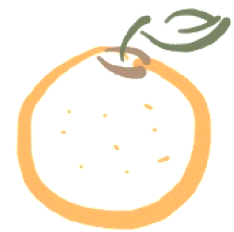

BUNK 742: Advanced Marketing Analytics

# Project #1: Pricing Decisions for Tropicana Orange Juice

Group #1
Shiyun Shen, Yingzheng Xue, Michael Rees, Zihan Cai, Jana Haas, Ziying Yan

October 8th, 2019

**Honor Pledge:**
**We pledge on our honor that we have not given or received any unauthorized assistance on this assignment**.

# Executive Summary

We were given this assignment to improve profits for Tropicana orange juice for Nick's grocery by finding the optimal price of a 64 oz bottle of orange juice. To find this, we looked at sales data from January 2009 to December 2010. We created a semi-log model examining how sales were affected by price, psychological pricing, deals, quarters of the year, each week, and individual differences between each store. This gave us the optimal price of $3.22, but because of psychological pricing, after further calculation, we determined the optimal price to be $3.19. By using this price, we could increase estimated profits for week 105 by 33.1% compared to if we used the price of week 104 and the exact profit is $174.

# Introduction and Background

Nick's is a regional grocery retailer with 15 stores with influence in the Midwest market and targets consumers with cheap but quality goods. The main competitors of Nick's include Kroger, Giant, Walmart, and Wegmans. Their competitors often make weekly promotions to lure consumers to purchase cheaper goods, luring away their consumers due to large price differences.

Since losing customers has been such a challenge within the last few years, Nick's is looking to not only retain their customers, but gain new ones - specifically in the upcoming summer. To overcome this challenge, Nick's CEO decided that it was no longer in the store's best interest to rely on each individual store's manager to assign prices as it appears to be one of the main causes for the decrease in sales and customers. As a result, Nick's reached out to our company to standardize the price of specific products in specific categories. The category that the team decided would be the most beneficial product to analyze is the 64 oz bottles of Tropicana orange juice. This is the chosen product because it is the best selling item in the refrigerated section. To resolve this scenario, we are planning to analyze the effects of price and other relevant variables on sales, and to determine the optimal retail price for Tropicana orange juice 64 oz. based on economic phenomena and consumer behaviors, in order to make a well researched and data-backed recommendation.

# Data and Methodology

The data we were given was gathered from January 2009 to December 2010. The data contained variables such as the store that sold the product, the quantity of the product sold (in ounces), the price of the product when it was sold, the sales for each week, and if the product had any ongoing deals at the time it was sold. In addition to including these variables, we created a few variables of our own to be used as dummy variables and to simplify the conclusions made from our results. To simplify our results from the data, we created a variable "bottle," which is quantity/64 because there were 64 ounces in each bottle. This allowed us to measure our findings in terms of each total bottles sold instead of each ounce. To simplify the analysis, we also took the log of the quantity, the price, and the number of bottles. We named these variables lq, lp, and lb, respectively. We also created the variable end9 to find out if the psychological pricing of ending the price of the product with a nine made a significant difference. We also split the years up into quarters and used each quarter as a dummy variable to investigate the potential effect of seasons on sales. Moreover, we added week as a continuous variable in our models because we are going to find optimal price for week 105, so using week as a unit will make results more precise.We accounted for the effects of different stores having differing sales by using stores as a dummy variable. These differences could exist due to differing market segments for each store or different locations.

To find the optimal model, we used SAS and tested many variations of linear, log-log, and semi-log models. These models varied based upon the number of dummy variables included such as psychological pricing and deal. Based on the p-value and adjusted R-squared , the models including variables price and week, and dummy variables like deal, end9, qrt1-qrt3, and store are the best among linear models (adjusted $R^2$=0.1950), semi-log models (adjusted $R^2$ =0.4740) and log-log models (adjusted $R^2$=0.4610). By comparing these models, we decided to use a semi-log model with price, week, and the dummy variables deal, psychological pricing, quarter, and store. Its adjusted-R Squared is the largest out of all of our models (adjusted $R^2$=0.4740) and every variable is significant with a P-value smaller than 0.05. Compared to price elasticity in the linear model and log-log model, the semi-log model reflects the elevation in price. The full results of our semi-log model can be found in figure one. According to our analysis, people tend to be more price-sensitive towards orange juice, meaning that any increase in price could hurt the sales of the product.

