4. ZÁVISLOST MEZI SPOJITÝMI VELIČINAMI

Naším cílem v dnešní lekci je vyšetřit faktory související s hmotností mladého člověka. Za faktory byly vybrány:

- Věk
- Výška
- Pohlaví
- Obvod pasu
- Obvod hrudníku
- Obvod boků
- Obvod krku
- Velikost bot
- Počet hodin, po který se respondent průměrně věnuje týdně sportu
- Počet hodin, který respondent průměrně denně stráví u PC a TV

Průzkum byl prováděn mezi studenty vybrané střední školy. Soubor Data-celek.xls obsahuje údaje o 110 respondentech.

Obsah kapitoly

4.1	Krát	ký teoretický úvod	2
4.1	1	Kvalita regresních modelů	4
4.2	Reg	resní analýza pomocí SW Excel	6
4.2	.1	Grafické znázornění	6
4.2	.2	Jednoduchá regresní analýza pomocí doplňku prostředku Excel – lineární model	. 10
4.2	.3	Jednoduchá regresní analýza pomocí doplňku prostředku Excel – obecný model	. 12
4.2	.4	Vícenásobná regresní analýza pomocí doplňku prostředku Excel	. 16
4.3	Reg	resní analýza v SW STATISTICA	. 24
4.3	.1	Grafické znázornění	. 24
4.3	.2	Jednoduchá regresní analýza – lineární model	. 27
4.3	.3	Jednoduchá regresní analýza – obecný model	. 32
4.3	.4	Vícenásobná regresní analýza pomocí SW STATISTICA	. 38

4.1 Krátký teoretický úvod

Závislost spojitých veličin se vyšetřuje pomocí dvojice metod, a to regrese a korelace. Úkolem regrese je najít vhodný funkční model této závislosti. Úkolem korelace je změřit sílu lineární závislosti. Jinými slovy, regrese popisuje daný vztah a korelace zjišťuje jeho těsnost. Známe dva základní typy regresní analýzy, a to jednoduchou a vícenásobnou.

Cílem jednoduché (simple) regrese je najít model funkční závislosti (spojité) veličiny Y na jedné (spojité) veličině (na tzv. regresoru) X. Tvar funkce často napoví bodový graf dat. Příkladem může být zkoumání závislosti mezi platem a výší úspor či mezi výší exportu a výší HDP.

Cílem vícenásobné (multiple) regrese je najít model funkční závislosti (spojité) veličiny Y na více (spojitých) veličinách (regresorech). Příkladem může být zkoumání závislosti výše úspor na platu, výdajích za potraviny, výdajích za spotřební zboží a výdajích za kulturu.

Při jednoduché regresi můžeme hledat modely různých typů. Mezi nejvíce používané patří:

Lineární model: $y = b_1 x + b_0$ (1)

Polynomický model:
$$y = \sum_{i=0}^{n} b_i x^i$$
 (2)

Mocninný model: $y = b_0 x^{b_1}$ (3)

Logaritmický model: $y = b_1 \ln x + b_0$

Hledáním regresního modelu, resp. regresní funkce rozumíme hledání regresních koeficientů b_i, přičemž typ regresního modelu musíme stanovit sami (na základě zkušeností a vzhledu bodových grafů), hodnoty jednotlivých regresních koeficientů pak nalezne metoda (např. implementovaná v SW).

(4)

Blíže se nyní seznámíme s nejjednodušším a nejčastěji využívaném typu, a to lineárním modelem, v němž hledáme funkci, jejímž grafem je přímka, viz obr.



Snažili jsme se "proložit" tři body A, B a C regresní přímkou. Hledáme funkci (přímku) ve tvaru $\hat{y} = b_1 x + b_0$. Vidíme, že platí $y = b_1 x + b_0 + \varepsilon$, neboli že naměřené hodnoty se "o něco" liší od vypočítaných hodnot odhadu. Tomuto rozdílu říkáme reziduum, značíme ε . V regresním modelu by měla mít rezidua normální rozdělení se střední hodnotou 0.

Hodnoty b_1 a b_0 odhadujeme pomocí Metody nejmenších čtverců, která je založena na principu hledání minima funkce více proměnných pomocí parciálních derivací. Z této metody je možno získat následující vzorce.

$$b_1 = \frac{x \cdot y - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2}$$
(5)

$$b_0 = \bar{y} - b_1 \cdot \bar{x} \tag{6}$$

Směrnice přímky b_1 odpovídá změně závislé proměnné při nárůstu nezávislé proměnné o jednu jednotku.

Výpočty regresních koeficientů pomocí uvedených vzorců jsou poněkud pracné. Proto se většinou v praxi využívá různých pomocníků. V případě lineárních modelů můžeme výpočty provést na kalkulačkách, a to pomocí speciálních funkcí. Ještě efektivnější je využití různých SW, například i velmi rozšířeného Microsoft Excelu, nebo mnoha komerčních statistických SW, jako je STATISTICA či SPSS.

Pro obecné modely jednoduché regrese můžeme využít maticového vzorce, který rovněž vychází z metody nejmenších čtverců.

$$\vec{b} = (F^T F)^{-1} F^T \vec{y} \,. \tag{7}$$

kde $\vec{b} = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_m \end{pmatrix}$ je vektor regresních koeficientů, $\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ yn \end{pmatrix}$ je vektor hodnot veličiny y, $F = \begin{pmatrix} f_1(x_1), \cdots, f_m(x_1) \\ f_1(x_2), \cdots, f_m(x_2) \\ \vdots \\ f_1(x_n), \cdots, f_m(x_n) \end{pmatrix}$ je regresní matice příslušná danému regresnímu modelu.

Například pro lineární jednoduchou regresi má matice F následující tvar: $F = \begin{pmatrix} 1, x_1 \\ 1, x_2 \\ \vdots \\ 1, x_n \end{pmatrix}$, pro

kvadratický model pak: $F = \begin{pmatrix} 1, x_1, x_1^2 \\ 1, x_2, x_2^2 \\ \vdots \\ 1, x_n, x_n^2 \end{pmatrix}$, nebo pro logaritmický model: $F = \begin{pmatrix} 1, \ln(x_1) \\ 1, \ln(x_2) \\ \vdots \\ 1, \ln(x_n) \end{pmatrix}$.

Při vícenásobné regresi převážně využíváme lineárních vztahů. Výpočet regresních koeficientů se provádí opět s využitím vzorce (7). V tomto případě má matice F tvar:

$$F = \begin{pmatrix} 1, x_{11}, x_{12}, \cdots, x_{1m} \\ 1, x_{21}, x_{22}, \cdots, x_{2m} \\ \vdots \\ 1, x_{n1}, x_{n2}, \cdots, x_{nm} \end{pmatrix}, \text{ kde } x_{ij} \text{ znamená } i\text{-tou hodnotu } j\text{-tého regresoru.}$$

4.1.1 Kvalita regresních modelů

Jak již bylo výše uvedeno, reziduum ε značí odchylku naměřené hodnoty od hodnoty vypočítané, čili $\varepsilon_i = y_i - \hat{y}_i$.

Ve výpočtech pak z důvodu odstranění znaménka (+, –) pracujeme s druhými mocninami těchto reziduí, neboli s reziduálními čtverci ε_i^2 . Metoda nejmenších čtverců hledá minimum tzv. součtu reziduálních čtverců Q_e .

$$Q_e = \sum_{i=1}^n \varepsilon_i^2 \tag{8}$$

Kvalitu regresního modelu vyhodnocujeme pomocí následujících charakteristik.

Reziduální rozptyl

$$s_e^2 = \frac{Q_e}{n-p} \quad , \tag{9}$$

kde *n* je počet měření (bodů) a *p* je počet parametrů modelu (pro lineární model p = 2). Platí, že $s_e^2 \ge 0$ a dále, že čím větší je hodnota s_e^2 , tím hůře model vystihuje data.

Index determinace

$$I^{2} = \frac{Q_{\hat{y}}}{Q_{y}} = 1 - \frac{Q_{e}}{Q_{y}},$$
(10)

kde $Q_{\hat{y}} = \sum_{i=1}^{n} (f(x_i) - \overline{y})^2$ a $Q_y = \sum_{i=1}^{n} (y_i - \overline{y})^2$.

Někdy se můžeme setkat s názvem Koeficient determinace a také s označením R^2 .

Platí, že $I^2 \in \langle 0; 1 \rangle$. Hodnotu indexu determinace pro interpretaci převádíme na procenta. Jeho hodnota nám pak říká, kolik procent rozptylu vysvětlované proměnné je vysvětleno modelem. Srozumitelněji a zjednodušeně jinými slovy můžeme říci, z kolika procent model vystihuje daná data. Je tedy zřejmé, že čím vyšší index determinace, tím lepší model.

Nutno ovšem podotknout, že index determinace závisí na počtu vysvětlujících proměnných a s růstem jejich počtu narůstá i jeho hodnota. V důsledku toho index determinace zvýhodňuje složitější modely (tj. modely s více parametry). Toto je nepříjemná vlastnost, která částečně snižuje jeho kvalitu. Pokud tedy využíváme indexu determinace k porovnání dvou modelů s různým počtem parametrů, měli bychom jeho vyhodnocení doplnit i vyhodnocením například pomocí reziduálního rozptylu.

Upravený index determinace

Porovnání modelů s různým počtem parametrů můžeme také provést pomocí upraveného indexu determinace.

$$I_{upr}^2 = 1 - (1 - I^2) \frac{n - 1}{n - p}$$
(11)

Pearsonův korelační koeficient

$$|r| = \sqrt{I^2} \operatorname{asgn}(r) = \operatorname{sgn}(b_1) \tag{12}$$

Korelační koeficient má smysl počítat pouze pro lineární model. Korelační koeficient má stejné znaménko jako směrnice regresní přímky.

Tento výpočet korelačního koeficientu (12) je velmi zdlouhavý, proto se více využívá následujícího upraveného vzorce.

$$r = \frac{x \cdot y - \overline{x} \cdot \overline{y}}{\sqrt{(\overline{x^2} - \overline{x}^2) \cdot (\overline{y^2} - \overline{y}^2)}}$$
(13)

Platí, že $r \in \langle -1; 1 \rangle$. Čím blíže je jeho hodnota ke krajním hodnotám tohoto rozmezí, tím je lepší model. Pro vyhodnocení hodnot korelačního koeficientu existuje speciální test hypotéz. Zjednodušeně lze však říci, že pokud je jeho hodnota blízká 1 (pro dostatečné množství dat se většinou uvádí podmínka větší než 0,8), pak mluvíme o silné přímé lineární závislosti. Je-li jeho hodnota blízko -1 (menší než -0,8), pak mluvíme o silné nepřímé lineární závislosti. Pokud je jeho hodnota blízko 0 (v rozmezí od -0,3 do +0,3), pak říkáme, že není lineární závislost. Slovo lineární v poslední větě je velmi důležité. Je nutno si uvědomit, že neexistence lineární závislosti nevylučuje existenci funkční závislosti jiného druhu (kvadratické, logaritmické, ...)

Příklady vybraných situací v datech



Spearmanův koeficient pořadové korelace

Pokud data nesplňují předpoklady rozložení dat (jiné než normální rozložení proměnných, nelinearita vztahu, data obsahující odlehlá pozorování, ordinální data) je vhodnější použít neparametrický ekvivalent, a to Spearmanův koeficient pořadové korelace. Jsou-li hodnoty proměnných x_i a y_i seřazeny vzestupně do dvou řad a každé hodnotě je přiděleno pořadí, pak koeficient pořadové korelace je dán vztahem:

$$r_{s} = 1 - \frac{6\sum_{i=1}^{n} D_{i}^{2}}{n(n^{2} - 1)},$$
(14)

kde D_i je rozdíl pořadí hodnot x_i a y_i .

Pokud se vyskytuje v souboru více stejných hodnot, pak všechny z nich obdrží hodnotu pořadí vypočítanou jako průměr z hodnot jednotlivých pozic (př. pokud se v souboru vyskytují na pozicích 5-8 čtyři stejné hodnoty, pak všechny tyto hodnoty obdrží pořadí $\frac{5+6+7+8}{4} = 6,5$.)

Spearmanův koeficient pořadové korelace nabývá stejně jako korelační koeficient hodnot z intervalu $r_s \in \langle -1; 1 \rangle$. Vyhodnocení síly závislosti pak probíhá obdobně jako u korelačního koeficientu.

Kromě Spearmanova korelačního koeficientu existují i další neparametrické korelační koeficienty jako např. Kendallovo τ .

4.2 Regresní analýza pomocí SW Excel

4.2.1 Grafické znázornění

Pokud chceme provést regresní analýzu, vytvoříme si nejprve graf. V Excelu je pro tyto účely nejvhodnější bodový graf. V nabídce zvolíme *Vložit – Bodový graf (pouze se značkami)*.

9	Dom	- Cr - L	ni Rozk	ožení stránk	y Vzorce	Data	Revize	student Zobrazeni d	ske dota Acrobat	zniky (Režim kompatibility	/] - Microsoft Excel
Ko	ntingenční 1 tabulka *	abulka 0	Dbrázek Kilp	art Tvary	SmartArt SI	oupcový Spo	inicový Výse	cový Pruhový Plo	iný Bod	ovy Dalši grafy -	A Textové Záhlaví W pole a zápatí
	Tabulky		1	ustrace				Grafy	Bo	odový	
	08	6	• ()	fre					0	* 19-P	
12	A	В	C	D	E	F	G	H	1 .	· / / /	L M
1	věk	výška	boty	prsa	pas	boky	krk	sport spár	nek -		
2	18	19	3 4	5 10	8 97	96	38	6	1.	Bodový pouze se značka	mi
3	20	17	5 4	3 9	6 80	91	37	1		Umožňuje porovnávat o	tvolice
4	20	17	8 4	3 11	0 85	90	46	7		hodnot.	
5	20	18	1 4:	2 9	5 87	106	37	6	10.		change -
6	20	18	0 4	10	7 92	97	37	6		Tuto moznost pouzijte,	pokud
7	20	19	1 4	3 9	8 93	99	34	21	11	sefazeny podle pořadi i	nebo
8	19	18	5 4	1 11	2 94	118	40	2	0.	pokud představují sam	ostatné
9	20	19	0 4	5 11	0 95	120	41	2	2003	hodnoty.	
10	22	18	1 4	3 9	6 90	100	41	1	8	13 85	
11	22	18	3 4	3 9	8 93	100	40	1	8	12 86	
12	21	18	0 4	1 11	1 98	120	42	0	7	12 102	
13	20	18	2 4	5 11	0 102	121	43	0	7	9 110	
			-								

Po této volbě nám Excel vytvoří prázdnou plochu pro graf. V menu pro tuto oblast vybereme možnost *Vybrat data*.

6		- 00- 🗋		stude	ntske dotazr	iky (Režim	kompatibili	ty] - Micros	soft Excel		Nástroje grafu	Sec.		
1	Dom	ů Vlože	ni Rozlo	žení stránky	Vzorce	Data	Revize	Zobrazen	i Acrol	bat	Návrh Rozložení Formát			
z	měnit Ulož o grafu šat	it jako Pie Ionu d	pnaut řádek 5 sloupes	Vybrat data				•	•	• •	• • • • • •		•	•
	Тур		Data		Rozi	ožení grafu	N				Styty	r grafů		
	Graf		• (3)	Sec						X	∑yjmout			
120	A	В	C	D	E	F	G	Н	1	20	Kopirovat	N	0	
1	věk	výška	boty	prsa	pas	boky	krk	sport	spánek		Vložit			
2	2	193	40	108	97	96	30	,	1	6 .7	Obnovit a srovnat se stylem			
4	2	178	43	110	85	90	46	5	7	8	Dime			
5	21	181	42	95	87	106	37	1	6	9	PISMO			
6	21	180	44	107	92	97	37		6	9	Zménit typ grafu			
7	21	191	43	98	93	99	36	2	1	7 3	Vybrat data		-4	
8	1	185	44	112	94	118					Přesunout graf			
9	2	190	40	110	95	120	2			0	Prostorové otočení			
11	2	183	43	90	93	100				125	Shimina b			
12	2	180	44	111	98	120					all the second sec			
13	21	182	45	110	102	121				-10	Prenest do popredi			
14	2	187	42	107	96	113				-	Přenést do poza <u>d</u> i 🔹 🕨			
15	2	180	44	110	96	120					Přiřagit makro			
16	2	192	46	120	100	121	1				Formát oblacti orafu			
17	2	2 190	45	123	112	125	§				Louist courts diarow			
18	2	2 176	42	100	98	120	1			-				
19	2	2 175	43	99	94	100				Calit	bri (2 - 10 - A A 🔐 - 🖉 - 🥑			
20	2	186	44	100	99	110				B	I # # # # # # # # #			
21	2	199	46	108	102	120								
22	2	198	45	110	100	118	3							
23	2	185	44	105	89	102	2							
24	24	189	45	107	95	110	5-						-4	
25	21	193	45	110	100	115	43	3	3	9	9 95			

V následujícím dialogovém okně vybereme možnost Přidat Položky legendy (řady).



Následně vyplníme požadovaný název řady a tažením myší označíme sloupce s daty.

