

федеральное государственное бюджетное образовательное учреждение

высшего профессионального образования
**«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ПУТЕЙ СООБЩЕНИЯ»**

Кафедра «Математика»

Л.Ф. Кочнева, З.С. Липкина, В. И. Новосельцева

Теория вероятностей и математическая статистика

Часть III

Математическая статистика

Учебное пособие

МОСКВА - 2012

федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования

**«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ПУТЕЙ СООБЩЕНИЯ»**

Кафедра «Математика»

Л.Ф. Кочнева, З.С. Липкина, В. И. Новосельцева

Теория вероятностей и математическая статистика

Часть III

Математическая статистика

**Рекомендовано редакционно-издательским
советом университета
в качестве учебного пособия
для бакалавров направления «Экономика»**

МОСКВА - 2012

Введение

При исследовании экономических процессов используются методы и модели математической статистики, позволяющие получить необходимые знания об исследуемом объекте.

Одним из основных методов математической статистики является выборочный метод. При этом используются понятия генеральной совокупности и выборочной совокупности.

Генеральная совокупность – это случайная величина X , распределенная по неизвестному закону и имеющая такое достаточно большое количество элементов, что сплошное обследование трудоемко или невозможно.

Выборочная совокупность или просто выборка – это совокупность случайно отобранных объектов из генеральной совокупности. К выборочной совокупности предъявляются определенные математические требования: выборка должна обладать всеми свойствами генеральной совокупности, т. е. должна быть представительной (репрезентативной); на основании закона больших чисел теории вероятностей отбор объектов для исследования должен носить случайный характер; каждый объект генеральной совокупности должен иметь одну и ту же вероятность того, что он попадет в выборку.

Таким образом, задача математической статистики состоит в следующем:

- 1) Осуществление математически правильного отбора данных для обследования;
- 2) Получение оценки функции распределения вероятностей и ее параметров;
- 3) Проверка статистических гипотез о виде законов распределения и величине параметров распределения, вид которых известен.

По результатам статистического исследования получают научные и практические выводы

1. Статистическое распределение выборки

Пусть из генеральной совокупности извлечена выборка. При этом x_1 наблюдалось n_1 раз, x_2 наблюдалось n_2 раз, ..., x_k наблюдалось n_k раз. Сумма $\sum_{i=1}^k n_i = n$ называется объемом выборки. Наблюдаемые значения называют вариантами. Последовательность вариантов, записанных в возрастающем порядке, называют вариационным рядом.

Статистическим распределением выборки называют таблицу частот:

x_i	x_1	x_2	...	x_k	варианты
n_i	n_1	n_2	...	n_k	частоты

$$\sum_{i=1}^k n_i = n$$

или относительных частот:

x_i	x_1	x_2	...	x_k	варианты
$\frac{n_i}{n}$	$\frac{n_1}{n}$	$\frac{n_2}{n}$...	$\frac{n_k}{n}$	относительные частоты

$$\sum_{i=1}^k \frac{n_i}{n} = 1$$

Если объем выборки достаточно большой, то весь массив значений разбивается на k интервалов и подсчитывается число элементов выборки, попавших в соответствующий интервал. Элемент, попадающий на границу двух интервалов, учитывается в левом интервале.

Статистическое распределение в этом случае имеет вид:

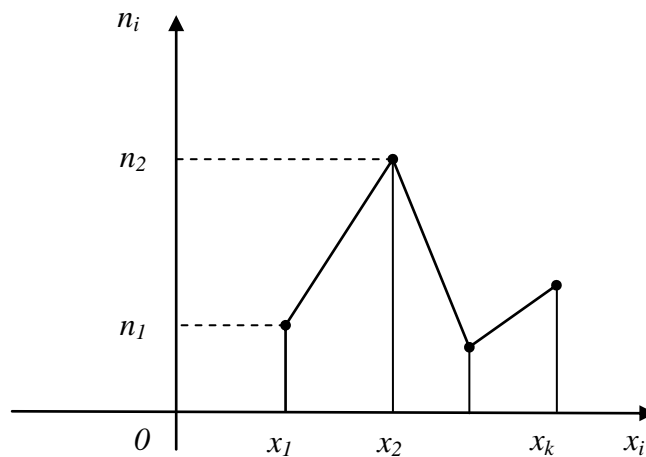
$(x_{i-1}; x_i)$	$(x_0; x_1)$	$(x_1; x_2)$...	$(x_{k-1}; x_k)$	варианты
n_i	n_1	n_2	...	n_k	частоты
$\frac{n_i}{n}$	$\frac{n_1}{n}$	$\frac{n_2}{n}$...	$\frac{n_k}{n}$	относительные частоты

2. Полигон и гистограмма

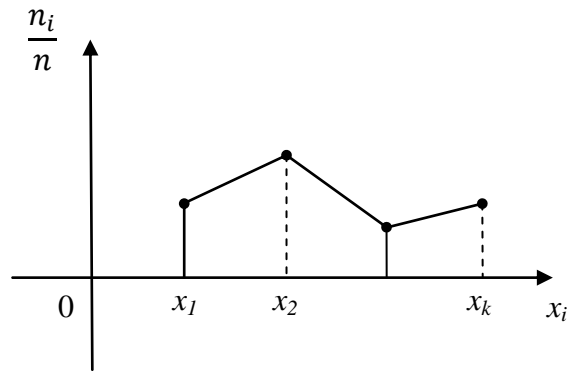
Для наглядности строят различные графики статистических распределений.

Полигоном частот называют ломаную, соединяющую точки $(x_i; n_i)$. По оси абсцисс откладывают x_i , а по оси ординат n_i .

Полигоном относительных частот называют ломаную, соединяющую точки $(x_i; \frac{n_i}{n})$.

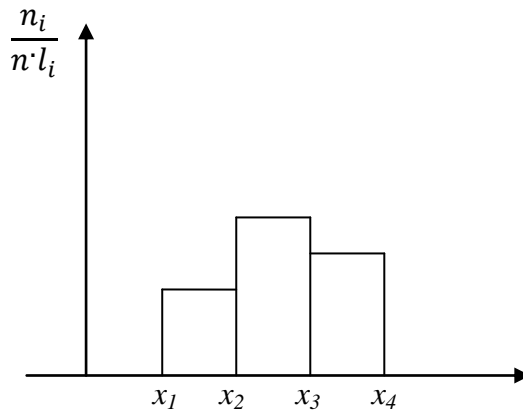


Полигон частот



Полигон относительных частот

Гистограмма строится по данным интервального статистического распределения и представляет собой набор прямоугольников, основаниями которых являются интервалы, а высоты равны: $h_i = \frac{n_i}{n \cdot l_i}$, где l_i -длина i -го интервала, то есть площадь i -го прямоугольника равна относительной частоте попадания элементов выборки в i -й интервал. Площадь всей гистограммы равна 1.



Таким образом, гистограмма относительных частот является аналогом плотности вероятностей.

Замечание. Возможно построение гистограммы частот. В этом случае площадь i -го прямоугольника равна частоте попадания элементов выборки в i -й интервал, а площадь всей гистограммы равна объему выборки n .

Гистограмма частот и гистограмма относительных частот отражают один и тот же характер плотности вероятностей генеральной совокупности. Разница только в выборе масштаба.

Пример 1. Дана выборка: 1, 3, 2, 1, 0, 4, 3, 1, 0, 2.

Составить статистическое распределение (таблицу частот) и построить полигон.

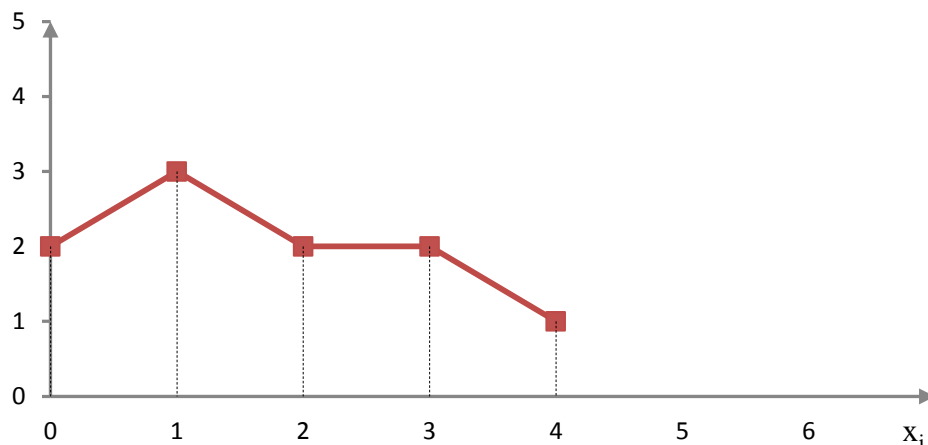
Решение.

Таблица частот:

x_i	0	1	2	3	4
n_i	2	3	2	2	1

$n=10$

Полигон частот:



Пример 2. Дана выборка: 13, 15, 20, 12, 11, 35, 17, 38, 23, 27, 24, 39, 22, 25, 36, 18, 16, 23, 14, 27, 24, 26, 18, 37, 23, 14, 39, 27, 13, 22, 33, 31, 18, 35, 24, 14, 23, 34, 37, 12, 11, 11, 23, 13, 34, 25, 15, 23, 13, 17.

Требуется составить интервальное статистическое распределение и построить гистограмму относительных частот, разбив выборку на 5 интервалов.

Решение. Выберем сначала длину интервала. Так как размах выборки $R=x_{\text{наиб}}-x_{\text{наим}}=$

$=39-11=28$, то длина интервала должна быть больше, чем $\frac{28}{5}=5,6$. Поэтому возьмем длину интервала $l=6$. Начало первого интервала нужно выбрать так, чтобы $x_1 \leq x_{\text{min}}$. В примере возьмем $x_1=10 < 11$. Записав интервалы, определим количество чисел, попавших в каждый интервал. Если число совпадает с границей интервала, то запишем его в левый интервал.

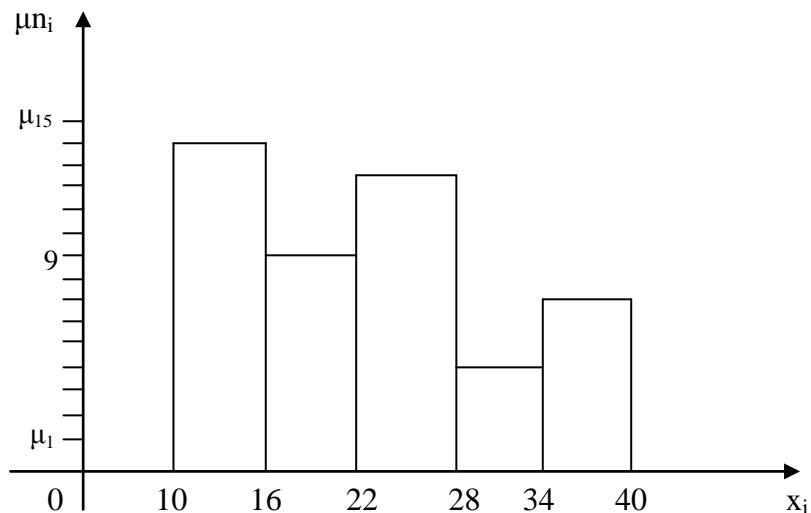
Таким образом, получим интервальное статистическое распределение:

(x_i-x_{i+1})	10-16	16-22	22-28	28-34	34-40	
n_i	15	9	14	4	8	$n=50$
$\frac{n_i}{n}$	0,30	0,18	0,28	0,08	0,16	$\sum \frac{n_i}{n} = 1$

Для построения гистограммы высоту i -го прямоугольника вычисляем по формуле:

$$h_i = \frac{n_i}{n \cdot l_i} = \frac{n_i}{50 \cdot 6} = \frac{1}{300} n_i = \mu n_i.$$

Число $\mu = \frac{1}{300}$ будем считать масштабной единицей на оси ординат.



На полученном рисунке гистограмма интервального статистического распределения данной выборочной совокупности.

3. Эмпирическая функция распределения

Напомним, что в теории вероятностей функцией распределения $F(x)$ называется $P(X < x)$ – вероятность того, что случайная величина X (дискретная или непрерывная) расположена слева от фиксированного значения x , то есть $F(x) = P(X < x)$.

В математической статистике случайная величина X – это генеральная совокупность, которую мы изучаем по выборке. И по выборке строим аналогичную функцию:

$F^*(x) = \frac{\sum n_i}{n}$ для всех $x_i < x$, где x – фиксированное значение из выборки, в числителе $\sum n_i$ означает количество элементов выборки с учетом частоты по величине меньших выбранного элемента x .

Другими словами теоретическая функция $F(x)$ определяет вероятность события $X < x$, а эмпирическая $F^*(x)$ определяет относительную частоту этого же события.

Доказано (см. т. Бернулли в законе больших чисел), что относительная частота и вероятность события при больших n мало отличаются одно от другого, то есть выполняется:

$$\lim_{n \rightarrow \infty} P(|F(x) - F^*(x)| < \varepsilon) = 1, \varepsilon > 0.$$

Из определения функции $F^*(x)$ вытекают ее свойства, аналогичные свойствам $F(x)$:

1. $0 \leq F^*(x) \leq 1$;
2. $F^*(x)$ – неубывающая функция;
3. Если x_1 наименьшая варианта, то $F^*(x) = 0$ для всех $x \leq x_1$;
4. Если x_k наибольшая варианта, то $F^*(x) = 1$ для всех $x > x_k$.

Таким образом, эмпирическая функция распределения выборки служит для оценки теоретической функции распределения генеральной совокупности.

