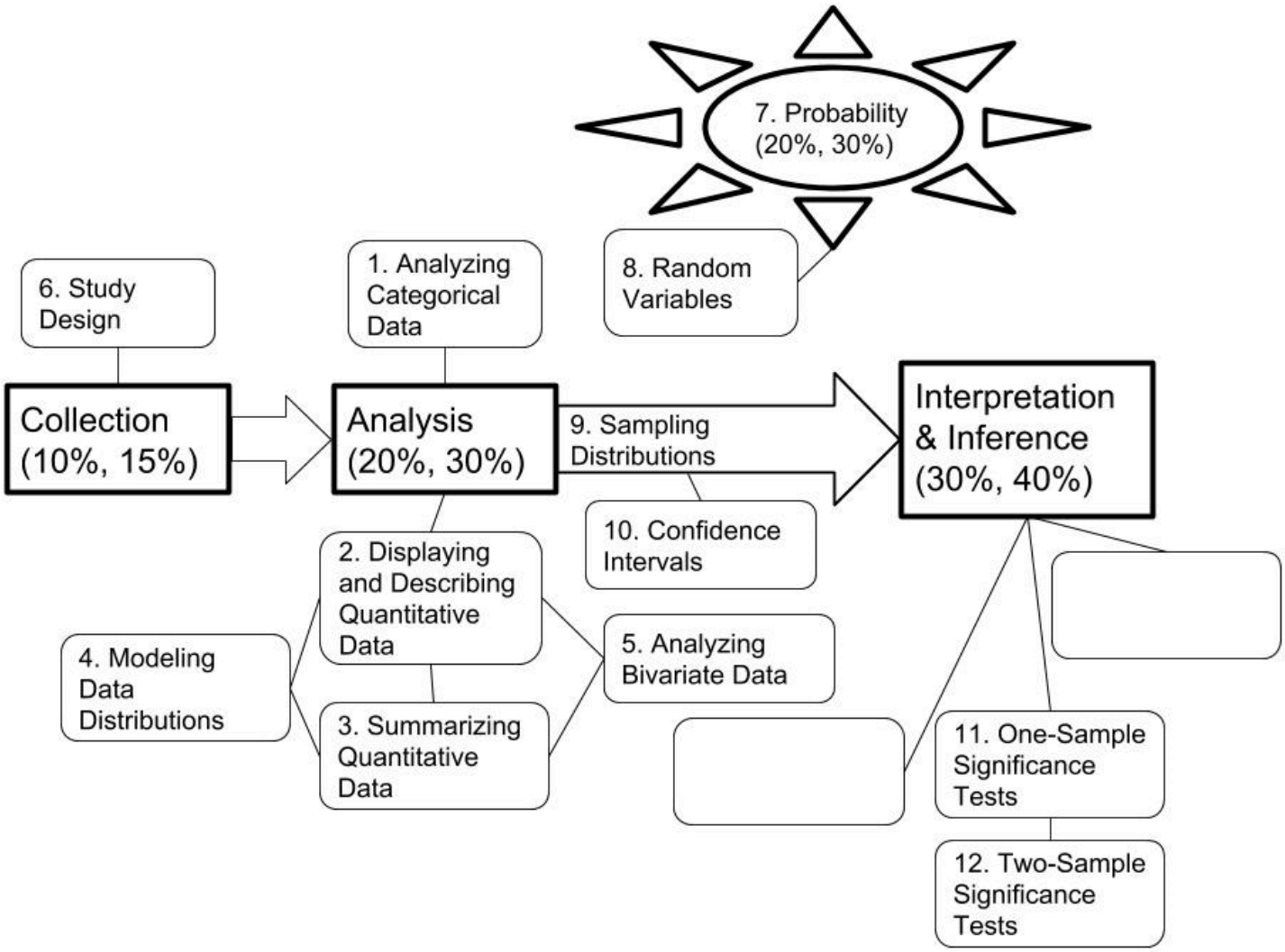
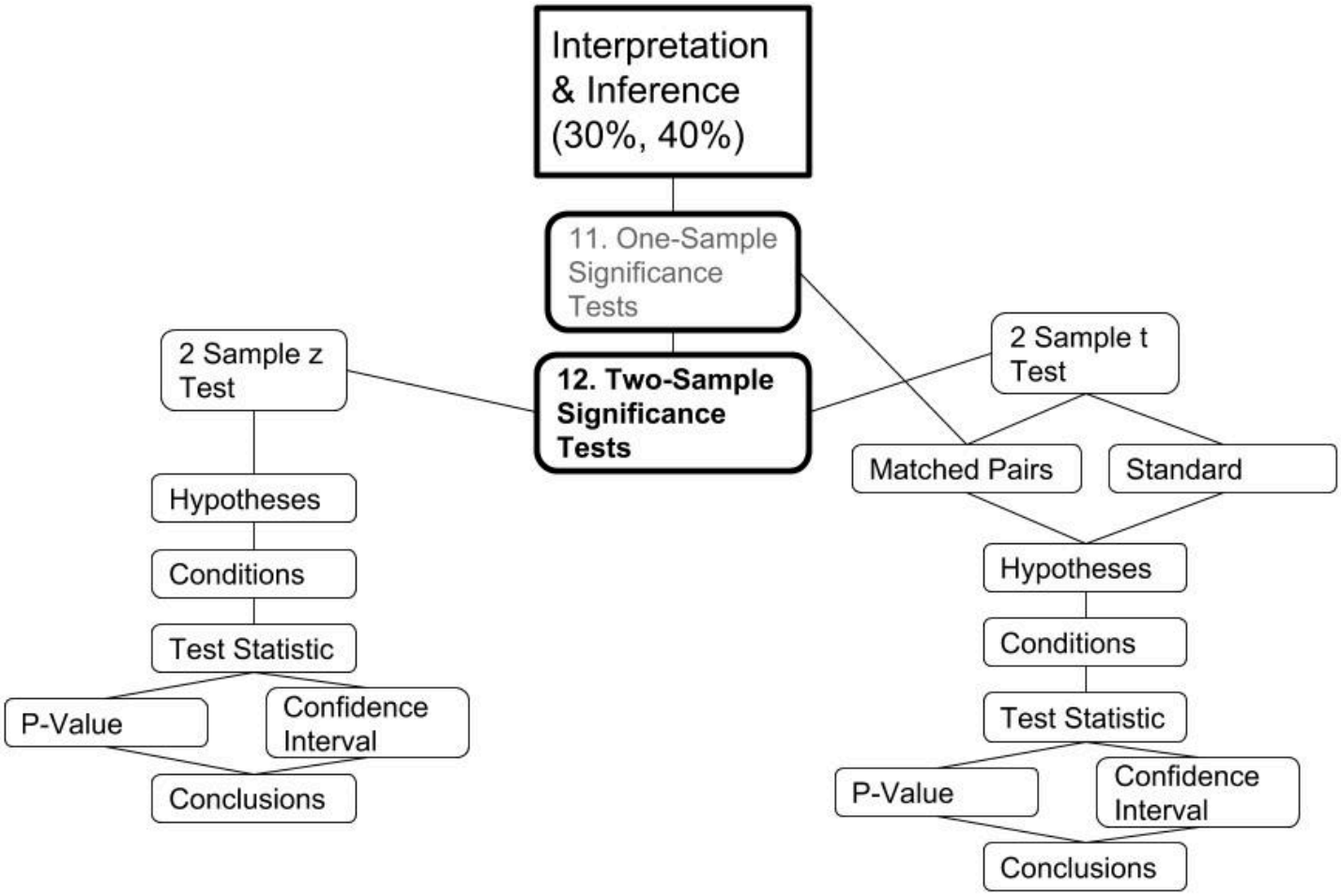
Unit 12: Two-Sample Significance Tests

Unit Objectives

- Test hypotheses in a statistically valid manner





Unit 12 Lesson 01: Check the Conditions for a Two-Sample z Test for the Difference of Proportions

What are the conditions for two-sample z tests for the difference of proportions, and how do we check them?

- Check the three conditions necessary for a two-sample z test for the difference of proportions

Conditions
Same as one sample z tests about a proportion, just for both samples.

1. _____

2. _____

3. _____

1. [Source: Khan Academy] A US market researcher was curious if there was a difference in electric vehicle ownership rates between people living on the East coast versus those on West coast. They obtained separate random samples of 500 car owners from each region. They found that 52 in the West coast sample owned an electric car compared to 38 in the East coast sample. They plan to use these results to do a two-sample z test to see if these results show a significant difference. Which conditions for this type of test did their samples meet? Justify your answers.

   (A) Both samples were randomly selected from their populations.

