BZAN 535 **Homework 7** **Due: October 26, 2016**

The journey_transaction_data table contains a column for retail discount. If someone buys a quantity of 3 items for a total sales_value of $3.75 but this would regularly cost $6, the retail discount would be -$2.25. In that case, the full retail price per item is $2. Since this information is provided for every transaction, you can calculate the % of items that are sold at a sale price each week. To answer this question, you will need to select all transactions for 1053690 and, in a separate file, acquire all the transactions for 844165.

1. On the midterm Exam, you explored sales of 2-liter soft drinks for two national brand products – 1053690 and 844165 (Presumably one is Coke and the other is Pepsi). For 1053690, discuss the pricing policy. The full retail price is ostensibly $1.59. Report the frequency and duration of major unit price discounts, and the amount of those discounts. How often has the product been at the full price of $1.59? Provide a graphic to aid in your answer. Highlight the few weeks that had both retail discount for some sales and no discount for other sales of 1053690.

2. For product 1053690, fit both a spline (adjusting the flexibility so that it bends but is a decreasing function) and a straight-line regression model for Weekly quantity sold vs. Mean Unit Sales_Value.

    (a) Present a single plot with both the spline model and the linear regression model for product 1053690. Report and interpret R^2 for each of the two models, showing how the sum of squares error figures in the calculation of R^2.
    (b) Display residual plots (residual vs. X and residual vs. time order) and comment about any problems or inadequacies evident in the residuals. Conduct a Durbin Watson test and report the conclusion from this test as well.
    (c) Report the F for lack of fit and the conclusion from this test. Discuss how residuals at X=$1.50 and $1.59 contribute to this lack of fit.

3. Exclude weeks 1-21 and 82-102, to focus on the year when $1.34 is the typical price.
    (a) Fit a straight-line regression model for Weekly quantity sold vs. Mean Unit Sales_Value
    (b) Report and interpret R^2, the RMSE, and the slope coefficient.
    (c) What is the predicted amount of weekly unit sales at $1.59? Show how you can use the model equation to obtain this value.
    (d) Show a plot of residuals vs. predicted. What does this reveal? Be sure to comment on the residuals at $1.59.

4. Finally, excluding weeks 1-21 and 82-102, fit a regression model for log(Weekly quantity) vs. log(Mean Unit Sales_Value).
    (a) Show a scatterplot that includes this fitted model. Write out the equation of the model.

(b) Interpret the slope and provide a 95% confidence interval for the true price elasticity. (Note: For a log quantity vs. log price model such as this, the slope coefficient represents the % change in units sold for each 1% increase in price. This is the price elasticity and a linear relationship between log quantity and log price implies that the price elasticity is constant at all prices.)
(c) Provide a 95% confidence interval for the mean weekly sales at price of $1
(d) Provide a 95% prediction interval for the weekly sales for a week with price = $1.
(e) Explain the difference of interpretation for the intervals in 4c and 4d.

Extra Credit: Show a plot of residuals (from the model in #3 or #4) vs. the average price of 844165. Does the sales of 1053690 seem to be impacted by the price of 844165?