Пример. Составить эмпирическую функцию распределения для выборки из предыдущего параграфа:

x_i	0	1	2	3	4
n_i	2	3	2	2	1

$$n = 10$$

Решение. Из определения F^* следует:

$$F^*(0)=0, x \leq 0;$$

$$F^*(2)=\frac{2+3}{10}=0,5;$$

$$F^*(1)=\frac{2}{10}=0,2;$$

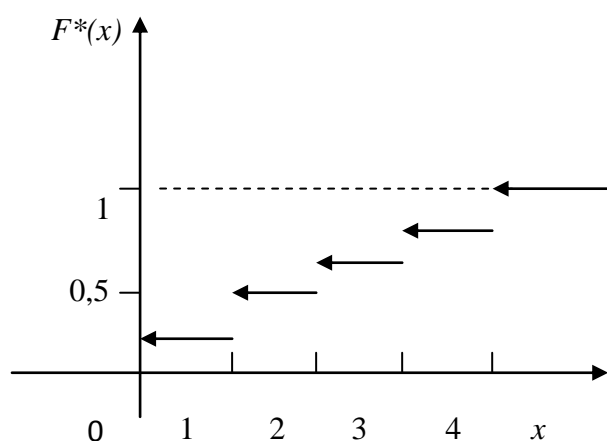
$$F^*(3)=\frac{2+3+2}{10}=0,7;$$

$$F^*(4)=\frac{2+3+2+2}{10}=0,9;$$

$$F^*(x)=1, x > 4.$$

Таким образом,

График $F^*(x)$



$$F^*(x) = \begin{cases} 0, & x \leq 0, \\ 0,2, & 0 < x \leq 1, \\ 0,5, & 1 < x \leq 2, \\ 0,7, & 2 < x \leq 3, \\ 0,9, & 3 < x \leq 4, \\ 1, & x > 4. \end{cases}$$

5. Примеры теоретических законов распределения

В теории вероятностей законы распределения дискретных и непрерывных случайных величин определяются параметрами и числовыми характеристиками: математическим ожиданием MX и дисперсией DX .

Например, дискретные законы распределения вероятностей:

1. Биномиальный закон.

$$P(X=k) = C_n^k p^k q^{n-k}, \quad k=0, 1, 2, \dots, n.$$

n -количество испытаний;

p -вероятность успеха в одном испытании;

$q=1-p$, k -количество успехов;

C_n^k -число сочетаний из n по k элементов.

$$M(X) = np, \quad DX = npq.$$

2. Закон Пуассона.

$$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k=0, 1, 2, \dots$$

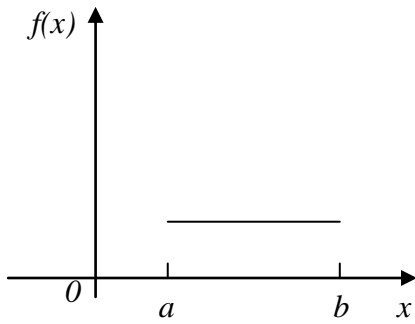
Число $\lambda = np$, где n велико, вероятность p мала.

$$MX = DX = \lambda.$$

Наиболее распространенными законами распределения непрерывных случайных величин являются следующие законы:

1. Равномерный закон, плотность вероятностей которого:

График



$$f(x) = \begin{cases} 0, & x < a, \\ \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & x > b. \end{cases}$$

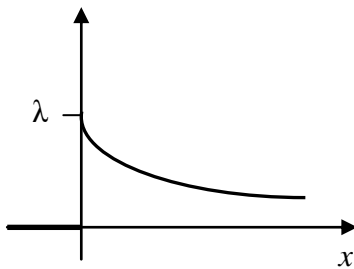
$MX = \frac{a+b}{2}$, $DX = \frac{(b-a)^2}{12}$. Вероятность попадания X в интервал (α, β) вычисляется по формуле:

$$P(\alpha \leq X < \beta) = \frac{\beta - \alpha}{b - a} \text{ при } a \leq \alpha \leq \beta \leq b.$$

2. Показательный закон.

Плотность вероятностей:

График



$$f(x) = \begin{cases} 0, & x < 0, \\ \lambda e^{-\lambda x}, & x \geq 0. \end{cases}$$

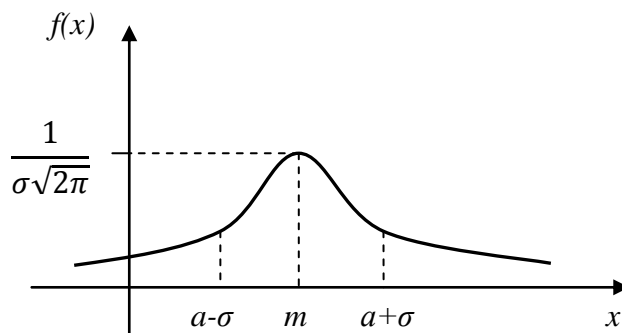
$$MX = \frac{1}{\lambda}, \quad DX = \frac{1}{\lambda^2}.$$

$$P(\alpha \leq X < \beta) = e^{-\lambda \alpha} - e^{-\lambda \beta} \text{ при } 0 \leq \alpha \leq \beta.$$

3. Нормальный закон.

Плотность вероятностей: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$,

График



$$MX = m, \quad DX = \sigma^2.$$

Вероятность попадания X в интервал (α, β)

$$P(\alpha \leq X < \beta) = \Phi\left(\frac{\beta - m}{\sigma}\right) - \Phi\left(\frac{\alpha - m}{\sigma}\right), \text{ где } \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt.$$

6. Статистические оценки параметров распределения

Параметры генеральной совокупности это $\bar{x}_Г$ -средняя генеральная величина, D_σ -дисперсия генеральная и $\sigma_\sigma = \sqrt{D_\sigma}$ -среднее квадратическое отклонение. Эти параметры неизвестны.

По случайной выборочной совокупности находят среднюю выборочную $\bar{x}_В$, дисперсию выборочную D_σ и среднее выборочное отклонение $\sigma_\sigma = \sqrt{D_\sigma}$. Выборочные параметры являются случайными числами, их называют оценками неизвестных параметров генеральной совокупности.

К оценкам предъявляются определенные статистические требования: несмещенность, состоятельность и эффективность.

Введем обозначения: пусть θ -неизвестный параметр генеральной совокупности, θ^* -оценка этого параметра, вычисленная по выборочной совокупности.

1. Несмещенность. Оценка θ^* называется несмещенной, если $M\theta^* = \theta$, то есть математическое ожидание несмещенной оценки равно ее истинному значению.
2. Состоятельность. Оценка θ^* называется состоятельной, если с ростом объема выборки значение $\theta^* \rightarrow \theta$ по вероятности, то есть выполняется:

$$\lim_{n \rightarrow \infty} P(|\theta^* - \theta| < \varepsilon) = 1, \varepsilon > 0.$$

При этом достаточным условием состоятельности является стремление дисперсии $D\theta^* \rightarrow 0$ при $n \rightarrow \infty$.

3. Эффективность. Оценка θ^* называется эффективной, если дисперсия $D\theta^*$ принимает минимальное значение.

Говорят θ^*_1 эффективнее θ^*_2 , если $D\theta^*_1 < D\theta^*_2$.

7. Оценка генеральной средней по выборочной средней

Пусть для изучения генеральной совокупности X извлечена выборка объема n . Различные значения x_1, x_2, \dots, x_k и соответствующие частоты n_1, n_2, \dots, n_k помещены в таблицу:

x_i	x_1	x_2	...	x_k	варианты
n_i	n_1	n_2	...	n_k	частоты

$$\sum n_i = n - \text{объем выборки}$$

$$\text{Выборочная средняя } \bar{x}_В = \frac{x_1 n_1 + x_2 n_2 + \dots + x_k n_k}{n} = \frac{1}{n} \sum x_i n_i.$$

Так как при повторении опыта получается другая выборка и другое значение $\bar{x}_В$, то можно говорить о математическом ожидании средней выборочной $M\bar{x}_В$ и дисперсии $D\bar{x}_В$.

Покажем, что $\bar{x}_В$ является несмещенной оценкой генеральной средней $\bar{x}_Г$. Заметим, что $\bar{x}_Г$ совпадает с математическим ожиданием генеральной совокупности X , то есть

$$\bar{x}_Г = MX.$$

Выборки, полученные в результате независимых испытаний, можно считать случайными величинами X_1, X_2, \dots, X_n , имеющими одинаковое распределение, совпадающее

с распределением генеральной совокупности X , то есть числовые характеристики случайных величин $X_i (i=1, 2, 3, \dots, n)$ одинаковые:

$$MX_1 = MX_2 = \dots = MX_n = a = MX = \bar{x}_r,$$

$$DX_1 = DX_2 = \dots = DX_n = DX.$$

Рассмотрим математическое ожидание средней выборочной:

$$M\bar{x}_b = M \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} (MX_1 + MX_2 + \dots + MX_n) = \frac{na}{n} = a = MX.$$

Поэтому имеем $M\bar{x}_b = \bar{x}_r$, следовательно, \bar{x}_b является несмещенной оценкой генеральной средней.

Состоятельность средней выборочной \bar{x}_b доказывает теорема Чебышева из закона больших чисел:

$$\lim_{n \rightarrow \infty} P(|\bar{x}_b - \bar{x}_r| < \varepsilon) = 1, \varepsilon > 0.$$

То есть $\bar{x}_b \rightarrow \bar{x}_r$ по вероятности при $n \rightarrow \infty$.

Таким образом, средняя выборочная является несмещенной состоятельной и, можно доказать, эффективной оценкой генеральной средней.

8. Среднее квадратическое отклонение

Так же, как в предыдущем параграфе 7, будем обозначать генеральную совокупность через дискретную случайную величину X достаточно большого объема. Тогда генеральная средняя $\bar{x}_r = MX$ и $D_r = DX$ и по определению:

$$DX = M(X - MX)^2 = MX^2 - (MX)^2.$$

Эта формула выведена при изучении случайных величин (см. раздел «Теория вероятностей»). Выборочная дисперсия находится аналогично по выборочной совокупности:

$$D^* = \frac{1}{n} \sum (x_i - \bar{x}_b)^2 n_i, \text{ среднее выборочное отклонение } \sigma^* = \sqrt{D^*}.$$

Обратим внимание на то, что сумма произведений отклонений $x_i - \bar{x}$ на соответствующие частоты n_i равна нулю.

Действительно:

$$\sum (x_i - \bar{x}) n_i = \sum x_i n_i - \sum \bar{x} n_i = n\bar{x} - \bar{x} \sum n_i = n\bar{x} - n\bar{x} = 0.$$

Пример. Имеется выборка:

x_i	1	2	3	
n_i	2	3	5	$n=10$

Вычислить:

- 1) Среднюю величину отклонения;
- 2) Среднее квадратическое отклонение.

Решение:

$$\bar{x} = \frac{1 \cdot 2 + 2 \cdot 3 + 3 \cdot 5}{10} = 2,3.$$

$$\begin{aligned} \frac{1}{n} \sum (x_i - \bar{x}) n_i &= \frac{1}{10} [(1 - 2,3) \cdot 2 + (2 - 2,3) \cdot 3 + (3 - 2,3) \cdot 5] \\ &= \frac{1}{10} (-1,3 \cdot 2 - 0,3 \cdot 3 + 0,7 \cdot 5) = \frac{1}{10} (-2,6 - 0,9 + 3,5) \\ &= \frac{1}{10} (-3,5 + 3,5) = 0 \end{aligned}$$

$$D^* = \frac{1}{10} [(1 - 2,3)^2 \cdot 2 + (2 - 2,3)^2 \cdot 3 + (3 - 2,3)^2 \cdot 5] = \frac{1}{10} [(-1,3)^2 \cdot 2 + (-0,3)^2 \cdot 3 + 0,7^2 \cdot 5] = \frac{1}{10} [1,69 \cdot 2 + 0,09 \cdot 3 + 0,49 \cdot 5] = \frac{1}{10} (3,38 + 0,27 + 2,45) = \frac{6,1}{10} =$$

$$0,61. \sigma^* = \sqrt{0,61} = 0,78.$$

Ответ: 1. Среднее отклонение равно нулю.

2. Среднее квадратическое отклонение равно 0,78.

Из примеров видно, что вычисление дисперсии по определению достаточно громоздко. Выведем формулу, которой обычно пользуются на практике:

$$D^* = \frac{1}{n} \sum n_i (x_i - \bar{x})^2 = \frac{1}{n} \sum n_i (x_i^2 - 2x_i \bar{x} + \bar{x}^2) = \frac{1}{n} (\sum n_i x_i^2 - 2 \sum x_i n_i \bar{x} + \sum n_i \bar{x}^2) =$$

$$\sum \frac{x_i^2 n_i}{n} - 2\bar{x} \sum \frac{x_i n_i}{n} + (\bar{x})^2 \sum \frac{n_i}{n} = \overline{x^2} - 2(\bar{x})^2 + (\bar{x})^2 = \overline{x^2} - (\bar{x})^2, \text{ то есть } D^* = \overline{x^2} - (\bar{x})^2$$

Выборочная дисперсия равна усредненному квадрату выборочных значений минус средняя выборочная величина в квадрате.

Пример. Возьмем выборку из предыдущего примера:

x_i	1	2	3	
n_i	2	3	5	$n=10$

Средняя величина $\bar{x}=2,3$.

Средний квадрат выборочных значений:

$$\overline{x^2} = \frac{1}{10} (1^2 \cdot 2 + 2^2 \cdot 3 + 3^2 \cdot 5) = \frac{1}{10} (2 + 12 + 45) = 5,9;$$

$$D^* = 5,9 - (2,3)^2 = 5,9 - 5,29 = 0,61.$$

1. Оценка генеральной дисперсии по исправленной выборочной

Покажем, что выборочная дисперсия D^* является смещенной оценкой генеральной дисперсии DX . Так же, как и в пункте 6, будем считать выборочные совокупности одинаково распределенными случайными величинами, имеющими одинаковые числовые характеристики.

