

PSTAT 126 - Homework 3

Due: 11:55 p.m. Monday, August 26

Do the following problems. Submit the homework on Gauchospace before Tuesday.

Problem Set:

1. This problem uses the data set **prostate** from the **faraway** package (see problem 2 from HW 2).
 - a) Using the variable **lpsa** as the response and **lcavol** as the predictor, use R to produce an ANOVA table for this regression fit.
 - b) In the ANOVA table from part a), which quantity represents the variability in **lpsa** which is left unexplained by the regression?
2. This problem uses the data set **baeskel** from the **alr4** package.
 - a) Fit the regression model with **Tension** as response and **Sulfur** predictor, and produce three diagnostic plots: Residuals vs. Fitted, Scale-Location and a QQ-plot. Comment on any violation of the standard linear model assumptions seen in these plots.
 - b) Consider two alternative models given by the predictor transformations $1/\text{Sulfur}$ and $\log(\text{Sulfur})$: With **Sulfur** on the horizontal axis and **Tension** on the vertical axis, fit these two alternatives and plot the regression fits along with the fit from part a). Note that the two fits from this part will not be linear, since the predictor was transformed.
Hint: The R function `invTranPlot` is useful here - see Section 8.1 of the R primer.
 - c) Replace **Sulfur** by its logarithm, and consider transforming the response **Tension**. To do this, find and report the optimal power transformation, $\hat{\lambda}_{ML}$ using the Box-Cox procedure discussed in class. Should you transform the variable? Explain.
3. This problem uses the data set **UN11** from the **alr4** package.
 - a) Examine the scatterplot matrix for (**fertility**, $\log(\text{ppgdp})$, **pctUrban**), and comment on the marginal relationships.
 - b) Fit the two simple regressions $\text{fertility} \sim \log(\text{ppgdp})$ and $\text{fertility} \sim \text{pctUrban}$, and verify that the slopes are significantly different from zero at any conventional level of significance.
 - c) Obtain the added-variable plots for both predictors. Based on the added-variable plots, does $\log(\text{ppgdp})$ seem to be useful after adjusting for **pcturban**, and similarly, does **pcturban** seem to be useful after adjusting for $\log(\text{ppgdp})$?