



Projekt z MSP

Číslo zadání: 24, 6

Cvičení – skupina: středa, 8:00

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	1,38	26	0,78
2	0,72	27	-0,65
3	0,18	28	0,39
4	-0,11	29	0,49
5	1,05	30	0,86
6	1,45	31	1,88
7	-0,28	32	0,47
8	2,49	33	0,51
9	-0,54	34	0,93
10	1,88	35	3,21
11	2,37	36	0,60
12	0,95	37	0,40
13	-0,55	38	1,14
14	2,05	39	2,82
15	1,29	40	1,07
16	0,07	41	1,32
17	1,59	42	0,66
18	-0,20	43	-0,53
19	-0,79	44	-0,76
20	0,37	45	2,80
21	0,95	46	0,40
22	0,97	47	0,70
23	-2,28	48	-0,68
24	2,88	49	0,19
25	0,78	50	-0,48

Uspořádaný statistický soubor

(1)	-2,28	(26)	0,72
(2)	-0,79	(27)	0,78
(3)	-0,76	(28)	0,78
(4)	-0,68	(29)	0,86
(5)	-0,65	(30)	0,93
(6)	-0,55	(31)	0,95
(7)	-0,54	(32)	0,95
(8)	-0,53	(33)	0,97
(9)	-0,48	(34)	1,05
(10)	-0,28	(35)	1,07
(11)	-0,20	(36)	1,14
(12)	-0,11	(37)	1,29
(13)	0,07	(38)	1,32
(14)	0,18	(39)	1,38
(15)	0,19	(40)	1,45
(16)	0,37	(41)	1,59
(17)	0,39	(42)	1,88
(18)	0,40	(43)	1,88
(19)	0,40	(44)	2,05
(20)	0,47	(45)	2,37
(21)	0,49	(46)	2,49
(22)	0,51	(47)	2,80
(23)	0,60	(48)	2,82
(24)	0,66	(49)	2,88
(25)	0,70	(50)	3,21

a) Proveďte roztřídění statistického souboru, vytvořte tabulku četností a nakreslete histogramy pro relevantní četnosti a relevantní kumulativní četnosti.

počet prvků statistického souboru $n = 50$

$$x_{(1)} = \min_i x_i = -2,28$$

$$x_{(n)} = \max_i x_i = 3,21$$

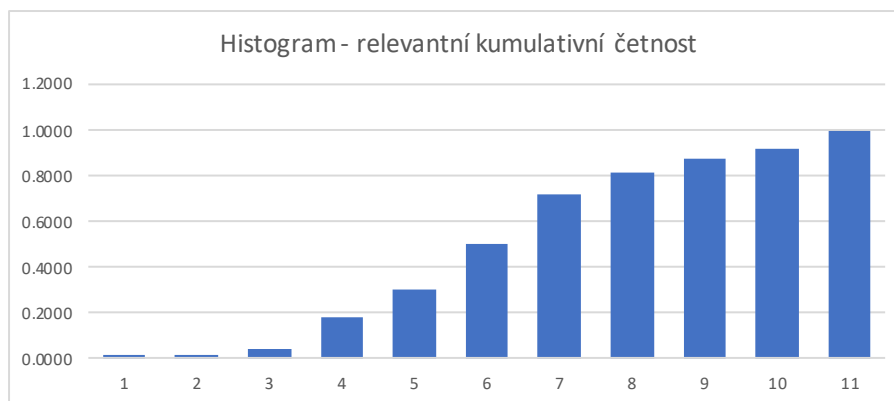
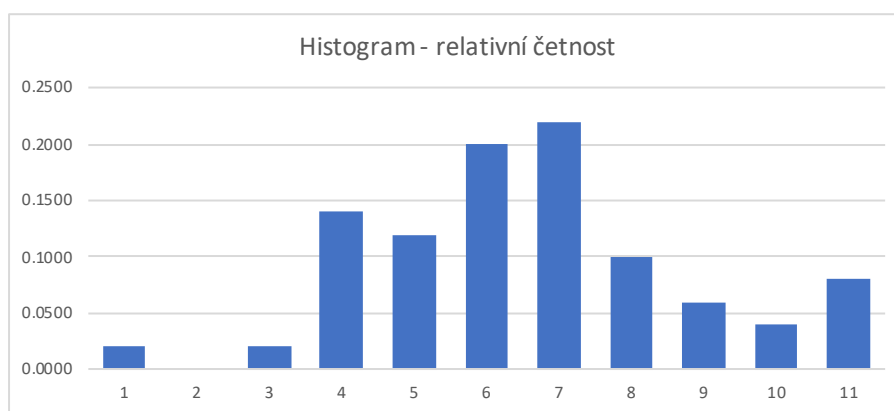
$$\text{variační obor: } \langle x_{(1)}, x_{(n)} \rangle = \langle -2,28; 3,21 \rangle$$

