Задача 3. Вариант 8 Зубарева Наталия БПИ195

В былые дни возле станции метро «Площадь Революции» проводились различные обследования (опросы и дегустации), за участие в которых можно было получить коробку конфет, шоколадку, бутылку пива или даже торт. Представьте себя на месте добрых людей, раздающих эти продукты.

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).

Требуется ответить на два вопроса.

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.

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

Наши данные - три массива по 45 позиций соответственно для полов респондентов, их оценок воды А и воды В:

people=[0 1 1 1 0 1 0 0 1 1 1 1 0 0 1 0 0 1 0 0 0 1 1 0 0 0 1 1 0 0 1 1 1 1 1 1 0 1 1 1 0 0 1 0 0]
statA = [5 3 6 5 4 5 2 7 7 6 4 6 4 7 7 3 5 1 4 2 4 4 2 6 5 4 3 3 4 5 6 3 4 4 3 5 4 5 3 2 4 1 4 5 2]
statB = [5 5 5 5 5 5 6 5 4 5 6 5 5 5 5 7 6 7 5 3 6 5 4 5 7 6 6 3 6 4 6 6 4 5 5 5 6 4 6 4 6 6 5 5 7]

1) Ответим на вопрос, есть ли основания считать, что потребитель предпочитает воду А воде В. Для этого выдвинем гипотезы и проверим их на уровне значимости 10%:

H0: P(statA(i) - statB(i) > 0) = P(statA(i) - statB(i) < 0) - потребитель не отдаёт предпочтение никакой воде
H1: P(statA(i) - statB(i) > 0) > P(statA(i) - statB(i) < 0) - потребитель предпочитает воду А воде В

где statA(i) и statB(i) - оценки i-тым респондентом воды А и В соотв.
(Случай равенства оценок statA(i) и statB(i) мы учитывать не будем)

Статистика: $Z = \frac{2n^- - n}{\sqrt{n}}$ ~N(0,1), где n - общее число респондентов, для которых ai - bi ≠ 0, n^- - число респондентов, для которых ai - bi < 0

Найдём критическое значение статистики - у нас односторонняя критическая область на уровне доверия α = 0.1, найдём квантиль в таблице:
$Z_\alpha = Z_{0.1} = - Z_{0.9} \approx -1.282$

Если наша статистика окажется меньше данного критического значения, мы отвергнем основную гипотезу, потому что это будет значить, что воде А отдаётся предпочтение, если же значение будет правее критического, отвергать H0 не будем

Введём наши массивы в матлаб и найдём n и n^-:
n = 37
n^- = 27

```
>> findSigns(statA, statB)

ans =

    37    27
```

```
function res = findSigns(statA, statB)
numberNotEquals = 0;
numberLess = 0;
for i = 1:length(statA)
    if statA(i) ~= statB(i)
    numberNotEquals = numberNotEquals + 1;
    if statA(i) - statB(i) < 0
        numberLess = numberLess+1;
    end
    end
    res = [numberNotEquals numberLess];
end
```

Вычислим статистику Z:
$Z = (2*27 - 37)/\sqrt{37} \approx 2.795 > -1.282 > Z\alpha$

Z > Zα, статистика попала в критическую область, значит у нас нет оснований для того, чтобы отвергнуть основную гипотезу, на уровне значимости 10% мы считаем, что потребители не предпочитают воду A воде B

2) Чтобы узнать, связано ли отношение к воде A с полом потребителя, составим гипотезы и проверим их по критерию ранговых сумм Уилкоксона:

H0 - основная гипотеза: отношение к воде A не связано с полом потребителя
HA - альтернативная гипотеза: отношение к воде A связано с полом потребителя

Разобьём наши данные в соответствии с полом каждого потребителя:

```
function res = getFMStat(people, statA, pplCode)
statFM = [];
for i = 1:length(people)
    if people(i) == pplCode
    statFM(length(statFM)+1) = statA(i);
    end
end
  res = statFM;
end
```

Получилось, что у нас 21 потребитель и 24 потребительницы

```
>> males = getFMStat(people, statA, 0)

males =

     5     4     2     7     4     7     3     5     4     2     4     6     5     4     4     5     4     4     1     5     2

>> females = getFMStat(people, statA, 1)

females =

  Columns 1 through 23

     3     6     5     5     7     6     4     6     7     1     4     2     3     3     6     3     4     4     3     5     5     3     2

  Column 24

     4
```

Теперь соединим эти массивы назад, чтобы посчитать ранги оценок, для этого отсортируем матрицу из значений полов и оценок воды А в соответствии с оценкой и каждой оценке сопоставим ранг в зависимости от ее индекса в отсортированной матрице (усредняя ранги равных оценок) (матлаб это делает сам отдельной функцией)

```
>> ratingToRank = [statA tiedrank(statA)]

ratingToRank =

    5.0000   32.0000
    3.0000   11.0000
    6.0000   39.0000
    5.0000   32.0000
    4.0000   21.0000
    5.0000   32.0000
    2.0000    5.0000
    7.0000   43.5000
    7.0000   43.5000
    6.0000   39.0000
    4.0000   21.0000
    6.0000   39.0000
    4.0000   21.0000
    7.0000   43.5000
    7.0000   43.5000
    3.0000   11.0000
    5.0000   32.0000
    1.0000    1.5000
    4.0000   21.0000
    2.0000    5.0000
    4.0000   21.0000
    4.0000   21.0000
    2.0000    5.0000
    6.0000   39.0000
    5.0000   32.0000
    4.0000   21.0000
    3.0000   11.0000
    3.0000   11.0000
    4.0000   21.0000
    5.0000   32.0000
    6.0000   39.0000
    3.0000   11.0000
    4.0000   21.0000
    4.0000   21.0000
    3.0000   11.0000
    5.0000   32.0000
    4.0000   21.0000
    5.0000   32.0000
    3.0000   11.0000
    2.0000    5.0000
    4.0000   21.0000
    1.0000    1.5000
    4.0000   21.0000
    5.0000   32.0000
    2.0000    5.0000
```

Теперь для массивов отдельно для каждого пола (males и females) найдём сумму рангов оценок в них:

```
>> ranksSummF = findRanksSumm(females,ratingToRank)

ranksSummF =

  553.5000

>> ranksSummM = findRanksSumm(males,ratingToRank)

ranksSummM =

  481.5000
```

findRanksSumm.m

```
function summ = findRanksSumm(statFM, generalRanks)
summ = 0;
for i = 1:length(statFM)
    for j = 1:length(generalRanks)
    if statFM(i) == generalRanks(j,1)
        summ = summ + generalRanks(j,2);
        break;
    end
    end
end
```

Теперь сравним статистику W и критическое значение для критерия Уилкоксона:

Найдём критическое значение статистики в таблице по длинам массивов и уровню значимости: Wкрит (21 24 0.1) = [410; 556]. Если статистика не попадёт в этот интервал, мы отвергнем H0.

Сама статистика W будет равна сумме рангов того массива, в котором было меньше изначальных элементов (у нас это массив потребителей-мужчин)

W = 481.5 ∈ [410; 556] ∈ Wкрит

Статистика попадает в этот интервал, значит, по критерию Уилкоксона на уровне значимости 10% мы не отвергаем основную гипотезу и считаем, что отношение к воде А не зависит от пола потребителя