Проверка непараметрических гипотез Критерии согласия.

Предположение о свойствах распределения вероятностей наблюдаемых явлений называется **статистической гипотезой**.

Среди статистических гипотез можно выделить гипотезы о виде закона неизвестного распределения случайной величины или гипотезы о числовых значениях параметров известных распределений случайной величины.
Выдвинутую гипотезу называют **нулевой** (**основной**) гипотезой и обозначают, как правило, H_0. Гипотеза H_1, противоречащая основной называется **альтернативной** (**конкурирующей**).

Проверка нулевой гипотезы основано на исследовании статистических данных.

При проверке статистических гипотез могут возникать ошибки первого или второго рода. Вероятность отвергнуть правильную гипотезу называется **ошибкой первого ряда** или **уровнем значимости** и обозначается α. Вероятность $\gamma = 1 - \alpha$ принятия верной гипотезы называется **доверительной вероятностью**. Вероятность принять неправильную гипотезу называется **ошибкой второго ряда** и обозначается β. Вероятность $1 - \beta$ принять гипотезу H_1, если она верна, называется **мощностью критерия.**

Случайная величина *K*, используемая для проверки нулевой гипотезы, называется **критерием**.

Совокупность значений критерия, при которых гипотезу отвергают, называется **критической областью**. Совокупность значений критерия, при которых гипотезу принимают (не отвергают), называется **областью принятия гипотез**.

Одной из главных задач математической статистики является установление истинного закона распределения случайной величины на основании эмпирических данных. Вид закона распределения определяют, учитывая исследованные ранее эксперименты, или из теоретических предпосылок.
Проверка гипотезы о предполагаемом законе неизвестного распределения производится при помощи специально подобранной случайной величины, называемой **критерием согласия**. Критерия согласия не доказывает справедливость гипотезы, а лишь устанавливает на принятом уровне значимости ее согласие или несогласие с данными наблюдений.

Пусть $\{x_1, x_2, \dots, x_n\}$ - выборка из некоторой генеральной совокупности *X*. Рассмотрим механизм проверки статистических гипотез с помощью критерия χ^2 Пирсона.
Для того, чтобы при заданном уровне значимости α проверить нулевую гипотезу H_0: генеральная совокупность распределена по предполагаемому закону, необходимо:

1. Вычислить наблюдаемое значение критерия

$$\chi^2_{набл} = \sum_{i=1}^{n} \frac{(n_i - n'_i)^2}{n'_i},$$

где n_i - эмпирические (наблюдаемые) частоты, n'_i - теоретические частоты.

2. По таблице критических точек распределения χ^2(приложение 3), используя заданный уровень значимости α и число степеней свободы $k = s - r - 1$, где s - число групп или интервалов эмпирического распределения выборки, r - число параметров предполагаемого распределения, оцениваемых по данным выборки, находят критическое значение критерия

$$\chi^2_{кр} = \chi^2(\alpha, k) = \chi^2(\alpha, s - r - 1).$$

3. Если $\chi^2_{набл} > \chi^2_{кр}$ нулевая гипотеза H_0 отвергается; если $\chi^2_{набл} < \chi^2_{кр}$ нет оснований отвергнуть нулевую гипотеза H_0.

Замечание 1. При использовании критерия согласия χ^2 объем выборки должен быть достаточно велик (не менее 50 наблюдений), в каждой группе должно быть не менее 5 вариант. Малочисленные группы необходимо объединить, суммируя частоты.

Замечание 2. Эмпирические частоты определяют по экспериментальным данным, а теоретические частоты, если они неизвестны по условиям задания, необходимо вычислить в предположении определенного закона распределения.

Задание 10

Для контроля взяты 200 узлов, собранных на ученическом конвейере. Число узлов m_i, при сборке которых пропущено i операций, сведено в таблицу

Таблица 3

i	0	1	2	3	4	5	6	7	
m_i	41	62	45	22	16	8	4	2	Всего 200

Согласуются ли полученные результаты с распределение Пуассона ($P(\xi = i) = \dfrac{a^i}{i!}e^{-a}$, где ξ - случайное число пропущенных операций) по критерию χ^2 при уровне значимости $\alpha = 0{,}05$? Решить задачу для a=1,85 и для случая, когда параметр a оценивается по данным выборки.

