Министерство образования Республики Беларусь
Учреждение образования
«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ»

Индивидуальная практическая работа №4

по дисциплине

“ Теория вероятностей и математическая статистика “

Вариант № 07

Выполнил: Малащицкая Е.М

студент группы 114351

Минск 2023

Задача 10. Обработка одномерной выборки

Условие задачи

По выборке одномерной случайной величины:

- получить вариационный ряд;

- построить на масштабно-координатной бумаге формата А4 график эмпирической функции распределения $F^*(x)$;

- построить гистограмму равно интервальным способом;

- построить гистограмму равновероятностным способом;

- вычислить точечные оценки математического ожидания и дисперсии;

- вычислить интервальные оценки математического ожидания и дисперсии (γ = 0,95);

- выдвинуть гипотезу о законе распределения случайной величины и проверить ее при помощи критерия согласия χ^2 и критерия Колмогорова (α = 0,05). График гипотетической функции распределения $F_0(x)$ построить совместно с графиком $F^*(x)$ в той же системе координат и на том же листе.

Решение.

1,39
1,55
1,8
1,97
2,15
2,19
2,22
2,37
2,41
2,45
2,47
2,62
2,64
3,05
3,11
3,3

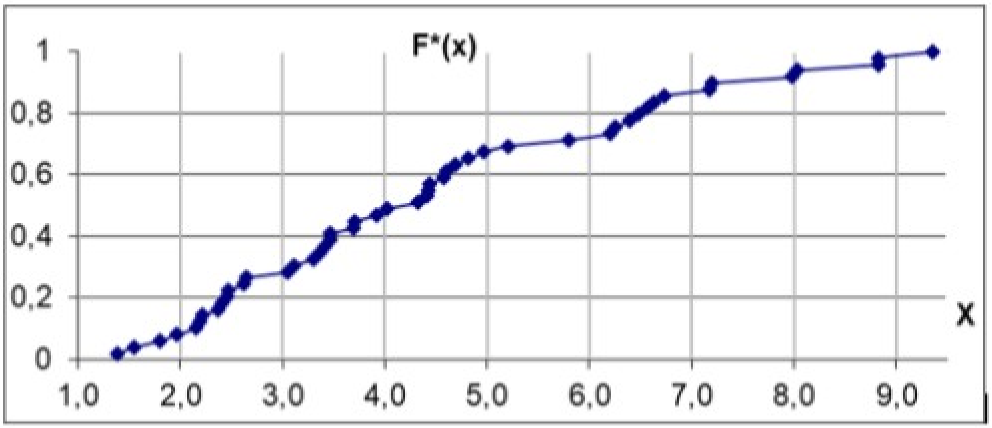
3,36
3,42
3,46
3,47
3,69
3,71
3,92
4,02
4,32
4,41
4,43
4,44
4,58
4,6
4,69
4,82
4,96
5,21
5,8
6,2
6,25
6,39
6,48
6,57
6,63
6,73
7,17
7,2
7,98
8,03

8,82
8,82
9,35

Эмпирическая функция распределения случайной величины X равна частоте того, что X примет значение меньшее, чем аргумент функции x, и определяется формулой

$$F^*(x) = p^*(X < x) = \begin{cases} 0, & x \le \widehat{x_1}, \\ \vdots & \\ \dfrac{i}{n}, & \widehat{x_i} < x \le \widehat{x_{i+1}} \\ \vdots & \\ 1, & x > \widehat{x_n}. \end{cases}$$

При $n \to \infty$ эмпирическая функция распределения $F^*(x)$ сходится по вероятности к теоретической функции распределения $F(x)$.



Количество интервалов *M*, необходимое для построения гистограмм, определим по объему выборки:

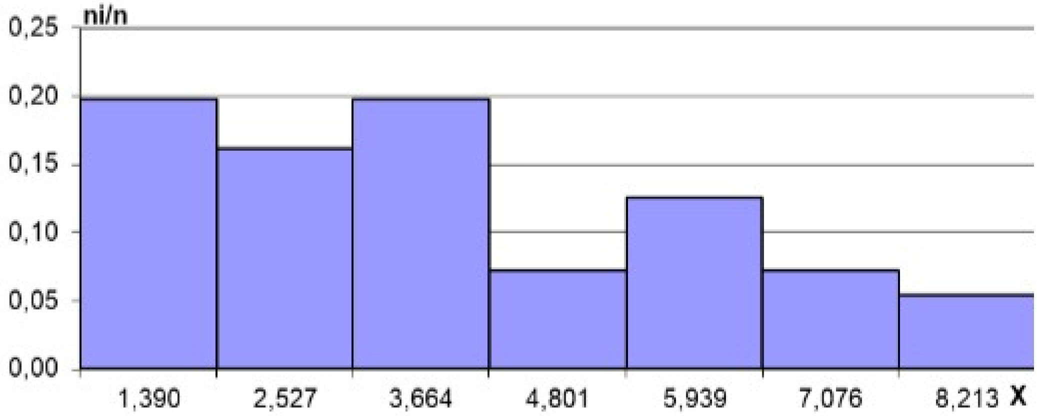
$$M \approx \sqrt{n} = \sqrt{49} = 7.$$

Шаг интервала

$$h = \frac{x_{max} - x_{min}}{7}$$

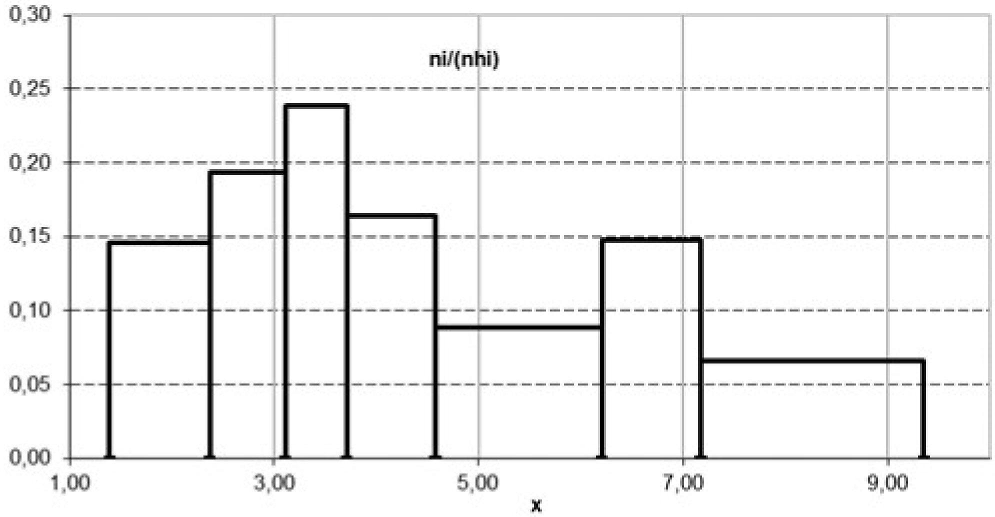
h= (9,35-1,39)/7=1,137

A_i	B_i	h_i	ν_i	p_i *	$f^*_i = \frac{\nu_i}{nh_i}$	p_x
1,390	2,527	1,137	11	0,224	0,1974	0,224
2,527	3,664	1,137	9	0,184	0,1615	0,408
3,664	4,801	1,137	11	0,224	0,1974	0,633
4,801	5,939	1,137	4	0,082	0,0718	0,714
5,939	7,076	1,137	7	0,143	0,1256	0,857
7,076	8,213	1,137	4	0,082	0,0718	0,939
8,213	9,350	1,137	3	0,061	0,0538	1,000



гистограмма, построенная равно интервальным способом

A_i	B_i	h_i	ν_i	p_i *	$f^*_i = \frac{\nu_i}{nh_i}$
1,39	2,37	0,98	7	0,1429	0,146
2,37	3,11	0,74	7	0,1429	0,193
3,11	3,71	0,60	7	0,1429	0,238
3,71	4,58	0,87	7	0,1429	0,164
4,58	6,20	1,62	7	0,1429	0,088
6,20	7,17	0,97	7	0,1429	0,147
7,17	9,35	2,18	7	0,1429	0,066



гистограмма построенная равновероятностным способом;

Вычислим точечную оценку математического ожидания по формуле:

$$m_X^* = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 4,523$$

Вычислим точечную оценку дисперсии по формуле:

$$D_X^* = S_0^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\overline{x}^2 = 4,457$$

.

Построим доверительный интервал для математического ожидания с надежностью γ = 0,95. Для этого в таблице функции Лапласа найдем значение, равное $\frac{\gamma}{2}$ = 0,475, и определим значение аргумента, ему соответствующее: $z_{0,95} = \arg\Phi(0,475) = 1,96$.

Затем вычислим $z_{0,95}\frac{S_0}{\sqrt{n}} = 0,591$ и получим доверительный интервал для математического ожидания:

$$I_{0,95}(m_X) = \left[3,932; 5,114\right].$$

Построим доверительный интервал для дисперсии с надежностью γ = 0,95 по формуле.

