Урок 9. Тестирование гипотез (часть 1)

Хакимов Р.И. + ChatGPT

Введение в тестирование гипотез

Тестирование гипотез — это статистический метод, используемый для проверки предположений (гипотез) о параметрах генеральной совокупности на основе выборочных данных. Этот процесс помогает определить, насколько убедительными являются результаты исследования, и позволяет сделать выводы о популяции, исходя из данных выборки.

Основные концепции тестирования гипотез

Гипотеза: Предположение о параметре генеральной совокупности, которое проверяется на основе данных выборки.
Нулевая гипотеза (H_0): Основное предположение, которое мы тестируем. Обычно это гипотеза о том, что нет эффекта или различий. Например, "средний доход мужчин и женщин одинаков".
Альтернативная гипотеза (H_1 или H_a): Гипотеза, которая предполагает наличие эффекта или различий. Это гипотеза, которая рассматривается в случае, если нулевая гипотеза отвергнута. Например, "средний доход мужчин и женщин различен".

Основные концепции тестирования гипотез

Уровень значимости (α): Вероятность ошибки первого рода (α), то есть вероятность отклонения нулевой гипотезы, когда она на самом деле верна. Обычно выбирается значение 0.05 или 0.01.

P-значение: Вероятность получения результатов, которые столь же экстремальны, как и наблюдаемые, при условии, что нулевая гипотеза верна. Если P-значение меньше уровня значимости, нулевая гипотеза отвергается.

Основные концепции тестирования гипотез

Ошибка первого рода (α): Ошибка, при которой нулевая гипотеза отвергается, хотя она верна.
Ошибка первого рода важна для понимания, так как её последствия могут быть серьезными, и она учитывается при планировании экспериментов и интерпретации их результатов.

Ошибка второго рода (β): Ошибка, при которой нулевая гипотеза не отвергается, хотя альтернативная гипотеза верна.
Ошибка второго рода особенно критична, когда важно не пропустить реальные эффекты или различия, например, в медицине, правосудии или научных исследованиях.

Процесс тестирования гипотез

1. Формулировка гипотез:
- Нулевая гипотеза (H_0): Предположение о том, что никакого эффекта или различия нет.
- Альтернативная гипотеза (H_1 или H_a): Альтернативное предположение, которое предполагает наличие эффекта или различия.

2. Выбор уровня значимости (α):
- Обычно выбирается 0.05, что означает 5% вероятность ошибки первого рода.

3. Сбор и анализ данных:
- Собираются данные и рассчитываются статистики теста (например, выборочное среднее, выборочная дисперсия).

4. Расчет тестовой статистики:
- Вычисляется статистика теста (например, t-статистика, Z-статистика) в зависимости от типа теста.

Процесс тестирования гипотез

5. Определение P-значения:
- *P-значение* - это вероятность получения результатов, которые столь же экстремальны, как и наблюдаемые, при условии, что нулевая гипотеза верна. Оно сравнивается с уровнем значимости.
- P-значение вычисляется на основе тестовой статистики и распределения.

6. Принятие решения:
- Если P-значение меньше уровня значимости (α), нулевая гипотеза отвергается.
- Если P-значение больше уровня значимости, нет оснований для отклонения нулевой гипотезы.

7. Интерпретация результатов:
- Делается вывод о том, поддерживается ли нулевая гипотеза, или есть основания для её отклонения в пользу альтернативной гипотезы.

Типы тестов гипотез

Тесты для среднего значения:

- Тест z: Используется, когда размер выборки большой и дисперсия генеральной совокупности известна.
- Тест t: Используется, когда размер выборки малый и дисперсия генеральной совокупности неизвестна.

Тесты для пропорций:

- Тест для одной пропорции: Используется для проверки гипотез о пропорциях в одной выборке.
- Тест для разности пропорций: Используется для проверки гипотез о разнице пропорций между двумя выборками.

Тесты для дисперсий:

- Тест χ^2 (хи-квадрат): Используется для проверки гипотез о дисперсиях.

Тесты для независимости:

- Тест χ^2 на независимость: Проверяет зависимость между двумя категориальными переменными.

Одновыборочный и двухвыборочный t-тест

t-тест — это статистический метод, используемый для проверки гипотез о среднем значении генеральной совокупности или для сравнения средних значений двух выборок. t-тест особенно полезен, когда размер выборки небольшой и дисперсия генеральной совокупности неизвестна.
Выделяют два основных типа t-тестов: одновыборочный и двухвыборочный.

