Bus 701:
Advanced Statistics

Harald Schmidbauer

About These Slides

- The present slides are not self-contained; they need to be explained and discussed. This will be done in the lectures.

- Even though being a "work in progress" and subject to revision, the slides constitute copyrighted material.
  If you want to reproduce or copy anything from the slides, please ask:

Harald Schmidbauer	**harald** at **hs-stat** dot **com**
Angi Rösch	**angi** at **angi-stat** dot **com**

- The slides were produced using LaTeX and R (the R project; website: www.R-project.org) on a GNU/Linux system.

- R files used for this course are available upon request.

About using statistical methods.

Advice.

- **If you have a statistics problem in your own research, please consult a statistician.**
- **Begin consulting the statistician before you start collecting your data set.**

PART I:

Introduction to Statistics;

Basic Concepts;

Descriptive Statistics

Chapter 1:

Introduction

1.1 Some Examples

Example 1: Scratching down numbers.

Students in a first course of statistics were asked to enter their gender and body-height into a list. The result was:

(m,167)(m,181)(m,178)(m,180) (f,160) (f,174) (f,170) (f,173) (f,168)
(f,165) (m,195)(m,185)(m,183)(m,192)(m,195)(m,188) (f,168) (f,160)
(f,169) (f,175) (f,157) (f,162) (f,172) (f,160) (m,173) (f,157) (m,170)
(f,170) (f,168) (m,182)(m,175) (f,165) (m,185) (f,170) (m,182)(m,190)
(m,188) (f,160) (m,188)(m,182) (f,177) (f,165) (f,163) (f,160) (f,160)
(f,170) (m,175)(m,176)(m,183) (f,170) (f,170) (f,169) (m,181)(m,177)

Here, f=female and m=male, and the body-height is in centimeters.

What can you tell from these data?

1.1 Some Examples

Example 1: Scratching down numbers.

- stem-and-leaf display (stemplot):

female:

15	77
16	0000002355588899
17	00000023457

male:

16	7
17	0355678
18	0112223355888
19	0255

- Here, **15** | 7 = 157 cm.

- . . . and what can you tell now from the data?

1.1 Some Examples

Example 2: Averages.

Averages are very important. They beware us from getting lost in information. But we have to be careful with averages.

Consider a share of stock. Its price. . .

. . . 2 years ago:	\$ 100
. . . 1 year ago:	\$ 150
. . . today:	\$ 90

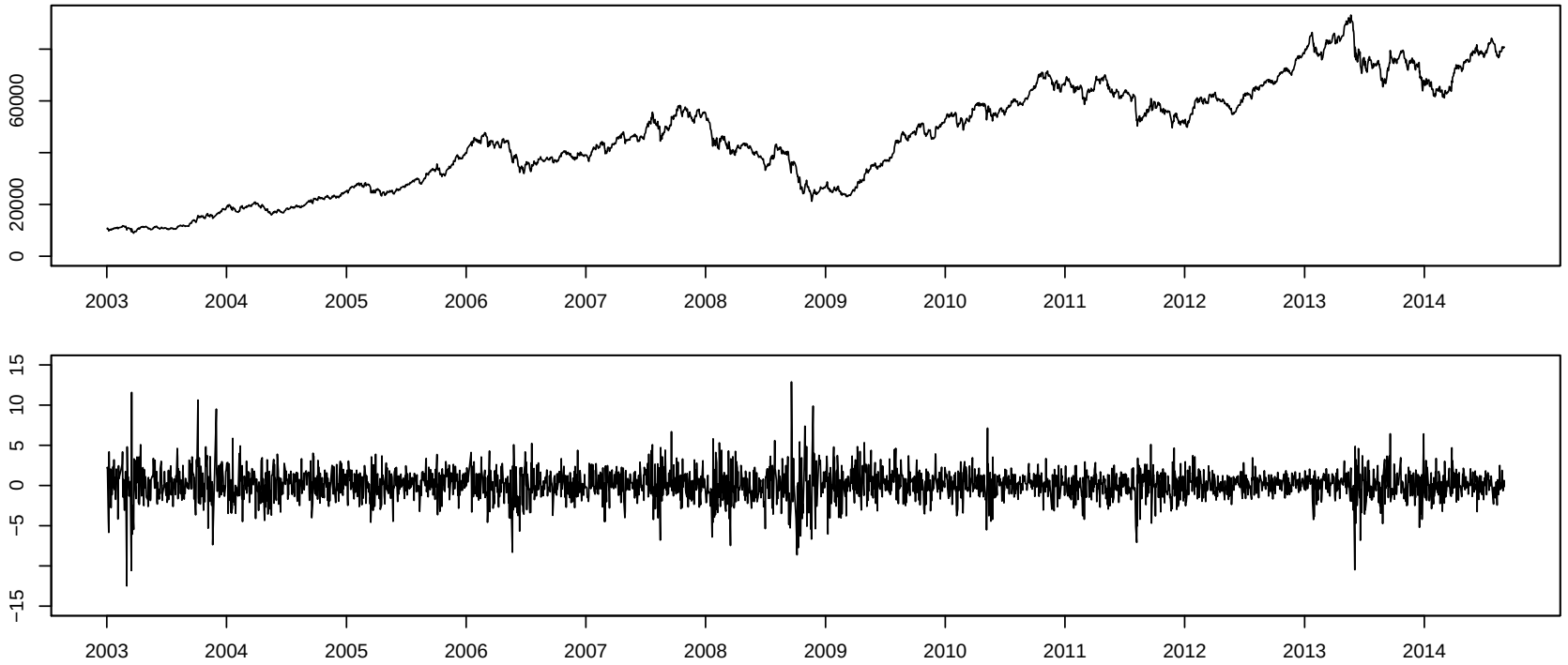
What is the average annual gain or loss of the stock in percent?