Вычислим $z_{0,95}\cdot\sqrt{\frac{2}{n-1}}\cdot S_0^2 = 0,364$ и получим доверительный интервал для дисперсии:

$$I_{0,95}(D_x) = \left[4,093; 4,821\right].$$

По виду гистограммы выдвинем гипотезу о равномерном распределении СВХ. Проверим гипотезу о равномерном распределении СВХ при помощи критерия χ^2

H_0: $f(x) = f_0(x)$ F(x)=F_0(x)

H_1: $f(x) \neq f_0(x)$ F(x)≠F_0(x)

Где $f_0(x)$,F_0(x)– теоретическая плотность и функция распределения

$$f_0(x)=\begin{cases}0; x<a^*\\ \dfrac{1}{b^*-a^*}\\ 0; x>b^*\end{cases}\quad F_0(x)=\begin{cases}0; x<a^*\\ \dfrac{x-a^*}{b^*-a^*}\\ 1; x>b^*\end{cases}$$

Где $a^*=\bar{X}-\sqrt{3}S_0=4{,}523-\sqrt{3\cdot 4{,}457}=0{,}866$ $b^*=\bar{X}+\sqrt{3}S_0=4{,}523+\sqrt{3\cdot 4{,}457}=8{,}179$

$$f_0(x)=\begin{cases}0; x<0{,}866\\ \dfrac{1}{8{,}179-0{,}866}; 0{,}866\le x\le 8{,}179\\ 0; x>8{,}179\end{cases}\quad f_0(x)=\begin{cases}0; x<0{,}866\\ 0{,}137; 0{,}866\le x\le 8{,}179\\ 0; x>8{,}179\end{cases}$$

$$F_0(x)=\begin{cases}0; x<0{,}866\\ \dfrac{x-0{,}866}{8{,}179-0{,}866}; 0{,}866\le x\le 8{,}179\\ 0; x>8{,}179\end{cases}\quad F_0(x)=\begin{cases}0; x<0{,}866\\ \dfrac{x-0{,}866}{7{,}313}; 0{,}866\le x\le 8{,}179\\ 0; x>8{,}179\end{cases}$$

$\chi^2=n\sum\dfrac{(p_i-p_i^*)^2}{p_i^*}$

$p_i^*=0{,}143$

$n_i'=49*0{,}143=7$

v_i	A_i	B_i	p_i^*	p_i	$\frac{(p_i-p_i^*)^2}{p_i^*}$
11	1,390	2,527	0,1429	0,2245	0,0466
9	2,527	3,664	0,1429	0,1837	0,0117
11	3,664	4,801	0,1429	0,2245	0,0466
4	4,801	5,939	0,1429	0,0816	0,0262
7	5,939	7,076	0,1429	0,1429	0,0000
4	7,076	8,213	0,1429	0,0816	0,0262
3	8,213	9,350	0,1429	0,0612	0,0466
		сумма	1,0000	1,0000	**0,2041**

$\chi^2=49*0.2041=10{,}0$

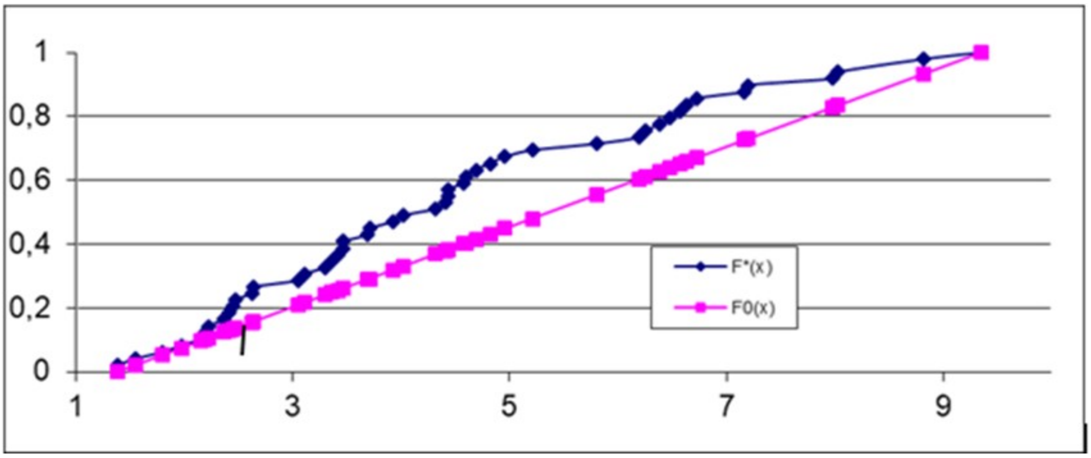
По таблице найдем критическое значение критерия $\chi^2_{кр}(5;0{,}05)=11{,}1$, так как $\chi^2_{кр}>\chi^2$ то гипотеза о равномерном распределении СВ X принимается.

