Statistical Analysis of Science Education Data -
What Does It Mean to Quantify Data?
Joseph M. Wisenbaker, Ph.D.

joe_wisenbaker@charter.net
Issues in Conducting Quantitative Research in Science Education
Handouts

• Short list of suggested readings (with links)
• Guide to finding the ‘right’ statistical test (with link to more complete source)
• Background related to thinking about statistical power (or, How many subjects do I need?)
Suggested Readings

• NOT comprehensive by any means
• Some things that jumped out for me as I explored possible things to talk with you about
Suggested Readings

• When Numbers Mislead - NYTimes SundayReview opinion page column by Stephanie Coontz

• http://www.nytimes.com/2013/05/26/opinion/sunday/when-numbers-mislead.html?src=recg
• How averages can mislead (and many of the kinds of statistics of interest are often averages or highly related to them)
• Average might not be very Typical
• Effects of outliers can be profound
Case in point:

In 2011, for example, the average income of the 7,878 households in Steubenville, Ohio, was $46,341. But if just two people, Warren Buffett and Oprah Winfrey, relocated to that city, the average household income in Steubenville would rise 62 percent overnight, to $75,263 per household.
• Combined household incomes of Warren and Oprah for 2011 must have been something like $230,834,842!
Coontz

- Interesting graphic

mean

variation

teased for being different or the tyranny of the average
• There are often many important facts about data beyond simple averages
• Variation is a very big one
Suggested Readings

• The Role of Qualitative Research in Science Education by Iztok Devetak, Saša A. Glažar and Janez Vogrinc in Eurasia Journal of Mathematics, Science & Technology Education, 2010, 6(1), 77-84.

• www.ejmste.com/v6n1/EURASIA_v6n1_Devetak.pdf
Devetak, et al

• Their interest was in the extent to which qualitative methods were used in articles appearing between 2006-08 in three prominent science education journals:

• Journal of Research in Science Teaching, Science Education and the International Journal of Science Education
They classified the 461 articles published there as being qualitative, quantitative, mixed or theoretical/review.

Their findings:
Devetak, et al

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Relative Prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualitative</td>
<td>26%</td>
</tr>
<tr>
<td>Mixed</td>
<td>45%</td>
</tr>
<tr>
<td>Quantitative</td>
<td>21%</td>
</tr>
<tr>
<td>Theoretical</td>
<td>8%</td>
</tr>
</tbody>
</table>

- Their interest - 71% used qualitative in some way
- But, 66% made some use of quantitative methods
So, it seems reasonable to conclude that quantitative approaches find wide use among science education researchers either alone or in combination with qualitative approaches.
Suggested Readings

• One of the articles I found reporting the results of quantitative research in science education was Effects of Small-Group Learning on Undergraduates in Science, Mathematics, Engineering, and Technology: A Meta-Analysis by Leonard Springer, Mary Elizabeth Stanne, and Samuel S. Donovan.
• This was a nice example illustrating one of the ways in which quantitative methods can find application in science education research - a meta-analysis of research looking at small group instruction
Suggested Readings

• The other two articles in my short list of suggested readings are each reviews conducted by the U.S. Department of Education’s Institute of Education Sciences in their What Works Clearinghouse.

• These are illustrative of very quantitative work in the context of an organization that often serves as a gatekeeper for federal dollars for education.
What Works Clearinghouse

• Those kinds of reports are based on strict review criteria carefully made public in their WWC Procedures and Standards Handbook and, given their purpose, are often far more demanding than most reviewed journals. As such, they provide some interesting guidance that you all might want to be aware of as you design and implement your own research.
WWC Criteria

  
• Design
• Measurement
• Analysis
WWC Criteria

• There are five categories of studies rejected for inclusion in one of their reviews (that comes with a carefully defined protocol):
  • The study is not a primary analysis of the effect of an intervention.
  • The study does not have an eligible design.
  • The study does not use a sample aligned with the protocol.
WWC Criteria

• The study does not include an outcome within a domain specified by the protocol.

• The study was not published within the relevant time frame.
WWC Criteria

• Focus:
  • Focus must be on looking at the impact or effectiveness of an intervention.

