

## **0.1 Theory of Data Analysis ?!**

a Courses in statistics:

Collection of methods for a defined type of problems.

Exercises: Datasets asking for this methodology.

## 0.1 Theory of Data Analysis ?!

- a Courses in statistics:  
Collection of methods for a defined type of problems.  
Exercises: Datasets asking for this methodology.
  
- b Practice: Start from (scientific) question and (potential) data.  
→ Choose appropriate statistical methods

In reality:

- Questions are often vague  
→ Sharpen to identify suitable methods
- Data are distorted, missing, ...

- Questions are often vague  
→ Sharpen to identify suitable methods
  
  - Data are distorted, missing, ...
- c Practice of Data Analysis is difficult to teach consistently.  
→ “Apprenticeship”, “Statistics Lab”

- Questions are often vague  
→ Sharpen to identify suitable methods
  
  - Data are distorted, missing, ...
- c Practice of Data Analysis is difficult to teach consistently.  
→ “Apprenticeship”, “Statistics Lab”
- d **Theory of Data Analysis?**  
would need “soft science” methods. First step: Concepts, like

- **Statistical Problem:**  
("Scientific") Question and (potential) Data.
- **Strategy**, consisting of **steps** (see course on Regression).
- **Quality of a result:**  
How to measure? → Ranking of Strategies.

- **Statistical Problem:**  
("Scientific") Question and (potential) Data.
- **Strategy**, consisting of **steps** (see course on Regression).
- **Quality of a result:**  
How to measure? → Ranking of Strategies.

e **Case Studies** are useful!

- **Statistical Problem:**  
("Scientific") Question and (potential) Data.
- **Strategy**, consisting of **steps** (see course on Regression).
- **Quality of a result:**  
How to measure? → Ranking of Strategies.

e **Case Studies** are useful!

f General Strategy to define and tackle a Statistical Problem?

# 1 Statistical Studies and Consulting

## 1.1 Experience

a Statistical Consulting covers a wide spectrum:

- Knowledge and skill of the client and the consultant
- Goals of the project

“How should I interpret this output?”

“A reviewer has criticized the following: ...

How can I justify my approach?"

"I have been told to do a *Conjoint Analysis*.

Which program does this for me?"

"I have here an interesting data set.

I would like to apply multivariate methods."

# 1 Statistical Studies and Consulting

## 1.1 Experience

- a Statistical Consulting covers a wide spectrum:
- Knowledge and skill of the client and the consultant
  - Goals of the project

“How should I interpret this output?”

“A reviewer has criticized the following: ...

How can I justify my approach?”

“I have been told to do a *Conjoint Analysis*.

Which program does this for me?”

“I have here an interesting data set.

I would like to apply multivariate methods.”

b **Roles:** Statistician is

- consultant with a limited task (limited liability)
- responsible for adequate statistical analysis of the data
- partner in the project, from planning to reporting / article

1.1

## c Communication

- Who poses the (scientific) problem, and to whom should the answer be targeted?  
What effort will the targeted persons make to understand the answer?  
—→ Type of answer, approaches that can be understood.
- What knowledge do the targeted persons – the specialists – the statisticians have?
- How much effort can be invested? (Money, time, energy)

1.1

- d **Critical points.** At the beginning clarify:
- the problem (“scientific” question).  
Informal questions are ok. at the outset, but must be made precise!  
→ leads to methodology = models and procedures
  - structure of the data: How have / will they be generated?  
Search for groupings
    - blocks,
    - “main/subplot” / “within/between subjects”
    - “closeness” in space or time.

1.1

## e Why is data structure important?

Statistics = model + estimation of parameters

estimation without indication of precision is meaningless → "standard error"

→ confidence interval or test.

Determination of s.e. needs independent observations

(or a good model for the dependencies)

If (positive) correlations among observations are neglected,

→ s.e. will be under-estimated

→ too short confidence intervals or "liberal tests" (=wrong !)

Look for independent groups of observations!

Do any analysis you like, as for individual observations,  
then use bootstrap over groups to get precision.

## 2.5 Structure of a Report

- a **Target audience:** Client and his/her “environment”.

**Theory:** Sketch main ideas, give basic reference.

## 2.5 Structure of a Report

- a **Target audience:** Client and his/her “environment”.

**Theory:** Sketch main ideas, give basic reference.

- b **Structure:** in scientific articles, the common structure is fixed: “Introduction”, “Material and Methods”, “Results”, “Discussion”  
This may be adequate. Sometimes preferable:  
Integrate statistical methods with results

## 2.5 Structure of a Report

- a **Target audience:** Client and his/her “environment”.

**Theory:** Sketch main ideas, give basic reference.

- b **Structure:** in scientific articles, the common structure is fixed: “Introduction”, “Material and Methods”, “Results”, “Discussion”  
This may be adequate. Sometimes preferable:  
Integrate statistical methods with results

- c Do not write a detective story! Do not withhold the results!

2.5

d Proposal for the structure:

- **Introduction:** Problem, background, earlier work.
  
- **Data**
  
- **Statistical methods and results**
  
- **Discussion and outlook**
  - Summary and interpretation of results,
  - Outlook on open questions and extensions of the analyses

2.5

- e Write **summary** in the end. Should be readable for itself.

- 2.5
- e Write **summary** in the end. Should be readable for itself.
  
  - f **Appendices:**
    - detailed description of data,
    - exact description of statistical methods
    - “complaints”
    - repetitive analyses
    - analyses with uninteresting results

- 2.5
- e Write **summary** in the end. Should be readable for itself.
  
  - f **Appendices:**
    - detailed description of data,
    - exact description of statistical methods
    - “complaints”
    - repetitive analyses
    - analyses with uninteresting results
  
  - g **Length of report:** 5 - 10 pages (without appendices)