



## Regression Fundamentals

[www.BradfordSoftware.com/Redstone](http://www.BradfordSoftware.com/Redstone)



## Regression Overview

Property selection for the regression analysis includes ALL of the sales of like type, SFR or Condo, in the market area for the previous twelve months. Properties are not filtered relative to the subject property's size, age, number of stories or configuration. This is counter intuitive to most appraisers as appraisers are accustomed to narrowing a search to only include comps.

## Objective vs. Subjective Qualities within a Regression Analysis

Regression analysis does not take into account, condition, location, quality or views. It is advisable not to combine properties with substantially different qualities with other types of properties due to their having qualities not accounted for within the regression. Foreclosures, fixers, lake view properties, gated communities and exclusive communities may be removed from the analysis, unless these types of properties are representative of the market. Exclusive communities may include but are not limited to senior housing, low income housing or golf course development. In some circumstances perhaps when there is not a lot of activity in a market, the appraiser may go back 24-months (or further) instead of 12 months to obtain an adequate number of sales in order to complete an analysis on exclusive communities. Sample sizes can range from 40 to 300 properties. Normally, the larger the sample the more accurate the analysis.

## Using the Regression Analysis for the Subject Property

When the analysis is done, the subject property is not even a consideration. Basically, components of value are determined by using all the properties in the area to determine a value range for each component. In order to complete an analysis, variety not similarity is needed. Regression works best when the neighborhood is made up of various types of homes rather than similar types of home.

Once the regression analysis is complete, the value of the Subject Property is calculated using the values determined by the analysis.

## Some Limitations of Redstone and CompCruncher Regression Analyses

Rural areas with limited number of sales may need to cover a larger area than what would be used for selecting comparables. Often several communities are combined.

Manufactured homes, multi-family, income and commercial properties are outside the scope of the regression analysis used in Redstone and CompCruncher. Properties of this type require a different configuration of components of value to complete a viable analysis on properties of this type.

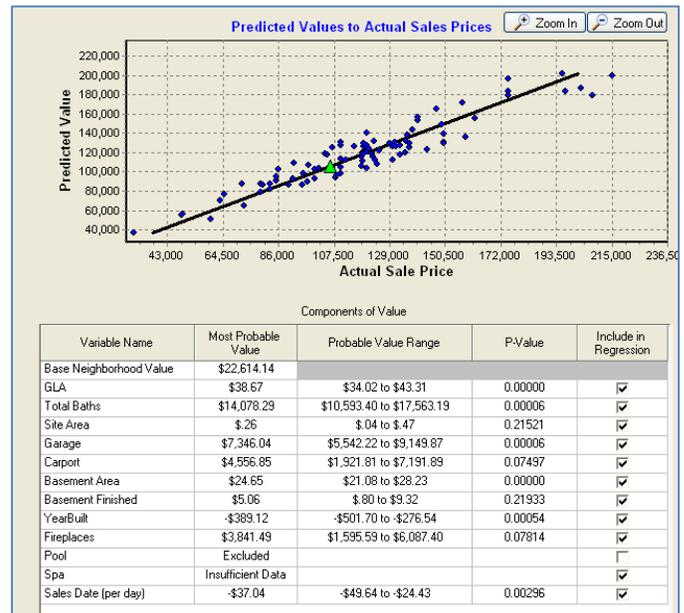
Regression analysis is not appropriate for all SFRs and condos in all situations. Properties located at the high or low end of a value range are much harder to evaluate using regression analysis. Acreage properties with additional attributes, such as, out-buildings are inappropriate for regression analysis.

## Utilizing Regression Analysis to Help Substantiate Adjustments for Components of Value

Regression analysis is a tool appraisers can use to assist them in calculating market adjustments. Regression analysis values are determined using a standard regression calculation. Redstone and CompCruncher apply a multi-linear regression to imported MLS data to calculate a value range for each Component of Value. A value range may not be calculated for one or more components of value due to insufficient data.

