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

HUNGARY'S RENEWAL

The project is supported by the European Union.

Author: Anikó Bíró
Supervised by Anikó Bíró
June 2010

Week 2

Data types, descriptive statistics, indices

Data types I

- Time series:
  - Variables ordered in time
  - Freqency of observations (e.g. monthly, yearly)
  - Notation: Y_t
  - Examples (macroeconomic, financial – individual?)
- Cross sectional:
  - Sample of economic agents at a given time point
  - Examples (individuals, enterprises, countries)
  - Notation: Y_i
  - Random sample

Data types II

- Panel:
  - Time series + cross sectional jointly
  - Observation of the cross sectional sample throughout more time periods
  - Notation: Y_{it}
  - Examples (GDP of European countries, panel of individual households)

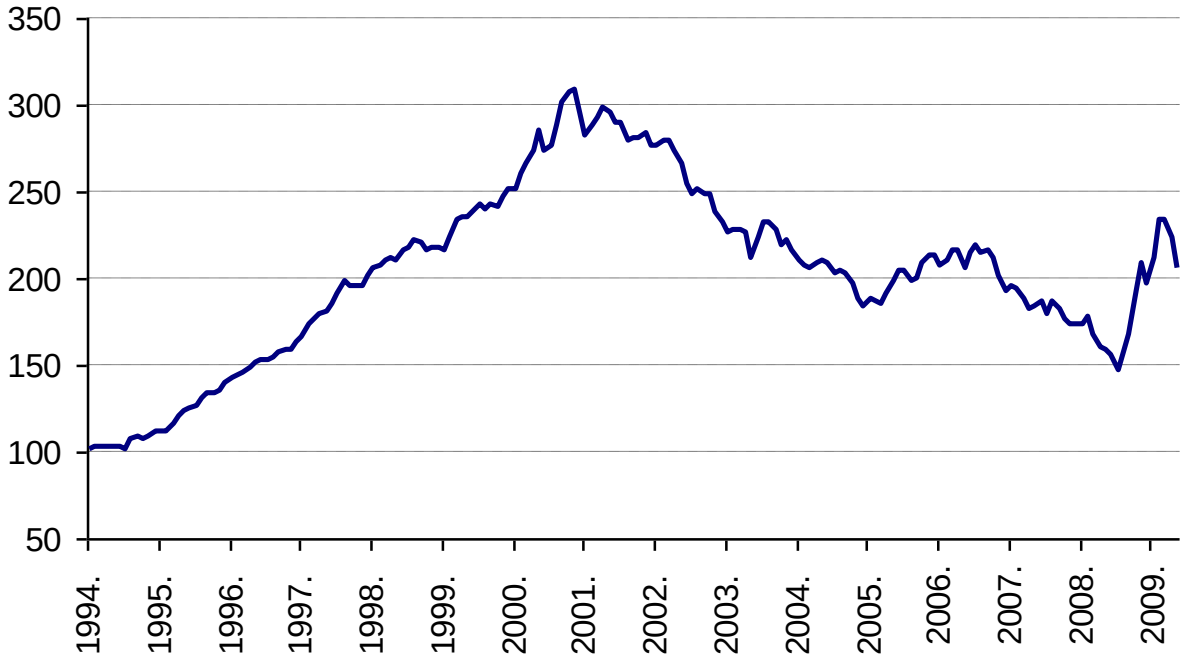
Data types III

- Quantitative and qualitative
  - Quantitative: e.g. inflation, income
  - Qualitative: e.g. male/female, education level – code as numbers
- Level and dynamics
  - E.g. number of employed vs. change in employment

