BZAN 535 - Homework 7

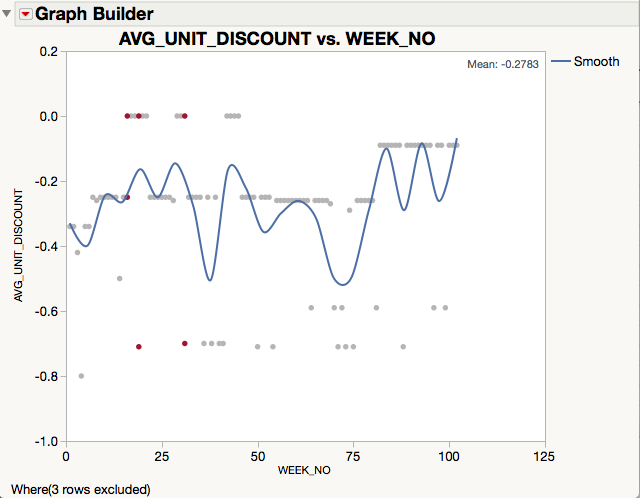
Kevin Gardner

10/25/2016

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.

# Problem 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.



The pricing discount policy generally seems to go between $0.0, -$0.25, -$0.50 and -$0.70. The duration for the $0.0 and -$0.25 discount appear to be about 5 weeks, whereas the duration for -$0.50 and -$0.70 discount is usually only one week. In weeks 16, 19 and 31 (highlighted in red) units were both sold at a discount and at no discount.

# Problem 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.

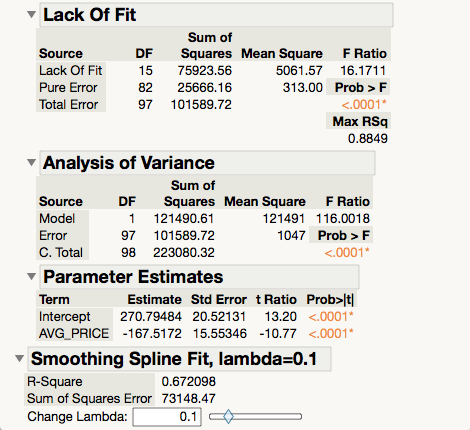
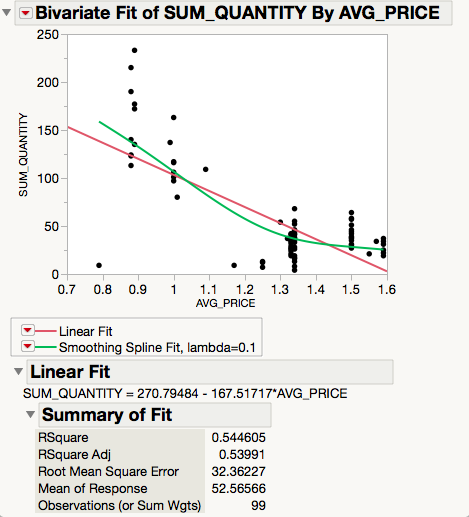
1. Present a single plot with both the spline model and the linear regression model for product 1053690. Report and interpret R2 for each of the two models, showing how the sum of squares error figures in the calculation of R2.

The R2 for the linear regression and the spline model are .54 and .67, respectively. According to the linear regression model, 54% of the variation weekly units sold is explained by the variation in average unit price. Using the splines model, 67% of the variation weekly units sold is explained by the variation in average unit price. R2 is the model sum of squares error divided by total sum of squares error. Hence for the linear model we have

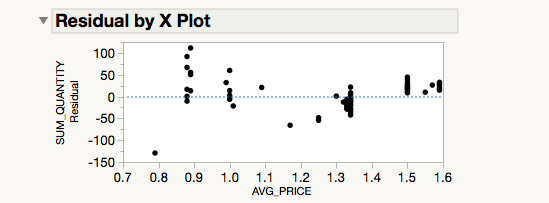
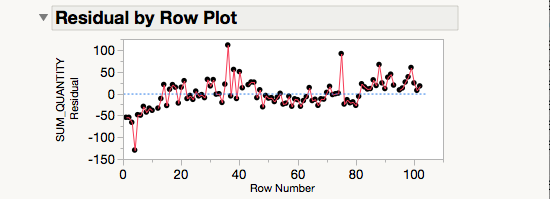
121490.61/223080.32 = 0.544605

and for the linear model we have

1-(73148.47/223080.32) = 0.672098



1. 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.



