



Projekt z MSP

Vypracoval: Tomas Lapsansky, xlapsa00
Číslo zadania: 3, 28
Cvičenie - skupina: utorok, 12:00
Dátum: 8.12.2020

Fakulta Informačních technologií VUT v Brně
2019/2020

Vypracovanie

1. Při kontrole výrobků byla sledována odchylka $X[\text{mm}]$ jejich rozměru od požadované velikosti. Naměřené hodnoty tvoří statistický soubor v listu Data_př. 1.

1	-0.42	26	1.14
2	0.07	27	-0.05
3	0.38	28	0.32
4	-0.60	29	0.89
5	0.17	30	0.33
6	-0.63	31	0.31
7	-0.66	32	-0.23
8	-0.15	33	0.13
9	-1.87	34	-0.10
10	0.27	35	-0.34
11	-0.67	36	1.09
12	0.09	37	-0.12
13	0.35	38	-0.20
14	0.25	39	0.17
15	0.92	40	0.36
16	-0.23	41	0.57
17	0.16	42	0.17
18	0.30	43	0.16
19	-0.51	44	-0.30
20	-0.90	45	1.22
21	0.08	46	-0.80
22	0.65	47	-0.18
23	0.45	48	0.16
24	-1.13	49	-0.78
25	0.38	50	-1.24

1	-1.87	26	0.13
2	-1.24	27	0.16
3	-1.13	28	0.16
4	-0.90	29	0.16
5	-0.80	30	0.17
6	-0.78	31	0.17
7	-0.67	32	0.17
8	-0.66	33	0.25
9	-0.63	34	0.27
10	-0.60	35	0.30
11	-0.51	36	0.31
12	-0.42	37	0.32
13	-0.34	38	0.33
14	-0.30	39	0.35
15	-0.23	40	0.36
16	-0.23	41	0.38
17	-0.20	42	0.38
18	-0.18	43	0.45
19	-0.15	44	0.57
20	-0.12	45	0.65
21	-0.10	46	0.89
22	-0.05	47	0.92
23	0.07	48	1.09
24	0.08	49	1.14
25	0.09	50	1.22

- a) Proved'te roztrídění statistického souboru, vytvořte tabulku četností a nakreslete histogramy pro relativní četností a relativní kumulativní četností.

$$x_{(1)} = \min_i x_i = -1,87$$

$$x_{(n)} = \max_i x_i = 1,22$$

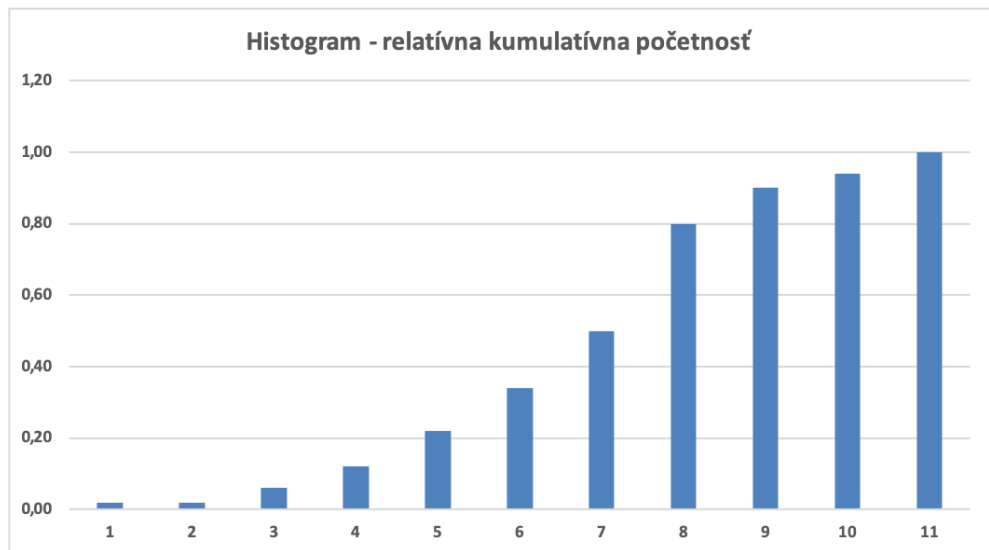
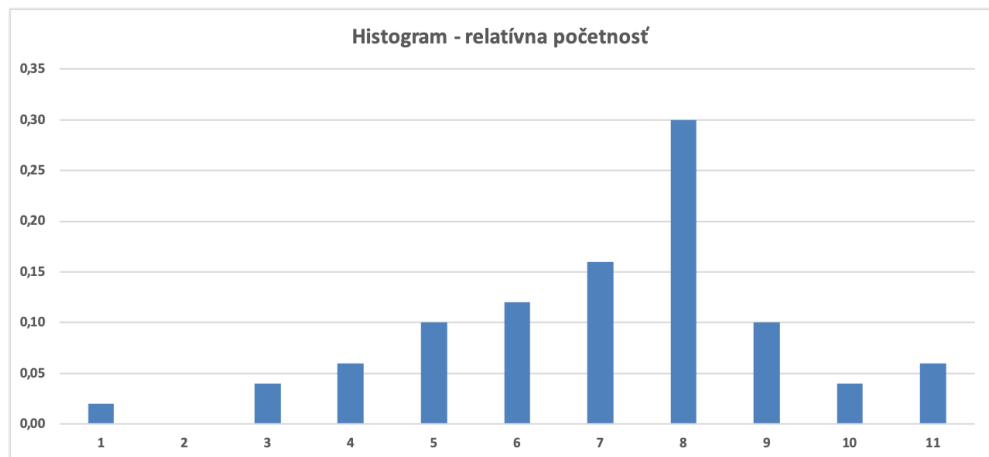
$$\text{Variačný odbor: } \langle x_{(1)}, x_{(n)} \rangle = \langle -1,87; 1,22 \rangle$$

