T distribution and Inference for μ_{diff} and $\mu_1 - \mu_2$

https://www.wsj.com/articles/parsing-the-gender-pay-gap-1542917969

New Type of Research Question we can Ask: Is there a relationship between

- gender (categorical independent variable-2 levels)
- and income (numerical dependent variable)?

https://www.wsj.com/articles/parsing-the-gender-pay-gap-1542917969

New Type of Research Question we can Ask: Is there a relationship between

- ground water zinc concentration and surface water zinc concentration (two paired/dependent numerical variables)?

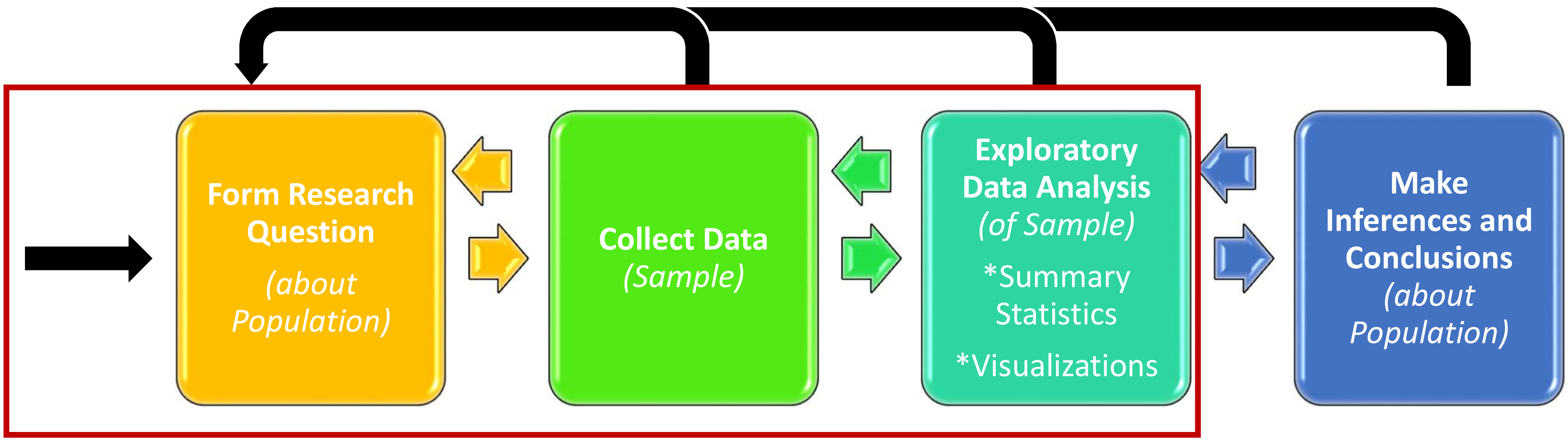
Coming up…

- Lab Assignment 6 is due **Thursday just before your lab section time.**

- Peer Evaluations is due **Thursday 2/28 11:55pm** *(part of your participation grade)*

- Read over Project Stage 1 statement before **Thursday 2/28**
- Project Stage 1 is due **Thursday 3/7**

Project Stage 1 due Thursday 3/7



Making a Confidence Interval for μ with CLT

Independence

1. ✓Random sampling/assignment is used.
2. ✓ Sample size n < 10% of population

✓ One of the available **Sample Size/Skewness "Scenarios"** is met

SCENARIOS	σ is known	σ is not known (have s)
Scenarios: n >30	$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$	$\bar{x} \pm t^*_{n-1} \frac{s}{\sqrt{n}}$
Scenarios: • n ≤ 30 AND • population distribution IS approximately normal.	$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$	$\bar{x} \pm t^*_{n-1} \frac{s}{\sqrt{n}}$ *or not extremely skewed
Scenarios: • n ≤ 30 AND • population distribution IS NOT approximately normal.	X	X

Hypothesis Testing for μ with CLT

Independence

1. ✓Random sampling/assignment is used.
2. ✓ Sample size n < 10% of population

✓ One of the available **Sample Size/Skewness "Scenarios"** is met

SCENARIOS	σ is known	σ is not known (have s)
Scenarios: n >30	Test Stat $Z = \frac{\bar{x} - (null\ value)}{\sigma/\sqrt{n}}$	Test Stat $T_{n-1} = \frac{\bar{x} - (null\ value)}{s/\sqrt{n}}$
Scenarios: • n ≤ 30 AND • population distribution IS approximately normal.	Test Stat $Z = \frac{\bar{x} - (null\ value)}{\sigma/\sqrt{n}}$	Test Stat $T_{n-1} = \frac{\bar{x} - (null\ value)}{s/\sqrt{n}}$ *or not extremely skewed
Scenarios: • n ≤ 30 AND • population distribution IS NOT approximately normal.	X	X

Unit 4

Confidence Intervals for μ_{diff} with CLT

Independence

1. ✓Random sampling/assignment is used to collect the pairs of data
2. ✓ Sample size n < 10% of population of pairs.

