

Cvičení 14.: Mnohonásobná lineární regresní analýza

Příklad 1.: Známe údaje o 20 starších automobilech stejné značky. Veličina Y je cena vozu (v tisících Kč), veličina X_1 udává stáří vozu v letech a veličina X_2 udává počet ujetých km (v tisících).

	1	2	3
	X1	X2	Y
1	3	106	167
2	4	134	139
3	4	51	159
4	5	102	135
5	5	125	139
6	6	104	139
7	6	49	139
8	6	74	145
9	7	156	109
10	7	147	119
11	7	59	129
12	7	83	135
13	8	137	99
14	8	91	99
15	8	114	109
16	8	97	119
17	9	298	63
18	9	165	69
19	9	172	76
20	9	145	77

Předpokládáme, že závislost ceny vozu na jeho stáří a počtu ujetých km lze modelovat regresní rovinou.

Jsou k dispozici tyto výstupy ze systému STATISTICA:

Korelace (ojeta_auta.sta) Označ. korelace jsou významné na hlad. $p < ,00000$ N=20 (Celé případy vynechány u ChD)		
Proměnná	X1	X2
Y	-0,907679	-0,728202

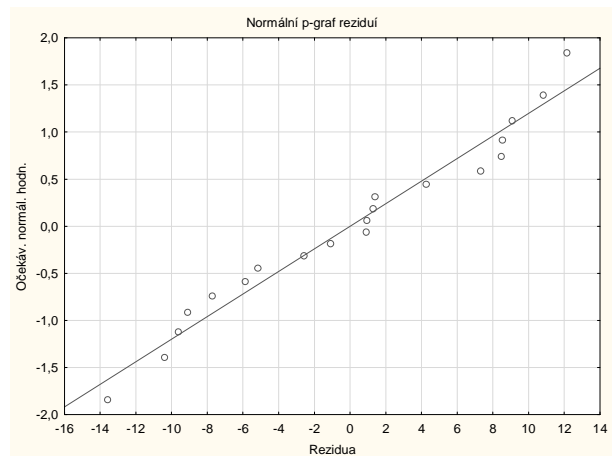
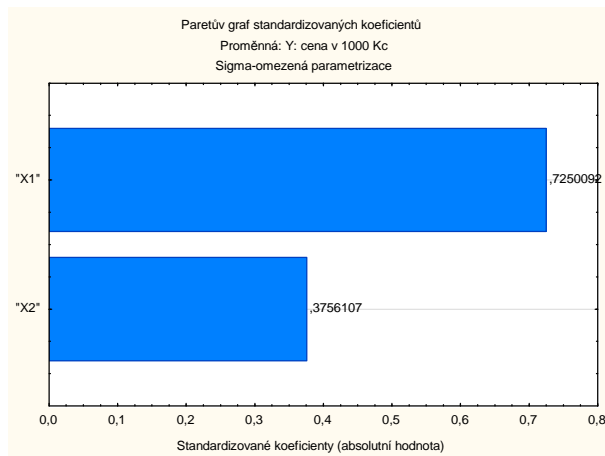
Parciální korelace (ojeta_auta.sta) S vyloučením vlivu: X2 Označ. korelace jsou významné na hlad. $p < ,00000$ N=20 (Celé případy vynechány u ChD)		
Proměnná	Y	X1
Y	1,000000	-0,924323
X1	-0,924323	1,000000

Parciální korelace (ojeta_auta.sta) S vyloučením vlivu: X1 Označ. korelace jsou významné na hlad. $p < ,00000$ N=20 (Celé případy vynechány u ChD)		
Proměnná	Y	X2
Y	1,000000	-0,782051
X2	-0,782051	1,000000

Statistiky kolineace za daných podmínek (ojeta_auta.sta) Sigma-omezená parametrizace								
Efekt	Toler.	Rozptyl Infl fak	R^2	Y Beta v	Y Parciál.	Y Semipar.	Y t	Y p
"X1"	0,7634866	1,3097807	0,2365134	-0,7250092	-0,9243230	-0,6334965	-9,986823	0,0000000
"X2"	0,7634866	1,3097807	0,2365134	-0,3756107	-0,7820508	-0,3282001	-5,173944	0,0000763

Výsledky regrese se závislou proměnnou : Y (ojeta_auta.sta) R= ,96519211 R2= ,93159581 Upravené R2= ,92354825 F(2,17)=115,76 $p < ,00000$ Směrod. chyba odhadu : 8,2779						
N=20	b*	Sm.chyba z b*	b	Sm.chyba z b	t(17)	p-hodn.
Abs.člen			222,5918	7,284279	30,55784	0,000000
X1	-0,725009	0,072597	-11,8497	1,186534	-9,98682	0,000000
X2	-0,375611	0,072597	-0,2022	0,039083	-5,17394	0,000076

