ИЗСЛЕДВАНЕ НА ДВЕ ИЗВАДКИ

§32. Зависими и независими извадки. Изследване на две зависими извадки.

В много случаи с помощта на статистически методи се търси оценка на различията между две отделни групи от обществото, на ефекта от прилагането на даден метод за лечение или технологично нововъедение. Например, въздействието на дадено лекарство се изследва с два подхода:

1) Взимат се две отделни съвкупности от доброволци като на едните се дава новото лекарство, а на другите се дава безобидно лекарство, без истински ефект, чиято роля е само психологическа.
2) Изследва се една и съща съвкупност от доброволци, като се отчитат резултатите преди и след приемане на лекарството.

Двата подхода се различават главно по избора на обектите на наблюденията: в първия случай наблюденията се извършват върху две случайни извадки, които нямат общ елемент помежду си. Извадките могат да бъдат от една или от две различни генерални съвкупности и да имат различен обем. Такива извадки се наричат ***независими*** и те ще бъдат разгледани в **§33**.

При втория подход наблюденията се извършват върху елементите от една и съща съвкупност или върху елементите на зависими съвкупности. Такива извадки се наричат ***зависими.*** Получените съвкупности от наблюдавани стойности $(x_1,\ldots,x_n)$ и $(y_1,\ldots,y_n)$ се разглеждат като наблюдавани стойности на различни признаци X и Y на генералната съвкупност.

Зависимите извадки имат еднакъв обем и се изследват като се съпоставят резултатите от наблюденията x_i и y_i на i-тия обект, при изучаваните условия (напр. преди и след прилагането на лекарството). За целта се разглеждат разликите $d_i = x_i - y_i$, чрез които се получава се нова съвкупност от данни. Прилагайки различни статистически методи за данните $(d_1,\ldots,d_n)$ – точкови оценки, доверителни интервали и хипотези се правят оценки и изводи за влиянието на новите условия върху изучаваната съвкупност от обекти, а от там и на цялата генерална съвкупност.

Ще разгледаме случая, когато X и Y са ***нормално разпределени*** с математически очаквания $EX=\mu_x$ и $EY=\mu_y$ и дисперсии $DX=\sigma_x^2$ и $DY=\sigma_y^2$. Разликата $D=X-Y$ също има нормално разпределение и затова може да приложим методите за изследване на нормално разпределени признаци.

Пример 32.1. Изучава се влиянието на лекарство върху диастоличното кръвно налягане. Измерено е диастоличното кръвно

налягане на 8 пациента преди и след приемането на лекарството. Получени се следните данни

x_i	93	106	87	102	95	88	110	92
y_i	92	102	89	101	96	88	105	92.
$d_i = x_i - y_i$	1	4	−2	1	−1	0	5	0

а) да се намери доверителен интервал за средното изменение на диастоличното кръвно налягане с доверителна вероятност $\gamma = 0{,}99$;

б) Може ли да се твърди на базата на наличните данни, че лекарството не влияе на диастоличното кръвно налягане, ако приемем ниво на значимост $\alpha = 0{,}01$?

Решение. Изчисляваме: $\overline{d} = \frac{1}{8}(1+4-2+1-1+0+5+0) = 1$.

$$\widetilde{s}_d^2 = \frac{1}{n-1}\sum (d_i - \overline{d})^2 = 5{,}71, \quad \widetilde{s}_d = \sqrt{\widetilde{s}_d^2} = \sqrt{5{,}71} = 2{,}39 .$$

а) $\overline{d} = 1$ е точкова оценка за разликата $\mu_x - \mu_y$, а доверителният интервал за тази разлика ще намерим по формула (28.2):

1) намираме квантила от ред $(1+\gamma)/2 = 0{,}995$ на t-разпределението с $n-1 = 8-1 = 7$ степени на свобода: $t_{0,995}(7) = 3{,}50$.

2) Представителната грешка е $\delta = t_{0,995}(7)\frac{\widetilde{s}_d}{\sqrt{n}} = 3{,}50\frac{2{,}39}{\sqrt{8}} = 2{,}957$.

