

# 4EK211 Základy ekonometrie

ZS 2016/17 Cvičení 5: Vícenásobná regrese



---

LENKA FIŘTOVÁ

KATEDRA EKONOMETRIE, FAKULTA INFORMATIKY A STATISTIKY

VYSOKÁ ŠKOLA EKONOMICKÁ V PRAZE

# 1. Jednoduchá regrese opakování

---

**Zdroj:** ECON2300, University of Queensland, Australia, 2012.

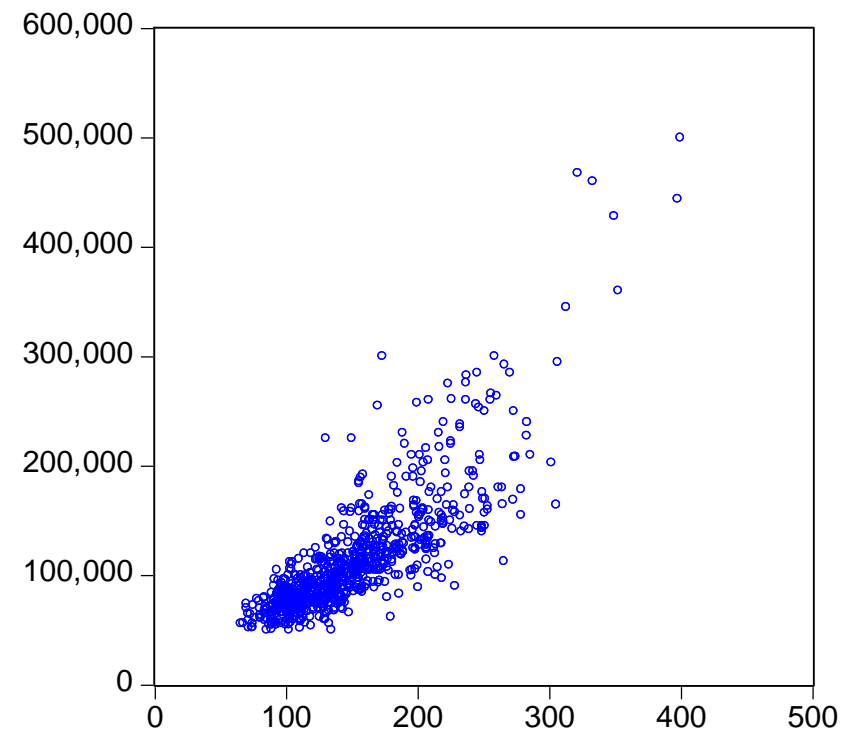
**Data:** domy.wf1

**Zadání:** Zkoumáme závislost ceny domu (v dolarech, proměnná *cena*) na jeho obytném prostoru (v m<sup>2</sup>, proměnná *rozloha*).

# 1. Jednoduchá regrese opakování

---

Vykreslete bodový graf závislosti ceny domů na obytném prostoru.



# 1. Jednoduchá regrese opakování

---

1. Odhadněte regresi:  $cena = \beta_0 + \beta_1 rozloha + u$

2. Interpretujte odhadnutý parametr  $\beta_1$ .

3. Jaký je koeficient determinace?

4. Jakou byste předpověděli cenu domu o rozloze 200 m<sup>2</sup>?

5. Na 5% hladině významnosti otestujte nulovou hypotézu o nevýznamnosti  $\beta_1$ .

# 1. Jednoduchá regrese opakování

---

1. Odhadněte regresi:  $\widehat{cena} = -18385,65 + 876 \cdot rozloha$

2. Interpretujte odhadnutý parametry  $\beta_1$ .

S každým metrem čtverečním vzroste cena domu o 876 dolarů.

3. Jaký je koeficient determinace? 0,67

4. Jakou byste předpověděli cenu domu o rozloze 200 m<sup>2</sup>? 156 814 dolarů

5. Na 5% hladině významnosti otestujte nulovou hypotézu o nevýznamnosti  $\beta_1$ .

$t = \frac{876}{20,65} = 42,4$ . Kritická hodnota: 1,96. Zamítáme nulovou hypotézu.

