Vysoké učení technické v Brně
Fakulta informačních technologií

MSP - Štatistika a pravdepodobnosť

Cvičenie - streda, 08:00

# Projekt – zadanie 28, 14

8. decembra 2019 **Matúš Liščinský**, (xlisci02)

## Zadanie projektu:

**1. Pri kontrole výrobkov bola sledovaná odchýlka X [mm] ich rozmerov od požadovanej veľkosti. Namerané hodnoty tvoria štatistický súbor v liste Data_pr.1.**

(a) Urobte roztriedenie štatistického súboru, vytvorte tabuľku početností a nakreslite histogramy pre relatívnu početnosť a relatívnu komutatívnu početnosť.

(b) Vypočítajte aritmetický priemer, medián, modus, rozptyl a smerodajnú odchýlku.

(c) Vypočítajte bodové odhady strednej hodnoty, rozptylu a smerodajnej odchýlky.

(d) Testujte predpoklad o výbere z normálneho rozdelenia Pearsonovým (chi-kvadrát) testom na hladine významnosti $0,05$.

(e) Za predpokladu (bez ohľadu na výsledok časti d)), že štatistický súbor bol získaný náhodným výberom z normálneho rozdelenia, určte intervalové odhady strednej hodnoty, rozptylu a smerodajnej odchýlky so spoľahlivosťou $0,95$ a $0,99$.

(f) Testujte hypotézu optimálneho nastavenia stroja, tj. že stredná hodnota odchýlky je nulová, proti dvojstrannej alternatívnej hypotéze, že stredná hodnota odchýlky je rôzna od nuly, a to na hladine významnosti $0,05$.

(g) Overte štatistickým testom na hladine významnosti $0,05$, či nastavenie stroja ovplyvnilo kvalitu výrobu, ak viete, že vyššie uvedený štatistický súbor 50-tich hodnôt vznikol spojením dvoch čiastočných štatistických súborov, tak, že po nameraní prvých 20-tich hodnôt bolo vykonané nové nastavenie stroja a následne bolo nameraných zvyšných 30 hodnôt.

**Návod:** Oba súbory spracujte nezotriedené. Testujte najskôr rovnosť rozptylov odchýliek pred a po nastavení stroja. Podľa výsledkov potom zvoľte vhodný postup pre testovanie rovnosti stredných hodnôt odchýliek pred a po nastavení stroja.

**2. Nameraním dvojice (Výška[cm], Váha[kg]) u vybraných študentov z fakulty VUT FIT bol získaný dvojrozmerný štatistický súbor zapísaný po dvojiciach v riadkoch v liste Data_pr.2.**

(a) Vypočítajte bodový odhad koeficientu korelácie.

(b) Na hladine významnosti $0,05$ testujte hypotézu, že náhodné veličiny *Výška* a *Váha* sú lineárne nezávislé.

(c) **Regresná analýza** - dáta preložte priamkou: Váha $= \beta_1 + \beta_1 *$ Výška

a) Bodovo odhadnite $\beta_1$, $\beta_1$ a rozptyl $s^2$.

b) Na hladine významnosti $0,05$ otestujte hypotézy:

$$H : \beta_1 = -100, \quad H_A : \beta_1 \neq -100, \tag{1}$$
$$H : \beta_1 = 1, \quad H_A : \beta_1 \neq 1, \tag{2}$$

c) Vytvorte graf bodov spolu s regresnou priamkou a pásom spoľahlivosti pre individuálnu hodnotu výšky.

# Vypracovanie:

**1) Pri kontrole výrobkov bola sledovaná odchýlka X [mm] ich rozmerov od požadovanej veľkosti. Namerané hodnoty tvoria štatistický súbor v liste Data_pr.1, zadanie č. 28.**

**Štatistický súbor**

| $n$ | $x_n$ | $n$ | $x_n$ |
|---|---|---|---|
| 1 | -0,43 | 26 | 0,22 |
| 2 | -0,12 | 27 | -0,37 |
| 3 | 0,04 | 28 | 0,30 |
| 4 | -1,05 | 29 | 0,74 |
| 5 | 0,30 | 30 | -0,02 |
| 6 | 0,27 | 31 | 0,36 |
| 7 | 0,18 | 32 | -0,19 |
| 8 | 0,86 | 33 | 0,49 |
| 9 | 0,40 | 34 | 0,12 |
| 10 | 0,39 | 35 | 0,00 |
| 11 | -0,25 | 36 | 0,39 |
| 12 | -0,73 | 37 | 0,70 |
| 13 | 0,25 | 38 | 0,00 |
| 14 | -0,25 | 39 | -0,50 |
| 15 | -0,40 | 40 | 0,05 |
| 16 | 0,59 | 41 | -0,47 |
| 17 | -0,19 | 42 | -0,59 |
| 18 | 0,09 | 43 | 0,29 |
| 19 | -0,27 | 44 | -0,15 |
| 20 | -0,07 | 45 | 0,00 |
| 21 | 0,31 | 46 | 0,40 |
| 22 | 0,31 | 47 | -0,09 |
| 23 | -1,16 | 48 | 0,08 |
| 24 | 0,70 | 49 | 0,08 |
| 25 | 0,78 | 50 | 0,56 |

