Chapter 10
Study Guide

Comparing Two Populations or Treatments

In this section, we will turn our attention to dealing with two populations or treatments. A typical question that arises involves how one treatment or population compares to another. This chapter deals with the inference methods used to answer this question.

Objectives

- Correctly set up and carry out a hypothesis test for the difference in two population or treatment means using independent or paired samples.
- Correctly set up and carry out a hypothesis test for the difference in two population or treatment proportions.
- Construct and interpret a confidence interval for a difference in two means.
- Construct and interpret a confidence interval for a difference in two proportions.

Inferences Concerning the Difference Between Two Population or Treatment Means Using Independent Samples

(*Introduction to Statistics & Data Analysis* 3rd ed. pages 583–597/4th ed. pages 638–656)

To begin, let's take a look at inference methods based on two *independent* samples. Independent samples are samples where knowing the individuals selected for one sample does not tell you anything about which individuals are selected into the other sample. In contrast, in paired samples, an observation from one sample is paired in some meaningful way with an observation in the other sample.

As we saw in the previous review section, a hypothesis test can be consolidated into four key parts: hypotheses, test statistic and assumptions, computations, and a conclusion in context. Following this process, or *test procedure*, will be helpful in all other testing situations as well. The question of interest for a two independent sample procedure will be, "Are the two population means or proportions equal to each other or not?" However, since there are now two treatments or populations, it is important to define the symbols used to represent the populations or treatments. One way to manage this is by using subscripts to represent each population and sample separately. In the general notation shown, everything from the same population or treatment uses the same subscript.

Notation	**Mean**	**Variance**	**Standard Deviation**	
Population or Treatment 1	μ_1	σ_1^2	σ_1	
Population or Treatment 2	μ_2	σ_2^2	σ_2	
	Sample Size	**Mean**	**Variance**	**Standard Deviation**
Sample from Population or Treatment 1	n_1	$\bar{x}_1$	s_1^2	s_1
Sample from Population or Treatment 2	n_2	$\bar{x}_2$	s_2^2	s_2

When working with independent samples, hypotheses that compare the means would be written as one of the following:

$\mu_1 - \mu_2 = 0$, which is the same as saying $\mu_1 = \mu_2$

or if one mean were larger than the other, it would be written as

$\mu_1 - \mu_2 > 0$, which is the same as saying $\mu_1 > \mu_2$ or
$\mu_1 - \mu_2 < 0$, which is the same as saying $\mu_1 < \mu_2$

When the samples are independent, the sampling distribution of the difference in sample means, $\bar{x}_1 - \bar{x}_2$, has a mean that is equal to the difference of the two population means. The variance of the sampling distribution of $\bar{x}_1 - \bar{x}_2$ is the sum of the two population variances.

Given this behavior, we get the following sampling distribution properties.

If the random samples are selected independently, then

1. $\mu_{\bar{x}_1-\bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2} = \mu_1 - \mu_2$
   so this means the distribution will be centered around $\mu_1 - \mu_2$, making $\bar{x}_1 - \bar{x}_2$ an unbiased estimate for $\mu_1 - \mu_2$.

2. $\sigma^2_{\bar{x}_1-\bar{x}_2} = \sigma^2_{\bar{x}_1} + \sigma^2_{\bar{x}_2} = \dfrac{\sigma^2_1}{n_1} + \dfrac{\sigma^2_2}{n_2}$ and $\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\dfrac{\sigma^2_1}{n_1} + \dfrac{\sigma^2_2}{n_2}}$

3. Finally, if both n_1 and n_2 are large or the population distributions are approximately normal, then both $\bar{x}_1$ and $\bar{x}_2$ are approximately normal and this implies that $\bar{x}_1 - \bar{x}_2$ is also approximately normal.

Simply interpreted, means that the general rules for two independent samples are

1. The mean value of a difference in means is the difference of the two individual mean values.
2. The variance of a difference in *independent* quantities is the *sum* of the two individual variances.

With this in mind, the statistic

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma^2_1}{n_1} + \dfrac{\sigma^2_2}{n_2}}}$$

will have a standard normal distribution. However, because the values of σ^2_1 and σ^2_2 are rarely known, we will typically use a *t* statistic:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s^2_1}{n_1} + \dfrac{s^2_2}{n_2}}}$$

$$**df = \frac{(V_1 + V_2)^2}{\dfrac{V_1^2}{n_1 - 1} + \dfrac{V_2^2}{n_2 - 1}}, \text{ where } V_1 = \frac{s_1^2}{n_1} \text{ and } V_2 = \frac{s_2^2}{n_2}$$

If you are using *t* table, the result for *df* should be rounded down (truncated) to get an integer for *df*.

**You are not required to know this formula for degrees of freedom on the AP exam. Instead, you are allowed to use the *df* that the calculator gives you.

Test Procedures

Once we have established that the two samples are independently selected, the null hypothesis of interest would be of the general form

$H_0 : \mu_1 - \mu_2 =$ some hypothesized value (very often, this value = 0)

We will follow the same basic hypothesis testing procedure as in previous sections. Two key differences in comparing two populations will occur in the procedures. First, in the hypotheses and test statistic sections there will be two populations and two sets of sample statistics used in the calculations and claims. Second, in the assumptions section it is necessary to check assumptions with each sample.