The available Components of Value within the Redstone and CompCruncher regression analysis include:

- Gross Living Area (per sq. ft.)
- Bathrooms (each)
- Site Area (per sq. ft.)
- Garage (per bay)
- Carport (per bay)
- Basement Area (per sq. ft.)
- Basement Finished (per sq. ft.)
- Year Built (per year and should be a negative adjustment)
- Fireplaces (each)
- Pool
- Spa
- Sales Date (per day on market, can be negative or positive adjustment)



In addition, there are other market factors that are not included or taken into account within a regression analysis. These factors/attributes include:

- Condition
- Location
- View
- Quality of construction
- Proximity to areas that increase or decrease the marketability of a given property  
i.e. busy streets, commercial property, schools and parks

Many factors contribute to the results obtained from the data:

- Variety of property configurations
- Accuracy
- Quantity of properties included in the sample
- Market variability

Note: When selecting the neighborhood for the analysis, select **ALL** properties of the same type, Single Family, Condominium, etc. The analysis is of the surrounding properties and is unrelated to the selection of comparable properties. DO NOT exclude properties based on the attributes of the subject property.

To obtain an adequate sample, a larger geographic area may be needed that is larger than would be considered appropriate for the selection of comparable properties. A larger sample is recommended. A larger sample size is more likely to provide adjustment ranges for more Components than a smaller sized sample.

Properties to consider removing from the sample include:

- Foreclosures and other non-arms-length sales may be removed from the sample if they are an insignificant part of the market.
- Properties significantly different from the other properties in the market due to: view, location or attribute (Gated Community, Senior Community and Golf Course Communities).
- Ranch properties and Farms are different than smaller lot properties and are difficult to compare using a regression analysis due to the complexity of their configurations.

Given these variables, the regression will calculate the PVR, Probable Value Range, for each Component included in the analysis. When evaluating p-scores, a value closer to 0.00000 is more reliable than a score closer to 0.99999. A p-value of <.05 means the probability that the recommended adjustment is random, is quite small. For each of the calculated Components, the significance of the Component of Value will range from:

- Very High      Less than 0.01
- High            0.05 through 0.0099
- Moderate      0.15 through 0.0499
- Low             0.3 through 0.1499
- Very low      0.99 through 0.2999
- Insufficient Data
- Excluded (due to MLS inaccuracies or negative value. Negative values may also be calculated for Components. It is up to the appraiser's discretion to either include or exclude the estimated value range.)

As the PVR for specific Component's is based on a mathematical calculation, a PVR may not appear rational. Individual Components PVR may seem unreasonable in comparison to the total value of the property. The calculations are not incorrect; rather it is a reflection of the variability of the data used in calculating the results. Aspects of the properties not included in the analysis may be hidden within the calculations. These include condition, location and etc. as previously listed.

When applying the results of a regression analysis to a specific valuation report:

- Consider using PVRs for Components with p-scores with a green rating. (Within the range)
- Yellow PVRs may provide guidance in determining adjustments
- Red regression PVRs are not as reliable based on the data provided
- Use "typical" adjustments for Components that are:
  - Excluded
  - Insufficient Data
  - High p-value
- Appraisers will develop a feel for typical PVRs upon completing numerous regression analyses within their market area where sufficient data is available to calculate PVRs with a satisfactory p-value.

- Unusually high or low adjustment values may be calculated through regression analysis. Though the calculations are correct mathematically, the appraiser may opt to exclude the component of value from the analysis. If used, the appraiser may state that the adjustment may be affected by other factors that result in an unusually high or low value for the specific component of value.

