

Swiss Statistics Meeting 2004

* * *

The Perception of Statistics: Statistical Thinking and Quality

Yves-Laurent Grize
Basel, Switzerland

`yves-laurent.grize@baloise.ch`

Overview

- Perception of Statistics
- What is Statistical Thinking?
- Promotion of Statistical Thinking
- Statistical Thinking and Quality

A Conversation

Example: a knowledgeable scientist comes and asks a statistician for help:

Scientist: I will bother you just for a couple of minutes,
I know that I need to do this and that analysis, but what do you
think, is it correct?

Statistician (trying to be diplomatic):

Well, pretty good, but let me ask you something:
What is really the problem you want to solve?

Scientist (frowning the eyebrows):

Aren't you a statistician?

(and starts to wonder if he was right to come and ask for help).

**Scientists don't want us statisticians
to be „involved“ in their business!**

A paradoxical Situation

On the one hand Statistics IS about:

- learning from data
- the discovery process (G. Box)
- THE scientific method (the science of doing science)
- etc...

And yet on the other hand:

**there is no appreciation by users of Statistics
of the role and power of statistical thinking!**

Only as a Tool the usefulness of Statistics is well recognized:
e.g. FDA approval requirements, Quality Certification, journal
publication requirements, etc...

Why???

Perception of Statistics



because Statistics is perceived as a set of tools
tools and not as a set of problems

J.H. Friedman

Example:

the use of statistical experimental design in pharmaceutical & chemical research

From my experience:

Users tend to be mostly in Engineering (process optimization), sometimes in Development (analytical methods) and surprisingly rarely in Research (discovery)!

-> Tool perception dominates and is not perceived as a fundamental element for discovery/learning/thinking

This 'tool-perception' has been accentuated over the last 25 years due to the increase in computing power and the broad availability of software and PCs.

What is Statistics about“?

■ The Fundamental Problem of Statistics:

- How to learn from data (information)
- despite of errors in data (variability)

-> Separation of systematic errors and random errors

Uncertainty and Variation

The logical Challenge: how to draw conclusions from certain premises

The statistical Challenge: how to draw conclusions from uncertain premises



Key: Quantification of Uncertainty!

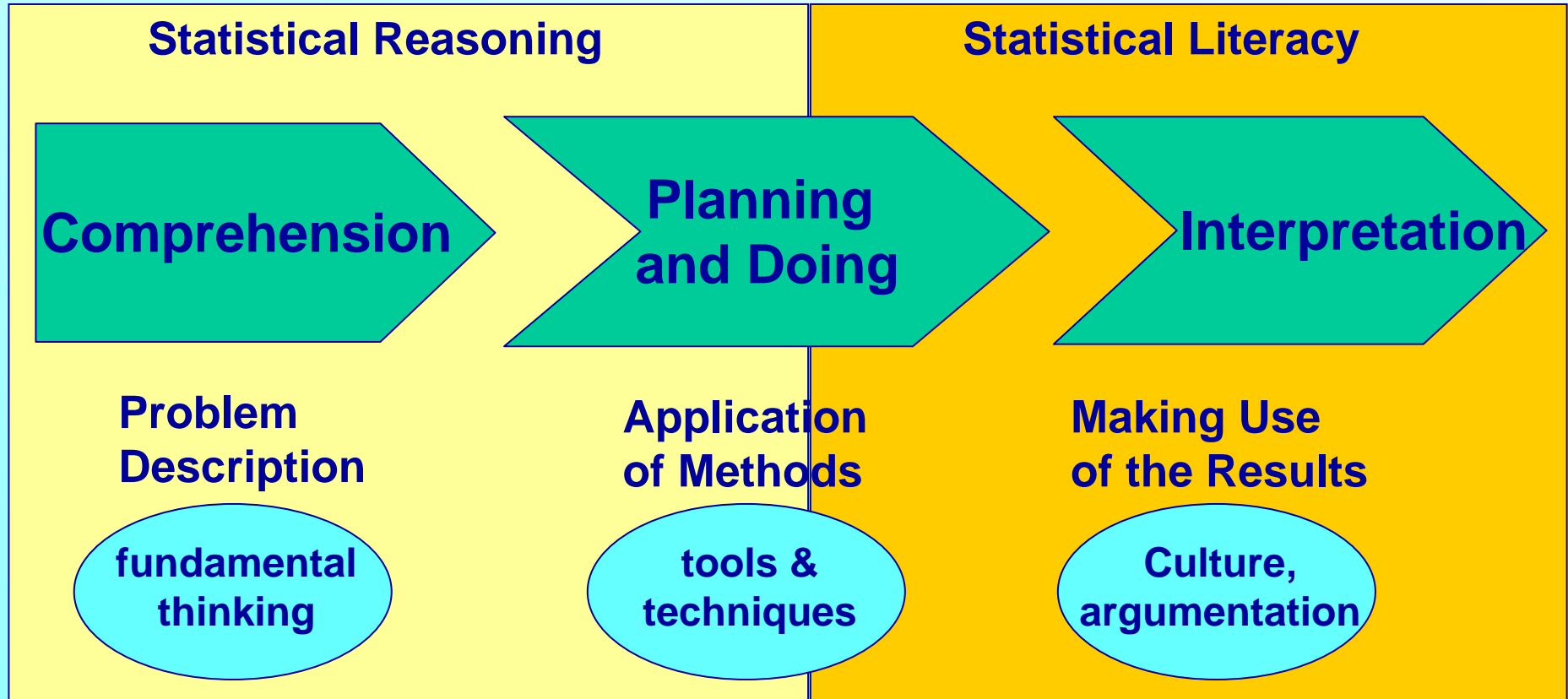
What is „Statistical Thinking“?

