ST625

QUANTITATIVE ANALYSIS FOR BUSINESS

Qian Yu

**PAM AND SUSAN’S: LOCATING NEW STORES**

PART 1 (20%)

1. *Find the mean, standard deviation, and range of the Sales variable (your dependent variable in regression analyses for all questions below, except #7).*

For sales variable: its mean is $12,140,028; standard deviation is $5,452,301; and range is $30,710,000.

1-1

|  |  |  |  |
| --- | --- | --- | --- |
| **Statistics** | | | |
| var00031 | | | |
| N | Valid | | 250 |
| Missing | | 0 |
| Mean | | 12140.0280 | | |
| Std. Deviation | | | 5452.30077 |
| Range | | | 30710.00 |

1. *Envisioning the data of 250 store sales as a sample of a much larger set of stores, find a 95% confidence interval for the true mean Sales per store.*

The 95% confidence interval for the true mean sales per store is from $11,460,865 to $12,819,191.

2-1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **One-Sample Test** | | | | | | |
|  | Test Value = 0 | | | | | |
| t | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference | |
| Lower | Upper |
| var00031 | 35.205 | 249 | .000 | 12140.02800 | 11460.8652 | 12819.1908 |

1. *Prior to the study, it was believed that average family size would be a significant predictor of sales. What percent of the variability in sales is associated with the variability in average family size?*

We can see the output from SPSS that R Square is 0.078, which means there is 7.8 percentage of the variability in sales is associated with the variability in average family size.

3-1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model Summary** | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1 | .280a | .078 | .074 | 5245.47462 |
| a. Predictors: (Constant), var00029 | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | 18838.269 | 1498.073 |  | 12.575 | .000 |
| var00029 | -1849.117 | 403.290 | -.280 | -4.585 | .000 |
| a. Dependent Variable: var00031 | | | | | | |

3-2

1. *What would you predict the change in Sales to be if average family size were a half- of-a-person larger (i.e., .5 higher)?*

In Q3: sales is Yc1 =18838.269-1849.117X

In Q4: sales is Yc2 =18838.269-1849.117\*(1.5X)

Yc2 - Yc1 = 924.559X

So the change is 924.559X; therefore, for the average family size increased by 1, the predict the change in sales is $924,559.

1. *What would you predict Sales to be for a particular store/trading zone that has a value of average family size of 5.0?*

The regression model is : Yc = 18838.269 – 1849.117\*X

We set average family size (var00029) =5, Yc = 18838.269 – 1849.117\*5 = 9592.684

Therefore, when average family size of 5.0, the predict Sales to be for a particular store/trading zone $9,592,684.

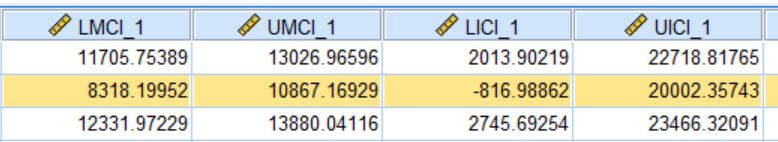
5-1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | 18838.269 | 1498.073 |  | 12.575 | .000 |
| var00029 | -1849.117 | 403.290 | -.280 | -4.585 | .000 |
| a. Dependent Variable: var00031 | | | | | | |

1. *Find a 95% confidence interval for this Sales value in question 5.*

After run the SPSS, from the data set, we can found the corresponding values that when var00029 = 5.0, which is high lined as below.

6-1



95% confidence interval: $-816,989 to $20,002,357

Refer to LICI 1 (Lower Individual Confidence Interval) and UICI 1 (Upper Individual Confidence Interval).

95% confidence interval for mean: $8,318,200 to $10,867,169

By using he columns LMCI 1 and UMCI 1

1. *The relationship between total population and average family size.*
2. The relationship between total population and average family size is significant in 0.01 level;
3. The correlation coefficient between the two variables is negative;
4. The sign of the correlation coefficient is minus, which means when total population goes up, then average family size goes down ( or average family size goes up, then total population goes down);
5. It is not make sense to me. When the average family size goes up, it is hard to say if total population will goes up or down. Just as the r is small, which mean those two variables do not have a very tired relationship.

