

BRNO UNIVERSITY OF TECHNOLOGY

FACULTY OF INFORMATION TECHNOLOGY



Štatistika a pravdepodobnosť - MSP

Cvičenie - streda, 08:00

Projekt - zadanie 23

1 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) Vykonajte 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 výroby, 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 zvolte 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** - data preložte priamkou: $Váha = \beta_0 + \beta_1 * Výška$
 - (a) Bodovo odhadnite β_0 , β_1 a rozptyl s^2 .
 - (b) Na hladine významnosti 0.05 otestujte hypotézy:

$$H : \beta_0 = -100, \quad H_A : \beta_0 \neq -100, \quad (1)$$

$$H : \beta_1 = 1, \quad H_A : \beta_1 \neq 1, \quad (2)$$

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

2 VYPRACOVANIE 1.ÚLOHY

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

Štatistický súbor - X

	x_n		x_n
n = 1	1.94	26	2.41
2	1.40	27	0.85
3	-0.42	28	-0.23
4	0.18	29	-0.07
5	1.39	30	0.75
6	1.72	31	2.71
7	1.77	32	1.71
8	-0.40	33	3.06
9	1.91	34	-1.26
10	0.63	35	-0.66
11	0.69	36	-0.77
12	0.10	37	0.98
13	2.06	38	1.42
14	1.71	39	0.96
15	0.17	40	0.10
16	0.78	41	0.41
17	0.15	42	-0.26
18	2.45	43	-0.32
19	3.51	44	-0.16
20	1.71	45	2.53
21	0.30	46	2.19
22	1.38	47	0.00
23	1.11	48	1.01
24	1.27	49	3.65
25	1.11	50	0.90

Usporiadáný štatistický súbor - $X_{<}$

	x_i		x_i
n = (1)	-1.26	(26)	0.98
(2)	-0.77	(27)	1.01
(3)	-0.66	(28)	1.11
(4)	-0.42	(29)	1.11
(5)	-0.40	(30)	1.27
(6)	-0.32	(31)	1.38
(7)	-0.26	(32)	1.39
(8)	-0.23	(33)	1.40
(9)	-0.16	(34)	1.42
(10)	-0.07	(35)	1.71
(11)	0.00	(36)	1.71
(12)	0.10	(37)	1.71
(13)	0.10	(38)	1.72
(14)	0.15	(39)	1.77
(15)	0.17	(40)	1.91
(16)	0.18	(41)	1.94
(17)	0.30	(42)	2.06
(18)	0.41	(43)	2.19
(19)	0.63	(44)	2.41
(20)	0.69	(45)	2.45
(21)	0.75	(46)	2.53
(22)	0.78	(47)	2.71
(23)	0.85	(48)	3.06
(24)	0.90	(49)	3.51
(25)	0.96	(50)	3.65

2.1 Vykonajte roztriedenie štatistického súboru, vytvorte tabuľku početností a nakreslite histogramy pre relatívnu početnosť a relatívnu komutatívnu početnosť.

Neroztriedený štatistický súbor: $X = \{x_1, x_2, \dots, x_n\}$, $1 \leq n \leq 50$.

Rozsah štatistického súboru: $|X| = n$.

Usporiadáný štatistický súbor: $X_{<} = \{x_{(1)}, x_{(2)}, \dots, x_{(n)}\}$, kde $\forall 1 \leq i \leq n : x_{(i)} \leq x_{(i+1)}$.

Variačný obor: $\langle x_{(1)}, x_{(n)} \rangle = \langle \min_i x_i, \max_i x_i \rangle = \langle -1.26, 3.65 \rangle$.

Rozpätie štatistického súboru: $|\langle x_{(1)}, x_{(n)} \rangle| = x_{(n)} - x_{(1)} = 3.65 - (-1.26) = 4.91$.

Počet tried: $m = \left\lceil \frac{\sqrt{n} + 2\sqrt{n}}{2} \right\rceil = \left\lceil \frac{\sqrt{50} + 2\sqrt{50}}{2} \right\rceil = \left\lceil \frac{15\sqrt{2}}{2} \right\rceil = 11$.

Dĺžka triedy: $h \approx \frac{x_{(n)} - x_{(1)}}{m} = \frac{3.65 - (-1.26)}{11} = 0.44636$.

