

**SUPPLEMENT**

**USING NCSS 2004/2007 IN BUSI 444**

## TABLE OF CONTENTS

<b>Introduction .....</b>	<b>1</b>
<b>LESSON 1.....</b>	<b>2</b>
Page 1.12: Scatterplots .....	2
Page 1.15: Fit line for a second degree polynomial .....	2
Page 1.17: Setting a Filter .....	3
Page 1.21: Boxplots.....	3
Page 1.22: Crosstabulation.....	4
Page 1.23: Compare Means .....	4
Page 1.24: Transformations .....	4
Page 1.28: Transformations (Recode).....	5
Page 1.29: Multiple Regression.....	5
Page 1.33: Correlation Matrix .....	5
Page 1.37: Ratio Statistics .....	5
Page 1.36: Exponential Transformations .....	5
Page 1.38: Kruskal-Wallis Test .....	6
Page 1.39: Mann-Whitney Test .....	6
<b>LESSON 2.....</b>	<b>7</b>
Page 2.5: Deleting Duplicate Rows .....	7
Page 2.5: Descriptive Statistics: Frequency Tables.....	7
Page 2.5: Descriptive Statistics: Descriptive Tables.....	7
Page 2.9: Ratio Statistics .....	7
Page 2.11: Kruskal-Wallis Test .....	7
Page 2.12: Recode Transformations.....	7
Page 2.23: Recode Transformations.....	8
Page 2.26: Recode Transformations.....	8
Page 2.27: Recode Transformations.....	9
Page 2.27: Compute Transformation.....	9
<b>LESSON 3.....</b>	<b>9</b>
Page 3.6: Extract Month .....	9
Page 3.10: Multiple Regression .....	9
Page 3.13: Multiple Regression.....	9
Page 3.15: Stepwise Regression .....	9
Page 3.17: Casewise Diagnostics, Storing Regression Results and filtering out “outliers” .....	10
Page 3.34: Cubic and Quadratic Curve Estimation .....	10
<b>LESSON 4.....</b>	<b>10</b>
Page 4.5: Cubic and Quadratic Curve Estimation .....	10
<b>LESSON 5.....</b>	<b>10</b>
Page 5.9: Multiple Regression .....	10
Page 5.10: Standard Error of the Estimate.....	13
Page 5.10: Multicollinearity .....	13
Page 5.15: Transformations .....	13
Page 5.16: Separating Model and Test Databases .....	13
Page 5.17: Stepwise Regression .....	13
Page 5.19: Multiple Linear Regression .....	14
Page 5.19: Filter sales outside 3 standard deviations.....	16

<b>LESSON 6</b> .....	<b>19</b>
Page 6.5: Descriptive Statistics: .....	19
Page 6.9: Transformations .....	19
Page 6.13: Non-Linear Regression .....	19
Page 6.15: Non-Linear Regression .....	21
Page 6.21: Non-Linear Regression .....	28
<b>LESSON 7</b> .....	<b>32</b>
Page 7.9 to 7.10: Correlation Matrix .....	32
Page 7.14: Saving Predicted Values.....	32
Page 7.15: Comparing Groups .....	32



# SUPPLEMENT

## USING NCSS 2004/2007 IN BUSI 444

---

### Introduction

Several lessons of the BUSI 444 course workbook use the SPSS software program to explain and illustrate statistical analysis. This document explains equivalent instructions for students who wish to instead use the NCSS software program. The document follows the BUSI 444 course workbook, explaining equivalent NCSS commands the first time a new SPSS command is introduced. Where an NCSS technique is required and has already been introduced earlier in this supplement, the instructions are not provided. Where the NCSS output is different from the SPSS output, this supplement provides and interprets the NCSS output. This document is based on NCSS Versions 2004 and 2007.

NOTE: this Supplement assumes that students have the NCSS software installed and a working knowledge of how to operate the software program and interpret its results. If you require instruction on basic operations, you should review the document "NCSS 2004/2007 Orientation". This can be found on the BUSI 444 "Online Readings" webpage.

#### **ALERT! Proceed at your own risk!**

This supplement will attempt to explain the working of NCSS as an alternative to the SPSS Version 15.0 software illustrated in the course workbook. However, it is ultimately the student's responsibility to reconcile software differences. If you find you are unable to reconcile these differences, you are strongly recommended to instead use the SPSS Version 15.0 software.

#### **NCSS Instructions: Help Files**

The steps required for the NCSS software will be illustrated throughout this NCSS supplement. As a first step, you may wish to review the NCSS help menu. The help menu provides detailed information on all statistics generated by the software. In particular the tutorials may be extremely useful for understanding some of the output generated by the various commands.

Choose Help from the top menu and select Help from the drop-down menu. Select Index and type Tutorial. A list of all tutorials available in NCSS is shown. Select the tutorial you are interested in and all necessary instructions will be explained.

## LESSON 1

### Page 1.12: Scatterplots

The NCSS Scatterplots menu allows you to create scatterplots with a variety of features.

- **Graphics → Scatterplots**
- *SALEPRICE* as the **Vertical variable** and enter *AREAI\_2* as the **Horizontal variable**.
- Click on the **Lines 1** tab and check the **Regression Line** box. This will display the best fit regression line.
- In the **Confidence Limits** section, select the **C.L. Means** box and set the **Confidence Limit Alpha** to 0.050.
- R-square [feature not available in NCSS 2004]: in NCSS 2007, to display the R-square, click on the **Titles and Misc.** tab, and specify the following:
  - under **Top Title Line 1**:  $R\text{-Square} = \{R\}$
  - note that R-Square is displayed as a percent, not a decimal (e.g., 77.1% rather than 0.771)
- Regression Equation: to display the regression equation, click on the **Lines 1** tab [NCSS 2004] or the **Titles and Misc.** tab [NCSS 2007], and specify the following:
  - under **Top Title Line 2**:  $\{Y\} = \{A\} + (\{B\})\{X\}$
- Press the **Play** button (or go to **Run > Run Procedure**)

#### *NCSS Tip: Templates*

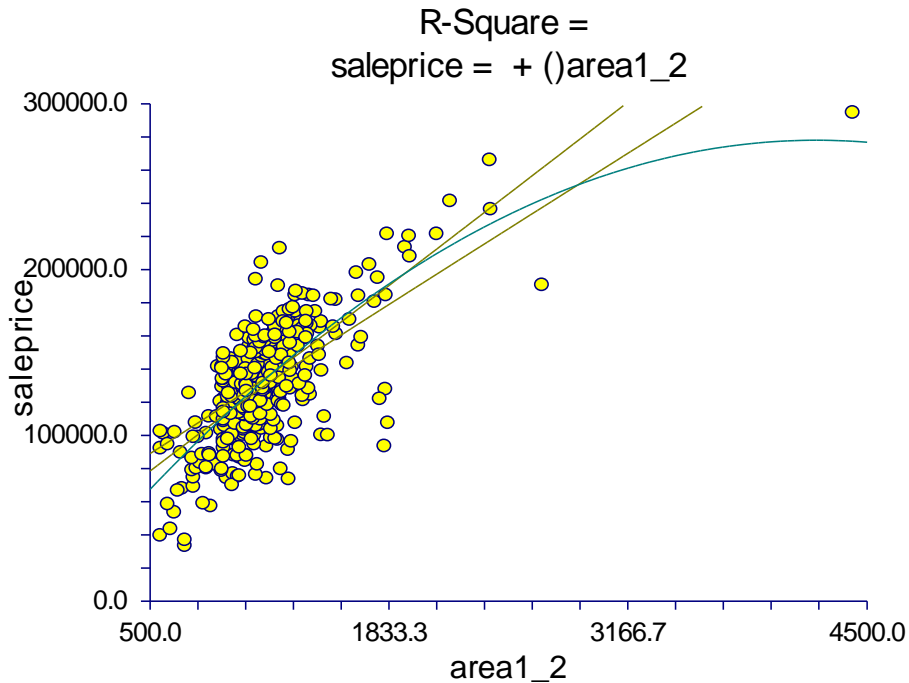
If you are carrying out repeated commands with the same settings, you may wish to save a **Template**. This will allow you to load the template later and easily re-create output or run the same test on different databases in future. To create and open a template:

- In any NCSS module, click on the **Template** tab.
- Under File Name, type in descriptive title you will remember, e.g., *Scatter\_with\_line*.
- Click **Save Template**.

To load this template in the future, click on the Template tab, select the template name you want to load, and click **Load Template**.

### Page 1.15: Fit line for a second degree polynomial

- Return to the **Scatter Plots** module. The data you entered for the previous step should still be inputted.
- In the **Lines 1** tab, uncheck the **Regression Line** box
- Check the **Polynomial Fit** box
- Press the **Play** button



NCSS doesn't display the R-Square value for this procedure. To find the R-Square value of the polynomial fit, follow these steps:

- **Analysis** → **Curve Fitting** → **One Independent Variable** → **Curve Fitting - General**
- Select *saleprice* as the **Y variable** and *area1\_2* as the **X variable**
- In the **Model box**, select Quadratic
- Press the **Play** button

R2 is shown in the Parameter Estimates for All Groups

#### Page 1.17: Setting a Filter

- **Select Data** → **Filter**.
- Under **Keep Spreadsheet Row If:**, select **If at least one statement is true (OR)**
- In the **Filter Statement** box, enter: `saleprice <= 250000`
- Check the **Filter Statement Active** checkbox, then press **Play**
- Remember to **turn off the filter** for future analyses.

#### Page 1.21: Boxplots

- **Graphics** → **Box plots**.
- Enter *saleprice* in **Variable(s)** and *carptfl* as **Grouping Variable**.
- Data Label Variable: NCSS does not have a feature that allows you to identify specific data points in graphics. However, in NCSS, if you want to see data for the plots on the graph, you can specify a **Data Label Variable**. The points on the plots will display the values for whatever variable was chosen.
- Confidence Limits: you may add notches to show the 95% confidence limit around the median; if the notches do not overlap, then you can conclude the groups are statistically different. In NCSS 2004, click the **Box Plot** tab and under **Shape**, select **Notched - Old**. In NCSS 2007, this selection is found on the **Variables** tab.
- Press **Play** button or F9.

**Page 1.22: Crosstabulation**

- **Analysis → Descriptive Statistics → Cross Tabulation.**
- Enter *CULDESAC* under **Table Columns** and *NBHD* under **Table Rows**.
- Run the report.
- The report shows the Crosstabulation, Chi-Square, and probability, but not the correlations.
- To view percentages, in the Crosstabulation window click on the **Reports** tab, and change **Row Percents**, **Column Percents** and **Table Percents** from **Omit** to **Report**.

**Page 1.23: Compare Means**

Descriptive Statistics is the closest equivalent to the SPSS Compare Means function.

- **Select Analysis → Descriptive Statistics → Descriptive Statistics.**
- Under **Data Variables**, select *MBYRBLT*
- Under **Grouping variables** select *NBHD* as **group1 variable**.
- Click on the **Reports** tab, select only Summary Section.
- Run the report.

**Page 1.24: Transformations**

In NCSS, there are two alternative methods for creating a transformation. The NCSS Help menu offers good advice on transformations.

