Introduction to econometrics

I. Introduction to econometrics and working with data

Content

1. Course organization

2. Introduction to econometrics
   - Economic model
   - Econometric model

3. Working with data
   - Types of economic data
   - Working with data – graphical methods
   - Descriptive statistics and correlation

4. Exercises

Content

Contact informations

- E-mail: nemecd@econ.muni.cz
- Office hours: Wed. 13:00–15:00 and by appointment.
- Department of Economics, room 612.

Course books and software

- Main book: Gary Koop (2008), „Introduction to Econometrics“.
- Other econometric books available (cover similar topics) – Hill et al. (nice exercises), Wooldridge, Gujarati, Dougherty etc.
- Main software: GRETL (other possibilitie available – e.g. Matlab).

Software review

Platform – Name	Developer
Open source	
gretl	„By econometricians, for econometricians“
JMulTi	Benkwitz, Krätzig
Octave	University of Wisconsin
R/Rmetrics	Free Software Foundation, Inc.
Commercial	
EViews	QMS Software, Inc.
GAUSS	Aptech Systems, Inc.
LIMDEP	Econometric Software, Inc.
Matlab	MathWorks, Inc.
RATS	Estima
SAS	SAS Institute
SHAZAM	Northwest Econometrics, Ltd.
SPSS	SPSS, Inc.
Stata	StataCorp

Literature review

Level – Authors	Title
For beginners	
Koop (2009)	Analysis of Economic Data
Introductory	
Koop (2008)	Introduction to Econometrics
Hill, Griffiths, Lim (2008)	Principless of Econometrics
Stock, Watson (2007)	Introduction to Econometrics
Stock, Watson (2008)	Introduction to Econometrics
Wooldridge (2009)	Introductory Econometrics (A Modern Approach)
Intermediate	
Brooks (2008)	Introductory Econometrics for Finance
Dougherty (2007)	Introduction to Econometrics
Enders (2005)	Applied Econometric Time Series
Gujarati, Porter (2009)	Basic Econometrics
Kennedy (2008)	A Guide to Econometrics
Verbeek (2008)	A Guide to Modern Econometrics
Advanced	
Baltagi (2008)	Econometric Analysis of Panel Data
Hayashi (2001)	Econometrics
Greene (2008)	Econometric Analysis
Davidson, Russel, MacKinnon (2004)	Econometric Theory and Methods

Course grading

- 50 % – homeworks (mostly computer exercises).
- 50 % – final (written) exam (including econometric theory and interpreting estimation results).

Final grading

Grade	Points
A	86–100
B	80–85
C	73–79
D	67–72
E	60–66
F	0–59

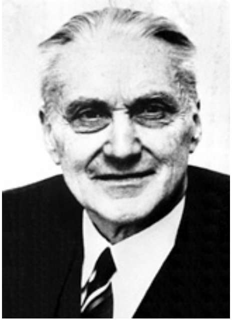
Content

What is econometrics?

- „Econometrics is what the econometricians do."
- „Econometrics is the study of the application of statistical methods to the analysis of economic phenomena."
- „Econometrics is based upon the development of statistical methods for estimating economic relationships, testing economic theories, and evaluating and implementing government and business policy." Wooldridge (2009).
- The 1930s: foundation of Econometric Society (journal *Econometrica*).
- Ragnar Frisch (1933) explains in the first issue of Econometrica: „. . . it is the unification of statistics, economic theory and mathematics that constitutes econometrics."

„Nobel Prize“ in economics

- 1969 – first „Nobel Prize“ in economics.
- „for having developed and applied dynamic models for the analysis of economic processes.“



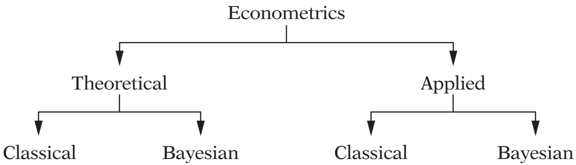
Ragnar Frisch (1895–1973)

Jan Tinbergen (1903–1994)

Who are econometricians?

