



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta informačních technologií

Projekt z MSP

Zpracovala: Alena Tesařová (xtesar36)

Číslo zadání: 27, 8

Cvičení – skupina: čtvrtek, 11:00

Datum: 30. 11. 2019

Zadání projektu z předmětu MSP

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.
 - a. Proveďte roztržidění statistického souboru, vytvořte tabulku četností a nakreslete histogramy pro relativní četnosti a relativní kumulativní četnosti.
 - b. Vypočítejte aritmetický průměr, medián, modus, rozptyl a směrodatnou odchylku.
 - c. Vypočítejte bodové odhady střední hodnoty, rozptylu a směrodatné odchylky.
 - d. Testujte předpoklad o výběru z normálního rozdělení Pearsonovým (chí-kvadrát) testem na hladině významnosti 0,05.
 - 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.
 - 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.
 - 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.

Návod: Oba soubory zpracujte neroztržiděné. Testujte nejprve rovnost rozptylů odchylek před a po seřízení stroje. Podle výsledku pak zvolte vhodný postup pro testování rovnosti středních hodnot odchylek před a po seřízení stroje.

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.
 - a. Vypočítejte bodový odhad koeficientu korelace.
 - b. Na hladině významnosti 0,05 testujte hypotézu, že náhodné veličiny Výška a Váha jsou lineárně nezávislé.
 - c. **Regresní analýza** - data proložte přímkou: $Váha = \beta_0 + \beta_1 \cdot Výška$
 - i. Bodově odhadněte β_0, β_1 a rozptyl s^2 .
 - ii. Na hladině významnosti 0,05 otestujte hypotézy:
 $H: \beta_0 = -100, \quad H_A: \beta_0 \neq -100,$
 $H: \beta_1 = 1, \quad H_A: \beta_1 \neq 1,$
 - iii. Vytvořte graf bodů spolu s regresní přímkou a pásem spolehlivosti pro individuální hodnotu výšky.

Termín pro odevzdání práce je 11 týden výuky zimního semestru ve cvičení.

Vypracování

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

1	2,13	26	-0,53
2	-0,66	27	2,48
3	1,06	28	2,25
4	-0,15	29	0,21
5	0,80	30	0,58
6	1,87	31	0,35
7	-0,08	32	0,26
8	1,29	33	0,43
9	1,07	34	1,38
10	1,21	35	2,29
11	1,91	36	0,49
12	1,10	37	2,31
13	0,64	38	2,82
14	3,07	39	0,62
15	0,26	40	0,94
16	0,16	41	1,40
17	1,00	42	0,71
18	1,39	43	3,44
19	0,81	44	1,61
20	-0,55	45	1,94
21	0,67	46	2,64
22	0,60	47	1,78
23	1,67	48	1,14
24	2,14	49	1,51
25	1,56	50	1,05

Uspořádaný soubor

1	-0,66	26	1,10
2	-0,55	27	1,14
3	-0,53	28	1,21
4	-0,15	29	1,29
5	-0,08	30	1,38
6	0,16	31	1,39
7	0,21	32	1,40
8	0,26	33	1,51
9	0,26	34	1,56
10	0,35	35	1,61
11	0,43	36	1,67
12	0,49	37	1,78
13	0,58	38	1,87
14	0,60	39	1,91
15	0,62	40	1,94
16	0,64	41	2,13
17	0,67	42	2,14
18	0,71	43	2,25
19	0,80	44	2,29
20	0,81	45	2,31
21	0,94	46	2,48
22	1,00	47	2,64
23	1,05	48	2,82
24	1,06	49	3,07
25	1,07	50	3,44

- a) Proveďte roztržení statistického souboru, vytvořte tabulku četností a nakreslete histogramy pro relativní četnosti a relativní kumulativní četnosti.

$$x_{(1)} = \min x_i = -0,66$$

$$x_{(n)} = \max x_i = 3,44$$

$$\text{Varianční obor: } \langle x_{(1)} | x_{(n)} \rangle = \langle -0,66 | 3,44 \rangle$$