Trieda	(x_i^-, x_i^+)	Stred	KP	P	RP	RKP
1	$\langle -1.2600, -0.8136 \rangle$	-1.0368	1	1	0.02	0.02
2	$\langle -0.8136, -0.3673 \rangle$	-0.5905	5	4	0.08	0.1
3	$\langle -0.3673, 0.0791 \rangle$	-0.1441	11	6	0.12	0.22
4	$\langle 0.0791, 0.5255 \rangle$	0.3023	18	7	0.14	0.36
5	$\langle 0.5255, 0.9718 \rangle$	0.7486	25	7	0.14	0.5
6	$\langle 0.9718, 1.4182 \rangle$	1.1950	33	8	0.16	0.66
7	$\langle 1.4182, 1.8645 \rangle$	1.6414	39	6	0.12	0.78
8	$\langle 1.8645, 2.3109 \rangle$	2.0877	43	4	0.08	0.86
9	$\langle 2.3109, 2.7573 \rangle$	2.5341	47	4	0.08	0.94
10	$\langle 2.7573, 3.2036 \rangle$	2.9805	48	1	0.02	0.96
11	$\langle 3.2036, 3.6500 \rangle$	3.4268	50	2	0.04	1

Table 1. Tabuľka početnosti k danému štatistickému súboru, kde **KP** je *kumulatívna početnosť*, **P** je *početnosť*, **RP** je *relatívna početnosť* a **RKP** je *relatívna kumulatívna početnosť*.

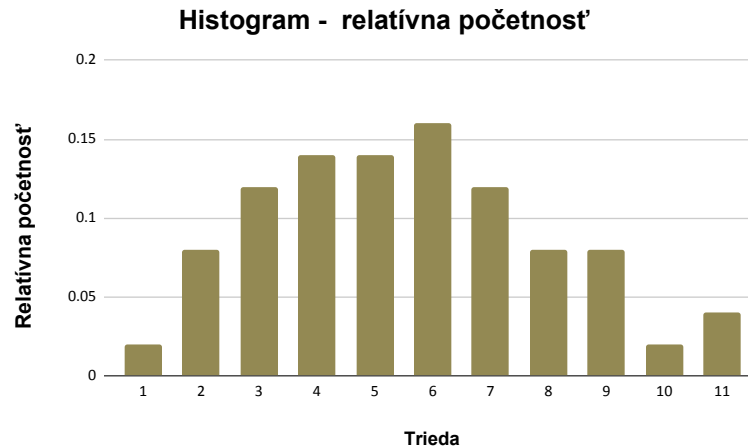


Figure 1. Histogram relatívnej početnosti vytvorený z údajov v Tabuľke 1.

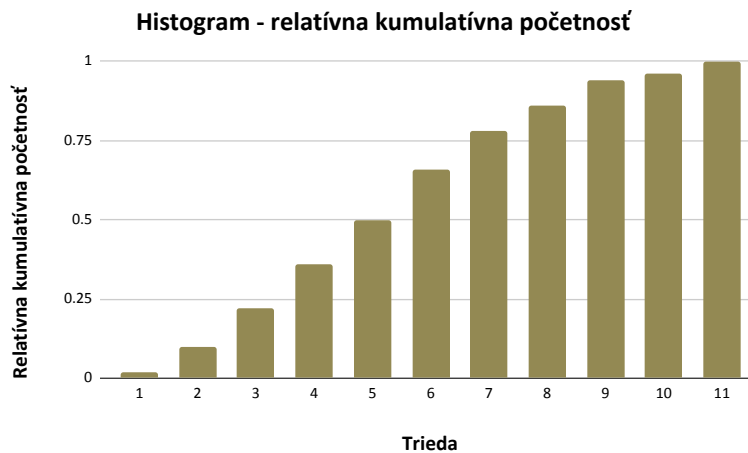


Figure 2. Histogram relatívnej kumulatívnej početnosti vytvorený z údajov v Tabuľke 1.

2.2 Vypočítajte aritmetický priemer, medián, modus, rozptyl a smerodajnú odchýlku.

Aritmetický priemer: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 1.0106$

Medián: $\tilde{x} = 0.9700$

Modus: $\hat{x} = 1.7100$

Rozptyl: $\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = 1.25494964$

Smerodajná odchýlka: $\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} = 1.1202453481$

2.3 Vypočítajte bodové odhady strednej hodnoty, rozptylu a smerodajnej odchýlky.

Bodový odhad strednej hodnoty: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 1.0106$

Bodový odhad rozptylu: $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 1.280560857$

Bodový odhad smerodajnej odchýlky: $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = 1.131618689$

2.4 Testujte predpoklad o výbere z normálneho rozdelenia Pearsonovým (chi-kvadrát) testom na hladine významnosti 0.05.

Uvažujme nasledovnú Hypotézu: $H : X \sim N(\mu, \sigma^2)$, kde μ, σ^2 sú neznáme parametre. Využijeme preto ich bodový odhad z predchádzajúceho kroku:

Bodový odhad strednej hodnoty: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 1.0106$

Bodový odhad rozptylu: $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 1.280560857$

Použijeme zotriedený štatistický súbor z kroku (a), v ktorom sme pôvodný nezotriedený štatistický súbor roztriedili do 11 tried ($m = 11$) s nasledujúcimi vlastnosťami:

Variačný obor: $\langle x_{(1)}, x_{(n)} \rangle = \langle \min_i x_i, \max_i x_i \rangle = \langle -1.26, 3.65 \rangle$.