- Economists – utilizing economic theory to improve their empirical analyses of the problems they address.
- Mathematicians – formulating economic theory in ways that make it appropriate for statistical testing.
- Accountants – finding and collecting economic data and relating theoretical economic variables to observable ones.
- Applied statisticians – trying to estimate economic relationships or predict economic events.
- Theoretical statisticians – applying their skills to the development of statistical techniques appropriate to the empirical problems characterizing the science of economics.

Categories of econometrics



- Classical – mainstream.
- Bayesian – rising importance and attractiveness in economic applications.

Why to use econometrics?

- Econometrics shows us how to use data in a sensible ans systematic manner to shed light on economic questions.
    - Testing whether financial markets are weak-form informationally efficient.
    - Measuring and forecasting the volatility of bond returns.
    - Explaining the determinants of bond credit ratings used by the ratings agencies.
    - Modelling long-term relationships between prices and exchange rates.
    - Forecasting the correlation between the stock indices of two countries.
    - Examining the effects of job training on worker productivity.
    - Estimating the effect of the minimum wage on unemployment.
    - Estimating the effect of law enforcement on city crime levels. Does the presence of more police officers on the street deter crime? How much?
    - Investigating why some people choose to travel to work by car and others choose to travel by public transport.
    - Estimating the effect of advertising on sales.
    - Estimating factors which determine consumer behaviour.

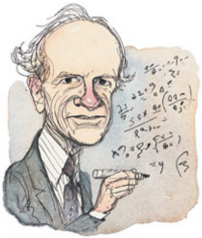
Content

Economic models – examples

- Empirical analysis – using data to test theories or to estimate relationships among variables of interest.
- Economic model $\rightarrow$ to test economic theories.
- Examples:
    - Economic model of crime.
    - Job training and worker productivity.
- Formal economic model = a background for econometric analysis.

Economic model of crime – introduction

- Gary Becker – Nobel prize winner, model desribing individual's participation in crime (1968).



Gary S. Becker (*1930)

- Utility maximization framework – costs and rewards of criminal activities.
- Model describing the amount of time spent in criminal activity as a function of various factors.

Economic model of crime

$$y = f(x_1, x_2, x_3, x_4, x_5, x_6, x_7)$$

$$\begin{aligned}
y &= \text{hours spent in criminal activities,} \\
x_1 &= \text{„wage“ for an hour spent in criminal activity,} \\
x_2 &= \text{hourly wage in legal employment,} \\
x_3 &= \text{income other than from crime or employment,} \\
x_4 &= \text{probability of getting caught,} \\
x_5 &= \text{probability of being convicted if caught,} \\
x_6 &= \text{expected sentence if convicted,} \\
x_7 &= \text{age.}
\end{aligned}$$

Economic model of crime – specification

- Representative list of the main factors affecting a person's decisionto participate in crime.
- Function $f(\cdot)$ not specified (depends on an underlying utility function, rarely known).
- Using economic theory to predict the effect that each variable would have on = the basis for an econometric analysis of individual criminal activity.

Job training and worker productivity – introduction

- Effectiveness of a publicly funded job training program (teaching computer use in the manufacturing process, the twenty-week programme, any worker may participate manufacturing worker may participate, enrollment in all or part of the program is voluntary).
- To examine the effects of job training on worker productivity (represented by hourly wage).
- Little need for formal economic theory.
- Basic economic understanding sufficient $\rightarrow$ factors such as education, experience, and training affect worker productivity.
- Reasonable belief – workers are paid commensurate with their productivity.

Job training and worker productivity – model

$$wage = f(educ, exper, training)$$

$$\begin{aligned} wage &= \text{hourly wage}, \\ educ &= \text{years of formal education}, \\ exper &= \text{years of workforce experience}, \\ training &= \text{weeks spent in job training}. \end{aligned}$$

- Other factors might generally affect the wage rate $\times$ model captures the essence of the problem.