0	16	3 17 -	61 -	0.0		÷		student	tske dota	zniky [Re	žim kompa	tibility] - M	licrosoft Exc	el	Nástroje	grafu	-					
	2	Domů	Vie	zení	8	loziože	ni st	ránky	Vzore	e Dat	ta Revis	te Zobi	azeni i	Acrobat	Návrh	Rozie	žení For	mát				
VB	al l	Х Ц Ј В	1	<u>U</u>	• 10		A'	x") (1 2-) (1		≫ ••)	amovat text učit a zarov	nat na střed	- Egg	cný - % 000	- ***	Podminěn formátován	e Formát (* jako tab	tovat Styly	Viož	e Odita	amit Form
Sch	ánka	r.		Pin	mo						Zarovnáni				Ċislo			Styly			Buñi	cy:
		N2		(*)	3	19	fx															
	в	С	D	E	F	G	н	1	J	K	L	M	N	0	Р	Q	R	S	ा	U		V
2	18	193	46	108	97	96	38	6	8	4	24,16172	normální	90	1			14 C 12					
3	17	160	38	86	73	86	31	3	7	3	20,3125	normální	52			výs	ska					
4	16	171	41	85	68	93	33	5	7	3,5	21,20311	normální	62	140								
5	17	160	39	82	63	86	33	10	7	4	19,33594	podváha	49,5									
6	17	169	40	79	64	90	31	3,5	6	1,5	17,50639	podváha	50	120						-		
7	17	163	39	92	68	100	34	5	8	5	21,82995	normální	58					.*				
8	16	169	39	89	65	96	3	Harnit	Techy .			5	2 ×	100			• •		7 A 1 1	-		
9	17	168,5	39	88	77	90	3	opravit	Touy		1000			80				*****		-		
10	17	160	37	89	65	90	3	Název	łady:								5 m	š • `	**			_
11	18	167	40	86	63	85	3	=\vis	ka*			= výška		60 -			-	4				
12	17	167	39	92	70	90	3	Hodoo	ty X čad-		100			40		* · * ·	15 C 1 C 1 C 1					
13	1/	165	39	82	65	89	3	[date	10000.00		6	- 102	10. 171.	40								
14	18	178	40	86	67	90	3	=0313	130.32:30	.\$111	6	193;	100; 1/1;	20 -								-
15	18	160	31	66	15	18	2	Hodno	ty <u>i</u> rad:		100	-										-
10	19	109	41	104	10	95	3	=data	1\$N\$2:\$N	\$111	1	= 90; 5	2; 62; 49	0				100	100	200		
10	20	1/5	43	90	00	100	2				OK		Storno	1	50 1	60	1/0	180	140	200	210	1
10	17	176	41	100	70	100	2															
20	17	164	20	100	60	90	20	6	5	6	10 00102	naduáha	1 541									

Dále do grafu vložíme křivku regresní funkce, a to tak, že si otevřeme menu řady (např. zmačknutím pravého tlačítka myši, pokud kurzor ukazuje na libovolný bod řady). V tomto menu vybereme možnost *Přidat spojnici trendu*.



Tato volba nám do grafu vloží regresní přímku a zároveň se nám otevře dialogové okno pro úpravu této přímky. Ve spodní části dialogového okna zaškrtneme možnosti *Zobrazit rovnici regrese* a *Zobrazit hodnotu spolehlivosti R* (= hodnotu indexu determinace).



Dále můžeme pomocí tohoto okna volit typ regresní funkce. Na výběr máme lineární model (implicitní možnost), dále model exponenciální, logaritmický, mocninný a různé modely polynomické, a to až do stupně 6.

V levé části tohoto okna můžeme volit různé možnosti pro úpravu vzhledu křivky regresní funkce, a to jak barvu, styl i tloušťku čáry.

Х Ча 1 В Z Ц - 2	10 · A' A' 学生 後) 学Zstamovat t - <u>2</u> · <u>A</u> · 学生 读 读 完 Stoudt a zar	eit rovnat na střed =	Obecný	000 (56 49	Podminěr	ié Formát ni jako tab	tovat Sty	Viož	t Odstran
ormát spojnice trendu		• •	Ćísl	o 19		Styly			Buňky
Možnosti spojnice trendu	Možnosti spojnice trendu			0	D	0			
Barva čáry	Typ trendu a regrese	90	/ P		R	0	1	0	
Stvl čárv		52 (4)-			_				
Stin	C Lipeárni	62 49,5 50	140	výš	ka ^v	= -0,0017x ³	³ + 0,9007x ² R ² = 0,60	- 158,9x + 9 15	326,5
	🖉 💿 Logaritnický	58 58,5 60	120			•	••	•	
	a planid a stat	00	00					*	
	Porag: 3	53 55	80			in		• ••	
	Bouranný Bouranný Bouranný Bouranný Bouraný průměr Obdobi: 2 ++	53 55 54 54 58	80		-			• ••	_
	Sourinný Souraný příměr Obdobi: 2 **	53 55 54 58 52 52	80 60 40		-	in the second		• •	_
	Sovjetanov Porsej: 3 v	53 55 54 54 58 52 62 70	80 60 40 20	1				• •	_
	Byrnamov Porse: 3	53 56 54 54 58 52 62 70 75	80 60 40 20 0			÷ ·		• • •	
	Begynamický porsej: 3	53 56 54 58 52 62 70 75 70	80 60 40 20 0 150	160	170	180	190	200	210
	Beynomocy Porse: 3 Beynomocy Porse:	53 55 54 54 58 52 62 70 75 70 75 70 51 34	80 60 40 20 0 150	160	170	180	190	200	210
	Beynamicky Porse: 3 Boanny Sourcevy promer globob: 2	53 55 54 54 58 52 62 70 75 70 51 51 54	80 60 40 20 0 150	160	170	180	190	200	210
	Beynamický Porsej: 3 Boanny Sousavý průměr Období: 2 Název spojnice trendu Mazev spojnice trendu M	53 55 ii 54 58 52 62 70 75 70 51 54 54 58 70	80 60 40 20 0 150	160	170	180	190	200	210
	Gymanický porsej: 3 Godobi: 2	53 55 ii 54 58 52 62 70 75 70 51 54 54 54 58 70 58 870 58	80 80 60 40 20 0 150	160	170	180	190	200	210
	Gynamický porsej: 3 Georgenný Souzevý průměr období: 2	53 55 ii 54 58 52 62 70 51 51 54 58 70 51 58 58 70 58 59 59	80 60 40 20 0 150	160	170	180	190	200	210

Pokud vyzkoušíme všechny volby typu regresního modelu, vidíme, že tři z nich by mohly být dobrými modely pro vyjádření vztahu mezi veličinami Výška a Váha. Těmito modely jsou lineární model, kvadratický model a polynomický model 3. stupně.





4.2.2 Jednoduchá regresní analýza pomocí doplňku prostředku Excel – lineární model

Abychom však správně vybrali z těchto modelů ten nejlepší, nemůžeme se spolehnout pouze na porovnání indexů determinace jednotlivých modelů, ale potřebujeme podrobnější regresní analýzu. K tomu musíme využít speciálního doplňku Excelu, a to *Analýzu dat*. Nejprve si ukážeme využití tohoto prostředku na jednoduché lineární regresi. Výsledkem by měla být regresní přímka s rovnicí, která je vidět na výše uvedeném obrázku.

Ca	6	7 - 194	1	2)	7.30						10.00	studer	ntske dot	azniky [R	ežim kompati	bility] - Micro	soft Excel		the surface	and have	-	أيسور	
	Dom	6 N	/ložení	R	ozložer	ni strá	inky 👘	Vzorce	Dat	Re	rize Zobr	azení	Acrobat										<u>v</u> _ = = ×
Z aplik Acces	ace 3			iných rojů -	Exist	ajici jeni	Aktua	C Ilizovat	Phpoje Vlastn 4 Upræi	ni osti t odkazy	21 2X X1 Sefadit	Filtr	i K. Vym Sy Zna Sy Upi	azət vu poulit esnit	Text do sloupců.	Debrat Ovële stejnë dat	ni Sloučit	Analýza hypotéz -	Seskupit C	ddēlīt so	*32 =35	obrazit p	notomosti 🦓 Analýza dat Nonosti 2. Řešitel
		Naőst e	steini d	ata				Phy	pojeni			Seladit a f	filtrovat			Datové na	istroje			0	snovs	1	G Analýza
	T17		(j,	s																	Nástroje pro analýzu dat
В	C	D	E	F	G	H	1	J	К	L	M	N	0	P	Q	R	S	Т	U	V	W	1 3	Nástroje pro analýzu finančnich a
1 vés	c výšk	a boty	prsa	pas	boky	krk s	sport s	pánek p	oc+tel	BMI	identifika	ceváha										1	védeckých dat
2 1	8 15	93 4	5 108	97	96	38	6	8	4	24,1617	2 normální	90									-	1	G FUNCRES
3 1	7 16	50 31	8 86	73	86	31	3	1	3	20,312	5 normální	52				ve	-0.0017x ³	0.9007x2	158 9x + 932	6.5	-		Další nápovědu zobrazite
6 1	7 40	C0 21	1 05	63	93	33	10	7	3,5	10 2260	1 normaini	10.6	-			1		R ² = 0,601	5	-	-		stisknutim klávesy F1.
6 1	7 16	50 J	70	64	00	33	3.6	6	1.6	17 6061	19 podváha	43,5	_	140						-	140 -		
7 1	7 16	63 3	9 92	68	100	34	5	8	5	21 8299	5 normální	58	_	120						_			
8 1	6 16	59 3	9 89	65	96	33	2	7	5	20 4824	8 normální	58.5	-	120				** [*]			120		
9 1	7 168	5 3	9 88	77	90	31	5	7	2.5	21.1325	3 normální	60		100					· · · · ·		100		
10 1	7 16	50 3	7 89	65	90	32	4	9	3	20,7031	13 normální	53		2.5					-		80		· · · · · · · · · · ·
11 1	8 16	67 40	86 0	63	85	31	4,5	5	5	19,7210	4 podváha	65		80	+ +	A 12				-		• • •	A States .
12 1	7 16	57 3	9 92	70	90	30	0	6	5	19,3624	7 podváha	54	4	Same 1	2. 5		* *	5			00	*4	3 Cong
13 1	7 16	66 35	9 82	65	89	31	3	4	3	19,5964	6 podváha	64	1	60		S. A.	4			-	40		
14 1	8 17	78 40	0 86	67	90	33	4,5	6	3,5	18,3057	7 podváha	58		40						_	20		
40 4	0 40	00 2	7 00	75	07	20	- 2	0	2	20.244	for the second fact	60		TV							11.00751		

V dialogovém okně, které se nám otevře po výběru tohoto doplňku, vybereme možnost *Regrese*.

0	216	19	01-	00	2)	Ŧ		students	ske dotaz	niky (Re	žim kon	npatibility] - Mi	crosoft E	xcel	Nästroje	e grafu								and the second	- 0
	2	Domů	Vi	zení	R	ozlože	ni str	ánky	Vzorce	Dat	a P	tevize Zobra	izení	Acrobat	Návrh	Rozlože	ni Fo	ormát							
Z aş	ilkace xess	Z z webu	Z	Z ji u zdi	iných rojů *	Exist	tujići ojeni	Aktur	C discovat) Pfipoj Vlastn © Uprav	eni Iosti It odkaz	24 2 Å Å↓ Sefadit	Filtr	K Vyma: ∳Znovi ∳Upřes	rat s použit init	Text do O	debrat Qv tejné d	effeni Sloučit.	Analýza hypotéz -	Seskupit C	Pall (●∃ Zol ●∃ Skr uhm	irazit podr jt podrobr	obnosti iosti	Analýza dat ? Řešitel
		Na	cist ext	erni di	ata			1	Pfij	po)eni		3	ieradit a f	filtrovat			Datove	e nastroje			0	snova		(#)	Analyza
	G	iraf 1		• (2	0	f _x																		
3	B	C	D	E	F	G	Н	1	J	K	L	M	N	0	Р	Q	R	S	T	U	V	W	X	Y	Z
2	18	193	46	108	97	96	38	6	8	4	24,16	172 normální	90												
3	17	160	38	86	73	86	31	3	7	3	20,3	125 normální	52	1							13			0	2
4	16	171	41	85	68	93	33	5	7	3,5	21,20	311 normální	62			with	a v	= -0,0017x3	Analýza dat	-				(B)	-
5	1/	160	39	32	63	86	33	10	1	4	19,33	594 podvaha	49,5			Vysk	a	20402533	Analytické n	astroie:				-	
0	17	163	40	02	69	100	31	3,5	0	1,5	21.92	039 pouvaria	00	140	1				Dyouvýběr	ový F-test or	o rozotvi			OK	
8	16	160	30	92	66	90	33	2	7	5	21,02	248 normální	69.6	120					Fourierova	analýza				Storno	
9	17	168.5	39	88	77	90	31	5	7	25	21 13	253 normální	60	***					Histogram	Amile					-
10	17	160,5	37	89	65	90	32	4	9	2.5	20 70	313 normální	53	100					Generátor	rumer oseudonáhod	ných čísel		- m 3	Nápověd	3
11	18	167	40	86	63	85	31	4.5	5	5	19 72	104 nodváha	55					+ 12 44 10	Pořadová s	tatistika a pe	ercently		1		
12	17	167	39	92	70	90	30	0	6	5	19.36	247 podváha	54	80		. 11	1 A	1. 1	Varkovini						
13	17	166	39	82	65	89	31	3	4	3	19.59	646 podváha	54	60			1000		Dvouvýběr	ový párový t	test na stře	dní hodnotu			
14	18	178	40	86	67	90	33	4.5	6	3.5	18,30	577 podváha	58	-	·	2 2.000	2.04		Dvouvýběr	ový t-test s r	rovnosti rozpi	tylů	-		
15	18	160	37	88	75	87	32	3	9	3	20,3	125 normální	52	40					<u></u>			_	_	_	
16	19	169	41	104	76	95	35	5	7	4	21,70	792 normální	62	20							-		_	_	
17	20	175	43	96	80	91	37	1	6	7	22,85	714 normální	70	1.55.0											
18	19	187	41	96	85	100	36	12	7	3	21,44	757 normální	75	0		See. 1	Sec.		10		-				
19	17	175	41	106	70	96	35	2,5	8	3	22,85	714 normální	70	1	50	160	170	180	190 2	00 2	210				
20	17	164	38	89	69	88	29	5	5	6	18,96	193 podváha	51	24							- 2				
2734	40	460	44	70	C0.	0.4	24	0	c	c .	10 0	nco estáte	6.8												

Následně vyplníme potřebné údaje v dialogovém okně. Povinně vybereme oblasti obsahující data. Doporučujeme do výběru zahrnout i hlavičky daných sloupců, ulehčí nám to následnou orientaci ve výstupu. V tomto případě však musíme zaškrtnout pole *Popisky*. Někdy můžeme mít potřebu upravit *Hladinu spolehlivosti*. Pokud bychom potřebovali velmi přesné a spolehlivé výsledky, můžeme tuto hodnotu navýšit až na 99 %, v případě nedostatku dat můžeme naopak tuto hodnotu snížit až na 90 %.

Dále si můžeme vybrat, kam nám má Excel umístit výstup. Tuto možnost však doporučujeme nevyužít a ponechat implicitní nastavení, které výstup umístí do nového listu.

Někdy můžeme potřebovat i informace o hodnotách reziduí, případně jejich grafické znázornění, v takovém případě zaškrtneme vybraná požadovaná pole.

Cin	16	17 -	01 -	0.0		Ŧ							stude	ntske dota	zniky (Reži	m kompatibi	lity] - Micros	oft Excel	8	- 3
-	D	omů	Vic	žení	R	ozlože	ni stra	ánky	Vzorce	Dat	a Revi	e Zobri	azení	Acrobat						
Z apli Acci	kace	Z webu Nac	Z test	Z ji zdr	ných njů *	Exist	ujid ojeni	Aktur	alizovat le -	D Připoji Vlastn Př Upravi pojení	ení osti t odkazy	24 2 A	Filtr Seřadit a	€ Vyma {> Znov ∑⁄ Upře filtrovat	izat u použit snit	Text do Oc sloupcû st	sebrat. Ověře tejně dat - Datové ná	ni Sloučil Istroje	I. Analyza hypotez	obnosti 🦓 Analýza dat nosti 2. kešitel
	8	81		• (2	10	f.e												Regrese	(international second
84 85 86 87 88 89 90 91 92 93 94 95 94 95 96 97 98 99 100	B 19 18 19 17 19 20 21 20 21 20 21 20 21 23 22 23 18 17 18 19	C 180 165 152 156 185 189 163 159 163 190 163 193 171 164 165 176 169 196	D 44 39 36 37 42 45 37 38 45 40 46 40 38 39 39 40 39 46	E 100 98 95 90 96 98 91 90 98 89 97 91 90 91 90 89 102	F 96 86 85 73 88 87 75 67 95 80 94 70 68 63 74 70 98	G 112 100 98 91 94 95 90 90 90 90 90 90 90 90 90 90 90 90 90	H 42 33 32 39 40 31 31 31 41 37 40 32 32 31 33 31 41	1 0 0 4 10 11 12 10 13 12 11 10 9 10 12 13 14 14 14 14 14 14 14 14 14 14	J 99 88 87 78 87 77 88 77 88 77 87 78 87 78 87 78 87 78 87 78 87 78 87 78 87 78 87 78 78	K 11 9 9 8 5 3 3 4 5 4 4 5 4 4 5 4 4 5 4 4 5 4 4 5 4 4 5 4 4 5 4 4 5 4 4 5 4 4 5 4 5 4 5 4 5 4 5 4 5 5 4 5	L 30,24691 31,2213 30,73064 22,60026 24,54346 23,71184 24,93076 24,4646 22,39076 24,4646 22,22906 22,30816 22,230816 22,230816 21,67126 20,98395 20,30741 21,3274	M obezita obezita obezita normální normální normální normální normální normální normální normální normální normální	N 98 85 71 55 84 85 63 60 90 65 60 65 90 65 60 59 90 55 8 90	0	P	Q	R	S	Vstup Vstupni oblast Y: Vstupni oblast Y: Vstupni oblast Y: Staffing Staffing St	OK Z Storno Népogèda
102 103	21	175 184	41 45	87 98	81 95	88 100	40	3	8	7	25,14286	nadváha nadváha	77							

Pro potvrzení zadání zvolíme *OK* a poté nám Excel vloží do nového listu následující tabulku. VÝSLEDEK

Regresní statistika	
Násobné R	0,738967
Hodnota spolehlivosti R	0,546073
Nastavená hodnota	
spolehlivosti R	0,54187
Chyba stř. hodnoty	11,66689
Pozorování	110

ANOVA

	Rozdíl	SS	MS	F	Významnost F	
Regrese	1	17684,71	17684,71	129,9236	3,09E-20	
Rezidua	108	14700,55	136,1162			
Celkem	109	32385,26				
		Chyba stř.		Hodnota		Horní
	Koeficienty	hodnoty	t Stat	Р	Dolní 95%	95%
Hranice	-110,739	16,2265	-6,82459	5,34E-10	-142,903	-78,5756
výška	1,061048	0,093087	11,39841	3,09E-20	0,876533	1,245564

Ve výstupu jsme obarvili pár důležitých polí. Žluté pole (Násobné R) nám udává hodnotu korelačního koeficientu. Hodnota 0,739 značí středně silnou lineární závislost mezi veličinami Výška a Váha.