Analýza rozptylu (ojeta_auta.sta)					
Efekt	Součet čtverců	sv	Průměr čtverců	F	p-hodn.
Regres.	15864,84	2	7932,422	115,7614	0,000000
Rezid.	1164,91	17	68,524		
Celk.	17029,75				



Durbin-Watsonovo d (ojeta_auta.sta) a sériové korelace reziduí		
	Durbin- Watson.d	Sériové korelace
Odhad	1,600032	0,173167

Úkoly:

- a) Porovnejte párové a parciální koeficienty korelace.
- b) Zjistěte, zda se v modelu vyskytuje multikolinearita.
- c) Napište rovnici regresní roviny a vysvětlete význam odhadů regresních parametrů.
- d) Interpretujte index determinace.
- e) Najděte odhad rozptylu náhodných odchylek.
- f) Interpretujte výsledky celkového F-testu a dílčích t-testů.
- g) Má větší vliv stáří nebo počet ujetých kilometrů.
- h) Jaký je regresní odhad ceny auta, které je staré 6 let a má najeto 106 000 km?
- i) Je splněn předpoklad normality reziduí?
- j) Lze na hladině významnosti 0,05 zamítnout hypotézu o kladné autokorelaci reziduí?

Výsledky:

Ad a) Mezi cenou vozu a jeho stářím existuje velmi silná nepřímá lineární závislost (-0,91), která se ještě nepatrně zesílí, když eliminujeme vliv počtu ujetých km (-0,92).
Mezi cenou vozu a počtem ujetých km existuje silná nepřímá lineární závislost (-0,73), která se ještě zesílí při eliminaci vlivu stáří vozu (-0,78).

Ad b) V modelu se multikolinearita nevyskytuje, koeficient VIF nabývá hodnoty pouze 1,31.

Ad c) Cena nového vozu je 222 590 Kč
Zvýší-li se stáří vozu o rok jeho cena klesne o 11 850 Kč
Zvýší-li se počet ujetých km o 1000 km, jeho cena klesne 200 Kč.

Ad d) Variabilita ceny auta je z 93,2 % vysvětlena zvoleným regresním modelem.

Ad e) $s^2 = 68,524$

Ad f) Celkový F test je významný, dílčí t-testy rovněž.

Ad g) Větší vliv na cenu vozu má jeho stáří než počet najetých km.

Ad h)) cena vozu = 130 290 Kč

Ad i) Tečky v N-P grafu jsou blízko ideální přímky, rezidua splňují předpoklad normality.

Ad j) Vzhledem k tomu, že kritické hodnoty D-W testu pro $\alpha = 0,05$, $n = 2$, $p = 2$ jsou 1,1 a 1,54 a testová statistika $D = 1,6$, nelze na hladině významnosti 0,05 prokázat pozitivní autokorelaci.

Příklad 2.: U 19 vzorků potravinářské pšenice byl zjišťován obsah zinku v zrně (proměnná Y), v kořenech (proměnná X_1), v otrubách (X_2) a ve stonku a listech (X_3).

Y	X_1	X_2	X_3
175	164	198	162
169	160	198	159
175	158	211	164
181	162	211	162
539	520	567	523
526	502	540	491
344	339	355	334
475	460	500	446
820	683	813	695
841	731	832	714
828	710	846	697
775	716	818	709
622	543	635	563
661	577	712	580
579	505	596	531
936	790	946	814
903	806	946	834
927	793	912	824
889	820	919	807

