MMA 860 – Assessing & Testing Models Exercises

Please complete the following list of exercises. All data is stored in the Assessing & Testing models data file.

1. **Basic Regression**
    a. Import the data from tab 'Sales Data'.
    b. Run a linear regression model to explain order size in terms of Ad_Budget and Distance.
    c. Assess the results. Are there problems?
    d. Determine if Distance belongs in the model.
    e. Rerun the model without distance.
2. **Bimodal Error 1**
    a. Import the data from tab 'Bimodal Error 1'.
    b. Run the linear regression model y = b0 + b1 x1 + b2 x2 with the default settings.
    c. Examine the regression output table. Does it look OK?
    d. Now look at the distribution of the error term. What problems do you see here?
3. **Bimodal Error 2**
    a. Import the data from tab 'Bimodal Error 2'. This data is the same as the Bimodal Error 1 data, however we have added a 'dummy' variable (more on this next class).
    b. Run the linear regression model y = b0 + b1 x1 + b2 x2 + b3 US.
    c. Examine the regression output table and error distribution. Suppose this reflected store sales data for stores in Canada and the US where x1 and x2 were measures of local customer base and economic activity. Explain to your manager why the results were different.
4. **Non-Normal Errors**
    a. Import the data from the tab 'Nonlinear'.
    b. Run the linear regression y = b0 + b1x1.
    c. Examine the residual plot. Notice the pattern.
    d. Re-run the regression including the X1Squared variable. Note the difference in the residual plot.
5. **Outliers**
    a. Import the data from the tab 'Outliers'.
    b. Run the linear regression y = b0 + b1x1 + Outlier.
    c. Examine the Cook's D plot and observe the single outlier.
6. **Heteroskedasticity**
    a. Import the data from tab 'Heteroskedasticity'.
    b. Run a linear regression model on the data using y = b0 + b1 x1 + b2 x2. Assess for heteroskedasticity.
7. **Collinearity**
    a. Import the data from the tab 'Collinear'.
    b. Run a linear regression to explain y in terms of experience and height. Does height appear to explain y?
    c. Run a linear regression to explain y in terms of experience and weight. Does weight appear explain y?

d. Run a linear regression to explain y in terms of experience and height and weight. Explain the results in language a manager would understand.

8. **Diminishing Returns**
    a. Import the data from the tab 'Diminishing Returns'
    b. Run a linear regression to explain sales in terms of price and ad_budget. Observe the plots – does anything look strange?
    c. Create a new variable called ln_ad_budget which is the natural log of ad_budget (in Python, the command is 'log' in the math package)
    d. Run a linear regression to explain sales in terms of price and ln_ad_budget. Observe the plots – what has changed?