Решение

а) 1. Предполагаемый закон Пуассона имеет вид $P_{200}(i) = \dfrac{1{,}85^i}{i!}e^{-1{,}85}$. Найдем вероятности того, что при сборке пропущено i=0,1,2,3,4,5,6,7 операций:

$$P_0 = P_{200}(0) = \frac{1{,}85^0}{0!}e^{-1{,}85} = 0{,}1572, \quad P_1 = P_{200}(1) = \frac{1{,}85^1}{1!}e^{-1{,}85} = 0{,}2909,$$

$$P_2 = P_{200}(2) = \frac{1{,}85^2}{2!}e^{-1{,}85} = 0{,}2691, \quad P_3 = P_{200}(3) = \frac{1{,}85^3}{3!}e^{-1{,}85} = 0{,}1659,$$

$$P_4 = P_{200}(4) = \frac{1{,}85^4}{4!}e^{-1{,}85} = 0{,}0767, \quad P_5 = P_{200}(5) = \frac{1{,}85^5}{5!}e^{-1{,}85} = 0{,}0284,$$

$$P_6 = P_{200}(6) = \frac{1{,}85^6}{6!}e^{-1{,}85} = 0{,}0088, \quad P_7 = P_{200}(7) = \frac{1{,}85^7}{7!}e^{-1{,}85} = 0{,}0023.$$

2. Найдем теоретические частоты по формуле $m_i' = mP_i = 200P_i$. Используя значения P_i, найденные в п.1, получим

$$m_0' = 200 \cdot 0{,}1572 = 31{,}44, \qquad m_1' = 200 \cdot 0{,}2909 = 58{,}18,$$

$$m_2' = 200 \cdot 0{,}2691 = 53{,}82, \qquad m_3' = 200 \cdot 0{,}1659 = 33{,}18,$$

$$m_4' = 200 \cdot 0{,}0767 = 15{,}34, \qquad m_5' = 200 \cdot 0{,}0284 = 5{,}68,$$

$$m_6' = 200 \cdot 0{,}0088 = 1{,}76, \qquad m_7' = 200 \cdot 0{,}0023 = 0{,}46.$$

3. Сравним теоретические и эмпирические частоты с помощью критерия Пирсона. Для этого найдем наблюдаемое значение критерия $\chi^2_{набл} = \sum_{i=0}^{7} \frac{(m_i - m_i')^2}{m_i'}$, составив расчетную таблицу (табл. 4). Учитывая замечание 1, объединим малочисленные теоретические и эмпирические частоты при i=6 и i=7. Из расчетной таблицы находим $\chi^2_{набл} = 15{,}7824$.

Таблица 4

i	m_i	m_i'	$\frac{(m_i - m_i')^2}{m_i'}$
0	41	31,44	2,9069
1	62	58,18	0,2508
2	45	53,82	1,4454
3	22	33,18	3,7671
4	16	15,34	0,0284
5	8	5,68	0,9476
6	6	2,22	6,4362
Сумма			$\chi^2_{набл} = 15{,}7824$

4. По таблице критических точек распределения χ^2(приложение 3), используя заданный уровень значимости α=0,05 и число степеней свободы $k = s - r - 1$, где число групп эмпирического распределения выборки s=7, число параметров распределения Пуассона r=1, найдем критическую точку правосторонней критической области

$$\chi^2_{кр} = \chi^2(\alpha, k) = \chi^2(0{,}05;5) = 11{,}07.$$

5. В данном случае $\chi^2_{набл} > \chi^2_{кр}$, поэтому нулевая гипотеза H_0 о распределении Пуассона отвергается, то есть эмпирические и теоретические частоты различаются значимо. Это означает, что данные наблюдений не согласуются с гипотезой о распределении Пуассона.

