AFM – Unit 1, Statistics | Name _____

Notes 1.1, Central Tendency, ~~Box~~ Plots | Date _____ Period _____

Lesson 2.1, Measures of Central Tendency and Box Plots (AA text pp. 76 – 80)

- Measures of central tendency, also referred as measures of center, refer to different types of averages.
- The most common measures of central tendency are mean, median, and mode.

MEAN

- The symbol for the mean is $\bar{x}$, which is read as x-bar.
- Another symbol for the mean is μ, which is read as mew.

MEDIAN

→ Numbers must be in order from least to greatest

- Median refers to the middle 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 not a number in the data set. {4, 7, 9, 15, 24, 40}

{4, 7, 9, 15, 24} → 9 is median

$\frac{9+15}{2} = \frac{24}{2} = 12$ is the median

MODE

- Mode refers to the number that appears most frequently in a set of data. Data sets with two modes are said to be bi-modal. 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

mean : $54,000

median : $59,500

mode : $ none → since no # repeats

* put in calculator as a list set.

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

- Mean – changes significantly, because
- Median –
- Mode – does not change

8/27

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? 9 to 10 lbs.

{ 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 }

Q_1 Q_2 Q_3

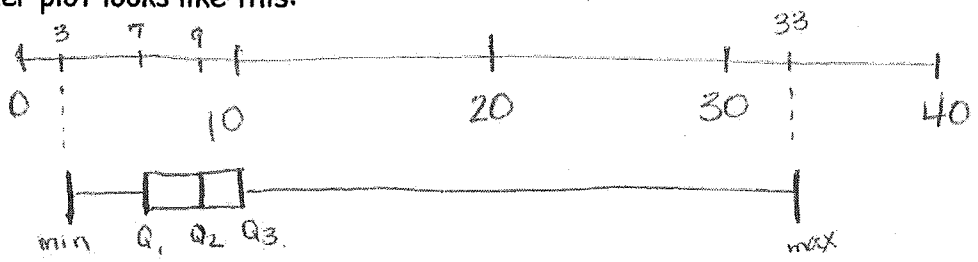
min max

mean : 10.2 lbs. mode: 10 lbs.

median: 9

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 variability (spread) 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 minimum, 1st quartile, 2nd quartile (median), 3rd quartile, and maximum.

1st: Arrange the data in order and find the median. This separates the data into 2 groups.

2nd: Find the median of the 1st half and 2nd half of the data set.

Now your data set is divided into four groups, and each of these four groups is called a quartiles. There are 3 points called quartile points, (Q_1, Q_2, and Q_3) that denote the breaks in the data for each quartile.

- Q_1 is the median of the first half of the data set
- Q_2 is the median of the entire data set
- Q_3 is the median of the second half of the data set
- The difference between Q1 and Q3 (i.e., Q3-Q1) is called the interquartile range
- The difference between the maximum and minimum values is called the range

Box-and-~~Whisk~~Whiskers plots...

- can be drawn vertically or horizontally
- consists of a rectangular box with the ends, or edges, located at the first and third quartiles
- the segments extending from the ends of the box are called Whiskers
- the whiskers stop at the minimum and maximum values of a data set unless it contains outliers.

Outliers

- Outliers are extreme 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 extreme values 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.

Examples:

MATH III – Unit 1, Statistics

Notes 8-2, Dot Plots and Distributions

Name _____

Date _____ Period _____

A data set is a set of related numbers, often called data points

A distribution is the way the numbers in a data set are distributed

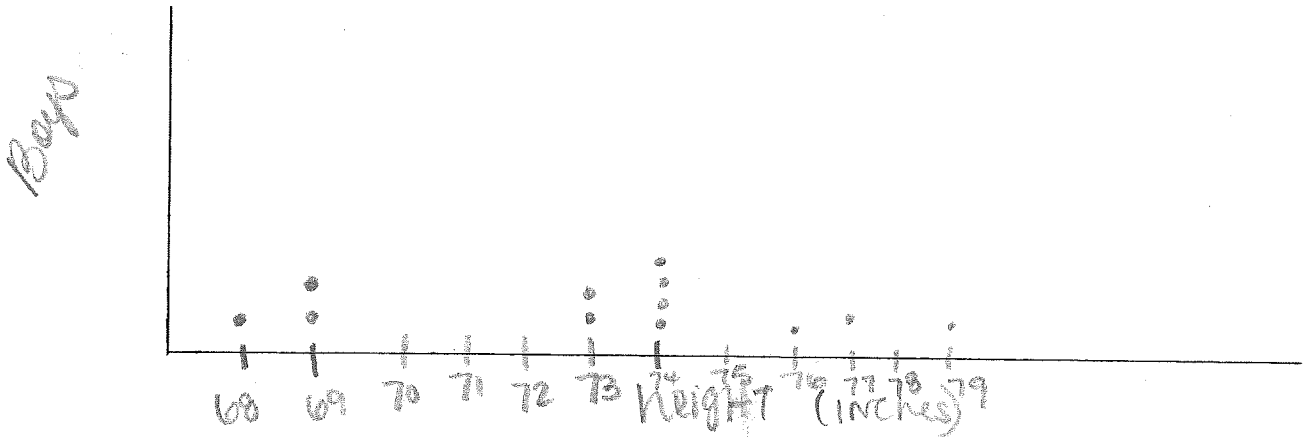
A dot plot is a way of representing a distribution in graphical form

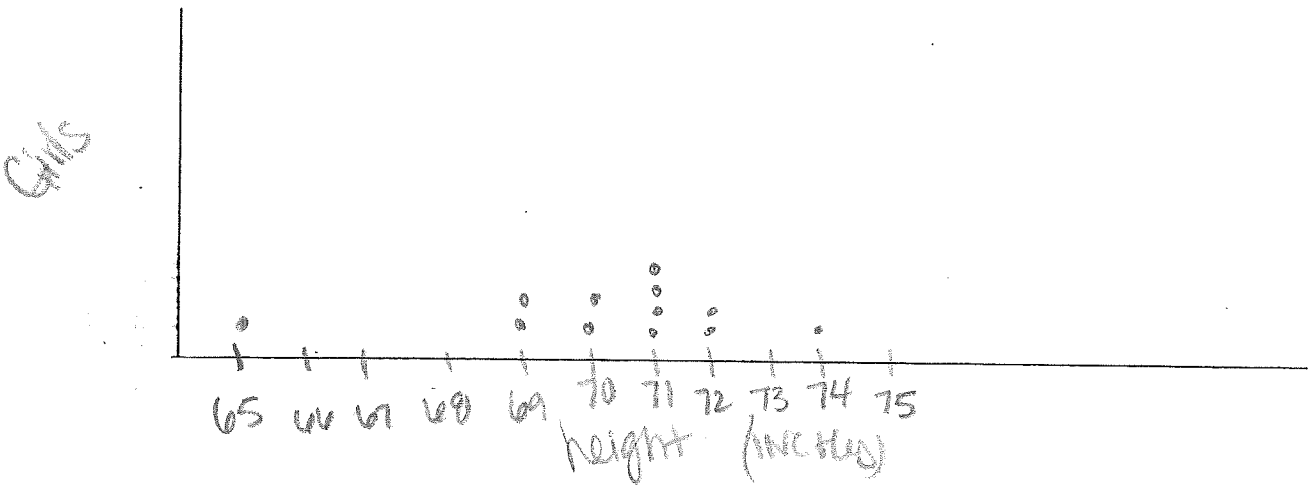
Example 1) The heights (in inches) of each member of the girls' and boys' basketball teams at Holbrook High School are shown below.

