Галанов Ю.И.

Лабораторный практикум по мат. статистике

Однофакторный анализ

ПРИМЕР.

Для выяснения влияния денежного стимулирования на производительность труда шести однородным группам из 5 человек каждая были предложены задачи одинаковой трудности. Задачи предлагались каждому испытуемому независимо от всех остальных. Группы отличаются между собой величиной денежного вознаграждения за решаемую задачу.

В таблице приведено число решенных задач членами каждой группы.

Проверяем гипотезу об отсутствии влияния денежного вознаграждения на число решенных задач.

Вводим данные:

$$M := \mathrm{READPRN}(\text{"krask.prn"})$$

$$M = \begin{pmatrix} 10 & 8 & 12 & 12 & 24 & 19 \\ 11 & 10 & 17 & 15 & 16 & 18 \\ 9 & 16 & 14 & 16 & 22 & 27 \\ 13 & 13 & 9 & 16 & 18 & 25 \\ 7 & 12 & 16 & 19 & 20 & 24 \end{pmatrix}$$

$$k := 0 .. \mathrm{cols}(M) - 1$$

$$j := 0 .. \mathrm{cols}(M) - 1$$

$$l := 0 .. \mathrm{rows}(M) - 1$$

$$i := 0 .. \mathrm{rows}(M) - 1$$

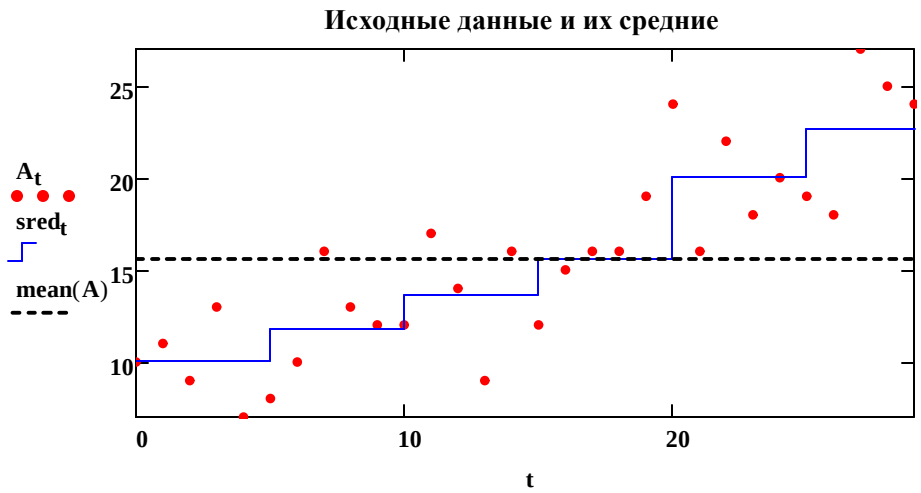
Представим их в виде одной выборки

$$A_{i+5\cdot j} := M_{i,j}$$

Посчитаем средние для каждого столбца:

$$c_j := \frac{1}{5}\cdot\sum_i M_{i,j} \qquad t := 0 .. last(A)$$

$$sred_t := if\left(t < 5, c_0, if\left(t < 10, c_1, if\left(t < 15, c_2, if\left(t < 20, c_3, if\left(t < 25, c_4, if\left(t < 30, c_5, c_5\right)\right)\right)\right)\right)\right)$$



Уже на данной стадии анализа виден сдвиг средних при переходе от первой группы к последней, однако необходимо убедиться в том, что эти сдвиги значимы.

Перейдем к ранговому представлению данных:

$$R_{i,j} := \sum_l \sum_k if\left(M_{l,k} - M_{i,j} > 0, 0, 1\right)$$

Получим таблицу рангов

$$M = \begin{pmatrix} 10 & 8 & 12 & 12 & 24 & 19 \\ 11 & 10 & 17 & 15 & 16 & 18 \\ 9 & 16 & 14 & 16 & 22 & 27 \\ 13 & 13 & 9 & 16 & 18 & 25 \\ 7 & 12 & 16 & 19 & 20 & 24 \end{pmatrix} \qquad R = \begin{pmatrix} 6 & 2 & 10 & 10 & 28 & 24 \\ 7 & 6 & 20 & 14 & 19 & 22 \\ 4 & 19 & 13 & 19 & 26 & 30 \\ 12 & 12 & 4 & 19 & 22 & 29 \\ 1 & 10 & 19 & 24 & 25 & 28 \end{pmatrix}$$