- Statistical thinking =
the mental process used to solve the fundamental problem of statistics, i.e. of understanding variation in data.
- Statistics is the language of data
and Statistical Thinking is how to think in this language, hence it enables us to GRASP this problem and hopefully to SOLVE it.
- But one needs a good ear:

***„Although we often hear that data speak for themselves,
their voices can be soft and sly“.***

Mosteller, Fienberg, Rourke (1983)

„Statistical Thinking“ as a global process



Statistical Reasoning Skills

- 1. Critical Thinking** Ability to think logically and to assess the quality of evidence. Identify logical fallacies and well-posed questions amenable to scientific inquiry.
- 2. Hypothetical Thinking** Think in terms of hypothetical implications, scientific method, causality, use concepts of counterfactuals.
- 3. Stochastic Understanding** Understand probability, variation, uncertainty, sample space.
- 4. Statistical Modeling** Create and evaluate statistical models. Apply useful statistical summary measures to descriptions and explanations.
- 5. Inferential Reasoning** Think inferentially, to recognize the limits of generalization. The basics of inferential statistics, sampling distributions, confidence intervals, p-values, hypothesis testing.

Modified from: Study „Using Interactive Tutors to Overcome Cognitive Learning Hurdles in Quantitative Reasoning“ (C.L. Cappell, 2001, Tech. Rep., Northern Illinois University)

Statistical Literacy Skills

- **Statistics is the language of data
and like a language, it can be used to argue....**

- **Some major argumentation skills:**
 - **Determine how data might confirm or challenge a conclusion**
 - **Assess statistical information used as evidence in an argument**
 - **Determine if a conclusion is based on an adequate/representative sample**
 - **Assess the meaningfulness of the collected data in view of the problem**
 - **Employ effective communication tools:
graphs, diagrams, trees, etc...**

Promotion of „Statistical Thinking“

- Tools
- Education
- Responsibility of Statistical Associations

Tools

Statistical Tools should:

- ◆ not present an definitive answer but let the data speak for itself
- ◆ be graphical, hence subjective and ambiguous
- ◆ always show the variability of the data
- ◆ be easy to understand, to use & to communicate with



they should stimulate statistical thinking!!

