

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Майкопский государственный технологический университет»

БИГАНОВА С.Г.

БИОМЕТРИЯ

Учебное пособие

Майкоп – 2015

УДК 57.087.1 (07)
ББК 28.0
Б-59

Печатается по решению Научно-технического совета ФГБОУ ВО «Майкопский государственный технологический университет».

Рецензент – доктор с.-х. наук, профессор **Сухоруких Ю.И.**

Автор – кандидат с.-х. наук, доцент **Биганова С.Г.**

БИОМЕТРИЯ. Учебное пособие. – Майкоп: изд-во МГТУ, 2015. – 100 с.

В учебном пособии изложены методологические и методические основы биометрии, современные представления о системе статистических методов исследования биологических объектов в научных и практических целях. Дан обзор законов распределения, приведены основные критерии и методы анализа.

Пособие предназначено для аспирантов, обучающихся по биологическим, сельскохозяйственным и лесохозяйственным специальностям.

За стилистику и орфографию ответственность несет автор.

© Биганова С.Г.,
Майкоп: МГТУ, 2015

ВВЕДЕНИЕ

Для изучения и освоения методов биометрии предназначено предлагаемое учебное пособие, которое подготовлено в соответствии с программой учебной дисциплины «Биометрия».

Учебное пособие состоит из введения, восьми глав, списка литературы, приложений.

Первая глава посвящена взаимоотношениям биологии и математической статистики. В ней показано единство объекта их исследования, раскрыты представления о биологических объектах и математическом подходе исследования, охарактеризована история развития биометрии как науки.

Вторая глава вводит студента в сферу вариативных основ биометрии. В ней дан также обзор средних величин, показаны структурные средние, их способы вычисления. Охарактеризованы показатели вариации, раскрыты закономерности варьирования.

В третьей главе раскрыты понятия теории вероятности. Кроме того, здесь рассматриваются основные законы распределений случайных величин. Особое внимание уделено нормальному распределению, как наиболее часто встречающемуся в природе.

Четвертая глава посвящена описанию структуры совокупности и выборки. Раскрыты понятия точечных и интервальных оценок. Дано понятие доверительного интервала.

Пятая глава дает представление о критериях достоверности оценок. Раскрыты сущность и содержание нулевой гипотезы и альтернативной. Дан обзор параметрических критериев, показаны возможности их применения. Для распределений, не следующих нормальному закону, приведены непараметрические критерии.

Шестая глава продолжает раскрывать широкий спектр применения критериев. В ней рассказывается о проверке гипотез о законах распределения. Дано применение коэффициентов асимметрии и эксцесса для проверки нормальности распределения. Рассмотрены хи-квадрат распределение, критерии Ястремского и Уилка-Шапиро. Дано представление об оценке регрессии рядов и проверки сомнительных вариантов.

Седьмая глава посвящена дисперсионному анализу. Раскрыта суть метода. Даны основные понятия и символы. Показан анализ однофакторных комплексов, ранговый анализ, и применение корреляционных таблиц. Рассмотрены методы Плохинского и Снедекора для оценки силы влияния факторов.

Восьмая глава даёт основы корреляционного анализа. Показаны взаимосвязь функциональной зависимости и корреляция. Дано определение коэффициента корреляции как показателя связи, его преимущества и недостатки. Рассмотрены способы вычисления коэффициента корреляции. Дано понятие коэффициента детерминации.

В приложениях приведены греческий и латинский алфавиты, используемые в биометрии. Даны таблицы с критическими значениями.

Глава 1. ОСНОВНЫЕ ПОНЯТИЯ

Принимающий всё более высокие темпы развития научно-технический прогресс предъявляет к специалистам в области биологии, лесного хозяйства новые, повышенные требования. Современные специалисты должны не только хорошо знать свою специальность, но и вести исследовательскую работу.

Биометрия – наука, изучающая биологические объекты и явления, анализирующая их и позволяющая делать научно-обоснованные выводы.

Предметом биометрии служит любой биологический объект. Он изучается с помощью количественной оценки его качественных свойств. При этом имеются в виду массовые явления, в которых проявляются статистические законы.

Обычно наблюдения проводят на групповых объектах, например на особях одного и того же вида, пола, возраста, которые рассматривают как составные элементы группового объекта и называют *единицами наблюдения*. Множество относительно однородных, но индивидуально различимых единиц, объединенных для группового изучения, называют *статистической совокупностью*.

Понятие статистической совокупности – одно из фундаментальных понятий в биометрии. Оно основано на принципе качественной однородности состава совокупности. Нельзя объединять в одну совокупность особей разного пола или возраста. Неправомерно изучать закономерность модификационной изменчивости на генетически неоднородном материале, объединяя в одну совокупность чистопородных и гибридных особей.

Существует также понятие *статистического комплекса*. Он складывается из разнородных групп, объединяемых для совместного изучения. При этом каждая группа, входящая в состав комплекса, должна состоять из однородных элементов. Например, при испытании различных доз удобрений каждую опытную делянку рассматривают как отдельную группу, входящую в состав статистического комплекса.

Вопрос об объединении биометрических данных экспериментатор решает сам, исходя из имеющегося объекта и цели исследования.

Необходимым условием грамотного использования в биометрии статистических методов является чёткое осознание смысла тех терминов и понятий, которые используются в математической статистике и теории вероятностей и конкретного применения их к исследуемым объектам и процессам.

Величины, которые не меняют своих значений в процессе наблюдений, называются *постоянными* (константами). Если величина может принимать два или более различных значений, то она называется *переменной*.

Явления, происходящие в результате наблюдений, в математической статистике и теории вероятностей принято называть *событиями*. *Случайными событиями* называются события, которые не могут быть точно предсказаны. Переменная величина, значение которой до проведения испытания заранее не может быть точно предсказано, называется *случайной величиной*. В биометрии исследователь постоянно имеет дело со случайными величинами, каковыми являются свойства и признаки биологических объектов.

В ряде случаев значения переменной прямо или косвенно выбирают или устанавливают исследователем и тогда переменная является неслучай-

ной или *детерминированной*. Такими переменными являются, например нормы полива, а также свойства специально подобранных тел (содержание питательных элементов в участках, получивших разные нормы удобрений и т.д.).

Статистическая совокупность всех возможных значений случайной величины называется *генеральной статистической совокупностью*. Число компонентов совокупности называется *объёмом совокупности*. Объём физической и соответственно статистической совокупности всегда конечен. В отличие от них, генеральная совокупность имеет бесконечный объём. Для биологических объектов, имеющих непрерывные значения признаков, объём всегда бесконечен.

Когда объект дискретный, объём генеральной совокупности конечен. Например, при изучении состава древостоя некоторого участка леса, генеральная физическая совокупность определяется однозначно по виду деревьев, и объём её конечен и равен числу деревьев на участке.

1.1. Признаки и их классификация

Под словом *признак* понимают свойство, проявлением которого один объект отличается от другого. В области биологии признаками, по которым проводят наблюдения над объектами, служат такие характерные особенности в строении и функциях живого, которые позволяют отличать одну единицу наблюдения от другой. Например, исследователя интересует годовой прирост на яблоне, выращиваемой на специально отведенном участке. Массив данной культуры будет объектом наблюдения, а признаком – годовой прирост ветвей на отдельных яблонях. Которые являются единицами наблюдения, составляя в общей массе статистическую совокупность.

Характерным свойством биологических признаков является *варьирование* величины признака в определённых пределах при переходе от одной единицы наблюдения к другой. Например, измеряя годовой прирост ветвей яблонь нетрудно заметить, что величина признака колеблется, образуя совокупность числовых значений признака, по которому проводят наблюдение. Колебания величины одного и того же признака, наблюдаемые в массе однородных членов статистической совокупности, называют *вариациями* (от лат VARIATIO – различимый, меняющийся), а отдельные числовые значения варьирующего признака принято называть *вариантами*.

Все биологические признаки варьируют. И все они поддаются измерению в каких-либо величинах. Признаки принято делить на:

- *Качественные* или *атрибутивные*
- *Количественные*

Качественные признаки не поддаются непосредственному измерению и определяются по силе проявления признака или по наличию их свойств у отдельных членов изучаемой группы. Например, среди растений подсчитывают количество экземпляров с разной окраской цветов (белой, красной, желтой, фиолетовой и т.д.), среди плодов – вкус (очень хороший, хороший, посредственный, плохой, очень плохой).

Количественные признаки поддаются непосредственному измерению или счёту. Их делят на *мерные* или *метрические* и *счётные* или *меристические*. Длина годового прироста ветвей, урожайность яблони – это всё мерные призна-

ки. Они варьируют непрерывно: их величина может принимать в некоторых пределах любые числовые значения. Счётные признаки – число листьев на ветви, число плодов с одного дерева и др. варьируют дискретно (прерывисто). Их числовые значения могут принимать вид только целых чисел (1, 2, 3, 4 и т.д.).

Частным случаем качественных признаков являются *альтернативные*, когда признак может принимать только два значения: признак имеется и признак отсутствует. Например: больные – здоровые, высокие – низкие, гнедые – не гнедые и т.д.

Биологические признаки варьируют под влиянием самых различных, в том числе случайных, причин. Помимо естественного варьирования, на величине признака сказываются и ошибки, неизбежно возникающие при измерении изучаемых объектов. *Погрешностью* или *ошибкой* называется разница между результатами измерений и действительно существующими значениями измеряемой величины.

Ошибки возникают из-за неточности или неисправности приборов измерения (технические ошибки), личных качеств исследователя, его мастерства и навыков (личные ошибки), и от целого ряда других неустранимых и не поддающихся регулированию причин (случайные ошибки). Первые две категории ошибок можно преодолеть, совершенствуя технические средства, условия работы и опыт. Однако случайные ошибки остаются. Но по сравнению с естественным варьированием случайные ошибки, как правило, невелики. Поэтому варьирование результатов испытаний рассматривают как естественное варьирование.

1.2. Точность измерений

При проведении исследований мы имеем дело с измерениями биологических объектов. Обычно измерения проводят с точностью до десятых, сотых или тысячных долей единицы. Более точные измерения проводят достаточно редко. Результат вычисления не может быть точнее тех данных, на которых оно основано.

Обычно исследователю приходится оперировать приближёнными числами, полученными в результате измерений. Многие математические операции, такие как деление, нахождение корня чисел, логарифма чисел и т.д. также дают приближённые числа.

Для того чтобы не допускать в работе ошибок и получать сопоставимые результаты, необходимо соблюдать правила записи и округления приближённых чисел. *Числа, занесённые в документы учёта измерений должны соответствовать точности, принятой при измерении варьирующих объектов.* Например, если измерения проводят с точностью до двух знаков, то результаты нельзя записать так: 5,2; 6; 7,69; 9,037 и т.д. Правильно будет записать так: 5,20; 6,00; 7,69; 9,04.

Числа округляют следующим образом: если за последней сохраняемой цифрой следуют цифры 0, 1, 2, 3, 4, они отбрасываются (*округление с недостатком*); если же за последней сохраняемой цифрой следуют цифры 5, 6, 7, 8 и 9, то последняя сохраняемая цифра увеличивается на единицу (*округление с избытком*). Например, числа 8,266; 45,123; 32,788; 5,981 округляются до двух десятичных знаков следующим образом: 8,27; 45,12; 32,79; 5,98.

1.3. Репрезентативность выборки и рандомизация

Целью исследователя является изучение биологического объекта с помощью свойств генеральной совокупности. При этом генеральные совокупности почти никогда не анализируются целиком, это технически сложно и не всегда нужно. В большинстве случаев о свойствах объектов судят на основании испытаний над некоторым конечным числом компонентов физической совокупности, называемом *выборкой*. С её помощью исследователь судит о свойствах генеральной совокупности, и тем самым об объекте. Именно поэтому крайне важно, чтобы выборка правильно отражала свойства объекта, чтобы выборка была представительна, *репрезентативна*. Лучшим способом обеспечения репрезентативности выборки считается рандомизация, т.е. случайность отбора компонентов физической совокупности. Случайность отбора компонентов на анализ совсем не означает, что надо брать что попало и где попало. При соблюдении общих условий проведения испытания для обеспечения случайности пользуются некоторыми правилами, регламентирующими выбор компонента в пространстве объекта (а иногда и во времени).

Существует два основных способа отбора компонентов генеральной совокупности: повторный и бесповторный. *Повторный отбор* производят по схеме «возвращения» учтённых единиц в генеральную совокупность, поэтому одна и та же единица может попасть в выборку повторно. При *бесповторном отборе* учтённые единицы в генеральную совокупность не возвращаются, поэтому каждая компонента регистрируется только один раз. На практике обычно применяют *бесповторный случайный отбор*.

Идеальный случайный отбор производится с помощью таблицы случайных чисел (Таблица 1 ПРИЛОЖЕНИЙ). В практике применяют и другие виды выборки из генеральной совокупности.

Типический отбор используют в тех случаях, когда генеральная совокупность расчленяется на отдельные типические группы. Например, в питомнике произрастают саженцы 1-х годовичные, 2-х годовичные, 3-х годовичные и т.д. В таких случаях из каждой группы случайным способом отбирают одинаковое, а чаще пропорциональное число единиц. Затем вычисляют характеристики, объединяемые в общую характеристику генеральной совокупности.

При *серийном отборе*, генеральную совокупность делят на группы (серии, гнёзда), образуемые обычно по территориальному принципу. Затем из общего количества серий отбирают их некоторое количество, для совместной обработки. Размерность серий может быть и одинаковой и различной. Например, из 40 групп студентов в возрасте от 18 до 19 лет намечено обследовать 7 групп выборочно. Членов этих групп и объединяют в выборку для совместного изучения. В отличие от типического отбора при серийном отборе из генеральной совокупности выбирают не отдельные единицы, а целые серии или гнёзда относительно однородных единиц.

При *механическом отборе* генеральная совокупность разбивается на несколько равных частей или групп. Затем из каждой части случайным образом отбирают по одной единице. Например, при обследовании посева пшеницы на урожайность намечено отобрать 100 растений. В этом случае поле должно быть разбито на 100 равных делянок. При механическом отборе число единиц равно численности групп, на которые разбита генеральная совокупность.

Может использоваться и другая схема. В выборку должны попадать каждая 10 или 100 или другая единица генеральной совокупности. Например, при проведении ботанических или зоологических исследований можно регистрировать каждый десятый экземпляр искомого вида растения или животного.

1.4. Способы группировки первичной информации

Результаты наблюдений или измерений должны заноситься в дневник, журнал, бланк или другую форму учёта. Эти сведения представляют собой первичный материал, который необходимо обработать. Обработка начинается с упорядочения данных. Систематизация данных, их объединение в относительно однородные группы по некоторому признаку называется *группировкой*.

Группировка направлена на выявление связей между явлениями. Это ответственный этап, от него зависят выводы о природе изучаемого явления. Нельзя группировать в одну совокупность неоднородные по составу данные.

Таблицы. Наиболее распространённой формой группировки являются статистические таблицы. Они бывают простыми и сложными. К простым относятся четырёхпольные применяемые при альтернативной группировке.

Например:

Сорт	Число плодов		Всего
	одномерных	Не одномерных	
№1	85	15	100
№5	88	12	100
№8	70	30	100
Итого	243	57	300

К сложным относятся многопольные таблицы, применяемые при изучении корреляционной зависимости и при анализе варьирующих признаков. Таблицы имеют не только иллюстративное, но и аналитическое значение, позволяющее обнаружить связь между варьирующими признаками.

Статистические ряды. *Статистическим* называется ряд числовых значений признака, расположенных в определённом порядке. В зависимости от изучаемых признаков ряды делятся на атрибутивные, вариационные, ряды динамики, и регрессии, а также ряды ранжированных значений признаков, ряды накопленных частот, являющимися производными вариационных рядов. Примером атрибутивного ряда служат данные, показывающие зависимость плотности древесины от породы дерева.

Порода	дуб	бук	пихта	сосна	липа
Плотность древесины	6	5	4,2	3,8	2,9

Вариационным рядом или рядом распределения называют двойной ряд чисел, показывающий, каким образом числовые значения признака связаны с их повторяемостью в данной статистической совокупности. Например, из урожая картофеля, собранного на опытной делянке, случайным образом отобрано 25 клубней, в которых подсчитывали число глазков. Результаты следующие: 6, 9, 5, 7, 10, 8, 9, 10, 8, 11, 9, 12, 9, 8, 10, 11, 9, 10, 8, 10, 7, 9, 11, 9, 10. Расположим эти данные в ряд с учётом повторяемости вариантов в этой совокупности:

Варианты x_i	6	9	5	7	10	8	11	12
Число вариант f_i	1	7	1	2	6	4	3	1

Это и есть вариационный ряд. Числа, показывающие, сколько раз отдельные варианты встречаются в данной совокупности, называются *частотами* или *весами* вариант и обозначаются строчной буквой латинского алфавита f . Общая сумма частот вариационного ряда равна объёму данной совокупности

$$\sum_{i=1}^k f_i = n,$$

где $\sum_{i=1}^k$ – (греческая буква сигма прописная) обозначает действие суммирования, в данном случае суммирование частот вариационного ряда от первого ($i=1$) до k -го класса, а n – общее число наблюдений, или объём совокупности.

Частоты (веса) выражают и в относительных величинах – в долях единицы или в процентах от общей численности вариант. В таких случаях веса называют *частостями* или *относительными частотами*.

$$\sum \frac{f_i}{n} = 1 \quad \text{или} \quad \sum \left(\frac{f_i}{n} \right) \cdot 100 = 100\%,$$

если частоты выражены в процентах от общего числа наблюдений n . Относительные частоты удобно применять для сопоставления вариационных рядов различных по объёмам.

Вариационный ряд строят по ранжированным значениям признака. Под *ранжированием* (от франц. ranger – выстраивать в ряд по ранжиру, т.е. по росту) понимают расположение членов ряда в возрастающем или убывающем порядке. Наш пример будет выглядеть следующим образом

Варианты x_i	5	6	7	8	9	10	11	12
Число вариант f_i	1	1	2	4	7	6	3	1

Этот упорядоченный ряд хорошо иллюстрирует закономерность варьирования признака. Если признак варьирует дискретно, частоты относятся непосредственно к ранжированным значениям признака, которые приобретают положение отдельных групп или классов вариационного ряда. Если признак варьирует непрерывно, то подсчитывают частоты, относящиеся к отдельным промежуткам или интервалам (от – до) на которые разбивается общая вариация признака в пределах от минимальной варианты до максимальной варианты данной совокупности. Классовые интервалы могут быть равными и не равными по ширине. Отсюда различают *равно- и неравноинтервальные вариационные ряды*. Как правило, биометрические данные распределяются в равноинтервальные ряды. Это облегчает вычисление сводных характеристик варьирования и сопоставление рядов друг с другом.

При построении вариационного ряда важно правильно определить ширину классового интервала. Очень широкий классовый интервал ведёт к искажению типичных черт варьирования и снижению точности числовых характеристик ряда. Чрезмерно узкий интервал даёт слишком растянутый ряд и неясную картину варьирования. Чтобы правильно разбить ряд на интервалы, делят размах варьирования на число классов, намечаемых при построении ряда.

$$\lambda = \frac{x_{\max} - x_{\min}}{K} \quad (1)$$

где λ – величина классового интервала; x_{\max} , x_{\min} – максимальная и минимальная варианты совокупности, K – число классов.

Величину K можно определить по формуле Стерджеса : $K = 1 + 3,32 \lg n$ если в совокупности членов более 100 то можно применять формулу

$$K = 5 \lg n$$

Если признак варьирует слабо ($\lambda = 1$), то данные распределяются в безинтервальный ряд. Если же признак варьирует в широких границах, то по данным строят интервальный вариационный ряд. Для удобства соотношение числа наблюдений и количества классов вариационного ряда приведены в таблице 1.

Таблица 1 - Зависимость количества классов от числа наблюдений

Число наблюдений	Число классов
25-40	5-6
40-60	6-8
60-100	7-10
100-200	8-12
>200	10-15

Вопросы к главе 1.

1. Что такое качественные, количественные признаки?
2. Привести примеры альтернативных признаков.
3. Предложить альтернативу к понятиям:
 - ранораспускающиеся формы дуба
 - быстрорастущие гибриды
 - устойчивые формы
 - повреждённые плоды
 - опущённые листья
4. Привести пример дискретных и непрерывных признаков.
5. В чём различие дискретных и непрерывных величин?
6. Привести примеры качественных, количественных (мерных и счётных) признаков.
7. Что такое событие?
8. Какое событие называется случайным?
9. Когда можно переменную величину считать случайной?
10. Что такое статистическая совокупность?
11. Когда можно считать генеральную статистическую совокупность элементов конечной и когда бесконечной:
 - а) для дискретных величин
 - б) для непрерывных величин
12. Что такое генеральная совокупность и что такое выборка?
13. Каковы причины варьирования результатов повторных испытаний?
14. Случайная и детерминированная переменные величины, в чём их различия?
15. Что такое объём совокупности?
16. Виды отбора из генеральной совокупности.
17. Что такое статистический ряд?
18. Что такое абсолютная частота?
19. Что такое относительная частота?

Глава 2. ОСНОВНЫЕ ХАРАКТЕРИСТИКИ ВАРЬИРУЮЩИХ ОБЪЕКТОВ

2.1. Средние величины

Вариационные ряды недостаточны для полного описания варьирующих объектов. Для этой цели служат особые логически и теоретически обоснованные числовые показатели, называемые статистическими характеристиками.

В отличие от индивидуальных числовых характеристик *средние величины* обладают большей устойчивостью, способностью характеризовать целую группу однородных единиц одним числом. Значение средних величин заключается в их свойстве аккумулировать или уравнивать все индивидуальные отклонения, в результате чего проявляется то наиболее устойчивое и типичное, что характеризует качественное своеобразие варьирующего объекта.

В зависимости от того, как распределены первичные данные в равно- или в неравноинтервальный вариационный ряд, для их характеристики применяют разные средние величины. Именно при распределении собранных данных в неравноинтервальный вариационный ряд более подходящей обобщающей характеристикой изучаемого объекта служит так называемая плотность распределения, т.е. отношение частот или *частостей* к ширине классовых интервалов. Кроме того, числовыми характеристиками таких рядов могут служить средние из абсолютных или относительных показателей плотности распределения.

В качестве статистических характеристик равноинтервальных вариационных рядов применяют *степенные* и *структурные* (нестепенные) *средние величины*.

$$M = \sqrt[R]{\frac{\sum x_i^R}{n}} \quad (2)$$

Средние величины могут характеризовать только однородную совокупность вариантов. При наличии разнородных по составу данных их необходимо группировать в отдельные качественно, однородные группы и вычислять групповые или частные средние.

Средняя арифметическая \bar{x} . Из общего семейства степенных средних наиболее часто используют *среднюю арифметическую*. Этот показатель является центром распределения, вокруг которого группируются все варианты статистической совокупности. Средняя арифметическая может быть *простой* и *взвешенной*. Простую среднюю арифметическую определяют как сумму всех членов совокупности, деленную на их общее число:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3)$$

В этой формуле X_i – значения вариантов; Σ – знак суммирования вариантов в пределах от первой (x_1) до n -й варианты; n – общее число вариантов, или объем данной совокупности.

Когда отдельные варианты повторяются, среднюю арифметическую вычисляют по формуле:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^R x_i f_i \quad (4)$$

и называют *взвешенной средней*, причем весами, как это показывает формула (5), служат частоты вариант f_i . При объединении групповых средних их весами будут объемы групп n_i , по которым эти средние вычислены. Общую (взвешенную) среднюю арифметическую нескольких однородных групп определяют по формуле

$$\bar{x} = \frac{\bar{x}_1 n_1 + \bar{x}_2 n_2 + \bar{x}_3 n_3 + \dots + \bar{x}_k n_k}{n_1 + n_2 + n_3 + \dots + n_k} = \frac{\sum (\bar{x}_i n_i)}{\sum n_i} \quad (5)$$

Средняя арифметическая – одна из основных характеристик варьирующих объектов. Она обладает рядом важных свойств.

1. Если каждую варианту статистической совокупности уменьшить или увеличить на некоторое произвольно взятое положительное число A , то и средняя уменьшится или увеличится на это число.

2. Если каждую варианту разделить или умножить на какое-либо одно и то же число A , то средняя арифметическая изменится во столько же раз.

3. Сумма произведений отклонений вариант от их средней арифметической на соответствующие им частоты равна нулю.

4. Сумма квадратов отклонений вариант от их средней \bar{x} меньше суммы квадратов отклонений тех же вариант от любой другой величины A , не равной \bar{x} , т.е.

$$\sum (x_i - \bar{x})^2 < \sum (x_i - A)^2 \quad (6)$$

Средняя гармоническая \bar{x}_h . Эту характеристику в отличие от средней арифметической, представляющей сумму вариант, отнесенную к их числу, определяют как *сумму обратных значений вариант, деленную на их число*. Для определения простой и *взвешенной* средней гармонической применяют формулы

$$\bar{x}_h = \frac{n}{\sum (1/x_i)} \quad \bar{x}_h = \frac{n}{\sum (1/x_i) f_i} \quad (7)$$

в которых n – число произведенных наблюдений; x_i – значения вариант; f_i – частоты.

Средняя гармоническая применяется тогда, когда результаты наблюдений обнаруживают обратную зависимость, заданы обратными значениями вариант.

Средняя квадратическая x_q . Для более точной числовой характеристики мер площади применяется средняя квадратическая. Этот показатель вычисляют по формулам

$$x_q = \sqrt{\frac{\sum x_i^2}{n}} \quad x_q = \sqrt{\frac{\sum f_i x_i^2}{n}} \quad (8)$$

Средняя кубическая x_o . В качестве характеристики объемных признаков более точной является *средняя кубическая*, определяемая по формулам

$$x_o = \sqrt[3]{\frac{\sum x_i^3}{n}} \quad x_o = \sqrt[3]{\frac{\sum f_i x_i^3}{n}} \quad (9)$$

Средняя геометрическая x_g : Этот показатель представляет собой корень n -й степени из произведений членов ряда. $x_g = \sqrt[n]{x_1 x_2 x_3 \dots x_n}$, где n – объем совокупности; при этом $x_i > 0$.

Обычно среднюю геометрическую вычисляют с помощью десятичных логарифмов по следующим рабочим формулам:

$$\overline{\lg x_g} = \frac{\sum \lg x_i}{n} \quad (10)$$

$$\lg \overline{x_g} = \frac{\sum \lg(x_2/x_1)}{n} \quad (11)$$

$$\lg \overline{x_g} = \frac{\lg x_k - \lg x_n}{n} \quad (12)$$

Формулу (10) применяют для вычисления средней геометрической из абсолютных прибавок величины признака; формулу (11) служит для вычисления средней геометрической из относительных прибавок величины признака за равные промежутки времени, а формулу (12) используют для вычисления средней геометрической по разности между конечной x_k и начальной x_n прибавками величины признака.

Средняя геометрическая – более точная характеристика рядов динамики, чем средняя арифметическая.

В заключение обзора степенных средних необходимо отметить, что между средними существуют определенные соотношения, выражаемые следующим рядом мажорантности (неравенства):

$$X_Q > X_q > X > X_g > X_h$$

2.2. Структурные средние и способы их вычисления

Медиана (Me). Средняя арифметическая очень чувствительна к увеличению числа наблюдений или к уменьшению за счет вариант, резко отличающихся по своей величине от основной массы.

Во многих случаях в качестве обобщающих характеристик совокупности более полезными могут оказаться так называемые *структурные средние*. Эти величины обычно представляют собой конкретные варианты имеющейся совокупности, которые занимают особое место в ряду распределения.

Одной из таких характеристик является медиана – средняя, относительно которой ряд распределения делится на две равные части: в обе стороны от медианы располагается одинаковое число вариант.

При четном числе членов ряда медиана определяется по полусумме двух соседних вариант, расположенных в центре ранжированного ряда.

Для данных, сгруппированных в вариационный ряд, медиана определяется следующим образом. Сначала находят класс, в котором содержится медиана. Для этого частоты ряда кумулируют в направлении от меньших к большим значениям классов до величины, превосходящей половину всех членов данной совокупности, т.е. $n/2$. Первая величина в ряду накопленных частот $\sum f_i$ которая превышает $n/2$, соответствует медианному классу. Затем берут разность между $n/2$ и суммой; накопленных частот $\sum f_i$, предшествующей медианному классу, которая относится к частоте медианного класса f_{Me} ; результат умножают на величину классового интервала λ . Найденную таким способом величину прибавляют к нижней границе x_h медианного класса. Описанные действия выражаются в виде следующей формулы:

$$Me = x_h \lambda \left(\frac{\frac{n}{2} - \sum f_i}{f_{Me}} \right) \quad (14)$$

где x_h – нижняя граница классового интервала, содержащего медиану или полусумма соседних классов безинтервального ряда, в промежутке между ко-

торыми находится медиана; Σf_i – сумма накопленных частот, стоящая перед медианным классом; f_{me} – частота медианного класса; λ – величина классового интервала; n – общее число наблюдений.

Мода (Mo). Модой называется величина, наиболее часто встречающаяся в данной совокупности. Класс с наибольшей частотой называется модальным.

$$Mo = x_n + \lambda \left(\frac{f_2 - f_1}{2f_2 - f_1 + f_3} \right) \quad (15)$$

где x_n – нижняя граница модального класса, т.е. класса с наибольшей частотой f_2 ; f_1 – частота класса, предшествующего модальному; f_3 – частота класса, следующего за модальным; λ – ширина классового интервала.

Квантили. Наряду с медианой и модой к структурным характеристикам вариационного ряда относятся так называемые квантили, отсекающие в пределах ряда определенную часть его членов. К ним относятся квартили, децили и перцентили (процентили). Квартили – это три значения признака (Q_1, Q_2, Q_3), делящие ранжированный вариационный ряд на четыре равные части. Аналогично, девять децилей делят ряд на 10 равных частей, а 99 перцентилей – на 100 равных частей. В практике используют обычно перцентили $P_3, P_{10}, P_{25}, P_{50}, P_{75}, P_{90}$ и P_{97} . Любой перцентиль определяется рядом последовательных действий, которые можно выразить в виде следующей формулы:

$$P_i = x_n + \lambda \left(\frac{K - \Sigma f_i}{f_p} \right) \quad (16)$$

где x_n – нижняя граница класса, содержащего перцентиль P_i ; она определяется по величине $K = Li_p/100$, превосходящей или равной Σf_i в ряду накопленных частот. Здесь P_i – выбранный перцентиль; n – общее число наблюдений; λ – ширина классового интервала; f_p – частота класса, содержащего искомый перцентиль; Li – так называемый порядок перцентилей, показывающий, какой процент наблюдений имеет меньшую величину, чем P_i .

2.3. Показатели вариации

Средние величины не являются универсальными характеристиками варьирующих объектов. При одинаковых средних признаки могут отличаться по величине и характеру варьирования. Поэтому наряду со средними для характеристики варьирующих признаков используют и *показатели вариации*. Одним из таких показателей являются *лимиты* (от лат. *limes* – предел), обозначаемые символом *lim*. В биометрии под этим термином понимают значения минимальной x_{min} и максимальной x_{max} вариант совокупности.

Размах вариации R. Это показатель, представляющий собой разность между максимальной и минимальной вариантами совокупности, т.е. $R = x_{min} - x_{max}$. Чем сильнее варьирует признак, тем больше размах вариации, и, наоборот, чем слабее вариация признака, тем меньше будет размах вариации.

Лимиты и размах вариации имеют существенные недостатки: при повторных измерениях одного и того же группового объекта они могут значительно изменяться; кроме того, они не отражают существенные черты варьирования.

