Практическое занятие № 5
по курсу «Статистические методы обработки данных»

Тема: Идентификация форм распределения экспериментальных данных.

Цель работы: Овладеть методами идентификация форм распределения

Краткая теория

Идентификация формы распределения экспериментальных данных.

Критерии согласия.

Наиболее распространенным является критерий согласия Пирсона (χ^2 - критерий).

Суть критерия состоит в вычислении величины

$$\chi^2_{ýê} = n\sum_{j=1}^{m}\frac{\left(\nu_j - p_j\right)^2}{p_j} \qquad (1),$$

где m – число интервалов, ν_j - частота попадания в j-й интервал, p_j - вероятности в том же j-м столбце, рассчитанном по заданной модели.

Вычисленная $\chi^2_{экспер}$ сравнивается с $\chi^2(\kappa,\alpha)$ с числом степеней свободы $\kappa = m-1-r$, где r – число определяемых по статистике параметров, необходимых для совмещения модели и гистограммы.

Если $\chi^2_{экс} \le \chi^2_{\alpha,\kappa}$, то гипотеза о распределении при данном уровне значимости α не противоречит модели, если $\chi^2 > \chi^2_{\alpha,\kappa}$, то гипотеза отклоняется.

Как правило, χ^2 дает лишь отрицательный ответ, однако на практике используется и как для принятия положительного решения.

Критерий Колмогорова-Смирнова.

В этом критерии используется максимальное значение модуля разности между экспериментальным и теоретическим распределением интегральных функций распределений. Математически это означает:

$$D = \max | F_{экспер}(x) - F_{теор}(x) | / n$$

и сравнивается с граничным значением вероятности

$$P \ge 2\exp\left(-2nD^2\right) = 2\exp(-2\Delta^2 / n)$$

$$\Delta = | F_{экспер} - F_{теор} |$$

Здесь P – искомая вероятность того, что искомая выборка может соответствовать полученной модели.

Можно построить и доверительный интервал следующего типа:

$$P_{дов}[F_{экспер} - d_\alpha \le F_{теор} < F_{экспер} + d_\alpha] = 1 - \alpha,$$

где критические значения $d\alpha$

$$d_\alpha \approx \sqrt{-0.5\ln\left(\frac{\alpha}{2}\right)/n} \qquad (n>35)$$

Здесь α - уровень значимости.

Доверительный интервал представляет собой полосу с шириной $\pm d_\alpha$ около выбранного нами (измеренного) $F_{экспер}$ и с вероятностью $1-\alpha$ истинная функция $F_{теор}(x)$ лежит внутри полосы.

Преимуществом перед χ^2 состоит в том, что не требуется группирование данных в интервалы.

<u>Критерий ω^2 (критерий Мизеса).</u>

Этот критерий также использует не сгруппированные данные. Рассчитывается величина

$$\omega^2 = \int_{-\infty}^{\infty} \left[F_{экспер} - F_{теор}\right]^2 dF_{теор}$$

$$dF_{теор} = f_{теор}(x)dx$$

В итоге для

$$F_{экспер} = \begin{cases} 0, & x < x_1 \\ k/n, & x_k \le x < x_{k+1}, \; k=1\ldots n-1 \\ 1, & x > x_n \end{cases}$$

<u>Получаем</u>

$$\omega^2 = \frac{1}{n}\{\frac{1}{12n} + \sum_{k=1}^{n}\left[F(x_k) - \frac{2k-1}{2n}\right]^2$$

Затем сравнивают $n\omega^2$ с $\left(n\omega^2\right)_{крит}$ с заданным уровнем значимости α.

α	$\left(n\omega^2\right)_{крит}$
0.5	0.1184
0.4	0.1467
0.3	0.1843
0.2	0.2412
0.1	0.3473

α	$\left(n\omega^2\right)_{крит}$
0.05	0.4614
0.03	0.5489
0.02	0.6198
0.01	0.7435
0.001	1.1679

(Эта таблица для n>40).

Здесь более полно используется информация о выборке.

<u>Замечания для критериев согласия</u>.

При использовании критериев согласия обычно оговаривается, что их применение корректно лишь при достаточно «больших» выборках и, как правило, указывается n>200.

Располагая соотношением для выбора m, увидим это более наглядно

$$m = \frac{\varepsilon + 1.5}{6} n^{0.4} \Rightarrow n = \left[\frac{6m}{\varepsilon + 1.5}\right]^{2.5}$$

При использовании χ^2 желательно иметь число отклонений (1-3, как обычно), а хотя бы 7-9-11.

Тогда для нормального распределения ($\varepsilon = 3$) при m=7-11 имеем n=170-800 отсчетов; для равномерного $(\varepsilon = 1.8)$ при m=7-11 необходимо уже n=600-2700 отсчетов. Аналогично оценим объем выборки с использованием критерия Колмогорова. Так для

$$\alpha = 0.05 \qquad d_{\alpha=0.05} = \frac{0.61}{\sqrt{n}}$$

$$n = 100 \qquad d_{\alpha=0.05} = 0.061$$

Но объем 500-2500 для экспериментаторов практически не достижим.

Задание

Задача 1. С помощью критерия χ^2 определить является ли случайная физическая величина нормальным распределением т.е. определить с какой вероятностью данная выборка будет нормальным распределением. Проверку выполнить для <u>каждой</u> из 3-х выборок по варианту.

Задача 2. Проделать задачу 1 с помощью критерия Колмогорова-Смирнова.

Задача 3. Построить графики и сделать визуальное сравнение выбранной модели и экспериментальной кривой. Сделать выводы.

Задача 4. Подготовить отчет.

Составители: Андреев В.В. (2014),
Бабич К.С. (2016)