- Normalitu proměnných Y , X_1 , X_2 , X_3 posuďte pomocí Lilieforsovým testem s hladinou významnosti 0,05.
- Závislost mezi dvojicemi proměnných (Y, X_1) , (Y, X_2) , (Y, X_3) znázorněte dvourozměrnými tečkovými diagramy.
- Vypočítejte výběrovou korelační matici všech čtyř proměnných a pro $\alpha = 0,05$ otestujte významnost jednotlivých korelačních koeficientů.
- Vypočítejte výběrové parciální korelační koeficienty $r_{Y, X_1.(X_2, X_3)}$, $r_{Y, X_2.(X_1, X_3)}$, $r_{Y, X_3.(X_1, X_2)}$ a porovnejte je s výběrovými párovými korelačními koeficienty r_{YX_1} , r_{YX_2} , r_{YX_3} . Na hladině významnosti $\alpha = 0,05$ testujte hypotézy o nevýznamnosti parciálních korelačních koeficientů $\rho_{Y, X_1.(X_2, X_3)}$, $\rho_{Y, X_2.(X_1, X_3)}$, $\rho_{Y, X_3.(X_1, X_2)}$.
- V první fázi zpracování předpokládejte, že je vhodný regresní model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$. Vypočítejte index determinace a interpretujte ho. Proveďte celkový F-test. Odhadněte parametry regresního modelu. Proveďte dílčí t-testy pro regresní koeficienty. Zjistěte odhad rozptylu. Vypočítejte parciální indexy determinace. (Hladinu významnosti volte $\alpha = 0,05$.)
- Posuďte pomocí B-koeficientů vliv jednotlivých nezávisle proměnných veličin na regresní model.
- Z regresního modelu odstraňte ty proměnné, jejichž regresní koeficienty se neprokázaly významné pro $\alpha = 0,05$. Sestavte nový regresní model a proveďte v něm tytéž úkoly jako v bodě e).
- Normalitu reziduí v tomto novém regresním modelu posuďte Lilieforsovým testem na hladině významnosti $\alpha = 0,05$.
- V novém regresním modelu najděte 95% interval spolehlivosti pro teoretickou regresní funkci a 95% predikční interval.
- Proveďte regresi metodou STEPWISE, a to jak Forward, tak Backward.

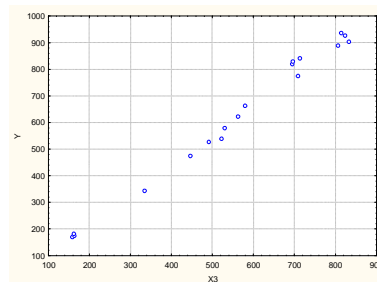
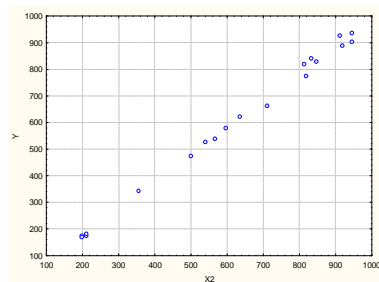
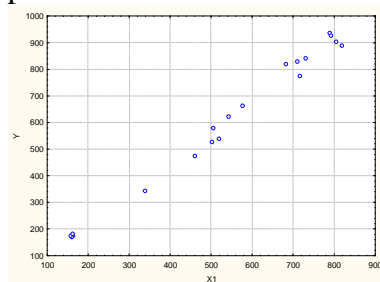
Řešení: Načteme datový soubor zinek.sta.
ad a) Výsledky Lilieforsova testu normality

proměnná	testová statistika	p-hodnota
Y	0,15792	> 0,2
X ₁	0,15613	> 0,2
X ₂	0,18177	< 0,1
X ₃	0,16420	< 0,2

Na hladině významnosti 0,05 nelze ani v jednom případě zamítnout hypotézu o normalitě.

ad b)

Dvourozměrné tečkové diagramy dvojic (Y,X₁), (Y,X₂), (Y,X₃) svědčí o existenci dosti silné přímé lineární závislosti.



ad c) Výběrová korelační matice proměnných Y, X₁, X₂, X₃ spolu s odpovídajícími p-hodnotami:

Proměnná	Y	X1	X2	X3
Y	1,0000	,9947	,9981	,9959
	p= ---	p=,000	p=0,00	p=0,00
X1	,9947	1,0000	,9954	,9980
	p=,000	p= ---	p=,000	p=0,00
X2	,9981	,9954	1,0000	,9962
	p=0,00	p=,000	p= ---	p=0,00
X3	,9959	,9980	,9962	1,0000
	p=0,00	p=0,00	p=0,00	p= ---

Na hladině významnosti 0,05 zamítáme hypotézu o nevýznamnosti jednotlivých korelačních koeficientů.

ad d)

Výběrový koeficient parciální korelace $r_{Y,X_1,(X_2,X_3)}$

Proměnná	Y	X1
Y	1,0000	-,0390
	p= ---	p=,882
X1	-,0390	1,0000
	p=,882	p= ---

Výběrový koeficient korelace r_{YX_1} je 0,9947, zatímco $r_{Y,X_1,(X_2,X_3)}$ je -0,039.

Pokud eliminujeme vliv proměnných X₂, X₃, tak mezi proměnnými Y a X₁ existuje velmi slabá nepřímá lineární závislost, která není na hladině 0,05 významná.

Výběrový koeficient parciální korelace $r_{Y,X_2,(X_1,X_3)}$

Proměnná	Y	X2
Y	1,0000	,7515
	p= ---	p=,001
X2	,7515	1,0000
	p=,001	p= ---

Výběrový koeficient korelace r_{YX_2} je 0,9981, zatímco $r_{Y,X_2,(X_1,X_3)}$ poklesl na 0,7515.

