Institutional Comparisons of NSSE Results: Alternatives to t-tests and Cohen's d for Discrete Ordinal Data

Presented by

Jeanine Romano

University of Tampa

James Coraggio

University of South Florida

Florida Association for Institutional Research, Cocoa Beach, Florida, 2006

Paper

Appropriate statistics for ordinal level data : Should we really be using t-test and Cohen's d for evaluating group differences on the NSSE and other surveys?

- **Authors:**
- **Jeanine Romano, University of Tampa**
- **Jeffrey D. Kromrey, University of South Florida**
- **Jesse Coraggio, University of South Florida**
- **Jeff Skowronek, University of Tampa**

Statistics is like a bikini.
What they reveal is suggestive.
What they conceal is vital.
(Arthur Koestler)

Introduction

- This presentation is a culmination of research performed by four individuals interested in the affect of using *t*-tests and Cohen's *d* on ordinal survey data.
- The completed research paper provides a more in-depth analysis of the issues and explanations of alternative statistical methodologies.
- The paper will be provided to participants on request.

Purpose

- Are statistics such as *t*-tests and Cohen's *d* appropriate to evaluate differences between group responses on ordinal-type survey data such as the *National Survey of Student Engagement* (NSSE)?
- What other statistics may be more appropriate for such analysis?

Levels of Measurement

- Stevens (1946) proposed that there are four levels that classify the nature of information

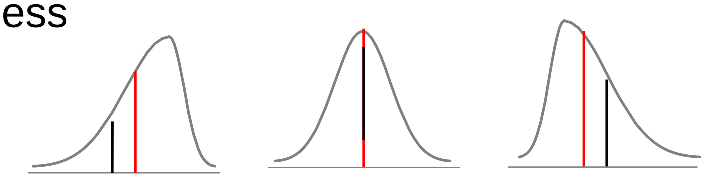
Nominal	Characteristic is used to classify individuals into categories which are not meaningfully ordered
Ordinal	Characteristic is used to order or rank individuals
Interval	Characteristic is used to order individuals, and the distances between numbers are equal
Ratio	Characteristic is used to order individuals, the distances between numbers are equal, and the 0 is meaningful

Identifying Differences (*t*-tests)

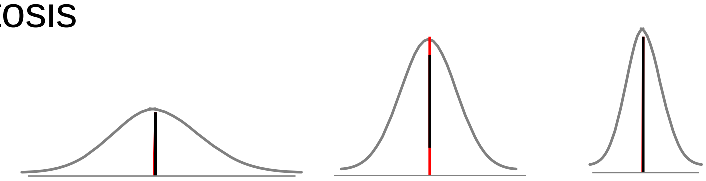
- A parametric test used to determine whether or not two independent sample means are significantly different from one another.
- Data should be interval or ratio level
- Assumptions:
  - Distributions are normal
  - Variances are homogeneous
  - Observations are independent samples

Normality Assumption

- Skewness



- Kurtosis



Equal Variances

- If this assumption is violated, there is little value in pooling the variances and the interpretation of any *t*-test results can be misleading or completely uninformative.
- There are a number of analyses that can be used to determine if the homogeneity assumption has been violated, including Hartley's *F*-max test or Levene's equality test.

Variance Violation Example

- Results of this analysis would lead a researcher to conclude there is a significant difference between the means of the two samples under investigation.

t-value	df	*p-value*
2.034	50	.047

Variance Violation Example

- When a test for equal variances is analyzed, it becomes clear that some adjustments need to be made.
- Results of Levene's test for equal variances, i.e. the significant *p*-value indicates that the variances are unequal and therefore the assumption is violated.

F value	*p*-value
4.563	.038

Variance Violation Example

- An adjustment must be made to the degrees of freedom corresponding to the *t*-test.
- With the adjustment made to the degrees of freedom, the results of the *t*-test were the following:

t-value	Adjusted df	*p*-value
1.815	27.326	.080

- After the adjustment is made the researcher would now fail to reject the hypotheses that the means are equal.

Identifying Differences (Chi-Square)

- A non-parametric test used to determine whether the two sample proportions (or frequencies) are different from one another
- More appropriate test for categorical or qualitative data (*NSSE* data)
- Assumptions:
  - Observations are independent samples
  - Categories are mutually exclusive and "exhaustive"
  - Minimum number in each cell (5)

Chi-square Comparisons to *t*-test

- No emphasis on a particular parameter (non-parametric)
- Data in analysis can be ordinal
- No normality or homogeneity of variance Assumption

Contingency Table

- Frequency data are generally displayed in a contingency table format.

