ФИО студента

Личный номер студента

Номер варианта (15)

23.1 Данные выборочного обследования жилищных условий жителей района:

Площадь жилья на 1 чел,$м^2$	до 5,0	5,0-10,0	10,0-15,0	15,0-20,0	20,0-25,0	25,0-30,0	30,0 и более	Всего
Число жителей	8	95	204	270	210	130	83	1000

1) составить эмпирическую функцию распределения случайной величины $X -$ площадь жилья на одного человека и построить её график;

2) найти границы, в которых с вероятностью 0,98 заключен средний размер площади на одного человека в целом по району.

23.2 По данным задачи 1, используя $\chi^2 -$ критерий Пирсона, на уровне значимости $\alpha = 0{,}05$ проверить гипотезу о том, что случайная величина $X -$ размер площади на одного человека в целом по району – распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.

23.3 Совместное распределение двух случайных факторов X и Y представлено в таблице

$X\|Y$	42	44	46	48
1.3	5	2	-	-
1.4	-	7	3	1
1.5	-	3	8	2
1.6	-	1	3	5

Необходимо:

1. Вычислить групповые средние $\overline{y_i}$, построить эмпирическую линию регрессии Y на X.

2. Предполагая, что между переменными X и Y существует линейная корреляционная зависимость:

1) найти уравнение прямой регрессии Y на X, построить её график на одном чертеже с эмпирической линией регрессии;

2) вычислить выборочный коэффициент корреляции и сделать вывод о тесноте и направлении связи между переменными X и Y;

3) используя полученное уравнение регрессии вычислить прогнозное значение Y при росте фактора X на 10% от среднего значения.

23.1 Данные выборочного обследования жилищных условий жителей района:

Площадь жилья на 1 чел,м^2	до 5,0	5,0-10,0	10,0-15,0	15,0-20,0	20,0-25,0	25,0-30,0	30,0 и более	Всего
Число жителей	8	95	204	270	210	130	83	1000

1) составить эмпирическую функцию распределения случайной величины X — площадь жилья на одного человека и построить её график;

2) найти границы, в которых с вероятностью 0,98 заключен средний размер площади на одного человека в целом по району.

Решение.

1) Прейдем к точечному выборочному распределению, взяв за значения признака середины частичных интервалов.

x_i	2,5	7,5	12,5	17,5	22,5	27,5	32,5
n_i	8	95	204	270	210	130	83
w_i	0,008	0,095	0,204	0,270	0,210	0,130	0,083
p_i	0,008	0,103	0,307	0,577	0,787	0,917	1

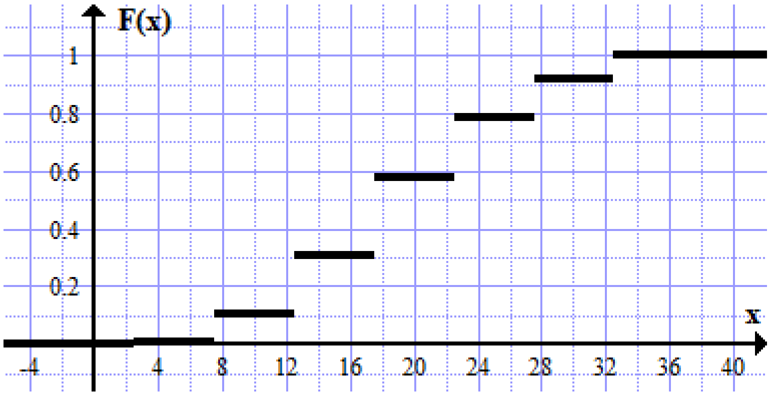
n_i — частоты

$w_i = n_i/n$ — относительные частоты

p_i — накопленные частоты

Эмпирическая функция распределения $F(x)$ в зависимости от значения вариант x_i равна соответствующей накопленной относительной частоте p_i. Тогда эмпирическая функция распределения $F(x)$ имеет вид:

$$F(x) = \begin{cases} 0 & x \leq 2{,}5 \\ 0{,}008 & 2{,}5 < x \leq 7{,}5 \\ 0{,}103 & 7{,}5 < x \leq 12{,}5 \\ 0{,}307 & 12{,}5 < x \leq 17{,}5 \\ 0{,}577 & 17{,}5 < x \leq 22{,}5 \\ 0{,}787 & 22{,}5 < x \leq 27{,}5 \\ 0{,}917 & 27{,}5 < x \leq 32{,}5 \\ 1 & x > 32{,}5 \end{cases}$$



2) Вычислим числовые характеристики.

