Teaching Statistics to Social Science Students

Particular focus on teaching mixed ability and less enthusiastic students

“Trying to teach me statistics? That is like trying to fit a square peg into a round hole”

First-year student in Anthropology

Manfred te Grotenhuis
My Own Educational Career

• Elementary School (age 6-12)
• Lower Secondary School (age 12-16)
• Vocational School to become an electronic engineer (age 16-19)
• Worked in engineering for two years, then decided to go to College (age 21)
• Entered Radboud University Nijmegen in the Netherlands to study Sociology (age 23)
• Graduated in 1994 and began a thesis, which was a mix of statistics and sociology (age 27)
• Got a job as a junior lecturer on Statistics in 1999 after completing PhD (age 32)
My first two years as a lecturer

• We used McClave, Dietrich & Sincich: *Statistics*. A good book, covers statistics well, clear, but technical, in English with crowded pages.

• My students come from Sociology, Anthropology, Communication Sciences. They hate, do not like statistics much.

• We did not use SPSS in exams, students used a calculator.

• We had rather poor results: about 40-50% past the test.

• The definite blow came from a student in Anthropology who said that my way of explaining was ok, and that she might be able to calculate Chi-Square but that she had no clue whatsoever what it meant in her field of research or how to use it…
What went wrong?

I guess that we did everything to live up to their (worst) expectations:

• They expected something with formulas and calculations: we did

• They expected that it was going to be difficult, hard to understand and hardly would have any connection to their own study: so it was

• They expected theory on statistics and no practical tools to use it: that was (almost) true.
Looking for a solution (back in 2000)
What kind of students do we teach statistics?
What do they have to know?
What do they have to be capable of?
How to fit square pegs: adapt the holes

On the basis of my experiences, I reworked the courses on statistics:

- Started using examples that fall into the field of interest

I started talking about Aboriginals (Anthropology), about watching tv (Communication Sciences) and life expectancy in 1899 (Sociology). These themes are from courses they already had been taken.

- The examples (real data) are used in SPSS, and the SPSS output and the way to obtain the output (menu & syntax) are discussed during class.
I have also reworked the assignments / exams:

- Weekly assignments: related examples and students are invited to analyze, interpret and to write about it!
- Attaining classes, and weekly assignments are compulsory
- An exam in which both statistical theory and practical skills are tested

In the next slides I walk you through an example, assignment and exam
Example: Age distribution in the Netherlands, 1899

Age distribution (ages 0-101) in the Netherlands from 1899, 5.5 million cases (source: Statistics Netherlands, http://statline.cbs.nl/statweb)

mean age (μ) = 27.1 years
standard deviation (σ) = 20.6 years
Let’s play darts! A Random Sample…

Playing darts blindfolded…

Question to students: if every dart picks one person out of this distribution, what will the distribution look like in a sample of 100 darts/persons?
The Age Distribution in a Sample…

Answer: it will look very similar to the triangle in the population: many young people and the older they are, the less they will be found in the sample.

So:

The distribution in a sample tends to look like the distribution in the population!

Question: if we take many of these ‘100 darts samples´, and we calculate the mean age for each sample, what will the distribution of mean ages look like then?
Mean Age in 100,000 Samples (1000 units each) with $\mu = 27.1$ and $\sigma = 20.7$

Answer: a normal distribution!

$E(\bar{x}) \approx 27.1$

(= mean age in population)

$\sigma(\bar{x}) \approx 0.65$

Depends on sample size (here: 1000 individuals) and on standard deviation in population (20.7). Calculation: $20.7 / \text{square root of 1000} = 0.65$. 
Sampling Distribution: 90% Confidence Interval

Mean Age in 100,000 samples with $\mu=27.1$ and $\sigma=20.7$

The mean age is $\approx 27.1$ and the standard error is $\approx 0.65$.

1.65 standard error away from mean

Yes, but why then?

Between these two boundaries you will find 90% of all samples!!

27.1 + 1.65 * 0.65 = 28.2
27.1 - 1.65 * 0.65 = 26.0
To test Hypotheses with $p$ and $\alpha$

Research hypothesis: in 1930 the mean age has risen compared to 1899

Null hypothesis: the mean age did not rise between 1899-1930

- Rejection Area
  $\alpha = 5\% \ (1.65 \times 0.65)$

- Probability
  $P=0.0228$

In the 1930 sample the mean age is 28.4! This outcome is within the rejection area of 5% (0.05).

