36-303: Sampling, Surveys & Society
HW04: Hand in on paper Thurs Apr 1, 2010

Reminders:

- Please see Lecture 20, Week 10, for an outline of events for the rest of the semester.
- Reading:
  - Hill Handout on Oral Presenations
  - Bem Handout on Writing Papers

  Additional information and examples are on the course website, under the links `giving talks` and `writing`.

Exercises to Turn In (there are 4 exercises):

1. **Robert M. Groves Visit**.

   - Campus-wide lectured Noon-1pm, McConomy. Student reception 1-2pm. *I strongly encourage you to attend.*
   - Visiting our class for Q&A 3-430pm. *You must attend.*

   Attend one or both of the above events, and then answer the following questions:

   (a) Which event will you write about below? (The *noon lecture*, or the *class visit*)

   (b) Write a one-paragraph summary of the talk or visit, in your own words.

   (c) Identify one interesting or noteworthy thing that you learned, that you didn't know before, explain why it was interesting or noteworthy.

   Your description should be clear enough that I can easily remember it from the talk or class visit.

   (d) Identify one thing that you heard, that you didn't really understand, or you think you understood but it doesn't sound right to you. If it's something you didn't understand, list one or more things you think you would need to know to understand it. If it's something that didn't sound right to you, explain why it didn't sound right.

   Your description should be clear enough that I can easily remember it from the talk or class visit.

2. **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	$\overline{y}_i$ = (#errors)/(#fields)
1	4	0.01860
1	3	0.01395
4	2	0.09302
22	1	0.04561
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 828×215=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)?

3. **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
M	Eng	28
M	Eng	29
M	Eng	23
M	Eng	35
M	Eng	29
M	Eng	30
M	Eng	34
M	Eng	31
F	Eng	30
F	Eng	31

Sex	College	Hrs/Wk
F	Eng	36
F	Eng	33
M	Lib	27
M	Lib	28
F	Lib	29
F	Lib	30
F	Lib	28
F	Lib	28
F	Lib	32
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).

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

Notes on #3 and #4: 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.