10-2016

Multiple Linear Regression Applications in Real Estate Pricing

Ceyhun Ozgur
Valparaiso University

Zachariah Hughes
Valparaiso University

Grace Rogers
Valparaiso University

Sufia Parveen
Valparaiso University

Follow this and additional works at: https://scholar.valpo.edu/cba_fac_pub

Recommended Citation
Ozgur, Ceyhun; Hughes, Zachariah; Rogers, Grace; and Parveen, Sufia, "Multiple Linear Regression Applications in Real Estate Pricing" (2016). Business Faculty Publications. 61.
https://scholar.valpo.edu/cba_fac_pub/61

This Article is brought to you for free and open access by the College of Business at ValpoScholar. It has been accepted for inclusion in Business Faculty Publications by an authorized administrator of ValpoScholar. For more information, please contact a ValpoScholar staff member at scholar@valpo.edu.
Multiple Linear Regression Applications in Real Estate Pricing

Ceyhun Ozgur, Ph.D., CPIM\(^1\), Zachariah Hughes\(^2\), Grace Rogers\(^3\), Sufia Parveen\(^4\)

\(^1\)Professor, Valparaiso University College of Business Information & Decision Sciences Urschel Hall 223 – Valparaiso University Valparaiso, IN 46383
\(^2\)Undergraduate Research Assistant Valparaiso University Finance – College of Business Economics – College of Arts and Sciences
\(^3\)Undergraduate Research Assistant Valparaiso University Actuarial Science - College of Arts and Sciences Business Analytics - College of Business
\(^4\)Graduate Research Assistant Valparaiso University M.S. in Analytics and Modeling

ABSTRACT: In this paper, we attempt to predict the price of a real estate individual homes sold in North West Indiana based on the individual homes sold in 2014. The data/information is collected from realtor.com. The purpose of this paper is to predict the price of individual homes sold based on multiple regression model and also utilize SAS forecasting model and software. We also determine the factors influencing housing prices and to what extent they affect the price. Independent variables such square footage, number of bathrooms, and whether there is a finished basement, and whether there is brick front or not and the type of home: Colonial, Contemporary or Tudor. How much does each type of home (Colonial, Contemporary, Tudor) add to the price of the real estate.

I. INTRODUCTION

This paper is about 480 houses selected as a representative sample of all real estate sold in the State Indiana in 2014. The information used in this paper was taken from realtor.com. The purpose of this paper is to develop a relatively good regression equation for predicting the price of these houses. It is known that there are many factors to influence housing price, but we do not know for certain what factors will influence the price of the houses and to what extent these factors will impact the price. As a buyer, one can judge whether the price of a house of interest is rational or not. As a seller, one would be able to select a rational price according to the equation helping to optimize potential sales and correctly forecast demand.

According to one of the similar papers by Anpalaki J. Ragavan(2008), independent variables were categorized as two types: a) continuous independent variables (e.g: number of bedrooms, number of bathrooms, size of the property), and b) indicator independent variables that provide supporting information about the unit in the form of an item or facility that is either present (1) or not present (0) in the unit (e.g: built in dishwasher (DW), refrigerator (fridge), laundry facilities (WD)). There were 6 continuous independent variables namely, i) number of bedrooms (BED), ii) number of bathrooms (BATH), iii) square footage or size (SIZE), iv) sale price (PRICE), v) age of the property (AGE), and v) size of the yard (LOT).

According to Leslie A. Christensen, a linear model has the form \( Y = b_0 + b_1X + \varepsilon \). The constant \( b_0 \) is called the intercept and the coefficient \( b_1 \) is the parameter estimate for the variable \( X \). The \( \varepsilon \) is the error term. \( \varepsilon \) is the residual that cannot be explained by the variables in the model. Most of the assumptions and diagnostics of linear regression focus on the assumptions of \( \varepsilon \). The following assumptions must hold when building a linear regression model.

1. The dependent variable must be continuous. If we are trying to predict a categorical variable, linear regression is not the correct method. We can investigate discrim, logistic, or some other categorical procedure.
2. The data we are modeling meets the "iid" criterion. That means the error terms, \( \varepsilon \), are: a) independent from one another and b) identically distributed. If assumption 2a does not hold, we need to investigate time series or some other type of method. If assumption 2b does not hold, we need to investigate methods that do not assume normality such as non-parametric procedures.
3. The error term is normally distributed with a mean of zero and a standard deviation of \( \sigma \), \( N(0,\sigma^2) \). Although not an actual assumption of linear regression, it is good practice to ensure the data we are modeling came from a random sample or some other sampling frame that will be valid for the conclusions we wish to make based on our model.

