Beer, Wine, and the Thanksgiving Holiday

In ‘92 at the Chicago Dominick’s Finer Foods Store

Olivia Leeson

December 12, 2016

Contents

List of Figures

1 Introduction

This paper aims to understand which food-store factors influence daily beer sales in the Chicago area. Specifically we are interested in whether or not wine sales are correlated with beer sales. If beer sales and wine sales are correlated, then what does that relationship look like? The answer to this would determine whether beer and wine act as complementary or substitution goods in the grocery store environment.

We will also want to test whether other factors such as frozen food sales, customer count, spirits sales, and others have any relationship with total daily beer sales. The goal of this analysis is to better understand aggregated consumer behaviour to aid in decision making for any local food store in the Chicago area.

We will also examine the affect that the Thanksgiving Holiday season may have on beer sales. We would like to test the average change in beer sales due to the Thanksgiving season and examine whether the relationships between beer sales and wine sales changes due to the fact Thanksgiving is near.

1.1 Data

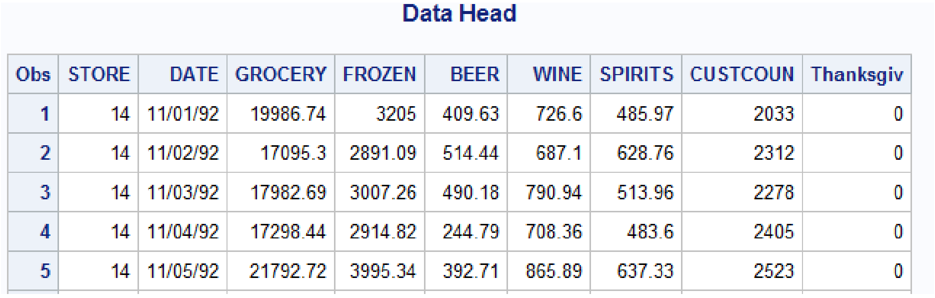
The original data set is pulled from the Dominick's Database. This data set contains store-level data on sales of goods at Dominick's Fine Foods stores in the Chicago Area. The Dominick's data set contains daily sales data of goods with over 3,500 Universal Product Codes (UPCs). This sales data was collected from 1989 to 1994.

In order to answer our questions of interest, we make this massive data set much more manageable. After subsetting and cleaning the original data set, we narrow our scope to 4 stores in the Chicago area. We further narrow our data to the month of November in 1992.

1992 was an a presidential election year and the month of November hosts the Thanksgiving holiday. Thanksgiving in 1992 fell on the 26th of November.

We have coded the variable "Thanksgiv" to be 1 if the day in date is Thanksgiving Day (11/26/1992) or a day in the week leading up to Thanksgiving. We code the weekend prior and the days leading up to Thanksgiving as "1" because many people do their food/grocery shopping in the week or days prior to the Thanksgiving Holiday.

The entire final data set on which we run our analysis contains 120 observations and 9 variables.



Data Head

Obs	STORE	DATE	GROCERY	FROZEN	BEER	WINE	SPIRITS	CUSTCOUN	Thanksgiv
1	14	11/01/92	19986.74	3205	409.63	726.6	485.97	2033	0
2	14	11/02/92	17095.3	2891.09	514.44	687.1	628.76	2312	0
3	14	11/03/92	17982.69	3007.26	490.18	790.94	513.96	2278	0
4	14	11/04/92	17298.44	2914.82	244.79	708.36	483.6	2405	0
5	14	11/05/92	21792.72	3995.34	392.71	865.89	637.33	2523	0

Figure 1: Head of Dominick's Data Set

1.2 Questions of Interest

Our larger question of interest is: how do sales of other goods relate to beer sales? and, how do purchases change due to shopping for Thanksgiving?

1.2.1 Prediction

We would like to predict expected beer sales on a day in the week leading up to Thanksgiving given that the store sold $1500 worth of wine.

1.2.2 Coefficient Interpretation

Given a dollar increase in wine sales in a Dominick's Fine Food store holding all else constant, what is the expected change in beer sales?

