует остаточная система.

Отсюда следует и соотношение (1.21). Чтобы найти соответствующее выражение для отрицательных значений коэффициента $\tau$, заменим данную последовательность соответствующих рангов сопряженной. Это сведется просто к изменению знака у коэффициентов $\rho$ и $\tau$.

### Стандартный коэффициент Спирмэна

**2.20.** В заключение коротко рассмотрим коэффициент, основанный не на определении величины $S\,(d^2)$ (как при исчислении коэффициента $\rho$), а на отыскании значения $S\,|d|$, где $|d|$ означает абсолютное значение $d$, т. е. в расчет входит лишь величина $d$ независимо от знака. Этот показатель, который иногда называют стандартным коэффициентом Спирмэна, уже нельзя вывести как частный случай обобщенного коэффициента, заданного соотношением (2.1). Зададим величину коэффициента $R$ следующим образом:

$$R=1-\frac{3S\,|d|}{n^2-1}. \tag{2.47}$$

Если две последовательности рангов идентичны, $S\,|d|=0$ и коэффициент $R$, как мы обычно и предполагаем, равен 1.

Рассмотрим случай, когда одна последовательность представляет собой инверсию другой. Предположим, что $n$ — нечетное число, равное $2m+1$. Тогда, в соответствии с (1.15),

$$S\,|d| = 2\,[2m+(2m-2)+(2m-4)+\ldots+4+2] = 2\,m\,(m+1).$$

Таким образом,

$$R = 1 - \frac{6m(m+1)}{(2m+1)^2 - 1} = -0{,}5. \qquad (2.48)$$

Теперь предположим, что $n$ — четное число, равное 2 $m$. Тогда, рассуждая аналогичным образом, можно показать, что

$$S\,|d| = 2m^2$$

и, следовательно,

$$R = 1 - \frac{6m^2}{4m^2 - 1} = -0{,}5\left(1 + \frac{3}{n^2 - 1}\right). \qquad (2.49)$$

Таким образом, если $n \neq 2$, коэффициент $R$ не может достичь своего минимума, равного —1. При больших четных $n$ значения этого коэффициента сходятся к —0,5, а если величины $n$ нечетны, коэффициент $R$ всегда равен —0,5. Это, конечно, недостаток выбранного нами коэффициента, и мы не сможем полностью устранить его даже в том случае, если попытаемся использовать вместо $3/(n^2 - 1)$ другой множитель, стоящий при $S\,|d|$ в выражении (2.47).

Прибегнем, например, к следующей формуле:

$$R' = 1 - \frac{4S\,|d|}{n^2}.$$

В этом случае коэффициент $R'$ будет равен —1 при четных $n$ и $-1 + 2/n^2$ при нечетных $n$.

**2.21.** Наряду с этим коэффициент $R$ намного менее чувствителен, чем $\tau$ или $\rho$. Выпишем, например, цифры 1, 2, 3, 4 и, рассмотрев все 24 возможные перестановки из этих элементов, сопоставим их с порядком натурального ряда. В таком случае окажется:

| Значения $\rho$ | Значения $R$ | Частоты |
|---|---|---|
| 1,0 | 1,0 | 1 |
| 0,8 | 0,6 | 3 |
| 0,6 | 0,2 | 1 |
| 0,4 | 0,2 | 4 |
| 0,2 | 0,2 | 2 |
| 0,0 | —0,2 | 2 |
| —0,2 | —0,2 | 2 |
| —0,4 | —0,2 | 4 |
| —0,6 | —0,2 | 1 |
| —0,8 | —0,6 | 3 |
| —1,0 | —0,6 | 1 |
| | | 24 |

Коэффициент $\rho$ меняется, например, от 0,2 до 0,6; коэффициент $R$ продолжает сохранять значение, равное 0,2; или же, скажем, $\rho$ может менять значения от 0,0 до —0,6, тогда как величина $R$ остается постоянной и равной — 0,2.

Учитывая все эти обстоятельства и принимая во внимание аналитические трудности, которые возникают при работе с выборками, характеризующими распределение величин $R$, мы можем с полным основанием отказаться от его применения, предпочитая коэффициенты $\tau$ или $\rho$.

## Библиография

Основная работа, посвященная построению обобщенных коэффициентов, принадлежит Дэниелсу [9] (см. также ссылки, приведенные в библиографии к гл. 1). В чрезвычайно важной, но сложной статье [39] рассмотрен весьма общий класс статистических характеристик, который включает как частный случай коэффициенты ранговой корреляции. См. также [25].

Свойство 2.8 описано в [10]. Неравенства, характеризующие коэффициенты $\rho$ и $\tau$, описаны в работах, которые указаны в библиографии к гл. 1.

## *ГЛАВА 3.* СВЯЗАННЫЕ РАНГИ

**3.1.** В практических приложениях методов, основанных на ранжировании, иногда сталкиваются со случаями, когда два или несколько объектов настолько подобны, что не удается отдать предпочтение одному из них. Когда исследователь ранжирует объекты на основе субъективных суждений, то это свойство (отсутствие предпочтений) связано с истинной их неразличимостью или неспособностью исследователя найти существующие различия. В этих случаях говорят, что такие объекты являются *связанными*. Расположение студентов в соответствии с их достоинствами или экзаменационными баллами является известным примером такого рода связей.

**3.2.** Метод, который мы примем для приписывания численных значений рангов связанным объектам, заключается в усреднении рангов, которые они имели бы, если были бы различимы. Например, если исследователь связывает третий и четвертый объекты, то каждому приписывается число $3^1/_2$, а если связываются все объекты от второго до седьмого включительно, то каждый получает ранг $\frac{1}{6}(2+3+4+5+6+$ $+7)=4^1/_2$. Иногда такой подход называют «методом средних рангов». Когда нет основы для выбора между объектами, то ясно, что в этом случае мы должны приписать всем одинаковые ранги, как если бы рассматривали их в виде одного объекта. Преимущество нашего метода состоит в том, что сумма рангов для всех объектов остается точно такой, как и при ранжировании без связей.

**3.3.** Теперь мы должны рассмотреть влияние связей на расчет $\tau$ и $\rho$.

В 1.9 было показано, что парам объектов приписывались значения, равные $+1$ и $-1$, в соответствии с тем, находились ли их ранги в полном соответствии друг с другом или нет. Если они связаны, то мы должны приписать им нулевую оценку, т. е. серединное значение между двумя этими величинами, которых следовало бы ожидать, если бы объекты не были связаны. Теперь легко определить величину $S$.

**3.4.** В связи с расчетом знаменателя, на который должна быть поделена величина $S$ при определении $\tau$, возникает новая проблема. При этом существуют две возможности:

а) взять в качестве знаменателя $^1/_2\, n\,(n-1)$, как и для несвязанной формы $\tau$:

б) заменить $^1/_2\, n\,(n-1)$ на $^1/_2\sqrt{\Sigma a_{ij}^2 \Sigma b_{ij}^2}$, где $a_{ij}$ — оценка объектов $i$ и $j$ в одном ранжировании и $b_{ij}$ — корреспондирующая оценка в другом.

При отсутствии связей любое $a_{ij}^2$ равно единице, так что $\Sigma a_{ij}^2$ равно числу возможных оценок, а именно $n\,(n-1)$; аналогично положение для $\Sigma b_{ij}^2$, так что выражение $^1/_2\sqrt{\Sigma a_{ij}^2 \Sigma b_{ij}^2}$ сводится к $^1/_2\, n(n-1)$, как это и должно быть. Причина выбора этого выражения для случая, когда имеется связь, станет ясной, если вернуться к 2.2 и 2.3.

Если наблюдается связь $t$ последовательных членов, то все оценки, относящиеся к любой выбранной из них паре, равны нулю. Таких пар насчитывается $t\,(t-1)$. Соответственно, сумма $\Sigma a_{ij}^2$ равна $n\,(n-1) - \sum_t t\,(t-1)$, где суммирование производится только для различных комбинаций связей. Поскольку $n\,(n-1)$ равно сумме, которую получили бы, если не было бы связей, то эта величина должна быть уменьшена в связи с тем, что связи дают нулевые оценки. Поэтому запишем:

$$T = \frac{1}{2}\sum_t t\,(t-1) \tag{3.1}$$

для связей в одной последовательности рангов и

$$U = \frac{1}{2}\sum_u u\,(u-1) \tag{3.2}$$

для связей в другой последовательности. Теперь альтернативная форма коэффициента $\tau$ для связанных рангов может быть записана следующим образом:

$$\tau = \frac{S}{\sqrt{\frac{1}{2}n\,(n-1)-T}\;\sqrt{\frac{1}{2}n\,(n-1)-U}}\,. \tag{3.3}$$

До обсуждения других альтернативных форм рассмотрим простой пример.

*Пример 3.1*

Пусть заданы две последовательности рангов:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $A$ | 1 | $2\frac{1}{2}$ | $2\frac{1}{2}$ | $4\frac{1}{2}$ | $4\frac{1}{2}$ | $6\frac{1}{2}$ | $6\frac{1}{2}$ | 8 | $9\frac{1}{2}$ | $9\frac{1}{2}$ |
| $B$ | 1 | 2 | $4\frac{1}{2}$ | $4\frac{1}{2}$ | $4\frac{1}{2}$ | $4\frac{1}{2}$ | 8 | 8 | 8 | 10 |

Если не принимать во внимание связи, то обе последовательности имеют один и тот же порядок, корреляция здесь высокая. Рассматривая первый член в связи с остальными девятью, находим, что его вклад в $S$ равен 9; второй и третий члены последовательности $A$ связаны так, что они ничего не дают для $S$ вне зависимости от того, какие оценки имеет эта пара в последовательности $B$. Оценка объектов, ассоциированных со вторым членом, равна 7 и т. д. Суммарная оценка ($S$) составит: $9+7+4+4+4+3+1+1+0=33$.

Если мы принимаем для определения $\tau$ варианта «а» (стр. 46), то

$$\tau_a = \frac{33}{45} = +0{,}733.$$

В соответствии с альтернативой «б» получим для последовательности $A$:

$$\Sigma a_{ij}^2 = 45 - \frac{1}{2}(2\times 1) - \frac{1}{2}(2\times 1) - \frac{1}{2}(2\times 1) - \frac{1}{2}(2\times 1) = 41$$

и для последовательности $B$:

$$\Sigma b_{ij}^2 = 45 - \frac{1}{2}(4\times 3) - \frac{1}{2}(3\times 2) = 36.$$

Отсюда

$$\tau_b = \frac{33}{\sqrt{(41\times 36)}} = +0{,}859.$$

Во всех случаях значения $\tau_b$ должно быть, конечно, больше, чем $\tau_a$. В данном случае $\tau_b$ существенно выше $\tau_a$.

**3.5.** Исходя из общих соображений, развитых в гл. 2, приемлемой формой коэффициента — действительной меры корреляции между двумя рядами чисел — является $\tau_b$. Например, если мы измерим согласие между двумя суждениями о распределении группы кандидатов в порядке их достоинств (и нет никакого объективного их ранжирования), то мы должны воспользоваться $\tau_b$. Оба суждения могут быть ошибочными относительно некоторой объективной последовательности или они могут не совпадать с другими суждениями, однако, мы не будем это обстоятельство принимать во внимание. Мы измеряем согласованность последовательностей, а не их точность.

Предположим, что обе последовательности совпадают, последним членом каждой из них является $n$, а все остальные члены связаны и, таким образом, имеют ранг, равный $^1/_2\, n$*. Тогда $\tau_b = 1$, поскольку этот коэффициент должен выражать полную согласованность между последовательностями. Однако из (3.1) и (3.2) вытекает, что

$$T = U = \frac{1}{2}(n-1)(n-2),$$

откуда

$$\frac{1}{2}n(n-1) - T = n-1.$$

---

* В самом деле сумма $1+2+\ldots+(n-1) = \frac{(n-1)\,n}{2}$, откуда среднее значение ранга для $n-1$ членов составит: $\frac{(n-1)\,n}{2} : (n-1) = \frac{1}{2}\,n$. — *Прим. перев.*

Следовательно, поскольку оценка равна здесь $n-1$ (подтверждая, что $\tau_b = 1$), имеем:

$$\tau_a = \frac{n-1}{\frac{1}{2}n(n-1)} = \frac{2}{n}. \tag{3.4}$$

При больших $n$ эта величина близка к нулю. Становится ясным, что $\tau_a$ не является приемлемой мерой *согласованности*.

**3.6.** Однако могут встречаться случаи, когда $\tau_a$ будет представлять собой лучший измеритель, чем $\tau_b$. Предположим, что реально существует некоторая объективная последовательность членов ряда. В этом случае целью коррелирования приписанных исследователем рангов является измерение точности этого ряда. Если исследователь ввел связанные ранги, то форма $\tau_b$ учитывает тот факт, что дисперсия его оценок уменьшена. Такой расчет принимает во внимание так называемое «пучкование» оценок в то время как исследователю *не следовало создавать связи рангов, поскольку существует некоторая объективная последовательность*. В этом случае можно согласиться с тем, что при определении $\tau$ следует применить полный делитель, т. е. $^1/_2\, n\ (n-1)$, иначе говоря, $\tau_a$ является приемлемой формой коэффициента.

Рассмотрим случай, когда одна наша последовательность представлена натуральным числом чисел $1, \ldots, n$, а другая — имеет связь в качестве первых $n-1$ членов (ранг каждого равен $^1/_2\, n$), конечный член последовательности равен $n$. Из (3.4), имеем:

$$\tau_a = \frac{2}{n},$$

тогда как

$$\tau_b = \frac{n-1}{\sqrt{\left[\frac{1}{2}n(n-1)^2\right]}} = \sqrt{\frac{2}{n}}. \tag{3.5}$$

Например, при $n = 9$ $\tau_a = 0{,}22$ и $\tau_b = 0{,}47$.

По-видимому, первая величина ближе к тому, что мы должны ожидать от меры согласия для такой объективной последовательности. Исследователь не получил ни одной неправильно расположенной пары и правильно ранжировал один член ряда. Однако он не в состоянии сделать выбор между первыми девятью членами, и величина 0,22 как мера его способности* представляется заслуженной. С другой стороны, если первая последовательность не объективна, а является только выражением мнения другого исследователя, причем его оценки не обладают большей надежностью, то представляется, что значение 0,47 есть хорошая мера согласия.

**3.7.** Есть и другой интересный подход при рассмотрении проблемы связанных рангов. Предположим, что мы имеем дело с некоторым свя-

* Способности различать, выделять объекты. — *Прим. перев.*

занным множеством, состоящим из $t$ членов, которое возникло в связи с тем, что исследователь был неспособен уловить реальные различия. Тогда можно спросить: каково *среднее значение* $\tau$ для всех $t!$[1] возможных путей приписывания целочисленных рангов связанным членам множества?

Если мы заменим некоторое связанное множество $t$ целочисленными рангами и средней всех $t!$ возможных вариантов последовательностей, то получим тот же результат, что и при замене оценок $a_{ij}$ связанных членов нулем, поскольку в $t!$ вариантах упорядочения каждая пара встретится одинаковое число раз в последовательности $XY$ и в последовательности $YX$. Поэтому выбор $+1$ в одном случае и $-1$ в другом—эквивалентен выбору 0 для средней. Таким образом, мы можем рассматривать $\tau_a$ как средний коэффициент, какой был бы получен, если бы связанные ранги были замещены всеми возможными вариантами целочисленных рангов и для каждого варианта было подсчитано $\tau$, а затем исчислена средняя всех результирующих величин.

3.8. Обратимся теперь к рассмотрению аналогичных проблем, связанных с коэффициентом ранговой корреляции $\rho$. Опять мы должны сделать выбор между двумя знаменателями и двумя коэффициентами, которые обозначим как $\rho_a$ и $\rho_b$. Если в двух последовательностях имеется несколько групп связей, представленных $t$ и $u$ членами, то

$$\left.\begin{aligned} T' &= \frac{1}{12}\sum_t (t^3-t) \\ U' &= \frac{1}{12}\sum_u (u^3-u) \end{aligned}\right\}. \tag{3.6}$$

Тогда получим:

$$\rho_a = 1 - \frac{6\,[S(d^2)+T'+U']}{n^3-n}. \tag{3.7}$$

$$\rho_b = \frac{\frac{1}{6}(n^3-n) - S(d^2) - T' - U'}{\sqrt{\left[\frac{1}{6}(n^3-n)-2T'\right]\left[\frac{1}{6}(n^3-n)-2U'\right]}}. \tag{3.8}$$

Прежде чем мы докажем эти формулы, рассмотрим следующий пример.

*Пример 3.2.*

Возьмем опять две последовательности из примера 3.1

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $A$ | 1 | $2\frac{1}{2}$ | $2\frac{1}{2}$ | $4\frac{1}{2}$ | $4\frac{1}{2}$ | $6\frac{1}{2}$ | $6\frac{1}{2}$ | 8 | $9\frac{1}{2}$ | $9\frac{1}{2}$ |
| $B$ | 1 | 2 | $4\frac{1}{2}$ | $4\frac{1}{2}$ | $4\frac{1}{2}$ | $4\frac{1}{2}$ | 8 | 8 | 8 | 10 |

[1] Символ $t!$ (читается как «факториал $t$») означает произведение $1 \times 2 \times 3 \times \ldots \times (t-1) \times t$. Он равен числу возможных способов расположения $t$ объектов. Например, четыре объекта $A$, $B$, $C$, $D$ могут быть расположены 24 различными способами.

В первом ряду рангов имеется четыре связанные пары ($t = 2$), отсюда

$$T' = \frac{4}{12}(2^3-2)=2.$$

Во втором — только одно множество связей, у которого $t = 4$, и одно с $t = 3$, следовательно,

$$U' = \frac{1}{12}(4^3-4+3^3-3)=7.$$

Находим также

$$S(d^2)=13.$$

Отсюда на основе (3.7) получим:

$$\rho_a=1-\frac{6(13+7+2)}{990}=0{,}867$$

и на основе (3.8)

$$\rho_b=\frac{165-22}{\sqrt{(161\times151)}}=0{,}917.$$

**3.9.** Полезно отметить, что (3.8) может быть представлено в форме

$$\rho_b=\frac{\frac{1}{6}(n^3-n)-S(d^2)-T'-U'}{\left[\frac{1}{6}(n^3-n)-(T'-U')\right]\sqrt{1-\frac{(T'-U')^2}{\left[\frac{1}{6}(n^3-n)-(T'+U')\right]^2}}}. \qquad (3.9)$$

Таким образом, если $T'$ и $U'$ являются небольшими относительно $\frac{1}{6}(n^3-n)$ величинами, то получим приближенное соотношение (при $T' = U'$ оно точное):

$$\rho_b=1-\frac{S(d^2)}{\frac{1}{6}(n^3-n)-(T'+U')} \qquad (3.10)$$

или приближенно:

$$\rho_b=1-\frac{S(d^2)}{\frac{1}{6}(n^3-n)}. \qquad (3.11)$$

Эта формула обычно и применяется при расчете $\rho$ в случае отсутствия связей. Поэтому следует ожидать, что когда число связей не слишком велико, применение формул (3.9) и (3.10) приводит к незначительным различиям в числовых результатах сравнительно с теми, которые дает формула (3.11).

Например, используя данные примера 3.2, находим по формуле (3.10)

$$\rho_b = 1 - \frac{13}{165-9} = 0{,}9167$$

и по формуле (3.11)

$$\rho_b = 1 - \frac{13}{165} = 0{,}9212.$$

Формула (3.8) дает 0,9171. Все три результата равны при округлении до второго десятичного знака.

**3.10.** При выводе формул (3.7) и (3.8) мы воспользовались некоторыми результатами гл. 2. В 2.6 было показано, что $\rho$ может трактоваться как коэффициент корреляции между рангами. Предположим, что мы приняли ту же самую точку зрения и для случая, когда некоторые ранги связаны.

Для множества несвязанных рангов сумма квадратов рангов равна

$$\Sigma p_i^2 = \frac{1}{6} n(n+1)(2n+1),$$

где $p_i$ — ранг $i$-го объекта. Сумма рангов составит:

$$\Sigma p_i = \frac{1}{2} n(n+1).$$

Если $t$ рангов связаны, то сумма рангов остается той же самой, однако изменяется сумма квадратов. Предположим, что ранги $p_k + 1, \ldots, \ldots, p_k + t$ связаны. Тогда сумма квадратов сокращается на:

$$(p_k+1)^2 + (p_k+2)^2 + \ldots + (p_k+t)^2 - t\left[p_k + \frac{1}{2}(t+1)\right]^2 =$$
$$= tp_k^2 + 2p_k(1+2+\ldots+t) + 1^2 + 2^2 + \ldots + t^2 -$$
$$- t\left[p_k^2 + p_k(t+1) + \frac{1}{4}(t+1)^2\right] = \frac{1}{12}(t^3 - t).$$

Сейчас удобно (для читателя, который пропустил гл. 2) ввести понятие дисперсии. Эта величина определяется как средний квадрат отклонений совокупности значений от их средней арифметической. Это квадрат стандартного отклонения. Таким образом, дисперсия для несвязанной последовательности равна:

$$\frac{1}{n}\Sigma\left[p_i - \frac{1}{2}(n+1)\right]^2 = \frac{1}{n}\Sigma p_i^2 - \frac{1}{4}(n+1)^2 = \frac{1}{12}(n^2-1),$$

где суммирование производится по $n$ рангам. Отсюда следует, что для связанной последовательности дисперсия составит величину

$$\frac{1}{12}(n^2-1) - \frac{1}{12n}\Sigma(t^3-t) = \frac{1}{2n}\left[\frac{1}{6}(n^3-n) - 2T'\right]. \qquad (3.12)$$

Если две величины $x$ и $y$ измерены в виде отклонений от их средних, то подобным же образом можно определить их ковариацию, cov $(x, y)$, равную $\frac{1}{n}\Sigma(xy)$. Поскольку

$$x^2 + y^2 - 2xy = (x - y)^2,$$

то

$$\operatorname{var} x + \operatorname{var} y - 2\operatorname{cov}(x, y) = \operatorname{var}(x - y).$$

Коэффициент корреляции $\rho_b$ может быть определен:

$$\rho_b = \frac{\operatorname{cov}(p, q)}{\sqrt{\operatorname{var} p \operatorname{var} q}} = \frac{1}{2}\,\frac{\operatorname{var} p + \operatorname{var} q - \operatorname{var}(p - q)}{\sqrt{\operatorname{var} p \operatorname{var} q}}. \tag{3.13}$$

Легко проверить, что для случая, когда ранги не связаны, это дает величину $\rho_b$. Применим такой же подход и к случаю, когда ранги связаны. Поскольку var $(p - q)$ опять равно $\frac{1}{n}S(d^2)$, то из (3.13), применяя (3.12), находим, что

$$\rho_b = \frac{1}{2}\,\frac{\frac{1}{6}(n^3 - n) - 2T' + \frac{1}{6}(n^3 - n) - 2U' - 2S(d^2)}{\sqrt{\left[\frac{1}{6}(n^3 - n) - 2T'\right]\left[\frac{1}{6}(n^3 - n) - 2U'\right]}}.$$

Это выражение легко преобразуется в (3.8). Формулу (3.7) для $\rho_a$ получаем таким же образом.

**3.11.** То, что было сказано выше о различных обстоятельствах, в которых могут предпочитаться $\tau_a$ или $\tau_b$, применимо и к выбору между $\rho_a$ и $\rho_b$. Для того чтобы аналогия была полной, нам необходимо доказать, что $\rho_a$ является средней из значений коэффициентов, которые были бы получены, если бы связи были заменены на целочисленные ранги всеми возможными способами их размещения.

Если последовательность $A$ остается постоянной, то средняя из ковариаций, исчисленных для всех $t!$ размещений множества из $t$ рангов другой последовательности $B$, равна ковариации постоянных рангов в $A$ и среднего ранга $B$. Однако последняя величина определяет значения связанных рангов. Из этого следует, что средняя ковариация есть ковариация связанных рангов, поскольку влияние различных множеств связей аддитивно. Результатом этого является сказанное выше.

*Пример 3.3*

Если две последовательности идентичны, то последний член в каждой из них имеет ранг $n$, а остальные $n - 1$ являются связанными, их ранг равен $\frac{1}{2}n$. Отсюда очевидно, что $\rho_b = 1$. Однако для $\rho_a$ находим:

$$U' = T' = \frac{1}{12}[(n-1)^3 - (n-1)] = \frac{1}{12}n(n-1)(n-2)$$

и $S(d^2) = 0$.

Таким образом, из (3.7) вытекает, что

$$\rho_a = \frac{\frac{1}{6}(n+1)(n)(n-1) - \frac{1}{6}n(n-1)(n-2)}{\frac{1}{6}(n+1)(n)(n-1)} = \frac{3}{n+1}.$$

Итак, различие между двумя видами $\rho$ такое же, что и между двумя видами $\tau$ в (3.5).

**3.12.** В психологии часто сталкиваются с проблемой, которая заключается в измерении взаимосвязи между двумя качественными характеристиками, одна из которых дает возможность осуществить ранжирование, а вторая — *дихотомию*, или группировку на две группы (класса), в соответствии с тем, обладают ли единицы наблюдения определенным свойством или нет. Обращаемся к следующему ранжированному ряду 15 девочек (д) и мальчиков (м) в соответствии с их успехами на экзамене:

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ранг: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| Пол: | м | м | д | м | д | д | м | м | м | д | м | д | м | д | д |

Здесь нас интересует, имеется ли связь между полом и успехом на экзамене — действительно ли мальчики в среднем получили лучшие оценки, чем девочки, или наоборот.

Вообразим, что деление по полу само является некоторым ранжированием. Обследование охватило 8 мальчиков и 7 девочек. Предположим теперь, что первые 8 членов ряда, ранжированного по полу, являются связями, то же самое справедливо для следующей группы из 7 членов. Действительное значение связанных рангов в первом случае составляет $4\,^1/_2$ и 12 — во втором, так что пары рангов могут быть записаны следующим образом:

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Последовательность $A$: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| Последовательность $B$: | $4\frac{1}{2}$ | $4\frac{1}{2}$ | 12 | $4\frac{1}{2}$ | 12 | 12 | $4\frac{1}{2}$ | $4\frac{1}{2}$ | $4\frac{1}{2}$ | 12 | $4\frac{1}{2}$ | 12 | $4\frac{1}{2}$ | 12 | 12 |

Для данных последовательностей мы можем подсчитать теперь $\tau_b$. Находим:

$S = 7 + 7 - 6 + 6 - 5 - 5 + 4 + 4 + 4 - 2 + 3 - 1 + 2 + 0 = 18$; $T = 0$;

$$U = \frac{1}{2}(8 \times 7) + \frac{1}{2}(7 \times 6) = 49;$$

$$\tau = \frac{18}{\sqrt{105 \times 56}} = +0{,}24.$$

Ответ указывает на некоторую положительную корреляцию между последовательностью успеха и последовательностью принадлежности к полу, которую мы приняли, поставив мальчиков первыми. (Конечно, расположив первыми девочек, мы должны будем получить $\tau = -0{,}24$, что приводит к тому же заключению). Короче, представляется, что мальчики имеют более высокие оценки, чем девочки; является ли это свидетельство достаточным индикатором действительной корреляции или нет, зависит от существенности результата. Эту проблему мы обсудим в следующей главе.

*Пример 3.4*

На фабрике опросили группу рабочих; на этой основе было вынесено суждение об их адаптации к условиям жизни. Рабочие квалифицировались как «хорошо приспособляемые» и «сверхактивные». Кроме того, имелись свидетельства опрошенных о частоте никтурии. Для мужчин в возрасте от 50 до 59 лет были получены следующие результаты наблюдения. Ранжирование в соответствии с частотой никтурии (наибольшая никтурия получила наивысший ранг):

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| «Хорошо приспособляемые» | $2\frac{1}{2}$ | $2\frac{1}{2}$ | $2\frac{1}{2}$ | $2\frac{1}{2}$ | $6\frac{1}{2}$ | $6\frac{1}{2}$ | 10 | 10 | 10 | 10 | 14 | 14 |
| «Сверхактивные» | 5 | 10 | 14 | 16 | 17 | | | | | | | |

Беглое ознакомление с этими данными указывает на то, что группа «сверхактивных» имеет большую частоту никтурии. Измерим теперь взаимосвязь с помощью $\tau_b$. Обозначим символом $E$ «хорошо приспособляемых» и $O$ «сверхактивных», получим:

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Последовательность $A$ | $2\frac{1}{2}$ | $2\frac{1}{2}$ | $2\frac{1}{2}$ | $2\frac{1}{2}$ | 5 | $6\frac{1}{2}$ | $6\frac{1}{2}$ | 10 | 10 | 10 | 10 | 10 | 14 | 14 | 14 | 16 | 17 |
| Последовательность $B$ | $E$ | $E$ | $E$ | $E$ | $O$ | $E$ | $E$ | $O$ | $E$ | $E$ | $E$ | $E$ | $E$ | $E$ | $O$ | $O$ | $O$ |

Можно было бы заменить символы $E$ и $O$ рангами, однако для подсчета $S$ это не является необходимым. Находим:

$$S = 5+5+5+5-8+4+4-2+3+3+3+3+2+2 = 34;$$

$$T = \frac{1}{2}(4\times 3)+\frac{1}{2}(2\times 1)+\frac{1}{2}(5\times 4)+\frac{1}{2}(3\times 2) = 20;$$

$$U = \frac{1}{2}(12\times 11)+\frac{1}{2}(5\times 4) = 76.$$

Таким образом,

$$\tau_b = \frac{34}{\sqrt{116\times 60}} = +0{,}41.$$

Читатель должен бы проверить расчет $S$. Взяв, например, первый член последовательности $A$, мы увидим, что остальные ее члены, имеющие ранг $2^1/_2$, не могут внести свой вклад в $S$, поэтому мы рассмотрим следующие члены. Первый член последовательности $B$, соответствующий $E$, дает $+1$ в тех случаях, когда остальные члены обозначены как $O$; в остальной части ряда имеется пять таких членов.

Аналогичное положение наблюдается в отношении второго, третьего и четвертого членов последовательности $A$. Все члены ряда $A$, следующие за пятым, имеющим ранг 5, могут внести свой вклад. В последовательности $B$ пятый член обозначен $O$, отсюда вклад, равный —1, дает каждый из восьми членов, обозначенных $E$, и т. д.

**3.13.** Если ранжирование приводит к дихотомии с числом членов в двух группах: $x$ и $n - x = y$, то

$$\frac{1}{2} n(n-1) - T = \frac{1}{2} n(n-1) - \frac{1}{2} x(x-1) - \frac{1}{2}(n-x)(n-x-1) = xy.$$

Следовательно, получим:

$$\tau_b = \frac{S}{\sqrt{xy\left[\frac{1}{2} n(n-1) - U\right]}}. \tag{3.14}$$

**3.14.** Рассмотрим экстремальный случай, когда обе последовательности характеризуются дихотомией; одна сгруппирована на $x$ и $n—x=y$ членов, другая на $p$ и $n—p$ членов. Тогда в знаменателе $\tau_b$ имеем простое выражение $\sqrt{xypq}$. Данные, включаемые в таблицу (обычно известную как таблица $2 \times 2$), можно записать следующим образом:

Первое качество

| | | обладает | не обладает | всего |
|---|---|---|---|---|
| Второе качество | обладает | $a$ | $b$ | $p$ |
| | не обладает | $c$ | $d$ | $q$ |
| | всего | $x$ | $y$ | $n$ |

(3.15)

Любой член группы (класса), обладающий обоими качествами (число $a$), взятый с любым членом класса, не обладающим ни одним качеством (число $d$), представляют собой пару, имеющую одно и то же положение в любом ранжированном ряде и, следовательно, увеличивающую $S$ на $+1$. Подобно этому член $b$-класса в сочетании с любым членом $c$-класса имеет вклад в величину $S$, равный —1. Остальные варианты сочетаний не дают ничего. Следовательно,

$$S = ad - bc$$

и

$$\tau_b = \frac{ad-bc}{\sqrt{xypq}}. \qquad (3.16)$$

Приведенная формула является одной из полезных форм коэффициента, измеряющего взаимосвязь в таблице $2 \times 2$. Имеются и другие коэффициенты такого же рода, однако интересно, что для экстремального случая, когда обе последовательности настолько связаны, что представляют дихотомию, в качестве коэффициента $\tau$ берут измеритель взаимосвязи, который был создан для других целей [1].

*Пример 3.5*

Вернемся к данным примера 3.4. Предположив, что никтурию можно классифицировать только как нормальную и чрезмерную, получим следующие результаты (см. табл. 3.1)

Таблица 3.1

| | | Никтурия: нормальная | чрезмерная | всего |
|---|---|---|---|---|
| Аттестация | хорошая приспособляемость | 10 | 2 | 12 |
| | сверхактивность | 2 | 3 | 5 |
| | итого | 12 | 5 | 17 |

На основе этих данных имеем:

$$\tau_b = \frac{30-4}{\sqrt{12\times5\times12\times5}} = +0{,}43,$$

что хорошо согласуется со значением 0,41, которое было получено в примере 3.4.

## Применение к порядковым таблицам взаимной сопряженности

**3.15.** Идеей, подобной той, что мы применили при анализе таблицы $2 \times 2$, можно воспользоваться и для создания измерителя, оценивающего взаимосвязь в таблице взаимной сопряженности, в которой группировка по строкам и столбцам соответствует натуральному ряду чисел. Рассмотрим, например, данные табл. 3.2, которые характеризуют степень зоркости правого и левого глаз 3242 мужчин в возрасте 30—39 лет.

---

[1] См., например, [58, vol. 2, § 33.6]. Если подсчитанный обычным путем $\chi^2$ является коэффициентом квадратичной сопряженности для таблицы $2\times2$, то $\tau_b^2$, определенное для этой таблицы, равно $\frac{\chi^2}{n}$.

Таблица 3.2 **Степень зоркости 3242 мужчин в возрасте 30—39 лет [93]**

Левый глаз

| | степень | высшая | вторая | третья | низшая | всего |
|---|---|---|---|---|---|---|
| Правый глаз | высшая | 821 | 112 | 85 | 35 | 1053 |
| | вторая | 116 | 494 | 145 | 27 | 782 |
| | третья | 72 | 151 | 583 | 87 | 893 |
| | низшая | 43 | 34 | 106 | 331 | 514 |
| | всего | 1 052 | 791 | 919 | 480 | 3 242 |

Нас интересует взаимосвязь между степенями зоркости правого и левого глаза. Отметим, что строки и столбцы расположены в правильной последовательности.

Теперь качество зрения (его зоркость) может рассматриваться для этих целей как признак, по которому человеку может быть приписан ранг; любой человек будет иметь ранги в двух последовательностях соответственно для правого и левого глаза. Мы можем рассматривать группировку по степеням зоркости как сравнительно интенсивное связывание 3242 рангов. Так, например, первые 1053 человека «связаны» в одну группу (по зоркости правого глаза), затем следующие 782, следующие 893 и, наконец, последние 514. Значение $\tau$, подсчитанное по данным таблицы с должным учетом связей, измерит взаимосвязь между зоркостью правого и левого глаза для группы в целом.

**3.16.** Для того чтобы определить такой коэффициент, нам требуется сумма $S$. Каждый член любой клетки таблицы в сочетании с любым членом, расположенным ниже и правее этой клетки, внесет положительный вклад в $S$; например, показатель 821, приведенный в верхнем левом углу таблицы, дает следующий вклад:

$$821\ (494 + 145 + 27 + 151 + 583 + 87 + 34 + 106 + 331).$$

Аналогично получим отрицательные вклады, связанные с элементами таблицы, находящимися ниже и левее данного элемента; например, оценка, относящаяся к элементу, лежащему в третьем столбце и второй строке, равна:

$$145\ (-72 - 151 - 43 - 34 + 87 + 331).$$

Никакие оценки не возникают в связи с нижней итоговой строкой таблицы. Находим:

$$S = 821\ (1958) + 112\ (1048) + 85\ (-465) + 35\ (-1744) +$$
$$+ 116\ (1292) + 494\ (992) + 145\ (118) + 27\ (-989) + 72\ (471) +$$
$$+ 151\ (394) + 583\ (254) + 87\ (-183) = 2\ 480\ 223.$$

Для знаменателя $\tau_b$ имеем две величины:

$$\frac{1}{2}(3242)(3241)-\frac{1}{2}(1053)(1052)-\frac{1}{2}(782)(781)-$$
$$-\frac{1}{2}(893)(892)-\frac{1}{2}(514)(513)=3\,864\,293;$$
$$\frac{1}{2}(3242)(3241)-\frac{1}{2}(1052)(1051)-\frac{1}{2}(791)(790)-$$
$$-\frac{1}{2}(919)(918)-\frac{1}{2}(480)(479)=3\,851\,609.$$

Отсюда мера взаимосвязи определяется как:

$$\tau_b=\frac{2\,480\,223}{\sqrt{3\,864\,293\times 3\,851\,609}}=0{,}643.$$

**3.17.** Рассмотрим в качестве меры взаимосвязи альтернативный вариант — коэффициент $\tau_a$, который обладает одним весьма нежелательным свойством (разделяемым и некоторыми другими коэффициентами сопряженности), а именно коэффициент не может достигнуть единицы, если таблица не содержит равное число строк и столбцов. Для случаев, когда связи относительно редки, это не является серьезной помехой, однако для существенно связанных последовательностей может оказаться предпочтительнее применение коэффициента, который удовлетворяет этому условию.

Возьмем максимальную оценку, которую можно получить для $n$ членов, размещенных в таблице, состоящей из $u$ строк и $v$ столбцов. Она достижима тогда, когда все наблюдения лежат на наибольшей диагонали таблицы, а частоты, расположенные в диагональных клетках, близки к тому, чтобы быть равными. Наибольшая диагональ содержит, скажем, $m$ клеток, где $m$ — наименьшее из $u$ и $v$. Если $n$ кратно $m$, то максимальная оценка есть:

$$2\left(\frac{n}{m}\right)^2[1+2+\ldots+(m-1)]=n^2\frac{m-1}{m}. \qquad (3.17)$$

Когда $n$ не кратно $m$, то такая оценка не достигается, однако она очень близка к ней при большом $n$ и малом $m$. Следовательно, находим:

$$\tau_c=\frac{2S}{n^2\frac{m-1}{m}}. \qquad (3.18)$$

Отсюда немедленно следует:

$$\tau_c=\frac{n-1}{n}\,\frac{m}{m-1}\tau_a. \qquad (3.19)$$

В нашем примере $n = 3242$ и $m = 4$, так что

$$\tau_c = \frac{2 \times 2\,480\,223}{\frac{3}{4}(3242)^2} = 0{,}629,$$

в то время как ранее была получена величина $\tau_b = 0{,}643$.

**3.18.** Для таблицы с одинаковым числом строк и столбцов (как, например, в таблице $2 \times 2$) коэффициент $\tau_b$ может достичь единицы, если частоты будут лежать только на главной диагонали. Действительно, пусть частоты имеют значения $f_1, f_2, \ldots, f_m$. Величина $S$ составит:

$$f_1(n-f_1)+f_2(n-f_1-f_2)+ \ldots +f_m(n-f_1-f_2- \ldots -$$
$$-f_m)=n\sum f_i-\sum f_i^2-\sum_{i<j} f_i f_j=$$
$$=\frac{1}{2}(n^2-\Sigma f_i^2).$$

Знаменатель $\tau_b$, подсчитанный по итогам как строк, так и столбцов, равен:

$$\frac{1}{2}n(n-1)-\sum \frac{1}{2}f_i(f_i-1)=\frac{1}{2}(n^2-\Sigma f_i^2),$$

отсюда $\tau_b$ равно единице. Коэффициент $\tau_c$ применяется главным образом тогда, когда число строк и столбцов не равно друг другу.

## Библиография

См. [96], [18], [118], [51], [52], [82], [114], [6]. Приложение к таблицам сопряженности признаков см. [58, vol. 2, ch. 32].

Другой тип коэффициентов $\tau$ был предложен в [32] и рассмотрен в [58].

## *ГЛАВА 4.* ПРОВЕРКА СУЩЕСТВЕННОСТИ

**4. 1.** Последовательности, с которыми мы имеем дело на практике, обычно основываются на совокупности объектов, которая сама является лишь выборкой из более многочисленной совокупности. Так, возможность измерения взаимосвязи между математическими и музыкальными способностями в данной группе учеников представляет определенный интерес, однако существенно больший интерес заключается в возможности ответить на вопрос, в какой мере результат, полученный на основе выборки, проливает свет на взаимосвязь в исходной совокупности, если данная группа учеников отобрана случайно. В этой главе рассмотрим следующий вопрос: с какой степенью надежности мы можем полагаться на заключение о том, что в совокупности, из которой произведен отбор, существует корреляция, если получен некоторый выборочный коэффициент корреляции рангов. Короче, мы попытаемся проверить *существенность* наблюдавшихся корреляций рангов в свете статистической теории выборки.

**4.2.** Предположим, что в представленной совокупности нет взаимозависимости между двумя изучаемыми качествами. Тогда, если выборка случайная, какая-либо последовательность качественных характеристик $A$ должна появиться в сочетании с данной последовательностью $B$ как, вероятно, любая другая последовательность. Если мы приняли любое произвольное расположение $B$ (не имеет значения, какое, поэтому возьмем натуральный ряд чисел, т. е. 1, 2, ..., $n$), то все $n!$ возможных последовательностей чисел от 1 до $n$ для $A$ являются одинаково вероятными. Соответственно каждое из них имеет вероятность $\frac{1}{n!}$ (в данном случае ограничимся случаем, когда связи отсутствуют).

Каждой возможной последовательности $A$ теперь корреспондирует некоторое значение $\tau$ и $\rho$. Совокупность таких величин, общим числом $n!$, может быть сгруппирована в соответствии с действительными значениями $\tau$ или $\rho$, которые охватывают диапазон от $-1$ до $+1$. Результат такой группировки называют *распределением частот*. Это распределение является фундаментальным в настоящем исследовании, и мы рассмотрим некоторые его частные свойства.

**4.3.** При ранжировании четырех единиц наблюдения существует 24 возможных последовательности. Если читатель выпишет их и сопоставит с натуральным рядом чисел 1, 2, 3, 4, то получит следующие значения $S$:

| Значение $S$ | Число последовательностей с данным значением $S$ (частота) |
|---|---|
| —6 | 1 |
| —4 | 3 |
| —2 | 5 |
| 0 | 6 |
| 2 | 5 |
| 4 | 3 |
| 6 | 1 |
| | 24 |

Таким образом, наибольшая частота значений $S$, равная 6, соответствует $S = 0$. Данное распределение симметрично относительно этой величины; когда абсолютное значение $S$ превышает 0, частоты падают до единицы.

**4.4.** Такого же рода распределение для $n = 8$ имеет следующий вид (ниже приводится только нулевое и положительные значения $S$, частоты отрицательных значений симметричны относительно приведенных ниже величин).

| Значение $S$ | Число последовательностей с данным значением $S$ (частота) | Значение $S$ | Число последовательностей с данным значением $S$ (частота) |
|---|---|---|---|
| 0 | 3 826 | 16 | 602 |
| 2 | 3 736 | 18 | 343 |
| 4 | 3 450 | 20 | 174 |
| 6 | 3 017 | 22 | 76 |
| 8 | 2 493 | 24 | 27 |
| 10 | 1 940 | 26 | 7 |
| 12 | 1 415 | 28 | 1 |
| 14 | 961 | Итого для распределения (в целом) | 40 320 |

На рис. 4.1 показан график частот этого распределения; здесь частоты ($f$) показаны в виде ординат соответствующих абсцисс ($S$).

Опять нами найдено, что максимальное значение приходится на $S = 0$, затем при росте $S$ значение частоты неуклонно падает.

**4.5.** В следующей главе мы покажем, как получить такие распределения для различных значений $n$. Здесь же мы утверждаем без доказательства следующее:

а) рассматриваемое распределение всегда симметрично. Если величина $\frac{1}{2}n(n-1)$ четная, то $S$ принимает только четные значения и максимум частоты приходится на $S=0$. Если же величина $\frac{1}{2}n(n-1)$ нечетная, то $S$ может принимать только нечетные значения и имеется два максимума частот в точках $S=\pm 1$;

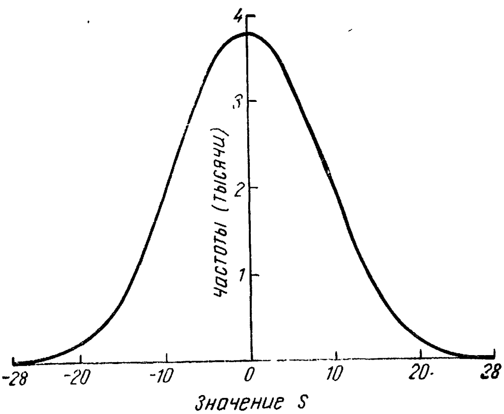


Рис. 4.1.

б) частоты монотонно уменьшаются от максимума до единицы в точках $S = \pm\frac{1}{2}n(n-1)$;

в) по мере роста $n$ очертания полигона частот приближаются к *нормальной кривой*

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}, \qquad (4.1)$$

и если $n$ превышает 10, эта кривая обеспечивает удовлетворительное приближение к полигону. Параметр $\sigma$ кривой, равный ее стандартному отклонению, определяется:

$$\sigma^2=\frac{1}{18}n(n-1)(2n+5). \qquad (4.2)$$

Если нормальную кривую со стандартным отклонением, полученным по данной формуле, изобразить на графике (см. рис. 4.1), то она пройдет настолько близко от полигона, что будет при принятой шкале едва отличима от него.

**4.6.** Приближение данного распределения частот к нормальному является полезным свойством, которое позволяет нам избежать расчета действительного распределения для $n > 10$. Для $n \leqslant 10$ распределения найдены, они положены в основу табл. 1 приложения. Эта таблица показывает относительные частоты (фактические частоты, деленные на $n!$), которые соответствуют величине $S$, равной или превышающей некоторые фиксированные значения. Такая форма представления данных требуется наиболее часто.

**4.7.** Теперь рассмотрим пути использования данных распределений при проверке существенности $\tau$. Проверка $\tau$ эквивалентна проверке соответствующего значения $S$, поскольку одна из этих величин кратна другой; легко прийти к выводу о том, что более удобно (в расчетном отношении) иметь дело с $S$.

Если нет взаимосвязи между двумя качественными характеристиками, то пара случайно выбранных последовательностей даст некоторое значение $S$, лежащее в пределах $\pm\frac{1}{2}n(n-1)$. Подавляющая часть таких значений будет концентрироваться вокруг нуля. Примем следующий критерий: если наблюдаемые значения $S$ таковы, что весьма маловероятно, чтобы такое или большее абсолютное значение могло появиться случайно, то мы отклоняем гипотезу о том, что данные две качественные характеристики являются независимыми. Иначе говоря, если наблюдаемое значение $S$ лежит в «хвостах» распределения $S$ далеко от средней, то мы отвергнем эту гипотезу. Решение вопроса о том, где провести линии, для того чтобы отделить «хвосты», или, что является весьма маловероятным, дело соглашения. Обычно вероятность 0,05 или 0,01 считают малой, а 0,001 — очень малой. Иногда говорят «пятипроцентный уровень вероятности $S$». Это означает, что некоторое значение $S$ достигают или превышают с вероятностью 0,05, аналогично употребляют выражения «однопроцентный уровень вероятности» и «0,1-процентный уровень вероятности». Соответствующие значения $S$ могут быть определены, например, как «пятипроцентный предел существенности». Фраза «наблюденные $S$ лежат вне пятипроцентного предела существенности» означает, что вероятность появления такого или большего (абсолютного) значения $S$ меньше, чем 0,05.

Если мы заранее подозреваем существование положительной корреляции, то можно рассматривать вероятность того, что $S$ лежит только в верхнем «хвосте» распределения; аналогично можно ограничиться только нижним «хвостом», если ожидается отрицательная корреляция. Это равнозначно расчету вероятности того, что $S$ достигает некоторого заданного значения его, или превышает, или, как это иногда бывает, падает ниже какой-нибудь величины вместо того, чтобы достигать или превышать некоторое число без учета его знака.

*Пример* 4.1

Для случайной выборки, давшей последовательность из 10 объектов, найдено значение $\tau$, равное —0,11. Является ли эта величина существенной?

Соответствующее значение $S$ равно —5. Из табл. 1 приложения видно, что доля последовательностей, которые дают значение —5 и меньше, +5 и больше составляет 0,364, взятое дважды, т. е. 0,728. Эта значительная величина, и мы не можем отвергнуть гипотезу независимости двух качественных характеристик. Другими словами, наблюденная величина $\tau$ не является существенной. Она могла появиться случайно в ходе выборки из исходной совокупности, в которой математические и музыкальные способности не находятся во взаимосвязи.

Если найденное значение $\tau$ было бы равно +0,56, что соответствует $S = 25$, то вероятность того, что $|S| \geqslant 25$ (больше или равно 25 по абсолютной величине) составит 0,028. Это незначительная вероятность; следовало бы прийти к выводу о том, что способности не являются независимыми в исходной совокупности.

Мы основывали наше заключение на *абсолютных* значениях $S$; представляется, что такой подход является наилучшим, особенно тогда, когда выборочное распределение симметрично. Однако, если необходимо, заключение можно основывать на фактических значениях. Если вероятность того, что $S$ больше, чем некоторая абсолютная величина $S_0$, равна $P_0$, то это записывается так:

$$\text{Вероятность } \{|S| > S_0\} = P_0.$$

Тогда вероятность того, что $S > S_0$ (где $S_0$ — положительная величина), равна $\frac{1}{2} P_0$; этому равна и вероятность того, что $S < S_0$ (где $S_0$ — отрицательная величина). Например, для $S_0 = 25$

$$\text{вероятность } \{S \geqslant S_0\} = 0{,}014$$

а для $S = -5$

$$\text{вероятность } \{S \leqslant -5\} = 0{,}364.$$

Пользуясь таблицами вероятностей, читатель должен помнить, какая вероятность рассматривается. Так, 5%-ный предел, который применяется для абсолютного значения, представляет собой лишь 2,5%-ный предел для фактической величины.

### Поправка на непрерывность

4.8. Если $n$ больше 10, то мы применяем таблицу площадей под нормальной кривой как приближение к точным значениям распределения $S$. Соответствующие площади даны в табл. 3 приложения. Они характеризуют вероятность того, что данное число стандартных отклонений (а не абсолютная величина) будет достигнуто или превышено. Полезно запомнить, что у нормального распределения вероятность превышения в 1,96 раза абсолютного значения стандартного отклонения[1] равна 0,05, вероятность превышения стандартного отклонения в 2,58 раза составит 0,01 и превышения в 3,3 раза — 0,001.

При применении этой таблицы следует помнить, что мы заменяем распределение $S$, которое дискретно, непрерывным распределением, отличающимся от первого на две единицы. Для того чтобы улучшить приближение, будем считать, что частота $f$, относящаяся к $S$, не сконцентрирована в точке $S$, а равномерно распределена на протяжении от $S - 1$ до $S + 1$. Для того чтобы сопоставить с площадями нормальной кривой, примем, что «хвост» распределения начинается в точке $S - 1$, иначе говоря, мы вычтем единицу из наблюдаемого $S$, если оно положительно (и прибавим единицу, если оно отрицательно), прежде чем выразим этот показатель в стандартных отклонениях. Данная процедура

[1] Стандартное отклонение распределения выборочных величин (т. е. выборочного отклонения) называется «стандартной ошибкой». Квадрат этой величины известен как «выборочная дисперсия», или кратко как «дисперсия» и записывается «var». Таким образом $\operatorname{var} S = \sigma_S^2 = \frac{1}{18} n (n - 1) (2n + 5)$.

известна как «поправка на непрерывность». Следующий пример прояснит эту процедуру.

*Пример* 4.2.

Для двух ранжирований 20 объектов получено значение $S$, равное 58, соответственно $\tau = 0{,}31$. Является ли эта величина существенной?

На основе (4.2) находим:

$$\sigma^2 = \operatorname{var} S = \frac{1}{18}(20 \times 19 \times 45) = 950,$$

$$\sigma = 30{,}82.$$

Для $S$, скорректированного на непрерывность, получим величину 57, и, следовательно,

$$S \text{ (скоррект.)} = \frac{57}{30{,}82}\sigma = 1{,}85\,\sigma.$$

В табл. 3 приложения находим, что вероятность отклонения меньшего чем $1{,}85\,\sigma$ составляет около 0,9678. Вероятность того, что отклонение будет равно или превысит эту величину, найдем как $2(1 - 0{,}9678) = 0{,}064$. Эта величина небольшая, но не очень. Мы подозреваем, что наблюдаемое значение $\tau$ является существенным, однако мы не можем прийти к вполне определенному решению.

Попробуем теперь сравнить величину, которую дает нормальное приближение при $n = 9$. Предположим, что наблюдаемое $S$ равно 20, соответственно $\tau = 0{,}56$. На основе (4.2) имеем:

$$\sigma = \sqrt{\frac{1}{18}(9 \times 8 \times 23)} = 9{,}592.$$

При корректировке на непрерывность получим:

$$\frac{S-1}{\sigma} = \frac{19}{9{,}592} = 1{,}981.$$

Как видно из табл. 3 приложения, вероятность того, что эта величина будет достигнута или превзойдена по абсолютному значению, равна примерно 0,048. Точное значение, взятое из табл. 1 приложения, равно 0,044. Если бы мы не произвели поправку на непрерывность, то нашли бы, что искомое значение равно 0,037.

**4.9.** Формула (4.2) для стандартной ошибки $S$ требует некоторой перестройки для случаев, когда имеются связи. Если связи охватывают $t$ членов в одной последовательности и $u$ в другой, то дисперсия распределения, полученного коррелированием одной последовательности со всеми $n!$ возможными вариантами второй последовательности, определяется как

$$\operatorname{var} S = \frac{1}{18}[n(n-1)(2n+5)-\sum_t t(t-1)(2t+5)-$$
$$-\sum_u u(u-1)(2u+5)] + \frac{1}{9n(n-1)(n-2)}[\sum_t t(t-1)(t-2)] \times$$
$$\times [\sum_u u(u-1)(u-2)] + \frac{1}{2n(n-1)}[\sum_t t(t-1)][\sum_u u(u-1)]. \quad (4.3)$$

Если связи содержатся только в одной последовательности и, соответственно, все $u$ суть нули, то

$$\operatorname{var} S = \frac{1}{18}[n(n-1)(2n+5)-\sum_t t(t-1)(2t+5)]. \quad (4.4)$$

Мы докажем это в следующей главе.

**4.10.** По мере увеличения $n$ распределение $\tau$ для любого фиксированного числа связей, как и при отсутствии связей, стремится к нормальному. По-видимому, возникает не очень серьезная ошибка при применении нормального приближения для $n \geqslant 10$, тем не менее если связи многочисленны, то, может быть, следует предпринять специальное исследование. Для случаев, когда $n < 10$, полных таблиц для большого числа вариантов нет. Однако Силлито [82] создал таблицы для любого числа связанных пар или связанных троек до $n = 10$ включительно.

*Пример 4.3*

Рассмотрим две последовательности, состоящие из двенадцати членов:

$$\begin{array}{lcccccccccccc} A & 1\frac{1}{2} & 1\frac{1}{2} & 3 & 4 & 6 & 6 & 6 & 8 & 9\frac{1}{2} & 9\frac{1}{2} & 11 & 12 \\ B & 2\frac{1}{2} & 2\frac{1}{2} & 7 & 4\frac{1}{2} & 1 & 4\frac{1}{2} & 6 & 11\frac{1}{2} & 11\frac{1}{2} & 8\frac{1}{2} & 8\frac{1}{2} & 10 \end{array}$$

Находим

$$S = 8+8+1+5+5+5+5-3-2+1+1 = 34.$$

В первой последовательности связи охватывают 2, 2 и 3 члена, во второй — 2, 2, 2 и 2 члена. Теперь на основе (4.3) получим:

$$\operatorname{var} S = \frac{1}{18}[(12\times 11\times 29)-6(2\times 1\times 9)-(3\times 2\times 11)]+0+$$
$$+\frac{1}{2\times 12\times 11}[2(2\times 1)+(3\times 2)][4(2\times 1)] = 203{,}30.$$

Таким образом, с поправкой на непрерывность

$$S = \frac{33}{\sqrt{203{,}30}}\sigma = 2{,}31\,\sigma.$$

Шанс достигнуть или превзойти эту величину по абсолютному значению составит около 0,021. Это незначительная вероятность, и мы склоняемся считать величину $S$ существенной.

**4.11.** Если одна последовательность вырождается в дихотомию с числом членов $x$ и $n - x = y$, как, например, в 3.13 и со связями в другой последовательности, представляемыми $t$ членами, то после подстановки в (4.3) соответствующих величин получим уравнение

$$\operatorname{var} S = \frac{xy}{3n(n-1)}\left[n^3 - n - \sum_t (t^3 - t)\right]. \qquad (4.5)$$

Наконец, если в обеих последовательностях, как, например, в 3.14, наблюдается дихотомия, то после подстановки в 4.5 соответствующих величин имеем:

$$\operatorname{var} S = \frac{xypq}{n-1}. \qquad (4.6)$$

**4.12.** Эти уравнения позволяют нам проверить существенность коэффициентов $\tau_a$, $\tau_b$ и $\tau_c$ или значений $S$, на основе которых они получены. Имеется, однако, одно затруднение, связанное с поправкой на непрерывность, а именно:

а) в случае, когда имеется дихотомия и несвязанная последовательность, интервал между соседними значениями $S$ равен двум. Поправка на непрерывность равна половине этого интервала, т. е. единице;

б) для дихотомии и последовательности, целиком состоящей из связей, охватывающих $t$ членов, интервал равен $2t$ и соответствующая поправка на непрерывность составляет величину $t$;

в) когда обе переменные представлены дихотомиями, интервал равен $n$ и поправка на непрерывность составляет $\frac{1}{2}n$;

г) если последовательность для одной переменной расчленена на две группы, а вторая содержит связи различной протяженности, то в различных частях диапазона наблюдений интервалы между соседними значениями $S$ будут варьировать. В этом случае мы можем прибегнуть к приближенному методу, как это показано ниже в примере 4.5.

*Пример 4.4*

В примере 3.5 мы нашли величину

$$\tau_b = 0{,}43,$$

которая базировалась на $S = 26$. На основе (4.6) определяем дисперсию:

$$\sigma_s^2 = \operatorname{var} S = \frac{12 \times 5 \times 12 \times 5}{16},$$

$$\sigma = 15.$$

Поправка на непрерывность равна 17/2, следовательно, имеем:

$$\frac{S-8,5}{15}=1,167.$$

Вероятность того, что эта величина будет достигнута или превышена по абсолютному значению, равна 0,24, $\tau_b$ не является существенным.

*Пример 4.5*

Для данных примера 3.4 при условии, что одна переменная не разбита на две группы, находим, что $\tau_b = 0,41$ и $S = 34$. На основе (4.5) определим дисперсию:

$$\sigma^2 = \operatorname{var} S = \frac{12\times 5}{3\times 17\times 16}[17^3-17-(4^3-4)-$$
$$-(2^3-2)-(5^3-5)-(3^3-3)] = 344,6,$$
$$\sigma = 18,56.$$

Приступая к определению поправки на непрерывность, заметим, что в последовательности $A$ имеется «прыжок» с $2\frac{1}{2}$ на 5, а это приводит к интервалу в значении $S$, равному 5. Если же мы заменим одно из $E$ на $O$, которое корреспондирует рангу 5 в последовательности $A$, то $S$ сократится на 5 и первые пять параметров составят $4+4+4-9+4$ вместо $5+5+5+5-8$. Аналогично «прыжок» с 5 до $6\frac{1}{2}$ дает интервал 3 и т. д. Мы можем подсчитать средний интервал без расчета каждого отдельного интервала. Общий интервал параметра $S$ равен удвоенному числу членов в последовательности минус продолжительность связей, включающих первый и последний член. В рассматриваемом случае он составит $34 - 4 - 1 = 29$. Таким образом, средний интервал равен $29/6 = 4,833$. В качестве поправки на непрерывность возьмем половину этой величины, откуда

$$\frac{S-2,416}{18,56}=1,70.$$

Это дает возможность установить, что вероятность того, что $S$ будет равно данной величине или превысит ее по абсолютному значению, составит 0,089. Величина $\tau_b$ опять-таки является несущественной.

В предыдущем примере для таблицы $2\times 2$ мы нашли, что $\tau_b = 0,43$ и что эта величина не является существенной. Результат, вообще говоря, мало чего стоит. В примере, который был только что рассмотрен, $\tau_b = 0,41$, т. е. несколько меньшая величина, вероятность превзойти которую по абсолютному значению равна 0,089, в то время как для этой же величины, полученной для таблицы $2\times 2$, вероятность составила 0,24. Здесь нет расхождения. В настоящем примере мы не расчленили на две группы вторую последовательность и приняли во внимание все ранги. Наш метод позволил проявить себя величинам, которые могли сформироваться случайно, и, следовательно, вероятность превышения данной величины в этой, более обширной области, может отличаться от величины, полученной для более ограниченной области определения при дихотомии.

**Существенность $\rho$**

4.13. Совокупность $n!$ величин, полученных коррелированием всех возможных (несвязанных) последовательностей с произвольной последовательностью, дает множество значений $\rho$, которое может быть использовано для проверки существенности этого коэффициента так же, как это было сделано для $\tau$. Распределение $\rho$ симметрично и имеет тенденцию к нормальному при больших $n$, однако его менее удобно применять, чем распределение $\tau$, по следующим причинам:

а) действительное распределение значительно труднее определить; полностью оно получено только для $n = 10$ включительно;

б) распределение приближается к нормальному медленнее, чем распределение $\tau$, и необходима некоторая промежуточная форма, для того чтобы восполнить разрыв между значениями для $n = 11$ и $n$ (довольно сомнительным), при котором приближение к нормальному распределению достаточно;

в) для такой промежуточной области не разработаны простые методы расчета, учитывающие связи.

4.14. При отсутствии связей стандартное отклонение распределения $\rho$ рассчитывается по простой формуле

$$\operatorname{var} \rho = \sigma_\rho^2 = \frac{1}{n-1}. \tag{4.7}$$

Соответствующее выражение для $S(d^2)$ имеет вид:

$$\operatorname{var} S(d^2) = \left(\frac{n^3-n}{6}\right)^2 \frac{1}{n-1}. \tag{4.8}$$

Если $n > 20$, то, вероятно, применение этих выражений при предположении о наличии нормального выборочного распределения будет вполне правильным.

*Пример 4.6*

При ранжировании 20 объектов наблюденное значение $S(d^2)$ составило 840.

Имеем:

$$\rho = 1 - \frac{6 \times 840}{20 \times 21 \times 19} = 0{,}3684.$$

Стандартная ошибка по формуле (4.7) равна $1/\sqrt{19} = 0{,}2294$. Таким образом, наблюденное значение равно $0{,}3684/0{,}2294 = 1{,}61$ стандартной ошибки и едва ли может быть существенным.

Для такой большой величины $S(d^2)$ поправка на непрерывность не представляет большой важности, однако если мы все же хотим ее осуществить, то должны вычесть единицу из $S(d^2)$. Тогда по формуле (4.8) получим стандартную ошибку $S(d^2)$:

$$\sigma = \frac{20^3-20}{6}\sqrt{\frac{1}{19}} = 305{,}1.$$

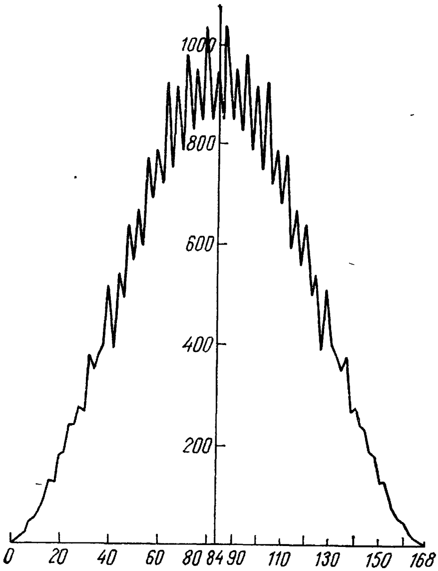


Рис. 4.2

Теперь $S(d^2)$ варьирует от 0 до $\frac{1}{3}(n^3-n)$ со средней $\frac{1}{6}(n^3-n)$; в данном случае она равна 1330. Получим следующее отклонение наблюденной величины от средней: $840 - 1330 = -490$. Абсолютное отклонение с учетом поправки на непрерывность составит: $\frac{489}{305,1} = 1,60$ стандартного отклонения, что приводит к тому же заключению, к которому мы пришли относительно $\rho$.

На рис. 4.2 показано распределение $S(d^2)$ в виде полигона частот для $n = 8$. Пилообразный профиль графика является необычным, однако соответствие нормальной кривой умеренно хорошее, хотя не вполне хорошее для наших целей, поскольку при проверке существенности нас интересует, главным образом, соответствие кривой в области «хвостов» распределения.

