



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta informačních technologií

Projekt z MSP

Zpracoval: Bc. Jakub Sadílek
Číslo zadání: 4, 12
Cvičení – skupina: čtvrtek, 10:00
Datum: 23.11.2020

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

Statistický soubor				Uspořádaný statistický soubor			
1	0,50	26	-1,08	(1)	-1,45	(26)	-0,12
2	-0,04	27	-0,22	(2)	-1,31	(27)	-0,12
3	-0,77	28	-0,24	(3)	-1,20	(28)	-0,11
4	-0,39	29	-0,91	(4)	-1,08	(29)	-0,10
5	-0,93	30	0,05	(5)	-1,07	(30)	-0,07
6	-0,07	31	-0,57	(6)	-0,93	(31)	-0,07
7	0,85	32	-0,18	(7)	-0,93	(32)	-0,06
8	-0,70	33	-1,45	(8)	-0,91	(33)	-0,06
9	0,56	34	-0,02	(9)	-0,87	(34)	-0,04
10	-0,32	35	-0,51	(10)	-0,81	(35)	-0,02
11	-0,87	36	0,28	(11)	-0,77	(36)	0,00
12	-0,11	37	-1,31	(12)	-0,70	(37)	0,05
13	0,23	38	-0,07	(13)	-0,57	(38)	0,13
14	-0,12	39	-0,25	(14)	-0,51	(39)	0,23
15	-0,06	40	0,46	(15)	-0,47	(40)	0,28
16	0,00	41	-0,81	(16)	-0,39	(41)	0,29
17	-0,21	42	-0,39	(17)	-0,39	(42)	0,46
18	0,13	43	0,29	(18)	-0,32	(43)	0,50
19	0,91	44	0,64	(19)	-0,32	(44)	0,56
20	-0,32	45	-0,47	(20)	-0,27	(45)	0,60
21	0,84	46	-0,12	(21)	-0,25	(46)	0,63
22	-0,10	47	0,63	(22)	-0,24	(47)	0,64
23	0,60	48	-1,07	(23)	-0,22	(48)	0,84
24	-1,20	49	-0,06	(24)	-0,21	(49)	0,85
25	-0,93	50	-0,27	(25)	-0,18	(50)	0,91

