

Разработка практического задания на формирование УМЕНИЙ
Кейс-задание «Применение вероятностных методов для принятия
экономических решений»
(макс. – 15 баллов)

Задание 1. (Проверка параметрических гипотез)

Цель работы – ознакомление с методикой проверки гипотезы о равенстве двух математических ожиданий и ее использование на конкретном примере.

Краткие теоретические сведения

Обозначим через n и m объемы независимых выборок, по которым найдены соответствующие выборочные средние $\bar{x}_B$ и $\bar{y}_B$. Генеральные дисперсии σ_1^2 и σ_2^2 известны. Требуется по выборочным средним и при заданном уровне значимости проверить нулевую гипотезу, состоящую в том, что генеральные средние (математические ожидания) рассматриваемых совокупностей равны между собой, т.е. $H_0: a_1 = a_2$.

Если окажется, что нулевая гипотеза справедлива, т.е. генеральные средние одинаковы, то различие выборочных средних незначимо и объясняется случайными причинами и, в частности, случайным отбором объектов выборки. Если нулевая гипотеза отвергнута, т.е. генеральные средние неодинаковы, то различие выборочных средних значимо и не может быть объяснено случайными причинами, а объясняется тем, что сами генеральные средние (математические ожидания) различны.

В качестве критерия проверки нулевой гипотезы принимается случайная

$$Z = \frac{\bar{x}_B - \bar{y}_B}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

величина:

Значение критерия, вычисленное по данным наблюдений, обозначается через $Z_{\text{набл}}$.

Критическая область строится в зависимости от вида конкурирующей гипотезы.

Сформулируем правила проверки нулевой гипотезы.

Правило 1. Для того чтобы при заданном уровне значимости проверить нулевую гипотезу $H_0: a_1 = a_2$ о равенстве математических ожиданий (генеральных средних) двух нормальных генеральных совокупностей с известными дисперсиями при конкурирующей гипотезе $H_a: a_1 \neq a_2$, надо вычислить наблюдаемое значение критерия

$$Z_{\text{набл}} = \frac{\bar{x}_B - \bar{y}_B}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

и по таблице функции Лапласа (приложение №2) найти критическую точку $z_{\text{кр}}$ из

$$\Phi(z_{\text{кр}}) = \frac{1-\alpha}{2}.$$

Если $|Z_{\text{набл}}| < z_{\text{кр}}$ – нет оснований отвергнуть нулевую гипотезу. Если $|Z_{\text{набл}}| > z_{\text{кр}}$ – нулевую гипотезу отвергают.

Правило 2. При конкурирующей гипотезе $H_a: a_1 > a_2$ находят критическую точку $z_{\text{кр}}$ по

$$\Phi(z_{\text{кр}}) = \frac{1-\alpha}{2}.$$

Если $z_{\text{набл}} < z_{\text{кр}}$ – нет оснований отвергнуть нулевую гипотезу.

Если $z_{\text{набл}} > z_{\text{кр}}$ – нулевую гипотезу отвергают.

Правило 3. При конкурирующей гипотезе $H_a: a_1 < a_2$ находят критическую точку $z_{кр}$ по правилу 2.

Если $z_{набл} > -z_{кр}$ – нет оснований отвергнуть нулевую гипотезу.

Если $z_{набл} < -z_{кр}$ – нулевую гипотезу отвергают.

Условие задачи. Фирма предлагает автоматы по розливу напитков. При выборке $n=16$ найдена средняя величина $\bar{X}=182$ г дозы, наливаемой в стакан автоматом №1. При выборке $m=9$ найдена средняя величина $\bar{Y}=185$ дозы, наливаемой в стакан автоматом №2. По утверждению изготовителя, случайная величина наливаемой дозы имеет нормальный закон распределения с дисперсией, равной $\sigma_x^2 = \sigma_y^2 = 25$. Можно ли считать отличия выборочных средних случайной ошибкой при уровне значимости $\alpha=0,01$?

Поставлены задачи (ОПК-2, У3):

- 1) Проверить гипотезу: $H_0: a_1 = a_2$
- 2) Проверить гипотезу $H_a: a_1 > a_2$
- 3) Проверить гипотезу $H_a: a_1 < a_2$

Задание 2. (Проверка гипотезы о распределении генеральной совокупности)

Цель работы – ознакомление с методикой проверки гипотезы о распределении генеральной совокупности по критерию Пирсона ее использование на конкретном примере.