## Поправка на непрерывность для $\rho$

**4.15.** Подобно поправкам на непрерывность для $S$, рассмотренным в 4.12, можо предложить следующие поправки для $S(d^2)$ при наличии связей:

а) если имеются дихотомия и последовательность без связей, то интервал между последовательными величинами равен $n$, и поэтому мы вычитаем $\frac{1}{2}n$ из $S(d^2)$;

б) если имеются дихотомия и последовательность, которая целиком состоит из связей, охватывающих $t$ членов, то интервал равен $nt$ и соответственно надо вычитать $\frac{1}{2}nt$;

в) при наличии двойной дихотомии поправка на непрерывность равна $\frac{1}{4}n^2$.

Однако эта поправка мало чего стоит, поскольку при двойной дихотомии $\rho_b$ точно такое же, что и $\tau_b$; формула для последней характеристики обычно и применяется в расчетах.

Если в последовательности присутствуют связи, то не требуется никакой корректировки в значении дисперсии $\rho$, она остается равной $1/(n-1)$.

### Проверка в нулевом случае

**4.16.** До сих пор мы применяли проверки, основанные на распределении корреляций в совокупности величин, полученных при перестановке рангов всеми возможными путями. Такой подход эффективен, если проверяется гипотеза о том, что исследуемые качественные признаки в генеральной совокупности независимы. Тогда наша проверка покажет, является ли наблюдаемая корреляция существенным указанием на отличие корреляции генеральной совокупности *от нуля*. Однако нам может понадобиться испытать корреляции с другой точки зрения или приписать пределы в вероятностном смысле этой генеральной величине. Например, предположим, что мы получили значение $\tau$, равное 0,6, и найдено, что оно существенно. Можем ли мы сказать, в каких пределах находится величина корреляции? Допустим еще раз, что для другой выборки эта величина равна 0,8, причем она также существенна. Можем ли мы сказать, что второе значение более существенно, чем первое, или разность между ними можно приписать действию случая?

**4.17.** Предположим, что целая совокупность, охватывающая $N$ членов, ранжирована по первой переменной в порядке 1, ..., $N$. Очевидно, что это можно допустить без потери общности доказательства. Расположим теперь в этой последовательности данную совокупность. Пусть ранг $i$-го члена, соответствующий второй переменной, будет равен $p_i$. Тогда в соответствии с нашим обычным методом подсчитаем значение $\tau$ для этой совокупности.

Теперь предположим, что из $N$ выбрано $n$ членов. Эти члены будут иметь натуральный порядок по первой переменной. Тогда можно определить величину, допустим, $t$, которая будет означать выборочный коэффициент $\tau$. Итак, вместо $\tau$ запишем $t$ для обозначения того, что мы рассматриваем выборочную величину, а не показатель, относящийся к генеральной совокупности. Для каждой возможной выборки имеется соответствующая величина $t$, и поскольку существует $\binom{N}{n} = \frac{N!}{n!\,(N-n)!}$ выборок, то имеется равное количество значений $t$.

**4.18.** В следующей главе мы покажем, что для любой совокупности распределение $t$ стремится к нормальному распределению, по мере того как размер выборки $n$ возрастает при условии, что генеральное $\tau$ не очень близко к единице и удовлетворяются некоторые условия, не носящие характер ограничений. Мы также покажем, что средняя этого распределения равна $\tau$. До сих пор все шло хорошо. Однако дальше мы сталкиваемся с рядом затруднений. Если бы стандартное отклонение распределения зависело только от $\tau$, то мы легко могли бы проверить наблюдаемую величину так, как нам нужно; однако в действительности стандартное отклонение зависит от других неизвестных величин.

**4.19.** Простая иллюстрация прояснит эту мысль. Рассмотрим последовательность из 9 членов (в соответствии со значением второй переменной):

5 2 3 1 6 7 8 9 4

Имеется $\binom{9}{3} = 84$ возможных варианта выборок, состоящих из 3 членов. Найдем значение $P$ для каждого варианта выборки. Получим следующее распределение:

| Значение $P$ | Численность |
|---|---|
| 0 | 2 |
| 1 | 15 |
| 2 | 34 |
| 3 | 33 |
| | Итого 84 |

Находим, что средняя этого распределения равна 13/6, отсюда среднее значение для 84 величин $t$ составит:

$$E(t) = \frac{26/6}{3} - 1 = 0{,}44.$$

Можно найти, что для генерального значения $\tau$ $P = 26$ и, следовательно,

$$\tau = \frac{52}{36} - 1 = 0{,}44 = E(t).$$

Таким образом проверяется наше утверждение о том, что среднее значение $t$ равно генеральному $\tau$.

Последовательность

1 2 5 9 3 6 7 8 4

также имеет $\tau = 0{,}44$, однако распределение $P$ для выборок по 3 члена теперь следующее:

| Значение $P$ | Численность |
|---|---|
| 0 | 3 |
| 1 | 16 |
| 2 | 29 |
| 3 | 36 |
| | Итого 84 |

Опять среднее значение $t$ равно генеральному $\tau$, распределение $P$ во втором случае отличается от распределения в первом примере, его дисперсия равна 0,734, в то время как дисперсия в первом примере составила 0,639.

**4.20.** Мы находимся в трудном положении, поскольку, если мы ничего не знаем о последовательности рангов в исходной совокупности (обычно так и бывает), не можем выразить дисперсию $t$ через известные факторы. В следующей главе, тем не менее, мы покажем, что для любой исходной совокупности дисперсия $t$ не может превышать определенное значение, а именно:

$$\operatorname{var} t \leqslant \frac{2}{n}(1 - \tau^2). \tag{4.9}$$

Этот результат можно с большой надежностью применить для проверки. Данное выражение также пригодно для случаев, когда имеются связи. Соответствующее приближенное выражение для $\rho$ при больших $n$ (здесь $r_s$ означает выборочное значение $\rho$) имеет вид:

$$\operatorname{var} r_s \leqslant \frac{3}{n}(1-\rho^2). \tag{4.10}$$

Однако это выражение нельзя считать обоснованным для случаев, когда имеются связи. Данное замечание будет более понятным при рассмотрении следующего примера.

*Пример 4.7*

Найдено, что для последовательности из 30 рангов величина $t$ равна 0,816. Допустим, что этот ранжированный ряд является случайной выборкой. Что можно сказать теперь о значении $\tau$ в исходной совокупности?

Для выборок, состоящих из 30 единиц, распределение $t$ можно принять нормальным и взять в виде оценки среднего значения величину 0,816. Из (4.9) находим:

$$\operatorname{var} t \leqslant \frac{2}{30}(1-0{,}816^2) = 0{,}022276,$$

иначе говоря, стандартная ошибка *меньше* или равна 0,149.

Вероятность отклонения от средней на 1,96 стандартного отклонения или больше (по абсолютной величине) равна 0,05. Поэтому мы можем сказать, что в худшем случае вероятность того, что действительная величина $\tau$ лежит в пределах $\pm 0{,}149 \times 1{,}96 = 0{,}292$ вокруг 0,816, равна 0,95. Иначе говоря, вероятность того, что действительная величина $\tau$ лежит вне интервала $0{,}524 \div 1{,}0$, меньше или равна 0,05. Вместо определения пределов обычным статистическим путем с соответствующей степенью вероятности, мы устанавливаем пределы с максимальной вероятностью; иначе говоря, мы устанавливаем *внешние* границы для диапазона генерального $\tau$. Это может привести к излишней строгости: мы делаем ошибку, повышающую надежность в том смысле, что избегаем опасность приписывания существенности несущественным результатам. Правда, в некоторых случаях мы можем не выявить существенность там, где она действительно имеется.

**4.21.** Выше связь со стандартной ошибкой была показана так, как это обычно практикуется в элементарном курсе статистики[1]. Для того чтобы применить формулу (4.9), нам нужно заменить неизвестное генеральное значение $\tau$ на выборочное $t$, как это было сделано в предыдущем примере. Этой процедуры можно избежать, если обратиться к теории доверительных интервалов. Если $x$ — нормированное отклонение, соответствующее уровню вероятности $P$ про-

[1] См., например, Юл Д. Э., Кендэл М. Дж. Теория статистики. М., Госстатиздат, 1960, гл. 18 и 19. О теории доверительных интервалов см. [58, vol. 2, ch. 20].

центов (т. е. вероятность того, что отклонение от средней на $x$ стандартных отклонений или больше по абсолютному значению возникнет в связи со случайным отбором, равна $0,0P$), то самое большее с этой вероятностью можно утверждать, что

$$\tau_1 \leqslant \tau \leqslant \tau_2, \qquad (4.11)$$

где $\tau_1$, $\tau_2$ — корни следующего выражения:

$$t - \tau = x \sqrt{\frac{2}{n}(1-\tau^2)}\,.$$

Отсюда

$$\tau = \frac{t \pm x\sqrt{\frac{2}{n}}\sqrt{1 + \frac{2x^2}{n} - t^2}}{1 + \frac{2x^2}{n}}\,. \qquad (4.12)$$

*Пример 4.8*

По данным примера 4.7 имеем $n = 30$, $t = 0,816$. Нормированное отклонение, соответствующее вероятности 0,05, равно 1,96. Подстановка данных в (4.12) дает:

$$\tau = \frac{t \pm 0,50607\sqrt{1,2561 - t^2}}{1,2561}$$

и $\tau_1 = 0,34$, $\tau_2 = 0,96$.

Таким образом, можно утверждать, что $\tau$ лежит между 0,34 и 0,96, причем можно быть уверенным, что это справедливо в среднем по крайней мере в 95% случаев. Этот метод более точен, чем метод, рассмотренный в 4.20. Ограничения, разумеется, здесь остаются максимальными.

*Пример 4.9*

Для выборки из 20 объектов получено значение $\tau$, равное 0,8. Вторая выборка дала значение этого коэффициента, равное 0,6. Имеется ли указание на то, что выборки принадлежат различным совокупностям; иначе говоря, могут или нет различные значения $\tau$ возникнуть благодаря случайности?

Имеем следующую величину максимальной дисперсии для первого случая:

$$\operatorname{var} t = \frac{2}{n}(1-\tau^2) = 0,036\,.$$

Стандартная ошибка равна 0,19. Значение стандартной ошибки для второго случая отличается примерно на эту же величину и, следовательно, мы не может признать существенность различия.

Альтернативный подход заключается в следующем. Найдем дисперсию для второго коэффициента, она равна 0,064. Таким образом, максимальная дисперсия разности коэффициентов (равна сумме дисперсий) составит 0,100, а стандартная ошибка — 0,32. Эта величина превышает действительную разность, равную 0,2. Отсюда опять следует, что различие коэффициентов не является существенным.

Вообще, если имеется ряд значений $S$ (даже для последовательностей различной протяженности), то можно суммировать их и проверить существенность всей совокупности, сопоставляя с суммой дисперсий отдельных последовательностей. Ценность этой процедуры основывается на том факте, что дисперсия суммы *независимых* переменных равна сумме их дисперсий. Она оказывается очень полезной в экспериментальной работе.

**4.22.** Предыдущий пример подчеркнул одну довольно разочаровывающую особенность коэффициента ранговой корреляции — ему свойственны относительно большие стандартные ошибки. Какой бы ни была величина $\tau$, стандартная ошибка имеет величину порядка $\sqrt{2/n}$. Этот недостаток характерен для большинства коэффициентов корреляции. Так, например, стандартная ошибка коэффициента парной корреляции в выборках, следующих нормальному закону распределения, равна $(1-\rho^2)/\sqrt{n}$ и, таким образом, имеет в общем величину порядка $1/\sqrt{n}$. Очевидно, что невозможно оценить коэффициент корреляции для генеральной совокупности очень точно, если ранжированный ряд состоит менее чем из 30 или 40 членов. В связи с этим необходима осторожность при приписывании реального смысла коэффициентам корреляции, подсчитанным для последовательностей небольшой протяженности, если нет нескольких выборочных величин.

Например, при ранжировании 32 объектов максимальная стандартная ошибка равна $\frac{1}{4}\sqrt{1-\tau^2}$. Если в этом случае $t$ близко к нулю, то на этой основе мы не можем определить генеральное $\tau$ в пределах, меньших чем удвоенная стандартная ошибка, т. е. $\pm 0,5$. Если же, допустим, $t$ равно 0,8, то пределы сужаются, однако область неопределенности, охватывающая действительное значение, все еще простирается от 0,5 до 1,0.

**4.23.** Как будет показано в следующей главе, нельзя значительно уменьшить максимальные пределы, получаемые на основе (4.9) или (4.11). Однако мы увидим, что весьма существенное улучшение возможно, если имеется первичное ранжирование и исследователь обладает терпением для выполнения необходимых расчетов. Если не принимать во внимание сказанное выше, то представляется невероятным, что возможно какое-либо простое улучшение без принятия предположений о существе переменной и ее генеральном распределении, как это и будет показано в гл. 9.

## Библиография

О распределении $\tau$ см. [48], [14], [9], [82]; о распределении $\rho$ см. [73], [55], [78], [44], [16]. Об общем совместном распределении коэффициентов корреляции см. [9], [13] и [39].

Более полные таблицы распределения $t$ для $n$, достигающие 40 единиц, опубликованы в [47].

Формула (4.10) является следствием более общей формулы, полученной Хоффдингом [39]. Если $r_s$ и $t$ есть выборочные значения $\rho$ и $\tau$, то

$$k = \frac{1}{n-2}\left[(n+1)\,r_s - 3t\right]$$

и

$$\operatorname{var} k \leqslant \frac{3}{n}\left(1-\rho_s^2\right),$$

где $\rho_s$ — коэффициент ранговой корреляции Спирмэна для непрерывной совокупности (см. 9.5). Для больших $n$ $k$ имеет тенденцию быть равным $r_s$. См. также [6], [31], [27], [28] и [74].

## *ГЛАВА 5.* ДОКАЗАТЕЛЬСТВА РЕЗУЛЬТАТОВ ГЛАВЫ 4

**5.1.** Формулы, приведенные в предыдущей главе, при проверке существенности $\tau$ и $\rho$, требуют четыре типа результатов, а именно:

а) определение точного распределения для невысоких значений $n$;

б) доказательство того, что распределение стремится к нормальному при больших $n$;

в) определение средних и дисперсий для ограничивающих распределений;

г) определение поправок на непрерывность.

Рассмотрим названные проблемы, сохраняя их порядок. В конце главы более детально проанализируем ненулевой случай.

### Точное распределение $\tau$ в нулевом случае

**5.2.** Если мы коррелируем фиксированную последовательность из $n$ членов с $n!$ возможными последовательностями (исключая связи), то получим одно и то же распределение (значений $S$. — *Прим. перев.*), какой бы ни была эта фиксированная последовательность; это справедливо для любых возможных ее вариантов. Поэтому мы не теряем общности доказательства, предположив, что нашей фиксированной последовательностью является натуральный ряд чисел $1, \ldots, n$. Пусть $u\,(n,\ S)$ есть число значений $S$ для совокупности из $n!$ возможных значений, полученных путем коррелирования этой последовательности со всеми возможными последовательностями.

Рассмотрим теперь, какое влияние оказывает введение нового члена $(n+1)$ в различные места такой последовательности от первого (т. е. предшествующего первому члену $p_1$) до последнего (т. е. следующего за членом $p_n$). Введение $(n+1)$ члена в начало ранжированного ряда добавит к $S$ величину, равную $-n$; введение его на второе место добавит $-(n-2)$; на третье $-(n-4)$ и т. д.; введение этого члена на последнее место добавляет величину $n$. Отсюда следует, что

$$u\,(n+1,\ S)=u\,(n, S-n)+u\,(n,\ S-n+2)+u\,(n,\ S-n+4)+ \\ +\ldots+u\,(n,\ S+n-4)+u\,(n,\ S+n-2)+ \\ +u\,(n,\ S+n). \tag{5.1}$$

Данная рекуррентная формула позволяет определить распределение частот для $S$ при числе членов $n+1$, когда известно $n$. Исходя из этого мы можем построить ряд распределений, начиная с простейших случаев, т. е. для $n = 2$, $n = 3$ и т. д.

**5.3.** Для практических целей эта процедура может быть упрощена. Так, для $n = 2$ имеется два значения $S$, а именно: $-1$ и $+1$. Если мы три раза запишем частоты одни под другими, со сдвигом вправо каждый раз на один шаг, то получим:

$$\begin{array}{cccc}
1 & 1 & & \\
 & 1 & 1 & \\
 & & 1 & 1 \\
\hline
1 & 2 & 2 & 1
\end{array}$$

Суммы по столбцам характеризуют частоты $S$ для $n = 3$. Значения $S$ находятся в диапазоне от $-\frac{1}{2}n(n-1)$ до $\frac{1}{2}n(n-1)$ с интервалом в две единицы, т. е. $-3$, $-1$, $+1$, $+3$.

Аналогично выпишем четыре раза ряд частот и соответствующие суммы для $n = 3$:

$$\begin{array}{ccccccc}
1 & 2 & 2 & 1 & & & \\
 & 1 & 2 & 2 & 1 & & \\
 & & 1 & 2 & 2 & 1 & \\
 & & & 1 & 2 & 2 & 1 \\
\hline
1 & 3 & 5 & 6 & 5 & 3 & 1
\end{array}$$

получим частоты для значений $S$ от $-6$ до 6 с интервалом в две единицы.

Ценность этого правила, учитывая содержание предыдущего раздела, очевидна. Так, любой совокупности значений $S$ при заданном $n$ будет корреспондировать распределение $S$ с $n+1$ членами, причем значения $S$ увеличены на $-n$, $-(n-2)$, ..., $(n-2)$, $n$. Все, что следует выполнить, — это выписать эти частоты (т. е. частоты распределения для $n$ членов. — *Прим. перев.*); один ряд для $-n$, второй для $-(n-2)$ и т. д. и суммировать их.

**5.4.** В свою очередь и эта процедура может быть упрощена. Составим следующим образом числовой треугольник:

| $n$ | Частоты | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | | | | | | | | | |
| 2 | 1 | 1 | | | | | | | | |
| 3 | 1 | 2 | 2 | 1 | | | | | | |
| 4 | 1 | 3 | 5 | 6 | 5 | 3 | 1 | | | |
| 5 | 1 | 4 | 9 | 15 | 20 | 22 | 20 | 15 | 9 | 4 | 1 |

и т. д.

В этом расположении чисел элемент, находящийся в строке $r$, есть сумма числа, стоящего непосредственно над ним, и $(r-1)$ членов слева от этого верхнего числа; например, в пятой строке стоит число 22, получим: $22 = 3 + 5 + 6 + 5 + 3$. Частоты значений $S$ для $n = r$ получают как числа строки $r$. Описанный метод был использован для получения частот табл. 1 приложения.

5.5. Распределение $S$ симметричное; для любой последовательности, дающей определенное значение $S$, существует сопряженный ранжированный ряд, который дает $-S$. Поэтому, как и следует ожидать, среднее значение $S$ равно нулю. То, что $S$ находится в диапазоне от $-\frac{1}{2}n(n-1)$ до $\frac{1}{2}n(n-1)$, очевидно из того, что экстремальные значения получают тогда, когда последовательность находится в обратном порядке к натуральному ряду чисел или сама является натуральным рядом. Далее, интервал между последовательными значениями $S$ для заданного $n$ равен 2. Это следует из (5.1), или, что, возможно, проще — из учета того факта, что перемена местами двух членов в последовательности изменяет величину $P$ на $+1$ или $-1$. Отсюда разность между $P$ и $Q$*, которая равна $S$, изменяется на $+2$ или $-2$.

## Стремление распределения $\tau$ к нормальному в нулевом случае

5.6. Нам остается определить дисперсию $S$ и доказать, что распределение стремится к нормальному при росте $n$.

Приняв обозначения гл. 2, запишем:

$$c_{ij} = a_{ij} b_{ij}, \tag{5.2}$$

где $a_{ij}$ и $b_{ij}$ — оценки, относящиеся к двум последовательностям. Пусть

$$c = \sum_{i,j=1}^{n} c_{ij}. \tag{5.3}$$

Тогда

$$c = 2S. \tag{5.4}$$

Теперь

$$\sum_{l=1}^{n} a_{il} = n + 1 - 2i, \tag{5.5}$$

поскольку эта величина является оценкой $i$-го члена и равна $(n-i)-(i-1)$.

Далее

$$\sum_{i,l=1}^{n} a_{il}^2 = n(n-1), \tag{5.6}$$

* См. раздел 1.9. — *Прим. перев.*

так как $a^2$ равно $+1$, а сумма представляет собой просто число возможных путей образования пар из $m$ членов, причем каждая пара учитывается дважды: один раз как $XY$, другой — $YX$. Из (5.5) вытекает, что

$$\sum_{i,\,l=1}^{n} a_{il}=\sum_{i=1}^{n}(n+1-2i)=n(n+1)-2\sum_{i=1}^{n} i=0,$$

как и следовало ожидать.

Найдем также

$$\begin{aligned}\sum_{i,\,l,\,t=1}^{n} a_{il}\,a_{it}&=\sum_{i,\,l} a_{il}(n+1-2i)=\sum_{i}(n+1-2i)^2=\\&=\sum(n+1)^2-4(n+1)\sum i+4\sum i^2=\\&=n(n+1)^2-2n(n+1)^2+\frac{2}{3}n(n+1)\times\\&\times(2n+1)=\frac{1}{3}n(n^2-1).\end{aligned}\qquad(5.7)$$

Обозначим теперь символом $E$ среднюю, полученную при суммировании по всем возможным перестановкам:

$$E(c)=E\sum_{i,\,j=1}^{n} a_{ij}\,b_{ij}=\sum_{i,j}E(a_{ij}\,b_{ij}).$$

Поскольку $a_{ij}$ и $b_{ij}$ являются независимыми и фиксированное значение одной из них берется со всеми возможными значениями другой, а также поскольку среднее значение $a_{ij}$ или $b_{ij}$ равно нулю, имеем:

$$E(c)=0.\qquad(5.8)$$

Таким образом, подтверждается, что средняя для $c$ (или $S$) равна нулю. Для определения дисперсии $c$ нам необходимо найти

$$\begin{aligned}E(c^2)=E\Big[\sum_{i,\,j}(a_{ij}\,b_{ij})\Big]^2&=E\,[\sum(a_{ij}^2\,b_{ij}^2)+\Sigma'(a_{ij}\,b_{ij}\,a_{ik}\,b_{ik})+\\&+\Sigma''(a_{ij}b_{ij}a_{kl}b_{kl})],\end{aligned}\qquad(5.9)$$

где $i\neq j\neq k\neq l$. Здесь суммирование производится отдельно для выделенных членов, причем любой член при разложении $(\Sigma a_{ij}b_{ij})^2$ может встретиться более одного раза.

1. Член $E\Sigma''$ обращается в нуль. Для того чтобы продемонстрировать это, достаточно показать, что $E\Sigma'' a_{ij}a_{kl}$ принимает нулевое значение. Теперь получим (символ $\Sigma$ указывает на суммирование по всем переменным):

$$\begin{aligned}\Sigma\,a_{ij}\,a_{kl}&=\Sigma''\,a_{ij}\,a_{kl}+\Sigma'\,a_{ij}\,a_{il}+\Sigma'\,a_{ij}\,a_{ki}+\Sigma'\,a_{ij}\,a_{jl}+\\&+\Sigma'\,a_{ij}\,a_{kj}+\Sigma a_{ij}\,a_{ij}+\Sigma a_{ij}\,a_{ji}.\end{aligned}$$

При определении математических ожиданий учтем, что члены правой стороны этого выражения, не считая первого, обращаются в нуль в силу соотношения $a_{ij}=-a_{ji}$. Член, стоящий в левой стороне, обращается в нуль, поскольку $E\Sigma a_{ij}a_{kl} = E\Sigma a_{ij}E\Sigma a_{kl}$ и $\Sigma a_{ij} = 0$ (суммирование производится по всем переменным). Следовательно, первый член правой стороны также равен нулю.

2. Теперь рассмотрим сумму $E\Sigma a_{ij}^2 b^2_{ij}$. Отдельные слагаемые получают как квадрат $a_{ji}\,b_{ij}$ или как произведение этого члена с $a_{ij}b_{ji}$. Таким образом, общая сумма равна удвоенной сумме членов $a^2_{ij}b^2_{ij}$, где $i, j = 1, \ldots, n$. Эта сумма состоит из $n\,(n-1)$ членов и, следовательно, математическое ожидание равно:

$$2n\,(n-1)\,E\,(a_{12}^2\,b_{12}^2) = 2n\,(n-1)\,E\,(a_{12}^2)\,E\,(b_{12}^2) = \\ = 2\Sigma a_{ij}^2\,\Sigma b_{ij}^2\,/n\,(n-1).$$

3. Аналогично этому определим второй член правой стороны выражения (5.9). Он равен:

$$\frac{4}{n\,(n-1)\,(n-2)}\,\Sigma'\,a_{ij}\,a_{ik}\,\Sigma'\,b_{ij}\,b_{ik}.$$

Произведение $a_{ij}a_{ik}b_{ij}b_{ik}$ может быть получено четырьмя различными путями, поэтому если фиксировать $i$, $j$, $k$ для значений $a$, то членам $b$ можно приписать индексы

$$(ij,\ ik),\ (ji,\ ik),\ (ij,\ ki)\ \text{и}\ (ji,\ ki).$$

Имеется $n\,(n-1)\,(n-2)$ возможностей для приписывания различных индексов, если последовательность содержит $n$ членов. Читателю предлагается выписать их для случая $n = 4$, если его настораживает этот вывод.

Кроме того,

$$\Sigma'\,a_{ij}\,a_{ik} = \Sigma a_{ij}\,a_{ik} - \Sigma a_{ij}^2\,.$$

Подставляя результаты в (5.9), получим:

$$E\,(c^2) = \frac{4}{n\,(n-1)\,(n-2)}\,[\Sigma a_{ij}\,a_{ik} - \Sigma a_{ij}^2\,]\,[\Sigma b_{ij}\,b_{ik} - \Sigma b_{ij}^2\,] + \\ + \frac{2}{n\,(n-1)}\,\Sigma a_{ij}^2\,\Sigma b_{ij}^2\,. \qquad (5.10)$$

Теперь, воспользовавшись результатами (5.6) и (5.7), определяем:

$$E\,(c^2) = \frac{4}{n\,(n-1)\,(n-2)}\left[\frac{1}{3}\,n\,(n^2-1) - n\,(n-1)\right]^2 + \\ + \frac{2}{n\,(n-1)}\,[n\,(n-1)]^2 = \frac{2n\,(n-1)\,(2n+5)}{9}\,. \qquad (5.11)$$

Отсюда следует, что

$$E(S^2) = \operatorname{var} S = \frac{n(n-1)(2n+5)}{18}, \tag{5.12}$$

$$\operatorname{var} t = \frac{2(2n+5)}{9n(n-1)}. \tag{5.13}$$

Выражение (5.12) и есть результат, который был приведен в (4.2).

5.7. Если в последовательностях находятся связи, то уравнение (5.10) сохраняет справедливость, однако выражения (5.6) и (5.7) требуют уточнений. Так, для (5.6) имеем:

$$\Sigma a_{ij}^2 = n(n-1) - \sum_t t(t-1), \tag{5.14}$$

где суммирование $\sum_t$ производится по всем связям. Это следует из того, что пары связанных рангов $a_{ij} = 0$ и, следовательно, сумма квадратов вкладов от связанного ряда имеет тот же вид, что и от последовательности в целом.

Для (5.7) получаем:

$$\Sigma a_{ij} a_{ik} = \frac{1}{3} n(n^2 - 1) - \frac{1}{3} \sum_t t(t^2 - 1). \tag{5.15}$$

Этот результат не совсем очевиден. Рассмотрим влияние связывания нескольких рангов. Вклад в сумму (т. е. в левую сторону (5.15)) не будет изменяться, если подписной индекс $i$ исключить из этой совокупности. Если $i$ ввести в нее, а $j$ и $k$ исключить, то снова вклад останется неизменным. Если включить и $j$ и $k$, то это не приведет к увеличению вклада, поэтому нам следует вычесть величину $\frac{1}{3} t(t^2 - 1)$ Если же один из этих подписных индексов включить, а один исключить, то вклад останется неизменным. Поскольку он равен нулю в исходном случае при отсутствии связей, каждой возможной паре один раз приписывается $+1$ и один раз $-1$.

Если (5.14) и (5.15) подставить в (5.10), то найдем, что дисперсия $S$, когда последовательности имеют связи $t$ и $u$, равна:

$$\begin{aligned}\operatorname{var} S = {} & \frac{1}{18}\Big[n(n-1)(2n+5) - \sum_t t(t-1)(2t+5) - \\ & - \sum_u u(u-1)(2u+5)\Big] + \\ & + \frac{1}{9n(n-1)(n-2)}\,[\Sigma t(t-1)(t-2)]\,[\Sigma u(u-1)(u-2)] + \\ & + \frac{1}{2n(n-1)}\,[\Sigma t(t-1)]\,[\Sigma u(u-1)].\end{aligned} \tag{5.16}$$

Этот результат приведен в уравнении (4.3). Уравнения (4.4), (4.5) и (4.6) непосредственно вытекают из него.

5.8. При доказательстве того, что распределение $S$ стремится к нормальному, мы будем следовать процедуре, которая с незначительными изменениями доказывает нормальность $S(d^2)$ и представляет собой введение в более общий анализ, относящийся к ограниченным формам общего коэффициента ранговой корреляции (см. гл. 2), развитый в [9].

Мы докажем, что моменты распределения $S$ стремятся к моментам нормального распределения. Это следует из положения о том, что данное распределение стремится к нормальному, известного как вторая предельная теорема (см. [58, 4.29—4.30]). Поскольку распределение $S$ симметрично, нечетные моменты относительно средней равны нулю. Следовательно, мы должны только показать, что для четных моментов

$$\mu_{2r} = \frac{(2r)!}{2^r\, r!} (\mu_2)^r. \tag{5.17}$$

Рассмотрим среднюю величины $(\Sigma a_{ij} b_{ij})^{2r}$. При разложении этого выражения получим следующие члены, которые имеют вид:

$$\Sigma' a_{ij}\, a_{kl}\, a_{mn} \dots b_{ij}\, b_{kl}\, b_{mn}.$$

Суммирование здесь проводится по индексам, исключая некоторые величины (например, при $i \neq k$); мы можем заменить эти суммы выражениями, характеризующими полное суммирование. Тогда разложенное выражение будет, кроме уже указанных числовых членов, содержать члены такого вида:

$$\Sigma a_{ij}\, a_{kl}\, a_{mn} \dots \Sigma b_{ij}\, b_{kl}\, b_{mn}, \tag{5.18}$$

где одни подписные индексы могут совпадать или быть «связанными», а другие будут различными или «свободными». Рассмотрим член, в котором $4r$ подписных индекса у $a$ являются связанными в пары, т. е.

$$\Sigma a_{ij}\, a_{ik}\, a_{lm}\, a_{ln}. \tag{5.19}$$

Тогда имеется $3r$ независимых индекса. Величина $\Sigma a_{ij} a_{ik}$ теперь имеет порядок $\frac{1}{3} n^3$, следовательно, (5.19) имеет порядок $\left(\frac{1}{3} n^3\right)^r$. В выражении $(\Sigma a_{ij} b_{ij})^{2r}$ такого рода члены будут появляться с частотой, равной числу путей связывания $2r$ пар индексов. Это равно произведению трех величин, а именно: 1) числа вариантов выбора $r$ величин $a$ из $2r$, т. е. $\binom{2r}{r}$; 2) числа вариантов связи выбранных величин с остальными $r$ факторами, т. е. $r!$; 3) в связи с тем, что любой подписной индекс может быть связанным, возникает величина $2^r$. Числовой коэффициент тогда равен:

$$\binom{2r}{r} r!\, 2^r = \frac{(2r)!\, 2^r}{r!}$$

Кроме того, если $3r$ индексов фиксированы, то остальные члены последовательности могут варьировать $(n-3r)!$ путями. Отсюда $\mu_{2r}$ содержит член

$$\frac{(n-3r)!}{n!}\,\frac{2^r(2r)!}{r!}\left(\frac{1}{3}\,n^3\right)^{2r}\sim\frac{(2r)!}{2^r r!}\left(\frac{4}{9}\,n^3\right)^r.$$

Однако $\mu_2$ имеет величину $\frac{4}{9}n^3$ (для этого надо подставить $r=1$), следовательно, $\mu_{2r}$ содержит член

$$\frac{(2r)!}{2^r r!}\,(\mu_2)^r.$$

Доказательство будет полным, если мы покажем, что все остальные члены имеют меньший порядок относительно $n$.

Если какой-либо член (5.18) содержит пару индексов, каждый из которых не появляется еще где-либо, то он равен нулю, поскольку $\Sigma a_{ij}=0$. Рассмотрим член, в котором связано более двух индексов. Тогда суммирование (5.19) производится не более чем по $3r-1$ индексам и порядок результата не может превысить величину $(n^{3r-1})^2$. Если $3r-1$ или менее индексов фиксированы, то эта величина не превосходит

$$\frac{(n-3r+1)!}{n!}\,n^{6r-2}\sim n^{3r-1}$$

и следовательно, оказывается меньше уже найденного члена[1].

## Распределение $\rho$ в нулевом случае

**5.9.** Действительное распределение $\rho$ труднее определить, чем распределение $\tau$; метод конструирования распределения для $n+1$ единиц при наличии распределения для $n$ единиц неизвестен. Рассмотрим следующее расположение чисел:

$$\begin{array}{ccccccc}
0 & 1 & 2 & \ldots & n-3 & n-2 & n-1\\
-1 & 0 & 1 & \ldots & n-4 & n-3 & n-2\\
-2 & -1 & 0 & \ldots & n-5 & n-4 & n-3\\
\vdots & \vdots & \vdots & \ldots & \vdots & \vdots & \vdots\\
-(n-2) & -(n-3) & -(n-4) & \ldots & -1 & 0 & 1\\
-(n-1) & -(n-3) & -(n-3) & \ldots & -2 & -1 & 0
\end{array}\qquad(5.20)$$

Любые допустимые расхождения между последовательностью 1, 2,… …, $n$ и произвольным порядком чисел задается путем выбора члена из этой таблицы, в которой два одинаковых члена не встречаются ни

[1] Простые доказательства приводятся в [68 и 83]. Эти доказательства исходят из более тонких идей и не обобщаются для других ранговых коэффициентов, однако они непосредственно определяют моменты распределения любого порядка.

в столбце, ни в строке. Рассмотрим теперь числа:

$$\begin{matrix} a^0 & a^1 & a^4 & \ldots & a^{(n-3)^2} & a^{(n-2)^2} & a^{(n-1)^2} \\ a^1 & a^0 & a^1 & \ldots & a^{(n-4)^2} & a^{(n-3)^2} & a^{(n-2)^2} \\ \vdots & \vdots & \vdots & \ldots & \vdots & \vdots & \vdots \\ a^{(n-2)^2} & a^{(n-3)^2} & a^{(n-4)^2} & \ldots & a^1 & a^0 & a^1 \\ a^{(n-1)^2} & a^{(n-2)^2} & a^{(n-3)^2} & \ldots & a^4 & a^1 & a^0 \end{matrix} \qquad (5.21)$$

Показатели степеней, приведенные в (5.21), являются квадратами элементов, показанных в (5.20); если мы обобщим (5.21), то получим совокупность величин $S(d^2)$. Под «обобщением» мы понимаем построение рядов путем выбора из таблицы $n$ коэффициентов (5.21) $n!$ возможными способами так, что ни один из них дважды не входит в один столбец или строку. Умножение их на каждую совокупность коэффициентов и суммирование дает $n!$ результатов. Этот метод был применен для получения распределений, образующих основу табл. 2 приложения.

**5.10.** Некоторые простые свойства распределения $S(d^2)$ очевидны уже при элементарном анализе. Во-первых, величина $S(d^2)$ должна быть четной; так как $\Sigma d = 0$, число нечетных значений $d$ и, следовательно, $d^2$ должно быть четным. Во-вторых, распределение симметричное, поскольку любое значение $S(d^2)$ корреспондирует величине $\frac{1}{3}(n^3-n)-S(d^2)$, полученной для сопряженной последовательности (2.9). В-третьих, средняя этого распределения равна $\frac{1}{6}(n^3-n)$. В-четвертых, оно простирается от 0 до $\frac{1}{3}(n^3-n)$.

**5.11.** Мы можем определить дисперсию $S(d^2)$, так, как это показано в 5.6. Из того подхода, которым было определено (5.10), содержащее оценки $a$ и $b$, становится ясным, что это уравнение справедливо также для $c$, когда эти величины относятся к коэффициенту $\rho$ Спирмэна. Поэтому без потери общности мы можем написать:

$$a'_{ij} = j - i. \qquad (5.22)$$

Теперь непосредственно получаем:

$$\sum_{j=1}^{n} a'_{ij} = \frac{1}{2} n(n+1-2i) \qquad (5.23)$$

и

$$\sum_{i,j=1}^{n} a'^2_{ij} = \sum (j-i)^2 = 2n \sum_{j=1}^{n} j^2 - 2\left(\sum_{j=1}^{n} j\right)^2 = \frac{1}{6} n^2 (n^2-1),$$

а также

$$\sum_{i,j,k=1}^{n} a'_{ij} a'_{ik} = \frac{1}{12} n^3 (n^2-1). \qquad (5.24)$$

Подстановка в (5.10) теперь дает

$$E\,(c^2)=\frac{n^4\,(n-1)\,(n+1)^2}{36}\,. \tag{5.25}$$

Вспомнив, что

$$\rho=1-\frac{6S\,(d^2)}{n^3-n}$$

и что в соответствии с (2.9)

$$c=\frac{1}{6}\,n^2\,(n^2-1)-nS\,(d^2),$$

находим на основе (5.25)

$$\operatorname{var}\rho=E\,(\rho^2)=\frac{1}{n-1}\,, \tag{5.26}$$

что отвечает формуле (4.7)[1].

**5.12.** Аналогичными методами может быть показано, что четвертый момент $\rho$ определяется формулой

$$\mu_4=\frac{3\,(25n^3-38n^2-35n+72)}{25n\,(n+1)\,(n-1)^3}\,. \tag{5.27}$$

Третий момент в силу симметрии распределения, естественно, равен нулю. Для $\mu_6$ и $\mu_8$ выражения известны (см. [16]), однако они довольно громоздки.

В нормальном распределении

$$\mu_4-3\mu_2^2=0. \tag{5.28}$$

Для распределения $\rho$ на основе (5.27) и (5.26) находим:

$$\mu_4-3\mu_2^2=-\frac{114n^2+30n-216}{25n\,(n+1)\,(n-1)^3}= \tag{5.29}$$

$$=-\frac{4{,}56}{n^3}+0\,(n^{-4}). \tag{5.30}$$

Эта величина меньшего порядка, чем $\mu_2^2$. Соответственно имеем:

$$\frac{\mu_4}{\mu_2^2}-3=-\frac{4{,}56}{n}+0\,(n^{-2}). \tag{5.31}$$

Отсюда видно, что для больших $n$ распределение $\rho$ близко к нормальному.

---

[1] То, что эта формула не требует модификации для связанных рангов, следует из (7.14), где $m=2$, имея в виду, что (6.6) должно быть модифицировано для случая, когда имеются связи. В противном случае можно показать, что при $c$, определенном соответственно (5.3), и оценках $x_j-x_i$ и $y_j-y_i$, $E\,(c^2)=4n^2\Sigma\,(x-\bar{x})^2\Sigma\,(y-\bar{y})^2/(n-1)$, откуда следует (5.26), вне зависимости от того, являются ранги связанными или нет.

Имеющиеся таблицы точных значений и нормальных приближений для большинства обычных разработок являются достаточными. Для более точного анализа в [16] получены обобщения, дающие распределение как $\rho$, так и $\tau$ с точностью до четвертого знака.

## Совместное распределение $\tau$ и $\rho$

**5.13.** Тенденция распределения $\rho$ к нормальному может быть продемонстрирована путем некоторого изменения доказательства нормальности распределения $\tau$, приведенного в (5.8). Оба результата являются частными случаями более общей теоремы, развитой в [9], которую мы теперь докажем.

Совместное распределение $\tau$ и $\rho$ стремится к двумерному нормальному распределению при стремлении $n$ к бесконечности. В самом деле, при наличии некоторых неограничивающих условий любые два коэффициента общего типа, определенные в 2.2, имеют тенденцию к совместному двумерному нормальному распределению.

Предположим, что $a$, $a'$ являются величинами двух таких коэффициентов, аналогично $b$ и $b'$. Теперь покажем, что моменты совместного распределения корреспондирующих значений $c$ и $c'$ также стремятся к такой же нормальной двумерной форме.

Моменты порядка $p$ этого совместного распределения равны суммам членов, содержащих

$$\Sigma' a_{gh} a_{ij} a_{hl} \ldots \Sigma' b_{rs} b_{tu} b_{vw} \ldots, \tag{5.32}$$

где группы подписных индексов в пределах сумм $\Sigma'$ могут быть связанными или свободными.

Каждая сумма $\Sigma'$ содержит произведение $p$ оценок, которые могут принадлежать той или другой системе. В свою очередь каждая такая сумма $\Sigma'$ является линейной комбинацией соответствующих $\Sigma$, имеющих те же подписные индексы, и других $\Sigma$, в которых содержатся дополнительные связанные индексы. Ни одна из сумм $\Sigma$ не может содержать пары несвязанных индексов, представляющих одну оценку, поскольку тогда она превращается в нуль в связи с тем, что $\Sigma a_{ij} = 0$.

Рассмотрим сперва моменты четного порядка. Пусть $p = 2m$. Возьмем $\Sigma$, в которой $2m$ оценок расчленено на $m$ пар, каждая из которых имеет один связанный индекс, так что всего имеется $3m$ независимых индексов, а именно:

$$\Sigma a_{ij} a_{ik} a_{lr} a_{ls} a_{tu} a_{tv}. \tag{5.33}$$

Это выражение может быть переписано как

$$(\Sigma a_{ij} a_{ik})^{\lambda} (\Sigma a_{ij} a'_{ik})^{\mu} (\Sigma a'_{ij} a'_{ik})^{\nu}, \tag{5.34}$$

где $\lambda + \mu + \nu = m$ и $\lambda$, $\mu$, $\nu$ — показатели, характеризующие число раз, когда оценки объединялись в пары в указанных сочетаниях.

Предположим, что наибольшее значение $a_{ij}$ приравнено единице. Примем теперь условие, согласно которому $\Sigma a_{ij}\, a_{ik}$ имеет величину порядка $n^3$ вне зависимости от того, принадлежат или нет $a_{ij}$ и $a_{ik}$ к одной или различным системам оценок. Это условие, в частности, удовлетворяется коэффициентами $\tau$ и $\rho$, когда $\max a_{ij} = 1$. Представляется, что данное условие приводит к тому, что суммы ($\Sigma$) указанного выше вида равны величинам порядка $n^{3m}$.

Другие варианты связывания индексов дают суммы меньшего порядка. Порядок величины этого выражения не уменьшается и при замене каждого $a_{ij}$ на $+1$; следовательно, если оказываются связанными и другие индексы, то порядок суммы становится меньшим, чем $n^{3m}$, поскольку имеется меньше $3m$ суммирований от 1 до $n$. Отсюда следует, что доминирующий член в $\Sigma'$ соответствует сумме ($\Sigma$), имеющей тот же самый ряд индексов.

Более того, каждая не сходящаяся к нулю сумма, охватывающая $3m$ независимых индекса, может представлять собой лишь перестановку типа (5.32), тогда как суммы с более чем $3m$ различными индексами все должны быть равными нулю. Это становится ясным при рассмотрении того, как $3m$ индексов может быть расположено для $2m$ оценок. Для начала разместим в случайном порядке $3m$ различных индекса по $4m$ имеющимся местам. По крайней мере $m$ оценок получат свой полный комплект индексов, причем все индексы будут различными. Более $m$ таких укомплектованных оценок не может быть, поскольку если $\Sigma$ не сводится к нулю, то один индекс каждой пары может быть связан. Это достигается только при повторении одного индекса из каждой пары в каждом из оставшихся мест, которое должно быть заполнено, а их имеется только $m$. Таким образом, мы приходим к перестановке типа $\Sigma$, которая обсуждалась выше. Если бы мы начали с более чем $3m$ различных индексов, то не осталось бы достаточного числа незанятых мест для того, чтобы предотвратить появление по крайней мере одной оценки с парой свободных индексов. Отсюда все $\Sigma$ с более чем тремя индексами должны обратиться в нули.

Любой смешанный момент равен сумме членов, подобных выражению

$$\frac{(n-f)!}{n!} A\Sigma' a_{ij}\, a_{kl} \ldots \Sigma' b_{rs}\, b_{tu} \ldots,$$

где $f$ — число независимых индексов в $\Sigma'$, а $A$ — коэффициент, равный единице, коль скоро затрагивается $n$ членов последовательности. Из сказанного выше следует, что максимальное значение $f$ равно $3m$; в данном случае этот член имеет порядок $n^{-3m} \times n^{3m} \times n^{3m} = n^{3m}$. Когда $f \leqslant 3m - 1$, порядок рассматриваемого члена не превышает величину $n^{-3m+1} \times (n^{3m-1})^2 = n^{3m-1}$, и, следовательно, подобными членами можно пренебречь. Выпишем теперь выражения

$$\left.\begin{aligned} h_{11} &= \Sigma a_{ij}\, a_{ik}\, \Sigma b_{iu}\, b_{iv} \\ h_{12} &= \Sigma a_{ij}\, a'_{ik}\, \Sigma b_{tu}\, b'_{tv} \\ h_{22} &= \Sigma a'_{ij}\, a'_{ik}\, \Sigma b'_{iu}\, b'_{iv} \end{aligned}\right\}. \tag{5.35}$$

Если пренебречь наименьшим по величине членом, то четный смешанный момент $\mu_{rs}$, где $r+s=2m$, определяется суммой следующих членов:

$$n^{-3m} A_{\lambda\mu\nu} h_{11}^{\lambda} h_{12}^{\mu} h_{22}^{\nu},$$
$$2\lambda+\mu=r, \quad \mu+2\nu=s \tag{5.36}$$

для всех возможных значений $\lambda$, $\mu$ и $\nu$. Коэффициенты $A_{\lambda\mu\nu}$ определяются исходя из следующего соображения. Рассмотрим $\Sigma$, последовательность индексов которой такова, что она может быть разложена на множители $(\Sigma a_{ij}a_{ik})^{\lambda}(\Sigma a_{ij}a'_{ik})^{\mu}(\Sigma a'_{ij}a'_{ik})^{\nu}$. Пары индексов могут быть переставлены $r!s!$ способами в пределах совокупности оценок типа $a$ и $a'$, однако $\lambda!\ (2!)^{\lambda}\ \mu!\nu!\ (2!)^{\nu}$ из них дают почти такие же $\Sigma$. Парные индексы каждой оценки могут быть также переупорядочены $2^{2m}$ способами без изменения результатов.
Отсюда

$$A_{\lambda\mu\nu}=\frac{r!\,s!\,2^{2m}}{\lambda!\,\mu!\,\nu!\,2^{\lambda+\nu}}=\frac{r!\,s!\,2^{m+\mu}}{\lambda!\,\mu!\,\nu!}. \tag{5.37}$$

Из (5.35), (5.36) и (5.37) следует, что расчет $\mu_{rs}$ равносилен определению коэффициента $t_1^r$ и $t_2^s$ в

$$\frac{2^m}{n^{3m}\,m!}(h_{11}t_1^2+2h_{12}t_1t_2+h_{22}t_2^2)^m. \tag{5.38}$$

Рассмотрим теперь нечетные моменты. Для $\rho$ и $\tau$ они обращаются в нуль в силу симметрии, однако можно показать, что даже в более общем случае ими можно пренебречь. Сумма $(\Sigma)$, содержащая $2m+1$ оценок, не может охватывать более $3m+1$ индексов. Это следует из того же аргумента, который мы приводили выше для четных моментов. Отсюда порядок величины $(2m+1)$-го момента равен самое большее $n^{3m+1}$. Как было показано, $2m$-е моменты должны быть порядка $n^{3m}$. Порядок $(2m+1)$-го момента не превышает $n^{\frac{3}{2}(2m+1)}\times n^{-\frac{1}{2}}$. Поэтому нечетные моменты имеют порядок меньший (на множитель $n^{-\frac{1}{2}}$), чем четные моменты.

Наконец, из (5.38) следует, что моменты $\mu_{rs}$ есть коэффициенты $t_1^r$, $t_2^s$ в

$$\exp\frac{2}{n^3}(h_{11}t_1^2+2h_{12}t_1t_2+h_{22}t_2^3), \tag{5.39}$$

а это производящая функция моментов для двумерного нормального распределения (см. пример 15.1 в [58, т. 1]). Этот результат доказан.

5.14. Представляет некоторый интерес изучение коэффициента корреляции между $\rho$ и $\tau$. Точно таким же путем, как было выведено

(5.10), находим среднюю $E(cc')$:

$$E(cc') = \frac{4}{n(n-1)(n-2)}[\Sigma a_{ij} a'_{ik} - \Sigma a_{ij} a'_{ij}] \times$$
$$\times [\Sigma b_{ij} b'_{ik} - \Sigma b_{ij} b'_{ij}] +$$
$$+ \frac{2}{n(n-1)} \Sigma a_{ij} a'_{ij} \Sigma b_{ij} b'_{ij}. \tag{5.40}$$

В частном случае, когда $c$ относится к $S$ и $c'$ к $S(d^2)$, получим:

$$\sum_{i,j,k} a_{ij} a'_{ik} = \frac{1}{6} n^2 (n^2 - 1), \tag{5.41}$$

$$\sum_{i,j} a_{ij} a'_{ij} = \frac{1}{3} n (n^2 - 1). \tag{5.42}$$

После подстановки в (5.40) имеем:

$$E(cc') = \frac{1}{9} n^2 (n-1)(n+1)^2. \tag{5.43}$$

Таким образом, коэффициент корреляции между $S$ и $S(d^2)$, который также является коэффициентом корреляции между $\tau$ и $\rho$, определяется выражением:

$$\frac{E(cc')}{\sqrt{E(c^2) E(c'^2)}} = \frac{2(n+1)}{\sqrt{2n(2n+5)}}. \tag{5.44}$$

При больших $n$ эта величина стремится к единице. Даже для средних значений $n$ она довольно высока. Для $n = 5$ она составит 0,980, а для $n = 20$ равна 0,990.

### Поправка на непрерывность

5.15. Вернемся теперь к проблеме поправки на непрерывность, правило для определения которой было изложено в 4.12 и 4.15.

а. Рассмотрим вариант, когда одна последовательность является несвязанной, а вторая имеет связи и в крайнем случае может быть представлена дихотомией. Можно допустить, что несвязанная последовательность представлена натуральным рядом чисел, а вторая имеет любой произвольный порядок. Если мы поменяем местами пару соседних членов несвязанной последовательности, то это окажет влияние только на те оценки, которые включают оба эти члена. Соответствующие ранги во второй последовательности являются несвязанными либо связанными. В первом случае величина $S$ изменится на две единицы, во втором — она не изменится. Если во второй последовательности много связей (исключая случай, когда весь ряд полностью связан), то должна быть одна перестановка соседних членов в первой последовательности, которая изменяет величину $S$ на две единицы. Таким образом, все интервалы между последовательными значениями $S$ в распределении этой величины равны двум единицам и соответствующая поправка на непрерывность равна единице.

б. Если теперь первая последовательность целиком состоит из связей протяженностью $t$, а вторая представляет собой дихотомию, то перемена местами двух соседних членов из разных связанных групп может изменить $S$ (и изменяет при наличии некоторых последовательностей второй переменной) самое большее на $2t$. Поправка на непрерывность равна $t$.

в. Если обе переменные представлены дихотомиями, то, как показано в 3.14, $S = ad - bc$. Наименьшее изменение, которое может произойти, связано с увеличением или уменьшением $a$ на единицу. В этом случае (скажем, при увеличении) рост значения $S$ составит:

$$(a+1)(d+1)-(b-1)(c-1)-(ad-bc) = a+b+c+d=n.$$

Поправка на непрерывность равна, следовательно, $\frac{1}{2}n$.

г. Когда оба ранжированных ряда содержат связи, то невозможно сформулировать общее правило для определения поправки на непрерывность. Если эта сторона анализа является существенной, то необходимо специальное исследование, подобное приведенному в примере 4.5.

**5.16.** Для определения поправки на непрерывность для $\rho$ рассмотрим последовательность натуральных чисел и дихотомию, состоящую из $k$ и $n-k$ членов со средними рангами, равными $\frac{1}{2}(k+1$ и $\frac{1}{2}(n+k+1)$. Если два члена первой последовательности есть $x$ и $y$, то их перестановка приводит к перестановке двух членов второй последовательности, по одному на каждую часть дихотомии*, тогда изменение в $S$ $(d^2)$ составит:

$$\left[x-\frac{1}{2}(k+1)\right]^2+\left[y-\frac{1}{2}(n+k+1)\right]^2-\left[x-\frac{1}{2}(n+k+1)\right]^2-$$
$$-\left[y-\frac{1}{2}(k+1)\right]^2=n(x-y).$$

Таким образом, имеется одна перестановка соседних членов ($y = x + 1$), которая увеличит или уменьшит $S$ $(d^2)$ на $n$. Соответствующая поправка равна $\frac{1}{2}n$.

Если члены первой последовательности все связаны, а связи охватывают $t$ членов, то минимальное (и достижимое) изменение равно $nt$, давая поправку $\frac{1}{2}nt$.

Наконец, если первая последовательность является дихотомией, то минимальное изменение равно $\frac{1}{2}n^2$ и поправка составит $\frac{1}{4}n^2$.

---

* Последнее, естественно, не всегда имеет место. — *Прим. перев.*

**Ненулевой случай**

**5.17.** Обратимся теперь к более сложному случаю — к корреляции в выборке и генеральной совокупности. Обозначим генеральное значение коэффициента корреляции через $\tau$, а выборочное его значение — через $t$. Прежде всего мы докажем, что среднее значение $t$ для всех возможных выборок есть $\tau$.

Рассмотрим $\binom{N}{n}$ выборок (т. е. выборок из совокупности объемом $N$, состоящих из $n$ единиц). Каждая специфическая пара членов совокупности будет содержаться в $\binom{N-2}{n-2}$ выборках, так что все пары встречаются с одинаковой частотой в общей совокупности выборок. Таким образом, итоговая оценка для всех выборок составит $\binom{N-2}{n-2}$-кратную оценку для этой совокупности, скажем $\Sigma$. Следовательно,

$$E(t)=\frac{\binom{N-2}{n-2}\Sigma}{\frac{1}{2}n(n-1)\binom{N}{n}}=\frac{\Sigma}{\frac{1}{2}N(N-1)}=\tau. \tag{5.45}$$

**5.18.** Выведем теперь выражение для дисперсии $t$. Пусть $c^{(n)}$ означает значение $c$ для выборочной последовательности, состоящей из $n$ единиц, а $c$ — генеральное значение. Тогда

$$t=\frac{c^{(n)}}{n(n-1)} \tag{5.46}$$

и

$$c^{(n)}=\Sigma^{(n)}c_{ij}, \tag{5.47}$$

где $\Sigma^{(n)}$ означает суммирование выборочных значений $c$ по $i$ и $j$.

Поскольку нам требуется найти $E(t^2)$, рассмотрим

$$\sum_n (c^{(n)})^2=\sum_n \Sigma^{(n)} c_{ij}c_{kl}, \tag{5.48}$$

где $\Sigma$ означает суммирование по всем вариантам выборок объемом $n$ из совокупности, содержащей $N$ членов. Теперь перечислим случаи возможных появлений в данной сумме величин $c_{ij}c_{kl}$ и подобных произведений со связанными индексами.

1. Если $i$, $j$, $k$ и $l$ все являются различными, то член $c_{ij}c_{kl}$ может появиться в $\binom{N-4}{n-4}$ выборках из остающихся членов; вклад таких членов в $\sum\limits_n$ составит:

$$\binom{N-4}{n-4}\Sigma' c_{ij}c_{kl},$$

где $\Sigma'$, как и раньше, означает суммирование от 1 до $N$ по несовпадающим значениям $i$, $j$, $k$ и $l$.

2. Аналогично член $c_{ij}c_{il}$ может появиться $\binom{N-3}{n-3}$ путями, и существует четыре варианта связывания одного индекса. Таким образом, вклад этих членов составит:

$$4\binom{N-3}{n-3}\Sigma' c_{ij}c_{il}.$$

3. Члены, подобные $c_{ij}^2$, дают аналогичный вклад, т. е.

$$2\binom{N-2}{n-2}\Sigma' c_{ij}^2.$$

$$\sum_n (c^{(n)})^2=\binom{N-4}{n-4}\Sigma' c_{ij}c_{kl}+4\binom{N-3}{n-3}\Sigma' c_{ij}c_{ik}+$$
$$+2\binom{N-2}{n-2}\Sigma' c_{ij}^2.$$

Выражая $\Sigma'$ в виде членов $\Sigma$ и поделив на $\binom{N}{n}$, получим

$$E(c^{(n)})^2=\frac{n^{(4)}}{N^{(4)}}(\Sigma c_{ij}c_{kl}-4\Sigma c_{ij}c_{il}+2\Sigma c_{ij}^2)+$$
$$+\frac{4n^{(3)}}{N^{(3)}}(\Sigma c_{ij}c_{il}-\Sigma c_{ij}^2)+\frac{2n^{(2)}}{N^{(2)}}\Sigma c_{ij}^2, \tag{5.49}$$

где $n^{(r)}=n(n-1)\ \dots\ (n-r+1)$. Поскольку $\Sigma c_{ij}^2=N(N-1)$ и $\Sigma c_{ij}c_{kl}=c^2$, дисперсия $t$ для заданных $\tau$ и $n$, как видно, зависит от $\Sigma c_{ij}c_{il}=\Sigma c_i^2$, где $c_i=\sum_{j=1}^{N}c_{ij}$.

Пусть $N$ — очень большое число. Поскольку величины $c$ и $\Sigma c_i^2$ имеют порядок $N^2$ и $N^3$ соответственно, можно написать

$$\tau_i=\frac{c_i}{N}. \tag{5.50}$$

Находим теперь, что

$$E(t^2)\sim\frac{(n-2)(n-3)}{n(n-1)}\tau^2+\frac{4(n-2)\Sigma\tau_i^2}{n(n-1)N}+\frac{2}{n(n-1)}. \tag{5.51}$$

Таким образом, в пределе

$$\operatorname{var} t=\frac{4(n-2)}{n(n-1)}\operatorname{var}\tau_i+\frac{2}{n(n-1)}(1-\tau^2). \tag{5.52}$$

**5.19.** Рассмотрим теперь величину $c_{ij}=a_{ij}b_{ij}$. Пусть $b_{ij}$ равно $\pm1$, а коэффициенты $a$ могут принимать любые значения при соблюдении следующих условий:

$$\Sigma a_{ij}^2=N(N-1),\quad \Sigma a_{ij}b_{ij}=c=N(N-1)\tau.$$

Стационарные значения $\Sigma c_i^2$ имеют место тогда, когда величины $a$ удовлетворяют условию:

$$b_{ij}(c_i+c_j)-\lambda a_{ij}-\mu b_{ij}=0, \tag{5.53}$$

где $\lambda$ и $\mu$ неопределенные коэффициенты. Умножая на $b_{ij}$ и суммируя по всем $j$, находим:

$$c_i=\frac{\mu(N-1)-c}{N-2-\lambda}.$$

Следовательно, если все $c_i$ не будут равными (если они равные, то $\Sigma c_i^2$ является *минимальной*), то $\lambda$ и $\mu$ должны иметь следующие значения:

$$\lambda=N-2, \quad \mu=c/(N-1).$$

Умножая (5.53) на $a_{ij}$ и суммируя по $j$ и $i$, получим:

$$2\,\Sigma c_i^2-\lambda N(N-1)-\mu c=0.$$

Откуда следует, что $\Sigma c_i^2$ не может превысить

$$\frac{1}{2}N(N-1)(N-2)+\frac{1}{2}c^2/(N-1).$$

Для больших $N$ это означает, что

$$\Sigma\tau_i^2/N\leqslant\frac{1}{2}(1+\tau^2).$$

Следовательно,

$$\operatorname{var}\tau_i\leqslant\frac{1}{2}(1-\tau^2)$$

и, таким образом, из (5.52) вытекает, что

$$\operatorname{var}t\leqslant\left[\frac{2(n-2)}{n(n-1)}+\frac{2}{n(n-1)}\right](1-\tau^2)\leqslant\frac{2}{n}(1-\tau^2), \tag{5.54}$$

т. е. получим уравнение (4.9) предыдущей главы.

5.20. Полученный результат предполагает использование преобразования

$$w=\sin^{-1}t. \tag{5.55}$$

С той же степенью приближения мы можем принять, что $w$ нормально распределено относительно $w=\sin^{-1}\tau$ и что дисперсия $w$ удовлетворяет отношению

$$\operatorname{var}w\leqslant\frac{2}{n}, \tag{5.56}$$

преимущество которого заключается в независимости от $w$. Неизвестно, ближе ли в данном случае это распределение к нормальному, чем распределение $t$.

5.21. После доказательств, помещенных выше в данной главе, приведем в основных чертах доказательство того, что распределение $t$ стремится к нормальному для больших $n$. Рассмотрим лишь общие контуры такого доказательства.

Запишем

$$g_{ij}=c_{ij}-c/N^2,$$

так что

$$g_{ij}=g_{ji}, \quad \Sigma g_{ij}=0 \text{ и } g_{ii}=-c/N^2=-(N-1)\tau/N.$$

Момент величины $c^{(n)}$ $r$-го порядка относительно средней есть $E\,(\Sigma^{(n)} g_{ij})^r$, поэтому рассмотрим

$$\sum_n (\Sigma^{(n)} g_{ij})^r=\sum_n \Sigma^{(n)} g_{ij}\, g_{hl}\, g_{uv}. \tag{5.57}$$

Основное условие заключается в том, что

$$N^3\, \Sigma g_{ij}\, g_{ik} \text{ имеет ненулевую границу,} \tag{5.58}$$

это справедливо только в том случае, если $1-\tau^2$ находится около 1, следовательно, тенденция к нормальности может исчезнуть при высокой степени корреляции. Мы также предполагаем, что $n/N$ стремится к нулю.

Как и в (5.13), при выводе момента степени $2m$ основной член определяется выражением, подобным $(\Sigma g_{ij}\, g_{ik})^m$, другие члены имеют порядок, меньший $m$. При приписывании $3m$ подписных индексов имеется $\binom{N-3m}{n-3m}$ вариантов выбора остающихся $n-3m$ членов, а подписные индексы могут быть связаны $\frac{(2m)!\,2^{2m}}{m!\,(2!)^m}$ путями для того, чтобы дать тот же самый результат. Разделив на $\binom{N}{n}$ и имея в виду, что для больших $N$ и $n$

$$\binom{N-3m}{n-3m}\Big/\binom{N}{n}\sim\frac{n^{3m}}{N^{3m}},$$

находим выражение для главного члена момента степени $2m$

$$\frac{n^{3m}}{N^{3m}}\,\frac{(2m)!}{m!}\,2^m\,(\Sigma g_{ij}\, g_{ik})^m,$$

величина которого имеет порядок, равный $n^{3m}$. Применяя такое же доказательство, находим, что члены с $f<3m$ различными индексами имеют порядок $n^f$ и ими можно пренебречь. Таким образом,

$$\mu_{2m}\sim\frac{n^{3m}}{N^{3m}}\,\frac{(2m)!}{m!}\,2^m\,(\Sigma g_{ij}\, g_{ik})^m\sim\frac{(2m)!}{2^m\, m!}\,(\mu_2)^m. \tag{5.59}$$

Для сравнения отметим, что величина $\mu_{2m+1}$ имеет порядок, равный $n^{-\frac{1}{2}}$, стремление к нормальности сохраняется и здесь. Дисперсия $c^{(n)}$ равна:

$$\frac{4n^3}{N^3}(\Sigma g_{ij}\,g_{ik}) = 4n^3 \operatorname{var}\tau_i, \qquad (5.60)$$

дисперсия $t$ соответственно составит $4/n \operatorname{var}\tau_i$, что согласуется с (5.52).

**5.22.** Теперь перейдем к причинам введенного выше предположения о том, что пределы для дисперсии $t$, заданные (5.54), не могут быть очень узкими.

Рассмотрим последовательность:

5 2 3 1 6 7 8 9 4

Число положительных пар равно 26, так что $t = 0{,}44$. Преобразуем его так, чтобы единица была в начале последовательности, при этом для сохранения оценки 26 передвинем число 9. Для того чтобы достичь начала последовательности, число 1 перескочит через три члена и, следовательно, это добавит к оценке величину 3. Тогда число 9 должно быть сдвинуто влево на 3 члена так, чтобы уменьшить оценку на 3. При этом получим следующую последовательность:

1 5 2 3 9 6 7 8 4

Проделав то же самое с 2, получим:

1 2 5 9 3 6 7 8 4

Для того чтобы 9 не стала рядом с 1 и в связи с невозможностью дальнейшего сдвига влево, мы далее должны сдвигать число 8 и т. д. Продолжая процесс, мы в конце концов получим:

1 2 3 4 9 8 7 6 5

Числа до 4 представляют натуральный ряд, а остальные имеют обратный порядок расположения. Можно назвать такую последовательность «каноническим порядком» для данного $S$. Не всегда возможно привести заданную последовательность к канонической, однако не может быть случая, когда более одного числа окажется не на своем месте в канонической форме.