Following the above procedures in this paper, the response variable selected is the “price” of these houses and the price is a numeric variable. At the same time, there are 12 variables used as the potential predictors and they
are: Size (in square feet), Region (Urban, Suburban or Rural) Type (Condominium, Townhouse or SFH), Yard (in square feet), Bedrooms (number of), Bathrooms (number of), Garage (Attached, Detached or No), Floors, (number of), HOA (Homeowner Association in dollars), Tax (in percentage), Basement (Yes or No), Age (in years). The variables region, type, garage and basement are specifically numeric variables and are measured by digits, while the remainder are character variables. These character variables can be changed to numeric variables for the purpose of performing regression analysis. In this study, there are 480 houses and they are observational subjects.

To assess the validity of the regression assumptions, residual plots are used. Appropriate transformations of the dependent or independent variable can be done by making changes to the model until they are improved upon from previous problems. Estimated regression equation is then reported. Any outliers or influential values are being noted. If we find any influential values, we remove them from the dataset and a new report of estimated regression equation is made without the influential values. Finally comparison of the new estimated regression equation to the original one is done. We have used the Sherwin Rosen (1974) paper on hedonic pricing and implicit markets as the foundation of this paper regarding pricing in real estate. We also used the paper by Lipsey and Rosenbluth (1971). Both of these papers, showed that the Real Estate market is competitively priced.

II. PRELIMINARY DATA EXPLORATION

For quantitative predictors, scatterplot and correlation can be used to assess; for categorical predictors, boxplots can be used to assess. At the same time, the method of correlation can be used to find if there are some high correlated potential predictors.

2.1 The Basic Data

Table 2: Tests for Location

<table>
<thead>
<tr>
<th>Test</th>
<th>Statistic</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student's t</td>
<td>t 57.06253</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Sign</td>
<td>M 240</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Signed Rank</td>
<td>S 57720</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Quantiles (Definition 5)

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% Max</td>
<td>1148962</td>
</tr>
<tr>
<td>99%</td>
<td>827428</td>
</tr>
<tr>
<td>95%</td>
<td>619739</td>
</tr>
</tbody>
</table>
From this result, we can get some basic data about the response variable (the price of these houses). Specially, Mean=344294.4 Median=320063.0 Standard deviation=132190

2.2: Correlation between variables

Table 3: Correlation between variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Sum</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>price</td>
<td>480</td>
<td>344294</td>
<td>132190</td>
<td>165261327</td>
<td>89145</td>
<td>1149862</td>
</tr>
<tr>
<td>size</td>
<td>480</td>
<td>2265</td>
<td>583.09745</td>
<td>1059861</td>
<td>950.0069</td>
<td>33860</td>
</tr>
<tr>
<td>Region_h</td>
<td>480</td>
<td>2.21975</td>
<td>0.72556</td>
<td>1115</td>
<td>1.00000</td>
<td>3.00000</td>
</tr>
<tr>
<td>Type_h</td>
<td>480</td>
<td>2.16937</td>
<td>0.66293</td>
<td>1040</td>
<td>1.00000</td>
<td>3.00000</td>
</tr>
<tr>
<td>yards</td>
<td>480</td>
<td>15.00</td>
<td>11.8225</td>
<td>1977466</td>
<td>0</td>
<td>306790</td>
</tr>
<tr>
<td>bedrooms</td>
<td>480</td>
<td>3.03958</td>
<td>0.72253</td>
<td>1747</td>
<td>1.00000</td>
<td>6.00000</td>
</tr>
<tr>
<td>bathrooms</td>
<td>480</td>
<td>2.30021</td>
<td>0.81018</td>
<td>1115</td>
<td>1.00000</td>
<td>4.00000</td>
</tr>
<tr>
<td>garage_h</td>
<td>480</td>
<td>1.77590</td>
<td>0.62649</td>
<td>85220000</td>
<td>0</td>
<td>3.00000</td>
</tr>
<tr>
<td>floors</td>
<td>480</td>
<td>1.57708</td>
<td>0.72241</td>
<td>80500000</td>
<td>0</td>
<td>3.00000</td>
</tr>
<tr>
<td>basement_h</td>
<td>480</td>
<td>1.53125</td>
<td>0.49554</td>
<td>72500000</td>
<td>0</td>
<td>2.00000</td>
</tr>
<tr>
<td>age</td>
<td>480</td>
<td>37.38195</td>
<td>30.11915</td>
<td>17943</td>
<td>0</td>
<td>162.0000</td>
</tr>
<tr>
<td>hoa</td>
<td>480</td>
<td>1639</td>
<td>938.49569</td>
<td>786598</td>
<td>0</td>
<td>5400</td>
</tr>
</tbody>
</table>