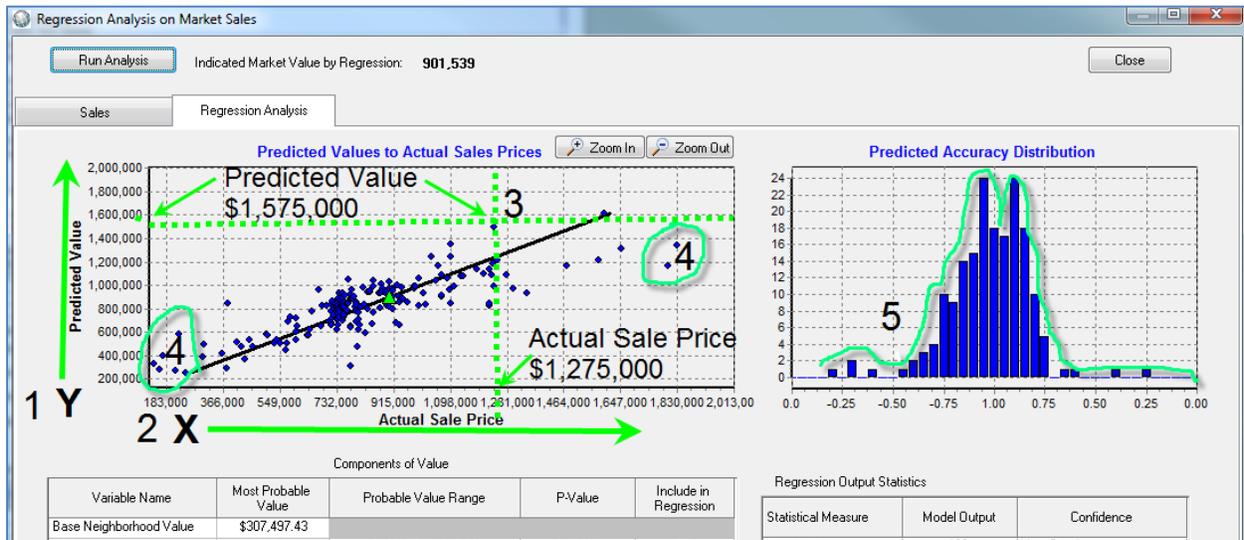
In conclusion, the PVRs calculated by a regression analysis provide the appraiser with useful information for setting adjustment amounts for Components. The regression analysis will not provide information for determining all adjustment due to the quantity and quality of available data. Some adjustments, such as condition, location, view and etc., are beyond the scope of available data obtained from your local MLS.

The appraiser will need to combine personal knowledge of the market, results of other relevant regression analyses, and the data provided within a given report to determine the adjustment amounts used for each property valuation.

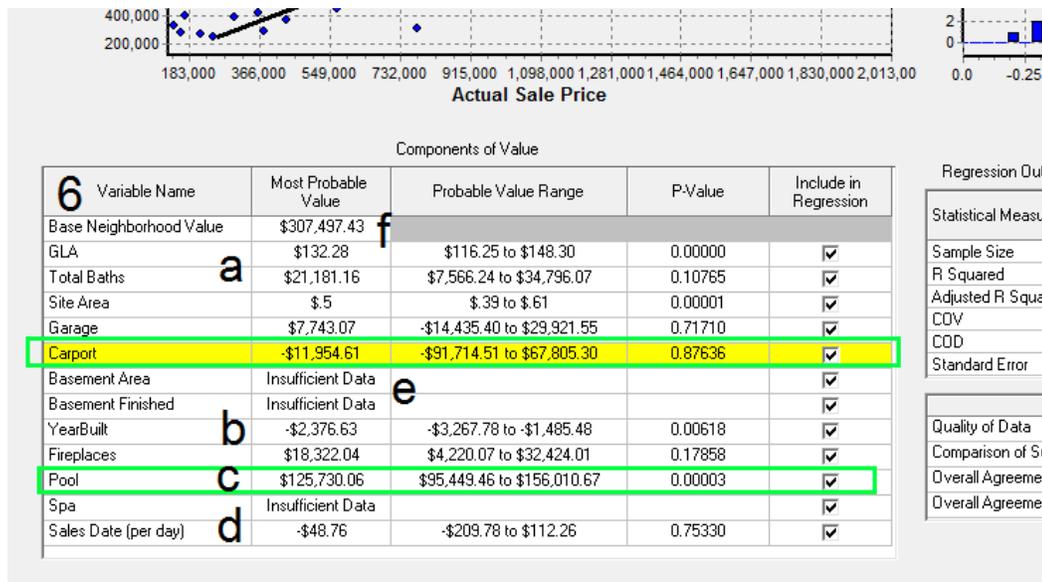
Though regression may not always provide an adjustment for every Component of Value it is a methodology the appraiser may use as evidence of applying mathematical analysis to the available market data to determine adjustments.

