МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ВЫПОЛНЕНИЮ РАСЧЕТНО-ГРАФИЧЕСКОГО ЗАДАНИЯ

дисциплина «Математика»

раздел «Математическая статистика»

Постановка задачи

Произведена выборка объема n из некоторой генеральной совокупности. Получены статистические данные, содержащие пары значений величин X и Y.

Требуется:

1 часть.

1) построить эмпирическую функцию распределения, полигон, гистограмму для случайной величины X;
2) построить точечные и интервальные оценки для мат. ожидания и дисперсии генеральной совокупности X;
3) сделать проверку статистической гипотезы о законе распределения случайной величины X;

2 часть.

1) нанести на координатную плоскость данные выборки $(x;y)$ и по виду корреляционного облака подобрать вид функции регрессии;
2) составить корреляционную таблицу по сгруппированным данным;
3) вычислить коэффициент корреляции;
4) получить уравнение регрессии;

Сформулировать выводы по результатам работы.

<u>Образец выполнения работы</u>

Для представленного примера получена выборка:

Таблица 1 «Выборочные данные X и Y»

N	106	493	66	201	274	158	223	336	362	162	96	20
X	162	166	172	169	176	167	167	168	167	169	167	69
Y	100	84	82	91	86	90	92	88	89	88	89	83

N	288	251	257	152	279	478	86	439	368	203	271	395
X	169	163	164	164	164	178	176	167	165	172	168	170
Y	91	92	84	89	85	91	82	85	90	87	88	88

N	396	94	305	341	12	128	492	407	172	87	441	29
X	187	165	171	171	169	163	161	175	172	163	180	172
Y	86	87	94	91	79	80	88	95	89	91	98	90

N	140	59	70	453	487	447	105	232	95	456	80	225
X	174	164	169	157	178	176	161	176	165	161	182	176
Y	97	89	88	90	90	93	94	90	87	84	90	93

N	147	101	373	51	343	355	195	463	260	183	326	282
X	168	164	160	178	170	168	173	176	170	163	165	165
Y	93	91	83	89	90	81	89	95	81	93	84	88

N	139	483	399	467	266	372	356	290	241	273	450	329
X	170	166	165	181	172	165	172	178	173	165	174	159
Y	86	84	85	92	88	91	98	90	90	87	96	81

N	469	423	242	475	168	365	107	428	367	457	224	199
X	171	169	169	170	170	165	190	175	157	148	172	159
Y	92	92	87	91	88	94	105	91	82	87	99	83

N	404	363	192	109	429	60	13	291	400	337	100	187
X	162	167	167	160	175	163	164	180	164	169	169	170
Y	92	85	88	87	90	91	89	85	84	87	91	93

N	88	292	283	52	45	358	252	62	130	286	361	184
X	179	167	162	169	172	166	164	173	161	159	166	158
Y	99	81	80	91	99	82	84	84	82	86	84	91

N	79	371	378	419	307	56	374	169	43	298	239	145
X	163	165	170	172	161	171	166	164	183	173	166	167
Y	88	87	91	94	84	97	87	97	90	90	89	85

N	325	65	153	375	9	340	142	193	261	116	26	253
X	162	156	167	168	170	171	174	179	161	170	172	166
Y	89	88	86	92	90	91	90	85	79	95	91	88

N	61	202	440	21	200	221	332	275	287	108	468	103
X	173	172	179	155	175	173	170	171	171	167	165	173
Y	89	96	85	86	89	96	96	83	90	91	91	90

N	240	110	424	414	296	284	83	435	81	54	397	134
X	167	165	169	171	181	164	164	176	163	165	174	177
Y	89	94	82	89	89	86	91	87	88	93	86	87

N	303	430	34	144	277	451	179	472	342	293	327	448
X	180	170	168	175	171	170	168	160	169	164	171	164
Y	90	91	82	85	89	90	87	85	91	87	91	83

N	154	438	297	219	196	204	230	258	262	213	89	357
X	164	163	170	174	161	167	173	164	174	168	176	156
Y	83	88	92	88	91	91	87	90	91	83	93	85

N	426	480	156	127	295	115	36	7	473	376	157	254
X	162	168	176	184	165	176	163	167	169	186	172	175
Y	90	93	88	98	94	92	89	88	89	92	91	90

N	98	126	265	443	82	110	432	479
X	170	173	160	171	169	165	185	168
Y	90	91	89	85	87	94	91	90

1.Составим ранжированный (по увеличению) ряд для случайной величины X.

Таблица 2 «Ранжированный ряд случайной величины X»

