Homework 8

Kevin Gardner

Due 11/02/2016

Use the following query to acquire the data. Export the 74,596 records to JMP.

```
SELECT * from carbo_transactions WHERE upc in (3000005970, 3000005300,
9999967727) and geography = 1 ORDER BY upc;
```

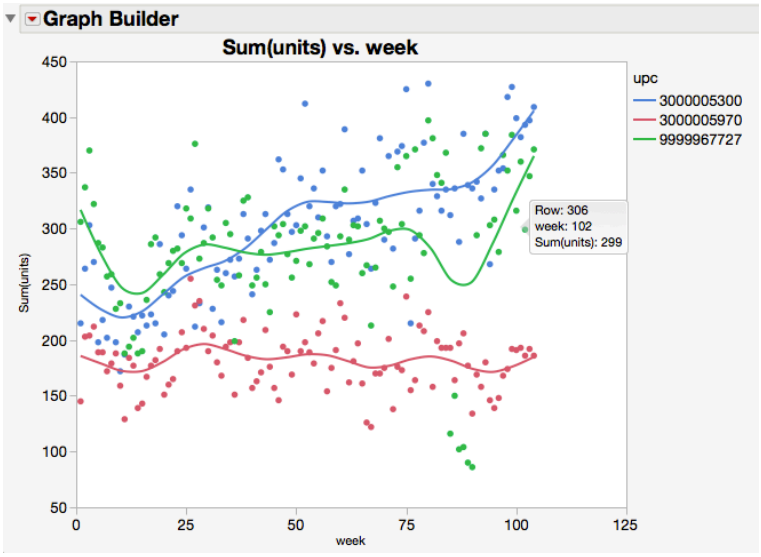
Question 1

Identify what products these three upc denote. Using Graph Builder, plot the weekly sales volume (number of units) for each of these three products over time. Fit a spline to each. Describe what you see.

Querying the carbo_product_lookup for the three upc returns the following result. From this we see that 3000005300, 3000005970 and 9999967727 are Aunt Jemima pancake mix, Aunt Jemima syrup and Private Label syrup, respectively.

upc	product_description	commodity	brand	product_size
3000005300	AJ BUTTERMILK PANCAKE MIX	pancake mixes	Aunt Jemima	32 OZ
3000005970	AUNT JEMIMA ORIGINL SYRUP	syrups	Aunt Jemima	24 OZ
9999967727	PRIVATE LABEL SYRUP PLASTIC BOTTLE	syrups	Private Label	24 OZ

Below is a plot of weekly sales volume for each of these three products over time. From this we see that Aunt Jemima pancake mix (3000005300) tends to be the highest selling of the three products and is closely followed by Aunt Jemima syrup (3000005970). Weekly units sold of 3000005970 and 3000005300 are trending upwards over time, whereas the trend is slightly negative for 9999967727.



Question 2

Using median unit price each week has advantages over using mean unit price per week. Answer the following questions using simple regression models with X = ln(weekly median unit price).

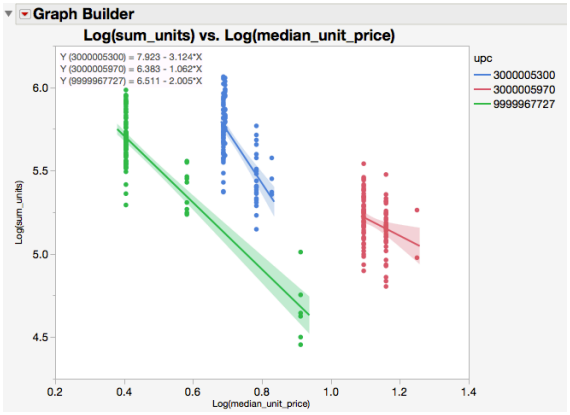
a. Does Aunt Jemima syrup weekly unit sales show price sensitivity? Give a 95% confidence interval for price elasticity of demand.

Aunt Jemima syrup weekly unit sales does shows some price sensitivity in that its demand curve has a negative slope. From the graph this is indicated by the slope of the red line, which smaller larger negative slope coefficient than the other two products.

Interpreting the model for 3000005970 we have that a 1% increase in log median unit price corresponds with about a 1.062% decreases in log sum units sold. We are 95% confident that the price elasticity of demand is between -1.836032 and -0.288879.

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	6.3833847	0.437285	14.60	<.0001*	5.5160316	7.2507377
Log(median_unit_price)	-1.062456	0.390007	-2.72	0.0076*	-1.836032	-0.288879



b. Estimate the price sensitivity for the Kroger store brand (Private Label) syrup and compare this result with that for Aunt Jemima syrup. Does the difference make sense?

