Stats Review
Chapter 11

Revised 8/16

Note:

This review is meant to highlight basic concepts from the course.   It does not cover all concepts presented by your instructor. Refer back to your notes, unit objectives, handouts, etc. to further prepare for your exam.

The questions are displayed on one slide followed by the answers are displayed in red on the next.

This review is available in alternate formats upon request.

Is the sample independent or dependent?
Is the response variable qualitative or quantitative?

- A researcher wants to compare the mean allowances between siblings.

- In 2005, a random sample of 100 adults were asked what is their favorite color. The process was repeated in 2009 with 100 different adults.

Is the sample independent or dependent?
Is the response variable qualitative or quantitative?

- A researcher wants to compare the mean allowances between siblings.
  - Dependent, qualitative

- In 2005, a random sample of 100 adults were asked what is their favorite color. The process was repeated in 2009 with 100 different adults.
  - Independent, quantitative

Inference on 2 Population Proportions: Independent samples

A teacher wants to test the pass rates of two delivery methods. A random sample of 100 students was selected. Group 1 had an online delivery with 50 students, 25 of whom passed. Group 2 had a traditional delivery with 50 students, 30 of whom passed. Is there evidence that the traditional delivery method had a higher pass rate?

Inference on 2 Population Proportions: Independent samples

A teacher wants to test the pass rates of two delivery methods. A random sample of 100 students was selected. Group 1 had an online delivery with 50 students, 25 of whom passed. Group 2 had a traditional delivery with 50 students, 30 of whom passed. Is there evidence that the traditional delivery method had a higher pass rate?

Are the conditions met?

1. Independent random sample? Yes
2. $n_1\hat{p}_1(1-\hat{p}_1) \geq 10$ $\quad$ $n_2\hat{p}_2(1-\hat{p}_2) \geq 10$

   $\hat{p}_1 = \frac{25}{50} = .5$ $\quad$ $\hat{p}_2 = \frac{30}{50} = .6$

   $50 * .5(1-.5) = 12.5 \geq 10$ $\quad$ $50 * .6(1-.6) = 12 \geq 10$

   Both are greater than 10, so this condition is met.

3. Both n's are less than 5% of the population. This condition is met.

Step 1: Determine the Null and Alternative Hypothesis

$H_0: p_1 = p_2$ or $H_0: p_1 - p_2 = 0$

$H_0: p_1 < p_2$

We use < because we think the traditional delivery (group 2) has a higher pass rate.

Step 2: Select a level of significance.

None was given, so we use α=.05

Continues

Inference on 2 Population Proportions

Step 3: Compute test statistic

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{25 + 30}{50 + 50} = .55$$

Test statistic: $z_o = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{.5 - .6}{\sqrt{.55\,(1-.55)}\sqrt{\frac{1}{50}+\frac{1}{50}}} = -1.01$

Classical Approach

Determine the Critical Value with α=.05

$-z_\alpha$ is used for Left-tailed

(look up .05 inside the table to find z)

$-z_\alpha$=-1.645

Step 4: Compare the critical value to the test statistic. For left tailed: Reject if $z_0 < -z_\alpha$

-1.01>=1.645, we do not reject

P-value approach

Looking up z_o=-1.01 in the table we get

p-value =.1562

If you do the entire thing with technology, p-value=.1574

Step 4: If p-value<α, reject the null hypothesis

.1562>.05, we do not reject

Step 5: State your conclusion

We do *not* reject the null hypothesis. There is *not* sufficient evidence that the pass rate of the students with the traditional delivery is greater than the pass rate of the students with the online delivery.

Two Population Proportions Confidence Intervals

A survey of 900 males found that 802 had the chicken pox as a kid while a survey of 700 females found that 545 had the chicken pox as a kid. Construct a 95% confidence interval to determine if there is a difference in the proportions of males and females that had the chicken pox as kids.

