

Statistické zpracování dat

4. prezentace

Vícenásobná regresní analýza I

Mgr. Radmila Krkošková, Ph.D.



**SLEZSKÁ
UNIVERZITA**

OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ



- **Vícerozměrná (vícenásobná, mnohorozměrná, mnohonásobná, n-rozměrná, n-násobná) lineární regresní analýza**
- **Populační a výběrová regresní funkce**
- **Přiléhavost regresní nadroviny k datům**
- **Koeficient determinace R^2**
- **Klasický vícerozměrný lineární regresní model**



- **Multikolinearita**
- **Heteroskedasticita (H-S)**
- **Testy H-S a její odstraňování**
- **Autokorelace**

Příklad 1.



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Zajímají nás:

Y - tržby prodejny (např. spotřební elektroniky OK)

v závislosti na:

X_1 - výdaje na reklamu

X_2 - počet kolemjdoucích

X_3 - průměrný plat prodavačů

X_4 – počet konkurenčních prodejen v místě

Příklad 2.



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Zajímá nás:

Y – dětská úmrtnost /v promile/

v závislosti na:

X_1 – gramotnost žen /v procentech/

X_2 – HDP na hlavu /v USD/

X_3 – porodnost /v procentech/

Vícenásobná regresní analýzy



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- Grafické znázornění v dimenzích $m > 2$ - obtížné(?)
- Jediné kritérium = závislá proměnná: Y
- Více prediktorů = nez. prom.: X_1, X_2, \dots, X_m ($m = 2, 3, \dots$)

regresní *nadrovina*:

$$Y = B_0 + B_1 X_1 + B_2 X_2 + \dots + B_m X_m$$

regresní model:

Cíl: nalezení **nejlepších** odhadů regresních koeficientů (Excel)

$$Y_i = B_0 + B_1 X_{i1} + B_2 X_{i2} + \dots + B_m X_{im} + u_i$$

Příklad 1 – řešení v Excelu



<i>Regresní statistika</i>	
Násobné R	0,969
Hodnota spolehlivosti R	0,940
Nastavená hodnota spolehlivosti	0,927
Chyba stř. hodnoty	780,552
Pozorování	25

	<i>Koeficienty</i>	<i>Chyba stř. hodnoty</i>	<i>t stat</i>	<i>Hodnota P</i>
Hranice	1642,641	932,471	1,762	0,093
Poč. kolemjdoucích/hod.	81,899	36,855	2,222	0,038
Velikost prodejny m ²	19,893	8,496	2,342	0,030
Prům. plat prodavačů/měs.	241,002	70,482	3,419	0,003
Přítomnost konkurence v místě	-171,803	399,213	-0,430	0,672

Násobné R	= R - koeficient korelace
Hodnota spolehlivosti R	= R ² - koeficient determinace
Nastavená hodnota spolehlivosti R	= R ² _{adj} - upravený koeficient determinace
Chyba stř. hodnoty	= s ² - směrodatná chyba (odhad směrodatné odchylky náhod. složky)

Příklad 1 – řešení – interpretace výsledků



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Kritérium: Y - tržby z prodeje (v tis.Kč/rok)

Prediktory: X_1 - poč. kolemjdoucích

X_2 - velikost prodejny v m²

X_3 - průměrný plat prodavačů v tis.Kč/měs.

X_4 - přítomnost konkurence (binární)

Regresní rovnice:

$$y = 1642,6 + 81,9x_1 + 19,9x_2 + 241,0x_3 - 171,8x_4$$

Příklad 1 – řešení – interpretace výsledků



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Hypotézy o statistické významnosti regres. koeficientů a R^2 :

H_0 : koeficient = 0

$b_0 = 1642,6$ (p -hodnota = 0,093 $\Rightarrow H_0$ zamítáme)

$b_1 = 81,9$ (p -hodnota = 0,038 $\Rightarrow H_0$ zamítáme)

$b_2 = 19,9$ (p -hodnota = 0,030 $\Rightarrow H_0$ zamítáme)

$b_3 = 241,0$ (p -hodnota = 0,003 $\Rightarrow H_0$ zamítáme)

$b_4 = -171,8$ (p -hodnota = 0,672 $\Rightarrow H_0$ nezamítáme)

Koeficient determinace (přiléhavost): $R^2 = 0,940$

(p -hodnota = 0,005 $\Rightarrow H_0$ zamítáme)

Závěr: Přítomnost konkurence nemá na tržby prodejny vliv. Tržby nové prodejny jsou na základě modelu prognózovány ve výši 10700 tis. Kč.

Předpoklady lineárního regresního modelu



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVÍNĚ

1. Střední hodnota náhodné poruchy u je 0, tj. $E(u) = 0$
2. Náhodná chyba má **normální rozdělení**, tj. $u \sim N(0, \sigma^2)$
3. Vysvětlující proměnné X_1, X_2, \dots, X_m **nejsou kolineární**
(JINAK: **MULTIKOLINEARITA**)
4. Rozptyl náhodné chyby u je konstantní - **homoskedasticita** tj.
$$\text{Var}(u) = \sigma^2$$

(JINAK: **HETEROSKEDASTICITA**)
5. Náhodné chyby u **jsou nekorelované**, tj.
$$\text{Cov}(u_i, u_j) = 0 \text{ pro } i \neq j$$

(JINAK: **AUTOKORELACE**)

Co se může stát, když některý z předpokladů není splněn?

Poznámky:



1. Předpoklady kromě 3. jsou stejné jako v jednoduchém lineárním regresním modelu.
2. Kolinearita znamená, že vysvětlující proměnná je přesnou lineární kombinací některých ostatních vysvětlujících proměnných.

Příklad: $X_{1i} = 2X_{2i} + X_{3i}$ pro všechna $i=1,2,\dots,n$

3. Problém tzv. **multikolinearity** spočívá v tom, že některé vysvětlující proměnné jsou **téměř** kolineární (lin. kombinacemi jiných proměnných).



Děkuji Vám za pozornost!!!