

Measures of Variability

Table 2.11: Annual Payouts for Two Different Investment Funds



Measures of Variability

Range Variance

Standard Deviation Coefficient of Variation

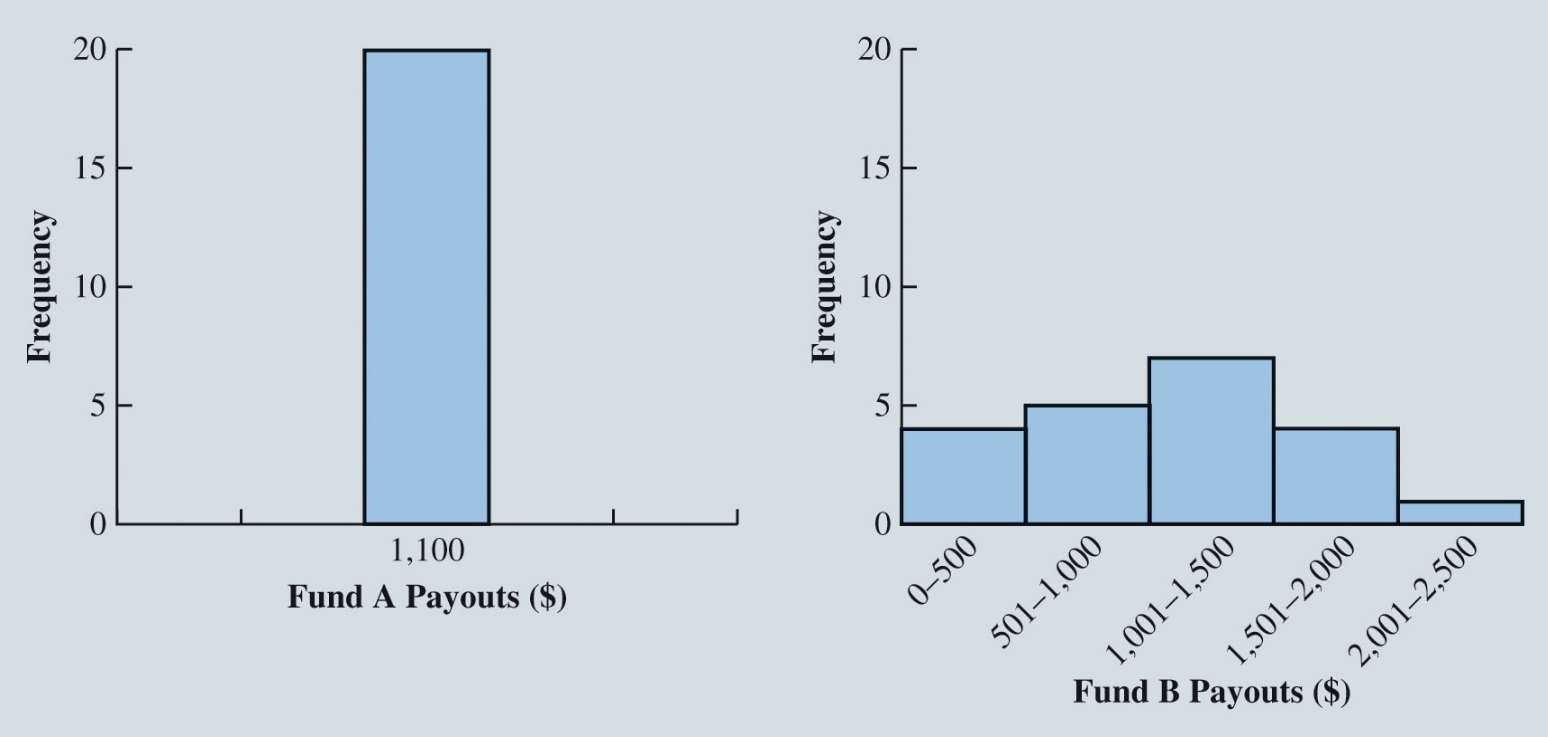
|  |  |  |
| --- | --- | --- |
| **Year** | **Fund A ($)** | **Fund B ($)** |
| 1 | 1,100 | 700 |
| 2 | 1,100 | 2,500 |
| 3 | 1,100 | 1,200 |
| 4 | 1,100 | 1,550 |
| 5 | 1,100 | 1,300 |
| 6 | 1,100 | 800 |
| 7 | 1,100 | 300 |
| 8 | 1,100 | 1,600 |
| 9 | 1,100 | 1,500 |
| 10 | 1,100 | 350 |
| 11 | 1,100 | 460 |



Measures of Variability

Table 2.11: Annual Payouts for Two Different Investment Funds (cont.)

|  |  |  |
| --- | --- | --- |
| **Year** | **Fund A ($)** | **Fund B ($)** |
| 12 | 1,100 | 890 |
| 13 | 1,100 | 1,050 |
| 14 | 1,100 | 800 |
| 15 | 1,100 | 1,150 |
| 16 | 1,100 | 1,200 |
| 17 | 1,100 | 1,800 |
| 18 | 1,100 | 100 |
| 19 | 1,100 | 1,750 |
| 20 | 1,100 | 1,000 |
| Mean | 1,100 | 1,100 |



Measures of Variability

Figure 2.18: Histograms for Payouts of Past 20 Years from Fund A and Fund B



Measures of Variability

Range:

 The **range** can be found by subtracting the smallest value from the largest value in a data set.

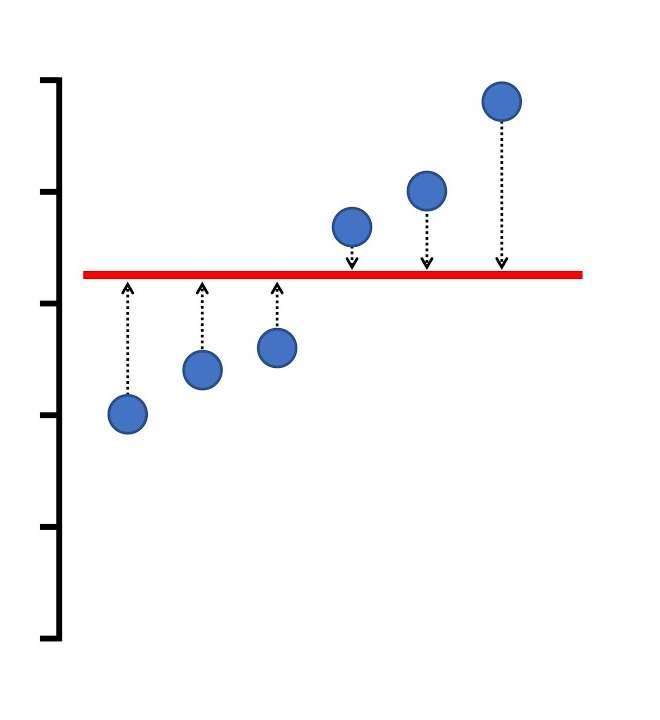
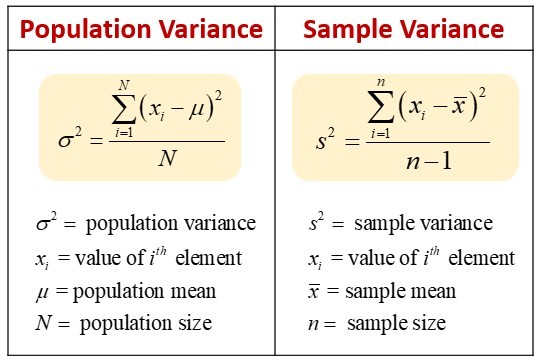
 Illustration: Consider the data on home sales in a Cincinnati, Ohio, suburb.