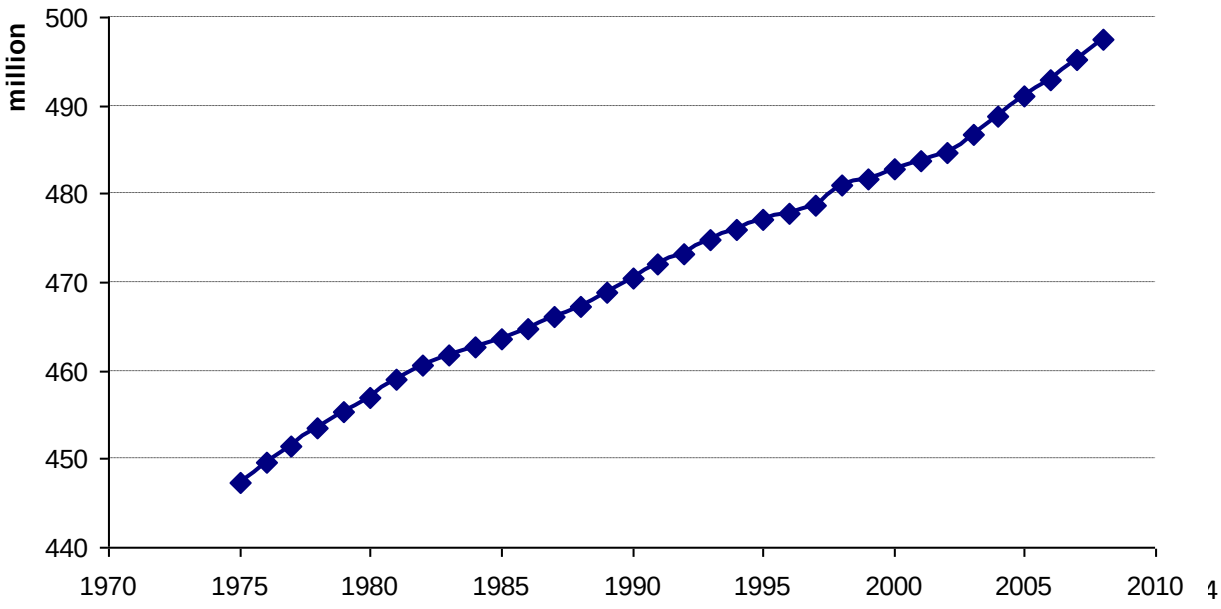
$$\%\text{change} = \frac{(Y_{t+1} - Y_t)}{Y_t} \cdot 100$$

