α = 0,05

ТР - 10: "Выравнивание статистического ряда". Варианты **1 - 16.**

В индивидуальном задании приведен статистический ряд: отрезок **[a, b]** разб **12** равных интервалов длины **h** для каждого такого интервала указана частота **n**

№	**- номер варианта.**	**a =**			**b =**			**h**			
n	**n**	...	...	...	...	...	...	...	...	...	**n**
1		a =	75	b =	675	h =	50				
2	4	8	16	18	14	12	10	6	4	4	2
2		a =	60	b =	180	h =	10				
4	9	23	32	52	61	50	28	19	16	4	2
3		a =	52	b =	100	h =	4				
1	10	27	35	56	61	46	32	17	10	4	1
4		a =	6,5	b =	24,5	h =	1,5				
5	11	27	36	67	80	56	34	19	10	3	2
5		a =	2,5	b =	6,1	h =	0,3				
4	9	13	22	33	40	31	20	13	8	5	2
6		a =	3,9	b =	5,1	h =	0,1				
6	12	23	38	58	70	55	36	24	13	10	5
7		a =	0,85	b =	2,65	h =	0,15				
12	19	34	48	71	83	65	50	34	23	20	16
8		a =	1,2	b =	19,2	h =	1,5				
8	14	28	38	56	70	55	41	30	18	11	6
9		a =	4,75	b =	7,15	h =	0,2				
7	17	34	54	70	94	89	73	55	34	16	7
10		a =	114,5	b =	133	h =	1,5				
5	11	26	40	52	62	80	60	48	22	13	6
11		a =	98	b =	140	h =	3,5				
7	16	31	52	70	84	92	78	57	36	18	9
12		a =	11,2	b =	32,8	h =	1,8				
8	17	32	54	74	96	90	77	56	40	22	9
13		a =	120	b =	138	h =	1,5				
3	8	14	29	42	49	40	30	18	9	5	3
14		a =	36	b =	66	h =	2,5				
2	11	24	36	52	66	50	37	22	14	8	3
15		a =	36,6	b =	67,8	h =	2,6				
7	16	25	40	65	78	53	35	16	9	4	2
16		a =	42,3	b =	74,7	h =	2,7				
15	25	36	47	65	86	84	64	40	28	19	16

$\alpha = 0,05$

ТР - 10: "Выравнивание статистического ряда". Варианты 17 - 30.

В индивидуальном задании приведен статистический ряд: отрезок **[a, b]** разбит на **12** равных интервалов длины **h** для каждого такого интервала указана частота **n**

№	**- номер варианта.**	**a =**				**b =**		**h =**			
n	n	...	...	...	...	...	...	...	...	...	n

17		a =	36	b =	132	h =	8				
14	19	35	48	69	88	92	80	58	33	24	15
18		a =	78	b =	132	h =	4,5				
2	5	11	20	35	50	62	51	38	28	17	6
19		a =	98	b =	113	h =	1,25				
15	19	28	40	64	85	90	80	59	36	20	14
20		a =	75	b =	675	h =	50				
5	11	29	51	66	78	90	72	51	30	12	5
21		a =	60	b =	180	h =	10				
2	4	15	19	25	47	64	52	34	23	11	4
22		a =	52	b =	100	h =	4				
2	4	10	17	30	45	60	50	35	27	15	5
23		a =	2,5	b =	6,1	h =	0,3				
2	4	7	13	22	33	37	31	22	14	10	5
24		a =	3,9	b =	5,1	h =	0,1				
7	11	15	24	40	58	72	61	47	36	20	9
25		a =	0,85	b =	2,65	h =	0,15				
16	20	23	34	50	65	80	66	50	34	22	15
26		a =	1,2	b =	19,2	h =	1,5				
6	11	18	29	41	55	73	56	38	26	14	8
27		a =	4,75	b =	7,15	h =	0,2				
9	17	35	56	74	93	101	71	55	35	19	10
28		a =	115	b =	127	h =	1				
6	11	25	43	65	79	62	52	40	26	11	5
29		a =	100	b =	292	h =	16				
9	18	38	60	80	98	93	78	62	34	18	12
30		a =	11,2	b =	32,8	h =	1,8				
8	15	33	55	70	88	93	76	56	32	16	8

