МИНОБРАНАУКИ РФ

Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования

«Тульский государственный университет»
КАФЕДРА МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ

«Прикладная математика»
для студентов-заочников, обучающихся по направлению
190700 Технология транспортных процессов

Тула 2013

Задание 1. Первичная обработка выборок

По выборке А составить вариационный ряд, вычислить относительные частоты и накопленные частоты, построить полигон и гистограмму, составить эмпирическую функцию распределения и построить ее график, вычислить числовые характеристики вариационного ряда: выборочное среднее, дисперсию, стандартное отклонение.

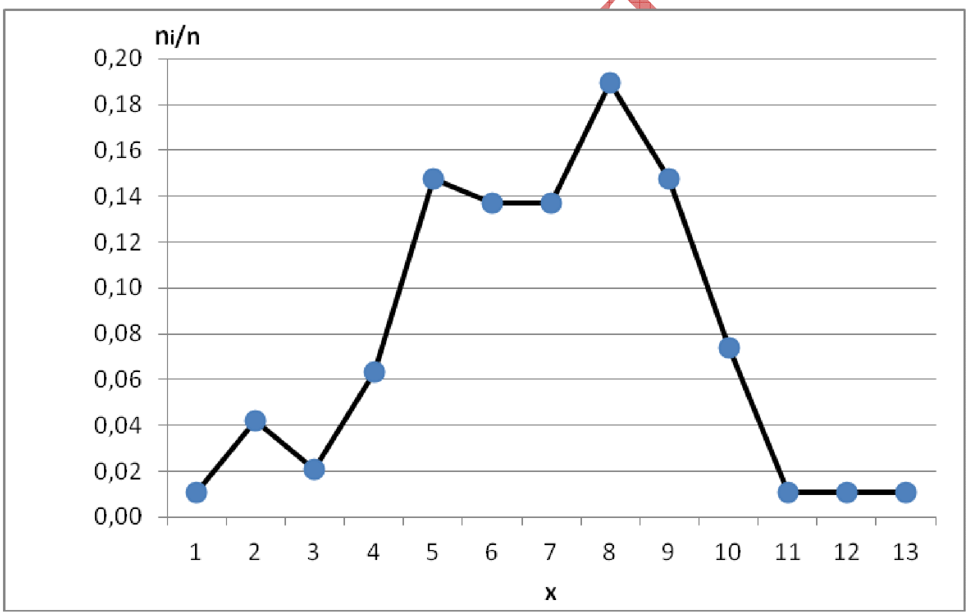
1.1. По выборке А случайной величины X составим вариационный ряд по значениям, так как размах $X_{max} - X_{min} + 1 = 12 - 0 + 1 = 9$ достаточно мал (табл. 1).