Oranžové pole (Hodnota spolehlivosti R) udává hodnotu indexu determinace (můžeme zkontrolovat s hodnotou, kterou uvedl Excel v případě práce s grafy – viz výše). Uvědomme si, že platí $0,739^2 = 0,546$ (index determinace je druhou mocninou korelačního koeficientu). Hodnotu 0,546 můžeme interpretovat: "Variabilita veličiny Váha je z 54,6 % popsána veličinou Výška. Popis zbytku, neboli 45,4 % variability Váhy, je nutno hledat jinde." Zjednodušeně bychom také mohli říci, že model vystihuje data z 54,6 %.

Hnědé pole (Nastavená hodnota spolehlivosti R) udává hodnotu upraveného indexu determinace, který slouží k porovnání modelů s rozdílným počtem parametrů (regresních koeficientů).

Zelené pole udává hodnotu reziduálního rozptylu (tj. Q_e).

Vidíme, že obě růžová pole mají stejnou hodnotu. Toto platí pouze v případě lineární jednoduché regrese. Jejich hodnota vypovídá o významnosti modelu, respektive regresního koeficientu. Horní růžové pole "*Významnost F*" se týká významnosti modelu jako celku (testuje se nulová hypotéza, že všechny regresní koeficienty kromě absolutního členu jsou nulové). Spodní růžové pole "Hodnota P" se týká významnosti pouze jednoho regresního koeficientu, v našem případě koeficientu pro Výšku (testuje se nulová hypotéza, že tento koeficient je roven nule). Pokud je tato hodnota menší než zvolená hladina významnosti (v našem případě 0,05), tj. doplňková hodnota k hladině spolehlivosti, pak je model významný, respektive regresní model je statisticky významně odlišný od 0. V opačném případě (p-hodnota je větší než 0,05) je model nevýznamný. V našem případě je p-hodnota velmi malá $(3,09*10^{-20})$, tedy model významný je.

Modrá pole udávají hodnoty regresních koeficientů (možno opět zkontrolovat s grafem – viz výše). Na základě těchto hodnot tedy můžeme napsat rovnici regresní funkce, v našem případě přímky y = 1,061x - 110,739. Hodnota směrnice (tj. 1,061) nám říká, že o 1 cm větší osoba v průměru váží o 1,061 kg více.

Hodnota 1,061 je bodovým odhadem regresního koeficientu. Přesnější je však intervalový odhad, který nám určují červená pole. V našem případě vidíme, že směrnice má s 95% spolehlivostí hodnotu v rozmezí mezi 0,877 a 1,246.

4.2.3 Jednoduchá regresní analýza pomocí doplňku prostředku Excel – obecný model

Poněkud složitější bude tvorba obecnějšího modelu jednoduché regrese. K tomu účelu si nejprve musíme speciálním způsobem upravit data, speciálně oblast obsahující údaje o veličině x. Tato úprava vychází z tvaru matice F, o které jsme mluvili v teoretickém úvodu v části o maticovém vzorci pro výpočet vektoru regresních koeficientů.

Oblast dat obsahující údaje o veličině x bude obsahovat tolik sloupců, kolikrát se x objevuje v rovnici požadovaného regresního modelu, každý z těchto výskytů je ve tvaru nějaké funkce, označme ji $f_i(x)$. Jednotlivé sloupce tedy budou obsahovat $f_i(x)$.

Uveď me si několik příkladů:

Lineární model má tvar $y = b_1 x + b_0$, proto bude oblast obsahovat jeden sloupec a v něm hodnoty *x*.

Kvadratický model má tvar $y = b_2 x^2 + b_1 x + b_0$, proto bude oblast obsahovat dva sloupce, v prvním z nich hodnoty x, v druhém x^2 .

Polynomický model 3. stupně má tvar $y = b_3 x^3 + b_2 x^2 + b_1 x + b_0$, proto bude oblast obsahovat tři sloupce, v prvním z nich hodnoty x, v druhém x^2 , ve třetím x^3 .

Logaritmický model má tvar $y = b_1 \ln x + b_0$, proto bude oblast obsahovat jeden sloupec a v něm hodnoty ln x.

My nyní chceme vytvořit kvadratický model, proto do listu vložíme nový sloupec obsahující hodnoty x^2 (sloupec obsahující x již v souboru máme).

0	10.	7 - (4 - 🗋	<u>a)</u> =			
6	Dom	u Vložení	Rozio	žení stránky	Vzorce	Data
Vio	and	Arial B I U -	- 10 - 10	A' A' - A -	= = <mark>-</mark> = = =	≫~ 课 键
Schr	ánka 🕞	Pi	ismo	6		Za
	J19	•	(*	f _x		
	A	В	С	D	E	F
1	výška		váha			
2	х	x ²	y			
3	193	37249	90			
4	160	25600	52			
5	171	29241	62			
6	160	25600	49,5			
7	400	00564	50			

Nyní v nabídce *Data* vybereme doplněk *Analýza dat* a v následném dialogovém okně vybereme možnost *Regrese*. Jako vstupní oblast Y vybereme sloupec obsahující údaje o váze respondentů, jako vstupní oblast X vyznačíme oba sloupce obsahující informace o položce výška. Pozor – jako popisky můžeme vyznačit pouze jeden řádek. Ostatní části dialogového okna vyplníme obdobně, jako jsme již vyplňovali v případě tvorby lineárního modelu.

Cin	100	• (* •]]	•					3	studentske	dotazniky [Re	tim komp	atibility]
-	Domů	vložení	Rozlož	eni stri	inky Vzorce	Data	Revize	Zobraze	ni Acr	obat		
Z ap Ao	likace Z cess wet	z zj su testu zdr lačist externi d	ných Eo při	itujid pojeni	Aktualizovat vše - sa U Připoje	ipojeni astnosti pravit odka ni	24 %1	2 A Sefadit	Filt S	Vymazst Znovu použit Upřesnit vat	Test do sloupcù	Oidebri stejne
	B112	• (9	fx								
	A	B	C	D	E	F	G	H	1	J	K	
1	výška	1.000000000	váha	-	6					2	D	
2	x	x ²	y		Regrese					(con)		
3 1	193	37249	90		Vstup							
4 1	160	25600	52		Vstunní oblast Y		ene	2-606112		OK		
5 ;	171	29241	62					LIPOPTIL	(1469)	Storno		
6	160	25600	49.5		Vstupní oblast X		SAS	2:\$8\$112	1			
7	169	28561	50		and the second second				100	Ninwida		
8	163	26569	58		Popisky		Konsta	nta je nula		[aboTcon		
9	169	28561	58,5		🔄 Hadina spok	shilvosti	95 9	Ne				
10 ;	168,5	28392,25	60									
11	160	25600	53		Možnosti výstupu				-			
12 }	167	27889	55		Výstupní obla	ist:			1			
13	167	27889	54		A Neuró lete							
14	166	27556	54		C mort inc		-					
15	178	31684	58		O Novy sesit							
16	160	25600	52		Rezidua							
17	169	28561	62		Bezidua		Grafs	rezigui				
18	175	30625	70		Standardní re	zidua	Grafre	gresní pří <u>m</u> k	У			
19	187	34969	75		Normální pravdě	teondoboost						
20	175	30625	70		Graf pravděc	odobnosti						
21	164	26896	51		T Tan brandt							
22	169	28561	54								-	
23	172	29584	54			_	_		-		THE R	
0.1		anner										

Excel nám do nového listu umístí následující výstup.

VÝSLEDEK

Regresní statist	ika					
Násobné R	0,753741					
Hodnota spolehlivosti R	<mark>0,568126</mark>					
Nastavená hodnota						
spolehlivosti R	<mark>0,560053</mark>					
Chyba stř. hodnoty	11,43301					
Pozorování	110					
ANOVA					14/	
	Dordíl	<u> </u>	MC	r	vyznamnost	
	Rozali	33	MS	F	F	
Regrese	2	18398,91	9199,453	70,37868	<mark>3,1E-20</mark>	
Rezidua	107	13986,36	<mark>130,7136</mark>			
Celkem	109	32385,26				
		Chyba stř.		Hodnota		Horní
	Koeficienty	hodnoty	t Stat	Р	Dolní 95%	95%
Hranice	<mark>415,5727</mark>	225,7234	1,84107	0,068382	-31,8976	863,043
Х	<mark>-4,97572</mark>	2,584211	-1,92543	0,05683	-10,0986	0,147181
x2	0,017229	0,007371	2,337475	0,021275	0,002617	0,03184

Ve výstupu jsme barevně vyznačili nejdůležitější pole pro vyhodnocení modelu a následný výběr modelu nejlepšího. Tato pole později zaneseme do přehledné tabulky obsahující údaje o všech vytvářených modelech. Jen si nyní uvědomme, že model je významný (p-hodnota jeho významnosti je rovna 3,1*10⁻²⁰) a taktéž je významný kvadratický koeficient (p-hodnota jeho významnosti je rovna 0,021275), tedy se opravdu jedná o kvadratický model.

Nyní obdobně vytvoříme polynomický model 3. stupně. Nejprve upravíme vstupní oblast X, a to tak, že přidáme další sloupec. Tento sloupec bude obsahovat třetí mocniny *x*. Uvědomme si, že celá vstupní oblast X musí být kompaktní, tedy všechny její sloupce musí být bezprostředně vedle sebe.

8	Domú	· OI · D	Rotiožen	i stránky	studentske dotazniky [Režim kompatibility] - N Vrove Data Pestre Zobrazeni Acrohat
Z ap Ac	Alikase Z cess webs	Z Z u textu zo ačist externí o	jiných Existu připid tata	jid Aktua eni vii	Pripojeri 24 214 Y Kymazat Vlastnosti 24 Sefadt Futr V Ustrosti Stovat - Upravit odkazy 24 Sefadt Futr V Upfernit Soupcia, stojne Data
	C112	•	6 1	c	Regrese
1	A	В	С	D	Vstup Vstupní oblast Y: SD52: SD5112 K
2	X 193	x ² 37249	x ³ 7189057	y 90	Vstupní oblast <u>X</u> : \$A\$2:\$C\$112 Storno
4	160	25600	4096000	52	Popisky Gonstanta je nula Nápověda
6	160	25600	4096000	49,5	E Hadina spolehlivosti 95 %
7 :	169	28561	4826809	50	Možnosti výstupu
9	169	28561	4826809	58,5	Výstupní oblast:
10	168,5	28392,25	4/84094,1	53	Nový sešt
12	167	27889	4657463	55	Rezidua
13	167	27889	4657463	54	🔄 🔄 Bezdua 🔄 Graf s rezigui
14	166	27556	4574296	54	🔄 Standardní rezidua 📃 Graf regresní přigky
15	178	31684	5639752	58	Number and descent
16 ;	160	25600	4096000	52	Promain pravoepodonost
17	169	28561	4826809	62	El X al higrachandrivea
18	175	30625	5359375	70	
19	187	34969	6539203	75	

Excel nám do nového listu umístí následující výstup.

VÝSLEDEK

Regresní statistika	
Násobné R	0,77559
Hodnota spolehlivosti R	<mark>0,60154</mark>
Nastavená hodnota spolehlivosti R	<mark>0,590263</mark>
Chyba stř. hodnoty	11,03349
Pozorování	110

ANOVA

	Rozdíl	SS	MS	F	Významnost F
Regrese	3	19481,04	6493,68	53,34145	<mark>4,27E-21</mark>
Rezidua	106	12904,22	<mark>121,738</mark>		
Celkem	109	32385,26			

	Koeficienty	Chyba stř. hodnoty	t Stat	Hodnota P	Dolní 95%	Horní 95%
Hranice	<mark>9326,493</mark>	2996,716	3,112238	0,002387	3385,212	15267,77
x	<mark>-158,902</mark>	51,68809	-3,07424	0,002684	-261,378	-56,425
x2	<mark>0,900658</mark>	0,296394	3,038719	0,002992	0,313028	1,488289
x3	<mark>-0,00168</mark>	0,000565	-2,98145	<mark>0,003559</mark>	-0,0028	-0,00056

Ve výstupu jsme opět vyznačili žlutě nejdůležitější pole pro vyhodnocení modelu a následný výběr modelu nejlepšího. Tato pole opět zaneseme do závěrečné tabulky obsahující údaje o všech vytvářených modelech. Vidíme, že tento model je významný (p-hodnota jeho významnosti je rovna $4,27*10^{-21}$) a taktéž je významný vedoucí koeficient, tedy koeficient u x^3 (p-hodnota jeho významnosti je rovna 0,003559), tedy se opravdu jedná o polynomický model stupně 3.

A nyní si již můžeme vytvořit již zmíněnou závěrečnou tabulku.

	Významnost modelu	Významnost vůdčího koef.	Index det./ upr.index det.	Rezid. rozptyl	Rovnice
lineární	3,09*10 ⁻²⁰	3,09*10 ⁻²⁰	0,546/ 0,542	136,12	<i>y</i> = 1,061 <i>x</i> -110,739
kvadratický	3,1*10 ⁻²⁰	0,021275	0,568/ 0,560	130,71	$y = 0,017229 x^2 - 4,97572 x + 415,5727$
polynom 3.st.	4,27*10 ⁻²¹	0,003559	0,602/ 0,590	121,74	$y = -0,00168 x^{3} + 0,900658 x^{2} - 158,902 x + 9326,493$

Jednotlivé indexy determinace a rovnice jednotlivých regresních modelů můžeme porovnat s výsledky v rámci grafického zpracování. Je však přirozené, že index determinace je vyšší, čím vyšší je mocnina v polynomické funkci. Zvýhodňování složitějších modelů (tj. modelů s vyšším počtem regresních koeficientů) je základní vlastnost a nevýhoda indexu determinace. Proto jsme nemohli udělat hodnověrný závěr už v první fázi zpracování.

Nyní však již máme všechny potřebné údaje. Vidíme, že všechny vytvořené modely jsou významné. Významné jsou i regresní koeficienty u nejvyšší mocniny v polynomu. Jedná se tedy o využitelné modely. Rozhodujícím faktorem vyhodnocení tedy bude porovnání hodnot upravených indexů determinace a reziduálních rozptylů jednotlivých modelů. Z teoretického úvodu připomeňme tři kritéria pro porovnání modelů. Víme, že

- čím nižší reziduální rozptyl, tím lepší model;
- čím vyšší upravený index determinace, tím kvalitnější model.
- čím vyšší upravený index determinace, tím lepší model.

Ze závěrečného vyhodnocení tedy vychází vítězně polynom 3. stupně. Vidíme, že v rámci tohoto modelu je 60,2 % změn veličiny váha vysvětleno změnami veličiny výška a 39,8 % zůstalo nevysvětleno.

4.2.4 Vícenásobná regresní analýza pomocí doplňku prostředku Excel

Z výše uvedeného je zřejmé, že ještě musíme hledat další faktory mající vliv na váhu člověka. V našem průzkumu byly sledovány ještě veličiny Pohlaví, Věk, Velikost bot, Obvod pasu, Obvod prsou, Obvod boků, Obvod krku, Týdenní počet hodin strávený sportem, Týdenní počet hodin strávený u TV a PC, Denní počet hodin strávený spánkem.

Budeme sledovat vliv všech těchto faktorů. K tomu využijeme vícenásobné regrese, speciálně Stepwise regrese. Než však začneme dělat samotnou regresi, měli bychom zkontrolovat vzájemnou korelaci mezi jednotlivými regresory. K tomu využijeme další z možností doplňku *Analýza dat*, a to *Korelace*.

