4EK211 Základy ekonometrie

ZS 2016/17 Cvičení 6: Multikolinearita, umělé proměnné

LENKA FIŘTOVÁ

KATEDRA EKONOMETRIE, FAKULTA INFORMATIKY A STATISTIKY

VYSOKÁ ŠKOLA EKONOMICKÁ V PRAZE

1. Multikolinearita

Otevřete si data z minula. Upravte proměnnou price (vydělte 1000).

Data: pizza.wf1

Zdroj: ECON2300, University of Queensland, 2012, upraveno

Co budeme zkoumat: kolik utrácí lidi za pizzu v závislosti na různých faktorech

1. Multikolinearita

Proměnné:

- pizza:	roční útrata za pizzu v dolarech
- zena:	= 1 pro ženy, jinak 0 (umělá proměnná, dummy variable)
- muz:	= 1 pro muže, jinak 0 (umělá proměnná, dummy variable)
- prijem	roční příjem v dolarech
- vek	věk (v letech)
- hranolky	roční útrata za hranolky v dolarech
- hamburgery	roční útrata za hamburgery v dolarech
- salaty	roční útrata za saláty v dolarech

1. Multikolinearita

Budeme zkoumat vliv pohlaví na útratu za pizzu.

Odhadněte model:

$$pizza = \beta_0 + \beta_1 prijem + \beta_2 vek + \beta_3 zena + \beta_4 muz + u$$

V čem je problém? Který G-M předpoklad je porušen? Jakou úpravu modelu byste navrhli?

1. Multikolinearita

Odhadněte následující modely a posuďte, zda jsou proměnné v modelu významné.

$$pizza = \beta_0 + \beta_1 hranolky + u$$

$$pizza = \beta_0 + \beta_1 hranolky + \beta_2 hamburgery + u$$

$$pizza = \beta_0 + \beta_1 hranolky + \beta_2 hamburgery + \beta_3 salaty + u$$

Může zde hrát roli multikolinearita?

1. Multikolinearita

→ jde o lineární závislost vysvětlujících proměnných

→ je pak obtížné poznat, jak každá z vysvětlujících proměnných ovlivňuje vysvětlující proměnnou (poznáme, jak ji ovlivňují dohromady)

→ příčiny:

- Tendence časových řad vyvíjet se stejným směrem
- Průřezová data
- Zpožděné hodnoty proměnných
- Nesprávný počet dummy proměnných - kdy jsme se s tím dnes setkali?

1. Multikolinearita

→ netestujeme ji, nýbrž ji měříme v jednom konkrétním souboru

→ důsledky:

- Odhady jsou nestranné i vydatné, ale…
- Odhady nejsou stabilní, jsou citlivé i na malé změny v matici **X**
- Směrodatné chyby koeficientů jsou velké - proměnná se může jevit jako nevýznamná, i když to nemusí být pravda

1. Multikolinearita

Měření - 2 proměnné:

multikolinearita je v modelu únosná, pokud platí současně:

$|r_{x_1,x_2}| \leq 0,9$

$r_{x_1,x_2}^2 \quad \leq R^2$

Kde r_{x_1,x_2} je párový korelační koeficient mezi dvěma vysvětlujícími proměnnými

R^2 je koeficient determinace z modelu

2. Multikolinearita

Měření - více než 2 proměnné:

Tabulka párových korelačních koeficientů

(Quick → Group Statistics → Correlations)

	HAMBURGE...	HRANOLKY	SALATY
HAMBURGE...	1.000000	0.828942	-0.751122
HRANOLKY	0.828942	1.000000	-0.734738
SALATY	-0.751122	-0.734738	1.000000

Odhalí lineární závislost mezi jednotlivými dvojicemi proměnných. To ale někdy nestačí…

V případě více proměnných používáme pomocné regrese.

1. Multikolinearita

Měření - více než 2 proměnné:

Původní regrese:

$y = f(x_1,x_2,x_3) \rightarrow R^2$

Pomocné regrese:

$x_1 = f(x_2,x_3) \rightarrow R_1^2$

$x_2 = f(x_1,x_3) \rightarrow R_2^2$

$x_3 = f(x_1,x_2) \rightarrow R_3^2$

Jsou-li všechny dílčí koeficienty determinace z pomocných regresí menší než koeficient determinace z původní regrese, je multikolinearita v modelu únosná.