*Transformation Method 1: Variable Info*

- Click on the **Variable Info** tab at the bottom left of the NCSS window.
- Under **Name**, for the next available variable type in name desired, e.g., *SP\_SQFT*.
- Under **Transformation**, type the formula desired, using either variable names or numbers: e.g., *C2/C7* or *SALEPRICE/TOTAREA*.
- Choose **Data** from the top menu and select **Recalc Current** or **Recalc All** to run the transformation OR click the yellow calculator icon on NCSS menu bar.
- Click on **Sheet1** at the bottom left of the NCSS window to view the newly transformed variable.

*Transformation Method 2: Data → Enter Transform*

- Select **Data → Enter Transform**. (Note: you can open the "Enter Transform" window by double clicking on the Transformation cell in the "Variable Info" window).
- Specify Result Variable: choose next available variable number. \*\*\*important, otherwise you may transform over existing variables\*\*\*.
- Under **Transformation Formula**, either type the formula in the box (e.g., *SALEPRICE/TOTAREA*) or use **Available Functions** and **Argument Variables** to create the formula (helpful for complicated formulas with many variables).
- Click OK.
- Click on **Variable Info** tab at the bottom left of the NCSS window; you will see the formula beside the result variable selected (if formula is beside the wrong variable, you forgot to specify Result Variable; delete the formula and start again!)
- Under **Name**, for the result variable type in name desired (*SP\_SQFT*).
- Choose **Data** from the top menu and select **Recalc Current** or **Recalc All** to run the transformation OR click the **yellow calculator** icon on NCSS menu bar.
- Click on **Sheet1** at bottom left of NCSS window to view the transformed variable.



**Page 1.28: Transformations (Recode)**

- First, create the new variables in **Variable Info**.
- Then transform the necessary variables using either: (a) the **Transformation** column in the **Variable Info** worksheet; (b) **Data → If-Then Transform**; or (c) **Data → Enter Transform**.
- Create the following transformations using the Transformation column in **Variable Info** or **Data → Enter Transform** (ensure you type the formulas exactly as shown):

<u>Variable Name</u>	<u>Transformation Formula</u>
<i>lincpls</i> :	RECODE(cp_mancl;(911=.437) (921=1) (931=1.332)(else=1))

Use **Data → Recalc Current** to run the transformations. Confirm they ran correctly in **Sheet1**.

**Page 1.29: Multiple Regression**

- The NCSS regression module provides all of the necessary output by default.
- **Analysis → Regression/Correlation → Multiple Regression**.
- For **Variables**, select *cpmktval* as the **Dependent Variable** and *crpt\_low* and *crpt\_hi* as the Independent Variables.
- Click on Reports tab:
- NCSS 2004: Under **Choose a Group of Reports and Plots**, select **Display only those items that are CHECKED BELOW**. Select **Run Summary, Means-Std Dev's, Correlations, Equation, Coefficient, ANOVA Summary, R-Squared, Multicollinearity, Reg Diagnostics, Histogram, and Res vs Yhat Plot**.
- NCSS 2007: Under **Select a Group of Reports and Plots**, select **Display only those items that are CHECKED BELOW**. Select **Run Summary, Descriptive Statistics, Correlations, Equation, Regression Coefficients, ANOVA Summary, R-Squared, Multicollinearity, Regression Diagnostics, Histogram, and Residuals vs Yhat Plots**.
- Click the **Play** icon to run the procedure.

**Page 1.33: Correlation Matrix**

- **Analysis → Regression / Correlation → Correlation Routines → Correlation Matrix**.
- In **Correlation Variables** select (or type) *bedrooms, baths, familyrm, fireplcs, flrarea1, flrarea2, bsmfjn, curb\_gut, conduit, alley, sidewalk, culdesac, ndhd* and *linmcdv*.
- Under **Correlation Type** select **Pearson Product Moment**
- Run the report.

**Page 1.37: Ratio Statistics**

- **Analysis → Mass Appraisal → Appraisal Ratios**.
- Specify *totalval* as the **Appraisal Variable**, *saleprice* as the **Sale Price Variable**, and *nbhd* as the **Break Variable**.
- Click the **Report** tab and select the statistics of interest (**Median, Mean, Confidence Limits for Mean and Median, PRD, COD and COV** (note that Range, Minimum, and Maximum are not included with this report)).
- In the **Report Options** section you can set the decimal places and column widths for your report
- Run the report.

**Page 1.36: Exponential Transformations**

- In the **Variable Info** tab, enter *ADJLOTSZ* in the next available row under the **Name** column.

- In the appropriate **Transformation cell**, enter  $LOT\_SIZE^{0.75}$  and hit the Enter key on your keyboard (the ^ symbol is a shift of the number 6 key...it represents exponentiation in NCSS)
- Highlight the **Transformation cell**, and select **Data → Recalc Current** from the main menu.
- View the **Sheet1** tab to make sure the transformation was successful.

### Page 1.38: Kruskal-Wallis Test

- **Analysis → Analysis of Variance (ANOVA) → One-Way Analysis of Variance.**
- Select ASR as the **Response Variable** and nbhd as the **Factor Variable**.
- Under **Reports**, select only Kruskal-Wallis Report.
- Run the report.

### Kruskal-Wallis One-Way ANOVA on Ranks

#### Hypotheses

H<sub>0</sub>: All medians are equal.

H<sub>a</sub>: At least two medians are different.

#### Test Results

Method	DF		Chi-Square (H)	Prob Level	Decision(0.05)
Not Corrected for Ties	6		67.15916	0.000000	Reject H <sub>0</sub>
Corrected for Ties	6		67.15917	0.000000	Reject H <sub>0</sub>

Number Sets of Ties	4
Multiplicity Factor	24

#### Group Detail

Group	Count	Sum of Ranks	Mean Rank	Z-Value	Median
1	78	24141.00	309.50	6.2742	0.9229962
2	48	12163.00	253.40	1.5724	0.8817468
3	42	9216.00	219.43	-0.3178	0.8577755
4	31	7884.00	254.32	1.2788	0.8833497
5	87	16090.00	184.94	-3.2388	0.8432237
6	99	16333.00	164.98	-5.2428	0.8262587
7	65	15648.00	240.74	1.0213	0.8695332

### Page 1.39: Mann-Whitney Test

- **Analysis → T-Tests → T-Test - Two Sample.**
- Select *saleprice* as the **Response Variable** and *culdesac* as the **Group Variable**.
- Run the report.

This report will provide much more detail than the results shown in the workbook. Make a note of where to find the key data within this report so that you can locate these areas in future uses of the M-W test.

## LESSON 2

### Page 2.5: Deleting Duplicate Rows

- There is no **Identify Duplicate Cases** module in NCSS. Instead, you must carry this out manually.
- **Data → Sort.** Sort by *PID*, *Sale Month*, and *Sale Day*, then click **OK** to run.
- In the **Sheet1** window, scroll through the data and look for duplicate addresses. Delete the first (or oldest) sale for any pairs found.

### Page 2.5: Descriptive Statistics: Frequency Tables

- The best NCSS alternative for producing these statistics is **Analysis → Descriptive Statistics → Frequency Tables.**
- In the **Variables** tab, enter the *SaleYear*, *Stories*, and *Bedrooms* variables in the **Discrete Variables** box. Enter the *LotSizeSqft*, *YearBuilt*, *EffectiveYear*, *FinishedArea*, and *SalePrice* variables in the **Numeric Variables (Width)** box. Note that mean, median, standard deviation, range, minimum, and maximum are not included with this report (see the following step).
- Press **Play** to run the report.

### Page 2.5: Descriptive Statistics: Descriptive Tables

- **Analysis → Descriptive Statistics → Descriptive Tables**
- In the **Variables** tab, enter the *LotSizeSqft*, *YearBuilt*, *EffectiveYear*, *FinishedArea*, and *SalePrice* variables in the **Response Variables** box.
- In the **Reports** tab, select the statistics you wish to display
- Press **Play** to run the report.

### Page 2.9: Ratio Statistics

- **Analysis → Mass Appraisal → Appraisal Ratios.**
- Specify *SalePrice* as the **Appraisal Variable**, *AssessedTotal* as the **Sale Price Variable**, and *SaleYear* as the **Break Variable**. (All sales were from 2006, so this will ensure that the ratio analysis is summarized in just one row.)
- Click the **Report** tab and select the statistics of interest (Median, Mean, Confidence Limits for Mean and Median, COD, PRD, and COV (note that Range, Minimum, and Maximum are not included with this report).
- Run the report.

### Page 2.11: Kruskal-Wallis Test

- **Analysis → Analysis of Variance (ANOVA) → One-Way Analysis of Variance.**
- Select *SAR* as the **Response Variable** and *SaleMonth* as the **Factor Variable**.
- Under the **Reports** tab, select only Kruskal-Wallis Report.
- Run the report.

### Page 2.12: Recode Transformations

In NCSS, the *SPRange* variable can be populated using **Data → If-Then Transform**. First, go to **Variable Info** and create the new variable *SPRange* in the next available row (C15). Then go to **Data → If-Then Transform** and enter the necessary transformation:

Type following into IF box:

SalePrice <= 500000  
 (SalePrice > 500000) AND (SalePrice <= 700000)  
 SalePrice > 700000

Type following into THEN box:

SPRange = 1  
 SPRange = 2  
 SPRange = 3

Run the Transformation

*Helpful Hint:* In NCSS, you may wish to save your transformations into a template file (this is the closest NCSS equivalent to the syntax files in SPSS). This allows you to easily run them again later, should you need to re-create new variables in a new database or in restoring a backup database (e.g., if you need to start over from scratch). In the **If-Then** transformation window, click on the **Template** tab. Under **File Name**, enter "Lesson2", then click **Save Template** and it will save the transformation into a template file in the **Template Directory** noted. You can then re-load this template later. When you are completing your projects, this will be a very helpful tool.

**Page 2.23: Recode Transformations**

- First, go to **Variable Info** and create the new variables *crawl*, *partbsmt*, *fullbsmt*, and *slab* in the next available rows.
- Then go to **Data → If-Then Transform** and enter the necessary transformations.
- Notes:
  - a) where a variable's value is text rather than a number, the text entries must be enclosed in double quotation marks; also, upper or lower case is important, they must be exactly the same as the database value (e.g., "CRAWL" not "crawl"); and
  - b) < > means "not-equal-to" or "less than or greater than".

Enter into IF box:

- Foundation = "CRAWL"
- Foundation < > "CRAWL"
- Foundation = "PBSMT"
- Foundation < > "PBSMT"
- Foundation = "BSMT"
- Foundation < > "BSMT"
- Foundation = "SLAB"
- Foundation < > "SLAB"

Enter into THEN box:

*crawl* = 1  
*crawl* = 0  
*partbsmt* = 1  
*partbsmt* = 0  
*fullbsmt* = 1  
*fullbsmt* = 0  
*slab* = 1  
*slab* = 0

- Confirm the transformation results appear in the **Sheet1** worksheet or using crosstabs.

**Page 2.26: Recode Transformations**

- First, go to **Variable Info** and create the new variables *story15* and *story20* in the next available rows.
- Then go to **Data → If-Then Transform** and enter the necessary transformations.

Enter into IF box:

- Stories = 1.5
- Stories < > 1.5
- Stories = 2.0
- Stories < > 2.0

Enter into THEN box:

*story15* = 1  
*story15* = 0  
*story20* = 1  
*story20* = 0

Confirm the transformation results appear in the **Sheet1** worksheet or using crosstabs.