Cho	120	- 0	12)	÷		-	-	-	st	udentske de	otazniky [f	Režim kompa	tibility] -	Microsoft Excel							- 9	E
-	Domi	Viože	ni s	Rozložení s	tránky	Vzorce	Data	Revize	Zobrazeni	Acrob	at											- 0
Z ap Ac	a Li Ikace z rss we	bu tetu	Z jiných zdrojů -	Existujio	i Ak	tualizovat vše -) Pfipojeni Vlastnost P Upravit or	± tary ⊼↓	ZZA Sefadit P		mazat Iovu použit přesnit	Test do	Odebrat stejné	Ověření Sloučit dat	Analýza hypotéz	4[] Seskupit	Pall (Oddělin – Sk	100 e1 euton	Zobrazit podr Skrjt podrobr	obnosti nosti	Panalýza dat ?	
	1	Načist extern	i data			Při	pojeni		Seřad	iit a filtrovat	8		Dat	tové nástroje			0	snova		(8)	Analýza	
	08	1 10	- (a	fx																		
2	A	В	C		D	E	F	G	Н	1	J	K	L	. M	N	0	P	Q	R	S	T	U
1 1	ék	výška	boty	prsa		pas	boky	krk	sport	spánek	pc+tel	váha										
2	18	193		46	108	97	96	30		5	8	4	90									
3	20	1/:	1	43	96	80	9 9	31			6	1	10					10	A m			
4	20	1/8		43	110	85	90	46			8	3	10	Korelace				(ing				
5	20	181		42	35	6/	100	31			9	11	0	Vstup				-		-		
0	20	100	-	44	107	92	2 91	31		2	3	11	5/	Vstupní oblast:		taet-orea	G 🛒		×	-		
0	10	19		43	112	93	440	34	4		0	44	00			Carrie and	- (C158	Ste	orno	-		
0	13	100		44	110	04	400	41			9	0 1	10	Sdružit:		 Soupce 		Land Sector		-		
10	20	190	-	40	00	90	100	4		1	0	12 1	26			Rágky		Nápo	oyěda	-		
11	22	183		43	98	93	3 100	40			8	12	36	Popisky v prvi	nim řádku							
12	21	180		44	111	95	120	4		1	7	12 1	12	Malazzati solati mu								
13	20	182	1	45	110	102	12	4		1	7	9 1	10	mou kulo vystopu	22	-	100	1				
14	21	187		42	107	96	113	42		1	7	7 1	00	Vystupni oblac	st:		1.001					
15	22	180	1	44	110	96	5 120	40	0)	8	11 1	00	Nový list:								
16	21	192	1	46	120	100	12	43		1	8	11 1	10	🔘 Nový sešit								
17	22	190	Ň.	45	123	112	2 125	44	()	7	11 1	13	1.00								
40	22	470		42	100	00	100	45		1	0	11 1	10		_	_	_	_				

V dialogovém okně jednak označíme celou souvislou oblast s daty, tedy všechny sloupce obsahující údaje o jednotlivých zmíněných veličinách. Zde je víceméně nutno vybrat i hlavičky sloupců a tedy zatrhnout pole *Popisky v prvním řádku*. Opět doporučujeme ponechat implicitní možnost výstupu do nového listu.

Výsledkem je následující korelační matice (tj. tabulka) obsahující hodnoty vzájemných korelačních koeficientů. Matice je pouze trojúhelníková, a to z důvodu, že korelace je vztah vzájemný, symetrický, nemusíme tedy vyplňovat zbylé údaje. Jedničky na diagonále znamenají, že každá z veličin je sama se sebou dokonale korelovaná, což je naprosto přirozené a nemůže tomu být jinak.

	pohlaví	věk	výška	boty	prsa	pas	boky	krk	sport	spánek	pc+tel
pohlaví	1,00										
věk	0,23	1,00									
výška	0,80	0,16	1,00								
boty	0,87	0,16	0,93	1,00							
prsa	0,72	0,19	0,70	0,74	1,00						
pas	0,84	0,19	0,71	0,79	0,82	1,00					
boky	0,64	0,22	0,62	0,66	0,84	0,80	1,00				
krk	0,92	0,18	0,78	0,85	0,80	0,87	0,74	1,00			
sport	0,07	0,02	0,20	0,08	-0,07	-0,12	-0,20	-0,01	1,00		
spánek	0,17	0,07	0,07	0,12	0,37	0,33	0,32	0,25	-0,03	1,00	
pc+tel	0,31	0,11	0,10	0,24	0,36	0,49	0,47	0,34	<mark>-0,5</mark> 7	0,33	1,00

Žlutě jsou vyznačeny hodnoty korelačních koeficientů, které jsou větší než 0,5 – tedy značí určitou míru korelace. Hodnoty nad 0,7 znamenají již významnou korelaci. Znaménko mínus u určitých hodnot značí, že daný vztah je nepřímý. Například hodnota -0,57 korelačního koeficientu mezi veličinami Sport a PC+TV značí středně silnou závislost a dále nám dává informaci, že čím více osoby sportují, tím méně tráví času u PC+TV a naopak. Vidíme, že veličina Věk není korelována s žádnou z ostatních veličin. Naopak veličiny Pohlaví a Krk

jsou korelovány s většinou ostatních veličin. Není vhodné, aby model vícenásobné regrese obsahoval vzájemně korelované faktory, protože toto dává zkreslenou informaci o síle působení těchto faktorů. Pokud data obsahují vzájemně korelované faktory, je toto nutno nějak řešit. Možností je mnoho. Nejjednodušší, ne vždy však dobře použitelné je některé z těchto faktorů z analýzy vypustit. Někdy ale nelze dobře určit, který z faktorů vypustit. Pokud nepoužijeme jiný způsob řešení této situace, musíme si aspoň dát pozor, aby se ve výsledném modelu tyto faktory nevyskytovaly společně. Složitějšími možnostmi jsou některé vícerozměrné metody, např. Metoda hlavních komponent, které nám na základě našich vstupních faktorů vytvoří umělé faktory, které jsou již nekorelované.

V případě silné korelovanosti naší veličiny Pohlaví můžeme využít možnosti rozdělení souboru na dvě části – muže a ženy. Další šetření pak budeme provádět zvlášť pro muže a zvlášť pro ženy. Vytvoříme tak dva regresní modely, které pak můžeme vzájemně porovnávat.

ženy	věk	výška	boty	prsa	pas	boky	krk	sport	spánek	pc+tel
věk	1,00									
výška	0,05	1,00								
boty	-0,04	0,76	1,00							
prsa	0,10	0,05	0,00	1,00						
pas	0,09	-0,23	-0,09	0,44	1,00					
boky	0,15	-0,02	0,04	0,60	0,58	1,00				
krk	-0,12	0,07	0,15	0,45	0,41	0,50	1,00			
sport	-0,04	0,26	0,07	0,03	-0,34	-0,11	0,10	1,00		
spánek	0,07	-0,21	-0,16	0,51	0,48	0,48	0,36	-0,07	1,00	
pc+tel	0,21	-0,32	-0,09	0,23	0,62	0,45	0,05	-0,48	0,37	1,00

Následující dvě tabulky obsahují korelační matice v případě žen a mužů odděleně.

Vidíme, že v souboru obsahujícím údaje o ženách jsou již faktory korelovány minimálně. Nejzávažnější je korelace mezi veličinami Boty a Výška. V tomto případě můžeme situaci řešit vypuštěním veličiny Boty z další analýzy.

muži	věk	výška	boty	prsa	pas	boky	krk	sport	spánek	pc+tel
věk	1,00									
výška	-0,15	1,00								
boty	-0,17	0,82	1,00							
prsa	-0,03	0,54	0,59	1,00						
pas	-0,15	0,63	0,61	0,73	1,00					
boky	0,10	0,42	0,42	0,78	0,78	1,00				
krk	0,01	0,32	0,37	0,56	0,56	0,53	1,00			
sport	0,06	0,19	0,00	-0,34	-0,32	-0,43	-0,44	1,00		
spánek	-0,09	0,10	0,11	0,22	0,06	0,24	0,12	-0,01	1,00	
pc+tel	-0,17	-0,17	-0,06	0,20	0,26	0,36	0,23	-0,75	0,21	1,00

V souboru obsahujícím údaje o mužích se vyskytuje více významných hodnot korelačních koeficientů, musíme si tedy v dalším zpracování dávat větší pozor.

Nyní budeme provádět vícenásobnou regresi v souboru žen. Nejprve použijeme tzv. metodu Enter, tedy do regrese zahrneme všechny sledované faktory. Opět využijeme prostředek *Regrese* v rámci doplňku *Analýza dat*. Abychom mohli dobře označit *Vstupní oblast dat X*, musíme si dát pozor, aby sloupce obsahující tyto faktory byly všechny vedle sebe. Do *Vstupní oblast X* pak tažením myši vyznačíme všechny tyto sloupce naráz.

0	200	7 - 00 -	14)		-		-		stud	lentske dot	azniky [Rež	im kompatibili	ty] - Microsoft Excel			-	0	1
	Dor	ů Viož	ení R	ozložení stri	ánky Vz	orce	Data R	evize	Zobrazeni	Acrobat							0 -	
Z a	Dilkace ccess w	2 P z z tbu textu	Z jiných zdrojů *	Existující připojení	Aktualizov vie *		Připojení Vlastnosti Upravit odkazy	24 24	2 Z	r X.Vym ≦sZno ⊻ Upř	azāt vu použit esnit	Text do Od sloupcû sti	ebrat Ověření Sloučit Anal giné dat	gas Seskupit Oddélit Soul	9 Zobrazit pod 9 Skrjit podrob	irobnosti inosti	Analýza dat	
	50	macist exter	ni data.	6	1	Phipoj	Jean		Setadit	a filtrovat			Datove nastroje	Un	ova	7//5/1	мпануга	
	23		• (1	Jx				0	н			V	(.		0 7	0	T	_
1	věk	výška	boty	orsa	nas	h	rokv kri	e e	snort s	nánek	nc+tel	váha	Regrese		(eren)	3		
2	1	7 16	0	38	86	73	86	31	3	7	3	52	Vstup					
3		6 17	1	41	85	68	93	33	5	7	3.5	62	Vstupní oblast <u>Y</u> :	\$K\$1:\$K\$67	OK			
4	1	7 16	0	39	82	63	86	33	10	7	4	49,5	Web and able at Mr.		Storno			
5	1	7 16	9	40	79	64	90	31	3,5	6	1,5	50	vstupni obiast <u>X</u> :	\$A\$1:\$J\$67				
6	1	7 16	3	39	92	68	100	34	5	8	5	58	Denistry	Koostanta la mila	Nápoyěda			
7	1	6 16	9	39	89	65	96	33	2	7	5	58,5	E Copiery					
8	1	7 168	5	39	88	77	90	31	5	7	2,5	60	Ciaona spolenivoso	23 76				
9	3	7 16	0	37	89	65	90	32	4	9	3	53	Možnosti výstupu					
10	1	8 16	7	40	86	63	85	31	4,5	5	5	55	Contraction of	1941				
11	61	7 16	7	39	92	70	90	30	0	6	5	54	 vystupni obiast: 	E (1996)				
12	1	7 16	6	39	82	65	89	31	3	4	. 3	54	Nový list:					
13	1	8 17	8	40	86	67	90	33	4,5	6	3,5	58	🔿 Nový sejšit					
14	1	8 16	0	37	88	75	87	32	3	9	3	52	Rezidua					
15		9 16	9	41	104	76	95	35	5	7	4	62	Rezidua	Graf s rezidui				
16		9 18	7	41	96	85	100	36	12	7	3	75	Standardní rezidua	Graf regresni přímky				
17	1	7 17	5	41	106	70	96	35	2,5	8	3	70						
18	1	7 16	4	38	89	69	88	29	5	5	6	51	Normainí pravděpodobno	ost				
19		8 16	9	41	70	68	84	31	0	6	6	54	Graf pravdépodobno	50				
20	2	1 17	2	40	86	63	89	29	9	5	5	54						
21	2	1 16	5	38	92	71	93	31	3	8	8	58	<u> </u>					
22	2	0 17	1	39	97	75	97	34	5	7	5	70						

Výstup jsme si opět nechali umístit do nového listu. Vzniklá tabulka je nyní poněkud složitější než v případě jednoduché regrese.

VÝSLEDEK

Regresní statistika								
Násobné R	<mark>0,886917</mark>							
Hodnota spolehlivosti R	0,786623							
Nastavená hodnota spolehlivosti R	0,747827							
Chyba stř. hodnoty	5,123442							
Pozorování	66							

ANOVA

	Rozdíl	SS	MS	F	Významnost F
Regrese	10	5322,36	532,236	20,27592	<mark>4,77E-15</mark>
Rezidua	55	1443,731	26,24966		
Celkem	65	6766,091			

	Koeficientv	Chyba stř. hodnotv	t Stat	Hodnota P	Dolní 95%	Horní 95%
Hranice	-159 507	24 33852	-6 55366	2.03E-08	-208 282	-110 731
věk	0 522515	0.2400002	1 525607	0.420246	0 15066	1 206690
VER	0,525515	0,340090	1,555697	<mark>0,130340</mark>	-0,15900	1,200009
výška	0,238201	0,15998	1,488941	<mark>0,142215</mark>	-0,08241	0,558808
boty	1,404614	0,79341	1,770351	<mark>0,082211</mark>	-0,18542	<mark>2,994644</mark>
prsa	0,094408	0,171444	0,550662	<mark>0,584095</mark>	-0,24917	0,437988
pas	<mark>0,711686</mark>	0,135465	5,253662	2,49E-06	<mark>0,440208</mark>	<mark>0,983163</mark>
boky	0,680456	0,225227	3,021206	0,003816	0,229092	1,131821
krk	-0,56238	0,579918	-0,96975	<mark>0,336415</mark>	-1,72456	0,599804
sport	-0,11926	0,213659	-0,55817	<mark>0,578995</mark>	-0,54744	0,308925
spánek	1,838993	0,634988	2,896105	0,005412	0,566448	3,111538
pc+tel	-0,02808	0,391058	-0,07182	<mark>0,943008</mark>	-0,81178	0,755613

Tmavě modré pole značí hodnotu vícenásobného korelačního koeficientu, hodnotícího působení všech faktorů dohromady. Zelené pole značí hodnotu indexu determinace. V našem případě lze tedy říci, že variabilita veličiny Váha je modelem vystižena z 78,66 %. Růžové pole určuje p-hodnotu významnosti celého modelu. Její hodnota (4,77*10⁻¹⁵) je velmi nízká, model je tedy významný.

Žlutě jsou vyznačeny p-hodnoty, které jsou vyšší než 0,05, tedy ukazují, že příslušný koeficient není významný. P-hodnoty menší než 0,05 jsme nevyznačili, tyto příslušné koeficienty významné jsou. Povšimněme si modře a červeně vyznačených polí. Červeně vyznačený je koeficient s hodnotou 1,404614, modře pak koeficient s hodnotou 0,711686. Vidíme, že červený koeficient má vyšší hodnotu, přesto dle p-hodnoty není významný. Oproti tomu modrý koeficient má nižší hodnotu a přesto dle p-hodnoty významný je. Toto je

příkladem, že významnost koeficientů nelze posuzovat "od oka". Opravdu se může stát, že hodnota 1000 bude nevýznamná, naopak hodnota 0,001 významná bude.

Z výstupu vidíme, že se v modelu vyskytují nevýznamné koeficienty, tedy model není dobrý. Použijeme tedy metodu stepwise regrese k postupnému budování výsledného modelu. V prvním kroku vytvoříme tolik jednoduchých lineárních regresí, kolik máme faktorů. Každá z těchto regresí vytvoří výstup do samostatného listu. Do přehledné tabulky si z těchto výstupů přeneseme p-hodnoty významnosti jednotlivých modelů. V následující tabulce jsme si takto vytvořili první sloupec. Řádek Boty je prázdný, protože jsme výše zjistili, že tato veličina je silně korelovaná s veličinou výška a z další analýzy jsme ji vyjmuli. Žlutě jsou v tomto sloupci vyznačeny p-hodnoty větší než 0,05, které ukazují na nevýznamnost příslušného modelu. Z ostatních p-hodnot (tedy hodnot, které ukazovaly na významné modely) jsme vybrali tu nejmenší a označili ji modře. Tato hodnota (1,8295*10⁻¹²) náleží k modelu popisujícímu vztah váhy a obvodu pasu a ukazuje, že tento model je nejvýznamnější. Proto veličinu Obvod pasu vybereme jako nejsilnější faktor do dalšího zpracování. Výsledkem 1. kroku je tedy model s jedním faktorem.

V druhém kroku budeme vytvářet několik modelů vícenásobné regrese, každý z nich bude obsahovat dva regresory. V každém z těchto modelů bude jedním ze dvou regresorů faktor Obvod pasu. Druhým regresorem budou postupně všechny faktory, jejichž modely byly v 1. kroku identifikované jako významné. Do části tabulky obsahující informace o tomto 2. kroku jsme si zaznamenali jednak p-hodnoty regresních koeficientů příslušných přidávaných regresorů, jednak p-hodnoty významnosti celého modelu.

	1.krok	2.krok = pas	;+	3.krok = pa	s+boky+
	jednoduch.	p-hodnota	významnost	p-hodnota	významnost
věk	0,0718902				
výška	0,29303342				
boty					
prsa	3,1714E-06	0,003808	2,985E-13	0,207638	1,2229E-14
pas	1,8295E-12				
boky	8,0392E-11	2,94E-05	3,071E-15		
krk	0,00098189	0,213433	<mark>9,3E-12</mark>		
sport	0,08071724				
spánek	1,7736E-06	0,006516	4,869E-13	0,063926	4,8609E-15
pc+tel	1,6632E-05	0,512528	<mark>1,64E-11</mark>		

Výsledky stepwise regrese - ženy

Žlutě jsme vyznačili případy, kdy daný model nepřináší vylepšení modelu z předchozího kroku. Poznáme to jak na p-hodnotě daného koeficientu, a to tak, že tato hodnota je větší než 0,05. Také to poznáme z p-hodnoty významnosti daného modelu (např. u faktoru Krk je tato

hodnota $9,3*10^{-12}$) – ta je větší než p-hodnota významnosti modelu z předchozího kroku (1,8295*10⁻¹²).