 Largest home sales price: $456,250. Range = Largest value – *S*mallest value

 Smallest home sales price: $108,000. = $456,250 – $108,000

= $348,250

 Drawback: Range is based on only two of the observations and thus is highly influenced by extreme values.

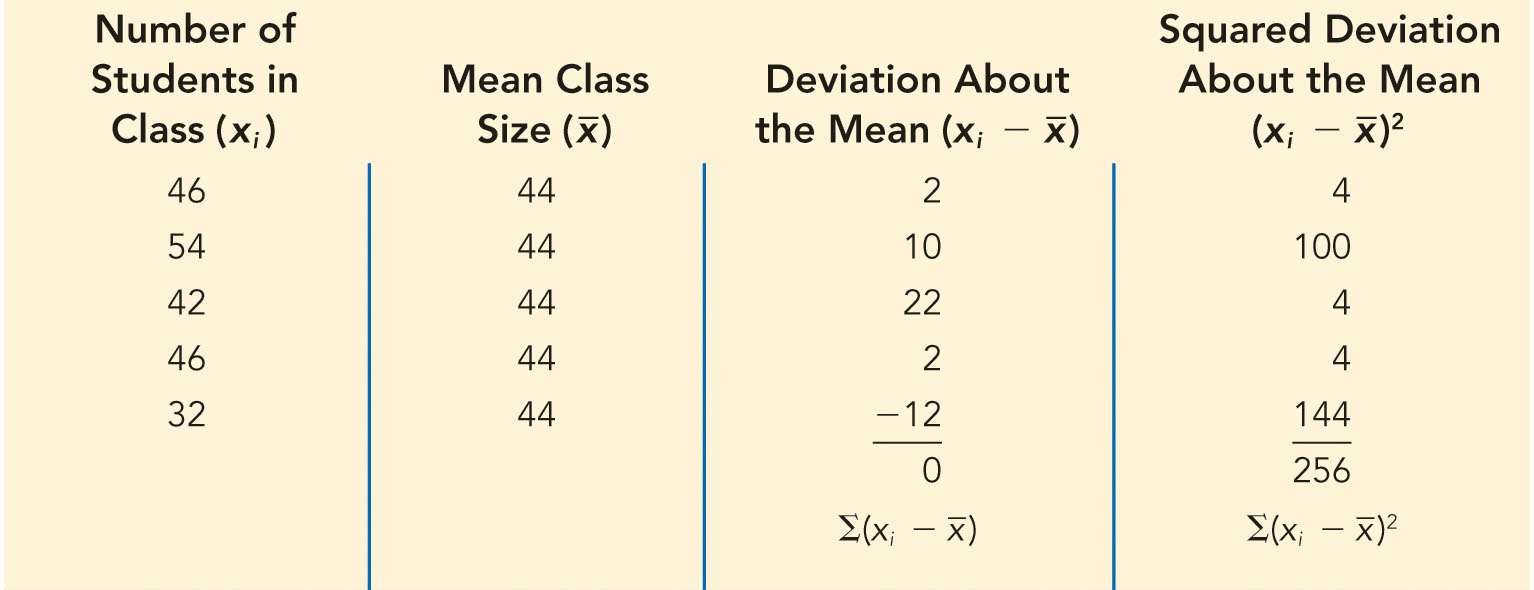


Measures of Variability

 Variance:

 is a measure of variability that utilizes all the data.

 It is based on the deviation about the mean, which is the difference between the value of each observation



Measures of Variability

Table 2.12: Computation of Deviations and Squared Deviations About the Mean for the Class Size Data

 Computation of Sample Variance:

𝑠2 =

Σ 𝑥i − 𝑥 2

𝑛 − 1

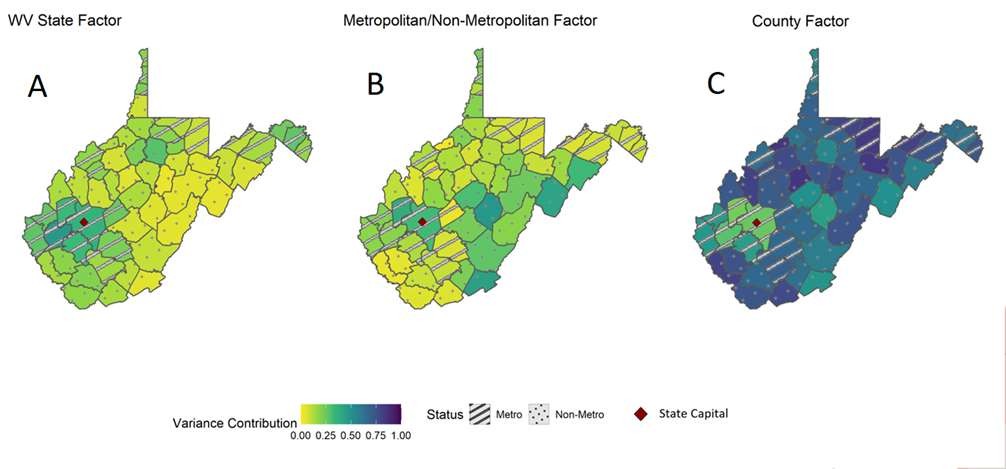
= = 64 students 2

256

4



Variation in the Change in Labor Force Participation Rates



Variation in the Change in Labor Force Participation Rates



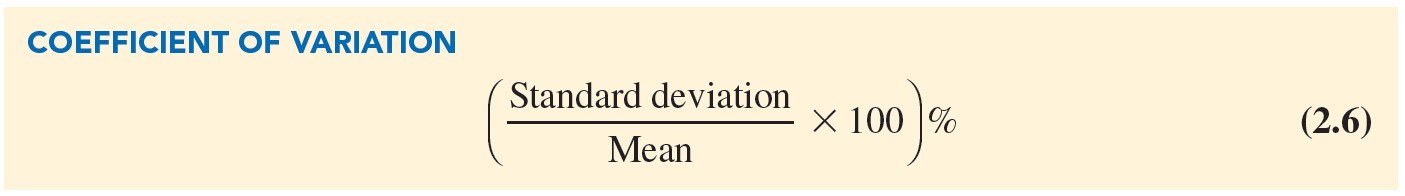
Measures of Variability

 **Standard deviation:**

is the positive square root of the variance.

