## Introduction to Randomization Test

*Scenario*: An ***experiment*** with two treatments randomly assigned to experimental units

*Goal*: To compare the mean effects $\mu_1$ & $\mu_2$ of Treatments 1 and 2 based on the data

*Analysis*: Generally… can use two sample t-test

But, the ***randomization test*** is an ***important alternative*** to the two-sample t-test for testing hypotheses about treatment means

## So why the randomization test?

1. The randomization test can be a more intuitive approach to assessing evidence (p-values) compared to t-test.

2. Two sample t-test requires assumptions for using t-ratios and t-distributions.

These assumptions are often valid for your data analysis…. ***but not always***! (We'll discuss this.)

…. but the randomization test ***does not*** use a t-distribution assumption and so can apply to cases where the two sample t-test is not valid.

Main Disadvantage: the test is computationally a lot more work than the two sample t-test.

While the two sample t-test and randomization tests are different, ***both*** tests often give ***similar p-values when the t-test is valid***. (This gives more support to using the two sample t-test whenever possible.)

## Example of Reasoning Behind the Randomization Test

- Suppose you have developed a treatment (drug, genetic alteration, feed, method of raising, etc.) that you believe will increase the lean percentage of hogs at slaughter.

  *(Lean percentage is often estimated as a function of carcass weight, back fat, and loin eye depth.)*

- You have 100 pigs to use in an experiment to test your claim that the treatment will increase lean percentage.
- You randomly allocate 50 pigs to receive the treatment and the other 50 to be the control group.
- You treat all pigs in the same manner aside from giving the one group of 50 the treatment and not the other. *(Control group should get a placebo.)*

- You record the lean percentages of the 100 pigs at the time of slaughter.

| Treatment | Control |
|---|---|
| 49.2 52.4 50.0 54.4 51.0 | 49.5 53.5 46.7 52.1 49.2 |
| 54.3 52.5 52.9 56.0 54.4 | 48.2 51.5 51.7 50.4 52.5 |
| 54.8 53.8 53.0 49.7 50.4 | 52.5 49.8 50.5 51.2 49.3 |
| 53.4 52.2 56.1 52.4 48.9 | 48.5 52.8 51.7 53.1 47.1 |
| 50.9 52.5 51.9 56.5 49.4 | 52.6 49.7 47.9 54.6 53.0 |
| 53.6 52.1 50.4 49.1 53.4 | 52.9 52.3 52.3 50.1 51.1 |
| 47.8 54.6 53.2 54.4 51.6 | 49.8 53.8 51.1 52.2 50.9 |
| 54.3 47.6 50.6 51.0 50.5 | 50.0 50.1 50.3 50.5 49.7 |
| 50.0 53.6 50.2 51.1 53.3 | 50.0 49.7 48.1 50.0 51.8 |
| 53.0 52.9 54.4 53.7 54.0 | 52.3 49.1 48.7 51.8 51.8 |

# Summary Statistics

| | Treatment | Control |
|---|---|---|
| Sample Size | 50 | 50 |
| Mean | 52.3 | 50.8 |
| Standard Deviation | 2.1 | 1.8 |
| Maximum | 56.5 | 54.6 |
| Q3 | 53.77 | 52.27 |
| Median | 52.5 | 50.7 |
| Q1 | 50.52 | 49.7 |
| Minimum | 47.6 | 46.7 |

# Is there evidence that the treatment increased lean percentage?

- If the treatment works, we should expect the 50 treatment hogs to have a _______ sample average than the 50 control hogs
- If the treatment does NOT work, we would expect the sample averages of both groups to be about the same (i.e., difference of the two sample averages should be about ___)
- Observed that the average lean percentage for the treatment hogs (52.3) was higher than for control hogs (50.8).
- A difference of 52.3-50.8=1.5 points. So what?

## Suppose the treatment has no effect

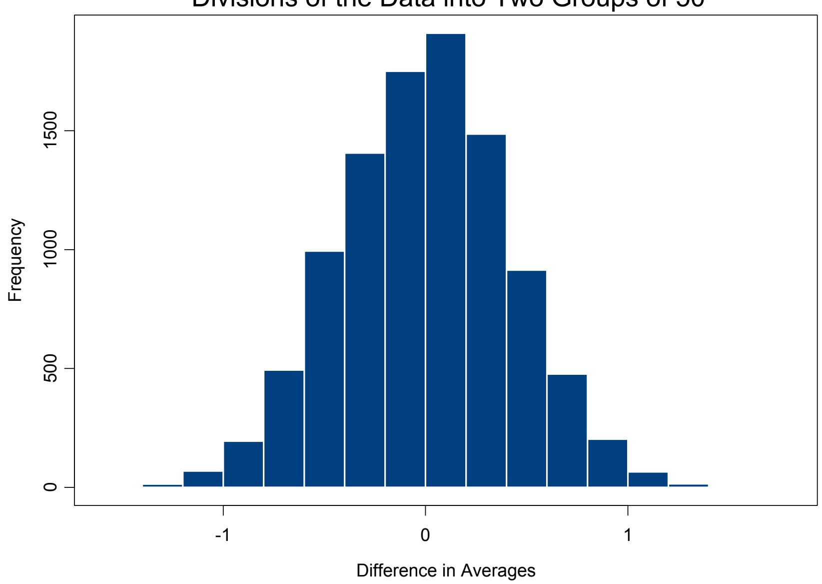
- The two treatment ``groups" made were really artificial… each hog would grown the same regardless of which ``group" it was in
- Your random split was just artificial labeling and maybe the difference of 52.3-50.8=1.5 in averages is *just due to the chance* way hogs were randomly divided into two groups (by chance more lean hogs ended up in the treatment group)
- If the treatment had no effect, then really you just randomly split 100 hog observations into two groups. So ask next:

  ``If these 100 hog observations were randomly split into two groups, what is the chance that 50 hog observations randomly assigned to the treatment group would have a sample average at least 1.5 points higher than the sample average for the hog observations assigned to the control group?"

## Randomization Test

1. Randomly divide the 100 lean percentages that we observed in this experiment into two groups of 50.
2. Note the difference in sample averages between the two groups.
3. Repeat 1 and 2 a total of 10,000 times.
4. Note the proportion of times that the difference in sample averages was at least as great (as far from 0) as it was in the original experiment 52.3-50.8=1.5.

Histogram of Mean Differences from 10,000 Random Divisions of the Data into Two Groups of 50




- The mean difference was as far from 0 as 1.5 (at least as great) for only 2 out of the 10,000 random divisions of the data into two groups of 50.
- If the treatment really had no effect, then it would be very unusual to find a difference in means as large as our observed difference of 52.3-50.8=1.5 simply by a chance labeling of hogs.

  What we actually observed in our experiment is so unlikely to occur under the assumption of ``no treatment effect" that we start to doubt this assumption (the null hypothesis)
- It seems reasonable to believe that the treatment caused the difference in mean lean percentages.