$$\text{Rozpätie: } x_{(n)} - x_{(1)} = 3,09$$

$$\text{Počet tried } m = 11 \text{ (zvolené)}$$

$$\text{Dĺžka triedy} = \frac{x_{(n)} - x_{(1)}}{m} = 0,280909091$$

Trieda	xi-	xi+	Stred triedy	Kumulatívna početnosť	Početnosť	Relatívna početnosť	Relatívna kumulatívna početnosť
1	-1,8700	-1,5891	-1,7295	1	1	0,02	0,02
2	-1,5891	-1,3082	-1,4486	1	0	0,00	0,02
3	-1,3082	-1,0273	-1,1677	3	2	0,04	0,06
4	-1,0273	-0,7464	-0,8868	6	3	0,06	0,12
5	-0,7464	-0,4655	-0,6059	11	5	0,10	0,22
6	-0,4655	-0,1845	-0,3250	17	6	0,12	0,34
7	-0,1845	0,0964	-0,0441	25	8	0,16	0,50
8	0,0964	0,3773	0,2368	40	15	0,30	0,80
9	0,3773	0,6582	0,5177	45	5	0,10	0,90
10	0,6582	0,9391	0,7986	47	2	0,04	0,94
11	0,9391	1,2200	1,0795	50	3	0,06	1,00



- b) **Vypočítejte aritmetický průměr, medián, modus, rozptyl a směrodatnou odchylku.**

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = -0,0114$$

$$\text{medián: } \tilde{x} = 0,11$$

$$\text{modus: } \hat{x} = 0,16$$

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \tilde{x})^2 = 0,37424804$$

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \tilde{x})^2} = 0,611758155$$

- c) **Vypočítejte bodové odhady střední hodnoty, rozptylu a směrodatné odchylky.**

$$\text{Bodový odhad střednej hodnoty: } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = -0,0114$$

$$\text{Bodový odhad rozptylu: } s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \tilde{x})^2 = 0,3819$$

$$\text{Bodový odhad smerodatnej odchýľky: } s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \tilde{x})^2} = 0,617969057$$

- d) **Testujte předpoklad o výběru z normálního rozdělení Pearsonovým (chí-kvadrát) testem na hladině významnosti 0,05.**

