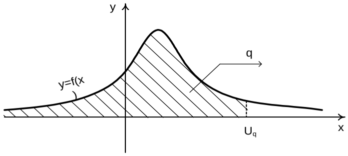
ФГБОУ ВПО «Воронежский государственный технический университет»

Кафедра высшей математики
и физико-математического моделирования

МЕТОДИЧЕСКИЕ УКАЗАНИЯ
для организации самостоятельной работы
по изучению дисциплины
«Теория вероятностей и математическая статистика»
для студентов направления подготовки бакалавров 230100
«Информатика и вычислительная техника»
(профиль «Системы автоматизированного проектирования»)
очной формы обучения
Часть 3



Воронеж 2013

Составители: канд. физ.-мат. наук Е.Г. Глушко,
канд. физ.-мат. наук А.П. Дубровская

УДК 517.9

Методические указания для организации самостоятельной работы по изучению дисциплины «Теория вероятностей и математическая статистика» для студентов направления подготовки бакалавров 230100 «Информатика и вычислительная техника» (профиль «Системы автоматизированного проектирования») очной формы обучения. Ч. 3 / ФГБОУ ВПО «Воронежский государственный технический университет»; сост. Е.Г. Глушко, А.П. Дубровская. Воронеж, 2013. 55 с.

В методических указаниях содержатся основные теоретические положения по дисциплине «Теория вероятностей и математическая статистика». Приводится большое количество решенных типовых задач, задач для самостоятельного решения.

Методические указания подготовлены в электронном виде в текстовом редакторе MS Word 2003 и содержатся в файле ТВ3.doc .

Библиогр.: 6 назв.

Рецензент канд. физ.-мат. наук, доц. В.В. Ломакин

Ответственный за выпуск зав. кафедрой
д-р физ.-мат. наук, проф. И.Л. Батаронов

Издается по решению редакционно-издательского совета Воронежского государственного технического университета

Занятие №14. Элементы математической статистики

14.1. Основные определения

Математической статистикой называется наука, которая основывается на методах теории вероятностей, занимается систематизацией, обработкой и использованием экспериментальных данных для получения научных и практических выводов.

Одним из основных методов исследования случайных явлений в математической статистике является *выборочный метод*. Рассмотрим основные понятия этого метода.

Пусть рассматривается некоторый случайный эксперимент, связанный с СВ X, имеющей ФР F(x). Полный набор всех возможных результатов измерений СВ X в эксперименте называют *генеральной совокупностью* (ГС) с ФР F(x).

Число членов N, образующих генеральную совокупность, называют *объемом генеральной совокупности*.

Отметим, что объем генеральной совокупности может быть как конечным, так и бесконечным.

Выборкой (*выборочной совокупностью* (ВС)) объемом n из N генеральной совокупности (ГС) называется последовательность $x_1,x_2,...,x_n$ наблюдаемых значений СВ X, соответствующих n независимым повторениям эксперимента .

Метод, состоящий в том, что на основании изучения характеристик и свойств выборки $x_1,x_2,...,x_n$ даются заключения о числовых характеристиках и законе распределения СВ X, называется *выборочным методом*. Выборка может быть записана в виде вариационного ряда или в виде статистического ряда.

Вариационным рядом выборки $x_1,x_2,...,x_n$ называется способ ее записи, при котором элементы x_i упорядочиваются по величине, то есть записываются в виде последовательности $x^1, x^2, ..., x^n$, причем $x^1 \le x^2 \le ... \le x^n$.

Разность между максимальным и минимальным элементами выборки $x^n - x^1 = \omega$ называется *размахом выборки*.

Пусть в выборке объемом n элемент x_i встречается n_i раз. Число n_i называется *частотой элемента* x_i. Очевидно, что

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

Статистическим рядом называется последовательность пар (x_i, n_i), которая записывается обычно в виде таблицы

x_i	x_1	x_2	…	x_k	
i	n_1	n_2	…	n_k	$\sum_{i=1}^{k} n_i = n$

Отношение $\omega_i = n_i/n$ называется *относительной частотой*, или *частостью элемента* x_i выборки.

Статистическим распределением СВ X называется последовательность пар $(x_i, n_i / n)$, которая также записывается в виде таблицы

x_i	x_1	x_2	…	x_k	
$\omega_i = \frac{n_i}{n}$	$\frac{n_1}{n}$	$\frac{n_2}{n}$	…	$\frac{n_k}{n}$	$\sum_{i=1}^{k} \frac{n_i}{n} = 1 = \sum_{i=1}^{k} \omega_i$

При большом объеме выборки ее элементы объединяют в группы (разряды), представляя результаты опытов в виде группированного статистического ряда. Для этого интервал, содержащий все элементы выборки, разбивается на k частичных непересекающихся интервалов.

Обычно выбирают частичные интервалы одинаковой длины $b = \omega/k$. После того, как частичные интервалы выбраны, определяют частоты - количество n_i^* элементов выборки, попавших в i-тый интервал (элемент, совпадающий с верхней границей интервала, относится к следующему интервалу). В группированный статистический ряд в верхнюю строку записываются середины x_i^* интервалов группировки, а в нижний - частоты n_i^*.

В зависимости от объема выборки число k интервалов группировки берется от 6 до 20. Наряду с частотами n_i^* удобно одновременно подсчитывать также *накопленные частоты* $\sum_{j=1}^{i} n_j^*$, *относительные частоты* и *накопленные относительные частоты* $\sum_{j=1}^{i} \omega_j^*$, $\quad i = 1, 2, ..., k$.

Полученные результаты сводятся в таблицу, называемую таблицей частот группированной выборки.

Следует отметить, что группировка выборки вносит погрешность в последующие вычисления, которая становится тем больше, чем меньше выбирается число интервалов.

Пример 14.1. Дана выборка из некоторой ГС 11,15,12,9,13, 12,6,11,12,13,15,8,9,14,9,11,6. Определить объем, размах выборки, а также построить вариационный и статистический ряды.

Решение. Объем n=17.Вариационный ряд 6,6,8,9,9,11,11,11, 12,12,12,13,13,14,15,15. Размах ω=15-6=9.Статистический ряд и статистическое распределение имеют вид

x_i	6	8	9	11	12	13	14	15
n	2	1	3	3	3	2	1	2
ω_i	2/17	1/17	3/17	3/17	3/17	2/17	1/17	2/17

$\sum n_i = n; \qquad \sum \omega_i = 1$

Пример 14.2. Время решений контрольной работы студентами 2-го курса дается выборкой

38	60	41	51	33	42	45	21	53	60
68	52	47	46	49	49	10	57	54	59
79	47	28	48	58	32	42	58	61	30
61	35	47	72	41	45	44	55	30	40
67	65	39	48	43	60	54	42	59	50

Построить выборку в виде таблицы частот группированной выборки, используя 7 интервалов группировки.

Решение. Размах выборки 79-10=69. Длина интервала b=69/7$\cong$10.

Инт.	[10-20)	[20-30)	[30-40)	[40-50)	[50-60)	[60-70)	[70-80)	
x_i^*	15	25	35	45	55	65	75	
n_i^*	1	2	7	18	12	8	2	$\sum n_i = 50$
$\sum n_i^*$	1	3	10	28	40	48	50	
ω_i	0,02	0,04	0,14	0,36	0,24	0,16	0,04	$\sum \omega_i = 1$
$\sum \omega_i^*$	0,02	0,06	0,20	0,56	0,80	0,96	1,00	

14.2. Графическое представление выборки

Для наглядности сгруппированные статистические ряды представляются графиками и диаграммами.

Полигоном частот группированной выборки называется ломаная с вершинами в точках (x_i^*, n_i^*), i=1, 2,...,k.

Полигоном относительных частот группированной выборки называется ломаная с вершинами в точках (x_i^*, ω_i^*).

Гистограммой частот группированной выборки называется ступенчатая фигура, составленная из прямоугольников, построенных на интервалах так, что площадь каждого прямоугольника равна частоте n_i^*, i=1, 2,...,k. Отсюда следует, что площадь гистограммы частот равна объему выборки n. В том случае, когда длины всех интервалов одинаковы и равны b, высоты прямоугольников равны $h_i = n_i^* / b$, i=1, 2,...,k. Аналогично строится гистограмма *относительных частот*. Площадь гистограммы относительных частот равна единице.

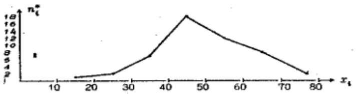
Полигоном накопленных частот группированной выборки называется ломаная с вершинами в точках ($x_i^* + b/2; \sum_{j=1}^{i} n_j^*$).

Полигоном относительных накопленных частот (кумулятивной кривой, кумулятой) называется ломаная с вершинами в точках ($x_i^* + b/2;\ (\sum_{j=1}^{i} n_j^*)/n$).

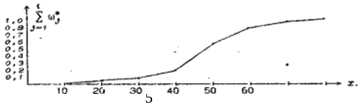
Замечание. Перечисленные графические представления аналогичным образом определяются и в случае негруппированной выборки.

Пример 14.3. Для выборки примера 14.2 построить гистограмму, полигон частот и кумулятивную кривую.

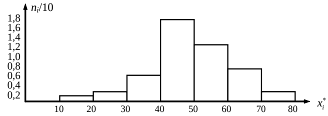
Полигон



Кумулята



Гисто-
грамма



14.3. Эмпирическая функция распределения (ЭФР)

Эмпирической функцией распределения СВ X называется функция F*(x), определяющая для каждого значения x относительную частоту события (X<x) $F^*(x)=n_x/n$, где n_x - число выборочных значений, меньших x, а n - объем выборки.

По значениям накопленных относительных частот ЭФР определяется следующим образом:

$$F^*\ x\ =\sum_{x_i<x} n_i\,/\,n\,.$$

В отличие от ЭФР, функция распределения генеральной совокупности F(x)=P(X<x) называется *теоретической функцией распределения* (ТФР).

Отметим, что разница между ТФР и ЭФР состоит в том, что ТФР определяет вероятность события (X<x), а ЭФР определяет относительную частоту этого же события.

ЭФР обладает всеми свойствами ТФР, то есть:

1) значения F*(x) принадлежат отрезку [0,1];
2) F*(x)- неубывающая функция аргумента x;
3) F*(x)=0, если $x<x_1$, и F*(x)=1, если $x>x_n$, где x_1 - наименьшее, а x_n - наибольшее наблюдаемые значения СВ X .

Из закона больших чисел, а именно из теоремы Бернулли, следует, что при объеме выборки $n\to\infty$ ЭФР сходится по вероятности к ТФР. Это означает, что при достаточно большом объеме выборки ЭФР $F^*(x)$ и ТФР мало отличаются друг от друга.

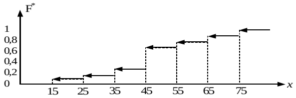
Основное значение ЭФР состоит в том, что она используется в качестве оценки ТФР.

Пример 14.4. Построить график ЭФР по выборке примера 14.2.

