Math 243

Inv. 4.8 - Matched Pairs Designs

Ch. 4: Compare a **quantitative variable** between two groups (**binary variable**)

In general, there are 3 *methods* of testing $H_0: \mu_1 = \mu_2$

1. Simulation of null distribution

2. Exact mathematical formula

3. Approximate formula for null distribution

Ch. 4: Compare a **quantitative variable** between two groups (**binary variable**)

In general, there are 3 *methods* of testing $H_0: \mu_1 = \mu_2$

1. Simulation of null distribution

2. Exact mathematical formula

3. Approximate formula for null distribution

Warning: the actual form of these depends on how the study was designed – experimental (how was random assignment used?) or observational study

Inv. 4.4 Sleep Deprivation: part j

Subjects were randomly assigned to be sleep deprived or not. Their improvements in reaction times on visual discrimination task are given below.

Sleep deprivation group ($n = 11$): −10.7, 4.5, 2.2, 21.3, −14.7, −10.7, 9.6, 2.4, 21.8, 7.2, 10.0
Unrestricted sleep group ($n = 10$): 25.2, 14.5, −7.0, 12.6, 34.5, 45.6, 11.6, 18.6, 12.1, 30.5

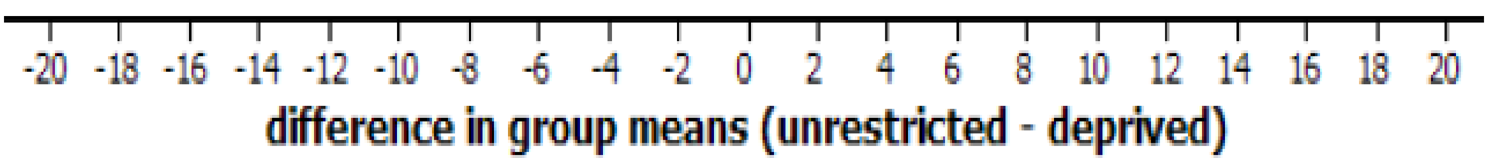
1. Write each of these numbers on an index card, shuffle the 21 cards and randomly deal into two groups.

