VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta informačních technologií

Projekt z MSP

Zpracovala: Kateřina Fořtová (xforto00)

Čísla zadání: příklad 1: 27, **příklad 2:** 4

Cvičení: středa od 8.00

Datum: 23. 11. 2020

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

Netříděný statistický soubor:

27

Př. 1

X [mm]
2,13
-0,66
1,06
-0,15
0,80
1,87
-0,08
1,29
1,07
1,21
1,91
1,10
0,64
3,07
0,26
0,16
1,00
1,39
0,81
-0,55
0,67
0,60
1,67
2,14
1,56
-0,53
2,48
2,25
0,21
0,58
0,35
0,26
0,43
1,38
2,29
0,49
2,31
2,82
0,62
0,94
1,40
0,71
3,44
1,61
1,94
2,64
1,78
1,14
1,51
1,05

Tříděný statistický soubor:

27

Př. 1

X [mm]
-0,66
-0,55
-0,53
-0,15
-0,08
0,16
0,21
0,26
0,26
0,35
0,43
0,49
0,58
0,60
0,62
0,64
0,67
0,71
0,80
0,81
0,94
1,00
1,05
1,06
1,07
1,10
1,14
1,21
1,29
1,38
1,39
1,40
1,51
1,56
1,61
1,67
1,78
1,87
1,91
1,94
2,13
2,14
2,25
2,29
2,31
2,48
2,64
2,82
3,07
3,44

Proveďte roztřídění statistického souboru, vytvořte tabulku četností a nakreslete histogramy pro relativní četnosti a relativní kumulativní černosti.

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

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

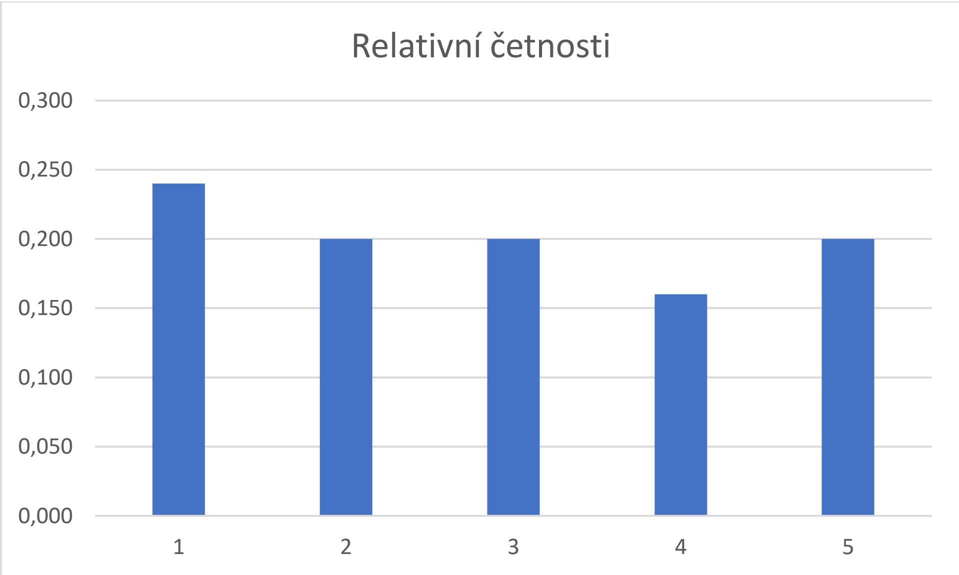
Variační obor: $\langle x_{(1)}, x_{(n)} \rangle = \langle -0{,}66; 3{,}44 \rangle$