A summary of the two-sample t test for comparing two population means is outlined in the steps below.

STATE THE HYPOTHESIS

$H_0 : \mu_1 - \mu_2 =$ some hypothesized value

and the alternative hypothesis will be one of the following:

$H_a : \mu_1 - \mu_2 >$ some hypothesized value

$H_a : \mu_1 - \mu_2 <$ some hypothesized value

$H_a : \mu_1 - \mu_2 \neq$ some hypothesized value

state the α level

ASSUMPTIONS

Simple random sample: Check that *both* samples came from a random sample.

Large sample or normal population distributions: again, check *both* samples to insure that n_1 and n_2 are sufficiently large; if the samples sizes are not large, construct a plot to verify the plausibility that each population distribution is approximately normal.

TEST STATISTIC

given $\bar{x}_1, \bar{x}_2, s_{x_1}, s_{x_2}, n_1, n_2$, then

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s^2_1}{n_1} + \frac{s^2_2}{n_2}}}$$

and

$$df = \frac{(V_1 + V_2)^2}{\frac{V_1^2}{n_1 - 1} + \frac{V_2^2}{n_2 - 1}}, \text{ where } V_1 = \frac{s_1^2}{n_1} \text{ and } V_2 = \frac{s_2^2}{n_2}$$

(Note: if a t table is used, remember to truncate the df to an integer. Or, a more conservative *df* can be used by using either $(n_1 - 1)$ or $(n_2 - 1)$, whichever is smaller.)

CONCLUSION Once the test statistic and P-value are calculated, the P-value can be compared to the significance level, α. If P-value $< \alpha$ there is sufficient evidence to reject the null hypothesis in favor of the alternative hypothesis. In other words, we would have sufficient evidence to say that the difference in the two means is not equal to the hypothesized value and favor whichever direction the alternative hypothesis claim takes.

SAMPLE PROBLEM 1 Students want to see if regular and low-fat chocolate chip cookies have the same number of chips in a cookie, on average. They suspect that the way low-fat cookies are made is by simply reducing the number of chips in each cookie. To test this theory, they selected random samples of cookies of each type and counted the number of chips in each cookie. Summary statistics are shown in the table below. Boxplots of the two samples were approximately symmetric and there were no outliers.

Cookie	**Sample Size**	**Mean Chips**	**Standard Deviation**
Regular	25	16.3	2.29
Low-Fat	25	15.2	3.25

Based on the data provided, is there evidence that the mean number of chips for regular cookies is greater than the mean number for low-fat cookies? Justify your answer.

SOLUTION TO PROBLEM 1 To answer this question, we will need to carry out an appropriate hypothesis test. The steps are shown here.

HYPOTHESES

$H_0 : \mu_r - \mu_l = 0$ where μ_r = mean number of chips for regular chocolate chip cookies
μ_l = mean number of chips for low-fat chocolate chip cookies

$H_A : \mu_r - \mu_l > 0$ to test to see if regular cookies have more chips, on average, than low-fat

$\alpha = 0.05$ (since no significance level is specified, we can choose a value.)

TEST: two sample t test

ASSUMPTIONS

1. The problem states that the samples were random samples.
2. The samples sizes are not greater than 30, but the problem states that the two boxplots were approximately symmetric and there were no outliers. So, it is reasonable to assume that the two population distributions are approximately normal.

TEST STATISTIC

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s^2_1}{n_1} + \frac{s^2_2}{n_2}}} = \frac{(16.3 - 15.2) - (0)}{\sqrt{\frac{2.69^2}{25} + \frac{3.25^2}{25}}} = \frac{(16.3 - 15.2) - (0)}{\sqrt{0.289 + 0.423}} = 1.33$$

$$df = \frac{(V_1 + V_2)^2}{\frac{V_1^2}{n_1 - 1} + \frac{V_2^2}{n_2 - 1}} = \frac{(0.289 + 0.423)^2}{\frac{0.289^2}{24} + \frac{0.423^2}{24}} = \frac{0.50694}{0.01096} = 46.3$$

$P - value = 0.095$

CONCLUSION Since 0.095 > 0.05, we fail to reject the null hypothesis. In other words, there is not convincing evidence that the mean number of chocolate chips in regular cookies is greater than the mean number in low-fat cookies, based on the samples.

POOLED *T* TEST

The pooled t test is used when it is known that the two population variances are $(\sigma_1^2 = \sigma_2^2)$. However, this rarely is known. So, unless there is reason to know the variances are the same, use an *unpooled* test. In the case of the AP Exam, we advise you to always use the unpooled test when testing hypotheses about the difference in means using independent samples.

CONFIDENCE INTERVALS

A confidence interval can be used to estimate a difference in means. Similar to the one-sample interval, the two-sample interval would be calculated in the following manner.

$$\bar{x}_1 - \bar{x}_2 \pm (t \text{ critical value})\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

with

$$df = \frac{(V_1 + V_2)^2}{\frac{V_1^2}{n_1 - 1} + \frac{V_2^2}{n_2 - 1}}, \text{ where } V_1 = \frac{s_1^2}{n_1} \text{ and } V_2 = \frac{s_2^2}{n_2}$$

The same assumptions used for the two-sample t test still apply and will need to be verified before computing the confidence interval. Remember, for a two-sample confidence interval for independent samples, both samples should be random samples and both sample sizes must be sufficiently large or the populations approximately normal for the interval to be appropriate.