   (B) The expected counts of successes and failures are sufficiently large in each sample.

   (C) Individual observations in each sample can be considered independent, and the samples themselves are independent.

2. [Source: Khan Academy] A medical researcher suspects that there is a difference between the proportions of men and women that are diagnosed with high blood pressure. They want to sample people and create a two-sample z interval to estimate the difference. Which of the following are conditions for this type of interval? Justify your answers.

   (A) The men and women within each sample are independent of each other.

   (B) They take separate random samples of men and women.

   (C) They sample an equal number of men and women.

3. [Source: Khan Academy] Trevor thinks that left-handed pitchers in a baseball league are more likely to strike out batters than right-handed pitchers are. He's considering each "at-bat" as an opportunity for a strikeout. He takes a random sample of at-bats where the pitcher was left-handed, and he records the proportion of those at-bats that result in a strikeout. He repeats this process for at-bats with right-handed pitchers. He wants to use the results to test if the proportion of at-bats resulting in a strikeout is significantly higher for left-handed pitchers. Which of the following is an appropriate set of hypotheses for Trevor's significance test?

   (A) $H_0 : p_R - p_L = 0$ $H_a : p_R - p_L > 0$
   (B) $H_0 : p_L - p_R = 0$ $H_a : p_L - p_R < 0$
   (C) $H_0 : p_L - p_R = 0$ $H_a : p_L - p_R > 0$
   (D) $H_0 : p_R - p_L = 0$ $H_a : p_R - p_L \neq 0$

~~~U12L01 Homework~~~

1. Inference comparing two groups or populations: Writing hypotheses for testing the difference of proportions
2. Inference comparing two groups or populations: Conditions for inference on two proportions

~~~U12L01 Classwork~~~

1. [Source: Khan Academy] Tatiana wonders if the same proportion of teens and adults check social media at least once per day. She wants to obtain a random sample of people from each group to test if there is a significant difference between the proportion of teens and adults that check social media at least once per day. Which of the following is an appropriate set of hypotheses for Tatiana's significance test?

   (A) $H_0 : p_T - p_A = 0$ $\quad H_a : p_T - p_A < 0$

   (B) $H_0 : p_A - p_T = 0$ $\quad H_a : p_A - p_T < 0$

   (C) $H_0 : p_T - p_A = 0$ $\quad H_a : p_T - p_A \neq 0$

   (D) $H_0 : p_A - p_T \neq 0$ $\quad H_a : p_A - p_T = 0$

2. [Source: Khan Academy] A sociologist suspects that men are more likely to have received a ticket for speeding than women are. The sociologist wants to sample people and create a two-sample z interval to estimate the difference between the proportion of men who have received a speeding ticket and the proportion of women who have received a speeding a ticket. Which of the following are conditions for this type of interval? Justify your answers.

   (A) They take separate random samples of men and women.

   (B) The people in each sample can be considered independent.

   (C) The samples both include at least 10 people who have received a speeding ticket and least 10 people who haven't.

3. **AP**: A physician believes that adults in California develop skin cancer at higher rates than adults in other states because of California's abundance of sunny days. To study this she gathers information on the number of hours of sun exposure per week from a random sample of Californians and a random sample of Americans from other parts of the country. Which of the following might be an appropriate null hypothesis for this study?
    (A) The average number of hours of sun exposure per week for adult Californians is different from the average number of hours of sun exposure per week for adults in the rest of America.
    (B) The average number of hours of sun exposure per week for adult Californians is the same as the average number of hours of sun exposure per week for adults in the rest of America.
    (C) The average number of hours of sun exposure per week for adult Californians is greater than the average number of hours of sun exposure per week for adults in the rest of America.
    (D) The average number of hours of sun exposure per week for adult Californians is less than the average number of hours of sun exposure per week for adults in the rest of America.
    (E) The probability is 0.5 than an adult from California and an adult from the rest of America receive an equal number of hours of sun exposure per week.

4. **AP**: The 2020 election in the United States features legalized marijuana on nearly all of the Democratic candidates' platforms. A pollster is interested in comparing the proportions of Democrats and Republicans in favor of legalization. The pollster plans to test the hypothesis that the proportion of Democrats in favor of legalization is different from the proportion of Republicans in favor of legalization. There are 141 million registered Democrats and 184 million registered Republicans in the United States. From a simple random sample of 50 registered democrats, the pollster finds that 42 are in favor of legalization. From an independent simple random sample of 60 registered republicans, the pollster finds that 29 are in favor of legalization. Which of the following statements is true about this situation?
    (A) Because the samples are from normal populations, a two-proportion z-test would be valid.
    (B) Because the size of each sample is greater than 30, a two-proportion z-test would be valid.
    (C) Because the number who favored the ban is greater than 10 in both groups, a two proportion z-test would be valid.
    (D) Because of the relative sizes of the populations and samples, a two-proportion z-test would be valid.
    (E) A two-proportion z-test would not be valid for these data.

Unit 12 Lesson 02: Calculate a Two-Sample z Interval for the Difference of Proportions

What's the difference between calculating a one-sample z interval and a two-sample z interval regarding proportions?

- Return to confidence intervals--calculate the values for the specified interval for two-sample proportions

1. [Source: Khan Academy] Marta created an app, and she suspected that teens were more likely to use it than adults. She obtained separate random samples of teens and adults. Here are the results:

Use app?	Teens	Adults
Yes	54	26
No	66	104
Total	120	130

She wants to use these results to construct a 95% confidence interval to estimate the difference between the proportions of teens and adults that use her app (p_T - p_A). Assume that all of the conditions for inference have been met. Which of the following is a correct 95% confidence interval based on her data? And draw a little diagram.

(A) 28 ± 0.148
(B) 0.25 ± 0.057
(C) 0.25 ± 0.094
(D) 0.25 ± 0.112
(E) 0.25 ± 0.148

2. **AP**: About 35% of American adults are obese. As part of a national obesity study, a random sample of adults was selected and surveyed about their dietary habits and their weight. Of the 192 adults who did not regularly consume soda (nondrinkers), 21 percent of them were obese. Of the 154 adults who did regularly consume soda, 49 percent were obese. Which of the following is the most appropriate standard error for a confidence interval for the difference in proportions of adults who are obese among soda drinkers and nondrinkers?

(A) $\sqrt{\frac{(0.21)(0.79)}{192} + \frac{(0.49)(0.51)}{154}}$

(B) $\sqrt{\frac{(0.21)(0.79) + (0.49)(0.51)}{192 + 154}}$

(C) $\sqrt{(0.33)(0.67)\left(\frac{1}{192} + \frac{1}{154}\right)}$

(D) $\sqrt{(0.5)(0.5)\left(\frac{1}{192} + \frac{1}{154}\right)}$

(E) $\sqrt{(0.5)(0.5)\left(\frac{1+1}{192+154}\right)}$

3. [Source: Khan Academy] A member of congress is curious how voters in two major political parties support a certain policy change. They obtained separate random samples of voters from each party. Here are the results:

Supports policy change?	Democrat	Republican
Yes	80	150
No	320	450
Total	400	600

They want to use these results to construct a 99% confidence interval to estimate the difference in the proportion of voters in these parties who support the policy change (p_R - p_D). Assume that all of the conditions for inference have been met. Which of the following is a correct 99% confidence interval based on these samples?

(A) (-0.019, 0.119)
(B) (-0.002, 0.102)
(C) (0.006, 0.094)
(D) (0.023, 0.077)

~~~U12L02 Homework~~~ 1. Inference comparing two groups or populations: Two-sample z interval for the difference of proportions

~~~U12L02 Classwork~~~

1. **AP**: In 1960 sociologists studied a random sample of 1,552 adults and found that 72 percent of adults were married. In 2017 the study was replicated with a random sample of 1,529 adults found that 50 percent of adults were married. Which of the following represents a 99 percent confidence interval for the difference between the proportions of adults from 1960 to 2017 that were married?

(A) $(0.72 - 0.50) \pm 1.96\sqrt{\frac{(0.72)(0.28)}{1,552} + \frac{(0.50)(0.50)}{1,529}}$

(B) $(0.72 - 0.50) \pm 2.326\sqrt{\frac{(0.72)(0.28)}{1,552} + \frac{(0.50)(0.50)}{1,529}}$

(C) $(0.72 - 0.50) \pm 2.576\sqrt{\frac{(0.72)(0.28)}{1,552} + \frac{(0.50)(0.50)}{1,529}}$

(D) $(0.72 - 0.50) \pm 2.326\sqrt{\frac{1,882}{3,079}\left(\frac{1}{1,552} + \frac{1}{1,529}\right)}$

(E) $(0.72 - 0.50) \pm 2.576\sqrt{\frac{1,882}{3,079}\left(\frac{1}{1,552} + \frac{1}{1,529}\right)}$

2. [Source: Khan Academy] A sociologist was curious about how much smartphone access has changed for American teenagers. They obtained separate random samples of teenagers from 2017 and 2013. Here are their results:

Has access to smartphone?	2017	2013
Yes	450	120
No	150	280
Total	600	400

They want to use these results to construct a 95% confidence interval to estimate the difference in the proportion of teens each year who had access to a smartphone (p_{2017} - p_{2013}). Assume that all of the conditions for inference have been met. Which of the following is a correct 95% confidence interval based on their data?