Time series graphs

HUF/USD monthly exchange rate



EU27 population

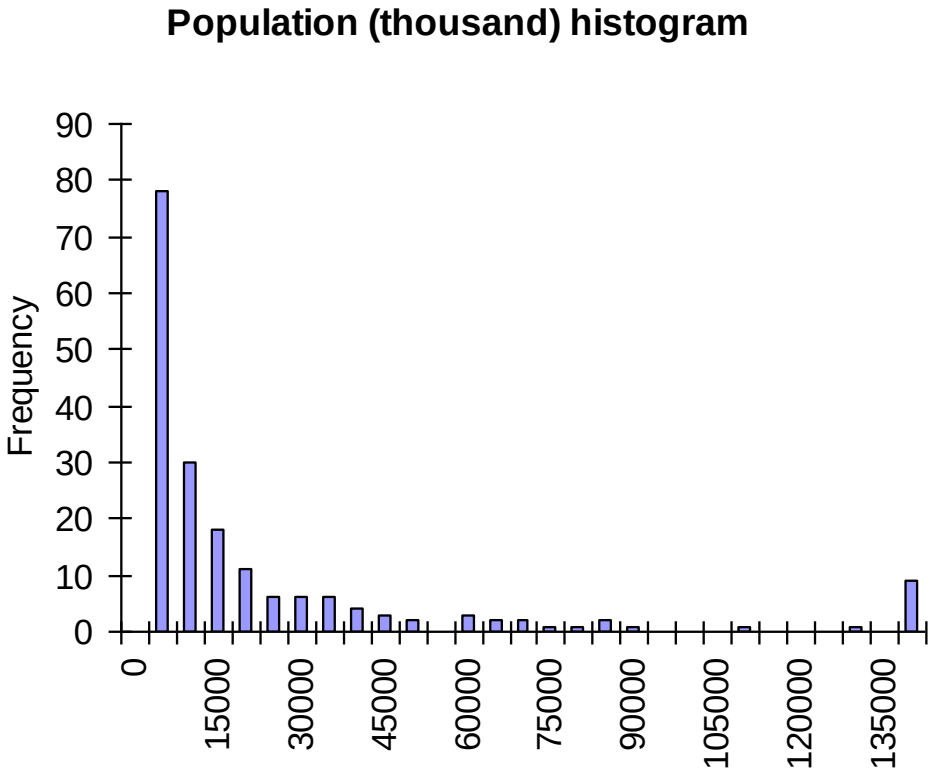


Histograms

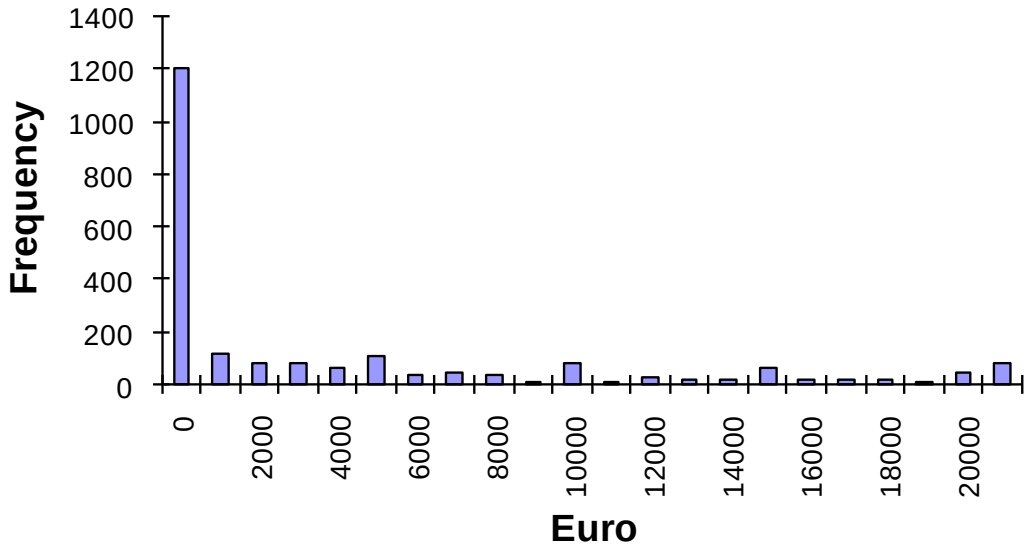
- Plotting cross sectional data
- Example: distribution of income per capita
- Equal intervals (brackets) – determine it in Excel according to the data
- Frequency within the intervals

Histogram, example

- Penn World: distribution of countries according to population (bracket size: 5000)



