MATH III - Unit 1, Statistics

Name ____________________

Notes 1-1

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 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

- 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 the mean of the 2 middle numbers.

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

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

- Mean - $54,667
- Median - $59,500
- Mode - none

Outliers will cause the spread to be larger

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 }

mean: 10.2 lbs   median: 9 lbs   mode: 10 lbs

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 1st Quartile, 2nd Quartile, 3rd Quartile, Minimum, 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-Whiskers plots...

- can be drawn vertically or horizontally
- consists of a rectangular box with the ends, or medians, 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

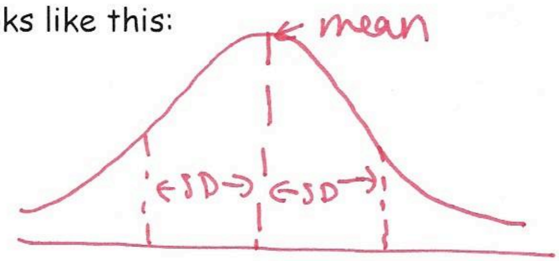
- Outliers are outside values
- The technical definition of an outlier is a data point that is more than 1.5 of the interquartile range beyond the upper or lower quartiles. That is, any number less than $Q_1 - 1.5(IQR)$ or greater than $Q_3 + 1.5(IQR)$ is considered an outlier.
- Outliers are value represented 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 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.

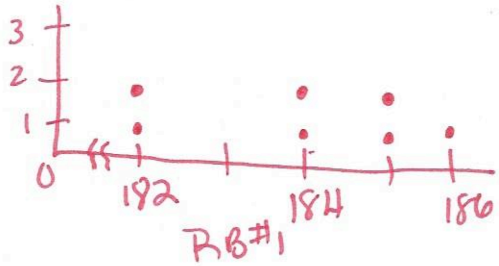
For a normally distributed data set, there are two values that we can calculate that will tell us a GREAT DEAL about the data set.

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.) our data is with respects to the mean

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

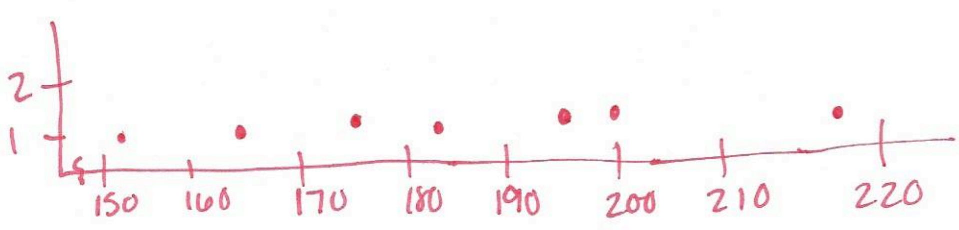
You want to find out how consistently rubber bands will travel when launched, so you use a ruler to launch two rubber bands seven times each. You generate the following data sets:

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



Mean: 184

Mean 184



RB# 2

Data Point	Mean	Deviation from Mean	Squared Deviation From Mean
182	184	-2	4
186	184	+2	4
182	184	-2	4
184	184	0	0
185	184	+1	1
184	184	0	0
185	184	+1	1
			14

$$\text{variance} = \frac{\text{squared deviation}}{\# dp - 1} = \frac{14}{6} = 2.33$$

$$SD = \sqrt{\text{variance}} = \sqrt{2.33} = 1.53$$

KB #2 SD = 21.57

σx

Sx

MATH III Notes, Unit 1, Statistics	Name Key

Notes 1-3, Normal Distributions and Percentages	Date ____________ Period ____________

Normal Distributions and Percentages

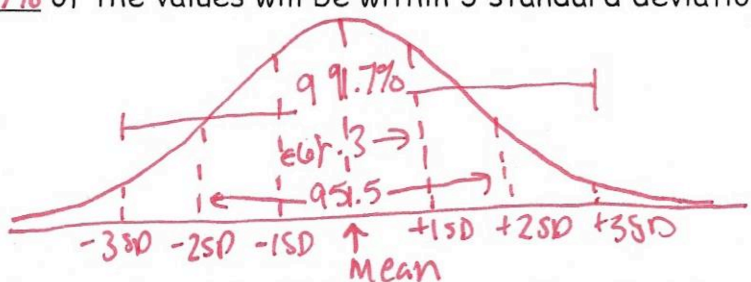
The Empirical Rule - States nearly all values lie within 3 standard deviations from the mean in normal distribution