Выведено в 7, что для средней \bar{X} одинаково распределенных величин выполняется:

$$\begin{aligned} M\bar{X} &= MX, \\ M\bar{X}^2 &= MX^2, \\ D\bar{X} &= \frac{DX}{n}. \end{aligned} \quad (1)$$

$$\text{Рассмотрим } M(D^*) = M(\overline{X^2} - (\bar{X})^2) = M(\overline{X^2}) - M(\bar{X})^2; \quad (2)$$

По определению $D(\bar{X}) = M(\bar{X})^2 - (M\bar{X})^2$, поэтому получим равенство:

$$D(\bar{X}) = M(\bar{X})^2 - (M\bar{X})^2 \quad (3)$$

Из (3) следует, учитывая (1)

$$M(\bar{X})^2 = \frac{DX}{n} + (MX)^2 \quad (4)$$

Подставим теперь в (4) в (2):

$$M(D^*) = M(\overline{X^2}) - \frac{DX}{n} - (MX)^2 = DX - \frac{DX}{n} = \frac{n-1}{n} DX.$$

Таким образом, $M(D^*) \neq DX$ и, следовательно, дисперсия выборочная является смещенной оценкой генеральной дисперсии. Однако можно заметить, что $\frac{n-1}{n} \rightarrow 1$ при $n \rightarrow \infty$, то есть для больших объемов выборки D^* фактически будет несмещенной.

При $n \leq 30$ используют исправленную дисперсию:

$$S^2 = \frac{n}{n-1} D^*,$$

которая является несмещенной оценкой DX .

$$\text{Действительно: } M(S^2) = M\left(\frac{n}{n-1} D^*\right) = \frac{n}{n-1} M D^* = \frac{n}{n-1} \cdot \frac{n-1}{n} DX = DX.$$

Вычисляется S^2 по формуле:

$$S^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 n_i.$$

$S = \sqrt{S^2}$ называется стандартным отклонением.

10. Статистические (точечные) оценки параметров основных законов распределения

1. Нормальное распределение: $X=N(m,\sigma)$. Теоретические параметры: $m=MX$, $\sigma^2=DX$.
Оценки параметров $m^*=\bar{X}$ -средняя арифметическая величина, определяемая по выборке и $S^2=D^*$ -исправленная выборочная дисперсия.

$$\text{Асимметрия } A_s^* = \frac{(\overline{X-\bar{X}})^3}{(\sigma^*)^3} \text{ и эксцесс } E_x^* = \frac{(\overline{X-\bar{X}})^4}{(\sigma^*)^4} - 3.$$

Оценка асимметрии закона вычисляется по формуле:

$$A_s^* = \frac{(\overline{X-\bar{X}})^3}{(\sigma^*)^3} = \frac{1}{(\sigma^*)^3} (\overline{x^3} - 3\overline{x^2} \cdot \bar{x} + 2(\bar{x})^3).$$

Точечная оценка коэффициента эксцесса находится по формуле:

$$E_x^* = \frac{1}{(\sigma^*)^4} (\overline{(x - \bar{x})^4} - 3).$$

2. Показательное распределение: $X=E(\lambda)$

$$\text{Параметр закона } \lambda, M(X)=\frac{1}{\lambda}; D(X)=\frac{1}{\lambda^2}.$$

Оценка параметра $\lambda^*=\frac{1}{\bar{X}}$; \bar{X} -средняя выборочная.

3. Равномерное распределение: $X=R[a,b]$

$$\text{Теоретические параметры, } MX=\frac{a+b}{2}, DX=\frac{(b-a)^2}{12}.$$

Оценки параметров:

$$\text{средняя выборочная } \bar{X} = \frac{a^*+b^*}{2} \text{ и исправленная выборочная дисперсия } S^2 = \frac{(b^*-a^*)^2}{12}.$$

11. Метод вычисления точечных оценок параметров распределения с помощью введения условных вариантов

В теории вероятностей введены понятия начальных и центральных моментов для дискретных случайных величин.

Начальным моментом порядка k случайной величины X называют математическое ожидание величины X^k :

$$v_k = M(X^k)$$

Заметим, что $v_1 = M(X)$, $v_2 = M(X^2)$.

Центральным моментом k -го порядка случайной величины X называют математическое ожидание величины $(X-MX)^k$, т. е.:

$$\mu_k = M(X-MX)^k.$$

Очевидно, что:

$$\begin{aligned} \mu_1 &= M(X-MX) = 0, \\ \mu_2 &= M(X-MX)^2 = DX. \end{aligned}$$

По аналогии в математической статистике вводятся эмпирические моменты:

а) начальные моменты: $v_k^* = \frac{\sum x_i^k n_i}{n}$ и

б) центральные моменты: $\mu_k^* = \frac{1}{n} \sum (x_i - \bar{x})^k n_i$, при этом $v_1^* = \bar{x}$ - средняя величина, $\mu_2^* = D^*$ -дисперсия выборочная.

Центральные моменты выражаются через начальные:

$$\begin{aligned} \mu_2^* &= v_2^* - (v_1^*)^2 \\ \mu_3^* &= v_3^* - 3v_2^*v_1^* + 2(v_1^*)^3, \\ \mu_4^* &= v_4^* - 4v_3^*v_1^* + 6v_2^*v_1^{*2} - 3(v_1^*)^4. \end{aligned}$$

Условные варианты вводятся следующим образом. Пусть имеется вариационный ряд с равноотстоящими значениями с разностью h между двумя соседними значениями. Условными вариантами называются величины $U_i = \frac{x_i - c}{h}$, где в качестве c -ложного нуля можно выбрать произвольную варианту x_m .

Покажем, что варианты U_i являются целыми числами.

Действительно:

$$U_i = \frac{x_i - x_m}{h} = \frac{[x_1 + (i-1) \cdot h] - [x_1 + (m-1) \cdot h]}{h} = \frac{ih - h - mh + h}{h} = i - m.$$

Так как i, m -целые числа, то U_i будут также целыми числами.

На практике нужно в качестве ложного нуля выбирать варианту с наибольшей частотой.

Рассмотрим пример вычисления средней выборочной и дисперсии с помощью введения условных вариантов.

Пусть имеется выборка:

x_i	16,5	19,5	22,5	25,5	28,5
n_i	5	10	20	15	10
U_i	-2	-1	0	1	2

$$h=3, n=60.$$

Выберем $c=22,5$.

$$\text{Тогда } U_1 = \frac{16,5 - 22,5}{3} = -2; U_2 = \frac{19,5 - 22,5}{3} = -1; U_3 = 0; U_4 = \frac{25,5 - 22,5}{3} = 1; U_5 = \frac{28,5 - 22,5}{3} = 2.$$

Используя условные варианты, вычислим начальный момент:

$$v_1^* = \frac{1}{n} \sum U_i n_i = \frac{1}{60} (-2 \cdot 5 - 10 + 15 + 20) = \frac{15}{60} = 0,25;$$

$$v_2^* = \frac{1}{n} \sum U_i^2 n_i = \frac{1}{60} ((-2)^2 \cdot 5 + 10 + 15 + 4 \cdot 10) = 1,42.$$

Покажем теперь, что средняя величина вычисляется по формуле:

$$\bar{x} = v_1^* \cdot h + c.$$

Действительно:

$$v_1^* = \frac{1}{n} \sum U_i n_i = \frac{1}{n} \sum \frac{x_i - c}{h} n_i = \sum \frac{1}{h} (x_i - c) \frac{n_i}{n} = \frac{1}{h} (\sum \frac{x_i n_i}{n} - c \sum \frac{n_i}{n}) = \frac{1}{h} (\bar{x} - c), \sum \frac{n_i}{n} = 1.$$

Таким образом, $v_1^* = \frac{1}{h} (\bar{x} - c) \Rightarrow \bar{x} = v_1^* \cdot h + c$.

В примере $\bar{x} = 0,25 \cdot 3 + 22,5 = 23,25$.

Для вычисления дисперсии с использованием условных вариантов возьмем формулу:

$$\mu_2^* = v_2^* - (v_1^*)^2$$

$$v_2^* = \frac{1}{n} \sum U_i^2 n_i = \frac{1}{60} (4 \cdot 5 + 1 \cdot 10 + 1 \cdot 15 + 4 \cdot 10) = 1,42.$$

Тогда $\mu_2^* = 1,42 - (0,25)^2 = 1,3575 \approx 1,36$.

Аналогично можно доказать, что $D_e = \mu_2^* h^2$. Вообще, для любого k , $\mu_k(X) = h^k \mu_k(U)$ и $A_s^*(U) = A_s^*(X)$; $E_x^*(U) = E_x^*(X)$.

Дисперсия выборочная будет равна:

$$D_e = \mu_2^* h^2 = 1,36 \cdot 9 = 12,24.$$

12. Метод наибольшего правдоподобия для точечной оценки параметров распределения

Пусть X -дискретная случайная величина и выборка из нее: x_1, x_2, \dots, x_n . Допустим, что предполагаемый закон распределения вероятностей выбран и требуется найти

точечную оценку параметра θ закона. Обозначим через $p(x_i; \theta)$ вероятность того, что величина X примет значение x_i ($i=1, 2, \dots, n$)

Функцией правдоподобия дискретной случайной величины называется функция:

$$L(x_1, x_2, \dots, x_n, \theta) = p(x_1, \theta) p(x_2, \theta) \dots p(x_n, \theta) \dots$$

Находят такое значение θ^* , при котором функция достигает максимума и θ^* называют оценкой наибольшего правдоподобия.

Замечено, что функции L и $\ln L$ достигают максимума при одном и том же значении θ , поэтому исследуем на экстремум логарифмическую функцию правдоподобия:

$$\ln L = \ln p(x_1, \theta) + \ln p(x_2, \theta) + \dots + \ln p(x_n, \theta).$$

Для этого по известному правилу математического анализа нужно:

1. Найти производную $\frac{d(\ln L)}{d\theta}$;
2. Приравнять производную к нулю и, решая уравнение $\frac{d(\ln L)}{d\theta} = 0$, вычислить θ^* ;
3. Найти производную второго порядка $\frac{d^2(\ln \theta)}{d\theta^2}$; подставив в нее значение $\theta = \theta^*$, убедиться в том, что вторая производная отрицательна, и, следовательно, θ^* -точка максимума, которую принимают в качестве оценки наибольшего правдоподобия параметра θ .

Метод наиболее применим для малых выборок. Недостаток метода состоит в том, что приводит к сложным вычислениям.

Пример. Найти методом наибольшего правдоподобия оценку параметра λ в распределении вероятностей Пуассона:

$$P_n(m) = \frac{\lambda^m}{m!} e^{-\lambda}$$

Решение. Пусть проводится k опытов, каждый из которых состоит из n испытаний. В каждом опыте фиксируется число появлений рассматриваемого события. Обозначим число появлений события в i -м опыте через m_i и подставим в формулу Пуассона:

$$P_n(m_i) = \frac{\lambda^{m_i}}{m_i!} e^{-\lambda}, \quad i=1, 2, \dots, k.$$

Составим теперь функцию наибольшего правдоподобия:

$$\begin{aligned} L &= P_n(m_1, \lambda) P_n(m_2, \lambda) \dots P_n(m_k, \lambda) = \\ &= \frac{\lambda^{m_1}}{m_1!} e^{-\lambda} \frac{\lambda^{m_2}}{m_2!} e^{-\lambda} \dots \frac{\lambda^{m_k}}{m_k!} e^{-\lambda} = \frac{\lambda^{(m_1+m_2+\dots+m_k)}}{m_1! m_2! \dots m_k!} e^{-k\lambda} = \frac{\lambda^{\sum_{i=1}^k m_i}}{m_1! m_2! \dots m_k!} e^{-k\lambda}. \end{aligned}$$

Запишем логарифмическую функцию наибольшего правдоподобия:

$$\ln L = (\sum_{i=1}^k m_i) \ln \lambda - \ln(m_1! m_2! \dots m_k!) - k\lambda.$$

Исследуем функцию на экстремум.

- 1) Возьмем производную по λ и приравняем ее к нулю:

$$\frac{d \ln L}{d \lambda} = \frac{1}{\lambda} \sum_{i=1}^k m_i - k = 0.$$

Получим $\lambda^* = \frac{1}{k} \sum_{i=1}^k m_i = \bar{m}_B$, то есть λ^* равна среднему выборочному значению числа появлений события в k опытах.

- 2) Возьмем производную второго порядка:

$$\frac{d^2 \ln L}{d \lambda^2} = -\frac{1}{\lambda^2} \sum_{i=1}^k m_i.$$

Очевидно, что она отрицательна, следовательно, $\lambda^* = \bar{m}_B$ - точка максимума. Таким образом, в качестве оценки наибольшего правдоподобия параметра λ распределения Пуассона принимается выборочная средняя числа появлений события в k опытах.

13. Число степеней свободы

Статистические оценки параметров предполагаемого теоретического распределения вычисляются по случайным выборкам и поэтому являются случайными числами. Они подвергаются дальнейшему анализу с помощью различных критериев согласия: например, хи-квадрат (χ^2), Фишера-Снедекора, Стьюдента и другие.

Для применения этих стандартных законов распределения вводится понятие числа степеней свободы изучаемого статистического материала.

Различные независимые измерения X_1, X_2, \dots, X_k одной и той же величины можно рассматривать как различные случайные величины, имеющие одно и то же распределение и имеющие k степеней свободы. Сумма $X_1 + X_2 + \dots + X_k$ и сумма квадратов также будет иметь число степеней свободы, равное k .

