# Warm Up

1. The Null hypotheses of $H_0: \mu_1 - \mu_2 = 0$ could also be written as $H_0: \mu_1 = \mu_2$.
   As far as the alternative hypotheses $H_a: \mu_1 - \mu_2 < 0$ what is another way to write $H_a$?

2. What is the critical value for a 99% Confidence interval for a difference between two proportions?

$H_0: \mu_1 - \mu_2 = 0$ $x - y = 0$

$H_a: \mu_1 - \mu_2 < 0$ $x - y < 0$

$H_a: \mu_1 < \mu_2$

## Two-Sample z Interval for a Difference Between Two Proportions

When the conditions are met, a $C\%$ confidence interval for $p_1 - p_2$ is

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

where $z^*$ is the critical value for the standard Normal curve with $C\%$ of the area between $-z^*$ and $z^*$.


Go the Tables but use .005 not .01

## Note:

For AP Exam and for tomorrow's test

If you are ever asked just to state the null and alternative hypotheses in a question, be sure to also define your parameters.

## Note #2

You have been using technology a lot lately to do CI's and significance tests. On multiple choice questions you will asked to show an understanding of the formulas.

## Test Advice #3

Only calculate combined proportions if you are running a significance test (not a CI) for a difference in proportions.

# Significance Tests for $p_1 - p_2$

A significance test begins by assuming that $H_0: p_1 - p_2 = 0$ is true. In that case, $p_1 = p_2$. We call the common value of these two parameters $p$.

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}}}$$

Unfortunately, we don't know the common value of $p$. To estimate $p$, we combine (or "pool") the data from the two samples as if they came from one larger sample.

$$\hat{p} = \frac{\text{number of successes in both samples combined}}{\text{number of individuals in both samples combined}} = \frac{X_1 + X_2}{n_1 + n_2}$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}}$$

# Warm Up #2

Finish #3 and #4 from yesterday's handout

Proportions (z) or means (t) ?

Estimate (CI) or a Test (Signif. Test) ?

one sample or two ?
- Unpaired ?
- Paired ?

A test prep course takes a random sample of $50$ its customers. Researchers score each customer on a diagnostic test before taking the course and again on a similar test after taking the course. They want to use these results to estimate the average difference between the before and after scores.

**Which of these inference procedures is most appropriate?**

3

- (A) A $t$-interval for slope
- (B) A paired $t$-interval for the mean difference
- (C) A two-sample $z$-interval for the difference of proportions
- (D) A $z$-interval for a proportion
- (E) A two-sample $t$-interval for the difference of means

A test prep course takes a random sample of $50$ its customers. Researchers score each customer on a diagnostic test before taking the course and again on a similar test after taking the course. They want to use these results to estimate the average difference between the before and after scores.

**Which of these inference procedures is most appropriate?**

3

- (A) A $t$-interval for slope
- (B) A paired $t$-interval for the mean difference
- (C) A two-sample $z$-interval for the difference of proportions
- (D) A $z$-interval for a proportion
- (E) A two-sample $t$-interval for the difference of means

Proportions (z) or means (t) ?

Estimate (CI) or a Test (Signif. Test) ?

one sample or two ?
- Unpaired ?
- Paired ?

A test prep course takes a random sample of 50 its customers. Researchers score each customer on a diagnostic test before taking the course and again on a similar test after taking the course. They want to use these results to estimate the average difference between the before and after scores.

CI mean Paired 3

**Which of these inference procedures is most appropriate?**

- (A) A $t$-interval for slope
- (B) A paired $t$-interval for the mean difference
- (C) A two-sample $z$-interval for the difference of proportions
- (D) A $z$-interval for a proportion
- (E) A two-sample $t$-interval for the difference of means

Lylah created an app, and she recently updated the app. She randomly samples a group of users to estimate what percentage of all users are using the updated version.

**Which of these inference procedures is most appropriate?**

4

**Choose 1 answer:**

- (A) A $t$-test for a mean
- (B) A $z$-interval for a proportion
- (C) A paired $t$-test for the mean difference
- (D) A $z$-test for a proportion
- (E) A $t$-interval for a mean

INCORRECT

A $t$-test for a mean

Lylah wants to *estimate*, so she should use an *interval*, not a test. She's tallying a *categorical* variable, so looking at a *proportion* is more appropriate than a mean.

