Statistics 2023
Thursday, November 20, 2014

EXTRA Point Opportunity
Link for Course Evaluation WILL BE
sent to your OKSTATE address by OSU
Complete the evaluation for 10 points

HW 18 due on Tuesday, 11pm
SLIC is open Noon-4pm MRTW
SLIC will open early at 10:30 on Tuesday, November 25
No Class on Tuesday, November 25, 2014

Lesson 17, Comparing the Means of Two Populations

Comparing the mean cost of rotary versus standard engines.

SLIC is open
Noon - 4pm MTWR

Tuesday, Nov. 25. SLIC
opens early at 10:30.

SLIC IS CLOSED DURING
FINALS WEEK.

A 95% confidence interval to estimate $\mu_1 - \mu_2$ when σ_1^2 and σ_2^2 are known:

$$\bar{x}_1 - \bar{x}_2 \pm z_{\frac{.05}{2}} \, \sigma_{\bar{x}_1 - \bar{x}_2}$$

$$\bar{x}_1 - \bar{x}_2 \pm z_{.025} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Using the data from the sliding door example on Tuesday:

$$815 - 800 \pm 1.96 \sqrt{\frac{1600}{100} + \frac{900}{100}}$$

$$15 \pm 1.96 \; (5)$$

$$15 \pm 9.8$$

$$(5.2, \quad 24.8)$$

This is a set of reasonable, plausible, and not rejectable values for the parameter, $\mu_1 - \mu_2$.

Would $H_0: \mu_1 - \mu_2 = 0$ testing equal means be rejected at the 5% significance level.

3 Cases that occur when a confidence interval is used to estimate a difference parameter, such as $\mu_1 - \mu_2$.

1. $(+\#, +\#) \Longrightarrow \mu_1 > \mu_2$

2. $(-\#, -\#) \Longrightarrow \mu_2 > \mu_1$

3. $(-\#, +\#) \Longrightarrow$ no evidence that $\mu_1 \neq \mu_2$

Lesson 17. Comparing two means.

Parameter is $\mu_1 - \mu_2$

Point Estimator is $\bar{x}_1 - \bar{x}_2 = \widehat{\mu_1 - \mu_2}$

Point Estimator is not precise, so consider

Standard Error of $\bar{x}_1 - \bar{x}_2$

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

But what if the σ's are unknown?

The standard error of $\bar{X}_1 - \bar{X}_2$ is

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

2 ways to estimate the above.

1. No assumption on $\sigma_1^2 = \sigma_2^2$ OR $\sigma_1^2 \neq \sigma_2^2$

$$S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

2. $$S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$$

where $$S_p^2 = \frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{(n_1 - 1) + (n_2 - 1)}$$

$$S_p^2 = \frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n - 2}$$

Use S_p^2 only when $\sigma_1^2 = \sigma_2^2$ is assumed.

Example. Lesson 17. $\mu_1 - \mu_2$

Difference between mean operating cost per 1000 miles of cars with rotary and standard engines. Refer to page 11 of 31 in Lesson 17.

Rotary	Standard
$n_1 = 8$	$n_2 = 12$
$\bar{x}_1 = \$56.90$	$\bar{x}_2 = \$52.73$
$S_1 = \$4.85$	$S_2 = \$6.35$

indicates use S_p^2 see pg 11 of 31 Lesson 17.

Assuming equal variances, that is, $\sigma_1^2 = \sigma_2^2$ what is the 90% confidence interval to estimate $\mu_1 - \mu_2$?

$$\bar{x}_1 - \bar{x}_2 \pm t_{\frac{.10}{2}(n_1+n_2-2)} \, S_{\bar{x}_1 - \bar{x}_2}$$

$$\bar{x}_1 - \bar{x}_2 \pm t_{.05(8+12-2)} \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$$

where $S_p^2 = \dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \dfrac{(8-1)4.85^2 + (12-1)6.35^2}{8+12-2}$

$$S_p^2 = 33.789$$

$$56.96 - 52.73 \pm 1.734 \sqrt{\frac{33.789}{8} + \frac{33.789}{12}}$$

$$4.23 \pm 1.734(2.6532)$$

$$(-0.3706, \; 8.8306)$$

90% $(-.3, 8.8)$ $\alpha = .10$

Reject $H_0: \mu_1 - \mu_2 = 0$ at $\alpha = .10$

when $p > .10 = \alpha \Rightarrow$

Do not Reject H_0.

Mean operating costs of rotary versus standard engines.

① $H_0: \mu_1 - \mu_2 = 0$

$H_A: \mu_1 - \mu_2 \neq 0$

② $t_{calc} = \dfrac{\bar{X}_1 - \bar{X}_2 - D_0}{S_{\bar{X}_1 - \bar{X}_2}} = \dfrac{\bar{X}_1 - \bar{X}_2 - D_0}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}} = 1.59.$

③ If H_0 is true $t \sim t(18)$.

④ P-value $= P(t(18) < -1.59) + P(t(18) > 1.59) \Rightarrow$

$\Rightarrow .05 < \frac{P}{2} < .10 \Rightarrow .10 < P < .20$.

