

STAT6038 week 6 lecture 18

Rui Qiu

2017-03-31

Summary So far, what have we learned? The general procedure for examining the relationship between two variables proceeds as follows:

1. **Start with a little exploratory data analysis.** Examine a plot of the data to see if a linear association seems a plausible explanation of the scatterplot.
2. Using scientific or other background information, transform the data appropriately.
3. Fit a linear regression model to the two variables using least-square estimates.
4. Use residual plots and normal q-q plots to examine the plausibility of the basic assumptions of the model.
5. If necessary (based on the plots of previous step) transform the data again (or move to a more complicated model structure).
Note: We cycle through step 3,4,5 until we are happy with the residual plots.
6. Re-fit the regression and again examine the residuals thoroughly.
Note: After 5, 6 we get an “appropriate” model, do step 7 and look at the first ANOVA table, if F test okay then it is an “adequate” model, then we look at the summary output.
7. Test the significance of the regression and make required predictions.
8. When satisfactory residual analyses have been reached, re-test and re-predict as required (remembering to transform back to appropriate scales if necessary).

We will now take up the point mentioned parenthetically in step 5, and examine more complicated, though still linear models, where we will allow for the possibility of more than one predictor variable.