Более удобной характеристикой вариации мог бы служить показатель, который строится на основании отклонений вариант от их средней, т.е. $(x_i - \bar{x}) = d$. Сумма таких отклонений, взятая без учета знаков и отнесенная к числу наблюдений n , называется *средним линейным отклонением*

$$\bar{d} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n} \quad (13)$$

Несмотря на явное преимущество среднего линейного отклонения перед лимитами и размахом вариации, этот показатель не получил широкого применения в биометрии. Наиболее подходящим оказался показатель, построенный не на отклонениях вариант от их средних, а на квадратах этих отклонений, его называют **дисперсией** (от лат. *dispersion* – рассеяние) и выражают формулами

$$\sigma^2_x = \frac{\sum_{i=1}^k (x_i - \bar{x})^2}{n} \quad \text{или} \quad \sigma^2_x = \frac{\sum_{i=1}^k f_1 (x_1 - \bar{x}_2)^2}{n} \quad (17)$$

где Σ – знак суммирования произведений отклонений вариант x_i от их средней \bar{x} на веса или частоты f_i , этих отклонений в пределах от первого до k -го класса; n – общее число наблюдений. Индекс x у символа дисперсии обозначает, что этот показатель характеризует варьирование числовых значений признака вокруг их средней величины.

Ценность дисперсии заключается в том, что, являясь мерой варьирования числовых значений признака вокруг их средней арифметической, она измеряет и внутреннюю изменчивость значений признака, зависящую от разностей между наблюдениями. Преимущество дисперсии перед другими показателями вариации состоит также и в том, что она разлагается на составные компоненты, позволяя тем самым оценивать влияние различных факторов на величину учитываемого признака.

Установлено, что рассчитываемая по формуле (17) дисперсия оказывается смещенной по отношению к своему генеральному параметру на величину, равную $n/(n-1)$. Чтобы получить несмещенную дисперсию, нужно в формулу (1) ввести в качестве множителя поправку на смещенность, называемую поправкой Бесселя. В результате формула (17) преобразуется следующим образом:

$$\sigma^2_x = \frac{\sum_{i=1}^k f_1 (x_1 - \bar{x}_2)^2}{n} \cdot \frac{n}{n-1} = \frac{\sum_{i=1}^k f_1 (x_1 - \bar{x}_2)^2}{n-1} \quad (18)$$

Разность $n-1$, обозначаемую в дальнейшем строчной буквой латинского алфавита k , называют **числом, степеней свободы**, под которым понимают число свободно варьирующих единиц в составе численно ограниченной статистической совокупности.

Так, если совокупность состоит из n -го числа членов и характеризуется средней величиной \bar{x} , то любой член этой совокупности может иметь какое угодно значение, не изменяя при этом среднюю \bar{x} , кроме одной, варианты, значение которой определяется разностью между суммой значений всех остальных вариантов и величиной $n\bar{x}$. Следовательно, одна варианта численно ограниченной статистической совокупности не имеет свободы вариации. Отсюда число степеней свободы для такой совокупности будет равно ее объему n без единицы. А при наличии не одного, а нескольких ограничений свободы вариации число степеней свободы вариации будет равно $k = n - v$, где v – (греческая буква ню) обозначает число ограничений свободы вариации.

Дисперсия обладает рядом важных свойств, из которых необходимо отметить следующие.

1. Если каждую варианту совокупности уменьшить или увеличить на одно и то же постоянное число A то дисперсия не изменится:

$$\sigma^2_x = \frac{1}{n-1} \sum [(x_i - A) - (\bar{x} - A)]^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 \quad (19)$$

2. Если каждую варианту совокупности разделить или умножить на одно и то же постоянное число A , то дисперсия уменьшится или увеличится в A^2 раз.

Среднее квадратическое отклонение σ_x . Наряду с дисперсией важнейшей характеристикой варьирования является среднее квадратическое отклонение – показатель, представляющий корень квадратный из дисперсии:

$$\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \quad (20)$$

Эта величина в ряде случаев оказывается более удобной характеристикой варьирования, чем дисперсия, так как выражается в тех же единицах, что и средняя арифметическая величина.

Дисперсия и среднее квадратическое отклонение наилучшим образом характеризуют не только величину, но и специфику варьирования признаков.

В. Шеппард (1898) установил, что разность между *расчетной* и фактической величиной дисперсии составляет $\frac{1}{2}$ квадрата классового интервала. Следовательно, при вычислении дисперсии по формуле (13) следует вносить поправку Шеппарда, т.е. вычитать эту величину ($1/12 \lambda^2$) из σ_x^2 .

Поправка Шеппарда вносится далеко не всегда. Ее обычно применяют или при высокой точности расчетов, или при наличии большого числа наблюдений ($n > 500$), распределяемых в интервальный вариационный ряд. Для получения обобщающих числовых характеристик небольших и средних по объему ($n < 500$) совокупностей поправку Шеппарда не вносят.

Коэффициент вариации V Дисперсия и среднее квадратическое отклонение применимы и для сравнительной оценки одноименных средних величин. В практике же довольно часто приходится сравнивать изменчивость признаков, выражаемых разными единицами. В таких случаях используют не абсолютные, а относительные показатели вариации. Дисперсия и среднее квадратическое отклонение как величины, выражаемые теми же единицами, что и характеризующий признак, для оценки изменчивости разноименных величин непригодны. Одним из относительных показателей вариации является коэффициент вариации. Этот показатель представляет собой среднее квадратическое отклонение, выраженное в процентах от величины средней арифметической

$$V = \frac{\sigma_x}{\bar{x}} 100\% \quad (21)$$

Кроме того, не вполне корректно сравнивать величины, выраженные разными единицами меры. Именно поэтому в подобных случаях уместно использовать безразмерные значения коэффициентов вариации.

Значение этого показателя V остается более или менее устойчивым и при симметричных распределениях обычно не превышает 50%. При сильно асимметричных рядах распределения коэффициент вариации может достигать 100% и даже выше. Варьирование считается слабым, если не превосходит 10%, средним когда V составляет 11-25%, и значительным при $V > 25\%$.

Нормированное отклонение. Отклонение той или иной варианты от средней арифметической, отнесенное к величине среднего квадратического отклонения, называют *нормированным отклонением*:

$$t = \frac{(x_i - \bar{x})}{\sigma_x} \quad (22)$$

Этот показатель позволяет «измерять» отклонения отдельных вариантов от среднего уровня и сравнивать их для разных признаков.

2.4. Статистические характеристики при альтернативной группировке вариантов

В данной главе уже упоминалось об альтернативной группировке исходных данных, когда одна группа вариантов противопоставляется другой. Так, число мужских особей в популяции может противопоставляться числу женских особей, группа здоровых индивидов – группе больных индивидов и т.д. При альтернативной группировке данных их статистическими характеристиками будут как абсолютные, так и относительные численности противопоставляемых друг другу групп (альтернатив). Если абсолютную численность вариант, обладающих данным признаком, обозначить через m , то численность вариант противоположной группы, не имеющих данного признака, будет равна $n-m$, где n – общее число членов рассматриваемой совокупности. Например, среди $n = 408$ новорожденных оказалось $m = 208$ мальчиков. Тогда число новорожденных девочек $n-m = 408 - 208 = 200$.

Численность альтернатив можно выразить в долях единицы, а также в процентах от их общего числа n . Обозначив долю вариант, обладающих учитываемым признаком, через p , получим $p = m/n$. Тогда доля вариант, не обладающих этим признаком, обозначаемая буквой q , выразится – $(n-m)/n = 1 - (m/n) = 1 - p$. Для того чтобы численность противопоставляемых групп была выражена в процентах, достаточно каждую долю умножить на 100:

$$P = (m/n) * 100; \quad (23)$$

$$q = \frac{n-m}{n} * 100\% = 100 - p \quad (24)$$

Очевидно, $p + q = 1$ и $p + q = 100\%$.

Относительные частоты или доли вариант при альтернативной группировке выполняют такую же роль, как средние величины для рядовой изменчивости признаков, когда исходные данные распределяются в вариационный ряд.

В качестве *характеристики альтернативного варьирования* служит среднее квадратическое отклонение σ_p , которое определяют по формуле

$$\sigma_p = \sqrt{p * (1 - p)} = \sqrt{p * q} \quad (25)$$

Этот показатель одинаково характеризует варьирование обеих альтернативных групп. Если же численность групп (альтернатив) выражена в процентах от их общего числа n , формула (33) принимает следующее выражение:

$$\sigma_p = \sqrt{p * (100 - p)} \quad (26)$$

Когда альтернативы выражены абсолютными числами, среднее квадратическое отклонение определяют по формуле

$$\sigma_p = \sqrt{n * p * q} \quad (27)$$

Пример 1. Из общего числа $n = 408$ новорожденных доля мальчиков составила $p = 208/408 = 0,51$, а доля новорожденных девочек $q = 200/408 = 0,49$.

Среднее квадратическое отклонение долей выразится величиной: 0,5. Если выразить в процентах соотношение между количеством новорожденных мальчиков (51%) и девочек (49%), этот показатель будет равен

$$\sigma_p = \sqrt{51 * (100 - 51)} = \sqrt{51 * 49} = \sqrt{0,2499} = 50,0\%$$

Для абсолютных значений альтернатив среднее квадратическое отклонение выразится величиной

$$\sigma_p = \sqrt{408 * 0,51 * 0,49} = \sqrt{101,96} = 10,1.$$

Вопросы к главе 2.

1. Назвать различные виды средних.
2. Когда применяется средняя арифметическая?
3. Когда применяется средняя гармоническая?
4. Когда применяется средняя квадратическая?
5. Когда применяется средняя кубическая?
6. Когда применяется средняя геометрическая?
7. В чём проявляется абстрактность средних величин?
8. Дать определение взвешенной средней арифметической.
9. Что такое математический вес?
10. Что может выступать в качестве математического веса?
11. Что такое мода, модальный класс?
12. Для каких признаков может быть указана мода?
13. С чем связана полимодальность совокупности?
14. Что такое медиана, медианный класс?
15. Что такое квантили?
16. Назовите основные квантили?
17. Когда применяется квантильный способ представления данных?
18. Что такое лимиты и амплитуда колебаний?
19. В чём недостаток лимитов как показателей степени варьирования?
20. Что такое дисперсия, какова её размерность?
21. Свойства дисперсии.
22. Что такое среднее квадратическое отклонение, какова его размерность и знак?
23. Что такое число степеней свободы?
24. Какие оценки являются смещёнными?
25. Какие оценки являются состоятельными?
26. Почему смещённая оценка дисперсии является в то же время состоятельной?
27. Почему сравнение variability признаков по их средним квадратическим отклонениям затруднительно?
28. Что такое коэффициент вариации?
29. В каких случаях коэффициент вариации применять нельзя?
30. Что такое нормированное отклонение и чем оно нормировано?
31. Как рассчитывается среднее квадратическое отклонение для альтернативных признаков?
32. Какова особенность стандартного отклонения для альтернативных признаков?

Глава 3. ЗАКОНЫ РАСПРЕДЕЛЕНИЯ

3.1. Вероятность события и ее свойства

Случайное событие можно предсказать лишь с некоторой уверенностью или вероятностью, которую данное событие имеет. При этом *вероятность* рассматривают как числовую меру объективной возможности осуществления события A при единичном испытании и обозначают символом $P(A)$. Согласно классическому определению, вероятность события A выражается отношением числа благоприятствующих осуществлению этого события исходов m к числу всех равновозможных и несовместных исходов n , т.е.

$$P(A) = \frac{m}{n} \quad (28)$$

Например, в урне находится 5 белых и 10 черных шаров. Наугад вынимают один шар. Какова вероятность, что вынутый шар окажется белым? Так как из общего числа 15 шаров в урне 5 белых, то из $n = 15$ возможных исходов лишь $m = 5$ «благоприятствуют» осуществлению ожидаемого события, т.е. появлению в однократном испытании белого шара. Отсюда вероятность этого события $P = 5/15 = 1/3 \approx 0,33$, т.е. чем больше шансов, благоприятствующих наступлению ожидаемого события, тем выше его вероятность.

Из «классического» определения вероятности следует, что она представляет *число, заключенное между нулем и единицей*, т.е. выражается в долях единицы (может быть выражена в процентах от общего числа испытаний). Очевидно, вероятность достоверного события A равна единице, а вероятность невозможного события A_1 равна нулю. Из этих аксиоматических свойств вероятности следует, что вероятность события A и вероятность противоположного события A_1 (не A) в сумме равны единице, т.е. $P(A) + P(A_1) = 1$.

Считают, что события, имеющие очень малую вероятность, в единичных испытаниях не произойдут, т.е. такие события рассматривают как *практически невозможные*. Если же вероятность события достаточно велика, его принято считать *практически достоверным*. Этот принцип практической уверенности в прогнозировании исходов случайных событий позволяет использовать теорию вероятностей в практических целях.

Для упрощения символики принято значение вероятности ожидаемого события обозначать строчной латинской буквой p , т.е. тем же знаком, которым обозначается частость, а значение вероятности противоположного события – буквой q , т.е. $P(A) = p$ и $P(A) = q$, откуда $p + q = 1$.

Вероятность, которую можно указать до опыта, называют *априорной*. Например, при подбрасывании монеты заранее известно, что она может упасть вверх гербом или решкой. Здесь только два возможных исхода, вероятность которых одна и та же: $p = 1/2 = 0,5$. Иное дело, например, испытание действия на Организм различных доз лекарственных или токсических веществ. В таких случаях результаты отдельных испытаний заранее, до опыта, указать невозможно; вероятность осуществления таких событий может быть установлена только на основании опыта, т.е. *апостериорно*.

Встречаются, однако, и такие события (и их немало), исход которых, как правило, отклоняется от их вероятности. Ярким примером такого рода собы-

тий служит соотношение полов в потомстве многих животных и человека. Известно, что пол потомства определяется в момент оплодотворения, когда в зиготу привносятся либо XX-, либо XY-хромосомы. Вероятность появления в потомстве мужских и женских особей одна и та же: $p = 1/2$. Это означает, что на 1000 новорожденных детей следует ожидать примерно равное число мальчиков и девочек. В действительности же равновеликого соотношения полов в потомстве не наблюдается. Так, по данным шведской статистики за 1935 г. на каждую тысячу новорожденных число родившихся девочек составляло на протяжении года от 473 до 491 при средней частоте равной 482; относительная частота или частость новорожденных девочек $482/1000 = 0,482$, а частота рождения мальчиков – $(1000 - 482)/1000 = 0,518$. Можно также привести данные австрийской статистики о новорожденных детях за период с 1866 по 1905 г., т.е. за 40 лет. Доля новорожденных мальчиков за этот период колебалась от 0,512 до 0,518 и составляла в среднем 0,515.

3.2. Характерные черты варьирования

В любом более или менее симметричном вариационном ряду заметна одна характерная особенность - накапливание вариантов в центральных классах и постепенное убывание их численности по мере удаления от центра ряда. Эта особенность варьирования количественных признаков встречается довольно часто. Так, например, среди взрослого населения чаще встречаются люди среднего роста, а индивиды очень высокого или очень низкого роста – значительно реже. Однако в массе одновозрастных индивидов людей выше и ниже среднего роста оказывается примерно одинаковое количество.

Пусть большая масса людей одного пола и возраста разделена на отдельные группы так, чтобы в каждую группу вошли индивиды приблизительно одинакового роста. Пусть затем от каждой группы выделено по одному представителю, которые построены в одну шеренгу по ранжиру – от самого низкорослого до самого высокорослого. Если в затылок им поставить членов группы, получается «живая диаграмма» распределения, более или менее симметричная. На рис. 1 проиллюстрирована эта закономерность. Отмеченная черта варьирования обнаруживается не только в распределении людей по росту, но и по многим другим признакам, в частности по размерам обуви (рис. 2).

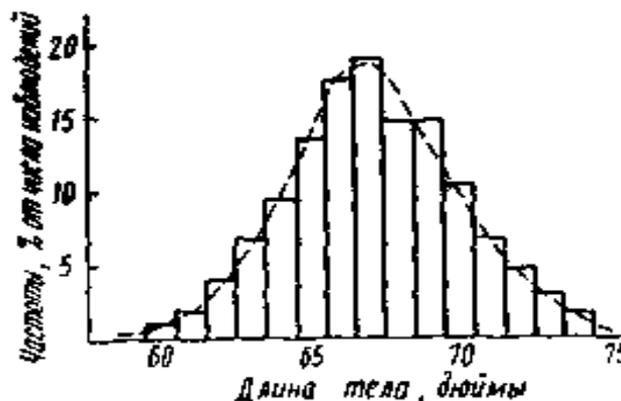


Рис. 1. Гистограмма изменчивости длины тела у 117 мужчин

(по Н. Бейли, 1959)

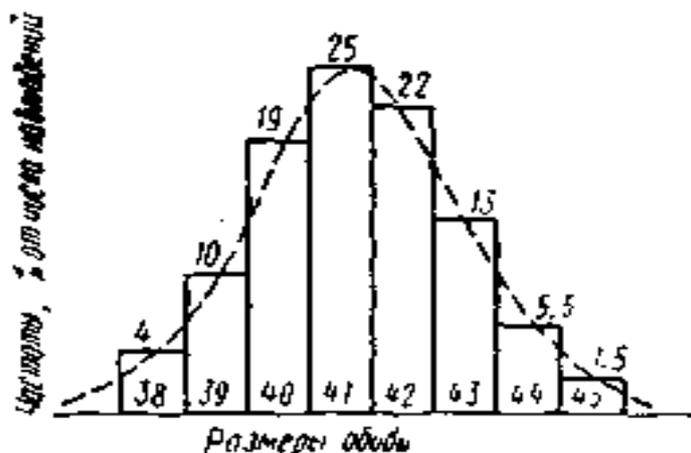


Рис. 2. Гистограмма распределения размеров мужской обуви среди населения центральных областей РСФСР

Впервые на эту закономерность варьирования обратил внимание А. Кетле (1835), исследовавший распределение нескольких тысяч американских солдат по росту (длине тела). «Человеческий рост, - писал он, - изменяющийся, по-видимому, самым случайным образом, тем не менее подчиняется самым точным законам; и эта особенность свойственна не только росту; она проявляется также и в весе, силе, быстроте передвижении человека, во всех его физических и нравственных способностях. Этот великий принцип разнообразия проявление человеческих способностей кажется нам одним из самых удивительных законов мира».

Описанная закономерность относится не только к человеку. Выше рассмотрены варьирование глазков в клубнях картофеля, Подобных примеров можно привести много. Особенно примечательно, что не только распределение живых существ и продуктов их жизнедеятельности, но и случайные ошибки измерений подчиняются этой закономерности. «Не удивительно ли, - писал А. Кетле, - что случайные ошибки располагаются в таком совершенном порядке, и наши бессознательные промахи Проявляются с такой симметрией, которая, кажется, могла бы быть результатом тщательно обдуманных расчетов».

Таким образом, прослеживается широко распространенная в природе закономерность: в массе относительно однородных единиц, составляющих статистическую совокупность, большинство членов оказывается среднего или близкого к нему размера, в чем дальше они отстоят от среднего уровня варьирующего признака, тем реже встречаются в данной совокупности. И это независимо от формы распределения, что указывает на определенную связь между числовыми значениями варьирующих признаков и частотой их встречаемости в данной совокупности. Наглядным выражением этой связи и служат *вариационный ряд* и его линейный график – *вариационная кривая*. Эту закономерность можно воссоздать априори в виде математической модели, не опасаясь впасть в противоречие с фактами.

3.3. Случайные события

Подброшенный камень падает вниз, брошенный в воду – тонет. По длине одной из сторон куба можно точно определить объем. На языке теории вероятностей всякий результат, или исход, однократного испытания называется *событием*.

ем. Под испытанием, которое может повторяться бесконечно большое число раз, подразумевают комплекс условий, необходимых для того, чтобы тот или иной исход мог осуществиться. В таких случаях о вероятности событий принято судить по предельным значениям частоты, обладающей устойчивостью. Отсюда в отличие от «классической» вероятности частоты случайных событий, а точнее их предельные значения, обладающие определенной устойчивостью, принято называть *статистической вероятностью* этих событий. Так что число 0,482 или округленно 0,48, возле которого колеблются значения частоты рождения девочек, можно принять за вероятность этого события. В то же время статистическая вероятность появления в потомстве мальчиков будет равной 0,52.

3.4. Закон больших чисел

Многочисленные опыты и наблюдения показали, что частоты ожидаемых случайных событий приближаются к их вероятности по мере увеличения числа испытаний n . Так, если одну и ту же монету подбрасывать большое число раз, то невозможно ожидать, чтобы во всех без исключения случаях выпадал только герб или только решка. Ясно, что в каком-то числе случаев выпадет герб, а в других случаях – решка. Примечательно, что чем больше число испытаний, тем ближе к единице оказывается отношение выпавших гербов и решек, а частота каждого события становится ближе к его вероятности. Подтверждением тому служат результаты опытов с метанием монет, проведенные разными лицами.

Опыт показал, что с увеличением числа испытаний отклонение частоты ожидаемого результата от его вероятности ($p = 0,5$) уменьшается. В этом факте проявляется действие *закона больших чисел*, теоретическое обоснование которым, было дано Я. Бернулли (1713), а также П.Л. Чебышевым и другими математиками XIX столетия. Этот закон утверждает, что частота m/n события A будет сколь угодно близкой к его вероятности p , если число испытаний неограниченно возрастает. Как было показано выше, частота события и его вероятность не совпадают. Разница между ними уменьшается при увеличении числа испытаний. Можно взять сколь угодно малое число ε и сравнить его с разницей между частотой и вероятностью события. Вероятность того, что эта разница превысит число ε , будет стремиться к нулю при стремлении числа испытаний n к бесконечности, т.е.

$$P = \left\{ \left| \frac{m}{n} - p \right| > \varepsilon \right\}_{n \rightarrow \infty} \rightarrow 0 \quad (29)$$

Этот вывод подтверждается и опытом Кетле: в урну помещали 20 белых и 20 черных шаров, затем извлекали из нее наугад один шар, регистрировали его и возвращали обратно. Каждое испытание повторяли многократно. Вероятность появления белого или черного шара оставалась при этом постоянной, равной 1/2. Результаты опыта Кетле показали, что с увеличением числа испытаний соотношение белых и черных шаров приближается к единице.

Закон больших чисел, как и другие статистические законы, о которых речь пойдет ниже, имеет объективный характер: их действие не зависит от сознания и воли людей. В качестве примера, иллюстрирующего действие за-

кона больших чисел, можно рассмотреть русскую почтовую статистику за период с 1906 по 1910 г., приведенную А.А. Кауфманом в 1916 г.

Письмо без адреса или без указания места назначения, опущенное в почтовый ящик, явление случайное. А между тем, как явствует из данных табл. 21, число таких случаев из года в год оставалось относительно постоянным. Такие факты были известны еще П. Лапласу. Затем А. Кетле и другие статистики собрали большой материал, убедительно свидетельствующий о наличии внутренней связи между случайностью и закономерностью, существующей в сфере массовых явлений.

Расчет теоретических частот можно упростить, применяя табл. Приложений, в которой содержатся значения вероятности $P(m)$ для каждого класса испытаний t и средней величины $a = x$. Чтобы получить теоретические частоты', достаточно значения вероятности $P(m)$, приведенные в табл. Приложений для t и x (вместо $a = np$), умножить на общее число наблюдений n . Так, для $x = 1,5$ и $m = 0$ в табл. Приложений находим $P(m) = 0,2231$. Умножая эту величину на n , равное 517, получаем $f_0' = 115,34$. Затем для $m = 1$ и $x = 1,5$ в той же таблице находим $P(m) = 0,3347$ и $f_1' = 517 \cdot 0,3347 = 173,04$ и так поступаем до конца ряда, как это показано в табл. 27.

При сравнении (визуальном) эмпирических частот с частотами, вычисленными по закону Пуассона, видно, что они согласуются между собой.

3.5. Нормальное распределение

Случайные величины. Как было показано выше, варьирующие признаки в математике рассматривают как переменные случайные величины, способные в одних и тех же условиях испытания принимать различные числовые значения, которые заранее невозможно предсказать. Случайные величины делят на дискретные и непрерывные. Случайная величина называется *дискретной*, если она может принимать только определенные фиксированные значения, которые обычно выражаются целыми числами. Если же случайная величина способна принимать любые числовые значения, она называется *непрерывной*. Очевидно, что счетные признаки относятся к дискретным случайным величинам, тогда как признаки мерные, варьирующие непрерывно, являются величинами непрерывными.

Случайная величина X в серии независимых повторных испытаний может принимать самые различные значения, но в каждом отдельном испытании она принимает единственное из возможных значений.

Закон распределения случайных величин. Функция $f(x)$, связывающая значения x_i переменной случайной величины x с их вероятностями p_i , называется *законом распределения этой величины*. Закон распределения случайной величины можно задать таблично, выразить графически в виде кривой вероятности и описать соответствующей формулой. Закон распределения дискретной случайной величины может, например, выражаться в виде биномиальной кривой и описываться формулой Бернулли, которая позволяет находить вероятные значения этой величины в серии независимых испытаний. В отношении же непрерывной случайной величины речь может идти лишь о тех значениях, которые она способна принять с той или иной вероятностью в интервале от и до. Этот интервал может быть каким угодно: и

большим, и малым. Выдающиеся математики – А. Муавр (1733), И.Г. Ламберт (1765), П. Лаплас (1795) и К. Гаусе (1821) – установили, что очень часто вероятность P любого значения x непрерывно распределяющейся случайной величины x находится в интервале от x до $x + dx$ и выражается формулой

$$P(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}} dx \quad (30)$$

где dx – малая величина, определяющая ширину интервала; π и e – математические константы (π – отношение длины окружности к ее диаметру, равное 3,1416...; $e = 2,7183$ – основание натуральных логарифмов); σ – стандартное отклонение, характеризующее степень рассеяния значений x случайной величины X вокруг средней μ , называемой *математическим ожиданием*. В показателе степени числа e входит *нормированное отклонение* величина, играющая важную роль в исследовании свойств нормального распределения, описываемого формулой (30).

Как видно из этой формулы, закон нормального распределения (нормальный закон) выражает функциональную зависимость между вероятностью $P(X)$ и нормированным отклонением. Он утверждает, что вероятность отклонения любой варианты X от Центра распределения μ , где $x - \mu = 0$, определяется функцией нормированного отклонения I . Графически эта функция выражается в виде кривой вероятности, называемой *нормальной кривой*. Форма и положение этой кривой определяются только двумя параметрами: μ и σ . При изменении величины μ форма нормальной кривой не меняется, лишь график ее смещается вправо или влево. Изменение же величины σ влечет за собой изменение только ширины кривой: при уменьшении, а кривая делается более узкой за счет меньшего рассеяния вариант вокруг средней, а при увеличении σ кривая расширяется. Во всех случаях, однако, нормальная кривая остается строго симметричной относительно центра распределения, сохраняя правильную колоколообразную форму (рис. 3).

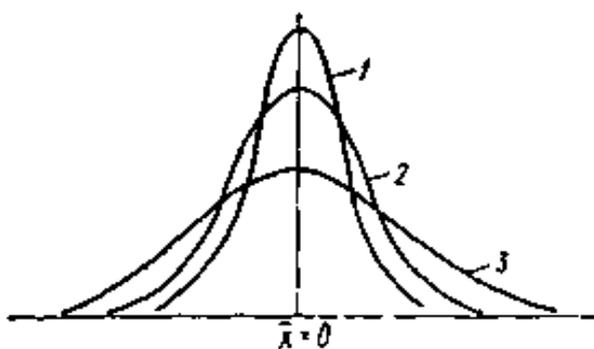


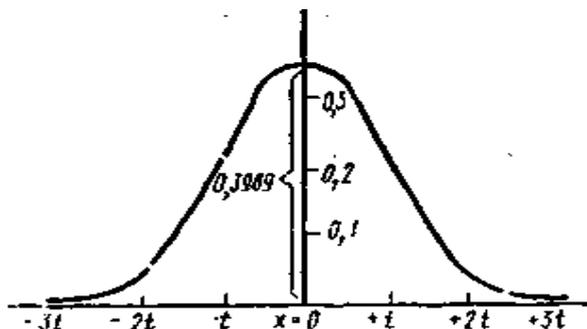
Рис. 3. Нормальные кривые (1, 2, 3) параметра σ ($\sigma_1 < \sigma_2 < \sigma_3$)

Нормальная кривая с параметрами $\mu = 0$ и $\sigma = 1$ называется *нормальной* или *стандартизованной кривой*. Она описывается формулой

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \quad (31)$$

Любую нормальную кривую можно привести к стандартной (вычитанием μ из X и делением на σ). Стандартная кривая (рис. 4) имеет площадь, равную

единице. Ее вершина, т.е. максимальная ордината y_{\max} , соответствует началу прямоугольных координат, перенесенному в центр распределения, где $x_i - \mu = 0$. Вправо и влево от этого центра случайная величина X может принимать любые



значения, и величина каждого отклонения ($x_i - \mu$) определяется функцией его нормированного отклонения $t(f)$. Вероятности P таких отклонений, соответствующие разным значениям, приведены в таблице 2 Приложений.

Рис. 4. Стандартизированная форма нормальной кривой (при $\sigma = 1$)

Для того чтобы ордината выражала не вероятности, а абсолютные числовые значения случайной величины, т.е. выравнивающие частоты вариант эмпирического распределения, нужно в правую часть формулы (31) внести дополнительные множители: в числитель – общее число наблюдений n , умноженное на величину классового интервала, а в знаменатель – величину среднего квадратического отклонения эмпирического ряда распределения σ_x . В результате можно записать формулу

$$f' = \frac{n\lambda}{\sigma_x} f(t) \quad (32)$$

Здесь f' – теоретические (выравнивающие) частоты вариационного ряда, а $f(t)$ функции нормированного отклонения, рассчитанные по формуле (32). Эти значения содержатся в таблице 3 Приложений. Применяя таблицы 2 и 3. Приложений, можно по двум показателям (средней арифметической x и среднему квадратическому отклонению) вычислить теоретические частоты эмпирического вариационного ряда, рассчитать ординаты и построить график нормальной кривой. Сравнивая частоты эмпирического вариационного ряда с частотами, вычисленными по формуле (32), можно проверить, следует ли эмпирическое распределение нормальному закону.

3.6. Параметры нормального распределения.

Основные свойства нормального распределения

Как было показано, нормальное распределение характеризуется двумя параметрами: средней величиной или математическим ожиданием μ , и дисперсией σ_x^2 случайной величины X . Первый параметр равен сумме произведений отдельных значений x , случайной величины X на их вероятности p_i , т.е.

$$\mu = x_1 p_1 + x_2 p_2 + x_3 p_3 + \dots + x_n p_n = \sum_{i=1}^n x_i p_i$$

Второй параметр равен сумме квадратов отклонений отдельных значений x_i случайной величины X от ее математического ожидания μ , т.е.

$$\sigma_x^2 = \sum ([x_i - \mu(x)]^2 f_i) \quad (33)$$

Формально математическое ожидание соответствует средней величине эмпирического распределения, однако, по существу, эти показатели отождествлять нельзя. Среднюю величину определяют как сумму всех членов ряда, отнесенную к их общему числу, а математическое ожидание представляет собой сумму произведений членов ряда на их вероятности. Эмпирическая средняя стремится к математическому ожиданию случайной величины по мере увеличения числа испытаний; при небольшом числе испытаний средняя может значительно отклоняться от своего математического ожидания.

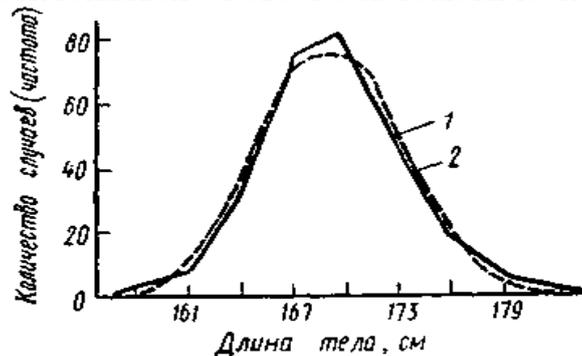


Рис. 5. Эмпирическая (1) и вычисленная по нормальному закону (2) кривые распределения длины тела у 267 мужчин

Для нормального распределения характерно совпадение по абсолютной величине средней арифметической, медианы и моды. Равенство между этими показателями указывает на *нормальность данного распределения*. Вероятность отклонения любой варианты в ту или другую сторону от средней μ , на как это видно из таблицы 2 Приложений, следующая:

$$P \{ -t < |x - \mu| < +t \} = 0,6827$$

$$P \{ -2t < |x - \mu| < +2t \} = 0,9545$$

$$P \{ -3t < |x - \mu| < +3t \} = 0,9973$$

Это означает, что при распределении совокупности наблюдений по нормальному закону из 10000 вариант в интервале от $\mu - t$ до $\mu + t$ окажется 6827 вариант, или 68,3% от общего числа вариантов, составляющих данную совокупность. В интервале от $\mu - 2t$ до $\mu + 2t$ будет находиться 9545 вариант, или 95,4% от числа всех вариантов совокупности. И в интервале от $\mu - 3t$ до $\mu + 3t$ окажется 9973, или 99,7% от общего объема совокупности. Следовательно, с вероятностью $P = 0,6827$ можно утверждать, что наугад отобранная из нормально распределяющейся совокупности варианта не выйдет за пределы от $\mu - t$ до $\mu + t$, или в компактной форме $\mu \pm t$. Вероятность того, что случайно отобранная варианта не отклонится от средней μ , более чем на $\mu \pm 3t$ равна $P = 0,9973$. Это означает, что 99,7% от всех вариантов нормально распределяющейся совокупности находится в пределах $\mu \pm 3\sigma$. Этот важный вывод известен в биометрии как правило *плюс-минус трех сигм*.