---

CORRECT (SELECTED)

A $z$-interval for a proportion

Lylah wants to *estimate*, so she should use an *interval*. She's tallying a *categorical* variable, so looking at a *proportion* is appropriate.

---

INCORRECT

A paired $t$-test for the mean difference

Lylah collected one data point (whether or not they have the updated version) for each user, so she doesn't have paired data. Also, this data is *categorical*, so looking at a *proportion* is more appropriate than a mean.

---

INCORRECT

A $z$-test for a proportion

Lylah wants to *estimate*, so she should use an *interval*, not a test.

---

INCORRECT

A $t$-interval for a mean

Lylah is tallying a *categorical* variable, so looking at a *proportion* is more appropriate than a mean.

I need to get a count of how many of you will be signing up to take the AP Statistics exam

(I need to let Voss know)

- [ ] Definitely 12
- [ ] Not sure 4
- [ ] prob. not

See your LCQ's

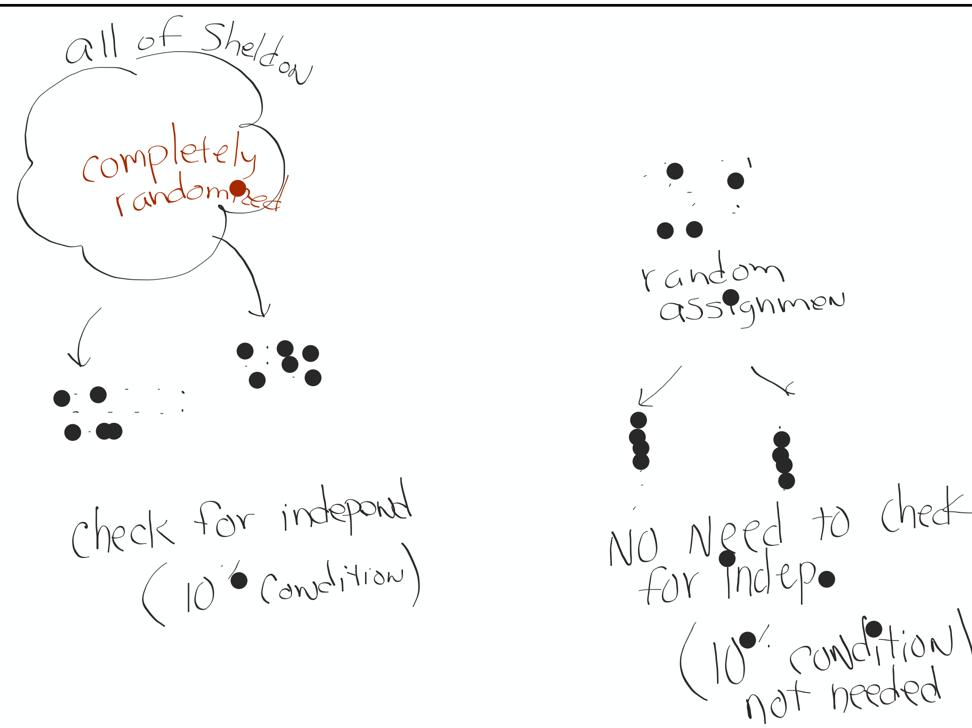
Preparing

Summary of formulas

Frappy !

