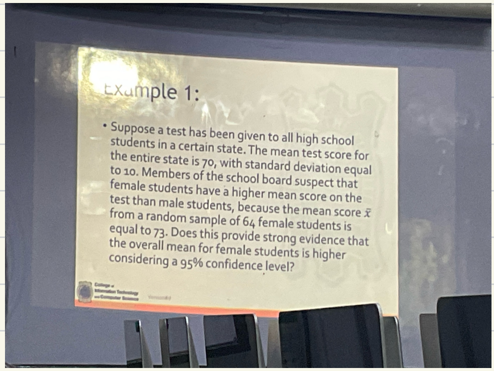
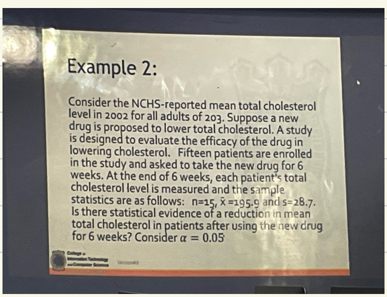
ST $\rightarrow$ Mean $<$ μ - population
$\bar{x}$ - sample

Interpret - pval } z-test
- test stat
- test-stat $\Rightarrow$ t-test



$\mu_0 = 70$
$\sigma = 10$
$n = 64$
$\bar{x} = 73$
$C = 95\%$
$a = 0.05$ or 5%

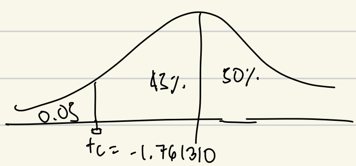


$\mu_0 = 203$
$\sigma = 28.7$
$n = 15$ $(n<30) \Rightarrow$ t-test
$\bar{x} = 195.9$
$a = 0.05$

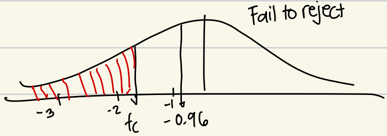
$H_0: \mu \geq 203$
$H_a: \mu < 203$ (left-tailed)
$df = n-1$
$= 15-1$
$= 14$

$se_0 = \frac{s}{\sqrt{n}} = \frac{28.7}{\sqrt{15}} = 7.41$

$t = \frac{\bar{x} - \mu_0}{se_0} = \frac{195.9-203}{7.41} = -0.96$



Fail to reject



$s = 4.99$

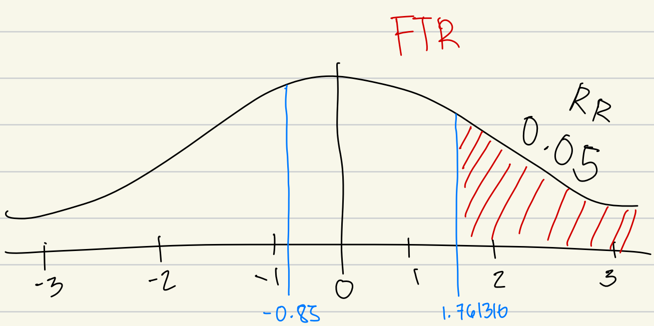
$\mu_0 = 7$ $\qquad H_0 : \mu = 7$

$n = 15$ $\qquad H_a : \mu > 7$

$\bar{x} = 5.9$

$\alpha = 0.03$

$$se_0 = \frac{s}{\sqrt{n}} = \frac{4.99}{\sqrt{15}} = 1.288409873$$



Two Groups — Proportions (Categorical - Qualitative) — Variable
Means (Numerical - Quantitative)

Comparing Means

- independent
  - ↳ samples that have no direct associative
  - ex. → method 1 → students A
  - ↳ method 2 → students B
- paired
  - ↳ connected — Diet program: pre-diet weight, post-diet weight
  - 1 sample