3.7. Интеграл вероятности нормального распределения

Учитывая симметрию кривой нормального распределения, можно заключить, что при этом распределении вероятности значений, превышающих

среднее и меньших него, одинаковы и равны 0,5. Поскольку графически выражением вероятности служит площадь под кривой распределения, это проявляется в равенстве площадей под кривой справа и слева от наибольшей ординаты, соответствующей значению μ .

Вероятность того, что нормально распределенная случайная величина X с параметрами μ и σ^2 примет значение в интервале от x_1 до x_2 ($x_1 < x_2$), можно найти графически как долю площади под кривой распределения в соответствующем интервале значений x от общей площади под кривой распределения. Однако более точно у го можно сделать, вычислив интеграл вероятности, который для нормального закона имеет вид

$$\varphi[x_1, x_2] = \varphi(x_2) - \varphi(x_1) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad (34)$$

Предпочтительнее находить интеграл вероятности не вычислением по формуле (34), а с помощью заранее составленных таблиц. Однако ясно, что табулировать интегралы вероятности для различных переменных с разными средними и дисперсиями совершенно нереально, да в этом и нет надобности, поскольку всегда имеется простая возможность от распределения конкретной величины X перейти к распределению величины Z , с параметрами 0 и 1, и наоборот.

Существует несколько подходов к табулированию интегралов вероятности нормального распределения. Наиболее часто в таблицу помещают значение $\Phi(z)$ для интервалов значений переменной от 0 до z или от $-z$ до z .

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{z^2}{2}} dz \quad (35)$$

Функция распределения $F(z)$, как следует из формулы (35), показывает долю, которая приходится на значения переменных, меньших z , от общего объема совокупности.

Так как $P(-\infty < z < +\infty) = 1$, то, зная $F(z_1) = P(z < z_1)$, не трудно найти вероятность $p(z_1 < z)$:

$$P(z < z) = 1 - F(z_1) \quad (36)$$

Зная функцию распределения для значений z_1 и z_2 . ($z_1 < z_2$), легко найти вероятность того, что случайная величина примет значение в интервале от z_1 до z_2 .

$$P(z_1 < z < z_2) = F(z_2) - F(z_1) \quad (37)$$

Так, для $z_1 = -1,75$ и $z_2 = 0,38$ значения $F(z)$ соответственно равны 0,04006 и 0,64803; поэтому вероятность того, что случайно полученное значение z будет находиться в интервале от $-1,75$ до $0,38$, равна $0,64803 - 0,04006 = 0,60797$. В этом случае искомая вероятность графически представляет собой разность между площадями под кривой распределения, ограниченной справа ординатами, соответствующими значениям z_1 и z_2 (рис. б).

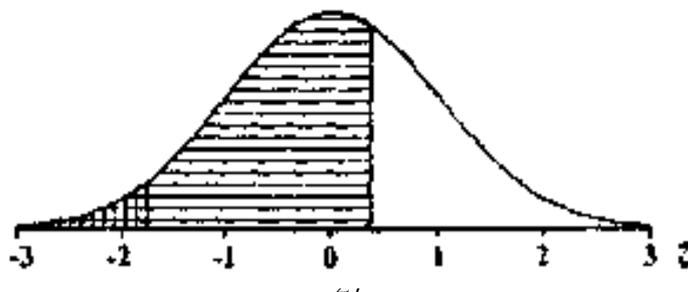


Рис. 6. Графическое выражение вероятности как разности интегралов (площади, соответствующие $F(z)$ заштрихованы вертикально, для $z_1 = -1,75$ и горизонтально для $z_2 = 0,38$)

Если z_1 и z_2 имеют одинаковую абсолютную величину и отличаются только знаками (то в силу симметрии нормального распределения имеем $F(z_1) = 1 - F(z_2) = P(z < z_1) = P(z > z_2)$), а вероятность того, что случайная величина примет значение в интервале от z_1 до z_2 , составит

$$P(z_1 < z < z_2) = F(z_2) - F(z_1) = 1 - 2F(z_1) \quad (38)$$

Интервал, ограниченный значениями, равноотстоящими по обе стороны от среднего, при нормальном распределении соответствует середине ряда распределения, т.е. наиболее часто встречающимся значениям переменной. При этом отклонения в большую или меньшую сторону от указанного интервала равновероятны. Все это обуславливает особый интерес к интервалам значений от $-z$ до z . Можно показать, что значения нормально распределенной величины при $P = 0,50$ заключены в пределах от $-0,674$ до $0,674$. Вероятности $0,90$ соответствует интервал от $-1,64$ до $1,64$, а 95% объема совокупности не выходит за пределы значений $z = -1,96$ и $z = 1,96$. Вероятностям $0,99$ и $0,999$ соответствуют интервалы, ограниченные значениями $z = \pm 2,58$ и $z = \pm 3,29$.

Приведенные данные наглядно показывают, как с увеличением абсолютной величины z растет вероятность, соответствующая выделяемому интервалу значений. Из них также видно, что при нормальном распределении большая часть значений переменной не слишком сильно отклоняется от среднего (если отклонения рассматриваются как стандартные).

Учитывая, что для нормально распределенной величины X с параметрами μ и σ^2 нормированные отклонения распределены как величина Z , можно найти не только вероятность значений в заданном интервале от x_1 до x_2 , но и интервал значений $x_1 < x < x_2$, соответствующий заданной вероятности. В первом случае для значений находят нормированные отклонения u , которые, будучи нормально распределенными, представляют собой величину z , а затем с помощью $F(z)$ определяют искомые вероятности; во втором случае для заданной вероятности P по таблице отыскивают соответствующие значения $\pm z$ и по формуле $x = \mu + z\sigma$ находят искомые значения x_1 и x_2 .

3.8. Логнормальное распределение

Если закону нормального распределения подчиняется не сама случайная величина X , а ее логарифм $l = \log_a x$ (a – основание логарифма), то величина X называется логарифмически нормально (логнормально) распределенной. Логнормальный закон описывается обычным уравнением Гаусса-Лапласа:

$$\omega(x) = \frac{1}{x\sigma_l\sqrt{2\pi}} e^{-\frac{(l-\mu_l)^2}{2\sigma_l^2}} \quad (39)$$

где μ_l и σ_l^2 – параметры нормально распределенной величины L , т.е. $\mu_l = E(l) = E(\log_a x)$ – средний логарифм, а $\sigma_l^2 = E[(l - \mu_l)^2]$ – средний квадрат отклонения логарифма значений от среднего логарифма (дисперсия ряда логарифмов). При логнормальном законе распределения величина X может принимать значения от 0 до $+\infty$. Заметим, что антилогарифм величины μ_l , т.е. $g_x = a^{\mu_l}$ называется средним

геометрическим величины X , которое отличается от математического ожидания (среднего арифметического) $\mu_x = E(x)$ случайной величины X причем $g_x < \mu_x$. Соответственно антилогарифм величины σ^2 не равен дисперсии σ_x^2 величины x .

3.9. Биномиальное распределение

Биномиальное распределение возникает в том случае, когда каждое испытание имеет два возможных несовместимых друг с другом исхода, как например, в случае, когда каждое из зерен песчаной фракции может оказаться кварцевым или не кварцевым, а тот или иной агрегат – водопрочным или нет. Если в условиях проведения испытаний вероятность наличия признака (например "кварца") обозначить через P , а "его отсутствия – через Q (при этом $P + Q = 1$), то вероятность $P_n(x)$ в n испытаниях получить x раз ($x = 0, 1, 2, \dots, n$) наличие признака определяется выражением

$$P_n(x) = \frac{n!}{x!(n-x)!} P^x Q^{n-x} \quad (40)$$

где знак «!» показывается факториалом и означает, что следует перемножить целые числа от 1 до числа, стоящего под знаком факториала.

Так, $n! = 1 * 2 * 3 * \dots * (n-1) * n$. Заметим, что факториал нуля считается равным единице ($0! = 1$).

При биномиальном распределении в качестве случайной величины выступает число случаев с заданным исходом при фиксированном объеме выборки n . Единственным параметром этого распределения является P (поскольку $Q = 1 - P$), который может принимать значения в интервале от 0 до 1. Для выборки объема n дисперсия биномиального распределения находится по формуле:

$$\sigma^2 = nPQ \quad (41)$$

из которой следует, что дисперсия не может превосходить $0,25n$. Это наибольшее значение дисперсия принимает при $P = Q = 0,5$. Математическое ожидание случайной величины в биномиальном распределении представляет собой среднее число появлений признака в выборках объема n .

$$\mu = E(x) = nP \quad (42)$$

При больших значениях n для нахождения $P_n(x)$ прибегают к логарифмированию выражения (40):

$$\lg P_n = (\lg n)! - \lg(n-x)! + x \lg P + (n-x) \lg Q \quad (43)$$

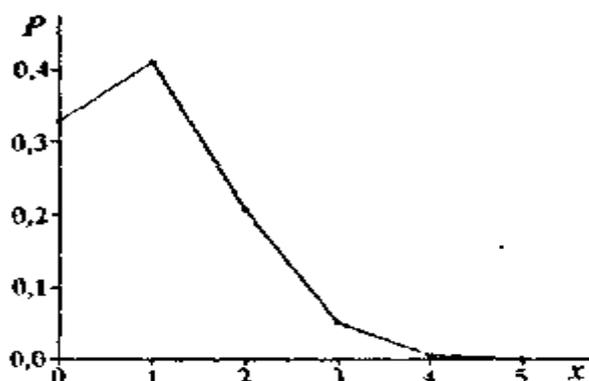


Рис. 7. Распределение числа x водопрочных агрегатов в выборке объема $n = 5$ при биномиальном распределении (водопрочных агрегатов 20%)

Поскольку логарифмы факториалов можно найти в справочниках по математике, нахождение $\lg P_n(x)$, а затем и искомой вероятности $P_n(x)$ не составляет большого труда.

Биномиальное распределение симметрично, когда $P = Q = 0,5$. При $P \neq Q$ распределение несимметрично и имеет одну моду, причем с увеличением x на единицу $P_n(x+1)$ возрастает сравнительно с $P_n(x)$ до тех пор, пока выполняется условие $x < P_n - Q$.

При больших n максимум $P_n(x)$ расположен в окрестности значения $x = Pn$, причем при $n \rightarrow \infty$ и любых x величина $P_n(x)$ также стремится к нулю.

Симметричность биномиального распределения зависит не только от величины P , но и от объема выборки n . Действительно, так как при этом распределении центральный момент третьего порядка определяется выражением

$$\mu_{(3)} = nPQ(Q - P) \quad (44)$$

а $\sigma^2 = nPQ$, то согласно формуле, получим, что коэффициент асимметрии

$$\alpha = \frac{Q - P}{\sqrt{nPQ}} \quad (45)$$

Центральный момент четвертой степени при биномиальном распределении определяется по формуле

$$\mu_{(4)} = nPQ[3 \cdot PQ(n - 2) + 1] \quad (46)$$

а коэффициент эксцесса составляет

$$\varepsilon = \frac{\frac{1}{PQ} - 6}{n} = \frac{1 - 6P + 6P^2}{nPQ} \quad (47)$$

В данном примере имеем $e = 0,05$.

Коэффициент вариации наряду с обычным способом его вычисления можно найти из выражения:

$$V = 100 \sqrt{\frac{Q}{nP}} \% \quad (48)$$

По мере увеличения n биномиальное распределение приближаем к нормальному и тем быстрее, чем меньше отличаются друг от друга P и Q .

3.10. Распределение Пуассона

Распределение для описания случайных событий, происходящих в течение того или иного интервала времени (например, число α -частиц, излучаемых почвой в единицу времени), либо в пределах некоторого пространства (на плоскости или в объеме почвы, как, например, число ходов червей на некотором участке поверхности почвы, число ортштейнов в заданном в объеме, число зерен циркона на шлифе и т.д.). Пуассоновское распределение описывается выражением

$$P(x) = \frac{\mu^x}{x!} e^{-\mu} \quad (49)$$

где $P(x)$ – вероятность в единичном испытании получить x раз ($x = 0, 1, 2, \dots$) интересующее нас событие, μ – среднее число таких событий на одно испытание, e – основание натуральных логарифмов. Пуассоновское распределение можно использовать для аппроксимации биномиального распределения, ко-

гда вероятность P появления изучаемого признака очень мала (например, наличие семян сорняков среди группы из n взятых зерен посевного материала) и тогда $\mu = nP$.

Распределение Пуассона возникает в случае, когда вероятность обнаружения событиями бесконечно малом временном интервале (или области пространства) бесконечно мала, причем для любых двух непересекающихся интервалов времени (или областей пространства) вероятность обнаружения любого числа событий в одном не зависит от числа событий в другом.

Распределение Пуассона характерно тем, что среднее μ и дисперсия σ^2 численно равны и, как следует из формулы (49), это распределение описывается одним параметром.

В распределении Пуассона, как и при биномиальном распределении, случайной величиной является число случаев, а не значение таких свойств, как, например, содержание железа или водопроницаемость. Для этих последних может иметь место равенство средних и дисперсий, но это не должно служить достаточным основанием для аппроксимации их распределений законом Пуассона, который здесь неприменим.

Пуассоновское распределение вероятностей обычно очень асимметрично. При $\mu < 1$ максимальная вероятность $P(x)$ соответствует значению $x = 0$. С увеличением μ распределение становится более симметричным, причем при данном μ возрастание $P(x)$ с увеличением x наблюдается до тех пор, пока $x + 1 < \mu$ (рис. 8).

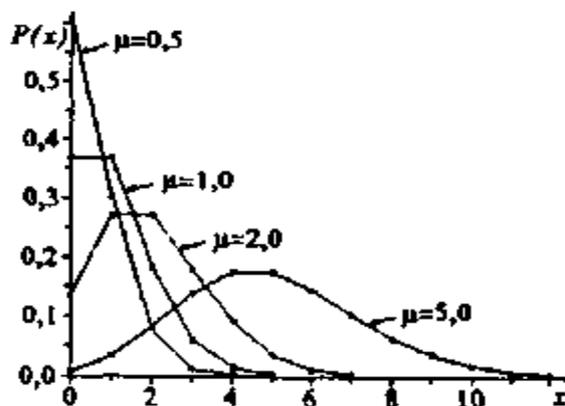


Рис. 8. Пуассоновское распределение вероятностей при различном среднем μ

При пуассоновском распределении равны между собой не только среднее и дисперсия (начальный момент первой степени и центральный момент второй степени), но и центральный момент третьей степени:

$$\mu = \mu_{(2)} = \mu_{(3)}. \quad (50)$$

В связи с этим, коэффициент асимметрии при пуассоновском распределении обратно пропорционален стандарту распределения (или квадратному корню из среднего):

$$\alpha = \frac{1}{\sqrt{\mu}} = \frac{1}{\sigma} \quad (51)$$

Коэффициент эксцесса обратно пропорционален среднему (или дисперсии)

$$\varepsilon = \frac{1}{\mu} = \frac{1}{\sigma^2} \quad (52)$$

Из рассмотрения формул (51) и (52) следует, что для распределения Пуассона асимметрия и эксцессивность положительны и с увеличением среднего уменьшаются.

3.11. Распределение Максвелла

По нормальному закону распределяются многие биологические признаки, но не все: нередко встречаются и асимметричные распределения, которые, однако, не следуют закону Пуассона. Одним из трех распределений является *распределение, описываемое формулой Максвелла*

$$P(X) = \frac{2}{\sqrt{2\pi}} \frac{t^2}{a} e^{-\frac{t^2}{2a}} dx \quad (53)$$

В этой формуле $a = 0,6267$, x – параметр распределения, определяемый через среднюю арифметическую \bar{x} варьирующего привязка; $t = x_i/a$, где x_i – числовые значения случайной величины X ; dx – разность между двумя смежными значениями переменной величины X .

Указанием на то, что эмпирическое распределение следует закону Максвелла, служит равенство между средним квадратическим отклонением и величиной $0,674 a$, т.е. $\sigma_x = 0,674a$, тогда как для распределения Пуассона характерно равенство. Чтобы рассчитать по формуле (34) теоретические (выравнивающие) частоты, нужно проделать следующее: 1. Определить среднюю арифметическую эмпирического вариационного ряда и Параметр a . 2. Разделить каждую классовую варианту x_i на величину a , что даст значения t . 3. Найти для каждого значения $t = x/a$ в таблице 2 Приложений значение функции $f(t)$. 4. Определить значения t^2/a . 5. Умножить значения t^2/a на удвоенную величину t и на величину классового промежутка ($\lambda = dx$), т.е. определить $P = (t^2/a)2f(t)\lambda$. 6. Умножить значения P на общее число наблюдений n , получить теоретические выравнивающие частоты данного вариационного ряда, т.е. $f' = Pn$.

3.12. Измерение асимметрии и эксцесса

Среди эмпирических распределений *асимметрия* и *эксцесс* встречаются довольно часто. Заметить асимметрию и эксцесс можно по характеру распределения частот в классах вариационного ряда. Графически асимметрия выражается в виде скошенной вариационной кривой, вершина которой может находиться левее или правее центра распределения. В первом случае асимметрия называется *правосторонней* или *положительной* а во втором – *левосторонней* или *отрицательной* (по знаку числовой характеристики). При правосторонней асимметрии ее пологая сторона находится правее, при левосторонней – левее центра распределения (рис. 9).

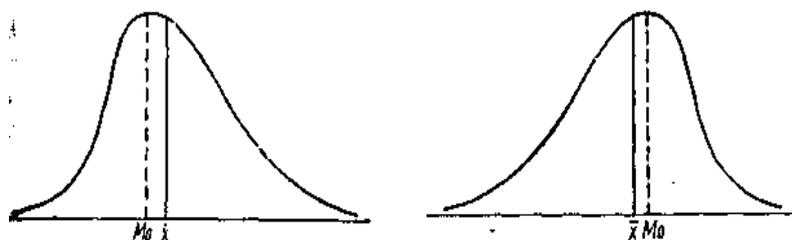


Рис. 9. Асимметричные кривые: справа – положительная асимметрия, слева – отрицательная асимметрия

Наряду с асимметричными, встречаются островершинные и плосковершинные распределения. Островершинность кривой распределения вызывается чрезмерным накапливанием частот в центральных классах вариационного ряда, вследствие чего вершина вариационной кривой оказывается сильно поднятой вверх. В таких случаях говорят о положительном эксцессе распределения (рис. 10.). Кроме одновершинных встречаются и двух- и многовершинные кривые, а также плосковершинные и двугорбые кривые, что свидетельствует о наличии у такого распределения отрицательного эксцесса.



Рис. 10. Эксцесс: слева – крутовершинная кривая – положительный эксцесс, справа – плосковершинная кривая – отрицательный эксцесс

Величина асимметрии и эксцесса может быть различной, поэтому важно ее не только обнаружить, но и измерить. Для измерения асимметрии и эксцесса используют центральные моменты распределения третьего и четвертого порядков. В качестве показателя асимметрии служит центральный момент третьего порядка, отнесенный к кубу среднего квадратического отклонения, т.е.

$$As = \frac{\mu_3}{s_x^3} = \frac{\sum fi(xi - x)^3}{n} / s_x^3 \quad (54)$$

Показатель эксцесса, обозначаемый символом E_x , выражается формулой:

$$E_x = \frac{\mu_4}{s_x^4} - 3 = \left[\frac{\sum fi(xi - x)^4}{n} / s_x^4 \right] - 3 \quad (55)$$

В случае положительного эксцесса этот показатель приобретает положительный знак (+) и может иметь самую различную величину. При плосковершинности и двугорбости вариационной кривой коэффициент имеет отрицательный знак (-); предельная величина отрицательного эксцесса равна минус двум.

Вычисление показателей асимметрии и эксцесса по формулам (54) и (55), т.е. способом произведений непосредственно по центральным моментам распределения, оказывается довольно трудоемким особенно при наличии в выборке многозначных чисел. Поэтому центральные моменты обычно вычисляют, используя стандартные прикладные программы.

3.13. Проверка нормальности распределения

Существует несколько методов проверки нормальности распределения. При исследовании распределения отмечается, что если оно близко к нормальному, то значение медианы находится между значениями моды и сред-

ней величины $Mo > Me > X$. В умеренно асимметричных распределениях должно соблюдаться следующее равенство:

$$Mo - x = 3 * (Me - x) \quad (56)$$

Если коэффициент асимметрии положителен, то асимметрия правосторонняя, если отрицателен – то левосторонняя.

Оценка существенности As производится на основе средней квадратической ошибки коэффициента асимметрии (σ_{As}), которая зависит от числа наблюдений (n) и рассчитывается по формуле

$$\sigma_{As} = \sqrt{\frac{6(n-1)}{(n-1)(n+3)}} \quad (57)$$

Если $As/\sigma_{As} > 3$, то асимметрия существенна, распределение признака в генеральной совокупности несимметрично.

Если $As/\sigma_{As} < 3$, то асимметрия несущественна, распределение признака симметрично.

Значительный отрицательный эксцесс может говорить о качественной неоднородности исследуемой выборки. В литературе нет однозначной интерпретации значений As и Ek . При этом ориентируются на следующие величины. Если:

- $Ek, < 1$, то эксцессивность слабая;
- $5 < Ek > 1$, то эксцессивность средняя;
- $Ek > 5$, то эксцессивность сильная.

Пример 2. Произведён статистический анализ выборки орехов оценённых по общему баллу качества плодов. Проверить нормальность распределения.

Общий балл

Среднее _____	31,74
Медиана _____	31,73
Мода _____	31,55
Стандартное отклонение _____	1,62
Дисперсия выборки _____	2,62
Эксцесс _____	0,80
Асимметричность _____	-0,67
Минимум _____	26,91
Максимум _____	34,98
Счет _____	70

1. В нашем случае значение медианы находится между значениями моды и средней величины: $Mo > Me > X$.

$$31,55 > 31,73 > 31,74$$

$$\sigma_{As} = \sqrt{\frac{6(70-1)}{(70-1)(70+3)}} = \sqrt{\frac{414}{5037}} = 0,29$$

2. $As/\sigma_{As} = -0,67/0,29 = -2,3$, то есть левосторонняя асимметрия несущ-

щественна, распределение признака довольно симметрично.

3. $E_k = 0,8$ – значит эксцессивность слабая.

Учитывая, что все три условия соблюдены, можно сделать вывод о том, что распределение нормально.

Вопросы к главе 3.

1. Каковы особенности нормального распределения?
2. Каковы условия возникновения нормального распределения?
3. Какому значению случайной величины при нормальном распределении соответствует максимальная плотность вероятности?
4. Укажите пределы колебания случайной величины при нормальном распределении.
5. Укажите точки на оси соответствующие при нормальном распределении среднему, моде и медиане.
6. Какому значению соответствует точки перехода выпуклого участка кривой в вогнутый?
7. Что такое интеграл вероятности?
8. Что такое уровень значимости?
9. Что такое логнормальный закон распределения?
10. Что общего и в чём различия биномиального распределения и распределения Пуассона?
11. Распределение какой величины известно как распределение t-Стьюдента?
12. Что такое асимметрия?
13. Когда асимметрия называется положительной и когда отрицательной?
14. Размерность коэффициента асимметрии.
15. Что такое эксцесс?
16. Когда эксцессивность называется положительной и когда отрицательной?
17. Размерность коэффициента эксцесса.
18. Каковы причины возникновения асимметрии и эксцесса?
19. В каких пределах могут принимать значения коэффициенты асимметрии и эксцесса?
20. Каковы придержки, которыми пользуются при определении степени асимметрии и эксцесса?
21. При каком объёме совокупности есть смысл рассчитывать коэффициенты асимметрии и эксцесса и с чем это связано?

Глава 4. ВЫБОРОЧНЫЕ ОЦЕНКИ И ОШИБКИ РЕПРЕЗЕНТАТИВНОСТИ

На практике исследователь имеет дело не с генеральными совокупностями, а с выборками, содержащими информацию о свойствах генеральной совокупности. Так как выборка является частью совокупности, то значения изучаемых параметров не могут являться точными. Они позволяют вычислить лишь *оценки* параметров. Всегда следует учитывать, что выборка позволяет получить оценки с некоторой погрешностью. Так как оценки случайны, невозможно сказать в большую или меньшую сторону отклоняется оценка от истинного значения исследуемого параметра. Случайность оценок не позволяет отождествлять их с константами.

Часто константы характеризуются оценками, выраженными одним числом. Они получили название *точечные*. Точечными оценками являются выборочные значения среднего, дисперсии, среднеквадратичного отклонения, коэффициента вариации и т.д.

4.1. Точечные оценки

Мода – это значение, встречающееся наиболее часто (для дискретных величин), или имеющее наибольшую плотность вероятности (для непрерывных величин). Поэтому оценку моды целесообразно находить, если в выборке не менее 30 единиц наблюдения и можно построить ряд распределения. В этом случае за оценку моды принимают середину модального класса, т.е. класса, имеющего наибольшую частоту. Если ширина класса λ равна точности измерений Δ и середина любого класса является единственным значением случайной величины, то выборочное значение моды приравнивают тому значению, которое встречается наиболее часто.

Пример 2.

x_j	0	1	2	3	4	5
f_j	3	8	13	4	0	2

Здесь $Mo = 2$, если $\lambda > \Delta$, то в качестве Mo принимают середину класса, но есть и более точная оценка

$$Mo = x_{(m-1)/m} + \lambda \frac{f_m - f_{m-1}}{2f_m - (f_{m-1} + f_{m+1})} \quad (58)$$

где индексы m , $m-1$, $m+1$ соответствуют модальному, предмодальному и послемодальному классам, а $x_{(m-1)/m}$ – граница между модальным и предмодальным классами.

Пример 3.

x_j	1,08	1,13	1,18	1,23	1,28	1,33	1,38	1,43	1,48	1,53
f_j	7	19	21	29	26	18	12	9	6	3

Если за оценку принять середину класса с наибольшей частотой, то $Mo = 1,23$. По формуле $Mo = 1,205 + 0,05 \cdot (29 - 21) / (2 \cdot 29 - (21 + 26)) = 1,24$.

Оценку **медианы** можно получить и по малой выборке, при этом данные должны быть ранжированы. Для вычисления выборочной медианы при чётном n :

$$Med = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} \quad (59)$$

При нечётном n :

$$Med = x_{\frac{n+1}{2}} \quad (60)$$

Пример 4.

x_j 7,4 7,8 8,2 8,9 9,8 так как n нечётно, то $Med = 8,2$.

Пример 5.

x_j 1,35 1,75, 1,85 1,9 2,05 2,2 $n = 6$ – чётное и $Med = 0,5*(1,85 + 1,9) = 1,88$.

Если $\lambda > \Delta$, то в качестве Med принимают середину медианного класса, для примера 2 $n = 150$. Середина ранжированного ряда приходится на 75 и 76 места. Соответствующие значения принадлежат 4 классу, для которого $x_4 = 1,23$, поэтому $Med = 1,23$. Но можно при помощи линейной интерполяции получить более точную оценку

$$Med = x_{(m-1)/m} + \lambda \frac{\frac{n}{2} - \sum_{j=1}^{m-1} f_j}{f_m} \quad (61)$$

где m – номер медианного класса, $x_{(m-1)/m}$ – граница между медианным и предмедианным классами, f_j частота медианного класса, $\sum f_j$ сумма частот классов, предшествующих медианному.

В примере 2 $\lambda = 2$, $m = 4$, $x_{3/4} = 1,205$, $\sum f_j = 47$, тогда по формуле $Med = 1,205 + 0,05*((150/2)-47)/29 = 1,25$.

Оценку среднего можно получить разными способами. Самый простой

$$\bar{x} = \frac{x_{\min} + x_{\max}}{2}$$

Эта оценка пригодна для симметричных распределений. Недостаток её в ненадёжности. При симметричности распределения в качестве оценки можно использовать и медиану. Наиболее часто оценку среднего находят как среднее арифметическое всех значений. Эта оценка и будет использоваться в дальнейшем.

$$\bar{x} = \frac{\sum x_i}{n} \quad (62)$$

4.2. Статистические оценки при объединении выборок

Две или более выборок могут принадлежать одной генеральной совокупности. При этом очевидно, что полученные в разных выборках оценки, например среднего или дисперсии оценивают одно и то же среднее, дисперсию. Если выборки действительно получены из одной генеральной совокупности, то выборки можно просуммировать. По такой выборке находятся единые оценки констант.

Если исследователя интересуют среднее и дисперсия, и они уже вычислены для отдельных, не объединённых в одну выборку, то не обязательно возвращаться к исходным данным. Это важно и тогда, когда данные берутся

из литературных источников, и имеются сведения лишь о среднем и дисперсии. Если объём суммарной выборки $n = \sum n_j$, то

$$\bar{x} = \frac{\sum_j n_j \bar{x}_j}{n} \quad (63)$$

Среднее объединённой выборки вычисляется как среднее взвешенное, причём для каждого среднего в качестве веса берут соответствующий объём выборки.

Для объединённой выборки дисперсия находится из выражения

$$\sigma^2 = \frac{\sum_j \sigma_j^2 (n_j - 1) + \sum_j n_j (\bar{x} - \bar{x}_j)^2}{n - 1} \quad (64)$$

Стандартное отклонение объединённой выборки находят как корень из дисперсии, причём число степеней свободы с которыми находятся эти оценки равно сумме объёмов частных выборок минус один ($n - 1$).

4.3. Ошибки репрезентативности

Оценки, получаемые на основе выборочных наблюдений, являются случайными величинами. В связи с этим каждая такая величина имеет некоторый закон распределения и характеризуется некоторыми константами (средним, дисперсией, стандартным отклонением, коэффициентом вариации и т.д.). Существенным при этом является то, что ряд констант, характеризующих распределение оценок, зависит от объёма выборки, по которой вычисляется оценка. Очевидно, чем больше выборка, тем с большей точностью получаются оценки. При равенстве объёма выборки объёму генеральной совокупности, оценка среднего равна истинному среднему, то есть превращается в константу, а дисперсия, соответственно = 0. Аналогичная зависимость дисперсии оценок от объёма выборки наблюдается и для оценок других констант. Так как характеристикой варьирования оценок около своего среднего может быть стандартное отклонение, то применительно к оценкам его называют *ошибкой* соответствующей *оценки*. Так как размер этой ошибки является функцией объёма выборки, соответствующие ошибки получили название *ошибок выборочности*, или *ошибок репрезентативности*.

Наиболее интересна ошибка среднего. Она вычисляется по формуле

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad (65)$$

Здесь имеется ввиду выборочная дисперсия.

Пример 6. Если высота деревьев липы определяется с $n = 100$, и характеризуется $x = 19,8$ м и $\sigma = 0,52$ м, то

$$\sigma_{\bar{x}} = \frac{0,52}{\sqrt{100}} = \frac{0,52}{10} = 0,052 \text{ м}$$

Это означает, что если проводить измерения высоты с $n = 100$ многократно, то оценки средней высоты будут варьировать со средним квадратическим отклонением $= 5,2$ см.