Since we chose the semi-log model to optimize price, the price will be calculated in a formula that includes cost=$2.57 and the coefficient of price. However, after we found the optimal price is 3.22, we decided to then test the prices of 3.19 and 3.29 to see if those might be better due to psychological pricing. Then after comparing the profit from different prices (profit with price $3.22 in week105 is $144, profit with $3.29 in week105 is $136), we found the optimal price is $3.19, generating an estimated profit of $174 in week 105.

# Key Findings

From figure two, we can see that sales generally decrease when the price is higher, which matches the price elasticity of the product. The graph shows that there were more sales when the price was between $3 and $3.5, with one particularly notable data point where the quantity was significantly higher than at any other points. Furthermore, the third figure shows that the sales vary significantly between quarters. There were more sales in the first quarter than other quarters, and the other three quarters displayed similar levels of sales. In addition, when we take into account if there is a running deal in figure four, we can see that sales rise significantly when there is a deal. By combining the above two variables quarter and deal, we created figure five to show that across quarters one, two, and three, the effect of having a deal is strong, but it is slightly weaker in quarter four, possibly implying that less deals should be run in quarter four because it would impact sales less than in other quarters. Figure six shows the number of sales for each store, and it shows that store 137 is performing extremely well, but stores 2 and 95 may be struggling and may need more assistance.

Figure eleven describes that the average price in these two years is $3.60, the maximum price is $4.18, and the minimum price is $2.74. It also states that the amount of variation in price is $0.35. Figure ten shows that the average bottles sold are 639 each week; however, the amount of variation of bottle sales is 769, which means that sales greatly fluctuated overtime.

In figure one, through our chosen semi-log model, we can see that the sales of Tropicana Orange Juice 64 oz. will decrease as the retail price increases. The data shows that if price increased by $1, you would only get 21% of the sales you would have gotten had the price not increased. Moreover, sales of Tropicana Orange Juice appear to increase gradually over time. On average, each week sales of Tropicana orange juice increased by .35%. If there is a deal, sales will increase by 8.41%. When the price ends with 9, the sales of Tropicana Orange Juice 64 oz. will increase by 20.62%.

Each store generated different sales, with some performing significantly better than others, as seen in figure 6. Store 137 performed better than the other stores, but we do not have sufficient information such as location to determine why this store is performing so much better. Looking at figure three, the first quarter has the largest sales compared with the other three quarters, but we do not have sufficient information to determine why that is the case. Figure eight shows that there is a good fit using the model and figure nine shows that the model yields a mostly normal distribution. The initial optimal price we had was $3.22, but when we took into account the effect of psychological pricing, $3.19 was revealed to be the more optimal price.

# Conclusions and Recommendations

Based on our extended analysis we have come up with a number of results to help us make a recommendation for an optimal price along with some to further develop Nick's.

Primarily, we decided to use the semi-log model due to a large adjusted r-squared and a small p-value. In the data we were given, sales were not consistent and fluctuated greatly over time. The optimal price for Nick's is $3.19 per 64oz. bottle of Tropicana Orange Juice. This maximizes the number of sales and profits across all 15 stores. Some stores did not do as well as others, for example, store 137 did extremely well in comparison to all other stores like 2 and 95 (see figure 6). Due to this great gap, we were led to our greatest limitation. If there was more information about the locations of the stores, we could have taken that into consideration when determining the optimal price. To maximize the accuracy of the research there should have been a couple of changes made. The locations for each individual store should be added to increase the certainty of our findings and recommendations and a variable such as population density should be added to understand why sales might lower or higher. The location of products on shelves might be different in different stores, which will also affect people's willing to purchase.

Another potential limitation involves the pricing of Tropicana Orange Juice. We do not know if some of the prices given were during a price discount or promotion at that store. Because we do not have this information, it's effect is beyond our measurement.

The first recommendation we made to Nick's was to increase the amount of deals offered for Tropicana Orange Juice. When there was a deal running, sales increased by 8.1%, so this could be used to consistently increase sales. Based on our findings we have developed the conclusion of the optimal price for Nick's grocery store; therefore, we have developed a series of recommendations to optimize sales. Sales will increase if there are more in-store displays, or featured advertising. The second recommendation was to expand their own marketing research and analyzing their stores at a more independent level. Many of the stores varied in how successful they were; therefore, Nick's should investigate and question targeted groups to find out why certain stores such as 137 are so successful and should open future stores with conditions such as location or management style that are similar to store 137. They can also use their findings to individualize the methods used at each store to maximize sales.

