Lecture 06 - Two Sample Hypothesis Testing

Samuel E DeWitt, PhD

Organization of Lecture

1) Logic of two sample inference
2) Sampling distribution of mean differences
3) Independent samples hypothesis testing
4) Non-independent sample hypothesis testing

Logic of Two Sample Inference

- Research Question
  - Are these two samples drawn from a single population with mean μ, or are they subsets of two different populations with means μ_1 and μ_2, respectively?

Two Possible Answers

- Yes, the samples are drawn from the same population.
  - The observed difference between $\overline{x}_1$ and $\overline{x}_2$ reflects a **random** difference.
- No, the samples are drawn from different populations.
  - The observed difference between $\overline{x}_1$ and $\overline{x}_2$ reflects a **systematic** difference.

Sampling Distribution of Mean Differences

- Probability theory
  - How likely is it that we would observe the difference between the two sample means if the population means are actually equal?
  - Is the difference random or systematic?

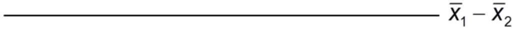
- New probability distribution
  - **Sampling distribution of mean differences** (not a sampling distribution of means)

Thought Experiment

Random samples of size n_1 and n_2, compute mean difference: $\bar{x}_1 - \bar{x}_2$

Compute mean differences for an infinite number of independent samples...

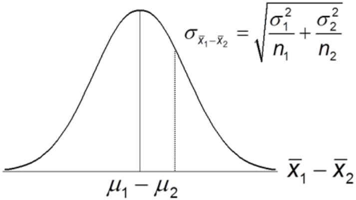
What will happen?



Sampling Distribution of Mean Differences

Distribution of all possible mean differences between two independent samples

- Defined by two parameters:
  - Mean $= \mu_1 - \mu_2$
  - Standard error $= \sigma_{\overline{x}_1 - \overline{x}_2}$



Central Limit Theorem Redux

With large samples, the sampling distribution of mean differences is approximately normal, even if the characteristic is not normally distributed in the populations from which the samples are drawn.

We can conduct hypothesis testing since we know that the probability distribution is normal.

Standard Score for Mean Difference

- General form for a standard score:

$$z = \frac{\text{estimate - parameter}}{\text{standard deviation or error}}$$

- Standard score for mean difference with independent samples:

$$z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\overline{x}_1 - \overline{x}_2}} = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Hypothesis Testing

- Test statistic when σ_1 and σ_2 are **known**:

$$TS = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- Test statistic when σ_1 and σ_2 are **unknown**:

$$TS = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}}$$

Social Disorganization & Violent Victimization

- Low-SES Neighborhoods
  - $n_L = 41$, $\overline{x}_L = 12.4$, $\sigma_L^2 = 18.49$

- High-SES neighborhoods
  - $n_H = 21$, $\overline{x}_H = 5.2$, $\sigma_H^2 = 2.25$

- Research Question
  - Do people that live in low-SES neighborhoods have a different risk of victimization than people that live in high-SES neighborhoods?

Social Disorganization & Violent Victimization (cont)

- **Step 1 - Formally state hypotheses:**

Social Disorganization & Violent Victimization (cont)

- **Step 1 - Formally state hypotheses:**
  - $\text{H}_1 : \mu_L \neq \mu_H \therefore \mu_L - \mu_H \neq 0$; $\text{H}_0 : \mu_L = \mu_H \therefore \mu_L - \mu_H = 0$

- **Step 2 - Obtain a probability distribution:**

Social Disorganization & Violent Victimization (cont)

- **Step 1 - Formally state hypotheses**:
  - $H_1 : \mu_L \neq \mu_H \therefore \mu_L - \mu_H \neq 0$; $H_0 : \mu_L = \mu_H \therefore \mu_L - \mu_H = 0$
- **Step 2 - Obtain a probability distribution**:
  - z-distribution
- **Step 3 - Make Decision Rules**:
  - α=.01 (two-tailed), $z_{crit} = 2.576$, reject H_0 if $|TS| > 2.576$

Social Disorganization & Violent Victimization (cont)