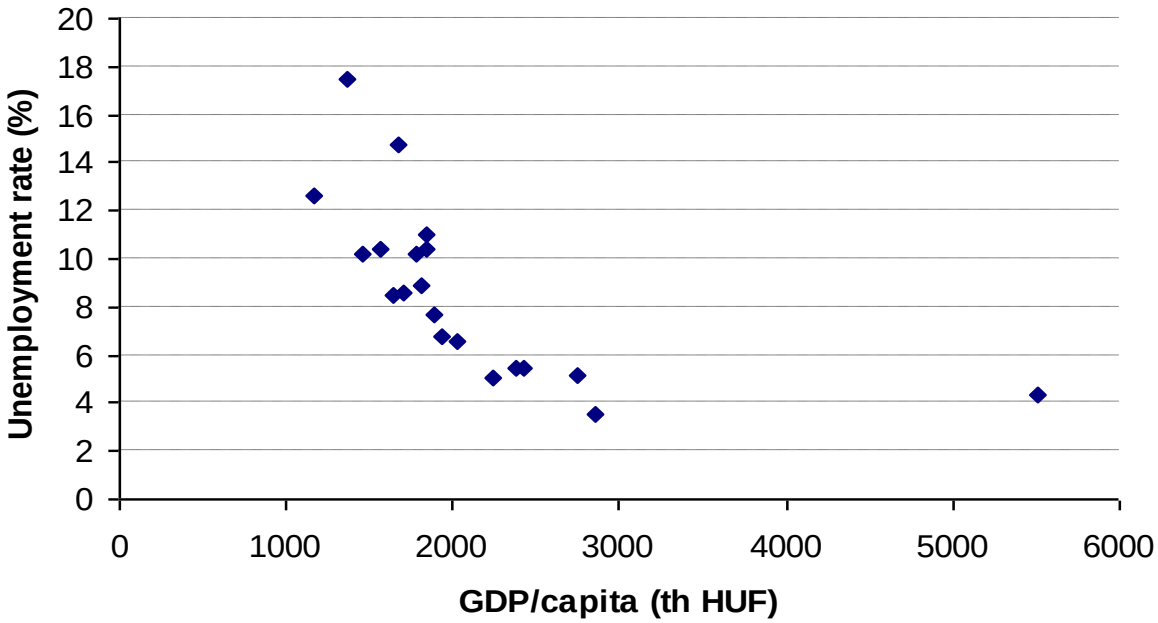
Histogram of car value – Austria, 50+



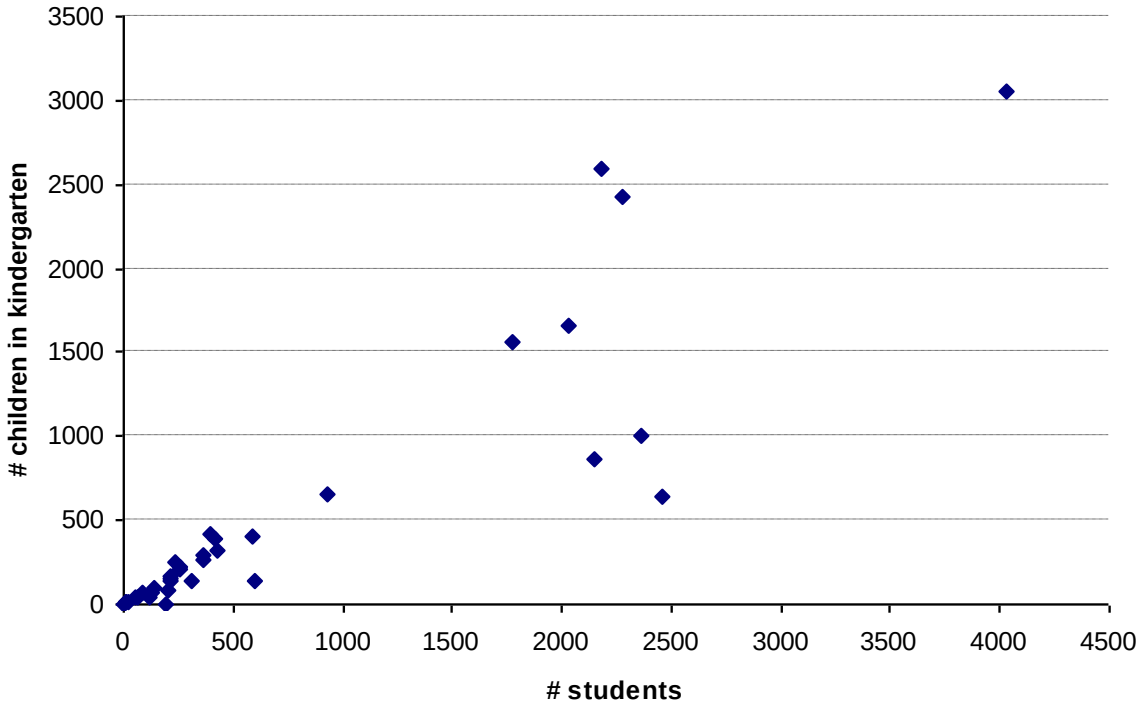
- SHARE: cross sectional sample of people aged 50+
- Value of the car, Austrian subsample (bracket size: 1000)

Point diagram

- Relationship between two variables
- KSH: data on counties



Eurostat: number of students and children in kindergarten by countries, 2007



Descriptive statistics

- Up to now: graphical methods
- Descriptive statistics: numerical summary of some characteristics of the variables
  - Level? – mean, median, mode
  - Variability? – standard deviation, range

