

Практика 7.

Задания по темам:

Критерии сравнения зависимых выборок: парный t-критерий Стьюдента и критерий Вилкоксона.

Критерии сравнения независимых выборок:

t-критерий Стьюдента,

U-критерий Манна–Уитни, H-критерий

Краскала–Уоллиса.

Обработка данных на компьютере (MS Excel).

Блок 1. Сравнение зависимых (парных) выборок

Задание 1 Постановка задачи для парного t-критерия

В группе из 8 человек измерили давление до и после приёма препарата. Данные (систолическое давление):

До	140	145	138	150	142	139	148	144
После	135	140	136	142	138	137	143	140

- (a) Сформулируйте нулевую гипотезу для проверки эффективности препарата (снижение давления).
- (b) Вычислите разности $d_i = \text{После} - \text{До}$. Найдите среднее разностей $\bar{d}$ и стандартное отклонение s_d .
- (c) Вычислите парный t-критерий: $t = \frac{\bar{d}}{s_d/\sqrt{n}}$. Сравните с критическим значением при $\alpha = 0.05$ (двусторонний или односторонний?).

Задание 2 Выбор между парным t и критерием Вилкоксона

В задании 1 предположим, что распределение разностей не является нормальным (гистограмма показывает сильную асимметрию).

- (a) Какой критерий следует применить вместо парного t-критерия? Почему?
- (b) Какое предположение (менее строгое) требуется для этого непараметрического критерия?

Задание 3 Критерий знаковых рангов Вилкоксона (Т-критерий)

Даны пары наблюдений (оценки удовлетворённости до и после тренинга, баллы 1–10):

До	5	6	4	7	5	6	5	8
После	6	7	5	7	6	8	7	9

- Вычислите разности и их знаки. Отбросьте нулевые разности (если есть).
- Проранжируйте абсолютные величины разностей (от меньшей к большей).
- Отдельно сложите ранги для положительных разностей (R^+) и для отрицательных (R^-). Статистика критерия $T = \min(R^+, R^-)$.
- По таблице критических значений критерия Вилкоксона для $n = 8$ и $\alpha = 0.05$ определите, отвергается ли гипотеза об отсутствии эффекта.

Задание 4 Сравнение мощности на примере

Почему при нормально распределённых разностях парный t-критерий предпочтительнее критерия Вилкоксона? А в каких ситуациях Вилкоксон оказывается более мощным (или единственно возможным)?

Блок 2. Сравнение независимых выборок (две группы)

Задание 5 t-критерий Стьюдента для независимых выборок

Две группы студентов: контрольная (традиционное обучение) и экспериментальная (новая методика). Баллы за экзамен:

Контрольная: 72, 68, 75, 70, 74, 69, 73, 71

Экспериментальная: 78, 82, 80, 79, 81, 77, 83, 80

- Найдите средние $\bar{x}_1$, $\bar{x}_2$ и выборочные дисперсии s_1^2 , s_2^2 .
- Вычислите t-статистику для независимых выборок (предполагая равенство дисперсий): $t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$, где $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$.
- Определите число степеней свободы $df = n_1 + n_2 - 2$. Сравните с критическим значением $t_{0.05, df}$. Сделайте вывод.

Задание 6 U-критерий Манна–Уитни

Та же задача, но данные не удовлетворяют предположению о нормальности.

- Объедините обе выборки и проранжируйте все значения (от меньшего к большему).
- Найдите сумму рангов для каждой группы (R_1 , R_2).
- Вычислите $U_1 = R_1 - \frac{n_1(n_1 + 1)}{2}$ и $U_2 = R_2 - \frac{n_2(n_2 + 1)}{2}$; статистика $U = \min(U_1, U_2)$.
- Для $n_1 = n_2 = 8$ критические значения U для $\alpha = 0.05$ равны, например, 13. Сравните. Какой вывод о различии групп?

Задание 7 Сравнение t- и U-критериев на одних данных

В задании 5 и 6 выполните оба теста (даже если условия нарушены).

- Совпадают ли выводы о значимости?

- (b) Что изменится, если в экспериментальной группе появится выброс (например, один студент набрал 50 баллов)? Пересчитайте (хотя бы идея) t - и U -критерии. Какой критерий более устойчив?

Блок 3. Сравнение трёх и более независимых выборок (Н-критерий Краскала–Уоллиса)

Задание 8 Переход от U к H

H -критерий Краскала–Уоллиса является обобщением U -критерия Манна–Уитни на $k \geq 3$ групп. Сформулируйте нулевую гипотезу для H -теста. Что проверяется?

