**Due Monday, February 26**

Complete exercises below. For all questions, give the SAS or R output with estimates.

The shapespop and baseball data are available as a SAS data file (.sas7bdat) and as a comma delimited ASCII file (.csv). These data sets are provided as attachments to the homework.

The SAS program is called Homework\_5\_SAS\_code is also provided as an attachment.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

1. *Objects in bin*. In Homework 3, you selected an SRS of 200 objects drawn from the population in the data file, shapespop. But an SRS does not take advantage of the knowledge that the population contains 15,000 squares and 5,000 circles.
2. Draw a stratified sample of 200 objects form the population with proportional allocation, stratified by *shape*. Answer the questions below for the stratified sample. These questions are like (c) – (d) of Homework 2.
3. Using the sample, estimate the average area for objects in the bin. Give a 95% CI.
   1. How does the 95% CI for this sample compare to the 95% CI from your SRS sample results in Homework 3?
4. Using the sample, estimate the total number of gray objects in the population, along with the 95% CI.
5. How does the 95% CI for this sample compare to the 95% CI from your SRS sample results in Homework 3?
6. *Baseball data*. The data file baseball contains statistics on 797 baseball players from the roster of all major league teams in November 2004. In this exercise, you will treat the file baseball as a population and draw samples from it using different sampling designs.
7. Take an SRS of 150 players from the file.
8. Take a stratified random sample of 150 players from the file, using proportional allocation with the different teams as strata.
9. For each sample, estimate the proportion of players who are pitchers and give a 95% CI.
   1. How do the estimates compare across the two samples? How do the 95% CI’s compare?
10. For each sample, estimate the mean of the *logsal* variable and give a 95% CI.
    1. How do the estimates compare across the two samples? How do the 95% CI’s compare?
11. Examine the variances of *logsal* in each stratum using the population data. Do you think optimal allocation would be worthwhile for this problem? Why or why not?