ПЕРЕВІРКА НЕПАРАМЕТРИЧНИХ СТАТИСТИЧНИХ ГІПОТЕЗ

№1 Изучается процентное отношение номинальной и рыночной цен на акции на фондовом рынке (*X*) за некоторый период. Выборка получена случайным способом по акциям 50 разных предприятий и задается данными, приведенными в таблице.

N	X	N	X	N	X	N	X	N	X
1	98,06	**11**	100,4	**21**	97,9	**31**	96,32	**41**	99,97
2	101,25	**12**	99,4	**22**	103,9	**32**	102,72	**42**	103,97
3	101,25	**13**	99,6	**23**	105,9	**33**	100,72	**43**	98,97
4	96,25	**14**	98,4	**24**	100,9	**34**	100,72	**44**	99,97
5	100,85	**15**	98,4	**25**	99,9	**35**	101,72	**45**	96,27
6	101,25	**16**	97,4	**26**	101,9	**36**	96,42	**46**	96,97
7	98,85	**17**	94,4	**27**	101,9	**37**	96,42	**47**	98,97
8	99,25	**18**	100,8	**28**	99,9	**38**	98,72	**48**	100,97
9	100,55	**19**	98,4	**29**	103,3	**39**	100,22	**49**	98,97
10	100,45	**20**	96,4	**30**	101,9	**40**	101,72	**50**	100,97

Необходимо при помощи критерия Пирсона проверить гипотезу о нормальном законе распределения процентного отношения номинальной и рыночной цен на акции на фондовом рынке при уровне значимости $\alpha = 0{,}05$.

Решение. Разобьем интервал для X на 5 равных подинтервалов. Сюда же отнесем статистическое распределение для усредненных вариант x_i^*.

Размах вариаций: $R_x = 105{,}9 - 94{,}4 = 11{,}5$. Длины подинтервалов: $h_x = \frac{R_x}{5} = 2{,}3$.

Соответствующие усредненные варианты: $x_i^* = \frac{x_i + x_{i+1}}{2}$.

x_i	x_{i+1}	x_i^*	n_i
94,400	96,700	95,550	7
96,700	99,000	97,850	12
99,000	101,300	100,150	21
101,300	103,600	102,450	7
103,600	105,900	104,750	3

Для вычисления теоретических частот n^t используем формулу: $n^t = nP_i$, где $P_i = \Phi\left(\frac{x_i - \bar{x}}{\sigma_x}\right) - \Phi\left(\frac{x_{i-1} - \bar{x}}{\sigma_x}\right)$ (вероятность попадания случайной величины в данный интервал), значения функции $\Phi(z)$ находим по таблице. Вычисления наблюдаемого значения $\chi^2_{спост}$ статистики χ^2- критерия оформим в виде таблицы:

Интервал	$\Phi(z_{i-1})$	$\Phi(z_i)$	n	P_i	n^t	$\frac{(n-n^t)^2}{n}$
$(-\infty;-1{,}36)$	-0,5000	-0,4099	7	0,090	4,51	0,889
$(-1{,}36;-0{,}35)$	-0,4099	-0,1368	12	0,273	13,66	0,228
$(-0{,}35;0{,}66)$	-0,1368	0,2454	21	0,382	19,11	0,170
$(0{,}66;1{,}66)$	0,2454	0,4515	7	0,206	10,31	1,560
$(1{,}66;+\infty)$	0,4515	0,5000	3	0,049	2,43	0,110

$\sum$			**50**		**50,00**	**2,96**

Итак, $\chi^2_{спост} = 2{,}96$. Для уровня значимости $\alpha = 0{,}05$ критическая точка χ^2- критерия для $k = s - 3 = 5 - 3 = 2$ степей свободы $\chi^2_{кр}(0{,}05;2) = 6{,}0$. Т.к. наблюдаемое значение критерия меньше критического, то нулевая гипотеза H_0 о нормальном распределении принимается.

№2 Пусть задана выборка *A*. 2, 4, 2, 4, 3, 3, 3, 2, 0, 6, 1, 2, 3, 2, 2, 4, 3, 3, 5, 1, 0, 2, 4, 3, 2, 2, 3, 3, 1, 3, 3, 3, 1, 1, 2, 3, 1, 4, 3, 1, 7, 4, 3, 4, 2, 3, 2, 3, 3, 1, 4, 3, 4, 4, 5, 3, 4, 2, 4, 5, 3, 6, 4, 1, 3, 2, 4, 1, 3, 1, 0, 0, 4, 6, 4, 7, 4, 1, 3.

$n = 79$. Начало первого интервала: 0. Длина интервала 7. Требуется с уровнем значимости $\alpha = 0{,}01$, проверить гипотезу о распределении Пуассона генеральной совокупности по данным выборки *A*.