Ze zbylých modelů, tedy modelů, které přinášejí určité vylepšení, vybereme ten, který má p-hodnotu významnosti ze všech nejmenší, daný model je tedy nejvýznamnější. V našem případě je to modře vyznačená hodnota 3,071*10⁻¹⁵, která určuje, že dalším faktorem vstupujícím do modelu je veličina Boky.

Obdobně postupujeme i v 3. kroku. Zde vytváříme modely vícenásobné regrese obsahující tři regresory, dvěma z nich jsou vybrané faktory Obvod pasu a Obvod boků. Třetím faktorem jsou postupně všechny faktory, jejichž modely v 2. kroku byly významné (v našem případě to budou postupně faktory Obvod prsou a Množství spánku). Do tabulky si opět zaznamenáme dvě vybrané hodnoty pro každý z těchto modelů.

Jak vidíme, pro oba tyto modely platí, že nejsou významnější než model z 2. kroku (jejich p-hodnoty významnosti nejsou menší, respektive p-hodnoty regresních koeficientů jsou větší než 0,05).

Jako příklad ze všech výstupů jednotlivých regresí si uvedeme výstup závěrečného kroku stepwise regrese.

VÝSLEDEK

Regresní statistik	a
Násobné R	0,808604
Hodnota spolehlivosti R	<mark>0,65384</mark>
Nastavená hodnota spolehlivosti R	0,642851
Chyba stř. hodnoty	6,097295
Pozorování	66

ANOVA

	Rozdíl	SS	MS	F	Významnost F	
Regrese	2	4423,94	2211,97	59,49834	3,07E-15	
Rezidua	63	2342,151	37,177			
Celkem	65	6766,091				
	Koeficienty	Chyba stř. hodnoty	t Stat	Hodnota P	Dolní 95%	Horní 95%
Hranice	<mark>-78,7155</mark>	16,73245	-4,70436	1,44E-05	-112,153	-45,2783
boky	0,979094	0,217337	4,504949	2,94E-05	0,54478	1,413408
pas	0,698306	0,126248	5,531227	6,54E-07	0,44602	0,950592

Zeleně jsou vyznačeny důležité hodnoty. Jednak jsme vyznačili hodnotu indexu determinace, která nám říká, že výsledný model vystihuje variabilitu veličiny Váha z 65,38 %, tedy zbývajících 34,62 % popisují ještě jiné faktory.

Výsledný model má tedy tvar:

V = -78,7155 + 0,979094*B + 0,698306*P,

kde V je váha v kilogramech, B je obvod boků v centimetrech a P je obvod pasu v centimetrech.

Vidíme, že pokud má žena v pase o 1 cm více, bude těžší průměrně o 0,698 kg a pokud bude mít žena přes boky o 1 cm více, bude těžší průměrně o 0,979 kg.

4.3 Regresní analýza v SW STATISTICA

4.3.1 Grafické znázornění

Pokud chceme provést regresní analýzu, vytvoříme si nejprve graf. V SW STATISTICA je pro tyto účely nejvhodnější bodový graf. V nabídce zvolíme *Grafy - Bodové grafy*. V dialogovém okně vybereme jako nezávislou proměnnou výšku a jako závislou proměnnou váhu. Zkontrolujeme, zda je zaškrtnuto pole *Typ proložení: Lineární*.



Pokud bychom chtěli mít graf s podrobnějšími informacemi, zvolíme záložku *Detaily* a nastavíme své požadavky. Mezi možnostmi je např. zobrazení regresních pásů. V tomto případě můžeme volit spolehlivost pro intervalové odhady. Volit můžeme mezi regresními pásy pro střední hodnotu (v dialogovém okně označeno *Spolehlivost –* naše volba), nebo regresní pásy pro individuální hodnotu (v dialogovém okně označeno *Predikce*). Zároveň můžeme zvolit možnost zobrazení regresní rovnice, indexu determinace, korelačního koeficientu a p-hodnoty významnosti modelu.





Z následujících výstupů je vidět, jak s rostoucí spolehlivostí vzrůstá i šíře těchto odhadů, tedy i zobrazených pásů.





Případně můžeme změnit typ regresního modelu, např. na kvadratický.



4.3.2 Jednoduchá regresní analýza – lineární model

Grafický výstup sám o sobě není dostačující pro určení modelu. V následujícím textu si tedy ukážeme, jak provést podrobnou regresní analýzu. K tomu vybereme v záložce *Statistiky* možnost *Vícenásobná regrese*. Opět zpracováváme závislost váhy na výšce u sledovaných osob. Jako závislou proměnnou vybereme veličinu Váha, nezávislou pak Výška. Pak v obou dialogových oknech zvolíme možnost *OK*.



Poté nám SW STATISTICA zobrazí následující výstup.

A B P P G	🕸 🖥 =			STATISTICA Cz - [data-celek* - List1]	
Soubor Domů	Upravit Zol	brazit	Formát Stati	stiky Data-mining Grafy Nástroje Data Sešit	
Základní Vícenásobná statistiky regrese	ANOVA Neparam statis	netrické Pro stiky ro	okládání Rozděle ozdělení simula	k. ² . Pokročilé modely - 14 ² . Neuron. sitě	u 📑 ST
	Základ			🖉 Výsledky - vícenásobná regrese: List1 v data-celek	8 - 33
a-celek* <		List1			
List1 잘 2D bodové g		witte	váha	Výsledky- vicerozm, regrese	
- Bodový g	1	192	90	Záv.prom. :váha vicenás. R = ,73896743 F = 129,9236	
🔉 Základní stat 👘	2	155	52	R2= ,54607286 sv = 1,108 Poč. připadů: 110 upravené R2= ,54186983 p = 0,000000	
🖹 🎯 Dialog ko 🚽	3	171	62	Smězodatná chyba odhadu :11,666885017	
Koreli -	4	160	49.5	Abs.clen: -110,7392991 Sm. chyba: 16,22650 t(108) = -6,825 p =	,0000
Vicenasoona	5	169	50	výška b*=, 739	
Analý	6	163	58		
Statist	7	169	58,5		
Výsler	8	168,5	60		
-Prom	9	160	53	(významná b* jsou zvýrazněna červeně)	
Souče	10	167	55		20
Koreli	11	167	54	Afa pro zvýraznění efektů: .05 🖨	OK
Statist	12	166	54	783 de l'actual de la Constante de la Constant	
Výslet	13	178	58	zakadni vyseuky oktarni vyseuky nesiona/predpokady/predpokedi	Stomo
Statist	14	160	52	Wipočet: Výsledky regrese	Možnosti
Výslec	15	169	62		
🔐 Dialog re:	16	175	70	2	Anal.Sku
- Norm	17	187	75		
Predp	18	175	70		
Statist	19	164	51		
vysier	20	169	54		

Výsledné okno obsahuje např. následující důležité informace.

Záv. prom. - obsahuje jméno závislé proměnné – v našem případě Výška.

Vicenás. R - hodnota koeficientu vicerozměrné korelace, což je odmocnina hodnoty R^2 (resp. I^2 , neboli koeficientu determinace).

R2 – hodnota indexu determinace. Jak už jsme v teoretické části psali, tato veličina nám udává, jaký podíl celkové variability závisle proměnné je vysvětleno naším modelem.

Upravené R2 – hodnota upraveného indexu determinace, který bere do úvahy také počet regresorů zahrnutých v modelu.

F – hodnota testového kritéria týkajícího se testu významnosti celého modelu

P - odpovídající p-hodnota, tedy opět slouží k vyhodnocení významnosti celého modelu

Směrodatná chyba odhadu - představuje míru rozptýlení pozorovaných hodnot okolo regresní přímky.

Abs. člen. - obsahuje odhad b₀ (tj. absolutního členu) regresní rovnice.

*Výška b** - koeficient(y) nezávisle proměnné (resp. proměnných). Toto však neodpovídá odhadům parametrů z uvažovaného regresního modelu. Jedná se o speciálně upravené odhady parametrů z jiného modelu, které nám umožňují porovnat relativní vliv jednotlivých regresorů na závisle proměnnou. Statisticky významné regresní koeficienty jsou zvýrazněny červenou barvou.

Pro podrobnější informace zvolíme záložku Detailní výsledky.



Souhrnné výsledky regresní analýzy obdržíme zvolením možnosti Výpočet, výsledky regrese

	9 @ C	御習す					STATISTICA Ca	- [data-celek	* - Výsledky regrese s
Soubor	Domů	Upravit	Zobrazit	Formát	Statistiky	D	ata mining	Grafy Ná	stroje Data :
Základni V statistiky	/icenásob regrese	ná ANOVA N	eparametrické P statistiky r čáklad	rokládání Roz rozdělení si	dêlenî a mulace	Po R Vi	kročilé modely cerozm. anal. + nalýza síly testu Pokročilé/Více	Neuron. PLS, PCA	sité Diagramy ři Multivariati Predictive
😂 data-ce	elek* < 1 2D bod		Výsledky reg R= ,7389674 F(1,108)=12	grese se záv 13 R2= ,5460 9,92 p<0,000	islou pron)7286 Upr)0 Směro	něnno avené d. chy	u : váha (List R2= ,5418698 ba odhadu : 1	1 v data-cele 33 1,667	k)
E De	Boc Základi	N=110	b*	Sm.chyba z b*	b		Sm.chyba z b	t(108)	p-hodn.
Te	Dial	Abs.člen	and the second		-110	,739	16,22650	-6,82459	0.000000
	1.000	výška	0,738967	0,06483	1	.061	0.09309	11,39841	0,000000
6-00	Vicená:								

V této tabulce již obdržíme všechny koeficienty tak, jak očekáváme, jak známe např. z Excelu. Můžeme tedy určit výsledný tvar regresní rovnice: y = -110,739 + 1,061x

Tabulku ANOVA obsahující informace o významnosti modelu obdržíme volbou *ANOVA* (*celk. vhodnost modelu*).



Zde opět vidíme jednak hodnotu testového kritéria a p-hodnotu, na jejichž základě můžeme vyhodnotit významnost modelu. V našem případě vidíme, že p-hodnota je po zaokrouhlení na šest desetinných míst rovna nule (uvědomme si však, že nule rovna ve skutečnosti není), tedy je menší než hladina významnosti (ať už 5% nebo 10%). Znamená to, že námi vytvořený model je významný.

Pro podrobnější vyhodnocení vhodnosti modelu provedeme ještě verifikaci chování reziduí. K tomuto účelu zvolíme záložku *Rezidua/předpoklady/předpovědi*.



Asi nejjednodušším způsobem zjištění, zda se rezidua chovají "rozumně", tedy zda mají normální rozdělení, je vytvoření Normálního p-grafu reziduí. Toto provedeme výběrem možnosti *Reziduální analýza* v zobrazeném dialogovém okně a následně výběrem možnosti *Normální p-graf reziduí* na záložce *Základní*.

A) 🏟 🖥 =				STATISTICA C	z - [data-celek	* - Výsledky r	regrese se závislou pr
Souber Domů	Upravit	Zobrazit	Formát	Statistiky	Data mining	Grafy Ná	stroje Da	sta Sešit
Základní Vicenásob statistiky regrese	oná ANOVA I	Veparametrické statistiky Základ	Prokládání I rozdělení	Rozdělení a simulace	Pokročilé modely Vicerozm. anal. • Ma Analýza síly testu Pokročilé/Vice	 Neuron. PLS, PCA VEPAC verozměrné 	sitë 🔛 Dia 🚟 Mi	sgramy řízení kvality • ultivariate edictive Průmyslová si
Bodon - < Bodon - < Bodon Základní : Dialog Ko Vicenásot Dialog Ar Stu Vý Pr So Ko Ko Stu Vý	N=110 Abs.čl výška	Výsledky r R= 73896 Ø Reziduální a Záv. prom. Poč. připa Sm Abs. člen: Porode Základ Imi Microte	egrese se z Z43 R2= 54 halýza: List1 :váha dů: 110 á:codatná 4 -110,7392 p. grafy Detaly a Reactor or semální p.graf	távislou pror 1607286 Up v data-celek vícená uprave hyba od 2991 Sm. Rezk reziduí	měnnou : váha (Lis ravené R2= 541869 s. R = ,7389674 22 : ,5460728 mé 22 : ,5486928 mé 22 : ,5486928 Cdehlé hodnoty dua Předpověc	11 v data-cele 83 5	k) 25, 9236 1, 108 ,000000 -6, 825 5 02k	2 x ,000 k ,000 Somo Možnost v 2 énal Skup.
□ Oialoc ■ - Control								

Rezidua by měla mít normální rozdělení, což znamená, že v zobrazeném grafu by měla ležet co nejblíže přímky.



V našem případě vidíme, že se body ve spodní části grafu od přímky relativně významně odchylují, tedy normalita není zcela splněna, stav je minimálně hraniční.

Pro podrobnější vyhodnocení ještě můžeme vytvořit histogram reziduí. K tomuto účelu si musíme sloučit soubory, a to jednak původní soubor a dále soubor reziduí. Ten získáme pomocí možnosti *Předpověď závislé proměnné* v záložce *Rezidua/předpoklady/předpovědi*.

Vlastní sloučení pak provedeme přes záložku *Data*, kde v nabídce vybereme možnost *Sloučit*. V následujícím dialogovém okně pak pomocí tlačítek *1. soubor* a *2. soubor* postupně vybereme oba slučované soubory. Na následujícím obrázku je znázorněn výběr souboru s rezidui.

۲۹۲ Přepočítat المنظر المنظ المنظر المنظر	tandardizov řekódovat ořadí hodn	vat 1/11 Pře 11/= Tra ot 🛄 Da	skupování 🕼 Box-Cox nsponovat * 🔻 Filtrovat/překód tum/čas T _A Automatický filtr Transformace	्राष्ट्र ETL + lovat + 🔟 Reportovací tabulky r +	 Třídit Jména Případy * Případy 	Specifikace Sechny specifik Textové popisky Prom	llill Posun cace 🍈 Svazky M Proměnné • těnné	I+I Sloučit	data • % Vzorková 1% Seminko sžina 🔍 Ověřit • Spravovat
Případ	List1 1 výška	2 váha	Možnosti sloučení	omite Pinate X Mai	2 S	a labulku	9	10	11
2 3 4 5 6 7 7 8 9 9 10 11 12 13	160 160 169 163 169 163,5 160 163,5 160 167 167 166 178	52 62 49,5 50 58 58,5 60 53 55 54 54 54 58	Mód Sogit Nantézské Porovnat jm. případů Porovnat proměrné Hil Zivelit Kritéria porovnání Automaticky v	Nesouhl, případy Wyphat ChO Odstrant případy Vygenerovat katézské Přeušt spojování Vicenásobné případy Wyphat ChD Upustt ví Zkopír, dolů Uppustt ví	cenás cenás	Uicedat Dal	Výsledky regrese se z og rezidu i regrese Předpověd a rezidua Statistické shmut i ZP Výsledky regrese (List 1 og výsledků regrese Statistické slymut i ZP výsledků regrese Předpověd a rezidua	ávislou proměr (List1 v data c vsňa (List1 v ávislou proměr data celek) : váha (List1 v ávislou proměr (List1 v data c	OK Storno
14 15 16 17 18 19	160 169 175 187 175 164	52 62 70 75 70 51	1. soubor 2. soubor Podr. po mistria tačitka OK bude stoušeni pode aktivni záloby.	relek rspušséno OK	Z sktu Soubo	uślnich tabulek a sešnů zvo ory' můžete použit pro vý 1,117058 1,444813	ite tabulku, která se má j tér souboru s tabulkou 0,008327 0,680719	-12,4638	0,000838

Následně nám vznikne a zobrazí se soubor, který jednak obsahuje sloupce našeho zpracovávaného souboru a jednak sloupce ze souboru obsahujícího informace výsledcích předpovědí a o reziduích.

Na základě tohoto souboru vytvoříme histogram reziduí. K tomuto účelu v záložce *Grafy* vybereme v nabídce možnost *Histogram*. V následném dialogovém okně pak vybereme proměnnou pro histogram. Touto proměnnou bude veličina Reziduum. Dále si vybereme *Typ* proložení Normální a můžeme též zaškrtnout možnost *Shapiro-Wilkův test* pro vyhodnocení normality reziduí.



Jak z grafického zobrazení, tak z p-hodnoty Shapiro-Wilkova testu je vidět, že rezidua v našem případě normální rozdělení nemají, což není na závadu při odhadování regresních koeficientů a můžeme tedy říci, že regresní rovnice je v pořádku (významnost modelu byla potvrzena). Nemůžeme se však zcela spolehnout na významnost regresních koeficientů a především na správnost intervalů spolehlivosti. Toto bychom se měli snažit odstranit, např. vhodnou transformací dat, nebo zjištěním a odstraněním odlehlých pozorování. Ale těchto možných příčin a tím i řešení je mnoho.



4.3.3 Jednoduchá regresní analýza – obecný model

Ukázali jsme si jeden ze způsobů, jak je v SW STATISTICA vytvořit a vyhodnotit lineární regresní model vyjadřující vztah mezi dvěma veličinami. Jak už jsme v úvodní teoretické části tak i v části o zpracování pomocí Excelu, ne vždy je však lineární model jediný možný a nejlepší.