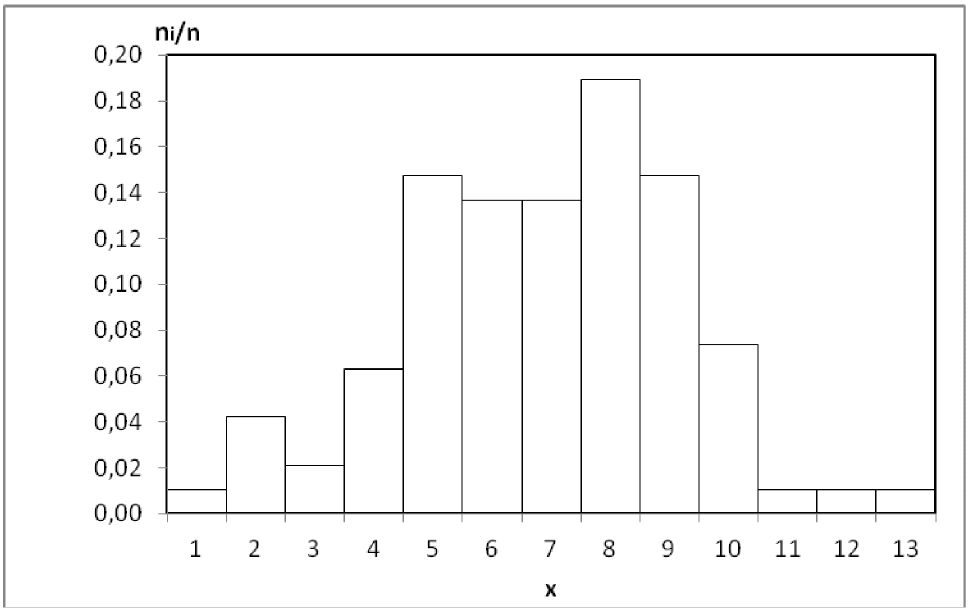
Таблица 1

Вариационный ряд выборки А

x_i	n_i	n_i/n	w_i
0	1	0,0105	0,0105
1	4	0,0421	0,0526
2	2	0,0211	0,0737
3	6	0,0632	0,1368
4	14	0,1474	0,2842
5	13	0,1368	0,4211
6	13	0,1368	0,5579
7	18	0,1895	0,7474
8	14	0,1474	0,8947
9	7	0,0737	0,9684
10	1	0,0105	0,9789
11	1	0,0105	0,9895
12	1	0,0105	1,0000
Σ	95	1	-

Полигон и гистограмма показаны на рисунках 1 и 2.



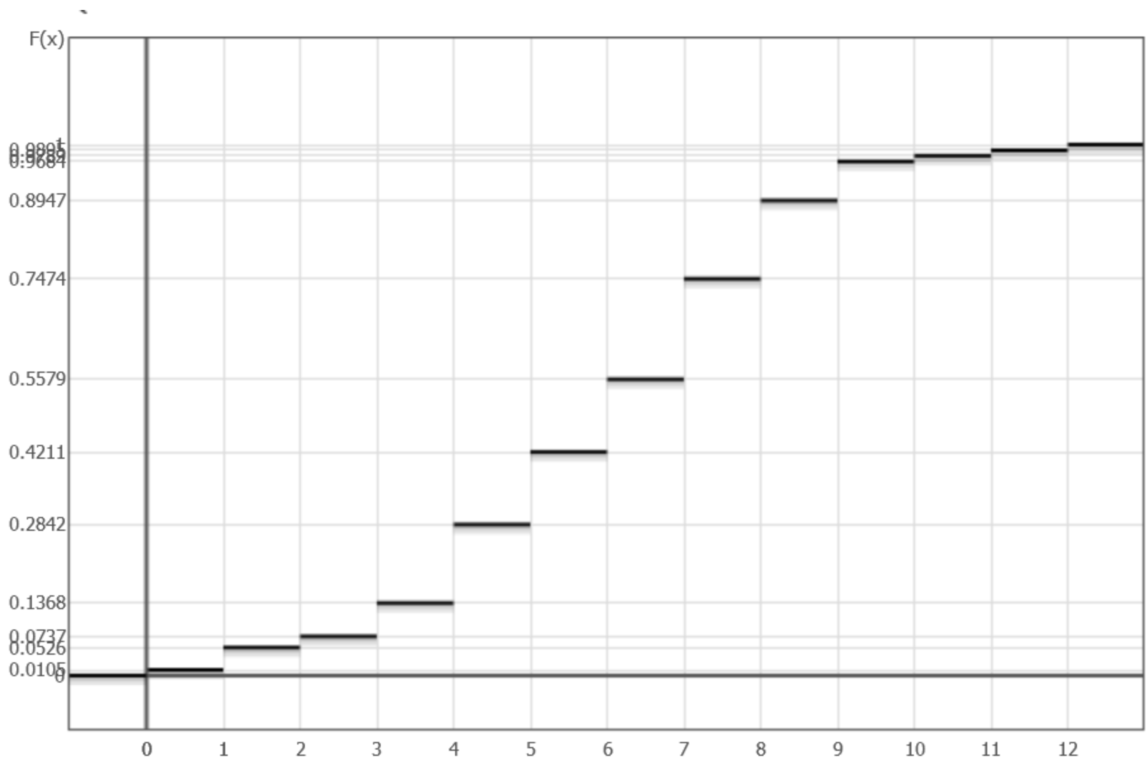


Эмпирическая (выборочная) функция распределения случайной величины X определяется по накопленным частотам (табл. 1):

$$\hat{F}(x)=\begin{cases} 0, & x\le 0 \\ 0{,}0105, & 0<x\le 1 \\ 0{,}0526, & 1<x\le 2 \\ 0{,}0737, & 2<x\le 3 \\ 0{,}1368, & 3<x\le 4 \\ 0{,}2842, & 4<x\le 5 \\ 0{,}4211, & 5<x\le 6 \\ 0{,}5579, & 6<x\le 7 \\ 0{,}7474, & 7<x\le 8 \\ 0{,}8947, & 8<x\le 9 \\ 0{,}9684, & 9<x\le 10 \\ 0{,}9789, & 10<x\le 11 \\ 0{,}9895, & 11<x\le 12 \\ 1, & x>12 \end{cases}$$

График эмпирической функции распределения показан на рисунке 3[1].

