

36-303: Sampling, Surveys & Society
HW05: Hand in on paper Thu Apr 07, 2011

Reminders:

- Please see Lecture 18, Week 09, for an outline of events for the rest of the semester.
- Reading:
 - Hill Handout on Oral Presentations
 - Bem Handout on Writing Papers

Examples of past 303 papers, are on the course website, under the link **Handouts on Writing & Speaking**.

- Exercises #2 and #3 depend on material I will present in lecture Mar 31 and Apr 05.

Exercises to Turn In (there are 3 exercises):

1. **Cluster Sampling (Lohr, 1999, p. 170).** An accounting firm is interested in estimating the error rate in a compliance audit it is conducting. The population consists of 828 claims and the firm audits an SRS of 85 of those claims. In each of the 85 sampled claims, 215 fields are checked for errors. We can think of the claims as psu's (clusters) and the fields as ssu's (units), in a single-stage cluster sample.

Several clusters had the same number of errors, so the results of this single-stage cluster sample can be summarized as follows:

Number of claims (clusters)	Number of errors out of 215 fields	\bar{y}_i = (#errors)/(#fields)
1	4	0.01860
1	3	0.01395
4	2	0.00930
22	1	0.00465
57	0	0.00000

(data courtesy of Fritz Scheuren).

- (a) Use the ideas about cluster sampling discussed in class and in the handout on cluster sampling, to estimate the *proportion* of all $828 \times 215 = 178,020$ fields with errors in them. Give a standard error for your estimate.
 - (b) Use your answer in part (a) to estimate, and give a standard error for, the *total* number of errors in all 178,020 fields.
 - (c) Suppose that, instead of taking a cluster sample, the firm takes an SRS of $(85) \cdot (215) = 18,275$ fields from the 178,020 fields in the population. If the estimated error rate is the same as in part (a), what would the estimated variance be? How does this compare to the estimated variance in part (a)?
2. **Taylor Series Variance Estimates.** At a certain university there are 1998 undergraduates. An SRS without replacement, of size $n = 20$ is taken, to find out what is the average number of hours per week the students study. Some demographic data was collected also. Here is the data from 20 respondents in the survey:

Sex	College	Hrs/Wk	Sex	College	Hrs/Wk
M	Eng	28	F	Eng	36
M	Eng	29	F	Eng	33
M	Eng	23	M	Lib	27
M	Eng	35	M	Lib	28
M	Eng	29	F	Lib	29
M	Eng	30	F	Lib	30
M	Eng	34	F	Lib	28
M	Eng	31	F	Lib	28
F	Eng	30	F	Lib	32
F	Eng	31	F	Lib	30

- (a) Estimate the mean number of hours/wk students study at this university, and give the standard error.
 (b) Administrative records show that the total undergrad enrollment at the university is

Sex	Eng	Lib
M	617	380
F	450	551

Make four post-strata by combining sex and college of the students, compute post-stratum weights to make the sample more representative of the population, and compute the weighted average number of hours/wk students study.

- (c) Use the Taylor Series method discussed in class to compute a standard error for your weighted estimate in part (b). Compare to the standard error you computed in part (a).
3. **Jackknife Variance Estimates.** Use the jackknife method discussed in class to compute a standard error for the weighted estimate in problem 2(b). Compare to the standard errors you calculated in problems 2(a) and 2(c).

Notes on #2 and #3 : The sample size $n = 20$ is small enough that you can do all the computations “by hand” using Minitab, R, Excel, etc. If you know R and you want to compare your answers with the results you get from the R functions I introduced in class, please feel free to do so.