# Inference About the Difference Between Means $(\mu_1 - \mu_2)$

Lecture 17

Reading: Chapter 14

1

## Quote from Textbook, p. 452

*A hypothesis test really says nothing about the size of the difference. All it says is that the observed difference is large enough that we can be confident that it isn't zero. That's what the term "statistically significant" means. It doesn't say that the difference is important, financially significant, or interesting. Rejecting a null hypothesis simply says that the observed statistic is unlikely to have been observed if the null hypothesis were true.*

2

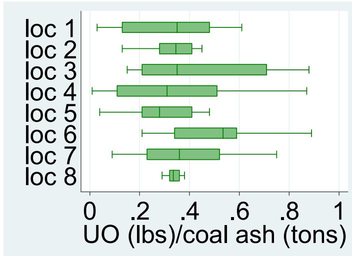
## Overview: Inference for $(\mu_1 - \mu_2)$

- Inference about comparing means $(\mu_1 - \mu_2)$
  - How do UO concentrations compare (Sparton)?
  - How do parents' beliefs compare with kids' scores?
- Hypothesis testing and CI estimation:
  - Independent samples ("unequal variances")
  - [Book] Independent samples ("equal variances")
    - **Note:** If $n_1 = n_2$, then $SE[\bar{X}_1 - \bar{X}_2]$ is same whether or not you pool (i.e. is same for unequal and equal cases)
  - Paired samples

3

## Recall Sparton Resources (p. 430)

| | n | mean | s.d. |
|---|---|---|---|
| **loc 1** | 10 | 0.325 | 0.204 |
| **loc 2** | 10 | 0.332 | 0.102 |
| **loc 3** | 10 | 0.437 | 0.270 |
| **loc 4** | 10 | 0.335 | 0.274 |
| **loc 5** | 10 | 0.283 | 0.147 |
| **loc 6** | 10 | 0.484 | 0.208 |
| **loc 7** | 10 | 0.383 | 0.200 |
| **loc 8** | 10 | 0.337 | 0.028 |




Last week investigated if each location met 0.32 threshold, but what about comparing locations with each other?

"Rising from the ashes" http://www.economist.com/node/15865280

4

## How to Compare Two Locations?

- Q1: What parameters we are interested in?
- A1: Mean conc. of UO at one location ($\mu_1$) vs. mean conc. of UO at another location ($\mu_2$)
- Q2: What's basis for inference re: $(\mu_1 - \mu_2)$?
- A2: $(\bar{X}_1 - \bar{X}_2)$
- Q3: CI estimation or hypothesis testing (HT)?
- A3: CI if we wish to measure how much two locations differ; If they differ (yes/no), then HT

5

## If Confidence Intervals Overlap?

- The 90% CI estimate for Loc 5 & Loc 6
  - Loc 5: LCL= 0.198; UCL = 0.368
    - $\bar{X} \pm t_{\alpha/2}\frac{s}{\sqrt{n}} = 0.283 \pm 1.833\frac{0.147}{\sqrt{10}} = 0.283 \pm 0.085$
  - Loc 6: LCL = 0.363; UCL = 0.605
    - Do these confidence intervals overlap?
  - BUT there *is* a statistically significant difference between these locations at a *5% significance level*
    - Checking if CI's overlap is wrong approach; Chapter 14 gives the correct approach

6

## Sampling Distribution

- Sampling distribution of $(\bar{X}_1 - \bar{X}_2)$ tells how sampling error affects $(\bar{X}_1 - \bar{X}_2)$:
  - $E[\bar{X}_1 - \bar{X}_2] = \mu_1 - \mu_2$
  - $V[\bar{X}_1 - \bar{X}_2] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$; $SD[\bar{X}_1 - \bar{X}_2] = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
  - If have a sufficiently large sample for each then $\bar{X}_1$ and $\bar{X}_2$ are Normal (CLT). Because $(\bar{X}_1 - \bar{X}_2)$ is a linear combination, then $(\bar{X}_1 - \bar{X}_2)$ also Normal
  - But we don't know $\mu_1$, $\mu_2$, $\sigma_1^2$, and $\sigma_2^2$

7

## HT & CI Est. w/ Independent Samples

- To test $H_0$: $(\mu_1 - \mu_2) = \Delta_0$ use the $t$ test statistic

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad \nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