Решение. За меру отклонения распределения выборки от предлагаемого распределения Пуассона принимается величина χ^2. Параметром распределения Пуассона является $\lambda = M(X) = D(X)$, его значение неизвестно, поэтому по данным выборки вычислим оценку $M(X)$ по формуле:

$$\overline{x_B} = \frac{1}{79}(0\cdot 4 + 1\cdot 13 + 2\cdot 14 + 3\cdot 24 + 4\cdot 16 + 5\cdot 3 + 6\cdot 3 + 7\cdot 2) = 2{,}84$$

и оценку $D(X)$ по формуле

$$D_B = \frac{1}{79-1}\Big(4\cdot(0-2{,}84)^2 + 13\cdot(1-2{,}84)^2 + 14\cdot(2-2{,}84)^2 + 24\cdot(3-2{,}84)^2 +$$
$$+16\cdot(4-2{,}84)^2 + 3\cdot(5-2{,}84)^2 + 3\cdot(6-2{,}84)^2 + 2\cdot(7-2{,}84)^2\Big) = 2{,}3668.$$

Вычислим несмещенную оценку дисперсии

$$S^2 = \frac{79}{79-1}\cdot 2{,}3668 = 2{,}3971.$$

Оценки среднего значения случайной величины и дисперсии близки по значению, но не равны.

В таблице распределения Пуассона ближайшими к ним значениями λ являются 2 и 3. проверим гипотезу при $\lambda = 3$.

Частоты последних значений вариационного ряда малы, поэтому объединим их в один интервал (≥ 5). Количество интервалов $k = 6$. Следовательно статистика χ^2 имеет распределение с числом степеней свободы $k_1 = 6 - 1 - 1 = 4$.

Критическая область определяется неравенством

$$P\left(\chi^2 > \chi^2_{0{,}01;\,4}\right) = 0{,}01.$$

Выполним необходимые вычисления при помощи таблицы. Вероятность P_i получим по таблице распределения Пуассона.

x_i	n_i	P_i	nP_i	$n_i - nP_i$	$(n_i - nP_i)^2$	$\frac{(n_i - nP_i)^2}{nP_i}$
0	4	0,0498	3,9	0,1	0,01	0,0026
1	13	0,1494	11,8	1,2	1,44	0,1220
2	14	0,2240	17,7	–3,7	13,69	0,77
3	24	0,2240	17,7	6,3	36,69	0,2424
4	16	0,1680	13,3	2,7	7,29	0,5481
≥ 5	8	0,1847	14,6	–6,6	43,56	2,9836

Σ	**79**	**0,9999**	**79**	**–**	**–**	**6,6721**

№3 Монету бросали 4040 раз (Бюффон). Получили $n_1 = 2048$ выпадений герба и $n_2 = 1992$ выпадений решки. Проверить, используя критерий Колмогорова, согласуются ли эти данные с гипотезой H_0 о симметричности монеты ($\alpha = 0,05$).

Решение.

Случайная величина X принимает два значения: $x_1 = -1$ (решка) и $x_2 = 1$ (герб). Гипотеза H_0: $P\{x = -1\} = P\{x = 1\} = 0,5$.

По таблице распределения Колмогорова находим корень уравнения $K(x) = 1 - \alpha$ при $\alpha = 0,05$. Значит, $x_0 = 1,358$. Тогда, $D_0 = \frac{x_0}{\sqrt{n}} = \frac{1,358}{\sqrt{4040}} \approx 0,021$.

Для нахождения по выборке D_n строим функции $F_0(x)$, $F_n^*(x)$ и вычисляем величину $D_n = \max\limits_{-\infty<x<\infty} \left| F_n^*(x) - F_0(x) \right|$.

x_i	решка $x_1 = -1$	герб $x_2 = 1$
p_i	0,5	0,5

$$\Rightarrow F_0(x) = \begin{cases} 0, & \text{при } x \le -1, \\ 0,5, & \text{при } -1 < x \le 1, \\ 1, & \text{при } 1 < x; \end{cases}$$

x_i	решка $x_1 = -1$	герб $x_2 = 1$
n_i	1992	2048
p_i^*	0,493	0,507

$$\Rightarrow F_n^*(x) = \begin{cases} 0, & \text{при } x \le -1, \\ 0,493, & \text{при } -1 < x \le 1, \\ 1, & \text{при } 1 < x; \end{cases}$$

Максимальное отклонение $F_0(x)$ от $F_n^*(x)$ равно 0,007, т.е. $D_n = 0,007$. Поскольку $D_n < D_0$, то нет оснований отвергать гипотезу H_0 о симметричности монеты.

№4 Використовуючи критерій Пірсона (*хі*-квадрат) з рівнем значущості $\alpha = 0.05$, перевірити гіпотезу про нормальний розподіл генеральної сукупності X, якщо відомі емпіричні n_k та теоретичні n_k' частоти

n_i	5	10	20	8	7
n_i'	6	14	18	7	5