Lecture 14

Inference for Two Means

These slides are the property of Dr. Wendy Rummerfield ©

agenda

Reminders	Mid-Semester survey 1 due tonight (mistake about STEM LAB resource)
lecture part 1	hypothesis testing for a difference in means - independent groups
lecture part 2	hypothesis testing for paired means
R activity	hypothesis testing for two means (`ht_two_means.Rmd`)

Hypothesis testing for two means (independent groups)

1) Hypotheses

$$H_0 : \mu_1 - \mu_2 = 0 \quad \mu_1 = \mu_2$$

$$H_A : \mu_1 - \mu_2 < / > / \neq 0 \quad \mu_1 \lessgtr \mu_2$$

2) Test conditions:
   - Independence
   - Normality

3) Calculate test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t_{df}$$

$$df = \min\{n_1, n_2\} - 1$$

4) Compute p-value

```
<: pt(t, df)
>: pt(t, df, lower.tail = FALSE)
≠: 2*pt(-abs(t), df)
```

5) Decision and conclusion in context

Do people living in California make more money, on average, than people living in Ohio?

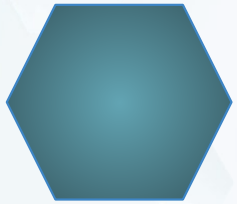
What is the parameter of interest? Define in symbols and interpret the parameter in context.

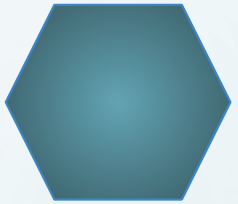
$\mu_1 - \mu_2$ = true difference in avg income for people in CA minus OH.

↓

$\bar{x}_1 - \bar{x}_2$

Hypothesis testing





null **alternative**

1 = CA , 2 = OH

$$H_0 : \mu_1 - \mu_2 = 0 \qquad H_A : \mu_1 - \mu_2 \neq 0$$

$$\mu_{CA} - \mu_{OH} = 0$$

$<$

$>$

$$\mu_1 - \mu_2 > 0$$

Check conditions

$n_1 = 150$
$n_2 = 160$

Independence ✓

observational units are randomly selected, but each group is independent of one another

10%: $10 \times n_1 = 1500 <$ pop'n size
$10 \times n_2 = 1600 <$ " " "

Normality ✓

- $n_1, n_2 \geq 30$ and no extreme outliers OR
- $n_1, n_2 < 30$ and data are approximately normal

$$t = \frac{\text{statistic - null value}}{\text{standard error of statistic}}$$

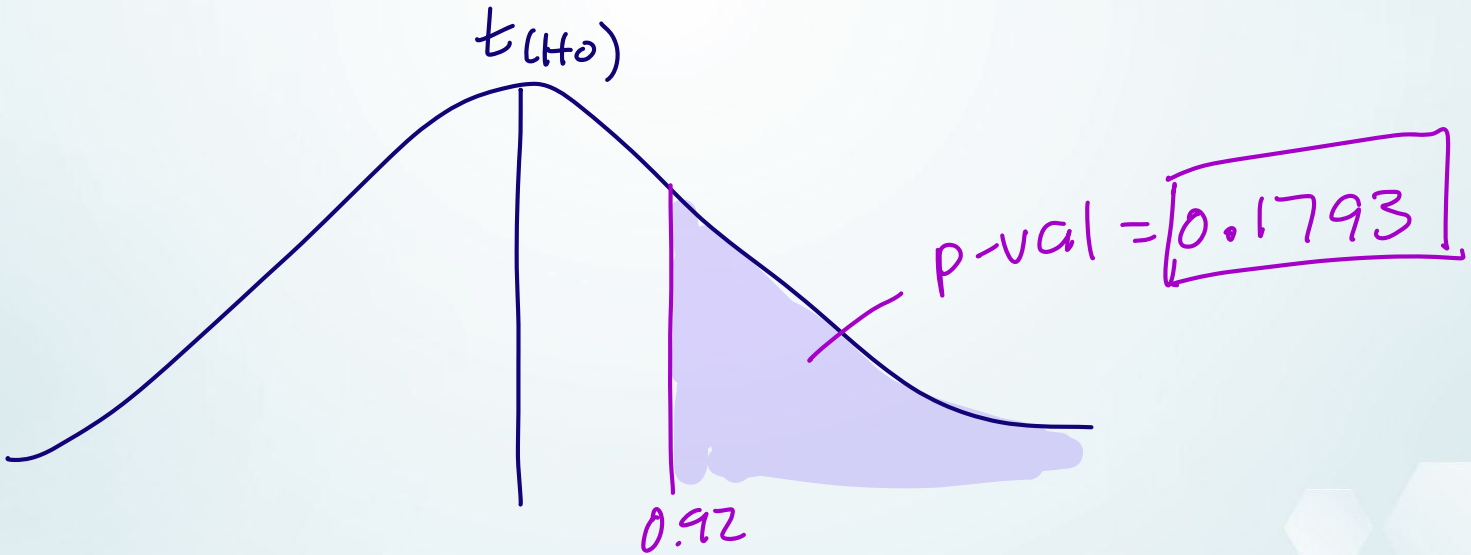
$$t = \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$= 0.9207$ (see Rmd)

