Math 361

Tests for Experimental Studies-Means from Two Independent Groups - 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**

Study designs

We'll see how to perform simulation-based tests and approximate tests for data for **three** types of study designs.

1. Observational study of samples from two populations (Inv. 4.2)

Experimental study with random assignment into

2. two independent groups (Inv. 4.4), or
3. matched pairs (Inv. 4.8)

Inv. 4.4: 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? sleep binary

RV? difference in reaction time numerial

c) H_0 $\mu_{res} = \mu_{unres}$

H_A $\mu_{res} \neq \mu_{unres}$

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

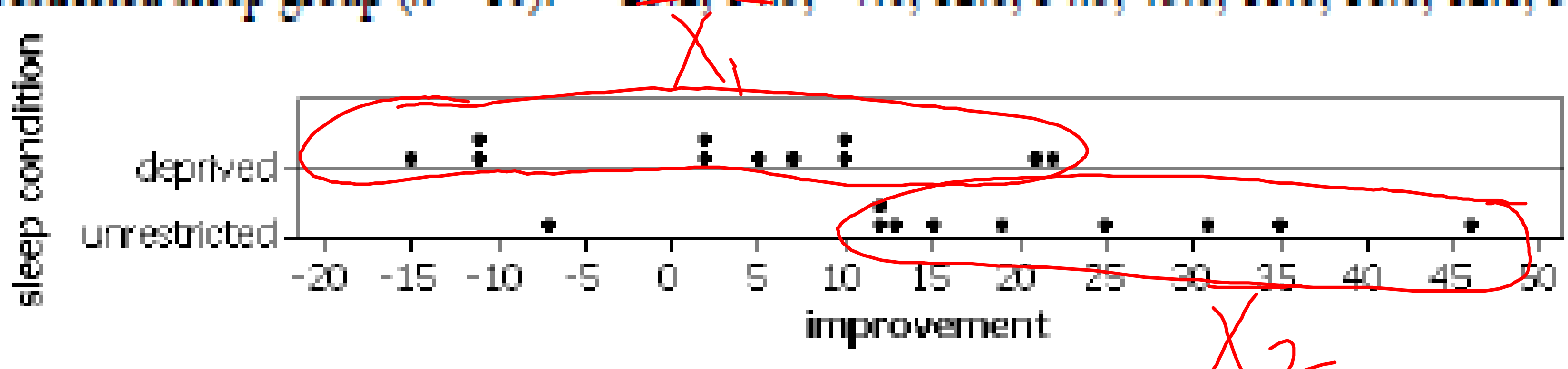
2 dotplots, histograms, boxplots *on the same scale*

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

21 cards dealt into 2 piles

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

21

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.

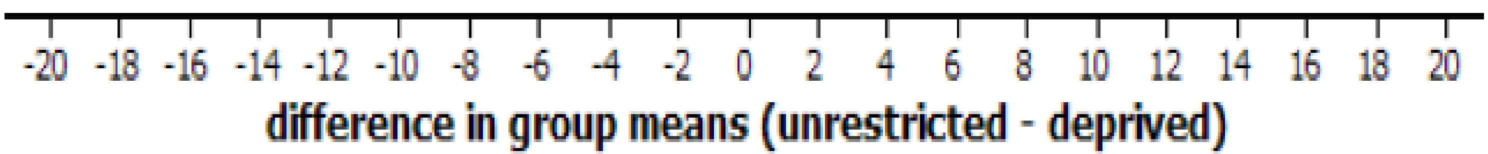
Sleep deprivation group ($n = 11$): −10.7, 4.5, 2.2, 21.3, −14.7, −10.7, 9.6, 2.4, 21.8, 7.2, 10.0
Unrestricted sleep group ($n = 10$): 25.2, 14.5, −7.0, 12.6, 34.5, 45.6, 11.6, 18.6, 12.1, 30.5

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
(Simulated) 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 of 0.008.

“Exact” Randomization Test

p-value = proportion of all possible divisions of the 21 subjects into groups of size 10 and 11 that are at least as extreme as the observed study difference (15.92)

Recall $\binom{n}{k}$ = number of ways to choose k items from n unique items when order doesn’t matter

“Exact” Randomization Test

p-value = proportion of all possible divisions of the 21 subjects into groups of size 10 and 11 that are at least as extreme as the observed study difference (15.92)

Recall $\binom{n}{k}$ = number of ways to choose k items from n unique items when order doesn’t matter:

There are $\binom{21}{10} = 352716$ ways to divide 21 people into groups of size 10 and 11.