Решение. ЭФР имеет вид

$$F^*(x)=\begin{cases}0 & x\le 15;\\ 0,02 & 15<x\le 25;\\ 0,06 & 25<x\le 35;\\ 0,20 & 35<x\le 45;\\ 0,56 & 45<x\le 55;\\ 0,80 & 55<x\le 65;\\ 0,96 & 65<x\le 75;\\ 1 & 75<x.\end{cases}$$

График $F^*(x)$ имеет вид



Задачи для самостоятельного решения:

14.1. В течение суток измеряют напряжение X тока в электросети в вольтах. В результате опыта получена выборка объема $n=30$:

107	108	110	109	110	111	109	110	111	107
108	109	110	108	107	110	109	111	111	110
109	112	113	110	106	110	109	110	108	112

Построить статистический ряд этой выборки. Построить полигон относительных частот. Найти эмпирическую функцию и построить ее график.

14.2. По данному распределению выборки

x_i	1	3	6
n_i	10	25	15

найти эмпирическую функцию и построить ее график.

14.3. Дана выборка:

x_i	2	4	5	7	10
n_i	15	20	10	10	45

Найти эмпирическую функцию распределения, построить ее график. Построить полигон относительных частот выборки.

14.5. Измерен рост n=500 студентов. Результаты измерений представлены в виде интервального статистического ряда:

[145;150)	[150;155)	[155;160)	[160;165)	[165;170)
1	2	28	90	169
[170;175)	[175;180)	[180;185)	[185;190)	[190;195]
132	55	16	6	1

Построить гистограмму относительных частот.

14.6. По данным выборки построить гистограмму относительных частот:

1.

Номер интервала	Интервал	Число вариант в интервале
1	(1;5)	10
2	(5;9)	20
3	(9;13)	50
4	(13;17)	12
5	(17;21)	8

2.

Номер интервала	Интервал	Число вариант в интервале
1	(2;5)	6
2	(5;8)	10
3	(8;11)	5
4	(11;14)	4

14.7. Дана выборка:

38	60	41	51	33	42	45	21	53	60
68	52	47	46	42	43	57	44	54	59
77	47	28	27	49	49	14	28	61	30
61	35	47	46	58	45	42	21	30	40
67	65	39	35	41	60	54	42	59	60

Построить гистограмму относительных частот.

14.8. Построить полигон относительных частот следующей выборки:

x_i	4	6	10	12
n_i	10	15	5	20

Занятие №15. Статистическая оценка неизвестных параметров распределения. Точечные оценки

15.1. Постановка задачи

При обработке опытных данных часто бывает так, что вид закона распределения генеральной совокупности (ЗРГС) известен, а требуется найти только некоторые параметры, от которых он зависит. Например, если известно, что СВ X распределена по нормальному закону, то на основании опытных данных (по выборке) необходимо "оценить", то есть найти приближенное значение двух параметров - МО и СКО.

Одна из задач математической статистики и состоит в нахождении оценок неизвестных параметров по выборке наблюдений.

Пусть из ГС с ФР F(x,θ), где θ - неизвестный параметр, произведена выборка x_1, x_2,…,x_n объемом n. В качестве оценки параметра θ рассматривают *функции элементов выборки* $\tilde{\theta} = \mathrm{U}\ \ x_1, x_2, ..., x_n$, которые называются *статистиками*.

Задача оценки неизвестного параметра θ сводится к нахождению таких статистик (выборочных функций) $\tilde{\theta} = \mathrm{U}\ \ x_1, x_2, ..., x_n$, которые могут быть использованы для приближенного определения значения неизвестного параметра θ.

Оценки параметров подразделяются на точечные и интервальные.

15.2. Основные свойства точечных статистических оценок распределения

Точечная оценка параметра θ определяется одним числом $\tilde{\theta} = \mathrm{U}\ \ x_1, x_2, ..., x_n$.

Качество оценок характеризуется некоторыми *свойствами*. Сформулируем основные из них.

Свойство 1. Оценка $\tilde{\theta}$ называется *несмещенной*, если ее МО равно оцениваемому параметру, то есть если М($\tilde{\theta}$)=θ. Разность М($\tilde{\theta}$)-θ называется *смещением*.

Свойство 2. Оценка $\tilde{\theta}\ \ n\ \ = \mathrm{U}\ \ x_1, x_2, ..., x_n$ называется *состоятельной*, если при увеличении объема выборки n оценка $\tilde{\theta}\ \ n$ сходится по вероятности к θ, то есть $\forall \varepsilon > 0$ $\lim \mathrm{P}\ \ \left|\tilde{\theta}\ \ n\ \ -\theta\right| < \varepsilon\ \ = 1$.

Свойство 3. Пусть $\tilde{\theta}_1$ и $\tilde{\theta}_2$ - две различные несмещенные оценки параметра θ. Если D($\tilde{\theta}_1$)<D($\tilde{\theta}_2$), то говорят, что оценка $\tilde{\theta}_1$ *более эффективна*, чем оценка $\tilde{\theta}_2$.

Требование несмещенности устраняет систематические ошибки в определении оценок, обусловленных ограниченным объемом выборки.

Требование состоятельности гарантирует от совершения грубых ошибок ε в определении θ при достаточно большом объема n выборки.

Свойство эффективности используется для выбора оценки, обладающей наименьшим разбросом.

15.3. Статистическая оценка МО

В качестве статистической оценки МО выбирается выборочное среднее.

Выборочным средним $\overline{x}$ называется среднее арифметическое элементов выборки

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i . \qquad (15.1)$$

Если x_i - варианты выборки, n_i -частоты вариант x_i , $i=\overline{1,k}$, $n=\sum_{i=1}^{k} n_i$ - объем выборки, то

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i \qquad (15.2)$$

Для группированной выборки это соотношение принимает вид

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i^* n_i^* . \qquad (15.2)$$

Выборочное среднее $\overline{x}$ является несмещенной и состоятельной оценкой.

По поводу эффективности $\overline{x}$ заметим, что если СВ X распределена по нормальному закону, то выборочное среднее является эффективной оценкой МО.

15.4. Статистическая оценка дисперсии

В качестве статистической оценки дисперсии D(X) СВ X примем *выборочную дисперсию*

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2, \qquad (15.4)$$

или

$$s^2 = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i - \overline{x})^2 \qquad (15.5)$$

для группированной выборки

$$s^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i^* - \overline{x})^2 \cdot n_i^*. \qquad (15.6)$$

Но оценка s^2 является смещенной. Для получения несмещенной оценки дисперсии выборочную дисперсию s^2 исправляют, умножая ее на множитель $n/(n\text{-}1)$.

Исправленная дисперсия

$$s_0^2 = \frac{n}{n-1}\cdot s^2 = \frac{1}{(n-1)}\cdot\sum_{i=1}^{n}(x_i - \overline{x})^2 \qquad (15.7)$$

Или

$$s_0^2 = \frac{n}{n-1}\cdot s^2 = \frac{1}{(n-1)}\cdot\sum_{i=1}^{k} n_i (x_i - \overline{x})^2 \qquad (15.8)$$

является уже несмещенной оценкой дисперсии.

Непосредственно из определений следует, что

$$s_0^2 = \frac{1}{(n-1)}\cdot\sum_{i=1}^{k} n_i x_i^2 - \frac{n}{n-1}\overline{x}^2. \qquad (15.9)$$

Величина s_o называется *исправленным средним квадратическим отклонением*.

Оценка s^2 (а вместе с ней и s_0^2) состоятельна.

Пример 15.1. Оценить математическое ожидание и дисперсию случайной величины X по результатам ее независимых наблюдений: 7, 3, 4, 8, 4, 6, 3.

Решение. По формулам (15.1) и (15.4) имеем

$$M\left(X\right) \approx \tilde{X} = \frac{7+3+4+8+4+6+3}{7} = 5,$$

$$D\left(X\right) \approx s^2 = \frac{\left(7-5\right)^2 + \left(3-5\right)^2 + \left(4-5\right)^2 + \ldots + \left(3-5\right)^2}{6} = \frac{25}{6} \approx 4{,}17.$$

Пример 15.2. Данные 25 независимых наблюдений случайной величины X представлены в сгруппированном виде:

Границы интервалов	5-7	7-9	9-11	11-13	13-15
Число наблюдений	2	4	9	7	3

Оценить математическое ожидание и дисперсию этой случайной величины.

Решение. Представителем каждого интервала можно считать его середину. С учетом этого формулы (15.3) и (15.6) дают следующие оценки:

$$M\left(X\right) \approx \overline{x} = \frac{6\cdot 2+8\cdot 4+10\cdot 9+12\cdot 7+14\cdot 3}{25} = \frac{260}{25} = 10{,}4;$$

$$D\left(X\right) \approx s^2 = \frac{\left(6-10{,}4\right)^2\cdot 2+\left(8-10{,}4\right)^2\cdot 4+\ldots+\left(14-10{,}4\right)^2\cdot 3}{24} = 5.$$

Пример 15.3. По выборке признака X, заданной следующей таблицей:

x_i	45	50	55	60	65	70	75
n_i	4	6	10	40	20	12	8

найти выборочное среднее, дисперсию и среднее квадратическое отклонение.

Решение. Объем выборки $n=\sum_{i=1}^{7} n_i = 100$. По формуле (15.2) выборочное среднее

$$\overline{x} = \frac{45\cdot 4 + 50\cdot 6 + 55\cdot 10 + 60\cdot 40 + 65\cdot 20 + 70\cdot 12 + 75\cdot 8}{100} = 61{,}7.$$

Для вычисления дисперсии составляем таблицу квадратов значений СВ *X*:

x_i^2	2025	2500	3025	3600	4225	4900	5625
n_i	4	6	10	40	20	12	8

По формуле (15.9) имеем

$$s_0^2 = \frac{1}{99}\cdot\sum_{i=1}^{7} n_i {x_i}^2 - \frac{100}{99}(61{,}7)^2 = 50{,}11,$$

откуда $s_0 = \sqrt{50{,}11} \approx 7{,}08.$

Получили несмещенные оценки для дисперсии и среднего квадратического отклонения. Соответствующие смещенные оценки $s^2 = 49{,}61$ и $s = 7{,}04.$

Пример 15.4. Проведено несколько измерений расстояния. Результаты измерений в метрах представлены в виде ряда:

i	x_i	i	x_i	i	x_i	i	x_i
1	1235,6	5	1238,5	9	1234,5	13	1234,3
2	1237,5	6	1234,2	10	1236,8	14	1237,5
3	1232,9	7	1235,9	11	1237,6	15	1235,4
4	1236,2	8	1233,3	12	1233,1	16	1234,7

Вычислить оценки математического ожидания, дисперсии и среднего квадратического отклонения измеренного расстояния.