- **Step 4 - Calculate the test statistic**:

$$z = \frac{\overline{\mathrm{x}}_1 - \overline{\mathrm{x}}_2}{\sqrt{\frac{\sigma_1^2}{\mathrm{n}_1} + \frac{\sigma_2^2}{\mathrm{n}_2}}} = \frac{12.4 - 5.2}{\sqrt{\frac{18.49}{41} + \frac{2.25}{21}}} =$$

Social Disorganization & Violent Victimization (cont)

- **Step 4 - Calculate the test statistic:**

$$z = \frac{\overline{\mathrm{x}}_1 - \overline{\mathrm{x}}_2}{\sqrt{\frac{\sigma_1^2}{\mathrm{n}_1} + \frac{\sigma_2^2}{\mathrm{n}_2}}} = \frac{12.4 - 5.2}{\sqrt{\frac{18.49}{41} + \frac{2.25}{21}}} = 9.638$$

- **Step 5 - Make a decision about the null hypothesis:**

Social Disorganization & Violent Victimization (cont)

- **Step 4 - Calculate the test statistic**:

$$z = \frac{\overline{\mathrm{x}}_1 - \overline{\mathrm{x}}_2}{\sqrt{\frac{\sigma_1^2}{\mathrm{n}_1} + \frac{\sigma_2^2}{\mathrm{n}_2}}} = \frac{12.4 - 5.2}{\sqrt{\frac{18.49}{41} + \frac{2.25}{21}}} = 9.638$$

- **Step 5 - Make a decision about the null hypothesis**:
  - Reject H_0, low-SES and high-SES neighborhoods have significantly different rates of victimization.

Youth Employment & Problem Behavior

- Worker sample
  - $n_1 = 2243$, $\overline{x}_1 = 3.73$, $s_1^2 = 12.95$
- Non-worker sample
  - $n_2 = 6679$, $\overline{x}_2 = 2.22$, $s_2^2 = 8.19$
- Research Question
  - Do working youths engage in more problem behavior than non-working youths?

Youth Employment & Problem Behavior - Hypothesis Test

- **Step 1 - Formally state hypotheses:**

Youth Employment & Problem Behavior - Hypothesis Test

- **Step 1 - Formally state hypotheses:**
  - $\mathrm{H}_1 : \mu_W > \mu_N \therefore \mu_W - \mu_N > 0$; $\mathrm{H}_0 : \mu_W \leq \mu_N \therefore \mu_W - \mu_N \leq 0$

- **Step 2 - Obtain a probability distribution:**

Youth Employment & Problem Behavior - Hypothesis Test

- **Step 1 - Formally state hypotheses**:
  - $H_1 : \mu_W > \mu_N \therefore \mu_W - \mu_N > 0$; $H_0 : \mu_W \leq \mu_N \therefore \mu_W - \mu_N \leq 0$
- **Step 2 - Obtain a probability distribution**:
  - t-distribution, $df = 2243 + 6679 - 2$ =8920
- **Step 3 - Make decision rules**:
  - α=.05 (one-tailed), $t_{crit} = 1.645$, reject H_0 if TS>1.645.

Youth Employment & Problem Behavior (cont)

Step 4 - Calculate the test statistic:

$$t = \frac{(\bar{\mathrm{x}}_1 - \bar{\mathrm{x}}_2)}{\sqrt{\frac{s_1^2}{\mathrm{n}_1 - 1} + \frac{s_2^2}{\mathrm{n}_2 - 1}}} = \frac{3.73 - 2.22}{\sqrt{\frac{12.95}{2243-1} + \frac{8.19}{6679-1}}} =$$

Youth Employment & Problem Behavior (cont)

Step 4 - Calculate the test statistic:

$$t = \frac{(\overline{\mathrm{x}}_1 - \overline{\mathrm{x}}_2)}{\sqrt{\frac{s_1^2}{\mathrm{n}_1 - 1} + \frac{s_2^2}{\mathrm{n}_2 - 1}}} = \frac{3.73 - 2.22}{\sqrt{\frac{12.95}{2243-1} + \frac{8.19}{6679-1}}} = 18.048$$

- **Step 5 - Make a decision about the null hypothesis:**

Youth Employment & Problem Behavior (cont)