Pokud eliminujeme vliv proměnných X_1 , X_3 , tak mezi proměnnými Y a X_2 existuje silná přímá lineární závislost, která je na hladině 0,05 významná.

Výběrový koeficient parciální korelace $r_{Y,X_3,(X_1,X_2)}$

Proměnná	Y	X3
Y	1,0000	,2230
	p= ---	p=,390
X3	,2230	1,0000
	p=,390	p= ---

Výběrový koeficient korelace r_{YX_3} je 0,99589, zatímco $r_{Y,X_3,(X_1,X_2)}$ je pouze 0,223.

Pokud eliminujeme vliv proměnných X_1 , X_2 , tak mezi proměnnými Y a X_3 existuje slabá přímá lineární závislost, která není na hladině 0,05 významná.

Vidíme, že existují značné rozdíly mezi párovými a parciálními výběrovými korelačními koeficienty. Lze tedy soudit na existenci multikolinearity. O tom svědčí i koeficienty VIF:

Statistiky kolineace za daných podmínek (zinek.sta)								
Sigma-omezená parametrizace								
Efekt	Toler.	Rozptyl Infl fak	R ²	Y Beta v	Y Parciál.	Y Semipar.	Y t	Y p
X1	0,003802	262,9861	0,996198	-0,037425	-0,038960	-0,002308	-0,151006	0,881983
X2	0,007214	138,6290	0,992786	0,793836	0,751501	0,067422	4,411716	0,000505
X3	0,003120	320,5035	0,996880	0,242409	0,223005	0,013540	0,886006	0,389598

ad e) Výsledky pro regresní model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$

Výsledky regrese se závislou proměnnou : Y (zinek.sta)						
R= ,99824679 R ² = ,99649665 Upravené R ² = ,99579598						
F(3,15)=1422,2 p<,00000 Směrod. chyba odhadu : 18,094						
N=19	b*	Sm.chyba z b*	b	Sm.chyba z b	t(15)	p-hodn.
Abs.člen			-28,7607	10,60478	-2,71205	0,016066
X1	-0,037425	0,247835	-0,0439	0,29089	-0,15101	0,881983
X2	0,793836	0,179938	0,8079	0,18312	4,41172	0,000505
X3	0,242409	0,273598	0,2802	0,31623	0,88601	0,389598

Adjustovaný index determinace je 0,9958, tedy zvolený regresní model s proměnnými X_1 , X_2 , X_3 vysvětluje variabilitu proměnné Y z 99,58 %. Testová statistika pro celkový F-test nabývá hodnoty 1422,2, odpovídající p-hodnota je velmi blízká 0, tedy model jako celek je významný na hladině 0,05.

Odhad rozptylu získáme z tabulky analýzy rozptylu:

Efekt	Součet čtverců	sv	Průměr čtverců	F	Úroveň p
Regres.	1396846	3	465615,2	1422,205	0,000000
Rezid.	4911	15	327,4		
Celk.	1401757				

$$s^2 = 327,4$$

Odhadnutá regresní funkce má tvar: $\hat{Y} = -28,7607 - 0,0439x_1 + 0,8079x_2 + 0,2802x_3$.

Dílčí t-testy pro jednotlivé regresní koeficienty:

testová statistika pro test hypotézy $H_0: \beta_0 = 0$ je -2,71205, p-hodnota je 0,016066, tedy H_0 zamítáme na hladině významnosti 0,05;

testová statistika pro test hypotézy $H_0: \beta_1 = 0$ je -0,15101, p-hodnota je 0,881983, tedy H_0 nezamítáme na hladině významnosti 0,05;

testová statistika pro test hypotézy $H_0: \beta_2 = 0$ je 4,41172, p-hodnota je 0,000505, tedy H_0 zamítáme na hladině významnosti 0,05;

testová statistika pro test hypotézy $H_0: \beta_3 = 0$ je 0,88601, p-hodnota je 0,389598, tedy H_0 nezamítáme na hladině významnosti 0,05.

Výpočet parciálních indexů determinace:

$r_{Y,X_1}^2 = 0,9947^2 = 0,9894$ (Pokud do modelu $Y = \beta_0 + \varepsilon$ zařadíme veličinu X_1 , pak bude vysvětlovat variabilitu hodnot veličiny Y z 98,94 %.)

$r_{Y,X_2,X_1}^2 = 0,8079^2 = 0,6527$ (Pokud do modelu $Y = \beta_0 + \beta_1 X_1 + \varepsilon$ zařadíme veličinu X_2 , pak bude vysvětlovat variabilitu hodnot veličiny Y z 65,27 %.)

