Лекция №7 Непараметрические тесты

Цель лекции:

- ✓ Научиться тестировать гипотезы, когда не выполняются условия применимости параметрических тестов
- ✓ Протестировать гипотезу с помощью критерия Манна – Уитни
- ✓ Провести тест **Уилкоксона**
- ✓ Применить тест Крускала –Уоллиса для проверки гипотезы
- ✓ Выполнить тест Фридмана

Материал прошлого урока:

На прошлых 2х занятиях мы рассмотрели критерии Z и t, которые являются параметрическими.Но использование данных критериев предполагает соблюдение следующих условий

1. выборки взяты из нормально распределенной генеральной совокупности
2. дисперсии сравниваемых выборок равны.

Если мы столкнулись с ситуацией, когда какое-то из выше перечисленных условий грубо нарушается, при этом мы работаем с количественными или порядковыми данными, тогда следует использовать непараметрические критерии. Именно изучением этих тестов мы займемся на этом уроке.

План урока:

Критерий Манна-Уитни

Критерий Уилкоксона

Критерий Крускала –Уоллиса

Критерий Фридмана

Критерий Манна – Уитни U

Критерий Манна – Уитни основан на рангах. Данный критерий считается аналогом критерия Стьюдента t. Его следует использовать для сравнения двух независимых выборок, если не соблюдается условие нормальности, или, если выборки очень маленького объема и мы не можем проверить соблюдение данного условия, и в то же время у нас нет уверенности, что выборки взяты именно из нормально распределенной генеральной совокупности. И также этот критерий применяется, если дисперсии в сравниваемых выборках различны. Т.о. критерий Манна-Уитни может применяться при работе с количественными и порядковыми данными.

Для тестирования гипотезы о различиях между двумя независимыми выборками,

Т.е. мы можем проверять наличие различий между двумя независимыми выборками с использованием параметрических тестов (при соблюдении определенных условий) и непараметрических тестов (при грубом нарушении хотя бы одного условия). Когда мы использовали параметрические тесты, мы формулировали нулевую гипотезу, как равенство двух средних арифметических $\mu_1 = \mu_2$, а альтернативную гипотезу, как неравенство двух средних арифметических ($\mu_1 \neq \mu_2$, $\mu_1 > \mu_2$, $\mu_1 < \mu_2$). С непараметрическими тестами мы вместо средних арифметических сравниваем медианы. Мы помним с третьего урока, что медиана также является мерой центральной тенденции.

Рассмотрим, как работает критерий Манна-Уитни. При его расчете значения случайной величины (СВ) заменяются рангами, т.е. порядковыми значениями, где 1й ранг соответствует самому маленькому значению, а последний ранг соответствует самому большому значению., т.е. мы присваиваем порядковые номера. Поступая таким образом, мы сохраняем максимально возможное количество информации о изучаемой СВ. Но если с параметрическими критериями нас интересовал тип распределения, то с непараметрическим критерием Манна-Уитни мы не заботимся об этом вопросе. Давайте рассмотрим конкретную задачу.

Предположим, у нас есть две выборки, которые мы хотим сравнить, чтобы понять, а есть ли между ними статистически значимые различия. Эти данные являются количественными, но нарушено какое-то из условий применимости t критерия, поэтому мы будем использовать критерий Манна-Уитни. Чтобы не утомлять Вас расчетами, но дать представление, по какому принципу работает непараметрический тест, берем совсем маленькие объемы выборок. Еще есть одно условие, которое должно соблюдаться при использовании данного критерия – это объемы выборок должны быть не меньше трех.

Выборка 1: 47, 75, 90

Выборка2: 58, 60, 77

Сформулируем нулевую и альтернативную гипотезы, где μ_1 и μ_2 – медианы 1-й и 2-й выборок.

$H_0: \mu_1 = \mu_2$

$$H_1: \mu_1 \neq \mu_2$$

Теперь создадим таблицу, куда занесем значения, содержащиеся в первой и второй выборке и присвоенные им ранги.