Step 4 - Calculate the test statistic:

$$t = \frac{(\overline{\mathrm{x}}_1 - \overline{\mathrm{x}}_2)}{\sqrt{\frac{s_1^2}{\mathrm{n}_1 - 1} + \frac{s_2^2}{\mathrm{n}_2 - 1}}} = \frac{3.73 - 2.22}{\sqrt{\frac{12.95}{2243-1} + \frac{8.19}{6679-1}}} = 18.048$$

- **Step 5 - Make a decision about the null hypothesis**:
  - Reject H_0, working youths commit significantly more crimes than non-working youths.

Criminal Lifestyle and Victimization

Victimization among individuals who do and do not report being criminally active.

- Criminally-active sample
  - $n_C = 10$, $\overline{x}_C = 5.9$, $s_C^2 = 5.29$
- Non-criminally-active sample
  - $n_N = 10$, $\overline{x}_N = 2.8$, $s_N^2 = 5.36$
- Research Question
  - Does a criminal lifestyle increase the likelihood of victimization?

Criminal Lifestyle and Victimization - Hypothesis Test

- **Step 1 - Formally state hypotheses:**

Criminal Lifestyle and Victimization - Hypothesis Test

- **Step 1 - Formally state hypotheses:**
  - $\text{H}_1 : \mu_C > \mu_N \therefore \mu_C - \mu_N > 0$; $\text{H}_0 : \mu_C \leq \mu_N \therefore \mu_C - \mu_N \leq 0$

- **Step 2 - Obtain a probability distribution:**

Criminal Lifestyle and Victimization - Hypothesis Test

- **Step 1 - Formally state hypotheses**:
  - $H_1 : \mu_C > \mu_N \therefore \mu_C - \mu_N > 0$; $H_0 : \mu_C \leq \mu_N \therefore \mu_C - \mu_N \leq 0$
- **Step 2 - Obtain a probability distribution**:
  - t-distribution; df$=10 + 10 - 2 = 18$
- **Step 3 - Make decision rules**:
  - α=.10 (one-tailed), $t_{crit} = 1.33$, reject H_0 if TS>1.33.

Criminal Lifestyle and Victimization (cont)

- **Step 4 - Calculate the test statistic:**

$$t = \frac{(\overline{\mathrm{x}}_1 - \overline{\mathrm{x}}_2)}{\sqrt{\frac{s_1^2}{\mathrm{n}_1 - 1} + \frac{s_2^2}{\mathrm{n}_2 - 1}}} = \frac{5.9 - 2.8}{\sqrt{\frac{5.29}{10-1} + \frac{5.36}{10-1}}} =$$

Criminal Lifestyle and Victimization (cont)

- **Step 4 - Calculate the test statistic:**

$$t = \frac{(\overline{\mathrm{x}}_1 - \overline{\mathrm{x}}_2)}{\sqrt{\frac{s_1^2}{\mathrm{n}_1 - 1} + \frac{s_2^2}{\mathrm{n}_2 - 1}}} = \frac{5.9 - 2.8}{\sqrt{\frac{5.29}{10-1} + \frac{5.36}{10-1}}} = 2.85$$

- **Step 5 - Make a decision about the null hypothesis:**

Criminal Lifestyle and Victimization (cont)

- **Step 4 - Calculate the test statistic**:

$$t = \frac{(\overline{x}_1 - \overline{x}_2)}{\sqrt{\frac{s_1^2}{n_1-1} + \frac{s_2^2}{n_2-1}}} = \frac{5.9 - 2.8}{\sqrt{\frac{5.29}{10-1} + \frac{5.36}{10-1}}} = 2.85$$

- **Step 5 - Make a decision about the null hypothesis**: Reject H_0, people who engage in crime at a high rate are victimized significantly more often.

Another Example - Dating and Delinquency

- Daters
  - $n_D = 8$, $\overline{x}_D = 2.38$, $s_D^2 = 2.46$
- Non-daters
  - $n_N = 6$, $\overline{x}_N = 1.67$, $s_N^2 = 1.21$
- Research Question
  - Is there an association between dating and delinquency?