[1] Необходимо поставить на всех горизонтальных линиях стрелку влево.



Для вычисления числовых характеристик вариационного ряда составим таблицу 2.

Таблица 2

Таблица для вычисления числовых характеристик вариационного ряда выборки А

x_i	n_i	$x_i\ n_i$	$x_i^2 n_i$
0	1	0	0
1	4	4	4
2	2	4	8
3	6	18	54
4	14	56	224
5	13	65	325
6	13	78	468
7	18	126	882
8	14	112	896
9	7	63	567
10	1	10	100
11	1	11	121
12	1	12	144
Σ	95	559	3793

Выборочное среднее находится по формуле $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i n_i = \frac{1}{95}\cdot 559 = 5{,}88$,

а выборочная дисперсия $S^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 n_i - \overline{x}^2 = \frac{1}{95}\cdot 3793 - 5{,}88^2 = 5{,}35$.

Тогда стандартное отклонение

$$S = \sqrt{S^2} = \sqrt{5{,}35} = 2{,}31.$$

1.2. По выборке B случайной величины X составим вариационный ряд (табл. 3) по интервалам значений, так как размах

$$X_{max} - X_{min} + 1 = -14 - (-84) + 1 = 99$$

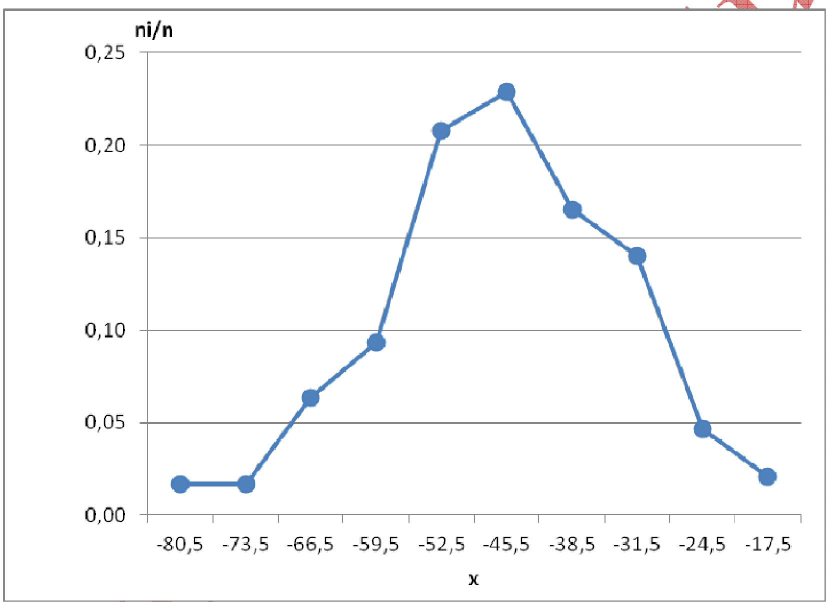
достаточно велик. Выберем длину интервала, равную 2.

