# MATH 3070 Introduction to Probability and Statistics Lecture notes Measures of Position

# **Objectives:**

- 1. Compute individual items for a five number summary
- 2. Compute and use standard variance and deviation

# Quartiles

We can measure the spread of the data by identifying the quartiles. As the word implies, a quartile is one-quarter (25%) of the data. We begin by arranging the data points in order. We then find the middle, or median, point. The median is, conveniently, the second quartile since it is at the 50% (one-half) point in the data. The median point between the true median and the lower end of the list of data items is the first quartile. This is where 25% of the data is below this point. The median point between the true median and the upper end of the list of data items is the third quartile. This is where 75% of the data is below this point.

If we divide the data set into 100 equal slices, we call these percentiles. "Per cent" means "of 100" so each percentile represents 1/100th of the data set. The quartiles are, then, the 25th, 50th, and 75th percentiles, respectively.

Another definition of percentile is

The 100pth percentile of a data set is a value of y located so that 100p% of the data lies to the left of the 100pth percentile and 100(1-p)% of the data lies to the right.

Example: If your grade in an industrial engineering class was located at the 84th percentile, then 84% of the grades were lower than your grade and 16% were higher. (Mendenhall)

#### Five Number Summary

A quick way to summarize a data set is by creating a five number summary. This is the minimum data point, the first, second, and third quartile, and the maximum data point.

If we use a graphical display of this summary, it is called a **box-and-whisker plot**. The box of the graph is bound by the first and third quartiles. The median is drawn as a line in the middle of the box. The minimum and maximum data values are located to the appropriate side of the box and connected to the box with lines (the whiskers). This is a compact and convenient display of the numeric data in the summary.

# Interquartile Range

If we take the difference between the first and third quartiles, this is called the **in-terquartile range**. This is the range of the middle 50% of the data. Johnson and Kuby

## Standardization

The phrase, "Comparing apples to oranges," is used to imply that two things being compared really can't be. An example would be SAT to ACT scores. The tests have different measuring scales and it's hard to know if a score of 620 on the SAT verbal section is equal to a score of 18 on the ACT verbal. So how can we do this?

If the data being compared come from normal distributions, we can transform the data so that the two sets are equivalent. We call this transformation **standardization**. By doing this we change the values from their original units into standard deviation units. We also change the starting measuring point to be the mean of the distribution and convert that value to zero.

The result of this operation is called a **z**-score. The z-score tells us how far above or below the mean (in units of standard deviations) an observation is. It also allows us to compare the values since both now are measured the same (in terms of standard deviations) and measured from the same starting point (the mean, zero). Taking advantage of the fact that the area under the normal curve is equal to one (1), we can measure the percentage area for each value.

The formula for this transformation is

$$z = \frac{x - \bar{x}}{s}$$

### Area under the curve (optional)

To compute the area, or percentage, at a given z-score, we must always remember that the mean (0) is at 50%. This means that any value to the left (below) the mean will have less than 50% area below it and any value to the right (above) the mean will have more than 50% area below it. The area above the z-score will be the inverse.

We find the area under the curve using a table. This table can be either one sided or two sided, depending upon the author. Ours is two sided.

First we compute the z-score to two decimal places. Then we look at the table. The column lists the values for the gross measurement, the one's digit and the tenth's. The columns list the fine measurement, the hundredth's. Where the two intersect is the area under the curve for that z-score.