Ch. 10 Review Exercises

Ch. 10 Practice Test



| Lesson | 10.1 – Difference of Proportions | 10.2 – Difference of Means | 10.3 – Mean Difference |
|---|---|---|---|
| Symbol for parameter (population) | $p_1-p_2 \rightarrow$ The diff. in proportions | $\mu_1-\mu_2 \rightarrow$ The difference in means | $\mu_{diff} \rightarrow$ The mean difference |
| Symbol for statistic (sample) | $\hat{p}_1-\hat{p}_2$ | $\bar{x}_1-\bar{x}_2$ | $\bar{x}_{diff}$ |
| Name the procedure | Two Sample Z _ for $p_1-p_2$ | Two Sample t _ for $\mu_1-\mu_2$ | One Sample t _ for $\mu_{diff}$ |
| Conditions | Random<br>10%<br>Normals<br>$n_1p_1$, $n_1(1-p_1)$, $n_2p_2$, $n_2(1-p_2) \geq 10$ | ① Pop. is Normal<br>② $n \geq 30$ CLT<br>③ No strong skew or outliers<br>*True for BOTH groups | ① Pop is Normal<br>② $n \geq 30$ CLT<br>③ No strong skew or outliers.<br>*Only needed for difference |
| Formula for mean of the sampling distribution | $\mu_{\hat{p}_1-\hat{p}_2} = p_1-p_2$ | $\mu_{\bar{x}_1-\bar{x}_2}=\mu_1-\mu_2$ | $\mu_{\bar{x}_{diff}}=\mu_{diff}$ |
| Formula for standard deviation of the sampling distribution | $\sigma_{\hat{p}_1-\hat{p}_2}=\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}$ | $\sigma_{\bar{x}_1-\bar{x}_2}=\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$ | $\sigma_{\bar{x}_{diff}}=\frac{s_{diff}}{\sqrt{n}}$ |
| Formulas for Confidence Intervals | $(\hat{p}_1-\hat{p}_2) \pm z^*\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ | $(\bar{x}_1-\bar{x}_2)\pm t^*\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$ | $\bar{x}_{diff} \pm t^*\frac{s_{diff}}{\sqrt{n}}$ |
| Formulas for Significance Tests | $z=\frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\frac{\hat{p}_c(1-\hat{p}_c)}{n_1}+\frac{\hat{p}_c(1-\hat{p}_c)}{n_2}}}$ | $t=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$ | $t=\frac{\bar{x}_{diff}-\mu_{diff}}{\frac{s_{diff}}{\sqrt{n}}}$ |
| How to find P-value | Table A or normcdf | Table B or tcdf | Table B or tcdf |

Researchers were studying how playing a dancing video game impacts heart rate. They measured the heart rates (in beats per minute) of $15$ subjects before they danced a song and again after they finished dancing the song. They want to use these results to estimate the average difference between before and after heart rates.

**Which of these inference procedures is most appropriate?**

**Choose 1 answer:**

(A) A $z$-interval for a proportion

(B) A two-sample $t$-interval for the difference of means

(C) A paired $t$-interval for the mean difference

(D) A two-sample $z$-interval for the difference of proportions

(E) A $t$-interval for slope

INCORRECT

A $z$-interval for a proportion

The researchers aren't categorizing the heart rates, so proportions wouldn't be appropriate.

INCORRECT

A two-sample $t$-interval for the difference of means

The before heart rates are **not** independent of the after heart rates, so we shouldn't treat them as two separate samples.

CORRECT (SELECTED)

A paired $t$-interval for the mean difference

The researchers recorded two measurements on each subject. They should calculate the difference between the two heart rates for each subject, and do a test on the mean of those differences.

f

February 05, 2019

A website streams movies and television shows to millions of users. Employees know that the average time a user spends per session on their website is $2$ hours. The website changed its design, and they wanted to know if the average session length was longer than $2$ hours. They randomly sampled $100$ users and recorded their session lengths.

**Which of these inference procedures is most appropriate?**

**Choose 1 answer:**

- (A) A paired $t$-test for the mean difference
- (B) A two-sample $t$-test for the difference of means
- (C) A $t$-test for a mean
- (D) A $z$-test for a proportion
- (E) A two-sample $z$-test for the difference of proportions

---

INCORRECT

A paired $t$-test for the mean difference

The employees collected one data point (the session length) for each user in the sample, so they don't have paired data.

INCORRECT

A two-sample $t$-test for the difference of means

The employees are looking at one sample of data, not two.

CORRECT (SELECTED)

A $t$-test for a mean

The employees are interested in the *average* session length, so $t$ procedures for a *mean* are appropriate. They are comparing the mean of a single sample to a hypothesized value, so two-sample procedures aren't appropriate.

Felipe is curious if there is a relationship between a runner's age and their finishing time in a recent marathon. He takes a random sample of finishers and records the age (in years) and the finishing time (in minutes) for each of those sampled.