Для проверки гипотезы о распределении случайной величины X проведем выборку, которую оформим в виде статистического ряда:

x_i	x_1	x_2	$\ldots$	x_m
n_i	n_1	n_2	$\ldots$	n_m

(8.11)

где $\sum\limits_{i=1}^{m} n_i = n$ — объем выборки.

Требуется сделать заключение: согласуются ли результаты наблюдений с высказанным предположением. Для этого используем специально подобранную величину — критерий согласия.

Критерием согласия называют статистический критерий проверки гипотезы о предполагаемом законе неизвестного распределения. (Он используется для проверки согласия предполагаемого вида распределения с опытными данными на основании выборки.)

Существуют различные критерии согласия: Пирсона, Колмогорова, Фишера, Смирнова и др.

Критерий согласия Пирсона — наиболее часто употребляемый критерий для проверки простой гипотезы о законе распределения.

Критерий χ^2 Пирсона

Для проверки гипотезы H_0 поступают следующим образом.

Разбивают всю область значений с. в. X на m интервалов $\Delta_1, \Delta_2, \ldots, \Delta_m$ и подсчитывают вероятности p_i $(i = 1, 2, \ldots, m)$ попадания с. в. X (т. е. наблюдения) в интервал Δ_i, используя формулу $P\{\alpha \leqslant X \leqslant \beta\} = F_0(\beta) - F_0(\alpha)$. Тогда теоретическое число значений с. в. X, попавших в интервал Δ_i, можно рассчитать по формуле $n \cdot p_i$. Таким образом, имеем статистический ряд распределения с. в. X (8.11) и теоретический ряд распределения:

Δ_1	Δ_2	$\ldots$	Δ_m
$n'_1 = np_1$	$n'_2 = np_2$	$\ldots$	$n'_m = np_m$

(8.12)

Если эмпирические частоты (n_i) сильно отличаются от теоретических $(np_i = n'_i)$, то проверяемую гипотезу H_0 следует отвергнуть; в противном случае — принять.

Каким критерием, характеризующим степень расхождения между эмпирическими и теоретическими частотами, следует воспользоваться? В качестве меры расхождения между n_i и np_i для $i = 1, 2, \ldots, m$

К. Пирсон (1857–1936; англ. математик, статик, биолог, философ) предложил величину («критерий Пирсона»):

$$\chi^2 = \sum_{i=1}^{m} \frac{(n_i - np_i)^2}{np_i} = \sum_{i=1}^{m} \frac{n_i^2}{np_i} - n. \tag{8.13}$$

Согласно теореме Пирсона, при $n \to \infty$ статистика (8.13) имеет χ^2-распределение с $k = m - r - 1$ степенями свободы, где m — число групп (интервалов) выборки, r — число параметров предполагаемого распределения. В частности, если предполагаемое распределение нормально, то оценивают два параметра (a и σ), поэтому число степеней свободы $k = m - 3$.

Правило применения критерия χ^2 сводится к следующему:

1. По формуле (8.13) вычисляют $\chi^2_{\text{набл}}$ — выборочное значение статистики критерия.
2. Выбрав уровень значимости α критерия, по таблице χ^2-распределения находим критическую точку (квантиль) $\chi^2_{\alpha,k}$.
3. Если $\chi^2_{\text{набл}} \leqslant \chi^2_{\alpha,k}$, то гипотеза H_0 не противоречит опытным данным; если $\chi^2_{\text{набл}} > \chi^2_{\alpha,k}$, то гипотеза H_0 отвергается.

Необходимым условием применения критерия Пирсона является наличие в каждом из интервалов не менее 5 наблюдений (т. е. $n_i \geqslant 5$). Если в отдельных интервалах их меньше, то число интервалов надо уменьшить путем объединения (укрупнения) соседних интервалов.

