36-303: Sampling, Surveys & Society
HW02: Due Thursday Feb 11, 2010 in class

Reminders:

- Things to turn in:
  - Due Tonight (Feb 4) by Midnight: Team Assig. I.2 (revising I.1, especially target pop, frame, mode) – email to brian@stat.cmu.edu
  - Due Mon Feb 8: Team Assig. I.3 (choosing which project to do) – email to brian@stat.cmu.edu
    - **Please wait for email feedback from me on I.2 (and instructions on I.3) before starting to work on I.3.**
- Things to read:
  - Done already: Groves, Ch 1, 2, 4 (first part), 5, 11 (first part); and Lohr Appx B
  - For next week (and this hw assignment): Groves Ch 7; after that, parts of 8 and 9
- Clear, careful writing and interpretation of results is an important part of both weekly homeworks and the projects. *I always expect neatly typed or neatly handwritten work.*
- Always be judicious about including computer output and graphs: show enough that we can clearly see what you are doing, but not so much that we will get lost or bored leafing through your work!
- You are operating under the same ethical guidelines as any intellectual worker: *beware of falsification, fabrication and plagiarism.* Citing the source (person, book, etc) of material you use usually protects you against plagiarism claims.

Exercises to Turn In (there are 4 excercises):

1. Let X and Y be discrete random variables with finite sample spaces $\{x_1, \ldots, x_K\}$ and $\{y_1, \ldots, y_K\}$, and let $p_{ij} = P[X = x_i \textit{ and } Y = Y_j]$. Use the definitions of $E[\ ]$, $Var(\ )$, and summation notation ($\sum_{i=1}^{K}$), to show
   (a) $E[aX + bY + c] \; = \; aE[X] + bE[Y] + c$
   (b) $Var(aX + bY + c) = a^2 Var(X) + 2abCov(X, Y) + b^2 Var(Y)$
   (c) If X and Y are independent, then $E[X|Y = y] = E[X]$, for any y.

2. Recall the "Randomized response" model, from lecture, for the question:

> Flip a coin, but dont tell me whether its heads or tails.
>
> - If heads, answer truthfully: have you ever cheated in a CMU class?
> - If tails, answer truthfully: is the last digit of your SSN odd?

Recall from lecture that

$$\pi = \frac{\lambda - (1/2) \cdot (1 - p)}{p}$$

where $p = P[Heads]$; $\pi = P[Cheat]$; and $\lambda = P[Yes]$. Consider a SRS of n students <u>*with replacement*</u>[1]. Let $\hat{\lambda}$ be the fraction of "Yes" answers in the survey, and let $\hat{\pi} = (\hat{\lambda} - \frac{1}{2}(1 - p))/p$.

(a) Show that $E[\hat{\pi}] = \pi$.

(b) Express $Var(\hat{\pi})$ in terms of $Var(\hat{\lambda})$ and show that, as p gets closer and closer to 1, $Var(\hat{\pi})$ gets closer and closer to $Var(\hat{\lambda})$.

(c) Suppose you use a fair coin, so that $p = \frac{1}{2}$, and you think the true rate of cheating on campus is around 0.10. How large a sample would you need, so that a 95% confidence interval for π would be only 0.02 wide?

3. Groves, Ch 2, pp. 64–65. #1.

4. For the following, do not provide responses based on politics, social desirability, etc. Instead, provide analytic responses based on the principles outlined in Groves, Ch 7, and/or in class.

(a) Groves, Ch 7, pp. 238–239, #4.

(b) Groves, Ch 7, p. 239, #5.

(c) Groves, Ch 7, p. 240, #6.

[1] Surveys are seldom conducted this way, but it is easier for the math.