Краткие теоретические сведения

Критерий согласия χ^2 Пирсона. Критерий Пирсона позволяет производить проверку согласия эмпирической функции распределения с гипотетической функцией, принадлежащей к некоторому множеству Ω функций определенного вида (нормальных, показательных, биномиальных и т.д.).

Пусть СВ имеет функцию распределения, принадлежащую некоторому классу функций. Из генеральной совокупности извлечена выборка объема $|n \geq 50|$. Разобьем весь диапазон полученных результатов на частичных интервалов равной длины, и пусть в каждом частичном интервале оказалось измерений, причем. Составим сгруппированный статистический ряд распределения частот:

Интервалы значений СВ	наблюдаемых			...	$[x_{i-1}; x_i)$	...	
Частоты				...		...	

Требуется на основе имеющейся информации проверить нулевую гипотезу о том, что гипотетическая функция распределения значимо представляет данную выборку, т.е. $H_0: F(x) \in \Omega$.

При проверке нулевой гипотезы с помощью критерия согласия придерживаются следующей последовательности действий:

1) на основании гипотетической функции вычисляют вероятности попадания СВ в частичные интервалы:

$$p_i = P[x_{i-1} \leq X < x_i] = \int_{x_{i-1}}^{x_i} f(x) dx = F(x_i) - F(x_{i-1}); \quad i = 1, 2, \dots, k$$

2) умножая полученные вероятности p_i на объем выборки, получают теоретические частоты np_i частичных интервалов, т.е. частоты, которые следует ожидать, если нулевая гипотеза справедлива;

$$\chi^2 = \sum_{i=1}^k \frac{|m_i - np_i|^2}{np_i}$$

3) вычисляют выборочную статистику (критерий) :

Замечание 1. При проверке гипотезы о нормальном распределении СВ вероятности попадания СВ в частичные интервалы находят по формуле:

$$p_i = P(x_{i-1} \leq X < x_i) = \Phi\left(\frac{x_i - a}{\sigma}\right) - \Phi\left(\frac{x_{i-1} - a}{\sigma}\right), \text{ где } \Phi - \text{функция Лапласа (приложение).}$$

Если нулевая гипотеза верна, то при распределении выборочной статистики независимо от вида функции стремится к распределению с $v = k - r - 1$ степенями свободы (k – число частичных интервалов; r – число параметров гипотетической функции , оцениваемых по данным выборки).

Критерий сконструирован таким образом, что чем ближе к нулю наблюдаемое значение критерия χ^2 , тем вероятнее, что нулевая гипотеза справедлива. Поэтому для проверки нулевой гипотезы применяется критерий с правосторонней критической областью. Следовательно, для того, чтобы проверить нулевую гипотезу, необходимо найти по таблицам квантилей χ^2 -распределения по заданному уровню значимости и числу степеней свободы критическое значение $\chi_{\alpha;v}^2$, удовлетворяющее условию $P(\chi^2 \geq \chi_{\alpha;v}^2) = \alpha$. Сравнивая наблюдаемое значение выборочной статистики $\chi_{\text{набл}}^2$ с критическим значением $\chi_{\text{кр}}^2$, принимаем одно из двух решений:

1) если $\chi_{\text{набл}}^2 \geq \chi_{\text{кр}}^2$, то нулевая гипотеза отвергается в пользу альтернативной $H_a: F(x) \notin \Omega$, т.е. считается, что гипотетическая функция не согласуется с результатами эксперимента;

2) если $\chi_{\text{набл}}^2 < \chi_{\text{кр}}^2$, то считается, что нет оснований для отклонения нулевой гипотезы, т.е. гипотетическая функция согласуется с результатами эксперимента.

Замечание 2. При применении критерия необходимо, чтобы в каждом частичном интервале было не менее 5 элементов. Если число элементов (частота) меньше 5, то рекомендуется объединять такие частичные интервалы с соседними.

Условие задачи. Масса (в граммах) 30 пачек полуфабриката «Геркулес» такова: 503,509,495,493,489,485,507,511,487,485,506,504,507,511,499,491,494,518,506,515, 487,509,507,488,495,490,498,497,492,495.

