

4EK211 Základy ekonometrie

ZS 2016/17 Cvičení 3: Lineární regresní model



LENKA FIŘTOVÁ

KATEDRA EKONOMETRIE, FAKULTA INFORMATIKY A STATISTIKY

VYSOKÁ ŠKOLA EKONOMICKÁ V PRAZE

1. Seznámení s EViews

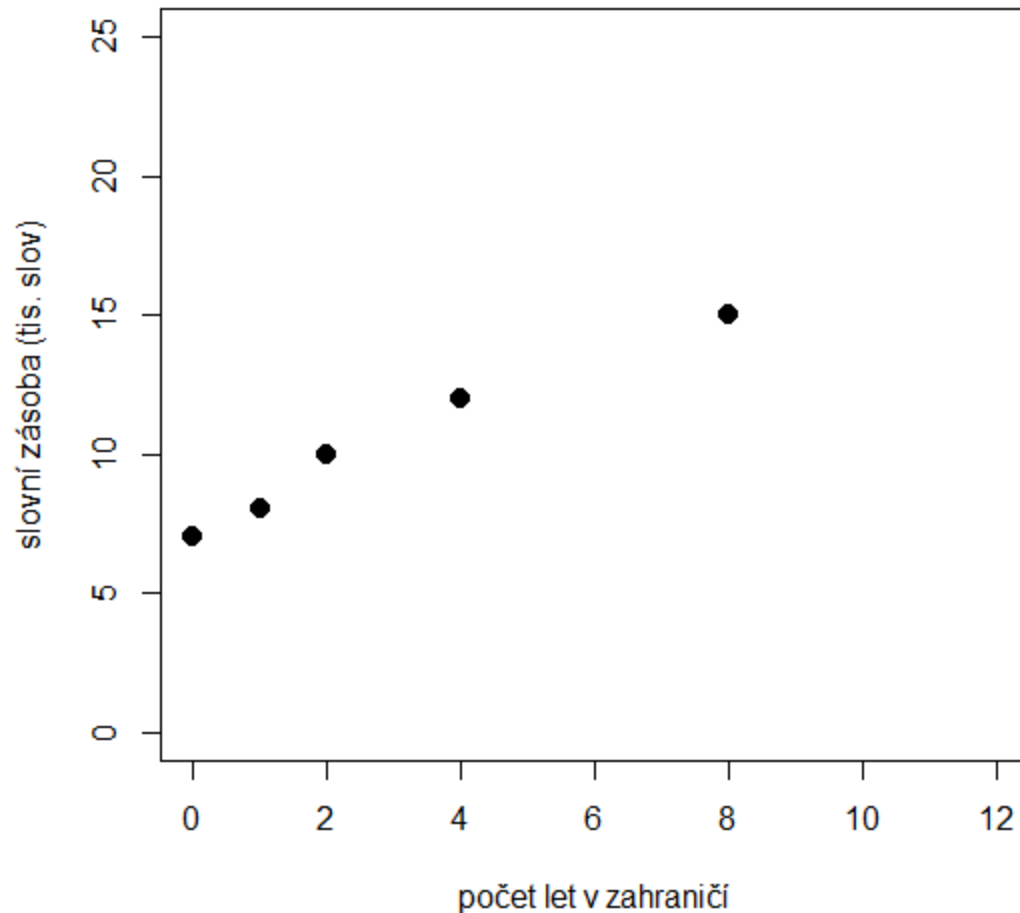
Upřesnění ke značení:

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

→ odhad

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t, \hat{y}_t = b_0 + b_1 x_t$$

1. Seznámení s EViews



DATA Z MINULA

počet let v zahraničí	0	1	2	4	8
slovní zásoba (tis. slov)	7	8	10	12	15

1. Seznámení s EViews

Odhadněte regresi v Eviews.

Quick → Estimate equation → $y = c + x$

Zapište odhadnutou regresní funkci.