Mean

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

- N: number of observations
- Example: mean of country population (Penn World Table) – ca. 34 million

Mode

Mode: most frequent observation

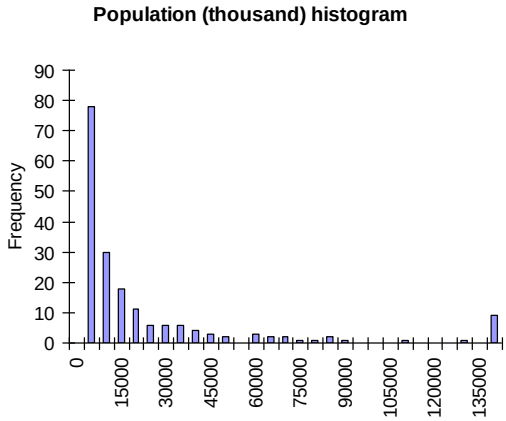
- Problem: does not always exist (e.g. one from each value), there can be more modes
- Possible solution: highest point of the histogram (depends on brackets) – center of the interval

Median, percentile

- Median: value in the middle – half of the observations below the median
- Xth percentile: X% of the observations below X
- Quartile: cuts the observations into four
  - 1st quartile: 25% below, 2nd quartile = median

Skewness

- Example: mean > median
- Some large values – mean is large
- Skewed to the left
- Long right tail



Standard deviation

- Range: difference between maximum and minimum
  - Not reliable (outlier values)
- Variance: mean of squared differences from the mean
- Standard deviation:

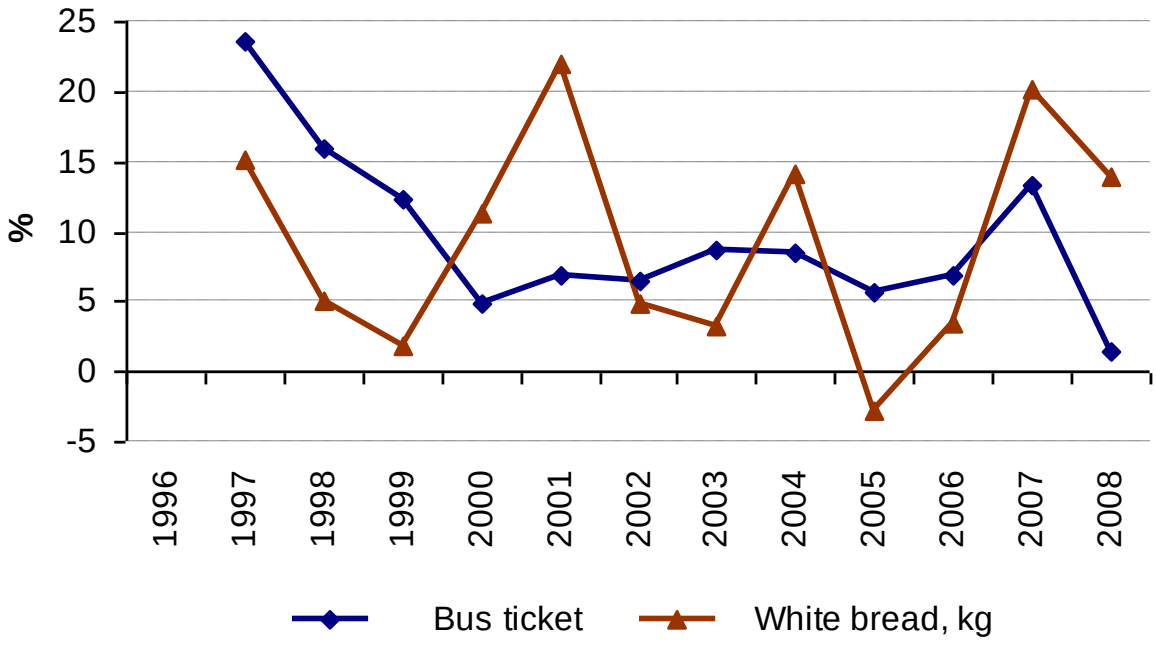
$$s = \sqrt{Var} = \sqrt{\frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}{N-1}}$$

