Name: ____________________________________Date: _________________

8-1 REVIEW STATISTICS

When we conduct a study, the complete set of all subjects that share a common characteristic that is being studied is known as the **population**. All populations have **natural or inter-individual variability**. Most of the time, the entire population is not measured, but a sample is taken to infer characteristics of a population. Still, all populations in theory have **population parameters** that describe the population, such as its mean, standard deviation, and interquartile range.

***Exercise* #1:** 18 students in Mr. Weiler's Advanced Calculus class took a quiz with the following results in ascending order.

56, 68, 72, 72, 75, 78, 80, 84, 84, 85, 88, 88, 90, 93, 95, 99, 100, 100

(a) Use your calculator to determine the mean, the median, and the quartiles for this data set. Then, construct a simple box-and-whiskers (box plot) for this data set.

(b) What is the interquartile range of this data set? In theory, what percent of the data set should lie between the first and third quartiles? Is that true for this data set?

(c) What is the population standard deviation for this data set to the nearest tenth? How do you interpret the standard deviation?

(d) What percent of the scores were within one standard deviation of the mean? Within two standard deviations of the mean? Round your percents to the nearest percent and show your work.

Within One Standard Deviation of the Mean		Within Two Standard Deviations of the Mean

Sometimes data is grouped in a frequency chart. We still should be able to calculate the basic population parameters when the information is given in this form.

***Exercise* #2:** A small company has salaries for their 50 employees as given in the table below

(a) Find the mean and standard deviations of the salary range.

(b) What is the median of this data set? Why is the median considerably lower than the mean in this data set?

Salary $\left(x_i\right)$	Frequency $\left(f_i\right)$
25,000	5
32,000	21
45,000	14
58,000	7
75,000	2
120,000	1

(c) Does more or less than 50% of the data set fall within one standard deviation of the mean? Show the analysis that leads to your answer.

Although we have often concentrated on experimental studies where data is collected and means are found, many times we use statistics to represent results of a survey where we are interested in what **proportion** of a **population** share a certain characteristic. These proportions are most expressed as decimals, but sometimes are represented by fractions or percent.

***Exercise* #4:** A questionnaire went home to all juniors concerning their ability to bring and use mobile devices at school. The questionnaires constituted a **census** since all of the juniors were surveyed. Of the 742 juniors, 564 of them reported having web-enabled mobile devices. What was the population proportion for web-enabled devices? Express your answer as a decimal and as a percent.

***Exercise* #5:** The proportion of eggs that get cracked in a local egg handling facility is 0.023. If 2,500 dozen eggs are packaged in the factor per day, what should we expect to be the number of eggs cracked per day?

(1) 350 (3) 230

(2) 450 (4) 690

______

8-1 Statistics Vocabulary

Definitions:

Population-____

Sample-____

Parameter of Interest-____

For each of the scenarios below, identify the population, the sample and the population parameter of interest.

1. A grocery store wants to know the average number of items that shoppers purchase in each visit to the store. They decide to count the items in the cart of every twentieth person through the check stand.

   Population ____

   Sample ____

   Parameter of interest ____

2. A team of biologist wants to know the average weight of fish in a lake. They decide to drop a net and measure all the fish caught in three different locations in the lake.

   Population ____

   Sample ____

   Parameter of interest ____

5 Different Sampling Methods

Sampling Method	Definition
Simple Random	Each individual is chosen by chance. Everybody has the same chance of being chosen.
Systematic Random	Individuals are chosen by a system.
Stratified Random	Certain individuals are chosen from subgroups.
Cluster	Everybody is chosen from a subgroup.
Convenience	Individuals are chosen based upon how easy it is to ask them to participate.

Examples:

____

You are interested in finding out the percent of residents in the city that have experienced a robbery in the past year. Using the city property records, you assign each residence a number. You use a random number generator to give you a list of numbers. You contact the residence that corresponds to that number to ask your questions.

____

You are in charge of school activities. You want to know what activities students would prefer to participate in during the school year. You assign each student in the school a number. You randomly select a starting number among the first 10 numbers and then select every tenth student in the list from that point forward.

____

An auto analyst is conducting a satisfaction survey, sampling from a list of 10,000 new car buyers. The list includes 2,500 Ford buyers, 2,500 GM buyers, 2,500 Honda buyers, and 2,500 Toyota buyers. The analyst selects a sample of 400 car buyers, by randomly sampling 100 buyers of each brand.

____

You are in charge of school activities. You want to know what activities students would prefer to participate in during the school year. You get the list of all the homeroom classes and randomly select 5 classes. You go to each of the classes selected and survey all the students in that class.

____

You are in charge of school activities. You want to know what activities students would prefer to participate in during the school year. You stand in the cafeteria during your lunch break and ask students in they would be willing to participate in your survey as they walk by.

8-2 THE NORMAL DISTRIBUTION

Many populations have a distribution that can be well described with what is known as **The Normal Distribution** or the **Bell Curve**. This curve, as seen in the accompanying handout to this lesson, shows the percent or proportion of a normally distributed data set that lies certain amounts from the mean.

***Exercise* #1:** For a population that is normally distributed, find the percentage of the population that lies

(a) within one standard deviation of the mean.

(b) within two standard deviations of the mean.

(c) more than three standard deviations away from the mean.

(d) between one and two standard deviations above the mean.

As can be easily seen from *Exercise* #1, the majority of any normally distributed population will lie within one standard deviation of its mean and the vast majority will lie within two standard deviations. A whole variety of problems can be solved if we know that a population is normally distributed.