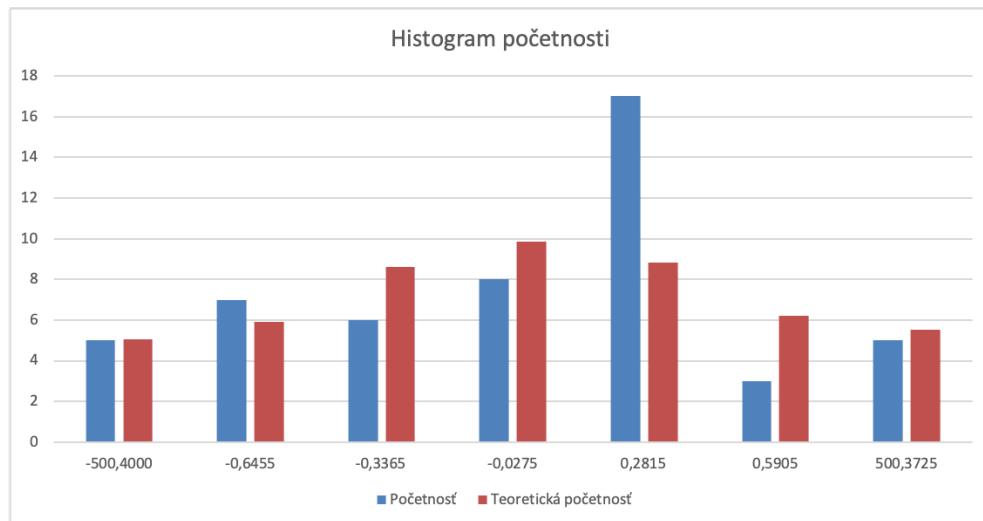
Trieda	xi-	xi+	Stred triedy	Kumulativna početnost	Početnost	Teoretická početnost	roz ² /teoretická početnost
1	-1000,0000	-0,8000	-500,4000	5	5	5,047865783	0,000453882
2	-0,8000	-0,4910	-0,6455	12	7	5,894520982	0,207325390
3	-0,4910	-0,1820	-0,3365	18	6	8,620055226	0,796362576
4	-0,1820	0,1270	-0,0275	26	8	9,867836293	0,353553942
5	0,1270	0,4360	0,2815	43	17	8,842839981	7,524648159
6	0,4360	0,7450	0,5905	46	3	6,203165700	1,654037792
7	0,7450	1000,0000	500,3725	51	5	5,523716036	0,049654704

$$\text{Testovacie kritérium: } t = \sum_{j=1}^m \frac{(f_j - \hat{f}_j)^2}{\hat{f}_j} = 10,586036445$$

$$\chi^2_{1-\alpha} \text{ pre } k = 7 - 2 - 1 \text{ stupňov voľnosti: } 9,487729037,$$

$$\text{doplňok kritického odboru: } \bar{W}_\alpha = \left\langle 0, \chi^2_{1-\alpha} \right\rangle = \left\langle 0, 9,487729037 \right\rangle.$$

$$t \in \bar{W}_\alpha \implies H : X \sim N(-0,0806; 0,2116) \text{ sa } \mathbf{zamieta}.$$



- e) Za predpokladu (bez ohľadu na výsledok časti d)), že statistický súbor bol získan náhodným výberom z normálneho rozdelenia, určete intervalové odhady strednej hodnoty, rozptylu a smerodatnej odchylky se spoľehľivosťou 0,95 a 0,99.

Predpoklad: $X \sim N(\mu, \sigma^2)$, σ^2 - neznáme

Bodový odhad strednej hodnoty: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = -0,0114$

Bodový odhad rozptylu: $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 0,3819$

Bodový odhad smerodatnej odchylky: $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = 0,617969057$

Intervalový odhad parametra μ :

0,975 kvantil Studentovho rozdelenia $t_{1-\alpha/2}$ s $k = n - 1 = 50 - 1 = 49$ stupňami voľnosti = 2,009575237

0,995 kvantil Studentovho rozdelenia $t_{1-\alpha/2}$ s $k = n - 1 = 50 - 1 = 49$ stupňami voľnosti = 2,679951964

$$\alpha = 0,05 : \left\langle \bar{x} - t_{1-\alpha/2} \frac{s}{\sqrt{n}}; \bar{x} + t_{1-\alpha/2} \frac{s}{\sqrt{n}} \right\rangle = \left\langle -0,187024863; 0,164224863 \right\rangle$$

$$\alpha = 0,01 : \left\langle \bar{x} - t_{1-\alpha/2} \frac{s}{\sqrt{n}}; \bar{x} + t_{1-\alpha/2} \frac{s}{\sqrt{n}} \right\rangle = \left\langle -0,245611781; 0,222811781 \right\rangle$$

Intervalový odhad parametra σ^2 :

0,975 kvantil Pearsovho rozdelenia $\chi^2_{\alpha/2}$ s $k = n - 1 = 50 - 1 = 49$ stupňami voľnosti = 31,55491646