"Unequal variances" test: doesn't assume $\sigma_1^2 = \sigma_2^2$

- For a CI estimate of $(\mu_1 - \mu_2)$ with confidence level $(1 - \alpha)$ use same $\nu$ given above and

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

8

## The Degrees of Freedom Complication

- Make sure to see the box "An Easier Rule?" on page 446 of textbook
- Also, recall that as dof gets large (>1000) you can use the Normal table: Student $t$ converges to Normal

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

*Obviously you do not want to deal with this dof formula (see footnote 1 on page 446) by hand unless necessary*

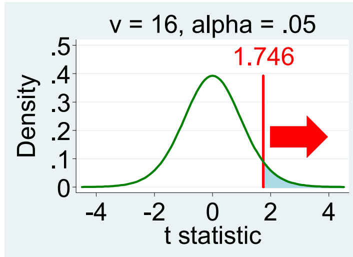
9

## Are you sure Loc 6 better than Loc 5?

$$H_0: \mu_6 - \mu_5 = 0 \; vs. H_1: \mu_6 - \mu_5 > 0$$

$$t = \frac{(\bar{X}_6 - \bar{X}_5) - \Delta_0}{\sqrt{\frac{s_6^2}{n_6} + \frac{s_5^2}{n_5}}}$$

$$t = \frac{0.201 - 0}{0.0805} = 2.50$$

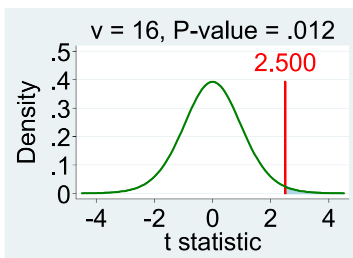



$$v = \frac{\left(\frac{s_6^2}{n_6} + \frac{s_5^2}{n_5}\right)^2}{\frac{1}{n_6 - 1}\left(\frac{s_6^2}{n_6}\right)^2 + \frac{1}{n_5 - 1}\left(\frac{s_5^2}{n_5}\right)^2} = \frac{\left(\frac{0.208^2}{10} + \frac{0.147^2}{10}\right)^2}{\frac{1}{9}\left(\frac{0.208^2}{10}\right)^2 + \frac{1}{9}\left(\frac{0.147^2}{10}\right)^2} = 16.195 \approx 16$$



## Same Question with P-value

- $H_0: \mu_6 - \mu_5 = 0$
- $H_1: \mu_6 - \mu_5 > 0$
- $t = 2.50$
- $P - value =$ $P(t > 2.50) = 0.012$
- Statistically significant?




Student *t* table tells us:
$P(t > 2.583 \mid v = 16) = 0.01$
$P(t > 2.120 \mid v = 16) = 0.025$



## How big is difference between best and worst locations? Use CI

$$(\bar{X}_6 - \bar{X}_5) \pm t_{\alpha/2}\sqrt{\frac{s_6^2}{n_6} + \frac{s_5^2}{n_5}}$$

$$(0.484 - 0.283) \pm 2.120\sqrt{\frac{0.147^2}{10} + \frac{0.208^2}{10}}$$

Already calculated $v = 16$

$$0.201 \pm 2.120 * 0.081 = 0.201 \pm 0.171$$

We're 95% confident that the mean UO concentration at Loc 6 is between 0.03 and 0.37 lbs/ton *higher* than at Loc 5. The point estimate is that the concentration at Loc 6 is a whopping 0.201 *higher* with a big margin of error of 0.171 lbs/ton.

But 99% CI is $0.201 \pm 0.235$. What does that mean?

# Paired Data

- For example, paired data would compare:
  - Employee satisfaction for 20 employees before and after a change of management and policies
  - Salaries of a random sample of 150 Ontario public sector employees in 2018 versus 2017
- With paired data, samples *not* independent
- From Lecture 16, recall the Dizon-Ross paper

*Source:* Rebecca Dizon-Ross, "Parents' Beliefs About Their Children's Academic Ability: Implications for Educational Investments" forthcoming, *American Economic Review* https://www.aeaweb.org/articles?id=10.1257/aer.20171172

13

| Summary Statistics | | |
|---|---|---|
| | Mean | SD |
| *Academic Performance (Average Achievement Scores)* | | |
| Overall score | 46.8 | 17.5 |
| Math score | 44.9 | 20.2 |
| English score | 44.2 | 20.1 |
| Chichewa score | 51.2 | 22.5 |
| (Math – English) score | 0.71 | 19.5 |
| *Respondent's Beliefs about Child's Academic Performance* | | |
| Believed Overall score | 62.4 | 16.5 |
| Believed Math score | 64.7 | 19.0 |
| Believed English score | 55.3 | 20.9 |
| Believed Chichewa score | 66.8 | 19.4 |
| Beliefs about (Math – English) score | 9.48 | 21.5 |
| Sample size (number of kids) | 5,268 | |

Excerpt from Online Appendix Table C.25, Dizon-Ross (2019); From 39 randomly selected primary schools in two districts (Machinga and Balaka) in Malawi.