(A) 0.45 ± 0.029
(B) 0.45 ± 0.048
(C) 0.45 ± 0.057
(D) 330 ± 0.048
(E) 330 ± 0.057

Unit 12 Lesson 03: Calculate the Test Statistic in a Two-Sample z Test for the Difference of Proportions

What is a "test statistic", and how does one calculate it?

- Calculate test statistics for differences of proportions

1. [Source: Khan Academy] A large school district was considering changing the start times to its schools. The change would involve high schools starting later and elementary schools starting earlier. The school district was curious if parents supported this change, so they surveyed separate random samples of parents from both types of school in the district. Here are their results:

Supports change?	High school	Elementary
Yes	54	24
No	46	26
Total	100	50

The school district wants to use these results to carry out a two-sample z test to determine if the proportion of parents who support the change is significantly different for each type of school. Assume that all conditions have been met. Calculate the test statistic for the district's test.

2. [Source: Khan Academy] Caroline thinks she can flip a coin so it lands showing heads more often with her right hand. She flipped a coin 40 times with each hand. Here the results:

Result	Right	Left
Heads	26	18
Tails	14	22
Total	40	40

Caroline wants to use these results to test $H_0 : p_R - p_L = 0$ versus $H_a : p_R - p_L > 0$. Assume that all conditions have been met. Calculate the test statistic for her test.

~~~U12L03 Homework~~~ 1. Inference comparing two groups or populations: Test statistic in a two-sample z test for the difference of proportions

~~~U12L03 Classwork~~~

1. [Source: Khan Academy] A political consultant wondered if support for a candidate was significantly different between voters who owned a landline telephone and voters who did not. The consultant surveyed a random sample of voters and found 36 of 75 who owned a landline supported the candidate, while 20 of 100 who did not own a landline supported the candidate. The consultant wants to test if these results suggest a significant difference in support between voters who do and do not own a landline. Assume that all conditions have been met. Calculate the test statistic for the consultant's test.

2. [Source: Khan Academy] Researchers studying osteoporosis (bone loss) suspected that women over the age of 50 in the United States are diagnosed with the disease more often than women over 50 in Mexico. They took a random sample of 200 women over the age of 50 from each country. Here are the results:

Has osteoporosis?	US	Mexico
Yes	40	20
No	160	180
Total	200	200

The researchers want to use these results to test $H_0 : p_{US} - p_M = 0$ versus $H_a : p_{US} - p_M > 0$. Assume that all conditions have been met. Which of the following would be an appropriate test statistic for their test? Calculate the test statistic for the researchers' test.

Unit 12 Lesson 04: Find the P-Value in a Two-Sample z Test for the Difference of Proportions

What if I want to go one step further, and find the probability that a certain test statistic would occur?

- Find P-values for z tests for differences of proportions

1. [Source: Khan Academy] An economist was curious if women were more satisfied with their jobs than men. A random sample of 220 workers showed that 46 of 100 women were satisfied with their jobs, and 42 of 120 men were satisfied. The economist wants to use these results to test $H_0 : p_W - p_M = 0$ versus $H_a : p_W - p_M > 0$. Assume that all conditions have been met. What is the P-value associated with these sample results?
   (A) P-value < 0.01
   (B) 0.01 ≤ P-value < 0.05
   (C) 0.05 ≤ P-value < 0.10
   (D) 0.10 ≤ P-value < 0.20
   (E) P-value ≥ 0.20

2. [Source: Khan Academy] A political consultant wondered if support for a candidate was significantly different between voters who owned a landline telephone and voters who did not. The consultant surveyed a random sample of voters and found 36 of 75 who owned a landline supported the candidate, while 27 of 100 who did not own a landline supported the candidate. The consultant wants to test if these results suggest a significant difference in support between voters who do and do not own a landline. They will test $H_0 : p_{\text{Landline}} = p_{\text{No landline}}$ versus $H_a : p_{\text{Landline}} \neq p_{\text{No landline}}$. Assume that all conditions have been met. What is the P-value associated with these sample results?
   (A) P-value < 0.01
   (B) 0.01 ≤ P-value < 0.05
   (C) 0.05 ≤ P-value < 0.10
   (D) 0.10 ≤ P-value < 0.20
   (E) P-value ≥ 0.20

~~~U12L04 Homework~~~ 1. Inference comparing two groups or populations: P-value in a two-sample z test for the difference of proportions

~~~U12L04 Classwork~~~

1. [Source: Khan Academy] An economist was comparing unemployment rates between states. A random sample of 800 adults in Wisconsin showed that 20 were unemployed, and a random sample of 800 adults in Minnesota showed that 28 were unemployed. The economist want to use these results to test $H_0 : p_{MN} - p_{WI} = 0$ versus $H_a : p_{MN} - p_{WI} > 0$. Assume that all conditions have been met. What is the P-value associated with these sample results?
   (A) P-value < 0.01
   (B) 0.01 ≤ P-value < 0.05
   (C) 0.05 ≤ P-value < 0.10
   (D) 0.10 ≤ P-value < 0.15
   (E) P-value ≥ 0.15

2. [Source: Khan Academy] A large school district was considering changing the start times to its schools. The change would involve high schools starting later and elementary schools starting earlier. The school district was curious if parents supported this change, so they surveyed separate random samples of parents from both types of school in the district. Here are their results:

Supports change?	High school	Elementary
Yes	54	24
No	46	26
Total	100	50

   The school district wants to use these results to test if the proportion of parents who support the change is different between these types of schools. They will test $H_0 : p_{HS} = p_E$ 0 versus $H_a : p_{HS} \neq p_E$. Assume that all conditions have been met. What is the P-value associated with these sample results?
   (A) P-value < 0.01
   (B) 0.01 ≤ P-value < 0.05
   (C) 0.05 ≤ P-value < 0.10
   (D) 0.10 ≤ P-value < 0.15
   (E) P-value ≥ 0.15

Unit 12 Lesson 05: Draw Conclusions about Differences of Proportions

Now that I've run my test and gotten my P-value, what EXACTLY can I conclude?

- Identify appropriate and inappropriate conclusions resulting from a given two-sample z test

1. **AP**: Phlebotomists in California and Maine routinely draw blood from volunteers for use in hospitals. Phlebotomists investigating blood type obtained random samples of patient medical records and recorded the blood type of each patient in the sample. A test was conducted of $H_0 : p_1 = p_2$ versus $H_a = : p_1 \neq p_2$, where p_1 represents the proportion of all patients with Rh positive blood in California and p_2 represents the proportion of all patients with Rh positive blood in Maine. The resulting test statistic for a two-sample z-test for a difference between proportions was 1.84. At the significance level $\alpha = 0.05$ which of the following is a correct conclusion?
   (A) There is not sufficient statistical evidence to conclude that the proportion of all patients with Rh positive blood in California is different from the proportion of all patients with Rh positive blood in Maine because the *p*-value is greater than 0.05.
   (B) There is not sufficient statistical evidence to conclude that the proportion of all patients with Rh positive blood in California is different from the proportion of all patients with Rh positive blood in Maine because the *z*-test statistic is greater than 0.05.
   (C) There is sufficient statistical evidence to conclude that the proportion of all patients with Rh positive blood in California is different from the proportion of all patients with Rh positive blood in Maine because the *p*-value is greater than 0.05.
   (D) There is sufficient statistical evidence to conclude that the proportion of all patients with Rh positive blood in California is different from the proportion of all patients with Rh positive blood in Maine because the *p*-value is less than 0.05.
   (E) There is sufficient statistical evidence to conclude that the proportion of all patients with Rh positive blood in California is greater than the proportion of all patients with Rh positive blood in Maine because the *z*-test statistic is positive.