Boys' team: 68, 69, 69, 73, 73, 74, 74, 74, 74, 76, 77, 79

Girls' team: 65, 69, 69, 70, 70, 71, 71, 71, 71, 72, 72, 74

Sketch a dot plot for each data set.





Dot Plots & Distributions

The two dot plots show how the heights of the two groups of basketball players are distributed. How would you describe, in words, these two distributions?

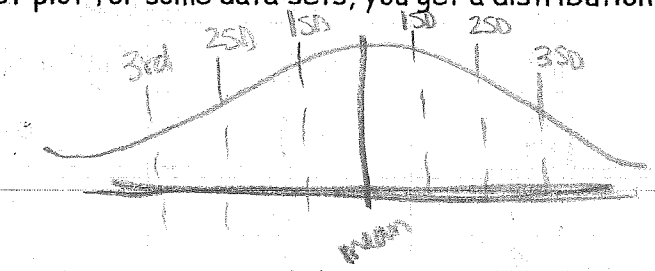
	Boys	Girls
mean :	73.3	70.4
median :	74	71
mode :	74	71

MATH III Unit 8, Statistics Name _____

Notes 8-3, 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:



3 standard deviation to the right of the mean

3 stan. dev. to the left of the mean

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 normal distribution. (You may have also heard it called a bell-shaped curve.)

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.

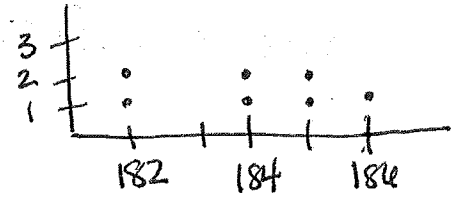
For a normally distributed data set, there are two values that we can calculate that will tell us a GREAT DEAL about the data set.

1. The value of the mean, which is a measure of central tendency
2. The value of the standard deviation (SD), which is a measure of spread or how spread out. (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:

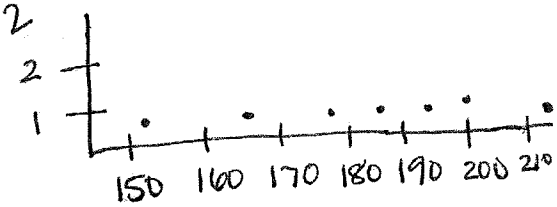
- Rubber band #1 distances (cm): {182, 186, 182, 184, 185, 184, 185}
- Rubber band #2 distances (cm): {152, 194, 166, 216, 200, 176, 184}

RB#1



mean 184

Rb#2



mean 184

RB#1

Data Point	Mean	Deviation from Mean	Squared Deviation From Mean
182	184	−2	$(-2)^2 = 4$
186	184	+2	$(2)^2 = 4$
182	184	−2	4
184	184	0	0
185	184	+1	1
184	184	0	0
185	184	+1	1
		total	14

only make table if finding standard deviation by hand

$$\text{Variance} = \frac{14}{7-1} = \frac{14}{6} = 2.33$$

(14 ← squared deviation; 7 − 1 ← # of data points)

$$\text{Standard deviation} = \sqrt{\text{variance}} = \sqrt{2.33} = 1.53$$

↳ S_x or σ_x

*Variance: the average of the squared differences from the mean

$$*\text{Variance} = (\text{S.D.})^2$$

↑ standard deviation

MATH III Notes, Unit 8, Statistics — Name _____

Notes 8-4, Normal Distributions and Percentages — Date _____ Period _____

Normal Distributions and Percentages

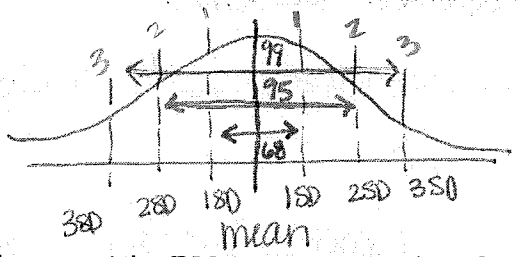
The Empirical Rule - (the 68-95-99 Rule)