$r_{Y,X_3,(X_1,X_2)}^2 = 0,223^2 = 0,0497$ (Pokud do modelu $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ zařadíme veličinu X_3 , pak bude vysvětlovat variabilitu hodnot veličiny Y z 4,97 %.)

ad f) Interpretace B-koeficientů:

$b_1^* = -0,037425$, $b_2^* = 0,793836$, $b_3^* = 0,242409$. V absolutní hodnotě je největší b_2^* , tedy obsah zinku v otrubách má největší vliv na obsah zinku v zrně.

ad g) Protože dílčí t-testy prokázaly, že na hladině 0,05 nejsou proměnné X_1 a X_3 významné, sestavíme nový regresní model $Y = \beta_0 + \beta_2 X_2 + \varepsilon$.

Výsledky regrese se závislou proměnnou : Y (zinek.sta)						
R= ,99807615 R2= ,99615600 Upravené R2= ,99592988						
F(1,17)=4405,5 p<0,0000 Směrod. chyba odhadu : 17,803						
N=19	b*	Sm.chyba z b*	b	Sm.chyba z b	t(17)	p-hodn.
Abs.člen			-30,2507	10,31117	-2,93378	0,009274
X2	0,998076	0,015037	1,0157	0,01530	66,37372	0,000000

Adjustovaný index determinace je 0,9959, tedy zvolený regresní model s proměnnou X_2 vysvětluje variabilitu proměnné Y z 99,59 %. Testová statistika pro celkový F-test nabývá hodnoty 4405,5, odpovídající p-hodnota je velmi blízká 0, tedy model jako celek je významný na hladině 0,05.

Vidíme, že $\hat{Y} = -30,2507 + 1,0157 X_2$.

Dílčí t-testy pro jednotlivé regresní koeficienty:

testová statistika pro test hypotézy $H_0: \beta_0 = 0$ je -2,93378, p-hodnota je 0,009274, tedy H_0 zamítáme na hladině významnosti 0,05;

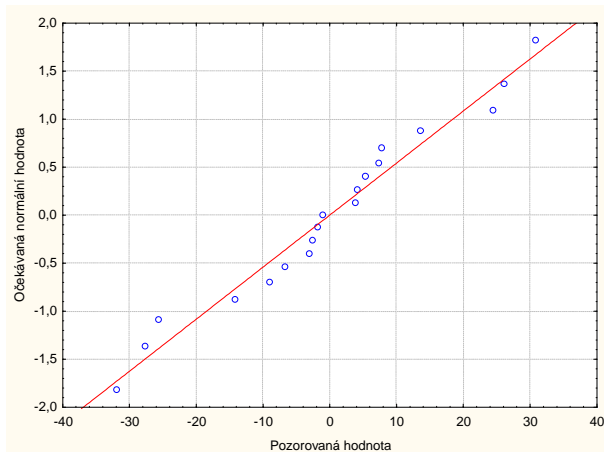
testová statistika pro test hypotézy $H_0: \beta_2 = 0$ je 66,37372, p-hodnota je 0,000000, tedy H_0 zamítáme na hladině významnosti 0,05.

ad h) Ověření normality reziduí

Abychom mohli analyzovat rezidua, musíme je uložit. Ve výstupní tabulce zvolíme Rezidua/předpoklady/předpovědi – Reziduální analýza – Uložit – Uložit rezidua& předpovědi - OK.

Testová statistika pro K-S test nabývá hodnoty 0,1163, odpovídající p-hodnota je větší než 0,20, tedy hypotézu o normalitě reziduí nezamítáme na hladině významnosti 0,05.

Pro úplnost ještě posoudíme vzhled N-P plotu:

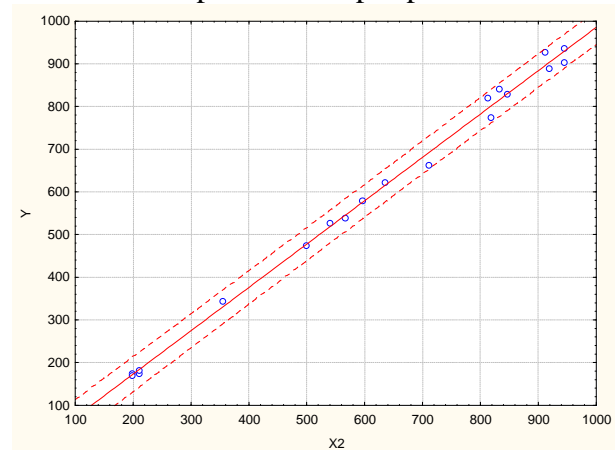
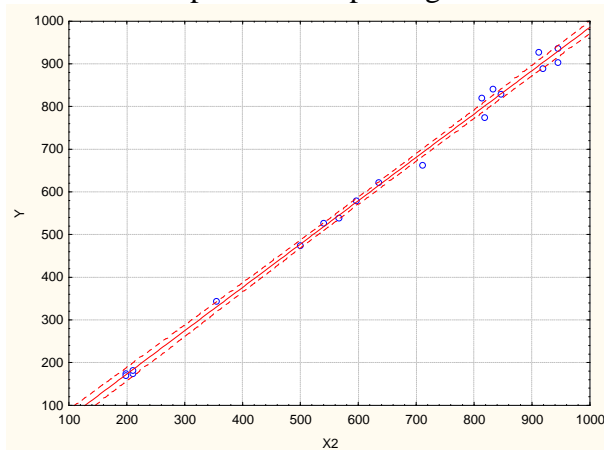