Interpreting the model for 9999967727 we have that a 1% increase in log median unit price corresponds with about a 2.005% decrease in log sum units sold. We are 95% confident that the price elasticity of demand is between -2.227121 and -1.782028.

Demand for private label syrup is more sensitive to price than Aunt Jemima syrup and this does not surprise us.

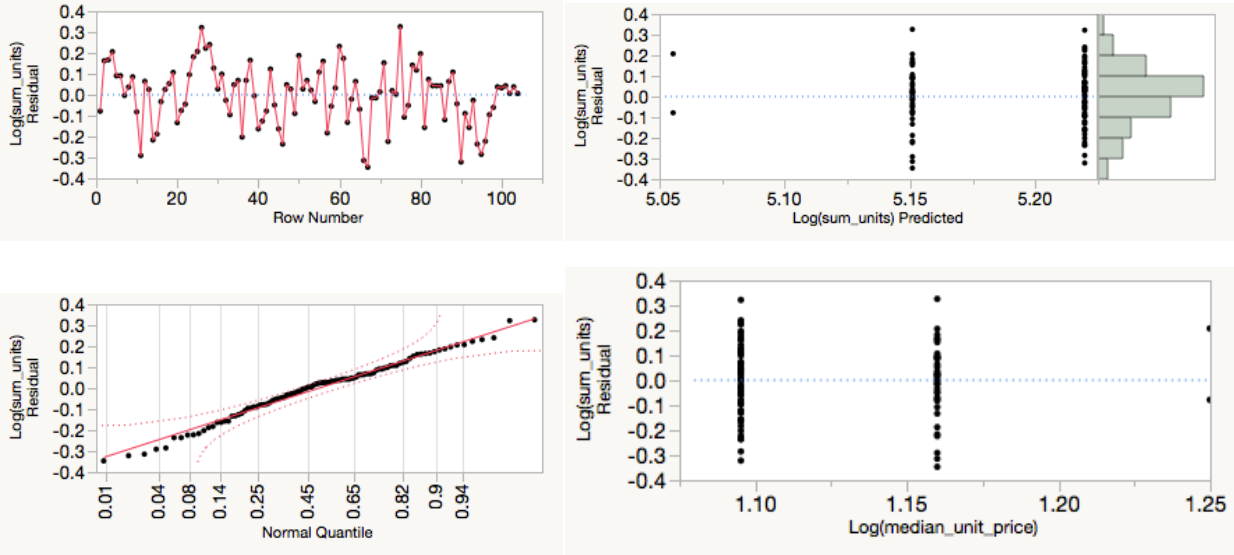
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	6.5111347	0.052592	123.80	<.0001*	6.4068181	6.6154513
Log(median_unit_price)	-2.004575	0.112199	-17.87	<.0001*	-2.227121	-1.782028

Question 3

Show residual plots for the regression models fit for 2a and 2b and try to validate whether these models do or do not satisfy the assumptions of constant price elasticity, constant variance, and independently distributed normal errors (residuals).

Residual plots for 2a - Aunt Jemima syrup (3000005970)

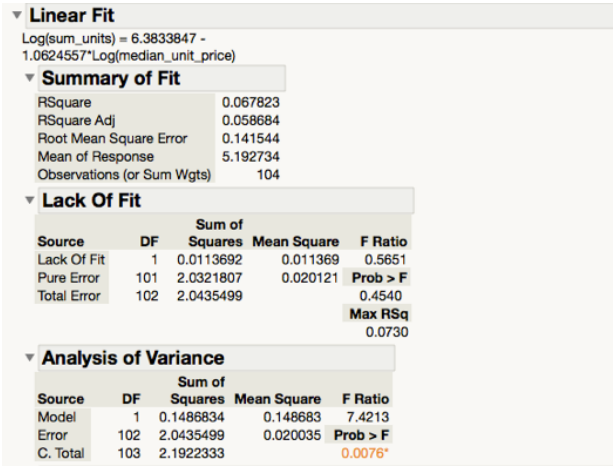


Constant price elasticity

From the Residual by Predicted plot we do not see any strong patterns that suggest nonlinearity. The residuals are somewhat smaller when predicted log sum units is 5.05, but this could be due to the fact that there are only two weeks with this level of predicted log sum units.

The linear model only explains about 7% of the variation in log sum units, but the most complex model does not provide significant improvement. The p-value for lack of fit is 0.4540, so we fail to reject the null hypothesis and conclude that the linear model is

adequate.