This procedure gives information about the 12 variables with their simple statistics and Pearson’s correlation coefficient.
2.3: Scattered plots
Graph 1: Scatter plot between Price and hoa

Graph 2: Scatter plot between Price and size:

Graph 3: Scatter plot between Price and yards:
Multiple Linear Regression Applications in Real Estate Pricing

Graph 4: Scatter plot between Price and age:

Graph 5: Scatter plot between Price and floors:

Graph 6: Scatter plot between Price and bedrooms:
By observing the correlation between the variables and the scatterplots the following conclusion was drawn. The correlation coefficient between price and HOA is 0.16628 and the scatterplot shows they have positive relationship. From the second scatterplot, we can find that the price and size have positive relationship; the price increases as the size increases. In detail, the correlation coefficient between price and size is 0.04523. From the third scatterplot it has been observed that price and yards have positive relationship just like it has with size. The correlation coefficient between price and yards is 0.028, indicating that the correlation is significant. The correlation coefficient between price and age is -0.00313, indicating that the correlation between price and age is not significant and the scatterplot between them shows a negative relationship, that is, as the age of the house increases the price obviously decreases.

The correlation coefficient between price and floors is 0.04369 and the scatterplot shows there is a significant positive correlation between them. The correlation coefficient between price and bedrooms is -0.05749 and the scatterplot shows as the number of bedrooms increases the price gradually decreases which is a negative correlation. Finally, the correlation coefficient between price and bathrooms is 0.01755 and the scatterplot indicates there is a positive correlation meaning, as the number of bathrooms increases the price increases initially but later decreases with an increase in number of bathrooms.

There also exist relationships among the other potential predictors through the table. For example the correlation coefficient between size and yards is found to be 0.58639 and the correlation coefficient between garage_h and type_h is 0.33917, these two absolute values of correlation coefficient are both greater than 0.3 and less than 0.8, so they have moderate correlation. Other absolute values of correlation coefficients, which were found to be less than 0.3, were found to have low correlation relationships.

2.4: Box plots

Graph 8: Box plot between Price and type:

From the box-plot above, it can be observed that when price range is between 20,000 to 60,000, the type of the house available is townhouse, when it is between 15,000 to 65,000 the house type is SFH and when the price range is between 10,000 to 55,000 the house type available is condominium. All the three types of houses have some outliers.
Graph 9: Box-plot between Price and region_h:

From the box-plot between price and region_h, it has been observed that when the price range is between 20,000 and 55,000 the region is rural, when the price range is between 20,000 and 53,000 the region is urban, and when the price range is between 15,000 and 55,000 the region is suburban. All three regions have some outliers.

Graph 10: Box-plot between Price and garage_h:

From the box-plot between price and garage_h it has been observed that when the price range is between 10,000 and 55,000 the garage is detached, when the price range is between 20,000 and 57,000 the garage is attached, and when the price range is between 11,000 and 60,000 there is no garage. All three garage types have outliers.

Graph 11: Box-plot between Price and basement_h:

From the box-plot between price and basement_h it has been observed that when the price range is between 10,000 and 60,000 there is a basement and when the price range is between 12,000 and 58,000 there is not a basement. Both of the basement types have some outliers.

As we can see from the figures, the various categorical predictors all have the outliers. Now as we did some primary analysis for different predictors, we cannot give up any predictors from the information above.
III. REGRESSION

3.1. Based on all the information and outputs above, our initial model would be:

\[
\text{Price} = \beta_0 + \beta_1 \text{size} + \beta_2 \text{region}_h + \beta_3 \text{type}_h + \beta_4 \text{yards} + \beta_5 \text{age} + \beta_6 \text{floors} + \beta_7 \text{hoa} + \beta_8 \text{bedrooms} + \beta_9 \text{bathrooms} + \beta_{10} \text{garage}_h + \beta_{11} \text{basement}_h + \epsilon.
\]

From the scattered plots, it is clearly seen that the linear relation is negative between price and age. The normal thinking is that the price will go down with the age goes up. That is to say, the coefficient between price and age is negative. Although the dots are dispersed and the linear relation is not obvious, we still put that in into the initial model. Secondly, the relationship between price and size is positive, as observed in the scatterplot, so we include this in the initial model. We also include all other numeric predictors into the initial model. As for the four categorical predictors, we use the side-by-side box plots to see the distribution. By observing the boxplots, some outliers were existent. But we still include them in the initial model. Initially, we include all potential predictors in the model, assuming that they may have an influence on the price. (Articles & Statistical Papers, December 1992)