$$\text{rozpětí: } x_{(n)} - x_{(1)} = 5,49$$

počet tříd $m = 11$ (zvoleno)

$$\text{délka třídy} = \frac{x_{(n)} - x_{(1)}}{m} = 0,49909$$

třída	$x_i -$	$x_i +$	střed třídy	kum. četnost	četnost	rel. četnost	rel. kum. četnost
1	-2,2800	-1,7809	-2,0305	1	1	0,02	0,02
2	-1,7809	-1,2809	-1,5309	1	0	0,00	0,02
3	-1,2809	-0,7809	-1,0309	2	1	0,02	0,04
4	-0,7809	-0,2809	-0,5309	9	7	0,14	0,18
5	-0,2809	0,2191	-0,0309	15	6	0,12	0,30
6	0,2191	0,7191	0,4691	25	10	0,20	0,50
7	0,7191	1,2191	0,9691	36	11	0,22	0,72
8	1,2191	1,7191	1,4691	41	5	0,10	0,82
9	1,7191	2,2191	1,9691	44	3	0,06	0,88
10	2,2191	2,7191	2,4691	46	2	0,04	0,92
11	2,7191	3,2100	2,9645	50	4	0,08	1,00



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

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 0,7438$$

$$\text{medián: } \hat{x} = 0,71$$

$$\text{modus: } \hat{x} = 0,4$$

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = 1,212$$

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} = 1,1009$$

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

bodový odhad střední hodnoty: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 0,7438$

bodový odhad rozptylu: $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 1,2368$

bodový odhad směrodatné odchylky: $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = 1,1121$

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

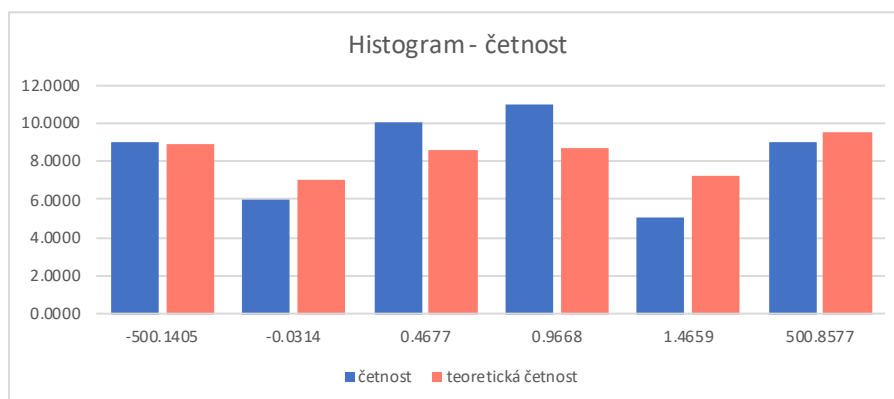
třída	$x_i -$	$x_i +$	střed třídy	kum. čet.	četnost	teor. čet.	rozdíl ² /teor.cet.
1	-1000,0000	-0,2809	-500,1405	9	9	8,9208	0,0007
2	-0,2809	0,2182	-0,0314	15	6	6,9910	0,1405
3	0,2182	0,7173	0,4677	25	10	8,6124	0,2236
4	0,7173	1,2164	0,9668	36	11	8,7036	0,6059
5	1,2164	1,7155	1,4659	41	5	7,2153	0,6802
6	1,7155	1000,0000	500,8577	50	9	9,5569	0,0324

testovací kritérium: $t = \sum_{j=1}^m \frac{(f_j - \hat{f}_j)^2}{\hat{f}_j} = 1,6833$

$\chi^2_{1-\alpha}$ pro $k = 6 - 2 - 1$ stupňů volnosti: 7,8147

doplňek kritického oboru $\bar{W}_\alpha = \langle 0, \chi^2_{1-\alpha} \rangle = \langle 0; 7,8147 \rangle$

