THE NATIONAL UNIVERSITY OF SINGAPORE
Master of Science Business Analytics

DSC5101 Analytics in Managerial Economics
AY2017/2018 SEM I

Project 3
Estimating the Price Discrimination of US Pumpkin Price

Name	:	Dong Gang	(E0229781)
		Swetha Narayanan	(A0074604J)
		Wang Shenghao	(A0105772H)
		Wang Shuai Max	(A0092174E)

Date of Submission : 7 November 2017

1 Executive Summary

The objective of this project is to evaluate the level of price discrimination in the US pumpkin markets. We first establish linear regression models to study the relationship between the net price of pumpkins and their physical properties including pumpkin variety and size, which directly determines the pumpkin prices. The residual net prices are derived and regressed against US cities and pumpkin origins, to examine the potential third-degree price discrimination. Robustness tests are further performed to address the time effect and the variation of transportation cost. Second-degree price discrimination with respect to pumpkin variety and third-degree price discrimination with respect to US cities are identified. We conclude that consumers in higher-paying cities tend to be charged higher prices for same types of pumpkins. The corresponding price elasticities are derived in the remaining part of the study

Key Findings

- Third price discrimination is found between different cities. The net price difference has significant relationship with the city's' personal income and consumer price index(CPI).
- Second price discrimination is identified between different packages, the possible reason is that the seller use quantity-discounting menu price to increase the profit.
- The Demand Elasticity matches closely with the personal income. Higher income generally implies relatively lesser elasticity of demand.

Implication

- To maximize the product profit, seller could utilize third degree price discrimination of the net price based on the city's' economic factor such as personal income and consumer price index(CPI).
- For fresh perishable product, transportation method is important. Seller should select the most efficient and worthy transportation method to balance the transportation cost and the transportation duration.
- Based on the customer demand, the seller can package the pumpkin and offer a quantity-discounting menu price to convert some consumers surplus into profit.

2 Introduction

Price discrimination refers to the exercise of offering a same product to different consumers at different prices. It is a common pricing strategy that is frequently adopted by most businesses to achieve the purpose of profit maximization. In general, price discrimination can be classified into three categories, including personalized pricing (first-degree), menu-pricing (second-degree), and group-pricing (third-degree). Second-degree and third-degree price discriminations are common in agricultural markets.

In order to study the price discrimination in the US pumpkin market, we constructed the dataset of pumpkin prices with the recent one year's Specialty Crops Terminal Markets Standard Reports distributed by the United States Department of Agriculture. We estimate the net prices of pumpkins by subtracting the approximate transportation cost from the average pumpkin prices. Linear regression models are built to evaluate the significance of correlation between pumpkin price and city as well as pumpkin variety. Furthermore, we examined the robustness of our analysis with respect to the time effect and the variation of unit shipping cost. In the last part of our study, the demand elasticity of different cities are evaluated and correlated with the city income.

3 Baseline Model and Tests

3.1 Third Degree Price Discrimination

The baseline model of pumpkin prices is derived based on the following assumptions.

- The change of pumpkin price over time is negligible.
- The transportation of pumpkins between US cities is via truck with a uniform rate of 0.37.
- Supply of pumpkins is constant over the time period under study.
- The cost incurred from growing pumpkins is considered at the same level over different origins.

The baseline model is formulated with two steps of linear regression. Firstly, we calculate the net price of pumpkins by subtracting the estimated shipping cost from the average pumpkin prices. We make a hypothesis that the pumpkin price has no correlation with the cities. The net price of pumpkins is regressed against the pumpkin variety and size. From the results shown in the following table, it can be observed that the net price of pumpkins has close relationship with the variety of pumpkins. After counteracting the effect of variety, we regress the net price residual against city and origin separately. Half of the cities under study have significant coefficients. Additionally, cities with larger coefficients correspond to higher average annual salary, which indicates the existence of third-degree discrimination with respect to cities. It is easy to understand that people with better standard of living tend to have higher willingness to pay for their food and commodities.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	161.099	26.038	6.187	8.18E-10	***
VarietyBLUE TYPE	49.692	16.591	2.995	0.0028	**
VarietyCINDERELLA	23.15	11.91	1.944	0.05214	.
VarietyFAIRYTALE	27.004	11.13	2.426	0.0154	*
VarietyHOWDEN TYPE	-30.945	9.609	-3.22	0.00131	**
VarietyHOWDEN WHITE TYPE	-9.746	12.571	-0.775	0.43833	
VarietyKNUCKLE HEAD	30.855	16.298	1.893	0.05855	.
VarietyMINIATURE	-115.744	11.348	-10.199	< 2e-16	***
VarietyMIXED HEIRLOOM VARIETIES	10.526	14.632	0.719	0.47205	
VarietyPIE TYPE	-29.056	10.183	-2.853	0.00439	**
Item.Sizejbo	-20.847	24.92	-0.837	0.40299	
Item.Sizelge	-21.87	24.551	-0.891	0.37319	
Item.Sizemed	-34.714	24.585	-1.412	0.15818	
Item.Sizemed-lge	18.052	25.013	0.722	0.4706	
Item.Sizesml	-25.974	24.79	-1.048	0.29495	
Item.Sizexlge	14.204	24.699	0.575	0.56534	