0,975 kvantil Pearsovho rozdelenia $\chi^2_{1-\alpha/2}$ s $k = n - 1 = 50 - 1 = 49$ stupňami voľnosti = 70,22241357

0,995 kvantil Pearsovo rozdělení $\chi^2_{\alpha/2}$ s $k = n - 1 = 50 - 1 = 49$ stupňami volnosti = 27,24934921

0,995 kvantil Pearsovo rozdělení $\chi^2_{1-\alpha/2}$ s $k = n - 1 = 50 - 1 = 49$ stupňami volnosti = 78,23070806

$$\alpha = 0,05 : \left\langle \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}; \frac{(n-1)s^2}{\chi^2_{\alpha/2}} \right\rangle = \left\langle 0,266473353; 0,593010665 \right\rangle$$

$$\alpha = 0,01 : \left\langle \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}; \frac{(n-1)s^2}{\chi^2_{\alpha/2}} \right\rangle = \left\langle 0,2391951; 0,686710052 \right\rangle$$

Intervalový odhad parametra σ :

$$\alpha = 0,05 : \left\langle \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}; \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} \right\rangle = \left\langle 0,51621057; 0,770071857 \right\rangle$$

$$\alpha = 0,01 : \left\langle \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}; \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} \right\rangle = \left\langle 0,489075761; 0,828679704 \right\rangle$$

- f) Testujte hypotézu optimálního seřazení stroje, tj. že střední hodnota odchylky je nulová, proti dvoustranné alternativní hypotéze, že střední hodnota odchylky je různá od nuly, a to na hladině významnosti 0,05.

Studentov jednovýberový test:

Testujeme hypotézu $H_0 : \mu = 0$:

testovací kritérium: $t = \frac{\bar{x} - \mu_0}{s} \sqrt{n} = \frac{\bar{x} - 0}{s} \sqrt{n} = -0,130443705$

doplnek kritického odboru: $\bar{W}_\alpha = \left\langle -t_{1-\alpha/2}, t_{1-\alpha/2} \right\rangle$ pre alternativnu hypotézu: $H_A : \mu \neq \mu_0$,

0,975 kvantil Studentovho rozdělení $t_{1-\alpha/2}$ s $k = n - 1 = 50 - 1 = 49$ stupňami volnosti = 2,009575237

$$\bar{W}_\alpha = \left\langle -t_{1-\alpha/2}, t_{1-\alpha/2} \right\rangle = \left\langle -2,0095752; 2,0095752 \right\rangle$$

$t \in \bar{W}_\alpha \implies H : H_0 : \mu = 0$ sa **nezamieta**.

- g) Ověřte statistickým testem na hladině významnosti 0,05, zda seřazení stroje ovlivnilo kvalitu výroby, víte-li, že výše uvedený statistický soubor 50-ti hodnot vznikl spojením dvou dílčích statistických souborů tak, že po naměření prvních 20-ti hodnot bylo provedeno nové seřazení stroje a pak bylo naměřeno zbývajících 30 hodnot.

x1:20 - X		x21:50 - Y	
1	-0,42	1	0,08
2	0,07	2	0,65
3	0,38	3	0,45
4	-0,60	4	-1,13
5	0,17	5	0,38
6	-0,63	6	1,14
7	-0,66	7	-0,05
8	-0,15	8	0,32
9	-1,87	9	0,89
10	0,27	10	0,33
11	-0,67	11	0,31
12	0,09	12	-0,23
13	0,35	13	0,13
14	0,25	14	-0,10
15	0,92	15	-0,34
16	-0,23	16	1,09
17	0,16	17	-0,12
18	0,30	18	-0,20
19	-0,51	19	0,17
20	-0,90	20	0,36
		21	0,57
		22	0,17
		23	0,16
		24	-0,30
		25	1,22
		26	-0,80
		27	-0,18
		28	0,16
		29	-0,78
		30	-1,24

	X	Y
n=	20	30
priemer=	-0,1840	0,1037
rozptyl s²=	0,3595	0,3510
smer _odch=	0,5996	0,5925