Для удобства вычисления составим расчетную таблицу:

x_i	n_i	$x_i n_i$	${x_i}^2 n_i$
2,5	8	20	50
7,5	95	712,5	5343,75
12,5	204	2550	31875
17,5	270	4725	82687,5
22,5	210	4725	106312,5
27,5	130	3575	98312,5
32,5	83	2697,5	87668,75
122,5	1000	19005	412250

$$\bar{x} = \frac{1}{n}\sum x_i n_i$$

$$\bar{x} = \frac{1}{1000} * 19005 = 19{,}005$$

$$\sigma^2 = \overline{x^2} - (\bar{x})^2;$$

$$\sigma^2 = \frac{1}{1000} * 412250 - (19{,}005)^2 = 51{,}06$$

Средняя ошибка выборки для среднего значения признака составляет:

$$\mu = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{51{,}06}{1000}} \approx 0{,}226.$$

Предельная ошибка выборки составляет:

$$\Delta = t \cdot \mu,$$

где t – коэффициент доверия; для доверительной вероятности $\gamma = 0{,}98$ из уравнения

$$\Phi(t) = \frac{\gamma}{2} = \frac{0{,}98}{2} = 0{,}49,$$

где $\Phi(t) -$ функция Лапласа,

по таблице значений функции Лапласа находим:

$$t = 2{,}33.$$

Тогда

$$\Delta = 2{,}33 \cdot 0{,}226 \approx 0{,}527.$$

Границы, в которых с вероятностью 0,98 заключен средний размер площади на одного человека в целом по району, находим из следующего двойного неравенства:

$$\bar{x} - \Delta \leq x \leq \bar{x} + \Delta,$$

подставляем значения:

$$19{,}005 - 0{,}527 \leq x \leq 19{,}005 + 2{,}034,$$

$$18{,}478 \leq x \leq 19{,}532.$$

С вероятностью 0,98 можно ожидать, что средний размер площади на одного человека составит от 18,478 до 19,532 м^2.

23.2 По данным задачи 1, используя $\chi^2 -$ критерий Пирсона, на уровне значимости $\alpha = 0{,}05$ проверить гипотезу о том, что случайная величина $X -$ размер площади на одного человека в целом по району – распределена по нормальному закону. Построить на одном чертеже гистограмму эмпирического распределения и соответствующую нормальную кривую.

Решение.

Проверку гипотезы о виде закона распределения проведем, используя критерий согласия Пирсона. Суть проверки гипотезы о том, что случайная величина распределена по нормальному закону, состоит в том, что сравниваются наблюдаемое значение статистики и критическое: $\chi^2_{набл}$ и $\chi^2_{крит}$.

Наблюдаемое значение статистики определяется по эмпирическим и теоретическим частотам по формуле:

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

где n_i– эмпирические, а np_i– теоретические частоты.

Критическое значение статистики определяется в зависимости от уровня значимости α и числа степеней свободы $r = k - 1 - l$, где $k -$ число интервалов; $l -$ число параметров закона распределения (в нормальном распределении $l = 2$)

Для определения теоретических частот нам нужны параметры закона распределения, а именно - математическое ожидание a и среднее квадратическое отклонение σ

$$\bar{x} = a = 19{,}005; \;\; \sigma = \sqrt{51{,}06}$$

Для расчета вероятностей p_i используем функцию Лапласа

$$P(\alpha \le x < \beta) = \Phi\left(\frac{\beta - \bar{x}}{\sigma}\right) - \Phi\left(\frac{\alpha - \bar{x}}{\sigma}\right)$$

$$P(0 \le x < 5) = \Phi\left(\frac{5 - 19{,}005}{\sqrt{51{,}06}}\right) - \Phi\left(\frac{0 - 19{,}005}{\sqrt{51{,}06}}\right) =$$

