## STATISTICS AND THE APPRAISAL PROCESS

### INTRODUCTION

Statistics offer a way for the appraiser to quantify many of the heretofore qualitative decisions used in the mass appraisal process. Statistics can be applied to interpret available data and to support value conclusions. Refer to Chapter 1 for information on Assessment Uniformity.

Level of assessment (appraisal) in relation to market is usually expressed as a ratio of assessed (appraised) value to sale (market) value. Three common statistical (appraisal) measures of this ratio are Measures of Central Tendency, the mean, median and mode.

Examples follow.

### UNWEIGHTED MEAN (AVERAGE)

This measure is found by dividing the sum of all variates in a population by the number of variates. The following list of sales illustrates how to compute different means (averages).

| OBSERVATION NUMBER | SALEPRICE | APPRAISED VALUE | SALES RATIO |
|---|---|---|---|
| 1 | $22,600 | $21,500 | 95 % |
| 2 | 31,000 | 28,600 | 92 |
| 3 | 37,800 | 34,000 | 90 |
| 4 | 38,400 | 33,000 | 86 |
| 5 | 34,300 | 29,500 | 86 |
| 6 | 20,000 | 16,000 | 80 |
| 7 | 13,000 | 9,800 | 75 |
| 8 | 18,700 | 13,500 | 72 |
| 9 | 26,900 | 17,200 | 64 |
| 10 | 40,800 | 24,500 | 60 |
| | $283,500 | $227,600 | 800 % |

Mean Sale Ratio = 800/10 = 80%

Mean Appraised Value = $227,600/10 = $22,760

Mean Sales Price = $283,500/10 = $28,350

**WEIGHTED MEAN**

The WEIGHTED MEAN is an aggregate ratio, the sum of all appraised values divided by the sum of all sales values.

**MEDIAN**

The MEDIAN is the measure of central tendency which is least influenced by extreme values as it is based upon position rather than on level. It is the value half-way from either end of a list of values when the list is arrayed in ascending (or descending) order. If the list contains an odd number of sales then the median is the middle value in the list. However, if there are even numbers of sales in the list then it is the average of the two values on either side of the theoretical mid point in the list. Using our example it is:

MEDIAN = (TOTAL NUMBER OF SALES + 1) / 2 or (10 + 1) / 2 = 5.5th item in the list

| That is in our list: | Sales | Sales Ratio |
|---|---|---|
| | 1 | 95% |
| | 2 | 92 |
| | 3 | 90 |
| | 4 | 86 |
| | 5 | 86 |
| Median 5.5 Sales----------> | | |
| | 6 | 80 |
| | 7 | 75 |
| | 8 | 72 |
| | 9 | 64 |
| | 10 | 60 |

The median is, therefore, halfway between the ratio 86 and 80 or:

MEDIAN = (86 + 80) / 2 = 166 / 2 = 83%

**MODE**

The mode is the variate or attribute that appears most frequently in a population. In the example above the mode of sales ratios would be 86% (occurs 2 times).

**RANGE**

The range is the difference between the highest and lowest variates. The range may be reported as the minimum and maximum values themselves. In the preceding example, the range (for the sales ratios) is:

35% or from 60% to 95%

As a general statement it is not useful in analysis due to dependence on extreme values.

**COEFFICIENT OF DISPERSION (COD)**
The COD is the average absolute deviation expressed as a percentage of the level of assessment, and is calculated by dividing the average absolute deviation by the median. The average deviation is calculated by subtracting the median from each ratio, summing the absolute values of the computed differences, and dividing that there is a large variance between the highest ratios and the lowest ratios compared to the median. A low COD represents conformity amongst the county. In a reappraisal year, it is expected the COD would be lower than in non-reappraisal years.

**MEAN (AVERAGE) DEVIATION**
This is a measure of how much the actual values of a population or sample deviate from the mean. It is the mean of the sum of the absolute differences of each of the variates from the mean of the variates.

Example:

SALES RATIO (x) – MEAN (Sales Ratio) ($\bar{x}$) = ABSOLUTE DEVIATION BETWEEN EACH SALES RATIO & MEAN SALES RATIO [ $x - \bar{x}$ ]

| **x** | | $\bar{x}$ | | |
|---|---|---|---|---|
| 95 | – | 80 | = | 15 |
| 92 | – | 80 | = | 12 |
| 90 | – | 80 | = | 10 |
| 86 | – | 80 | = | 6 |
| 86 | – | 80 | = | 6 |
| 80 | – | 80 | = | 0 |
| 75 | – | 80 | = | 5 |
| 72 | – | 80 | = | 8 |
| 64 | – | 80 | = | 16 |
| 60 | – | 80 | = | 20 |
| | | | | 98 Total deviation from mean $\sum [x - \bar{x}]$ |

Mean Deviation $\frac{\sum [x - \bar{x}]}{N} = \frac{98}{10} = 9.8\%$

This ratio expresses the mean (average) deviation of the individual ratios in the population from the mean is 9.8%. It is influenced by extremes in the mean.

**STANDARD DEVIATION**
To overcome the handicaps of the mean deviation, the standard deviation is used. It is a numerical measure of the degree of dispersion, variability, or non-homogeneity of the data to which it is applied.