Rozpätie štatistického súboru: $|\langle x_{(1)}, x_{(n)} \rangle| = x_{(n)} - x_{(1)} = 3.65 - (-1.26) = 4.91$.

Dĺžka triedy: $h \approx \frac{x_{(n)} - x_{(1)}}{m} = \frac{3.65 - (-1.26)}{11} = 0.44636$.

Trieda	$\langle x_i^-, x_i^+ \rangle$	P	TP
1	$\langle -1.2600, -0.8136 \rangle$	1	1.553670
2	$\langle -0.8136, -0.3673 \rangle$	4	2.910558
3	$\langle -0.3673, 0.0791 \rangle$	6	4.676067
4	$\langle 0.0791, 0.5255 \rangle$	7	6.442857
5	$\langle 0.5255, 0.9718 \rangle$	7	7.613315
6	$\langle 0.9718, 1.4182 \rangle$	8	7.715586
7	$\langle 1.4182, 1.8645 \rangle$	6	6.706008
8	$\langle 1.8645, 2.3109 \rangle$	4	4.998703
9	$\langle 2.3109, 2.7573 \rangle$	4	3.195543
10	$\langle 2.7573, 3.2036 \rangle$	1	1.751944
11	$\langle 3.2036, 3.6500 \rangle$	2	0.823713
Σ		50	48.387967

Figure 3. Tabuľka teoretických četností (TP) vyrátaná na variačnom obore zotriedeného štatistického súboru.

Trieda	$\langle x_i^-, x_i^+ \rangle$	P	TP
1	$\langle -\infty, -0.8136 \rangle$	1	2.673726
2	$\langle -0.8136, -0.3673 \rangle$	4	2.910558
3	$\langle -0.3673, 0.0791 \rangle$	6	4.676067
4	$\langle 0.0791, 0.5255 \rangle$	7	6.442857
5	$\langle 0.5255, 0.9718 \rangle$	7	7.613315
6	$\langle 0.9718, 1.4182 \rangle$	8	7.715586
7	$\langle 1.4182, 1.8645 \rangle$	6	6.706008
8	$\langle 1.8645, 2.3109 \rangle$	4	4.998703
9	$\langle 2.3109, 2.7573 \rangle$	4	3.195543
10	$\langle 2.7573, 3.2036 \rangle$	1	1.751944
11	$\langle 3.2036, \infty \rangle$	2	1.315690
Σ		50	50

Figure 4. Tabuľka teoretických četností (TP) vyrátaná na celom obore \mathbb{R} .

Aby bola splnená podmienka pre teoretické početnosti, ktorá vraví o tom, že všetky teoretické početnosti sú **väčšie než 1** a aspoň **80%** z nich je **väčších ako 5**, zlúčime krajné triedy do spoločnej triedy:

Trieda	(x_i^-, x_i^+)	P	TP	$\frac{(TP-P)^2}{TP}$
1	$(-10000, -0.3673)$	5	5.584284	0.061134
2	$(-0.3673, 0.0791)$	6	4.676067	0.374844
3	$(0.0791, 0.5255)$	7	6.442857	0.048179
4	$(0.5255, 0.9718)$	7	7.613315	0.049408
5	$(0.9718, 1.4182)$	8	7.715586	0.010484
6	$(1.4182, 1.8645)$	6	6.706008	0.074329
7	$(1.8645, 2.3109)$	4	4.998703	0.199533
8	$(2.3109, 10000)$	7	6.263261	0.086662
Σ		50	50	0.904572

Testovacie kritérium: $t = \sum_{j=1}^m \frac{(f_j - \hat{f}_j)^2}{\hat{f}_j} = 0.904571943939$

Zvolená hladina významnosti: $\alpha = 0.05$

Počet stupňov voľnosti: $k = 8 - 2 - 1 = 5$

Kvantil: $\chi^2_{1-\alpha}(k) = \chi^2_{0.95}(5) = 11.0704976935$

Doplnok kritického oboru: $\bar{W}_\alpha = \langle 0, \chi^2_{1-\alpha}(k) \rangle = \langle 0, \chi^2_{0.95}(5) \rangle = \langle 0, 11.0704976935 \rangle$

Pretože $t \in \bar{W}_\alpha$, teda hypotéza: $H: X \sim N(\mu, \sigma^2) = X \sim N(1.0106, 1.280560857)$ sa **nezamieta**.

