Inference for Two Population Means

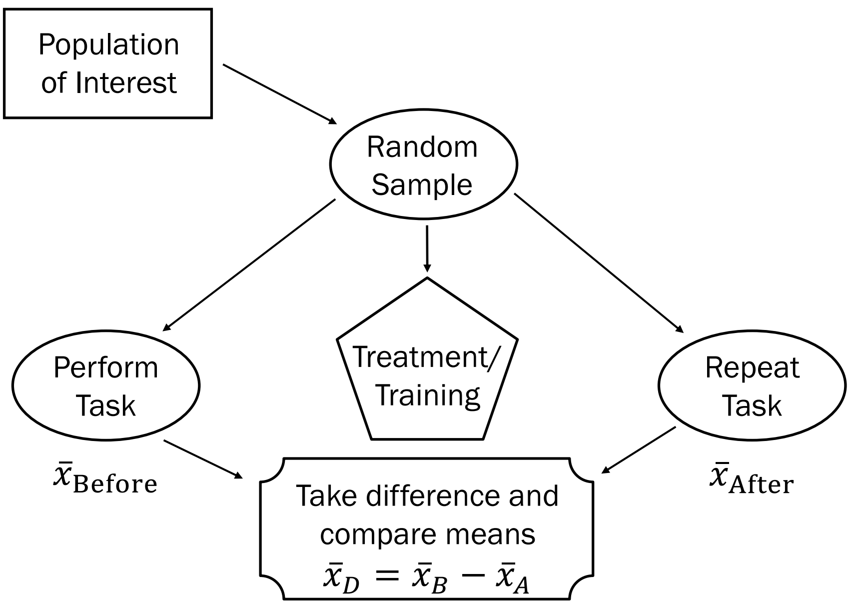
- Testing for a Difference in Paired Data
- Testing the Difference Between Two Means in Independent Samples
- Pooled Two-Sample t-Test

University of Pittsburgh

Motivation: Matched Pairs Test

- **Scenario:** Recall the example from Lecture 7 where we want to test the effectiveness of a new weight loss program. Recruit 25 people who are weighed at the beginning of the study, follow the plan for 6 weeks, and get weighed again at the end.
- **Question:** Did people lose a significant amount of weight while on the weight loss program?
- **Observations:**
  - Comparing ____________ data across ____ levels of a __________ variable:
    - **Categorical:** ______ weight was taken (________________)
    - **Quantitative:** ________
  - Have two samples that are ________________
    - Using the ___________ in each sample causes ___________ → Data is _______
  - Comparing __________ across samples

Visualizing the Matched Pairs Test



1. Take random sample from the population
2. Subjects do three things
   - Perform the task
   - Undergo training or treatment
   - Repeat the task
3. Match up each subjects' observations and take difference
4. Compare before and after means

Goal: Determine if there is a significant difference in the ________________ means

Matched Pairs Test

Step	Description
Used for	Comparing the means of two dependent population means
Conditions	Shape of the sampling distribution of the differences between the means must be normal
Notation	$\bar{X}_d = \bar{X}_{\text{Before}} - \bar{X}_{\text{After}}$ μ_d: Hypothesized difference between means – usually 0
Test Statistic	$t = \frac{\bar{X}_d - \mu_d}{s_d / \sqrt{n_d}}$ with $n-1$ degrees of freedom
Confidence Interval	$\bar{X}_d \pm t_{n-1} \frac{s_d}{\sqrt{n_d}}$ with $n-1$ degrees of freedom

Note: *Once the differences between paired observations have been taken, the matched pairs test is identical to a one-sample t-test.*

Example: Matched Pairs Test

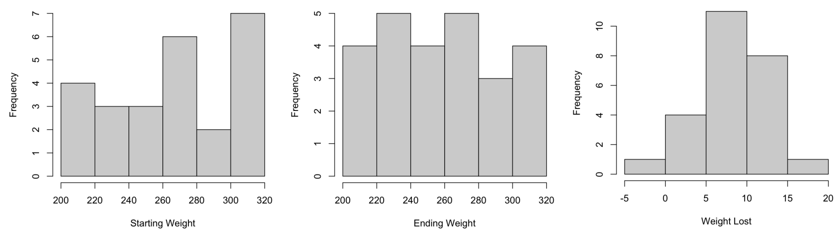
- **Scenario:** Recall the example from Lecture 7 where we want to test the effectiveness of a new weight loss program. Recruit 25 people who are weighed at the beginning of the study, follow the plan for 6 weeks, and get weighed again at the end.
- **Question:** Did people lose a significant amount of weight while on the weight loss program using a 5% level of significance?
- **Hypotheses:**
  - H_0: ________ → H_0: __________ → H_0: ______
    - ___________ in starting and ending weights
  - H_A: ________ → H_A: __________ → H_A: ______
    - Starting weight is _______ than ending weight
- **Note:** All three forms of these hypotheses are acceptable.