	Estimate	Std. Error	t value	Pr(>|t|)	
City.NameATLANTA	-23.969	7.555	-3.173	0.00155	**
City.NameBALTIMORE	9.937	9.185	1.082	0.27949	
City.NameBOSTON	51.84	8.259	6.277	4.68E-10	***
City.NameCHICAGO	9.002	8.535	1.055	0.29177	

City.NameCOLUMBIA	22.903	8.382	2.732	0.00637	**
City.NameDALLAS	20.133	9.16	2.198	0.02813	*
City.NameDETROIT	-4.611	11.818	-0.39	0.69649	
City.NameLOS ANGELES	43.065	13.912	3.095	0.00201	**
City.NameNEW YORK	25.323	9.373	2.702	0.00698	**
City.NamePHILADELPHIA	21.391	11.262	1.899	0.05773	.
City.NameSAN FRANCISCO	59.902	11.56	5.182	2.54E-07	***
City.NameST. LOUIS	-6.844	10.16	-0.674	0.5007	

In order to gain more evidence on the third-degree discrimination over cities, we select a few origins with more data points such as Michigan. The same conclusion can be reached with the aforementioned regressions.

3.2 Second Degree Price Discrimination

When the firm lack of detailed information of each customer, firm utilizes second degree discrimination with different menu price designed to sort consumers. The result shows that the pumpkin price has second price discrimination where unit net price is different from each package. At first, we made hypothesis that there is no price difference between packages. The pumpkin weight is estimated based on different package and Item size. The unit package net price can be calculated by divided the estimated weight from the net price. However, it is found that the hypothesis fails because the price difference has significant differences between different packages. As a result, the pumpkin price model has second price discrimination on different package.

One possible reason for the second price discrimination is because the different customers prefer different packages. Retailer shop will order large quantity of pumpkin and enjoy cheaper unit price. On the other hand, personal customers prefer to pay more unit price for small packages.

Package	Estimate	Std. Error	t value	Pr(>|t|)	
Package 1 1/9 bushel cartons	0.019792	0.008354	2.369	0.01797	*
Package1 1/9 bushel crates	-0.163707	0.022653	-7.227	8.36E-13	***
Package1/2 bushel cartons	0.029899	0.011035	2.71	0.006826	**
Package20 lb cartons	0.585355	0.036414	16.075	< 2e-16	***
Package24 inch bins	0.008039	0.009275	0.867	0.386267	
Package36 inch bins	-0.060739	0.009146	-6.641	4.54E-11	***
Package50 lb cartons	0.01001	0.03386	0.296	0.767553	
Package50 lb sacks	-0.074594	0.044205	-1.687	0.09175	.
Packagebushel cartons	-0.061047	0.018411	-3.316	0.000939	***

4 Robustness Tests

4.1 Time effect

The first robustness test is carried out to study the time effect on price discrimination. Fixed effect regression is performed over the same sets of control variables, as used in the baseline model. The results of regression indicate that the aforementioned second degree and thrid degree price discrimination generally exist over the time period under study.

4.2 Varying transportation cost

In the Baseline model, assumption is made that the pumpkin transportation is only made by truck. From the research, fresh vegetables are transported through trucks and trains between the states. Ship transport is most common transport ways between countries. As a result, to improve the robustness of the transportation fee calculation. Following assumption is made.

- Distance < 1500 KM, truck is used, the transportation uniform rate is 0.37
- 1500KM<= Distance< 3000 KM, train is used, the transportation uniform rate is 0.03
- Distance >= 3000 KM, ship is used, the transportation uniform rate is 0.10

Introduced the nonlinear transportation cost, the result indicates that the price discrimination is more related to CPI with p value 0.0378.

5 Demand Elasticity Analysis

To find the pumpkin demand elasticity for each city. Following assumptions are made:

- For every date, the demanded quantity represents unit quantity for each package.
- Pumpkin supply is is inelastic to price change since pumpkins is seasonal product so that the price is mainly controlled by the harvest season.

In the month of October, the quantity of pumpkin is highest mainly related to the Halloween festival. We have data for 4 months (Sep, Oct, Nov, Dec) for years 2016 and 2017. Due to paucity of data across Oct months in 2016 and 2017, we compare change in aggregate quantity demanded with respect to change in price across Cities.

From Year 2016 to 2017, the Price Elasticity of Demand for San Francisco is 0.88 which is inelastic price elasticity of demand.

At the same time, the Price Elasticity of Demand for San Francisco is 4.856638 which is high elasticity of demand.

The results match closely with the incomes of each city. The San Francisco which has the highest per capita income have the smallest demand elasticity. On the contrast, the lowest income city - Columbia, has the highest price elasticity. The results meet our expectation that higher income generally implies relatively lesser elasticity of demand.

6 Conclusion

On our analysis on pumpkin price differentiation across U.S cities' terminal market, we observe second-degree and third-degree price discrimination. We observe menu pricing based on unit net price difference across packages. We observe group pricing where prices are different across cities based on the average income in the city. We also performed robustness tests to verify the price discrimination model and to validate our assumptions. We also performed demand-elasticity analysis and concluded that San-Francisco (with highest net income) is relatively inelastic to changing pumpkin prices compared to Columbia (with lowest net income).