При нормальном распределении случайной величины и больших объемах выборки, $n = 100$ и более, оценка дисперсии имеет стандартное отклонение

$$\sigma_{\sigma^2} = \sigma^2 \sqrt{\frac{2}{n-1}} \quad (66)$$

Для оценки стандартного отклонения ошибку можно найти по формуле

$$\sigma_{\sigma} = \frac{\sigma}{\sqrt{2(n-1)}} \quad (67)$$

Для коэффициента вариации ошибка составляет

$$\sigma_v = v * \sqrt{\frac{0,5 + 0,0001 * v^2}{n}} \quad (68)$$

Ошибку медианы можно вычислить по формуле

$$\sigma_{Med} = c * \frac{\sigma}{\sqrt{n}} \quad (69)$$

где c – некоторый коэффициент, зависящий от n . При $n \rightarrow \infty$ $c \rightarrow \sqrt{\pi/2}$

n	4	6	8	10	12	20	∞
c	1,09	1,14	1,16	1,18	1,19	1,21	1,15

Ошибка выборочной доли в формуле 70

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}} \quad (70)$$

Если применить поправку Пирсона, она примет вид

$$\sigma_p = \sqrt{\frac{pq}{n} \left(1 - \frac{n}{N}\right)} \quad (71)$$

Судить о точности, с которой определено выборочное среднее, позволяет отношение ошибки репрезентативности к своей средней. Этот показатель обозначается Cs и выражается обычно в процентах. Определяют его по одной из формул

$$Cs = \frac{\sigma_x}{x} * 100 \quad (72)$$

$$Cs = \frac{V}{\sqrt{n}} \quad (73)$$

где V – коэффициент вариации, выраженный в процентах, n – объем выборки. Показатель точности применяют часто при сравнительной оценке результатов сельскохозяйственных опытов. Точность средних показателей считают приемлемой, если показатель точности не превышает 3-5 %. Показатель точности имеет ошибку, вычисляемую по формуле

$$\sigma_{Cs} = Cs \sqrt{\frac{1}{2n} + \left(\frac{Cs}{100}\right)^2} \quad (74)$$

Для итога статистического анализа вполне достаточно тех показателей, с которыми мы познакомились. Вычислять, и придавать гласности нужно те показатели, которые необходимы и достаточны для решения тех задач, ради которых проводилось исследование.

Минимум статистических сведений должен включать:

Объём выборки n .

Среднее X .

Стандартное отклонение σ .

Коэффициент вариации v .

Ошибка среднего σ_x .

Наряду с необходимым набором данных не менее важно уметь приводить результаты статистического анализа с нужной степенью точности. Точность оценки констант определяется точностью ошибок. Не стоит проводить оценки с большей точностью, чем их ошибки, но вполне допустимо ограничиться меньшей точностью, если большая точность практически не нужна.

4.4. Интервальные оценки

По известным выборочным характеристикам можно построить интервал, в котором с той или иной вероятностью находится генеральный параметр. Вероятности, признанные достаточными для уверенного суждения о генеральных параметрах на основании известных выборочных характеристиках, называют *доверительными*. Понятие это предложено Фишером. Обычно в качестве доверительных используют вероятности $P_1 = 0,95$; $P_2 = 0,99$ и $P_3 = 0,999$. Доверительным вероятностям соответствуют следующие величины нормированных отклонений.

$$P_1 = 0,95 \quad t_1 = 1,96$$

$$P_2 = 0,99 \quad t_2 = 2,58$$

$$P_3 = 0,999 \quad t_3 = 3,29$$

С доверительной вероятностью тесно связан уровень значимости α под которым понимают разность $\alpha = 1 - P$.

Выборочное распределение некоторого показателя, например средней арифметической при достаточно больших объёмах имеет нормальную форму, можно утверждать, что

$$-t \leq \frac{\bar{x} - \mu}{\sigma_x} \leq t \quad (75)$$

Это означает, что вероятность того, что средняя выборки отклонится от генеральной средней на долю квадратической ошибки, может быть оценена через нормированное значение по таблицам нормального распределения. Отсюда *доверительный интервал для генеральной средней*

$$\bar{x} - t\sigma_x \leq \mu \leq \bar{x} + t\sigma_x \quad (76)$$

или

$$\bar{x} - \frac{t\sigma_x}{\sqrt{n}} \leq \mu \leq \bar{x} + \frac{t\sigma_x}{\sqrt{n}} \quad (77)$$

Среднюю и ошибку определяют по выборке, а t зависит только от значений доверительной вероятности.

Пример 7.

Распределение характеризуется средней $x = 11,94$, ошибка – 1,27, $n = 100$. Построить 95 % доверительный интервал для генеральной средней этого распределения.

$$11,94 - 1,96 * 1,27/10 \leq \mu \leq 11,94 + 1,96 * 1,27/10$$

$$11,70 \leq \mu \leq 12,18$$

Это довольно узкий интервал, значит выборочная средняя хорошо оценивает генеральную среднюю.

Доверительный интервал для дисперсии нормально распределённой генеральной совокупности имеет вид

$$\sigma_x^2 - t\sigma_x^2 * \sqrt{\frac{2}{n}} \leq \sigma^2 \leq \sigma_x^2 + t\sigma_x^2 * \sqrt{\frac{2}{n}} \quad (78)$$

Пример 8.

Определить границы доверительного интервала для дисперсии. $n = 100$, дисперсия = 1,60. Ошибка выборочной дисперсии $\sigma_{\sigma^2} = 1,6 * \sqrt{2}/100 = 0,226$. Отсюда границы доверительного интервала для дисперсии $1,16 \leq \sigma^2 \leq 2,04$, а для стандартного отклонения $1,08 \leq \sigma \leq 1,43$.

Более точно доверительный интервал строят с применением критерия χ^2 Пирсона (хи-квадрат). Критические точки приведены в таблице приложений. При использовании критерия χ^2 для построения доверительного интервала применяют двусторонний уровень значимости, т.е. 2,5 % и 97,5 % для 95 % доверительного интервала. Границы интервала определяют по формулам.

$$P_H = \frac{(n-1)\sigma_x^2}{\chi_{2,5}^2} \quad P_B = \frac{(n-1)\sigma_x^2}{\chi_{97,5}^2} \quad (79)$$

Пример 9.

Для примера 7 применим этот способ. $n = 100$; $(n - 1) \sigma^2 = 99 * 1,60 = 158,4$. В таблице приложений для $n - 1 = 99$ и $\alpha = 2,5$ % находим $\chi^2 = 128,42$ и для $P = 97,5$ % $\chi^2 = 73,36$. Отсюда границы доверительного интервала для дисперсии $1,23 \leq \sigma^2 \leq 2,16$, а для стандартного отклонения $1,11 \leq \sigma \leq 1,47$.

Границы *доверительного интервала для генерального коэффициента вариации* определяют по формулам

$$P_H = \frac{Cv}{1 + K\sqrt{1 + 2Cv^2}} \quad P_B = \frac{Cv}{1 - K\sqrt{1 + 2Cv^2}} \quad (80)$$

где

$$Cv = \frac{\sigma_x}{x} \quad K = \frac{t}{\sqrt{2(n-1)}}$$

Пример 10.

Коэффициент вариации равен $V = 10,6\%$ или $Cv = 0,106$. Определим границы доверительного интервала для генерального коэффициента вариации.

$$K = \frac{1,96}{\sqrt{2(100-1)}} = 0,139$$

Подставляем значения в формулы

$$P_H = \frac{0,106}{1 + 0,139\sqrt{1 + 2(0,106)^2}} = 0,093$$

$$P_B = \frac{0,106}{1 - 0,139\sqrt{1 + 2(0,106)^2}} = 0,123$$

Это значит, что $9,3\% \leq V \leq 12,3\%$.

Границы доверительного интервала для доли определяют также, как и для генеральной средней. $P_H = p - t^* \sigma_m$ и $P_B = p + t^* \sigma_m$. Эти формулы пригодны тогда, когда доли равны между собой или незначительно отличаются от 50 %. Если же $75\% \leq P \leq 25\%$, то границы определяют по формуле

$$P = \frac{1}{n + t^2} \left[\left(m + \frac{t^2}{2} \right) \pm t^2 \sqrt{\frac{m(n-m)}{n} + \frac{t^2}{4}} \right] \quad (81)$$

где n – объём выборки, m – абсолютная численность одной из групп, t – нормированное отклонение.

Пример 11.

Обследовались 150 студентов, из которых оказалось 20 больных. Определить границы генеральной доли больных в вузе.

Выборочная доля больных $p = 20/150 = 0,13$ или 13 %. При уровне значимости 0,05 $t = 1,96 \approx 2$. Подставляем данные в формулу.

$$P = \frac{1}{150 + 4} \left[\left(20 + \frac{4}{2} \right) \pm 4 \sqrt{\frac{20(150-20)}{150} + \frac{4}{4}} \right] = 0,143 \pm 0,11$$

Отсюда $3,2\% \leq P \leq 25,4\%$.

Вопросы к главе 4.

1. Что такое ошибки репрезентативности, с чем они связаны?
2. Зависят ли ошибки репрезентативности от объёма выборки?
3. Ошибка среднего арифметического и её размерность, по какой формуле она вычисляется?
4. Можно ли характеризовать среднее арифметическое генеральной совокупности по отдельному значению случайной величины, и если можно, то чему будет равна в этом случае ошибка среднего?
5. Каков смысл ошибок дисперсии, коэффициента вариации и доли?
6. Для чего сравнивают выборочные оценки с теоретически предполагаемым значением параметра?
7. В чём состоит нулевая гипотеза при сравнении оценки исследуемого параметра с некоторой константой?

8. Что такое константа в рамках вариационной статистики?
9. Между чем признаются различия?
10. Какие различия признаются статистическими значимыми?
11. Как зависит значимость оцениваемой выборочной разности от объёма выборки?
12. В чём особенности сравнения средних двух независимых выборок, имеющих разные дисперсии?
13. В чём особенность сравнения выборок с одинаковыми дисперсиями?
14. Что такое интервальная оценка среднего?

Глава 5. КРИТЕРИИ ДОСТОВЕРНОСТИ ОЦЕНОК

Выборочные характеристики являются оценками генеральных параметров. В исследованиях также важно оценить генеральные параметры по разности выборочных оценок. Приходится в каждом исследовании сравнивать между собой какие-либо выборочные показатели (средние и др.). О преимуществе какой либо выборки судят обычно по разности. Вопрос о достоверности выборочной разности с её неизбежной ошибкой решают исходя из предположения, относительно параметров сравниваемых групп, которое выражено в терминах вероятности и может быть проверено.

В биометрии применяют так называемую *нулевую гипотезу* H_0 . Она заключается в том, что разница между генеральными параметрами сравниваемых выборок равна нулю, а имеющиеся различия носят чисто случайный характер. То есть $\sigma_x = \sigma_y$.

Противоположная нулевой – альтернативная гипотеза H_a при ней $\sigma_x \neq \sigma_y$. Для проверки принятой гипотезы используют *критерии достоверности*. Это величины, у которых известны функции распределения. Значения функций распределения приведены в специальных таблицах, т.е. табулированы. Значения приведены для разных степеней свободы и уровней значимости.

Обычно при проверке статистических гипотез применяют три уровня значимости $\alpha = 0,05$; $\alpha = 0,01$ $\alpha = 0,001$. В биологических исследованиях часто считают достаточным 5 % уровень значимости. При этом нулевая гипотеза не отвергается, если вероятность ошибочной оценки, относительно принятой гипотезы превышает 5 %, т.е. $P > 0,05$. Если $P < 0,05$, то гипотезу следует отвергнуть на взятом уровне значимости.

При исследованиях требующих особой точности уровень значимости может быть уменьшен до 1 % или до 0,1 %.

В биометрии применяют два вида статистических критериев: *параметрические* и *непараметрические*. Параметрические построены на основе параметров данной совокупности (среднего, дисперсии). Непараметрические – функции, зависящие от вариант данной совокупности с их частотами. Первые проверяют гипотезы для нормально распределённых совокупностей, вторые – независимо от вида распределения.

Если распределение нормально – параметрические критерии более мощны, чем непараметрические. В случае больших отличий распределений признак от нормального вида следует применять непараметрические критерии. Когда варьирующие признаки выражаются не числами, а условными знаками, применение непараметрических критериев – единственно возможный выход.

5.1. Параметрические критерии.

5.2. t-критерий Стьюдента (t-распределение)

В 1908 г. английский математик В. Госсет (Стьюдент) нашёл закон распределения величины t .

$$t = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad f(t) = C \left(1 + \frac{t^2}{n-1} \right) \quad (82)$$

где C – константа, которая зависит от числа степеней свободы $k = n-1$.

Этот закон служит основой теории малой выборки, которая характеризует распределение выборочных средних в нормально распределяющихся совокупностях в зависимости от объёма выборки. Для практического использования t -распределения составлена специальная таблица 4 Приложений, в которой содержатся значения критических точек для разных уровней значимости и числа степеней свободы.

Оценка разности средних

При сравнении друг с другом двух независимых выборок используют следующее отношение

$$t = \frac{d}{\sigma_d}$$

где, d – разность выборочных средних, σ_d – ошибка разности средних.

H_0 гипотезу отвергают, если фактически установленная величина критерия t_ϕ превосходит или окажется равной критическому значению t_{st} , т.е. $t_\phi \geq t_{st}$.

Ошибку разности средних определяют по формулам:

1) для выборок с одинаковым объёмом $n_1 = n_2$

$$\sigma_d = \sqrt{\frac{\sum (x_i - \bar{x}_1)^2 + \sum (x_i - \bar{x}_2)^2}{(n-1)n}} \quad (83)$$

2) для выборок с разным объёмом $n_1 \neq n_2$

$$\sigma_d = \sqrt{\frac{\sum (x_i - \bar{x}_1)^2 + \sum (x_i - \bar{x}_2)^2}{n_1 + n_2 - 2} \left(\frac{n_1 + n_2}{n_1 n_2} \right)} \quad (84)$$

Если H_0 гипотеза не опровергается, это не является доказательством равенства между неизвестными параметрами совокупностей. Напомним, что применение критерия предполагает равенство дисперсий для двух выборок $\sigma_1^2 = \sigma_2^2$. Если это равенство не соблюдается, то величину критерия находят по формуле

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (85)$$

А число степеней свободы k

1) для выборок с одинаковым объёмом $n_1 = n_2$

$$k = n - 1 + \frac{2n - 2}{\frac{\sigma_1^2}{\sigma_2^2} + \frac{\sigma_2^2}{\sigma_1^2}} \quad (86)$$

2) для выборок с разным объёмом $n_1 \neq n_2$

$$k = \frac{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^2}{\left(\frac{\left(\frac{\sigma_1^2}{n_1}\right)^2}{n_1 + 1} + \frac{\left(\frac{\sigma_2^2}{n_2}\right)^2}{n_2 + 1}\right)} - 2 \quad (87)$$

В программе Microsoft Excel реализован двухвыборочный t -тест для средних. Для этого обращаются к пункту меню СЕРВИС. Команда АНАЛИЗ ДАННЫХ открывает окно со списком инструментов анализа.

Следует выбрать из списка пункт ДВУХВЫБОРОЧНЫЙ ТЕСТ С ОДИНАКОВЫМИ ДИСПЕРСИЯМИ, ИЛИ ДВУХВЫБОРОЧНЫЙ ТЕСТ С РАЗЛИЧНЫМИ ДИСПЕРСИЯМИ, в зависимости от того какие пары выборок анализируются.

Правильное применение критерия предполагает нормальность распределения и равенство генеральных дисперсий. Если эти условия не выполняются, то более эффективно применить непараметрические критерии.

Оценка средней разности между выборками с попарно связанными вариантами

Довольно часто приходится сравнивать выборки с попарно связанными вариантами, т.е. *зависимыми*. В таких случаях оценкой разности между генеральными средними будет *средняя разность*, определяемая из суммы разностей между попарно связанными вариантами сравниваемых групп

$$d = \frac{\sum d_i}{n} \quad (88)$$

$$\sigma^2 = \frac{\sum (d_i - \bar{d})^2}{n - 1} \quad (89)$$

Оценкой генеральной дисперсии разности средних служит выборочная дисперсия

$$\bar{d} = \frac{\sum d_i}{n} = \bar{x}_1 - \bar{x}_2$$

где n – число парных наблюдений; $d_i = x_i - y_i$;

Ошибку средней разности определяют по формулам

$$\sigma_d = \sqrt{\frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{(n-1)n}} \quad \sigma_d = \sqrt{\frac{1}{n-1} \left(\frac{\sum d_i^2}{n} - \bar{d}^2 \right)} \quad (90)$$

Если $t_{\phi} \geq t_{st}$ то нулевая гипотеза должна быть отвергнута.

Пример 12. На одних и тех же девяти скальных участках изучали количество двух видов лишайников, находящихся на 1 м². Следует установить, имеются ли различия средней разности между ними

Участок	Вид А(x ₁)	Вид В(x ₂)	Разность d	D ₂
1	13,1	15,6	-2,5	6,25
2	12,8	17,2	-4,4	19,36
3	14	16,1	-2,1	4,41
4	15,9	18,3	-2,4	5,76
5	12,5	14,8	-2,3	5,29
6	16,1	17,7	-1,6	2,56
7	14,2	16,8	-2,6	6,76
8	12,9	14,9	-2	4
9	14,2	17,5	-3,3	10,89
Сумма 9	125,7	148,9	-23,2	65,28

Среднее значение для вида А

$$\bar{x}_1 = \sum \frac{x_1}{n} = \frac{125,7}{9} = 13,97 \text{ шт.}$$

Среднее значение для вида В

$$\bar{x}_2 = \sum \frac{x_2}{n} = \frac{148,9}{9} = 16,54 \text{ шт.}$$

Средняя разность

$$\bar{d} = \sum \frac{d}{n} = \frac{23,2}{9} = 2,58 \text{ шт.}$$

Средняя ошибка выборочной оценки

$$s_{\bar{d}} = \sqrt{\frac{\sum d^2 - (\sum d)^2/n}{n(n-1)}} = \sqrt{\frac{65,28 - (23,2)^2/9}{9(9-1)}} = 0,2758$$

Фактическое значение критерия

$$t_{\text{факт.}} = \frac{\bar{d}}{s_{\bar{d}}} = \frac{2,58}{0,2758} = 9,35$$

Число степеней свободы при 9 сопряжённых парах $\nu = n - 1 = 9 - 1 = 8$

Табличное значение критерия Стьюдента при 8 степенях свободы $t_{0,95} = 2,31$, $t_{0,01} = 3,36$ (табл. 4 Приложения). Это не превышает фактическое значение $t_{\text{факт.}} = 9,35$, следовательно, установлено отличие между численностью видов с вероятностью не ниже 0,99.

Оценка разности между долями

Выборочная доля зависит от числа единиц в выборке, имеющих оцениваемый признак, а общее число таких единиц составляет *генеральную долю*. Оценкой разности между генеральными долями служит разность между вы-

борочными долями. $p_1 - p_2 = d$. Отношение этой разности к своей ошибке даёт случайную величину $t = d/s_d$, которая следует распределению Стьюдента. H_0 гипотезу отвергают, если $t_\phi \geq t_{st}$ для $k = n_1 + n_2 - 2$ и принятого уровня значимости α . Ошибка разности между долями (при разнице в численности не более 25 %), вычисляют по формулам

$$\sigma_{d_p} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \quad \sigma_{d_p} = \sqrt{n_1 p_1 q_1 + n_2 p_2 q_2} \quad (91)$$

где, $q = 1 - p$

Когда сравнивают доли из неравновеликих выборок (при разнице в численности $75\% < p < 25\%$), ошибку разности определяют по формуле

$$\sigma_{d_p} = \sqrt{pq \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \quad (92)$$

P определяют как средневзвешенную из $p_1 p_2$ долей, или же из абсолютных численностей групп

$$p = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2} = \frac{m_1 + m_2}{n_1 + n_2} \quad (93)$$

где n_1 и n_2 численности групп, из которых определяют доли $p_1 = m_1/n_1$ и $p_2 = m_2/n_2$. Если неравновеликие группы выражены абсолютными числами, то ошибку разности определяют по формуле

$$\sigma_{d_p} = \sqrt{\frac{m_1 + m_2}{n_1 + n_2} \left(1 - \frac{m_1 + m_2}{n_1 + n_2} \right) \frac{n_1 + n_2}{n_1 n_2}} \quad (94)$$

Пример 13. В потомстве от скрещивания жёлтого и зелёного гороха было получено в первой группе 14 – жёлтых и 9 – зелёных. Во второй группе – 12 жёлтых и 8 зелёных. Разница составила $14 - 12 = 2$ единицы. Определить ошибку этой разницы.

$$\sigma_{d_p} = \sqrt{\frac{14 * 9}{23} + \frac{12 * 8}{20}} = \sqrt{10,28} = 3,2$$

Критерий $t_\phi = 2/3,2 = 0,62$. По таблице находим для 5 % уровня значимости и $k = 23 + 20 - 2 = 41$ $t_{st} = 2,02$. Так как фактическое значение не превосходит табличное, то H_0 гипотезу отвергнуть нельзя, разница в потомстве оказалась статистически недостоверной.

Описанные критерии пригодны при ($25\% < p < 75\%$). Особенно важно это для небольших выборок. Более универсальным и свободным от подобных ограничений является способ проверки равенства долей, основанный на использовании угловой трансформации (ϕ – преобразования Фишера).

Этот метод предусматривает сравнивать доли в процентах с введением поправки Йейтса на непрерывность, равной $1/2n$, которую вычитают из большей и прибавляют к меньшей доле. Затем по таблице значений $\phi = 2 \arcsin \sqrt{p}$ находят величины для исправленных долей: $p_1 \% + 100/(2n)$ и $p_2 \% - 100/(2n)$, берут их разность и делят её на ошибку, определяемую по формуле

$$\sigma_{d_p} = \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (95)$$

Условием для непринятия нулевой гипотезы служит следующее выражение $t_{\phi} \geq t_{st}$, для $k = n_1 + n_2 - 2$ и принятого уровня значимости α .

Пример 14. Для условий примера 2 имеем $p_1 = 0,608$, $p_2 = 0,6$. Тогда (p_1 %) $60,8 + 100/(2 * 23) = 62,97$ %. (p_2 %) $60 - 100/(2 * 20) = 57,2$ %. По таблице приложений для $p_1 = 62,97$ находим $\varphi_1 = 1,832$ и для $\varphi_2 = 1,721$. Отсюда критерий достоверности

$$t = \frac{\varphi_1 - \varphi_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = (\varphi_1 - \varphi_2) \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \geq t_{st}$$

$$t = \frac{1,832 - 1,721}{\sqrt{\frac{1}{23} + \frac{1}{20}}} = 0,36$$

Критерий $t_{\phi} = 0,36$. По таблице находим для 5 % уровня значимости и $k = 23 + 20 - 2 = 41$ $t_{st} = 2,02$. Так как фактическое значение не превосходит табличное, то H_0 гипотезу отвергнуть нельзя, разница в потомстве оказалась статистически недостоверной.

Оценка разности между выборочной и генеральной долями

При оценке разности между известной генеральной долей P и выборочной долей p , H_0 сводится к предположению, что разница между ними возникла случайно. Критерий Стьюдента имеет вид

$$t = \frac{d_p}{\sqrt{\frac{p(1-p)}{n}}} \quad (96)$$

где знаменатель – ошибка доли. Условием для непринятия H_0 служит критерий

$$t_{\text{факт.}} = \frac{\bar{P} - p}{\sigma_{d_p}} \geq t_{st}$$

5.3. F-критерий Фишера (F-распределение)

Для проверки нулевой гипотезы о равенстве генеральных дисперсий нормально распределённых совокупностей при малом объёме выборки t-критерий недостаточно точен. Д. Снедекор предложил использовать отношения выборочных дисперсий, назвав его в честь Фишера:

$$F = \frac{\sigma_1^2}{\sigma_2^2} \quad (97)$$

Условие для соотношения дисперсий следующее: $\sigma_1^2 \geq \sigma_2^2$.

Если сравниваемые выборки взяты из одной совокупности, или из разных совокупностей с одинаковыми дисперсиями, то величина критерия не превысит критические точки, указанные в таблице приложений для k_1 k_2 и уровня значимости α . Если же выборки взяты из разных совокупностей, с различными по значению дисперсиями, то $F_{\phi} \geq F_{st}$ и нулевая гипотеза должна быть отвергнута.

F – критерий применяют для оценки разности между долями из неравновеликих выборок. Нулевая гипотеза отвергается при условии, что

$$F = (\varphi_1 - \varphi_2)^2 \frac{n_1 n_2}{n_1 + n_2} \geq F_{st} \quad \text{для} \left\{ \begin{array}{l} k_1 = 1 \\ k_2 = n_1 + n_2 - 2. \end{array} \right\}$$

$$F = \frac{\bar{x}_1}{x_2 + 1} \geq F_{st} \quad \text{для} \left\{ \begin{array}{l} k_1 = 2(\bar{x}_2 + 1) \\ k_2 = 2\bar{x}_1. \end{array} \right\}$$

При этом k_1 находят по горизонтали, а k_2 в первом столбце таблицы 5 Приложений. Если оценивают разность между средними двух выборок, извлечённых из совокупностей, которые распределяются по закону Пуассона, F -критерий строят следующим образом и принятого уровня значимости α .

Оценка разности между коэффициентами вариации

Разность между коэффициентами вариации сравниваемых групп, взятых из нормально распределённых совокупностей можно оценить с помощью t -критерия Стьюдента. Оценкой разности служит

$$t = \frac{d_V}{\sqrt{\sigma^2_{v_1} + \sigma^2_{v_2}}}$$

где d_V – разность коэффициентов вариации сравниваемых групп; $\sigma^2_{v_1}$ – ошибка коэффициентов вариации сравниваемых групп.

Нулевую гипотезу отвергают, если $t_{\phi} > t_{st}$ для $k = n_1 + n_2 - 2$ и принятого уровня значимости α .

Разность между коэффициентами вариации можно оценить сопоставлением доверительных интервалов, построенных для генеральных параметров сравниваемых групп. Близость границ доверительных интервалов показывает отсутствие существенных различий между коэффициентами вариации групп.

5.4. Непараметрические критерии

Не всегда биологические признаки распределяются по нормальному закону. Это ограничивает применение параметрических критериев. Кроме того, часто исследуются качественные признаки, которые выражаются порядковыми номерами, индексами, баллами и др. в таких случаях надо использовать *непараметрические критерии*.

Среди непараметрических критериев видное место принадлежит *ранговым критериям*. Их применение основано на ранжировании членов сравниваемых групп. При этом сравниваются не сами члены, а их порядковые номера или ранги.

5.5. X-критерий Ван-дер-Вардена

Его применяют для проверки нулевой гипотезы при сравнении друг с другом независимых выборок. Сравнимые выборки ранжируют в один общий ряд по возрастающим значениям признака и нумеруют. Затем по порядковым номерам меньшей выборки находят значение

$$\frac{R}{N+1}$$

где $N+1 = n_1 + n_2 + 1$, R = порядковый номер ранжированного ряда, их ранг.

С помощью таблицы 9 Приложений находят значения функции $\psi[R/(N+1)]$ для каждого значения $R/(N+1)$. Суммируя значения функции с

учётом знаков, получают величину $X_\phi = \sum \psi [R/(N+1)]$, которую сравнивают с критической точкой этого критерия. X_{st} для принятого уровня значимости α и $N = n_1 + n_2$. Критические точки X -критерия с учётом разности в объёмах выборок приведены в таблице 10 Приложений. Нулевая гипотеза сводится к предположению, что сравниваемые выборки извлечены из генеральных совокупностей с одинаковыми функциями распределения. Если окажется, что $X_\phi \geq X_{st}$ то H_0 должна быть отвергнута на принятом уровне значимости.

5.6. U-критерий Уилкоксона (Манна-Уитни)

H_0 гипотезу можно проверить и с помощью *рангового критерия Уилкоксона (Манна-Уитни)*. Для расчёта этого критерия необходимо:

1. Расположить числовые значения выборок в возрастающем порядке в один ряд и пронумеровать.
2. Отдельно для каждой выборки найти суммы рангов R и определить величины, которые отображают связь между суммами рангов первой и второй выборки.
3. В качестве U -критерия использовать меньшую величину U_ϕ , которую сравнить с табличным значением U_{st} , взятую из таблицы 11 или 6 Приложений. Условием для принятия H_0 гипотезы служит неравенство $U_\phi > U_{st}$.

$$U_1 = R_1 - \frac{n_1(n_1 + 1)}{2} \quad (98)$$

$$U_2 = R_2 - \frac{n_2(n_2 + 1)}{2} \quad (99)$$

5.7. Критерий знаков Z

Если результаты исследований выражаются не числами, а знаками (+) или (-), различия между попарно связанными членами выборок сравнивают с помощью критерия знаков Z . Если сравниваемые значения двух выборок существенно не отличаются друг от друга, то число плюсовых и минусовых разностей окажется одинаковым. Если плюсы или минусы будут заметно преобладать, то это указывает на положительное или отрицательное действие изучаемого фактора на результативный признак. Z_ϕ служит большее число разностей с одинаковым знаком. При этом нулевые разности в расчёт не берутся. H_0 отвергается если $Z_\phi \geq Z_{st}$ для принятого уровня значимости и числа парных наблюдений без учёта нулевых разностей. Критические точки Z_{st} приведены в таблице 12 Приложений.

5.8. T-критерий Уилкоксона

Для зависимых выборок (попарно связанных общими условиями) можно для оценки различий применить *ранговый T-критерий Уилкоксона*. Он более мощен, чем критерий знаков. Рассчитываю его по нижеприведённому алгоритму:

1. Ранжируют попарные разности в один общий ряд. При этом нулевые разности в расчёт не принимают. Все остальные, независимо от знака, ранжируют так, чтобы наименьшая абсолютная разность получила первый ранг. Одинаковым по величине разностям присваивают один и тот же ранг.
2. Находят суммы рангов положительных и отрицательных разностей. Меньшую сумму без учёта знака используют в качестве величины T -критерия.
3. Сравнивают эту величину T_ϕ с критическим T_{st} для принятого уровня значимости α и числа парных наблюдений, которое берут без учёта нуле-

вых разностей. Нулевую гипотезу отвергают, если $T_{\phi} > T_{st}$. Критические значения парного критерия Уилкоксона приведены в таблице 13 Приложений.

Вопросы к главе 5.

1. Что такое нулевая гипотеза?
2. Что такое альтернативная гипотеза?
3. Дать понятие параметрического критерия.
4. Сформулировать закон t-распределения.
5. Что такое зависимые выборки?
6. Почему критерий Стьюдента непригоден для проверки равенства долей при малых значениях p ($p < 0.25\%$)?
7. В чём заключается способ угловой трансформации?
8. Почему критерий Фишера применяют для оценки дисперсий малочисленных выборок?
9. От чего зависит величина F-критерия?
10. Когда следует применять непараметрические критерии?
11. На чём основано применение ранговых критериев?
12. Когда применяется критерий Ван-дер-Вардена?
13. Как можно проверить гипотезу о принадлежности сравниваемых независимых выборок к одной генеральной совокупности?
14. Как произвести сравнение двух выборок, если результаты наблюдений выражаются знаками «плюс» (+) и «минус» (-)?

Глава 6. ПРОВЕРКА ГИПОТЕЗ О ЗАКОНАХ РАСПРЕДЕЛЕНИЯ

6.1. Критерий хи-квадрат (χ^2 распределение)

Проверку гипотез о законах распределения производят с помощью специально разработанных критериев. Один из них нашёл широкое применение в биометрии. Это *критерий согласия* или *соответствия* χ^2 (1900, Пирсон).

$$\chi^2 = \sum_{i=1}^k \frac{(f - f')^2}{f'} = \sum_{i=1}^k \left(\frac{d^2}{f'} \right) \quad (100)$$

где f – вычисленные частоты; f' – теоретические частоты.

Величина χ^2 всегда положительна. При полном совпадении эмпирических частот с вычисленными $\chi^2 = 0$. Распределение вероятных значений случайной величины χ^2 является непрерывным и асимметричным. Оно зависит от числа степеней свободы k и приближается к нормальному по мере увеличения объёма выборки n . Поэтому не рекомендуется применять этот критерий к дискретным распределениям. Он сопровождается в этом случае погрешностями, особенно при малых выборках.