Решение. Так как значение вариант x_i большие, то удобно ввести условные варианты $u_i = x_i - a$, где в качестве a возьмем среднее число 1235, т.е. a=1235. В результате получим таблицу для условных вариант:

i	u_i	i	u_i	i	u_i	i	u_i
1	0,6	5	3,5	9	-0,5	13	-0,7
2	2,5	6	-0,8	10	1,8	14	2,5
3	-2,1	7	0,9	11	2,6	15	0,4
4	1,2	8	-1,7	12	-1,9	16	-0,3

Выборочное среднее в данном случае вычисляется по формуле

$$\bar{x} = a + \frac{1}{n}\sum_{i=1}^{k} n_i u_i = 1235 + \frac{1}{16}(16-8) = 1235,5.$$

Статистическая дисперсия

$$s_0^2 = \frac{1}{(n-1)} \cdot \sum_{i=1}^{7} n_i u_i^{\,2} - \frac{n}{n-1}(\bar{x}-a)^2 = \frac{1}{n}\sum_{i=1}^{k} n_i (u_i - \bar{u})^2,$$

где $\bar{u}$ - среднее значение условных вариант. Отсюда

$$s_0^2 = \frac{1}{15}(0,01 + 4 + 6,76 + 0,49 + 9 + 1,69 + 0,16 + 4,48 + 1 +$$

$$+1,69 + 4,41 + 5.76 + 1,44 + 4 + 0,01 + 0,64) = 3,06,$$

$$s_0 = \sqrt{s_0^2} = \sqrt{3,06} \approx 1,75 \text{ м.}$$

Задачи для самостоятельного решения:

15.1. Оценить математическое ожидание и дисперсию случайной величины X по результатам ее независимых наблюдений: 9, 3, 7, 4, 3, 8, 7, 3.

15.2. Найти выборочное среднее по заданному распределению выборки объемом $n = 20$:

x_i	2560	2600	2620	2650	2700
n_i	2	3	10	4	1

15.3. Найти выборочную дисперсию по данному распределению выборки объемом $n = 10$:

x_i	186	192	194
n_i	2	5	3

По данным выборкам найти выборочные средние и средние квадратические отклонения.

15.4.

x_i	1250	1275	1280	1300
n_i	20	25	50	5

15.5.

x_i	12,5	17,5	22,5	27,5	32,5
n_i / n	0,1	0,2	0,4	0,2	0,1

15.6.

Интер-вал	(28;30)	(30;32)	(32;34)	(34;36)
n_i	8	15	15	12
Интер-вал	(36;38)	(38;40)	(40;42)	(42;44)
n_i	15	20	10	5

15.7. Данные 30 независимых наблюдений случайной величины X представлены в сгруппированном виде:

Границы интервалов	4-8	8-12	12-16	16-20
Число наблюдений	5	8	14	3

Оценить математическое ожидание и дисперсию этой случайной величины.

15.8. При сверлении отверстий одним и тем же сверлом и последующем измерении диаметров отверстий получены данные, представленные в виде интервального статистического ряда

[40,25;40,28)	[40,28;40,31)	[40,31;40,34)	[40,34;40,37)
2	10	18	25
[40,37;40,40)	[40,40;40,43)	[40,43;40,46)	
12	8	5	$\sum_{i=1}^{n} n_i = 80.$

По данным выборкам найти выборочное среднее и среднее квадратическое отклонение.

Ответы:

15.1. $M(X) \approx 5,5, D(X) \approx 7,43$; 15.2. 2621. 15.3. 8,93.

15.4. $\bar{x} = 1273,75, s = 13,054$. 15.5. $\bar{x} = 22,5, s = 5,45$.

15.6. $\bar{x} = 35,72, s = 4,012$. 15.7. $M(X) \approx 12, D(X) \approx 12,97$;

15.8. $\bar{x} = 40,355, s = 0,04$.

Рассмотрены оценки неизвестных параметров распределения по выборке для частного случая, когда оцениваемый параметр θ является математическим ожиданием или дисперсией распределения, то есть первым начальным и вторым центральным моментами распределения. Приведем теперь общие методы нахождения точечных оценок.

15.5. Метод моментов

Пусть известен вид распределения ГС F(x, θ_1, θ_2,..., θ_k), зависящего от k параметров. Идея метода моментов заключается в следующем: по выборке вычисляют k выборочных моментов и приравнивают их к соответствующим моментам распределения ГС. Искомые оценки каждого из k параметров находятся как решение полученной системы k уравнений. При этом оценки начальных и центральных моментов k-го порядка вычисляются по формулам

$$\tilde{\alpha}_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k, \qquad \tilde{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^k \qquad (15.6)$$

Замечание. Оценки математического ожидания и дисперсии, рассмотренные выше, получены по методу моментов.

Пример 15.5. СВ X распределена по показательному закону с плотностью вероятностей $f(x,\lambda) = \begin{cases} 0, & x < 0, \\ \lambda e^{-\lambda x}, & x \ge 0. \end{cases}$

Требуется по результатам наблюдений $x_1, x_2, \ldots, x_n$ оценить параметр λ.

Решение. Ранее мы находили $M(X)=\alpha_1=1/\lambda$. Оценка математического ожидания есть выборочное среднее $\tilde{\alpha}_1 = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Приравнивая теоретический момент α_1 эмпирическому $\tilde{\alpha}_1$, получаем $1/\lambda = \overline{x}_1$. Отсюда $\lambda = 1/\overline{x} = n/\sum_{i=1}^{n} x_i$.

Пример 15.6. Случайная величина X имеет пуассоновский закон распределения: $P\left(X = k\right) = \frac{\lambda^k}{k!}e^{-\lambda}$, $k = 0, 1, 2, \ldots$. Найти оценку параметра λ по методу моментов.

Решение. Так как $M\left(X\right) = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!} = \lambda$, то оценкой параметра по методу моментов будет величина $\tilde{\lambda} = \overline{x}$.

Метод моментов обладает тем недостатком, что оценки, полученные по нему, вообще говоря, не являются асимптотически эффективными и могут быть смещенными.

15.6. Метод максимального правдоподобия

Метод максимального правдоподобия наиболее распространен при нахождении точечных оценок параметров. Будем рассматривать результаты выборки как реализацию n-мерной случайной величины $(X_1, X_2,...X_n)$ с независимыми компонентами. Для получения оценки неизвестного параметра θ естественно попытаться найти такое значение $\tilde{\theta}$, при котором вероятность реализации этой выборки x_1, x_2,..., x_n была бы максимальной.

Если СВ X дискретна, то закон распределения ее имеет вид $P(X=x_i)=p_i(\theta)$, $i=1,2,...,n$. Тогда вероятность при n независимых наблюдениях СВ X получить выборку (x_1, x_2,..., x_n) равна

$$\mathrm{L}(\theta) = \mathrm{L}(x_1, x_2, ..., x_n, \theta) = \mathrm{P}(\mathrm{X}_1 = x_1, \mathrm{X}_2 = x_2, ..., \mathrm{X}_n = x_n) =$$

$$= \mathrm{P}(\mathrm{X}_1 = x_1)\,\mathrm{P}(\mathrm{X}_2 = x_2)\,\mathrm{P}(\mathrm{X}_n = x_n) = p_1(\theta) \cdot p_2(\theta) \cdot ... \cdot p_n(\theta).$$

Функция $\mathrm{L}(\theta)$ называется функцией *правдоподобия*, а величина $\tilde{\theta}$, являющаяся точкой максимума этой функции, есть оценка параметра θ, полученная по методу максимального правдоподобия (сокращенно МП-оценкой).

Если определяется оценка непрерывной СВ X с плотностью распределения $f(x,\theta)$, то функция правдоподобия определяется так:

$$\mathrm{L}(\theta) = \mathrm{L}(x_1, x_2, ..., x_n, \theta) = f(x_1, \theta)\, f(x_2, \theta)\, f(x_n, \theta). \quad (15.7)$$

Если функция правдоподобия дифференцируема по θ и при любых возможных значениях x_i достигает максимума по θ внутри интервала возможных значений параметра θ, то $\tilde{\theta}$ находят, решая уравнение $\frac{\partial \mathrm{L}(x_1, x_2, ..., x_n, \theta)}{\partial \theta} = 0.$

Поскольку точки максимума функции $\mathrm{L}\left(x_1, x_2, ..., x_n, \theta\right)$ и $\ln \mathrm{L}\left(x_1, x_2, ..., x_n, \theta\right)$ при фиксированных $x_1,..., x_n$ совпадают, то в некоторых случаях удобно решать уравнение

$$\frac{\partial \ln \mathrm{L}\left(x_1, x_2, ..., x_n, \theta\right)}{\partial \theta} = 0.$$

Если требуется оценить не один, а k неизвестных параметров $\theta_1, \theta_2, ..., \theta_k$, то оценки максимального правдоподобия для этих параметров находят, решая систему уравнений

$$\frac{\partial \mathrm{L}\left(x_1, x_2, ..., x_n, \theta_1, ..., \theta_k\right)}{\partial \theta_j} = 0, \qquad j = 1, 2, ..., k.$$

Пример 15.7. Пусть СВ X распределена по нормальному закону N(m,σ^2) с неизвестными параметрами m и σ^2. Найти МП-оценку параметров нормального распределения.

Решение. Рассмотрим выборку x_1, x_2,…, x_n как реализацию n-мерного СВ (X_1, X_2,…, X_n). Тогда составляющие X_i также распределены по закону N(m,σ^2). Запишем функцию правдоподобия

$$\mathrm{L}\left(x_1, x_2, ..., x_n, m, \sigma^2\right) = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} \exp\left\{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i - m\right)^2\right\}.$$

Здесь удобнее перейти к $\ln \mathrm{L}$:

$$\ln \mathrm{L}\left(x_1, x_2, ..., x_n, m, \sigma^2\right) = -\left(n/2\right)\cdot\ln\left(2\pi\right) - \left(n/2\right)\ln\sigma^2 - \\ -(1/2\sigma^2)\sum_{i=1}^{n}\left(x_i - m\right)^2. \qquad (15.8)$$

Дифференцируя (15.8) по m и σ^2, получаем систему

$$\begin{cases} \dfrac{\partial \ln \mathrm{L}}{\partial m} = \dfrac{-1}{\tilde{\sigma}^2}\sum_{i=1}^{n} \left(x_i - \tilde{m}\right) = 0; \\ \dfrac{\partial \ln \mathrm{L}}{\partial \left(\sigma^2\right)} = -\dfrac{n}{2\tilde{\sigma}^2} + \dfrac{1}{2\tilde{\sigma}^4}\sum_{i=1}^{n} \left(x_i - \tilde{m}\right)^2 = 0. \end{cases}$$

Из первого уравнения находим $\tilde{m} = (1/n)\cdot\sum_{i=1}^{n} x_i = \overline{x}$. Подставив это значение во второе уравнение, получим $\tilde{\sigma}^2 = 1/n \cdot \sum_{i=1} \left(x_i - \overline{x}\right)^2$. Заметим, что оценка $\tilde{m}$ совпадает с оценкой, полученной по методу моментов, а оценка $\tilde{\sigma}^2$ не совпадает.

Пример 15.8. Пусть имеется простейший поток событий неизвестной интенсивности λ. Для оценки параметра λ проведено наблюдение потока и зарегистрированы $x_1, x_2, \ldots x_n$ - длительности n последовательных интервалов времени между моментами наступления событий. Найти оценку для λ.