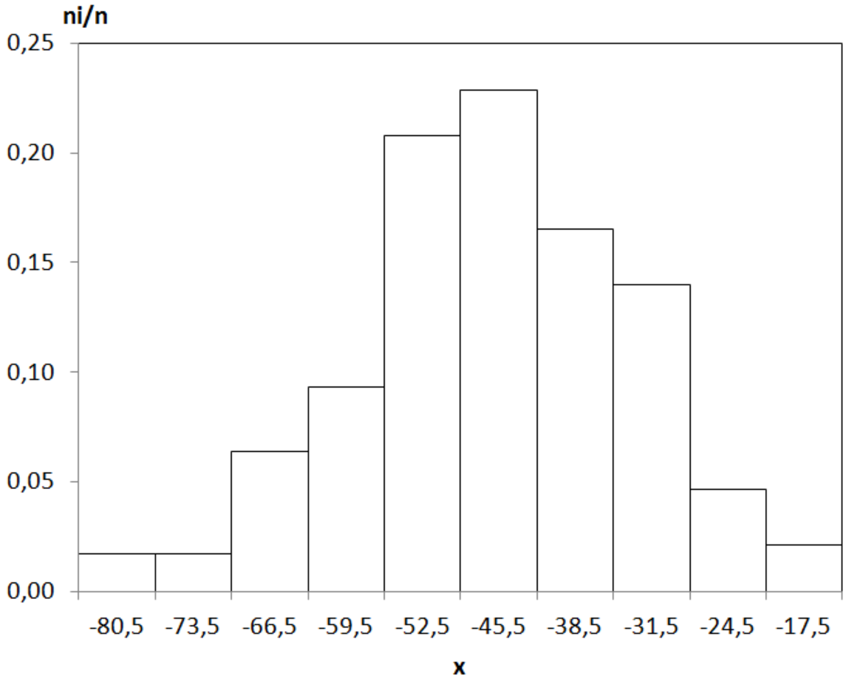
Таблица 3

Вариационный ряд выборки B

Номер интервала	Границы интервала	Середина интервала x_i	n_i	n_i/n	w_i
1	-84—-77	-80,5	4	0,0169	0,0169
2	-77—-70	-73,5	4	0,0169	0,0339
3	-70—-63	-66,5	15	0,0636	0,0975
4	-63—-56	-59,5	22	0,0932	0,1907
5	-56—-49	-52,5	49	0,2076	0,3983
6	-49—-42	-45,5	54	0,2288	0,6271
7	-42—-35	-38,5	39	0,1653	0,7924
8	-35—-28	-31,5	33	0,1398	0,9322
9	-28—-21	-24,5	11	0,0466	0,9788
10	-21—-14	-17,5	5	0,0212	1,0000
Σ	-	-	236	1,0000	-

Полигон и гистограмма показаны на рис. 4, 5.



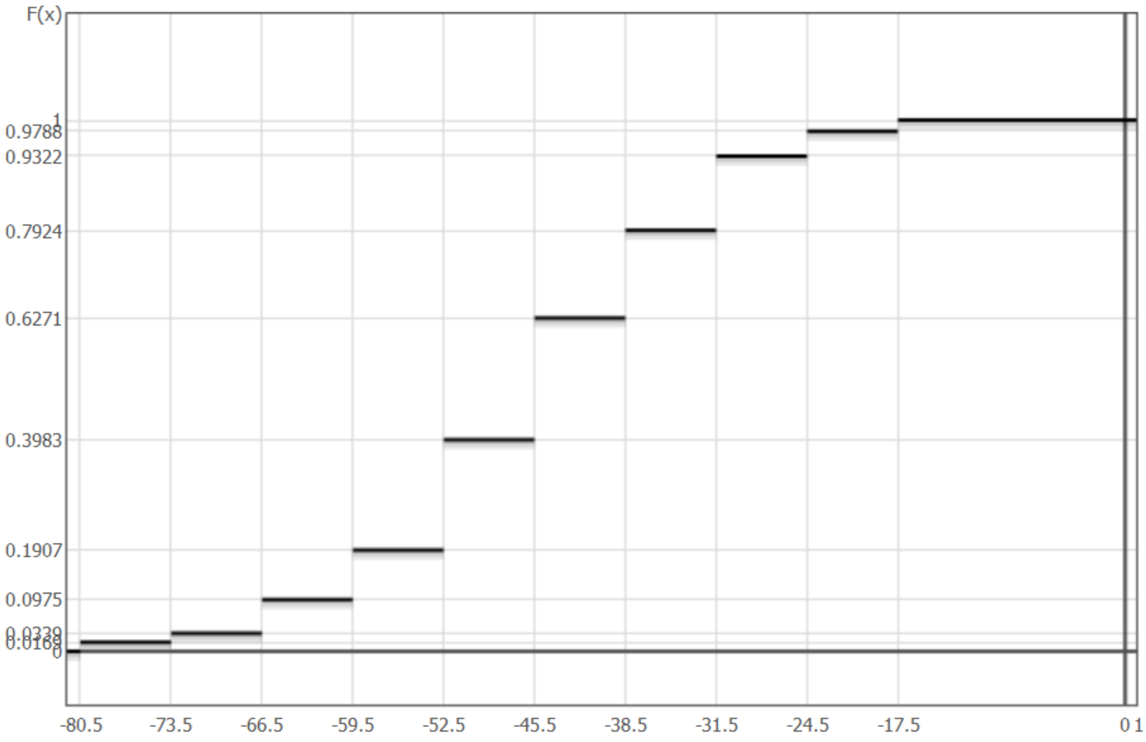


Эмпирическая (выборочная) функция распределения случайной величины X определяется по накопленным частотам (табл. 3):