Linear Fit

Log(sum_units) = 6.3833847 - 1.0624557*Log(median_unit_price)

Summary of Fit

RSquare	0.067823
RSquare Adj	0.058684
Root Mean Square Error	0.141544
Mean of Response	5.192734
Observations (or Sum Wgts)	104

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	1	0.0113692	0.011369	0.5651
Pure Error	101	2.0321807	0.020121	Prob > F
Total Error	102	2.0435499		0.4540
				Max RSq
				0.0730

Analysis of Variance

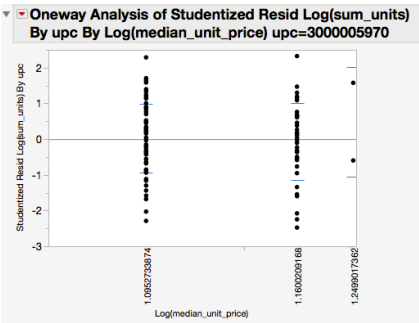
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.1486834	0.148683	7.4213
Error	102	2.0435499	0.020035	Prob > F
C. Total	103	2.1922333		0.0076*

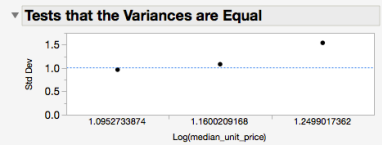
Constant variance

From the Residual by log median unit price plot (bottom right) it looks like the residuals are evenly distributed for the first 2 levels of log median unit price. The residuals are smaller when log median unit price is $1.25, but this could be due to only having two weeks with this level of log median unit price.

By performing a one-way analysis of Studentized Residual by log median unit price we see that Studentized residuals become slightly more positive as log median unit price increases and the second plot we see the standard deviation of the Studentized residual may be increasing as a function of log median unit price. However, all four test tests fail to reject the null hypothesis of constant variance. In particular O'Brien test, which is an ANOVA on the group sample variances, has a p-value of 0.4189. This is not significant so we fail to

reject the null hypothesis and conclude that the variances are not unequal.





Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
1.0952733874	66	0.963482	0.765806	0.749580
1.1600209168	36	1.079684	0.822032	0.815676
1.2499017362	2	1.537246	1.086997	1.086997

Test	F Ratio	DFNum	DFDen	Prob > F
O'Brien[.5]	0.6589	1	100	0.4189
Brown-Forsythe	0.3497	2	101	0.7057
Levene	0.3336	2	101	0.7171
Bartlett	0.4613	2	.	0.6305

Warning: Small sample sizes. Use Caution.

Welch's Test

Welch Anova testing Means Equal, allowing Std Devs Not Equal

F Ratio	DFNum	DFDen	Prob > F
0.1372	2	2.6692	0.8775

Independently distributed normal errors (residuals)

From the Residual by Row Number plot (top left) it looks like positive residuals are more likely to be followed by negative residuals and negative residuals are more likely to be followed by positive residuals, but this is difficult to determine visually.

From the Durbin-Watson test we find the autocorrelation coefficient is 0.3549. The p-value is statistically significant so we reject the null hypothesis and conclude there is positive autocorrelation. This suggests that the errors are not independent.

Durbin-Watson

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
1.2871663	104	0.3549	<.0001*

From the Residual by Normal Quantile plot we do not observe any severe departures from normality. The residual quantiles bend away slightly from the normal quantiles near the tails of the distribution but this does not appear to be systematic.

From the Goodness-of-Fit Test we fail to reject the null hypothesis so can conclude that the errors are approximately normally distributed.

