Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего образования

«Южно-Уральский государственный университет (национальный исследовательский университет)»

Филиал ФГАОУ ВО «ЮУрГУ (НИУ)» в г. Златоусте

Факультет Техники и технологии

Кафедра «Математика и вычислительная техника»

Индивидуальное домашнее задание

по дисциплине «Теория вероятностей и математическая статистика»

Выполнил:
студент группы ФТТ-207

_________________/Елисеев П.Е./

"5" Мая 2025г.

Проверил:
к.ф.-м.н., доцент

_______________/ Тарасова О.Ю./

"____"_________________20___г.

Златоуст 2025

Вариант 3

№6

Построить доверительный интервал для среднего значения (математического ожидания) случайной величины, распределенной по нормальному закону с неизвестными $M(X)$ и $\sigma^2(X)$ по данным выборки ($n = 50$).

Отсортировать выборку по возрастанию, составить интервальный статистический ряд, построить гистограмму частот, построить кумуляту; вычислить выборочное среднее, выборочную дисперсию, исправленную выборочную дисперсию, оценки моды и медианы. Найти моду и медиану графически. Вычислить асимметрию и эксцесс.

Используя критерий Пирсона, проверить гипотезу о том, что генеральная совокупность X распределена нормально. Уровень значимости $\alpha = 0,05$.

–0,517	0,465	0,568	0,217	0,212
–0,639	0,743	0,826	1,049	0,459
0,068	–0,454	1,616	1,398	1,729
0,311	0,617	0,915	1,191	0,379
0,521	0,214	0,554	2,440	–0,840
1,146	0,475	0,966	0,168	–0,591
0,627	1,133	0,281	2,635	0,411
0,111	0,460	–0,494	0,460	1,201
0,294	0,448	0,822	1,310	0,372
–0,279	0,545	2,376	0,002	0,499

Решение:

Сначала отсортировали выборку по возрастанию и вычислили количество интервалов, на которые нужно разбить выборку.

Количество интервалов k найдем по формуле:

$$k = 1 + \log_{10} N,$$

где N – объем выборки.

$$k = 1 + \log_{10} 50$$

$$k = 6,64397$$

Округляем до 7. Получаем $k = 7$

Найдем шаг интервала h по формуле:

$$h = \frac{x_{max} - x_{min}}{k}$$

$$h = \frac{2{,}635 - (-0{,}84)}{7}$$

$$h = 0{,}496$$

-0,84		k=	6,64397835
-0,639		k=	7
-0,591		h=	0,49642857
-0,521			0,496

Построим интервальный ряд:

	x_i	x_{i+1}
1	-0,840	-0,344
2	-0,344	0,152
3	0,152	0,648
4	0,648	1,144
5	1,144	1,640
6	1,640	2,136
7	2,136	2,635

Теперь найдем относительные частоты n_i . Будем находить по количеству значений, входящих в интервал $x_i - x_{i+1}$.

Например, $n_1 = 10$

-0,84
-0,639
-0,591
-0,521
-0,52
-0,519
-0,518
-0,517
-0,494
-0,454

Получаем:

	x_i	x_{i+1}	n_i
1	-0,840	-0,344	10
2	-0,344	0,152	4
3	0,152	0,648	19
4	0,648	1,144	7
5	1,144	1,640	6
6	1,640	2,136	1
7	2,136	2,635	3

Объем выборки n = 50. Из суммы n_i должны получить 50.

$$n = \sum n_i = 10 + 4 + 19 + 7 + 6 + 1 + 3 = 50$$

Далее найдем отношения $\frac{n_i}{h}$ для построения гистограммы частот.

n_i/h
20,161
8,065
38,306
14,113
12,097
2,016
6,048

Далее вычислим X_i^*– принятое в качестве варианты среднее арифметическое концов интервала. Будет находится по формуле:

$X_i^* = \frac{X_i + X_{i+1}}{2}$.

