Math II Statistics Packet Name ______________________

Objectives: 1. To find the measures of central tendency.

2. To construct box and whisker plots and use a box plot to find the lower quartile, upper quartile, median and outliers.

3. To find variation and standard deviation.

4. To understand the normal curve.

In this section we are going to examine data and their centers and spread. We will be doing some “number crunching” with data sets to help us to get “a feel” for the data:

We are going to find values that help to answer the following questions:

1. What is the "typical" grade for the class?
2. How did most students do?
3. Did many of the students perform markedly different from most of the class?

CENTER OF THE DATA

Where is the center of the data set? Do the numbers tend to converge about a certain value? Is there a single value that describes the data set?

We call these **Measures of Central Tendency** and we will find 3 main types.

They are the **mean, median, and mode**.

ARITHMETIC MEAN ($\bar{x}$)

Susan earned the following percents on her seven tests during sophomore chemistry. What is her average?

75, 76, 80, 80, 86, 88, 96

$$\bar{x} = mean = \frac{75+76+80+80+86+88+96}{7} = 83$$

DEFINITION of MEAN

The arithmetic mean of the numbers x_1, x_2, x_3, …,x_n, denoted by $\bar{x}$ and read "x bar" is given by

$$\bar{x} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n}.$$

UNDERSTANDING THE MEAN

The mean is the most commonly used type of central tendency.

Write the name of each of the following states on a strip of graph paper. Use one strip for each state and one square for each letter in the name. Cut off the unused squares on the end of each strip.

NEVADA MONTANA VIRGINIA WISCONSIN MINNESOTA

To find the **mean** of the numbers of letters in the names, Cut off letters from the longer names and move then to fill in the shorter ones. Continue cutting off and moving letters until all the rows have as close to the same number of letters as possible.

N	E	V	A	D	A			
M	O	N	T	A	N	A		
V	I	R	G	I	N	I	A	
W	I	S	C	O	N	S	I	N
M	I	N	N	E	S	O	T	A

Try this--The mean average of 21 people in a room is $21,000. The President of the company walks in and his salary is $109,000. What is the mean salary of the people in the room?

MEDIAN

The value that is exactly in the middle of an ordered set of data

The median is found by:

1. arranging the numbers in order from least to greatest.
2. If n is odd, the median is the middle number. If n is even the median is the mean of the two middle numbers.

Find the median for the following scores

74, 79, 83, 85, 90, 99

The median is 84

MODE.

The mode of a set of numbers (data) is the number that appears the most frequently.

Can there be more than one mode? If a set of data has two values that appear the most often we say the data is **<u>bimoda</u>**l. It is possible for a set of data to have no mode.

Find the mode of the following test scores:

62, 94, 95, 98, 98, 98

<u>TRY THESE</u>

For #1 and 2 find the mean, median and mode.

1. 50, 60, 70, 85, 90, 90, 95, 100

2. 80, 82, 84, 84, 86, 88

WHICH IS BEST?

3. Susan worked for a small company with five employees. The following are their salaries. Find the mean, median and mode, then decide which is the best average.

$22,000, $14,000, $22,000, $20,000, $157,000

NOTE--the mean is more affected by extreme values.
the mode is always a member of the original set of data

Try these:

4. Suppose a company employs 20 people. The president of the company earns $200,000, the VP earns $75,000, and the 18 employees each earn $10,000. What is the mean? Is this the best number to represent the "average" salary of the company? What is the median and mode?

5. Suppose you own a hat shop and decide to order hats in only *one* size for the coming season. To decide which size to order, you look at last year's sales figures, which are itemized according to size. Should you find the mean, median or mode for the data? Why?

6. Students of Dr. Van Horn were asked to keep track of their own grades. One day, Dr. Van Horn asked the students to report their grades. One student had lost the papers but claims to remember the grades on 4 of the 6 assignments: 100, 82, 74, and 60. In addition, the student remembered that the mean of all 6 papers was 69 and the other 2 papers had identical grades. What were the grades on the other 2 papers?

7. Faith has an average of 76% of her 3 chemistry tests. What grade would she have to make on the 4^{th} test to have an average of 80%?

