Name ______________________

Date ____________ Period ___________

Lesson 1-1 , Measures of Central Tendency and Box Plots

- Measures of central tendency, also referred as measures of center, refer to different types of ____________________.
- The most common measures of central tendency are ______________, _______________, and ____________ .

MEAN

- The symbol for the mean is ______ , which is read as __________ .
- Another symbol for the mean is _______ , which is read as __________ .

MEDIAN

- Median refers to the ___________ value of a set of data once it has been ordered from least to greatest. The median of a set of data with an even number of values is ____ .

MODE

- Mode refers to the number that appears ______________________ in a set of data. Data sets with two modes are said to be _______________ . Sets have no mode when each item of the set has equal frequency

Ex. 1: Salary Data

Find the mean, median, and mode of the salaries for the corporate employees listed below. Which measure of central tendency appears to most accurately represent the set of data?

Allen: $40,000
Baker: 42,000
Chase: 59,000
Deitz: 60,000
Eckerd: 62,000
Francis: 65,000

How do extreme values (outliers) affect the measures of central tendency?

- Mean -
- Median -
- Mode -

Ex. 2: Backpack Weights
Owen is a member of the student council and wants to present data about backpack safety to the school board. He collects data on the weights of backpacks of 30 randomly chosen students. How much does the typical backpack weigh at Owen's school?

{ 3, 4, 4, 4, 6, 7, 7, 7, 7, 7, 8, 8, 9, 9, 9, 9, 9, 10, 10, 10, 10, 10, 10, 13, 15, 15, 16, 17, 20, 33 }

Box and Whisker Plots and the Five Number Summary

Next we will look at Box and Whisker Plots (aka Box Plots). They are used to summarize a data set and to visually illustrate the ______________________________ of the data. A Box and Whisker plot looks like this:

The five parts of a Box and Whisker plot for a particular data set correspond to the Five Number Summary for that data set. The five numbers in the Five Number Summary are the ______________, __________________, __________________, __________________, and ______________ .

1st: Arrange the data in order and find the median. This separates the data into 2 groups.
2nd: Find the median of the _________________ and ___________________ of the data set. Now your data set is divided into four groups, and each of these four groups is called a ________________ . There are 3 points called __________________________________, (Q_1, Q_2, and Q_3) that denote the breaks in the data for each quartile.

- Q_1 is the median of the ____
- Q_2 is the median of the _____________________________________
- Q_3 is the median of the ____
- The difference between Q1 and Q3 (i.e., Q3-Q1) is called the ______________________________
- The difference between the maximum and minimum values is called the __________

Box-and-Whiskers plots…
- can be drawn vertically or horizontally
- consists of a rectangular box with the ends, or _______________, located at the first and third quartiles

- the segments extending from the ends of the box are called ______________
- the whiskers stop at the minimum and maximum values of a data set unless it contains ______________.

Outliers

- Outliers are ______________ values
- The technical definition of an outlier is a data point that is more than 1.5 of the interquartile range beyond the upper or lower quartiles. That is, any number less than $Q_1 - 1.5(IQR)$ or greater than $Q_3 + 1.5(IQR)$ is considered an outlier.
- Outliers are ________________ represented by single points on a box plot.
- If outliers exist, each whisker is extended to the last value of the data set that is not an outlier.

MATH III Unit 1, Statistics | Name ______________________________

Notes 1-2, Normal Distributions, Mean, S.D. | Date ____________ Period ___________

The Normal Distribution

When you draw a dot plot for some data sets, you get a distribution that has a particular shape. It looks like this:

This distribution shape is so common, and there are so many different data sets that produce it, that it is given a special name. It is called a ___________________ distribution. (You may have also heard it called a ____________________________ .)

When you have a data set that is normally distributed, that means that if you were to draw a dot plot of the data set, it would have this characteristic "bell" shape.

For a normally distributed data set, there are two values that we can calculate that will tell us a GREAT DEAL about the data set.

1. The value of the mean, which is a measure of ____________________________
2. The value of the standard deviation (SD), which is a measure of ___________ or ___________________. (The greater the SD, the greater the spread of the data about the mean.)

Example 1: The Rubber Band Launch (P. 85-86 in Green AA text)

You want to find out how consistently rubber bands will travel when launched, so you use a ruler to launch two rubber bands seven times each. You generate the following data sets:

- Rubber band #1 distances (cm): {182, 186, 182, 184, 185, 184, 185}
- Rubber band #2 distances (cm): {152, 194, 166, 216, 200, 176, 184}

Data Point	Mean	Deviation from Mean	Squared Deviation From Mean

Name ______________________________
Date ____________ Period ___________

Normal Distributions and Percentages

The Empirical Rule - ____
____

In any ___________ data set that is ___________________________ distributed:

Approx. ______ of the values will be within 1 standard deviation of the mean

Approx. ______ of the values will be within 2 standard deviations of the mean

Approx. ______ of the values will be within 3 standard deviations of the mean

Ex. 1: A group of students weighs 500 US pennies. They find that the pennies have normally distributed weights with a mean of 3.1g and a standard deviation of 0.14g.

a) Sketch the normal curve for this distribution below. Label the mean and three standard deviations above and below the mean.

b.) What percent of the pennies have a weight that lies between:
2.96g and 3.24g (i.e., within one standard deviation of the mean)? ________
2.82g and 3.38g (i.e., within two standard deviations of the mean)? ___________
2.68g and 3.52g (i.e., within three standard deviations of the mean)? _________

c.) **How many** pennies have a weight that lies within
2.96g and 3.24g (i.e., within one standard deviation of the mean)? ________
2.82g and 3.38g (i.e., within two standard deviations of the mean)? ___________
2.68g and 3.52g (i.e., within three standard deviations of the mean)? _________

What if I wanted to know the percentage of pennies that had a weight between 3g and 3.2g?

Calculator Function: **normalcdf()**

The TI83/TI84 calculators have a function called normalcdf() which will tell you:

____

and all you have to give it is: ______________________

____

____

(Note that normalcdf assumes that your data set is ____________________________.)

The format of the normalcdf() function is:

normalcdf(_____________, ___________, __________, ___________)

So if we wanted to know the percentage of pennies from our data set that had a weight between 3g and 3.2g, we would enter the following into our calculator:

normalcdf (_______, ________, ________, _________) MATH III Notes, Unit 1, Statistics

Notes 1-4, Z-scores

Name ______________________

Date ____________ Period ____________

The z-score of a data point: ____.

A z-score or z-value can be calculated for ________________________________.

To calculate the z-value for a given data point:

Ex 1: A group of students weighs 500 US pennies.
They find that the pennies have normally distributed weights with a mean of 3.1g and a standard deviation of 0.14g

a) What is the z-score for a penny that weighs 3.24g?

b) What is the z-score for a penny that weighs 2.96g?

c) What is the z-score for a penny that weighs 3.31g?

d) What is the z-score for a penny that weighs 2.89g?

A positive z-score indicates ____.

A negative z-score indicates ____.

Ex. 2: For the data set in Example 1:

a.) If a penny has a z-score of .64, how much does it weigh?

b.) If a penny has a z-score of -2.8, how much does it weigh?

MATH III Notes, Unit 1, Statistics

Notes 1-5 Samples and Surveys

Name ______________________________

Date ____________ Period ___________

A ____________________ is all the members of a set.

A ____________________ is part of a population.

If you determine a sample carefully, it can give a good estimate of the total population.

Sampling Types and Methods

1. ______________________________ - select any members of the population who are conveniently and readily available.
2. ______________________________ - select only members of the population who volunteer for the sample.
3. ______________________________ - order the population in some way, and then select from it at regular intervals.
4. ______________________________ - all members of the population are equally likely to be chosen.

A _______ is a systematic error introduced by the sampling method.

Example 1 Analyzing Sampling Methods

A newspaper wants to find out what percent of the city population favors a property tax increase to raise money for local parks. What is the sampling method used for each situation? Does the sample have a bias? Explain.

A. A newspaper article on the tax increase invites readers to call the paper and express their opinions.

B. A reporter interviews people leaving the city's largest park.

C. A survey service calls every 50^{th} listing from the local phone book.

Study Methods

1. ______________________________ - measure or observe members of a sample in such a way that they are not affected by the study.
2. ______________________________ - divide the sample into two groups. You impose a treatment on one group but not on the other "control" group. Then you compare the effect on the treated group to the control group.
3. ___________________________ - ask every member of the sample a set of questions.
4. ______________________________ - uses a probability experiment to mimic a real-life situation.

A poorly written survey question can introduce bias. It should avoid:

-
-
-
-
-

Example 2 Analyzing Survey Questions

Is there any bias in the survey question? Explain.

A. Do you think farmers should use poison to control insects on crops?

B. Don't you agree that most childcare workers are underpaid?

C. Do you think teachers should communicate frequently with students and their parents about class grade?

Name ______________________________

Date ____________ Period ___________

Margin of Error

Margin of Error is a ____________ that tells the uncertainty in an estimate.

It is a measure of how close we believe the ____________ proportion is to the ___________________ proportion.

The formula used to predict MOE with 95% confidence ≈ __________________________.

The margin of error is roughly two standard deviations away from the mean.

Example 1

During the week of 08/10/2001, CNN conducted a poll asking 1000 Americans whether they approve of President Bush's performance as President. The approval rating was 57%. In the next poll conducted during the week of 09/21/2001, CNN conducted the same poll asking 100 Americans whether they approve of President Bush's performance as President. The approval rating was 90%.

1. Why the difference in ratings?

2. Find the MOE in the August poll.

3. Find the MOE in the September poll.

4. Explain why the MOE for the August poll is less than that in September.