2.5 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

Bodový odhad strednej hodnoty: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 1.0106$

Bodový odhad rozptylu: $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 1.280560857$

Bodový odhad smerodajnej odchýlky: $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = 1.131618689$

2.5.1 Intervalový odhad parametra μ

$$\left\langle \bar{x} - t_{1-\alpha/2} \frac{s}{\sqrt{n}}; \bar{x} + t_{1-\alpha/2} \frac{s}{\sqrt{n}} \right\rangle, \text{ kde } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \text{ a}$$

$t_{1-\alpha/2}$ je $1 - \alpha/2$ - kvantil Studentovho rozdelenia s $k = n - 1$ stupňov voľnosti

0.975 kvantil Studentovho rozdelenia $t_{1-\alpha/2}$ s $k = n - 1 = 50 - 1 = 49$ stupňov voľnosti = **2.009575237**.

0.995 kvantil Studentovho rozdelenia $t_{1-\alpha/2}$ s $k = n - 1 = 50 - 1 = 49$ stupňov voľnosti = **2.679951974**.

$$\alpha = 0.05 : \left\langle \bar{x} - t_{1-\alpha/2} \frac{s}{\sqrt{n}}; \bar{x} + t_{1-\alpha/2} \frac{s}{\sqrt{n}} \right\rangle = \left\langle 0.688997527; 1.332202473 \right\rangle$$

$$\alpha = 0.01 : \left\langle \bar{x} - t_{1-\alpha/2} \frac{s}{\sqrt{n}}; \bar{x} + t_{1-\alpha/2} \frac{s}{\sqrt{n}} \right\rangle = \left\langle 0.581713753; 1.439486247 \right\rangle$$

2.5.2 Intervalový odhad parametra σ^2

$$\left\langle \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}; \frac{(n-1)s^2}{\chi^2_{\alpha/2}} \right\rangle, \text{ kde } s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \text{ a } \chi^2_{1-\alpha/2}, \chi^2_{\alpha/2} \text{ je } 1 - \alpha/2 \text{ (} \alpha/2 \text{) -}$$

kvantil Pearsonovho rozdelenia s $k = n - 1$ stupňov voľnosti.

0.025 kvantil Pearsonovho rozdelenia $\chi^2_{\alpha/2}$ s $k = n - 1 = 50 - 1 = 49$ stupňov voľnosti = **31.55491646**.
0.975 kvantil Pearsonovho rozdelenia $\chi^2_{1-\alpha/2}$ s $k = n - 1 = 50 - 1 = 49$ stupňov voľnosti = **70.22241357**.
0.005 kvantil Pearsonovho rozdelenia $\chi^2_{\alpha/2}$ s $k = n - 1 = 50 - 1 = 49$ stupňov voľnosti = **27.24934907**.
0.995 kvantil Pearsonovho rozdelenia $\chi^2_{1-\alpha/2}$ s $k = n - 1 = 50 - 1 = 49$ stupňov voľnosti = **78.23070809**.

$$\alpha = 0.05 : \left\langle \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}; \frac{(n-1)s^2}{\chi^2_{\alpha/2}} \right\rangle = \left\langle 0.945279578; 1.410147799 \right\rangle$$

$$\alpha = 0.01 : \left\langle \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}; \frac{(n-1)s^2}{\chi^2_{\alpha/2}} \right\rangle = \left\langle 0.895590588; 1.517469899 \right\rangle$$

2.5.3 Intervalový odhad parametra σ

$$\alpha = 0.05 : \left\langle \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}; \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} \right\rangle = \left\langle \sqrt{0.945279578}; \sqrt{1.410147799} \right\rangle = \left\langle 0.972254893; 1.187496442 \right\rangle$$

$$\alpha = 0.01 : \left\langle \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}; \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} \right\rangle = \left\langle \sqrt{0.895590588}; \sqrt{1.517469899} \right\rangle = \left\langle 0.946356480; 1.231856282 \right\rangle$$

2.6 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.

Bodový odhad strednej hodnoty: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 1.0106$

Bodový odhad rozptylu: $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 1.280560857$

Studentov jednovýberový test

Testujeme hypotézu $H_0 : \mu = \mu_0, \mu_0 = 0$:

Testovacie kritérium: $t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \sqrt{n} = \frac{\bar{x} - 0}{\frac{s}{\sqrt{n}}} \sqrt{n} = \frac{1.0106}{1.131618689} \sqrt{50} = \mathbf{6.31486666}$

Doplňok kritického oboru: $\bar{W}_\alpha = \langle -t_{1-\alpha/2}; t_{1-\alpha/2} \rangle$ pre alternatívnu hypotézu: $H_A : \mu \neq \mu_0, \mu_0 = 0$

0.975 kvantil Studentovho rozdelenia $t_{1-\alpha/2}$ s $k = n - 1 = 50 - 1 = 49$ stupňov voľnosti = **2.009575237**.