X	148	155	156	156	157	157	158	159	159	159	160	160
Y	87	86	85	88	82	90	91	81	83	86	83	85

X	160	161	161	161	161	161	161	162	162	162	162	162
Y	87	79	82	84	84	88	91	80	89	90	92	94

X	162	163	163	163	163	163	163	163	163	163	164	164
Y	100	80	88	88	88	89	91	91	92	93	83	83

X	164	164	164	164	164	164	164	164	164	164	164	164
Y	84	84	84	85	86	87	89	89	89	90	90	91

X	164	164	165	165	165	165	165	165	165	165	165	165
Y	91	97	84	85	87	87	87	87	88	90	91	91

X	165	165	165	165	165	166	166	166	166	166	166	166
Y	93	94	94	94	94	82	84	84	84	87	88	89

X	166	167	167	167	167	167	167	167	167	167	167	167
Y	89	81	85	85	85	86	88	88	89	89	89	90

X	167	167	167	168	168	168	168	168	168	168	168	168
Y	91	91	92	81	82	83	87	88	88	90	92	93

X	168	169	169	169	169	169	169	169	169	169	169	169
Y	93	79	83	87	87	87	88	88	89	91	91	91

X	169	169	169	169	170	170	170	170	170	170	170	170
Y	91	91	92	92	81	86	88	88	90	90	90	90

X	170	170	170	170	170	170	170	171	171	171	171	171
Y	91	91	91	92	93	95	96	83	85	89	89	90

X	171	171	171	171	171	171	172	172	172	172	172	172
Y	91	91	91	92	94	97	82	87	88	89	90	91

X	172	172	172	172	172	172	173	173	173	173	173	173
Y	91	94	96	98	99	99	84	87	89	89	90	90

X	173	173	173	174	174	174	174	174	174	175	175	175
Y	90	91	96	86	88	90	91	96	97	85	89	90

X	175	175	175	176	176	176	176	176	176	176	176	176
Y	90	91	95	82	86	87	88	90	92	93	93	93

X	176	177	178	178	178	178	179	179	179	180	180	180
Y	95	87	89	90	90	91	85	85	99	85	90	98

X	181	181	182	183	184	185	186	187	190
Y	89	92	90	90	98	91	92	86	105

2. Составим новую таблицу, в которой отразим частоты n_i *появления случайных величин* X_i *и относительные частоты* $p_i = \dfrac{n_i}{n}$.

Таблица 3 «Дискретный вариационный ряд»

i	1	2	3	4	5	6	7	8	9	10	11	12
X_i	148	155	156	157	158	159	160	161	162	163	164	165
n_i	1	1	2	2	1	3	3	6	6	9	15	15
p_i	$\frac{1}{200}$	$\frac{1}{200}$	$\frac{2}{200}$	$\frac{2}{200}$	$\frac{1}{200}$	$\frac{3}{200}$	$\frac{3}{200}$	$\frac{6}{200}$	$\frac{6}{200}$	$\frac{9}{200}$	$\frac{15}{200}$	$\frac{15}{200}$

i	13	14	15	16	17	18	19	20	21	22	23	24
X_i	166	167	168	169	170	171	172	173	174	175	176	177
n_i	8	14	10	15	15	11	12	9	6	6	10	1
p_i	$\frac{8}{200}$	$\frac{14}{200}$	$\frac{10}{200}$	$\frac{15}{200}$	$\frac{15}{200}$	$\frac{11}{200}$	$\frac{12}{200}$	$\frac{9}{200}$	$\frac{6}{200}$	$\frac{6}{200}$	$\frac{10}{200}$	$\frac{1}{200}$

i	25	26	27	28	29	30	31	32	33	34	35
X_i	178	179	180	181	182	183	184	185	186	187	190
n_i	4	3	3	2	1	1	1	1	1	1	1
p_i	$\frac{4}{200}$	$\frac{3}{200}$	$\frac{3}{200}$	$\frac{2}{200}$	$\frac{1}{200}$	$\frac{1}{200}$	$\frac{1}{200}$	$\frac{1}{200}$	$\frac{1}{200}$	$\frac{1}{200}$	$\frac{1}{200}$

3. Составим интервальный вариационный ряд

В данном примере случайные величины сплошь заполняют промежуток (148;190). Число возможных значений велико. Их нельзя представить в виде случайных величин, принимающих отдельные, изолированные значения, тем самым отделить одно возможное значение от другого промежутком, не содержащим возможных значений случайной величины. Поэтому для построения вариационного ряда будем использовать интервальный ряд распределения. Весь возможный интервал варьирования разобьём на конечное число интервалов и подсчитаем частоту попадания значений величины в каждый интервал. Минимальное и максимальное значения случайной величины: $x_{\min} = 148, x_{\max} = 190$ Тогда интервал варьирования R («размах») будет равен R=$x_{\max} - x_{\min} = 42.$ Длину интервала рассчитывают по формуле:

$$h = \frac{x_{\max} - x_{\min}}{1 + 3{,}2 \lg n} \qquad (1)$$

При этом значение признака, находящегося на границе интервалов относят к правой границе интервала.

На практике считают, что правильно составленный ряд распределения содержит от 6 до 15 частичных интервалов. Для данного примера $h = \frac{190 - 148}{1 + 3{,}28 * \ln 200} = 2{,}285$, округлим до 3, т.е. размер интервала *h*=3, а число интервалов будет равно 14. Соответствующий интервальный вариационный ряд приведён в таблице №4.

Таблица 4 «Интервальный вариационный ряд»

Индекс интервала i	Интервалы $x_i < X \le x_{i+1}$	Частота n_i	Относительная частота $p_i^* = \frac{n_i}{n}$
1	148-151	1	1/200
2	151-154	0	0
3	154-157	5	5/200
4	157-160	7	7/200
5	160-163	21	21/200
6	163-166	38	38/200
7	166-169	39	39/200
8	169-172	38	38/200
9	172-175	21	21/200
10	175-178	15	15/200
11	178-181	8	8/200
12	181-184	3	3/200
13	184-187	3	3/200
14	187-190	1	1/200

$$\sum \frac{200}{200} = 1$$

4. Расчёт эмпирической функции распределения

После составления вариационного ряда необходимо построить функцию распределения выборки или эмпирическую функцию $F^*(x) = \frac{n_x}{n}$, то есть функцию найденную опытным путём. Здесь n_x – относительная частота события X< x, n - общее число значений.

Эмпирическое распределение можно изобразить в виде полигона, гистограммы или ступенчатой кривой.