Решение. В простейшем потоке интервалы времени между последовательными моментами наступления событий потока имеют показательный закон распределения

$$F\left(x\right) = 1 - \exp\left(-\lambda x\right), x \geq 0.$$

Так как плотность вероятности показательного закона распределения равна $f\left(x,\lambda\right) = F'\left(x,\lambda\right) = \lambda \cdot \exp\left(-\lambda x\right)$, то функция правдоподобия (15.7) имеет вид:

$$L = f\left(x_1,\lambda\right) \cdot f\left(x_2,\lambda\right) \cdot f\left(x_3,\lambda\right) \cdot \ldots \cdot f\left(x_n,\lambda\right) =$$

$$= \lambda \cdot \exp\left(-\lambda x_1\right) \cdot \lambda \cdot \exp\left(-\lambda x_2\right) \cdot \ldots \cdot \lambda \cdot \exp\left(-\lambda x_n\right) = \lambda^n \exp\left\{-\lambda \sum_{i=1}^{n} x_i\right\}.$$

Тогда $\ln L = n\ln\lambda - \lambda\sum_{i=1}^{n} x_i$ и уравнение правдоподобия $\frac{\partial \ln L}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$ имеет решение $\tilde{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}}$.

При таком значении $\lambda = \tilde{\lambda}$ функция правдоподобия действительно достигает наибольшего значения, так как

$$\frac{\partial^2 \ln L}{\partial \lambda^2} = -\frac{n}{\lambda^2} < 0.$$

Пример 15.9. Случайная величина X имеет равномерное распределение на отрезке $\theta - b; \theta + b$, где θ и b неизвестны. Пусть $x_1, x_2, \ldots, x_n$ - результаты n независимых наблюдений. Найти оценку параметра θ.

Решение. Функция плотности вероятности величины X имеет вид

$$f\ \ x, b, \theta\ \ = \begin{cases} 1/(2b), & x \in (-b, b), \\ 0, & x \notin (-b, b). \end{cases}$$

В этом случае функция правдоподобия $L = \ 1/(2b)\ ^n$ от θ явно не зависит. Дифференцировать по θ такую функцию нельзя и нет возможности записать уравнение правдоподобия. Однако, легко видеть, что L возрастает при уменьшении b. Все результаты наблюдений лежат в $\theta - b; \theta + b$, поэтому можно записать, что

$$\theta - b \le x^{\ 1}\ , x^{\ n}\ \ \le \theta + b,$$

где x^{1} - наименьший, а x^{n} - наибольший из результатов наблюдений. При минимально возможном b

$$\theta - b = x^{1},\ x^{n} = \theta + b,$$

откуда $\theta - b + \theta + b = x^{1} + x^{n}$ или $2\theta = x^{1} + x^{n}$. Оценкой наибольшего правдоподобия для параметра θ будет величина

$$\tilde{\theta} = \frac{x^{1} + x^{n}}{2}.$$

Примеры для самостоятельного решения

15.1. Методом моментов найти оценку параметра $\theta = p$, где *p*- есть вероятность «успеха» в любом из *n* независимых повторных наблюдений, а случайная величина *k*- число «успехов».

15.2. Случайная величина X (число появлений события А в *n* независимых испытаниях) подчинена биномиальному распределению с неизвестным параметром распределения *p*. Проведено 10 опытов по 5 испытаний в каждом. В результате получено эмпирическое распределение

x_i	0	1	2	3	4	5
n_i	4	2	1	1	1	1

где x_i - число появлений события А в одном опыте; n_i - количество опытов, в которых А появилось x_i раз. Методом моментов найти точечную оценку параметра *p* биномиального распределения.

15.3. Пусть дана случайная выборка $(x_1, x_2, ..., x_n)$ объема *n* из ГС X, имеющей равномерный закон распределения

$$f\;x = \begin{cases} 1/(b-a),\ x \in (a,b); \\ 0,\ x \notin (a,b), \end{cases}$$

с неизвестными параметрами a и b. Найти методом моментов точечные оценки этих параметров.

15.4. Случайная величина X- ошибка измерения дальности радиодальнометра - подчинена равномерному распределению с неизвестными параметрами a и b. Статистическое распределение СВ X имеет вид

x_i	3	5	7	9	11	13	15	17	19	21
n_i	21	16	15	26	22	14	21	22	18	25

где x_i - средняя ошибка измерений; n_i - количество измерений, имеющих среднюю ошибку x_i. Методом моментов найти точечные оценки неизвестных параметров a и b равномерного распределения.

15.5. Методом наибольшего правдоподобия найти по выборке $x_1, x_2, ..., x_n$ точечную оценку параметра p геометрического распределения $P(X = x_i) = (1 - p)^{x_i - 1} p$, где p- вероятность появления события в отдельном испытании.

15.6. Случайная величина имеет плотность вероятности $f\ x\ = \frac{x^{m-1}\lambda^m}{m-1\ !} e^{-\lambda x}$, где λ и m - параметры. Найдите оценку наибольшего правдоподобия для параметра λ.

15.7. При испытании 10 однотипных датчиков импульсного питания зафиксирована их наработка в часах до первого отказа: $x_1 = 52$, $x_2 = 354$, $x_3 = 600$, $x_4 = 418$, $x_5 = 97$, $x_6 = 452$, $x_7 = 553$, $x_8 = 127$, $x_9 = 211$, $x_{10} = 136$. Из теоретических соображений известно, что время безотказной работы датчика имеет функцию распределения $F\ x\ = 1 - \exp\ -\lambda x\ $, $0 \le x, \lambda > 0$

(показательный закон распределения). Найдите на основе опытных данных наиболее правдоподобное значение λ.

15.8. Случайная величина имеет закон распределения Рэлея. Функция плотности вероятности этого закона распределения имеет вид

$$f\left(x\right)=\frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right),\ x\geq 0\,.$$

Найдите оценку наибольшего правдоподобия для σ, если результаты наблюдений дали следующие значения случайной величины: $x_1=1{,}5$, $x_2=1{,}1$, $x_3=1{,}2$.

15.9. Отказ устройства произошел при k-м по счету испытании. Найдите оценку наибольшего правдоподобия для вероятности отказа устройства при одном испытании.

Ответы: 15.1. $\tilde{p}=k/n.$ 15.2. $p=0{,}32.$

15.3. $\tilde{b}=\overline{x}+\sqrt{3}\tilde{\sigma},\ \tilde{a}=\overline{x}-\sqrt{3}\tilde{\sigma}.$ 15.4. $a=2{,}24; b=2{,}38.$

15.5. $p=1/\overline{x}.$ 15.6. $\lambda=mn/\sum_{i=1}^{n}X_i;$ 15.7. $\tilde{\lambda}=1/300;$

15.8. $\tilde{\sigma}=1{,}4;$ 15.9. $1/k.$

Занятие 16. Интервальные оценки параметров распределения

16.1. Доверительный интервал и доверительная вероятность

Интервальной называют оценку, которая определяется двумя числами $\tilde{\theta}_1$ и $\tilde{\theta}_2$ - концами интервала, накрывающего оцениваемый параметр θ.

На основании выборки $x_1, x_2,\ldots, x_n$ указывают два значения $\theta_1(x_1, x_2,\ldots, x_n)$ и $\theta_2(x_1, x_2,\ldots, x_n)$, с помощью которых можно сделать статистический вывод о том, что истинное значение параметра θ лежит в интервале (θ_1, θ_2).

Доверительным интервалом для параметра θ называется интервал (θ_1, θ_2), содержащий истинное значение параметра с заданной вероятностью p=1-α. Таким образом,

$$P(\theta_1<\theta<\theta_2)=1-\alpha. \tag{16.1}$$

Число 1-α называется *доверительной вероятностью* (надежностью), а значение α - *уровнем значимости*.

Часто применяют односторонние доверительные интервалы, границы которых определяются из условия $P(\theta<\theta_2)=1-\alpha$ или $P(\theta_1<\theta)=1-\alpha$.

Эти интервалы называются соответственно *левосторонними* и *правосторонними* доверительными интервалами. Выбор доверительной вероятности определяется в каждом случае конкретными условиями. Обычно используемые значения 1-α равны 0,90; 0,95; 0,99. На практике часто рассматривают *симметричные доверительные интервалы* длиной 2δ. Соотношение (16.1) в этом случае записывается в виде

$$P\left(\left|\tilde{\theta}-\theta\right|<\alpha\right)=1-\alpha \text{ или } P\left(\tilde{\theta}-\delta<\theta<\tilde{\theta}+\delta\right)=1-\alpha. \tag{16.2}$$

Длина доверительного интервала играет важную роль: чем меньше длина доверительного интервала, тем точнее оценка. Если же длина доверительного интервала велика, то оценка малопригодна для практики.

Из соотношения (16.1) или (16.2) и из того, что θ_1 и θ_2 являются функциями выборки, следует, что длина доверительного интервала определяется двумя величинами: доверительной вероятностью 1-α и объемом n выборки.

Таким образом, величины δ, (1-α), n взаимосвязаны и, задавая определенные значения двум из них, можно определить значение третьей. Процедура нахождения границ доверительного интервала для параметра θ по заданной доверительной вероятности в простейшем случае состоит в следующем.

1. Из ГС с ФР $F(x,\theta)$ извлекается выборка объемом n. По этой выборке методом моментов или методом максимального правдоподобия находится точечная оценка $\tilde{\theta}$ неизвестного параметра θ.

2. Составляется некоторая функция элементов выборки - статистика $Y(\tilde{\theta},\theta)$, связанная с параметром θ, такая, что ее распределение не зависит от θ и других неизвестных параметров.

3. Задается доверительная вероятность $(1-\alpha)$.

4. Зная распределение статистики Y, определяют два числа y_1 и y_2, удовлетворяющих условию $P(y_1<Y<y_2)=1-\alpha$.

5. Границы доверительного интервала для параметра θ определяются из решения относительно θ неравенства

$$y_1<Y(\tilde{\theta},\theta)<y_2.$$

Используя указанную схему, можно получить доверительные интервалы параметров m и σ нормального закона распределения.

16.2. Доверительный интервал для математического ожидания СВ X, распределенной по закону N(m, σ) при известном σ

Пусть СВ X имеет нормальное распределение $N(m,\sigma)$. Тогда доверительный интервал для параметра m по результатам выборки $x_1,x_2,...,x_n$ объемом n при условии, что дисперсия σ^2 известна, а доверительная вероятность равна 1-α, имеет вид

$$\bar{X}-\sigma/\sqrt{n}\cdot u_{1-\alpha/2};\ \bar{X}+\sigma/\sqrt{n}\cdot u_{1-\alpha/2}\quad. \tag{16.3}$$

Здесь $u_{1-\alpha/2}$, квантиль стандартизованного нормального распределения, определяется как решение уравнения

$$\Phi\left(u_{1-\alpha/2}\right)=1-\alpha/2 \Rightarrow 2/\sqrt{2\pi}\int_0^{u_{1-\alpha/2}} e^{-t^2/2}dt=1-\alpha .$$

Для $u_{1-\alpha/2}$ имеется таблица

$1-\alpha$	0,90	0,95	0,99	0,997	0,999
$u_{1-\alpha/2}$	1,64	1,96	2,58	3,00	3,37

Из анализа полученных соотношений можно сделать следующие *выводы*.

1. *Увеличение объема n выборки приводит к уменьшению длины доверительного интервала*.

