

МЭСИ

Индивидуальное домашнее задание

Часть 2. Математическая статистика

$$K=6$$

1. Распределение случайной величины X - заработной платы сотрудников на фирме (в у.е.) - задано в виде интервального ряда:

$X_{\min i} (a_i)$	300	$310+10*k$	$320+20*k$	$330+30*k$	$340+40*k$	$350+50*k$
$X_{\max i} (b_i)$	$310+10*k$	$320+20*k$	$330+30*k$	$340+40*k$	$350+50*k$	$360+60*k$
Частота m_i	10	20	30	25	10	5

Найти: \bar{X} , S_x . Построить теоретическое нормальное распределение и сравнить его с эмпирическим с помощью критерия согласия Пирсона χ^2 при $\alpha=0,05$.

Решение. Получим таблицу:

X_{\min}	300	370	440	510	580	650
X_{\max}	370	440	510	580	650	720
m	10	20	30	25	10	5

Найдем точечные оценки математического ожидания и среднего квадратического отклонения: выборочное среднее \bar{X} и исправленное выборочное среднее квадратическое отклонение s . Для этого перейдем к простому вариационному ряду, выбрав в качестве вариант середины интервалов.

x_i	m_i	$x_i m_i$	$(x_i - \bar{X})^2 m_i$
335	10	3350	237160
405	20	8100	141120
475	30	14250	5880
545	25	13625	78400
615	10	6150	158760
685	5	3425	192080
Сумма	100	48900	813400

Выборочное среднее $\bar{X} = \frac{1}{m} \sum x_i m_i = \frac{1}{100} 48900 = 489$.

Выборочная дисперсия $\bar{D} = \frac{1}{m} \sum (x_i - \bar{X})^2 m_i = \frac{1}{100} 813400 = 8134$

Исправленная дисперсия $s^2 = \frac{m}{m-1} \bar{D} = 8216,2$

Исправленное выборочное среднеквадратическое отклонение $s = 90,64$

Теоретическое нормальное распределение имеет плотность

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right) = \frac{1}{90,64\sqrt{2\pi}} \exp\left(-\frac{(x-489)^2}{2*90,64^2}\right)$$

С помощью критерия Пирсона сравним теоретическое распределение с эмпирическим. Проверим гипотезу H_0 о нормальном законе распределения наблюдаемой случайной величины X при уровне значимости $\alpha = 0,05$.

Пронормируем случайную величину X , то есть перейдем к величине $Z = \frac{x - \bar{x}}{\sigma}$,

вычислим концы интервалов по формулам $z_i = \frac{x_i - \bar{X}}{s}$, $z_{i+1} = \frac{x_{i+1} - \bar{X}}{s}$.

Вычислим теоретические (выравнивающие частоты) $m_i' = mP_i$, где $m = 100$,

$P_i = \Phi(z_{i+1}) - \Phi(z_i)$ - вероятность попадания в интервал (x_i, x_{i+1}) , $\Phi(z)$ - функция Лапласа. Для нахождения значений составим расчетную таблицу:

Интервал		Нормированный интервал		$\Phi(z_i)$	$\Phi(z_{i+1})$	P_i	m_i'
x_i	x_{i+1}	z_i	z_{i+1}				
300	370	$-\infty$	-1,313	-0,5	-	0,0951	9,51
370	440	-1,313	-0,541	0,4049	-	0,1995	19,95
440	510	-0,541	0,232	0,2054	0,091	0,2964	29,64
510	580	0,232	1,004	0,091	0,3413	0,2503	25,03
580	650	1,004	1,776	0,3413	0,4625	0,1212	12,12
650	720	1,776	∞	0,4625	0,5	0,0375	3,75

Теперь сравним эмпирические и теоретические частоты, используя критерий Пирсона:

$$\chi^2 = \sum \frac{(m_i - m_i')^2}{m_i'}$$

Результаты вычислений занесем в таблицу:

m_i	m_i'	$\frac{(m_i - m_i')^2}{m_i'}$
10	9,51	0,025
20	19,95	0,000
30	29,64	0,004
25	25,03	0,000
10	12,12	0,371
5	3,75	0,417

0,817

Получили наблюдаемое значение критерия $\chi^2_{\text{набл.}} = 0,817$. По таблице критических точек распределения χ^2 по уровню значимости $\alpha = 0,05$ и числу степеней свободы $k = 6 - 3 = 3$, находим $\chi^2_{\text{кр.}} = 7,8$. Так как $\chi^2_{\text{набл.}} = 0,817 < \chi^2_{\text{кр.}} = 7,8$, то нет оснований отвергнуть гипотезу о нормальном распределении данной величины.