Например, для i = 1:

$X_1^* = \frac{X_1 + X_2}{2} = \frac{-0,84 + (-0,344)}{2} = -0,592.$

x_i^*
-0,592
-0,096
0,400
0,896
1,392
1,888
2,386

Далее вычислим $X_i^* \cdot n_i$– произведение варианты на относительную частоту:

Например, для i = 1: $X_1^* \cdot n_1 = -0,592 * 10 = -5,920.$

Сумма $X_i^* \cdot n_i$:

$$\sum X_i^* \cdot n_i = 24,965$$

$x_i^* * n_i$
-5,920
-0,384
7,600
6,272
8,352
1,888
7,157
24,965

Вычислим $(X_i^*)^2 \cdot n_i$ – произведение квадрата варианты на относительную частоту:

Например, для i = 1: $(X_1^*)^2 \cdot n_1 = -5,920^2 * 10 = 3,505$

Сумма $(X_i^*)^2 \cdot n_i =$

$$\sum (X_i^*)^2 \cdot n_i = 44,464$$

$(x_i^*)^2 * n_i$
3,505
0,037
3,040
5,620
11,626
3,565
17,072
44,464

Вычислим выборочное среднее ($\overline{X^*}$), которое вычисляется по формуле: $\overline{X^*} = \frac{\sum X_i^* \cdot n_i}{n}$:

$$\overline{X^*} = \frac{24{,}965}{50} = 0{,}499$$

Вычислим выборочное среднее квадратическое отклонение (выборочная дисперсия σ), которая вычисляется по формуле: $\sigma = \frac{\sum (X_i^*)^2 \cdot n_i}{n}$:

$$\sigma = \frac{44{,}464}{50} = 0{,}889$$

Вычислим корень из выборочной дисперсии:

$\sqrt{\sigma} = \sqrt{0{,}889} = 0{,}94301.$

Вычислим исправленную дисперсию S^2 – вычисляется по формуле: $S^2 = \frac{n}{n-1} \cdot \sigma$

$$S^2 = \frac{50}{50-1} \cdot 0{,}889 = \frac{50}{49} \cdot 0{,}889 = 0{,}907$$

Вычислим корень из исправленной дисперсии $S = \sqrt{S^2}$

$$S = \sqrt{0{,}907} = 0{,}953$$

Найдем доверительный интервал $\left(\overline{X^*} - \frac{t_\gamma \cdot S}{\sqrt{n}}; \overline{X^*} + \frac{t_\gamma \cdot S}{\sqrt{n}}\right)$, где t_γ будем искать по таблице значений из приложения 3 учебника. Тогда при $\gamma = 0{,}95$, $t_\gamma = 2{,}009$

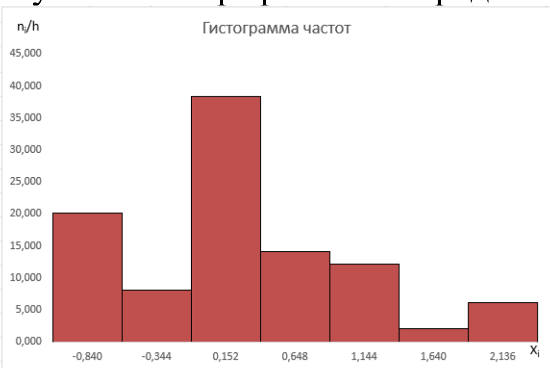
1) $0{,}499 - \frac{2{,}009 \cdot 0{,}953}{\sqrt{50}} = 0{,}229$

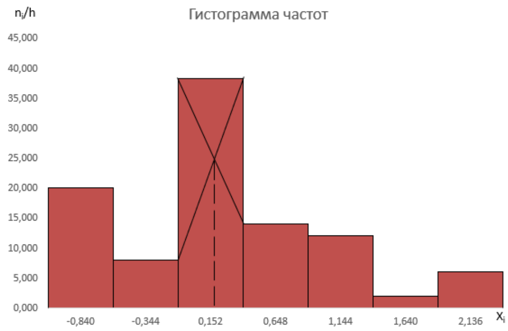
2) $0{,}499 + \frac{2{,}009 \cdot 0{,}953}{\sqrt{50}} = 0{,}770$

Найденный доверительный интервал: (0,229; 0,770).

Выборочное среднее:			0,499	
Выборочная дисперсия:			0,889	0,94301
Исправленная дисперсия S^2:			0,907	
		S =	0,953	
Доверительный интервал:			0,229	0,770

Построим гистограмму частот и графически определим моду.





Мода лежит в промежутке от –0,344 до 0,152. Вычислим моду по формуле:

$$M_0 = x_0 + h \cdot \frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})}$$

где:

x_0 – нижняя граница интервала, где находится мода;

f_m – частота интервала, где находится мода;

f_{m-1} – частота интервала, предшествующего тому, где находится мода.

f_{m+1} – частота интервала, после того, где находится мода.

$$M_0 = -0{,}344 + 0{,}496 \cdot \frac{4 - 10}{(4 - 10) + (4 - 19)}$$

$$M_0 = -0{,}202$$

-0,344<= M0<= 0,152

M_0= -0,202

Вычисленное значение моды сходится со значением на графике.