Dating and Delinquency - Hypothesis Test

- **Step 1 - Formally state hypotheses:**

Dating and Delinquency - Hypothesis Test

- **Step 1 - Formally state hypotheses:**
  - $\mathrm{H}_1 : \mu_N \neq \mu_D \therefore \mu_N - \mu_D \neq 0$; $\mathrm{H}_0 : \mu_N = \mu_D \therefore \mu_N - \mu_D = 0$

- **Step 2 - Obtain a probability distribution:**

Dating and Delinquency - Hypothesis Test

- **Step 1 - Formally state hypotheses**:
  - $H_1 : \mu_N \neq \mu_D \therefore \mu_N - \mu_D \neq 0$; $H_0 : \mu_N = \mu_D \therefore \mu_N - \mu_D = 0$

- **Step 2 - Obtain a probability distribution**:
  - t-distribution; df$=6 + 8 - 2 = 12$

- **Step 3 - Make decision rules**:
  - α=.01 (two-tailed), $t_{crit} = 3.055$, reject H_0 if $|TS|>3.055$.

Dating and Delinquency (cont)

- **Step 4 - Calculate the test statistic**:

$$t = \frac{(\overline{\mathrm{x}}_1 - \overline{\mathrm{x}}_2)}{\sqrt{\frac{s_1^2}{\mathrm{n}_1 - 1} + \frac{s_2^2}{\mathrm{n}_2 - 1}}} = \frac{2.38 - 1.67}{\sqrt{\frac{2.46}{8-1} + \frac{1.21}{6-1}}} = 0.922$$

- **Step 5 - Make a decision about the null hypothesis**:
  - Accept H_0, youth who date do not engage in a different amount of problem behaviors than youth who do not date.

Non-Independent Samples Hypothesis Testing

- Assumption of independent samples hypothesis test
  - The two samples must be independent of each other (i.e., both samples are randomly selected) and the fact that one case is selected into one sample does not affect the selection of a different case into the other sample

Non-Independent Samples Hypothesis Testing

- We **intentionally** violate this assumption in the following designs:
  - *Before-and-after* design: dependent variable measured for each case at two points in time, with introduction of some treatment in between
  - *Matched samples* design: individuals in treatment and control conditions matched on important characteristics (e.g., age, gender, race)

Test Statistics for Independent Samples Tests

- Test statistic for independent samples
  - Based on the difference between **two sample means**:

$$\frac{\sum x_1}{n_1} - \frac{\sum x_2}{n_2} = \overline{x}_1 - \overline{x}_2 \Rightarrow \frac{(\overline{\mathrm{x}}_1 - \overline{\mathrm{x}}_2)}{\sqrt{\frac{s_1^2}{\mathrm{n}_1 - 1} + \frac{s_2^2}{\mathrm{n}_2 - 1}}}$$

- Test statistic for non-independent samples
  - Based on the difference between the **scores for each pair of observations**:

$$\frac{\sum (x_2 - x_1)}{n} = \overline{x}_D \Rightarrow \frac{\overline{x}_D}{\sqrt{\frac{s_D^2}{n-1}}}$$

Non-Independent - Computing the Mean & Standard Deviation

- As in the prior slide, the mean is computed slightly differently for a non-independent samples hypothesis test
  - We now want the average **difference** across matched pairs or between before-and-after values.

$$\frac{\sum(x_1 - x_2)}{n}$$

Non-Independent - Computing the Mean & Standard Deviation

- The computation for the standard deviation is also slightly different, becoming the standard deviation of the differences between matched pairs or before-and-after values.

$$\sqrt{\frac{\sum(x_D - \overline{x}_D)^2}{n}}$$

I'll show you the procedure for obtaining both later in this lecture. Basically, this amounts to a slight change in notation, not in the underlying procedure to obtain these sample statistics.

An Example - Does Crime Run in Families?

- Youths with criminal parents matched with youths with non-criminal parents
  - $n_{Pairs} = 121$, $\overline{x}_D = 5.75$, $s_D = 4.25$
- Research Question
  - Are youths with criminal parents more likely to be arrested than youths without criminal parents?