**Page 2.27: Recode Transformations**

- First, go to **Variable Info** and create the new variable *poolyes* in the next available row.
- Then go to **Data → If-Then Transform** and enter the necessary transformations.

Enter into IF box:

- Pool="Y"
- Pool="N"

Enter into THEN box:

- poolyes=1
- poolyes=0

- Confirm the transformation results in the **Sheet1** worksheet or using crosstabs.

**Page 2.27: Compute Transformation**

- First, go to **Variable Info** and create the new variable *totalbath* in the next available row.
- Then, in the **Transformation** column, enter the following:

*FullBath+ThreeQtrBath\*.75+HalfBath\*.5*

- Confirm the transformation results appear in the **Sheet1** worksheet or using crosstabs.

**LESSON 3****Page 3.6: Extract Month**

- Enter a new variable called *Month* in the next available row in the **Variable Info** worksheet
- Enter "Month(saledate+2415019)" into the **Transformation cell** (leave out the quote marks)
- With the **Transformation cell** still highlighted, click **Data → Recalc Current**

**Page 3.10: Multiple Regression**

- **Analysis → Regression/Correlation → Multiple Regression**
- Select *TASPRICE* as the **Dependent Variable** and *landval* and *imprval* as the **Independent Variables**.
- In the **Model** tab, select **Remove Intercept**
- Click the **Play** icon to run the procedure.

**Page 3.13: Multiple Regression**

- **Analysis → Regression/Correlation → Multiple Regression**
- Select *LANDRSID* as the **Dependent Variable** and the listed 17 variables as the **Independent Variables**.
- In the **Model** tab, de-select **Remove Intercept**
- Select the same Reports as listed for Page 1.29
- Click the **Play** icon to run the procedure.

**Page 3.15: Stepwise Regression**

- **Analysis → Regression/Correlation → Variable Selection Routines → Stepwise Regression**
- Select *LANDRSID* as the **Dependent Variable** and the listed 16 variables shown on Page 3.14 as the **Independent Variables**.

- In the **Model Selection** section, choose **Stepwise**, under **Prob to Enter** select 0.15, under **Prob to Remove** select 0.20, under **Maximum Iterations** choose 40
- Click the **Play** icon to run the procedure.

### Page 3.17: Casewise Diagnostics, Storing Regression Results and filtering out “outliers”

- **Analysis → Regression/Correlation → Multiple Regression**
- Select *LANDRSID* as the **Dependent Variable** and the eight variables from Page 3.16 as the **Independent Variables**.
- In the **Storage** tab under **Storage Option** select **Store in empty columns only**, under **Store First Variable In** select the first available empty column (e.g. C100) and under **Select Items to Store on the Spreadsheet** select **Residuals**.
- The Report Section called Regression Diagnostics is equivalent to the Casewise Diagnostics table. Search for those records with a Standardized Residual outside of  $\pm 3.00$
- Record 382 is the least “bad” of the outliers so create filter based on the new Residual column. Filter out those records outside of  $\pm 37800$  (to set the filter use the AND option with two rows in the filter dialog box **C100 > -37800** and **C100 < 37800** – replace **C100** with the actual column number of the Residual column)

### Page 3.34: Cubic and Quadratic Curve Estimation

- **Analysis → Curve Fitting → One Independent Variable → Curve Fitting - General**
- In the Variables tab, select *SIZEFACT* as the **Y Variable** and *LOTSIZE2* as the **X Variable**. For **Preset Model** (in the **Model box**, still in **Variables tab**), choose **Cubic** then press **Play**. Repeat with **Quadratic**.
- The number of decimal places for the coefficients can be changed under the **Reports** tab, **Report Options** section

**NOTE:** Regression reports will be covered in more detail in Lesson 5.

## LESSON 4

### Page 4.5: Cubic and Quadratic Curve Estimation

- First, set the filter to *OBSDEPRN* > 0
- **Analysis → Curve Fitting → One Independent Variable → Curve Fitting - General**
- In the Variables tab, select *OBSDEPRN* as the **Y Variable** and *EFFAGE* as the **X Variable**. For **Preset Model** (in the **Model box**, still in **Variables tab**), choose **Cubic** then press **Play**. Repeat with **Quadratic**.

## LESSON 5

### Page 5.9: Multiple Regression

- **Analysis → Regression/Correlation → Multiple Regression**
- For **Variables**, select *Adj\_Price* as the **Dependent Variable** and place all other variables in the Numeric Independent Variables section.
- Click on **Reports tab**:
- NCSS 2004: Under **Choose a Group of Reports and Plots**, select **Display only those items that are CHECKED BELOW**. Select Run Summary, Means-Std Dev's, Correlations, Equation,

Coefficient, ANOVA Summary, R-Squared, Multicollinearity, Reg Diagnostics, Histogram, and Res vs Yhat Plot.

- NCSS 2007: In the Reports tab, under **Select a Group of Reports and Plots**, choose **Display only those items that are CHECKED BELOW**. Select Run Summary, Regression Coefficients, ANOVA Summary, and Multicollinearity.
- Click the **Play** icon to run the procedure.

### Multiple Regression Report

Page/Date/Time 1 20/03/2008 12:35:55 PM  
 Database C:\Program Files\NCSS2007\444 Data 2008\Midsize700.S0  
 Dependent Adj\_Price  
 Warning: At least one value was reset to 0.0 because it was less than the machine zero of 1E-16.

#### Run Summary Section

Parameter	Value	Parameter	Value
Dependent Variable	Adj_Price	Rows Processed	700
Number Ind. Variables	22	Rows Filtered Out	0
Weight Variable	None	Rows with X's Missing	0
R2	0.7994	Rows with Weight Missing	0
Adj R2	0.7929	Rows with Y Missing	0
Coefficient of Variation	0.0806	Rows Used in Estimation	700
Mean Square Error	1.541353E+09	Sum of Weights	700.000
Square Root of MSE	39260.07	Completion Status	Normal Completion
Ave Abs Pct Error	6.097		

#### Regression Coefficient Section

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	180050.2456	14764.4604	151112.4349	208988.0563	0.0000
ADJLOTSZ	32.9554	2.1048	28.8300	37.0807	0.3964
Bedrooms	-2117.2803	1950.6718	-5940.5268	1705.9662	-0.0282
CarPort corner	5302.3059	7157.0735	-8725.3005	19329.9123	0.0150
crawl	-4920.2018	5933.7730	-16550.1832	6709.7796	-0.0146
DeckAreaCovered	-33394.9947	4380.4483	-41980.5156	-24809.4738	-0.1858
DeckAreaUncovered	-32.5769	19.0521	-69.9184	4.7646	-0.0372
effage	26.0025	12.0022	2.4785	49.5265	0.0492
effage	-1694.1277	259.0516	-2201.8596	-1186.3959	-0.2389
Fireplcs	7303.4953	2492.4080	2418.4654	12188.5252	0.0636
FullBath	-0.0746	-5668.7579	3168.3105	-11878.5324	541.0167
HalfBath	-6911.6300	3687.8718	-14139.7258	316.4659	-0.0461
linbsmtfin	-10.8899	3.4030	-17.5597	-4.2201	-0.1184
linfinarea	25.5040	2.2174	21.1578	29.8501	0.5843
MultiCarGarage	17799.8748	6225.5818	5597.9587	30001.7908	0.0841
outbldgbin	307.7169	8929.2650	-17193.3209	17808.7548	0.0006
partbsmt	-6600.3492	6425.6411	-19194.3742	5993.6759	-0.0195
poolyes	32506.0836	14238.1894	4599.7452	60412.4220	0.0401
SingleCarGarage	-716.3324	6733.8424	-13914.4210	12481.7562	-0.0023
slab	-37460.9229	8291.5203	-53712.0040	-21209.8418	-0.0866
story15	45904.3142	9689.6043	26913.0387	64895.5897	0.0887
story20	38723.2770	5669.0713	27612.1014	49834.4527	0.2160
ThreeQtrBath	-6535.3104	4092.9083	-14557.2633	1486.6426	-0.0388

Note: The T-Value used to calculate these confidence limits was 1.960.

**Multiple Regression Report**

Page/Date/Time 2 20/03/2008 12:35:55 PM  
 Database C:\Program Files\NCSS2007\444 Data 2008\Midsize700.S0  
 Dependent Adj\_Price  
 Warning: At least one value was reset to 0.0 because it was less than the machine zero of 1E-16.

**Analysis of Variance Section**

Source	DF	Sum of R2	Mean Squares	Prob Square	Power F-Ratio	Level	(5%)
Intercept	1	1.660808E+14	1.660808E+14				
Model	22	0.7994	4.158931E+12	1.890423E+11	122.647	0.0000	1.0000
Error	677	0.2006	1.043496E+12	1.541353E+09			
Total(Adjusted)	699	1.0000	5.202427E+12	7.442671E+09			

**Storage Variables Section**

Stored Item	First Variable	Last Variable
Predicted	Y	C70
Residuals		C71

This report lists the variables on the database into which various statistics were stored.

**Multicollinearity Section**

Independent Variable	Variance Inflation Factor	R2 Versus Other I.V.'s	Tolerance	Diagonal of X'X Inverse
ADJLOTSZ	2.1631	0.5377	0.4623	2.874194E-09
Bedrooms	2.2811	0.5616	0.4384	2.468688E-03
CarPort	1.3785	0.2746	0.7254	3.323294E-02
corner	1.0410	0.0394	0.9606	2.284335E-02
crawl	2.0043	0.5011	0.4989	1.244901E-02
DeckAreaCovered	1.5982	0.3743	0.6257	2.354967E-07
DeckAreaUncovered	1.7404	0.4254	0.5746	9.345942E-08
effage	4.5025	0.7779	0.2221	4.353821E-05
Fireplcs	1.5903	0.3712	0.6288	4.030288E-03
FullBath	5.8677	0.8296	0.1704	6.512584E-03
HalfBath	2.0403	0.5099	0.4901	8.823674E-03
linbsmtfin	4.6223	0.7837	0.2163	7.513227E-09
linfinarea	8.7097	0.8852	0.1148	3.190104E-09
MultiCarGarage	2.9169	0.6572	0.3428	2.514535E-02
outbldgbin	1.1507	0.1309	0.8691	5.172843E-02
partbsmt	1.2207	0.1808	0.8192	2.678741E-02
poolyes	1.0402	0.0386	0.9614	0.1315247
SingleCarGarage	1.6382	0.3896	0.6104	2.941872E-02
slab	1.2399	0.1935	0.8065	4.460322E-02
story15	1.1835	0.1550	0.8450	6.091299E-02
story20	3.3744	0.7036	0.2964	2.085075E-02
ThreeQtrBath	1.9975	0.4994	0.5006	1.086831E-02



**Page 5.10: Standard Error of the Estimate**

- In the NCSS regression results, the SEE is referred to as "Square Root of MSE".

**Page 5.10: Multicollinearity**

- The NCSS regression results show the VIF (Variance Inflation Factor) and Tolerance for each variable. The R2 versus other IV's column is equivalent to 1 - VIF. You can ignore the last column.

**Page 5.15: Transformations**

- A Transformation types that have not previously been explained are below:
- Deck: Enter *Deck* in the next available row in the Variable info sheet, then enter the following in the **Transformation cell**:  $DeckAreaCovered + DeckAreaUncovered * .75$ . The variable *Garages* would be done in a similar fashion:  $1.75 * MultiCarGarage + SingleCarGarage + .3 * Carport$ .