Two Population Proportions Confidence Intervals

A survey of 900 males found that 802 had the chicken pox as a kid while a survey of 700 females found that 545 had the chicken pox as a kid. Construct a 95% confidence interval to determine if there is a difference in the proportions of males and females that had the chicken pox as kids.

Check the conditions: see previous problem or page 531.

Step 1: State the Null and Alternative hypothesis

$H_0: p_1 = p_2$ or $H_0: p_1 - p_2 = 0$

$H_1: p_1 \neq p_2$ or $H_1: p_1 - p_2 \neq 0$

Step 2: Construct the Confidence interval

For a 95% C.I., $z_{\alpha/2}$=1.96 (see chapter 9)

$\hat{p}_1 = \frac{x_1}{n_1} = \frac{802}{900} = .8911$ $\hat{p}_2 = \frac{x_2}{n_2} = \frac{545}{700} = .7786$

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_1}}$$

$$(.8911 - .7786) \pm 1.96\sqrt{\frac{.8911(1-.8911)}{900} + \frac{.7786(1-.7786)}{700}} = (.07566, .14942)$$

Step 3: Reject H_0 if 0 (or the difference in the null hypothesis) is in the confidence interval.

0 is in the interval, so we reject the null hypothesis.

Step 4: State your conclusion.

We reject the null hypothesis. There is sufficient evidence to suggest that is a difference in the proportions of males and females that had the chicken pox.

Inference on 2 Population Proportions: Dependent samples

Assuming the samples are randomly selected and are dependent. Test if the whether the population proportions differ at the α=.01 level.

		Treatment A	
		success A	failure A
Treatment B	success B	2415	10
	failure B	3542	70

Inference on 2 Population Proportions: Dependent samples

Assuming the samples are randomly selected and are dependent. Test if the whether the population proportions differ at the α=.01 level.

		Treatment A	
		success A	failure A
Treatment B	success B	2415	10
	failure B	3542	70

Conditions met?

Samples random and dependent? Yes

The total number of observations where the outcome (success/failure) differ must be greater than 10? $(f_{12} + f_{21} \geq 10)$ 10+3542 ≥10 so yes

Step 1: Determine the Null and Alternative hypothesis

$H_0: p_1 = p_2$

$H_1: p_1 \neq p_2$

Step 2: Select the significance level, α

α=.01 was given

Continues

Inference on 2 Population Proportions: Dependent samples

Step 3: Computer the Test Statistic

$$z_0 = \frac{|f_{12} - f_{21}| - 1}{\sqrt{f_{12} + f_{21}}} = \frac{|10 - 3542| - 1}{\sqrt{10 + 3542}} = 59.246$$

Classical Approach

Find $z_{\alpha/2}$ where α=.01

$z_{\alpha/2}$= 2.576

Step 4: Reject if $z_{\alpha/2}$<z_0

2.576<59.249 so we reject H_0

P-value approach

Look up z_0 in the table.

z_0 is off the chart so the p-value =0.

Step 4: Reject if the p-value<α

0<.01, so we reject H_0

Step 5: State your conclusion

We reject the null hypothesis. There is sufficient evidence that the proportion of people who had success with treatment A is different that those that had success with treatment B.

Sample Size for Estimates $p_1 - p_2$

You wish to estimate the difference in proportions of girls and boys who hit puberty before the age of 11. She wants to be within 4% points and have 95% confidence level.

a) If 30% of boys and 52% of girls reach puberty by age 11, what must the sample size be?

b) What would the sample size be without prior estimates?

Sample Size for Estimates $p_1 - p_2$

You wish to estimate the difference in proportions of girls and boys who hit puberty before the age of 11. She wants to be within 4% points and have 95% confidence level.