Для этого объединим все данные из обеих выборок в один ряд и отсортируем их по возрастанию.

47, 58, 60, 75, 77, 90

Теперь присвоим индексы-ранги.

1 , 2, 3, 4, 5, 6

Именно эти ранги занесем в таблицу.

И после этого останется посчитать суммы рангов для каждой выборки.

1 выборка		2 выборка	
значения	ранг	значения	ранг
75	4	60	3
90	6	58	2
47	1	77	5
	$\sum = 11$		$\sum = 10$

У нас получились суммы рангов для выборок 11 и 10, т.о. нам надо понять, а является ли различие между 10 и 11 статистически значимым. Чтобы ответить на этот вопрос, нам нужно перебрать все возможные комбинации из шести рангов.

Достаточно перебрать комбинации только для первой выборки. Их количество мы найдем, как сочетания из 6 по 3.

$$C_6^3 = \frac{6!}{3! * (6-3)!} = 20$$

И в таблице ниже приведены все возможные 20 сочетаний из 3х рангов, которые мы могли бы встретить в 1-й выборке.

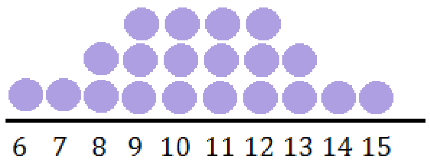
№	комбинация	Сумма чисел комбинации
1	123	6
2	124	7
3	125	8
4	126	9
5	134	8
6	135	9
7	136	10
8	145	10
9	146	11
10	156	12

11	234	9
12	235	10
13	236	11
14	245	11
15	246	12
16	256	13
17	245	12
18	346	13
19	356	14
20	456	15

Теперь построим сводную таблицу, где посчитаем, сколько раз встретилась та или иная сумма рангов.

сумма	6	7	8	9	10	11	12	13	14	15
частота	1	1	2	3	3	3	3	2	1	1

Теперь построим график частот ниже.



Теперь можно продолжить. Мы не обсудили уровень статистической значимости α. Но как и в любом тестировании гипотезы, нам заранее нужно определиться с его значением. Возьмем α = 0.1. Теперь еще раз вспомним, что вероятность – это доля. Вероятность получить сумму рангов $6 \text{ и } 15 = \frac{1}{20} = 0.05$ (так как встречается 1 раз из 20). 6 и 15 – и будут критическими значениями, т.е. 0.05+0.05 = 0.1. Ведь у нас двусторонний тест. Если сумма рангов в меньшей выборке попадает в интервал между критическими значениями, причем сами критические значения не включаются в этот интервал, то верна нулевая гипотеза. Т.к. у нас одинаковые объемы выборок, то мы можем брать любую выборку. Мы брали 1ю с суммой рангов 11. Мы видим, что 11 попадает в интервал (6; 15), следовательно, верна нулевая гипотеза, о том, что статистически значимых различий между двумя выборками не обнаружены. Но если бы сумма рангов получилась 6 или 15, то нулевую гипотезу следовало отклонить, что означало бы, что мы нашли статистически значимые различия между выборками. В нашем случае $\frac{1}{20} = 0.05$, но обычно эта дробь отличается от заданных точных значений α 0.05 или 0.01. Тогда за критическое значение выбирается наиболее близкое к заданному уровню статистической значимости. Т.е. в отличие от распределения Стьюдента, которое непрерывно, полученное распределение для суммы рангов является дискретным.