Если исходная последовательность записана в обратном порядке, то $\tau$ становится $-\tau$. Мы можем привести ее к канонической форме и повторно инверсировать результат, тогда коэффициент снова станет равным $\tau$. Такую последовательность можно назвать «обратная каноническая форма»[1].

**5.23.** Теперь обратимся к каноническому случаю, когда имеется всего $N$ членов, причем в начале находится $R$ членов, расположенных

[1] В [97] опубликовано доказательство того, что каноническая последовательность имеет минимальную дисперсию, однако проф. Хоффдинг привел возражения, которые мне представляются ценными.

в виде натурального ряда; $N-R$ членов имеют обратный порядок расположения. Если мы отберем $n-j$ членов из $R$ и $j$ членов из $N-R$, то величина $S$ для выборки составит $\frac{1}{2}n(n-1)+\frac{1}{2}j(j-1)$, а относительная частота величины $Q=\frac{1}{2}n(n-1)-S$ равна:

$$\binom{R}{n-j}\binom{N-R}{j}\bigg/\binom{N}{n}.$$

Теперь предположим, что $N$ бесконечно большая величина, а $R/N$ стремится к пределу $p$. Тогда относительная частота величины $Q$, которая равна $\frac{1}{2}j(j-1)$, будет стремиться к $\binom{n}{j}p^{n-j}q^{j}$, где $q=1-p$. Среднее значение $Q$ равно:

$$\sum_{0}^{n}\frac{1}{2}j(j-1)\binom{n}{j}p^{n-j}q^{j}=\frac{1}{2}n(n-1)q^{2}$$

и поскольку

$$t=1-\frac{2Q}{\frac{1}{2}n(n-1)},$$

то мы должны получить

$$q=\left[\frac{1}{2}(1-\tau)\right]^{\frac{1}{2}}.$$

Дисперсия $Q$ находится как

$$\operatorname{var}Q=n(n-1)pq^{2}\left[nq+\frac{1}{2}(1-3q)\right]$$

и, следовательно,

$$\operatorname{var}t=\frac{16pq^{2}\left[nq+\frac{1}{2}(1-3q)\right]}{n(n-1)}. \tag{5.61}$$

Если обратная генеральная последовательность приведена к канонической форме и при этом имеются отношения $p'$ и $q'$, то получим:

$$q'=\left[\frac{1}{2}(1+\tau)\right]^{\frac{1}{2}}$$

и

$$\operatorname{var}t'=\frac{16p'q'^{2}\left[nq'+\frac{1}{2}(1-3q')\right]}{n(n-1)}. \tag{5.62}$$

Тогда, поскольку $q^{2}+q'^{2}=1$,

$$\operatorname{var}t'-\operatorname{var}t=\frac{16(n-2)}{n(n-1)}(q'-q)(1-q)(1-q').$$

Если $\tau$ — положительная величина, то $q' > q$, и тогда $\operatorname{var} t' > \operatorname{var} t$. Взяв обратную каноническую последовательность (когда $\tau > 0$) и прямую каноническую последовательность ($\tau < 0$), находим для дисперсии $t$ при больших $n$

$$\operatorname{var} t \sim \frac{4\sqrt{2}}{n}(1+|\tau|)^{\frac{3}{2}}\left\{1-\sqrt{\left[\frac{1}{2}(1+|\tau|)\right]}\right\}. \qquad (5.63)$$

Отношение этой величины к верхнему пределу, равному $\frac{2}{n}(1-\tau^2)$, изменяется от $2(\sqrt{2}-1) = 0{,}83$ (когда $\tau = 0$) до 1 (когда $\tau = 1$) Очевидно, что верхний предел дисперсии не может быть существенно снижен, поскольку дисперсия действительной генеральной последовательности приближается к этому пределу для всех значений $\tau$, когда $n$ не очень мало.

## Более точный анализ для ненулевого случая

**5.24.** В доказательстве того, что распределение $t$ стремится к нормальному в ненулевом случае, мы пренебрегли членом порядка $1/\sqrt{n}$, а это означает, что нормальное приближение, справедливое для больших $n$, может оказаться неправомерным для малых или средних $n$. Кратко проверим возможность улучшения приближения для средних $n$.

Рассматривая снова (5.52)

$$\operatorname{var} t = \frac{4(n-2)}{n(n-1)}\operatorname{var}\tau_i + \frac{2}{n(n-1)}(1-\tau^2), \qquad (5.64)$$

мы видим, что действительная дисперсия $t$ для больших $N$ зависит от неизвестных функций $\tau_i$ и $\tau$. При отсутствии точных знаний об этих величинах мы можем оценить их на основе выборки, взяв выборочные значения $c_i$ и $c$ вместо неизвестных генеральных значений. Однако, выполняя это, лучше слегка модифицировать нашу формулу так, чтобы устранить смещение. Читатель, по-видимому, помнит, что в обычной теории статистики, как правило, предпочитают применять оценку дисперсии, т. е. $\Sigma(x-\bar{x})^2/(n-1)$, а не действительную величину выборочной дисперсии, равную $\Sigma(x-\bar{x})^2/n$, поскольку средняя первой величины для всех выборок равна генеральной дисперсии. Подобно этому лучше не подставлять в (5.64) выборочные величины $c_i$ и $c$, а применить формулу, которая дает точное значение $\operatorname{var} t$ для всех выборок. Такой формулой, дающей несмещенный результат, является

$$\operatorname{var} t = \frac{1}{n(n-1)(n-2)(n-3)}\left[4\Sigma c_i^2 - \frac{2(2n-3)}{n(n-1)}c^2 - 2n(n-1)\right], \qquad (5.65)$$

где $c_i$ и $c$ — выборочные значения. Приведенная формула дает нам наилучшую оценку $\operatorname{var} t$.

5.25. Теперь мы можем немного продвинуться далее, рассматривая третий момент $t$, с тем чтобы можно было отойти от нормального выборочного распределения. Подробное доказательство можно найти в [13]. Здесь мы приводим лишь результат. Если

$$\gamma_1 = \frac{\mu_3(t)}{[\mu_2(t)]^{\frac{3}{2}}}, \tag{5.66}$$

то частота распределения величины

$$x = \frac{t - \tau}{\sqrt{\mu_2(t)}}$$

есть

$$f(x) = \left(1 - \frac{\gamma_1}{6}\frac{d^3}{dx^3}\right)\frac{e^{-\frac{1}{2}x^2}}{\sqrt{(2\pi)}}\,[1 + 0(n^{-1})]. \tag{5.67}$$[1]

Обозначим через $\xi$ нормальное отклонение, вероятность превышения которого равна $P(\xi)$. Вероятность того, что $x$ превысит $\xi$, составит:

$$F(\xi) = P(\xi) + \frac{\gamma_1}{6}(\xi^2 - 1)\,\frac{e^{-\frac{1}{2}\xi^2}}{\sqrt{(2\pi)}}.$$

Если $X$ представляет собой точный предел, такой, что $F(X) = P(\xi)$, то с помощью последовательного приближения легко доказывается, что

$$X = \xi + \frac{\gamma_1}{6}(\xi^2 - 1) \tag{5.68}$$

стремится к $n^{-1}$. Например, при 5% величина $\xi$ равна $\pm 1{,}96$. Соответствующее значение $X$ составит:

$$\pm 1{,}96 + \frac{(1{,}96)^2 - 1}{6}\gamma_1 = \pm 1{,}96 + 0{,}474\gamma_1. \tag{5.69}$$

$X$, соответствующий 1%, равен:

$$\pm 2{,}58 + 0{,}941\,\gamma_1. \tag{5.70}$$

В следующем примере показано, как можно использовать эти результаты.

*Пример 5.1*

Следующие данные характеризуют ранги, приписанные двум признакам $A$ и $B$, в выборке из 30 единиц, извлеченных из совокупности неизвестного свойства. Нам необходимо оценить корреляцию в генеральной совокупности.

---

[1] Этот результат более строг, чем можно было бы получить путем обычного разложения функции частот в ряд Грема—Шарлье, основанного только на первых трех моментах. Подобного рода разложения рассмотрены в [58, т. 1,6.17].

Таблица 5.1

| $A$ | $B$ | $A$ | $B$ | $A$ | $B$ |
|---|---|---|---|---|---|
| 1 | 5 | 11 | 17 | 21 | 21 |
| 2 | 4 | 12 | 13 | 22 | 29 |
| 3 | 9 | 13 | 24 | 23 | 28 |
| 4 | 3 | 14 | 14 | 24 | 19 |
| 5 | 6 | 15 | 1 | 25 | 23 |
| 6 | 2 | 16 | 12 | 26 | 20 |
| 7 | 15 | 17 | 10 | 27 | 7 |
| 8 | 18 | 18 | 30 | 28 | 26 |
| 9 | 8 | 19 | 22 | 29 | 27 |
| 10 | 11 | 20 | 16 | 30 | 25 |

Найдено, что коэффициент корреляции $t$ равен $+0{,}490$.

а. Прежде всего рассмотрим максимальные доверительные интервалы, определяемые в (4.12). Пятипроцентные пределы составят:

$$-0{,}02 \leqslant t \leqslant 0{,}80.$$

б. $\sin^{-1}$ — преобразование (согласно 5.20) дает:

$$w = \sin^{-1} t = 0{,}512,$$
$$0{,}01 \leqslant t \leqslant 0{,}85,$$

т. е. получили результат, который не очень сильно отличается от результата, приведенного в «а».

в. Для того чтобы продвинуться дальше, нам необходимо получить величины $c_i$ и $c$. В табл. 5.2 приведена матрица значений $c_i$ и $c$, полученная для наших данных. Находим:

$$c = 426,$$
$$\Sigma c_i^2 = 7470.$$

Затем на основе (5.65) получим:

$$\operatorname{var} t = \frac{1}{30 \times 29 \times 28 \times 27}\left[4 \times 7470 - \frac{2 \times 57}{30 \times 29} + 426^2 - 60 \times 29\right] =$$
$$= 0{,}006630,$$

откуда оценка стандартной ошибки равна 0,0814. Пятипроцентные доверительные пределы при предположении о нормальном распределении тогда равны:

$$0{,}33 \leqslant t \leqslant 0{,}65.$$

Полученный интервал значительно уже, чем в «а» и «б».

г. Мы можем не связывать себя требованием о нормальном распределении, если оценим $\gamma_1$, которое зависит от $\mu_3(t)$. Следующая формула дает приближение для выборок средних объемов:

$$\mu_3(t) = \frac{8}{n^6}\left[\sum_{i>j} c_{ij}(c_i + c_j)^2 - \frac{5c\Sigma c_i}{n} + \frac{3c^2}{n^3}\right], \tag{5.71}$$

Таблица 5.2

| $c_{ij}$ | $c_i$ |
|---|---|
| 0 − + − + − + + + + + + + + − + + + + + + + + + + + + + + + | 21 |
| − 0 + − + − + + + + + + + + − + + + + + + + + + + + + + + + | 21 |
| + + 0 − − − + + − + + + + + − + + + + + + + + + + + − + + + | 17 |
| − − − 0 + − + + + + + + + + − + + + + + + + + + + + + + + + | 19 |
| + + − + 0 − + + + + + + + + − + + + + + + + + + + + + + + + | 23 |
| − − − − − 0 + + + + + + + + − + + + + + + + + + + + + + + + | 17 |
| + + + + + + 0 + − − + − + − − − − + + + + + + + + + − + + + | 13 |
| + + + + + + + 0 − − − − + − − − − + + − + + + + + + − + + + | 9 |
| + + − + + + − − 0 + + + + + − + + + + + + + + + + + − + + + | 19 |
| + + + + + + − − + 0 + + + + − + − + + + + + + + + + − + + + | 19 |
| + + + + + + + − + + 0 − + − − − − + + − + + + + + + − + + + | 13 |
| + + + + + + − − + + − 0 + + − − − + + + + + + + + + − + + + | 15 |
| + + + + + + + + + + + + 0 − − − − + − − − + + − − − − + + + | 7 |
| + + + + + + − − + + − + − 0 − − − + + + + + + + + + − + + + | 13 |
| − − − − − − − − − − − − − − 0 + + + + + + + + + + + + + + + | 1 |
| + + + + + + − − + + − − − − + 0 − + + + + + + + + + − + + + | 13 |
| + + + + + + − − + − − − − − + − 0 + + + + + + + + + − + + + | 11 |
| + + + + + + + + + + + + + + + + + 0 − − − − − − − − − − − − | 5 |
| + + + + + + + + + + + + − + + + + − 0 − − + + − + − − + + + | 15 |
| + + + + + + + − + + − + − + + + + − − 0 + + + + + + − + + + | 17 |
| + + + + + + + + + + + + − + + + + − − + 0 + + − + + − + + + | 19 |
| + + + + + + + + + + + + + + + + + − + + + 0 − − − − − − − − | 11 |
| + + + + + + + + + + + + + + + + + − + + + − 0 − − − − − − − | 11 |
| + + + + + + + + + + + + − + + + + − − + − − − 0 + + − + + + | 15 |
| + + + + + + + + + + + + − + + + + − + + + − − + 0 − − + + + | 17 |
| + + + + + + + + + + + + − + + + + − − + + − − + − 0 − + + + | 15 |
| + + − + + + − − − − − − − − + − − − − − − − − − − − 0 + + + | −11 |
| + + + + + + + + + + + + + + + + + − + + + − − + + + + 0 + − | 21 |
| + + + + + + + + + + + + + + + + + − + + + − − + + + + + 0 − | 21 |
| + + + + + + + + + + + + + + + + + − + + + − − + + + + − − 0 | 19 |
| | $c = 426$ <br> $n = 30$ <br> $\sum c^2 = 7470$ |

где первый член в квадратных скобках предполагает суммирование величин $c_{ij}(c_i + c_j)^2$ по всем значениям $i > j$, т. е. суммирование величин, лежащих ниже диагонали в табл. 5.2. После ряда утомительных вычислений находим:

$$\gamma_1 = -0{,}32.$$

Теперь скорректированные 5%-ные пределы из (5.69) будут равны:

$$0{,}32 \leqslant t \leqslant 0{,}64.$$

Поправка на отсутствие нормального распределения мала и пределы весьма схожи с пределами, приведенными в «в».

**5.26.** Расчеты, которые нужно было выполнить в предыдущих примерах, требуют большего терпения, чем то, которым мы обычно обладаем. В [98] исследованы третий и четвертый моменты $\tau$ в ненулевом случае и показано, что последний зависит от 10 параметров, которые могут быть оценены на основе имеющихся данных. Этот результат представляет значительный теоретический интерес, однако затраты труда опять выступают как препятствие, если только расчет не будет осуществляться на быстродействующем компьютере.

У нас осталось еще одна возможность. Если можно предположить, что последовательность основана на переменной, имеющей нормальное распределение (или, возможно, в качестве приближения, эта переменная имеет распределение, близкое к нормальному), то дисперсия уменьшается весьма значительно. Об этом смотри в 9.6 и следующих параграфах.

### $\rho$ в ненулевом случае

5.27. Как обычно, теория выборки $\rho$ более сложна, чем соответствующая теория, относящаяся к $\tau$. Для ненулевого случая мало что известно о распределении этого коэффициента, помимо того, что мы можем определить выборочное значение $\rho$, равное математическому ожиданию $r_s$.

Рассмотрим функцию $V$ из (2.27):

$$V=\sum_{i>j} m_{ij}(j-i). \tag{5.72}$$

Значение $m_{ij}$ для любой конкретной пары членов является одинаковым как для выборки, так и для генеральной совокупности — это и служит основной причиной того, что $E(t)=\tau$. Однако сказанное не относится к сомножителю $(j-i)$.

Рассмотрим пару членов, имеющих ранги $I$ и $J$ в совокупности, состоящей из $N$ единиц. Вероятность попасть в выборку для любого члена этой совокупности $\frac{n-2}{N-2}$. Поэтому среднее значение ранга в выборке равно:

$$\frac{n-2}{N-2}(J-I-1).$$

Отсюда выборочное среднее значение $j-i$ на единицу больше, чем эта величина, а именно:

$$\frac{N-n}{N-2}+\frac{n-2}{N-2}(J-I). \tag{5.73}$$

Таким образом, среднее значение $V$ определяется формулой

$$E(V)=\frac{\binom{N-2}{n-2}}{\binom{N}{n}}\sum_{I<J} m_{IJ}\left[\frac{N-n}{N-2}+\frac{n-2}{N-2}(J-I)\right],$$

где суммирование производится для всей совокупности по всем парам рангов; последние же могут быть образованы $\binom{N-2}{n-2}$ способами. Отсюда, выражая $V$ через $r_s$, находим:

$$E(1-r_s)=\frac{12}{N(N-1)(n+1)}\Sigma m_{IJ}\left[\frac{N-n}{N-2}+\frac{n-2}{N-2}(I-J)\right]. \tag{5.74}$$

Теперь определим для нашей совокупности

$$\Sigma m_{IJ} = \frac{N(N-1)}{4}(1-\tau),$$

$$\Sigma m_{IJ}(J-I) = \frac{N(N^2-1)}{12}(1-\rho)$$

и, подставляя в (5.74), находим:

$$E(r_s) = \frac{1}{(n+1)(N-2)}[3(N-n)\tau + (n-2)(N+1)\rho] \qquad (5.75)$$

общую формулу (см. [20] и [12]).

5.28. Заметим, что для больших $N$ это выражение стремится к результату, описанному в [39], а именно:

$$E(r_s) = \frac{1}{n+1}[3\tau + (n-2)\rho]. \qquad (5.76)$$

Таким образом,

$$E(r_s) - \rho = \frac{3}{n+1}(\tau - \rho). \qquad (5{,}77)$$

Пользуясь статистической терминологией, назовем $r_s$ смещенной оценкой $\rho$. Для некоторых совокупностей это смещение может быть заметным. Вероятно, уместно скорректировать $r_s$, вычитая «выборочное смещение», равное $3(t - r_s)/(n-2)$.

## Библиография

См. библиографию к гл. 4. О распределении $\tau$ в нулевом случае см. [68], [83] и [16]. О среднем значении выборочного коэффициента Спирмэна см. [39], [11], [12] и [20]. Этот результат для выборки из нормальной совокупности получен в [67].

Нормальное распределение $t$ в случае, когда имеют место связи, является следствием общего результата, описанного в [39]. Простое доказательство этого не так-то легко получить. Однако для случая, когда связи имеются только в одной последовательности, см. [102] и [59].

## *ГЛАВА 6.* ПРОБЛЕМА $m$ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

**6.1.** До сих пор мы рассматривали корреляцию двух последовательностей рангов. Теперь рассмотрим случай, когда имеется несколько последовательностей; их число обозначим буквой $m$, а количество рангов в каждой последовательности — буквой $n$. Рассмотрим общие соотношения между указанными последовательностями. Предположим, что четыре эксперта упорядочили шесть объектов следующим образом:

| | Объект | | | | | |
|---|---|---|---|---|---|---|
| | $A$ | $B$ | $C$ | $D$ | $E$ | $F$ |
| Эксперт $P$ | 5 | 4 | 1 | 6 | 3 | 2 |
| » $Q$ | 2 | 3 | 1 | 5 | 6 | 4 |
| » $R$ | 4 | 1 | 6 | 3 | 2 | 5 |
| » $S$ | 4 | 3 | 2 | 5 | 1 | 6 |
| Общая сумма рангов | 15 | 11 | 10 | 19 | 12 | 17 |

(6.1)

Воспользовавшись рассмотренными ранее методами, мы можем рассчитать коэффициенты корреляции между рангами каждых двух экспертов; следуя по этому пути, можно получить $\binom{4}{2} = 6$ коэффициентов корреляции. Однако обычно нам требуются не такие показатели. Нас интересует общая мера согласованности (конкордации) внутри группы экспертов.

**6.2.** Наиболее наглядно такую меру, вероятно, можно было бы получить, усреднив все возможные значения коэффициентов $\tau$ и $\rho$, исчисленные для каждой пары экспертов; но когда число $m$ велико, такая процедура, понятно, потребует чрезвычайно больших затрат труда. В случаях, подобных описанному выше, проще всего сопоставить суммы рангов, приписываемых каждому объекту различными экспертами; такие данные приведены в последней строке (6.1). Сумма этих чисел составляет 84; в общем случае она равна $\frac{1}{2}mn(n+1)$. Эта величина получена путем суммирования $m$ слагаемых, каждое из которых представляет собой сумму натуральных чисел от 1 до $n$.

Тогда среднее значение суммы рангов одного объекта составляет $\frac{1}{2} m(n+1)$; в нашем примере оно равно 14. Рассмотрим индивидуальные отклонения от этого среднего значения:

$$1, -3, -4, 5, -2, 3. \tag{6.2}$$

Если бы все последовательности совпадали, суммы рангов в (6.1) выглядели бы следующим образом:

$$m, 2m, \ldots, nm$$

(эти суммы, разумеется, необязательно располагались бы именно в таком порядке), а их отклонения равнялись бы соответственно

$$-\frac{1}{2} m(n-1), \qquad -\frac{1}{2} m(n-3), \ldots, \quad \frac{1}{2} m(n-1).$$

Сумма квадратов этих отклонений составляет:

$$\frac{1}{12} m^2 (n^3-n). \tag{6.3}$$

Это максимальное значение, которое может принимать сумма квадратов рассматриваемых отклонений. Другое экстремальное значение может быть равно нулю, когда $m$ четно либо $n$ нечетно (либо и то и другое одновременно), в остальных случаях значение суммы квадратов хотя и не равно нулю, но сравнительно мало.

Обозначим буквой $S$ сумму квадратов фактически встречающихся отклонений, в нашем примере (см. (6.2)) она составляет 64. В таком случае величину

$$W = \frac{12S}{m^2 (n^3-n)} \tag{6.4}$$

будем называть *коэффициентом конкордации* (коэффициентом согласованности). В нашем примере

$$W = \frac{12 \times 64}{16 \times 210} = 0{,}229.$$

**6.3.** В некотором смысле $W$ служит мерой общности суждений $m$ экспертов. Если все эти суждения совпадают, то $W = 1$. Если различия между ними очень велики, суммы рангов окажутся более или менее близки друг к другу по своей величине и поэтому сумма квадратов $S$ будет меньше ее возможного максимального значения, а следовательно, и коэффициент $W$ составит малую величину. Увеличение $W$ от 0 до 1 означает, что различия между отклонениями «усиливаются», и в этом проявляется все бо́льшая согласованность рангов.

**6.4.** Читатель может спросить: почему выбранный нами коэффициент меняет свои значения в границах от 0 до 1, а не от —1 до 1, как, например, коэффициент ранговой корреляции? Ответ заключается в следующем: когда имеется более двух экспертов, противоположные понятия согласованности и несогласованности утрачивают прежнюю симметричность. Оценочные суждения $m$ экспертов могут полностью совпадать, однако эти суждения не могут полностью не совпадать между собой в том смысле, который мы вкладывали ранее в это понятие. Если рассматриваются оценки трех экспертов $P$, $Q$ и $R$, причем мнения $P$ и $Q$ расходятся между собой, а оценки $P$ и $R$ также не совпадают, то суждения $Q$ и $R$ должны совпадать.

**6.5.** Обозначим символом $\rho_{av}$ среднее значение коэффициента Спирмэна, исчисленного для $\binom{m}{2}$ возможных пар экспертов; в таком случае

$$\rho_{av} = \frac{mW-1}{m-1}. \tag{6.5}$$

Действительно, воспользуемся символом $x_{ij}$ для обозначения ранга $j$-го объекта, проставленного $i$-м экспертом (причем эти оценки отсчитываются относительно среднего значения $\frac{1}{2}(n+1)$).

В таком случае среднее значение $\rho$ составляет:

$$\rho_{av}=\frac{1}{m(m-1)}\,\frac{\sum\limits_{i,\,k=1}^{m}\sum\limits_{j=1}^{n} x_{ij}x_{kj}}{\frac{1}{12}(n^3-n)}=$$

$$=\frac{12}{m(m-1)(n^3-n)}\left[\sum_{j=1}^{n}\left(\sum_{i=1}^{m} x_{ij}\right)^2-\sum_{j=1}^{n}\sum_{i=1}^{m} x_{ij}^2\right]=$$

$$=\frac{12}{m(m-1)(n^3-n)}\left[S-\frac{1}{12}m(n^3-n)\right]=\frac{mW-1}{m-1}. \tag{6.6}$$

Если $\rho_{av} = +1$, то $W = 1$. Если $W = 0$, то $\rho_{av} = -1/(m-1)$, т. е. наименьшая величина, которую принимает среднее значение $\rho$. Сказанное может служить еще одним аргументом, относящимся к вопросу, рассмотренному в конце 6.4.

В рассматриваемом случае применить коэффициент $\rho$ удобнее, чем $\tau$. Дело в том, что среднее значение $\tau$ не так-то легко выразить через сумму соответствующих рангов[1].

**6.6.** Если некоторые последовательности рангов содержат связи, величину $T'$ можно выразить следующим образом, в соответствии с (3.6):

$$T'=\frac{1}{12}\sum_{t}(t^3-t). \tag{6.7}$$

В этом случае мы можем определить коэффициент $W$ с помощью следующей формулы:

$$W=\frac{S}{\frac{1}{12}m^2(n^3-n)-m\sum\limits_{T'} T'}, \tag{6.8}$$

[1] Об использовании коэффициента $\tau$ в задачах множественной корреляции рангов см. [23], а также работы, указанные в гл. 13.

где знак $\sum\limits_{T'}$ означает суммирование по различным последовательностям рангов.

В этом случае нужно несколько модифицировать и формулу (6.6).

**6.7.** Следует несколько уточнить сказанное, поскольку знаменатель в выражении (6.8) может отличаться от максимального значения суммы квадратов всех рангов (измеряемых относительно их среднего значения). Действительно, в том случае, когда последовательности рангов не содержат связи, мы можем определить значение $W$ по новой формуле

$$W = \frac{S}{mS'}, \tag{6.9}$$

где $S$ представляет собой сумму квадратов отклонений всех рангов от их среднего значения. Это явно согласуется с нашим прежним определением, потому что сумма квадратов отклонений, подсчитанных по любой последовательности рангов, равна $\frac{1}{12}(n^3 - n)$, а всего насчитывается $m$ таких последовательностей. Выражение (6.8) также соответствует определению (6.9), поскольку, как было показано в 3.10, влияние связей, существующих в последовательностях рангов и характеризуемых величиной $T'$, проявляется в том, что общая сумма квадратов отклонений от средних значений уменьшается на величину $T'$.

Запись выражения (6.9) особенно удобна: дело в том, что в дисперсионном анализе используется аналогичная формула. Предположим, что нам дана таблица рангов (измеряемых относительно их средних значений)

$$\begin{matrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{matrix} \qquad (6.10)$$

Сумма элементов любой строки равна нулю, и, следовательно, средние значения этих сумм также равны нулю. Суммы элементов столбцов обозначим через $S_1, \ldots, S_n$, а средние значения этих сумм будут соответственно равны $S_1/m, \ldots, S_n/m$.

Дисперсия, исчисленная по всему массиву рассматриваемых величин, по определению равна $S'/mn$. Дисперсия средних значений, исчисляемых по столбцам, равна $\frac{1}{n}\Sigma\,(S_i/m)^2 = S/m^2n$. Нетрудно видеть, что отношение второй величины к первой равно $S/mS' = W$, так что $W$ представляет собой отношение дисперсии средних значений, исчисляемых по столбцам, к дисперсии, исчисляемой по всему массиву рассматриваемых величин. Это отношение, а тем более отношение $S/(mS' - S)$ хорошо известны читателю, знакомому с дисперсионным анализом.

*Пример 6.1*

Рассмотрим три последовательности рангов:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $P$ | 1 | $4\frac{1}{2}$ | 2 | $4\frac{1}{2}$ | 3 | $7\frac{1}{2}$ | 6 | 9 | $7\frac{1}{2}$ | 10 |
| $Q$ | $2\frac{1}{2}$ | 1 | $2\frac{1}{2}$ | $4\frac{1}{2}$ | $4\frac{1}{2}$ | 8 | 9 | $6\frac{1}{2}$ | 10 | $6\frac{1}{2}$ |
| $R$ | 2 | 1 | $4\frac{1}{2}$ | $4\frac{1}{2}$ | $4\frac{1}{2}$ | $4\frac{1}{2}$ | 8 | 8 | 8 | 10 |
| Сумма | $5\frac{1}{2}$ | $6\frac{1}{2}$ | 9 | $13\frac{1}{2}$ | 12 | 20 | 23 | $23\frac{1}{2}$ | $25\frac{1}{2}$ | $26\frac{1}{2}$ |
| Отклонения от средней $\left(16\frac{1}{2}\right)$ | $-11$ | $-10$ | $-7\frac{1}{2}$ | $-3$ | $-4\frac{1}{2}$ | $3\frac{1}{2}$ | $6\frac{1}{2}$ | 7 | 9 | 10 |

Сумма квадратов отклонений равна 591. Подсчитаем теперь соответствующие величины $T'$:

$$P: \quad \frac{1}{12}\times 2\,(2^3-2)=1;$$

$$Q: \quad \frac{1}{12}\times 3\,(2^3-2)=1\frac{1}{2};$$

$$R: \quad \frac{1}{12}\times (4^3-4+3^3-3)=7.$$

Таким образом, из (6.8) получим:

$$W=\frac{591}{742,5-28,5}=0,828.$$

Из этого непосредственно следует, что вносимая нами поправка на наличие связей в данном случае не оказывает существенного влияния на результаты.

6.8. Перейдем теперь к вопросу о том, как проверить существенность наблюдаемого значения $W$. Если эксперты независимы друг от друга в своих суждениях, то та или иная система рангов столь же вероятна, как и любая другая. В связи с этим будем рассматривать распределения $W$ на множестве, состоящем из $(n!)^m$ возможных систем рангов. Прибегая к обычным методам проверки статистических гипотез, будем полагать, что суждения экспертов не характеризуются общностью предпочтений.

Действительно, существующие распределения $W$ были исследованы лишь для малых $m$ и $n$: например, при $n = 3$ и $m$, изменяющемся от 2 до 10; при $n = 4$ и $m$, принимающем значения от 2 до 6; при $n = 5$ и $m = 3$. Соответствующие данные можно найти в табл. 5, приведенной в приложении.

Для характеристики распределения при больших значениях $m$ и $n$ можно использовать следующие два приближенных метода вычислений.

1. Если нам приходится иметь дело с любыми значениями*, отличными от приведенных в табл. 5 приложения, мы можем прибегнуть к аппроксимации, основанной на использовании известного в статистике $z$-распределения Фишера. Будем полагать

$$z = \frac{1}{2} \log_e \frac{(m-1)W}{1-W}, \tag{6.11}$$

$$\left.\begin{aligned} \nu_1 &= n - 1 - \frac{2}{m} \\ \nu_2 &= (m-1)\nu_1 \end{aligned}\right\}. \tag{6.12}$$

Тогда, зная «степени свободы» $\nu_1$ и $\nu_2$, можно проверить значение $z$ пользуясь приведенными в ряде источников таблицами распределения Фишера. Можно, однако, обойтись без этого, прибегнув непосредственно к данным приведенной в приложении табл. 5; в ней содержатся значения $S$, соответствующие 5%-му и 1%-му уровням существенности величины $z$, если при этом $m$ принимает значения от 3 до 20, а $n$ — от 3 до 7. Несколько ниже будет приведен пример, показывающий, как пользоваться этой таблицей.

2. Выше был предложен общий метод проверки; однако, если $n > 7$, можно прибегнуть к более простой процедуре. Будем полагать

$$\chi_r^2 = m(n-1)W = \frac{S}{\frac{1}{12}mn(n+1)}. \tag{6.13}$$

В таком случае распределение $\chi_r^2$ совпадает с известным в статистике распределением $\chi^2$ с $\nu = n - 1$ степенями свободы.

*Пример 6.2*

Предположим, что мы имеем дело с 18 последовательностями, каждая из которых содержит по 7 рангов. При этом значение $S$ равно 1620, следовательно,

$$W = \frac{1620}{\frac{1}{12} \times 18^2 \times 336} = 0{,}179.$$

В табл. 6, приведенной в приложении, можно найти следующие величины $S$, соответствующие 5%-му уровню существенности:

$$m = 15, \quad S = 864{,}9,$$
$$m = 20, \quad S = 1158{,}7$$

и величины $S$, соответствующие 1%-му уровню существенности:

$$m = 15, \quad S = 1129{,}5,$$
$$m = 20, \quad S = 1521{,}9.$$

* Имеются в виду значения $m$ и $n$. — *Прим. перев.*

Величина $S$ при $m = 18$ лежит между значениями, соответствующими $m = 15$ и $m = 20$, и, стало быть, рассчитанная нами величина $S$ оказывается больше, чем табличное значение $S$, соответствующее 1%-му уровню существенности. Из этого следует, что вероятность получения числа, равного наблюдаемому или превосходящего его, составляет менее 0,01, или, как иногда говорят, значение лежит выше 1%-го уровня. Таким образом, исчисленная величина существенна, если мы исходим из того, что столь малые вероятности могут характеризовать статистическую существенность.

*Пример 6.3*

Пусть даны 28 последовательностей, каждая из которых содержит по 13 членов. При этом величина $S = 11\,440$ и, следовательно,

$$W = \frac{11\,440}{\frac{1}{12} \times 28^2 \times 2184} = 0{,}080.$$

Мы можем проверить эти вычисления, пользуясь методом (6.13). Действительно,

$$\chi_r^2 = \frac{11\,440}{\frac{1}{12} \times 28 \times 13 \times 14} = 27.$$

В приведенной в приложении табл. 8 можно найти соответствующую 1%-му уровню существенности величину $\chi^2$; при $\nu = n - 1 = 12$ степенях свободы $\chi^2 = 26{,}217$. Полученное нами значение несколько больше этого числа, стало быть, оно «в точности существенно» при уровне существенности в 1%.

*Пример 6.4*

На практике часто приходится иметь дело с ситуациями, когда $m$ столь велико, что не имеет смысла делать поправки для всех промежуточных значений $m$. Если поправка все же требуется, ее можно ввести, добавив 2 к знаменателю и уменьшив на 1 числитель в формуле

$$W = \frac{S}{\frac{1}{12} m^2 (n^3 - n)}.$$

Рассмотрим в качестве примера случай, когда $n = 3$, $m = 9$; предположим, что $S = 78$. С помощью приведенной в приложении табл. 5А, можно выяснить, что вероятность такого или большего значения $S$ равна 0,010, так что это приблизительно соответствует 1%-му уровню существенности. Предположим теперь, что, проверяя существенность этих данных, мы воспользовались $z$-критерием. Тогда, введя поправку на непрерывность значений, можно получить:

$$W = \frac{78-1}{\frac{1}{12}(81 \times 24) + 2} = 0{,}4695;$$

$$z = 0{,}979; \quad \nu_1 = \frac{16}{9}; \quad \nu_2 = \frac{128}{9}.$$

Проведя линейную интерполяцию указанных в приведенной в приложении табл. 7В обратных величин, мы можем по значениям $\nu_1$ и $\nu_2$ отыскать табличную величину $z$, равную 0,954 (в то время как вычисленное нами значение равно 0,979). Даже при столь малом значении $n$ ($n = 3$) проверка по $z$-критерию обеспечивает удовлетворительную аппроксимацию.

**6.9.** Наличие связей в последовательностях рангов не требует изменений в методах проверки по $z$-критерию, за исключением тех случаев, когда число связей велико или значительна их протяженность. В последнем случае процедура проверки усложняется. Обозначим через $\mu_{2i}$ дисперсию, исчисленную для $i$-й последовательности, характеризуемой выражением $\frac{1}{12}(n^2 - 1) - \frac{1}{n}T'$. Запишем выражение

$$\mu_2(W) = \frac{4}{m^2(n-1)} \frac{\sum_{i,j} \mu_{2i}\mu_{2j}}{(\Sigma\mu_{2i})^2}, \tag{6.14}$$

причем суммирование производится по $\frac{1}{2}m(m-1)$ значениям $i \neq j$. Тогда статистическая существенность $W$ может быть проверена с помощью $z$-критерия по формуле (6.11), причем в расчет числа степеней свободы вносятся следующие изменения:

$$\left.\begin{aligned} \nu_1 &= \frac{2(m-1)}{m^3\mu_2(W)} - \frac{2}{m} \\ \nu_2 &= (m-1)\nu_1 \end{aligned}\right\}. \tag{6.15}$$

В таком случае соответствующее значение $\chi_r^2$ может быть определено из формулы

$$\chi_r^2 = \frac{S}{\frac{1}{12}mn(n+1) - \frac{1}{n-1}\Sigma T'}. \tag{6.16}$$

### Статистическая оценка

**6.10.** Теперь предположим, что исчисленный нами коэффициент $W$ обладает статистической существенностью и, следовательно, можно полагать, что между оценками экспертов существует некоторая согласованность. Попытаемся сделать еще один шаг в указанном направлении: предположим, что можно более или менее точно упорядочить суждения экспертов в соответствии с некоторой объективной шкалой.

Тогда мы можем поставить следующий вопрос: как установить «подлинную» последовательность объектов или как можно получить наилучшую оценку такой «подлинной» последовательности?

**6.11.** Предположим, что мы имеем дело с тремя последовательностями, каждая из которых содержит по восемь рангов.

| | | Объект | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* | |
| Эксперт | *P* | 4 | 2 | 1 | 7 | 6 | 3 | 5 | 8 | |
| » | *Q* | 7 | 2 | 1 | 6 | 4 | 5 | 3 | 8 | (6.17) |
| » | *R* | 7 | 4 | 2 | 6 | 5 | 3 | 1 | 8 | |
| Суммарное значение оценок | | 18 | 8 | 4 | 19 | 15 | 11 | 9 | 24 | |

Нельзя ли упорядочить объекты, исходя из числа «первых мест», «вторых мест» и т. д., занимаемых каждым объектом в последовательностях различных экспертов? Например, мы могли бы приписать объекту *C* общую ранговую оценку 1, поскольку этот объект у двух экспертов был выдвинут на первое место. Объект *G* один раз занял первое место, и мы могли бы приписать ему общую ранговую оценку, равную 2. Затем обратимся к объектам, занявшим вторые места: объект *B* занял второе место у двух экспертов; поэтому ему приписывается общая ранговая оценка, равна трем. Один из экспертов отвел второе место *C*, но этот объект уже проранжирован. Затем переходим к объектам, занявшим третьи места, и т. д. В конечном счете получаем ряд:

$$C\ G\ B\ F\ E\ A\ D\ H. \tag{6.18}$$

Два объекта *E* и *A* заняли четвертое место по одному разу, однако мы поставили *E* перед *A*, поскольку *E* у одного из оставшихся экспертов занял пятое место, а объект *A* у двух оставшихся экспертов занял лишь седьмое место.

Однако от такого подхода приходится отказаться, поскольку он содержит внутренние противоречия. Предположим, мы начали упорядочение с другого конца, так что ранг 8 присвоен объекту, занявшему наибольшее число раз восьмое место. В таком случае рассматривавшиеся ранее объекты в конце концов окажутся расположенными в такой последовательности:

$$C\ B\ G\ F\ E\ D\ A\ H. \tag{6.19}$$

Полученные последовательности отличаются друг от друга. Вообще говоря, у нас нет особых причин для того, чтобы всегда начинать упорядочение с какого-то одного конца (а не с другого). Поскольку два способа применения рассмотренной процедуры упорядочения приводят к различным результатам, эту процедуру, разумеется, приходится признать неудовлетворительной.

**6.12.** Более эффективным оказывается метод упорядочения, основанный на использовании сумм рангов, приписываемых каждому объекту. Следуя такой процедуре, можно заключить, что объект $C$ в (6.17) обладает наименьшей суммой рангов, так что общий ранг можно положить равным единице, объект $B$ занимает следующее место и т. д. В итоге получаем последовательность

$$C\ B\ G\ F\ E\ A\ D\ H, \tag{6.20}$$

которая, как нетрудно заметить, отличается от последовательности (6.18) и от последовательности (6.19).

Можно доказать (мы докажем это утверждение в следующей главе), что с точки зрения некоторого критерия, используемого обычно в расчетах по методу наименьших квадратов, такая оценка является «наилучшей».

Действительно, рассмотрим сумму квадратов разностей между действительно наблюдаемыми суммами рангов и значениями таких величин в ситуации, когда все последовательности совпадают между собой; эта сумма принимает минимальное значение в том случае, когда совокупные ранги оцениваются в соответствии с описанной выше процедурой. Кроме того, если измерить соответствие между построенной таким образом последовательностью рангов и наблюдаемыми последовательностями с помощью коэффициента $\rho$ Спирмэна, среднее значение этого коэффициента окажется больше, чем при сравнении наблюдаемых последовательностей с любой другой построенной последовательностью рангов. Сравнение указанных последовательностей с помощью коэффициента $\tau$ может не обнаружить такого свойства, однако если число сравниваемых между собой объектов не слишком велико, то приведенное выше утверждение оказывается верным и применительно к коэффициенту $\tau$.

**6.13.** Несколько слов следует сказать также в отношении проблемы связей рангов.

а. Если наблюдаемые последовательности не содержат связей и мы исключаем возможность, что такие связи все же существуют, в некоторых случаях могут возникать дополнительные проблемы.

Предположим, что некоторые три объекта были упорядочены следующим образом:

| | | $X$ | $Y$ | $Z$ |
|---|---|---|---|---|
| Эксперт | $P$ | 7 | 8 | 10 |
| » | $Q$ | 9 | 8 | 6 |
| » | $R$ | 3 | 7 | 6 |
| » | $S$ | 5 | 1 | 2 |
| Сумма рангов | | 24 | 24 | 24 |

(6.21)

Сумма рангов всех трех объектов совпадает, наш метод не дает критерия, с помощью которого можно провести упорядочение. Если мы допускаем наличие связей между рангами, то правомерно полагать, что все объекты имеют одинаковый ранг. Если все же возможность существования связей отвергается, то наиболее логичным представляется следующее решение: мы отдаем предпочтение тому объекту, у которого ранги в наименьшей степени различаются. Так как суммы рангов одинаковы, приведенное выше утверждение эквивалентно такому: первое место приписывается объекту с наименьшей суммой квадратов рангов. В примере (6.21) суммы квадратов рангов для объектов $X$, $Y$, $Z$ соответственно равны 164, 178, 176, так что их можно упорядочить следующим образом: $X$, $Z$, $Y$.

б. Если рассматривать ранг объекта, равный $j$, как выражение того факта, что $j-1$ элементов последовательности получили предпочтение, то сумма рангов (за вычетом $m$) характеризует как бы сумму предпочтений и, следовательно, описанный метод позволяет ранжировать объекты в соответствии с предпочтениями. Если некоторые ранги связаны между собой, то замещение суммы рангов связями не искажает соотношений с другими членами группы, а просто снимает вопрос о соотношении предпочтений внутри связанной группы. Таким образом, описанный метод позволяет упорядочить объекты в соответствии с предпочтениями даже в тех случаях, когда существуют связи.

**6.14.** Мало разработана следующая проблема: как распределена величина $W$ в случае, когда существует некоторая общность предпочтений. Поэтому мы не можем проверить статистическую существенность различий между системами, каждая из которых содержит несколько последовательностей рангов. Предположим, что каждый из 8 преподавателей-мужчин ранжировал учащихся в классе, состоящем из 30 мальчиков; результирующее значение $W$ для всех последовательностей оказалось равным 0,5. Предположим также, что 6 преподавателей-женщин проранжировали учащихся того же класса, величина $W$ в этом случае оказалась равной 0,7. Приведенные значения показывают, что преподаватели-женщины обладают более общей системой предпочтений, чем преподаватели-мужчины; однако если мы будем рассматривать этих мужчин как выборку из всей совокупности учителей-мужчин (то же самое относится и к женщинам), то нет способа, с помощью которого удалось бы проверить статистическую существенность гипотезы о большей общности суждений преподавателей-женщин по сравнению с преподавателями-мужчинами. Здесь налицо пробел в наших знаниях, и было бы весьма полезным его заполнить.

**6.15.** Методы статистической проверки гипотез в условиях существования связей между рангами пока, как отмечалось выше, мало разработаны; но в ряде случаев удается обнаружить важные эффекты, прибегнув к непосредственному анализу исходных данных. Если заранее (a priori) существует «подозрение», что вкусы внутри некоторой группы экспертов совершенно разнородны, объединение оценок различных экспертов в одну группу может привести к искажению выявленных предпочтений. Рассмотрим, например, предельный случай: десять экспертов оказались единодушными в своих оценках шести объектов, суммы рангов соответственно составляют:

$$10,\ 20,\ 30,\ 40,\ 50,\ 60,$$

так что $W = 1$. Предположим при этом, что десять других экспертов также совершенно одинаково оценили эти объекты, но суммы рангов оценки теперь расположены в противоположном порядке:

$$60,\ 50,\ 40,\ 30,\ 20,\ 10,$$

причем вновь $W = 1$. Рассмотрим теперь суммарные ранговые оценки всех 20 экспертов: для каждого объекта оценка, разумеется, равна 70, а $W = 0$. Из последнего расчета должен следовать вывод о том, что среди группы экспертов не существует никакой общности предпочтений, тогда как в действительности единодушие одной части экспертов полностью скрадывается общностью суждений другой группы экспертов.

## Неполные последовательности рангов

**6.16.** В ряде случаев приходится иметь дело с неполными последовательностями рангов. Представляют интерес, в частности, проблемы использования неполных последовательностей при планировании эксперимента, поскольку мы можем «сконструировать» такие неполные последовательности, которые будут обладать свойством симметрии.

Рассмотрим, например, стратегию предпринимателя, выпускающего мороженое. Он хотел бы выявить предпочтения покупателей относительно семи сортов мороженого, исходя из упорядочения суждений нескольких потребителей. При этом предприниматель может не захотеть давать на пробу каждому дегустатору все сорта мороженого — то ли потому, что он стремится экономить время, то ли потому, что вкус человека как бы «притупляется» после дегустации нескольких сортов мороженого. Назовем сорта $A$, $B$, $C$, $D$, $E$, $F$, $G$ и образуем следующие группы, каждая из которых включает три сорта мороженого:

$$\begin{matrix} A & B & C & D & E & F & G \\ B & C & D & E & F & G & A \\ D & E & F & G & A & B & C \end{matrix} \qquad (6.22)$$

Будем полагать, что каждая группа, содержащая по три сорта, выражает предпочтения одного из семи дегустаторов. Смысл такого разбиения состоит в следующем: каждый объект подвергается экспериментальной оценке одинаковое число раз (в нашем случае оценивается трижды), кроме того, в эксперименте сопоставляется каждая пара объектов по одному разу, так что все элементы охвачены сопоставлением одинаковое число раз. Указанную систему можно рассматривать как семь последовательностей рангов семи объектов, но при этом в каждой последовательности отсутствуют оценки четырех элементов[1].

[1] Поэтому подобный план эксперимента называют также неполным латинским квадратом.

**6.17.** Предположим, что дегустаторы расположили свои оценки указанных выше сортов следующим образом:

$$\begin{matrix} 2 & 2 & 1 & 1 & 2 & 2 & 3 \\ 1 & 3 & 2 & 3 & 3 & 3 & 1 \\ 3 & 1 & 3 & 2 & 1 & 1 & 2 \end{matrix} \tag{6.23}$$

Выпишем суммы рангов каждого из семи объектов:

$$\begin{matrix} A & B & C & D & E & F & G \\ 4 & 4 & 6 & 6 & 6 & 8 & 8 \end{matrix} \tag{6.24}$$

Если бы индивидуальные ранговые оценки были случайно распределены между объектами, то существовала бы тенденция к выравниванию сумм рангов по различным группам; с другой стороны, если бы ранги однозначно соответствовали месту данного объекта при строгом упорядочении системы, то суммы рангов были бы равны соответственно 3, 4, ..., 9 и различались бы между собой настолько, насколько это вообще возможно. Поэтому, как обычно, будем полагать, что $S$ представляет собой сумму квадратов отклонений чисел, приведенных в последовательности (6.24), от их среднего значения; тогда можно построить коэффициент $W$, разделив эту сумму на максимальное значение $S$. В нашем примере численная величина $W$ может быть определена следующим образом:

$$W = \frac{16}{28} = 0{,}571.$$

**6.18.** Схему эксперимента, подобную (6.22), называют квадратом Юдена[1]. Если из общего числа объектов, равного $n$, в эксперименте каждый раз представлено $k$ объектов и при этом любой из объектов участвует в эксперименте в общей сложности $m$ раз, то всего потребуется образовать $mn/k$ групп. При этом в каждой группе будет сопоставляться между собой $\frac{1}{2}k(k-1)$ пар, следовательно, общее число таких сопоставлений составит $\frac{1}{2}mn(k-1)$. Буквой $\lambda$ обозначим общее число групп, в которых встречается некая заданная пара элементов. В таком случае

$$\frac{1}{2}n(n-1)\lambda = \frac{1}{2}mn(k-1)$$

или

$$\lambda = \frac{m(k-1)}{n-1}. \tag{6.25}$$

$\lambda$ — это целое число и поэтому произведение $m(k-1)$ должно содержать в качестве множителя число $n-1$. Кроме того, $mn$ должно нацело делиться на $k$. Эти условия налагают существенные ограничения на числа, используемые при составлении такого рода

[1] Ряд схем такого вида приводится в книге: W. G. Cochran and G. M. Cox, Experimental Design, 1951, John Wiley and Sons, New York.

схем. В приведенном выше примере $n = 7$, $k = 3$ и, следовательно, $m$ должно делиться на 3 (в нашем примере оно просто было равно 3); поэтому из соотношения (6.25) следовало: $\lambda = 1$.

Допустим, что в последовательностях рангов нет связей, обозначим через $x_{ij}$ ранг, приписанный $j$-му объекту $i$-м экспертом (тем самым предполагается наличие $i$-го блока, если данное разбиение на блоки однократно используется в эксперименте, и $i$-х блоков, если эксперимент повторяется). $S$ будет означать сумму квадратов отклонений $x_j = \sum_{i=1}^{m} x_{ij}$ от среднего значения этих величин. Она принимает максимальное значение в том случае, когда суммы рангов распределяются по объектам следующим образом:

$m$, $m + \lambda$, $m + 2\lambda$, ..., $m + (n - 1)\lambda$ и, следовательно, максимальное значение $S^2 = \frac{\lambda^2 n (n^2 - 1)}{12}$.

Таким образом, в общем случае

$$W = \frac{12S}{\lambda^2 n (n^2 - 1)} = \tag{6.26}$$

$$= \frac{12 \sum_{1}^{n} x_j^2 - 3nm^2 (k+1)^2}{\lambda^2 n (n^2 - 1)}. \tag{6.27}$$

Не всегда легко ввести поправку, когда имеются связи, поэтому нам остается либо игнорировать их, если они не слишком многочисленны, либо просто «на глаз» приписать им некоторые суммарные ранги. Это не отразится на проверке существенности.

**6.19.** При ограниченном $n$ и больших $m$ статистическую существенность $W$ можно проверить следующим образом: будем полагать, что величина

$$\chi^2 = \frac{\lambda (n^2 - 1)}{k+1} W \tag{6.28}$$

распределена как $\chi^2$ с $n - 1$ степенью свободы. Для более точной проверки можно использовать $z$-распределение.

$$z = \frac{1}{2} \log \frac{\left(\frac{\lambda (n+1)}{k+1} - 1\right) W}{1 - W} \tag{6.29}$$

при

$$\left.\begin{aligned} \nu_1 &= \frac{mn\left(1 - \frac{k+1}{\lambda (n+1)}\right)}{\left(\frac{nm}{n-1} - \frac{k}{k-1}\right)} - \frac{2(k+1)}{\lambda (n+1)} \\ \nu_2 &= \left(\frac{\lambda (n+1)}{k+1} - 1\right) \nu_1 \end{aligned}\right\} \tag{6.30}$$

степенях свободы.

Если $k = n$ и, следовательно, $\lambda = m$, эти выражения, как и следовало ожидать, сводятся к соотношениям (6.11) и (6.13).

*Пример* 6.5 [19]

Предположим, что мы трижды воспользовались планом эксперимента (6.22) и что в результате мы будем располагать 21 блоком оценок, каждый из этих блоков принадлежит соответствующему эксперту. Суммы рангов равны:

| *A* | *B* | *C* | *D* | *E* | *F* | *G* |
|---|---|---|---|---|---|---|
| 20 | 13 | 18 | 25 | 22 | 12 | 16 |

Исчисленная на основе этих данных средняя ранговая оценка равна 18. $S = 134$, $n = 7$, $m = 9$, $k = 3$ и, следовательно, $\lambda = 3$. Таким образом,

$$W = \frac{12 \times 134}{9 \times 7 \times 48} \doteq 0{,}532.$$

Для проверки существенности подставляем соответствующие числовые значения в (6.29) и (6.30).

$$\nu_1 = 5{,}5;\ \nu_2 = 27{,}5;\ z = 0{,}87.$$

В приведенной в приложении таблице 7B находим следующие значения для $\nu_1 = 5$ и $\nu_1 = 6$; $\nu_2 = 27$ и $\nu_2 = 28$:

| | |
|---|---|
| 0,6655 | 0,6346 |
| 0,6614 | 0,6303 |

Следовательно, полученная оценка характеризуется 1%-ным уровнем статистической существенности. Отсюда следует вывод, что такое упорядочение предпочтений не является случайным.

## Библиография

См. [29], [30], [56] и [110].

Квадраты Юдена описываются в [19]; в этой работе показано также, что «истинное» упорядочение в использовании сбалансированных планов указанного типа можно получить методом, описанным в 6.10 (упорядочение объектов проводится в соответствии с суммой приписанных им рангов). Подобная процедура обеспечивает минимизацию суммы взвешенных инверсий в наблюдаемых последовательностях рангов относительно оцененной последовательности.

Общий случай (когда в каждой из последовательностей может отсутствовать любое количество рангов) описан в работе [1].

Методы, изложенные в 6.16—6.19, применимы не только к квадратам Юдена, но ко всяким планам проведения эксперимента с неполными блоками, включая планы попарных сопоставлений. Более подробно эти вопросы рассматриваются в [19]. Эффективность этих методов анализируется в [109].

## *ГЛАВА 7.* ДОКАЗАТЕЛЬСТВО РЕЗУЛЬТАТОВ ГЛАВЫ 6

**7.1.** Рассмотрим следующий вопрос: насколько правомерно использовать $z$-распределение для статистической проверки существенности коэффициента конкордации $W$ на множестве $(n!)^m$ возможных последовательностей рангов. Поскольку нам потребуются некоторые общие результаты, предполагающие наличие связанных последовательностей, мы приведем здесь, следуя [78], постановку задачи в общем виде.

**7.2.** Предположим, что нам дано $m$ последовательностей чисел:

$$\begin{matrix} a_1 & a_2 & \dots & a_n \\ b_1 & b_2 & \dots & b_n \\ \vdots & \vdots & \dots & \vdots \\ k_1 & k_2 & \dots & k_n \end{matrix} \tag{7.1}$$

В дальнейшем будем полагать, что каждое число характеризует отклонение от среднего значения элементов соответствующей строки, следовательно, средняя из элементов любой строки и среднее значение всех чисел данного массива равны нулю. В таком случае мы можем записать:

$$W=\frac{\frac{1}{m}\sum_{1}^{n}(a_j+b_j+\ldots+k_j)^2}{\sum_{1}^{n}a_j^2+\sum_{1}^{n}b_j^2+\ldots+\sum_{1}^{n}k_j^2}. \tag{7.2}$$

Знаменатель этой дроби представляет собой постоянную величину, так что изменения коэффициента $W$ могут быть связаны только с изменениями в числителе дроби. Обозначим через $\alpha_2$ вторые моменты строк, содержащих элементы $a$, $b$, $c$ и т. д.; в этом случае знаменатель рассматриваемой дроби можно записать так: $n\Sigma\alpha_2$. Кроме того, введем следующие обозначения:

$$R_{ab} = \sum_{1}^{n} a_i\, b_i\,, \tag{7.3}$$

$$U = \sum_{1}^{\binom{m}{2}} R_i, \tag{7.4}$$

причем $R_i$ представляет собой обозначение любого $R_{ab}$ при всех возможных сочетаниях элементов $a$ и $b$. Тогда $W$ записывается следующим образом:

$$W = \frac{1}{m} + \frac{2U}{mn\Sigma\alpha_2}\,. \tag{7.5}$$

Теперь определим моменты: сначала для $R_{ab}$, затем для $U$ и, наконец, для $W$.

7.3. Буквой $E$ будем обозначать математическое ожидание соответствующих величин. В этом случае

$$\begin{aligned}
E(R_{ab}) &= 0,\\
E(R_{ab}^2) &= E(\Sigma a_i\, b_i)^2 =\\
&= E(\Sigma a_i^2\, b_i^2 + \Sigma'\, a_i\, b_i\, a_j\, b_j) =\\
&= nE(a_i^2\, b_i^2) + n(n-1)\, E(a_i\, a_j\, b_i\, b_j) =\\
&= n\, E(a_i^2)\, E(b_i^2) + \frac{n(n-1)}{n^2(n-1)^2} \times\\
&\times [E(\Sigma a_i)^2 - E\Sigma a_i^2]\, [E(\Sigma b_i)^2 - E\Sigma b_i^2] =\\
&= n\alpha_2\, \beta_2 + \frac{n}{n-1}\, \alpha_2\, \beta_2 = \frac{n^2\, \alpha_2\, \beta_2}{n-1}\,.
\end{aligned} \tag{7.6}$$

Аналогично можно показать (промежуточные вычисления опущены), что

$$E(R_{ab}^3) = \frac{n^3\, \alpha_3\, \beta_3}{(n-1)(n-2)} = \frac{(n-1)(n-2)}{n}\, \alpha_3'\, \beta_3', \tag{7.7}$$

$$E(R_{ab}^4) = \frac{3n^4\, \alpha_2^2\, \beta_2^2}{(n-1)(n+1)} + \frac{(n-1)(n-2)(n-3)}{n(n+1)}\, \alpha_4'\, \beta_4', \tag{7.8}$$

где символ $\alpha_3$ обозначает третий момент $a$ и где $\alpha_3'$, $\alpha_4'$ характеризуют ее третью и четвертую $k$-статистику*, которые можно выразить через соответствующие моменты:

$$\alpha_3' = \frac{n^2}{(n-1)(n-2)}\, \alpha_3,$$

$$\alpha_4' = \frac{n}{(n-1)(n-2)(n-3)}\, [(n+1)\, \alpha_4 - 3(n-1)\, \alpha_2^2].$$

---

* Более подробно свойства $k$-статистик для выборок из конечных совокупностей рассматриваются, например, в книге Кендалл М., Стьюарт А. Теория распределений. М., «Наука», 1966, с. 385—394, 415—419. — *Прим. ред.*

Смысл применения $k$-статистик состоит в следующем: для совокупностей нормально распределенных величин все $k$-статистики выше второго порядка обращаются в нуль; поэтому можно полагать, что для достаточно близких к нормальному распределений значения $k$-статистик будут малы.

7.4. Прежде чем приступить к вычислению моментов для $U$, отметим, что

$E(R_{ij}R_{kl}) = 0$ при всех значениях индексов, за исключением случаев, когда $i = k$, $j = l$;

$E(R_{ij}R_{kl}R_{mn}) = 0$, если только значения индексов не образуют «замкнутой» последовательности типа $ij$, $jk$, $ki$. Если рассматривается произведение, содержащее четыре множителя, то аналогичным образом в нуль не будут обращаться лишь те из них, индексы которых образуют замкнутую последовательность $ij$, $jk$, $kl$, $li$. Следовательно,

$$E(U) = E(\Sigma R) = 0, \tag{7.9}$$

$$E(U^2) = E(\Sigma R_i^2 + \Sigma' R_i R_j) = \frac{n^2}{n-1} \Sigma \alpha_2 \beta_2. \tag{7.10}$$

Далее

$$\begin{aligned} E(R_{ij} R_{jk} R_{ki}) &= E(\Sigma a_i b_i \, \Sigma b_i c_i \, \Sigma c_i a_i) = E\{[\Sigma b_i^2 a_i c_i + \\ &+ \Sigma' b_i b_j (a_i c_j + a_j c_i)] \, \Sigma c_i a_i\} = E[E b_i^2 \Sigma a_i c_i + \\ &+ E b_i b_j \Sigma (a_i c_j + a_j c_i)] \, \Sigma c_i a_i = \\ &= (E b_i^2 - E b_i b_j) \, E(\Sigma a_i c_i)^2 = \left(\beta_2 + \frac{\beta_2}{n-1}\right) \frac{n^2 \alpha_2 \gamma_2}{n-1} = \\ &= \frac{n^3 \alpha_2 \beta_2 \gamma_2}{(n-1)^2}, \end{aligned}$$

$$\Sigma E(R_i^3) = \frac{(n-1)(n-2)}{n} \Sigma \alpha_3' \beta_3'.$$

Следовательно,

$$E(U^3) = \frac{6n^3}{(n-1)^2} \Sigma \alpha_2 \beta_2 \gamma_2 + \frac{(n-1)(n-2)}{n} \Sigma \alpha_3' \beta_3' \tag{7.11}$$

и, опуская промежуточные вычисления, выпишем конечные результаты:

$$\begin{aligned} E(U^4) &= \frac{3n^4}{(n-1)(n+1)} \Sigma \alpha_2^2 \beta_2^2 + \frac{(n-1)(n-2)(n-3)}{n(n+1)} \Sigma \alpha_4' \beta_4' + \\ &+ \frac{3n^4}{(n-1)^2} [(\Sigma \alpha_2 \beta_2)^2 - \Sigma \alpha_2^2 \beta_2^2] + 12(n-2) \Sigma \alpha_3' \beta_3' \gamma_2 + \\ &+ \frac{72n^4}{(n-1)^3} \Sigma \alpha_2 \beta_2 \gamma_2 \delta_2. \end{aligned} \tag{7.12}$$

Наконец, перейдем к рассмотрению моментов $W$:

$$E(W) = \frac{1}{m}, \tag{7.13}$$

обозначим эту величину $\overline{W}$.

$$E(W-\overline{W})^2 = \frac{4}{m^2(n-1)} \frac{\Sigma\alpha_2\beta_2}{(\Sigma\alpha_2)^2}. \tag{7.14}$$

Пренебрегая слагаемыми, содержащими $\alpha_3'$ и $\alpha_4'$ (поскольку они представляют собой величины меньшего порядка относительно $n$), можно утверждать, что

$$E(W-\overline{W})^3 = \frac{48}{m^3(n-1)^2} \frac{\Sigma\alpha_2\beta_2\gamma_2}{(\Sigma\alpha_2)^3}; \tag{7.15}$$

$$E(W-\overline{W})^4 = \frac{48}{m^4(n-1)^2} \frac{(\Sigma\alpha_2\beta_2)^2}{(\Sigma\alpha_2)^4} - \frac{96}{m^4(n-1)^2(n+1)} \frac{\Sigma\alpha_2^2\beta_2^2}{(\Sigma\alpha_2)^4} + \\ + \frac{1152}{m^4(n-1)^3} \frac{\Sigma\alpha_2\beta_2\gamma_2\delta_2}{(\Sigma\alpha_2)^4}. \tag{7.16}$$

**7.5.** Теперь рассмотрим случай, когда не существует связей. В таком случае дисперсии всех последовательностей равны между собой и, слеоовательно,

$$E(W) = \frac{1}{m}; \tag{7.17}$$

$$\mu_2(W) = E(W-\overline{W})^2 = \frac{2(m-1)}{m^3(n-1)}, \tag{7.18}$$

$$\mu_3(W) = E(W-\overline{W})^3 = \frac{8(m-1)(m-2)}{m^5(n-1)^2}, \tag{7.19}$$

$$\mu_4(W = EW-\overline{W})^4 = \frac{12(m-1)^2}{m^6(n-1)^2} + \frac{48(m-1)(m-2)(m-3)}{m^7(n-1)^3} - \\ - \frac{48(m-1)}{m^7(n-1)^2(n+1)}. \tag{7.20}$$

Рассмотрим теперь распределение

$$dF = \frac{1}{B(p, q)} W^{p-1}(1-W)^{q-1}\, dW. \tag{7.21}$$

Первые два момента равны:

$$E(W) = \frac{p}{p+q}, \tag{7.22}$$

$$\mu_2(W) = \frac{pq}{(p+q)^2(p+q+1)}. \tag{7.23}$$

Приравняв между собой соотношения (7.17) и (7.22), а также (7.18) и (7.23), можно найти, что

$$\left.\begin{aligned} p &= \frac{1}{2}(n-1) - \frac{1}{m} \\ q &= (m-1)p \end{aligned}\right\}. \tag{7.24}$$

Таким образом, распределение $W$ служит аппроксимацией распределения (7.21), если $p$ и $q$ принимают указанные значения. Для (7.21) третий момент равен:

$$\frac{8(m-1)(m-2)}{m^4(n-1)(mn-m+2)}=\frac{8(m-1)(m-2)}{m^5(n-1)^2}\left[1-\frac{2}{m(n-1)+2}\right].$$

Сопоставляя это выражение с (7.19), можно видеть, что для величин, характеризующихся распределением (7.21), третий момент можно полагать приблизительно равным третьему моменту $W$, если только произведение $m(n-1)$ не представляет собой малой величины. Аналогичным образом вычисляем четвертый момент для (7.21), равный

$$\frac{12(m-1)}{m^6(n-1)^2}\left[\frac{(m-1)y^2+4m^2y-14(m-1)y}{(y+2)(y+4)}\right],$$

где $y=m(n-1)$. Это выражение приблизительно равно четвертому моменту $W$, если только произведение $m(n-1)$ принимает достаточно большие значения.