14

# Excerpt of Raw Data, $n = 5{,}268$

| hhid | refchild | ave | math | engl | chich | b_ave | b_math | b_engl | b_chich |
|---|---|---|---|---|---|---|---|---|---|
| 4293 | 1 | 58 | 57 | 61 | 57 | 65 | 60 | 65 | 70 |
| 4420 | 2 | 89 | 92 | 100 | 75 | 50 | 40 | 20 | 60 |
| 14298 | 1 | 48 | 47 | 49 | 48 | 60 | 60 | 50 | 70 |
| 4102 | 1 | 59 | 80 | 53 | 43 | 75 | 65 | 75 | 85 |
| 4018 | 2 | 71 | 70 | 61 | 83 | 60 | 60 | 65 | 60 |
| 14123 | 2 | 47 | 60 | 37 | 43 | 50 | 70 | 45 | 30 |
| 4100 | 2 | 48 | 4 | 64 | 77 | 80 | 90 | 85 | 85 |
| 14477 | 1 | 50 | 51 | 59 | 40 | 65 | 40 | 75 | 50 |
| … | | | | | | | | | |
| 9626 | 1 | 59 | 60 | 40 | 78 | 80 | 80 | 60 | 80 |
| 9628 | 2 | 26 | 45 | 25 | 8 | 55 | 85 | 45 | 50 |

First two variables are *identifier variables*.

Next four variables measure achievement scores: overall is an average of the three subjects. The last four variables are beliefs.

15

## HT & CI Estimation w/ Paired Data

- To test $H_0: \mu_d = \Delta_0$ where $\mu_d = (\mu_1 - \mu_2)$:
  - Use test statistic: $t = \frac{\bar{d} - \Delta_0}{\frac{s_d}{\sqrt{n}}}$
  - where $d = (X_1 - X_2)$ and $\bar{d}$ sample mean difference; $s_d$ s.d. of $d$
- For CI estimate of $\mu_d$:
  - Use $\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$

| hhid | refchild | chich | b_chich | d |
|---|---|---|---|---|
| 4293 | 1 | 57 | 70 | 13 |
| 4420 | 2 | 75 | 60 | -15 |
| 14298 | 1 | 48 | 70 | 22 |
| 4102 | 1 | 43 | 85 | 42 |
| 4018 | 2 | 83 | 60 | -23 |
| … | | | | |
| 9626 | 1 | 78 | 80 | 2 |
| 9628 | 2 | 8 | 50 | 42 |

Define $\mu_1$ as pop. mean *belief* and $\mu_2$ as mean *actual* score. $\mu_d$ is the mean difference.

For both hypothesis testing and CI estimation: $\nu = n - 1$

16

## Describe Data

```
. summarize chich b_chich d if b_chich~=.

Variable |       Obs        Mean    Std. Dev.       Min        Max
---------+--------------------------------------------------------
   chich |     5,258    51.20426    22.52733          0        100
 b_chich |     5,258    66.76436    19.42828          0        100
       d |     5,258     15.5601    25.15133        -75        100
```

Recall Section 9.3 & Lecture 8: $s_{X-Y}^2 = s_X^2 + s_Y^2 - 2 * r * s_X * s_Y$

```
. correlate chich b_chich
(obs=5,258)
             |    chich  b_chich
-------------+------------------
       chich |   1.0000
     b_chich |   0.2883   1.0000
```

In other words, if independent, then mean of d would still be 15.5601 but s.d. would be 29.747919.

