**Activity 7.4.5 – Testing Differences in Two Population Means**

**Does Playing Casual Video Games Reduce Anxiety Among Teenagers?**

A researcher conducted a randomized experiment to determine if playing casual video games reduces anxiety among teenager who suffer from anxiety. The study was conducted on 20 teenagers. 10 teenagers were randomly assigned to an experimental group and played casual video games for one hour, three times a week. 10 teenagers were assigned to a control group and did not play any video games. After one month each participant was given an anxiety post-test. The researcher calculated the change in each teenager’s anxiety scores (post score – pre score). A positive change represents an increase in anxiety; a negative change represents a decrease in anxiety.

The changes are shown below. The dotplot compares the changes in the two groups.

Change in Anxiety Scores for Control Group, *n* = 10

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| -5 | 0 | -3 | -4 | 1 | 2 | -2 | -1 | 5 | 3 |

Change in Anxiety Scores for Treatment Group, *n* = 10

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| -7 | 0 | -6 | -6 | 1 | -5 | -5 | -3 | -2 | 1 |



1. Use technology to calculate the mean change for each group.
2. On average, which group experienced a greater decrease in anxiety?

There are two possible scenarios that could explain the difference in the observed sample means.

**Scenario 1:** The population means are the same (i.e. the change in anxiety levels of teenagers who play casual video games and the change in anxiety levels of those who don't are the same) and the difference in sample means is due to chance alone. This implies that playing casual video games has *no effect* on anxiety levels in teenagers who suffer from anxiety.

**Scenario 2:** The population means are different (i.e. change in anxiety levels of teenagers who play casual video games is greater than the change in anxiety levels of those who don’t) and the difference in sample means reflects the actual difference that exists between the population means. This implies that playing casual video games has *an effect* on anxiety levels in teenagers who suffer from anxiety.

To make a decision on which scenario is most probable, we can perform a randomization hypothesis test using a randomization distribution of *differences in sample means.*

**Randomization Test for a Difference in Population Means**

* Assume the population means are equal.
* Construct a randomization distribution of differences in sample means from random samples from the *same* population
	+ Combine the data from both samples to form a population
	+ Randomly reallocate data values to samples and calculate differences in sample means
* Find the probability of observing a difference in sample means as extreme as the one we observed.
* State a conclusion about the population means.

**Constructing a Randomization Distribution of Differences in Sample Means**

We assume that that the mean change in anxiety scores for teenagers who play casual video games is the *same* as the mean change in anxiety scores for teenagers who don’t. In other words, playing casual video games has *no effect* on teenagers’ anxiety levels. Thus, we assume that the two samples came from the same population. This population is shown below.

Population of changes in anxiety scores

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| -5 | 0 | -3 | -4 | 1 | 2 | -2 | -1 | 5 | 3 |
| -7 | 0 | -6 | -6 | 1 | -5 | -5 | -3 | -2 | 1 |

The population consists of all 20 data values. We can construct new samples (re-randomized samples) by randomly assigning 10 values to each group. Once we have two groups of 10 values, we can calculate the difference between the means of the re-randomized samples.

1. *Group 1* – Randomly select 10 values from the population above. Calculate the sample mean. Round to the nearest tenth.
2. *Group 2* – Select the remaining 10 values from the population that were not selected for Group 1. Calculate the sample mean. Round to the nearest tenth.
3. Calculate the difference in sample means from the re-randomized samples.
4. Repeat Questions 3 – 5 to calculate a second difference in sample means from re-randomized samples.
5. Create a dotplot of the differences in sample means from the re-randomized samples. Use all the differences found by students in your class. This distribution is formed under the assumption that the samples come from the *same* population, which means playing casual video games has no effect on anxiety among teenagers who suffer from anxiety.



1. What is the difference between the observed sample means from the treatment group and control group? Use the sample means that you calculated in Question 1 to answer this question. Calculate *mean of control – mean of treatment*.
2. The randomization distribution above assumes the two groups come from the same population. Based on the randomization distribution above, what is the probability of obtaining two random samples whose sample means have a difference greater than or equal to the value computed in Question 8? This probability is called a *P*-value (probability value).
3. Is the observed difference in sample means statistically significant?
4. What can we conclude about the population means? Is there convincing evidence that playing casual video games reduces anxiety among teenagers who suffer from anxiety?

**Who Exercises More?**

The following dotplots display the results of a sample survey that asked students at one high school how much time they spend exercising each week. Each sample contains data from a random sample of fifteen students.



The fifteen juniors have a sample mean of 3.5 hours per week. The fifteen seniors have a sample mean of 6.5 hours per week. Does the sample data provide evidence that seniors exercise more than juniors? We can use a randomization test to answer this question.

1. Find the difference in the observed sample means (*mean of seniors – mean of juniors*).

The randomization distribution below displays 300 differences in sample means from re-randomized samples under the assumption that the mean weekly exercise amount for juniors and seniors are the same.

Randomization Distribution of Differences in Sample Means



1. The randomization distribution assumes that the mean weekly exercise amount for juniors and seniors are the same. If this is true, what is the probability of obtaining two random samples from this population whose sample means differ by the amount observed or more?
2. Is the observed difference in sample means statistically significant?
3. What can we conclude about the mean weekly exercise amount for juniors and seniors at this high school?