Tools

Statistical Tools should be easy to:

- ◆ Easy to understand: - black boxes are the ,opposite‘ of ,thinking‘
- ◆ Easy to use: - must be used by problem owner
- ◆ Easy to communicate: - must be graphics oriented (essential for communication & impact)

Examples: treatment comparison

Problem

Analyze the effect of 4 different diets on the blood coagulation time.

24 animals of two age categories distributed into 4 diet-groups (duration : 1 week)

Classical solution: Perform a 2-way ANOVA with factor interactions

Age: slightly not significant (P-value=6.7%)

Diet: highly significant (P- value <<1%)

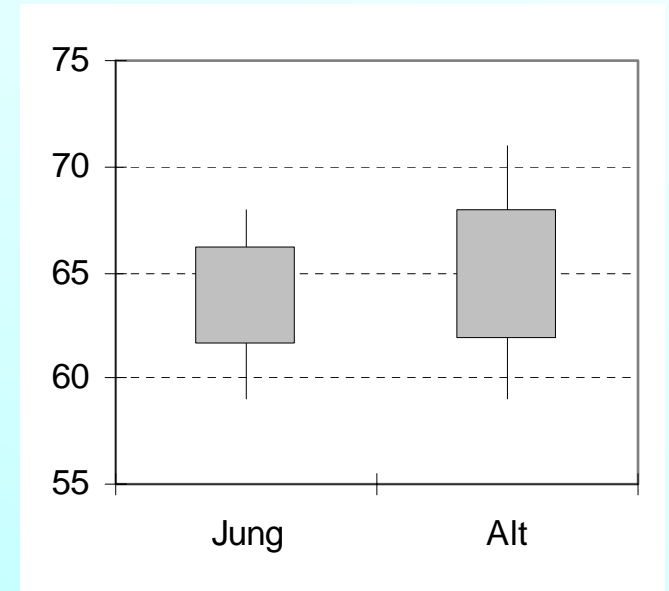
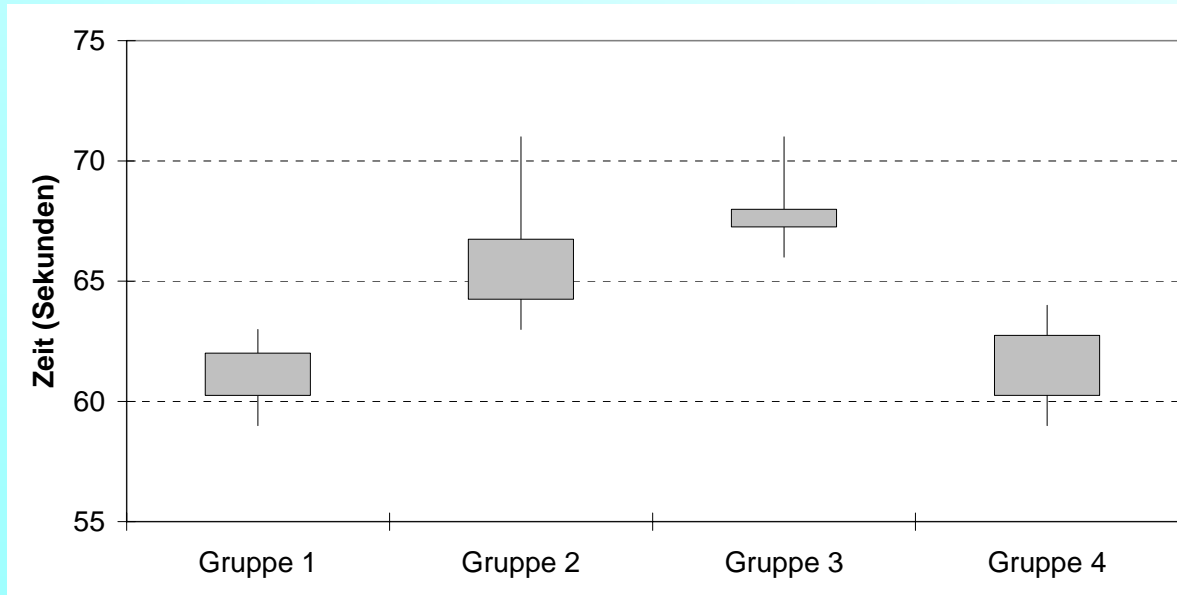
Interaction: not significant (P- value =33%)