 Measured in the same units as the original data.

 For population, *σ* = *σ* 2 .



Measures of Variability

The **coefficient of variation**

is a descriptive statistic that indicates how large the standard deviation is relative to the mean.

Expressed as a percentage.



Measures of Variability

Illustration:

 Consider the class size data:

46 54 42 46 32

 Mean,

*x* = 44.

 Coefficient of variation =

( 8 





 44

100 % = 18.2%.



 Standard deviation, *s* = 8 (students).

The coefficient of variation tells that the sample standard deviation is 18.2% of the value of the sample mean.

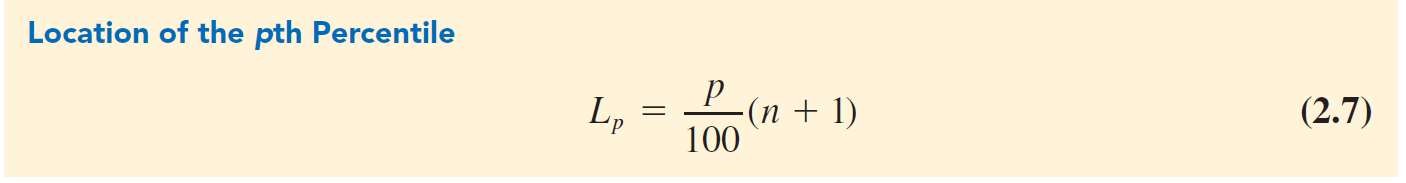


Analyzing Distributions

Percentiles Empirical Rule

Quartiles Identifying Outliers

z-Scores Boxplots



Analyzing Distributions

 A **percentile**

 is the value of a variable at which a specified (approximate) percentage of observations are below that value.

 The *p*th percentile tells us the point in the data where:

 Approximately *p* percent of the observations have values less than the *p*th percentile.

 Approximately:

(100 – *p*) percent of the observations have values greater than the *p*th percentile.



Analyzing Distributions

Illustration:

 To determine the 85th percentile for the home sales data:

 Arrange the data in ascending order:

2. Compute

*L* =

85

*p*

100

(*n* + 1) =  (12 + 1) = 11.05.

( 85 

 100 

3. The interpretation of *L*85 = 1 1 .0 5

is that the 85th percentile is 5% of the way

between the value in position 11 and value in position 12.

|  |  |  |  |
| --- | --- | --- | --- |
| 108,000 | 138,000 | 138,000 | 142,000 186,000 199,500 |
| 208,000 | 254,000 | 254,000 | 257,500 298,000 456,250 |



Analyzing Distributions

Illustration (cont.):

 To determine the 85th percentile for the home sales data:

 The value in the 11th position is 298,000.

 The value in the 12th position is 456,250.

 $305,912.50 represents the 85th percentile of the home sales data:

85th percentile = 298,000 + 0.05(456,250 – 298,000)

= 298,000 + 0.05(158,250)

= 305,912.50



Analyzing Distributions

**Quartiles:** When the data is divided into four equal parts:

 Each part contains approximately 25% of the observations.

 Division points are referred to as quartiles.

*Q*1 = first quartile, or 25th percentile.

*Q*2 = second quartile, or 50th percentile (also the median).

*Q*3 = third quartile or 75th percentile.

 The difference between the third and first quartiles is often referred to as the **interquartile range**, or IQR.



Analyzing Distributions

 The **z-score:**

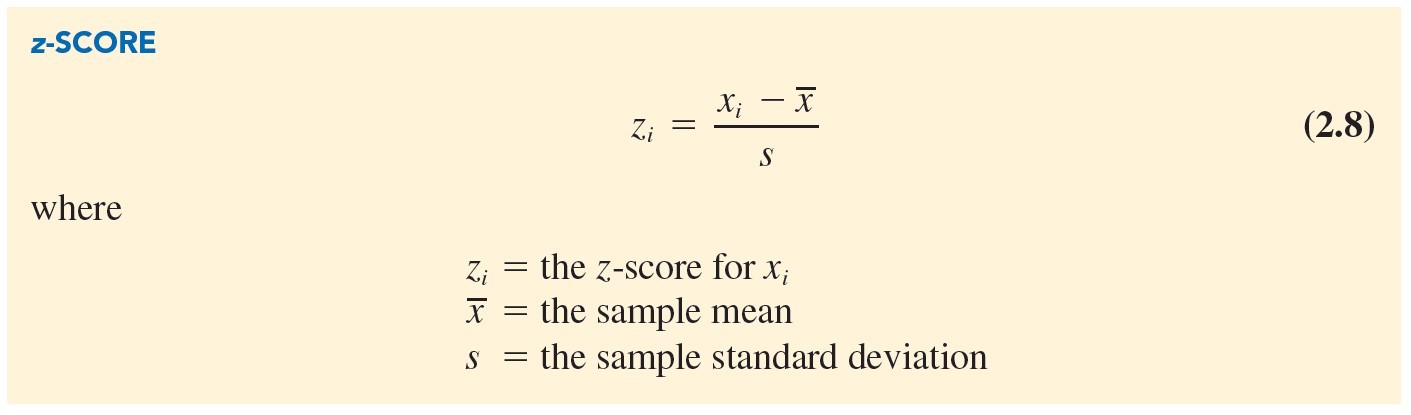
 measures the relative location of a value in the data set.

 Helps to determine how far a particular value is from the mean relative to the data set’s standard deviation.

 Often called the standardized value.