Test rovnosti rozptylov - F-test:**Testujeme hypotézu** $H_0 : \sigma_x^2 = \sigma_y^2$:testovacie kritérium: $t = \frac{s^2(X)}{s^2(Y)} = \frac{0,3595}{0,3510} = 1,024104559$ doplňok kritického odboru: $\overline{W}_\alpha = \left\langle F_{\alpha/2}(n-1, m-1), F_{1-\alpha/2}(n-1, m-1) \right\rangle$ pre $H_A : \sigma_x^2 \neq \sigma_y^2$, $F_{\alpha/2}(k_1, k_2), F_{1-\alpha/2}(k_1, k_2)$ sú kvantily Fisherovho-Snedecorovho rozdelenia s $k_1 = n - 1$ a $k_2 = m - 1$ stupňami voľnosti. $F_{\alpha/2}(19, 29) = 0,416329667$ $F_{1-\alpha/2}(19, 29) = 2,231274$ $\overline{W}_\alpha = \left\langle F_{\alpha/2}(n-1, m-1), F_{1-\alpha/2}(n-1, m-1) \right\rangle = \left\langle 0,416329667; 2,231274 \right\rangle$ $t \in \overline{W}_\alpha \implies H_0 : \sigma_x^2 = \sigma_y^2$ sa **nezamieta**.**Studentov dvojvýberový test:****Testujeme hypotézu** $H_0 : \mu_x - \mu_y = 0$ za podmienky $\sigma_x^2 = \sigma_y^2$ testovacie kritérium: $t = \frac{\bar{x} - \bar{y} - \mu_0}{\sqrt{(n-1)s^2(X) + (m-1)s^2(Y)}} \sqrt{\frac{n \cdot m(n+m-2)}{n+m}} = -1,674027$ doplňok kritického odboru: $\overline{W}_\alpha = \left\langle -t_{1-\alpha/2}, t_{1-\alpha/2} \right\rangle$ pre $H_A : \mu_x - \mu_y \neq 0$, $t_{1-\alpha/2}$ je kvantil Studentovho rozdelenia s $k = n + m - 2 = 20 + 30 - 2 = 48$ stupňami voľnosti. $t_{1-\alpha/2} = 2,010634758$ $\overline{W}_\alpha = \left\langle -t_{1-\alpha/2}, t_{1-\alpha/2} \right\rangle = \left\langle -2,010634758; 2,010634758 \right\rangle$ $t \in \overline{W}_\alpha \implies H_0 : \mu_x - \mu_y = 0$ sa **nezamieta**.

2. Měřením dvojice (Výška[cm], Váha[kg]) u vybraných studentů z FIT byl získán dvourozměrný statistický soubor zapsaný po dvojicích v řádcích v listu Data_př. 2.

Výška [cm]	Váha [kg]
174	81
162	71
154	63
186	94
176	81
170	75
159	63
174	76
161	77
175	82
200	119
170	78
150	60
183	89
197	111
179	92
184	84
181	91
170	71
163	66

$$n = 20$$

$$\bar{x} = 173,4$$

$$\bar{y} = 81,2535$$

$$\sum_{i=1}^n x_i^2 = 604636$$

$$\sum_{i=1}^n y_i^2 = 136427,0225$$

$$\sum_{i=1}^n x_i y_i = 285384,2547$$

-
- a) Vypočtěte bodový odhad koeficientu korelace.

$$r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{(\sum_{i=1}^n x_i^2 - n \bar{x}^2)(\sum_{i=1}^n y_i^2 - n \bar{y}^2)}} = 0,947858976$$

- b) Na hladině významnosti 0,05 testujte hypotézu, že náhodné veličiny Výška a Váha jsou lineárně nezávislé.

Testujeme hypotézu $H_0 : \rho = 0$:

testovací kritérium: $t = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} = 12,61860796$

doplnok kritického odboru: $\bar{W}_\alpha = \langle 0, t_{1-\frac{\alpha}{2}} \rangle$ pre alternatívnu hypotézu: $H_A : \rho \neq 0$,
 $t_{1-\frac{\alpha}{2}}(n-2) = t_{0,975}(20-2) = 2,1009$

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

- c) **Regresní analýza** - data proložte přímkou: Váha = $\beta_0 + \beta_1 \cdot$ Výška

