

Методические материалы
для выполнения расчетно-графической
работы по математической статистике

Задание на работу

1. Генерировать две повторные выборки с нормальным распределением.
2. Для первой выборки найти выборочное среднее и дисперсию, экстремальные значения, размах, выборочную медиану.
3. Построить для первой выборки график эмпирической функции распределения и гистограмму. Сгладить гистограмму с помощью нормальной плотности, оценив предварительно параметры.
5. Построить оценки параметров для нормального распределения по МНП.
6. Построить доверительные интервалы для среднего и дисперсии.
7. По критерию χ^2 проверить согласие эмпирических данных с нормальным распределением.
8. Две выборки сгруппировать и проверить однородность по критерию χ^2 .
9. Для первой выборки из нормального распределения проверить гипотезы о параметрах нормального распределения.
10. Для двух независимых выборок из нормального распределения проверить гипотезу о равенстве средних (при условии равенства дисперсий).

Для вычисления критических констант необходимо использовать либо таблицы распределений вероятностей, либо ППП по статистике.

Работа может выполняться либо вручную, либо с помощью ППП по статистике. Но в любом случае в отчете необходимо подробно описать все этапы вычислений и выписать все формулы.

Выполнение работы является **допуском** к сдаче экзамена по

математической статистике.

Ниже приводятся теоретические результаты, необходимые для выполнения работы.

Гистограмма

Гистограмма строится по выборке $x = (x_1, \dots, x_N)$ для наглядного представления экспериментальных данных. Она служит оценкой плотности распределения изучаемой количественной характеристики. Вид графика гистограммы позволяет выдвинуть гипотезу о том, к какому классу принадлежит истинное распределение. Обычно, для сравнения на одном рисунке изображают и график гистограммы и график "сглаженной" плотности распределения из выбранного семейства. Для построения последнего сначала оценивают параметры, а затем вычисляют значения плотности (аналитически и по таблицам) в некоторых точках y_k . После этого приближенно строят график плотности.

Для построения гистограммы выбирают некоторый интервал $[a, b]$, содержащий все выборочные точки x_k . Затем образуют некоторое разбиение этого отрезка:

$$a = a_0 < a_1 < \dots, a_k < a_{k+1} < \dots < a_{r-1} < a_r = b .$$

Это разбиение строят таким образом, чтобы в каждый подотрезок попадало не менее 5 точек. **Гистограммой** называется функция, определяемая по правилу

$$\rho_N(y) = h_k = \frac{N_k}{N \cdot \Delta_k} , \quad a_k < y \leq a_{k+1} , \quad k = \overline{0, r-1} ,$$

где N_k – число элементов выборки, попавших в интервал $(a_k, a_{k+1}]$, $\Delta_k = (a_{k+1} - a_k)$.

Для построения "сглаженной" плотности нужно оценить параметры и вычислить ее значения в некоторых точках y_k . Обычно, в качестве этих точек выбирают середины интервалов разбиения. Например, для семейства нормальных плотностей с параметрами (a, σ^2) оценки строятся по правилу

$$\hat{a} = \bar{x}, \quad \hat{\sigma}^2 = S^2 , q$$

а значения "сглаженной" плотности берут равными

$$\varphi\left(\frac{y_k - \bar{x}}{S}\right) ,$$

где $y_k = (a_k + a_{k+1})/2$, $\varphi(y)$ – стандартная нормальная плотность.

Доверительные интервалы

С. в. ξ имеет нормальное распределение с параметрами (a, σ^2) .
 $X = (X_1, \dots, X_N)$ – повторная выборка.

1. $\theta = a$, σ^2 – известно.

Доверительный интервал уровня γ для параметра a строится по правилу:

$$\bar{X} - \frac{C_\gamma \sigma}{\sqrt{N}} < a < \bar{X} + \frac{C_\gamma \sigma}{\sqrt{N}},$$

где константа C_γ определяется по таблицам нормального распределения из соотношения

$$2\Phi_0(C_\gamma) = \gamma.$$

2. $\theta = a$, σ^2 – неизвестно.

Доверительный интервал уровня γ для параметра a строится по правилу:

$$\bar{X} - \frac{t_{N-1}(\gamma) S_1}{\sqrt{N}} < a < \bar{X} + \frac{t_{N-1}(\gamma) S_1}{\sqrt{N}}$$

где константа $t_{n-1}(\gamma)$ определяется по таблицам распределения Стьюдента с $N - 1$ степенью свободы из соотношения

$$P(|t_{N-1}| < t_{n-1}(\gamma)) = \gamma.$$

3. $\theta = \sigma^2$.

Доверительный интервал уровня γ для параметра σ^2 строится по правилу:

$$\frac{(N-1)S_1^2}{C_2(\gamma)} < \sigma^2 < \frac{(N-1)S_1^2}{C_1(\gamma)},$$

где константы $C_1(\gamma)$ и $C_2(\gamma)$ находим из таблиц χ^2 -распределения с $N - 1$ степенью свободы с помощью соотношений:

$$P(\chi^2 < C_1(\gamma)) = \frac{1 - \gamma}{2} , \quad P(\chi^2 > C_2(\gamma)) = \frac{1 - \gamma}{2} .$$

χ^2 -критерий Пирсона

1. Критерий согласия.

С. в. ξ имеет неизвестную ф. р. $F(y)$. Мы хотим проверить гипотезу

$$H_0 : F(y) = F_0(y) ,$$

где $F_0(y)$ —некоторая фиксированная ф. р.

