Continuous data – 2 group estimation, testing, power and sample size

Fundamental Rule of Data Analysis

Different types of data require different statistical analyses.

One-group vs. Two-group

- We've discussed estimation and testing for a continuous variable in one group
  - Compare the mean of one population to a known value

- Often we're more interested in comparing a continuous variable in two groups
  - Compare the mean of one population to the mean of another population

Two Groups of Interest

Paired samples

- Each observation in the first group has a corresponding observation in the second group
- Oftentimes the two observations are within a single subject
- Examples:
  - Test scores before vs. after a course
  - Balance on leg with torn ACL vs. leg without torn ACL

Independent samples

- Observations in the two groups are not related to each other
- Cannot match up or pair the observations together in any way
- Examples:
  - LDL cholesterol level in the United States vs. Australia
  - Life expectancy in men vs. women

Two Groups of Interest

Paired samples

- Each observation in the first group has a corresponding observation in the second group
- Oftentimes the two observations are within a single subject
- Examples:
  - Test scores before vs. after a course
  - Balance on leg with torn ACL vs. leg without torn ACL

Independent samples

- Observations in the two groups are not related to each other
- Cannot match up or pair the observations together in any way
- Examples:
  - LDL cholesterol level in the United States vs. Australia
  - Life expectancy in men vs. women

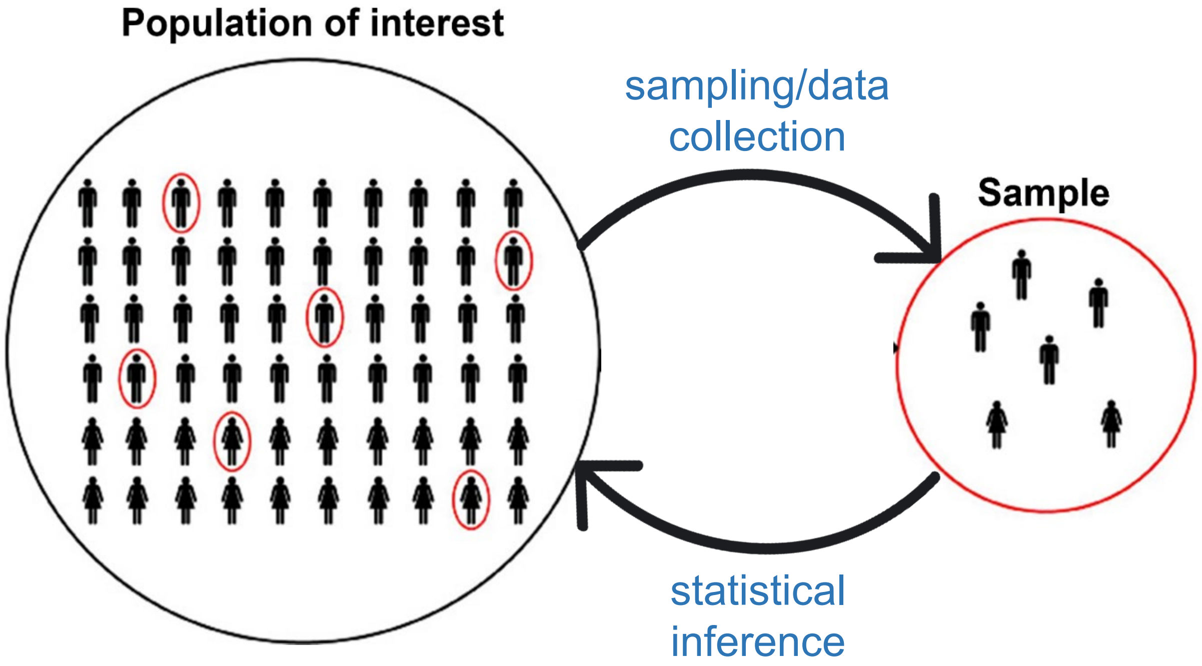
Paired Data

- Make sure the data are "paired up"
- Calculate the differences
- Treat the differences like a single variable/group
  - What value for the mean would indicate no difference between the groups?
- Use everything you already know about one-group estimation and testing

Student	Pre-module score	Post-module score	Difference
1	18	22	+4
2	21	25	+4
3	16	17	+1
4	22	24	+2
5	19	16	-3
6	24	29	+5
7	17	20	+3
8	21	23	+2
9	23	19	-4
10	18	20	+2
11	14	15	+1
12	16	15	-1
13	16	18	+2
14	19	26	+7
15	18	18	0
16	20	24	+4
17	12	18	+6
18	22	25	+3
19	15	19	+4
20	17	16	-1

Paired Data: Notation

- Δ = mean of differences in population
- σ_d = standard deviation of differences in population
- $\bar{d}$ = mean of differences in sample
- s_d = standard deviation of differences in sample



Paired Data: Estimation

- The best estimate of Δ (population mean of differences) is $\bar{d}$ (sample mean of differences)
- Can make a confidence interval for Δ just like we did with one group
  - Centered at $\bar{d}$
  - Width depends on confidence level, sample size, and s_d (standard deviation **of the differences**)

Paired Example: Test Scores

- Suppose that a professor is interested in the efficacy of a learning module. She has 20 students take a pre-test, then complete the learning module, then take a post-test. Provide an estimate and 95% confidence interval for the average change in test scores after completing the learning module.

