ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

**«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ПУТЕЙ СООБЩЕНИЯ ИМПЕРАТОРА НИКОЛАЯ II»**

КАФЕДРА «МАТЕМАТИКА»

М.В. Ишханян

Введение в эконометрику

Учебное пособие

Москва – 2016

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПУТЕЙ СООБЩЕНИЯ ИМПЕРАТОРА НИКОЛАЯ II»

КАФЕДРА «МАТЕМАТИКА»

М.В. Ишханян

Введение в эконометрику

Учебное пособие
для студентов направления
«Экономика»

Москва – 2016

УДК – 330
И 97

Ишханян М.В. Введение в эконометрику: Учебное пособие. – М.: МГУПС (МИИТ), 2016. – 117 с.

Учебное пособие предназначено для студентов направления «Экономика», обучающихся по дисциплинам «Теория вероятностей и математическая статистика» или «Эконометрика». Содержит подробное изложение основных понятий, определений и положений математической статистики, теории оценивания, проверки гипотез, основ эконометрики. Теоретический материал сопровождается решением типовых практических задач, содержит примеры реализации решения прикладных задач на компьютере. Каждый из разделов снабжен перечнем заданий для самостоятельной работы. Может быть использовано для проведения практических занятий, а так же для организации самостоятельной работы студентов.

Рецензенты:

О.А. Платонова, к.ф.-м.н., доцент, зав. кафедрой «Высшая и вычислительная математика» МГУПС (МИИТ).
Л.А. Климина, к.ф.-м.н., с.н.с. НИИ механики МГУ имени М.В. Ломоносова.

Оглавление

§1. Введение

При анализе реальных социально-экономических процессов приходится обрабатывать большие объемы статистических данных по различным показателям, которые по своей сути являются случайными величинами. По ходу проводимого исследования часто возникает необходимость оценивания числовых значений различных параметров, неоднократно приходится выдвигать и проверять различные предположения, устанавливать наличие и силу зависимости между разнообразными факторами. На практике мы сталкиваемся с конкретными реализациями рассматриваемых случайных величин. Количество таких реализаций носит ограниченный характер, что не позволяет применять напрямую теоретические методы анализа. Поэтому здесь в первую очередь используются методы и модели математической статистики, позволяющие получить необходимые знания об исследуемом объекте, осуществить направленный анализ и сделать обоснованные выводы.

Одной из центральных задач математической статистики является выявление закономерностей в статистических данных, на базе чего можно будет строить соответствующие модели для принятия обдуманных решений. Под статистическими данными подразумеваются данные наблюдений за значениями некоторой случайной величины или совокупности случайных величин, характеризующих изучаемый процесс.

Первая задача математической статистики − указать способы сбора и группировки статистических данных, полученных в результате наблюдений или испытаний.

Вторая задача математической статистики – разработать методы анализа статистических данных в зависимости от целей исследования.

Сюда относятся:

a) оценки неизвестной вероятности события; неизвестной функции распределения; неизвестных параметров известного распределения; зависимости двух или нескольких случайных величин и т. п.;

b) проверка статистических гипотез о виде неизвестного распределения; о величинах параметров известного распределения; о виде и силе зависимости между рассматриваемыми случайными величинами.

Таким образом, основная задача математической статистики состоит в создании методов сбора и обработки статистических данных для получения научных и практических выводов.

Знание методов математической статистики и умение ими оперировать являются необходимой предпосылкой для успешного эконометрического анализа.

§2. Генеральная совокупность и случайные выборки

Приведем некоторые понятия статистического наблюдения, необходимые для дальнейшего изложения.

Объект наблюдения - это совокупность предметов и явлений, объединенных каким-либо общим признаком или свойством качественного или количественного характера.

Всякий объект наблюдения состоит из ***единиц наблюдений (единиц отбора)***, содержащих первичные данные, и ***единиц совокупности (элемент выборки)***, являющихся носителями признаков, подлежащих наблюдению.

Пример 1. Происходит перепись населения страны. Тогда, единицей наблюдения может быть каждый человек, которые является и источником сведений, и носителем признаков.

Статистическое наблюдение можно проводить как для всей, так и для части совокупности. Однако, при этом возникает ряд вопросов об организации такого отбора: сколько единиц совокупности следует взять и насколько достоверны и надежны будут полученные результаты. Ответы на данные вопросы дает выборочный метод.

Если наблюдение организовано так, что анализу подлежат все элементы выборки, то в этом случае статистическую совокупность называют генеральной.

Генеральной совокупностью называется вся подлежащая изучению какого-либо свойства (говорят, признака) совокупность объектов.

Выборочная совокупность (или выборка) – это часть объектов, которая случайным образом отобрана для непосредственного изучения какого-либо признака генеральной совокупности.

Объем генеральной совокупности и объем выборки – это количество элементов в них. Обозначаются, соответственно, N и n.

В дальнейшем будем считать, что объем выборки существенно меньше объема генеральной совокупности.

Например, $N = 100\ 000, n = 1000$.

Замечание. Если объем совокупности велик, то его полагают равным бесконечности.

Случайная выборка из n элементов - это такой отбор, при котором элементы извлекаются по одному из всей генеральной совокупности и каждый из них имеет равный быть отобранным. Такая выборка называется ***собственно-случайной.***

Когда исследователи используют слово «***выборка***», они, как правило, имеют в виду ***репрезентативную***, т.е. представительную выборку (хорошо представляет элементы генеральной совокупности).

Для того, чтобы выборка была ***репрезентативной***, она должна быть отобрана случайно. Нарушение принципов случайного выбора приводит к серьезным ошибкам.

Различают ***повторную*** и ***бесповторную*** выборки. В первом случае отобранный объект перед отбором следующего возвращается в генеральную совокупность. Во втором, отобранный в выборку объект не возвращается в генеральную совокупность. Если выборка составляет незначительную часть генеральной совокупности, то различие между повторной и бесповторной выборками стирается.

Способы отбора собственно-случайной выборки

Различают следующие способы отбора:

1) механический;
2) стратифицированный (типический);
3) серийный (гнездовой);
4) комбинированный.

Механический отбор всегда бесповторный. В данном случае генеральная совокупность делится на столько групп, сколько единиц должно войти в выборку, и из каждой группы отбирается одна единица.

Пример 2.

- Учебный отдел проводит оценку успеваемости студентов. Проведена 5%-ная механическая выборка. Для этого из алфавитного списка отбирается каждый 20-ый студент. Например, 1-й, 21-й, 41-й, ...
- Если выборка 10%-ная, то отбору подлежит каждый 10-й студент.

При ***типическом отборе*** генеральная совокупность делится по некоторому признаку на однотипные группы и затем из каждой группы собственно-случайным или механическим отбором выбирают количество единиц, пропорциональных весу группы во всей совокупности.

Пример 3. Планируется провести опрос 1500 жителей небольшого города для прогноза результатов предстоящих выборов. Списки четырех избирательных участков города охватывают 100 000 избирателей. Расчет структуры объема выборки производится по следующей схеме:

$$n_g = n\frac{N_g}{N}$$

Решение примера 3.

Проведем расчет структуры объема выборки по данным примера

Участок	**Число избирателей**	**Доля избирателей в общей численности**	**Число отобранных для выборки избирателей**
1-й	18 000	0,18	$1500 \cdot 0{,}18 = 270$
2-й	24 000	0,24	360
3-й	32 000	0,32	480
4-й	26 0000	0,26	390
ИТОГО	100 000	1	1 500

Серийным называют отбор, при котором объекты выбираются из генеральной совокупности не по одному, а сериями, которые подвергаются сплошному обследованию.

Пример 4. Таможенная служба выборочно вскрывает каждый сотый контейнер из прибывающих в порт, а в нем проверяется полностью весь груз.

Нередко на практике применяется ***комбинированный отбор***, при котором сочетаются описанные выше способы.

Рассмотренные примеры иллюстрируют общие принципы различных способов отбора, на практике же для реализации выборки применяются более сложные приемы, обеспечивающие случайность.

Способы представления данных. Первичная обработка статистических данных

После получения (тем или иным способом) выборочной совокупности все ее объекты обследуются по отношению к определенной случайной величине, т.е. обследуемому признаку объекта. В результате этого получают наблюдаемые данные, которые представляют собой множество чисел, расположенных в беспорядке. Анализ таких данных весьма затруднителен, и для изучения закономерностей полученные данные подвергаются определенной обработке.

Для этого применяются различные формы упорядочения данных в выборке по возрастанию, по совпадающим значениям, по интервалам и т. п.

Пример 5. На телефонной станции проводились наблюдения над числом X неправильных соединений в минуту. Наблюдения в течение часа дали следующие 60 значений: 2; 1; 0; 3; 4; 1; 2; 4; 0; 3; 3; 1; 3; 1; 4; 0; 2; 2; 0; 1; 1; 4; 3; 1; 1; 4; 2; 2; 1; 1; 1; 3; 2; 7; 2; 0; 0; 1; 3; 3; 1; 2; 1; 2; 0; 2; 3; 1; 2; 5; 1; 2; 4; 2; 0; 2; 3; 1; 2; 5

Очевидно, что число *X* является дискретной случайной величиной, а полученные данные есть значения этой случайной величины. Анализ исходных данных в таком виде весьма затруднителен.

Простейшая операция – ***ранжирование*** статистических данных.

Ранжированием – называется операция расположения экспериментальных данных в порядке *неубывания*. Если среди элементов встречаются одинаковые, то они объединяются в одну группу. Значение случайной величины, соответствующее отдельной группе сгруппированного ряда наблюдаемых данных, называется ***вариантом***, а изменение этого значения – ***варьированием***. Варианты будем обозначать строчными буквами с соответствующими порядковому номеру группы индексами $x_1, x_2, \dots, x_k$, где k – число групп.

При этом, $x_1 < x_2 < x_3 < \dots < x_k; k \le n,$

причем значение x_i встречается n_i раз:

$$n_1 + n_2 + \dots + n_k = n$$

x_i –вариант, n_i - ***частота варианта***, $w_i = n_i/n$ – ***относительная частота варианта (частностью)*** $\left(\sum_{i=1}^{k} w_i = 1\right)$.

Разность между максимальным и минимальным значениями вариант называется ***размахом выборки***:

$$R = x_{max} - x_{min}$$

Вариационным рядом называется ранжированная последовательность вариант.

Статистическим распределением выборки (рядом) называется ранжированный в порядке возрастания ряд значений (вариантов) с соответствующими им частотами.

Значения x_i	x_1	x_2	...	x_k
Частоты n_i	n_1	n_2	...	n_k
Относительные частоты $w_i = n_i/n$	w_1	w_2	...	w_k

Пример 6. Для данных ***примера 5*** были выполнены операции ранжирования и группировки. В результате были получены семь значений случайной величины (варианты): 0; 1; 2; 3; 4; 5; 7. При этом значение 0 в этой группе встречается 8 раз, значение 1 – 17 раз, значение 2 – 16 раз, значение 3 – 10 раз, значение 4 – 6 раз, значение 5 – 2 раза, значение 7 – 1 раз. Составьте статистический ряд.

Решение примера 6. Статистический ряд имеет вид:

Значения x_i	0	1	2	3	4	5	7
Частоты n_i	8	17	16	10	6	2	1
$w_i = \frac{n_i}{n}$	$\frac{8}{60}$	$\frac{17}{60}$	$\frac{16}{60}$	$\frac{10}{60}$	$\frac{6}{60}$	$\frac{2}{60}$	$\frac{1}{60}$

При большом объеме выборки ее элементы могут быть сгруппированы в ***интервальный статистический ряд***. Для этого все n наблюдаемых значений выборки разбивают по k непересекающимся подынтервалам равной длины h (h - шаг разбиения).

Как правило, частичные интервалы, на которые разбивается весь интервал варьирования, имеют одинаковую длину и представимы в виде $[z_i, z_i + h), i = 1, \dots, m,$ где m – число интервалов.

Длину h следует выбирать так, чтобы построенный ряд не был громоздким, но в то же время позволял выявлять характерные изменения случайной величины.

Для вычисления h рекомендуется использовать следующую формулу (***формулу Стэрджеса)***:

$$h = \frac{x_{max} - x_{min}}{1 + 3{,}222 \lg n}$$

где x_{max}, x_{min} – наибольшее и наименьшее значения случайной величины, n - объем выборки.

Если окажется, что h – дробное число, то за длину интервала следует принять либо ближайшую простую дробь, либо ближайшую целую величину. При этом необходимо выполнение условий:

$$z_1 \leq x_{min} \text{ и } z_m + h \geq x_{max}$$

После нахождения частных интервалов определяется, сколько значений случайной величины попало в каждый конкретный интервал. При этом в интервал включают значения, большие или равные нижней границе и меньшие верхней границы.

Полученный при этом ряд ***называют интервальным статистическим распределением (рядом)***:

Интервалы Δ_i	Δ_1	Δ_2	...	Δ_k
Частоты n_i	n_1	n_2	...	n_k
Относительные частоты $w_i = n_i/n$	w_1	w_2	...	w_k

Здесь n_i – количество вариант, попавших в интервал Δ_i.

Пример 7. Для данных о числе товаров, проданных 26 продавцам универмага были выполнены операции ранжирования и группировки. В результате был получен статистический ряд имеет вид:

x_i		2	3	4	5	6	7	9	1	3	7
n_i											

Постройте интервальный статистический ряд.

Решение примера 7. Согласно формуле Стэрджеса длина интервала:

$$h = \frac{x_{max} - x_{min}}{1 + 3{,}222 \lg n} = \frac{27 - 9}{1 + 3{,}222 \lg 26} = 3{,}2380$$

Округляем до целого числа: $h = 3$. Интервальный ряд имеет вид:

Δ_i	9-12	12-15	15-18	18-21	21-24	24-27
n_i	1	11	10	1	2	1

Разность между верхней и нижней границами интервала называется ***интервальной разностью или длиной (величиной) интервала***.

Если интервал имеет обе границы, то его называют ***закрытым***. Первый и последний интервалы могут быть ***открытыми***, т. е. иметь только одну границу.

Например, 1-й интервал может быть задан как «до 100», 2-й — «100-110», ... предпоследний — «190-200», последний — «200 и более». Очевидно, что 1-й интервал не имеет нижней границы, а последний — верхней, оба они — открытые.

Часто открытые интервалы приходится условно закрывать. Обычно для этого величину 1-го интервала принимают равной величине 2-го, а величину последнего — величине предпоследнего. В нашем примере величина 2-го интервала равна 110 - 100 = 10, следовательно, нижняя граница 1-го условно составит 100 - 10 = 90; величина предпоследнего равна 200 - 190 = 10, значит, верхняя граница последнего условно составит 200 + 10 = 210.

Кроме этого в интервальном вариационном ряде могут встречаться ***интервалы разной длины***. Если интервалы в вариационном ряде имеют одинаковую длину (интервальную разность), их называют ***равновеликими***, в противном случае — ***неравновеликими***.

Визуализация данных

Важный инструмент представления числовых данных - использование графических изображений. Рассмотрим некоторые графические методы: гистограммы, полигоны частот, кумуляты.

Равновеликие интервалы

Гистограмма частот есть графическое представление, в котором по оси X откладываются значения переменной, а по оси У соответствующие им частоты. Гистограмма строится в виде прямоугольников, высота которых соответствует частоте значения переменной.

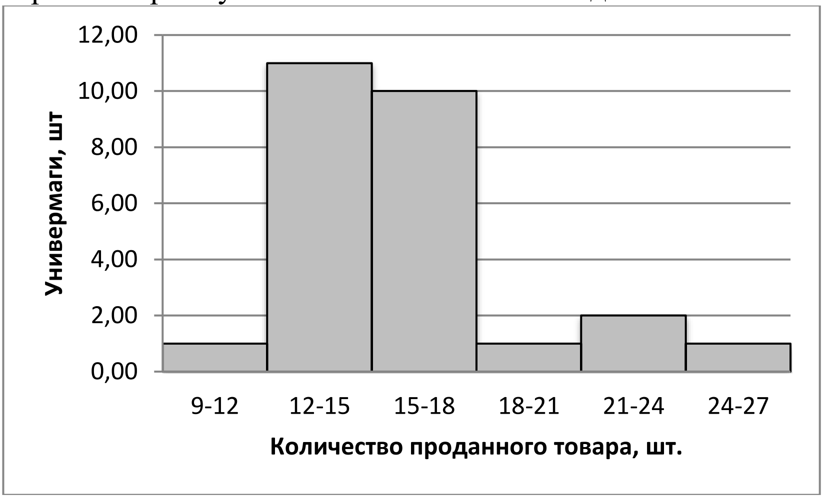
Гистограмма относительных частот есть графическое представление, в котором по оси X откладываются значения переменной, а по оси Y соответствующие им доли или проценты. Гистограмма строится в виде прямоугольников, высота которых соответствует доле или проценту для значения переменной

Неравновеликие интервалы

Гистограмма частот есть графическое представление, в котором по оси X откладываются значения переменной, а по оси У соответствующие им значения абсолютной плотности распределения $f(\Delta_i) = \frac{n_i}{l_i}$, где l_i - длина интервала Δ_i;n_i - его частота. Гистограмма строится в виде прямоугольников, высота которых соответствует значению $f(\Delta_i)$.

Гистограмма относительных частот есть графическое представление, в котором по оси X откладываются значения переменной, а по оси У соответствующие им значения относительной плотности распределения $f(\Delta_i) = \frac{w_i}{l_i}$, где l_i - длина интервала Δ_i ; w_i - его частость. Гистограмма строится в виде прямоугольников, высота которых соответствует значению $f(\Delta_i)$.

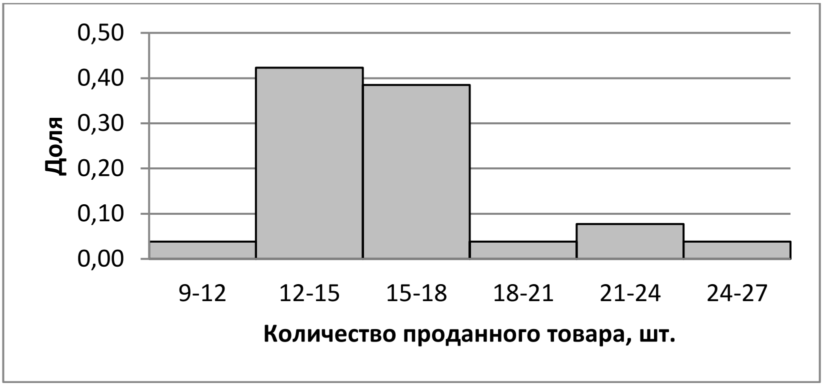
Построим гистограмму частот для примера о проданном товаре (***пример 7***). По оси X откладываем соответствующие интервалы: 9-12; 12-15; 15-18 и.т.д., всего шесть. В интервале 9-12 строим прямоугольник высотой 1, что означает, что лишь один продавец из 26 купил количество товара, попавшее в первый интервал, во втором интервале - прямоугольник высотой 11 и так далее.



Гистограмма частот полезна для визуальной оценки данных.

Можно отметить, что наибольшим спросом пользуется количество товара от 12 до 15 шт., в этот интервал попало 11 универмагов, это больше, чем в любой другой интервал.