INFERENCES CONCERNING THE DIFFERENCE BETWEEN TWO POPULATION OR TREATMENT MEANS USING PAIRED SAMPLES

(*Introduction to Statistics & Data Analysis* 3rd ed. pages 606–614/4th ed. pages 658–666)

The two-sample t test just described is used with two independent samples. However, sometimes there are other variables that should be considered. For example, tracking weight loss of individuals who are exercising versus those who are not may be confounded by the individual's caloric intake as well. So just having a mean weight loss for those who exercise and those who don't may not be the best comparison. However, if the two groups were *paired* by caloric intake, then the weight loss difference in the pairs could be calculated. This would allow us to see if the exercise had an effect on weight loss for individuals who ate roughly the same number of calories per day.

If the two sample sizes are the same, think about whether or not the samples are paired. If the sample sizes are different, then the samples are not paired samples. To perform the paired t test, you will use the same steps as for the one-sample t test where:

1. The hypothesis is based on the mean difference in the pairs (μ_d = hypothesized value).
2. The *df* is $n - 1$, where the n is the number of pairs we have.

The paired t test is summarized in these simple steps.

HYPOTHESES

$H_0 : \mu_d =$ hypothesized value, where $\mu_d = \mu_1 - \mu_2$

alternative hypothesis will be either

$H_a : \mu_d >$ some hypothesized value

$H_a : \mu_d <$ some hypothesized value

$H_a : \mu_d \neq$ some hypothesized value

state the α level

TEST: paired t test

ASSUMPTIONS

1. The samples are *paired*, and therefore, *not independent.*
2. The sample differences came from a random sample of population differences.
3. The sample size (number of differences) is large or it is reasonable to assume that the distribution of differences in the population is approximately normal.

TEST STATISTIC

Given $\bar{x}_d$ = mean of the sample difference and s_d = standard deviation of sample differences

$$t = \frac{\bar{x}_d - \text{hypothesized value}}{\frac{s_d}{\sqrt{n}}}$$

and

$df = n - 1$, where n is the number of sample differences.

CONCLUSION As before, if P-value $< \alpha$ there is sufficient evidence to reject the null hypothesis in favor of the alternative hypothesis. In other words, we would have sufficient evidence to reject the hypothesis that the mean difference equaled the hypothesized value. We use the same four steps as in other tests.

To better understand this procedure, let's reconsider the chocolate chip cookie scenario in the first part of the chapter. If you recall, we were given the mean number of chips in samples of 25 regular and low-fat cookies. Suppose that each student in the class counted the number of chips in one cookie of each type. Since everyone may count what constitutes a chip a little differently, a better way to analyze this data is to incorporate the pairing. The following sample problem uses the same set of data used for the two-sample test earlier, but instead of only having the means for each type of cookie, this data also contains information on the pairing. Notice how this test differs from the two-sample problem presented earlier.

SAMPLE PROBLEM 2 Twenty-five students counted the number of chocolate chips in a low-fat and a regular chocolate chip cookie and recorded the counts as seen in the table below. Prior to counting, each student flipped a coin to decide if they would count the regular or the low-fat cookie first. Is there evidence that, on average, regular chocolate chip cookies contain more chips than low-fat cookies? Defend your answer.

R = regular chocolate chip cookies
L = low-fat chocolate chip cookies
Diff. = R – L

Person	R	L	Diff	Person	R	L	Diff
1	20	17	3	14	21	14	7
2	17	13	4	15	17	11	6
3	12	10	2	16	13	16	–3
4	15	13	2	17	16	16	0
5	16	12	4	18	17	16	1
6	20	23	–3	19	15	15	0
7	15	16	–1	20	18	20	–2
8	19	18	1	21	14	12	2
9	18	16	2	22	13	15	–2
10	17	15	2	23	21	19	2
11	15	19	–4	24	12	10	2
12	18	18	0	25	15	13	2
13	13	12	1				

$\bar{x}_d = 1.12,\ s_d = 2.68$

SOLUTION TO PROBLEM 2 To answer this question, we need to carry out a hypothesis test. Since the samples are paired, we will use a paired t test.

HYPOTHESES

$H_0 : \mu_d = 0$ where μ_d = difference in mean number of chips(regular – low-fat)

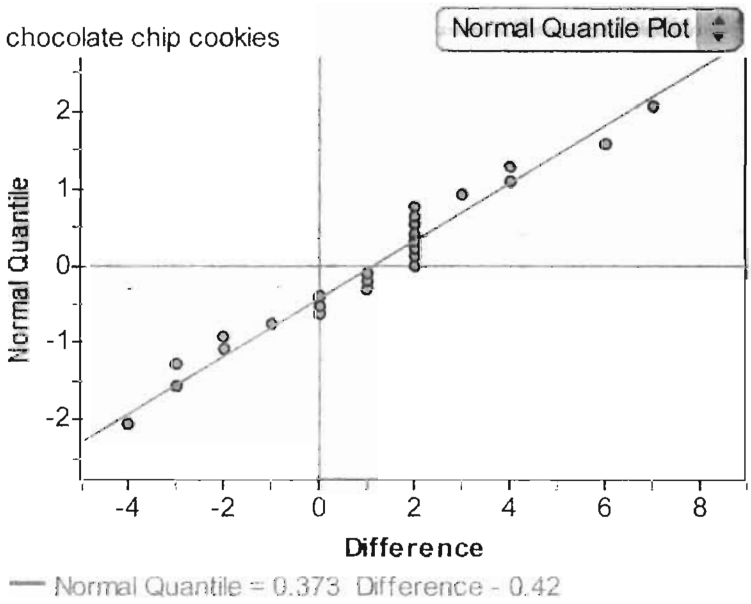
$H_a : \mu_d > 0$ since they want to know if regular have more chips