  - Difficult to interpret on its own

Indices

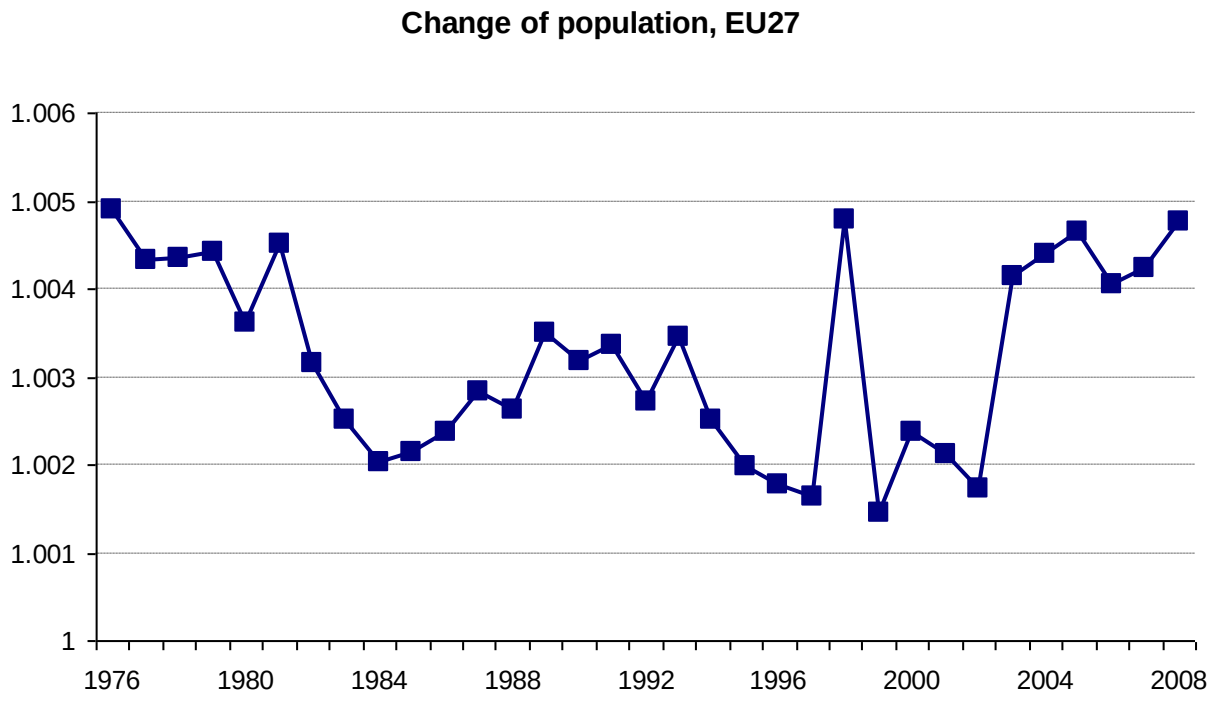
- Price index
  - Price level, average price are difficult to interpret
  - Price index: price level as % of price level at the basic period
  - Annual inflation: basic period changes yearly