Построим выборочную функцию распределения. Очевидно, что для $x \in (-\infty, 148]$ функция $F^*(x) = 0$, так как $n_x = 0$. На концах интервалов значения функции $F^*(x)$ рассчитаем в виде «нарастающей относительной частоты» (таблица 5).

Таблица 5 «Расчёт эмпирической функции распределения»

Индекс интервала i	$F^*(x)$
1	1/200
2	1/200
3	1/200+5/200=6/200
4	6/200+7/200=13/200
5	13/200+21/200=34/200
6	34/200+38/200=72/200
7	72/200+39/200=111/200
8	111/200+38/200=149/200
9	149/200+21/200=170/200
10	170/200+15/200=185/200
11	185/200+8/200=193/200
12	193/200+3/200=196/200
13	196/200+3/200=199/200
14	199/200+1/200=200/200

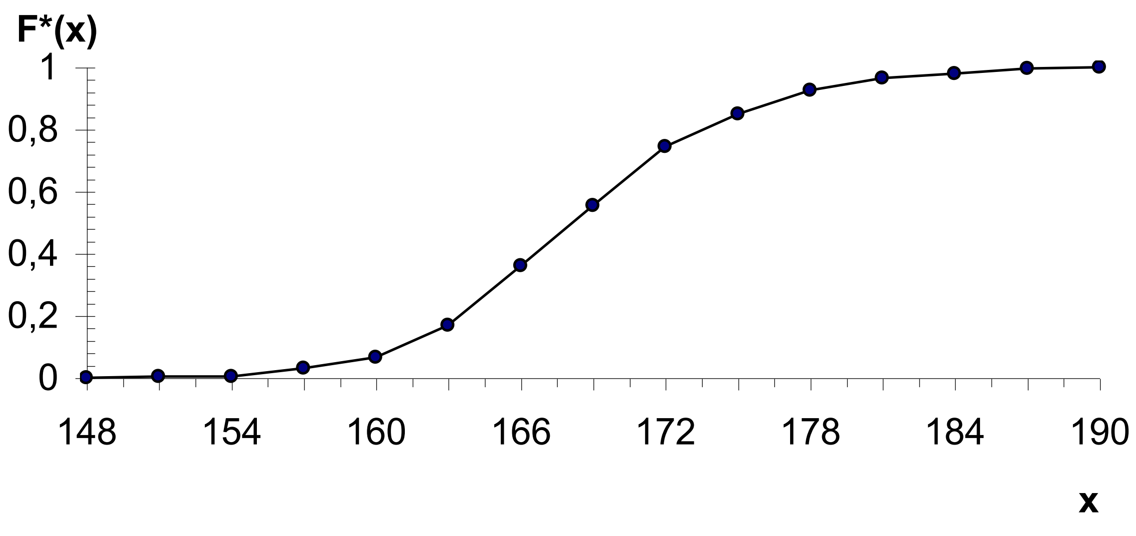

Рис.1

5. Расчёт выравнивающих частот

Полученные данные, представим в виде дискретного вариационного ряда (таблица 6) и изобразим графически в виде ломаной линии (полигона), связывающей на плоскости точки с координатами $(X_i; p_i^*)$, где X_i - среднее значение интервала $x_i < \mathrm{X} \le x_{i+1}$, а p_i^* - относительная частота.(таблица 6 и рис.2).

Таблица 6 “Дискретный вариационный ряд”

Номер интервала i	Среднее значение интервала X_i	Относительная частота p_i^*	Выборочная оценка плотности вероятности $\frac{n_i}{h \cdot n}$
1	149,5	0,005	0,002
2	152,5	0	0
3	155,5	0,025	0,008
4	158,5	0,035	0,012
5	161,5	0,105	0,035
6	164,5	0,19	0,063
7	167,5	0,195	0,065
8	170,5	0,19	0,063
9	173,5	0,105	0,035
10	176,5	0,075	0,025
11	179,5	0,04	0,013
12	182,5	0,015	0,005
13	185,5	0,015	0,005
14	188,5	0,005	0,002

На основании полученных выборочных данных необходимо сделать предположение, что изучаемая величина распределена по некоторому определённому закону. Для того чтобы проверить, согласуется ли это предположение с данными наблюдений, вычисляют частоты полученных в наблюдениях значений, т.е. находят теоретически сколько раз величина *X* должна была принять каждое из наблюдавшихся значений, если она распределена по предполагаемому закону. Для этого находят выравнивающие (теоретические) частоты по формуле:

$$n_i' = np_i, \qquad (2)$$

где n – число испытаний,

p_i - вероятность наблюдаемого значения x_i, вычисленная при допущении, что X имеет предполагаемое распределение.

Эмпирические (полученные из таблицы) и выравнивающие частоты сравнивают, и при небольшом расхождении данных делают заключение о выбранном законе распределения.

Предположим, что случайная величина X распределена нормально. В этом случае выравнивающие частоты находят по формуле:

$$n'_i = \frac{nh}{\sigma_B}\varphi(u_i), \qquad (3)$$

где n-число испытаний,

h-длина частичного интервала,

σ_B-выборочное среднее квадратичное отклонение,

$u_i = \dfrac{x_i - \bar{x}_B}{\sigma_B}$ (x_i - середина i – го частичного интервала)

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \quad \text{– функция Лапласа} \qquad (4)$$

Результаты вычислений отобразим в таблице №7.