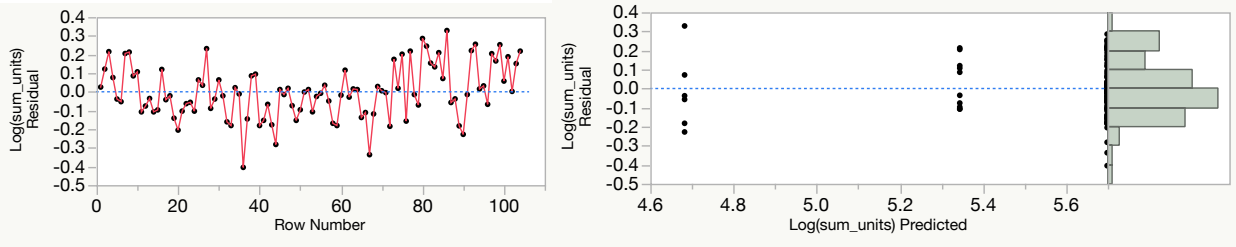
Goodness-of-Fit Test

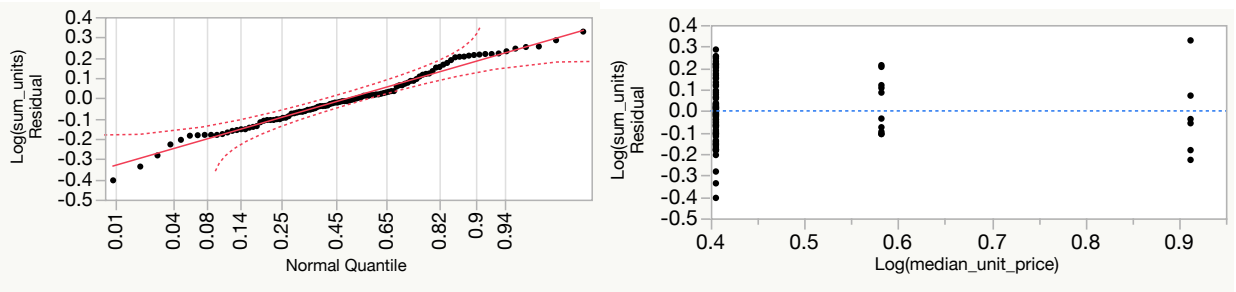
Shapiro-Wilk W Test

W	Prob<W
0.986524	0.3782

Note: Ho = The data is from the Normal distribution. Small p-values reject Ho.

Residual plots for 2b – Private label syrup (9999967727)





Constant price elasticity

From the Residual by Predicted plot___.

Plotting residual log mean units versus the independent variable we see the residuals are less widely distributed for larger values log median unit price. As log median unit price increases, the residuals become more concentrated around 0. This indicates that the relationship between log mean units sold and log median unit price might not be linear and suggests that constant price elasticity is not satisfied.

Linear Fit

Log(sum_units) = 6.5111347 - 2.0045745*Log(median_unit_price)

Summary of Fit

RSquare	0.757836
RSquare Adj	0.755462
Root Mean Square Error	0.143302
Mean of Response	5.605669
Observations (or Sum Wgts)	104

Lack Of Fit

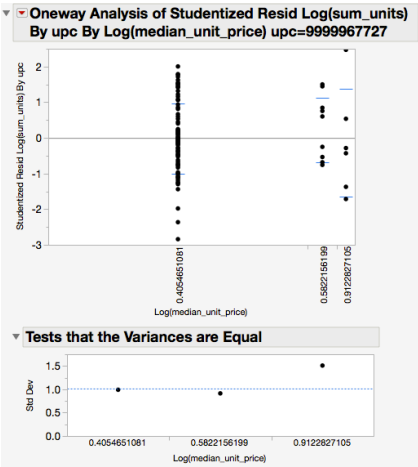
Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	1	0.0116940	0.011694	0.5670
Pure Error	101	2.0829311	0.020623	Prob > F
Total Error	102	2.0946251		0.4532
				Max RSq
				0.7592

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	6.5549890	6.55499	319.2022
Error	102	2.0946251	0.02054	Prob > F
C. Total	103	8.6496142		<.0001*

Constant variance

From the Residual by log median unit price plot (bottom right)



Level	Count	Std Dev	MeanAbsDif to Mean	MeanAbsDif to Median
0.4054651081	88	0.989808	0.762640	0.761351
0.5822156199	10	0.912171	0.813551	0.813551
0.9122827105	6	1.508430	1.090828	1.040087

Test	F Ratio	DFNum	DFDen	Prob > F
O'Brien[.5]	2.2012	2	101	0.1160
Brown-Forsythe	0.5569	2	101	0.5747
Levene	0.7906	2	101	0.4563
Bartlett	1.1314	2	.	0.3226

Welch's Test

Welch Anova testing Means Equal, allowing Std Devs Not Equal

F Ratio	DFNum	DFDen	Prob > F
0.2907	2	9.6174	0.7541

Independently distributed normal errors (residuals)

From the Residual by Row Number plot (top left)

From the Durbin-Watson test we find the autocorrelation coefficient is 0.3642. The p-value is of <.0001 is significant, so we reject the null hypothesis and conclude that positive autocorrelation is present. This suggests that the residuals are not independent.

Durbin-Watson

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
1.248606	104	0.3642	<.0001*

From the Residual by Normal Quantile plot we observe some slight departures from normality. The residual quantiles bend away from the normal quantiles near the tails of the distribution, but our sample may be large enough to accommodate these apparent departures.

From the Goodness-of-Fit Test we fail to reject the null hypothesis, so we can conclude that the errors are approximately normally distributed.