**Usporiadaný štatistický súbor**

| $n$ | $x_n$ | $n$ | $x_n$ |
|---|---|---|---|
| (1) | -1,16 | (26) | 0,08 |
| (2) | -1,05 | (27) | 0,09 |
| (3) | -0,73 | (28) | 0,12 |
| (4) | -0,59 | (29) | 0,18 |
| (5) | -0,50 | (30) | 0,22 |
| (6) | -0,47 | (31) | 0,25 |
| (7) | -0,43 | (32) | 0,27 |
| (8) | -0,40 | (33) | 0,29 |
| (9) | -0,37 | (34) | 0,30 |
| (10) | -0,27 | (35) | 0,30 |
| (11) | -0,25 | (36) | 0,31 |
| (12) | -0,25 | (37) | 0,31 |
| (13) | -0,19 | (38) | 0,36 |
| (14) | -0,19 | (39) | 0,39 |
| (15) | -0,15 | (40) | 0,39 |
| (16) | -0,12 | (41) | 0,40 |
| (17) | -0,09 | (42) | 0,40 |
| (18) | -0,07 | (43) | 0,49 |
| (19) | -0,02 | (44) | 0,56 |
| (20) | 0,00 | (45) | 0,59 |
| (21) | 0,00 | (46) | 0,70 |
| (22) | 0,00 | (47) | 0,70 |
| (23) | 0,04 | (48) | 0,74 |
| (24) | 0,05 | (49) | 0,78 |
| (25) | 0,08 | (50) | 0,86 |

---

*a) Urobte roztriedenie štatistického súboru, vytvorte tabuľku početností a nakreslite histogramy pre relatívnu početnosť a relatívnu komutatívnu početnosť.*

$x_{(1)} = \min_i x_i = -1,16$

$x_{(n)} = \max_i x_i = 0,86$

**Variačný obor:** $\langle x_{(1)}; x_{(n)} \rangle = \langle -1,16; 0,86 \rangle$

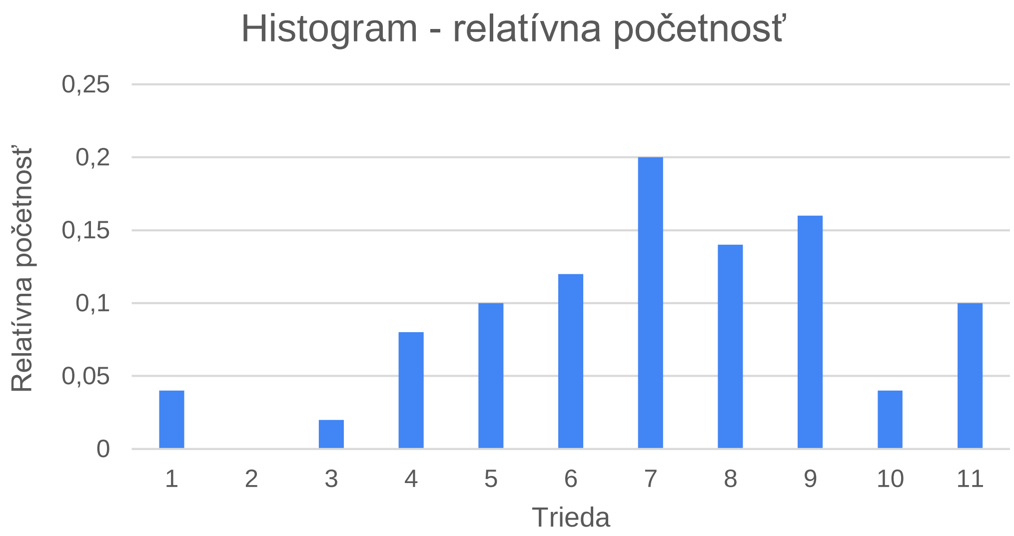
**Rozpätie:** $x_{(n)} - x_{(1)} = 2,02$