Annual price indices, example



Change of population

- Base: previous year



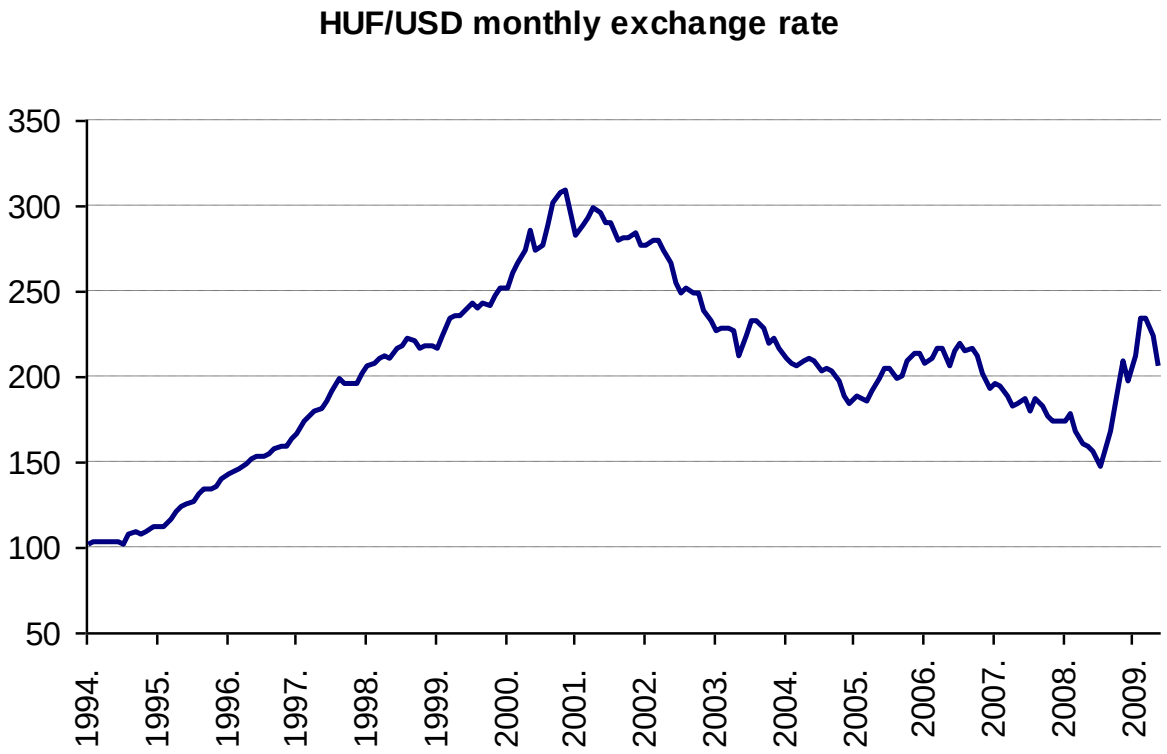
Summary

- Data types:
  - Time series, cross sectional, panel
  - Quantitative, qualitative
- Graphical methods: time series, histogram, point diagram
- Descriptive statistics:
  - Mean, mode, median
  - Skewness
  - Standard deviation

Data types, descriptive statistics, indices

Seminar 2

Time series graphs



HUF/EUR exchange rate? – graph, analysis

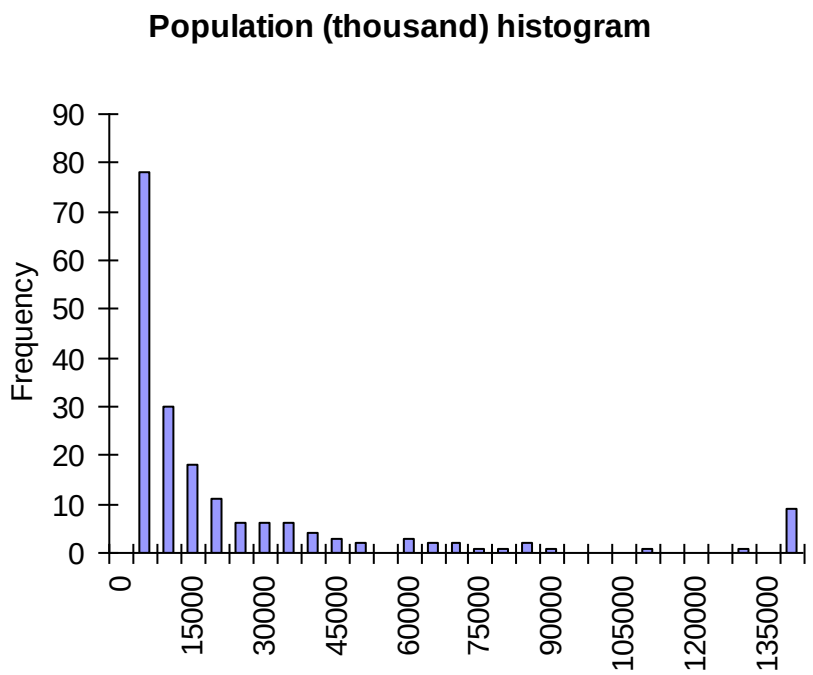
Histograms

- Graphical analysis of cross sectional data
- Excel: Analysis ToolPak extension
- Equal intervals (brackets) – determine it in Excel according to the
- Frequency within the brackets