**7-1**

|  |  |  |  |
| --- | --- | --- | --- |
| **Correlations** | | | |
|  | | var00028 | var00029 |
| var00028 | Pearson Correlation | 1 | -.211\*\* |
| Sig. (2-tailed) |  | .001 |
| N | 250 | 250 |
| var00029 | Pearson Correlation | -.211\*\* | 1 |
| Sig. (2-tailed) | .001 |  |
| N | 250 | 250 |
| \*\*. Correlation is significant at the 0.01 level (2-tailed). | | | |

PART 2 (80%)

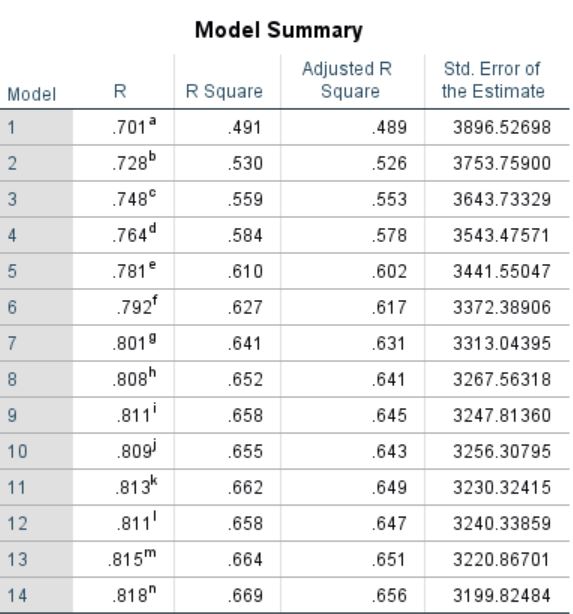
1. *Using the variables listed at the top of page 2 (i.e., using all variables except var 33 - NOT using comtype), use stepwise regression to develop the best model you can to predict store sales. How good is this model?*

First of all, we drop var00010 and var00027.

The model is good, because of R Square 0.669, which means there is

67 percentages of data can be explained by model 13.

8-1



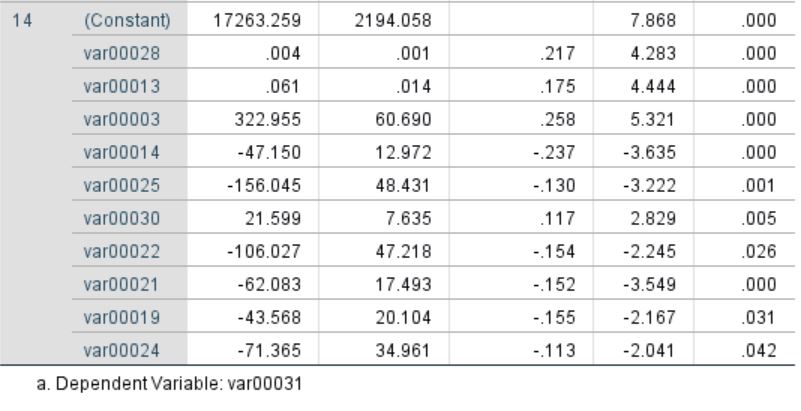
*Based on this model, how would you characterize/describe the nature of location sites that are likely to have higher sales?*

The more total population (var00028), higher median home value (var00013), more percentage of Spanish speakers (var00003), larger square feet of selling area (var00030) are likely to have higher sales.

Also, less percentage of home owners (var00014), adult is with more than 11 years of education (var00024 and var00025), households without freezer or air conditioner, dryer (var00019, var00021, and var00022) are likely to have higher sales.

The SPSS output is in next page





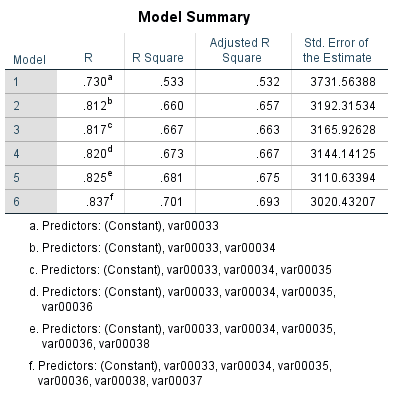
1. *A group within the planning department has developed a more subjective approach in which potential sites are classified according to an assessment of the “competitive type” of the trading zone.*

The output from SPSS is showed below: 9-1 & 9-2

*How does the model in question 8 compare to how well you can predict sales based (solely) on the “competitive type” classifications?*