## Manually Processing the Regression Analysis

If we were to include all the sales in the regression and ran an analysis without the automatic removal of outliers we might see something that causes the analysis to look sloppy and causes the person viewing it to be confused. Here are some things that we can see:



1. The Values on the Y, perpendicular axis, represent the regression predicted value.
2. The Values on the X, horizontal axis, represent the actual sale price.
3. The point where each property meets sales price and predicted value becomes a point on the scatter plot. For example this point labeled #3, as shown by the dotted green lines:
  - a. Predicted Value \$1,575,000 / Actual Value \$1,275,000 = 80.95% Predicted Accuracy
4. In the right hand corner there are some properties that are way out to the right. There are also some lower priced properties clustered to the left.
5. The graph on the upper right shows the distribution. Notice that this raw data does not form a bell shaped curve. The larger the sample, the more apt you are to see a bell shaped curve. Smaller samples may look much different when first analyzed.



6. In the lower left we see components of value that are confusing.
  - a. Components of value should mostly be positive
  - b. Year built is a negative value as for each year of age takes away value while each bathroom, for example, adds value
  - c. There are some components of value that have highly exaggerated values and the appraiser may decide to remove them.
  - d. As this is a rising market, the Sales Date (per day) adjustment should not be a negative number. As the component often causes more confusion for the customer than helps them understand the analysis, many appraisers opt to remove this component of value, especially if it is going in the wrong direction.
  - e. Components of value with Insufficient Data should not be removed from the analysis as this is a significant fact.
  - f. In a deflated market(s) where homes included in the analysis are selling way below market value, you may see a negative Base Neighborhood value. This is acceptable as the addition of the combined values of the components of value accurately reflect the value of the included properties.

7

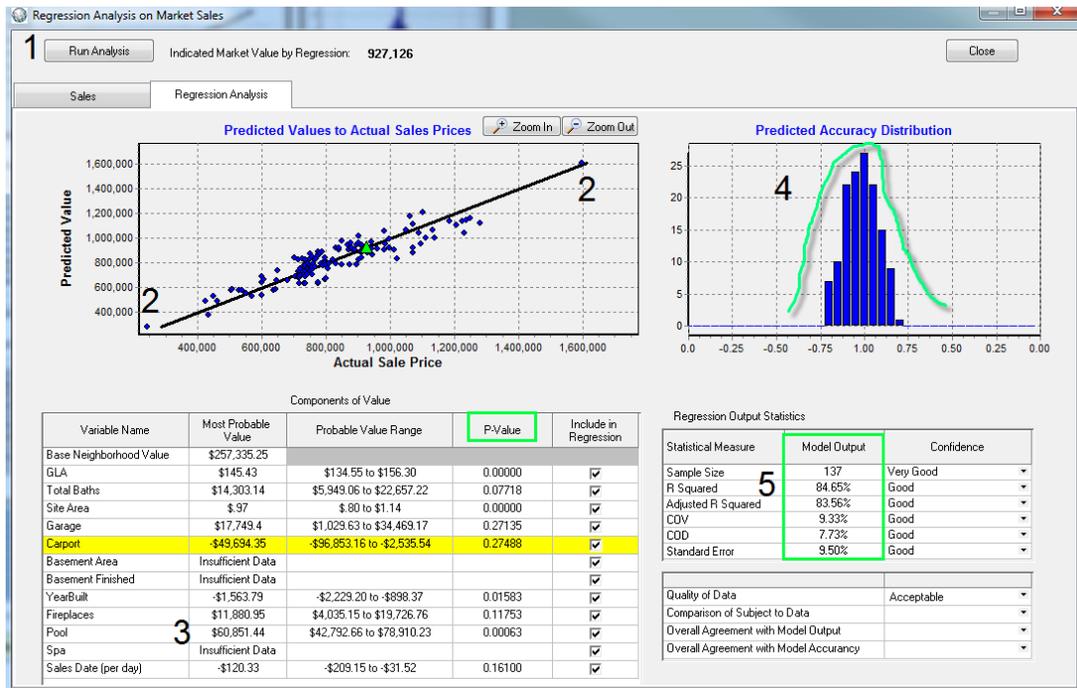
Statistical Measure	Model Output	Confidence
Sample Size	182	Very Good
R Squared	68.33%	Acceptable
Adjusted R Squared	66.68%	Poor
COV	22.04%	Poor
COD	15.23%	Poor
Standard Error	19.28%	Poor