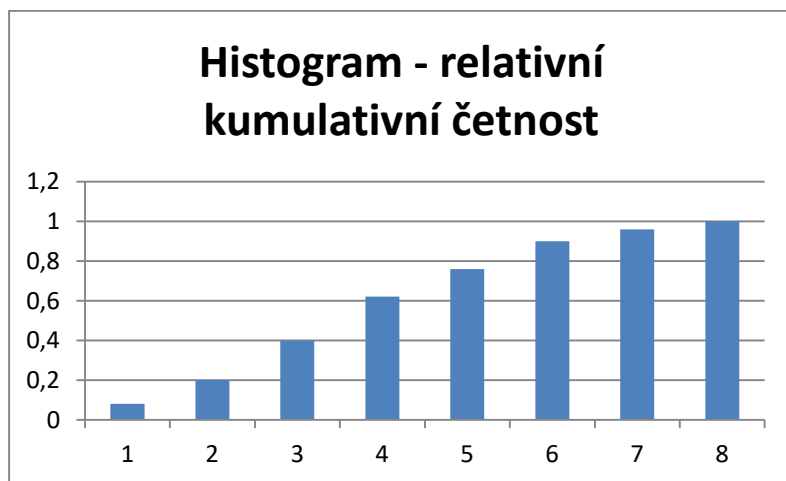
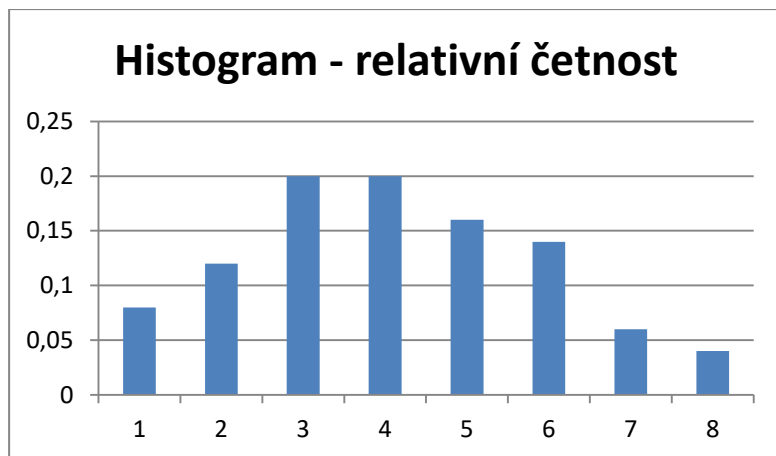
$$\text{Rozpětí: } x_{(n)} - x_{(1)} = 4,1$$

Počet tříd: 8 (zvolíme)

Délka tříd: 0,5125

tříd a	xi-	xi+	střed tříd	Kumulat čet.	četnos t	Relat. Čet.	Relat. Kum. Čet.
1	-0,6600	-0,1475	-0,4038	4	4	0,08	0,08
2	-0,1475	0,3650	0,1088	10	6	0,12	0,2
3	0,3650	0,8775	0,6213	20	10	0,2	0,4

4	0,8775	1,3900	1,1338	30	10	0,2	0,6
5	1,3900	1,9025	1,6463	38	8	0,16	0,76
6	1,9025	2,4150	2,1588	45	7	0,14	0,9
7	2,4150	2,9275	2,6713	48	3	0,06	0,96
8	2,9275	3,4400	3,1838	50	2	0,04	1



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

Aritmetický průměr: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 1,1814$

Medián: $\tilde{x} = 1,085$

Modus: $\hat{x} = 0,26$

Rozptyl: $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = 0,87$

Směrodatná odchylka: $s = \sqrt{s^2} = 0,93$

c) Vypočtete bodové odhady střední hodnoty, rozptylu a směrodatné odchylky
bodový odhad střední hodnoty: $\mu = \bar{x} = 1,1814$

bodový odhad rozptylu: $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 0,891943$