Reference

Torian, R. (2017). *Cost Per Ton Mile for Four Shipping Modes*. *Richardtorian.blogspot.sg*. Retrieved 6 November 2017, from http://richardtorian.blogspot.sg/2012/01/cost-per-ton-mile-for-four-shipping.html

Ahmad, M., & Siddiqui, M. (2017). Mode of Transportation of Fresh Produce. Retrieved 6 November 2017, from https://link.springer.com/chapter/10.1007/978-3-319-21197-8_12

DSC5101 Group Project 3 Estimating Price Discrimination of Pumpkin

by Dong Gang, Swetha Narayanan, Wang Shenghao, Wang Shuai Max

Load Data

The dataset used in this project is stored in **agg_pumpkin_price.csv**. Import the data and remove the invalid entries.

```
raw.data <- read.csv("agg_pumpkin_price.csv")
rate_truck <- 0.37
rate_rail <- 0.03
rate_air <- 4.63
rate_water <- 0.10
lb_to_ton <- 0.0005
km_to_ml <- 0.621371
pumpkin.data <- raw.data[raw.data$Variety!="" & raw.data$Item.Size!="" &
raw.data$Origin!="" & raw.data$Package!='each' & raw.data$Package!='bins',]
pumpkin.data <- pumpkin.data[pumpkin.data$Origin!='CANADA' &
pumpkin.data$Origin!='MEXICO' & pumpkin.data$Origin!='COSTA RICA',]
pumpkin.data$Variety <- as.factor(pumpkin.data$Variety)
pumpkin.data$Item.Size <- as.factor(pumpkin.data$Item.Size)
pumpkin.data$City.Name <- as.factor(pumpkin.data$City.Name)
pumpkin.data$Origin <- as.factor(pumpkin.data$Origin)
pumpkin.data$avg_price <- (pumpkin.data$Low.Price + pumpkin.data$High.Price)
/ 2
pumpkin.data$wt_in_ton <- pumpkin.data$Weight * lb_to_ton
pumpkin.data$dist_in_ml <- pumpkin.data$Distance * km_to_ml
pumpkin.data$shipping_cost <- pumpkin.data$dist_in_ml *
pumpkin.data$wt_in_ton * rate_truck
pumpkin.data$net_price <- pumpkin.data$avg_price - pumpkin.data$shipping_cost
pumpkin.data <- pumpkin.data[pumpkin.data$net_price > 0,]
```

Derive Baseline Model with Linear Regression

The following assumptions are made to derive the baseline model. 1. The time effect on pumpkin price is negligible. 2. Shipment of pumpkins is via truck, with a uniform rate of 0.37 US$/(ton.ml).

Hypothesis: Price is same across all the cities

```
price.base.mod <- lm(net_price ~ Variety + Item.Size, data = pumpkin.data)
summary(price.base.mod)

##
## Call:
```

```
## lm(formula = net_price ~ Variety + Item.Size, data = pumpkin.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -149.224  -31.326   -2.923   30.326  313.717
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      161.099     26.038   6.187 8.18e-10 ***
## VarietyBLUE TYPE                  49.692     16.591   2.995  0.00280 **
## VarietyCINDERELLA                 23.150     11.910   1.944  0.05214 .
## VarietyFAIRYTALE                  27.004     11.130   2.426  0.01540 *
## VarietyHOWDEN TYPE               -30.945      9.609  -3.220  0.00131 **
## VarietyHOWDEN WHITE TYPE          -9.746     12.571  -0.775  0.43833
## VarietyKNUCKLE HEAD               30.855     16.298   1.893  0.05855 .
## VarietyMINIATURE                -115.744     11.348 -10.199  < 2e-16 ***
## VarietyMIXED HEIRLOOM VARIETIES   10.526     14.632   0.719  0.47205
## VarietyPIE TYPE                  -29.056     10.183  -2.853  0.00439 **
## Item.Sizejbo                     -20.847     24.920  -0.837  0.40299
## Item.Sizelge                     -21.870     24.551  -0.891  0.37319
## Item.Sizemed                     -34.714     24.585  -1.412  0.15818
## Item.Sizemed-lge                  18.052     25.013   0.722  0.47060
## Item.Sizesml                     -25.974     24.790  -1.048  0.29495
## Item.Sizexlge                     14.204     24.699   0.575  0.56534
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 59.28 on 1307 degrees of freedom
## Multiple R-squared:   0.38,  Adjusted R-squared:  0.3729
## F-statistic:  53.4 on 15 and 1307 DF,  p-value: < 2.2e-16
```

```
price.base.mod.res <- lm(resid(price.base.mod) ~ City.Name, data =
pumpkin.data)
summary(price.base.mod.res)
```

```
##
## Call:
## lm(formula = resid(price.base.mod) ~ City.Name, data = pumpkin.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -174.56  -30.81    3.59   29.47  277.78
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             -23.969      7.555  -3.173  0.00155 **
## City.NameBALTIMORE        9.937      9.185   1.082  0.27949
## City.NameBOSTON          51.840      8.259   6.277 4.68e-10 ***
## City.NameCHICAGO          9.002      8.535   1.055  0.29177
## City.NameCOLUMBIA        22.903      8.382   2.732  0.00637 **
```

```
## City.NameDALLAS          20.133      9.160   2.198  0.02813 *
## City.NameDETROIT         -4.611     11.818  -0.390  0.69649
## City.NameLOS ANGELES     43.065     13.912   3.095  0.00201 **
## City.NameNEW YORK        25.323      9.373   2.702  0.00698 **
## City.NamePHILADELPHIA    21.391     11.262   1.899  0.05773 .
## City.NameSAN FRANCISCO   59.902     11.560   5.182 2.54e-07 ***
## City.NameST. LOUIS       -6.844     10.160  -0.674  0.50070
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 56.03 on 1311 degrees of freedom
## Multiple R-squared:  0.1039, Adjusted R-squared:  0.09641
## F-statistic: 13.82 on 11 and 1311 DF,  p-value: < 2.2e-16
```

However, it can be observed that the residuals of the net price difference have significant relations with cities. The initial hypothesis that there is no price difference between different cities fails.