17

## Beliefs versus Actual Performance

- Are parents' beliefs biased up on average?
  - $H_0: \mu_d = 0$ vs. $H_1: \mu_d > 0$
  - $t = \frac{\bar{d} - \Delta_0}{\frac{s_d}{\sqrt{n}}} = \frac{15.5601 - 0}{\frac{25.15133}{\sqrt{5,258}}} = 44.86$
    - P-value? Statistically sig.? Economically sig.?
- How much are parents' biased on average?
  - $\bar{d} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}} = 15.5601 \pm 1.960 \frac{25.15133}{\sqrt{5,258}} =$ $15.6 \pm 0.7$ with a 95% confidence level

18

# Application to U.S. Health Policy

- "Medicaid Increases Emergency-Department Use: Evidence from Oregon's Health Insurance Experiment" Taubman et al (2013)
  - An important goal of this course: you are ready to read and understand empirical papers and research that use methods we have covered
  - We will look at the abstract and a table of results from this paper to practice these skills

http://www.sciencemag.org/content/343/6168/263.abstract

19

**ABSTRACT:** In 2008, Oregon initiated a limited expansion of a Medicaid program for uninsured, low-income adults, drawing names from a waiting list by lottery. This lottery created a rare opportunity to study the effects of Medicaid coverage using a randomized controlled design. Using the randomization provided by the lottery and emergency department records from Portland-area hospitals, we study the emergency department use of about 25,000 lottery participants over approximately 18 months after the lottery. We find that Medicaid coverage significantly increases overall emergency use by 0.41 visits per person, or 40 percent relative to an average of 1.02 visits per person in the control group. We find increases in emergency-department visits across a broad range of types of visits, conditions, and subgroups, including increases in visits for conditions that may be most readily treatable in primary care settings.

Which kind of data: observational or experimental? Causality?

20

**Table 2. Emergency-department use**

| | | *Percent with any visits*[1] | | | *Number of visits*[2] | | |
|---|---|---|---|---|---|---|---|
| | N | Percent in Control Group | Effect of Medicaid Coverage | P-value | Mean Value in Control Group | Effect of Medicaid Coverage | P-value |
| **Panel A:** Overall | | | | | | | |
| All visits | 24,646 | 34.5 | 7.0 (2.4) | 0.003 | 1.022 (2.632) | 0.408 (0.116) | <0.001 |

*Notes:* We report the estimated effect of Medicaid on emergency department use over our study period (March 10, 2008 – September 30, 2009). We report the sample size, the control mean of the dependent variable (with standard deviation for continuous outcomes in parentheses), the estimated effect of Medicaid coverage (with standard error in parentheses), and the p-value of the estimated effect. Sample consists of individuals in Portland-area zip codes (N=24,646).
[1] For the percent-with-any-visits measures, the estimated effects of Medicaid coverage are shown in percentage points.
[2] The number-of-visits measures are unconditional, including those with no visits.

21

# Panel A (entire sample) and
# Panel B (subgroups of sample)

"We report the estimated effect of Medicaid on emergency department use over our study period (March 10, 2008 – September 30, 2009) in the entire sample and in subpopulations based on *pre-randomization emergency department use*. For each subpopulation, we report ...."

*In other words, does the effect of Medicaid coverage on emergency department use vary across types of people: sicker people vs. heathier people?*

*One way to measure whether a person is sicker or healthier is by previous use of the emergency department: heavy users are likely sicker than those using it less or not at all.*

22

**Table 2. Emergency-department use**

| | | *Percent with any visits*[1] | | | *Number of visits*[2] | | |
|---|---|---|---|---|---|---|---|
| | N | Percent in Control Group | Effect of Medicaid Coverage | P-value | Mean Value in Control Group | Effect of Medicaid Coverage | P-value |
| **Panel A:** Overall | | | | | | | |
| All visits | 24,646 | 34.5 | 7.0 (2.4) | 0.003 | 1.022 (2.632) | 0.408 (0.116) | <0.001 |
| **Panel B:** By emergency department use in the pre-randomization period | | | | | | | |
| No visits | 16,930 | 22.5 | 6.7 (2.9) | 0.019 | 0.418 (1.103) | 0.261 (0.084) | 0.002 |
| One visit | 3,881 | 47.2 | 9.2 (6.0) | 0.127 | 1.115 (1.898) | 0.652 (0.254) | 0.010 |
| Two+ visits | 3,835 | 72.2 | 7.1 (5.6) | 0.206 | 3.484 (5.171) | 0.380 (0.648) | 0.557 |

[1] For the percent-with-any-visits measures, the estimated effects of Medicaid coverage are shown in percentage points.
[2] The number-of-visits measures are unconditional, including those with no visits.

23