UDC 519.233.33

Elshin V.R., Chekhlystova Yu.A., Glazkova M.Yu. Checking static hypotheses. Comparison of the sample average with the mathematical expectation.

Проверка статических гипотез. Сравнение выборочной средней с математическим ожиданием.

Yelshin Vladislav Romanovich
2nd year student of the specialty Construction
of unique buildings and structures (SUZ-232)
Faculty of Civil Engineering,
Voronezh State Technical University
Chehlystova Julia Alexandrovna
2nd year student of the specialty Construction
of unique buildings and structures (SUZ-233)
Faculty of Civil Engineering,
Voronezh State Technical University
Glazkova Maria Yurievna
Ph.D, Associate Professor
Department of Applied Mathematics and Mechanics
Voronezh State Technical University
Ельшин Владислав Романович
студент 2 курса специальности Строительство
уникальных зданий и сооружений (СУЗ-232)
строительный факультет, Воронежский
государственный технический университет
Чехлыстова Юлия Александровна
студентка 2 курса специальности Строительство
уникальных зданий и сооружений (СУЗ-233)
строительный факультет, Воронежский
государственный технический университет
Глазкова Мария Юрьевна
к. ф.-м. н., доцент кафедры прикладной
математики и механики
Воронежский государственный
технический университет

Abstract. *Testing statistical hypotheses is an extremely urgent task in various spheres of life, since it allows us to draw conclusions based on sample data. Medicine: testing hypotheses about the effects of new drugs, treatment methods, or risk factors affecting human health. Biology: analysis of data on the impact of the environment on animal and plant populations. Physics: Testing theories and models based on experimental data.*

In the course of the work, the basic concepts of mathematical statistics are considered, the stages of testing the statistical hypothesis are described and several mathematical problems are solved, during the solution of which the hypothesis of comparing the sample average with the mathematical expectation is confirmed.

Keywords*: mathematical statistics, general population, sample, statistical hypothesis, statistical criterion.*

Аннотация. *Проверка статистических гипотез является чрезвычайно актуальной задачей в различных сферах жизни, поскольку позволяет нам делать выводы на основе выборочных данных. Медицина: проверка гипотез о влиянии новых лекарств, методик лечения или факторов риска влияющих на здоровье человека. Биология: анализ данных о влиянии окружающей среды на популяции животных и растений. Физика: проверка теорий и моделей, основанных на экспериментальных данных.*

В ходе работы рассмотрены основные понятия математической статистики, описаны этапы проверки статистической гипотезы и решены несколько математических задач, в ходе решения которых подтверждена гипотеза о сравнении выборочной средней с математическим ожиданием.

Ключевые слова: *математическая статистика, генеральная совокупностью, выборка, статистическая гипотеза, статистический критерий.*

Рецензент: Мартеха Александр Николаевич – кандидат технических наук, доцент. Доцент ФГБОУ ВО «РГАУ-МСХА им. К.А. Тимирязева»

Целью работы является подтвердить одну из статических гипотез -сравнить выборочную среднюю с математическим ожиданием.

Определим основные понятия и этапы проверки статистических гипотез, на основе которых будет строится решение задач.

Математическая статистика – это раздел математики, который изучает методы сбора, систематизации, обработки результатов наблюдений массовых случайных явлений.

Любое множество, подлежащее изучению в статистике, называется генеральной совокупностью.

Любое подмножество генеральной совокупности называется выборкой.

Статистической гипотезой называется предположение относительно параметров или вида распределения случайной величины X , где X — наблюдаемая дискретная или непрерывная случайная величина.

Основной или нулевой гипотезой H_0 называют выдвинутую гипотезу, а гипотезу H_1 , ей противоречащую — конкурирующей или альтернативной.

Правило, по которому принимается решение принять или отклонить гипотезу H_0 называют статистическим критерием (K) .

Проверка статистической гипотезы основывается на принципе, соответствии с которым маловероятные события считаются *невозможными*, а события, имеющие большую вероятность, считаются *достоверными*. Этот принцип можно реализовать следующим образом.

