# Téma 9: Vícenásobná regrese

# 1) Vytvoření modelu

V menu *Statistika* zvolíme nabídku *Vícerozměrná regrese*. Aktivujeme kartu *Detailní nastavení* – viz obr.1. Nastavíme *Proměnné* tak, že v příslušném okně – viz. obr.2, v levém sloupci označíme závisle proměnnou a v pravém sloupci označíme nezávisle proměnné. Při nesouvislém výběru nezávislé proměnných použijeme tlačítko CTRL <u>Kdybychom chtěli</u> vytvořit prostý lineární model, označíme v pravém sloupci pouze jedinou nezávisle proměnnou.

### Obr. 1

📝 Vícerozměrná lineární regrese: auto	?_×
Základní nastavení Detailní nastavení	E OK
🗩 <u>P</u> roměnné	Storno
Závislé: žádné Nazávislá: žádné	🔈 Možnosti 🔻
Vstupní soubor: Data	🗁 O <u>t</u> evři data
Další možnosti (kroková nebo hřeben. regrese)	SELECT f
🔲 Zobrazit popisné statistiky, korelační matici	□ Vážené momentu
Rozšířená přesnost výpočtů	SV =
🗖 Dávkové zpracování/protokol	© W-1 O N-1
Tisk/protokol reziduální analýzy	ChD vynechána
Vyberte všechny proměnné na analýzu. Zaškrtněte "Další možecsti" na krekovou analýzu atd	Celé případy
	C Párově
	C Substituce
Mz také modul Obecné regresní modely (GRM).	prumerem

Vyberte závislé a nezávislé promění	né	<u>? ×</u>
1-výkon 2-cena 3-max.rychlost 4-spotřeba	1-výkon 2-cena 3-max.rychlost 4-spotřeba	OK Storno
Vybrat vše DI. názvy Detaily Závislá prom. (n. seznam pro dávku): 2	Vybrat vše DI. názvy Detaily Seznam nezáv. proměnných: 1 3-4	

Výběr potvrdíme tlačítkem **OK**. Poté zajistíme, aby na kartě **Detailní nastavení** (obr.1) byla zaškrtnuta volba **Další možnosti (kroková nebo hřeben. regrese)**. Stiskneme – li opět tlačítko **OK**, otevře se okno **Definice modelu** – viz. obr.3. Zkontrolujeme, zda je nastavena **Standardní metoda** a zda je **absolutní člen zahrnut v modelu**.

🚀 Definice modelu: auto	? _ ×
Základ Detaily Metoda Popisné	📰 OK
💁 <u>P</u> roměnné	Storno
Závislé: cena Nezávislé: 1 3-4	Možnosti▼
Metoda: Standardní 💌	
Abs. clen: Zahrnut v modelu Tolerance: ,0001 (Pomocí 0.0 nastavíte	
Hřebenová regrese; lamb,10	
🗖 Zpracování/tisk po dávkách	
Tisk/protokol reziduální analýzy	

Stiskneme-li potřetí tlačítko tlačítko *OK*, otevře se okno *Výsledky – vícerozměrná regrese* –viz obr.4.

Obr.4
-------

📈 Výsledky - vícerozměrná regrese: auto	?X
Výsleduv- vícerozy regrese	
Vysredny Vicerozm. regrese	
Záv.prom. : cena vícenás. R = ,90483644 F = 39,14388	
R^2= ,81872899 sv = 3,26 Poč nřínadů: 30 unrav R^2 = 79781310 n = 000000	
Směrodatní chyba odhadu :62812, 307268	
Abs. člen: 116707,40762 Sm. chyba: 198440,5 t( 26) = ,58812 p =	,5615
výkon beta=,833 max.rychlost beta=,328 spotřeba beta=-,	36
(významná beta jsou zvýraznění)	
	Eg ±
Alfa na zvýraznění efektů: 05 🚔	OK
Základní výsledky Detailní výsledky Residua/předpoklady/předpovědi	Storno
Výpočet: Výsledky regrese	Možnosti <del>v</del>

Vybereme kartu *Základní výsledky*. V ní stiskneme tlačítko *Výpočet: výsledky regrese*. Objeví se výstupní sestava - viz obr. 5.

Obr.	5
------	---

	Výsledky regrese se závislou proměnnou : cena (auto) R= ,90483644 R^2= ,81872899 uprav. R^2= ,79781310 F(3,26)=39,144 p<,00000 Směrod. chyba odhadu : 62812,								
	Beta	Beta Sm.chyba B Sm.chyba t(26) Úroveň p							
N=30		beta		В					
Abs.člen			116707,4	198440,4	0,58812	0,561524			
výkon	0,833407	0,215908	4860,1	1259,1	3,86002	0,000673			
max.rychlost	0,328018	0,200519	2225,5	1360,5	1,63584	0,113923			
spotřeba	-0,356197	-0,356197 0,126022 -55252,4 19548,3 -2,82646 0,0089							

V její horní polovině je řada užitečných informací. Pro nás jsou důležité hodnoty

R – totální (vícenásobný) koeficient korelace a  $R^2$  totální (vícenásobný) koeficient determinace.

Chceme-li porovnat totální koeficienty determinace v modelech s různým počtem proměnných, používáme k tomuto účelu nezkreslený odhad totál. koeficientu determinace -  $uprav. R^{2}$ .

Ve výstupní sestavě –viz obr. 5 - jsou pro vytvoření modelu důležité hodnoty, uvedené ve sloupci **B**. Jedná se o koeficienty u proměnných, jejichž název je uveden v prvním sloupci –tj. na obr. 5 se jedná o *Abs.člen*, *výkon, max. rychlost, spotřeba*. Ve výstupní sestavě uvedené na obr. 5 se tedy jedná o model

 $y' = 116707, 4 + 4860, 1x_1 + 2225, 5x_2 - 55252, 4x_3$ 

s vícenásobným koeficientem korelace  $R_{y,x1, x2, x3} = 0,905$  a s vícenásobným koeficientem determinace  $R_{y,x1, x2, x3}^2 = 0,819$ .