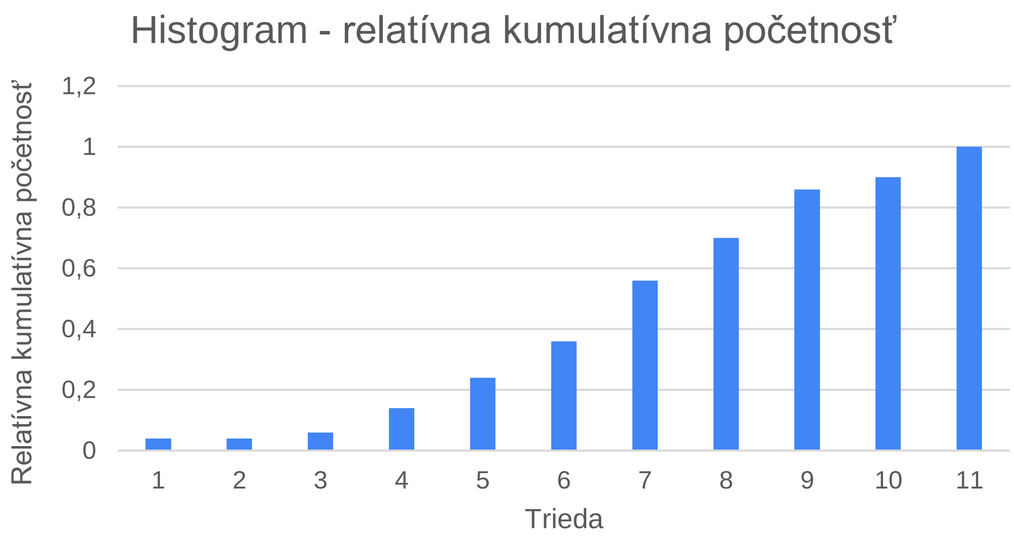
**Počet tried** $m = \dfrac{\sqrt{n} + 2\sqrt{n}}{2} \approx 11$

**Dĺžka triedy** $h = \dfrac{x_{(n)} - x_{(1)}}{m} = 0,183636364$

| trieda | xi- | xi+ | stred triedy | kp | p | rp | rkp |
|---|---|---|---|---|---|---|---|
| 1 | -1,1600 | -0,9764 | -1,0682 | 2 | 2 | 0,04 | 0,04 |
| 2 | -0,9764 | -0,7927 | -0,8845 | 2 | 0 | 0 | 0,04 |
| 3 | -0,7927 | -0,6091 | -0,7009 | 3 | 1 | 0,02 | 0,06 |
| 4 | -0,6091 | -0,4255 | -0,5173 | 7 | 4 | 0,08 | 0,14 |
| 5 | -0,4255 | -0,2418 | -0,3336 | 12 | 5 | 0,1 | 0,24 |
| 6 | -0,2418 | -0,0582 | -0,1500 | 18 | 6 | 0,12 | 0,36 |
| 7 | -0,0582 | 0,1255 | 0,0336 | 28 | 10 | 0,2 | 0,56 |
| 8 | 0,1255 | 0,3091 | 0,2173 | 35 | 7 | 0,14 | 0,7 |
| 9 | 0,3091 | 0,4927 | 0,4009 | 43 | 8 | 0,16 | 0,86 |
| 10 | 0,4927 | 0,6764 | 0,5845 | 45 | 2 | 0,04 | 0,9 |
| 11 | 0,6764 | 0,8600 | 0,7682 | 50 | 5 | 0,1 | 1 |

Tabuľka početností štatistického súboru, kde **kp** je *kumulatívna početnosť*, **p** je *početnosť*, **rp** je *relatívna početnosť* a **rkp** je *relatívna kumulatívna početnosť*.



*b) Vypočítajte aritmetický priemer, medián, modus, rozptyl a smerodajnú odchýlku.*

**Aritmetický priemer:** $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 0,059$

**Medián:** $\widetilde{x} = 0,08$

**Modus:** $\hat{x} = 0$

**Rozptyl:** $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = 0,18864258$

**Smerodajná odchýlka:** $s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} = 0,434330036$

*c) Vypočítajte bodové odhady strednej hodnoty, rozptylu a smerodajnej odchýlky.*

**Bodový odhad strednej hodnoty:** $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 0,059$

**Bodový odhad rozptylu:** $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = 0,200924490$

**Bodový odhad smerodajnej odchýlky:** $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} = 0,4482460148$

*d) Testujte predpoklad o výbere z normálneho rozdelenia Pearsonovým (chi-kvadrát) testom na hladine významnosti 0,05.*

Uvažujme hypotézu: $H : X \sim N(\mu, \sigma^2)$, kde $\mu, \sigma^2$ sú neznáme parametre. Využijeme ich bodové odhady z predchádzajúcej úlohy:

$$\overline{x} = \tfrac{1}{n}\sum_{i=1}^{n} x_i = 0,059 \qquad s^2 = \tfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = 0,200924490$$

Ďalej použijeme zoradený štatistický súbor z kroku **(a)**, kde sme štatistický súbor roztriedili do celkovo **11** tried nasledovne:

**Variačný obor:** $\langle x_{(1)}; x_{(n)}\rangle = \langle -1,16; 0,86\rangle$

**Rozpätie:** $x_{(n)} - x_{(1)} = 2,02$

**Dĺžka triedy** $h = \frac{x_{(n)} - x_{(1)}}{m} = 0,183636364$

K početnostiam spočítame teoretické početnosti $\hat{f}_j$ - tj. početnosti za podmienky, že by mala náhodná premenná $X$ distribučnú funkciu normálneho rozdelenia s parametrami $(\overline{x}, s^2)$.

