Math 243

Randomization test and t-test - Inv. 4.4-4.5

Ch. 4 – Comparing two groups (binary EV) on single quantitative response (RV)

- What are appropriate **graphs** to look at?
- What are appropriate **statistics** for summarizing the data numerically?
- How to **test** H_0?

  - Randomization test (simulation)

  - t-test (approximate)
- How **estimate** a difference in population or treatment means?
- Scope of conclusions based on **study design**

Inv. 4.4 & 4.5: Sleep Deprivation

When a quantitative variable is measured in an experimental study in which groups were randomly assigned, it is often of interest to test whether the means of each group are equal.

Today, we will investigate a “randomization” test and see that the results are close to the two-sample t-test.

Inv. 4.4: parts a-d

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

a) Experiment or Observational Study?
b) EV? RV?
c) H_0 H_A

Inv. 4.4: parts a-d

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

a) Experiment or Observational Study?

b) EV: Sleep Deprived (binary)

RV: Improvement in reaction time (quantitative)

a) H_0: $\mu_{unrestricted} = \mu_{deprived}$

H_a: $\mu_{unrestricted} > \mu_{deprived}$

Descriptive Statistics for one quantitative variable, two independent groups

Statistic

difference in sample means, $\bar{x}_1$ - $\bar{x}_2$

Graphs

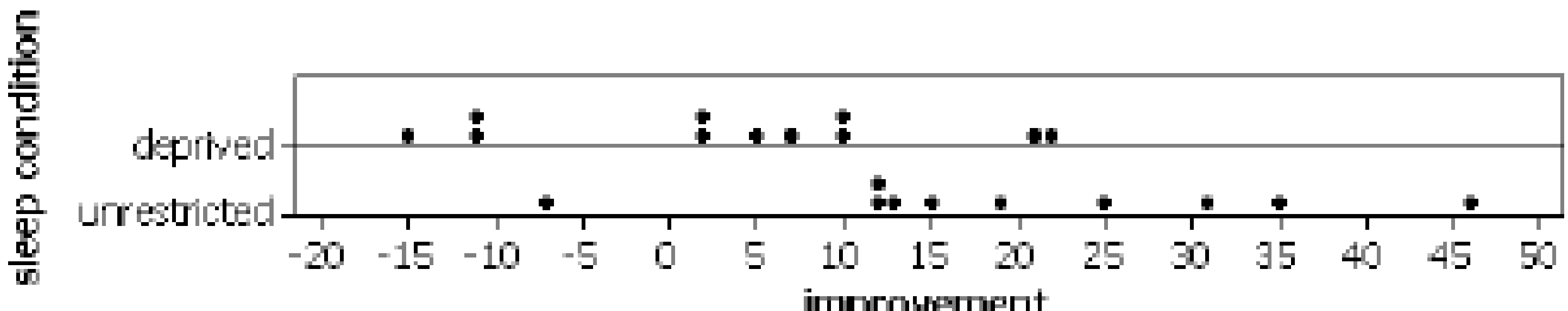
2 dotplots, histograms, boxplots *on the same scale*