Student	Pre-module score	Post-module score	Difference
1	18	22	+4
2	21	25	+4
3	16	17	+1
4	22	24	+2
5	19	16	-3
6	24	29	+5
7	17	20	+3
8	21	23	+2
9	23	19	-4
10	18	20	+2
11	14	15	+1
12	16	15	-1
13	16	18	+2
14	19	26	+7
15	18	18	0
16	20	24	+4
17	12	18	+6
18	22	25	+3
19	15	19	+4
20	17	16	-1

n=20 or n=40?

Let $\Delta =$

How would we calculate $\bar{d}$ and s_d?

Paired Example: Test Scores

Sample summary statistics for differences:

```
mean        sd  IQR 0%  25% 50% 75% 100%  n
2.05 2.837252 3.25 -4 0.75   2   4    7 20
```

95% confidence interval for differences:

```
95 percent confidence interval:
 0.7221251 3.3778749
sample estimates:
mean of the differences
                   2.05
```

- Best estimate of average change in test scores:
  $\bar{d} = 2.05$ points

- CI interpretation: “We are 95% confident that the average change in test scores from before the module to after the module is between 0.72 and 3.38 points.”

Paired Data: Testing

- Testing whether there's a difference in the two groups when the data are paired is the same as testing whether the difference equals 0
  - Called a paired t-test
  - Exactly the same as one-group t-test using the differences

$$H_0: \Delta = 0$$
$$H_A: \Delta \neq 0$$

(Or one-sided H_A if interest is only in a difference in one direction)

- Power and sample size calculations are the same as we did with one group
  - Remember: variable of interest is the **difference**
  - Standard deviation is SD of the **differences**
  - Effect size is the expected (average) **difference** between pairs
  - Sample size is number of **pairs of observations**

Paired Example: Test Scores

- Now perform a hypothesis test to test whether there is a significant change in test scores after taking the module. Use α=0.05.

Student	Pre-module score	Post-module score	Difference
1	18	22	+4
2	21	25	+4
3	16	17	+1
4	22	24	+2
5	19	16	-3
6	24	29	+5
7	17	20	+3
8	21	23	+2
9	23	19	-4
10	18	20	+2
11	14	15	+1
12	16	15	-1
13	16	18	+2
14	19	26	+7
15	18	18	0
16	20	24	+4
17	12	18	+6
18	22	25	+3
19	15	19	+4
20	17	16	-1

Paired Example: Test Scores

$H_0\text{: } \Delta = 0 \qquad H_A\text{: } \Delta \neq 0 \qquad \alpha = 0.05$

Paired t-test for change in test scores:

```
        Paired t-test

data:  postmodule and premodule
t = 3.2313, df = 19, p-value = 0.004395
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.7221251 3.3778749
sample estimates:
mean of the differences
                   2.05
```

$\bar{d} = 2.05$

p-value $= 0.004$

Since the p-value is less than 0.05, we reject H_0 and conclude that there is sufficient evidence to say that there is a difference in test scores after completing the learning module. In fact, the data suggest that test score increase.

Paired Example: Test Scores

- How many subjects would we need to have 80% power to detect a change in test scores of 1 point? Use $\alpha = 0.05$, and recall that we previously calculated that the standard deviation for the difference in test scores is 2.84 points.

Power =

Effect size =

Standard deviation =

Significance level =

Paired Example: Test Scores

- How many subjects would we need to have 80% power to detect a change in test scores of 1 point? Use $\alpha = 0.05$, and recall that we previously calculated that the standard deviation for the difference in test scores is 2.84 points.

```
     Paired t test power calculation

              n = 65.25272
          delta = 1
             sd = 2.84
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number of *pairs*, sd is std.dev. of *differences* within pairs
```

Two Groups of Interest

Paired samples

- Each observation in the first group has a corresponding observation in the second group
- Oftentimes the two observations are within a single subject
- Examples:
  - Test scores before vs. after a course
  - Balance on leg with torn ACL vs. leg without torn ACL

Independent samples

- Observations in the two groups are not related to each other
- Cannot match up or pair the observations together in any way
- Examples:
  - LDL cholesterol level in the United States vs. Australia
  - Life expectancy in men vs. women