	Behandlung			
Alter	Gruppe 1	Gruppe 2	Gruppe 3	Gruppe 4
J	62	63	68	59
J	60	64	67	62
J	63	67	66	60
A	59	71	68	61
A	61	65	71	63
A	62	66	68	64

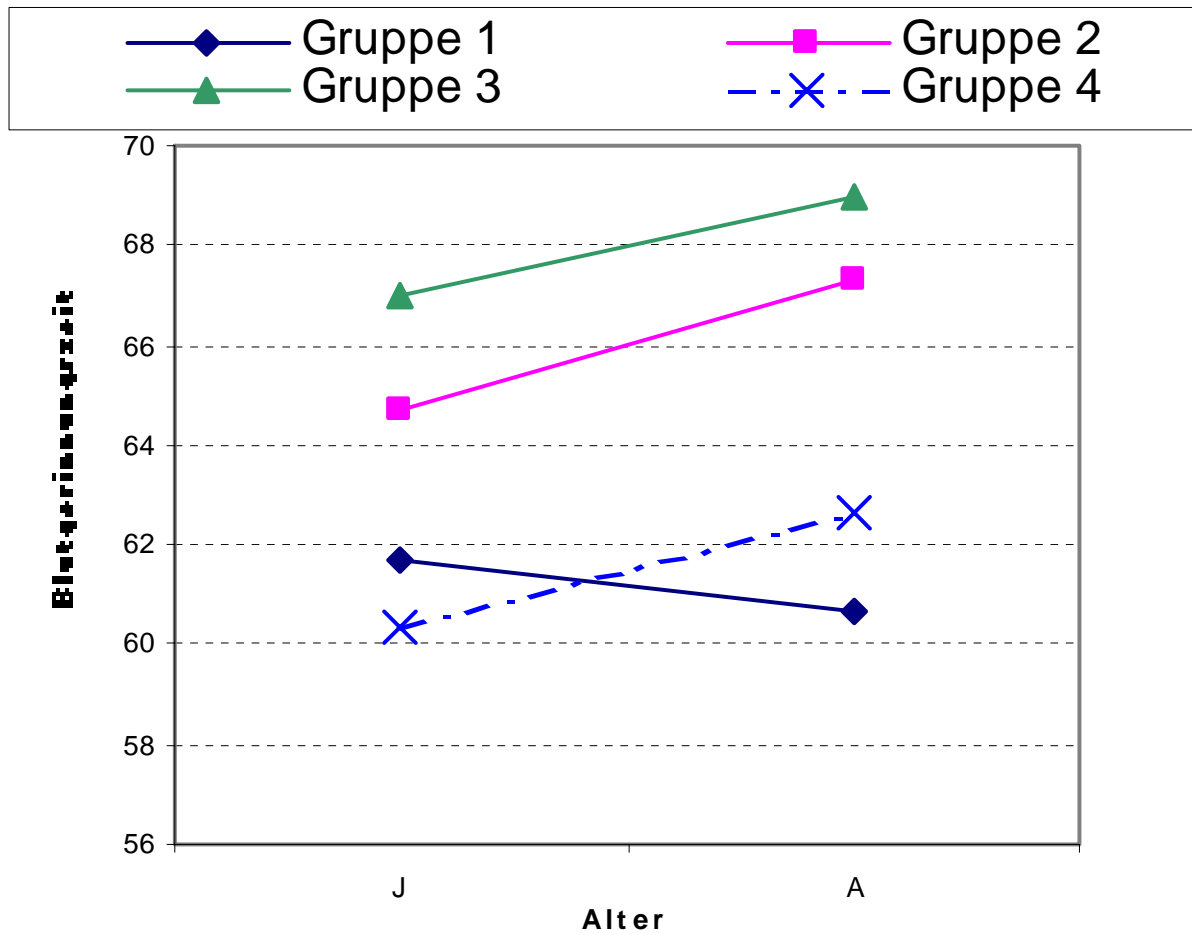


Don't give them ANOVA-Tables'!

