07/06/2016

Лекція 5. /2год./ Гіпотези однорідності. Параметричні та непараметричні. Т-критерій Стьюдента. Критерій Колмогорова-Смирнова. Критерій Вілкоксона. Критерій знаків. Критерій Мана-Уітні.
До лекції №5 практична №6.

5.1 Класифікація критеріїв.

Статистичні критерії підрозділяють на такі категорії:

- **Критерій значущості.** Перевірка за значущістю припускає перевірку гіпотези про числові значення відомого закону розподілу:

  $H_0 : \quad a = a_0$— нульова гіпотеза
  $H_1 : \quad a > a_0 \quad (a < a_0)$або $a \neq a_0$—альтернативна, що конкурує.
- **Критерій узгодженості.** Перевірка на узгодженість має на увазі, що випадкова величина, що досліджується,підкорюється закону, що розглядається. Критерій узгодженості можна також сприймати, як критерій значущості. Усі критерії узгодженості будуються за єдиним принципом: задається деяка *функція відхилення* емпіричного розподілу від теоретичного. Вимагають, щоб ця функція збігалась до якогось власного розподілу, якщо справедлива гіпотеза, що перевіряється, і необмежено зростала, якщо гіпотеза не справедлива. Гіпотеза приймається або відкидається в залежності від величини даної функції відхилення. Наприклад, критерій Колмогорова та критерій «хі-квадрат» Пірсона (параметричний та непараметричний), Т-критерій Стьюдента.
- **Критерій однорідності.**При перевірці на однорідність випадкові величини досліджуються на факт взаємної відповідності їх законів розподілу (чи підкорюються ці величини одному і тому ж закону). Також використовуються у Факторному аналізі для визначення наявності залежностей. Наприклад, критерій Колмогорова-Смірнова, критерій Вілкоксона, критерій знаків, критерій Мана-Уітні.

Цей розділ умовний, і часто один і той же критерій може бути використаний в різних якостях. Наприклад, Т-критерій узгодженості використовують для перевірки гіпотези однорідності.

Приклад. Нехай дано незалежну вибірку $\mathbf{X} = (X_1, \ldots, X_n)^\top$, де $X_i \sim \mathrm{N}(\mu, 1), \quad i = 1, \ldots, n$. Нехай маємо дві прості гіпотези:

$$H_0 : \quad \mu = 0,$$
$$H_1 : \quad \mu = 1.$$

Тоді можна визначити наступний статистичний критерій значущості :

$$f(x_1, \ldots, x_n) = \begin{cases} H_0, & \bar{x} \le 0.5 \\ H_1, & \bar{x} > 0.5, \end{cases}$$ де $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ - вибіркове середнє.

Статистичні критерії також підрозділяють на такі види:

- **параметричні**
- **непараметричні**

Група статистичний критеріїв, які включають в розранунок параметри ймовірнісного розподілу ознаки (середнього і дисперсії) називаються **параметричними.**

- t-критерій Стьюдента
- Критерій «хі-квадрат» Пірсона

Група статистичних критеріїв, які не включають в розрахунок параметри ймовірнісного розподілу і засновані на оперуванні частотами або рангами називаються **непараметричними.**

5.2. Параметричні критерії для перевірки гіпотези однорідності: Т-критерій Стьюдента.

Постановка задачі. Розглянемо дві незалежні вибірки $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ і $\boldsymbol{y} = (y_1, y_2, ..., y_m)$, причому x_i розподілено за нормальним законом з параметрами (a_1, σ^2), y_j розподілено за нормальним законом з параметрами (a_2, σ^2). Дисперсія σ^2 однакова для обох розподілів, але, загалом, невідома. Перевіряється складна гіпотеза: $H_1 : a_1 = a_2$ проти складної альтернативи H_2: H_1 не справедлива.

Побудуємо Т-критерій Стьюдента точного рівня ε.