Two Groups of Interest

Paired samples

- Each observation in the first group has a corresponding observation in the second group
- Oftentimes the two observations are within a single subject
- Examples:
  - Test scores before vs. after a course
  - Balance on leg with torn ACL vs. leg without torn ACL

Independent samples

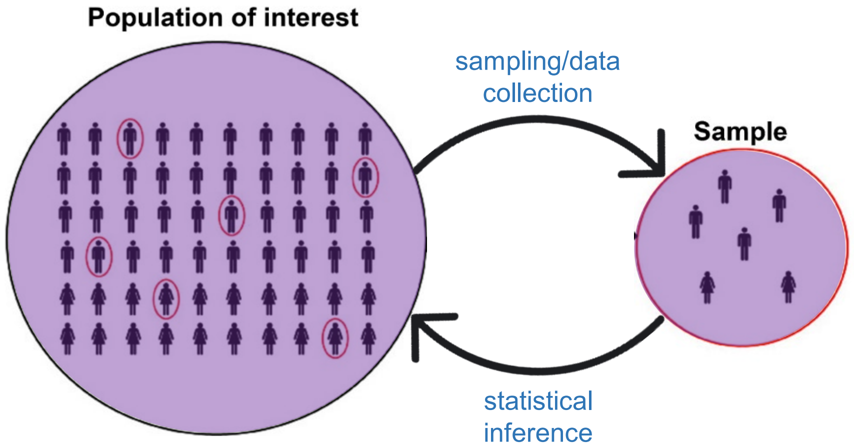
- Observations in the two groups are not related to each other
- Cannot match up or pair the observations together in any way
- Examples:
  - LDL cholesterol level in the United States vs. Australia
  - Life expectancy in men vs. women

Two Independent Groups

- Observations in the two groups are independent (not related to each other)
- Ask yourself if you would be able to pair or match up the observations in one group to the observations in the other group
  - If not, groups are independent

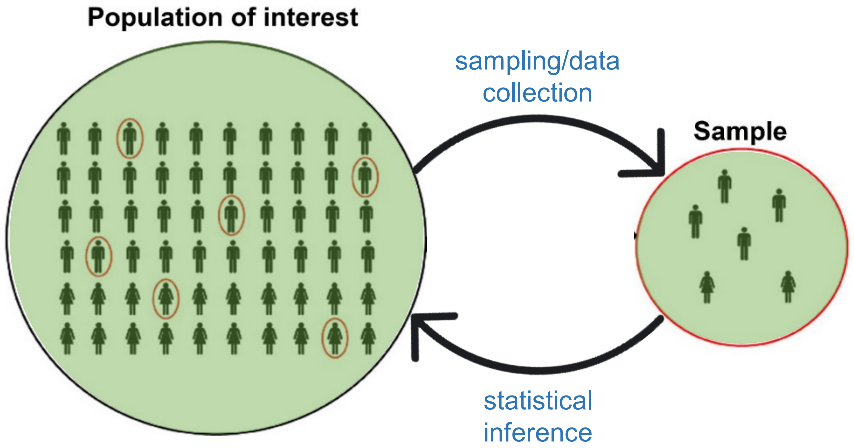
Two Independent Groups: Notation

- μ_1 = population mean in group 1
- σ_1 = population SD in group 1



- $\bar{x}_1$ = sample mean in group 1
- s_1 = sample SD in group 1

- μ_2 = population mean in group 2
- σ_2 = population SD in group 2



- $\bar{x}_2$ = sample mean in group 2
- s_2 = sample SD in group 2

Two Independent Groups: Estimation

- Interest is in the difference between group means
- Best estimate of difference between means in each population ($\mu_1 - \mu_2$) is difference between means in each sample ($\bar{x}_1 - \bar{x}_2$)
- Can make a confidence interval for $\mu_1 - \mu_2$
  - Centered at $\bar{x}_1 - \bar{x}_2$
  - Width depends on confidence level, sample size in each group, and s_1 and s_2 (sample standard deviations in each group)

Low Birth Weight Data

- Information on 100 low birth weight infants born in two teaching hospitals in Boston, Massachusetts

Variable	Description
sex	Sex of the baby (Male, Female)
gestage	Gestational age at time of birth (weeks)
length	Length of the baby (cm)
birthwt	Birth weight of the baby (g)
apgar	Apgar score (integers, min=0, max=10). This is a scoring system used for assessing the clinical status of a newborn. 7 or higher is generally considered normal, 4-6 is low, and 3 or below is critically low.
apgarlow	Categorizes the Apgar score into two categories: Normal (7-10) or Low (0-6)

Find the dataset (lowbwt.xlsx) and the full data dictionary (lowbwt Data Dictionary.pdf) in the Data Module on the Canvas site

Two-Group Example: Birth Wt/Apgar Score

- Researchers are interested in whether the average birth weight among babies with an Apgar score classified as “Low” is different from the average birth weight among babies with an Apgar score classified as “Normal”. Provide an estimate and 95% confidence interval for the difference in birth weight between the groups.