	Agree	Disagree	Total
Group 1	*80*	*20*	*100*
Group 2	*50*	*50*	*100*
Total	*130*	*70*	*200*

Chi-square Calculation

- The chi-square is calculated with the following formula:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Where O_i represents the observed frequency, E_i represents the expected frequency (see below), and k represents the number of cells in the table.

$$E = \frac{row\ total * column\ total}{total}$$

Dichotomizing Data

- For purposes of the analysis individual item data were analyzed in both original response categories and artificially categorized into two groups (chi-square-2, log-odds, and raw differences),
  - Even categories (4 option example): *Strongly Agree* and *Agree* into top half and *Disagree* and *Strongly Disagree* into lower half
  - Odd categories (5 option example): *Strongly Agree* and *Agree* into top half and *Undecided*, *Disagree,* and *Strongly Disagree* into lower half

Odds Ratio

- A measure of the strength of relationship between two variables of interest.
- Also more appropriate test for categorical or qualitative data (*NSSE* data)
- Odds ratio (OR) is essentially the ratio of the odds for each individual event.

Odds Calculation

- Probability (or number) of the event (*p*) divided by the probability (or number) of the non-event (*q*). Odds = *p/q*
- For example, to calculate the odds of a pregnant woman having a boy…
  - Divide the number of boy births (51 male births in 100) by the number of non-boy births (49 female in 100).
  - So the odds of having a baby boy would be 51/49 or 1.04.

Odds Calculation

- If the odds are greater than 1, than the event is more likely to occur.
- Odds are not the same as the probability of occurring.
- If the odds =1, then the probability of the event occurring is the same as the probability of the event not occurring.
- Odds ratios are essentially the ratio of two independent events

Odds Ratio Calculation

- Divide the odds in one group (group of interest) by the odds in the other group (control group)
- OR = $odds_1/odds_2$
- Male Birth rate in Country 1 is 1.04, Country 2 is 1.22, then OR is 1.04/1.22 or .852. Odds are greater in Country 1

Odds Ratio Interpretation

- OR must be a non-negative number
- If OR = 1 then two groups have the same odds of occurring
- If OR > 1 then the group of interest has higher odds of occurring
- If OR <1 then the control or comparison group has higher odds of occurring

Odds Ratio Statistical Significance

- OR can be used as a significance test to evaluate population differences using this Confidence Interval (CI) formula (see paper)

$$\log(\hat{\theta}) \pm z_{1-\alpha/2}\sqrt{\frac{1}{n_{11}}+\frac{1}{n_{10}}+\frac{1}{n_{01}}+\frac{1}{n_{00}}}$$

- If the created CI includes 1 then there is no “statistical difference” between the two groups.

Cohen’s *d*

- A traditional measure of effect size is the standardized mean difference (Cohen’s d or Hedge’s g). These effect sizes may be used to describe differences in means relative to an assumed common variance. Cohen’s d is given by

$$d = \frac{\bar{X}_1 - \bar{X}_2}{\hat{\sigma}}$$

- Where $\hat{\sigma}$ is the pooled standard deviation.

Cohen's *d*

- This effect size index represents the difference between sample means in standard deviation units.
- That is, an effect size of 1.0 simply indicates that the sample means are one standard deviation apart

Interpreting Cohen's *d*

- Cohen's d = .20 small effect
  - "Noticeably smaller than medium but not so small as to be trivial"
- Cohen's d= .50 medium effect
  - "An effect likely to be visible to the naked eye of a careful observer"
- Cohen's d= .80 large effect
  - A large effect size is the "same distance above medium as small was below it."

Cliff's Delta

- Cliff's delta represents the degree of overlap between the two distributions of scores.
- More appropriate for ordinal data
- It is non-parametric and has no assumptions about underlying distributions (normality or homogeneity of variance)

Cliff's Delta Calculation

- A sample estimate of the parameter is obtained by enumerating the number of occurrences of an observation from one group having a higher response value than an observation from the second group, and the number of occurrences of the reverse.

$$\hat{\delta} = \frac{\#(x_{i1} > x_{j2}) - \#(x_{i1} < x_{j2})}{n_1 n_2}$$

- where $\# x_{i1} > x_{j2}$ is the number of comparisons between observations in the two groups for which the Group 1 observation is larger than the Group 2 observation.

Cliff's Delta Example

Table 2

Sample of Two Groups of Responses to a Single Item.