N-P plot svědčí o tom, že rozložení reziduí se příliš neliší od normálního rozložení.

ad i) Intervaly spolehlivosti pro regresní funkci a pro predikci získáme pomocí dvourozměrných tečkových diagramů, kde v Detailích vybereme lineární proložení a zvolíme regresní pásy.

95% interval spolehlivosti pro regresní funkci

95% interval spolehlivosti pro predikci



ad j) Nejprve aplikujeme metodu Forward:

Statistiky – Vícerozměrná regrese – Proměnné – Závisle proměnná Y, Nezávisle proměnné X1, X2, X3 – OK – Detailní nastavení – zaškrtneme Další možnosti – OK – Metoda – zvolíme Kroková dopředná – na záložce Metoda zvolíme Zobrazit výsledky Po každém kroku – OK (V kroku 0 nejsou v regresní rovnici žádné proměnné.) Klikneme na Další – Výpočet: Výsledky regrese.

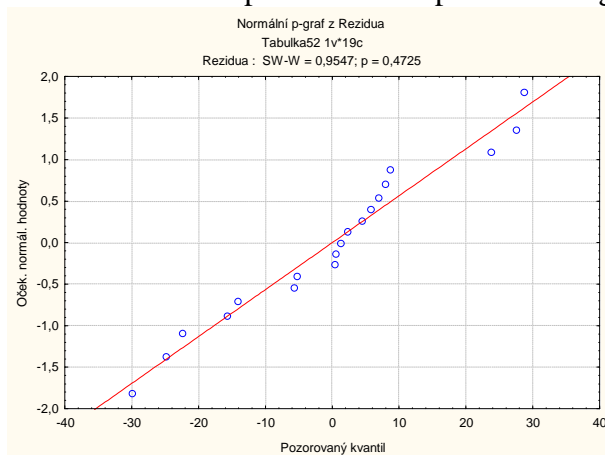
Výsledky regrese se závislou proměnnou : Y (zinek.sta)						
R= ,99807615 R2= ,99615600 Upravené R2= ,99592988						
F(1,17)=4405,5 p<0,0000 Směrod. chyba odhadu : 17,803						
N=19	b*	Sm.chyba z b*	b	Sm.chyba z b	t(17)	p-hodn.
Abs.člen			-30,2507	10,31117	-2,93378	0,009274
X2	0,998076	0,015037	1,0157	0,01530	66,37372	0,000000

V prvním kroku byla vybrána proměnná X2. Opět klikneme na Další a dostaneme výsledky kroku 2, který je již konečný:

Výsledky regrese se závislou proměnnou : Y (zinek.sta) R= ,99824412 R2= ,99649132 Upravené R2= ,99605274 F(2,16)=2272,1 p<0,0000 Směrod. chyba odhadu : 17,533						
N=19	b*	Sm.chyba z b*	b	Sm.chyba z b	t(16)	p-hodn.
Abs.člen			-28,9426	10,20929	-2,83493	0,011948
X2	0,788109	0,170440	0,8020	0,17345	4,62396	0,000282
X3	0,210764	0,170440	0,2436	0,19700	1,23659	0,234086

Empirická regresní funkce má tvar $\hat{Y} = -28,9426 + 0,802x_2 + 0,2436x_3$.
Model jako celek je významný na hladině 0,05, avšak nezávisle proměnná X_3 významná není. Přispívá však k vysvětlení variability hodnot závisle proměnné veličiny Y. Adjustovaný index determinace je 0,9961. V modelu s nezávisle proměnnou X_2 byl 0,9959 a v modelu se všemi třemi nezávisle proměnnými byl 0,9958.

Normalitu reziduí prozkoumáme pomocí N-P grafu a S-W testu:



Rezidua neporušují předpoklad normality.