**Page 5.16: Separating Model and Test Databases**

- **Data → Sort**. Select Random and Ascending. Click **OK** to run.
- **Save** the database as "midsizecomplete.S0"
- In the **Sheet1** window, scroll through the data in the Random column. At row 500, random shows 704.
- Use your mouse to select rows 501 to 700 (to end of list). Right-click on the selected block and select Delete from the Edit menu.
- You should now have only 500 rows of data remaining.
- **File → Save As**. Save the file as "midsizemodel.S0" (that's "Szero").
- To create the test database, re-open the "midsizecomplete.S0" file.
- **Sort by Random** and ensure row 500 is 704.
- Use your mouse to select rows 1 to 500. Right-click on the selected block and select Delete.
- **File → Save As**. Save the file as "midsizetest.S0"
- Open the "midsizemodel.s0"

**Page 5.17: Stepwise Regression**

- **Analysis → Regression/Correlation → Variable Selection Routines → Stepwise Regression**.
- For Variables, select *Adj\_Price* as the **Dependent Variable** and *adjlotsz*, *slab*, *poolyes*, *partbsmt*, *Fireplcs*, *crawl*, *Bedrooms*, *effage*, *outbldgbin*, *story15*, *story20*, *corner*, *linfinarea*, *DECKS*, and *GARAGES* as **Independent Variables**.
- In the **Selection Method pulldown menu**, choose Min MSE
- In the **Min RMSE change pulldown menu** choose 0.000
- Click **Reports** - Under **Report Format** select **Brief**
- Click the **Play** icon to run the procedure.

**Stepwise Regression Report**

Page/Date/Time	1 26/03/2008 2:52:29 PM
Database	C:\Program Files\NCSS2007\444 Data 2008\Midsizemodel.S0
Dependent	Adj_Price

## Iteration Detail Section

Iter. No.	Action	Variable	R-Squared	Sqrt(MSE)	Max R-Squared Other X's
0	Unchanged	0.000000	86599	0.000000	
1	Added	linfinarea	0.635721	52319.77	0.000000
2	Added	adjlotsz	0.681540	48968.04	0.000144
3	Added	effage	0.744327	43920.27	0.580803
4	Added	story20	0.760089	42587.83	0.622178
5	Added	crawl	0.769102	41822.48	0.622472
6	Added	slab	0.775686	41263.68	0.624482
7	Added	poolyes	0.779366	40965.33	0.624760
8	Added	Fireplcs	0.782717	40694.46	0.637578
9	Added	story15	0.785908	40435.73	0.642609
10	Added	Bedrooms	0.787255	40349.48	0.728349
11	Added	garages	0.788581	40264.75	0.728650
12	Added	partbsmt	0.790020	40168.69	0.728726
13	Unchanged	0.790020	40168.69	0.728726	

corner, prtbsmt, Outbuildingbin are not included in the stepwise iterations.

## Page 5.19: Multiple Linear Regression

- **Analysis → Regression/Correlation → Multiple Regression**
- Select *Adj\_Price* as the **Dependent Variable** and input *adjlotsz*, *slab*, *poolyes*, *partbsmt*, *Fireplcs*, *crawl*, *Bedrooms*, *effage*, *story15*, *story20*, *linfinarea*, *GARAGES* and *corner* (*corner* was added in error in the workbook, so it is included here so that students following this supplement get similar results) as **Numeric Independent Variables**.
- In the **Reports** tab, choose Select **Display only those items that are CHECKED BELOW** in the **Select Report** section.
- In the **Select Reports** box, choose Run Summary, Regression Coefficients, ANOVA Summary, Multicollinearity and Regression Diagnostics box
- In NCSS there is no option to select outliers, so the Regression Diagnostics report includes all residuals outside  $\pm 2$  Std. Dev. Looking at the Row statistics at the end of the report, note that those outside  $\pm 3$  std dev are the same as those shown in the SPSS report on page 5.19 of the workbook. The NCSS statistic *Rstudent* has similar values to SPSS standardized residuals, so saving it in NCSS and using it as the basis to filter outliers is a similar method.
- In the **Storage** tab select **Store in empty columns only**, then check **Rstudent** to be stored.
- Click the **Play** icon to run the procedure. The results, before we filter out the Standard Residuals greater than 3 and less than -3, are as follows:

## Multiple Regression Report

Page/Date/Time 1 26/03/2008 3:13:39 PM  
 Database C:\Program Files\NCSS2007\444 Data 2008\Midsizemodel.S0  
 Dependent Adj\_Price  
 Warning: At least one value was reset to 0.0 because it was less than the machine zero of 1E-16.

## Run Summary Section

Parameter	Value	Parameter	Value
Dependent Variable	Adj_Price	Rows Processed	500
Number Ind. Variables	13	Rows Filtered Out	0

Weight Variable	None	Rows with X's Missing	0
R2	0.7904	Rows with Weight Missing	0
Adj R2	0.7848	Rows with Y Missing	0
Coefficient of Variation	0.0827	Rows Used in Estimation	500
Mean Square Error	1.613757E+09	Sum of Weights	500.000
Square Root of MSE	40171.59	Completion Status	Normal Completion
Ave Abs Pct Error	6.334		

Regression Coefficient Section

<b>Independent Variable</b>	<b>Regression Coefficient</b>	<b>Standard Error</b>	<b>Lower 95% C.L.</b>	<b>Upper 95% C.L.</b>	<b>Standardized Coefficient</b>
Intercept	176013.4648	16130.3962	144398.4692	207628.4604	0.0000
adjlotsz	34.2558	2.3278	29.6933	38.8182	0.4095
Bedrooms	-4348.7683	2217.8498	-8695.6741	-1.8625	-0.0576
corner	-6957.4491	7216.5314	-21101.5908	7186.6926	-0.0205
crawl	-29824.1983	4780.9414	-39194.6712	-20453.7253	-0.1655
effage	-1812.1868	281.3259	-2363.5754	-1260.7981	-0.2524
Fireplcs	7588.8985	2962.6276	1782.2552	13395.5418	0.0656
garages	8443.5735	4101.4850	404.8106	16482.3364	0.0564
linfinarea	19.9309	1.7687	16.4643	23.3976	0.4506
partbsmt	-14376.2268	7924.7454	-29908.4425	1155.9888	-0.0401
poolyes	42234.5020	16602.2013	9694.7855	74774.2185	0.0532
slab	-31843.1928	8924.1127	-49334.1322	-14352.2534	-0.0787
story15	34723.1482	11785.4624	11624.0664	57822.2300	0.0639
story20	40622.6119	5427.7184	29984.4792	51260.7445	0.2267

Note: The T-Value used to calculate these confidence limits was 1.960.

Analysis of Variance Section

<b>Source</b>	<b>DF</b>	<b>Sum of R2</b>	<b>Mean Squares</b>	<b>Prob F-Ratio</b>	<b>Power Level</b>	<b>(5%)</b>
Intercept	1	1.178364E+14	1.178364E+14			
Model	13	0.7904	2.957908E+12	2.275314E+11	140.995	0.0000
Error	486	0.2096	7.842859E+11	1.613757E+09		1.0000
Total(Adjusted)	499	1.0000	3.742194E+12	7.499387E+09		

Multiple Regression Report

Page/Date/Time 2 26/03/2008 3:13:39 PM  
 Database C:\Program Files\NCSS2007\444 Data 2008\Midsizemodel.S0  
 Dependent Adj\_Price  
 Warning: At least one value was reset to 0.0 because it was less than the machine zero of 1E-16.

Multicollinearity Section

<b>Independent Variable</b>	<b>Variance Inflation Factor</b>	<b>R2 Versus Other I.V.'s</b>	<b>Tolerance</b>	<b>Diagonal of X'X Inverse</b>
adjlotsz	1.7957	0.4431	0.5569	3.357858E-09
Bedrooms	2.0005	0.5001	0.4999	3.048078E-03
corner	1.0504	0.0480	0.9520	3.227148E-02
crawl	1.6317	0.3871	0.6129	1.416409E-02
effage	3.5613	0.7192	0.2808	4.904348E-05
Fireplcs	1.5192	0.3418	0.6582	5.438962E-03
garages	1.7415	0.4258	0.5742	1.042423E-02

linfinarea	3.7074	0.7303	0.2697	1.938583E-09
partbsmt	1.1316	0.1163	0.8837	3.891639E-02
poolyes	1.0125	0.0124	0.9876	0.1708021
slab	1.1276	0.1131	0.8869	4.935054E-02
story15	1.0898	0.0824	0.9176	8.607066E-02
story20	2.1277	0.5300	0.4700	1.825562E-02

### Multiple Regression Report

Page/Date/Time 3 26/03/2008 3:13:39 PM

Database C:\Program Files\NCSS2007\444 Data 2008\Midsizemodel.S0

Dependent Adj\_Price

Warning: At least one value was reset to 0.0 because it was less than the machine zero of 1E-16.

### Regression Diagnostics Section

Row	Standardized Residual	RStudent	Hat Diagonal	Cook's D	Dffits	CovRatio
11	2.3441	2.3550	0.0105	0.0042	0.2428	0.8871
50	-2.7315	-2.7499	0.0238	0.0130	-0.4295	0.8490
58	3.3802	3.4171	0.0102	0.0084	0.3476	0.7453
137	2.1775	2.1859	0.0222	0.0077	0.3294	0.9176
251	-2.2613	-2.2709	0.0200	0.0075	-0.3244	0.9057
260	2.3506	2.3616	0.0067	0.0027	0.1936	0.8829
270	2.0343	2.0410	0.0065	0.0019	0.1651	0.9191
275	-2.1134	-2.1210	0.0119	0.0038	-0.2325	0.9153
304	2.1063	2.1138	0.0106	0.0034	0.2186	0.9149
327	-2.8145	-2.8348	0.0418	0.0247	-0.5922	0.8534
351	2.1201	2.1278	0.0114	0.0037	0.2280	0.9141
375	-2.4476	-2.4603	0.0244	0.0107	-0.3889	0.8868
377	-2.2402	-2.2495	0.0064	0.0023	-0.1805	0.8958
380	-2.0783	-2.0855	0.0116	0.0036	-0.2256	0.9190
392	-3.3411	-3.3767	0.0395	0.0328	-0.6846	0.7740
408	2.3470	2.3580	0.0183	0.0073	0.3221	0.8938
429	2.7308	2.7492	0.0128	0.0069	0.3127	0.8396
439	-2.3180	-2.3286	0.0398	0.0159	-0.4738	0.9174
452	2.5068	2.5206	0.0343	0.0159	0.4749	0.8883
453	-2.0799	-2.0871	0.0230	0.0073	-0.3203	0.9295
483	3.6540	3.7014	0.0427	0.0425	0.7818	0.7280
486	4.5356	4.6300	0.0207	0.0311	0.6731	0.5736
498	-3.9822	-4.0446	0.0073	0.0083	-0.3469	0.6516

The highlighted values are the same rows as shown in SPSS report.

### Page 5.19: Filter sales outside 3 standard deviations

- Go to the **Variable Info** worksheet and name the last variable RStudent.
- **Data** → **Filter** choose to **Keep Spreadsheet Row If**: “Only if all statements are true (AND)”
- Enter  $RStudent > -3$  into one field and  $RStudent < 3$  in another.
- Make sure the **Filter System Active box** is checked.
- Run the procedure.