Если для рассмотренных случайных величин используется какая-либо связь, то число степеней свободы уменьшается. Например, если по результатам выборки вычислена средняя величина:

$$\bar{x}_B = \frac{x_1 + x_2 + \dots + x_k}{k},$$

и зафиксирована для всех X_1, X_2, \dots, X_k , то есть $\bar{x}_B = \frac{X_1 + X_2 + \dots + X_k}{k}$, то тогда одну из величин можно выразить через остальные. Значит, система оказалась связанной и потеряла одну степень свободы.

Если по результатам выборки вычислена выборочная дисперсия и присвоена любой из случайных величин X_i ($i=1, 2, \dots, k$), то число степеней свободы станет равно $k-2$.

В некоторых статистических исследованиях фиксируются не только общая средняя выборочная и дисперсия выборочная, но и групповые средние и групповые дисперсии. Тогда число степеней свободы у системы случайных величин уменьшается еще и на число таких связей.

14. Основные стандартные законы распределения, их таблицы и применение этих таблиц

Для практического применения статистических расчетов служат таблицы основных стандартных законов распределения: нормального, χ^2 , Стьюдента и Фишера.

Введем понятие квантиля и критической точки. Пусть X -СВ, имеющая стандартное распределение. P -Квантилем распределения X называется такое значение x_p , что $F(x_p)=p$, где $F(x)$ -функция распределения СВ X . Часто вместо квантиля используется критическая точка. Критической точкой уровня α распределения X называется такое значение $x_{kp(\alpha)}$, что $P(X > x_{kp(\alpha)}) = \alpha$. Ясно, что квантиль уровня p равен критической точке уровня $1-p$. Если распределение X симметрично относительно оси ординат, то $x_\alpha = -x_{1-\alpha}$ (см. рис.1)

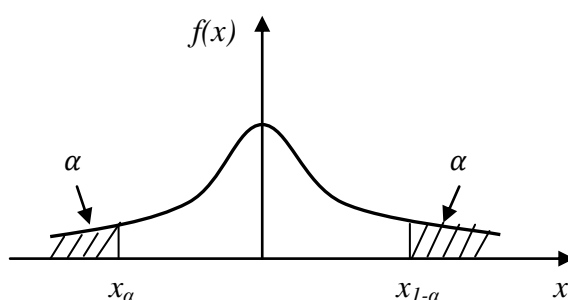


Рис. 1

1. Работа с таблицами стандартного нормального распределения

Стандартным нормальным распределением называется нормальное распределение $N[0,1]$ с параметрами $m=0, \sigma=1$, p -Квантиль нормального распределения обозначается u_p . Обычно в приложениях для отыскания u_p используются таблицы значений функции Лапласа:

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_0^u e^{-\frac{t^2}{2}} dt.$$

Функция Лапласа связана с функцией распределения стандартной нормальной СВ равенством: $F(u) = \Phi(u) + 0,5$ (см. рис.2)

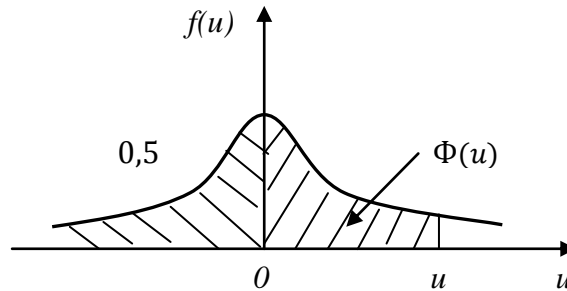


Рис. 2

Если $p > 0,5$, то, используя таблицы функции Лапласа, находим квантиль u_p из равенства $\Phi(u_p) = p - 0,5$.

Например, найдем $u_{0,95}$. Из таблиц имеем $\Phi(1,64) = 0,4495$, $\Phi(1,65) = 0,4505$. Полагаем $\Phi(1,645) = 0,45 = 0,95 - 0,5$, т. е. $u_{0,95} = 1,645$. Если же $p < 0,5$, то $1 - p > 0,5$. Находим u_{1-p} . Тогда $u_p = -u_{1-p}$.

Пример. Найти $u_{0,05}$. Тогда $1 - 0,05 = 0,95$; $u_{0,95} = 1,645$. Значит, $u_{0,05} = -1,645$.

Замечание. Если в приложениях даются таблицы функции $F(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{t^2}{2}} dt$, то тогда сразу по заданному уровню p находим u_p .

2. Распределение χ^2 (хи квадрат)

Пусть U_1, U_2, \dots, U_n – взаимно независимые случайные величины, имеющие стандартные нормальные распределения. Тогда $\chi^2 = \sum_{i=1}^n U_i^2$ имеет $\chi^2(n)$ распределение с n степенями свободы. Отметим, что число степеней свободы (v) определяется числом случайных величин, ее составляющих, уменьшенному на число линейных связей между ними. Случайная величина, имеющая χ^2 распределение, может принимать только неотрицательные значения, ее график плотности имеет асимметрический вид с вытянутым правым «хвостом». Однако, с увеличением числа степеней свободы распределение $\chi^2(n)$ постепенно приближается к нормальному.

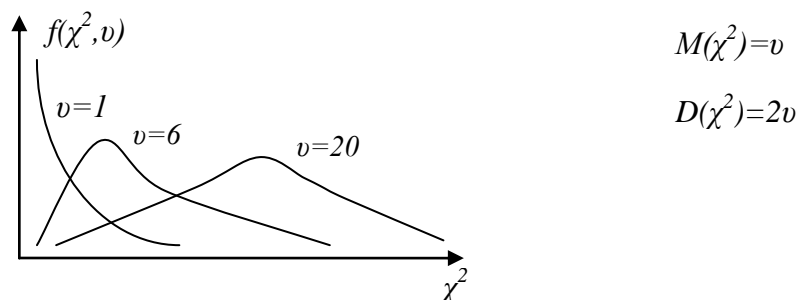


Рис. 3

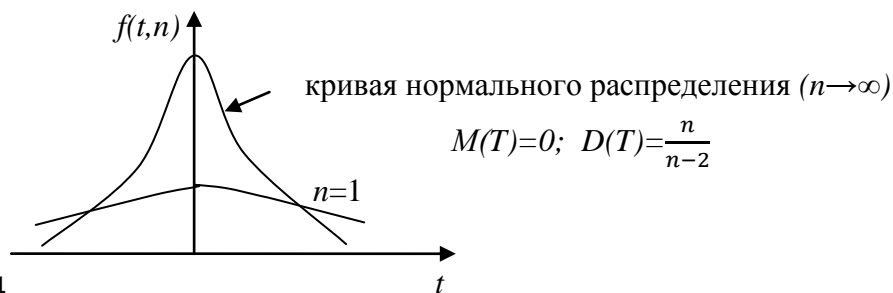
Распределение χ^2 применяется для нахождения интервальных оценок и проверки статистических гипотез. В приложениях обычно для данного уровня p и числа степеней свободы v даются значения критических точек $\chi^2_{кр}(p, v)$.

Например, $\chi^2_{кр}(0,995,4)=0,207$, т. е. $P(\chi^2(4)>0,207)=0,995$.

3. Распределение Стьюдента

Пусть СВ U имеет стандартное нормальное распределение, а СВ V не зависит от U и распределена по закону χ^2 с n степенями свободы.

Тогда СВ $T=\frac{U}{\sqrt{V/n}}$ имеет распределение Стьюдента (t распределение) с n степенями свободы $T=St(n)$. Распределение Стьюдента имеет один параметр, график функции плотности симметричен относительно оси ординат.



С увеличением числа степеней свободы распределения Стьюдента приближается к нормальному и при $n>30$ его практически можно заменить нормальным. Так же, как и распределение χ^2 его применяют для построения интервальных оценок и проверки статистических гипотез.

При этом используют либо таблицы квантилей, либо критических точек. Поскольку распределение Стьюдента симметрично, то $t_{1-p}=-t_p$. Значения квантиля находятся на пересечении строки с числом степеней свободы и столбца-вероятности. Например, $t_{0,95}(5)=2,015$, т. е. $P(St(5)<2,015)=0,95$. Это же значение $t=2,015$ является критической точкой для уровня $0,05$ $t_{кр}(0,05;5)=2,015$, т. е. $P(St(5)>2,015)=0,05$.

4. Распределение Фишера-Снедекора

Пусть $V=\chi^2(m)$, $W=\chi^2(n)$ независимые СВ, распределенные по закону χ^2 со степенями свободы $v_1=m$ и $v_2=n$. Тогда СВ $F=\frac{V/m}{W/n}$ имеет распределение Фишера со степенями свободы $v_1=m$ и $v_2=n$, $F=F(m, n)$. Таким образом, распределение Фишера (F -распределение) определяется двумя параметрами, причем $M(F)=\frac{n}{n-2}$ ($n>2$); $D(F)=\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$ $n>4$.

СВ, распределенное по закону Фишера, принимает только неотрицательные значения, применяется для проверки статистических гипотез в дисперсионном и регрессионном анализе. Таблицы квантилей (критических точек) должны иметь три входа, поэтому в приложениях приводятся таблицы для некоторых наиболее употребительных значений уровней значимости: $\alpha=0,01$; $\alpha=0,05$; $\alpha=0,1$. Для каждого из этих значений приводятся значения критических точек (или квантилей) для распределения Фишера с числом степеней свободы v_1 и v_2 . Например, для $\alpha=0,05$ $F_{кр}(0,05; (4,2))=19,2$. Это значит,

что $P(F(4,2) > 19,2) = 0,05$. Если же $\alpha = 0,1$, то $F_{кр}(0,1; (4,2)) = 9,24$, т. е. $P(F(4,2) > 9,24) = 0,1$. В то же время эти критические точки являются квантилями уровней 0,95 и 0,9 соответственно. $F_{0,95}(4,2) = 19,2$ и $F_{0,9}(4,2) = 9,24$, т. е. $P(F(4,2) < 19,2) = 0,95$ и $P(F(4,2) < 9,24) = 0,9$.

График плотности распределения Фишера имеет вид:

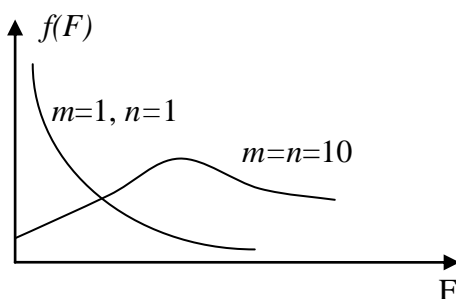


Рис. 5

15. Интервальные оценки

Точечная оценка θ^* параметра θ является функцией выборки и зависит в том числе от ее объема. Поскольку обычно оценки берутся несмещенные и состоятельные, то если объем выборки достаточно большой, оценка дает хорошее приближение оцениваемому параметру. Однако нередко, особенно, если невозможно в силу условий построить выборку большого объема, возникает необходимость задать некоторые границы, в которых заключается истинное значение параметра с некоторой вероятностью (надежностью).

Это приводит нас к задаче построения интервальных оценок или доверительных интервалов. Дадим точное определение.

Пусть θ -неизвестный параметр и γ -вероятность, тогда интервал $[\theta_1(n); \theta_2(n)]$, здесь n – объем выборки, покрывающий параметр θ с вероятностью γ , называется доверительным интервалом, или интервальной оценкой параметра θ , а γ -надежностью или доверительной вероятностью. Мы говорим «интервал покрывает параметр θ », поскольку θ -неслучайная величина, а случайными являются концы интервала. Очевидно, интервал зависит от оценки θ^* параметра. Иногда, но не всегда, он имеет вид $\theta^* \pm \Delta$. В этом случае Δ -предельная абсолютная погрешность, а величина Δ или 2Δ называется точностью оценки. Ясно, что при одном и том же объеме выборки, если мы хотим увеличить точность оценки (т. е. уменьшить длину интервала), то это приведет к уменьшению ее надежности и наоборот. Погрешность Δ (ошибка Δ) возникает вследствие того, что вместо всей генеральной совокупности исследуется только ее часть – случайная выборка. Она называется ошибкой репрезентативности.

Алгоритм построения доверительного интервала:

1. Задается уровень значимости $\alpha = 1 - \gamma$, где γ -надежность оценки. Обычно для α используется значение 0,1; 0,05; 0,01; 0,001.
2. Пусть θ -неизвестный параметр распределения СВ X . Обычно предполагается, что X имеет нормальное распределение. В силу Ц. П. Т. для больших объемов n выборки эти же формулы используются и для других распределений.
3. Строится СВ Z , являющаяся функцией выборки, зависящая от θ и ее точечной оценки θ^* и имеющая стандартное распределение. Она называется статистикой.

4. С помощью таблиц, имеющих для этих стандартных распределений, находятся такие z_1 и z_2 , что $P(z_1 < Z < z_2) = \gamma = 1 - \alpha$.
5. Используя связь между Z и θ , строится равносильное неравенство $\theta_1 < \theta < \theta_2$, так что $P(\theta_1 < \theta < \theta_2) = \gamma$. Если распределение Z симметрично, то интервал имеет вид $\theta^* \pm \Delta$, где Δ -предельная абсолютная погрешность.