So, between 1899 and 1930 the mean age most probably did rise! This confirms the research hypothesis and consequently the null hypothesis is rejected.
The example is from Statistical Tools, book & webpage:

- This site supports our textbook STATISTICAL TOOLS (click here for more information and/or to order).
- Free examination copies available, email for request: click here.
- All additional information can be found in the six toolboxes below.
- The boxes 'Figures and Tables', 'Calculate Probabilities', 'Exercises' and 'More Tools' are ready. The other two boxes ('Mostly Theory' and 'More Theories') will be filled in the near future.
- I presented Statistical Tools during the 5th BSSR conference at the University of Maastricht. The presentation can be found at the 'More Theories' box.

http://www.ru.nl/mt/statistics/home/
The example in SPSS: syntax (or menus) and outcomes

***** Open file 'figure 3.2.sav'.
GRAPH line(area)=age.
Example II: a t-test for number of kids in the Netherlands

**** Open file 'figure 3.8.sav'.

T-TEST
/TESTVAL = 2
/VARIABLES = number_kids.

<table>
<thead>
<tr>
<th>One-Sample Test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test Value = 2</strong></td>
</tr>
<tr>
<td><strong>95% Confidence Interval of the Difference</strong></td>
</tr>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>number_kids</td>
</tr>
<tr>
<td>Number of Children</td>
</tr>
</tbody>
</table>

T-TEST
/TESTVAL = 2
/VARIABLES = number_kids.
• Check **Discrete missing values**, click in first empty field and type 999 (see Figure 2.7).
• Click **OK** to have 999 missing.
This is the first set of exercises related to Statistical Tools. You need the program SPSS or another software package that can read and analyze SPSS sav files. Computer related operations are indicated with a bullet (•). Relevant Computer Tools are indicated with ST.

- Download the data file PRACTICE.SAV from the web page (URL = http://www.ru.nl/mt/statistics/home/exercises). You may download this file by a mouse click together with Ctrl key on download or by returning to the main page ➔ exercises. Start SPSS (or any other program that can read SPSS sav files) and open PRACTICE.SAV. Open also a text file in the program Word or any other word processor where you store your answers to the questions below.

1. How many units of analysis are stored in the data file?
2. How many men and women are in the file? (Suggestion: to answer this question a frequency table comes in handy (see ST: page 28).
Question 5

Someone argues that the effect of someone's body length on his/her weight you just found in your research implies that the effect also exists in the Netherlands as a whole. Please reply to this on the basis of the regression analysis you just did.

I'm not sure about the effect as only 1000 respondents took part in the sample.

I'm pretty sure the effect exists in the Netherlands as a whole as I have used a random sample which is sufficiently large.

I'm not sure about the effect because the statistical probability is very low.

I'm pretty sure the effect exists in the Netherlands as a whole, because the t-value is very large.
An introductory course in Statistics

9 weeks: lectures 2 hours a week (90 minutes on ‘theory’, 30 minutes on assignments), take home assignments (3 hours work)

• Week 1: Introduction to descriptive statistics & SPSS
• Week 2 and 3: SPSS training (using menus only)
• Week 4: Graphical descriptions of qualitative & quantitative variables
• Week 5: Numerical descriptions of qualitative & quantitative variables
• Week 6: Relationships between two qualitative variables
• Week 7: Relationships between two quantitative variables
• Week 8: Likert-scales + how to write a paper with descriptive statistics
• Week 9: Summary
A follow-up course in Statistics

9 weeks: lectures 2 hours a week theory, 2 hours a week assignment (take home assignments, 3 hours work)

- Week 1: Introduction to inferential statistics & SPSS using syntax
- Week 2: One-Sample tests for Mean
- Week 3: One-Sample tests for Proportion
- Week 4: Tests for Comparing Two Means (paired / not paired)
- Week 5: Analysis of Variances (ANOVA)
- Week 6: Associations in Contingency Tables
- Week 7: Linear regression analysis (simple)
- Week 8: Linear regression analysis (multivariate)
- Week 9: Summary
Summary

Teaching Statistics to mixed ability and less enthusiastic students

Square pegs & round holes?

No, I do not believe that is true. In my experience one can teach even quite complex statistics to these students when one:

• starts with topics from their research field
• shows how statistics can be a useful tool
• avoids Mathematics as much as possible
• uses weekly assignments with a clear link to the fields of interest, with SPSS as a workhorse.
Teaching Statistics to Social Science Students

Particular focus on teaching mixed ability and less enthusiastic students

Manfred te Grotenhuis
Radboud University Nijmegen
The Netherlands

Thank you!