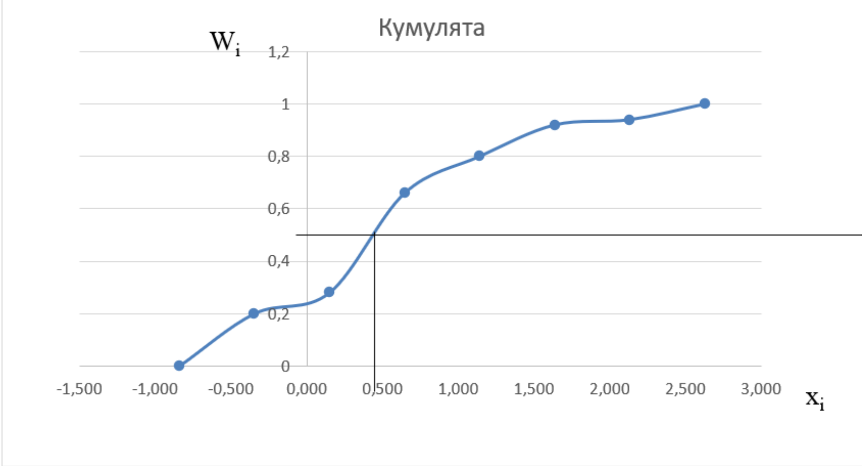
Далее по кумуляте графически найдем медиану. Для этого составим расчетную таблицу:

x_i	n_i	s_i	Wi
-0,840	0	0	0
-0,344	10	10	0,2
0,152	4	14	0,28
0,648	19	33	0,66
1,144	7	40	0,8
1,640	6	46	0,92
2,136	1	47	0,94
2,635	3	50	1

где s_i – сумма текущей относительной частоты n_i и предыдущей;

$W_i = \frac{s_i}{N}$ – отношение относительных частот к объему выборки

Медиана находится следующим образом: ординату, пропорциональную сумме всех частот делим пополам. Далее из точки пересечения с графиком проводим абсциссу.



Из графика видно, что медиана лежит в промежутке от 0 до 0,500. Ближе к 0,500. Вычислим медиану по формуле:

$$Me = x_0 + h \cdot \frac{\frac{n}{2} - S_{m-1}}{f_m}$$

где S_{m-1} – сумма частот всех предшествующих интервалов.

$$Me = 0{,}152 + 0{,}496 \cdot \frac{25 - 14}{19}$$

$$Me = 0{,}43916$$

Значение медианы сходится со значением, определенным графически.

Найдем интервалы $(Z_i; Z_{i+1})$. Для этого объединим интервалы:

1 с 2 и интервалы 5, 6, 7.

Получим:

x_i	x_{i+1}	n_i
-0,840	0,152	14
0,152	0,648	19
0,648	1,144	7
1,144	2,6350	10
		50

Далее произведем аналогичные вычисления, что и ранее.

Получим:

x_i	x_{i+1}	n_i	x_i^*	$x_i^* * n_i$	$(x_i^*)^2 * n_i$	
-0,840	0,152	14	-0,344	-4,816	1,656704	
0,152	0,648	19	0,400	7,6	3,04	
0,648	1,144	7	0,896	6,272	5,619712	
1,144	2,6350	10	1,890	18,895	35,7021	
		50		27,951	46,01852	
		Выборочное среднее:			0,55902	
		Выборочная дисперсия:			0,9204	0,95936

Интервалы $(Z_i; Z_{i+1})$ находятся по следующим формулам:

$$Z_i = \frac{X_i - \overline{X^*}}{\sqrt{\sigma}}$$

$$Z_{i+1} = \frac{X_{i+1} - \overline{X^*}}{\sqrt{\sigma}}$$

Левый конец первого интервала примем равным $-\infty$, а правый конец последнего интервала примем равным $+\infty$.