Задание 9 Расчёт H -критерия вручную

Три метода лечения: результаты (баллы улучшения): Метод А: 5, 7, 6, 4
Метод В: 8, 9, 7, 10
Метод С: 3, 5, 4, 4

- (a) Объедините все выборки и проранжируйте значения.
(b) Найдите сумму рангов для каждой группы (R_A, R_B, R_C); $n_A = n_B = n_C = 4$.
(c) Вычислите $H = \frac{12}{N(N+1)} \sum \frac{R_i^2}{n_i} - 3(N+1)$, где $N = 12$.
(d) Сравните полученное H с критическим значением $\chi_{0.05, df=2}^2$ (степеней свободы $k - 1 = 2$). Сделайте вывод.

Задание 10 Пост-хок анализ

Если H -тест отвергает гипотезу о равенстве всех групп, что делают дальше? Назовите один из методов попарных сравнений с коррекцией на множественные тесты (например, Данна или Манна–Уитни с поправкой Бонферрони).

Блок 4. Обработка данных на компьютере (MS Excel)

Задание 11 Парный t -тест в Excel

Опишите (пошагово) как в MS Excel провести парный t -критерий для данных из задания 1, используя «Пакет анализа» → « t -тест: парный двухвыборочный для средних». Что нужно указать в диалоговом окне? Как интерпретировать вывод (t -статистика, p -значение)?

Задание 12 U -тест Манна–Уитни в Excel

В стандартном «Пакете анализа» Excel нет U -теста. Предложите способ вычислить U -критерий с помощью средств Excel (ранжирование с использованием функций РАНГ.СР, СУММЕСЛИ, и т.д.). Напишите последовательность действий (без макросов).

Задание 13 Расчёт H -критерия в Excel

Для данных задания 9:

- (a) Покажите формулы Excel для вычисления рангов объединённой выборки (используйте функцию РАНГ.СР).

- (b) Как получить суммы рангов по группам (например, через СУММЕСЛИ)?
- (c) Вычислите Н по формуле. Какую функцию Excel использовать для получения р-значения из распределения χ^2 ?

Задание 14 Работа с надстройками

Какие надстройки или статистические пакеты для Excel позволяют выполнять непараметрические тесты (Манна–Уитни, Вилкоксона, Краскала–Уоллиса) без ручных вычислений? Назовите хотя бы один (например, «Real Statistics» или «XLSTAT»). Опишите, как установить и запустить хотя бы один тест.

Задание 15 Сравнение результатов: ручной расчёт и Excel

Возьмите любые данные из предыдущих заданий (например, из задания 5). Проведите t-тест в Excel с помощью «Пакета анализа» и сравните полученное р-значение с табличным результатом, который вы получили вручную. Должно ли оно совпадать? Почему возможно небольшое расхождение?

Комплексное задание

Задание 16 Полный цикл анализа данных в Excel

Исследователь провёл эксперимент: три группы (А, В, С) по 10 человек в каждой. Измерялся показатель концентрации внимания (баллы). Данные в условной электронной таблице.

- (a) Сформулируйте задачу: нужно ли применять параметрический или непараметрический критерий? Какие условия необходимо проверить (нормальность, однородность дисперсий)?
- (b) В Excel постройте гистограммы для каждой группы и выполните тест на нормальность (например, с помощью критерия Шапиро–Уилка через надстройки). Если распределения ненормальные, какой критерий следует использовать для сравнения трёх групп?
- (c) Проведите Н-критерий Краскала–Уоллиса в Excel (ручной расчёт или через надстройку).
- (d) Если Н-тест показал значимые различия, выполните попарные сравнения (U-тест Манна–Уитни) с поправкой Бонферрони. Как в Excel скорректировать уровень значимости?
- (e) Оформите итоговый отчёт: таблица средних, медиан, результатов тестов, вывод.

Задание 17 Анализ реальных (тренировочных) данных

Ниже приведены результаты теста для двух независимых групп (10 человек в каждой). Попробуйте вручную применить t-критерий, затем U-тест, а затем сделайте то же самое в Excel (хотя бы мысленно следуя инструкциям).

Группа 1: 23, 25, 27, 24, 26, 22, 28, 25, 23, 27

Группа 2: 30, 32, 29, 31, 33, 28, 30, 31, 32, 29

- (a) Проверьте предположение о нормальности (визуально – гистограмма или Q-Q график).
- (b) Выберите подходящий критерий. Обоснуйте.
- (c) Выполните вычисления (можно с помощью Excel, но опишите формулы).
- (d) Сделайте вывод о значимости различий ($\alpha = 0.05$).
- (e) Какой из критериев дал бы более строгий (меньший) p-value? Почему?