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:

$$X_1 = (X_1^1, ..., X_1^{14}),$$
$$X_2 = (X_2^1, ..., X_2^{14})$$

Выборки связанные

Гипотезы:

$$H_0 : \textit{Экспериментальное одеяло не помогает},$$
$$H_1 : \textit{Экспериментальное одеяло влияет на время восстановления}$$

В данном случае мы ничего не можем сказать про тип распределения явно, поэтому впервую 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.

$T(X_1, X_2) = \sum_i rank(|X_{1i} - X_{2i}|) sign(X_{1i} - X_{2i})$ 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 a1, a2, a3, a4. 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:

$$X_1 = (X_1^1, ..., X_1^6),$$
$$X_2 = (X_2^1, ..., X_2^6),$$
$$X_3 = (X_3^1, ..., X_3^6),$$
$$X_4 = (X_4^1, ..., X_4^6),$$
Выборки связанные

Гипотезы:

$$H_0 : \textit{Все модели классифицируют с одинаковым качеством},$$
$$H_1 : \textit{Какая-то модель лучше справляется с задачей}$$

Так как ситуация, когда качество моделей растет по типу:

$\text{quality}(\text{model}_1) \leq \text{quality}(\text{model}_2) \leq \text{quality}(\text{model}_3) \leq \text{quality}(\text{model}_4) \leq \text{quality}(\text{model}_1)$, то целесообразнее делать сравнение попарные сравнения и выявлять различие в качестве попарно.

Решение о различии в качестве можно принимать основываясь на критерии знаков (Нулевое распределение лежит в классе $Bi(6, \frac{1}{2})$ для каждой пары моделей), решение о выборе лучше модели принимается в случае принятия статистикой значения 6, $\mathbf{P}\{T = 6\} = \frac{1}{32} \approx 0.03$.

При данном уровне значимосте можно сказать, что 4я модель хуже 2й и 3й, но данный критерий не учитывает сами величины, из-за чего теряется дополнительная информация о качестве. С данной проблемой лучше справляется критерий знаковых рангов Уилкоксона,

Поэтому воспользуемся критерием знаковых рангов Уилкоксона: нулевое распределение определяется таблично... После проверки оказалось, что данный критерий отделяет так же не смог отличить модели. Параметрические критерии лучше не пытаться применять, потому что слишком мало данных для точечных оценок.

Таким образом, исследователь может утверждать, что модель 4 справляется с классификацей хуже 2й и 3й. Так как было бы странно утверждать, что модель справляется хуже других равных в разной степени, положим, что модель 4 справляется и хуже 1й.