2. Compute the simulated difference in means.

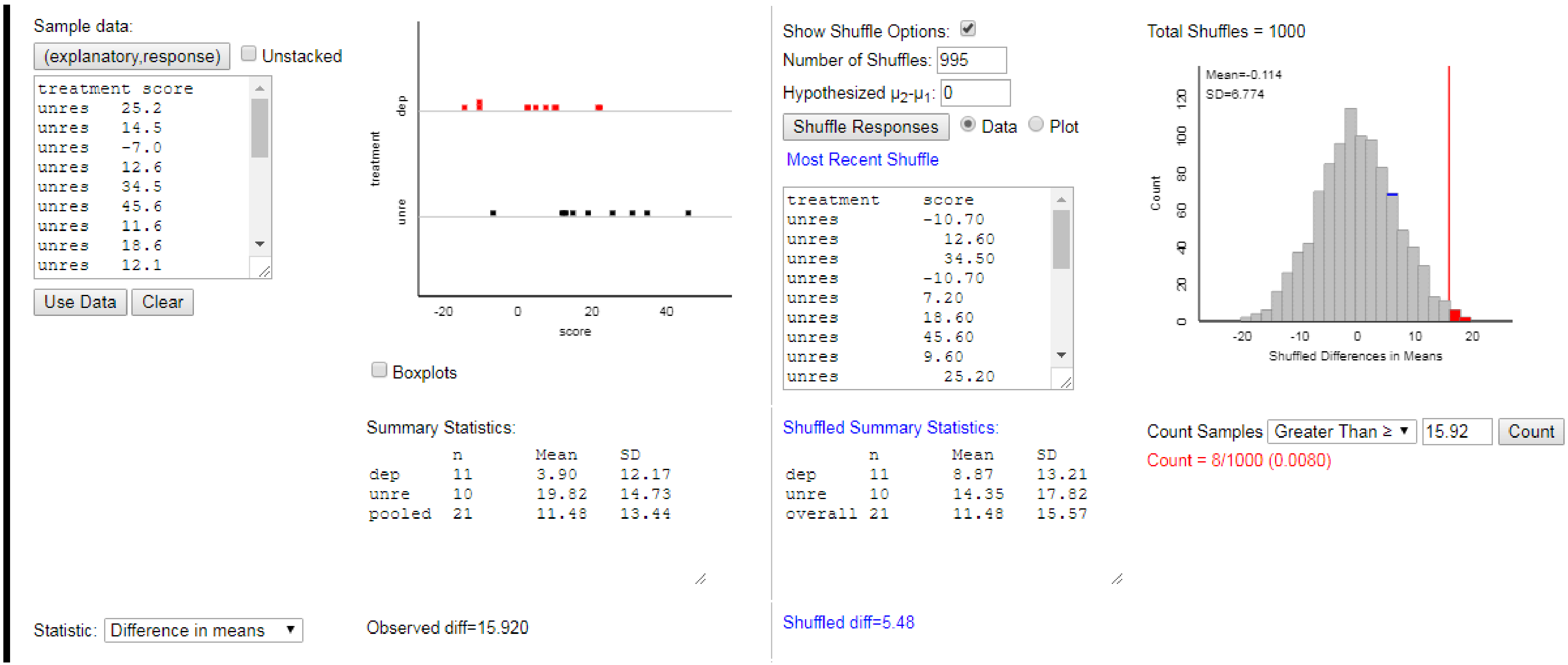
3. Repeat steps 1-2 many times

Inv. 4.4: part l
Randomization Test

(l) Combine your results with your classmates to produce a dotplot of the *difference in group means*.



Inv. 4.4: part p-q



Shuffle 21 subjects and randomly assign to two groups many (1000) times – this gives us the null distribution and allows us to compute the “empirical” p-value.

