Adding Context to Introductory Statistics

MILO SCHIELD

Augsburg College

Member: International Statistical Institute

US Rep: International Statistical Literacy Project

Director, W. M. Keck Statistical Literacy Project

May 22, 2013

Slides at www.StatLit.org/pdf/2013Schield-ASA-TC6up.pdf

Statistics Education has "issues"

1. Students see less value in statistics after finishing the intro statistics course than before they started.
2. Six months after completing a statistics course, students forget half of what they learned.
3. Statistics courses are largely irrelevant—not just boring or technically difficult, but irrelevant. Enhrenberg (1954)
4. "become more difficult to provide an agreed-upon list of … topics … that all students should learn." Pearl et al (2012).

Why does Introductory Stats have these Issues?

Traditional introductory statistics courses focus on variability – they are not math courses.

But they don't focus on context. Once the median is jettisoned in place of the mean, context is absent.

The lack of context may explain:

- why students see less value after a course than before.
- why students forget half of what they learn in 6 mos.
- why students consider statistics irrelevant.
- why statistical educators cannot agree on topics.

Thesis

Adding context to introductory statistics will

- uphold context as the essence of statistics (e.g., statistics are numbers in context),
- more clearly separate statistics as a liberal art from mathematical statistics,
- improve student retention of key ideas, and improve student attitudes on the value of statistics.

Consider five examples of context influencing statistics

Influence of Context #1: Subject Bias*

When asked their income, men over-stated by about 10% on average; women told the truth.

When asked their weight, women understated by 10# on average; men typically told the truth.

* Made-up statistics to illustrate the point.

Influence of Context #2: Defining Groups or Conditions

Number of US children with elevated lead:

- 27,000 in 2009
- 259,000 in 2010

CDC changed the standard in 2010 from 10 micrograms of lead per dl of blood to five.

www.cdc.gov/nceh/lead/data/StateConfirmedByYear1997-2011.htm

Influence of Context #3: What is taken into account

- The chance of a run of k heads in n flips of a fair coin depends on the context: “place pre-specified” versus “somewhere in the series.”
- The accuracy of a medical test depends on the context: confirming versus predicting.
- The predictive accuracy of a medical test depends on the context: the percentage of subjects tested that have the disease.

Influence of Context #4: Choice of Population

In predicting or explaining grade differences among first-year college students:

- SAT scores do a poor job for students at colleges that admit a narrow range of scores (highly selective colleges).
- SAT scores do a good job for students at colleges that admit a wide-range of scores.

Influence of Context #5: Confounding

The male-female difference in median* **weights** among 20-year-olds is 27 pounds.

27#: Male median wt: 156#; Female median wt: 129#

Male median height: 70"; Female median height: 64"

Median weight of 70” high females is 142# est.

The male-female difference in median weight for 20-year olds is 14 pounds **after controlling for height**.

* www.cdc.gov/growthcharts/html_charts/bmiagerev.htm

Influence of Context on Statistical Significance

The foregoing shows how context can influence a statistic, but the focus of the intro statistics course is statistical significance.

Q1. Can we show how each of these can influence statistical significance??? **ABSOLUTELY!!!**

Q2. Can it be done with minimal math and time?

ABSOLUTELY!!! Do everything with tables and confidence intervals. Non-overlap means statistical significance.

Influence of Bias on Significance

Response bias: Men likely to overstate income

$5,000 is the 95% margin of error					
Income	Men	Women	Diff	Overlap	Stat. Sig
Stated	$62,000	$51,000	$11,000	No	Yes
Actual	$53,000	$51,000	$2,000	Yes	No

Sample bias: Rich less likely to do surveys

$3,000 is the 95% margin of error					
Income	Men	Women	Diff	Overlap	Stat. Sig
Responders	$53,000	$51,000	$2,000	Yes	No
Population	$62,000	$55,000	$7,000	No	Yes

Influence of Assembly on Significance

Two definitions of "bullying"

Middle-school kids 5% is the 95% margin of error

BULLYING	Boys	Girls	Diff	Overlap	Stat.Sig
1) Physical only	40%	10%	30%	No	Yes
2) Physical & Social	42%	40%	2%	Yes	No

Two ways to combine subgroups to form groups

6% is the 95% margin of error

Fishing	Dislike	Neutral	Like	% who like*	% who like**
Men	30%	30%	40%	40%	70%
Women	50%	20%	30%	30%	50%
			Overlap	Yes	No
			Statistical significance	No	Yes