Такие таблицы составлять очень сложно, потому что, объемы выборок обычно намного больше. Давайте произведем тест Манна- Уитни в Python с помощью функции $mannwhitheyu()$.

```
import numpy as np
import scipy.stats as stats

x1= np.array([47, 90, 75])

x2 = np.array([ 58, 60, 77])

stats.mannwhitneyu(x1, x2)
MannwhitneyuResult(statistic=5.0, pvalue=1.0)
```

И чтобы интерпретировать результат, как и в любом тесте, смотрим на p-value

Рассмотрим пример из книги Стентона Гланца, где даны значения для двух групп

```
Рассмотрим пример из книги Стентона Гланца

group_1= np.array ([1000, 1380, 1200])

group_2 = np.array ([1400, 1600, 1180, 1220])

stats.mannwhitneyu(group_1, group_2)
MannwhitneyuResult(statistic=3.0, pvalue=0.4)
```

Ранги для 1й группы: 1, 5, 3. Сумма рангов 9

Ранги для 2й группы: 6, 7, 2, 4. Сумма рангов 19

p-value = 0.4, что дает нам право сделать выбор в пользу нулевой гипотезы, т.к. даже при α=0.1 p-value > α.

Выше мы обсудили, что критерий Манна-Уитни U проводится для независимых выборок, но что делать, если у нас есть 2 зависимые выборки, например, измерения какого-то параметра у одних и тех же пациентов до и после принятия лекарства. В этих случаях используется тест Уилкоксона, при условии, что объемы выборок небольшие. При больших объемах выборок, свыше 50, используется критерий знаков, который мы рассматривать не будем, потому что все критерии мы в одном курсе рассмотреть не сможем, но зная условия применимости, вы сможете безошибочно выбрать нужный тест.

Критерий Уилкоксона

Этот ранговый непараметрический критерий является также аналогом критерия Стьюдента. Но используется для парных измерений, т.е. для зависимых выборок. Давайте его также сразу рассмотрим на конкретном примере.

Предположим, исследуется влияние некоторой диеты на вес пациентов. В исследовании участвуют 10 пациентов. Массив x1 будет представлять их веса до начала диеты, а массив x2 – веса тех же пациентов, но после окончания диеты. Т.е. до и после мы имеем по 10 весов пациентов. Мы находим разницу между их весами x2 – x1. У нас есть положительные и отрицательные значения. Мы видим, что подавляющее большинство разностей имеют отрицательное значение и только 2 из них положительное. Чтобы найти расчетный критерий Уилкоксона, нам необходимо сложить ранги этих несвойственных значений для массива $x2 - x1$, обозначим эти значения Δ. Давайте произведем эти расчеты.

```
import numpy as np
import scipy.stats as stats

x1= np.array ([70,74, 74.5, 79, 85, 93, 94, 98, 106.5, 107])

x2= np.array ([64, 76.5, 67, 73.5, 89, 85, 89.5, 91, 98, 100.5])

x2 - x1

array([-6. ,  2.5, -7.5, -5.5,  4. , -8. , -4.5, -7. , -8.5, -6.5])
```

Выше мы нашли десять значений Δ (разница между измерениями до и после диеты), теперь надо присвоить ранги. Обратите внимание, что здесь мы ранги присваиваем по абсолютной величине.

Значение Δ	-6	2.5	-7.5	-5.5	4	-8	-4.5	-7	-8.5	-6.5
ранг	5	1	8	4	2	9	3	7	10	6

Складываем ранги 1+2 =3 => W =3. Проверим наши расчеты функцией из Python.

```
stats.wilcoxon(x1, x2)
WilcoxonResult(statistic=3.0, pvalue=0.009765625)
```

Статистик также получился равен 3. Смотрим на p-value, которое приблизительно равно 0.01. Предположим, что мы выбирали уровень статистической значимости α= 0.05. Следовательно, принимаем альтернативную гипотезу, о том, что статистически значимые различия между группами обнаружены.

Давайте теперь в массиве x2 последнее значение поменяем на 113.5, т.е. у этого пациента после диеты вес не уменьшился на 6.5 кг, а наоборот увеличился. Теперь это будет массив x3. Поскольку ранги назначаются по абсолютным значениям, то значения ранга не меняется. Но теперь эту ячейку в таблице ниже мы тоже выделим, т.к. количество рангов со знаком минус все еще в меньшинстве

Значение Δ	-6	2.5	-7.5	-5.5	4	-8	-4.5	-7	-8.5	6.5
ранг	5	1	8	4	2	9	3	7	10	6

Теперь осталось сложить значения выделенных в таблице рангов и получить расчетный критерий.