2. В процессе исследования среднедушевого дохода (в руб.) обследовано 100 семей. Выявлены оценки: $\bar{X} = (1500 + 10 \cdot k)$, $S = (200 + k)$. В предположении о нормальном законе найти долю семей, чей среднедушевой доход находится в пределах от 1200 до 1800.

2. В процессе исследования среднедушевого дохода (в руб.) обследовано 100 семей. Выявлены оценки: $\bar{X} = 1560$, $S = 206$. В предположении о нормальном законе найти долю семей, чей среднедушевой доход находится в пределах от 1200 до 1800.

Решение.

Используем формулу для нормального распределения:

$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right)$. Подставляем наши значения (берем точечные оценки, полученные по выборке) $a = 1560$, $\sigma = 206$, $\alpha = 1200$, $\beta = 1800$ и получаем:

$$P(1200 < X < 1800) = \Phi\left(\frac{1800-1560}{206}\right) - \Phi\left(\frac{1200-1560}{206}\right) = \\ = \Phi(1,17) - \Phi(-1,75) = \Phi(1,17) + \Phi(1,75) = 0,3780 + 0,45977 = 0,83777 = 83,77\%$$

Ответ: 83,77% семей.

3. Объем дневной выручки в 5 торговых точках (в тыс. у.е.) составил:
 $x_1=(10+k)$, $x_2=(15+k)$, $x_3=(20+k)$, $x_4=(17+k)$, x_5 . Учитывая, что $\bar{X}=(16+k)$, найти выборочную дисперсию S^2 .

3. Объем дневной выручки в 5 торговых точках (в тыс. у.е.) составил:
16, 21, 26, 23, x_5

Учитывая, что $\bar{X}=22$, найти выборочную дисперсию S^2 .

Решение.

Так как $\bar{X} = \frac{1}{n}(X_1 + X_2 + X_3 + X_4 + X_5)$, получаем уравнение относительно неизвестного X_5 :

$$22 = \frac{1}{5}(16 + 21 + 26 + 23 + X_5),$$

$$X_5 = 24.$$

Теперь найдем выборочную дисперсию

$$S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2 = \frac{1}{4}(36 + 1 + 16 + 1 + 4) = \frac{58}{4} = 14,5$$

Ответ: 14,5.

4. По данным 17 сотрудников фирмы, где работает $(200+10*k)$ человек, среднемесячная заработная плата составила $(300+10*k)$ у.е., при $S=(70+k)$ у.е.

Какая минимальная сумма должна быть на счету фирмы, чтобы с

вероятностью $\gamma=0,98$ гарантировать выдачу заработной платы всем
сотрудникам?

4. По данным 17 сотрудников фирмы, где работает 260 человек,
среднемесячная заработная плата составила 360 у.е., при $S=76$ у.е. Какая
минимальная сумма должна быть на счету фирмы, чтобы с вероятностью
 $\gamma=0,98$ гарантировать выдачу заработной платы всем сотрудникам?