$\alpha = 0.05$

TEST: paired *t* test

ASSUMPTIONS

1. The samples are paired.
2. The differences can be regarded as a random sample of differences.
3. The sample size is only 25, which is not large. A normal probability plot (below) is reasonably straight and so it is reasonable to assume that the distribution of differences is approximately normal.



TEST STATISTIC

$$t = \frac{\bar{x}_d - 0}{\frac{s_d}{\sqrt{n}}} = \frac{1.12 - 0}{\frac{2.68}{\sqrt{25}}} = \frac{1.12}{\frac{2.68}{5}} = \frac{1.12}{0.536} = 2.09$$

and

$df = n - 1 = 24$

$P - value = 0.024$

CONCLUSION Since 0.024 < 0.05, there is sufficient evidence to reject the null hypothesis in favor of the alternative hypothesis. In other words, there is reason to believe that the mean difference in chips is greater than 0 so there does appear to be more chocolate chips, on average, in regular cookies than in low-fat cookies.

What is important to notice in the sample problem given is that when using the differences instead of two independent samples, our conclusion changed significantly. Both tests used the same samples; however, the paired t test took the pairing into account. This illustrates why it is important to use an appropriate test when the samples are paired.

LARGE-SAMPLE INFERENCES CONCERNING A DIFFERENCE BETWEEN TWO POPULATION OR TREATMENT PROPORTIONS

(*Introduction to Statistics & Data Analysis* 3rd ed. pages 619–626/4th ed. pages 671–678)

As we saw in the previous chapters, categorical data involves working with proportions instead of means. We also learned that a hypothesis test and confidence interval could be generated for this type data. In this section, we will see how this concept can also be extended to compare two population proportions. For example, is the proportion of males who would be willing to pay an extra $300 to upgrade to a first class airline seat equal to the proportion of females who would be willing to pay extra to upgrade? This question and others that involve comparing two population proportions can be answered using a two proportions z test.

This test is based on the following properties of the sampling distribution of $\hat{p}_1 - \hat{p}_2$:

1. $\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2$, in other words, since the distribution of $\hat{p}_1 - \hat{p}_2$ is centered at $p_1 - p_2$, it is an unbiased estimator of $p_1 - p_2$
2. $\sigma^2_{\hat{p}_1-\hat{p}_2} = \sigma^2_{\hat{p}_1} + \sigma^2_{\hat{p}_2} = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$

   making

   $$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$
3. When the sample sizes are both large, the sampling distribution of $\hat{p}_1 - \hat{p}_2$ is approximately normal. This MUST be checked by verifying all of the following:

   $n_1 p_1 \ge 10$ $\quad$ $n_2 p_2 \ge 10$

   $n_1(1-p_1) \ge 10$ $\quad$ $n_2(1-p_2) \ge 10$

AP Tip

It is not enough to just state the large sample conditions—you must show that you have actually checked these conditions.

Once we have verified that the sampling distribution is roughly normal and the samples were random samples, we can then form a test statistic:

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}, \text{ which simplifies to } z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

when the null hypothesis specifies $p_1 - p_2 = 0$.

The actual test procedure uses a slightly different denominator. Because the values of p_1 and p_2 are not known, the denominator cannot be computed. Because the null hypothesis $H_0 : p_1 - p_2 = 0$ specifies that p_1 and p_2 are equal, we estimate this common value

p_c = combined estimate of the common population proportion,

using $$p_c = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{\text{total S's in two samples}}{\text{total of the sample sizes}}$$

So given this information, we can now describe a two proportions z test.

HYPOTHESIS

$H_0 : p_1 - p_2 = 0,$

the alternative will again be one of

$H_a : p_1 - p_2 \neq 0$

$H_a : p_1 - p_2 > 0$

$H_a : p_1 - p_2 < 0$

ASSUMPTIONS

1. Independent random samples
2. Large sample sizes. Verify

   $n_1\hat{p}_1 \geq 10$ $\quad$ $n_2\hat{p}_2 \geq 10$

   $n_1(1-\hat{p}_1) \geq 10$ $\quad$ $n_2(1-\hat{p}_2) \geq 10$

TEST STATISTIC

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

This will allow us to find the related p-value using the standard normal distribution.

CONCLUSION As with other tests, the conclusion is based on comparing the P-value to α.

AP Tip

In any two-sample hypothesis test, you MUST be sure to check the standard assumptions for *both* samples. If you don't, you will not get credit for checking assumptions.

A great way to be sure you have checked everything is to use a table to insure that you have included a check for each of the samples on every assumption. Sketch something like the ones shown here and be sure you plug in the actual values or sketches as needed.

