

In []:

ДЗ 1 задача 3

Юдина Екатерина БПИ198

Вариант 1, проверяющая Дубина Н.А., nadubina@edu.hse.ru.

У с л о в и е :

—

Производитель газированной воды, планирующий продвижение нового товара (вода А), заказал вам малое обследование потенциальных потребителей. Вы уже собрали нужные данные и записали их в файл «Данные к задаче 3.ods»:

a — общая оценка респондентом воды А по семибалльной шкале (1 — совсем не понравилось, 7 — превосходно);

b — оценка респондентом воды В (предполагаемого конкурента);

sex — пол респондента (0 — мужской, 1 — женский).

1) Есть ли основание считать, что потенциальный потребитель предпочитает воду А воде В?

2) Связано ли отношение к воде А с полом потребителя? От этого зависит стратегия продвижения товара

P.S.

Для ответа на первый вопрос решено использовать критерий знаков, для ответа на второй — критерий ранговых сумм Уилкоксона. Выбран уровень значимости 10%

В каждом случае сформулируйте основную и альтернативную гипотезы, рассчитайте статистику, приведите критическое значение (или значения) и сделайте вывод

1) Есть ли основание считать, что потенциальный потребитель предпочитает воду А воде В?

Р е ш е н и е

—

1. Составим таблицу
2. Сформулируем гипотезы
3. Найдём разности
4. Распишем знаки
5. Выберем статистику
6. Посчитаем критическое и эмпирическое значение критерия
7. Сформулируем вывод

► Таблица знаков:

sex1	a1	b1	diff	sign	total +	total -	total =	total
1	7	4	3	+	26	14	4	44
0	3	5	-2	-				
0	7	5	2	+				
1	7	5	2	+				
1	5	4	1	+				

0	7	4	3	+
0	4	5	-1	-
1	7	5	2	+
1	6	4	2	+
0	5	6	-1	-
0	4	4	0	0
0	7	4	3	+
1	4	6	-2	-
1	7	6	1	+
0	2	5	-3	-
1	5	4	1	+
1	7	5	2	+
1	7	5	2	+
0	3	4	-1	-
0	5	6	-1	-
0	4	4	0	0
1	4	3	1	+
1	4	5	-1	-
1	4	4	0	0
1	5	6	-1	-
1	6	5	1	+
0	3	6	-3	-
1	6	5	1	+
0	5	4	1	+
1	6	3	3	+
1	7	4	3	+
0	5	4	1	+
0	6	5	1	+
1	7	6	1	+
0	4	5	-1	-
0	2	5	-3	-
0	3	4	-1	-
1	6	2	4	+
0	4	5	-1	-
1	5	4	1	+
1	5	4	1	+
0	4	4	0	0
1	7	5	2	+
1	6	4	2	+

Попробуем проверить гипотезу о том, что потенциальный потребитель предпочитает воду А воде В на уровне значимости 10%. В 26 случаях предпочтение отдано производителю А («+»), в 14 — производителю В («-»). Обозначим за X оценку варианта А, а за Y — В.

► Сформулируем гипотезы:

$$H_0: P(X_i > Y_i) = P(X_i < Y_i)$$

$$H_A: P(X_i > Y_i) > P(X_i < Y_i)$$

Пусть пары таковы, что $(X_1, Y_1), \dots, (X_{40}, Y_{40})$

$$d_i = X_i - Y_i \sim i. i. d.$$

$n^+ \sim Bi(40, 0.5)$ - число наблюдений, где $d_i > 0 (X_i > Y_i)$

$n^- \sim Bi(40, 0.5)$ число наблюдений, где $d_i < 0 (X_i < Y_i)$

Выявляем наблюдения где $P(X_i = Y_i) = 0$

► Возьмём в качестве статистики $T = n^+$

► Критическая область: $T \leq T_{n,\alpha}$

В пользу будут говорить значения, близкие к 0. Выберем критическую область возле 0 так, чтобы вероятность попасть туда при H_0 не превосходила уровень значимости 10%.

$$n^+ = 24,$$

$$n^- = 16,$$

$$n = n^+ + n^- = 40$$

$$E(n^+) = n/2$$

$$D(n^+) = n/4$$

► n - достаточно большое

По интегральной теореме Муавра-Лапласа:

$$Z = \frac{2n^+ - n}{\sqrt{n}} \sim N(0, 1)$$

$$Z = \frac{2 \cdot 24 - 40}{\sqrt{40}} = -1.26$$

$$z_\alpha = z_{0,1} = -1.23$$

► $Z < z_\alpha \rightarrow$ Попали в критическую область \rightarrow отвергаем основную в пользу альтернативной, что значит, что потенциальный потребитель предпочитает воду А воде В на уровне значимости 10%.