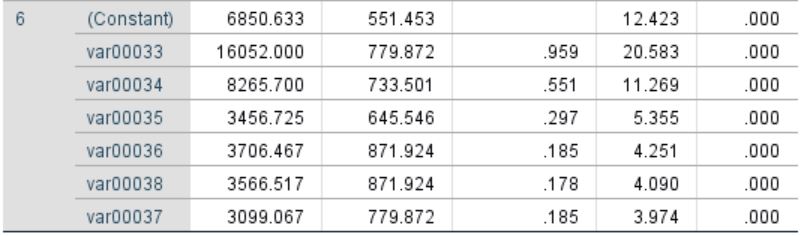
In this question, the best model has 0.701 for its R Square; in contract, the best R Square is 0.669 in question 8. So we can see that this model which solely based on competitive type classifications is slightly better than other factors to predict sales.

9-1



9-2





1. *In the model developed in question 9, interpret very precisely the coefficient of the “competitive type 4” variable. Based on the model in question 9, which is the worst competitive environment for (higher) sales?*

In question 9, the coefficient of the “competitive type 4” (var00036) is 3706.467.

Since we dropped “competitive type 7”, so it is the base for other types. Because all other types are positive, which means compare with “competitive type 7”, they have positive influence to sales (higher). So in overall, “competitive type 7” is the worst environment for (higher) sales.

At the same time, if we do not consider “competitive type 7”, we should compare with slope is smaller, which is the var00037, it has the smallest slope: 3099.067.

1. *Using as eligible the “winners” (i.e., the variables that were in the final equation) determined in question 8, combined with the “winners” determined in question 9, perform another stepwise regression to develop a final model for predicting sales*

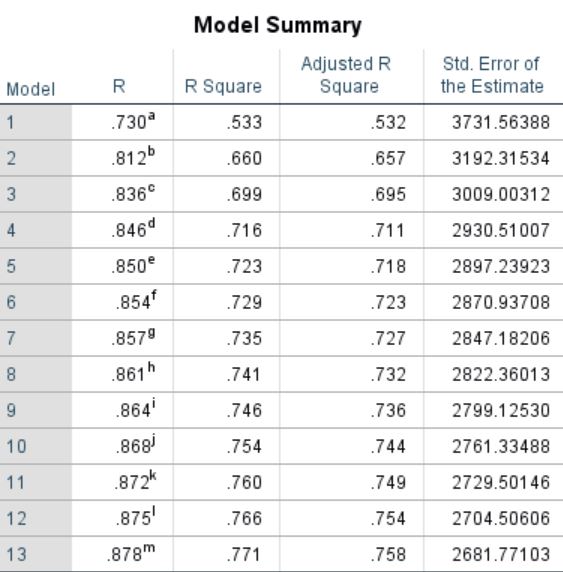
The output from SPSS is showed below: 11-1, 11-2, &11-3

*To what extent does the “competitive type” classification add to the predictive power of the specific model developed in question 8?*

In this question, the final model has 0.771 for its R Square, which is improved by 0.102 (the R Square in question 8 is 0.669.

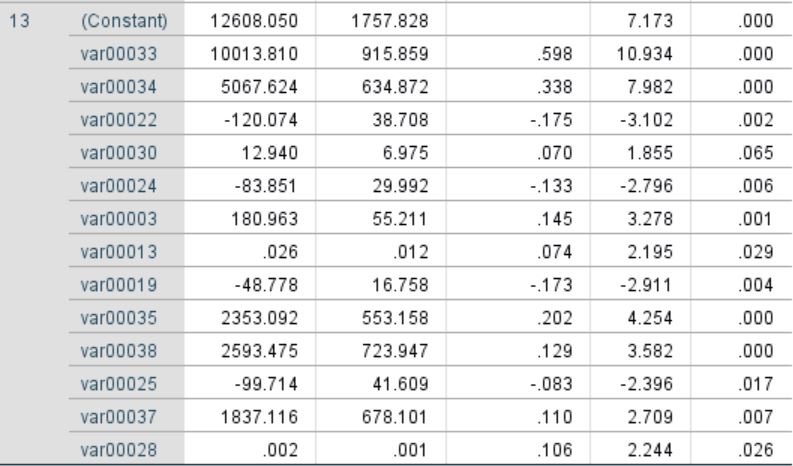
So we can say that this model which solely based on competitive type classifications is slightly better than other factors to predict sales.

