1. A 1994 survey of 250 local snowmobilers in St. Louis County Minnesota asked respondents to estimate per person per day expenditures ($) on snowmobile trips. Relevant data are in Assign5_1.csv. Licensing and ancillary records suggest there were 10,000 local snowmobilers in St. Louis County Minnesota at the time. Assume simple random sampling was used for the survey.

   a. Estimate the mean per person per day expenditure and provide a standard error estimate (use the fpc).
   b. If we assume it is known that the mean trip length is 1.7 days for these snowmobilers and that they take on average 4.3 trips per season, estimate total expenditures for the 10,000 local snowmobilers in St. Louis County Minnesota for the 1994 season. Provide a standard error for the estimate of total. Compute a 90% confidence interval for the true total expenditure.
   c. Using this sample to estimate variability, how many snowmobilers need to be surveyed if it is desired to obtain a standard error of the estimate of the mean of $1.50.
   d. Can you identify any means of improving this sampling effort through use of stratified random sampling?

2. This question uses the data on wood testing from exercise 2.100 on page 182 of your textbook. Read the text of the exercise for background. The data are in Assign5_2.csv.

   a. Fit a simple linear regression with MOR as the y-variable and MOE as the x-variable. Identify the intercept and slope estimates, the standard error of the regression and \( R^2 \). Examine the assumptions of linear regression with a standardized residual plot. Is the assumption required to conduct formal inference seemingly met?
   b. What is the P-value for the test of the null hypothesis intercept=0? Is there especially strong evidence against the null hypothesis? Refit the regression without an intercept. Compute an “\( R^2 \) comparable to the one in a. Based on fit statistics (s and \( R^2 \)), is this model practically inferior to the one in a.? Examine the standardized residuals for the intercept-less regression. Is there a pattern of concern in the residuals?
   c. There would appear to be one data point that does not conform to the overall pattern of the bulk of the MOR/MOE data. Remove the point and repeat the analysis of a. Comment on the impact of the point.
   d. Find a 90% confidence interval for the mean MOR at MOE=1,800,000. Find a 90% prediction interval for an individual MOR at MOE=1,800,000. Repeat these two steps for MOE=2,400,000. Why are these latter intervals (where MOE=2,400,000) wider than those where MOE=1,800,000?

3. This question uses the perch size data from exercise 10.38 on page 678 of your textbook. We will only use the Weight (variable of interest) and Length data. The data on
the smallest fish are potentially problematic but are legitimate and can not be deleted. The data are in Assign5_3.csv.

a. Fit a simple linear regression with Weight as the y-variable and Length as the x-variable. Identify the intercept and slope estimates, the standard error of the regression and $R^2$. Examine the assumptions of linear regression with a standardized residual plot. Is the assumption required to conduct formal inference seemingly met?

b. Fit a simple linear regression with log(Weight) as the y-variable and log(Length) as the x-variable. Identify the intercept and slope estimates, the standard error of the regression and $R^2$. How does this model compare to that in a.? Be careful what/how you interpret.

c. Use nonlinear least squares (function nls in R) to fit the allometric model:
   \[ \text{Weight} = b_1 \times \text{Length}^{b_2} + \varepsilon \]
to the data. Identify the parameter estimates and the standard error of the regression. Compute an equivalent of $R^2$. What does the P-value on $b_1$ imply (think carefully)? Does a residual plot indicate any problems? How’s our “smallest fish” looking? How does this model compare to that in a.?

d. Compare the predictions of Weight for the smallest and largest fishes using the three models (a., b., c.).

4. This is an extension of question 1. on assignment 3. In addition to phosphorus, nitrogen concentration was also observed in the 25 lakes. Interest lies in the relationship between chlorophyll-a (CH) and phosphorus (P) and nitrogen (N) simultaneously. The data are in Assign5_4.csv.

a. Plot CH against P and N separately. Which indicates the strongest relationship? Compute correlations to back this up.

b. One possible equation for this relationship is:
   \[ \text{CH} = \beta_0 + \beta_1 \text{P} + \beta_2 \text{N} + \varepsilon \]
Estimate the parameters of this equation. Comment on the tests on the parameters. Complete a residual analysis.

c. Another possible equation for this relationship is:
   \[ \log(\text{CH}) = \beta_0 + \beta_1 \log(\text{P}) + \beta_2 \log(\text{N}) + \varepsilon \]
Estimate the parameters of this equation. Comment on the tests on the parameters. Complete a residual analysis.

d. Compare the results of b. and c. in terms of which provides the best description of the relationship between chlorophyll-a and phosphorus and nitrogen simultaneously.

e. Use both equations b. and c. to find a 90% for true mean chlorophyll-a when phosphorus=150 and nitrogen=12. This will require some extra work when using equation c.