Final Review

Final Exam

- Part II (Take Home)

  Saturday April 15, 9.00 a.m. CT

  Submit on Monday April 17, midnight CT

{Text book, Live session notes, Labs, HWs, asynchronous videos (week1-week14)}

- Part I (In-class)

  Wednesday April 19, during the live session

  Live session notes (week1-5, 8-10)

- Office hour : April 13, 10.00p.m.-11.00p.m. CT

Conclusion

Carry out a Chi-square test of the following hypothesis:

H_0: Self-reported bad health is independent of marital status.

H_a: Self-reported bad health is not independent of marital status.

Conclusion

H_0: Self-reported bad health is independent of marital status.

H_a: Self-reported bad health is not independent of marital status.

eg:

- Reject the null
- There is sufficient evidence at 5% significance level (p-value=?) to conclude that self reported bad health is not independent of marital status.

- Be able to define and distinguish sampling error and non-sampling error, and to give 3 examples of the latter.
- Be able to define and distinguish a probability and non-probability sample
- Be able to identify whether or not a specified sampling procedure is a probability sample or not.
- Be able to describe, identify, and illustrate the concept of a sampling distribution. For example, be able to specify the sampling distribution of a defined estimator for a simple sample design from a small population, such as we did for the tax audit population.

- Be able to define the concept of standard error of an estimator, and to explain how it would be determined from the sampling distribution.
- Know how to estimate a mean, total, and proportion from a simple random, stratified, cluster, two-stage, and systematic sample designs, given either raw data or appropriate summary statistics. Know how to compute margin of error for these estimators from a simple random sample.
- Be able to define a finite population correction factor, and know how it affects the standard error of an estimator from a simple random sample.
- Know how to determine the sample size required to achieve a specified margin of error for a simple random sample.

- Be able to state the advantages of a stratified sample design over a simple random sample design.
- Be able to state the advantages of a cluster or two-stage sample design over a simple random sample design.
- Be able to describe and compute sample sizes for both proportional allocation and optimal (Neyman) allocation in stratified sampling. Know when each is most effective.
- Know the advantages and disadvantages of a ratio estimator over stratified sampling where strata are formed by categories of auxiliary variable.

- Be able to describe why it is necessary to use a sampling software such as SAS PROC SURVEYMEANS to properly estimate means and totals rather than software that is not specific to finite population samples, such as SAS PROC MEANS.
- Given a described sample design (even if you don't recognize it as a srs, stratified, cluster, two-stage, or systematic design), know how to compute base weights to achieve an unbiased estimator of mean or total.
- Know how to write proc surveymeans code in order to calculate estimates and their standard errors from any design whose unit selection probabilities, psu's, and strata are known. (This includes all the designs we covered in class.)

- Know how and when to use the ratio statement in proc surveymeans.
- Know how to read a dataset into SAS using both the datalines statement and proc import for an excel file.
- Know how to write proc surveyselect code to select srs and stratified samples.
- Know how to write proc surveyfreq code to produce frequency tables and perform a chi-square test with weighted data.
- Know how to specify all 6 steps in hypothesis testing.