Protože $t \in \bar{W}_\alpha$, tedy hypotéza: $X \sim N(0,7438; 1,1121)$ 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: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 0,7438$

bodový odhad rozptylu: $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 1,2368$

bodový odhad směrodatné odchylky: $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = 1,1121$

Intervalový odhad parametru μ :

0,975 kvantil Studentova rozdělení $t_{1-\frac{\alpha}{2}}$ s $k = n - 1 = 50 - 1 = 49$ stupni volnosti = 2,0096

0,995 kvantil Studentova rozdělení $t_{1-\frac{\alpha}{2}}$ s $k = n - 1 = 50 - 1 = 49$ stupni volnosti = 2,68

$\alpha = 0,05$: $\left\langle \bar{x} - t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}; \bar{x} + t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right\rangle = \langle 0,4277; 1,0599 \rangle$

$\alpha = 0,01$: $\left\langle \bar{x} - t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}; \bar{x} + t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right\rangle = \langle 0,3223; 1,1653 \rangle$

Intervalový odhad parametru σ^2 :

0,975 kvantil Pearsonova rozdělení $\chi^2_{\frac{\alpha}{2}}$ s $k = n - 1 = 50 - 1 = 49$ stupni volnosti = 31,5549

0,975 kvantil Pearsonova rozdělení $\chi^2_{1-\frac{\alpha}{2}}$ s $k = n - 1 = 50 - 1 = 49$ stupni volnosti = 70,2224

0,995 kvantil Pearsonova rozdělení $\chi^2_{\frac{\alpha}{2}}$ s $k = n - 1 = 50 - 1 = 49$ stupni volnosti = 27,2493

0,995 kvantil Pearsonova rozdělení $\chi^2_{1-\frac{\alpha}{2}}$ s $k = n - 1 = 50 - 1 = 49$ stupni volnosti = 78,2307

$\alpha = 0,05$: $\left\langle \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}; \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}} \right\rangle = \langle 0,7760; 1,7269 \rangle$

$\alpha = 0,01$: $\left\langle \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}; \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}} \right\rangle = \langle 0,6966; 1,9998 \rangle$

Intervalový odhad parametru σ :

$\alpha = 0,05$: $\left\langle \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}}; \sqrt{\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}} \right\rangle = \langle \sqrt{0,7760}; \sqrt{1,7269} \rangle = \langle 0,8809; 1,3141 \rangle$

$\alpha = 0,01$: $\left\langle \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}}; \sqrt{\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}} \right\rangle = \langle \sqrt{0,6966}; \sqrt{1,9998} \rangle = \langle 0,8346; 1,4141 \rangle$

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 $H_0 : \mu = 0$:

testovací kritérium: $t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \sqrt{n} = \frac{\bar{x} - 0}{\frac{s}{\sqrt{n}}} \sqrt{n} = 4,7293$

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

0,975 kvantil Studentova rozdělení $t_{1-\frac{\alpha}{2}}$ s $k = n - 1 = 50 - 1 = 49$ stupni volnosti = 2,0096

$\bar{W}_\alpha = \left\langle -t_{1-\frac{\alpha}{2}}, t_{1-\frac{\alpha}{2}} \right\rangle = \langle -2,0096; 2,0096 \rangle$

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 neměřeno zbývajících 30 hodnot.

$x \ 1 : 20 - X$		$x \ 21 : 50 - Y$	
1	1,38	21	0,9500
2	0,72	22	0,9700
3	0,18	23	-2,2800
4	-0,11	24	2,8800
5	1,05	25	0,7800
6	1,45	26	0,7800
7	-0,28	27	-0,6500
8	2,49	28	0,3900
9	-0,54	29	0,4900
10	1,88	30	0,8600
11	2,37	31	1,8800
12	0,95	32	0,4700
13	-0,55	33	0,5100
14	2,05	34	0,9300
15	1,29	35	3,2100
16	0,07	36	0,6000
17	1,59	37	0,4000
18	-0,20	38	1,1400
19	-0,79	39	2,8200
20	0,37	40	1,0700
		41	1,3200
		42	0,6600
		43	-0,5300
		44	-0,7600
		45	2,8000
		46	0,4000
		47	0,7000
		48	-0,6800
		49	0,1900
		50	-0,4800