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

$$x_{(1)} = \min_i x_i = -1,45$$

$$x_{(n)} = \max_i x_i = 0,91$$

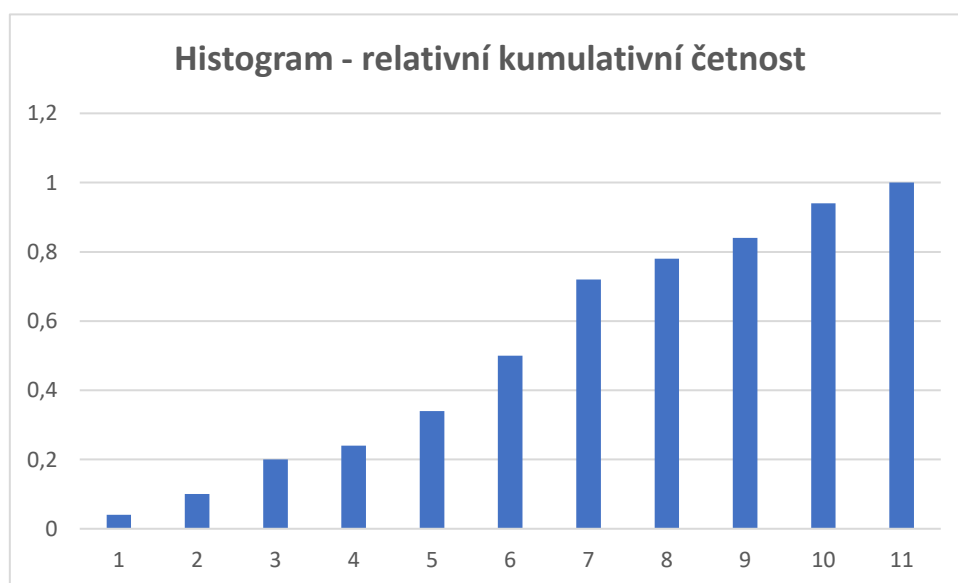
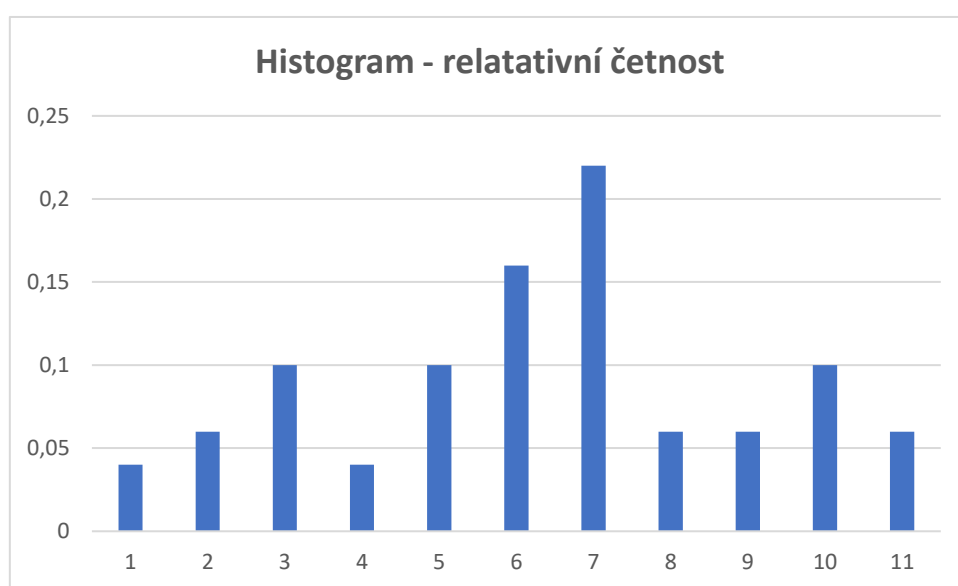
$$\text{Variační obor: } \langle x_{(1)}, x_{(n)} \rangle = \langle -1,45; 0,91 \rangle$$

$$\text{Rozpětí: } x_{(n)} - x_{(1)} = 2,36$$

$$\text{Počet tříd } m = 11 \quad (\text{zvoleno})$$

$$\text{Délka třídy} = \frac{x_{(n)} - x_{(1)}}{m} = 0,21455$$

Třída	xi-	xi+	Střed třídy	Kumulat. čet.	četnost	Relat. Čet.	Relat. Kum. Čet.
1	-1,45	-1,2355	-1,3428	2	2	0,04	0,04
2	-1,2355	-1,021	-1,1283	5	3	0,06	0,1
3	-1,021	-0,8065	-0,9138	10	5	0,1	0,2
4	-0,8065	-0,592	-0,6993	12	2	0,04	0,24
5	-0,592	-0,3775	-0,4848	17	5	0,1	0,34
6	-0,3775	-0,163	-0,2703	25	8	0,16	0,5
7	-0,163	0,0516	-0,0557	36	11	0,22	0,72
8	0,0516	0,2662	0,1589	39	3	0,06	0,78
9	0,2662	0,4808	0,3735	42	3	0,06	0,84
10	0,4808	0,6954	0,5881	47	5	0,1	0,94
11	0,6954	0,91	0,8027	50	3	0,06	1



b) Vypočtěte aritmetický průměr, medián, modus, rozptyl a směrodatnou odchylku.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = -0,2034$$

$$\text{Medián: } \tilde{x} = -0,15$$

$$\text{Modus: } \hat{x} = -0,0557$$

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = 0,33381$$

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} = 0,577763$$

c) Vypočtěte bodové odhady střední hodnoty, rozptylu a směrodatné odchylky.