Позначимо через $s_0^2(\boldsymbol{x}), s_0^2(\boldsymbol{y})$ незміщені вибіркові дисперсії:

$$s_0^2(\boldsymbol{x}) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2,\ s_0^2(\boldsymbol{y}) = \frac{1}{m-1}\sum_{j=1}^{m}(y_i - \bar{y})^2 .$$

Теорема 5. *Випадкова величина*

$$t_{n+m-2} = \sqrt{\frac{nm}{n+m}}\frac{(\bar{x} - a_1) - (\bar{y} - a_2)}{\sqrt{\dfrac{(n-1)s_0^2(\boldsymbol{x}) + (m-1)s_0^2(\boldsymbol{y})}{n+m-2}}}$$

має розподіл Стьюдента з $n+m-2$ *степенями вільності.*

Доведення. Домовимось символом $\xi \sim F$ позначати факт, що випадкова величина ξ має закон розподілу F. Запис $N(a, \sigma^2)$ буде позначати нормальний закон розподілу з параметрами (a, σ^2). Якщо ξ_1 и ξ_2 - независимые нормально распределенные случайные величины с параметрами a_1, σ_1^2 и a_2, σ_2^2 соответственно, то их сумма $\xi_1 + \xi_2$ тоже распределена по нормальному закону, притом с параметрами $a_1 + a_2$ и $\sigma_1^2 + \sigma_2^2$.

Зрозуміло, що $M(\bar{x} - a_1) = M(\bar{x}) - M(a_1) = a_1 - a_1 = 0$.

$$D(\bar{x} - a_1) = D(\bar{x}) - D(a_1) = D(\frac{\sum_{i=1}^{n} x_i}{n}) = \frac{1}{n^2}D(\sum_{i=1}^{n} x_i) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}.$$

Отже,

$$\bar{x} - a_1 \sim N(0, \frac{\sigma^2}{n}),\ \bar{y} - a_2 \sim N(0, \frac{\sigma^2}{m}).$$

Таким чином,

$$(\bar{x} - a_1) - (\bar{y} - a_1) \approx N(0, \frac{\sigma^2}{n} + \frac{\sigma^2}{m}) = N(0, \frac{\sigma^2(n+m)}{nm}).$$

Звідси

$$\xi_0 = \frac{1}{\sigma}\sqrt{\frac{nm}{n+m}}((\bar{x} - a_1) - (\bar{y} - a_2)) \sim N(0,1).$$

Відомо (лема Фішера), що

$$\frac{n-1}{\sigma^2}s_0^2(\boldsymbol{x}) \sim \chi^2_{n-1},\ \frac{m-1}{\sigma^2}s_0^2(\boldsymbol{y}) \sim \chi^2_{m-1}.$$

Тоді

$$S = \frac{1}{\sigma^2}((n-1)s_0^2(\boldsymbol{x}) + (m-1)s_0^2(\boldsymbol{y})) \sim \chi^2_{n+m-2}$$

і не залежать від $\bar{x}$ і $\bar{y}$.

Тому випадкова величина

$$\frac{\xi_0}{\sqrt{S/(n+m-2)}}$$

має розподіл Стьюдента з $n+m-2$ степенями вільності. Залишилось підставити ξ_0 і S і переконатись, що цей дріб співпадає з t_{n+m-2}.

Алгоритм Т-критерію Стьюдента:

К1. Розглянемо функцію відхилення

$$\rho = \sqrt{\frac{nm}{n+m}} \frac{(\bar{x}-\bar{y})}{\sqrt{\frac{(n-1)s_0^2(\boldsymbol{x})+(m-1)s_0^2(\boldsymbol{y})}{n+m-2}}}.$$

З попередньої теореми випливає властивість $K1(a)$, тобто якщо гіпотеза H_1 справедлива, то функція відхилення збігається до відомого розподілу. Тоді ρ має розподіл Стьюдента з $n+m-2$ степенями вільності. Для доведення властивості *К1(б)* можна скористатися законом великих чисел, відповідно до якого, у випадку справедливості гіпотези H_2, величини $\bar{x}-\bar{y}$ і $\frac{(n-1)s_0^2(\boldsymbol{x})+(m-1)s_0^2(\boldsymbol{y})}{n+m-2}$ будуть мати скінченні, відмінні від нуля границі, у той час як $\frac{nm}{n+m}\to\infty$. Таким чином, у випадку справедливості гіпотези H_2 виконується співвідношення $|\rho|\to\infty$.