Проверим гипотезу о равномерном распределении СВ X при помощи критерия Колмогорова

H_0: $F(x)=F_0(x)$

H_1: $F(x)\neq F_0(x)$

Где $F_0(x)$– теоретическая функция распределения



По графику определим максимальное по модулю отклонение между функциями $F^*(x)$ и $F_0(x)$:

$$Z = \max_{i=1}^{n} \left| F^*(x_i) - F_0(x_i) \right| = 0,22$$

Вычислим значение критерия Колмогорова по формуле:

$$\lambda = \sqrt{n} \cdot Z = \sqrt{49} \cdot 0,22 = 1,54.$$

Из таблицы Колмогорова по заданному уровню значимости $\alpha = 0,05$ выбираем критическое значение $\lambda_\gamma = \lambda_{1-\alpha} = \lambda_{0,95} = 1,36.$

Так как $\lambda = 1,54 > \lambda_{0,95} = 1,36$, то гипотезу H_0 о равномерном законе распределения не принимаем.

Задача 11. Обработка *двухмерной выборки*

Условие ***задачи***

По выборке двухмерной случайной величины:
- вычислить точечную оценку коэффициента корреляции;
- вычислить интервальную оценку коэффициента корреляции ($\gamma = 0,95$);
- проверить гипотезу об отсутствии корреляционной зависимости;
- вычислить оценки параметров a_0 и a_1 линии регрессии $\overline{y}(x) = a_0^* + a_1^* x$;
- построить диаграмму рассеивания и линию регрессии.

Решение.

Состоятельная оценка коэффициента корреляции

$$R_{XY}^* = \frac{\frac{n}{n-1}\left(\overline{XY} - \overline{X}\,\overline{Y}\right)}{s_X s_Y}$$

Расчетная таблица

	X	Y	X*Y	X^2	Y^2
	-11,95	10,25	-122,4875	142,802 5	105,0625
	-6,99	6,28	-43,8972	48,8601	39,4384

	-5,34	3,56	-19,0104	28,5156	12,6736
	-4,37	4,32	-18,8784	19,0969	18,6624
	-5,29	4,57	-24,1753	27,9841	20,8849
	-8,53	7,53	-64,2309	72,7609	56,7009
	-10,68	8,91	-95,1588	114,0624	79,3881
	-7,14	6,66	-47,5524	50,9796	44,3556
	-9,2	8,44	-77,6480	84,6400	71,2336
	-7,69	5,83	-44,8327	59,1361	33,9889
	-0,18	-0,6	0,108	0,0324	0,36
	-2,88	3,02	-8,6976	8,2944	9,1204
	-4,76	3,8	-18,088	22,6576	14,44
	-9,47	8,95	-84,7565	89,6809	80,1025
	-3,39	3,09	-10,4751	11,4921	9,5481
	-1,96	1,77	-3,4692	3,8416	3,1329
	-1,8	1,34	-2,412	3,24	1,7956
	-4,81	3,75	-18,0375	23,1361	14,0625
	-5,2	2,92	-15,184	27,04	8,5264
	0,45	-1,26	-0,567	0,2025	1,5876
	-2,73	1,24	-3,3852	7,4529	1,5376
	-4,68	3,76	-17,5968	21,9024	14,1376
	-3,86	2,85	-11,001	14,8996	8,1225
	1,69	-2,69	-4,5461	2,8561	7,2361
	-2,08	0,8	-1,664	4,3264	0,64
сумма	**-122,84**	**99,09**	**-757,6436**	**889,8932**	**656,7387**
среднее	**-4,9136**	**3,9636**	**-30,3057**	**35,5957**	**26,2695**

$\overline{X}$ =-4,9136

$\overline{Y}$ =3,9636

$\overline{XY}$ =-30,3057

$\overline{X^2}$ =35,5957

$\overline{Y^2}$ =26,2695

$$S_X = \sqrt{\frac{n}{n-1}\left(\overline{X^2} - (\overline{X})^2\right)}$$

$$s_Y = \sqrt{\frac{n}{n-1}\left(\overline{Y^2} - (\overline{Y})^2\right)}$$

$$S_Y = \sqrt{\frac{25}{25-1}\left(26,2695 - (3,9636)^2\right)}$$ =3,317

$$S_x = \sqrt{\frac{25}{25-1}\left(35,5957 - (-4,9136)^2\right)}$$ =3,454

дисперсия

$S_x^2 =$ 11,929

S_y^2 =10,999

Состоятельная оценка коэффициента корреляции

$$R^*_{XY}=\frac{(-30{,}3057-(-4{,}9136)\cdot(3{,}9636))\frac{25}{25-1}}{3{,}317\cdot 3{,}454}=-0{,}985$$