Bodový odhad střední hodnoty:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = -0,2034$$

Bodový odhad rozptylu:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 0,3406229$$

Bodový odhad směrodatné odchylky:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = 0,58362908$$

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

Třída	xi-	xi+	Střed třídy	Kumulat. čet.	četnost	Teor. čet.	roz^2/teor čet
1	-1000	-0,8065	-500,40325	10	10	7,53586525	0,805741594
2	-0,8065	-0,59195	-0,699225	12	2	5,10343642	1,887222023
3	-0,59195	-0,3774	-0,484675	17	5	6,50070989	0,346443728
4	-0,3774	-0,16285	-0,270125	25	8	7,24478096	0,078726438
5	-0,16285	0,0517	-0,055575	37	12	7,06410419	3,448854483
6	0,0517	0,26625	0,158975	39	2	6,02637386	2,690122923
7	0,26625	1000	500,13313	50	11	10,5247294	0,021462035

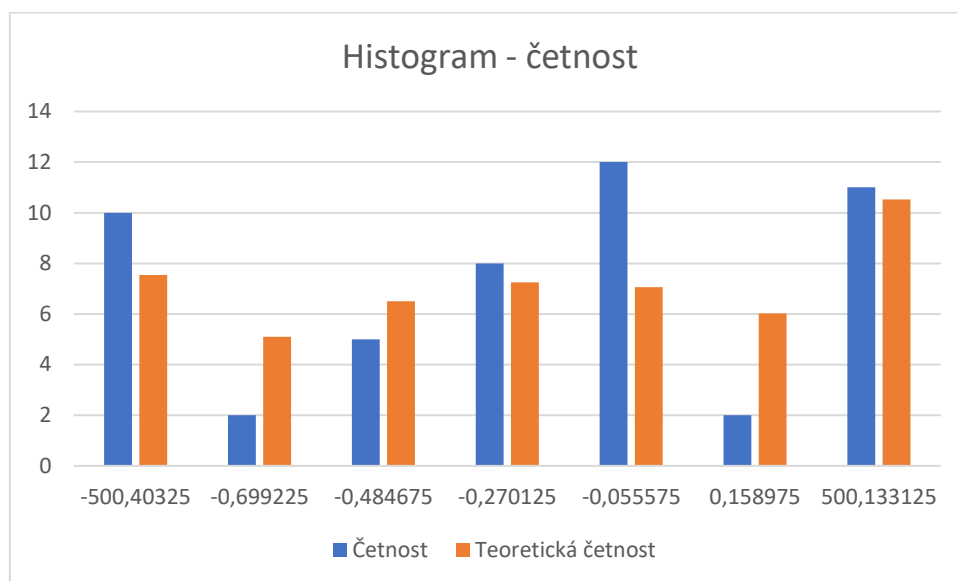
Testovací kritérium:

$$t = \sum_{j=1}^m \frac{(f_j - \hat{f}_j)^2}{\hat{f}_j} = 9,278573225$$

$\chi^2_{1-\alpha}$ pro $k = 7-2-1$ stupňů volnosti: 9,487729037

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

Protože $t \in \overline{W}_\alpha$, tedy hypotéza: $X \sim N(-0,2034; 0,3406229)$ se **nezamítá**.



- e) Za předpokladu (bez ohledu na výsledek části d)), že statistický soubor byl získán náhodným výběrem z normálního rozdělení, určete intervalové odhady střední hodnoty, rozptylu a směrodatné odchylky se spolehlivostí 0,95 a 0,99.