Example: Matched Pairs Test (Cont.)

Original Data

	starting	ending	difference
1	256	258	−2
2	319	315	4
3	240	234	6
4	285	279	6
5	320	313	7
6	209	202	7
7	277	269	8
8	230	222	8
9	263	253	10
10	306	296	10
11	274	264	10
12	217	207	10
13	245	234	11
14	245	234	11
15	314	302	12
16	240	228	12
17	282	279	3
18	303	290	13
19	214	205	9
20	266	252	14
21	311	297	14
22	320	305	15
23	268	263	5
24	261	245	16
25	218	217	1

Perform inference on __________



Observation: The differences between paired observations can be _______ even if the original samples are not.

Normality Condition: __________

- Rule of Thumb ___: Histogram of differences in starting and ending weights is close to ______/slightly ____-skewed, but the sample size is ___________

Example: Matched Pairs Test (Cont.)

- **Summary:**

Statistic	Starting	Ending	Difference
Mean	267.3	258.5	8.8
Std. Dev.	35.95	35.26	4.425

- **Test Statistic:** $t =$ ____________ $=$ ______ $=$ _____
  - **Interpretation:** If the weight loss program had __________, then an average weight loss of ____ pounds from a random sample of ___ people is _____ ______________________ what we expected.

- **R Code:**

Before sample — After sample

```
t.test(weight$starting, weight$ending,
    alternative = "greater", mu = 0, paired = TRUE)
```

Alternative hypothesis form — Hypothesized difference — Tell R the samples are paired

Example: Matched Pairs Test (Cont.)

- **Output:**

```
data:  weight$starting and weight$ending
t = 9.9428, df = 24, p-value = 2.749e-10
alternative hypothesis: true difference in means is greater than 0
```

- **P-Value:** _________________
- **Confidence Interval:** _____________________ $=$ ______________
- **Conclusions:**
  - __________ and conclude that the average amount of __________ while on the program for 6 weeks is _________________________.
    - $p = 2.749 \times 10^{-10} < 0.05$
  - We are 95% confident that the ________________________ during 6 weeks on the program is _____________________ pounds.

Motivation: Difference of Two Means Test

- **Scenario:** Recall the example from Lecture 7 where we compared the lengths of business and personal Uber rides. Uber driver tracks the length (in miles) of 32 business rides and 50 personal rides.
- **Question:** Is there a significant difference in the lengths of the two ride types?
- **Observations:**
  - Comparing ____________ data across two levels of a __________ variable:
    - **Categorical:** ____________ (business or personal)
    - **Quantitative:** ______ (in miles)
  - Two ___________ samples
    - Ride cannot be for ____________
    - Data cannot be _______

Difference of Two Means Test

Step	Description
Used for	Comparing the means of two independent population means
Conditions	Shape of the sampling distribution must be normal for both means
Notation	Subscripts 1 and 2 refer to the groups being compared
Test Statistic	$t = \frac{(\bar{X}_1 - \bar{X}_2) - \mu_d}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ with $\frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$ degrees of freedom
Confidence Interval	$(\bar{X}_1 - \bar{X}_2) \pm t_{df}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ with above number of degrees of freedom

__Note:__ You will __never__ need to calculate the degrees of freedom by hand for this test. They will always be calculated using R.

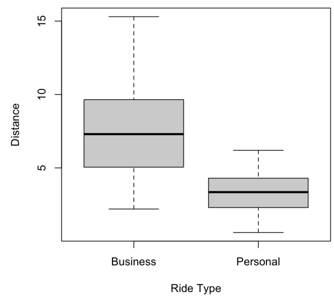
Example: Difference of Two Means Test

- **Scenario:** Recall the example from Lecture 7 where we compared the lengths of business and personal Uber rides. Uber driver tracks the length (in miles) of 32 business rides and 50 personal rides.
- **Question:** Is there a significant difference in the lengths of the two ride types at the 5% level of significance?
- **Hypotheses:**
  - H_0: ____________ → H_0: ______________
    - ______________ in mean Uber ride length for business and personal trips
  - H_A: ____________ → H_A: ______________
    - Mean business and personal Uber ride lengths __________________, but we do not have a ________ as to which is _________

Example: Difference of Two Means Test (Cont.)