MEANS		PROPORTIONS	
Sample 1	Sample 2	Sample 1	Sample 2
Random sample?	Random sample?	Random sample?	Random sample?
Is $n_1 > 30$?	Is $n_2 > 30$?	Check-	
If not, show a plot for normality	If not, show a plot for normality	$n_1p_1 \geq 10$ $n_1(1-p_1) \geq 10$	$n_2p_2 \geq 10$ $n_2(1-p_2) \geq 10$

EXAMPLE An airline wishes to know if the proportion of passengers who would pay $300 extra to upgrade to a first class airline seat is greater for international flights than for flights within the U.S. To investigate, they asked each person in a random sample of passengers on international flights and in a random sample of flights within the U.S. if they would pay extra. The resulting data are summarized in the following table:

Flight Type	***n***	**# who would pay extra**	**Proportion who would pay extra**
International	99	73	$\hat{p}_i = \frac{73}{99} = 0.74$
Within U.S.	91	56	$\hat{p}_w = \frac{56}{92} = 0.62$

let $\hat{p}_i$ = proportion of international passengers who would pay extra

$\hat{p}_w$ = proportion of passengers on flights within U.S. who would pay extra

Is there convincing evidence that the proportion of international passengers who would pay extra is greater than this proportion for passengers on flights within the U.S.? Use a significance level of 0.10. Would the conclusion be different if a significance level of 0.05 had been used?

HYPOTHESIS

$H_0 : p_i - p_w = 0$

$H_a : p_i - p_w > 0$

$\alpha = 0.10$

ASSUMPTIONS

Sample 1	Sample 2
▪ problem states it's a random sample	▪ problem states it's a random sample
▪ $100(0.7) \geq 10$	▪ $92(0.61) \geq 10$
▪ $100(30) \geq 10$	▪ $92(0.39) \geq 10$
So sample size is large enough	So sample size is large enough

TEST STATISTIC

$$p_c = \frac{99(0.74) + 91(0.62)}{190} = 0.68$$

$$z = \frac{0.74 - 0.62}{\sqrt{\frac{0.68(0.32)}{99} + \frac{0.68(0.32)}{91}}} = 1.799$$

P-value $= 0.072$

CONCLUSION At $\alpha = 0.10$: since $0.072 < 0.10$, we would reject the null hypothesis and conclude that there is evidence that the proportion who would pay extra is greater for international passengers. However, at $\alpha = 0.05$: since 0.072 is not smaller than 0.05, we would fail to reject the null hypothesis.

AP Tip

If a question does not specify a significance level, you can select one. Just be sure to state what you have chosen and then use it to reach a conclusion. (Note: 0.05 is often selected.)

A CONFIDENCE INTERVAL

Again, a confidence interval can be used to estimate the difference between two population proportions. The following confidence interval can be used when you have independent random samples and both sample sizes are large. The large sample conditions to be checked are the same as for the two proportions z test.

$$(\hat{p}_1 - \hat{p}_2) \pm (z \text{ critical value})\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Using our previous example of international passengers who would be willing to pay extra for a first class seat and passengers on flights within the U.S. who would pay extra, let's generate a 90% confidence interval estimate of the difference in population proportions. The assumptions were addressed in the hypothesis-testing example, so it is reasonable to proceed.

$$(\hat{p}_i - \hat{p}_w) \pm (z \text{ critical value})\sqrt{\frac{\hat{p}_i(1-\hat{p}_i)}{n_i} + \frac{\hat{p}_w(1-\hat{p}_w)}{n_w}}$$

$$= (0.74 - 0.62) \pm 1.645\sqrt{\frac{0.74(0.26)}{99} + \frac{0.62(0.38)}{91}}$$

$$= 0.12 \pm 1.645\sqrt{0.0019 + 0.0026}$$

$$= 0.12 \pm 1.645(0.0671)$$

$$= 0.12 \pm 0.1104$$

$$= (0.01, 0.23)$$

Notice that the interval does not contain the value of 0. A difference of 0 corresponds to the case of no difference in the population proportions. We are 90% confident that the proportion who would pay extra for a first class seat is greater for international passengers than for passengers on flights within the U.S. by somewhere between 0.01 and 0.23. We used a method that will capture the true difference 90% of the time in repeated sampling.

INTERPRETING AND COMMUNICATING THE RESULTS OF STATISTICAL ANALYSES

(*Introduction to Statistics & Data Analysis* 3rd ed. pages 629–632/4th ed. pages 683–685)

As we saw with the one-sample hypothesis tests, the conclusion is based on the P-value that is generated from our hypothesis test.

Two-sample tests are widely used as many studies are designed to compare two populations or two treatments. With this in mind, there are several things to remember when reading descriptions of statistical studies. Make sure you ask yourself the following questions.

- Are only two groups being compared? If there are more, we will need a different method.
- Are the samples independent or paired?
- What are the hypotheses being tested?
- Is the test appropriate? In other words, have all the needed assumptions been checked?
- If a confidence interval is calculated, has it been correctly interpreted?
- What is the P-value and does it lead us to reject the null hypothesis?
- Are the conclusions made consistent with the test results?

If you keep these questions in mind as you read descriptions of statistical studies or carry out two-sample hypothesis tests, you will avoid many common mistakes.