2. *Увеличение доверительной вероятности (1-α) приводит к увеличению длины доверительного интервала, то есть к уменьшению точности δ.*

3. *Если задать точность δ, то есть предельную погрешность интервальной оценки, по формуле (2) и доверительную вероятность 1-α, то из соотношения* $\delta=\left(\sigma.\sqrt{n}\right)\cdot u_{1-\alpha/2}$ *можно найти минимальный объем выборки, который обеспечивает заданную точность:*

$$n=\sigma^2/\delta\cdot u_{1-\alpha/2}^2 .$$

16.3. Доверительный интервал для МО СВ Х, распределенной по нормальному закону при неизвестном σ

Если генеральная совокупность $\xi\in N(m,\sigma^2)$ и σ неизвестно, то с вероятностью $P=1-\alpha$

$$m\in\left(\overline{x}-\frac{s_0}{\sqrt{n}}t_{1-\alpha/2,n-1},\ \overline{x}+\frac{s_0}{\sqrt{n}}t_{1-\alpha/2,n-1}\right), \qquad (16.4)$$

$$\frac{(n-1)\tilde{s}^2}{\chi^2_{1-\alpha}(n-1)} \le \sigma^2 \le \frac{(n-1)\tilde{s}^2}{\chi^2_{\alpha}(n-1)},$$

где $s_0 = \frac{1}{n-1}\sum (x_i - \overline{x})^2$, $t_{1-\alpha/2,n-1}$ $\chi^2_{\alpha}(n-1)$ – квантиль распределения Стьюдента (Пирсона) с $(n-1)$ степенью свободы уровня α, n – объем выборки .

Для нахождения квантилей распределения Стьюдента t_p имеется таблица. Приведем ее для двух значений доверительной вероятности.

1-α \ n	5	10	20	30	∞
0,95	2,571	2,228	2,086	2,042	1,960
0,99	4,032	3,169	2,845	2,750	2,576

16.4. Доверительный интервал для σ² СВ X, распределенной по нормальному закону

Пусть СВ X имеет нормальное распределение N(m,σ), причем m и σ неизвестны. Тогда доверительный интервал для параметра σ² по выборке $x_1x_2,..,x_n$ объемом n с доверительной вероятностью 1-α имеет вид

$$\left(\frac{n-1\ s_0^2}{\chi^2_{1-\alpha/2,n-1}};\ \frac{n-1\ s_0^2}{\chi^2_{\alpha/2,n-1}}\right). \qquad (16.5)$$

Для квантилей $\chi^2_{p,n-1}$ имеются таблицы.

Замечание. Так как при $n\to\infty$ распределение χ^2_n приближается к нормальному, то при достаточно большом объеме выборки ($n\ge50$) доверительный интервал можно найти по формуле

$$\frac{s_0}{1+u_{1-\alpha/2}/\sqrt{2n}} < \sigma < \frac{s_0}{1-u_{1-\alpha/2}/\sqrt{2n}},$$

где $u_{1-\alpha/2}$ квантиль стандартизованного нормального распределения, соответствующая доверительной вероятности $1\text{-}\alpha$.

Пример 16.1. Измерения сопротивления резистора дали следующие результаты (в омах): $x_1 = 592$, $x_2 = 595$, $x_3 = 594$, $x_4 = 592$, $x_5 = 593$, $x_6 = 597$, $x_8 = 589$, $x_9 = 590$. Известно, что ошибки измерения имеют нормальный закон распределения. Систематическая ошибка отсутствует. Построить доверительный интервал для истинного сопротивления резистора с надежностью 0,99 в предположении: а) дисперсия ошибки измерения известна и равна 4; б) дисперсия ошибки измерения неизвестна.

Решение. В данной серии из девяти наблюдений

$$\bar{X} = \frac{592+595+\ldots+590}{9} = 593.$$

Если дисперсия ошибки измерения известна, то можно воспользоваться формулой (16.3). Для этого из таблицы функции Лапласа (см. приложение) находим, что $2\Phi\ 2,58\ = 0,99$, т.е. уровню надежности 0,99 соответствует значение $t_\gamma = 2,58$. Тогда по формуле (16.3)

$$593 - 2,58\frac{2}{\sqrt{9}} < M\ \ X\ \ < 593 + 2,58\frac{2}{\sqrt{9}}$$

или $591,28 < M\ \ X\ \ < 594,72$ с вероятностью 0,99.

В случае неизвестной дисперсии ее можно оценить на основе тех же опытных данных:

$$\sigma^2 \approx s^2 = \frac{(592-593)^2 + (595-593)^2 + \ldots + (590-593)^2}{8} = 6{,}5,$$

$s = \sqrt{6{,}5} \approx 2{,}55$. По таблице распределения Стьюдента (см. приложение, табл. 3) для $n-1=9-1=8$ степеней свободы и заданной вероятности $\gamma = 0{,}99$ находим $t_\gamma = 3{,}355$. Тогда по формуле (16.4)

$$593 - 3{,}355\frac{2{,}55}{\sqrt{9}} < M(X) < 593 + 3{,}355\frac{2{,}55}{\sqrt{9}}$$

или $590{,}15 < M(X) < 595{,}85$ с вероятностью $0{,}99$.

Пример 16.2. В таблице приведены сгруппированные данные измерений роста у 50 наугад выбранных студентов:

Рост	166 – 170	170 – 174	174 – 178	178 – 182	182 – 186	186 – 190
Число Студентов	3	7	15	13	11	1

Оценить средний рост и дисперсию роста студентов. Построить доверительный интервал для среднего роста студентов с надежностью 0,9.

Решение. Так как данные сгруппированы, то в качестве представителя каждого интервала можно взять середину этого интервала. Тогда

$$M(X) \approx \bar{X} = \frac{\sum_k \tilde{X}_k n_k}{n} =$$

$$= \frac{168\cdot 3 + 172\cdot 7 + 176\cdot 15 + 180\cdot 13 + 184\cdot 11 + 188\cdot 1}{50} = 178,$$

$$D(X) \approx s^2 = \frac{\sum_k (X_k - \overline{X})^2 \cdot n_k}{n-1} =$$

$$= \frac{(168-178)^2 \cdot 3 + (172-178)^2 \cdot 7 + \ldots + (188-178)^2 \cdot 1}{49} \approx 23{,}76,$$

Так как $2\Phi(1{,}65) = 0{,}9$, то по формуле (16.4) имеем

$$178 - 1{,}65\sqrt{\frac{23{,}67}{50}} < M(X) < 178 + 1{,}65\sqrt{\frac{23{,}76}{50}}$$

или $176{,}86 < M(X) < 179{,}14$ с вероятностью 0,9.►

Пример 16.3. По результатам девяти измерений емкости конденсатора получена оценка $\overline{X} = 20$ мкФ. Среднеквадратическая ошибка измерения известна и равна 0,04 мкФ. Построить доверительный интервал для емкости конденсатора с надежностью 0,95.

Решение. В предположении, что ошибки измерения имеют нормальный закон распределения можно воспользоваться формулой (16.3). Так как $2\Phi(1{,}96) = 0{,}95$, то

$$20 - 1{,}96\frac{0{,}04}{3} < M(X) < 20 + 1{,}96\frac{0{,}04}{3}$$

или $19{,}974 < M(X) < 20{,}026$ с вероятностью 0,95.

вероятности 1-α.

Примеры для самостоятельного решения

16.1. Стрелок 20 раз попал в цель при 100 выстрелах. Построить доверительный интервал для вероятности попадания в цель при одном выстреле для уровня надежности $\gamma = 0,90$.

16.2. Для проверки всхожести посеяли 900 семян. Из них проросло 810. Постройте доверительный интервал для доли всхожих семян с надежностью 0,95.

16.3. Для изучения общественного мнения было опрошено наугад 1600 жителей нашего города. Деятельность мэра города одобрили 1200 из них. Постройте с надежностью 0,95 доверительный интервал для доли жителей нашего города, одобряющих деятельность мэра.

16.4. По данным 16 наблюдений нормально распределенной случайной величины найдены ее среднее арифметическое $\bar{x} = 15,6$ и оценка среднего квадратического отклонения $s = 0,06$. Построить доверительный интервал для математического ожидания этой случайной величины при уровне надежности 0,95.

16.5. По результатам 10 измерений некоторой физической величины найдены среднее арифметическое результатов измерений $\bar{x} = 8,4$ и оценка среднего квадратического отклонения $s = 0,06$. Считая, что ошибки измерений имеют нормальный закон распределения, найдите интервальную оценку для измеряемой величины с вероятностью 0,95.

16.6. Из большой партии однотипных транзисторов наугад отобрали и проверили 100 штук. У 36 из них оказался коэффициент усиления меньше стандартного. Постройте 95%-й доверительный интервал для доли транзисторов с недостаточным коэффициентом усиления во всей партии.

Ответы:
16.1.(0,134;0,266); 16.2.(0,88;0,92); 16.3.От 73% до 77%; 16.4.(15,385;15,815);16.5.(8,383;8,437); 16.6.(0,26;0,45).

Занятие № 17. Проверка статистических гипотез

Статистическими гипотезами называют любые предложения относительно параметров (такие гипотезы называют параметрическими) или вида функции распределения случайной величины.

Наряду с выдвинутой гипотезой обычно рассматривают одну или несколько альтернативных (конкурирующих) гипотез. Если выдвинутая гипотеза будет отвергнута, то ее место занимает альтернативная.

Основную (выдвинутую) гипотезу называют *нулевой* и обозначают через H_0.

Альтернативную гипотезу обозначают через H_α.

Выбор альтернативной гипотезы определяется конкретной формулировкой задачи.

Пример 17.1. В теории надежности для многих классов изделий установлен экспоненциальный закон отказов: если X – время безотказной работы изделия, то

$$\mathrm{P}\left(\mathrm{X}>x\right)=\mathrm{e}^{-x/T},\quad x\geq 0,$$

где T – среднее время безотказной работы изделия. Допустим, что техническими условиями предусмотрено обеспечение безотказной работы с надежностью 1-α в течение времени T_0.

$$\mathrm{P}\left(\mathrm{X}>T_0\right)\geq 1-\alpha \Leftrightarrow \mathrm{e}^{-T_0/T}\geq 1-\alpha \quad \text{или} \quad T\geq \frac{T_0}{\ln\frac{1}{1-\alpha}}=T_1.$$

Тогда

$$\mathrm{H}_0 = \mathrm{F}_\chi \; x \; \in \; 1-\mathrm{e}^{-x/T}, \;\; T \geq T_1 \quad .$$

Альтернативная гипотеза

$$\mathrm{H}_\alpha = \mathrm{F}_\chi \; x \; \in \; 1-\mathrm{e}^{-x/T}, \;\; T < T_1 \quad .$$

Задача проверки статистической гипотезы H_0 относительно генеральной совокупности ξ ставится так:

найти правило, позволяющее по выборке обоснованно решить вопрос о принятии или отклонении гипотезы H_0.

Это правило называется статистическим критерием или просто критерием К проверки гипотезы.