Пример 8.8. Измерены 100 обработанных деталей; отклонения от заданного размера приведены в таблице:

$[x_i, x_{i+1})$	$[-3,-2)$	$[-2,-1)$	$[-1,0)$	$[0,1)$	$[1,2)$	$[2,3)$	$[3,4)$	$[4,5)$
n_i	3	10	15	24	25	13	7	3

Проверить при уровне значимости $\alpha = 0{,}01$ гипотезу H_0 о том, что отклонения от проектного размера подчиняются нормальному закону распределения.

Число наблюдений в крайних интервалах меньше 5, поэтому объединим их с соседними. Получим следующий ряд распределения ($n = 100$):

$[x_i, x_{i+1})$	$[-3,-1)$	$[-1,0)$	$[0,1)$	$[1,2)$	$[2,3)$	$[3,5)$
n_i	13	15	24	25	13	10

Случайную величину — отклонение — обозначим через X. Для вычисления вероятностей p_i необходимо вычислить параметры, определяющие нормальный закон распределения (a и σ). Их оценки вычислим по выборке: $\overline{x} = \frac{1}{100} \cdot (-2 \cdot 13 + (-0{,}5) \cdot 15 + \ldots + 4 \cdot 10) = 0{,}885 \approx 0{,}9$, $D_в = \frac{1}{100}(4 \cdot 13 + 0{,}25 \cdot 15 + \ldots + 16 \cdot 10) - (0{,}885)^2 \approx 2{,}809$, $\sigma \approx 1{,}676 \approx 1{,}7$.

Находим p_i ($i = \overline{1,6}$). Так как с.в. $X \sim N(a, \sigma)$ определена на $(-\infty, \infty)$, то крайние интервалы в ряде распределения заменяем, соответственно, на $(-\infty, -1)$ и $(3, +\infty)$. Тогда $p_1 = P\{-\infty < X < -1\} = \Phi_0\left(\frac{-1-0{,}9}{1{,}7}\right) - \Phi_0(-\infty) = \frac{1}{2} - \Phi_0(1{,}12) = 0{,}1314$. Аналогично получаем: $p_2 = 0{,}1667$, $p_3 = 0{,}2258$, $p_4 = 0{,}2183$, $p_5 = 0{,}1503$, $p_6 = p\{3 \leqslant X < \infty\} = \Phi_0(\infty) - \Phi_0\left(\frac{3-0{,}9}{1{,}7}\right) = 0{,}5 - \Phi_0(1{,}24) = 0{,}1075$.

Полученные результаты приведем в следующей таблице:

$[x_i, x_{i+1})$	$(-\infty, -1)$	$[-1, 0)$	$[0, 1)$	$[1, 2)$	$[2, 3)$	$[3, \infty)$
n_i	13	15	24	25	13	10
$n' = np_i$	13,14	16,67	22,58	21,83	15,03	10,75

Вычисляем $\chi^2_{набл}$:

$$\chi^2_{набл} = \sum_{i=1}^{6} \frac{n_i^2}{np_i} - n =$$

$$= \left(\frac{13^2}{13{,}14} + \frac{15^2}{16{,}67} + \ldots + \frac{10^2}{10{,}75}\right) - 100 = 101{,}045 - 100,$$

т. е. $\chi^2_{набл} \approx 1{,}045$.

Находим число степеней свободы; по выборке рассчитаны два параметра, значит, $r = 2$. Количество интервалов 6, т. е. $m = 6$. Следовательно, $k = 6 - 2 - 1 = 3$. Зная, что $\alpha = 0{,}01$ и $k = 3$, по таблице χ^2-распределения находим $\chi^2_{\alpha,k} = 11{,}3$. Итак, $\chi^2_{набл} < \chi^2_{\alpha,k}$, следовательно, нет оснований отвергнуть проверяемую гипотезу.