Čísla ve sloupci **Úroveň** *p* značí nejmenší hladiny významnosti, pro něž lze zamítnout hypotézy o nulových hodnotách regresních koeficientů, tj. výsledky testů H<sub>0</sub>:  $\beta i = 0$  proti

 $H_A$ : non  $H_0$ , i = 0, 1, 2, 3. (Interpretujeme–li tedy hodnoty p v obr. 5, hypotézu o nulové hodnotě  $\beta_0$  a  $\beta_2$  zamítnout nelze a hypotézu o nulové hodnotě  $\beta_1$  a  $\beta_2$  zamítnout lze.)

Čísla ve sloupci *t(B)* jsou pak hodnoty testovacích kritérií.

Pro nás z toho plyne, že model lze zjednodušit tím, že vypustíme absolutní člen. Aktivujeme lištu *Výsledky-vícerozměrné*. Ta je zobrazena *v levé spodní části obrazovky*. Objeví se okno, uvedené na obr. 4. ( V něm můžeme v kartách *Detailní výsledky* či *Rezidua/předpoklady/předpovědi* zvolit provedení dalších výpočtů. ) V okně, uvedeném na obr. 4, stiskneme tlačítko *Storno*. Vrátíme se tak do okna *Definice modelu* ( obr. 2). V něm v kartě *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 na sestavě zjednodušené –viz obr. 6. Vícenásobný

#### Obr. 6

# STATISTICA Cz

POZN.: R^2 v případě regresní přímky procházející počátkem (bez abs. členu) nelze porovnávat s R^2 v případě, že absolutní člen je zahrnut v modelu.



koeficient determinace je v modelu bez absolutního členu počítán podle jiného vzorce, než v modelu s absolutním členem.

Nová výstupní sestava již absolutní člen nebude mít. Interpretace výsledků je stejná jako interpretace výstupní sestavy uvedené na obr.5.

# 2) Predikce

Predikci umožní provést nastavení karty **Residua/předpoklady/předpovědi** v okně **Výsledky** – vícerozměrná regrese- viz obr. 4. Do tohoto okna se nejrychleji vrátíme stiskem tlačítka Výsledky –vícerozmě., umístěného ve spodní části obrazovky.

Zvolíme-li nabídku *Výpočet interv. spolehlivosti*, aktualizujeme-li hladinu významnosti *Alfa* a zadáme-li hodnotu *nezávisle proměnné*, pro niž chceme predikci provést – viz obr.7, pak vlastní predikci provedeme stiskem tlačítka *OK*. Výstupní sestava je uvedena na obr. 8.



Obr. 8

	Předpovězené hodnoty (Tabulka1) proměnné: cena				
	B-váž.	Hodnota	B-váž.		
Proměnná			* Hodnot		
výkon	5065,7	60,0000	303940		
rychlost	3788,2	190,0000	719751		
spotřeba	-52486,6	6,0000	-314920		
Abs. člen			-252745		
Předpověď			456026		
-95,0%LS			390126		
+95,0%LS			521926		

Obsahuje informace o příspěvcích nezávisle proměnných k odhadu závisle proměnné ( sloupec B-váž.), bodový a intervalový odhad hodnot závisle proměnné (sloupec Bváž.\*Hodnot).

## 3) Parciální korelace

Potřebujeme-li znát charakteristiky parciální korelace, vrátíme se do okna *Výsledky – vícerozměrná regrese*- viz obr. 9. (Do tohoto okna se nejrychleji vrátíme stiskem tlačítka *Výsledky –vícerozmě*., umístěného ve spodní části obrazovky.) Na kartě *Detailní výsledky* zvolíme tlačítko *Parciální korelace*. V sestavě, která se objeví - viz obr. 10, najdeme potřebné informace.



🖌 Výsledky - vícerozměrná regrese: Tabulka1	? _ 🔀
Výsledky- vícerozm. regrese	
Záv.prom. : cena vícenás. R = ,99216682 F = 103 R^2= ,98439500 sv = 3	5,1367 8,5
Poč. případů: 9 uprav. R^2 = ,97503200 p = ,6 Směrodatní chyba odhadu :24113,563955	00062
Abs. člen: -252745,0344 Sm. chyba: 132707,1 t( 5) = -1,908	5 p = ,1152
výkon beta=,808 rychlost beta=,437 spotřeba b	oeta=-,34
(významná beta jsou zvýraznění)	
	<u>e</u> ±
Alfa na zvýraznění efektů: 🛛 💭 🚔	CK OK
Základní výsledky Detailní výsledky Residua/předpoklady/předpovědi	Storno
Výpočet: výsledky regrese Parciální korelace	Možnosti▼
ANDVA (Celk. vhodnost modelu) Redundance	
Koyariance koeficientů Výsledky krokov <u>é</u> regrese	
Současná swejep matice ANDVA upravená pro průměry	

	Proměnné obsažené v rovnici ; ZP: cena (auta)							
	Beta	Beta Parciál. Semipar. Tolerance R^2 t(27) Úroveň p						
Proměnná		korelace	korelace					
výkon	0,681920	0,763478	0,144495	0,044899	0,955101	6,14267	0,000001	
max.rychlost	1,098386	0,582953	0,087697	0,006375	0,993625	3,72811	0,000904	
spotřeba	-0,788903	-0,492409	-0,069152	0,007683	0,992317	-2,93972	0,006657	