For 95% Confidence, $z_{\alpha/2}$=2.576

a) If 30% of boys and 52% of girls reach puberty by age 11, what must the sample size be?

$$n = n_1 = n_2 = [\hat{p}_1(1-\hat{p}_1)+\hat{p}_2(1-\hat{p}_2)]\left(\frac{z_{\alpha/2}}{E}\right)^2$$

$$n = n_1 = n_2 = [.3(1-.3)+.52(1-.52)]\left(\frac{2.576}{.04}\right)^2 = 1906.1 \approx \mathbf{1907} \textbf{ people}$$

b) What would the sample size be without prior estimates?

$$n = n_1 = n_2 = .5\left(\frac{z_{\alpha/2}}{E}\right)^2$$

$$n = n_1 = n_2 = .5\left(\frac{2.576}{.04}\right)^2 = 2073.7 \approx \mathbf{2074} \textbf{ people}$$

Difference of 2 Means Using Matched Pairs

A researcher wanted to know if the breaking distance was different from a wet road and an icy road. 9 different cars were used with the same driver and tires. The results are below. Assume conditions of a matched-pair design were met.

Car	1	2	3	4	5	6	7	8	9
Wet	102	106	110	108	100	106	107	102	106
Icy	116	112	114	115	117	120	115	112	113
Difference	-14	-6	-4	-7	-17	-14	-8	-10	-7

a) Is there a difference in the breaking distance at the α=.05 level. Use the p-value approach, $\bar{d} = -9.67$ and $s_d = 4.39$

b) Construct a 95% confidence interval.

**The difference will most likely not be given to you. You will may need to find $\bar{d}$ (average difference) and s_d (sample standard deviation) will most likely be given on a test.*

Difference of 2 Means Using Matched Pairs

a) Is there a difference in the breaking distance at the α=.05 level given $\bar{d} = -9.67$ and $s_d = 4.39$. Use the p-value approach,

Step 1: Determine the null/alternative hypothesis

$H_0: \mu_d = 0$

$H_1: \mu_d \neq 0$

Step 2: Select the level of significance. α=.05 was given

Step 3: Compute test statistic with Degrees of freedom = n-1=9-1=8

$t_o = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{-9.67}{4.39/\sqrt{9}}$=-6.61

Step 4: reject if p-value< α

Using technology p-value=.0002 and using the table p-value <.0005

Reject the null hypothesis

Step 5: State the conclusion.

Reject the null hypothesis. There is sufficient evidence to suggest that there is a difference in breaking speed between wet and icy roads.

b) Construct a 95% confidence interval.

Critical value $t_{\alpha/2} = t_{\frac{.05}{2}}$ = $t_{.025}$and n-1=8 degrees of freedom is 2.306

$\bar{d} \pm t_{\alpha/2} \cdot \frac{s_d}{\sqrt{n}}$=-9.67 $\pm$ 2.306 $\cdot \frac{4.39}{\sqrt{9}}$**=(-13.044,-6.296)**

Inference About 2 Means: Independent Samples

Researchers conducted a random sample of 75 people to try out a new weight loss drug. There weight loss was calculated after 6 months. The data below was collected.

	Received new drug	Received Placebo
n	45	30
Mean weight loss	57.6	52.6
Sample standard deviation	9.75	9.27

Is there sufficient evidence to conclude that the people who received the new drug had a higher mean weight loss? Use the classical approach

Inference About 2 Means: Independent Samples

Researchers conducted a random sample of 75 people to try out a new weight loss drug. There weight loss was calculated after 6 months. The data below was collected.