3) Следователно, с вероятност 0,99 разликата $\mu_x - \mu_y$ е число от интервала $(1-2{,}957; 1+2{,}957)$ = $(-1{,}957; 3{,}597)$;

б) Проверяваме хипотезата $H_0 = \{\mu_x = \mu_y\}$, т.е. $\mu_x - \mu_y = 0$, при ниво на значимост 0,01. Ако конкуриращата хипотеза е $H_1 = \{\mu_x \neq \mu_y\}$, то можем да използваме получения в а) резултат. Наистина, тъй като $0 \in (-1{,}957; 3{,}957)$ с вероятност 0,99, то вероятността да приемем вярна хипотеза H_0 е 0,99, а вероятността да отхвърлим вярна хипотеза H_0 ще бъде $1-0{,}99 = 0{,}01$, което съвпада с нивото на значимост. Следователно, може да приемем с ниво на значимост 0,01, че лекарството не влияе на диастоличното кръвно налягане (до същия извод се достига, и ако се приложи проверка на хипотезата H_0 по правилата, описани в **§31**).

Пример.32.2. Изследване, базирано на две зависими извадки с обем 17, има за цел да отхвърли или приеме нулевата хипотеза за равенство на генералните средни ($H_0 = \{\mu_x = \mu_y\}$) при конкурираща хипотеза $H_1 = \{\mu_x > \mu_y\}$. Намерено е, че средната на разликите d_i е $\overline{d} = 5{,}7$, а поправеното средно квадратично отклонение на разликите е

$\widetilde{s}_d = 4{,}8$. Достатъчни ли са наличните данни за отхвърляне на нулевата хипотеза при ниво на значимост 0,05?

Решение. Проверяваме хипотезата $H_0 = \{\mu_d = 0\}$ при $H_1 = \{\mu_d > 0\}$:

1) Наблюдаваната стойност е $t_{набл.} = \dfrac{\overline{d} - 0}{\widetilde{s}_d / \sqrt{n}} = \dfrac{5{,}7.\sqrt{17}}{4{,}8} = 4{,}896$.

2) Критичната област се определя по формула (31.2), съгласно която изчисляваме $t_{кр.} = t_{1-\alpha}(n-1) = t_{0,95}(16) = 1{,}75$. Тъй като $t_{набл.} > t_{кр.}$, то данните са достатъчни за отхвърляне на нулевата хипотеза.

Упражнение.

1. Сравнява се времето, необходимо за изпълнение на програми, написани на два различни програмни езика. За 12 програми времето (в минути) за изпълнение е следното:

x_i	17	16	21	14	18	24	13	23	14	16	18	21
y_i	18	14	19	11	23	21	15	24	13	10	20	19

Да се намери доверителен интервал за средната разлика $d_i = x_i - y_i$ с доверителна вероятност $\gamma = 0{,}95$. Може ли да се направи извод за по-голяма ефективност на един от двата езика?

§33. Критерии за сравняване на дисперсиите и математическите очаквания на две независими нормално разпределени съвкупности.

Нека за изследване на зависимостите между признаците X и Y на две генерални съвкупности са дадени независимите извадки $(x^{(1)},\dots,x^{(n_x)})$ и $(y^{(1)},\dots,y^{(n_y)})$ с обеми съответно n_x и n_y (разглеждаме негрупирани извадки с цел опростяване на означенията).

Сравняването на признаците X и Y се извършва чрез:

- Диаграмите тип „кутия" на двете извадки за сравняване на общите закономерности (виж пример **33.1**);
- Разликата на извадъчните средни $\overline{x} - \overline{y}$ за сравняване на генералните средни $EX = \mu_x$ и $EY = \mu_y$;
- Отношенията на дисперсиите им $\widetilde{s}_x^2$ и $\widetilde{s}_y^2$ за сравняване на дисперсиите $DX = \sigma_x^2$ и $DY = \sigma_y^2$.