Effect of cities incomes and CPI on price discrimination.

```r
income.data <- read.csv("Income.csv")
income.data <- income.data[income.data$City.Name != "MIAMI",]
income.data$coef <- as.numeric(coef(price.base.mod.res))
income.model <- lm(coef ~ Income, data = income.data)
summary(income.model)
```

```
##
## Call:
## lm(formula = coef ~ Income, data = income.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -30.322  -9.806   1.470  10.490  37.966
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.036e+02  5.466e+01  -1.895   0.0873 .
## Income       2.555e-03  1.132e-03   2.257   0.0476 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.99 on 10 degrees of freedom
## Multiple R-squared:  0.3374, Adjusted R-squared:  0.2711
## F-statistic: 5.092 on 1 and 10 DF,  p-value: 0.04764
```

```r
income.cpimodel <- lm(coef ~ CPI, data = income.data)
summary(income.cpimodel)
```

```
##
## Call:
## lm(formula = coef ~ CPI, data = income.data)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -39.210 -10.620  -1.258  17.381  27.130
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -91.854     51.247  -1.792   0.1033
## CPI            0.475      0.218   2.179   0.0543 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.23 on 10 degrees of freedom
## Multiple R-squared:  0.3219, Adjusted R-squared:  0.2541
## F-statistic: 4.748 on 1 and 10 DF,  p-value: 0.05435
```

Comparative Study on same origin (Michigan)

```r
michigan.data <- pumpkin.data[pumpkin.data$Origin=='MICHIGAN',]
michigan.lm.mod1 <- lm(net_price ~ City.Name + Weight + Variety + Item.Size,
data = michigan.data)
summary(michigan.lm.mod1)
```

```
##
## Call:
## lm(formula = net_price ~ City.Name + Weight + Variety + Item.Size,
##     data = michigan.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -51.637  -6.342   0.957   8.788  48.134
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    32.970367  17.986430   1.833 0.068325 .
## City.NameBOSTON               -19.363104   4.274902  -4.529 1.03e-05 ***
## City.NameCHICAGO              -16.955441   4.785366  -3.543 0.000495 ***
## City.NameCOLUMBIA             -32.339994   5.256560  -6.152 4.28e-09 ***
## City.NameDALLAS               -11.973396   7.871504  -1.521 0.129861
## City.NameDETROIT                8.896444   4.800059   1.853 0.065343 .
## City.NameNEW YORK             -22.985946   6.765000  -3.398 0.000824 ***
## City.NamePHILADELPHIA         -15.936436  16.325439  -0.976 0.330195
## City.NameST. LOUIS            -46.667065   6.087762  -7.666 8.33e-13 ***
## Weight                          0.196562   0.007028  27.969  < 2e-16 ***
## VarietyHOWDEN TYPE            -52.859102  16.340408  -3.235 0.001431 **
## VarietyHOWDEN WHITE TYPE      -68.818582  20.153518  -3.415 0.000777 ***
## VarietyMINIATURE               17.325982  17.430630   0.994 0.321463
## VarietyMIXED HEIRLOOM VARIETIES  -5.941704  17.140336  -0.347 0.729230
## VarietyPIE TYPE                -8.347130  16.293368  -0.512 0.609022
## Item.Sizelge                  -18.358926   5.850151  -3.138 0.001965 **
```

```
## Item.Sizemed                          -5.060321   6.550376  -0.773 0.440744    
## Item.Sizemed-lge                      16.604810   5.938490   2.796 0.005692 ** 
## Item.Sizesml                         -15.216172   5.793944  -2.626 0.009323 ** 
## Item.Sizexlge                         -3.106896   5.331494  -0.583 0.560742    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 15.45 on 194 degrees of freedom
## Multiple R-squared:  0.9066, Adjusted R-squared:  0.8975 
## F-statistic: 99.12 on 19 and 194 DF,  p-value: < 2.2e-16
```

