EDP308: STATISTICAL LITERACY

The University of Texas at Austin, Fall 2020

RAZ: Rebecca A. Zárate, MA

Overview

- Dependent Samples t-tests
  - Repeated Measures
  - Hypotheses Formulation
    - The Null – Zero Enough
  - Variation of the Difference
- Full Example: Atkin's Diet
- Confidence Intervals for Dependent Samples
- Relationship of Standard Error and Sample Size
- Dependent Samples t-tests in R
  - Summary Statistics
  - Data

Dependent Samples t-test

Dependent Samples t-tests

- With dependent samples, we are interesting in comparing two means (like we've been doing along) which are now *related* in some way ("depend" on each other in some way)
- Matched Pairs of Data
  - Repeated Samples:
    - Taking two measures from one person
      - Ex. Before and After treatment
      - Ex. With notifications on the phone enable and disabled

Independent and Dependent

- In independent samples, we take the averages of both groups and take the difference between them and compare it to the Null (i.e. zero difference)
- In dependent samples, we compute the difference (ex. After Treatment – Before Treatment) then take the average of those difference and compare it to…?

What do you think our NULL hypothesis is here?

Dependent Samples NULL

- Just like in independent samples t-tests, we are going to compare to the NULL hypothesis assumption that there is NO DIFFERENCE between before and after treatment scores. We always start with the assumption there is no difference, no treatment affect, no change, etc.
  - Subscript is now D for difference $\mu_{Post} - \mu_{Pre} = \mu_D$

$$H_0: \mu_{Post} - \mu_{Pre} = 0$$
$$H_1: \mu_{Post} - \mu_{Pre} \neq 0$$

$$H_0: \mu_D \geq 0$$
$$H_1: \mu_D < 0$$

One sided.

$$H_0: \mu_D = 0$$
$$H_1: \mu_D \neq 0$$

Two sided.

$$H_0: \mu_D \leq 0$$
$$H_1: \mu_D > 0$$

One sided.

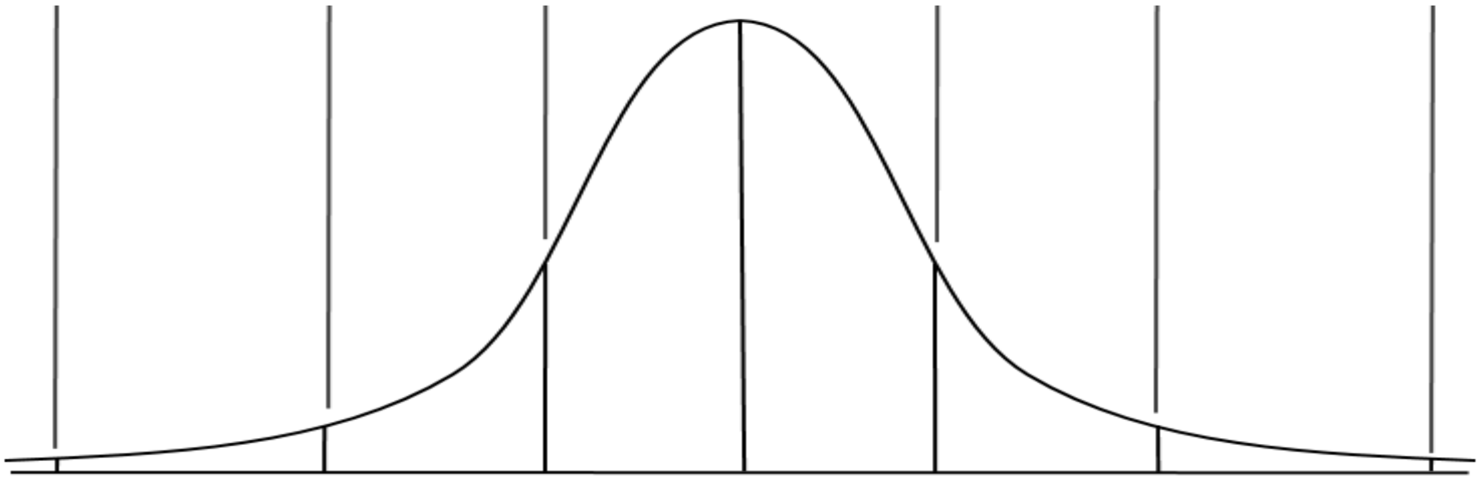
Dependent Samples Variables

- Say we have a sample of 10 students from this class. I measure your Stats Literacy knowledge at the beginning of the semester then at the end.
  - What is my NULL hypothesis?
  - What are my degrees of freedom?
  - What mean am I going to compare against the null?
  - What is my standard error?

Dependent Samples Variables

- Say we have a sample of 10 students from this class. I measure your Stats Literacy knowledge at the beginning of the semester then at the end.
  - What is my NULL hypothesis?
    - Null: There is no difference in Pre and Post Stats Literacy semester course scores. Difference will be zero.
  - What are my degrees of freedom?
    - $n - 1 = 9$
      - n is the number of SUBJECTS, not observations (20)
  - What mean am I going to compare against the null?
    - The average of difference scores $\bar{x}_D$
  - What is my standard error?
    - ${s_D}/{\sqrt{n}}$

Sampling Distribution of Differences in Means



$(\mu_{\bar{x}_{Post}} - \mu_{\bar{x}_{Pre}}) = \mu_{Post} - \mu_{Pre} = \mu_D = 0$

Same curve we see in sampling distributions, just now our sampling distribution representing $(\bar{x}_{Post}$ - $\bar{x}_{Pre}) = \bar{x}_D$