.**1. Сравняване на дисперсии.** Нека за изследване на величините $X \sim N(\mu_x, \sigma_x)$ и $Y \sim N(\mu_y, \sigma_y)$ разполагаме с две независими извадки като ***с X е означена тази величина, за която*** $\widetilde{s}_x^2 > \widetilde{s}_y^2$. За съпоставяне на дисперсиите на величините се използва, че отношението $\dfrac{\widetilde{s}_x^2/\sigma_x^2}{\widetilde{s}_y^2/\sigma_y^2}$ има разпределение на Фишер и се извършва по следния начин:

Намиране на доверителни интервали за отношението $\dfrac{\sigma_x^2}{\sigma_y^2}$ на $X \sim N(\mu_x, \sigma_x)$ и $Y \sim N(\mu_y, \sigma_y)$ с доверителна вероятност γ :

1) μ_x ***и*** μ_y ***са известни***: $\dfrac{\sigma_x^2}{\sigma_y^2} \in \left(\dfrac{s_{0x}^2 / s_{0y}^2}{F_1}, \dfrac{s_{0x}^2 / s_{0y}^2}{F_2} \right)$, *където* $s_{0x}^2 > s_{0y}^2$ *са:*

$$s_{0x}^2 = \frac{1}{n_x - 1} \sum (x^{(i)} - \mu_x)^2 \,,\; s_{0y}^2 = \frac{1}{n_y - 1} \sum (y^{(i)} - \mu_y)^2 \qquad (33.1)$$

$F_1 = F_{\frac{1+\gamma}{2}}(n_x, n_y)$, $F_2 = F_{\frac{1-\gamma}{2}}(n_x, n_y)$ ***-*** *квантили на F-разпределението с степени на свобода n_x и n_y.*

2) μ_x ***и*** μ_y ***са неизвестни***: $\dfrac{\sigma_x^2}{\sigma_y^2} \in \left(\dfrac{\widetilde{s}_x^2 / \widetilde{s}_y^2}{F_1}, \dfrac{\widetilde{s}_x^2 / \widetilde{s}_y^2}{F_2} \right)$, ($\widetilde{s}_x^2 > \widetilde{s}_y^2$) (33.2)

където: $F_1 = F_{\frac{1+\gamma}{2}}(n_x - 1, n_y - 1)$, $F_2 = F_{\frac{1-\gamma}{2}}(n_x - 1, n_y - 1)$ *са квантили на F-разпределението със степени на свобода* $n_x - 1$ *и* $n_y - 1$.

Забележка 33.1. ***Обикновено са съставени таблици за критичните точки на F-разпределението, като квантилите и критичните точки са свързани по формули (14.2) и (14.3).***

Пример 33.1. Изследва се диаметърът на детайли, произведени от две машини. От произведените детайли са получени две извадки с обеми $n_x = 25$ и $n_y = 13$ и са изчислени, $\widetilde{s}_x^2 = 0{,}35$, $\widetilde{s}_y^2 = 0{,}40$. Да се намери 90% доверителен интервал за отношението σ_y^2 / σ_x^2.

Решение. . 1) Изчисляваме $\dfrac{\widetilde{s}_y^2}{\widetilde{s}_x^2} = \dfrac{0{,}40}{0{,}35} = 1{,}1428$ (тук $\widetilde{s}_y^2 > \widetilde{s}_x^2$)

2) На стр.170 са дадени *критичните точки* на *F*-разпределението за ниво на значимост 0,05, затова (виж формули (14.2) и (14.3)):

$$F_1 = F_{\frac{1+\gamma}{2}}(n_y - 1, n_x - 1) = F_{\frac{1,9}{2}}(13-1, 25-1) = F_{0,95}(12,24) = F_{0,05}^{kp}(12,24) = 2{,}18$$

$$F_2 = F_{\frac{1-\gamma}{2}}(n_y - 1, n_x - 1) = F_{0,05}(12,24) = \frac{1}{F_{0,05}^{kp}(24,12)} = \frac{1}{2{,}51} = 0{,}398$$

По формула (33.2) получаваме $\dfrac{\sigma_y^2}{\sigma_x^2} \in \left(\dfrac{1{,}1428}{2{,}18}, \dfrac{1{,}1428}{0{,}40} \right) = (0{,}524; 2{,}857)$ ♦

В зависимост от това дали μ_x и μ_y са известни или не, се определят следните правила за проверка на хипотезата за равенство на дисперсиите на две независими извадки:

Проверка на хипотезата $H_0=\{\sigma_x^2=\sigma_y^2\}$ при ниво на значимост α:

1) μ_x ***и*** μ_y ***са известни***:

- *Наблюдавана стойност:* $F_{набл.}=\dfrac{s_{0x}^2}{s_{0y}^2}$ *като* $s_{0x}^2>s_{0y}^2$ (формули (33.1)).
- *Критична област:*
  $F_{набл.}>F_{кр.}$ *при* $H_1=\{\sigma_x^2>\sigma_y^2\}$, *където* $F_{кр.}=F_{1-\alpha}(n_x,n_y)$.
  $F_{набл.}>F_{кр.}$ *при* $H_1=\{\sigma_x^2\neq\sigma_y^2\}$, *където* $F_{кр.}=F_{1-\frac{\alpha}{2}}(n_x,n_y)$.