 The z -score can be interpreted as the number of standard deviations

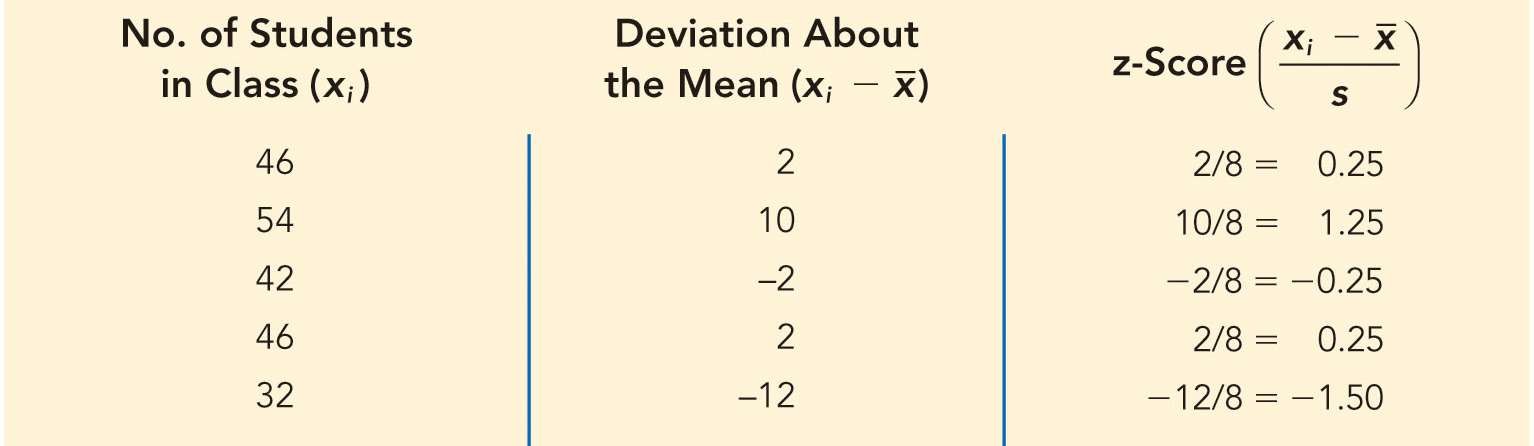
𝑥\_𝑖 is from the mean 𝑥 ̅.



Analyzing Distributions

 z-Scores (cont.):

If *x*1 , *x*2 , , *xn* is a sample of *n* observations:



Analyzing Distributions

Table 2.13: *z*-Scores for the Class Size Data

For class size data, *x* = 44 and *s* = 8.

For observations with a value > mean, *z*-score > 0. For observations with a value < mean, *z*-score < 0.



Analyzing Distributions

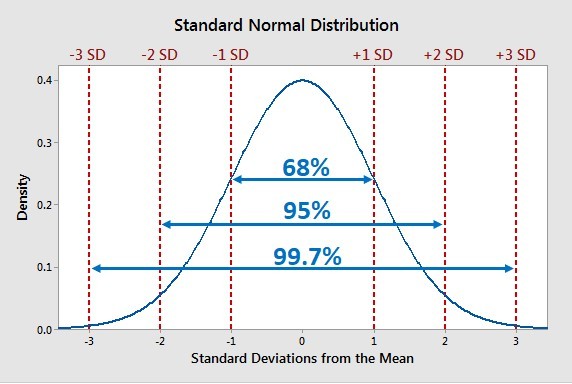
Empirical Rule:

 For data having a bell-shaped distribution:

 Approximately 68% of the data values will be within 1 standard deviation.

 Approximately 95% of the data values will be within 2 standard deviations.

 Almost all the data (99.7%) values will be within 3 standard deviations.



Analyzing Distributions

A Symmetric Bell-Shaped Distribution

 The height of adult males in the United States

 Mean 69.5 inches and

 Standard deviation of 3 inches.

 1 SD -> 68% of U.S. males are between 66.5 and 72.5 in tall

 2 SD -> 95% of U.S. males are between 63.5 and 75.5 in tall

 3 SD -> 99.7% of U.S. males are between 60.5 and 78.5 in tall



Analyzing Distributions

**Outliers**:

Extreme values in a data set.

 They can be identified using standardized values (*z*-scores).

 Any data value with a *z-*score less than –3 or greater than +3 is an outlier.

 Such data values can then be reviewed to determine their accuracy and whether they belong in the data set.



Analyzing Distributions

**Outliers:**

 Data value incorrectly recorded;

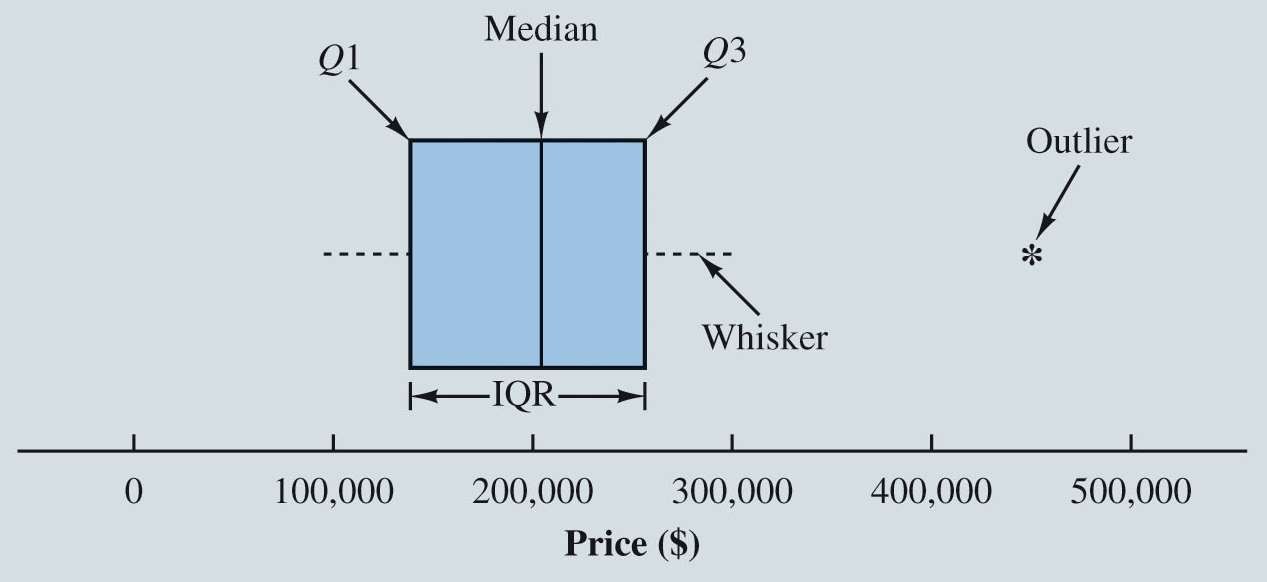
 Correct before further analysis.

 Data value incorrectly included

 It can be removed.

 Unusual data value that has been recorded correctly

 The observation should remain.