Dependent Samples t-statistic

Here is the test statistic when testing for mean differences:

$$t_{stat} = \frac{\bar{x}_D - \mu_D}{s_D/\sqrt{n}}$$

- $\bar{x}_D$ is the sample average of the difference scores for the matched pairs, ($\bar{x}_{Post}$- $\bar{x}_{Pre}$) = $\bar{x}_D$
- s_D is the sample standard deviation of the difference scores for the matched pairs
- n is the number of matched PAIRS in the sample
- $s_D/\sqrt{n}$ is the standard error of the mean difference scores

Dependent Samples t-test

Post Mean – Pre Mean

Null Hypothesis…

$$t_{stat} = \frac{(\bar{x}_{Post} - \bar{x}_{Pre}) - (\mu_{Post} - \mu_{Pre})}{s_D / \sqrt{n}}$$

Standard Deviation of Difference Scores

Standard Error

Dependent Samples t-test NULL

- The NULL hypothesis for repeated measures t-test is: "There is no difference between the pre-post means."

$$H_0: \mu_{Post} - \mu_{Pre} = 0$$

So this will always be equal to zero for two tailed tests. And the equation is actually this…

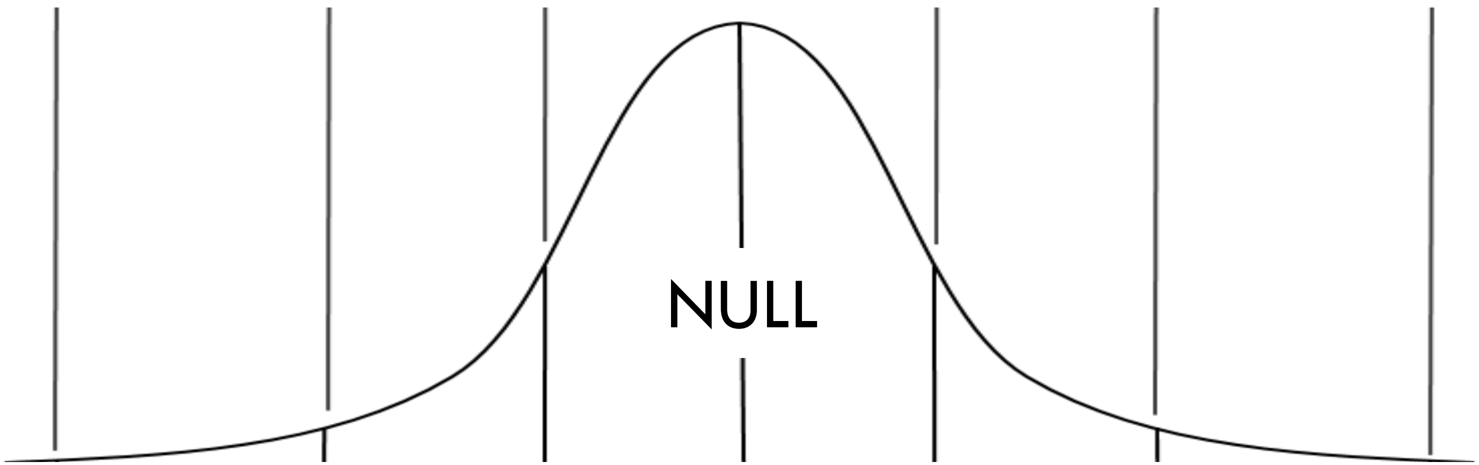
$$t_{stat} = \frac{(\bar{x}_{Post} - \bar{x}_{Pre}) - (\mu_{Post} - \mu_{Pre})}{s_D / \sqrt{n}}$$