Does Crime Run in Families? - Hypothesis Test

- **Step 1 - Formally state hypotheses:**

Does Crime Run in Families? - Hypothesis Test

- **Step 1 - Formally state hypotheses:**
  - $H_1 : \mu_D > 0$; $H_0 : \mu_D \leq 0$
- **Step 2 - Obtain probability distribution:**

Does Crime Run in Families? - Hypothesis Test

- **Step 1 - Formally state hypotheses:**
  - $\text{H}_1 : \mu_D > 0$; $\text{H}_0 : \mu_D \leq 0$
- **Step 2 - Obtain probability distribution:**
  - t-distribution; df$=121 - 1 = 120$
- **Step 3 - Make decision rules:**

Does Crime Run in Families? - Hypothesis Test

- **Step 1 - Formally state hypotheses**:
  - $H_1 : \mu_D > 0$; $H_0 : \mu_D \leq 0$
- **Step 2 - Obtain probability distribution**:
  - t-distribution; df$=121 - 1 = 120$
- **Step 3 - Make decision rules**:
  - α=.01 (one-tailed), $t_{crit} = 2.358$, reject H_0 if TS>2.358.

Does Crime Run in Families? (cont)

- **Step 4 - Calculate the test statistic**:

$$t = \frac{\bar{\mathrm{x}}_D - \mu_D}{\mathrm{s}_D/\sqrt{\mathrm{n}-1}} = \frac{5.75}{4.25/\sqrt{121-1}} =$$

Does Crime Run in Families? (cont)

- **Step 4 - Calculate the test statistic:**

$$t = \frac{\overline{x}_D - \mu_D}{s_D/\sqrt{n-1}} = \frac{5.75}{4.25/\sqrt{121-1}} = 14.821$$

- **Step 5 - Make a decision about the null hypothesis:**

Does Crime Run in Families? (cont)

- **Step 4 - Calculate the test statistic**:

$$t = \frac{\bar{x}_D - \mu_D}{s_D/\sqrt{n-1}} = \frac{5.75}{4.25/\sqrt{121-1}} = 14.821$$

- **Step 5 - Make a decision about the null hypothesis**:
  - Reject H_0, having criminal parents significantly increases the likelihood of being arrested.

Another Example - Do Formal Sanctions Deter Crime?

x_1	x_2	$x_D = x_2 - x_1$	$x_D - \bar{x}_D$	$(x_D - \bar{x}_D)^2$
3	1	-2	-1.6	2.56
2	2	0	0.4	0.16
1	0	-1	-0.6	0.36
1	1	0	0.4	0.16
2	1	-1	-0.6	0.36
4	4	0	0.4	0.16
6	6	0	0.4	0.16
1	0	-1	-0.6	0.36
1	1	0	0.4	0.16
2	3	1	1.4	1.96
$\bar{x}_1 = 2.2$	$\bar{x}_2 = 1.8$	$\sum x_D = -4$	$\sum x_D - \bar{x}_D = 0$	$\sum(x_D - \bar{x}_D)^2 = 6.4$
		$\bar{x}_D = -0.4$		$s_D = \sqrt{\frac{6.4}{10}} = 0.8$

Do Formal Sanctions Deter Crime? - Hypothesis Test

- **Step 1 - Formally state hypotheses:**

Do Formal Sanctions Deter Crime? - Hypothesis Test

- **Step 1 - Formally state hypotheses:**
  - $H_1 : \mu_D < 0$; $H_0 : \mu_D \geq 0$
- **Step 2 - Obtain a probability distribution:**

Do Formal Sanctions Deter Crime? - Hypothesis Test

- **Step 1 - Formally state hypotheses:**
  - $\text{H}_1 : \mu_D < 0$; $\text{H}_0 : \mu_D \geq 0$
- **Step 2 - Obtain a probability distribution:**
  - t-distribution; df$=10 - 1 = 9$
- **Step 3 - Make decision rules:**

Do Formal Sanctions Deter Crime? - Hypothesis Test

- **Step 1 - Formally state hypotheses**:
  - $H_1 : \mu_D < 0$; $H_0 : \mu_D \geq 0$
- **Step 2 - Obtain a probability distribution**:
  - t-distribution; df$=10 - 1 = 9$
- **Step 3 - Make decision rules**:
  - α=.05 (one-tailed), $t_{crit} = -1.833$, reject H_0 if TS<-1.833.