**7.6.** Таким образом, два первых момента распределения (7.21) всегда в точности совпадают с соответствующими моментами действительного распределения $W$, а третий и четвертый моменты распределения (7.21) приблизительно равны соответствующим моментам $W$; в связи с этим мы предполагаем, что распределение (7.21) достаточно хорошо аппроксимирует действительное распределение $W$. На самом деле точность аппроксимации выше, чем можно было бы предположить, глядя на приведенные выше довольно громоздкие доказательства.

В теоретической статистике распределение (7.21) называют $\beta$-распределением или распределением типа I*. С помощью несложных преобразований можно свести это распределение к $z$-распределению Фишера. Действительно, полагая

$$z=\frac{1}{2}\log\frac{(m-1)W}{1-W},$$

находим, что выражение (7.21) приводится к виду:

$$dF\propto\frac{e^{2pz}\,dz}{[(m-1)+e^{2z}]^{p+q}},$$

т. е. к распределению Фишера, при

$$\nu_1=2p,$$
$$\nu_2=2q.$$

---

* Более подробная характеристика таких распределений приведена в книге Кендалл М., Стьюарт А. Теория распределений. М., «Наука» 1966, гл. 6. — *Прим. перев.*

Таким образом, из (7.24) следует

$$\nu_1=(n-1)-\frac{2}{m},$$
$$\nu_2=(m-1)\nu_1,$$

что совпадает со значениями, которые были приведены в формуле (6.12).

7.7. Более подробно следует рассмотреть случай, когда имеют место связи.

а. В предыдущих рассуждениях мы пользовались отсутствием связей лишь тогда, когда вычисляли величины $\alpha_2$, так что если последовательности имеют одинаковые связи, то дисперсии по-прежнему будут оставаться равными друг другу и наши выводы сохраняют справедливость.

б. Если числа $T'$, соответствующие связям, малы по сравнению с величиной $N=\frac{1}{12}(n^3-n)$, то процедура статистической проверки не требует изменений. Тогда, ограничиваясь лишь теми слагаемыми, которые содержат $\frac{\Sigma T'}{mN}$ в первой степени, можно записать:

$$\frac{\Sigma\alpha_2\beta_2}{(\Sigma\alpha_2)^2}=\frac{\Sigma(N-T'_\alpha)(N-T'_\beta)}{\Sigma(N-T'_\alpha)^2}=\frac{(m-1)m}{m^2}\left[1-\frac{2}{m\,N}\Sigma T'\right]\times$$
$$\times\left[1-\frac{2}{mN}\Sigma T'\right]^{-1}=\frac{m-1}{m}.$$

Таким образом, с указанной выше степенью точности можно полагать, что второй момент $W$ остается неизменным. Влияние связей на величину третьего и четвертого моментов при заданном порядке величин также пренебрежимо мало. Из этого следует, что приведенные выше выводы сохраняют свое значение.

в. Если числа $T'$ велики, то нам придется вычислить $\mu_2(W)$. В таком случае

$$\left.\begin{aligned}\frac{1}{2}\nu_1&=p=\frac{m-1}{m^3\mu_2(W)}-\frac{1}{m}\\ \frac{1}{2}\nu_2&=q=(m-1)\,p\end{aligned}\right\}. \qquad (7.25)$$

Следовательно, можно использовать описанную выше процедуру статистической проверки, подставив соответствующие значения $\nu_1$ и $\nu_2$.

7.8. Докажем теперь, что распределение величин

$$\chi_r^2=m(n-1)\,W=\frac{S}{\frac{1}{12}mn(n+1)} \qquad (7.26)$$

с ростом $m$ стремится к распределению $\chi^2$.

$$dF\propto e^{-\frac{1}{2}\chi^2}\chi^{\nu-1}\,d\chi \qquad (7.27)$$

при $\nu=n-1$.

Обратимся к суммам элементов любого столбца из (7.1). Рассмотрим, например, сумму элементов первого столбца, обозначив ее буквой $p$. Тогда

$$E(p)=E(a_1+b_1+\ldots+k_1)=0,$$

$$E(p^2)=E(\Sigma a)^2=\Sigma\alpha_2.$$

Порядок $E(p^{2r+1})$ равен $\Sigma\alpha_{2r+1}+\Sigma\alpha_{2r-1}\ \beta_2+$ и т. д., а порядок $E(p^{2r})$ равен $\Sigma\alpha_{2r}+\ldots+\Sigma\alpha_2\beta_2+\varkappa_2$. С помощью рассуждений, аналогичных приведенным в (5.21), можно показать, что выражение $E(p^{2r+1})$ имеет меньший порядок и что главный член в выражении $E(p^{2r})$ образует сумма $\Sigma\alpha_2\beta_2\ldots\varkappa_2$. Таким образом, в пределе нечетные моменты обращаются в нуль и

$$\frac{\mu_{2r}}{(\mu_2)^r}\sim\frac{(2r)!}{2^r}\,\frac{\Sigma\alpha_2\,\beta_2\ldots\varkappa_2}{(\Sigma\alpha_2)^r}. \tag{7.28}$$

Если все элементы $\alpha$ равны между собой или близки к этому, то

$$\frac{\mu_{2r}}{(\mu_2)^r}\sim\frac{(2r)!}{r!\,2^r}$$

и, следовательно, распределение величин $p$ стремится к нормальному распределению с нулевой средней.

Обозначим через $S$ сумму $n$ случайных величин $p$, подчиненных только одному ограничению: $\Sigma(p)=0$. Тогда значения $kS$ распределены как $\chi^2$ с $\nu=n-1$ степенями свободы, причем $k$ представляет собой коэффициент, определяемый из следующего условия: средняя величина $kS$ равна $n-1$, среднему значению $\chi^2$. Однако

$$E(kS)=kn\Sigma\alpha_2,$$

таким образом,

$$k=\frac{n-1}{n\Sigma\alpha_2}$$

и, следовательно, величины

$$\chi_r^2=\frac{(n-1)\,S}{n\Sigma\alpha_2}$$

распределены по закону $\chi^2$.

Если последовательности рангов не содержат связей, то дисперсия любой из них равна $\frac{1}{12}(n^2-1)$, откуда следует, что

$$\chi_r^2=\frac{S^1}{\frac{1}{12}nm(n+1)},$$

что совпадает с выражением (7.26).

Если имеют место связи, характеризующиеся соответствующими числами $T'$, то $\chi_r^2$ определяется из следующего выражения:

$$\chi_r^2 = \frac{(n-1)S}{n\left[\frac{1}{12}m(n^2-1) - \frac{1}{n}\sum T'\right]} = \frac{S}{\frac{1}{12}mn[n+1) - \frac{1}{n-1}\sum T'}. \quad (7.29)$$

Если связи не имеют широкого распространения, то второй член в знаменателе оказывает сравнительно малое влияние на величину $\chi_r^2$

**7.9.** Теперь изложим основное содержание методов, с помощью которых рассчитывается действительное распределение $W$ или — эквивалентная задача — распределение $S$ при условии, что значения $m$ и $n$ невелики.

Заметим, что при $m = 2$ можно воспользоваться $\rho$-распределением Спирмэна. Будем исходить из того, что нам даны $m$ и $n$, требуется рассмотреть переход к значениям $m + 1$, $n$. Пусть, например, при $m = 2$, $n = 3$ заданы следующие значения сумм рангов (измеренных относительно средней):

| Тип | Частота |
|---|---|
| —2 0 2 | 1 |
| —2 1 1 | 2 |
| —1 0 1 | 2 |
| 0 0 0 | 1 |

Здесь —2; 1; 1 и 2; —1; —1 можно полагать идентичными, поскольку величины $S$, рассчитанные для этих последовательностей, совпадают между собой; они будут играть одинаковую роль и после рассматриваемого ниже перехода к $m = 3$.

При $m = 3$ приведенные выше типы могут быть прибавлены к шести перестановкам из —1, 0, 1; например, на основе типа —2, 0,2 появятся следующие типы: —3, 0, 3; —3, 1, 2; —2, —1, 3; —2, 1, 1; —1, —1, 2 и —1, 0, 1. Эти типы будут появляться также при переходе каждого из остальных «базисных» ($m=2$) типов к $m=3$. В результате мы получили

| Тип | Частота |
|---|---|
| —3 0 3 | 1 |
| —3 1 2 | 6 |
| —2 0 2 | 6 |
| —2 1 1 | 6 |
| —1 0 1 | 15 |
| 0 0 0 | 2 |
| | 36 |

При $n$, равном пяти или превышающем 5, трудоемкость расчетов резко возрастает в связи с тем, что на каждом этапе приходится принимать во внимание все большее число различных типов. И все же в обычных задачах, связанных со статистической проверкой существенности, вероятно, можно полагать, что при достаточно больших значениях $n$ $z$-распределение служит достаточно точной аппроксимацией.

**7.10.** Теперь мы можем показать, что предложенный метод оценивания (как отмечалось в 6.12) максимизирует среднее значение $\rho$, коэффициента корреляции между оцениваемыми и наблюдаемыми последовательностями.

Предположим, что оцениваемая последовательность имеет вид $X_1, \ldots, X_n$, а суммарные ранги равны $S_1, \ldots, S_n$. Тогда среднее значение коэффициента $\rho$ можно получить из следующего выражения:

$$\frac{1}{m}\sum_{k=1}^{m}\rho_k=\frac{12}{m(n^3-n)}\sum_{k=1}^{m}\sum_{j=1}^{n}\left[X_j-\frac{1}{2}(n+1)\right]\left[x_{jk}-\frac{1}{2}(n+1)\right],$$

где $x_{jk}$ представляет собой ранг $k$-го объекта в $j$-й последовательности. Это выражение равносильно следующему:

$$\frac{12}{m(n^3-n)}\sum_{j=1}^{n}\left[X_j-\frac{1}{2}(n+1)\right]\left[S_j-\frac{1}{2}m(n+1)\right]=$$

$$=\frac{12}{m(n^3-n)}\left[\sum_{j=1}^{n}(X_jS_j)-\frac{1}{4}mn(n+1)^2\right]. \qquad (7.30)$$

Ясно, что это выражение достигает максимальной величины в том случае, когда наибольшее значение принимает $\Sigma\ (XS)$, т. е. когда наибольшее $S$ умножается на наибольшее $X$, затем эта процедура продолжается до тех пор, пока наименьшее $S$ не умножается на наибольшую величину $X$. Предполагавшееся выше правило оценивания, по существу, обеспечивает необходимый для максимизации порядок умножения, а отсюда непосредственно следует требуемый результат.

Рассмотрим теперь сумму

$$U=\sum_{j=1}^{n}(S_j-mX_j)^2=\Sigma S_j^2+m^2\Sigma X^2-2m\Sigma(XS).$$

Поскольку первые два члена в правой части представляют собой постоянные величины, $U$ достигает минимума в том случае, когда $\Sigma\ (XS)$ принимает максимальное значение. Поэтому наш метод оценивания основан на минимизации $U$ или, другими словами, мы минимизируем сумму квадратов разностей между действительными значениями сумм $S$ и значениями $mX$, которые они принимают в том случае, когда все последовательности совпадают между собой.

**7.11.** Теперь выясним методы статистической проверки существенности в случае неполных последовательностей рангов, рассматривавшихся в 6.16. Придерживаясь прежних обозначений, будем исходить из того, что

$$U=\sum_{j}\sum_{i<p}x_{ij}x_{pj}. \qquad (7.31)$$

Тогда, поскольку

$$\sum x_{ij}^2 = \frac{mn(k^2-1)}{12},$$

можно определить $W$ с помощью следующего выражения:

$$W = \frac{12}{\lambda^2 n(n^2-1)}\left[\sum x_{ij}^2 + 2\sum_j \sum_{i<p} x_{ij} x_{pj}\right] = \frac{k+1}{\lambda(n+1)} + \frac{24U}{\lambda^2 n(n^2-1)}. \tag{7.32}$$

В приведенном выражении объект $j$ встречается дважды: он фигурирует в оценках экспертов $i$ и $p$, причем величина любого элемента $x$ из одной группы не зависит от элементов $x$, содержащихся в любой другой группе. Следовательно,

$$E(U) = 0 \tag{7.33}$$

и

$$U^2 = \Sigma x_{ij}^2 x_{pj}^2 + 2\Sigma x_{ij} x_{pj} x_{ql} x_{rl}. \tag{7.34}$$

Первая сумма в этом выражении содержит $\frac{1}{2} nm(m-1)$ слагаемых, математическое ожидание каждого из этих слагаемых $[E(x_{ij}^2)]^2$ равно $\frac{1}{144}(k^2-1)^2$. Во вторую сумму, фигурирующую в выражении (7.34), входят произведения таких элементов $x$, у которых $j$-й и $l$-й объекты совместно встречаются лишь в различных группах.

Далее,

$$E(x_{ij} x_{ql}) = E(x_{pj} x_{rl}) = -\frac{1}{12}(k+1)$$

в тех случаях, когда $j$-й и $l$-й объекты входят в одну и ту же группу; вместе с тем $E(x_{ij}x_{ql})$ равно нулю во всех остальных случаях. Общее число случаев, когда $j$-й и $l$-й объекты входят в одну группу, равно $\frac{1}{4} n(n-1)\lambda(\lambda-1)$. Подставив приведенные данные в выражение (7.34), можно получить

$$E(U^2) = \frac{mn(k+1)(k^2-1)}{288}[(m-1)(k-1)+\lambda-1]. \tag{7.35}$$

Таким образом,

$$E(W) = \frac{k+1}{\lambda(n+1)}, \tag{7.36}$$

$$\operatorname{var} W = \frac{2(k+1)^2}{mn\lambda^2(n+1)^2}\left(m-1+\frac{\lambda-1}{k-1}\right). \tag{7.37}$$

**7.12.** Следует отметить, что третий и четвертый моменты мы не можем вычислить тем же способом, который использовался при определении обычных коэффициентов. Ведь ранее можно было воспользоваться свойством симметрии, поскольку объекты или пары объектов появлялись одинаково часто. При переходе к наборам, содержащим три объекта, или к более сложным системам симметричность исчезает, однако по аналогии с величинами $W$ можно полагать (отчасти исходя из эвристических соображений), что в обычном случае рассматриваемое нами распределение совпадает с распределением (7.21), а первый и второй моменты нашей совокупности можно приравнять соответственно (7.36) и (7.37). В таком случае можно записать следующие соотношения:

$$\left.\begin{aligned} p &= \frac{mn\left(1-\dfrac{k+1}{\lambda(n+1)}\right)}{2\left(\dfrac{mn}{n-1}-\dfrac{k}{k-1}\right)} - \frac{k+1}{\lambda(n+1)} \\ q &= \left(\frac{\lambda(n+1)}{k+1}-1\right)p \end{aligned}\right\}. \tag{7.38}$$

Из этого непосредственно вытекает методика проверки, предлагавшаяся в 6.19.

7.13. Положим, $W = a\chi^2$ с тем, чтобы обеспечить наибольшее соответствие распределения величин $W$ $\chi^2$-распределению с $n$—1 степенями свободы. В таком случае

$$E(W) = \frac{k+1}{\lambda(n+1)} = a(n-1).$$

Из этого следует, что, исключив $a$, мы можем записать выражение для $\chi^2$:

$$\chi^2 = \frac{\lambda(n^2-1)}{k+1} W. \tag{7.39}$$

Отсюда видно, что распределение этих величин стремится к распределению $\chi^2$ с $n$ — 1 степенями свободы. Дисперсия нашего выражения $\chi^2$ равна:

$$2(n-1)\left[1-\frac{k(n-1)}{mn(k-1)}\right],$$

т. е. приблизительно равна:

$$2(n-1)\left(1-\frac{1}{m}\right). \tag{7.40}$$

При больших $m$ этой формуле будет соответствовать соотношение (7.37).

## Библиография

Здесь можно вновь сослаться на статьи, упоминавшиеся в библиографии к предшествующей главе, а также указать работы [78], [113], [50] и [51].

## *ГЛАВА 8.* ЧАСТНАЯ РАНГОВАЯ КОРРЕЛЯЦИЯ

**8.1.** Интерпретируя сопоставление наблюдаемых качественных характеристик двух объектов $A$ и $B$, мы часто сталкиваемся со следующим вопросом: действительно ли существует определенная связь или корреляция между характеристиками $A$ и $B$, либо же эта связь на самом деле объясняется зависимостью или корреляцией каждой из этих характеристик с третьей характеристикой $C$. Методы, с помощью которых обычно исследуются проблемы такого рода, в теоретической статистике носят название частной корреляции или анализа частных связей; с помощью методов частной корреляции пытаются оценить те области общей совокупности, внутри которых устранено влияние изменений в характеристике объекта $C$. Аналогичные проблемы возникают и при использовании методов ранговой корреляции. Если, например, у определенного числа людей можно наблюдать статистически существенную корреляцию между математическими и музыкальными способностями, возникает вопрос, не объясняется ли такая корреляция зависимостью этих способностей от некоторого более существенного качества, скажем, такого, как интеллигентность. Перейдем к рассмотрению метода, с помощью которого в теории ранговой корреляции можно исследовать проблемы такого рода.

**8.2.** Предположим, что нам пришлось столкнуться со следующими тремя последовательностями (каждая из них содержит по 6 чисел):

$$\begin{aligned} P &= 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \\ Q &= 3 \quad 1 \quad 4 \quad 2 \quad 6 \quad 5 \\ R &= 4 \quad 2 \quad 1 \quad 6 \quad 3 \quad 5 \end{aligned} \tag{8.1}$$

Всего можно образовать $\binom{6}{2}$ пар. Одну из этих последовательностей будем считать стандартной (поскольку выбор ее произволен, мы остановимся на той последовательности, элементы которой расположены в натуральном порядке). Выписав затем все возможные пары, будем отмечать знаком «+» пары рангов, расположенные в том же порядке, что и в стандартной последовательности, и знаком «—» пары, расположенные в обратном порядке.

| | (12) | (13) | (14) | (15) | (16) | (23) | (24) | (25) | (26) | (34) | (35) | (36) | (45) | (46) | (56) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $P$ | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + |
| $Q$ | — | + | — | + | + | + | + | + | + | — | + | + | + | + | — |
| $R$ | — | — | + | — | + | — | + | + | + | + | + | + | — | — | + |

Теперь вычислим соответствующие коэффициенты корреляции:

$$\tau_{PQ}=\frac{7}{15},\quad \tau_{PR}=\frac{3}{15},\quad \tau_{QR}=-\frac{1}{15}.$$

Далее в таблицу размером $3\times3$ выпишем показатели, характеризующие степень согласованности рангов $Q$ и $R$ с рангами $P$.

Ранги $R$

| Ранги $Q$ | пары с «+» (согласованность с $P$) | пары с «—» (несогласованность с $P$) | общее количество |
|---|---|---|---|
| пары с «+» (согласованность с $P$) | 6 | 5 | 11 |
| пары с «—» (несогласованность с $P$) | 3 | 1 | 4 |
| общее количество | 9 | 6 | 15 |

(8.2)

Таким образом, в 11 случаях ранги $Q$ согласованы с $P$, причем в 6 случаях и ранги $R$ также согласуются с $P$, а в остальных пяти случаях ранги $R$ оказываются несогласованными.

Вообще если имеются три последовательности рангов, каждая из которых содержит $n$ чисел, то наша таблица будет иметь следующий вид:

$$\begin{array}{c|c|c} a & b & a+b \\ \hline c & d & c+d \\ \hline a+c & b+d & N=\binom{n}{2}=a+b+c+d \end{array} \tag{8.3}$$

Введем теперь следующее определение коэффициента частной ранговой корреляции $Q$ и $R$ с $P$:

$$\tau_{QR\cdot P}=\frac{ad-bc}{\left|\sqrt{[(a+b)\,(c+d)\,(a+c)\,(b+d)]}\right|}. \tag{8.4}$$

В нашем примере его численное значение равно:

$$\frac{6-15}{\sqrt{(11\times4\times9\times6)}}=-0{,}185,$$

что сравнимо с $\tau_{QR}=-0{,}067$.

8.3. Выражение (8.4) представляет собой коэффициент связи в таблице $2\times 2$; мы уже использовали его в другом контексте в 3.14. Значения коэффициента меняются от $-1$ до $+1$, но не могут выходить за эти пределы, с его помощью измеряется интенсивность связей между двумя величинами: согласованностью рангов $Q$ с $P$ и согласованностью рангов $R$ с $P$.

Если указанный коэффициент равен единице, то

$$(ad-bc)^2=(a+b)(a+c)(b+d)(c+d);$$

отсюда следует

$$4abcd+a^2(bc+bd+cd)+b^2(ac+ad+cd)+$$
$$+c^2(ab+ad+bd)+d^2(ac+ab+bc)=0.$$

Так как ни одна из величин $a$, $b$, $c$, $d$ не может принимать отрицательных значений, это равенство оказывается справедливым только в том случае, когда по меньшей мере две из них равны нулю. Если эти две величины расположены в одной строке или в одном столбце, то перед нами совершенно тривиальная ситуация, когда предпочтения, выраженные в одной из последовательностей ранговых оценок, — $Q$ или $R$ — полностью согласованы или, напротив, полностью противоположны (рассогласованы) с предпочтениями $P$. Нам остается теперь рассмотреть только те случаи, когда нулю равны величины $a$ и $d$ или величины $b$ и $c$. В последнем случае последовательности рангов $Q$ и $R$ полностью согласуются с последовательностью $P$ и $\tau_{QR\cdot P}=1$. А в том случае, когда $Q=0$ и $d=0$, эти предпочтения полностью противоположны (рассогласованы), и соответствующий коэффициент равен $-1$.

8.4. Читатель, который знаком с величиной, называемой $\chi^2$, может без труда убедиться в том, что

$$\tau_{QR.\,P}=\sqrt{\frac{\lambda^2}{N}}. \tag{8.5}$$

Мы используем $\chi^2$, а следовательно, и коэффициент $\tau$ для того, чтобы установить, насколько велики различия между рассматриваемым случаем и ситуацией, когда дихотомия предпочтений эксперта $Q$ полностью независима от дихотомии предпочтений эксперта $R$.

Действительно, допустим, что они независимы. Тогда наша таблица будет содержать следующие частоты:

$$\frac{(a+b)(a+c)}{N}; \qquad \frac{(a+b)(b+d)}{N};$$
$$\frac{(a+c)(c+d)}{N}; \qquad \frac{(c+d)(b+d)}{N}.$$

В таком случае можно подсчитать разности между наблюдаемыми значениями и значениями, соответствующими случаю «полной независимости». Ограничимся здесь одним примером:

$$\frac{(a+b)(a+c)}{N}-a=\frac{(a+b)(a+c)-a(a+b+c+d)}{N}=\frac{bc-ad}{N}.$$

Таким образом, $\chi^2$ будет представлять собой сумму четырех слагаемых вида

$$\frac{(bc-ad)^2}{N^2} \Big/ \frac{(a+b)(a+c)}{N}$$

и эту сумму можно представить следующим образом:

$$\frac{(bc-ad)^2 (a+b+c+d)}{(a+b)(a+c)(c+d)(b+d)},$$

а отсюда следует соотношение (8.5).

8.5. Итак, мы построили коэффициент, меняющий значения от —1 до +1, с помощью которого можно измерять степень согласованности последовательностей $Q$ и $R$ в совпадении порядка их рангов с порядком рангов в $P$. Если коэффициент равен +1, это означает, что указанные ранги полностью согласуются между собой; если он равен нулю, $ad - bc = 0$ и $a/b = c/d$, так что предпочтения оказываются совершенно независимыми; если же коэффициент равен —1, предпочтения совершенно противоположны (рассогласованы).

Но в таком случае можно утверждать, что определяемый таким образом коэффициент частной корреляции рангов $\tau$ измеряет степень согласованности между рангами $Q$ и $R$ независимо от влияния последовательности ранговых оценок $P$. Коэффициент частной корреляции рангов $\tau$ растет при увеличении согласованности между рангами $Q$ и $R$ независимо от того, согласуются они с рангами $P$ или не согласуются. Это станет яснее, если вновь обратиться к данным, приведенным в (8.3). Считая обычный коэффициент корреляции рангов между последовательностями $Q$ и $R$, мы бы получили:

$$\tau_{QR} = \frac{(a+d)-(b+c)}{N}. \tag{8.6}$$

Однако в (8.3) мы выделяли случаи согласованности между последовательностями $Q$ и $R$, исходя из следующего критерия: согласуются ли эти ранги с рангами, содержащимися в последовательности $P$. Строка, в которой записаны числа $a$ и $b$, показывает, сколько плюсов или минусов будет иметь последовательность $R$ в тех случаях, когда последовательность $Q$ содержит только плюсы. Если эта строка аналогична строке, содержащей $c$ и $d$ (другими словами, если элементы этих строк пропорциональны), в последовательности $Q$ плюсы и минусы будут распределяться примерно в той же пропорции независимо от того, имеет ли в этих случаях последовательность $R$ знак плюс или не имеет. Но в подобной ситуации мы уже не можем полагать, что последовательности $Q$ и $R$ обнаруживают такое взаимное соответствие, которое не удастся полностью объяснить согласованностью каждой из этих последовательностей с последовательностью $P$. Коэффициент частной корреляции $\tau$ показывает, насколько рассматриваемая ситуация отличается от описанной выше, он показывает, в частности, как велики различия в количестве знаков плюс и минус в последовательности $Q$ (проставляемых вне зависимости от того, имеет ли последовательность в каждом из этих случаев знак плюс). Другими словами, коэффициент частной корреляции рангов является лучшим показателем того, насколько последовательности $Q$ и $R$ согласуются друг с другом, вне зависимости от согласованности каждой из них с последовательностью $P$.

**8.6.** В дополнение к (8.6) можно привести следующие соотношения:

$$\tau_{PQ}=\frac{(a+b)-(c+d)}{N}\,; \tag{8.7}$$

$$\tau_{PR}=\frac{(a+c)-(b+d)}{N}\,. \tag{8.8}$$

Как отмечалось выше, $N = a + b + c + d$ и, следовательно,

$$1-\tau_{PQ}^2=\frac{4}{N^2}(a+b)(c+d),$$

$$1-\tau_{PR}^2=\frac{4}{N^2}(a+c)(b+d),$$

$$\tau_{QR}-\tau_{PQ}\tau_{PR}=\frac{1}{N^2}\left[(a+b+c+d)\left\{\overline{a+d}-\overline{b+c}\right\}-\right.$$

$$\left.-\left\{\overline{a+b}-\overline{c+d}\right\}\left\{\overline{a+c}-\overline{b+d}\right\}\right]=\frac{4}{N^2}(ad-bc).$$

Таким образом, на основании (8.4) можно заключить:

$$\tau_{QR\cdot P}=\frac{\tau_{QR}-\tau_{PR}\,\tau_{QP}}{\sqrt{[(1-\tau_{PR}^2)(1-\tau_{QP}^2)]}}\,. \tag{8.9}$$

Эта формула позволяет выразить коэффициент частной корреляции $\tau$ через коэффициенты корреляции рангов $\tau$, полученные в результате сопоставления исходных последовательностей. Любопытно, что это соотношение полностью соответствует формуле коэффициента частной корреляции, выраженного через коэффициенты корреляции низших порядков [1] (однако, здесь мы имеем дело, вероятно, просто с совпадением соответствующих формул).

*Пример* 8.1

Рассмотрим оценки качеств определенной группы людей, упорядоченные по трем различным признакам: по общему уровню развития интеллекта (1), по способностям, проявляемым при изучении математических дисциплин (2), и по музыкальным способностям (3).

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| (1) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| (2) | 1 | 4 | 5 | 6 | 2 | 7 | 3 | 9 | 8 | 10 |
| (3) | 4 | 1 | 3 | 5 | 2 | 6 | 7 | 10 | 9 | 8 |

Определим значения исходных коэффициентов корреляции рангов:

$$\tau_{12}=0{,}644;\ \tau_{13}=0{,}644;\ \tau_{23}=0{,}556.$$

[1] См., например, Юл Д. Э. и Кендэл М. Дж. Теория статистики. М., Госстатиздат, 1960, 12.15.

Таким образом, в силу (8.9) можно получить:

$$\tau_{23,1}=\frac{0,556-(0,644)^2}{1-(0,644)^2}=0,24.$$

Частная корреляция рангов оказывается слабее, чем рассчитанная выше корреляция между ранговыми оценками (2) и (3); можно предположить, что корреляция, существующая между оценками (1) и (2), с одной стороны, и оценками (1) и (3) — с другой, как бы «накладывается» на характеристики действительного соотношения между оценками (2) и (3). Однако подобные выводы можно сделать лишь со значительными оговорками. Эти соображения, которые, разумеется, желательно иметь в виду в ходе дальнейшего исследования, но подобные соображения представляют собой не более чем догадки, если только мы не располагаем более основательными аргументами, позволяющими предсказать наличие подобного эффекта.

8.7. Методы проверки статистической существенности коэффициентов частной корреляции $\tau$ до настоящего времени не разработаны. Непригодной здесь оказывается проверка по критерию $\chi^2$, поскольку между некоторыми рангами, входящими в величины $a$, $b$, $c$, $d$, существует взаимная зависимость, например, если ранговая оценка $A$ меньше ранга $B$, а ранг $B$ меньше ранга $C$, то ранг $A$ должен быть меньше ранга $C$. Здесь мы вновь сталкиваемся с тем разделом статистической теории, который требует дальнейших исследований.

## Библиография

См. [49]. В [40] приводится довольно сложное выражение, характеризующее дисперсию коэффициента частной корреляции, причем рассматривается предельный случай, когда значение $n$ велико. Если $\tau_{13} = \tau_{23}=0$, распределение $\sqrt{n\ (t_{12.3}-\tau_{12.3})}$ в пределе будет совпадать с распределением $\sqrt{n\ (t_{12}-\tau_{12})}$.

В [70] также рассмотрены проблемы исчисления коэффициента частной корреляции $\tau$; не получив каких-либо четких результатов, автор ограничился выводом о том, что присутствующие в этой задаче проблемы распределения случайных величин чрезвычайно сложны.

Кроме того, в [70] рассмотрена также методика подсчета коэффициента множественной корреляции для случая трех случайных величин по той же формуле, что и в обычной статистической теории:

$$1-R^2_{1(23)}=(1-\tau^2_{13})(1-\tau^2_{12.3}).$$

Но при малых $n$ мы вновь сталкиваемся с трудной проблемой распределения случайных величин. В [70] предложена методика статистической проверки, в которой используется отношение соответствующих дисперсий.

См. также [43].

## *ГЛАВА 9.* РАНГИ И ЗНАЧЕНИЯ ПРИЗНАКА

**9.1.** До сих пор мы рассматривали ранги как основные данные об имеющейся статистической ситуации, не связывая их со способами, которыми они были получены. Во многих таких ситуациях, однако, ранжирование производится (или можно предположить, что производится) в соответствии с величиной статистической переменной или *признака*.

Безусловно, интересно рассмотреть взаимосвязь между рангами и значениями соответствующих переменных или между измерителями корреляции, основанными на рангах, и измерителями, базой которых являются значения признака.

### Конкордации

**9.2.** Нам будет нужно в общем рассмотреть непрерывную совокупность, иначе говоря, совокупность, значения признака которой могут находиться в любых точках непрерывного диапазона. Обратим внимание на одну деталь. Строго говоря, такая совокупность не может содержать ранговую корреляцию, поскольку существо ранжирования заключается в том, что объекты должны быть перечисляемыми, а совокупность значений непрерывного признака не может быть в этом смысле упорядочена.

**9.3.** Тем не менее мы можем соединить идеи ранговой корреляции и корреляции переменных при рассмотрении *свойств упорядочения*. Допустим, что из непрерывной совокупности в случайном порядке отобраны два члена: $x_i$ и $x_j$. Вероятность того, что они совпадают, равна нулю и, таким образом, допустимо не учитывать возможность $x_i = x_j$. Рассмотрим теперь вероятность того, что $x_i < x_j$ и вероятность противоположного события, т. е. того, что $x_i > x_j$. Кроме того, если мы извлекли два члена $x_i$, $y_i$ и $x_i$, $y_j$ из двумерной совокупности, то мы можем проанализировать вероятности *конкордации* типа 1:

$$\pi_1 = \text{Prob}\,(y_i < y_i \mid x_i < x_j), \tag{9.1}$$

иначе говоря, вероятность того, что $y_i < y_j$, если $x_i < x_j$. Вероятность противоположного события составит:

$$\chi_1 = 1 - \pi_1 = \text{Prob}\,(y_i > y_j \mid x_i < x_j). \tag{9.2}$$

Число $\pi_1$ характеризует собой свойство совокупности.

**9.4.** Теперь предположим, что выборка, состоящая из $n$ величин, случайным образом извлечена из совокупности и произведено упорядочение $x$-ов в порядке возрастания их значений. Из $\frac{1}{2}n(n-1)$ пар $x$-ов, которые могли быть отобраны для сравнений, некоторым сопутствовали $y$-ки, расположенные в порядке возрастания. Очевидно, что число пар, у которых наблюдалось это соответствие, деленное на $\frac{1}{2}n(n-1)$, является оценкой $\pi_1$. Кроме того, как будет доказано в следующей главе, данная оценка является несмещенной. Если $p_1$ равно этой доле, а $q_1 = 1 - p_1$, то сразу видно, что коэффициент $\tau$ для данной выборки (обозначим его через $t$) определяется как

$$t_1 = p_1 - q_1 = 2p_1 - 1. \tag{9.3}$$

Теперь мы можем определить $t_1$, как это уже было отмечено в 2.14, в терминах согласованности (конкордации) и перейти к коэффициенту, который имеет аналог в случае непрерывного распределения.

**9.5.** Предположим теперь, что у нас имеется триада величин $(x_i, y_i)$, $(x_j, y_j)$ и $(x_k, y_k)$, $i \neq j \neq k$. Рассмотрим вероятность конкордации типа 2:

$$\pi_2 = \mathrm{Prob}\,(y_i < y_k \mid x_i < x_j), \tag{9.4}$$

иначе говоря, вероятность того, что $y_i$ меньше $y_k$ при условии, что $x_i < x_j$. Определим выборочную величину $p_2$ в виде числа конкордаций типа 2, имеющихся в выборке, деленного на общее возможное их число. В отличие от $p_1$, которое могло изменяться от 0 до 1, $p_2$ (как это будет показано в следующей главе) может варьировать только в пределах от $\frac{1}{3}$ до $\frac{2}{3}$. В связи с некоторыми обстоятельствами, речь о которых будет идти ниже, мы не будем использовать в качестве другого коэффициента $6\left(p_2 - \frac{1}{2}\right)$ (такой коэффициент менял бы свои значения от $-1$ до $+1$). Вместо этого, следуя (2.35), определим выборочную величину

$$r_s = \frac{3t}{n+1} + \frac{6(n-2)}{n+1}\left(p_2 - \frac{1}{2}\right). \tag{9.5}$$

Величина $r_s$ представляет собой коэффициент Спирмэна для выборки. Снова обнаруживаем, что коэффициент ранговой корреляции можно определять через конкордации. Заметим, что для больших $n$ выражение (9.5) становится равным:

$$\rho_s = 6\left(\pi_2 - \frac{1}{2}\right). \tag{9.6}$$

Данную величину можно интерпретировать как определение коэффициента Спирмэна для непрерывных совокупностей.

## Отношение между рангами и значениями переменной

**9.6.** Предположим, что мы извлекли выборку из совокупности скалярных значений $x$-ов и ранжировали ее в порядке возрастания. Представляет некоторый интерес рассмотрение коэффициента корреляции между значениями $x$-ов и рангами. Действительно, этот коэффициент иногда бывает удивительно высоким. А. Стюартом получены следующие основные результаты:

а) пусть коэффициент корреляции для совокупности из $n$ единиц составляет $C_n$, а предельное его значение при $n \to \infty$ равно $C$, тогда во всех случаях

$$C_n = \left(\frac{n-1}{n+1}\right)^{\frac{1}{2}} C; \tag{9.7}$$

б) если генеральная совокупность имеет равномерное распределение, иначе говоря, значения оказываются одинаковыми в некотором конечном диапазоне, то $C = 1$ и

$$C_n = \left(\frac{n-1}{n+1}\right)^{\frac{1}{2}}; \tag{9.8}$$

в) если генеральная совокупность является нормальной и $C = \sqrt{(3/\pi)}$, то

$$C_n = 0{,}9772\left(\frac{n-1}{n+1}\right)^{\frac{1}{2}}; \tag{9.9}$$

г) если генеральная совокупность следует так называемому гамма-распределению, то

$$dF = \frac{1}{\Gamma(m)} e^{-x} x^{m-1}\, dx, \quad 0 \leqslant x \leqslant \infty,$$

следовательно,

$$C_n = \left(\frac{n-1}{n+1}\right)^{\frac{1}{2}} \left(\frac{3m}{\pi}\right)^{\frac{1}{2}} \frac{\Gamma\left(m+\frac{1}{2}\right)}{\Gamma(m+1)}. \tag{9.10}$$

Например, при ранжировании 10 единиц, взятых из нормальной совокупности, коэффициент корреляции между рангами и значениями переменной составит:

$$0{,}9772 \times \sqrt{\frac{9}{11}} = 0{,}884.$$

**9.7.** Благодаря такой довольно тесной взаимосвязи между рангами и значениями переменной следует ожидать, что если бы мы заменили бы значения переменной числовыми рангами и затем оперировали с ними так, как если бы они были исходными переменными, то мы пришли бы во множестве случаев к тем же самым выводам. В большинстве практических случаев именно так и бывает, однако следовать этой процедуре нужно с известной долей осторожности. Переход к рангам означает эффективное нормирование переменной и фиксирование средней величины; иногда эта процедура может вывести на ложный путь.

9.8. Некоторые авторы рекомендуют обратную процедуру. Соответственно ей в последовательности рангов, извлеченных из нормальной совокупности, ранги общим числом $n$ единиц заменяются значениями переменной $x$. Величина $x_i$ есть средняя величина $i$-го члена в выборке объемом $n$. Это, разумеется, не устраняет затруднения, которое было рассмотрено в 9.7. Однако, как было показано в [41], когда проверка связана с гипотезами о нормальном распределении, такая процедура обладает свойством оптимальности[1].

### Отношение между $t$ и генеральным коэффициентом корреляции в случае нормального распределения

9.9. Теперь для того чтобы избежать путаницы, нам необходимо немного модифицировать и расширить систему обозначений:

1) будем, как и прежде, обозначать коэффициент $\tau$ в выборке через $t$;

2) обозначим выборочный коэффициент Спирмэна $\rho$ как $r_s$, а соответствующий параметр генеральной совокупности (9.6) — как $\rho_s$;

3) параметр двумерной нормальной совокупности, характеризующий коэффициент корреляции, обозначим через $\rho$, а выборочный коэффициент корреляции — через $r$;

4) иногда оценки $\rho$ будут рассчитываться на основе значений $t$ или $r_s$, их мы будем отмечать штрихом над $r$, т. е. $r'$.

9.10. Можно показать, что для выборок из нормальной совокупности

$$E(t)=\frac{2}{\pi}\sin^{-1}\rho. \tag{9.11}$$

Например, если ,

$$\rho=1/\sqrt{2}=0{,}707, \text{ то } E(t)=0{,}5.$$

Поэтому мы можем сконструировать оценку для $\rho$, скажем, $r'$, приняв

$$r'=\sin\frac{1}{2}\pi t. \tag{9.12}$$

Полученная величина не является несмещенной оценкой $\rho$, поскольку для этого необходимо, чтобы $E(r')=\rho$. Однако это не следует из (9.11). Тем не менее данная процедура представляется приемлемой. Выраже-

[1] Замена количественных признаков рангами эквивалентна аппроксимированию функции выборочного распределения на диаграмме прямой линией, а их замена эквивалентными нормальными отклонениями означает аппроксимирование этой функции кривой нормального распределения.

ние (9.11) впервые было предложено в [33], а выражение (9.13) для ее дисперсии — в [24].

**9.11.** В следующей главе мы покажем, что для выборки, следующей нормальному распределению:

$$\operatorname{var} t=\frac{2}{n\,(n-1)}\left\{1-\left(\frac{2}{\pi}\sin^{-1}\rho\right)^2+2\,(n-1)\left[\frac{1}{9}-\left(\frac{2}{\pi}\sin^{-1}\frac{1}{2}\,\rho\right)^2\right]\right\}. \tag{9.13}$$

Нам известно, что у больших выборок $t$ распределено нормально и, следовательно, этот результат можно использовать для проверки существенности наблюденных $t$, соотнося с интегралом нормального распределения. Однако, выполняя эту операцию, нам следует придать некоторое значение неизвестной величине $\rho$. В соответствии с обычной практикой, принятой в теории больших выборок, мы должны заменить $\rho$ на $r'$ уравнения (9.12).

Если $p$ и $q$ характеризуют доли положительных и отрицательных вкладов в $t$, то имеем:

$$t = p - q$$

и

$$1 - t^2 = 4pq. \tag{9.14}$$

Таким образом, на основе (9.13) получим для больших выборок

$$\operatorname{var} t=\frac{8}{n\,(n-1)}\left\{pq+\frac{1}{2}\,(n-2)\left[\frac{1}{9}-\left(\frac{2}{\pi}\sin^{-1}\frac{1}{2}\,r'\right)^2\right]\right\}. \tag{9.15}$$

Можно также показать, что

$$0\leqslant\frac{1}{9}-\left(\frac{2}{\pi}\sin^{-1}\frac{1}{2}\,r'\right)^2\leqslant\frac{4}{9}\,pq \tag{9.16}$$

и для $n \geqslant 10$

$$\frac{2}{n\,(n-1)}\left[1+\frac{2}{9}\,(n-2)\right]\leqslant\frac{5}{9\,(n-1)}\,. \tag{9.17}$$

Подставляя в (9.15), находим:

$$\operatorname{var} t\leqslant\frac{20\,pq}{9\,(n-1)}=\frac{5\,(1-t^2)}{9\,(n-1)}\,. \tag{9.18}$$

Верхний предел действительно достигается в целом ряде случаев, так что неравенство (9.18) дает нам достаточно хорошую оценку фактической дисперсии.

**9.12.** С той степенью приближения, которую мы здесь приняли, мы можем сравнить (9.18) с (4.9); последнее запишем как

$$\operatorname{var} t\leqslant\frac{2\,(1-t^2)}{n}\,, \tag{9.19}$$

Если не принимать во внимание разницу в сомножителях знаменателей, $n$ и $n-1$, которая не играет роли для больших выборок, то при сравнении выражения (9.18) с (9.19) видно, что первое определяет предел, составляющий всего 0,278 от предела, рассчитываемого в соответствии с последним. Соответственно стандартная ошибка в первом случае будет равна 0,53 или чуть больше половины стандартной ошибки во втором случае. Этот выигрыш в точности достигается благодаря допущению, что совокупность имеет нормальное распределение.

**9.13.** Поскольку

$$r' = \sin \frac{1}{2} \pi t,$$

для небольших изменений имеем:

$$\delta r' = \frac{1}{2} \pi \cos \left( \frac{1}{2} \pi t \right) \delta t.$$

Возведя в квадрат и суммируя все такие изменения, получим:

$$\operatorname{var} r' = \frac{1}{4} \pi^2 (1 - r'^2) \operatorname{var} t. \tag{9.20}$$

Воспользуемся выражением (9.18). Тогда

$$\operatorname{var} r' \leqslant \frac{5\pi^2}{9} pq \frac{1-r'^2}{n-1} \leqslant (2,34)^2 pq \frac{1-r'^2}{n-1}, \quad n \geqslant 10. \tag{9.21}$$

Если мы взяли $r'$ в качестве оценки $\rho$, то (9.16) дает нам оценку верхнего предела стандартной ошибки этой оценки. Интересно сравнить эту ошибку со стандартной ошибкой выборочного коэффициента $r'$, которая определяется формулой

$$\operatorname{var} r = \frac{(1-r^2)^2}{n}. \tag{9.22}$$

Взяв верхний предел по (9.21), получим, пренебрегая разницей между $n$ и $n-1$,

$$\sqrt{\frac{\operatorname{var} r'}{\operatorname{var} r}} = \frac{2,34 \sqrt{(pq)}}{\sqrt{(1-r^2)}} \tag{9.23}$$

Если генеральное $\rho$ равно нулю, то $p$ примерно равно $q$, так что отношение стандартных ошибок, задаваемое (9.23), примерно равно 1,17. Если $\rho = 0,9$, то приближенно можно положить, что

$$r = 0,9; \quad t = \frac{2}{\pi} \sin^{-1} 0,9 = 0,713; \quad pq = \frac{1}{4}(1 - t^2) = 0,123.$$

Отношение стандартных ошибок теперь равно примерно 1,88. Коэффициент корреляции $r$ более точен, чем $r'$, в том смысле, что он имеет меньшую стандартную ошибку и поэтому более вероятно, что он ближе к истинной величине $\rho$.

**Отношение между $\rho_s$ и $\rho$ в случае нормального распределения**

**9.14.** Если мы определяем $\tau$ для непрерывной совокупности с помощью выражения, аналогичного (9.3), а именно

$$\tau = 2\pi_1 - 1, \tag{9.24}$$

то $t$ является несмещенной оценкой $\tau$. Процедура, подобная той, которая вытекает из (9.5), дает нам

$$E(r_s) = \frac{3\tau}{n+1} + \frac{6(n-2)}{n+1}\left(\pi_2 - \frac{1}{2}\right),$$

а если $\rho_s$ определяется для совокупности с помощью (9.6), то это дает нам

$$E(r_s) = \frac{n-2)}{n+1}\rho_s + \frac{3\tau}{n+1} = \rho_s + \frac{3(\tau-\rho_s)}{n+1}. \tag{9.25}$$

Так подтверждается результат (5.76), согласно которому $r_s$ не является несмещенной оценкой $\rho_s$.

**9.15.** Рассмотрим теперь взаимосвязь между $\rho_s$ и $\rho$ в нормальной совокупности. Можно показать, что

$$\rho = 2\sin\frac{1}{6}\pi\rho_s.$$

и, следовательно, в качестве оценки $\rho$ для больших выборок мы можем взять величину $r''$

$$r'' = 2\sin\frac{1}{6}\pi r_s. \tag{9.26}$$

Однако в связи с тем, что имеется смещение, обнаруживаем с помощью (9.25), по-видимому, лучше взять

$$r'' = 2\sin\frac{1}{6}\pi\left[r_s - \frac{3(t-r_s)}{n-2}\right]. \tag{9.27}$$

Формула (9.26) принадлежит К. Пирсону [75]. Она была выведена им при рассмотрении корреляций степеней интенсивности следующим образом.

**9.16.** Определим *степень интенсивности* (grade), связанную с конкретным значением переменной $x$, как долю совокупности значений переменных меньших $x$ или равных $x$. Коэффициент корреляции между степенями интенсивности переменных $x$ и $y$ в двумерной взаимосвязи называется *коэффициентом корреляции степеней интенсивности.* Эта величина существует для непрерывной совокупности. Нетрудно убедиться, что она приводит к коэффициенту ранговой корреляции Спирмэна, когда применяется для конечной выборки. В следующей главе мы докажем, рассматривая коэффициенты конкордации, что коэффициент корреляции степеней интенсивности, определяемый таким путем, по существу является величиной $\rho_s$, найденной с помощью (9.6).

**9.17.** Можно предположить, что имеется некоторая разумным образом интерпретируемая формула, подобная (9.13), дающая дисперсию $r_s$ в случае нормального распределения. Однако в действительности это не так. Доказано, что такой формулы, выраженной с помощью элементарных функций, не существует.

Для больших выборок на основе (9.26) имеем:

$$\delta r'' = \frac{1}{3}\pi \cos\frac{1}{6}\pi r_s$$

и, следовательно,

$$\operatorname{var} r'' = \frac{1}{9}\pi^2\left(1 - \frac{1}{4}r^2\right)\operatorname{var} r_s. \tag{9.28}$$

Эта формула поможет нам тогда, когда мы знаем $\operatorname{var} r_s$. В случае, когда $\rho = 0$, она сокращается до $1/(n-1)$ и, таким образом,

$$\operatorname{var} r'' = \frac{\pi^2}{9(n-1)} = \frac{1,047}{n-1}.. \tag{9.29}$$

Когда $\rho$ не равно нулю, необходимое выражение (имеющее практический смысл при больших $n$) может быть выведено в виде бесконечного ряда следующим образом:

$$\operatorname{var} r_s = \frac{1}{n}(1 - 1,5635\rho^2 + 0,3047\rho^4 + 0,1553\rho^6 + $$
$$+ 0,0616\rho^8 + 0,0242\rho^{10} + \ldots). \tag{9.30}$$

Уточненные варианты такой формулы см. в [27], [15] и [28]. В [28] вычислены некоторые таблицы и проверена формула с помощью выборочных экспериментов. Авторы также рассмотрели результаты нормализации $z_s = \operatorname{tg}^{-1} r_s$ и $z_t = \operatorname{tg}^{-1} t$ и пришли к выводу, что вполне достаточно испытать эти преобразованные величины в случае нормального распределения с помощью

$$\operatorname{var} z_s = \frac{1,060}{n-3}, \quad \operatorname{var} z_t = \frac{0,437}{n-4}.$$

**9.18.** Если мы сопоставим стандартную ошибку $r''$ с ошибкой, определяемой (9.22), то находим при $\rho = 0$, что это отношение равно 1,047, а при $\rho = 0,5$ оно составит 1,137. Чаша весов склоняется в пользу коэффициента $r$ (как более точного), однако сказанное нельзя воспринимать очень строго. Оценка $\rho$, основанная на $r_s$, также достаточно хороша, если учитывать все обстоятельства. Это вытекает из замечания, приведенного в 9.7. Мы не можем вдаваться здесь в более подробные объяснения, но не случайно дисперсия $r''$ в (9.29) имеет множитель $\pi^2/9$, тогда как коэффициент $C$ в (9.9) равен обратной величине корня четвертой степени из этой величины.

**9.19.** Интересно рассмотреть взаимосвязь между $t$ и $r_s$ в случае нормального распределения. Опять у нас нет точного выражения, однако для больших выборок разложение, подобное тому, что было приведено в (9.30), дает нам

$$\operatorname{cov}(r_s, t) = \frac{2}{3n}(1 - 1{,}2486\rho^2 + 0{,}0683\rho^4 + 0{,}0728\rho^6 + $$
$$+ 0{,}0403\rho^8 + 0{,}0164\rho^{10} + \ldots). \qquad (9.31)$$

Находим, например, что для $\rho = 0$ коэффициент корреляции между $r_s$ и $t$ равен 1, как это уже известно из 5.14; для $\rho = 0{,}2$ он равен 0,99955; для $\rho = 0{,}4$ составляет 0,9981; хотя он и стремится к нулю при приближении $\rho$ к единице, даже для $\rho = 0{,}8$ он равен 0,9843. Для больших $n$ отношение $r_s$ к $t$ приближается к отношению их математических ожиданий, несмотря на то, что корреляция между ними может быть мала для больших $\rho$, поскольку их дисперсии приближаются к нулю. Это отношение, равное $3\sin^{-1}\frac{1}{2}\rho/\sin^{-1}\rho$, изменяется от 1,3 в окрестности $\rho = 0$ до 1,42, когда $\rho = 0{,}6$, и до единицы, когда $\rho = 1$. Это и является одной из причин, способствующих тому, что на практике значение $r_s$, часто на 40 или 50% превосходит $t$, если только каждая из этих величин не близка к единице.

9.20. Может сложиться неправильное понимание оценок $\rho$, основанных на $t$ и $r_s$, связанное с относительными значениями стандартных ошибок $r'$ и $r''$. Тот факт, что эти оценки различаются, указывает на действительную разницу в эффективности процессов оценивания. Процесс с меньшей дисперсией более эффективен.

Для больших выборок дисперсии $t$ и $r_s$, заданные (9.13) и (9.30), также различаются между собой. Однако это не означает, что $t$—лучшая или худшая оценка $\tau$ (по сравнению с тем, как $r_s$ оценивает значение $\rho_s$). Различие в дисперсии связано с различием в самих шкалах измерения. Легко проверить, что для $\rho$, которое не близко к единице:

$$\frac{E(t)}{\sqrt{\operatorname{var} t}} = \frac{E(r_s)}{\sqrt{\operatorname{var} r_s}}. \qquad (9.32)$$

Последнее находится в соответствии с нашим общим результатом, согласно которому корреляция между $t$ и $r_s$ является высокой по крайней мере для случаев, когда $n$ велико, а $\rho$ не близко к единице.

## Библиография

О средней и дисперсии $t$ см. [24] и [33]. О средней для $r_s$ см. [40], [66] и [67]; о дисперсии см. [53] и [16]. См. также библиографию к гл. 10. О корреляции степеней интенсивности см. [75].

О взаимосвязи между рангами и значениями признака см. [94]. Асимптотическое отношение для случая с нормальным распределением первоначально было получено (но не опубликовано) К. Бертом (C. Burt).

О конкордации первого и второго рода см. [97].

## *ГЛАВА 10.* ДОКАЗАТЕЛЬСТВО РЕЗУЛЬТАТОВ ГЛАВЫ 9

### Корреляция между рангами и значениями признака

**10.1.** Пусть $N$ выборок каждая объемом $n$ единиц отобраны из непрерывной совокупности со средней $\mu$ и дисперсией $\sigma^2$. В каждой выборке наблюдения ранжированы. Наименьшей $i$-й величине присваивается ранг $i$. Оценим теперь для совокупности, состоящей из $Nn$ наблюдений, ковариацию значений признака и рангов ($\mu_{11}$) и дисперсии значений признака ($\mu_{20}$) и рангов ($\mu_{02}$).

Если $i$-я наименьшая величина в $j$-й выборке есть $x_{(i)j}$, то имеем:

$$\mu_{02}=\frac{1}{12}(n^2-1), \tag{10.1}$$

$$\mu_{20}=\frac{1}{Nn}\sum_{i=1}^{n}\sum_{j=1}^{N}\left[x_{(i)j}-\frac{1}{Nn}\sum_{i=1}^{n}\sum_{j=1}^{N}x_{(i)j}\right]^2=$$

$$=\frac{1}{Nn}\sum_{i}\sum_{j}x_{(i)j}^2-\left[\frac{1}{Nn}\sum_{i}\sum_{j}x_{(i)j}\right]^2, \tag{10.2}$$

$$\mu_{11}=\frac{1}{Nn}\sum_{i}\sum_{j}\left[i-\frac{1}{2}(n+1)\right]\left[x_{(i)j}-\frac{1}{Nn}\sum_{i}\sum_{j}x_{(i)j}\right]^2. \tag{10.3}$$

Если $N$ стремится к бесконечности, то

$$\lim\frac{1}{N}\sum_{j}x_{(i)j}^r=E\left(x_{(i)}^r\right),$$

следовательно,

$$\lim_{N\to\infty}\mu_{20}=\frac{1}{n}\sum_{i}E\left(x_{(i)}^2\right)-\left[\frac{1}{n}\sum_{i}E\left(x_{(i)}\right)\right]^2, \tag{10.4}$$

$$\lim_{N\to\infty}\mu_{11}=\frac{1}{n}\sum_{i}E\left(x_{(i)}\right)-\left[\frac{n+1}{2n}\Sigma E\left(x_{(i)}\right)\right]^2. \tag{10.5}$$

Распределение $x_{(i)}$, скажем $G_{(x)}$, в выборках, состоящих из $n$ единиц, из совокупности с функцией распределения $F(x)$ задается[1] как

$$dG(x) = n\binom{n-1}{i-1}[F(x)]^{i-1}[1-F(x)]^{n-i}\,dF(x). \qquad (10.6)$$

Таким образом,

$$\frac{1}{n}\sum_i E\left(x_{(i)}^r\right) = \int_{-\infty}^{\infty}\left[\sum\binom{n-1}{i-1}F^{i-1}(1-F)^{n-i}\right]x^r\,dF(x). \qquad (10.7)$$

Однако сумма в квадратных скобках есть разложение бинома $[F + (1-F)]^{n-1}$ и, следовательно, равна единице. Отсюда

$$\left.\begin{aligned}\frac{1}{n}\sum_i E\left(x_{(i)}^r\right) &= \int_{-\infty}^{\infty} x^r\,dF\\ &= \mu \qquad\quad (r=1)\\ &= \sigma^2+\mu^2 \;(r=2)\end{aligned}\right\}. \qquad (10.8)$$

Предположим, что эти моменты существуют. Аналогично находим:

$$\frac{1}{n}\sum_i iE\left(x_{(i)}\right) = (n-1)\int_{-\infty}^{\infty} xF\,dF + \mu. \qquad (10.9)$$

Отсюда на основе (10.4) и (10.5) получим:

$$\lim_{N\to\infty}\mu_{20} = \sigma^2, \qquad (10.10)$$

$$\lim \mu_{11} = (n-1)\left[\int_{-\infty}^{\infty} xF(x)\,dF(x) - \frac{1}{2}\mu\right]. \qquad (10.11)$$

Таким образом, искомый коэффициент корреляции определяем как

$$C_n = \lim_{N\to\infty}\frac{\mu_{11}}{\sqrt{(\mu_{20}\mu_{02})}} = \left[\frac{12(n-1)}{\sigma^2(n+1)}\right]^{\frac{1}{2}}\left[\int_{-\infty}^{\infty} xF(x)\,dF - \frac{1}{2}\mu\right]. \qquad (10.12)$$

Если теперь предположить, что $n$ стремится к бесконечности, то получим:

$$C = \lim_{n\to\infty} C_n = \left(\frac{12}{\sigma^2}\right)^{\frac{1}{2}}\left[\int_{-\infty}^{\infty} xF(x)\,dF - \frac{1}{2}\mu\right], \qquad (10.13)$$

так, что

$$C_n = \left(\frac{n-1}{n+1}\right)^{\frac{1}{2}} C. \qquad (10.14)$$

---

[1] См. [58; 1; 14.2].

**10.2.** Рассмотрим теперь частные случаи.

Если выражение в квадратных скобках в (10.13) обозначить как $\frac{1}{4}\Delta$, то имеем:

$$\Delta = 4\int_{-\infty}^{\infty} x\left[F(x) - \frac{1}{2}\right]dF(x).$$

Интегрирование по частям дает

$$\Delta = 2\int_{-\infty}^{\infty} F(x)[1 - F(x)]\,dx = 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[|x - y|\,dF(y)]\,dF(x). \qquad (10.15)$$

В действительности величина $\Delta$ является коэффициентом рассеяния, известным под названием средняя разность Джини [58; 1; 2.21]. Это неотрицательная величина. Затем получим формулу Стюарта

$$C = \frac{\Delta\sqrt{3}}{2\sigma}. \qquad (10.16)$$

**10.3.** Для равномерного распределения

$$dF = \frac{dx}{k}, \quad 0 \leqslant x \leqslant k \qquad (10.17)$$

имеем

$$\sigma^2 = \frac{1}{12}k^2, \quad \Delta = \frac{1}{3}k, \qquad (10.18)$$

и, следовательно, $C = 1$.

Этого и следовало ожидать для распределения, в котором появление любой из случайных величин равновероятно.

**10.4.** Рассмотрим нормальное распределение. Без потери общности результатов можно предположить, что оно имеет дисперсию, равную единице, и нулевую среднюю. Тогда мы имеем:

$$dF = \frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2}dx, \quad -\infty \leqslant x \leqslant \infty, \qquad (10.19)$$

и

$$\sigma^2 = 1$$

$$\Delta = \frac{2}{\sqrt{\pi}}$$

Следовательно,

$$C = \sqrt{\frac{3}{\pi}}. \qquad (10.20)$$

**10.5.** Для распределения

$$dF=\frac{1}{\Gamma(m)}e^{-x}x^{m-1}dx, \quad 0\leqslant x\leqslant\infty, \qquad (10.21)$$

имеем:

$$\sigma^2=m,$$

$$\Delta=\frac{m\Gamma(2m+1)}{2^{2m-1}[\Gamma(m+1)]^2},$$

откуда, используя формулу

$$\pi^{\frac{1}{2}}\Gamma(2m+1)=2^{2m}\Gamma\left(m+\frac{1}{2}\right)\Gamma(m+1),$$

находим:

$$C=\sqrt{\frac{3m}{\pi}\frac{\Gamma\left(m+\frac{1}{2}\right)}{\Gamma(m+1)}}. \qquad (10.22)$$

Этот результат приведен в 9.6. Для $m=\frac{1}{2}$ (наименьшее значение, представляющее интерес для статистики); $C=0{,}78$; для $m=1$, $C=0{,}87$; для $m=4$, $C=0{,}95$.

## Конкордация

**10.6.** Теперь мы покажем, что $p_1$ и $p_2$—выборочные характеристики конкордации частот — являются несмещенными оценками генеральных величин $\pi_1$ и $\pi_2$. Это, вообще говоря, очевидно исходя из некоторых соображений, используемых в теории вероятностей, однако возможно, что простое доказательство не помешает.

Припишем любой паре членов совокупности некоторую переменную, которая равна 1, если наблюдается конкордация типа 1, и нулю, если ее нет. Математическое ожидание $p_1$ является тогда математическим ожиданием значения этой переменной для любой данной пары, поскольку математическое ожидание суммы есть сумма математических ожиданий даже тогда, когда слагаемые являются зависимыми величинами. Однако математическое ожидание этой переменной равно $\pi_1$, откуда и следует искомый результат. Такой путь доказательства применим также и к $p_2$ и $\pi_2$.

**10.7.** Пусть функция распределения $x$, $y$ будет $F(x, y)$. Функция распределения только $x$ или $y$ пусть будет $F(x, \infty)$ и $F(\infty, y)$ соответственно.

Теперь для любого фиксированного значения $x_j$ вероятность того, что $x_i<x_j$, есть $F(x_j, \infty)$ и, следовательно, вероятность того, что $y_i<y_j$ при данном $x_i<x_j$, есть $F(x_j, y_j)/F(x_j, \infty)$. Для того чтобы получить вероятности того, что для любых двух пар $(x_i, y_i)$ и $(x_j, y_j)$

$y_i < y_j$, если $x_i < x_j$, интегрируем соответствующие функции по $x_j$ и $y_j$:

$$\pi_1 = \frac{\int\int F(x_j, y_j)\, dF(x_j, y_j)}{\int\int F(x_j, \infty)\, dF(x_j, y_j)}. \qquad (10.23)$$

Теперь мы можем опустить подстрочный индекс $j$. Кроме того,

$$\int\int F(x, \infty)\, dF(x, y) = \int F(x, \infty)\, dF(x, \infty) =$$
$$= \left[\frac{1}{2} F^2(x, \infty)\right]_0^1 = \frac{1}{2}$$

и, следовательно,

$$\pi_1 = 2\int\int F(x, y)\, dF(x, y), \qquad (10.24)$$

интегрирование производится по всему диапазону значений $x$ и $y$.

Опираясь на аналогичные аргументы, находим:

$$\pi_2 = 2\int\int F(x, \infty)\, F(\infty, y)\, dF(x, y). \qquad (10.25)$$

10.8. Рассмотрим случай, когда переменные полностью связаны линейным отношением. Совместное распределение $x$ и $y$ становится одномерным и без потери общности можно предположить, что это распределение будет равномерным в диапазоне от 0 до 1. Тогда $F(x, y)$ приводится к распределению одной переменной, скажем, $z$, и

$$\pi_1 = 2\int_0^1 z dz = 1.$$

В случае независимости переменных $\pi_1 = 0$, следовательно, $\pi_1$ варьирует от 0 до 1. Поскольку рассматриваемая величина является вероятностью, она и не может лежать вне этого диапазона.

Для $\pi_2$ находим в случае полной линейной зависимости:

$$\pi_2 = 2\int_0^1 z^2\, dz = \frac{2}{3},$$

если эта зависимость отрицательная, то

$$\pi_2 = 2\int_0^1 z(1 - z)\, dz = \frac{1}{3}.$$

Эти величины представляют собой крайние значения, и легко видеть (ср. с. 31—32), что $\pi_2$ не может лежать вне интервала $\frac{1}{3}$, $\frac{2}{3}$. Соответственно коэффициент $6\left(\pi_2 - \frac{1}{2}\right)$ находится в пределах $-1 \div +1$.

**10.9.** Из (10.25) следует, что ковариация накопленных сумм частот $x$ и $y$ есть

$$\operatorname{cov} = \iint F(x, \infty) F(\infty, y)\, dF(x, y) -$$

$$- \left[\int F(x, \infty)\, dF(x, \infty)\right]\left[\int F(\infty, y)\, dF(\infty, y)\right] = \frac{1}{2}\pi_2 - \frac{1}{4},$$

а дисперсия составит величину: $\frac{1}{3} - \frac{1}{4} = \frac{1}{12}$. Отсюда следует, что корреляция накопленных частот равна:

$$\frac{1}{2}\left(\pi_2 - \frac{1}{2}\right) \Big/ \frac{1}{12} = 6\left(\pi_2 - \frac{1}{2}\right). \tag{10.26}$$

Эту величину мы и определили выше как $\rho_s$.