$\bar{W}_\alpha = \langle -t_{1-\alpha/2}; t_{1-\alpha/2} \rangle = \langle -2.009575237; 2.009575237 \rangle$

Pretože $t \notin \bar{W}_\alpha$, tak hypotéza $H_0 : \mu = 0$ sa **zamiet**a alternatívna hypotéza $H_A : \mu \neq 0$ sa **nezamiet**a.

2.7 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

1	1.94	11	0.69
2	1.40	12	0.10
3	-0.42	13	2.06
4	0.18	14	1.71
5	1.39	15	0.17
6	1.72	16	0.78
7	1.77	17	0.15
8	-0.40	18	2.45
9	1.91	19	3.51
10	0.63	20	1.71

Figure 5. $X_1 = \{x_1, x_2, \dots, x_{20}\}$

1	0.30	11	2.71	21	0.41
2	1.38	12	1.71	22	-0.26
3	1.11	13	3.06	23	-0.32
4	1.27	14	-1.26	24	-0.16
5	1.11	15	-0.66	25	2.53
6	2.41	16	-0.77	26	2.19
7	0.85	17	0.98	27	0.00
8	-0.23	18	1.42	28	1.01
9	-0.07	19	0.96	29	3.65
10	0.75	20	0.10	30	0.90

Figure 6. $X_2 = \{x_{21}, x_{22}, \dots, x_{50}\}$

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

	n	\bar{X}	s^2	s
X₁	20	1.17250	1.01129875	1.005633507
X₂	30	0.90267	1.38825956	1.178244268

Test rovnosti rozptylov – F-test

Testujeme hypotézu $H_0 : \sigma_{X_1}^2 = \sigma_{X_2}^2$:

Testovacie kritérium: $t = \frac{s^2(X_1)}{s^2(X_2)} = \frac{1.01129875}{1.38825956} = 0.7284651821$

Doplnok kritického oboru: $\bar{W}_\alpha = \langle F_{\alpha/2}(n_1 - 1, n_2 - 1); F_{1-\alpha/2}(n_1 - 1, n_2 - 1) \rangle$ pre $H_A : \sigma_{X_1}^2 \neq \sigma_{X_2}^2$, kde

$F_{\alpha/2}(k_1, k_2)$, $F_{1-\alpha/2}(k_1, k_2)$ sú kvantily Fischerovo-Snedecorovho rozdelenia s $k_1 = n_1 - 1$ a $k_2 = n_2 - 1$ stupňov voľnosti.

$F_{\alpha/2}(n_1 - 1, n_2 - 1) = F_{\alpha/2}(20 - 1, 30 - 1) = F_{\alpha/2}(19, 29) = 0.4163296676$

$F_{1-\alpha/2}(n_1 - 1, n_2 - 1) = F_{1-\alpha/2}(20 - 1, 30 - 1) = F_{1-\alpha/2}(19, 29) = 2.231273833$

$\bar{W}_\alpha = \langle F_{\alpha/2}(n_1 - 1, n_2 - 1); F_{1-\alpha/2}(n_1 - 1, n_2 - 1) \rangle = \langle 0.4163296676; 2.231273833 \rangle$

Pretože $t \in \bar{W}_\alpha$, tak hypotéza $H_0 : \sigma_{X_1}^2 = \sigma_{X_2}^2$ sa **nezamieta**.

Studentov dvojvýberový test

Testujeme hypotézu $H_0 : \mu_{X_1} - \mu_{X_2} = \mu_0$ za podmienky $\sigma_{X_1}^2 = \sigma_{X_2}^2$

Testovacie kritérium: $t = \frac{\bar{x}_1 - \bar{x}_2 - \mu_0}{\sqrt{(n_1 - 1)s^2(X_1) + (n_2 - 1)s^2(X_2)}} \sqrt{\frac{n_1 * n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} = 0.8397355191$

Doplnok kritického oboru: $\bar{W}_\alpha = \langle -t_{1-\alpha/2}; t_{1-\alpha/2} \rangle$ pre $H_A : \mu_{X_1} - \mu_{X_2} \neq 0$, kde

$t_{1-\alpha/2}$ – kvantil Studentovho rozdelenia s $k = n_1 + n_2 - 2 = 20 + 30 - 2 = 48$ stupňami voľnosti.

$t_{1-\alpha/2}(48) = 2.010634758$

$\bar{W}_\alpha = \langle -t_{1-\alpha/2}; t_{1-\alpha/2} \rangle = \langle -2.010634758; 2.010634758 \rangle$

Pretože $t \in \bar{W}_\alpha$, tak hypotéza $H_0 : \mu_{X_1} - \mu_{X_2} = 0$ sa **nezamieta**.