| **trieda** | **xi-** | **xi+** | $f_j$ | $\hat{f}_j$ |
|---|---|---|---|---|
| 1 | $-\infty$ | -0,9764 | 2 | 0,5225 |
| 2 | -0,9764 | -0,7927 | 0 | 0,9129 |
| 3 | -0,7927 | -0,6091 | 1 | 1,9672 |
| 4 | -0,6091 | -0,4255 | 4 | 3,5923 |
| 5 | -0,4255 | -0,2418 | 5 | 5,5590 |
| 6 | -0,2418 | -0,0582 | 6 | 7,2902 |
| 7 | -0,0582 | 0,1255 | 10 | 8,1023 |
| 8 | 0,1255 | 0,3091 | 7 | 7,6313 |
| 9 | 0,3091 | 0,4927 | 8 | 6,0913 |
| 10 | 0,4927 | 0,6764 | 2 | 4,1204 |
| 11 | 0,6764 | $\infty$ | 5 | 4,2106 |
| **Σ** | | | **50** | **50** |

Následne potrebujeme zlúčiť triedy tak, aby boli splnené podmienky pre teoretické početnosti:
1) všetky teoretické početnosti majú byť **väčšie než 1**, a
2) aspoň **80%** z nich je **väčších ako 5**.
Triedy, ktoré sa zlúčili do jednej triedy sú farebne vyznačené.

| **trieda** | **xi-** | **xi+** | $f_j$ | $\hat{f}_j$ |
|---|---|---|---|---|
| 1 | $-\infty$ | -0,4255 | 7 | 6,9949 |
| 2 | -0,4255 | -0,2418 | 5 | 5,5590 |
| 3 | -0,2418 | -0,0582 | 6 | 7,2902 |
| 4 | -0,0582 | 0,1255 | 10 | 8,1023 |
| 5 | 0,1255 | 0,3091 | 7 | 7,6313 |
| 6 | 0,3091 | 0,4927 | 8 | 6,0913 |
| 7 | 0,4927 | $\infty$ | 7 | 8,3310 |
| **Σ** | | | **50** | **50** |

**Testovacie kritérium:** $t = \sum_{j=1}^{m} \frac{(f_j - \hat{f}_j)^2}{\hat{f}_j} = 1,59200788$

**Počet stupňov voľnosti:** $k = m - q - 1 = 7 - 2 - 1 = 4$

**Kvantil:** $\chi^2_{1-\alpha}(k) = \chi^2_{1-\alpha}(4) = 9,487729037$

**Doplnok kritického oboru:** $\overline{W}_\alpha = \langle 0;\, \chi^2_{1-\alpha} \rangle = \langle 0;\, 9,487729037 \rangle$

Pretože $t \in \overline{W}_\alpha$, hypotéza $H : X \sim N(0,059;\, 0,200924490)$ sa **nezamieta**.

*e) Za predpokladu (bez ohľadu na výsledok časti (d)), že štatistický súbor bol získaný náhodným výberom z normálneho rozdelenia, určte intervalové odhady strednej hodnoty, rozptylu a smerodajnej odchýlky so spoľahlivosťou 0, 95 a 0, 99.*

**Predpoklad:** $X \sim N(\mu, \sigma^2), \sigma^2$ - nepoznáme

**Bodový odhad strednej hodnoty:** $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 0,059$

**Bodový odhad rozptylu:** $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = 0,200924490$

**Bodový odhad smerodajnej odchýlky:** $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} = 0,4482460148$

**Intervalový odhad** parametra $\boldsymbol{\mu}$:
$0,975$ kvantil Studentovho rozdelenia s $k = n - 1 = 49$ stupňami voľnosti:
$t_{0,975}(49) = 2,009575237$

$0,995$ kvantil Studentovho rozdelenia s $k = n - 1 = 49$ stupňami voľnosti:
$t_{0,995}(49) = 2,679951974$

$\alpha = 0,05 : \left\langle \overline{x} - t_{1-\alpha/2}\frac{s}{\sqrt{n}}; \overline{x} + t_{1-\alpha/2}\frac{s}{\sqrt{n}} \right\rangle = \big\langle -0,068390108; 0,186390108\big\rangle$

$\alpha = 0,01 : \left\langle \overline{x} - t_{1-\alpha/2}\frac{s}{\sqrt{n}}; \overline{x} + t_{1-\alpha/2}\frac{s}{\sqrt{n}} \right\rangle = \big\langle -0,110886335; 0,228886335\big\rangle$

**Intervalový odhad** parametra $\boldsymbol{\sigma^2}$:
$0,025$ kvantil Pearsonovho rozdelenia $\chi^2_{\alpha/2}$ s $k = n - 1 = 49$ stupňami voľnosti:
$\chi^2_{0,025}(49) = 31,55491646$

$0,975$ kvantil Pearsonovho rozdelenia $\chi^2_{1-\alpha/2}$ s $k = n - 1 = 49$ stupňami voľnosti:
$\chi^2_{0,975}(49) = 70,22241357$

