Lab 12a for Math 17: Regression Investigations

1 Tasks

We have multiple regression examples to consider (class examples determined last week). For each example, a series of questions is provided to guide your analysis. You may want to work with a partner so you can compare results. For tests and CIs, the other assumptions will be checked in the last step, so do not worry about checking them before you run the regression.

2 Regression Commands Review

There are no new commands. Here are the regression commands you have seen.
Obtaining a scatterplot: Graphs>Scatterplot. Remember to turn off the smooth line option.
Obtaining a correlation: Statistics>Summaries>Correlation matrix
Fitting the regression model (i.e. getting the basic output): Statistics>Fit models>Linear regression
Obtaining residuals: Models>Add observation statistics to data. Then select residuals or studentized residuals (you can turn the rest off).
Making residual plots: Make a scatterplot using the residuals you obtained versus the predictor variable.
Shortcuts: Rather than saving the residuals and then making the scatterplot, click Models>Graphs>Basic Diagnostic Plots. You will get a set of 4 plots. The first 2 are the residual plot (residual vs. fitted - this is the SAME as a residual plot for simple linear regression) and qq plot of the residuals.

3 Class Data

The multiple pairs of variables were collected by the class for use in regression. Note that this is a convenience sample. The data is online as classregdataS11.txt. Everyone will be choosing a pair of variables to investigate after some preliminary analysis as a class. Note: This is going to be tricky with our class data. There are a few large correlations, and we still need to check for linear relationships in those cases. You should choose a pair that is reasonable for the analysis (i.e. do NOT pick a case where the linear model fails to be appropriate at first glance).

4 Heavy Metals

A second example on heavy metals has its own set of guided questions.

5 Class Example

Choose one of the pairs of variables collected by the class. Which variables did you choose to investigate?

1. What does a basic data analysis reveal about the variables? Are there interesting features to the data?

2. Identify the response and explanatory variables.

3. Estimate the correlation and find the exact value using Rcmdr. Does a linear regression appear appropriate? It may not appear appropriate for all examples.

4. Fit a regression model if appropriate. Report the fitted regression line. Compute *your* residual.

5. Does the model look like it fits well? Explain.

6. Test to see whether or not the predictor is a significant predictor of the response.

7. Make a confidence interval for the slope.

8. Do the regression assumptions appear valid? Explain.

6 Heavy Metals

Metals.txt online under labs contains information on 17 core samples from Louisiana. Each sample's depth was recorded in meters, and then each core was tested for zinc (ppm) and iron concentration (percentage). You will be exploring the data to determine if depth can be used to predict either zinc or iron concentration.

1. What does a basic data analysis reveal about the variables? Are there interesting features to the data?

2. Identify the response and explanatory variables. Note there are 2 different regressions to consider here.

3. Estimate the correlations and find exact values using Rcmdr. Does a linear regression appear appropriate for predicting either zinc or iron? It may not appear appropriate for all examples.

4. Fit a regression model if appropriate. Report the fitted regression line. If someone wanted predictions for a depth of 50 m, what would you tell them?

5. Does the model look like it fits well? Explain.

6. Test to see whether or not the predictor is a significant predictor of the response.

7. Make a confidence interval for the slope.

8. What are the four regression assumptions? Do the regression assumptions appear valid? Explain.