3 VYPRACOVANIE 2.ÚLOHY

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.

Výška [cm]	Váha [kg]	Výška [cm]	Váha [kg]	Výška [cm]	Váha [kg]	Výška [cm]	Váha [kg]
161	81.4617	187	102.7276	151	58.9631	190	107.8464
180	92.2658	169	87.6579	182	112.4235	158	77.6862
179	87.3327	169	91.9425	193	110.0129	200	126.3419
196	118.8958	180	93.7274	154	64.4363	197	115.5837
198	114.1958	175	90.2365	182	106.1934	176	106.2360

n	\bar{x}	\bar{y}	$\sum_{i=1}^n x_i^2$	$\sum_{i=1}^n y_i^2$	$\sum_{i=1}^n x_i y_i$
20	178.85	97.308361	643961	195492.194386	352843.46909

3.1 Vypočítajte bodový odhad koeficientu korelácie

$$r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{(\sum_{i=1}^n x_i^2 - n \bar{x}^2)(\sum_{i=1}^n y_i^2 - n \bar{y}^2)}} = 0.9399797064$$

3.2 Na hladine významnosti 0.05 testujte hypotézu, že náhodné veličiny Výška a Váha sú lineárne nezávislé.

Testujeme hypotézu $H_0 : \rho = 0$:

Testovanie kritérium: $t = \frac{|r| \sqrt{n-2}}{\sqrt{1-r^2}} = 11.68710858$

Doplnok kritického oboru: $\bar{W}_\alpha = \langle 0; t_{1-\alpha/2} \rangle$, pre alternatívnu hypotézu $H_A : \rho \neq 0$, kde r je bodový odhad koeficientu korelácie a $t_{1-\alpha/2}$ je kvantil Studentovho rozdelenia s $k = n - 2$ stupňami voľnosti.

$$t_{1-\alpha/2}(n-2) = t_{0.975}(20-2) = 2.10092204$$

$$\bar{W}_\alpha = \langle 0; t_{1-\alpha/2} \rangle = \langle 0; 2.10092204 \rangle$$

Pretože $t \notin \bar{W}_\alpha$, tak hypotéza $H_0 : \rho = 0$ sa **zamietá**.

3.3 Regresná analýza - data preložte priamkou: Váha = $\beta_1 + \beta_1 * \text{Výška}$

n	$\sum_{i=1}^n x_i$	$\sum_{i=1}^n y_i$	\bar{x}	\bar{y}	$\sum_{i=1}^n x_i^2$	$\sum_{i=1}^n y_i^2$	$\sum_{i=1}^n x_i y_i$
20	3577	1946.167210	178.85	97.308361	643961	195492.194386	352843.46909

$$\det(\mathbf{H}) = n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 = 84291$$

3.3.1 Bodovo odhadnite β_1, β_1 a rozptyl s^2

$$b_2 = \frac{1}{\det(\mathbf{H})} \left(n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i \right) = 1.132140691$$

$$b_1 = \bar{y} - b_2 \bar{x} = -105.175002$$

$$y = b_1 + b_2 x = -105.175002 + 1.132140691x$$

$$S_{min}^* = \sum_{i=1}^n y_i^2 - b_1 \sum_{i=1}^n y_i - b_2 \sum_{i=1}^n x_i y_i = 711.885846$$

$$s^2 = \frac{S_{min}^*}{n-2} = \frac{S_{min}^*}{20-2} = 39.54921366$$

3.3.2 Na hladine významnosti 0.05 otestujte nasledujúce hypotézy

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

$$h^{11} = \frac{\sum_{i=1}^n x_i^2}{\det(H)} = 7.639736152$$

$$\text{Testovacie kritérium } t = \frac{b_j - \beta_{j-1}^0}{s\sqrt{h^{jj}}} = \frac{b_1 - \beta_0^0}{s\sqrt{h^{11}}} = \frac{b_1 - (-100)}{s\sqrt{h^{11}}} = -0.2977161712$$

$$\text{Doplňok kritického oboru: } \bar{W}_\alpha = \left\langle -t_{1-\alpha/2}(n-2); t_{1-\alpha/2}(n-2) \right\rangle = \left\langle -t_{1-\alpha/2}(20-2); t_{1-\alpha/2}(20-2) \right\rangle$$

$$t_{1-\alpha/2}(20-2) = t_{0.975}(18) = 2.10092204$$

$$\bar{W}_\alpha = \left\langle -t_{0.975}(18); t_{0.975}(18) \right\rangle = \left\langle -2.10092204; 2.10092204 \right\rangle$$

Pretože $t \in \bar{W}_\alpha$, tak hypotéza $H : \beta_0 = -100$ sa **nezamieta**.

