SDS 439 - Homework 02

Due Feb 10, 1:00 pm

Minnesota Water Usage

This homework concerns analysis of the minnesota water dataset, which is included in the R package for the textbook "Applied Linear Regression" by Sanford Weisberg. The dataset has municipal and agricultural water usage data for 24 years, along with agricultural and municipal precipitation data for the same time period. We are interested in studying how water usage is changing over time.

The dataset is located in the course repository in the file `datasets/mn_water.csv`.

A thorough description of the data can be found on page 32 of the `alr4` documentation: https://cran.r-project.org/web/packages/alr4/alr4.pdf

1. You should try to start your analyses by looking at the data, so you have an idea of what to expect from your analysis. Make 4 informative plots of the data, to familiarize yourself with the major features of the data. Use good plotting practice, labeling your axes, using legends when appropriate, etc. A good plot should show several aspects of the data, without overwhelming you.
2. Use `lm` to regress the total water usage on the municipal usage and the irrigation usage, and print out the summary table. What do the results tell you about the relationship between these 3 variables?
3. Let's first study the municipal water usage to understand how it is changing over time. Make a plot of municipal water usage over time.
4. Regress municipal water usage on year. Print the results and answer the following questions: How do you interpret the slope? How do you interpret the intercept?
5. To facilitate interpretation of the intercept, create a new variable counting the years since 1988. Then regress municipal use on your new variable and print the results. Except for the intercept, does anything else change in the model? What is the interpretation of the intercept in this model? Verify that the intercept for this model is equal to the fitted value for 1988 from the previous model.
6. Perhaps the changes in municipal water usage can be explained by changes in municipal population. Regress municipal water use *per person* on years since 1988. BUT make sure to use an appropriate unit for the response (billions of gallons per person is hard to interpret). Make a plot with the data and the regression line, and summarize your results in words.
7. Let's pivot to the irrigation usage. Make two plots of the irrigation usage, one against year, and one against agricultural precipitation
8. Perform a multiple regression of irrigation usage on years since 1988 and agricultural precipitation. Perform the following steps: (1) Write out the model in mathematical notation (like we do in class), (2) Print the summary table, (3) provide an interpretation of each of the 3 regression coefficients.
9. What do the regression results tell us about how irrigation usage is changing over time?
10. Provide an estimate of how much water would have been used in 2012 if the precipitation were 12 inches. Write down the estimate in terms of your model parameter estimates and provide the numerical value. Can you provide a standard error for your estimate?

11. Provide an estimate of how much less water would be used for irrigation in a rainy year (say 16 inches) versus in a dry year (say 8 inches). Write down the estimate in terms of your model parameters and provide the numerical value. Can you provide a standard error for your estimate?