Content

Economic and econometric model

- Specification of economic model $\Rightarrow$ econometric model.
- Specification of relationships among variables, $f(\cdot)$.
- Solving the problem with „unobservable“ variables, e.g.:
  - „wage“ that an individual can earn from criminal activity – well defined variable $\times$ difficult to observe;
  - probability of being arrested – for a given individual canot be obtained $\times$ observable arrest statistics $\rightarrow$ „proxy“ variable.

Econometric model of crime

$$
\begin{aligned}
crime = \alpha + \beta_1 wage_m + \beta_2 othinc + \beta_3 freqarr + \beta_4 freqconv \\
+ \beta_5 avgsen + \beta_6 age + \epsilon
\end{aligned}
$$

$$
\begin{aligned}
crime &= \text{some measure of the frequency of criminal activity,,} \\
wage_m &= \text{wage that can be earned in legal employment,} \\
othinc &= \text{income from other sources (assets etc.),} \\
freqarr &= \text{frequency of arrests for prior infractions} \\
&\quad \text{(aprroximated probability of arrest),} \\
freqconv &= \text{frequency of conviction,} \\
avgsen &= \text{average sentence length after conviction,} \\
age &= \text{age.}
\end{aligned}
$$

Econometric model of crime – explanation

- Choice of variables determined by the economic theory as well as data considerations.
- error term: $\epsilon =$ unobserved factors („wage“ for criminal activity, moral character, family background) and errors in measuring things (criminal activity, probability of arrest).
- parameters: $\alpha, \beta_1, \ldots, \beta_6 =$ directions and strengths of the relationship between explained variable, *crime*, and explaining variables, factors used to determine *crime* in model.

Econometric model of job training and worker productivity

$$wage = \alpha + \beta_1 educ + \beta_2 exper + \beta_3 training + \epsilon$$

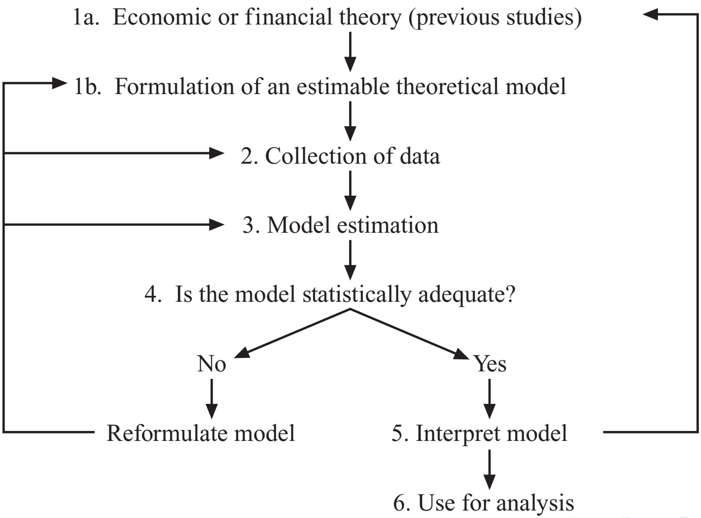
$$\begin{aligned} wage &= \text{hourly wage}, \\ educ &= \text{years of formal education}, \\ exper &= \text{years of workforce experience}, \\ training &= \text{weeks spent in job training}. \end{aligned}$$

- ϵ = factors such as quality of education, family background and other factors influencing person's wage.
- β_3 = parameter measuring effects of job training on wage.

Other issues

- Econometric analysis begins by specifying an econometric models without consideration of the details of the model's creation (careful derivation might be very often difficult.
- Possible use of economic reasoning and common sense as guides for choosing the variables.
- This approach loses some of the richness of economic analysis $\times$ commonly and effectively applied by careful researchers.
- Stating hypothesis in terms of the unknown parameters (e.g. wage that can be earned in legal employment, $wage_m$, has no effect on criminal behaviour $\Leftrightarrow \beta_1 = 0$).