Independent

	Groups	n	$\bar{x}$	SD
①	Cars	40	5.3	2.2
②	Trucks	40	7.1	3.0

Is there a difference between the two means consider

$c = 90\%$ two tailed

$\alpha = 10\% \rightarrow 10/2 = 5\% / 0.05$

① Assumption

$n_1 = 40 \rightarrow$ z-test

$n_2 = 40 \rightarrow$ z-test

② Hypothesis

$H_0: \mu_1 = \mu_2 \ (\mu_1 - \mu_2 = 0)$

$H_a: \mu_1 \neq \mu_2$

③ Test statistic

$$se_0 = \sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}} \qquad z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{se_0}$$

$$= \sqrt{\frac{(2.2)^2}{40} + \frac{(3.0)^2}{40}} \qquad = \frac{(5.3 - 7.1) - 0}{0.5882}$$

$$= 0.5882 \qquad = -3.06 \approx 0.4989$$

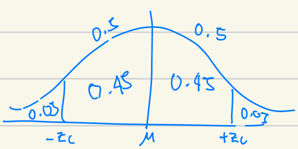
Interpret the
④ P-value

$2(0.5 - 0.4989)$

$= 0.0022$ - pval

$= 0.05$ - α → bigger

⑤ ∴ Reject H_0 that states that



(selling price) Groups	n	$\bar{x}$	s	s^2
1) PHP 13	10 (camara)	171	26	676
2) PHP 10	5 (Turon)	198	29	841

Is there a difference between the means

Consider

$c = 90\%$

$a = 10\% \approx 10\%/2 = 5\% / 0.05$

① $n_1 = 10 \rightarrow$ t-test

$n_2 = 5 \rightarrow$ t-test

② $H_0 = \mu_1 = \mu_2 \; (\mu_1 - \mu_2 = 0)$

$H_a = \mu_1 \neq \mu_2$ (two-tailed)

③ $se_0 = \sqrt{\dfrac{(s_1)^2}{n_1} + \dfrac{(s_2)^2}{n_2}}$

$= \sqrt{\dfrac{(26)^2}{10} + \dfrac{(29)^2}{5}}$

$= 15.3558$

$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{se_0}$

$= \dfrac{(171 - 198) - 0}{15.3558}$

$= -1.76$

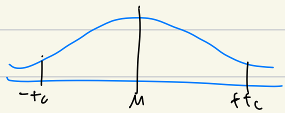
④ $df = \dfrac{\left[\dfrac{(s_1)^2}{n_1} + \dfrac{(s_2)^2}{n_2}\right]^2}{\dfrac{1}{n_1 - 1}\left(\dfrac{(s_1)^2}{n_1}\right)^2 + \dfrac{1}{n_2 - 1}\left(\dfrac{(s_2)^2}{n_2}\right)^2}$ (unequal)

$= \dfrac{\left[\dfrac{(26)^2}{10} + \dfrac{(29)^2}{5}\right]^2}{\dfrac{1}{10-1}\left(\dfrac{(26)^2}{10}\right)^2 + \dfrac{1}{5-1}\left(\dfrac{(29)^2}{5}\right)^2}$

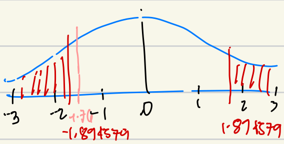
$= 7.33$

$= 7$

5.

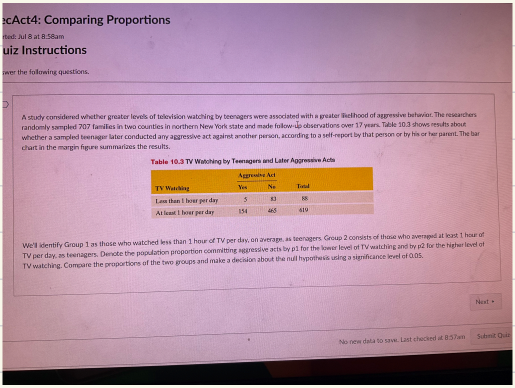


$df = 7$
$a = 0.05$ } $t_c = 1.894579$



6. ∴ FTR the H_0 because ...