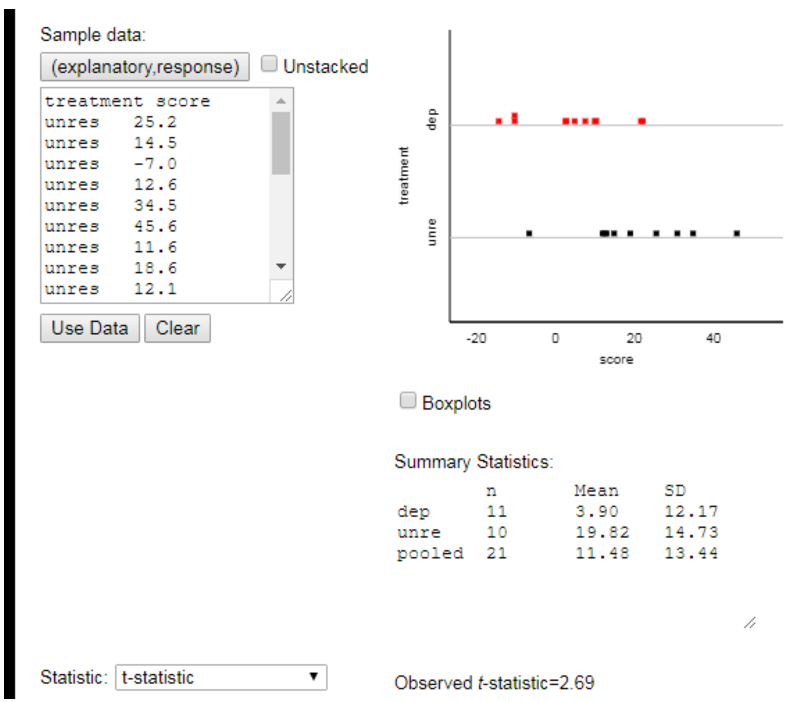
Inv. 4.4: part w

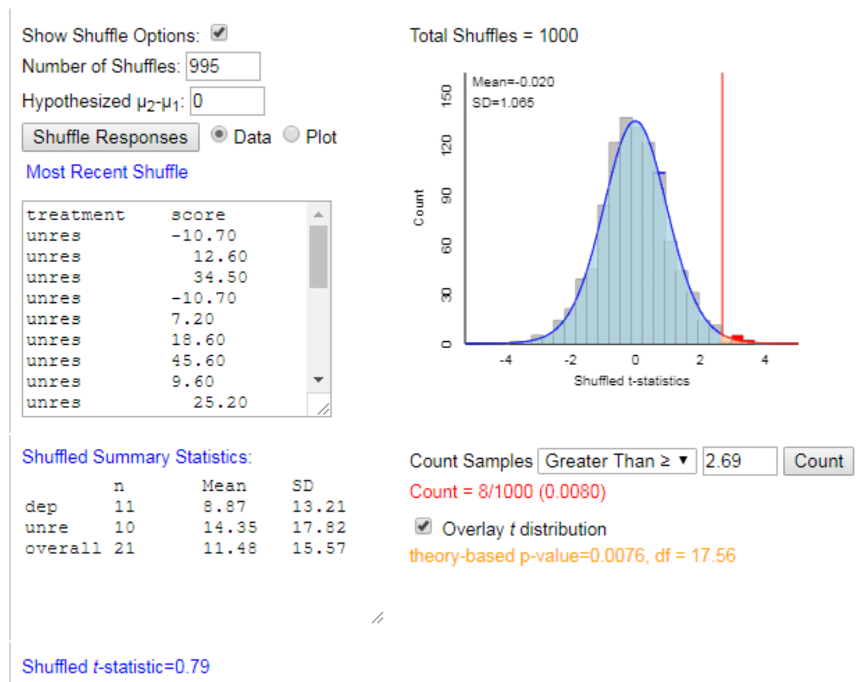
The “exact” p-value is

2456/352716 = 0.0070

Inv. 4.5: part w

The approximate p-value from the two-sample t-test is 0.0076





Which test to use?

- **Exact Test** is exactly correct, but can take some time to run for large datasets
- **Randomization Test** with a large number of shuffles is always possible and can always be made “close enough” to the Exact Test
- **Two sample t-test** – will be “close enough” to the Exact Test if either
  - The two populations are normally distributed, or
  - The sample sizes are both large (>20)

Investigation 4.8 (p. 289)

<u>**Research Question:**</u>

Compare melting times of **dark** and **milk** chocolate chips.

<u>**Parameter of interest**</u>

The difference in mean melting times between dark and milk chocolate chips, $\mu_D - \mu_M$

<u>**Hypotheses:**</u>

$$H_0: \mu_D - \mu_M = 0 \quad \text{vs.} \quad H_a: \mu_D - \mu_M \neq 0$$

Let’s collect some data…

Unfortunately, I have only a limited number of chocolate chips and I need the answer to my research question by the end of class...

Please work in pairs to come up with a study design that we can implement in class.

Study designs?

<u>Considerations:</u>

We have access to

- Students,
- Timers on their cellphones,
- A limited number of milk and dark chips

We want to be able to conclude that any observed difference in ***melting times*** was caused by a difference in ***chip type*** (milk or dark)

Study designs?

Considerations:

We have access to

- Students,
- Timers on their cellphones,
- A limited number of milk and dark chips

We want to be able to conclude that any observed difference in ***melting times*** was caused by a difference in ***chip type*** (milk or dark)…

…so we should use random assignment to eliminate the possibility of confounding variables.

An additional consideration: power

Power is the probability that our test will detect a difference in melting times of milk and dark chips if a difference truly exists.

We'd like to design our experiment to have as high a power as possible.

Given a fixed sample size (number of students/chips available) and a fixed level of significance (0.05), we should seek out the most powerful **design** for our experiment.

Which design do you think is more **powerful?**

Each student tosses a coin to **randomly pick either a dark or milk chip**.

Each student then

1. places their chip on their tongue,
2. starts a timer and holds the chip against the roof of their mouths until the chip is melted, and
3. the time until the chip is completely melted, without any "encouragement" by the student, is recorded in seconds.

Each student tosses a coin to **randomly pick to melt a dark or milk chip first**.

Each student then

1. places the first chip on their tongue,
2. starts a timer and holds the chip against the roof of their mouths until the chip is melted, and
3. the time until the chip is completely melted, without any "encouragement" by the student, is recorded in seconds.

The steps 1-3 are repeated for the **second chip by each student**.

A Matched Pair Design

The melting times of dark and milk chocolate chips are “paired” by student.

This design results in a more powerful experiment because the variation in melting times due to differences in students is eliminated.

Each student tosses a coin to **randomly pick to melt a dark or milk chip first**.

Each student then

1. places the first chip on their tongue,
2. starts a timer and holds the chip against the roof of their mouths until the chip is melted, and
3. the time until the chip is completely melted, without any “encouragement” by the student, is recorded in seconds.

The steps 1-3 are repeated for the **second chip by each student**.

Collect Data using the “matched pair” design

<u>Instructions:</u>

1. Randomly decide which chip to melt first (toss a coin)
2. Melt each chip:
   - Put the chip on your tongue
   - Press to the roof of your mouth
   - Record time (in seconds)
3. Record both melting times by chip type on the sheet.