2 Multiple Linear Regression

2.1 Plots

First we would like to visualize the relationship between beer and wine sales given whether or not it's close to Thanksgiving

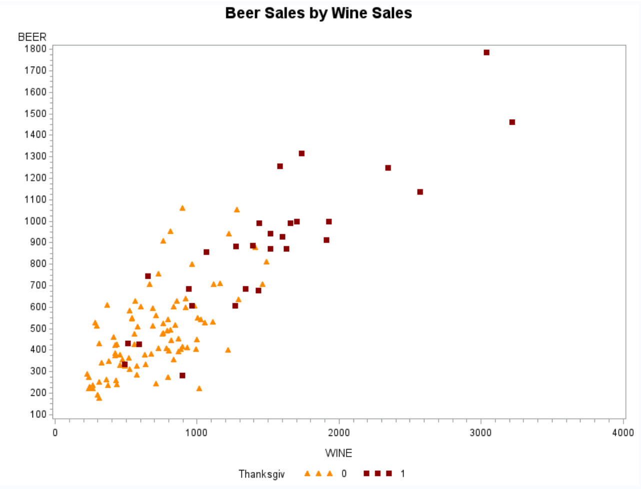


Figure 2: Beer by Wine

There appears to be a linear relationship among beer and wine sales.

We want to view the scatter plot of relationships among all variables in our data. To do this we create a matrix that also includes the histograms for each variable.

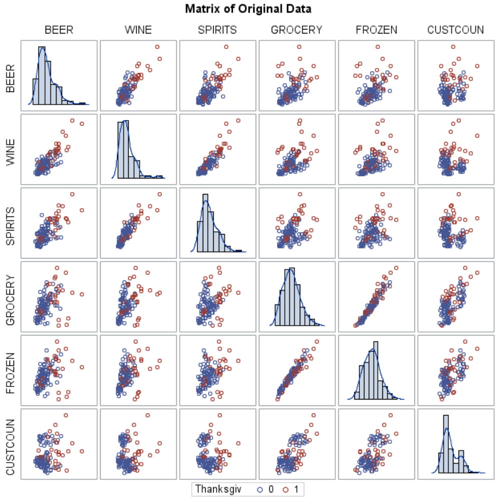


Figure 3: Scatter Plot Matrix of all Variables

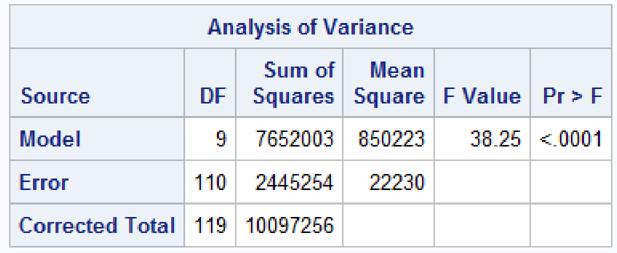
The histograms within the matrix of plots shows us that beer, wine, and spirits have a prominent right skew. The grocery and frozen variables are distributed more normally. The customer count seems to not follow a specific distribution.

Our scatter plots indicate that most variables have a positive relationship with one another. 'Customer count' seems to be the least strongly correlated with the other variables. The grocery and frozen food sales are very tightly correlated. Due to this strong correlation between grocery and food sales we will omit the grocery sales parameter from our full model, as it is captured in the frozen food parameter. Wine, beer, and spirits are all strongly positively correlated. We will leave spirits in the full model for now, but it is likely that its correlation with wine sales will cause us to drop 'spirits' from the final model.

2.2 Model

$$\begin{multline} Beer' = \beta_0' + \beta_1' Wine + \beta_2' Spirits + \beta_3' Custcoun + \beta_4' Frozen \\ + \beta_5' Thanksgiv + \beta_6'(Thanksgiv)(Wine) + \beta_7'(Thanksgiv)(Spirits) \\ + \beta_8'(Thanksgiv)(Custcoun) + \beta_9'(Thanksgiv)(Frozen) \quad (1) \end{multline}$$

We use SAS to run a multiple linear regression model on the data.



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	7652003	850223	38.25	<.0001
Error	110	2445254	22230		
Corrected Total	119	10097256			

Figure 4: ANOVA for Full Model

Our ANOVA gives us an F-value of 38.25 and a Pvalue of less than .0001, meaning that our model significantly accounts for the variability in beer sales.

Our R-Squared value is 0.7578 and the adjusted R-Squared is 0.7380. The parameters estimated in our model account for about 75.78% of the variance in beer sales.

2.3 Fit Assessment

We take a look at the plots of residuals for the regression. We have a cluster of points with high leverage as evidenced in the leverage plot. The studentized residuals are mostly between -2 and 2, with a few larger residuals around -3 and 2.5.