1.1 Some Examples

Example 3:

Daily closing quotations and returns on a stock index.

- BIST 100:

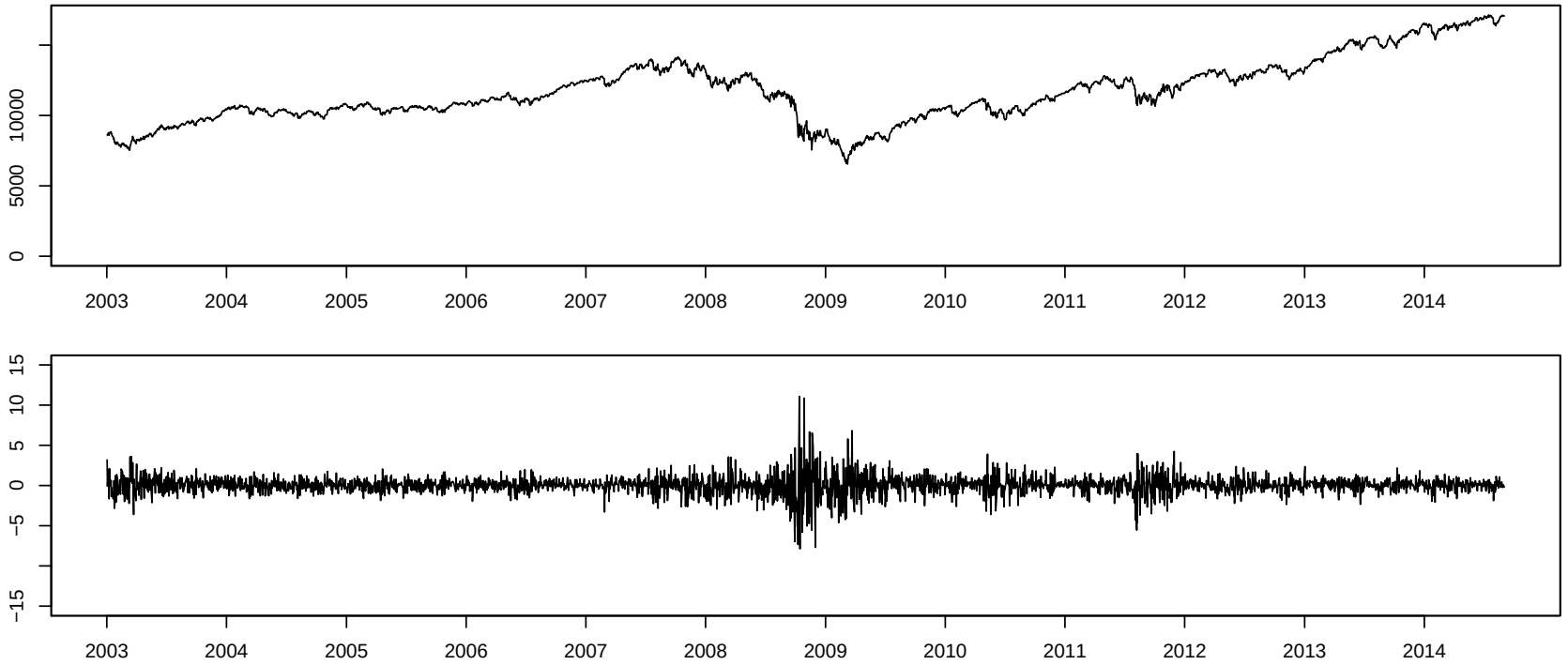


1.1 Some Examples

Example 3:

Daily closing quotations and returns on a stock index.