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

$$h^{22} = \frac{n}{\det(H)} = \frac{20}{\det(H)} = 0.0002372732557$$

$$\text{Testovacie kritérium } t = \frac{b_j - \beta_{j-1}^0}{s\sqrt{h^{jj}}} = \frac{b_2 - \beta_1^0}{s\sqrt{h^{22}}} = \frac{b_2 - 1}{s\sqrt{h^{22}}} = 1.364090716$$

$$\text{Doplňok kritického oboru: } \bar{W}_\alpha = \left\langle -t_{1-\alpha/2}(n-2); t_{1-\alpha/2}(n-2) \right\rangle = \left\langle -t_{1-\alpha/2}(20-2); t_{1-\alpha/2}(20-2) \right\rangle$$

$$t_{1-\alpha/2}(20-2) = t_{0.975}(18) = 2.10092204$$

$$\bar{W}_\alpha = \left\langle -t_{0.975}(18); t_{0.975}(18) \right\rangle = \left\langle -2.10092204; 2.10092204 \right\rangle$$

Pretože $t \in \bar{W}_\alpha$, tak hypotéza $H : \beta_1 = 1$ sa **nezamieta**.

3.3.3 Vytvorte graf bodov spolu s regresnou priamkou a pásom spoľahlivosti pre individuálnu hodnotu výšky

x_i	y_i	h^*	$S y_i^-$	$S y_i^+$	$I y_i^-$	$I y_i^+$
150	64.646102	0.2474878694	58.073220	71.218983	49.889137	79.403066
155	70.306805	0.1849663665	64.624491	75.989119	55.924388	84.689222
160	75.967508	0.1343085264	71.125443	80.809574	61.895877	90.039140
165	81.628212	0.0955143491	77.544894	85.711530	67.799304	95.457120
170	87.288915	0.0685838346	83.828807	90.749024	73.631039	100.946791
175	92.949619	0.0535169828	89.893117	96.006121	79.388372	106.510866
180	98.610322	0.0503137939	95.646703	101.573941	85.069707	112.150938
185	104.271026	0.0589742677	101.062466	107.479586	90.674700	117.867352
190	109.931729	0.0794984043	106.206457	113.657002	96.204279	123.659179
195	115.592433	0.1118862037	111.172994	120.011872	101.660575	129.524291
200	121.253136	0.1561376659	116.032388	126.473884	107.046749	135.459523

Table 2. Tabuľka zobrazujúca hodnoty priamky (y_i) určenej vypočítanými koeficientami b_1 a b_2 v daných bodoch x_i určených na základe hodnôt vstupných dát. Tabuľka taktiež obsahuje vypočítané hodnoty intervalových odhadov v daných bodoch x_i : Intervalový odhad individuálnej hodnoty y_i : $\langle I y_i^-; I y_i^+ \rangle$, Intervalový odhad strednej hodnoty y_i : $\langle S y_i^-; S y_i^+ \rangle$.

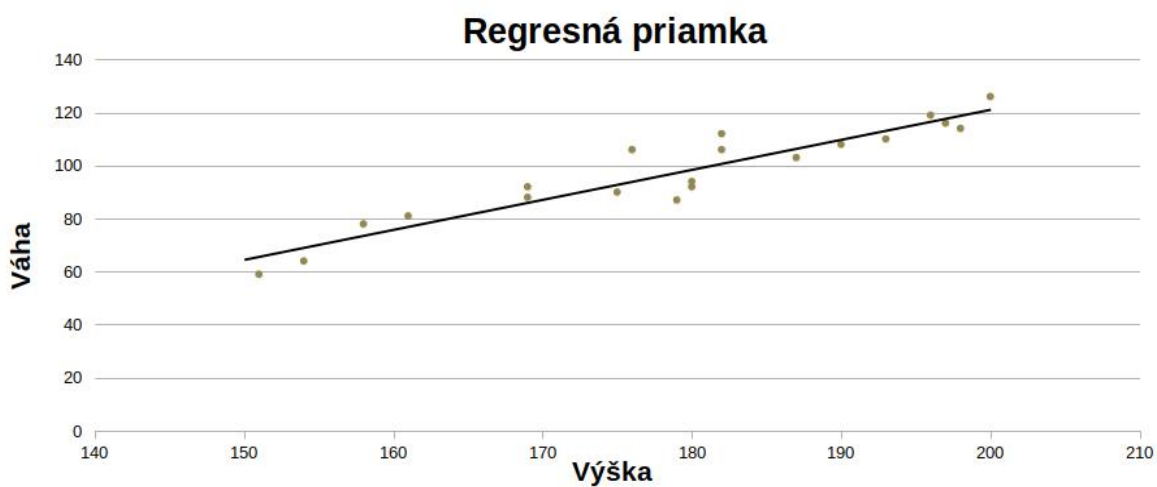


Figure 7. Graf zobrazujúci hodnoty vstupných dát a regresnú priamku vyjadrenú bodmi určenými v Tabuľke 2.