Econometric modelling



Steps in econometric modelling

1. General statement of the problem – formulation of a theoretical model, or intuition from economic theory that two or more variables should be related to one another in a certain way.
2. Collection of data – external or internal (our own surveys) sources.
3. Choice of estimation method – single equation or multiple equation technique?.
4. Statistical evaluation of the model – What assumptions were required to estimate the parameters of the model optimally? Were these assumptions satisfied by the data or the model? Does the model adequately describe the data?
5. Evaluation of the model from a theoretical perspective – Are the parameter estimates of the sizes and signs that the theory or suggested?.
6. Use of model – testing the theory, formulating forecasts or suggested courses of action (e.g. „if inflation and GDP rise, buy stocks in sector X“), as an input to government policy.

Content

Content

Times series

- Data collected at specific points in time.
- Observations on variable at time (for an individual – country, firm, household, etc.).

$$Y_t \quad t = 1, \ldots, T$$

- Observations are not independent (working with times series often requires special tools).
- Many frequencies.
- Seasonal adjustment – if necessary.

Time series – example

Tabulka: Minimal wage, unemployment and other data for Puerto Rico

Obsno	Year	Avgmin	Avgcov	Unemp	GNP
1	1950	0.20	20.1	15.4	878.7
2	1951	0.21	20.7	16.0	925.0
3	1952	0.23	22.6	14.8	1015.9
⋮	⋮	⋮	⋮	⋮	⋮
37	1986	3.35	58.1	18.9	4281.6
38	1987	3.35	58.2	16.8	4496.7

Time series – example (comments)

- Data from Wooldridge (2009), available in gretl (*prminwge.gdt*).
- Originally: Castillo-Freeman and Freeman (1992) – effects of minimal wage in the Puerto Rico.

$$
\begin{aligned}
\mathit{Obsno} &= \text{observation;}\\
\mathit{Year} &= \text{rok;}\\
\mathit{Avgmin} &= \text{average (hourly) minimal wage for the year;}\\
\mathit{Avgcov} &= \text{average coverage rate;}\\
&\quad \text{(\% of workers covered by the minimum wage)}\\
\mathit{Unemp} &= \text{unemployment rate;}\\
\mathit{GNP} &= \text{gross national product.}
\end{aligned}
$$

Cross-sectional data

- Sample of individuals, households, firms, cities or a variety of other units taken at a given point in time.

$$Y_i \quad i = 1, \ldots, N$$

- Assumptions – obtained by random sampling from underlying population (may be a problem).
- Pooled cross sections – different, randomly sampled individuals at different time periods (effective to analyse structural breaks in time.

Cross-sectional data – example 1

Tabulka: Cross-sectional data on wage and other individual characteristics.

Obsno	Wage	Educ	Exper	Female	Married
1	3.10	11	2	1	0
2	3.24	12	22	1	1
3	3.00	11	2	0	0
4	6.00	8	44	0	1
5	5.30	12	7	0	1
⋮	⋮	⋮	⋮	⋮	⋮
525	11.56	16	5	0	1
526	3.50	14	5	1	0

Cross-sectional data – example 1 (comments)

- Data from Wooldridge (2009), available in gretl (*wage1.gdt*).
- Abbreviated form on 526 working individuals for the year 1976.

$$
\begin{aligned}
Obsno &= \text{observation;} \\
Wage &= \text{hourly wage;} \\
Educ &= \text{years of education;} \\
Exper &= \text{years of potential labor force experience;} \\
Female &= \text{an indicator for gender;} \\
Married &= \text{martial status.}
\end{aligned}
$$

Cross-sectional data – example 2

Tabulka: Data set on economic growth and country characteristics

Obsno.	Country	Gpcrgdp	Govcons60	Second60
1	Argentina	0.89	9	32
2	Austria	3.32	16	50
3	Belgium	2.56	13	69
4	Bolivia	1.24	18	12
⋮	⋮	⋮	⋮	⋮
61	Zimbabwe	2.30	17	6

Cross-sectional data – example 2 (comments)

- Data from Wooldridge (2009).
- Originally: De Long a Summers (1991) – the study of cross-country growth rates.