Výpočty:

xi	yi	xi ²	yi ²	xi*yi
174	81	30276,0000	6640,3505	14178,9722
162	71	26244,0000	5106,9792	11577,0274
154	63	23716,0000	3916,6226	9637,7706
186	94	34596,0000	8838,8897	17486,8588
176	81	30976,0000	6559,2747	14254,1255
170	75	28900,0000	5642,8545	12770,2191
159	63	25281,0000	4011,2126	10070,1274
174	76	30276,0000	5750,4853	13194,7601
161	77	25921,0000	5897,5468	12364,0734
175	82	30625,0000	6779,9097	14409,5363
200	119	40000,0000	14138,5988	23781,1680
170	78	28900,0000	6130,4057	13310,4743
150	60	22500,0000	3648,8361	9060,8395
183	89	33489,0000	7858,9105	16223,0409
197	111	38809,0000	12319,3241	21865,5128
179	92	32041,0000	8543,0493	16544,7225
184	84	33856,0000	7021,7504	15418,4429
181	91	32761,0000	8252,9771	16443,1074
170	71	28900,0000	5039,7461	12068,4987
163	66	26569,0000	4329,2986	10724,9771
3468,0000	1625,0701	604636,0000	136427,0225	285384,2547
173,4000	81,2535			

$$n = 20$$

$$\bar{x} = 173,4$$

$$\bar{y} = 81,2535$$

$$\sum_{i=1}^n x_i = 3468,0000$$

$$\sum_{i=1}^n y_i = 1625,0701$$

$$\begin{aligned}\sum_{i=1}^n x_i^2 &= 604636,0000 \\ \sum_{i=1}^n y_i^2 &= 136427,0225 \\ \sum_{i=1}^n x_i y_i &= 285384,2547 \\ \det(H) &= n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2 = 65696\end{aligned}$$

1. Bodově odhadněte β_0, β_1 a rozptyl s^2

$$\begin{aligned}b_2 &= \frac{1}{\det(H)} \left(n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i \right) = 1,09507368496826 \\ b_1 &= \bar{y} - b_2 \bar{x} = -108,632271587546 \\ y &= b_1 + b_2 x = y = -108,632271587546 + 1,09507368496826x \\ S_{min}^* &= \sum_{i=1}^n y_i^2 - b_1 \sum_{i=1}^n y_i - b_2 \sum_{i=1}^n x_i y_i = 445,2923432491 \\ s^2 &= \frac{S_{min}^*}{n-2} = 24,7384635138\end{aligned}$$

2. Na hladině významnosti 0,05 otestujte hypotézy:

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

$$\begin{aligned}h^{11} &= \frac{\sum_{i=1}^n x_i^2}{\det(H)} = 9,203543595 \\ t &= \frac{b_2 - 1}{s\sqrt{h^{11}}} = -0,572085838743498 \\ t_{1-\frac{\alpha}{2}}(n-2) &= t_{0,975}(18) = 2,100922037 \\ \overline{W} &= \langle -2, 100922037, 2, 100922037 \rangle\end{aligned}$$

$t \in \overline{W}$, takže $H : \beta_1 = -100$ nezamietame

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

$$\begin{aligned}h^{22} &= \frac{n}{\det(H)} = 0,000304433 \\ t &= \frac{b_2 - 1}{s\sqrt{h^{22}}} = 1,09554048653776 \\ t_{1-\frac{\alpha}{2}}(n-2) &= t_{0,975}(18) = 2,100922037 \\ \overline{W} &= \langle -2, 100922037, 2, 100922037 \rangle\end{aligned}$$

$t \in \overline{W}$, takže $H : \beta_1 = 1$ nezamietame

3. Vytvořte graf bodů spolu s regresní přímkou a pásem spolehlivosti pro individuální hodnotu výšky.

x_i	y_i	Ey_-	Ey_+	Y_-	Y_+	h^*
150	55,62878116	50,5067	60,7509	43,9914	67,2661	0,2403
155	61,10414958	56,7733	65,4350	49,7927	72,4156	0,1718
160	66,57951801	62,9824	70,1766	55,5282	77,6308	0,1185
165	72,05488643	69,0911	75,0186	61,1932	82,9166	0,0804
170	77,53025486	75,0221	80,0384	66,7839	88,2766	0,0576
175	83,00562328	80,6690	85,3422	72,2981	93,7132	0,0500
180	88,48099171	85,9729	90,9891	77,7347	99,2273	0,0576
185	93,95636013	90,9926	96,9201	83,0947	104,8180	0,0804
190	99,43172856	95,8346	103,0288	88,3804	110,4830	0,1185
195	104,907097	100,5762	109,2380	93,5957	116,2185	0,1718
200	110,3824654	105,2604	115,5045	98,7451	122,0198	0,2403