Calculate test statistic

Compute p-value

```
<: pt(t, df)  |  >: pt(t, df, lower.tail = FALSE)  |  ≠: 2*pt(-abs(t), df)
```



Conclusion in context

Reject the null: We have enough evidence that the true mean of _____ for [group 1] is less than/different from/ more than the mean for [group 2].

Fail to reject the null: We do not have enough evidence that the true mean of _____ for [group 1] is less than/different from/ more than the mean for [group 2].

Confidence intervals

$$\bar{x}_1 - \bar{x}_2 \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

BP relax	BP Test	Test - relax
X_1	Y_1	$Y_1 - X_1 = d_1$
X_2	Y_2	$Y_2 - X_2 = d_2$
X_3	Y_3	
$\vdots$	$\vdots$	$\vdots$
X_n	Y_n	$Y_n - X_n = d_n$

Hypothesis testing for two means (dependent groups)

1) Hypotheses $H_0 : \mu_d = 0$
   $H_A : \mu_d < / \neq / > 0$
2) Test conditions:
   - Independence
   - Normality
3) Calculate test statistic

$$t = \frac{\bar{x}_d - 0}{s_d / \sqrt{n_d}} \sim t_{df}$$

df = # of pairs − 1

4) Compute p-value

```
<: pt(t, df = n_d - 1)
>: pt(t, df = n_d - 1,
      lower.tail = FALSE)
≠: 2*pt(-abs(t), df = n_d - 1)
```

5) Decision and conclusion in context

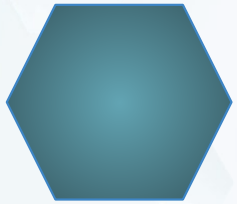
After being primed with sad memories, does anti-anxiety medication positively affect memory?
Kaggle data

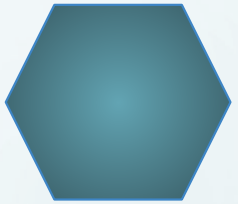
What is the parameter of interest? Define in symbols and interpret the parameter in context.

μ_d = mean of diff. in memory score after - before sad mems.

d = after - before

Hypothesis testing





null

alternative

$$H_0 : \mu_d = 0$$

$$H_A : \mu_d < 0$$

$$H_A : \mu_d = 0$$

$$H_A : \mu_d > 0$$

Check conditions

independence ✓

observational units are paired, but randomly selected

randomly sampled

normality ✓

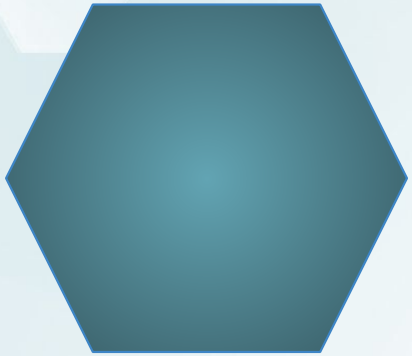
$n_d \geq 30$ and no extreme outliers

$$t = \frac{\text{statistic - null value}}{\text{standard error of statistic}}$$

$$t = \frac{\bar{x}_d - 0}{s_d / \sqrt{n_d}}$$ = # of pairs

$= 4.6179$

Calculate test statistic



Compute p-value

```
<: pt(t, df)   |   >: pt(t, df, lower.tail = FALSE)   |   ≠: 2*pt(-abs(t), df)
```

0.000028

Conclusion in context

Reject the null: We have enough evidence that the true differences of the mean _____ for [group 1] compared to [group 2] is less than/different from/ more than 0.

Fail to reject the null: We do not have enough evidence that the true differences of the mean _____ for [group 1] compared to [group 2] is less than/different from/ more than 0.

Confidence intervals

$$\bar{x}_d \pm t^* \frac{s_d}{\sqrt{n_d}}$$

to R!

ht_two_means.Rmd