Чтобы оценки были более точными выборка должна содержать не менее 50 вариантов. В крайних классах должно быть не менее 5 вариант. Число степеней свободы:

- При нормальном – $k = N - 3$.
- При пуассоновском – $k = N - 2$.

□ В других случаях – особо.

Нулевая гипотеза сводится к предположению, что различия между частотами носят случайный характер.

Для проверки служит неравенство $\chi^2 \Phi \geq \chi^2_{st}$. При его выполнении нулевая гипотеза должна быть отвергнута на принятом уровне значимости с числом степеней свободы k . Критические точки приведены в таблице 8 Приложений.

Критерий хи-квадрат χ^2 применяют и для оценки сходства между вариационными рядами, частоты которых распределяются в границах одних и тех же классов. В таких случаях критерий определяется по формулам

При $n_1 = n_2$

$$\chi^2 = 4 \left(\sum_{i=1}^k \frac{f_i^2}{f_1 + f_2} \right) - N \quad (101)$$

При $n_1 \neq n_2$

$$\chi^2 = \frac{N^2}{n_1 n_2} \left(\sum_{i=1}^k \frac{f_i^2}{f_1 + f_2} - \frac{n^2}{N} \right) \quad (102)$$

где f_1, f_2 частоты сравниваемых распределений.

$n_1 = \sum f_1$ $n_2 = \sum f_2$ – объём рядов распределения.

$N = n_1 + n_2$.

Число степеней свободы определяют по числу классов $N - 1$.

При группировке выборки в четырёхпольную или многопольную таблицу критерий хи-квадрат определяют по формуле

$$\chi^2 = \sum \frac{d_1^2}{f_1'} + \sum \frac{d_2^2}{f_2'} + \dots + \sum \frac{d_n^2}{f_n'} \quad (103)$$

где d – разница между эмпирическими и вычисленными частотами; f' – теоретические частоты.

Число степеней свободы $k = (c-1)(z-1)$, где c – число строк, z – число граф или столбцов таблицы.

6.2. Критерий Ястремского J.

Профессор Ястремский исследовал критерий Пирсона хи-квадрат и обнаружил, что он не даёт информации о степени близости между частотами. Поэтому он построил другой критерий согласия

$$J = \frac{|C - N|}{\sqrt{2N + 4\theta}} \quad (104)$$

где N – число классов ряда; θ – величина, зависящая от числа классов N . Если $N \leq 20$, то $\theta \leq 0,6$. Поэтому величину 4θ можно считать равной 2,4.

$q = 1 - p$, где $p = f(t)$, т.е. функция нормированного отклонения J имеет непрерывную функцию распределения и подчинена нормальному закону. С вероятностью $P = 99,5\%$ можно утверждать что H_0 верна, если $J \leq 3$.

6.3. Критерий Уилка-Шапиро

Для проверки нормальности распределения можно использовать критерий Уилка-Шапиро. Условием применимости этого критерия служит объём выборки $3 \leq n \leq 50$ и несгруппированность исходных данных.

H_0 сводится к предположению о нормальности распределения рассматриваемой совокупности. Алгоритм расчёта критерия следующий:

1. Строят ранжированный ряд по возрастанию x_i .
2. По выборочным данным вычисляют величину C

$$C = \sum x_i^2 - \frac{(\sum x_i)^2}{n} \quad C = \sum \frac{(f - f')^2}{f'q} \quad (105)$$

Находят величину k :

- Если n – чётное, то $k = n/2$
- Если n – нечётное, то $k = (n - 1)/2$

3. Вычисляют вспомогательную величину B по формуле

$$B = \sum_{i=1}^{i=k} a_{n-i+1} (x_{n-i+1} - x_i) = a_n (x_n - x_1) + a_{n-1} (x_{n-1} - x_2) + \dots + a_{n-k+1} (x_{n-k+1} - x_k)$$

где a_{n-i+1} – некоторые коэффициенты, значения которых в зависимости от n приведены в табл. 26 Приложений. При нечётном n медианное значение занимающее $[(n+1)/2]$ место при вычислении B не используется.

4. Вычисляют величину W по формуле

$$W = \frac{B^2}{C}, \quad (106)$$

которая служит статистикой для проверки H_0 гипотезы. Если $W_{\phi} \leq W_{st}$ то с уровнем значимости α нулевая гипотеза отвергается, т.е. распределение считается отличным от нормального. Критические значения W_{st} находят из таблицы 7 Приложений.

6.4. Оценка трансгрессии рядов

При распределении независимых выборок взятых из разных генеральных совокупностей, случается, что некоторая часть этих выборок оказывается в одних и тех же классах вариационного ряда. Ряды, у которых часть классов оказывается общей, называются *трансгрессирующими*. А сам факт – *трансгрессией*. График трансгрессии изображён на рисунке 11. Величину трансгрессии можно измерить, выразив сумму трансгрессирующих вариантов в процентах от общей численности обеих выборок. Но этот способ неточен и не даёт истинного представления о её размере.

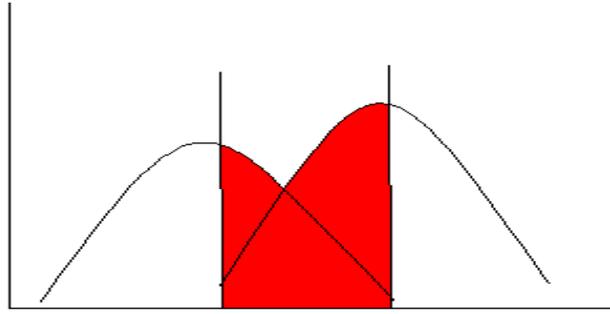


Рис. 11. Трансгрессия рядов

Более точно рассчитать величину трансгрессии можно по формуле

$$Tr = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} \quad (107)$$

где n_1 и n_2 – объёмы распределений;

$$P_1 = 0,5 + 0,5\Phi(t_1) \text{ и } P_2 = 0,5 + 0,5\Phi(t_2);$$

$$t_1 = (\min_2 - x_1)/\sigma_1 \text{ и } t_2 = (\max_1 - x_2)/\sigma_2$$

$$\min_2 = x_2 - 3\sigma_2 \text{ и } \max_1 = x_1 + 3\sigma_1$$

x – средние арифметические выборок.

Показатель трансгрессии выражают в долях единицы или процентах. Значения функции $\Phi(t)$ приведены в таблице 3 Приложений. Эти значения нужно брать с отрицательным знаком, если $\min_2 > x_1$ и $\max_1 < x_2$. Т.е. исходить из

$$P = 0,5 - 0,5 \Phi(t).$$

6.5. Проверка сомнительных вариантов

В отдельных случаях в выборке могут встречаться варианты, сильно отличающиеся по значению от всех остальных. Сомнительные варианты могут попасть в выборку ошибочно и влиять на результаты анализа, особенно если выборка мала по объёму. Но нельзя исключать из выборки сомнительные варианты без статистической проверки.

Для проверки нулевой гипотезы (сомнительные варианты принадлежат к одной генеральной совокупности) применяют особые критерии. Один из них – тау-критерий:

$$\tau_{\max} = \frac{|x_{\text{сомн}} - \bar{x}|}{\sigma} \quad (108)$$

Если $\tau_{\max} \geq \tau_{\max \alpha}$, то нулевая гипотеза может быть отвергнута, и допускается выбраковка.

Если же $\tau_{\max} < \tau_{\max \alpha}$ то следует считать, что статистических оснований для выбраковки нет. Критические значения $\tau_{\max \alpha}$ приведены для уровней значимости α и разных объёмов выборки в таблице 14 Приложений.

6.6. Применение коэффициентов асимметрии и эксцесса для проверки нормальности распределения

Гипотезу о законе распределения можно проверить разными способами, в частности при помощи коэффициентов асимметрии и эксцесса. Показа-

тели асимметрии и эксцесса – это случайные величины, сопровождающиеся ошибками. Возможно определить эти показатели по формулам, однако более удобный способ содержится в таблицах 15 и 16 Приложений. В них даны критические значения этих коэффициентов для конкретного объёма выборки и уровня значимости α .

Вопросы к главе 6.

1. Указать величины коэффициентов асимметрии и эксцесса для нормального распределения.
2. Можно ли утверждать, что
 - а) распределение отлично от нормального, если A (или E) $\neq 0$;
 - б) распределение нормально, если $A = E = 0$
3. Какое соотношение между значениями средней арифметической, медианой и модой должно соблюдаться в нормальных и близких к ним распределений?
4. Какой знак имеет величина критерия χ^2 и почему?
5. Сколько вариантов должна содержать выборка, чтобы оценка была точной?
6. Как устанавливается число степеней свободы?
7. Что сказывается на величине критерия χ^2 ?
8. Чем отличается критерий Ястремского от критерия χ^2 , объясните?
9. При каком объёме выборки можно применять критерий Уилка-Шапиро?
10. Условия применимости критерия Уилка-Шапиро.
11. Что такое трансгрессия?
12. Какой показатель используют для оценки величины трансгрессии, в каких единицах он выражается?
13. Почему сомнительные варианты необходимо проверять статистически на принадлежность к генеральной совокупности?

Глава 7. ДИСПЕРСИОННЫЙ АНАЛИЗ

Выборки можно сравнивать между собой простыми способами, рассмотренными в предыдущих главах. Однако встречаются более сложные задачи, когда приходится сравнивать одновременно несколько выборок, объединённых в один статистический комплекс. Фишером был предложен метод комплексной оценки сравниваемых средних. Его назвали *дисперсионным анализом*. Метод основан на разложении общей дисперсии статистического комплекса на составляющие её компоненты. Сравнивая их друг с другом, с помощью F-критерия определяют, какую долю общей вариации обуславливает действие регулируемых и случайных факторов.

В простейшем случае межгрупповая девиата определяется по формуле

$$D_x = \sum_{i=1}^a \frac{n(\bar{x}_i - \bar{x})^2}{N} \quad (109)$$

где $N = \sum n$; a – общее число групп; \bar{x}_i – групповая средняя; \bar{x} – общая средняя комплекса.

Внутригрупповая девиата

$$D_e = \sum_i^a \left[\sum_{j=1}^a (x_j - \bar{x}_i)^2 \right] \quad (110)$$

x_j – отдельные варианты; \bar{x}_i – их групповые средние.

Общая девиата

$$D_T = \sum_{j=1}^a (x_j - \bar{x})^2 \quad (111)$$

x_j – отдельные варианты; \bar{x} – их общая средняя комплекса

$$D_T = D_x + D_e$$

Деление девиат на число степеней свободы даёт выборочные дисперсии

$$\sigma^2_T = \frac{D_T}{k_T} \quad \sigma^2_x = \frac{D_x}{k_x} \quad \sigma^2_e = \frac{D_e}{k_e}$$

Отношение межгрупповой дисперсии (факториальной) к внутригрупповой (остаточной) служит критерием оценки влияния регулируемых и случайных факторов на результативный признак.

$$F = \frac{\sigma^2_x}{\sigma^2_e} \quad (112)$$

При $\sigma^2_x \geq \sigma^2_e$. Нулевая гипотеза сводится к предположению, что генеральные межгрупповые средние и дисперсии равны между собой а различия обусловлены случайными факторами. Её отвергают, если $F_{\text{ф}} \geq F_{\text{ст}}$ для принятого уровня значимости α и чисел степеней свободы k_x и k_e .

Применение дисперсионного анализа предъявляет определённые требования к группировке выборочных данных (например, по высоте, по диаметру, по возрастному составу и т.д.). Важно, чтобы распределение было нормальным или близко к нормальному. При этом дисперсии выборочных групп должны не слишком отличаться друг от друга. Не менее важным является условие равенства объёмов выборок, что значительно облегчает дисперсионный анализ.

7.1. Основные понятия и символы

Признаки, изменяющиеся под воздействием каких-либо причин, называют *результативными*. Причины, вызывающие изменение признака называют *факторами*. Например, масса, линейный размер и т.п. – это признаки, на которые могут оказывать влияние различные факторы: питание, упражнения, удобрения и т.д. Факторы принято обозначать буквами латинского алфавита А, В, С, D. А признаки – X, Y, Z.

На признак оказывает влияние много факторов. В опыте обычно учитывается один, он называется *регулируемым* или *организованным*. Обычно каждый регулируемый фактор испытывают серийно, в виде нескольких испытаний или градаций. Градации обозначают индексами A_1, A_2 , и т.д. число градаций фактора определяется условиями опыта: количеством сортов, дозами удобрений, и т.д. Результативные признаки тоже могут иметь свои градации, на которых испытывают действие регулируемых факторов.

При образовании статистических комплексов необходимо соблюдать два важных условия, гарантирующих правильное применение дисперсионного анализа.

1. Регулируемые факторы, действующие на признак, должны быть независимы друг от друга.

2. Выборки, входящие в статистический комплекс должны быть сделаны рандомизировано.

Форму статистического комплекса задают таблицей, строки её – число подразделений результативного признака, а столбцы – число градаций регулируемого фактора, или факторов.

Если на признак действует один регулируемый фактор, комплекс наз. *однофакторным*. Если одновременно исследуется два, три, или большее число факторов, комплекс наз. *двух-, трёх-, и многофакторным*. Варианты могут располагаться по градациям комплекса равномерно, пропорционально и неравномерно, поэтому статистические комплексы называются *равномерными, пропорциональными и неравномерными*. Равномерные и пропорциональные комплексы называют *ортогональными*, а неравномерные – *неортогональными*.

В ортогональных комплексах соблюдается равенство $D_T = D_x + D_e$. в двухфакторных: $D_T = D_A + D_B + D_{AB} + D_e$. В неортогональных это равенство нарушается. Это надо учитывать при планировании опытов.

7.2. Анализ однофакторных комплексов

Техника дисперсионного анализа однофакторного комплекса сводится к расчёту показателей варьирования. Дисперсионный анализ однофакторных равномерных комплексов удобно проводить по следующей схеме:

1. Первичные данные группируют в виде комбинационной таблицы. Градации регулируемого (организованного) фактора А располагают по горизонтали в первой строке, а числовые значения признака Х т.е. варианты размещаются по градациям фактора А.

2. Сгруппировав данные, приступают к расчёту вспомогательных величин $\sum x_i$, $\sum(\sum x_i)^2$, $\sum(x_i)^2$.

3. Затем рассчитывают девиаты

$$D_A = \sum^a \frac{(\sum x_i)^2}{n} - H \quad D_T = \sum_{i=1}^N x_i^2 - H \quad (113)$$

Или в случае равномерного комплекса

$$D_A = \frac{\sum^a (\sum x_i)^2}{n} - H \quad (114)$$

$$D_e = D_T - D_A$$

Величина $H = (\sum x_i)^2/N$, где x_i – варианты, входящие в состав комплекса; N – общее число наблюдений или объём комплекса; n – численность вариант в каждой из градаций комплекса; D_A – факториальная девиата, характеризующая варьирование обусловленное фактором А.

4. Рассчитав девиаты, переходят к определению чисел степеней свободы k:

➤ $k_T = N - 1$ для общего варьирования
 ➤ $k_A = a - 1$ для факториального варьирования
 ➤ $k_e = (N - 1) - (a - 1) = N - a$ для остаточной вариации через a обозначено число градаций фактора А. Число степеней свободы подчиняется следующему соотношению $k_T = k_x + k_e$.

5. Вычисляют выборочные дисперсии

$$\sigma^2_T = \frac{D_T}{N-1} \quad \sigma^2_A = \frac{D_A}{a-1} \quad \sigma^2_e = \frac{D_e}{N-a}$$

6. Определяют дисперсионное соотношение при $\sigma^2_x \geq \sigma^2_e$. по которому судят о воздействии фактора на результативный признак.

$$F = \frac{\sigma^2_A}{\sigma^2_e}$$

Так как полученное отношение является случайной величиной, то его сравнивают с критическим F_{st} по таблице 5 Приложений для принятого уровня значимости и уровней свободы k_A и k_e . При этом число степеней свободы для большей дисперсии находят в верхней строке, а для меньшей – в первом столбце таблицы Фишера.

Нулевую гипотезу отвергают и эффективность действия фактора А на результативный признак Х признают статистически достоверной, если $F_{\Phi} \geq F_{st}$.

Если нет сведений о характере распределения, то анализ дисперсионного комплекса проводят непараметрическими методами. Если комплекс равномерен, ранговый анализ проводят с помощью критерия Фридмана по формуле

$$\chi^2_R = \frac{12 \sum (\sum R_i)^2}{na(n+1)} - 3n(a+1)$$

где $\sum R_i$ – сумма рангов в каждой градации; n – численность вариантов в каждой градации; a – число градаций. Полученное значение χ^2_R сравнивают с критическим, взятым из таблицы 20 Приложений. Нулевая гипотеза отвергается если $\chi^2_R \geq \chi^2_{st}$.

7.3. Неравночисленные комплексы

Дисперсионный анализ однофакторных не равных по числу комплексов, принципиально не отличается от анализа равномерных комплексов. Но так как групповые средние неравномерных комплексов имеют разный статистический вес, то факториальную девиату следует вычислять по формуле

$$D_A = \sum_j^a \frac{(\sum x_i)^2}{n_j} - H \quad \text{или} \quad (115)$$

$$D_A = \sum_j^a \left(n_j \bar{x}_j^2 \right) - H \quad D_A = \sum_j^a \left[n_j (\bar{x}_j - \bar{x})^2 \right] \quad (116)$$

Применение корреляционных таблиц

В выборках большого объема варианты приходится размещать в вариационный ряд или в ряд ранжированных значений признака. В таких случаях

удобно группировать данные для дисперсионного анализа с помощью *корреляционной решётки*. Она образуется сочетанием строк и столбцов, число которых равно числу групп или классов сопряжённых рядов. Классы располагаются в верхней строке и в первом столбце таблицы. Общие частоты, обозначаемые f_{xy} , распределяются по ячейкам решётки.

Классы или значения признаков, помещённые в верхней строке таблицы, располагаются обычно слева направо в возрастающем порядке, а в первом столбце таблицы – в убывающем порядке, т.е. сверху вниз. При этом промежутки между классами могут быть равно- и неравновеликими. В случае неравновеликих промежутков девиаты рассчитывают по формулам

$$D_T = \sum_j^a (f_x x_i^2) - H \quad (117)$$

$$D_A = \sum \frac{(f_{xy} x_i^2)^2}{f_T} - H \quad (118)$$

где $H = (\sum f_x x_i)^2 / N$; f_x – частоты ряда X; f_y – частоты ряда Y; x_i – значения классовых вариантов ряда X. Остаточную девиату определяют по разности между общей (тотальной) и факториальной девиатой.

7.4. Оценка силы влияния факторов

Метод Плохинского. После того как достоверно установлено действие регулируемого фактора, можно измерить силу его влияния на результативный признак. Последнюю определяют как долю межгрупповой вариации в общем варьировании результативного признака. Для измерения силы влияния предложено несколько методов. Наибольшее признание получили методы Плохинского (1966, 1970) и Снедекора (1961). Метод Плохинского базируется на равенстве девиат $D_y = D_x + D_e$, которое осуществляется в любом дисперсионном комплексе. Показатель силы влияния, обозначаемый символом h_x^2 , строят следующим образом: все члены указанного равенства делят на D_y , что дает

$$\frac{D_x}{D_y} + \frac{D_e}{D_y} = 1$$

Отсюда вычисляют показатель силы влияния:

$$h_x^2 = \frac{D_x}{D_y} \text{ или } h_x^2 = 1 - \frac{D_e}{D_y} \quad (119)$$

Критерием достоверности этого показателя служит его отношение к своей ошибке, которую определяют по следующей приближенной формуле:

$$\sigma_{h_x^2} = (1 - h_x^2) \frac{a-1}{N-a} \quad (120)$$

где a – число градаций регулируемого фактора; N – объем дисперсионного комплекса. Нулевую гипотезу отвергают, если

$F_\phi = h_x^2 / \sigma_{h_x^2} \geq F_{st}$ для принятого уровня значимости a и чисел степеней свободы $k1 = a - 1$ (находится в верхней строке таблице 5 Приложений) и $k2 = N - a$ (находится в первом столбце той же таблицы).

Метод Снедекора. В отличие от метода Плохинского этот метод основан на применении дисперсий, причем показатель силы влияния строят с учетом действия на признак не только регулируемых, но и нерегулируемых в опыте факторов, оценкой которых служит внутригрупповая дисперсия.

$$h_x^2 = \frac{\widehat{\sigma}_x^2}{\widehat{\sigma}_x^2 + \sigma_e^2} \quad (121)$$

где $\widehat{\sigma}_x^2 = (\sigma_x^2 - \sigma_e^2)/n$ – «исправленная» межгрупповая дисперсия, равная разности между дисперсиями межгрупповой (неисправленной) и внутригрупповой, или остаточной, отнесенной к числу членов n в градациях комплекса. Если комплекс неравномерный, величину n определяют по формуле

$$n = \frac{1}{a-1} \left(N - \frac{\sum n_i^2}{N} \right) \quad (122)$$

где a – число градаций регулируемого фактора; n – численность вариант x_i в отдельных градациях фактора; $N = \sum n$ – общее число вариант, или объем комплекса. Достоверность оценок силы влияния, определяемых по методу Снедекора, устанавливают обычным в дисперсионном анализе способом – посредством F-критерия Фишера. Величину F-критерия сравнивают с критическим значением этого показателя для принятого уровня значимости a и числе степеней свободы k_x и k_e , причем последние определяют так же, как и при оценке критерия Плохинского.

Определение показателя силы влияния по Плохинскому и по Снедекору не приводит к идентичному результату. Поэтому при определении силы влияния следует указывать, каким методом вычислен этот показатель. Показатель силы влияния фактора, найденный по методу Плохинского, является смещенной оценкой. Поэтому лучше пользоваться аналогичным показателем Снедекора.

7.5. Сравнение групповых средних дисперсионного комплекса

После того как достоверно установлено влияние регулируемого фактора или факторов на результативный признак, при необходимости прибегают к сравнению групповых средних друг с другом или с какой-либо другой величиной, например с контролем, стандартом, установленной нормой и т.п. Разность между средними величинами, оценивают по t-критерию Стьюдента, т.е. по отношению указанной разности к её ошибке. Этот способ, неприменим к сравнительной оценке средних в дисперсионном комплексе, так как наряду с межгрупповой дисперсией на величине ошибки разности σ_d между групповыми средними комплекса сказывается и влияние внутригрупповой дисперсии σ_e^2 , величина которой зависит и от численности вариант x_i в группах, и от количества групп a , входящих в данный комплекс. Эти обстоятельства ограничивают применение критериев Стьюдента и Фишера. Поэтому в качестве ошибки разности между групповыми средними дисперсионного комплекса принят корень квадратный из отношения внутригрупповой

или остаточной, дисперсии к числу вариант, входящих в состав градаций фактора A

$$\sigma_d = \sqrt{\frac{\sigma_e^2}{n}} \quad (123)$$

Для оценки разности между групповыми средними дисперсионного комплекса применяют специальные методы, созданные на базе критериев Стьюдента и Фишера. Из них наиболее подходящими считают методы множественных сравнений, разработанные Дж. Тьюки (1949) и Г. Шеффе (1953).

Метод Тьюки. Этот метод применяют для проверки нулевой гипотезы при сравнении групповых средних x_1 и x_2 равных по величине групп, т.е. при $n_1 = n_2 = n$. Критерием оценки служит отношение разности сравниваемых средних к своей ошибке:

$$t_Q = \frac{|x_1 - x_2|}{\sqrt{\frac{\sigma_e^2}{n}}} \quad (124)$$

Величину t_Q сравнивают с критической точкой Q_{st} для k_e и 5% уровня значимости с учетом числа групп или градаций a регулируемого фактора A . Критические значения Q_{st} содержатся в таблице 25 Приложений. Нулевую гипотезу отвергают, если $t_Q \geq Q_{st}$ или $|x_1 - x_2| \geq \sigma_d Q_{st}$.

Метод Шеффе. В отличие от метода Тьюки этот метод множественных сравнений одинаково применим и к равно- и к неравновеликим по составу группам. Критерием достоверности различий, наблюдаемых между групповыми средними дисперсионного комплекса, служат следующие отношения:

$$F = \frac{|x_1 - x_2|}{\sqrt{\sigma_e^2}} \sqrt{\frac{n}{2}} \quad \text{при } n_1 = n_2 \quad (125)$$

$$F = \frac{|x_1 - x_2|}{\sqrt{\sigma_e^2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \quad \text{при } n_1 \neq n_2 \quad (126)$$

где a - число градаций фактора A ; F_{st} определяют с помощью таблицы Фишера (таблица 5 Приложений) для принятого уровня значимости α и чисел степеней свободы $k_1 = a - 1$ и $k_2 = N - a$, где N - объем дисперсионного комплекса.

Метод Шеффе отличается от метода Тьюки, поскольку последний неприменим к оценке групповых средних дисперсионного комплекса, вычисленных на разных объемах групп. Однако при сравнении средних x_1 и x_2 равных по величине групп предпочтение следует отдавать методу Тьюки.

В программе Excel реализован однофакторный дисперсионный анализ. Для этого обращаются к пункту меню СЕРВИС. Команда АНАЛИЗ ДАННЫХ открывает окно со списком инструментов анализа.

Следует выбрать из списка пункт **ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ**.

Вопросы к главе 7.

1. Для чего применяется дисперсионный анализ?
2. Что такое дисперсионный комплекс?
3. Когда дисперсионный комплекс считается однофакторным?
4. Что такое объём дисперсионного комплекса?
5. Какой комплекс называют равномерным и какой неравномерным?
6. В чём состоит нулевая гипотеза при дисперсионном анализе данных?
7. Что оценивает в дисперсионном анализе общая, факториальная и остаточная дисперсии?
8. В каких случаях целесообразно вычислять факториальные дисперсии?
9. Каким статистическим критерием оценивают значимость влияния фактора на результативный признак?
10. Чем отличается дисперсионный анализ неравночисленных комплексов?
11. Когда применяют непараметрические методы анализа?
12. В чём сущность рангового анализа?
13. Что такое сила влияния на результативный признак?
14. Чем отличается метод Плохинского от метода Снедекора?
15. В чём особенность сравнения средних в дисперсионном анализе?
16. Чем метод Тьюки отличается от метода Шеффе?

Глава 8. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Ещё в 6 веке до нашей эры Гиппократ обратил внимание на связь между телосложением и темпераментом людей, строением тела и предрасположенностью к тем или иным заболеваниям. Связь между варьирующими признаками обнаруживается на всех уровнях организации живого. Поэтому эту закономерность издавна стремились выразить в количественных выражениях.

Для описания связи между переменными применяют математические понятия. $y = f(x)$, здесь x – аргумент, y – соответствующее ему значение функции $f(x)$. Такого рода однозначные зависимости между величинами называют функциональными. Но однозначные функциональные связи в природе встречаются

далеко не всегда. Каждый биологический признак является функцией многих переменных. На него влияют и генетические и средовые факторы, что и обуславливает варьирование признаков. Зависимость между биологическими признаками носит статистический характер. При этом определённом значении одного признака соответствует целый вариационный ряд значений другого признака, рассматриваемого в качестве зависимой переменной или функции. Такого рода зависимость называется корреляционной, или *корреляцией*.

$$\bar{y}_x = f(x) \quad \text{и} \quad \bar{x}_y = f(y)$$

Корреляционные связи можно изучить на групповых объектах методами математической статистики. Корреляционная связь бывает линейной и нелинейной, положительной и отрицательной. Задача корреляционного анализа состоит в установлении направления и формы связи между варьирующими признаками, измерению её тесноты и к проверке достоверности выборочных показателей корреляции.

Зависимость между переменными X и Y можно выразить аналитически (формулой) и графически. График функции строят по уравнению функции, которая с 1886 г. называется *регрессией*. Здесь y_x и x_y средние арифметические, найденные при условии, что X и Y примут некоторые значения x и y . Эти средние наз. *условными*.

8.1. Параметрические показатели связи

В случае варьирующих признаков исследуют сопряжённую вариацию (ковариацию) выражая её в виде взаимно связанных отклонений вариант от их средних. *Ковариация* (cov) – это усреднённая величина отклонения каждой пары от их средних

$$\text{cov} = \frac{1}{n} \left[\sum (x_i - \bar{x})(y_i - \bar{y}) \right] \quad (127)$$

Недостаток коэффициента ковариации заключается в невозможности сравнивать коррелируемые признаки, выражаемые разными единицами измерения. Поэтому в практике применяют показатель, который называют эмпирическим коэффициентом корреляции r .

$$r_{xy} = \frac{\frac{1}{n} \left[\sum (x_i - \bar{x})(y_i - \bar{y}) \right]}{\sigma_x \sigma_y} \quad (128)$$

Коэффициент корреляции изменяется от -1 до $+1$. При независимом варьировании признаков $r = 0$. Чем сильнее сопряжённость признаков, тем ближе к 1 значение коэффициента корреляции. При положительной или прямой связи $r > 0$. Если связь отрицательна или обратная, то $r < 0$. Однако этот коэффициент характеризует лишь линейные связи, т.е. выражаемые уравнением линейной регрессии. При наличии нелинейной связи следует применять другие показатели связи.

8.2. Вычисление коэффициента корреляции

Малые выборки. При этом коэффициент корреляции вычисляют по приведённым формулам. При наличии многозначных или дробных числах применяют и следующие рабочие формулы

$$r_{xy} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \sqrt{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}} \quad (129)$$

$$r_{xy} = \frac{\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}}{\sqrt{D_x D_y}} \quad (130)$$

$$r_{xy} = \frac{D_x + D_y - D_d}{2\sqrt{D_x D_y}} \quad (131)$$

$$D_x = \sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$D_y = \sum (y_i - \bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

$$D_d = \sum (d_i - \bar{d})^2 = \sum d_i^2 - \frac{(\sum d_i)^2}{n}$$

где здесь x_i, y_i – парные варианты сопряжённых признаков X и Y; \bar{x}, \bar{y} – средние арифметические; $d = (x_i - y_i)$ – разность между парными вариантами сопряжённых признаков; n – общее число парных наблюдений. Для вычисления коэффициента корреляции необходимо вычислить следующие величины: $\sum x_i; \sum y_i; \sum x_i y_i; \sum x_i^2; \sum y_i^2$. А для последней формулы ещё и $\sum x_i$ (обязательно с учётом знака) и $\sum x_i^2$.

Эмпирический коэффициент корреляции, как и любой другой выборочный показатель, служит оценкой своего генерального параметра. Он сопровождается ошибкой

$$\sigma_r = \sqrt{\frac{1-r^2}{n-2}} \quad (132)$$

Отношение коэффициента корреляции к своей ошибке служит критерием для проверки нулевой гипотезы – предположения, что в генеральной совокупности он равен 0. Нулевую гипотезу отвергают на принятом уровне значимости, если

$$t_\phi = r \sqrt{\frac{n-2}{1-r^2}} \geq t_{st}$$

Критические значения для разных уровней значимости и числа степеней свободы $k = n - 2$ приведены в табл. 4 Приложений.

Достоверность выборочного коэффициента корреляции можно проверить по специальной таблице 22 Приложений.

Установлено, что при обработке малых выборок, особенно если $n < 30$ расчёт коэффициента корреляции даёт заниженные оценки генерального параметра. Для поправки на точность вводят величину

$$r^*_{xy} = r_{xy} \left[1 + \frac{1-r^2}{2(n-3)} \right] \quad (133)$$

Было установлено, что на малочисленной выборке коэффициент корреляции по значению превосходящий 0,5, не является точной оценкой гене-

рального параметра. Фишер нашёл более точный способ оценки генерального коэффициента корреляции. Он сводится к замене r_{xy} на преобразованную величину z , которая связана с ним отношением

$$z = \frac{1}{2} \ln \frac{1+r}{1-r} \quad \text{или} \quad z = 1,15129 \lg \frac{1+r}{1-r}$$

Преобразование коэффициента корреляции в величину z производят по специальной таблице приложений, критерием достоверности служит отношение

$$t_z = \frac{z}{\sigma_z} = z\sqrt{n-3}$$

Этот критерий используют, когда вместо коэффициента корреляции используют величину z . Критические значения для разных уровней значимости и числа степеней свободы $k = n - 2$ приведены в таблице 23 Приложений.