8. A class of 23 students had a mean of 78 on a chemistry exam. The 10 boys in the class had a mean of 76.2. What was the mean of the girls' test scores? Round to 2 decimal places

GRADE POINT AVERAGE GPA

Find the grade point average for the following report:
Assume that the grade point values are 4.0 for an A, 3.0 for a B, etc.

Course	Credits	Grade
T102	3	B
FYS	1	A
M333	2	C
P170	3	C
E241	3	A

Try this—
1. What if this student put more effort into the P170 course and less in the FYS?

Course	Credits	Grade
T102	3	B
FYS	1	C
M333	2	C
P170	3	A
E241	3	A

2. Find the mean, median and mode for sets A and B.

Set A	Set B
28	24
25	23
22	22
22	22
19	21
16	20

The mean, median, and mode provide limited information about the whole distribution of the data. To tell how scattered or how far dispersed or spread out data are, we use measures of spread or dispersion. The Measures of Dispersions that we are using are the range, interquartile range, variance, and standard deviation.

Measures of Dispersion

Range-the difference between the greatest and the least value in a data set. But be careful, just because the range of one data set equals that of another does not mean that they have the same dispersion.

For example: Data Set 1: 1, 3, 7, 8, 10 Range =

Data Set 2: 1, 10, 10, 10, 10 Range =

Note that the spread of these 2 data sets is quite different although they have the same range. For this reason we need other measurements of dispersion besides the range.

Interquartile range(IQR)-is the range of the middle half of the data.

Consider the numbers:

18 27 34 5 2 54 59 61 6 8 78 82 85 8 7 91 93 100

The range is:________ Median is:_____ Mark this Q_2

Median of left scores is:______ Mark this lower quartile-Q_1-(first)

Median of the right scores is: _____ Mark this upper quartile-Q_3-(3^{rd})

We have now divided the data into four groups separated by 3 marking points. Notice that 50% of the data lie between Q_1 and Q_3. This is the **INTERQUARTILE**. The IQR can then be calculated by finding $Q_3 – Q_1$. Find the IQR.

IQR = Q_3-Q_1=______

25% of the scores lie at or below Q_1

75% of the scores lie at or below Q_3

The IQR contains the middle 50% of the values.(data)

This IQR is a useful measure of dispersion because it is less influenced by extreme values. They are typically used instead of line plots when there is a large amount of data. They are very useful when comparing 2 sets of data.

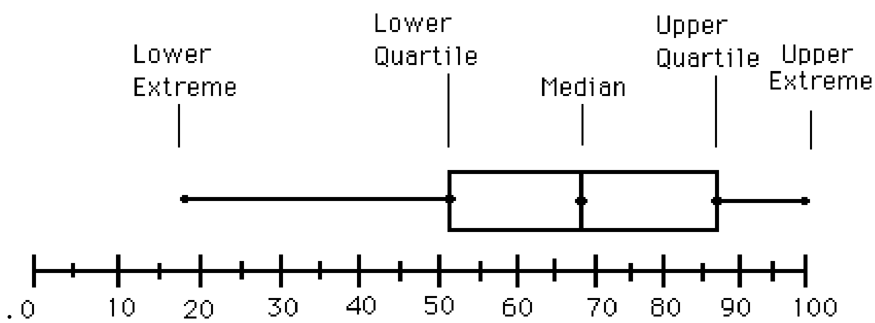
Box and Whisker Plot

A box and whisker plot is a way to display data visually and draw informal conclusions. The box and whisker plot shows only certain data-**five number summary**.

Median, upper and lower quartile, least and greatest values in the distribution.

18 27 34 5 2 54 59 61 6 8 78 82 85 8 7 91 93 100

Draw a box and whisker plot for the above data.



TRY THIS

Make a box and whisker plot for this data:

20, 25, 40, 50, 50, 60, 70, 75, 80, 80, 90, 100, 100

Outliers

Outliers are values that are widely spread from the rest of a group.
Outliers are numbers that are:
technically defined as any number that is **more than 1.5 times the IQR above Q_3 or below Q_1.**
Make a box and whisker plot for this data. Find any outliers.

