Разбор типового расчета по статистике

Задание №12. Ниже приведены 50 значений случайной величины X, а также вычисленные по выборке среднее a и среднее квадратическое s. Используя критерий хи-квадрат определить, будет ли X нормальной случайной величиной. В качестве уровня значимости взять вероятность 0,2.

41,11	41,55	50,72	49,82	40,53	41,29	36,03	19,88	32,36	31,28
32,36	31,10	56,22	40,39	62,24	34,66	25,87	8,37	33,69	55,75
24,31	37,76	33,74	35,02	35,44	45,36	21,51	24,80	48,62	62,30
32,15	51,24	68,09	25,61	19,22	55,61	17,81	43,84	24,12	15,95
37,90	34,76	20,77	37,60	33,58	39,28	59,88	36,26	10,61	8,31

$$a = 36{,}13; \qquad s = 14{,}192$$

Решение

Построим вариационный ряд – выборку в порядке возрастания:

8,31	8,37	10,61	15,95	17,81	19,22	19,88	20,77	21,51	24,12
24,31	24,80	25,61	25,87	31,10	31,28	32,15	32,36	32,36	33,58
33,69	33,74	34,66	34,76	35,02	35,44	36,03	36,26	37,60	37,76
37,90	39,28	40,39	40,53	41,11	41,29	41,55	43,84	45,36	48,62
49,82	50,72	51,24	55,61	55,75	56,22	59,88	62,24	62,30	68,09

Найдем размах выборки R.

$$x_{min} = 8{,}31; \qquad x_{max} = 68{,}09$$

$$R = x_{max} - x_{min} = 68{,}09 - 8{,}31 = 59{,}78$$

Число интервалов N, на которые следует разбить интервал значений признака, найдём по формуле Стерджесса:

$$N = 1 + 3{,}322 lg(n)$$

где n – объём выборки, то есть число единиц наблюдения.

В данном случае $n = 50$. Получим:

$$N = 1 + 3{,}322 \cdot lg(50) = 6{,}64 \approx 7$$

Рассчитаем шаг (длину частичного интервала) h по формуле:

$$h = \frac{R}{N} = \frac{59{,}78}{7} = 8{,}54$$

Подсчитаем частоту n_i каждого интервала, то есть число вариант, попавших в этот интервал. Построим интервальный вариационный ряд частот с равными интервалами.

Интервал		Частота n_i
8,37	16,91	3
16,91	25,45	7
25,45	33,99	9
33,99	42,53	14
42,53	51,07	4
51,07	59,61	4
59,61	68,15	4

Вероятность попадания случайной величины в каждый интервал равна приращению функции распределения:

$$p_i = P(x_1 < X < x_2) = \Phi\left(\frac{x_2 - a}{s}\right) - \Phi\left(\frac{x_1 - a}{s}\right)$$

$$P_1 = P(-\infty < X < 16{,}91) = \Phi\left(\frac{16{,}91 - 36{,}13}{14{,}192}\right) - \Phi(-\infty) =$$

$$= -0{,}4115 + 0{,}5 = 0{,}0885$$

$$P_2 = P(16{,}91 < X < 25{,}45) = \Phi\left(\frac{25{,}45 - 36{,}13}{14{,}192}\right) - \Phi\left(\frac{16{,}91 - 36{,}13}{14{,}192}\right) =$$

$$= -0{,}2734 + 0{,}4115 = 0{,}1381$$

$$P_3 = P(25{,}45 < X < 33{,}99) = \Phi\left(\frac{33{,}99 - 36{,}13}{14{,}192}\right) - \Phi\left(\frac{25{,}45 - 36{,}13}{14{,}192}\right) =$$

$$= -0{,}0596 + 0{,}2734 = 0{,}2138$$

$$P_4 = P(33{,}99 < X < 42{,}53) = \Phi\left(\frac{42{,}53 - 36{,}13}{14{,}192}\right) - \Phi\left(\frac{33{,}99 - 36{,}13}{14{,}192}\right) =$$

$$= 0{,}1736 + 0{,}0596 = 0{,}2332$$

$$P_5 = P(42{,}53 < X < 51{,}07) = \Phi\left(\frac{51{,}07 - 36{,}13}{14{,}192}\right) - \Phi\left(\frac{42{,}53 - 36{,}13}{14{,}192}\right) =$$

$$= 0{,}3531 - 0{,}1736 = 0{,}1795$$

$$P_6 = P(51{,}07 < X < 59{,}61) = \Phi\left(\frac{59{,}61 - 36{,}13}{14{,}192}\right) - \Phi\left(\frac{51{,}07 - 36{,}13}{14{,}192}\right) =$$

$$= 0{,}4505 - 0{,}3531 = 0{,}0974$$

$$P_7 = P(59{,}61 < X < +\infty) = \Phi(+\infty) - \Phi\left(\frac{59{,}61 - 36{,}13}{14{,}192}\right) =$$

$$= 0{,}5 - 0{,}4505 = 0{,}0495$$

Теоретические частоты определим по формуле: $n'_i = n \cdot p_i$ и вычислим значения:

$$\frac{(n_i - n'_i)^2}{n'_i}$$

Результаты запишем в таблицу

Интервалы		n_i	p_i	n'_i	$\frac{(n_i - n'_i)^2}{n'_i}$
8,37	16,91	3	0,0885	3,9825	0,2424
16,91	25,45	7	0,1381	6,2145	0,0993
25,45	33,99	9	0,2138	9,621	0,0401
33,99	42,53	14	0,2332	10,494	1,1713
42,53	51,07	4	0,1795	8,0775	2,0583
51,07	59,61	4	0,0974	4,383	0,0335
59,61	68,15	4	0,0495	2,2275	1,4104
$\sum$			1		5,0553

Используя критерий Пирсона при уровне значимости $\alpha = 0{,}2$, проверим гипотезу о нормальном распределении случайной величины X. Получили $\chi^2_{\text{наб}} = 5{,}0553$. Число степеней свободы нормального распределения $v = 7 - 2 - 1 = 4$. По таблице при уровне значимости $\alpha = 0{,}2$ находим $\chi^2_{\text{кр}}(0{,}2; 4) = 5{,}99$. Так как $\chi^2_{\text{наб}} = 5{,}0553 < \chi^2_{\text{кр}}(0{,}2; 4) = 5{,}99$, то нет оснований отвергать гипотезу о нормальном распределении при заданном уровне значимости.