К2: Тому залишається за заданим ε вибрати таке c, що для величини t_{n+m-2}, у силу симетричності розподілу Стьюдента,

$$\varepsilon = \mathbf{P}\{|t_{n+m-2}| \geq c\} = 2\mathbf{P}\{t_{n+m-2} \geq c\},$$

тобто $\mathbf{P}\{t_{n+m-2} > c\} = \varepsilon/2$. Звідси $c=\tau_{1-\varepsilon/2}$, де $\tau_{1-\varepsilon/2}$ – квантиль розподілу Стьюдента зі $n+m-2$ степенем вільності. І критерій Стьюдента має вигляд як усі критерії узгодженості:

$$K(\boldsymbol{x},\boldsymbol{y}) = \begin{cases} H_1, & \text{åñëè } |\rho(\boldsymbol{x},\boldsymbol{y})| < c; \\ H_2, & \text{åñëè } |\rho(\boldsymbol{x},\boldsymbol{y})| \geq c. \end{cases}$$

Приклад 4.2.1 (с.106-108, В.М. Турчин):

Чи можна вважати, що між даними оцінювання однієї і тієї ж групи людей вома експертами нема систематичної розбіжності? Використаємо ***Т-критерій Стьюдента***

Перша вибірка: 2, 6, 7, 7, 9,5,3, 3, 2, 1,.4,2

Друга вибірка: 1, 4, 2, 2, 3, 5, 7, 5, 7, 3, 3, 5.

Сформулюємо гіпотези:

H_0: $a_1 = a_2$

H_1: $a_1 \neq a_2$

n=m=12

Знайдемо середні значення :

$\bar{x}$=(2+6+7+7+9+5+3+3+2+1+4+2) = 4,25

$\bar{y}$= (1+4+2+2+3+5+7+5+7+3+3+5) = 3,92

Знайдемо незміщені вибіркові дисперсії :

$$s_0^2(\boldsymbol{x}) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2,\ s_0^2(\boldsymbol{y}) = \frac{1}{m-1}\sum_{j=1}^{m}(y_i-\bar{y})^2.$$

$$s_0^2(x) = \frac{1}{11}\sum_{i=1}^{12}(x_i - 4{,}25)^2 = \quad 6{,}39$$

$$s_0^2(y) = \frac{1}{11}\sum_{i=1}^{12}(x_i - 3{,}92)^2 = \quad 3{,}72$$

Порахуємо емпіричне значення відхилення $t_{n+m-2} = \sqrt{\frac{nm}{n+m}}\dfrac{(\bar{x} - a_1) - (\bar{y} - a_2)}{\sqrt{\dfrac{(n-1)s_0^2(\boldsymbol{x}) + (m-1)s_0^2(\boldsymbol{y})}{n+m-2}}}$

За гіпотезою H_0: $a_1 = a_2$ Тому у зазначеній формулі замість a_2 запишемо a_1:

$$t_{12+12-2} = t_{22} = \sqrt{\frac{12*12}{12+12}}\frac{(4{,}25 - a_1) - (3{,}92 - a_1)}{\sqrt{\dfrac{11*s_0^2(x) + 11*s_0^2(y)}{12+12-2}}} \approx 2{,}45*\frac{0{,}33}{2{,}25} \approx 0{,}36$$

Тепер знайдемо теоретичне значення відхилення. Для $\lambda = 0{,}05{:}\,2 = 0{,}025$ та ступенів свободи 22 воно рівне 2,074.

$$0{,}36 < 2{,}074$$

Отже, можна вважати, що між показниками двох вибірок про вплив ремонту на кількість аварій по місяцях суттєвих розходжень немає.

5.3. Непараметричні критерії для перевірки гіпотези однорідності: критерій Колмогорова-Смирнова, критерій Вілкоксона, критерій знаків, критерій Мана-Уітні.

Постановка задачі.

Є дві вибірки $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ і $\boldsymbol{y} = (y_1, y_2, ..., y_m)$, причому x_i розподілені за законом $F_\xi(x)$, y_j – за законом $F_\eta(y)$, і розподіли $F_\xi(x)$ і $F_\eta(y)$, загалом, невідомі. Обєми вибірок не великі. Перевіряється складна гіпотеза про однорідність $H_1 : F_\xi = F_\eta$ проти складної альтернативи $H_2 : H_1$ не вірна.

Непараметричні критерії розрізняються для:

- залежних вибірок (студенти на початку року і в кінці року). Це – критерій знаків, критерій Вілкоксона,
- незалежник вибірок (чоловіки – жінки) Це критерії Вальда-Волловітця, Манна-Уітні, Колмогорова-Сміркова
- для перевірки гіпотези про залежність.