$$
\begin{aligned}
\mathit{Obsno} &= \text{observation;}\\
\mathit{Country} &= \text{country;}\\
\mathit{Gpcrgdp} &= \text{growth in real per capita GDP from 1960 to 1985;}\\
\mathit{Govcons}60 &= \text{government consumption as a percentage of GDP;}\\
\mathit{Second}60 &= \text{adult secondary education rates.}
\end{aligned}
$$

Pooled cross-sectional data – example

Tabulka: Two Years of Housing Prices

Obsno.	Year	Hprice	Proptax	Sqrft	Bdrms
1	1993	85500	42	1600	3
2	1993	67300	36	1440	3
3	1993	134000	38	2000	4
⋮	⋮	⋮	⋮	⋮	⋮
250	1993	243600	41	2600	4
251	1995	65000	16	1250	2
252	1995	182400	20	2200	4
⋮	⋮	⋮	⋮	⋮	⋮
520	1995	57200	16	1100	2

Pooled cross-sectional data – example (comments)

- Data from Wooldridge (2009).
- Data on houses sold in 1993 (250 observations) and 1995 (270 observations).

$$\begin{aligned} Obsno &= \text{observation;} \\ Year &= \text{years;} \\ Hprice &= \text{house price;} \\ Proptax &= \text{property tax;} \\ Sqrft &= \text{lot size in square feet;} \\ Bdrms &= \text{number of bedrooms.} \end{aligned}$$

Panel data

- Time series and corss-sectional component.
- Data on the same individuals (countries, cities, firms, etc.) over a given time period.

$$Y_{it} \quad i = 1, \dots, N \quad t = 1, \dots T$$

- Some difficulties to collect them.
- To estimate „individual effects“.

Panel data – example

Tabulka: A Two-Year Panel Data Set on City Crime Statistics

Obsno.	Město	Rok	Murders	Population	Unem	Police
1	1	1986	5	350000	8.7	440
2	1	1990	8	359200	7.2	471
3	2	1986	2	64300	5.4	75
4	2	1990	1	65100	5.5	75
⋮	⋮	⋮	⋮	⋮	⋮	⋮
297	149	1986	10	260700	9.6	286
298	149	1990	6	245000	9.8	334
299	150	1986	25	543000	4.3	520
300	150	1990	32	546200	5.2	493

Panel data – example (comments)

- Data from Wooldridge (2009).
- Two-year panel data set on crime and related statistics for 150 cities in the United States (1986 and 1990).

$$
\begin{aligned}
\textit{Obsno} &= \text{observation;} \\
\textit{City} &= \text{city;} \\
\textit{Year} &= \text{year;} \\
\textit{Murders} &= \text{number of murders;} \\
\textit{Population} &= \text{population number;} \\
\textit{Unem} &= \text{unemployment rate;} \\
\textit{Police} &= \text{number of policemen.}
\end{aligned}
$$

Data transformation

- Depends on purposes of economic analysis.
- Difference: $\Delta Y_t = Y_t - Y_{t-1}$.
- Growth rate (% change):

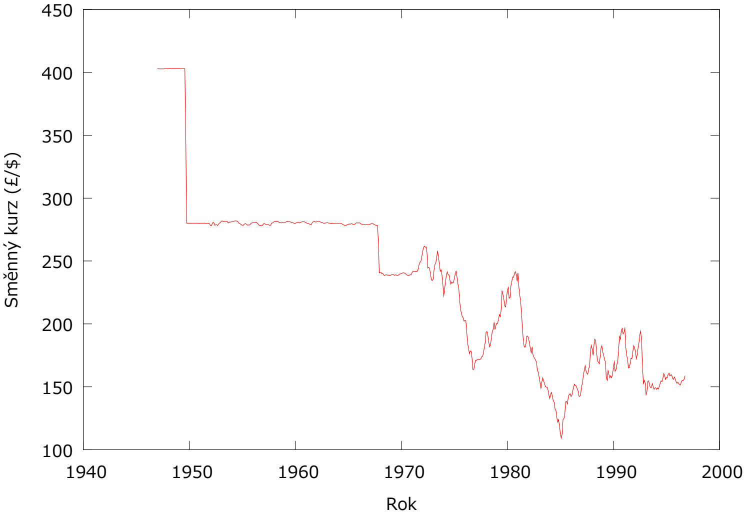
$$\%\Delta Y_t = \frac{Y_t}{Y_{t-1}} - 1 = \frac{Y_t - Y_{t-1}}{Y_{t-1}} \quad (\times 100[\%])$$