Now re-run the Multiple Regression Report:

**Multiple Regression Report**

Page/Date/Time 1 27/03/2008 12:53:07 PM  
 Database C:\PROGRAM FILES\NCSS2007\444 DATA 2008\MIDSIZEMODEL.S0  
 Filter Rstudent > -3 and Rstudent < 3  
 Dependent Adj\_Price

**Run Summary Section**

Parameter	Value	Parameter	Value
Dependent Variable	Adj_Price	Rows Processed	500
Number Ind. Variables	13	Rows Filtered Out	5
Weight Variable	None	Rows with X's Missing	0
R2	0.8172	Rows with Weight Missing	0
Adj R2	0.8123	Rows with Y Missing	0
Coefficient of Variation	0.0769	Rows Used in Estimation	495
Mean Square Error	1.389607E+09	Sum of Weights	495.000
Square Root of MSE	37277.43	Completion Status	Normal Completion
Ave Abs Pct Error	6.101		

**Regression Coefficient Section**

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	183191.1025	15172.5667	153378.4023	213003.8027	0.0000
adjlotsz	33.0489	2.2204	28.6860	37.4117	0.3911
Bedrooms	-4581.2972	2058.6462	-8626.3479	-536.2465	-0.0612
corner	-6213.9993	6699.1052	-19377.1257	6949.1272	-0.0185
crawl	-30919.4144	4462.7006	-39688.2113	-22150.6175	-0.1726
effage	-1827.1647	262.3723	-2342.7021	-1311.6273	-0.2548
Fireplcs	8277.2614	2754.3672	2865.1828	13689.3400	0.0722
garages	9601.3144	3824.4692	2086.5837	17116.0452	0.0645
linfinarea	20.4194	1.6501	17.1771	23.6618	0.4647
partbsmt	-14427.0690	7361.2955	-28891.3386	37.2006	-0.0407
poolyes	41987.6414	15406.9134	11714.4715	72260.8112	0.0535
slab	-30847.4378	8296.6763	-47149.6448	-14545.2308	-0.0771
story15	35949.2569	10940.9136	14451.3664	57447.1474	0.0669
story20	37879.8398	5083.4621	27891.3036	47868.3760	0.2128

Note: The T-Value used to calculate these confidence limits was 1.965.

**Analysis of Variance Section**

Source	DF	R2	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (5%)
Intercept	1		1.164139E+14	1.164139E+14			
Model	13	0.8172	2.987878E+12	2.298368E+11	165.397	0.0000	1.0000
Error	481	0.1828	6.684007E+11	1.389607E+09			
Total(Adjusted)	494	1.0000	3.656279E+12	7.401374E+09			

**Storage Variables Section**

Stored Item	First Variable	Last Variable
Rstudent		C70

This report lists the variables on the database into which various statistics were stored.

**Multiple Regression Report**

Page/Date/Time 2 27/03/2008 12:53:07 PM  
 Database C:\PROGRAM FILES\NCSS2007\444 DATA 2008\MIDSIZEMODEL.S0  
 Filter Rstudent > -3 and Rstudent < 3  
 Dependent Adj\_Price

**Multicollinearity Section**

<b>Variance Independent Variable</b>	<b>R2 Inflation Factor</b>	<b>Versus Other I.V.'s</b>	<b>Diagonal Tolerance</b>	<b>of X'X Inverse</b>
adjlotsz	1.8162	0.4494	0.5506	3.547831E-09
Bedrooms	1.9912	0.4978	0.5022	3.049801E-03
corner	1.0504	0.0480	0.9520	3.229548E-02
crawl	1.6337	0.3879	0.6121	0.0143319
effage	3.5232	0.7162	0.2838	4.953863E-05
Fireplcs	1.5190	0.3417	0.6583	5.459487E-03
garages	1.7372	0.4244	0.5756	1.052569E-02
linfinarea	3.7099	0.7304	0.2696	1.959485E-09
partbsmt	1.1332	0.1175	0.8825	3.899569E-02
poolyes	1.0125	0.0123	0.9877	0.1708203
slab	1.1312	0.1160	0.8840	4.953549E-02
story15	1.0904	0.0829	0.9171	8.614207E-02
story20	2.1450	0.5338	0.4662	1.859633E-02

**Multiple Regression Report**

Page/Date/Time 3 27/03/2008 12:53:07 PM  
 Database C:\PROGRAM FILES\NCSS2007\444 DATA 2008\MIDSIZEMODEL.S0  
 Filter Rstudent > -3 and Rstudent < 3  
 Dependent Adj\_Price

**Regression Diagnostics Section**

<b>Row</b>	<b>Standardized Residual</b>	<b>RStudent</b>	<b>Hat Diagonal</b>	<b>Cook's D</b>	<b>Dffits</b>	<b>CovRatio</b>
11	2.5348	2.5492	0.0107	0.0050	0.2651	0.8621
27	-2.0005	-2.0068	0.0134	0.0039	-0.2338	0.9283
50	-2.9668	-2.9912	0.0240	0.0155	-0.4692	0.8146
131	-2.1755	-2.1841	0.0159	0.0055	-0.2777	0.9109
137	2.4240	2.4364	0.0223	0.0096	0.3683	0.8866
188	-2.0066	-2.0129	0.0946	0.0301	-0.6507	1.0109
236	-2.0899	-2.0972	0.0108	0.0034	-0.2194	0.9160
245	-2.1162	-2.1239	0.0528	0.0178	-0.5012	0.9535
251	-2.4722	-2.4855	0.0201	0.0089	-0.3559	0.8784
260	2.4959	2.5096	0.0068	0.0030	0.2070	0.8637
270	2.1787	2.1872	0.0066	0.0022	0.1777	0.9020
275	-2.3060	-2.3165	0.0120	0.0046	-0.2553	0.8919
304	2.3104	2.3209	0.0107	0.0041	0.2409	0.8901
327	-2.9408	-2.9645	0.0426	0.0275	-0.6257	0.8342
351	2.2361	2.2455	0.0115	0.0042	0.2426	0.8998
374	-2.0733	-2.0805	0.0272	0.0086	-0.3476	0.9333
375	-2.6339	-2.6504	0.0245	0.0124	-0.4197	0.8611
377	-2.4359	-2.4486	0.0065	0.0028	-0.1975	0.8710
380	-2.2005	-2.2093	0.0116	0.0041	-0.2397	0.9041
398	2.0754	2.0826	0.0335	0.0107	0.3879	0.9392

---

408	2.4943	2.5080	0.0184	0.0083	0.3435	0.8741
429	2.9130	2.9360	0.0129	0.0079	0.3351	0.8129
439	-2.3380	-2.3490	0.0410	0.0167	-0.4859	0.9149
452	2.7371	2.7558	0.0351	0.0195	0.5255	0.8565
453	-2.2384	-2.2479	0.0234	0.0086	-0.3477	0.9104
499	-2.0513	-2.0582	0.0120	0.0036	-0.2266	0.9214

## LESSON 6

### Page 6.5: Descriptive Statistics:

- Select **Analysis** → **Descriptive Statistics** → **Descriptive Statistics**.
- Under Variables, select *LOTSIZESQFT*, *EFFECTIVEYEAR*, *FINISHEDAREA*, *BASEMENTFINISHEDAREA*, and *UPPERFINISH*.
- In the **Reports** tab, select Summary Section, Counts Section, Means Section, and Percentile Section.
- Also enter 5 and 95 for the **Smallest Percentile** and **Largest Percentile** drop down menus, respectively.
- Run the report.

### Page 6.9: Transformations

- Transformation types that have not previously been explained are below (many were covered in Lesson 1):
- *Nbhd36*: Enter *Nbhd36* in the next available row in the **Variable Info** sheet, then double click on the **Transformation cell**. Press **If-Then**, then enter *Nbhd=36* and (*Nbhd* < > 36) in the next **If box**. Enter *Nbhd36=1* and *Nbhd36=0* in the next **Then box**. Use similar syntax for *NBHD42*.
- *Fp*: Enter *Fireplcs* into the **Transformation cell** (this is essentially a rename).
- Recall that the **\*\*** notation in SPSS is replaced by a **^** symbol in NCSS

### Page 6.13: Non-Linear Regression

- **Analysis** → **Mass Appraisal** → **Nonlinear Regression**
- Enter *SalePrice* as the dependent variable and the following in the Model box:  
 $L*LotSizeSqft + UF*upperfinish + BFA*BasementFinishedArea$
- In **Model Parameters**, enter *L*, *UF* and *BFA* in the first three **Parameter** boxes (overwrite the defaults)
- Click the **Storage** tab
- Select an empty column under **Predicted Values**
- Run the report.
- After running this procedure, go to the **Variable Info** sheet and name the new column *ESTCRATE*

#### Nonlinear Regression Report

Page/Date/Time            1    08/04/2008 1:45:31 PM  
 Database                    C:\Users\saxon\Desktop\444 supplement\ReginaNLR.S0  
 Dependent                   SalePrice

**Minimization Phase Section**

<b>Itn No.</b>	<b>Error Sum Lambda</b>	<b>L</b>	<b>UF</b>	<b>BFA</b>	
0	9.946274E+13	0.00004	1	1	1
1	2.929555E+12	0.000016	16.00284	148.2683	54.34361
2	2.929555E+12	0.0000064	15.99804	148.2929	54.33177

Convergence criterion met.

**Model Estimation Section**

<b>Parameter Name</b>	<b>Parameter Estimate</b>	<b>Asymptotic Standard Error</b>	<b>Lower 95% C.L.</b>	<b>Upper 95% C.L.</b>
L	15.99804	1.055004	13.92522	18.07086
UF	148.2929	4.142214	140.1545	156.4313
BFA	54.33177	6.184043	42.18168	66.48186

Model SalePrice = L\*LOTSIZESQFT + UF\*UPPERFINISH + BFA\*BASEMENTFINISHEDAREA

R-Squared 0.092938  
Iterations 2

Estimated Model:

$(15.99804) * (\text{LOTSIZESQFT}) + (148.2929) * (\text{UPPERFINISH}) + (54.33177) * (\text{BASEMENTFINISHEDAREA})$

**Analysis of Variance Table**

<b>Source</b>	<b>Sum of DF</b>	<b>Mean Squares</b>	<b>Square</b>
Mean	1	1.00672E+14	1.00672E+14
Model	3	1.009721E+14	3.365737E+13
Model (Adjusted)	2	3.001634E+11	1.500817E+11
Error	497	2.929555E+12	5.894476E+09
Total (Adjusted)	499	3.229718E+12	
Total	500	1.039017E+14	