Малая вероятность α , называемая уровнем значимости, и равная вероятности отвергнуть правильную H_0 гипотезу.

Уровень значимости α определяет размер «критической области».

Критическая область V_K — те значения критерия K , при которых гипотезу H_0 отвергают.

Генеральная дисперсия (σ^2) — мера рассеивания данных во *всей* генеральной совокупности. Показывает, насколько сильно значения в генеральной совокупности разбросаны вокруг среднего значения.

Выборочная дисперсия (S^2) — мера рассеивания данных в *выборке*. Показывает, насколько сильно значения в выборке разбросаны вокруг среднего значения выборки.

В зависимости от принятого уровня значимости из области допустимых значений функции критерия K выделяют критическую область V_k. Далее руководствуются следующим правилом: если вычисленное по выборке значение критерия K попадает в критическую область, то от H_0 отвергается и принимается гипотеза H_1. При этом возможно, что H_0 справедлива и, следовательно, совершена ошибка первого рода, вероятность которой α.

Возможны 3 варианта расположения критической области:

Правосторонняя критическая область, состоящая из интервала ($k_{\text{кр}}^{\text{п}} > +\infty$), где $k_{\text{кр}}^{\text{п}}$ определяется из условия $P(K > k_{\text{кр}}^{\text{п}}) = \alpha$.

Левосторонняя критическая область, состоящая из интервала ($-\infty > k_{\text{кр}}^{\text{л}}$), где $k_{\text{кр}}^{\text{л}}$ определяется из условия $P(K < k_{\text{кр}}^{\text{л}}) = \alpha$

Двусторонняя критическая область, состоящая из интервалов ($-\infty$; $k_{\text{кр}}^{\text{л}}$) и ($k_{\text{кр}}^{\text{п}}$; $+\infty$), где точки $k_{\text{кр}}^{\text{л}}$ и $k_{\text{кр}}^{\text{п}}$ определяются из условий $P(K < k_{\text{кр}}^{\text{л}}) = \frac{\alpha}{2}$ и $P(K > k_{\text{кр}}^{\text{п}}) = \frac{\alpha}{2}$

Рассмотрим этапы проверки статической гипотезы.

1. Назначить уровень значимости (α).
2. Выбрать статистический критерий.
3. Сформулировать проверяемую (H_0) и альтернативную (H_1) гипотезы.
4. Определить критическую область V_K.
5. Определить теоретическое (K_T) и выборочное (K_B) значения критерия.
6. Принять статистическое решение: если $K_B \notin V_K$, то гипотезу H_0 принять, т. е. считать, что гипотеза H_0 не противоречит результатам наблюдений.
7. Если $K_B \in V_K$, то отклонить гипотезу H_0 как не согласующуюся с результатами наблюдений.

На практике часто требуется оценить, соответствует ли действительности регламентированные данные о параметрах того или иного товара. В это случае возникает задача сравнения выборочной средней с заявленным значением этого параметра. Рассмотрим один из примеров.

Фармацевтическая компания разработала новый препарат для снижения артериального давления. Известно, что у здоровых людей среднее артериальное давление составляет 120 мм рт. ст. (это математическое ожидание). Для проверки эффективности препарата был проведен эксперимент с участием 20 пациентов. После приема препарата в течение месяца у них были замерены следующие значения артериального давления: 115, 118, 112, 125, 110, 116, 119, 108, 114, 122, 117, 109, 113, 120, 111, 121, 115, 107, 118, 123 (мм. рт. ст.)

Генеральная дисперсия артериального давления после приема препарата неизвестна.

Задача: На уровне значимости α = 0.05, можно ли утверждать, что новый препарат снижает артериальное давление? Проверить статистический анализ и сформулировать вывод.

Решаем пример в соответствии с этапами проверки статистической гипотезы:

1. $\alpha = 0{,}05$
2. Предположим, что случайная величина артериального давления подчинена нормальному закону распределению (при котором значения случайной величины распределяются *симметрично*.)