$$t_{stat} = \frac{(\bar{x}_{Post} - \bar{x}_{Pre})}{s_D / \sqrt{n}}$$

$$t = \frac{(\bar{x}_{Post} - \bar{x}_{Pre}) - 0}{s_D / \sqrt{n}}$$

NULL is Zero

SAMPLING distribution of $(\bar{x}_{Post} - \bar{x}_{Pre})$



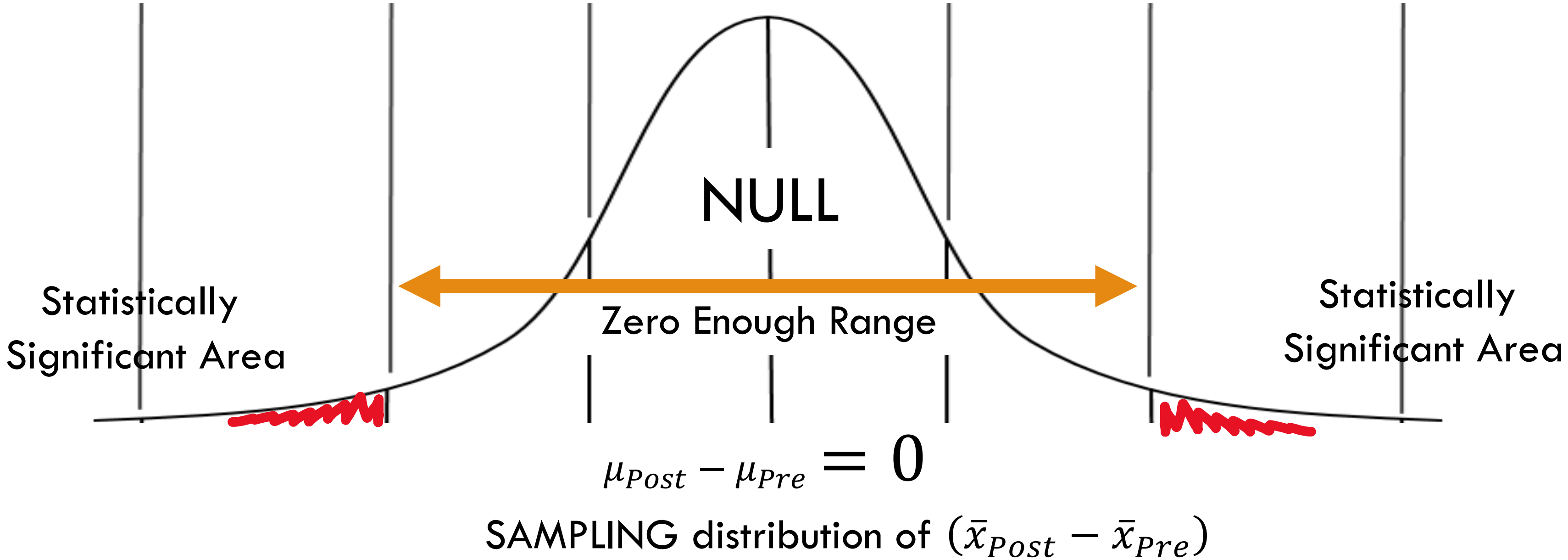
$\mu_{Post} - \mu_{Pre} = 0$

If the NULL were true, would you expect that every sample of $(\bar{x}_{Post} - \bar{x}_{Pre}) = 0$?

Why or why not?

NULL is Zero

- Because of random error and chance, the difference will not always be zero.
  - Just like before, we want to know a range of reasonable zero enough values (a range of "zero enough")
  - If our t-statistic is far enough outside the "zero enough" range, we can reject the null and conclude there is a difference in pre and post scores



Atkins Diet

- Proponents of the Atkins Diet claim that after 6 months of no carbs, dieters will have substantial change in weight. A random sample of eight dieters showed the following weights before and after completing the diet regimen. At the .01 level of significance, test whether the Atkins Diet has an effect on weight.

DIETER	BEFORE(X_1)	AFTER (X_2)
JC	155	154
LT	228	207
TJ	141	147
MT	162	157
ZB	211	196
DD	164	150
RI	184	170
PT	172	165

Step 1: State the Hypotheses

Step 1:

$$H_0\!: \mu_D = 0$$
$$H_1\!: \mu_D \neq 0$$

H_0: The true average weight change after 6 months of a no carb diet is 0.

H_1: The true average weight change after 6 months of a no carb diet is not 0.

This is a two-tailed test because we are asking about weight change– you could gain or lose weight.

Step 2 and 3: Significance and Test

Step 2:

$$\alpha = .01$$

Step 3:

$$t = \frac{\bar{x}_D - \mu_D}{{s_D}/{\sqrt{n}}}$$

Step 4: Get Critical Value(s)

Step 4:

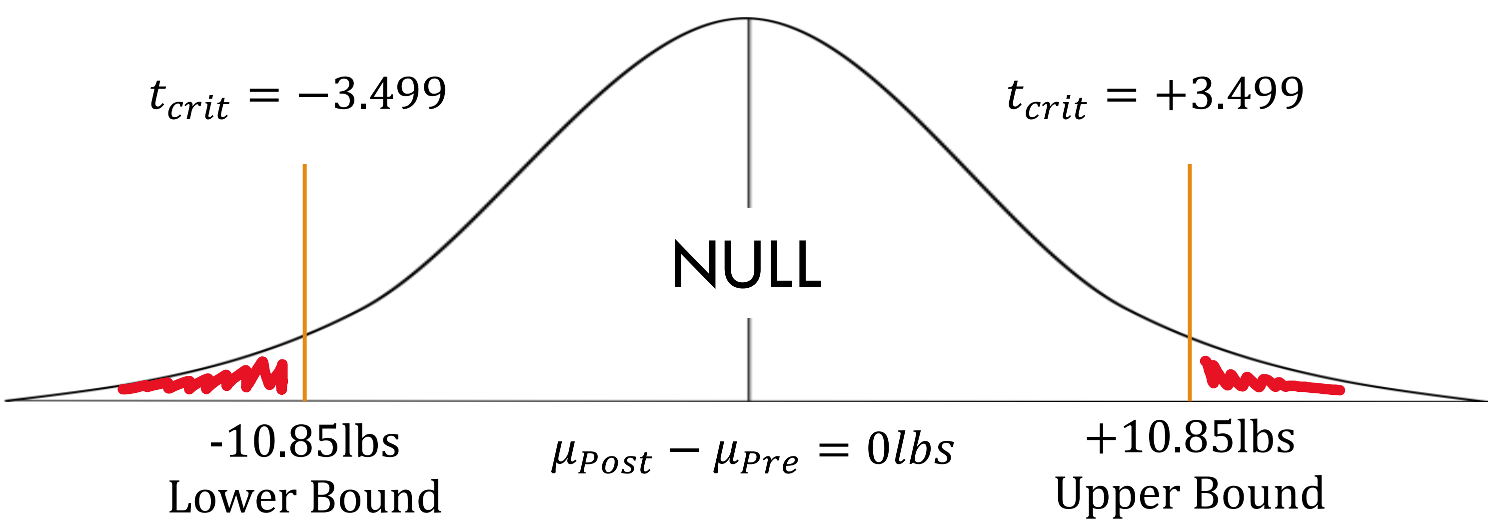
$$\alpha = .01$$
$$n = 8$$
$$df = 8 - 1 = 7$$

Two-tailed test, meaning that our two critical values should cut off tails with probabilities of .005 each, half a percent in each tail.