$$\hat{F}(x)=\begin{cases} 0, & x \le -84 \\ 0{,}0169, & -84 < x \le -77 \\ 0{,}0339, & -77 < x \le -70 \\ 0{,}0975, & -70 < x \le -63 \\ 0{,}1907, & -63 < x \le -56 \\ 0{,}3983, & -56 < x \le -49 \\ 0{,}6271, & -49 < x \le -42 \\ 0{,}7924, & -42 < x \le -35 \\ 0{,}9322, & -35 < x \le -28 \\ 0{,}9788, & -28 < x \le -21 \\ 1, & x > -21 \end{cases}$$

График эмпирической функции распределения изображен на рис. 6[2].

[2] Необходимо поставить на всех горизонтальных линиях стрелку влево.



Для вычисления числовых характеристик вариационного ряда положим значение c = -45,5, соответствующее середине интервала, обладающего максимальной частотой. Шаг таблицы примем *k*= 7 и составим таблицу 4.

Таблица 4

Таблица для вычисления числовых характеристик вариационного ряда выборки В

Середина интервала x_i	n_i	$\frac{x_i - c}{k}$	$\frac{x_i - c}{k} n_i$	$\left(\frac{x_i - c}{k}\right)^2$	$\left(\frac{x_i - c}{k}\right)^2 n_i$
-80,5	4	-5	-20	25	100
-73,5	4	-4	-16	16	64
-66,5	15	-3	-45	9	135
-59,5	22	-2	-44	4	88
-52,5	49	-1	-49	1	49
-45,5	54	0	0	0	0
-38,5	39	1	39	1	39
-31,5	33	2	66	4	132
-24,5	11	3	33	9	99
-17,5	5	4	20	16	80
Σ	236	-	-16	-	786

Выборочное среднее найдем по формуле

$$\bar{x} = \frac{k}{n}\sum_{i=1}^{n}\frac{x_i - c}{k}n_i + c = \frac{7}{236}\cdot(-16) + (-45{,}5) = -45{,}97.$$

Выборочная дисперсия

$$S^2 = \frac{k^2}{n}\sum_{i=1}^{n}\left(\frac{x_i - c}{k}\right)^2 n_i - \left(\overline{x} - c\right)^2 = \frac{49}{236}\cdot 786 - (-45,97 - (-45,5))^2 = 162,97 .$$

Стандартное отклонение

$$S = \sqrt{S^2} = \sqrt{162,97} = 12,76 .$$

Задание 2. Точечное оценивание

2.1. Пусть случайная величина имеет распределение Пуассона. Используя метод моментов и метод максимального правдоподобия получения точечных оценок, найти по выборке А значение оценки $\hat{\lambda}$ неизвестного параметра λ.

<u>Метод моментов</u>

Математическое ожидание распределения Пуассона $P(X=k)=\frac{a^k}{k!}e^{-a}$ равно MX = a, где a – параметр распределения. Но выборочное среднее $\bar{x}$ = 5,88 (см. задание 1). По методу моментов приравниваются теоретические моменты и их выборочные значения. В нашем случае $\hat{a}$ = 5,88.