11-1



11-2

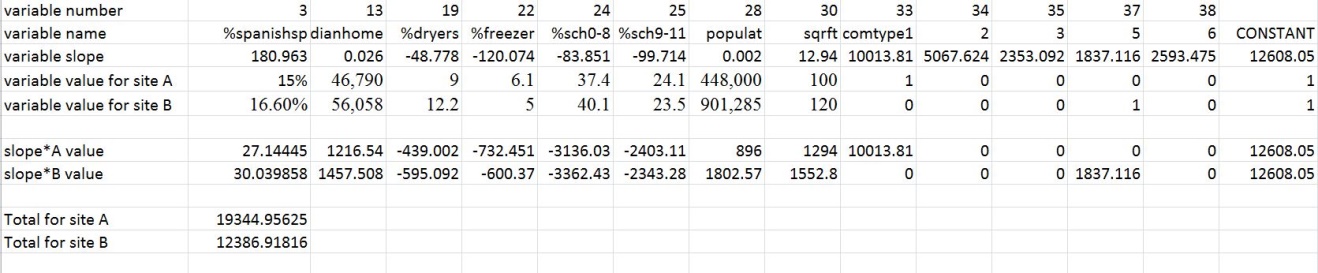




1. *Two sites, A and B, are currently under consideration for the next new store opening. Characteristics of the two sites are provided in “Attachment B.” Which site would you recommend? Justify your choice.*

Put site A and site B value into final model, the detail is at below.

Therefore, site A is recommend because the predict sale is 19,344,956.25, which is much higher than 12,386,918.16 of the sale of site B.



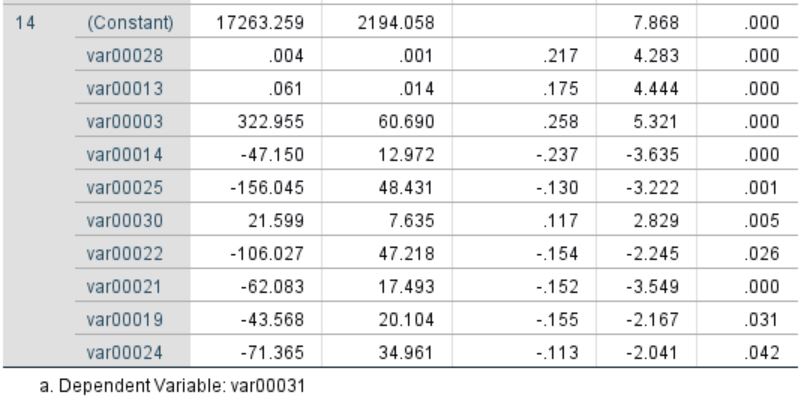
1. *Two of the variables in the database are under managerial control.*

The size of the store (square feet of selling area) is in the final model; percentage of hard goods is NOT in the final model.

*If it is, describe its relationship to sales after adjusting for the location-based variables*

Since the slope is 21.599, so when the selling area increase by 1000 square feet, the sales increase 21,599 dollars. And it is significant ( 0.005 <0.05)

13-1



*If it isn’t, is it because the variable is redundant or because it is irrelevant?*

Percentage of hard goods is not exist because it is irrelevant.

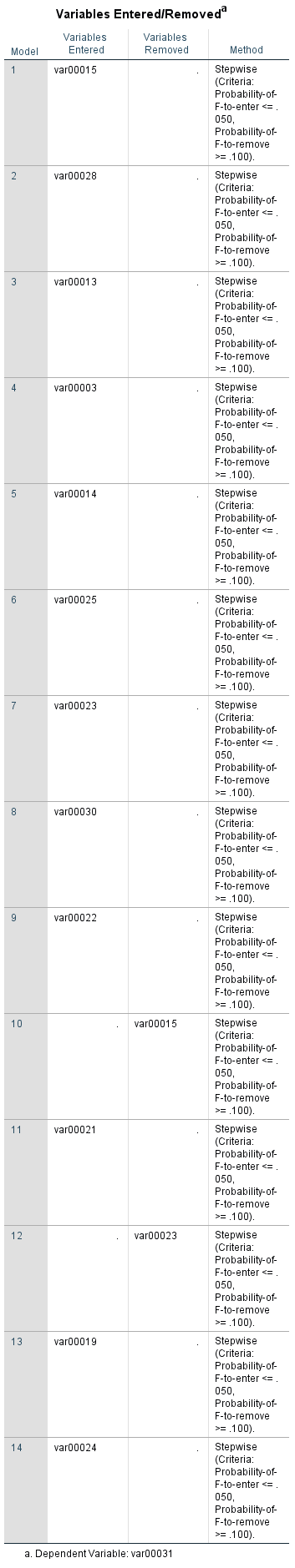
When I check the output from SPSS, var00032 has never be added in the

regression ( not like Var00015 or var00023, they were been added, then removed,

I think this situation is probably because of redundant, otherwise, it should

be considered as irrelevant).