The t-critical values that satisfies $df = 7$ and $t_{.005}$ are:
$$t_{crit} = \pm 3.499$$

Visualizing Critical Cut Offs

- If our test statistic is greater or less than the critical t of ± 3.499, or more than 10.85 pounds of difference, then we can be confident that the true difference between pre and post measurements is not equal to zero.



Upper and Lower Bounds $= 0 \pm 3.499 * \underbrace{(3.1)}_{\text{Standard Error}}$

Step 5: Compute the Test Statistic

Step 5a: First, calculate the mean difference

$$\bar{x}_D = \frac{-1+-21+6+-5+-15+-14+-14+-7}{8} = \frac{-71}{8} = -8.875 \approx -8.9$$

Mean Difference

DIETER	AFTER (X_2)	BEFORE(X_1)	D ($X_2 - X_1$)
JC	154	155	-1
LT	207	228	-21
TJ	147	141	+6
MT	157	162	-5
ZB	196	211	-15
DD	150	164	-14
RI	170	184	-14
PT	165	172	-7

Add these together and divide by n to get the average amount of difference between pre and post weight, ≈ -8.9 lbs, i.e. people lost on average about -8.9 pounds.

Variance of a Dependent Samples t-test

1) Get the differences (D) between Pre and Post

3) Subtract average difference ($\bar{x}_D$) from each difference (D) score

DIETER	AFTER (X_2)	BEFORE (X_1)	D ($X_2 - X_1$)	D - $\bar{x}_D$	(D - $\bar{x}_D$)^2	Squared Deviations
JC	154	155	-1	-8.9	7.9^2	62.41
LT	207	228	-21	-8.9	-12.1^2	146.41
TJ	147	141	+6	-8.9	14.9^2	222.01
MT	157	162	-5	-8.9	3.9^2	15.21
ZB	196	211	-15	-8.9	-6.1^2	37.21
DD	150	164	-14	-8.9	-5.1^2	26.01
RI	170	184	-14	-8.9	-5.1^2	26.01
PT	165	172	-7	-8.9	1.9^2	3.61

2) Average the differences
$\bar{x}_D$ = -8.875, rounded to -8.9

4) Square the differences $(D - \bar{x}_D)^2$

Variance of a Dependent Samples t-test

DIETER	AFTER (X_2)	BEFORE (X_1)	D ($X_2 - X_1$)	D - $\bar{x}_D$	(D - $\bar{x}_D$)^2	Squared Deviations
JC	154	155	-1	-8.9	7.9^2	62.41
LT	207	228	-21	-8.9	-12.1^2	146.41
TJ	147	141	+6	-8.9	14.9^2	222.01
MT	157	162	-5	-8.9	3.9^2	15.21
ZB	196	211	-15	-8.9	-6.1^2	37.21
DD	150	164	-14	-8.9	-5.1^2	26.01
RI	170	184	-14	-8.9	-5.1^2	26.01
PT	165	172	-7	-8.9	1.9^2	3.61

5) Add these together and divide by (n-1) to get the variance (s^2) of amount of weight loss amongst the participants, $= 76.98$

$$\text{Variance} = 538.88 \,/\, (8\text{-}1) = 76.98$$

$$\text{Standard Deviation} = \sqrt{76.98} = 8.77$$

6) To get the standard deviation (the amount of variation in the sample in the original units) square root the variance, $\mathrm{SD}\ (s) = 8.77$

Step 5: Compute the Test Statistic

- Step 5b: Then, calculate the standard deviation

$$s_D = \sqrt{\frac{(-1+8.9)^2+(-21+8.9)^2+(6+8.9)^2+(-5+8.9)^2+(-15+8.9)^2+(-14+8.9)^2+(-14+8.9)^2+(-7+8.9)^2}{8-1}}$$

