

Suggestions for Analysis & Reporting of TAH Program Data

John Hansen, M.S.

Introduction

The following is a set of general guidelines to be used for general data analysis under general circumstances. The most appropriate statistical analysis plan for any one set of data can only be derived on a case by case basis with considerations given to all the salient facets of the investigation. For specific recommendations as to appropriate analysis for specific research questions we direct the reader to resources including: (Michael, 2005); (Campbell and Stanley, 1966); (Cook and Campbell, 1979). The recommendations given here are based on commonly understood protocols and metrics in the statistics and psychometric literature.

In this module we will be outlining a general approach to statistical analysis that can be used regarding a general research question about Teaching American History.

The research questions under investigation are:

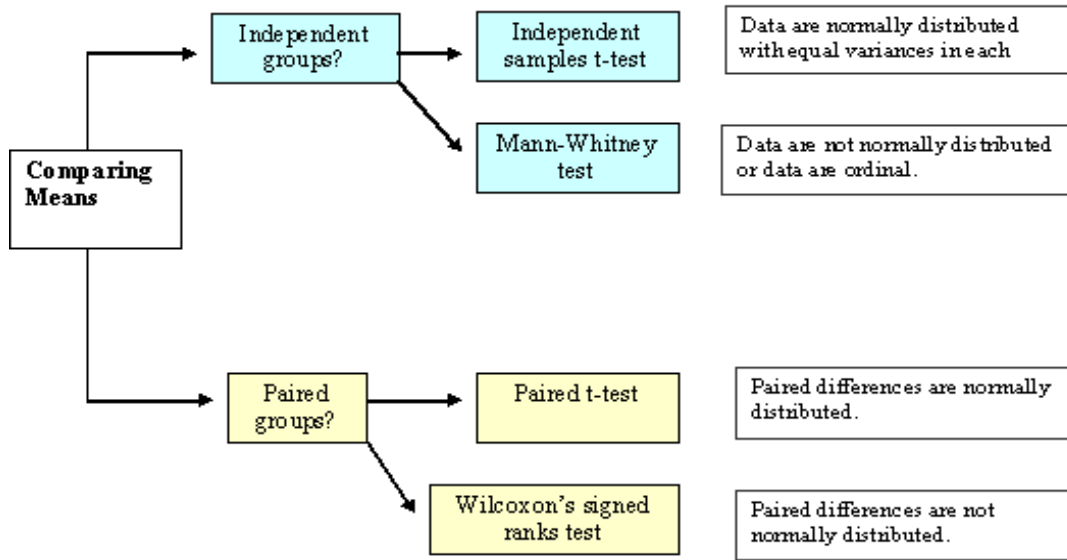
1. How do teachers (Level 2) who receive enrichment training differ with respect to their peers who do not receive this training?
2. How do students (Level 1) of teachers who receive enrichment training differ with respect to students of teachers who do not receive this training?

Three conditions for comparing group differences will be presented in this module.

1. Small sample size of both treatment teachers and control teachers.
2. Large sample size of both treatment teachers and control teachers.
3. Unequal sample size of treatment teachers and control teachers.

This discussion is followed with a basic flowchart, Figure 1.0, for use by researchers in determining appropriate statistical tests for data with qualities that go beyond those presented in this guide

Figure 1.0 General Guidelines for Selection of Statistical Significance Test



1.0 Small sample size of both treatment teachers and control teachers.

To address the two research questions under investigation:

1. How do teachers (Level 2) that receive enrichment training differ with respect to their peers that do not receive this training?
2. How do students (Level 1) of teachers that receive enrichment training differ with respect to students of teachers that do not receive this training?

We consider the hypothetical situation where:

1. Six teachers received enrichment training and six teachers received no enrichment training, Table 1.0.
2. Each of the teachers in the experiment has a classroom of 25 students, Table 1.0.
3. Both teachers and students have pre/post scores on a performance instrument.

Table 1.0 Experimental Design with Small Sample of Teachers (Level 2)

Treatment (enrichment training)	Control (no enrichment training)
Teachers $n_{tt} = 6$	Teachers $n_{tc} = 6$
Students $n_{st} = 150$	Students $n_{sc} = 150$

Analysis of these sample data at level 2, teachers, would amount to only a comparison of descriptive statistics. That is, with a small sample of only six observations per group there is no universe to which generalizations can be made. Additionally, if an inferential comparison was made, the denominator in the ratio of mean difference to standard error, would likely be large enough to eclipse any difference found in the means (Cohen, 1988). Descriptive statistics will allow for the researcher to communicate the means and standard deviations of the two teacher groups in their pretest scores, posttest scores, and in the normalized gain score that represents change.¹

Analysis of these sample data at level 1, students, would entail both the reporting of descriptive statistics as well as a test for statistical significance for the difference between mean scores for students of teachers that received enrichment training and students of those teachers who did not. This test of difference in means would be an independent t-test using the normalized gain scores from students pre to post examination.

These analysis do not take into effect classroom by classroom variation due to the prohibitively small sample size which would not allow for developing a multi-level regression model.

