A. Using Ratio Estimators for an SRS from Stephens County (Use the data, **Stephens.dat** provided)

1. Inference on the population ratio

   (a) Estimate the ratio of hours spent watching news over hours spent watching sports during the week prior to the study.

   (b) Estimate the standard error of the estimator of the ratio in (1a).

   (c) Calculate a 95% confidence interval for the ratio we are interested in.

   (d) Interpret this confidence interval in the context of the problem.

2. Let us return to the topic of willingness to pay to see if we can use known information to increase the precision of our estimate for the average willingness to pay.

   (a) Plot willingness to pay versus assessed house value.

   (b) Obtain the sample correlation between willingness to pay and assessed house value.

   (c) Given (2a) and (2b), does using a ratio estimator for the mean willingness to pay seem sensible? Why or why not?

   (d) Regardless of your answer to (2c), use a ratio estimator to estimate the mean willingness to pay for cable television in Stephens County. (Note that the population mean assessed house value is 68,045.)

   (e) Estimate the standard error of the ratio estimator in (2d). (Note that the total number of households in Stephens county is 31,989.)

   (f) Obtain the unbiased estimate for the mean willingness to pay and the standard error of the mean.

   (g) Create a table that uses sample quantities from these data to compare the properties of the ratio-adjusted estimate for the mean willingness to pay with that of the sample mean estimate. Using what you have learned in class and from the text, provide an interpretation of the results displayed this table.

B. Estimation for an SRS from Golf Courses

3. The data used in this problem are posted as **golfsrs.dat**. The documentation(**golfsrs.pdf**) for the data-set used in this problem is also given. The course name variable was omitted from the documentation, so we have provided the correct input statement:

```
INPUT rn state $ course $ holes type $ yearblt wkday18 wkday9
      wkend18 wkend9 backtee rating par cart18 cart9 caddy $ pro $ ;
```

   The data set contains data on a SRS of 120 golf courses. The sample was selected from the list of 16,883 US golf courses, obtained from the website www.golfcourse.com.

   (a) Estimate the average greens fee to play 9 holes on a weekend.

(b) Consider the relationship between the greens fee to play 9 holes on a weekend and the back-tee yardage.

i. Create a plot for these two variables.

ii. Estimate the correlation coefficient for these two variables

iii. Estimate a regression model for predicting weekend greens fees for 9 holes on a weekend from the back-tee yardage.

(c) Suppose you were asked to obtain a regression estimate for the mean greens fee to play 9 holes on a weekend for the all golf courses listed on the website, along with a 90% confidence interval. Do you have all of the information you need? If not, what information is missing?

(d) Based on what you know about regression estimation, information from (3b), and the description of this dataset, would you expect the regression estimator to provide a etter?estimate of the mean greens fee for 9 holes on a weekend than the sample mean would? Justify your answer.

C. Cherry Problem: Use the attached data (cherry.csv) and SAS program below

4. The data set *cherry.dat* from Hand et al.(1994), contains measurements of diameter(inches), height(feet), and timber volume(cubic feet) for a sample of 31 black cherry trees. Diameter and height of trees are easily measured but volume is more difficult to measure.

(a) Suppose that these trees are an SRS from a forest of $N = 2,967$ trees and that the sum of diameters for all trees in the forest is $t_x = 41,385$ inches. Use ratio estimation to estimate the total volume for all trees in the forest. Give a 95% CI.

(b) Use regression estimation to estimate the total volume for all trees in the forest. Give a 95% CI.

```
filename cherry 'C:\Stat311\cherry.csv';

data cherry;
infile cherries delimiter=',' firstobs=2;
input diam height vol;
sampwt = 2967/31;
obsnum = _n_;
label diam      = 'diam (in) at 4.5 feet'
      height    = 'height of tree (feet)'
      vol       = 'volume of tree (cubic feet)'
      sampwt    = 'sampling weight'
;

RUN;
```