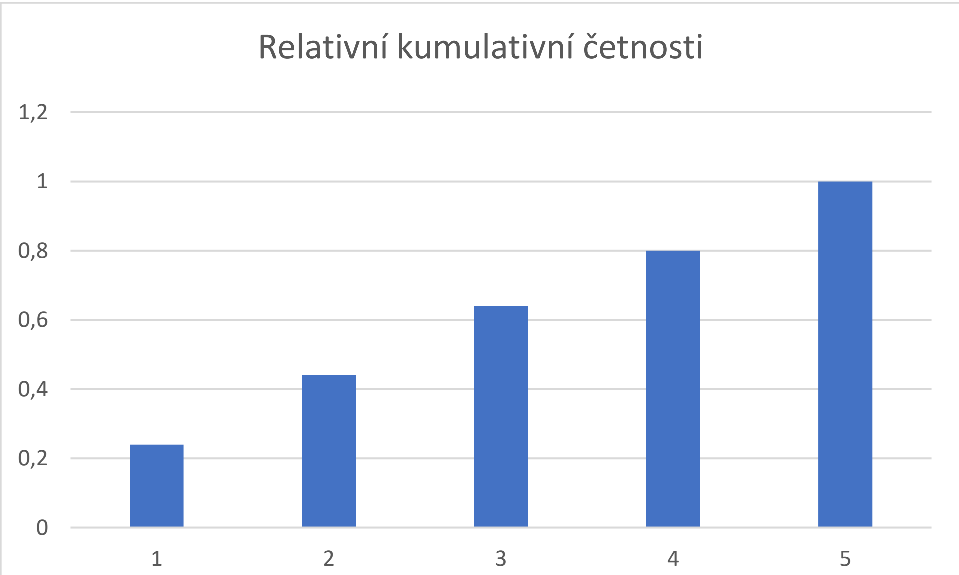
Rozpětí: $x_{(n)} - x_{(1)} = 4{,}1$

Počet tříd (zvoleno): $m = 5$

Délka třídy: $\frac{x_{(n)}-x_{(1)}}{m} = 0{,}82$

Třída	Interval hodnot třídy	Četnost (fi)	Interval hodnot třídy	Kumulativní četnosti	Relativní četnost	Relativní kumulativní četnost
1	(-∞, 0,5>	12	(-∞, 0,5>	12	0,24	0,24
2	(0,5, 1>	10	(0,5, 1>	22	0,2	0,44
3	(1, 1,5>	10	(1, 1,5>	32	0,2	0,64
4	(1,5, 2>	8	(1,5, 2>	40	0,16	0,8
5	(2, +∞)	10	(2, +∞)	50	0,2	1
Suma		50			1	





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

Aritmetický průměr: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 1{,}181$

Medián: $\tilde{x} = 1{,}085$

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

Rozptyl: $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 0{,}874$

Směrodatná odchylka: $s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} = 0{,}935$

Vypočtěte bodové odhady střední hodnoty, rozptylu a směrodatné odchylky.

Bodový odhad střední hodnoty: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 1{,}181$

Bodový odhad rozptylu: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 0{,}892$

Bodový odhad směrodatné odchylky: $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = 0{,}944$

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

Třída	**Interval hodnot třídy**	**Četnost (fi)**	**Kumulativní četnost**	**Relativní četnost**	**Relativní kumulativní četnost**	**Teoretická pravděpodobnost**	**Teoretická četnost (^fi)**	**(fi - ^fi)^2 / ^fi**
1	(-∞, 0,5>	12	12	0,24	0,24	0,233	11,653	0,010
2	(0,5, 1>	10	22	0,2	0,44	0,190	9,501	0,026
3	(1, 1,5>	10	32	0,2	0,64	0,210	10,514	0,025
4	(1,5, 2>	8	40	0,16	0,8	0,176	8,800	0,073
5	(2, +∞)	10	50	0,2	1	0,191	9,532	0,023
Suma		50		1		1	50	0,157

H – jedná se o výběr z normálního rozdělení

Testovací kritérium: $t = \sum_{i=1}^{n}\frac{(f_i - \hat{f_i})^2}{\hat{f_i}} = 0{,}157$

Počet stupňů volnosti: $k = 5 - 2 - 1 = 2$

$\chi^2_{1-\alpha}$pro 2 stupně volnosti: 5,991

Doplněk kritického oboru: $\overline{W_\alpha} = \langle 0, \chi^2_{1-\alpha}\rangle = \langle 0; 5{,}991\rangle$

Protože $t \in \overline{W_\alpha}$, tedy hypotéza $X \sim N(1{,}181; 0{,}892)$ **se nezamítá.**

Za předpokladu (bez ohledu na výsledek předchozí části), 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.

