

# 8-4

## Significance of Experimental Results Going Deeper

**Essential question:** In an experiment, when is the difference between the control group and treatment group likely to be caused by the treatment?

**Essential question:** In an experiment, when is a difference between the control group and treatment group likely to be caused by the treatment?

**COMMON CORE Standards for Mathematical Content**

CC.9-12.5.IC.5 Use data from a randomized experiment to compare two treatments; use simulations to decide if differences between parameters are significant.\*

When conducting a z-test, we are actually assuming that the data (or the random errors) follow a normal distribution.

Permutation tests (randomization tests) can be used without the normal assumption for the distribution of data.

### Using a Permutation Test (P-Test)

	SAT Scores				
Treatment Group	1440	1610	1430	1700	1690
	1570	1480	1620	1780	2010
Control Group	1150	1500	1050	1600	1460
	1860	1350	1750	1680	1330

Do students that take an SAT prep course actually improve their SAT scores?

### Using a Permutation Test (P-Test)

	SAT Scores				
Treatment Group	1440	1610	1430	1700	1690
	1570	1480	1620	1780	2010
Control Group	1150	1500	1050	1600	1460
	1860	1350	1750	1680	1330



a.) State the null and alternative hypothesis in terms of the difference of the 2 group means.

**H<sub>0</sub>:** The difference of the 2 group means is 0.

**H<sub>a</sub>:** The difference between the 2 group means (treatment - control) is positive.

### Using a Permutation Test (P-Test)

	SAT Scores				
Treatment Group	1440	1610	1430	1700	1690
	1570	1480	1620	1780	2010
Control Group	1150	1500	1050	1600	1460
	1860	1350	1750	1680	1330



b.) Calculate the mean score for each group

$$\bar{x}_T = 1633$$

$$\bar{x}_C = 1473$$

$$\bar{x}_T - \bar{x}_C = 160$$

test statistic

If the PREP course has NO effect at all, then if we rearrange the data in the two group, we should expect the same results.

**SIMULATION:**

We are going to randomize the data before putting them into 2 groups. We then will find the difference in the means between the 2 groups.

Random number generator on graphing calc to select 10 data to be on group A. The rest will be in group B.

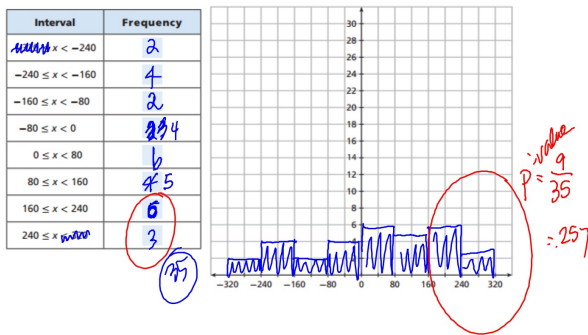
Make sure you select 10 unique random numbers. Hit enter until you do.

		SAT Scores				
Treatment Group		1440	1610	1430	1700	1690
		1570	1480	1620	1780	2010
Control Group		1150	1500	1050	1600	1460
		1860	1350	1750	1680	1330

		SAT SCORES				
A	1690					} A $\bar{x}$ A-B
B						

**Using a Permutation Test (P-Test)**

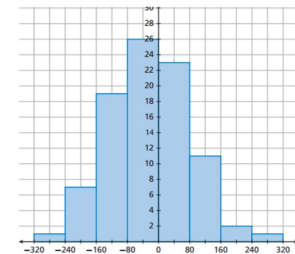
d.) Collect the data from students



**Using a Permutation Test (P-Test)**

e.) Look at the histogram or frequency table to see how many times our test statistic of 160 or more occurred. (this is a sample of what the simulations would also look like)

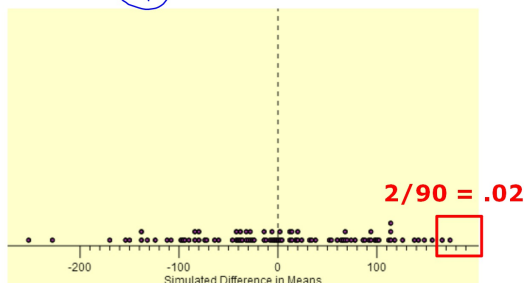
Interval	Frequency
$-320 \leq x < -240$	1
$-240 \leq x < -160$	7
$-160 \leq x < -80$	19
$-80 \leq x < 0$	26
$0 \leq x < 80$	23
$80 \leq x < 160$	11
$160 \leq x < 240$	2
$240 \leq x < 320$	1



p value:  $\frac{3}{90} = 3.3\%$   
.033

**Using a Permutation Test (P-Test)**

d.) Here is another way to represent the data. This is a DOTPLOT representing how many times our simulated value of 160 or more could occur in 90 simulations

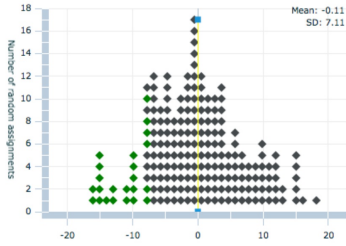


**Using a Permutation Test (P-Test)**

f.) Use the p-value to determine the significance of the experiment.