Dependent Variable: Y
Method: Least Squares
Date: 09/26/14 Time: 18:22
Sample: 1 5
Included observations: 5

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.400000	0.412311	17.94764	0.0004
X	1.000000	0.100000	10.00000	0.0021
R-squared	0.970874	Mean dependent var	10.40000	
Adjusted R-squared	0.961165	S.D. dependent var	3.209361	
S.E. of regression	0.632456	Akaike info criterion	2.210761	
Sum squared resid	1.200000	Schwarz criterion	2.054536	
Log likelihood	-3.526902	Hannan-Quinn criter.	1.791469	
F-statistic	100.0000	Durbin-Watson stat	1.666667	
Prob(F-statistic)	0.002128			

1. Seznámení s EViews

Ve výstupu najděte:

1) b_0, b_1

2) R^2

3) RSS

Dopočtěte

1) TSS

2) ESS

Dependent Variable: Y
Method: Least Squares
Date: 09/26/14 Time: 18:22
Sample: 1 5
Included observations: 5

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.400000	0.412311	17.94764	0.0004
X	1.000000	0.100000	10.00000	0.0021

R-squared	0.970874	Mean dependent var	10.40000
Adjusted R-squared	0.961165	S.D. dependent var	3.209361
S.E. of regression	0.632456	Akaike info criterion	2.210761
Sum squared resid	1.200000	Schwarz criterion	2.054536
Log likelihood	-3.526902	Hannan-Quinn criter.	1.791469
F-statistic	100.0000	Durbin-Watson stat	1.666667
Prob(F-statistic)	0.002128		

1. Seznámení s EViews

Ve výstupu najděte:

1) b_0, b_1

2) R^2

3) RSS

Dopočtěte

1) TSS

2) ESS

Dependent Variable: Y
Method: Least Squares
Date: 09/26/14 Time: 18:22
Sample: 1 5
Included observations: 5

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.400000	0.412311	17.94764	0.0004
X	1.000000	0.100000	10.00000	0.0021
R-squared	0.970874	Mean dependent var		10.40000
Adjusted R-squared	0.961165	S.D. dependent var		3.209361
S.E. of regression	0.632456	Akaike info criterion		2.210761
Sum squared resid	1.200000	Schwarz criterion		2.054536
Log likelihood	-3.526902	Hannan-Quinn criter.		1.791469
F-statistic	100.0000	Durbin-Watson stat		1.666667
Prob(F-statistic)	0.002128			

1. Seznámení s EViews

1) Jaké jsou hodnoty reziduí?
Porovnejte je s ručním výpočtem.

2) Uložte vyrovnané hodnoty a porovnejte je s ručním výpočtem.
◦ *Forecast*

3) Jaký je výběrový průměr reziduí?
Je to tak vždy?

4) Čemu se rovná korelační koeficient skutečných a vyrovnaných hodnot y ? Jaký je jeho vztah k R^2 ?

Dependent Variable: Y
Method: Least Squares
Date: 09/26/14 Time: 18:22
Sample: 1 5
Included observations: 5

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.400000	0.412311	17.94764	0.0004
X	1.000000	0.100000	10.00000	0.0021
R-squared	0.970874	Mean dependent var	10.40000	
Adjusted R-squared	0.961165	S.D. dependent var	3.209361	
S.E. of regression	0.632456	Akaike info criterion	2.210761	
Sum squared resid	1.200000	Schwarz criterion	2.054536	
Log likelihood	-3.526902	Hannan-Quinn criter.	1.791469	
F-statistic	100.0000	Durbin-Watson stat	1.666667	
Prob(F-statistic)	0.002128			

1. Seznámení s EViews

Otestujte, zda se β_1 statisticky významně liší od nuly na 5% hladině významnosti.

Definujte nulovou hypotézu.

Spočítejte testovou statistiku.

Porovnejte ji s kritickými hodnotami t -rozdělení s $n - k - 1$ stupni volnosti.

Zjistěte, jaká je p -hodnota.
Porovnejte s výstupem s EViews.

Dependent Variable: Y
Method: Least Squares
Date: 09/26/14 Time: 18:22
Sample: 1 5
Included observations: 5

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.400000	0.412311	17.94764	0.0004
X	1.000000	0.100000	10.00000	0.0021
R-squared	0.970874	Mean dependent var	10.40000	
Adjusted R-squared	0.961165	S.D. dependent var	3.209361	
S.E. of regression	0.632456	Akaike info criterion	2.210761	
Sum squared resid	1.200000	Schwarz criterion	2.054536	
Log likelihood	-3.526902	Hannan-Quinn criter.	1.791469	
F-statistic	100.0000	Durbin-Watson stat	1.666667	
Prob(F-statistic)	0.002128			

1. Seznámení s EViews

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

$$t_0 = \frac{b_1 - \beta_1}{s_{b_1}} = \frac{1 - 0}{0,1} = 10$$

Kritická hodnota: 3,18

p-hodnota: 0,0021

Dependent Variable: Y
Method: Least Squares
Date: 09/26/14 Time: 18:22
Sample: 1 5
Included observations: 5