Inv. 4.4: part f

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

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

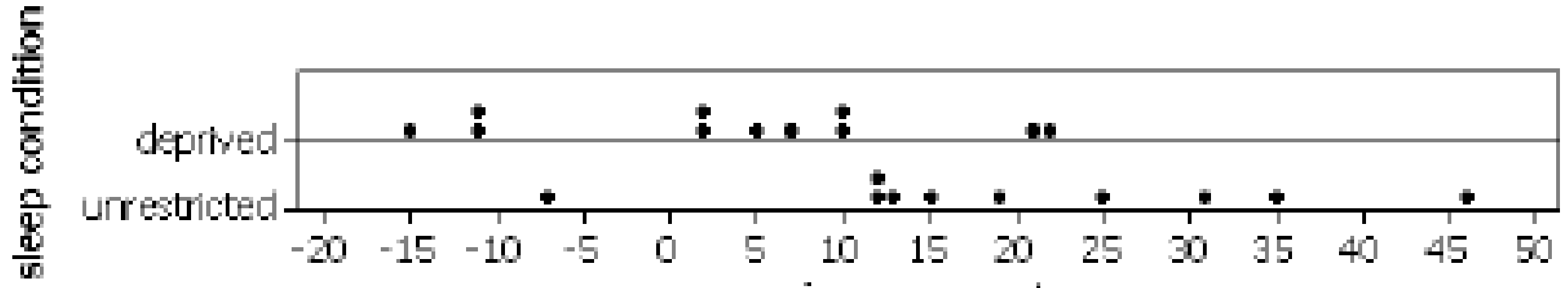


Inv. 4.4: part f

Subjects were randomly assigned to be sleep deprived or not. Their improvements in reaction times on visual discrimination task are given below. Compute the mean difference in reaction times (15.92).

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



Simulation-based approach for p-value

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

How could we simulate the method of data collection if the null hypothesis were true?

Inv. 4.4: part j

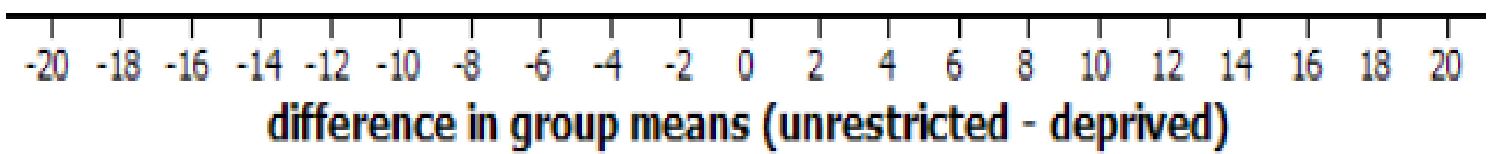
Subjects were randomly assigned to be sleep deprived or not. Their improvements in reaction times on visual discrimination task are given below.

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.