- **Summary:**

Statistic	Business	Personal
Mean	7.59	3.33
Std. Dev.	3.29	1.36
Sample Size	32	50



- **Normality Condition:** _____________
  - Rule of Thumb ___: Both sample sizes are ____________
    - Neither sample is _________ skewed– Rule of Thumb ___ also applies
- **Test Statistic:** $t =$ _________________ $=$ __________ $=$ ________
  - **Interpretation:** If there is actually ______________ in the length of business and personal Uber rides, then our sample mean ___________________ is ______________________________ what we expected.

Difference of Two Means Test Using R

- **Function:**

```
t.test(uber$distance ~ uber$type,
  alternative = "two.sided", mu = 0, paired = FALSE)
```

- uber$distance: Quantitative variable
- uber$type: Categorical variable
- alternative = "two.sided": Alternative hypothesis form
- mu = 0: Hypothesized difference
- paired = FALSE: Tell R the samples are not paired

```
        Welch Two Sample t-test

data:  uber$distance by uber$type
t = 6.9466, df = 37.891, p-value = 2.949e-08
alternative hypothesis: true difference in means between group Business and group Personal is not equal to 0
95 percent confidence interval:
 3.015233 5.495767
sample estimates:
mean in group Business mean in group Personal
                7.5875                 3.3320
```

Note: R is very specific in calculating degrees of freedom. We can round this to 38.

Example: Difference of Two Means Test (Cont.)

- **P-Value:** $p =$ ____________
  - **Interpretation:** If there is actually no difference in the length of business and personal Uber rides, then the probability we would get a sample mean difference ________________________________.

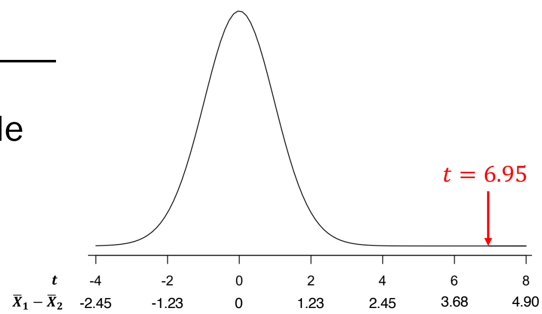
- **95% Confidence Interval:** ______________ $\pm$ ____________________

  ____

  - **Note:** 38 df does not appear in the t-table
    - Use the **qt** function to get the critical value

```
qt(0.975, 38)
```

    0.975: Percentile; 38: df



Example: Difference of Two Means Test (Cont.)

- **Conclusions:**
  - _________ and conclude that the average length of business and personal Uber rides are _____________________.
    - $p =$ ___________ $<$ _____
  - We are 95% confident that the true ________________ in the length of _______________________ Uber rides is between _________________ miles.
    - This is consistent with the hypothesis test as the interval _________________.
  - Business Uber rides appear to be ______ as the entire confidence interval is __________.
    - Hypotheses were set up as: Business - Personal

POOLED STANDARD DEVIATION

- If the standard deviations of the two samples are approximately the same, the individual standard deviations in the two-sample t-test may be replaced with a pooled estimate of the variation in the groups.

- **Pooled standard deviation:** $s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{(n_1-1)+(n_2-1)}}$
  - Weights the standard deviations according to the sample size and returns an estimate of the combined variability that is in between s_1 and s_2.

- ***Rule of Thumb:*** *Standard deviations can be pooled if* $0.5 < \frac{s_1^2}{s_2^2} < 2$.