Intervalový odhad střední hodnoty:

$$\langle \bar{x} - t_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}; \bar{x} + t_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\rangle$$

Intervalový odhad rozptylu:

$$\langle \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}; \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}\rangle$$

Intervalový odhad směrodatné odchylky:

$$\left\langle \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}};\ \sqrt{\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}} \right\rangle$$

Intervalové odhady pro spolehlivost 0,95 ($\boldsymbol{\alpha = 0,05}$)

Počet stupňů volnosti: $k = n - 1 = 50 - 1 = 49$

Intervalový odhad střední hodnoty:

0,975 kvantil Studentova rozdělení $t_{1-\frac{\alpha}{2}}$ s 49 stupni volnosti: 2,01

Výsledný interval: $\langle 0{,}913; 1{,}45 \rangle$

Intervalový odhad rozptylu:

0,975 kvantil Pearsova rozdělení $\chi^2_{\frac{\alpha}{2}}$ s 49 stupni volnosti: 31,555

0,975 kvantil Pearsova rozdělení $\chi^2_{1-\frac{\alpha}{2}}$ s 49 stupni volnosti: 70,222

Výsledný interval: $\langle 0{,}622; 1{,}385 \rangle$

Intervalový odhad směrodatné odchylky:

Výsledný interval: $\langle 0{,}789; 1{,}177 \rangle$

Intervalové odhady pro spolehlivost 0,99 ($\boldsymbol{\alpha = 0,01}$)

Počet stupňů volnosti: $k = n - 1 = 50 - 1 = 49$

Intervalový odhad střední hodnoty:

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

Výsledný interval: $\langle 0{,}823; 1{,}539 \rangle$

Intervalový odhad rozptylu:

0,995 kvantil Pearsova rozdělení $\chi^2_{\frac{\alpha}{2}}$ s 49 stupni volnosti: 27,249

0,995 kvantil Pearsova rozdělení $\chi^2_{1-\frac{\alpha}{2}}$ s 49 stupni volnosti: 78,231

Výsledný interval: $\langle 0{,}559; 1{,}604 \rangle$

Intervalový odhad směrodatné odchylky:

Výsledný interval: $\langle 0{,}747; 1{,}266 \rangle$

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: $t = \frac{\bar{x}-\mu_0}{s}\sqrt{n} = \frac{\bar{x}-0}{s}\sqrt{n} = 8{,}845$

Doplněk kritického oboru: $\overline{W_\alpha} = \langle -t_{1-\frac{\alpha}{2}}, t_{1-\frac{\alpha}{2}} \rangle$ pro alternativní hypotézu: $H_A: \mu \neq \mu_0$:

Počet stupňů volnosti: $k = n - 1 = 50 - 1 = 49$

0,975 kvantil Studentova rozdělení $t_{1-\frac{\alpha}{2}}$ s 49 stupni volnosti: 2,01

Doplněk kritického oboru: $\overline{W_\alpha} = \langle -t_{1-\frac{\alpha}{2}}, t_{1-\frac{\alpha}{2}} \rangle = \langle -2{,}01; 2{,}01 \rangle$

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

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.

Naměření prvních 20 hodnot (X)	Naměření zbývajících 30 hodnot (Y)
2,13	0,67
-0,66	0,60
1,06	1,67
-0,15	2,14
0,80	1,56
1,87	-0,53
-0,08	2,48
1,29	2,25
1,07	0,21
1,21	0,58
1,91	0,35
1,10	0,26
0,64	0,43
3,07	1,38
0,26	2,29
0,16	0,49
1,00	2,31
1,39	2,82
0,81	0,62
-0,55	0,94
	1,40
	0,71
	3,44
	1,61
	1,94
	2,64
	1,78
	1,14
	1,51
	1,05

