Fall 2015 Test

You may use SAS help at `sas.support.com` as well as **R** online help; either package is appropriate for any question. You may also review your PROC IMPORT statements for tab-delimited text files with headers. Both data sets used in this exam (RichlandAADT.txt and MTQ2.txt) are available in Blackboard and on the course website as tab-delimited text files.

1. Average daily traffic intensity (`AADT`) in 2012 for pre-selected road reaches in Richland County has been saved in file `RichlandAADT.txt`. The file includes a second variable, `Type`, that classifies the road reach. Types include 1=Interstate, 2=Expressway, 4=Major arterial road, 7=Minor arterial road, 9=Local road.

   (a) Create side-by-side boxplots of AADT by Type. Comment on any patterns you see there.

   (b) We will focus on a two-sample comparison of AADT for Type 4 (Major arterial road) and Type 7 (Minor arterial road). Create a data set with the AADT for only these two road types. Look at normal quantile plots to see if the data appears normal for both levels of road type.

   (c) Before proceeding with testing or estimation, experiment with a transformation of AADT and inspect normal quantile plots to see if the transformed data appears normal. What do you observe? Why might a transformation be unnecessary for this data?

   (d) On the original data scale, create and interpret a 90 % nonparametric confidence interval for $\mu_{Major} - \mu_{Minor}$. A test seems unnecessary here, but if you were to carry one out, which alternative would you use if your null is $\mathrm{H}_0 : \mu_{Major} = \mu_{Minor}$?

   (e) Now construct a pooled t-test on the scale you chose (transformed or untransformed) in 1(b) and 1(c). Justify your choice of a standard error for $\bar{x}_{Major} - \bar{x}_{Minor}$. Compare results from 1(d) and 1(e), using a back-transformation for 1(e) if needed.

2. Suppose we choose to regress a response Y on a pre-test predictor variable X_1 and a second predictor variable X_2 in the model $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \epsilon_i$.

   (a) Fix $\beta_1 = 1$ and derive the least-squares estimator for β_2, then show that it is the difference between the least squares slope estimators for the regression of Y on X_2 and the regression of X_1 on X_2.

   (b) See the attached data set MTQ2.txt. Carry out all three regressions mentioned in 2(a) and confirm the relationship between the slope estimators.

   (c) Now regress Y on X_1 and X_2 for the unconstrained model. Test the hypothesis $\mathrm{H}_0 : \beta_1 = 1$ against a two-sided alternative.

   (d) Create a scatterplot matrix and other appropriate scatterplots of Y, X_1 and X_2. Do any of the variables need to be transformed? If so, which variable(s) and what transformation(s) would you choose?

   (e) Make any appropriate transformations and create a new regression model. What is the coefficient of multiple determination? Are both X_1 and X_2 (or their transformations) needed in the model? Comment on plots of residuals against $\hat{Y}$ and each of the predictor variables.