Chapter 6: Unit B – Measure of Variation

**Why Variation Matters?**

Imagine customers waiting in line for tellers at 2 different banks. Customer at Bank of Arizona can enter any one of three different lines leading to three different tellers. Bank of the West also has 3 tellers, but all customers wait in a single line and are called to the next available teller. The following values are waiting times, in minutes, for 11 customers at each bank. Times are arranged in ascending order.

Bank of Arizona (3 lines):

4.1 5.2 5.6 6.2 6.7 7.2 7.7 7.7 8.5 9.3 11.0

Bank of the West (1 line):

6.6 6.7 6.7 6.9 7.1 7.2 7.3 7.4 7.7 7.8 7.8

Bank of AZ has greater variation than Bank of the West. Now consider several places where you commonly wait in lines, such as grocery, a concert ticket outlet, or a fast food restaurant. Does each place use a single line or multiple lines? If the places use multiple lines, do you think a single line would be better?

The simplest way to describe the variation of data set is to compute its **range**, the difference between the highest and the lowest data values.

Example 1: Consider 2 sets of quiz scores for students.

Quiz 1: 1 8 8 8 8 8 8

Quiz 2: 2 3 4 5 6 7 8

Which set has the greater range? Would you say that the scores in this set are more varied?

**Quartiles and Five-Number Summary**

**Quartiles** are the values that divide the data distribution into quarters.

**Lower quartile** (first quartile) divides the lowest fourth of a data set from the upper three-fourth. It is the median of the data values in the *lower half* of a data set.

**Middle quartile** (second quartile) is the overall median.

**Upper quartile** (third quartile) divides the lowest three-fourth of a data set from the upper fourth. It is the median of the data values in the *upper half* of a data set.

Bank of Arizona (3 lines):

4.1 5.2 5.6 6.2 6.7 7.2 7.7 7.7 8.5 9.3 11.0

Bank of the West (1 line):

6.6 6.7 6.7 6.9 7.1 7.2 7.3 7.4 7.7 7.8 7.8

Five-number summary describe the data set consisting of the lowest value, the lower quartile, the median, the upper quartile, and the highest value.

|  |  |
| --- | --- |
| Bank of Arizona | Bank of the West |
| Low = 4.1  Lower quartile = 5.6  Median = 7.2  Upper quartile = 8.5  High = 11 | Low = 6.6  Lower quartile = 7.2  Median = 7.2  Upper quartile = 7.7  High = 7.8 |

Boxplot (box-and-whisker- plot) shows the five-number summary visually, with a rectangular box enclosing the lower and upper quartiles, a line marking the median, and whiskers extending to the low and high values.

Example 2: Considering the following two data sets of twenty 100-meter running times (in seconds):

Set 1: 9.92 9.97 9.99 10.01 10.06 10.07 10.08 10.10 10.13 10.13 10.14 10.15 10.10 10.17 10.18 10.21 10.24 10.26 10.31 10.38

Set 2: 9.89 9.90 9.98 10.5 10.35 10.41 10.54 10.76 10.93 10.98 11.05 11.21 11.30 11.46 11.55 11.76 11.81 11.85 11.87 12.00

|  |  |
| --- | --- |
| Set 1 | Set 2 |
| Low = 9.92  Lower quartile = 10.065  Median = 10.135  Upper quartile = 10.195  High = 10.38 | Low = 9.89  Lower quartile = 10.38  Median = 11.015  Upper quartile = 11.655  High = 12.00 |

Five-number summary also characterizes the variation of data values. The single number most commonly used to describe variation is called **standard deviation**. Standard deviation is a measure of how far data values are spread around the mean of a data set.

**Interpreting Standard Deviation**

**Range rule of thumb** is a method to approximate standard deviation.

If we know the standard deviation for a data set, we can estimate the low and high values as follow:

**Homework: Quick Quiz #1-10**

**Review Questions: 1-5**

**Does It Make Sense? # 7, 9, 11**

**Basic Skills & Concept: 15, 17, 19, 21, 23, 25**