Second Degree Price Discrimination over pumpkin packages Hypothesis: all packages share the same unit price

```
#find the Unit Price
pumpkin.data$unitprice <- pumpkin.data$net_price / pumpkin.data$Weight
price.lm.mod3 <- lm(unitprice ~ Variety + City.Name, data = pumpkin.data)
summary(price.lm.mod3)
```

```
## 
## Call:
## lm(formula = unitprice ~ Variety + City.Name, data = pumpkin.data)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -0.44259 -0.06006  0.00005  0.04525  0.69078 
## 
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                     0.15478    0.02217   6.982 4.63e-12 ***
## VarietyBLUE TYPE                0.22761    0.02934   7.758 1.73e-14 ***
## VarietyCINDERELLA               0.16109    0.02136   7.540 8.78e-14 ***
## VarietyFAIRYTALE                0.18313    0.01993   9.187  < 2e-16 ***
## VarietyHOWDEN TYPE             -0.01744    0.01753  -0.995 0.320073    
## VarietyHOWDEN WHITE TYPE        0.06898    0.02303   2.995 0.002794 ** 
## VarietyKNUCKLE HEAD             0.16752    0.02883   5.811 7.80e-09 ***
## VarietyMINIATURE                0.55511    0.01856  29.909  < 2e-16 ***
## VarietyMIXED HEIRLOOM VARIETIES  0.09740    0.02644   3.684 0.000239 ***
## VarietyPIE TYPE                 0.10485    0.01752   5.984 2.81e-09 ***
## City.NameBALTIMORE              0.07128    0.01771   4.025 6.02e-05 ***
## City.NameBOSTON                 0.11338    0.01568   7.232 8.11e-13 ***
## City.NameCHICAGO                0.06181    0.01628   3.796 0.000154 ***
## City.NameCOLUMBIA               0.01995    0.01588   1.256 0.209213    
## City.NameDALLAS                 0.05892    0.01750   3.367 0.000781 ***
## City.NameDETROIT                0.01753    0.02267   0.773 0.439554    
## City.NameLOS ANGELES            0.06665    0.02607   2.557 0.010681 *  
## City.NameNEW YORK               0.02739    0.01766   1.551 0.121067    
## City.NamePHILADELPHIA           0.02419    0.02099   1.152 0.249409    
## City.NameSAN FRANCISCO          0.07746    0.02163   3.581 0.000355 ***
## City.NameST. LOUIS             -0.02506    0.01914  -1.309 0.190752    
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1044 on 1302 degrees of freedom
## Multiple R-squared:  0.7606, Adjusted R-squared:  0.7569
## F-statistic: 206.8 on 20 and 1302 DF,  p-value: < 2.2e-16
```

```
price.lm.mod3.res <- lm(resid(price.lm.mod3) ~ Package, data = pumpkin.data)
summary(price.lm.mod3.res)
```

```
##
## Call:
## lm(formula = resid(price.lm.mod3) ~ Package, data = pumpkin.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.40164 -0.05146 -0.00382  0.05769  0.66295
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               0.019792   0.008354   2.369 0.017970 *
## Package1 1/9 bushel crates -0.163707   0.022653  -7.227 8.36e-13 ***
## Package1/2 bushel cartons   0.029899   0.011035   2.710 0.006826 **
## Package20 lb cartons        0.585355   0.036414  16.075  < 2e-16 ***
## Package24 inch bins         0.008039   0.009275   0.867 0.386267
## Package36 inch bins        -0.060739   0.009146  -6.641 4.54e-11 ***
## Package50 lb cartons        0.010010   0.033860   0.296 0.767553
## Package50 lb sacks         -0.074594   0.044205  -1.687 0.091750 .
## Packagebushel cartons      -0.061047   0.018411  -3.316 0.000939 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08682 on 1314 degrees of freedom
## Multiple R-squared:  0.3021, Adjusted R-squared:  0.2979
## F-statistic:  71.1 on 8 and 1314 DF,  p-value: < 2.2e-16
```

The results indicate that the unit price of pumpkins is significantly different across different Packages. The hypothsis is rejected.