Для рассмотренного нами примера гистограмма относительных частот имеет вид:



По гистограмме относительных частот можно сказать, что около 0,42 (или 42%) универмагов купило товара в количестве от 12 до 15 шт.

Полигон частот есть графическое представление, в котором по оси X откладываются значения переменной, а по оси У соответствующие им частоты. Полигон строится в виде области ограниченной линией, проходящей по точкам: для интервального ряда (середина интервала, частота), для статистического ряда (значение варианта, частота).

Полигоны, также как и гистограммы, можно строить для частот, относительных частот и процентов. Во всех трех случаях график остается прежним, за исключением оси Y.

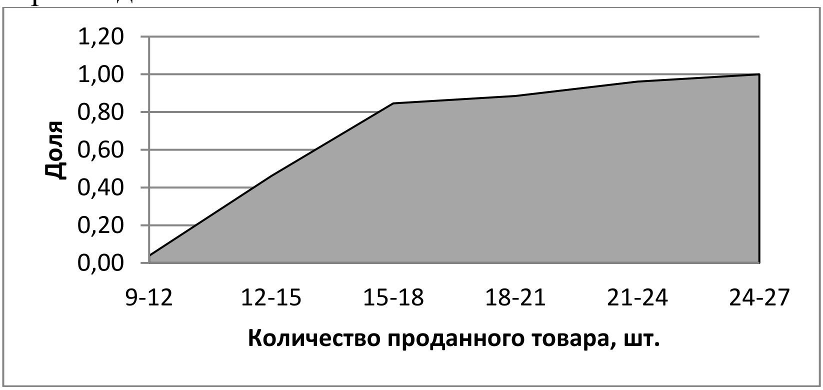
Полигон дает зрительное представление о распределении частот, которое сильно отличается от гистограммы при одних и тех же данных.

Не существует правил, которые говорили бы, какое из представлений лучше. Все зависит от конкретной ситуации и от вкусов исследователя, который выбирает вид представления их существующих альтернатив.

Для нашего примера полигон частот будет выглядеть следующим образом:

Еще одним часто используемым графическим представлением данных является ***кумулята*** - от слова "кумулятивный", что означает накапливаемый. В этом случае по оси Y откладываются накопленные частоты или накопленные проценты.

Для рассматриваемого нами примера кумулята построена для накопленных относительных частот:



По графику можно сказать, что около 80% (или 0,80) универмагов покупает товара в количестве менее 15шт., зато менее 12 единиц товара покупает около 20% универмагов.

Кумуляты могут быть построены как по возрастанию значений признака, так и по убыванию, в зависимости от того, какой анализ мы предполагаем проводить с использованием графического изображения.

Рассмотренные графические способы представления данных позволяют ответить на следующие вопросы:

- Какие значения являются максимальным и минимальным?
- Каков размах имеющихся данных?
- Какие значения встречаются в выборке чаще других?
- Какие значения являются наиболее типичными?
- Какой вид имеет распределение?
- Где сосредоточена основная часть данных?
- Симметрично ли они расположены вокруг типичного значения? В какую сторону смешены?
- Имеются ли характерные особенности? Выбросы? Есть ли значения признака, которые пропущены?

§3. Выборочные характеристики

Для любой случайной величины кроме определения ее функции распределения желательно указать ее числовые характеристики, важнейшими из которых являются математическое ожидание, дисперсия, среднее квадратическое отклонение (или стандартное отклонение).

Параметры и статистики

Мы будем использовать слово ***параметры***, когда речь будет идти о характеристиках генеральной совокупности, и слово ***статистики***, когда речь будет идти о характеристиках выборки.

Параметры - числовые характеристики генеральной совокупности.
Статистики – числовые характеристики выборки.

Любое число, полученное на основе выборки, носит название ***«выборочная статистика»*** (или просто ***«статистика»***).

Полигоны частот, гистограммы и кумуляты позволяют делать выводы о закономерностях исследуемого массового явления. Однако они неудобны для описания группирования и рассеивания наблюдаемых данных. Для этого используются так называемые *числовые характеристики выборочной совокупности*:

Выборочное среднее – среднее арифметическое вариантов.

Обозначение:

$M(X)$ - теоретическое обозначение;

$\bar{x}$ – статистическое обозначение.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot n_i = \sum_{i=1}^{k} x_i \cdot \underbrace{\frac{n_i}{n}}_{=w_i} = \sum_{i=1}^{k} x_i \cdot w_i$$

Выборочная дисперсия – среднее арифметическое квадратов отклонений вариантов от их среднего:

$$D_{\text{в}} = \sigma_{\text{в}}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot n_i$$
$$= \sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot w_i$$

Исправленная выборочная дисперсия

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot n_i$$
$$= \frac{n}{n-1}\sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot w_i$$

Выборочное стандартное отклонение (выборочное среднее квадратическое отклонение)

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot n_i}$$

Коэффициент вариации – величина, характеризующая насколько сильно элементы выборки отличаются друг от друга.

$$v = \frac{S}{|\bar{x}|} \cdot 100\%$$

Замечание. Использование коэффициента вариации имеет смысл только при изучении вариации признака, принимающего только положительные (отрицательные) значения. Если признак принимает и положительные, и отрицательные значения вычисление коэффициента вариации не имеет смысла

Принять решение о степени вариабельности переменной можно с помощью таблицы, либо с помощью экспертного задания некоторого критического значения коэффициента вариации v^*, например $v^* = 35\%$.

Переменные, удовлетворяющие неравенству $v \leq v^*$ признаются квазинеизменными (условно неизменными) и исключаются из множества потенциальных объясняющих переменных, поскольку не несут значимой информации.

Значение коэффициента вариации v	**Характер распределения переменной**
0%-20%	Весьма равномерный
20%-40%	Равномерный
40%-100%	Неравномерный
100%-150%	Весьма неравномерный
Более 150%	Крайне неравномерный

Медина – это значение признака, приходящееся на середину ранжированного ряда наблюдений.

Обозначение:

MeX- теоретическое обозначение;

$\widetilde{Me}$ - статистическое обозначение.

Если число вариант четное, т.е. $n = 2m$, то

$$\widetilde{Me} = \frac{x_m + x_{m+1}}{2}$$

Если число вариант нечетное, т.е. $n = 2m + 1$,

$$\widetilde{Me} = x_{m+1}$$

Мода – значение признака, наиболее часто встречающееся в выборке. Т.е. это варианта с самой большой частотой.

Обозначение:

MoX – теоретическое;

$\widetilde{Mo}$ – статистическое обозначение.

§4. Точечные оценки неизвестных параметров

При построении эконометрических моделей часто используются так называемые ***точечные*** (или ***выборочные***) оценки различных коэффициентов модели. Поэтому кратко остановимся на понятии точечной оценки, ее свойствах.

Пусть $x_1, x_2, \dots, x_n$ - выборка из генеральной совокупности, n – объем выборки.

Будем рассматривать эту выборку как систему случайных величин X_1, X_2, …, X_n.

Точечной оценкой $\tilde{\theta}$ неизвестного параметра θ теоретического закона распределения называют всякую функцию результатов наблюдений случайной величины X, значение которой принимают в качестве приближенных значений параметра θ:

$$\widetilde{\theta_n} = f(x_1, x_2, \dots, x_n)$$

Требования к точечным оценкам:

- ✓ **Несмещенность:** оценка $\widetilde{\theta_n}$ параметра θ называется ***несмещенной***, если $M\left(\widetilde{\theta_n}\right) = \theta$
- ✓ **Эффективность:** оценка $\widetilde{\theta_n}$ параметра θ называется ***эффективной***, если она имеет наименьшую дисперсию среди всех оценок параметра по выборкам одного и того же объема: $D\widetilde{\theta_n} \to \min$ при фиксированном n
- ✓ **Состоятельность:** оценка $\widetilde{\theta_n}$ параметра θ называется ***состоятельной***, если она удовлетворяет закону больших чисел (ЗБЧ):

$$\widetilde{\theta_n} \xrightarrow[n \to \infty]{P} \theta$$

✓ **Устойчивость:** при незначительных изменениях в исходной информации значение оценки не должно существенно изменяться.

На практике ищут оценки удовлетворяющие условию несмещенности или оценки с несколько большей дисперсией по сравнению с эффективной оценкой.

Рассмотрим часто используемые в эконометрике точечные оценки числовых характеристик случайной величины X.

Точечные оценки для числовых характеристик случайной величины

Оценкой для математического ожидания $M(X)$ случайной величины является ***выборочное среднее***

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Можно показать, что ***оценка $\overline{\mathbf{x}}$ является несмещенной, эффективной и состоятельной***, т.е. удовлетворяет всем требованиям «хорошей» оценки. В дальнейшем операцию усреднения каких-либо значений будем обозначать горизонтальной чертой над обозначением этих значений. Например,

$$\overline{x^2} = \frac{1}{\mathrm{n}}\sum_{i=1}^{n} x_i^2$$

Оценкой для дисперсии $\sigma_X^2 = D(X)$ случайной величины X является ***выборочная дисперсия***

$$\sigma_{\mathrm{X}}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

На практике для вычисления σ_X^2 часто используют следующую формулу:

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2 = \overline{x^2} - (\bar{x})^2$$

Оценка $\boldsymbol{\sigma_X^2}$ является состоятельной, но смещенной. Несмещенная оценка имеет вид:

$$\mathrm{s}_{\mathrm{X}}^2 = \frac{\mathrm{n}}{\mathrm{n}-1}\sigma_{\mathrm{X}}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

При большом объеме выборки n отличие между этими оценками пренебрежимо мало.

Рассмотрим точечную оценку m_{xy} для корреляционного момента μ_{XY} и точечную оценку r_{XY} для коэффициента корреляции ρ_{XY} случайных величин *X*, *Y*, определяемых по выборке объемом n. Оценки вычисляются по следующим формулам:

$$m_{XY} = \overline{xy} - \bar{x}\cdot\bar{y}$$

$$r_{XY} = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{s_X \cdot s_Y}$$

где $\overline{xy} = \frac{1}{\mathrm{n}}\sum_{i=1}^{n}(x_i \cdot \mathrm{y_i})$

Методы получения точечных оценок параметров

Для нахождения оценок характеристик генеральной совокупности использует ряд методов: метод наибольшего правдоподобия; **м**етод моментов; метод наименьших квадратов.

§5. Распределения, связанные с нормальным законом распределения

Генеральные совокупности часто имеют нормальный закон распределения. В этом случае многие выборочные характеристики выражаются через небольшое число распределений. Как правило, в математической статистике используются не плотности этих распределений, а некоторые характеристики, представленные таблицами. Чаще всего в качестве такой характеристики выступает ***квантиль распределения***.

> ***Квантилем уровня*** p $(0 < p < 1)$ или ***p-квантилем*** случайной величины X называется такое число d_p, что вероятность $P(X < d_p)$ равна заданной величине p.

Рассмотрим несколько распределений, которым подчиняются выборочные характеристики и которые используются для построения интервальных оценок.

Распределение χ^2- квадрат («хи - квадрат») или распределение Пирсона

> Распределением χ^2 с n степенями свободы называется распределение суммы квадратов n независимых случайных величин, распределенных по стандартному закону:
>
> $$\chi^2(n) = X_1^2 + X_2^2 + \cdots + X_n^2$$

Плотность χ^2- распределения:

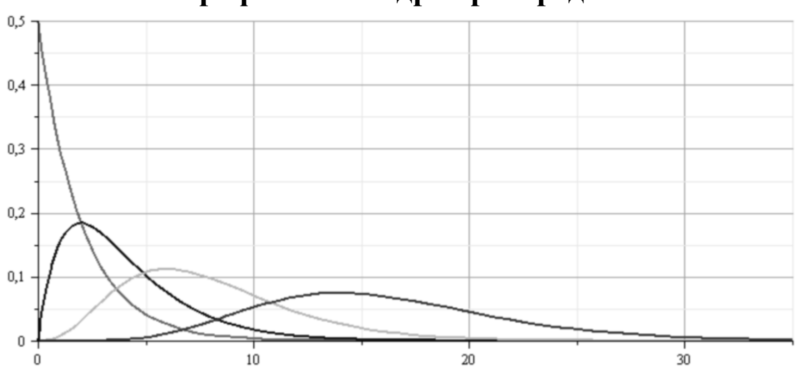
$$f(x) = \begin{cases} 0, & x \le 0 \\ \dfrac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} e^{-\frac{x}{2}} x^{\frac{k}{2}-1}, & x > 0, \end{cases}$$

где $\Gamma(x) = \int\limits_0^\infty t^{x-1} e^{-t} dt -$ гамма $-$ функция Эйлера.

Свойства

1. Всегда неотрицательно.
2. Зависит от k - числа степеней свободы.
3. Математическое ожидание $M(\chi^2) = n$.
4. Дисперсия $D(\chi^2) = 4n^2$.
5. С увеличением числа степеней свободы χ^2 - распределение стремится к нормальному.

График хи-квадрат распределения



Все кривые обладают положительной асимметрией (смещение кривой в правую сторону). С увеличением числа степеней свободы распределение постепенно приближается к нормальному.

Остановимся более подробно на понятии **«число степеней свободы».** Дело в том, что оценка вариации зависит, не только от объема выборки, но и от того, как много параметров должно оцениваться в выборке.

Чем больше данных, тем больше мы можем доверять полученным результатам. Чем больше параметров мы должны оценить, тем меньше мы им доверяем. Эти два момента в статистике учитываются при вычислении числа степеней свободы:

число степеней свободы = число наблюдений - число параметров, которые должны быть оценены заранее.

Смысл числа степеней свободы покажем с помощью следующего примера:

Пример 8. Менеджер компании имеет бюджет 150 млн. руб. на четыре различных проекта. Сколькими степенями свободы располагает менеджер?

Решение примера 8. Общий бюджет 4-х различных проектов можно рассматривать как их среднюю арифметическую, умноженную на число проектов:

$$x_1 + x_2 + x_3 + x_4 = n \cdot \bar{X}$$

Когда мы говорим, что бюджет для 4-х проектов составляет 150 млн. руб., это то же самое как определение среднего 150 000/4=37,5 млн. руб.

Менеджер имеет 3 возможности распределения средств на любые 3 из 4 проекта.

Как только на первые 3 проекта средства распределены, ему не остается выбора при распределении средств на 4 проект.

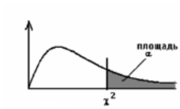
Он может выделить на него только средства, равные разности между 150 млн. руб. (общим бюджетом) и суммой, выделенной на 3 предыдущих проекта.

Следовательно, менеджер располагает 3 степенями свободы.

Функция распределения имеет сложное аналитическое выражение, поэтому для непосредственного подсчета вероятностей используется специальная таблица (см. Приложение 4).

Нахождение табличных значений

Например, при $\alpha = 0{,}05$ и числа степеней свободы, равного 10, критическая точка $\chi^2_{0,05;10} = 18{,}3070$.



Приложение 4

Критические значения распределения Пирсона

Число степеней свободы k	Уровень значимости					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,6349	5,0239	3,8415	0,0039	0,0010	0,0002
2	9,2103	7,3778	5,9915	0,1026	0,0506	0,0201
3	11,3449	9,3484	7,8147	0,3519	0,2158	0,1148
4	13,2767	11,1433	9,4877	0,7107	0,4844	0,2971
5	15,0863	12,8325	11,0705	1,1455	0,8312	0,5543
6	16,8119	14,4494	12,5916	1,6354	1,2373	0,8721
7	18,4753	16,0128	14,0671	2,1674	1,6899	1,2390
8	20,0902	17,5346	15,5073	2,7326	2,1797	1,6465
9	21,6660	19,0228	16,9190	3,3251	2,7004	2,0879
10	23,2093	20,4832	18,3070	3,9403	3,2470	2,5582
11	24,7250	21,9201	19,6751	4,5748	3,8158	3,0535

t-распределение или распределение Стьюдента

Распределением Стьюдента (t-распределением) называется распределение случайной величины:

$$t = \frac{Y}{\sqrt{\frac{1}{k}\chi^2(k)}},$$

где Y - имеет стандартное нормальное распределение, а $\chi^2(k)$ - распределение хи-квадрат с k степенями свободы.

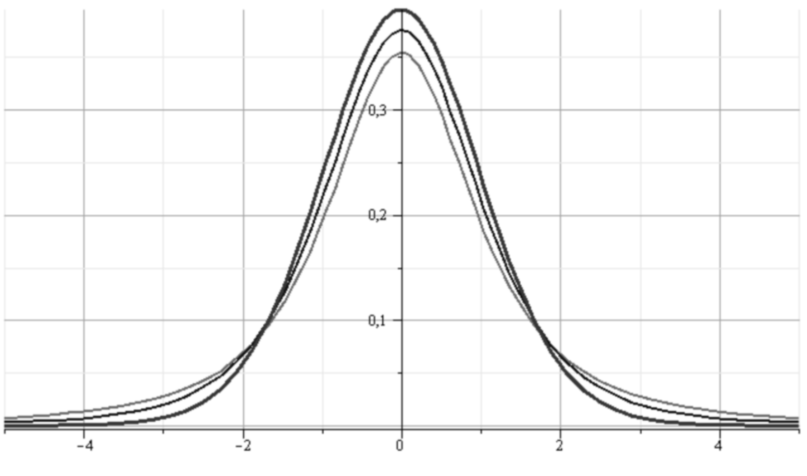
Распределение Стьюдента было введено в 1908 году В.С. Госсетом, ирландским служащим пивоваренного завода, который участвовал в разработке новых технологий производства пива.

Поскольку самостоятельно публиковать результаты исследований работникам завода не разрешалось, Госсет напечатал свои материалы под псевдонимом Стьюдент, поэтому t-распределение часто называют распределением Стьюдента.

Свойства

1. Симметрично относительно вертикальной оси.
2. Зависит от k - числа степеней свободы.
3. Математическое ожидание $M(t) = 0$.
4. Дисперсия $D(t) = \frac{n}{n-2}$.
5. С увеличением числа степеней свободы t - распределение стремится к нормальному.

График распределения Стьюдента



Аналитическое выражение функции плотности распределения Стьюдента имеет довольно сложную форму записи. Для определения вероятностей случайных величин, подчиняющихся распределению Стьюдента, обычно пользуются готовыми таблицами (см. Приложение 2).

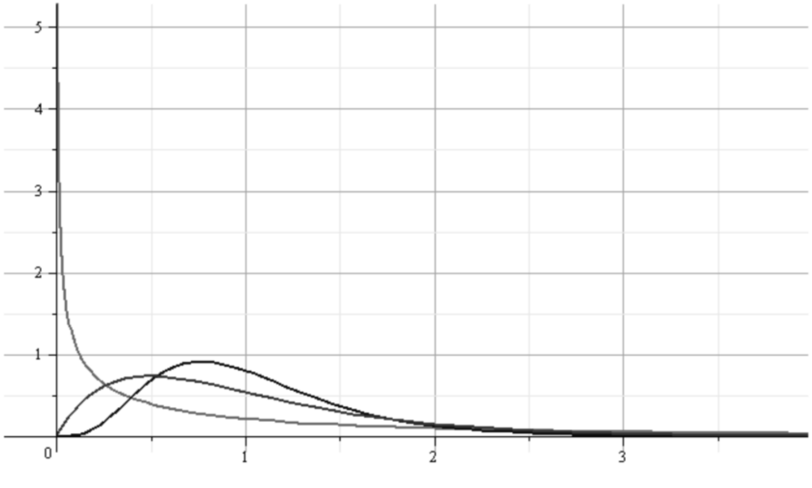
Распределение Фишера

Распределением Фишера называется распределение случайной величины:

$$F = \frac{\frac{1}{k_1}\chi^2(k_1)}{\frac{1}{k_2}\chi^2(k_2)},$$

где $\chi^2(k_1), \chi^2(k_2)$ – случайные величины, имеющие распределение хи-квадрат с k_1 и k_2 степенями свободы соответственно.

