

## Math 114 – Review for Exam III – Chapter 9 – Statistics

For Exam III you should understand the following concepts and be able to do problems like the examples shown.

(1.) **Identifying a variable as discrete or continuous.**

- a. Discrete Variable: a variable that can take on only a finite number of values or a countable number of values.
- b. Continuous Variable: a variable that can take on any of the countless number of values in a line interval – i.e. real numbers

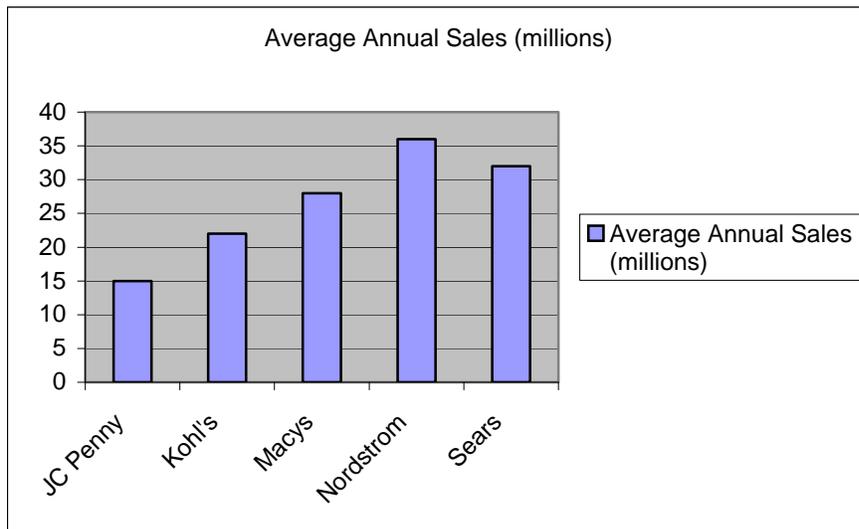
Examples: Identifying a variable as discrete or continuous

Variable	Continuous or Discrete?
<b>Measure</b> the time it takes a student selected at random to register for the fall term	Time can take on any value, so this is a <b>continuous</b> variable
<b>Count</b> the number of bad checks drawn on Upright Bank on a day selected at random	The number of bad checks can only be a whole number such as 0,1,2,... This is a <b>discrete</b> variable
<b>Measure</b> the amount of gasoline needed to drive your car 200 miles	Measuring volume, which can assume any value, this is a <b>continuous</b> variable
Pick a random sample of 50 registered voters in a district and find the number who voted in the last presidential election	This is a count so the variable is <b>discrete</b> .

- (2.) **Random/Unbiased Sample** – To ensure you have a random or unbiased sample you need to make sure that every member of the population has an equal chance of being included in the sample. A biased sample favors some parts of the population over others.
- (3.) **Constructing A Bar Graph** – In a bar graph the horizontal axis is used to define the category and the vertical axis is used to define the frequency. For instance, the following data can be arranged into a bar graph:

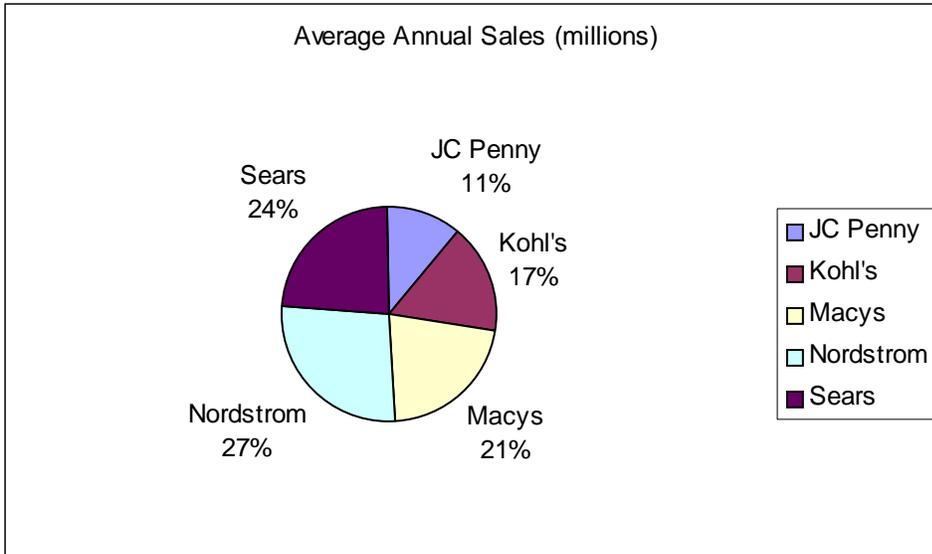
### Average Annual Store Sales

Store	Average Annual Sales (millions)
JC Penny	15
Kohl's	22
Macys	28
Nordstrom	36
Sears	32
Totals	133