2) Связано ли отношение к воде А с полом потребителя? От этого зависит стратегия продвижения товара

Р е ш е н и е

—

Воспользуемся критерием 'Ранговых сумм' ('критерий Манна — Уитни')

► Составим единый ранжированный ряд из обеих сопоставляемых выборок, расставив их элементы по степени нарастания признака и приписав меньшему значению меньший ранг (при наличии повторяющихся элементов в выборке используем средний ранг)

Общее количество рангов получится равным $N = n_1 + n_2$

где n_1 — количество элементов в первой выборке,

а n_2 — количество элементов во второй выборке.

Единый ранжированный ряд:

(все данные записаны в порядке возрастания так как переносились вручную, НО потом проверялись)

total	Rg					
2	1,5	(1+2)/2				
2	1,5					
3	4,5	(3+4+5+6)/4				
3	4,5					
3	4,5					

3	4,5					
4	11,50	(7+8+9+10+11+12+13+14+15+16)/10				
4	11,50	115/10				
4	11,50					
4	11,50					
4	11,50					
4	11,50					
4	11,50					
4	11,50					
4	11,50					
4	11,50					
5	21,00	(17+18+19+20+21+22+23+24+25)/9				
5	21,00	21				
5	21,00					
5	21,00					
5	21,00					
5	21,00					
5	21,00					
5	21,00					
5	21,00					
5	21,00					
6	29,00	(26+27+28+29+30+31+32)/7				
6	29,00	29				
6	29,00					
6	29,00					
6	29,00					
6	29,00					
6	29,00					
6	29,00					
7	38,50	(33+34+35+36+37+38+39+40+41+42+43+44)/12				
7	38,50	38,5				
7	38,50					
7	38,50					
7	38,50					
7	38,50					
7	38,50					
7	38,50					
7	38,50					
7	38,50					
7	38,50					
7	38,50					
7	38,50					
Все возможные значения из двух выборок						

► Таблица рангов для каждой из выборок:

	men	Rg	women	Rg
	2	1,5	4	11,50
	2	1,5	4	11,50
	3	4,50	4	11,50
	3	4,50	4	11,50

	3	4,50	5	21,00
	3	4,50	5	21,00
	4	11,50	5	21,00
	4	11,50	5	21,00
	4	11,50	5	21,00
	4	11,50	6	29,00
	4	11,50	6	29,00
	4	11,50	6	29,00
	5	21,00	6	29,00
	5	21,00	6	29,00
	5	21,00	6	29,00
	5	21,00	7	38,50
	6	29,00	7	38,50
	7	38,50	7	38,50
	7	38,50	7	38,50
	7	38,50	7	38,50
			7	38,50
			7	38,50
			7	38,50
			7	38,50
total	20	318,5	24	671,50
	Количество	Сумма	Количество	Сумма
	во	рангов	во	рангов
	мужчин	для	женщин	для
		мужчин		женщин

► Сформулируем гипотезы:

H_0 : Пол не влияет на ответы респондентов

H_A : Пол влияет на ответы респондентов

► Сумма рангов для мужчин: $R_m = \sum_{i=1}^{20} Rg_i = 318.5$

► Сумма рангов для женщин: $R_w = \sum_{i=1}^{24} Rg_i = 671.5$

► Большая ранговая сумма: $R_w = 671.5$

► Вычислим значение статистики $U = n_1 \cdot n_2 + n_x \cdot (n_x + 1)/2 - T_x$,

где n_1 – объем выборки №1;

n_2 – объем выборки №2;

T_x – большая из двух ранговых сумм;

n_x – объем максимальной выборки: $n_x = \max(n_1, n_2)$.

В нашем случае: $U = 20 \cdot 24 + 24 \cdot (24 + 1)/2 - 671.5 = 108.5$

► Критическое правило: $U_{эмп} \leq U_{кр}$ -- H_0 отвергается.

Определим критические значения $U_{кр}$ по таблице. Так как гипотеза двусторонняя, то ищем $U_{0,05} = 169$

$U_{эмп} < U(0.05)$

↔

$108.5 < 169$

► Наше значение попадает в зону значимости (в критическую зону), что говорит о том, что пол влияет на выставление оценки воде производителя А. Следовательно и стратегию продвижения товара надо выработать согласно тому, что пол влияет