$0,005$ kvantil Pearsonovho rozdelenia $\chi^2_{\alpha/2}$ s $k = n - 1 = 49$ stupňami voľnosti:
$\chi^2_{0,005}(49) = 27,24934907$

$0,995$ kvantil Pearsonovho rozdelenia $\chi^2_{1-\alpha/2}$ s $k = n - 1 = 49$ stupňami voľnosti:
$\chi^2_{0,995}(49) = 78,23070809$

$\alpha = 0,05 : \left\langle \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}; \frac{(n-1)s^2}{\chi^2_{\alpha/2}} \right\rangle = \big\langle 0,140201675; 0,312005263\big\rangle$

$\alpha = 0,01 : \left\langle \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}; \frac{(n-1)s^2}{\chi^2_{\alpha/2}} \right\rangle = \big\langle 0,125849558; 0,361304043\big\rangle$

**Intervalový odhad** parametra $\boldsymbol{\sigma}$:

$\alpha = 0,05 : \left\langle \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}; \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} \right\rangle = \big\langle 0,374435141; 0,558574313\big\rangle$

$\alpha = 0,01 : \left\langle \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}; \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} \right\rangle = \big\langle 0,354752812; 0,601085721\big\rangle$

*f) Testujte hypotézu optimálneho nastavenia stroja, tj. že stredná hodnota odchýlky je nulová, proti dvojstrannej alternatívnej hypotéze, že stredná hodnota odchýlky je rôzna od nuly, a to na hladine významnosti 0, 05.*

Bodové odhady **strednej hodnoty** a **smerodajnej odchýlky**:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 0,059 \qquad s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} = 0,4482460148$$

**Studentov jednovýberový test**:

**Testujeme hypotézu** $H_0 : \mu = 0$:

**Testovacie kritérium:** $t = \frac{\overline{x} - \mu_0}{s}\sqrt{n} = \frac{\overline{x} - 0}{s}\sqrt{n} = 0,930723279$

**Doplnok kritického oboru:** $\overline{W}_\alpha = \big\langle -t_{1-\alpha/2}; t_{1-\alpha/2}\big\rangle$ pre alternatívnu hypotézu: $H_A : \mu \neq \mu_0$

$0,975$ kvantil Studentovho rozdelenia $t_{1-\alpha/2}$ s $k = n - 1 = 49$ stupňami voľnosti:

$t_{0,975} = -2,009575237$

$\overline{W}_\alpha = \big\langle -t_{1-\alpha/2}; t_{1-\alpha/2}\big\rangle = \big\langle -2,009575237; 2,009575237\big\rangle$

Pretože $t \in \overline{W}_\alpha$, tak hypotéza $H_0 : \mu = 0$ sa **nezamieta** a alternatívna hypotéza $H_A : \mu \neq 0$ sa **zamieta**.

*g) Overte štatistickým testom na hladine významnosti 0, 05, či nastavenie stroja ovplyvnilo kvalitu výrobku, ak viete, že vyššie uvedený štatistický súbor 50-ich hodnôt vznikol spojením dvoch čiastočných štatistických súborov, tak, že po nameraní prvých 20-ich hodnôt bolo vykonané nové nastavenie stroja a následne bolo nameraných zvyšných 30 hodnôt.*

**Štatistický súbor X**

| $n$ | $x_n$ | $n$ | $x_n$ |
|---|---|---|---|
| 1 | -0,43 | 11 | -0,25 |
| 2 | -0,12 | 12 | -0,73 |
| 3 | 0,04 | 13 | 0,25 |
| 4 | -1,05 | 14 | -0,25 |
| 5 | 0,30 | 15 | -0,40 |
| 6 | 0,27 | 16 | 0,59 |
| 7 | 0,18 | 17 | -0,19 |
| 8 | 0,86 | 18 | 0,09 |
| 9 | 0,40 | 19 | -0,27 |
| 10 | 0,39 | 20 | -0,07 |

**Štatistický súbor Y**

| $n$ | $y_n$ | $n$ | $y_n$ | $n$ | $y_n$ |
|---|---|---|---|---|---|
| 21 | 0,31 | 31 | 0,36 | 41 | -0,47 |
| 22 | 0,31 | 32 | -0,19 | 42 | -0,59 |
| 23 | -1,16 | 33 | 0,49 | 43 | 0,29 |
| 24 | 0,70 | 34 | 0,12 | 44 | -0,15 |
| 25 | 0,78 | 35 | 0,00 | 45 | 0,00 |
| 26 | 0,22 | 36 | 0,39 | 46 | 0,40 |
| 27 | -0,37 | 37 | 0,70 | 47 | -0,09 |
| 28 | 0,30 | 38 | 0,00 | 48 | 0,08 |
| 29 | 0,74 | 39 | -0,50 | 49 | 0,08 |
| 30 | -0,02 | 40 | 0,05 | 50 | 0,56 |

| | ***X*** | ***Y*** |
|---|---|---|
| ***n*** | 20 | 30 |
| ***priemer*** | -0,0195000 | 0,111333 |
| $s^2$ | 0,19466475 | 0,185751556 |
| $s$ | 0,441208284 | 0,430989043 |