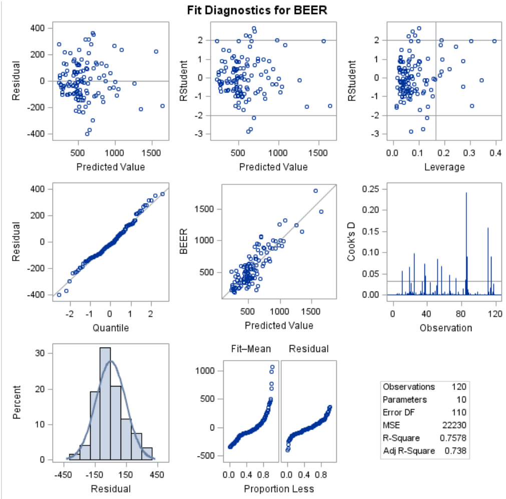


Figure 5: ANOVA for Full Model

The residual qqplot follow a straight 45 degree line. The histogram is also evidence for a normal distribution of residuals. The normal distribution of residuals is evidence that the linear regression is a good fit for the data.

We see from the Cook's D plot that there is one observation that has a significantly larger Cook's D than the rest. This point has high influence and leverage for the model.

In order to seek out the observation, we plot the observations by Cook's D.

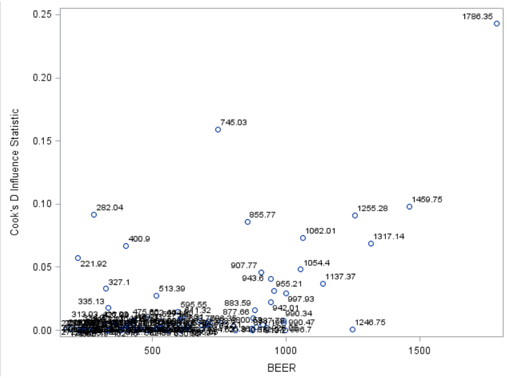


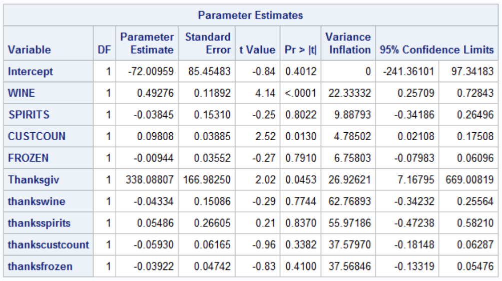
Figure 6: Scatter Plot of Cook's D

The observation in question is observation 85: Beer sales are $1,786.35 and the date of the observation is 11/25/1992. This is the day before Thanksgiving. We should not be too surprised that there is a statistically larger value of beer sales on the day before Thanksgiving.

We decide to include this influential observation in our analysis, as it is an important example of the phenomenon we are trying to capture and there is no observation to suggest that this observation is an error.

2.4 Parameter Estimate Interpretation

Our parameter estimates for the model are shown below.



Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation	95% Confidence Limits	
Intercept	1	-72.00959	85.45483	-0.84	0.4012	0	-241.36101	97.34183
WINE	1	0.49276	0.11892	4.14	<.0001	22.33332	0.25709	0.72843
SPIRITS	1	-0.03845	0.15310	-0.25	0.8022	9.88793	-0.34186	0.26496
CUSTCOUN	1	0.09808	0.03885	2.52	0.0130	4.78502	0.02108	0.17508
FROZEN	1	-0.00944	0.03552	-0.27	0.7910	6.75803	-0.07983	0.06096
Thanksgiv	1	338.08807	166.98250	2.02	0.0453	26.92621	7.16795	669.00819
thankswine	1	-0.04334	0.15086	-0.29	0.7744	62.76893	-0.34232	0.25564
thanksspirits	1	0.05486	0.26605	0.21	0.8370	55.97186	-0.47238	0.58210
thankscustcount	1	-0.05930	0.06165	-0.96	0.3382	37.57970	-0.18148	0.06287
thanksfrozen	1	-0.03922	0.04742	-0.83	0.4100	37.56846	-0.13319	0.05476

Figure 7: Parameter Estimates for Full Model

Substituting the parameter estimates into the model, the full model including coefficients is :

$$\begin{aligned} Beer' = -72 + 0.493Wine - 0.038Spirits + 0.98Custcoun - 0.009Frozen \\ + 338.09Thanksgiv - 0.04(Thanksgiv)(Wine) + 0.055(Thanksgiv)(Spirits) \\ - 0.059(Thanksgiv)(Custcoun) - 0.04(Thanksgiv)(Frozen) \quad (2) \end{aligned}$$

The parameters deemed statistically significant at the 95% level are shown in the table 1 below.

