36-303: Sampling, Surveys & Society
HW02: Due Tue Feb 8, 2011 in class

Reminders:

- Things to do:

  Also Due Tues Feb 8: Team Assig. I.3

  - Choose single project to do based on my feedback for I.2
  - Turn in revised versions of A, B, C, D, E, F, for the single project you will do this semester. Details in "project-schedule.pdf" handout (available at the class website, `http://www.stat.cmu.edu/~brian/303`).
  - Start on IRB application (but don't turn it in yet). See irb materials at `http://www.stat.cmu.edu/~brian/303`.

  Get started on... Team Working Agreement, Due Feb 15.

  - Read and do the exercises in the twa.pdf file available at `http://www.stat.cmu.edu/~brian/303`.
  - Turn in a completed team working agreement
    - On paper and signed by everyone on your team. Every team member should also have a signed copy.
    - In email to me (perhaps not signed).

- Things to read:

  - You should be reading Lohr, Appendix B (handout, also available at class website).
  - For next week: Groves, Ch's 7 & 8.

- 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!

- Mostly homeworks will be submitted on paper in class.

Exercises to Turn In (there are 4 excercises):

1. Which of the following modes of data collection is most likely to produce question-order effects:

   - Face-to-face
   - telephone
   - Mail-out/mail-back
   - Web-based survey (like surveymonkey.com)

   Why?

2. 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 $p_{ij} = P[X = x_i \text{ and } Y = Y_j]$. 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) $Var(aX + bY + c) = a^2Var(X) + 2abCov(X, Y) + b^2Var(Y)$

   (c) If X and Y are independent, then $E[X|Y = y] = E[X]$, for any y.

3. 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 *with replacement*[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?

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, p. 255 #1.

   (b) Groves, Ch 7, pp. 255–256, #5.

   (c) Groves, Ch 7, p. 256, #6.

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