

Лабораторная работа 1

Титов Вадим, 676 группа, titov.vn@phystech.edu

Задача 2.4

Одеяла с электрообогревом применяются в хирургии для восстановления температуры тела пациента после операции. Имеются два вида одеял: стандартный (b0) и экспериментальный (b1).

Для 14 пациентов известно время, за которое нормальная температура тела восстанавливается при использовании одеяла каждого из видов. Как понять, отличаются ли экспериментальные одеяла от стандартного?

Требуется:

1. Записать задачу формально в виде проверяемой гипотезы и альтернативы.
2. Предложить не менее 2-х критериев и соответствующих статистик для проверки этой гипотезы и описать:
 - при каких дополнительных условиях (если они есть) стоит применять тот или иной критерий
 - в чём преимущества/недостатки того или иного критерия
3. Аналитически выразить достигаемый уровень значимости каждого критерия на выборке или опишите, как его получить с помощью табличных данных.

Постановка задачи:

Выборка:

$$\begin{aligned}X_1 &= (X_1^1, \dots, X_1^{14}), \\X_2 &= (X_2^1, \dots, X_2^{14}) \\&\text{Выборки связанные}\end{aligned}$$

Гипотезы:

H_0 : Экспериментальное одеяло не помогает,

H_1 : Экспериментальное одеяло влияет на время восстановления

В данном случае мы ничего не можем сказать про тип распределения явно, поэтому в первую очередь хочется проверить непараметрическими критериями: критерий знаков, критерий знаковых рангов Уилкоксона и двухвыборочный перестановочный критерий (для связанных выборок).

1) Для критерия знаков особых предположений не нужно, нулевое распределение принадлежит классу $Bi(14, \frac{1}{2})$

Будем рассматривать двустороннюю альтернативу:

$$T(X_1, X_2) = \sum_{i=1}^{14} [X_{1i} > X_{2i}]$$

достигаемый уровень значимости: $p = P(\{T \geq t\} \cup \{T \leq t\})$, при $p \leq \alpha = 0.05$ отвергаем гипотезу о том, что экспериментальное одеяло не влияет

2) Данный критерий лучше использовать, если нет нулевой разности какой-то из пар.

$T(X_1, X_2) = \sum_i \text{rank}(|X_{1i} - X_{2i}|) \text{sign}(X_{1i} - X_{2i})$ Уровень значимости определяется таблично: определяется критическое значение для числа попарных разностей $n = 14$ и $\alpha = 0.05$.

3) Перестановочный критерий лучше использовать в случае, когда разность выборок распределена симметрично относительно выборочного среднего.

$$T(X_1, X_2) = \sum_i (X_{1i} - X_{2i})$$

Достигаемый уровень значимости (p) — доля перестановок знаков, на которых получилось такое же или ещё более экстремальное значение статистики.

Задача 4.2

Рассмотрим некоторую задачу классификации. Пусть задано качество 4 моделей a_1, a_2, a_3, a_4 . Качество полученных моделей показано в таблице.

Исследователю требуется выбрать наилучшую модель. Для выбора лучшей модели исследовать требуется попарно сравнить среднее значение качества всех моделей. Может ли исследователь утверждать что какая-то из моделей лучше другой?

Требуется:

- записать задачу формально;
- предложить статистику для решения данной задачи;
- записать нулевое распределение данной статистики;
- записать явно правило принятия решения на основе статистики и нулевого распределения для обеспечения уровня значимости $\alpha = 0.05$;
- проверить гипотезу по записанному критерию, для данных из условия. Противоречат ли они гипотезе?

Постановка задачи:

Выборка:

$$\begin{aligned}X_1 &= (X_1^1, \dots, X_1^6), \\X_2 &= (X_2^1, \dots, X_2^6), \\X_3 &= (X_3^1, \dots, X_3^6), \\X_4 &= (X_4^1, \dots, X_4^6), \\&\text{Выборки связанные}\end{aligned}$$

Гипотезы:

$$\begin{aligned}H_0 &: \text{Все модели классифицируют с одинаковым качеством,} \\H_1 &: \text{Какая-то модель лучше справляется с задачей}\end{aligned}$$

Так как ситуация, когда качество моделей растет по типу:

$quality(model_1) \leq quality(model_2) \leq quality(model_3) \leq quality(model_4) \leq quality(model_1)$, то целесообразнее делать сравнение попарные сравнения и выявлять различие в качестве попарно.

Решение о различии в качестве можно принимать основываясь на критерии знаков. В данном случае при данном уровне значимости можно сказать, что 4я модель хуже 2й и 3й, но в данном случае мы не учитываем количественную разность, из-за чего критерий теряет часть информации и не выбирает мощнее.

Поэтому воспользуемся критерием знаковых рангов Уилкоксона: критические значения определяем с помощью таблицы;

Можно утверждать с уровнем значимости $\alpha = 0.05$, что модели 2 и 3 классифицируют лучше чем 1 и 4.