ECONOMIC STATISTICS

Sponsored by a Grant TÁMOP-4.1.2-08/2/A/KMR-2009-0041

Course Material Developed by Department of Economics,

Faculty of Social Sciences, Eötvös Loránd University Budapest (ELTE)

Department of Economics, Eötvös Loránd University Budapest

Institute of Economics, Hungarian Academy of Sciences

Balassi Kiadó, Budapest

Author: Anikó Bíró
Supervised by Anikó Bíró
June 2010

Week 6

Multiple regression

Discussion of the 1st exam

Several explanatory variables – examples

- County level unemployment: number of enterprises, geographical location, average education level…
- Sales: advertisement expenditures, hours worked, quality of the product…
- Real estate prices: lot size, number of rooms, location…

Estimation, interpretation

- Regression with k regressors:

$$Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + e_i$$

$$SSR = \sum_{i=1}^{N} (Y_i - \hat{\alpha} - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} - \ldots - \hat{\beta}_k X_{ik})^2$$

- OLS: minimal sum of squared residuals
- Interpretation of the coefficients:
  - Marginal effect
  - Other explanatory variables held constant
  - ”Ceteris paribus”

Hypothesis testing

- Confidence interval: analogously to univariate case
- Significance of coefficients: t-test, p-value
- R^2 = 1-SSR/TSS
  - Measure of fit
  - What % of the variation of the dependent variable is explained by the explanatory variables
  - Testing R^2=0: F-test

$$F = \frac{(N-k-1)R^2}{1-R^2}$$

Example 1: earnings

Wage tariff subsample, 2003 (monthly gr. earnings – age – education year)

Regression statistics	
r-squared	0,46

ANALYSIS OF VARIANCIE

	df	*SS*	*MS*	*F*	*F sign.*
Regression	2	2,87E+13	1,44E+13	2170,7	0
Residual	4997	3,30E+13	6,61E+09		
Total	4999	6,18E+13			

	Coeff.	*St. dev.*	*t stat.*	*p-value*	*Bottom 95%*	*Top 95%*
Intercept	-328321,34	8040,13	-40,84	0,00	-344083,52	-312559,16
Education	27250,22	452,97	60,16	0,00	26362,20	28138,24
Age	3171,29	109,05	29,08	0,00	2957,52	3385,07

- Coefficient interpretations: marginal effect!
- Incorrect: ”older people generally earn more”!

Example 2: housing prices

Housing prices (CAD) – lot size (sq. foot) – number of bedrooms, bathrooms, stories (source: Koop)

Regression statistics	
r-squared	0,54

ANALYSIS OF VARIANCE

	df	SS	MS	F	F sign.
Regression	4	2,08E+11	5,2E+10	155,95	0,00
Residual	541	1,80E+11	3,34E+08		
Total	545	3,89E+11			

	Coeff.	St. dev.	t stat.	p-value	Bottom 95%	Top 95%
Intercept	-4009,55	3603,11	-1,11	0,27	-11087,35	3068,25
Lot size	5,43	0,37	14,70	0,00	4,70	6,15
#bedrooms	2824,61	1214,81	2,33	0,02	438,30	5210,93
#bathrooms	17105,17	1734,43	9,86	0,00	13698,12	20512,22
#stories	7634,90	1007,97	7,57	0,00	5654,87	9614,92

Multiple regression

Seminar 6

OLS estimation

- Regression with k regressors:

$$Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + e_i$$

$$SSR = \sum_{i=1}^{N} (Y_i - \hat{\alpha} - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} - \ldots - \hat{\beta}_k X_{ik})^2$$

- OLS: SSR → min

Example 1

105 countries: 1960–85 average GDP growth rate, average investment/GDP, average population growth rate

- Data in %
- Interpretation of coefficients? (percentage points)

Example 2

- Electricity companies (Koop, electric.xls)
- Dependent variable: production cost
- Explanatory variables: output, unit costs: labor, capital, heating material

- Estimation in logarithmic form
  - Coefficients: elasticity

Simulation with Excel

- Y=a+bX
  - Regression: estimated = true
- Random number generation: e~N(0,1)
- Y=a+bX+e
  - Regression: estimated ≠ true
- Increasing the sample size?
- Increasing the standard deviation of the error term?