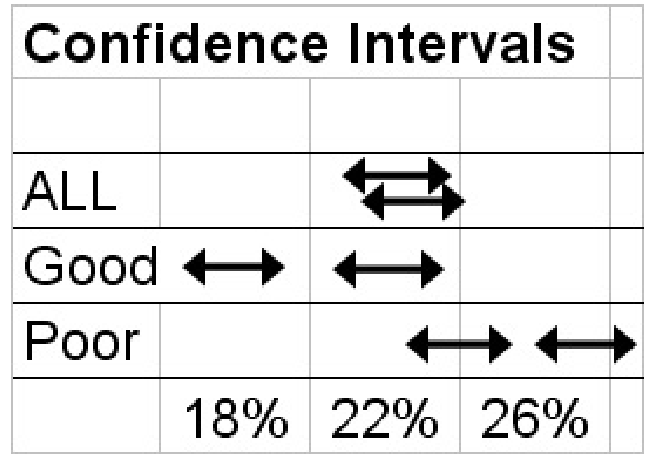
* Exclude neutral

** Include neutral

Confounder Influence: Insignificance to Significance

Necessary: Confounding must increase gap.

Death	1.4%	Margin of Error (95%)			
Rate	City	Rural		Diff	Compare
ALL	22.8%	23.0%		0.2%	Basis
Good	18.0%	22.0%		4.0%	larger
Poor	24.0%	27.0%		3.0%	larger



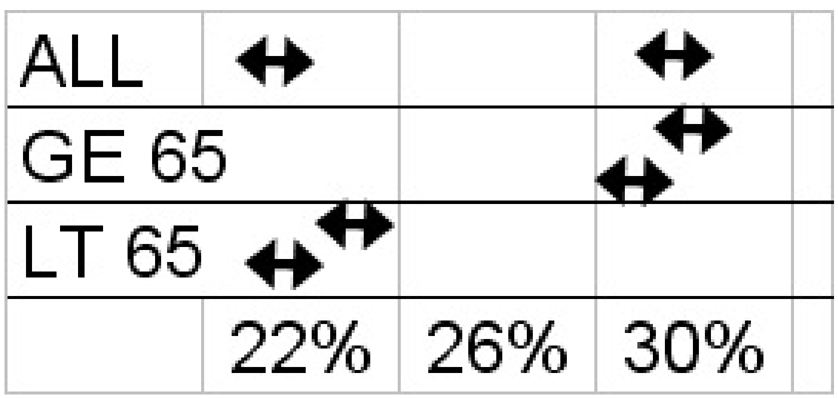
Theorem: If the confidence intervals don't overlap for the two values of the binary confounder and the order never reverses, then the confidence intervals at any standardized value will not overlap.

Confounder Influence: Significance to Insignificance

Necessary:

Confounding must decrease the predictor gap.

Location & age	1.5%	The 95% Margin of Error		
Death Rate	City	Rural	Diff	Compare
ALL	22.7%	29.4%	6.7%	Standard
Over 65	29.0%	30.0%	1.0%	smaller
Under 65	22.0%	24.0%	2.0%	smaller



Conclusion #1

To uphold statistics as mathematics with a context, the introductory statistics course must be redesigned.

The intro course needs much more focus on big ideas:

- **Context** (what is controlled), **assembly** (definitions) and **bias** are big ideas for non-statisticians.
- **Randomness** and **statistical significance** are big ideas for statisticians.
- **Seeing how confounding, assembly and bias can influence statistical significance should be central for a "statistics-in-context" course**.

Conclusion #2

Thesis: Adding context to introductory statistics will

- improve student retention of key ideas,
- improve attitudes on the value of studying statistics,
- uphold context – not variability – as the essential difference between statistics and mathematics.

Since this can be done with minimal math and very little time, the introductory statistics course should be re-designed as a "statistics-in-context" course!

References

ASA (2012). GAISE Report.

Ehrenberg, A. S. C. (1976). We must preach what is practised: a radical review of statistical teaching. *Journal of the Royal Statistical Society*, Series D, 25(3),195–208.

Pearl, D., Garfield, J., delMas, R., Groth, R., Kaplan, J. McGowan, H., and Lee, H.S. (2012). Connecting Research to Practice in a Culture of Assessment for Introductory College-level Statistics. www.causeweb.org/research/guidelines/ResearchReport_Dec_2012.pdf

Schield, M. (2006). Presenting Confounding and Standardization Graphically. *STATS Magazine*, American Statistical Association. Fall 2006. pp. 14-18. Copy at www.StatLit.org/pdf/2006SchieldSTATS.pdf.