Lab 12b for Math 17: Regression Investigations

Only one example and then turn-in portion today, the rest of the time you may devote to homework or projects, etc.

1 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.

2 Forest Fires

A data set on forest fires lists number of fires in thousands along with number of acres burned in millions. Does the number of fires appear to be a significant predictor for number of acres burned? (Data from Triola). Online as FFx2.txt.

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. What is your best estimate of the number of acres burned if there were 50,000 fires?

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.

3 Turn-In

Determine if each statement is true or false for 1-8. The rest are short answer. You may tear off this page to turn in. Be sure your name is on it!

1. Most of the assumptions for simple linear regression are about the error terms, but are checked using the residuals.

2. A residual plot can only indicate problems with nonconstant variance.

3. Standard error of the residuals is roughly the average size of a residual.

4. The distribution of the test statistic when testing for a significant predictor in OLS under the null hypothesis is a t with n-1 degrees of freedom.

5. A fitted linear regression model results in an R-squared value of .869. This means the linear regression was appropriate and yields very good predictions.

6. A 95 percent prediction interval for an x-value close to the mean will be wider than a 95 percent prediction interval for an x-value further from the mean.

7. A 90 percent confidence interval for the mean response will be narrower than a 90 percent prediction interval for an individual response assuming both are for the same x-value.

8. There are no tests or CIs in regression associated with the intercept term.

9. Suppose you know a linear regression is appropriate and that all assumptions check out. Based on the sample R output below and assuming you want to test for a significant positive linear relationship between x and y, the p-value for this test would be what?

	Estimate	SE	t value	p-value
(Intercept)	-.63652	1.49784	-.425	.673
x	-.07329	.05887	-1.245	.219

10. Again suppose you know a linear regression is appropriate and all assumptions check out. Using the sample R output below, find the test statistic and p-value for testing to see if the slope is equal to 2 or is greater than 2. You also know the degrees of freedom is 48.

	Estimate	SE	t value	p-value
(Intercept)	-.4.32742	1.97891	-2.187	.0337
x	3.13653	.07778	40.328	<2e-16