Для решения этой задачи выбирают критерий проверки К, т.е. некоторую функцию (статистику) от выборки

$$z = z(x_1, \dots x_n) \qquad (17.1),$$

которая является случайной величиной, так как все x_i есть случайные величины. Предполагается, что для этой функции известны плотности распределения вероятностей $f_1(z/H_0)$ и $f_2(z/H_a)$, где H_a – альтернативная гипотеза. Зададим уровень значимости $\alpha = (0{,}1;\, 0{,}05;\, 0{,}01)$. Эта вероятность такова, что событиями, происходящими с такой вероятностью в данной ситуации можно пренебречь.

Критической областью G называют совокупность значений критерия z, при которой гипотезу H_0 отвергают.

Область G находят из условия

$$P \;\; z \in G/H_0 \;\; = \int_G f_1 \;\; z/H_0 \;\; dz = \alpha \qquad (17.2)$$

Отметим, что условием (17.2) область G определяется неоднозначно.

Основной принцип проверки статистической гипотезы состоит в следующем: по выборке и формуле (17.1) считают величину $z = z_{\text{набл.}}$

Если $z_{\text{набл.}} \in G$, то H_0 отвергают в пользу альтернативной гипотезы H_a. Если же $z_{\text{набл.}} \notin G$, то оснований отвергнуть H_0 нет, так как выборочные данные не противоречат гипотезе H_0.

Число $r = P\left(z \in G / H_a\right) = \int_G f_2\left(z / H_a\right) dz$ называют *мощностью критерия*.

$$\beta = 1 - r = P\left(z \in G / H_a\right) \qquad (17.3)$$

При принятии или отклонении гипотезы H_0 возможны ошибки двоякого рода: 1) ошибка первого рода – H_0 отвергают, а она верна.

Вероятность ошибки 1 рода $P\left(z \in G / H_0\right) = \alpha$;

2) ошибка второго рода – H_0 принимают, а она не верна.

Вероятность ошибки второго рода $P\left(z \in G / H_a\right) = \beta$.

Из формулы (17.3) видно, что чем больше мощность r, тем меньше ошибка 2 рода. Обычно поступают следующим образом: фиксируют уровень значимости α, т.е. фиксируют приемлемую вероятность ошибки 1 рода, а затем ищут критерий z_α с наибольшей мощностью, то есть с наименьшей ошибкой 2 рода.

Таким образом, проверка параметрической статистической гипотезы может быть разбита на следующие этапы:

1) формулируем гипотезы H_0 и H_a;

2) назначаем уровень значимости α;

3) выбираем статистику z (17.1) для проверки гипотезы H_0;

4) находим плотности распределения $f_1 \ z/H_0$ и $f_2 \ z/H_a$;

5) в зависимости от гипотезы H_a находим критическую область G;

6) по выборке вычисляем $z_{\text{набл.}} = z(x_1, x_2, \ldots, x_n)$;

7) принимаем решение: если $z_{\text{набл.}} \notin G$, гипотезу H_0 оставляем. Если $z_{\text{набл.}} \in G$, гипотезу H_0 отклоняем в пользу альтернативной H_a.

Модель 1. Пусть известно, что генеральная совокупность $\xi \in N(m, \sigma^2)$, σ – известно. Требуется по выборке и уровню значимости α проверить нулевую гипотезу $H_0 : m = m_0$.

Решение. Для этой модели статистика $z = \dfrac{\overline{x} - m_0}{\sigma}$, где $\overline{x} = \dfrac{1}{n}\sum x_i$.

Тогда для альтернативной гипотезы

$H_a: \ m > m_0$ критическая область $G = \ z > u_{1-\alpha}$,

$H_a: \ m < m_0$ $\qquad G = \ z < u_\alpha$,

$H_a: \ m \neq m_0$ $G = \ |z| > |u_{\alpha/2}|$.

Для простой альтернативной гипотезы

$H_a: \ m = m_1, m_1 < m_0$ мощность $r = F\left(u_\alpha + \dfrac{m_0 - m_1}{\sigma}\sqrt{n}\right)$;

Для гипотезы $H_a: \ m = m_1 \ m_1 > m_0$,

$$r = 1 - F\left(u_{1-\alpha} - \frac{a_1 - a_0}{\sigma_x}\sqrt{n} \right).$$

Здесь $F(x) = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{t^2/2}dt$, u_α – квантиль уровня α случайной величины $\eta \in N(0,1)$.

Модель 2. Пусть генеральная совокупность $\xi \in N(m,\sigma^2)$, но оба параметра неизвестны. По выборке найдем точечные оценки

$$\overline{x} = \frac{1}{n}\sum x_i \quad \text{и} \quad s_0^2 = \frac{1}{(n-1)}\sum_{i=1}^{n} x_i - \overline{x}^{\;\;2}$$

неизвестных параметров.

По уровню значимости α проверить нулевую гипотезу $H_0 : m = m_0$:

Решение. Для альтернативных гипотез $H_a: \;\; m > m_0$ критическая область $G = \; z > t_{1-\alpha}(n-1)$.

$H_a: \;\; m < m_0 \qquad G = \; z < t_\alpha(n-1)$.

$H_a: \;\; m \neq m_0 \;\; G = \;\; |z| > |t_{\alpha/2}(n-1)|$,

где $z = \sqrt{n}\,\frac{\overline{x} - m_0}{s_0}$, $\quad t_\alpha(n-1)$ – квантиль уровня α распределения Стьюдента с $(n-1)$ степенью свободы.

Для $H_a: \;\; m = m_1, \;\; m_1 < m_0$ мощность

$$r = T_{n-1}\left(t_\alpha + \frac{m_0 - m_1}{\overline{s}}\sqrt{n} \right),$$

где $T_{n-1}(t)$– функция распределения Стьюдента с $(n-1)$ степенью свободы.

Модель 3. Пусть имеем две независимые выборки $x_1, x_2, ... x_n$ и $y_1 \dots y_m$ объемом n и m из нормальных генеральных совокупностей $\xi \in N(m_1, \sigma_1^2)$ и $\eta \in N(m_2, \sigma_2^2)$. Предположим, что σ_1 и σ_2 известны. Требуется на уровне значимости α проверить гипотезу $H_0 : m_x = m_y$:

Решение. Для этой модели статистика $z = \dfrac{\overline{x} - \overline{y}}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}}$,

где $\overline{x} = \dfrac{1}{n}\sum x_i$, $\overline{y} = \dfrac{1}{m}\sum y_i$.

Для альтернативных гипотез $H_a: \ m_1 > m_2$ область $G = \ z > u_{1-\alpha}$, $H_a: \ m_1 \neq m_2$ область $G = \ |z| > |u_{\alpha/2}|$, $H_a: \ m_1 < m_2$ $G = \ z < u_\alpha$, где u_α квантиль уровня α случайной величины $\eta \in N(0,1)$.

Пример 17.2. По паспортным данным автомобильного двигателя расход топлива на 100 км пробега составляет 10 л. Ожидается, что после модернизации двигателя расход топлива уменьшится. Для проверки производятся испытания 25 случайно отобранных автомобилей с модернизированным двигателем. По результатам испытаний выборочная средняя расходов топлива на 100 км пробега составила $\overline{x} = 9,3$ л. Предполагая, что расход топлива есть нормальная случайная величина с $\sigma = 2,$ проверить гипотезу H_0 утверждающую, что изменение конструкции двигателя не повлияет на расход топлива при уровне значимости $\alpha = 0,05$.

Решение. Дано: $\bar{x} = 9,3$, $\sigma = 2$, $n = 25$, $\alpha = 0,05$, $H_0 : a_0 = 10$, $H_0 : a < 10$.

Вычислим статистику

$$z_{\text{набл.}} = \sqrt{n}\,\frac{\bar{x} - a_0}{\sigma} = \sqrt{25} \cdot \frac{9,3 - 10}{2} = -1,75.$$

Область $G = \{z < u_\alpha\}$, где u_α – квантиль уровня α случайной величины $\xi \in N(0,1)$. Имеем $u_{0,05} = -u_{0,95} = -1,65$, т.е. $G = z < -1,65$. Так как $z_{\text{набл.}} \in G$

$-1,75 < -1,65$, то гипотезу H_0 отвергаем в пользу альтернативной. То есть из опытных данных следует, что модернизация двигателя привела к уменьшению расхода топлива.

Замечание. Пусть в условиях задачи $H_a : a_1 = 9$. Вычислим мощность критерия r, вероятность ошибки второго рода $\beta = 1 - r$, и ответим на вопрос, какой минимальный объем выборки нужно взять, чтобы $\beta \le 0,05$.

Имеем

$$r = F\left(u_\alpha + \frac{a_0 - a_1}{\sigma}\sqrt{n}\right) = F\left(-1,65 + \frac{10 - 9}{2}\sqrt{25}\right) = F(0,85) =$$

$0,802$.

$\beta = 1 - r = 0,198$, где $F(x) = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-t^2/2}dt$.

$\beta \le 0,05 \Rightarrow r = 1 - \beta \ge 0,95$.

Решим уравнение $0,95 = F\left(u_\alpha + \frac{a_0 - a_1}{\sigma}\sqrt{n}\right)$,

$$u_\alpha + \frac{a_0 - a_1}{\sigma}\sqrt{n} = u_{0,95} = 1,65, \quad -1,65 + \frac{10-9}{2}\sqrt{n} = 1,65$$

$$\sqrt{n} \ge 6,6, \qquad n \ge 44.$$

Пример 17.3. Из продукции двух станков-автоматов, выпускающих однотипные изделия, взяты выборки объемов $n_1 = 15$ и $n_2 = 18$. По результатам выборок найдены $\overline{x}_1 = 32$ мм, $\overline{x}_2 = 35$ мм. Дисперсии генеральных совокупностей известны $\sigma_1^2 = 1,5$, $\sigma_2^2 = 2,1$. В предположении о нормальном законе распределения погрешностей изготовления требуется на уровне значимости $\alpha = 0,05$ проверить гипотезу $H_0 : m_1 = m_2$ при альтернативной гипотезе $H_a : m_1 \ne m_2$.

Решение. У нас $n_1 = 15$, $n_2 = 18$, $\overline{x}_1 = 32$, $\overline{x}_2 = 35$, $\sigma_1^2 = 1,5$, $\sigma_2^2 = 2,1$, $\alpha = 0,05$.

Статистика $u = \dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$. $u_{\text{набл.}} = \dfrac{32-35}{\sqrt{\dfrac{1,5}{15} + \dfrac{2,1}{18}}} = \dfrac{-3}{\sqrt{0,1+0,117}} =$

$$= \frac{-3}{0,465} = -6,45.$$

Критическая область G для альтернативной гипотезы $H_a : m_1 \ne m_2$ имеет вид

$$G = \{|u| > |u_{\alpha/2}|\}.$$

$$u_{\alpha/2} = u_{0,025} = -u_{0,975} = -1,96, \qquad G = \{|u| > 1,96\}.$$

Т.к. $u_{\text{набл.}} \in G$ $(6,45 > 1,96)$, то отклоняем гипотезу H_0 в пользу альтернативной H_a.

Задачи для самостоятельного решения.