- $p > .10$ : **NOT significant**  
support null, reject alternative
- $0.05 < p \leq .10$ : **Marginally significant**  
support alternative, reject null
- $0.01 < p \leq .05$ : **Significant**  
support alternative, reject null
- $p \leq .01$ : **Highly significant**  
support alternative, reject null

In reality, a program is used to do the simulation. The simulation data will look something like this:



★ A study investigates how many numbers in a list a person can memorize in order in 30 seconds. Volunteers are randomly assigned either to a group simply shown the list or to a group taught a memorization strategy before being shown the list. A report on the study's results includes the statement, "After comparing the means of the groups, the null hypothesis is rejected with a  $P$ -value of 0.01." What does this statement say about the results of the study?

Teaching memorization strategy has a significant effect on how a person can memorize a list in a short amount of time.

The results of an experiment are resampled 80 times by randomly assigning them to a "control" or a "treatment" group, and the differences in the means are found for each resampling. The table gives the results of the differences  $x$ .

4. In the original experiment, the difference of the means was 8. What is the  $P$ -value for this result?

★ 0.1125  $\frac{9}{80}$

5. Using a one-tailed test, what is the significance of the experimental result?

\_\_\_\_\_

Because  $P$  is  $> 0.1$ , there's not enough evidence to reject the null hypothesis.

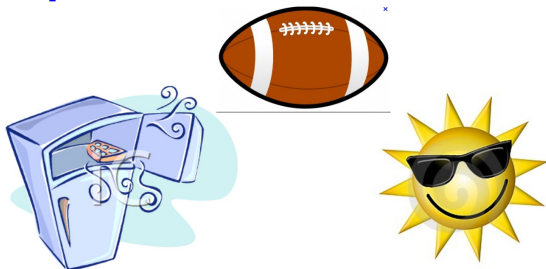
Interval	Frequency
$-16 \leq x < -12$	1
$-12 \leq x < -8$	6
$-8 \leq x < -4$	10
$-4 \leq x < 0$	18
$0 \leq x < 4$	23
$4 \leq x < 8$	13
$8 \leq x < 12$	7
$12 \leq x < 16$	2

## Using a Permutation Test (P-Test)

**The placekicker on a high school football team notices that he seems to kick the ball farther when the weather is warmer compared to when it is cold.**

## Using a Permutation Test (P-Test)

**To conduct the experiment, the kicker will take 10 similar footballs and pick 5 to put in a refrigerator and 5 to put in direct sun for 1 hour.**



## Using a Permutation Test (P-Test)

**After being heated or cooled for 1 hour, each ball will be placed in an identical insulated box to keep the temperature in tact.**



## Using a Permutation Test (P-Test)

A fellow player randomly selects a ball and places it on the tee so the kicker will be blind, thus the kicker will not know if they are kicking a hot or cold ball.

## Data Collection

Here are the results:

Dist (Warm FB): 61 58 63 55 60  $\bar{X}=59.4$   
Dist (Cold FB): 60 52 53 57 59  $\bar{X}=56.2$

Test Statistic (warm-cold) =  $59.4 - 56.2 = 3.2$

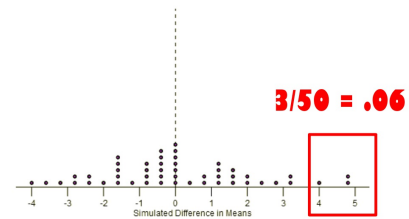
## Using a Permutation Test (P-Test)

The following hypotheses will be tested using the difference in means (warm - cold) as the test statistic.

$H_0$ : there is no difference between the means

$H_a$ : the difference between the means (warm-cold) is a positive number

Here is a sample of 50 simulations!



## Using a Permutation Test (P-Test)

Use the p-value to determine the significance of the experiment.

- $p > .10$ : NOT significant  
support null, reject alternative
- $0.05 < p \leq .10$ : Marginally significant  
support alternative, reject null
- $0.01 < p \leq .05$ : Significant  
support alternative, reject null
- $p \leq .01$ : Highly significant  
support alternative, reject null

With a p-value of 6% our results are **MARGINLY SIGNIFICANT**. Thus we can reject our null hypothesis and support our alternative hypothesis. Thus its plausible that kicking warm footballs go further than cold footballs.