Robustness Tests

1. Robustness test with respect to variation of shippping cost

```
pumpkin.data$Date <- as.Date(pumpkin.data$Date, format = "%m/%d/%Y")
pumpkin.data$shipping_cost2 <- pumpkin.data$dist_in_ml *
pumpkin.data$wt_in_ton * pumpkin.data$Rate
pumpkin.data$net_price2 <- pumpkin.data$avg_price -
pumpkin.data$shipping_cost

price2.lm.mod2 <- lm(net_price2 ~ Variety + Item.Size, data = pumpkin.data)
summary(price2.lm.mod2)
```

```
##
## Call:
## lm(formula = net_price2 ~ Variety + Item.Size, data = pumpkin.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -149.224  -31.326   -2.923   30.326  313.717
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                        161.099     26.038   6.187 8.18e-10 ***
## VarietyBLUE TYPE                    49.692     16.591   2.995  0.00280 **
## VarietyCINDERELLA                   23.150     11.910   1.944  0.05214 .
## VarietyFAIRYTALE                    27.004     11.130   2.426  0.01540 *
## VarietyHOWDEN TYPE                 -30.945      9.609  -3.220  0.00131 **
## VarietyHOWDEN WHITE TYPE            -9.746     12.571  -0.775  0.43833
## VarietyKNUCKLE HEAD                 30.855     16.298   1.893  0.05855 .
## VarietyMINIATURE                  -115.744     11.348 -10.199  < 2e-16 ***
## VarietyMIXED HEIRLOOM VARIETIES     10.526     14.632   0.719  0.47205
## VarietyPIE TYPE                    -29.056     10.183  -2.853  0.00439 **
## Item.Sizejbo                       -20.847     24.920  -0.837  0.40299
## Item.Sizelge                       -21.870     24.551  -0.891  0.37319
## Item.Sizemed                       -34.714     24.585  -1.412  0.15818
## Item.Sizemed-lge                    18.052     25.013   0.722  0.47060
## Item.Sizesml                       -25.974     24.790  -1.048  0.29495
## Item.Sizexlge                       14.204     24.699   0.575  0.56534
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 59.28 on 1307 degrees of freedom
## Multiple R-squared:   0.38,  Adjusted R-squared:  0.3729
## F-statistic:  53.4 on 15 and 1307 DF,  p-value: < 2.2e-16
```

```
price2.lm.mod2.res <- lm(resid(price2.lm.mod2) ~ City.Name, data =
pumpkin.data)
summary(price2.lm.mod2.res)
```

```
##
## Call:
## lm(formula = resid(price2.lm.mod2) ~ City.Name, data = pumpkin.data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -174.56  -30.81    3.59   29.47  277.78
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -23.969      7.555  -3.173  0.00155 **
## City.NameBALTIMORE         9.937      9.185   1.082  0.27949
## City.NameBOSTON           51.840      8.259   6.277 4.68e-10 ***
```

```
## City.NameCHICAGO          9.002      8.535   1.055  0.29177    
## City.NameCOLUMBIA        22.903      8.382   2.732  0.00637 ** 
## City.NameDALLAS          20.133      9.160   2.198  0.02813 *  
## City.NameDETROIT         -4.611     11.818  -0.390  0.69649    
## City.NameLOS ANGELES     43.065     13.912   3.095  0.00201 ** 
## City.NameNEW YORK        25.323      9.373   2.702  0.00698 ** 
## City.NamePHILADELPHIA    21.391     11.262   1.899  0.05773 .  
## City.NameSAN FRANCISCO   59.902     11.560   5.182 2.54e-07 ***
## City.NameST. LOUIS       -6.844     10.160  -0.674  0.50070    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 56.03 on 1311 degrees of freedom
## Multiple R-squared:  0.1039, Adjusted R-squared:  0.09641 
## F-statistic: 13.82 on 11 and 1311 DF,  p-value: < 2.2e-16
```

```r
income.data$coef2 <- as.numeric(coef(price2.lm.mod2.res))
income.model2 <- lm(coef2 ~ Income, data = income.data)
summary(income.model2)
```

```
## 
## Call:
## lm(formula = coef2 ~ Income, data = income.data)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -30.322  -9.806   1.470  10.490  37.966 
## 
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -1.036e+02  5.466e+01  -1.895   0.0873 .
## Income       2.555e-03  1.132e-03   2.257   0.0476 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 20.99 on 10 degrees of freedom
## Multiple R-squared:  0.3374, Adjusted R-squared:  0.2711 
## F-statistic: 5.092 on 1 and 10 DF,  p-value: 0.04764
```

```r
income.cpimodel2 <- lm(coef2 ~ CPI, data = income.data)
summary(income.cpimodel2)
```

```
## 
## Call:
## lm(formula = coef2 ~ CPI, data = income.data)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -39.210 -10.620  -1.258  17.381  27.130 
## 
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -91.854     51.247  -1.792   0.1033
## CPI            0.475      0.218   2.179   0.0543 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.23 on 10 degrees of freedom
## Multiple R-squared:  0.3219, Adjusted R-squared:  0.2541
## F-statistic: 4.748 on 1 and 10 DF,  p-value: 0.05435
```

```
pumpkin.data$unitprice <- pumpkin.data$net_price2 / pumpkin.data$Weight
price2.lm.mod3 <- lm(unitprice ~ Variety + City.Name, data = pumpkin.data)
summary(price2.lm.mod3)
```

```
##
## Call:
## lm(formula = unitprice ~ Variety + City.Name, data = pumpkin.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.44259 -0.06006  0.00005  0.04525  0.69078
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     0.15478    0.02217   6.982 4.63e-12 ***
## VarietyBLUE TYPE                0.22761    0.02934   7.758 1.73e-14 ***
## VarietyCINDERELLA               0.16109    0.02136   7.540 8.78e-14 ***
## VarietyFAIRYTALE                0.18313    0.01993   9.187  < 2e-16 ***
## VarietyHOWDEN TYPE             -0.01744    0.01753  -0.995 0.320073
## VarietyHOWDEN WHITE TYPE        0.06898    0.02303   2.995 0.002794 **
## VarietyKNUCKLE HEAD             0.16752    0.02883   5.811 7.80e-09 ***
## VarietyMINIATURE                0.55511    0.01856  29.909  < 2e-16 ***
## VarietyMIXED HEIRLOOM VARIETIES  0.09740    0.02644   3.684 0.000239 ***
## VarietyPIE TYPE                 0.10485    0.01752   5.984 2.81e-09 ***
## City.NameBALTIMORE              0.07128    0.01771   4.025 6.02e-05 ***
## City.NameBOSTON                 0.11338    0.01568   7.232 8.11e-13 ***
## City.NameCHICAGO                0.06181    0.01628   3.796 0.000154 ***
## City.NameCOLUMBIA               0.01995    0.01588   1.256 0.209213
## City.NameDALLAS                 0.05892    0.01750   3.367 0.000781 ***
## City.NameDETROIT                0.01753    0.02267   0.773 0.439554
## City.NameLOS ANGELES            0.06665    0.02607   2.557 0.010681 *
## City.NameNEW YORK               0.02739    0.01766   1.551 0.121067
## City.NamePHILADELPHIA           0.02419    0.02099   1.152 0.249409
## City.NameSAN FRANCISCO          0.07746    0.02163   3.581 0.000355 ***
## City.NameST. LOUIS             -0.02506    0.01914  -1.309 0.190752
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1044 on 1302 degrees of freedom
```

```
## Multiple R-squared:  0.7606, Adjusted R-squared:  0.7569
## F-statistic: 206.8 on 20 and 1302 DF,  p-value: < 2.2e-16
```

```
price2.lm.mod3.res <- lm(resid(price2.lm.mod3) ~ Package, data =
pumpkin.data)
summary(price2.lm.mod3.res)
```

```
##
## Call:
## lm(formula = resid(price2.lm.mod3) ~ Package, data = pumpkin.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.40164 -0.05146 -0.00382  0.05769  0.66295
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 0.019792   0.008354   2.369 0.017970 *
## Package1 1/9 bushel crates -0.163707   0.022653  -7.227 8.36e-13 ***
## Package1/2 bushel cartons   0.029899   0.011035   2.710 0.006826 **
## Package20 lb cartons        0.585355   0.036414  16.075  < 2e-16 ***
## Package24 inch bins         0.008039   0.009275   0.867 0.386267
## Package36 inch bins        -0.060739   0.009146  -6.641 4.54e-11 ***
## Package50 lb cartons        0.010010   0.033860   0.296 0.767553
## Package50 lb sacks         -0.074594   0.044205  -1.687 0.091750 .
## Packagebushel cartons      -0.061047   0.018411  -3.316 0.000939 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08682 on 1314 degrees of freedom
## Multiple R-squared:  0.3021, Adjusted R-squared:  0.2979
## F-statistic:  71.1 on 8 and 1314 DF,  p-value: < 2.2e-16
```