3.2. Then let us begin to do the multiple regression analysis.

Graph 12: Scatter plot between predicted value and studentized residual

Table 4: Regression Analysis of Real Estate predictor variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>t</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>504928</td>
<td>47452</td>
<td>10.64</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region_h</td>
<td></td>
<td>-10228</td>
<td>8875</td>
<td>-1.16</td>
<td>0.2471</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type_h</td>
<td></td>
<td>-4499.2</td>
<td>7684</td>
<td>-0.60</td>
<td>0.5498</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basement_h</td>
<td></td>
<td>-52498</td>
<td>16027</td>
<td>-3.15</td>
<td>&lt;.0017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garage_h</td>
<td></td>
<td>9526.91148</td>
<td>9438.85735</td>
<td>1.01</td>
<td>0.3134</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td></td>
<td>-65.05399</td>
<td>27.25726</td>
<td>-2.41</td>
<td>0.0182</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td>33.05113</td>
<td>195.3357</td>
<td>0.17</td>
<td>0.8657</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years</td>
<td></td>
<td>5.68943</td>
<td>2.57136</td>
<td>2.21</td>
<td>0.0275</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoa</td>
<td></td>
<td>77.00611</td>
<td>12.40116</td>
<td>6.29</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bedrooms</td>
<td></td>
<td>-1776.7</td>
<td>8466.62355</td>
<td>-0.10</td>
<td>0.9357</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bathrooms</td>
<td></td>
<td>5130.44606</td>
<td>8471.92452</td>
<td>0.91</td>
<td>0.3650</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Floors</td>
<td></td>
<td>-11624</td>
<td>17136</td>
<td>-0.68</td>
<td>0.4975</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

www.ijmsi.org 46 | Page
3.3. The fitted quadratic model:

\[
\text{Price} = 504928 - 10228 \cdot \text{Region}_h - 4498.217 \cdot \text{Type}_h - 52498 \cdot \text{Basement}_h + 9526.911 \cdot \text{Garage}_h - 65.854 \cdot \text{Size} + 33.051 \cdot \text{Age} + 5.684 \cdot 4 \cdot \text{Yards} + 77.091 \cdot \text{Hoa} - 17727 \cdot \text{Bedrooms} + 5130.446 \cdot \text{Bathrooms} - 17727 \cdot \text{Floors}
\]

The adjusted $R^2$ is 0.0745, which indicates this model fits the data well. The $F$-test (ANOVA) tests whether \( \beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0 \). The p-Value is $< 0.0001$, indicating that overall the explanatory variables are significant to the response variable.

Parameter estimates and standard errors of the estimates, along with t-tests and p-values are also given. The t-test tests whether each explanatory variable is zero. Hence, only HOA is significant because the p-value $< 0.0001$. We can then put the HOA into this model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>305911</td>
<td>11994</td>
<td>25.50</td>
<td>$&lt;0.001$</td>
<td></td>
</tr>
<tr>
<td>hoa</td>
<td>1</td>
<td>23.42238</td>
<td>6.35320</td>
<td>3.69</td>
<td>0.003</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Analysis of Variance for dependent variable price

Graph 13: Scatter plot between predicted value and studentized residual

From the result, the p-value=0.0003, which means that HOA variables are significant in relation to the response variable. At the same time, the adjusted R-Sq=0.0256, which means that it explains 2.56% of the change in price. Finally, we get the following equation:

\[
\text{Price} = 305911 + 23.422 \cdot \text{hoa}.
\]

However, the above residual plot shows that the assumption of constant variance is not met. So we need to check for multicollinearity as well as influential observations.

3.4: Multicollinearity

Table 6: Multicollinearity
Multiple Linear Regression Applications in Real Estate Pricing

Table 7: Regression Procedure collinearity diagnostics

<table>
<thead>
<tr>
<th>Number</th>
<th>Eigenvalue</th>
<th>Condition Index</th>
<th>Proportion of Variation-- intercept hoa</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.86801</td>
<td>1.00000</td>
<td>0.06600</td>
</tr>
<tr>
<td>2</td>
<td>0.13199</td>
<td>3.76197</td>
<td>0.93400</td>
</tr>
</tbody>
</table>

Tolerance is the proportion of each variable’s variance not shared with the other explanatory variables. Small tolerance values indicate collinearity. In general, we should ensure the tolerance is greater than 0.2. If we notice the tolerance value of HOA its 1.000 which is greater than 0.2, this variable has no tolerance problem.