- DJIA:



1.1 Some Examples

Example 3:
Daily closing quotations and returns on a stock index.

- What can you say about the average daily return?
- Is the "risk" the same when investing in BIST 100 or DJIA?

1.1 Some Examples

Example 4: A public opinion poll.

- After hurricane Katrina, are people in favour of rebuilding New Orleans?
- Result of a poll: 384 of 609 adults (that is, about 63%) polled by telephone September 5-6, 2005, said they believe New Orleans should be rebuilt.
- What can we do with this information? — What does it mean? — Can it answer our initial question?

1.1 Some Examples

Example 5: Television audience rating.

- Ayşe hanım is the program manager of a television channel.
- Her goal is that, in the future, the rating of "Çiçek Taksi" should be at least 10%.
- One evening, 350 televisions in 4000 randomly selected households were tuned into this program.
- Can we conclude that Ayşe hanım has not reached her goal?

1.1 Some Examples

Example 6:
Credit card management: Is a person credit-worthy?

A bank cannot see the person in detail. . . But there are clues:

- age, education, professional environment
- previous payment behaviour
- stability of residential area
- number of cellular phone contracts

How can we exploit these clues? Data protection issues?!?

1.2 Statistics as a Science

"Statistics".

The word "statistics" can refer to:

- a science ("statistics" is singular)
- results of this science ("statistics" is plural)

1.2 Statistics as a Science

Statistics is the science of reasoning with numbers.

Statistics is concerned with

- detecting the structure in data sets.
- facilitating the communication between people.
- making well-founded decisions.
- forecasting the future.
- providing a link between theory and observations.
- determining what (numerical) information is needed to solve a given problem.

1.2 Statistics as a Science

On being misled by numbers.

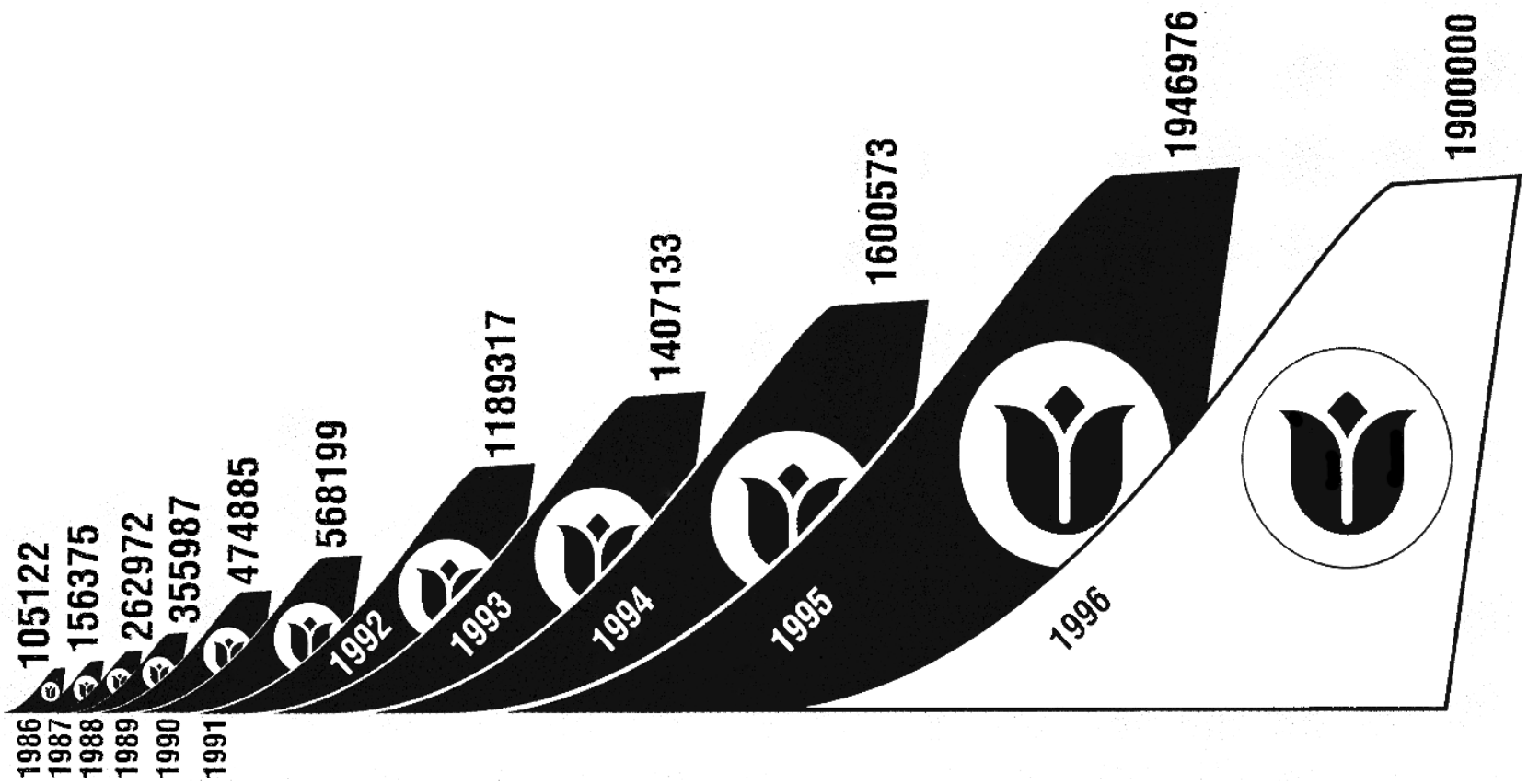
İstanbul'un nüfusu (2013):