График распределения Фишера



Для определения вероятностей случайных величин, подчиняющихся распределению Фишера, обычно пользуются готовыми таблицами (см. Приложение 3)

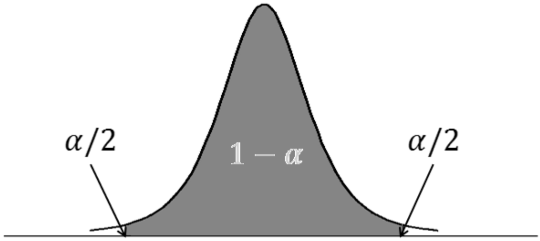
§6 Интервальные оценки параметров генеральной совокупности

После получения точечной оценки желательно иметь данные о надежности такой оценки. Особенно важно иметь сведения о точности оценок для небольших выборок. Поэтому точечная оценка может быть дополнена интервальной оценкой - интервалом $(\theta_1;\theta_2)$, внутри которого, с наперед заданной вероятностью γ, находится точное значение оцениваемого параметра θ.

Задачу определения такого интервала называют ***интервальным оцениванием***, а сам интервал − ***доверительным интервалом***. При этом $\gamma = 1 - \alpha$- называется **дове-**

рительной вероятностью, или надежностью, или уровнем доверия, α – **уровень значимости**.

Доверительная вероятность представляет собой площадь под графиком (1 - α).



Сам интервал называется доверительный интервал или интервальной оценкой параметра θ.

Задача состоит в поиске такого значения ε, что выполняется следующее равенство:

$$P\left(\left|\theta-\bar{\theta}\right|<\varepsilon\right)=\gamma$$

или

$$P(\bar{\theta}-\varepsilon<\theta<\bar{\theta}+\varepsilon)=\gamma$$

Пусть $\theta_1=\bar{\theta}-\varepsilon$, $\theta_2=\bar{\theta}+\varepsilon$. Тогда доверительный интервал – это интервал следующего вида:

$$I_\gamma=(\theta_1;\theta_2)$$

ε - называется «точность оценки», или «предельная ошибка выборки».

Выбор α (или $\gamma=1-\alpha$) определяется конкретными условиями. Обычно используется $\alpha=0{,}1; 0{,}05; 0{,}01$, что соответствует 90, 95, 99%-ным доверительным интервалам.

Доверительный интервал для среднего

Построим доверительный интервал для среднего генеральной совокупности, имеющей нормальный закон распределения.

Предположим, у нас имеется простая случайная выборка из этой генеральной совокупности. Объем выборки равен n. Требуется построить доверительный интервал, который с доверительной вероятностью будет содержать среднее генеральной совокупности:

$$I_\gamma = (\bar{x} - \varepsilon; \bar{x} + \varepsilon)$$

По выборке мы можем вычислить выборочное среднее. Это есть точечная оценка для среднего генеральной совокупности. Наша задача сводится к вычислению **точности интервальной оценки** ε, что позволит вычислить границы доверительного интервала. При построении доверительного интервала мы будем основываться на известных нам свойствах нормального закона распределения.

Существует два различных случая, которые мы последовательно рассмотрим.

Первый случай: σ^2 известна или $n \geq 30$

Цель. Оценить среднее для генеральной совокупности, имеющей нормальный закон распределения с параметрами m,σ.

Что мы имеем. Имеем случайную выборку объема n из генеральной совокупности. Стандартное отклонение σ предполагается известным или объем выборки $n \geq 30$.

Требуется. Построить доверительный интервал для среднего:

$$I_\gamma = (\bar{x} - \varepsilon; \bar{x} + \varepsilon), \varepsilon = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Метод:

1. По выборке вычислить выборочное среднее.

2. По таблице нормального закона найти z-значение для доверительной вероятности $1-\alpha$.
3. Вычислить точность интервальной оценки по формуле:

$$\varepsilon = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Если значение σ неизвестно в случае $n \geq 30$, тогда вместо σ в формулу подставляется ее выборочная оценка s.
4. Вычислить доверительный интервал. Написать ответ.

Второй случай: σ^2 неизвестна и $n \leq 30$

Цель. Оценить среднее для генеральной совокупности, имеющей нормальный закон распределения с параметрами m,σ.

Что мы имеем. Имеем случайную выборку объема n из генеральной совокупности. Стандартное отклонение σ неизвестно или объем выборки $n \leq 30$.

Требуется. Построить доверительный интервал для среднего:

$$I_\gamma = (\bar{x}-\varepsilon; \bar{x}+\varepsilon), \varepsilon = t_{\text{кр}}\frac{s}{\sqrt{n}}$$

Метод:

1. По выборке вычислить выборочное среднее и стандартное отклонение.
2. По таблице t-распределения найти $t_{\text{кр}}$ для доверительной вероятности $1-\alpha$ и числа степеней свободы $df = n-1$ (двусторонняя критическая область).

3. Вычислить точность интервальной оценки по формуле:

$$\varepsilon = t_{\text{кр}} \frac{s}{\sqrt{n}}$$

4. Вычислить доверительный интервал. Написать ответ.

Доверительный интервал для дисперсии

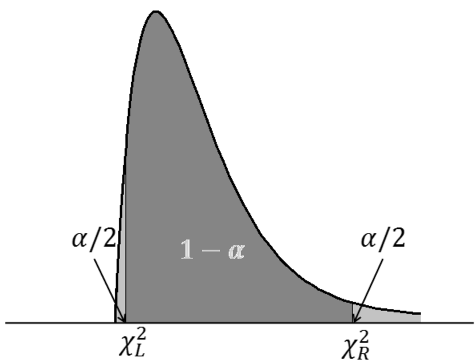
Построим доверительный интервал для неизвестной дисперсии нормально распределенной генеральной совокупности. Оценкой для генеральной дисперсии является выборочная дисперсия. **Доверительный интервал для дисперсии** находится по следующей формуле:

$$I_\gamma = \left(\frac{(n-1) \cdot s^2}{\chi_R^2} ; \frac{(n-1) \cdot s^2}{\chi_L^2} \right)$$

Значения χ_L^2 и χ_R^2 находятся по таблице критических точек распределения Пирсона, исходя из следующих условий:

$$P(\chi^2 > \chi_L^2) = 1 - \frac{\alpha}{2}$$

$$P(\chi^2 > \chi_R^2) = \frac{\alpha}{2}$$



Доверительный интервал для стандартного отклонения имеет вид:

$$I_\gamma = \left(s\sqrt{\frac{n-1}{\chi_R^2}}; s\sqrt{\frac{n-1}{\chi_L^2}} \right)$$

Метод:

1. По выборке вычислить дисперсию.
2. По таблице хи-квадрат распределения найти значения χ_L^2 и χ_R^2 для доверительной вероятности $1-\alpha$ и числа степеней свободы $df = n - 1$.
3. Вычислить доверительный интервал. Написать ответ.

§7 Проверка статистических гипотез

Общие принципы проверки гипотез

Перед тем, как рассматривать и проверять конкретные гипотезы, нам необходимо привести основные понятия, а также обсудить общие принципы, которые применяются при проверке статистических гипотез.

Гипотеза – утверждение, которое надо либо доказать, подтвердить, исходя из разумных предположений, либо опровергнуть.

Статистической гипотезой называют гипотезу о виде неизвестного распределения или о параметрах известного распределения.

Проверка статистических гипотез состоит из шести этапов:

Формулируем основную и альтернативную гипотезы

↓

Задаем уровень значимости

↓

Выбираем статистику - критерий проверки гипотезы

↓

Определяем критическую область

↓

Вычисляем значение статистики по выборке

↓

Сравниваем значение статистики с критической областью, делаем вывод

Основная и альтернативная гипотезы

Проверяемая гипотеза в статистике называется **основной (или нулевой) гипотезой**.
Обозначение: H_0

Основная гипотеза H_0 подлежит проверке, по результатам которой ее можно принять либо отклонить. «Принять» означает «не получить убедительных аргументов для «отклонения гипотезы».

Альтернативной (или конкурирующей) называют гипотезу, которая противоречит гипотезе H_0.
Обозначение: H_1.

Альтернативная гипотеза H_1 принимается только тогда, когда есть убедительное статистическое доказательство для отклонения основной гипотезы.

В качестве иллюстрации приведем пример:

Пример 9. Менеджер бюро переводов хочет снизить расходы компании на канцелярские принадлежности. В среднем эти расходы составляют 5 300 рублей в неделю. После принятия определенных мер по экономии бумаги и скрепок менеджер хотел бы проверить, снизились ли расходы или остались на прежнем уровне. Гипотезы о величине расходов выглядят следующим образом:

$$H_0: m \geq 5300$$
$$H_1: m < 5300$$

Виды критериев

От решаемой задачи зависит, какой из критериев будет выбран:

Двусторонний	Левосторонний	Правосторонний
$\boldsymbol{H_0}: =$ $\boldsymbol{H_1}: \neq$	$\boldsymbol{H_0}: \geq$ $\boldsymbol{H_1}: <$	$\boldsymbol{H_0}: \leq$ $\boldsymbol{H_1}: >$

Выбор гипотез зависит от задачи, которую решает исследователь. В реальных исследованиях, чтобы получить результаты, которые окажутся важными и полезными, приходится не один раз формулировать и проверять различные гипотезы. Это выглядит как сложный путь проб и ошибок. Исследователь напоминает в этот момент скульптора, который отсекает от камня лишнее, оставляя только безукоризненные формы своего будущего творения.

Чтобы быть корректными, статистики в случае принятия основной гипотезы не утверждают, что она верна, а говорят, что в результате эксперимента не появилось оснований ее отклонить. При принятии альтернативной гипотезы статистики с уверенностью отклоняют основную гипотезу раз и навсегда, окончательно и бесповоротно. В этом случае статистическими методами нельзя доказать какое-либо утверждение, можно лишь отвергнуть ошибочные предположения.

Ошибки первого и второго рода

Статистические гипотезы проверяются статистическими методами, на основании выборки, полученной из генеральной совокупности. Из-за случайности выборки в результате проверки могут возникать ошибки и приниматься неправильные решения.

Назовем ошибкой первого рода ситуацию, в которой мы отвергаем верную гипотезу H_0. При ошибке второго рода принимается гипотеза H_0 в то время, как она неверна.

Ошибка первого рода (type I error) происходит, если мы отвергаем верную нулевую гипотезу.

Ошибка второго рода (type II error) происходит, если мы принимаем нулевую гипотезу, когда она неверна.

	Основная гипотеза верна	**Основная гипотеза неверна**
Мы приняли основную гипотезу	**Верное решение**	**Ошибка II рода**
Мы отклонили основную гипотезу	**Ошибка I рода**	**Верное решение**

Уровень значимости

Поскольку существуют ненулевые вероятности совершить ошибки первого или второго рода, следует устанавливать разумные значения этих вероятностей. До начала эксперимента принято задавать значение вероятности ошибки первого рода, называемое уровнем значимости.

Уровнем значимости (level of significance) гипотезы называют вероятность совершить ошибку первого рода, то есть отклонить верную нулевую гипотезу.

Обозначение: α.

Значение α обычно выбирают небольшим: 10%, 5% или 1%.

Желательно, чтобы вероятность совершения ошибок была минимальна. Тем не менее, уровень значимости не может быть выбран очень малым, поскольку уменьшение вероятности совершить ошибку первого рода вызывает увеличение вероятности совершить ошибку второго рода.

Статистический критерий

Статистическим критерием или просто критерием называют случайную величину K, которая служит для проверки гипотезы H_0.

Множество значений статистики включает два непересекающихся подмножества: ***область принятия гипотезы***, то есть множество тех значений статистики, при которых гипотеза H_0 принимается, и ***критическую область***, то есть множество тех значений статистики, при которых гипотеза H_0 отклоняется и принимается альтернативная гипотеза H_1.

Областью принятия гипотезы (областью допустимых значений критерия) называют совокупность значений критерия, при которых нулевую гипотезу принимают.

Критической областью называют совокупность значений критерия, при которых нулевую гипотезу опровергают.

Критические точки отделяют критическую область от области принятия гипотезы.

Критическая область

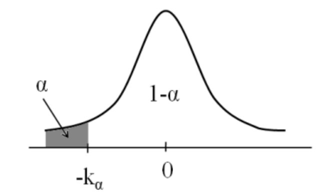
Критическая область строится, исходя из имеющихся знаний о законе распределения статистики, и зависит от:

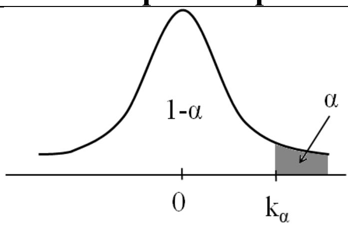
-объема выборки,

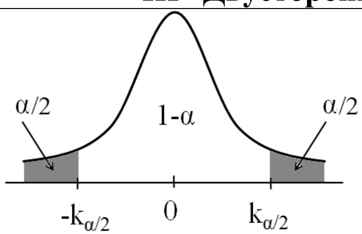
-уровня значимости, задаваемого исследователем,

-вида альтернативной гипотезы.

Различают одностороннюю и двустороннюю критическую область. В свою очередь, односторонняя критическая область может быть левосторонней и правосторонней.

<table>
<tr><th colspan="2">I – Левосторонняя критическая область</th></tr>
<tr><td>
</td><td>Критическая область определяется уравнением:

$$P(k < -k_\alpha) = \alpha$$</td></tr>
<tr><th colspan="2">II – Правосторонняя критическая область</th></tr>
<tr><td>
</td><td>Критическая область определяется уравнением:

$$P(k > k_\alpha) = \alpha$$</td></tr>
<tr><th colspan="2">III –Двусторонняя критическая область</th></tr>
<tr><td>
</td><td>Критическая область определяется уравнением:

$$P\left(k > k_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2},$$

$$P(k < -k_{\alpha/2}) = \alpha/2$$</td></tr>
</table>

Критические точки находятся по таблицам.

После построения критической области значение статистики по выборке сравнивают с критической областью. Если значение статистики попало в область принятия гипотезы, то основная гипотеза принимается. Если значение статистики попало в критическую область, то основная гипотеза отклоняется и принимается альтернативная гипотеза.

Гипотеза о среднем

Первый случай: σ^2 известна или $n \geq 30$

Что мы имеем?

Имеется генеральная совокупность с нормальным законом распределения. Параметры m и σ.

Что мы хотим?

На основе анализа простой случайной выборки проверить гипотезу о среднем значении генеральной совокупности m.

Гипотезы:

Альтернативные гипотезы могут быть трех различных видов:

I	II	III
Нулевая гипотеза: $H_0: m = m_0$ Альтернативная гипотеза: $H_1: m < m_0$	Нулевая гипотеза: $H_0: m = m_0$ Альтернативная гипотеза: $H_1: m > m_0$	Нулевая гипотеза: $H_0: m = m_0$ Альтернативная гипотеза: $H_1: m \neq m_0$

Критерий:

$$z = \frac{\bar{x} - m_0}{\sigma/\sqrt{n}}$$

Здесь $\bar{x}$ – среднее значение выборки, m_0 – гипотетическое генеральное среднее, σ – стандартное отклонение генеральной совокупности, n – объем выборки.

Используемая статистика имеет стандартное нормальное распределение.

Критическая область:

Построим критическую область для каждого из трех вариантов гипотез.

I – Левосторонняя критическая область
Нулевая гипотеза: $$H_0: m = m_0$$ Альтернативная гипотеза: $$H_1: m < m_0$$ Критическая область: $$U: z < -z_\alpha$$ Критическое значение находится по таблице значений функции Лапласа (см. приложение 1): $$\Phi_0(z_\alpha) = 0{,}5 - \alpha$$
II – Правосторонняя критическая область
Нулевая гипотеза: $$H_0: m = m_0$$ Альтернативная гипотеза: $$H_1: m > m_0$$ Критическая область: $$U: z > z_\alpha$$ Критическое значение находится по таблице значений функции Лапласа (см. приложение 1): $$\Phi_0(z_\alpha) = 0{,}5 - \alpha$$

III –Двусторонняя критическая область
Нулевая гипотеза: $$H_0: m = m_0$$ Альтернативная гипотеза: $$H_1: m \neq m_0$$ Критическая область: $$U: \lvert z \rvert > z_{\alpha/2}$$ Критическое значение находится по таблице значений функции Лапласа(см. приложение 1): $$\Phi_0\left(z_{\alpha/2}\right) = 0,5 - \alpha/2$$

Правило принятия гипотезы:

> Если значения критерия z попадает в критическую область, то гипотеза H_0 отвергается.

Второй случай: $\boldsymbol{\sigma^2}$ неизвестно и $\boldsymbol{n} < 30$

В первом случае проверка гипотезы о среднем проводилась при условии, что нам известно стандартное отклонение генеральной совокупности. Теперь рассмотрим проверку гипотезы, если стандартное отклонение неизвестно. Вместо стандартного отклонения мы будем использовать стандартное отклонение выборки.

Если объем выборки достаточно велик ($n \geq 30$), мы использовали z-статистику.

Критерий:

$$t = \frac{\bar{x} - m_0}{s/\sqrt{n-1}}$$

Здесь $\bar{x}$ – среднее значение выборки, m_0 – гипотетическое генеральное среднее, s – стандартное отклонение выборки, n – объем выборки

Используемая статистика имеет t-распределение Стьюдента с числом степеней свободы $df = n - 1$.

Критическая область:

Построим критическую область для каждого из трех вариантов гипотез.

I – Левосторонняя критическая область

Нулевая гипотеза:

$$\boldsymbol{H_0: m = m_0}$$

Альтернативная гипотеза:

$$\boldsymbol{H_1: m < m_0}$$

Критическая область:

$$\boldsymbol{U: t < -t_\alpha}$$

Критическое значение $\mathbf{t}_{\boldsymbol{\alpha}}$ находится по таблице критических значений t-критерия Стьюдента (см. приложение 2, односторонняя критическая область) по заданному уровню значимости $\boldsymbol{\alpha}$ и числу степеней свободы $\boldsymbol{df = n - 1}$.

II – Правосторонняя критическая область

Нулевая гипотеза:

$$\boldsymbol{H_0: m = m_0}$$

Альтернативная гипотеза:

$$\boldsymbol{H_1: m > m_0}$$

Критическая область:

$$\boldsymbol{U: t > t_\alpha}$$

Критическое значение $\mathbf{t}_{\boldsymbol{\alpha}}$ находится по таблице критических значений t-критерия Стьюдента (см. приложение 2, односторонняя критическая область) по заданному уровню значимости $\boldsymbol{\alpha}$ и числу степеней свободы $\boldsymbol{df = n - 1}$.