Требуется проверить гипотезу о числовом значении математического ожидания нормально распределенной величины при неизвестной генеральной дисперсии

1) $\bar{X} = \frac{\sum_{k=1}^{m} X_k}{n} = \frac{2300}{20} = 115$ (мм. рт. ст.)

2) $S^2 = \frac{\sum_{i=1}^{m}(x_i - \bar{x})^2}{n-1} = \frac{(115\text{-}115)^2 + (118\text{-}115)^2 + \ldots + (123\text{-}115)^2}{20-1} = 22{,}53$

–выборочная дисперсия

3) S = $\sqrt{s^2} = \sqrt{22{,}53} = 4,\ 75$ (мм. рт. ст.) – исправленное выборочное стандартное отклонение.

$K_B = \frac{\bar{X} - m_0}{S/\sqrt{n-1}}$ — функция для подсчета статистического критерия для выборки, где $\bar{X}$ – выборочная средняя, m_0 –математическое ожидание, S– исправленное выборочное стандартное отклонение.

Случайная величина K_B имеет t-распределение (распределение Стьюдента) с $l = n - 1$, где l – степень свободы, n – выборка.

3. Нулевая гипотеза (H_0): Препарат не влияет на артериальное давление (m_0 = 120).

Альтернативная гипотеза (H_1): Препарат снижает артериальное давление (m_0 < 120).

Очевидно, что другие альтернативные гипотезы (m_0 >120, m_0 ≠120) нецелесообразны, так как нас интересует только снижение давления.

4. Следующим шагом требуется найти критическую область для нулевой гипотезы. Критическая область левосторонняя. $k_{\text{кр}}^{\text{л}}$ находим из условия P(K < $k_{\text{кр}}^{\text{л}}$)= α.

5.1 При α = 0,05 и $l = 20 - 1 = 19$ в таблице t-распределения находим $k_{\text{кр}}^{\text{л}}$ = -$k_{\text{кр}}^{\text{п}}$= -1,73 (по таблице t-распределения Стьюдента)

Таким образом, критическая область V_K= (-∞; −1,73).

5.2 Рассчитаем $K_B = \frac{\bar{X}-m_0}{S/\sqrt{n-1}} = \frac{115-120}{4,75/\sqrt{20-1}} = -4.71 \in$ (-∞; −1,73).

6. Значение попадает в критическую область, поэтому нулевая гипотеза H_0 должна быть отвергнута. Следовательно, на уровне значимости 0.05 есть статистически значимые основания утверждать, что новый препарат снижает артериальное давление. Полученное значение t-статистики значительно меньше критического значения, что свидетельствует о высокой степени достоверности результата.

Рассмотрим еще один пример.

Фармацевтическая компания производит лекарство для снижения уровня холестерина. В результате многолетних исследований и огромного числа пациентов, компания точно знает, что уровень холестерина после курса лечения составляет m_0 = 4,2 (ммоль/л) при известной генеральной дисперсии $\sigma^2 = 0,25$ (ммоль/л).

Новая партия лекарства была произведена с использованием незначительно измененного технологического процесса. Для проверки качества новой партии было отобрано случайная выборка из n = 100 пациентов. После курса лечения с новой партией лекарства, средний уровень холестерина в этой выборке составил $\overline{X}$ = 3,8 (ммоль/л).

На уровне значимости α = 0.05, можно ли утверждать, что новый технологический процесс повлиял на эффективность препарата, то есть изменил средний уровень холестерина после лечения?

Решаем задачу.

1.α = 0.05

2.Предположим, что уровень холестерина в крови, есть случайная величина, подчиненная нормальному закону распределения. Речь идет о проверке гипотезы о числовом значении математического ожидания (m_0) нормального распределения при известном среднем квадратичном отклонении (σ). Составим гипотезы.