	X	Y
n	20	30
průměr \bar{x}	0,7685	0,7273
rozptyl s^2	0,9994	1,3531
směrodatná odchylka s	0,9997	1,1632

Test rovnosti rozptylů – F-test:

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

testovací kritérium: $t = \frac{s^2(X)}{s^2(Y)} = \frac{0,9994}{1,3531} = 0,7386$

doplňek 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

$$F_{\frac{\alpha}{2}}(19, 20) = 0,4163$$

$$F_{1-\frac{\alpha}{2}}(19, 20) = 2,2313$$

$$\langle F_{\frac{\alpha}{2}}(n-1, m-1), F_{1-\frac{\alpha}{2}}(n-1, m-1) \rangle = \langle 0,4163; 2,2313 \rangle$$

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

Studentův dvouvýběrový test:

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

testovací kritérium: $t = \frac{\bar{x} - \bar{y} - \mu_0}{\sqrt{(n-1)s^2(X) + (m-1)s^2(Y)}} \sqrt{\frac{n \cdot m \cdot (n+m-2)}{n+m}} = 0,1295$

doplňek kritického oboru: $\bar{W}_\alpha = \langle -t_{1-\frac{\alpha}{2}}, t_{1-\frac{\alpha}{2}} \rangle$ pro $H_A : \mu_X - \mu_Y \neq 0$

$t_{1-\frac{\alpha}{2}}$ – kvantil Studentova rozdělení s $k = n + m - 2 = 20 + 30 - 2 = 48$ stupni volnosti

$t_{1-\frac{\alpha}{2}} = 2,0106$

$\bar{W}_\alpha = \langle -t_{1-\frac{\alpha}{2}}, t_{1-\frac{\alpha}{2}} \rangle = \langle -2,0106; 2,0106 \rangle$

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**.

X – Výška [cm]	Y – Váha [kg]
150	50,4397652181239
177	73,1879124409093
154	53,2231188041242
152	43,7631039035112
169	68,5086360050001
200	94,4644499790565
196	99,4245453695221
181	73,7032922682493
152	49,8882219051297
172	73,9958901681090
152	58,0373066459991
150	46,0077991234941
178	77,6318937228627
154	57,2090559679648
190	90,1798535187189
195	98,1416023757658
182	79,7087348561839
184	88,4058636653169
156	41,7956696089958
154	65,8182027724775

$n = 20$

$\bar{x} = 169,9$

$\bar{y} = 69,1767$

$\sum_{i=1}^n x_i^2 = 583236$

$\sum_{i=1}^n y_i^2 = 102343,9642$

$\sum_{i=1}^n x_i y_i = 241030,3296$

a) Vypočtete bodový odhad koeficientu korelace.

$$r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sqrt{\left(\sum_{i=1}^n x_i^2 - n \bar{x}^2\right) \left(\sum_{i=1}^n y_i^2 - n \bar{y}^2\right)}} = 0,9525$$

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: $t = \frac{|r| \sqrt{n-2}}{\sqrt{1-r^2}} = 13,27$

doplňk kritického oboru: $\bar{W}_\alpha = \langle 0, t_{1-\frac{\alpha}{2}} \rangle$ pro alternativní hypotézu: $H_A : \rho \neq 0$

$t_{1-\frac{\alpha}{2}}(n-2) = t_{0,975}(20-2) = 2,1009$

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

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

Pomocné výpočty:

x_i	y_i	x_i^2	y_i^2	$x_i y_i$	
150	50,4397652181239	22500	2544,1699	7565,9648	
177	73,1879124409093	31329	5356,4705	12954,2605	
154	53,2231188041242	23716	2832,7004	8196,3603	
152	43,7631039035112	23104	1915,2093	6651,9918	
169	68,5086360050001	28561	4693,4332	11577,9595	
200	94,4644499790565	40000	8923,5323	18892,8900	
196	99,4245453695221	38416	9885,2402	19487,2109	
181	73,7032922682493	32761	5432,1753	13340,2959	
152	49,8882219051297	23104	2488,8347	7583,0097	
172	73,9958901681090	29584	5475,3918	12727,2931	
152	58,0373066459991	23104	3368,3290	8821,6706	
150	46,0077991234941	22500	2116,7176	6901,1699	
178	77,6318937228627	31684	6026,7109	13818,4771	
154	57,2090559679648	23716	3272,8761	8810,1946	
190	90,1798535187189	36100	8132,4060	17134,1722	
195	98,1416023757658	38025	9631,7741	19137,6125	
182	79,7087348561839	33124	6353,4824	14506,9897	
184	88,4058636653169	33856	7815,5967	16266,6789	
156	41,7956696089958	24336	1746,8780	6520,1245	
154	65,8182027724775	23716	4332,0358	10136,0032	
3398	1383,53490000000	583236	102343,9642	241030,3296	suma
169,9	69,1767				průměr