1. **AP**: A new lubricant was devised that is expected to be better than the older petroleum based lubricant. A large number of randomized experiments were conducted to determine whether the new lubricant can keep a bearing lubricated longer than the old lubricant. For each of the experiments, the lubricants efficacy is the difference between the proportion of bearings with the new lubricant that seized and the proportion of bearings using the old lubricant that seized. If the new lubricant was no better than the old lubricant, which of the following experimental results would be anticipated?
   I. *p*-values will be greater than 0.05 for about 95 percent of the experiments.
   II. There will be about an equal number of experiments showing positive and negative values of lubricant effect.
   III. When 95 percent confidence intervals for the population drug effect are constructed, those confidence intervals include 0 about 95 percent of the time.
   (A) I only
   (B) II only
   (C) III only
   (D) I and II only
   (E) I, II, and III

~~~U12L05 Homework~~~

1. Inference comparing two groups or populations: Making conclusions about the difference of proportions

~~~U12L05 Classwork~~~

1. [Source: Khan Academy] A sociologist took a random sample of 1200 drivers and found that 59 of the 610 men in the sample had received a speeding ticket, while 28 of the 590 women in the sample had received a speeding ticket. The sociologist used those results to make a 99% confidence interval to estimate the difference between the proportion of male and female drivers who have received a speeding ticket (p_M - p_W). The resulting interval was (0.011, 0.087). They want to use this interval to test $H_0 : p_M = p_W$ versus $H_a : p_M \neq p_W$ at the $\alpha = 0.01$ significance level. Assume that all conditions for inference have been met. Based on the interval, what do we know about the corresponding P-value and conclusion at the $\alpha = 0.01$ level of significance?
   (A) The P-value is greater than $\alpha = 0.01$, and they should conclude that there is a difference between the proportions.
   (B) The P-value is greater than $\alpha = 0.01$, and they cannot conclude that there is a difference between the proportions.
   (C) The P-value is less than $\alpha = 0.01$, and they should conclude that there is a difference between the proportions.
   (D) The P-value is less than $\alpha = 0.01$, and they cannot conclude that there is a difference between the proportions.

2. **AP**: In November 2016, California voters approved Proposition 56, which increases the cigarette tax rate. Before the election a poll was conducted across the state to try to predict the election result. A random sample of 240 California residents who were smokers were selected, and another random sample of 274 city residents who were not smokers were selected. Of the non-smokers, 187 approved of the tax increase, and of the smokers 167 approved of the tax increase. An appropriate hypothesis test was conducted to investigate whether there was a difference between the two groups of residents in their approval of the tax increase. Is there convincing statistical evidence of the difference between the two population proportions at the significance level of 0.10?
   (A) Yes, because the sample proportions are different.
   (B) Yes, because the probability of observing a difference at least as large as the sample difference is greater than 0.10.
   (C) Yes, because the probability of observing a difference at least as large as the sample difference, if the two population proportions are the same, is less than 010.
   (D) No, because the probability of observing a difference at least as large as the sample difference, if the two population proportions are the same, is greater than 0.10.
   (E) No, because the probability of observing a difference at least as large as the sample difference is less than 0.10.

3. **AP**: Alice selected a random sample of 100 moviegoers from the 1,400 customers her theater received on a randomly chosen day to investigate preferences for screening of the classics vs. modern films for matinee screenings. The results are shown in the table below.

Preference	Number of Customers
Prefers classic films	59
Prefers modern films	41

Alice incorrectly performed a large sample test of the difference in two proportions using $\frac{59}{100}$ and $\frac{41}{100}$ and calculated a *p*-value of 0.01. Consequently, she concluded that there was a significant difference in preference for the two options. Which of the following best describes his error in the analysis of these data.

(A) No statistical test was necessary because 0.59 is clearly larger than 0.41.
(B) The results of the test were invalid because less than 10% of the population was sampled.
(C) Alice performed a two-tailed test and should have performed a one-tailed test.
(D) A one-sample test for a proportion should have been performed because only one sample was used.
(E) More options should have been included, and a chi-square test should have been performed.

4. **AP**: A survey organization conducted internet interviews in November 2015 in which 1,315 randomly selected adults in the United States responded to the following question.

> At the present time, do you think antioxidants are beneficial your health?

Of the 1,315 adults surveyed, 723 responded "yes." In November 2016, 712 of 1,401 randomly selected adults in the United States had responded "yes" to the same question. Do the data provide convincing evidence that the proportion of adults in the United States who would respond "yes" to the question changed from November 2015 and November 2016?

Unit 12 Lesson 06: Check the Conditions for a Two-Sample t Test for the Difference of Means

What are the conditions for two-sample t tests for the difference of means, and how do we check them?

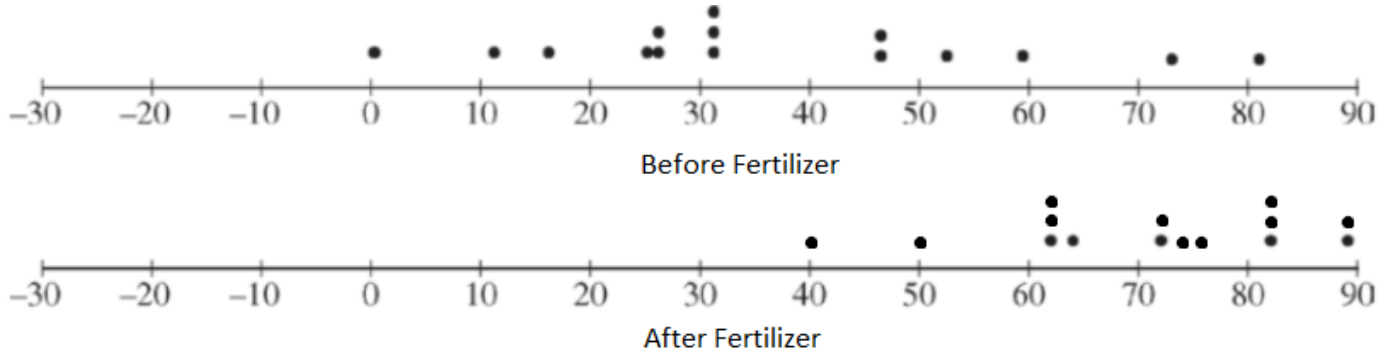
- Check the three conditions necessary for a two-sample t test for the difference of means

1. **AP**: A farmer recorded the percentage growth rate of a sample of his plants every day for a month. The sample came from a population of thousands of plants. He wanted to know how effective a new fertilizer was.

Percent Growth During Month before Fertilizer														
0	11	16	25	26	26	31	31	31	46	46	52	59	73	81

Percent Growth During Month with Fertilizer														
40	50	62	62	62	64	72	72	73	75	83	83	83	89	89

The dotplots below display the distributions of percentage growth before and after fertilizer was applied.



(a) Based on the dotplots, does it appear that the distribution of growth rates changed after fertilizer was applied? Explain.

(b) The farmer wants to perform a 2 sample t-test to determine whether the average percentage growth rate changed after fertilizer was applied. State the conditions for that test. For each condition, comment on whether it appears to be met.

2. [Source: Khan Academy] Eira works for a consumer advocacy firm. She would like to determine whether there is a difference in the battery life of two different brands of batteries, each of which sells thousands of batteries per year. She obtains separate random samples of 12 batteries from one brand and 20 batteries from the other, then tests the amount of time each battery lasts. There are outliers in her data. Eira plans to use this information to perform a *t* test to determine whether there is a difference in the average battery life of the two brands. Which conditions for performing this type of test did Eira's samples meet? Justify your choice for each condition.

(A) Both samples were randomly selected from their populations.

(B) Both samples satisfy the normal condition.