3. Нулевая гипотеза *H₀ :* $m_0 = 4{,}2$ (ммоль/л)

Альтернативная гипотеза *H₁ :* $m_0 \neq 4{,}2$ (ммоль/л)

4. Критическая область будет двусторонней. Ее образуют интервалы

(- $\infty < z^{\text{л}}_{\text{кр}}$)= $\frac{\alpha}{2}$ и ($z^{\text{п}}_{\text{кр}} > \infty$)= $\frac{\alpha}{2} = \frac{0{,}05}{2} = 0{,}025$. Это вероятность попадания случайной величины Z в левостороннюю или правостороннюю области. В это случае вероятность непопадания случайной величины Z в правостороннюю критическую область (1-$\frac{\alpha}{2}$) можно представить следующим образом: P(- $\infty < Z < z^{\text{п}}_{\text{кр}}$)= P(- $\infty < Z$ <0)+ P(0< $Z < z^{\text{п}}_{\text{кр}}$)

Так как P(- $\infty < Z$ <0) = 0,5, а P(0< $Z < z^{\text{п}}_{\text{кр}}$) = Φ($z^{\text{п}}_{\text{кр}}$) – функция Лапласа в точке $z^{\text{п}}_{\text{кр}}$. Тогда Φ($z^{\text{п}}_{\text{кр}}$) = 1- $\frac{\alpha}{2}$ -0,5= 0,475.

5.1 На основании таблицы значений функции Лапласа находим $z^{\text{п}}_{\text{кр}} = -z^{\text{л}}_{\text{кр}} =$ 1,96. Следовательно, критическая область состоит из интервалов (- ∞ < -1,96) и (1,96 < + ∞)

5.2 Рассчитаем Z_r:

$$Z_r = \frac{\bar{X} - m_0}{\sigma/\sqrt{n}} = \frac{175 - 180}{10/\sqrt{100}} = -8$$

6. Так как рассчитанное значение z-статистики (-8) попадает в критическую область, мы отвергаем нулевую гипотезу и принимается альтернативная гипотеза $m_0 \neq 4{,}2$ (ммоль/л).

Вывод: На уровне значимости 0.05 есть статистически значимые основания утверждать, что новый технологический процесс изменил средний уровень холестерина после лечения. Более того, так как z-статистика отрицательная, можно заключить, что новый технологический процесс, вероятно, привёл к *снижению* уровня холестерина.

В заключение, мы рассмотрели ключевые аспекты сравнения выборочной средней с математическим ожиданием. Мы выяснили, что эта процедура является основополагающей для проверки статистических гипотез и позволяет нам делать выводы о генеральной совокупности на основе данных выборки. Использование z-критерия или t-критерия Стьюдента зависит от того, известна генеральная дисперсия или нет.

Мы также обсудили практическое применение этой процедуры, показав, как сравнение выборочной средней с математическим ожиданием помогает принимать обоснованные решения на основе статистического анализа данных.

References

Чернова, Н.И., Математическая статистика : учеб. пособие / Н. И. Чернова, Новосиб. гос. ун-т — 2-е изд., испр. и доп. — Новосибирск : РИЦ НГУ, 2014 — 150 с. – Текст: непосредственный.— ISBN 978-5-4437-0304-6

Элементы математической статистики: учебное пособие по дисциплине «Теория вероятностей и математическая статистика» : / сост.: Н.И. Головко, Л.С. Ксендзенко, Г.С. Полещук, В.И. Рукавишникова, О.В. Бондрова, Ю.И. Коробецкая, Д.Б. Прокопьева, П.Н. Французова, Е.С. Фролова, Д.С. Шунскайте. – Владивосток : Издательство Дальневосточного федерального университета, 2021 – [101 с.]. – ISBN 987-5-7444-5225-4. – URL: https://www.dvfu.ru/science/publishing-activities/catalogue-of-books-fefu/. – Дата публикации: 28.12.2021. – Текст: электронные.

Григорьев Л.И., Подгорнов В.М., Фастовец Н.О. «Основы математической статистики в задачах нефтегазовой отрасли» – М.: ГАНГ, 1995. – Текст: электронный.