

Теория вероятностей и математическая статистика

Задание 1.

Даны две независимые выборки. Не соблюдается условие нормальности

x1 380,420,290

y1 140,360,200,900

Сделайте вывод по результатам, полученным с помощью функции

Подсказка № 1

Тест Манна-Уитни — это непараметрический тест, который используется для сравнения двух независимых выборок. Он применяется, когда не соблюдается условие нормальности данных.

Подсказка № 2

Перед выполнением теста нужно определить уровень значимости (**alpha**). Обычно используется значение 0.05, что соответствует 95% уровню доверия. Если р-значение теста меньше выбранного уровня значимости, это означает, что существует статистически значимое различие между выборками.

Подсказка № 3

После выполнения теста Манна-Уитни обратите внимание на значения U-статистики и р-значения. U-статистика показывает степень различия между выборками. Однако для интерпретации важнее р-значение: если оно меньше уровня значимости, значит, различия между выборками статистически значимы.

Эталонное решение:

```
import scipy.stats as stats

# Данные задач
x1 = [380, 420, 290]
y1 = [140, 360, 200, 900]

# Тест Манна-Уитни
```

```
u_statistic, p_value = stats.mannwhitneyu(x1, y1,
alternative='two-sided')

print(f"U-статистика: {u_statistic}")

print(f"P-значение: {p_value}")

# Вывод результатов

alpha = 0.05

if p_value < alpha:

    print("Есть статистически значимые различия между выборками.")

else:

    print("Нет статистически значимых различий между выборками.")
```

Задача 2.

Исследовалось влияние препарата на уровень давления пациентов. Сначала измерялось давление до приема препарата, потом через 10 минут и через 30 минут. Есть ли статистически значимые различия?

1е измерение до приема препарата: 150, 160, 165, 145, 155

2е измерение через 10 минут: 140, 155, 150, 130, 135

3е измерение через 30 минут: 130, 130, 120, 130, 125

Подсказка № 1

Парный t-тест используется для сравнения средних значений двух зависимых выборок. В данном случае измерения давления до и после приема препарата проводятся на одних и тех же пациентах, поэтому выборки зависимы. Важно понимать, что тест проверяет наличие значимых различий между двумя средними значениями.

Подсказка № 2

Перед выполнением теста нужно определить уровень значимости (**alpha**). Обычно используется значение 0.05, что соответствует 95% уровню доверия. Если p-значение теста меньше **alpha**, это указывает на наличие статистически значимых различий между выборками.

Подсказка № 3

После выполнения t-теста важно правильно интерпретировать p-значения. Если p-значение меньше уровня значимости, это означает, что различия между выборками статистически значимы. В коде анализируются различия между измерениями до приема препарата и через 10 минут, до приема препарата и через 30 минут, а также между измерениями через 10 минут и через 30 минут.

Подсказка № 4

Обратите внимание, что в задаче проверяются все три возможные пары измерений: (1) до приема препарата и через 10 минут, (2) до приема препарата и через 30 минут, (3) через 10 минут и через 30 минут. Это позволяет комплексно оценить влияние препарата на давление во

Эталонное решение:

```
import numpy as np

import scipy.stats as stats

# Данные задач

before = [150, 160, 165, 145, 155]

after_10 = [140, 155, 150, 130, 135]

after_30 = [130, 130, 120, 130, 125]

# Тесты для парных выборок

t_stat_10, p_value_10 = stats.ttest_rel(before, after_10)

t_stat_30, p_value_30 = stats.ttest_rel(before, after_30)

t_stat_10_30, p_value_10_30 = stats.ttest_rel(after_10, after_30)

print(f"t-статистика для давления до и через 10 минут: {t_stat_10},  
p-значение: {p_value_10}")

print(f"t-статистика для давления до и через 30 минут: {t_stat_30},  
p-значение: {p_value_30}")
```

```
print(f"t-статистика для давления через 10 и 30 минут:  
{t_stat_10_30}, p-значение: {p_value_10_30}")  
  
alpha = 0.05  
  
if p_value_10 < alpha:  
    print("Есть статистически значимые различия между давлением до  
приема и через 10 минут.")  
else:  
    print("Нет статистически значимых различий между давлением до  
приема и через 10 минут.")  
  
if p_value_30 < alpha:  
    print("Есть статистически значимые различия между давлением до  
приема и через 30 минут.")  
else:  
    print("Нет статистически значимых различий между давлением до  
приема и через 30 минут.")  
  
if p_value_10_30 < alpha:  
    print("Есть статистически значимые различия между давлением  
через 10 и 30 минут.")  
else:  
    print("Нет статистически значимых различий между давлением через  
10 и 30 минут.")
```

Задача 3.

Сравните 1 и 2 е измерения, предполагая, что 3го измерения через 30 минут не было.

Подсказка № 1

В t-тесте для независимых выборок сравниваются средние значения двух разных групп. В данном случае нужно будет рассмотреть, как отличаются результаты двух измерений и проверить их статистическую значимость.

Подсказка № 2

Перед использованием t-теста для независимых выборок важно проверить, равны ли дисперсии выборок. Если дисперсии не равны, при вычислении t-статистики используется параметр `equal_var=False`, чтобы учесть это.

Подсказка № 3

После вычисления t-статистики и p-значения, нужно проверить, меньше ли p-значение уровня значимости (`alpha`). Если да, то есть статистически значимые различия между двумя выборками; если нет, то различия незначимы.