**10.10.** Теперь выведем выражения для средних и дисперсий $t$ и $r_s$ в случае нормального распределения.

Пусть $\operatorname{sgn} \xi$ равно $+1$, если $\xi$ положительно, 0 — если $\xi$ равно нулю, и наконец, $-1$, если $\xi$ отрицательно. Распространим этот результат для действительных значений $\xi$ следующим образом:

$$\left.\begin{aligned} \operatorname{sgn} \xi &= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{e^{it\xi}\, dt}{it} = \\ &= +1, \quad \xi > 0 \\ &= 0, \quad \xi = 0 \\ &= -1, \quad \xi < 0 \end{aligned}\right\}. \tag{10.27}$$

Интеграл здесь следует понимать как главную компоненту выражения, иначе говоря,

$$\int_{-\infty}^{\infty} = \lim_{\substack{c \to \infty \\ \varepsilon \to 0}} \left[\int_{-c}^{-\varepsilon} + \int_{\varepsilon}^{c}\right].$$

Выражение (10.27) эквивалентно действительному интегралу

$$\operatorname{sgn} \xi = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\sin t\xi\, dt}{t}. \tag{10.28}$$

Из определения главной компоненты становится ясно, что, если $\xi = 0$, то этот интеграл обращается в нуль в силу симметрии подынтегрального выражения. По-видимому, наикратчайший путь получения выражения (10.27) заключается в рассмотрении комплексного интеграла

$$\int \frac{e^{iz\xi}\, dz}{iz}.$$

Если $\xi$ положительна, мы берем этот интеграл по контуру, состоящему из отрезка на действительной оси с концами $-R$ и $-\varepsilon$, малой полуокружности радиуса $\varepsilon$, лежащей выше оси, отрезка на действительной оси с концами $\varepsilon$ и $R$ и большой полуокружности радиуса $R$, лежащей выше оси. Этот интеграл обращается в нуль, так как подынтегральное выражение не имеет «полюсов» внутри контура. Интеграл вдоль действительной оси стремится к

$$\int_{-\infty}^{\infty} \frac{e^{it\xi}\,dt}{it}.$$

Интеграл по большой полуокружности стремится к нулю при $R$, стремящемся к бесконечности. Интеграл по малой полуокружности является на самом деле просто интегралом от $dz/iz$ по этой полуокружности, пройденной по часовой стрелке, и равен $-\pi$. Таким образом,

$$\int_{-\infty}^{\infty} \frac{e^{it\xi}\,dt}{it} - \pi \simeq 0,$$

откуда следует результат (10.27) для $\xi > 0$. Если $\xi < 0$, мы рассматриваем интеграл, в котором $t$ берется с другим знаком.

**10.11.** Рассмотрим теперь нормальную совокупность переменных $x$, $y$, коэффициент корреляции которых равен $\rho$. Запишем для них:

$$dF = \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\left[-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right] dx\,dy \qquad (10.29)$$

Без потери общности можно предположить, что случайные величины измерены как отклонения от нулевой средней, их дисперсии равны единице. Если мы возьмем пары значений, скажем, $x_1$ и $x_2$, то можно использовать оценку, базирующуюся на измерении разности между $x_1$ и $x_2$. При построении коэффициента $\tau$ будем исходить из того, что такая оценка определяется знаком при $(x_1 - x_2)$ или при соответствующем положительном численном значении этой разности. Распределение пар независимых значений $x_1$ и $x_2$, $y_1$ и $y_2$ характеризуется следующим выражением:

$$dF \propto \exp\left\{-\frac{1}{2(1-\rho^2)}[x_1^2 + x_2^2 - 2\rho(x_1 y_1 + x_2 y_2) + y_1^2 + y_2^2]\right\} dx_1\,dx_2\,dy_1\,dy_2. \qquad (10.30)$$

Положим:

$$u_1 = \frac{1}{\sqrt{2}}(x_1 - x_2), \qquad u_2 = \frac{1}{\sqrt{2}}(x_1 + x_2),$$

$$v_1 = \frac{1}{\sqrt{2}}(y_1 - y_2), \qquad v_2 = \frac{1}{\sqrt{2}}(y_1 + y_2).$$

Распределение тогда примет вид:

$$dF \propto \exp\left[-\frac{1}{2(1-\rho^2)}(u_1^2 - 2\rho u_1 v_1 + v_1^2)\right] du_1\,dv_1 \times$$
$$\times \exp\left[-\frac{1}{2(1-\rho^2)}(u_2^2 - 2\rho u_2 v_2 + v_2^2)\right] du_2\,dv_2. \qquad (10.31)$$

Следовательно, $u_1$ и $v_1$ также нормально распределены и имеют корреляцию $\rho$, не зависимую от $u_2$ и $v_2$. Опуская подстрочные значки, получим:

$$dF \propto \exp\left\{-\frac{1}{2(1-\rho^2)}\left[u^2-2\rho uv+v^2\right]\right\} du\, dv. \tag{10.32}$$

**10.12.** Если $t$ есть выборочное значение $\tau$, $E(t)$ — математическое ожидание суммы $\frac{1}{2} n(n-1)$ членов, каждый из которых может быть записан как $\operatorname{sgn} u \operatorname{sgn} v$, то

$$E(t)=E(\operatorname{sgn} u \ \operatorname{sgn} v)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \operatorname{sgn} u \ \operatorname{sgn} v\, dF.$$

Это выражение, в соответствии с (10.27), примет вид:

$$\frac{1}{\pi^2}\int_{-\infty}^{\infty}\frac{dt_1}{it_1}\int_{-\infty}^{\infty}\frac{dt_2}{it_2}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{iut_1+ivt_2}\, dF\right]. \tag{10.33}$$

Выражение в квадратных скобках есть характеристическая функция $u$ и $v$, оно равно[1]:

$$\exp-\frac{1}{2}\left(t_1^2+2\rho t_1 t_2+t_2^2\right). \tag{10.34}$$

Отсюда

$$E(t)=\frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{dt_1}{it_1}\frac{dt_2}{it_2}\exp-\frac{1}{2}\left(t_1^2+2\rho t_1 t_2+t_2^2\right).$$

Таким образом,

$$\frac{\partial E(t)}{\partial\rho}=\frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dt_1\, dt_2 \exp-\frac{1}{2}\left(t_1^2+2\rho t_1 t_2+t_2^2\right)=\frac{2}{\pi\sqrt{(1-\rho^2)}}. \tag{10.35}$$

Следовательно, простое интегрирование по $\rho$, учитывая, что $E(t)$ обращается в нуль, когда $\rho=0$, дает

$$E(t)=\frac{2}{\pi}\sin^{-1}\rho. \tag{10.36}$$

Этот результат и приведен в (9.11).

**10.13.** Для того чтобы найти дисперсию $t$ во всех возможных выборках, нам необходимо определить

$$E(t^2)=E(\Sigma \operatorname{sgn} u \ \operatorname{sgn} v)^2,$$

---

[1] См. [58; 1; пример 3.17]. Этот интеграл легко определяется непосредственно.

где суммирование производится по всем $\binom{n}{2}$ значениям $u$ и $v$. Можно написать:

$$\Sigma (\operatorname{sgn} u \ \operatorname{sgn} v)^2 = \Sigma \operatorname{sgn} u_{ij} \operatorname{sgn} u_{kl} \operatorname{sgn} v_{ij} \operatorname{sgn} v_{kl}.$$

Здесь возможны три случая:

1) если $i = k$, $j = l$, то этот член равен $+1$ и математическое ожидание каждого слагаемого равно $+1$;

2) если $i \neq k$, $j \neq l$, то математическое ожидание этого произведения в силу (10.36) сводится к

$$E (\operatorname{sgn} u_{ij} \operatorname{sgn} v_{ij}) E (\operatorname{sgn} u_{kl} \operatorname{sgn} v_{kl}) = \left(\frac{2}{\pi} \sin^{-1} \rho\right)^2; \qquad (10.37)$$

3) если только $i = k$ или только $j = l$, то получим тот тип, который может быть оценен при рассмотрении случая $i = k = 1$, $j = 2$, $k = 3$. Для удобства будем использовать единственный общий знак. Пусть

$$M = E (\operatorname{sgn} u_{12} \operatorname{sgn} v_{12} \operatorname{sgn} u_{13} \operatorname{sgn} v_{13}) =$$
$$= \frac{1}{\pi^4} \int \frac{dt_1}{it_1} \frac{dt_2}{it_2} \frac{dt_3}{it_3} \frac{dt_4}{it_4} \int f e^{i\Omega} \, dx_1 \, dx_2 \, dx_3 \, dy_1 \, dy_2 \, dy_3, \qquad (10.38)$$

куда теперь, опуская $\sqrt{2}$, введем $u_{12} = x_1 - x_2$, и т. д. Таким образом,

$$\Omega = (x_1 - x_2) t_1 + (y_1 - y_2) t_2 + (x_1 - x_3) t_3 + (y_1 - y_3) t_4 =$$
$$= x_1 (t_1 + t_3) - x_2 t_1 - x_3 t_3 + y_1 (t_2 + t_4) - y_2 t_2 - y_3 t_4. \qquad (10.39)$$

Для интегрирования $f e^{i\Omega}$ по $x$ и $y$ мы можем воспользоваться известными свойствами характеристических функций или интегрировать непосредственно. Находим:

$$T = \int f e^{i\Omega} \, dx_1 \ldots dy_1 \ldots = \exp \left\{ -\frac{1}{2} [(t_1 + t_3)^2 + (t_2 + t_4)^2 + t_1^2 + t_2^2 + \right.$$
$$\left. + t_3^2 + t_4^2] + 2\rho' [(t_1 + t_3)(t_2 + t_4) + t_1 t_2 + t_3 t_4] \right\}$$

и, следовательно,

$$\frac{\partial T}{\partial \rho'} = -T [2t_1 t_2 + 2t_3 t_4 + t_1 t_4 + t_2 t_3]. \qquad (10.40)$$

Если мы продифференцируем $M$ в (10.38) по $\rho$ и используем соотношение (10.40), то получим выражение

$$\frac{\partial M}{\partial \rho} = \frac{2}{\pi^4} \int dt_1 \, dt_2 \frac{dt_3}{it_3} \frac{dt_4}{it_4} T + \frac{2}{\pi^4} \int dt_3 \, dt_4 \frac{dt_1}{it_1} \frac{dt_2}{it_2} T +$$
$$+ \frac{1}{\pi^4} \int dt_1 \, dt_4 \frac{dt_2}{it_2} \frac{dt_3}{it_3} T + \frac{1}{\pi^4} \int dt_2 \, dt_3 \frac{dt_1}{it_1} \frac{dt_4}{it_4} T.$$

В силу симметрии $T$ относительно $t_1$ и $t_3$, $t_2$ и $t_4$ это выражение может быть сведено к

$$\frac{\partial M}{\partial\rho}=\frac{4}{\pi^4}\int dt_1\,dt_2\,\frac{dt_3}{it_3}\,\frac{dt_4}{it_4}\,T+\frac{2}{\pi^4}\int dt_1\,dt_4\,\frac{dt_2}{it_2}\,\frac{dt_3}{it_3}\,T. \qquad (10.41)$$

Произведем интегрирование в первой части относительно $t_1$ и $t_2$. Получаем:

$$\int dt_1\,dt_2\,T=\frac{\pi}{\sqrt{1-\rho^2}}\exp-\frac{3}{4}(t_3^2+2\rho t_3t_4+t_4^2).$$

Оставшаяся часть интегрирования может быть выполнена способом, который привел к оцениванию (10.33).

Аналогичным путем может быть исчислен второй интеграл в (10.41). Окончательно

$$\frac{\partial M}{\partial\rho}=\frac{8\sin^{-1}\rho}{\pi^2\sqrt{1-\rho^2}}-\frac{4}{\pi^2}\,\frac{\sin^{-1}\frac{1}{2}\rho}{\sqrt{1-\left(\frac{1}{2}\rho\right)^2}}. \qquad (10.42)$$

Когда $\rho=1$, $M=1$. Вновь обращаясь к интегрированию, находим, что:

$$M=\left(\frac{2}{\pi}\sin^{-1}\rho\right)^2-\left(\frac{2}{\pi}\sin^{-1}\frac{1}{2}\rho\right)^2+\frac{1}{9}. \qquad (10.43)$$

Имеется $\binom{n}{2}$ случаев типа I, $\binom{n}{2}\binom{n-2}{2}$ случаев типа II и 6 $\binom{n}{3}$ случаев типа III. Таким образом,

$$E(t^2)=\frac{1}{\binom{n}{2}^2}\left\{\binom{n}{2}+\binom{n-2}{2}\left(\frac{2}{\pi}\sin^{-1}\rho\right)^2+\right.$$
$$\left.+6\binom{n}{3}\left[\frac{1}{9}+\left(\frac{2}{\pi}\sin^{-1}\rho\right)^2-\left(\frac{2}{\pi}\sin^{-1}\frac{1}{2}\rho\right)^2\right]\right\}. \qquad (10.44)$$

После вычитания квадрата $E(t)$ и небольших преобразований находим:

$$\operatorname{var} t=\frac{1}{\binom{n}{2}}\left\{1-\left(\frac{2}{\pi}\sin^{-1}\rho\right)^2+2(n-2)\times\right.$$
$$\left.\times\left[\frac{1}{9}-\left(\frac{2}{\pi}\sin^{-1}\frac{1}{2}\rho\right)^2\right]\right\}. \qquad (10.45)$$

Этот результат приведен в (9.13).

**10.14.** Указанную формулу можно аппроксимировать следующим образом. Если

$$\sin^{-1}\frac{1}{2}\rho=\alpha,$$
$$\frac{1}{3}\sin^{-1}\rho=\beta,$$

то

$$2\sin\alpha = \sin 3\beta = 3\sin\beta\left(1 - \frac{4}{3}\sin^2\beta\right).$$

Однако

$$|\beta| \leqslant \frac{\pi}{6}$$

и, следовательно,

$$1 - \frac{4}{3}\sin^2\beta \geqslant \frac{2}{3}.$$

Таким образом,

$$|\alpha| \geqslant |\beta|,$$

поэтому

$$0 \leqslant \frac{1}{9} - \left(\frac{2}{\pi}\sin^{-1}\frac{1}{2}\rho\right)^2 \leqslant \frac{1}{9}\left[1 - \left(\frac{2}{\pi}\sin^{-1}\rho\right)^2\right] \leqslant \frac{4}{9}pq, \quad (10.46)$$

где $p$ и $q$ определены в 9.11. Кроме того,

$$\frac{2}{n(n-1)}\left[1 + \frac{2}{9}(n-2)\right] \leqslant \frac{5}{9(n-1)}, \quad n \geqslant 10. \quad (10.47)$$

Применяя этот результат в (10.45), находим:

$$\operatorname{var} t \leqslant \frac{20pq}{9(n-1)} = -\frac{5(1-t^2)}{9(n-1)}, \quad (10.48)$$

как это показано в (9.18).

**10.15.** Остальные нужные нам результаты выводятся таким же методом, каким выводились $E(t)$ и $\operatorname{var} t$; однако здесь вычисления становятся более громоздкими.

Рассмотрим прежде всего $E(r_s)$. Как и в (2.35), имеем:

$$E(r_s) = \frac{3}{n^3-n} E\left[\Sigma a_{ij} b_{ij} + \Sigma a_{ij} b_{ik}\right] =$$

$$= \frac{3}{n^3-n}\left[n(n-1)E(a_{ij}b_{ij}) + n(n-1)(n-2)E(a_{ij}b_{ik})\right] =$$

$$= \frac{3}{n+1}E(a_{ij}b_{ij}) + \frac{3(n-2)}{n+1}E(a_{ij}b_{ik}), \quad j \neq k. \quad (10.49)$$

Мы уже определили, что

$$E(a_{ij}b_{ij}) = \frac{2}{\pi}\sin^{-1}\rho.$$

Для оценивания $E(a_{ij}b_{ik})$ рассмотрим случай, когда $i=1$, $j=2$, $k=3$. Таким же способом, что и в 10.11, находим

$$E(a_{12}b_{13})=\frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{dt_1}{it_1}\frac{dt_2}{it_2}E\{\exp[it(x_1-x_2)+it_2(y_1-y_2)]\}=$$

$$=\frac{1}{\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{dt_1}{it_1}\frac{dt_2}{it_2}\exp\left[-\frac{1}{2}(t_1^2+\rho t_1t_2+t_2^2)\right]. \qquad (10.50)$$

Это выражение аналогично интегралу, приводившемуся в 10.12, только здесь вместо $2\rho$ стоит $\rho$ и поэтому мы получаем $\frac{2}{\pi}\sin^{-1}\frac{1}{2}\rho$. Отсюда находим, что

$$E(r_s)=\frac{6}{\pi(n+1)}\left[\sin^{-1}\rho+(n-2)\sin^{-1}\frac{1}{2}\rho\right]. \qquad (10.51)$$

**10.16.** Оценивание $\operatorname{var} r_s$ и $\operatorname{cov}(r_s, t)$ осуществляется тем же способом. Однако в первом случае мы сталкиваемся с интегралом эллиптического вида. Следует разложить полученную функцию в ряд по степеням $\rho$. Это приводит к уравнениям (9.30) и (9.31). Подробности можно найти в [53] и [16].

## Библиография

См. список литературы к гл. 9. В [53] рассматривается взаимосвязь между ранговыми коэффициентами и исходными параметрами для генеральных совокупностей, не являющихся нормальными и представленными в виде рядов Грема—Шарлье. В [98] приводятся третьи и четвертые моменты относительно $t$ и найден метод их оценивания для нормального случая.

В [27] получены некоторые важные формулы для асимптотической дисперсии $r_s$. В [15] выведены некоторые точные выражения, которые были табулированы в [28].

## *ГЛАВА 11.* ПАРНЫЕ СРАВНЕНИЯ

**11.1.** До сих пор предполагалось, что последовательности рангов содержатся в самой постановке задачи, и мы не задавались вопросом о том, в какой степени рассматриваемые данные удовлетворяют требованиям, предъявляемым к последовательностям рангов. Однако во многих случаях (особенно часто — в психологических исследованиях) возникают некоторые сомнения относительно правомерности такой постановки вопроса. Допустим, мы попросили эксперта ранжировать десять человек по степени развития их интеллектуальных способностей. Предположим даже, что эксперт смог справиться с этой задачей (хотя, вообще говоря, природа интеллекта настолько неясна, что мы не можем быть уверенными в самой возможности адекватного упорядочения индивидуумов по этому признаку). Мы допускаем какую-то произвольность в наших рассуждениях, предполагая, что переменная, характеризующая интеллектуальные способности, линейна. Другой пример: мы можем предложить эксперту упорядочить некоторое число географических районов в соответствии со своими предпочтениями: насколько привлекателен каждый из них в качестве места жительства. Однако в этом случае предпочтения будут зависеть от взаимодействия различных факторов, таких, как стоимость жизни, наличие удобного транспорта, высота над уровнем моря, близость к торговым центрам; поэтому нет никакой уверенности в том, что наблюдатель сможет выразить свои окончательные предпочтения с помощью линейной шкалы. Даже в тех случаях, когда мы настаиваем на том, чтобы эксперт упорядочил свои предпочтения, а он стремится выполнить указанную просьбу, будучи уверенным в том, что располагает соответствующими возможностями, эти попытки могут свестись к тому, что мы будем, так сказать, втискивать наши данные в рамки чрезмерно узкой схемы, которая исказит реальную действительность. В этой главе рассматривается метод, с помощью которого можно преодолеть подобные трудности.

**11.2.** Предположим, что имеется $n$ объектов и эксперту поочередно представляют каждую из $\frac{1}{2}n(n-1)$ пар объектов. Эксперт сопоставляет между собой представленные объекты и по каждой паре записывает свои предпочтения. Если объект $A$ предпочитают объекту $B$, мы можем записать это следующим образом: $A \rightarrow B$ или $B \leftarrow A$.

*Пример* 11.1.

В экспериментальных целях было подготовлено шесть различных видов корма для собак. Собаке поочередно предлагалась каждая из $\binom{6}{2} = 15$ возможных пар, составленных из этих видов корма; при этом каждый раз фиксировалось, какую пищу собака ест в первую очередь. (Подобный опыт проводился с целью иллюстрации справедливости некоторых предположений, и к нему, разумеется, нельзя относиться как к некой серьезной попытке выявить собачьи предпочтения.) Обозначив виды корма с помощью букв латинского алфавита от $A$ до $F$, объединим полученные результаты в таблицу:

Таблица 11.1

Реакции собаки при сопоставлении шести видов корма

| | $A$ | $B$ | $C$ | $D$ | $E$ | $F$ |
|---|---|---|---|---|---|---|
| $A$ | — | 1 | 1 | 0 | 1 | 1 |
| $B$ | 0 | — | 0 | 1 | 1 | 0 |
| $C$ | 0 | 1 | — | 1 | 1 | 1 |
| $D$ | 1 | 0 | 0 | — | 0 | 0 |
| $E$ | 0 | 0 | 0 | 1 | — | 1 |
| $F$ | 0 | 1 | 0 | 1 | 0 | — |

Запись 1 в клетке, находящейся на пересечении, скажем, столбца $Y$ и строки $X$, означает $X \to Y$; разумеется, в клетке, находящейся на пересечении строки $Y$ и столбца $X$, должен стоять 0. Таким образом, в нашей таблице $A \to B$, $A \to C$, $A \leftarrow D$ и т. д. Диагональные элементы таблицы исключены.

Последовательность, в которой различные виды корма записываются по строкам и столбцам таблицы, произвольна; понятно, однако, что удобнее всего располагать объекты по строкам и по столбцам в одинаковом порядке.

Всю систему предпочтений можно охарактеризовать с помощью диаграммы. Для этого расположим шесть объектов от $A$ до $F$ по вершинам правильного многоугольника (как показано на рис. 11.1). Проведем все возможные линии, соединяющие вершины этого многоугольника, и если $X \to Y$, направим стрелку из $X$ в $Y$ вдоль линии $XY$.

**11.3.** Предположим, что нам даны три объекта и эксперт высказал следующие предпочтения: $X \to Y \to Z \to X$ или $X \leftarrow Y \leftarrow Z \leftarrow X$. В этом случае будем говорить, что предпочтения по трем объектам характеризуются цикличностью или что предпочтения «несовместны». С помощью диаграммы (подобной 11.1) можно показать, что цикличные предпочтения по трем объектам позволяют, двигаясь вдоль стрелок, последовательно обойти все вершины треугольника $XYZ$. На рис. 11.1 цикличными оказываются предпочтения по объектам $ACD$ и $BEF$ (можно найти также три другие цикличные тройки объектов).

Ясно, что цикличность предпочтений не может возникать в обычных последовательностях рангов, так как если $X \to Y$ и $Y \to Z$, то $X \to Z$. Необходимым и достаточным условием того, что предпочтения могут быть выражены с помощью последовательности рангов, служит отсутствие цикличных предпочтений по трем объектам. Чем более интенсивно (если можно так выразиться) проявляется цикличность предпочтений по трем объектам, тем меньше возможностей составить последовательность рангов и тем ближе мы оказываемся к «несовместной» ситуации, когда объект $X$ может предпочитаться объекту $Y$, объект $Y$ — объекту $Z$ и тем не менее объект $Z$ предпочитается объекту $X$.

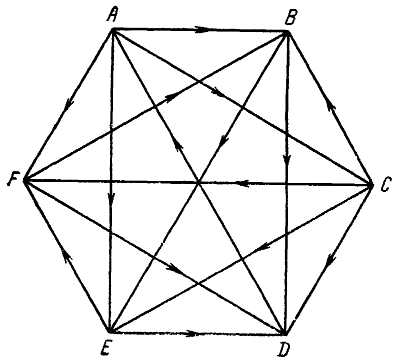

Рис. 11.1

**11.4.** Иногда встречаются циклы предпочтений, содержащие более трех объектов. Например, пусть $A \to B \to C \to D \to A$; здесь цикличностью характеризуются предпочтения по четырем объектам. Цикл предпочтений по $n$ объектам должен содержать по крайней мере $n - 2$ циклов по трем объектам (их может быть и больше); но если в состав системы предпочтений, включающей $n$ объектов, входят циклы по трем объектам, из этого еще не следует, что цикличны предпочтения по всей системе. Предположим, например, что предпочтения по объектам $ABCD$ образуют цикл. Тогда или $A \to C$, или $C \to A$. В первом случае цикличны предпочтения по объектам $ACD$, во втором — по объектам $ABC$. Рассуждая аналогично, можно показать, что должны быть цикличны предпочтения либо по объектам $ABD$, либо по объектам $BCD$. Таким образом, цикл предпочтений по четырем объектам должен включать по меньшей мере два цикла предпочтений по трем объектам. Допустим теперь, что имеют место соотношения следующего типа: $A \to B \to C \leftarrow D \to A$, $B \to D$, $C \to A$; такая система предпочтений содержит циклы по трем объектам: $ABC$ и $ABD$, однако предпочтения по всей системе $ABCD$ нецикличны.

В дальнейшем изложении мы будем рассматривать прежде всего циклы предпочтений по трем объектам (поскольку они образуют «элементарные клеточки» несовместной ситуации). Мы не будем прибегать к более сложным (и менее определенным) критериям выявления циклов по совокупностям высшего порядка.

**11.5.** В следующей главе будет показано, что когда величина нечетна, максимальное количество циклов по трем объектам равно $\frac{1}{24}(n^3 - n)$, а при четном $n$ равно $\frac{1}{24}(n^3 - 4n)$. Минимальное их число равно нулю. Исходя из этих соображений, можно ввести коэффициент совместности, определив его с помощью следующих уравнений:

$$\begin{aligned} \zeta &= 1-\frac{24d}{n^3-n}, \text{ если } n \text{ нечетно;} \\ \zeta &= 1-\frac{24d}{n^3-4n}, \text{ если } n \text{ четно,} \end{aligned} \tag{11.1}$$

где $d$ — число обнаружившихся циклов предпочтений по трем объектам. Циклы предпочтений по трем объектам отсутствуют и ранговые оценки удается упорядочить тогда и только тогда, когда коэффициент $\xi$ равен единице.

Например, оценки, приведенные в примере 11.1, содержат 5 циклов предпочтений по трем объектам. Максимальное число таких циклов равно 8, следовательно, $\zeta = 0{,}375$.

**11.6.** Для того чтобы как-то проверить существенность коэффициента $\zeta$, выясним, как будут распределены эти величины в том случае, когда все предпочтения складываются случайным образом. Благодаря этому мы сможем узнать, может ли наблюдаемое значение коэффициента $\zeta$ представлять собой случайную величину (такая ситуация имеет место, когда эксперт совершенно некомпетентен), либо же, напротив, предпочтения эксперта характеризуются некоторой, хотя, возможно, и не идеальной согласованностью.

Предположив, что предпочтения носят чисто случайный характер (и, следовательно, одна система предпочтений столь же вероятна, как и любая другая), можно подсчитать вероятности, характеризующие различные значения коэффициента $d$. В приложении в табл. 9 приведены вероятности того, что будут достигнуты или превышены некоторые величины $d$ при $n$, меняющем значения от 2 до 7. На практике редко приходится иметь дело с большими значениями $n$. Допустим, однако, что нам требуется провести статистическую проверку при больших значениях $n$; поскольку расчет подобных распределений — дело довольно трудоемкое, можно воспользоваться распределением $\chi^2$, к которому с ростом $n$ стремится распределение $d$. Действительно, выписав значения $\nu$ и $\chi^2$:

$$\left.\begin{aligned} \nu &= \frac{n(n-1)(n-2)}{(n-4)^2} \\ \chi^2 &= \frac{8}{n-4}\left\{\frac{1}{4}\binom{n}{3}-d+\frac{1}{2}\right\}+\nu \end{aligned}\right\} \tag{11.2}$$

можно видеть, что величины $\chi^2$ с $\nu$ степенями свободы распределены примерно таким же образом. Однако в распределении величин $d$ отсчет ведется от большего значения к меньшему, поэтому и вероятность того, что $d$ достигает некоторой величины или превосходит ее, не совпадает с соответствующей вероятностью для $\chi^2$, а представляет собой ее дополнение (до единицы). Следующий пример поможет лучше понять технику вычислений.

*Пример 11.2*

Предположим, что мы имеем дело с системой, состоящей из 7 объектов, причем $d = 13$. На основании (11.2) можно записать:

$$\nu = \frac{7\times6\times5}{9} = 23{,}33;$$

$$\chi^2 = \frac{8}{3}\left(8{,}75 - 13 + \frac{1}{2}\right) + 23{,}33 = 13{,}33.$$

В приведенной в приложении табл. 8, найдем, что этим значениям примерно соответствует уровень существенности, равный 0,95. В таком случае соответствующий уровень существенности для $d$ равен 1—0,95 = 0,05. Приведенное в табл. 9 приложения точное значение вероятности равно 0,36. Следовательно, даже для столь малого значения $n$, как 7, результаты аппроксимации вполне удовлетворительны.

**11.7.** Если мы располагаем таблицей, подобной 11.1, общую численность циклов предпочтений по трем объектам можно определить и более простым путем. Обозначим, например, суммы чисел по строкам с помощью букв $a_1, \ldots, a_n$.

Тогда

$$d = \frac{1}{6}n(n-1)(n-2) - \frac{1}{2}\sum_{i=1}^{n} a_i(a_i - 1). \tag{11.3}$$

Так, в табл. 11.1 суммы оценок, исчисленные по строкам, соответственно равны 4, 2, 4, 1, 2, 2, итого $15 = \binom{6}{2}$ единиц.

Таким образом,

$$\frac{1}{2}\sum a(a-1) = \frac{1}{2}\sum a^2 - \frac{1}{2}\sum a = \frac{1}{2}(45 - 15) = 15$$

и, следовательно, $d = 20 - 15 = 5$.

Можно использовать эту же формулу применительно к суммам оценок по столбцам (обозначим эти суммы буквами $b_1, \ldots, b_n$). Исходя из самого построения таблицы, можно записать:

$$b_i = (n-1) - a_i.$$

Таким образом,

$$\sum b_i = \sum a_i;$$

$$\sum b_i^2 = n(n-1)^2 - 2(n-1)\sum a + \sum a^2 = \sum a_i^2$$

и, следовательно,

$$\sum b(b-1) = \sum a(a-1).$$

Соотношение (11.3) можно записать также в следующем виде:

$$d = \frac{1}{12}n(n-1)(2n-1) - \frac{1}{2}\sum a_i^2. \tag{11.4}$$

Вероятно, это простейший путь исчисления $d$.

## Коэффициент согласия

11.8. Предположим, в опыте участвуют $m$ экспертов, каждый из которых рассмотрел все возможные пары из $n$ объектов и высказал $\binom{n}{2}$ предпочтений. Построим таблицу, подобную 11.1, причем во всех случаях, когда $X \rightarrow Y$, мы будем вписывать единицу в клетку, расположенную на пересечении $x$-й строки и $y$-го столбца, а затем подсчитаем количество единиц в каждой клетке. В ней может оказаться любое число единиц — от 0 до $m$. Если наблюдатели единодушны, то $\binom{n}{2}$ клеток будут содержать по $m$ единиц, а остальные клетки — нули. Согласие между наблюдателями может быть полным даже в том случае, когда встречаются циклы предпочтений.

Предположим, что в клетке, находящейся на пересечении $x$-й строки и $y$-го столбца, вписано число $\gamma$. Введем теперь следующее соотношение:

$$\Sigma = \sum \binom{\gamma}{2}, \tag{11.5}$$

причем суммирование ведется по $n\,(n-1)$ клеткам таблицы (диагональные клетки не учитываются). Тогда $\Sigma$ будет представлять собой общее число случаев, в которых пара экспертов согласна между собой. Положим

$$u = \frac{2\Sigma}{\binom{m}{2}\binom{n}{2}} - 1 = \frac{8\Sigma}{m\,(m-1)\,n\,(n-1)} - 1. \tag{11.6}$$

Величину $u$ мы будем называть *коэффициентом согласия*; $u = 1$ в том и только в том случае, когда имеет место полная согласованность в предпочтениях экспертов. Чем меньше степень согласованности (измеряемая совпадением или несовпадением предпочтений каждой пары экспертов), тем меньшие значения принимает коэффициент $u$. В таком случае число единиц в клетке, характеризующее минимальную согласованность предпочтений, составляет $\frac{1}{2}m$, если величина $m$ четна, а в противном случае $\frac{1}{2}(m \pm 1)$. Тогда в первом случае $u$ равно $-1/(m-1)$, во втором $-1/m$.

Только при $m = 2$ $u$ достигает своего минимального значения, равного $-1$. Поскольку в опыте участвуют всего два эксперта, наша формула принимает следующий вид:

$$u = \frac{2\Sigma}{\frac{1}{2}n\,(n-1)} - 1. \tag{11.7}$$

В этом случае $u$ можно представить как некое обобщение коэффициента $\tau$.

*Пример 11.3.*

В классе мальчиков (в возрасте от 11 до 13 лет) был проведен опрос, в ходе которого им предложили сформулировать свои предпочтения относительно некоторых учебных дисциплин. Каждый из опрашиваемых получил лист бумаги, на котором были выписаны все возможные пары дисциплин; в каждом случае требовалось подчеркнуть название дисциплины, которой оказывалось предпочтение. В классе насчитывался 21 ученик, а число учебных дисциплин равно 13. В результате опроса были получены результаты, сведенные в табл. 11.2.

Таблица 11.2. **Предпочтения, высказанные 21 учеником относительно 13 предметов**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | Сумма |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Ручной труд | — | 14 | 20 | 15 | 15 | 16 | 16 | 18 | 18 | 18 | 20 | 21 | 20 | 211 |
| 2. Гимнастика | 7 | — | 14 | 12 | 13 | 18 | 14 | 16 | 16 | 20 | 16 | 18 | 19 | 183 |
| 3. Рисование | 1 | 7 | — | 10 | 14 | 10 | 16 | 18 | 16 | 16 | 17 | 16 | 19 | 160 |
| 4. Естественные науки | 6 | 9 | 11 | — | 11 | 12 | 15 | 14 | 13 | 13 | 17 | 17 | 16 | 154 |
| 5. История | 6 | 8 | 7 | 10 | — | 14 | 11 | 12 | 14 | 15 | 13 | 14 | 16 | 140 |
| 6. География | 5 | 3 | 11 | 9 | 7 | — | 14 | 14 | 13 | 13 | 16 | 15 | 17 | 137 |
| 7. Арифметика | 5 | 7 | 5 | 6 | 10 | 7 | — | 9 | 11 | 13 | 15 | 13 | 15 | 116 |
| 8. Религия | 3 | 5 | 3 | 7 | 9 | 7 | 12 | — | 12 | 14 | 14 | 16 | 14 | 116 |
| 9. Английская литература | 3 | 5 | 5 | 8 | 7 | 8 | 10 | 9 | — | 10 | 13 | 13 | 15 | 106 |
| 10. Экономические дисциплины | 3 | 1 | 5 | 8 | 6 | 8 | 8 | 7 | 11 | — | 10 | 10 | 14 | 91 |
| 11. Алгебра | 1 | 5 | 4 | 4 | 8 | 5 | 6 | 7 | 8 | 11 | — | 10 | 13 | 82 |
| 12. Английский язык | 0 | 3 | 5 | 4 | 7 | 6 | 8 | 5 | 8 | 11 | 11 | — | 13 | 81 |
| 13. Геометрия | 1 | 2 | 2 | 5 | 5 | 4 | 6 | 7 | 6 | 7 | 8 | 8 | — | 61 |
| | | | | | | | | | | | Итого | | | 1 638 |

Наименования учебных дисциплин в этой таблице упорядочены в соответствии с общим числом выявленных предпочтений, поэтому величину $\Sigma$ легче вычислить следующим образом: соотношение (11.5) можно привести к виду

$$\Sigma = \Sigma(\gamma)^2 - m\,\Sigma(\gamma) + \binom{m}{2}\binom{n}{2},$$

а суммирование будем проводить по той половине таблицы, что лежит ниже диагонали. Благодаря тому, что расположенные здесь числа меньше, чем числа в верхней половине таблицы, удается несколько облегчить соответствующие арифметические операции.

Подсчеты показывают, что

$$\Sigma = 9\,718$$

и, следовательно,

$$u = \frac{2 \times 9\,718}{\binom{21}{2}\binom{13}{2}} - 1 = 0{,}186.$$

Таким образом, мы исчислили некоторую величину, характеризующую степень согласованности между предпочтениями учеников; она измеряется положительным значением коэффициента $u$.

Циклы предпочтений по трем объектам в этом примере распределяются следующим образом:

| Число циклов предпочтений по трем объектам | Частота появления | Число циклов предпочтений по трем объектам | Частота появления |
|---|---|---|---|
| 0 | 1 | 12 | 1 |
| 1 | 1 | 17 | 3 |
| 4 | 5 | 21 | 1 |
| 6 | 2 | 25 | 1 |
| 7 | 2 | 29 | 1 |
| 8 | 1 | 39 | 1 |
| 10 | 1 | | |
| | | Общее количество | 21 |

Общее число таких циклов предпочтений равно 242, а их среднее значение — 11,5. Лишь у одного ученика была выявлена полностью совместная (непротиворечивая) система предпочтений. С другой стороны, для $n = 13$ максимальное число циклов предпочтений по трем объектам равно 91 при среднем значении, равном 71,5. Ясно, что за исключением, быть может, одного ученика, все остальные высказывали такие предпочтения, которые нельзя полагать случайными. Следовательно, мы вновь можем убедиться в том, что ребята действительно способны формулировать свои предпочтения. Это проявилось, в частности, в том, что у половины мальчиков значения коэффициента $\zeta$ превышают 0,92.

*Пример 11.4*

На практике часто встречается ситуация, когда эксперты отказываются высказать отчетливое предпочтение одному из двух рассматриваемых объектов. В этом случае мы сталкиваемся с такими же трудностями, как при анализе связанных рангов. Для того чтобы каким-то образом преодолеть затруднения, мы при составлении таблицы типа 11.1 будем использовать следующий прием. Если при сравнении объектов $X$ и $Y$ не выявлено никакого предпочтения, будем вписывать величину «$\frac{1}{2}$» как в клетку, находящуюся на пересечении строки $X$ и столбца $Y$, так и в клетку, находящуюся на пересечении строки $Y$ и столбца $X$. Проиллюстрируем этот метод.

На предприятии был проведен опрос 46 работников: им предложили сравнить между собой попарно ряд возможных улучшений (всего 66 пар), выделив в каждой паре ту перспективу, которая каждому из них представляется более важной.

(*А*) Улучшение вентиляции
(*Б*) Улучшение питания в столовой предприятия
(*В*) Переход на работу, предполагающую большую ответственность
(*Г*) Улучшение пенсионного обеспечения
(*Д*) Хорошие условия для продвижения по службе
(*Е*) Улучшение раздевалок, туалетов и т. д.
(*Ж*) Переход на работу, не требующую размышлений
(*З*) Лучшее освещение
(*И*) Переход на более интересную работу
(*К*) Сокращение рабочего времени
(*Л*) Гарантия занятости (т. е. гарантия работы вообще)
(*М*) Гарантия работы именно на данном предприятии

Результаты опроса сведены в таблицу (см. стр. 166).

Подсчитав суммы квадратов приведенных в таблице значений, мы найдем:

$$\sum \gamma^2 = 86\,392, \quad \sum \gamma = 3036.$$

Таким образом,

$$\frac{1}{2} \sum \gamma(\gamma - 1) = 41\,678.$$

Если, невзирая на то, что $\gamma$ теперь может принимать дробные значения, мы будем пользоваться прежней формулой (11.6), то получим:

$$u = \frac{2 \times 41\,678}{\binom{46}{2}\binom{12}{2}} - 1 = 0{,}220.$$

Рассмотрим теперь одну из оценок, например число, которое записано в клетке, находящейся на пересечении строки *Б* и столбца *В* $\left(\text{т. е. } 24\frac{1}{2}\right)$, а также его дополнение — число, которое расположено в клетке, находящейся на пересечении столбца *Б* и строки *В* $\left(\text{т. е. } 21\frac{1}{2}\right)$. Эти дробные величины дадут следующие оценки, включаемые в общие результаты наших расчетов:

$$\binom{24\frac{1}{2}}{2} + \binom{21\frac{1}{2}}{2} = \frac{1}{2}\left\{\binom{25}{2} + \binom{24}{2}\right\} + \frac{1}{2}\left\{\binom{22}{2} + \binom{21}{2}\right\} - \frac{1}{4}.$$

Итак, применяемый метод приближенного оценивания в случае отсутствия четкого предпочтения эквивалентен определению среднего значения, причем сначала мы полагаем, что (в данном случае $B \to Б$) соответствующие мнения характеризуются числами 25 и 21, а затем, что $B \leftarrow Б$ и число предпочтений соответственно равно 24 и 22; при этом мы не принимаем в расчет слагаемое, равное $-\frac{1}{4}$, но его влияние срав-

| | А | Б | В | Г | Д | Е | Ж | З | И | К | Л | М | Итого |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| А | — | 14 | 10 | 10 | 20 | 16 | 3 | 20 | 20 | 24 | $28\frac{1}{2}$ | 27 | $192\frac{1}{2}$ |
| Б | 32 | — | $24\frac{1}{2}$ | $26\frac{1}{2}$ | $30\frac{1}{2}$ | 25 | 0 | 35 | 28 | 30 | 34 | $32\frac{1}{2}$ | 298 |
| В | 36 | $21\frac{1}{2}$ | — | 21 | 40 | 33 | 0 | $35\frac{1}{2}$ | 36 | 32 | 37 | 28 | 320 |
| Г | 36 | $19\frac{1}{2}$ | 25 | — | $31\frac{1}{2}$ | 26 | 2 | 32 | 32 | 29 | 32 | $29\frac{1}{2}$ | $294\frac{1}{2}$ |
| Д | 26 | $15\frac{1}{2}$ | 6 | $14\frac{1}{2}$ | — | 23 | 1 | 27 | $28\frac{1}{2}$ | 26 | $25\frac{1}{2}$ | 23 | 216 |
| Е | 30 | 21 | 13 | 20 | 23 | — | 0 | 18 | $22\frac{1}{2}$ | 22 | 24 | $30\frac{1}{2}$ | 224 |
| Ж | 43 | 46 | 46 | 44 | 45 | 46 | — | 46 | 46 | 44 | 46 | $44\frac{1}{2}$ | $496\frac{1}{2}$ |
| З | 26 | 11 | $10\frac{1}{2}$ | 14 | 19 | 28 | 0 | — | 26 | 25 | $27\frac{1}{2}$ | 20 | 207 |
| И | 26 | 18 | 10 | 14 | $17\frac{1}{2}$ | $23\frac{1}{2}$ | 0 | 20 | — | 14 | 33 | 23 | 199 |
| К | 22 | 16 | 14 | 17 | 20 | 24 | 2 | 21 | 32 | — | 32 | $18\frac{1}{2}$ | $218\frac{1}{2}$ |
| Л | $17\frac{1}{2}$ | 12 | 9 | 14 | $20\frac{1}{2}$ | 22 | 0 | $18\frac{1}{2}$ | 13 | 14 | — | $26\frac{1}{2}$ | 167 |
| М | 19 | $13\frac{1}{2}$ | 18 | $16\frac{1}{2}$ | 23 | $15\frac{1}{2}$ | $1\frac{1}{2}$ | 26 | 23 | $27\frac{1}{2}$ | $19\frac{1}{2}$ | — | 203 |
| Итого | $313\frac{1}{2}$ | 208 | 186 | $211\frac{1}{2}$ | 290 | 282 | $9\frac{1}{2}$ | 299 | 307 | $287\frac{1}{2}$ | 339 | 303 | 3036 |

нительно невелико. Это согласуется с изложенным выше методом изучения случаев, когда существуют связи между рангами.

В нашей таблице в клетке, находящейся на пересечении строки *В* и столбца *Г*, стоит число 21. Оно равно $20\frac{2}{2}$, т. е. включает две дроби, каждая из которых равна $\frac{1}{2}$. Аналогично на пересечении столбца *В* и строки *Г* записано число, которое фактически составляло $24\frac{2}{2}$. Если бы и в этом случае нам потребовалось рассчитать соответствующие средние значения, то они были бы равны

$$\frac{1}{4}(20;\ 26)+\frac{1}{2}(21;\ 25)+\frac{1}{4}(22;\ 24).$$

Разность между этой суммой и числами, которые мы действительно использовали в расчетах (21; 25), составляет:

$$\frac{1}{8}\left\{\binom{20}{2}+\binom{26}{2}+\binom{22}{2}+\binom{24}{2}-2\binom{21}{2}-2\binom{25}{2}\right\}=\frac{1}{2}.$$

Такая разность вновь очень невелика.

**11.9.** Попытаемся найти способ для проверки статистической существенности. Для этого рассмотрим распределение и в том случае, когда все предпочтения носят случайный характер. Такие распределения построены для $m=3$ и $n=2\div 8$; $m=4$ и $n=2\div 6$; $m=5$ и $n=2\div 5$; $m=6$ и $n=2\div 4$. На основе этих расчетов исчислены данные, приведенные в приложении в табл. 10.

Для больших значений $m$ и $n$ вполне удовлетворительные результаты обеспечивает аппроксимация с помощью распределения $\chi^2$. В таком случае можно записать:

$$\chi^2=\frac{4}{m-2}\left\{\Sigma-\frac{1}{2}\binom{n}{2}\binom{m}{2}\frac{m-3}{m-2}\right\}, \tag{11.8}$$

$$\nu=\binom{n}{2}\frac{m(m-1)}{(m-2)^2} \tag{11.9}$$

и провести проверку статистической существенности по критериям распределения $\chi^2$ с $\nu$ степенями свободы. Можно ввести поправку на непрерывность, для этого из $\Sigma$ нужно соответственно вычитать единицу.

Например, при $m=3$, $n=8$ мы получили следующие значения $\chi^2$ и $\nu$:

$$\chi^2=4\sum,\quad \nu=168.$$

В приведенной в приложении табл. 10А мы находим:

$$\text{для } \sum=54,\quad P=0{,}011,$$

$$\text{для } \sum=58,\quad P=0{,}0011.$$

Соответствующие этим значениям $\Sigma$ (поправка на непрерывность введена) величины $\chi^2$ равны 212 и 228. При $\nu = 168$ мы можем полагать, что величина $\sqrt{2\chi^2}$ распределена по нормальному закону с единичной дисперсией, составляющей примерно $\sqrt{2\nu-1} = 18{,}30$, так что наши отклонения равны:

$$\sqrt{(2\times 212)} - 18{,}30 = 2{,}29$$

и

$$\sqrt{(2\times 228)} - 18{,}30 = 3{,}05.$$

С помощью приведенной в приложении табл. 3 можно установить, что такие отклонения соответствуют вероятностям 0,011 и 0,00114, т. е. полученные результаты весьма близки к истинным значениям.

Допустим теперь, что $m = 6$ и $n = 4$, тогда с помощью соотношений (11.8) и (11.9) можно аналогичным образом найти, что распределение величины ($\Sigma$ — 33,75) характеризуется 11, 25 степенями свободы. С помощью табл. 10D (см. приложение) мы можем определить, что 1%-ный уровень существенности предполагает значения $\Sigma$, лежащие между числами 59 и 60. Тогда (введя поправку на непрерывность) мы сможем отыскать соответствующие значения $\chi^2$, они равны 24, 25 и 25, 25. Обратившись к приведенной в приложении табл. 8, мы увидим, что эти значения действительно вполне соответствуют 1%-му уровню существенности, поскольку соответствующая величина $\chi^2$ при $\nu = 11, 25$ должна находиться в следующих границах: 24,725 ($\nu = 11$) и 26,217 ($\nu = 12$).

*Пример 11.5*

В примере 11.3 мы рассчитали значения $\Sigma = 9718$ и $u = 0{,}186$, соответствующие $n = 13$ и $m = 21$. Это свидетельствовало о том, что имеет место некоторая (хотя и не очень большая) согласованность предпочтений. Прибегнем теперь к проверке статистической существенности полученных величин.

С помощью формул (11.8) и (11.9) можно рассчитать значения $\chi^2$ и $\nu$:

$$\chi^2 = \frac{4}{21}\left\{9718 - \frac{1}{2}\,\frac{18}{19}\binom{13}{12}\binom{21}{2}\right\} = 421{,}4;$$

$$\nu = \binom{13}{2}\frac{21\times 20}{19^2} = 90{,}7;$$

$$\sqrt{(2\chi^2)} - \sqrt{(2\nu-1)} = 15{,}3.$$

Полученные величины значительно превосходят любые общепринятые стандарты статистической значимости. Из этого можно заключить, что коэффициент $u$ не мог бы принять значение, подобное тому, которое мы рассчитали, если бы совокупность всех высказываемых учениками предпочтений действительно содержала лишь случайные величины. Этот расчет еще раз подтверждает справедливость выводов, сформулированных на основе анализа примера 11.3.

Проделав аналогичные расчеты в примере 11.4, получаем:

$$\chi^2 = 754{,}5; \quad \nu = 70{,}6;$$

$$\sqrt{(2\chi^2)} - \sqrt{(2\nu - 1)} = 27{,}02.$$

И здесь полученные результаты свидетельствуют о том, что лишь с ничтожной вероятностью можно предполагать чисто случайный характер предпочтений. Поэтому можно сделать следующий вывод: рассчитанное значение коэффициента $u$ существенно.

## Библиография

См. [57], а также [65]. Парные сравнения (ненулевой случай) рассматриваются в [85] и [23].

В [23], кроме того, приводится уточнение предложенного Кендэлом и Бэбингтоном Смитом метода, с помощью которого проводится статистическая проверка существенности коэффициента $u$ в случае упорядоченности исходных данных. Если $M = \binom{m}{2}$ и $N = \binom{n}{2}$, то можно провести статистическую проверку существенности коэффициента $u$, используя для этого распределение $\chi^2$; в таком случае

$$\chi^2 = \frac{6(2n+5)MN}{(m-2)(2n^2+6n+17)}u + \nu$$

с $\nu = \dfrac{2(2n+5)^3 MN}{(m-2)^2(2n^2+6n+7)^2}$ степенями свободы.

См. также монографию H. A. D a v i d «The Method of Paired Comparisons» (Griffin and Co., London, 1963, 1969).

## *ГЛАВА 12.* ДОКАЗАТЕЛЬСТВО ТЕОРЕМ, ПРИВЕДЕННЫХ В ГЛАВЕ 11

**12.1.** Прежде всего докажем утверждение, согласно которому при полном попарном сравнении объектов максимальное число циклов предпочтений по трем объектам равно $\frac{1}{24}(n^3 - n)$, если $n$ нечетно, и $\frac{1}{24}(n^3 - 4n)$, если $n$ четно.

Рассмотрим многоугольник с $n$ вершинами, аналогичный тому, который приведен на рис. 11.1. Каждая вершина служит началом $n - 1$ линий. С помощью букв $a_1$, ..., $a_n$ обозначим количество стрелок, «выходящих» из каждой вершины. В таком случае

$$\sum_{j=1}^{n} a_j = \binom{n}{2} \qquad (12.1)$$

и среднее значение $a$ равно $\frac{1}{2}(n - 1)$.

Введем теперь функцию

$$T = \sum_{j=1}^{n} \left[a_j - \frac{1}{2}(n - 1)\right]^2 . \qquad (12.2)$$

Указанное выражение можно рассматривать как $n$ дисперсий величин $a$, поэтому его можно переписать следующим образом:

$$T = \sum a_j^2 - \frac{1}{4} n (n - 1)^2. \qquad (12.3)$$

**12.2.** Покажем теперь, что если направление предпочтения изменилось и если в результате этого количество циклов предпочтений по трем объектам увеличилось на величину $p$, то выражение $T$ уменьшится на $2p$, и наоборот. Рассмотрим, например, предпочтение $A \to B$. Допустим, что мы изменили направление этого предпочтения, и теперь $B \to A$; это повлияет только на те треугольники, которые содержат линию $AB$. Предположим, что общее число предпочтений типа $A \to X$ (включая $A \to B$) равно $\alpha$, а предпочтений типа $B \to X$ равно $\beta$. Тогда возможны 4 различных соотношения между тремя объектами:

$A \to X \leftarrow B$, предположим, что общее число таких соотношений равно $x$,

$A \leftarrow X \to B$,

$A \to X \to B$, число таких соотношений должно равняться $\alpha - x - 1$,

$A \leftarrow X \leftarrow B$, число таких соотношений должно равняться $\beta - x$. Когда мы обращаем предпочтение $A \to B$, два первых соотношения попрежнему остаются нецикличными, соотношения третьего вида становятся цикличными, а четвертый вид утрачивает цикличность. В таком случае увеличение общего числа циклов предпочтений по трем объектам составляет:

$$(\alpha - x - 1) - (\beta - x) = \alpha - \beta - 1 = p.$$

Величина $T$ уменьшится на величину

$$\alpha^2 - (\alpha - 1)^2 + \beta^2 - (\beta + 1)^2 = 2(\alpha - \beta - 1) = 2p.$$

Приведенные выше соображения нетрудно обобщить: ведь каждое последующее изменение индивидуальных предпочтений снова будет вызывать соответствующие изменения в величинах $T$ и $d$. Следовательно наше исходное положение справедливо.

12.3. Из определения $T$ следует, что это выражение достигает своего наибольшего значения, когда исходные данные упорядочены. Таким образом, максимальная величина $T$ равна $\frac{1}{12}(n^3 - n)$. Чему же равно его минимальное значение? Для того чтобы найти ответ, рассмотрим многоугольник с вершинами $A_1, \ldots, A_n$. Введем сначала предпочтения $A_1 \to A_2 \to \ldots \to A_n$, а затем предпочтения $A_1 \to A_3 \to A_5 \to$ . Допустим, что при этом мы не получили замкнутого контура, который позволял бы нам переходить из любой точки многоугольника к ближайшей несмежной вершине, скажем, к $A_k$, в таком случае введем предпочтения $A_k \to A_{k+2} \to$ и т. д. Затем аналогично установим предпочтения типа $A_1 \to A_4 \to A_7$ и т. д. Подобную процедуру будем продолжать до тех пор, пока вся система парных предпочтений не будет полностью завершена.

Допустим, что величина $n$ нечетна; тогда описанная система предпочтений будет содержать ряд находящихся внутри многоугольника циклических контуров. При этом каждый элемент $a$ равен $\frac{1}{2}(n-1)$ и, следовательно, $T = 0$; ясно, что здесь $T$ достигает своего минимального значения. Если же величина $n$ четна, то «завершающее» предпочтение $A_1 \to A_{\frac{1}{2}n+1}$ не образует замкнутого контура; оно представляет собой просто линию, соединяющую одну из вершин с симметричной ей (противоположной) вершиной. Построив систему предпочтений, мы обнаружим, что $\frac{1}{2}n$ вершин будет характеризоваться величинами $a$, равными $\frac{1}{2}n$, и $\frac{1}{2}n$ вершин — величинами $a$, равными $\frac{1}{2}n - 1$. В этом случае $T = \frac{1}{4}n$, здесь мы вновь имеем дело с минимальным значением $T$.

Таким образом, выражение $T$ может менять свое численное значение от 0 $\left(\text{или } \frac{1}{4}n\right)$ до $\frac{1}{12}(n^3-n)$. Если учесть к тому же, что увеличению величины $T$ на две единицы соответствует уменьшение $d$ на единицу, можно записать следующие максимальные значения, которые принимает $d$:

$$\frac{1}{24}(n^3-n), \text{ если } n \text{ нечетно;}$$

$$\frac{1}{24}(n^3-4n), \text{ если } n \text{ четно.}$$

**12.4.** В таблицах типа 11.1 числа $a$ представляют собой не что иное, как суммы чисел по строкам; поэтому, принимая во внимание сказанное выше, можно заключить, что общее число циклов предпочтений по трем объектам определяется из следующих соотношений:

$$d=\frac{1}{2}\left[\frac{1}{12}(n^3-n)-T\right]=\frac{1}{2}\left[\frac{1}{12}(n^3-n)+\frac{1}{4}n(n-1)^2-\sum a_j^2\right]=$$
$$=\frac{1}{12}n(n-1)(2n-1)-\frac{1}{2}\sum a_j^2=\frac{1}{6}n(n-1)(n-2)-$$
$$-\frac{1}{2}\sum a_j(a_j-1).$$

Приведенные соотношения соответствуют формулам (11.3) и (11.4).

**12.5.** Рассмотрим теперь аппроксимацию распределения $d$ с помощью распределения $\chi^2$ при больших величинах $n$.

Перенумеруем наши объекты числами от 1 до $n$. При этом будем полагать, что $P_{ijk}=1$, если предпочтения по трем объектам $(i, j, k)$ образуют цикл, и $P_{ijk}=0$ в противном случае.

Тогда

$$d=\sum P_{ijk}, \tag{12.4}$$

причем суммирование здесь проводится по всем возможным наборам из трех объектов. Таким образом,

$$E(d)=\binom{n}{3}E(P_{ijk}).$$

Перебирая различные соотношения, которые могут сложиться между предпочтениями в рамках одного набора из трех объектов, находим, что $E(P_{ijk})=\frac{1}{4}$. Следовательно, среднее значение, которое принимает $d$, равно

$$E(d)=\frac{1}{4}\binom{n}{3}. \tag{12.5}$$

Рассмотрим теперь $E(\Sigma P_{ijk})^2$. Записанная в этом выражении сумма включает $\binom{n}{3}$ слагаемых вида $P_{ijk}^2$; $3\binom{n}{3}\binom{n-3}{2}$ слагаемых вида

$P_{ijk}\,P_{ilm}$, где $j\neq l$, $k\neq m$; $3\binom{n}{3}(n-3)$ слагаемых вида $P_{ijk}\,P_{ijl}$, где $k\neq l$, и $\binom{n}{3}\binom{n-3}{3}$ слагаемых вида $P_{ijk}\,P_{lmn}$, где ни одна пара индексов не совпадает. Обратившись к рассмотрению соответствующих конфигураций, можно подсчитать, что в первом случае математическое ожидание будет составлять $\frac{1}{4}$, а в каждом из остальных случаев — $\frac{1}{16}$. Таким образом,

$$E\left(\sum P_{ijk}\right)^2=\frac{1}{16}\binom{n}{3}\left[4+\frac{3}{2}(n-3)(n-4)+3(n-3)+\binom{n-3}{3}\right]=$$
$$=\frac{1}{16}\binom{n}{3}\left[\binom{n}{3}+3\right].$$

Следовательно,

$$\mu_2(d)=E(d-\bar{d})^2=\frac{3}{16}\binom{n}{3}. \qquad (12.6)$$

Вычисление третьего и четвертого моментов — дело гораздо более сложное; однако в принципе общий ход рассуждений не меняется. В результате расчетов можно получить следующие выражения:

$$\mu_3=-\frac{3}{32}\binom{n}{3}(n-4); \qquad (12.7)$$

$$\mu_4=\frac{1}{55\,296}\binom{n}{3}[972n^3+972n^2-36\,936n+80\,352]. \qquad (12.8)$$

**12.6.** Для распределения $\chi^2$

$$dF\propto e^{-\frac{1}{2}\chi^2}\chi^{\nu-1}\,d\chi$$

первый, второй и третий моменты имеют следующие значения:

$$\begin{aligned}\mu_1^1 \text{ (относительно нуля)} &= \nu,\\ \mu_2 &= 2\nu,\\ \mu_3 &= 8\nu.\end{aligned}$$

Отметим, что $\mu_3(d)$ отрицателен, тогда как $\mu_3(\chi^2)$ положителен. Для того чтобы привести в соответствие оба распределения, мы будем отсчитывать величины $d$ «с другого конца» их распределения. Введя поправку на непрерывность, мы можем записать следующее выражение:

$$x=k\left[\frac{1}{4}\binom{n}{3}-d+\frac{1}{2}\right]+\nu. \qquad (12.9)$$

Среднее значение величин $x$ равно $\nu$, т. е. среднему из значений $\chi^2$. Величину $k$ будем выбирать так, чтобы наши распределения имели одинаковую дисперсию; такое равенство, очевидно, будет иметь место, когда

$$k=\sqrt{\frac{\mu_2(\chi^2)}{\mu_2(d)}}=\sqrt{\frac{2\nu}{\frac{3}{16}\binom{n}{3}}}. \qquad (12.10)$$

Для того чтобы совпадали третьи моменты, должно выполняться следующее соотношение:

$$\frac{\frac{3}{32}\binom{n}{3}(n-4)}{\left[\frac{3}{16}\binom{n}{3}\right]^{\frac{3}{2}}}=\frac{8\nu}{(2\nu)^{\frac{3}{2}}}.$$

Преобразовав это выражение, получаем:

$$\nu=\frac{n(n-1)(n-2)}{(n-4)^2}. \qquad (12.11)$$

Полученное значение $\nu$ подставим в соотношение (12.10); в таком случае первые три момента для

$$\frac{8}{n-4}\left[\frac{1}{4}\binom{n}{3}-d+\frac{1}{2}\right]+\nu \qquad (12.12)$$

совпадают с соответствующими моментами распределения $\chi^2$.

12.7. Можно показать, что распределение величин $d$ стремится к нормальному, когда $n\to\infty$. Общая последовательность доказательства совпадает с аналогичными рассуждениями, приведенными в гл. 5 — подробности можно найти в [65]. В этой главе мы покажем, что нечетные моменты представляют собой величины меньшего порядка относительно $n$, чем четные. Так, в выражении

$$(\sum Q_{ijk})^{2m},$$

где $Q_{ijk}=P_{ijk}-\frac{1}{4}$, доминирующее слагаемое, имеющее вид $Q^2_{ijk}\,Q^2_{lmn}\ldots$, содержит $m$ множителей; математическое ожидание этих величин равно $\frac{3}{32}$, а частота составляет $\frac{(2m)!}{2^m m!}$. Таким образом, для доминирующего слагаемого можно записать:

$$\mu_{2m}\sim\frac{(2m)!}{2^m\,m!}\left(\frac{3}{32}\right)^m n^{3m}\sim\frac{(2m)!}{2^m\,m!}(\mu_2)^m,$$

а отсюда следует, что распределение стремится к нормальному, если $n$ неограниченно возрастает.

12.8. Докажем, наконец, возможность использования распределения $\chi^2$ в качестве аппроксимирующего при проверке статистической значимости коэффициента согласия $u$ в том случае, когда все предпочтения складываются случайным образом.

Включение в $\Sigma$ чисел, содержащихся в двух клетках (строка $X$ и столбец $Y$, строка $Y$ и столбец $X$), изменяет величину $\Sigma$ на

$$\binom{\gamma}{2}+\binom{m-\gamma}{2}. \tag{12.13}$$

Всего насчитывается $2^m$ способов размещения $m$ предпочтений по клеткам, в том числе имеется $\binom{m}{\gamma}$ вариантов, при которых в клетке содержится $\gamma$ единиц. Следовательно, частота, характеризующая появление того или иного слагаемого в выражении $\Sigma$, совпадает с соответствующим коэффициентом при $t$ в выражении

$$f=t^{\binom{m}{2}}+\binom{m}{1}t^{\binom{m-1}{2}}+\ldots+\binom{m}{\gamma}t^{\binom{m-\gamma}{2}+\binom{\gamma}{2}}+\ldots+t^{\binom{m}{2}}. \tag{12.14}$$

Далее, поскольку предпочтения складываются случайным образом, входящие в состав $\Sigma$ слагаемые, которые соответствуют $\binom{n}{2}$ парам клеток, независимы друг от друга. Следовательно, распределение величин $\Sigma$ задается функцией

$$f^{\binom{n}{2}}, \tag{12.15}$$

причем частота $\Sigma$ совпадает с коэффициентом при $t^{\Sigma}$ в этом выражении. Пусть, например, $m=3$, $n=4$. Тогда

$$f=t^3+3t+3t+t^3=2t\,(3+t^2)$$

и распределение можно охарактеризовать следующим образом:

$$[2t\,(3+t^2)]^6=2^6t^6\,(729+1458t^2+1215t^4+540t^6+135t^8+18t^{10}+t^{12})$$

с общей частотой, равной $2^6\times4^6$. С помощью подобных расчетов исчислены приведенные в приложении данные (табл. 10).

**12.9.** Если величина $m$ постоянна, то распределение $\chi^2$ с ростом $n$ будет стремиться к нормальному: ведь оно представляет собой среднее арифметическое из $\binom{n}{2}$ составляющих, характеризующихся конечными равными моментами [1]. При постоянной величине $n$ с увеличением $m$ распределение стремится к одной из форм распределения $\chi^2$, а так как каждой из $\binom{n}{2}$ клеток в выражении $\Sigma$ соответствует слагаемое, образуемое значением переменной $\binom{\gamma}{2}+\binom{m-\gamma}{2}$, которая распределена подобно $\gamma^2$, а распределение $\gamma$ стремится к нормальному.

[1] Здесь рассматривается по существу частный случай центральной предельной теоремы [58,1, русск. перев. стр. 268—270].