б) 1. Найдем значение выборочного среднего

$$\overline{\xi}_B = \frac{\sum_{i=0}^{7} \xi_i m_i}{m} = \frac{41 \cdot 0 + 62 \cdot 1 + 45 \cdot 2 + 22 \cdot 3 + 16 \cdot 4 + 8 \cdot 5 + 4 \cdot 6 + 2 \cdot 7}{200} = 1{,}8$$

2. В качестве оценки параметра a распределения Пуассона примем выборочное среднее значение $\overline{\xi}_B = 1{,}8$, поэтому предполагаемый закон Пуассона принимает вид $P_{200}(i) = \frac{1{,}8^i}{i!} e^{-1{,}8}$. Найдем вероятности того, что при сборке пропущено i=0,1,2,3,4,5,6,7 операций:

$$P_0 = P_{200}(0) = \frac{1{,}8^0}{0!} e^{-1{,}8} = 0{,}1653, \quad P_1 = P_{200}(1) = \frac{1{,}8^1}{1!} e^{-1{,}8} = 0{,}2975,$$

$$P_2 = P_{200}(2) = \frac{1{,}8^2}{2!} e^{-1{,}8} = 0{,}2678, \quad P_3 = P_{200}(3) = \frac{1{,}8^3}{3!} e^{-1{,}8} = 0{,}1607,$$

$$P_4 = P_{200}(4) = \frac{1{,}8^4}{4!} e^{-1{,}8} = 0{,}0723, \quad P_5 = P_{200}(5) = \frac{1{,}8^5}{5!} e^{-1{,}8} = 0{,}026,$$

$$P_6 = P_{200}(6) = \frac{1{,}8^6}{6!} e^{-1{,}8} = 0{,}0078, \quad P_7 = P_{200}(7) = \frac{1{,}8^7}{7!} e^{-1{,}8} = 0{,}002.$$

2. Найдем теоретические частоты по формуле $m_i' = mP_i = 200P_i$. Используя значения P_i, найденные в п.1, получим

$$m_0' = 200 \cdot 0{,}1653 = 33{,}06, \qquad m_1' = 200 \cdot 0{,}2975 = 59{,}5,$$

$$m_2' = 200 \cdot 0{,}2678 = 53{,}56, \qquad m_3' = 200 \cdot 0{,}1607 = 32{,}14,$$

$$m_4' = 200 \cdot 0{,}0723 = 14{,}46, \qquad m_5' = 200 \cdot 0{,}026 = 5{,}2,$$

$$m_6' = 200 \cdot 0{,}0078 = 1{,}56, \qquad m_7' = 200 \cdot 0{,}002 = 0{,}4.$$

3. Сравним теоретические и эмпирические частоты с помощью критерия Пирсона. Для этого найдем наблюдаемое значение критерия

$\chi^2_{набл} = \sum_{i=0}^{7} \frac{(m_i - m_i')^2}{m_i'}$, составив расчетную таблицу (табл. 4). Учитывая замечание 1, объединим малочисленные теоретические и эмпирические частоты при $i=6$ и $i=7$. Из расчетной таблицы находим $\chi^2_{набл} = 16{,}5781$.

Таблица 4

i	m_i	m_i'	$\frac{(m_i - m_i')^2}{m_i'}$
0	41	33,06	1,9069
1	62	59,5	0,105
2	45	53,56	1,3681
3	22	32,14	3,1991
4	16	14,46	0,164
5	8	5,2	1,5077
6	6	1,96	8,3273
Сумма			$\chi^2_{набл} = 16{,}5781$

4. По таблице критических точек распределения χ^2 (приложение 3), используя заданный уровень значимости $\alpha = 0{,}05$ и число степеней свободы $k = s - r - 1$, где число групп эмпирического распределения выборки $s=7$, число параметров распределения Пуассона $r=1$, найдем критическую точку правосторонней критической области

$$\chi^2_{кр} = \chi^2(\alpha, k) = \chi^2(0{,}05;5) = 11{,}07.$$

5. В данном случае $\chi^2_{набл} > \chi^2_{кр}$, поэтому нулевая гипотеза H_0 о распределении Пуассона отвергается, то есть эмпирические и теоретические частоты различаются значимо. Это означает, что данные наблюдений не согласуются с гипотезой о распределении Пуассона.

<u>Задача.</u> Используя критерий Пирсона, при уровне значимости проверить, согласуется ли гипотеза о нормальном распределении признак генеральной совокупности с заданным распределением выборки (в табл указаны частичные интервалы, во втором – соответствующие им частоты)

1. $\alpha=0{,}05$

(2;12)	7
(12;22)	8
(22;32)	15
(32;42)	36
(42;52)	15
(52;62)	11
(62;72)	8

2. $\alpha=0{,}05$

(1,5;3,5)	4
(3,5;5,5)	18
(5,5;7,5)	12
(7,5;9,5)	35
(9,5;11,5)	15
(11,5;13,5)	10
(13,5;15,5)	6