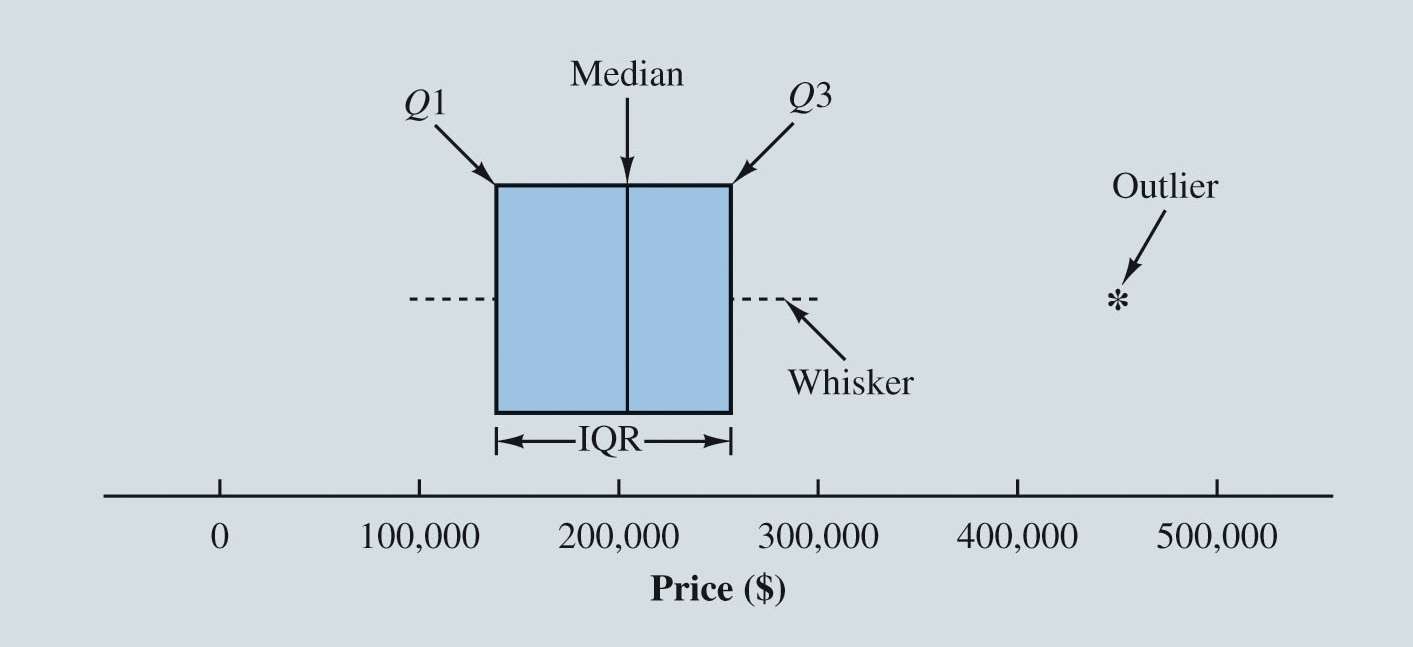
Analyzing Distributions

 A **boxplot:**

 is a graphical summary of the distribution of data.

 Developed from the quartiles for a data set.

Figure 2.22: Boxplot for the Home Sales Data



Analyzing Distributions

 Steps used to male boxplot:



 Draw box with ends located at the first and third quartiles.

 E.g. Q1 = 139,000 and Q3 = 256,625.

Draw Whiskers

 Dashed lines from the ends of the box to the smallest and largest values inside the limits computed in Step 3.

 Values of 108,000 and 298,000.



 Draw a line at the median (203,750)

Finally, the location of each outlier is shown with an asterisk

 456,250.