1. Multikolinearita

$$pizza = \beta_0 + \beta_1 hranolky + \beta_2 hamburgery + \beta_3 salaty + u \rightarrow R^2 = 0{,}16$$

$$hranolky = \beta_0 + \beta_1 hamburgery + \beta_2 salaty + u \rightarrow R^2 = 0{,}72$$

$$hamburgery = \beta_0 + \beta_1 hranolky + \beta_2 salaty + u \rightarrow R^2 = 0{,}73$$

$$salaty = \beta_0 + \beta_1 hranolky + \beta_2 hamburgery + u \rightarrow R^2 = 0{,}60$$

Multikolinearita není v modelu únosná.

1. Multikolinearita

→ řešení:

- Získat další pozorování
- Použít jiný model (jiná formulace, vypuštění proměnné), pozor na specifikační chybu
- Transformace pozorování (první diference, podíly)

2. Umělé proměnné

Zkuste nyní odhadnout následující dva modely:

$$(1)\ \ pizza = \ \beta_0 + \beta_1 prijem + \beta_2 zena + u$$

$$(2)\ \ pizza = \ \beta_0 + \beta_1 prijem + \beta_2 (prijem \cdot zena) + u$$

Interpretujte koeficienty a nakreslete v obou případech regresní přímku pro muže a pro ženy.

2. Umělé proměnné

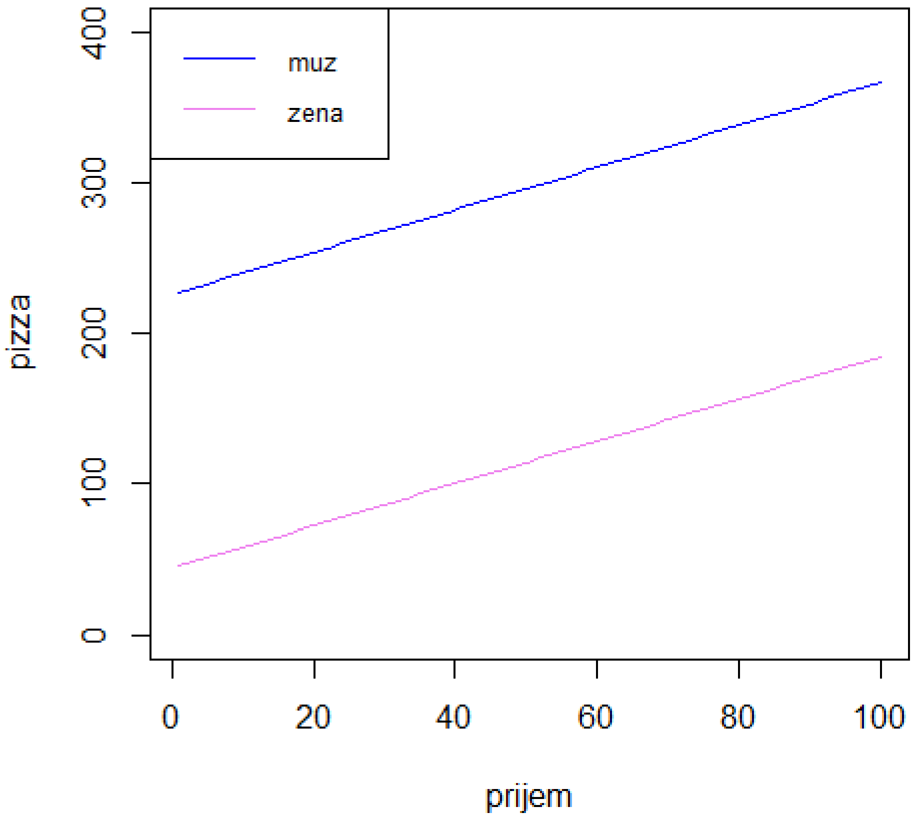
$$(1)\ pizza = \beta_0 + \beta_1 prijem + \beta_2 zena + u$$

$$\widehat{pizza} = 226 + 1{,}41 \cdot prijem - 182 \cdot zena$$

Střední hodnota vysvětlované proměnné:

Muž: $\mathrm{E}(pizza) = 226 + 1{,}41 \cdot prijem$

Žena: $\mathrm{E}(pizza) = 44 \ \ + 1{,}41 \cdot prijem$



2. Umělé proměnné

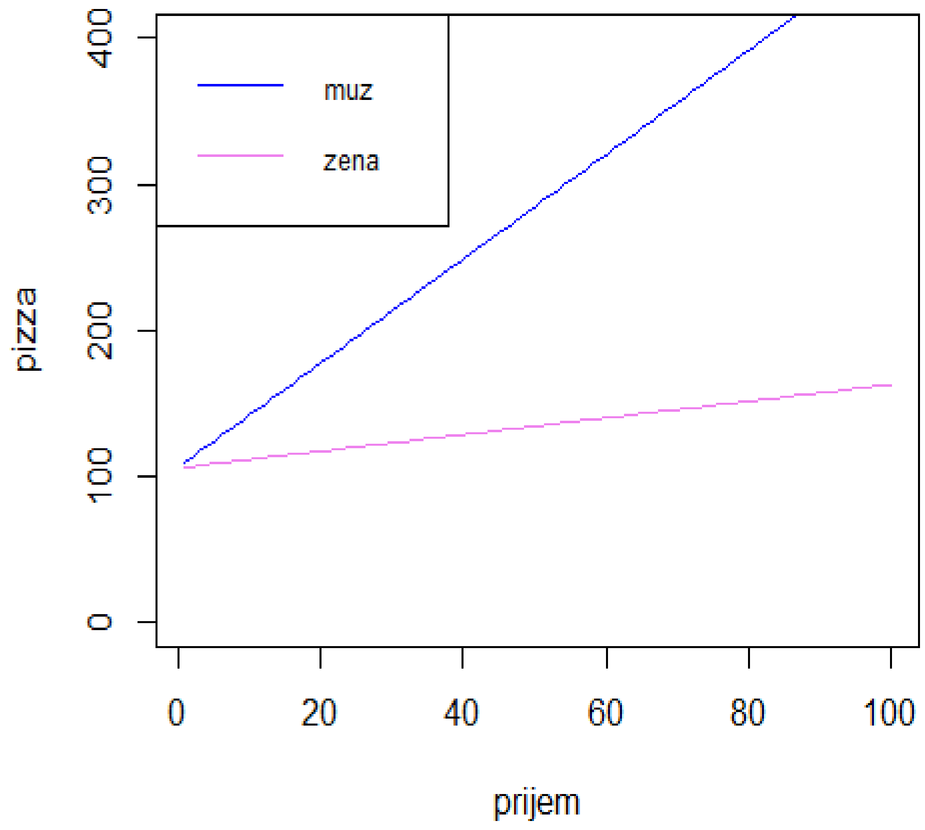
(2) $pizza = \beta_0 + \beta_1 prijem + \beta_2 (prijem \cdot zena) + u$

$\widehat{pizza} = 106 + 3{,}57 \cdot prijem - 3 \cdot prijem \cdot zena$

Střední hodnota vysvětlované proměnné:

Muž: $\mathrm{E}(pizza) = 106 + 3{,}57 \cdot prijem$

Žena: $\mathrm{E}(pizza) = 106 + 0{,}57 \cdot prijem$



2. Umělé proměnné

1. Kdybyste chtěli zkoumat útratu za pizzu v závislosti na tom, zda má člověk základní, střední či vyšší vzdělání, jaká data byste museli nasbírat a jak byste takový model specifikovali?
2. Napadá vás, jak by se mohly použít umělé proměnné při analýze časových řad?

Na doma: Co byste měli umět

1. Co je to multikolinearita, co je její příčinou?
2. Jak se měří multikolinearita v daném výběru?
3. Co je důsledkem multikolinearity?
4. Co jsou umělé proměnné, jak s nimi pracovat?