W = 1+2+6 = 9

Проверим вычисления в Python

```
А теперь поменяем последнее значение в массиве x2  на 113.5

x1= np.array ([70,74, 74.5, 79, 85, 93, 94, 98, 106.5, 107])

x3= np.array ([64, 76.5, 67, 73.5, 89, 85, 89.5, 91, 98, 113.5])

stats.wilcoxon(x1, x3)
WilcoxonResult(statistic=9.0, pvalue=0.064453125)
```

Теперь pvalue $\approx$ 0.065, следовательно, принимаем нулевую гипотезу о том, что различий между группами не найдено.

В настройках функции по умолчанию выполняется двусторонний тест (two-sided). Также можно выбрать опции меньше «less» и больше «grater».

При двустороннем тесте проверяется, что распределение для величины Δ не будет симметрично относительно нуля. Для одностороннего теста «grater» проверяется, будет ли распределение величины Δ лежать выше распределения, симметричного нуля , ну а для теста «less», соответственно ниже, чем распределение, симметрично относительно нуля.

Критерий Крускала –Уоллиса (Краскела – Уоллиса)

Для сравнения нескольких групп на последнем уроке мы будем использовать дисперсионный анализ, в основе которого лежит параметрический критерий Фишера. Но параметрические тесты, как мы уже знаем, требуют соблюдения условия нормальности и равенства дисперсий. В противном случае, если есть грубые нарушения, данных условий, мы прибегаем к непараметрическим тестам. И поскольку сегодняшний урок посвящен непараметрическим тестам, логично будет продолжить их изучение.

Значит, теперь мы будем работать с множественными сравнениями, когда нам требуется сравнить несколько групп. Опять же мы можем работать с несколькими независимыми выборками, тогда нужно будет воспользоваться критерием Крускала – Уоллиса, а можем столкнуться с повторными измерениями, когда какой-то параметр измерялся у одних и тех же пациентов несколько раз. И в этом случае понадобится критерий Фридмана. Рассмотрим по порядку оба этих критерия.

Итак, начнем с критерия Крускала- Уоллиса H . Это аналог дисперсионно анализа, который нам еще предстоит изучить. Данный тест проверяет равенство медиан трех и более выборок.

Сначала рассмотрим, идею, которая легла в основу вычисление критерия H.

Мы сейчас будем работать с задачей, где будут приведены заработные платы людей из трех разных профессий. И с подобной задачей мы будем работать, изучая дисперсионный анализ. Только сейчас мы будем исходить из предположения, что распределения в подгруппах отлично от нормального, поэтому нам понадобится тест Крускала - Уоллиса.

Как и любое тестирование гипотезы, нам надо сравнить табличное значение с расчетным значением. Мы сделаем следующее, посчитаем расчетное значение. Затем воспользуемся функцией в Python и убедимся, что наши расчеты верны и сделаем вывод, опираясь на p-value, рассчитанное функцией. Обратите внимания, что в этой лекции мы не ищем табличные значения для критериев, чтобы потом с ним сравнить расчетное значение. Поскольку сегодня и так много информации о четырех разных критериев, не будем путать вас с таблицами, тем более в современных реалиях пользуются готовыми функциями. А функции предоставляют нам pvalue, по которым и интерпретируют результат. Единственное, что мы делаем, мы смотрим, как функция получает то или иное значение статистика.

Чтобы рассчитать критерий Крускала-Уоллиса H делаем следующее:

1) Обобщим все данные в один ряд
2) Присвоим ранги в этом ряду
3) Посчитаем сумму рангов, присвоенных в общем ряду, но теперь уже в отдельных группах. Т.е. получим сумму рангов для каждой отдельной группы.
4) Воспользуемся формулой:

$$H = \frac{12}{N * (N+1)} * \sum_{i=1}^{k_j} \frac{{T_j}^2}{n_j} - 3(N+1),$$

где N – общее число измерений во всех сравниваемых выборках,
k_j- объем j-ой выборки
T_j- сумма рангов в каждой выборке.