Tedy:

$$\sum_{i=1}^n x_i = 3398, \sum_{i=1}^n y_i = 1383,5349, \sum_{i=1}^n x_i^2 = 583236, \sum_{i=1}^n y_i^2 = 102343,9642, \sum_{i=1}^n x_i y_i = 241030,3296$$

$$\det(H) = n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 = 118316$$

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

$$b_2 = \frac{1}{\det(H)} \left(n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i \right) = 1,0088$$

$$b_1 = \bar{y} - b_2 \bar{x} = -102,2152$$

$$y = b_1 + b_2 x = -102,2152 + 1,0088x$$

$$S_{\min}^* = \sum_{i=1}^n y_i^2 - b_1 \sum_{i=1}^n y_i - b_2 \sum_{i=1}^n x_i y_i = 615,3705$$

$$s^2 = \frac{S_{\min}^*}{n-2} = \frac{S_{\min}^*}{20-2} = 34,1872$$

2) Na hladině významnosti 0,05 otestujte hypotézy:

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

$$h^{11} = \frac{\sum_{i=1}^n x_i^2}{\det(H)} = 4,9295$$

$$t = \frac{b_1 - (-100)}{s\sqrt{h^{11}}} = -0,1706$$

$$t_{1-\frac{\alpha}{2}}(n-2) = t_{0,975}(20-2) = 2,1009$$

$t \in \overline{W} = \langle -2, 1009; 2, 1009 \rangle$, a tedy $H : \beta_0 = -100$ se **nezamítá**.

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

$$h^{22} = \frac{n}{\det(H)} = 0,0002$$

$$t = \frac{b_2 - 1}{s\sqrt{h^{22}}} = 0,1155$$

$$t_{1-\frac{\alpha}{2}}(n-2) = t_{0,975}(20-2) = 2,1009$$

$t \in \overline{W} = \langle -2, 1009; 2, 1009 \rangle$, a tedy $H : \beta_1 = 1$ se **nezamítá**.

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

Výpočet pásu spolehlivosti

X – Výška [cm]	Y – Váha [kg]	střední y		individuální y		h^*
		dolní	horní	dolní	horní	
150	49,1020	44,9013	53,3027	36,1195	62,0845	0,1169
177	76,3391	73,3674	79,3107	63,7007	88,9775	0,0585
154	53,1371	49,3963	56,8779	40,2961	65,9781	0,0927
152	51,1196	47,1550	55,0841	38,2116	64,0275	0,1042
169	68,2688	65,5183	71,0194	55,6806	80,8571	0,0501
200	99,5411	94,0044	105,0778	86,0669	113,0152	0,2032
196	95,5059	90,5138	100,4980	82,2462	108,7656	0,1652
181	80,3742	77,1050	83,6434	67,6626	93,0859	0,0708
152	51,1196	47,1550	55,0841	38,2116	64,0275	0,1042
172	71,2952	68,5280	74,0624	58,7033	83,8871	0,0507
152	51,1196	47,1550	55,0841	38,2116	64,0275	0,1042
150	49,1020	44,9013	53,3027	36,1195	62,0845	0,1169
178	77,3479	74,3117	80,3841	64,6942	90,0016	0,0611
154	53,1371	49,3963	56,8779	40,2961	65,9781	0,0927
190	89,4532	85,2283	93,6782	76,4629	102,4436	0,1183
195	94,4972	89,6376	99,3567	81,2868	107,7075	0,1565
182	81,3830	78,0245	84,7415	68,6481	94,1179	0,0747
184	83,4006	79,8486	86,9525	70,6133	96,1878	0,0836
156	55,1547	51,6229	58,6864	42,3730	67,9364	0,0827
154	53,1371	49,3963	56,8779	40,2961	65,9781	0,0927