2) μ_x ***и*** μ_y ***са неизвестни:***

- *Наблюдавана стойност:* $F_{набл.}=\dfrac{\tilde{s}_x^2}{\tilde{s}_y^2}$ *като* ($\tilde{s}_x^2>\tilde{s}_y^2$).
- *Критична област:*
  $F_{набл.}>F_{кр.}$ *при* $H_1=\{\sigma_x^2>\sigma_y^2\}$, *където* $F_{кр.}=F_{1-\alpha}(n_x-1,n_y-1)$
  $F_{набл.}>F_{кр.}$ *при* $H_1=\{\sigma_x^2\neq\sigma_y^2\}$, *където.* $F_{кр}=F_{1-\frac{\alpha}{2}}(n_x-1,n_y-1)$

За изследване на разликата $\mu_x-\mu_y$ за нормално разпределени величини $X\sim N(\mu_x,\sigma_x)$ и $Y\sim N(\mu_y,\sigma_y)$, се използват Z и *t*- разпределения по следния начин:

Доверителни интервали за разликата $\mu_x-\mu_y$:

1) σ_x^2 ***и*** σ_y^2 ***-известни:***

$$\mu_x-\mu_y\in\left(\bar{x}-\bar{y}-Z_{\frac{1+\gamma}{2}}\sigma,\ \bar{x}-\bar{y}+Z_{\frac{1+\gamma}{2}}.\sigma\right),\qquad \text{където } \sigma=\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}.$$

2) $\sigma_x^2=\sigma_y^2$, ***но не са известни***:

$$\mu_x-\mu_y\in\left(\bar{x}-\bar{y}-t.\sqrt{s^2\left(\frac{1}{n_x}+\frac{1}{n_y}\right)},\ \bar{x}-\bar{y}+t.\sqrt{s^2\left(\frac{1}{n_x}+\frac{1}{n_y}\right)}\right),$$

където $s^2=\dfrac{(n_x-1)\tilde{s}_x^2+(n_y-1)\tilde{s}_y^2}{n_x+n_y-2}$, $t=t_{\frac{\gamma+1}{2}}(n_x+n_y-2)$

3) $\sigma_x^2\neq\sigma_y^2$ ***-неизвестни:*** $\mu_x-\mu_y\in\left(\bar{x}-\bar{y}-t.\sqrt{\dfrac{\tilde{s}_x^2}{n_x}+\dfrac{\tilde{s}_y^2}{n_y}},\bar{x}-\bar{y}+t.\sqrt{\dfrac{\tilde{s}_x^2}{n_x}+\dfrac{\tilde{s}_y^2}{n_y}}\right)$,

където $t=t_{\frac{\gamma+1}{2}}(k^*)$, k^* *е цялата част на израза* $\dfrac{\left[(\tilde{s}_x^2/n_x)+(\tilde{s}_y^2/n_y)\right]^2}{\dfrac{(\tilde{s}_x^2/n_x)^2}{n_x-1}+\dfrac{(\tilde{s}_y^2/n_y)^2}{n_y-1}}$.

Проверка на хипотезата $H_1=\{\mu_x=\mu_y\}$ при ниво на значимост α :

1) σ_x^2 ***и*** σ_y^2 ***-известни:***

- *Наблюдавана стойност:* $Z_{набл.}=\dfrac{\bar{x}-\bar{y}}{\sigma}$, $\sigma=\sqrt{\dfrac{\sigma_x^2}{n_x}+\dfrac{\sigma_y^2}{n_y}}$
- *Критична област:*
  $Z_{набл.}>Z_{кр.}$ *при* $H_1=\{\mu_x>\mu_y\}$, *където* $Z_{кр.}=Z_{1-\alpha}$
  $Z_{набл.}<Z_{кр.}$ *при* $H_1=\{\mu_x<\mu_y\}$, *където* $Z_{кр.}=-Z_{1-\alpha}$
  $|Z_{набл.}|>Z_{кр.}$ *при* $H_1=\{\mu_x\neq\mu_y\}$, *където* $Z_{кр.}=Z_{1-\frac{\alpha}{2}}$

