



Significance Testing For Two Cluster Samples With Identical Clusters And Different Units

***Pedro J Saavedra, R .Lee Harding,
Ronaldo Iachan***

Comparing independent samples

- This research considers two studies or two samples that generate estimates on the same (or a similar) variable.
- This could arise from the following different scenarios:
 - two populations,
 - two years for the same population,
 - each study asked the questions differently
 - each study used different edits or imputations

The independence assumption

- Comparisons of the two study estimates can be performed from the study reports, if they provide variances, but this assumes that the samples are independent.
- We need more information if the samples are not independent
- Incorrect (conservative) estimates result from an incorrect independence assumption

Purpose of this Research

- Implications of not accounting for the lack of independence between the two samples
- Focus on the case where the two samples use the same PSUs.
- Examine differences between the two estimates—assuming or not assuming independence

Same PSUs, different units

- One situation where two different surveys are not quite independent is when they have the same PSUs, but different units.
- This happens often, for example, when a survey of students selects a sample of schools, and administers half the students one survey and a different survey to the other half.
- It also happens when two surveys sample different kinds of establishments using the same geographical PSUs.
- It also happens when two cycles of the same survey use the same PSUs but different units.

Examples

- Concurrent surveys, piggybacking on each other:
 - Florida Youth Tobacco Survey (FYTS) and Florida Youth Substance Abuse Survey (FYSAS)
 - HUD QC (three assisted housing surveys)
- Same PSUs in different cycles
 - Some of the CDC National HIV surveillance systems
 - FYTS about a decade ago

Main Example

- The FYTS and FYSAS surveys include surveys designed at the State level as well as the county level. Most of the data used in this paper come from the 2009 high school level State Survey.
- Our estimates may differ from those in the published reports because we used different edits.
- On account of this, we will not identify the actual variables used.

FYTS and FYSAS

- The FYTS and FYSAS are classic examples of piggybacking surveys.
- A sample of 80 high schools was selected.
- Classes and students were selected within the high schools and one of three surveys (one not included in this study) was randomly assigned to each selected student.

Weighting

- Each survey was weighted separately.
- Non-response adjustments were conducted at different levels (schools, classes and students within schools).
- Trimming and post-stratification were conducted using grade, race and gender as weighting classes for both purposes.

Variables

- The editing method was different in both surveys, so that respondents with a large number of missing values were considered unit nonrespondents in one and not in the other.
- There were no variables that appeared in both surveys with the same exact wording.
- Two variables were compared, one which yielded large differences and one which did not.
- An additional variable was created through simulation, by adding noise to one of the variables (i.e., changing the answer through a random process).

Variance Estimates

- Either survey can be seen as standing alone.
- In order to obtain variance estimates for either survey, the schools should be seen as primary sampling units.
- Variances can be easily calculated with two methods: a) A jackknife approach that drops one PSU at a time to create replicate weights, or b) A Taylor Series linearization method that uses schools as clusters
- The jackknife can begin with the final weights which are then post-stratified for each replicate. This captures most of the variance due to weighting.

Building Common PSUs into the Variance Estimation Procedure

- The PSUs referred to in the previous slide are common to both surveys.
- Calculation of separate variances for each survey to determine if there are significant differences is likely to overestimate the variance of the difference.
- This phenomenon may not be apparent if one is not aware of the relationship between the surveys.

Building the Design into the Analysis

- There are several ways of doing this.
- Some prefer including the weight variable and the PSUs as covariates, along with the survey.
- The use of PROC SURVEYFREQ and the definition of PSUs as clusters or replicate groups is an alternative.
- Or a jackknife can simple estimate the variance of the difference between proportions.

Comparison through PROC SURVEYFREQ

- Assume you have data properly weighted for the two surveys.
- Let schools be the clusters, and only one stratum.
- Use Taylor Series with no fpc.
- Cross survey indicator with a particular variable that is asked in both surveys.
- The next four slides reflect analyses using the same variable.

Method 1

- Combine the two surveys in one file.
- Use Taylor Series defining the school as a PSU
- The two surveys are included with the weights adding to the post-stratification cells for each survey.
- PROC SURVEYFREQ used.
- Calculate Rao-Scott Chi-Square

Method 2

- The PSUs (schools) of the FYTS are numbered 1 through 80 in order of their county-school ID.
- The schools of the FYSAS are randomly assigned numbers from 1 to 80.
- The two schools assigned the same number are assigned the same cluster.
- Taylor Series are used in PROC SURVEYFREQ.