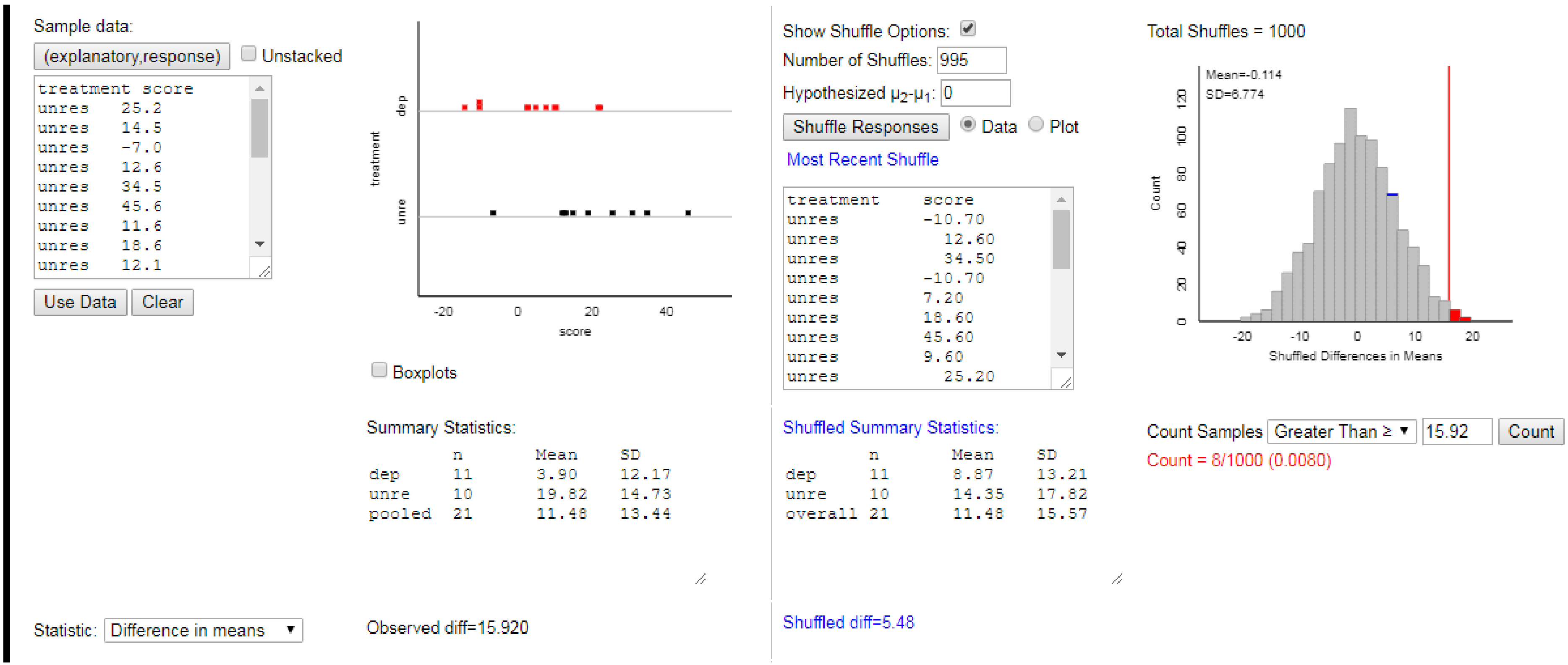
Inv. 4.4: part l
Randomization Test

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



3. Repeat steps 1-2 many times

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.