 Find Outliers

 IQR = Q3 – Q1

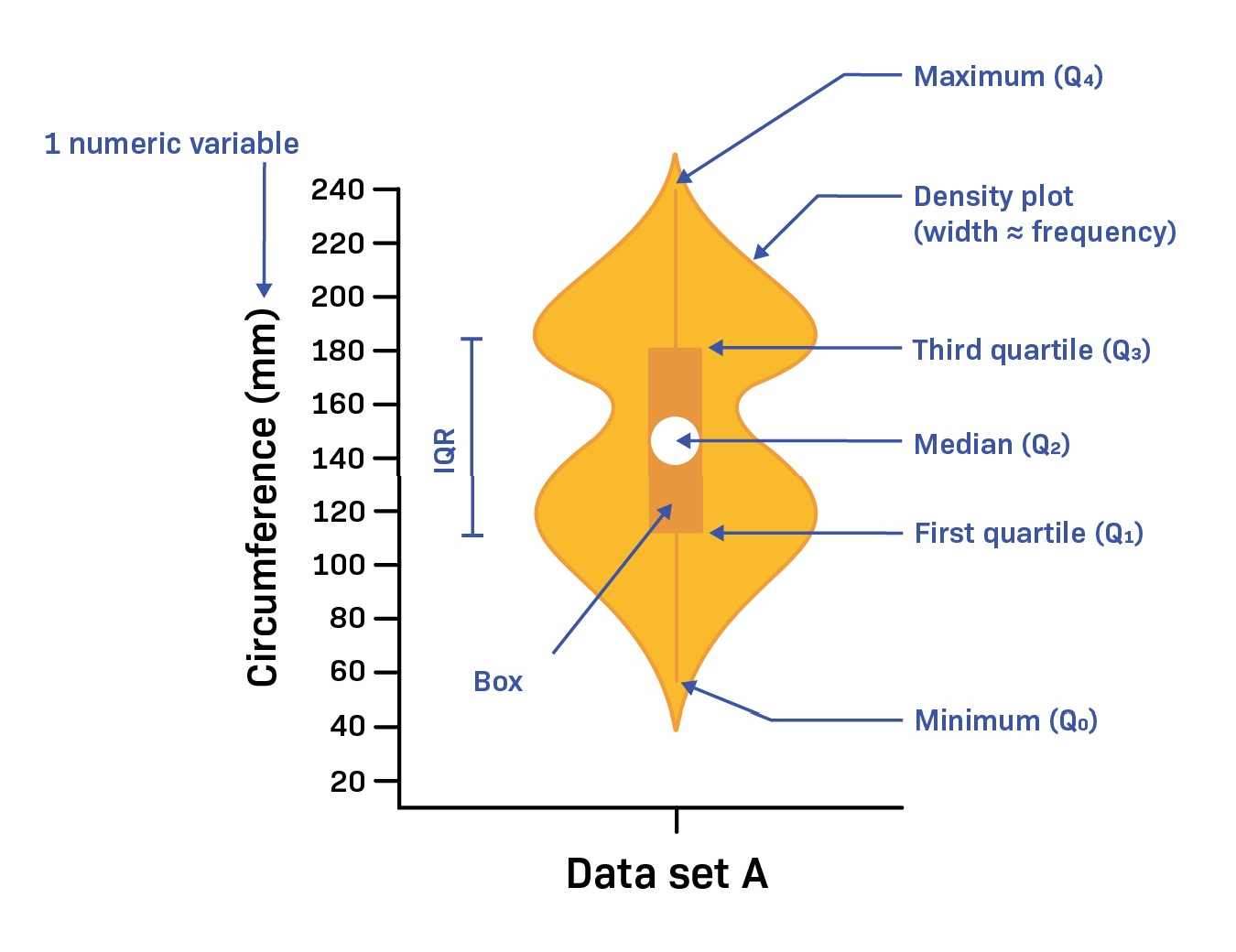
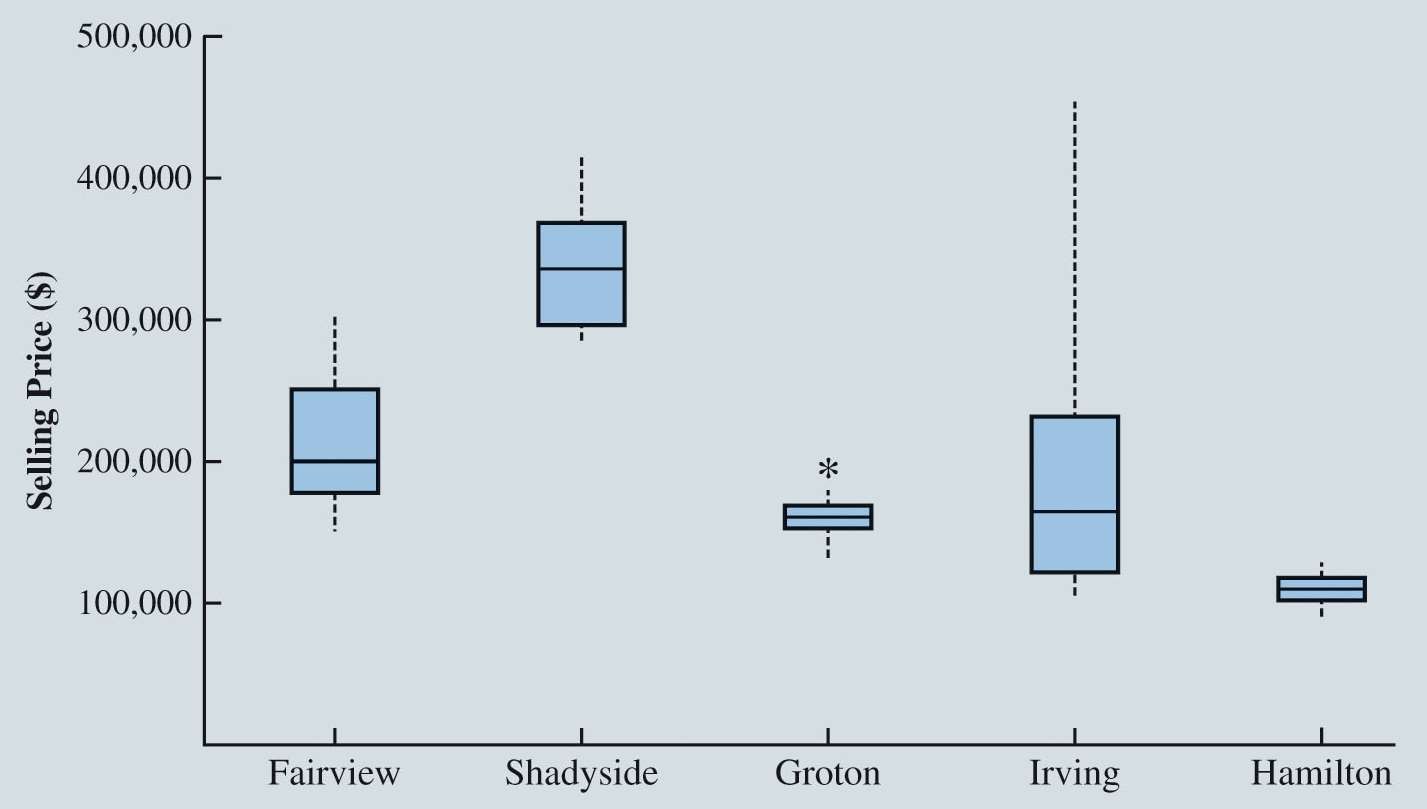
 Limits = Q1 – 1.5(IQR), Q3 + 1.5(IQR)

 IQR = 117,625.