∴ Reject the H_a because ...


ecAct4: Comparing Proportions

rted: Jul 8 at 8:58am

uiz Instructions

swer the following questions.

A study considered whether greater levels of television watching by teenagers were associated with a greater likelihood of aggressive behavior. The researchers randomly sampled 707 families in two counties in northern New York state and made follow-up observations over 17 years. Table 10.3 shows results about whether a sampled teenager later conducted any aggressive act against another person, according to a self-report by that person or by his or her parent. The bar chart in the margin figure summarizes the results.

Table 10.3 TV Watching by Teenagers and Later Aggressive Acts

	Aggressive Act		
TV Watching	Yes	No	Total
Less than 1 hour per day	5	83	88
At least 1 hour per day	154	465	619

We'll identify Group 1 as those who watched less than 1 hour of TV per day, on average, as teenagers. Group 2 consists of those who averaged at least 1 hour of TV per day, as teenagers. Denote the population proportion committing aggressive acts by p1 for the lower level of TV watching and by p2 for the higher level of TV watching. Compare the proportions of the two groups and make a decision about the null hypothesis using a significance level of 0.05.

Next ▸

No new data to save. Last checked at 8:57am

Submit Quiz

① $n_1 = 88$ ② $p_1 = p_2 \; (p_1 - p_2 = 0)$

$n_2 = 619$ $\quad p_1 \neq p_2$

③ $se_0 = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$

$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$

$= \frac{5 + 154}{88 + 619}$

$= \frac{159}{707}$

$= 0.22489392$

$\sqrt{(0.2248\ldots)(0.77510608)(0.011363636 + 0.0016155897)}$

0.012979145

$\sqrt{0.002262491}$

0.04756554

$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{se_0}$

$= 0.056818182 - 0.24878837$

$= -0.19197019$

Paired Samples

> 1 sample group

> direct association

$\bar{x}_d$ = difference of n (before / after)

$H_0 = \mu_1 = \mu_2 \ (\mu_1 - \mu_2 = 0)$

Example:

① $n = 8$ (not large enough) t-test

② $H_0 = \mu_1 = \mu_2 \ (\mu_1 - \mu_2 = 0)$

$H_a = \mu_1 \neq \mu_2$ (two-tailed test)

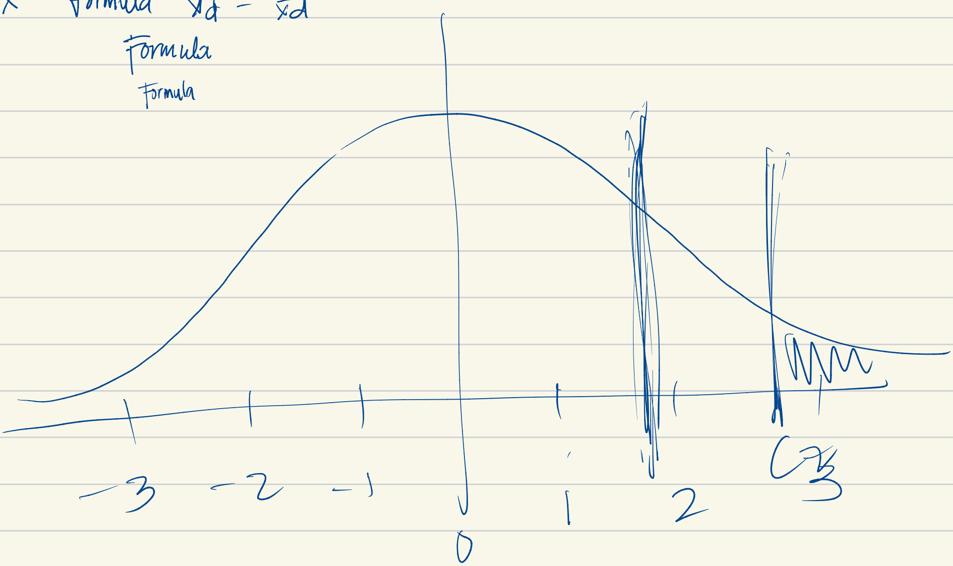
③ $SE_D = \dfrac{s_d}{\sqrt{n}}$ $\qquad s = \sqrt{\dfrac{\Sigma(x - \bar{x})^2}{n-1}}$