Поставлены задачи (ОПК-2, У3):

- 4) Построить статистический ряд распределения относительных частот;
- 5) Построить гистограмму выборочной оценки плотности вероятности;
- 6) Найти выборочную среднюю и выборочную дисперсию по распределению выборки
- 7) Найти несмещенную оценку математического ожидания дисперсии
- 8) Найти доверительный интервал с надежностью 0,95 для оценки математического ожидания нормально распределенной случайной величины;
- 9-10) Проверить гипотезу о нормальном законе распределения случайной величины на уровне значимости $\alpha = 0,05$.

Задание 3. (Корреляционный анализ данных)

Цель работы – ознакомление с методом корреляционно-регрессионного анализа и его применение при анализе линейной связи между негруппированными данными.

Краткие теоретические сведения

Корреляционный анализ — метод статистического исследования экспериментальных данных, позволяющий определить степень линейной зависимости между переменными.

Парная линейная корреляция — простейшая система корреляционной связи, представляющая линейную связь между двумя признаками. Ее практическое значение состоит в выделении одного важнейшего фактора, который и определяет вариацию результативного признака.

Для определения степени тесноты парной линейной зависимости служит *линейный коэффициент корреляции*, который был впервые введен в начале 1890-х гг. Пирсоном, Эджуортом и Велдоном. В теории разработаны и на практике применяются различные варианты формул расчета данного коэффициента:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y} = \frac{\sum (x - \bar{x})(y - \bar{y})}{n \sigma_x \sigma_y};$$

$$\sigma_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}; \quad \sigma_y = \sqrt{\frac{\sum (y - \bar{y})^2}{n}},$$

При малом числе наблюдений для практических вычислений линейный коэффициент корреляции удобнее исчислять по следующей формуле:

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n} \right) \left(\sum y^2 - \frac{(\sum y)^2}{n} \right)}},$$

$$-1 \leq r \leq +1.$$

Линейный коэффициент корреляции принимает значения:

Чем ближе линейный коэффициент корреляции по абсолютной величине к 1, тем теснее связь. С другой стороны, если он равен 1, то зависимость является не стохастической, а функциональной.

Знак при нем указывает направление связи: знак «-» соответствует обратной зависимости, «+» — прямой. Величина коэффициента корреляции служит также оценкой соответствия уравнения регрессии выявленным причинно-следственным связям.

Степень взаимного влияния факторов в зависимости от коэффициента корреляции приведена в табл. 1.

Таблица 1. Количественная оценка тесноты связи при различных значениях коэффициента корреляции

Величина коэффициента корреляции	0,1-0,3	0,3-0,5	0,5-0,7	0,7-0,9	0,9-0,99
Теснота связи	Слабая	Умеренная	Заметная	Высокая	Весьма высокая

Линейная регрессия сводится к нахождению уравнения вида:

$$y_x = a_0 + a_1 x,$$

где x — индивидуальное значение факторного признака; a_0, a_1 — параметры уравнения прямой (уравнения регрессии); y_x — теоретическое значение результирующего фактора.

Данное уравнение показывает среднее значение изменения результативного признака x на одну единицу его измерения. Знак параметра показывает направление этого изменения.

На практике построение линейной регрессии сводится к оценке ее параметров a_0, a_1

При классическом подходе параметры уравнения a_0, a_1 , находятся методом наименьших квадратов, который позволяет получить такие оценки параметров, при которых сумма

квадратов отклонений фактических значений результативного признака y от расчетных, теоретических была бы минимальной.

$$\text{где } a_0 = \frac{\sum y \sum x^2 - \sum xy \sum x}{n \sum x^2 - (\sum x)^2};$$

$$a_1 = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}.$$

Условие задачи(ОПК-2, У3): С целью анализа взаимного влияния зарплаты и текучести рабочей силы на пяти однотипных фирмах с одинаковым числом работников проведены измерения уровня месячной зарплаты X и числа уволившихся за год рабочих Y .

Поставлены задачи:

- 11) Построить поле корреляции
- 12) Найти выборочное уравнение линейной регрессии Y на X
- 13) Найти выборочное уравнение линейной регрессии X на Y
- 14) Найти выборочный коэффициент корреляции
- 15) Построить линии регрессии в поле корреляции.

Внимание!

1. Округление результатов проводить до сотых долей.
2. В СДО Moodle необходимо загрузить отчет с подробным обоснованием результатов анализа экономико-математических моделей. .
Графическое представление результатов исследования может быть выполнено в любом графическом редакторе и представлено сканами или выполняется на клетчатой бумаге.
3. Правильное выполнение каждой задачи оценивается в один балл.