Give them boxplots!



and interaction diagrams!



Education

Much has already been said

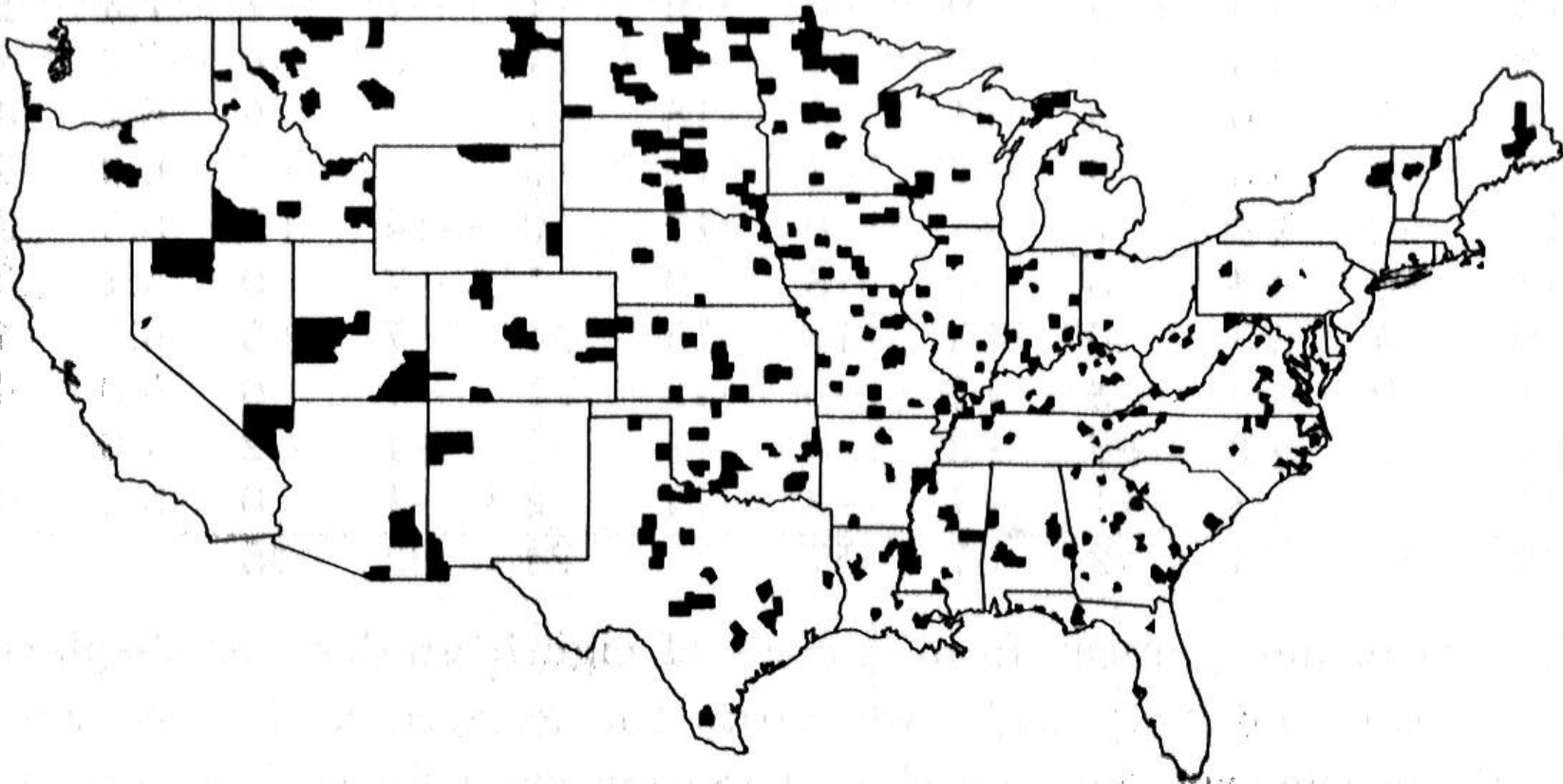
- Focus more on the reasoning and literacy skills
- Relevance of solving practical problems
- Obligation to know one other field of study where statistics is used as a research tool
- Teach the ideas and not the formulas
- More courses in sampling and design