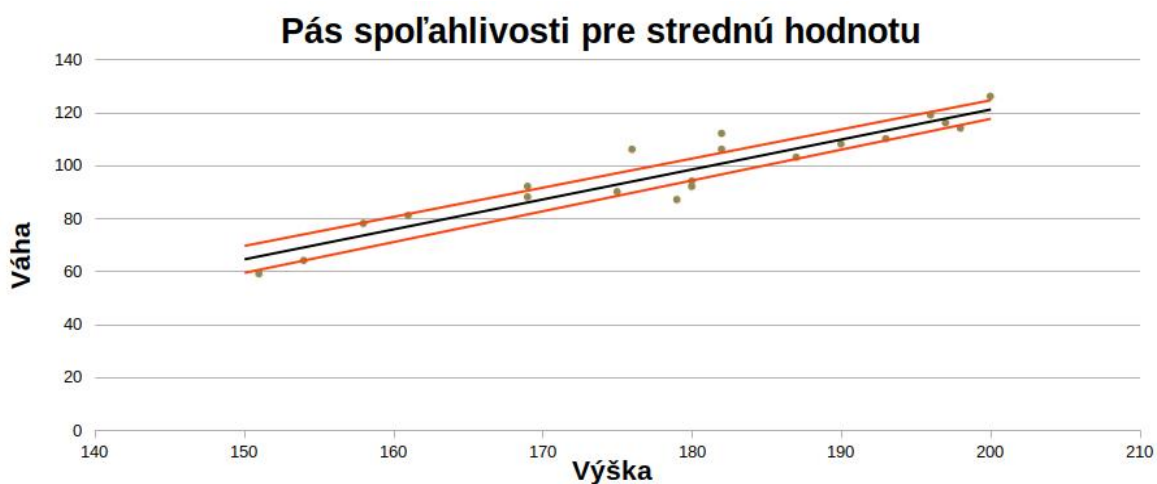


Figure 8. Graf zobrazujúci hodnoty vstupných dát, regresnú priamku vyjadrenú bodmi určenými v Tabuľke 2 a pás spoľahlivosti pre strednú hodnotu y_i na základe hodnôt $\langle S y_i^-; S y_i^+ \rangle$ v Tabuľke 2.

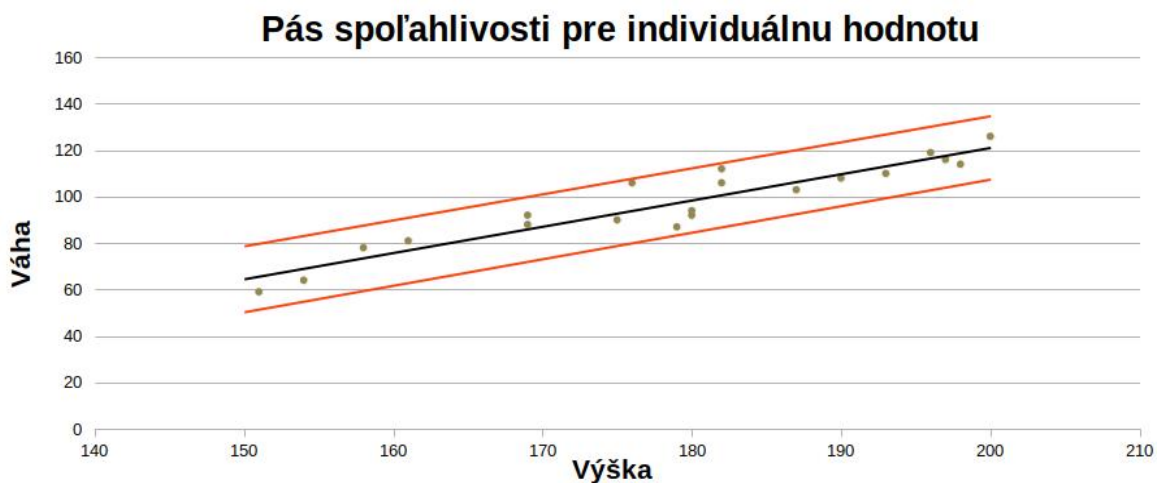


Figure 9. Graf zobrazujúci hodnoty vstupných dát, regresnú priamku vyjadrenú bodmi určenými v Tabuľke 2 a pás spoľahlivosti pre individuálnu hodnotu y_i na základe hodnôt $\langle I y_i^-; I y_i^+ \rangle$ v Tabuľke 2.