Тогда, например для i = 2, получим:

$$Z_2 = \frac{X_2 \cdot \overline{X^*}}{\sqrt{\sigma}} = \frac{0{,}152 - 0{,}55902}{0{,}95936} = -0{,}368$$

Найдем значения функций $\Phi(Z_i)$ и $\Phi(Z_{+1})$ по таблице значений в приложении 2. Так как функция нечетная, то при отрицательных значениях Z берем отрицательное значение функции. Левый конец первого интервала будет равен -0,5, а правый конец последнего интервала будет равен 0,5:

Например, для i = 2:

$$\Phi(-0{,}368) = -0{,}1443$$

Найдем теоретические вероятности P_i и теоретические частоты n_i':

$$P_i = \Phi(Z_{i+1}) - \Phi(Z_i)$$

$$n_i' = P_i \cdot n$$

Тогда, например, для i = 2, получим:

$$P_2 = -0{,}0636 - (-0{,}1443) = -0{,}2823 + 0{,}1443 = 0{,}208$$

$$n_2' = P_2 \cdot 50 = 0{,}208 \cdot 50 = 10{,}395$$

Для проверки, вычислим сумму теоретических вероятностей P_i. Она должна быть равна 1:

$$\sum P_i = 1$$

Вычислим сумму теоретических частот n_i'. Она должна быть равна объему выборки n = 50:

$$\sum n_i' = 50$$

Z_i	Z_{i+1}	$\Phi(Z_i)$	$\Phi(Z_{i+1})$	P_i	n_i'
$-\infty$	-0,368	-0,5	-0,1443	0,356	17,785
-0,368	0,158	-0,1443	0,0636	0,208	10,395
0,158	0,684	0,0636	0,2517	0,188	9,405
0,684	$+\infty$	0,2517	0,5	0,248	12,415
				1,000	50

Теперь сравним эмпирически и теоретические частоты, используя критерий Пирсона:

1) Вычислим наблюдаемое значение критерия Пирсона по формуле:

$$\chi^2_{\text{набл}} = \sum \frac{(n_i - n_i')^2}{n_i'}$$

Тогда получим:

$$\chi^2_{\text{набл}} = 9{,}0135$$

2) По таблице критических точек распределения χ^2(приложение 5 учебника), по уровню значимости $\alpha = 0{,}05$ и числу степеней свободы $k = s - 3 = 4 - 3 = 1$, (где s – число интервалов) находим критическую точку правосторонней области:

$$\chi^2_{\text{кр}}(0{,}05; 1) = 3{,}8$$

Z_i	Z_{i+1}	$\Phi(Z_i)$	$\Phi(Z_{i+1})$	P_i	n_i'	$(n_i-n_i')^2/n_i'$	
$-\infty$	-0,368	-0,5	-0,1443	0,356	17,785	0,8055	
-0,368	0,158	-0,1443	0,0636	0,208	10,395	7,1232	
0,158	0,684	0,0636	0,2517	0,188	9,405	0,615	
0,684	$+\infty$	0,2517	0,5	0,248	12,415	0,4698	
				1,000	50	9,0135	
k=	1					хи-квадрат критическое	
						3,8	

Так как $\chi^2_{\text{набл}} > \chi^2_{\text{кр}}$, тогда есть основания отвергнуть гипотезу о нормальном распределении генеральной совокупности X. Это означает, что эмпирические и теоретические частоты различаются значимо, и данные наблюдений не согласуются с гипотезой о нормальном распределении генеральной совокупности.

Для проверки вычислим асимметрию и эксцесс.

Асимметрия находится по формуле:

$$A = \frac{\mu_3}{\sigma^3}$$

где μ_3 – центральный момент 3-го порядка.

σ – среднее квадратическое отклонение.

Центральный момент 3-го порядка найдем по формуле:

$$\mu_3 = \frac{\sum n_i \cdot (x_i - x_i^*)^3}{n}$$

Получим:

$$A = \frac{0{,}26868}{0{,}9204^3}$$

$$A = 0{,}34463$$

Асимметрия=	-0,9030	-0,7364	-10,3091	0,26868
	-0,1590	-0,0040	-0,0764	0,34463
	0,3370	0,0383	0,2679	
	1,3305	2,3552	23,5519	
	-0,5590	-0,17		

Теперь вычислим эксцесс. Он находится по формуле:

$$E = \frac{\mu_4}{\sigma^4}$$

где μ_4 – центральный момент 4-го порядка.

Центральный момент 4-го порядка найдем по формуле:

$$\mu_3 = \frac{\sum n_i \cdot (x_i - x_i^*)^4}{n}$$

Получим:

$$E = \frac{0{,}81494}{0{,}9204^4}$$

$$E = 1{,}13573$$

Эксцесс=	-0,9030	0,6650	9,3093	0,81494
	-0,1590	0,0006	0,0121	1,13573
	0,3370	0,0129	0,0903	
	1,3305	3,1335	31,3353	

Ответ: так как асимметрия и эксцесс велики, можно сделать вывод, что эмпирические и теоретические частоты действительно различаются значимо, и данные наблюдений не согласуются с гипотезой о нормальном распределении генеральной совокупности.