Таблица 7 «Расчёт выравнивающих частот»

x_i	$x_i - \bar{x}_B$	$u_i = \dfrac{x_i - \bar{x}_B}{\sigma_B}$	$\varphi(u_i)$	$n'_i = \dfrac{nh}{\sigma_B}\varphi(u_i)$	n'_i	$p_i^{*\prime}$
149,5	-19,5	-3	0,004	0,42	1	0,05
152,5	-16,5	-2,53	0,02	1,55	2	0,01
155,5	-13,5	-2,06	0,048	4,54	5	0,025
158,5	-10,5	-1,59	0,11	10,68	11	0,055
161,5	-7,05	-1,11	0,22	20,37	20	0,1
164,5	-4,05	-0,64	0,33	31,0	31	0,155
167,5	-1,05	-0,17	0,396	37,48	37	0,185
170,5	1,95	0,31	0,38	36,0	36	0,18
173,5	4,95	0,78	0,3	28,0	28	0,14
176,5	7,95	1,25	0,18	17,34	17	0,085
179,5	10,95	1,73	0,09	8,44	8	0,04
182,5	13,95	2,2	0,04	3,37	3	0,015
185,5	16,95	2,67	0,011	1,06	1	0,005
188,5	19,95	3,15	0,003	0,26	0	0

$$\sum n'_i = 200$$

На этом же рисунке 2 отобразим пунктирной линией выравнивающие (теоретические) частоты из таблицы 7. Сравнение графиков (рис.2) наглядно показывает близость выравнивающих частот к наблюдавшимся и подтверждает правильность допущения о том, что обследуемый признак распределён нормально.

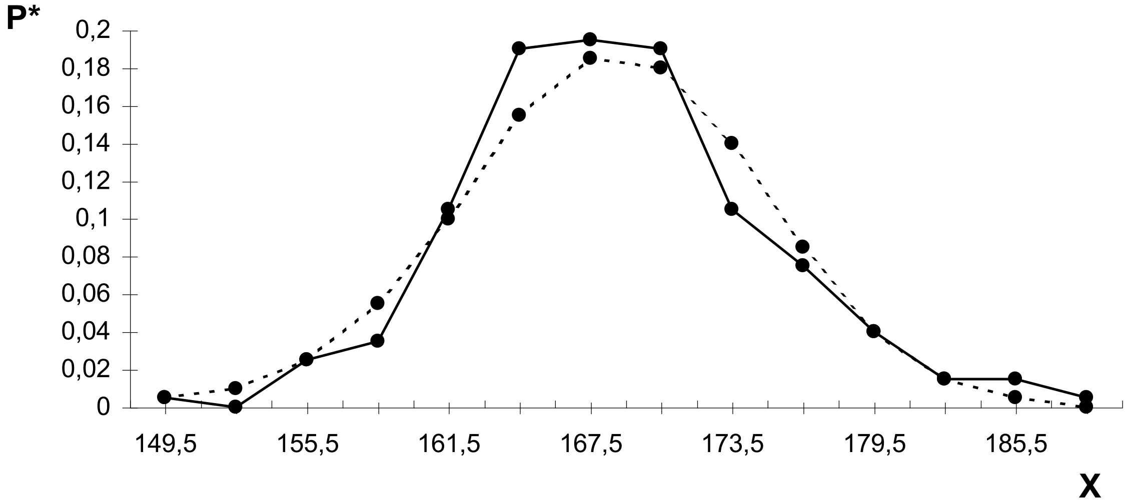

Рис.2

Интервальный вариационный ряд графически изобразим в виде гистограммы (рис.3). На оси X отложим интервалы длиной h=3, а на оси Y значения $\frac{n_i}{h \cdot n}$,расчёт которых представлен в таблице №6. Площадь под гистограммой равна сумме всех относительных частот, т.е. единице.

Графическое изображение вариационных рядов в виде полигона и гистограммы позволяет получать первоначальное представление о закономерностях, имеющих место в совокупности наблюдений.

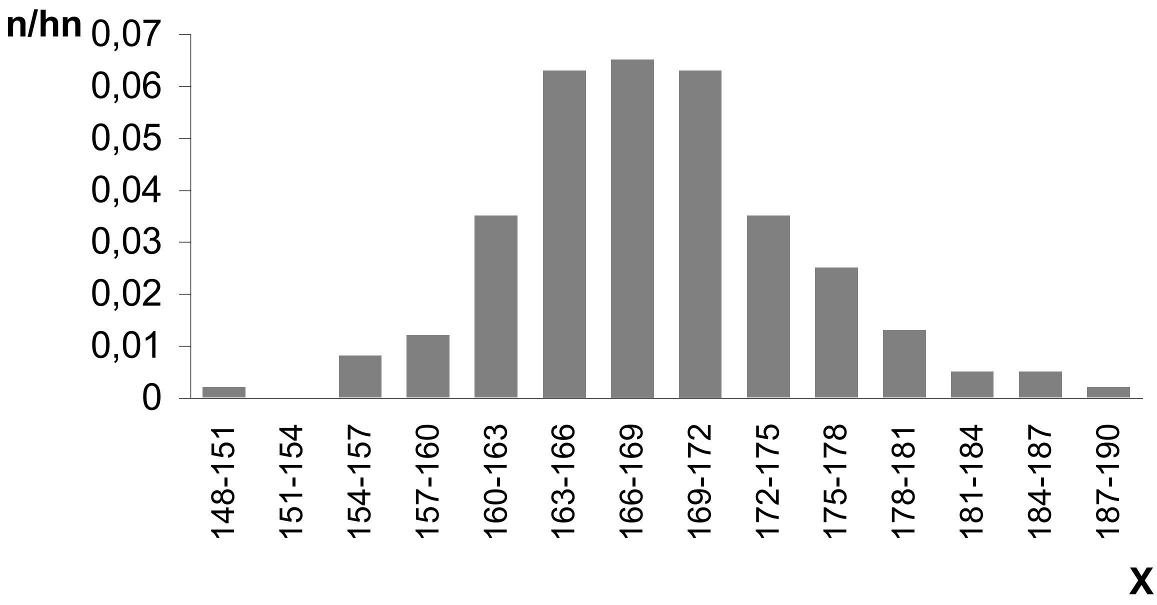

Рис.3

6. Найдём числовые характеристики вариационного ряда, используя таблицу №3.