Схема построения доверительных интервалов

Оцениваемый параметр	Статистика и ее распределение	Доверительный интервал
1. Математическое ожидание нормальной СВ $X=N[m, \sigma]$ $MX=m$, σ известно	$U = \frac{(\bar{x}-m)\sqrt{n}}{\sigma} = N[0,1]$	$\Delta = u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$ $P(\bar{x} - \Delta < m < \bar{x} + \Delta) = 1 - \alpha$
2. Математическое ожидание нормальной СВ $X=N[m, \sigma]$ $MX=m$, σ неизвестно	$T = \frac{(\bar{x}-m)\sqrt{n}}{s} = St[n-1]$ s -исправленное среднее квадратическое отклонение	$\Delta = t_{1-\alpha/2}(n-1) \frac{s}{\sqrt{n}}$ $P(\bar{x} - \Delta < m < \bar{x} + \Delta) = 1 - \alpha$
3. Дисперсия нормальной СВ $X=N[m, \sigma]$ $DX=\sigma^2$, m неизвестно	$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \chi^2[n-1]$	$P\left(\frac{s^2(n-1)}{\chi^2_{1-\alpha/2}(n-1)} < \sigma^2 < \frac{s^2(n-1)}{\chi^2_{\alpha/2}(n-1)}\right) = 1 - \alpha$ $P\left(s \sqrt{\frac{(n-1)}{\chi^2_{1-\alpha/2}(n-1)}} < \sigma < s \sqrt{\frac{(n-1)}{\chi^2_{\alpha/2}(n-1)}}\right) = 1 - \alpha$
4. Вероятность p биномиального распределения (доля), n велико	$U = \frac{(P-W)}{\sqrt{\frac{W(1-W)}{n}}} = N[0,1]$	$\Delta = u_{1-\alpha/2} \sqrt{\frac{W(1-W)}{n}}$ $P(w - \Delta < p < w + \Delta) = 1 - \alpha$

Здесь u_p , $t_p(k)$, $\chi^2_p(k)$ - квантили соответственно стандартного нормального распределения, распределения Стьюдента и χ^2 с k степенями свободы.

$W = \frac{m}{n}$ - относительная частота (доля).

Примеры решения задач

Задача №1

В результате предварительного опроса больших групп населения по 50 избирательным округам установлено, что в среднем 35,1% готовы проголосовать за кандидата данной партии.

а) Считая, что СВ X -процент проголосовавших в каждом округе за данного кандидата имеет нормальное распределение со средним квадратическим отклонением $\sigma=1,5\%$, указать с 90% уверенностью, в каких пределах будет заключаться процент проголосовавших на предстоящих выборах.

б) Решить ту же задачу, заменяя заданное σ стандартным отклонением, полученным в данной выборке объема 50 и составляющим 2,1%.

с) На уровне значимости $\alpha=0,1$ проверить, не является ли стандартное отклонение в 1,5% заниженной оценкой?

Решение

а) $\gamma=0,9$, $1-\alpha/2=0,95$; $u_{0,95}=1,645$, $\Delta=\frac{1,645 \cdot 1,5}{\sqrt{50}}=0,35$. Интервал $35,1 \pm 0,35$. Итак, в 90% случаев можно быть уверенным, что процент проголосовавших за данного кандидата будет заключен в интервале (34,75%;35,45%).

б) Поскольку σ неизвестно, пользуемся распределением Стьюдента $t_{0,95}(49)=1,675$, $\Delta=\frac{1,675 \cdot 2,1}{\sqrt{50}}=0,5$. Интервал (34,6%;35,6%) оказался шире.

с) Построим доверительный интервал для σ .

$$\chi_{0,95}^2(49) = 67,5, \chi_{0,05}^2(49) = 34,8,$$

$$s \sqrt{\frac{49}{67,5}} = 2,1 \cdot 0,85 = 1,79, s \sqrt{\frac{49}{34,8}} = 2,1 \cdot 1,9 = 2,49,$$

$$P(1,79 < \sigma < 2,49) = 0,9.$$

Итак, на уровне значимости $\alpha=0,1$ можно утверждать, что оценка стандартного отклонения в 1,5% занижена, т. к. 1,5 не принадлежит интервалу (1,79;2,49).

Задача №2

Из большой партии деталей для проверки было отобрано 100 деталей, среди которых оказалось 96 стандартных. Найти: а) вероятность того, что доля нестандартных деталей во всей партии отличается от полученной доли в выборке по абсолютной величине не более, чем на 0,01; б) границы, в которых с надежностью 0,95 заключена доля нестандартных деталей во всей партии; с) определить число деталей, которое надо отобрать в выборку, чтобы с вероятностью 0,9 доля нестандартных деталей в выборке отличалась от генеральной доли (по абсолютной величине) не более, чем на 0,02.

Решение

Поскольку n велико, статистика $\frac{(P-W)\sqrt{n}}{\sqrt{W(1-W)}}$ имеет нормальное стандартное распределение.

$$а) W = \frac{100-96}{100} = 0,04; \sigma_w = \sqrt{\frac{W(1-W)}{n}} = \sqrt{\frac{0,04 \cdot 0,96}{100}} = \frac{0,196}{10} = 0,0196;$$

$$P(|W-p| < 0,01) = 2\Phi\left(\frac{0,01}{0,0196}\right) = 2\Phi(0,51) = 2 \cdot 0,0196 = 0,39.$$

$$б) \Delta = u_{1-\alpha/2} \cdot \sqrt{\frac{W(1-W)}{n}}; \alpha/2=0,025; 1-\alpha/2=0,975; U_{0,975}=1,96$$

$$\Delta = 1,96 \cdot 0,0196 = 0,038; p \in W \pm \Delta = 0,04 \pm 0,038;$$

$$P(0,002 < p < 0,078) = 0,95;$$

$$с) P(|W-p| < 0,02) = 2\Phi\left(\frac{0,02}{\sigma'_w}\right) = 0,9; \Phi(1,645) = 0,45;$$

$$0,012 = \frac{0,02}{1,645} = \sigma'_w, \text{ т. к. } \frac{0,02}{\sigma'_w} = 1,645; \text{ откуда } \sqrt{\frac{0,04 \cdot 0,96}{n}} = 0,012$$

$$\text{Значит, } n > \frac{0,04 \cdot 0,96}{0,012^2} = 1361.$$

Замечание. Для решения задачи с) мы предположили, что доля нестандартных деталей не изменилась. Иногда ее полагают равной средней доли данного производства. Если о доли ничего, даже приблизительно, неизвестно, то полагают:

$W(1-W) = W(1-W)_{max} = 0,25$, что увеличивает объем выборки.

Вывод: с надежностью 0,39 можно полагать, что доля нестандартных деталей всей партии лежит в границах (3%,5%) и с надежностью 0,95 лежит в границах (0,2%,7,8%).

Объем выборки должен быть не менее, чем 1361, чтобы с вероятностью 0,9 гарантировать отклонение доли нестандартных деталей от генеральной доли не более, чем на 0,02.

16. Проверка статистических гипотез

Статистической называют гипотезу о виде закона распределения или о параметре известного распределения. В первом случае гипотеза называется непараметрической, а во втором – параметрической.

Рассмотрим сначала случай параметрических гипотез. Пусть θ -неизвестный параметр. Выдвигается гипотеза $H_0: \theta = \theta_0$. Эта гипотеза называется нулевой или основной. В качестве конкурирующих или альтернативных гипотез рассматриваются гипотезы $\theta \neq \theta_0$, $\theta > \theta_0$, $\theta < \theta_0$. С логической точки зрения альтернативной является только $\theta \neq \theta_0$. Но часто, исходя из смысла параметра θ , оказывается, что одна из них, например, $\theta > \theta_0$ не имеет смысла, и тогда $\theta < \theta_0$ является альтернативной гипотезой. Заметим, что обычно в качестве нулевой гипотезы берется та, которую собираются отвергнуть. Причина в том, что для доказательства некоторого утверждения требуется рассмотреть все случаи, а для его отрицания достаточно одного опровергающего примера.

Алгоритм проверки параметрической гипотезы

1. Выбирается уровень значимости α ;
2. Выдвигаются основная H_0 гипотеза и альтернативная H_1 ;
3. Строится случайная величина Z , которая при условии, что гипотеза H_0 справедлива, имеет стандартное распределение;
4. Используя стандартное распределение и уровень значимости α , область изменения СВ Z разбивается на две области: критическую, вероятность попадания в которую равна α , и область принятия гипотезы, вероятность попадания в которую СВ Z равна $\gamma = 1 - \alpha$;
5. Вычисляется выборочное значение критерия Z . Если оно попадает в критическую область, то гипотеза H_0 отвергается. В противном случае говорят, что гипотеза не противоречит эксперименту и может быть принята.

Следующий пример позволяет лучше понять задачу проверки гипотезы.

Пусть известно, что во время эпидемии гриппа заболевает в среднем 30%. Была разработана вакцина от гриппа. Разработчики утверждают, что в результате прививки данной вакцины человек либо не заболевает, либо вероятность заболеть резко уменьшается. Для подтверждения этого было отобрано 10 групп добровольцев, подвергшихся вакцинации. В результате в пяти из них никто не заболел, в двух группах процент заболевших был равен двум и в двух группах – 4% заболевших, но в одной группе оказалось 27% заболевших. Можно ли утверждать, что вакцина действует, и какой на самом деле средний процент заболевших?

Для решения этой задачи была выдвинута гипотеза $H_0: MX=30$, где X -процент заболевших. В качестве альтернативной гипотезы принята гипотеза $H_1: MX < 30$. В результате проверки гипотеза H_0 , означающая, что вакцина не работает, была отвергнута. Значит процент заболевших уменьшился, но как? Последовательно проверялись гипотезы: $MX=20$, $MX=10$, $MX=5$. В результате проверки была принята гипотеза $MX < 5$, что означает, что после вакцинации на уровне значимости 0,05 можно утверждать, что процент вакцинированных заболевших не превышает 5.

Статистическая проверка гипотез основывается на выбранных данных, а потому не исключена возможность принятия неправильного решения. При этом возможны ошибки двух родов: ошибка первого рода состоит в том, что будет отвергнута правильная нулевая гипотеза, ошибка второго рода заключается в том, что будет принята ошибочная нулевая гипотеза.

Результаты статистических выводов представлены следующей таблицей:

Результаты проверки гипотезы	Возможные состояния гипотезы	
	Верна H_0	Верна H_1
Гипотеза H_0 отклоняется, принимается H_1	Ошибка первого рода $P(H_1/H_0)=\alpha$	Правильный вывод $P(H_1/H_1)=1-\beta$
Гипотеза H_1 отклоняется, принимается H_0	Правильный вывод $P(H_0/H_0)=1-\alpha$	Ошибка второго рода $P(H_0/H_1)=\beta$

$P(H_1/H_1)=1-\beta$ называется мощностью критерия, т. е. вероятностью отвергнуть неверное решение.

Вероятность α совершить ошибку первого рода называют также уровнем значимости.

Невозможно исключить ошибки первого и второго рода, поскольку проверка осуществляется на основании выборки, но можно попытаться уменьшить их вероятности. Однако одновременное уменьшение α и β достигается только увеличением объема выборки.

Вероятность α ошибки первого рода называют также риском продавца, а β – вероятность ошибки второго рода – риском покупателя.

В самом деле, пусть продавец продает оптом большую партию ящиков с апельсинами. Так как невозможно проверить все ящики, покупатель выбирает несколько из них. Если апельсины в них оказались второго сорта (H_1), хотя фактически подавляющее большинство апельсинов партии первого сорта (H_0), то вся партия считается вторым сортом и проигрывает продавец. Его риск равен α . Если, напротив, в выбранных ящиках апельсины оказались первого сорта (H_0), хотя в действительности почти все они второго сорта (H_1), то покупатель платит за них по цене первого сорта и проигрывает. Его риск равен β .

Итак, пусть нулевая гипотеза $H_0: \theta=\theta_0$, α – уровень значимости, Z – статистика, K_p – квантиль этой статистики уровня p , V – область принятия гипотезы H_0 . следующая таблица устанавливает связь между конкурирующей гипотезой H_1 и областью V .

Конкурирующая гипотеза	Область V принятия гипотезы
$H_1: \theta \neq \theta_0$	$K_{\alpha/2} < Z_{наб} < K_{1-\alpha/2}$
$H_1: \theta > \theta_0$	$Z_{наб} < K_{1-\alpha}$
$H_1: \theta < \theta_0$	$Z_{наб} > K_{\alpha}$

Примеры проверки гипотез

1. Проверка гипотезы о значении математического ожидания нормальной СВ при известной и неизвестной дисперсии.

$$\sigma^2 \text{ известно } H_0: MX=m_0, \quad Z=\frac{(\bar{x}-m_0)\sqrt{n}}{\sigma} = N[0,1], \quad K_p=U_p$$

$$\sigma^2 \text{ неизвестно } H_0: MX=m_0, \quad Z=\frac{(\bar{x}-m_0)\sqrt{n}}{s} = St[n-1], \quad K_p=t_p(n-1)$$

s - исправленное среднее квадратическое отклонение

$$s = \sqrt{\frac{n}{n-1} \widehat{D}(x)} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

Замечание 1

u_p – квантиль нормального распределения, $t_p(n-1)$ – квантиль распределения Стьюдента с $(n-1)$ степенью свободы. Поскольку стандартное нормальное распределение и распределение Стьюдента симметричны относительно $x=0$, то $u_p = -u_{1-p}$ и тоже самое $t_p = -t_{1-p}$.

Замечание 2

Если альтернативная гипотеза $\theta \neq \theta_0$, то проверка гипотезы может быть проведена следующим образом: строится доверительный интервал параметра θ с надежностью $\gamma = 1 - \alpha$, а затем, если значение θ_0 попадает в данный интервал, то гипотеза H_0 не отвергается, если нет – то отвергается.

2. Проверка гипотезы о величине дисперсии нормальной СВ.

$$H_0: \sigma^2 = \sigma_0^2$$

$$Z = \frac{(n-1)s^2}{\sigma_0^2} = \chi^2[n-1]; \quad \chi_p^2[n-1] = K_p.$$

Здесь снова s^2 -исправленная дисперсия $\chi_p^2[n-1]$ – квантиль распределения χ^2 с $(n-1)$ степенью свободы.