**Test rovnosti rozptylov – F-test:**

**Testujeme hypotézu** $H_0 : \sigma_X^2 = \sigma_Y^2$:

**Testovacie kritérium:** $t = \dfrac{s^2(X)}{s^2(Y)} = \; = 1,047984494$

**Doplnok kritického oboru:** $\overline{W}_\alpha = \langle F_{\alpha/2}(n-1, m-1);\, F_{1-\alpha/2}(n-1, m-1)\rangle$ pre $H_A : \sigma_X^2 \neq \sigma_Y^2$

$F_{\alpha/2}(n-1, m-1) = F_{0,025}(19, 29) = 0,416329668$

$F_{1-\alpha/2}(n-1, m-1) = F_{0,975}(19, 29) = 2,231273833$

$\overline{W}_\alpha = \Big\langle F_{\alpha/2}(n-1, m-1);\, F_{1-\alpha/2}(n-1, m-1)\Big\rangle = \Big\langle 0,416329668;\, 2,231273833\Big\rangle$

Pretože $t \in \overline{W}_\alpha$, hypotéza $H_0 : \sigma_X^2 = \sigma_Y^2$ sa **nezamieta**.

**Studentov dvojvýberový test:**

**Testujeme hypotézu** $H_0 : \mu_X - \mu_Y = 0$ za podmienky $\sigma_X^2 = \sigma_Y^2$

**Testovacie kritérium:** $t = \dfrac{\overline{x} - \overline{y} - \mu_0}{\sqrt{(n-1)s^2(X) + (m-1)s^2(Y)}} \sqrt{\dfrac{nm(n+m)}{n+m}} = -1,041734424$

**Doplnok kritického oboru:** $\overline{W}_\alpha = \Big\langle -t_{1-\alpha/2};\, t_{1-\alpha/2}\Big\rangle$ pre alternatívnu hypotézu:
$H_A : \mu_X - \mu_Y \neq 0$,

$t_{1-\alpha/2}$ - kvantil Studentovho rozdelenia s $k = n + m - 2 = 48$ stupňami voľnosti:

$t_{0,975}(48) = 2,010634758$

$\overline{W}_\alpha = \Big\langle -t_{1-\alpha/2};\, t_{1-\alpha/2}\Big\rangle = \Big\langle -2,010634758;\, -2,010634758\Big\rangle$

Pretože $t \in \overline{W}_\alpha$, hypotéza $H_0 : \mu_X - \mu_Y = 0$ sa **nezamieta.**

**2. Meraním dvojice (Výška[cm], Váha[kg]) u vybraných študentov z fakulty VUT FIT bol získaný dvojrozmerný štatistický súbor zapísaný po dvojiciach v riadkoch v liste Data_pr.2, zadanie č.14.**

| X - Výška [cm] | Y - Váha [kg] | X - Výška [cm] | Y - Váha [kg] |
|---|---|---|---|
| 170 | 59 | 170 | 66 |
| 197 | 100 | 175 | 65 |
| 168 | 75 | 178 | 97 |
| 157 | 52 | 177 | 68 |
| 182 | 78 | 153 | 43 |
| 190 | 89 | 168 | 44 |
| 188 | 90 | 168 | 66 |
| 187 | 84 | 167 | 61 |
| 198 | 97 | 176 | 87 |
| 193 | 109 | 151 | 54 |

$n = 20$, $\overline{x} = 175,65$, $\overline{y} = 74,257642$

$$\sum_{i=1}^{n} x_i^2 = 620633, \sum_{i=1}^{n} y_i^2 = 117452,39139, \sum_{i=1}^{n} x_i y_i = 265291,9784$$

*a) Vypočítajte bodový odhad koeficientu korelácie*

$$r = \frac{\sum\limits_{i=1}^{n} x_i y_i - n\overline{x}\overline{y}}{\sqrt{\left(\sum\limits_{i=1}^{n} x_i^2 - n\overline{x}^2\right)\left(\sum\limits_{i=1}^{n} y_i^2 - n\overline{y}^2\right)}} = 0,874135371$$

*b) Na hladine významnosti 0,05 testujte hypotézu, že náhodné veličiny Výška a Váha sú lineárne nezávislé.*

**Testujeme hypotézu** $H_0 : \rho = 0 :$

**Testovacie kritérium:** $t = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} = 7,635941928$

**Doplnok kritického oboru:** $\overline{W}_\alpha = \langle 0;\, t_{1-\alpha/2} \rangle$ pre alternatívnu hypotézu: $H_A : \rho \neq 0$,

$t_{1-\alpha/2}(n-2) = t_{0,975}(18) = 2,10092204$

Pretože $t \notin \overline{W}_\alpha$, teda hypotéza $H_0 : \rho = 0$ sa **zamieta.**