Выборочная средняя ($\bar{x}_B$):

$$\bar{x}_B = \frac{x_1 + x_2 + \ldots . + x_k}{n}$$

$$\text{или} \quad \bar{x}_B = \sum_{i=1}^{k} \frac{n_i x_i}{n}, \qquad (5)$$

где n_1, n_2, n_3- частоты,

а $n_1 + n_2 + \ldots + n_k = n$-объём выборки. Выборочная средняя является оценкой математического ожидания (среднего значения теоретического закона распределения).

В некоторых случаях $\bar{x}_B$ удобнее рассчитать с помощью условных вариант. В нашем случае варианты x_i - большие числа, поэтому используем разность:

$$u_i = x_i - C, \qquad (6)$$

где C – произвольно выбранное число (ложный нуль). В этом случае

$$\bar{x}_B = C + \sum_{i=1}^{k} \frac{n_i u_i}{n}. \qquad (7)$$

Для изменения значения варианты можно ввести также условные варианты путём использования масштабного множителя:

$$\mathrm{u_i} = \mathrm{C_0 x_i}, \qquad (8)$$

где $\mathrm{C_0} = 10^{\mathrm{b}}$ (*b* выбирается положительным или отрицательным числом).

$$\bar{x}_B = 170{,}5 + \frac{1}{200}(-21*1 - 18*0 - ... + 18*1) = 170{,}5 - 1{,}95 = 168{,}55.$$

Здесь С – середина 8-го интервала.

Выборочная дисперсия (d_B):

$$\bar{d}_B = \sum_{i=1}^{k} \frac{n_i (x_i - \bar{x}_B)^2}{n} \qquad (9)$$

d_B также может быть рассчитана с помощью условных вариант:

$$d_B = d_{BU} = \sum_{i=1}^{k} \frac{n_i u_i^2}{n} - (\sum_{i=1}^{k} \frac{n_i u_i}{n})^2 \qquad (10)$$

$d_B = \frac{1}{200}$(1*441+0*324+…+1*324)- 1,95²=40,21

Среднеквадратическое отклонение:

$$\bar{\sigma}_B = \sqrt{\bar{d}}_B \qquad (11)$$

$\bar{\sigma}_B = \sqrt{40{,}21}$=6,34

Найдем несмещённую оценку дисперсии и среднеквадратического отклонения («исправленную» выборочную дисперсию и среднеквадратическое отклонение) по формулам:

$$S^2 = \frac{n}{n-1} \bar{d}_B \text{ и } S = \sqrt{\frac{n}{n-1}} \bar{\sigma}_B \qquad (12)$$

$S^2 = \frac{200}{199} \cdot 40{,}21$=40,41 и $S = \sqrt{\frac{200}{199}} \cdot 6{,}34$=6,36

Доверительный интервал для оценки математического ожидания с надёжностью $\gamma = 0{,}95$ определяют по формуле:

$$P(\bar{x}_B - t\frac{\bar{\sigma}_B}{\sqrt{n}} < a < \bar{x}_B + t\frac{\bar{\sigma}_B}{\sqrt{n}}) = 2\,\Phi(t) = \gamma \qquad (13)$$

Из соотношения $\Phi(z)=\gamma/2$ вычисляют значение функции Лапласа: $\Phi(z)$=0,475. По таблице значений функции Лапласа (Приложение 1) находят z=1,96. Таким образом,

$$168{,}55\text{-}1{,}96\frac{6{,}34}{\sqrt{200}} < a < 168{,}5 + 1{,}96\frac{6{,}34}{\sqrt{200}},$$

$$167{,}67 < a < 169{,}43.$$

Доверительный интервал для оценки среднего квадратичного отклонения случайной величины находят по формуле:

$$\frac{S}{1+q} < \sigma_x < \frac{S}{1-q}, \qquad (14)$$

где S – несмещённое значение выборочного среднего квадратичного отклонения;

q – параметр, который находится по таблице (Приложение 2) на основе известного объёма выборки n и заданной надёжности оценки γ.

На основании данных значений γ=0,95 и n=200 по таблице (Приложение В) можно найти значение q=0,099. Таким образом,

$$\frac{6{,}34}{1+0{,}099} < \sigma_x < \frac{6{,}34}{1-0{,}099},$$

$$5{,}79 < \sigma_x < 7{,}06$$

$$V = \frac{\sigma_B(x)}{\bar{x}_B} 100\% \qquad (15)$$

7. *Проведём проверку статистической гипотезы о нормальном распределении.*

Нормальный закон распределения имеет два параметра (r=2): математическое ожидание и среднее квадратическое отклонение. По выборочным данным (таблицы 4 и 6) полученные оценки параметров нормального распределения, вычисленные выше:

$$\bar{x}_B = 168{,}55,\ \bar{d}_B = 40{,}21,\ S^2 = \frac{n}{n-1}\bar{d}_B = 40{,}41,\ S=6{,}36.$$

Для расчёта теоретических частот p_i^m используют табличные значения функции Лапласа $\Phi(z)$. Алгоритм вычисления p_i^m состоит в следующем:

- по нормированным значениям случайной величины Z находят значения $\Phi(z)$, а затем $F_N(x_i')$:

$$z_i = \frac{X_i - \bar{x}_B}{S},\ F_N(x_i) = 0{,}5 + \Phi(z_i).$$

Например,

$$X_1 = 149{,}5;\ z_1 = \frac{149{,}5 - 168{,}55}{6{,}36} = -3{,}0;\ \Phi(-3{,}0) = -0{,}4987;$$

$$F_N(149{,}5) = 0{,}0013;$$

- далее вычисляют вероятности $p_i^m = P(z_i \le X < z_{i+1}) = F_N(x_{i+1}) - F_N(x_i)$;

- находят числа $n_i^m = p_i^m n$, и если некоторое $n_i^m<5$, то соответствующие группы объединяются с соседними.