Quality of Data	Acceptable
Comparison of Subject to Data	
Overall Agreement with Model Output	
Overall Agreement with Model Accuracy	

7. The lower right hand Statistical Measure is out of bounds. Minimal standards are:
- Sample size can be as low as ten if all other statistical minimums are missed. The reliability of the regression is increased as the sample size increases. Sample sizes of 60 to 200 work well. Too many and the data become hard to manage.
  - R-Squared and Adjusted R-Squared relate the variety of the configurations within the sample. This is a value that can only be increased through a larger sample size and the number of components used in the analysis. Removing too many components of value may reduce these numbers. 30% is the minimum. The higher the number towards 100% increases the Confidence level
  - Coefficient of Variance, Coefficient of Distribution and Standard Error result as the percentage amount decreases. 15% is considered an upper limit. COV = Coefficient of Variance COD = Coefficient of Distribution
  - The objective is to reflect the neighborhood, not to move all the Confidence ratings to Very Good.
  - As all the qualities that make up the value of the homes included in the analysis as compared to the subject property are not available for analysis, it is recommended that a 8% to 12% Standard Error is needed to produce a value range that provide for individual property differences. Two properties may have the same components of value, but may have a higher or lower value depending on condition, location, quality of construction and general desirability. These factors are not available within available data to be included in the regression analysis.

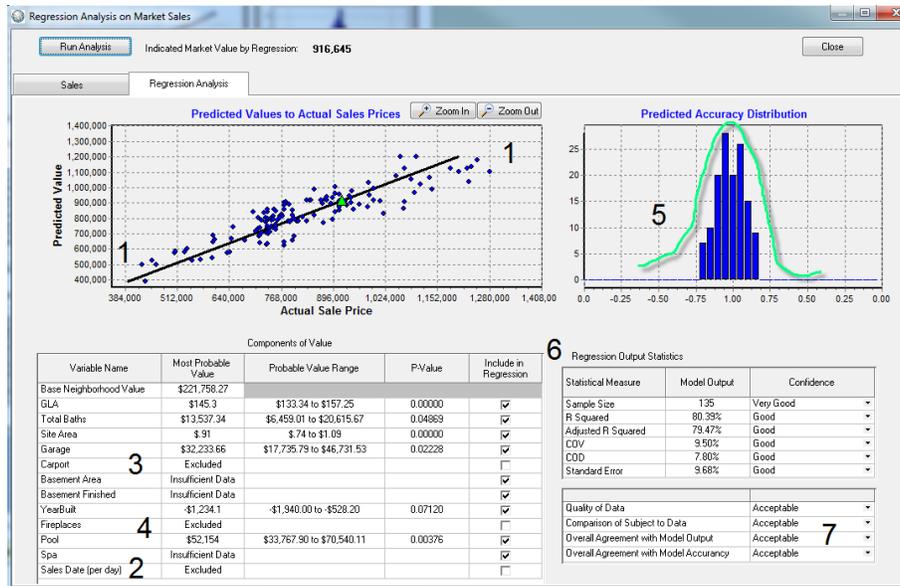
Here is a regression after pressing both the <Ctrl> and Run Analysis at the same time.



Applying the automated process to the regression helps to better refine the analysis but there are still remaining issues to address.