Nyní provedeme metodu Backward:

Statistiky – Vícerozměrná regrese – Proměnné – Závisle proměnná Y, Nezávisle proměnné X_1, X_2, X_3 – OK – Detailní nastavení – zaškrtneme Další možnosti – OK – Metoda – zvolíme Kroková zpětná – na záložce Metoda zvolíme Zobrazit výsledky Po každém kroku – OK – Výpočet:Výsledky regrese.

Výsledky regrese se závislou proměnnou : Y (zinek.sta) R= ,99824679 R2= ,99649665 Upravené R2= ,99579598 F(3,15)=1422,2 p<,00000 Směrod. chyba odhadu : 18,094						
N=19	b*	Sm.chyba z b*	b	Sm.chyba z b	t(15)	p-hodn.
Abs.člen			-28,7607	10,60478	-2,71205	0,016066
X1	-0,037425	0,247835	-0,0439	0,29089	-0,15101	0,881983
X2	0,793836	0,179938	0,8079	0,18312	4,41172	0,000505
X3	0,242409	0,273598	0,2802	0,31623	0,88601	0,389598

V prvním kroku byly zařazeny všechny proměnné.

Klikneme na Další – Výpočet: Výsledky regrese.

Výsledky regrese se závislou proměnnou : Y (zinek.sta) R= ,99824412 R2= ,99649132 Upravené R2= ,99605274 F(2,16)=2272,1 p<0,0000 Směrod. chyba odhadu : 17,533						
N=19	b*	Sm.chyba z b*	b	Sm.chyba z b	t(16)	p-hodn.
Abs.člen			-28,9426	10,20929	-2,83493	0,011948
X2	0,788109	0,170440	0,8020	0,17345	4,62396	0,000282
X3	0,210764	0,170440	0,2436	0,19700	1,23659	0,234086

V tomto kroku byla vyloučena proměnná X1.

Opět klikneme na Další – Výpočet: Výsledky regrese a dostaneme konečnou tabulku:

Výsledky regrese se závislou proměnnou : Y (zinek.sta) R= ,99807615 R2= ,99615600 Upravené R2= ,99592988 F(1,17)=4405,5 p<0,0000 Směrod. chyba odhadu : 17,803						
N=19	b*	Sm.chyba z b*	b	Sm.chyba z b	t(17)	p-hodn.
Abs.člen			-30,2507	10,31117	-2,93378	0,009274
X2	0,998076	0,015037	1,0157	0,01530	66,37372	0,000000

Vidíme, že metoda STEPWISE, Backward poskytla stejné výsledky jako metoda ENTER.

Příklad k samostatnému řešení:

Byla změřena výška 20 osmnáctiletých chlapců (proměnná Y) a dále byly zjištěny výšky jejich příbuzných ve věku 18 let: X1 ... výška matky, X2 ... výška otce, X3 ... výška babičky z matčiny strany, X4 ... výška dědečka z otcovy strany, X5 ... výška babičky z otcovy strany, X6 ... výška dědečka z otcovy strany, X7 ... výška chlapce při narození.

Data jsou uložena v souboru vysky_pribuznych.sta.

Nejprve proveďte korelační analýzu: vypočtete koeficienty korelace proměnné Y se všemi nezávisle proměnnými.

Korelace (vysky_pribuznych.sta) Označ. korelace jsou významné na hlad. p < ,05000 N=20 (Celé případy vynechány u ChD)							
Proměnná	x1	x2	x3	x4	x5	x6	x7
y	0,395117	0,791297	-0,211974	-0,260861	0,193163	-0,230270	0,264300

Dále vypočtete parciální korelační koeficienty proměnné Y s proměnnou X1 při vyloučení vlivu X3 a X4 a s proměnnou X2 při vyloučení vlivu X5 a X6.

Parciální korelace (vysky_pribuznych.sta) S vyloučením vlivu: x3 a x4 Označ. korelace jsou významné na hlad. p < ,05000 N=20 (Celé případy vynechány u ChD)		
Proměnná	y	x1
y	1,0000	,3600
	p= ---	p=,142
x1	,3600	1,0000
	p=,142	p= ---