Результаты вычисления p_i^m, n_i^m, и χ_r^2 приведены в таблице 8.

По формуле

$$\chi_r^2 = \sum \frac{n_i^2}{n_i^m} - n \qquad (27)$$

можно сделать проверку расчетов.

$$\chi_r^2 = \frac{6^2}{11{,}38} + \frac{7^2}{15{,}4} + \ldots + \frac{15^2}{14} + \frac{15^2}{8} - 200 = 15{,}61$$

По таблице (приложения 3) можно найти число $\chi_{кр}^2$ по схеме: для уровня значимости α=0,05 и числа степеней свободы $l=k-r-1$=9-2-1=6$\Rightarrow \chi_{кр}^2$=12,6. Следовательно, критическая область - (12,6;∞). Величина χ_r^2=15,61 входит в критическую область, поэтому гипотеза о том, что случайная величина X подчинена нормальному закону распределения, отвергается.

При α=0,1 $\chi_{кр}^2$=10,6. Критическая область - (10,6;∞). Величина χ_r^2=15,61 также входит в критическую область и гипотеза о нормальном законе распределения величины X отвергается.

При α=0,01 $\chi_{кр}^2$=16,8, (16,8;∞). В этом случае нет оснований отвергать гипотезу о нормальном законе распределения.

Таблица 8 «Определение χ_r^2»

i	$x_i \div x_{i+1}$	n_i	$\Phi(z_i)$	$F_N(x_i)$	$F_N(x_{i+1})$	$p_i^m = F_N(x_{i+1}) - F_N(x_i)$	$n_i^m = p_i^m n$	$\frac{(n_i - n_i^m)^2}{n_i^m}$
0	$-\infty \div$ 149,5	0	-0,500	0,000	0,0013	0,0013	0,26	-
1	149,5÷ 152,5	1	-0,449	0,0013	0,0059	0,0046	0,92	-
2	152,5÷ 155,5	0	-0,494	0,0059	0,02	0,014	2,8	-
3	155,5÷ 158,5	5	-0,48	0,02	0,057	0,037	7,4	2,54
4	158,5÷ 161,5	7	-0,44	0,057	0,134	0,077	15,4	4,58
5	161,5÷ 164,5	21	-0,37	0,134	0,26	0,126	25,2	0,7
6	164,5÷ 167,5	38	-0,24	0,26	0,433	0,1725	34,5	0,36
7	167,5÷ 170,5	39	-0,07	0,433	0,62	0,188	37,6	0,06
8	170,5÷ 173,5	38	0,12	0,62	0,78	0,16	32	1,125

9	173,5÷ 176,5	21	0,28	0,78	0,89	0,11	22	0,045
10	176,5÷ 179,5	15	0,39	0,89	0,96	0,07	14	0,071
11	179,5÷ 182,5	8	0,46	0,96	0,99	0,03	6	6,125
12	182,5÷ 185,5	3	0,49	0,99	0,996	0,006	1,2	-
13	185,5÷ 188,5	3	0,496	0,996	0,999	0,003	0,6	-
14	188,5÷ ∞	1	0,5	0,999	1,0	0,001	0,2	-

$\sum = 200$ $\sum = 1,0000$ $\sum = 15,61$

<u>2 часть</u>

1) Данные таблицы №2 сгруппируем в корреляционную таблицу №9.

Таблица 9 “Корреляционная таблица”

n_{x_i}	10	10	99	98	97	96	95	94	93	92	91	90	89	88	87	86	85	84	83	82	81	80	79	Y/X
1															1									148
1																1								155
2														1			1							156
2												1								1				157
1											1													158
3																1			1		1			159
3															1		1		1					160
6											1			1				2		1			1	161
6		1						1		1		1	1									1		162
9									1	1	2		1	3								1		163
15					1						2	2	3		1	1	1	3	2					164
15								4	1		2			1	4		1	1						165
8													2	1	1			3		1				166
14										1	2	1	3	2		1	3				1			167
10									2	1		1		2	1				1	1	1			168
15										2	5		1	2	3				1				1	169
15						1	1		1	1	3	4		2		1					1			170
11					1			1		1	3	1	2				1		1					171
12			2	1		1		1			2	1	1	1	1					1				172
9						1					1	3	2		1			1						173
6					1	1					1	1		1		1								174
6							1				1	2	1				1							175
10							1		3	1		1		1	1	1				1				176
1															1									177
4											1	2	1											178
3			1														2							179
3				1								1					1							180
2										1			1											181
1												1												182
1												1												183
1				1																				184
1											1													185
1										1						1								186
1																								187
1	1																							190
200	1	1	3	3	3	4	3	7	8	11	28	24	19	18	17	7	12	10	7	6	4	2	2	n_{y_i}

2) Строим в системе координат множество, состоящее из 200 экспериментальных точек (рисунок 4).