Goodness-of-Fit Test

Shapiro-Wilk W Test

W	Prob<W
0.986524	0.3782

Note: Ho = The data is from the Normal distribution. Small p-values reject Ho.

Question 4

Fit a multiple regression model using Fit Model for Y = ln(weekly Private Label syrup units sold) with X1 = ln(weekly median unit price) and include Week and Week^2 in the models as well. Interpret the following output, explaining how the addition of the terms involving Week.

a. Changes R2.

Summary of Fit

RSquare	0.757836
RSquare Adj	0.755462
Root Mean Square Error	0.143302
Mean of Response	5.605669
Observations (or Sum Wgts)	104

Summary of Fit

RSquare	0.715607
RSquare Adj	0.707075
Root Mean Square Error	34.58089
Mean of Response	281.5288
Observations (or Sum Wgts)	104

b. Changes the RMSE.

c. Changes autocorrelation in the residuals.

Durbin-Watson

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
1.248606	104	0.3642	<.0001*

Durbin-Watson

Durbin-Watson	Number of Obs.	AutoCorrelation	Prob<DW
1.607988	104	0.1950	0.0103*

d. Changes the estimated price elasticity.

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	6.5111347	0.052592	123.80	<.0001*	6.4068181	6.6154513
Log(median_unit_price)	-2.004575	0.112199	-17.87	<.0001*	-2.227121	-1.782028

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%
Intercept	6.4436687	0.050143	128.51	<.0001*	6.3441873	6.5431502
Log(median_unit_price)	-2.139124	0.102564	-20.86	<.0001*	-2.342609	-1.93564
week	0.0011429	0.000417	2.74	0.0072*	0.0003162	0.0019696
(week-52.5)*(week-52.5)	7.5716e-5	1.588e-5	4.77	<.0001*	4.4216e-5	0.0001072

Question 5

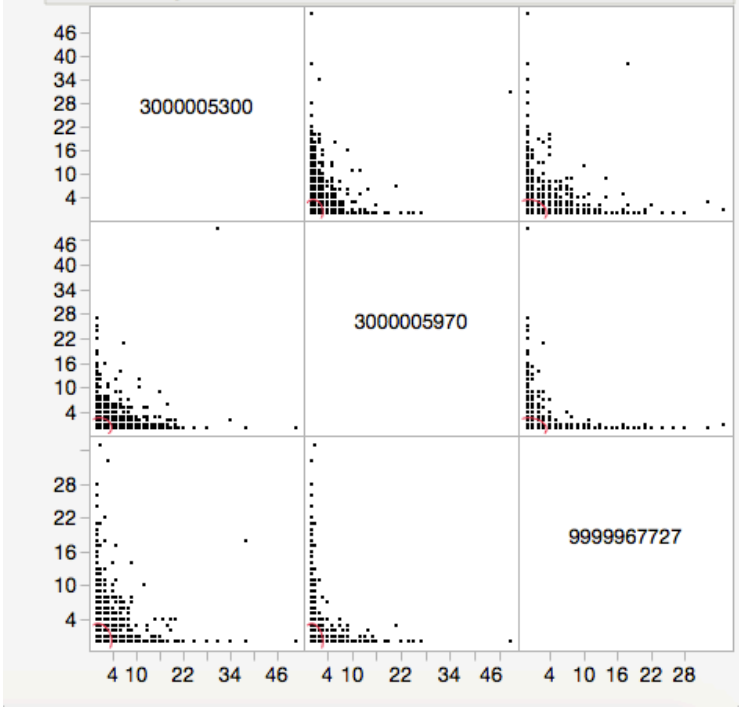
The file HW8 sales by household.jmp lists the number of times each household bought these items. Use Analysis – Multivariate Methods – Multivariate to produce a matrix of scatterplots and correlations. Interpret these three coefficients, explaining why they have the sign and magnitude that they do. What would be the effect of adding (0 0 0) for households with no trips on the correlation coefficients?

Multivariate

Correlations

	3000005300	3000005970	9999967727
3000005300	1.0000	-0.0181	-0.1283
3000005970	-0.0181	1.0000	-0.1673
9999967727	-0.1283	-0.1673	1.0000

Scatterplot Matrix



Pairwise Correlations

Variable	by Variable	Correlation	Count	Lower 95%	Upper 95%	Signif Prob	-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
3000005970	3000005300	-0.0181	44580	-0.0274	-0.0088	0.0001*	
9999967727	3000005300	-0.1283	44580	-0.1374	-0.1192	<.0001*	
9999967727	3000005970	-0.1673	44580	-0.1763	-0.1582	<.0001*	

Extra credit:

Do you see evidence of cross-elasticity for any of these products?