Процедура одновыборочного t-теста

Одновыборочный t-тест используется для проверки гипотезы о среднем значении генеральной совокупности на основе одной выборки.

1. Формулировка гипотез:

- Нулевая гипотеза (H_0): $\mu = \mu_0$, где μ — среднее значение генеральной совокупности, μ_0 — предполагаемое среднее значение по нулевой гипотезе.

- Альтернативная гипотеза (H_1): $\mu \neq \mu_0$ (двусторонний тест), или $\mu > \mu_0$ (односторонний тест), или $\mu < \mu_0$ (односторонний тест).

2. Расчет тестовой статистики:

- Формула для t-статистики:

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

где: - $\bar{X}$ — выборочное среднее,
- μ_0 — значение среднего по нулевой гипотезе,
- S — выборочное стандартное отклонение,
- n — размер выборки.

Процедура одновыборочного t-теста

3. Определение степени свободы: - Степени свободы (df) для одновыборочного t-теста: $n - 1$.

4. Сравнение с критическим значением t:

- Определите критическое значение t из таблицы распределения t на основе уровня значимости (α) и степеней свободы.
- Если вычисленное t-значение больше критического (или если P-значение меньше α), нулевая гипотеза отвергается.

Процедура одновыборочного t-теста

Пример. Предположим, что вы хотите проверить, отличается ли среднее значение роста людей в вашей выборке от 170 см. Вы собрали выборку из 25 человек, средний рост составил 172 см, а выборочное стандартное отклонение — 8 см. Уровень значимости — 0.05.

- Нулевая гипотеза: $\mu = 170$.
- Альтернативная гипотеза: $\mu \neq 170$.

Расчет t-статистики:

$$t = \frac{172 - 170}{8/\sqrt{25}} = \frac{2}{1.6} = 1.25$$

Если для 24 степеней свободы критическое значение t для двустороннего теста на уровне значимости 0.05 составляет 2.064, то поскольку $1.25 < 2.064$, мы не отвергаем нулевую гипотезу.

Процедура двухвыборочного t-теста

Двухвыборочный t-тест используется для сравнения средних значений двух независимых выборок, чтобы определить, есть ли статистически значимая разница между ними.

1. Формулировка гипотез:

Нулевая гипотеза (H_0): $\mu_1 = \mu_2$, где μ_1 и μ_2 — средние значения двух генеральных совокупностей.

Альтернативная гипотеза (H_1): $\mu_1 \neq \mu_2$ (двусторонний тест), или $\mu_1 > \mu_2$ (односторонний тест), или $\mu_1 < \mu_2$ (односторонний тест).

Процедура двухвыборочного t-теста

2. Расчет тестовой статистики:

- Формула для t-статистики:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

где:

- $\bar{X}_1$ и $\bar{X}_2$ — выборочные средние двух групп,
- S_p^2 — объединенная выборочная дисперсия,
- n_1 и n_2 — размеры выборок.
- Объединенная выборочная дисперсия рассчитывается как:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

где S_1^2 и S_2^2 — выборочные дисперсии двух групп.

Процедура двухвыборочного t-теста

3. Определение степени свободы:

- Степени свободы (df) для двухвыборочного t-теста: $n_1 + n_2 - 2$.

4. Сравнение с критическим значением t:

- Определите критическое значение t из таблицы распределения t на основе уровня значимости (α) и степеней свободы.
- Если вычисленное t-значение больше критического (или если P-значение меньше α), нулевая гипотеза отвергается.

Процедура двухвыборочного t-теста

Пример. Предположим, что вы хотите сравнить средний рост мужчин и женщин. Вы имеете две выборки: рост 20 мужчин (средний 175 см, стандартное отклонение 10 см) и рост 25 женщин (средний 165 см, стандартное отклонение 8 см). Уровень значимости — 0.05.

Нулевая гипотеза: $\mu_1 = \mu_2$.

Альтернативная гипотеза: $\mu_1 \neq \mu_2$.

Расчет объединенной дисперсии:

$$S_p^2 = \frac{(20-1)\cdot 100 + (25-1)\cdot 64}{20+25-2} = \frac{1900+1536}{43} = 67.0$$

Расчет t-статистики:

$$t = \frac{175-165}{\sqrt{67.0\left(\frac{1}{20}+\frac{1}{25}\right)}} = \frac{10}{\sqrt{67.0\cdot 0.089}} = \frac{10}{1.76} = 5.68$$

Если для 43 степеней свободы критическое значение t для двустороннего теста на уровне значимости 0.05 составляет 2.016, то поскольку $5.68 > 2.016$, нулевая гипотеза отвергается.