2. Robustness test with respect to time effect

```
library(plm)
```

```
## Loading required package: Formula
```

```
## Warning: package 'Formula' was built under R version 3.4.1
```

```
library(Formula)
pumpkin.data$Date <- as.Date(pumpkin.data$Date, format = "%m/%d/%Y")
price.lm.mod2 <- plm(net_price ~ Variety + Item.Size, data = pumpkin.data,
index = c("Date"), model = "within")
```

```
## These series are NA and have been removed: Grade, Environment, Quality,
Condition, Appearance, Storage, Crop, Trans.Mode
## These series are constants and have been removed: Commodity.Name, Q,
Repack
```

```
summary(price.lm.mod2)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = net_price ~ Variety + Item.Size, data = pumpkin.data,
##     model = "within", index = c("Date"))
##
## Unbalanced Panel: n=18, T=4-158, N=1323
##
## Residuals :
##       Min.    1st Qu.     Median    3rd Qu.       Max.
## -157.81269  -32.52252   -0.68812   29.38474  316.03748
##
## Coefficients :
##                                    Estimate Std. Error t-value  Pr(>|t|)
## VarietyBLUE TYPE                    53.9944    16.6123  3.2503  0.001183 **
## VarietyCINDERELLA                   30.0774    11.9602  2.5148  0.012031 *
## VarietyFAIRYTALE                    35.5493    11.2680  3.1549  0.001643 **
## VarietyHOWDEN TYPE                 -27.7224     9.6669 -2.8678  0.004201 **
## VarietyHOWDEN WHITE TYPE            -4.9975    12.5565 -0.3980  0.690693
## VarietyKNUCKLE HEAD                 35.6620    16.3369  2.1829  0.029222 *
## VarietyMINIATURE                  -109.7898    11.4522 -9.5868 < 2.2e-16 ***
## VarietyMIXED HEIRLOOM VARIETIES     15.7334    14.6793  1.0718  0.284006
## VarietyPIE TYPE                    -21.9685    10.2530 -2.1426  0.032329 *
## Item.Sizejbo                       -23.0635    24.7232 -0.9329  0.351063
## Item.Sizelge                       -23.0581    24.3800 -0.9458  0.344438
## Item.Sizemed                       -34.5970    24.3936 -1.4183  0.156350
## Item.Sizemed-lge                    15.5853    24.8375  0.6275  0.530448
## Item.Sizesml                       -27.3339    24.5978 -1.1112  0.266675
## Item.Sizexlge                       12.3382    24.5138  0.5033  0.614828
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    7221700
## Residual Sum of Squares: 4454600
## R-Squared:      0.38317
## Adj. R-Squared: 0.36787
## F-statistic: 53.4224 on 15 and 1290 DF, p-value: < 2.22e-16
```

```
price.lm.mod2.res <- plm(resid(price.lm.mod2) ~ City.Name, data =
pumpkin.data, index = c("Date"), model = "within")
```

```
## These series are NA and have been removed: Grade, Environment, Quality,
Condition, Appearance, Storage, Crop, Trans.Mode
## These series are constants and have been removed: Commodity.Name, Q,
Repack
```

```
summary(price.lm.mod2.res)
```