1. See Hake, R. (1998), Relationship of Individual Student Normalized Learning Gains in Mechanics with Gender, High-School Physics, and Pretest Scores on Mathematics and Spatial Visualization. A copy of this article is distributed by the author at: <http://www.physics.indiana.edu/~hake/PERC2002h-Hake.pdf>



1.1 Large sample size of both treatment teachers and control teachers.

To address the two research questions under investigation:

1. How do teachers (Level 2) that receive enrichment training differ with respect to their peers that do not receive this training?
2. How do students (Level 1) of teachers that receive enrichment training differ with respect to students of teachers that do not receive this training?

We consider the hypothetical situation where:

1. Greater than thirty teachers received enrichment training and greater than thirty teachers received no enrichment training, Table 1.1.
2. Each of the teachers in the experiment has a classroom of 25 students, Table 1.1.
3. Both teachers and students have pre/post scores on a performance instrument.

Table 1.1 Experimental Design with Large Sample of Teachers (Level 2)

Treatment (enrichment training)	Control (no enrichment training)
Teachers $n_{tt} = 30$	Teachers $n_{tc} = 30$
Students $n_{st} = 750$	Students $n_{sc} = 750$

Given a sample as described in Table 1.1, analysis of these data at level 2, teachers, would allow for both the reporting of descriptive statistics as well as a test for statistical significance for the difference between mean scores for teachers that received enrichment training and those teachers who did not. This test of difference in means would be an independent t-test using the normalized gain scores from teachers pre to post examination.

Analysis of these data at level 1, students, would include both the reporting of descriptive statistics as well as a test for statistical significance for the difference between mean scores for students of teachers that received enrichment training and for students of those teachers who did not receive the training. This test of difference in means would be an independent t-test using the normalized gain scores from students pre to post examination.

Finally, these data would lend themselves to multi-level modeling with would take into account classroom by classroom variation. The procedure for using a multilevel model as well as for interpreting the results are detailed in the four module series by Hansen (2005).¹

1. For additional considerations regarding sample size and multilevel modeling readers are directed to (Maas & Hox, 2003).



1.2 Unequal sample size of treatment teachers and control teachers.

To address the two research questions under investigation:

1. How do teachers (Level 2) that receive enrichment training differ with respect to their peers that do not receive this training?
2. How do students (Level 1) of teachers that receive enrichment training differ with respect to students of teachers that do not receive this training?

We consider the hypothetical situation where:

1. Thirty teachers received enrichment training and six teachers received no enrichment training, Table 1.2.
2. Each of the teachers in the experiment has a classroom of 25 students, Table 1.2.
3. Both teachers and students have pre/post scores on a performance instrument.

Table 1.2 Experimental Design with Unequal Sample of Teachers (Level 2)¹

Treatment (enrichment training)	Control (no enrichment training)
Teachers $n_{tt} = 30$	Teachers $n_{tc} = 6$
Students $n_{st} = 750$	Students $n_{sc} = 150$

Analysis of these sample data at level 2, teachers, would amount to only a comparison of descriptive statistics. That is, with a small sample of only six observations in the control group the universe to which we are comparing the treatment teachers is ill defined. Additionally, if an inferential comparison was made, the denominator in the ratio of mean difference to standard error, would likely be large enough to eclipse any difference found in the means (Cohen, 1988).² Descriptive statistics will allow for the researcher to communicate the means and standard deviations of the two teacher groups in their pretest scores, posttest scores, and in the normalized gain score that represents change.

Analysis of these sample data at level 1, students, would entail both the reporting of descriptive statistics as well as a test for statistical significance for the difference between mean scores for students of teachers that received enrichment training and students of those teachers who did not. This test of difference in means would be an independent t-test using the normalized gain scores from students pre to post examination.

These analysis do not take into effect classroom by classroom variation due to the prohibitively small sample size for teachers that did not receive training. These data would not warrant developing a multi-level regression model.

-
1. Recommendations presented here also encompass the inverse scenario where teachers receiving treatment number six and teachers in the control group number thirty.
 2. For comparison of teacher treatment and control groups when group sizes are unequal and of size greater size than that presented in this section we direct the reader to (Yang, Sackett, & Arvey, 1996).

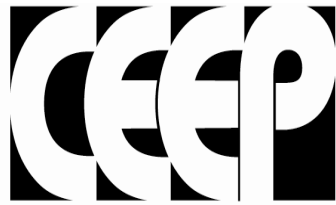


Works Cited

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, second edition. Erlbaum. Mahwah, NJ.
- Campbell, D.T. & Stanley, J.C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally College Pub. Co.
- Cook, T.D. & Campbell, D.T. (1979). *Quasi-experimentation: design and analysis issues for field settings*. Chicago: Rand McNally College Pub. Co.
- Hansen, J. (2005). *Multilevel regression models*. Internal publication. Bloomington, IN: Center for Evaluation and Education Policy.
- Maas, C.J.M., Hox, J.J. (2003). *Sufficient sample sizes for multilevel modeling*. Department of Methodology and Statistics, Utrecht University, NL, submitted for publication.
- Michael, R. (2005). *Teaching American history: Data analysis resources*. Internal publication. Bloomington, IN: Center for Evaluation and Education Policy
- Yang, H., Sackett, R. & Arvey, R. (1996). *Statistical power and cost in training evaluation: Some new considerations*. *Personnel Psychology*, 49, 651-668.



Produced by:



**CENTER FOR EVALUATION
& EDUCATION POLICY**

On the web at:

www.ceep.indiana.edu