Пусть $X = (X_1, \dots, X_N)$ есть повторная выборка. Разобьем множество значений с. в. ξ на несколько интервалов:

$$-\infty = a_0 < a_1 < \dots < a_k < a_{k+1} < \dots < a_r < a_{r+1} = \infty . \quad (1)$$

Всего мы имеем $r + 1$ интервал. Пусть n_k равно числу элементов выборки, попавших в интервал $[a_k, a_{k+1})$, а

$$p_k = F_0(a_{k+1}) - F_0(a_k)$$

есть вероятность попадания в этот интервал, если верна H_0 .

При верной гипотезе H_0 с.в.

$$\chi_N^2 = \sum_{k=0}^r \frac{(n_k - Np_k)^2}{Np_k}$$

имеет асимптотически при $N \rightarrow \infty$ χ^2 -распределение с $(r+1) - 1 = r$ степенями свободы. Для заданного α по таблицам χ^2 -распределения найдем константу $\chi_r^2(\alpha)$ из соотношения

$$P(\chi^2 > \chi_r^2(\alpha)) = \alpha .$$

Если реально полученное $\chi_N^2 > \chi_r^2(\alpha)$, то H_0 —неверна. В противном случае говорят, что она не противоречит экспериментальным данным.

В случае сложной гипотезы

$$H_0 : F(y) = F_0(y, \theta_1, \dots, \theta_m) ,$$

где $F_0(y, \theta_1, \dots, \theta_m)$ —известная ф. р., а $\theta = (\theta_1, \dots, \theta_m)$ —неизвестные параметры, мы вначале оцениваем параметры (например, по МНП), а затем вычисляем

$$p_k^* = F_0(a_{k+1}, \theta^*) - F_0(a_k, \theta^*) .$$

Далее все делаем также как в случае простой гипотезы, но χ_N^2 имеет асимптотически χ^2 -распределение с $(r+1) - 1 - m = r - m$ степенями свободы.

2. Критерий однородности.

Пусть мы имеем две независимые выборки $X = (X_1, \dots, X_{N_1})$ и $Y = (Y_1, \dots, Y_{N_2})$ из генеральных совокупностей с ф. р. $F_1(z)$ и $F_2(z)$ соответственно. Мы хотим проверить гипотезу

$$H_0 : F_1(z) = F_2(z) .$$

Для этого вновь образуем разбиение (1) множества значений случайных величин. Пусть n_{k1} есть число элементов выборки X , попавших в интервал $[a_k, a_{k+1})$, а n_{k2} аналогичные числа для второй выборки.

Если гипотеза H_0 верна, то с. в.

$$\chi_N^2 = \sum_{k=0}^r \sum_{j=1}^2 \frac{(n_{kj} - n_k \frac{N_j}{N})^2}{n_k \frac{N_j}{N}} ,$$

где $n_k = n_{k1} + n_{k2}$, имеет асимптотически при $N_1, N_2 \rightarrow \infty$ χ^2 -распределение с r степенями свободы.

Далее процедура проверки гипотезы об однородности строится также как в случае критерия согласия.

Проверка гипотез

Пусть с. в. ξ имеет нормальное распределение с параметрами (a, σ^2) , $X = (X_1, \dots, X_{N_1})$ —повторная выборка.

1. Проверяется гипотеза

$$H_0 : a = a_0$$

против альтернативы

$$H_1 : a \neq a_0 .$$

Для этого используют статистику

$$T = \frac{\bar{X} - a_0}{S_1} \sqrt{N} ,$$

которая при гипотезе H_0 имеет распределение Стьюдента с $N - 1$ степенью свободы. Для заданного α по таблицам находим константу $t_{n-1}(\alpha)$ из соотношения

$$P(|T| > t_{n-1}(\alpha)) = \alpha .$$

Если реально полученное $|T| > t_{n-1}(\alpha)$, то отвергаем H_0 . В противном случае она не противоречит экспериментальным данным.

2. Проверяется гипотеза

$$H_0 : \sigma^2 = \sigma_0^2$$

против альтернативы

$$H_1 : \sigma^2 \neq \sigma_0^2$$

Для этого используют статистику

$$\chi^2 = (N - 1)S_1^2/\sigma_0^2 ,$$

которая при гипотезе H_0 имеет χ^2 -распределение с $N - 1$ степенью свободы. Для заданного α по таблицам находим константы $C_1(\alpha)$ и $C_2(\alpha)$ из соотношений

$$P(\chi^2 < C_1(\alpha)) = \alpha/2 , P(\chi^2 > C_2(\alpha)) = \alpha/2 .$$

Если реально полученное χ^2 таково, что

$$\chi^2 < C_1(\alpha) \text{ или } \chi^2 > C_2(\alpha) .$$

то отвергаем H_0 . В противном случае она не противоречит экспериментальным данным.

3. Пусть мы имеем две независимые повторные выборки $X = (X_1, \dots, X_{N_1})$ и $Y = (Y_1, \dots, Y_{N_2})$ из генеральных совокупностей с нормальными распределениями, параметры которых есть (a_1, σ^2) и (a_2, σ^2) .

Мы хотим проверить гипотезу

$$H_0 : a_1 = a_2$$

против альтернативы

$$H_1 : a_1 \neq a_2 .$$

Для этого используют статистику

$$T = \frac{\sum_{i=1}^{N_1} X_i - \sum_{i=1}^{N_2} Y_i}{\sqrt{(N_1 S_1^2 + N_2 S_2^2)/(N-2)}} \frac{1}{\sqrt{N}} ,$$

где $N = N_1 + N_2$, $N_1 S_1^2 = \sum_{i=1}^{N_1} (X_i - \bar{X})^2$ $N_2 S_2^2 = \sum_{i=1}^{N_2} (Y_i - \bar{Y})^2$, которая при гипотезе H_0 имеет распределение Стьюдента с $N-2$ степенями свободы.

Далее процедура проверки гипотезы H_0 строится также как и в задаче 1.