# 1. Jednoduchá regrese opakování

---

1. Odhadněte regresi:  $\ln(\text{cena}) = \beta_0 + \beta_1 \text{rozloha} + u$

2. Interpretujte odhadnutý parametr  $\beta_1$ .

3. Jaký je koeficient determinace?

4. Jakou byste předpověděli cenu domu o rozloze 200 m<sup>2</sup>?

# 1. Jednoduchá regrese opakování

---

1. Odhadněte regresi:  $\ln(\widehat{cena}) = 10,59 + 0,0064 \cdot rozloha$

Quick  $\rightarrow$  Estimate equation  $\rightarrow$   $\log(cena)$  c rozloha

2. Interpretujte odhadnutý parametr  $\beta_1$ .

S každým metrem čtverečním vzroste cena domu o 0,64 %.

3. Jaký je koeficient determinace? 0,71

4. Jakou byste předpověděli cenu domu o rozloze 200 m<sup>2</sup>?

$\exp(11,87) = 142\,914$  dolarů

# 1. Jednoduchá regrese opakování

---

1. Odhadněte regresi:  $\ln(\text{cena}) = \beta_0 + \beta_1 \ln(\text{rozloha}) + u$

2. Interpretujte odhadnutý parametr  $\beta_1$ .

3. Jaký je koeficient determinace?

4. Jakou byste předpověděli cenu domu o rozloze 200 m<sup>2</sup>?



# 1. Jednoduchá regrese opakování

---

1. Odhadněte regresi:  $\ln(\widehat{cena}) = 6,56 + 1 \cdot \ln(rozloha)$

Quick  $\rightarrow$  Estimate equation  $\rightarrow$  log(cena) c log(rozloha)

2. Interpretujte odhadnutý parametr  $\beta_1$ .

S každým růstem rozlohy o 1 % vzroste cena domu o 1 %.

3. Jaký je koeficient determinace? 0,69

4. Jakou byste předpověděli cenu domu o rozloze 200 m<sup>2</sup>?

$\exp(11,86) = 141\,492$  dolarů

# 1. Jednoduchá regrese opakování

---

Vysvětlovaná proměnná	Vysvětlující proměnná	Interpretace $\beta_1$
$y$	$x$	$\Delta y = \beta_1 \Delta x$
$y$	$\ln(x)$	$\Delta y = (\beta_1 / 100) \% \Delta x$
$\ln(y)$	$x$	$\% \Delta y = (100 \beta_1) \Delta x$
$\ln(y)$	$\ln(x)$	$\% \Delta y = \beta_1 \% \Delta x$

## 2. Vícenásobná regrese příklad 1

---

Data: sleep.wf1

Zdroj: Zouhar, <http://nb.vse.cz/~zouharj/zek.html>

Původní zdroj: model vychází z článku Biddleho a Hamermeshe (1990)

Co budeme zkoumat: Kompenzují lidé delší pracovní dobu zkrácením délky spánku?

## 2. Vícenásobná regrese příklad 1

---

Proměnné:

- *totwrk*: celková doba spánku za týden (v minutách)
- *sleep*: celková doba práce za týden (v minutách)
- *educ*: počet let vzdělání (v letech)
- *age*: věk (v letech)

Regresní přímka:

$$\textit{sleep} = \beta_0 + \beta_1 \textit{totwrk} + \beta_2 \textit{educ} + \beta_3 \textit{age} + u$$

## 2. Vícenásobná regrese příklad 1

---

Regresní přímka:

$$\textit{sleep} = \beta_0 + \beta_1 \textit{totwrk} + \beta_2 \textit{educ} + \beta_3 \textit{age} + u$$

1. Jaká znaménka byste očekávali u koeficientů  $\beta_1, \beta_2, \beta_3$ ?
2. Může u  $b_1$  vyjít jiné znaménko, než jste očekávali, i v případě, že je model správně specifikován a jsou splněny G-M předpoklady?
3. Odhadněte regresní přímku.

## 2. Vícenásobná regrese příklad 1

---

Regresní přímka:

$$\widehat{sleep} = 3638 - 0,15 \cdot totwrk - 11,1 \cdot educ + 2,2 \cdot age$$

Interpretujte odhadnuté koeficienty.

Jak se změní doba spánku, začneme-li pracovat o 10 hodin týdně více?

Kolik hodin spánku denně byste dle modelu předpověděli sobě?

Jaký je koeficient vícenásobné determinace? Připomeňte, co vyjadřuje.

Jaký je korigovaný koeficient vícenásobné determinace? Co to je?