13-2

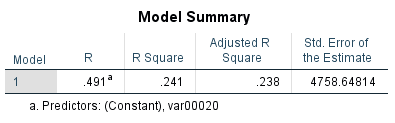


1. *Management is interested in the impact of %dishwasher on sales.*

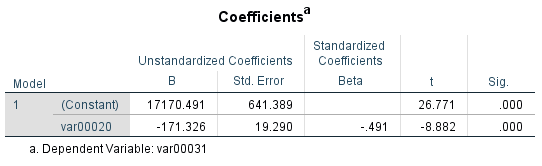
*To a lesser extent, they are also interested in the impact of %dryers on sales.*

I run the simple regression of sales vs. %dishwasher, just as Vivian said, it is negative relation. But there is higher percentage dishwasher will hurt the price of sales doesn’t make sense to me. I also notice that the R Square is just 0.241, so the dishwasher’s percentage CANNOT well explain on sales. Therefore, I think Vivian is wrong.

14-1-1



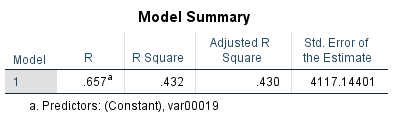
14-1-2



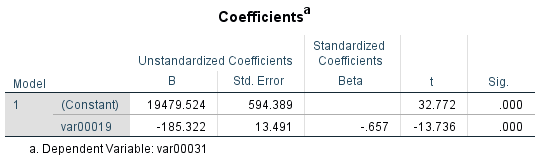
I also run both %dryers alone, and both %dryers and %dishwasher. %dryers is significant both results, and as Debbie says %dishwasher was not related to sales (not significant).

Compare with R Square, it just improves from 0.432 to 0.435, so %dishwasher is NOT important. I think Debbie conclusion is right.

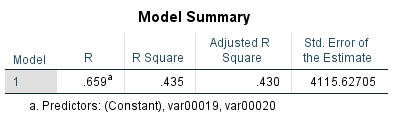
14-2-1



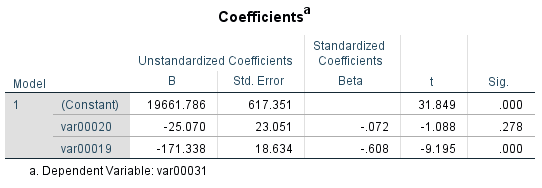
14-2-2



14-3-1

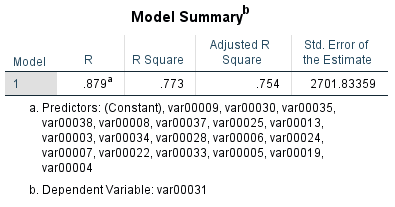


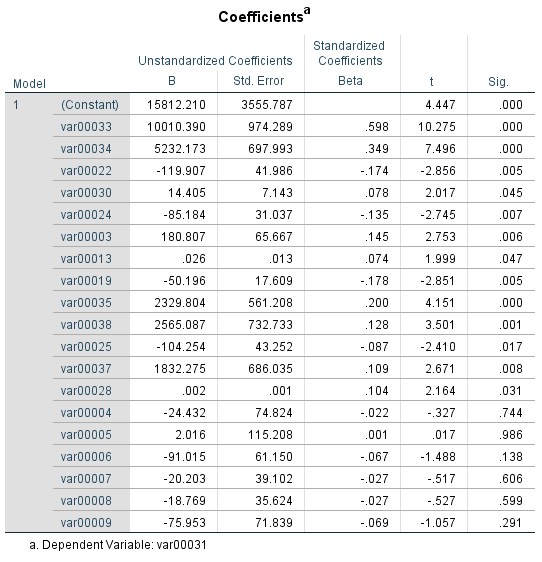
14-3-2



1. *Suppose that Management notes that none of the 7 income variables are in the final model of question 11*

The R Square 0.773, compare to the final model (0.771) , which added 0.002 more.