In calculation, it is similar to the average deviation but differs in its method of averaging differences from the mean. It does this by squaring each difference and eventually summing all squared differences averaging them and taking the square root thereof giving an "average deviation" from the mean.

STANDARD DEVIATION = $\sqrt{\frac{\sum(X-U)^2}{N}}$ or $\sqrt{\frac{\sum(x-u)}{N-1}}$ Where u = "mu" (arithmetic mean)

$$\sqrt{\frac{\text{Sum of the individual differences squared}}{\text{Number of observations}}}$$

The second formula using N-1 is most often used when dealing with sample data and is used in our sales ratio reports.

In this example, using sales ratios it would be:

| Observation | X | (X-u) | $(X-u)^2$ |
|---|---|---|---|
| 1 | 95% | 15 | 225 |
| 2 | 92 | 12 | 144 |
| 3 | 90 | 10 | 100 |
| 4 | 86 | 6 | 36 |
| 5 | 86 | 6 | 36 |
| 6 | 80 | 0 | 0 |
| 7 | 75 | 5 | 25 |
| 8 | 72 | 8 | 64 |
| 9 | 64 | 16 | 256 |
| 10 | 60 | 20 | 400 |

X = 800% $(X-u)^2$ = 1286

Arithmetic Mean (u) Sales Ratio = 800 / 10 = 80%

Hence: SD = $\sqrt{\frac{(X-u)^2}{N}}$ OR SD = $\sqrt{\frac{\sum(X-u)^2}{N-1}}$

$\sqrt{\frac{1286}{10}}$ = $\sqrt{\frac{1286}{10-1}}$

$\sqrt{128.6}$ = $\sqrt{142.89}$

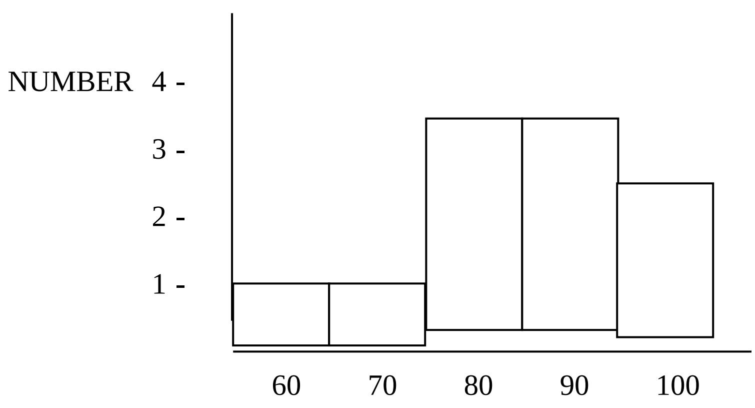
$\sqrt{11.34}$ = $\sqrt{11.95}$

The standard deviation is useful in that it is logical mathematically and may be used satisfactorily in further calculations. This is its outstanding superiority over the other measures of dispersion.

**FREQUENCY DISTRIBUTIONS**

Frequency distributions are an arrangement of numerical data according to size or magnitude. Distributions are normally presented as tables or graphs. The following table and graph is taken from the example:

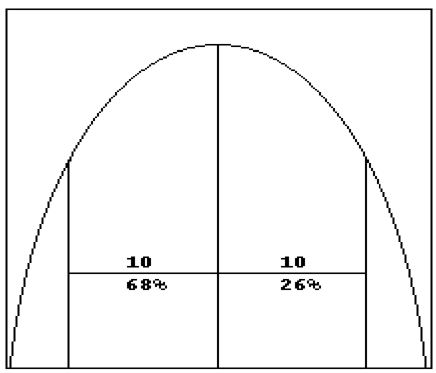
| SALES RATIO CLASS INTERVAL | NUMBER OF OCCURENCES |
|---|---|
| 91 - 100 | 2 |
| 81 - 90 | 3 |
| 71 - 80 | 3 |
| 61 - 70 | 1 |
| 51 - 60 | 1 |
| TOTAL: | **10** |




**SALES RATIOS**

When describing observations, numbers [mean, median, mode, standard deviation, average deviation, etc.] are used to give a picture of what frequency distribution would look like if drawn on a graph.

A particularly shaped distribution is the one from which we depart when trying to visualize the shape of a distribution when given such statistics as the mean, median and mode for information. The reference point is what is called the "NORMAL DISTRIBUTION". It has some particular features by which it is characterized and referred to. This is what it looks like:




"Normal" Distribution Showing the Percentage of the Area Included Within One Standard Deviation Measured Both Plus and Minus About the Arithmetic Mean.

The MEAN, MEDIAN, and MODE are all equal. It also possesses some traits which make it statistically useful in making decisions about differences in distributions.

One of these properties is that one may determine what percent of the observations lie within; one, two, or three times the calculated standard deviation by using pre-computed tables. (In fact, any fractional part of the standard deviation may also be used.)