In any 68-95-99 data set that is normally distributed:

Approx. 68.3% of the values will be within 1 standard deviation of the mean

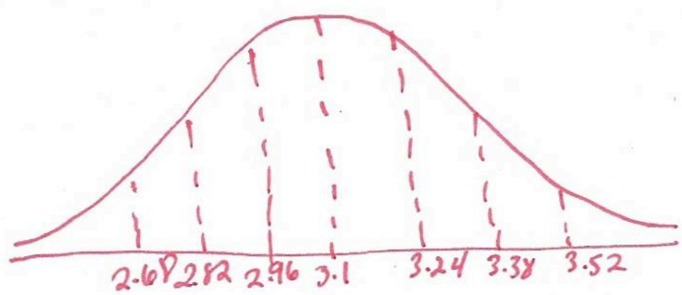
Approx. 95.5% of the values will be within 2 standard deviations of the mean

Approx. 99.7% 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)? 342
2.82g and 3.38g (i.e., within two standard deviations of the mean)? 478
2.68g and 3.52g (i.e., within three standard deviations of the mean)? 499

$\frac{x}{500} = \frac{\%}{100}$

What if I wanted to know the percentage of pennies that had a weight between 3g and 3.2g? 52.5% $\frac{x}{100} = \frac{52.5}{100}$ 263 pennies

Calculator Function: **normalcdf()**

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

and all you have to give it is: interval 3 - 3.2

mean 3.1

SD 0.14

(Note that normalcdf assumes that your data set is normally distributed.)

The format of the normalcdf() function is:

normalcdf(lower bound, upper bound, mean, SD)

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, 0.14)

Name ______________________
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

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

Convenience

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

Systematic

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 2 issues
- Using double negatives
- Overlapping answer choices
- words that cause strong reaction (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

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

Leading

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

2 Issues

Name ______________________
Date ____________ Period ____________

Margin of Error

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

It is a measure of how close we believe the sample proportion is to the population proportion.

The formula used to predict MOE with 95% confidence ≈ $\frac{1}{\sqrt{n}}$.

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?

   The attacks on September 11th

2. Find the MOE in the August poll.

   $\frac{1}{\sqrt{n}} = \frac{1}{\sqrt{1000}} = \frac{1}{31.622} \approx .032$ or 3.2%

3. Find the MOE in the September poll.

   $\frac{1}{\sqrt{n}} = \frac{1}{\sqrt{100}} = \frac{1}{10} = .10$ or 10%

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

   The total number of people asked was greater.

Example 3: Fundraisers

At a raffle, 500 tickets are sold at $1 each for three prizes of $100, $50, and $10. What is the expected value of your net gain if you buy a ticket?

Gain, X	$100 - $1 or $99	$50 - $1 or $49	$10 - $1 or $9	$0 - 1 or -$1
Probability P(x)	$\frac{1}{500} = 0.002$	$\frac{1}{500} = 0.002$	$\frac{1}{500} = 0.002$	$\frac{497}{500} = 0.994$

EV = (0.002*99) + (0.002*49) + (0.002*9) + (0.994*(-1)) = -$0.68

Example 4: Water Park

A water park makes $350,000 when the weather is normal and loses $80,000 per season when there are more bad weather days than normal. If the probability of having more bad weather days than normal this season is 35%, find the park's expected profit.

Gain, x	350,000	-80,000
Probability, P(x)	.65	.35

EV = (.65 * 350,000) + (.35 * -80,000)
= $199,500

Example 5: MP3 Players

Construct a probability distribution and find the expected value:

Students were asked how many MP3 players they own.

Players, x	Frequency	P(x)
0	9	0.214
1	17	0.405
2	9	0.214
3	5	0.119
4	2	0.048

EV = (.214*0) + (.405*1) + (.214*2) + (.119*3) + (0.048*4)

= 1.382

Calculator Stat, Enter, Put values in L1, L2, Stat, calc, 2varstat
Down
$\sum xy$

Probability Distribution

Name ______________________

Date __________ Period __________

Probability Distribution can be a table, graph, or equation that links each possible outcome of an event with its probability of occurring.