А) критерій Колмогорова – Смірнова (чутливий до форми розподілу);

Якщо $F_\xi(x)$ і $F_\eta(y)$ мають неперервні функції розподілу, можливо використати критерій Колмогорова-Смирнова.

Нехай $F^*_{n,x}$ і $F^*_{m,y}$ – емпіричні функції розподілу, побудовані за вибірками $\boldsymbol{x}$ і $\boldsymbol{y}$,

$$\rho(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\frac{mn}{m+n}}\sup_t\left|F^*_{n,x}(t) - F^*_{m,y}(t)\right|.$$

Теорема 5.3. (без доведення) *Якщо гіпотеза H_1 справедлива, тоді $\rho(\boldsymbol{x}, \boldsymbol{y})$ при $m, n \to \infty$ збігається за ймовірністю до випадкової величини ζ, розподіленої за законом Колмогорова.*

Нехай знову випадкова величина ζ має розподіл з функцією розподілу $K(t)$. За заданим $\varepsilon > 0$ знайдемо c таке, що $\varepsilon = \mathbf{P}\{\zeta \geq c\}$, і побудуємо критерій узгодженості **Колмогорова-Смирнова:**

$$K(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} H_1, & \text{åñëè } \rho(\boldsymbol{x}, \boldsymbol{y}) < c; \\ H_2, & \text{åñëè } \rho(\boldsymbol{x}, \boldsymbol{y}) \geq c. \end{cases}$$

Б) критерій знаків;

Найпростішим критерієм однорідності є ***критерій знаків.*** Якщо вибірки, що порівнюються $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ і $\boldsymbol{y} = (y_1, y_2, ..., y_n)$ мають однаковий об'єм і розподілені за одним й тим самим законом розподілу, то значення x_i і y_i взаємозамінні. Тобто не може бути. Що більшість значеннь однієї вибірки більше (менше) відповідних значень іншої вибіркиі. Отже, ймовірності появи позитивних і негативних різниць $x_i - y_i$ рівні. Якщо додатково зробити припущення про неперервність розподілу вимірюваної ознаки, то ймовірність появи нульових різниць дорівнює нулю. Таким чином,

$$\mathbf{P}\{x_i - y_i > 0\} = \mathbf{P}\{x_i - y_i < 0\} = 1/2,\ i = 1, 2, ..., l\ ,$$

де l – число ненульових різниць, $l \le n$.

Статистикою критерію є число знаків «+» у послідовності знаків різниць парних вибірок (x_i, y_i). Якщо гіпотеза H_1, що перевіряється, вірна, то пари (x_i, y_i) спостережень і, отже, знаки $x_i - y_i$ різниць незалежні, число знаків «+» має біноміальний розподіл з параметрами $p = 1/2$ і l.

Нехай r – кількість знаків «+», що спостерігаються, α – заданий рівень значущості і альтернативна гіпотеза сформульована як $H_2: p \ne 1/2$. Тоді гіпотеза H_1 відхиляється, якщо виконується одна з нерівностей

$$\sum_{i=r}^{l} C_l^i \left(\frac{1}{2}\right)^l \le \frac{\alpha}{2},\quad \sum_{i=0}^{r} C_l^i \left(\frac{1}{2}\right)^l \le \frac{\alpha}{2}.$$

Більш зручно проводити перевірку зазначеної гіпотези, використовуючи статистику Фішера. Гіпотеза H_1 відхиляється, якщо виконується одна з нерівностей

$$\frac{r}{l - r + 1} \ge F_{1-\alpha/2}(k_1, k_2),\ k_1 = 2(l - r + 1),\ k_2 = 2r,$$

$$\frac{l - r}{r + 1} \ge F_{1-\alpha/2}(k_1, k_2),\ k_1 = 2(r + 1),\ k_2 = 2(l - r).$$

Приклад 4.2.1 Критерій знаків (с.175-177, В.М. Турчин або с 73-74 Шпортюк-Оленко):
Розв'язання вручну(Остаматій)

Перша вибірка: 2, 6, 7, 7, 9,5,3, 3, 2, 1,.4,2
Друга вибірка: 1, 4, 2, 2, 3, 5, 7, 5, 7, 3, 3, 5.