**Asymptotic Correlation Matrix of Parameters**

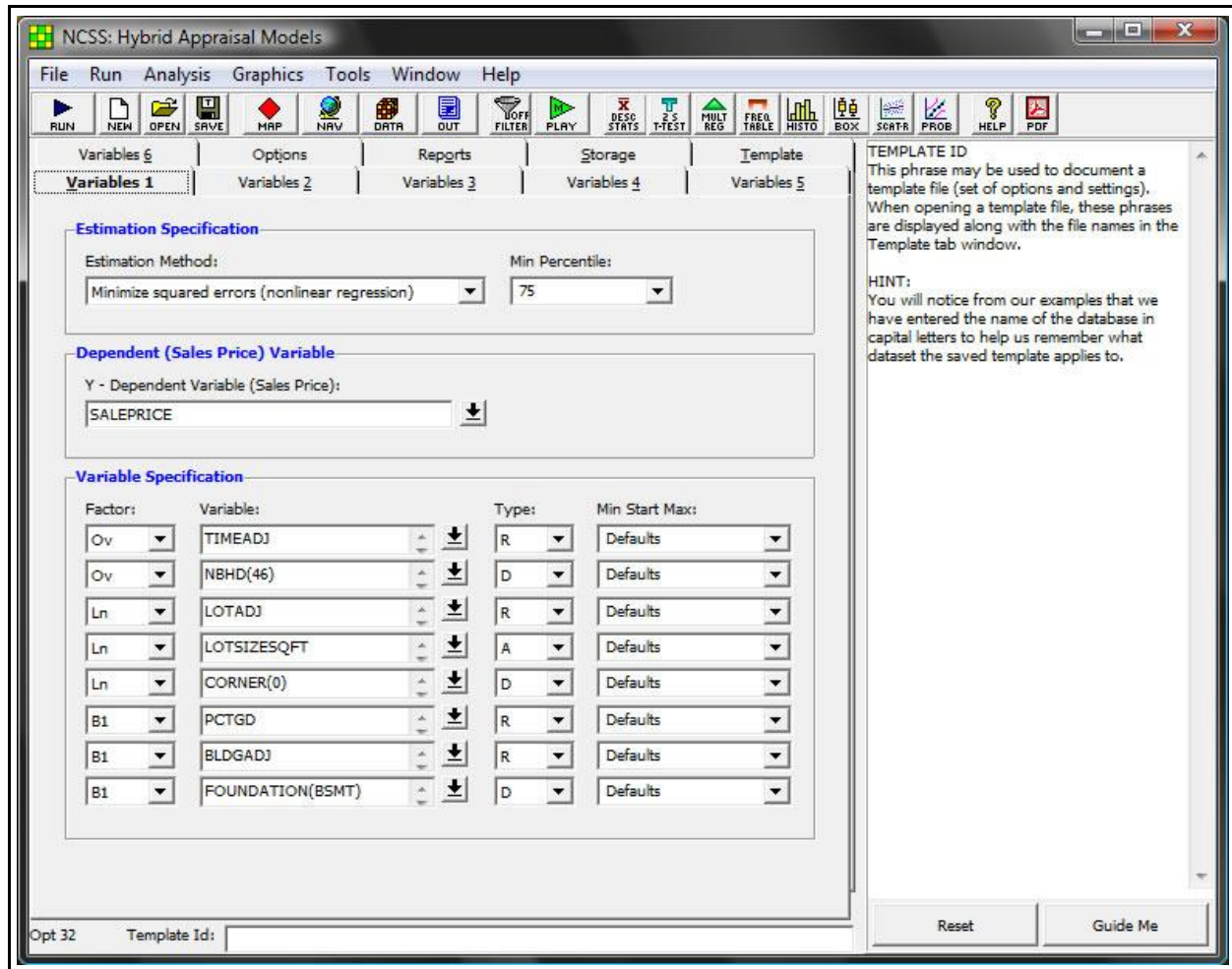
	<b>L</b>	<b>UF</b>	<b>BFA</b>
<b>L</b>	1.000000	-0.876138	-0.008309
<b>UF</b>	-0.876138	1.000000	-0.265805
<b>BFA</b>	-0.008309	-0.265805	1.000000



## Page 6.15: Non-Linear Regression

This model is much more complex than the one we ran on page 6.13, and we are using a different NCSS module to run it.

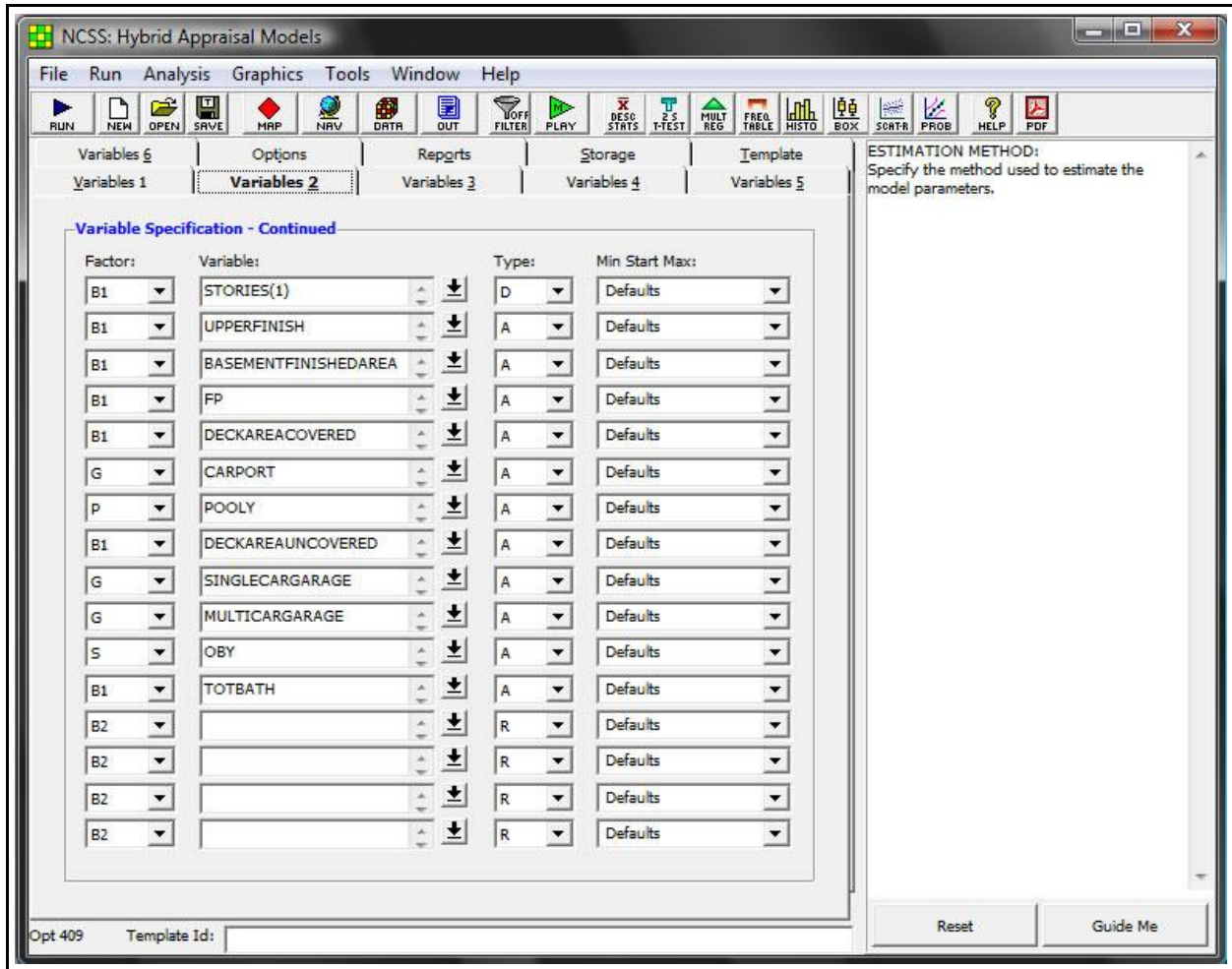
- **Analysis → Mass Appraisal → Hybrid Appraisal Models.**
- Fill out the **Variables 1 tab** as shown in the screenshot below



The factors used here are “Ov” for overall, “Ln” for Land, “B1” for main buildings. An “R” type allows a variable to be entered into the model with an exponential coefficient. A “D” type results in the variable being entered as an exponent in the model. An “A” type results in the variable being entered as a standard multiplicative item in the model.

Note, too, that we have typed “(46)” after the *NBHD* variable. Because of this, NCSS will enter the *NBHD* variable into the model, using *neighbourhood 46* as the reference value. Similar logic applies to *Corner* and *Foundation*.

- Fill out the **Variables 2** tab as shown in the screenshot below:



The factors used here are “B1” for main buildings, “G” for garages, “P” for pools, and “S” for sheds (outbuildings). Notice that *CarPort*, *SingleCarGarage*, and *MultiCarGarage* are entered as “A” (multiplicative) types, rather than a “D” (exponent) types as in the SPSS model. The reason is that NCSS will not allow the use of R or D type if the coefficient turns out to be negative.

- Fill out the **Options** tab as shown in the screenshot below

NCSS: Hybrid Appraisal Models

File Run Analysis Graphics Tools Window Help

Variables 1 Variables 2 Variables 3 Variables 4 Variables 5  
Variables 6 **Options** Reports Storage Template

**Nonlinear Regression Options**

Lambda: 0.1 Lambda Inc.: 10 Max Iterations: 40  
Nash Phi: 1 Lambda Dec.: 4 Min Iterations: 10

**Differential Evolution Options**

Max Generations: 50 Mutation Rate: 0.3 Min Amount: 5  
Individuals: 20 Grid Range: 4 Seed: 12346  
Inheritance: 0.85 Min Percent: 10

**'Min Start Max' Default Options**

Default 'Min Start Max' for Type = (A)mounts: -B 1 B Default 'Min Start Max' for Type = (D)iscrete: 0 1 5  
Default 'Min Start Max' for Type = (R)ate: -5 0 5

VARIABLE:  
Specify one or more variables with the same factor, type, minimum, maximum, and starting values. Usually, you will specify only one variable per line, but more are allowed if you desire.