Group 1	Group 2
1	1
1	2
2	3
2	4
2	4
3	5
3	
3	
4	
5	

Dominance Matrix

	1	**2**	**3**	**4**	**4**	**5**	*di*
1	0	-1	-1	-1	-1	-1	-0.833
1	0	-1	-1	-1	-1	-1	-0.833
2	1	0	-1	-1	-1	-1	-0.500
2	1	0	-1	-1	-1	-1	-0.500
2	1	0	-1	-1	-1	-1	-0.500
3	1	1	0	-1	-1	-1	-0.167
3	1	1	0	-1	-1	-1	-0.167
3	1	1	0	-1	-1	-1	-0.167
4	1	1	1	0	0	-1	0.333
5	1	1	1	1	1	0	0.833
d.j.	0.8	0.3	-0.3	-0.7	-0.7	-0.9	***-0.250***

Cliff's Delta Values

- It ranges from –1 to +1
- If delta = -1 then all observations in Group 1 are larger than all observations in Group 2
- If delta = 0 then the two distributions are identical
- If delta = 1 then all observations in Group 1 are smaller than all observations in Group 2.

Interpreting Cliff's Delta

- Cohen (1988) presents interpretations of the effect size index *d* in terms of the non-overlap between two normal distributions.
- This provides a direct bridge between *d* and delta

Interpreting Cliff's Delta

- With two normal distributions:
  - *d* effect size of 0.20 corresponds to a delta value of 0.147
  - *d* effect size of 0.50 corresponds to a delta value of 0.33
  - *d* effect size of 0.80 corresponds to a delta of 0.474.

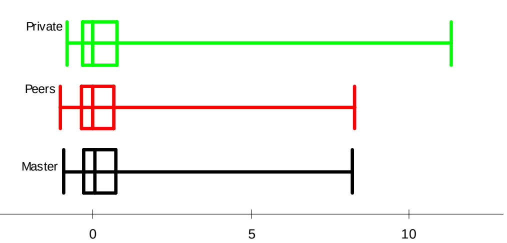
Raw Differences

- Used artificially dichotomized variables as mentioned earlier
- Decided that a cut off value of a 10% difference between two groups on an item was an item that warranted concern for the user
- To calculate a statically significant difference between two proportions one would need to conduct a z test that actually considers standard error as well as the proportion difference

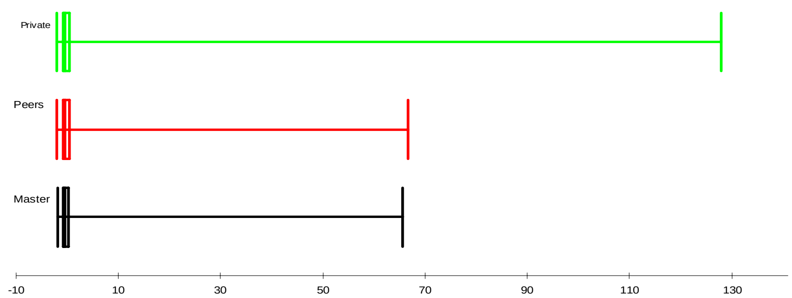
Results of One Evaluation

- To demonstrate these various methods we used data from a small private university's 2005 NSSE report.
- Only the response of first year students were evaluated
- The private institution data was compared to both to the set of its selected peers group to other masters institutions, i.e. its Carnegie Classification group

Skewness



Kurtosis



Evaluation of Normality for Private

Small Private	**Kurtosis**				
Skewness	-.5 < x< .5	-1<x<-.5 or .5 <x<1	-2<x<-1 or 1<x<2	x< -2 or x>2	Grand Total
-.5 < x< .5	7	29	11		47
-1<x<-.5 or .5 <x<1	14	6			20
-2<x<-1 or 1<x<2	2	2	5		9
x< -2 or x>2				9	9
Grand Total	23	37	16	9	85

Evaluation of Normality for Peers

Selected Peers	Kurtosis				
Skewness	-.5 < x< .5	-1<x<-.5 or .5 <x<1	-2<x<-1 or 1<x<2	x< -2 or x>2	Grand Total
-.5 < x< .5	6	37	9		52
-1<x<-.5 or .5 <x<1	11	2	1		14
-2<x<-1 or 1<x<2	3	3	2	3	11
x< -2 or x>2			2	6	8
Grand Total	20	42	14	9	85