III –Двусторонняя критическая область
Нулевая гипотеза: $$\boldsymbol{H_0: m = m_0}$$ Альтернативная гипотеза: $$\boldsymbol{H_1: m \neq m_0}$$ Критическая область: $$\boldsymbol{U: \lvert t \rvert > t_{\alpha/2}}$$ Критическое значение $\mathbf{t}_{\boldsymbol{\alpha/2}}$ находится по таблице критических значений t-критерия Стьюдента (см. приложение 2, двусторонняя критическая область) по заданному уровню значимости $\boldsymbol{\alpha}$ и числу степеней свободы $\boldsymbol{df = n - 1}$.

Правило принятия гипотезы:

Если значения критерия t попадает в критическую область, то гипотеза H_0 отвергается.

Пример 10. Менеджер кредитного отдела нефтяной компании хотел бы выяснить, является ли среднемесячный баланс владельцев кредитных карточек равным 75 тыс. руб.
Для анализа он случайным образом отобрал 100 счетов и нашел, что среднемесячный баланс владельцев составил 83,4 тыс. руб. с выборочным стандартным отклонением, равным 23,65 тыс. руб.
Определить на 5% уровне значимости, может ли этот менеджер утверждать, что средний баланс отличен от 75 тыс. руб.

Решение примера 10.

1) Исходя из условия задачи сформулируем гипотезы:

$$H_0: m = 75 \text{ тыс. руб.}$$

$$H_1: m \neq 75 \text{ тыс. руб.}$$

σ^2 – неизвестно, объем $n > 30$.

Уровень значимости $\alpha = 0{,}05$.

2) Для проверки гипотезы H_0 применим критерий $z = \frac{\bar{x} - m_0}{\sigma/\sqrt{n}}$ с двусторонней критической областью.

3) Вычислим z:

$$z = \frac{\bar{x} - m_0}{\sigma/\sqrt{n}} = ?$$

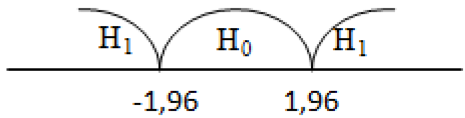
$$m_0 = 75; n = 100; \sigma = s = 23{,}65; \bar{x} = 83{,}4$$

$$z = \frac{83{,}4 - 75}{23{,}65/\sqrt{100}} = 3{,}5518$$

4) Критическая область: U: $|\boldsymbol{z}| > \boldsymbol{z_{\alpha/2}}$

$$\boldsymbol{\Phi_0(z_{\alpha/2}) = 0{,}5 - \alpha/2} \Rightarrow \boldsymbol{\Phi_0(z_{\alpha/2})} = 0{,}475 \Rightarrow \boldsymbol{z_{\alpha/2}} = 1{,}96$$

5)



$$z \in U$$

Нулевая гипотеза отвергается, т.е. на 5% уровне значимости предположение о балансе карт неверно.

Гипотеза о дисперсии

Что мы имеем?

Имеется генеральная совокупность с нормальным законом распределения. Параметры m и σ.

Что мы хотим?

На основе анализа простой случайной выборки проверить гипотезу о значении неизвестной дисперсии генеральной совокупности σ^2.

Гипотезы:

Альтернативные гипотезы могут быть трех различных видов:

I	II	III
Нулевая гипотеза: $H_0: \sigma^2 = \sigma_0^2$ Альтернативная гипотеза: $H_1: \sigma^2 < \sigma_0^2$	Нулевая гипотеза: $H_0: \sigma^2 = \sigma_0^2$ Альтернативная гипотеза: $H_1: \sigma^2 > \sigma_0^2$	Нулевая гипотеза: $H_0: \sigma^2 = \sigma_0^2$ Альтернативная гипотеза: $H_1: \sigma^2 \neq \sigma_0^2$

Для проверки гипотезы используется следующий **критерий:**

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

Здесь σ_0 – гипотетическое генеральное стандартное отклонение, s – стандартное отклонение выборки, n – объем выборки

Эта статистика имеет χ^2-распределение с числом степеней свободы $df = n - 1$.

Критическая область:

Построим критическую область для каждого из трех вариантов гипотез.

I – Левосторонняя критическая область
Альтернативная гипотеза: $$H_1: \sigma^2 < \sigma_0^2$$ Критическая область: $$U: \chi^2 < \chi^2{}_{1-\alpha}$$ Критическое значение $\chi^2{}_{1-\alpha}$ находится по таблице критических значений распределения Пирсона (см. приложение 4) по заданному уровню значимости $1-\alpha$ и числу степеней свободы $df = n-1$.
II – Правосторонняя критическая область
Альтернативная гипотеза: $$H_1: \sigma^2 > \sigma_0^2$$ Критическая область: $$U: \chi^2 > \chi^2{}_{\alpha}$$ Критическое значение $\chi^2{}_{\alpha}$ находится по таблице критических значений распределения Пирсона (см. приложение 4) по заданному уровню значимости α и числу степеней свободы $df = n-1$.
III –Двусторонняя критическая область
Альтернативная гипотеза: $$H_1: \sigma^2 \neq \sigma_0^2$$ Критическая область: $$U: \chi^2 < \chi^2{}_{1-\alpha} \text{ и } \chi^2 > \chi^2{}_{\alpha}$$ Критические значения $\chi^2{}_{\alpha}$, $\chi^2{}_{1-\alpha}$ находится по таблице критических значений распределения Пирсона (см. приложение 4) по заданному уровню значимости, соответственно, α и $1-\alpha$ и числу степеней свободы $df = n-1$.

Замечание. Если $k > 30$, то

$$\chi_{\alpha}^{2} = k\left(1 - \frac{2}{9k} + z_{\alpha}\sqrt{\frac{2}{9k}}\right)^{3}$$

$$\Phi_0(z_\alpha) = 0{,}5 - \alpha$$

Правило принятия гипотезы:

Если значения критерия χ^2 попадает в критическую область, то гипотеза H_0 отвергается.

Пример 11. Считается, что среднее значение IQ теста равно 100 со стандартным отклонением 15. Предположим, мы хотим проверить гипотезу о дисперсии. Создали случайную выборку из 20 человек и вычислили стандартное отклонение этой выборки. Пусть оно оказалось равно 19,4.

Есть ли у нас основания считать, что стандартное отклонение генеральной совокупности отличается от 15? Требуется проверить на уровне значимости 5%.

Решение примера 11.

1) Исходя из условия задачи сформулируем гипотезы:

$$H_0\colon \sigma^2 = 15^2$$

$$H_1\colon \sigma^2 > 15^2$$

Уровень значимости $\alpha = 0{,}05$.

2) Для проверки гипотезы H_0 применим критерий $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$

3) Вычислим χ^2:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} =?$$

$$n = 20; s = 19{,}4;\ \sigma_0 = 15$$

$$z = \frac{19 \cdot 19{,}4^2}{15^2} = 31{,}7815$$

4) Критическая область: $U: \chi^2 > {\chi^2}_{\boldsymbol{\alpha}}$

$${\chi^2}_{\boldsymbol{\alpha}} = \chi^2_{\text{кр}}(0{,}05; 19) = 30{,}1435$$

5) $\chi^2 \in U = (30{,}1435; +\infty)$

Нулевая гипотеза отвергается, т.е. на 5% уровне значимости у нас есть основание считать, что стандартное отклонение генеральной совокупности отличается от 15.

Гипотеза о равенстве двух средних

Что мы имеем?

1. Две генеральные совокупности, из которых получены простые случайные выборки. Выборки независимые.

2. Обе выборки имеют объем $n \geq 30$. Если нет, то обе выборки взяты из нормально распределенных генеральных совокупностей.

Что мы хотим?

Проверить гипотезу о равенстве средних двух исследуемых генеральных совокупностей.

Нулевая гипотеза:

$$H_0: m_1 = m_2$$

Альтернативная гипотеза:

$$H_1: m_1 \neq m_2$$

Нулевая гипотеза может быть записана иначе:

$$H_0: m_1 - m_2 = 0$$

Такая гипотеза называется гипотезой о разности средних и может быть проверена для любого значения разности, в том числе не нулевого.

Мы ограничим себя исключительно гипотезой о равенстве средних.

По виду альтернативной гипотезы можно заключить, что критическая область будет **двусторонней**.

Для сравнения средних существуют также односторонние критерии, **правосторонний**:

$$H_1\colon m_1 > m_2$$

и **левосторонний**:

$$H_1\colon m_1 < m_2$$

Наблюдаемым значением является разность двух выборочных средних, ожидаемым значением – разность двух генеральных средних. В нашем случае, при проверке гипотезы о равенстве средних, ожидаемое значение равно нулю. Стандартная ошибка вычисляется по-разному, в зависимости от того, известным нам дисперсии двух генеральных совокупностей или не известны.

Мы рассмотрим три различные ситуации и соответствующий каждой из них метод проверки гипотез и построения доверительных интервалов:

- Дисперсии генеральных совокупностей известны.
- Дисперсии не известны, но у нас есть основания считать их равными.
- Дисперсии неизвестны и мы не можем считать их равными.

Ситуация 1. Дисперсии известны

В этом случае для проверки гипотезы применяется статистика:

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Здесь

$\bar{x}_1, \bar{x}_2$ – выборочные средние,

σ_1^2, σ_2^2 – известные генеральные дисперсии,

n_1, n_2 – объемы выборок.

Статистика имеет стандартное нормальное распределение.

Ситуация 2. Дисперсии не известны, но равны

В этом случае для проверки гипотезы применяется статистика:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

Здесь

$\bar{x}_1, \bar{x}_2$ – выборочные средние,

s_1^2, s_2^2 – выборочные дисперсии,

n_1, n_2 – объемы выборок.

Используемая статистика имеет t-распределение Стьюдента с числом степеней свободы $df = n_1 + n_2 - 2$.

Ситуация 3. Дисперсии не известны и не предполагаются равными

В этом случае для проверки гипотезы применяется статистика:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Здесь

$\bar{x}_1, \bar{x}_2$ – выборочные средние,

s_1^2, s_2^2 – выборочные дисперсии,

n_1, n_2 – объемы выборок.

Используемая статистика имеет t-распределение Стьюдента с числом степеней свободы $df = \min(n_1 - 1; n_2 - 1)$.

Пример 12. Исследователь предполагает, что среди студентов университета девушки чаще, чем юноши, прогуливают занятия. Выборочное исследование 16-ти девушек показало, что их не бывает в университете примерно 3,9 дня в году, а юношей (22 человека) 3,6 дня. Стандартные отклонения составили 0,6 и 0,8 дня соответственно. Проверьте предположение исследователя на уровне значимости α=0,01. Предполагается, что дисперсии равны.

Решение примера 12.

1) Исходя из условия задачи сформулируем гипотезы:

$$H_0: m_1 \leq m_2$$
$$H_1: m_1 > m_2$$

Здесь

m_1 - генеральное среднее для прогулов девушек,

m_2 - генеральное среднее для прогулов юношей.

Уровень значимости $\alpha = 0{,}01$.

Дисперсии считаются равными, но неизвестны.

2) Для проверки гипотезы H_0 применим критерий

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

3) Вычислим t:

$$n_1 = 16; n_2 = 22; s_1 = 0{,}6; s_2 = 0{,}8; m_1 = 3{,}9; m_2 = 3{,}6$$

$$t = \frac{3{,}9 - 3{,}6}{\sqrt{(16-1)\cdot 0{,}6^2 + (22-1)\cdot 0{,}8^2}} \cdot \sqrt{\frac{16\cdot 22\cdot(16+22-2)}{16+22}} = \frac{0{,}3}{\sqrt{18{,}84}}\sqrt{\frac{12672}{38}} = 0{,}3\cdot 4{,}2072 = 1{,}2622$$

4) Критическая область - **правосторонняя**: $U: t > t_\alpha$

$t_\alpha = t_{\text{кр}}(0{,}01; 15+22-2) = t_{\text{кр}}(0{,}01; 35) = 2{,}42$

5) $t \notin U = (2{,}42; +\infty)$

Нулевая гипотеза принимается, т.е. на 1% уровне значимости у нас нет оснований думать, что девушки прогуливают занятия больше юношей.

Гипотеза о равенстве двух дисперсий

Что мы имеем?

1. Две простые случайные выборки, полученные из двух нормально распределенных генеральных совокупностей.

2. Выборки являются независимыми. Это значит, что между субъектами выборок нет связи.

Что мы хотим?

Проверить гипотезу о равенстве дисперсий генеральных совокупностей.

Нулевая гипотеза:

$$H_0: \sigma_1^2 = \sigma_2^2$$

Здесь

σ_1^2, σ_2^2 – генеральные дисперсии,

s_1^2, s_2^2 – выборочные дисперсии,

n_1, n_2 – объемы выборок.

Для того, чтобы воспользоваться критерием необходимо, чтобы $s_1^2 > s_2^2$

Альтернативная гипотеза:

$$H_1: \sigma_1^2 > \sigma_2^2$$

Другие альтернативные гипотезы в данном случае не рассматриваются.

В качестве **критерия** проверки этой гипотезы применяется статистика:

$$F = \frac{s_1^2}{s_2^2}$$

Эта статистика имеет F-распределение с числом степеней свободы $df_1 = n_1 - 1$ и $df_2 = n_2 - 1$.

Критическая область – правосторонняя.

Уравнение критической области:

$$P\left(F > F_{\text{кр}}\right) = \alpha$$

Критическая точка $F_{\text{кр}}$ находится по таблице критических значений распределения Фишера-Снедекора (см. приложение 3) по заданному уровню значимости α и числу степеней свободы $df_1 = n_1 - 1$ и $df_2 = n_2 - 1$.

Правило принятия гипотезы:

Если значения критерия F попадает в критическую область, то гипотеза H_0 отвергается.

§8 Задачи

Вариационные ряды и их характеристики

Задача 1. Имеются выборочные данные (табл. 1.) о стоимости потребительской корзины из 19 основных продуктов по городам Ростовской области (на начало апреля 1996 г.).

Таблица 1

Стоимость потребительской корзины, тыс. руб.	196	208	216	222	227	240
Число городов области	2	3	4	4	5	7

Требуется:

1) Постройте полигон распределения частот.
2) Найдите среднюю стоимость потребительской корзины в выборке, дисперсию, среднее квадратическое отклонение, коэффициент вариации. Объясните полученные результаты.

Решение.

Объем выборки: $n = 2 + 3 + 4 + 4 + 5 + 7 = 25.$

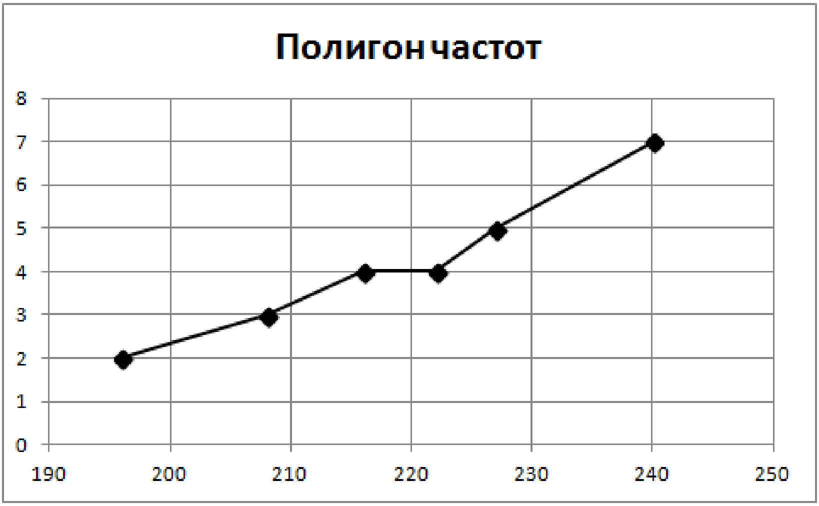
1) Данные о стоимости потребительской корзины представлены в виде дискретного вариационного ряда.

Значения изучаемого признака — стоимость потребительской корзины — обозначим x_i, частоты — n_i.

Получаем

x_i	196	208	216	222	227	240
n_i	2	3	4	4	5	7

Дискретный вариационный ряд можно представить графически, построив полигон распределения частот:



2) Рассчитаем среднюю стоимость потребительской корзины. Расчет средней арифметической произведем по следующей формуле:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot n_i$$

Для простоты расчетов составим вспомогательную расчетную таблицу 2.

Таблица 2

№ п/п	x_i	n_i	$x_i \cdot n_i$
1	196	2	392
2	208	3	624
3	216	4	864
4	222	4	888
5	227	5	1135
6	240	7	1680
Сумма			**5583**

$$\bar{x} = \frac{1}{25}\sum_{i=1}^{6} x_i \cdot n_i = \frac{5583}{25} = 223{,}32$$

Средняя стоимость потребительской корзины составляет 223,32 тыс.руб.

Для расчета дисперсии воспользуемся следующей формулой:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{k} (x_i - \bar{x})^2 \cdot n_i$$

Составим вспомогательную расчетную таблицу 3.

Таблица 3

№ п/п	$\boldsymbol{x_i}$	$\boldsymbol{n_i}$	$\boldsymbol{x_i - \bar{x}}$	$\boldsymbol{(x_i - \bar{x})^2}$	$\boldsymbol{(x_i - \bar{x})^2 \cdot n_i}$
1	196	2	-27,32	746,3824	1492,7648
2	208	3	-15,32	234,7024	704,1072
3	216	4	-7,32	53,5824	214,3296
4	222	4	-1,32	1,7424	6,9696
5	227	5	3,68	13,5424	67,7120
6	240	7	16,68	278,2224	1947,5568
Сумма					**4433,4400**

$$S^2 = \frac{1}{25-1}\sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{4433{,}44}{24} = 184{,}7267$$

Дисперсия стоимости потребительской корзины составляет 184,7267 (тыс.руб.)^2.

Найдем среднее квадратическое отклонение стоимости потребительской корзины по формуле

$$s = \sqrt{S^2} = \sqrt{184{,}7267} = 13{,}5914$$

Среднее квадратическое отклонение стоимости потребительской корзины составляет 13,5914 тыс. руб.

Найдем коэффициент вариации стоимости потребительской корзины по формуле:

$$v = \frac{S}{|\bar{x}|} \cdot 100\%$$

$$v = \frac{13,5914}{223,32} \cdot 100\% = 6,086 \approx 6,1\%$$

Коэффициент вариации составляет 6,1%. Будем считать, что критическое значение коэффициента вариации v^* составляет 35%. Так как коэффициент вариации меньше 35%, можно сделать вывод о том, что изучаемая совокупность городов является однородной, чем и объясняется низкая колеблемость стоимости потребительской корзины в данной совокупности.

Ввиду однородности городов, попавших в выборку, использование средней арифметической для характеристики наиболее типичного уровня стоимости потребительской корзины оправданно — средняя арифметическая типична для изучаемой совокупности.

Задача 2. Имеются выборочные данные о числе сделок (табл. 4.), заключенных брокерскими фирмами и конторами города в течение месяца.

Таблица 4

Число заключенных сделок	10-30	30-50	50-70	70-90
Число брокерских фирм и контор	20	18	12	5

Требуется:

1) Постройте гистограмму распределения частот.
2) Найдите среднее число заключенных сделок, дисперсию, среднее квадратическое отклонение, коэффициент вариации, размах вариации. Объясните полученные результаты.