It is useful in measuring uniformity of values. For instance, a set of sales with a mean of 87% and a Standard Deviation of 10%, concludes that 95.46% of all sales would fall between the limits of 75.46% and 115.46%. Extrapolating that sales represent the rest of the parcels in the county distribute themselves in relation to the market values of the parcels.

It is important to remember: (1) The distribution must be normal or approximately so for this to be true and (2) if there is ever a source of disagreement, sales ratio studies are primary.

**RELATIVE MEASURE OF VARIATION**
Statistical tools are the relative measures, which relate back to the mean or median in discussing the degree of variance in a set of observations. Three common ones are:

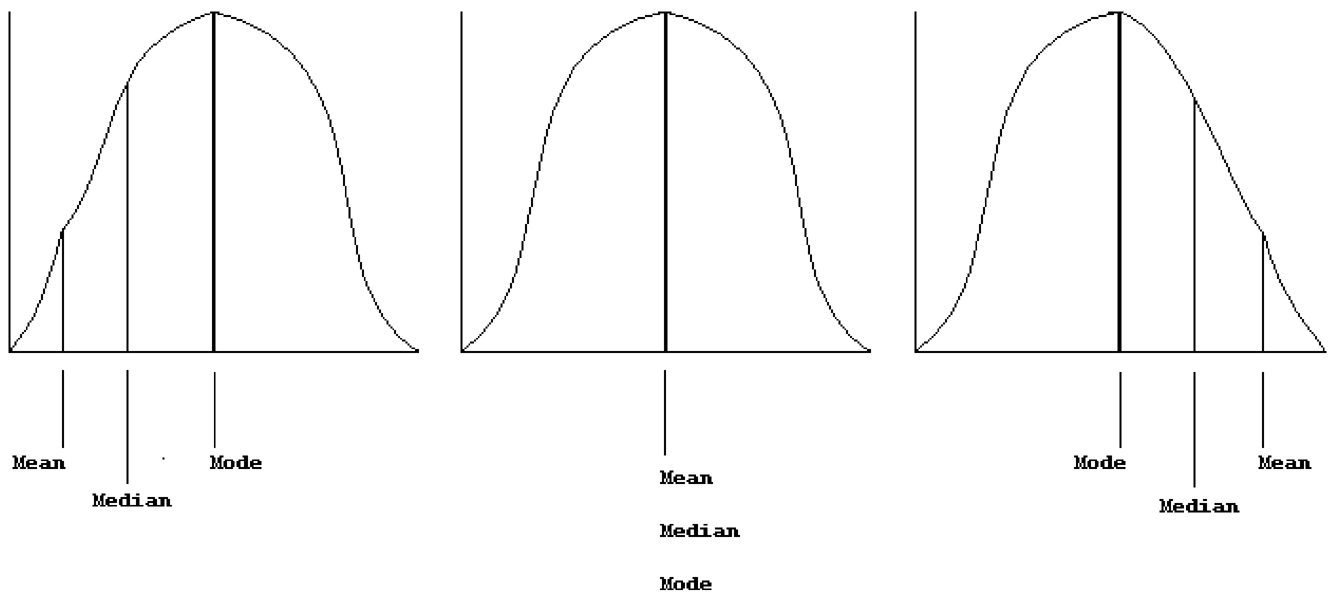
| | | |
|---|---|---|
| $\frac{\text{AVERAGE DEVIATION ABOUT THE MEAN}}{\text{MEAN}}$ X 100 | = | Coefficient of dispersion of the average deviation |
| $\frac{\text{STANDARD DEVIATION}}{\text{MEAN}}$ X 100 | = | Coefficient of dispersion of the standard deviation |
| STANDARD DEVIATION ABOUT THE MEDIAN X 100 | = | Coefficient of dispersion of the median deviation |

The last two yield the most useful statistic in that the standard deviation is significant in appraising in relationship to the level as there are few who would want a ratio to go consistently over 100% (which is one use of the standard deviation) or whom would want a mean of 70% with a relative error of 35% on 68% of all parcels.

**SHAPE**
How do you describe the shape of a distribution? We have used the mean, median, mode, average and standard deviation. We also would like to be able to tell the extent to which our values were consistently biased either high or low. The statistics measuring this are the coefficients of skewness. That is, a measure of the degree to which the distribution departs from the normal distribution.

There are three, more or less, classic shapes a distribution may take:

SKEWED LEFT | NORMAL | SKEWED RIGHT




Skewness is a term for the degree of distortion from symmetry exhibited by a frequency distribution. This means if graphed the sales ratios it could be expected that all errors should be random and hence symmetrical and not biased either low or high for certain properties. This can be checked by using the common measures of degree of skewness.

$$SK_1 = \frac{3\ (\text{MEAN-MODE})}{\text{STANDARD DEVIATION}}$$

Note: (Pearson's Coefficient of Skewness)

And

$$SK_2 = \frac{(Q3 - \text{MEDIAN}) - (\text{MEDIAN} - Q1)}{(Q3 - Q1)}$$

The second measure uses a "QUARTILE" which is something like the median (in fact, the median is the Q2 or second quartile or quarter, EG 50% of the way through the list, item) but is the item 25% (Q1) down the list and the 75% (Q3) item down the list of ordered observations and may be determined much as is the median.