Решение. Найдем доверительный интервал для математического ожидания
(средней зарплаты на одного работника) по формуле:

$$\bar{X} - t \frac{S}{\sqrt{n}} \sqrt{1 - \frac{n}{N}} < a < \bar{X} + t \frac{S}{\sqrt{n}} \sqrt{1 - \frac{n}{N}},$$

где $n=17$, $N=260$, $\bar{X}=360$, $S=76$, t определяется по доверительной вероятности
из Лапласа $t(0,98) = \Phi^{-1}(0,98/2) = \Phi^{-1}(0,49) = 2,33$. Получаем:

$$360 - 2,33 \frac{76}{\sqrt{17}} \sqrt{1 - \frac{17}{260}} < a < 360 + 2,33 \frac{76}{\sqrt{17}} \sqrt{1 - \frac{17}{260}}$$

$$318,48 < a < 401,52.$$

Тогда для всех сотрудников фонд заработной платы находится в пределах от
82804,7 у.е. до 104395,3 у.е. На счету должна быть сумма 104395,3 у.е., чтобы
обеспечить выплату зарплаты всем сотрудникам с вероятностью 0,98.

Ответ: 104395,3 у.е.

5. С целью размещения рекламы опрошено $(400+10*k)$ телезрителей, из которых данную передачу смотрят $(150+10*k)$ человек. С доверительной вероятностью $\gamma=0,91$ найти долю телезрителей, охваченных рекламой в лучшем случае.

5. С целью размещения рекламы опрошено 460 телезрителей, из которых данную передачу смотрят 210 человек. С доверительной вероятностью 0,91 найти долю телезрителей, охваченных рекламой в лучшем случае.

Решение. Для оценки неизвестной доли телезрителей используем формулу:

$w-t\sqrt{\frac{w(1-w)}{n}} < p < w+t\sqrt{\frac{w(1-w)}{n}}$, где $n=460$, $w=\frac{210}{460}=\frac{21}{46}$ - относительная частота,

$t=\Phi^{-1}(\gamma/2)=\Phi^{-1}(0,91/2)=\Phi^{-1}(0,455)=1,695$. Подставим данные:

$$\frac{21}{46}-1,695\sqrt{\frac{21/46(1-21/46)}{460}} < p < \frac{21}{46}+1,695\sqrt{\frac{21/46(1-21/46)}{460}},$$

$$0,417 < p < 0,496.$$

Лучший вариант охвата телезрителей – 49,6% телезрителей.

Ответ: 49,6%.

6. Согласно рекламе автомобиль должен расходовать на 100 км пробега не более 8 л бензина. Проведено 10 испытаний, по результатам которых найден средний расход бензина $\bar{X}=(10+0,1*k)$ л на 100 км, при среднеквадратическом отклонении $S=(1+0,1*k)$ л на 100 км. Проверить справедливость рекламы при $\alpha=0,05$.

6. Согласно рекламе автомобиль должен расходовать на 100 км пробега не более 8 л бензина. Проведено 10 испытаний, по результатам которых найден средний расход бензина $\bar{X}=10,6$ л на 100 км, при среднеквадратическом отклонении $S=1,6$ л на 100 км. Проверить справедливость рекламы при $\alpha=0,05$.

Решение. Пусть a - средний расход бензина на 100 км пробега. Введем гипотезу $H_0: a=8$ при конкурирующей гипотезе $H_1: a>8$ и проверим ее. Вычислим наблюдаемое значение критерия:

$$U_{\text{набл.}} = \frac{\bar{X}-a}{S}\sqrt{n} = \frac{10,6-8}{1,6}\sqrt{10} \approx 5,14.$$

Найдем критическую точку односторонней критической области $U_{\text{кр.}}$ из соотношения:

$$\Phi(U_{\text{эд.}}) = \frac{1-2\alpha}{2} = \frac{1-0,1}{2} = 0,45, \Rightarrow U_{\text{эд.}} = 1,645.$$

Так как $U_{\text{набл.}} > U_{\text{кр.}}$, нулевую гипотезу следует отвергнуть при данном уровне значимости, данные рекламы не соответствуют действительности.

7. Фирма утверждает, что контролирует 40% регионального рынка. Проверить справедливость этого утверждения при $\alpha=0,05$, если услугами этой фирмы пользуются $(100+10 \cdot k)$ человек из $(300+10 \cdot k)$ опрошенных.

7. Фирма утверждает, что контролирует 40% регионального рынка. Проверить справедливость этого утверждения при $\alpha=0,05$, если услугами этой фирмы пользуются 160 человек из 360 опрошенных.