Do Formal Sanctions Deter Crime? (cont)

- **Step 4 - Calculate the test statistic**:

Do Formal Sanctions Deter Crime? (cont)

- **Step 4 - Calculate the test statistic**:

$$t = \frac{\bar{\mathrm{x}}_D - \mu_D}{\mathrm{s}_D/\sqrt{\mathrm{n}-1}} = \frac{-0.40}{.80/\sqrt{10-1}} =$$

Do Formal Sanctions Deter Crime? (cont)

- **Step 4 - Calculate the test statistic:**

$$t = \frac{\bar{\mathrm{x}}_D - \mu_D}{\mathrm{s}_D/\sqrt{\mathrm{n}-1}} = \frac{-0.40}{.80/\sqrt{10-1}} = -1.500$$

- **Step 5 - Make a decision about the null hypothesis:**

Do Formal Sanctions Deter Crime? (cont)

- **Step 4 - Calculate the test statistic**:

$$t = \frac{\bar{\mathrm{x}}_D - \mu_D}{\mathrm{s}_D/\sqrt{\mathrm{n}-1}} = \frac{-0.40}{.80/\sqrt{10-1}} = -1.500$$

- **Step 5 - Make a decision about the null hypothesis**:
  - Accept H_0, formal adjudication does not appear to deter crime.

Another Example - After-School Tutoring

Before-and-after design for five high-risk middle-school students with a history of poor academic performance. Did the program improve test scores?

x_1	x_2	$x_D = x_2 - x_1$	$x_D - \bar{x}_D$	$(x_D - \bar{x}_D)^2$
60	62	2	-0.8	0.64
66	66	0	-2.8	7.84
65	67	2	-0.8	0.64
60	64	4	1.2	1.44
59	65	6	3.2	10.24
$\overline{x}_1 = 62$	$\overline{x}_1 = 64.8$	$\sum x_D = 14$	$\sum x_D - \bar{x}_D = 0$	$\sum (x_D - \bar{x}_D)^2 = 20.8$
		$\overline{x}_D = 2.8$		$s_D = \sqrt{\frac{60}{5} - 2.8^2} = 2.04$

After-School Tutoring - Hypothesis Test

- **Step 1 - Formally state hypotheses:**

After-School Tutoring - Hypothesis Test

- **Step 1 - Formally state hypotheses:**
  - $H_1 : \mu_D > 0$; $H_0 : \mu_D \leq 0$
- **Step 2 - Obtain a probability distribution:**

After-School Tutoring - Hypothesis Test

- **Step 1 - Formally state hypotheses:**
  - $H_1 : \mu_D > 0$; $H_0 : \mu_D \leq 0$
- **Step 2 - Obtain a probability distribution:**
  - t-distribution; df$=5-1=4$
- **Step 3 - Make decision rules:**

After-School Tutoring - Hypothesis Test

- **Step 1 - Formally state hypotheses**:
  - $H_1 : \mu_D > 0$; $H_0 : \mu_D \leq 0$

- **Step 2 - Obtain a probability distribution**:
  - t-distribution; df$=5-1=4$

- **Step 3 - Make decision rules**:
  - α=.05 (one-tailed), $t_{crit} = 2.132$, reject H_0 if TS>2.132.

After-School Tutoring (cont)

- **Step 4 - Calculate the test statistic**:

$$t = \frac{\overline{\mathrm{x}}_D - \mu_D}{\mathrm{s}_D/\sqrt{\mathrm{n}-1}} = \frac{2.8}{2.04/\sqrt{5-1}} =$$

After-School Tutoring (cont)

- **Step 4 - Calculate the test statistic:**

$$t = \frac{\overline{\mathrm{x}}_D - \mu_D}{\mathrm{s}_D/\sqrt{\mathrm{n}-1}} = \frac{2.8}{2.04/\sqrt{5-1}} = 2.745$$

- **Step 5 - Make a decision about the null hypothesis:**

After-School Tutoring (cont)

- **Step 4 - Calculate the test statistic**:

$$t = \frac{\overline{\mathrm{x}}_D - \mu_D}{\mathrm{s}_D/\sqrt{\mathrm{n}-1}} = \frac{2.8}{2.04/\sqrt{5-1}} = 2.745$$

- **Step 5 - Make a decision about the null hypothesis**:
  - Reject H_0, the tutoring program significantly raised test scores.