- The probability of each outcome must be between 0 and 1.
- The sum of all the probabilities must equal 1.

Making a Probability Distribution

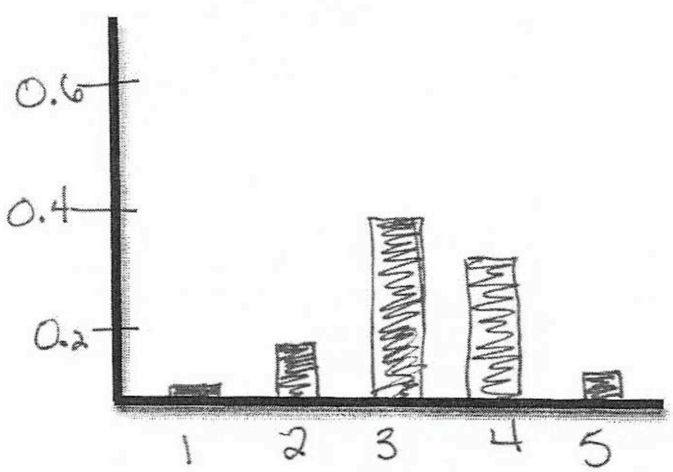
Example 1: Bakery

A bakery is trying a new recipe for the fudge deluxe cake. Customers were asked to rate the flavor of the cake on a scale of 1 to 5, with 1 being not tasty, 3 being okay, and 5 being delicious. Use the frequency distribution show to construct and graph a probability distribution.

Step 1: Find the proabability of each score.

Score, x	Frequency	P(x)
1	1	1/50 = .02
2	8	8/50 = .16
3	20	20/50 = .40
4	16	16/50 = .32
5	5	5/50 = .10
	50	

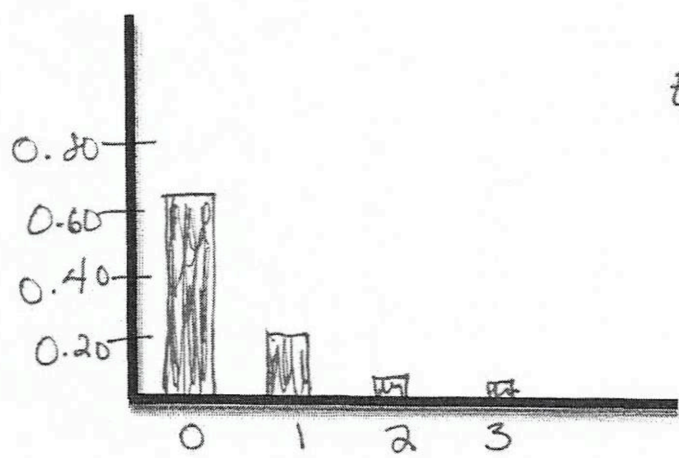
Step 2: Graph the score versus the probability.



Example 2: Car Sales

A car salesperson tracked the number of cars she sold each day during a 30-day period. Use the frequency distribution of the results to construct and graph a probability didstribution for the random variable x, rounding each proability to the nearest hundreth.

Cars Sold, x	Frequency	P(x)
0	20	$20/30 = 0.667$
1	7	$7/30 = 0.233$
2	2	$2/30 = 0.067$
3	1	$1/30 = 0.033$



$EV = P_1 X_1 + P_2 X_2 + P_3 X_3 + P_4 X_4$

Expected Value

$= (.667 * 0) + (.233 * 1) + (.067 * 2) + (0.033 * 3)$

$= 0 + .233 + .0335 + .099$

$= 0.466$

Expected Value

In a random experiment, the values of the n outcomes are $X_1, X_2, X_3, \cdots X_n$ and the corresponding probabilities of the outcomes occurring are $p_1, p_2, p_3, \cdots p_n$.

The expected value (EV) of the experiment is given by:

$EV = p_1 X_1 + p_2 X_2 + p_3 X_3 + \cdots + p_n X_n$

To calculate expected value:

- Start with the probability Distribution or create it if you don't have it.
- Multiply the value of each outcome by it's probability.
- Add up all those products.
- The sum is the expected value.