X (měření před seřízením stroje):

$$n = 20$$

$$průměr = 0{,}917$$

$$rozptyl\ (s^2) = 0{,}821$$

$$směrodatná\ odchylka\ (s) =\ 0{,}906$$

Y (měření po seřízení stroje):

$$n = 30$$

$$průměr = 1{,}358$$

$$rozptyl\ s^2 = 0{,}832$$

$$směrodatná\ odchylka = 0{,}912$$

Test rovnosti rozptylů – F-test:

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

Testovací kritérium: $t = \frac{s^2(X)}{s^2(Y)} = 0{,}987$

Doplněk kritického oboru: $\overline{W_\alpha} = \langle F_{\frac{\alpha}{2}}(n-1, m-1), F_{1-\frac{\alpha}{2}}(n-1, m-1)\rangle$ 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

$$F_{\frac{\alpha}{2}}(19{,}29) = 0{,}416$$

$$F_{1-\frac{\alpha}{2}}(19{,}29) = 2{,}231$$

$$\overline{W_\alpha} = \langle F_{\frac{\alpha}{2}}(n-1, m-1), F_{1-\frac{\alpha}{2}}(n-1, m-1)\rangle = \langle 0{,}416; 2{,}231\rangle$$

Protože $t \in \overline{W_\alpha}$, 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 $\sigma_X^2 = \sigma_Y^2$

Testovací kritérium: $t = \frac{\bar{x}-\bar{y}-\mu_0}{\sqrt{(n-1)s^2(X)+(m-1)s^2(Y)}}\sqrt{\frac{n*m(n+m-2)}{n+m}} = -1{,}681$

Doplněk kritického oboru: $\overline{W_\alpha} = \langle -t_{1-\frac{\alpha}{2}}, t_{1-\frac{\alpha}{2}}\rangle$ 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.

$$t_{1-\frac{\alpha}{2}} = 2{,}011$$

$$\overline{W_\alpha} = \langle -t_{1-\frac{\alpha}{2}}, t_{1-\frac{\alpha}{2}}\rangle = \langle -2{,}011; 2{,}011\rangle$$

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

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.

4	
Př. 2 **Výška [cm]**	**Váha [kg]**

167	61	191	101
198	125	179	106
176	97	166	76
183	97	170	82
182	106	195	124
156	72	194	106
165	79	158	95
181	105	198	125
182	98	192	105
196	106	182	82

$$n = 20$$

$$\bar{x} = 180{,}55$$

$$\bar{y} = 97{,}4$$

$$\sum_{i=1}^{n} x_i^2 = 655339$$

$$\sum_{i=1}^{n} y_i^2 = 195706$$

$$\sum_{i=1}^{n} x_i y_i = 355372{,}2115$$

Vypočtěte bodový odhad koeficientu korelace.

$$r = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sqrt{(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2)(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2)}} = 0{,}816$$

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

Testujeme hypotézu $H_0 : \rho = 0$:

Testovací kritérium: $t = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} = 5{,}989$

Doplněk kritického oboru: $\overline{W_\alpha} = \langle 0, t_{1-\frac{\alpha}{2}} \rangle$ pro alternativní hypotézu $H_A : \rho \neq 0$

$$t_{1-\frac{\alpha}{2}}(n-2) = t_{0{,}975}(20-2) = 2{,}101$$

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

Regresní analýza – data proložte přímkou: $\boldsymbol{V\acute{a}ha = \beta_0 + \beta_1 * V\acute{y}\check{s}ka}$

	Výška [cm]	Váha [kg]	Výška^2	Výška * váha	Váha^2
	167	61	27889	10144,15806	3690
	198	125	39204	24822,06972	15716
	176	97	30976	17033,12934	9366
	183	97	33489	17777,50638	9437
	182	106	33124	19238,29114	11174
	156	72	24336	11241,75711	5193
	165	79	27225	12982,20198	6191
	181	105	32761	19069,94267	11100
	182	98	33124	17881,82074	9653
	196	106	38416	20860,5999	11328
	191	101	36481	19230,76805	10137
	179	106	32041	18909,87122	11160
	166	76	27556	12588,81142	5751
	170	82	28900	13899,44784	6685
	195	124	38025	24158,93094	15349
	194	106	37636	20616,25178	11293
	158	95	24964	15027,29909	9046
	198	125	39204	24741,39454	15614
	192	105	36864	20239,28418	11112
	182	82	33124	14908,67537	6710
Sumy:	3611	1948	655339	355372,2115	195706