COMPARING TWO POPULATIONS OR TREATMENTS: STUDENT OBJECTIVES FOR THE AP EXAM

- You will be able to write the null and alternative hypothesis for two-sample tests.
- You will be able to carry out a two-sample hypothesis test for a difference in means.
- You will be able to carry out a two-sample hypothesis test for a difference in proportions.
- You will be able to distinguish between independent samples and paired samples.
- You will be able to construct and interpret a confidence interval for a difference in population means.
- You will be able to construct and interpret a confidence interval for a difference in population proportions.

MULTIPLE-CHOICE QUESTIONS

1. Insurance companies charge higher monthly premiums for male drivers than female drivers between the ages of 16–25. Their rationale is that male drivers have more traffic violations than females. A consumer group believes this rationale is no longer appropriate. To investigate, a random sample of male drivers and a random sample of female drivers in the target age groups were selected. Each person was asked how many traffic violations they had received. The resulting data was used to compute the given summary statistics. A 98% confidence interval for the difference in population mean number of traffic violations is (-1.3, 1.5). Based on this confidence interval, is there evidence that the population means are different?

Gender	***n***	**Mean Violations**	**Standard Deviation**
Male	73	4.2	1.8
Female	56	3.5	2.1

(A) Yes Since the interval contains 0, there is strong evidence that the mean number of traffic violations is greater for males.
(B) No. Since the interval contains 0, it is plausible that the mean number of traffic violations is the same for males and females.
(C) Yes. Since the interval does not contain 0, there is strong evidence that the mean number of traffic violations is greater for males.
(D) No. Since the interval does not contain 0, there is insufficient evidence that the number of traffic violations is the same for males and females.
(E) Yes. Because the sample mean for males is greater than the sample mean for females.

2. In a random sample of 300 elderly men, 65% were married, while in a random sample of 400 elderly women, 48% were married. Which of the following is the 99% confidence interval estimate for the difference between the proportion of elderly men and the proportion of elderly women who are married?
   (A) 0.17 ± 0.073
   (B) 0.17 ± 0.096
   (C) 0.55 ± 0.067
   (D) 0.56 ± 0.067
   (E) 0.565 ± 0.096

3. Which of the following statements are *not* true?
   I. When samples are paired, a two-sample *t* test is appropriate.
   II. The degrees of freedom for a paired *t* test is $n - 1$, where n is the sum of the two sample sizes.
   III. A two-sample *z*-test is used to test hypotheses about $p_1 - p_2$ with large independent samples.
   (A) I only
   (B) II only
   (C) III only
   (D) I and II
   (E) I and III

4. What would be the appropriate hypotheses for a research company who wants to see if there is a difference in the amount of vitamin D in a brand name multi-vitamin and generic brand multi-vitamin?
   (A) $H_0 : \mu_b - \mu_g = 0, \quad H_a : \mu_b - \mu_g > 0$
   (B) $H_0 : \mu_b - \mu_g = 0, \quad H_a : \mu_b - \mu_g < 0$
   (C) $H_0 : \mu_b - \mu_g = 0, \quad H_a : \mu_b - \mu_g \neq 0$
   (D) $H_0 : \pi_b - \pi_g = 0, \quad H_a : \pi_b - \pi_g > 0$
   (E) $H_0 : \pi_b - \pi_g = 0, \quad H_a : \pi_b - \pi_g \neq 0$

Question 5–6 refer to the following set of data:
When a virus is placed on a tobacco leaf, small lesions appear on the leaf. To compare the mean number of lesions produced by two different strains of virus, one strain is applied to half of each of 8 tobacco leaves, and the other strain is applied to the other half of each leaf. The strain that goes on the right half of the tobacco leaf is decided by a coin flip. The lesions that appear on each half are then counted. The data are given below.

Leaf	1	2	3	4	5	6	7	8
Strain 1	31	20	18	17	9	8	10	7
Strain 2	18	17	14	11	10	7	5	6

5. What is the number of degrees of freedom associated with the appropriate t test for testing to see if there is a difference between the mean number of lesions per leaf produced by the two strains?
(A) 7
(B) 8
(C) 11
(D) 14
(E) 16

6. Using a significance level of 0.01, is there statistical evidence that the mean number of lesions is not the same for the two strains?
(A) Yes, $H_0: \mu_1 - \mu_2 = 0, H_a: \mu_1 - \mu_2 \neq 0$, P-value = 0.262
(B) No, $H_0: \mu_1 - \mu_2 = 0, H_a: \mu_1 - \mu_2 \neq 0$, P-value = 0.262
(C) No, $H_0: \mu_1 = \mu_2$, $H_a: \mu_1 \neq \mu_2$, P-value = 0.262
(D) Yes, $H_0: \mu_d = 0$, $H_a: \mu_d \neq 0$, P-value = 0.034
(E) No, $H_0: \mu_d = 0$, $H_a: \mu_d \neq 0$, P-value = 0.034

7. A new classroom strategy uses interactive technology to review information for a test. A random sample of 40 classroom teachers has been asked to use the technology with one of their classes and not with another class. The average test score for each class is then recorded for each teacher. Which would be the appropriate test to run to see if the technology helped?
(A) One-sample z test
(B) One-sample proportions test
(C) Two-sample t test
(D) Paired t test
(E) Two-sample z test

8. Two independent samples were studied, resulting in the following summary statistics:

$\bar{x}_1 = 10,\ s_{\bar{x}_1} = 2.1,\ n_1 = 108$ and $\bar{x}_2 = 15,\ s_{\bar{x}_2} = 2.9,\ n_2 = 78.$

Which of the following could be an appropriate test statistic for testing $H_0 : \mu - \mu = 0$?