$r$-й момент величины $\Sigma$, измеряемый относительно начала отсчета, может быть выражен следующим образом:

$$2^{m\binom{n}{2}}\mu_r' = \left[\left(t\frac{\partial}{\partial t}\right)^r f^{\binom{n}{2}}\right]_{t=1}. \tag{12.16}$$

Действительно, при дифференцировании слагаемое, содержащее $t^{\Sigma}$ умножается на $\Sigma^r$; если же мы положим $t = 1$, то наш ряд будет представлять собой сумму частот, каждая из которых умножается на $\Sigma^r$, т. е. перед нами не что иное, как $r$-й момент $\Sigma$. Исчислим таким способом первый момент:

$$2^m\mu_1' = \binom{n}{2}\sum_{j=0}^{m}\binom{m}{j}\left[j^2 - mj + \frac{1}{2}(m^2-m)\right] =$$

$$= \binom{n}{2}\left[2^m\binom{m}{2} + \sum\binom{m}{j}(j^2 - jm)\right]$$

и, следовательно,

$$\mu_1' = \frac{1}{2}\binom{m}{2}\binom{n}{2}. \tag{12.17}$$

Аналогично (промежуточные вычисления опущены) мы можем найти:

$$\mu_2 = \frac{1}{4}\binom{n}{2}\binom{m}{2}, \tag{12.18}$$

$$\mu_3 = \frac{3}{4}\binom{n}{2}\binom{m}{3}, \tag{12.19}$$

$$\mu_4 = \binom{m}{2}\binom{n}{2}\left[\frac{3m^2 - 15m + 17}{8} + \frac{3}{32}\binom{n}{2}m(m-1)\right]. \tag{12.20}$$

Пользуясь методом, изложенным в 12.6, мы выясняем, что

$$\chi^2 = \frac{4}{m-2}\left[\Sigma - \frac{1}{2}\binom{n}{2}\binom{m}{2}\frac{m-3}{m-2}\right]. \tag{12.21}$$

Эта величина характеризуется обычным распределением с

$$\nu = \binom{n}{2}\frac{m(m-1)}{(m-2)^2} \tag{12.22}$$

степенями свободы.

## Библиография

См. библиографию к гл. 11.

## *ГЛАВА 13.* НЕКОТОРЫЕ ДАЛЬНЕЙШИЕ ПРИЛОЖЕНИЯ

**13.1.** В данной, заключительной главе необходимо кратко обрисовать ряд недавних достижений в теории ранжирования. Некоторые из идей и методов, которые будут здесь затронуты, являются новыми и еще не полностью разработаны. Развитию других препятствует сложность возникающих математических проблем. Поэтому данная глава, в отличие от предыдущих, не содержит систематического введения в раздел статистической теории, который исследован в такой степени, что может быть применен на практике в качестве рутинной процедуры. Скорее это обзор последних достижений и руководство для дальнейшего чтения.

### Ситуации, требующие применения рангов

**13.2.** До сих пор мы имели дело главным образом с применением рангов в ситуациях, в которых не было генеральной корреляции (так называемый нулевой случай), или с одной особенной моделью (5.17), в которой наши наблюдения рассматривались как извлечение из исходной последовательности большего объема. В некоторых условиях ни один из этих подходов не является пригодным. Например, пусть нам известно, что имеется некоторая исходная взаимосвязь, тогда больший интерес представляет характеристика ее степени, чем выявление факта существования. В иных случаях мы можем изучать генеральную совокупность последовательностей, имеющих одинаковую протяженность, причем некоторые из них представлены в выборке.

**13.3.** В двух ситуациях, имеющих довольно общий характер, был достигнут некоторый прогресс.

*а. Модель Стюарта*

Стюарт [92] изучил проблемы выборки из группы экспертов. Предположим, что каждый член группы располагает $n$ объектов в порядке предпочтения; например, совокупность видов деятельности может быть упорядочена в соответствии с общественным престижем каждого из них. Если случайным образом из множества $M$ отобрано $m$ членов, то получим $m$ совокупностей последовательностей, состоящих из $n$ объектов. Наша проблема теперь заключается не в испытании независимости (подобно тому, как это изложено в гл. 6), хотя эта проблема также возникает, а главным образом в том, чтобы увидеть, насколько подобная общность предпочтений, существующая в $m$ наблюденных членах, может быть принята в качестве представителя предпочтений множества $M$.

**13.4.** При исследовании этой проблемы Стюарт применил коэффициент конкордации рангов $W$. В ненулевом случае распределение подобной функции весьма сложно и для оценивания даже первых четырех моментов требует трудоемких расчетов. Однако затраты труда не являются непреодолимым препятствием, и соответствующие результаты были успешно использованы для решения практических задач, связанных с социальными группировками. Так, в результате опроса примерно 1000 человек, охваченных выборкой, были получены ранги, характеризующие 30 видов занятий в соответствии с их неопределимым качественным признаком, отражающим «общественное положение». Далее был подсчитан средний ранг, приписанный каждому занятию. Возникает вопрос: насколько существенно расхождения в выборочных средних рангах отражают реальные расхождения в совокупности? Подробности об этом можно найти в работе Стюарта.

**13.5.** В этой связи можно упомянуть об одной важной проблеме в теории конкордации. Предположим, что у нас имеется два набора, содержащих $m_1$ и $m_2$ последовательностей. Испытания, рассмотренные в гл. 6, могут показать, что коэффициенты конкордации каждого из них, скажем $W_1$ и $W_2$, являются существенными. Как нам проверить, является или нет разность $W_1 — W_2$ существенной, или, что более обще, может ли какой-либо метод быть использован в качестве инструмента проверки того, существенно или нет степень согласованности между $m_1$ последовательностями отличается от степени согласованности между $m_2$ последовательностями? Представляется, что по этой проблеме еще ничего не сделано.

*б. Модель Тэрстоуна—Мостеллера—Дэниелса.*

**13.6.** Допустим, что одна последовательность определяется на основе объективного критерия, например, объекты могут быть упорядочены во времени или в пространстве в соответствии с некоторым измеримым критерием. Отдельный эксперт, не зная истинного порядка объектов, выносит суждения о них и приписывает им некоторый порядок. Совокупность значений, с которой теперь мы имеем дело, представляет собой (бесконечное) множество последовательностей, которое мог воспроизвести эксперт при повторении опыта; элементы этого множества различаются между собой, поскольку он мог допускать ошибки или неправильные суждения, или, более обще, последовательности рангов могут быть воспроизведены разными экспертами. Модель, рассмотренная в [11], предполагает наличие непрерывной переменной $y$ в качестве источника отклонения от истинного значения $m$; функция плотности последовательности $y_1, \dots, y_n$ есть

$$f(y_1-m_1)f(y_2-m_2)\dots f(y_n-m_n), \qquad (13.1)$$

и проблема заключается в оценивании значения $m$ или по крайней мере определении их порядка на основе наблюдений. Таким образом, возникает проблема, связанная с регрессией, и Дэниелс разработал ее значительно дальше, чем можно было ожидать.

## Модели парных сравнений

**13.7.** Весьма обобщенные модели могут быть разработаны для парных сравнений. Однако их практическое применение не легко осуществить. Например, предположим, что у нас имеется $n$ объектов и вероятность того, что объекту $i$ будет приписан более высокий ранг, чем объекту $j$, равна $\pi_{ij}$. Такая модель была бы пригодна для ситуаций, в которых эксперт выражает предпочтение, причем оно может быть ошибочным, как и в модели последовательности 13.6 или в случае, когда ряд экспертов высказывает различные предпочтения. Такая модель предложена в [23]. Коэффициент согласованности $u$ (11.8) имеет среднее значение, определяемое как

$$E(u)=1-\frac{4}{m(m-1)}\sum_{i<j}^{n}\pi_{ij}(1-\pi_{ij}), \tag{13.2}$$

и дисперсию

$$\operatorname{var} u=\frac{64}{m(m-1)n^2(n-1)^2}\sum_{i<j}\{(m-1)[\pi_{ij}(1-\pi_{ij})]-4\pi_{ij}^2(1-\pi_{ij})^2\}. \tag{13.3}$$

Трудность применения данного результата заключается в том, что мы не знаем значений $\pi$.

**13.8.** Другого рода подход развит в [3]. Предположим, что $n$ объектов имеют «истинную» последовательность, охарактеризованную числами $\pi_1, \ldots, \pi_n$ (причем $\sum_{j=1}^{n}\pi=1$). Когда объекты $i$ и $j$ сравниваются, то вероятность того, что объект $i$ предпочитается $j$, равна $\pi_i/(\pi_i+\pi_j)$. Данная модель, уже заслужившая определенное доверие во многих предшествующих работах, связанных с парными сравнениями, проще, чем предыдущая, поскольку в ней представлено $n$ параметров $\pi_i$, вместо $\frac{1}{2}n(n-1)$ параметров $\pi_{ij}$. В [3] разработан метод оценивания параметров $\pi$ на основе наблюдений и проверки существенности полученных результатов. Были опубликованы таблицы, необходимые для практического приложения данного метода (см. [3] и [4]).

## Ранговые критерии стандартных статистических ситуаций

**13.9.** Идеи, лежащие в основе формирования ранговых коэффициентов, особенно коэффициентов $\tau$, могут быть приложимы к ряду ситуаций, с которыми часто сталкиваются в статистическом анализе. Можно выделить следующие пять случаев.

*а. Проверка независимости.* В случае, когда имеется $n$ пар наблюдений $(x_1, y_1), \ldots, (x_n, y_n)$, может возникнуть необходимость проверки независимости распределения $x$ от распределения $y$. Для этого может быть применен простой критерий, основанный на $\tau$ или $\rho$. Более сложный критерий разработан в [39].

*б. Проверка однородности двух выборок.* Критерий, приведенный в 3.12, был разработан в [116] и [64] исходя из разных соображений. Обычно он называется тестом Уилкоксона. Он эффективен при проверке однородности двух выборок путем рассмотрения доли случаев, в которых значение членов одной выборки превышает значение членов другой. Эта проверка формально может быть связана с проверкой $\tau$ для дихотомии.

*в. Проверка однородности для $k$ выборок.* Метод проверки для двух выборок не может быть с легкостью распространен на случай с числом выборок выше двух. В [102] и [46] некоторым образом обобщен критерий Уилкоксона. В основу полученного критерия положена величина $\tau$, учитывающая связи в одной последовательности. В [60] также рассмотрен этот случай и сделан обзор имевшейся в то время литературы.

*г. Проверка на наличие тренда во временных рядах.* Совокупность наблюдений, упорядоченных во времени, может быть проверена на наличие тренда путем анализа корреляции между рядом рангов наблюденных величин и показателями времени (см. пример 1.1). В [63] изучены две возможности такого анализа и оценена их относительная эффективность. В [94] и [95] рассмотрены другие подходы. См. также [8].

*д. Подбор функциональных взаимосвязей.* Одна из наиболее трудных проблем статистики возникает при подгонке функциональных взаимосвязей к переменным, содержащим ошибки. Классические методы решения этой проблемы предполагают принятие определенных допущений относительно дисперсии ошибок. Тейл [104] показал, что эта проблема может быть успешно разрешена без таких допущений с помощью методов, базирующихся на свойствах, связанных с порядком расположения величин.

Довольно полное изложение всех этих методов приводится в [58, гл. 29 и 31].

## Преобразование парных сравнений в ранги

**13.10.** В некоторых ситуациях, когда проводятся многочисленные попарные сравнения, возникает необходимость обобщения окончательных результатов в форме последовательностей. Например, при исследовании вкусов потребителей может оказаться желательным в конце концов ранжировать множество объектов в порядке общего предпочтения. В [112] и [54] обсужден метод, который можно было применить в данном случае. Проблема сокращения числа парных сравнений требует для своего решения применения сбалансированных схем экспериментов*.

---

* Описание сбалансированных схем проведения эксперимента читатель может найти в работах, посвященных теории планирования экспериментов. См., например, Финни Д. Введение в теорию планирования экспериментов (Перев. с англ.) М., «Наука», 1970. — *Прим. ред.*

## Эффективность и мощность ранговых методов

13.11. Рамки настоящей книги не позволяют охватить вопросы, связанные с эффективностью и мощностью ранговых методов в свете современной теории проверки гипотез, за исключением эпизодического рассмотрения относительной эффективности $\tau$ и $\rho$ при оценивании параметров корреляции в двумерных нормальных совокупностях. В последние годы в этой области предприняты большие работы. Вообще говоря, метод оценивания или проверки, использующий ранги, выигрывает в общности, поскольку он не зависит от формы распределения в генеральной совокупности, однако он может быть связан с потерями эффективности и мощности. Это вполне естественно. Средство, употребляемое для многих целей, обычно не так эффективно, как специально созданное ради одной цели. С другой стороны, существуют ситуации, когда мощность проверок, основанных на рангах, удивительно высока и потери от применения ранговых методов незначительны. О некоторых результатах в данной области см. [58, гл. 31].

13.12. Библиография, которая следует далее (см. с. 182), охватывает соответствующую область достаточно полно, однако читатель, который желает пойти дальше в изучении данного предмета, может получить дополнительные сведения в [106] и [80].

## Ранжирование и измерение

13.13. Автором разработан метод определения шкал измерения в тех случаях, когда не только объекты, но и различия между ними могут быть ранжированы [*Biometrika*, *July*, 1962]. Некоторые связанные с этим задачи были исследованы Дэниелсом и Пэрри (там же). Все же проблема еще не полностью разработана, однако возможности в этом отношении представляются интересными.

## БИБЛИОГРАФИЯ


1. Benard A., and van Elteren Ph. (1953). A generalization of the method of *m* rankings. *Kon. Ned. Ak. Wet.*, A, 56, No. 4 and *Indag. Math.*, **15**, No. 4.
2. Bose R. C. (1956). Paired comparison designs for testing concordance between judges. *Biometrika*, **43**, 113.
3. Bradley R. A., and Terry M. E. (1952). The rank analysis of incomplete block designs. **I.** The method of paired comparisons. *Biometrika*, **39**, 324.
4. Bradley R. A. (1954). The rank analysis of incomplete block designs. II. Additional tables for the method of paired comparisons. *Biometrika*, **41**, 502.
5. Bradley R. A. (1955). Rank analysis of incomplete block designs. III. Some large sample results on estimation and power for a method of paired comparisons. *Biometrika*, **42**, 450.
6. Burr E. J. (1960). The distribution of Kendall's score $S$ for a pair of tied rankings. *Biometrika*, 47, 151.
7. Chambers E. G. (1946). Statistical techniques in applied psychology. *Biometrika*, 33, 269.
8. Cox D. R., and Stuart A. (1955). Some quick sign tests for trend in location and dispersion. *Biometrika*, 42.
9. Daniels H. E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika*, 33, 129.
10. Daniels H. E. (1948). A property of rank correlations. *Biometrika*, 35, 416.
11. Daniels H. E. (1950). Rank correlation and population models. *J. Roy. Statist. Soc.*, B, 12, 171.
12. Daniels H. E. (1951). Note on Durbin and Stuart's formula for $E(r_s)$. *J. Roy. Statist. Soc.*, B, **13**, 310.
13. Daniels H. E. and Kendall M. G. (1947). The significance of rank correlations where parental correlation exists. *Biometrika*, **34**, 197.
14. Dantzig G. B. (1939). On a class of distributions that approach the normal distribution function. *Ann. Math. Statist.*, **10**, 247.
15. David F. N., and Mallows C. L. (1961). The variance of Spearman's rho in normal samples. *Biometrika*, **48**, 19.
16. David S. T., and others (1951). Some questions of distribution in the theory of rank correlation. *Biometrika*, **38**, 131.
17. Dixon W. J. (1953). Power functions of the sign test and power efficiency for normal alternatives. *Ann. Math. Statist.*, **24**, 467.
18. Dubois P. (1939). Formulas and tables for rank correlation. *Psych. Rec.*, **3**, 46.
19. Durbin J. (1951). Incomplete blocks in ranking experiments. *Brit. J. Psych.* (Stat. Section), **4**, 85.
20. Durbin J., and Stuart A. (1951). Inversions and rank correlation coefficients. *J. Roy. Statist. Soc.*, B, **13**, 303.

21. Dykstra O. (1956). A note on the rank analysis of incomplete block designs, applications beyond the scope of existing tables. *Biometrics*, **12**, 301.
22. Eelis W. C. (1929). Formulas for probable errors of coefficients of correlation. *J. Am. Stat. Ass.*, **24**, 170.
23. Ehrenberg A. S. C. (1952). On sampling from a population of rankers. *Biometrika*, **39**, 82.
24. Esscher F. (1924). On a method of determining correlation from the ranks of variates. *Skand. Akt.* **7**, 201.
25. Farlie D. J. G. (1963). The asymptotic efficiency of Daniels's generalised correlation coefficient. *Biometrika*, **50**, 499.
26. Feller W. (1945). Limit theorems in probability. *Bull. Am. Math. Soc.*, **51**, 800.
27. Fieller E. C., Hartley H. O., and Pearson, E. S. (1957). Tests for rank. correlation coefficients: I. *Biometrika*, **44**, 470.
28. Fieller E. C., and Pearson E. S. (1961). Tests for rank correlation coefficients: II. *Biometrika*, **48**, 29.
29. Friedman M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Am. Stat. Ass.*, **32**, 675.
30. Friedman M. (1940). A comparison of alternative tests of significance for the problem of *m* rankings. *Ann. Math. Statist.*, **11**, 86.
31. Glasser G. J., and Winter R. F. (1961). Critical values of the coefficient of rank correlation for testing the hypothesis of independence. *Biometrika*, **48**, 444.
32. Goodman L. A., and Kruskal W. H. (1954). Measures of association for cross classification. J. *Am. Stat. Ass.*, **49**, 732.
33. Greiner R. (1909). Uber das Fehlersystem der Kollektivmasslehre. *Zeitschrift für Mathematik und Physik*, **57**, 121, 225 and 337.
34. Griffin H. D. (1958). Graphic computation of tau as a coefficient of disarray. *J. Am. Stat. Ass.*, **53**, 441.
35. Guttman L. (1946). An approach for quantifying paired comparisons and rank order. *Ann. Math. Statist.*, **17**, 144.
36. Haden H. G. (1947). A note on the distribution of the different orderings of *n* objects. *Proc. Camb. Phil. Soc.*, **43**, 1.
37. Harter H. L. (1961). Expected values of normal order statistics. *Biometrika*, **48**, 151.
38. Hoeffding W. (1947). On the distribution of the rank correlation coefficient $\tau$ when the variates are not independent. *Biometrika*, **34**, 183.
39. Hoeffding W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.*, **19**, 293.
40. Hoeffding W. (1948). A non-parametric test of independence. *Ann. Math. Statist.*, **19**, 546.
41. Hoeffding W. (1951). Optimum non-parametric tests. *Second Berkeley Symposium*, University of California Press.
42. Hoeffding W. (1952). The large sample power of tests based on permutations of observations. *Ann. Math. Statist.*, **23**, 169.
43. Hoflund O. (1963). Simulated distributions for small *n* of Kendall's partial rank correlation coefficient. *Biometrika*, **50**, 520.
44. Hotelling H., and Pabst M. R. (1936). Rank correlation and tests of significance involving no assumptions of normality. *Ann. Math. Statist.*, **7**, 29.
45. Irwin J. O. (1925). The further theory of Francis Galton's individual difference problem. *Biometrika*, **17**, 100.
46. Jonkheere A. R. (1954). A distribution-free *k*-sample test against ordered alternatives. *Biometrika*, **41**, 133.
47. Kaarsemaker L., and Van Wijngaarden A. (1952). *Tables for use in rank correlation*. Report R 73, Computation Mathematical Centre, Amsterdam.
48. Kendall M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**, 81.

49. Kendall M. G. (1942a). Partial rank correlation. *Biometrika*, **32**, 277.
50. Kendall M. G. (1942б). Note on the estimation of a ranking. *J. Roy. Statist. Soc.*, **105**, 119.
51. Kendall M. G. (1945). The treatment of ties in ranking problems. *Biometrika*, **33**, 239.
52. Kendall M. G. (1947). The variance of τ when both rankings contain ties. *Biometrika*, **34**, 297.
53. Kendall M. G. (1949). Rank and product-moment correlation. *Biometrika*, **36**, 177.
54. Kendall M. G. (1955) Further contributions to the theory of paired comparisons. *Biometrika*, **11**, 43.
55. Kendall M. G., and others (1939). The distribution of Spearman's coefficient of rank correlation, etc. *Biometrika*, **30**, 251.
56. Kendall M. G., and Babington Smith B. (1939). The problem of *m* rankings. *Ann. Math Statist.*, **10**, 275.
57. Kendall M. G., and Babington Smith B. (1940). On the method of paired comparisons. *Biometrika*, **31**, 324.
58. Kendall M. G., and Stuart A. (1958, 1961). *The Advanced Theory of Statistics*, Vols. 1 and **2**. London: Charles Griffin & Co. Русский перевод: Кендалл М. Д., Стьюарт А. «Теория распределений». М., «Наука», 1966 и Кендалл М. Д., Стьюарт А. «Статистические выводы и связи». М., «Наука», 1973.
59. Kruskal W. H. (1952). A non-parametric test for the several sample problem. *Ann. Math. Statist.*, **23**, 525.
60. Kruskal W. H., and Wallis W. A. (1952). Use of ranks in one-criterion variance-analysis. *J. Am. Stat. Ass.*, **47**, 583.
61. Lehmann E. L. (1951). Consistency and unbiasedness of certain non-parametric tests. *Ann. Math. Statist.*, **22**, 155.
62. Lehmann E. L. (1953). The power of rank tests. *Ann. Math. Statist.*, **24**, 23.
63. Mann H. B. (1945). Non-parametric tests against trend. *Econometrica*, **13**, 245.
64. Mann H. B., and Whitney D. R. (1947). On a test of whether one of two random variables is larger than the other. *Ann. Math. Statist.*, **15**, 50.
65. Moran P. A. P. (1947). On the method of paired comparisons. *Biometrika*, **34**, 363.
66. Moran P. A. P. (1948). Rank correlation and permutation distributions. *Proc. Camb. Phil. Soc.*, **44**, 142.
67. Moran P. A. P. (1948). Rank correlations and product-moment correlation. *Biometrika*, **35**, 203.
68. Moran P. A. P. (1950). Recent developments in ranking theory. *J. Roy. Statist. Soc.*, B, **12**, 153.
69. Moran P. A. P. (1950). A curvilinear ranking test. *J. Roy. Statist. Soc.*, **12**, 292.
70. Moran P. A. P. (1951). Partial and multiple rank correlation. *Biometrika*, **38**, 26.
71. Mosteller F. (1951). Remarks on the method of paired comparisons. *Psychometrika*, **15**, 3; 16, 203; **16**, 207.
72. Muhsam H. V. (1954). A probability approach to ties in rank correlation. *Bull. Res. Council of Istrael*, **3**, 321.
73. Olds E. G. (1938). Distribution of sums of squares of rank differences for small numbers of individuals. *Ann. Math. Statist.*, **9**, 133.
74. Pearson E. S., and Snow Barbara A. S. (1962). Tests for rank correlation coefficients. III: Distribution of the transformed Kendall coefficient. *Biometrika*, **49**, 185.
75. Pearson K. (1907). Mathematical contributions to the theory of evolution. XVI. On further methods of determining correlation. *Drapers' Co. Res. Mem., Biometric Series IV*. Cambridge University Press.

76. Pearson K. (1914, 1921). On an extension of the method of correlation by grades or ranks. *Biometrika*, **10**, 416, and Second note, *ibid.*, **13**, 302.
77. Pearson K., and Pearson M. V. (1931, 1932). On the mean character and variance of a ranked individual, and on the mean and variance of the intervals between ranked individuals. *Biometrika*, **23**, 364, and **24**, 203.
78. Pitman E. J. G. (1937, 1938). Significance tests which may be applied to samples from any populations. *Supp. J. R. Stat. Soc.*, **4**, 119. II. The correlation coefficient test. *Supp. J. R. Stat. Soc.*, **4**, 225. III. The analysis of variance test. *Biometrika*, **29**, 332.
79. Savage I. R. (1952). Bibliography of non-parametric statistics and related topics. *National Bureau of Standards, Washington D. C., Report* 182.
80. Savage I. R. (1954). Contributions to the theory of rank order statistics. *Memorandum for private circulation.*
81. Shah S. M. (1961). A note on Griffin's paper «Graphic computation of tau as a coefficient of disarray». *J. Am. Stat. Ass,*, **56**, 736.
82. Sillitto G. P. (1947). The distribution of Kendall's $\tau$ coefficient of rank correlation in rankings containing ties. *Biometrika*, **34**, 36.
83. Silverstone H. (1950). A note on the cumulants of Kendall's S-distribution. *Biometrika*, **37**, 231.
84. Slater P. (1961). Inconsistencies in a schedule of paired comparisons. *Biometrika*, **48**, 303.
85. Smith B. Babington (1950). Discussion of Professor Ross's paper. *J. Roy. Statist. Soc.*, B, **12**, 54.
86. Snow Barbara A. S. (1962). The third moment of Kendall's tau in normal samples. *Biometrika*, **49**, 177.
87. Snow Barbara A. S. (1963). The distribution of Kendall's tau for samples of four from a normal bivariate population with correlation *p*. *Biometrika*, **50**, 538.
88. Spearman C. (1904). The proof and measurement of association between two things . *Am. J. Psych.*, **15**, 88.
89. Spearman C. (1906). A footrule for measuring correlation. *Brit. J. Psych.*, **2**, 89.
90. Spearman C. (1910). Correlation calculated from faulty data. *Brit. J. Psych.*, **3**, 271.
91. Starks T. H. and David H. A. (1961). Significance tests for paired-comparison experiments, *Biometrika*, **48**, 95.
92. Stuart A. (1951). An application of the distribution of the ranking concordance coefficient. *Biometrika*, **38**, 33.
93. Stuart A. (1953). The estimation and comparison of strengths of association in contingency tables. *Biometrika*, **40**, 105.
94. Stuart A. (1954*a*). The correlation between variate-values and ranks in samples from a continuous distribution. *Brit. J. Psych.* (Stat. Section), **7**, 37.
95. Stuart A. (1954*b*). The asymptotic relative efficiency of distribution-free tests of randomness against normal alternatives. *J. Am. Stat. Ass.*, **49**, 147.
96. «Student» (1921). An experimental determination of the probable error of Dr. Spearman's correlation coefficient. *Biometrika*, **13**, 263.
97. Sundrum R. M. (1953). *Theory and Applications of Distribution-Free Methods in Statistics*, Ph. D. Thesis, London University.
98. Sundrum R. M. (1953*a*). Moments of the rank correlation coefficient $\tau$ in the general case. *Biometrika*, **40**, 409.
99. Sundrum R. M. (1953*b*). A method of systematic sampling based on order properties. *Biometrika*, **40**, 452.
100. Sundrum R. M. (1953*c*). The power of Wilcoxon's two-sample test. *J. Roy. Soc.*, B, **15**, 246.
101. Teegarden K. L. (1960). Critical values for rank order correlations. *Industrial Quality Control*, **16**, 11, 48.

102. T e r p s t r a T. J. (1952). The asymptotic normality and consistency of Kendall's test against trend when ties are present in one ranking. *Proc. Kon. Ned. Ak. Wet.*, A, 55, *Indag. Math.*, **14**, 327.
103. T e r r y M. E. (1952). Some rank order tests which are most powerful against specific alternatives. *Ann. Math. Statist.*, **23**, 346.
104. T h e i l H. (1950). A rank invariant method of linear and polynomial regression analysis. *Proc. Kon. Ned. Ak. v. Wet.*, **53**, 386, 521, 1397.
105. V a n D a n t z i g D. (1951). On the consistency and the power of Wilcoxon's two-sample test. *Proc. Kon. Ned. Ak. v. Wet.*, **54**, 172.
106. V a n D a n t z i g D. and H e m e l r i j k J. (1954). Statistical methods based on few assumptions. *Bull. Int. Stat. Inst.*, **34**, 2me livraison, 239.
107. V a n d e r V a a r t H. R. (1950). Some remarks on the power function of Wilcoxon's test for the problem of two samples. *Proc. Kon. Ned. Ak. v.* Wet., **53**, 494 and 507.
108. V a n d e r W a e r d e n B. L. (1952). Order tests for the two-sample problem and their power. *Proc. Kon. Ned. Ak. v. Wet.*, **55**, 453 (Corrigenda, 56, 80).
109. V a n E l t e r e n Ph., and N o e t h e r G. E. (1959). The asymptotic efficiency of the $\chi_r^2$ test for a balanced incomplete block design. *Biometrika*, **46**, 475.
110. W a l l i s W. A. (1939). The correlation ratio for ranked data. *J. Am. Stat. Ass.*, **34**, 533.
111. W a t k i n s G. P. (1933). An ordinal index of correlation. *J. Am. Stat. Ass.*, **28**, 139.
112. W e i T. H. (1952). The Algebraic Foundations of Ranking Theory. Ph. D. thesis. Cambridge University.
113. W e l c h B. L. (1937). On the *z*-test in randomised blocks and Latin Squares. *Biometrika*, **29**, 21.
114. W h i t f i e l d J. W. (1947). Rank correlation between two variables, one of which is ranked, the other dichotomous. *Biometrika*, **34**, 292.
115. W h i t f i e l d J. W. (1950). Uses of the ranking method in psychology *J. Roy. Statist. Soc.*, **12**, 163.
116. W i l c o x o n F. (1945). Individual comparisons by ranking methods. *Biometriks*, **1**, 80.
117. W i l k i n s o n J. L. (1957). An analysis of paired comparison designs with incomplete repetitions. *Biometrika*, **44**, 97.
118. W o o d b u r y M. A. (1940). Rank correlation when there are equal variates. *Ann. Math. Statist.*, **11**, 358.

## ПРИЛОЖЕНИЯ

Таблица 1. **Вероятность того, что $S$ (для $\tau$) достигнет или превзойдет заданное значение** (показаны только положительные величины; отрицательные величины определяются по симметрии)

| $S$ | Значения $n$ | | | | $S$ | Значения $n$ | | |
|---|---|---|---|---|---|---|---|---|
| | 4 | 5 | 8 | 9 | | 6 | 7 | 10 |
| 0 | 0,625 | 0,592 | 0,548 | 0,540 | 1 | 0,500 | 0,500 | 0,500 |
| 2 | 0,375 | 0,408 | 0,452 | 0,460 | 3 | 0,360 | 0,386 | 0,431 |
| 4 | 0,167 | 0,242 | 0,360 | 0,381 | 5 | 0,235 | 0,281 | 0,364 |
| 6 | 0,042 | 0,117 | 0,274 | 0,306 | 7 | 0,136 | 0,191 | 0,300 |
| 8 | | 0,042 | 0,199 | 0,238 | 9 | 0,068 | 0,119 | 0,242 |
| 10 | | $0{,}0^{2}83$ | 0,138 | 0,179 | 11 | 0,028 | 0,068 | 0,190 |
| 12 | | | 0,089 | 0,130 | 13 | $0{,}0^{2}83$ | 0,035 | 0,146 |
| 14 | | | 0,054 | 0,090 | 15 | $0{,}0^{2}14$ | 0,015 | 0,108 |
| 16 | | | 0,031 | 0,060 | 17 | | $0{,}0^{2}54$ | 0,078 |
| 18 | | | 0,016 | 0,038 | 19 | | $0{,}0^{2}14$ | 0,054 |
| 20 | | | $0{,}0^{2}71$ | 0,022 | 21 | | $0{,}0^{3}20$ | 0,036 |
| 22 | | | $0{,}0^{2}28$ | 0,012 | 23 | | | 0,023 |
| 24 | | | $0{,}0^{3}87$ | $0{,}0^{2}63$ | 25 | | | 0,014 |
| 26 | | | $0{,}0^{3}19$ | $0{,}0^{2}29$ | 27 | | | $0{,}0^{2}83$ |
| | | | | | 29 | | | $0{,}0^{2}46$ |
| 28 | | | $0{,}0^{4}25$ | $0{,}0^{2}12$ | 31 | | | $0{,}0^{2}23$ |
| 30 | | | | $0{,}0^{3}43$ | 33 | | | $0{,}0^{2}11$ |
| 32 | | | | $0{,}0^{3}12$ | 35 | | | $0{,}0^{3}47$ |
| | | | | | 37 | | | $0{,}0^{3}18$ |
| 34 | | | | $0{,}0^{4}25$ | 39 | | | $0{,}0^{4}58$ |
| 36 | | | | $0{,}0^{5}28$ | 41 | | | $0{,}0^{4}15$ |
| | | | | | 43 | | | $0{,}0^{5}28$ |
| | | | | | 45 | | | $0{,}0^{6}28$ |

Примечание. Повторяющиеся нули заменены степенями; например, $0{,}0^{3}47$ означает 0,00047.

Таблица 2. **Функция вероятности $S$ ($d^2$) (для $\rho$)**

| $n=4$ | | $n=5$ | | $n=6$ | | $n=7$ | | $n=8$ | | $n=9$ | | $n=10$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $S$ | $P$ | $S$ | $P$ | $S$ | $P$ | $S$ | $P$ | $S$ | $P$ | $S$ | $P$ | $S$ | $P$ |
| 12 | 458 | 22 | 475 | 36 | 500 | 58 | 482 | 86 | 488 | 122 | 491 | 166 | 500 |
| 14 | 375 | 24 | 392 | 38 | 460 | 60 | 453 | 88 | 467 | 124 | 474 | 168 | 486 |
| 16 | 208 | 26 | 342 | 40 | 401 | 62 | 420 | 90 | 441 | 126 | 456 | 170 | 473 |
| 18 | 167 | 28 | 258 | 42 | 357 | 64 | 391 | 92 | 420 | 128 | 440 | 172 | 459 |
| 20 | 042 | 30 | 225 | 44 | 329 | 66 | 357 | 94 | 397 | 130 | 422 | 174 | 446 |
| | | 32 | 175 | 46 | 282 | 68 | 331 | 96 | 376 | 132 | 405 | 176 | 433 |
| | | 34 | 117 | 48 | 249 | 70 | 297 | 98 | 352 | 134 | 388 | 178 | 419 |
| | | 36 | 067 | 50 | 210 | 72 | 278 | 100 | 332 | 136 | 372 | 180 | 406 |
| | | 38 | 042 | 52 | 178 | 74 | 249 | 102 | 310 | 138 | 354 | 182 | 393 |
| | | 40 | $0^{2}83$ | 54 | 149 | 76 | 222 | 104 | 291 | 140 | 339 | 184 | 379 |
| | | | | 56 | 121 | 78 | 198 | 106 | 268 | 142 | 322 | 186 | 367 |
| | | | | 58 | 088 | 80 | 177 | 108 | 250 | 144 | 307 | 188 | 354 |
| | | | | 60 | 068 | 82 | 151 | 110 | 231 | 146 | 290 | 190 | 341 |
| | | | | 62 | 051 | 84 | 133 | 112 | 214 | 148 | 276 | 192 | 328 |
| | | | | 64 | 029 | 86 | 118 | 114 | 195 | 150 | 260 | 194 | 316 |
| | | | | 66 | 017 | 88 | 100 | 116 | 180 | 152 | 247 | 196 | 304 |
| | | | | 68 | $0^{2}83$ | 90 | 083 | 118 | 163 | 154 | 231 | 198 | 292 |
| | | | | 70 | $0^{2}14$ | 92 | 069 | 120 | 150 | 156 | 218 | 200 | 280 |
| | | | | | | 94 | 055 | 122 | 134 | 158 | 205 | 202 | 268 |
| | | | | | | 96 | 044 | 124 | 122 | 160 | 193 | 204 | 257 |
| | | | | | | 98 | 033 | 126 | 108 | 162 | 179 | 206 | 246 |
| | | | | | | 100 | 024 | 128 | 098 | 164 | 168 | 208 | 235 |
| | | | | | | 102 | 017 | 130 | 085 | 166 | 156 | 210 | 224 |
| | | | | | | 104 | 012 | 132 | 076 | 168 | 146 | 212 | 214 |
| | | | | | | 106 | $0^{2}62$ | 134 | 066 | 170 | 135 | 214 | 203 |
| | | | | | | 108 | $0^{2}34$ | 136 | 057 | 172 | 125 | 216 | 193 |
| | | | | | | 110 | $0^{2}14$ | 138 | 048 | 174 | 115 | 218 | 184 |
| | | | | | | 112 | $0^{3}20$ | 140 | 042 | 176 | 106 | 220 | 174 |
| | | | | | | | | 142 | 035 | 178 | 097 | 222 | 165 |
| | | | | | | | | 144 | 029 | 180 | 089 | 224 | 156 |
| | | | | | | | | 146 | 023 | 182 | 081 | 226 | 148 |

Примечание. Знаки целой части опущены у $P$, например, для $n=4$ и $S=20$ значение $P$ равно 0,042.

Поскольку распределение $S$ ($d^2$) является симметричным, приведены только $P \geqslant \frac{1}{2}$. Для значений $S$ ($d^2$) меньших, чем те, что показаны в таблице, для $n=4$, находим $22-S$ ($d^2$), для $n=5$ $42-S(d^2)$, для $n=6$ $72-S$ ($d^2$), для $n=7$ $114-S(d^2)$, для $n=8$ $170-S$ ($d^2$), для $n=9$ $242-S$ ($d^2$) и для $n=10$ $332-S$ ($d^2$). Например, при $n=7$ и $S(d^2)=24$ значение $P$ равно $1-0,083=0,917$ (т. е. $114-24=90$, для $S(d^2)=90$ табличное значение $P=0,083$; откуда для $S$ ($d^2$)$=24$ $P=1-0,083$. —*Прим. перев.*)

Таблица 2 (продолжение)

| $n=8$ | | $n=9$ | | $n=10$ | | $n=10$ | |
|---|---|---|---|---|---|---|---|
| $S$ | $P$ | $S$ | $P$ | $S$ | $P$ | $S$ | $P$ |
| 148 | 018 | 184 | 074 | 228 | 139 | 286 | 010 |
| 150 | 014 | 186 | 066 | 230 | 132 | 288 | $0^{2}87$ |
| 152 | 011 | 188 | 060 | 232 | 124 | 290 | $0^{2}75$ |
| 154 | $0^{2}77$ | 190 | 054 | 234 | 116 | 292 | $0^{2}63$ |
| 156 | $0^{2}54$ | 192 | 048 | 236 | 109 | 294 | $0^{2}53$ |
| 158 | $0^{2}36$ | 194 | 043 | 238 | 102 | 296 | $0^{2}44$ |
| 160 | $0^{2}23$ | 196 | 038 | 240 | 096 | 298 | $0^{2}36$ |
| 162 | $0^{2}11$ | 198 | 033 | 242 | 089 | 300 | $0^{2}29$ |
| 164 | $0^{3}57$ | 200 | 029 | 244 | 083 | 302 | $0^{2}24$ |
| 166 | $0^{3}20$ | 202 | 025 | 246 | 077 | 304 | $0^{2}19$ |
| 168 | $0^{4}25$ | 204 | 022 | 248 | 072 | 306 | $0^{2}14$ |
| | | 206 | 018 | 250 | 067 | 308 | $0^{2}11$ |
| | | 208 | 016 | 252 | 062 | 310 | $0^{3}80$ |
| | | 210 | 013 | 254 | 057 | 312 | $0^{3}57$ |
| | | 212 | 011 | 256 | 052 | 314 | $0^{3}40$ |
| | | 214 | $0^{2}86$ | 258 | 048 | 316 | $0^{3}27$ |
| | | 216 | $0^{2}69$ | 260 | 044 | 318 | $0^{3}17$ |
| | | 218 | $0^{2}54$ | 262 | 040 | 320 | $0^{3}10$ |
| | | 220 | $0^{2}41$ | 264 | 037 | 322 | $0^{4}54$ |
| | | 222 | $0^{2}30$ | 266 | 033 | 324 | $0^{4}25$ |
| | | 224 | $0^{2}23$ | 268 | 030 | 326 | $0^{4}10$ |
| | | 226 | $0^{2}15$ | 270 | 027 | 328 | $0^{5}28$ |
| | | 228 | $0^{2}10$ | 272 | 024 | 330 | $0^{6}28$ |
| | | 230 | $0^{3}66$ | 274 | 022 | | |
| | | 232 | $0^{3}37$ | 276 | 019 | | |
| | | 234 | $0^{3}18$ | 278 | 017 | | |
| | | 236 | $0^{4}83$ | 280 | 015 | | |
| | | 238 | $0^{4}25$ | 282 | 013 | | |
| | | 240 | $0^{5}28$ | 284 | 012 | | |

Таблица 2 (продолжение)

| $n=11$ | | $n=11$ | | $n=11$ | | $n=12$ | | $n=12$ | |
|---|---|---|---|---|---|---|---|---|---|
| $S$ | $P$ | $S$ | $P$ | $S$ | $P$ | $S$ | $P$ | $S$ | $P$ |
| 222 | 495 | 320 | 082 | 418 | $0^{3}19$ | 288 | 496 | 386 | 133 |
| 224 | 484 | 322 | 077 | 420 | $0^{3}14$ | 290 | 487 | 388 | 128 |
| 226 | 473 | 324 | 073 | 422 | $0^{4}95$ | 292 | 478 | 390 | 123 |
| 228 | 462 | 326 | 069 | 424 | $0^{4}64$ | 294 | 469 | 392 | 118 |
| 230 | 452 | 328 | 065 | 426 | $0^{4}41$ | 296 | 460 | 394 | 114 |
| 232 | 441 | 330 | 061 | 428 | $0^{4}24$ | 298 | 452 | 396 | 109 |
| 234 | 430 | 332 | 057 | 430 | $0^{4}14$ | 300 | 443 | 398 | 105 |
| 236 | 419 | 334 | 054 | 432 | $0^{5}69$ | 302 | 435 | 400 | 100 |
| 238 | 409 | 336 | 050 | 434 | $0^{5}30$ | 304 | 426 | 402 | 096 |
| 240 | 398 | 338 | 047 | 436 | $0^{5}12$ | 306 | 417 | 404 | 092 |
| 242 | 388 | 340 | 044 | 438 | $0^{6}28$ | 308 | 409 | 406 | 088 |
| 244 | 377 | 342 | 041 | 440 | $0^{7}25$ | 310 | 400 | 408 | 084 |
| 246 | 367 | 344 | 038 | | | 312 | 391 | 410 | 081 |
| 248 | 357 | 346 | 035 | | | 314 | 383 | 412 | 077 |
| 250 | 347 | 348 | 033 | | | 316 | 375 | 414 | 074 |
| 252 | 337 | 350 | 030 | | | 318 | 366 | 416 | 070 |
| 254 | 327 | 352 | 028 | | | 320 | 358 | 418 | 067 |
| 256 | 317 | 354 | 026 | | | 322 | 350 | 420 | 064 |
| 258 | 307 | 356 | 024 | | | 324 | 342 | 422 | 061 |
| 260 | 298 | 358 | 022 | | | 326 | 334 | 424 | 058 |
| 262 | 288 | 360 | 020 | | | 328 | 325 | 426 | 055 |
| 264 | 279 | 362 | 018 | | | 330 | 318 | 428 | 052 |
| 266 | 270 | 364 | 017 | | | 332 | 310 | 430 | 049 |
| 268 | 260 | 366 | 015 | | | 334 | 302 | 432 | 047 |
| 270 | 252 | 368 | 014 | | | 336 | 294 | 434 | 044 |
| 272 | 243 | 370 | 013 | | | 338 | 287 | 436 | 042 |
| 274 | 234 | 372 | 011 | | | 340 | 279 | 438 | 040 |
| 276 | 226 | 374 | 010 | | | 342 | 272 | 440 | 037 |
| 278 | 217 | 376 | $0^{2}91$ | | | 344 | 264 | 442 | 035 |
| 280 | 209 | 378 | $0^{2}81$ | | | 346 | 257 | 444 | 033 |
| 282 | 201 | 380 | $0^{2}72$ | | | 348 | 250 | 446 | 031 |
| 284 | 193 | 382 | $0^{2}64$ | | | 350 | 243 | 448 | 030 |
| 286 | 186 | 384 | $0^{2}56$ | | | 352 | 235 | 450 | 028 |
| 288 | 178 | 386 | $0^{2}49$ | | | 354 | 229 | 452 | 026 |
| 290 | 171 | 388 | $0^{2}43$ | | | 356 | 222 | 454 | 024 |
| 292 | 163 | 390 | $0^{2}37$ | | | 358 | 215 | 456 | 023 |
| 294 | 157 | 392 | $0^{2}32$ | | | 360 | 208 | 458 | 021 |
| 296 | 150 | 394 | $0^{2}27$ | | | 362 | 202 | 460 | 020 |
| 298 | 143 | 396 | $0^{2}23$ | | | 364 | 196 | 462 | 019 |
| 300 | 137 | 398 | $0^{2}20$ | | | 366 | 189 | 464 | 017 |
| 302 | 130 | 400 | $0^{2}16$ | | | 368 | 183 | 466 | 016 |
| 304 | 124 | 402 | $0^{2}14$ | | | 370 | 177 | 468 | 015 |
| 306 | 118 | 404 | $0^{2}11$ | | | 372 | 171 | 470 | 014 |
| 308 | 112 | 406 | $0^{3}91$ | | | 374 | 166 | 472 | 013 |
| 310 | 107 | 408 | $0^{3}73$ | | | 376 | 160 | 474 | 012 |
| 312 | 102 | 410 | $0^{3}58$ | | | 378 | 154 | 476 | 011 |
| 314 | 096 | 412 | $0^{3}45$ | | | 380 | 149 | 478 | 010 |
| 316 | 091 | 414 | $0^{3}35$ | | | 382 | 143 | 480 | $0^{2}93$ |
| 318 | 087 | 416 | $0^{3}26$ | | | 384 | 138 | 482 | $0^{2}85$ |

| $n=12$ | | $n=13$ | | $n=13$ | | $n=13$ | | $n=13$ | |
|---|---|---|---|---|---|---|---|---|---|
| $S$ | $P$ | $S$ | $P$ | $S$ | $P$ | $S$ | $P$ | $S$ | $P$ |
| 484 | $0^{2}78$ | 366 | 496 | 464 | 182 | 562 | 029 | 660 | $0^{3}60$ |
| 486 | $0^{2}71$ | 368 | 489 | 466 | 177 | 564 | 028 | 662 | $0^{3}53$ |
| 488 | $0^{2}65$ | 370 | 482 | 468 | 172 | 566 | 026 | 664 | $0^{3}46$ |
| 490 | $0^{2}59$ | 372 | 475 | 470 | 167 | 568 | 025 | 666 | $0^{3}40$ |
| 492 | $0^{2}53$ | 374 | 468 | 472 | 162 | 570 | 024 | 668 | $0^{3}34$ |
| 494 | $0^{2}48$ | 376 | 460 | 474 | 158 | 572 | 022 | 670 | $0^{3}29$ |
| 496 | $0^{2}43$ | 378 | 453 | 476 | 153 | 574 | 021 | 672 | $0^{3}25$ |
| 498 | $0^{2}39$ | 380 | 446 | 478 | 149 | 576 | 020 | 674 | $0^{3}21$ |
| 500 | $0^{2}35$ | 382 | 439 | 480 | 144 | 578 | 019 | 676 | $0^{3}18$ |
| 502 | $0^{2}31$ | 384 | 432 | 482 | 140 | 580 | 018 | 678 | $0^{3}15$ |
| 504 | $0^{2}28$ | 386 | 425 | 484 | 136 | 582 | 017 | 680 | $0^{3}13$ |
| 506 | $0^{2}25$ | 388 | 417 | 486 | 132 | 584 | 016 | 682 | $0^{3}10$ |
| 508 | $0^{2}22$ | 390 | 410 | 488 | 128 | 586 | 015 | 684 | $0^{4}86$ |
| 510 | $0^{2}19$ | 392 | 403 | 490 | 124 | 588 | 014 | 686 | $0^{4}70$ |
| 512 | $0^{2}17$ | 394 | 396 | 492 | 120 | 590 | 013 | 688 | $0^{4}56$ |
| 514 | $0^{2}15$ | 396 | 389 | 494 | 116 | 592 | 013 | 690 | $0^{4}45$ |
| 516 | $0^{2}13$ | 398 | 382 | 496 | 112 | 594 | 012 | 692 | $0^{4}35$ |
| 518 | $0^{2}11$ | 400 | 375 | 498 | 108 | 596 | 011 | 694 | $0^{4}27$ |
| 520 | $0^{3}93$ | 402 | 369 | 500 | 105 | 598 | 010 | 696 | $0^{4}21$ |
| 522 | $0^{3}80$ | 404 | 362 | 502 | 101 | 600 | $0^{2}97$ | 698 | $0^{4}16$ |
| 524 | $0^{3}67$ | 406 | 355 | 504 | 098 | 602 | $0^{2}91$ | 700 | $0^{4}12$ |
| 526 | $0^{3}57$ | 408 | 348 | 506 | 094 | 604 | $0^{2}85$ | 702 | $0^{5}85$ |
| 528 | $0^{3}47$ | 410 | 341 | 508 | 091 | 606 | $0^{2}79$ | 704 | $0^{5}61$ |
| 530 | $0^{3}39$ | 412 | 335 | 510 | 088 | 608 | $0^{2}74$ | 706 | $0^{5}42$ |
| 532 | $0^{3}32$ | 414 | 328 | 512 | 085 | 610 | $0^{2}68$ | 708 | $0^{5}28$ |
| 534 | $0^{3}26$ | 416 | 321 | 514 | 082 | 612 | $0^{2}64$ | 710 | $0^{5}18$ |
| 536 | $0^{3}21$ | 418 | 315 | 516 | 079 | 614 | $0^{2}59$ | 712 | $0^{5}11$ |
| 538 | $0^{3}17$ | 420 | 308 | 518 | 076 | 616 | $0^{2}55$ | 714 | $0^{6}67$ |
| 540 | $0^{3}13$ | 422 | 302 | 520 | 073 | 618 | $0^{2}51$ | 716 | $0^{6}37$ |
| 542 | $0^{3}10$ | 424 | 296 | 522 | 070 | 620 | $0^{2}47$ | 718 | $0^{6}19$ |
| 544 | $0^{4}77$ | 426 | 289 | 524 | 068 | 622 | $0^{2}43$ | 720 | $0^{7}85$ |
| 546 | $0^{4}58$ | 428 | 283 | 526 | 065 | 624 | $0^{2}40$ | 722 | $0^{7}34$ |
| 548 | $0^{4}42$ | 430 | 277 | 528 | 062 | 626 | $0^{2}36$ | 724 | $0^{7}11$ |
| 550 | $0^{4}30$ | 432 | 271 | 530 | 060 | 628 | $0^{2}33$ | 726 | $0^{8}21$ |
| 552 | $0^{4}21$ | 434 | 265 | 532 | 057 | 630 | $0^{2}31$ | 728 | $0^{9}16$ |
| 554 | $0^{4}14$ | 436 | 259 | 534 | 055 | 632 | $0^{2}28$ | | |
| 556 | $0^{5}90$ | 438 | 253 | 536 | 053 | 634 | $0^{2}25$ | | |
| 558 | $0^{5}55$ | 440 | 247 | 538 | 051 | 636 | $0^{2}23$ | | |
| 560 | $0^{5}32$ | 442 | 241 | 540 | 049 | 638 | $0^{2}21$ | | |
| 562 | $0^{5}17$ | 444 | 235 | 542 | 046 | 640 | $0^{2}19$ | | |
| 564 | $0^{6}80$ | 446 | 230 | 544 | 044 | 642 | $0^{2}17$ | | |
| 566 | $0^{6}34$ | 448 | 224 | 546 | 043 | 644 | $0^{2}15$ | | |
| 568 | $0^{6}12$ | 450 | 218 | 548 | 041 | 646 | $0^{2}14$ | | |
| 570 | $0^{7}25$ | 452 | 213 | 550 | 039 | 648 | $0^{2}12$ | | |
| 572 | $0^{8}21$ | 454 | 208 | 552 | 037 | 650 | $0^{2}11$ | | |
| | | 456 | 202 | 554 | 035 | 652 | $0^{3}99$ | | |
| | | 458 | 197 | 556 | 034 | 654 | $0^{3}88$ | | |
| | | 460 | 192 | 558 | 032 | 656 | $0^{3}78$ | | |
| | | 462 | 187 | 560 | 031 | 658 | $0^{3}68$ | | |

**Таблица 3. Площади под нормальной кривой (функция нормального распределения)**

В таблице показаны значения площади под кривой $y=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$, лежащие левее точки $x$, например для $x=1{,}86$ $(=1{,}5+0{,}36)$, $F=0{,}9686$

| $x$ | 0, + | 0,5+ | 1,0+ | 1,5+ | 2,0+ | 2,5+ | 3,0+ | 3,5+ |
|---|---|---|---|---|---|---|---|---|
| 0,00 | 5000 | 6915 | 8413 | 9332 | 9772 | $9^2379$ | $9^2865$ | $9^377$ |
| 0,01 | 5040 | 6950 | 8438 | 9345 | 9778 | $9^2396$ | $9^2869$ | $9^378$ |
| 0,02 | 5080 | 6985 | 8461 | 9357 | 9783 | $9^2413$ | $9^2874$ | $9^378$ |
| 0,03 | 5120 | 7019 | 8485 | 9370 | 9788 | $9^2430$ | $9^2878$ | $9^379$ |
| 0,04 | 5160 | 7054 | 8508 | 9382 | 9793 | $9^2446$ | $9^2882$ | $9^380$ |
| 0,05 | 5199 | 7088 | 8531 | 9394 | 9798 | $9^2461$ | $9^2886$ | $9^381$ |
| 0,06 | 5239 | 7123 | 8554 | 9406 | 9803 | $9^2477$ | $9^2889$ | $9^381$ |
| 0,07 | 5279 | 7157 | 8577 | 9418 | 9808 | $9^2492$ | $9^2893$ | $9^382$ |
| 0,08 | 5319 | 7190 | 8599 | 9429 | 9812 | $9^2506$ | $9^2897$ | $9^383$ |
| 0,09 | 5359 | 7224 | 8621 | 9441 | 9817 | $9^2520$ | $9^2900$ | $9^383$ |
| 0,10 | 5398 | 7257 | 8643 | 9452 | 9821 | $9^2534$ | $9^303$ | $9^384$ |
| 0,11 | 5438 | 7291 | 8665 | 9463 | 9826 | $9^2547$ | $9^306$ | $9^385$ |
| 0,12 | 5478 | 7324 | 8686 | 9474 | 9830 | $9^2560$ | $9^310$ | $9^385$ |
| 0,13 | 5517 | 7357 | 8708 | 9484 | 9834 | $9^2573$ | $9^313$ | $9^386$ |
| 0,14 | 5557 | 7389 | 8729 | 9495 | 9838 | $9^2585$ | $9^316$ | $9^386$ |
| 0,15 | 5596 | 7422 | 8749 | 9505 | 9842 | $9^2598$ | $9^318$ | $9^387$ |
| 0,16 | 5636 | 7454 | 8770 | 9515 | 9846 | $9^2609$ | $9^321$ | $9^387$ |
| 0,17 | 5675 | 7486 | 8790 | 9525 | 9850 | $9^2621$ | $9^324$ | $9^388$ |
| 0,18 | 5714 | 7517 | 8810 | 9535 | 9854 | $9^2632$ | $9^326$ | $9^388$ |
| 0,19 | 5753 | 7549 | 8830 | 9545 | 9857 | $9^2643$ | $9^329$ | $9^389$ |
| 0,20 | 5793 | 7580 | 8849 | 9554 | 9861 | $9^2653$ | $9^331$ | $9^389$ |
| 0,21 | 5832 | 7611 | 8869 | 9564 | 9864 | $9^2664$ | $9^334$ | $9^390$ |
| 0,22 | 5871 | 7642 | 8888 | 9573 | 9868 | $9^2674$ | $9^336$ | $9^390$ |
| 0,23 | 5910 | 7673 | 8907 | 9582 | 9871 | $9^2683$ | $9^338$ | $9^404$ |
| 0,24 | 5948 | 7704 | 8925 | 9591 | 9875 | $9^2693$ | $9^340$ | $9^408$ |
| 0,25 | 5987 | 7738 | 8944 | 9599 | 9878 | $9^2702$ | $9^342$ | $9^412$ |
| 0,26 | 6026 | 7764 | 8962 | 9608 | 9881 | $9^2711$ | $9^344$ | $9^415$ |
| 0,27 | 6064 | 7794 | 8980 | 9616 | 9884 | $9^2720$ | $9^346$ | $9^418$ |
| 0,28 | 6103 | 7823 | 8997 | 9625 | 9887 | $9^2728$ | $9^348$ | $9^422$ |
| 0,29 | 6141 | 7852 | 9015 | 9633 | 9890 | $9^2736$ | $9^350$ | $9^425$ |
| 0,30 | 6179 | 7881 | 9032 | 9641 | 9893 | $9^2744$ | $9^352$ | $9^428$ |
| 0,31 | 6217 | 7910 | 9049 | 9649 | 9896 | $9^2752$ | $9^353$ | $9^431$ |
| 0,32 | 6255 | 7939 | 9066 | 9656 | 9898 | $9^2760$ | $9^355$ | $9^433$ |
| 0,33 | 6293 | 7967 | 9082 | 9664 | 9901 | $9^2767$ | $9^357$ | $9^436$ |
| 0,34 | 6331 | 7995 | 9099 | 9671 | 9904 | $9^2774$ | $9^358$ | $9^439$ |
| 0,35 | 6368 | 8023 | 9115 | 9678 | 9906 | $9^2781$ | $9^360$ | $9^441$ |
| 0,36 | 6406 | 8051 | 9131 | 9686 | 9909 | $9^2788$ | $9^361$ | $9^443$ |
| 0,37 | 6443 | 8078 | 9147 | 9693 | 9911 | $9^2795$ | $9^362$ | $9^446$ |
| 0,38 | 6480 | 8106 | 9162 | 9699 | 9913 | $9^2801$ | $9^364$ | $9^448$ |
| 0,39 | 6517 | 8133 | 9177 | 9706 | 9916 | $9^2807$ | $9^365$ | $9^450$ |
| 0,40 | 6554 | 8159 | 9192 | 9713 | 9918 | $9^2813$ | $9^366$ | $9^452$ |
| 0,41 | 6591 | 8186 | 9207 | 9719 | 9920 | $9^2819$ | $9^368$ | $9^454$ |
| 0,42 | 6628 | 8212 | 9222 | 9726 | 9922 | $9^2825$ | $9^369$ | $9^456$ |
| 0,43 | 6664 | 8238 | 9236 | 9732 | 9925 | $9^2831$ | $9^370$ | $9^458$ |
| 0,44 | 6700 | 8264 | 9251 | 9738 | 9927 | $9^2836$ | $9^371$ | $9^459$ |
| 0,45 | 6736 | 8289 | 9265 | 9744 | 9929 | $9^2841$ | $9^372$ | $9^461$ |
| 0,46 | 6772 | 8315 | 9279 | 9750 | 9931 | $9^2846$ | $9^373$ | $9^463$ |
| 0,47 | 6808 | 8340 | 9292 | 9756 | 9932 | $9^2851$ | $9^374$ | $9^464$ |
| 0,48 | 6844 | 8365 | 9306 | 9761 | 9934 | $9^2856$ | $9^375$ | $9^466$ |
| 0,49 | 6879 | 8389 | 9319 | 9767 | 9936 | $9^2861$ | $9^376$ | $9^467$ |

Примечание. Знаки целой части, т. е. 0, в таблице опущены. Повторяющиеся девятки заменены степенями; например, $9^371$ означает 0,99971.

**Таблица 4. Случайные последовательности рангов от 1 до 20 (случайные перестановки первых двадцати чисел натурального ряда)**

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 19 | 15 | 10 | 13 | 3 | 11 | 18 | 1 | 8 | 7 | 2 | 14 | 17 | 12 | 9 | 6 | 20 | 5 | 16 |
| 16 | 14 | 2 | 17 | 8 | 11 | 12 | 13 | 4 | 3 | 7 | 10 | 5 | 15 | 1 | 6 | 20 | 18 | 19 | 9 |
| 6 | 5 | 10 | 12 | 17 | 7 | 9 | 2 | 1 | 15 | 19 | 11 | 13 | 14 | 16 | 3 | 20 | 4 | 18 | 8 |
| 18 | 9 | 10 | 19 | 6 | 12 | 4 | 1 | 20 | 11 | 13 | 16 | 15 | 3 | 2 | 5 | 14 | 17 | 8 | 7 |
| 11 | 12 | 8 | 9 | 1 | 15 | 13 | 2 | 14 | 17 | 3 | 19 | 6 | 7 | 16 | 4 | 5 | 20 | 18 | 10 |
| | | | | | | | | | | | | | | | | | | | |
| 1 | 15 | 18 | 4 | 8 | 20 | 19 | 9 | 11 | 3 | 17 | 10 | 13 | 14 | 6 | 7 | 5 | 2 | 12 | 16 |
| 18 | 2 | 16 | 12 | 8 | 5 | 11 | 17 | 4 | 10 | 3 | 9 | 7 | 13 | 1 | 14 | 6 | 19 | 20 | 15 |
| 2 | 3 | 18 | 19 | 11 | 1 | 7 | 12 | 20 | 8 | 15 | 5 | 14 | 4 | 13 | 6 | 9 | 17 | 16 | 10 |
| 15 | 16 | 12 | 13 | 17 | 19 | 1 | 2 | 20 | 7 | 4 | 11 | 10 | 9 | 6 | 8 | 5 | 18 | 14 | 3 |
| 8 | 18 | 3 | 4 | 10 | 7 | 16 | 17 | 9 | 6 | 11 | 15 | 19 | 2 | 13 | 12 | 14 | 1 | 20 | 5 |
| | | | | | | | | | | | | | | | | | | | |
| 8 | 6 | 13 | 1 | 10 | 11 | 15 | 20 | 7 | 14 | 3 | 18 | 2 | 12 | 16 | 17 | 4 | 19 | 5 | 9 |
| 10 | 5 | 17 | 2 | 15 | 9 | 16 | 3 | 14 | 1 | 6 | 12 | 19 | 13 | 20 | 4 | 11 | 8 | 7 | 18 |
| 16 | 1 | 18 | 11 | 2 | 14 | 13 | 15 | 20 | 7 | 9 | 3 | 19 | 17 | 5 | 12 | 6 | 10 | 4 | 8 |
| 12 | 7 | 5 | 20 | 15 | 19 | 13 | 10 | 18 | 1 | 8 | 16 | 9 | 2 | 6 | 11 | 4 | 3 | 17 | 14 |
| 20 | 1 | 19 | 5 | 9 | 6 | 11 | 10 | 7 | 3 | 2 | 18 | 4 | 14 | 8 | 17 | 15 | 12 | 13 | 16 |
| | | | | | | | | | | | | | | | | | | | |
| 9 | 6 | 17 | 15 | 10 | 12 | 18 | 13 | 1 | 19 | 4 | 5 | 14 | 3 | 2 | 16 | 7 | 11 | 20 | 8 |
| 4 | 2 | 3 | 19 | 11 | 10 | 7 | 16 | 13 | 18 | 15 | 1 | 12 | 17 | 9 | 8 | 6 | 5 | 20 | 14 |
| 9 | 18 | 19 | 14 | 10 | 12 | 4 | 16 | 5 | 2 | 6 | 17 | 1 | 8 | 20 | 11 | 13 | 7 | 3 | 15 |
| 13 | 12 | 8 | 7 | 18 | 4 | 16 | 10 | 17 | 19 | 5 | 1 | 20 | 3 | 14 | 11 | 2 | 6 | 9 | 15 |
| 6 | 1 | 4 | 5 | 8 | 9 | 10 | 18 | 15 | 20 | 13 | 7 | 12 | 2 | 3 | 19 | 17 | 16 | 11 | 14 |
| | | | | | | | | | | | | | | | | | | | |
| 8 | 20 | 4 | 5 | 15 | 10 | 11 | 6 | 16 | 18 | 19 | 9 | 7 | 14 | 17 | 1 | 12 | 3 | 2 | 13 |
| 19 | 16 | 3 | 4 | 7 | 6 | 12 | 1 | 13 | 2 | 10 | 17 | 8 | 20 | 11 | 9 | 18 | 5 | 15 | 14 |
| 19 | 18 | 13 | 8 | 15 | 9 | 10 | 2 | 17 | 14 | 16 | 7 | 5 | 1 | 20 | 4 | 11 | 3 | 6 | 12 |
| 2 | 11 | 13 | 5 | 7 | 18 | 20 | 6 | 16 | 14 | 1 | 3 | 15 | 8 | 19 | 9 | 12 | 17 | 4 | 10 |
| 2 | 12 | 16 | 10 | 17 | 13 | 20 | 11 | 3 | 4 | 8 | 14 | 15 | 19 | 7 | 9 | 5 | 6 | 18 | 1 |
| | | | | | | | | | | | | | | | | | | | |
| 13 | 11 | 4 | 3 | 7 | 5 | 20 | 15 | 1 | 12 | 9 | 6 | 14 | 17 | 16 | 2 | 18 | 19 | 10 | 8 |
| 15 | 1 | 16 | 7 | 3 | 14 | 8 | 11 | 2 | 13 | 10 | 6 | 18 | 5 | 9 | 12 | 19 | 4 | 17 | 20 |
| 3 | 20 | 1 | 11 | 17 | 16 | 4 | 14 | 6 | 2 | 9 | 19 | 15 | 12 | 10 | 8 | 5 | 13 | 7 | 18 |
| 18 | 16 | 3 | 2 | 12 | 19 | 4 | 20 | 14 | 11 | 8 | 10 | 7 | 17 | 1 | 5 | 15 | 13 | 9 | 6 |
| 3 | 4 | 12 | 5 | 10 | 14 | 15 | 1 | 7 | 2 | 13 | 19 | 17 | 18 | 6 | 11 | 20 | 16 | 9 | 8 |
| | | | | | | | | | | | | | | | | | | | |
| 13 | 20 | 11 | 14 | 3 | 9 | 4 | 5 | 19 | 17 | 18 | 1 | 16 | 15 | 8 | 10 | 6 | 7 | 12 | 2 |
| 17 | 16 | 13 | 18 | 7 | 2 | 9 | 15 | 10 | 6 | 11 | 4 | 12 | 19 | 14 | 3 | 20 | 5 | 1 | 8 |
| 13 | 7 | 3 | 4 | 8 | 6 | 15 | 2 | 18 | 16 | 5 | 19 | 9 | 12 | 10 | 20 | 11 | 1 | 14 | 17 |
| 2 | 19 | 17 | 7 | 18 | 20 | 15 | 5 | 9 | 11 | 3 | 16 | 13 | 6 | 12 | 1 | 4 | 10 | 14 | 8 |
| 4 | 6 | 7 | 10 | 20 | 18 | 3 | 5 | 11 | 12 | 2 | 1 | 14 | 9 | 19 | 13 | 15 | 17 | 8 | 16 |
| | | | | | | | | | | | | | | | | | | | |
| 19 | 3 | 15 | 12 | 10 | 17 | 16 | 1 | 8 | 7 | 2 | 6 | 14 | 9 | 11 | 18 | 4 | 5 | 20 | 13 |
| 18 | 14 | 9 | 2 | 8 | 6 | 7 | 1 | 3 | 10 | 5 | 17 | 20 | 4 | 13 | 15 | 11 | 12 | 16 | 19 |
| 18 | 5 | 17 | 1 | 4 | 7 | 11 | 20 | 9 | 10 | 15 | 3 | 19 | 14 | 16 | 13 | 6 | 12 | 8 | 2 |
| 5 | 1 | 10 | 6 | 19 | 7 | 15 | 11 | 2 | 4 | 3 | 9 | 20 | 18 | 14 | 17 | 13 | 16 | 8 | 12 |
| 19 | 17 | 1 | 16 | 10 | 9 | 8 | 18 | 13 | 15 | 5 | 2 | 7 | 20 | 14 | 12 | 11 | 4 | 3 | 6 |
| | | | | | | | | | | | | | | | | | | | |
| 13 | 6 | 14 | 4 | 9 | 17 | 19 | 7 | 12 | 8 | 18 | 16 | 11 | 1 | 3 | 5 | 20 | 10 | 12 | 15 |
| 10 | 16 | 9 | 6 | 12 | 17 | 4 | 19 | 11 | 1 | 20 | 13 | 14 | 18 | 15 | 7 | 2 | 3 | 8 | 5 |
| 14 | 8 | 7 | 16 | 18 | 6 | 17 | 19 | 11 | 20 | 4 | 13 | 2 | 10 | 1 | 15 | 9 | 3 | 5 | 12 |
| 16 | 2 | 8 | 14 | 4 | 10 | 19 | 11 | 7 | 6 | 12 | 17 | 5 | 20 | 13 | 15 | 1 | 9 | 3 | 18 |
| 14 | 5 | 9 | 8 | 17 | 10 | 6 | 12 | 19 | 13 | 3 | 20 | 2 | 18 | 4 | 7 | 1 | 11 | 15 | 16 |

Примечание. Случайные последовательности, охватывающие меньше чем 20 первых чисел натурального ряда, могут быть получены путем отбрасывания ненужных рангов.