Matice H: $H = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n {x_i}^2 \end{pmatrix} = \begin{pmatrix} 20 & 3611 \\ 3611 & 655339 \end{pmatrix}$

Determinant matice H: $\det(H) = 67459$

Bodově odhadněte $\boldsymbol{\beta_0}$, $\boldsymbol{\beta_1}$ a rozptyl $\boldsymbol{s^2}$.

$$\beta_1 = b_2 = \frac{1}{\det(H)}\left(n\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i\right) = 1{,}085$$

$$\beta_0 = b_1 = \bar{y} - b_2\bar{x} = -98{,}547$$

Rovnice pro podmíněnou střední hodnotu: $V\acute{a}ha = -98{,}547 + 1{,}085 * V\acute{y}\check{s}ka$

$$S_{min}^* = \sum_{i=1}^n y_i^2 - b_1\sum_{i=1}^n y_i - b_2\sum_{i=1}^n x_i y_i = 1997{,}743$$

$$s^2 = \frac{S_{min}^*}{n-2} = \frac{S_{min}^*}{20-2} = 110{,}986$$

Na hladině významnosti 0,05 otestujte hypotézy.

$$h^{11} = \frac{\sum_{i=1}^n x_i^2}{\det(H)} = 9{,}715$$

$$h^{22} = \frac{n}{\det(H)} = 0{,}000296$$

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

Testovací kritérium: $t = \frac{b_1 - (-100)}{s\sqrt{h^{11}}} = 0{,}044$

$$t_{1-\frac{\alpha}{2}}(n-2) = t_{0,975}(20-2) = 2{,}101$$

$t \in \overline{W} = \langle -2{,}101; 2{,}101 \rangle$, a tedy $H: \beta_0 = -100$ **se nezamítá.**

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

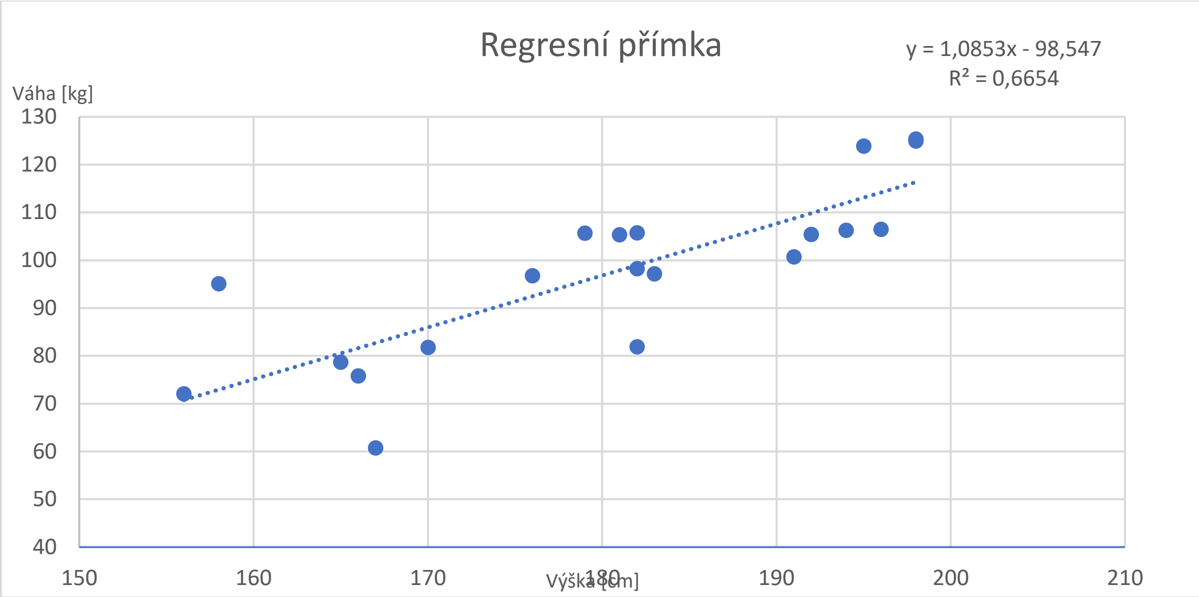
Testovací kritérium: $t = \frac{b_2 - 1}{s\sqrt{h^{22}}} = 0{,}047$