Решение.

Объем выборки: $n = 20 + 18 + 12 + 5 = 55$;

Изучаемый признак: число заключенных сделок;

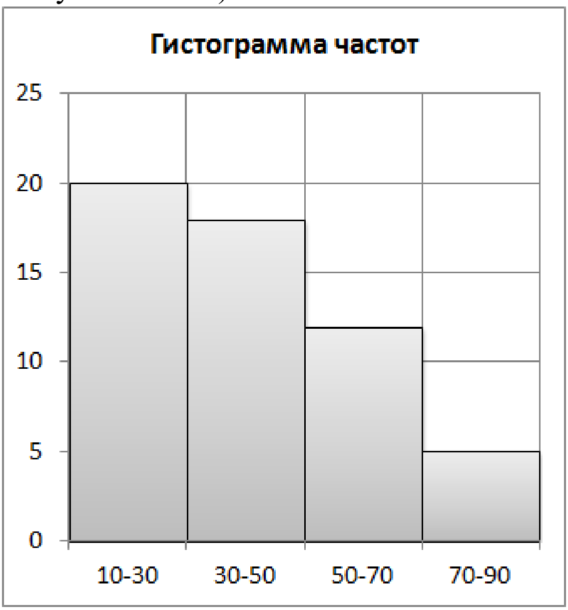
Частоты ($\boldsymbol{n_i}$): число брокерских фирм и контор.

1) Данные о числе заключенных сделок представлены в виде интервального вариационного ряда — значения признака заданы в виде интервалов.

Данный интервальный вариационный ряд — с равными интервалами: интервальные разности (разность между верхней и нижней границами) интервалов одинаковы:

$$l_i = 30 - 10 = 50 - 30 = 70 - 50 = 90 - 70 = 20$$

Построим гистограмму частот (высота прямоугольника соответствует частоте).



2) Рассчитаем среднее число заключенных сделок. Так как частоты интервалов разные, используем для расчета средней арифметической формулу

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot n_i$$

При расчете числовых характеристик интервального вариационного ряда в качестве значений признака принимаются середины интервалов, найдем их.

$$x_1 = \frac{30+10}{2} = 20; x_2 = \frac{30+50}{2} = 40;$$

$$x_3 = \frac{70+50}{2} = 60; x_4 = \frac{90+70}{2} = 80.$$

Для простоты расчетов составим вспомогательную расчетную таблицу 5.

Таблица 5

№ п/п	$\boldsymbol{x_i}$	$\boldsymbol{n_i}$	$\boldsymbol{x_i \cdot n_i}$
1	20	20	400
2	40	18	720
3	60	12	720
4	80	5	400
Сумма			**2240**

$$\bar{x} = \frac{1}{55}\sum_{i=1}^{4} x_i \cdot n_i = \frac{2240}{55} = 40{,}7273 \approx 41$$

Среднее число заключенных сделок составляет 41 шт.

Для расчета дисперсии воспользуемся следующей формулой:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{k} (x_i - \bar{x})^2 \cdot n_i$$

Составим вспомогательную расчетную таблицу 6.

Таблица 6

№ п/п	x_i	n_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot n_i$
1	20	20	-21	441	8820
2	40	18	-1	1	18
3	60	12	19	361	4332
4	80	5	39	1521	7605
Сумма					**20775**

$$S^2 = \frac{1}{55-1}\sum_{i=1}^{4}(x_i - \bar{x})^2 = \frac{20775}{54} = 384{,}72 \approx 385$$

Дисперсия числа заключенных сделок составляет 385 шт^2.

Найдем среднее квадратическое отклонение числа заключенных сделок по формуле

$$s = \sqrt{S^2} = \sqrt{385} = 19{,}6214 \approx 20$$

Среднее квадратическое отклонение числа заключенных сделок составляет 20 шт.

Найдем коэффициент вариации числа заключенных сделок по формуле

$$v = \frac{S}{|\bar{x}|} \cdot 100\%$$

$$v = \frac{20}{41} \cdot 100\% = 0{,}49 \approx 49\%$$

Коэффициент вариации составляет 49%. Так как коэффициент вариации больше 35%, можно сделать вывод о том, что изучаемая совокупность брокерских фирм и контор является неоднородной, в ее состав вошли и крупные, и мелкие фирмы и конторы, что и обусловило высокую колеблемость числа заключенных сделок.

Ввиду неоднородности фирм и контор, попавших в выборку, использование средней арифметической для характеристики наиболее типичного числа заключенных сделок не вполне оправданно — средняя арифметическая нетипична для изучаемой совокупности, в качестве характеристики наиболее типичного числа заключенных сделок в данной совокупности лучше использовать моду или медиану.

Задача 3. Имеются данные о годовой мощности предприятий цементной промышленности:

Таблица 7

Предприятия с годовой мощностью, тыс. т	Количество предприятий
До 500	27
500 - 1 000	11
1 000 - 2 000	8
2 000 - 3 000	8
Свыше 3 000	2

Требуется:

1) Постройте гистограмму распределения частот.
2) Рассчитайте среднюю мощность предприятий.
3) Найдите дисперсию, среднее квадратическое отклонение, коэффициент вариации.
4) Объясните полученные результаты, сделайте выводы.

Решение.

Объем выборки: $\boldsymbol{n = 27 + 11 + 8 + 8 + 2 = 56}$**.**

Изучаемый признак: годовая мощность предприятий цементной промышленности;

Частоты ($\boldsymbol{n_i}$ **):** количество предприятий.

1) Данные о годовой мощности предприятий цементной промышленности представлены в виде интервального вариационного ряда — значения признака заданы в виде

интервалов. При этом первый и последний интервалы — открытые: оба интервала не имеют одной из границ.

Наконец, данный интервальный вариационный ряд — с неравными интервалами: интервальные разности (разность между верхней и нижней границами) интервалов неодинаковы. Условно закроем границы открытых интервалов.

Интервальная разность 2-го интервала:

$$1\,000 - 500 = 500$$

Следовательно, нижняя граница 1-го интервала

$$500 - 500 = 0.$$

Интервальная разность предпоследнего интервала

$$3\,000 - 2\,000 = 1\,000.$$

Следовательно, верхняя граница последнего интервала

$$3\,000 + 1\,000 = 4\,000.$$

В результате, получим следующий вариационный ряд:

Таблица 8

x_i	n_i
0-500	27
500 - 1 000	11
1 000 - 2	8
2 000 - 3	8
3 000 - 4	2

Учитывая неодинаковую величину интервалов, для построения гистограммы рассчитаем абсолютные плотности распределения по формуле $f(\Delta_i) = \frac{n_i}{l_i}$:

$$f(\Delta_1) = \frac{27}{500 - 0} = 0{,}054; f(\Delta_2) = \frac{11}{1000 - 500} = 0{,}022, \text{и. т. д.}$$

Таблица 9

x_i	n_i	$f(\Delta_i)$
0-500	27	0,054
500 - 1 000	11	0,022
1 000 - 2	8	0,008
2 000 - 3	8	0,008
3 000 - 4	2	0,002

Построим гистограмму:

2) Рассчитаем среднюю мощность предприятий цементной промышленности. Так как частоты интервалов разные, используем для расчета средней арифметической формулу

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot n_i$$

При расчете числовых характеристик интервального вариационного ряда в качестве значений признака принимаются середины интервалов, найдем их.

$$x_1 = \frac{0 + 500}{2} = 250; x_2 = \frac{500 + 1000}{2} = 750;$$

$$x_3 = \frac{1000 + 2000}{2} = 1500; x_4 = \frac{2000 + 3000}{2} = 2500;$$

$$x_5 = \frac{4000 + 3000}{2} = 3500.$$

Для простоты расчетов составим вспомогательную расчетную таблицу 10.

Таблица 10

№ п/п	$\boldsymbol{x_i}$	$\boldsymbol{n_i}$	$\boldsymbol{x_i \cdot n_i}$
1	250	27	6750
2	750	11	8250
3	1500	8	12000
4	2500	8	20000
5	3500	2	7000
Сумма			**54000**

$$\bar{x} = \frac{1}{56}\sum_{i=1}^{5} x_i \cdot n_i = \frac{54000}{56} = 964{,}29.$$

Средняя мощность предприятий цементной промышленности составляет 964,29 тыс. т.

Следует отметить, что использование с той или иной целью средней арифметической, рассчитанной по данным интервального ряда с открытыми интервалами, может привести к серьезным ошибкам. Это связано с тем, что открытые интервалы закрываются условно, в действительности значения признака у объектов, попадающих в открытые интервалы, могут выходить далеко за их условные границы.

В связи с этим для оценки наиболее типичного уровня изучаемого признака по данным интервального ряда с открытыми интервалами лучше использовать моду или медиану.

3) Оценим колеблемость мощности предприятий цементной промышленности.

Так как частоты неодинаковы, для расчета дисперсии используем формулу

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot n_i$$

Составим вспомогательную расчетную таблицу 11.

Таблица 11

№ п/п	$\boldsymbol{x_i}$	$\boldsymbol{n_i}$	$\boldsymbol{x_i - \bar{x}}$	$\boldsymbol{(x_i - \bar{x})^2}$	$\boldsymbol{(x_i - \bar{x})^2 \cdot n_i}$
1	250	27	-714,29	510204,08	13775510,20
2	750	11	-214,29	45918,37	505102,04
3	1500	8	535,71	286989,80	2295918,37
4	2500	8	1535,71	2358418,37	18867346,94
5	3500	2	2535,71	6429846,94	12859693,88
Сумма					**48303571,43**

$$S^2 = \frac{1}{56-1}\sum_{i=1}^{5}(x_i - \bar{x})^2 = \frac{48303571{,}43}{56} = 862563{,}78.$$

Дисперсия мощности предприятий — 862 563,78 (тыс. т)^2.

Найдем среднее квадратическое отклонение по формуле $s = \sqrt{S^2} = \sqrt{862\ 563{,}78} = 928{,}74$.

Среднее квадратическое отклонение мощности предприятий — 928,74 тыс. т.

Найдем коэффициент вариации мощности предприятий по формуле

$$v = \frac{S}{|\bar{x}|} \cdot 100\%$$

$$v = \frac{928,74}{964,29} \cdot 100\% = 0,9631 \approx 96,31\%$$

Коэффициент вариации годовой мощности предприятий цементной промышленности составляет 96,31%. Поскольку он больше 35%, можно сделать вывод о том, что изучаемая совокупность предприятий является неоднородной, в ее состав вошли и крупные и мелкие предприятия, что и обусловило высокую колеблемость годовой мощности.

Следовательно, использование средней арифметической для характеристики наиболее типичного уровня годовой мощности предприятий цементной промышленности неверно — средняя арифметическая нетипична для изучаемой совокупности. Это еще раз подтверждает необходимость использования моды или медианы для характеристики наиболее типичного уровня годовой мощности данной совокупности предприятий цементной промышленности.

Задача 4. Имеются данные о годовом доходе (X) на душу населения в некотором городе:

8,50 10,50 12,25 7,00 17,00 8,75 10,00 9,30
8,00 11,50 10,00 12,00 9,00 6,50 13,00 10,20

Оцените среднедушевой доход в городе и разброс в доходах.

Решение.

Составим вспомогательную расчетную таблицу в MS Excel:

№	Xi	Xi^2
1,00	8,50	=B4^2
2,00	10,50	110,25
3,00	12,25	150,06
4,00	7,00	49,00
5,00	17,00	289,00
6,00	8,75	76,56
7,00	10,00	100,00
8,00	9,30	86,49
9,00	8,00	64,00
10,00	11,50	132,25
11,00	10,00	100,00
12,00	12,00	144,00
13,00	9,00	81,00
14,00	6,50	42,25
15,00	13,00	169,00
16,00	10,20	104,04
Сумма	163,50	1770,16
Среднее	10,22	110,63
Дисперсия	6,21	
Ср.кв.откл.	2,49	

Сумма	163,50	1770,16
Среднее	10,22	110,63
Дисперсия	=C21-B21^2	
Ср.кв.откл.	2,49	

Дисперсия	6,21
Ср.кв.откл.	=B22^0,5

Таким образом, получаем следующие результаты:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{16} \cdot 163{,}50 = 10{,}22;$$

$$\sigma_X^2 = \overline{x^2} - (\bar{x})^2;$$

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{1}{16} \cdot 1770{,}16 = 110{,}63$$

$$\sigma_X^2 = 110{,}63 - (10{,}22)^2 = 6{,}21$$

$$\sigma_x = \sqrt{6{,}21} = 2{,}49$$

Вывод: Средний душевой доход в городе составляет 10,22 у.е. Разброс дохода составляет 2,49 у.е.

Задача 5 . Туристическую фирму крупного курортного города интересует связь между числом отпускников, останавливающихся в отелях, и расходами на рекламу отелей. Взято случайное число отелей 8, сходных по размерам (см. табл. 4.4).

Таблица 12

Отель	1	2	3	4	5	6	7	8
Реклама (тыс. долл.)	9	6	10	8	7	4	11	9
Число гостей	1100	1200	1600	1300	1100	800	1200	1500

Определите степень влияния рекламы на выбор отеля (вычислить коэффициент парной корреляции).

Решение.

Составим вспомогательную расчетную таблицу в MS Excel:

	A	B	C	D	E	F
2	У- кол-во гостей отеля					
3	№	X	Y	X*Y	X^2	Y^2
4	1.00	9.00	1100.00	9900.00	81.00	1210000.00
5	2.00	6.00	1200.00	7200.00	36.00	1440000.00
6	3.00	10.00	1600.00	16000.00	100.00	2560000.00
7	4.00	8.00	1300.00	10400.00	64.00	1690000.00
8	5.00	7.00	1100.00	7700.00	49.00	1210000.00
9	6.00	4.00	800.00	3200.00	16.00	640000.00
10	7.00	11.00	1200.00	13200.00	121.00	1440000.00
11	8.00	9.00	1500.00	13500.00	81.00	2250000.00
12	**Сумма**	**64.00**	**9800.00**	**81100.00**	**548.00**	**12440000.00**
13	Среднее	8.00	1225.00	10137.50	68.50	1555000.00
14	Ср.кв.откл.	2.12	233.18			
15						
16	**Коэффициент парной корреляции**					
17						
18	**r_xy**	0.68		=(D13-B13*C13)/(B14*C14)		

Таким образом, коэффициент парной корреляции $r_{xy} = 0{,}68$, что свидетельствует о наличии средней корреляционной связи между изучаемыми показателями.

Интервальное оценивание

Задача 6. Руководство фирмы провело выборочное обследование 900 своих служащих. Средний стаж их работы в фирме равен 8,70 года, а среднее квадратическое (стандартное) отклонение – 2,70 года. Считается, что стаж работы служащих фирмы распределен по нормальному закону.

С вероятностью 0,95 постройте доверительный интервал, в котором окажется средний стаж работы всех служащих фирмы.

Решение.

Объем выборки $n = 900$, т. е. выборка большая.

Найдем границы доверительного интервала среднего стажа работы всего коллектива фирмы, т. е. границы доверительного интервала для генеральной средней:

$$I_\gamma = (\bar{x} - \varepsilon; \bar{x} + \varepsilon), \varepsilon = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

По условию

$$\bar{x} = 8{,}70; \sigma = 2{,}70; \gamma = 0{,}95; \alpha = 1 - \gamma = 0{,}05; z_{\alpha/2} = z_{0,025} = 1{,}96.$$

Найдем точность интервальной оценки:

$$\varepsilon = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1{,}96 \cdot \frac{2{,}70}{\sqrt{900}} = 1{,}96 \cdot 0{,}09 = 0{,}1764.$$

Доверительный интервал:

$$I_{0,95} = (8{,}70 - 0{,}1764; 8{,}70 + 0{,}1764)$$

$$I_{0,95} = (8{,}5236; 8{,}8764).$$

С вероятностью 0,95 можно ожидать, что средний стаж работы всего коллектива фирмы находится в интервале от 8,5236 до 8,8764 года.

Замечание. Значение $z_{\alpha/2}$ можно вычислить с помощью встроенной функции Excel «НОРМСТОБР». Для задачи 6 получаем

$$\mathrm{z}_{0,025} = \text{НОРМСТОБР}(\underbrace{1 - 0{,}0025}_{1-\alpha/2}) = 1{,}96.$$

Задача 7. Владелец автостоянки опасается обмана со стороны своих служащих (охраны автостоянки). В течение года (365 дней) владельцем автостоянки проведено 40 проверок. По данным проверок среднее число автомобилей, оставляемых на ночь на охрану, составило 400 единиц, а среднее квадратическое (стандартное) отклонение их числа – 10 автомобилей.

С вероятностью 0,99 оцените с помощью доверительного интервала истинное среднее число автомобилей, оставляемых на ночь на охрану.

Обоснованы ли опасения владельца автостоянки, если по отчетности охранников среднее число автомобилей, оставляемых на ночь на охрану, составляет 395 автомобилей?

Решение.

Объем выборки $n = 40$, что больше 30 единиц, т. е. выборка большая.

Объем генеральной совокупности $N = 365$.

Найдем границы доверительного интервала для оценки среднего числа автомобилей, оставляемых на ночь на охрану, т. е. границы доверительного интервала для генеральной средней:

$$I_\gamma = (\bar{x} - \varepsilon; \bar{x} + \varepsilon), \varepsilon = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

По условию

$$\bar{x} = 400; \sigma = 10; \gamma = 0{,}99; \alpha = 1 - \gamma = 0{,}01;$$
$$z_{\alpha/2} = \mathrm{z}_{0{,}005} = \text{НОРМСТОБР}(1 - 0{,}005) = 2{,}575.$$

Найдем точность интервальной оценки:

$$\varepsilon = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 2{,}575 \cdot \frac{10}{\sqrt{40}} = 2{,}575 \cdot 1{,}5811 = 4{,}0713.$$

Доверительный интервал:

$$I_{0{,}99} = (400 - 4{,}0713; 400 + 4{,}0713)$$
$$I_{0{,}99} = (395{,}9287; 404{,}0713) \quad \text{или} \quad I_{0{,}99} = (396; 404)$$

С уверенностью в 99% можно ожидать, что среднее число автомобилей, оставляемых на ночь на охрану, находится в интервале от 396 до 404.

Таким образом, можно утверждать, что служащие автостоянки обманывают ее владельца.

Задача 8. Служба контроля энергосбыта провела выборочную проверку расхода электроэнергии жителями одного из многоквартирных домов. Случайным образом выбрано 10 квартир и определен расход электроэнергии в течение одного из летних месяцев (кВт · ч):

$$125;\ 78;\ 102; 140;\ 90;\ 45;\ 50; 125;\ 115; 112.$$

С вероятностью 0,95 определите доверительный интервал для оценки среднего расхода электроэнергии на 1 квартиру во всем доме при условии, что в доме 70 квартир.

Решение.

Объем выборки $n =$ 10 единиц, т. е. выборка малая.