*c) Regresná analýza - dáta preložte priamkou: Váha = $\beta_1 + \beta_1 * $ Výška*

$$n = 20, \sum_{i=1}^{n} x_i = 3513; \sum_{i=1}^{n} y_i = 1485,152831; \sum_{i=1}^{n} x_i^2 = 620633; \sum_{i=1}^{n} y_i^2 = 117452,391388;$$

$$\sum_{i=1}^{n} x_i y_i = 265291,978445; \qquad det(H) = n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 = 71491$$

*1) Bodovo odhadnite $\beta_1$, $\beta_1$ a rozptyl $s^2$*

$$b_2 = \frac{1}{det(H)}\left(n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i\right) = 1,237885505$$

$$b_1 = \overline{y} - b_2\overline{x} = -143,1769473$$

$$y = b_1 + b_2 x = -143,1769473 + 1,237885505x$$

$$S^*_{min} = \sum_{i=1}^{n} y_i^2 - b_1\sum_{i=1}^{n} y_i - b_2\sum_{i=1}^{n} x_i y_i = 1690,94547$$

$$s^2 = \frac{S^*_{min}}{n-2} = \frac{S^*_{min}}{20-2} = 93,94141499$$

*2) Na hladine významnosti 0,05 otestujte nasledujúce hypotézy:*

**Testujeme hypotézu** $H: \beta_0 = -100$, $H_A: \beta_0 \neq -100$,

$$h^{11} = \frac{\sum_{i=1}^{n} x_i^2}{det(H)} = 8,681274566$$

**Testovacie kritérium:** $t = \dfrac{b_1 - (-100)}{s\sqrt{h^{11}}} = -1,511929842$

$$t_{1-\alpha/2}(n-2) = t_{0,975}(18) = 2,10092204$$

$t \in \overline{W} = \langle -2,10092204; -2,10092204 \rangle$, a teda $H: \beta_1 = -100$ sa **nezamieta**.

**Testujeme hypotézu** $H: \beta_1 = 1$, $H_A: \beta_1 \neq 1$,

$$h^{22} = \frac{n}{det(H)} = 0,000279755$$

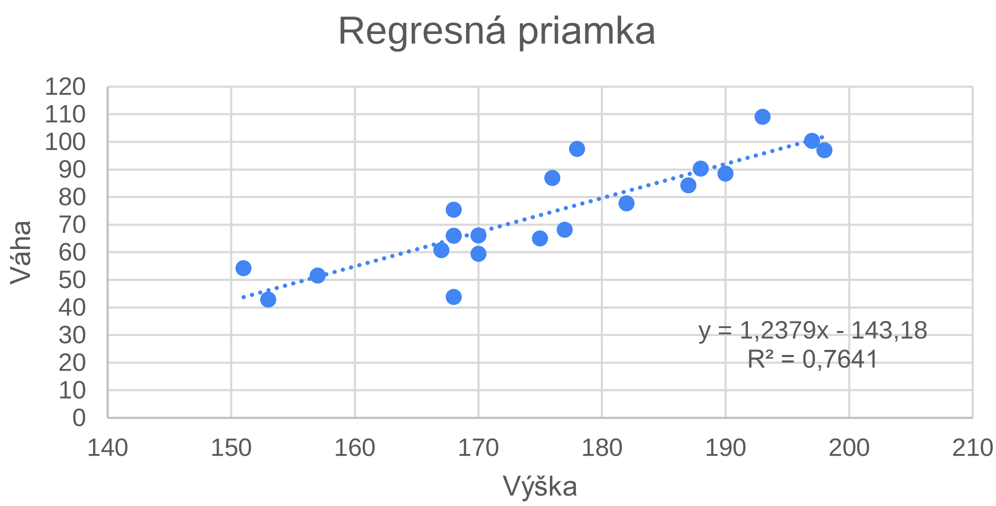
**Testovacie kritérium:** $t = \dfrac{b_2 - 1}{s\sqrt{h^{22}}} = 1,46740542$

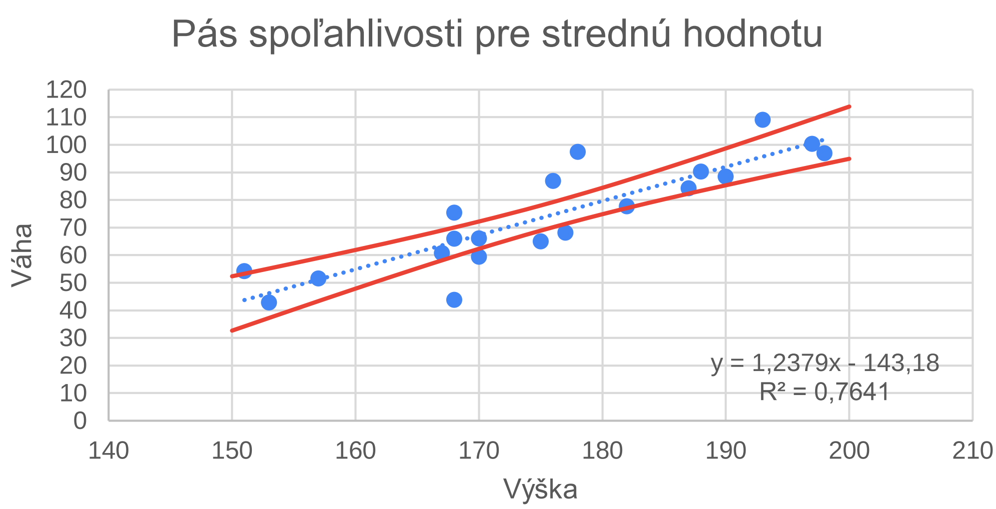
$$t_{1-\alpha/2}(n-2) = t_{0,975}(18) = 2,10092204$$

