Name: _____ Date: _____

THE DIFFERENCE IN SAMPLE MEANS
COMMON CORE ALGEBRA II

In a classic **experiment** two or more **treatment groups** are randomly created and then subjected to the **treatments**. The question then is how to determine if the variability seen between the two groups is due to the treatment.

***Exercise* #1:** Suppose 50 people were chosen to try out a new diet pill. The people are randomly divided into two groups. One is given the pill while the other is given a **placebo** (a pill designed to look the like real one, but with no medicine). There are two main ways variability can be introduced to the results. Discuss each.

Induced (Variability created because of the treatment the subject was placed in):

Natural (Variability just because people, animals, plants, etcetera, are naturally different):

The question, then, is how we can distinguish between the two types. Let's look at a case study.

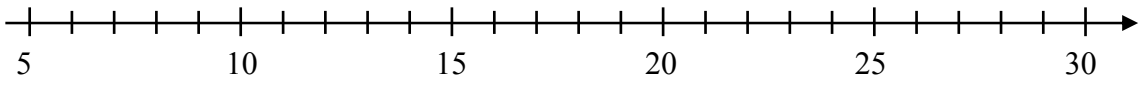
***Exercise* #2:** A seed company is trying to determine the effect of synthetic nutrients versus organic nutrients on the growth rate of corn plants. They select 40 seeds and randomly distribute the seeds to two groups of 20. The seeds in Group 1 are given the organic nutrients and the seeds in Group 2 are given the synthetic nutrients. After three weeks each plant's growth is measured in centimeters.

Group 1 (Organic): 6, 8, 10, 12, 12, 12, 13, 13, 14, 14, 16, 16, 17, 17, 18, 18, 20, 20, 22, 25

Group 2 (Synthetic): 9, 11, 12, 12, 15, 15, 15, 16, 17, 18, 18, 18, 19, 19, 19, 21, 21, 22, 24, 28

Enter these two lists in your calculator. State the mean of each. Then, create a box plot for each using the grid below. You may want to summarize the information you need under each heading.

Group 1 Group 2



***Exercise* #3:** Let's look at the **descriptive statistics** we have so far: the sample means and the box-plot. What does this data suggest? Does it give a clear indication that one treatment resulted in greater plant growth?

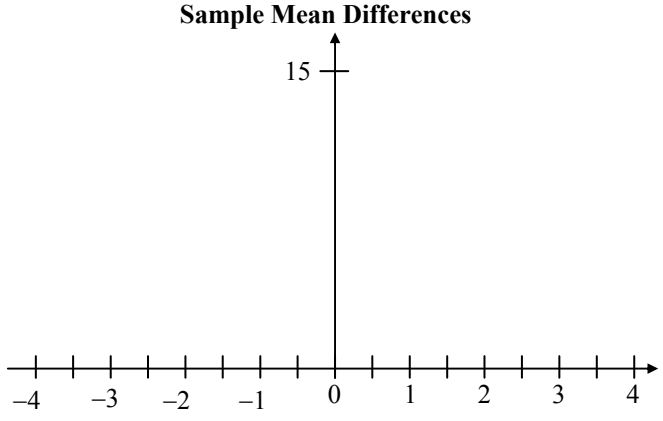
There are very sophisticated techniques to probabilistically determine what portion of the variability in these two data sets is due to natural causes and what is **induced**. But, we can run a simple simulation which can give us a very good sense. Consider just the question of the **difference in the sample means**. The program MEANCOMP will take our two groups of data and randomly scramble them up into two new groups. It will do that over and over again and calculate the difference in the means of the groups.

***Exercise* #4:** If $\overline{x_1}$ represents the mean of Group 1 and $\overline{x_2}$ represents the mean of Group 2, do the following.

(a) Find the **observed difference** in the sample means:

$\overline{x_2} - \overline{x_1} =$

(b) Run the program MEANCOMP with 100 simulations. Use your calculator to create a frequency histogram on the axes below for the sample mean differences. Point out where on the histogram the **observed difference** falls.



(c) Look at the data list containing the sample mean differences. Given there are 100 differences, what percent of the differences were at or above the observed difference?

(d) How confident are you about the **observed difference in sample means** being due to **treatment variability** and not **natural variability**? Justify.

Name: _____ Date: _____

THE DIFFERENCE IN SAMPLE MEANS
COMMON CORE ALGEBRA II HOMEWORK

APPLICATIONS

1. In an experiment, two main types of variability are introduced. Explain how both affect the results of the experiment. Give examples to support your descriptions.

   Treatment Variability

   Natural Variability

2. A simulator takes the data from the various treatments, randomly scrambles them together to create groups that contain mixed treatments. Explain how this helps quantify the question of natural variability.

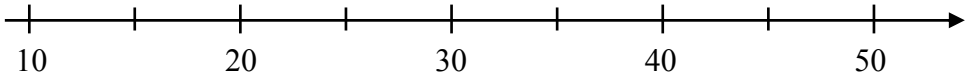
3. Researchers in a sleep lab at a college decide to see how a night of no sleep affected the ability of volunteers to answer 50 addition problems in a minute time span. Thirty volunteers were randomly assigned to two groups. Group 1 was not allowed sleep and Group 2 slept normally. Their results, in terms of questions answered out of 50, are given below. As in the lesson, find the sample means and graph a box plot for each.

   Group 1: 11, 14, 16, 17, 23, 25, 25, 27, 30, 31, 33, 34, 34, 36, 38

   Group 2: 18, 22, 24, 25, 30, 30, 32, 33, 34, 34, 36, 37, 42, 44, 48

   Group 1

   Group 2



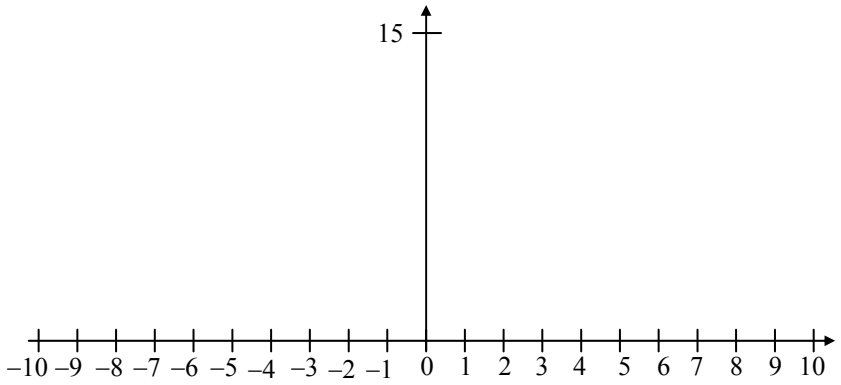
4. Does it appear that getting sleep helps in the ability to answer addition problems? What descriptive statistics can you use to strengthen your argument?

5. Is it true that a person who gets sleep will always answer more addition problems than a person who has not gotten any sleep? Support your answer from the experimental results.

6. Run the program MEANCOMP with 100 simulations on these two data sets. Using your calculator, create a frequency histogram for the sample mean differences on the axes below. Mark on the distribution where the **observed difference** in the sample means lies.

Observed Difference:

$\overline{x}_2 - \overline{x}_1 =$



7. What percent of the simulated differences was our observed difference greater than or equal to? Show your calculation below.

8. Can we **confidently** conclude that the variability in sample means is due to the **treatment** or due to **natural variability**? Support your argument using the distribution above and your answer to 7.