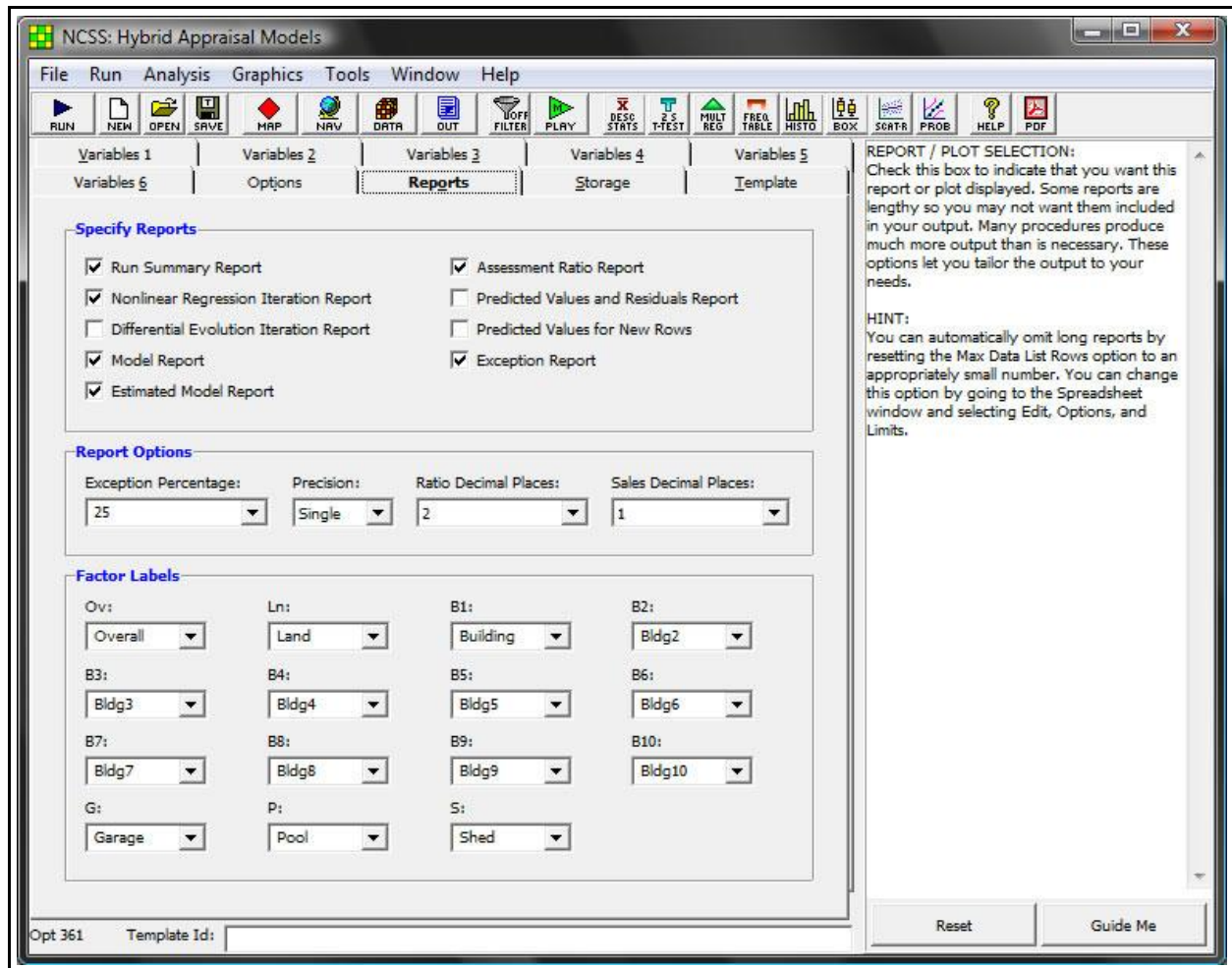
Binaries  
If this is a type D (discrete) variable, you can specify a reference value in parentheses after the variable name. If you do not specify a reference value, the program sorts the values and picks the last value as the reference value. The reference value is that value for which no binary variable is generated. The number of binary (0-1) variables generated is always one less than the number of unique values.  
For example, suppose you will use a variable called ExtType that has three possible values: B for brick, U for stucco, or S for siding. Further suppose that in your area, siding is the most common exterior type. Hence, siding would be by most obvious choice for the reference value. You would enter ExtType(S) for this variable. The program would generate two binary variables: one for brick and the other for stucco.

Single Binary  
It is possible to specify that only a single binary be generated for a type D (discrete) variable. This is done by adding a comma and an I after the reference value. When you do this, only a single binary variable is generated for the value indicated.  
For example, using the exterior type example given above, the statement ExtType(S,I) would cause the program to generate a single indicator variable that is '1' when the value is

Opt 35 Template Id:

Reset Guide Me

- Fill out the **Reports tab** as shown in the screenshot below



- Save the template
- Run the model.

The results we get are slightly different from those gathered from the SPSS model, probably because *SingleCarGarage*, *MultiCarGarage*, and *CarPort* were entered as standard multiplicative variables rather than exponents. Nevertheless, the results of the model below are very good and are worthy substitutes for the SPSS output. The R-Square value is approximately 0.81, the Assessment Ratio section shows an excellent result with mean 1.01, median 1.00, COD 6.00% and PRD 1.01.

**Hybrid Appraisal Model (Nonlinear Regression)**

Page/Date/Time            1   10/04/2008 9:48:31 AM  
 Database                    C:\Users\saxon\Desktop\444 supplement\ReginaNLR.S0  
 Dependent                  SalePrice

**Run Summary Report**

<b>Item</b>	<b>Value</b>
Model	SalePrice=Overall(Land+Building+Garage+Pool+Shed)
Estimation Method	Nonlinear Regression (SSE)
Final Value of SSE	627676577699.64
R-Squared (from NonLinReg)	0.8056559
Random Number Seed	-12346
Number of Variables Used	21
Number of Parameters in Model	24
Number of Rows Used	500
Number of N.R. Iterations	25
Number of D.E. Iterations	0

**Nonlinear Regression Iteration Section**

<b>Itn No.</b>	<b>Sum of Squared Errors</b>	<b>Message</b>
0	9.938559E+13	
1	8.73199E+13	Stepsize reduced to 5.956195E-02 by bounds.Lower bound active on B11.
2	2.64281E+13	Freeing parameter B11.Stepsiz reduced to 0.4875837 by bounds.Upper bound active on B4.
3	1.587726E+13	Freeing parameter B4.Stepsiz reduced to 0.77057 by bounds.Lower bound active on B13.
4	1.413987E+12	Freeing parameter B13.
5	1.223099E+12	
6	7.456335E+11	
7	6.283636E+11	
8	6.276769E+11	
9	6.276766E+11	
10	6.276766E+11	
11	6.276765E+11	
12	6.276765E+11	
13	6.276765E+11	
14	6.276765E+11	
15	6.276765E+11	
16	6.276765E+11	
17	6.276765E+11	
18	6.276765E+11	
19	6.276765E+11	
20	6.276765E+11	
21	6.276765E+11	
22	6.276765E+11	
23	6.276765E+11	
24	6.276765E+11	
25	6.276765E+11	

Convergence criterion met.

**Hybrid Appraisal Model (Nonlinear Regression)**

Page/Date/Time 2 10/04/2008 9:48:31 AM  
 Database C:\Users\saxon\Desktop\444 supplement\ReginaNLR.S0  
 Dependent SalePrice

**Model Specification and Estimation Section**

Parm Name	Variable Type	Variable Name	Parameter Estimate	Starting Value	Parameter Bounds
B1	Overall Rate	TimeAdj	1.507549	0	-5 to 5
B2	Overall Binary	Nbhd=36	0.9894186	1	0 to 5
B3	Overall Binary	Nbhd=42	0.9890087	1	0 to 5
B4	Land Rate	LotAdj	1.281385	0	-5 to 5
B5	Land Amount	LotSizeSqft	36.7611	1	-B to B
B6	Land Binary	Corner=1	0.9898927	1	0 to 5
B7	Building Rate	PctGd	1.162047	0	-5 to 5
B8	Building Rate	BldgAdj	0.828653	0	-5 to 5
B9	Building Binary	Foundation="CRAWL"	0.8060232	1	0 to 5
B10	Building Binary	Foundation="PBSMT"	0.913833	1	0 to 5
B11	Building Binary	Foundation="SLAB"	0.7906931	1	0 to 5
B12	Building Binary	Stories=1.5	1.154645	1	0 to 5
B13	Building Binary	Stories=2	1.138379	1	0 to 5
B14	Building Amount	upperfinish	117.8183	1	-B to B
B15	Building Amount	BasementFinishedArea	64.63383	1	-B to B
B16	Building Amount	Fp	18049.9	1	-B to B
B17	Building Amount	DeckAreaCovered	-47.93469	1	-B to B
B18	Garage Amount	CarPort	10120.65	1	-B to B
B19	Pool Amount	PoolY	37334.36	1	-B to B
B20	Building Amount	DeckAreaUncovered	38.95527	1	-B to B
B21	Garage Amount	SingleCarGarage	3987.354	1	-B to B
B22	Garage Amount	MultiCarGarage	7694.658	1	-B to B
B23	Shed Amount	ObY	-8683.398	1	-B to B
B24	Building Amount	TotBath	-6370.971	1	-B to B

**Model**

Factor	Detail
Overall	TimeAdj <sup>(B1)</sup> *(B2) <sup>(Nbhd=36)</sup> *(B3) <sup>(Nbhd=42)</sup>
Land	LotAdj <sup>(B4)</sup> *(B6) <sup>(Corner=1)</sup> *((B5) *LotSizeSqft)
Building	PctGd <sup>(B7)</sup> *BldgAdj <sup>(B8)</sup> *(B9) <sup>(Foundation="CRAWL")</sup> *(B10) <sup>(Foundation="PBSMT")</sup> *(B11) <sup>(Foundation="SLAB")</sup> *(B12) <sup>(Stories=1.5)</sup> *(B13) <sup>(Stories=2)</sup> *((B14) *upperfinish + (B15) *BasementFinishedArea + (B16) *Fp + (B17) *DeckAreaCovered + (B20) *DeckAreaUncovered + (B24) *TotBath)
Garage	(B18) *CarPort + (B21) *SingleCarGarage + (B22) *MultiCarGarage
Pool	(B19) *PoolY
Shed	(B23) *ObY

**Estimated Model**

TimeAdj^(1.50754868629256) \*(0.989418561945171)^(Nbhd=36)  
 \*(0.989008732555124)^(Nbhd=42)\*((LotAdj^(1.28138466873103) \*(0.989892670807007)^(Corner=1)  
 \*((36.7610952335244) \*LotSizeSqft))+(PctGd^(1.16204720256929) \*BldgAdj^(0.82865306140076)  
 \*(0.806023180185531)^(Foundation="CRAWL") \*(0.913832950693523)^(Foundation="PBSMT")  
 \*(0.790693129996248)^(Foundation="SLAB") \*(1.15464476601521)^(Stories=1.5)  
 \*(1.13837891681141)^(Stories=2) \*((117.818265614701) \*upperfinish +(64.6338378605057)  
 \*BasementFinishedArea +(18049.895842006) \*Fp +(-47.9346932529567) \*DeckAreaCovered  
 +(38.9552732639984) \*DeckAreaUncovered +(-6370.97129806837) \*TotBath))+((10120.6543889986)  
 \*CarPort +(3987.35393753278) \*SingleCarGarage +(7694.65765396906)  
 \*MultiCarGarage)+((37334.3649201748) \*PoolY)+((-8683.39892468647) \*ObY))

**Hybrid Appraisal Model (Nonlinear Regression)**

Page/Date/Time 3 10/04/2008 9:48:31 AM  
 Database C:\Users\saxon\Desktop\444 supplement\ReginaNLR.S0  
 Dependent SalePrice

**Assessment Ratio Section**

Statistic Name	Actual SalePrice	Predicted SalePrice	Percent Ratio	Error
Number of Cases	500	500	500	500
Mean	448713.6	448707.8	1.01	5.99
Minimum	250000.0	285696.8	0.71	0.08
Lower Quartile	387500.0	393027.4	0.96	2.16
Median	445000.0	448087.3	1.00	4.37
Upper Quartile	504836.0	495563.6	1.04	7.67
Maximum	690000.0	689607.7	1.36	36.35
Range	440000.0	403910.9	0.65	36.27
I. Q. Range	117336.0	102536.2	0.09	5.51
Variance	6472380559.4	5219708383.9	0.01	32.65
Std. Deviation	80451.1	72247.5	0.08	5.71
Ave  Dev. from Median	64090.5	57559.3	0.06	3.83
Coef. of Variation x 100	17.93	16.10	8.20	95.38
Coef. of Dispersion x 100	14.40	12.85	6.00	87.65
Weighted Mean			1.00	
Price Related Differential			1.01	

**Hybrid Appraisal Model (Nonlinear Regression)**

Page/Date/Time 4 10/04/2008 9:48:31 AM  
 Database C:\Users\saxon\Desktop\444 supplement\ReginaNLR.S0  
 Dependent SalePrice

**Exception Report Section**

Row No.	Actual SalePrice	Predicted SalePrice	Actual-Predicted (Residual)	Predicted/Actual (Ratio)	Percent  Error
50	308000.0	410056.4	-102056.4	1.33	33.14
275	330000.0	413806.0	-83806.0	1.25	25.40
327	250000.0	335044.5	-85044.5	1.34	34.02
392	345900.0	464655.6	-118755.6	1.34	34.33
439	250000.0	340869.9	-90869.9	1.36	36.35
453	318000.0	400411.7	-82411.7	1.26	25.92
486	595000.0	424109.7	170890.3	0.71	28.72
498	386250.0	525205.6	-138955.6	1.36	35.98

**Page 6.21: Non-Linear Regression**

- **Analysis → Mass Appraisal → Hybrid Appraisal Models.**
- **Load** the template from the previous model
- Replace *upperfinish* with *SflaGrng*, add *SflaUp* as B1 with A type, replace *DeckAreaCovered* with *Deck*,
- replace *CarPort* with *Garage*, remove *DeckAreaUncovered*, *SingleCarGarage*, *MultiCarGarage*, *TotBath*, and *OBY*.
- Save the template with a new name
- Run the model

The following are the final NCSS reports. As with the model on page 6.15, the results here are not identical to the SPSS output. Still, this model is a good one. The R-Square value is approximately 0.80, the Assessment Ratio section shows an excellent result with mean 1.01, median 1.00, COD 6.00% and PRD 1.01.