Подсказка № 4

Определение уровня значимости. Обычно уровень значимости (`alpha`) задается значением 0.05, что соответствует 95% уровню доверия.

Эталонное решение:

```
# Данные задач

first_measurement = [150, 160, 165, 145, 155]
second_measurement = [140, 155, 150, 130, 135]

# Тест для независимых выборок (хотя данные парные, используем
t-тест для независимых выборок)

t_stat, p_value = stats.ttest_ind(first_measurement,
second_measurement, equal_var=False)

print(f"t-статистика: {t_stat}")

print(f"p-значение: {p_value}")

if p_value < alpha:

    print("Есть статистически значимые различия между первым и
вторым измерением.")
```

```
else:  
  
    print("Нет статистически значимых различий между первым и вторым  
измерением.")
```

Задача 4.

Даны 3 группы учеников плавания.

В 1 группе время на дистанцию 50 м составляют:

56, 60, 62, 55, 71, 67, 59, 58, 64, 67

Вторая группа : 57, 58, 69, 48, 72, 70, 68, 71, 50, 53

Третья группа: 57, 67, 49, 48, 47, 55, 66, 51, 54

Подсказка № 1

ANOVA (анализ вариаций) используется для сравнения средних значений трех и более групп. Этот метод позволяет определить, есть ли статистически значимые различия между группами.

Подсказка № 2

Перед использованием ANOVA важно убедиться, что данные в группах распределены нормально. Если нормальность не соблюдается, может потребоваться использование непараметрических тестов, таких как тест Краскела-Уоллиса.

Подсказка № 3

ANOVA предполагает, что дисперсии внутри групп равны. Если дисперсии существенно отличаются, результаты теста могут быть недостоверными. Для проверки этого предположения можно использовать тест Левена.

Подсказка № 4

F-статистика указывает на соотношение межгрупповой дисперсии к внутригрупповой. Если p-значение меньше уровня значимости (обычно 0.05), это свидетельствует о наличии статистически значимых различий между группами.

Подсказка № 5

Если ANOVA показывает значимые различия, рекомендуется провести пост-хок тесты (например, тест Тьюки), чтобы определить, какие именно группы отличаются друг от друга.

Эталонное решение:

```

import numpy as np

import scipy.stats as stats

# Данные задач

group1 = [56, 60, 62, 55, 71, 67, 59, 58, 64, 67]

group2 = [57, 58, 69, 48, 72, 70, 68, 71, 50, 53]

group3 = [57, 67, 49, 48, 47, 55, 66, 51, 54]

# ANOVA

f_statistic, p_value = stats.f_oneway(group1, group2, group3)

print(f"F-статистика: {f_statistic}")

print(f"p-значение: {p_value}")

alpha = 0.05

if p_value < alpha:

    print("Есть статистически значимые различия между группами.")

else:

    print("Нет статистически значимых различий между группами.")

```

Задача 5.

Заявляется, что партия изготавливается со средним арифметическим 2,5 см. Проверить данную гипотезу, если известно, что размеры изделий подчинены нормальному закону распределения. Объем выборки 10, уровень статистической значимости 5%

2.51, 2.35, 2.74, 2.56, 2.40, 2.36, 2.65, 2.7, 2.67, 2.34

Подсказка № 1

В этом задании проверяется нулевая гипотеза, которая утверждает, что среднее значение размера изделий равно 2,5 см. Альтернативная гипотеза утверждает, что среднее значение отличается от 2,5 см. Это двухсторонний тест, так как мы проверяем отклонение в обе стороны (больше или меньше 2,5 см).

Подсказка № 2

Поскольку объем выборки мал ($n = 10$), и известно стандартное отклонение выборки, используется t-тест для проверки гипотезы. Важно понимать, что t-тест лучше всего подходит для небольших выборок.

Подсказка № 3

t-статистика показывает, насколько сильно среднее значение выборки отклоняется от гипотетического среднего (2,5 см), выраженное в единицах стандартной ошибки. Чем больше значение t, тем сильнее отклонение.

Подсказка № 4

Критическое значение t зависит от уровня значимости (**alpha**) и числа степеней свободы ($n - 1$). Если абсолютное значение t-статистики превышает критическое значение, то нулевая гипотеза отвергается.

Подсказка № 5

После вычисления t-статистики и сравнения с критическим значением, результат позволяет сделать вывод о том, есть ли статистически значимые доказательства того, что среднее значение размера изделий отличается от заявленного 2,5 см. Если t-статистика выходит за пределы критического значения, это указывает на значимые различия.

Эталонное решение:

```
import numpy as np

import scipy.stats as stats

# Данные задач
data = [2.51, 2.35, 2.74, 2.56, 2.40, 2.36, 2.65, 2.7, 2.67, 2.34]

mu_0 = 2.5 # Среднее значение гипотезы
alpha = 0.05

n = len(data)
```



```
# Среднее значение и стандартное отклонение выборки

mean_X = np.mean(data)

std_dev_X = np.std(data, ddof=1)

# Т-статистика

t_stat = (mean_X - mu_0) / (std_dev_X / np.sqrt(n))

# Критическое значение для t-распределения

t_crit = stats.t.ppf(1 - alpha / 2, df=n - 1)

print(f"t-статистика: {t_stat}")

print(f"Критическое значение t: {t_crit}")

# Проверка гипотезы

if abs(t_stat) > t_crit:

    print("Отвергаем нулевую гипотезу.")

else:

    print("Нет оснований отвергать нулевую гипотезу.")
```