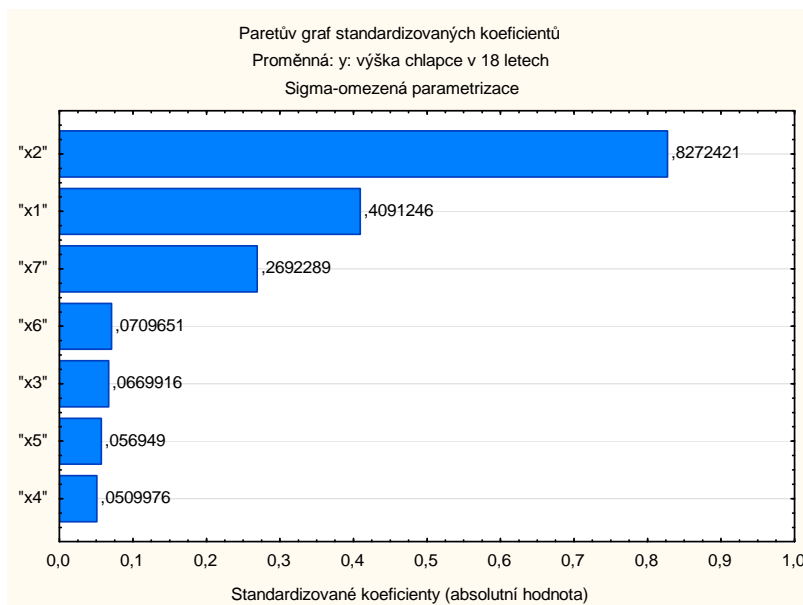
Parciální korelace (vysky_pribuznych) S vyloučením vlivu: x5 a x6 Označ. korelace jsou významné na hlad. $p < ,05000$ N=20 (Celé případy vynechány u ChD)		
Proměnná	y	x2
y	1,000000	0,763265
x2	0,763265	1,000000

Pomocí koeficientů VIF prověřte, zda v modelu, který vysvětluje proměnnou Y pomocí proměnných X1 až X7, mezi nezávisle proměnnými veličinami existuje multikolinearita.

Statistiky kolineace za daných podmínek (vysky_pribuznych.sta) Sigma-omezená parametrizace								
Efekt	Toler.	Rozptyl Infl fak	R ²	y Beta v	y Parciál.	y Semipar.	y t	y p
"x1"	0,6093157	1,6411853	0,3906843	0,4091246	0,7398021	0,3193573	3,8089422	0,0024890
"x2"	0,6095369	1,6405897	0,3904631	0,8272421	0,9120212	0,6458515	7,7030062	0,0000055
"x3"	0,8091856	1,2358105	0,1908144	-0,0669916	-0,2031556	-0,0602621	-0,7187401	0,4860614
"x4"	0,8192387	1,2206455	0,1807613	0,0509976	0,1569551	0,0461588	0,5505319	0,5920558
"x5"	0,6902695	1,4487094	0,3097305	0,0569490	0,1607847	0,0473146	0,5643166	0,5829329
"x6"	0,6513272	1,5353267	0,3486728	-0,0709651	-0,1934629	-0,0572722	-0,6830803	0,5075302
"x7"	0,6796494	1,4713469	0,3203506	0,2692289	0,6071912	0,2219546	2,6472305	0,0212876

V uvedeném regresním modelu posuďte pomocí B-koeficientů vliv jednotlivých nezávisle proměnných na Y. Použijte také Paretův diagram.

N=20	b*
Abs. člen	
x1	0,409125
x2	0,827242
x3	-0,066992
x4	0,050998
x5	0,056949
x6	-0,070965
x7	0,269229



Nyní pro výstavbu modelu použijte dopřednou i zpětnou krokovou metodu a jejich výsledky porovnejte.

Dopředná metoda:

Výsledky regrese se závislou proměnnou : y (vysky_pribuznych.sta) R= ,95111334 R2= ,90461658 Upravené R2= ,88673219 F(3,16)=50,581 p<,00000 Směrod. chyba odhadu : 2,3719						
N=20	b*	Sm.chyba z b*	b	Sm.chyba z b	t(16)	p-hodn.
Abs.člen			-199,722	33,79003	-5,91069	0,000022
x2	0,873161	0,078496	1,106	0,09942	11,12369	0,000000
x1	0,364507	0,085585	0,688	0,16156	4,25898	0,000600
x7	0,263581	0,086598	1,373	0,45103	3,04372	0,007740

Zpětná metoda:

Výsledky regrese se závislou proměnnou : y (vysky_pribuznych.sta) R= ,92162267 R2= ,84938835 Upravené R2= ,83166934 F(2,17)=47,937 p<,00000 Směrod. chyba odhadu : 2,8915						
N=20	b*	Sm.chyba z b*	b	Sm.chyba z b	t(17)	p-hodn.
Abs.člen			-156,313	37,34335	-4,18582	0,000620
x1	0,474630	0,094553	0,896	0,17849	5,01972	0,000105
x2	0,836417	0,094553	1,059	0,11976	8,84599	0,000000

Mezi těmito dvěma modely rozhodněte na základě reziduální analýzy a adjustovaného indexu determinace.