Předpoklad: $X \sim N(\mu, \sigma^2)$, σ^2 – neznámé

Bodový odhad střední hodnoty:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = -0,2034$$

Bodový odhad rozptylu:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 0,3406229$$

Bodový odhad směrodatné odchylky:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = 0,58362908$$

Intervalový odhad parametru μ :

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

0,995 kvantil Studentova rozdělení $t_{1-\alpha/2}$ s $k = n - 1 = 50 - 1 = 49$ stupni volnosti = 2,679951964

$$\alpha = 0,05: \langle \bar{x} - t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}; \bar{x} + t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \rangle = \langle -0,369265548; -0,03753445 \rangle$$

$$\alpha = 0,01: \langle \bar{x} - t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}; \bar{x} + t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \rangle = \langle -0,424596845; 0,017796845 \rangle$$

Intervalový odhad parametru σ^2 :

0,975 kvantil Pearsova rozdělení $\chi^2_{\alpha/2}$ s $k = n - 1 = 50 - 1 = 49$ stupni volnosti = 31,55491646

0,975 kvantil Pearsova rozdělení $\chi^2_{1-\alpha/2}$ s $k = n - 1 = 50 - 1 = 49$ stupni volnosti = 70,22241357

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

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

$$\alpha = 0,05: \langle \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}; \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} \rangle = \langle 0,237680837; 0,528935705 \rangle$$

$$\alpha = 0,01: \langle \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}; \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} \rangle = \langle 0,213350005; 0,612510851 \rangle$$

Intervalový odhad parametru σ :

$$\alpha = 0,05: \langle \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}; \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} \rangle = \langle \sqrt{0,237680837}; \sqrt{0,528935705} \rangle = \langle 0,4875252; 0,7272797 \rangle$$

$$\alpha = 0,01: \langle \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}; \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}} \rangle = \langle \sqrt{0,213350005}; \sqrt{0,612510851} \rangle = \langle 0,4618983; 0,7826307 \rangle$$

- f) Testujte hypotézu optimálního seřízení 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.

Studentův jednovýběrový test:

Testujeme hypotézu $H_0 : \mu = 0$:

$$\text{Testovací kritérium: } t = \frac{\bar{x} - \mu_0}{s} \sqrt{n} = \frac{\bar{x} - 0}{s} \sqrt{n} = -2,464330947$$

doplňek kritického oboru: $\bar{W}_\alpha = \langle -t_{1-\alpha/2}, t_{1-\alpha/2} \rangle$ pro alternativní hypotézu: $H_A : \mu \neq \mu_0$,

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

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

Protože $t \notin \bar{W}_\alpha$, tak hypotéza $H_0 : \mu = 0$ se **zamítá** a alternativní hypotéza $H_A : \mu \neq 0$ se **nezamítá**.

- g) Ověřte statistickým testem na hladině významnosti 0,05, zda seřízení 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 seřízení stroje a pak bylo naměřeno zbývajících 30 hodnot.

x1:20 - X		x21:50 - Y	
1	0,50	1	0,84
2	-0,04	2	-0,10
3	-0,77	3	0,60
4	-0,39	4	-1,20
5	-0,93	5	-0,93
6	-0,07	6	-1,08
7	0,85	7	-0,22
8	-0,70	8	-0,24
9	0,56	9	-0,91
10	-0,32	10	0,05
11	-0,87	11	-0,57
12	-0,11	12	-0,18
13	0,23	13	-1,45
14	-0,12	14	-0,02
15	-0,06	15	-0,51
16	0,00	16	0,28
17	-0,21	17	-1,31
18	0,13	18	-0,07
19	0,91	19	-0,25
20	-0,32	20	0,46
		21	-0,81
		22	-0,39
		23	0,29
		24	0,64
		25	-0,47
		26	-0,12
		27	0,63
		28	-1,07
		29	-0,06
		30	-0,27

	X	Y
n =	20	30
průměr =	-0,0865	-0,28133
rozptyl s ² =	0,258833	0,368612
směrodat. odch. =	0,508756	0,607134

Test rovnosti rozptylů – F-test:

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

$$\text{testovací kritérium: } t = \frac{s^2(X)}{s^2(Y)} = \frac{0,258833}{0,368612} = 0,702182951$$

doplňk kritického oboru: $\bar{W}_\alpha = \langle F_{\frac{\alpha}{2}}(n-1, m-1), F_{1-\frac{\alpha}{2}}(n-1, m-1) \rangle$ pro $H_A: \sigma_X^2 \neq \sigma_Y^2$,
 $F_{\frac{\alpha}{2}}(k_1, k_2), F_{1-\frac{\alpha}{2}}(k_1, k_2)$ jsou kvantily Fischerova-Snedecorova rozdělení s $k_1 = n-1$ a $k_2 = m-1$
stupni volnosti.

$$F_{\frac{\alpha}{2}}(19, 29) = 0,416329668$$

$$F_{1-\frac{\alpha}{2}}(19, 29) = 2,231273833$$

$$\bar{W}_\alpha = \langle F_{\frac{\alpha}{2}}(n-1, m-1), F_{1-\frac{\alpha}{2}}(n-1, m-1) \rangle = \langle 0,416329668, 2,231273833 \rangle$$

Protože $t \in \bar{W}_\alpha$, tedy hypotéza: $H_0 : \sigma_X^2 = \sigma_Y^2$ se **nezamítá**.

Studentův dvouvýběrový test:

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

$$\text{testovací kritérium: } t = \frac{\bar{x} - \bar{y} - \mu_0}{\sqrt{(n-1)s^2(X) + (m-1)s^2(Y)}} \sqrt{\frac{n*m(n+m)-2}{n+m}} = 1,183605658$$

doplňk kritického oboru: $\bar{W}_\alpha = \langle -t_{1-\alpha/2}, t_{1-\alpha/2} \rangle$ pro $H_A : \mu_X - \mu_Y \neq 0$,
 $t_{1-\alpha/2}$ - kvantil Studentova rozdělení s $k = n + m - 2 = 20 + 30 - 2 = 48$ stupni volnosti.

$$t_{1-\alpha/2} = 2,010634758$$

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

Protože $t \in \bar{W}_\alpha$, tedy hypotéza: $H_0 : \mu_X - \mu_Y = 0$ se **nezamítá**.

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.

12	
Př. 2	
X - Výška [cm]	Y - Váha [kg]
197	112
196	114
173	98
171	89
160	95
188	101
199	111
177	101
187	83
169	98
200	124
189	102
170	86
158	82
179	92
167	86
199	111
181	89
194	110
177	79

$$n = 20$$

$$\bar{x} = 181,55$$

$$\bar{y} = 98,15$$

$$\sum_{i=1}^n x_i^2 = 662641$$

$$\sum_{i=1}^n y_i^2 = 195649$$

$$\sum_{i=1}^n x_i y_i = 358823$$

-
- a) Vypočtěte bodový odhad koeficientu korelace.

$$r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{(\sum_{i=1}^n x_i^2 - n \bar{x}^2)(\sum_{i=1}^n y_i^2 - n \bar{y}^2)}} = 0,762904494$$

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

Testujeme hypotézu $H_0 : \rho = 0$:

$$\text{Testovací kritérium: } t = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} = 5,006455641$$

doplňk kritického oboru: $\bar{W}_\alpha = \langle 0, t_{1-\alpha/2} \rangle$ pro alternativní hypotézu: $H_0 : \rho \neq 0$,
 $t_{1-\frac{\alpha}{2}}(n-2) = t_{0,975}(20-2) = 2,100922037$

Protože $t \notin \bar{W}_\alpha$, tedy hypotéza $H_0 : \rho = 0$ se **zamítá**.