10 20 50 80 85 88 88 90 92 94 96 96 98 100

You try this:
Construct a box and whisker plot. Identify any outliers.

Etruscan skull sizes

126	132	138	140	141	141
142	143	144	144	144	
145	146	147	148	148	149
149	150	150	150	154	
155	158	158			

So the boxplot is: (Note 126 is an outlier)

```
                               --------------
    *         -----------------I     +      I-------------
                               --------------
  --+---------+---------+---------+---------+---------+----
126.0     132.0     138.0     144.0     150.0     156.0
```

Look at the box plot in your book page 640-641. Note how it allows the observer to compare sets of data.

Variance and Standard Deviation

Variance and standard deviation-are closely related and measure how far the data is from the mean.

Let's look at the following set of numbers. Fill in the table below.

<u>Set A</u>	$x - \overline{x}$	$(x - \overline{x})^2$
28		
25		
22		
22		
19		
16		

Variance-is the mean of the squared deviations or the sum of the squares of the differences between each data value and the mean divided by the number of values in the data set, *n*.

$$V = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + ... + (x_n - \overline{x})^2}{n}$$

STANDARD DEVIATION Denoted by the letter *s*, the standard deviation is the square root of the *variance*.

$$S = \sqrt{v} = \sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + ... + (x_n - \overline{x})^2}{n}}$$

Find v and s for Set A above v = _____ s = _____

Try this--

Find v and s for Set B

Set B	$x - \overline{x}$	$(x - \overline{x})^2$
24		
23		
22		
22		
21		
20		

The standard deviation is a larger number when the values from a set of data are widely spread and a smaller number (close to 0) when the data values are close together.

Mean Absolute Deviation MAD

The mean absolute deviation makes use of the absolute value to find the distance each data point is away from the mean.

We find the mean absolute value by:

1. Measure the distance from the mean $(x - \overline{x})$
2. Find the absolute value of each difference
3. Sum the differences
4. Divide this sum by the number of scores

MAD = $$\frac{|x_1 - \overline{x}| + |x_2 - \overline{x}| + |x_3 - \overline{x}| + |x_4 - \overline{x}| ... + |x_n - \overline{x}|}{n}$$

TRY THIS: Find the MAD for the numbers:

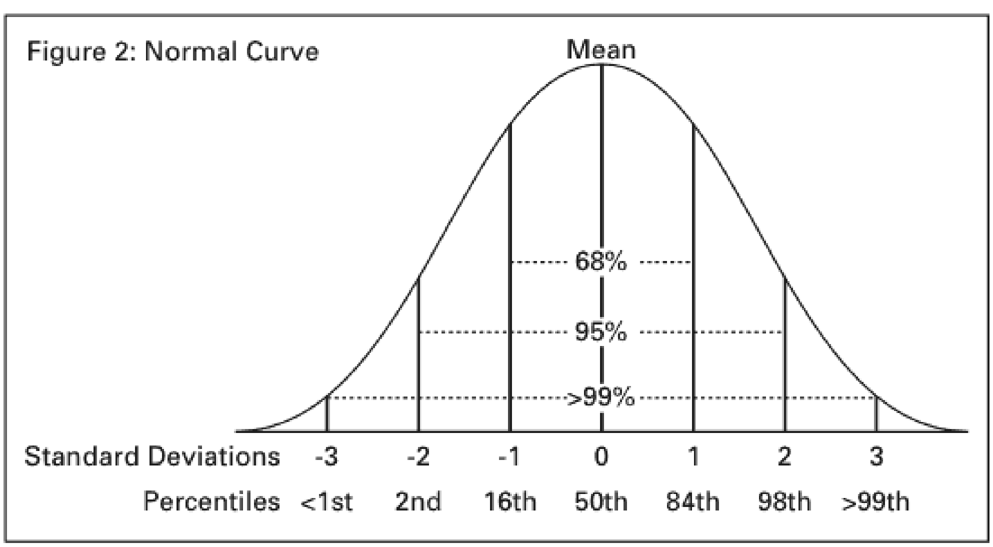
96, 71, 43, 77, 75, 76, 61. 93

Normal Distribution

To better understand how standard deviations are used as measures of dispersion, we next consider normal distributions. The graphs of normal distributions are the normal curves that we call bell curves. These curves are used when describing IQ scores or standardized test scores.