$\Rightarrow$ Do not Reject H_0 w/ reasonable error rate.

⑤ These data do not provide evidence that $\mu_1 - \mu_2 \neq 0$ or in words that the data do not indicate a significant difference in mean cost of rotary versus standard engine.

Hypothesis Test on $\mu_1 - \mu_2$

not assuming $\sigma_1^2 = \sigma_2^2$

Rotary vs. Standard engine expenses

$H_0: \mu_1 - \mu_2 = 0$

① $H_A: \mu_1 - \mu_2 \neq 0$

What if example exam question

What is H_A if one is seeking evidence in the data that μ_2 exceeds μ_1 by more than 5 units?

$H_A: \mu_2 - \mu_1 > 5$

Define the parameter as $\mu_1 - \mu_2$

$H_A: \mu_1 - \mu_2 < -5$

The test statistic for not assuming equal variances is

$$t = \frac{\bar{X}_1 - \bar{X}_2 - D_0}{S_{\bar{X}_1 - \bar{X}_2}} = \frac{\bar{X}_1 - \bar{X}_2 - D_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$= \frac{56.9 - 52.73 - 0}{\sqrt{\frac{4.85^2}{8} + \frac{6.35^2}{12}}} = \frac{4.17}{2.51}$$

$$= 1.66$$

$t_{calc} \sim t(df)$ when Ho is true,

$$\text{where } df. = \frac{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}{\frac{\left(\frac{S_1^2}{n_1}\right)}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)}{n_2 - 1}} = 17.55 \Rightarrow$$

$df = 18$

② $$t^*_{calc} = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{S_{\bar{x}_1 - \bar{x}_2}}$$

$$= \frac{56.90 - 52.73 - 0}{\sqrt{\frac{4.85^2}{8} + \frac{6.35^2}{12}}}$$

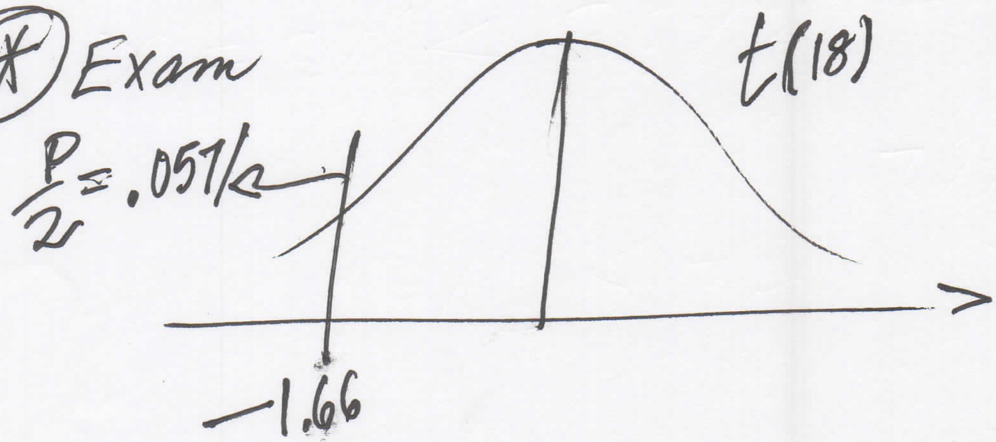
$$= \frac{\cancel{4.77}\ 4.17}{2.51}$$

$$= 1.66$$

③ $t^*_{calc} \sim t^*(df)$ if H_0 is true

④ $P\text{-value} = P(t(18) < -1.66) + P(t(18) > 1.66)$

(*) Exam



$$\frac{P}{2} = .0571 \Rightarrow P = 2(.0571) = .1142$$

Do not Reject H_0 with $p = .1142$

What if $n_1 = n_2$?

$$S_p^2 = \frac{S_1^2 + S_2^2}{2}$$

and

$$\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

<u>ONLY</u> $n_1 = n_2$

Not assuming equal variances leads to using separate variance estimates in the standard error calculation.

Not assuming $\sigma_1^2 = \sigma_2^2$ leads to using $S_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

90% Confidence Interval is

$$\bar{x}_1 - \bar{x}_2 \pm t^*_{\frac{\alpha}{2}(df)} \cdot S_{\bar{x}_1-\bar{x}_2}$$

$$\bar{x}_1 - \bar{x}_2 \pm t^*_{\frac{\alpha}{2}(df)} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$56.9 - 52.73 \pm t^*_{.05(18)} \sqrt{\frac{4.85^2}{8} + \frac{6.35^2}{12}}$$

$$df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}} = \frac{\left(\frac{4.85^2}{8} + \frac{6.35^2}{12}\right)^2}{\frac{\left(\frac{4.85^2}{8}\right)^2}{8-1} + \frac{\left(\frac{6.35^2}{12}\right)^2}{12-1}} = 17.55$$

use $df = 18$

When Z? when t

If σ_1^2 and σ_2^2 are known use Z

Tuesday Nov 25 HW

If σ_1^2 and σ_2^2 are unknown

then use t

But, if $df > 30$ then

use Z to estimate t.