3. Проверка гипотезы о значении вероятности в независимых испытаниях Бернулли.

$$H_0: p = p_0$$

$$Z = \frac{(W-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} = N[0,1] \quad K_p = u_p$$

Здесь $W = \frac{m}{n}$ -относительная частота. Из теоремы Бернулли следует, что при достаточно больших значениях n относительная частота W сходится по вероятности к вероятности события. При этом $M(W) = p$, $\sigma(W) = \sqrt{\frac{pq}{n}}$. Величина $U = \frac{(W-p)\sqrt{n}}{\sqrt{pq}}$ имеет приближенно стандартное нормальное распределение.

4. Проверка гипотезы о равенстве математических ожиданий.

Пусть генеральные совокупности X и Y распределены нормально и их дисперсии известны: σ_x^2 и σ_y^2 . По независимым выборкам, объемы которых равны n и m соответственно найдены выборочные средние \bar{X} и \bar{Y} .

$$H_0: MX = MY$$

а) σ_x^2 и σ_y^2 известны $Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} = N[0,1] \quad K_p = u_p$

в) $H_0: MX = MY$

σ_x^2 и σ_y^2 неизвестны,

X и Y имеют нормальное распределение или

σ_x^2 и σ_y^2 неизвестны, n и $m > 30$ и выборки независимы

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}} = N[0,1] \quad K_p = u_p$$

с) $H_0: MX = MY$

σ_x^2 и σ_y^2 неизвестны,

X и Y нормально распределены

m, n малы, $\sigma_x^2 = \sigma_y^2$

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{(n-1)s_x^2 + (m-1)s_y^2}} \sqrt{\frac{nm(n+m-2)}{m+n}} = St(n+m-2)$$
$$K_p = t_p(n+m-2)$$

5. Проверка гипотезы о равенстве двух дисперсий нормальных распределений.

Пусть генеральные совокупности X и Y распределены нормально. По независимым выборкам с объемами соответственно n и m найдены исправленные выборочные дисперсии s_x^2 и s_y^2 и пусть для определенности $s_x^2 > s_y^2$.

$H_0: DX=DY \quad Z = s_x^2/s_y^2 = F(n-1, m-1); \quad K_p = F_p(n-1, m-1)$

Замечание. Если наоборот $s_x^2 < s_y^2$, то $Z = s_y^2/s_x^2 = F(m-1, n-1)$.

6. Проверка равенства двух вероятностей биномиальных распределений (равенстве долей).

Пусть в двух генеральных совокупностях проведены независимые испытания: n_1 – в первой и n_2 – во второй. При этом частота появления события A в первой равна $W_1 = \frac{m_1}{n_1}$, а во второй – $W_2 = \frac{m_2}{n_2}$. Обозначим неизвестные вероятности появления события A в первой совокупности через p_1 , а во второй – p_2 .

$H_0: p_1 = p_2 = p_0$; $Z = \frac{W_1 - W_2}{\sqrt{\frac{m_1+m_2}{n_1+n_2} (1 - \frac{m_1+m_2}{n_1+n_2}) (\frac{1}{n_1} + \frac{1}{n_2})}} = N[0,1]$; $K_p = u_p$

n_1, n_2 велики

Замечание. В качестве наилучшего значения берут $p_0 = \frac{m_1+m_2}{n_1+n_2}$, если две выборки смешать в одну.

Примеры решения задач

Задача №3.

Анализируется доход X фирм в некоторой отрасли. Предполагается, что СВ X имеет нормальное распределение и при этом средний доход в отрасли не менее 1 млн. а среднее квадратическое отклонение не более 0,1 млн (в \$). По выборке из 50 фирм получены следующие данные $\bar{X}=0,9$ млн, $s=0,15$ млн. На уровне значимости $\alpha=0,01$ проверить эти предположения.

Проверим, что средний доход не менее 1 млн при известном и неизвестном σ .

$H_0: MX = m_0 = 1$. Поскольку $\bar{X} = 0,9 < 1$, то в качестве гипотезы $H_1: MX < 1$

σ известно $Z_{набл} = \frac{(0,9-1)\sqrt{50}}{0,1} = -7,07$;

Область левосторонняя. $u_\alpha = u_{0,01} = -u_{0,99}$.

$\Phi(2,3) = 0,49 = 0,99 - 0,5$, т. е. $u_{0,01} = -2,3$.

Поскольку $Z_{набл} = -7,07 < -2,3$, то гипотеза H_0 отвергается, принимается H_1 , т. е. средний доход в отрасли меньше 1 млн \$.

$$\sigma \text{ неизвестно } Z_{набл} = \frac{(\bar{X} - m_0)\sqrt{n}}{s} = \frac{(0,9-1)\sqrt{50}}{0,15} = -4,67;$$

$V = t_{кр}(0,01;49) = -2,4$. Снова $Z_{набл} = -4,67 < t_{кр} = -2,4$ и делается тот же вывод.

Проверим теперь, что среднее квадратическое не более, чем 0,1; $\sigma_{выб} = 0,15$, поэтому

$$H_0: \sigma^2 = \sigma_0^2 = 0,1^2$$

$$H_1: \sigma^2 > \sigma_0^2 = 0,1^2 \quad Z_{набл} = \frac{(n-1)s^2}{\sigma_0^2} = \frac{49 \cdot 0,15^2}{0,1^2} = 110,25$$

Область правосторонняя: $\chi_{кр}^2(0,01; 49) = 60 = \chi_{0,99}^2(49) < Z_{набл} = 110,25$, то принимается гипотеза H_1 , т. е. $\sigma > 0,1$

Вывод. Средний доход фирм в этой отрасли меньше 1 млн\$, а среднее квадратическое отклонение $\sigma > 0,1$.

Задача №4.

В рекламе фирмы A утверждается, что месячный доход по ее акциям в среднем равен 0,75% и превышает доход по акциям фирмы B более, чем на 0,3%. И при этом ее риски меньше. В течение годового периода средний месячный доход по акциям B составил 0,4%, а по акциям A – 0,65%, а их средние квадратические отклонения 1,9% для B и 2% для A . Полагая распределение доходности по каждой акции нормальными, на уровне значимости $\alpha = 0,05$ проверить все три утверждения, содержащиеся в рекламе.

Пусть X и Y месячный доход по акциям фирм A и B соответственно и пусть $Y_I = Y + 0,3$. Тогда $\bar{Y}_I = \bar{Y} + 0,3$;

$$s_{y1}^2 = s_y^2.$$

$$a) H_0: MX = 0,75\% = m_0 \quad Z_{набл} = \frac{(0,65 - 0,75)\sqrt{12}}{2} = -0,17; s = 2\%$$

$$H_1: MX < 0,75\%,$$

$t_{0,05}(11) = -1,796$; $-0,17 > -1,796$, принимается гипотеза H_0 , т. е. $MX = 0,75\%$.

$$в) H_0: MX = MY + 0,3\%; \quad Z_{набл} = \frac{0,65 - 0,4 - 0,3}{\sqrt{\frac{1,9^2}{2} + \frac{2^2}{12}}} = \frac{-0,05 \cdot \sqrt{12}}{\sqrt{7,61}} = -0,06$$

$$H_1: MX > MY + 0,3\%; \quad Z_{набл} = -0,06 < u_{0,95} = 1,645$$

$$S_x = 2\%, S_y = 1,9\%,$$

$$u_{1-\alpha} = u_{0,95} = 1,645$$

Принимается гипотеза H_0 : $MX = MY + 0,3\%$

$$c) H_0: \sigma_x^2 = \sigma_y^2 \quad Z_{набл} = \frac{2^2}{1,9^2} = 1,1$$

$$H_1: \sigma_x^2 > \sigma_y^2$$

$$F_{кр}(11,11) = F_{0,05}(11,11) = 2,82 > Z_{набл} = 1,1$$

Принимается гипотеза H_0 .

Вывод: Действительно, средний доход по акциям в фирме A равен 0,75% и он на 0,3% больше, чем в фирме B , но риски в обеих фирмах примерно одинаковые.

Задача №5

В университете из 102 студентов специальности A 62 успешно сдали все экзамены, а из 50 студентов специальности B сдали все экзамены 32 студента. На уровне значимости $\alpha=0,05$ проверить гипотезу об отсутствии существенных различий в успеваемости студентов этих двух специальностей.

Решение:

$$p_0 = \frac{32+62}{50+102} = 0,625; \quad W_1 = \frac{62}{102} = 0,608; \quad W_2 = \frac{32}{50} = 0,64$$

$$Z_{набл} = \frac{0,64 - 0,608}{\sqrt{0,625(1-0,625)\left(\frac{1}{102} + \frac{1}{50}\right)}} = \frac{0,032}{0,0836} = 0,38$$

$$H_0: p_1 = p_2 \quad \alpha = 0,05; \quad \alpha/2 = 0,025; \quad 1 - \alpha/2 = 0,975;$$

$$H_1: p_1 \neq p_2 \quad u_{0,025} = -1,96; \quad u_{0,975} = 1,96$$

$$-1,96 < Z_{набл} = 0,38 < 1,96.$$

Следовательно, гипотеза H_0 принимается, т. е. не существует значимого различия в успеваемости студентов специальностей A и B .

17. Непараметрические гипотезы. Критерий Пирсона

Пусть по виду гистограммы или на основании теоретического анализа выдвигается гипотеза о виде закона распределения СВ X . Для ответа на вопрос, объясняются ли расхождения между гипотетическим и эмпирическим распределениями случайными обстоятельствами или гипотеза неверна, служат критерии согласия. На практике чаще всего используется критерий χ^2 -Пирсона.

Схема применения критерия Пирсона. Пусть X – непрерывная СВ. Полученная выборка объема n сформирована в виде интервального вариационного ряда.

1. Выбирается уровень значимости α .
2. Число интервалов. Необходимо, чтобы в каждый интервал попадало не менее 5 наблюдений, т. е. $n_i \geq 5$. Если это условие не выполнено, то объединяют соседние интервалы. Пусть число интервалов равно k .
3. Параметры распределения. Число степеней свободы.

Параметры гипотетического распределения могут быть известными и неизвестными. Если они неизвестны, то их заменяют точечными оценками. Оценки чаще всего находят методом моментов, т. е. приравнивают теоретические и эмпирические моменты. Число степеней свободы $\nu=k-1$, если параметры известны и $\nu=k-s-1$, если приходится оценивать s параметров.

Заметим, что всегда должно быть выполнено соотношение: $\sum_{i=1}^k n'_i = n$

4. Теоретические частоты

На основании гипотетического распределения находятся вероятности $p_i = P(x \in \Delta_i)$ и теоретические частоты: $n'_i = np_i$

5. Выдвигается гипотеза H_0 : исследуемая СВ величина распределена по данному закону. Строится статистика:

$$Z_{наб} = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}$$

Эта статистика имеет χ^2 распределение с числом степеней свободы $\nu=k-s-1$ при достаточно большом n . Для данного α находим $\chi_{кр}^2(\alpha, k-s-1)$

6. Если $Z_{наб} > \chi_{кр}^2(\alpha, k-s-1)$, то гипотезу H_0 отвергают, в противном случае считают, что расхождение между теоретическим и эмпирическим распределениями несущественно, и гипотезу H_0 можно считать правдоподобной.

Комментарии

1. При проверке гипотезы о нормальном распределении СВ X , параметры m и σ могут быть как известными, так и неизвестными. В этом случае полагают $m = \bar{X}$, $\sigma = s$.

$$\text{Тогда } n'_i = np_i = n \left(\Phi \left(\frac{x_{i+1} - \bar{x}}{s} \right) - \Phi \left(\frac{x_i - \bar{x}}{s} \right) \right).$$

Так как $x \in (-\infty, +\infty)$, то $x_0 = -\infty$, а $x_k = +\infty$, поскольку должно быть выполнено равенство $\sum_{i=1}^k p_i = 1$, или иначе $\sum_1^k n'_i = n$. Возможна небольшая погрешность, вызванная ошибками округления

2. если проверяется гипотеза о равномерном распределении СВ $X=R[a, b]$ и концы интервала известны, то $n'_i = \frac{n}{k}$, где k – число интервалов одинаковой длины, или же $n'_i = n \frac{\Delta_i}{b-a}$, если пришлось объединять интервалы. В случае, когда концы интервала неизвестны, их заменяют оценками $a^* = \bar{X} - s\sqrt{3}$, $b^* = \bar{X} + s\sqrt{3}$.

$$\text{Тогда } n'_1 = \frac{x_1 - a^*}{b^* - a^*} \cdot n; \quad n'_k = \frac{b^* - x_{k-1}}{b^* - a^*} \cdot n.$$

$$\text{Для внутренних интервалов } n'_i = \frac{x_{i+1} - x_i}{b^* - a^*} \cdot n$$

3. Показательное распределение $X \sim E(\lambda)$ имеет один параметр λ . Если он неизвестен, его заменяют оценкой $\lambda^* = \frac{1}{\bar{x}}$. Так как $x \in [0, +\infty]$, то полагают $x_0 = 0$ и $x_k = +\infty$, чтобы

$$\sum_1^k n'_i \text{ была равна } n. \text{ Частоты } n'_i = n(e^{-\lambda^* x_i} - e^{-\lambda^* x_{i+1}})$$

Пример 1.

Для оценки месячного дохода на душу населения X (тыс.руб.) была произведена случайная выборка объема $n=100$ и сформирован интервальный вариационный ряд.

x_i	Менее 7,3	7,3-8,5	8,5-9,7	9,7-10,9	10,9-12,1	12,1-13,3	13,3-14,5	Свыше 14,5
n_i	10	10	22	18	20	6	8	6

На уровне значимости $\alpha=0,05$ проверить гипотезу о нормальном распределении СВ X .