- Diference of logarithms (to approximate growth rate):

$$\%\Delta Y_t \approx \ln(Y_t) - \ln(Y_{t-1}) \quad (\times 100[\%])$$

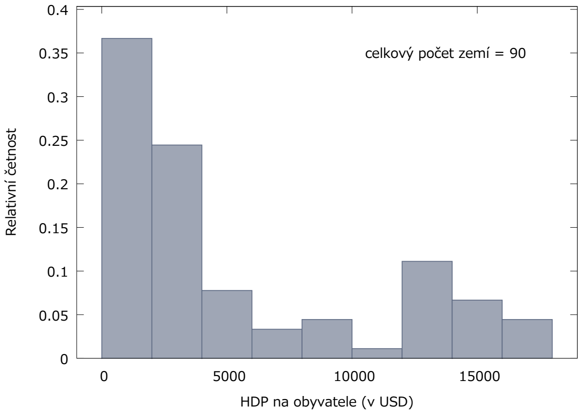
Content

Time series graphs



Obrázek: Time series plot of UK pound/US dollar exchange rate

Histograms



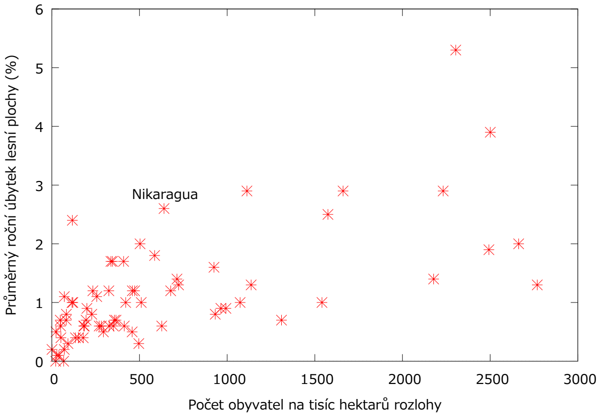
Obrázek: Histogram of GDP per capita for 90 countries.

Absolute and relative frequencies

Tabulka: Frequency table for GDP per capita data.

Interval (USD)	Frequency	
	Absolute	Relative
0-2000	33	36.67 %
2001-4000	22	24.44 %
4001-6000	7	7.78 %
6001-8000	3	3.33 %
8001-10000	4	4.44 %
10001-12000	2	2.22 %
12001-14000	9	10.00 %
14001-16000	6	6.67 %
16001-18000	4	4.44 %

Scatter plots (XY plots)



Obrázek: XY plot of population density against deforestation.

Content

Sample mean and variance

- Random sample: $Y_1, \ldots, Y_N$.
- Sample mean:

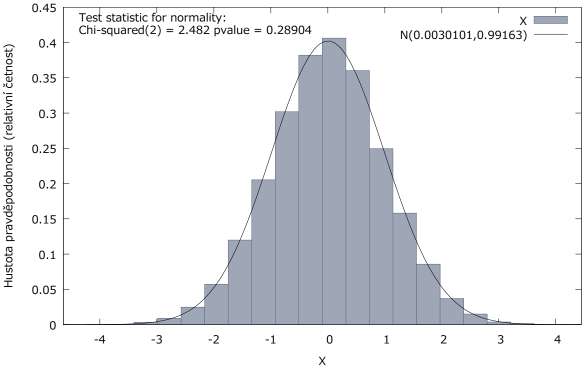
$$\overline{Y} = \frac{\sum_{i=1}^{N} Y_i}{N}$$

- Sample standard deviation:

$$s_Y = \sqrt{\frac{\sum_{i=1}^{N} \left(Y_i - \overline{Y}\right)^2}{N-1}}$$

- Sample variance: s_Y^2.