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.400000	0.412311	17.94764	0.0004
X	1.000000	0.100000	10.00000	0.0021
R-squared	0.970874	Mean dependent var		10.40000
Adjusted R-squared	0.961165	S.D. dependent var		3.209361
S.E. of regression	0.632456	Akaike info criterion		2.210761
Sum squared resid	1.200000	Schwarz criterion		2.054536
Log likelihood	-3.526902	Hannan-Quinn criter.		1.791469
F-statistic	100.0000	Durbin-Watson stat		1.666667
Prob(F-statistic)	0.002128			

1. Seznámení s EViews

Spočítejte 95% interval
spolehlivosti pro parametr β_1 .

Dependent Variable: Y
Method: Least Squares
Date: 09/26/14 Time: 18:22
Sample: 1 5
Included observations: 5

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.400000	0.412311	17.94764	0.0004
X	1.000000	0.100000	10.00000	0.0021
R-squared	0.970874	Mean dependent var	10.40000	
Adjusted R-squared	0.961165	S.D. dependent var	3.209361	
S.E. of regression	0.632456	Akaike info criterion	2.210761	
Sum squared resid	1.200000	Schwarz criterion	2.054536	
Log likelihood	-3.526902	Hannan-Quinn criter.	1.791469	
F-statistic	100.0000	Durbin-Watson stat	1.666667	
Prob(F-statistic)	0.002128			

1. Seznámení s EViews

Spočítejte 95% interval spolehlivosti pro parametr β_1 .

$$\langle b_1 - s_{b_1} t^*_{0,025}; b_1 + s_{b_1} t^*_{0,025} \rangle$$

$$\langle 1 - 0,1 \cdot 3,18; 1 + 0,1 \cdot 3,18 \rangle$$

$$\langle 0,682; 1,318 \rangle$$

Jaký závěr byste učinili v předchozím testu hypotézy na základě znalosti tohoto intervalu?

Dependent Variable: Y
Method: Least Squares
Date: 09/26/14 Time: 18:22
Sample: 1 5
Included observations: 5

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.400000	0.412311	17.94764	0.0004
X	1.000000	0.100000	10.00000	0.0021
R-squared	0.970874	Mean dependent var	10.40000	
Adjusted R-squared	0.961165	S.D. dependent var	3.209361	
S.E. of regression	0.632456	Akaike info criterion	2.210761	
Sum squared resid	1.200000	Schwarz criterion	2.054536	
Log likelihood	-3.526902	Hannan-Quinn criter.	1.791469	
F-statistic	100.0000	Durbin-Watson stat	1.666667	
Prob(F-statistic)	0.002128			

2. EViews: příklad 2

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

Data: salary.wf1

Zadání: Zkoumáme závislost výše měsíčního příjmu vyučujících na vysoké škole v dolarech v závislosti na tom, před kolika lety získali titul PhD.