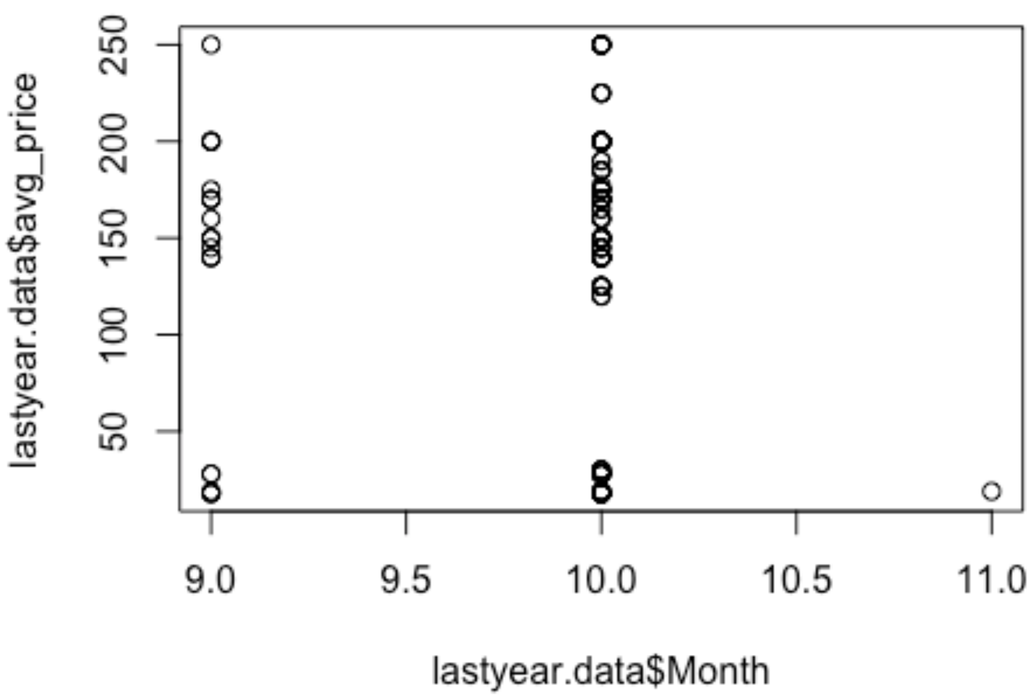
```
## Oneway (individual) effect Within Model
##
## Call:
```

```
## plm(formula = resid(price.lm.mod2) ~ City.Name, data = pumpkin.data,
##     model = "within", index = c("Date"))
##
## Unbalanced Panel: n=18, T=4-158, N=1323
##
## Residuals :
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -187.3014  -30.9868    3.6196   30.6002  280.5111
##
## Coefficients :
##                          Estimate Std. Error t-value  Pr(>|t|)
## City.NameBALTIMORE         6.3211     9.1750  0.6889   0.49098
## City.NameBOSTON           49.1575     8.2177  5.9819 2.850e-09 ***
## City.NameCHICAGO          10.5036     8.6088  1.2201   0.22265
## City.NameCOLUMBIA         17.8365     8.4213  2.1180   0.03436 *
## City.NameDALLAS           15.8529     9.1190  1.7385   0.08237 .
## City.NameDETROIT         -10.7943    11.8416 -0.9116   0.36217
## City.NameLOS ANGELES      35.5077    13.9194  2.5509   0.01086 *
## City.NameNEW YORK         19.2056     9.4153  2.0398   0.04157 *
## City.NamePHILADELPHIA     15.7834    11.3394  1.3919   0.16419
## City.NameSAN FRANCISCO    55.8213    11.5816  4.8198 1.607e-06 ***
## City.NameST. LOUIS       -11.6150    10.1739 -1.1416   0.25381
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    4454600
## Residual Sum of Squares: 3992400
## R-Squared:      0.10376
## Adj. R-Squared: 0.084363
## F-statistic: 13.6185 on 11 and 1294 DF, p-value: < 2.22e-16
```

Demand Elasticity Analysis

```
#Demand Elasticity
nrow(pumpkin.data)

## [1] 1323

lastyear.data <- pumpkin.data[which(pumpkin.data$Year == "2016" &
pumpkin.data$City.Name == "SAN FRANCISCO" ) ,]
thisyear.data <- pumpkin.data[which(pumpkin.data$Year == "2017" &
pumpkin.data$City.Name == "SAN FRANCISCO" ) ,]
lastyear.data <- pumpkin.data[which(pumpkin.data$Year == "2016" &
pumpkin.data$City.Name == "COLUMBIA" ) ,]
thisyear.data <- pumpkin.data[which(pumpkin.data$Year == "2017" &
pumpkin.data$City.Name == "COLUMBIA" ) ,]


Q1 <- sum(lastyear.data$Q)
Price1 <- sum(lastyear.data$avg_price)/nrow(lastyear.data)
Q2 <- sum(thisyear.data$Q)
```

```
Price2 <- sum(thisyear.data$avg_price)/nrow(thisyear.data)
e1 = ((Q2 - Q1)/Q1)/((Price2 - Price1)/Price1)
e1

## [1] -4.856638

#SF inelastic - 0.88
#Columbia - Elastic - 4.856638
#Boston - Elastic - 3.827553
#New york - 3.392982
#LA - Inelastic - 0.4743102

plot(lastyear.data$Month,lastyear.data$avg_price)
```




```
unique(lastyear.data$Month)

## [1]  9 10 11

lastyear.september.data <- lastyear.data[lastyear.data$Month == "9",]
lastyear.october.data <- lastyear.data [lastyear.data$Month  == "10",]
lastyear.november.data <- lastyear.data [lastyear.data$Month  == "11",]
lastyear.december.data <- lastyear.data [lastyear.data$Month  == "12",]
```

***[THE END]**