(A) $z = \dfrac{10-15}{\sqrt{\dfrac{2.1}{108} + \dfrac{2.9}{78}}}$

(B) $z = \dfrac{10-15}{\sqrt{\dfrac{2.1^2}{108} + \dfrac{2.9^2}{78}}}$

(C) $t = \dfrac{10-15}{\sqrt{\dfrac{2.1}{108} + \dfrac{2.9}{78}}}$

(D) $t = \dfrac{10-15}{\sqrt{\dfrac{2.1^2}{107} + \dfrac{2.9^2}{77}}}$

(E) $t = \dfrac{10-15}{\sqrt{\dfrac{2.1^2}{108} + \dfrac{2.9^2}{78}}}$

9. A soup manufacturer is deciding which company to use for their mushroom purchases. A random sample of 20 mushrooms for each company found 30% of the mushrooms from one company were damaged and 35% from the other company were damaged. What assumptions for the two proportions z test would be a concern?

I. We don't know that they are random samples from both companies.
II. The samples may not be independent.
III. The sample size is not large enough.

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

10. In an experiment to test the effect of alcohol on fine motor skills, volunteers were randomly assigned to one of two groups, A and B. Everyone in group A drank 2 ounces of alcohol, and 20 minutes later everyone in both groups was timed on a manual dexterity test. The average completion times for the group A and B volunteers were 38 and 31 seconds, respectively. The 90% confidence interval estimate for the mean difference is (3,11). If μ_A and μ_B are the true mean completion times, respectively, for people who have and have not drunk 2 ounces of alcohol, how many of the following statements are reasonable conclusions?
    I. $\mu_A - \mu_B$ with probability 0.90
    II. There is a 0.90 probability that $3 < \mu_A - \mu_B < 11$.
    III. The interval (3,11) was calculated using a method that produces an interval that includes $\mu_A - \mu_B$ for 90% of all possible random samples.
    IV. We are 90% confident that $\mu_A - \mu_B$ lies between 3 and 11 seconds.
    (A) None
    (B) One
    (C) Two
    (D) Three
    (E) Four

11. In a test of $H_0 : \mu_1 - \mu_2 = 0$ versus $H_a : \mu_1 - \mu_2 > 0$, the value of the test statistic was $t = 2.34$ and the P-value was 0.01. What conclusion would be appropriate if $\alpha = 0.05$?
    (A) There was not a significant difference in the population means.
    (B) There is convincing evidence that there is a difference in the population means.
    (C) There is convincing evidence that the mean for population 1 is greater than the mean for population 2.
    (D) The proportion of successes in population is greater than the proportion of successes in population 2.
    (E) There is insufficient evidence to conclude that the proportions are different.

Questions 12–13 refer to the following set of data:
Ten men and women were given a supplement for weight loss, and the number of pounds lost by each person was measured at the end of one month. The data is shown below. The investigators would like to know if there is evidence that the mean weight loss for men is greater than the mean weight loss for women.

Subject	1	2	3	4	5	6	7	8	9	10
Males	5	8	12	7	9	11	10	16	8	14
Females	4	9	8	6	11	7	8	12	10	13

12. Which of the following is an appropriate set of hypotheses if μ_m = *men*, μ_w = *women*, μ_d = difference.
(A) $H_0 : \mu_m - \mu_w = 0, \quad H_a : \mu_m - \mu_w \neq 0$
(B) $H_0 : \mu_m - \mu_w = 0, \quad H_a : \mu_m - \mu_w > 0$
(C) $H_0 : \mu_m - \mu_w = 0, \quad H_a : \mu_m - \mu_w < 0$
(D) $H_0 : \mu_d = 0, \quad H_a : \mu_d \neq 0$
(E) $H_0 : \mu_d = 0, \quad H_a : \mu_d > 0$

13. Assuming that the conditions for inference are met, what would be an appropriate conclusion for a significance level of 0.10?
(A) With $t = 1.62$, $df = 9$, and $p = 0.140$, we would fail to reject the null hypothesis. There is no difference in men's and women's mean weight loss.
(B) With $t = 1.62$, $df = 9$, and $p = 0.140$, we would reject the null hypothesis. There is a significant difference in men's and women's mean weight loss.
(C) With $t = 1.62$, $df = 18$, and $p = 0.140$, we would fail to reject the null hypothesis. There is no difference in men's and women's mean weight loss.
(D) With $t = 0.87$, $df = 17$, and $p = 0.197$, we would fail to reject the null hypothesis. There is no difference in men's and women's mean weight loss.
(E) With $t = 0.87$, $df = 17$, and $p = 0.197$, we would reject the null hypothesis. There is a significant difference in men's and women's mean weight loss.

