МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Настоящее пособие является практическим продолжением учебного пособия [2]. Оно содержит подробный пример выполнения статистических расчетов.

В пособие включено 30 вариантов типового расчета по теме «Математическая статистика». Предназначено для студентов всех форм обучения.

Задание к типовому расчету (курсовой работе)

1. Составить вариационный ряд.
2. Составить сгруппированный статистический ряд.
3. Построить гистограмму выборки.
4. Построить график эмпирической функции распределения.
5. Найти выборочное среднее, выборочное среднеквадратическое отклонение, коэффициенты асимметрии и эксцесса.
6. Построить доверительный интервал для математического ожидания при доверительной вероятности $\gamma_1 = 0{,}9$.
7. Построить доверительный интервал для среднеквадратического отклонения при доверительной вероятности $\gamma_2 = 0{,}95$.
8. Проверить с помощью критерия Пирсона гипотезу H_0 о том, что генеральная совокупность распределена по нормальному закону.

Пример выполнения статистического расчета

Пусть в результате некоторого эксперимента получено n значений изучаемой случайной величины X. Данные записаны в виде таблицы и составляют первичную выборку объема n =100.

81,38	66,94	88,26	72,73	65,72	109,76	88,45	96,73	85,90	66,93
61,31	86,73	89,03	65,14	80,06	93,68	69,09	57,56	74,79	77,66
83,88	54,58	58,83	78,05	67,33	70,51	60,18	67,01	52,27	50,84
63,83	72,92	69,22	71,12	88,77	45,84	98,34	76,98	57,34	79,80
63,29	66,80	78,64	74,30	56,75	77,85	71,95	67,63	69,32	58,53
58,55	39,92	73,16	96,09	70,44	67,22	73,04	59,72	72,19	65,53
86,60	65,12	74,32	77,27	79,60	79,95	61,63	45,51	104,02	84,87
71,36	68,05	51,88	81,18	75,06	85,37	50,82	87,18	64,12	86,93
71,90	30,03	49,98	42,52	60,96	99,11	78,32	44,69	43,08	79,58
60,85	64,43	95,54	89,67	57,37	98,60	80,13	67,04	77,00	69,26

1. Представим выборку в виде вариационного ряда: последовательности исходных величин, записанных в возрастающем порядке.

30,03	61,63	71,36	80,13
39,92	63,29	71,90	81,18
42,52	63,83	71,95	81,38
43,08	64,12	72,19	83,88
44,69	64,43	72,73	84,87
45,51	65,12	72,92	85,37
45,84	65,14	73,04	85,90
49,98	65,53	73,16	86,60
50,82	65,72	74,30	86,73
50,84	66,80	74,32	86,93
51,88	66,93	74,79	87,18
52,27	66,94	75,06	88,26
54,58	67,01	76,98	88,45
56,75	67,04	77,00	88,77
57,34	67,22	77,27	89,03
57,37	67,33	77,66	89,67
57,56	67,63	77,85	93,68
58,53	68,05	78,05	95,54
58,55	69,09	78,32	96,09
58,83	69,22	78,64	96,73
59,72	69,26	79,58	98,34
60,18	69,32	79,60	98,60
60,85	70,44	79,80	99,11
60,96	70,51	79,95	104,02
61,31	71,12	80,06	109,76

2. Составим группированный статистический ряд. Найдем наименьший и наибольший элемент выборки: $x_{\min} = 30{,}03 \approx 30$; $x_{\max} = 109{,}76 \approx 110$. Разобьем отрезок $[x_{\min}; x_{\max}]$ на k равных по длине промежутков. При объеме выборки n=100 рекомендуется взять $k = 8$. Число n_j – частота попадания элементов выборки в j-ый промежуток.

Таблица 1

Интервал	[30; 40]	(40; 50]	(50; 60]	(60; 70]	(70; 80]	(80; 90]	(90; 100]	(100; 110]
n_j	2	6	14	26	28	15	7	2

3. Для построения гистограммы дополним таблицу 1 тремя строками: $h_j = x_{j+1} - x_j$, $w_j = n_j / n$ и w_j / h_j, где h_j – длина j-ого промежутка;

$w_j = n_j / n$ – относительная частота попадания элементов выборки в j-ый промежуток.