Another Example - Youth Employment and Problem Behavior

Before-and-after design: Is the transition to youth employment criminogenic?

Another Example - Youth Employment and Problem Behavior

x_1	x_2	$x_D = x_2 - x_1$	$x_D - \overline{x}_D$	$(x_D - \overline{x}_D)^2$
4	5	1	0.467	0.218
0	0	0	-0.533	0.284
1	2	1	0.467	0.218
1	1	0	-0.533	0.284
3	1	-2	-2.533	6.416
5	4	-1	-1.533	2.35
2	3	1	0.467	0.218
0	2	2	1.467	2.152
1	3	2	1.467	2.152
3	2	-1	-1.533	2.35
2	1	-1	-1.533	2.35
3	5	2	1.467	2.152
5	5	0	-0.533	0.284
3	4	1	0.467	0.218
2	5	3	2.467	6.086
$\overline{x}_1 = 2.33$	$\overline{x}_1 = 2.87$	$\sum x_D = 8$	$\sum x_D - \overline{x}_D = 0.005$	$\sum (x_D - \overline{x}_D)^2 = 27.732$
		$\overline{x}_D = 0.533$		$s_D = \sqrt{\frac{27.732}{15}} = 1.36$

Youth Employment and Problem Behavior - Hypothesis Test

- **Step 1 - Formally state hypotheses:**

Youth Employment and Problem Behavior - Hypothesis Test

- **Step 1 - Formally state hypotheses:**
  - $\mathrm{H}_1 : \mu_D > 0$; $\mathrm{H}_0 : \mu_D \leq 0$
- **Step 2 - Obtain a probability distribution:**

Youth Employment and Problem Behavior - Hypothesis Test

- **Step 1 - Formally state hypotheses**:
  - $\mathrm{H}_1 : \mu_D > 0$; $\mathrm{H}_0 : \mu_D \leq 0$
- **Step 2 - Obtain a probability distribution**:
  - t-distribution; df$=15 - 1 = 14$
- **Step 3 - Make decision rules**:

Youth Employment and Problem Behavior - Hypothesis Test

- **Step 1 - Formally state hypotheses**:
  - $H_1 : \mu_D > 0$; $H_0 : \mu_D \leq 0$
- **Step 2 - Obtain a probability distribution**:
  - t-distribution; df$=15-1=14$
- **Step 3 - Make decision rules**:
  - α=.05 (one-tailed), $t_{crit} = 1.761$, reject H_0 if TS>1.761.

Youth Employment and Problem Behavior (cont)

- **Step 4 - Calculate the test statistic:**

$$t = \frac{\overline{\mathrm{x}}_D - \mu_D}{\mathrm{s}_D/\sqrt{\mathrm{n}-1}} = \frac{.533}{1.36/\sqrt{15-1}} =$$

Youth Employment and Problem Behavior (cont)

- **Step 4 - Calculate the test statistic:**

$$t = \frac{\bar{\mathrm{x}}_D - \mu_D}{\mathrm{s}_D/\sqrt{\mathrm{n}-1}} = \frac{.533}{1.36/\sqrt{15-1}} = 1.466$$

- **Step 5 - Make a decision about the null hypothesis:**

Youth Employment and Problem Behavior (cont)

- **Step 4 - Calculate the test statistic**:

$$t = \frac{\bar{\mathrm{x}}_D - \mu_D}{\mathrm{s}_D/\sqrt{\mathrm{n}-1}} = \frac{.533}{1.36/\sqrt{15-1}} = 1.466$$

- **Step 5 - Make a decision about the null hypothesis**:
  - Accept H_0, youth employment is not criminogenic.

That's All Folks...

What are your two questions?