Найдем доверительный интервал для оценки среднего расхода электроэнергии на 1 квартиру во всем доме, т. е. границы доверительного интервала для оценки генеральной средней:

$$I_\gamma = (\bar{x} - \varepsilon; \bar{x} + \varepsilon), \varepsilon = t_{\text{кр}} \frac{s}{\sqrt{n}}$$

Для определения границ доверительного интервала необходимо рассчитать выборочные среднюю и среднее квадратическое (стандартное) отклонение. Составим вспомогательную расчетную таблицу 13.

Таблица 13

№ п/п	$\boldsymbol{x_i}$	$\boldsymbol{x_i - \bar{x}}$	$\boldsymbol{(x_i - \bar{x})^2}$
1	125	26,8	718,24
2	78	-20,2	408,04
3	102	3,8	14,44
4	140	41,8	1747,24
5	90	-8,2	67,24
6	45	-53,2	2830,24
7	50	-48,2	2323,24
8	125	26,8	718,24
9	115	16,8	282,24
10	112	13,8	190,44
Сумма	**982**		**9299,6**

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{\mathrm{n}} x_i = \frac{1}{10} \cdot 982 = 98,2;$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{9299,6}{9} = 1033,2889;$$

$$s = \sqrt{s^2} = 32,1448.$$

Итак, дано: $\bar{x} = 98,2; s = 32,1448; \gamma = 0,95; \alpha = 0,05$.

По таблице Стьюдента *(приложение 2)* найдем $t_{\text{кр}}$ по уровню значимости $\alpha = 0,05$ и числу степеней свободы $df = n - 1 = 10 - 1 = 9$ (двусторонняя критическая область):

$$t_{\text{кр}} = \mathrm{t}_{\text{кр}}(0,05; 9) = 2,26.$$

Найдем точность интервальной оценки:

$$\varepsilon = t_{\text{кр}} \frac{s}{\sqrt{n}} = 2{,}26 \cdot \frac{32{,}1448}{\sqrt{10}} = 22{,}9731.$$

Доверительный интервал:

$$I_{0,95} = (98{,}2 - 22{,}9731; 98{,}2 + 22{,}9731)$$

$$I_{0,95} = (75{,}2269; 121{,}1731).$$

С вероятностью 0,95 можно ожидать, что средний расход электроэнергии на 1 квартиру во всем доме находится в интервале от 75,2269 до 121,1731 кВт.ч.

Замечание. Значение $t_{\text{кр}}$ для **двусторонней критической области** можно вычислить с помощью встроенной функции Excel «СТЬЮДРАСПОБР».
Для задачи 8 получаем

$$t_{\text{кр}} = \text{СТЬЮДРАСПОБР}(0{,}05; 9) = 2{,}96.$$

Задача 9. На фабрике работает автоматическая линия по фасовке растворимого кофе в жестяные 100-грамовые банки. Если средний вес наполняемых банок отличается от точного, то линия налаживается для подгонки среднего веса в рабочем режиме. Если дисперсия веса банок превышает заданное значение, то линия должна быть остановлена на ремонт и переналадку. Время от времени производится отбор банок с кофе для проверки веса и его колеблемости. Предположим, что с линии случайным образом отобрано 30 банок с кофе и оценка дисперсии 18,540 г^2.

С вероятностью 0,95 определите доверительный интервал, в котором окажется генеральная дисперсия (считается, что вес банок распределен по нормальному закону).

Решение.

Объем выборки $n = 30$.

Доверительный интервал для генеральной дисперсии равен:

$$I_\gamma = \left(\frac{(n-1)\cdot s^2}{\chi_R^2}; \frac{(n-1)\cdot s^2}{\chi_L^2}\right)$$

$$P(\chi^2 > \chi_L^2) = 1 - \frac{\alpha}{2}$$

$$P(\chi^2 > \chi_R^2) = \frac{\alpha}{2}$$

По условию:

$n = 30; s^2 = 18{,}540; \gamma = 0{,}95; \alpha = 1 - \gamma = 0{,}05.$

Значения χ_L^2 и χ_R^2 найдем по таблице критических значений хи-квадрат распределения (приложение 4) из условий:

$$\chi_L^2 = \chi^2_{\text{кр}}\left(1 - \frac{\alpha}{2}; \mathrm{n} - 1\right) \text{ и } \chi_R^2 = \chi^2_{\text{кр}}\left(\frac{\alpha}{2}; \mathrm{n} - 1\right).$$

Получаем, что

$$\chi_L^2 = \chi^2_{\text{кр}}\left(1 - \frac{\alpha}{2}; \mathrm{n} - 1\right) = \chi^2_{\text{кр}}(0{,}975; 29) = 16{,}0471;$$

$$\chi_R^2 = \chi^2_{\text{кр}}\left(\frac{\alpha}{2}; \mathrm{n} - 1\right) = \chi^2_{\text{кр}}(0{,}0025; 29) = 45{,}7223.$$

Доверительный интервал:

$$I_{0,95} = \left(\frac{29 \cdot 18{,}540}{45{,}7223}; \frac{29 \cdot 18{,}540}{16{,}0471}\right);$$

$$I_{0,95} = (11{,}765; 33{,}604)$$

Замечание. Значение $\chi^2_{\text{кр}}$ можно вычислить с помощью встроенной функции Excel «ХИ2ОБР». Для задачи 9 получаем

$$\chi^2_{\text{кр}}(0{,}975; 29) = \text{ХИ2ОБР}(0{,}975; 29) = 16{,}0471.$$

$$\chi^2_{\text{кр}}(0{,}0025; 29) = \text{ХИ2ОБР}(0{,}0025; 29) = 45{,}7223.$$

Проверка статистических гипотез

Задача 10. Техническая норма предусматривает в среднем 40 с на выполнение определенной технологической операции на конвейере по производству часов. От работающих на этой операции поступили жалобы, что они в действительности затрачивают на нее больше времени. Для проверки данной жалобы произведены хронометрические измерения времени выполнения этой технологической операции у 16 работниц, занятых на ней, и получено среднее время выполнения операции $\bar{x} = 42$ с.

Можно ли по имеющимся хронометрическим данным на уровне значимости α = 0,01 отклонить гипотезу о том, что среднее время выполнения этой операции соответствует норме, если исправленное выборочное среднее квадратическое отклонение $s = 3{,}5$ с ?

Решение.

1) Для решения данной задачи необходимо проверить гипотезу о том, что неизвестная генеральная средняя нормальной совокупности точно равна определенному числу, когда ***дисперсия генеральной совокупности не известна*** (выборка мала, так как n = 16 меньше 30).

Сформулируем нулевую и альтернативную гипотезы согласно условию задачи.

$$H_0: m = m_0 = 40$$

Неизвестное математическое ожидание m (нормально распределенной генеральной совокупности с неизвестной дисперсией) равно гипотетически предполагаемому числовому значению m_0 (применительно к условию данной задачи – время выполнения технологической операции соответствует норме).

$$H_1: m > 40$$

Неизвестное математическое ожидание m (нормально распределенной генеральной совокупности с неизвестной дисперсией) больше числового значения m_0 (применительно к условию данной задачи – время выполнения технологической операции больше установленной нормы).

Так как конкурирующая гипотеза – правосторонняя, то и критическая область – правосторонняя.

В качестве критерия для сравнения неизвестного математического ожидания m (нормально распределенной генеральной совокупности с неизвестной дисперсией) с гипотетическим числовым значением m_0 используется случайная величина t-критерий Стьюдента.

2) Для проверки гипотезы H_0 применим критерий $t = \frac{\bar{x}-m_0}{s/\sqrt{n-1}}$ с односторонней критической областью.

3) Вычислим t:

$$t = \frac{\bar{x} - m_0}{s/\sqrt{n-1}} = ?$$

$$m_0 = 40; n = 16; s = 3{,}5; \bar{x} = 42$$

$$t = \frac{42 - 40}{3{,}5/\sqrt{15}} = 2{,}2131$$

4) Критическая область U является правосторонней:

$$U = \mathrm{t} > \mathrm{t}_\alpha, \text{ где}$$

$t_\alpha = t_{\text{кр}}(\alpha; n-1)$ следует находить по таблице Стьюдента (приложение 2) по уровню значимости $\alpha = 0{,}01$ и числу степеней свободы $df = n - 1 = 16 - 1 = 15$ (односторонняя критическая область):

$$t_{0{,}01} = t_{\text{кр}}(0{,}01; 15) = 2{,}58$$

Таким образом, критическая область имеет вид:

$$U = (2{,}58; +\infty)$$

5) Так как, $t \notin U$ на данном уровне значимости нет оснований отклонить нулевую гипотезу.

Вывод. По имеющимся хронометрическим данным на уровне значимости α = 0,01 нельзя отклонить гипотезу о том, что среднее время выполнения этой операции соответствует норме. Следовательно, жалобы работниц - необоснованны.

Задача 11. Техническая норма предусматривает в среднем 40 с на выполнение определенной технологической операции на конвейере по производству часов. От работающих поступили жалобы, что они в действительности затрачивают на эту операцию больше времени. Для проверки данной жалобы произведены хронометрические измерения времени ее выполнения у 36 работниц, занятых на этой операции, и получено среднее время выполнения операции $\bar{x} = 42$ с.

Можно ли (предполагая время выполнения технологической операции случайной величиной, подчиняющейся нормальному закону) по имеющимся хронометрическим данным на уровне значимости α = 0,01 отклонить гипотезу о том, что среднее время выполнения этой операции соответствует норме, если известно, что среднее квадратическое отклонение генеральной совокупности σ составляет 3,5 с?

Решение.

1) Для решения данной задачи необходимо проверить гипотезу о том, что неизвестная генеральная средняя нормальной совокупности точно равна числовому значению, когда ***дисперсия генеральной совокупности известна*** (большая выборка, так как n = 36 больше 30).

Сформулируем нулевую и альтернативную гипотезы согласно условию задачи:

$$H_0: m = m_0 = 40$$

Неизвестная генеральная средняя нормально распределенной совокупности с известной дисперсией равна числовому значению (применительно к условию данной задачи –время выполнения технологической операции соответствует норме).

$$H_1: m > 40$$

Неизвестная генеральная средняя нормально распределенной совокупности с известной дисперсией больше числового значения (применительно к условию данной задачи – время выполнения технологической операции больше установленной нормы).

2) Для проверки гипотезы H_0 применим критерий $z = \frac{\bar{x} - m_0}{\sigma/\sqrt{n}}$ с односторонней критической областью.

3) Вычислим z:

$$z = \frac{\bar{x} - m_0}{\sigma/\sqrt{n}} = ?$$

$$m_0 = 40; n = 36; s = 3{,}5; \bar{x} = 42$$

$$z = \frac{42 - 40}{3{,}5/\sqrt{36}} = 3{,}4286$$

4) Критическая область U является правосторонней:

$$U = z > z_\alpha, \text{ где}$$

$\Phi_0(z_\alpha) = 0{,}5 - \alpha$ следует находить по таблице значений функции Лапласа (приложение 1) по уровню значимости $\alpha = 0{,}01$:

$$\Phi_0(z_{0,01}) = 0{,}5 - 0{,}01 = 0{,}49; z_{0,01} = 2{,}33$$

При вычислении z_α с помощью встроенной функции Excel «НОРМСТОБР» также получим

$$z_{0,01} = \text{НОРМСТОБР}(1 - 0{,}01) = 2{,}33$$

Таким образом, критическая область имеет вид:

$$U = (2{,}33; +\infty)$$

5) Так как, $t \in U$ на данном уровне значимости нулевая гипотеза отвергается в пользу альтернативной.

Вывод. По имеющимся хронометрическим данным на уровне значимости α = 0,01 можно утверждать, что среднее время выполнения этой операции превышает норму, жалобы работниц – обоснованны.

Задача 12. Экономический анализ производительности труда предприятий отрасли позволил выдвинуть гипотезу о наличии 2 типов предприятий с различной средней величиной показателя производительности труда. Выборочное обследование 42 предприятий 1-й группы дало следующие результаты: средняя производительность труда $\overline{\mathrm{X}} = 119$ деталей. По данным выборочного обследования, на 35 предприятиях 2-й группы средняя производительность труда $\overline{Y} = 107$ деталей. Генеральные дисперсии известны: $D(X) = 126{,}91$ (дет.^{2}); $D(Y) = 136{,}1$ (дет.^{2}). Считая, что выборки извлечены из нормально распределенных генеральных совокупностей X и Y, на уровне значимости 0,05, проверьте, случайно ли полученное различие средних показателей производительности труда в группах или же имеются 2 типа предприятий с различной средней величиной производительности труда.

Решение.

1) Для решения данной задачи необходимо сравнить две средние нормально распределенных генеральных совокупностей, ***генеральные дисперсии которых известны*** (большие независимые выборки).

В данной задаче речь идет о больших выборках, так как $n_x = 42$ и $n_y = 35$ больше 30. Выборки – независимые, так как из контекста задачи видно, что они извлечены из непересекающихся генеральных совокупностей.

Сформулируем нулевую и альтернативную гипотезы согласно условию задачи.

$$H_0: m_x = m_y$$

Генеральные средние двух нормально распределенных совокупностей с известными дисперсиями равны (применительно к условию данной задачи — предприятия 2 групп относятся к одному типу предприятий: средняя производительность труда в 2 группах — одинакова).

$$H_1: m_x \neq m_y$$

Генеральные средние двух нормально распределенных совокупностей с известными дисперсиями неравны (применительно к условию данной задачи — предприятия 2 групп относятся к разному типу предприятий: средняя производительность труда в 2 группах — неодинакова).

Замечание. Выдвигаем двустороннюю конкурирующую гипотезу, так как из условия задачи не следует, что необходимо выяснить больше или меньше производительность труда в одной из групп предприятий по сравнению с другой.

2) Для проверки гипотезы H_0 применим критерий $z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ с двусторонней критической областью.

3) Вычислим z:

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = ?$$

$$\bar{x}_1 = 119; \bar{x}_2 = 107; \sigma_1^2 = 129{,}91; \sigma_2^2 = 136{,}1; \mathrm{n}_1 = 42; \mathrm{n}_2 = 35;$$

$$z = \frac{119 - 107}{\sqrt{\frac{129{,}91}{42} + \frac{136{,}1}{35}}} = \frac{12}{\sqrt{6{,}9817}} = 4{,}5415$$

4) Критическая область U является двусторонней:

$$U: |z| > z_{\alpha/2}, \text{ где}$$

$\Phi_0(z_{\alpha/2}) = 0{,}5 - \alpha/2$ следует находить по таблице значений функции Лапласа (приложение 1) по уровню значимости $\alpha = 0{,}01$:

$$\Phi_0(z_{0,025}) = 0{,}5 - 0{,}025 = 0{,}475; z_{0,025} = 1{,}96$$

Таким образом, критическая область имеет вид:

$$U = (-\infty; -1{,}96) \cup (1{,}96; +\infty)$$

5) Так как, $z \in U$ на данном уровне значимости нулевая гипотеза отвергается в пользу альтернативной.

Вывод. На уровне значимости α = 0,05 можно утверждать, что полученное различие средних показателей производительности труда в группах не случайно, имеются 2 типа предприятий с различной средней величиной производительности труда.

Задача 13. Предполагается, что применение нового типа резца сократит время обработки некоторой детали. Хронометраж времени обработки 9 деталей, обработанных старым типом резцов, дал следующие результаты: среднее время обработки детали $\overline{\mathrm{X}} = 57$ мин, исправленная выборочная дисперсия $\mathrm{s}^2_{\mathrm{X}} = 186{,}2$ (мин^2). Среднее время обработки 15 деталей, обработанных новым типом резцов, $\bar{Y}$ по данным хронометражных измерений составляет 52 мин, а исправленная выборочная дисперсия $\mathrm{s}^2_{Y} = 166{,}4$ (мин^2). На уровне значимости α = 0,01 ответьте на вопрос, позволило ли использование нового типа резцов сократить время обработки детали?

Решение.

1) Для решения данной задачи необходимо сравнить две средние нормально распределенных генеральных совокупностей, ***генеральные дисперсии которых не известны и не предполагаются равными***.

В этой задаче речь идет о малых выборках, так как $n_x = 9$ и $n_y = 15$ меньше 30. Выборки – независимые, поскольку из контекста задачи видно, что они извлечены из непересекающихся генеральных совокупностей.

Сформулируем нулевую и альтернативную гипотезы согласно условию задачи.

$$H_0: m_x = m_y$$

Генеральные средние двух нормально распределенных совокупностей с неизвестными дисперсиями (но предполагаемыми одинаковыми) равны (применительно к условию данной задачи — среднее время, затрачиваемое на обработку детали резцами нового и старого типа, — одинаково, т. е. использование нового типа резца не позволяет снизить время на обработку детали).

$$H_1: m_x > m_y$$

Генеральная средняя для X больше, чем генеральная средняя для Y (применительно к условию данной задачи – среднее время, затрачиваемое на обработку детали резцами старого типа, больше, чем – нового, т. е. использование нового типа резца позволяет снизить время на обработку детали).

Приступать к проверке гипотезы о равенстве генеральных средних двух нормально распределенных совокупностей с неизвестными дисперсиями можно лишь в том случае, если генеральные дисперсии равны. В противном случае, данная задача в теории неразрешима.

Поэтому, прежде чем проверять эту гипотезу, проверим гипотезу о равенстве генеральных дисперсий нормальных совокупностей.

2) Для проверки гипотезы H_0 применим критерий $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ с правосторонней критической областью.

3) Вычислим t:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = ?$$

$$\bar{x}_1 = 57; \bar{x}_2 = 52; \sigma_1^2 = 186{,}2; \sigma_2^2 = 166{,}4; \mathrm{n}_1 = 9; \mathrm{n}_2 = 15;$$

$$t = \frac{57 - 52}{\sqrt{\frac{186{,}2}{9} + \frac{166{,}4}{15}}} = \frac{5}{\sqrt{31{,}7822}} = 0{,}8869$$

4) Критическая область U является правосторонней:

$U = \mathrm{t} > \mathrm{t}_{\alpha}$, где

$t_{\alpha} = t_{\text{кр}}(\alpha; df)$ следует находить по таблице Стьюдента (приложение 2) по уровню значимости $\alpha = 0{,}01$ и числу степеней свободы $df = \min(n_1 - 1; n_2 - 1) = \min(8; 14) = 8$ (односторонняя критическая область):

$$t_{0{,}01} = t_{\text{кр}}(0{,}01; 8) = 2{,}90$$

Таким образом, критическая область имеет вид:

$$U = (2{,}90; +\infty)$$

5) Так как, $t \notin U$ на данном уровне значимости нет оснований отклонить нулевую гипотезу.

По имеющимся хронометрическим данным на уровне значимости $\alpha = 0{,}01$ нельзя отклонить гипотезу о том, что генеральные средние равны, т. е. среднее время, затрачиваемое на обработку детали старым и новым типом резцов, отличается незначимо, расхождения между средними – случайны, использование нового типа резцов не позволяет снизить время обработки детали.

Вывод. На уровне значимости $\alpha = 0{,}01$ нельзя утверждать, что использование нового типа резцов позволило сократить время обработки детали.

Задача 14. В рекламе фирмы А утверждается, что месячный доход по ее акциям в среднем равен 0,75% и превышает доход акция фирмы В, причем ее риски (фирмы А) меньше. По выборке из 12 месяцев средний месячный доход по акциям А составил 0,65%, а по акциям В – 0,4%. Среднее квадратическое отклонение для А составило 2%, а для В – 1,9%. Полагая уровень значимости равным 5%, проверить все три утверждения, содержащиеся в рекламе.