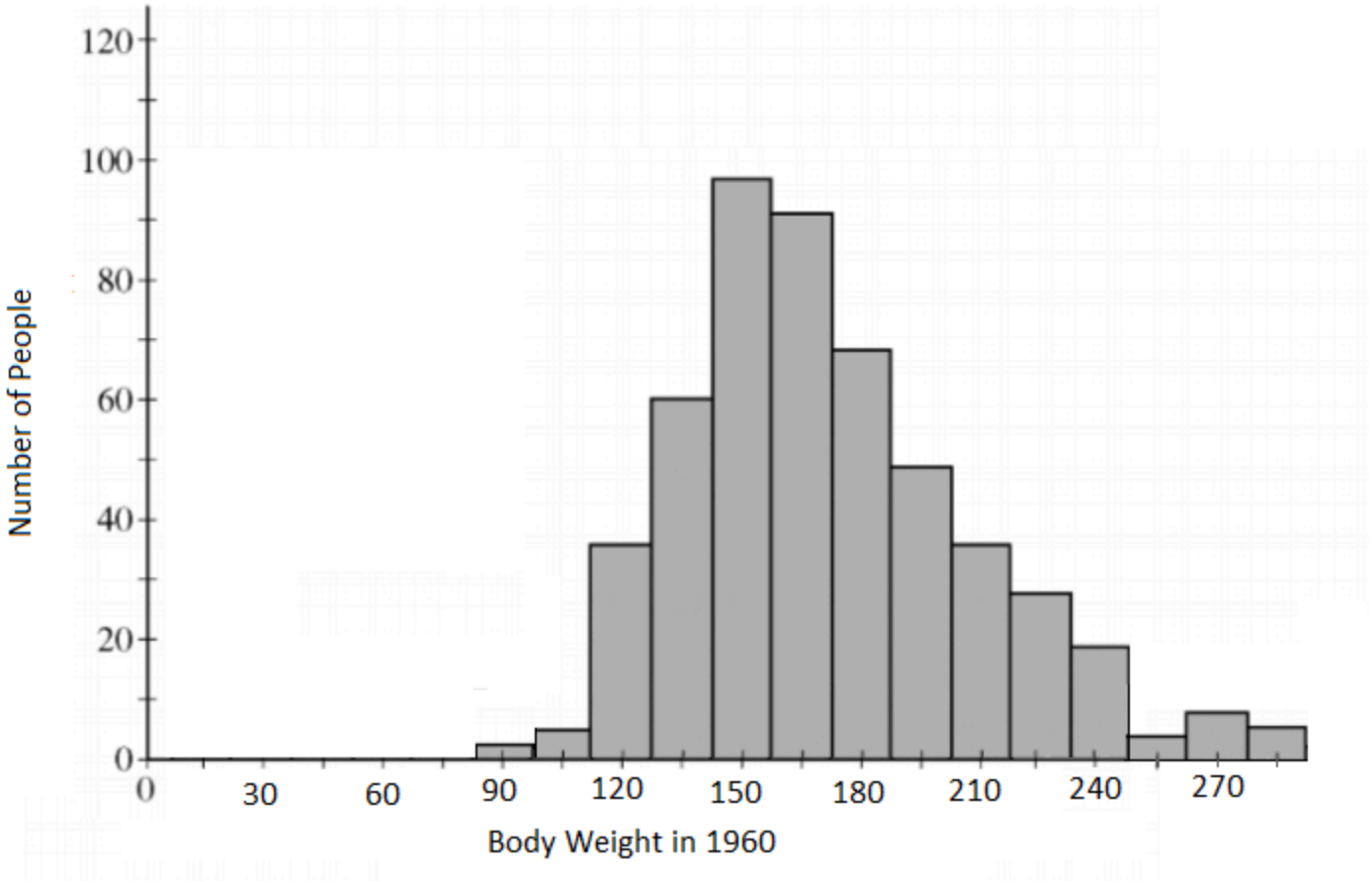
(C) Individual observations in each sample can be considered independent, and the samples themselves are independent.

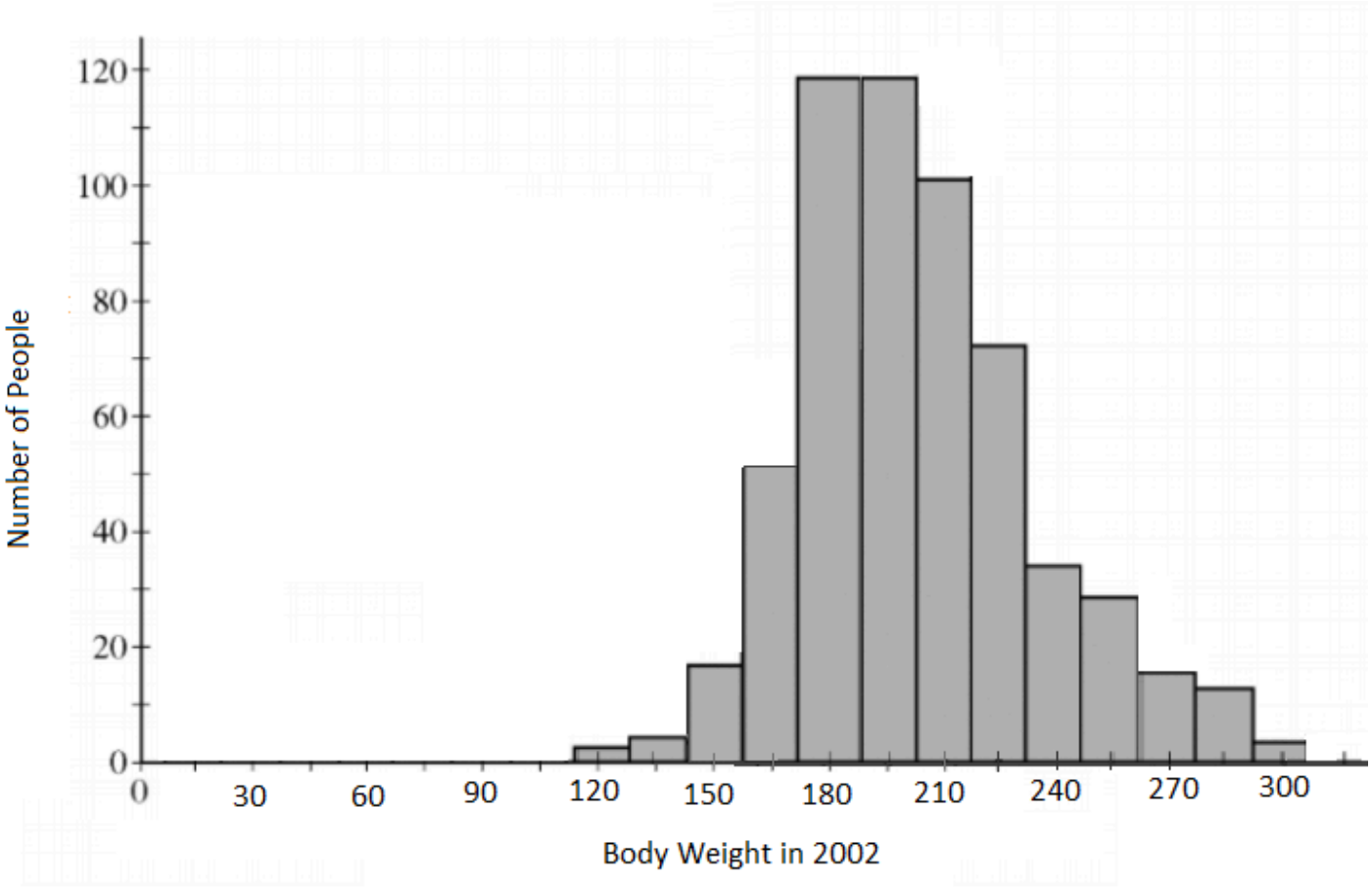
~~~U12L06 Homework~~~

1. Inference comparing two groups or populations: Writing hypotheses to test the difference of means
2. Inference comparing two groups or populations: Conditions for inference on two means

1. **AP**: An experiment will be conducted to test the effectiveness of a chemotherapy treatment for cancer. Volunteers will be randomly assigned to take either the supplement or a placebo for 6 months, with 14 volunteers in each group. The subjects will not know which treatment they receive. At the end of the experiment, researchers plan to calculate the mean tumor mass decrease for each of the two groups and to construct a two-sample *t*-confidence interval for the difference of the two treatment means. Which of the following assumptions is necessary for the confidence interval to be valid?
   (A) The sample size is greater than or equal to 10 percent of the population size.
   (B) Each of the two groups has at least 5 successes and at least 5 failures.
   (C) The distributions of tumor mass decrease of the two treatments are approximately normally distributed.
   (D) The volunteers in the chemotherapy group are paired with volunteers in the placebo group.
   (E) The expected number of people who lose tumor mass in each group is at least 5.

2. **AP**: Independent random samples of 500 men were taken from across the United States for the years 1960 and 2002. Histograms of body weight per person for the years are show below.





(a) Compare the distributions of body weight in the United States for the years 1960 and 2002.

(b) A researcher wants to use these data to construct a confidence interval to estimate the change in mean body weight in the United States from the year 1960 to the year 2002. State the conditions for using a two-sample *t*-procedure, and explain whether the conditions for inference were met.

Unit 12 Lesson 07: Calculate a Two-Sample t Interval for the Difference of Means

What's the difference between calculating a one-sample t interval and a two-sample t interval?

- Calculate the values for the specified interval for two-sample means

1. **AP**: A matched-pairs t-test is NOT an appropriate way to analyze data consisting of which of the following?
   (A) Measurements of resting heart rate taken both before and after a two-year fitness course for a random sample of 100 people who took the course
   (B) Measurements of resting heart rate for each twin for 100 randomly selected pairs of twins
   (C) Measurements of resting heart rate for both individual in pairs formed by matching 100 people from State A and 100 people from State B based on weekly hours spent exercising
   (D) Measurements of resting heart rate for both individual in pairs formed by assigning 100 people to pairs at random
   (E) Measurements of resting heart rate recorded for both spouses of 100 randomly selected married couples

2. [Source: Khan Academy] Kylie suspected that when people exercise longer, their body temperatures change. She randomly assigned people to exercise for 30 or 60 minutes, then measured their temperatures. The 18 people who exercised for 30 minutes had a mean temperature of $\bar{x}_{30} = 38.3^\circ C$ with a standard deviation $s_{30} = 0.27^\circ C$. The 24 people who exercised 60 minutes had a mean temperature of $\bar{x}_{60} = 38.9^\circ C$ with a standard deviation $s_{60} = 0.29^\circ C$. Assume that the conditions for inference have been met, and that Kylie will use the conservative degrees of freedom from the smaller sample size. Determine the 90% confidence interval for the difference in mean body temperature after exercising for the two amounts of time. What can she conclude based on this confidence interval?

3. **AP**: An economist wants to estimate the difference between the retirement savings of millennials and baby boomers. A random sample of 25 millennials has a mean retirement savings of $6,201 and a standard deviation of $1,205. A random sample of 25 baby boomers has a mean retirement savings of $89,931 and a standard deviation of $9,381. What is the standard error of the difference (boomers - millennials) between the sample means?

(A) $\sqrt{\frac{(1,205)^2 + (9,381)^2}{50}}$

(B) $\sqrt{\frac{(1,205)^2 - (9,381)^2}{50}}$

(C) $\sqrt{\frac{(1,205)^2 + (9,381)^2}{25}}$

(D) $\sqrt{\frac{(1,205)^2 - (9,381)^2}{25}}$

(E) $\sqrt{\frac{(1,205) - (9,381)}{25}}$