*Answer as much of **parts (h)-(j)** as you can while the data is being recorded.*

Results of Data Collection

What test of significance should we use?

Test: H_0: $\mu_D - \mu_M = 0$ vs. H_a: $\mu_D - \mu_M \neq 0$

What test of significance should we use?

<u>Test:</u> H_0: $\mu_D - \mu_M = 0$ vs. H_a: $\mu_D - \mu_M \neq 0$

A two-sample t-test or the Randomization Test from last time assume that the samples of milk and dark chocolate chips melting times are

independent

that is, the melting times of a milk chip are unrelated to the melting of any of the dark chips.

This assumption is not satisfied for our data because the melting times are related through students:

A student with a "cool" mouth will produce longer melting times for both milk and dark chips, while "hot" mouthed students will produce longer

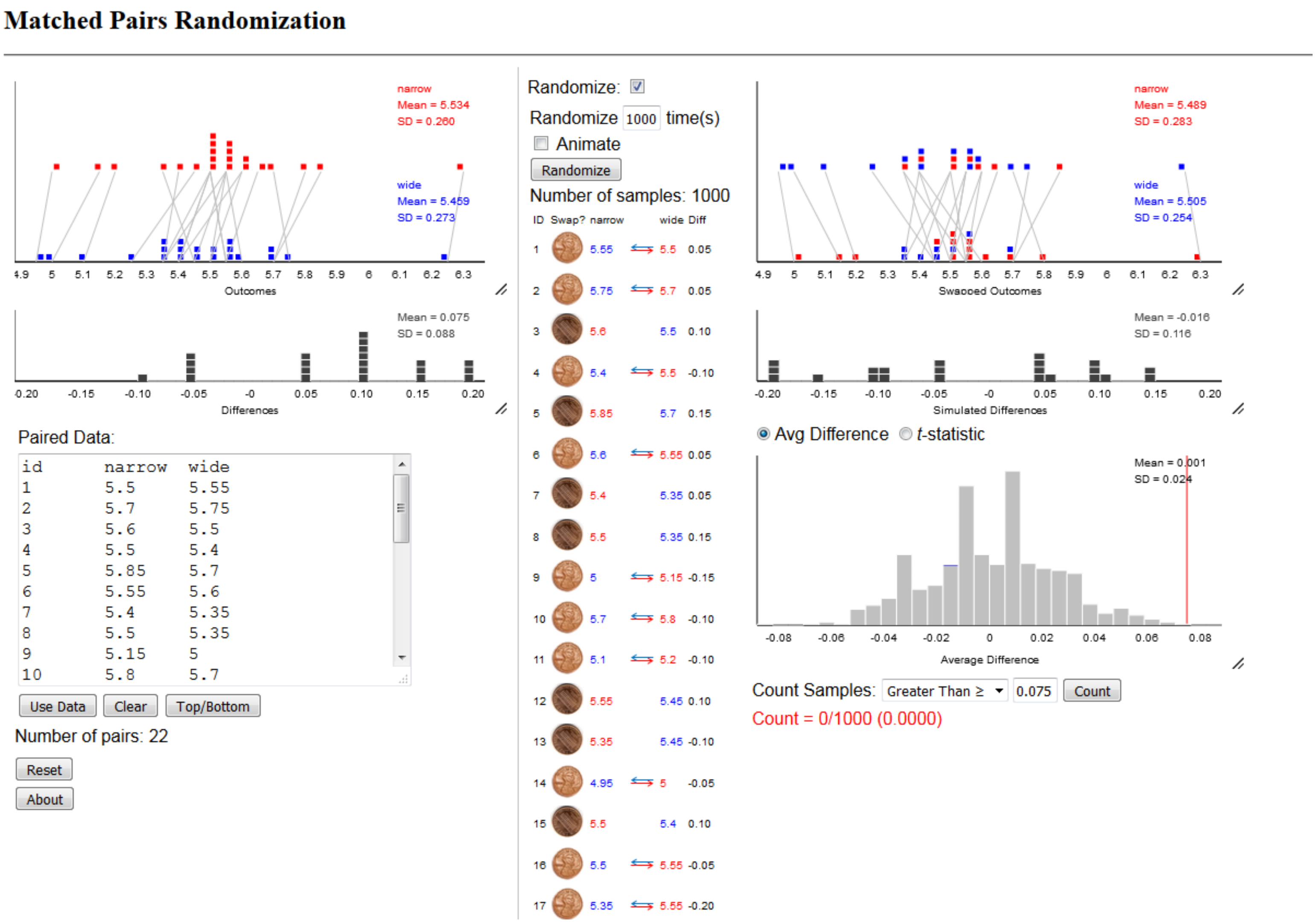
Tests of significance for "Matched Pairs"

<u>Test:</u> H_0: $\mu_D - \mu_M = 0$ vs. H_a: $\mu_D - \mu_M \neq 0$

It is appropriate to use a one-sample t-test, where the "one" sample consists of the "differences in melting times", if the number of students is more than 30.