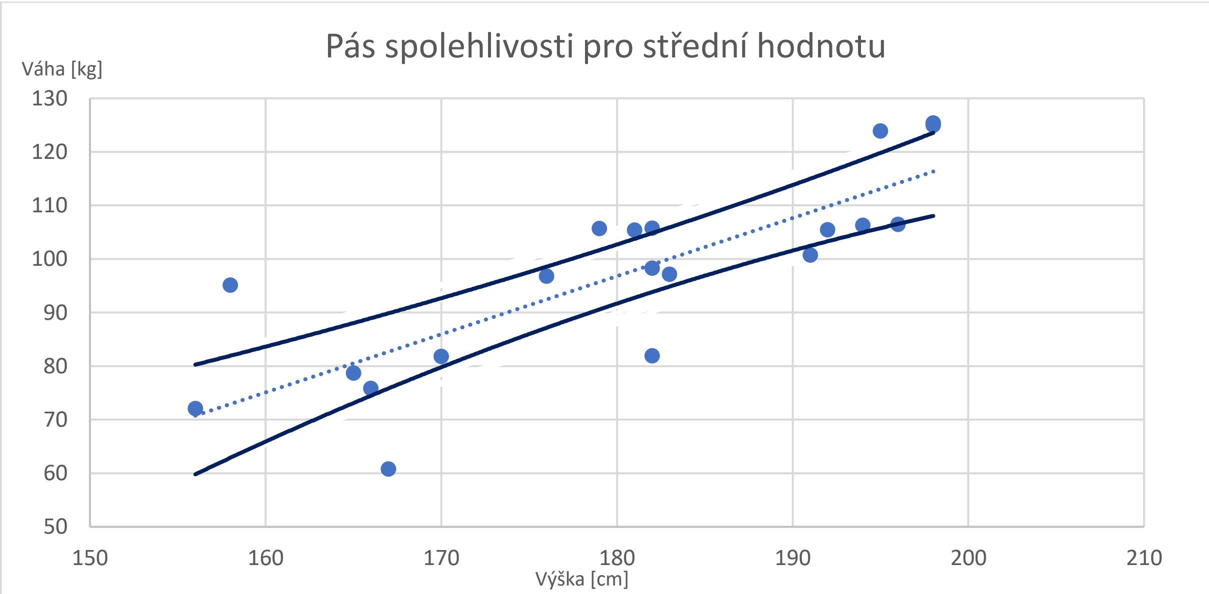
$$t_{1-\frac{\alpha}{2}}(n-2) = t_{0,975}(20-2) = 2{,}101$$