From the residual plots we see there are larger (negative) residuals for earlier weeks and there might be positive autocorrelation present. We also see that the size of the residuals might be decreasing and then increasing as a function of average unit price.

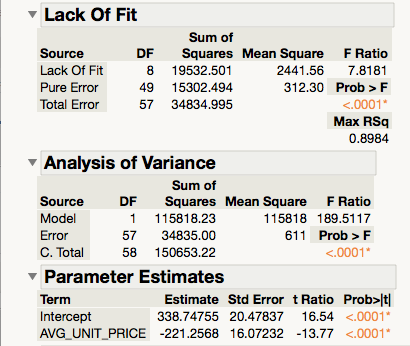
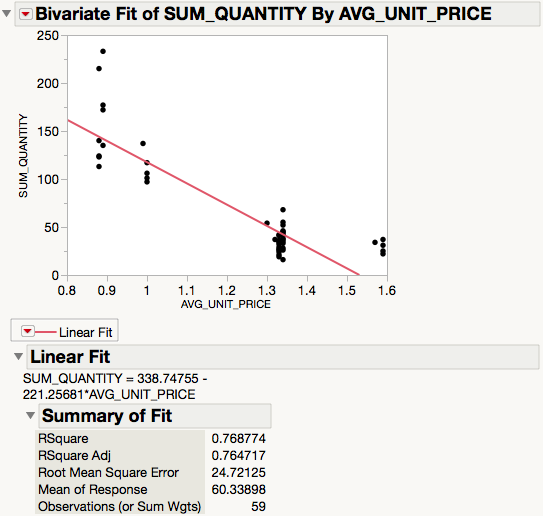
1. Report the F for lack of fit and the conclusion from this test. Discuss how residuals at X=$1.50 and $1.59 contribute to this lack of fit.

From the Lack-of-Fit, we see that the F-ratio is 16.1711. The corresponding p-value is >.0001, which is less than .05 so we reject the null hypothesis and conclude the linear model is not adequate. A more complex model should be considered in order to explain more of the variation in weekly unit sales.

# Problem 3

Exclude weeks 1-21 and 82-102, to focus on the year when $1.34 is the typical price.

1. Fit a straight-line regression model for Weekly quantity sold vs. Mean Unit Sales\_Value



1. Report and interpret R2, the RMSE, and the slope coefficient.

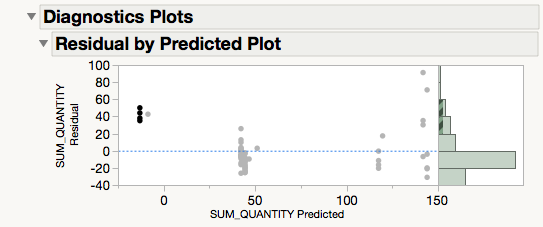
The R2 for this model is .768774. So about 77% of the variation in weekly units sold is explained by the variation in average unit price. The RMSE is 24.7125. So for an average week the prediction from the model is off by about 25 units sold. The slope of the regression -221.2568. So a $1 increase in average unit price is associated with decrease of about 221 units in weekly sales.

1. What is the predicted amount of weekly unit sales at $1.59? Show how you can use the model equation to obtain this value.