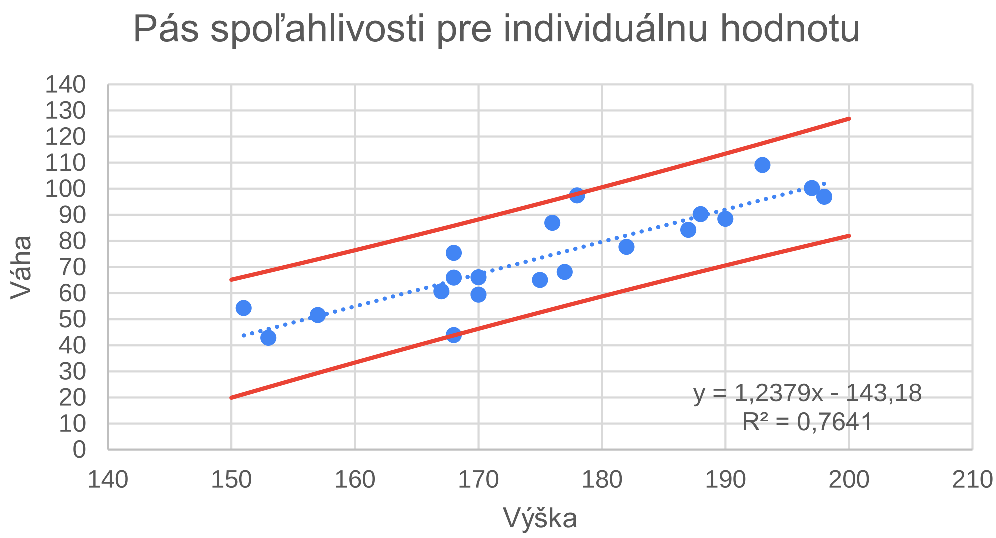
$t \in \overline{W} = \langle -2,10092204; -2,10092204 \rangle$, a teda $H: \beta_2 = 1$ sa **nezamieta**.

*3) Vytvorte graf bodov spolu s regresnou priamkou a pásom spoľahlivosti pre individuálnu hodnotu výšky*

| | | **stredná váha** | | **individuálna váha** | | |
|---|---|---|---|---|---|---|
| **výška$_i$** | **váha$_i$** | **váha$_i^-$** | **váha$_i^+$** | **váha$_i^-$** | **váha$_i^+$** | **h*** |
| 150 | 42,50587837 | 32,65443863 | 52,35731812 | 19,88517149 | 65,12658525 | 0,234057434 |
| 155 | 48,69530589 | 40,31694012 | 57,07367167 | 26,67616586 | 70,71444592 | 0,169294037 |
| 160 | 54,88473342 | 47,87451736 | 61,89494947 | 33,34898073 | 76,4204861 | 0,118518415 |
| 165 | 61,07416094 | 55,25271669 | 66,8956052 | 39,89552246 | 82,25279942 | 0,081730567 |
| 170 | 67,26358846 | 62,32038451 | 72,20679242 | 46,309334 | 88,21784292 | 0,058930495 |
| 175 | 73,45301599 | 68,89436646 | 78,01166551 | 52,58613313 | 94,31989884 | 0,050118197 |
| 180 | 79,64244351 | 74,85420024 | 84,43068678 | 58,72420295 | 100,5606841 | 0,055293673 |
| 185 | 85,83187103 | 80,27550263 | 91,38823944 | 64,72455608 | 106,939186 | 0,074456925 |
| 190 | 92,02129855 | 85,34154152 | 98,70105559 | 70,59083732 | 113,4517598 | 0,107607951 |
| 195 | 98,21072608 | 90,20041732 | 106,2210348 | 76,32898572 | 120,0924664 | 0,154746751 |
| 200 | 104,4001536 | 94,93913322 | 113,861174 | 81,94672577 | 126,8535814 | 0,215873327 |




Pás spoľahlivosti pre individuálnu hodnotu
Váha
Výška
y = 1,2379x - 143,18
R² = 0,7641
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
140
150
160
170
180
190
200
210