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 12 | 18 | 15 | 4 | 1 | 13 | 5 | 17 | 9 | 7 | 14 | 20 | 19 | 6 | 10 | 11 | 2 | 3 | 8 | 16 |
| 8 | 17 | 5 | 3 | 4 | 12 | 10 | 14 | 9 | 18 | 15 | 6 | 13 | 2 | 20 | 16 | 19 | 1 | 7 | 11 |
| 11 | 19 | 8 | 9 | 16 | 10 | 13 | 17 | 15 | 4 | 5 | 20 | 14 | 2 | 1 | 3 | 7 | 18 | 6 | 12 |
| 14 | 1 | 5 | 2 | 3 | 15 | 11 | 9 | 18 | 19 | 7 | 10 | 16 | 13 | 6 | 8 | 4 | 17 | 12 | 20 |
| 9 | 17 | 5 | 20 | 2 | 7 | 4 | 19 | 18 | 15 | 1 | 10 | 8 | 16 | 14 | 3 | 11 | 13 | 12 | 6 |
| 10 | 14 | 17 | 19 | 20 | 1 | 13 | 18 | 5 | 11 | 7 | 8 | 16 | 3 | 9 | 12 | 15 | 6 | 2 | 4 |
| 20 | 5 | 9 | 14 | 1 | 8 | 4 | 12 | 13 | 3 | 15 | 7 | 2 | 19 | 10 | 6 | 17 | 16 | 11 | 18 |
| 5 | 15 | 1 | 17 | 2 | 11 | 9 | 10 | 13 | 18 | 19 | 16 | 3 | 14 | 4 | 20 | 12 | 8 | 6 | 7 |
| 2 | 13 | 14 | 3 | 9 | 19 | 15 | 8 | 20 | 17 | 18 | 4 | 11 | 10 | 6 | 1 | 5 | 7 | 12 | 16 |
| 18 | 12 | 14 | 10 | 1 | 13 | 9 | 3 | 8 | 7 | 5 | 17 | 16 | 2 | 4 | 20 | 15 | 19 | 6 | 11 |
| 3 | 16 | 11 | 6 | 12 | 13 | 15 | 5 | 7 | 20 | 19 | 18 | 4 | 10 | 9 | 14 | 8 | 2 | 17 | 1 |
| 2 | 11 | 4 | 16 | 13 | 20 | 14 | 18 | 6 | 15 | 8 | 10 | 1 | 3 | 19 | 7 | 5 | 9 | 17 | 12 |
| 14 | 12 | 20 | 10 | 15 | 2 | 9 | 3 | 16 | 13 | 6 | 5 | 7 | 11 | 8 | 1 | 17 | 19 | 4 | 18 |
| 12 | 10 | 18 | 5 | 4 | 20 | 17 | 16 | 13 | 15 | 9 | 7 | 14 | 2 | 6 | 11 | 1 | 19 | 8 | 3 |
| 1 | 18 | 5 | 16 | 12 | 7 | 19 | 15 | 14 | 2 | 10 | 9 | 13 | 3 | 20 | 4 | 17 | 11 | 6 | 8 |
| 10 | 8 | 14 | 18 | 16 | 3 | 2 | 4 | 9 | 19 | 15 | 5 | 11 | 12 | 13 | 17 | 6 | 1 | 20 | 7 |
| 18 | 4 | 1 | 9 | 12 | 14 | 15 | 19 | 2 | 6 | 3 | 10 | 17 | 13 | 7 | 16 | 8 | 20 | 11 | 5 |
| 1 | 15 | 18 | 9 | 13 | 16 | 7 | 8 | 5 | 11 | 3 | 4 | 12 | 20 | 14 | 10 | 19 | 6 | 2 | 17 |
| 13 | 20 | 11 | 10 | 6 | 12 | 18 | 9 | 15 | 17 | 19 | 1 | 7 | 16 | 8 | 2 | 3 | 5 | 14 | 4 |
| 6 | 16 | 9 | 7 | 10 | 2 | 8 | 20 | 13 | 1 | 11 | 15 | 17 | 3 | 19 | 4 | 14 | 12 | 18 | 5 |
| 2 | 7 | 10 | 5 | 15 | 19 | 16 | 1 | 13 | 12 | 11 | 18 | 4 | 3 | 6 | 2 | 14 | 17 | 9 | 8 |
| 90 | 12 | 19 | 4 | 2 | 6 | 1 | 15 | 7 | 18 | 3 | 13 | 10 | 20 | 11 | 8 | 14 | 5 | 16 | 9 |
| 17 | 1 | 7 | 8 | 15 | 14 | 6 | 19 | 11 | 20 | 9 | 4 | 17 | 2 | 16 | 3 | 10 | 13 | 12 | 5 |
| 18 | 19 | 9 | 7 | 17 | 18 | 4 | 11 | 12 | 8 | 6 | 10 | 3 | 16 | 13 | 1 | 20 | 2 | 5 | 15 |
| 14 | 11 | 18 | 1 | 8 | 4 | 16 | 10 | 7 | 13 | 19 | 12 | 5 | 17 | 6 | 20 | 2 | 3 | 14 | 15 |
| 14 | 10 | 8 | 9 | 4 | 20 | 6 | 11 | 17 | 1 | 16 | 3 | 18 | 13 | 7 | 15 | 19 | 12 | 2 | 5 |
| 17 | 15 | 18 | 16 | 6 | 14 | 8 | 9 | 7 | 11 | 2 | 10 | 19 | 3 | 5 | 20 | 4 | 1 | 13 | 12 |
| 5 | 7 | 2 | 20 | 12 | 13 | 6 | 19 | 3 | 1 | 15 | 8 | 17 | 18 | 10 | 11 | 16 | 4 | 14 | 9 |
| 12 | 4 | 10 | 8 | 14 | 16 | 11 | 1 | 19 | 13 | 18 | 15 | 9 | 17 | 6 | 3 | 2 | 7 | 20 | 15 |
| 5 | 14 | 8 | 11 | 9 | 4 | 12 | 20 | 2 | 1 | 15 | 6 | 10 | 17 | 13 | 3 | 7 | 18 | 19 | 16 |
| 6 | 3 | 12 | 19 | 14 | 15 | 13 | 2 | 17 | 1 | 5 | 16 | 18 | 4 | 11 | 20 | 9 | 8 | 7 | 10 |
| 18 | 11 | 6 | 10 | 16 | 1 | 8 | 12 | 15 | 3 | 14 | 20 | 4 | 2 | 9 | 13 | 5 | 19 | 17 | 7 |
| 18 | 12 | 4 | 7 | 11 | 13 | 5 | 9 | 2 | 14 | 19 | 3 | 1 | 17 | 8 | 16 | 6 | 15 | 20 | 10 |
| 16 | 5 | 20 | 11 | 2 | 9 | 15 | 1 | 19 | 3 | 10 | 14 | 12 | 6 | 7 | 17 | 8 | 4 | 13 | 18 |
| 6 | 20 | 11 | 13 | 18 | 14 | 17 | 8 | 12 | 7 | 5 | 4 | 2 | 16 | 9 | 10 | 15 | 3 | 19 | 1 |
| 8 | 6 | 20 | 16 | 11 | 12 | 15 | 3 | 9 | 7 | 13 | 1 | 5 | 14 | 4 | 18 | 2 | 10 | 17 | 19 |
| 18 | 12 | 5 | 16 | 1 | 9 | 7 | 6 | 8 | 4 | 13 | 14 | 19 | 2 | 15 | 10 | 3 | 11 | 17 | 20 |
| 4 | 15 | 9 | 20 | 2 | 10 | 5 | 13 | 6 | 3 | 7 | 16 | 1 | 8 | 18 | 14 | 17 | 19 | 12 | 11 |
| 16 | 19 | 17 | 13 | 15 | 12 | 11 | 18 | 8 | 2 | 5 | 10 | 20 | 7 | 3 | 1 | 14 | 6 | 9 | 4 |
| 12 | 4 | 2 | 18 | 5 | 11 | 1 | 7 | 17 | 6 | 14 | 8 | 9 | 3 | 16 | 10 | 15 | 13 | 20 | 19 |
| 13 | 20 | 9 | 5 | 1 | 6 | 8 | 4 | 7 | 19 | 16 | 15 | 10 | 2 | 12 | 17 | 11 | 18 | 14 | 3 |
| 6 | 7 | 11 | 8 | 10 | 1 | 20 | 16 | 12 | 9 | 13 | 15 | 14 | 18 | 3 | 17 | 2 | 5 | 4 | 19 |
| 17 | 8 | 11 | 10 | 4 | 5 | 12 | 9 | 7 | 1 | 19 | 14 | 20 | 6 | 2 | 18 | 15 | 3 | 16 | 13 |
| 5 | 16 | 9 | 18 | 6 | 15 | 19 | 3 | 4 | 20 | 14 | 1 | 2 | 12 | 7 | 17 | 10 | 13 | 8 | 11 |
| 8 | 11 | 15 | 16 | 1 | 18 | 12 | 10 | 5 | 13 | 20 | 2 | 14 | 17 | 9 | 6 | 4 | 7 | 3 | 19 |
| 17 | 6 | 10 | 12 | 2 | 4 | 9 | 18 | 5 | 13 | 8 | 15 | 11 | 1 | 16 | 3 | 20 | 7 | 14 | 19 |
| 20 | 17 | 5 | 2 | 9 | 7 | 10 | 11 | 8 | 1 | 4 | 19 | 14 | 16 | 12 | 3 | 13 | 15 | 6 | 18 |
| 9 | 12 | 8 | 20 | 15 | 5 | 10 | 2 | 16 | 17 | 11 | 18 | 6 | 3 | 4 | 1 | 14 | 7 | 13 | 19 |
| 19 | 15 | 14 | 11 | 5 | 18 | 16 | 9 | 17 | 2 | 20 | 7 | 12 | 4 | 13 | 6 | 1 | 8 | 3 | 10 |
| 1 | 10 | 8 | 17 | 6 | 19 | 18 | 7 | 12 | 11 | 13 | 5 | 14 | 2 | 20 | 4 | 9 | 16 | 3 | 15 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 3 | 13 | 19 | 6 | 1 | 5 | 12 | 17 | 8 | 11 | 2 | 16 | 20 | 10 | 18 | 9 | 15 | 14 | 7 |
| 11 | 7 | 12 | 10 | 16 | 5 | 18 | 4 | 1 | 13 | 19 | 2 | 3 | 14 | 6 | 8 | 9 | 20 | 17 | 15 |
| 7 | 14 | 13 | 3 | 15 | 19 | 16 | 8 | 5 | 20 | 11 | 6 | 12 | 17 | 1 | 10 | 18 | 9 | 2 | 4 |
| 14 | 6 | 4 | 10 | 2 | 8 | 19 | 5 | 20 | 15 | 9 | 3 | 17 | 12 | 16 | 1 | 18 | 13 | 7 | 11 |
| 13 | 12 | 17 | 7 | 4 | 5 | 2 | 19 | 9 | 20 | 15 | 11 | 6 | 3 | 1 | 14 | 16 | 18 | 8 | 10 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 17 | 12 | 10 | 15 | 18 | 4 | 14 | 6 | 16 | 3 | 13 | 11 | 1 | 5 | 19 | 8 | 20 | 9 | 2 | 7 |
| 16 | 20 | 11 | 2 | 8 | 18 | 4 | 14 | 7 | 12 | 3 | 5 | 15 | 19 | 13 | 10 | 6 | 17 | 9 | 1 |
| 14 | 6 | 8 | 17 | 4 | 13 | 7 | 5 | 9 | 20 | 12 | 10 | 11 | 19 | 1 | 15 | 16 | 2 | 18 | 3 |
| 20 | 18 | 2 | 1 | 3 | 12 | 14 | 7 | 10 | 11 | 6 | 17 | 15 | 16 | 5 | 9 | 8 | 13 | 4 | 19 |
| 19 | 3 | 1 | 17 | 4 | 5 | 13 | 10 | 7 | 9 | 11 | 18 | 2 | 12 | 8 | 20 | 15 | 6 | 14 | 16 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | 20 | 1 | 15 | 19 | 16 | 12 | 3 | 5 | 11 | 4 | 13 | 6 | 10 | 8 | 9 | 17 | 2 | 14 | 18 |
| 9 | 7 | 12 | 4 | 5 | 19 | 18 | 15 | 2 | 11 | 1 | 17 | 8 | 16 | 20 | 13 | 3 | 6 | 14 | 10 |
| 12 | 10 | 5 | 3 | 19 | 1 | 2 | 14 | 6 | 16 | 9 | 8 | 15 | 11 | 13 | 4 | 20 | 17 | 18 | 7 |
| 8 | 15 | 13 | 4 | 9 | 11 | 17 | 18 | 7 | 6 | 12 | 16 | 14 | 5 | 2 | 3 | 1 | 20 | 10 | 19 |
| 10 | 6 | 1 | 3 | 16 | 12 | 14 | 11 | 17 | 4 | 8 | 9 | 5 | 19 | 2 | 7 | 18 | 15 | 13 | 20 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 16 | 14 | 8 | 12 | 1 | 10 | 19 | 5 | 2 | 17 | 11 | 15 | 7 | 3 | 4 | 13 | 18 | 20 | 9 | 6 |
| 9 | 12 | 19 | 18 | 16 | 5 | 14 | 20 | 15 | 8 | 13 | 11 | 10 | 7 | 3 | 2 | 6 | 1 | 17 | 4 |
| 5 | 10 | 8 | 12 | 9 | 18 | 17 | 6 | 13 | 2 | 1 | 3 | 15 | 20 | 7 | 16 | 19 | 11 | 4 | 14 |
| 12 | 9 | 4 | 1 | 3 | 8 | 10 | 18 | 13 | 14 | 2 | 6 | 15 | 7 | 20 | 19 | 17 | 16 | 5 | 11 |
| 7 | 18 | 2 | 10 | 15 | 5 | 9 | 17 | 3 | 20 | 16 | 6 | 12 | 14 | 8 | 1 | 11 | 19 | 4 | 13 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 17 | 2 | 5 | 12 | 8 | 10 | 14 | 16 | 7 | 3 | 20 | 11 | 4 | 1 | 9 | 19 | 18 | 15 | 6 | 13 |
| 8 | 20 | 13 | 10 | 9 | 19 | 15 | 2 | 7 | 17 | 14 | 16 | 6 | 11 | 4 | 12 | 1 | 5 | 3 | 18 |
| 13 | 15 | 4 | 7 | 1 | 8 | 14 | 12 | 2 | 6 | 20 | 9 | 16 | 5 | 18 | 19 | 3 | 10 | 11 | 17 |
| 16 | 6 | 13 | 4 | 2 | 19 | 8 | 7 | 1 | 9 | 14 | 11 | 15 | 18 | 12 | 5 | 20 | 10 | 3 | 17 |
| 5 | 15 | 7 | 16 | 13 | 20 | 2 | 14 | 8 | 19 | 10 | 6 | 9 | 3 | 11 | 12 | 17 | 18 | 1 | 4 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 14 | 8 | 5 | 6 | 16 | 2 | 7 | 11 | 9 | 18 | 3 | 1 | 13 | 15 | 12 | 17 | 19 | 20 | 4 |
| 5 | 2 | 16 | 11 | 6 | 3 | 17 | 10 | 19 | 9 | 14 | 8 | 20 | 13 | 15 | 1 | 12 | 7 | 18 | 4 |
| 9 | 16 | 4 | 10 | 7 | 17 | 11 | 15 | 8 | 18 | 19 | 3 | 14 | 1 | 5 | 20 | 6 | 2 | 12 | 13 |
| 4 | 9 | 11 | 10 | 5 | 19 | 20 | 6 | 7 | 3 | 13 | 8 | 15 | 17 | 1 | 14 | 16 | 18 | 2 | 12 |
| 16 | 14 | 15 | 2 | 18 | 5 | 12 | 3 | 17 | 1 | 6 | 9 | 7 | 8 | 20 | 10 | 11 | 13 | 4 | 19 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 16 | 6 | 11 | 17 | 18 | 1 | 14 | 2 | 15 | 20 | 12 | 7 | 19 | 8 | 9 | 13 | 10 | 5 | 3 |
| 19 | 14 | 2 | 17 | 3 | 12 | 1 | 18 | 6 | 10 | 13 | 7 | 5 | 4 | 8 | 11 | 9 | 15 | 16 | 20 |
| 10 | 7 | 1 | 9 | 19 | 18 | 15 | 12 | 3 | 20 | 4 | 6 | 17 | 5 | 8 | 14 | 11 | 16 | 13 | 2 |
| 11 | 7 | 16 | 3 | 15 | 9 | 19 | 20 | 12 | 6 | 4 | 10 | 13 | 2 | 18 | 1 | 8 | 14 | 17 | 5 |
| 18 | 9 | 10 | 11 | 14 | 12 | 6 | 19 | 5 | 4 | 2 | 13 | 15 | 20 | 8 | 1 | 16 | 3 | 7 | 17 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 17 | 12 | 9 | 18 | 20 | 11 | 7 | 8 | 3 | 19 | 1 | 16 | 15 | 4 | 6 | 5 | 14 | 13 | 10 |
| 9 | 4 | 2 | 10 | 14 | 5 | 7 | 17 | 8 | 18 | 20 | 6 | 16 | 15 | 19 | 13 | 1 | 12 | 11 | 3 |
| 6 | 13 | 18 | 9 | 20 | 2 | 17 | 3 | 10 | 12 | 19 | 5 | 7 | 4 | 8 | 1 | 15 | 14 | 11 | 16 |
| 14 | 4 | 16 | 13 | 10 | 5 | 7 | 17 | 20 | 12 | 1 | 6 | 2 | 11 | 15 | 19 | 3 | 18 | 9 | 8 |
| 9 | 15 | 19 | 3 | 5 | 8 | 14 | 11 | 1 | 4 | 18 | 20 | 16 | 7 | 10 | 17 | 12 | 2 | 6 | 13 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7 | 20 | 10 | 2 | 19 | 15 | 5 | 6 | 18 | 1 | 9 | 11 | 14 | 4 | 3 | 13 | 16 | 12 | 17 | 8 |
| 17 | 4 | 10 | 19 | 12 | 8 | 5 | 16 | 6 | 18 | 15 | 2 | 11 | 7 | 13 | 9 | 20 | 3 | 14 | 1 |
| 16 | 14 | 19 | 15 | 9 | 18 | 8 | 13 | 3 | 11 | 2 | 20 | 4 | 12 | 10 | 6 | 17 | 7 | 1 | 5 |
| 6 | 18 | 15 | 13 | 11 | 8 | 10 | 12 | 1 | 19 | 14 | 17 | 2 | 9 | 16 | 3 | 20 | 4 | 7 | 5 |
| 16 | 15 | 13 | 1 | 19 | 12 | 10 | 18 | 7 | 3 | 6 | 20 | 2 | 4 | 9 | 11 | 14 | 17 | 8 | 5 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 9 | 20 | 8 | 4 | 2 | 7 | 5 | 16 | 17 | 18 | 6 | 14 | 11 | 19 | 12 | 15 | 10 | 13 | 3 |
| 13 | 4 | 1 | 9 | 6 | 8 | 3 | 2 | 18 | 19 | 17 | 20 | 10 | 15 | 14 | 11 | 7 | 5 | 12 | 16 |
| 6 | 15 | 3 | 1 | 16 | 19 | 13 | 2 | 17 | 10 | 11 | 5 | 20 | 7 | 4 | 14 | 8 | 9 | 12 | 18 |
| 16 | 1 | 10 | 11 | 15 | 5 | 14 | 18 | 12 | 3 | 19 | 7 | 8 | 9 | 4 | 6 | 17 | 20 | 2 | 13 |
| 1 | 14 | 10 | 7 | 2 | 8 | 15 | 5 | 20 | 11 | 17 | 9 | 19 | 18 | 6 | 16 | 4 | 3 | 12 | 13 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 19 | 11 | 16 | 1 | 5 | 18 | 9 | 3 | 20 | 12 | 17 | 15 | 10 | 4 | 6 | 7 | 8 | 13 | 14 |
| 18 | 15 | 13 | 17 | 2 | 4 | 11 | 1 | 14 | 6 | 16 | 20 | 7 | 12 | 3 | 8 | 10 | 19 | 9 | 5 |
| 10 | 14 | 9 | 7 | 12 | 8 | 1 | 2 | 19 | 20 | 13 | 4 | 17 | 16 | 6 | 18 | 3 | 11 | 15 | 5 |
| 5 | 13 | 6 | 3 | 19 | 20 | 16 | 12 | 4 | 18 | 9 | 10 | 15 | 17 | 1 | 8 | 7 | 14 | 11 | 2 |
| 6 | 13 | 20 | 11 | 2 | 16 | 18 | 12 | 3 | 5 | 8 | 10 | 19 | 9 | 15 | 14 | 1 | 7 | 4 | 17 |
| | | | | | | | | | | | | | | | | | | | |
| 10 | 6 | 1 | 2 | 5 | 12 | 8 | 3 | 7 | 9 | 11 | 19 | 16 | 18 | 4 | 14 | 13 | 20 | 17 | 15 |
| 7 | 2 | 20 | 18 | 10 | 3 | 5 | 4 | 9 | 14 | 16 | 17 | 6 | 13 | 12 | 19 | 15 | 1 | 8 | 11 |
| 15 | 11 | 17 | 19 | 2 | 9 | 16 | 7 | 4 | 20 | 18 | 6 | 1 | 12 | 5 | 8 | 10 | 3 | 13 | 14 |
| 20 | 5 | 4 | 9 | 15 | 11 | 2 | 17 | 13 | 6 | 12 | 1 | 14 | 8 | 16 | 10 | 19 | 7 | 3 | 18 |
| 5 | 13 | 19 | 12 | 15 | 20 | 4 | 2 | 17 | 9 | 7 | 10 | 14 | 6 | 16 | 8 | 3 | 11 | 18 | 1 |
| | | | | | | | | | | | | | | | | | | | |
| 12 | 13 | 5 | 6 | 8 | 9 | 7 | 15 | 2 | 18 | 20 | 4 | 14 | 3 | 11 | 16 | 17 | 19 | 1 | 10 |
| 6 | 2 | 10 | 18 | 9 | 16 | 20 | 13 | 17 | 5 | 4 | 12 | 1 | 11 | 7 | 19 | 15 | 8 | 3 | 14 |
| 15 | 11 | 10 | 8 | 20 | 6 | 7 | 1 | 9 | 4 | 2 | 14 | 18 | 19 | 3 | 16 | 13 | 12 | 5 | 17 |
| 17 | 15 | 19 | 13 | 7 | 12 | 20 | 8 | 14 | 16 | 9 | 11 | 6 | 2 | 3 | 1 | 4 | 18 | 5 | 10 |
| 7 | 2 | 9 | 19 | 3 | 18 | 6 | 13 | 10 | 20 | 11 | 12 | 4 | 17 | 16 | 8 | 1 | 5 | 14 | 15 |
| | | | | | | | | | | | | | | | | | | | |
| 2 | 16 | 4 | 6 | 17 | 18 | 14 | 9 | 12 | 3 | 1 | 5 | 15 | 8 | 19 | 10 | 11 | 20 | 13 | 7 |
| 17 | 3 | 2 | 20 | 12 | 18 | 19 | 13 | 6 | 10 | 14 | 16 | 1 | 4 | 11 | 7 | 9 | 8 | 5 | 15 |
| 8 | 5 | 2 | 17 | 10 | 15 | 12 | 18 | 1 | 20 | 14 | 3 | 13 | 11 | 19 | 4 | 6 | 9 | 16 | 7 |
| 11 | 19 | 20 | 1 | 6 | 13 | 2 | 7 | 3 | 17 | 4 | 8 | 12 | 16 | 15 | 10 | 18 | 9 | 5 | 14 |
| 17 | 7 | 6 | 1 | 8 | 15 | 18 | 20 | 5 | 10 | 12 | 4 | 3 | 9 | 16 | 19 | 2 | 13 | 14 | 11 |
| | | | | | | | | | | | | | | | | | | | |
| 12 | 6 | 9 | 14 | 2 | 17 | 4 | 20 | 16 | 10 | 7 | 15 | 1 | 8 | 19 | 13 | 5 | 18 | 3 | 11 |
| 11 | 17 | 9 | 19 | 2 | 5 | 15 | 20 | 6 | 16 | 4 | 1 | 10 | 7 | 8 | 3 | 13 | 18 | 12 | 14 |
| 20 | 11 | 5 | 1 | 9 | 17 | 10 | 19 | 7 | 2 | 4 | 15 | 16 | 14 | 6 | 12 | 13 | 3 | 18 | 8 |
| 11 | 17 | 20 | 1 | 10 | 2 | 12 | 18 | 3 | 6 | 4 | 13 | 16 | 19 | 7 | 9 | 14 | 15 | 8 | 5 |
| 6 | 3 | 10 | 15 | 11 | 4 | 20 | 2 | 5 | 18 | 9 | 19 | 12 | 7 | 1 | 16 | 17 | 8 | 14 | 13 |
| | | | | | | | | | | | | | | | | | | | |
| 2 | 4 | 12 | 16 | 3 | 15 | 6 | 5 | 11 | 10 | 17 | 19 | 13 | 9 | 7 | 18 | 14 | 1 | 20 | 8 |
| 14 | 8 | 16 | 4 | 13 | 19 | 6 | 11 | 3 | 9 | 1 | 7 | 2 | 10 | 17 | 15 | 5 | 20 | 18 | 12 |
| 17 | 1 | 8 | 12 | 15 | 5 | 4 | 16 | 20 | 19 | 6 | 2 | 7 | 14 | 9 | 11 | 3 | 18 | 10 | 13 |
| 5 | 6 | 13 | 9 | 4 | 19 | 16 | 17 | 15 | 11 | 2 | 10 | 7 | 8 | 12 | 3 | 1 | 20 | 14 | 18 |
| 13 | 19 | 20 | 2 | 7 | 9 | 15 | 18 | 4 | 6 | 8 | 1 | 5 | 14 | 16 | 3 | 11 | 17 | 12 | 10 |
| | | | | | | | | | | | | | | | | | | | |
| 2 | 4 | 14 | 10 | 13 | 18 | 11 | 19 | 15 | 16 | 7 | 9 | 3 | 8 | 6 | 17 | 12 | 5 | 20 | 1 |
| 9 | 1 | 13 | 18 | 16 | 8 | 15 | 17 | 4 | 6 | 7 | 20 | 14 | 5 | 10 | 12 | 3 | 11 | 2 | 19 |
| 5 | 2 | 11 | 7 | 1 | 16 | 20 | 14 | 3 | 4 | 15 | 13 | 9 | 8 | 12 | 17 | 19 | 6 | 10 | 18 |
| 8 | 16 | 18 | 4 | 10 | 20 | 2 | 19 | 5 | 7 | 13 | 1 | 11 | 17 | 15 | 12 | 6 | 9 | 14 | 3 |
| 9 | 6 | 16 | 12 | 18 | 19 | 17 | 8 | 10 | 5 | 3 | 4 | 13 | 20 | 1 | 15 | 2 | 11 | 14 | 7 |
| | | | | | | | | | | | | | | | | | | | |
| 2 | 13 | 4 | 9 | 14 | 8 | 10 | 3 | 16 | 17 | 5 | 20 | 1 | 7 | 19 | 11 | 6 | 18 | 15 | 12 |
| 20 | 15 | 8 | 11 | 12 | 18 | 5 | 19 | 3 | 4 | 9 | 13 | 17 | 1 | 2 | 10 | 6 | 16 | 7 | 14 |
| 13 | 16 | 19 | 20 | 15 | 5 | 3 | 4 | 11 | 2 | 17 | 12 | 14 | 7 | 8 | 1 | 10 | 9 | 6 | 18 |
| 15 | 7 | 20 | 16 | 1 | 14 | 5 | 3 | 12 | 17 | 19 | 4 | 9 | 2 | 11 | 8 | 10 | 18 | 6 | 13 |
| 7 | 15 | 16 | 5 | 6 | 14 | 11 | 4 | 20 | 10 | 12 | 1 | 18 | 19 | 9 | 17 | 8 | 3 | 2 | 13 |
| | | | | | | | | | | | | | | | | | | | |
| 14 | 18 | 13 | 1 | 8 | 3 | 11 | 7 | 20 | 15 | 16 | 5 | 19 | 9 | 12 | 10 | 4 | 2 | 17 | 6 |
| 8 | 12 | 7 | 9 | 6 | 14 | 10 | 5 | 16 | 13 | 4 | 19 | 20 | 17 | 11 | 1 | 18 | 2 | 15 | 3 |
| 14 | 6 | 9 | 3 | 2 | 18 | 8 | 16 | 13 | 12 | 10 | 4 | 15 | 5 | 17 | 1 | 20 | 7 | 19 | 11 |
| 6 | 3 | 15 | 13 | 2 | 1 | 4 | 18 | 10 | 7 | 19 | 11 | 12 | 16 | 5 | 17 | 8 | 20 | 9 | 14 |
| 7 | 14 | 3 | 8 | 13 | 2 | 15 | 6 | 16 | 19 | 20 | 1 | 18 | 5 | 4 | 12 | 17 | 11 | 10 | 9 |
| | | | | | | | | | | | | | | | | | | | |
| 11 | 12 | 6 | 5 | 9 | 10 | 3 | 20 | 4 | 16 | 1 | 18 | 15 | 19 | 7 | 8 | 17 | 2 | 13 | 14 |
| 4 | 20 | 17 | 7 | 10 | 9 | 1 | 3 | 16 | 8 | 6 | 2 | 18 | 14 | 19 | 12 | 15 | 5 | 11 | 13 |
| 2 | 18 | 20 | 6 | 4 | 3 | 7 | 16 | 17 | 13 | 12 | 5 | 14 | 15 | 19 | 11 | 9 | 10 | 1 | 8 |
| 8 | 3 | 13 | 18 | 11 | 9 | 4 | 20 | 1 | 14 | 7 | 16 | 12 | 17 | 6 | 15 | 10 | 5 | 19 | 2 |
| 20 | 18 | 11 | 1 | 15 | 7 | 12 | 2 | 10 | 6 | 17 | 14 | 8 | 19 | 9 | 13 | 16 | 5 | 3 | 4 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 7 | 2 | 8 | 6 | 10 | 11 | 17 | 12 | 16 | 18 | 15 | 4 | 19 | 20 | 9 | 13 | 1 | 14 | 3 |
| 13 | 20 | 16 | 17 | 9 | 19 | 4 | 15 | 7 | 10 | 14 | 3 | 2 | 11 | 12 | 18 | 6 | 8 | 5 | 1 |
| 14 | 3 | 13 | 16 | 5 | 2 | 15 | 9 | 17 | 19 | 12 | 6 | 18 | 7 | 11 | 20 | 10 | 8 | 1 | 4 |
| 13 | 9 | 11 | 6 | 3 | 2 | 14 | 16 | 18 | 15 | 10 | 20 | 4 | 1 | 7 | 8 | 12 | 17 | 5 | 19 |
| 4 | 7 | 1 | 11 | 5 | 18 | 15 | 6 | 10 | 13 | 9 | 14 | 3 | 19 | 2 | 16 | 20 | 17 | 12 | 8 |
| | | | | | | | | | | | | | | | | | | | |
| 6 | 18 | 2 | 9 | 19 | 3 | 17 | 1 | 13 | 4 | 12 | 20 | 15 | 14 | 7 | 11 | 8 | 16 | 5 | 10 |
| 20 | 7 | 12 | 13 | 2 | 1 | 6 | 11 | 14 | 5 | 19 | 9 | 4 | 15 | 18 | 8 | 10 | 16 | 3 | 17 |
| 5 | 12 | 4 | 20 | 13 | 3 | 8 | 17 | 19 | 16 | 18 | 14 | 2 | 10 | 15 | 9 | 1 | 11 | 7 | 6 |
| 1 | 8 | 2 | 17 | 7 | 11 | 12 | 20 | 14 | 5 | 10 | 16 | 15 | 3 | 9 | 4 | 19 | 18 | 13 | 6 |
| 10 | 2 | 18 | 11 | 13 | 12 | 8 | 4 | 14 | 16 | 17 | 3 | 20 | 5 | 15 | 9 | 6 | 1 | 7 | 19 |
| | | | | | | | | | | | | | | | | | | | |
| 5 | 16 | 9 | 3 | 13 | 14 | 2 | 15 | 17 | 8 | 4 | 6 | 10 | 18 | 11 | 1 | 19 | 12 | 20 | 7 |
| 7 | 19 | 12 | 18 | 15 | 2 | 8 | 16 | 4 | 17 | 11 | 20 | 3 | 5 | 14 | 9 | 10 | 6 | 13 | 1 |
| 18 | 12 | 2 | 19 | 1 | 16 | 3 | 5 | 6 | 15 | 17 | 11 | 10 | 20 | 8 | 4 | 7 | 14 | 9 | 13 |
| 3 | 14 | 11 | 18 | 7 | 8 | 6 | 12 | 1 | 2 | 5 | 19 | 4 | 15 | 20 | 9 | 17 | 13 | 10 | 16 |
| 16 | 11 | 8 | 18 | 4 | 7 | 6 | 14 | 12 | 17 | 5 | 9 | 19 | 10 | 15 | 13 | 2 | 1 | 3 | 20 |
| | | | | | | | | | | | | | | | | | | | |
| 17 | 11 | 15 | 13 | 1 | 14 | 12 | 6 | 9 | 18 | 4 | 19 | 7 | 16 | 3 | 20 | 5 | 2 | 10 | 8 |
| 1 | 20 | 8 | 15 | 14 | 10 | 13 | 9 | 12 | 19 | 17 | 18 | 11 | 6 | 4 | 3 | 5 | 2 | 7 | 16 |
| 17 | 7 | 15 | 2 | 18 | 12 | 5 | 11 | 9 | 1 | 16 | 6 | 4 | 20 | 13 | 3 | 19 | 14 | 10 | 8 |
| 4 | 16 | 1 | 11 | 15 | 9 | 13 | 3 | 6 | 7 | 10 | 20 | 12 | 19 | 8 | 18 | 2 | 14 | 17 | 5 |
| 1 | 16 | 8 | 13 | 10 | 19 | 12 | 15 | 11 | 17 | 18 | 5 | 2 | 4 | 20 | 9 | 3 | 7 | 6 | 14 |
| | | | | | | | | | | | | | | | | | | | |
| 4 | 16 | 12 | 9 | 11 | 7 | 5 | 2 | 13 | 14 | 1 | 19 | 3 | 15 | 20 | 18 | 6 | 10 | 8 | 17 |
| 1 | 14 | 15 | 18 | 19 | 13 | 3 | 12 | 11 | 16 | 10 | 9 | 7 | 4 | 8 | 20 | 5 | 17 | 2 | 6 |
| 7 | 1 | 15 | 14 | 6 | 11 | 3 | 2 | 12 | 4 | 10 | 19 | 9 | 5 | 18 | 17 | 20 | 16 | 13 | 8 |
| 7 | 20 | 8 | 18 | 17 | 2 | 16 | 9 | 12 | 1 | 4 | 13 | 3 | 15 | 11 | 14 | 19 | 6 | 10 | 5 |
| 10 | 1 | 16 | 6 | 18 | 2 | 4 | 15 | 3 | 9 | 11 | 17 | 12 | 5 | 7 | 19 | 8 | 13 | 14 | 20 |
| | | | | | | | | | | | | | | | | | | | |
| 6 | 8 | 16 | 17 | 11 | 13 | 4 | 1 | 20 | 2 | 12 | 9 | 7 | 18 | 5 | 10 | 19 | 15 | 14 | 3 |
| 8 | 5 | 7 | 2 | 20 | 1 | 11 | 19 | 12 | 16 | 10 | 13 | 17 | 4 | 3 | 15 | 18 | 6 | 14 | 9 |
| 20 | 14 | 4 | 7 | 18 | 13 | 12 | 11 | 17 | 16 | 19 | 5 | 3 | 9 | 10 | 8 | 2 | 1 | 6 | 15 |
| 8 | 16 | 11 | 15 | 3 | 20 | 9 | 6 | 13 | 14 | 17 | 12 | 7 | 5 | 2 | 4 | 19 | 1 | 18 | 10 |
| 4 | 3 | 17 | 18 | 12 | 5 | 9 | 2 | 8 | 14 | 7 | 6 | 16 | 15 | 10 | 19 | 11 | 20 | 13 | 1 |
| | | | | | | | | | | | | | | | | | | | |
| 5 | 16 | 10 | 11 | 19 | 4 | 12 | 1 | 3 | 9 | 2 | 20 | 6 | 14 | 18 | 8 | 13 | 17 | 15 | 7 |
| 10 | 5 | 15 | 17 | 9 | 14 | 7 | 16 | 11 | 12 | 19 | 3 | 8 | 20 | 4 | 13 | 6 | 1 | 2 | 18 |
| 11 | 15 | 8 | 14 | 9 | 5 | 13 | 18 | 7 | 17 | 4 | 6 | 3 | 10 | 16 | 12 | 20 | 1 | 2 | 19 |
| 1 | 13 | 19 | 17 | 2 | 11 | 16 | 6 | 9 | 18 | 4 | 3 | 8 | 12 | 5 | 10 | 7 | 15 | 14 | 20 |
| 9 | 10 | 2 | 13 | 18 | 8 | 4 | 5 | 3 | 16 | 15 | 7 | 14 | 19 | 17 | 20 | 11 | 12 | 6 | 1 |
| | | | | | | | | | | | | | | | | | | | |
| 20 | 4 | 12 | 13 | 3 | 5 | 14 | 17 | 10 | 15 | 2 | 9 | 16 | 1 | 18 | 19 | 11 | 7 | 6 | 8 |
| 18 | 8 | 14 | 17 | 9 | 1 | 11 | 10 | 12 | 4 | 13 | 6 | 3 | 16 | 7 | 19 | 20 | 5 | 2 | 15 |
| 5 | 20 | 9 | 13 | 17 | 14 | 19 | 2 | 8 | 7 | 1 | 6 | 12 | 4 | 18 | 15 | 10 | 3 | 11 | 16 |
| 18 | 8 | 2 | 12 | 1 | 7 | 16 | 19 | 5 | 9 | 13 | 3 | 20 | 4 | 6 | 11 | 15 | 10 | 17 | 14 |
| 12 | 4 | 1 | 15 | 10 | 11 | 18 | 19 | 7 | 13 | 17 | 9 | 16 | 8 | 14 | 20 | 2 | 5 | 3 | 6 |
| | | | | | | | | | | | | | | | | | | | |
| 6 | 20 | 3 | 10 | 13 | 19 | 14 | 12 | 4 | 7 | 2 | 1 | 9 | 18 | 11 | 5 | 15 | 16 | 8 | 17 |
| 7 | 9 | 16 | 15 | 17 | 1 | 11 | 13 | 14 | 5 | 6 | 2 | 19 | 18 | 4 | 3 | 8 | 12 | 20 | 10 |
| 13 | 5 | 19 | 3 | 10 | 8 | 20 | 9 | 1 | 2 | 6 | 18 | 7 | 14 | 15 | 17 | 4 | 12 | 11 | 16 |
| 15 | 2 | 7 | 6 | 4 | 17 | 8 | 1 | 13 | 11 | 10 | 18 | 20 | 5 | 12 | 19 | 3 | 9 | 14 | 16 |
| 18 | 7 | 16 | 6 | 2 | 12 | 3 | 13 | 9 | 15 | 8 | 11 | 5 | 1 | 19 | 20 | 10 | 14 | 4 | 17 |
| | | | | | | | | | | | | | | | | | | | |
| 19 | 5 | 15 | 7 | 10 | 2 | 16 | 11 | 8 | 4 | 14 | 1 | 9 | 12 | 20 | 17 | 18 | 13 | 6 | 3 |
| 3 | 11 | 1 | 2 | 10 | 12 | 5 | 19 | 8 | 18 | 16 | 14 | 13 | 9 | 15 | 17 | 4 | 6 | 20 | 7 |
| 18 | 7 | 14 | 5 | 13 | 2 | 1 | 8 | 20 | 9 | 19 | 17 | 4 | 3 | 11 | 16 | 6 | 15 | 12 | 10 |
| 11 | 15 | 16 | 17 | 7 | 4 | 9 | 19 | 5 | 3 | 8 | 18 | 10 | 20 | 14 | 6 | 2 | 12 | 13 | 1 |
| 6 | 17 | 20 | 9 | 14 | 12 | 1 | 10 | 15 | 19 | 5 | 11 | 16 | 13 | 3 | 7 | 4 | 2 | 18 | 8 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 16 | 6 | 3 | 15 | 13 | 17 | 14 | 12 | 20 | 11 | 10 | 4 | 9 | 19 | 18 | 7 | 8 | 5 | 2 |
| 2 | 1 | 13 | 17 | 14 | 15 | 3 | 20 | 11 | 18 | 6 | 12 | 19 | 16 | 7 | 4 | 5 | 10 | 8 | 9 |
| 20 | 8 | 14 | 2 | 18 | 11 | 16 | 1 | 13 | 12 | 6 | 4 | 15 | 7 | 3 | 19 | 9 | 10 | 5 | 17 |
| 3 | 9 | 4 | 8 | 12 | 5 | 19 | 6 | 15 | 17 | 10 | 2 | 20 | 16 | 13 | 18 | 14 | 11 | 7 | 1 |
| 12 | 9 | 18 | 1 | 10 | 4 | 16 | 8 | 7 | 20 | 17 | 2 | 11 | 6 | 13 | 14 | 3 | 15 | 5 | 19 |
| 18 | 13 | 11 | 10 | 3 | 17 | 19 | 20 | 8 | 15 | 16 | 1 | 5 | 7 | 6 | 12 | 9 | 4 | 14 | 2 |
| 18 | 9 | 4 | 14 | 11 | 17 | 2 | 1 | 8 | 5 | 15 | 16 | 20 | 7 | 6 | 10 | 13 | 12 | 19 | 3 |
| 9 | 11 | 5 | 4 | 12 | 7 | 1 | 19 | 16 | 18 | 20 | 13 | 10 | 2 | 3 | 6 | 14 | 15 | 8 | 17 |
| 14 | 15 | 6 | 20 | 7 | 8 | 19 | 13 | 1 | 5 | 4 | 16 | 9 | 3 | 10 | 11 | 18 | 17 | 12 | 2 |
| 19 | 15 | 10 | 1 | 12 | 16 | 9 | 8 | 13 | 2 | 3 | 6 | 20 | 11 | 7 | 5 | 4 | 17 | 14 | 18 |
| 20 | 5 | 17 | 7 | 14 | 1 | 6 | 16 | 4 | 2 | 13 | 9 | 3 | 10 | 8 | 12 | 15 | 19 | 11 | 18 |
| 7 | 9 | 12 | 8 | 20 | 11 | 6 | 16 | 2 | 3 | 4 | 15 | 1 | 17 | 19 | 5 | 14 | 13 | 18 | 10 |
| 11 | 15 | 9 | 6 | 18 | 3 | 2 | 19 | 13 | 20 | 12 | 16 | 14 | 4 | 17 | 5 | 7 | 10 | 8 | 1 |
| 4 | 3 | 17 | 14 | 2 | 18 | 7 | 12 | 16 | 10 | 11 | 15 | 5 | 20 | 13 | 1 | 8 | 9 | 19 | 6 |
| 8 | 6 | 17 | 4 | 10 | 5 | 9 | 3 | 18 | 20 | 15 | 16 | 13 | 1 | 11 | 12 | 14 | 7 | 19 | 2 |
| 11 | 5 | 15 | 9 | 1 | 12 | 6 | 16 | 17 | 3 | 20 | 13 | 14 | 18 | 8 | 7 | 19 | 10 | 4 | 2 |
| 4 | 12 | 13 | 9 | 1 | 7 | 15 | 3 | 6 | 19 | 2 | 8 | 16 | 11 | 17 | 18 | 5 | 20 | 10 | 14 |
| 8 | 15 | 5 | 4 | 13 | 7 | 19 | 10 | 20 | 12 | 17 | 2 | 11 | 6 | 9 | 16 | 3 | 18 | 1 | 14 |
| 3 | 12 | 11 | 18 | 2 | 15 | 13 | 14 | 8 | 19 | 7 | 20 | 10 | 16 | 9 | 6 | 1 | 17 | 4 | 5 |
| 19 | 1 | 10 | 6 | 9 | 2 | 20 | 5 | 18 | 17 | 11 | 16 | 8 | 14 | 4 | 12 | 15 | 13 | 7 | 3 |
| 5 | 2 | 1 | 7 | 14 | 17 | 4 | 16 | 9 | 3 | 13 | 19 | 15 | 18 | 10 | 12 | 20 | 8 | 11 | 6 |
| 2 | 12 | 18 | 13 | 4 | 10 | 5 | 6 | 14 | 16 | 15 | 19 | 9 | 7 | 3 | 17 | 1 | 11 | 20 | 8 |
| 15 | 13 | 2 | 18 | 6 | 19 | 7 | 10 | 11 | 16 | 20 | 1 | 12 | 3 | 14 | 9 | 17 | 4 | 8 | 5 |
| 4 | 12 | 17 | 1 | 7 | 9 | 16 | 8 | 14 | 20 | 10 | 13 | 6 | 2 | 19 | 3 | 15 | 18 | 11 | 5 |
| 14 | 17 | 16 | 7 | 19 | 9 | 8 | 3 | 15 | 10 | 5 | 13 | 11 | 18 | 2 | 20 | 4 | 1 | 12 | 6 |
| 15 | 8 | 13 | 19 | 6 | 18 | 5 | 7 | 10 | 11 | 3 | 14 | 2 | 1 | 16 | 9 | 4 | 12 | 17 | 20 |
| 20 | 8 | 12 | 11 | 7 | 10 | 2 | 13 | 4 | 16 | 19 | 15 | 1 | 9 | 17 | 6 | 3 | 14 | 18 | 5 |
| 11 | 4 | 9 | 5 | 15 | 10 | 19 | 12 | 13 | 7 | 2 | 6 | 3 | 16 | 17 | 14 | 18 | 8 | 1 | 20 |
| 9 | 19 | 15 | 13 | 2 | 12 | 8 | 14 | 7 | 4 | 20 | 6 | 1 | 5 | 11 | 3 | 17 | 18 | 16 | 10 |
| 1 | 2 | 17 | 13 | 3 | 7 | 9 | 5 | 6 | 20 | 14 | 10 | 4 | 12 | 18 | 19 | 8 | 11 | 15 | 16 |
| 18 | 12 | 16 | 19 | 17 | 9 | 5 | 11 | 6 | 13 | 8 | 14 | 3 | 2 | 15 | 20 | 7 | 1 | 10 | 4 |
| 11 | 16 | 19 | 6 | 12 | 7 | 2 | 14 | 15 | 18 | 4 | 8 | 5 | 3 | 9 | 17 | 20 | 1 | 13 | 10 |
| 4 | 19 | 12 | 17 | 7 | 6 | 14 | 20 | 5 | 11 | 9 | 13 | 18 | 8 | 10 | 2 | 3 | 15 | 1 | 16 |
| 3 | 14 | 7 | 13 | 17 | 6 | 9 | 8 | 15 | 19 | 2 | 20 | 12 | 11 | 18 | 4 | 10 | 1 | 5 | 16 |
| 6 | 4 | 3 | 13 | 12 | 10 | 11 | 2 | 15 | 7 | 9 | 20 | 8 | 1 | 19 | 18 | 5 | 17 | 16 | 14 |
| 3 | 6 | 4 | 13 | 19 | 1 | 7 | 15 | 8 | 10 | 12 | 18 | 11 | 9 | 20 | 5 | 2 | 16 | 14 | 17 |
| 10 | 6 | 20 | 14 | 19 | 4 | 1 | 8 | 16 | 2 | 13 | 3 | 7 | 9 | 12 | 5 | 11 | 15 | 17 | 18 |
| 17 | 7 | 8 | 16 | 4 | 6 | 9 | 19 | 11 | 5 | 12 | 13 | 10 | 18 | 20 | 1 | 14 | 15 | 2 | 3 |
| 12 | 16 | 9 | 3 | 4 | 18 | 19 | 10 | 8 | 6 | 20 | 14 | 13 | 7 | 5 | 15 | 11 | 2 | 17 | 1 |
| 3 | 5 | 15 | 9 | 6 | 19 | 4 | 17 | 2 | 14 | 11 | 20 | 13 | 10 | 7 | 18 | 12 | 8 | 16 | 1 |
| 8 | 7 | 2 | 12 | 17 | 19 | 9 | 5 | 20 | 13 | 11 | 18 | 1 | 3 | 4 | 15 | 10 | 14 | 16 | 6 |
| 20 | 13 | 4 | 6 | 7 | 9 | 2 | 16 | 11 | 14 | 5 | 19 | 3 | 17 | 12 | 18 | 10 | 1 | 8 | 15 |
| 14 | 19 | 5 | 6 | 3 | 16 | 10 | 8 | 12 | 17 | 20 | 9 | 13 | 18 | 2 | 11 | 7 | 4 | 15 | 1 |
| 19 | 3 | 12 | 11 | 13 | 14 | 7 | 18 | 5 | 1 | 2 | 17 | 15 | 6 | 9 | 20 | 4 | 8 | 16 | 10 |
| 5 | 11 | 15 | 17 | 9 | 6 | 1 | 10 | 20 | 19 | 16 | 8 | 14 | 2 | 7 | 3 | 13 | 18 | 12 | 4 |
| 14 | 3 | 11 | 2 | 6 | 1 | 15 | 19 | 8 | 9 | 12 | 10 | 16 | 7 | 5 | 17 | 20 | 13 | 4 | 18 |
| 9 | 15 | 18 | 6 | 14 | 13 | 12 | 2 | 1 | 4 | 8 | 3 | 19 | 5 | 20 | 11 | 7 | 10 | 16 | 17 |
| 20 | 13 | 16 | 18 | 3 | 9 | 14 | 5 | 8 | 12 | 17 | 10 | 15 | 7 | 19 | 2 | 11 | 6 | 4 | 1 |
| 8 | 19 | 15 | 20 | 9 | 1 | 18 | 7 | 17 | 14 | 4 | 16 | 11 | 5 | 6 | 10 | 13 | 2 | 12 | 3 |
| 18 | 15 | 16 | 1 | 20 | 14 | 11 | 13 | 9 | 2 | 5 | 8 | 10 | 17 | 12 | 7 | 19 | 4 | 3 | 6 |

Таблица 4 (продолжение)

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 16 | 2 | 15 | 3 | 8 | 11 | 12 | 1 | 14 | 19 | 7 | 17 | 20 | 4 | 6 | 9 | 10 | 13 | 5 | 18 |
| 6 | 9 | 4 | 1 | 2 | 15 | 18 | 16 | 8 | 17 | 13 | 3 | 14 | 20 | 12 | 5 | 19 | 7 | 10 | 11 |
| 12 | 10 | 4 | 5 | 16 | 1 | 7 | 19 | 8 | 11 | 2 | 3 | 14 | 6 | 18 | 13 | 15 | 9 | 20 | 17 |
| 17 | 12 | 14 | 7 | 9 | 6 | 19 | 8 | 16 | 5 | 11 | 1 | 20 | 15 | 3 | 4 | 18 | 13 | 10 | 2 |
| 5 | 4 | 10 | 13 | 15 | 20 | 7 | 3 | 12 | 9 | 14 | 1 | 17 | 19 | 6 | 16 | 18 | 2 | 8 | 11 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 8 | 2 | 9 | 16 | 7 | 11 | 20 | 17 | 19 | 5 | 18 | 13 | 6 | 4 | 15 | 1 | 12 | 3 | 14 |
| 10 | 3 | 9 | 14 | 5 | 8 | 1 | 6 | 17 | 7 | 15 | 12 | 2 | 4 | 16 | 18 | 13 | 20 | 11 | 19 |
| 14 | 8 | 16 | 6 | 19 | 7 | 5 | 18 | 4 | 17 | 2 | 20 | 1 | 13 | 12 | 15 | 9 | 11 | 3 | 10 |
| 20 | 14 | 3 | 11 | 5 | 8 | 19 | 18 | 4 | 7 | 17 | 2 | 16 | 13 | 12 | 10 | 1 | 15 | 9 | 6 |
| 1 | 9 | 19 | 16 | 11 | 15 | 7 | 8 | 14 | 10 | 12 | 18 | 4 | 2 | 6 | 20 | 17 | 13 | 5 | 3 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | 3 | 7 | 18 | 17 | 1 | 5 | 4 | 8 | 14 | 20 | 2 | 12 | 15 | 10 | 19 | 11 | 13 | 9 | 16 |
| 18 | 13 | 4 | 16 | 1 | 3 | 6 | 9 | 11 | 8 | 19 | 5 | 17 | 2 | 20 | 7 | 14 | 15 | 12 | 10 |
| 13 | 12 | 3 | 18 | 19 | 10 | 6 | 9 | 14 | 15 | 20 | 7 | 17 | 2 | 8 | 11 | 16 | 5 | 4 | 1 |
| 12 | 18 | 5 | 1 | 16 | 11 | 6 | 4 | 2 | 17 | 14 | 3 | 20 | 10 | 7 | 13 | 19 | 8 | 15 | 9 |
| 12 | 6 | 13 | 4 | 2 | 1 | 18 | 20 | 14 | 9 | 3 | 16 | 8 | 17 | 19 | 10 | 5 | 11 | 15 | 7 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 18 | 3 | 4 | 7 | 15 | 14 | 16 | 1 | 10 | 12 | 6 | 13 | 20 | 8 | 2 | 19 | 9 | 17 | 11 | 5 |
| 10 | 9 | 8 | 4 | 5 | 13 | 19 | 12 | 2 | 6 | 20 | 1 | 15 | 18 | 16 | 7 | 14 | 3 | 11 | 17 |
| 15 | 16 | 17 | 4 | 11 | 20 | 10 | 2 | 12 | 3 | 13 | 19 | 6 | 18 | 7 | 8 | 1 | 5 | 9 | 14 |
| 5 | 18 | 11 | 16 | 10 | 3 | 13 | 7 | 4 | 12 | 8 | 6 | 17 | 2 | 15 | 20 | 19 | 14 | 1 | 9 |
| 13 | 10 | 17 | 6 | 12 | 9 | 3 | 5 | 16 | 18 | 19 | 7 | 1 | 8 | 2 | 20 | 14 | 11 | 15 | 4 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 6 | 19 | 8 | 3 | 17 | 20 | 4 | 2 | 18 | 15 | 14 | 10 | 7 | 11 | 16 | 9 | 1 | 13 | 12 |
| 2 | 7 | 10 | 17 | 15 | 3 | 18 | 5 | 11 | 12 | 4 | 14 | 9 | 20 | 13 | 16 | 1 | 6 | 19 | 8 |
| 8 | 2 | 10 | 13 | 1 | 11 | 15 | 6 | 9 | 19 | 17 | 16 | 20 | 7 | 12 | 5 | 18 | 14 | 4 | 3 |
| 8 | 1 | 3 | 15 | 10 | 2 | 7 | 13 | 19 | 9 | 12 | 11 | 18 | 17 | 20 | 16 | 14 | 4 | 6 | 5 |
| 15 | 2 | 19 | 1 | 16 | 18 | 13 | 8 | 5 | 14 | 4 | 7 | 11 | 3 | 17 | 12 | 9 | 10 | 20 | 6 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | 14 | 18 | 5 | 2 | 10 | 11 | 7 | 15 | 16 | 13 | 17 | 12 | 9 | 8 | 6 | 4 | 3 | 19 | 1 |
| 18 | 13 | 6 | 9 | 1 | 5 | 19 | 2 | 16 | 17 | 10 | 15 | 7 | 11 | 20 | 14 | 8 | 4 | 3 | 12 |
| 9 | 13 | 20 | 16 | 11 | 8 | 12 | 15 | 7 | 14 | 4 | 10 | 6 | 17 | 2 | 3 | 18 | 5 | 19 | 1 |
| 7 | 10 | 4 | 8 | 3 | 1 | 14 | 15 | 2 | 6 | 17 | 20 | 11 | 5 | 19 | 12 | 18 | 13 | 9 | 16 |
| 16 | 18 | 12 | 7 | 13 | 14 | 4 | 10 | 15 | 11 | 6 | 9 | 1 | 5 | 17 | 19 | 3 | 20 | 8 | 2 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | 8 | 11 | 2 | 5 | 19 | 15 | 3 | 6 | 12 | 10 | 18 | 4 | 16 | 14 | 9 | 13 | 1 | 17 | 7 |
| 12 | 15 | 16 | 8 | 7 | 2 | 18 | 4 | 6 | 1 | 19 | 20 | 3 | 10 | 5 | 17 | 13 | 14 | 11 | 9 |
| 11 | 13 | 2 | 10 | 3 | 17 | 1 | 7 | 15 | 5 | 8 | 6 | 12 | 18 | 4 | 9 | 14 | 20 | 19 | 16 |
| 1 | 8 | 5 | 12 | 9 | 16 | 11 | 20 | 3 | 10 | 7 | 18 | 2 | 17 | 14 | 19 | 6 | 4 | 15 | 13 |
| 14 | 10 | 17 | 19 | 5 | 11 | 7 | 6 | 1 | 12 | 16 | 2 | 20 | 15 | 3 | 8 | 13 | 9 | 4 | 18 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | 6 | 10 | 8 | 12 | 3 | 11 | 9 | 7 | 13 | 14 | 16 | 2 | 15 | 19 | 18 | 17 | 4 | 5 | 1 |
| 3 | 6 | 1 | 19 | 18 | 5 | 7 | 15 | 13 | 9 | 17 | 8 | 16 | 2 | 10 | 4 | 20 | 14 | 12 | 11 |
| 9 | 1 | 12 | 20 | 17 | 4 | 11 | 10 | 16 | 6 | 18 | 5 | 2 | 15 | 8 | 7 | 19 | 14 | 13 | 3 |
| 17 | 11 | 6 | 16 | 9 | 7 | 20 | 5 | 15 | 8 | 12 | 1 | 18 | 4 | 19 | 13 | 2 | 3 | 14 | 10 |
| 7 | 11 | 6 | 13 | 19 | 12 | 16 | 10 | 17 | 5 | 8 | 20 | 3 | 18 | 2 | 4 | 14 | 9 | 15 | 1 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 18 | 3 | 2 | 16 | 20 | 1 | 14 | 15 | 12 | 6 | 19 | 8 | 5 | 9 | 10 | 7 | 11 | 4 | 13 | 17 |
| 17 | 9 | 16 | 10 | 13 | 12 | 5 | 4 | 7 | 20 | 19 | 11 | 15 | 6 | 1 | 3 | 2 | 14 | 18 | 8 |
| 1 | 7 | 17 | 3 | 15 | 6 | 20 | 2 | 8 | 10 | 12 | 5 | 18 | 4 | 11 | 19 | 14 | 16 | 9 | 13 |
| 15 | 6 | 18 | 17 | 20 | 8 | 3 | 14 | 5 | 4 | 7 | 13 | 10 | 2 | 1 | 19 | 9 | 12 | 16 | 11 |
| 8 | 7 | 1 | 17 | 10 | 5 | 13 | 16 | 3 | 18 | 15 | 12 | 4 | 2 | 11 | 9 | 19 | 14 | 6 | 20 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 16 | 20 | 15 | 1 | 13 | 4 | 17 | 12 | 19 | 2 | 5 | 8 | 18 | 11 | 9 | 7 | 3 | 6 | 14 | 10 |
| 2 | 13 | 12 | 3 | 5 | 4 | 9 | 10 | 17 | 14 | 8 | 16 | 11 | 18 | 6 | 15 | 20 | 19 | 7 | 1 |
| 7 | 11 | 20 | 5 | 10 | 18 | 8 | 12 | 6 | 4 | 19 | 13 | 1 | 15 | 17 | 9 | 3 | 16 | 14 | 2 |
| 6 | 13 | 17 | 10 | 2 | 15 | 5 | 19 | 20 | 7 | 3 | 11 | 14 | 4 | 9 | 18 | 12 | 1 | 8 | 16 |
| 6 | 18 | 9 | 17 | 14 | 15 | 1 | 8 | 13 | 11 | 7 | 20 | 2 | 4 | 10 | 19 | 5 | 16 | 12 | 3 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 2 | 18 | 9 | 7 | 1 | 5 | 17 | 16 | 3 | 20 | 13 | 12 | 8 | 19 | 10 | 6 | 15 | 11 | 14 |
| 18 | 1 | 17 | 3 | 15 | 6 | 4 | 14 | 20 | 12 | 9 | 2 | 19 | 8 | 7 | 13 | 16 | 11 | 5 | 10 |
| 5 | 8 | 1 | 18 | 16 | 17 | 13 | 12 | 2 | 7 | 10 | 15 | 6 | 11 | 4 | 19 | 14 | 20 | 9 | 3 |
| 5 | 18 | 11 | 1 | 9 | 19 | 3 | 10 | 4 | 15 | 8 | 2 | 20 | 6 | 14 | 12 | 7 | 17 | 13 | 16 |
| 13 | 1 | 6 | 9 | 4 | 15 | 12 | 10 | 11 | 3 | 20 | 18 | 5 | 8 | 17 | 19 | 7 | 16 | 14 | 2 |
| 15 | 14 | 11 | 19 | 6 | 10 | 12 | 8 | 4 | 18 | 13 | 20 | 9 | 3 | 17 | 16 | 2 | 7 | 5 | 1 |
| 13 | 12 | 16 | 5 | 10 | 4 | 19 | 15 | 18 | 1 | 2 | 8 | 20 | 17 | 7 | 3 | 9 | 11 | 6 | 14 |
| 8 | 2 | 16 | 10 | 18 | 4 | 9 | 6 | 12 | 5 | 15 | 17 | 20 | 3 | 1 | 14 | 11 | 13 | 7 | 19 |
| 8 | 13 | 6 | 12 | 9 | 14 | 20 | 7 | 4 | 19 | 10 | 5 | 11 | 17 | 1 | 18 | 3 | 15 | 2 | 16 |
| 3 | 8 | 11 | 2 | 18 | 15 | 9 | 1 | 10 | 13 | 20 | 5 | 14 | 4 | 17 | 19 | 6 | 7 | 12 | 16 |
| 16 | 17 | 12 | 2 | 15 | 4 | 19 | 10 | 7 | 1 | 3 | 9 | 11 | 5 | 13 | 6 | 20 | 8 | 14 | 18 |
| 19 | 3 | 18 | 10 | 15 | 8 | 12 | 13 | 11 | 17 | 6 | 14 | 9 | 20 | 1 | 5 | 2 | 16 | 7 | 4 |
| 19 | 9 | 3 | 5 | 8 | 6 | 11 | 16 | 4 | 15 | 2 | 18 | 10 | 20 | 14 | 7 | 12 | 17 | 13 | 1 |
| 6 | 7 | 8 | 14 | 16 | 12 | 11 | 2 | 17 | 3 | 5 | 1 | 4 | 10 | 18 | 13 | 15 | 19 | 9 | 20 |
| 9 | 20 | 17 | 5 | 19 | 13 | 6 | 12 | 15 | 14 | 3 | 11 | 16 | 8 | 18 | 7 | 10 | 4 | 2 | 1 |
| 12 | 14 | 19 | 16 | 10 | 1 | 13 | 5 | 8 | 4 | 18 | 17 | 9 | 6 | 11 | 15 | 7 | 3 | 2 | 20 |
| 18 | 1 | 2 | 8 | 4 | 17 | 7 | 15 | 19 | 5 | 12 | 6 | 20 | 16 | 13 | 11 | 14 | 10 | 3 | 9 |
| 8 | 6 | 5 | 11 | 13 | 20 | 7 | 2 | 4 | 17 | 3 | 10 | 1 | 15 | 14 | 9 | 12 | 16 | 19 | 18 |
| 8 | 11 | 16 | 20 | 10 | 19 | 6 | 1 | 15 | 2 | 14 | 7 | 13 | 4 | 5 | 18 | 3 | 9 | 17 | 12 |
| 2 | 7 | 9 | 6 | 15 | 10 | 3 | 18 | 17 | 1 | 5 | 4 | 19 | 11 | 8 | 14 | 16 | 12 | 13 | 20 |
| 5 | 8 | 1 | 15 | 18 | 7 | 3 | 4 | 12 | 16 | 10 | 19 | 14 | 20 | 6 | 17 | 9 | 2 | 13 | 11 |
| 15 | 5 | 20 | 18 | 9 | 6 | 17 | 8 | 10 | 11 | 3 | 2 | 12 | 4 | 7 | 13 | 14 | 19 | 16 | 1 |
| 14 | 3 | 4 | 11 | 1 | 20 | 17 | 15 | 10 | 13 | 2 | 6 | 7 | 9 | 19 | 8 | 5 | 12 | 18 | 16 |
| 8 | 10 | 14 | 5 | 6 | 11 | 7 | 15 | 20 | 3 | 12 | 19 | 13 | 1 | 16 | 4 | 17 | 18 | 9 | 2 |
| 12 | 2 | 19 | 10 | 20 | 17 | 15 | 16 | 3 | 11 | 5 | 4 | 8 | 18 | 9 | 6 | 1 | 14 | 13 | 7 |
| 18 | 10 | 14 | 3 | 13 | 19 | 20 | 15 | 1 | 12 | 16 | 7 | 5 | 17 | 8 | 6 | 11 | 9 | 4 | 2 |
| 2 | 6 | 11 | 8 | 3 | 5 | 12 | 10 | 20 | 13 | 19 | 15 | 17 | 9 | 7 | 16 | 14 | 4 | 1 | 18 |
| 18 | 11 | 10 | 13 | 8 | 20 | 7 | 6 | 9 | 19 | 14 | 5 | 12 | 1 | 17 | 3 | 16 | 15 | 2 | 4 |
| 20 | 5 | 19 | 7 | 10 | 17 | 9 | 3 | 14 | 1 | 16 | 8 | 15 | 18 | 11 | 4 | 12 | 2 | 6 | 13 |
| 7 | 17 | 10 | 15 | 4 | 2 | 13 | 1 | 5 | 14 | 9 | 16 | 18 | 3 | 12 | 11 | 8 | 20 | 19 | 6 |
| 19 | 8 | 11 | 14 | 9 | 17 | 5 | 2 | 13 | 10 | 16 | 6 | 1 | 4 | 12 | 20 | 18 | 7 | 15 | 3 |
| 2 | 4 | 15 | 3 | 13 | 7 | 16 | 20 | 6 | 10 | 8 | 19 | 17 | 5 | 14 | 11 | 12 | 1 | 9 | 18 |
| 18 | 5 | 9 | 13 | 7 | 1 | 12 | 6 | 14 | 15 | 2 | 16 | 19 | 17 | 4 | 3 | 10 | 8 | 20 | 11 |
| 11 | 12 | 14 | 10 | 20 | 9 | 8 | 4 | 15 | 17 | 19 | 1 | 3 | 18 | 7 | 5 | 16 | 6 | 13 | 2 |
| 9 | 18 | 17 | 8 | 4 | 5 | 6 | 2 | 14 | 13 | 7 | 15 | 10 | 3 | 12 | 19 | 16 | 20 | 11 | 1 |
| 1 | 9 | 2 | 18 | 6 | 8 | 12 | 19 | 20 | 15 | 4 | 3 | 13 | 16 | 5 | 14 | 11 | 17 | 7 | 10 |
| 13 | 11 | 14 | 19 | 18 | 15 | 20 | 5 | 10 | 17 | 7 | 8 | 9 | 2 | 16 | 1 | 6 | 3 | 12 | 4 |
| 3 | 14 | 11 | 6 | 18 | 10 | 2 | 15 | 7 | 9 | 19 | 4 | 17 | 20 | 12 | 1 | 5 | 16 | 8 | 13 |
| 18 | 1 | 20 | 7 | 3 | 11 | 13 | 15 | 17 | 10 | 16 | 4 | 6 | 5 | 2 | 12 | 8 | 9 | 19 | 14 |
| 10 | 20 | 7 | 18 | 9 | 5 | 6 | 15 | 2 | 17 | 16 | 19 | 13 | 14 | 12 | 1 | 8 | 11 | 3 | 4 |
| 4 | 14 | 20 | 19 | 6 | 13 | 17 | 8 | 10 | 7 | 5 | 12 | 11 | 2 | 18 | 3 | 9 | 15 | 16 | 1 |
| 6 | 5 | 2 | 8 | 20 | 9 | 15 | 1 | 7 | 10 | 17 | 14 | 4 | 19 | 16 | 13 | 12 | 11 | 3 | 18 |
| 13 | 9 | 2 | 12 | 14 | 18 | 6 | 11 | 20 | 4 | 8 | 19 | 15 | 5 | 7 | 3 | 17 | 1 | 16 | 10 |
| 12 | 17 | 6 | 3 | 18 | 1 | 15 | 9 | 11 | 14 | 20 | 5 | 7 | 19 | 2 | 13 | 8 | 10 | 4 | 16 |
| 7 | 12 | 3 | 10 | 16 | 20 | 13 | 6 | 18 | 2 | 17 | 14 | 4 | 5 | 1 | 11 | 9 | 8 | 19 | 15 |
| 1 | 4 | 18 | 17 | 16 | 10 | 5 | 14 | 20 | 15 | 9 | 3 | 13 | 2 | 7 | 6 | 8 | 11 | 12 | 19 |
| 18 | 9 | 4 | 2 | 16 | 11 | 5 | 19 | 14 | 1 | 12 | 15 | 8 | 17 | 13 | 7 | 20 | 10 | 3 | 6 |
| 3 | 1 | 16 | 18 | 14 | 11 | 4 | 15 | 5 | 13 | 19 | 17 | 6 | 12 | 10 | 7 | 8 | 2 | 20 | 9 |
| 10 | 19 | 7 | 5 | 20 | 3 | 2 | 13 | 14 | 9 | 8 | 12 | 6 | 15 | 4 | 17 | 16 | 11 | 18 | 1 |
| 20 | 15 | 19 | 6 | 7 | 8 | 9 | 2 | 3 | 12 | 18 | 11 | 14 | 16 | 13 | 10 | 5 | 1 | 4 | 17 |