Решение. Введем нулевую гипотезу $H_0: p = 0,4$ при конкурирующей гипотезе $H_1: p \neq 0,4$. Вычислим наблюдаемое значение критерия по формуле:

$$U_{\text{набл.}} = \frac{m/n - p}{\sqrt{pq}} \cdot \sqrt{n},$$

где $n = 360$, $m = 160$, $p = 0,4$, $q = 1 - p = 0,6$. Получаем:

$$U_{\text{набл.}} = \frac{160/360 - 0,4}{\sqrt{0,4 \cdot 0,6}} \cdot \sqrt{360} \approx 1,721.$$

Найдем критическую точку $U_{\text{кр.}}$ из условия $\Phi(U_{\text{кр.}}) = \frac{1-\alpha}{2} = \frac{1-0,05}{2} = 0,475$, $U_{\text{кр.}} = 1,96$.

Так как $|U_{\text{набл.}}| = 1,721 < 1,96 = U_{\text{кр.}}$, нулевую гипотезу можно принять при данном уровне значимости, утверждение фирмы справедливо.

8. Для сравнения существующего технологического процесса с новым по себестоимости продукции было изготовлено $n_x = (5+k)$ изделий по существующей технологии и получена средняя себестоимость продукции $\bar{X} = (13+k)$, $S_x^2 = (1+k)$. Для нового технологического процесса после изготовления $n_y = (8+k)$ изделий получили $\bar{Y} = (9+k)$, $S_y^2 = (2+k)$. Целесообразно ли при $\alpha=0,05$ вводить новую технологию?

8. Для сравнения существующего технологического процесса с новым по себестоимости продукции было изготовлено $n_x=11$ изделий по существующей технологии и получена средняя себестоимость продукции $\bar{X}=19$, $S_x^2=7$. Для нового технологического процесса после изготовления $n_y=14$ изделий получили $\bar{Y}=15$, $S_y^2=8$. Целесообразно ли при $\alpha=0,05$ вводить новую технологию?

Решение. Исправленные дисперсии различны, поэтому предварительно нужно проверить гипотезу о равенстве генеральных дисперсий по критерию Фишера-Снедекора. Вычислим наблюдаемое значение критерия (отношение большей дисперсии к меньшей) $F_{набл.} = \frac{8}{7} = 1,143$.

Найдем критическую точку односторонней критической области: $F_{кр.}(0,05, 14-1, 11-1) = F_{кр.}(0,05, 13,10) = 2,67$. Поскольку $F_{набл.} = 1,143 < 2,67 = F_{кр.}$, нет оснований отвергнуть нулевую гипотезу о равенстве генеральных дисперсий.

Теперь сравним себестоимости продукции (средние). Введем нулевую гипотезу $H_0: MX = MY$ при конкурирующей $H_1: MX > MY$. Найдем наблюдаемое значение критерия Стьюдента по формуле

$$T_{набл.} = \frac{\bar{X} - \bar{Y}}{\sqrt{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}} \sqrt{\frac{n_x n_y (n_x + n_y - 2)}{n_x + n_y}} = \frac{19 - 15}{\sqrt{10 \cdot 7 + 13 \cdot 8}} \sqrt{\frac{11 \cdot 14 (11 + 14 - 2)}{11 + 14}} \approx 3,609$$

Найдем критическую точку $t_{прав.кр.}(0,05, 11+14-2) = t_{прав.кр.}(0,05, 23) = 1,714$.

Поскольку $T_{набл.} = 3,609 > 1,714 = t_{прав.кр.}$, нулевую гипотезу следует отвергнуть. Вводить новую технологию целесообразно из-за более низкой себестоимости.

9. Из $(200+10*k)$ задач по теории вероятностей студенты решили $(110+10*k)$ задач, а из $(300+20*k)$ задач по математической статистике они решили $(140+30*k)$ задач. Можно ли при $\alpha=0,05$ утверждать, что оба раздела усвоены одинаково?