	Received new drug	Received Placebo
n	45	30
Mean weight loss	57.6	52.6
Sample standard deviation	9.75	9.27

Is there sufficient evidence to conclude that the people who received the new drug had a higher mean weight loss? Use the classical approach

Conditions are met (see page 555)

Step 1: Determine the null/alternative hypothesis

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 > \mu_2$$

Step 2: Select the level of significance

None was given so use α=.05

Step 3: Calculate test statistic

$$t_o = \frac{\overline{x_1} - \overline{x_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{57.6 - 52.6 - (0)}{\sqrt{\frac{9.75^2}{45} + \frac{9.27^2}{30}}} = 2.412$$

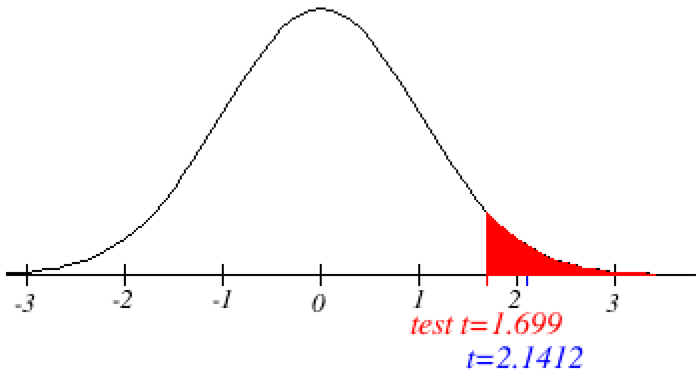
Degrees of freedom: Smaller of n_1-1 and n_2-1. DF=29

Find the critical value: $t_\alpha = t_{.05}\ with\ d.f. =$ 29 is 1.699 (From table).

Continues

Inference About 2 Means: Independent Samples

Step 4: Reject if $t_o > t_\alpha$ (in the rejection region)
2.412>1.699 so reject the null hypothesis
(the test statistic is in the rejection region)



Step 5: State the conclusion

Reject the null hypothesis. There is sufficient evidence that the people who took the new drug had a higher mean weight loss.

Inference about 2 population standard deviations

A researcher wanted to know if students who plan on applying for scholarships had more variability on the Accuplacer reading score than students who do not plan on applying for scholarships. The results of a random sample are found below. Do students who plan on applying for scholarships have a lower standard deviation on the reading Accuplacer score than those that are not going to apply at the α=.1 level. Accuplacer scores follow a normal distribution. Use the classical approach

plans on applying for scholarships	does not plan on applying for scholarships
$n_1 = 33$	$n_2 = 39$
$s_1 = 8.3$	$s_2 = 13.2$

Inference about 2 population standard deviations

A researcher wanted to know if students who plan on applying for scholarships had more variability on the Accuplacer reading score than students who do not plan on applying for scholarships. The results of a random sample are found below. Do students who plan on applying for scholarships have a lower standard deviation on the reading Accuplacer score than those that are not going to apply at the α=.1 level. Accuplacer scores follow a normal distribution. Use the classical approach

plans on applying for scholarships	does not plan on applying for scholarships
$n_1 = 33$	$n_2 = 39$
$s_1 = 8.3$	$s_2 = 13.2$

Conditions:

Independent simple random sample? Yes

Populations are normally distributed? Yes

Step 1: Determine the null/alternative hypothesis

$H_0: \sigma_1 = \sigma_2$

$H_1: \sigma_1 < \sigma_2$

Step 2: Select the level of significance. α=.1 was given

Continues

Inference about 2 population standard deviations

Step 3: Calculate the test statistic

$$F_0 = \frac{s_1^2}{s_2^2} = \frac{8.3^2}{13.2^2} = .395$$

Find the critical value

Left tailed critical value is $F_{1-\alpha,n_1-1,n_2-1} = F_{.9,32,38}$. However 32 and 38 are not in the table so we have to use the closest values which are 25 and 50. Then $F_{.1,25,50} = 1.53$ so that $F_{.9,25,50}=\frac{1}{1.53} = .653$ (see page 568)

Step 4: Compare test and critical values reject if $F_0 < F_{1-\alpha,n_1-1,n_2-1}$

.395<.653 so reject H_0

Step 5: State the conclusion

Reject the null hypothesis. There is sufficient evidence to suggest that those who plan on applying for scholarships have a smaller variability in Accuplacer reading score than whose who do not plan on applying for scholarships.