Таблица 2

Интервал	[30; 40]	(40; 50]	(50; 60]	(60; 70]	(70; 80]	(80; 90]	(90; 100]	(100; 110]
n_j	2	6	14	26	28	15	7	2
h_j	10	10	10	10	10	10	10	10
w_j	0,02	0,06	0,14	0,26	0,28	0,15	0,07	0,02
w_j / h_j	0,002	0,006	0,014	0,026	0,028	0,015	0,007	0,002

Последняя строка таблицы 2 определяет высоты столбцов гистограммы, приведенной на рисунке 1.

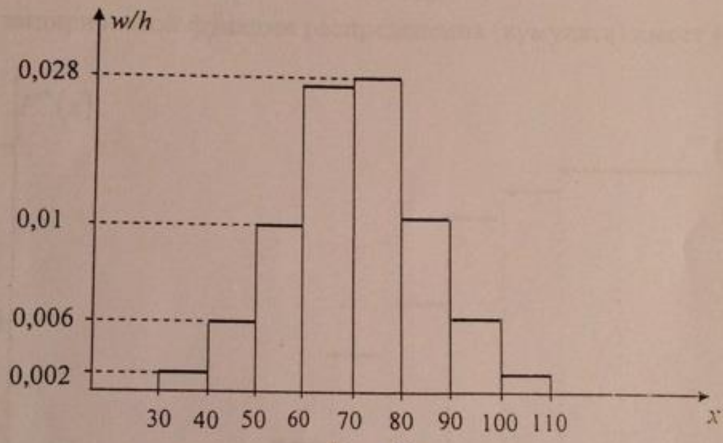

Рис. 1

Для непрерывной случайной величины гистограмма аппроксимирует плотность вероятности генеральной совокупности.

4. Для построения графика эмпирической функции распределения в таблицу 1 добавим две строки, в которых следует записать значения $N_j = n_1 + n_2 + \ldots + n_j$ и N_j / n.

Таблица 3

Интервал	[30; 40]	(40; 50]	(50; 60]	(60; 70]	(70; 80]	(80; 90]	(90; 100]	(100; 110]
n_j	2	6	14	26	28	15	7	2
N_j	2	8	22	48	76	91	98	100
N_j / n	0,02	0,08	0,22	0,48	0,76	0,92	0,98	1

Значения эмпирической функции распределения равны $F^{*}(x)=N_j/n$, если x принадлежит j - ому промежутку; $F^{*}(x)=0$, если $x \le x_{\min}$; и $F^{*}(x)=1$, если $x > x_{\max}$. Получим

$$F^{*}(x)=\begin{cases} 0 & x \le 30; \\ 0{,}02 & 30 < x \le 40; \\ 0{,}08 & 40 < x \le 50; \\ 0{,}22 & 50 < x \le 60; \\ 0{,}48 & 60 < x \le 70; \\ 0{,}76 & 70 < x \le 80; \\ 0{,}92 & 80 < x \le 90; \\ 0{,}98 & 90 < x \le 100; \\ 1 & x > 100. \end{cases}$$

График эмпирической функции распределения (кумулята) имеет вид:

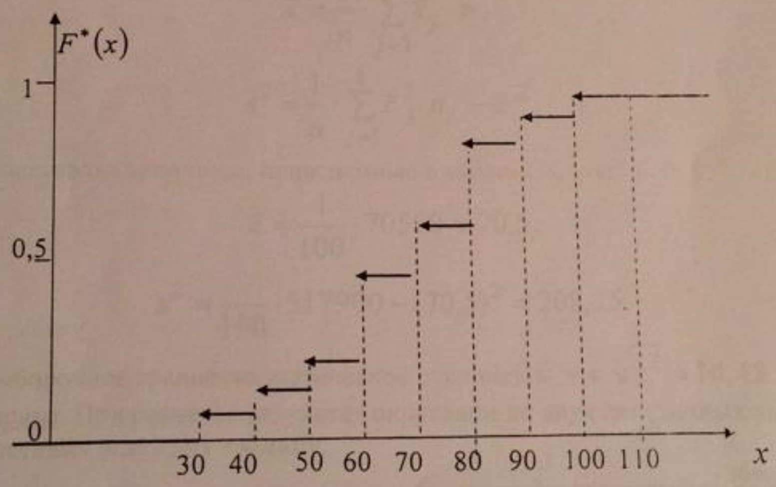

Рис. 2

Эмпирическая функция $F^{*}(x)$ аппроксимирует функцию распределения генеральной совокупности.