**Hybrid Appraisal Model (Nonlinear Regression)**

Page/Date/Time 1 10/04/2008 11:31:44 AM  
 Database C:\Users\saxon\Desktop\444 supplement\ReginaNLR.S0  
 Dependent SalePrice

**Run Summary Report**

Item	Value
Model	SalePrice = Overall(Land + Building + Garage + Pool)
Estimation Method	Nonlinear Regression (SSE)
Final Value of SSE	637737047692.05
R-Squared (from NonLinReg )	0.802541
Random Number Seed	-12346



---

Number of Variables Used	17
Number of Parameters in Model	20
Number of Rows Used	500
Number of N.R. Iterations	27
Number of D.E. Iterations	0

### Nonlinear Regression Iteration Section

<b>Itn No.</b>	<b>Sum of Squared Errors</b>	<b>Message</b>
0	9.939921E+13	
1	8.838306E+13	Stepsize reduced to 6.055745E-02 by bounds.Lower bound active on B11.
2	3.709993E+13	Freeing parameter B11.Stepsize reduced to 0.3947916 by bounds.Upper bound active on B4.
3	3.374455E+13	Freeing parameter B4.Stepsize reduced to 0.5916632 by bounds.Lower bound active on B13.
4	8.193334E+12	Freeing parameter B13.
5	3.367114E+12	Stepsize reduced to 0.6239424 by bounds.Lower bound active on B13.
6	3.153185E+12	Freeing parameter B13.Stepsize reduced to 0.1912342 by bounds.Lower bound active on B11.
7	2.657001E+12	Stepsize reduced to 0.2456735 by bounds.Lower bound active on B9.
8	2.429107E+12	Freeing parameter B11.Stepsize reduced to 0.1048149 by bounds.Lower bound active on B12.
9	1.235503E+12	Freeing parameter B9.Stepsize reduced to 8.492912E-02 by bounds.Stepsize reduced to 0.3847836 by bounds.Stepsize reduced to 0.5156957 by bounds.Lower bound active on B10.
10	1.020599E+12	Freeing parameter B10.
11	9.090545E+11	
12	7.669985E+11	Freeing parameter B12.
13	7.145771E+11	
14	6.429183E+11	
15	6.41564E+11	
16	6.402104E+11	
17	6.390784E+11	
18	6.381804E+11	
19	6.377579E+11	
20	6.377376E+11	
21	6.37737E+11	
22	6.37737E+11	
23	6.37737E+11	
24	6.37737E+11	
25	6.37737E+11	
26	6.37737E+11	
27	6.37737E+11	

Convergence criterion met.

**Hybrid Appraisal Model (Nonlinear Regression)**

Page/Date/Time 2 10/04/2008 11:31:44 AM  
 Database C:\Users\saxon\Desktop\444 supplement\ReginaNLR.S0  
 Dependent SalePrice

**Model Specification and Estimation Section**

Parm Name	Variable Type	Variable Name	Parameter Estimate	Starting Value	Parameter Bounds
B1	Overall Rate	TimeAdj	1.526985	0	-5 to 5
B2	Overall Binary	Nbhd=36	0.9889697	1	0 to 5
B3	Overall Binary	Nbhd=42	0.9895793	1	0 to 5
B4	Land Rate	LotAdj	1.222849	0	-5 to 5
B5	Land Amount	LotSizeSqft	34.31924	1	-B to B
B6	Land Binary	Corner=1	0.9868731	1	0 to 5
B7	Building Rate	PctGd	0.8844188	0	-5 to 5
B8	Building Rate	BldgAdj	0.833165	0	-5 to 5
B9	Building Binary	Foundation="CRAWL"	0.8121947	1	0 to 5
B10	Building Binary	Foundation="PBSMT"	0.9124588	1	0 to 5
B11	Building Binary	Foundation="SLAB"	0.7890924	1	0 to 5
B12	Building Binary	Stories=1.5	1.22433	1	0 to 5
B13	Building Binary	Stories=2	1.294076	1	0 to 5
B14	Building Amount	SflaGrnd	121.4899	1	-B to B
B15	Building Amount	BasementFinishedArea	56.09948	1	-B to B
B16	Building Amount	Fp	16361.92	1	-B to B
B17	Building Amount	Deck	5.130105	1	-B to B
B18	Garage Amount	Garage	3926.091	1	-B to B
B19	Pool Amount	PoolY	35983.61	1	-B to B
B20	Building Amount	SflaUp	79.88901	1	-B to B

**Model**

Factor	Detail
Overall	TimeAdj^(B1) *(B2)^(Nbhd=36) *(B3)^(Nbhd=42)
Land	LotAdj^(B4) *(B6)^(Corner=1) *(B5) *LotSizeSqft)
Building	PctGd^(B7) *BldgAdj^(B8) *(B9)^(Foundation="CRAWL") *(B10)^(Foundation="PBSMT") *(B11)^(Foundation="SLAB") *(B12)^(Stories=1.5) *(B13)^(Stories=2) *((B14) *SflaGrnd +(B15) *BasementFinishedArea +(B16) *Fp +(B17) *Deck +(B20) *SflaUp)
Garage	(B18) *Garage
Pool	(B19) *PoolY

**Estimated Model**

TimeAdj^(1.52698487977602) \*(0.988969679056857)^(Nbhd=36)  
 \*(0.98957927584407)^(Nbhd=42)\*((LotAdj^(1.22284927310725) \*(0.986873133217331)^(Corner=1)  
 \*((34.3192421522082) \*LotSizeSqft))+(PctGd^(0.884418855588737) \*BldgAdj^(0.833165025259994)  
 \*(0.812194699201455)^(Foundation="CRAWL") \*(0.912458820446835)^(Foundation="PBSMT")  
 \*(0.789092409711167)^(Foundation="SLAB") \*(1.22432983780895)^(Stories=1.5)  
 \*(1.29407555828283)^(Stories=2) \*((121.489868014199) \*SflaGrnd +(56.0994806386785)  
 \*BasementFinishedArea +(16361.9181127457) \*Fp +(5.13010470842204) \*Deck +(79.8890038770536)  
 \*SflaUp))+((3926.09093522549) \*Garage)+((35983.609502867) \*PoolY))

**Hybrid Appraisal Model (Nonlinear Regression)**

Page/Date/Time 3 10/04/2008 11:31:44 AM  
 Database C:\Users\saxon\Desktop\444 supplement\ReginaNLR.S0  
 Dependent SalePrice

**Assessment Ratio Section**

Statistic Name	Actual SalePrice	Predicted SalePrice	Percent Ratio	Error
Number of Cases	500	500	500	500
Mean	448713.6	448721.5	1.01	6.00
Minimum	250000.0	285252.9	0.72	0.02
Lower Quartile	387500.0	394228.2	0.96	2.09
Median	445000.0	447201.3	1.00	4.42
Upper Quartile	504836.0	495426.6	1.05	7.79
Maximum	690000.0	695051.8	1.37	37.11
Range	440000.0	409798.9	0.65	37.09
I. Q. Range	117336.0	101198.4	0.09	5.70
Variance	6472380559.4	5187303272.6	0.01	33.38
Std. Deviation	80451.1	72022.9	0.08	5.78
Ave  Dev. from Median	64090.5	57510.9	0.06	3.85
Coef. of Variation x 100	17.93	16.05	8.25	96.29
Coef. of Dispersion x 100	14.40	12.86	6.00	87.10
Weighted Mean Price Related Differential			1.00 1.01	

**Hybrid Appraisal Model (Nonlinear Regression)**

Page/Date/Time 4 10/04/2008 11:31:44 AM  
 Database C:\Users\saxon\Desktop\444 supplement\ReginaNLR.S0  
 Dependent SalePrice

**Exception Report Section**

Row No.	Actual SalePrice	Predicted SalePrice	Actual-Predicted (Residual)	Predicted/Actual (Ratio)	Percent  Error
50	308000.0	412552.3	-104552.3	1.34	33.95
327	250000.0	338000.1	-88000.1	1.35	35.20
392	345900.0	468426.3	-122526.3	1.35	35.42
439	250000.0	342770.7	-92770.7	1.37	37.11
453	318000.0	399196.2	-81196.2	1.26	25.53
486	595000.0	427704.2	167295.8	0.72	28.12
498	386250.0	528994.6	-142744.6	1.37	36.96

## LESSON 7

### Page 7.9 to 7.10: Correlation Matrix

- Analysis → Regression/Correlation → Correlation Routines → Correlation Matrix.
- Select units, effage, conditn, deprec, und\_park, elevator, CAP\_RATE, LOT\_AREA, AREA\_UNT, EXP\_EGI, NBHD1, NBHD3, NBHD4, STORY2, STORY4, and SML\_UNIT as Correlation Variables.
- Run the report.

### Page 7.14: Saving Predicted Values

(after entering the necessary variables in the Multiple Regression module)

- Click on the **Storage tab**
- Select **Store in empty columns only**
- check **Predicted Y** to store
- After running procedure to create the model and store values go to the Variable Info worksheet and name the new column **ESTCRATE**

### Page 7.15: Comparing Groups

- **Data → Filter**
- Make sure **Filter System Active** is checked, then enter "SML\_UNIT=1" into the **Filter Statements box**.
- Press the **Play** button
- **Analysis → Descriptive Statistics → Descriptive Statistics**
- Enter *CAP\_RATE* in the **Data Variables box** and *STORY* in the **Group 1 Variable box**
- Run the report
- Repeat the last three steps, inputting *NBHD*, *UND\_PARK*, *ELEVATOR*, AND *CONDITN* in the **Data Variables box** one at a time. Then repeat the whole process with a filter of SML\_UNIT=0