Then we need to find the outliers. Studentized residuals greater than 1.5 or less than -1.5 should be deleted because they are outliers. From these 480 data, #13, #24, #62, #63, #351, #371 should be deleted. In the second round, we need to delete #87, #96, #106, #108, #109, #118, #398. In the third round, we delete #135, #153, #157, #170, #211, #218, #224, #436. In the fourth round we delete #236, #242, #244, #268, #270, #284, #293, #313, #437, #217 and finally in total, we delete 26 data of 480.

After that, we need to check for multicollinearity again.

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>t</th>
<th>Tolerance</th>
<th>Variance Inflation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>31.2638</td>
<td>19.955</td>
<td>25.98</td>
<td>0.001</td>
<td></td>
<td>0.00000</td>
<td></td>
</tr>
<tr>
<td>Hoa</td>
<td>1</td>
<td>17.85426</td>
<td>6.78556</td>
<td>2.60</td>
<td>0.068</td>
<td>1.000</td>
<td>1.00000</td>
<td></td>
</tr>
</tbody>
</table>

After that, we need to check for multicollinearity again.

Table 8: Regression procedure dependent variable price

From the data, the tolerance of the HOA variable is greater than 0.2, which indicates that it has no tolerance problem.
Table 9: Regression Procedure Dependent Variable price

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>1.251753E11</td>
<td>1.251753E11</td>
<td>7.84</td>
<td>0.0053</td>
</tr>
<tr>
<td>Error</td>
<td>447</td>
<td>7.134306E12</td>
<td>15960416364</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>448</td>
<td>7.25948E12</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Variable hoa Entered: R-Square = 0.0172 and C(p) = 2.0000

Analysis of Variance

Table 10: Analysis of Variance

All variables left in the model are significant at the 0.1500 level.
All variables have been entered into the model.

The intercept for HOA is not significant because the p-value is 0.053 which is > 0.001.
From the multiple regression model we can get the equation as:

Price = 312638 + 17.854Ho.

Finally, we need to check the variance.
Multiple Linear Regression Applications in Real Estate Pricing

Graph 14: Scatter plot between predicted value and student zed residual

Through this diagram, it can meet the requirement of constant variance in general. So the best model is:

\[ \text{Price} = 312638 + 17.854 \text{Hoa} \]

IV. CONCLUSION

This paper is about what factors would have influence on the price of the houses. Firstly, we did the preliminary analysis and chose 12 variables to predict the price. Then we used the method of multiple linear regression to analyze how these factors affect the price of the houses. After the analysis, we chose seven variables (HOA, size, age, yards, floors, bedrooms, bathrooms) to include in our model. There were many outliers, and finally, we included HOA in our model. Because the easiest model is the best, we chose to use one variable to for prediction. Finally, we got the equation: \( \text{Price} = 312638 + 17.854 \text{Hoa} \)

Once we solved this paper question, we got a clear idea that HOA is one of the most useful indexes to estimate the price of house. At the same time, it has most important influence on the price. When we look forward to buying a house, we can include the information of HOA in this regression model, and construct an estimate quote for the price and compare that with the real price. Then we can decide whether to buy or not.

However, when you decide to buy a house you should consider other factors that influence the price, including the region, age, size, area, yards, number of bedrooms, floors and bathrooms etc. All of these elements will have a large or small influence on the price. If we want to get a more precise model, many other factors should be considered. However, it may consume too much time to collect the data and find the best regression model.

REFERENCES

[2]. How to use SAS® to fit Multiple Logistic Regression Models Anpalaki J. Ragavan, Department of Mathematics and Statistics, University of Nevada, Reno, NV 89557(369-2008)
[3]. Introduction to Linear Regression Analysis (By Douglas C. Montgomery, Elizabeth A. Peck, G.)
[5]. The Little SAS Book for Enterprise Guide 4.2 ( By Susan J. Slaughter, Lora D. Delwiche)
[8]. Regression with SAS (http://www.ats.ucla.edu/stat/sas/webbooks/reg/chapter1/sasreg1.htm )
[9]. ReliaSoft’s Experiment Design and Analysis Reference
[12]. Using Multivariate Statistics (FIJT H EDITION, Barbara G. Tabachnick California State University, Northridge Linda S. Fidell California State University; Northridge