- Excel: missing observations cause problems – solution: sorting

Histogram, example

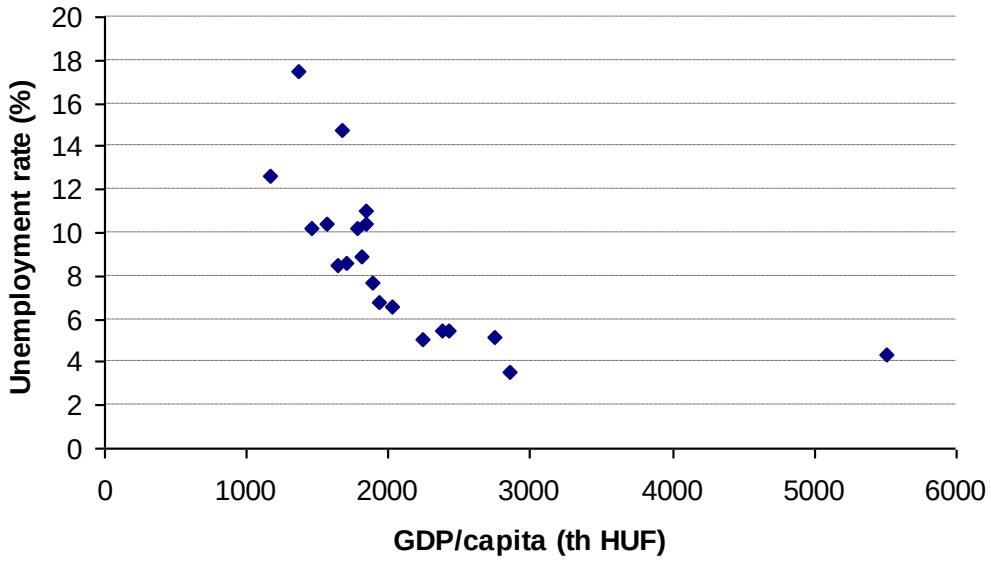
- Penn World: distribution of countries according to population



- Histogram of GDP/capital? Suggested bracket size: 2000

Point diagram

- Relationship between two variables
- KSH: data on counties



- GDP/capital and number of registered enterprises? – What is expected? What can be seen?

Mean

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

- N: number of observations
- Examples: average population of countries, average income/capital (Penn World Tables)

Mode

Mode: most frequent value

Examples:

- Country populations
- GDP per capita

- Based on histograms!

Median, percentile

- Median: middle value – half of the observations below
- Xth percentile: X% of the observations below
- Excel: descriptive statistics (median) + percentile function
  - Example: median, 3rd quartile of population and GDP/capita?
  - E.g. Median=PERCENTILE(B3:B189;0.5)

Standard deviation

- Range: difference between maximum and minimum
  - MIN(), MAX() functions
  - Not reliable (outliers)
- Variance: mean of squared differences

- Standard deviation:

$$s = \sqrt{Var} = \sqrt{\frac{\sum_{i=1}^{N} (Y_i - \overline{Y})^2}{N-1}}$$

- Range, variance, and standard deviation based on Penn World GDP/capita data (descriptive statistics table + functions)

Indices

KSH data

Price indices of bread and bus ticket

- Fix base
- Yearly changing base
- Graphical analysis

Homework 1 (groups)

1. Graphical analysis of a time series variable
2. Analysis of an economic indicator of a cross sectional sample with the help of histogram
3. Analysis of the relationship between two indicators of a cross sectional sample with the help of point diagram

For all three tasks: graph + one paragraph analysis!