1. Most of the outliers were removed by pressing both the <Ctrl> and the Run Analysis at the same time.
2. Properties on the far extreme edges that are separated from the primary grouping may need to be removed. Do not remove a property simply because it is at the high or low end of the market as they help define the extent of the price range and help to increase the R-Squared and Adjusted R-Squared percentages.
3. These components of value are exaggerated values. This is caused as a result of mathematical computations that are reflective of the properties within the analysis. The P-Value can indicate those values that are less accurate. The smaller the number the higher the P-Value.
4. The Predicted Accuracy Distribution resembles a bell curve.
5. All of the Statistical Measures are within acceptable ranges.

Below is the completed Regression Analysis on Market Sales



1. The properties on the far right and left on the graph have been removed
2. The Sale Date adjustment was removed first and the Analysis recalculated (It is best to remove components of value starting at the bottom and working your way up, rerunning the analysis with the removal of each component.)
3. Negative Carport value is removed (Some regression purist recommend leaving them in the analysis. Explaining this to the customer may be a challenge though it is mathematically appropriate.)
4. Fireplace was removed due to its exaggerated value of \$11,880.95
5. The Predicted Accuracy Distribution approximates a bell curve.
6. The Statistical Measures are within acceptable ranges.
  - Sample size is large (Above 30 is recommended and 10 is minimum)
  - R-Squared and Adjusted R-Squared is a high percentage (30% is a minimum)
  - COV, COD and Standard Error are below the maximum of 15%
  - COV = Coefficient of Variance COD = Coefficient of Distribution
7. Don't forget to set the four Appraiser valuations of the regression in the lower right hand corner of the screen.
8. Space is provided on the Regression worksheet for additional appraiser comments.

**For additional assistance, contact  
Technical Support at 1-800-622-8727 x313**

## DEFINITION OF STATISTICAL TERMS

### Definitions

**COLLINEARITY:** In statistics, **collinearity** is a phenomenon in which two or more predictor variables in a multiple regression model are highly correlated. For example, bedrooms and gross living area, one can be linearly predicted from the others with a non-trivial degree of accuracy.

**DEFINITION OF R-SQUARED (Coefficient of Determination):** In regression analysis, a measure of the strength of the relationship between the independent variables and the dependent variable. The measure ranges from 0 to 1 - the higher the number, the stronger the relationship (0 would indicate no relationship)

**DEFINITION OF ADJUSTED R-SQUARED:** R-Squared can overstate the goodness of fit in a model when insignificant variables are included, or the number of variables R-Squared appropriately.

**DEFINITION OF P-VALUE:** A particular statistic that measures the significance of a variable in a regression analysis. The statistic is important in inferential statistics for hypothesis testing in regression analysis. The P-Value is a measure of the randomness of a component used in the regression analysis.

**DEFINITION OF COV (Coefficient of Variation):** A standard statistical measure of the relative dispersion of the sample data about the mean of the data; the standard deviation expressed as a percentage of the mean.

**DEFINITION OF COD (Coefficient of Dispersion):** The average deviation of a group of numbers from the median expressed as a percentage of the median.

**DEFINITION OF STANDARD ERROR:** A measure of the precision of a measure of central tendency, the smaller the standard error, the more reliable the measure of standard tendency. In regression analysis, the standard deviation of a regression coefficient, the smaller the standard error relative to the coefficient, the more reliable the coefficient.

**DEFINITION OF SALES RATIO:** A measure of model accuracy that divides predicted values by sales prices. The closer to 1 that a sales ratio is, the better the model's predictive capabilities

**DEFINITION OF STANDARD DEVIATION:** The statistic calculated from a set number by subtracting the mean from each value and squaring the remainders, adding together all of the squares, dividing by the size of the sample less one, and taking the square root of the result. When the data are normally distributed, one can calculate the percentage of observations within any number of standard deviations of the mean from normal probability tables.

### Sources Documents

The Collateral Valuation Report (CVR) has been designed in conformance with all available technology, data and statistical processes, generally accepted to represent the state of industry, including:

#### Uniform Standards of Professional Appraisal Practice (USPAP):

Standard 1

Standard 2

Since specialized statistical and mass appraisal information is contained with Standard 6 and Advisory Opinion 18, these sources have also been considered in tandem with the Development and Reporting standards contained within Standard 1 and Standard 2. It is expressly understood that the Collateral Valuation Report is a summary appraisal report performed under the guidance of Standards 1 and 2 as noted above.