- (4.) **Constructing A Pie Chart** - A pie chart shows a circle divided into sectors - each sector is a category and the size of the sector depends on the frequency of that category. We can take the data from the bar graph shown and graph the data as a pie chart. You need to calculate the fractional part (sales for a specific store/total sales), Percentage, and Number of degrees (percentage x 360 degrees - total degrees in a circle)

Store	Average Annual Sales (millions)	Fractional Part	Percentage	# of Degrees Percentage * 360
JC Penny	15	15/133	11%	40
Kohl's	22	22/133	17%	61
Macys	28	28/133	21%	76
Nordstrom	36	36/133	27%	97
Sears	32	32/133	24%	86
Totals	133		100%	360



- (5.) **Frequency Charts** - A frequency chart shows the distribution of data into classes or intervals. The classes or intervals are constructed so that each data value falls into exactly one class.

For example the list below shows One-way Commuting Distances (in miles) for 60 workers in downtown Dallas.

13,7,12,6,34,14,47,25,45,2,13,26,10,8,1,14,41,10,3,21,8,13,28,24,16,19,4,7,36,37,20,  
 15,16,15,17,31,17,3,11,46,24,8,40,17,  
 18,12,27,16,4,14,23,9,29,12,2,6,12,18,9,16

To construct a frequency chart for this data first you need to define your class width.

**How to find the Class Width**

- 1.) **Class Width =  $\frac{\text{Largest Data value} - \text{Smallest Data Value}}{\text{Desired number of classes}}$**
- 2.) **Increase the computed value to the next highest whole number.**

So if I wanted 6 classes for my chart the class width would be equal to:

Class Width =  $(47-1)/6 = 7.7$ , now increase the computed value to next highest whole number, 8. Now construct the frequency chart by using the defined classes and listing the number of commuting distances within that specific class.

Class Intervals(miles)	Frequency
1-8	14
9-16	21
17-24	11
25-32	6
33-40	4
41-48	4

From this chart you can make a **histogram and a frequency polygon**. You should know how to construct these graphs from a frequency chart. Remember columns in a histogram should touch each other - no spaces.

(6.)

**Measures of Central Tendency** estimate the “center” of the data collected. The 3 most common measures of central tendency are Mean, Median, & Mode.

**Computing the Mean, Median, and Mode**

**a. Definitions**

**Mean** = Sum of all data points/Number of data points

The **Median** of a set of real numbers arranged in order of magnitude is the middle value if the number of items is odd, and it is the mean of the two middle values if the number of items is even.

**Mode** - the most frequent value in a set of data

**b. Calculating Mean, Median, & Mode for a set of data**

Find the mean, mode, and median for the following quiz scores:

2,4,4,6,7,8,5,3,7,9,10, 8,9,4,11

First list data in increasing order: 2,3,4,4,4,5,6,7,7,8,8,9,9,10,11

Mean =  $97/15 = 6.5$

Median = 7 (**must use data listed in increasing or decreasing order!**)

Mode = 4

c. **Calculating Mean & Mode for grouped data** – for this example let’s use a modified version of the one-way commute data.

Class Intervals(miles)	Frequency ( $f_i$ )	Midpoint ( $m_i$ )	$f_i m_i$
1-9	14	5	70
10-18	21	14	294
19-27	11	23	253
28-36	6	32	192
37-45	4	41	164
46-54	5	50	250
Totals	61		1223

$$\text{Mean} = \frac{\sum f_i m_i}{n} = 1223 / 61 = 20$$

$$\text{Midpoint} = m_i = \frac{\text{UpperLimit} + \text{LowerLimit}}{2}$$

The mean of two sets of data could be equal but the data itself could be very different. For example, suppose there were two different quizzes given to a class with nine students and suppose the quiz scores were as follows:

Quiz # 1 : 7,7,8,8,9,9,10,10,10

Quiz #2: 4,5,9,9,9,10,10,11,11

Both quiz scores have a mean of 8.7 but as you can see the scores for quiz #2 are more spread out. To measure the extent to which data is spread out we use **Measures of Dispersion**.

**(7.) Measures of Dispersion:** Measures of Dispersion include Range, Variance, and Standard Deviation.

**Range** = The simplest measurement of dispersion and defined as the difference between the largest data value and the smallest data value. For example: if the scores for a quiz were 4,6,7,7,8,9,9,10,10 then the range would = 10-4=6.