5. Для вычисления числовых характеристик выборки построим новый вариационный ряд. Обозначим $\overline{x}_j$ – середину j- того промежутка. Это значение присваивается всем элементам выборки, попавшим в j-ый интервал.

Коэффициенты асимметрии A и эксцесса E вычисляются по формулам:

$$A = \frac{1}{n \cdot s^3} \cdot \sum_{j=1}^{\kappa} (\bar{x}_j - \bar{x})^3 \cdot n_j;$$

$$E = \frac{1}{n \cdot s^4} \cdot \sum_{j=1}^{\kappa} (\bar{x}_j - \bar{x})^4 \cdot n_j - 3.$$

Подставляя величины, приведенные в таблице 4, получим:

$$A = \frac{1}{100 \cdot 14{,}45^3} \cdot (-12075) \approx -0{,}04;$$

$$E = \frac{1}{100 \cdot 14{,}45^4} \cdot 12575331{,}25 - 3 \approx -0{,}11$$

Если коэффициенты A и E значительно отклоняются от нулевого значения, то выборочное распределение отличается от нормального.

<u>*Замечание*</u>. При $n < 30$ в расчетах следует использовать исправленную оценку среднеквадратического отклонения $s^* = \sqrt{\frac{n}{n-1} \cdot s^2}$.

6. Оценку истинного значения параметра a (математического ожидания) дает доверительный интервал, который для случая большой выборки определяется формулой

$$\bar{x} - \frac{\tilde{t}s}{\sqrt{n}} < a < \bar{x} + \frac{\tilde{t}s}{\sqrt{n}},$$

где значение $\tilde{t}$ находится из условия $2 \cdot \Phi_0(\tilde{t}) = \gamma_1$.

Для доверительной вероятности (надежности) $\gamma_1 = 0{,}9$ по таблице значений функции Лапласа $\Phi_0(t)$, приведенной в приложении 2, находим число 0,4505, наиболее близкое к $\gamma_1/2$. Это число расположено в строке, именованной «1,6», и столбце с названием «5». Искомое значение $\tilde{t} = 1{,}6 + 0{,}05 = 1{,}65$, так как $2 \cdot \Phi_0(1{,}65) \approx 0{,}9$. При $\bar{x} = 70{,}5$ и точности оценки $\varepsilon = \tilde{t}s/\sqrt{n} = 2{,}61$ с надежностью 0,9 доверительный интервал для математического ожидания равен (67,89; 73,11).

7. Оценку истинного значения параметра σ дает доверительный интервал для среднеквадратического отклонения, который для случая большой выборки определяется по формуле

$$\frac{s}{\sqrt{1 + \tilde{t}\sqrt{2/n}}} < \sigma < \frac{s}{\sqrt{1 - \tilde{t}\sqrt{2/n}}}.$$

Таблица 6

Интервал		$(-\infty; 50]$		$(50; 60]$		$(60; 70]$		$(70; 80]$		$(80; 90]$		$(90; \infty)$	
Граница x_{j+1}			50		60		70		80		90		
n_j		8		14		26		28		15		9	
$y_{j+1} = \dfrac{x_{j+1} - \bar{x}}{s}$			− 1,42		− 0,73		− 0,03		0,66		1,35		
$\Phi_0(y_{j+1})$	− 0,5		− 0,42		− 0,27		− 0,01		0,25		0,41		0,5
$p_j = \Phi_0(y_{j+1}) - \Phi_0(y_j)$		0,08		0,15		0,26		0,26		0,17		0,09	
np_j		8		15		26		26		17		9	
$\lvert n_j - np_j \rvert$		0		1		0		2		2		0	
$(n_j - np_j)^2 / np_j$		0		0,07		0		0,15		0,24		0	

По заданной доверительной вероятности $\gamma_2 = 0{,}95$ по таблице значений функции Лапласа, находим $2\Phi_0(1{,}96) \approx 0{,}95$, следовательно, $\tilde{t} = 1{,}96$. Тогда с надежностью 0,95 доверительный интервал для среднеквадратического отклонения σ имеет вид (12,78; 16,99).