Ukážeme si nyní, jak bychom v SW STATISTICA vytvoříme takovýto model. Postup si ukážeme například na kvadratickém modelu. Grafický způsob jsme si již ukázali, nyní provedeme detailní regresní analýzu.

Opět se vrátíme k listu obsahujícímu data o všech studentech bez rozdílu pohlaví. V záložce *Statistiky* vybereme možnost *Pokročilé modely* a v otevřeném seznamu pak vybereme *Nelineární regrese – jednoduchá*. V zobrazeném dialogovém nastavíme proměnné regresního modelu.



V obou oknech pak zadávání potvrdíme stisknutím klávesy *OK*, následně se otevře okno *Nelineární komponenty regrese*.

区 タ マ C 会 習 マ Soubor Domů Upravit Zobrazit Fo	mát Statisti	ky Da	ta minir	ST/ ng Grafy Nástroj	ATISTICA Cz - je Data	(data-celek_1_5 - Lis Sešit	tl]	
Základní Vicenásobná ANOVA Neparametrické Prokli statistiky Základ	dání Rozdělení a lení simulace	A Pol	kročilé r erozm. alýza síl Pokro	modely • 😹 Neuron. sítě anal. • 🛞 PLS, PCA, ly testu 🔲 VEPAC cilé/Vicerozměrné	Diagram Multiva	ny řízení kvality • 満 riate 🛱 ve 60 Průmyslová statist	Analýza procesu DOE Six Sigma *	STATISTICA VB Dávk. analýza (dle si Kalkulátory • Nástro
ata-celek 1.5	L	ist1	_					
D bodové grafy (List1 v data-celek)	v	rýška	váha	Nelineární kompone	nty regrese: Li	st1 v data-celek 1 5	2	×
Zakladni statistiky a tabulky (List1 v	1	193			1-3		A secondary	
🕀 📴 Vicenásobná regrese (List1 v data-ce 🦳	2	160	_	Základní nastavení			. 🖂 🛛	ĸ
🖃 🎯 2D bodové grafy (List1 v data-celek) 👘 📖	3	171	-	Nelneámi transformace		Možný rozsah pův. ho	dnot Storm	
– Bodový graf z váha proti výška	4	160	- 1	Was da	(undur	5 0E+08 +15 0E	-08	
Bodový graf z váha proti výška	5	163		Vin di	-+ 0	E OC . OE . I E OC	Možno	sti ▼
Bodový graf z váha proti výška	7	169			ecij	-5.0E+05 az 5.0E	+05	
- Bodový graf z váha proti výška	8	168.5	-1	A 4 (Ana ct	(vitou)	-5.UE+04 az 5.UE	+04	
🖉 Bodový graf z váha proti výška	9	160		X 5 (X na pa	stou)	-0.UE+03 az 0.UE	+03	
	10	167		SQRT(X) (odmocr	nina)	X >= 0		
	11	167		LN(X) (přirozer	ný logaritmus)	X > 0		
	12	166		ELOG(X) (desitko	ový logantmus)	X>0		
	13	178		⊡e^X		-40 až +40		
	14	160		10°X		-18 až +18		
	15	169		1/X		X se nerovná 0		
	16	175						
	17	187					_	
	18	175		70				

V tomto okně volíme z nabídky nelineárních transformací. V našem případě je cílem vytvořit kvadratický model, proto zvolíme transformaci X^2 .

Kdybychom chtěli vytvořit polynomický model 3. stupně, zvolili bychom X^3, v případě, že bychom tímto způsobem tvořili lineární model, ignorujeme transformace a stiskneme OK ihned.

Následně se nám zobrazí dialogové okno *Definice modelu*. Nastává jedna z nejobtížnějších a nejméně přehledných, přitom velmi důležitých etap naší práce, a to výběr proměnných.

A B S CO	🗟 📅 Ŧ				STATIS	STICA Cz - [Data: Tabulka27 (4s krát 8i)]				
Soubor Domú	Upravit Zobr	azit Formát	Statistiky	Data mining	Grafy Nástroje	e Data				R Feature Finder	Mo
Základní Vícenásobní statistiky regrese	ANOVA Neparamet statistii Základ	trické Prokládán ky rozdělen	i Rozdělení a i simulace	Rokročilé modely R Vicerozm. anal. * Analýza sily testu Pokročilé/Vie	 Neuron. sitě PLS, PCA, VEPAC verozměrné 	Diagramy řízení kvality • E Multivariate Predictive Průmyslová st	Analýza procesu C DOE G Six Sigma * tatistika	STATISTICA VB Dávk. analýza (dle skupin) Kalkulátory * Nástroje	Statistiky bloku dat		
[List1 v data-cele	k_1_5									
	1 výška	2 váha	3 V1**2	4 V2**2							
výška váha V1**2 Průměry Smodch. Peč. přp Matice	1,00000 0,73897 0,99338 0,73311 173,90455 12,00468 110,00000 1,00000	0,73897 1,00000 0,74375 0,99410 73,78182 17,23695	0.9993 0.7437 1.0000 0.7386 30385,5931 4208,5076	8 0,73311 5 0,99410 0 0,73861 1 1,00000 8 5738,16818 5 2688,64742	Definice mo Základ Detai Demi Základ Detai Domi Základ Detai Domi Základ Domi Základ Nezávlád: Nezávlád: Nezávlád: Nezávlád: Tolerance: O Hébenová Zpracován Tale/protok	delu: List1 v data-celek,1.5 y Metoda Popuné statistky nné šádné šádné 8. efekty ▼ ahmut v modelu ▼ 001 ∰ (Premed: 0.0 rassavím Premere: 10-29) regrese: lambds: .10 ∰ //tak po dévlách ol resolution i anályzy	OK OK				

Jakmile stiskneme tlačítko *Proměnné*, objeví se nabídka všech variant výběru závisle i nezávisle proměnných.

Nevýhodou je, že transformované proměnné jsou označeny pořadím (např. V1**2 či V2**2), nikoli jménem (např. výška**2 nebo váha**2). Musíme být tedy velmi opatrní. V našem případě požadujeme, aby závislou proměnnou v modelu byla proměnná Váha a do pozice nezávislých proměnných se dostanou veličiny Výška a druhá mocnina výšky, tedy proměnná V1**2.

Můžeme si uvědomit paralelu s maticí F z maticového vzorce pro výpočet regresních koeficientů (viz teoretický úvod) či s přidávanými sloupci při tvorbě kvadratického modelu pomocí doplňku Analýza dat v Excelu (viz kapitola o zpracování regrese v Excelu).



Výběr proměnných potvrdíme stisknutím OK. Vrátíme se do okna Definice modelu.

Soubor	o (⇔ C		Zobraził	Format	Stati	ictiley	De	ta minino	STATISTICA Cz - [data-celek_1_5 - List1]	
Základní statistiky	Vicenásobní regrese	ANOVA	Neparametrické statistiky Základ	Prokládání rozdělení	Rozděle simula	ní a ce	Pol Vic	kročilé mode erozm. anal. alýza síły tes Pokročilé.^	(* 14 Neuron, sité ≧ Diagramy říčení kivality * 12 Analýza procesu 3 N NS, PCA, ≧ Multivariate ∰ DOE 10 VEPAC ≧ Predictive € St Sigma * Průmyslová statistika Nástroje Nástroje	tatistiky oku dat •
😂 data-c	elek 1.5			<		List	L .			
0-10 Ls	2D bodové	grafy (Lis	t1 v data-celek)			výšk	a	váha		
@ 📮	Základní st	atistiky a t	abulky (List1 v			1	193	90	C Definice modelu: List1 v data-celek 1.5	
@ 📮	Vicenásobr	ná regrese	(List1 v data-ci			2	160	52		
- -	Vicenasobr	nă regrese	(List1 v data-ci			3	171	62	Základ Detaly Metoda Popisné statistiky 🖂 OK	
	20 Bodove	graty (Lis	a protivoilka			4	160	49,5		
	Bodovi	graf z vál	na proti výška			5	169	50	Proměnné Stomo	
	Bodový	graf z vál	na proti výška		(6	163	58	Závislé: váha 🔊 Možnosti 🔹	
-	Bodový	graf z vál	na proti výška		1	7	169	58,5	Nezávidé: wilka V1**2	
	Bodový	graf z vál	na proti výška		1	B 1	168,5	60	Metoda Ma. efekty	
-	Bodový	graf z vál	na proti výška		9	9	160	53		
0-4	Jednoduch	a nelinear	ni regrese (List		10	D	167	55	gos. cien: Zahmut v modelu 🔹	
5	Prů	méry a sm	věrodatné odch		1	1	167	54	Tolerance: ,0001 (Pomoci 0.0 nastavite	
	Kor	elace (List	1 v data-celek	1	12	2	166	54	Withheast of managers lambda: 10	
	Reality				13	3	178	58	Debenova regrese, tambua. 10	
					14	4	160	52	Zpracováni/tisk po dávkách	
					1	5	169	62	Tisk/protokol reziduální analýzy	
				-	16	5	175	70		
					1	7	187	75		

Zde ještě zkontrolujeme, zda je nastavena metoda Vš. efekty a zda je absolutní člen zahrnut v modelu. Poté provedeme výpočet stiskem OK. Otevře se okno Výsledky – vícenásobná regrese. V záložce Základní výsledky zvolíme tlačítko Výpočet: Výsledky regrese.

Soubor Domú Upravit Zobrazit	Formát Statistiky Data míning	STATISTICA C2- (data-celek,1,5-List1) ng Grafy Nástroje Data Selit Resture Fr
Základní Vicenásobná ANOVA Neparametrické Pr statistiky regrese	rokládání Rozdělení a ozdělení simulace	nodely* 34 Neuron. sité 📄 Diagramy fizení tvality* 🚡 Analýza procesu 📮 STATISTICA VB anal. * 🛞 PLS, PCA, 🔚 Multivariate 🚱 DOE ly textu 🛄 VEPAC 😨 Predictive 😚 Six Sigma * 🗮 Kalkulátory * bloku dat *
Zálisd deta-celek 1,5 deta-celek 1,5 deta-c	Uist1 výška váha 1 193 9 2 160 5 3 171 6 4 160 49, 5 169 5 6 163 5 7 169 58, 8 168,5 6 9 100 55 10 167 5 12 166 5 13 178 3 14 160 5 16 175 7 17 187 7 18 175 7 19 164 5 20 169 5	Gall Výsledky-vicenšsobná regrese List v data-celek_1.5 90 Výsledky-vicenčsobná regrese 24v.prom. :váha vicenás. 8 = .76374123 F = 70,37668 252 poč. případů: 110 upravené R256012464 av = 2,100 90 sko-član: 115,9767699 Bn. chyla av = 2,100 90 vicenás. 8 = .76374123 F = 70,37668 av = 2,100 90 poč. případů: 110 upravené R256005042 p = 0,0654 90 adadaticí 11,143205646 av = 107) = 1,8411 p = .0654 90 výška b*=-3,5 V1==2 b==4,21 av = .0654 91 výška b*=-3,5 V1==2 b==4,21 av = .0654 92 Výsledy Datahí výsložy Besdus/předovědí Bo # av = .0654 93 (významná b* jsou zvýzazněna červeně) By # av = .0654 94 Ma por nýnamění efekti: 05 By # av = .0654 95 Vysočet: Wiskdy mgme av = .0654 av = .0654 96 José Justický v Datahí výsložy Besdus/předovědí av = .0654 av = .0654 97 José Justický v Datahí výsložy Besdus/předovědí Justický Justickýsloží Besdus/předovědí Justický Justický J

Zobrazí se následující tabulka. V její horní části je řada užitečných informací. Pro nás jsou důležité hodnoty R (korelační koeficient) a R2 (index determinace), podle toho, zda pracujeme s lineárním či nelineárním modelem.

	Výsledky regrese se závislou proměnnou : váha (List1 v data-celek_1_5) R= ,7537412 ,56812584 Upravené R2= ,56005342 F(2,107)=70,379 p										
N=110	b*	Sm.chyba z b*	b	Sm.chyba z b	t(107)	p-hodn.					
Abs.člen			415,5727	225,7234	1,84107	0,068382					
výška	-3,46534	1,799775	-4,9757	2,5842	-1,92543	0,056830					
V1**2	4,20693	1,799775	0,0172	0,0074	2,33747	0,021275					

Dále si v této tabulce budeme především všímat sloupce b, který obsahuje jednotlivé regresní koeficienty. V prvním sloupci poznáme příslušnost jednotlivých koeficientů. Náš model má tedy tvar:

 $y = 415,5727 - 4,9757x + 0,0172x^2$ s indexem determinace $I^2 = 0,5681$.

V posledním sloupci jsou uvedeny p-hodnoty významnosti jednotlivých koeficientů. Nejdůležitější je poslední uvedená hodnota náležející ke kvadratickému členu. Kdyby tento člen nebyl významný, nejednalo by se totiž o kvadratický model. V našem případě vidíme, že p-hodnota je 0,02, tedy koeficient významný je. Zbývající dvě p-hodnoty se rovnají 0,068, resp. 0,057, tedy příslušné koeficienty významné nejsou. Pro nás z toho plyne, že model lze zjednodušit tím, že vypustíme absolutní, resp. lineární člen.

Vsuvka:

To bychom provedli návratem do lišty *Výsledky-vícerozměrné*. Ta je zobrazena v levé spodní části obrazovky. Objeví se opět okno *Výsledky vícerozměrné regrese: Tabulka*, zde stiskneme tlačítko *Storno*. Vrátíme se tak do okna *Definice modelu*.

Vyloučení lineárního členu bychom provedli v tomto okně změnou výběru proměnných vstupujících do modelu, a to tak, že bychom odebrali nezávislou proměnnou Výška a ponechali pouze nezávislou proměnnou V1**2.

Vyloučení absolutního členu bychom provedli v okně *Definice modelu*, kde v záložce *Detaily* rozbalíme položku *Abs. člen* a vybereme nabídku *Nastaven na 0.* Klikneme na *OK*.

Provede se nový výpočet. Před tím však budeme upozorněni, že nelze srovnávat R^2 původní výstupní sestavy s hodnotou R^2 v sestavě zjednodušené. Je totiž počítán podle jiného vzorce. Nová výstupní sestava již nebude mít absolutní člen.

Tento krok však není nezbytně nutný.

Predikce

Predikci umožní provést nastavení záložky Residua/předpoklady/předpovědi v okně Výsledky –vícerozměrná regrese.

Marine Count Tobush	STATISTICA	Cz - [data-celek_1_5" - Vysledky regrese se závislou proměnnou : vaha (List1 v data-celek_1_5)]	S Feature Funder	Mahaati" B
Zastadni Vicenskobná ANOVA Reparametrické Pr Italistky regrese Zástad	vokládání Rozdělení a codělení simulace	dolly* > M Heurisn, ikk □ Diagnamy Rizeni kvality.* □ Analyiza piocesu □ STATISTICA V6 □ Inditusnivate □ DOL □ Disk. analyiza idle skupir Statistic □ Bredictive □ Bredicti	y.	
y data-celek 1,3 € intl 2 bodové grafy (List1 v data-celek) 2 Likádní statistiky a tabuliky (List1 v 3 Vicenšsobná regrese (List1 v data-ce 2 bodové grafy tjužt v data-celek) - 2 bodové grafy tjužt v data-celek) - 20 bodové grafy tjužt v data-celek)	Výsledky regrese se i R= 75374123 R2=,5 F(2,107)=70.379 p-0. b* Sm.chy b* Sm.chy Abs.člen y3ka -3.46534 1.79 V1**2 4.20693	Typesdy-vicencome segress that references Typesdy-vicencome segress Start Pytickly-vicencome segress Start Pytickly Segress Start Pytickly Segress Start Pytickly Segress Start Pytickly Segress Start Pytickly Segress Start Pytickly Segress Start Pytickly		
Bodový graf z váha proti výška Jododuchá nelineární regrese (List) Díalog popiných státistik regrese		výdža bra-2,5 V2+2 bras,21 (významá b* 3500 zvýznamána červená) <u>Naj</u>		
Průměry směrodane odch Romaec (tist) data-cekt.) Dialog výsteláň regrese Dialog výsteláň regrese Ustatické shrouž(ZP: vála () Výsteláky regrese se závislou j		Alls pro znýcazněl néteků. 25 gr Zákudní výskulty Detakin (výskulty <u>Pendus předscháda (předscháda (předscháda)</u> Bestudní máljas <u>Poprat attatiky</u> V poprat zákudá poseknet a V poprat rázen pogravos v Co		
· /	i * Statatické strmati 2P. válha (Lat 1 v d	sta colek, 1,9, 🔝 Vylody regrese se závsku proněvnou : víhra (Jat 1 v dzta c		

V zobrazeném okně vyplníme hodnotu Výšky, pro kterou nás zajímá předpověď Váhy. Nevýhodou je, že musíme vyplnit i druhou mocninu této hodnoty.