Variable	Parameter Estimate	P-Value
Wine Sales	0.49276	less than .0001
Customer Count	0.09808	0.0130
Thanksgiving	338.08807	0.0453

Table 1: Significant Parameters from Full Model

The intercept for this model is not statistically nor practically significant as it is a negative value. A negative value of beer sales would only be possible due to returned purchases. We are not interested with the returned beer purchases in this analysis.

Wine : For every \$1 increase in wine sales, we expect to see a corresponding increase of beer sales of 49 cents. We are 95% confident that the mean change in beer sales given a dollar increase in wine sales is between 26 cents and 73 cents.

Customer Count : For every 1 person increase in customer we expect to see about a 10 cent increase in beer sales. We are 95% confident that the mean change in mean beer sales given an extra customer in the store that day is between 2 cents and 18 cents.

Thanksgiving : For any of the 7 days leading up to and including Thanksgiving day, we expect to see an increase in mean beer sales of \$338.09. We are 95% confident that the increase in beer sales is between \$7.17 and \$669.00 given it is one of the seven days leading up to Thanksgiving.

We notice that many of our variable parameters are not statistically significant for the model. We will want to run a variable selection on our model to only capture parameters we find to be significant.

3 Variable Selection

In order to select only the statistically significant parameters in a systematic manner we run a variable selection procedure. A forward, stepwise, and backward selection process are all analyzed. We choose to use the model selected through the backward selection process. This model leaves us with the highest adjusted R-squared and the most significant parameters.

The adjusted R-squared of the full model is: 0.738 The adjusted R-squared for our new model is: 0.746

3.1 New Model

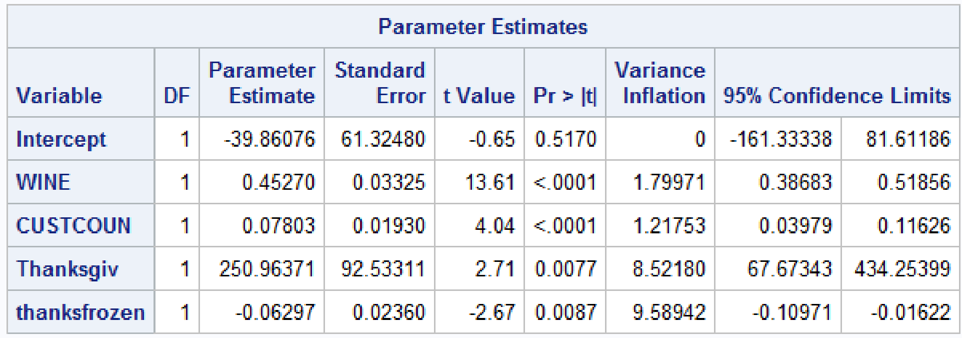
Our new model is:

$$\begin{aligned} Beer' = \beta_0' + \beta_1' Wine + \beta_2' Custcoun + \beta_3' Thanksgiv + \\ \beta_4' (Thanksgiv)(Frozen) \end{aligned} \quad (3)$$

The ANOVA for our new model gives us a much higher F-value of 88.28. This is extremely strong evidence that the model is statistically significant in explaining the variance in beer sales.

The adjusted R-squared is 0.746 and the multiple R-squared is 0.7543. 75.43% of the variance in beer sales is described by the model.

Our parameters are shown in the figure below.



Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation	95% Confidence Limits	
Intercept	1	-39.86076	61.32480	-0.65	0.5170	0	-161.33338	81.61186
WINE	1	0.45270	0.03325	13.61	<.0001	1.79971	0.38683	0.51856
CUSTCOUN	1	0.07803	0.01930	4.04	<.0001	1.21753	0.03979	0.11626
Thanksgiv	1	250.96371	92.53311	2.71	0.0077	8.52180	67.67343	434.25399
thanksfrozen	1	-0.06297	0.02360	-2.67	0.0087	9.58942	-0.10971	-0.01622

Figure 8: Parameter Estimates of New Model

Our model equation is then:

$$Beer' = -39.86 + 0.453Wine + 0.078Custcoun + 250.96Thanksgiv \\ - 0.063(Thanksgiv)(Frozen) \quad (4)$$

The intercept for this model is not statistically nor practically significant as was the case for the full model.