Сформулюємо гіпотези:
H_0: $a_1 = a_2$ (різниця у вибірках зумовлена випадковими помилками, суттєвої систематичної відмінності нема)
H_1: $a_1 \ne a_2$ (оцінки одного з експертів завищені або занижені)

n=m=12

Гіпотеза H_1 відхиляється, якщо виконується нерівность Фішера:

$$F_1 = \frac{r}{k - r + 1} \ge F_{1-\frac{\alpha}{2},\left(2(t-r+1), 2r\right)}$$ (альтернатива p>1/2)

або

$$F_2 = \frac{k - r}{r + 1} \ge F_{1-\frac{\alpha}{2},\left(2(r+1), 2(k-r)\right)}$$ (альтернатива p<1/2)

Складемо таблицю знаків різниць відповідних пар:

2	6	7	7	9	5	3	3	2	1	4	2
1	4	2	2	3	5	7	5	7	3	3	5
+	+	+	+	+	0	-	-	-	-	+	-

Проведемо підрахунок для нашої задачі:
Число нульових: 1
Число ненульових: k=11
Число додатних : r=6
$\alpha = 0{,}05$

Отже,
H_0: $a_1 = a_2$ (різниця у вибірках зумовлена випадковими помилками, суттєвої систематичної відмінності нема ймовірність зустріти + та – однакова, p=1/2)
H_1: $a_1 \neq a_2$

Використаємо другу нерівність
Емпіричне значення;
$$F_2 = \frac{11-6}{5+1} = \frac{5}{6} = 0{,}83$$
Теоретичне

$$F_{1-\frac{\alpha}{2},\left(2(r+1),2(k-r)\right)} = F_{1-0{,}05,(14,10)} = F_{0{,}05,(14,10)} \approx 2.85$$

Отже, маємо що 0,83<2,85. Це свідчить про те, що підстав відхиляти гіпотезу H_0 немає. Оцінки другого експерту не є завищеними. Можна ще перевірити, що не є заниженими.

В) критерій Вілкоксона; (Статистика – сума рангів вибірки меншого обсягу – характеризує міру «змішаності» вибірок), Спочатку формуємо загальний ряд з двох вибірок у порядку зростання.

критерій Вілксона(Турчин, с182-184)

1. Ручне розв'язання

Перша вибірка: 20,26,27.27.30.32,33,34,35,36.
Друга вибірка: 21,21,22.23.25,25, 25,25, 27,27,29,31.

n=10

m=12

Записуємо в один ряд, в зростаючому порядку елементи з обох вибірок

20 21 21 22 23 25 25 25 25 26 27 27 27 27 29 30 31 32 33 34 35 36

W – сума рангів вибірки меншого обсягу (емпіричне значення статистики)

W = 1+10+(11+12+13+14)/4*2+16+18+19+20+21+22=152

Теоретичне значення

$$w_{\alpha,n,m} = \frac{1}{2}n(n+m+1) + z_\alpha\sqrt{\frac{1}{12}nm(n+m+1)} = \frac{1}{2}\cdot 10\cdot 23 - 1{,}96\cdot\sqrt{\frac{1}{12}*120*23} = 85$$

Де $z_\alpha = z_{0,025} = -1{,}96$ - квантильнормального розподілу з параметрами (0,1).

Знайдемо теоретичний інтервал прийняття гіпотези

$$(w_{\alpha,n,m}, n(n+m+1) - w_{\alpha,n,m})$$

$$n(n+m+1) - w_{\alpha,n,m} = 10*23 - 85{,}27513 = 144{,}7249$$

Отже, інтервал (85; 144,7)

Гіпотеза про однорідність не приймається, тому що емпірична статистика не належить теоретичному інтервалу прийняття гіпотези. Отже, між показниками двох вибірок про ефективність раціонів годування скотини є суттєві розходження .

Г) критерій Мана-Уітні;

Цей метод виявлення відмінностей між вибірками був запропонований у 1945 році Френком Уілкоксоном (F. Wilcoxon). У 1947 році він був істотно перероблений і розширений Х. Б. Манном (H. B. Mann) і Д. Р. Уітні (D. R. Whitney), на честь яких сьогодні зазвичай і називається.

Простий непараметричний критерій. Цей метод визначає, чи досить мала зона значень, що перехрещуються, між двома рядами (ранжируваним рядом значень параметра в першій вибірці і таким же в другій вибірці). Чим менше значення критерію, тим вірогідніше, що відмінності між значеннями параметра у вибірках достовірні.