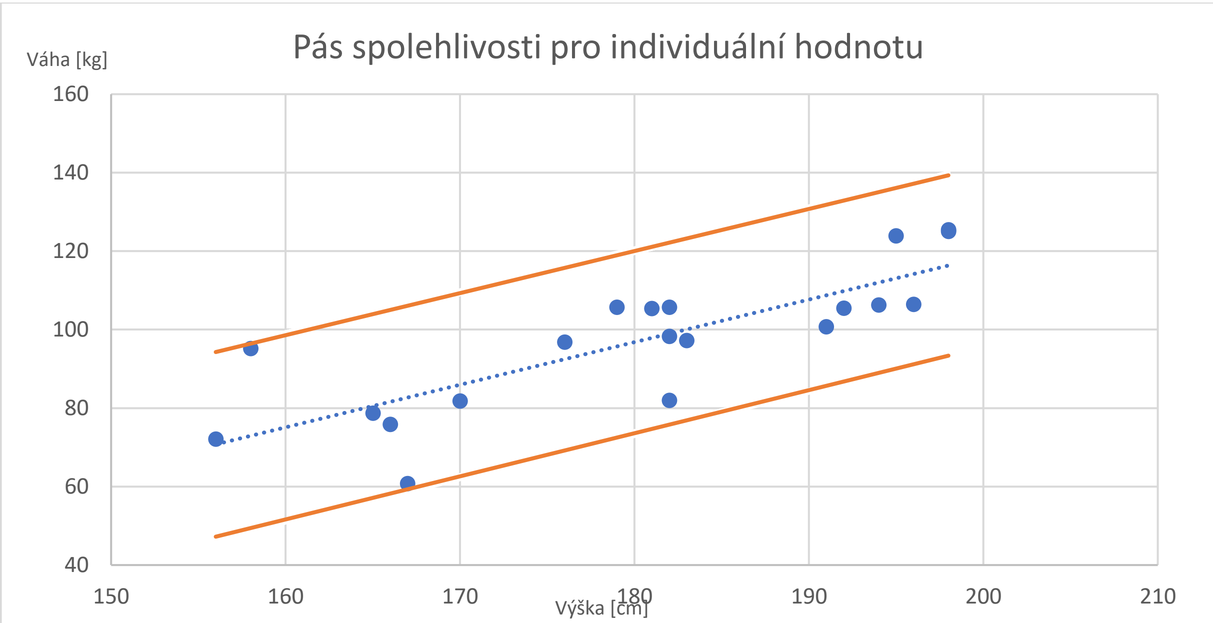
$t \in \overline{W} = \langle -2{,}101; 2{,}101 \rangle$, a tedy $H: \beta_1 = 1$ **se nezamítá.**

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

xi	yi (reálně naměřené)	yi (spočtené pomocí rovnice)	střední y dolní	střední y horní	individuální y dolní	individuální y horní	h*
167	61	82,698	75,545	89,851	59,438	105,958	0,104
198	125	116,342	108,053	124,632	92,708	139,977	0,140
176	97	92,466	87,222	97,710	69,720	115,212	0,056
183	97	100,063	95,026	105,099	77,364	122,762	0,052
182	106	98,978	93,998	103,957	76,291	121,664	0,051
156	72	70,760	60,175	81,344	46,226	95,294	0,229
165	79	80,528	72,807	88,248	57,086	103,969	0,122
181	105	97,892	92,940	102,844	75,212	120,573	0,050
182	98	98,978	93,998	103,957	76,291	121,664	0,051
196	106	114,172	106,480	121,864	90,740	137,603	0,121
191	101	108,745	102,393	115,098	85,719	131,772	0,082
179	106	95,722	90,737	100,706	73,034	118,409	0,051
166	76	81,613	74,180	89,045	58,265	104,961	0,113
170	82	85,954	79,578	92,330	62,921	108,987	0,083
195	124	113,087	105,682	120,491	89,748	136,425	0,112
194	106	112,001	104,876	119,126	88,749	135,253	0,104
158	95	72,930	63,013	82,847	48,677	97,184	0,201
198	125	116,342	108,053	124,632	92,708	139,977	0,140
192	105	109,831	103,233	116,429	86,735	132,926	0,089
182	82	98,978	93,998	103,957	76,291	121,664	0,051


Regresní přímka
y = 1,0853x - 98,547
R² = 0,6654
Váha [kg]
130
120
110
100
90
80
70
60
50
40
150
160
170
180
190
200
210
Výška [cm]


Pás spolehlivosti pro střední hodnotu
Váha [kg]
130
120
110
100
90
80
70
60
50
150
160
170
180
190
200
210
Výška [cm]


Pás spolehlivosti pro individuální hodnotu
Váha [kg]
160
140
120
100
80
60
40
150
160
170
180
190
200
210
Výška [cm]