2) $\sigma_x^2=\sigma_y^2$ ***- неизвестни и хипотезата*** $H_0'=\{\sigma_x^2=\sigma_y^2\}$ ***се приема :***

- *Наблюдавана стойност:*

$$t_{набл.}=\frac{\bar{x}-\bar{y}}{\sqrt{s^2\left(\frac{1}{n_x}+\frac{1}{n_y}\right)}},\quad s^2=\frac{(n_x-1)\tilde{s}_x^2+(n_y-1)\tilde{s}_y^2}{n_x+n_y-2}\qquad (33.2)$$

- *Критична област:*
  $t_{набл.}>t_{кр.}$ *при* $H_1=\{\mu_x>\mu_y\}$, *където* $t_{кр.}=t_{1-\alpha}(n_x+n_y-2)$
  $t_{набл.}<t_{кр.}$ *при* $H_1=\{\mu_x<\mu_y\}$, *където* $t_{кр.}=-t_{1-\alpha}(n_x+n_y-2)$
  $|t_{набл.}|>t_{кр.}$ *при* $H_1=\{\mu_x\neq\mu_y\}$, *където* $t=t_{\frac{\gamma+1}{2}}(n_x+n_y-2)$

3) $\sigma_x^2\neq\sigma_y^2$ ***неизвестни и хипотезата*** $H_0'=\{\sigma_x^2=\sigma_y^2\}$ ***се отхвърля:***

- *Наблюдавана стойност:* $t_{набл.}=\dfrac{\bar{x}-\bar{y}}{\sqrt{\dfrac{\tilde{s}_x^2}{n_x}+\dfrac{\tilde{s}_y^2}{n_y}}}$,
- *Критична област:*
  $t_{набл.}>t_{кр.}$ *при* $H_1=\{\mu_x>\mu_y\}$, *където* $t_{кр.}=t_{1-\alpha}(k^*)$
  $t_{набл.}<t_{кр.}$ *при* $H_1=\{\mu_x<\mu_y\}$, *където* $t_{кр.}=-t_{1-\alpha}(k^*)$
  $|t_{набл.}|>t_{кр.}$ *при* $H_1=\{\mu_x\neq\mu_y\}$, *където* $t_{кр.}=t_{1-\frac{\alpha}{2}}(k^*)$

  Тук k^* *е цялата част на израза* $\dfrac{\left[(\tilde{s}_x^2/n_x)+(\tilde{s}_y^2/n_y)\right]^2}{\dfrac{(\tilde{s}_x^2/n_x)^2}{n_x-1}+\dfrac{(\tilde{s}_y^2/n_y)^2}{n_y-1}}$.

Забележка 33.2. ***Напомняме отново:тъй като има таблици за критичните точки на F-разпределението за стойности, по-***

големи от единица, то приемаме, че в горните формули с X е означена тази величина, за която $\tilde{s}_x^2 > \tilde{s}_y^2$.

На практика сравняване на параметрите на две независими извадки се налага при сравняване на два метода на работа, на две технологии и т.н.

Пример 33.1. Предложена е нова технология за производство на лампи; за която се твърди, че обеспечава по-голямо светлоодаване и по-малко отклонение от средното светлоотдаване. За проверка на хипотезите са взети 8 лампи, произведени по старата технология, стойностите на светлоотдаване на които са (в лм/вт):

62, 73, 80, 79, 63, 77, 81, 75

и 9 лампи по новата технология, със светлоотдаване:

75, 78, 77, 68, 73, 79, 72, 71, 86.

а) да се сравнят диаграмите от тип бокс-плот;
б) да се проверят хипотезите с ниво на значимост $a=0{,}01$.

Решение.