Применение z -преобразования позволяет с большей точностью оценивать статистическую значимость коэффициента корреляции и разность $r_1 - r_2 = d$, когда в этом появляется надобность. Можно рассчитать объём выборки для заданного значения коэффициента корреляции, который был бы достаточен для опровержения нулевой гипотезы. Для этого служит следующая формула

$$n = \frac{t^2}{z^2} + 3 \quad (134)$$

где n – искомый объём выборки; t – величина заданная по принятому уровню значимости (лучше для $\alpha = 1\%$); z – преобразованный эмпирический коэффициент корреляции.

Большие выборки. При наличии многочисленных выборок их группируют в вариационный ряд. Затем разносят в корреляционную решётку общие частоты сопряжённых рядов. При наличии положительной связи между признаками общие частоты будут распределяться по клеткам решётки в виде эллипса по диагонали от левого нижнего к правому верхнему. Если же частоты распределяются по клеткам решётки более-менее равномерно, то это указывает на отсутствие корреляционной связи между признаками.

Корреляционная решётка даёт общее представление о наличии связи между признаками. Судить о её силе можно лишь по значению и знаку коэффициента корреляции. При построении корреляционной решётки не следует брать слишком широкие классовые интервалы. Для выбора оптимального интервала используют таблицу:

Объём выборки	Значение К
$50 \geq n > 30$	$K = 1 + 3,32 \lg n$
$100 \geq n > 50$	$K = 5 \lg n$
$200 \geq n \geq 100$	$K = 7 \lg n$
$300 \geq n \geq 200$	$K = 8 \lg n$

Способ произведений. Коэффициент корреляции можно вычислить, используя основные формулы, внося в них поправку на повторяемость вариант в димерной совокупности. Введём новые обозначения: $a_x = (x_i - \bar{x})$ и $a_y = (y_i - \bar{y})$, тогда формула примет вид

$$r_{xy} = \frac{\sum_{i=1}^K f_{xy} a_x a_y}{\sqrt{\sum f_x a_x^2 \sum f_y a_y^2}} \quad (135)$$

Для определения коэффициента корреляции этим способом необходимо рассчитать

$$\sum_{i=1}^K f_{xy} a_x a_y \quad \sum f_y a_y^2 \quad \sum f_x a_x^2$$

Применяя стандартные прикладные программы, в частности MICROSOFT EXCEL можно без труда и громоздких расчётов получить коэффициент корреляции для интересующих пар выборок. Пункт меню СЕРВИС подпункт АНАЛИЗ ДАННЫХ открывают список инструментов. В нём следует выбрать пункт КОРРЕЛЯЦИЯ.

8.3. Оценка разности между коэффициентами корреляции

При сравнении коэффициентов корреляции двух независимых выборок нулевая гипотеза сводится к предположению о том, что в генеральной совокупности разница между этими коэффициентами равна нулю. Для её проверки служит t – критерий Стьюдента

$$t_{\phi} = \frac{r_1 - r_2}{\sigma_d} \geq t_{st}$$

где

$$\sigma_d = \sqrt{\sigma_{r_1}^2 + \sigma_{r_2}^2}$$

а σ_{r_1} , σ_{r_2} – ошибки сравниваемых коэффициентов корреляции.

Для принятого уровня значимости α и числа степеней свободы $k = (n_1 - 2)(n_2 - 2) = n_1 + n_2 - 4$.

Для более точной оценки разности между выборочными коэффициентами корреляции, особенно в тех случаях, когда объём выборки менее 100 единиц, а коэффициенты более 0,5, следует переводить r_{xy} в число z . При этом

$$t_z = \frac{z_1 - z_2}{\sigma_d} \geq t_{st}$$

Для $k = n_1 + n_2 - 4$ и принятого уровня значимости.

Ошибка вычисляется по формуле

$$\sigma_d = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}} \quad (136)$$

8.4. Корреляционное отношение

Для измерения нелинейной зависимости используют предложенный Пирсоном показатель, называемый *корреляционным отношением* η (эта). Связь между переменными случайными величинами X и Y выражается по-разному в зависимости от того, по значениям какой величины ранжируется совокупность. Связь между признаками характеризуется двусторонне: X по Y и Y по X ; отсюда два коэффициента этого показателя: h_{yx} и h_{xy} . Корреляционное отношение – положительная величина, она принимает значения от 0 до 1. $h_{yx} \neq h_{xy}$. Если $h_{yx} = h_{xy}$ то зависимость строго линейная.

Коэффициенты корреляционных отношений определяют по формулам

$$h_{yx} = \sqrt{\frac{\sigma^2_{yx}}{\sigma^2_y}} \quad \text{и} \quad h_{xy} = \sqrt{\frac{\sigma^2_{xy}}{\sigma^2_x}} \quad (137)$$

где

$$\sigma^2_{yx} = \frac{1}{n} \sum_{i=1}^K f_x (\bar{y}_x - \bar{y})^2 \quad \text{и} \quad \sigma^2_{xy} = \frac{1}{n} \sum_{i=1}^K f_y (\bar{x}_y - \bar{x})^2 \quad \text{групповые}$$

$$\sigma^2_y = \frac{1}{n} \sum_{i=1}^K f_y (y_i - \bar{y})^2 \quad \text{и} \quad \sigma^2_x = \frac{1}{n} \sum_{i=1}^K f_x (x_i - \bar{x})^2 \quad \text{общие дисперсии}$$

или

$$h_{yx} = \sqrt{\frac{\sum f_x (\bar{y}_x - \bar{y})^2}{\sum f_y (y_i - \bar{y})^2}} \quad \text{и} \quad h_{xy} = \sqrt{\frac{\sum f_y (\bar{x}_y - \bar{x})^2}{\sum f_x (x_i - \bar{x})^2}} \quad (138)$$

Здесь y и x общие, а y_x и x_y – групповые средние арифметические f_x и f_y – частоты рядов x и y . Следовательно, чтобы вычислить корреляционное отношение необходимо:

1. Сгруппировать первичные данные в форме корреляционной таблицы.
2. Определить общие и групповые средние арифметические.
3. Возвести в квадрат отклонения групповых средних от общей средней данного ряда распределения, умножить на соответствующие частоты и результаты сложить.
4. Умножить суммы квадратов отклонений классовых вариантов от их средних на частоты этих отклонений и результаты сложить.
5. Подставить полученные данные в формулу и рассчитать корреляционное отношение.

8.5. Коэффициент детерминации

Показывают, какая доля вариации одного признака зависит от варьирования другого признака. При линейной связи коэффициентом детерминации служит квадрат коэффициента корреляции r^2_{xy} , а при нелинейной зависимости между признаками X и Y – квадрат корреляционного отношения h^2_{yx} . Коэффициенты детерминации позволяют построить примерную шкалу тесноты связи между признаками

- $R < 0.5$ – связь слабая; $R = 0.5 - 0.6$ – связь средняя
 $R \geq 0.7$ – связь сильная.

Вопросы к главе 8.

1. В чём отличие корреляционных связей от функциональных?
2. Что такое ковариация?
3. Что собой представляет коэффициент корреляции?
4. Каковы пределы изменения коэффициента корреляции?
5. Как оценивается статистическая значимость коэффициента корреляции?
6. Что показывает коэффициент детерминации?

СПИСОК ЛИТЕРАТУРЫ

1. Бард Й. Нелинейное оценивание параметров./ Пер. с англ. В.С. Дуженко и Е.С. Фоминой; Под ред. и с предисл. В.Г. Горского – М.: Статистика, 1979. – 349 с., ил. – (Математико-статистические методы за рубежом).

2. Венецкий И.Г., Венецкая В.И. Основные математико-статистические понятия и формулы в экономическом анализе. Справочник. - 2-е изд., перераб. и доп. – М.: Статистика, 1979.-47., ил. - (Мат статистика для экономистов).
3. Гмурман В.Е. Теория вероятностей и математическая статистика. Учеб. пособие для вузов. Изд.6-е, стер. – М.: Высш.шк.,1997. – 479.: ил.
4. Гусаров В.М. Статистика. Учебное пособие для студентов вузов. – М.: ЮНИТИ, 2001 – 463 с.
5. Демиденко Е.З. Линейная и нелинейная регрессии. – М.: Финансы и статистика, 1981. – 302 с.
6. Длин А.М. Математическая статистика в технике: Учеб. пособ. для высш. техн. учеб. завед. – 3-е изд., перераб. – М.: Советская наука, 1958, 461 с. ил.
7. Дмитриев Е.А. Математическая статистика в почвоведении: Учебник. - М.: Изд-во МГУ, 1995.- 320 с.: ил.
8. Доспехов Б.А. Методика полевого опыта: (С основами статистической обработки результатов исследований).- Изд.4-е, перераб. и доп. - М.: Колос,1979.- 416 с., ил. - (Учебники и учеб. пособия для высш. с.-х. учеб. заведений).
9. Доугерти К. Введение в эконометрику: Пер. с англ. – М.: ИНФРА-М, 1999. – XIV, 402 с.
10. Дрейпер Н., Смит Г. Прикладной регрессионный анализ: В2-х кн. Кн. 1/. Пер. с англ. Ю.П. Адлера и В.Г. Горского – Изд. 2-е, перераб. и доп. – М.: Финансы и статистика. 1986. – 366 с.: ил. – (Математико-статистические методы за рубежом).
11. Дрейпер Н., Смит Г. Прикладной регрессионный анализ: В2-х кн. Кн. 2/. Пер. с англ. Ю.П. Адлера и В.Г. Горского – Изд.2-е, перераб. и доп. – М.: Финансы и статистика. , 1987. – 351 с.: ил. – (Математико-статистические методы за рубежом).
12. Елисеева И.И., Юзбашев М.М.: Учебник/ Под ред. Чл. - корр. РАН И.И. Елисеевой. – М.: Финансы и статистика, 1996. – 368 с.: ил.
13. Ефимова М.Р. и др. Общая теория статистики: Учебник для студентов вузов – 2-е изд., испр. и доп. – М.: ИНФРА – М, 2001. – 416 с.
14. Лакин Г.Ф. Лакин Г.Ф. Биометрия: Учеб. Пособие для биол. спец. вузов- 4-е изд., перераб. и доп. - М.: Высш.шк.,1990.- 352 с.: ил.
15. Практикум по теории статистики: Учеб. пособие / Под ред. Р.А. Шмойловой. – М.: Финансы и статистика, 2000. – 416 с.: ил.
16. Свалов Н.Н. Вариационная статистика. Учебное пособие для вузов. – М., «Лесная промышленность», 1977. 176 с. Табл. 49, ил. 12, библиогр. – 21 назв.
17. Снедекор ДЖ.У. Статистические методы в применении к исследованиям в сельском хозяйстве и биологии. Пер. с англ. В.Н. Перегудова. – М., Сельхозиздат, 1961. 502 с.
18. Статистика: Курс лекций/Харченко Л.П., Долженкова В.Г., Ионин В.Г. и др.; Под ред. к.э.н. В.Г. Ионина. – Новосибирск: Изд-во НГАЭиУ.: ИНФРА-М, 1999. – 310 с.
19. Теория статистики: Учебник / Под ред. Проф. Р.А. Шмойловой. – 4-е изд., доп. и перераб. – М.: Финансы и статистика, 2004. – 656 с.: ил.
20. Теория статистики: Учебное пособие для студентов вузов / Под ред. Г.Л. Громько. – М.: ИНФРА – М., 2000. – 414 с.
21. Тереньтьев П.В., Ростова Н.С.: Практикум по биометрии. – Л., Изд-во Ленингр. Ун-та, 1977. –152 с.
22. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере. – М.: ИНФРА – М, Финансы и статистика, 1995. – 384 с.
23. Шмойлова Р.А. /Отв. ред. Практикум по теории статистики. Учеб. пособие для студентов вузов. – М., 2000.

ПРИЛОЖЕНИЯ

БУКВЫ ЛАТИНСКОГО АЛФАВИТА

Начертание	Произношение	Начертание	Произношение
A a A a	а	N n N n	эн
B b B b	бэ	O o O o	о
C c C c	цэ	P p P p	пэ
D d D d	дэ	Q q Q q	ку
E e E e	е	R r R r	эр
F f F f	эф	S s S s	эс
G g G g	же	T t T t	тэ
H h H h	аш	U u U u	у
I i I i	и	V v V v	вэ
J j J j	йот	W w W w	дубль-вэ
K k K k	ка	X x X x	икс
L l L l	эль	Y y Y y	игрек
M m M m	эм	Z z Z z	зэт

БУКВЫ ГРЕЧЕСКОГО АЛФАВИТА

Начертание	Произношение	Начертание	Произношение	Начертание	Произношение
A α	альфа	I ι	йота	P ρ	ро
B β	бета	K κ	каппа	Σ σ	сигма
Γ γ	гамма	Λ λ	лямбда	Τ τ	тау
Δ δ	дельта	Μ μ	мю	Υ υ	ипсилон
E ε	эпсилон	Ν ν	ню	Φ φ	фи
Z ζ	дзета	Ξ ξ	кси	Χ χ	хи
Η η	эта	Ο ο	омикрон	Ψ ψ	пси
Θ θ	тэта	Π π	пи	Ω ω	омега

Таблица 1

Таблица случайных чисел

0	0,21132	53	0,30492	106	0,35188
1	0,26215	54	0,16348	107	0,26331
2	0,79253	55	0,75307	108	0,00486
3	0,28952	56	0,40643	109	0,80191
4	0,93648	57	0,73857	110	0,81044
5	0,93726	58	0,25217	111	0,75385
6	0,35606	59	0,83369	112	0,82524
7	0,16043	60	0,32764	113	0,54294
8	0,40480	61	0,62633	114	0,49654
9	0,74225	62	0,96292	115	0,17114
10	0,70183	63	0,34499	116	0,28722

11	0,41904	64	0,31622	117	0,34354
12	0,75691	65	0,48381	118	0,30080
13	0,00524	66	0,49887	119	0,59332
14	0,59544	67	0,42757	120	0,90642
15	0,51846	68	0,70032	121	0,40683
16	0,38344	69	0,07664	122	0,36385
17	0,30438	70	0,31314	123	0,34851
18	0,05253	71	0,47206	124	0,44847
19	0,16183	72	0,05804	125	0,18594
20	0,92500	73	0,42046	126	0,07630
21	0,46777	74	0,10886	127	0,01483
22	0,33873	75	0,11909	128	0,92900
23	0,30228	76	0,21753	129	0,38400
24	0,27223	77	0,78087	130	0,07881
25	0,57355	78	0,83914	131	0,42041
26	0,96965	79	0,25929	132	0,61363
27	0,14291	80	0,25690	133	0,95413
28	0,56575	81	0,67351	134	0,26198
29	0,94983	82	0,70712	135	0,64337
30	0,71092	83	0,03327	136	0,01799
31	0,13687	84	0,50427	137	0,09945
32	0,19618	85	0,86400	138	0,76643
33	0,17474	86	0,16592	139	0,01184
34	0,57817	87	0,83168	140	0,90770
35	0,98727	88	0,53778	141	0,27310
36	0,80415	89	0,36797	142	0,98280
37	0,07641	90	0,91867	143	0,10394
38	0,83702	91	0,25512	144	0,29839
39	0,64725	92	0,18555	145	0,17819
40	0,28029	93	0,45103	146	0,55171
41	0,73297	94	0,91849	147	0,74780
42	0,00309	95	0,31422	148	0,45567
43	0,31992	96	0,52570	149	0,76785
44	0,76521	97	0,62883	150	0,36943
45	0,47253	98	0,36850	151	0,88635
46	0,84203	99	0,02961	152	0,36378
47	0,45840	100	0,47643	153	0,76584
48	0,64955	101	0,93607	154	0,66698
49	0,87323	102	0,48024	155	0,02154
50	0,74374	103	0,87140		
51	0,21248	104	0,56047		
52	0,47449	105	0,16733		

Таблица 2

Таблица значений локальной функции Лапласа

	0	1	2	3	4	5	6	7	8	9
--	----------	----------	----------	----------	----------	----------	----------	----------	----------	----------

0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3726	0,3712	0,3698
0,4	0,3683	0,3668	0,3652	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0395	0,0387	0,0379	0,0371	0,0363
2,2	0,0353	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001

Таблица 3

Значения стандартной нормальной функции распределения

	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
--	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------

0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Критические значения критерия Стьюдента

Число степеней свободы	Уровни статистической значимости			Число степеней свободы	Уровни статистической значимости		
	0,05	0,01	0,001		0,05	0,01	0,001
1	12,71	63,69	636,6	21	2,08	2,83	3,82
2	4,3	9,93	31,6	22	2,07	2,82	3,79
3	3,18	5,84	12,94	23	2,07	2,81	3,77
4	2,78	4,6	8,61	24	2,06	2,8	3,75
5	2,57	4,03	6,87	25	2,06	2,79	3,73
6	2,45	3,71	5,96	26	2,06	2,78	3,71
7	2,37	3,5	5,41	27	2,05	2,77	3,69
8	2,31	3,36	5,04	28	2,05	2,76	3,67
9	2,26	3,25	4,78	29	2,04	2,76	3,66
10	2,23	3,17	4,59	30	2,04	2,75	3,65
11	2,2	3,11	4,44	40	2,02	2,7	3,55
12	2,18	3,06	4,32	50	2,01	2,68	3,5
13	2,16	3,01	4,22	60	2	2,66	3,46
14	2,15	2,98	4,14	80	1,99	2,64	3,42
15	2,13	2,95	4,07	100	1,98	2,63	3,39
16	2,12	2,92	4,02	120	1,98	2,62	3,37
17	2,11	2,9	3,97	200	1,97	2,6	3,34
18	2,1	2,88	3,92	500	1,96	2,59	3,31
19	2,09	2,86	3,88	∞	1,96	2,58	3,29
20	2,09	2,85	3,85				

Таблица 5

Критические значения критерия F Фишера для уровней статистической значимости $p \leq 0,05$ и $p \leq 0,01$: df_1 - число степеней свободы в числителе, df_2 - число степеней свободы в знаменателе (начало)

df_1	1	2	3	4	5	6	7	8	9	10	11	12
df_2	$p \leq 0,05$											
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,37	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,27	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
	$p \leq 0,01$											
1	4052	4999	5403	5625	5764	5859	5928	5981	6022	6056	6082	6106
2	98,49	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40	99,41	99,42
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,29	10,15	10,05	9,96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	9,65	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55

Таблица 5 (продолжение)

df ₁	1	2	3	4	5	6	7	8	9	10	11	12
df ₂	p ≤ 0,05											
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34
19	4,38	3,52	3,13	2,90	2,74	2,63	2,55	2,48	2,43	2,38	2,34	2,31
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,35	2,31	2,28
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25
22	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40	2,35	2,30	2,26	2,23
23	4,28	3,42	3,03	2,80	2,64	2,53	2,45	2,38	2,32	2,28	2,24	2,20
24	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	2,30	2,26	2,22	2,18
25	4,24	3,38	2,99	2,76	2,60	2,49	2,41	2,34	2,28	2,24	2,20	2,16
26	4,22	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,30	2,25	2,20	2,16	2,13
28	4,20	3,34	2,95	2,71	2,56	2,44	2,36	2,29	2,24	2,19	2,15	2,12
29	4,18	3,33	2,93	2,70	2,54	2,43	2,35	2,28	2,22	2,18	2,14	2,10
30	4,17	3,32	2,92	2,69	2,53	2,42	2,34	2,27	2,21	2,16	2,12	2,09
32	4,15	3,30	2,90	2,67	2,51	2,40	2,32	2,25	2,19	2,14	2,10	2,07
34	4,13	3,28	2,88	2,65	2,49	2,38	2,30	2,23	2,17	2,12	2,08	2,05
	p ≤ 0,01											
17	8,40	6,11	5,18	4,67	4,34	4,10	3,39	3,79	3,68	3,59	3,52	3,45
18	8,28	6,01	5,09	4,58	4,25	4,01	3,85	3,71	3,60	3,51	3,44	3,37
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36	3,30
20	8,10	5,85	4,94	4,43	4,10	3,87	3,71	3,56	3,45	3,37	3,30	3,23
21	8,02	5,78	4,87	4,37	4,04	3,81	3,65	3,51	3,40	3,31	3,24	3,17
22	7,94	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,18	3,12
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,14	3,07
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,25	3,17	3,09	3,03
25	7,77	5,57	4,68	4,18	3,86	3,63	3,46	3,32	3,21	3,13	3,05	2,99
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,17	3,09	3,02	2,96
27	7,68	5,49	4,60	4,11	3,79	3,56	3,39	3,26	3,14	3,06	2,98	2,93
28	7,64	5,45	4,57	4,07	3,76	3,53	3,36	3,23	3,11	3,03	2,95	2,90
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,08	3,00	2,92	2,87
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,06	2,98	2,90	2,84
32	7,50	5,34	4,46	3,97	3,66	3,42	3,25	3,12	3,01	2,94	2,86	2,80
34	7,44	5,29	4,42	3,93	3,61	3,38	3,21	3,08	2,97	2,89	2,82	2,76

Таблица 5 (продолжение)

df ₁	1	2	3	4	5	6	7	8	9	10	11	12
df ₂	p≤0,05											
36	4,11	3,26	2,86	2,63	2,48	2,36	2,28	2,21	2,15	2,10	2,06	2,03
38	4,10	3,25	2,85	2,62	2,46	2,35	2,26	2,19	2,14	2,09	2,05	2,02
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,07	2,04	2,00
42	4,07	3,22	2,83	2,59	2,44	2,32	2,24	2,17	2,11	2,06	2,02	1,99
44	4,06	3,21	2,82	2,58	2,43	2,31	2,23	2,16	2,10	2,05	2,01	1,98
46	4,05	3,20	2,81	2,57	2,42	2,30	2,22	2,14	2,09	2,04	2,00	1,97
48	4,04	3,19	2,80	2,56	2,41	2,30	2,21	2,14	2,08	2,03	1,99	1,96
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,02	1,98	1,95
55	4,02	3,17	2,78	2,54	2,38	2,27	2,18	2,11	2,05	2,00	1,97	1,93
60	4,00	3,15	2,76	2,52	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92
65	3,99	3,14	2,75	2,51	2,36	2,24	2,15	2,08	2,02	1,98	1,94	1,90
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,01	1,97	1,93	1,89
80	3,98	3,11	2,72	2,48	2,33	2,21	2,12	2,05	1,99	1,95	1,91	1,88
100	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	1,97	1,92	1,88	1,85
125	3,92	3,07	2,68	2,44	2,29	2,17	2,08	2,01	1,95	1,90	1,86	1,83
150	3,91	3,06	2,67	2,43	2,27	2,16	2,07	2,00	1,94	1,89	1,85	1,82
200	3,89	3,04	2,65	2,41	2,26	2,14	2,05	1,98	1,92	1,87	1,83	1,80
400	3,86	3,02	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,81	1,78
1000	3,85	3,00	2,61	2,38	2,22	2,10	2,02	1,95	1,89	1,84	1,80	1,76
¥	3,84	2,99	2,60	2,37	2,21	2,09	2,01	1,94	1,88	1,83	1,79	1,75
	p≤0,01											
36	7,39	5,25	4,38	3,89	3,58	3,35	3,18	3,04	2,94	2,86	2,78	2,72
38	7,35	5,21	4,34	3,86	3,54	3,32	3,15	3,02	2,91	2,82	2,75	2,69
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,88	2,8	2,73	2,66
42	7,27	5,15	4,29	3,80	3,49	3,26	3,1	2,96	2,86	2,77	2,7	2,64
44	7,24	5,12	4,26	3,78	3,46	3,24	3,07	2,94	2,84	2,75	2,68	2,62
46	7,21	5,10	4,24	3,76	3,44	3,22	3,05	2,92	2,82	2,73	2,66	2,6
48	7,19	5,08	4,22	3,74	3,42	3,2	3,04	2,9	2,8	2,71	2,64	2,58
50	7,17	5,06	4,20	3,72	3,41	3,18	3,02	2,88	2,78	2,7	2,62	2,56
55	7,12	5,01	4,16	3,68	3,37	3,15	2,98	2,85	2,75	2,66	2,59	2,53
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,56	2,5
65	7,04	4,95	4,10	3,62	3,31	3,09	2,93	2,79	2,7	2,61	2,54	2,47
70	7,01	4,92	4,08	3,60	3,29	3,07	2,91	2,77	2,67	2,59	2,51	2,45
80	6,96	4,88	4,04	3,56	3,25	3,04	2,87	2,74	2,64	2,55	2,48	2,41
100	6,90	4,82	3,98	3,51	3,2	2,99	2,82	2,69	2,59	2,51	2,43	2,36
125	6,84	4,78	3,94	3,47	3,17	2,95	2,79	2,65	2,56	2,47	2,4	2,33
150	6,81	4,75	3,91	3,44	3,14	2,92	2,76	2,62	2,53	2,44	2,37	2,3
200	6,76	4,71	3,88	3,41	3,11	2,9	2,73	2,6	2,5	2,41	2,34	2,28
400	6,70	4,66	3,83	3,36	3,06	2,85	2,69	2,55	2,46	2,37	2,29	2,23
1000	6,66	4,62	3,80	3,34	3,04	2,82	2,66	2,53	2,43	2,34	2,26	2,2
¥	6,64	4,60	3,78	3,32	3,02	2,8	2,64	2,51	2,41	2,32	2,24	2,18

Таблица 5 (продолжение)

df ₁	14	16	20	24	30	40	50	75	100	200	500	∞
df ₂	p ≤ 0,05											
1	245	246	248	249	250	251	252	253	253	254	254	254
2	19,42	19,43	19,44	19,45	19,46	19,47	19,47	19,48	19,49	19,49	19,50	19,50
3	8,71	8,69	8,66	8,64	8,62	8,60	8,58	8,57	8,56	8,54	8,54	8,53
4	5,87	5,84	5,80	5,77	5,74	5,71	5,70	5,68	5,66	5,65	5,64	5,63
5	4,64	4,60	4,56	4,53	4,50	4,46	4,44	4,42	4,40	4,38	4,37	4,36
6	3,96	3,92	3,87	3,84	3,81	3,77	3,75	3,72	3,71	3,69	3,68	3,67
7	3,52	3,49	3,44	3,41	3,38	3,34	3,32	3,29	3,28	3,25	3,24	3,23
8	3,23	3,20	3,15	3,12	3,08	3,05	3,03	3,00	2,98	2,96	2,94	2,93
9	3,02	2,98	2,93	2,90	2,86	2,82	2,89	2,77	2,76	2,73	2,72	2,71
10	2,86	2,82	2,77	2,74	2,70	2,67	2,64	2,61	2,59	2,56	2,55	2,54
11	2,74	2,70	2,65	2,61	2,57	2,53	2,50	2,47	2,45	2,42	2,41	2,40
12	2,64	2,60	2,54	2,50	2,46	2,42	2,40	2,36	2,35	2,32	2,31	2,30
13	2,55	2,51	2,46	2,42	2,38	2,34	2,32	2,28	2,26	2,24	2,22	2,21
14	2,48	2,44	2,39	2,35	2,31	2,27	2,24	2,21	2,19	2,16	2,14	2,13
15	2,43	2,39	2,33	2,29	2,25	2,21	2,18	2,15	2,12	2,10	2,08	2,07
16	2,37	2,33	2,28	2,24	2,20	2,16	2,13	2,09	2,07	2,04	2,02	2,01
	p ≤ 0,01											
1	6142	6169	6208	6234	6261	6286	6302	6323	6334	6352	6361	6366
2	99,43	99,44	99,45	99,46	99,47	99,48	99,48	99,40	99,49	99,49	99,50	99,50
3	26,92	26,83	26,69	26,60	26,50	26,41	26,35	26,27	26,23	26,18	26,14	26,12
4	14,14	14,15	14,02	13,93	13,83	13,74	13,69	13,61	13,57	13,52	13,48	13,46
5	9,77	9,68	9,55	9,47	9,38	9,29	9,24	9,17	9,13	9,07	9,04	9,02
6	7,60	7,52	7,39	7,31	7,23	7,14	7,09	7,02	6,99	6,94	6,90	6,88
7	6,35	6,27	6,15	6,07	5,98	5,90	5,85	5,78	5,75	5,70	5,67	5,65
8	5,56	5,48	5,36	5,28	5,20	5,11	5,06	5,00	4,96	4,91	4,88	4,86
9	5,00	4,92	4,80	4,73	4,64	4,56	4,51	4,45	4,41	4,36	4,33	4,31
10	4,60	4,52	4,41	4,33	4,25	4,17	4,12	4,05	4,01	3,96	3,93	3,91
11	4,29	4,21	4,10	4,02	3,94	3,86	3,80	3,74	3,70	3,66	3,62	3,60
12	4,05	3,98	3,86	3,78	3,70	3,61	3,56	3,49	3,46	3,41	3,38	3,36
13	3,85	3,78	3,67	3,59	3,51	3,42	3,37	3,30	3,27	3,21	3,18	3,16
14	3,70	3,62	3,51	3,43	3,34	3,26	3,21	3,14	3,11	3,06	3,02	3,00
15	3,56	3,48	3,36	3,29	3,20	3,12	3,07	3,00	2,97	2,92	2,89	2,87

Таблица 5 (продолжение)

df ₁	14	16	20	24	30	40	50	75	100	200	500	∞
df ₂	p ≤ 0,05											
17	2,33	2,29	2,23	2,19	2,15	2,11	2,08	2,04	2,02	1,99	1,97	1,96
18	2,29	2,25	2,19	2,15	2,11	2,07	2,04	2,00	1,98	1,95	1,93	1,92
19	2,26	2,21	2,15	2,11	2,07	2,02	2,00	1,96	1,94	1,91	1,90	1,88
20	2,23	2,18	2,12	2,08	2,04	1,99	1,96	1,92	1,90	1,87	1,85	1,84
21	2,20	2,15	2,09	2,05	2,00	1,96	1,93	1,89	1,87	1,84	1,82	1,81
22	2,18	2,13	2,07	2,03	1,98	1,93	1,91	1,87	1,84	1,81	1,80	1,78
23	2,14	2,10	2,04	2,00	1,96	1,91	1,88	1,84	1,82	1,79	1,77	1,76
24	2,13	2,09	2,02	1,98	1,94	1,89	1,86	1,82	1,80	1,76	1,74	1,73
25	2,11	2,06	2,00	1,96	1,92	1,87	1,84	1,80	1,77	1,74	1,72	1,71
26	2,10	2,05	1,99	1,95	1,90	1,85	1,82	1,78	1,76	1,72	1,70	1,69
27	2,08	2,03	1,97	1,93	1,88	1,84	1,80	1,76	1,74	1,71	1,68	1,67
28	2,06	2,02	1,96	1,91	1,87	1,81	1,78	1,75	1,72	1,69	1,67	1,65
29	2,05	2,00	1,94	1,90	1,85	1,80	1,77	1,73	1,71	1,68	1,65	1,64
30	2,04	1,99	1,93	1,89	1,84	1,79	1,76	1,72	1,69	1,66	1,64	1,62
32	2,02	1,97	1,91	1,86	1,82	1,76	1,74	1,69	1,67	1,64	1,61	1,59
34	2,00	1,95	1,89	1,84	1,80	1,74	1,71	1,67	1,64	1,61	1,59	1,57
	p ≤ 0,01											
17	3,35	3,27	3,16	3,08	3,00	2,92	2,86	2,79	2,76	2,70	2,67	2,65
18	3,27	3,19	3,07	3,00	2,91	2,83	2,78	2,71	2,68	2,62	2,59	2,57
19	3,19	3,12	3,00	2,92	2,84	2,76	2,70	2,63	2,60	2,54	2,51	2,49
20	3,13	3,05	2,94	2,86	2,77	2,69	2,63	2,56	2,53	2,47	2,44	2,42
21	3,07	2,99	2,88	2,80	2,72	2,63	2,58	2,51	2,47	2,42	2,38	2,36
22	3,02	2,94	2,83	2,75	2,67	2,58	2,53	2,46	2,42	2,37	2,33	2,31
23	2,97	2,89	2,78	2,70	2,62	2,53	2,48	2,41	2,37	2,32	2,28	2,26
24	2,93	2,85	2,74	2,66	2,58	2,49	2,44	2,36	2,33	2,27	2,23	2,21
25	2,89	2,81	2,70	2,62	2,54	2,45	2,40	2,32	2,29	2,23	2,19	2,17
26	2,86	2,77	2,66	2,58	2,50	2,41	2,36	2,28	2,25	2,19	2,15	2,13
27	2,83	2,74	2,63	2,55	2,47	2,38	2,33	2,25	2,21	2,16	2,12	2,10
28	2,80	2,71	2,60	2,52	2,44	2,35	2,30	2,22	2,18	2,13	2,09	2,06
29	2,77	2,68	2,57	2,49	2,41	2,32	2,27	2,19	2,15	2,10	2,06	2,03
30	2,74	2,66	2,55	2,47	2,38	2,29	2,24	2,16	2,13	2,07	2,03	2,01
32	2,70	2,62	2,51	2,42	2,34	2,25	2,20	2,12	2,08	2,02	1,98	1,96
34	2,66	2,58	2,47	2,38	2,30	2,21	2,15	2,08	2,04	1,98	1,94	1,91