Soubor Domú Upravit Zobrazit	STATISTICA Cz - Formát Statistiky Data mining	[data-celek_1_5* - Výsledky regrese se závislou proměnnou : váha (Listl v data-celek_1_5)] Grafy Nástroje Data Sešit	C Feature Finder Možno
Soubor Domi Upret Zobract Základní Vicenásobná ANOVA Neparametricke F ztatistily regrese Vicenásobná Parovice Vicenásobná regrese (Listl v data-celek) Základní statistky a tabuláy (Listl v data-ce	Format Statistiky botkidaini Rozdeleni rozdeleni simulare Výsledky regrese se záki R=,75374123 R2=,5681 Výsledky regrese se záki R=,75374123 R2=,5681 F2(2107)=0.2787 pc.000 b* Sm.chyba Z b* Abs.Glen 2 b* 1 79977 V1*2 1 79977	Cardy Visited Date Selit Y* Neurons.ht Diagramy filterik kvality* Analyza procesu Y* Neurons.ht Diagramy filterik kvality* Analyza procesu Y* Neurons.ht Diagramy filterik kvality* Analyza procesu Y* Neurons.ht Diagramy filterik kvality* Date Y* Neurons.ht Diagramy filterik kvality* Date Y* Neurons.ht Diagramy filterik kvality* Date Y* Vicends.prom. Statistica Transferation Y* Vysledky- vicendscohnå regrese List vicends.prom. Y kalkuistory* Vysledky- vicendscohnå regrese Záv.prom.: velna vicends.pr 5603542 P = 0,0000 Abs. Slen: 110 uppravend 22	Peature Finder Modno kupini Statistiky bloku dat 0684
 Jednoduchá nelineární regrese (List) Dialog popinyich statistik regres Průměty a směrodatné odch Visladeka (List) v dsta-celek (Jist) Dialog výsledků regrese Statistické shrmuti ZP: váha (List) Výsledky regrese se závislou j 		(významá b* 3# Soděčná hod. Ala pro zvýsavání réda 0 Zákdaňi výsledky Progověd Bezdušíní snalýza Předpověd Popaně statistky Výpočet interv. předpověd Výpočet interv. předpověd 0	Ball + CK Stomo IMožnost • jinal Stop

Výsledkem je následující tabulka:

	Předpovězené hodnoty (List	Předpovězené hodnoty (List1 v data-celek_1_5) proměnné: váha									
Proměnná	b-váha	Hodnota	b-váha * Hodnot								
výška	-4,97572	172,00	-855,823								
V1**2	0,01723	29584,00	509,698								
Abs. člen			415,573								
Předpověď			<mark>69,448</mark>								
-95,0%LS			<mark>66,509</mark>								
+95,0%LS			<mark>72,387</mark>								

V této tabulce jsou pro nás nejdůležitější tři žlutá pole. Vidíme, že student/ka měřící 172 cm by průměrně měl/a vážit 69,45 kg. 95% intervalovým odhadem je rozmezí od 66,51 do 72,39 kg.

Grafické znázornění nalezeného regresního modelu

Známe-li předpis pro regresní model, můžeme si model graficky zobrazit. V záložce *Grafy*, zvolíme nabídku *2D grafy* a v něm nabídku *Grafy vlastních funkcí*.



V okně, které se otevře, zvolíme *Rozsah X* (tedy rozsah naší závisle proměnné), tj. nastavíme políčka *Min.*, *Max.* a předepíšeme tvar funkce. Zadání potvrdíme stisknutím *OK*.

	unit flatiat	Data mil	STATISTICA Cz - [data-celek,1_5* - List1]		Q Feature Finder Molace
Histogram Bodový Průměry Krabice Rozphyl Spojnice graf Běžný	2D • 3D Sekv. • 3D XYZ •	Matice	Image: Strate Strat		
data-celek 1 5*		Let 1	2D grafy uživatelských funkcí		
List1 20 bodové grafy (List1 v data-celek) 27 Zikladní statiritky a takulky (List1 v		výška váha	Vastni funkce Vzhled Možnosti 1 Možnosti 2	ОК	
Vicenásobná regrese (List) v data-ce	1	193	Nazev: Nova funkce Pňdat do seznamu	Stomo	
💮 🎇 Vícenásobná regrese (List1 v data-ce 🦳	2	160	DP Seznam funkcí	🔉 Možnosti 💌	
👜 🌆 2D bodové grafy (List1 v data-celek) 📃	3	171	ГЛЯКСЕ		
- Bodový graf z váha proti výška 📃	4	160	Bozsah X	Anal. skup.	
Bodový graf z váha proti výška	5	169	Min: 150 🖨 Max: 200 🖨	Baset Filtr případů	
Bodový graf z váha proti výška	6	163	V Prývodce funkcemi		
Bodový graf z vana proti výska	7	169	Fynkce:	C Vany	
Bodový graf z váha proti výška	8	168,5	Y = 415.5727-4.9757%+0.0172%*2	Galerie grafú	
E Jednoduchá nelineární regrese (Listi	9	160		Mitualizace: Auto	
😑 🎒 Dialog popisných statistik regres 🚽	10	16/		(manater res	
Průměry a směrodatné odch	11	167			
Korelace (List1 v data-celek_1	12	166			
📄 🎯 Jednoduchá nelineární regrese (List1 🔛	13	1/8	okné, pokud je typ funkce nastaven na parametrickou křivku).		
Dialog výsledků regrese	14	160	Pfiklady x**2		
Statisticke shrnuti;ZP: vaha (I	15	169	(Abs(x - 1) - 1) / CosH(x)		
Předpovězené hodnoty (List)	10	1/5	ault v) + coals x vi b - (1,X)		
E 20 grafy uživatelských funkcí (List)	1/	187			
Graf funkce	18	1/3			
- The Graf funkce	19	169			
- Graf funkce	20	109			

Následně se nám zobrazí graf, který můžeme upravovat dle vlastních potřeb (viz kapitola o grafech)



4.3.4 Vícenásobná regresní analýza pomocí SW STATISTICA

Nyní si ukážeme zpracování vícenásobné regresní analýzy v SW STATISTICA. V této části budeme zkoumat vliv různých faktorů (věk, výška, množství spánku, množství času u TV a PC, ...) na váhu u mladých mužů, přesněji řečeno u studentů středních škol.

Jak jsme si již v teoretické části i části zpracování pomocí Excelu řekli, nejdříve bychom měli zkontrolovat, zda jednotlivé faktory nejsou mezi sebou silně korelované. Tento fakt by mohl narušit kvalitu výsledného regresního modelu a analýzu síly vlivu jednotlivých faktorů.

K tomuto účelu zvolíme vytvoření korelační matice obsahující korelační koeficienty mezi jednotlivými faktory, a to tak, že v záložce *Statistiky* vybereme nabídku *Základní statistiky* a zde ze seznamu vybereme možnost *Korelační matice*. Pak už jen svůj výběr potvrdíme tlačítkem *OK*.

oor Domú l	Ipravit Zob	azit F	ormát	Statistiky	Data	mining	Grafy	Nástro	ije Data	Sešit				Q
dní Vicensobná Al tiky grese	IOVA Neparame statisti Zāklad	trické Prok ky roz	Sádání Ro dělení s	zdělení a imulace	Pokro	čilé model zm. anal. za síly test okročilé/V	y • 📢 No St PL 2 🔃 VE icerozmén	ruron. siti S, PCA, IPAC në	H Diagr Multi	amy fizeni kvi variate ctive Průmysle	ality • 🕋 Ar C D G S ová statistiku	nalýza procesu DE « Sigma *	STATISTICA VB	Statistiky bloku dat •
ata-muzi* <	Li	st1												
List1	vi	k v	ýška l	boty	prsa	pas	boky	krk	sport	spänek	pc+tel	váha		
Dial	1	18	193	46	108	9	7	96	38	6	8	4 90		
	2	20	175	43	96	84	5	91	37	1	6	7 70		
100	3	20	178	43	110	8	5	90	46	7	8	3 77		
	4	20	Zák	ladni stati	stiky a tabs	ulky: List1	v data	8 8		6	9	11 70		
	5	20	704	And a data of	1	and the second				6	9	11 87		
	6	20	2808	Cal Provide statistics			1 8	OK		21	7	2 85		
	7	19					•	0	5	2	9	11 96		
-01-	8	20		Karelaáni	natice		7	Storno		2	10	9 102		
	9	22	1	t-test, neza	Wale, die sk	upin	1 5	Možnosti		1	8	13 85		
🛛 🥁 Základr 📃 🔄	10	22		Hest, nezá	vislé, die pro	oměn.	C. C	Construction of the local sector		1	8	12 86		
🖂 🛶 Dial	11	21		Hest, závis	lé vzorky					0	7	12 102		
	12	20	8	Hest, same	ost. vzorek					0	7	9 110		
	13	21	25	Rozklad &	iednofakt. A	NOVA				1	7	7 100		
	14	22	推	Rozklad						0	8	11 100		
	15	21		Tabulicy de	tnosti					1	8	11 110		
	16	22	1000	Knotinneni	In I take dev					0	7	11 113		
	17	22	1221	Tabulley wi	centiach or	himme				0	8	11 95		
	18	22	100				-		-	3	-	4 75		
	19	20	2	resty rozd	iu.r, %, pror	nery		Otevn dat	0	4	0	3 85		
	20	21	-	mavdépod	obnostni ka	skulator	Dintr.	A		10	-	3 90		
	21	22					Contract of the local division of the local			3	78	2 901		

V následujícím dialogovém okně pak vybereme seznam proměnných, mezi kterými nás zajímá vzájemná korelovanost.



Po potvrzení našeho výběru pak v původním okně vybereme možnost *Výpočet*. Obdržíme korelační matici v následujícím tvaru. Jak vidíme, tato matice je obohacena o dva sloupce, které obsahují průměr a směrodatnou odchylku dané veličiny.

Korelace (Korelace (List1 v data-muzi) Označ. korelace jsou významné na hlad. p < ,05000 N=44												
	Průměry	Sm. odch.	věk	výška	boty	prsa	pas	boky	krk	sport	spánek	pc+tel	váha
věk	20,34	1,57	1,00	-0,15	-0,17	-0,03	-0,15	0,10	0,01	0,06	-0,09	-0,17	-0,21
výška	185,55	7,57	-0,15	1,00	0,82	0,54	0,63	0,42	0,32	0,19	0,10	-0,17	0,52
boty	44,07	1,61	-0,17	0,82	1,00	0,59	0,61	0,42	0,37	0,00	0,11	-0,06	0,59
prsa	103,52	8,16	-0,03	0,54	0,59	1,00	0,73	0,78	0,56	-0,34	0,22	0,20	0,71
pas	94,93	7,26	-0,15	0,63	0,61	0,73	1,00	0,78	0,56	-0,32	0,06	0,26	0,82
boky	106,05	11,20	0,10	0,42	0,42	0,78	0,78	1,00	0,53	-0,43	0,24	0,36	0,73
krk	40,66	2,28	0,01	0,32	0,37	0,56	0,56	0,53	1,00	-0,44	0,12	0,23	0,60
sport	4,64	4,79	0,06	0,19	0,00	-0,34	-0,32	-0,43	-0,44	1,00	-0,01	-0,75	-0,45
spánek	7,70	0,88	-0,09	0,10	0,11	0,22	0,06	0,24	0,12	-0,01	1,00	0,21	0,12
pc+tel	7,50	3,48	-0,17	-0,17	-0,06	0,20	0,26	0,36	0,23	-0,75	0,21	1,00	0,48
váha	90,30	11,43	-0,21	0,52	0,59	0,71	0,82	0,73	0,60	-0,45	0,12	0,48	1,00

V zobrazené tabulce jsou statisticky významné korelace označeny červeně. Poslední řádek a sloupec obsahuje údaje týkající se veličiny Váha, což je naše vysvětlovaná proměnná, nikoliv regresor, proto nás v tomto kroku příliš nezajímá. Pokud bychom chtěli mít analýzu a vývody "dokonale přesné", měli bychom si všímat všech červených hodnot. Většina učebnic však uvádí kritérium, že pro násobnou regresi jsou problémové hodnoty v absolutní hodnotě nad 0,8. Vidíme, že takovouto hodnotu má pouze korelační koeficient mezi veličinami Boty a Výška. Ze znalosti problematiky asi většina lidí usoudí, že spíše závisí velikost bot na výšce, než obráceně, proto by asi bylo nejlepší veličinu Boty z další analýzy vypustit.

Jednou z nejjednodušších ale i nejméně přesných možností vyhodnocení vhodnosti použití násobné regrese je grafické zobrazení vzájemných vztahů jednotlivých veličin.

Na diagonále grafické matice vidíme histogramy jednotlivých veličin. I nezkušeným okem můžeme vidět, že se většina z nich normálním rozdělením příliš neřídí. V ostatních políčkách matice vidíme bodové grafy popisující vztahy daných veličin s proloženými regresními přímkami. Zde vidíme (tak, jak už jsme přesněji viděli i v korelační matici), že na většině obrázků body příliš přímku nekopírují, že tedy korelace příliš silné nejsou (připomeňme, že si nemáme všímat posledního sloupce a řádku náležícího veličině Váha). Naopak, čeho si v posledním řádku můžeme všimnout, že nejsou zřejmé jiné než lineární závislosti (možná až na jedinou výjimku – viz osmý obrázek v posledním řádku), což by také mohlo být překážkou bezproblémového použití vícenásobné regrese.

Takže vidíme, že nejzásadnějším problémem je chování veličin odlišné od normálního rozdělení. Nejvýraznější je tento fakt v osmém řádku, kde je histogram silně asymetrický. Osmá veličina, což je veličina vyjadřující počet hodin strávených sportem, je tedy problematická ze dvou důvodů. To však neznamená, že násobnou regresi nemůžeme použít vůbec. Jednou z možností je veličinu Sport vyloučit ze zpracování, což by nám ale možná bylo líto, protože v povědomí lidí je, že množstvím sportu svou váhu ovlivňujeme. Proto ji v analýze ponecháme, ale budeme při případném vyhodnocení velmi opatrní.



Fakt, že jednotlivé veličiny nemají příliš normální rozdělení, je někdy považován za problém při vyhodnocování korelace pomocí Pearsonova korelačního koeficientu, což jsme právě prováděli. Proto ještě raději vyhodnotíme korelaci pomocí Spearmanova koeficientu pořadové korelace. K tomu účelu zvolíme v záložce *Statistiky* možnost *Neparametrické statistiky*. V zobrazeném dialogovém okně pak opět vybereme všechny veličiny, mezi nimiž chceme korelovanost vyšetřovat.

0 9 0 6 1 1	٠ 19					STA	TISTICA	Cz - [d	ata-muzi	" - List1)		
Soubor Domú	Upravit	Zobrazit Fo	rmat Statist	iky D ta mi	ning Grafy	Nästroje	Dat	a 5	elit			
Základni Vicenásobr statistiky regrese	as AlovA Nepar	rametrické Proje atistiky Proje	édání Rozdělení člení simulace	Pokročil Vicerozn Anatýza Pok	é modely = 😹 Neuro n. anal. = 🛞 PLS, P sily testu 🔛 VEPA ročité/Vicerozmérné	n. sitë CA,	Diag Mult	pramy říz tivariate lictive P	růmyslov	ty * Anal C DOE EG Six S à statistika	ýza procesu igma •	STATISTICA VB Dávk. analýza (dle sku Kalkulátory * Nástroje
😂 data-muzi* 🔺 <		List1										
E List1		vēk vý	Neparam	etrická korelace	: List1 v data-muzi	8	22	sp	ánek	pc+tel	váha	
e 🛶 🚦		1 18	Promě	mé	8	Speam	an B	6	8	3 4	90	
		2 20	Samt- Ula		-	1			6	5 7	70	
		3 20	Joebin. Vie		1.0	Stomo	20000	7	\$	3 3	77	
		4 20	1			Mažnos		6	5	9 11	70	
		5 20	Vytvořt:	Overcová matice	A Vuberte promi	noé na a	makimu	1.00	-	2 8	87	
		6 20	Záki, výsled	ky Další výsled	The Typerte prome	nore no a	manyzu			10704 10 000	85	
1		7 19		1.000	L - ves	-				OV.	96	
		8 20	Spec	amanúv koef. R	2 - vyška						102	
		9 22			3 - boby	3 - boty 4 - pros				Storno	85	
😑 🍑 Záki	1	0 22			5-045				500	[Svazky]	86	
6 🔤 🕻	1	1 21			5 - boky 7 - let				13		102	
-1	1	2 20			8 - sport				Pro		110	
-1	1	3 21	HE Mat	ce bodových grafi	9 - spánek				odp	evidajcich	100	
-1	1	4 22			11 - vaha				IVO	te Pouze	100	
- 4	1	5 21	Lare.		1.000				odp Pag	ovidajici vice	110	
- 4	1	6 22	190	45 123	Vubratude	Roztábno		PRACE	info	maci	113	
-9	1	7 22	176	42 100	(Terretory)			1 Aurent		KTIGE F1.	95	
- 4	1	8 22	175	43 99	Vyberte proměnné	£			_		75	
- 2	1	9 20	186	44 100	1-11						85	
-	2	0 21	199	46 108	Poure adaptit	dici momè	lond				90	
-9	2	1 22	198	45 110		the province	2005	_			90	
-9	2	2 24	185	44 105	89 102	4	0	7		2	79	
	2	3 24	189	45 107	95 110	4	1	7	\$	2 2	78	

Spearma	Spearmanovy korelace (List1 v data-muzi) Označ. korelace jsou významné na hl. p <,05000												
	věk	výška	boty	prsa	pas	boky	krk	sport	spánek	pc+tel	váha		
věk	1,00	-0,17	-0,18	-0,06	-0,09	0,19	0,04	-0,02	-0,11	-0,07	-0,16		
výška	-0,17	1,00	0,81	0,48	0,59	0,37	0,35	0,21	0,09	-0,25	0,47		
boty	-0,18	0,81	1,00	0,56	0,60	0,41	0,42	0,03	0,10	-0,11	0,56		
prsa	-0,06	0,48	0,56	1,00	0,72	0,75	0,64	-0,38	0,25	0,15	0,73		
pas	-0,09	0,59	0,60	0,72	1,00	0,79	0,73	-0,41	0,05	0,19	0,80		
boky	0,19	0,37	0,41	0,75	0,79	1,00	0,64	-0,53	0,17	0,35	0,75		
krk	0,04	0,35	0,42	0,64	0,73	0,64	1,00	-0,43	0,14	0,20	0,66		
sport	-0,02	0,21	0,03	-0,38	-0,41	-0,53	-0,43	1,00	0,07	-0,77	-0,60		
spánek	-0,11	0,09	0,10	0,25	0,05	0,17	0,14	0,07	1,00	0,22	0,14		
pc+tel	-0,07	-0,25	-0,11	0,15	0,19	0,35	0,20	-0,77	0,22	1,00	0,48		
váha	-0,16	0,47	0,56	0,73	0,80	0,75	0,66	-0,60	0,14	0,48	1,00		

Po potvrzení se nám zobrazí další tabulka.