In any large data set that is normally distributed:

Approx. 68% of the values will be within 1 standard deviation of the mean

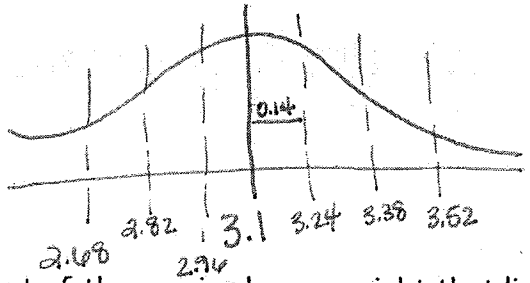
Approx. 95% of the values will be within 2 standard deviations of the mean

Approx. 99% 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)? 68.3%

2.82g and 3.38g (i.e., within two standard deviations of the mean)? 95.5%

2.68g and 3.52g (i.e., within three standard deviations of the mean)? 99.7%

c.) **How many** pennies have a weight that lies within

2.96g and 3.24g (i.e., within one standard deviation of the mean)? .68 * 500 = 342

2.82g and 3.38g (i.e., within two standard deviations of the mean)? .95 x 500 =

2.68g and 3.52g (i.e., within three standard deviations of the mean)? .99 x 500 =

multiply by the % you started with along w/ me decimal. This example started w/ 500 pennies

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

Calculator Function: **normalcdf()** 2nd Vars #2

The TI83/TI84 calculators have a function called normalcdf() which will tell you: the % of values that lie within a given interval

and all you have to give it is: the given interval
mean
standard deviation

(Note that normalcdf assumes that your data set is Normally Distributed.)

The format of the normalcdf() function is:

normalcdf(lower bound , upper bound , mean , standard deviation)

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:

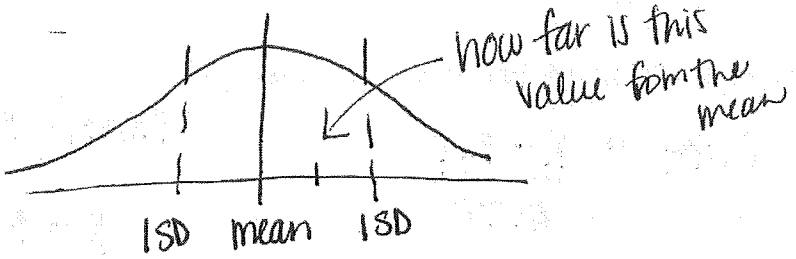
normalcdf (3 , 3.2 , 3.1 , .14)

this will give you a percentage

move decimal 2 places

MATH III Notes, Unit 8, Statistics Name _____

Notes 8-5, Z-scores Date _____ Period _____



The z-score of a data point: the number of standard deviations the point is from the mean

A z-score or z-value can be calculated for any point/value in a data set.

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

$$Z = \frac{\text{deviation from mean}}{\text{standard deviation}} = \frac{\text{data point} - \text{mean}}{\text{S.D.}} = \frac{x - \bar{x}}{s_x}$$

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?

$$Z = \frac{3.24 - 3.1}{0.14} = \frac{.14}{.14} = 1$$

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

$$Z = \frac{2.96 - 3.1}{0.14} = \frac{-0.14}{0.14} = -1$$

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

$$Z = \frac{3.31 - 3.1}{0.14} = \frac{0.21}{0.14} = 1.5$$

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

$$Z = \frac{2.89 - 3.1}{0.14} = \frac{-0.21}{0.14} = -1.5$$

A positive z-score indicates the data point lies above the mean

A negative z-score indicates the data point lies below the mean

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

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

$$0.64 = \frac{X - 3.1}{0.14}$$

$$0.64(0.14) = X - 3.1$$

$$X = 3.1896$$

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

$$-2.8 = \frac{X - 3.1}{0.14}$$

$$X = 2.708$$

$$-2.8(0.14) = X - 3.1$$

$$-0.392 = X - 3.1$$

$$+3.1 \qquad +3.1$$

$$2.708 = X$$

AFM Notes Statistics
Notes 1-7 Samples and Surveys

Name Key
Date _____ Period _____

A Population is all the members of a set.