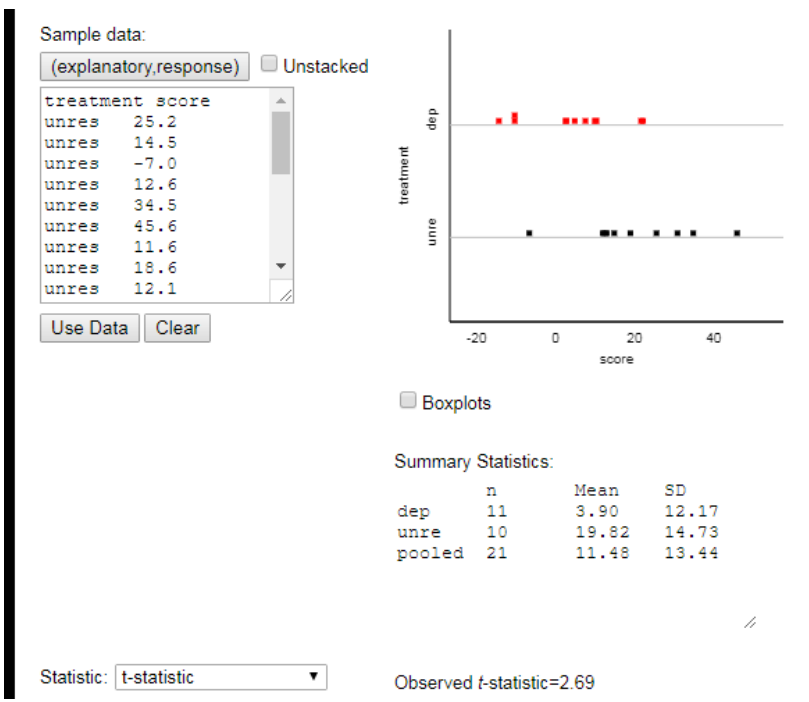
Inv. 4.4: part w

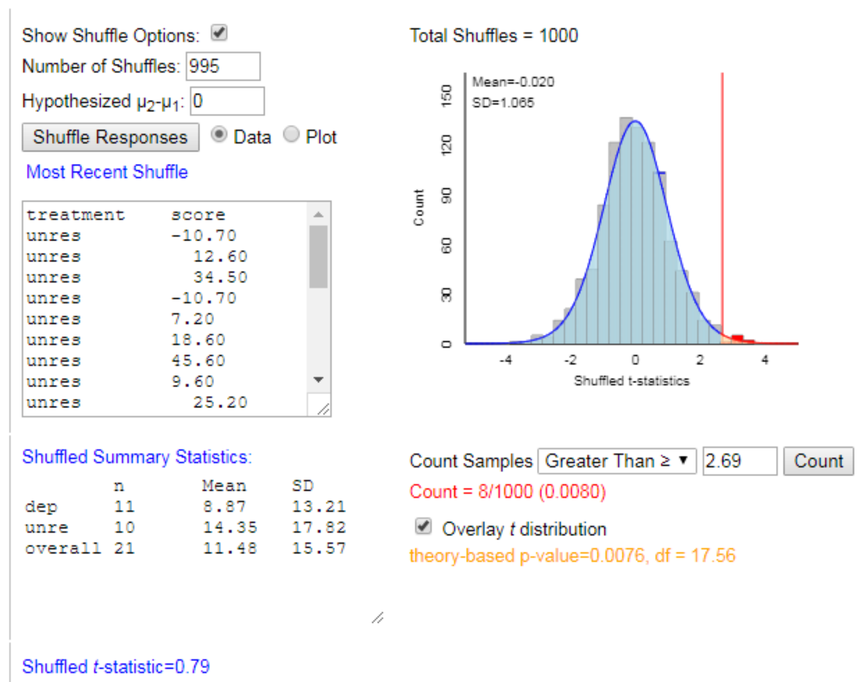
The “exact” p-value is

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)

Interpretations in context

- should be technically correct but free of statistical jargon

- Tip:
  - Start with the **definition** of the term then change the statistical jargon part to fit the “context”.
  - Add the **numbers** from your analysis
  - Then reword for readability

Give an *interpretation* of the **p-value**

P-value: the probability of seeing a sample result at least as extreme as our sample result if the null hypothesis were true.

Where's the statistical jargon?
What information do we have?

Give an *interpretation* of the **p-value**

P-value: the probability of seeing a sample result at least as extreme as our sample result if the null hypothesis were true.

Our sample result is that the difference in sample means of unrestricted minus deprived was 15.92

Null hypothesis is H_0: $\mu_{unrestricted} - \mu_{deprived} = 0$
If true, there is no difference in average improvement between unrestricted and deprived groups.

P-value from Exact Test was 0.007

Give an *interpretation* of the **p-value**

P-value: the probability of seeing a sample result at least as extreme as our sample result if the null hypothesis were true.

There is a 0.007 probability of seeing a sample result at least as large as a difference in sample means of unrestricted minus deprived of 15.92 if the there is no difference in improvement between unrestricted and deprived groups.

Reword for readability…

There is a 0.007 probability of seeing a sample result at least as large as a difference in sample means of unrestricted minus deprived of 15.92 if the there is no difference in improvement between unrestricted and deprived groups.

There is a 0.007 probability of seeing sample mean of the unrestricted group being 15.92 units or more than the deprived group if the sleep deprivation does not cause a change in reaction time.

Give an *interpretation* of the t-interval in context

95% confidence interval: I am 95% confident that the population parameter falls in the interval.

Where is the statistical jargon?

What other info do we know?

Give an *interpretation* of the t-interval in context

95% confidence interval: I am 95% confident that the population parameter falls in the interval.

95% confidence interval: I am 95% confident that the difference in means (unrestricted – deprived) falls in (3.44, 28.4).

Reword for readability

95% confidence interval: I am 95% confident that the difference in mean improvement (unrestricted – deprived) falls in (3.44, 28.4).

95% confidence interval: I am 95% confident that the mean improvement of young adults with unrestricted sleep is 3.44 to 28.4 units more than those deprived of sleep.