а) Намираме вариационните редове:

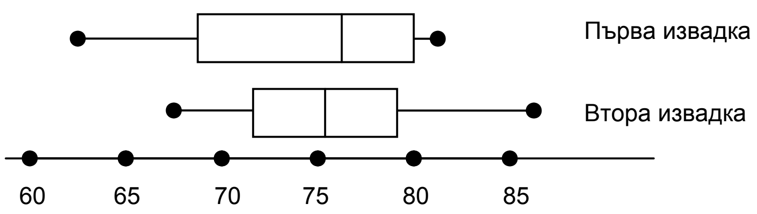
Първа извадка: 62, 63, 73, 75, 77, 79, 80, 81,

откъдето $Q_1=\dfrac{63+73}{2}=68,\ Q_2=\dfrac{75+77}{2}=76,\ Q_3=\dfrac{79+80}{2}=79{,}5$.

Втора извадка: 68, 71, 72, 73, 75, 77, 78, 79, 86,

откъдето: $Q_1'=72,\ Q_2'=75,\ Q_3'=78$.

Начертаваме диаграмите от тип бокс-плот:



Изводи: Медианите (които са оценки за средната стойност на величините) са близки $\Rightarrow$ средните стойности не се различват съществено.

$Q_3'-Q_1'=78-72=6$, $\quad Q_3-Q_1=79{,}5-68=11{,}5 \quad \Rightarrow$ отклоненията от средните стойности на втората извадка са по-малки от тези на първата.

б) Нека X е светлоотдаването на лампа, произведена по старата технология. Тогава EX е средното светлоотдаване на лампите, а σX е отклонението от средното светлоодаване. Аналогично, ако Y е светлоотдаването на лампа, произведена по новата технология, то EY и σY са съответно средното светлоодаване и разсейването около него на прозведените по втората технология лампи.

От извадките изчисляваме оценките на EX , σX , EY и σY :

$\bar{x}=73{,}75$, $\tilde{s}_x^2=55{,}07$; $\bar{y}=75{,}44$, $\tilde{s}_y^2=29{,}03$.

Съгласно условието на задачата, издигнати са хипотезите: при производството по втората технология имаме по-добро светлоодваване т.е. $EY>EX$ и по-малко разсейване, т.е. $\sigma^2 Y<\sigma^2 X$. Тъй като нулевата хипотеза трябва да е проста, то при ниво на значимост $\alpha=0{,}01$ ще проверим хипотезите $H_0=\{\sigma_x^2=\sigma_y^2\}$ срещу $H_1=\{\sigma_x^2>\sigma_y^2\}$ и

$H_0=(EX=EY\}$ срещу $H_1=\{EX<EY\}$.

- Проверка на хипотезата $H_0=\{\sigma_x^2=\sigma_y^2\}$ срещу $H_1=\{\sigma_x^2>\sigma_y^2\}$ и ниво на значимост $\alpha=0{,}01$:

1) Наблюдаваната стойност е

$$F_{набл.}=\frac{по-голяма\ дисперсия}{по-малка\ дисперсия}=\frac{\tilde{s}_x^2}{\tilde{s}_y^2}=\frac{55{,}07}{29{,}03}=1{,}897 .$$

2) При дадената конкурираща хипотеза критичната стойност определяме като критична точка от ред $\alpha=0{,}01$ на разпределението $F(n_x-1,n_y-1)=F(7,8)$, т.е. $F_{кр.}=F_{0{,}01}^{(кр)}(7,8)=6{,}18$.

3) Тъй като $F_{набл.}<F_{кр.}$, то хипотезата за равенство на дисперсиите се приема.

- Проверка на хипотезата $H_0=(EX=EY\}$ срещу $H_1=\{EX<EY\}$.

1) Тъй като се установи, че $H_0=\{\sigma_x^2=\sigma_y^2\}$ не може да се отхвърли на базата на наличните данни, то наблюдаваната стойност на критерия изчисляваме по формула (33.2). Изчисляваме обединената дисперсия

$$s^2=\frac{(n_x-1)\tilde{s}_x^2+(n_y-1)\tilde{s}_y^2}{n_x+n_y-2}=\frac{7.55{,}07+8.29{,}03}{8+9-2}=41{,}162 ,$$

откъдето $t_{набл.}=\dfrac{\bar{x}-\bar{y}}{\sqrt{s^2\left(\dfrac{1}{n_x}+\dfrac{1}{n_y}\right)}}=\dfrac{73{,}75-75{,}44}{\sqrt{41{,}162.\left(\dfrac{1}{8}+\dfrac{1}{9}\right)}}=-0{,}542$.