<u>Метод максимального правдоподобия</u>

Для получения оценки максимального правдоподобия составляется функция правдоподобия

$$L(a)=\prod_{i=1}^{n}\frac{a^{x_i}}{x_i!}e^{-an}=\frac{a^{\sum_{i=1}^{n}x_i}}{x_1!x_2!\ldots x_n!}e^{-an},$$

где n – объем выборки. Логарифмируя, получим

$$\ln L(a)=-\ln(x_i!x_2!\ldots x_n!)+\sum_{i=1}^{n}x_i\ln a-an.$$

Найдем максимум функции $\ln L(a)$. Используя необходимое условие экстремума, получим уравнение для нахождения оценки

$$\frac{d\ln L(a)}{da}=\sum_{i=1}^{n}x_i\frac{1}{a}-n=0,$$

откуда

$$\hat{a}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}=5,88.$$

Достаточное условие экстремума выполняется

$$\frac{d^2\ln L(a)}{da^2}=-\frac{\sum_{i=1}^{n}x_i}{a^2}<0.$$

2.2. Пусть случайная величина имеет нормальное распределение. Используя метод моментов и метод максимального правдоподобия получения точечных оценок, найти по выборке В значение оценки $\hat{\alpha}$, $\hat{\sigma}$ неизвестных параметров α, σ.

Метод моментов

Математическое ожидание распределения Пуассона $P(X=k)=\frac{a^k}{k!}e^{-a}$ равно МХ = a, где a – параметр распределения. Выборочное среднее $\bar{x}$ = -45,97 (см. задание 1). По методу моментов приравниваются теоретические моменты и их выборочные значения. В нашем случае $\hat{a}$ = -45,97.

Метод максимального правдоподобия

Для получения оценки максимального правдоподобия составляется функция правдоподобия

$$L(a)=\prod_{i=1}^{n}\frac{a^{x_i}}{x_i!}e^{-an}=\frac{a^{\sum_{i=1}^{n}x_i}}{x_1!x_2!...x_n!}e^{-an},$$

где n – объем выборки. Логарифмируя, получим

$$\ln L(a)=-\ln(x_i!x_2!...x_n!)+\sum_{i=1}^{n}x_i\ln a-an.$$

Найдем максимум функции $\ln L(a)$. Используя необходимое условие экстремума, получим уравнение для нахождения оценки

$$\frac{d\ln L(a)}{da}=\sum_{i=1}^{n}x_i\frac{1}{a}-n=0,$$

откуда

$$\hat{a}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}=-45,97.$$

Достаточное условие экстремума выполняется

$$\frac{d^2\ln L(a)}{da^2}=-\frac{\sum_{i=1}^{n}x_i}{a^2}<0.$$

Задание 3. Интервальное оценивание

3.1. Пусть генеральная совокупность имеет нормальное распределение. Найти доверительный интервалы для среднего значения α, дисперсии σ^2 и среднеквадратичного отклонения σ генеральной совокупности при доверительной вероятности $\gamma = 0{,}95$, если из генеральной совокупности сделана выборка В.

Для выборки В в задании 1 были получены несмещенная оценка математического ожидания $\bar{x}$=70,70 и смещенные оценки дисперсии S^2=16,91 и среднеквадратического отклонения S=4,11. Найдем несмещенные оценки дисперсии и среднеквадратического отклонения

$$\tilde{S}^2 = \frac{n}{n-1} S^2 = \frac{236}{235} \cdot 162{,}97 = 163{,}66.$$

$$\tilde{S} = \sqrt{\tilde{S}^2} = \sqrt{163{,}66} = 12{,}79\,.$$

Доверительный интервал для математического ожидания (среднего значения) $\hat{a}$ нормально распределенной случайной величины при неизвестной дисперсии σ^2 определяется из условия

$$P(X - t_\gamma \frac{\hat{S}}{\sqrt{n}} \le \hat{a} \le X + t_\gamma \frac{\hat{S}}{\sqrt{n}}) = \gamma,$$

где γ - заданная доверительная вероятность;

t_γ - квантили распределения Стьюдента.

Для выборки В объем n = 71. Из таблицы распределения Стьюдента [3] при доверительной вероятности γ = 0,95 при числе степеней свободы ν = n – 1 = 236 – 1 = 235 находим t_γ = 1,97 (интерполяция).