bodový odhad směrodatné odchylky: $s = \sqrt{s^2} = 0,944427$

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

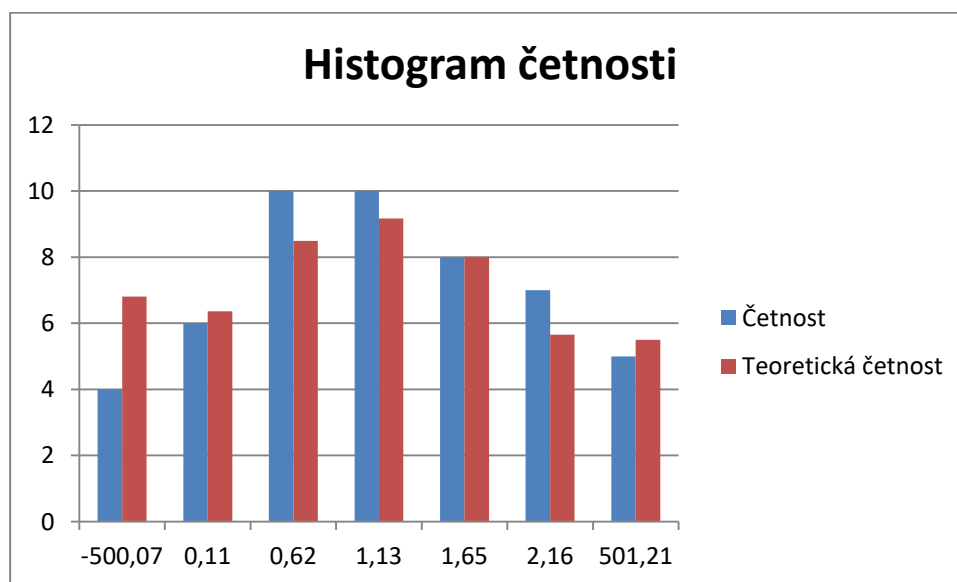
Třída	xi-	xi+	Střed třídy	četnost	Kumulativní četnost	Teoretická čet.	roz ² /teor. čet.
1	-1000	-0,1475	-500,07375	4	4	3,984971972	5,66733E-05
2	-0,1475	0,365	0,10875	6	10	5,698694218	0,015930873
3	0,365	0,8775	0,62125	10	20	9,0067741	0,109528415
4	0,8775	1,39	1,13375	10	30	10,67980785	0,043272194
5	1,39	1,9025	1,64625	8	38	9,501090049	0,237159244
6	1,9025	2,415	2,15875	7	45	6,341449708	0,068389486
7	2,415	1000	501,2075	5	50	4,787212102	0,009458258

Testovací kritérium: $t = \sum_{j=1}^m \frac{(f_j - \hat{f}_j)^2}{\hat{f}_j} = 0,483795144$

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

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

hypotézu $X \sim N(1,18; 0,944427)$ **nezamítáme** protože $t \in \overline{W}_\alpha$



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: $X \sim N(\mu, \sigma^2), \sigma^2$

bodový odhad střední hodnoty: $\mu = \bar{x} = 1,1814$

bodový odhad rozptylu: $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 0,891943$

bodový odhad směrodatné odchylky: $s = \sqrt{s^2} = 0,944427$

Intervalový odhad parametru μ :

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

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

$$\alpha = 0,05: \langle \bar{x} - t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}; \bar{x} + t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \rangle = \langle 0,912996735; 1,449803265 \rangle$$

$$\alpha = 0,01: \langle \bar{x} - t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}; \bar{x} + t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \rangle = \langle 0,823459753; 1,539340247 \rangle$$

Intervalový odhad parametru σ^2 :

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

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

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

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

$$\alpha = 0,05: \langle \frac{(n-1)s^2}{\chi^2_{1-a/2}}; \frac{(n-1)s^2}{\chi^2_{a/2}} \rangle = \langle 0,62238251; 1,385052059 \rangle$$

$$\alpha = 0,01: \langle \frac{(n-1)s^2}{\chi^2_{1-a/2}}; \frac{(n-1)s^2}{\chi^2_{a/2}} \rangle = \langle 0,558670669; 1,603898928 \rangle$$

Intervalový odhad parametru σ :

$$\alpha = 0,05: \langle \sqrt{\frac{(n-1)s^2}{\chi^2_{1-a/2}}}; \sqrt{\frac{(n-1)s^2}{\chi^2_{a/2}}} \rangle = \langle 0,788912232; 1,176882347 \rangle$$

$$\alpha = 0,01: \langle \sqrt{\frac{(n-1)s^2}{\chi^2_{1-a/2}}}; \sqrt{\frac{(n-1)s^2}{\chi^2_{a/2}}} \rangle = \langle 0,747442753; 1,266451313 \rangle$$

- 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

Test hypotézy $H_0 : \mu = 0$

$$\text{Testovací kritérium: } t = \frac{\bar{x} - \mu_0}{s} \sqrt{n} = \frac{\bar{x} - 0}{s} \sqrt{n} = 8,8453178$$

Doplňek kritického oboru: $\bar{W}_a = \langle -t_{1-\alpha/2}; t_{1-\alpha/2} \rangle$ pro alternativní hypotézu: $H_A : \mu \neq 0$

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

$$\bar{W}_{0,05} = \langle -2,0095752; 2,095752 \rangle \Rightarrow \text{hypotézu } H_0 \text{ zamítáme protože } t \notin \bar{W}_a$$