If less than 30 students were in the sample, then the matched pairs randomization test can be used. Here, the random choice of which chip to melt first is mimicked by randomly choosing the sign of the difference in melting times to be + or -.

Applet: Matched Pairs Randomization



Investigation 4.8 (p. 289)

Research Question:

Compare melting times of **dark** and **milk** chocolate chips.

Parameter of interest

The difference in mean melting times between dark and milk chocolate chips, $\mu_D - \mu_M$

Hypotheses:

$$H_0: \mu_D - \mu_M = 0 \text{ vs. } H_a: \mu_D - \mu_M \neq 0$$

Study designs?

<u>Considerations:</u>

We have access to

- Students,
- Timers on their cellphones,
- A limited number of milk and dark chips

We want to be able to conclude that any observed difference in ***melting times*** was caused by a difference in ***chip type*** (milk or dark)…

…so we should use random assignment to eliminate the possibility of confounding variables.

An additional consideration: power

Power is the probability that our test will detect a difference in melting times of milk and dark chips if a difference truly exists.

We'd like to design our experiment to have as high a power as possible.

Given a fixed sample size (number of students/chips available) and a fixed level of significance (0.05), we should seek out the most powerful **design** for our experiment.

Which design do you think is more **powerful?**

Each student tosses a coin to **randomly pick either a dark or milk chip**.

Each student then

1. places their chip on their tongue,
2. starts a timer and holds the chip against the roof of their mouths until the chip is melted, and
3. the time until the chip is completely melted, without any “encouragement” by the student, is recorded in seconds.

Each student tosses a coin to **randomly pick to melt a dark or milk chip first**.

Each student then

1. places the first chip on their tongue,
2. starts a timer and holds the chip against the roof of their mouths until the chip is melted, and
3. the time until the chip is completely melted, without any “encouragement” by the student, is recorded in seconds.

The steps 1-3 are repeated for the **second chip by each student**.

A Matched Pair Design

The melting times of dark and milk chocolate chips are "paired" by student.

This design results in a more powerful experiment because the variation in melting times due to differences in students is eliminated.

Each student tosses a coin to **randomly pick to melt a dark or milk chip first**.

Each student then

1. places the first chip on their tongue,
2. starts a timer and holds the chip against the roof of their mouths until the chip is melted, and
3. the time until the chip is completely melted, without any "encouragement" by the student, is recorded in seconds.

The steps 1-3 are repeated for the **second chip by each student**.

Collect Data using the “matched pair” design

<u>Instructions:</u>

1. Randomly decide which chip to melt first (toss a coin)
2. Melt each chip:
   - Put the chip on your tongue
   - Press to the roof of your mouth
   - Record time (in seconds)
3. Record both melting times by chip type on the sheet.

Answer as much of ***parts (h)-(j)*** *as you can while the data is being recorded.*

Results of Data Collection

What test of significance should we use?

<u>Test:</u> H_0: $\mu_D - \mu_M = 0$ vs. H_a: $\mu_D - \mu_M \neq 0$

A two-sample t-test or the Randomization Test from last time assume that the samples of milk and dark chocolate chips melting times are

independent

that is, the melting times of a milk chip are unrelated to the melting of any of the dark chips.

This assumption is not satisfied for our data because the melting times are related through students:

A student with a "cool" mouth will produce longer melting times for both milk and dark chips, while "hot" mouthed students will produce longer

Tests of significance for “Matched Pairs”

Test: H_0: $\mu_D - \mu_M = 0$ vs. H_a: $\mu_D - \mu_M \neq 0$

It is appropriate to use a one-sample t-test, where the “one” sample consists of the “differences in melting times”, if the number of students is more than 30.

If less than 30 students were in the sample, then the matched pairs randomization test can be used. Here, the random choice of which chip to melt first is mimicked by randomly choosing the sign of the difference in melting times to be + or -.

Applet: Matched Pairs Randomization