## 2. Vícenásobná regrese příklad 1

---

Regresní přímka:

$$\widehat{sleep} = 3638 - 0,15 \cdot totwrk - 11,1 \cdot educ + 2,2 \cdot age$$

1. Otestujte nulovou hypotézu, že  $\beta_2 = 0$ . Spočítejte 95 % interval spolehlivosti pro  $\beta_2$  a učiňte na základě něj nějaký závěr ohledně testované hypotézy.
2. Otestujte nulovou hypotézu, že  $\beta_2 < 0$ .
3. Otestujte významnost modelu jako celku.

## 2. Vícenásobná regrese příklad 1

---

Regresní přímka:

$$\widehat{sleep} = 3638 - 0,15 \cdot totwrk - 11,1 \cdot educ + 2,2 \cdot age$$

1. Otestujte nulovou hypotézu, že  $\beta_2 = 0$ .

- $H_0: \beta_2 = 0$
- $H_1: \beta_2 \neq 0$
- Testová statistika:  $-1,89$
- Kritická hodnota:  $1,96$
- $|-1,89| < 1,96 \rightarrow$  Nezamítáme nulovou hypotézu.



## 2. Vícenásobná regrese příklad 1

---

Regresní přímka:

$$\widehat{sleep} = 3638 - 0,15 \cdot totwork - 11,1 \cdot edu + 2,2 \cdot age$$

1. 95 % interval spolehlivosti pro  $\beta_2$

- $\langle -11,1 - 5,88 \cdot 1,96; -11,1 + 5,88 \cdot 1,96 \rangle$
- $\langle -22,6; 0,4 \rangle$
- Obsahuje nulu  $\rightarrow$  nezamítáme nulovou hypotézu.

## 2. Vícenásobná regrese příklad 1

---

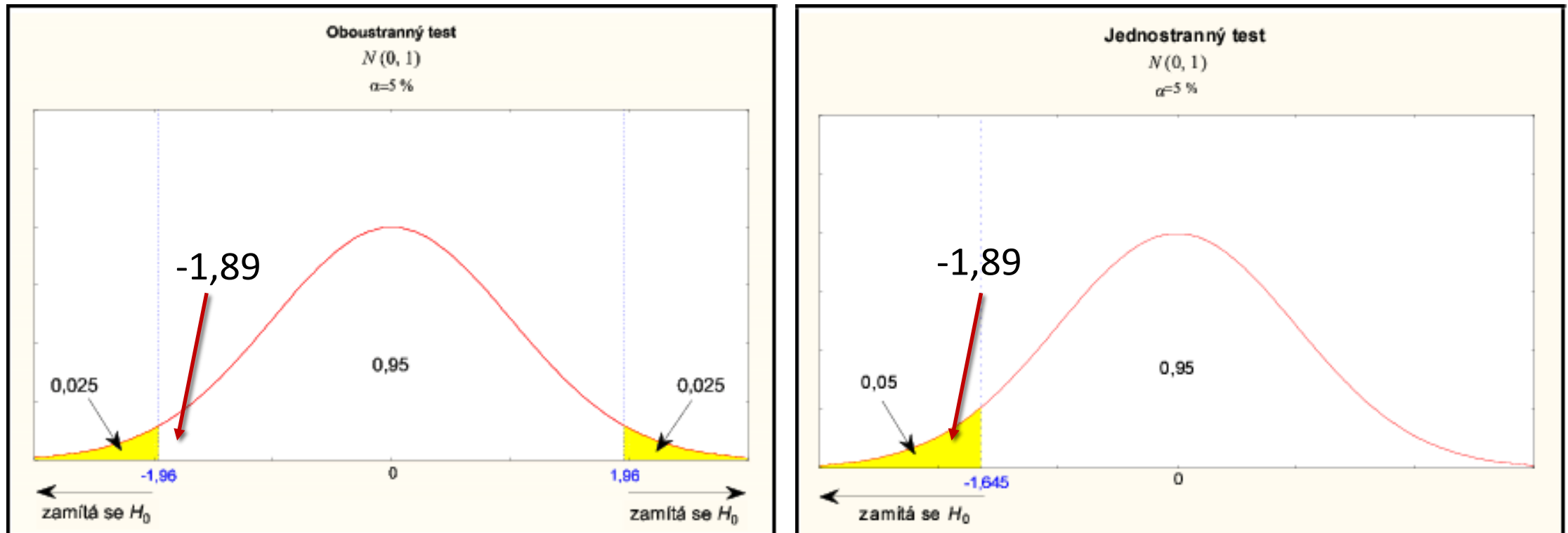
Regresní přímka:

$$\widehat{sleep} = 3638 - 0,15 \cdot totwork - 11,1 \cdot edu + 2,2 \cdot age$$

2. Otestujte nulovou hypotézu, že  $\beta_2 < 0$ .

- $H_0: \beta_2 = 0$
- $H_1: \beta_2 < 0$
- Testová statistika:  $-1,89$
- Kritická hodnota:  $1,64$
- $|-1,89| > 1,64 \rightarrow$  Zamítáme nulovou hypotézu.