$t = \dfrac{\bar{x}_d - 0}{SE_D} =$

$x - \bar{X}$ - formula $\quad x_d - \bar{x}_d$

Formula

Formula



Quizzes > Lesson05: Lecture Activity

on05: Lecture Activity

Jul 10 at 8:57am

Instructions

the following questions.

The following descriptive statistics were obtained from a study (Saud al-Obaidi et al., *Journal of Rehabilitation Research and Development*, vol. 40, 2003) that aimed to compare the weight of Kuwaiti men with Swedish men between the ages of 20 to 29 years.

	Group size	Mean weight (kg)	Standard deviation
Kuwaiti men	15	81.57	26.26
Swedish men	15	70.73	12.56

Interpret and explain what (if any) effect a country has on the mean weight of its men. Use a 95% confidence level.

Next ▸

Quiz saved at 8:56am Submit Qu

① $n_1 = 15$

$n_2 = 15$

②

Test for Independence

- ↳ crosstab, Contingency table
- ↳ Variables are either dependent or not
- ↳ Chi-squared test (χ^2)

Lung Cancer

Smoking	Present	Absent	Column_Total
Smoker	605 E_1	185 E_2	790
Non-Smoker	122 F_1	312 F_2	434
Raw_Total	727	497	1224

① x (cause) = smoking
y (effect) = lung cancer
$\alpha = 0.05$

② H_0: smoking $\perp$ lung cancer
H_a: Smoking $\pm$ lung cancer

$$E_{1,1} = \frac{727 \ast 790}{1{,}224} = 469.22 = 469$$

$$E_{1,2} = \frac{497 \ast 790}{1{,}224} = 320.78 = 321$$

$$E_{2,1} = \frac{727 \ast 434}{1{,}224} = 257.78 = 258$$

$$E_{2,2} = \frac{497 \ast 434}{1{,}224} = 176.22 = 176$$

* Slovins
for CC12 Research

Lung Cancer

Smoking	Present	Absent	Column Total
Smoker	469 E_1	321 E_2	790
Non-Smoker	238 F_1	176 F_2	429
Raw Total	727	497	1224

(3) $X^2 = \Sigma \frac{oc - ec}{ec}$

oc = Observed count
ec = Expected count

$$= \frac{((605 - 469^2))}{469} + \frac{(185 + 321)^2}{321} + \frac{(122 - 258)}{258} + \frac{(312 - 176)^2}{176}$$

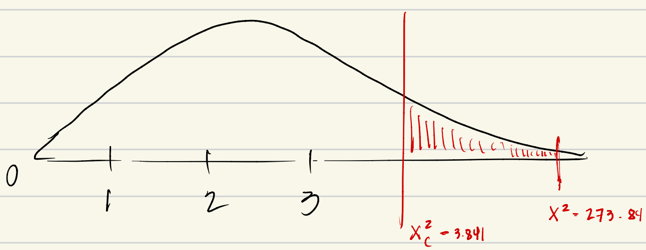
$= 273.84$

(4) $df = (r-1) \times (c-1)$
$= (2-1) \times (2-1)$
$= 1$

r = # of rows
c = # of columns

$X_c^2 = 3.841$

* right skewed



∴ Reject the Ho, that states that the variables smoking and lung cancer are independent, therefore we accept that s and l are associated

For Research:

- use Jforms
- methods learned on CC8 and CC12 only
- Descriptive vs. Inferential
  - describe within ○
  - measures of central tendencies

  ↳ assumption about
  ↳ outside the ○

POP