# Appendix:

## Figure1:

The REG Procedure
Model: MODEL1
Dependent Variable: lb

| | |
|---|---|
| Number of Observations Read | 1560 |
| Number of Observations Used | 1560 |

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 21 | 549.51784 | 26.16752 | 67.91 | <.0001 |
| Error | 1538 | 592.65683 | 0.38534 | | |
| Corrected Total | 1559 | 1142.17467 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 0.62076 | R-Square | 0.4811 |
| Dependent Mean | 6.07902 | Adj R-Sq | 0.4740 |
| Coeff Var | 10.21152 | | |

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | 11.76259 | 0.20390 | 57.69 | <.0001 |
| price | 1 | -1.53699 | 0.04952 | -31.04 | <.0001 |
| deal | 1 | 0.08074 | 0.03750 | 2.15 | 0.0315 |
| end9 | 1 | 0.18746 | 0.03882 | 4.83 | <.0001 |
| qrt1 | 1 | 0.21152 | 0.05037 | 4.20 | <.0001 |
| qrt2 | 1 | 0.17588 | 0.04756 | 3.70 | 0.0002 |
| qrt3 | 1 | 0.14630 | 0.04534 | 3.23 | 0.0013 |
| week | 1 | 0.00347 | 0.00063152 | 5.50 | <.0001 |
| store1 | 1 | -0.84167 | 0.08609 | -9.78 | <.0001 |
| store2 | 1 | -0.60828 | 0.08626 | -7.05 | <.0001 |
| store3 | 1 | -0.32561 | 0.08616 | -3.78 | 0.0002 |
| store4 | 1 | -0.43938 | 0.08620 | -5.10 | <.0001 |
| store5 | 1 | -0.54540 | 0.08615 | -6.33 | <.0001 |
| store6 | 1 | -0.63595 | 0.08615 | -7.38 | <.0001 |
| store7 | 1 | -0.68151 | 0.08609 | -7.92 | <.0001 |
| store8 | 1 | -0.55560 | 0.08616 | -6.45 | <.0001 |
| store9 | 1 | -0.54498 | 0.08625 | -6.32 | <.0001 |
| store10 | 1 | -0.90890 | 0.08618 | -10.55 | <.0001 |
| store11 | 1 | -0.61230 | 0.08620 | -7.10 | <.0001 |
| store12 | 1 | -0.48004 | 0.08611 | -5.57 | <.0001 |
| store13 | 1 | -0.66697 | 0.08613 | -7.74 | <.0001 |
| store14 | 1 | -0.62804 | 0.08611 | -7.29 | <.0001 |

Figure 2:

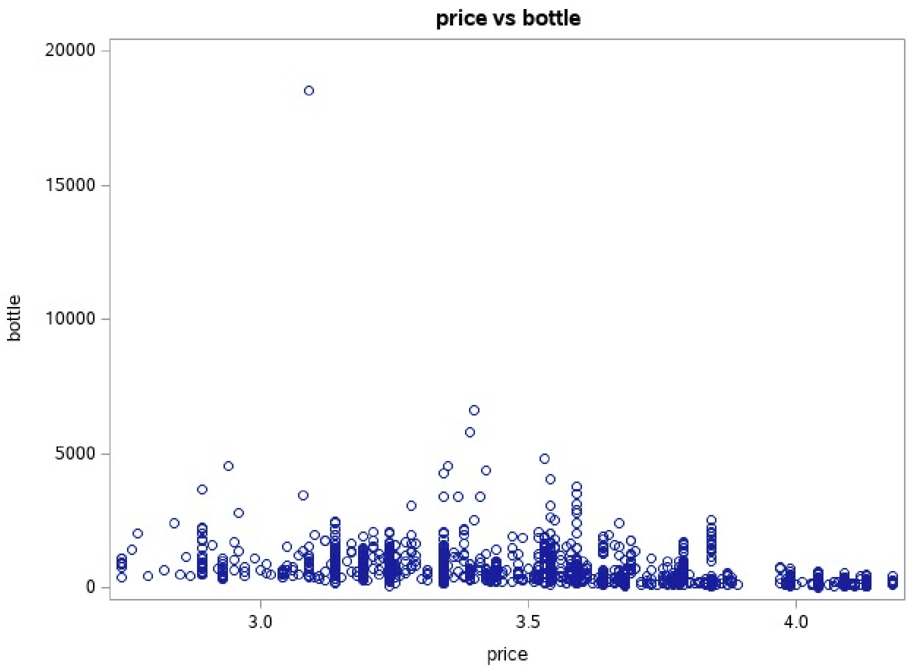


Figure 3:

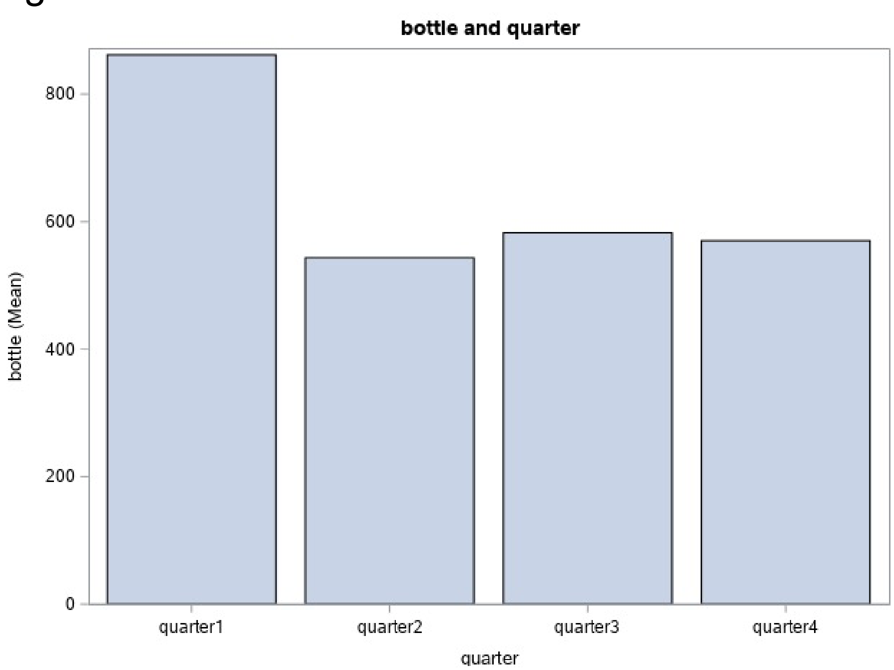

Figure 4:

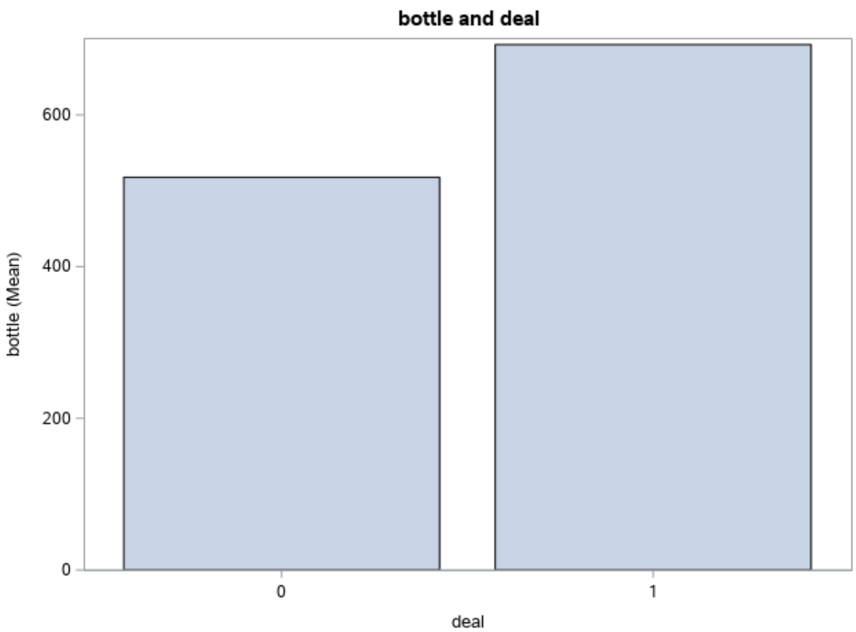


Figure 5:

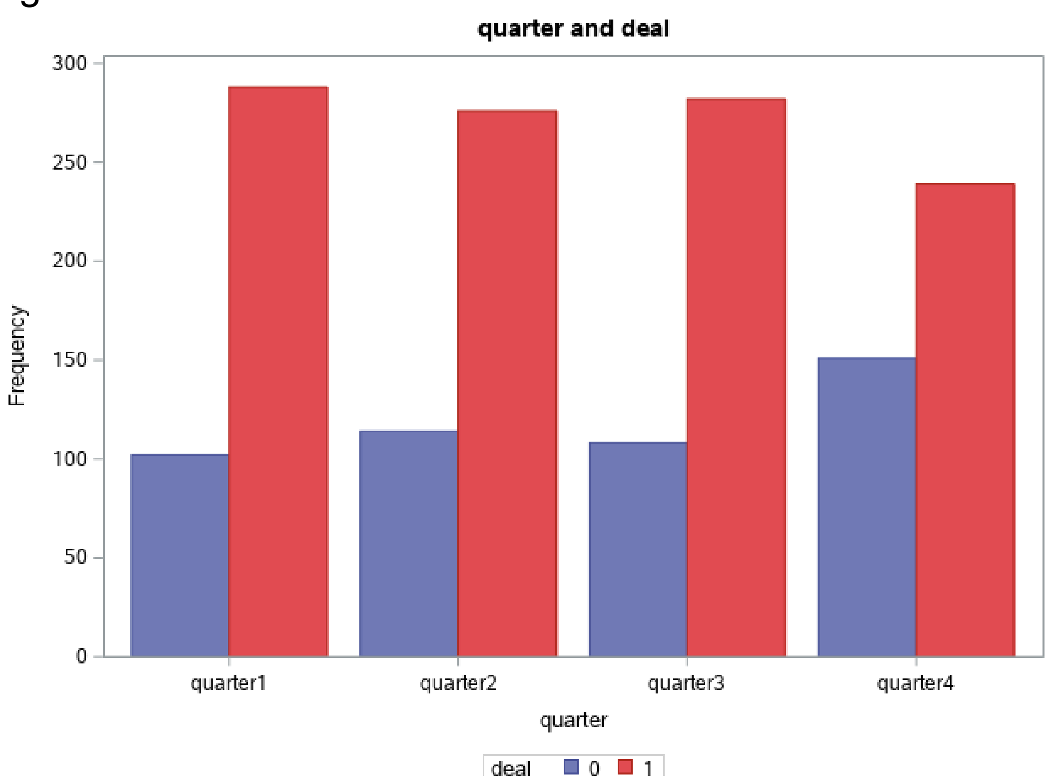

Figure 6:

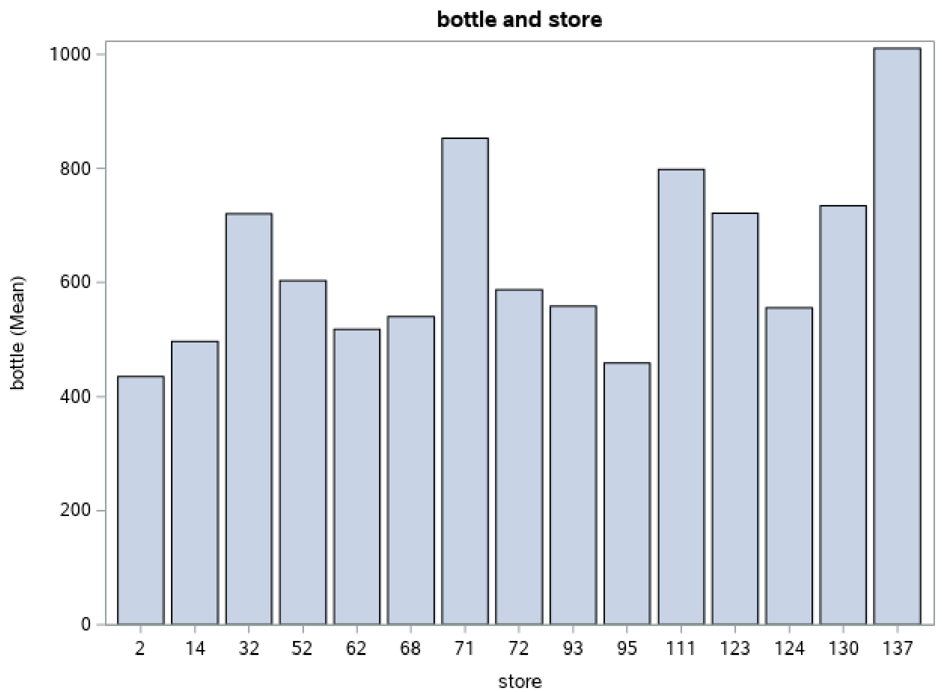


Figure 7:

The ANOVA Procedure

Dependent Variable: bottle

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 14 | 38260653.6 | 2732903.8 | 4.78 | <.0001 |
| Error | 1545 | 883676501.6 | 571958.9 | | |
| Corrected Total | 1559 | 921937155.1 | | | |

| R-Square | Coeff Var | Root MSE | bottle Mean |
|---|---|---|---|
| 0.041500 | 118.3000 | 756.2796 | 639.2897 |

| Source | DF | Anova SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| store | 14 | 38260653.59 | 2732903.83 | 4.78 | <.0001 |

[NOTE] In this ANOVA test, because the p-value is smaller than 0.05, we can conclude that different stores make great differences in sales.

Figure 8:

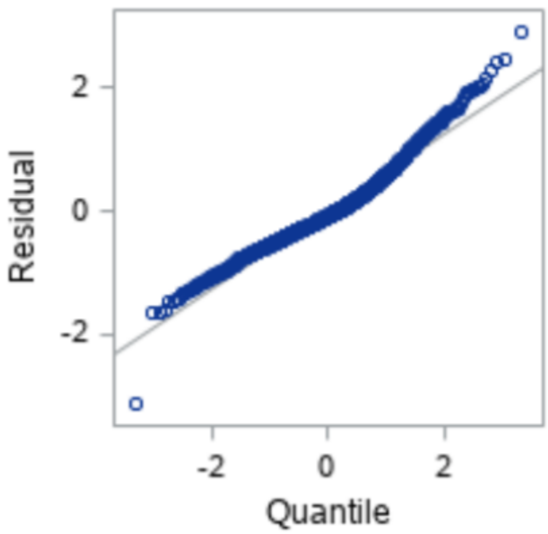


Figure 9:

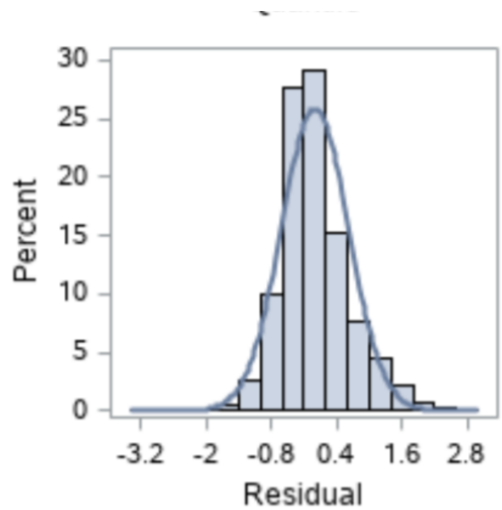


Figure 10:

The UNIVARIATE Procedure
Variable: bottle

| Moments | | | |
|---|---|---|---|
| N | 1560 | Sum Weights | 1560 |
| Mean | 639.289694 | Sum Observations | 997291.922 |
| Std Deviation | 769.002234 | Variance | 591364.436 |
| Skewness | 9.67240821 | Kurtosis | 193.869086 |
| Uncorrected SS | 1559495602 | Corrected SS | 921937155 |
| Coeff Variation | 120.290103 | Std Error Mean | 19.4699709 |

Figure 11:

The UNIVARIATE Procedure
Variable: price

| Moments | | | |
|---|---|---|---|
| N | 1560 | Sum Weights | 1560 |
| Mean | 3.60212179 | Sum Observations | 5619.31 |
| Std Deviation | 0.35077416 | Variance | 0.12304251 |
| Skewness | -0.1222403 | Kurtosis | -0.9012001 |
| Uncorrected SS | 20433.2623 | Corrected SS | 191.823277 |
| Coeff Variation | 9.73798726 | Std Error Mean | 0.00888107 |

| Quantiles (Definition 5) | |
|---|---|
| Level | Quantile |
| 100% Max | 4.18 |
| 99% | 4.13 |
| 95% | 4.13 |
| 90% | 4.09 |
| 75% Q3 | 3.84 |
| 50% Median | 3.59 |
| 25% Q1 | 3.34 |
| 10% | 3.14 |
| 5% | 3.04 |
| 1% | 2.87 |
| 0% Min | 2.74 |