Решение. Данная задача содержит три утверждения:

Утв.1. «Месячный доход по акциям фирмы А в среднем равен 0,75%.»

Утв.2. «Доход фирмы А превышает доход фирмы В»

Утв.3. «Риски фирмы А меньше рисков фирмы В »

Проведем проверку всех трех утверждений с помощью MS Excel.

	A	B	C	D	E	F	G	H
57								
58	**1. "Месячный доход по акциям фирмы А в среднем равен 0,75%."**							
59								
60	**Гипотеза о среднем**							
61	(дисперсия не известна, n=12 (<30))							
62								
63	H0: m=0,75							
64	H1: m≠0,75							
65								
66	m0	0,75						
67	Xcp	0,65						
68	s	2,00	=(B67-B66)/B68*(B69-1)^0,5					
69	n	12,00						
70	t	-0,17						
71	alpha	0,05						
72	tkp	2,20	=СТЬЮДРАСПОБР(B71;B69-1)					
73								

Таким образом, критическая область имеет вид:

$$U = (-\infty; -2,20) \cup (2,20; +\infty)$$

Статистика критерия $t \notin U$, следовательно, на данном уровне значимости нет оснований отклонить нулевую гипотезу. Другими словами, на уровне значимости 5% можно утверждать, что средний месячный доход акций для фирмы а составляет 0,75%.

	A	B	C	D	E	F	G	H
78	**2. "Доход фирмы А превышает доход фирмы В".**							
79								
80	Гипотеза о равенстве двух средних							
81	(дисперсии не известны и не предполагаются равными)							
82								
83	X - доход фирмы A							
84	У - доход фирмы B							
85								
86	H0: mx=my							
87	H1: mx>my							
88								
89	Xcp	0,65						
90	Ycp	0,40						
91	sx^2	4,00						
92	sy^2	3,61						
93	nx	12,00						
94	ny	12,00						
95	t	0,31	=(B89-B90)/(B91/B93+B92/B94)^0,5					
96	alpha	0,05						
97	df	11,00						
98	ta	1,80	(односторонняя критическая область)					
99								
100	Статистика критерия t=0,31							
101	Критическая область U: t>1,80							

Статистика критерия $t \notin U$, следовательно, на данном уровне значимости нет оснований отклонить нулевую гипотезу. Другими словами, на уровне значимости 5% можно утверждать, что среднемесячный доход по акциям в обеих фирмах примерно одинаковый.

	A	B	C	D	E	F	G
109	**3. " Риски фирмы А меньше рисков фирмы В"**						
110							
111	Гипотеза о равенстве дисперсий						
112	H0: Dx=Dy						
113	H1: Dx>Dy						
114							
115	Sx^2	4,00					
116	Sy^2	3,61					
117	F	1,11	=B115/B116				
118	alpha	0,05					
119	nx	12,00					
120	ny	12,00					
121	Fa	2,82	=FРАСПОБР(B118;B119-1;B120-1)				
122							
123	Статистика критерия F=1,11						
124	Критическая область U: F>2,82						

Статистика критерия $F \notin U$, следовательно, на данном уровне значимости нет оснований отклонить нулевую гипотезу. Другими словами, на уровне значимости 5% можно утверждать, что риски в обеих фирмах примерно одинаковые.

§9. Задания для самостоятельной работы

Упражнения и задачи

Упражнение 1. Предположим, сто выборка проведена из генеральной совокупности со средней 1,065 и средним квадратическим отклонением 500. Объем выборки 100. Чему равно ожидаемое значение и среднее квадратическое отклонение выборочной средней?

Упражнение 2. Имеются следующие результаты выборок, извлеченных из нормально распределенной генеральной совокупности:

- $n = 9; \sum x_i = 36; \sum (x_i - \bar{x})^2 = 288$;
- $n = 16; \sum x_i = 64; \sum (x_i - \bar{x})^2 = 180$;
- $n = 25; \sum x_i = 500; \sum x_i^2 = 11400.$

Чему равны оценки средней арифметической, дисперсии и среднего квадратического отклонения?

Упражнение 3. Аналитик рынка утверждает, что на 95% уверен в том, что истинное среднее значение суммы ежемесячных продаж продукции определенного типа находится между 170 000 и 200 000. Объясните, что аналитик имеет в виду? Почему аналитик не может утверждать, что он уверен в результате на 100%?

Упражнение 4. Постройте 95%-ный доверительный интервал для двух групп данных:

Группа 1: 1, 1, 1, 1, 8, 8, 8, 8.

Группа 2. 1, 2, 3, 4, 5, 6, 7, 8.

Объясните, почему доверительные интервалы различны, несмотря на то, что выборочные средние значения совпадают?

Задача 1. Брокер проводит случайную выборку четырех акций из большой генеральной совокупности акций с низким номиналом. Цены акции в генеральной совокупности подчиняются нормальному распределению. Цены акций в выборке составили: $5, $12, $17 и $10.

- Вычислите точечную оценку генеральной средней.
- Вычислите точечную оценку генеральной дисперсии. Какова ваша оценка генерального среднего квадратического отклонения?

Задача 2. Страховая компания оценивает среднюю сумму исков, предъявленных больными за врачебные ошибки. Компания осуществила случайную выборку 165 исков и нашла $\bar{x} = 16{,}530$ и $s = 5{,}542$. *Постройте* 95%-й и 99%-ный доверительные интервалы для средней суммы исков.

Задача 3. Крупный банк заинтересован в автоматизации кассовых операций в филиале, открывающемся в новом регионе. Для принятия обоснованного решения проводится эксперимент по определению среднего числа трансакций в условных денежных единицах на человека в день. Случайная выборка 10 экспериментальных трансакций, которые прошли через новые кассовые автоматы, дала следующие результаты: 53, 40, 39, 10, 12, 60, 72, 65, 50, 45. *Постройте* 95%-ный доверительный интервал для средней суммы трансакций.

Задача 4. Автотранспортная компания желает оценить среднее время транзита грузов из столицы в северные регионы страны. Случайная выборка 20 партий товаров дала: $\bar{x} = 2{,}6$ дней, $s = 0{,}4$ дня. *Постройте* 99%-ный доверительный интервал для среднего времени транзита товаров.

Задача 5. Оптовая фирма, торгующая моющими средствами, желает оценить объем ежедневной продажи упаковок мыла определенного сорта. Случайная выборка за 13 дней дала следующие результаты: 123, 110, 95, 120, 87, 89, 100, 105, 98, 88, 75, 125, 101. *Постройте* 90%-ный доверительный интервал числа ежедневной реализации упаковок мыла.

Задача 6. Бухгалтер желает оценить среднюю сумму счетов сервисной компании. Случайная выборка 46 счетов дала: $\bar{x} = 16{,}50$ условных денежных единиц, $s = 52{,}00$. *Постройте* 99%-ный доверительный интервал для средней суммы счетов.

Задача 7. Новая расфасовка сухих завтраков предлагается для проверки покупательского спроса в 16 магазинах города. Результаты месячного эксперимента дали следующий объем продаж в 1200 рублей со средним квадратическим отклонением в 180 рублей. Постройте 99%-ный доверительный интервал среднего объема продаж нового сорта сухих завтраков.

Задача 8. Предположим, что в магазин, торгующий эмалевыми красками для внутренних покрытий, начали поступать претензии от покупателей о том, что банки заполнены ниже нормы. Производитель красок утверждает, что среднее квадратическое отклонение объема краски в литровой банке составляет 0,02 литра. Случайная выборка 50 банок дала среднее значение объема 0,995 литра.

Требуется:

1) Постройте 99%-ный доверительный интервал для среднего значения объема краски в литровой банке.

2) Основываясь на выборочных результатах, объясните, должен ли владелец магазина подать рекламацию производителям краски? Почему?

3) Является ли генеральное распределение количества краски в банках нормальным? Объясните.

4) Объясните, почему наблюдаемое значение 0,98 литра краски в банке не является необычным, даже если находится вне вычисленного вами доверительного интервала.

Задача 9. Фирма-поставщик в рекламном букете утверждает, что средний срок безотказной работы предлагаемого изделия – 2900 ч. Для выборки из 50 изделий средний срок безотказной работы оказался равным 2720 ч. при выборочном среднем квадратическом отклонении 700 ч. При 5%-м уровне значимости проверить гипотезу о том, что значение 2900 ч является математическим ожиданием.

Задача 10. По результатам 10 замеров установлено, что среднее время обслуживания операционистом клиента $\bar{x} = 15$ мин. Предполагая, что время обслуживания клиента - нормально распределенная случайная величина с дисперсией $\sigma_x^2 = 9$ мин^2, при уровне значимости $\alpha = 0{,}05$ установить можно ли принять в качестве норматива (математического ожидания) для обслуживания одного клиента: а) 21 мин; б) 16 мин.

Задача 11. Годовой оборот 4 бирж в регионе A составил $12 \cdot 10^4$ у. е.; в регионе B годовой оборот 5 бирж —

$125 \cdot 10^3$ у. е. Исправленная выборочная дисперсия оборота в регионе А оказалась равной $3 \cdot 10^4$(у.е.)^2, в регионе B — $2 \cdot 10^4$ (у.е.)^2. Можно ли на уровне значимости $\alpha = 0,05$ утверждать, что средний оборот бирж в регионе A больше, чем в регионе B?

Вопросы и задания для самоконтроля

1. Что такое генеральная совокупность и выборка?
2. Назовите основные виды выборок и способы отбора элементов в них.
3. Что такое статистический ряд?
4. Что такое интервальный статистический ряд?
5. Что такое полигон частот и гистограмма? Для чего они используются?
6. Как вычисляются основные числовые характеристики по результатам выборки: выборочные среднее, дисперсия, среднее квадратическое отклонение?
7. Как вычисляется и где применяется выборочный коэффициент вариации?
8. Что такое точечная оценка и каковы желательные свойства?
9. Дайте определение несмещенности, эффективности и состоятельности оценок.
10. Что такое интервальная оценка? Как она строится?
11. Чем отличаются интервальные оценки для математического ожидания нормальной СВ при известной и неизвестной дисперсиях?
12. Как строятся доверительные интервалы для дисперсии и среднего квадратического отклонения нормальной СВ?

13. Что такое статистическая гипотеза?

14. Какова цель проверки гипотез?

15. В чем отличие параметрических и непараметрических гипотез?

16. Что такое нулевая и альтернативная гипотезы? Назовите принципы их построения.

17. Что такое статистический критерий? Приведите конкретные примеры критериев.

18. Сформулируйте общую схему проверки гипотез.

19. Что такое ошибки первого и второго рода? Как можно уменьшить вероятности этих ошибок?

20. Что такое уровень значимости?

21. Что такое уровень доверия?

22. Что определяет мощность критерия?

18.Что такое доверительный интервал и в чем его преимущества?

19. Приведите примеры проверки гипотез в экономике. Какими критериями можно воспользоваться при их проверке?

Ответы

Упражнения

1. 1,065; 2,500.

2. $\bar{x} = 4; s^2 = 36; s = 6;$

$\bar{x} = 4; s^2 = 12; s = 2\sqrt{3};$

$\bar{x} = 20; s^2 = 100; s = 10.$

Задачи

1. $\bar{x} = 11; s^2 = 24{,}67; s = 4{,}97.$

2. (15,68437; 17,375,63), (15,4186;17,6414).

3. (29,87; 59,33).

4. (2,34; 2,86).

5. (93,75; 108,71).

6. (15,86; 17,14).

7. (1067,40; 1332,60).

8. 1) (0,9877; 1,0023); 2) Поскольку доверительный интервал включает специфицированное значение 1,0 литр, то владелец магазина не имеет причин утверждать, что объем краски в банке меньше 1, 0 литра. 3) Поскольку σ_x известно и $n = 50$, то согласно центральной предельной теореме можно утверждать, что X распределено приблизительно нормально.

9. Фирма в рекламном буклете завышает срок безотказной работы изделия. $H_0: m = 2900; H_1: m < 2900; t = -1{,}8 \in (-\infty; 1{,}677)$

10. а) $H_0: m = 21$; $H_1: m < 21$; $z = -6{,}32 \in (-\infty; -1{,}645)$. Время обслуживания клиента, равное 21 мин, в качестве норматива опытными данными не подтверждается; б) $H_0: m = 16; H_1: m < 16; z = -1{,}054 \notin (-\infty; -1{,}645)$. Время обслуживания клиента, равное 16 мин, в качестве норматива не противоречит опытным данным.

11. $n_A = 4$; $n_B = 5$; $\bar{x}_A = 30000$; $\bar{x}_B = 25000$; $s_{\text{A}}^2 = 30000$; $s_{\text{B}}^2 = 20000$; $H_0\colon m_A = m_B$; $H_1\colon m_A > m_B$. Генеральные дисперсии не равны и не предполагаются равными. $t = 46{,}6252 \in U$; $U = (2{,}35; +\infty)$. На уровне значимости $\alpha = 0{,}05$ можно утверждать, что средний оборот бирж в регионе *A* больше, чем в регионе *B*.

Приложения

Приложение 1

Таблица значений функции Лапласа $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$

x	Φ0(x)	x	Φ0(x)	x	Φ0(x)	x	Φ0(x)
0	0	0,24	0,09483	0,48	0,18439	0,72	0,26424
0,01	0,00399	0,25	0,09871	0,49	0,18793	0,73	0,2673
0,02	0,00798	0,26	0,10257	0,5	0,19146	0,74	0,27035
0,03	0,01197	0,27	0,10642	0,51	0,19497	0,75	0,27337
0,04	0,01595	0,28	0,11026	0,52	0,19847	0,76	0,27637
0,05	0,01994	0,29	0,11409	0,53	0,20194	0,77	0,27935
0,06	0,02392	0,3	0,11791	0,54	0,2054	0,78	0,2823
0,07	0,0279	0,31	0,12172	0,55	0,20884	0,79	0,28524
0,08	0,03188	0,32	0,12552	0,56	0,21226	0,8	0,28814
0,09	0,03586	0,33	0,1293	0,57	0,21566	0,81	0,29103
0,1	0,03983	0,34	0,13307	0,58	0,21904	0,82	0,29389
0,11	0,0438	0,35	0,13683	0,59	0,2224	0,83	0,29673
0,12	0,04776	0,36	0,14058	0,6	0,22575	0,84	0,29955
0,13	0,05172	0,37	0,14431	0,61	0,22907	0,85	0,30234
0,14	0,05567	0,38	0,14803	0,62	0,23237	0,86	0,30511
0,15	0,05962	0,39	0,15173	0,63	0,23565	0,87	0,30785
0,16	0,06356	0,4	0,15542	0,64	0,23891	0,88	0,31057
0,17	0,06749	0,41	0,1591	0,65	0,24215	0,89	0,31327
0,18	0,07142	0,42	0,16276	0,66	0,24537	0,9	0,31594
0,19	0,07535	0,43	0,1664	0,67	0,24857	0,91	0,31859
0,2	0,07926	0,44	0,17003	0,68	0,25175	0,92	0,32121
0,21	0,08317	0,45	0,17364	0,69	0,2549	0,93	0,32381
0,22	0,08706	0,46	0,17724	0,7	0,25804	0,94	0,32639
0,23	0,09095	0,47	0,18082	0,71	0,26115	0,95	0,32894

Продолжение прилож.1

x	**Φ_0(x)**	**x**	**Φ_0(x)**	**x**	**Φ_0(x)**	**x**	**Φ_0(x)**
0,96	0,33147	1,24	0,39251	1,52	0,43574	1,8	0,46407
0,97	0,33398	1,25	0,39435	1,53	0,43699	1,81	0,46485
0,98	0,33646	1,26	0,39617	1,54	0,43822	1,82	0,46562
0,99	0,33891	1,27	0,39796	1,55	0,43943	1,83	0,46638
1	0,34134	1,28	0,39973	1,56	0,44062	1,84	0,46712
1,01	0,34375	1,29	0,40147	1,57	0,44179	1,85	0,46784
1,02	0,34614	1,3	0,4032	1,58	0,44295	1,86	0,46856
1,03	0,34849	1,31	0,4049	1,59	0,44408	1,87	0,46926
1,04	0,35083	1,32	0,40658	1,6	0,4452	1,88	0,46995
1,05	0,35314	1,33	0,40824	1,61	0,4463	1,89	0,47062
1,06	0,35543	1,34	0,40988	1,62	0,44738	1,9	0,47128
1,07	0,35769	1,35	0,41149	1,63	0,44845	1,91	0,47193
1,08	0,35993	1,36	0,41309	1,64	0,4495	1,92	0,47257
1,09	0,36214	1,37	0,41466	1,65	0,45053	1,93	0,4732
1,1	0,36433	1,38	0,41621	1,66	0,45154	1,94	0,47381
1,11	0,3665	1,39	0,41774	1,67	0,45254	1,95	0,47441
1,12	0,36864	1,4	0,41924	1,68	0,45352	1,96	0,475
1,13	0,37076	1,41	0,42073	1,69	0,45449	1,97	0,47558
1,14	0,37286	1,42	0,4222	1,7	0,45543	1,98	0,47615
1,15	0,37493	1,43	0,42364	1,71	0,45637	1,99	0,4767
1,16	0,37698	1,44	0,42507	1,72	0,45728	2	0,47725
1,17	0,379	1,45	0,42647	1,73	0,45818	2,02	0,47831
1,18	0,381	1,46	0,42785	1,74	0,45907	2,04	0,47932
1,19	0,38298	1,47	0,42922	1,75	0,45994	2,06	0,4803
1,2	0,38493	1,48	0,43056	1,76	0,4608	2,08	0,48124
1,21	0,38686	1,49	0,43189	1,77	0,46164	2,1	0,48214
1,22	0,38877	1,5	0,43319	1,78	0,46246	2,12	0,483
1,23	0,39065	1,51	0,43448	1,79	0,46327	2,14	0,48382

Продолжение прилож.1

x	$Ф_0(x)$	x	$Ф_0(x)$	x	$Ф_0(x)$	x	$Ф_0(x)$
2,16	0,48461	2,58	0,49506	3	0,49865	4,05	0,49997
2,18	0,48537	2,6	0,49534	3,05	0,49886	4,1	0,49998
2,2	0,4861	2,62	0,4956	3,1	0,49903	4,15	0,49998
2,22	0,48679	2,64	0,49585	3,15	0,49918	4,2	0,49999
2,24	0,48745	2,66	0,49609	3,2	0,49931	4,25	0,49999
2,26	0,48809	2,68	0,49632	3,25	0,49942	4,3	0,49999
2,28	0,4887	2,7	0,49653	3,3	0,49952	4,35	0,49999
2,3	0,48928	2,72	0,49674	3,35	0,4996	4,4	0,49999
2,32	0,48983	2,74	0,49693	3,4	0,49966	4,45	0,5
2,34	0,49036	2,76	0,49711	3,45	0,49972	4,5	0,5
2,36	0,49086	2,78	0,49728	3,5	0,49977	4,55	0,5
2,38	0,49134	2,8	0,49744	3,55	0,49981	4,6	0,5
2,4	0,4918	2,82	0,4976	3,6	0,49984	4,65	0,5
2,42	0,49224	2,84	0,49774	3,65	0,49987	4,7	0,5
2,44	0,49266	2,86	0,49788	3,7	0,49989	4,75	0,5
2,46	0,49305	2,88	0,49801	3,75	0,49991	4,8	0,5
2,48	0,49343	2,9	0,49813	3,8	0,49993	4,85	0,5
2,5	0,49379	2,92	0,49825	3,85	0,49994	4,9	0,5
2,52	0,49413	2,94	0,49836	3,9	0,49995	4,95	0,5
2,54	0,49446	2,96	0,49846	3,95	0,49996	5	0,5
2,56	0,49477	2,98	0,49856	4	0,49997		