 Limits =

 139,000 – 1.5(117,625) = –37,437.5

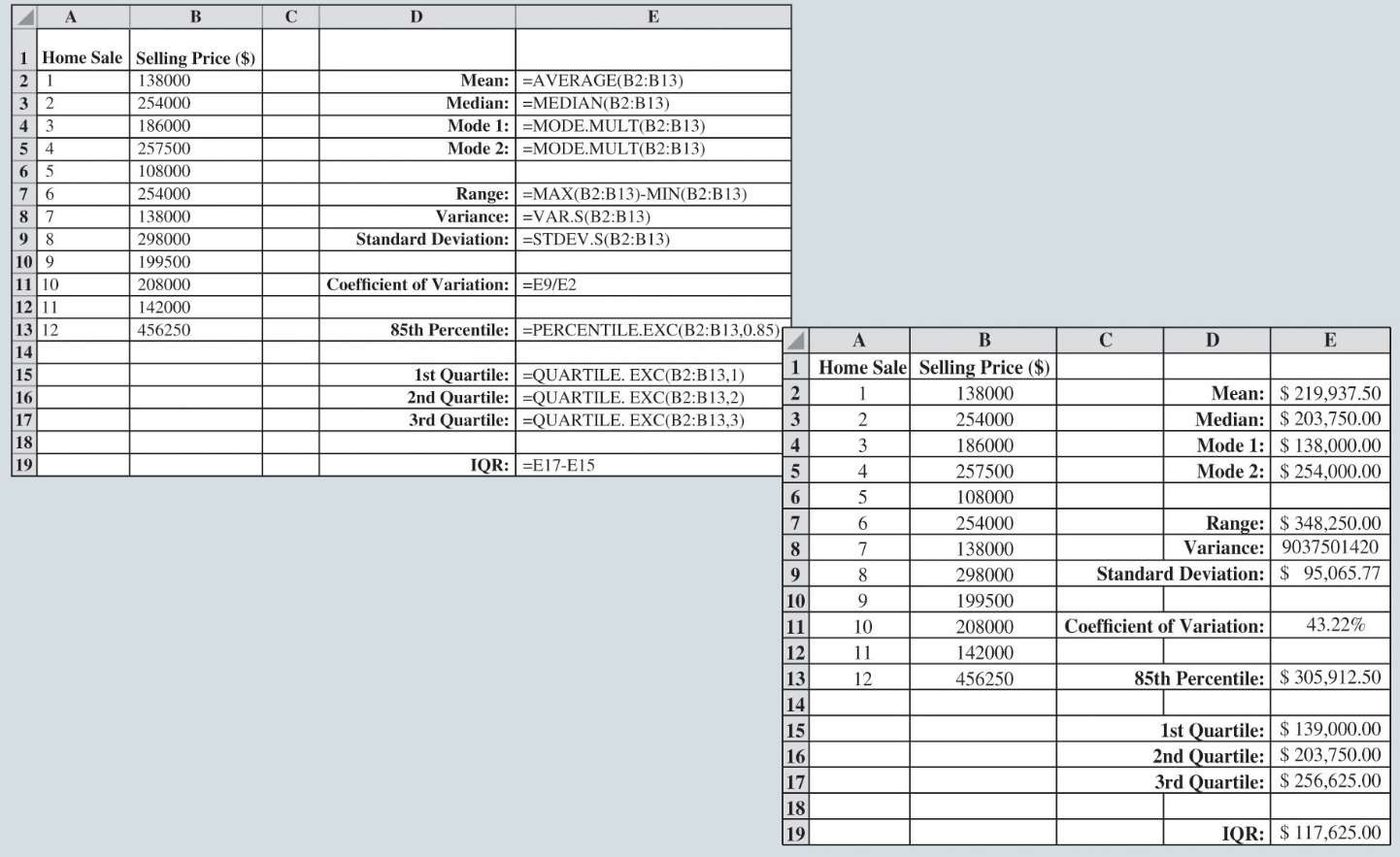
 256,625 + 1.5(117,625) = 433,062.5



Analyzing Distributions

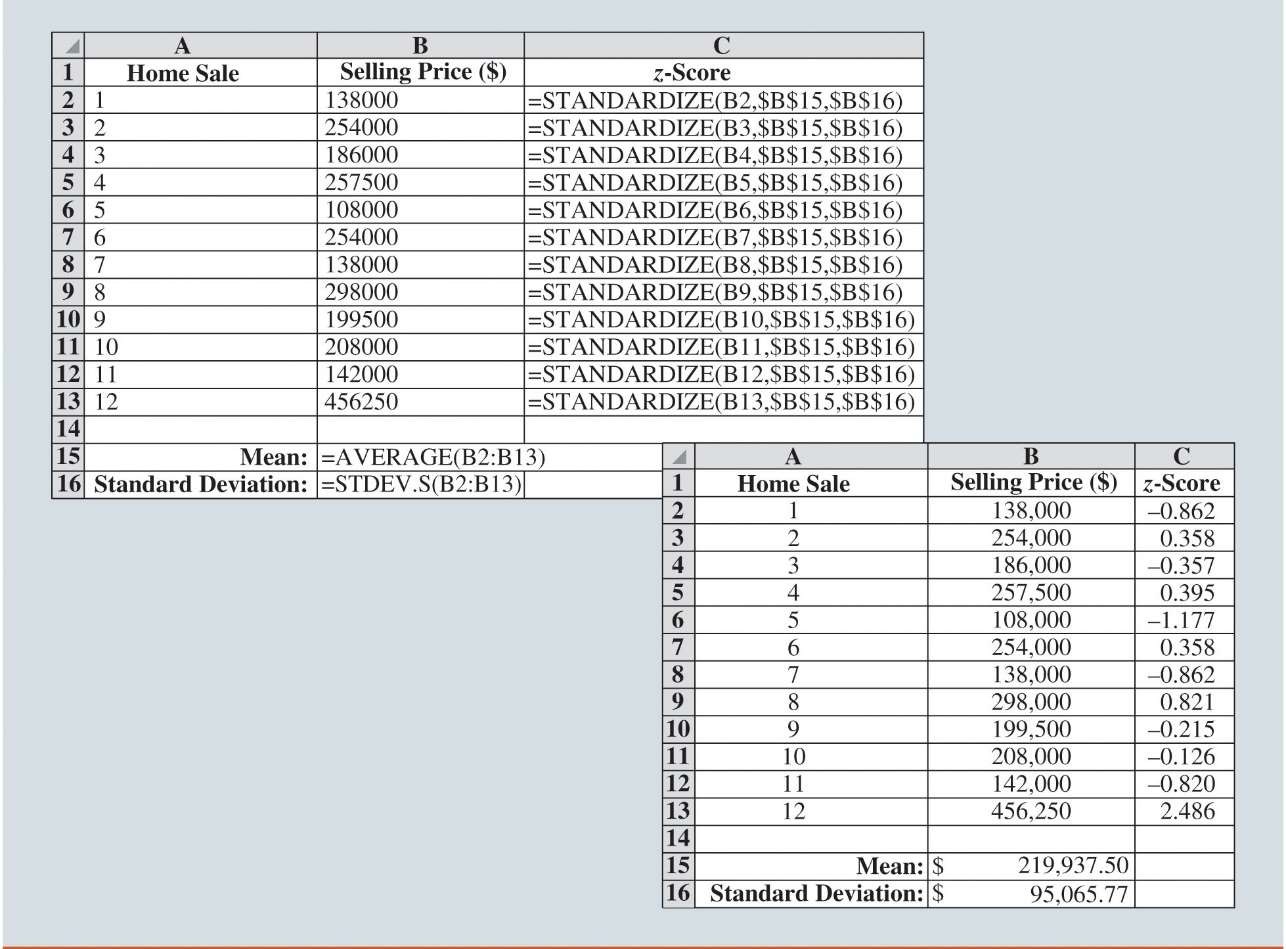
Figure 2.23: Boxplots Comparing Home Sale Prices in Different Communities