$s_D \approx 8.77$ $\bar{X}_D = -8.9$

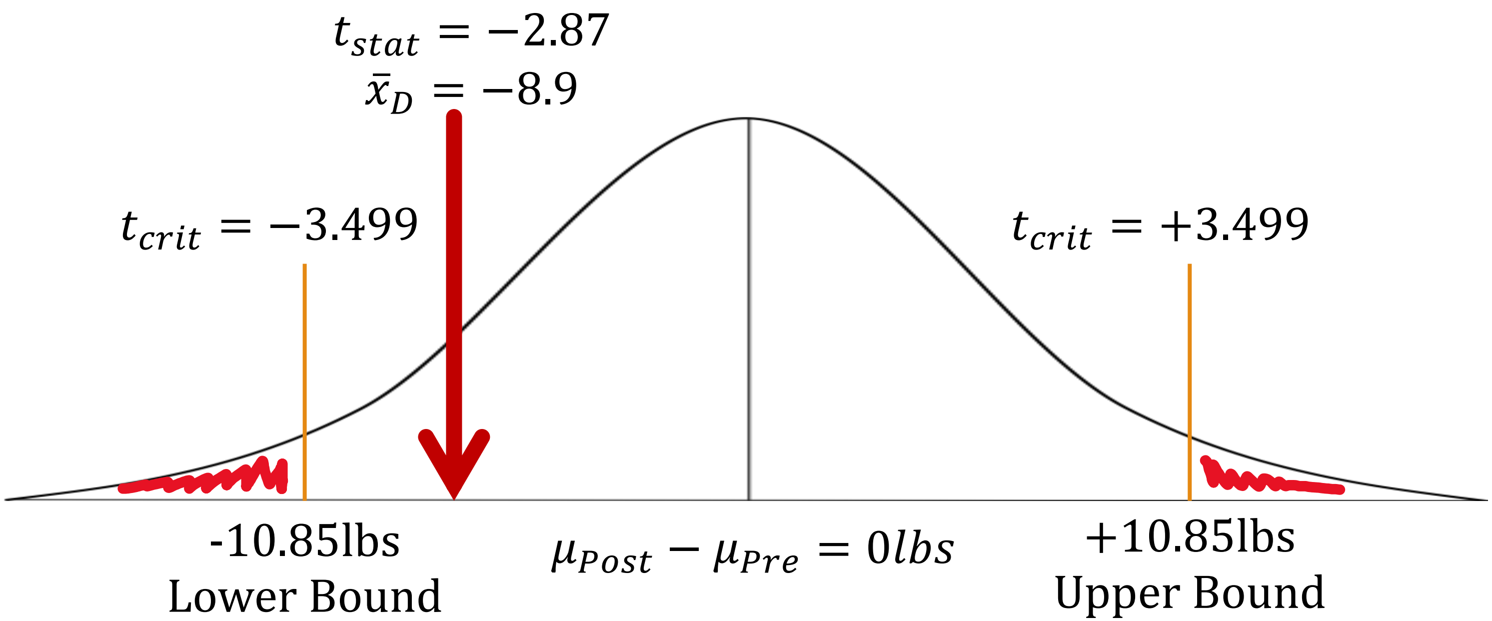
Substituting our sample statistics into the formula below:

$$t = \frac{\bar{x}_D - \mu_D}{s_D/\sqrt{n}} = \frac{-8.9 - 0}{8.77/\sqrt{8}} \approx \frac{-8.9}{3.1} \approx -2.87$$

So, $t_{stat} = -2.87$

Visualizing Critical Cut Offs

- If our test statistic is greater or less than ± 3.499, or more than 10.85 pounds of difference, then we can be confident that the true difference between the samples is not equal to zero.



Step 6: Draw Conclusions

Step 6:

Our $t_{stat} = -2.8$, and our $t_{crit} = \pm 3.499$.

Our t_{stat} is not past our t_{crit}, so we *fail to reject* H_0.

- "We fail to reject H_0. There is not enough evidence to reject the null hypothesis that the true mean difference in weight is equal to 0."
- Although we did see weight loss for most of the participants, we may have found a non-significant test result due to:
  - Inadequate power from small sample size
  - A relatively high amount of variation
    - Standard deviation and standard error are large
  - A strict alpha level of .01

Confidence Intervals for Dependent Samples

$$CI = \bar{x}_D \pm t_{critical} * \frac{s}{\sqrt{n}};$$

$$df = n - 1$$

Margin of Error:
The amount of error based on SE
and a desired level of confidence
in the original number scale

$$CI = Point\ Estimate \pm \overbrace{t_{critical} * Standard\ Error}^{\text{Margin of Error}}$$

Point Estimate:
Statistic (ex. sample mean value) in the original number scale, this is the mean difference $\bar{x}_D$

t-critical:
The level of confidence I want based on the t-table

Standard Error (SE):
Standard deviation of the sampling distribution

Confidence Intervals for Dependent Samples

99% CI = [-19.75, 1.95]

We are 99% confident that on average people will lose as much as 19.75 pounds OR gain as much as 1.95 on this diet.

Margin of Error:
The amount of error based on SE and a desired level of confidence in the original number scale

$$CI = -8.9 \pm \overbrace{3.499 * 3.1}$$

Point Estimate:
Statistic (ex. sample mean value) in the original number scale, this is the mean difference $\bar{x}_D$

t-critical:
The level of confidence I want based on the t-table

Standard Error (SE):
Standard deviation of the sampling distribution

Confidence Intervals for Dependent Samples

$$99\%\ \mathrm{CI} = [-19.75, 1.95]$$

0

We are 99% confident that on average people will lose as much as 19.75 pounds OR *gain* as much as 1.95 on this diet. Notice how the interval contains zero… If a confidence interval contains zero, it is not significant. This is because zero is the expected mean difference of the NULL. So this sample is not distinguishable from the NULL population. Fail to reject the null.

Standard Error and Sample Size, Again

What would happen to the standard error, t-statistic, and critical value if we increased the sample?

(Let's pretend that the variation and mean are the same for this example.)

Standard Error and Sample Size, Again

What would happen to the standard error and to the t-statistic if we increased the sample?

(Let's pretend that the variation and mean are the same for this example.)

Doubling sample from $\mathrm{n} = 8$ to $\mathrm{n} = 16$

$$t = \frac{\bar{x}_D - \mu_D}{s_D/\sqrt{n}} = \frac{-8.9 - 0}{8.77/\sqrt{8}} \approx \frac{-8.9}{3.1} \approx -2.86$$

So, $t_{stat} = -2.86$

Standard Error and Sample Size, Again

What would happen to the standard error and to the t-statistic if we increased the sample?

(Let's pretend that the variation and mean are the same for this example.)