Приложение 2

Критические значения t-критерия Стьюдента

Число степеней свободы df	**Уровень значимости α (двусторонняя критическая область)**					
	0,10	**0,05**	**0,02**	**0,01**	**0,002**	**0,001**
1	6,31	12,7	31,82	63,7	318,3	637,0
2	2,92	4,30	6,97	9,92	22,33	31,6
3	2,35	3,18	4,54	5,84	10,22	12,9
4	2,13	2,78	3,75	4,00	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,40
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,70
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,28	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,96
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73	2,09	2,53	2,85	3,55	3,85
Число степеней свободы df	**0,05**	**0,025**	**0,01**	**0,005**	**0,001**	**0,0005**
	Уровень значимости α (односторонняя критическая область)					

Продолжение прилож.2

Число степеней свободы df	Уровень значимости α (двусторонняя критическая область)					
	0,10	**0,05**	**0,02**	**0,01**	**0,002**	**0,001**
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,51	3,79
23	1,71	2,07	2,50	2,81	3,49	3,77
24	1,71	2,06	2,49	2,80	3,47	3,74
25	1,71	2,06	2,49	2,79	3,45	3,72
26	1,71	2,06	2,48	2,78	3,44	3,71
27	1,71	2,05	2,47	2,77	3,42	3,69
28	1,70	2,05	2,46	2,76	3,40	3,66
29	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,55
60	1,07	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
Число степеней свободы df	**0,05**	**0,025**	**0,01**	**0,005**	**0,001**	**0,0005**
	Уровень значимости α (односторонняя критическая область)					

Приложение 3

Критические значения распределения Фишера-Снедекора при уровне значимости α =0,05

df_1 / df_2	1	2	3	4	5	6	8	12	24	∞
1	161,4	199,5	215,7	224,6	230,2	234	238,9	243,9	249	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62

Продолжение прилож.3

df_1 / df_2	1	2	3	4	5	6	8	12	24	∞
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,60	1,21
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,10
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,53	1,03
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

Приложение 4

Критические значения распределения Пирсона

Число степеней свободы df	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,6349	5,0239	3,8415	0,0039	0,0010	0,0002
2	9,2103	7,3778	5,9915	0,1026	0,0506	0,0201
3	11,3449	9,3484	7,8147	0,3519	0,2158	0,1148
4	13,2767	11,1433	9,4877	0,7107	0,4844	0,2971
5	15,0863	12,8325	11,0705	1,1455	0,8312	0,5543
6	16,8119	14,4494	12,5916	1,6354	1,2373	0,8721
7	18,4753	16,0128	14,0671	2,1674	1,6899	1,2390
8	20,0902	17,5346	15,5073	2,7326	2,1797	1,6465
9	21,6660	19,0228	16,9190	3,3251	2,7004	2,0879
10	23,2093	20,4832	18,3070	3,9403	3,2470	2,5582
11	24,7250	21,9201	19,6751	4,5748	3,8158	3,0535
12	26,2170	23,3367	21,0261	5,2260	4,4038	3,5706
13	27,6883	24,7356	22,3620	5,8919	5,0088	4,1069
14	29,1412	26,1190	23,6848	6,5706	5,6287	4,6604
15	30,5779	27,4884	24,9958	7,2609	6,2621	5,2294
16	31,9999	28,8454	26,2962	7,9617	6,9077	5,8122
17	33,4087	30,1910	27,5871	8,6718	7,5642	6,4078
18	34,8053	31,5264	28,8693	9,3905	8,2308	7,0149
19	36,1909	32,8523	30,1435	10,1170	8,9065	7,6327
20	37,5662	34,1696	31,4104	10,8508	9,5908	8,2604
21	38,9322	35,4789	32,6706	11,5913	10,2829	8,8972
22	40,2894	36,7807	33,9244	12,3380	10,9823	9,5425
23	41,6384	38,0756	35,1725	13,0905	11,6886	10,1957
24	42,9798	39,3641	36,4150	13,8484	12,4012	10,8564
25	44,3141	40,6465	37,6525	14,6114	13,1197	11,5240
26	45,6417	41,9232	38,8851	15,3792	13,8439	12,1982
27	46,9629	43,1945	40,1133	16,1514	14,5734	12,8785
28	48,2782	44,4608	41,3371	16,9279	15,3079	13,5647
29	49,5879	45,7223	42,5570	17,7084	16,0471	14,2565
30	50,8922	46,9792	43,7730	18,4927	16,7908	14,9535

Число степеней свободы k	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,99
31	52,1914	48,2319	44,9853	19,2806	17,5387	15,6555
32	53,4858	49,4804	46,1943	20,0719	18,2908	16,3622
33	54,7755	50,7251	47,3999	20,8665	19,0467	17,0735
34	56,0609	51,9660	48,6024	21,6643	19,8063	17,7892
35	57,3421	53,2034	49,8019	22,4650	20,5694	18,5089
36	58,6192	54,4373	50,9985	23,2686	21,3359	19,2327
37	59,8925	55,6680	52,1923	24,0749	22,1056	19,9602
38	61,1621	56,8955	53,3835	24,8839	22,8785	20,6914
39	62,4281	58,1201	54,5722	25,6954	23,6543	21,4262
40	63,6907	59,3417	55,7585	26,5093	24,4330	22,1643
41	64,9501	60,5606	56,9424	27,3256	25,2145	22,9056
42	66,2062	61,7768	58,1240	28,1441	25,9987	23,6501
43	67,4594	62,9904	59,3035	28,9647	26,7854	24,3976
44	68,7095	64,2015	60,4809	29,7875	27,5746	25,1480
45	69,9568	65,4102	61,6562	30,6123	28,3662	25,9013
46	71,2014	66,6165	62,8296	31,4390	29,1601	26,6572
47	72,4433	67,8207	64,0011	32,2676	29,9562	27,4159
48	73,6826	69,0226	65,1708	33,0981	30,7545	28,1770
49	74,9195	70,2224	66,3387	33,9303	31,5549	28,9407
50	76,1539	71,4202	67,5048	34,7643	32,3574	29,7067

Приложение 5

Критические точки распределения Дарбина-Уотсона

n	k=1		k=2		k=3		k=4		k=5	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	0.61	1.40	-	-	-	-	-	-	-	-
7	0.7	1.36	0.47	1.9	-	-	-	-	-	-
8	0.76	1.33	0.56	1.78	0.37	2.29	-	-	-	-
9	0.82	1.32	0.63	1.7	0.46	2.13	-	-	-	-
10	0.88	1.32	0.7	1.64	0.53	2.02	-	-	-	-
11	0.93	1.32	0.66	1.6	0.6	1.93	-	-	-	-
12	0.97	1.33	0.81	1.58	0.66	1.86	-	-	-	-
13	1.01	1.34	0.86	1.56	0.72	1.82	-	-	-	-
14	1.05	1.35	0.91	1.55	0.77	1.78	-	-	-	-
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,1	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,9	1,71	0,78	1,9	0,67	2,1
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,4	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,2	1,41	1,1	1,54	1	1,68	0,9	1,83	0,79	1,99
21	1,2	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,22	1,43	1,15	1,54	1,05	1,66	0,96	1,8	0,86	1,94
23	1,24	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,9	1,92
24	1,26	1,45	1,19	1,55	1,1	1,66	1,01	1,78	0,93	1,9
25	1,27	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,29	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,3	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,32	1,48	1,26	1,56	1,18	1,65	1,1	1,75	1,03	1,85
29	1,33	1,48	1,27	1,56	1,2	1,65	1,12	1,74	1,05	1,84
30	1,34	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
31	1,35	1,5	1,3	1,57	1,23	1,65	1,16	1,74	1,09	1,83
32	1,36	1,5	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82
33	1,37	1,51	1,32	1,58	1,26	1,65	1,19	1,73	1,13	1,81
34	1,38	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,15	1,81
35	1,39	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,8
36	1,4	1,52	1,35	1,59	1,29	1,65	1,24	1,73	1,18	1,8
37	1,41	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,8

Продолжение прилож.5

N	k=1		k=2		k=3		k=4		k=5	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
38	1,42	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,21	1,79
39	1,43	1,54	1,38	1,6	1,33	1,66	1,27	1,72	1,22	1,79
40	1,44	1,54	1,39	1,6	1,34	1,66	1,29	1,72	1,23	1,79
41	1,45	1,55	1,4	1,6	1,35	1,66	1,29	1,72	1,24	1,78
42	1,46	1,55	1,41	1,61	1,36	1,66	1.31	1.72	1,25	1,78
43	1,46	1,56	1,41	1,61	1,37	1,66	1,32	1,72	1,27	1,78
44	1,47	1,56	1,42	1,61	1,37	1,66	1,33	1,72	1,28	1,78
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78
46	1,48	1,57	1,44	1,62	1,39	1,67	1,34	1,72	1,3	1,77
47	1,49	1,57	1,44	1,62	1,39	1,67	1,35	1,72	1,31	1,77
48	1,49	1,58	1,45	1,62	1.41	1,67	1,36	1,72	1,32	1,77
49	1,5	1,58	1,46	1,62	1,41	1,67	1,37	1,72	1,32	1,77
50	1,5	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
51	1,51	1,59	1,47	1,63	1,43	1,67	1,38	1,72	1,34	1,77
52	1,51	1,59	1,47	1,63	1,43	1,68	1,39	1,72	1,35	1,77
53	1,52	1,59	1,48	1,63	1,44	1,68	1,4	1,72	1,36	1,77
54	1,52	1,6	1,48	1,64	1,45	1,68	1,41	1,72	1,37	1,77
55	1,53	1,6	1,49	1,64	1,45	1,68	1,41	1,72	1,38	1,77
56	1,53	1,6	1,49	1,64	1,46	1,68	1,42	1,72	1,38	1,77
57	1,54	1,61	1,5	1,64	1,46	1,68	1,43	1,72	1,39	1,77
58	1,54	1,61	1,5	1,65	1,47	1,68	1,43	1,72	1,39	1,77
59	1,54	1,61	1,51	1,65	1,47	1,69	1,44	1,73	1,4	1,77
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,5	1,7	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,52	1,7	1,49	1,74	1,46	1,77
75	1,6	1,65	1,57	1,68	1,54	1,71	1,51	1,74	1,49	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
85	1,62	1,67	1,6	1,7	1,57	1,72	1,55	1,75	1,52	1,77
90	1,63	1,68	1,61	1,7	1,59	1,73	1,57	1,75	1,54	1,78
95	1,64	1,69	1,62	1,71	1,6	1,73	1,58	1,75	1,56	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78

Приложение 6

Критические значения коэффициентов корреляции для уровня значимости 0,05

k	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
$r_{кр}$	1,00	1,00	0,88	0,81	0,75	0,71	0,67	0,63	0,60	0,58	0,55
k	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**25**	**30**
$r_{кр}$	0,53	0,51	0,50	0,48	0,47	0,46	0,44	0,42	0,42	0,38	0,35
k	**35**	**40**	**45**	**50**	**60**	**70**	**80**	**90**	**100**		
$r_{кр}$	0,32	0,30	0,29	0,27	0,25	0,29	0,22	0,21	0,19		

Приложение 7

Таблица значений z-преобразования Фишера

r	**0**	**1**	**2**	**3**	**4**
0,0	0	0,0101	0,02	0,03	0,04
0,1	0,1003	0,1104	0,1206	0,1308	0,1409
0,2	0,2027	0,2132	0,2237	0,2342	0,2448
0,3	0,3095	0,3205	0,3316	0,3428	0,3541
0,4	0,4236	0,4356	0,4477	0,4599	0,4722
0,5	0,5493	0,5627	0,5764	0,5901	0,6042
0,6	0,6932	0,7089	0,725	0,7414	0,7582
0,7	0,8673	0,8872	0,9077	0,9287	0,9505
0,8	1,0986	1,127	1,1568	1,1881	1,2212
0,9	1,4722	1,5275	1,589	1,6584	1,7381
0,99	2,6466	2,6996	2,7587	2,8257	2,9031

r	**5**	**6**	**7**	**8**	**9**
0,0	0,0501	0,0601	0,0701	0,0802	0,0902
0,1	0,1511	0,1614	0,1717	0,182	0,1923
0,2	0,2554	0,2661	0,2769	0,2877	0,2986
0,3	0,3654	0,3767	0,3884	0,4001	0,4118
0,4	0,4847	0,4973	0,5101	0,523	0,5361
0,5	0,6184	0,6328	0,6475	0,6625	0,6777
0,6	0,7753	0,7928	0,8107	0,8291	0,848
0,7	0,973	0,9962	1,0203	1,0454	1,0714
0,8	1,2562	1,2933	1,3331	1,3758	1,4219
0,9	1,8318	1,9459	2,0923	2,2976	2,6467
0,99	2,9945	3,1063	3,2504	3,4534	3,8002

В качестве примера (для понимания, как получилась таблица) посчитаем значение z для r=0,997 (число из строки 0,99 и столбца 7): z(0,997)=3,2504.

Приложение 8

H_0 - нулевая гипотеза;
H_1 - альтернативная гипотеза;
K - статистика критерия;
U - критическая область.

Гипотезы о дисперсиях

H_0	H_1	K	U
$\sigma^2=\sigma_0^2$	$\sigma^2<\sigma_0^2$	$\chi^2=\dfrac{(n-1)s^2}{\sigma_0^2}$	$\chi^2<{\chi^2}_{1-\alpha}(n-1)$
$\sigma^2=\sigma_0^2$	$\sigma^2>\sigma_0^2$	$\chi^2=\dfrac{(n-1)s^2}{\sigma_0^2}$	$\chi^2>{\chi^2}_{\alpha}(n-1)$
$\sigma^2=\sigma_0^2$	$\sigma^2\neq\sigma_0^2$	$\chi^2=\dfrac{(n-1)s^2}{\sigma_0^2}$	$\chi^2<{\chi^2}_{1-\alpha}(n-1)$ и $\chi^2>{\chi^2}_{\alpha}(n-1)$
$\sigma_1^2=\sigma_2^2$	$\sigma_1^2>\sigma_2^2$	$F=\dfrac{s_1^2}{s_2^2}$ $s_1^2>s_2^2$	$F>F_\alpha(df_1;df_2)$; $df_1=n_1-1$; $df_2=n_2-1$.

Замечание. Для гипотезы о дисперсии при $k>30$ $(k=n-1)$, $\chi_\alpha^2=k\left(1-\frac{2}{9k}+z_\alpha\sqrt{\frac{2}{9k}}\right)^3$; $\Phi_0(z_\alpha)=0{,}5-\alpha$.

Продолжение прилож.8

Гипотезы о среднем

H_0	H_1	K	U
σ^2 - известна или $n \geq 30$			
$m = m_0$	$m < m_0$	$z = \dfrac{\bar{x} - m_0}{\sigma/\sqrt{n}}$	$z < -z_\alpha$; $\Phi_0(z_\alpha) = 0{,}5 - \alpha$
$m = m_0$	$m > m_0$	$z = \dfrac{\bar{x} - m_0}{\sigma/\sqrt{n}}$	$z > z_\alpha$; $\Phi_0(z_\alpha) = 0{,}5 - \alpha$
$m = m_0$	$m \neq m_0$	$z = \dfrac{\bar{x} - m_0}{\sigma/\sqrt{n}}$	$\lvert z \rvert > z_{\alpha/2}$; $\Phi_0\left(z_{\alpha/2}\right) = 0{,}5 - \alpha/2$
σ^2 - не известна и $n < 30$			
$m = m_0$	$m < m_0$	$t = \dfrac{\bar{x} - m_0}{s/\sqrt{n-1}}$	$t < -t_\alpha$ $t_\alpha = t_{кр}(\alpha; n-1)$ односторонняя критическая область
$m = m_0$	$m > m_0$	$t = \dfrac{\bar{x} - m_0}{s/\sqrt{n-1}}$	$t > t_\alpha$ $t_\alpha = t_{кр}(\alpha; n-1)$ односторонняя критическая область
$m = m_0$	$m \neq m_0$	$t = \dfrac{\bar{x} - m_0}{s/\sqrt{n-1}}$	$\lvert t \rvert > t_{\alpha/2}$ $t_{\alpha/2} = t_{кр}(\alpha; n-1)$ двусторонняя критическая область

Приложение 9

Названия некоторых функций в Excel 2010 были изменены по сравнению с более ранними версиями.

Чтобы повысить точность работы функций MS Excel, обеспечить их согласованность и привести имена функций в соответствии с их назначением, корпорация Microsoft изменила, переименовала и добавила несколько функций в библиотеку MS Excel 2010.

Для обеспечения обратной совместимости переименованные функции доступны также и по их старым именам.

Название функции в Excel более ранних версий	**Название функции в Excel 2010**	**Примечания**
ДИСП(число1,[число2],...])	ДИСП.В(число1,[число2],...])	Оценивает дисперсию по выборке
СТЬЮДРАСПОБР(вероятность; степени_свободы)	СТЬДЕНТ.ОБР.2Х(вероятность, степени_свободы)	Возвращает двустороннее обратное *t*-распределение Стьюдента
FРАСПОБР (вероятность;степени_свободы1;степени_свободы2)	F.ОБР.ПХ(вероятность,степени_свободы1, степени_свободы2)	Возвращает значение, обратное (правостороннему) *F*-распределению вероятностей
ХИ2ОБР(вероятность,степени_ свободы)	ХИ2.ОБР.ПХ(вероятность,степени_свободы)	Возвращает обратное значение односторонней вероятности распределения хи-квадрат

Литература

1. Бородич С.А. Вводный курс эконометрики: Учебное пособие. − Мн.: БГУ, 2000. − 354 с.
2. Вуколов Э.А. Основы статистическое анализа.- М., Форум, 2008. – 464 с.
3. Ниворожкина Л.П., Морозова 3.А., Герасимова И.А., Житников И.В. Основы статистики с элементами теории вероятностей для экономистов: Руководство для решения задач. — Ростов н/Д: Феникс, 1999. — 320 с.
4. Эконометрика: Учебник/И.И. Елисеева, С.В. Курышева, Т.В. Костеева и др.; под ред. И.И. Елисеевой. – 2-е изд., перераб. и доп. - М.: Юрайт, 2012.
5. Ишханян М.В. Математическое моделирование. – М.: МГУПС (МИИТ), 2015. – 150с.

Св. план 2016 г., поз.212

Ишханян Маргарита Владимировна

Введение в эконометрику

Учебное пособие для направления «Экономика»

Подписано в печать Формат 60 Х 84 / 16

Заказ № Усл. - печ. л. -

Тираж - 100 экз.