What Educated Citizens should know (J. Utts)

Cause and Effect	When it can be concluded that a relationship is one of cause and effect, and when it cannot
Statistical Significance and Practical Importance	Especially relevant by large sample sizes
Low Power vs. No Effect	Especially relevant by small sample sizes. Impact of sample size on variability
Biases in Surveys	Awareness of sources of bias such as wording or order of questions, volunteer responses, socially desirable answers
Probable coincidences	Seemingly very improbable events are not uncommon because there are so many possibilities
Confusion of the inverse	$P(A B)$ vs $P(B A)$
Average vs. normal	Variability is natural. Normal is not the same as average.

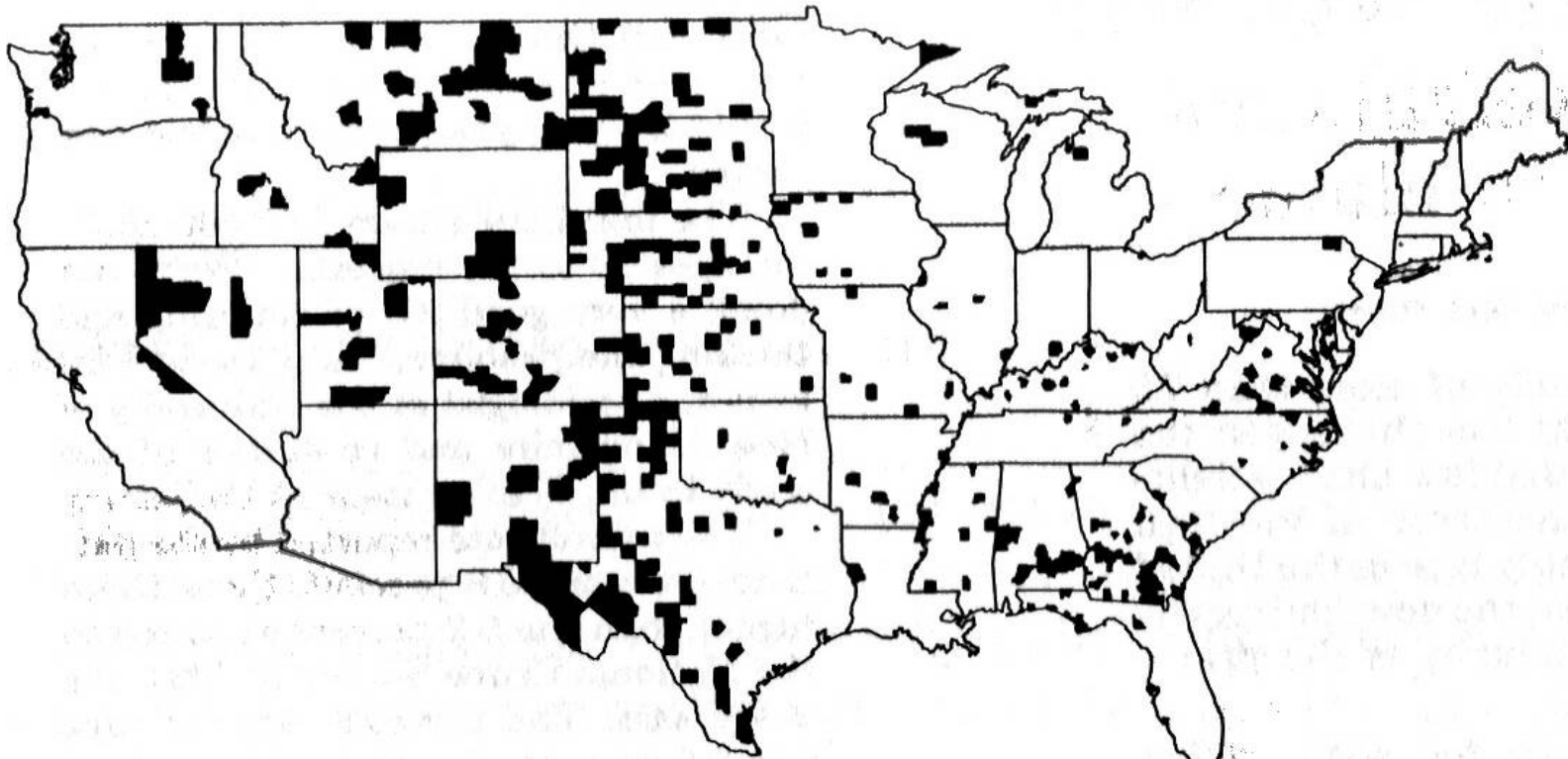
Testing statistical skills (example due to A.Gelman)