Method 3

- Each school-survey combination is designated a cluster. Thus a given school forms two clusters – one for the FYTS and one for the FYSAS. This assumes the clusters were sampled independently, and there is no link between a cluster in the FYTS and the same cluster in the FYSAS.
- Taylor Series are then used.

Enter the Jackknife

- Given the three methods described above, a question arises as to whether a jackknife procedure using the three definitions of clusters presented above as replicate groups would yield similar results.
- The replicates are adjusted by both survey and weighting cell.
- The jackknife is easier to adapt to different situations and to conduct simulations with.

Rao-Scott Chi-Square for Six Analyses (Real Data)

Method	PSUs	Rao-Scott Chi-Square	P <
Taylor Series	Matched by School	40.3219	.0001
Taylor Series	Random match	16.4338	.0001
Taylor Series	Independent	16.8318	.0001
Jackknife	Matched by School	41.2228	.0001
Jackknife	Random match	19.9722	.0001
Jackknife	Independent	22.9791	.0001

Adding Noise

- In the previous examples, the estimates examined would have been significantly different regardless of which approach was used.
- We also examined one variable for which differences would not have been significant regardless of method.
- The next table uses a variable where noise was added to a survey variable.

Rao-Scott Chi-Square for Six Analyses (Noise Added)

Method	PSUs	Rao-Scott Chi-Square	P =
Taylor Series	Matched by School	5.2448	.0220
Taylor Series	Random match	2.2990	.1285
Taylor Series	Independent	2.2863	.1305
Jackknife	Matched by School	5.2652	.0244
Jackknife	Random match	2.7964	.0945
Jackknife	Independent	3.0306	.0836

Standard Error of the Difference

- The jackknife can be effectively programmed to estimate the standard error of the difference between two proportions.
- Just to confirm the findings were not due to chance, 1,000 simulated random matches were done.
- Not one yielded a standard error as small as the actual matches.

Suppose each survey were treated separately

Survey	Proportion	Standard Error
Survey 1	0.67482	0.01001
Survey 2	0.60863	0.00927

Standard error of difference assuming independent samples = 0.01365

Standard error of difference with a single jackknife matching schools = 0.01015

Comparing Two Cycles

- 2009 the FYTS was a State survey.
- 2010 it was a county survey.
- In the county survey there were statewide weights.
- In 2010 there were schools that were certainty clusters.
- 62 of the 80 high schools in 2009 were in the 2010 survey.

Creating a Two Year Survey

- Treating the two years as part of the same survey with some PSUs overlapping and some not was beyond the scope of this study.
- The 62 common schools were selected, with a probability equal to the product of the probability in each cycle.
- Weights for each cycle were divided by the probability of selection in the other.
- Weights were then adjusted to add to population totals for the cycle, by grade, race and gender.

Comparison of Estimates

- One variable stood out as having between cycle differences.
- The variable had to do with noticing a particular ant-smoking campaign.
- We focused on that one variable.
- We also added noise to examine the effect of dependence on a marginal difference.

Comparison of the Two Cycles

Table of Matched and Independent Analyses

Variable	Method	Year 09	Year 10	Y09-Y10	Std Error	Lower 95%	Upper 95%
Actual	Matched	0.50321	0.39083	0.11238	0.011868	0.088643	0.13611
Actual	Indep.	0.50321	0.39083	0.11238	0.015122	0.082138	0.14261
Noise	Matched	0.50321	0.47567	0.02754	0.013048	.0014445	0.05362
Noise	Indep.	0.50321	0.47567	0.02754	0.015353	- .003164	0.05824

Conclusions

- Standard errors are consistently higher if one does not match.
- For a marginal result, one might detect significant differences using the combined database and matching PSUs which one might miss using the reported means and standard deviations.

Further Considerations

- This approach is useful when the schools constitute the PSUs in the design.
- If in two surveys a number of schools are selected in both surveys, but others are found in one only of them, the use of alternate approaches may be more effective. The results in such cases were mixed.
- Knowing the relationship between the sample designs of two surveys is essential.

- **Contacts:**

- Pedro.Saavedra@icfi.com
- Richard.L.Harding@icfi.com
- Ronaldo.lachan@icfi.com