По расположению точек делаем заключение о том, что математическую модель можно искать в виде $y = kx + b$.

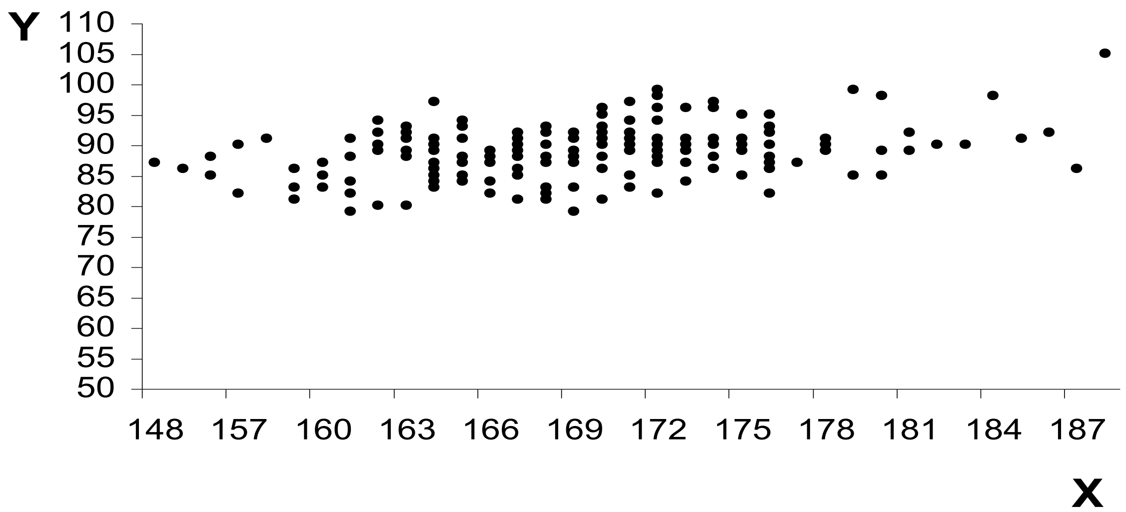

Рис.4

3) Найдём выборочные уравнения линейной регрессии.

Для упрощения расчётов разобьём случайные величины на интервалы и выберем средние значения. Для величины X указанные действия были выполнены в 1 части задания.

Для случайной величины Y, используя (1), получим h=2, число интервалов равно 13. Результаты внесём в таблицу со сгруппированными данными №10.

Находим средние значения $\bar{y}, \bar{x}^2, \bar{y}^2, \overline{xy}$, по формулам:

$$\bar{x} = \frac{1}{n}\sum x_i , \qquad (28)$$

$$\bar{y} = \frac{1}{n}\sum y_i , \qquad (29)$$

$$\bar{x}^2 = \frac{1}{n}\sum x^2{}_i , \qquad (30)$$

$$\overline{xy} = \frac{1}{n}\sum x_i y_i . \qquad (31)$$

$$\bar{y} = \frac{1}{200}(80*8 + 82*13 + ... + 100*1 + 104*1) = 88.53$$

$$\overline{x}^2 = \frac{1}{200}(149.5^2 * 1 + ... + 185.5^2 * 3 + 188.5^2 * 1) = 28449.31$$

$$\overline{y}^2 = \frac{1}{200}(80^2 * 8 + 82^2 * 13 + ... + 104^2 * 1) = 7856.02$$

$$\sum_{i=1}^{14} \sum_{j=1}^{13} n_{ij} x_i y_j = 149{,}5*86+155{,}5(82+\ldots+90)+\ldots+188{,}5*104=2986101$$

Используя формулы:

$$\sigma_{xB} = \sqrt{\overline{x}^2 - (\overline{x})^2}\ , \qquad (32)$$

$$\sigma_{yB} = \sqrt{\overline{y}^2 - (\overline{y})^2}\ , \qquad (33)$$

получим

$$\sigma_{xB} = \sqrt{28449{,}31 - 168{,}55^2} = 6{,}34\ , \sigma_{yB} = \sqrt{7856{,}02 - 88{,}53^2} = 4{,}297$$

Таблица 10 «Сгруппированные данные выборки»

№		1	2	3	4	5	6	7	8	9	10	11	12	13	14	
	XY	149,5	152,5	155,5	158,5	161,5	164,5	167,51 70,517	170,5	173,5	176,5	179,5	182,5	185,5	188,5	n_{y_j}
1	80				1	3		3	1							8
2	82			1	2	1	3	3	2		1					13
3	84			1	1	2	9	3	1	2		3				22
4	86	1		1	2		7	5	1	1	3			1		24
5	88			1		6	7	10	6	4	2	1				37
6	90			1	1	4	6	9	14	9	4	1	2	1		52
7	92					3	1	6	3		4	1		1		19
8	94					1	4		3	1	1					10
9	96						1		3	3						7
10	98								3			2	1			6
11	100					1										1
12	102															
13	104														1	1
	n_{x_i}	1		5	7	21	38	39	38	21	15	8	3	3	1	200

4) Вычисляем выборочный коэффициент корреляции r_B по формуле:

$$r_B = \frac{\sum_{i=1}^{k_1} \sum_{j=1}^{k_2} n_{ij} x_i y_j - \overline{x} n \overline{y}}{n \sigma_{xB} \sigma_{yB}}. \qquad (34)$$

$$r_B = \frac{2986101 - 200 \cdot 168{,}5 \cdot 88{,}53}{200 \cdot 6{,}34 \cdot 4{,}3} = 0{,}32$$

Принято считать, что если $0{,}1 < r_B < 0{,}3$ – связь слабая, если $0{,}3 < r_B < 0{,}5$ – связь умеренная, если $0{,}5 < r_B < 0{,}7$ – связь заметная, если $0{,}7 < r_B < 0{,}9$ – связь высокая, если $0{,}9 < r_B < 0{,}99$ – связь весьма высокая.