*If you (hypothesis) test whether the 6 added income variables as a group add to the prediction of sales, what is the p-value of the test?*

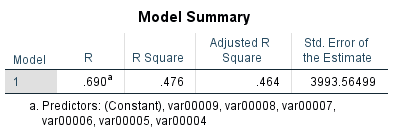
H0: b1=b2=b3=b4=b5=b6=0

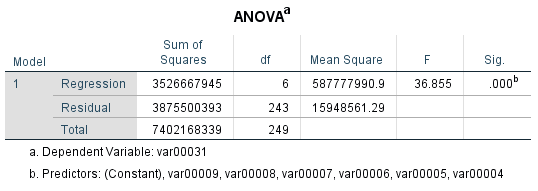
H1: at least one of b1, b2, b3, b4, b5, b6 is 0

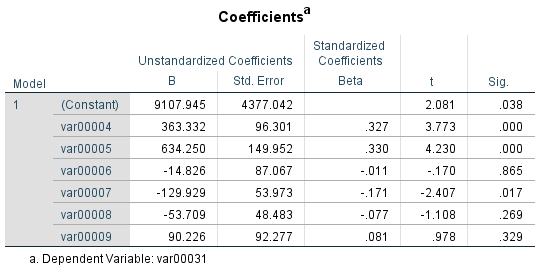
Since we see the 6 added income variables as a group, so it is a F test, F is 36.855,

its p-value is 0.000 which is smaller than 0.05. Therefore, we reject Ho, not every

beta of income is 0.







**ATTACHMENT A: COMPETITIVE TYPES**

Type 1: Densely populated areas, particularly good store sites with relatively little direct competition

Type 2: Good locations in relatively high-income areas, with little direct competition

Type 3: Locations near major shopping centers

Type 4: Stores in downtown areas of suburbs

Type 5: Stores with competition from discounters only (not from department stores)

Type 6: Stores in shopping centers

Type 7: Old stores located along the sides of major roads.

**ATTACHMENT B: PROPOSED NEW SITE LOCATIONS**

**Site A Site B**

Store size:

gross square feet 170,000 160,000

selling square feet 100,000 120,000

Competitive group 1 5

Population

black 40.0% 13.8%

Spanish speaking 15.0% 16.6%

Family income (000)

0-10 26.6% 19.2%

10+-14 14.0% 13.0%

14+-20 19.9% 22.2%

20+-30 23.9% 27.1%

30+-50 13.3% 15.7%

50+-100 2.0% 2.5%

> 100 0.3% 0.3%

Median yearly income $16,838 $18,802

Median rent per month $160 $166

Median home value $46,790 $56,058

% homeowners 10.1 10.7

% no cars 57.0 44.0

% 1 car 36.6 45.7

% TV 90.0 93.6

% washer 41.8 53.6

% dryer 9.0 12.2

% dishwasher 6.0 4.6

% air conditioner 17.9 39.3

% freezer 6.1 5.0

% second home 1.6 4.6

Education: %

0-8 37.4 40.1

9-11 24.1 23.5

12 29.0 25.5

12+ 9.5 10.9

Total population 448,000 901,285

# Average family size 3.7 3.5

Variables:

Var 2: population: % Black

Var 3: population: % Spanish speaking

Vars 4 – 10: % in each of the following family income categories (000s): 0-10; 10+-14; 14+-20; 20+-30; 30+-50; 50+-100; >100

Var 11: median yearly family income

Var 12: median rent per month

Var 13: median home value

Var 14: % home owners

Var 15: % with no cars

Var 16: % with one car

Var 17: % households with TV

Var 18: % households with washer

Var 19: % households with dryer

Var 20: % households with dishwasher

Var 21: % households with air conditioner

Var 22: % households with freezer

Var 23: % households with second home

Vars 24 - 27: % adults (over age 25) with the following years of education: 0-8; 9-11; 12, 12+

Var 28: total population

Var 29: average family size

The following data were collected on each store:

Var 30: square feet of selling area (000s)

Var 31: annual sales (000s of $)

Var 32: % hard goods

Var 33: Competitive type, a value from 1-7 as indicated in attachment A.