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

Pomocné výpočty:

	xi	yi	xi^2	yi^2	xi*yi
	197	112	38809	12544	22064
	196	114	38416	12996	22344
	173	98	29929	9604	16954
	171	89	29241	7921	15219
	160	95	25600	9025	15200
	188	101	35344	10201	18988
	199	111	39601	12321	22089
	177	101	31329	10201	17877
	187	83	34969	6889	15521
	169	98	28561	9604	16562
	200	124	40000	15376	24800
	189	102	35721	10404	19278
	170	86	28900	7396	14620
	158	82	24964	6724	12956
	179	92	32041	8464	16468
	167	86	27889	7396	14362
	199	111	39601	12321	22089
	181	89	32761	7921	16109
	194	110	37636	12100	21340
	177	79	31329	6241	13983
Suma	3631	1963	662641	195649	358823
Průměr	182	98			

Tedy:

$$n = 20, \sum_{i=1}^n x_i = 3631, \sum_{i=1}^n y_i = 1963, \sum_{i=1}^n x_i^2 = 662641,$$

$$\sum_{i=1}^n y_i^2 = 195649, \sum_{i=1}^n x_i y_i = 358823$$

$$\det(H) = n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 = 68659$$

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

$$b_2 = \frac{1}{\det(H)} \left(n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i \right) = 0,710860921$$

$$b_1 = \bar{y} - b_2 \bar{x} = -30,90680027$$

$$y = b_1 + b_2 x = -30,90680027 + 0,710860921x$$

$$S_{min}^* = \sum_{i=1}^n y_i^2 - b_1 \sum_{i=1}^n y_i - b_2 \sum_{i=1}^n x_i y_i = 1245,800551$$

$$s^2 = \frac{S_{min}^*}{n-2} = \frac{S_{min}^*}{20-2} = 69,2111417$$

2) Na hladině významnosti 0,05 otestujte hypotézy:

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

$$h^{11} = \frac{\sum_{i=1}^n x_i^2}{\det(H)} = 9,65118921$$

$$t = \frac{b_1 - (-100)}{s\sqrt{h^{11}}} = 2,673356541$$

$$t_{1-\frac{\alpha}{2}}(n-2) = t_{0,975}(20-2) = 2,100922037$$

$t \notin \bar{W} = \langle -2,100922037; 2,100922037 \rangle$, a tedy $H : \beta_0 = -100$ se **zamítá**.

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

$$h^{22} = \frac{n}{\det(H)} = 0,000291295$$

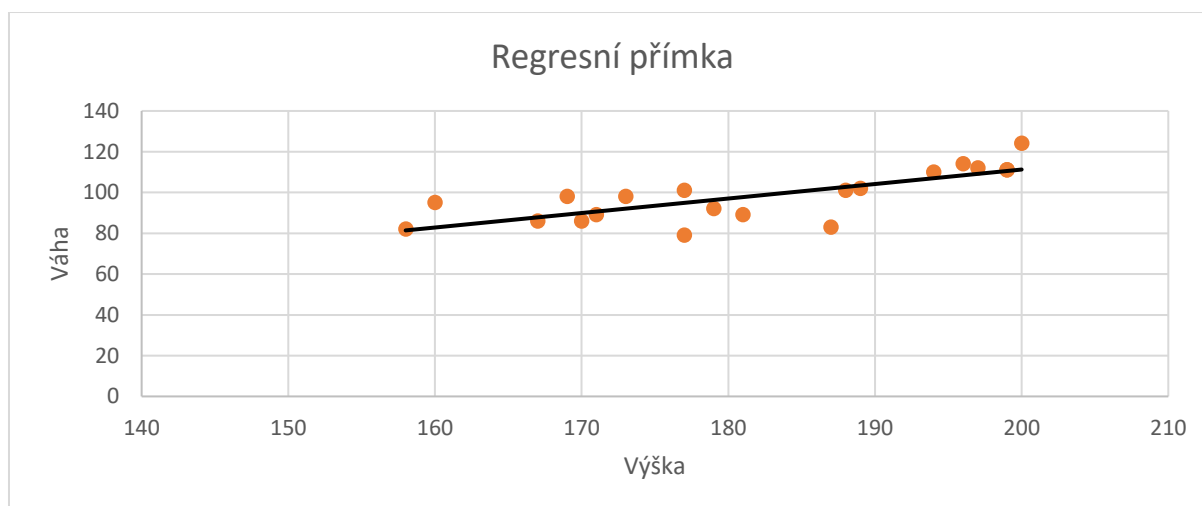
$$t = \frac{b_2 - 1}{s\sqrt{h^{22}}} = -2,03635047$$