Evaluation of Normality for Masters

Masters	Kurtosis				
Skewness	-.5 < x< .5	-1<x<-.5 or .5 <x<1	-2<x<-1 or 1<x<2	x< -2 or x>2	Grand Total
-.5 < x< .5	7	35	10		52
-1<x<-.5 or .5 <x<1	11	4			15
-2<x<-1 or 1<x<2	3	2	3	1	9
x< -2 or x>2				9	9
Grand Total	21	41	13	10	85

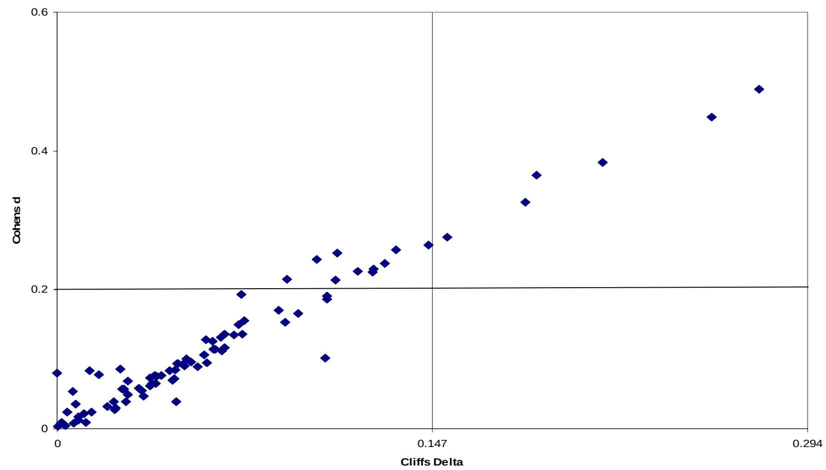
Equal Variances using Levene's Test

- Violation of the assumption of homogeneity of variances
  - 17 comparisons for Private vs. Peers
  - 5 comparisons for Private vs. Master

What is "significant"?

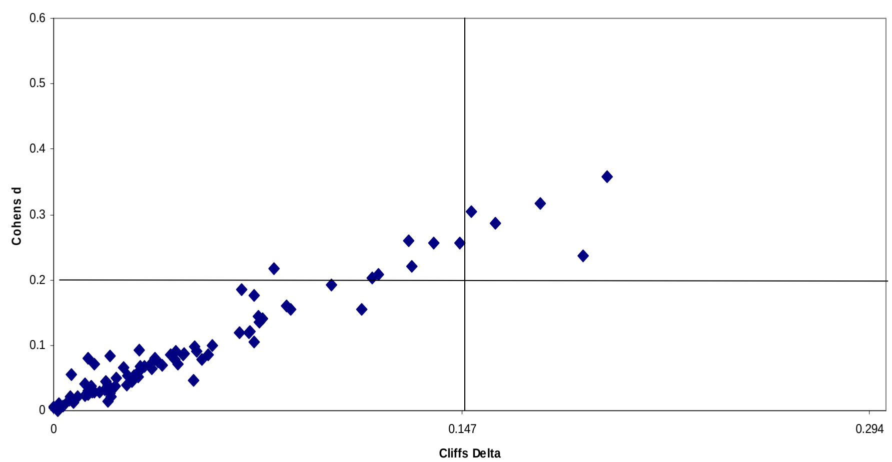
Statistic	**Criteria**
T test	$p \leq .05$
Cohen's d	$d \geq .20$
Chi Square	$p \leq .05$
Odds Ratio	Ratio ≤.50 or Ratio ≥2.0
Raw Difference	Raw difference ≥10 %
Cliff's Delta	Delta > .147

Cliff's Delta vs. Cohen's *d* for Peers



February 2, 2006 Florida Association for Institutional Research, Cocoa Beach, Florida 43

Cliff's Delta vs Cohen's *d* for Masters



February 2, 2006 Florida Association for Institutional Research, Cocoa Beach, Florida 44

Conclusion

Table 10

Contacts

- **Jeanine Romano, University of Tampa**
  jromano@ut.edu
- **Dr. Jeffrey D. Kromrey, University of South Florida**
  kromrey@tempest.coedu.usf.edu
- **Jesse Coraggio, University of South Florida**
  coraggio@coedu.usf.edu
- **Dr. Jeff Skowronek, University of Tampa**
  jskowronek@ut.edu

Institutional Comparisons of NSSE Results: Alternatives to t-tests and Cohen's d for Discrete Ordinal Data

Presented by

Jeanine Romano
University of Tampa

James Coraggio
University of South Florida

Florida Association for Institutional Research, Cocoa Beach, Florida, 2006