Using R or Minitab or Excel, of these 352716 ways, only 2456 have a difference in sample means as large as 15.92

So p-value = 2456/352716 = 0.007

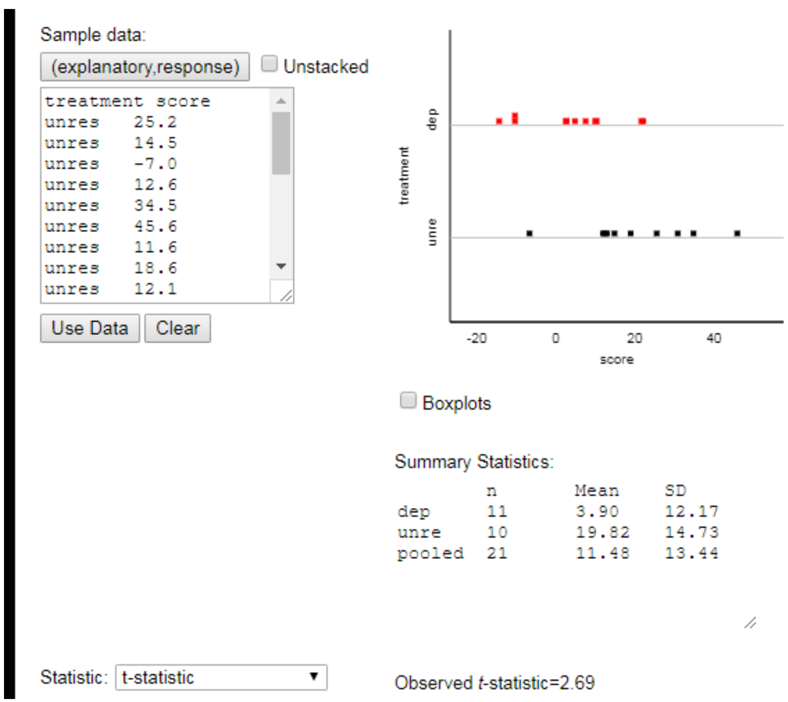
Inv. 4.4: part w

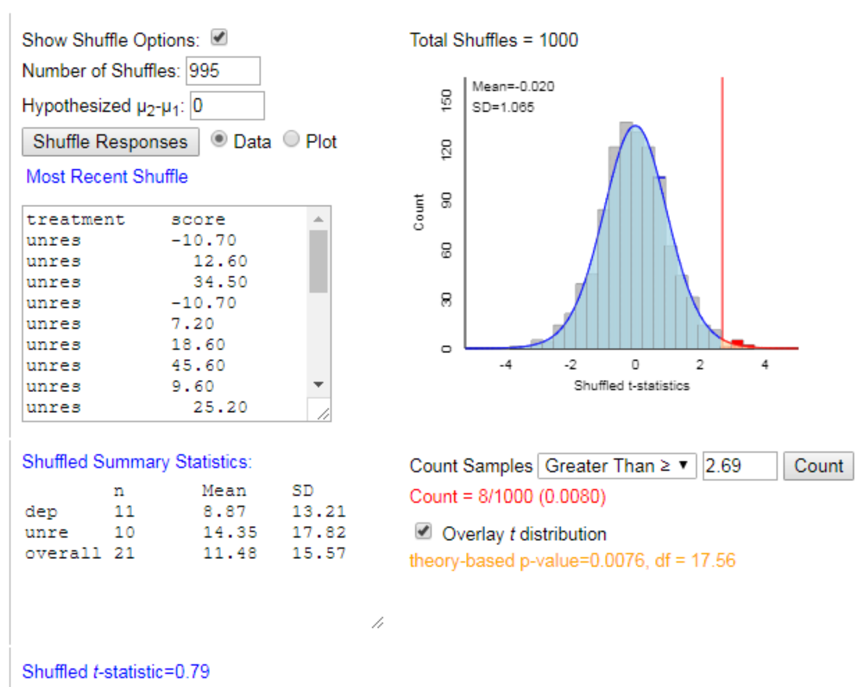
The “exact” p-value is

2456/352716 = 0.0070

Inv. 4.5: part w

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





Two sample t-test and t-interval

Summary of Two-sample *t* Procedures

Parameter: $\mu_1 - \mu_2$ = the difference in the population means

To test H_0: $\mu_1 - \mu_2 = \delta_0$

Test statistic:

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

***t*-Confidence interval for $\mu_1 - \mu_2$:**

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Approximate degrees of freedom: Compare this to a *t*-distribution with degrees of freedom equal to the smaller of the two samples sizes minus one: $min(n_1, n_2) - 1$.

Technical conditions: These procedures are considered valid if the sample distributions are reasonably symmetric or the sample sizes are both at least 20.

Three ways to compute a p-value to test a difference in means between two independent groups

- Exact
- Card Shuffle Simulation (also called “empirical)
- Two sample t-test
  - only if each sample size is at least 20