Таблица 5. (окончание)

df ₁	14	16	20	24	30	40	50	75	100	200	500	∞
df ₂	p≤0,05											
36	1,98	1,93	1,87	1,82	1,78	1,72	1,69	1,65	1,62	1,59	1,56	1,55
38	1,96	1,92	1,85	1,80	1,76	1,71	1,67	1,63	1,60	1,57	1,54	1,53
40	1,95	1,90	1,84	1,79	1,74	1,69	1,66	1,61	1,59	1,55	1,53	1,51
42	1,94	1,89	1,82	1,78	1,73	1,68	1,64	1,60	1,57	1,54	1,51	1,49
44	1,92	1,88	1,81	1,76	1,72	1,66	1,63	1,58	1,56	1,52	1,50	1,48
46	1,91	1,87	1,80	1,75	1,71	1,65	1,62	1,57	1,54	1,51	1,48	1,46
48	1,90	1,86	1,79	1,74	1,70	1,64	1,61	1,56	1,53	1,50	1,47	1,45
50	1,90	1,85	1,78	1,74	1,69	1,63	1,60	1,55	1,52	1,48	1,46	1,44
55	1,88	1,83	1,76	1,72	1,67	1,61	1,58	1,52	1,50	1,46	1,43	1,41
60	1,86	1,81	1,75	1,70	1,65	1,59	1,56	1,50	1,48	1,44	1,41	1,39
65	1,85	1,80	1,73	1,68	1,63	1,57	1,54	1,49	1,46	1,42	1,39	1,37
70	1,84	1,79	1,72	1,67	1,62	1,56	1,53	1,47	1,45	1,40	1,37	1,35
80	1,82	1,77	1,70	1,65	1,60	1,54	1,51	1,45	1,42	1,38	1,35	1,32
100	1,79	1,75	1,68	1,63	1,57	1,51	1,48	1,42	1,39	1,34	1,30	1,28
125	1,77	1,72	1,65	1,60	1,55	1,49	1,45	1,39	1,36	1,31	1,27	1,25
150	1,76	1,71	1,64	1,59	1,54	1,47	1,44	1,37	1,34	1,29	1,25	1,22
200	1,74	1,69	1,62	1,57	1,52	1,45	1,42	1,35	1,32	1,26	1,22	1,19
400	1,72	1,67	1,60	1,54	1,49	1,42	1,38	1,32	1,28	1,22	1,16	1,13
1000	1,70	1,65	1,58	1,53	1,47	1,41	1,36	1,30	1,26	1,19	1,13	1,08
∞	1,69	1,64	1,57	1,52	1,46	1,40	1,35	1,28	1,24	1,17	1,11	1,00
	p≤0,01											
36	2,62	2,54	2,43	2,35	2,26	2,17	2,12	2,04	2,00	1,94	1,90	1,87
38	2,59	2,51	2,40	2,32	2,22	2,14	2,08	2,00	1,97	1,90	1,86	1,84
40	2,56	2,49	2,37	2,29	2,20	2,11	2,05	1,97	1,94	1,88	1,84	1,81
42	2,54	2,46	2,35	2,26	2,17	2,08	2,02	1,94	1,91	1,85	1,80	1,78
44	2,52	2,44	2,32	2,24	2,15	2,06	2,00	1,92	1,88	1,82	1,78	1,75
46	2,50	2,42	2,30	2,22	2,13	2,04	1,98	1,90	1,86	1,80	1,76	1,72
48	2,48	2,40	2,28	2,20	2,11	2,02	1,96	1,88	1,84	1,78	1,73	1,70
50	2,46	2,39	2,26	2,18	2,10	2,00	1,94	1,86	1,82	1,76	1,71	1,68
55	2,43	2,35	2,23	2,15	2,06	1,96	1,90	1,82	1,78	1,71	1,66	1,64
60	2,40	2,32	2,20	2,12	2,03	1,93	1,87	1,79	1,74	1,68	1,63	1,60
65	2,37	2,30	2,18	2,09	2,00	1,90	1,84	1,76	1,71	1,64	1,60	1,56
70	2,35	2,28	2,15	2,07	1,98	1,88	1,82	1,74	1,69	1,62	1,56	1,53
80	2,32	2,24	2,11	2,03	1,94	1,84	1,78	1,70	1,65	1,57	1,52	1,49
100	2,26	2,19	2,06	1,98	1,89	1,79	1,73	1,64	1,59	1,51	1,46	1,43
125	2,23	2,15	2,03	1,94	1,85	1,75	1,68	1,59	1,54	1,46	1,40	1,37
150	2,20	2,12	2,00	1,91	1,83	1,72	1,66	1,56	1,51	1,43	1,37	1,33
200	2,17	2,09	1,97	1,88	1,79	1,69	1,62	1,53	1,48	1,39	1,33	1,28
400	2,12	2,04	1,92	1,84	1,74	1,64	1,57	1,47	1,42	1,32	1,24	1,19
1000	2,09	2,01	1,89	1,81	1,71	1,61	1,54	1,44	1,38	1,28	1,19	1,11
∞	2,07	1,99	1,87	1,79	1,69	1,59	1,52	1,41	1,36	1,25	1,15	1,00

Критические значения критерий Манна-Уитни (начало)

Критические значения для 5% ошибки																			
N1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
N2																			
3	...	0																	
4	...	0	1																
5	0	1	2	4															
6	0	2	3	5	7														
7	0	2	4	6	8	11													
8	1	3	5	8	10	13	15												
9	1	4	6	9	12	15	18	21											
10	1	4	7	11	14	17	20	24	27										
11	1	5	8	12	16	19	23	27	31	34									
12	2	5	9	13	17	21	26	30	34	38	42								
13	2	6	10	15	19	24	28	33	37	42	47	51							
14	3	7	11	16	21	26	31	36	41	46	51	56	61						
15	3	7	12	18	23	28	33	39	44	50	55	61	66	72					
16	3	8	14	19	25	30	36	42	48	54	60	65	71	77	83				
17	3	9	15	20	26	33	39	45	51	57	64	70	77	83	89	96			
18	4	9	16	22	28	35	41	48	55	61	68	75	82	88	95	102	109		
19	4	10	17	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123	
20	4	11	18	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138
Критические значения для 1% ошибки																			
5	0	1															
6	1	2	3														
7	...	0	1	3	4	6													
8	...	0	2	4	6	7	9												
9	...	1	3	5	7	9	11	14											
10	...	1	3	6	8	11	13	16	19										
11	...	1	4	7	9	12	15	18	22	25									
12	...	2	5	8	11	14	17	21	24	28	31								
13	0	2	5	9	12	16	20	23	27	31	35	39							
14	0	2	6	10	13	17	22	26	30	34	38	43	47						
15	0	3	7	11	15	19	24	28	33	37	42	47	51	56					
16	0	3	7	12	16	21	26	31	36	41	46	51	56	61	66				
17	0	4	8	13	18	23	28	33	38	44	49	55	60	66	71	77			
18	0	4	9	14	19	24	30	36	41	47	53	59	65	70	76	82	88		
19	1	4	9	15	20	26	32	38	44	50	56	63	69	75	82	88	94	101	
20	1	5	10	16	22	28	34	40	47	53	60	67	73	80	87	93	100	107	114

Таблица 6 (продолжение)

Критические значения для 5% ошибки																		
N1	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
N2																		
21	19	26	34	41	49	57	65	73	81	89	97	105	113	121	130	138	146	154
22	20	28	36	44	52	60	69	77	85	94	102	111	119	128	136	145	154	162
23	21	29	37	46	55	63	72	81	90	99	107	116	125	134	143	152	161	170
24	22	31	39	48	57	66	75	85	94	103	113	122	131	141	150	160	169	179
25	23	32	41	50	60	69	79	89	98	108	118	128	137	147	157	167	177	187
26	24	33	43	53	62	72	82	93	103	113	123	133	143	154	164	174	185	195
27	25	35	45	55	65	75	86	96	107	118	128	139	150	160	171	182	193	203
28	26	36	47	57	68	79	89	100	111	122	133	144	156	167	178	189	200	212
29	27	38	48	59	70	82	93	104	116	127	139	150	162	173	185	196	208	220
30	28	39	50	62	73	85	96	108	120	132	144	156	168	180	192	204	216	228
31	29	41	52	64	76	88	100	112	124	137	149	161	174	186	199	211	224	236
32	30	42	54	66	78	91	103	116	129	141	154	167	180	193	206	219	232	245
33	31	43	56	68	81	94	107	120	133	146	159	173	186	199	213	226	239	253
34	32	45	58	71	84	97	110	124	137	151	164	178	192	206	219	233	247	261
35	33	46	59	73	86	100	114	128	142	156	170	184	198	212	226	241	255	269
36	35	48	61	75	89	103	117	132	146	160	175	189	204	219	233	248	263	278
37	36	49	63	77	92	106	121	135	150	165	180	195	210	225	240	255	271	286
38	37	51	65	79	94	109	124	139	155	170	185	201	216	232	247	263	278	294
39	38	52	67	82	97	112	128	143	159	175	190	206	222	238	254	270	286	302
40	39	53	69	84	100	115	131	147	163	179	196	212	228	245	261	278	294	311
Критические значения для 1% ошибки																		
21	10	16	22	29	35	42	49	56	63	70	77	84	91	98	105	113	120	127
22	10	17	23	30	37	45	52	59	66	74	81	89	96	104	111	119	127	134
23	11	18	25	32	39	47	55	62	70	78	86	94	102	109	117	125	133	141
24	12	19	26	34	42	49	57	66	74	82	90	98	107	115	123	132	140	149
25	12	20	27	35	44	52	60	69	77	86	95	103	112	121	130	138	147	156
26	13	21	29	37	46	54	63	72	81	90	99	108	117	126	136	145	154	163
27	14	22	30	39	48	57	66	75	85	94	103	113	122	132	142	151	161	171
28	14	23	32	41	50	59	69	78	88	98	108	118	128	138	148	158	168	178
29	15	24	33	42	52	62	72	82	92	102	112	123	133	143	154	164	175	185
30	15	25	34	44	54	64	75	85	95	106	117	127	138	149	160	171	182	192
31	16	26	36	46	56	67	77	88	99	110	121	131	143	155	166	177	188	200
32	17	27	37	47	58	69	80	91	103	114	126	137	149	160	172	184	195	207
33	17	28	38	49	60	72	83	95	106	118	130	142	154	166	178	190	202	214
34	18	29	40	51	62	74	86	98	110	122	134	147	159	172	184	197	209	222
35	19	30	41	53	64	77	89	101	114	126	139	152	164	177	190	203	216	229
36	19	31	42	54	67	79	92	104	117	130	143	156	170	183	196	210	223	236
37	20	32	44	56	69	81	95	108	121	134	148	161	175	189	202	216	230	244
38	21	33	45	58	71	84	97	111	125	138	152	166	180	194	208	223	237	251
39	21	34	46	59	73	86	100	114	128	142	157	171	185	200	214	229	244	258
40	22	35	48	61	75	89	103	117	132	146	161	176	191	206	221	236	251	266

Таблица 6 (окончание)

Критические значения для 5% ошибки																			
N1	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
N2																			
21																			
22	171																		
23	180	189																	
24	188	198	207																
25	197	207	217	227															
26	206	216	226	237	247														
27	214	225	236	247	258	268													
28	223	234	245	257	268	279	291												
29	232	243	255	267	278	290	302	314											
30	240	252	265	277	289	301	313	326	338										
31	249	261	274	287	299	312	325	337	350	363									
32	258	271	284	297	310	323	336	349	362	375	389								
33	266	280	293	307	320	334	347	361	374	388	402	415							
34	275	289	303	317	331	345	359	373	387	401	415	429	443						
35	284	298	312	327	341	356	370	385	399	413	428	442	457	471					
36	292	307	322	337	352	367	381	396	411	426	441	456	471	486	501				
37	301	316	332	347	362	378	393	408	424	439	454	470	485	501	516	531			
38	310	325	341	357	373	388	404	420	436	452	467	483	499	515	531	547	563		
39	318	335	351	367	383	399	416	432	448	464	481	497	513	530	546	562	579	595	
40	327	344	360	377	394	410	427	444	460	477	494	511	527	544	561	578	584	611	628
Критические значения для 1% ошибки																			
21																			
22	142																		
23	150	158																	
24	154	166	174																
25	165	174	183	192															
26	173	182	191	201	210														
27	180	190	200	209	219	229													
28	188	198	208	218	229	239	249												
29	196	206	217	227	238	249	259	270											
30	203	214	225	236	247	258	270	281	292										
31	211	223	234	245	257	268	280	291	303	314									
32	219	231	242	254	266	278	290	302	314	326	338								
33	227	239	251	263	276	288	300	313	325	337	350	362							
34	234	247	260	272	285	298	311	323	336	349	362	375	387						
35	242	255	268	281	294	308	321	334	347	360	374	387	400	413					
36	250	263	277	290	304	318	331	345	358	372	386	399	413	427	440				
37	258	271	285	299	313	327	341	355	370	384	398	412	426	440	454	468			
38	265	280	294	308	323	337	352	366	381	395	410	424	439	453	468	482	497		
39	173	288	303	317	332	347	362	377	392	407	422	437	452	467	482	497	512	527	
40	281	296	311	326	342	357	372	388	403	418	434	449	465	480	495	511	526	542	557

Таблица 7

**Критические значения статистики Уилка-Шапиро W_α
в зависимости от объема выборки n . Вероятность $P(W < W_\alpha) = \alpha$**

n	α			n	α		
	0,01	0,05	0,10		0,01	0,05	0,10
3	0,753	0,767	0,789	27	0,894	0,923	0,935
4	0,687	0,748	0,792	28	0,896	0,924	0,936
5	0,686	0,762	0,806	29	0,898	0,926	0,937
6	0,713	0,788	0,826	30	0,900	0,927	0,939
7	0,730	0,803	0,838	31	0,902	0,929	0,940
8	0,749	0,818	0,851	32	0,904	0,930	0,941
9	0,764	0,829	0,859	33	0,906	0,931	0,942
10	0,781	0,842	0,869	34	0,908	0,933	0,943
11	0,792	0,850	0,876	35	0,910	0,934	0,944
12	0,805	0,859	0,883	36	0,912	0,935	0,945
13	0,814	0,866	0,889	37	0,914	0,936	0,946
14	0,825	0,874	0,895	38	0,916	0,938	0,947
15	0,835	0,881	0,901	39	0,917	0,939	0,948
16	0,844	0,887	0,906	40	0,919	0,940	0,949
17	0,851	0,892	0,910	41	0,920	0,941	0,950
18	0,858	0,897	0,914	42	0,922	0,942	0,951
19	0,863	0,901	0,917	43	0,923	0,943	0,951
20	0,868	0,905	0,920	44	0,924	0,944	0,952

Таблица 8

Таблица критических точек распределения Пирсона «хи-квадрат»)

k / α	0,01	0,025	0,05	0,95	0,975	0,99
1	6,63490	5,02389	3,84146	0,00393	0,00098	0,00016
2	9,21034	7,37776	5,99146	0,10259	0,05064	0,02010
3	11,34487	9,34840	7,81473	0,35185	0,21580	0,11483
4	13,2767	11,14329	9,48773	0,71072	0,48442	0,29711
5	15,08627	12,8325	11,0705	1,14548	0,83121	0,55430
6	16,81189	14,44938	12,59159	1,63538	1,23734	0,87209
7	18,47531	16,01276	14,06714	2,16735	1,68987	1,23904
8	20,09024	17,53455	15,50731	2,73264	2,17973	1,64650
9	21,66599	19,02277	16,91898	3,32511	2,70039	2,08790
10	23,20925	20,48318	18,30704	3,94030	3,24697	2,55821
11	24,72497	21,92005	19,67514	4,57481	3,81575	3,05348
12	26,21697	23,33666	21,02607	5,22603	4,40379	3,57057
13	27,68825	24,7356	22,36203	5,89186	5,00875	4,10692
14	29,14124	26,11895	23,68479	6,57063	5,62873	4,66043
15	30,57791	27,48839	24,99579	7,26094	6,26214	5,22935
16	31,99993	28,84535	26,29623	7,96165	6,90766	5,81221
17	33,40866	30,19101	27,58711	8,67176	7,56419	6,40776

18	34,80531	31,52638	28,86930	9,39046	8,23075	7,01491
19	36,19087	32,85233	30,14353	10,11701	8,90652	7,63273
20	37,56623	34,16961	31,41043	10,85081	9,59078	8,26040
21	38,93217	35,47888	32,67057	11,59131	10,2829	8,89720
22	40,28936	36,78071	33,92444	12,33801	10,98232	9,54249
23	41,63840	38,07563	35,17246	13,09051	11,68855	10,19572
24	42,97982	39,36408	36,41503	13,84843	12,40115	10,85636
25	44,31410	40,64647	37,65248	14,61141	13,11972	11,52398
26	45,64168	41,92317	38,88514	15,37916	13,84391	12,19815
27	46,96294	43,19451	40,11327	16,15140	14,57338	12,87850
28	48,27824	44,46079	41,33714	16,92788	15,30786	13,56471
29	49,58788	45,72229	42,55697	17,70837	16,04707	14,25645
30	50,89218	46,97924	43,77297	18,49266	16,79077	14,95346
31	52,19139	48,23189	44,98534	19,28057	17,53874	15,65546
32	53,48577	49,48044	46,19426	20,07191	18,29076	16,36222
33	54,77554	50,72508	47,39988	20,86653	19,04666	17,07351
34	56,06091	51,96600	48,60237	21,66428	19,80625	17,78915
35	57,34207	53,20335	49,80185	22,46502	20,56938	18,50893
36	58,61921	54,43729	50,99846	23,26861	21,33588	19,23268
37	59,89250	55,66797	52,19232	24,07494	22,10563	19,96023
38	61,16209	56,89552	53,38354	24,8839	22,87848	20,69144
39	62,42812	58,12006	54,57223	25,69539	23,65432	21,42616
40	63,69074	59,34171	55,75848	26,5093	24,43304	22,16426
41	64,95007	60,56057	56,94239	27,32555	25,21452	22,90561
42	66,20624	61,77676	58,12404	28,14405	25,99866	23,65009
43	67,45935	62,99036	59,30351	28,96472	26,78537	24,39760
44	68,70951	64,20146	60,48089	29,78748	27,57457	25,14803
45	69,95683	65,41016	61,65623	30,61226	28,36615	25,90127
46	71,20140	66,61653	62,82962	31,43900	29,16005	26,65724
47	72,44331	67,82065	64,00111	32,26762	29,95620	27,41585
48	73,68264	69,02259	65,17077	33,09808	30,75451	28,17701
49	74,91947	70,22241	66,33865	33,93031	31,55492	28,94065
50	76,15389	71,42020	67,50481	34,76425	32,35736	29,70668

Значения функции $\Psi(R/(n+1))$

$\frac{R}{n+1}$	0	1	2	3	4	5	6	7	8	9
0,00	$-\infty$	-3,09	-2,88	-2,75	-2,65	-2,58	-2,51	-2,46	-2,41	-2,37
0,01	-2,53	-2,29	-2,26	-2,23	-2,20	-2,17	-2,14	-2,12	-2,10	-2,07
0,02	-2,05	-2,03	-2,01	-2,00	-1,98	-1,96	-1,94	-1,93	-1,91	-1,90
0,03	-1,88	-1,87	-1,85	-1,84	-1,83	-1,81	-1,80	-1,79	-1,77	-1,76
0,04	-0,75	-1,74	-1,73	-1,72	-1,71	-1,70	-1,68	-1,67	-1,66	-1,65
0,05	-1,64	-1,64	-1,63	-1,62	-1,61	-1,60	-1,59	-1,58	-1,57	-1,57
0,06	-1,55	-1,55	-1,54	-1,53	-1,52	-1,51	-1,51	-1,50	-1,49	-1,48
0,07	-1,48	-1,47	-1,46	-1,45	-1,45	-1,44	-1,43	-1,43	-1,42	-1,41
0,08	-1,41	-1,40	-1,39	-1,39	-1,38	-1,37	-1,37	-1,36	-1,35	-1,35
0,09	-1,34	-1,33	-1,33	-1,32	-1,32	-1,31	-1,30	-1,30	-1,29	-1,29
0,10	-1,28	-1,28	-1,27	-1,26	-1,26	-1,25	-1,25	-1,24	-1,24	-1,23
0,11	-1,23	-1,22	-1,22	-1,21	-1,21	-1,20	-1,20	-1,19	-1,19	-1,18
0,12	-1,18	-1,17	-1,17	-1,16	-1,16	-1,15	-1,15	-1,14	-1,14	-1,13
0,13	-1,13	-1,12	-1,12	-1,11	-1,11	-1,10	-1,10	-1,09	-1,09	-1,09
0,14	-1,08	-1,08	-1,07	-1,17	-1,06	-1,06	-1,05	-1,05	-1,05	-1,04
0,15	-1,04	-1,03	-1,03	-1,12	-1,02	-1,02	-1,01	-1,01	-1,01	-1,00
0,16	-0,99	-0,99	-0,99	-0,98	-0,98	-0,97	-0,97	-0,97	-0,96	-0,96
0,17	-0,95	-0,95	-0,95	-0,94	-0,94	-0,93	-0,93	-0,93	-0,92	-0,92
0,18	-0,92	-0,91	-0,91	-0,90	-0,90	-0,90	-0,89	-0,89	-0,89	-0,88
0,19	-0,88	-0,87	-0,87	-0,87	-0,86	-0,86	-0,86	-0,85	-0,85	-0,85
0,20	-0,84	-0,84	-0,83	-0,83	-0,83	-0,82	-0,82	-0,82	-0,81	-0,81
0,21	-0,81	-0,80	-0,80	-0,80	-0,79	-0,79	-0,79	-0,78	-0,78	-0,78
0,22	-0,77	-0,77	-0,77	-0,76	-0,76	-0,76	-0,75	-0,75	-0,75	-0,74
0,23	-0,74	-0,74	-0,73	-0,73	-0,73	-0,72	-0,72	-0,72	-0,71	-0,71
0,24	-0,71	-0,70	-0,70	-0,70	-0,69	-0,69	-0,69	-0,68	-0,68	-0,68
0,25	-0,67	-0,67	-0,67	-0,67	-0,66	-0,66	-0,66	-0,65	-0,65	-0,65
0,26	-0,64	-0,64	-0,64	-0,63	-0,63	-0,63	-0,63	-0,62	-0,62	-0,62
0,27	-0,61	-0,61	-0,61	-0,60	-0,60	-0,60	-0,60	-0,59	-0,59	-0,59
0,28	-0,58	-0,58	-0,58	-0,57	-0,57	-0,57	-0,57	-0,56	-0,56	-0,56
0,29	-0,55	-0,55	-0,55	-0,54	-0,54	-0,54	-0,54	-0,53	-0,53	-0,53
0,30	-0,53	-0,52	-0,52	-0,52	-0,51	-0,51	-0,51	-0,50	-0,50	-0,50
0,31	-0,50	-0,49	-0,49	-0,49	-0,48	-0,48	-0,48	-0,47	-0,47	-0,47
0,32	-0,47	-0,46	-0,46	-0,46	-0,46	-0,45	-0,45	-0,45	-0,45	-0,44
0,33	-0,44	-0,44	-0,43	-0,43	-0,43	-0,43	-0,43	-0,42	-0,42	-0,42
0,34	-0,41	-0,41	-0,41	-0,40	-0,40	-0,40	-0,40	-0,39	-0,39	-0,39
0,35	-0,39	-0,38	-0,38	-0,38	-0,37	-0,37	-0,37	-0,37	-0,36	-0,36
0,36	-0,36	-0,36	-0,35	-0,35	-0,35	-0,35	-0,34	-0,34	-0,34	-0,33
0,37	-0,33	-0,33	-0,33	-0,32	-0,32	-0,32	-0,32	-0,31	-0,31	-0,31
0,38	-0,31	-0,30	-0,30	-0,30	-0,30	-0,29	-0,29	-0,29	-0,28	-0,28
0,39	-0,28	-0,28	-0,27	-0,27	-0,27	-0,27	-0,26	-0,26	-0,26	-0,26
0,40	-0,25	-0,25	-0,25	-0,25	-0,24	-0,24	-0,24	-0,24	-0,23	-0,23

Таблица 9 окончание

$\frac{R}{n+1}$	0	1	2	3	4	5	6	7	8	9
0,41	-0,23	-0,23	-0,22	-0,22	-0,22	-0,21	-0,21	-0,21	-0,21	-0,20
0,42	-0,20	-0,20	-0,20	-0,19	-0,19	-0,19	-0,19	-0,18	-0,18	-0,18
0,43	-0,18	-0,17	-0,17	-0,17	-0,17	-0,16	-0,16	-0,16	-0,16	-0,15
0,44	-0,15	-0,15	-0,15	-0,14	-0,14	-0,14	-0,14	-0,13	-0,13	-0,13
0,45	-0,13	-0,12	-0,12	-0,12	-0,12	-0,11	-0,11	-0,11	-0,11	-0,10
0,46	-0,10	-0,10	-0,10	-0,09	-0,09	-0,09	-0,09	-0,08	-0,08	-0,08
0,47	-0,08	-0,07	-0,07	-0,07	-0,07	-0,06	-0,06	-0,06	-0,06	-0,05
0,48	-0,05	-0,05	-0,05	-0,04	-0,04	-0,04	-0,04	-0,03	-0,03	-0,03
0,49	-0,03	-0,02	-0,02	-0,02	-0,02	-0,01	-0,01	-0,01	-0,01	-0,00
0,50	+0,00	+0,00	+0,01	+0,01	+0,01	+0,01	+0,02	+0,02	+0,02	+0,02
0,51	0,03	0,03	0,03	0,03	0,04	0,04	0,04	0,04	0,05	0,05
0,52	0,05	0,05	0,06	0,06	0,06	0,06	0,07	0,07	0,07	0,07
0,53	0,08	0,08	0,08	0,08	0,09	0,09	0,09	0,09	0,10	0,10
0,54	0,10	0,10	0,11	0,11	0,11	0,11	0,12	0,12	0,12	0,12
0,55	0,13	0,13	0,13	0,13	0,14	0,14	0,14	0,14	0,15	0,15
0,56	0,15	0,15	0,16	0,16	0,16	0,16	0,17	0,17	0,17	0,17
0,57	0,18	0,18	0,18	0,18	0,19	0,19	0,19	0,19	0,20	0,20
0,58	0,20	0,20	0,21	0,21	0,21	0,21	0,22	0,22	0,22	0,23
0,59	0,23	0,23	0,23	0,24	0,24	0,24	0,24	0,25	0,25	0,25
0,60	0,25	0,26	0,26	0,26	0,26	0,27	0,27	0,27	0,27	0,28
0,61	0,28	0,28	0,28	0,29	0,29	0,29	0,30	0,30	0,30	0,30
0,62	0,31	0,31	0,31	0,31	0,32	0,32	0,32	0,32	0,33	0,33
0,63	0,33	0,33	0,34	0,34	0,34	0,35	0,35	0,35	0,35	0,36
0,64	0,36	0,36	0,36	0,37	0,37	0,37	0,37	0,38	0,38	0,38
0,65	0,39	0,39	0,39	0,39	0,40	0,40	0,40	0,40	0,41	0,41
0,66	0,41	0,42	0,42	0,42	0,42	0,43	0,43	0,43	0,43	0,44
0,67	0,44	0,44	0,45	0,45	0,45	0,45	0,46	0,46	0,46	0,46
0,68	0,47	0,47	0,47	0,48	0,48	0,48	0,48	0,49	0,49	0,49
0,69	0,50	0,50	0,50	0,50	0,51	0,51	0,51	0,52	0,52	0,52
0,70	0,52	0,53	0,53	0,53	0,54	0,54	0,54	0,54	0,55	0,55
0,71	0,55	0,56	0,56	0,56	0,57	0,57	0,57	0,57	0,58	0,58
0,72	0,58	0,59	0,59	0,59	0,59	0,60	0,60	0,60	0,61	0,61
0,73	0,61	0,62	0,62	0,62	0,63	0,63	0,63	0,63	0,64	0,64
0,74	0,64	0,65	0,65	0,65	0,66	0,66	0,66	0,67	0,67	0,67
0,75	0,67	0,68	0,68	0,68	0,69	0,69	0,69	0,70	0,70	0,70
0,76	0,71	0,71	0,71	0,72	0,72	0,72	0,73	0,73	0,73	0,74
0,77	0,74	0,74	0,75	0,75	0,75	0,76	0,76	0,76	0,77	0,77
0,78	0,77	0,78	0,78	0,78	0,79	0,79	0,79	0,80	0,80	0,80
0,79	0,81	0,81	0,81	0,82	0,82	0,82	0,83	0,83	0,83	0,84
0,80	0,84	0,85	0,85	0,85	0,86	0,86	0,86	0,87	0,87	0,87
0,81	0,88	0,88	0,89	0,89	0,89	0,90	0,90	0,90	0,91	0,91
0,82	0,92	0,92	0,92	0,93	0,93	0,93	0,94	0,94	0,95	0,95

Таблица 9 окончание										
0,83	0,95	0,96	0,96	0,97	0,97	0,97	0,98	0,98	0,99	0,99
0,84	0,99	1,00	1,00	1,01	1,01	1,02	1,02	1,02	1,03	1,03
0,85	1,04	1,04	1,05	1,05	1,05	1,06	1,06	1,07	1,07	1,08
0,86	1,08	1,09	1,09	1,09	1,10	1,10	1,11	1,11	1,12	1,12
0,87	1,13	1,13	1,14	1,14	1,15	1,15	1,16	1,16	1,17	1,17
0,88	1,18	1,18	1,19	1,19	1,20	1,20	1,21	1,21	1,22	1,22
0,89	1,23	1,23	1,24	1,24	1,25	1,25	1,26	1,26	1,27	1,28
0,90	1,28	1,29	1,29	1,30	1,30	1,31	1,32	1,32	1,33	1,33
0,91	1,34	1,35	1,35	1,36	1,37	1,38	1,37	1,39	1,39	1,40
0,92	1,41	1,41	1,42	1,43	1,43	1,44	1,45	1,45	1,46	1,47
0,93	1,48	1,48	1,49	1,50	1,51	1,51	1,52	1,53	1,54	1,55
0,94	1,55	1,56	1,57	1,58	1,59	1,60	1,61	1,62	1,63	1,64
0,95	1,64	1,65	1,66	1,67	1,68	1,70	1,71	1,72	1,73	1,74
0,96	1,75	1,76	1,77	1,79	1,80	1,81	1,83	1,84	1,85	1,87
0,97	1,88	1,90	1,91	1,93	1,94	1,96	1,98	2,00	2,01	2,03
0,98	2,05	2,07	2,10	2,12	2,14	2,17	2,20	2,23	2,26	2,29
0,99	2,33	2,37	2,41	2,46	2,51	2,58	2,65	2,75	2,88	3,09