Doubling sample from $\mathrm{n} = 8$ to $\mathrm{n} = 16$

$$df = 15, t_{crit} = \pm 2.947$$

$$t = \frac{\bar{x}_D - \mu_D}{s_D / \sqrt{n}} = \frac{-8.9 - 0}{8.77 / \sqrt{16}} \approx \frac{-8.9}{2.19} \approx -4.06$$

So, $t_{stat} = -4.06$

Standard Error and Sample Size, Again

- Increasing the sample size:
  - Gives us more degrees of freedom and a lower critical value
  - Decreases standard error
  - Increases the t-statistic

Doubling sample from $\mathrm{n} = 8$ to $\mathrm{n} = 16$

$$df = 15,\ t_{crit} = \pm 2.947$$

$$t = \frac{\bar{x}_D - \mu_D}{s_D / \sqrt{n}} = \frac{-8.9 - 0}{8.77 / \sqrt{16}} \approx \frac{-8.9}{2.19} \approx -4.06$$

So, $t_{stat} = -4.06$

Try it. Journaling and Mood

- Journaling is suspected to be positively associated with wellbeing (i.e. increase wellbeing). To test this, I will give out a Wellbeing Scale to a group of 16 subjects at the start of the study. After one month of journaling, the Wellbeing Scale is readministered. The average difference ($\bar{x}_D$) in Wellbeing was 5.8 points, with a standard deviation of 1.2. At the .05 level of significance, test whether journaling increases wellbeing.

Step 1: State the Hypotheses

Step 1:

$$H_0\colon \mu_D \leq 0$$
$$H_1\colon \mu_D > 0$$

H_0: The true average increase in wellbeing after one month of journaling less than or equal to 0.

H_1: The true average increase in wellbeing after one month of journaling is greater than 0.

This is a one-tailed test, "increase" so right tail

Step 2 and 3: Significance and Test

Step 2:

$$\alpha = .05$$

Step 3:

$$t = \frac{\bar{x}_D - \mu_D}{s_D / \sqrt{n}}$$

Step 4: Get Critical Value(s)

Step 4:

$$\alpha = .05$$

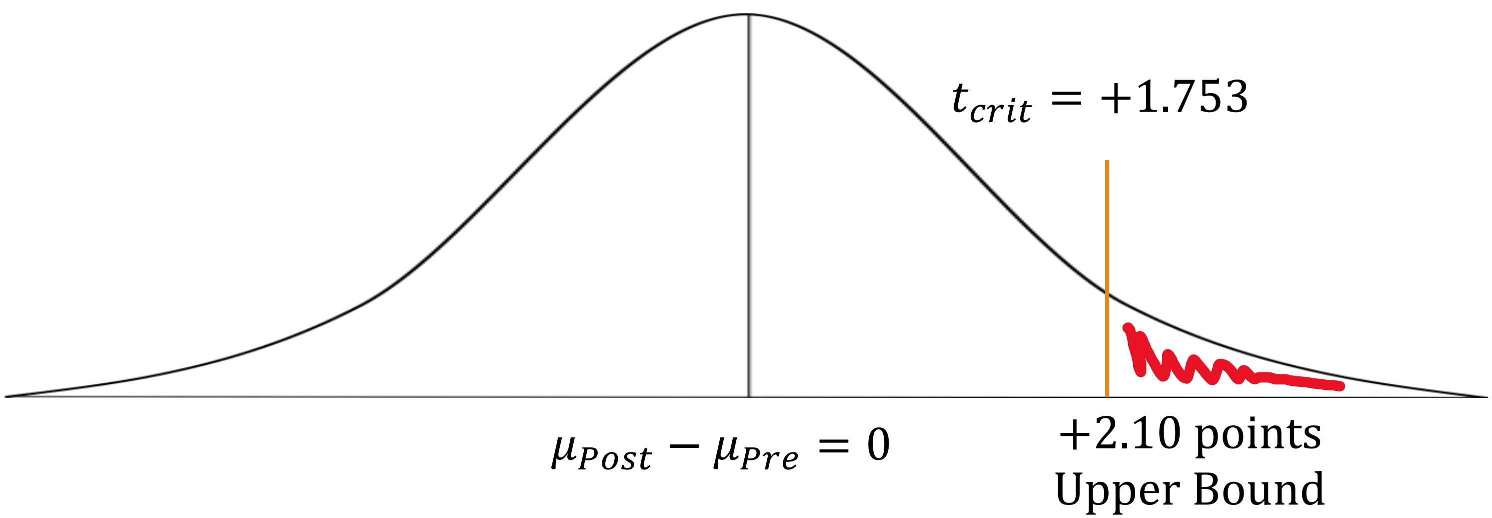
$$n = 16$$

$$df = 16 - 1 = 15$$

$$t_{crit} = +1.753$$

Visualizing Critical Cut Offs

- If our test statistic is greater than $+1.753$, or more than 2.10 points of difference, then we can be confident that the true difference between the samples is not equal to zero.



Upper Bounds $= 0 + 1.753 * (1.2)$

Step 5: Compute the Test Statistic

Substituting our sample statistics into the formula below:

$s_D = 1.2$

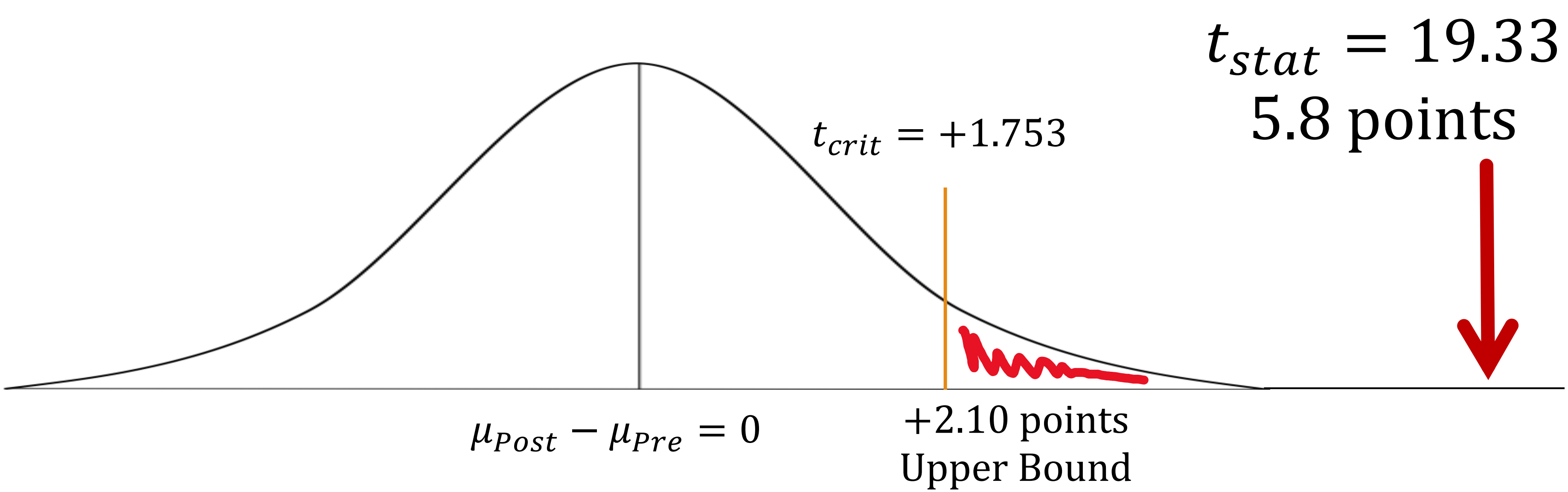
$\bar{x}_D = +5.8$

$$t = \frac{\bar{x}_D - \mu_D}{s_D / \sqrt{n}} = \frac{5.8 - 0}{1.2 / \sqrt{16}} \approx . \frac{5.8}{0.3} \approx 19.33$$

So, $t_{stat} = 19.33$

Visualizing Critical Cut Offs

- If our test statistic is greater or less than $+1.753$, or more than 2.10 points of difference, then we can be confident that the true difference between the samples is not equal to zero.



Upper Bounds $= 0 + 1.753 * (1.2)$

Step 6: Draw Conclusions

Step 6:

Our $t_{stat} = 19.33$, and our $t_{crit} = +1.753$.

Our t_{stat} is past our t_{crit}, so we to reject H_0.

$$95\% \text{ CI: } [5.27, 6.33]$$

- "We reject the H_0. There is enough evidence to reject the null hypothesis that the mean difference increase in wellbeing after one month of journaling is not equal to 0."
  - "Reject the null hypothesis. Journaling for one month appears to increases wellbeing on average between 5.27 to 6.33 points."

All t-test…

t-tests: Different Types, Same Logic

- Though they look different, the essence of the t-tests is the same…

$$\text{t-value} = \frac{\text{Difference between means}}{\text{Variation (SE) of the group(s)}}$$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

One Sample t-test

$$t = \frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$

Independent Samples t-test

$$t = \frac{\bar{x}_D - \mu_D}{s_D/\sqrt{n}}$$

Dependent Samples t-test

Which t-test do I use?

How many samples?

One Sample →

One Sample t-test:
Comparing one sample mean to some known mean
H_0: μ_1 $(=, \leq, \geq)$ μ_0
H_1: μ_1 $(\neq, <, >)$ μ_0
μ_1 = your sample, μ_0 some known mean
df = n-1

Two Samples ↓

Two Independent Groups or Repeated Measure of One Group?

Independent Groups ↓

Independent Samples t-test:
Comparing the difference between two groups' sample means to a Null
H_0: μ_1 $(=, \leq, \geq)$ μ_2 *or*
H_0: $\mu_1 - \mu_2 = 0$
H_1: μ_1 $(\neq, <, >)$ μ_2 or
H_1: $\mu_1 - \mu_2 \neq 0$
μ_1 = group 1, μ_2 group 2
df = n-2

Repeated Measure ↓

Dependent Samples t-test:
Comparing the mean difference between before and after scores to a Null
H_0: μ_{Post} $(=, \leq, \geq)$ μ_{Pre} *or*
H_0: $\mu_{Post} - \mu_{Pre} = 0$
H_1: μ_{Post} $(\neq, <, >)$ μ_{Pre} or
H_1: $\mu_{Post} - \mu_{Pre} \neq 0$
μ_{Post} = After, μ_{Pre} Before
df = n-1

- All t-tests…
  - Are hypothesis tests
  - That compare some sort of average score to another hypothesized average score
  - With an unknown σ
  - And use a Standard Error (SE)
    - With degrees of freedom to correct for bias
      - (n-1 or n-2)

Try it.

Which test and how how many tails should you use for the following examples?

- I want to compare the average exam grade of a sample of n=10 of my students in this Stats Literacy course to a sample of n=10 students in the other section of the Stats Literacy course.
- I want to compare a sample of UT students' average hours of sleep to the national average (unknown σ).
- I want to know if the Premont school district is performing lower than the state average on the STAR test.
- I want to know which drug is better at reducing flu symptoms. I create two groups, one receives drug A, the other drug B.
- I want to know if allowing people to fidget changes their average math score. I first tell the students they are not allowed to fidget for one week, then the following week allow them to fidget as much as they would like. I then compare their two averages.

Try it.

Which test and how how many tails should you use for the following examples?

