36-303: Sampling, Surveys & Society
HW02: Due Tue Feb 1, 2011 in class

Reminders:

- Things to do:

  Also Due Tues Feb 1: Team Assig. I.2: Elaborate your I.1 proposals by adding sections for D, E, F and G on the "project-schedule.pdf" handout (available at the class website, `http://www.stat.cmu.edu/~brian/303`).

- Things to read:
  - Done already: Groves Ch 1, 2, 3
  - This week: Groves, Ch 5, Ch 11 (sections 1–6), Ch 4 (sections 1–3)
  - Begin reading: Lohr, Appendix B (handout).
- 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. Please answer the following questions:
   (a) Groves, Ch 5, p. 179, #1. Please list two specific advantages, and two specific disadvantages, for *each* method discussed.
   (b) Groves, Ch 5, p. 179, #5.
   (c) Suppose you want to do a survey of undergraduates on the Pittsburgh campus of Carnegie Mellon. You are considering three possible modes of data collection:

   Email Email 100 students in C-Book and invite them to take your surveymonkey.com survey.

   Facebook Offer the survey to 100 CMU undergrads on Facebook.

   Fence Stop 100 people who look like undergrads, between Daugherty and the Fence on campus during the noon hour; ask them the questions orally if they agree to be interviewed.

   i. Rank the three modes from greatest to least, in likely coverage error
   ii. Rank the three modes from greatest to least, in likely nonresponse error
   iii. Rank the three modes from greatest to least, in likely measurement error
2. `http://www.cmu.edu/osp/regulatory-compliance/human-subjects.html` lists CMU IRB application forms and various training materials for ethical conduct of human subjects research. Under "Education Requirements" there is a link to "CITI's website", the training website of the Collaborative Institutional Training Initiative (CITI).

(a) Follow the "CITI's website" link.

(b) Select Carnegie Mellon University as your "participating institution" and create an account.

(c) When you have created an account you will be directed to a page titled "Select Curriculum" which displays a list of courses on Human Subjects Research: Biomedical Research Investigators; Social & Behavioral Investigators, and IRB Members. Choose the **Social & Behavioral Investigators** human subjects research module.

(d) The course may take a few hours to complete but can be done over a period of time. When you complete the course, CITI will e-mail your completion record to the CMU IRB.

(e) From the CITI main menu, you can also print a copy of the completion report for this course. *Turn in a printed copy of your completion report with this homework.*

If you already have an active certificate of completion from a course on ethical treatment of human subjects in research, that you took earlier, you can just attach a copy of that older certificate with your hw, without taking the CITI training course.

(You can use the certificate of completion for several years, in any future human subjects research approval process you may be involved in.)

3. Consider a set of n trials involving independent and identically distributed random variables $X_1, X_2, \ldots, X_n$, where the outcome of 1 corresponds to success and the outcome of 0 corresponds to failure, and the probability of success on each trial is p.

   (a) Show that $E(X_i) = p$, and $\text{Var}(X_i) = p(1-p)$ for $i = 1, 2, \ldots, n$.

   (b) Let $Y = X_1 + X_2 + \ldots + X_n$. Show that $E(Y) = np$, and $\text{Var}(Y) = np(1-p)$.

   (c) Let $\hat{p} = Y/n$. Show that $E(\hat{p}) = p$, and $\text{Var}(\hat{p}) = p(1-p)/n$.

4. We say that two discrete random variables are *independent* if $P[X = x, Y = y] = P[X = x] \times P[Y = y]$ for all specific values x and y that X and Y could be.

   (a) The table below shows the distribution for two random variables X abd Y. For example, $P[X = 1, Y = 3] = 3/8$.

x	y	$P[X = x, Y = y]$
1	4	$\frac{1}{8}$
1	3	$\frac{3}{8}$
2	4	$\frac{3}{8}$
2	3	$\frac{1}{8}$

   Use the information in the table to show whether X and Y are independent, or not.

   (b) For any two discrete random variables X and Y (not necessarily the ones in part (a)), show that **If** X and Y are independent, **then** $P[X = x|Y = y] = P[X = x]$, for any specific x and y that X and Y could be.