Решение.

При наличии открытых интервалов значений X типа «менее x_0 » или «свыше x_k » для проведения расчетов их условно заменяют интервалами той же длины. Интервал «менее 7,3» заменяем интервалом $[6,1;7,3]$, а интервал «свыше 14,5» - интервалом $[14,5;15,1]$. Беря в качестве представителя интервала его середину, получаем $\bar{X}=10,396 \approx 10,4$; $D_B(X)=5,242$, $s=2,3$.

Вычисления оформляем в виде таблицы:

№ интервала	x_i	x_{i+1}	n_i	a_i	a_{i+1}	$\Phi(a_i)$	$\Phi(a_{i+1})$	n'_i	$\frac{(ni - n'_i)^2}{n'_i}$
1	$-\infty$	7,3	10	$-\infty$	-1,35	-0,5	-0,4115	8,85	0,15
2	7,3	8,5	10	-1,35	-0,83	-0,4115	-0,2967	11,48	0,19
3	8,5	9,7	22	-0,83	-0,3	-0,2967	-0,1179	17,88	0,95
4	9,7	10,9	18	-0,3	0,22	-0,1179	0,0871	20,5	0,3
5	10,9	12,1	20	0,22	0,74	0,0871	0,2703	18,32	0,15
6	12,1	13,3	6	0,74	1,26	0,2703	0,3962	12,59	3,45
7	13,3	14,5	8	1,26	1,78	0,3962	0,4625	6,63	0,28
8	14,5	$+\infty$	6	1,78	$+\infty$	0,4625	0,5	3,75	1,35
									6,82

Здесь $a_i = \frac{x_i - 10,4}{2,3}$; $\Phi(x)$ – функция Лапласа

$$\chi^2_{\text{наб}} = 6,82; V = 8 - 1 - 2 = 5; \chi^2_{\text{кр}}(0,05; 5) = 11,1 > \chi^2_{\text{наб}} = 6,82.$$

Следовательно, гипотеза о нормальном распределении месячного дохода на душу населения не противоречит эксперименту и может быть принята.

Рассмотрим случай дискретной случайной величины.

Пример 2.

Анализируется количество договоров, заключенных страховыми агентами компании в течение недели, для чего была произведена выборка объема $n=50$. Результат оформлен в виде дискретного вариационного ряда, где x_i – количество заключенных договоров, а n_i – число агентов, заключивших x_i договоров.

x_i	0	1	2	3	4	5	6
n_i	1	10	15	9	6	5	4

На уровне значимости $\alpha=0,05$ проверить гипотезу о том, что СВ X – количество заключенных договоров страховым агентом, распределено по закону Пуассона.

Поскольку n_i должно быть ≥ 5 , то объединяем два значения $x=0$ и $x=1$, а также $x=5$ и $x \geq 6$. Распределение Пуассона имеет один параметр λ . В качестве его оценки берем $\lambda^* = \bar{X}$.

$$\text{Вычислим } \bar{X} = \frac{0 \cdot 1 + 1 \cdot 10 + 2 \cdot 15 + 3 \cdot 9 + 4 \cdot 6 + 5 \cdot 5 + 6 \cdot 4}{50} = 2,8$$

$$\lambda^* = 2,8 = \lambda.$$

Для распределения Пуассона $P(m) = \frac{\lambda^m e^{-\lambda}}{m!}; m=0, 1, \dots, n, \dots$

Вычисления оформим в виде таблицы, полагая $e^{-2,8} = 0,06$

№ варианта	x_i	n_i	n'_i	$\frac{(n_i - n'_i)^2}{n'_i}$
1	0 1	11	11,4	0,014
2	2	15	11,76	0,893
3	3	9	10,98	0,357
4	4	6	7,68	0,368
5	5 ≥ 6	9	8,18	0,082

$$\Sigma = 1,714.$$

$$\chi^2_{\text{наб}} = 1,714$$

Пояснения

$$P(X=0) + P(X=1) = \frac{\lambda^0 e^{-\lambda}}{0!} + \frac{\lambda e^{-\lambda}}{1!} = 0,06(1+\lambda) = 0,06 \cdot 3,8 = 0,228;$$

$$n'_1 = 50 \cdot p_1 = 50 \cdot 0,228 = 11,4;$$

$$P(X=2) = \frac{2,8^2 \cdot 0,06}{2!} = 0,2352; n'_2 = 0,2352 \cdot 50 = 11,76$$

$$P(X=3) = \frac{2,8^3 \cdot 0,06}{3!} = 0,2196; n'_3 = 10,98$$

$$P(X=4) = \frac{2,8^4 \cdot 0,06}{4!} = 0,1536; n'_4 = 7,68$$

$$P(X \geq 5) = 1 - P(X < 5) = 1 - (0,228 + 0,2352 + 0,2196 + 0,1537) = 0,1635$$

$$n'_5 = 8,18$$

$$V=5-1-1=3; \chi_{кр}^2(0,05;3)=7,8 > \chi_{наб}^2=1,714.$$

Следовательно, на уровне значимости $\alpha=0,05$ можно принять гипотезу H_0 о пуассоновском распределении числа страховых договоров, заключенных страховым агентом.

Вопросы для самопроверки

1. Какая связь между генеральной совокупностью и выборкой?
2. Дайте определение дискретного и интервального вариационных рядов?
3. Сформулируйте свойства эмпирической функции распределения?
4. Что такое полигон частот и гистограмма? Для чего они используются?
5. Приведите формулы для вычисления выборочных среднего, дисперсии и среднего квадратического отклонения.
6. Приведите формулу выборочного коэффициента корреляции и укажите его свойства. Что он характеризует?
7. Дайте определение точечной оценки параметра.
8. В чем разница между не смещенной и асимптотически несмещенной оценкой? Приведите пример смещенной, но асимптотически несмещенной оценки.
9. Какие оценки называются эффективными и какие состоятельными?
10. Что такое доверительный интервал и каков алгоритм его построения?
11. Как меняется надежность оценки при изменении её точности?
12. Если среднее выборочное стороны квадрата равно a , то будет ли среднее выборочное его площади равно a^2 ?
13. Перечислите основные стандартные распределения.
14. Как связаны квантиль и критическая точка?
15. Может ли квантиль распределения Стьюдента быть равным $-1,75$?
16. Как связаны выборочные средние, дисперсии и средние квадратические отклонения случайных величин χ и $U=2X$?
17. Как строятся доверительные интервалы для дисперсии и среднего квадратического отклонения нормальной случайной величины?
18. Чем отличаются интервальные оценки для математического ожидания нормальной СВ при известной и неизвестной дисперсиях?
19. Какая разница между параметрической и непараметрической статистическими гипотезами?
20. Какая связь между нулевой, альтернативной и конкурирующей гипотезами?
21. Какая статистика используется при проверке гипотезы о равенстве дисперсий нормальных СВ?
22. Приведите общую схему проверки гипотез.
23. Что такое ошибки первого и второго рода? Как можно уменьшить вероятность этих ошибок?
24. Что такое уровень значимости?
25. Какая статистика используется для проверки гипотезы о равенстве математических ожиданий двух нормальных СВ, если известны их дисперсии?
26. Как связаны квантили распределения Стьюдента для уровней $0,1$ и $0,9$?
27. Как проверяется гипотеза о нормальном распределении генеральной совокупности, если параметры нормального распределения неизвестны?
28. Как связаны квантили одного уровня для нормального распределения и распределения Стьюдента с большим числом степеней свободы?
29. Как меняется точность оценки при построении доверительного интервала для математического ожидания нормальной СВ с известной дисперсией при увеличении числа наблюдений в 2 раза?
30. Как меняются выборочное среднее и выборочная дисперсия при изменении масштаба?

18. Варианты заданий

Задача №1.

Для исследования доходов сотрудников предприятия было отобрано по 15 человек в каждом подразделении. В результате получены 2 выборки, где X и Y доходы подразделений (в тыс. руб.). Объединив обе выборки в одну:

1. Составить интервальный вариационный ряд, разбив значения варианты на 6 интервалов.
2. Построить гистограмму, кумуляту и эмпирическую функцию распределения.
3. Найти средний доход работников предприятия, медиану Me , моду Mo , дисперсию s^2 и среднее квадратическое отклонение.
4. (вар. 1,2,3). Найти вероятность того, что средний доход работников предприятия отличается от среднего дохода по объединенной выборке не более, чем на 1 тыс. рублей (по абсолютной величине). Определить границы, в которых с надежностью 0,99, заключен средний доход работников данного предприятия. Каким должен быть объем выборки, чтобы те же границы гарантировать с надежностью 0,997?
4. (вар. 4,5,6) Найти вероятность того, что доля работников с низкими доходами (меньше 15 тыс.) данного предприятия отличается от доли таких же работников в объединенной выборке не более, чем на 0,01 (по абсолютной величине). Определить границы, в которых с надежностью 0,95 заключена доля работников с низкими доходами. Каким должен быть объем выборки, чтобы те же границы доли бедных гарантировать с надежностью 0,997?
5. (вар.1,2,3) На уровне значимости $\alpha = 0,05$ проверить гипотезу о том, что средний доход работников обоих подразделений и разброс относительно среднего дохода отличаются незначимо.
5. (вар.4,5,6) На уровне значимости $\alpha = 0,05$ проверить гипотезу о том, что доли бедных для двух подразделений отличаются незначимо и гипотезу о том, что доля работников с низким доходом на данном предприятии составляет 15%.
6. На основании вида гистограммы выдвинуть гипотезу о виде закона распределения доходов работников предприятия и проверить ее критерием Пирсона на уровне значимости $\alpha = 0,05$.

Вариант 1.

X :	8,5	19,2	23,1	15,1	10,8	19,5	15,2	28,4	12,6	20,4	18,5	32,1	16,4	25,4	30,7
Y :	18,5	32,1	16,7	35,6	30,4	27,8	52,4	40,8	18,5	13,6	9,4	12,1	15,7	20,1	18,5

Вариант 2.

X :	30,4	17,5	15,2	18,1	23,4	24,3	52,1	32,4	10,8	33,5	20,5	15,7	12,6	14,8	20,7
Y :	52,1	32,4	10,8	32,5	20,5	15,7	10,2	52,1	50,4	32,4	26,4	18,1	35,2	27,2	40,1

Вариант 3.

X:	10,2	52,1	50,4	32,4	26,4	18,1	14,6	16,7	26,4	26,8	50,2	13,8	10,5	14,6	12,2
Y:	15,7	42,3	30,5	18,5	50,2	17,5	44,6	16,7	36,4	28,6	40,2	50,3	12,4	13,8	30,5

Вариант 4.

X:	8,5	15,2	18,5	30,4	52,1	10,2	18,1	20,3	15,7	14,6	19,2	20,4	32,1	17,5	32,4
Y:	52,1	40,7	45,1	42,3	16,7	23,1	25,2	16,4	15,2	10,8	50,4	46,2	39,7	30,5	16,7

Вариант 5.

X:	15,1	28,4	35,6	18,1	33,5	32,4	35,8	40,1	18,5	16,4	10,8	12,6	30,4	23,4	20,5
Y:	26,4	27,1	45,2	50,2	16,8	19,5	20,4	27,8	24,3	15,7	18,1	26,4	15,8	17,5	50,2

Вариант 6.

X:	32,8	31,5	12,8	16,1	25,4	40,7	42,8	14,9	38,5	36,4	39,1	40,4	18,1	20,5	40,1
Y:	40,2	14,3	15,8	34,5	50,7	52,1	14,8	12,3	16,7	18,2	25,1	18,2	30,5	16,7	26,4

Задача №2.

В университете проведен анализ успеваемости студентов по математике. Для этого случайным образом выбрано 30 студентов, отдельно 15 девушек (Д) и 15 юношей (Ю) и результаты их аттестации по системе РИТМ даны в следующих двух выборках.

Объединив обе выборки в одну:

1. Составить интервальный вариационный ряд, разбив значения варианты на 6 интервалов.
2. Построить гистограмму, эмпирическую функцию распределения, кумуляту.
3. Найти средний бал, выборочную дисперсию и среднее квадратическое отклонение для объединенной выборки и отдельно для юношей и девушек.
4. (вар.7,8) Найти вероятность того, что средний бал в выборке отличается от среднего балла всего курса не более, чем на 0,2 (по абсолютной величине) отдельно для юношей, девушек и в объединенной выборке. Определить границы, в которых с надежностью 0,95 заключен средний балл всех студентов данного курса.
(вар.9,10) Оценить вероятность того, что доля успевающих студентов (т. е. тех, у которых балл ≥ 3) среди всех студентов курса и в объединенной выборке отличаются по абсолютной величине не более, чем на 0,01. Каков должен быть минимальный объем выборки, чтобы эта вероятность была не меньше 0,99?
5. (вар.7,8) На уровне значимости $\alpha=0,05$ можно ли утверждать, что успеваемость не зависит от пола? Одинаковый ли для девушек и юношей разброс вокруг среднего балла?

(вар.9,10) Можно ли утверждать с надежностью 0,95, что доли успевающих студентов и студенток совпадают и что разброс вокруг среднего балла студентов курса (без различия пола) не превышает 0,2?

6. На основании вида гистограммы выдвинуть гипотезу о виде закона распределения аттестационного балла студентов данного курса и проверить ее с помощью критерия Пирсона на уровне значимости $\alpha=0,1$.

Вариант 7.