The equation of the for the model is given by

So the predicted amount of unit sales at $1.59 is

1. Show a plot of residuals vs. predicted. What does this reveal? Be sure to comment on the residuals at $1.59.

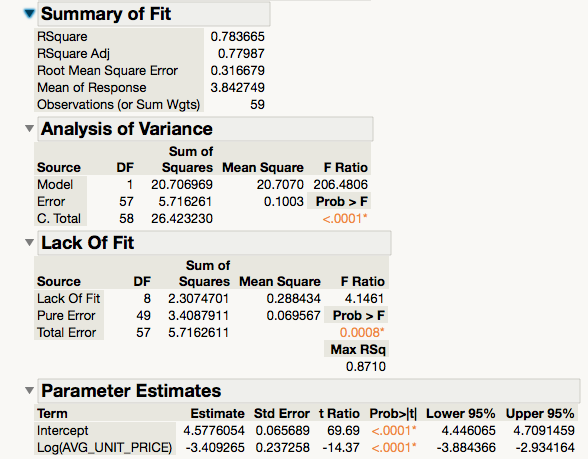
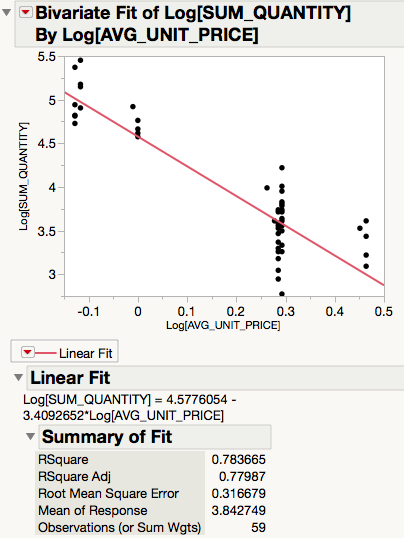


From the plot we see that the residuals are not evenly distributed vertically and there appears to be a U-shaped which suggests perhaps the relationship is not linear. The largest residual in the upper right corner of the plot looks like an outlier and corresponds to a week with 233 units sold at average unit price of $0.89. At $1.59 (highlighted in the plot) we see the predicted units sold is negative and so the positive and relatively large.

# Problem 4

Finally, excluding weeks 1-21 and 82-102, fit a regression model for log(Weekly quantity) vs. log(Mean Unit Sales\_Value).

1. Show a scatterplot that includes this fitted model. Write out the equation of the model.



1. 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.)

We are 95% confident that the true price elasticity (or slope of the demand curve) is between -3.88 and -2.93.

1. Provide a 95% confidence interval for the mean weekly sales at price of $1

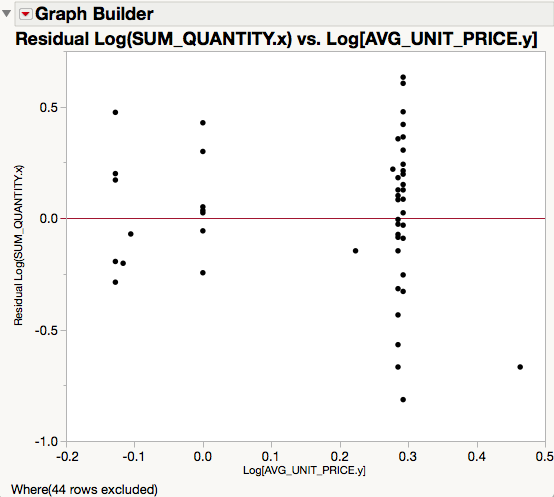
We are 95% confident that the true mean weekly units sold at price of 1$ is between $85.63 (=e^4.45) and $111.053 (=e^4.71).

1. Provide a 95% prediction interval for the weekly sales for a week with price = $1.

We predict with 95% probability that units sold for a week with average price = $1 will be between 50.905371605 and 185.90622204.

1. 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?

The confidence interval brackets the mean of the population and the prediction interval brackets the individual observations sampled from the population.



As Log(AVG\_UNIT\_PRICE) for 844165 increases, we the residuals from the model in #4 become more spread out vertically, so yes, it does appear that sales of 1053690 are impacted by the price of 844165.