Wine : For every \$1 increase in wine sales, we expect to see a corresponding increase of beer sales of 45 cents. We are 95% confident that the mean change in beer sales given a dollar increase in wine sales is between 38 cents and 52 cents.

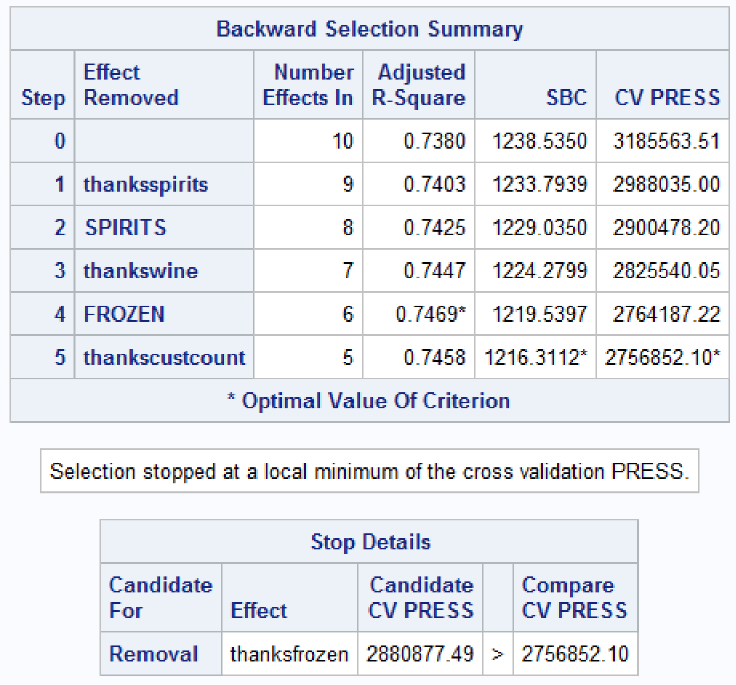
Customer Count : For every 1 person increase in customer we expect to see about a 8 cent increase in beer sales. We are 95% confident that the mean change in beer sales given an extra customer in the store that day is between 4 cents and 12 cents.

Thanksgiving : For any of the 7 days leading up to and including Thanksgiving day, we expect to see an increase in beer sales of \$250.96. We are 95% confident that the increase in mean beer sales is between \$67.67 and \$434.25 given it is one of the seven days leading up to Thanksgiving.

Thanksgiving and Frozen Foods Interaction : An interesting finding from this model selection is that the interaction term between Thanksgiving and Frozen Food sales is statistically significant. This means that if it is Thanksgiving, for a given dollar increase in frozen food sales, we expect to see a decrease in mean beer sales of about 6 cents. We are 95% confident that this decrease in mean beer sales is between 2 cents and 11 cents.

3.2 Internal Cross Validation

We will perform a five-fold internal cross validation to test our selected new model. The cross validation shows us the parameters dropped from the full model by CV PRESS comparison.



Backward Selection Summary					
Step	Effect Removed	Number Effects In	Adjusted R-Square	SBC	CV PRESS
0		10	0.7380	1238.5350	3185563.51
1	thanksspirits	9	0.7403	1233.7939	2988035.00
2	SPIRITS	8	0.7425	1229.0350	2900478.20
3	thankswine	7	0.7447	1224.2799	2825540.05
4	FROZEN	6	0.7469*	1219.5397	2764187.22
5	thankscustcount	5	0.7458	1216.3112*	2756852.10*

* Optimal Value Of Criterion

Selection stopped at a local minimum of the cross validation PRESS.

Stop Details				
Candidate For	Effect	Candidate CV PRESS		Compare CV PRESS
Removal	thanksfrozen	2880877.49	>	2756852.10

Figure 9: Parameters Dropped Through Cross Validation

The five-fold internal cross validation leaves us with the new model parameters described in the figure 8.

3.3 Answers to Question of Interest

We would like to predict expected beer sales on a day in the week leading up to Thanksgiving given that the store sold $1500 worth of wine. For the other parameters of Customer Count and Frozen Food sales we will use their mean observation in the model to answer our question.

$$\text{Mean Customer Count} = 2719.56$$
$$\text{Mean Frozen Food Sales} = 3417.98$$

We calculate our predicted beer sales following the equation:

$$\text{Beer'} = \text{-}39.86 + 0.453(1500) + 0.078(2719.56) + 250.96$$
$$- \ 0.063(3417.98)$$
$$\text{Beer'} = 887.13$$

In order to check this we add a row to our data with the given variables and 'beer' as unknown. We run the same model regression on this data and confirm the predicted mean beer sales of $887.13. We also collect the 95% confidence interval.