**Variance** = a measure of deviation from the mean. It is calculated by first determining the deviation from the mean for each data point and then squaring this value. Then adding the squares of the deviation and dividing by either the number of data points (N) for population data or the number of data points -1 (n-1) for sample data.

**Standard Deviation** = the square root of the variance.

### Calculating Sample Mean, Sample Variance, Sample Standard Deviation

Use Sample Mean, Sample Variance, and Sample Standard Deviation formulas when you are **working with a sample of data** from an entire population.

$$\text{Sample Mean} = \bar{X} = \frac{\sum x}{n} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

$$\text{Sample Variance} = S^2 = \frac{\sum (x - \bar{X})^2}{(n-1)} = \frac{(x_1 - \bar{X})^2 + (x_2 - \bar{X})^2 + (x_3 - \bar{X})^2 + \dots + (x_n - \bar{X})^2}{(n-1)}$$

$$\text{Sample Standard Deviation} = S = \sqrt{S^2}$$

where x = data point in sample & n = the number of data points in sample

### Calculating Population Mean, Population Variance, Population Deviation

Use Population Mean, Population Variance, and Population Standard Deviation formulas when you are **working with all the data** from a population.

$$\text{Population Mean} = \mu = \frac{\sum x}{N} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{N}$$

$$\text{Population Variance} = \sigma^2 = \frac{\sum (x - \mu)^2}{N} = \frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + (x_3 - \mu)^2 + \dots + (x_N - \mu)^2}{N}$$

$$\text{Population Standard Deviation} = \sigma = \sqrt{\sigma^2}$$

Where: N= the number of data in the entire population, x = data point in population

(9.)Normal Distribution - You should understand the properties of Normal Distribution (Section 9.6 pg 514) including how to calculate a Z score.

$$Z = \frac{x - \mu}{\sigma} = \text{(Difference between x and } \mu\text{)/Standard Deviation}$$

where:

x = Original data point

μ = Mean of the original data

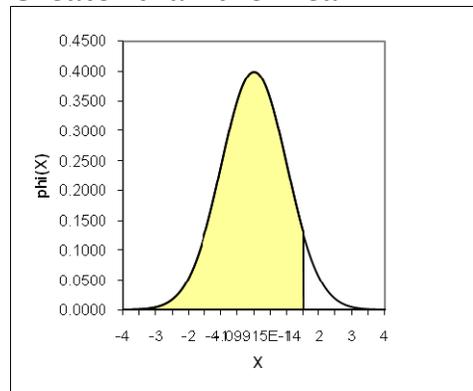
σ = Standard Deviation of the original data

You should also know how to use Z values to find the corresponding area under the Normal Distribution Curve by using the Standard Normal Curve Table in the back of your textbook. (See Example 5 - page 519 in your textbook)

Remember your table of Z values and area gives you the area between the mean and a Z value. Only positive Z values are given but the area for the corresponding negative Z values is the same as the positive Z value because the normal curve is symmetrical. (i.e  $Z = 1.5$ , area = .4332 and  $Z = -1.5$ , area = .4332)

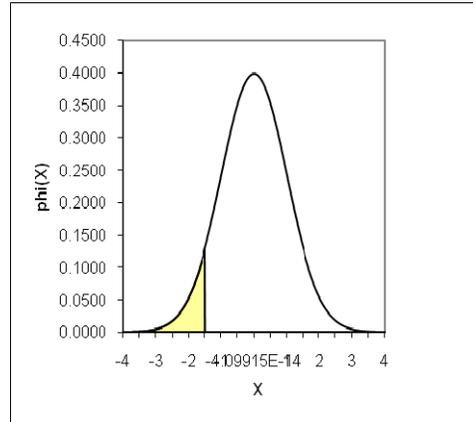
When asked to find the area under the curve for an x value or two x values always draw a picture first!! Then follow these rules, if x is:

Greater than the mean:



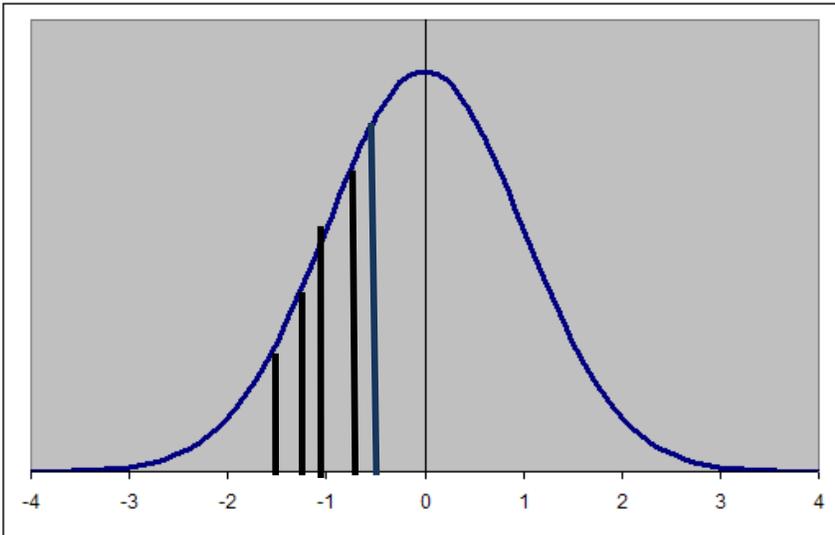
The Z score will be positive. The area to the left of the Z score (shaded) will be the area between your Z score and the mean (the value from your table) PLUS .50 and the area to the right of the Z score will be .50 minus the area between your Z score and the mean.

Less than the mean:

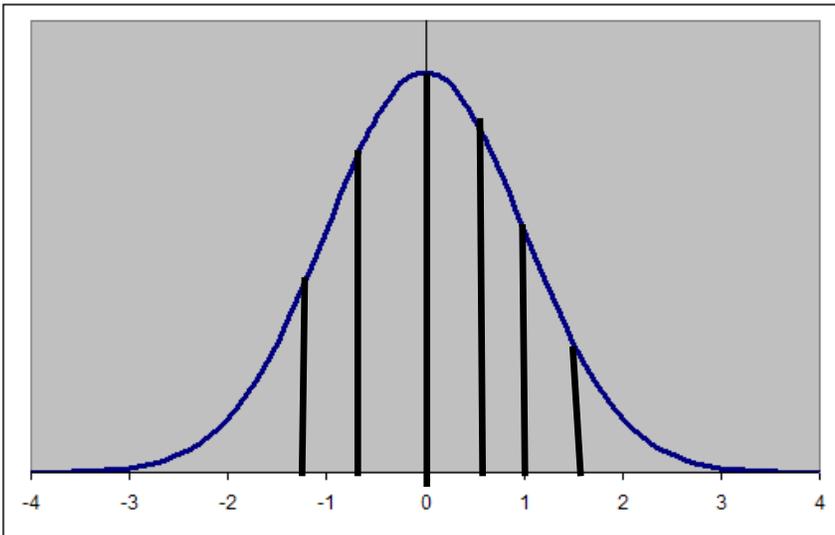


The Z score will be negative. The area to the left of the Z score (shaded) will be .50 minus the area between your Z score and the mean (take the absolute value of your Z score and read area from table) and the area to the right of the Z score will be .50 plus the area between your Z score and the mean (take the absolute value of your Z score and read area from table).

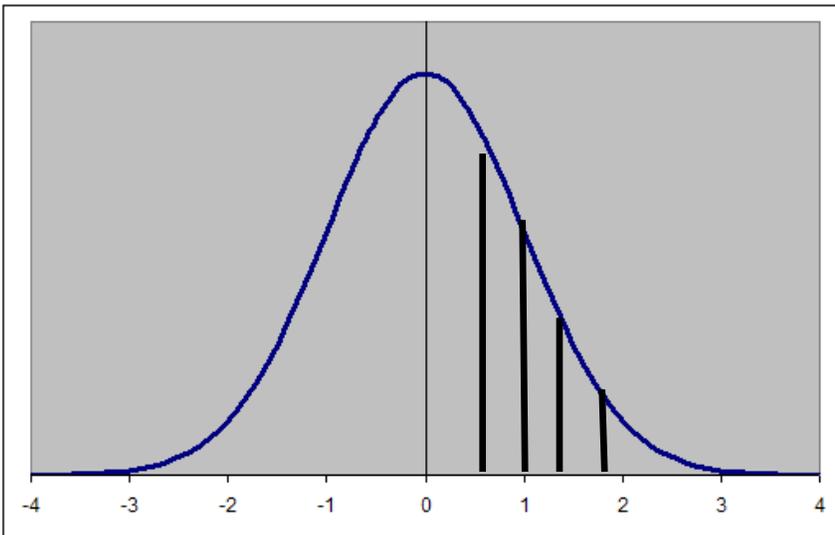
Between two values of  $x$ :



If both  $x_1$  and  $x_2$  are both less than the mean subtract the areas between your Z scores and the mean.



If  $x_1$  is less than the mean and  $x_2$  is greater than the mean add the areas between your Z scores and the mean.



If both  $x_1$  and  $x_2$  are both greater than the mean subtract the areas between your Z scores and the mean.