9. Из 260 задач по теории вероятностей студенты решили 170 задач, а из 420 задач по математической статистике они решили 320 задач. Можно ли при $\alpha=0,05$ утверждать, что оба раздела усвоены одинаково?

Решение. Пусть p_1 - процент студентов, решающих теорию вероятности, p_2 - процент студентов, решающих математическую статистику. Введем нулевую гипотезу $H_0: p_1 = p_2$ при конкурирующей гипотезе $H_1: p_1 \neq p_2$. Вычислим наблюдаемое значение критерия по формуле:

$$U_{\text{набл.}} = \frac{\frac{m_1}{n_1} - \frac{m_2}{n_2}}{\sqrt{\frac{m_1 + m_2}{n_1 + n_2} \left(1 - \frac{m_1 + m_2}{n_1 + n_2}\right) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \text{ где } m_1 = 170, n_1 = 260, m_2 = 320, n_2 = 420$$

Подставляем

$$U_{\text{набл.}} = \frac{\frac{170}{260} - \frac{320}{420}}{\sqrt{\frac{170 + 320}{260 + 420} \left(1 - \frac{170 + 320}{260 + 420}\right) \left(\frac{1}{260} + \frac{1}{420}\right)}} \approx -3,052$$

Найдем критическую точку $U_{\text{кр.}}$ из условия $\Phi(U_{\text{кр.}}) = \frac{1-\alpha}{2} = \frac{1-0,05}{2} = 0,475$, $U_{\text{кр.}} = 1,96$.

Так как $|U_{\text{набл.}}| = 3,052 > 1,96 = U_{\text{кр.}}$, нулевую гипотезу следует отвергнуть на данном уровне значимости, можно утверждать, что разделы усвоены неодинаково.

10. Исследование 27 семей по среднему доходу (X) и сбережениям (Y) дало результаты: $\bar{X} = (100 - 2 \cdot k)$ у.е., $S_x = (40 - k)$ у.е., $\bar{Y} = (30 + k)$ у.е., $S_y = (20 + k)$ у.е., $\overline{XY} = (3700 + k)$ (у.е.)². При $\alpha = 0,05$ проверить наличие линейной связи между X и Y . Определить размер сбережений семей, имеющих среднюю доход $X = 130$ у.е.

10. Исследование 27 семей по среднему доходу (X) и сбережениям (Y) дало результаты: $\bar{X} = 88$ у.е., $S_x = 34$ у.е., $\bar{Y} = 36$ у.е., $S_y = 26$ у.е., $\overline{XY} = 3706$ (у.е.)². При $\alpha = 0,05$ проверить наличие линейной связи между X и Y . Определить размер сбережений семей, имеющих среднюю доход $X = 130$ у.е.

Решение.

Вычислим выборочный коэффициент корреляции:

$$r = \frac{\overline{XY} - \bar{X} \cdot \bar{Y}}{S_x \cdot S_y} = \frac{3706 - 88 \cdot 36}{34 \cdot 26} \approx 0,609.$$

Проверим гипотезу о значимости коэффициента корреляции. Введем нулевую гипотезу $H_0: r = 0$ и вычислим наблюдаемое значение критерия Стьюдента

$$t_{\text{набл.}} = \sqrt{\frac{r^2}{1 - r^2}} (n - 2) = \sqrt{\frac{0,609^2}{1 - 0,609^2}} (27 - 2) = 4,92.$$

Найдем критическую точку $t_{\text{кр.}}(0,05, 27 - 2) = t_{\text{кр.}}(0,05, 25) = 2,06$. Так как $t_{\text{набл.}} = 4,92 > 2,06 = t_{\text{кр.}}$, нулевую гипотезу следует отвергнуть, коэффициент корреляции значим, между X и Y есть выраженная линейная связь.

Найдем уравнение регрессии Y на X :

$$y - 36 = 0,609 \cdot \frac{26}{34} (x - 88),$$

$$y = 0,466x - 4,982.$$

Определим размер сбережений семей, имеющих среднюю доход $X = 130$ у.е.

$$y(130) = 0,466 \cdot 130 - 4,982 \approx 55,56.$$