Обмеження застосованості критерію

1. У кожній з вибірок повинно бути не менше 3 значень ознаки. Допускається, щоб в одній вибірці були два значення, але в другій тоді не менше п'ять.
2. У вибіркових даних не повинно бути збігів (усі числа — різні) або таких збігів повинно бути дуже мало.

Алгоритм застосування U -критерію Манна — Уітні:

1. Скласти єдиний ранжируваний ряд з обох вибірок, що зіставляються, розставивши їхні елементи по мірі наростання ознаки і приписавши меншому значенню менший ранг. Загальна кількість рангів вийде рівною:

   $N = n_1 + n_2$,
   де n_1 — кількість одиниць в першій вибірці, а n_2 — кількість одиниць в другій вибірці.
2. Розділити єдиний ранжируваний ряд на два, що складаються відповідно з одиниць першої і другої вибірок. Підрахувати окремо суму рангів, що припали на долю елементів першої вибірки, і окремо — на долю елементів другої вибірки. Визначити **більшу** з двох рангових сум (T_x), таку, що відповідає вибірці з n_x одиниць.
3. Визначити значення U -критерію Манна — Уітні за формулою:

$$U = n_1 \cdot n_2 + \frac{n_x \cdot (n_x + 1)}{2} - T_x.$$

4. За таблицею для обраного рівня статистичної значущості визначити критичне значення критерію для даних n_1 і n_2. Якщо набуте значення *U* **менше** табличного або дорівнює йому, то визнається наявність істотної відмінності між рівнем ознаки в даних вибірках (приймається альтернативна гіпотеза). Якщо ж набуте значення *U* більше за табличне, приймається нульова гіпотеза. Достовірність відмінностей тим вище, чим менше значення *U*.
5. При справедливості нульової гіпотези критерій має математичне сподівання $M(U) = \frac{n_1 \cdot n_2}{2}$ і дисперсію $D(U) = \frac{n_1 \cdot n_2 \cdot (n_1 + n_2)}{12}$ і при достатньо великому об'ємі вибіркових даних $(n_1 > 19,\ n_2 > 19)$ розподілений практично нормально.

Приклад критерій Мана-Уітні.

Перша вибірка: 20,26,27.27.30.32,33,34,35,36.
Друга вибірка: 21,21,22.23.25,25, 25,25, 27,27,29,31.

n=10

m=12

Записуємо в один ряд, в зростаючому порядку елементи з обох вибірок

20 21 21 22 23 25 25 25 25 26 27 27 27 27 29 30 31 32 33 34 35 36

Для кожного елемента шукаємо його ранг.

20	1
21	(2+3)/2=2,5
21	2,5
22	4
23	5
25	(6+7+8+9)/4 =7,5
25	7,5
25	7,5
25	7,5
26	10
27	12,5
27	12,5
27	12,5
27	12,5
29	15
30	16
31	17
32	18
33	19
34	20
35	21

36 22

Виписуємо в два ряди разом із рангами

№	Вибірка 1	Ранг 1	Вибірка 2	Ранг2
1	20	1	21	2,5
2	26	10	21	2,5
3	27	12,5	22	4
4	27	12,5	23	5
5	30	16	25	7,5
6	32	18	25	7,5
7	33	19	25	7,5
8	34	20	25	7,5
9	35	21	27	12,5
10	36	22	27	12,5
			29	15
			31	17
Суми		152		101

Позначимо найбільшу суму, як $T_x=152$
Найбільшу вибірку позначимо, як $n_x=12$.

Тоді використовуючи формулу $U = n_1 \cdot n_2 + \frac{n_x \cdot (n_x+1)}{2} - T_x.$

$U = 10 * 12 + \frac{12 \cdot 13}{2} - 152 = 23$

При $p \le 0.05$ для $n=10$, $m=12$, за табличкою $U_{Кр}=29$.
Отже, U менше за табличне значення, тому нульова гіпотеза не приймається.

Знайдемо математичне сподівання і дисперсію.

$M(U) = \frac{n_1 \cdot n_2}{2}$ $\quad D(U) = \frac{n_1 \cdot n_2 \cdot (n_1 + n_2)}{12}$

M(U)=60; D(U)=220.