- 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	2,13	1	0,67
2	-0,66	2	0,60
3	1,06	3	1,67
4	-0,15	4	2,14
5	0,80	5	1,56
6	1,87	6	-0,53
7	-0,08	7	2,48
8	1,29	8	2,25
9	1,07	9	0,21
10	1,21	10	0,58
11	1,91	11	0,35
12	1,10	12	0,26
13	0,64	13	0,43
14	3,07	14	1,38
15	0,26	15	2,29
16	0,16	16	0,49
17	1,00	17	2,31
18	1,39	18	2,82
19	0,81	19	0,62
20	-0,55	20	0,94
		21	1,40
		22	0,71
		23	3,44
		24	1,61
		25	1,94
		26	2,64
		27	1,78
		28	1,14
		29	1,51
		30	1,05

	X	Y
n	20	30
průměr	0,92	1,36
Odhad bodového rozptylu s²	0,863792368	0,860485517
směr. odchylka	0,929404308	0,927623586

Test rovnosti rozptylů – F test

Testujeme hypotézu: $H_0 : \sigma^2_x = \sigma^2_y$

$$\text{Testovací kritérium: } t = \frac{s^2(X)}{s^2(Y)} = \frac{0,863792368}{0,860485517} = 1,003843006$$

Doplňek kritického oboru: $\bar{W}_a = \langle F_{a/2}(n-1, m-1), F_{1-a/2}(n-1, m-1) \rangle$ pro $H_A : \sigma^2_x \neq \sigma^2_y$

$F_{a/2}(k1, k2)$ a $F_{1-a/2}(k1, k2)$ jsou kvantily Fisherova-Snedecorova rozdělení s $k1 = n-1$ a $k2 = m-1$ stupni volnosti.

$$F_{a/2}(19, 29) = 0,4163297$$

$$F_{1-a/2}(19, 29) = 2,231274$$

$$\bar{W}_a = \langle F_{a/2}(n-1, m-1), F_{1-a/2}(n-1, m-1) \rangle = \langle 0,4163297; 2,231274 \rangle$$

\Rightarrow hypotézu H_0 **nezamítáme** protože $t \in \bar{W}_a$

Studentův dvouvýběrový test

Testujeme hypotézu $H_0 : \mu_x - \mu_y = 0$ za podmínky $\sigma^2_x = \sigma^2_y$

$$\text{Testovací kritérium: } t = \frac{\bar{x} - \bar{y} - \mu_0}{\sqrt{\frac{(n-1)s^2(X) + (m-1)s^2(Y)}{n+m}}} \sqrt{\frac{n*m(n+m-2)}{n+m}} = -1,647477443$$

Doplňek kritického oboru $\bar{W}_a = \langle -t_{1-\alpha/2}; t_{1-\alpha/2} \rangle$ 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.

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

$$\bar{W}_a = \langle -t_{1-\alpha/2}; t_{1-\alpha/2} \rangle = \langle -2,010634758; 2,01063475 \rangle$$

\Rightarrow hypotézu H_0 **nezamítáme** protože $t \in \bar{W}_a$

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]
173	88
168	32
188	83
196	88
150	40
176	63
181	96
168	59
195	102
174	53
158	37
153	40
187	66
198	88
174	68
166	63
199	100
182	80
156	30
166	42

$$n = 20$$

$$\bar{x} = 175,4$$

$$\bar{y} = 65,873$$

$$\sum_{i=1}^n x_i^2 = 619670$$

$$\sum_{i=1}^n y_i^2 = 97577,713$$

$$\sum_{i=1}^n x_i y_i = 237002,882$$

- a) Vypočítejte bodový odhad koeficientu korelace:

$$r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{(\sum_{i=1}^n x_i^2 - n \bar{x}^2)(\sum_{i=1}^n y_i^2 - n \bar{y}^2)}} = 0,86221773$$

- 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: $t = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} = 7,22173094$

Doplňk kritického oboru: $\bar{W}_a = \langle 0; t_{1-\alpha/2}(n-2) \rangle = \langle 0; 2,100922037 \rangle$ pro alternativní hypotézu:

$H_A: \rho \neq 0$

\Rightarrow hypotézu H_0 **zamítáme** protože $t \notin \bar{W}_a$

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

	Výška xi	Váha yi	xi^2	yi^2	xi*yi
	173	88	29929	7708	15188,273
	168	32	28224	1044	5428,5039
	188	83	35344	6873	15586,019
	196	88	38416	7775	17282,029
	150	40	22500	1575	5952,1526
	176	63	30976	3916	11013,615
	181	96	32761	9277	17433,34
	168	59	28224	3424	9830,4406
	195	102	38025	10410	19896,09
	174	53	30276	2775	9165,6619
	158	37	24964	1404	5921,2424
	153	40	23409	1638	6192,5838
	187	66	34969	4372	12364,064
	198	88	39204	7832	17522,335
	174	68	30276	4645	11858,696
	166	63	27556	3913	10384,26
	199	100	39601	10001	19900,895
	182	80	33124	6375	14531,002
	156	30	24336	873	4607,9715
	166	42	27556	1750	6943,7066
Suma:	3508	1317	619670	97578	237003
Průměr:	175	66			

$$\text{Det}(\mathbf{H}) = n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2 = 87336$$

1) Bodově odhadněte β_0, β_1, s^2

$$b_2 = \frac{1}{\text{det}(\mathbf{H})} (n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i) = \mathbf{1,35542956}$$

$$b_1 = \bar{y} - b_2 \bar{x} = \mathbf{-171,86891}$$

$$y = b_1 + b_2 x = -171,86891 + 1,35542956 x$$

$$s_{min}^* = \sum_{i=1}^n y_i^2 - b_1 \sum_{i=1}^n y_i - b_2 \sum_{i=1}^n x_i y_i = 2768,89893$$

$$s^2 = \frac{s_{min}^*}{n-2} = \mathbf{153,827718}$$

2) Na hladině významnosti 0,05 otestujte hypotézy:

$$H_0: \beta_0 = -100, H_A: \beta_0 \neq -100$$

$$h^{11} = \frac{\sum_{i=1}^n x_i^2}{\det(H)} = 7,09524137$$

$$t = \frac{b_1 - (-100)}{s\sqrt{h^{11}}} = -2,1754054$$

$$\bar{W}_a = \langle -t_{1-\alpha/2}(n-2); t_{1-\alpha/2}(n-2) \rangle = \langle -2,100922037; 2,100922037 \rangle$$

\Rightarrow hypotézu H_0 **zamítáme** protože $t \notin \bar{W}_a$

$$H_0: \beta_1 = 1, H_A: \beta_1 \neq 1$$

$$h^{22} = \frac{n}{\det(H)} = 0,000229$$

$$t = \frac{b_2 - (1)}{s\sqrt{h^{22}}} = 1,89372927$$

$$\bar{W}_a = \langle -t_{1-\alpha/2}(n-2); t_{1-\alpha/2}(n-2) \rangle = \langle -2,100922037; 2,100922037 \rangle$$

\Rightarrow hypotézu H_0 **nezamítáme** protože $t \in \bar{W}_a$

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

Výpočet pásu spolehlivosti

		střední y		individuální y		
xi	yi	dolní	horní	dolní	horní	h*
173	88	56,717484	68,52331389	35,90298013	89,337818	0,051319
168	32	49,326871	62,35963182	28,98362943	82,702873	0,0625401
188	83	75,294574	90,60911054	55,79286922	110,11082	0,0863561
196	88	83,798738	103,7918195	65,88637536	121,70418	0,1471787
150	40	19,858366	43,03267268	2,928187775	59,962851	0,1977421
176	63	60,855326	72,51804944	39,98498648	93,388389	0,0500824
181	96	67,232877	79,69479403	46,67202866	100,25564	0,0571815
168	59	49,326871	62,35963182	28,98362943	82,702873	0,0625401
195	102	82,760992	102,1187072	64,64315239	120,23655	0,1379729
174	53	58,123175	69,82848245	37,26946931	90,682188	0,0504488
158	37	33,287637	51,2902741	14,72086177	69,85705	0,1193322
153	40	24,930436	46,09318016	7,388125581	63,63549	0,1649034
187	66	74,188917	89,00390872	54,5068011	108,68602	0,0808143
198	88	85,858847	107,1534293	68,35758772	124,65469	0,1669644
174	68	58,123175	69,82848245	37,26946931	90,682188	0,0504488

166	63	46,226775	60,03800898	26,17569329	80,089091	0,0702345
199	100	86,88212	108,8410149	69,58570623	126,13743	0,1775442
182	80	68,437904	81,20062643	47,99208029	101,64645	0,0599753
156	30	29,962094	49,19409885	11,8032236	67,35297	0,1361867
166	42	46,226775	60,03800898	26,17569329	80,089091	0,0702345