$$t_{1-\frac{\alpha}{2}}(n-2) = t_{0,975}(20-2) = 2,100922037$$

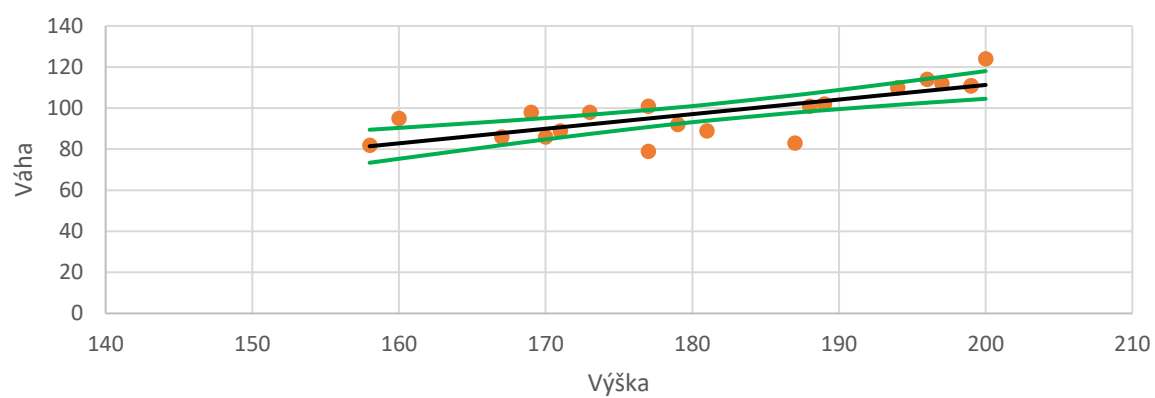
$t \in \bar{W} = \langle -2,100922037; 2,100922037 \rangle$, a tedy $H : \beta_1 = 1$ se **nezamítá**.

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

xi	yi	Střední y		individuální y		h*
		dolní	horní	dolní	horní	
158	81,40923	73,37013	89,44832	62,17082	100,6476	0,211553
160	82,83095	75,30762	90,35427	63,80229	101,8596	0,185278
167	87,80697	81,96631	93,64763	69,37867	106,2353	0,111668
169	89,2287	83,81665	94,64074	70,93172	107,5257	0,09588
170	89,93956	84,72941	95,1497	71,70128	108,1778	0,088859
171	90,65042	85,63255	95,66828	72,46613	108,8347	0,082422
173	92,07214	87,40527	96,73901	73,98156	110,1627	0,071294
177	94,91558	90,77835	99,05282	76,95435	112,8768	0,056031
177	94,91558	90,77835	99,05282	76,95435	112,8768	0,056031
179	96,3373	92,35571	100,3189	78,41128	114,2633	0,051894
181	97,75903	93,84733	101,6707	79,8484	115,6697	0,050088
187	102,0242	97,79127	106,2571	84,04068	120,0077	0,058652
188	102,7351	98,37884	107,0913	84,72212	120,748	0,062119
189	103,4459	98,94997	107,9419	85,39868	121,4931	0,066168
194	107,0002	101,6088	112,3917	88,70932	125,2911	0,095151
196	108,4219	102,6034	114,2405	90,00064	126,8432	0,110823
197	109,1328	103,09	115,1756	90,63942	127,6262	0,119533
199	110,5545	104,0452	117,0639	91,9035	129,2055	0,1387
199	110,5545	104,0452	117,0639	91,9035	129,2055	0,1387
200	111,2654	104,5151	118,0156	92,52892	130,0019	0,149157



Pás spolehlivosti pro střední hodnotu



Pás spolehlivosti pro individuální hodnotu