8. Проверка нулевой гипотезы H_0 по критерию Пирсона состоит из следующих этапов.

a) По выборке вычисляются точечные оценки математического ожидания $a \approx \bar{x}$ и среднеквадратического отклонения $\sigma \approx s$.

b) В каждом промежутке определяются эмпирические частоты и теоретические вероятности.

c) Для данной выборки вычисляется наблюдаемое значение $\chi^2_{\text{набл}}$ критерия Пирсона.

d) Задается уровень значимости α и подсчитывается количество степеней свободы.

e) По таблице приложения 3 определяется значение $\chi^2_{\text{крит}}$.

f) Если $\chi^2_{\text{набл}} > \chi^2_{\text{крит}}$, то гипотеза H_0 отвергается как маловероятная.

Значения коэффициентов асимметрии A и эксцесса E, близкие к нулю, а также вид гистограммы позволяют выдвинуть гипотезу H_0 о том, что генеральная совокупность распределена по нормальному закону. Проверим эту гипотезу с помощью критерия Пирсона.

a) Этот этап проделан в пункте 5.

b) Используя таблицу приложения 2, найдем теоретические вероятности попадания варианты в каждый промежуток по формуле

$$p_j = P(x_j \le x < x_{j+1}) = \Phi_0\left(\frac{x_{j+1} - \bar{x}}{s}\right) - \Phi_0\left(\frac{x_j - \bar{x}}{s}\right),$$

и вычислим $\chi^2_{\text{набл}} = \sum_{j=1}^{r} \frac{(n_j - np_j)^2}{np_j}$.

Вычисления удобно проводить по таблице 6. Предварительно следует изменить таблицу 1, объединив первый интервал со вторым и седьмой интервал с восьмым, так как в критерии Пирсона предполагается, что количество вариант в каждом интервале не меньше пяти. Крайние интервалы расширяются влево и вправо до бесконечности, причем $\Phi_0(-\infty) = -0{,}5$; $\Phi_0(+\infty) = 0{,}5$.

Сначала в граничных точках вычисляем аргументы функции Лапласа. Например, для промежутка (70; 80] имеем

$$y_4 = \frac{x_4 - \bar{x}}{s} = \frac{70 - 70{,}5}{14{,}45} = -0{,}0346 \approx -0{,}03,$$

$$y_5 = \frac{x_5 - \bar{x}}{s} = \frac{80 - 70{,}5}{14{,}45} = 0{,}6574 \approx 0{,}66.$$

По таблице приложения 2 вычисляем теоретическую вероятность p_j попадания варианты в промежуток (70; 80]

$$p_4 = \Phi_0(y_5) - \Phi_0(y_4) = \Phi_0(0{,}66) - \Phi_0(-0{,}03).$$

Функция $\Phi_0(x)$ является нечетной, следовательно,

$$p_4 = \Phi_0(0{,}66) + \Phi_0(0{,}03) \approx 0{,}2454 + 0{,}012 \approx 0{,}26.$$

В последней строке таблицы 6 помещены значения $\left(n_j - np_j\right)^2 / np_j$. Для промежутка [70; 80) эта величина принимает значение

$$(28 - 100 \cdot 0{,}26)^2 / (100 \cdot 0{,}25) \approx 0{,}15.$$

c) Суммируя все числа последней строки, получаем $\chi^2_{\text{набл}} = 0{,}46$. Полученное число необходимо сравнить с величиной $\chi^2_{\text{крит}}$.

d) Количество интервалов r вариационного ряда, приведенного в таблице 6, равно шести. Число степеней свободы $k = r - 3 = 3$.

