Homework #8

Answer the following questions in a single R script called `hw08.R`. Answers must be given by R commands. You cannot simply look at the data set and answer the question via direct inspection. Use comments (#) to indicate which portion of your code answers which question. Be sure that you obtain the correct solutions to each question when you execute your script one line at a time from top to bottom.

Each question will be graded according to the following criteria:

- 0%: No attempts is made to answer the question.
- 25%: An attempt is made that, although unsuccessful, revealed some understanding of what the question was asking.
- 50%: Solution is incorrect, but with some modifications, could be corrected.
- 75%: Solution is incorrect, but easily resolved with minor modifications **OR** solution is correct, but obtained via convoluted reasoning or by avoiding standard approaches.
- 100%: Solution is correct and uses standard approaches.

#1) A random sample of 130 healthy humans is taken and their body temperatures is recorded. Suppose the average temperature of the sample is $\overline{x} = 98.25$ with sample standard deviation of $s = 0.73$. Find a 99% confidence interval for the population mean μ of healthy human body temperature.

#2) Nutritional information provided by Kentucky Fried Chicken (KFC) claims that each small bag of potato wedges contains 4.8 ounces of food and 280 calories. A sample of ten orders from KFC restaurants in New York and New Jersey averaged 358 calories. If the sample standard deviation was $s = 54$, is there sufficient evidence to indicate that the average number of calories in small bags of KFC potato wedges is greater than advertised? Test at the 1% level of significance.

#3) The EPA has set a maximum noise level for heavy trucks at 83 decibels (dB). The manner in which this limit is applied will greatly affect the trucking industry and the public. One way to apply the limit is to require all trucks to conform to the noise limit. A second less satisfactory method is to require the truck fleet's mean noise level to be less than the limit. If the latter rule is adopted, variation in the noise level from truck to truck becomes important because a large value for σ^2 would imply that many trucks exceed the limit, even if the mean fleet level were 83 dB. A random sample of six heavy trucks produced the following noise levels (in decibels):

85.4, 86.8, 86.1, 85.3, 84.8, 86.0.

Use these data to construct a 90% confidence interval for σ^2, the variance of the truck noise-emission readings.

#4) Scholastic Assessment Test (SAT) scores, which have fallen slowly since the inception of the test, have begun to rise. Originally, a score of 500 was intended to be average. The mean scores for 2005 were approximately 508 for the verbal test and 520 for the mathematics test. A random sample of the test scores of 20 seniors from a large urban high school produced means and standard deviations listed in the accompanying table:

	Verbal	Mathematics
Sample means	505	495
Sample standard deviation	57	69

(a) Find a 90% confidence interval for the mean verbal SAT score for high school seniors from the urban high school.

(b) Find a 90% confidence interval for the mean math SAT score for high school seniors from the urban high school.

#5) Aptitude tests should produce scores with a large amount of variation so that an administrator can distinguish between persons with low aptitude and persons with high aptitude. The standard test used by a certain industry has been producing scores with a standard deviation of 10 points. A new test is given to 20 prospective employees and produces a sample standard deviation of 12 points. Are scores from the new test significantly more variable than scores from the standard. Use $\alpha = .01$.

#6) A study was conducted by the Florida Game and Fish Commission to assess the amounts of chemical residues found in the brain tissue of brown pelicans. In a test for DDT, random samples of $n_1 = 10$ juveniles and $n_2 = 13$ nestlings produced the results shown in the accompanying table (measurements in parts per million, ppm).

Juveniles	Nestlings
$n_1 = 10$	$n_2 = 13$
$\overline{y}_1 = .041$	$\overline{y}_2 = .026$
$s_1 = .017$	$s_2 = .006$

(a) Test the hypothesis that mean amounts of DDT found in juveniles and nestlings do not differ against the alternative that juveniles have a larger mean. Use $\alpha = .05$.

(b) Compute the p-value for the hypothesis test you performed in part (a).

#7) Let's revisit the dataset in `ws02_weather.csv`.

(a) Give a scatter plot for the `humidity` (horizontal axis) against the `cloudcover` (vertical axis).

(b) Using a simple linear model, compute estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ from the points in part (a). Then plot the resulting regression line together with the scatter plot.

(c) Looking at the plot in (b), do you believe that cloud cover is reasonably modeled using a simple linear model on the humidity? Be sure to justify your answer.

(d) Let's think of X as the random variables that measures the humidity on a given day. Let Y be the random variable that measures the cloud cover on a given day. Using our sample data, test the null hypothesis $H_0 : \rho = 0$ against the alternative hypothesis $H_a : \rho > 0$.