Точность оценки

$$\Delta = t_\gamma \frac{\hat{S}}{\sqrt{n}} = 1{,}97 \cdot \frac{12{,}79}{\sqrt{236}} = 1{,}64\,.$$

Тогда доверительный интервал для математического ожидания

$$\bar{x} - \Delta \le \hat{a} \le \bar{x} + \Delta\,,$$

$$-45{,}97 - 1{,}64 \le \hat{a} \le -45{,}97 + 1{,}64\,,$$

$$-47{,}61 \le \hat{a} \le -44{,}33\,.$$

Доверительный интервал для дисперсии $\hat{\sigma}^2$ определяется из условия

$$P\left\{\frac{(n-1)\hat{S}^2}{u_2} \le \hat{\sigma}^2 \le \frac{(n-1)\hat{S}^2}{u_1}\right\} = \gamma$$

Из таблицы χ^2 – распределения необходимо найти два числа u_1 и u_2, удовлетворяющих условию

$$P\,(u_1 \le \chi^2 \le u_2) = \gamma.$$

Так как таких пар чисел существует бесконечное множество, то используем условие симметрии по вероятности

$$P(\chi^2 < u_1) = P(\chi^2 > u_2).$$

Но $P(\chi^2 > u_2) = \alpha_2 = \frac{1}{2}(1-\gamma),\quad P(\chi^2 > u_1) = \alpha_1 = 1-\alpha_2 = 1-\frac{1}{2}(1-\gamma) = \frac{1}{2}(1+\gamma),$

Для выборки В $\nu = n - 1 = 236 - 1 = 235$. При $\gamma = 0{,}95$, $\alpha_1 = 0{,}975$, $\alpha_2 = 0{,}025$. По таблице χ - распределения $u_1 = 194{,}43$; $u_2 = 279{,}35$ (интерполяция). Тогда

$$\frac{(n-1)\tilde{S}^2}{u_2} \le \sigma^2 \le \frac{(n-1)\tilde{S}^2}{u_1}$$

или

$$\frac{235\cdot 163{,}66}{279{,}35} \le \sigma^2 \le \frac{235\cdot 163{,}66}{194{,}43},$$

$$137{,}68 \le \sigma^2 \le 197{,}81.$$

Доверительный интервал для среднеквадратического отклонения

$$\frac{\sqrt{n-1}\tilde{S}}{\sqrt{u_2}} \le \sigma \le \frac{\sqrt{n-1}\tilde{S}}{\sqrt{u_1}}$$

или $11{,}73 \le \hat{\sigma} \le 14{,}06$.

Задание 4. Статистическая проверка гипотез

4.1. По выборке А при уровне значимости $\alpha = 0{,}05$ проверить гипотезу о распределении Пуассона соответствующей генеральной совокупности.

При выполнении задания I для выборки А получено $\bar{x} = 5{,}88$, $S^2 = 5{,}35$. Найдем несмещенную оценку дисперсии

$$\tilde{S}^2 = \frac{n}{n-1}S^2 = \frac{95}{94}\cdot 5{,}35 = 5{,}41.$$

С уровнем значимости α = 0,05 проверим гипотезу о распределении Пуассона генеральной совокупности. Параметром распределения Пуассона

$$P(x = k) = \frac{a^k}{k!}e^{-a}$$

является величина a, при этом математическое ожидание и дисперсия MX = DX $= a$. Оценки среднего значения и дисперсии близки по значению, но не равны. В таблице распределения Пуассона ближайшим к ним значением a является 5.

Проверим гипотезу при a=6. Составим таблицу 5.

Объединим малочисленные частоты: (12,11,10,0,8) и соответствующие им теоретические частоты.

Таблица 5

Таблица для нахождения статистики Q^2 для выборки А при $a = 5$

i	x_i	n_i	p_i	$m_i = np_i$	$n_i - m_i$	$(n_i - m_i)^2$	$\frac{(n_i - m_i)^2}{m_i}$
1	0	1	0,0025	0,235	0,76	0,58	2,482
2	1	4	0,0149	1,413	2,59	6,69	4,737
3	2	2	0,0446	4,239	-2,24	5,01	1,182
4	3	6	0,0892	8,477	-2,48	6,14	0,724
5	4	14	0,1339	12,716	1,28	1,65	0,130
6	5	13	0,1606	15,259	-2,26	5,10	0,334
7	6	13	0,1606	15,259	-2,26	5,10	0,334
8	7	18	0,1377	13,079	4,92	24,21	1,851
8	8	14	0,1033	9,809	4,19	17,56	1,790
9	≥ 9	10	0,0688	6,540	3,46	11,97	1,831
Σ	-	95	1,0000	48	-	-	15,397

Статистика

$$Q^2 = \sum_{i=1}^{9}\frac{(n_i - m_i)^2}{m_i} = 15{,}397$$

имеет χ^2-распределение с числом степеней свободы ν = k – r – 1 = 9 – 1 – 1 = 7, где k – количество интервалов, r – число параметров распределения. Критическая область

$$P\,(Q^2 > \chi_\alpha^2) = \alpha = 0{,}005.$$

По таблице χ^2-распределения $\chi_\alpha^2 = 20{,}28$

Так как $Q^2 = 15{,}397$ не принадлежит критической области $Q^2 > 20{,}28$, то нет оснований отвергать гипотезу о распределении Пуассона генеральной совокупности.