$$= \Phi(-1{,}96) - \Phi(-2{,}66) = -0{,}4750 + 0{,}4961 = 0{,}0211$$

$$P(5 \le x < 10) = \Phi\left(\frac{10 - 19{,}005}{\sqrt{51{,}06}}\right) - \Phi\left(\frac{5 - 19{,}005}{\sqrt{51{,}06}}\right) =$$

$$= \Phi(-1{,}26) - \Phi(-1{,}96) = -0{,}3962 + 0{,}4750 = 0{,}0788$$

$$P(10 \le x < 15) = \Phi\left(\frac{15 - 19{,}005}{\sqrt{51{,}06}}\right) - \Phi\left(\frac{10 - 19{,}005}{\sqrt{51{,}06}}\right) =$$

$$= \Phi(-0,56) - \Phi(-1,26) = -0,2123 + 0,3962 = 0,1839$$

$$P(15 \le x < 20) = \Phi\left(\frac{20 - 19,005}{\sqrt{51,06}}\right) - \Phi\left(\frac{15 - 19,005}{\sqrt{51,06}}\right) =$$

$$= \Phi(0,14) - \Phi(-0,56) = 0,0557 + 0,2123 = 0,2680$$

$$P(20 \le x < 25) = \Phi\left(\frac{25 - 19,005}{\sqrt{51,06}}\right) - \Phi\left(\frac{20 - 19,005}{\sqrt{51,06}}\right) =$$

$$= \Phi(0,84) - \Phi(0,14) = 0,2995 - 0,0557 = 0,2438$$

$$P(25 \le x < 30) = \Phi\left(\frac{30 - 19,005}{\sqrt{51,06}}\right) - \Phi\left(\frac{25 - 19,005}{\sqrt{51,06}}\right) =$$

$$= \Phi(1,54) - \Phi(0,84) = 0,4382 - 0,2995 = 0,1387$$

$$P(30 \le x < 35) = \Phi\left(\frac{35 - 19,005}{\sqrt{51,06}}\right) - \Phi\left(\frac{30 - 19,005}{\sqrt{51,06}}\right) =$$

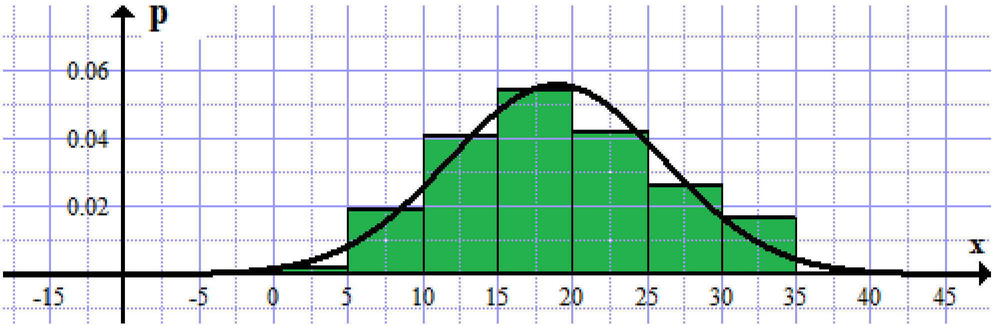
$$= \Phi(2,24) - \Phi(1,54) = 0,4875 - 0,4382 = 0,0493$$

i	интервал	n_i	p_i	np_i	$n_i - np_i$	$(n_i - np_i)^2/np_i$
1	0-5	8	0,0211	21,1	-13,1	8,1332
2	5-10	95	0,0788	78,8	16,2	3,3305
3	10-15	204	0,1839	183,9	20,1	2,1969
4	15-20	270	0,2680	268	2	0,0149
5	20-25	210	0,2438	243,8	-33,8	4,6860
6	25-30	130	0,1387	138,7	-8,7	0,5457
7	30-35	83	0,0493	49,3	33,7	23,0363
		1000				41,9434

Итого значение статистики $\chi^2_{набл} = 41,9434$;

Определим количество степеней свободы по формуле: $r = k - 1 - l.$ $r = 7 - 1 - 2 = 4.$ Соответствующее критическое значение статистики $\chi^2_{кр} = 9,5$

Так как $\chi^2_{набл} > \chi^2_{кр} \Rightarrow 41{,}9434 > 9{,}5$, то гипотеза о нормальном распределении отвергается.