Одновыборочный и двухвыборочный t-тест

Заключение

t-тесты являются мощным инструментом для проверки гипотез о средних значениях и сравнении групп.
Одновыборочный t-тест помогает проверить, отличается ли среднее значение одной выборки от известного значения.
Двухвыборочный t-тест позволяет сравнить средние значения двух независимых выборок, чтобы выявить статистически значимые различия.

Z-тест для сравнения средних значений

z-тест для сравнения средних значений используется для проверки гипотез о разнице между средними значениями двух независимых выборок.
Этот тест особенно полезен, когда размер выборок достаточно большой, и дисперсии генеральных совокупностей известны или можно считать их приближенными к известным значениям.

Процедура z-теста для сравнения средних значений

1. Формулировка гипотез:
Нулевая гипотеза (H_0): $\mu_1 = \mu_2$. Средние значения двух генеральных совокупностей равны.
Альтернативная гипотеза (H_1): $\mu_1 \neq \mu_2$ (двусторонний тест). Средние значения двух генеральных совокупностей различны. Или $\mu_1 > \mu_2$ (односторонний тест), или $\mu_1 < \mu_2$ (односторонний тест).
2. Выбор уровня значимости (α):
- Уровень значимости обычно устанавливается на уровне 0.05, 0.01 или другом значении, в зависимости от исследовательских целей.
3. Сбор и анализ данных:
- Собираются данные из двух независимых выборок.
- Вычисляются выборочные средние $\bar{X}_1$ и $\bar{X}_2$, а также стандартные отклонения σ_1 и σ_2 для каждой из выборок.

Процедура z-теста для сравнения средних значений

4. Расчет тестовой статистики:

Формула для z-статистики:

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

где:
- $\bar{X}_1$ и $\bar{X}_2$ — выборочные средние двух групп,
- σ_1^2 и σ_2^2 — дисперсии генеральных совокупностей (или выборочные дисперсии, если дисперсии генеральных совокупностей неизвестны),
- n_1 и n_2 — размеры выборок.

Если нулевая гипотеза предполагает равенство средних ($\mu_1 - \mu_2 = 0$), формула упрощается до:

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Процедура z-теста для сравнения средних значений

5. Определение критического значения и P-значения:

- Определите критическое значение z из таблицы распределения нормальных случайных величин на основе уровня значимости (α).
- Рассчитайте P-значение, которое соответствует вычисленному z-значению.
- Если P-значение меньше уровня значимости (α), отвергайте нулевую гипотезу.

6. Принятие решения:

- Отклонение нулевой гипотезы: Если вычисленное z-значение больше критического значения (или если P-значение меньше α), нулевая гипотеза отвергается.
- Не отклонение нулевой гипотезы: Если вычисленное z-значение меньше критического значения (или если P-значение больше α), нет оснований для отклонения нулевой гипотезы.

Процедура z-теста для сравнения средних значений

Пример. Предположим, вы хотите сравнить средний вес мужчин и женщин в двух независимых группах. В выборке из 30 мужчин средний вес составляет 80 кг, стандартное отклонение — 10 кг. В выборке из 40 женщин средний вес составляет 70 кг, стандартное отклонение — 8 кг. Уровень значимости — 0.05.

- Нулевая гипотеза: $\mu_1 = \mu_2$.
- Альтернативная гипотеза: $\mu_1 \neq \mu_2$.

Расчет z-статистики:

$$z = \frac{80 - 70}{\sqrt{\frac{10^2}{30} + \frac{8^2}{40}}} = \frac{10}{\sqrt{3.33 + 1.60}} = \frac{10}{2.50} = 4.00$$

Для уровня значимости 0.05 и двустороннего теста критическое значение z приблизительно равно ± 1.96. Поскольку $4.00 > 1.96$, нулевая гипотеза отвергается, что означает, что средние веса мужчин и женщин статистически значимо различаются.

Z-тест для сравнения средних значений

Заключение

z-тест для сравнения средних значений полезен при анализе разницы между двумя независимыми группами, когда размеры выборок достаточно велики и дисперсии генеральных совокупностей известны.
Этот тест позволяет определить, есть ли статистически значимая разница между средними значениями двух групп.