✓ One of the available **Sample Size/Skewness "Scenarios"** is met

SCENARIOS	σ_{diff} is known	σ_{diff} is not known (have s_{diff})
Scenarios: n >30 n=# of pairs	$\bar{x}_{diff} \pm z^* \frac{\sigma_{diff}}{\sqrt{n}}$	$\bar{x}_{diff} \pm t^*_{n-1} \frac{s_{diff}}{\sqrt{n}}$
Scenarios: n=# of pairs • n ≤ 30 AND • population distribution of differences IS approximately normal.	$\bar{x}_{diff} \pm z^* \frac{\sigma_{diff}}{\sqrt{n}}$	$\bar{x}_{diff} \pm t^*_{n-1} \frac{s_{diff}}{\sqrt{n}}$ *or not extremely skewed
Scenarios: n=# of pairs • n ≤ 30 AND • population distribution of differences IS NOT approximately normal.	X	X

Unit 4

Hypothesis Testing for μ_{diff} with CLT

Independence

1. ✓Random sampling/assignment is used to collect the pairs of data
2. ✓ Sample size n < 10% of population of pairs.

✓ One of the available **Sample Size/Skewness "Scenarios"** is met

SCENARIOS		σ_{diff} is known	σ_{diff} is not known (have s_{diff})
Scenarios: n >30	n=# of pairs	Test Stat $Z = \frac{\overline{x}_{diff} - (null\ value)}{\sigma_{diff}/\sqrt{n}}$	Test Stat $T_{n-1} = \frac{\overline{x}_{diff} - (null\ value)}{s_{diff}/\sqrt{n}}$
Scenarios: • n ≤ 30 AND • population distribution of differences IS approximately normal.	n=# of pairs	Test Stat $Z = \frac{\overline{x}_{diff} - (null\ value)}{\sigma_{diff}/\sqrt{n}}$	Test Stat $T_{n-1} = \frac{\bar{x}_{diff} - (null\ value)}{s_{diff}/\sqrt{n}}$ *or not extremely skewed
Scenarios: • n ≤ 30 AND • population distribution of differences IS NOT approximately normal.	n=# of pairs	X	X

Making a Confidence Interval for $\mu_1 - \mu_2$ with CLT

Independence between Groups

✓both populations are independent.

Independence within Groups

1. ✓Random sampling/assignment is used <u>for both samples</u>
2. ✓ Sample size n1 < 10% of <u>population 1</u> and sample size n2 < 10% of <u>population 2.</u>

✓ One of the available **Sample Size/Skewness "Scenarios"** is met

SCENARIOS	σ1 and σ2 known	σ1 and σ2 not known
<u>Condition</u>: n1 >30 AND n2 >30	$(\bar{x}_1 - \bar{x}_2) \pm z^* \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$(\bar{x}_1 - \bar{x}_2) \pm t^*_{\min(n1-1,n2-1)} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
<u>Condition</u>: n1 ≤ 30 OR n2 ≤ 30 and population distributions are approximately normal.	$(\bar{x}_1 - \bar{x}_2) \pm z^* \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$(\bar{x}_1 - \bar{x}_2) \pm t^*_{\min(n1-1,n2-1)} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ *Or not extremely skewed (if using t-distribution)
<u>Condition</u>: n1 ≤ 30 OR n2 ≤ 30 and population distributions are NOT approximately normal.	X	X

Making a Hypothesis Test for $\mu_1 - \mu_2$ with CLT

Independence between Groups

✓both populations are independent.

Independence within Groups

1. ✓Random sampling/assignment is used <u>for both samples</u>
2. ✓ Sample size n1 < 10% of <u>population 1</u> and sample size n2 < 10% of <u>population 2.</u>

✓ One of the available **Sample Size/Skewness "Scenarios"** is met

SCENARIOS	σ1 and σ2 known	σ1 and σ2 not known
<u>Condition</u>: $n_1 > 30$ AND $n_2 > 30$	Test Stat $Z = \frac{(\overline{x}_1 - \overline{x}_2) - (null\ value)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	Test Stat $t_{\min(n1-1, n2-1)} = \frac{(\overline{x}_1 - \overline{x}_2) - (null\ value)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
<u>Condition</u>: n1 ≤ 30 OR n2 ≤ 30 and population distributions are approximately normal.	Test Stat $Z = \frac{(\overline{x}_1 - \overline{x}_2) - (null\ value)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	Test Stat $t_{\min(n1-1, n2-1)} = \frac{(\overline{x}_1 - \overline{x}_2) - (null\ value)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ *Or not extremely skewed
<u>Condition</u>: n1 ≤ 30 OR n2 ≤ 30 and population distributions are NOT approximately normal.	X	X