## 2. Vícenásobná regrese příklad 1



<http://new.euromise.org/czech/tajne/ucebnice/html/html/node9.html>

## 2. Vícenásobná regrese příklad 1

---

Regresní přímka:

$$\widehat{sleep} = 3638 - 0,15 \cdot totwork - 11,1 \cdot edu + 2,2 \cdot age$$

3. Otestujte významnost modelu jako celku.

- $H_0: \beta_1 = \beta_2 = \beta_3 = 0$
- $H_1: non H_0$
- $F = \frac{R^2}{1-R^2} \frac{n-k-1}{k} = \frac{0,1134}{1-0,1134} \frac{706-3-1}{3} = 29,9$
- Porovnááme s kritickou hodnotou z Fisherova rozdělení:  $F^*(k, n - k - 1)$
- EViews uvádí p-hodnotu.

## 2. Vícenásobná regrese příklad 1

Dependent Variable: SLEEP  
Method: Least Squares  
Date: 10/17/14 Time: 15:13  
Sample: 1 706  
Included observations: 706

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3638.245	112.2751	32.40474	0.0000
TOTWRK	-0.148373	0.016694	-8.888075	0.0000
EDUC	-11.13381	5.884575	-1.892034	0.0589
AGE	2.199885	1.445717	1.521657	0.1285

R-squared	0.113364	Mean dependent var	3266.356
Adjusted R-squared	0.109575	S.D. dependent var	444.4134
S.E. of regression	419.3589	Akaike info criterion	14.92098
Sum squared resid	1.23E+08	Schwarz criterion	14.94681
Log likelihood	-5295.186	Hannan-Quinn criter.	14.93096
F-statistic	29.91889	Durbin-Watson stat	1.942609
Prob(F-statistic)	0.000000		

# 3. Vícenásobná regrese příklad 2

---

Data: pizza.wf1

Zdroj: ECON2300, University of Queensland, 2012, upraveno

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

# 3. Vícenásobná regrese příklad 2

---

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)
- příjem 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

### 3. Vícenásobná regrese příklad 2

---

Upravte proměnnou *prijem* tak, že ji vydělíte 1000.

Odhadněte tři modely:

(a)  $pizza = \beta_0 + \beta_1prijem + u$

(b)  $pizza = \beta_0 + \beta_1prijem + \beta_2vek + u$

(c)  $pizza = \beta_0 + \beta_1prijem + \beta_2vek + \beta_3vek \cdot prijem + u$



### 3. Vícenásobná regrese příklad 2

---

Odhadněte tři modely a vždy řekněte, které proměnné jsou v modelu významné. Interpretujte parametry.

(a)  $\widehat{pizza} = 129 + 1,46 \cdot příjem$

(b)  $\widehat{pizza} = 343 + 2,38 \cdot příjem - 7,58 \cdot vek$

(c)  $\widehat{pizza} = 162 + 9,07 \cdot příjem - 2,98 \cdot vek - 0,16 \cdot vek \cdot příjem$

Jak se ve třetím případě změní útrata za pizzu s 1 rokem věku navíc? Jak se změní s růstem ročního příjmu o 1 tisíc dolarů?

### 3. Vícenásobná regrese příklad 2

---

$$(c) \quad \widehat{pizza} = 162 + 9,07 \cdot příjem - 2,98 \cdot vek - 0,16 \cdot vek \cdot příjem$$

$$\frac{\partial pizza}{\partial příjem} = \beta_1 + \beta_3 \cdot vek$$

$$\frac{\partial pizza}{\partial vek} = \beta_2 + \beta_3 \cdot příjem$$

S rostoucím věkem útrata za pizzu klesá, a to tím více, čím vyšší má daná osoba příjem.

# Na doma: Co byste měli umět

---

1. Jak se interpretují odhadnuté koeficienty, jsou-li proměnné zlogaritmované?
2. Co je to koeficient determinace a korigovaný koeficient determinace?
3. Jak otestujeme významnost modelu jako celku?