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

A bias 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. Self-selected
Bias - who calls the newspaper. People who call may over represent or underrepresented some views.

B. A reporter interviews people leaving the city's largest park.
Convenience sample, since it is convenient for the reporter to stand in one place. May overrepresent park supporters.

C. A survey service calls every 50th listing from the local phone book.
Systematic Sample. If there is some link between people who are listed (or not listed) in a phone book and people who pay property tax

Study Methods

1. Observational Study - measure or observe members of a sample in such a way that they are not affected by the study.
2. Controlled Experiment - 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. Survey - ask every member of the sample a set of questions.
4. Simulation

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

- Combining two or more issues.
- Using double negatives
- overlapping answer choices
- words that cause strong reactions (loaded question)
- suggesting that you want a particular answer (leading question.

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?
Loaded Question
B. Don't you agree that most childcare workers are underpaid?
Leading Question
C. Do you think teachers should communicate frequently with students and their parents about class grade? 2 Issues

AFM Unit 2, Statistics

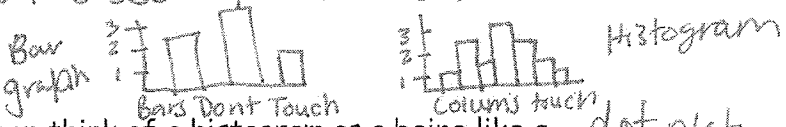
Notes 2-6, Histograms and Percentile Rank

Name Key

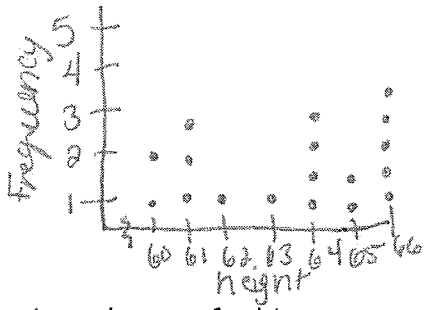
Date _____ Period _____

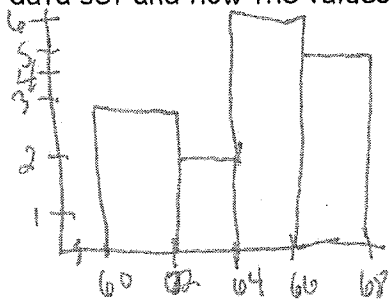
Lesson 2-6- Histograms and Percentile Ranks

Histogram - A way of displaying a distribution that use columns to show how the data points are distributed.



You can think of a histogram as a being like a dot plot, except that it doesn't show every single data point. For this reason, histograms are a good way to display information from very large data sets. Although you can't see individual data values, you can see the shape of the data set and how the values are distributed throughout the range.

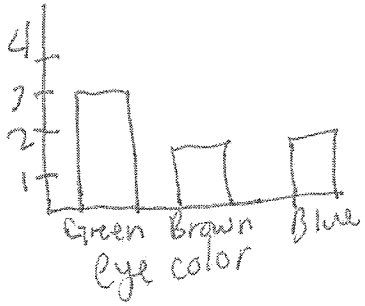




Intervals cannot change!

The columns of a histogram are called bins, which always have the same width. The height of the bins indicates how many data points fall within a given interval.

Note that a histogram is NOT the same as a bar graph. The bars of a bar graph indicate how many data points are in a particular category.



* The order of the categories can change!

All of the bins of a histogram should have the same width.
The bin width may change depending on how much detail you want your histogram to show.

Percentile Rank - For a given distribution, the percentile rank tells the percentage of data values that lie within a given value.