***Exercise* #2:** At Arlington High School, 424 juniors recently took the SAT exam. On the math portion of the exam, the mean score was 540 with a standard deviation of 80. If the scores on the exam were normally distributed, answer the following questions.

(a) What percent of the math scores fell between 500 and 660?

(b) How many scores fell between 500 and 660? Round your answer to the nearest whole number.

(c) If Evin scored a 740 on her math exam, what percent of the students who took the exam did better than her?

(d) Approximately how many students did better than Evin?

***Exercise* #3:** The heights of 16 year old teenage boys are normally distributed with a mean of 66 inches and a standard deviation of 3. If Jabari is 72 inches tall, which of the following is closest to his height's percentile rank?

(1) 85th (3) 98th

(2) 67th (4) 93rd

______

***Exercise* #4:** The amount of soda in a standard can is normally distributed with a mean of 12 ounces and a standard deviation of 0.6 ounces. If 250 soda cans were pulled by a company to check volume, how many would be expected to have less than 11.1 ounces in them?

(1) 17 (3) 28

(2) 23 (4) 11

______

***Exercise* #5:** Biologists are studying the weights of Red King Crabs in the Alaskan waters. They sample 16 crabs and compiled their weights, in pounds, as shown below.

9.8, 10.1, 11.1, 12.4, 11.8, 13.2, 12.8, 12.5, 13.7, 11.6, 13.4, 12.3, 12.6, 14.8, 14.2 15.1

(a) Determine the mean and sample standard deviation for this sample of crabs. Round both statistical measures to the nearest *tenth* of a pound.

(b) Why does this sample indicate that the population would be well modeled using a normal distribution? Explain. Hint – Use your calculator to sort this data in ascending order.

(c) Assuming your mean and standard deviation from part (a) apply to a normally distributed population of crabs caught in Alaska, what percent will fall between 9.6 pounds and 15.6 pounds?

(d) If fishermen must throw back any crab caught below 10.4 pounds, approximately what percent of the crabs caught will need to be thrown back if the weights are normally distributed?

8-3 THE NORMAL DISTRIBUTION AND Z-SCORES

The normal distribution can be used in increments other than half-standard deviations. In fact, we can use either our calculators or tables to determine probabilities (or proportions) for almost any data value within a normally distributed population, as long as we know the population mean, μ, and the population standard deviation, σ. But, first, we will introduce a concept known as a data value's z-score.

THE Z-SCORE OF A DATA VALUE

For a data point x_i, its z-score is calculated by: $z = \dfrac{x_i - \mu}{\sigma}$. It calculates how far from the mean, in terms of standard deviations, a data point lies. It can be positive if the data point lies above the mean or negative if the data point lies below the mean.

***Exercise* #1:** Boy's heights in seventh grade are normally distributed with a mean height of 62 inches and a standard deviation of 3.2 inches. Find z-scores, rounded to the nearest hundredth, for each of the following heights. Show the calculation that leads to your answer.

(a) $x_i = 66$ inches

(b) $x_i = 57$ inches

(c) $x_i = 70$ inches

Z-scores give us a way to compare how far a data point is away from its mean in terms of standard deviations. We should be able to compute a z-score for a data value and go in the opposite direction.

***Exercise* #2:** Jeremiah took a standardized test where the mean score was a 560 and the standard deviation was 45. If Jeremiah's score resulted in a z-value of 1.84, then what was Jeremiah's score to the nearest whole number?

With z-scores, we can then determine the probability that a subject picked from a normally distributed population would have a characteristic in a certain range. Z-score tables come in many different varieties. The one that comes with this lesson shows only the right hand side, so symmetry will have to be used to determine probabilities.

***Exercise* #3:** The lengths of full grown sockeye salmon are normally distributed with a mean of 29.2 inches and a standard deviation of 2.4 inches.

(a) Find z-scores for sockeye salmon whose lengths are 25 inches to 32 inches. Round to the nearest hundredth.

(b) Use the z-score table to determine the proportion of the sockeye salmon population, to the nearest percent, that lies between 25 inches and 32 inches. Illustrate your work graphically.

***Exercise* #4:** If the scores on a standardized test are normally distributed with a mean of 560 and a standard deviation of 75. Answer the following questions by using z-scores and the normal distribution table.

(a) Find the probability that a test picked at random would have a score larger than 720. Round to the nearest tenth of a percent.

(b) Find the probability that a completed test picked at random would have a score less than 500. Round to the nearest tenth of a percent.

(c) Find the probability that a completed test picked at random would have a score between 500 and 600.

(d) Find the probability that a completed test picked at random would have a score between 600 and 700.

This process is sometimes used to determine a particular data point's **percentile**, which is the **percent of the population less than the data point**.

***Exercise* #5:** The average weight of full grown beef cows is 1470 pounds with a standard deviation of 230 pounds. If the weights are normally distributed, what is the percentile rank of a cow that weighs 1,750 pounds?

(1) 89^{th} (3) 49^{th}

(2) 76^{th} (4) 35^{th}

____

Your graphing calculator can also find these proportions or percent values. Each calculator's inputs and language will be slightly different, although many will do much of the work for you, even allowing you to **not think about the z-scores**.

***Exercise* #6:** Given that the volume of soda in a 12 ounce bottle from a factory varies normally with a mean of 12.2 ounces and a standard deviation of 0.6 ounces, use your calculator to determine the probability that a bottle chosen at random would have a volume:

(a) Greater than 13 ounces. (b) Less than 11 ounces (c) Between 11.5 and 12.5 ounces

THE NORMAL DISTRIBUTION

BASED ON STANDARD DEVIATION