**Joint Industry Task Force on Automated Valuation Models:**

## Standards and Testing Guidelines

These standards and guidelines are instructive in the method of testing accuracy and identifies the statistics and outcome guidelines that can be relied upon in performance of statistical analysis.

**International Association of Assessing Officers:**

## Standard on Ratio Studies

## Mass Appraisal of Real Property

## Standard on Automated Valuation Models

**Appraisal Institute:**

## A Guide to Appraisal Valuation Modeling

## Practical Applications in Appraisal Valuation Modeling and Design

## The 13th Edition of the Appraisal of Real Estate

## Visual Valuation: Implementing Valuation Modeling and Geographic Information Systems

These texts form the body of knowledge that helps provide an understanding of the modeling process and the use of statistics in real estate.

**The Modeling Process**

An acceptable model will have both reasonable coefficients and satisfactory outcome statistics. The appraiser has been trained in a manner sufficient to understand the various statistical measures outlined in this report. The statistical measures defined within this analysis allow the appraiser to understand the data and draw certain conclusions based on the accuracy of the data, the amount and quality of the data, and the measures of statistical significance and accuracy of the analysis applied.

**Competence**

The appraiser completing the Collateral Valuation Report asserts that they have undergone sufficient training, and further, have an understanding of the statistical measures underlying the regression component of the process to generally understand the method and manner of analysis. The appraiser does not assert that they are statisticians. They are, however, aware of the basic guidelines pertaining to the use of CVR application as a tool to analyze small market datasets, and as such, are capable of understanding the analysis and methodology in a manner sufficient to render a credible estimate of value in tandem with the other data and analysis present in the report.

The final value conclusion is the appraiser's own, and is based on the appraiser's knowledge and experience in the field of appraisal. The data and analysis in this report, whether through direct information or through derived statistical information, aids the appraiser in understanding the dynamics of the neighborhood and market area.

**For additional assistance, contact  
Technical Support at 1-800-622-8727 x313**



# Support Your Adjustments Using Accepted Methodologies

Appraisers are being tasked with providing even greater detail and analytical support for their adjustments and comparable selection. Your local market experience and knowledge are invaluable to providing a thorough analysis, but lenders are demanding greater transparency and repeatable results using accepted methodologies — Redstone™ was developed specifically to address these challenges.

## Regression Analysis is at the Heart of Redstone

An accepted methodology for calculating your adjustments is Regression Analysis. It's the same methodology employed by most AVM's in calculating adjustments and values. Redstone's regression tool is intuitive and the appraiser is in control — this is not an AVM that is simply digesting your data. Redstone will also analytically support your comparables selection with three ranking scores that are prioritized by the appraiser to provide a final comp selection ranking that's repeatable and transparent.

## Redstone Adds Analytical Credibility for a Higher Caliber Report

Appraisers are categorized by the quality of their work and this directly affects their level of compensation. Redstone's advanced analytics and detailed addendum will provide the transparency and recognized methodologies that reviewers are looking for in a quality report. Unsupported adjustments are the fast track to censure and while your market experience may be accurate, it's not accepted as evidence to support an adjustment. Redstone gives you the power to analyze market data and combined with your market expertise, it helps support and defend your appraisal. The result is a stronger and more transparent appraisal.

### Try Redstone Appraisal Analytics — Special 10 Analyses for \$10

#### 10 for \$10 Special

This is a great starter pack to see if Redstone is right for you. After you have tried the sample properties that are built into Redstone, work with Redstone in your local market with your local data. This discount will give you 10 analyses to try Redstone with data you know and will be using in your reports. Purchase directly from the AppraisalWorld.com store or call your sales representative at 800-622-8727.



[www.BradfordSoftware.com/Redstone](http://www.BradfordSoftware.com/Redstone)