17.1. По результатам 230 замеров, установлено, что среднее время изготовления детали $\overline{x} = 48$ сек. Предполагая, что время изготовления есть нормальная случайная величина с $\sigma = 1$ сек., необходимо:

1) проверить на уровне значимости $\alpha = 0,05$ гипотезу H_0: $m = 50$ сек. против альтернативной гипотезы H_a: $m = 45$ сек.

2) вычислить мощность критерия и вероятность ошибки второго рода;

3) проверить на уровне значимости $\alpha = 0,05$ гипотезу H_0: $m = 50$ сек. против альтернативной гипотезы H_a: $m \neq 50$ сек.

17.2. По данным 170 рейсов установлено, что в среднем машина затрачивает на поездку до хлебоприемного пункта $\overline{x} = 73$ мин. Допустив, что время поездки есть нормальная случайная величина на уровнях значимости $\alpha = 0,05$ и $\alpha_1 = 0,1$ проверить гипотезу H_0: $a = 75$ мин.; при альтернативной гипотезе H_a: $a = 72$ мин.

1) если известно, что $\sigma = 1$ мин.;

2) если выборочное среднее квадратичное отклонение $s = 1$ мин.;

3) для условий 1) и 2) вычислить мощность критерия.

17.3. Из продукции двух автоматических линий, обрабатывающих корпуса вентилей одного типоразмера, взяты выборки объемов $n_1 = 150$ и $n_2 = 90$. По результатам выборочных наблюдений найдено $\overline{x}_1 = 182$ мм, $\overline{x}_2 = 185$ мм. Предварительно установлено, что погрешности изготовления есть нормальные случайные величины с дисперсиями $\sigma_1^2 = 9$ мм^2, $\sigma_2^2 = 15$

мм^2. Требуется на уровне значимости $\alpha = 0{,}05$ проверить гипотезу $H_0 : m_1 = m_2$:

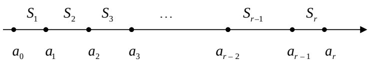
1) при альтернативной гипотезе $H_a :\ m_1 < m_2$;

2) при альтернативной гипотезе $H_a :\ m_1 \neq m_2$.

Занятие № 18. Проверка гипотезы о виде распределения случайной величины. Критерий χ^2

Пусть основная гипотеза H_0 состоит в том, что функция распределения случайной величины ξ есть функция $F(x)$, зависящая от ℓ неизвестных параметров. Наиболее часто применимым критерием проверки этой гипотезы является критерий, введенный К.Пирсоном. Его можно использовать для любых распределений, в том числе и многомерных.

Чтобы воспользоваться этим критерием, выборочные данные предварительно группируют следующим образом. Разбивают множество значений СВ X на r непересекающихся множеств S_i с помощью (r-1) чисел $a_0 < a_1 < a_2 < \ldots < a_r$:



Обозначим: $p_i = P(a_{i-1} < X < a_i) = F(a_i) - F(a_{i-1})$ - вероятность попадания X в интервал a_{i-1}, a_i в случае, когда предложенная гипотеза справедлива. Очевидно, что $\sum_{i=1}^{r} p_i = 1$. Пусть n_i, i=1,...,r - количество элементов выборки, попавших в интервал

a_{i-1}, a_i , $\sum_{i=1}^{r} n_i = 1$. Тогда n_i/n есть относительная частота попадания величины X в интервал $S_i = a_{i-1}, a_i$ при n наблюдениях. Очевидно, что $\sum_{i=1}^{r} n_i / n = 1$.

Для приведенного на рисунке разбиения p_i есть приращение гипотетической ФР F(x)на интервале S_i, а n_i/n - приращение эмпирической ФР F*(x) на том же интервале S_i. В качестве статистики принимают следующую величину:

$$\chi^2 = \sum_{i=1}^{r} \frac{n}{p_i}\left(\frac{n_i}{n} - p_i\right)^2 = \sum_{i=1}^{r} \frac{(n_i - np)^2}{np_i}, \qquad (18.1)$$

являющуюся мерой отклонения эмпирической ФР от теоретической, а критическую область задают в виде $V_k = \{Z \ge Z_\alpha\}$.

Таким образом, процедура применения критерия χ^2 для проверки гипотезы H_0 состоит из следующих этапов:

1. По выборке (2) найдем точечные оценки неизвестных параметров предполагаемого закона распределения $F(x)$.

2. Разобьем числовую ось на r промежутков $a_0, a_1 : a_1, a_2 , \ldots a_{r-1}, a_r$,
$a_0 = -\infty$, $a_r = +\infty$, $a_1 < a_2 < \ldots < a_{r-1}$.
Если гипотеза H_0 справедлива, то i – му промежутку a_{i-1}, a_i соответствует вероятность $p_i = F(a_i) - F(a_{i-1})$, $i = 1 \ldots r$.

3. Пусть из выборки (17.1) n_i значений попадает в i – ый промежуток a_{i-1}, a_i

$\sum n_i = n$. Тогда отношение n_i/n представляет собой частоту попадания выборочных значений в i – ый интервал. Близость частот $\frac{n_i}{n}$ к p_i свидетельствует в пользу гипотезы H_0.

4. Вычисляем выборочное значение статистики $\chi^2_{\text{набл.}} = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i}$, которая характеризует согласованность гипотезы H_0 с опытными данными.

5. Принимаем статистическое решение: гипотеза H_0 не противоречит опытным данным на заданном уровне значимости α, если $\chi^2_{\text{набл.}} < \chi^2_{1-\alpha}(r - \ell - 1)$;

если же $\chi^2_{\text{набл.}} \geq \chi^2_{1-\alpha}(r - \ell - 1)$, то гипотеза H_0 отклоняется. Здесь $\chi^2_{1-\alpha}$ – квантиль уровня $1-\alpha$ распределения Пирсона с $r-\ell-1$ степеней свободы, ℓ – число параметров распределения $F(x)$, которые оцениваются по выборке .

Замечание. Критерий χ^2 использует тот факт, что случайная величина $\frac{n_i - np_i}{\sqrt{np_i}}$

$i = 1,\ldots,r$, имеет распределение, близкое к нормальному $N(0,1)$. Чтобы это утверждение было достаточно точным, необходимо чтобы для всех интервалов выполнялось условие $np_i \geq 5$. Если в некоторых интервалах это условие не выполняется, то их следует объединить с соседними.

Задачи для самостоятельного решения

18.1. Используя критерий Пирсона, при уровне значимости $\alpha = 0,05$ установить, случайно или значимо расхождение между эмпирическими частотами n_i и теоретическими частотами n_i', которые вычислены исходя из гипотезы о нормальном распределении генеральной совокупности X, если

1.

n_i	5	10	20	8	7
n_i'	6	14	18	7	5

2.

n_i	6	8	13	15	20	16	10	7	5
n_i'	5	9	14	16	18	10	9	6	7

3.

n_i	14	18	32	70	20	36	10
n_i'	10	24	34	80	18	22	12

18.2. При уровне значимости $\alpha = 0,05$ проверить гипотезу о нормальном распределении генеральной совокупности, если известны эмпирические и теоретические частоты

1.

n_i	6	12	16	40	13	8	5
n_i'	4	11	15	43	15	6	6

2.

n_i	5	6	14	32	43	39	30	20	6	5
n_i'	4	7	12	29	48	35	34	18	7	6

3.

n_i	5	13	12	44	8	12	6
n_i'	2	20	12	35	15	10	6

18.3. Используя критерий Пирсона, при уровне значимости $\alpha = 0,05$ проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности X с заданным эмпирическим распределением, если

Номер интервала	1	2	3	4	5	6	7
Интервал	(-20;-10)	(-10;0)	(0;10)	(10;20)	(20;30)	(30;40)	(40;50)
Частот n_i	20	47	80	89	40	16	8

Ответы:

18.1.1. $\chi^2_{набл} = 2,47,\ \chi^2_{кр} = 6,0.$ 18.1.2 $\chi^2_{набл} = 1,52,\ \chi^2_{кр} = 12,6.$

18.1.3. $\chi^2_{набл} = 13,93,\ \chi^2_{кр} = 9,5.$ 18.2.1. $\chi^2_{набл} = 2,65,\ \chi^2_{кр} = 9,5.$

18.2.2. $\chi^2_{набл} = 3,\ \chi^2_{кр} = 14,1.$ 18.2.3. $\chi^2_{набл} = 13,0,\ \chi^2_{кр} = 9,5.$

18.3. $\bar{x} = 10,4,\ \sigma = 13,67,\ k = 4,\ \chi^2_{набл} = 1,52,\ \chi^2_{кр} = 9,5.$

Дополнение. Распределение χ^2

(Хи-квадрат с *n* степенями свободы)

Пусть $X_1, X_2, \ldots, X_n$ - независимые СВ, каждая из которых имеет стандартное нормальное распределение $X_i \in N(0,1)$. χ^2_n - *распределением* называется распределение СВ

$$\chi_n^2 = \mathrm{X}_1^2 + \mathrm{X}_2^2 + \ldots + \mathrm{X}_n^2 .$$

Плотность распределения χ^2 имеет вид

$$f_n(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} \mathrm{e}^{-x/2}, \qquad x > 0 .$$

Математическое ожидание и дисперсия определяются формулами: $\mathrm{M}(\chi_n^2) = n$, $\quad \mathrm{D}(\chi_n^2) = 2n$.

Индивидуальное домашнее задание

Хронометраж затрат времени на сборку узла машин у *n* слесарей дал следующее распределение (мин.)

а) Записать значения результатов эксперимента в виде вариационного ряда. Найти размах варьирования и разбить его на 5 интервалов. Построить гистограмму относительных частот.

б) Найти числовые характеристики выборки $\bar{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i$ и

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

где x_i' - середины интервалов ($x_i' = \frac{x_i + x_{i+1}}{2}$);

в) Определить доверительные интервалы для неизвестных математического ожидания m_x и среднего квадратического отклонения σ, отвечающие заданной доверительной вероятности $\gamma = 0,95$, в предположении, что выборка взята из нормальной генеральной совокупности;

1. Доверительный интервал для математического ожидания в случае нормального распределения

$\bar{x} - t_\gamma \dfrac{s}{\sqrt{n}} < m_x < \bar{x} + t_\gamma \dfrac{s}{\sqrt{n}}$, где n- объем выборки, $\bar{x}$ - выборочное среднее, $s = \sqrt{\dfrac{1}{n-1}\sum_{i=1}^{k} n_i (x_i' - \bar{x})^2}$ - исправленное среднее квадратическое отклонение выборки, γ - доверительная вероятность, значение параметра t_γ определяется из таблицы приложений по заданному уровню значимости $\alpha = 1-\gamma$ при числе степеней свободы $k = n-1$.

2. Доверительный интервал для среднего квадратического отклонения σ с заданной надежностью γ $s(1-q) < \sigma < s(1+q)$ при $q<1$ и $0 < \sigma < s(1+q)$ при $q>1$, где s - исправленное среднее квадратическое отклонение, параметр q находим из таблицы приложений.

г) Проверить гипотезу о нормальном законе распределения генеральной совокупности по критерию χ^2 при уровне значимости $\alpha = 0,05$.