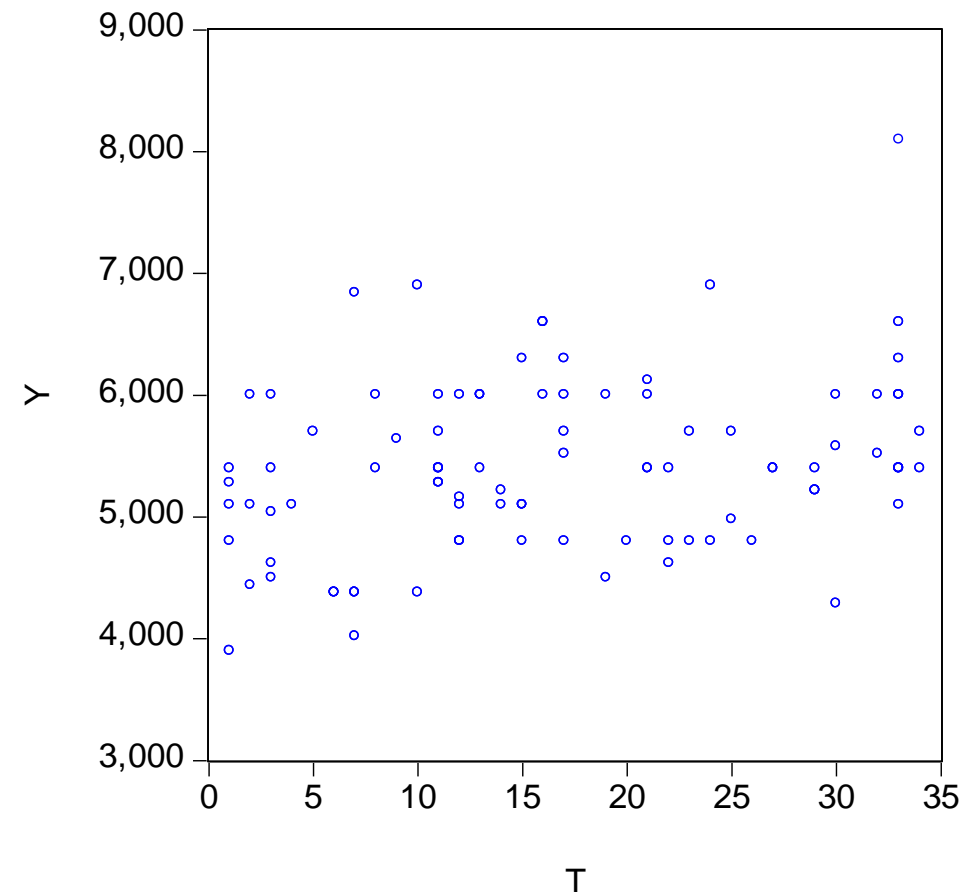
2. EViews: příklad 2

1. Nakreslete graf závislosti příjmu na počtu let od získání titulu.

2. EViews: příklad 2

1. Nakreslete graf závislosti příjmu na počtu let od získání titulu.

- Quick → Graph → t y → Scatter



2. EViews: příklad 2

2. Odhadněte a zapište regresní přímku.

Interpretujte odhadnuté koeficienty.

2. EViews: příklad 2

2. Odhadněte a zapište regresní přímku.
Interpretujte odhadnuté koeficienty.

■ Quick → Estimate Equation → y c t

■ $\hat{y} = 5089 + 19,78t$

Dependent Variable: Y				
Method: Least Squares				
Date: 10/05/14 Time: 19:39				
Sample: 1 93				
Included observations: 93				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5089.606	136.1416	37.38465	0.0000
T	19.77919	6.951048	2.845498	0.0055
R-squared	0.081707	Mean dependent var		5420.323
Adjusted R-squared	0.071615	S.D. dependent var		709.5872
S.E. of regression	683.7066	Akaike info criterion		15.91421
Sum squared resid	42538376	Schwarz criterion		15.96867
Log likelihood	-738.0106	Hannan-Quinn criter.		15.93620
F-statistic	8.096859	Durbin-Watson stat		0.265588
Prob(F-statistic)	0.005478			

2. EViews: příklad 2

3. Otestujte, zda se β_1 statisticky významně liší od nuly na 5% hladině významnosti.

- Definujte nulovou hypotézu.
- Spočítejte testovou statistiku.
- Porovnejte ji s kritickými hodnotami t -rozdělení s $n - k - 1$ stupni volnosti.
- Zjistěte, jaká je p -hodnota. Porovnejte s výstupem s Eviews.

Dependent Variable: Y				
Method: Least Squares				
Date: 10/05/14 Time: 19:39				
Sample: 1 93				
Included observations: 93				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5089.606	136.1416	37.38465	0.0000
T	19.77919	6.951048	2.845498	0.0055
R-squared	0.081707	Mean dependent var	5420.323	
Adjusted R-squared	0.071615	S.D. dependent var	709.5872	
S.E. of regression	683.7066	Akaike info criterion	15.91421	
Sum squared resid	42538376	Schwarz criterion	15.96867	
Log likelihood	-738.0106	Hannan-Quinn criter.	15.93620	
F-statistic	8.096859	Durbin-Watson stat	0.265588	
Prob(F-statistic)	0.005478			

2. EViews: příklad 2

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

$$t_1 = \frac{b_1 - \beta_1}{s_{b_1}} = \frac{19,78 - 0}{6,95} = 2,84$$

t -rozdělení s 91 stupni volnosti

kritická hodnota: 1,98638

p -hodnota: 0,0055

Dependent Variable: Y				
Method: Least Squares				
Date: 10/05/14 Time: 19:39				
Sample: 1 93				
Included observations: 93				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5089.606	136.1416	37.38465	0.0000
T	19.77919	6.951048	2.845498	0.0055
R-squared	0.081707	Mean dependent var		5420.323
Adjusted R-squared	0.071615	S.D. dependent var		709.5872
S.E. of regression	683.7066	Akaike info criterion		15.91421
Sum squared resid	42538376	Schwarz criterion		15.96867
Log likelihood	-738.0106	Hannan-Quinn criter.		15.93620
F-statistic	8.096859	Durbin-Watson stat		0.265588
Prob(F-statistic)	0.005478			