~~~U12L07 Homework~~~ 1. Inference comparing two groups or populations: Two-sample t interval for the difference of means

~~~U12L07 Classwork~~~

1. [Source: Khan Academy] Anouk conducted an experiment to see what difference plowing fields had on their total yield of crops. The sample of 6 fields that he plowed had a mean yield of $\bar{x}_P$ = 11,700 kilograms per hectare (kg/ha) and a standard deviation of 4,900 kg/ha. The 7 fields he didn't plow had a mean yield of $\bar{x}_N$ = 9,620 kg/ha and a standard deviation of 3,200 kg/ha. Assume that the conditions for inference have been met, and that Anouk will use the conservative degrees of freedom from the smaller sample size. Which of the following is a 95% confidence interval for the difference in mean yield (in kilograms per hectare) of the plowed and not plowed fields ($\mu_P - \mu_N$)? What can he conclude based on this confidence interval?

2. **AP**: A diet shake maker claims its Brand A weight loss shake has fewer calories than its competitor's Brand B weight loss shake. 8 shakes of each brand are tested independently, and the calories in each shake are shown in the table below.

Brand A	300	291	299	304	307	315	319	321
Brand B	320	310	331	314	302	326	311	313

Provided that the assumptions for inference are met, which of the following tests should be conducted to determine if Brand A weight loss shakes do, in fact, have fewer calories than Brand B weight loss shakes?

(A) A one-sided, paired *t*-test
(B) A one-sided, two-sample *t*-test
(C) A two-sided, two-sample *t*-test
(D) A one-sided, two-sample *z*-test
(E) A two-sided, two-sample *z*-test

3. **AP**: Researchers want to test the effect of a new drug on blood pressure. One group of volunteers was given the drug, and the other was given a placebo. Twelve people participated in the study. Each person took the drug for one month and the placebo for one month, and their blood pressure, in mm Hg, was recorded at the end of each month. The pill that was taken first was determined by the flip of a fair coin, and the people did not know which group they were in. What type of study was conducted by the researchers, and of the following, which is the appropriate *t*-interval for inference?
   (A) A completely randomized design and a *t*-interval for a difference between means for independent samples
   (B) A completely randomized design and a *t*-interval for a mean difference
   (C) A matched-pairs design and a *t*-interval for a difference between means for independent samples
   (D) A matched-pairs design and a *t*-interval for a mean difference
   (E) An observational study and a *t*-interval for a difference between means for independent samples

4. [Source: Khan Academy] Chana surveyed students from her school about how they travel to school and how long it takes them. The following tables summarizes her results.

	Cycling	Walking
Sample mean	$\bar{x}_C$ = 23 min	$\bar{x}_W$ = 20.375 min
Sample standard deviation	s_C = 5.9 min	s_W = 6.61 min
Sample size	n_C = 4	n_W = 8

   Assume that the conditions for inference have been met, and that Chana will use the conservative degrees of freedom from the smaller sample size. Which of the following is a 99% confidence interval for the difference in mean travel time (in minutes) for the different transportation methods ($\mu_C - \mu_W$)?
   (A) (-19.358, 24.608)
   (B) (-14.465, 19.715)
   (C) (-10.544, 15.794)
   (D) (-9.35, 14.6)
   (E) (-7.07, 12.32)

Unit 12 Lesson 08: Calculate the Test Statistic in a Two-Sample t Test for the Difference of Means

What is a "test statistic", and how does one calculate it?

- Calculate test statistics for differences of means

1. [Source: Khan Academy] Sophie was curious if her company's jumbo brown eggs and jumbo white eggs had different weights. She obtained two independent random samples of jumbo eggs, one of brown eggs and one of white eggs. Here is a summary of the results:

	Brown	White
Mean	71.2 g	70.5 g
Standard deviation	1.2 g	0.8 g
Number of eggs	490	496

Sophie wants to use these results to carry out a two-sample *t* test to determine if the mean weights are significantly different for these types of eggs. Assume that all conditions have been met. Determine an appropriate test statistic for Sophie's test.

2. [Source: Khan Academy] Researchers at a pharmaceutical company want to test a new pain-relief medicine. Specifically, they want to see if its effect takes less time than their old medicine. They take a sample of people who suffer from chronic pain and randomly assign them into two groups. The first group receive the old medicine while the second group receive the new medicine (the participants don't know which medicine they have). Then, the participants are asked to measure the time since they take the medicine until their pain is gone. Here is a summary of the results:

	Old medicine	New medicine
Mean	22 min	19 min
Standard deviation	4 min	2 min
Number of people	31	29

The researchers want to use these results to test $H_0 : \mu_{Old} - \mu_{New} = 0$ versus $H_a : \mu_{Old} - \mu_{New} > 0$. Assume that all conditions have been met. Determine an appropriate test statistic for the researchers' test.