23.3 Совместное распределение двух случайных факторов X и Y представлено в таблице

$X \vert Y$	42	44	46	48
1.3	5	2	-	-
1.4	-	7	3	1
1.5	-	3	8	2
1.6	-	1	3	5

Необходимо:

1. Вычислить групповые средние $\overline{y_i}$, построить эмпирическую линию регрессии Y на X.

2. Предполагая, что между переменными X и Y существует линейная корреляционная зависимость:

1) найти уравнение прямой регрессии Y на X, построить её график на одном чертеже с эмпирической линией регрессии;

2) вычислить выборочный коэффициент корреляции и сделать вывод о тесноте и направлении связи между переменными X и Y;

3) используя полученное уравнение регрессии вычислить прогнозное значение Y при росте фактора X на 10% от среднего значения.

Решение.

$X\|Y$	42	44	46	48	итого
1,3	5	2	-	-	7
1,4	-	7	3	1	11
1,5	-	3	8	2	13
1,6	-	1	3	5	9
итого	5	13	14	8	40

1. Вычисляем групповые средние по формуле:

$$\bar{y_j} = \frac{\sum y_i n_{ij}}{n_j};$$

$$\bar{y}_1 = \frac{1}{7}(42 * 5 + 44 * 2) = \frac{298}{7} \approx 42{,}57$$

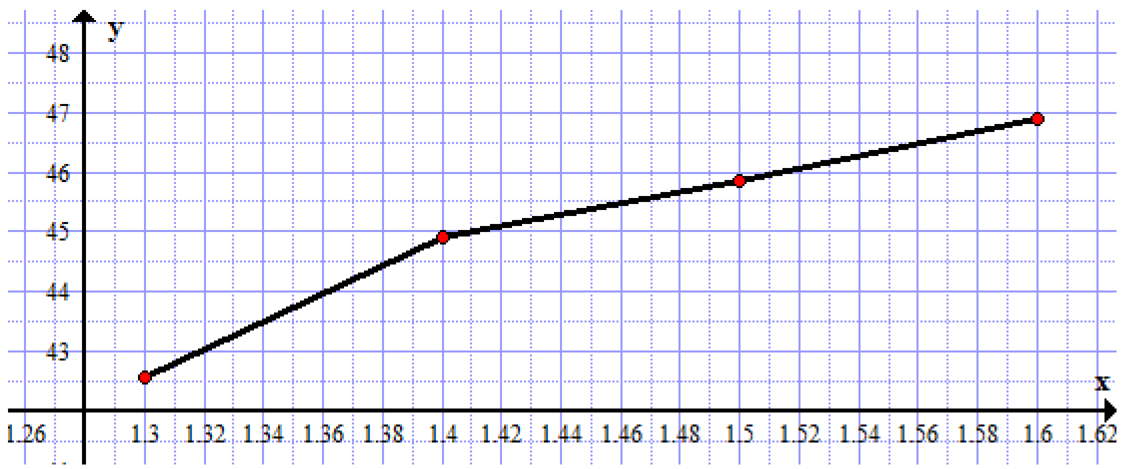
$$\bar{y}_2 = \frac{1}{11}(44 * 7 + 46 * 3 + 48 * 1) = \frac{494}{11} \approx 44{,}91$$

$$\bar{y}_3 = \frac{1}{13}(44 * 3 + 46 * 8 + 48 * 2) = \frac{596}{13} \approx 45{,}85$$

$$\bar{y}_4 = \frac{1}{9}(44 * 1 + 46 * 3 + 48 * 5) = \frac{422}{9} \approx 46{,}89$$

$X\|Y$	42	44	46	48	$\bar{y}_i$
1,3	5	2	-	-	42,57
1,4	-	7	3	1	44,91
1,5	-	3	8	2	45,85
1,6	-	1	3	5	46,89

Построим эмпирические линии регрессии Y на X в виде ломаной, соединяющей точки $(x_i; \bar{y_\iota})$.