2) Критичната област определяме съгласно конкуриращата хипотеза като лявостранна област $D=\{t_{набл.}<t_{кр.}\}$, т.е. област, за която $P(t_{набл.}<t_{кр.})=\alpha$. Изчисляваме: $t_{кр.}=-t_{1-\alpha}(n_x+n_y-2)=t_{0{,}99}(15)=-2{,}60$.

3) Тъй като $t_{набл.}>t_{кр.}$, то хипотезата за равенство на средните стойности също се потвърждава.

Следователно, на основа на наличните данни не може да се приеме, че новата технология ще подобри параметрите на произвежданите лампи.♦

Упражнения:

1. От продукцията на две автоматични линии са взети извадки от n_1=16 и n_2=12 детайли и са изчислени $\overline{x}_1$=180 мм и $\overline{x}_2$=186 мм. От предварителен анализ е установено, че неточността на изработката им се описва с нормално разпределени величини с дисперсии съответно $\sigma_1^2 = 6$ мм^2 и $\sigma_2^2 = 11$ мм^2. С ниво на значимост α=0,025 да се провери хипотезата H_0: $\mu_1 = \mu_2$ срещу хипотезата H_1: $\mu_1 < \mu_2$.

2. Предприятие има два автомата за производство на изделия, диаметърът на които се изследва. Взети са по 100 изделия за проверка. Средният диаметър на изработените от първия автомат изделия е 10,3 със стандартна грешка 1,2, а за изработените от втория – съответно 9,8 и 1,6. С ниво на значимост 0,01 да се проверят хипотезите: а) Точността на изработка на двата автомата е еднаква. б) Двата автомата произвеждат изделия с еднакви диаметри.

3. Две съвкупности от обекти, на които се изучава количественият признак X с нормално разпределение, са подложени на контрол. От първото множество е направена извадка с обем n_1=21 и е пресметната поправената дисперсия, която е равна на 0,75. От второто множество е получена извадка с обем n_2=11 с поправена дисперсия 0,25. Проверете хипотезата H_0: $DX_1=DX_2$. при конкурираща хипотеза H_1 :$DX_1 \neq DX_2$ и ниво на значимост 0,1.

4. При измерване на една и съща величина X , са получени следните резултати $x_1 = -4,\ x_2 = 0,\ x_3 = -2,\ x_4 = 2,\ x_5 = 6$.а) Да се намерят $\overline{x}$ и $\tilde{s}_X$. б) Да се определи доверителният интервал за EX на случайната величина X с доверителна вероятност $\gamma = 0{,}95$. в) По друг метод от извадка с обем $n = 9$ е получена извадъчна дисперсия $\tilde{s}_Y^2 = 16$. Да се провери с ниво на значимост $\alpha = 0{,}02$ хипотезата за еднаква точност на двата метода, т.е. хипотезата $H_0 = \{DX = DY\}$. Като конкурираща хипотеза да се разгледа $H_1 = \{DX \neq DY\}$ (предполага се, че X има нормално разпределение).

5. За сравняване на две системи за изучаване на чужд език са избрани случайно по 10 отлични и 10 посредствени студенти. По първата система студентите са усвоили предадения материал както следва (в проценти):

посредствени студенти (извадка X_1): 67, 56, 55, 61, 67, 56, 68, 53, 66, 65;

отлични студенти (извадка X_2): 87, 78, 86, 90, 77, 78, 81, 91, 82, 75.

След това същите групи студенти са обучавани по втората система като се получени следните резултати:

посредствени студенти (извадка Y_1): 32, 41, 51, 34, 55, 36, 39, 45, 36, 40;

отлични студенти (извадка Y_2): 90, 88, 83, 85, 94, 91, 95, 87, 90, 83.

а) Кои от двойките извадки са зависими и кои са независими?

Да се намерят: доверителен интервал с доверителна вероятност 0,9 за разликата в степента на усвояване по двете системи а) за посредствените студенти; б) за отличните студенти. Коя от системите е по-подходяща за посредствени студенти и коя за отлично работещи студенти. в) Да се провери с ниво на значимост 0,01 хипотезата, че обучаваните по система 1 по-равномерно усвояват знанията на студентите (отлични и посредствени) отколкото обучаваните по втората система. г) Да се сравнят с доверителен интервал при $\gamma = 0{,}95$ средните постижения на всички студенти по система 1 и по система 2.