~~~U12L08 Homework~~~ 1. Inference comparing two groups or populations: Test statistic in a two-sample t test

~~~U12L08 Classwork~~~

1. **AP**: A study was conducted using data collected on the average weight of a random sample of 12 pairs of oranges grown on trees older than 15 years vs oranges grown on trees younger than 15 years to determine whether the older trees' fruits weigh more than the younger trees' fruits. Let μ_Y represent the average weight of all oranges from younger trees, and μ_O represent the average weight of all oranges from older trees, and μ_D represent the average difference in weight (weight of older minus younger) for all pairs of oranges. Which of the following would be the null and alternative hypotheses for this study?
   (A) $H_0 : \mu_O = \mu_Y$ and $H_a : \mu_O < \mu_Y$
   (B) $H_0 : \mu_O = \mu_Y$ and $H_a : \mu_O \neq \mu_Y$
   (C) $H_0 : \mu_D = 0$ and $H_a : \mu_D > 0$
   (D) $H_0 : \mu_O - \mu_Y = \mu_D$ and $H_a : \mu_O - \mu_Y > \mu_D$
   (E) $H_0 : \mu_O - \mu_Y = \mu_D$ and $H_a : \mu_O - \mu_Y \neq \mu_D$

2. [Source: Khan Academy] Jenny wants to compare the amount of weight people lose at her gym using two popular, non-overlapping weight-loss programs, Shred It and Burn It. She randomly samples people who use each program and tracks their weight loss. Here is a summary of the results:

	Shred it	Burn it
Mean	10 kg	7 kg
Standard deviation	3.5 kg	5.7 kg
Number of people	35	35

   Jenny wants to use these results to carry out a two-sample *t* test to determine if the mean weight losses are significantly different for the two programs. Assume that all conditions have been met. What would be an appropriate test statistic for Jenny's test?

3. [Source: Khan Academy] Kaito grows tomatoes in two separate fields. When the tomatoes are ready to be picked, he is curious as to whether the sizes of his tomato plants differ between the two fields. He takes a random sample of plants from each field and measures the heights of the plants. Here is a summary of the results:

	Field A	Field B
Mean	1.3 m	1.6 m
Standard deviation	0.5 m	0.3 m
Number of plants	22	24

Kaito wants to use these results to test $H_0 : \mu_B - \mu_A = 0$ versus $H_a : \mu_B - \mu_A > 0$. Assume that all conditions have been met. Which of the following would be an appropriate test statistic for Kaito's test?

(A) $t = \dfrac{(1.3 - 1.6) - 0}{\sqrt{\frac{0.5^2}{22} + \frac{0.3^2}{24}}}$

(B) $t = \dfrac{(1.3 - 1.6) - 0}{0.6 \cdot \sqrt{\frac{1}{22} + \frac{1}{24}}}$

(C) $t = \dfrac{1.3 - 1.6}{\frac{0.6}{\sqrt{23}}}$

(D) $t = \dfrac{(1.6 - 1.3) - 0}{0.6 \cdot \sqrt{\frac{1}{24} + \frac{1}{22}}}$

(E) $t = \dfrac{(1.6 - 1.3) - 0}{\sqrt{\frac{0.3^2}{24} + \frac{0.5^2}{22}}}$

Unit 12 Lesson 09: Find the P-Value in a Two-Sample t Test for the Difference of Means

What if I want to go one step further, and find the probability that a certain test statistic would occur?

- Find P-values for z tests for differences of means

1. [Source: Khan Academy] A researcher studying education in the United Kingdom and Germany wanted to compare how many years, on average, women in each country spend in school. The researcher obtained a random sample of women from each country. Here is a summary of the number of years the women in each sample spent in school:

Years in school	UK	Germany
Mean	13.2	11.8
Standard deviation	2.8	3.3
Number of women	58	62

The researcher wants to use these results to test $H_0 : \mu_{UK} - \mu_G = 0$ versus $H_a : \mu_{UK} - \mu_G > 0$ Assume that all conditions have been met. What is the P-value associated with these sample results?

(A) *P*-value < 0.01
(B) 0.01 ≤ *P*-value < 0.05
(C) 0.05 ≤ *P*-value < 0.10
(D) 0.10 ≤ *P*-value < 0.20
(E) 0.20 ≥ *P*-value

2. [Source: Khan Academy] A sociologist studying fertility in Malaysia and South Korea wanted to compare how many babies, on average, women in each country have. The sociologist obtained a random sample of women from each country. Here is a summary of the number of babies for the women in each sample:

Number of babies	Malaysia	South Korea
Mean	1.98	1.32
Standard deviation	1.77	1.28
Number of women	46	58

They want to use these results to test $H_0 : \mu_M - \mu_{SK} = 0$ versus $H_a : \mu_M - \mu_{SK} \neq 0$. Assume that all conditions have been met. What is the P-value associated with these sample results?

3. **AP**: A food processing company has created a new food additive, monosodium glutamate (MSG). It tastes savory and delicious, but some people develop a headache shortly after eating it. To balance the flavor against the negative qualities, scientists conducted a double-blind experiment in which 12 people who experience headaches after eating MSG were randomly selected to take either a low dose or a high dose of the additive. Each of the 12 people recorded the time, in minutes, from eating the additive until headache formation. After an appropriate time period, each of the 12 people took the other dosage level and recorded the time from eating it until headache formation. The dosage level each person took first was randomly determined, and because both dosage levels look the same, the people in the study did not know which dosage level was eaten first. The table below shows summary statistics for the results.

	Minutes until Headache Formation		Difference (low dosage minus high dosage)
	Low Dosage	High Dosage	
Mean	21.08	17.67	1.58
Standard Deviation	5.09	3.42	2.31

Which of the following values is closest to the *p*-value of the appropriate t-test?

(A) 0.0186
(B) 0.3651
(C) 0.5822
(D) 0.7743
(E) 0.9814

4. **AP**: Neutrinos are a subatomic particle with near-zero mass that is notoriously hard to detect. They do not interact strongly with matter so their presence is hard to measure. In order to measure them, scientists build giant water tanks deep under the ground where cosmic rays cannot reach. They line the water tanks with photodetectors to pick up the photons released upon a neutrino's collision with a water particle. Another scientist developed a new kind of photodetector to try to more accurately measure the number of neutrinos passing through their detection tank for a given unit of time. 10 individual minutes of time chosen at random were used to compare the two kinds of photodetector. For each minute, the number of collisions was counted using both methods. The difference between the number of collisions detected was calculated for each minute. A test of the hypothesis that the population mean difference is zero had a p-value of 0.056.
The hypothesis test described had three components: the number of minutes, the sample standard deviation of the differences, and the magnitude of the sample mean difference. Compared to the test described, which of the following would have resulted in a smaller p-value?\

(A) The number of minutes sampled and the sample standard deviation of the differences remained the same, but the magnitude of the sample mean difference was smaller.
(B) The number of minutes sampled and the magnitude of the sample mean difference remained the same, but the sample standard deviation of the differences was larger.
(C) The number of minutes sampled remained the same, but the magnitude of the sample mean difference was smaller and the sample standard deviation of the difference was larger.
(D) The number of minutes sampled remained the same, but the magnitude of the sample mean difference was larger and the sample standard deviation of the difference was smaller.
(E) The magnitude of the sample mean difference and the sample standard deviation of the differences remained the same, but the number of minutes sampled was smaller.

~~~U12L09 Homework~~~ 1. Inference comparing two groups or populations: P-value in a two-sample t test

~~~U12L09 Classwork~~~

1. [Source: Khan Academy] A sociologist studying marriage in Spain and Italy wanted to compare how old, on average, women in each country are when they first get married. The sociologist obtained a random sample of married women from each country. Here is a summary of the ages at first marriage for the women in each sample:

Age at first marriage	Spain	Italy
Mean	29.5	28.8
Standard deviation	2.5	3.6
Number of women	84	73

The researcher wants to use these results to test $H_0 : \mu_S - \mu_I = 0$ versus $H_a : \mu_S - \mu_I > 0$. Assume that all conditions have been met. What is the P-value associated with these sample results?

(A) *P*-Value < 0.01
(B) 0.01≤ *P*-Value < 0.05
(C) 0.05 ≤ *P*-Value < 0.10
(D) 0.10 ≤ *P*-Value < 0.20
(E) 0.20 ≤ *P*-Value

2. **AP**: A food processing company has created a new food additive, monosodium glutamate (MSG). It tastes savory and delicious, but some people develop a headache shortly after eating it. To balance the flavor against the negative qualities, scientists conducted a double-blind experiment in which 12 people who experience headaches after eating MSG were randomly selected to take either a low dose or a high dose of the additive. Each of the 12 people recorded the time, in minutes, from eating the additive until headache formation. After an appropriate time period, each of the 12 people took the other dosage level and recorded the time from eating it until headache formation. The dosage level each person took first was randomly determined, and because both dosage levels look the same, the people in the study did not know which dosage level was eaten first. The table below shows summary statistics for the results.

	Minutes until Headache Formation		Difference (low dosage minus high dosage)
	Low Dosage	High Dosage	
Mean	21.08	17.67	1.58
Standard Deviation	5.09	3.42	2.31

Which of the following values is closest to the *p*-value of the appropriate t-test?

(A) 0.0186
(B) 0.3651
(C) 0.5822
(D) 0.7743
(E) 0.9814

Unit 12 Lesson 10: Draw Conclusions about Differences of Means

Now that I've run my test and gotten my P-value, what EXACTLY can I conclude?

- Identify appropriate and inappropriate conclusions resulting from a given two-sample t test

1. **AP**: To determine whether Beagles have higher body mass, on average, than Boston Terriers do, independent random samples of body mass were obtained for the two groups. The data are summarized below.

	Beagles	Boston Terriers
Mean	19.9	18
Standard Deviation	3.4	3.9
n	28	31

Based on the data, which of the following statements is true?

(A) At the 5% significance level, Beagles have a significantly higher mean body mass than Boston Terriers do.
(B) At the 1% significance level, Beagles have a significantly higher mean body mass than Boston Terriers do.
(C) At the 5% significance level, Boston Terriers have a significantly higher mean body mass than Beagles do.
(D) At the 1% significance level, Boston Terriers have a significantly higher mean body mass than Beagles do.
(E) At the 10% significance level, there is no significant difference in body mass between the two breeds.

2. **AP**: Researchers investigated whether a new chemical could slow the growth of grass in a typical lawn. From random samples of grass, the researchers found the mean reduction in growth rate by grass treated with the new chemical compared with grass not treated by the new chemical was 0.6 inches per week. All conditions for interference were met, and the *p*-value for the appropriate hypothesis test was 0.099. Which of the following statements is the best interpretation of the *p*-value?

(A) The probability that the null hypothesis is true is 0.099.
(B) The probability that the alternative hypothesis is true is 0.099.
(C) The probability of observing a mean reduction of growth rate of 0.6 inches per week is 0.099.
(D) If the null hypothesis is true, the probability of observing a mean reduction of at least 0.6 inches per week is 0.099.
(E) If the null hypothesis is true, the probability of observing a mean reduction of at most 0.6 inches per week is 0.099.

3. **AP**: All gasoline sold in the United States contains detergent to clean carbon deposits out of cars' engines. Chemical engineers have come up with a new detergent to put in gasoline. A large study was performed to test the effectiveness of the detergent on engine performance. The engineers first identified 77 variables that measure various aspects of proper engine performance. Because each car in the study was able to serve as its own control, the 77 variables were measured for each engine, both before adding the new detergent and after driving with the new detergent for 300 miles. The engineers then performed 77 matched-pair t-tests, one for each variable. They found statistically significant results at the 0.05 level in 2 of the variables, both in the direction of improved engine performance. Which of the following should the investigators conclude?
   (A) There is evidence that the new detergent improves the performance of engines because 2 of the 77 tests were statistically significant.
   (B) There is evidence that the new detergent improves the performance of engines because in studies investigators typically underestimate the proportion of the population that is helped by a detergent.
   (C) There is insufficient evidence that the new detergent improves the performance of engines because at the 0.05 significance level, one could easily get statistically significant results in 2 out of 77 tests just due to chance variability.
   (D) There is insufficient evidence that the new detergent improves the performance for engines because the sample size of 77 is not large enough to draw a conclusion.
   (E) No conclusion can be drawn because an even number of variables is needed for a matched-pairs design.

4. **AP**: A two-sample *t*-test of the hypotheses $H_0: \mu_1 - \mu_2 = 0$versus $H_0: \mu_1 - \mu_2 > 0$produces a *p*-value of 0.035. Which of the following must be true?

   I. A 90 percent confidence interval for the differences in means will contain the value 0.
   II. A 95 percent confidence interval for the difference in means will contain the value 0.
   III. A 99 percent confidence interval for the difference in means will contain the value 0.

   (A) I only
   (B) III only
   (C) I and II only
   (D) II and III only
   (E) I, II and III