Вычислить наблюдаемое значение критерия Пирсона

$$\chi^2_{набл} = \sum_{i=1}^{k} \frac{(n_i - n_i')^2}{n_i'},$$

где $n_i' = nP_i$, $P_i = \Phi(z_{i+1}) - \Phi(z_i)$, $z_i = \dfrac{x_i - \overline{x_в}}{\sigma_в}, z_{i+1} = \dfrac{x_{i+1} - \overline{x_в}}{\sigma_в}$.

$\Phi(x)$ - функция Лапласа, значения в таблице приложений . Для первого интервала левый конец устремляем в $-\infty$, для последнего интервала правый конец стремится к ∞. По таблице (приложений) критических точек распределения $\chi^2_{кр}$, уровню значимости $\alpha = 0,05$ и числу степеней свободы $k = l-3$, (l-

число интервалов) находим $\chi^2_{кр}$. Если $\chi^2_{набл} < \chi^2_{кр}$, то гипотеза H_0 о нормальном распределении генеральной совокупности принимается, если $\chi^2_{набл} > \chi^2_{кр}$, то гипотеза отвергается.

№ вар.	Выборка																
1	17	15	23	16	19	19	13	14	11	18	16	18	15	19	16	15	12
	18	19	10	8	19	16	12	12	17	17	11	16	18	22	14	18	10
	17	15	16	20	16	18	14	12	11	3	6	20	14	13	17	9	11
2	16	15	13	12	16	12	16	15	13	15	8	11	10	16	13	20	12
	14	7	5	16	9	13	9	8	14	13	14	15	19	11	14	15	13
	12	13	17	15	11	8	9	3	5	17	11	10	14	6	9	7	8
3	15	7	13	12	7	16	21	17	12	14	10	14	17	10	9	10	1
	1	12	2	15	4	9	15	18	16	9	12	14	11	14	16	16	15
	17	20	7	13	12	10	14	15	8	14	15	13	16	13	17	18	
4	14	13	10	11	4	9	12	9	11	1	5	11	14	5	12	8	10
	7	12	11	12	13	6	5	2	1	8	1	8	10	7	10	12	12
	11	13	16	3	13	8	6	11	5	14	12	8	10	10	12	11	12
5	17	19	25	12	10	21	18	21	15	16	20	13	20	18	21	17	20
	18	17	20	21	18	14	19	14	19	13	18	20	24	16	20	19	17
	22	18	20	18	16	14	13	5	22	18	15	19	14	11	12	10	15
6	14	16	9	22	15	13	7	12	13	10	17	15	17	18	14	15	18
	17	18	15	11	16	11	5	16	10	15	17	21	13	17	16	14	14
	15	17	13	11	10	2	19	18	13	12	16	8	11	13	18	15	19
7	18	11	17	16	24	17	20	14	15	15	12	19	17	19	20	16	20
	16	19	20	11	9	17	13	20	2	18	3	18	12	17	19	23	15
	16	17	17	21	16	19	15	13	12	4	7	15	21	11	13	9	17
8	20	18	16	17	20	15	19	18	17	19	12	15	14	19	16	22	14
	18	17	9	8	17	12	13	18	17	11	16	18	23	15	19	17	15
	16	17	21	11	12	10	18	14	15	20	4	16	18	15	13	12	7
	14	13	4														
9	15	17	24	17	19	14	11	9	20	19	15	16	20	15	19	17	15
	17	11	10	8	19	16	13	12	17	17	10	15	17	21	14	17	18
	9	16	21	16	19	12	15	12	11	4	5	19	14	13	18	10	12
	16	16	11														
10	15	16	11	9	11	10	14	10	11	17	5	17	13	16	11	18	12

	13	19	23	15	19	18	14	17	16	18	14	16	11	9	8	15	12
	15	20	14	18	15	17	8	14	16	7	20	13	16	9	16	9	13
11	14	17	6	16	20	14	17	11	16	21	9	16	12	8	1	8	14
	16	10	15	17	16	8	24	17	20	2	18	4	4	7	12	13	8
	7	11	19	13	12	17	19	20	7	17	2	10	12	15	19	28	20
12	23	19	16	9	14	13	8	13	15	9	8	11	10	11	13	6	18
	13	21	9	5	22	11	7	16	19	13	17	20	8	11	19	14	16
	11	14	18	7	8	18	21	8	16	15	19	15	18	10	15	13	17
	15	19	25	5	12	13											
13	20	21	17	19	23	15	12	21	11	16	11	12	21	17	9	15	13
	15	23	7	13	15	12	7	10	16	17	12	14	10	14	17	10	10
	12	2	1	15	4	15	9	16	18	9	12	14	16	16	10	9	17
14	16	13	18	13	12	15	9	14	11	18	16	6	7	5	16	13	10
	14	14	21	13	14	16	17	8	12	2	13	6	16	17	8	11	13
	16	20	9	8	15	9	12	10	1	11	12	16	10	11	15	7	13
	11	14															
15	17	26	18	37	12	34	30	27	12	28	14	17	25	32	23	28	30
	17	21	19	28	22	27	29	30	32	21	26	29	30	29	21	26	29
	23	27	30	18	21	29	24	26	21	15	24	22	18	28	31	18	26
16	9	12	26	24	27	16	20	17	15	23	19	20	17	22	18	18	23
	23	24	27	22	28	20	24	23	21	21	22	26	22	24	20	18	17
	21	16	23	14	29	22	25	19	17	20	21	24	22	24	25	21	21
	25																
17	25	14	14	17	16	24	23	20	19	20	17	18	23	21	9	11	13
	26	25	14	14	17	16	24	23	20	19	20	20	13	23	21	9	11
	23	26	24	22	17	18	22	23	21	16	23	28	18	24	21	19	20
18	7	10	5	1	6	14	13	11	10	11	5	12	10	14	8	13	6
	11	8	12	3	10	11	9	6	4	5	5	6	0	14	13	17	11
	11	9	17	12	11	13	6	4	3	10	2	7	6	18	9	7	12
	13																
19	10	8	2	8	7	10	1	13	9	7	6	4	16	2	1	6	11
	9	17	10	11	13	11	5	4	0	8	12	10	11	13	4	12	5
	10	13	12	9	14	9	10	4	10	0	10	5	4	11	1	9	6
	5	12															
20	12	9	15	4	8	6	6	9	13	2	3	6	8	4	1	5	4
	8	1	13	8	11	4	9	7	9	1	7	4	13	7	9	2	7
	10	1	4	8	6	4	2	1	2	11	7	4	7	7	7	1	10

	8																
21	7	13	5	5	9	4	12	6	8	5	0	6	8	0	2	4	2
	4	7	1	7	6	8	3	4	8	9	5	9	2	4	2	9	2
	10	3	7	6	7	3	6	1	10	6	7	9	6	5	9	1	7
	9	2	9														
22	14	17	6	16	20	14	17	11	16	21	9	16	12	8	1	8	14
	16	10	15	17	16	8	24	17	20	2	18	4	4	7	12	13	8
	7	11	19	13	12	17	19	20	7	17	2	10	12	15	19	28	20
23	1	0	7	5	2	0	1	8	5	0	1	0	8	7	8	0	2
	9	0	6	0	8	3	9	1	–5	6	3	5	2	0	7	27	4
	9	4	1	8	6	4	3	2	4	9	2	–2	8	0	7	–1	
24	9	5	7	6	7	6	4	1	4	0	5	6	3	6	–3	8	6
	1	2	1	4	0	5	12	6	1	2	1	11	7	–2	–4	7	5
	0	5	4	7	9	6	7	8	6	3	6	–1	3	4	10	8	0
	6																
25	5	8	4	7	9	2	13	–2	9	6	3	8	1	6	9	8	5
	2	7	2	7	1	6	8	0	8	7	9	4	1	8	7	5	6
	6	-1	2	0	3	6	1	12	4	7	5	10	8	12	1	10	–4
	–4	7															
26	1	8	1	–1	6	7	3	10	6	10	8	7	5	2	9	7	3
	1	9	10	5	9	7	6	4	7	10	7	4	8	8	5	6	2
	7	9	1	3	7	6	10	7	6	8	5	4	–1	3	-3	2	7
	5																
27	5	4	8	6	-1	1	5	3	12	5	5	5	2	2	11	0	7
	4	5	8	8	1	6	1	3	1	8	0	9	3	6	1	3	5
	3	4	6	4	9	5	4	1	0	6	9	3	2	3	–1	7	3
	6																
28	15	16	11	9	11	10	14	10	11	17	5	17	13	16	11	18	12
	13	19	23	15	19	18	14	17	16	18	14	16	11	9	8	15	12
	15	20	14	18	15	17	8	14	16	7	20	13	16	9	16	9	13
29	4	9	5	6	8	6	4	5	7	–3	5	6	3	8	4	6	0
	6	1	1	7	–1	0	1	–2	–1	6	1	6	10	5	1	10	2
	4	3	5	13	1	3	5	4	-1	8	7	6	3	2	6	0	8
	4	6															
30	9	1	5	0	5	3	7	2	4	8	2	2	4	1	9	2	5
	7	2	7	1	2	7	5	6	0	9	1	8	1	9	8	3	9
	7	5	6	4	7	2	6	8	7	8	3	1	7	4	3	9	2

	8																

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Математическая статистика: Учеб. Для вузов / В.Б. Горяинов, И.В. Павлов, Г.М. Цветкова и др.; Под ред. В.С. Зарубина, А.П. Крищенко.-М.: Изд-во МГТУ им. Н.Э. Баумана, 2001.-424 с.

2. Петрушко И.М. Курс высшей математики. Теория вероятностей: Лекции и практические занятия/ И.М. Петрушко, В.И. Афанасьев, А.А, Furthermore Бободжанов, В.Г. Крупин.-М.: Издательство МЭИ, 2004.-304 с.

3. Дубровская А.П. Курс теории вероятностей и математическая статистика: Учеб. пособие/ А.П. Дубровская, Е.Г. Глушко.- Воронеж: ВГТУ, 2004. 161 с.

4. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике/ В.Е. Гмурман. - М: Высш. шк., 1997.-400 с.

5. Сборник задач по математике для втузов. Ч.3. Теория вероятностей и математическая статистика: Учеб. Пособ./ Под ред. А.В. Ефимова. М.: Наука, 1990. 428 с.

6. Рябушко А.П. Индивидуальные задания по высшей математике: Операционное исчисление. Элементы теории устойчивости. Теория вероятностей. Математическая статистика: учеб. пособие/ А.П. Рябушко.-Мн.: Высш. шк., 2006.-336 с.

СОДЕРЖАНИЕ

МЕТОДИЧЕСКИЕ УКАЗАНИЯ

для организации самостоятельной работы
по изучению дисциплины «Теория вероятностей и
математическая статистика»
для студентов направления подготовки бакалавров 230100
«Информатика и вычислительная техника»
(профиль «Системы автоматизированного проектирования»)
очной формы обучения
Часть 3
Составители:
Глушко Елена Георгиевна
Дубровская Алевтина Петровна
В авторской редакции
Компьютерный набор Е.Г. Глушко
Подписано к изданию 26. 06. 2013.
Уч.- изд. л. 3,4.
ФГБОУ ВПО «Воронежский государственный
технический университет»
394026 Воронеж, Московский просп., 14