4.2. По выборке В при уровне значимости $\alpha = 0{,}1$ проверить гипотезу о нормальном распределении соответствующей генеральной совокупности.

При решении задания I для выборки В получены следующие числовые характеристики:
$\overline{x} = -45{,}97$, $S^2 = 163{,}66$, $S = 12{,}79$.

Проверим гипотезу о нормальном законе распределения генеральной совокупности при уровне значимости $\alpha = 0{,}1$. Примем следующие оценки параметров нормального распределения $\hat{a} = -45{,}97$, $\sigma = 12{,}79$. В вариационном ряде (табл. 3) частоты первых и последних интервалов слишком малы, поэтому объединим их с соседними интервалами так, чтобы осталось 8 интервалов. Для вычисления теоретических вероятностей составим табл. 7.

Значения функции

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

были взяты из таблицы 2 [2] $u_i = \dfrac{x_i - a}{\sigma}$

$$P_i = \Phi\left(\frac{x_{i+1} - a}{\sigma}\right) - \Phi\left(\frac{x_i - a}{\sigma}\right).$$

Статистика

$$Q^2 = \sum_{i=1}^{k} \frac{(n_i - m_i)^2}{m_i}$$

имеет χ^2-распределение с $\nu = k - r - 1 = 10 - 2 - 1 = 7$ степенями свободы. Здесь $r = 2$, так как нормальное распределение имеет два параметра. Критическую область найдем из условия $P(Q^2 > \chi_\alpha^2) = 0{,}1$. По таблице χ^2-распределения $\chi_\alpha^2 = 12{,}017$ и критическая область будет $Q^2 > 9{,}236$.

Таблица 7

Интервалы	Частоты m_i	Нормированные интервалы $[u_i; u_{i+1})$		P_i	nP_i	$(m_i - n_i P_i)^2$	$\frac{(n_i - m_i)^2}{m_i}$
-84—-77	4	1	0,0025	0,008	1,81	4,82	2,67
-77—-70	4	4	0,0149	0,023	5,32	1,74	0,33
-70—-63	15	2	0,0446	0,061	14,50	0,25	0,02
-63—-56	22	6	0,0892	0,125	29,50	56,24	1,91
-56—-49	49	14	0,1339	0,190	44,82	17,49	0,39
-49—-42	54	13	0,1606	0,215	50,85	9,94	0,20
-42—-35	39	13	0,1606	0,183	43,08	16,63	0,39
-35—-28	33	18	0,1377	0,115	27,25	33,03	1,21
-28—-21	11	14	0,1033	0,055	12,87	3,51	0,27
-21—-14	5	10	0,0688	0,025	6,01	1,02	0,17
Σ	236	95	1,0000	1,000	-	-	7,54

Значение статистики $Q^2 = 7{,}54$ (табл. 8) не принадлежит критической области $Q^2 < 12{,}017$, поэтому гипотеза о нормальном распределении генеральной совокупности не отвергается.

Библиографический список

1. Колде Я.И. Практикум по теории вероятностей и математической статистики. – М.: Высшая математика, 1991 – 157 с.
2. Чудесенко В.Ф. Сборник заданий по специальным курсам высшей математики. – М.: Высшая школа, 1983 – 112с.
3. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: ВЦ АН СССР, 1968 – 475 С.

В Ы Б О Р К И

ВАРИАНТ 17

Выборка A17(n=95)

x	0	1	2	3	4	5	6	7	8	9	10	11	12
k	1	4	2	6	14	13	13	18	14	7	1	1	1

Выборка B17(n=236)

x	(-84;-77]	(-77;-70]	(-70;-63]	(-63;-56]	(-56;-49]
k	4	4	15	22	49
x	(-49;-42]	(-42;-35]	(-35;-28]	(-28;-21]	(-21;-14]
k	54	39	33	11	5