**Which of these inference procedures is most appropriate?**

**Choose 1 answer:**

(A) A two-sample $z$-test for the difference of proportions

(B) A $z$-test for a proportion

(C) A paired $t$-test for the mean difference

(D) A $t$-test for slope

(E) A two-sample $t$-test for the difference of means

$150$ people in the $30$-$39$ age group about
e the difference between the percentage of

A $t$-interval for slope

A two-sample $z$-interval for the difference of proportions

A $z$-interval for a proportion

A two-sample $t$-interval for the difference of means

A $t$-interval for a mean

INCORRECT

A $t$-interval for slope

This type of interval is useful for estimating the slope of a regression line, but Angelica isn't looking at the relationship between two quantitative variables.

CORRECT (SELECTED)

A two-sample $z$-interval for the difference of proportions

Angelica has two groups ($150$ people from each age group) and she's comparing a *categorical* variable (vegetarian or not) between the two groups, so *proportions* are appropriate.

INCORRECT

A $z$-interval for a proportion

Angelica's data came from two groups ($150$ people from each age group), so two-sample procedures are appropriate.

A two-sample $t$-interval for the difference of means

Angelica is looking at a *categorical* variable, so using *proportions* is more appropriate than means.

→

INCORRECT

A $t$-interval for a mean

Angelica is looking at a *categorical* variable, so using *proportions* is more appropriate than means. Also, she has two groups, so two-sample procedures are appropriate.

INCORRECT

A $z$-test for a proportion

The employees are interested in the *average* session length, so they should use $t$ procedure for a *mean*. They aren't categorizing the lengths, so proportions wouldn't be appropriate.

INCORRECT

A two-sample $z$-test for the difference of proportions

The employees are looking at one sample of data, not two. Also, they are interested in the *average* session length, so they should use $t$ procedures for a *mean*. They aren't categorizin the lengths, so proportions wouldn't be appropriate.

INCORRECT

A two-sample $z$-interval for the difference of proportions

The researchers aren't categorizing the heart rates, so proportions wouldn't be appropriate.

INCORRECT

A $t$-interval for slope

This would be useful if the researchers were interested in the relationship between before and after heart rates in a scatter plot, but it wouldn't be the best way to estimate the average difference.

CORRECT (SELECTED)

A paired $t$-interval for the mean difference

The researchers recorded two measurements on each customer. They should calculate the difference between the two scores for each customer, and do a test on the mean of those differences.

INCORRECT

A two-sample $z$-interval for the difference of proportions

The researchers aren't categorizing the scores, so proportions wouldn't be appropriate.

INCORRECT

A $z$-interval for a proportion

The researchers aren't categorizing the scores, so proportions wouldn't be appropriate.

INCORRECT

A two-sample $t$-interval for the difference of means

The before scores are **not** independent of the after scores, so we shouldn't treat them as two separate samples.

A school counselor suspects that, on average, students at their school are sleeping less than $8$ hours per night. They survey a random sample of students about how many hours they slept the previous night to see if their average sleep amount is significantly less than $8$ hours.

**Which of these inference procedures is most appropriate?**

**Choose 1 answer:**

(A) A two-sample $z$-test for the difference of proportions

(B) A paired $t$-test for the mean difference

(C) A $z$-test for a proportion

(D) A $t$-test for a mean

(E) A two-sample $t$-test for the difference of means

CORRECT (SELECTED)

A $t$-test for a mean

The counselor is interested in the *average* sleep amount, so $t$ procedures for a *mean* are appropriate. They are comparing the mean of a single sample to a hypothesized value, so two-sample procedures aren't appropriate.

INCORRECT

A two-sample $z$-test for the difference of proportions

The counselor is looking at one sample of data, not two. Also, they are interested in the *average* sleep amount, so they should use $t$ procedures for a *mean*. They aren't categorizing the sleep amounts, so proportions wouldn't be appropriate.

INCORRECT

A paired $t$-test for the mean difference

The counselor collected one data point (the sleep amount) for each student in the sample, so they don't have paired data.

INCORRECT

A $z$-test for a proportion

The counselor is interested in the *average* sleep amount, so they should use $t$ procedures for a *mean*. They aren't categorizing the sleep amounts, so proportions wouldn't be appropriate.

INCORRECT

A two-sample $t$-test for the difference of means

The counselor is looking at one sample of data, not two.