2.1) Уравнение регрессии имеет вид:

$$y_x = \frac{\overline{xy} - \bar{x}\bar{y}}{S_x^2}(x - \bar{x}) + \bar{y};$$

x_i	n_i	y_j	n_j	$x_i n_i$	$y_j n_j$	${x_i}^2 n_i$	${y_j}^2 n_j$
1,3	7	42	5	9,1	210	11,83	8820
1,4	11	44	13	15,4	572	21,56	25168
1,5	13	46	14	19,5	644	29,25	29624
1,6	9	48	8	14,4	384	23,04	18432
5,8	**40**	**180**	**40**	**58,4**	**1810**	**85,68**	**82044**

$$\bar{x} = \frac{1}{n}\sum x_i n_i = \frac{58,4}{40} = 1,46$$

$$\bar{y} = \frac{1}{n}\sum y_j n_j = \frac{1810}{40} = 45,25$$

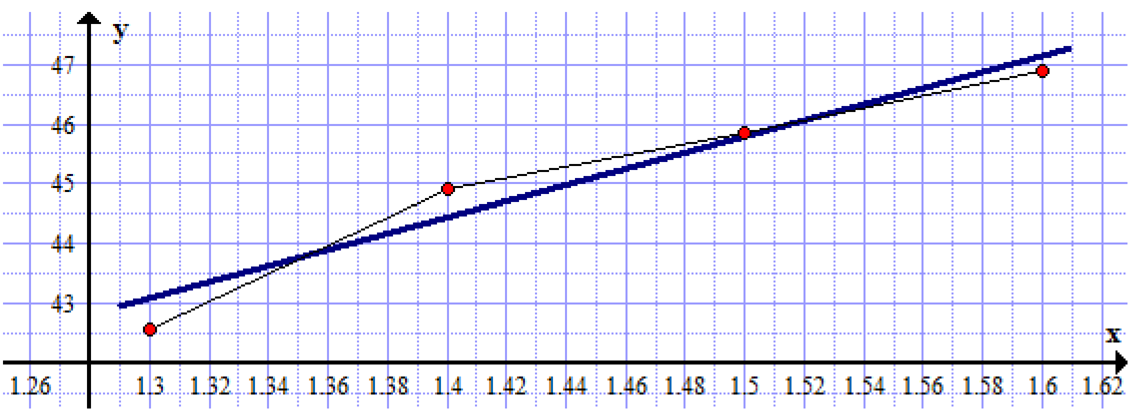
$$\overline{xy} = \frac{1}{n}\sum\sum x_i y_j n_{ij} = \frac{2648,2}{40} = 66,205$$

$$\mu = \overline{xy} - \bar{x}\bar{y} = 66,0925 - 1,46 * 45,25 = 0,14$$

$$S_x^2 = \frac{1}{n}\sum x_i^2 n_i - (\bar{x})^2 = \frac{85,68}{40} - (1,46)^2 = 0,0104$$

$$y_x = \frac{0,14}{0,0104}(x - 1,46) + 45,25$$

$y_x = 13,46x + 25,596$ уравнение регрессии Y на X.



2) Коэффициент корреляции

$$r = \frac{\overline{xy} - \bar{x}\bar{y}}{\sqrt{S_x^2 \cdot S_y^2}}$$

$$S_y^2 = \frac{1}{n}\sum y_i^2 n_i - (\bar{y})^2 = \frac{82044}{40} - (45{,}25)^2 = 3{,}5375$$

$$r = \frac{0{,}14}{\sqrt{0{,}0104 * 3{,}5375}} = 0{,}7299$$

Так как $0{,}7 < |r| < 0{,}9$ и $r > 0$, то связь между X и Y высока прямая.

3) При росте фактора X на 10% от среднего значения, он будет составлять $1{,}46 + 1{,}46 * 10\% = 1{,}606$.

Используя полученное уравнение регрессии:

$$y_x = 13{,}46x + 25{,}596$$

вычислим прогнозное значение Y:

$$y_{x=1{,}606} = 13{,}46 * 1{,}606 + 25{,}596 = 47{,}2128.$$