Ю:	2,1	1,5	2,8	4,6	4,2	3,7	2,5	0,1	1,4	0	2,8	4,3	3,8	2,5	4,3
Д:	4,8	2,5	3,1	0,5	2,1	3,1	2,4	4,2	1,5	0,8	2,1	1,6	2,4	5,0	4,1

Вариант 8.

Ю:	2,1	4,8	3,4	2,7	1,5	2,5	4,8	4,2	2,8	3,1	2,7	3,1	4,6	0,5	3,1
Д:	2,1	4,2	2,1	1,5	4,5	3,7	3,1	1,2	3,0	2,5	2,4	4,3	1,8	0,1	4,2

Вариант 9.

Ю:	2,1	4,2	1,4	1,5	4,1	3,2	0,3	0,8	2,3	2,7	5,0	2,8	2,1	2,7	0,5
Д:	4,3	1,6	3,1	2,1	3,8	2,4	1,7	3,8	2,5	4,1	1,9	3,1	4,3	4,1	2,5

Вариант 10.

Ю:	1,5	3,1	3,1	4,5	2,8	0,5	1,5	4,5	3,7	2,4	2,8	3,9	4,5	5,0	4,3
Д:	4,8	4,2	3,1	2,1	1,9	2,8	0,8	1,2	3,7	4,6	2,9	3,5	4,7	4,9	1,7

Задача №3.

Для изучения влияния двухнедельной диеты на изменение веса спортивный клуб провел анализ по случайной выборке из 15 человек до и после диеты. Получены следующие результаты. N – номер испытуемого, X (кг) его вес до диеты, Y (кг) – его же после диеты.

Вариант 11.

N :	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X :	85,5	92,7	79	68,6	102,5	88,3	82,7	90,5	77,5	85,3	72,5	108,7	80,3	79	90,4
Y :	83	90,5	77,5	68	101,1	88,1	80,1	89,2	76	85,1	69	100	75,1	71	82,7

Вариант 12.

N :	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X :	90,2	85,1	79,2	100,4	89	87,2	75,3	69,5	72,7	90,3	95,6	98,6	75,1	69,1	95,2
Y :	90	80,1	74	95,1	80,3	87	74,3	67,2	70,1	85,4	92,3	90,4	70,1	67,2	92,4

1. Построить графики эмпирических функций распределения и гистограммы (разбив на 5 интервалов) веса испытуемых до диеты (X) и потери веса ($Y-X$).
2. Найти средний вес и дисперсии веса испытуемых до и после диеты.
3. Определить 95% доверительные интервалы для: *a)* среднего веса до диеты, *в)* среднего веса после диеты, *с)* средней потери веса, *д)* дисперсии средней потери веса.
4. Можно ли по имеющимся данным достаточно объективно (с надежностью 0,99) оценить результаты диеты? Есть ли основание не доверять рекламному проспекту, обещающему среднюю потерю веса 3 кг?
5. Можно ли утверждать, что дисперсии среднего веса до и после диеты совпадают? ($\alpha=0,1$)
6. На уровне значимости $\alpha=0,1$ критерием Пирсона проверить гипотезу о равномерном распределении потери веса.

Задача №4.

Для выяснения объема валюты, необходимой банку для выплаты клиентам в течение одного дня, была произведена случайная выборка в период с 1 по 20 января, где X (1000\$) сумма выплат.

Вариант 13.

$X:$	5,5	8,7	9,5	10,1	6,8	9,2	8,5	8,9	6,5	5,9	7,3	10,2	9,8	8,5	8,7
	9,1	10,1	8,7	9,3	9,7										

Вариант 14.

$X:$	5,6	8,7	8,6	9,4	9,5	9,7	10,2	6,9	9,1	8,7	9,2	7,5	6,4	9,3	7,8
	10,6	9,3	10,1	8,4	6,8										

Вариант 15.

$X:$	8,9	7,8	7,6	6,9	10,5	10,1	9,2	8,5	7,6	9,2	7,5	9,2	7,5	8,5	8,7
	10,4	10,2	8,1	8,5	9,3										

1. Построить вариационный ряд СВ X разбив на 5 интервалов.
2. Построить графики гистограммы и кумуляты СВ X .
3. Найти средний дневной объем выплат, среднее квадратическое отклонение, моду и медиану.
4. Построить интервал для суммы выплат одного дня и интервал среднего квадратического отклонения с надежностью 0,99. Каким должен быть минимальный объем выборки, чтобы с той же надежностью гарантировать вдвое меньший интервал?

5. На уровне значимости $\alpha=0,05$ проверить гипотезу о том, что средний объем дневной выплаты равен 9000\$ и гипотезу о том, что доля дней, когда объем выплат не больше 9000\$ равна 50%.
6. Проверить с помощью критерия Пирсона на уровне значимости $\alpha=0,05$ гипотезу о равномерном распределении величины дневных выплат.

Задача №5.

Пусть X – число сделок на фондовой бирже за квартал для $n=400$ инвесторов. Здесь x_i – число заключенных сделок, а n_i – число инвесторов, заключивших такое количество сделок.

Вариант 16.

x_i	0	1	2	3	4	5	6	7	8	9	10
n_i	150	90	60	30	25	10	10	9	6	8	2

Вариант 17.

x_i	0	1	2	3	4	5	6	7	8	9	10
n_i	140	93	75	32	17	15	10	8	5	4	1

Вариант 18.

x_i	0	1	2	3	4	5	6	7	8	9	10
n_i	132	97	72	30	20	15	10	10	8	4	2

1. Построить полигон относительных частот, кумуляту, эмпирическую функцию распределения.
2. Найти среднее число сделок, приходящихся на одного инвестора, дисперсию этого числа, моду и медиану.
3. В каких пределах с вероятностью 0,95 можно гарантировать среднее число сделок, заключенных одним инвестором? Оценить вероятность того, что доля тех инвесторов, которые заключили не менее двух сделок в выборке отличается от той же доли для всей биржи по абсолютной величине не более, чем на 0,1.
4. На уровне значимости 0,05 проверить гипотезы: а) среднее количество сделок, заключенных одним инвестором за квартал равно двум, в) доля инвесторов, заключивших более одной сделки равна 50%.
5. С помощью критерия Пирсона на уровне значимости $\alpha=0,1$ проверить гипотезу о том, что закон распределения числа сделок, приходящихся на одного инвестора, является законом Пуассона (положить $\lambda=\bar{x}$)

Задача №6.

Для изучения времени безотказной работы сверлильных станков было случайным образом выбрано 100 станков. Результаты представлены в таблице:

Вариант 19.

Время безотказной работы	0-5	5-10	10-15	15-20	20-25	25-30
Число станков	66	22	6	3	2	1

Вариант 20.

Время безотказной работы	0-5	5-10	10-15	15-20	20-25	25-30
Число станков	60	20	8	6	5	1

1. Построить гистограмму, эмпирическую функцию распределения случайной величины T – время безотказной работы станка.
2. Найти среднее время безотказной работы, среднее квадратическое отклонение, моду и медиану.
3. Найти: а) интервалы, в котором с надежностью 95% будет заключено время безотказной работы и его среднее квадратическое отклонение, в) оценить вероятность того, что среднее время безотказной работы станков отклонится от среднего времени в выборке по абсолютной величине не более, чем на 2 часа.
4. На уровне значимости $\alpha=0,1$ проверить гипотезы:
 - а) среднее время безотказной работы равно 4 часам;
 - б) дисперсия T равна 1 час²;
 - с) процент станков, время безотказной работы которых не меньше, чем 20 часов, отличается незначимо от 3%.
5. Предполагая, что время безотказной работы станков имеет показательный закон распределения (в качестве оценки параметра λ взять $\frac{1}{\bar{T}}$), проверить эту гипотезу критерием Пирсона с уровнем значимости $\alpha=0,05$.

ЛИТЕРАТУРА

1. Гмурман В. Е. Теория вероятностей и математическая статистика. –М.: «Высшая школа». 1977.
2. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике. –М.: «Высшая школа». 2003.
3. Ермаков В. И. и др. Общий курс высшей математики для экономистов. –М.: «ИНФРА-М». 2002.
4. Красс М. С., Чупрынов Б. П. Основы математики и ее приложения в экономическом образовании. –М.: «Дело». 2000.
5. Дмитрий Письменный. Конспект лекций по теории вероятностей, математической статистике и случайным процессам. –М.: «Айрисс Пресс». 2010.
6. Кремер Н. Ш. Теория вероятностей и математическая статистика. –М.: «ЮНИТИ». 2000

Таблицы квантилей

Квантили u_p стандартного нормального закона распределения $N(1,0)$.

p	0,9	0,95	0,975	0,99	0,995	0,999	0,9995
u_p	1,282	1,645	1,960	2,325	2,576	3,090	3,291

Квантили $\chi_p^2(k)$ закона распределения $\chi^2(k)$,

p \ k	0,01	0,025	0,05	0,1	0,9	0,95	0,975	0,99
1	0,0002	0,001	0,004	0,0158	2,71	3,84	5,02	6,63
2	0,02	0,05	0,103	0,211	4,61	5,99	7,38	9,21
3	0,115	0,216	0,352	0,584	6,25	7,81	9,35	11,3
4	0,297	0,484	0,711	1,06	7,78	9,49	11,1	13,3
5	0,554	0,831	1,15	1,61	9,24	11,1	12,8	15,1
10	2,56	3,25	3,94	4,87	16	18,3	20,5	23,2
20	7,63	9,59	10,9	12,4	28,4	31,4	34,2	37,6
30	14,3	16,8	18,5	20,6	40,3	43,8	47,0	50,9
40	22,2	24,4	26,5	29,1	51,8	55,8	59,3	63,7
50	29,7	32,4	34,8	37,7	63,2	67,5	71,4	76,2
75	49,5	53	56,1	59,8	91,1	96,2	100,8	106,4
100	70,1	74,2	77,9	82,4	118,5	124,3	129,6	135,6

Квантили $t_p(k)$ закона распределения Стьюдента $T(k)$,

p \ k	0,9	0,95	0,975	0,99	0,995
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,92	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
10	1,372	1,812	2,228	2,764	3,169
20	1,325	1,725	2,086	2,528	2,845

30	1,312	1,697	2,042	2,457	2,750
40	1,303	1,684	2,021	2,423	2,704
120	1,289	1,658	2,98	2,358	2,617

Квантили $F_{0,9}(n_1, n_2)$ закона распределения Фишера

$n_1 \backslash n_2$	1	2	3	4	5	10	15	20	30	120
1	40	8,53	5,54	4,54	4,06	3,29	3,07	2,97	2,88	2,75
2	49,5	9	5,46	4,32	3,78	2,92	2,7	2,59	2,49	2,35
3	53,6	9,16	5,39	4,19	3,62	2,73	2,49	2,38	2,28	2,13
4	55,8	9,24	5,34	4,11	3,52	2,61	2,36	2,25	2,14	1,99
30	62,2	9,46	5,17	3,82	3,17	2,16	1,87	1,74	1,61	1,41

Квантили $F_{0,95}(n_1, n_2)$ закона распределения Фишера

$n_1 \backslash n_2$	1	2	3	4	5	10	15	20	30	120
1	161	18,5	10,13	7,71	6,61	4,96	4,54	4,35	4,24	3,92
2	199	19	9,55	6,94	5,79	4,1	3,68	3,49	3,39	3,07
3	216	19,16	9,28	6,59	5,41	3,71	3,29	3,1	2,99	2,68
4	225	19,25	9,12	6,39	5,19	3,48	3,05	2,87	2,76	2,45
30	250	19,46	8,62	5,75	4,5	2,7	2,25	2,04	1,84	1,55

ОГЛАВЛЕНИЕ

1. Введение.....	3
2. Статистическое распределение выборки.....	3
3. Полигон и гистограмма.....	4
4. Эмпирическая функция распределения.....	7
5. Примеры теоретических законов распределения.....	8
6. Статистические оценки параметров распределения.....	10
7. Оценка генеральной средней по выборочной средней.....	10
8. Среднее квадратическое отклонение.....	11
9. Оценка генеральной дисперсии по исправленной выборочной.....	12
10. Статистические точечные оценки параметров основных законов распределения.....	13
11. Метод вычисления точечных оценок параметров распределения с помощью введения условных вариантов.....	13
12. Метод наибольшего правдоподобия для оценки точечных параметров распределения.....	15
13. Число степеней свободы.....	16
14. Основные стандартные законы распределения, их таблицы и применение этих таблиц.....	16
15. Интервальные оценки.....	19
16. Проверка статистических гипотез.....	22
17. Непараметрические гипотезы. Критерий Пирсона.....	27
18. Вопросы для самопроверки.....	32
19. Варианты заданий.....	33
20. Литература.....	39
21. Приложение.....	40

УДК – 519.2

К - 55

Кочнева Л.Ф., Липкина З.С., Новосельцева В.И. ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА. Часть III. Математическая статистика: Учебное пособие. - М.: МИИТ, 2012 – 42с.

Предназначено для направления 080100.62 «Экономика», в учебных планах которого предусмотрена дисциплина «Теория вероятности и математическая статистика». Учебное пособие содержит краткое изложение основных понятий теории вероятностей, приводятся решенные примеры, задачи и индивидуальные задания для бакалавров направления «Экономика».

Рецензенты: к. ф.м. н. доцент МИИТ О.А.Платонова;

д.ф.-м.н. профессор МГУ

им.М.В. Ломоносова А.Л. Шмелькин.

© МИИТ, 2012

Св. план 2012г, 157поз.

Кочнева Людмила Федоровна, Липкина Зоя Семеновна,
Новосельцева Вера Ивановна

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Часть III. Математическая статистика.

Учебное пособие для направления 080100.62 «Экономика»

Подписано в печати

Формат 60 X 84/16

Заказ №

Усл. – печ.л. –

Тираж – 150 экз.