Two-Group Example: Birth Wt/Apgar Score

Sample summary statistics for birth weight in each Apgar score group:

```
              mean        sd
Low       1015.625  292.8949
Normal    1154.333  240.3836
```

95% confidence interval for difference in birth weight between Apgar groups:

```
95 percent confidence interval:
 -249.8317  -27.5850
sample estimates:
   mean in group Low mean in group Normal
            1015.625             1154.333
```

- Best estimate of difference in average birth weight between Apgar groups (Low-Normal):

$$\bar{x}_1 - \bar{x}_2 = 1015.6 - 1154.3$$
$$= -138.7 \text{ grams}$$

- CI interpretation: "We are 95% confident that the difference in average birth weight of babies with a low Apgar score and a normal Apgar score (Low-Normal) is between -249.8 and -27.6 grams."

Two Independent Groups: Testing

- We can test whether the two group means are equal with a two-group t-test (also known as independent samples t-test)

$H_0: \mu_1 = \mu_2$	$H_0: \mu_1 = \mu_2$	$H_0: \mu_1 = \mu_2$
$H_A: \mu_1 \neq \mu_2$	$H_A: \mu_1 < \mu_2$	$H_A: \mu_1 > \mu_2$

- Have to specify whether the population standard deviations are the same or different in the two groups → more on this later
- Power and sample size calculations for two-group t-test are very similar to what we've seen before, except:
  - Effect size is the expected difference between the two group means
  - Use the larger of the standard deviations from the two groups
  - Sample size is the number **in each group** (assuming it's the same)

Two-Group Example: Birth Wt/Apgar Score

- Now perform a hypothesis test to test whether the average birth weight among babies with an Apgar score classified as “Low” is different from the average birth weight among babies with an Apgar score classified as “Normal”. Use a significance level of 0.05.

Two-Group Example: Birth Wt/Apgar Score

H_0: $\mu_1 = \mu_2$ $\quad$ H_A: $\mu_1 \neq \mu_2$ $\quad$ $\alpha = 0.05$

Two-group t-test for difference in average birth weight between Apgar groups:

```
        Welch Two Sample t-test

data:  birthwt by apgarlow
t = -2.4882, df = 72.259, p-value = 0.01515
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -249.8317  -27.5850
sample estimates:
   mean in group Low mean in group Normal
            1015.625             1154.333
```

$\bar{x}_1 = 1015.6$ $\quad$ $\bar{x}_2 = 1154.3$

p-value $= 0.015$

Since the p-value is less than 0.05, we reject H_0 and conclude that there is sufficient evidence to say that the average birth weight among babies with a normal Apgar score is different from the average birth weight among babies with a low Apgar score.

Two-Group Example: Birth Wt/Apgar Score

- You're planning a new study that will recruit an equal number of infants with a low Apgar score and a normal Apgar score. Use the low birth weight dataset as pilot data to calculate how many subjects you would need to recruit in your study to have 80% power to detect a difference in mean birth weight between the two Apgar groups (using $\alpha = 0.05$).

Sample summary statistics for birth weight in each Apgar score group:

	mean	sd
Low	1015.625	292.8949
Normal	1154.333	240.3836

Power = 0.8

Effect size = 1015.625 - 1154.333 = -138.708

Standard deviation = 292.8949

Significance level = 0.05

Two-Group Example: Birth Wt/Apgar Score

- You're planning a new study that will recruit an equal number of infants with a low Apgar score and a normal Apgar score. Use the low birth weight dataset as pilot data to calculate how many subjects you would need to recruit in your study to have 80% power to detect a difference in mean birth weight between the two Apgar groups (using $\alpha = 0.05$).

```
     Two-sample t test power calculation

              n = 70.96684
          delta = 138.708
             sd = 292.8949
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

Equal vs. Unequal Variances

- For the two-group t-test, you have to specify if the two groups have equal variance or not
  - If equal: An estimate of the pooled standard deviation is used for the test
  - If unequal: The standard deviations of the two groups are used separately
- We can perform a separate hypothesis test *for the variances* to decide if the two population variances are equal → Stay tuned!

Important Points

- Difference between paired samples and independent samples
- Set up and interpretation of confidence intervals for difference in paired data
- Set up and interpretation of paired t-test
  - Power/sample size calculations
- Set up and interpretation of confidence intervals for difference in two independent samples
- Set up and interpretation of two-group (independent samples) t-test
  - Power/sample size calculations