Повторяющиеся ранги заменяем средними рангами

$R_{0,0} := 5.5$ $R_{1,1} := 5.5$ $R_{3,0} := 11.5$ $R_{3,1} := 11.5$

$R_{0,5} := 23.5$ $R_{4,3} := 23.5$ $R_{2,0} := 3.5$ $R_{3,2} := 3.5$

$R_{1,5} := 21.5$ $R_{3,4} := 21.5$ $R_{0,4} := 27.5$ $R_{4,5} := 27.5$

$R_{4,1} := 9$ $R_{0,2} := 9$ $R_{0,3} := 9$

$R_{2,1} := 17$ $R_{2,3} := 17$ $R_{4,2} := 17$ $R_{3,3} := 17$ $R_{1,4} := 17$

$$R = \begin{pmatrix} 5.5 & 2 & 9 & 9 & 27.5 & 23.5 \\ 7 & 5.5 & 20 & 14 & 17 & 21.5 \\ 3.5 & 17 & 13 & 17 & 26 & 30 \\ 11.5 & 11.5 & 3.5 & 17 & 21.5 & 29 \\ 1 & 9 & 17 & 23.5 & 25 & 27.5 \end{pmatrix}$$

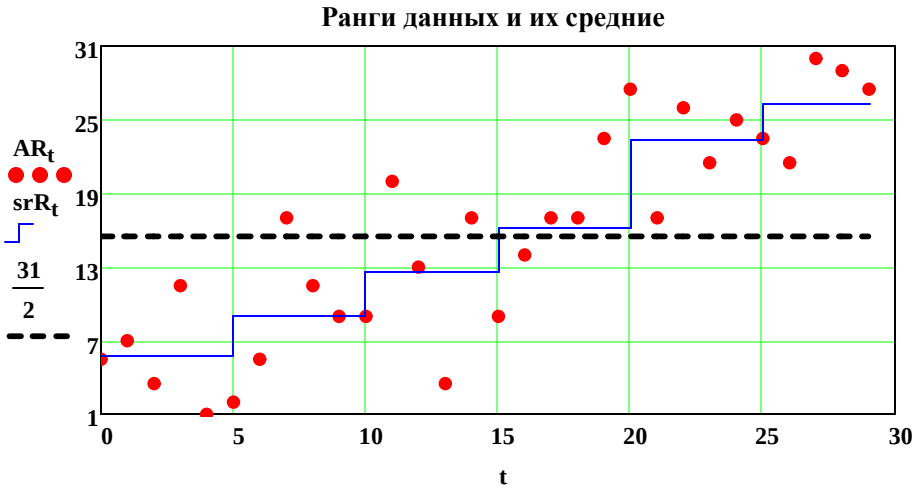
средние по столбцам ранги

$$Rj_j := \frac{1}{5} \cdot \sum_i R_{i,j}$$

$$Rj^T = (5.7 \quad 9 \quad 12.5 \quad 16.1 \quad 23.4 \quad 26.3)$$

$$srR_t := \text{if}\left(t < 5, Rj_0, \text{if}\left(t < 10, Rj_1, \text{if}\left(t < 15, Rj_2, \text{if}\left(t < 20, Rj_3, \text{if}\left(t < 25, Rj_4, \text{if}\left(t < 30, Rj_5, c_5\right)\right)\right)\right)\right)\right)$$

$$AR_{i+5\cdot j} := R_{i,j}$$



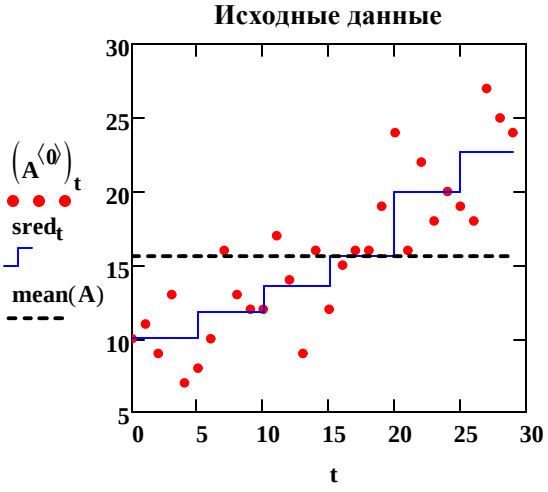
Проверим нулевую гипотезу:

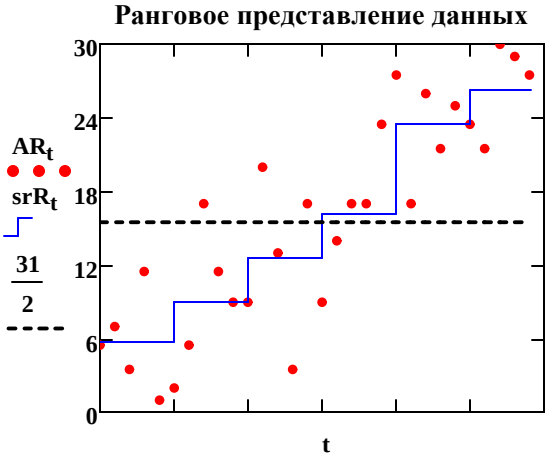
Данные, представленные в таблице принадлежат одной выборке.