We predict that the mean beer sales on a given day that Dominick's Fine Foods sells $1,500.00 in wine is $887.13. We are 95% confident that the mean beer sales will be between $828.18 and $946.0752.

4 Difference of Means

In order to determine whether there is a significant difference in mean beer sales during the week leading up to Thanksgiving Day we perform a difference of means test.

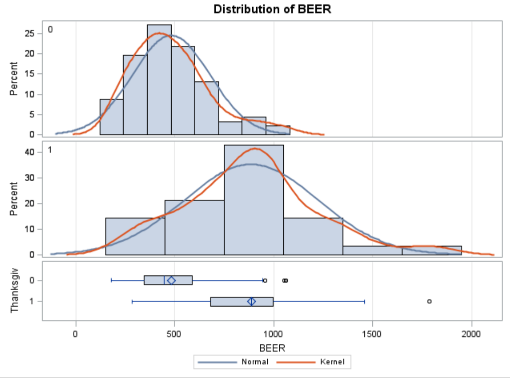


Figure 10: Histograms of Beer Sales by Thanksgiving

It appears that histograms for beer sales in both cases is fairly normal with a right skew. Our sample size should be large enough to ensure a parametric test would be robust to these deviation from the normal distribution.

The variances are not equal for the Thanksgiving and Not Thanksgiving group. We can tell that the distribution during the week leading up to Thanksgiving is wider and includes higher values for beer sales.

Each observation is independent. Dependency may arise if a store is out of stock of required item on a day given the sales of the day previous. We think this dependency is unlikely and proceed under the assumption that each days' sales are independent of one another.

In order to atone for our differences in variance, we will proceed with a Welch's T-Test. The Welch's T-Test does not assume equality of variances for the groups and keeps the power of being able to make inferences on the mean of the groups.

4.0.1 Step One: Formulate Hypothesis

Null Hypothesis
$H_0 : \mu_T \neq \mu_0$

Alternative Hypothesis
H$_a$: $\mu_T > \mu_0$

The null hypothesis is that the mean beer sales during the regular month of November and the week leading up to Thanksgiving is the same.

The alternative hypothesis is that the mean beer sales during the week of Thanksgiving is greater than the mean beer sales for the rest of the month.

4.0.2 Step Two: Critical Value

Under an assumption of equal variances we would be working with n-2 degrees of freedom, or 120-2 = 118. However, under a Welch's t-test we will use the Satterthwaite method to calculate our t-Value. The degrees of freedom for this method, calculated by SAS, will be 32.64. degrees of freedom. We would like to test the difference of means to a 95% significance level.

Our critical value is thus the t-value at a 95% significance level with 32.64 degrees of freedom.

The one-sided critical T value is :

$$1.693$$

4.0.3 Step Three: Test Statistic

We run the Welch's T-test on our data to the following results.

Method	Variances	DF	tValue	pValue
Satterthwaite	Unequal	32.64	6.03	less than .0001

Table 2: tValue for Welch's tTest

Our test statistic is:

$$6.03$$

4.0.4 Step Four: P-Value

Our one-sided pValue is:

less than .001

4.0.5 Step Five: Null Hypothesis Determination

We reject the null hypothesis that there is no difference in mean beer sales between the Thanksgiving days and the normal days.

4.0.6 Step Six: Conclusion

There is strong evidence to suggest that the mean beer sales made during the week leading up to Thanksgiving is statistically greater than the mean beer sales made during the rest of the month of November.

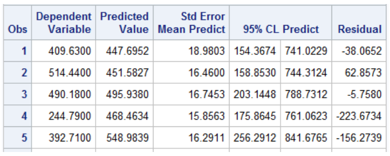
The difference in mean beer sales is $405.00 and we are 95% confident that this difference in beer sales is between $268.6 and $542.10. This finding is relevant to the population of Dominick's Fine Food Stores in the Chicago area during the month of November in 1992.

4.0.7 Means Inference versus Regression Inference

Our means test provides us with a mean difference of $405.00 given whether or not it is near Thanksgiving. Our final regression model indicates that the difference in mean beer sales is $250.96 (our parameter estimate for the Thanksgiving condition). The parameter estimate about the mean difference in the regression is not captured in our confidence interval about the mean for our Welch's Ttest. The differences between these two results are likely due to the additional variation and affects caused by the addition of other factors in the regression model not captured through the difference of means test.