Задача: Даны заработные платы людей, принадлежащих к трем разным профессиям (условия нормальности не соблюдается).

gr_1**:** 70, 50, 64, 61, 75, 67, 73
gr_2**:** 80, 78, 90, 68, 74, 65, 85
gr_3**:** 141, 142, 140, 152, 161, 163, 155

Требуется определить, влияет ли профессия на заработную плату.

Еще раз акцентирую внимание, мы рассмотрим на другом уроке, который будет посвящен дисперсионному анализу, подобную задачу, но уже с другим условиями применимости. В контрасте хорошо запоминаются условия применимости того или иного критерия.

Не отвлекаясь на новые условия, в памяти отложится два разных теста, которые казалось бы будут применены к почти одинаковым данным. И, скорее всего вы вспомните, что где-то нарушались условия применимости того или иного теста. И следовательно, вряд ли вы ошибетесь с выбором критерия.

Объединим все значения групп в один ряд по возрастанию и присвоим им ранги. Если значения повторяются, то тогда ранг находится, как среднее арифметическое мест повторяющихся значений. При этом номера мест должны быть присвоены в отсортированном ряду. Например, было бы два значения 65 с индексами 4 и 5 (т.е. на 4-м и 5-м местах) в отсортированном по возрастанию ряду. Тогда обоим значениям будет присвоен ранг $(4+5)/2\ = 4.5$ Вернемся к задаче.

Поместим в таблицу все значения зарплат по возрастанию и присвоим ранги:

50	61	64	65	67	68	70	73	74	75	78	80	85	90	140	141	142	152	155	161	163
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21

Теперь надо найти суммы рангов T_j для каждой группы. Составим таблицу:

Группа 1		Группа 2		Группа 3	
Зар.плата	Ранг	Зар.плата	Ранг	Зар.плата	Ранг
70	7	80	12	141	16
50	1	78	11	142	17
64	3	90	14	140	15
61	2	68	6	152	18
75	10	74	9	161	20
67	5	65	4	163	21
73	8	85	13	155	19

Сумма рангов T_1:	36	Сумма рангов T_2:	69	Сумма рангов T_3:	126

Теперь воспользуемся этой формулой. Общее число измерений N = 21

$$H = \frac{12}{N * (N+1)} * \sum_{i=1}^{k_j} \frac{{T_j}^2}{n_j} - 3 * (N+1),$$

$$H = \frac{12}{21 * (21+1)} * \left(\frac{36^2}{7} + \frac{69^2}{7} + \frac{126^2}{7}\right) - 3 * (21+1) = 15.38404$$

Давайте посмотрим, какое расчетное значение даст нам функция $kruskal()$

```
import numpy as np
import scipy.stats as stats

gr_1= np.array ([70,  50,  64,  61,  75,  67,  73])
gr_2=np.array([80,  78,  90,  68,  74,  65,  85])
gr_3 = np.array([141, 142, 140, 152, 161, 163, 155])

stats.kruskal(gr_1, gr_2, gr_3)

KruskalResult(statistic=15.384044526901675, pvalue=0.00045645416718036815)
```

Мы видим, что функция показала точно такое же значения, как и расчеты, сделанные вручную. p-value ≈ 0.0005. Это позволяет нам сделать вывод в пользу альтернативной гипотезы о том, что медианы в выборках различны. Как и в других непараметрических тестах, мы в нулевую гипотезу вкладываем предположение о равенстве медиан, а в альтернативную – предположение о том, что они не равны или для односторонних тестов (например, критерий Уилкоксона) можем делать предположение больше или меньше. Когда у нас несколько выборок и мы имеем дело со множественными сравнениями, мы использовали критерий Крускала- Уоллиса. Но, как мы теперь уже знаем, он применяется, если нарушаются условия для параметрических критериев (нормальность и равенство дисперсий) и группы независимы. Если, например, одни и те же пациенты для поддержания нормального уровня гемоглобина в крови сидят сначала на одной диете, а потом на другой. Т.е. у нас есть измерения до начала диет – это выборка 1, измерения после диеты А - это вторая выборка и значения после диеты В – это третья выборка. Это, так называемый анализ повторных измерений. Здесь понадобится критерий Фридмана.

