

Домашнее задание 3. Вариант 4. Мосолков Евгений Николаевич БПИ196

Данные:

```
sex data: 0,1,1,1,0,1,1,0,0,1,0,1,0,1,1,1,0,0,0,0,0,1,0,0,1,0,0,0,1,1,1,0,1,1,1,0,1,0,0,0,0,0,1,0,1,0,0
a data: 4,7,5,5,4,3,7,5,3,4,3,7,5,7,3,7,7,3,4,6,5,5,1,1,2,3,5,3,5,5,6,6,4,4,6,5,4,5,4,1,7,3,3,7,3,7,3,2
b data: 4,3,3,4,4,5,3,4,6,6,7,4,5,5,4,4,4,5,5,7,5,4,5,5,4,5,5,6,3,4,4,4,4,5,4,5,4,4,5,4,4,4,4,5,3,5,4
```

1. Есть ли основание считать, что потенциальный потребитель предпочтет воду А воде В?

Используем критерий знаков. Запишем гипотезы:

$$H_0: P(A > B) = 0$$

$$H_1: P(A > B) > 0$$

Сразу посчитаем разницу массива а и b, а также посчитаем количество чисел, где $a_i \neq b_i$ (как положительные, так и отрицательные числа):

```
Count of A = B = 8
Count of A > B = 19
Count of A < B = 21
```

Заранее определим критическое значение $z_{0.1}$ - воспользуемся `scipy.stats.norm.ppf`, вычислим критическое значение для уровня значимости в 10 процентов $z_{0.1} = -1.2815515655446004$

Далее посчитаем статистику по формуле $z = \frac{2n^- - n}{\sqrt{n}} = \frac{42 - 40}{\sqrt{40}} = 0.31622776601683794$

$$z_{0.1} < z$$

Получается у нас есть достаточно оснований отвергнуть нулевую гипотезу, следовательно гипотеза H_1 подтверждается, значит потенциальный потребитель предпочитает воду А воде В.

Ответ: Да – есть достаточные основания считать, что потенциальный потребитель предпочитает воду А воде В.

2. Связано ли отношение к воде А с полом потребителя?

Используем критерий ранговых сумм Уилкоксона. Запишем гипотезы:

$$H_0: M(x) = W(x)$$

$$H_1: M(x) \neq W(x)$$

Посчитаем количество мужчин и женщин, а так же отсортируем данные массива а по половому признаку:

```
Count of men = 26
Count of women = 22
Men votes = [4, 4, 5, 3, 3, 5, 7, 3, 4, 6, 5, 1, 1, 3, 5, 3, 4, 4, 4, 1, 7, 3, 3, 3, 3, 2]
Women votes = [7, 5, 5, 3, 7, 4, 7, 7, 3, 7, 5, 2, 5, 5, 6, 6, 4, 6, 5, 5, 7, 7]
```

Теперь выпишем все индексы женских оценок из массива sex, получаем:

```
[1, 2, 3, 5, 6, 9, 11, 13, 14, 15, 21, 24, 28, 29, 30, 31, 33, 34, 35, 37, 43, 45]
```

Теперь нам нужно найти

- W_n – значение критерия Уилкоксона

- W_r – правая критическая точка
- W_l – левая критическая точка

Тогда значение критерия Уилкоксона можно найти как сумму всех индексов женских оценок:

$$W_n = 469$$

А критические точки найдем по формулам:

- $W_l = \text{int}\left(\frac{m(m+n+1)-1}{2} - 1.644854 * \sqrt{\frac{m * n * (m+n+1)}{12}}\right)$
- $W_r = m * (m+n+1) - W_l$

В данных формулах m – наименьший размер выборки, а n – наибольший. В первой формуле 1.644854 – число обратной функции нормального распределения.

Получаем, что $W_l = 459$, $W_r = 619$, следовательно число W_n попадает в найденный интервал, значит у нас нет оснований отвергать нулевую гипотезу, но и нельзя отрицать, что отношение к воде А не связано с полом потребителя.

Ответ: Отношение к воде А может быть связано с полом потребителя.