**What do you observe about the differences in Home Prices for different communities?**



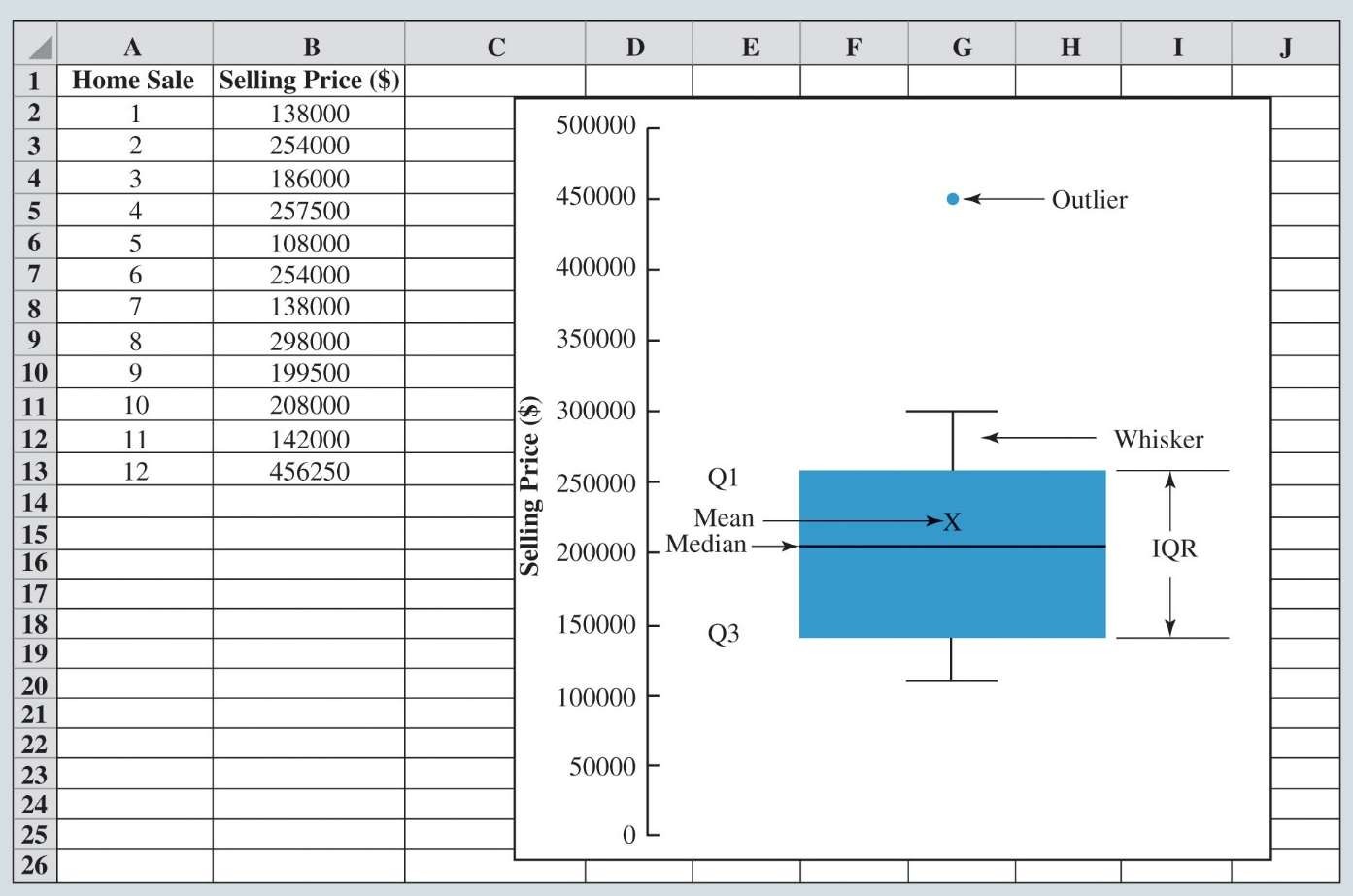
Measures of Variability

Figure 2.19: Calculating Variability Measures for the Home Sales Data in Excel



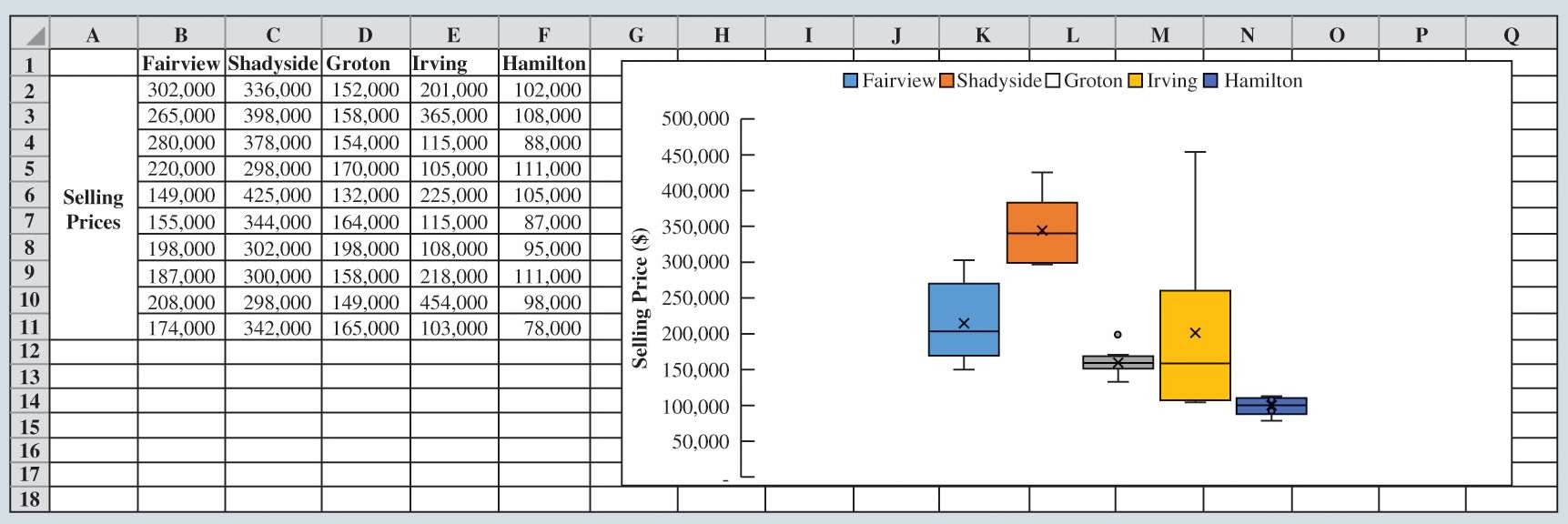
Analyzing Distributions

Figure 2.20: Calculating *z*-Scores for the Home Sales Data in Excel



Analyzing Distributions

Figure 2.24: Boxplot Created in Excel for Home Sales Data



Analyzing Distributions

Figure 2.25: Boxplots for Multiple Variables Created in Excel