Посчитаем статистику Краскела-Уоллеса:

$$H := \frac{12 \cdot 5}{30 \cdot 31} \cdot \sum_j \left(Rj_j - \frac{31}{2} \right)^2 \qquad H = 21.077419 \qquad qchisq(0.999, 5) = 20.515006$$

На уровне значимости 0.001 нулевая гипотеза отвергается





При переходе от количественой шкалы к порядковой наблюдается некоторая потеря информации, которая проявляется в большем рассеивании данных. Однако это не сказывается на окончательных результатах.

Особенность ранговых методов состоит в том, что распределение ранговых статистик не зависит от распределения исходных данных.

Поэтому они позволяют обрабатывать данные, изначально представленные в порядковых или номинальных шкалах.

Эта функция возвращает исправленное значене рангов

```
f3(x) := | n ← last(x)
         | for i ∈ 0..n
         |   x[i,1] ← i
         | x ← csort(x, 0)
         | for i ∈ 0..n
         |   x[i,2] ← i + 1
         | i ← 0
         | while i < n
         |   | if x[i,0] ≠ x[i+1,0]
         |   |   | x[i,3] ← 1
         |   |   | i ← i + 1
         |   | otherwise
         |   |   | k ← 1
         |   |   | M ← x[i,2]
         |   |   | while x[i,0] = x[i+1,0]
         |   |   |   | k ← k + 1
         |   |   |   | i ← i + 1
         |   |   |   | M ← M + x[i,2]
         |   |   | d ← k
         |   |   | for j ∈ 0..k − 1
         |   |   |   x[i−j,2] ← M/d
         |   |   | i ← i + 1
         | x ← csort(x, 1)
         | x⟨2⟩
```

$R1 := f3(A)$ $\qquad M1_{i,j} := R1_{i+5\cdot j}$

$$M1 = \begin{pmatrix} 5.5 & 2 & 9 & 9 & 27.5 & 23.5 \\ 7 & 5.5 & 20 & 14 & 17 & 21.5 \\ 3.5 & 17 & 13 & 17 & 26 & 30 \\ 11.5 & 11.5 & 3.5 & 17 & 21.5 & 29 \\ 1 & 9 & 17 & 23.5 & 25 & 27.5 \end{pmatrix}$$

$$R = \begin{pmatrix} 5.5 & 2 & 9 & 9 & 27.5 & 23.5 \\ 7 & 5.5 & 20 & 14 & 17 & 21.5 \\ 3.5 & 17 & 13 & 17 & 26 & 30 \\ 11.5 & 11.5 & 3.5 & 17 & 21.5 & 29 \\ 1 & 9 & 17 & 23.5 & 25 & 27.5 \end{pmatrix}$$

Проверочные суммы

$$\sum_t R1_t = 465 \qquad 30\cdot\frac{31}{2} = 465$$

$$BS_{t,2} := t + 1 \qquad BS = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & 3 \\ 0 & 0 & 4 \\ 0 & 0 & 5 \\ 0 & 0 & 6 \\ 0 & 0 & 7 \\ 0 & 0 & 8 \\ 0 & 0 & 9 \\ 0 & 0 & 10 \\ 0 & 0 & 11 \\ 0 & 0 & 12 \\ 0 & 0 & 13 \\ 0 & 0 & 14 \\ 0 & 0 & 15 \\ 0 & 0 & 16 \\ 0 & 0 & 17 \\ 0 & 0 & 18 \\ 0 & 0 & 19 \\ 0 & 0 & 20 \\ 0 & 0 & 21 \\ 0 & 0 & 22 \\ 0 & 0 & 23 \\ 0 & 0 & 24 \\ 0 & 0 & 25 \\ 0 & 0 & 26 \\ 0 & 0 & 27 \\ 0 & 0 & 28 \\ 0 & 0 & 29 \\ 0 & 0 & 30 \end{pmatrix} \qquad BS := csor$$

$$t(BS, 1) \qquad BS = \begin{pmatrix} 0 & 0 & 2 \\ 0 & 0 & 3 \\ 0 & 0 & 4 \\ 0 & 0 & 5 \\ 0 & 0 & 6 \\ 0 & 0 & 7 \\ 0 & 0 & 8 \\ 0 & 0 & 9 \\ 0 & 0 & 10 \\ 0 & 0 & 11 \\ 0 & 0 & 12 \\ 0 & 0 & 13 \\ 0 & 0 & 14 \\ 0 & 0 & 15 \\ 0 & 0 & 16 \\ 0 & 0 & 17 \\ 0 & 0 & 18 \\ 0 & 0 & 19 \\ 0 & 0 & 20 \\ 0 & 0 & 21 \\ 0 & 0 & 22 \\ 0 & 0 & 23 \\ 0 & 0 & 24 \\ 0 & 0 & 25 \\ 0 & 0 & 26 \\ 0 & 0 & 27 \\ 0 & 0 & 28 \\ 0 & 0 & 29 \\ 0 & 0 & 30 \\ 0 & 0 & 1 \end{pmatrix}$$