5 Prediction Intervals

In order to test the strength of individual predictions made by the model we calculate the prediction intervals for an individual observation of beer sales given wine sales.



Obs	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Predict		Residual
1	409.6300	447.6952	18.9803	154.3674	741.0229	-38.0652
2	514.4400	451.5827	16.4600	158.8530	744.3124	62.8573
3	490.1800	495.9380	16.7453	203.1448	788.7312	-5.7580
4	244.7900	468.4634	15.8563	175.8645	761.0623	-223.6734
5	392.7100	548.9839	16.2911	256.2912	841.6765	-156.2739

Figure 11: Prediction Intervals for New Model

From this table we can view the original observation of beer sales, the predicted value of beer sales calculated by our model, and the prediction interval for an individual days' beer sales.

After reviewing our actual observations against the prediction interval, we find in 6 cases (observations 19, 37, 38, 43, 66, and 114) that the observed beer sales for that day does not fall within our prediction interval.

There are 120 observations, which means that the observation falls within our prediction interval (6/120) = 95% of the time.

This is exactly what we would expect to observe in the data, as the prediction intervals are calculated based on a margin of error calculated from a t-statistic at the 95% significance level.

6 Conclusion

Beer sales and wine sales are positively correlated. We find that for a dollar increase in wine sales on a given day we would expect to see a 45 cent increase in mean beer sales. Beer sales are also correlated with the number of customers trafficking the store, whether or not it is the week leading up to Thanksgiving, and the interaction between the Thanksgiving holiday and frozen food sales. Our data is observational, so we cannot claim causation here. Luckily, we are less concerned with causation for this problem and are interested in how the variables correlate. This subset of data from which we ran our analysis is a sample of the entire data population that is available on the Dominick's database. The data was not selected randomly, as we deliberately chose to study the month of November in 1992 for a specific subset of stores. However, we can claim that our findings are valid for Dominick's Fine Food stores in Chicago in November of 1992.

6.1 Ideas for Further Analysis

As previously mentioned, this subset of data is only a sample of the data we have available. Our analysis brought up many questions that could be broached through further analysis. Some further questions of interest are:

- How do our relationships change if we group the Thanksgiving Days more precisely? That is, if we were to categorize our Thanksgiving variable into three categories instead of two for: regular day, 6 days leading up to Thanksgiving, and the day before Thanksgiving.

- Does this model hold if we apply to it a different subset of stores within our data set?

- How is the relationship of beer sales to other factors changed during other times of the year?

This dataset is exciting because there is much more analysis we can achieve by increasing our scope. This will be kept in mind for the future.

7 Code

7.1 Data Cleaning

```
data project.dominicks;
        set project.ccount;
        if Store= 8 or store=32;
        if Date =: ’9211’;
        if Date <= 921119 or Date >= 921127 then Thanksgiv = 0;
        else Thanksgiv=1;
        if Store = 8 then Price = ”L”;
        else Price = ”H”;
        Date = Input(PUT(Date,8.),YYMMDD8.);
        FORMAT Date MMDDYY8.;
        keep store date beer liqcoup spirits wine thanksgiv price;
run;
```

7.2 Plots

```
#MATRIX SCATTER
title "Matrix of Original Data";
proc sgscatter data = dominicks;
matrix beer wine spirits grocery frozen custcoun/ group=thanksgiv diagon
run;
title;

#WINE BY BEER SCATTER
SYMBOL2 V=squarefilled C=darkred I=none;
SYMBOL1 V=trianglefilled C=darkorange I=none;
title "Beer Sales by Wine Sales";
PROC GPLOT DATA=dominicks;
PLOT beer*wine=thanksgiv;
run; quit;
title;
```

7.3 Transformation Testing

```
###THIS IS BACKGROUND TESTING NOT SHOWN IN PAPER
data loglindom;
set dominicks;
logbeer= log(beer);
run;

data linlogdom;
set dominicks;
logwine = log(wine);
logspirits = log(spirits);
logfrozen = log(frozen);
loggrocery = log(grocery);
logcustcoun = log(custcoun);
run;

data loglogdom;
set dominicks;
logbeer = log(beer);
```

```
logwine = log(wine);
logspirits = log(spirits);
logfrozen = log(frozen);
loggrocery = log(grocery);
logcustcoun = log(custcoun);
run;


title ”no transform”;
proc reg data = dominicks;
model beer = wine;
run;
quit;
title;

title ”log−lin”;
proc reg data = loglindom;
model logbeer = wine;
run;
quit;
title;

title ”linlog”;
proc reg data = linlogdom;
model beer = logwine;
run;
quit;
title;

title ”loglog”;
proc reg data = loglogdom;
model logbeer = logwine;
run;
quit;
title;
```