POOLED TWO-SAMPLE T-TEST

Step	Description
Used for	Comparing the means of two independent population means when the variability in the groups is approximately equal
Conditions	Shape of the sampling distribution must be normal for both means and $0.5 < \frac{s_1^2}{s_2^2} < 2$
Test Statistic	$t = \frac{(\bar{X}_1-\bar{X}_2)-\mu_d}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ with $n_1 + n_2 - 2$ df
Confidence Interval	$(\bar{X}_1-\bar{X}_2) \pm t_{n_1+n_2-2} \times s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$ with $n_1 + n_2 - 2$ df

EXAMPLE: POOLED TWO-SAMPLE T-TEST

- **Scenario:** Tailed frogs have begun to disappear from certain habitats. Researchers identified 18 habitats with tailed frogs (Y) and 31 habitats without tailed frogs (N) and measured the water temperature in each (in degrees Fahrenheit), suspecting that differing water temperatures may contribute to the disappearance.
- **Question:** Is the mean water temperature in the areas without tailed frogs significantly colder than the areas with tailed frogs at the 5% level of significance?
- **Hypotheses:**
  - H_0: __________ → H_0: __________
  - H_A: __________ → H_A: __________

Summary Statistics by Group

- To obtain the summary statistics (mean, SD, and sample size) by group in R, use the **aggregate** function:

Quantitative variable · Categorical variable · Desired statistic

```
aggregate(frogs$temp, list(frogs$type), FUN = mean)
```

- The last argument can be changed to `FUN = sd` and `FUN = length` to find standard deviations and sample sizes by group respectively.

	temp	type
1	60.7	no
2	57.2	yes
3	58.9	yes
4	56.4	no
5	57.9	no
6	62.0	yes
7	59.3	yes
8	62.1	no

Mean

```
  Group.1        x
1      no 59.14839
2     yes 59.86667
```

Standard Deviation

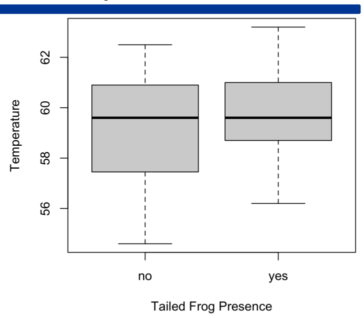
```
  Group.1        x
1      no 2.277992
2     yes 1.960192
```

Sample Size

```
  Group.1  x
1      no 31
2     yes 18
```


Example: Pooled Two-Sample T-Test (Cont.)

- **Normality Condition:** ___________
  - **No Tailed Frogs:** Rule of Thumb ___ applies as the sample size is __________ despite the __________
  - **Tailed Frogs:** Rule of Thumb ___ applies as the data is close to __________ with a sample size of at least ___



- **Question:** Should the standard deviations be pooled?
- **Answer:** _____ - Ratio is ______________________
  - **Check:** $\frac{s_1^2}{s_2^2} =$ ______ $=$ ______ $=$ _____
  - **Pooled SD:** $s_p = \sqrt{\frac{(31-1)(2.28)^2+(18-1)(1.96)^2}{(31-1)+(18-1)}} =$ _____

Standard Deviation

```
  Group.1        x
1      no 2.277992
2     yes 1.960192
```

Sample Size

```
  Group.1  x
1      no 31
2     yes 18
```


Pooled Two-Sample T-Test Using R

- **Function:**

Quantitative variable · Categorical variable · Alternative hypothesis form

```
t.test(frogs$temp ~ frogs$type, alternative = "less",
   mu = 0, paired = FALSE, var.equal = TRUE)
```

Hypothesized difference · Tell R the samples are not paired · Tell R to assume the spreads are equal

```
	Two Sample t-test

data:  frogs$temp by frogs$type
t = -1.1178, df = 47, p-value = 0.1347
alternative hypothesis: true difference in means between group no and group yes is less than 0
95 percent confidence interval:
      -Inf 0.3599186
sample estimates:
 mean in group no mean in group yes 
         59.14839          59.86667 
```

$n_1 + n_2 - 2$ (df = 47)

R will always do the order of subtraction in alphabetical order

EXAMPLE: POOLED TWO-SAMPLE T-TEST (CONT.)

- **Test Statistic:** $t = $ ____________ $=$ ______
- **P-Value:** $p = $ ________
- **95% Confidence Interval:** ____________ $\pm$ ____________________

  ____

- **Conclusions:**
  - _____________ and conclude that the average water temperature for habitats ________________________ are ____________________.
  - We are 95% confident that the true ______________ in the water temperature is between _____________ degrees.

FACTORS LEADING TO SMALLER P-VALUES

- **Question:** How can we get a smaller p-value in the difference of two means test?

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - \mu_d}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- **Answer:**
  - _________ difference between sample means
  - _________ sample standard deviations
  - _________ sample sizes
  - Use a _______ **two-sample t-test** if appropriate