• Design:
  • Experimental design using randomized control trials (strongest evidence)
  • Quasi-experimental design with comparisons to ‘units’ that are clearly similar or sufficiently similar to allow for statistical adjustments
Potential confounding effects that can eliminate studies include instruction delivered by a single teacher, in a single classroom, school or district.

Also, where teachers get to choose whether to deliver an intervention or are chosen due to particular factors (experience, education, etc.).
WWC Criteria

- Measurement of outcome must:
  - demonstrate either face validity or reliability,
  - not be over-aligned with the intervention, and
  - be collected in the same manner for both intervention and comparison groups.

- Standardized tests relevant to the topic are assumed to meet these criteria.
WWC Criteria

• WWC Analysis (assuming your analyses were correct):
  • One focus - were your comparisons of experimental and control groups statistically significant
  • Another focus - estimates of effect size for continuous or binary outcomes
  • Finally, estimates of improvement indices where standardized tests are used that have percentile ranks available
Methodological Considerations

- Adequate sample size leading to adequate levels of statistical power in the design of a quantitative is essential.

- Tools to help in determining sample size
  - [https://www.dssresearch.com/KnowledgeCenter/toolkitcalculators/statisticalpowercalculators.aspx](https://www.dssresearch.com/KnowledgeCenter/toolkitcalculators/statisticalpowercalculators.aspx)
  - A great repository of stat related links! [http://statpages.org](http://statpages.org)
  - Sites for power/sample size calculation including: [http://homepage.stat.uiowa.edu/~rlenth/Power/index.html](http://homepage.stat.uiowa.edu/~rlenth/Power/index.html)
  - a very good free program for PCs and Macs: [http://wwwpsycho.uni-duesseldorfd.de/abteilungen/aap/gpower3/](http://wwwpsycho.uni-duesseldorfd.de/abteilungen/aap/gpower3/)
Methodological Considerations

• Statistical Analysis Software
  • Commercial: SPSS, SAS, STATA
  • Free/open source:
    • OpenEpi: [http://www.OpenEpi.com/v37/Menu/OE_Menu.htm](http://www.OpenEpi.com/v37/Menu/OE_Menu.htm)
    • R with R-Commander
R version 3.0.1 (2013-05-16) -- "Good Night, Everyone"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin10.8.0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for more information.
Type 'citation()' for more information on citing R.
Type 'demo()' for some demos, 'help()' for help on using RStudio.
Type 'q()' to quit R.

[ R.app GUI 1.61 (6492) x86_64-apple-darwin10.8.0 ]
[ History restored from /Users/wisniewski/RData/...

> library(Rcmdr)
Loading required package: car
Loading required package: MASS
Loading required package: nnet

Rcmdr Version 1.9-6
>

> data(Animals, package="MASS")
> library(relimp, pos=4)
> showData(Animals, placement='20+200', font=getRcmdr('logFont'), maxwidth=80,
> + maxheight=30)
> cor(Animals[, c("body", "brain")], use="complete.obs")

body  brain
body 1.000000000 -0.005341163
brain -0.005341163 1.000000000

Some Deep Background Affecting the Validity of Inferential Tests and Estimates of Error

• Assumptions of population normality, homogeneity of variance and independence among analysis ‘units’ form the basis for many inferential tests.

• With enough ‘subjects’ per group (usually 30 or more), normality ceases to be a problem affecting the validity of statistical inferences though you may well still need more for adequate power)!
Deep Background ...

• Homogeneity of variance is less an issue with equal group sizes, but can be addressed explicitly through using non-parametric methods.

• Lack of independence among your study ‘units’ can be catastrophic! Actual alpha levels/risk of Type I error can be double or triple what you might think - and the larger your sample size, the worse this can be.
Deep Background ...

• What can lead to a lack of statistical independence?
  • Students interacting with each other as part of your treatment or incidental to it.
  • The presence of one or more disruptive students that detract from instruction and learning.
  • The presence of one or more students who ask particularly powerful questions.
Deep Background ...

- Teachers delivering your treatment (or in the control condition) who are differentially effective.

- Potential solutions:
  - Conduct your study with teacher/classroom as the ‘unit of analysis’.
  - More complex school/district issues may call for using school or district as the ‘unit of analysis’
  - This may require multi-site research as ‘sample size’ here becomes the number of teachers/classrooms/schools districts involved.