Highest kidney cancer death rates



Example (continued)

Lowest kidney cancer death rates



Statistical Professional Associations

The Promotion of Statistical thinking:

Raison d'être of an association

- Fight the misuse of Statistics in Medias and Politics
- Introduce the certification of Statisticians
- Explain Statistics to the Public (incl. Industry) / Education
-> Development of „awareness-programs“

A joint effort of academic and official institutions
is needed

Statistical Thinking & Quality: Society!

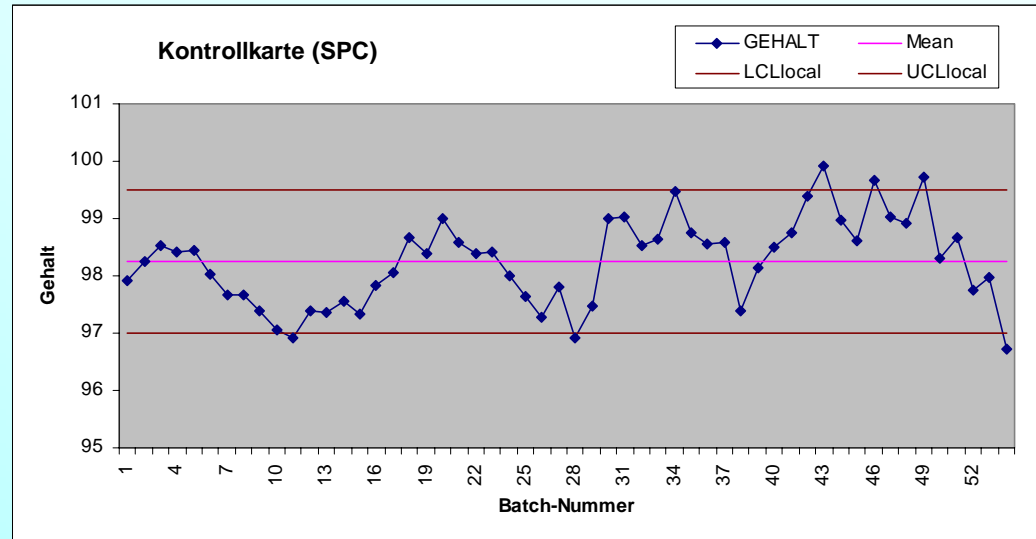
How Statistical Thinking is used can have dramatical consequences on Quality Issues

Some examples:

- Quality in Industry
- Quality in the Media
- Quality in Legal Issues

Quality in Industry

- A long tradition:
Shewart Control Chart:
(Separation of natural and special causes)
- QC during World War II



- TQM & the Quality Revolution in the 80's
Re-emergence of experimental design for solving industrial engineering problems
-> pro-active use of statistics for Continuous Quality Improvement
- Today: 6-Sigma program promotes statistical thinking
-> focus on the Omnipresence of Variation in industrial processes (also in services industry)

Conclusion: The 'Quality Revolution' and 6 Sigma are strong drivers for the promotion of Statistics

Quality in the Media

Headline: 'white males workers: an endangered species!'