A **normal curve** is a smooth bell-shaped curve that depicts the frequency of the data values symmetrically about the mean. (NOTE: When using a normal curve, the mean, median and mode all have the same value.)

Normal Distribution



NOTE:

1. Mean, median and mode have the same value

2. Curve never reaches zero

3. 68% of the value lie between one standard deviation of the mean
   This means 34% to the left and 34% to the right of the mean.

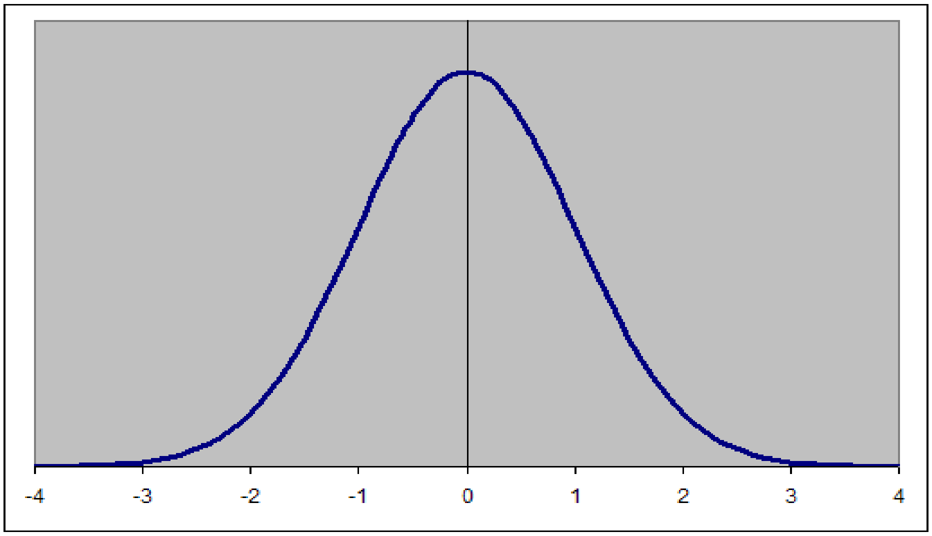
4. 95% of the value lie between two standard deviation of the mean
   This means ____% to the left and _____% to the right of the mean

5. 99.8% of the value lie between three standard deviation of the mean This means ________% to the left and ________% to the right of the mean

6. Total 100%

EXAMPLE: For a population with a normal distribution with a mean of 24 and a standard deviation of 3,

a. about 68% of the population lie between what limits?
b. about 95% of the population lie between what limits?
c. about 99.8% of the population lie between what limits?



EXAMPLES:

On a certain exam, the mean is 72 and the standard deviation is 9. If a grade of A is given to any student who scores at least two standard deviation above the mean, what is the lowest score that a person could receive and still get an A?

A tire company test a particular model of tire and found the tires to be normally distributed with respect to wear. The mean was 28,000 miles and the standard deviation was 2500 miles. If 2000 tires are tested, about how many are likely to wear out before 23,000 miles?

We have all taken standardized examination such as the SAT or the ACT. When given your score, you are also given a **percentile**. The **percentile** shows your score in relation to all the other scores. If you are in the 92nd percentile this means that approximately 92% of those taking the exam scored lower than you and 18% score higher than you.

A **percentile** shows a person's score relative to other scores. For example, if a student 's score is at the 75th percentile, this means that approximately 75% of those taking the test scored lower that the student and approximately 25% had a higher score.

Susan was upset because she scored in the 66th percentile on her SAT. She said, "When I was in school, 66% was a D or an F. I needs to do MUCH BETTER THAN THAT!"

a. Is her score:
   BETTER THAN AVERAGE OR AVERAGE OR BELOW AVERAGE?

b. What would you tell her about being in the 66th percentile?