Критерий Фридмана

Когда у нас было несколько независимых выборок, мы использовали критерий Крускала- Уоллиса. В случае повторных измерений и несоблюдения условий нормальности и равенства дисперсий в исследуемых выборках, применяем критерий Фридмана.

Порядок расчета для этого критерия следующий (будем сразу рассматривать на примере выше приведенной задачи с измерением гемоглобина):

1. Сначала назначаются ранги по каждому пациенту. Т.е. если у нас три раза брались измерения у одних и тех же пациентов, то для каждого пациента будут измерения с рангами от 1 до 3. Если бы 4 измерения у одного и того же пациента, тогда от 1 до 4.
2. Затем находи сумму рангов по выборкам. Не по пациентам, обратите внимание, а по выборкам. Чуть ниже будет пример.
3. Теперь нужно найти средний ранг $\underline{R}$

$$\underline{R} = \frac{n*(k+1)}{2}, \text{где } n - \text{объем выборки}, k - \text{число сравниваемых групп}$$

4. И последним действием производим расчет критерия Фридмана по формуле:

$\chi_r^2 = \frac{12}{n*k*(k+1)} * \sum \quad (R_i - \underline{R})^2$, где R_i – сумма рангов по подгруппам

Давайте продолжим рассматривать задачу. У нас есть 5 пациентов и для каждого из них три измерения гемоглобина. Для первого пациента это значения 123, 126 и 141, следовательно, для 126 –наименьший ранг 1, а для 141 –наибольший ранг 3.

Пациент	До диеты		Диета А		Диета В	
	значение	ранг	значение	ранг	значение	ранг
1	123	1	126	2	141	3
2	135	1	144	2	150	3
3	119	2	117	1	164	3
4	109	1	156	3	147	2
5	145	1	170	3	169	2
		$\sum$ =6		$\sum$ =11		$\sum$ =13

Найдем средний ранг $\underline{R} = \frac{n*(k+1)}{2} = \frac{5*(3+1)}{2} = 10$

$$\chi_r^2 = \frac{12}{n*k*(k+1)} * \sum \quad (R_i - \underline{R})^2 =$$

$$= \frac{12}{5*3*4} * [(6-10)^2 + (11-10)^2 + (13-10)^2] = \frac{1}{5} * (16+1+9) = 5.2$$

Произведем расчеты с помощью функции friedmanchisquare()

Получаем значение статистика 5.2 и p-value 0.07, что позволяет нам принять нулевую гипотезу об отсутствии различий между выборками.

```
import numpy as np

before= np.array([123,135,119,109, 145])

diet_1=np.array([ 126, 144, 117, 156, 170])

diet_2= np.array([ 141, 150, 164, 147, 169])

stats. friedmanchisquare(before, diet_1, diet_2)

FriedmanchisquareResult(statistic=5.200000000000003,
pvalue=0.07427357821433388)
```

Давайте подведем итог, мы сегодня изучили 4 непараметрических теста.

Критерий Манна-Уитни

Критерий Уилкоксона

Критерий Крускала –Уоллиса

Критерий Фридмана

Сравнение 2-х групп		Множественные сравнения	
Независимые выборки	Зависимые выборки	Независимые выборки	Анализ повторных измерений
Критерий Манна-Уитни	Критерий Уилкоксона	Крускала- Уоллиса	Критерий Фридмана

На этом мы закачиваем с тестированием гипотез и на следующем уроке переходим к изучению корреляционного анализа.