Histogram and bell-shaped distribution



Obrázek: Histogram for a bell-shaped distribution.

Expected values and variances

- Random variable Y.
- Expected value and population mean: $E(Y) \equiv \mu$

$$E(Y) = \sum_{i=1}^{N} y_i p(y_i) \quad \text{discrete random variable, sample space} \{y_1, \dots y_N\}$$

$$E(Y) = \int_{-\infty}^{\infty} y f(y) \mathrm{d}y \quad \text{continuous random variable}$$

- Population variance: $var(Y) = \sigma^2$

$$var(Y) = E[(Y - \mu)^2] = E(Y^2) - \mu^2$$

Example – expected return on the stock market

- Probability 70% (0.7) – stable markets (return 1%); probability 10% – falling markets (return −10%); probability 20% – rising markets (return 5%).
- $p(y_i) = \Pr(Y = y_i)$.

$$
\begin{aligned}
E(Y) &= p\,(0.05)\,0.05 + p\,(0.01)\,0.01 + p\,(-0.10)\,(-0.10) \\
&= 0.20 \times 0.05 + 0.70 \times 0.01 + 0.10 \times (-0.10) \\
&= 0.007.
\end{aligned}
$$

- Expected return on the stock next month is 0.7% (i.e. a bit less than 1%).

Example – expected ucertainty of stock returns

- Risk of expected returns $\rightarrow$ Y^2: $(0.05)^2 = 0.0025$, $(0.01)^2 = 0.0001$ a $(-0.10)^2 = 0.01$.

$$
\begin{aligned}
E(Y^2) &= p\,(0.0025) \times 0.0025 + p\,(0.0001) \times 0.0001 \\
&\quad + p\,(0.01) \times 0.01 \\
&= 0.20 \times 0.0025 + 0.70 \times 0.0001 + 0.10 \times 0.01 \\
&= 0.00157. \\
var(Y) &= E\left(Y^2\right) - [E(Y)]^2 \\
&= 0.00157 - (0.007)^2 \\
&= 0.001521.
\end{aligned}
$$

- Square root of variance $\rightarrow$ 0.039 (std. dev.) *Rightarrow* expected return 0.7% with uncertainty $\pm 3.9\%$.

Correlation

- Relationship between X a Y.
- Correlation (correlation coefficient):

$$r = \frac{\sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)\left(X_i - \overline{X}\right)}{\sqrt{\sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)^2}\sqrt{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2}}.$$

Properties of correlation

1. $r \in \langle -1, 1 \rangle$.
2. $r > 0$ ... positive correlation; $r < 0$ ... negative correlation; $r = 0$ ... no correlation; $r = 1$ or $r = -1$ ... perfect correlation.
3. $r_{XY} = r_{YX}$.
4. $r_{XX} = 1$.

Correlation – possible interpretations

- De Vaus (2002) – Analyzing Social Science Data: 50 Key Problems in Data Analysis.
- Take with reserve!

Correlation coefficient (abs. value)	Interpretation
0.01 – 0.09	trivial, no correlation
0.10 – 0.29	low
0.30 – 0.49	mild
0.50 – 0.69	substantial, strong
0.70 – 0.89	very strong
0.90 – 1.00	almost perfect