Reason: Executive Summary of Report commissioned by the US Dept. of Labor

"Over the next 13 years only 15% of the new entrants in the workforce will be white males instead of 47% today"

However

this was NOT the original statement in the text!

because complex problem (many causes for changes in the workforce composition)

-> modeling of the flows of workers: "net addition to the workforce"

e.g.: white males: +13.5 Mio (in) -11.3 Mio (out) -> net addition of +2.2 Mio.

Because the number of females and non-white workers are growing faster than for the white males, those contributed only 15% of the net addition to the workforce.

Many errors:

- 1) Confusion between new entrants and net addition
- 2) Confusion comparing entrants with existing labor force
- 3) Lack of critical thinking: such numbers are nonsense!

Too late! A new statistics had taken a life of its own: mutant statistics! (J. Best)

Conclusion: 1. Medias take advantage of the lack of statistical thinking of the public
2. Never overestimate the understanding of an innumerate public!

Quality in Legal Issues

The infamous “Prosecutor’s Fallacy”

Confusion between

- the probability of innocence of a suspect
- and the probability of an apparently rare event accusing the suspect

Example: DNA Matching in Forensic Evidence

Expert says: Probability of match by pure chance: 1 in 10 Mio.

And the suspect is matching! --> highly suspect, isn't it?

However....

in a country with 50 Mio. about 5 individuals will show a match

-> the probability of innocence given a match is.... 80%

Conclusion:

Improperly used Statistical Thinking can have dramatic consequences
(remember the Sally Clark Case in England!)

The future

- The amount of data keeps increasing
→ demand for stat. tools is on the rise
- Our society is becoming more and more evidence based
(e.g. use of travellers data to detect terrorists)
- The media (TV, internet, newspapers) are making use of statistics as never before
(e.g. nonsense internet surveys)
- More and more data-based crucial decisions are taken that affect mankind ...

 **a sense of urgency!!!**

Conlusions

1. Statistics is wrongly perceived as a set of tools instead of a set of problems
2. Let us promote tools that stimulate statistical thinking and not just provide packaged solutions
3. Statistical thinking is an essential part of quantitative reasoning. Good educational programs should include the development of these skills (argumentation and reasoning).
4. Statistical literacy is becoming increasingly important in today's society.

References

J. Best (2001) *Damned Lies and Statistics*, University of California Press.C.

J.H. Friedman, *The Role of Statistics in the Data Revolution?*, *Int. Stat. Review* 69 (2001), 5-10.

A. Gelman & D. Nolan: *Teaching Statistics: a bag of tricks*, Oxford Univ. Press, 2002.

Y.L. Grize ISI-Berlin 2003: *Statistics for Scientists: just another toolbox*

C. Sagan (reprint 1997), *The Demon-Haunted World: Science as a Candle in the Dark*. Ballantine Books.

J. Utts (2003), *What Educated Citizens Should Know About Statistics and Probability*, *American Statistician* Vol.57, pp 74-79.

Abstract

Many applied statisticians have during their consulting activities experienced the fact that 'statistical thinking' is simply not recognized as a key element of the problem solving process. The reason may be that statistics is often wrongly perceived as a set of tools instead of a set of problems.

The promotion of statistical thinking is urgent and efforts in that direction should include:

- providing tools to users that encourage thinking and not black box solutions,
- offering appropriate educational programs that develop statistical reasoning and argumentation skills and do not focuss on technical issues.

The importance of 'statistical thinking' in the modern society will be stressed and examples relating to quality in industry, law and the medias will be shown.

Professional statistical organizations are challenged to take a much more pro-active role in these matters. Statistics, Science and Society have all much to gain from a better appreciation of 'statistical thinking'.