Таблица 4 (продолжение)

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | 7 | 3 | 17 | 5 | 9 | 11 | 4 | 1 | 13 | 18 | 2 | 8 | 20 | 10 | 16 | 12 | 14 | 15 | 19 |
| 10 | 3 | 16 | 4 | 6 | 2 | 11 | 9 | 19 | 17 | 8 | 1 | 14 | 7 | 5 | 20 | 13 | 12 | 15 | 18 |
| 13 | 5 | 6 | 16 | 10 | 7 | 8 | 12 | 2 | 20 | 3 | 4 | 17 | 11 | 9 | 1 | 15 | 19 | 18 | 14 |
| 10 | 5 | 15 | 3 | 2 | 19 | 13 | 18 | 9 | 14 | 16 | 4 | 20 | 17 | 12 | 11 | 8 | 1 | 6 | 7 |
| 10 | 18 | 12 | 19 | 11 | 14 | 4 | 3 | 8 | 2 | 7 | 20 | 5 | 17 | 16 | 6 | 15 | 13 | 1 | 9 |

Таблица 5А. **Коэффициент конкордации $W$.**

Вероятность того, что данное значение $S$ будет достигнуто или превзойдено, для $n = 3$ и $m$ от 2 до 10

| $S$ | $m=2$ | $m=3$ | $m=4$ | $m=5$ | $m=6$ | $m=7$ | $m=8$ | $m=9$ | $m=10$ |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |
| 2 | 0,833 | 0,944 | 0,931 | 0,954 | 0,956 | 0,964 | 0,967 | 0,971 | 0,974 |
| 6 | 0,500 | 0,528 | 0,653 | 0,691 | 0,740 | 0,768 | 0,794 | 0,814 | 0,830 |
| 8 | 0,167 | 0,361 | 0,431 | 0,522 | 0,570 | 0,620 | 0,654 | 0,685 | 0,710 |
| 14 | | 0,194 | 0,273 | 0,367 | 0,430 | 0,486 | 0,531 | 0,569 | 0,601 |
| 18 | | 0,028 | 0,125 | 0,182 | 0,252 | 0,305 | 0,355 | 0,398 | 0,436 |
| 24 | | | 0,069 | 0,124 | 0,184 | 0,237 | 0,285 | 0,328 | 0,368 |
| 26 | | | 0,042 | 0,093 | 0,142 | 0,192 | 0,236 | 0,278 | 0,316 |
| 32 | | | 0,0046 | 0,039 | 0,072 | 0,112 | 0,149 | 0,187 | 0,222 |
| 38 | | | | 0,024 | 0,052 | 0,085 | 0,120 | 0,154 | 0,187 |
| 42 | | | | 0,0085 | 0,029 | 0,051 | 0,079 | 0,107 | 0,135 |
| 50 | | | | $0,0^{3}77$ | 0,012 | 0,027 | 0,047 | 0,069 | 0,092 |
| 54 | | | | | 0,0081 | 0,021 | 0,038 | 0,057 | 0,078 |
| 56 | | | | | 0,0055 | 0,016 | 0,030 | 0,048 | 0,066 |
| 62 | | | | | 0,0017 | 0,0084 | 0,018 | 0,031 | 0,046 |
| 72 | | | | | $0,0^{3}13$ | 0,0036 | 0,0099 | 0,019 | 0,030 |
| 74 | | | | | | 0,0027 | 0,0080 | 0,016 | 0,026 |
| 78 | | | | | | 0,0012 | 0,0048 | 0,010 | 0,018 |
| 86 | | | | | | $0,0^{3}32$ | 0,0024 | 0,0060 | 0,012 |
| 96 | | | | | | $0,0^{3}32$ | 0,0011 | 0,0035 | 0,0075 |
| 98 | | | | | | $0,0^{4}21$ | $0,0^{3}86$ | 0,0029 | 0,0063 |
| 104 | | | | | | | $0,0^{3}26$ | 0,0013 | 0,0034 |
| 114 | | | | | | | $0,0^{4}61$ | $0,0^{3}66$ | 0,0020 |
| 122 | | | | | | | $0,0^{4}61$ | $0,0^{3}35$ | 0,0013 |
| 126 | | | | | | | $0,0^{4}61$ | $0,0^{3}20$ | $0,0^{3}83$ |
| 128 | | | | | | | $0,0^{5}36$ | $0,0^{4}97$ | $0,0^{3}51$ |
| 134 | | | | | | | | $0,0^{4}54$ | $0,0^{3}37$ |
| 146 | | | | | | | | $0,0^{4}11$ | $0,0^{3}18$ |
| 150 | | | | | | | | $0,0^{4}11$ | $0,0^{3}11$ |
| 152 | | | | | | | | $0,0^{4}11$ | $0,0^{4}85$ |
| 158 | | | | | | | | $0,0^{4}11$ | $0,0^{4}44$ |
| 162 | | | | | | | | $0,0^{6}60$ | $0,0^{4}20$ |
| 168 | | | | | | | | | $0,0^{4}11$ |
| 182 | | | | | | | | | $0,0^{5}21$ |
| 200 | | | | | | | | | $0,0^{7}99$ |

Таблица 5В. **Коэффициент конкордации $W$.**

Вероятность того, что данное значение $S$ будет достигнуто или превзойдено, для $n = 4$, $m = 3$ и $m = 5$

| $S$ | $m = 3$ | $m = 5$ | $S$ | $m = 5$ |
|---|---|---|---|---|
| 1 | 1,000 | 1,000 | 61 | 0,055 |
| 3 | 0,958 | 0,975 | 65 | 0,044 |
| 5 | 0,910 | 0,944 | 67 | 0,034 |
| 9 | 0,727 | 0,857 | 69 | 0,031 |
| 11 | 0,608 | 0,771 | 73 | 0,023 |
| 13 | 0,524 | 0,709 | 75 | 0,020 |
| 17 | 0,446 | 0,652 | 77 | 0,017 |
| 19 | 0,342 | 0,561 | 81 | 0,012 |
| 21 | 0,300 | 0,521 | 83 | 0,0087 |
| 25 | 0,207 | 0,445 | 85 | 0,0067 |
| 27 | 0,175 | 0,408 | 89 | 0,0055 |
| 29 | 0,148 | 0,372 | 91 | 0,0031 |
| 33 | 0,075 | 0,298 | 93 | 0,0023 |
| 35 | 0,054 | 0,260 | 97 | 0,0018 |
| 37 | 0,033 | 0,226 | 99 | 0,0016 |
| 41 | 0,017 | 0,210 | 101 | 0,0014 |
| 43 | 0,0017 | 0,162 | 105 | $0{,}0^{3}64$ |
| 45 | 0,0017 | 0,141 | 107 | $0{,}0^{3}33$ |
| 49 | | 0,123 | 109 | $0{,}0^{3}21$ |
| 51 | | 0,107 | 113 | $0{,}0^{3}14$ |
| 53 | | 0,093 | 117 | $0{,}0^{4}48$ |
| 57 | | 0,075 | 125 | $0{,}0^{5}30$ |
| 59 | | 0,067 | | |

Таблица 5С. **Коэффициент конкордации $W$.**

Вероятность того, что данное значение $S$ будет достигнуто или превзойдено, для $n = 4$ и $m = 2$, $m = 4$ и $m = 6$

| $S$ | $m = 2$ | $m = 4$ | $m = 6$ | $S$ | $m = 6$ |
|---|---|---|---|---|---|
| 0 | 1,000 | 1,000 | 1,000 | 82 | 0,035 |
| 2 | 0,958 | 0,992 | 0,996 | 84 | 0,032 |
| 4 | 0,833 | 0,928 | 0,957 | 86 | 0,029 |
| 6 | 0,792 | 0,900 | 0,940 | 88 | 0,023 |
| 8 | 0,625 | 0,800 | 0,874 | 90 | 0,022 |
| 10 | 0,542 | 0,754 | 0,844 | 94 | 0,017 |
| 12 | 0,458 | 0,677 | 0,789 | 96 | 0,014 |
| 14 | 0,375 | 0,649 | 0,772 | 98 | 0,013 |
| 16 | 0,208 | 0,524 | 0,679 | 100 | 0,010 |
| 18 | 0,167 | 0,508 | 0,668 | 102 | 0,0096 |
| 20 | 0,042 | 0,432 | 0,609 | 104 | 0,0085 |
| 22 | | 0,389 | 0,574 | 106 | 0,0073 |
| 24 | | 0,355 | 0,541 | 108 | 0,0061 |
| 26 | | 0,324 | 0,512 | 110 | 0,0057 |
| 30 | | 0,242 | 0,431 | 114 | 0,0040 |
| 32 | | 0,200 | 0,386 | 116 | 0,0033 |
| 34 | | 0,190 | 0,375 | 118 | 0,0028 |
| 36 | | 0,158 | 0,338 | 120 | 0,0023 |

Продолжение табл. 5С

| S | m = 2 | m = 4 | m = 6 | S | m = 6 |
|---|---|---|---|---|---|
| 38 | | 0,141 | 0,317 | 122 | 0,0020 |
| 40 | | 0,105 | 0,270 | 126 | 0,0015 |
| 42 | | 0,094 | 0,256 | 128 | $0,0^{3}90$ |
| 44 | | 0,077 | 0,230 | 130 | $0,0^{3}87$ |
| 46 | | 0,068 | 0,218 | 132 | $0,0^{3}73$ |
| 48 | | 0,054 | 0,197 | 134 | $0,0^{3}65$ |
| 50 | | 0,052 | 0,194 | 136 | $0,0^{3}40$ |
| 52 | | 0,036 | 0,163 | 138 | $0,0^{3}36$ |
| 54 | | 0,033 | 0,155 | 140 | $0,0^{3}28$ |
| 56 | | 0,019 | 0,127 | 144 | $0,0^{3}24$ |
| 58 | | 0,014 | 0,114 | 146 | $0,0^{3}22$ |
| 62 | | 0,012 | 0,108 | 148 | $0,0^{3}12$ |
| 64 | | 0,0069 | 0,089 | 150 | $0,0^{4}95$ |
| 66 | | 0,0062 | 0,088 | 152 | $0,0^{4}62$ |
| 68 | | 0,0027 | 0,073 | 154 | $0,0^{4}46$ |
| 70 | | 0,0027 | 0,066 | 158 | $0,0^{4}24$ |
| 72 | | 0,0016 | 0,060 | 160 | $0,0^{4}16$ |
| 74 | | $0,0^{3}94$ | 0,056 | 162 | $0,0^{4}12$ |
| 76 | | $0,0^{3}94$ | 0,043 | 164 | $0,0^{5}80$ |
| 78 | | $0,0^{3}94$ | 0,041 | 170 | $0,0^{5}24$ |
| 80 | | $0,0^{4}72$ | 0,037 | 180 | $0,0^{6}13$ |

Таблица 5D. **Коэффициент конкордации** $W$.

Вероятность того, что данное значение $S$ будет достигнуто или превзойдено, для $n = 5$ и $m = 3$

| S | m = 3 | S | m = 3 | S | m = 3 | S | m = 3 |
|---|---|---|---|---|---|---|---|
| 0 | 1,000 | 22 | 0,649 | 44 | 0,236 | 66 | 0,038 |
| 2 | 1,000 | 24 | 0,595 | 46 | 0,213 | 68 | 0,028 |
| 4 | 0,988 | 26 | 0,559 | 48 | 0,172 | 70 | 0,026 |
| 6 | 0,972 | 28 | 0,493 | 50 | 0,163 | 72 | 0,017 |
| 8 | 0,941 | 30 | 0,475 | 52 | 0,127 | 74 | 0,015 |
| 10 | 0,914 | 32 | 0,432 | 54 | 0,117 | 76 | 0,0078 |
| 12 | 0,845 | 34 | 0,406 | 56 | 0,096 | 78 | 0,0053 |
| 14 | 0,831 | 36 | 0,347 | 58 | 0,080 | 80 | 0,0040 |
| 16 | 0,768 | 38 | 0,326 | 60 | 0,063 | 82 | 0,0028 |
| 18 | 0,720 | 40 | 0,291 | 62 | 0,056 | 86 | $0,0^{3}90$ |
| 20 | 0,682 | 42 | 0,253 | 64 | 0,045 | 90 | $0,0^{4}69$ |

Таблица 6. **Критические значения** $S$ **(для коэффициента конкордации** $W$**).** (Заимствовано из [30] с разрешения автора и издателя).

| m | n | | | | | Дополнительные значения для $n = 3$ | |
|---|---|---|---|---|---|---|---|
| | 3 | 4 | 5 | 6 | 7 | m | S |
| | Значения при 0,05 уровне существенности | | | | | | |
| 3 | | | 64,4 | 103,9 | 157,3 | 9 | 54,0 |
| 4 | | 49,5 | 88,4 | 143,3 | 217,0 | 12 | 71,9 |
| 5 | | 62,6 | 112,3 | 182,4 | 276,2 | 14 | 83,8 |
| 6 | | 75,7 | 136,1 | 221,4 | 335,2 | 16 | 95,8 |
| 8 | 48,1 | 101,7 | 183,7 | 299,0 | 453,1 | 18 | 107,7 |
| 10 | 60,0 | 127,8 | 231,2 | 376,7 | 571,0 | | |
| 15 | 89,8 | 192,9 | 349,8 | 570,5 | 864,9 | | |
| 20 | 119,7 | 258,0 | 468,5 | 764,4 | 1158,7 | | |

Продолжение

| $m$ | $n$ = 3 | 4 | 5 | 6 | 7 | Дополнительные значения для $n = 3$: $m$ | $S$ |
|---|---|---|---|---|---|---|---|
| | | | Значения при 0,01 уровне существенности | | | | |
| 3 | | | 75,6 | 122,8 | 185,6 | 9 | 75,9 |
| 4 | | 61,4 | 109,3 | 176,2 | 265,0 | 12 | 103,5 |
| 5 | | 80,5 | 142,8 | 229,4 | 343,8 | 14 | 121,9 |
| 6 | | 99,5 | 176,1 | 282,4 | 422,6 | 16 | 140,2 |
| 8 | 66,8 | 137,4 | 242,7 | 388,3 | 579,9 | 18 | 158,6 |
| 10 | 85,1 | 175,3 | 309,1 | 494,0 | 737,0 | | |
| 15 | 131,0 | 269,8 | 475,2 | 758,2 | 1129,5 | | |
| 20 | 177,0 | 364,2 | 641,2 | 1022,2 | 1521,9 | | |

Таблица 7А. **5%-ные точки распределения z.**

(Из табл. VI книги R. A. Fisher. Statistical Methods for Research Workers, Oliver and Boyd, Edinburgh. Фишер Р. Статистические методы для исследователей, Госстатиздат, М., 1958.)

| Значения $\nu_2$ | Значения $\nu_1$: 1. | 2. | 3. | 4. | 5. | 6. | 8. | 12. | 24. | $\infty$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2,5421 | 2,6479 | 2,6870 | 2,7071 | 2,7194 | 2,7276 | 2,7380 | 2,7484 | 2,7588 | 2,7693 |
| 2 | 1,4592 | 1,4722 | 1,4765 | 1,4787 | 1,4800 | 1,4808 | 1,4819 | 1,4830 | 1,4840 | 1,4851 |
| 3 | 1,1577 | 1,1284 | 1,1137 | 1,1051 | 1,0994 | 1,0953 | 1,0899 | 1,0842 | 1,0781 | 1,0716 |
| 4 | 1,0212 | 0,9690 | 0,9429 | 0,9272 | 0,9168 | 0,9093 | 0,8993 | 0,8885 | 0,8767 | 0,8639 |
| 5 | 0,9441 | 0,8777 | 0,8441 | 0,8236 | 0,8097 | 0,7997 | 0,7862 | 0,7714 | 0,7550 | 0,7368 |
| 6 | 0,8948 | 0,8188 | 0,7798 | 0,7558 | 0,7394 | 0,7274 | 0,7112 | 0,6931 | 0,6729 | 0,6499 |
| 7 | 0,8606 | 0,7777 | 0,7347 | 0,7080 | 0,6896 | 0,6761 | 0,6576 | 0,6369 | 0,6134 | 0,5862 |
| 8 | 0,8355 | 0,7475 | 0,7014 | 0,6725 | 0,6525 | 0,6378 | 0,6175 | 0,5945 | 0,5682 | 0,5371 |
| 9 | 0,8163 | 0,7242 | 0,6757 | 0,6450 | 0,6238 | 0,6080 | 0,5862 | 0,5613 | 0,5324 | 0,4979 |
| 10 | 0,8012 | 0,7058 | 0,6553 | 0,6232 | 0,6009 | 0,5843 | 0,5611 | 0,5346 | 0,5035 | 0,4657 |
| 11 | 0,7889 | 0,6909 | 0,6387 | 0,6055 | 0,5822 | 0,5648 | 0,5406 | 0,5126 | 0,4795 | 0,4387 |
| 12 | 0,7788 | 0,6786 | 0,6250 | 0,5907 | 0,5666 | 0,5487 | 0,5234 | 0,4941 | 0,4592 | 0,4156 |
| 13 | 0,7703 | 0,6682 | 0,6134 | 0,5783 | 0,5535 | 0,5350 | 0,5089 | 0,4785 | 0,4419 | 0,3957 |
| 14 | 0,7630 | 0,6594 | 0,6036 | 0,5677 | 0,5423 | 0,5233 | 0,4964 | 0,4649 | 0,4269 | 0,3782 |
| 15 | 0,7568 | 0,6518 | 0,5950 | 0,5585 | 0,5326 | 0,5131 | 0,4855 | 0,4532 | 0,4138 | 0,3628 |
| 16 | 0,7514 | 0,6451 | 0,5876 | 0,5505 | 0,5241 | 0,5042 | 0,4760 | 0,4428 | 0,4022 | 0,3490 |
| 17 | 0,7466 | 0,6393 | 0,5811 | 0,5434 | 0,5166 | 0,4964 | 0,4676 | 0,4337 | 0,3919 | 0,3366 |
| 18 | 0,7424 | 0,6341 | 0,5753 | 0,5371 | 0,5099 | 0,4894 | 0,4602 | 0,4255 | 0,3827 | 0,3253 |
| 19 | 0,7386 | 0,6295 | 0,5701 | 0,5315 | 0,5040 | 0,4832 | 0,4535 | 0,4182 | 0,3743 | 0,3151 |
| 20 | 0,7352 | 0,6254 | 0,5654 | 0,5265 | 0,4986 | 0,4776 | 0,4474 | 0,4116 | 0,3668 | 0,3057 |
| 21 | 0,7322 | 0,6216 | 0,5612 | 0,5219 | 0,4938 | 0,4725 | 0,4420 | 0,4055 | 0,3599 | 0,2971 |
| 22 | 0,7294 | 0,6182 | 0,5574 | 0,5178 | 0,4894 | 0,4679 | 0,4370 | 0,4001 | 0,3536 | 0,2892 |
| 23 | 0,7269 | 0,6151 | 0,5540 | 0,5140 | 0,4854 | 0,4636 | 0,4325 | 0,3950 | 0,3478 | 0,2818 |
| 24 | 0,7246 | 0,6123 | 0,5508 | 0,5106 | 0,4817 | 0,4598 | 0,4283 | 0,3904 | 0,3425 | 0,2749 |
| 25 | 0,7225 | 0,6097 | 0,5478 | 0,5074 | 0,4783 | 0,4562 | 0,4244 | 0,3862 | 0,3376 | 0,2685 |
| 26 | 0,7205 | 0,6073 | 0,5451 | 0,5045 | 0,4752 | 0,4529 | 0,4209 | 0,3823 | 0,3330 | 0,2625 |
| 27 | 0,7187 | 0,6051 | 0,5427 | 0,5017 | 0,4723 | 0,4499 | 0,4176 | 0,3786 | 0,3287 | 0,2569 |
| 28 | 0,7171 | 0,6030 | 0,5403 | 0,4992 | 0,4696 | 0,4471 | 0,4146 | 0,3752 | 0,3248 | 0,2516 |
| 29 | 0,7155 | 0,6011 | 0,5382 | 0,4969 | 0,4671 | 0,4444 | 0,4117 | 0,3720 | 0,3211 | 0,2466 |
| 30 | 0,7141 | 0,5994 | 0,5262 | 0,4947 | 0,4648 | 0,4420 | 0,4090 | 0,3691 | 0,3176 | 0,2419 |
| 60 | 0,6933 | 0,5738 | 0,5073 | 0,4632 | 0,4311 | 0,4064 | 0,3702 | 0,3255 | 0,2654 | 0,1644 |
| $\infty$ | 0,6729 | 0,5486 | 0,4787 | 0,4319 | 0,3974 | 0,3706 | 0,3309 | 0,2804 | 0,2085 | 0 |

Таблица 7В. **1%-ные точки распределения z.**

(Заимствованы из табл. VI книги R. A. Fisher. Statistical Methods for Research Workers, Oliver and Boyd, Edinburgh, русск. пер. см. на стр. 204).

| Значения $\nu_2$ | Значения $\nu_1$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1. | 2. | 3. | 4. | 5. | 6. | 8. | 12. | 24. | ∞ |
| 1 | 4,1535 | 4,2585 | 4,2974 | 4,3175 | 4,3297 | 4,3379 | 4,3482 | 4,3585 | 4,3689 | 4,3794 |
| 2 | 2,2950 | 2,2976 | 2,2984 | 2,2988 | 2,2991 | 2,2992 | 2,2994 | 2,2997 | 2,2999 | 2,3001 |
| 3 | 1,7649 | 1,7140 | 1,6915 | 1,6786 | 1,6703 | 1,6645 | 1,6569 | 1,6489 | 1,6404 | 1,6314 |
| 4 | 1,5270 | 1,4452 | 1,4075 | 1,3856 | 1,3711 | 1,3609 | 1,3473 | 1,3327 | 1,3170 | 1,3000 |
| 5 | 1,3943 | 1,2929 | 1,2449 | 1,2164 | 1,1974 | 1,1838 | 1,1656 | 1,1457 | 1,1239 | 1,0997 |
| 6 | 1,3103 | 1,1955 | 1,1401 | 1,1068 | 1,0843 | 1,0680 | 1,0460 | 1,0218 | 0,9948 | 0,9643 |
| 7 | 1,2526 | 1,1281 | 1,0672 | 1,0300 | 1,0048 | 0,9864 | 0,9614 | 0,9335 | 0,9020 | 0,8658 |
| 8 | 1,2106 | 1,0787 | 1,0135 | 0,9734 | 0,9459 | 0,9259 | 0,8983 | 0,8673 | 0,8319 | 0,7904 |
| 9 | 1,1786 | 1,0411 | 0,9724 | 0,9299 | 0,9006 | 0,8791 | 0,8494 | 0,8157 | 0,7769 | 0,7305 |
| 10 | 1,1535 | 1,0114 | 0,9399 | 0,8954 | 0,8646 | 0,8419 | 0,8104 | 0,7744 | 0,7324 | 0,6816 |
| 11 | 1,1333 | 0,9874 | 0,9136 | 0,8674 | 0,8354 | 0,8116 | 0,7785 | 0,7405 | 0,6958 | 0,6408 |
| 12 | 1,1166 | 0,9677 | 0,8919 | 0,8443 | 0,8111 | 0,7864 | 0,7520 | 0,7122 | 0,6649 | 0,6061 |
| 13 | 1,1027 | 0,9511 | 0,8737 | 0,8248 | 0,7907 | 0,7652 | 0,7295 | 0,6882 | 0,0386 | 0,5761 |
| 14 | 1,0909 | 0,9370 | 0,8581 | 0,8082 | 0,7732 | 0,7471 | 0,7103 | 0,6675 | 0,6159 | 0,5500 |
| 15 | 1,0807 | 0,9249 | 0,8448 | 0,7939 | 0,7582 | 0,7314 | 0,6937 | 0,6496 | 0,5961 | 0,5269 |
| 16 | 1,0719 | 0,9144 | 0,8331 | 0,7814 | 0,7450 | 0,7177 | 0,6791 | 0,6339 | 0,5786 | 0,5064 |
| 17 | 1,0641 | 0,9051 | 0,8229 | 0,7705 | 0,7335 | 0,7057 | 0,6663 | 0,6199 | 0,5630 | 0,4879 |
| 18 | 1,0572 | 0,8970 | 0,8138 | 0,7607 | 0,7232 | 0,6950 | 0,6549 | 0,6075 | 0,5491 | 0,4712 |
| 19 | 1,0511 | 0,8897 | 0,8057 | 0,7521 | 0,7140 | 0,6854 | 0,6447 | 0,5964 | 0,5366 | 0,4560 |
| 20 | 1,0457 | 0,8831 | 0,7985 | 0,7443 | 0,7058 | 0,6768 | 0,6355 | 0,5864 | 0,5253 | 0,4421 |
| 21 | 1,0408 | 0,8772 | 0,7920 | 0,7372 | 0,6984 | 0,6690 | 0,6272 | 0,5773 | 0,5150 | 0,4294 |
| 22 | 1,0363 | 0,8719 | 0,7860 | 0,7309 | 0,6916 | 0,6620 | 0,6196 | 0,5691 | 0,5056 | 0,4176 |
| 23 | 1,0322 | 0,8670 | 0,7806 | 0,7251 | 0,6855 | 0,6555 | 0,6127 | 0,5615 | 0,4969 | 0,4068 |
| 24 | 1,0285 | 0,8626 | 0,7757 | 0,7197 | 0,6799 | 0,6496 | 0,6064 | 0,5545 | 0,4890 | 0,3967 |
| 25 | 1,0251 | 0,8585 | 0,7712 | 0,7148 | 0,6747 | 0,6442 | 0,6006 | 0,5481 | 0,4816 | 0,3872 |
| 26 | 1,0220 | 0,8548 | 0,7670 | 0,7103 | 0,6699 | 0,6392 | 0,5952 | 0,5422 | 0,4748 | 0,3784 |
| 27 | 1,0191 | 0,8513 | 0,7631 | 0,7062 | 0,6655 | 0,6346 | 0,5902 | 0,5367 | 0,4685 | 0,3701 |
| 28 | 1,0164 | 0,8481 | 0,7595 | 0,7023 | 0,6614 | 0,6303 | 0,5856 | 0,5316 | 0,4626 | 0,3624 |
| 29 | 1,0139 | 0,8451 | 0,7562 | 0,6987 | 0,6576 | 0,6263 | 0,5813 | 0,5269 | 0,4570 | 0,3550 |
| 30 | 1,0116 | 0,8423 | 0,7531 | 0,6954 | 0,6540 | 0,6226 | 0,5773 | 0,5224 | 0,4519 | 0,3481 |
| 60 | 0,9784 | 0,8025 | 0,7086 | 0,6472 | 0,6028 | 0,5687 | 0,5189 | 0,4574 | 0,3746 | 0,2352 |
| ∞ | 0,9462 | 0,7636 | 0,6651 | 0,5999 | 0,5522 | 0,5152 | 0,4604 | 0,3908 | 0,2913 | 0 |

Таблица 8. **Табличные значения $\chi^2$.**

(Воспроизведено по табл. III книги R. A. Fisher Statistical Methods for
и издателей.)

| $\nu$ | $P=0,99$ | 0,98 | 0,95 | 0,90 | 0,80 | 0,70 |
|---|---|---|---|---|---|---|
| 1 | $0,0^3157$ | $0,0^3628$ | $0,0^2393$ | 0,0158 | 0,0642 | 0,148 |
| 2 | 0,0201 | 0,0404 | 0,103 | 0,211 | 0,446 | 0,713 |
| 3 | 0,115 | 0,185 | 0,352 | 0,584 | 1,005 | 1,424 |
| 4 | 0,297 | 0,429 | 0,711 | 1,064 | 1,649 | 2,195 |
| 5 | 0,554 | 0,752 | 1,145 | 1,160 | 2,343 | 3,000 |
| 6 | 0,872 | 1,134 | 1,635 | 2,204 | 3,070 | 3,828 |
| 7 | 1,239 | 1,564 | 2,167 | 2,833 | 3,822 | 4,671 |
| 8 | 1,646 | 2,032 | 2,733 | 3,490 | 4,594 | 5,527 |
| 9 | 2,088 | 2,532 | 3,325 | 4,168 | 5,380 | 6,393 |
| 10 | 2,358 | 3,059 | 3,940 | 4,865 | 6,179 | 7,267 |
| 11 | 3,053 | 3,609 | 4,575 | 5,578 | 6,989 | 8,148 |
| 12 | 3,571 | 4,178 | 5,226 | 6,304 | 7,807 | 9,034 |
| 13 | 4,107 | 4,765 | 5,892 | 7,042 | 8,634 | 9,926 |
| 14 | 4,660 | 5,368 | 6,571 | 7,790 | 9,467 | 10,821 |
| 15 | 5,229 | 5,985 | 7,261 | 8,547 | 10,307 | 11,721 |
| 16 | 5,812 | 6,614 | 7,962 | 9,312 | 11,152 | 12,624 |
| 17 | 6,408 | 7,255 | 8,672 | 10,085 | 12,002 | 13,531 |
| 18 | 7,015 | 7,906 | 9,390 | 10,865 | 12,857 | 14,440 |
| 19 | 7,633 | 8,567 | 10,117 | 11,651 | 13,716 | 15,352 |
| 20 | 8,260 | 9,237 | 10,851 | 12,443 | 14,578 | 16,266 |
| 21 | 8,897 | 9,915 | 11,591 | 13,240 | 15,445 | 17,182 |
| 22 | 9,542 | 10,600 | 12,338 | 14,041 | 16,314 | 18,101 |
| 23 | 10,196 | 11,293 | 13,091 | 14,848 | 17,187 | 19,021 |
| 24 | 10,856 | 11,992 | 13,848 | 15,659 | 18,062 | 19,943 |
| 25 | 11,524 | 12,697 | 14,611 | 16,473 | 18,940 | 20,867 |
| 26 | 12,198 | 13,409 | 15,379 | 17,292 | 19,820 | 21,792 |
| 27 | 12,879 | 14,125 | 16,151 | 18,114 | 20,703 | 22,719 |
| 28 | 13,565 | 14,847 | 16,928 | 18,939 | 21,588 | 23,647 |
| 29 | 14,256 | 15,574 | 17,708 | 19,768 | 22,475 | 24,577 |
| 30 | 14,953 | 16,306 | 18,493 | 20,599 | 23,364 | 25,508 |

Примечание. **Для значений $\nu$, превышающих 30, величину $\sqrt{2\chi^2}$ можно считать**

Research Workers, Oliver and Boyd, Edinburgh. Русский пер. см. на стр. 204.

| 0,50 | 0,30 | 0,20 | 0,10 | 0,05 | 0,02 | 0,01 |
|---|---|---|---|---|---|---|
| 0,455 | 1,074 | 1,642 | 2,706 | 3,841 | 5,412 | 6,635 |
| 1,386 | 2,408 | 3,219 | 4,605 | 5,991 | 7,824 | 9,210 |
| 2,366 | 3,665 | 4,642 | 6,251 | 7,815 | 8,837 | 11,345 |
| 3,357 | 4,878 | 5,989 | 7,779 | 9,488 | 11,668 | 13,277 |
| 4,351 | 6,064 | 7,289 | 9,236 | 11,070 | 13,388 | 15,086 |
| 5,348 | 7,231 | 8,558 | 10,645 | 12,592 | 15,033 | 16,812 |
| 6,346 | 8,383 | 9,803 | 12,017 | 14,067 | 16,622 | 18,475 |
| 7,344 | 9,524 | 11,030 | 13,362 | 15,507 | 18,168 | 20,090 |
| 8,343 | 10,656 | 12,242 | 14,684 | 16,919 | 19,679 | 21,666 |
| 9,342 | 11,781 | 13,442 | 15,987 | 18,307 | 21,161 | 23,209 |
| 10,341 | 12,899 | 14,631 | 17,275 | 19,675 | 22,618 | 24,725 |
| 11,340 | 14,011 | 15,821 | 18,549 | 21,026 | 24,054 | 26,217 |
| 12,340 | 15,119 | 16,985 | 19,812 | 22,362 | 25,472 | 27,688 |
| 13,339 | 16,222 | 18,151 | 21,064 | 23,685 | 26,873 | 29,141 |
| 14,339 | 17,322 | 19,311 | 22,307 | 24,996 | 28,259 | 30,578 |
| 15,338 | 18,418 | 20,465 | 23,542 | 26,296 | 29,633 | 32,000 |
| 16,338 | 19,511 | 21,615 | 24,769 | 27,587 | 30,995 | 33,409 |
| 17,338 | 20,601 | 22,760 | 25,989 | 28,869 | 32,346 | 34,805 |
| 18,338 | 21,689 | 23,900 | 27,204 | 30,144 | 33,687 | 36,191 |
| 19,337 | 22,775 | 25,038 | 28,412 | 31,410 | 35,020 | 37,566 |
| 20,337 | 23,858 | 26,171 | 29,615 | 32,671 | 36,343 | 38,932 |
| 21,337 | 24,939 | 27,301 | 30,813 | 33,924 | 37,659 | 40,289 |
| 22,337 | 26,018 | 28,429 | 32,007 | 35,172 | 38,968 | 41,638 |
| 23,337 | 27,096 | 29,553 | 33,196 | 36,415 | 40,270 | 42,980 |
| 24,337 | 28,172 | 30,675 | 34,382 | 37,652 | 41,566 | 44,314 |
| 25,336 | 29,246 | 31,795 | 35,563 | 38,885 | 42,856 | 45,642 |
| 26,336 | 30,319 | 32,912 | 36,741 | 40,113 | 44,140 | 46,963 |
| 27,336 | 31,391 | 34,027 | 37,916 | 41,337 | 45,419 | 48,278 |
| 28,336 | 32,461 | 35,139 | 39,087 | 42,557 | 46,693 | 49,588 |
| 29,336 | 33,530 | 36,250 | 40,256 | 43,773 | 47,962 | 50,892 |

распределенной нормально со средней $\sqrt{2\nu-1}$ и дисперсией, равной единице.

Таблица 9. **Частоты ($f$) для значений числа круговых триад ($d$) в парных сравнениях и вероятности того, что эти значения будут достигнуты или превышены**

| $d$ | $n=2$ | | $n=3$ | | $n=4$ | | $n=5$ | | $n=6$ | | $n=7$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $f$ | $P$ | $f$ | $P$ | $f$ | $P$ | $f$ | $P$ | $f$ | $P$ | $f$ | $P$ |
| 0 | 2 | 1,000 | 6 | 1,000 | 24 | 1,000 | 120 | 1,000 | 720 | 1,000 | 5 040 | 1,000 |
| 1 | | | 2 | 0,250 | 16 | 0,625 | 120 | 0,883 | 960 | 0,978 | 8 400 | 0,998 |
| 2 | | | | | 24 | 0,375 | 240 | 0,766 | 2 240 | 0,949 | 21 840 | 0,994 |
| 3 | | | | | | | 240 | 0,531 | 2 880 | 0,880 | 33 600 | 0,983 |
| 4 | | | | | | | 280 | 0,297 | 6 240 | 0,792 | 75 600 | 0,967 |
| 5 | | | | | | | 24 | 0,023 | 3 648 | 0,602 | 90 384 | 0,931 |
| 6 | | | | | | | | | 8 640 | 0,491 | 179 760 | 0,888 |
| 7 | | | | | | | | | 4 800 | 0,227 | 188 160 | 0,802 |
| 8 | | | | | | | | | 2 640 | 0,081 | 277 200 | 0,713 |
| 9 | | | | | | | | | | | 280 560 | 0,580 |
| 10 | | | | | | | | | | | 384 048 | 0,447 |
| 11 | | | | | | | | | | | 244 160 | 0,263 |
| 12 | | | | | | | | | | | 233 520 | 0,147 |
| 13 | | | | | | | | | | | 72 240 | 0,036 |
| 14 | | | | | | | | | | | 2 640 | 0,001 |
| Итого | 2 | — | 8 | — | 64 | — | 1 024 | — | 32 768 | — | 2 097 152 | — |

Таблица 9 (продолжение)

Для случаев, когда число объектов ($n$) равно 10, приближенное значение вероятности ($P'$) определяется на основе приближенного распределения $\chi^2$

| $d$ | $n=8$ |  | $n=9$ |  | $n=10$ |  |  |
|---|---|---|---|---|---|---|---|
|  | $f$ | $P$ | $f$ | $P$ | $f$ | $P$ | $P'$ |
| 0 | 40 320 | 1,000 | 362 880 | 1,000 | 3 628 800 | 1,0$^5$ | 1,000 |
| 1 | 80 640 | 0,9$^3$85 | 846 720 | 0,9$^5$ | 9 676 800 | 1,0$^5$ | 1,000 |
| 2 | 228 480 | 0,9$^3$55 | 2 580 480 | 0,9$^4$8 | 31 449 600 | 1,0$^5$ | 1,000 |
| 3 | 403 200 | 0,9$^2$87 | 5 093 760 | 0,9$^4$4 | 68 275 200 | 1,0$^5$ | 1,000 |
| 4 | 954 240 | 0,9$^2$72 | 12 579 840 | 0,9$^3$87 | 175 392 000 | 1,0$^5$ | 1,000 |
| 5 | 1 304 576 | 0,9$^2$36 | 19 958 400 | 0,9$^3$69 | 311 592 960 | 0,9$^5$ | 1,000 |
| 6 | 3 042 816 | 0,989 | 44 698 752 | 0,9$^3$40 | 711 728 640 | 0,9$^4$ | 1,000 |
| 7 | 3 870 720 | 0,977 | 70 785 792 | 0,9$^2$87 | 1 193 794 560 | 0,9$^4$ | 1,000 |
| 8 | 6 926 080 | 0,963 | 130 032 000 | 0,9$^2$77 | 2 393 475 840 | 0,9$^4$ | 1,000 |
| 9 | 8 332 800 | 0,937 | 190 834 560 | 0,9$^2$58 | 3 784 596 480 | 0,9$^3$ | 1,000 |
| 10 | 15 821 568 | 0,906 | 361 525 248 | 0,9$^2$31 | 7 444 104 192 | 0,9$^3$ | 1,000 |
| 11 | 14 755 328 | 0,847 | 443 931 264 | 0,988 | 10 526 745 600 | 0,9$^3$ | 1,000 |
| 12 | 24 487 680 | 0,792 | 779 950 080 | 0,981 | 19 533 696 000 | 0,9$^3$ | 1,000 |
| 13 | 24 514 560 | 0,701 | 1 043 763 840 | 0,970 | 27 610 168 320 | 0,9$^2$87 | 1,000 |
| 14 | 34 762 240 | 0,610 | 1 529 101 440 | 0,955 | 47 107 169 280 | 0,9$^2$79 | 0,999 |
| 15 | 29 288 448 | 0,480 | 1 916 619 264 | 0,933 | 64 016 040 960 | 0,9$^2$66 | 0,997 |
| 16 | 37 188 480 | 0,371 | 2 912 257 152 | 0,905 | 107 446 832 640 | 0,9$^2$47 | 0,996 |
| 17 | 24 487 680 | 0,232 | 3 078 407 808 | 0,862 | 134 470 425 600 | 0,9$^2$17 | 0,995 |
| 18 | 24 312 960 | 0,141 | 4 506 485 760 | 0,817 | 218 941 470 720 | 0,988 | 0,993 |
| 19 | 10 402 560 | 0,051 | 4 946 417 280 | 0,752 | 272 302 894 080 | 0,982 | 0,988 |
| 20 | 3 230 080 | 0,012 | 6 068 256 768 | 0,680 | 417 512 148 480 | 0,974 | 0,980 |
| 21 |  |  | 6 160 876 416 | 0,592 | 404 080 834 560 | 0,962 | 0,968 |
| 22 |  |  | 7 730 384 256 | 0,502 | 743 278 970 880 | 0,948 | 0,952 |
| 23 |  |  | 6 292 581 120 | 0,389 | 829 743 344 640 | 0,927 | 0,930 |
| 24 |  |  | 6 900 969 600 | 0,298 | 1 202 317 401 600 | 0,903 | 0,905 |
| 25 |  |  | 5 479 802 496 | 0,197 | 1 334 577 484 800 | 0,869 | 0,874 |
| 26 |  |  | 4 327 787 520 | 0,118 | 1 773 862 272 000 | 0,831 | 0,834 |
| 27 |  |  | 2 399 241 600 | 0,055 | 1 878 824 586 240 | 0,781 | 0,786 |
| 28 |  |  | 1 197 020 160 | 0,020 | 2 496 636 103 680 | 0,727 | 0,727 |
| 29 |  |  | 16 3094 400 | 0,0$^2$24 | 2 406 981 104 640 | 0,656 | 0,659 |
| 30 |  |  | 3 230 080 | 0,0$^4$5 | 3 032 021 672 960 | 0,588 | 0,583 |
| 31 |  |  |  |  | 2 841 072 675 840 | 0,502 | 0,500 |
| 32 |  |  |  |  | 3 166 378 709 760 | 0,421 | 0,413 |
| 33 |  |  |  |  | 2 743 311 191 040 | 0,331 | 0,325 |
| 34 |  |  |  |  | 2 877 794 035 200 | 0,253 | 0,243 |
| 35 |  |  |  |  | 2 109 852 702 720 | 0,171 | 0,170 |
| 36 |  |  |  |  | 1 840 136 336 640 | 0,111 | 0,109 |
| 37 |  |  |  |  | 1 109 253 196 800 | 0,059 | 0,063 |
| 38 |  |  |  |  | 689 719 564 800 | 0,028 | 0,032 |
| 39 |  |  |  |  | 230 683 084 800 | 0,0$^2$79 | 0,014 |
| 40 |  |  |  |  | 48 251 508 480 | 0,0$^2$14 | 0,005 |

**Таблица 10А. Согласованность парных сравнений.**

Вероятность ($P$) того, что значение $\Sigma$ (для $u$) будет достигнуто или превышено при $m=3$ и $n$ от 2 до 8.

| $n=2$ | | $n=3$ | | $n=4$ | | $n=5$ | | $n=6$ | | $n=7$ | | $n=8$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\Sigma$ | $P$ | $\Sigma$ | $P$ | $\Sigma$ | $P$ | $\Sigma$ | $P$ | $\Sigma$ | $P$ | $\Sigma$ | $P$ | $\Sigma$ | $P$ |
| 1 | 1,000 | 3 | 1,000 | 6 | 1,000 | 10 | 1,000 | 15 | 1,000 | 21 | 1,000 | 28 | 1,000 |
| 3 | 0,250 | 5 | 0,578 | 8 | 0,822 | 12 | 0,944 | 17 | 0,987 | 23 | 0,998 | 30 | 1,000 |
| | | 7 | 0,156 | 10 | 0,466 | 14 | 0,756 | 19 | 0,920 | 25 | 0,981 | 32 | 0,997 |
| | | 9 | 0,016 | 12 | 0,169 | 16 | 0,474 | 21 | 0,764 | 27 | 0,925 | 34 | 0,983 |
| | | | | 14 | 0,038 | 18 | 0,224 | 23 | 0,539 | 29 | 0,808 | 36 | 0,945 |
| | | | | 16 | 0,0046 | 20 | 0,078 | 25 | 0,314 | 31 | 0,633 | 38 | 0,865 |
| | | | | 18 | $0{,}0^{3}24$ | 22 | 0,020 | 27 | 0,148 | 33 | 0,433 | 40 | 0,736 |
| | | | | | | 24 | 0,0035 | 29 | 0,057 | 35 | 0,256 | 42 | 0,572 |
| | | | | | | 26 | $0{,}0^{3}42$ | 31 | 0,017 | 37 | 0,130 | 44 | 0,400 |
| | | | | | | 28 | $0{,}0^{4}30$ | 33 | 0,0042 | 39 | 0,056 | 46 | 0,250 |
| | | | | | | 30 | $0{,}0^{6}96$ | 35 | $0{,}0^{3}79$ | 41 | 0,021 | 48 | 0,138 |
| | | | | | | | | 37 | $0{,}0^{3}12$ | 43 | 0,0064 | 50 | 0,068 |
| | | | | | | | | 39 | $0{,}0^{4}12$ | 45 | 0,0017 | 52 | 0,029 |
| | | | | | | | | 41 | $0{,}0^{6}92$ | 47 | $0{,}0^{3}37$ | 54 | 0,011 |
| | | | | | | | | 43 | $0{,}0^{7}43$ | 49 | $0{,}0^{4}68$ | 56 | 0,0038 |
| | | | | | | | | 45 | $0{,}0^{9}93$ | 51 | $0{,}0^{4}10$ | 58 | 0,0011 |
| | | | | | | | | | | 53 | $0{,}0^{5}12$ | 60 | $0{,}0^{3}29$ |
| | | | | | | | | | | 55 | $0{,}0^{6}12$ | 62 | $0{,}0^{4}66$ |
| | | | | | | | | | | 57 | $0{,}0^{8}86$ | 64 | $0{,}0^{4}13$ |
| | | | | | | | | | | 59 | $0{,}0^{9}44$ | 66 | $0{,}0^{5}22$ |
| | | | | | | | | | | 61 | $0{,}0^{10}15$ | 68 | $0{,}0^{6}32$ |
| | | | | | | | | | | 63 | $0{,}0^{12}23$ | 70 | $0{,}0^{7}40$ |
| | | | | | | | | | | | | 72 | $0{,}0^{8}42$ |
| | | | | | | | | | | | | 74 | $0{,}0^{9}36$ |
| | | | | | | | | | | | | 76 | $0{,}0^{10}24$ |
| | | | | | | | | | | | | 78 | $0{,}0^{11}13$ |
| | | | | | | | | | | | | 80 | $0{,}0^{13}48$ |
| | | | | | | | | | | | | 82 | $0{,}0^{14}12$ |
| | | | | | | | | | | | | 84 | $0{,}0^{16}14$ |

Таблица 10B. **Согласованность парных сравнений.**

Вероятность ($P$) того, что значение $\Sigma$ (для $u$) будет достигнуто или превышено при $m=4$ и $n$ от 2 до 6 (для $n=6$ приведены только значения, превосходящие 1%-ный уровень)

| $n=2$ | | $n=3$ | | $n=4$ | | $n=5$ | | $n=5$ | | $n=6$ | | $n=6$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\Sigma$ | $P$ | $\Sigma$ | $P$ | $\Sigma$ | $P$ | $\Sigma$ | $P$ | $\Sigma$ | $P$ | $\Sigma$ | $P$ | $\Sigma$ | $P$ |
| 2 | 1,000 | 6 | 1,000 | 12 | 1,000 | 20 | 1,000 | 42 | 0,0048 | 57 | 0,014 | 79 | $0{,}0^{8}42$ |
| 3 | 0,625 | 7 | 0,947 | 13 | 0,997 | 21 | 1,000 | 43 | 0,0030 | 58 | 0,0092 | 80 | $0{,}0^{8}28$ |
| 6 | 0,125 | 8 | 0,736 | 14 | 0,975 | 22 | 0,999 | 44 | 0,0017 | 59 | 0,0058 | 81 | $0{,}0^{9}98$ |
| | | 9 | 0,455 | 15 | 0,901 | 23 | 0,995 | 45 | $0{,}0^{3}73$ | 60 | 0,0037 | 82 | $0{,}0^{9}15$ |
| | | 10 | 0,330 | 16 | 0,769 | 24 | 0,979 | 46 | $0{,}0^{3}41$ | 61 | 0,0022 | 83 | $0{,}0^{9}12$ |
| | | 11 | 0,277 | 17 | 0,632 | 25 | 0,942 | 47 | $0{,}0^{3}24$ | 62 | 0,0013 | 84 | $0{,}0^{10}51$ |
| | | 12 | 0,137 | 18 | 0,524 | 26 | 0,882 | 48 | $0{,}0^{4}90$ | 63 | $0{,}0^{3}76$ | 86 | $0{,}0^{11}30$ |
| | | 14 | 0,043 | 19 | 0,410 | 27 | 0,805 | 49 | $0{,}0^{4}37$ | 64 | $0{,}0^{3}44$ | 87 | $0{,}0^{11}17$ |
| | | 15 | 0,025 | 20 | 0,278 | 28 | 0,719 | 50 | $0{,}0^{4}25$ | 65 | $0{,}0^{3}23$ | 90 | $0{,}0^{13}28$ |
| | | 18 | 0,0020 | 21 | 0,185 | 29 | 0,621 | 51 | $0{,}0^{5}93$ | 66 | $0{,}0^{3}13$ | | |
| | | | | 22 | 0,137 | 30 | 0,514 | 52 | $0{,}0^{5}21$ | 67 | $0{,}0^{4}72$ | | |
| | | | | 23 | 0,088 | 31 | 0,413 | 53 | $0{,}0^{5}17$ | 68 | $0{,}0^{4}36$ | | |
| | | | | 24 | 0,044 | 32 | 0,327 | 54 | $0{,}0^{6}74$ | 69 | $0{,}0^{4}18$ | | |
| | | | | 25 | 0,027 | 33 | 0,249 | 56 | $0{,}0^{7}66$ | 70 | $0{,}0^{5}97$ | | |
| | | | | 26 | 0,019 | 34 | 0,179 | 57 | $0{,}0^{7}38$ | 71 | $0{,}0^{5}47$ | | |
| | | | | 27 | 0,0079 | 35 | 0,127 | 60 | $0{,}0^{9}93$ | 72 | $0{,}0^{5}20$ | | |
| | | | | 28 | 0,0030 | 36 | 0,090 | | | 73 | $0{,}0^{5}10$ | | |
| | | | | 29 | 0,0025 | 37 | 0,060 | | | 74 | $0{,}0^{6}51$ | | |
| | | | | 30 | 0,0011 | 38 | 0,038 | | | 75 | $0{,}0^{6}18$ | | |
| | | | | 32 | $0{,}0^{3}16$ | 39 | 0,024 | | | 76 | $0{,}0^{7}78$ | | |
| | | | | 33 | $0{,}0^{4}95$ | 40 | 0,016 | | | 77 | $0{,}0^{7}44$ | | |
| | | | | 36 | $0{,}0^{5}38$ | 41 | 0,0088 | | | 78 | $0{,}0^{7}15$ | | |

Таблица 10С. **Согласованность парных сравнений.**

Вероятность ($P$) того, что значение $\Sigma$ (для $u$) будет достигнуто или превышено при $m=5$ и $n$ от 2 до 5

| $n=2$ | | $n=3$ | | $n=4$ | | $n=5$ | | $n=5$ | |
|---|---|---|---|---|---|---|---|---|---|
| $\Sigma$ | $P$ | $\Sigma$ | $P$ | $\Sigma$ | $P$ | $\Sigma$ | $P$ | $\Sigma$ | $P$ |
| 4 | 1,000 | 12 | 1,000 | 24 | 1,000 | 40 | 1,000 | 76 | $0{,}0^{4}50$ |
| 6 | 0,375 | 14 | 0,756 | 26 | 0,940 | 42 | 0,991 | 78 | $0{,}0^{4}16$ |
| 10 | 0,063 | 16 | 0,390 | 28 | 0,762 | 44 | 0,945 | 80 | $0{,}0^{5}50$ |
| | | 18 | 0,207 | 30 | 0,538 | 46 | 0,843 | 82 | $0{,}0^{5}15$ |
| | | 20 | 0,103 | 32 | 0,353 | 48 | 0,698 | 84 | $0{,}0^{6}39$ |
| | | 22 | 0,030 | 34 | 0,208 | 50 | 0,537 | 86 | $0{,}0^{6}10$ |
| | | 24 | 0,011 | 36 | 0,107 | 52 | 0,384 | 88 | $0{,}0^{7}23$ |
| | | 26 | 0,0039 | 38 | 0,053 | 54 | 0,254 | 90 | $0{,}0^{8}53$ |
| | | 30 | $0{,}0^{3}24$ | 40 | 0,024 | 56 | 0,158 | 92 | $0{,}0^{8}12$ |
| | | | | 42 | 0,0093 | 58 | 0,092 | 94 | 0, 14 |
| | | | | 44 | 0,0036 | 60 | 0,050 | 96 | $0{,}0^{10}46$ |
| | | | | 46 | 0,0012 | 62 | 0,026 | 100 | $0{,}0^{12}91$ |
| | | | | 48 | $0{,}0^{3}36$ | 64 | 0,012 | | |
| | | | | 50 | $0{,}0^{3}12$ | 66 | 0,0057 | | |
| | | | | 52 | $0{,}0^{4}28$ | 68 | 0,0025 | | |
| | | | | 54 | $0{,}0^{5}54$ | 70 | 0,0010 | | |
| | | | | 56 | $0{,}0^{5}18$ | 72 | $0{,}0^{3}39$ | | |
| | | | | 60 | $0{,}0^{7}60$ | 74 | $0{,}0^{3}14$ | | |

Таблица 10D. **Согласованность парных сравнений.**

Вероятность ($P$) того, что значение $\Sigma$ (для $u$) будет достигнуто или превышено при $m=6$ и $n$ от 2 до 4

| $n=2$ | | $n=3$ | | $n=4$ | | $n=4$ | | $n=4$ | |
|---|---|---|---|---|---|---|---|---|---|
| $\Sigma$ | $P$ | $\Sigma$ | $P$ | $\Sigma$ | $P$ | $\Sigma$ | $P$ | $\Sigma$ | $P$ |
| 6 | 1,000 | 18 | 1,000 | 36 | 1,000 | 55 | 0,043 | 74 | $0{,}0^{4}12$ |
| 7 | 0,688 | 19 | 0,969 | 37 | 0,999 | 56 | 0,029 | 75 | $0{,}0^{5}89$ |
| 10 | 0,219 | 20 | 0,832 | 38 | 0,991 | 57 | 0,020 | 76 | $0{,}0^{5}49$ |
| 15 | 0,031 | 21 | 0,626 | 39 | 0,959 | 58 | 0,016 | 77 | $0{,}0^{5}32$ |
| | | 22 | 0,523 | 40 | 0,896 | 59 | 0,011 | 80 | $0{,}0^{6}68$ |
| | | 23 | 0,468 | 41 | 0,822 | 60 | 0,0072 | 81 | $0{,}0^{6}17$ |
| | | 24 | 0,303 | 42 | 0,755 | 61 | 0,0049 | 82 | $0{,}0^{6}12$ |
| | | 26 | 0,180 | 43 | 0,669 | 62 | 0,0034 | 85 | $0{,}0^{7}34$ |
| | | 27 | 0,147 | 44 | 0,556 | 63 | 0,0025 | 90 | $0{,}0^{8}93$ |
| | | 28 | 0,088 | 45 | 0,466 | 64 | 0,0016 | | |
| | | 29 | 0,061 | 46 | 0,409 | 65 | $0{,}0^{3}83$ | | |
| | | 30 | 0,040 | 47 | 0,337 | 66 | $0{,}0^{3}66$ | | |
| | | 31 | 0,034 | 48 | 0,257 | 67 | $0{,}0^{3}48$ | | |
| | | 32 | 0,023 | 49 | 0,209 | 68 | $0{,}0^{3}26$ | | |
| | | 35 | 0,0062 | 50 | 0,175 | 69 | $0{,}0^{3}16$ | | |
| | | 36 | 0,0029 | 51 | 0,133 | 70 | $0{,}0^{4}86$ | | |
| | | 37 | 0,0020 | 52 | 0,097 | 71 | $0{,}0^{4}68$ | | |
| | | 40 | $0{,}0^{3}58$ | 53 | 0,073 | 72 | $0{,}0^{4}48$ | | |
| | | 45 | $0{,}0^{4}31$ | 54 | 0,057 | 73 | $0{,}0^{4}16$ | | |

## ОГЛАВЛЕНИЕ

**Кендэл М.**

**К35** Ранговые корреляции. — Зарубежные статистические исследования. М., «Статистика», 1975.
216 с. с ил.

Цель книги — дать полную картину применения методов изучения ранговых корреляций в различных областях науки и практики, в том числе и в экономике.

Книга интересна всем, кто в той или иной степени сталкивается с экспертными оценками в своей работе.

К $\frac{10803-023}{008(01)-75}$ 127—75

*МОРИС ДЖ. КЕНДЭЛ*

**Ранговые корреляции**

---

Редактор *К. М. Чижевская*
Техн. редактор *Г. А. Сидорова*
Корректор *А. Т. Сидорова*
Худ. редактор *Т. В. Стихно*
Переплет художника *Л. С. Эрмана*

---

Сдано в набор 21/VIII 1974 г.
Подписано к печати 14/I—75 г.
Формат бумаги $60 \times 90^1/_{16}$ Бумага № 2
Объем 13,5 печ. л. Уч.-изд. л. 13,55
Тираж 8000 экз.
(Тематич. план 1975 г. № 127)

---

Издательство «Статистика»,
Москва, ул. Кирова, 39.
Заказ 1097 Цена 1 р. 02 к.

Московская типография № 4
Союзполиграфпрома при Государственном
комитете Совета Министров СССР
по делам издательств, полиграфии
и книжной торговли,
Москва, И-41, Б. Переяславская ул., д. 46

ЗАРУБЕЖНЫЕ СТАТИСТИЧЕСКИЕ ИССЛЕДОВАНИЯ