e) Выбираем уровень значимости $\alpha = 1 - \gamma_2 = 1 - 0{,}95 = 0{,}05$. В таблице приложения 3 параметрам $k = 3$ и $\alpha = 0{,}05$ соответствует значение $\chi^2_{\text{крит}} = 7{,}82$.

f) При выбранной надежности 0,95 $\chi^2_{\text{набл}} < \chi^2_{\text{крит}}$. Следовательно, отвергать гипотезу H_0 оснований нет. Предположение о том, что исследуемая физическая величина распределена по нормальному закону с параметрами $a = 70{,}5$; $\sigma = 14{,}45$ не противоречит результатам измерений.

Значит, можно считать, что функция плотности вероятности изучаемой физической величины имеет вид.

$$f(x) = \frac{1}{14{,}45}\varphi\left(\frac{x - 70{,}5}{14{,}45}\right) = \frac{1}{14{,}45 \cdot \sqrt{2\pi}} e^{-\frac{(x-70{,}5)^2}{2 \cdot 14{,}45^2}}.$$

Значения функции $\varphi(x)$ приведены в таблице приложения 1, а график $f(x)$ изображен на рисунке 3 сплошной линией. Отдельные точки на том же рисунке последней строке таблицы 1, по которой строилась гистограмма. Очевидно, что теоретическое распределение вполне согласуется с результатами выборки.

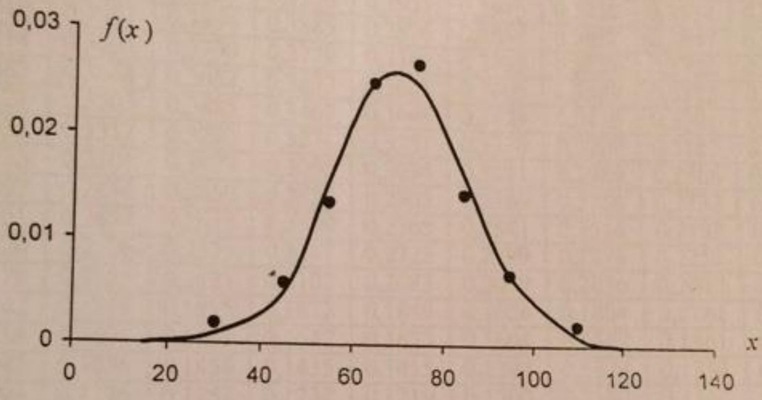

Рис. 3

Приложение 1. Таблица значений функции $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0.3939	0,3932	0.3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0.3857	0,3847	0.3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0.3739	0,3726	0.3712	0,3697
0,4	0,3683	0,3668	0,3652	0,3637	0,3621	0,3605	0,3589	0,3572	0.3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0.3410	0,3391	0.3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0.3209	0,3187	0.3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0.2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0.2756	0,2732	0.2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0.2516	0,2492	0.2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0.2227	0,2203
1,1	0,2178	0,2155	0,2131	0,2107	0,2083	0,2059	0.2036	0,2012	0.1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0.1804	0,1781	0.1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1623	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0.1374	0,1354	0.1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0.1182	0,1163	0.1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0.1006	0,0989	0.0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0.0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0.0707	0,0694	0.0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0.0584	0,0573	0.0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0.0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0.0387	0,0379	0.0371	0,0363
2,2	0,0355	0,0347	0,0339	0,0332	0,0323	0,0317	0.0310	0,0303	0.0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0.0246	0,0241	0.0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0.0194	0,0189	0.0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0.0143	0,0139
2,6	0,0135	0,0132	0,0129	0,0126	0,0122	0,0119	0.0116	0,0113	0.0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0.0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0.0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0.0050	0,0048	0.0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0.0037	0,0036	0.0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0.0020	0,0019	0.0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0.0014	0,0014	0.0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0.0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0.0007	0,0007	0.0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0.0005	0,0005	0.0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0.0003	0,0003	0.0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0.0002	0,0002	0.0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0.0002	0,0002	0.0001	0,0001

Примечание. В таблицах приложения 1 в верхней строке записаны третьи цифры аргумента функции, расположенного в первом столбце.