- I want to compare the average exam grade of a sample of n=10 of my students in this Stats Literacy course to a sample of n=10 students in the other section of the Stats Literacy course.
  - Independent Samples, Two Tailed
- I want to compare a sample of UT students' average hours of sleep to the national average (unknown σ).
  - One Sample, Two Tailed
- I want to know if the Premont school district is performing lower than the state average on the STAR test.
  - One Sample, One Tailed (Left)
- I want to know which drug is better at reducing flu symptoms. I create two groups, one receives drug A, the other drug B.
  - Independent Samples, Two Tailed
- I want to know if allowing people to fidget changes their average math score. I first tell the students they are not allowed to fidget for one week, then the following week allow them to fidget as much as they would like. I then compare their two averages.
  - Dependent Samples, Two Tailed

Up Next…

- We've been dealing with one or two samples thus far, but what if we ant to compare more than two groups?
- We need an…

ANOVA

Dependent Samples t-test in R

Dependent Samples t-test in R (Summary Statistics)

Using summary data.

```r
########################################
####### t-test Dependent Samples ########
############# TWO Tailed ################
########################################

# Using summary data, filling in the information

meand_dif <- -8.9
sd_d <- 8.77
n <- 8
df_dependent <- n -1

# Find the critical value, 99% confidence
two_tail_crit_t_99 <- qt(p = c(.005, .995), df = df_dependent) # t-critical: -3.499, 3.499

# Calculate the test statistic
t_stat_dependent <- meand_dif/(sd_d/sqrt(n)) #t-statistic = -2.87

# Our t-statistic (-2.87) is not further out than the critcal t values (-3.499, 3.499)
# We FAIL to reject the null hypothesis.
```

Dependent Samples t-test in R (Data)

```
# Same question but using data instead of summary staistics

pre <- c(155, 228, 141, 162, 211, 164, 184, 172)
post <- c(154, 207, 147, 157, 196, 150, 170, 165)

t.test(post, pre, paired = T, conf.level = .99)
# Same result, our t-statistic (-2.87) is not further out than the critcal t values (-3.499, 3.499)
# You can also see the confidence interval contains zero and is this not significant
# Notice the p-value is less than .05 but not less than .01, our level of significance for this example
# We FAIL to reject the null hypothesis.
```

Dependent Samples t-test in R (Data)

- We'll use another (bigger) diet weight loss data set to see if there is a statistically significant difference in the weight of participants before and after a 6-week diet.

Pre vs. Post Diet Weight t-test

- Does Pre vs. Post Diet weight differ significantly?
  - Test at .05 level.
- H_0: Pre vs. Post Diet weight do **NOT** differ significantly from each other (There is no difference in pre and post weight).
- H_1: Pre vs. Post Diet weight do differ significantly from each other (There is no difference in pre and post weight).

Descriptive Statistics

- First, let's look at some descriptive statistics.
- Does it look like pre and post weights differ *significantly*?

```
> mean(diet_data$pre.weight)
[1] 159.3947
> sd(diet_data$pre.weight)
[1] 17.54999
```

```
> mean(diet_data$weight6weeks)
[1] 150.6711
> sd(diet_data$weight6weeks)
[1] 17.82911
```

159.3947 lbs -150.6711 lbs = 8.7236

Pre vs. Post Diet Weight t-test, R Output

- Here we have output from a paired samples t-test in R.
- We have 75 degrees of freedom, so we know we had (N-1) 76 observations (people).
- Our t-value is 13.962, which even without looking at a t-table or p-value, we know is very big.

```
        Paired t-test

data:  diet_data$pre.weight and diet_data$weight6weeks
t = 13.962, df = 75, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 7.479002 9.968366
sample estimates:
mean of the differences
               8.723684
```

Pre vs. Post Diet Weight t-test, R Output

- The p-value is 0.000, so p < .05, we reject our null hypothesis that the pre and post weights do *not* differ.
- The confidence interval confirms this, too
  - 95% CI [7.48, 9.97]
  - Notice how the interval does not contain zero

```
        Paired t-test

data:  diet_data$pre.weight and diet_data$weight6weeks
t = 13.962, df = 75, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 7.479002 9.968366
sample estimates:
mean of the differences
               8.723684
```

Pre vs. Post Diet Weight t-test, R Output

- We would conclude that the pre and post diet weights are not the same, i.e. there is a statistically significant difference in weight.
- The confidence interval confirms this, too
  - 95% CI [7.48, 9.97]
  - $t(75) = 13.96, p = 0.00$

```
        Paired t-test

data:  diet_data$pre.weight and diet_data$weight6weeks
t = 13.962, df = 75, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 7.479002 9.968366
sample estimates:
mean of the differences
               8.723684
```

Pre vs. Post Diet Weight t-test, R Code

Using data.

```
#########################################
####### t-test Dependent Samples ########
############# TWO Tailed ################
############## with Data ################
#########################################

diet_data <- read.csv("dietdataset.csv")

# Looking at the mean and standard deviation of the weights before dieting
mean(diet_data$pre.weight)
sd(diet_data$pre.weight)

# Looking at the mean and standard deviation of the weights after dieting
mean(diet_data$weight6weeks)
sd(diet_data$weight6weeks)

# Conduct the t-test, to do a Dependendent samples, make sure to tell R
# "paired = TRUE" which let's R know it's a dependent samples test
t.test(diet_data$pre.weight, diet_data$weight6weeks, paired = TRUE)
```

Data Source: https://www.sheffield.ac.uk/polopoly_fs/1.570199!/file/stcp-Rdataset-Diet.csv.