Vidíme, že hodnoty obou druhů korelačních koeficientů vyšly obdobně, hodnoty Spearmanovy korelace vyšly poněkud nižší, ale nijak výrazně.

Nyní již tedy můžeme přistoupit k vlastní regresní analýze. Nejdříve provedeme Vícenásobnou regresi, a to metodu Enter. Postupujeme obdobně, jako už jsme postupovali při jednoduché lineární regresi. V záložce *Statistiky* vybereme možnost *Vícenásobná regrese*. Rozdíl však nastane v následujícím kroku, při výběru proměnných vstupujících do analýzy. Za závislou proměnnou vybereme opět veličinu Váha, ale za nezávislé proměnné vybereme všechny uvažované faktory.



S výstupem jsme se opět již setkali v části o jednoduché lineární regresi. Vidíme, že jediným rozdílem je seznam koeficientů b*, což jsou koeficienty nezávisle proměnných. Tyto

však neodpovídají odhadům parametrů z uvažovaného regresního modelu. Jedná se o speciálně upravené odhady parametrů z jiného modelu, které nám umožňují porovnat relativní vliv jednotlivých regresorů na závisle proměnnou. Statisticky významné regresní koeficienty jsou zvýrazněny červenou barvou.

≂ 🗊 🖓 🖓 🖓 🖨 🗑	STATISTICA Cz - [data-muzi - List1]	
Soubor Domů Upravit Zobrazit Formát Statistiky	Data mining Grafy Nástroje Data Sešit	🔍 Feature Finder 🛛 Možnosti 🕇 🏠 👘 👘
Základní Vícenásobná ANOVA Neparametrické Prokládání Rozdělení a statistiky rozdělení simulace	k Pokročilé modely - ¥ Neuron, sítě 🔛 Diagramy řízení kvality - 🚡 Analýza procesu 📑 ¥ Vicerozm. anal ŵ PLS, PCA, ₩ Multivariate 🚱 DOE 📲 Multivariate 🚱 Six Sigma - 🖬	STATISTICA VB Dávk. analýza (dle skupin) Statistiky blokku dat +
Základ	🖉 Výsledky - vícenásobná regrese: List1 v data-muzi	8 stroje
List1 List3 bit vik vyška boty 1 18 193 46 2 20 175 43 3 20 178 43 4 20 181 42 5 20 180 44 6 20 191 43 7 19 185 44 6 20 191 43 7 19 185 44 7 19 185 44 7 19 185 44 10 22 183 43 10 22 183 43 11 21 180 44 12 20 182 45 13 21 187 42 14 22 17	Výzledky- vicerozm. regrese Záv.prom. :vána vicenás. 2 = ,9021931 F = 14,6124 Záv.prom. :vána vicenás. 2 = ,5155615 sv = 10,33 Poč. připadki t4 upravená RZ= ,7592723 p = ,00000 Abs.člen: -61,046344 souží t4 = ,000000 p = ,000000 Abs.člen: -61,046344 souží t4 = ,000000 p = ,000000 vék t*=.00 výška t*=,021 boky b*=,17 preteľ b*=,180 sport b*=,143 spánek b*=.,13 pctteľ b*=,400 (významá b* jsou zvýraznána červeně) P Afs pro zvýrazňní dekki: .05 S S Zákdoní výsledky Detaní výsledky Residus (předpoklady/předpověd) P 105 89 102 40 7 7 2 79	- , 0194
23 24 189 45	107 95 110 41 7 8 2 78	La de la companya de
4 • • · · · · · · · · · · · · · · · · ·		
🕅 🗵 2D bodové grafy 📊 2D histogramy	eziduální analýza: List 1 v	
Pro nápovědu stiskněte F1		List1 v data-mu Filtr - Váhy: VYPN CAP NUM REC
📀 ⋵ 📋 o 🚺		CS 🔺 隆 🔐 📲 21:13 1.5.2016

Podrobnější výstup obsahující i hodnoty regresních koeficientů z hledaného modelu získáme výběrem možnosti *Výpočet: Výsledky regrese* v šedé části zobrazeného dialogového okna. Výsledkem je následující tabulka.

	Výsledky regre ,81575815 Upr	√ýsledky regrese se závislou proměnnou : váha (List1 v data-muzi) R= ,90319331 R2= ,81575815 Upravené R2= ,75992729 F(10,33)=14,611 p												
N=44	b*	Sm.chyba z b*	b	Sm.chyba z b	t(33)	p-hodn.								
Abs.čl			-81,1054	33,00325	-2,45750	0,019414								
věk	-0,094636	0,085961	-0,6895	0,62632	-1,10091	0,278899								
výška	0,020790	0,159101	0,0314	0,24016	0,13067	0,896826								
boty	0,199542	0,140520	1,4213	1,00090	1,42003	0,164978								
prsa	0,115434	0,141274	0,1618	0,19802	0,81710	0,419733								
pas	0,290648	0,165021	0,4575	0,25975	1,76128	0,087455								
boky	0,176413	0,165383	0,1801	0,16883	1,06669	0,293854								
krk	0,179674	0,102106	0,9003	0,51162	1,75969	0,087729								
sport	0,142656	0,145810	0,3402	0,34776	0,97837	0,335012								
spánek	-0,103072	0,087690	-1,3420	1,14173	-1,17541	0,248240								
pc+tv	0,399720	0,126830	1,3130	0,41660	3,15161	0,003445								

Z výsledných p-hodnot v posledním sloupci je zřejmé, že velká většina regresních koeficientů není významná, proto bychom se v dalším kroku měli pokusit vybrat pouze "důležité" faktory. K tomu využijeme krokové regrese (neboli stepwise regrese).

K tomu účelu se v dialogovém okně Vícenásobné regrese přepneme na záložku *Detailní výsledky*. Opakovaným stisknutím tlačítka *Další* pak budeme postupně spouštět jednotlivé regresní kroky. Postupně se nám budou zobrazovat další koeficienty b*, tedy další faktory

přidané do modelu. Dokud jsou výsledné b* zbarveny červeně, jsou dané faktory významné, jakmile se přidávané koeficienty b* zbarví namodro, příslušné faktory již významné nejsou, proto je do výsledného modelu nezapracujeme. Poslední krok zpracování poznáme jednoduše, a to tak, že se popis tlačítka *Další* změní na *OK*.

Soubor Domů	Upravit	Zobrazit Formát Statistiky	STATISTICA Cz - [data-mi Data mining Grafy Nástroje	uzi* - Průměry a směrodatné : Data Sešit	odchylky (List1 v data	-muzi)]	R Feature Finds
Základní Vícenásol statistiky regres	bnà ANOVA Ne	parametrické Prokládání Rozdělení a statistiky rozdělení simulace	Pokročilé modely - Meuron. sitě Vicerozm. anal Analýza síly testu UPAC	Diagramy řízení kvality v Multivariate	Analýza procesu C DOE G Six Sigma •	STATISTICA VB	Statistiky bloku dat -
E.	2	🖉 Výsledky - vícenásobná regrese: List	tl v data-muzi	2 🗙	N ^{istika}	Nástroje	
	Proměnná věk výška boty prsa pas boky krk spónt spónek pc+tel váha	Výsledky- vícerozm. regrese Záv.prom. :váha Poč. případů: 44 Směrodatná chyt Abs.člen: -01,64402254 Sm. pas b**,815 (významná b* jsou zvýrazněné					
94949		Afa pro zvýraznění efektů:	Residua/předpoklady/předpovědi	Delší Stomo			
		Wypočet: výsledky regrese	Pargiáln í korelace	🔁 Možnosti 🕶			2
E Nep		ANOVA (Celk. vhodnost modelu)	Redyndance	Anal Skup.			
		Kogariance koeficientů	Wisledky krokové regrese				
Vice		Současná s <u>weep</u> matice	ANOVA upravená pro grůměry		J		

Soubor	C C d	Upravit	Zobrazit Formát	Statistiky	STATIST Data mining	ICA Cz - [data-m	uzi" - Průměry a směrodatné (e Data Sešit	odchylky (List1 v data	-muzi)]	R Feature Find
Základní Víc statistiky	enásobná regrese	ANOVA Ne	parametrické Prokládání Ro rozdělení	ozdělení a simulace	Pokročilé modely * Vicerozm. anal. * Analýza síly testu	Neuron. sitě Gi PLS, PCA, VEPAC	Diagramy řízení kvality * Multivariate	Analýza procesu C DOE G Six Sigma •	STATISTICA VB	Statistiky bloku dat •
	J. ST		🔀 Výsledky - vícenásobr	ná regrese: Li	st1 v data-muzi		2	stika	Nastroje	
	Handrand H Andra Handrand Handran	roměnná ěk řška oty rsa as oky rk port pánek c+tel iha	Výsledky- vicerozm. zegrese(Krok 3) Záv.prom. :váha vicenás. 2 = .88083833 F = 46,15762 Poč. připadů: 44 upravené 227587623 sv = 3,40 Směrodstná chýba odhadu : 6,61218828 Abs.člen: -83,0590514 24,40702 : (40) = -3,405 p = .0015 pas b**,570 pc*tel b**,341 boty b**,254 (významá b* jsou zvýzazňan červeně)							
	0.0.0		Aťa pro zvýraznění efektů: Základní výsledky Deta	.05 🔮 Ini výsledky	Residua/předpoklady/j	oředpovědi	Storno			
	-		Wypočet: výsled	ky regrese	Pargiáin	i korelace	🔁 Možnosti 🕶			
	lep		ANOVA (Celk, vho	inost modelu)	Redy	ndance	Anal Skup.			
64			Koyariance ko	eficientú	Výsledky kro	kov <u>é</u> regrese				
-	1		Současná syre	ep matice	ANOVA uprav	ená pro grúměry				
E VI	ice			_						

Soubor Dor	C 🖨 📅 ∓ nú Upravit	Zobrazit Formát	STATIST Statistiky Data mining	TICA Cz - [data-mi Grafy Nástroje	uzi* - Průměry a směrodatné : Data Sešit	odchylky (List1 v data	-muzi)]	R Feature Finde
Základni Vicená statistiky regr	sobná ANOVA Ne	eparametrické Prokládání Roz statistiky rozdělení si	dělení a mulace 🕄 Pokročilé modely -	Neuron. sitě Sti PLS, PCA, VEPAC	Diagramy řízení kvality • Multivariate	Analýza procesu Analýza procesu DOE Six Sigma *	STATISTICA VB	Statistiky bloku dat -
E	2	Výsledky - vícenásobní	á regrese: List1 v data-muzi		8 ×	listika	Nástroje	
ورغارية رغارية والمرابع والمرابع	Proměnná věk výška boty prsa pas boky krk sport spórt spórt spórtel váha	Výsledky- viceroz již žádné F na za Závyprom. :váha Poč. případů: 44 Směro Abs.člen: -94,202 pas D**, krk D**, (významná D* jsou	m. regrese(Krók 5, kon hhrnuti nepřeshuje daná vicenás. R = upravené R2= upravené R2= datná chyba odhadu : 5, 42844 Sm. chyba: 25,778; 437 pettel ben; 129 prss btn; svýrazněna červeně)	ečné žešení) me "99557543 "79247705 m "77196088 455983650 26 t(38) = 319 126	F = 30,11283 W = 5,38 p = ,00000 -3,654 p = ,0008 hoty b*=,213			
and the second		Alfa pro zvýraznění efektů: Základní výsledky Detair	.05 🖨	předpověd	Storno			
-		time vypocet: vysledky	yregrese EE Pargian	i korelace	(24) Možnosti -			
E 🔐 Nep	-	BEB ANOVA (Celk. vhodr	nost modelu) (ISE Redu	indance	Anal Skup.	_		
Carpella		INTE Koyanance koe	ncientu (IBB Vysledky kn	okové regrese				
Uice		(IBIB Sonceaus all eet	DINADOR [ISB ANUVALIDIAN	vena pro grunary				

Po provedení posledního kroku zvolíme možnost *Výsledky krokové regrese* a tím obdržíme následující tabulku obsahující závěrečné informace o celé krokové regresi.

	Výsledky krokové regrese ; ZP: váha (List1 v data-muzi)								
Proměnná	Krok +do/-ven	Vícenás. R	Vícenás. R^2	R^2 změna	F zah/vyjm	p-hodn.	Zahrnuté proměnné		
pas	1	0,815387	0,664856	0,664856	83,31928	0,000000	1		
pc+tv	2	0,859609	0,738928	0,074072	11,63253	0,001468	2		
boty	3	0,880838	0,775876	0,036949	6,59434	0,014072	3		
krk	4	0,890004	0,792107	0,016231	3,04492	0,088866	4		
prsa	5	0,893575	0,798477	0,006370	1,20107	0,280005	5		

Dále v záložce Základní výsledky vybereme možnost Výpočet: Výsledky regrese.

ogram Bodový Průmě graf	Image:
B Ista-muri 🖉	ný 😥 Výsledky - vícenásobná regrese: List1 v data-muzi
Contraction of the second	Výsledky-vicerozm. regrese(Krok 5, kocedné felení) jiš žádof F na skruuti nepřesuky daná me Záv.prom. rváha vicenás. R = (8337643 F = 50,11283 Záv.prom. rváha vicenás. R = (8337643 F = 50,11283 Poč. připadi: 4 pravené 22 - 738705 sv = 5,38 Poč. připadi: 44 pravené 22 - 77136284 p = ,00008 Jak.čien -54,004284 Sm. dvýka 12,77124 kt 189 = -3,654 p = ,0008 pas bre.dl. pr.dl.
	1 (významná b. jsou zvýzaznéna čezvené)
	1 Afa pro zvýrazmění efektů: 0.05 💭 🖂 OK 1 Základní výsledky Detalní výsledky Residus/předpoklady/předpovědi Stomo
	1 Výpočet: Výsledky regrese
	1 drad State

	Výsledky regr ,79847705 Up	/sledky regrese se závislou proměnnou : váha (List1 v data-muzi) R= ,89357543 R2= 9847705 Upravené R2= ,77196088 F(5,38)=30,113 p									
N=44	b*	Sm.chyba z b*	b	Sm.chyba z b	t(38)	p-hodn.					
Abs. člen			-94,2024	25,77826	-3,65434	0,000776					
pas	0,436787	0,121088	0,6875	0,19060	3,60719	0,000888					
pc+tv	0,318795	0,079413	1,0471	0,26085	4,01440	0,000271					
boty	0,212814	0,099877	1,5158	0,71141	2,13076	0,039640					
krk	0,128902	0,091647	0,6459	0,45922	1,40650	0,167701					
prsa	0,126218	0,115169	0,1769	0,16143	1,09593	0,280005					

Poté obdržíme tabulku obsahující údaje o jednotlivých regresorech a jejich koeficientech.

Z obou tabulek můžeme vidět, že z devíti uvažovaných faktorů jsou významné pouze tři, a to Obvod pasu, Počet hodin strávených u TV+PC a Velikost bot. Pořadí jejich významnosti je shodné s pořadím uvedení.

Výsledný model má tedy tvar:

V = -94,2024 + 0,6875*P + 1,0471*T + 1,5158*O,

kde V je váha v kilogramech, P je obvod pasu v centimetrech, T je čas strávený u PC+TV a O je velikost bot.

Z této rovnice je například zřejmé, že pokud by průměrný mladý muž v pase přibral o 1 cm, pak se to na váze v průměru projeví přírůstkem 0,69 kg. Obdobně i pro ostatní faktory.

Vidíme, že i když jsme problémové faktory nevyřadili ze zpracování, k problémům nedošlo. Vysoká korelovanost mezi veličinami Boty a Výška nevadí, protože do výsledného modelu byla vybrána pouze jedna z nich. Také problematičnost veličiny Sport řešit nemusíme, ani ona nebyla do závěrečného modelu vybrána.

Vidíme, že modely žen (byl vytvářen pomocí Excelu) a mužů (právě jsme vytvořili) se liší. Pouze faktor Obvod pasu se ukázal významným v obou modelech. U žen byl jediným dalším významným faktorem identifikován Obvod boků, kdežto u mužů byly kromě Obvodu pasu identifikovány ještě dva významné faktory, a to Čas strávený u TV+PC a Velikost bot. Vzhledem k tomu, že mezi faktory Čas strávený u TV+PC a Čas strávený sportem byla identifikována silná nepřímá závislost, je vidět, že i tento čas je skrytě v modelu obsažen.