$$\boxed{14\,160\,467}$$

(Source: TÜİK)

1.2 Statistics as a Science

Images can be misleading, too.



1.2 Statistics as a Science

Statistics and the Computer.

- Although some techniques can be done using paper and pencil, statistics is a hi-tech science: It needs powerful software to be effective.
- The computer does the computation.
- You have to do the reasoning yourself.

1.2 Statistics as a Science

Statistics and the Computer.

- Elementary analysis: Use a spreadsheet program.
- Spredsheet programs:
  Gnumeric, MS-Excel, LibreOffice Calc, . . .
- Professional users: R. Please visit:

`www.R-project.org`

1.3 Descriptive and Inductive Statistics

The goals of descriptive and inductive statistics.

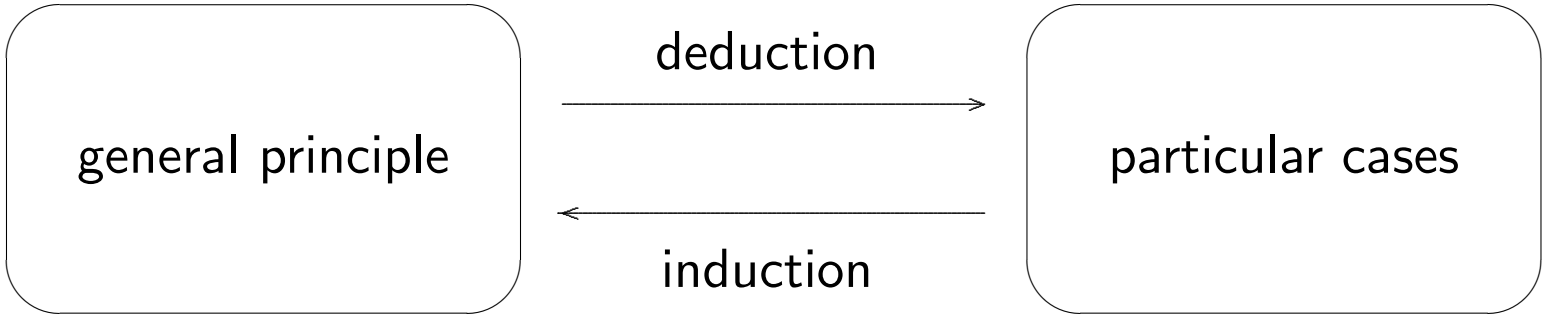
The goal of. . .

- . . . descriptive statistics is: Describe, summarize, display given data (data reduction!).
- . . . inductive statistics is: Draw conclusions from data (sample data, observations) to more general principles (the population). The process of drawing conclusions is called statistical inference.

1.3 Descriptive and Inductive Statistics

Conclusions.

There are two kinds of conclusions:



- In the context of inductive statistics, the "particular cases" are observed data (sample data).

- The "general principle" is a probability distribution, characterizing the entire population.