Уравнение регрессии имеет вид

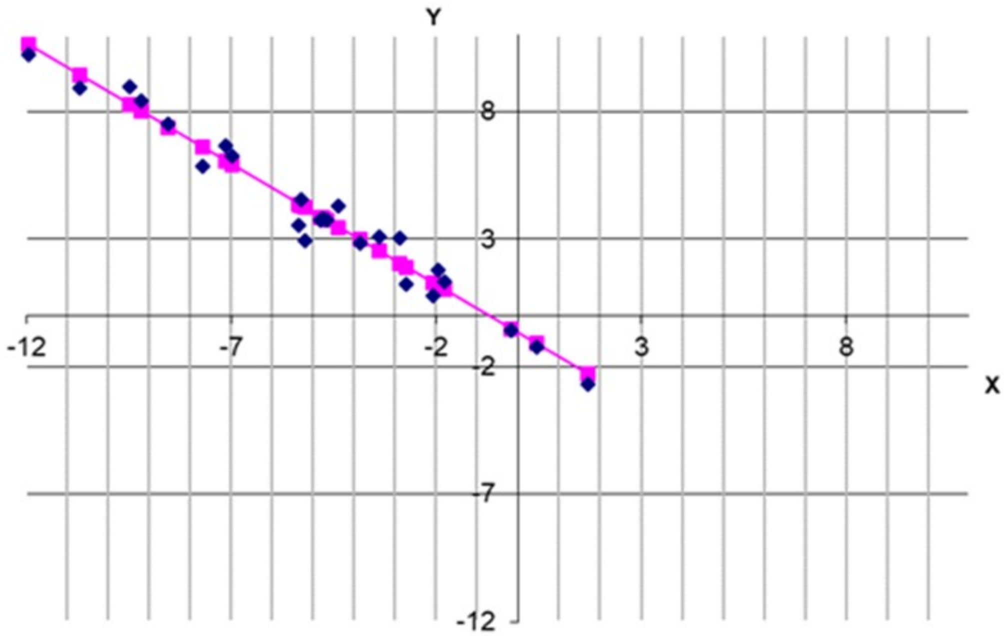
$$y-\bar{Y}=R^*_{XY}\frac{S_y}{S_X}(x-\bar{X})$$

Y-3,9636=-0,985*(3,317/3,454)(x+4,9136)

y=-0,946x-0,683

a_0=-0,946

a_1=-0,683



проверим значимость коэффициента корреляции, при помощи критерия t

H_0: $R^*_{XY}=0$

H_1: $R^*_{XY}\neq 0$

$$t=\frac{R^*_{XY}\sqrt{n-2}}{\sqrt{1-R^{*2}_{XY}}}$$

$$t=\frac{-0{,}985\sqrt{25-2}}{\sqrt{1-0{,}985^2}}=-27{,}24$$

по таблице найдем критическое значение $T_{кр}(0{,}05;23)=2{,}09$, так как $|t|>T_{кр}$ то коэффициент корреляции значим.

Доверительный интервал для коэффициента корреляции

$$I_y(R_{XY})=\left[\frac{e^{2a}-1}{e^{2a}+1};\frac{e^{2b}-1}{e^{2b}+1}\right]$$

Где

$$a=0.5\cdot\ln\left(\frac{1+R_{XY}}{1-R_{XY}}\right)-\frac{z_y}{\sqrt{n-3}}$$

$$b = 0.5 \cdot \ln\left(\frac{1 + R_{XY}}{1 - R_{XY}}\right) + \frac{z_\gamma}{\sqrt{n - 3}}$$

$$z_\gamma = arc\Phi\left(\frac{\gamma}{2}\right)$$

Для $\gamma = 0,95$ $z_\gamma = 1,96$

$$a = 0.5 \cdot \ln\left(\frac{1 - 0,985}{1 + 0,985}\right) - \frac{1,96}{\sqrt{25 - 3}} =$$ -2,8556

$$b = 0.5 \cdot \ln\left(\frac{1 - 0,985}{1 + 0,985}\right) + \frac{1,96}{\sqrt{25 - 3}} =$$ -2,0198

$$I_\gamma(R_{XY}) = \left[\frac{e^{2(-2,8556)} - 1}{e^{2(-2,8556)} + 1}; \frac{e^{2(-2,0198)} - 1}{e^{2(-2,0198)} + 1}\right] = [-0,993; -0,965]$$