~~~U12L10 Homework~~~ 1. Inference comparing two groups or populations: Making conclusions about the

difference of proportions

~~~U12L10 Classwork~~~

1. **AP**: A quality control officer wants to make sure his employer, company A, has an automated phone system that is competitive with their main competitor, company B. He needs to make sure company A's phone system is approximately as quick to navigate as company B's. Twenty five randomly selected callers called company A's service, and the time required to locate their desired information using the service had a mean time 2.1 minutes and a standard deviation of 0.7 minutes. Twenty five different randomly selected callers accessed company B's service, and the time required to locate their desired information using the service had a mean time of 2.0 minutes and a standard deviation of 1.4 minutes. Assuming that the conditions for interference are met, which of the following statements about the *p*-value obtained from the data and the conclusion of the significance test is true?
   (A) The *p*-value is less than 0.01; therefore, there is a significant difference in mean call times on the two services.
   (B) The *p*-value is greater than 0.01 but less than 0.05; therefore, there is a significant difference in mean call times on the two services.
   (C) The *p*-value is greater than 0.05 but less than 0.10; therefore, there is a significant difference in mean call times on the two services.
   (D) The *p*-value is greater than 0.10; therefore, there is no significant difference in mean call times on the two services.
   (E) Since this is a matched-pairs situation, additional information is needed to perform a test of significance.

2. **AP**: An artificial intelligence program is designed to identify images of animals. The creators claim that, given a set of 50 images, the AI can accurately identify at least 45 correctly. To test whether the claim is true, a software engineer took a random sample of 50 images, and computed the mean number of correctly labeled images. A test of the hypotheses $H_0 : \mu = 45$ versus $H_a : \mu < 45$ was conducted, where μ represents the population mean number of correctly labeled images. The *p*-value for the test was 0.07. Which of the following is the most appropriate conclusion to draw at the significance level of $\alpha = 0.05$?
   (A) Because the *p*-value is greater than the significance level, there is convincing evidence that the population mean number of correctly labeled images per category is 45.
   (B) Because the *p*-value is greater than the significance level, there is convincing evidence that the population mean number of correctly labeled images per category is less than 45.
   (C) Because the *p*-value is greater than the significance level, there is not convincing evidence that the population mean number of correctly labeled images per category is less than 45.
   (D) Because the *p*-value is less than the significance level, there is convincing evidence that the population mean number of correctly labeled images per category is 45.
   (E) Seven percent of the categories will have less than 49 correctly labeled images.

3. **AP**: A randomized experiment was performed to determine whether a species of banana tree give different yields when grown in brackish water than in regular fresh water. A total of 31 plants were grown; 15 using freshwater, and 16 using brackish water. The distributions of the data did not show marked skewness and there were no outliers in either data set. The results of the experiment are shown below.

	Freshwater	Brackish Water
Average number of bananas per plant	47.1	43.9
Standard deviation	5.41	3.97
Number of plants	15	16

Which of the following statements best describes the conclusion that can be drawn from this experiment?

(A) There is no statistical evidence of difference in the yields between brackish water and freshwater ($p > 0.15$)
(B) There is a borderline statistically significant difference in the yields between brackish water and freshwater ($0.10 < p < 0.15$)
(C) There is a evidence of a statistically significant difference in the yields between brackish water and freshwater ($0.05 < p < 0.10$)
(D) There is a evidence of a statistically significant difference in the yields between brackish water and freshwater ($0.01 < p < 0.05$)
(E) There is a evidence of a statistically significant difference in the yields between brackish water and freshwater ($p < 0.01$)

4. **AP**: Schools use standardized testing to compare their performance against other schools. An educational agency was concerned that the scores of students in Madera County were lower than the scores of students in Orange County. The agency took two independent random samples of scores of students, one from Madera county and one from Orange county. The data are summarized in the table below.

	Madera County	Orange County
Mean Score	81.4	82.9
Standard Deviation	5.3	6.2
Number of Students	119	81

(A) Yes, there is evidence at the significance level of $\alpha = 0.001$
(B) Yes, there is evidence at the significance level of $\alpha = 0.01$ but not at $\alpha = 0.001$.
(C) Yes, there is evidence at the significance level of $\alpha = 0.05$ but not at $\alpha = 0.01$.
(D) Yes, there is evidence at the significance level of $\alpha = 0.10$ but not at $\alpha = 0.05$.
(E) No, there is no evidence at the significance level of $\alpha = 0.10$.

Unit 12 Lesson 11: Review for Test on Two-Sample Significance Tests

Can I handle any situation that involves hypotheses, conditions, test statistics, P-values, and conclusions for two-sample significance tests?

- Ensure you've mastered the concepts and skills of two-sample significance tests
- Ensure you've retained mastery of previous units

1. **AP**: Japanese researchers investigated the relationship between fish consumption and schizophrenia. The researches gave a questionnaire about eating habits to a randomly selected sample of japanese men. Based on the responses to the questionnaire, the men were classified into two groups. Group A consisted of the 8150 men who ate the most fish per week, and group B consisted of the 8150 men who ate the least fish per week. The researchers tracked the men's health for ten years. During that time, there were 424 cases of schizophrenia among the men in group A and 521 cases of schizophrenia among the men in group B.

   (a) Do the data provide convincing statistical evidence Japanese men who would be classified into group A have a lower probability of becoming schizophrenic than Japanese men who would be classified into group B?

   (b) A report in a newspaper concluded that Japanese men can reduce their probability of becoming schizophrenic by eating more fish. Based on the description of the investigation, was the conclusion appropriate? Justify your answer.

5. **AP**: The successfulness of chemotherapies for cancer are gauged on mean years of survival after treatment. In a study of a new chemotherapy thought to improve patient outcomes, patients undergoing cancer treatment were randomly assigned to one of two groups. During treatment, patients in one group received standard care, and patients in the other group received the new chemotherapy.

   The number of years after treatment was recorded. A higher number means a greater number of years lived for the patient. Summary statistics for the data are shown in the table below.

	Mean (in years)	Standard Deviation (in years)	Sample Size
Standard care	5.8	2.3	58
New chemotherapy	6.9	2.4	58

   Do the data provide convincing statistical evidence that the new treatment results in more years of survival than does the standard care among people similar to the patients in the study?