Таблица 10

Критические значения Х-критерия Ван-дер-Вардена

n	n1-n2 = 0 или 1		Л,		n1-n2 = 2 или 3		n1-n2 = 4 или 5	
	Уровни значимости α %		Уровни значимости α %		Уровни значимости α %		Уровни значимости α %	
	5	1	5	1	5	1	5	1
8	2,40		2,30					
9	2,48	-	2,40	-				
10	2,60	3,20	2,49	3,10	2,30			
11	2,72	3,40	2,58	3,40	2,40			
12	2,86	3,60	2,79	3,58	2,68	3,40		
13	2,96	3,71	2,91	3,64	2,78	3,50		
14	3,11	3,94	3,06	3,88	3,00	3,76		
15	3,24	4,07	3,19	4,05	3,06	3,88		
16	3,39	4,26	3,36	4,25	3,28	4,12		
17	3,49	4,44	3,44	4,37	3,36	4,23		
18	3,63	4,60	3,60	4,58	3,53	4,50		
19	3,73	4,77	3,69	4,71	3,61	4,62		
20	3,86	4,94	3,84	4,92	3,78	4,85		
21	3,96	5,10	3,92	5,05	3,85	4,96		
22	4,08	5,26	4,06	5,24	4,01	5,17		
23	4,18	5,40	4,15	5,36	4,08	5,27		
24	4,29	5,55	4,27	5,53	4,23	5,48		
25	4,39	5,68	4,36	5,65	4,30	5,58		
26	4,50	5,83	4,48	5,81	4,44	5,76		
27	4,59	5,95	4,55	5,92	4,51	5,85		
28	4,68	6,09	4,68	6,07	4,64	6,03		
29	4,78	6,22	4,76	6,19	4,72	6,13		
30	4,88	6,35	4,87	6,34	4,84	6,30		

31	4,97	6,47	4,95	6,44	4,91	6,39
32	5,07	6,60	5,06	6,58	5,03	6,55
33	5,15	6,71	5,13	6,69	5,10	6,64
34	5,25	6,84	5,24	6,82	5,21	6,79
35	5,33	6,95	5,31	6,92	5,28	6,88
36	5,42	7,06	5,41	7,05	5,38	7,02
37	5,50	7,17	5,48	7,15	5,45	7,11
38	5,59	7,28	5,58	7,27	5,55	7,25
39	5,67	7,39	5,65	7,37	5,62	7,33
40	5,75	7,50	5,74	7,49	5,72	7,47
41	5,83	7,62	5,81	7,60	5,79	7,56
42	5,91	7,72	5,90	7,71	5,88	7,69
43	5,99	7,82	5,97	7,81	5,95	7,77
44	6,04	7,93	6,06	7,92	6,04	7,90
45	6,14	8,02	6,12	8,01	6,10	7,98
46	6,21	8,13	6,21	8,12	6,19	8,10
47	6,29	8,22	6,27	8,21	6,25	8,18
48	6,36	8,32	6,35	8,31	6,34	8,29
49	6,43	8,41	6,42	8,40	6,39	8,37
50	6,50	8,51	6,51	8,50	6,48	8,48

Таблица 11

**Критические значения U-критерия Уилкоксона (Манна-Уитни)
(односторонний критерий, P=0,01)**

n₁ \ n₂	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	n₂
3	0	0	0	0	1	1	1	2	2	2	3	3	4	4	4	5	3
4	0	1	1	9	3	3	4	5	5	6	7	7	8	9	9	10	4
5	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	10	5
6		3	4	6	7	8	9	11	12	14	15	16	18	19	20	22	6
7			6	7	9	11	12	14	16	18	19	21	23	24	26	28	7
8				9	11	13	15	17	20	22	24	26	28	30	32	34	8
9					14	16	19	21	23	26	28	31	33	36	38	48	9
10						19	22	24	27	30	33	36	38	41	44	47	10
11							25	28	31	34	37	41	44	47	50	53	11
12								31	35	38	42	46	49	53	56	60	12
13									39	43	47	51	55	59	63	67	13
14										47	51	56	60	65	69	73	14
15											56	61	66	70	75	80	15
16												66	71	76	82	87	16
17													77	82	88	94	17
18														88	94	100	18
19															101	107	19
20																114	20

Таблица 11 продолжение
(двусторонний критерий, P = 0,01)

n2 \ n1	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	n2
5	0	0	0																5
6	0	0	1	2															6
7	0	0	1	3	4														7
8	0	1	2	4	6	7													8
9	0	1	3	5	7	9	11												9
10	0	2	4	6	9	11	13	16											10
11	0	2	5	7	10	13	16	19	21										11
12	1	3	6	9	12	15	18	21	24	28									12
13	1	4	7	10	13	17	23	24	27	31	34								13
14	1	4	7	11	15	18	22	26	30	34	38	42							14
15	2	5	8	12	16	20	25	29	33	37	42	46	51						15
16	2	5	9	13	18	22	27	31	36	41	46	50	55	60					16
17	2	6	10	15	19	24	23	34	39	44	49	54	60	65	70				17
18	2	6	11	16	21	26	31	37	42	47	53	59	64	70	75	77			18
19	3	7	12	17	22	28	34	39	45	51	57	63	69	75	81	87	93		19
20	3	8	13	18	24	30	36	42	48	54	60	67	73	79	86	92	99	105	20
21	4	8	14	19	25	32	38	44	51	58	64	71	78	84	91	98	105	112	21
22	4	9	14	21	27	34	40	47	54	61	68	75	82	89	97	104	111	118	22
23	4	9	15	22	29	36	43	50	57	64	72	79	87	94	102	109	117	125	23
24	4	10	16	23	30	37	45	52	60	68	76	83	91	99	107	115	123	131	24
25	5	10	17	24	32	39	47	55	63	71	79	88	96	104	113	121	129	138	25

Таблица 12

**Критические значения z-критерия знаков
при разных уровнях значимости α и объеме выборки n**

n	$\alpha, \%$		n	$\alpha, \%$		n	$\alpha, \%$		n	$\alpha, \%$	
	5	1		5	1		5	1		5	1
6	6	-	30	21	23	54	35	37	78	49	51
7	7	-	31	22	24	55	36	38	79	49	52
8	8	8	32	23	24	56	36	39	80	50	52
9	8	9	33	23	25	57	37	39	81	50	53
10	9	10	34	24	25	58	37	40	82	51	54
11	10	11	35	24	26	59	38	40	83	51	54
12	10	11	36	25	27	60	39	41	84	52	55
13	11	12	37	25	27	61	39	41	85	53	55
14	12	13	38	26	28	62	40	42	86	53	56
15	12	13	39	27	28	63	40	43	87	54	56
16	13	14	40	27	29	64	41	43	88	54	57
17	13	15	41	28	30	65	41	44	89	55	58
18	14	15	42	28	30	66	42	44	90	55	58
19	15	16	43	29	31	67	42	45	91	56	59
20	15	17	44	29	31	68	43	46	92	56	59
21	16	17	45	30	32	69	44	46	93	57	60
22	17	18	46	31	33	70	44	47	94	57	60
23	17	19	47	31	33	71	45	47	95	58	61
24	18	19	48	32	34	72	45	48	96	59	62
25	18	20	49	32	34	73	46	48	97	59	62
26	19	20	50	33	35	74	46	49	98	60	63
27	20	21	51	33	36	75	47	50	99	60	63
28	20	22	52	34	36	76	48	50	100	61	64
29	21	22	53	35	37	77	48	51	—	—	—

P	0,05	0,01	—	0,05	0,01	—	0,05	0,01	—	0,05	0,01
---	------	------	---	------	------	---	------	------	---	------	------

Таблица 13

**Критические значения парного Т-критерия Уилкоксона
(односторонний критерий)**

Число парных наблюдений <i>n</i>	Уровни значимости α , %		Число парных наблюдений <i>n</i>	Уровни значимости	
	5	1		5	1
5	0	-	14	25	16
6	2	0	15	30	19
7	3	0	16	35	23
8	5	1	17	41	28
9	8	3	18	47	33
10	10	5	19	53	38
11	13	7	20	60	42
12	17	10	21	67	50
13	21	12	22	74	56
<i>P</i>	0,05	0,01	—	0,05	0,01

Таблица 13 продолжение
(двусторонний критерий)

Число парных наблюдений <i>n</i>	Уровни значимости α , %		Число парных наблюдений <i>n</i>	Уровни значимости α , %	
	5	1		5	1
6	1		16	31	21
7	3	—	17	36	24
8	5	1	18	41	29
9	7	3	19	47	33
10	9	4	20	53	39
11	12	e	21	60	44
12	15	8	22	67	50
13	18	11	23	74	56
14	22	14	24	82	62
15	26	17	25	90	69
<i>P</i>	0,05	0,01	—	0,05	0,01

Критические значения τ_{\max} в зависимости от объема выборки n . Р
 $(\tau_{\max} > \tau_{\max\alpha}) = \alpha$

α			α		
n	0,05	0,01	n	0,05	0,01
3	1,41	1,42	28	2,93	3,26
4	1,71	1,73	29	2,94	3,28
5	1,92	1,97	30	2,96	3,29
6	2,07	2,16	35	3,02	3,36
7	2,18	2,31	40	3,08	3,42
8	2,27	2,43	45	3,12	3,48
9	2,35	2,53	50	3,16	3,52
10	2,41	2,62	60	3,22	3,58
11	2,47	2,69	70	3,28	3,64
12	2,52	2,75	80	3,33	3,70
13	2,56	2,81	90	3,37	3,74
14	2,60	2,86	100	3,40	3,77
15	2,64	2,90	120	3,46	3,83
16	2,67	2,94	150	3,53	3,90
17	2,70	2,98	200	3,61	3,98
18	2,75	3,02	300	3,73	4,09
19	2,75	3,05	400	3,80	4,17
20	2,78	3,08	500	3,87	4,24
21	2,80	3,11	600	3,92	4,28
22	2,82	3,13	700	3,96	4,32
23	2,84	3,16	800	3,99	4,35
24	2,86	3,18	900	4,02	4,38
25	2,88	3,20	1000	4,05	4,41
26	2,90	3,22	1500	4,14	4,50
27	2,91	3,24	2000	4,21	4,56

Таблица 15

Критические значения коэффициента асимметрии A_s

Объём выборки n	Уровни значимости α , %		Объём выборки n	Уровни значимости %	
	5	1		5	1
25	0,711	1,061	250	0,251	0,360
30	0,661	0,982	300	0,230	0,329
35	0,621	0,921	350	0,213	0,305
40	0,587	0,869	400	0,200	0,285
45	0,558	0,825	450	0,188	0,269
50	0,533	0,787	500	0,179	0,255
60	0,492	0,723	550	0,171	0,243
70	0,459	0,673	600	0,163	0,233
80	0,432	0,631	650	0,157	0,224
90	0,409	0,596	700	0,151	0,215
100	0,389	0,567	750	0,146	0,208
125	0,350	0,508	800	0,142	0,202
150	0,321	0,464	850	0,138	0,196
175	0,298	0,430	900	0,134	0,190
200	0,280	0,403	950	0,130	0,185
			1000	0,127	0,180
P	0,05	0,01	—	0,05	0,01

Таблица 16

Критические значения коэффициента эксцесса E_x

Объём выборки n	Уровни значимости α , %		
	10	5	1
11	0,890	0,907	0,936
16	0,873	0,888	0,914
21	0,863	0,877	0,900
26	0,857	0,869	0,890
31	0,851	0,863	0,883
36	0,847	0,858	0,877
41	0,844	0,854	0,872
46	0,841	0,851	0,868
51	0,839	0,848	0,865
61	0,835	0,843	0,859
71	0,832	0,840	0,855
81	0,830	0,838	0,852
91	0,828	0,835	0,848
101	0,826	0,834	0,846
201	0,818	0,823	0,832
301	0,814	0,818	0,826
401	0,812	0,816	0,822
501	0,810	0,814	0,820
P	0,10	0,05	0,01

Таблица 17

Критические значения критерия $t_1 = \frac{X_2 - X_1}{X_{n-1} - X_1}$

n	Уровни значимости α , %		n	Уровни значимости α , %		n	Уровни значимости α , %	
	5	1		5	1		5	1
4	0,76	0,89	13	0,36	0,46	22	0,29	0,38
5	0,64	0,78	14	0,35	0,45	23	0,28	0,37
6	0,56	0,70	15	0,34	0,44	24	0,28	0,37
7	0,31	0,64	16	0,33	0,43	25	0,28	0,36
8	0,47	0,59	17	0,32	0,42	26	0,27	0,36
9	0,44	0,54	18	0,31	0,41	27	0,27	0,35
10	0,41	0,53	19	0,31	0,40	28	0,27	0,35
11	0,39	0,50	20	0,30	0,39	29	0,26	0,34
12	0,38	0,48	21	0,30	0,38	30	0,26	0,34

Таблица 18

Критические значения величины нормированного отклонения при оценке сомнительных вариантов с учетом объема выборки n и уровней значимости α

n	α %		n	α %		n	α %	
	5	1		5	1		5	1
4	1,71	1,73	13	2,56	2,81	23	2,84	3,16
5	1,92	1,97	14	2,60	2,86	24	2,86	3,18
6	2,07	2,16	15	2,64	2,90	25	2,88	3,20
7	2,18	2,31	16	2,67	2,95	26	2,90	3,22
8	2,27	2,43	17	2,70	2,98	27	2,91	3,24
9	2,35	2,53	18	2,73	3,02	28	2,93	3,26
10	2,41	2,62	19	2,75	3,05	29	2,94	3,28
11	2,47	2,69	20	2,78	3,08	30	2,96	3,29
12	2,52	2,75	21	2,80	3,11			
P	0,05	0,01	—	0,05	0,01	—	0,05	0,01

Таблица 19

Критические значения критерия $t_2 = \frac{X_n - X_{n-1}}{X_n - X_2}$

n	Уровни значимости α , %		n	Уровни значимости α , %		n	Уровни значимости α , %	
	5	1		5	1		5	1
4	0,96	0,99	13	0,41	0,52	22	0,32	0,41
5	0,81	0,92	14	0,40	0,50	23	0,31	0,41
6	0,69	0,80	15	0,38	0,49	24	0,31	0,40
7	0,61	0,74	16	0,37	0,47	25	0,30	0,39
8	0,55	0,68	17	0,36	0,46	26	0,30	0,39
9	0,51	0,64	18	0,35	0,45	27	0,30	0,38
10	0,48	0,60	19	0,34	0,44	28	0,29	0,38
11	0,45	0,57	20	0,33	0,43	29	0,29	0,37
12	0,43	0,54	21	0,33	0,42	30	0,28	0,37
P	0,05	0,01	—	0,05	0,01	—	0,05	0,01

Таблица 20

Критические значения критерия χ^2_R Фридмана

n	n _s = 3		n _s = 4		n	n _s = 3	
	α, %		α, %			α, %	
	5	1	5	1		5	1
2			6,0		9	6,22	8,67
3	6,0		7,4	9,0	10	6,20	9,60
4	6,5	8,0	7,8	9,6	11	6,54	9,46
5	6,4	8,4	7,8	9,96	12	6,17	9,50
6	7,0	9,0	7,6	10,20	13	6,0	9,38
7	7,14	8,86	7,8	10,37	14	6,14	9,00
8	6,25	9,00	7,65	10,35	15	6,40	8,93
P	0,05	0,01	0,05	0,01		0,05	0,01

Таблица 21

Критические значения критерия Н. Краскелла-Уоллиса

n ₁	n ₂	n ₃	Уровни значимости α, %		n ₁	n ₂	n ₃	Уровни значимости α, %	
2	2	3	4,71	—	3	4	4	5,62	7,14
2	2	4	5,33	—	3	4	5	5,63	7,45
2	2	5	5,16	6,53	3	4	6	5,61	7,41
2	2	6	5,35	6,65	3	5	5	5,71	7,54
2	3	3	5,36	—	3	5	6	5,60	7,59
2	3	4	5,44	6,44	3	6	6	5,63	7,45
2	3	5	5,25	6,82	4	4	4	5,69	7,65
2	3	6	5,35	6,97	4	4	5	5,62	7,76
2	4	4	5,45	7,04	4	4	6	5,68	7,94
2	4	5	5,27	7,12	4	5	5	5,64	7,77
2	4	6	5,34	7,34	4	5	6	5,66	7,94
2	5	5	5,34	7,27	4	6	6	5,72	8,00
2	5	6	5,34	7,38	5	5	5	5,78	8,00
3	3	3	5,60	7,20	5	5	6	5,73	8,03
3	3	4	5,73	6,75	5	6	6	5,76	8,12
3	3	5	5,65	7,08	6	6	6	5,80	8,22
3	3	6	5,62	7,41					

Таблица 22

Критические значения коэффициента корреляции r_{xy}

Степени свободы $k = n - 2$	Уровни значимости α , %		Степени свободы $k = n - 2$	Уровни значимости α , %	
	5	1		5	1
5	0,75	0,87	27	0,37	0,47
6	0,71	0,83	28	0,36	0,46
7	0,67	0,80	29	0,36	0,46
8	0,63	0,77	30	0,35	0,45
9	0,60	0,74	35	0,33	0,42
10	0,58	0,71	40	0,30	0,39
11	0,55	0,68	45	0,29	0,37
12	0,53	0,66	50	0,27	0,35
13	0,51	0,64	60	0,25	0,33
14	0,50	0,62	70	0,23	0,30
15	0,48	0,61	80	0,22	0,28
16	0,47	0,59	90	0,21	0,27
17	0,46	0,58	100	0,20	0,25
18	0,44	0,56	125	0,17	0,23
19	0,43	0,55	150	0,16	0,21
20	0,42	0,54	200	0,14	0,18
21	0,41	0,53	300	0,11	0,15
22	0,40	0,52	400	0,10	0,13
23	0,40	0,51	500	0,09	0,12
24	0,39	0,50	700	0,07	0,10
25	0,38	0,49	900	0,06	0,09
26	0,37	0,48	1000	0,06	0,09
P	0,05	0,01	—	0,05	0,01

Таблица 23

Значения z , соответствующие значениям выборочного коэффициента корреляции r_{xy}

r_{xy}	Сотые доли коэффициента корреляции									
	0	1	2	3	4	5	6	7	8	9
0,0	0,000	0,010	0,020	0,030	0,040	0,050	0,060	0,070	0,080	0,090
0,1	0,100	0,110	0,121	0,131	0,141	0,151	0,161	0,172	0,182	0,192
0,2	0,203	0,213	0,224	0,234	0,245	0,255	0,266	0,277	0,288	0,299
0,3	0,310	0,321	0,332	0,343	0,354	0,365	0,377	0,388	0,400	0,412
0,4	0,424	0,436	0,448	0,460	0,472	0,485	0,497	0,510	0,523	0,536
0,5	0,549	0,563	0,576	0,590	0,604	0,618	0,633	0,648	0,663	0,678
0,6	0,693	0,709	0,725	0,741	0,758	0,775	0,793	0,811	0,829	0,848
0,7	0,867	0,887	0,908	0,929	0,951	0,973	0,996	1,020	1,045	1,071
0,8	1,099	1,127	1,157	1,188	1,221	1,256	1,293	1,333	1,376	1,422
0,9	1,472	1,528	1,589	1,658	1,738	1,832	1,946	2,092	2,298	2,647
0,99	2,647	2,700	2,759	2,826	2,903	2,995	3,106	3,250	3,453	3,800

Таблица 24

Критические значения коэффициента корреляции рангов при различных уровнях значимости α и объемах выборки n

n	α, %		n	α, %		n	α, %	
	5	1		5	1		5	1
5	0,94		17	0,48	0,62	29	0,37	0,48
6	0,85	—	18	0,47	0,60	30	0,36	0,47
7	0,78	0,94	19	0,46	0,58	31	0,36	0,46
8	0,72	0,88	20	0,45	0,57	32	0,36	0,45
9	0,68	0,83	21	0,44	0,56	33	0,34	0,45
10	0,64	0,79	22	0,43	0,54	34	0,34	0,44
11	0,61	0,76	23	0,42	0,53	35	0,33	0,43
12	0,58	0,73	24	0,41	0,52	36	0,33	0,43
13	0,56	0,70	25	0,40	0,51	37	0,33	0,42
14	0,54	0,68	26	0,39	0,50	38	0,32	0,41
15	0,52	0,66	27	0,38	0,49	39	0,32	0,41
16	0,50	0,64	28	0,38	0,48	40	0,31	0,40
p	0,05	0,01	--	0,05	0,01	—	0,05	0,01

Таблица 25

Значения величины Q , соответствующие 5%-ному уровню значимости α и числу групп (градаций), входящих в дисперсионный комплекс

Число степеней свободы k	Число градаций a										
	2	3	4	5	6	7	8	9	10	11	12
6	3,5	4,3	4,9	5,3	5,6	5,9	6,1	6,3	6,5	6,7	6,8
7	3,3	4,2	4,7	5,1	5,4	5,5	5,8	6,0	6,2	6,3	6,4
8	3,3	4,0	4,5	4,9	5,2	5,4	5,6	5,8	5,9	6,1	6,2
9	3,2	4,0	4,4	4,8	5,0	5,2	5,4	5,6	5,7	5,9	6,0
10	3,1	3,9	4,3	4,7	4,9	5,1	5,3	5,5	5,6	5,7	5,8
11	3,1	3,8	4,2	4,6	4,8	5,0	5,2	5,4	5,5	5,6	5,7
12	3,1	3,8	4,2	4,5	4,8	5,0	5,1	5,3	5,4	5,5	5,6
13	3,1	3,7	4,2	4,5	4,7	4,9	5,1	5,2	5,3	5,4	5,5
14		3,7	4,1	4,4	4,6	4,8	5,0	5,1	5,3	5,4	5,5
15	3,0	3,7	4,1	4,4	4,6	4,8	4,9	5,1	5,2	5,3	5,4
16	3,0	3,7	4,1	4,3	4,6	4,7	4,9	5,0	5,2	5,3	5,4
17	3,0	3,6	4,0	4,3	4,5	4,7	4,9	5,0	5,1	5,2	5,3
18	3,0	3,6	4,0	4,3	4,5	4,7	4,8	5,0	5,1	5,2	5,3
19	3,0	3,6	4,0	4,3	4,5	4,6	4,8	4,9	5,0	5,1	5,2
20	3,0	3,6	4,0	4,2	4,5	4,6	4,8	4,9	5,0	5,1	5,2
24	2,9	3,5	3,9	4,2	4,4	4,5	4,7	4,8	4,9	5,0	5,1
30	2,9	3,5	3,8	4,1	4,3	4,5	4,6	4,7	4,8	4,9	5,0
40	2,9	3,4	3,8	4,0	4,2	4,4	4,5	4,6	4,7	4,8	4,9
60	2,8	3,4	3,7	4,0	4,2	4,3	4,4	4,6	4,7	4,7	4,8
120	2,8	3,4	3,7	3,9	4,1	4,2	4,4	4,5	4,6	4,6	4,7

**Коэффициент a_{n-i+1} для вычисления статистики Уилка-Шапиро
при n от 3 до 50 (ноль целых и запятая опущены)**

		n							
i	3	4	5	6	7	8	9	10	
1	7071	6872	6646	6431	6233	6052	5888	5739	
2		1677	2413	2606	3031	3164	3244	3291	
3				0875	1401	1743	1976	2141	
4						0561	0947	1224	
5								0399	
		n							
i	11	12	13	14	15	16	17	18	
1	5601	5475	5359	5251	5150	5056	4968	4886	
2	3315	3325	3325	3318	3306	3290	3273	3253	
3	2260	2347	2412	2460	2495	2521	2540	2553	
4	1429	1586	1707	1802	1878	1939	1988	2027	
5	0695	0933	1099	1240	1353	1447	1524	1587	
6		0303	0539	0727	0880	1005	1109	1197	
7				0240	0433	0593	0725	0837	
8						0196	0359	0496	
9								0164	
		n							
i	19	20	21	22	23	24	25	26	
1	4808	4734	4643	4590	4542	4493	4450	4407	
2	3232	3211	3185	3156	3126	3098	3069	3043	
3	2561	2565	2578	2571	2563	2554	2543	2533	
4	2059	2085	2119	2131	2139	2145	2148	2151	
5	1641	1686	1736	1764	1787	1808	1822	1836	
6	1271	1334	1399	1443	1480	1512	1539	1563	
7	0932	1013	1092	1150	1201	1245	1283	1316	
8	0612	0711	0804	0878	0941	0997	1046	1089	
9	0303	0422	0530	0618	0696	0764	0823	0876	
10		0140	0263	0368	0459	0539	0610	0672	
11				0122	0228	0321	0403	0476	
12						0107	0200	0284	
13								0094	
		n							
i	27	28	29	30	31	32	33	34	
1	4366	4328	4291	4254	4220	4188	4145	4127	
2	3018	2992	2968	2944	2921	2898	2876	2854	
3	2522	2510	2499	2487	2475	2463	2451	2439	
4	2152	2151	2150	2148	2145	2141	2137	2132	
5	1848	1857	1864	1870	1874	1878	1880	1882	
6	1584	1601	1616	1630	1641	1651	1660	1667	
7	1346	1372	1395	1415	1433	1449	1463	1467	
8	1128	1162	1192	1219	1243	1265	1284	1301	
9	0923	0965	1002	1036	1066	1093	1118	1140	
10	0728	0778	0822	0862	0899	0931	0961	0988	
11	0540	0598	0650-	0697	0739	0777	0812	0844	
12	0358	0424	0483	0537	0585	0629	0669	0706	
13	0178	0235	0320	0381	0435	0485	0530	0572	
14		0084	0159	0227	0289	0344	0395	0441	
15				0076	0144	0206	0262	0314	
16						0068	0131	0187	
17								0062	

Таблица 26 Окончание

i	n							
	35	36	37	38	39	40	41	42
1	4096	4068	4040	4015	3989	3964	3940	3917
2	2834	2813	2794	2774	2755	2737	2719	2701
3	2427	2415	2403	2391	2380	2368	2357	2345
4	2127	2121	2116	2110	2104	2098	2091	2085
5	1883	1883	1883	1880	1880	1880	1876	1874
6	1673	1678	1683	1686	1689	1691	1693	1694
7	1487	1486	1505	1513	1520	1526	1531	1535
8	1317	1331	1344	1356	1366	1376	1384	1392
9	1160	1179	1196	1211	1225	1237	1249	1259
10	1013	1036	1056	1075	1092	1108	1123	1136
11	0873	0900	0924	0947	0967	0986	1004	1020
12	0739	0770	0794	0824	0848	0870	0891	0909
13	0610	0645	0667	0706	0733	0759	0782	0804
14	0484	0523	0559	0592	0622	0651	0677	0701
15	0361	0404	0444	0481	0515	0546	0575	0602
16	0239	0287	0331	0372	0409	0444	0476	0506
17	0119	0172	0220	0264	0305	0343	0379	0411
18		0057	0110	0158	0203	0244	0283	0318
19				0053	0101	0146	0188	0227
20						0049	0094	0136
21								0045
i	n							
	43	44	45	46	47	48	49	50
1	3894	3872	3850	3830	3808	3789	3770	3751
2	2684	2667	2651	2635	2620	2604	2589	2574
3	2334	2323	2313	2302	2291	2281	2271	2260
4	2078	2072	2065	2058	2052	2045	2038	2032
5	1871	1868	1865	1865	1859	1855	1851	1847
6	1695	1695	1695	1695	1695	1693	1692	1691
7	1539	1542	1545	1548	1550	1551	1553	1554
8	1398	1405	1410	1415	1420	1423	1427	1430
9	1269	1278	1286	1293	1300	1306	1312	1317
10	1149	1160	1170	1180	1189	1197	1205	1212
11	1035	1049	1062	1073	1085	1095	1105	1113
12	0927	0943	0959	0972	0986	0998	1010	1020
13	0824	0842	0860	0876	0892	0906	0919	0932
14	0724	0745	0765	0783	0801	0817	0832	0846
15	0628	0651	0673	0694	0713	0731	0748	0664
16	0534	0560	0534	0607	0628	0648	0667	0685
17	0442	0471	0497	0522	0546	0568	0588	0603
18	0352	0383	0412	0439	0465	0489	0511	0532
19	0263	0286	0328	0357	0385	0411	0436	0459
20	0175	0211	0245	0277	0307	0335	0362	0389
21	0087	0126	0163	0197	0220	0259	0288	0314
22		0042	0081	0118	0153	0185	0215	0244
23				0039	0076	0111	0143	0174
24						0037	0071	0104
25								0035

СОДЕРЖАНИЕ

<i>ВВЕДЕНИЕ</i>	3
Глава 1. ОСНОВНЫЕ ПОНЯТИЯ	4
1.1. Признаки и их классификация	5
1.2. Точность измерений	6
1.3. Репрезентативность выборки и рандомизация	7
1.4. Способы группировки первичной информации	8
<i>Вопросы к главе 1</i>	10
Глава 2. ОСНОВНЫЕ ХАРАКТЕРИСТИКИ ВАРЬИРУЮЩИХ ОБЪЕКТОВ	11
2.1. Средние величины	11
2.2. Структурные средние и способы их вычисления	13
2.3. Показатели вариации	14
2.4. Статистические характеристики при альтернативной группировке вариант	17
<i>Вопросы к главе 2</i>	18
Глава 3. ЗАКОНЫ РАСПРЕДЕЛЕНИЯ	19
3.1. Вероятность события и ее свойства	19
3.2. Характерные черты варьирования	20
3.3. Случайные события	21
3.4. Закон больших чисел	22
3.5. Нормальное распределение	23
3.6. Параметры нормального распределения. Основные свойства нормаль- ного распределения	25
3.7. Интеграл вероятности нормального распределения	26
3.8. Логнормальное распределение	28
3.9. Биномиальное распределение	29
3.10. Распределение Пуассона	30
3.11. Распределение Максвелла	32
3.12. Измерение асимметрии и эксцесса	32
3.13. Проверка нормальности распределения	33
<i>Вопросы к главе 3</i>	35
Глава 4. ВЫБОРОЧНЫЕ ОЦЕНКИ И ОШИБКИ РЕПРЕЗЕНТАТИВНОСТИ	36
4.1. Точечные оценки	36
4.2. Статистические оценки при объединении выборок	37
4.3. Ошибки репрезентативности	38
4.4. Интервальные оценки	40
<i>Вопросы к главе 4</i>	42
Глава 5. КРИТЕРИИ ДОСТОВЕРНОСТИ ОЦЕНОК	43
5.1. Параметрические критерии	43
5.2. t-критерий Стьюдента (t-распределение)	43
5.3. F-критерий Фишера (F-распределение)	48
5.4. Непараметрические критерии	49
5.5. X-критерий Ван-дер-Вардена	49

5.6. U-критерий Уилкоксона (Манна-Уитни)	49
5.7. Критерий знаков z	50
5.8. T-критерий Уилкоксона	50
<i>Вопросы к главе 5</i>	50
Глава 6. ПРОВЕРКА ГИПОТЕЗ О ЗАКОНАХ РАСПРЕДЕЛЕНИЯ	51
6.1. Критерий хи-квадрат (χ^2 распределение)	51
6.2. Критерий Ястремского J.	52
6.3. Критерий Уилка-Шапиро	52
6.4. Оценка трансгрессии рядов	53
6.5. Проверка сомнительных вариант	53
6.6. Применение коэффициентов асимметрии и эксцесса для проверки нормальности распределения	54
<i>Вопросы к главе 6</i>	54
Глава 7. ДИСПЕРСИОННЫЙ АНАЛИЗ	55
7.1. Основные понятия и символы	56
7.2. Анализ однофакторных комплексов	56
7.3. Неравночисленные комплексы	58
7.4. Оценка силы влияния факторов	58
7.5. Сравнение групповых средних дисперсионного комплекса	59
<i>Вопросы к главе 7</i>	61
Глава 8. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ	62
8.1. Параметрические показатели связи	62
8.2. Вычисление коэффициента корреляции	63
8.3. Оценка разности между коэффициентами корреляции	65
8.4. Корреляционное отношение	65
8.5. Коэффициент детерминации	66
<i>Вопросы к главе 8</i>	66
СПИСОК ЛИТЕРАТУРЫ	67
ПРИЛОЖЕНИЯ	68

Биганова Светлана Герсановна

БИОМЕТРИЯ

Учебное пособие

Подписано в печать 21.03.2015 г.

Формат бумаги 60x84¹/₁₆. Бумага ксероксная. Гарнитура Таймс.

Усл. печ. л. 6,25. Заказ №0101. Тираж 100 экз.

Издательство МГТУ

385000, г. Майкоп, ул. Первомайская, 191