2. EViews: příklad 2

4. Spočítejte 95% interval spolehlivosti pro parametr β_1 .

Dependent Variable: Y				
Method: Least Squares				
Date: 10/05/14 Time: 19:39				
Sample: 1 93				
Included observation: 93				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5089.606	136.1416	37.38465	0.0000
T	19.77919	6.951048	2.845498	0.0055
R-squared	0.081707	Mean dependent var		5420.323
Adjusted R-squared	0.071615	S.D. dependent var		709.5872
S.E. of regression	683.7066	Akaike info criterion		15.91421
Sum squared resid	42538376	Schwarz criterion		15.96867
Log likelihood	-738.0106	Hannan-Quinn criter.		15.93620
F-statistic	8.096859	Durbin-Watson stat		0.265588
Prob(F-statistic)	0.005478			

2. EViews: příklad 2

4. Spočítejte 95% interval spolehlivosti pro parametr β_1 .

$$\langle b_1 - s_{b_1} t^*_{0,025}; b_1 + s_{b_1} t^*_{0,025} \rangle$$

$$\langle 19,78 - 6,95 \cdot 1,98; 19,78 + 6,95 \cdot 1,98 \rangle$$

$$\langle 6,02; 33,5 \rangle$$

Dependent Variable: Y				
Method: Least Squares				
Date: 10/05/14 Time: 19:39				
Sample: 1 93				
Included observation: 93				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5089.606	136.1416	37.38465	0.0000
T	19.77919	6.951048	2.845498	0.0055
R-squared	0.081707	Mean dependent var	5420.323	
Adjusted R-squared	0.071615	S.D. dependent var	709.5872	
S.E. of regression	683.7066	Akaike info criterion	15.91421	
Sum squared resid	42538376	Schwarz criterion	15.96867	
Log likelihood	-738.0106	Hannan-Quinn criter.	15.93620	
F-statistic	8.096859	Durbin-Watson stat	0.265588	
Prob(F-statistic)	0.005478			

3. Změny jednotek

Uvažujte odhadnutou regresní přímku:

$$\hat{y} = 10,52 + 0,52x$$

Zapište, jak by vypadala přímka v případě, že:

- a) bychom před odhadem všechny hodnoty x vydělili 10
- b) bychom před odhadem všechny hodnoty y vydělili 10
- c) bychom před odhadem všechny hodnoty x a y vydělili 10

Jak by se v každém případě změnil koeficient determinace?

3. Změny jednotek

Uvažujte odhadnutou regresní přímku:

$$\hat{y} = 10,52 + 0,52x$$

Zapište, jak by vypadala přímka v případě, že:

a) bychom před odhadem všechny hodnoty x vydělili 10

$$\hat{y} = 10,52 + 5,2x^*$$

3. Změny jednotek

Uvažujte odhadnutou regresní přímku:

$$\hat{y} = 10,52 + 0,52x$$

Zapište, jak by vypadala přímka v případě, že:

b) bychom před odhadem všechny hodnoty y vydělili 10

$$\hat{y}^* = 1,052 + 0,052x$$

3. Změny jednotek

Uvažujte odhadnutou regresní přímku:

$$\hat{y} = 10,52 + 0,52x$$

Zapište, jak by vypadala přímka v případě, že:

c) bychom všechny hodnoty x a y vydělili 10

$$\hat{y}^* = 1,052 + 0,52x^*$$

Jak by se v každém případě změnil koeficient determinace?

Zůstal by nezměněn.

Na doma: Co byste měli umět

1. Jak odhadnout regresi v Eviews?
2. Jak v Eviews vykreslit data do grafu?
3. Co je to směrodatná chyba odhadnutého koeficientu?
4. Jaký je vztah mezi TSS, RSS, ESS, R^2 ?
5. Jak otestovat, zda je proměnná v modelu významná?
6. Jak zkonstruovat interval spolehlivosti pro odhadnuté regresní koeficienty?
7. Jak se mění regresní koeficienty a R^2 se změnou jednotek proměnných?