7.4 Full Model Regression

```
###CREATE INTERACTION SET
data interact;
set dominicks;
thankswine = wine*thanksgiv;
thanksspirits = spirits*thanksgiv;
thankscustcount = custcoun*thanksgiv;
thanksgrocery = grocery*thanksgiv;
thanksfrozen = frozen*thanksgiv;
run;

###RUN REGRESSION
title "Full Model Regression";
proc reg data = interact;
model beer = wine spirits custcoun frozen  Thanksgiv thankswine
thanksspirits thankscustcount thanksfrozen / vif clb;
output out = t student=res cookd = cookd h = lev p =yhat
run;
title;
quit;
```

7.5 Influential Point Analysis

```
#FIND THE POINT
title "Finding Influential Obs";
proc sgplot data = t;
scatter y = cookd x = beer / datalabel = beer;
run;
quit;
title;

#REMOVE OBSERVATION FROM SET
data nohighinfluence;
set interact;
if beer = 1786.35 then delete;
run;

#RUN REGRESSION SANS OBSERVATION
title "Regression without Influential Obs";
```

```
proc reg data = nohighinfluence;
model beer = wine spirits custcoun frozen  Thanksgiv thankswine
thanksspirits thankscustcount thanksfrozen / vif clb;
output out = t student=res cookd = cookd h = lev p =yhat
run;
title;
quit;
```

7.6 Variable Selection

```
title”Variable Selction”;
proc glmselect data=interact plot=CriterionPanel;
model beer = wine spirits custcoun frozen  Thanksgiv thankswine
thanksspirits thankscustcount thanksfrozen
/ selection=backward stats=all;
run;
quit;
title;
```

7.7 New Model Regression

```
title ”New Model Regression”;
proc reg data = interact;
model beer = wine custcoun Thanksgiv thanksfrozen/clm vif clb;
output out = t student=res cookd = cookd h = lev p =yhat
run;
title;
quit;
```

7.8 Cross Validation

```
title”Cross Validation”;
proc glmselect data=interact plot=CriterionPanel;
model beer = wine spirits custcoun frozen  Thanksgiv thankswine
thanksspirits thankscustcount thanksfrozen
/ selection=backward(stop=cv) cvmethod=random(5) stats=all;
run;
quit;
title;
```

7.9 Prediction for QOI

```
#Create new set with missing beer sales to predict
data predict;
set interact end=eof;
output;
if eof then do;
        store = 14;
        date = .;
        frozen=3417.98;
        beer=.;
        wine=1500;
        spirits = .;
        custcoun = 2719.56;
        thanksgiv = 1;
        thanksfrozen =3417.98;
   output;
  end;
run;


#Run Regression
title "New Model Regression Prediction Intervals";
proc reg data = predict;
model beer = wine custcoun Thanksgiv thanksfrozen/ clm vif clb;
run;
title;
```

7.10 Ttest - Welch's

```
###New data set with just required variables
data justthanksgiving;
set dominicks;
keep beer thanksgiv;
run;

proc sort data=justthanksgiving;
by thanksgiv;
```

```
run;

###Ttest using Satterthwaite Output
title "Welch's ttest";
proc ttest data = justthanksgiving order=data sides=l;
class thanksgiv;
var beer;
run;
quit;
title;

###View plots for assumption testing
ods select histogram;
ods select qqplot;
proc univariate data = justthanksgiving plots;
by thanksgiv;
histogram beer;
qqplot beer;
run;
quit;

###Run GLM Welch's for confidence intervals
title "GLM Welch's";
proc glm data = justthanksgiving;
class thanksgiv;
model beer = thanksgiv;
means thanksgiv/hovtest=bf welch;
lsmeans thanksgiv/cl;
run;
quit;
title;
```

7.11 Print all Prediction Intervals

```
title "New Model Regression Prediction Intervals";
proc reg data = interact;
model beer = wine custcoun Thanksgiv thanksfrozen/ cli vif clb;
run;
```

```
title;
quit;
```