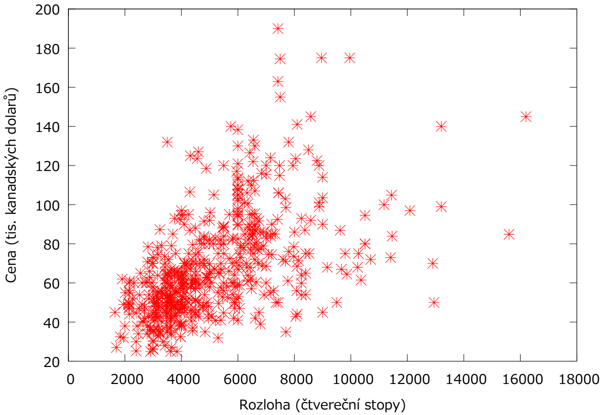
Example – houses prices

- Data set *hprice.gdt*, Koop (2008)
- Data relating to $N = 546$ houses sold in Windsoru, Kanadě, in the summer of 1987.
- Y = the sales price of the house,
  X = the size of its lot in square feet $\Rightarrow r_{XY} = 0.54$
  1. Houses with large lots tend to be worth more than thos with small lots.
  2. There is a positive relationship between lot size and sales price.
  3. The variation in lot size accounts for 29 % (i.e. $0.54^2 = 0.29$) of the variability in house prices.
- Z = number of bedrooms $\Rightarrow r_{YZ} = 0.37$: houses with more bedrooms tend to be worth more than houses with fewer bedrooms.
- $r_{XZ} = 0.15 \Rightarrow$ houses with larger lots tend to have more bedrooms (small correlation $\Rightarrow$ quite weak link between lot size and number of bedrooms).

Correlation and causality

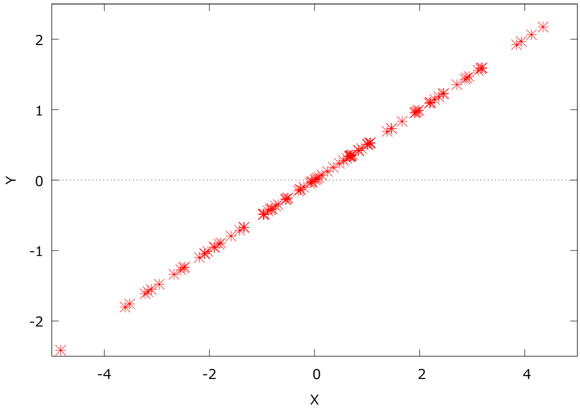
- Correlation does neccessarily imply causality.
- It may be the case that an underlyind third variable is responsible for correlation.
- Direct (no interventing variable) and undirect correlation.
- Use common sense or a convincingeconomic theory to establish causality.

Correlation and XY plot – example 1



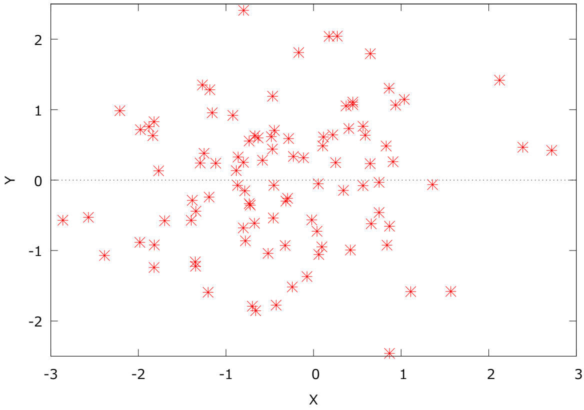
Obrázek: XY plot of lot size against house price.

Correlation and XY plot – example 2



Obrázek: XY plot of two perfectly correlated variables.

Correlation and XY plot – example 3



Obrázek: XY plot of uncorrelated variables.

Correlation matrix

- Correlation between several variables.

	X	Y	Z
X	1.000		
Y	0.318	1.000	
Z	-0.131	0.097	1.000

Population correlations and covariances

- Example: portfolio risk; investment over the summer months in the shares of two companies – an umbrella manufacturer and an ice cream maker.
- Overall portfolio might be much less risky than the individual stocks.
- Population covariance:

$$cov(X, Y) = E(XY) - E(X)E(Y)$$

- Population correlation:

$$corr(X, Y) = \frac{cov(X, Y)}{\sqrt{var(X)var(Y)}}$$

- Sample statistics $\rightarrow$ (consistent) estimates of population statistics.

Content

Exercises

- Koop (2008), chapter 1.