Для данного примера связь между X и Y умеренная.

Затем получают выборочное уравнение линейной регрессии Y на X в виде:

$$\bar{y}_x - \bar{y} = \frac{\sigma_{yB}}{\sigma_{xB}} r_B (x - \bar{x}) \qquad (35)$$

и выборочное уравнение линейной регрессии X на Y :

$$\bar{x}_y - \bar{x} = \frac{\sigma_{xB}}{\sigma_{yB}} r_B (y - \bar{y}). \qquad (36)$$

$$\bar{x}_y - 168{,}55 = \frac{6{,}34 \cdot 0{,}32}{4{,}3}(y - 88{,}53) \text{ и}$$

$$\bar{x}_y = 0{,}47y + 126{,}78$$

$$\bar{y}_x - 88{,}53 = \frac{4{,}3 \cdot 0{,}32}{6{,}34}(x - 168{,}55) \text{ или}$$

$$\bar{y}_x = 0{,}22x + 51{,}95$$

Самостоятельно сформулировать выводы:

1. **О результатах проверки гипотезы о законе распределения случайной величины *X*;**
2. **О характере связи между X и Y**
3. **Показать, что полученное уравнение линейной регрессии, адекватно описывает данные выборки.**

Приложение 1

«Нормированная функция Лапласа» $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-\frac{t^2}{2}} dt$

z	0	1	2	3	4	5	6	7	8	9
0,0	00000	00399	00789	01197	01595	01994	02392	02790	03188	03586
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1,1	36433	36650	36864	37076	37286	37493	37698	38000	38100	38298
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1,3	40320	40490	40658	40824	40988	41149	41308	41466	41621	41774
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2,2	48610	48645	48679	48713	48745	48778	48806	48840	48870	48899
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643

Продолжение приложения 1

z	0	1	2	3	4	5	6	7	8	9
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49896	49900
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997
4,0	49997									
5,0	49999									

Приложение 2

«Значения чисел q в зависимости от объёма выборки n и надёжности γ для определения доверительного интервала среднего квадратического отклонения σ_x»

n	γ				n	γ		
	0.95	0.99	0.999			0.95	0.99	0.999
7	0.92	-	-		25	0.32	0.49	0.73
8	0.80	-	-		30	0.28	0.43	0.63
9	0.71	-	-		35	0.26	0.38	0.56
10	0.65	-	-		40	0.24	0.35	0.50
11	0.59	0.98	-		45	0.22	0.32	0.46
12	0.55	0.90	-		50	0.21	0.30	0.43
13	0.52	0.83	-		60	0.188	0.269	0.38
14	0.48	0.78	-		70	0.174	0.245	0.34
15	0.46	0.73	-		80	0.161	0.226	0.31
16	0.44	0.70	-		90	0.151	0.211	0.29
17	0.42	0.66	-		100	0.143	0.198	0.27
18	0.40	0.63	0.96		150	0.115	0.160	0.211
19	0.39	0.60	0.92		200	0.099	0.136	0.185
20	0.37	0.58	0.88		250	0.089	0.120	0.162

Приложение 3

“Критические точки распределения χ^2”

Число Степеней Свободы	Уровень значимости α					
	0,01	0,05	0,1	0,90	0,95	0,99
1	6,6	3,8	2,71	0,02	0,004	0,0002
2	9,2	6,0	4,61	0,21	0,1	0,02
3	11,3	7,8	6,25	0,58	0,35	0,12
4	13,3	9,5	7,78	1,06	0,71	0,30
5	15,1	11,1	9,24	1,61	1,15	0,55
6	16,8	12,6	10,6	2,20	1,64	0,87
7	18,5	14,1	12,0	2,83	2,17	1,24
8	20,1	15,5	13,4	3,49	2,73	1,65
9	21,7	16,9	14,7	4,17	3,33	2,09
10	23,2	18,3	16,0	4,87	3,94	2,56
11	24,7	19,7	17,3	5,58	4,57	3,05
12	26,2	21,0	18,5	6,30	5,23	3,57
13	27,7	22,4	19,8	7,04	5,89	4,11
14	29,1	23,7	21,1	7,79	6,57	4,66
15	30,6	25,0	22,3	8,55	7,26	5,23
16	32,0	26,3	23,5	9,31	7,96	5,81
17	33,4	27,6	24,8	10,1	8,67	6,41
18	34,8	28,9	26,0	10,9	9,39	7,01
19	36,2	30,1	27,2	11,7	10,1	7,63
20	37,6	31,4	28,4	12,4	10,9	8,26
21	38,9	32,7	29,6	13,2	11,6	8,90
22	40,3	33,9	30,8	14,0	12,3	9,54
23	41,6	35,2	32,0	14,8	13,1	10,2
24	43,0	36,4	33,2	15,7	13,8	10,9
25	44,3	37,7	34,4	16,5	14,6	11,5
26	45,6	38,9	35,6	17,3	15,4	12,2
27	47,0	40,1	36,7	18,1	16,2	12,9
28	48,3	41,3	37,9	18,9	16,9	13,6
29	49,6	42,6	39,1	19,8	17,7	14,3
30	50,9	43,8	40,3	20,6	18,5	15,0