Statistics

Key Definitions

- **Population:** The population is the set of things being studied.
- **Sample:** A sample of that population is any subset of that group. A good sample is a subset that is chosen randomly.
- **Frequency:** Frequency is the number of times a particular data point occurs (the frequency of blue M&Ms in a bag is the number of blue M&Ms in that bag).

Measures of Central Tendency

- **Mean:** The mean of a data set is the average of all the points in that data set. The mean may be denoted by μ or $\bar{x}$. Mathematically,
$$\bar{x} = \frac{\Sigma x_i}{n}$$
where Σx_i means the sum of all data points and n is the total number of data points (the mean test score is the average of all the test scores for the class or the sum of all the scores divided by the number of scores).
Note: When dealing with grouped data, replace x with $f \cdot x$ in the formula
- **Median:** The median, referred to as the "middle number", is the data point in the middle of an ordered data set for sets with an odd number of elements and halfway between the two data points in the middle of an ordered data set with an even number of elements. The median can be found by using the formula
$$L = \frac{n+1}{2}$$
where L is the number element of the median and n is the size of the data set (in a set of 9 elements the median will be the 5^{th} element).
- **Mode:** The mode is the data point with the highest frequency in its data set. If all data points have the same frequency, then there is no mode. (In a survey customers were asked to give a rating from 1 to 5 stars of a particular service. The most frequently occurring rating was 4 stars, thus 4 is the mode of that data set.)

Measures of Dispersion

- **Variance:** The variance from the mean within a data set is used to find the standard deviation and is denoted σ^2or s^2. The variance is given by the formula
$$s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1}$$
where you take the difference between each data point and the mean, square each one, add them all, and then divide by one less than the size of the data set.

- **Standard Deviation:** The standard deviation from the mean within a data set is the average deviation of each data point from the mean, denoted by σ or s and computed using the following equation.

$$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}}$$

Note: The standard deviation formula is simply the square root of the variance formula.
Example: In a class of 5 students, there were the following test scores: 100, 95, 85, 80, and 70. The mean of these test scores is $\mu = \frac{100+95+85+80+70}{5} = 86$ (this can be done on excel using the "=AVERAGE" function on your data set. The standard deviation is $\sigma = \sqrt{\frac{(100-86)^2+(95-86)^2+(85-86)^2+(80-86)^2+(70-86)^2}{5-1}} = \sqrt{\frac{142.5}{4}} = 11.94$ (this can be accomplished on excel using the "=STDEV.S" function on your data set).

The Normal Distribution

- **The Normal Distribution:** Any distribution of data that is said to normal contains
- **The Normal Curve:** The normal curve is a bell-shaped curve that represents the quantity location of data in relation to the mean, which is located at the center of the curve. Most (68.26%) of a normal distribution's data is within one standard deviation of the mean, while 95.44% and 99.74% of the data is within 2 and 3 standard deviations respectively.
- ***Interpreting a Normal Curve*:** The area under the normal curve represents how much data has accumulated as you move from the mean which starts at 0 to the right. Area is accumulated at the same rate as you move to the left. The total area under the curve is 1 and the mean divides the curve in half.
- **Discrete Random Variable:** A discrete variable is any data point that is randomly selected from a data set that can be distinctly counted (like the number of students at Regent University).
- **Continuous Random Variable:** A continuous variable is any data point that is randomly selected from a data set that does not have a distinct number of elements (like how much water is in a bucket as you fill it with a hose).
- **Standard Normal Distribution:** The standard normal distribution is a normal distribution with $\mu = 0$ and $\sigma = 1$. We use a table found in Appendix F of your book to find the area under the curve (or the probability) given a random variable X. The formula

$$z = \frac{X - \mu}{\sigma}$$

provides you with a z-Score that you can look up in the table to find the area under the curve from the mean to that point.

Example: A population of race car top speeds is normally distributed with mean 250 mph and standard deviation 20 mph. Find the probability that a race car's top speed is greater than 255 mph.
Here we are asked to find probability from the right (see below). First, calculate the z-score.

$$z = \frac{255 - 250}{20} = .10$$

Now go to the table and find that the z-score of ,10 lines up with .0398 for area to the right of the mean. We add .5 to account for probability to the left to get .5398 and then subtract from 1 to get a final answer of .4602. There is about a 46% chance that a race car can go faster than 255 mph.

- ***Reading z-Scores***: Use the table in the back of the book to find area under the curve that corresponds to the z-Score that the above formula gives you. Always remember that the table starts at the mean (0) and works to the right. To find probabilities from the left (from true zero), add .5 to the area listed for the z-score you calculated. To find probabilities from the right, repeat the previous step, then subtract the probability related to your z-score from 1. To find probabilities between two z-scores, subtract the smaller area (lower z-score) from the larger area (higher z-score).

Margin of Error

- **Sample Proportion:** The sample proportion is the number of members of a sample, denoted as x, that share a distinguishing trait in a sample of size n. Mathematically, we express the proportion as a decimal or percent as found by $\frac{x}{n}$.
- **Population Proportion:** The population proportion is the actual number of members of the population that share the trait that the sample proportion estimates. The population proportion is denoted P and may larger or smaller than the sample proportion.
- **Margin of Error:** The margin of error is the range of possible error that a sample estimate may have compared to the population proportion. The book abbreviates it as "MOE". The margin of error formula is

$$\text{MOE} = \frac{z_{\alpha/2}}{2\sqrt{n}}$$

Where n is the sample size and $z_{\alpha/2}$ is the z-score for one half of the margin of error. Note that this section also uses the z distribution table in Appendix F

- **Level of Confidence:** The level of confidence is the probability that the sample estimate falls within the amount of possible error allowed for by the chosen MOE and is denoted α.

- **<u>Example</u>:** Assuming a 90% level of confidence, determine the margin of error for a sample of size 150 from a recent survey.

$$\text{MOE} = \frac{1.645}{2\sqrt{150}} \approx .067 \approx 7\%$$

Symbol Guide

Statistics Symbols		
Term	Symbol	Use
Frequency	f	
Data Set Size	n	
Change in (or Difference)	Δ	$\Delta x = x_2 - x_1$
Sum	Σ	$\Sigma x_i = x_1 + x_2 + x_3 \ldots$
Mean	μ	
Sample Mean	$\bar{x}$	
Standard Deviation	σ	
Sample Standard Deviation	s	
z-Score	z	
Level of Confidence	α	