14. A school district wishes to estimate the difference in the proportion of girls who purchase a school lunch and the proportion of boys who purchase a school lunch. Assuming that the conditions for inference are met, which of the following confidence intervals would be appropriate?
(A) A one-sample *t* interval
(B) A two-sample *t* interval
(C) A paired *t* interval
(D) A one proportion *z* test
(E) A two proportions *z* test

15. Time spent studying in a typical week was recorded for each student in a random sample of juniors and a random sample of seniors in a large school district. A 95% confidence interval for the difference in mean time (junior–senior) was (7.4, 14.7). Which of the following is a correct interpretation of this interval?

	n	**Mean Time (minutes)**	**Standard Deviation**
Juniors	152	168	12.5
Seniors	108	157	15.8

(A) I am 95% confident that the sample mean difference is between 7.4 and 14.7 minutes.
(B) I am 95% confident that the true mean difference in time spent studying is between 7.4 and 14.7 minutes.
(C) I am 95% confident that the difference in the proportion of students who study is between 7.4 and 14.7.
(D) I know that 95 out of 100 times the mean difference in time is between 7.4 and 14.7 minutes.
(E) 95 of the sample differences were between 7.4 and 14.7 minutes.

FREE-RESPONSE PROBLEMS

1. Two types of fertilizer for roses are being considered by a housing community for their landscaping needs. The community decided to test the fertilizer on 170 bushes to see if one yielded more rose growth than the other. Each rose bush was assigned at random to one of the two fertilizers. The average growth, in centimeters, for each fertilizer was recorded. Fertilizer A is less expensive and will be used unless there is convincing evidence that mean growth is greater for Fertilizer B. Carry out an appropriate hypothesis test using $\alpha = 0.05$, and make a recommendation as to which fertilizer should be used.

Fertilizer	**Sample Size**	**Mean Growth (cm)**	**Standard Deviation**
Type A	87	12.7	1.5
Type B	83	13.3	2.2

2. A recent study reported that in a random sample of 248 women, 58 had changed their political affiliation since the last election. It also reported that 120 in a random sample of 387 men had changed political affiliation. The researchers would like to know if these data provide convincing evidence that the proportion changing political affiliation is greater for men than for women.
   (a) State the hypotheses of interest.
   (b) Identify the appropriate test and verify the conditions that must be met.
   (c) Is there convincing evidence that the proportion changing political affiliation is greater for men than for women?

Answers

MULTIPLE-CHOICE QUESTIONS

1. **B.** When a two-sample interval is computed, if the value of 0 is within that interval then there appears to be no difference in the two samples. E is also incorrect as it was run as a one-sample confidence interval on males only (*Introduction to Statistics & Data Analysis* 3rd ed. pages 583–597/4th ed. pages 638–656).

2. **B.** The difference between 0.65 and 0.48 is 0.17, making A or B the solution. Answer choice A is for a 95% interval (*Introduction to Statistics & Data Analysis* 3rd ed. pages 619–626/4th ed. pages 671–678).

3. **D.** If the samples are paired, the appropriate test is a paired t test, and df = $n - 1$ where n is the number of pairs (*Introduction to Statistics & Data Analysis* 3rd ed. pages 583–597/4th ed. pages 638–656).

4. **C.** The researchers want to know if there is a difference in means. In this case, the alternative hypothesis would be two-sided (*Introduction to Statistics & Data Analysis* 3rd ed. pages 583–597/4th ed. pages 638–656).

5. **A.** This is a paired t test so the df is calculated as $n - 1$ where n is the number of sample pairs (*Introduction to Statistics & Data Analysis* 3rd ed. pages 583–614/4th ed. pages 638–666).

6. **D.** This would require a matched pairs test. Each virus is being tested on the same tobacco leaf and that makes the samples paired. When $P - value < \alpha$ there is strong evidence in favor of the alternative hypothesis. A and B are incorrect as they use a test for independent samples. Finally, C is incorrect as it is the same as B with the hypotheses written in an alternative format (*Introduction to Statistics & Data Analysis* 3rd ed. pages 583–614/4th ed. pages 638–666).

7. **D.** Each teacher is being assigned a class to use the technology with and a class to not use the technology. The samples are paired

by teacher. (*Introduction to Statistics & Data Analysis* 3rd ed. pages 583–614/4th ed. pages 638–666).

8. **E.** This is a two-sample t test. Notice the denominator must be set up correctly (*Introduction to Statistics & Data Analysis* 3rd ed. pages 583–597/4th ed. pages 638–656).

9. **C.** The sample size is not large enough to meet the conditions of the two proportions z test (*Introduction to Statistics & Data Analysis* 3rd ed. pages 619–626/4th ed. pages 671–678).

10. **C.** Both choice III and IV are reasonable statements based on the confidence interval (*Introduction to Statistics & Data Analysis* 3rd ed. pages 583–597/4th ed. pages 638–656).

11. **C.** Since the p-value is less than 0.05, the null hypothesis would be rejected in favor of the alternative $\mu_1 > \mu_2$ (*Introduction to Statistics & Data Analysis* 3rd ed. pages 583–597/4th ed. pages 638–656).

12. **B.** The samples are independent since there is no indication that the male and female subjects are paired in any meaningful way. Finally, the alternative is $\mu_M - \mu_F > 0$ since we are asking if the mean for men is greater than the mean for women (*Introduction to Statistics & Data Analysis* 3rd ed. pages 583–597/4th ed. pages 638–656).

13. **D.** The appropriate test is based on two independent samples with the alternative being that mean loss for men is greater than mean loss for women. A – C are incorrect as they use the test designed for paired (*Introduction to Statistics & Data Analysis* 3rd ed. pages 583–597/4th ed. pages 638–656).

14. **E.** The samples are independent and we are estimating the difference in population proportions (*Introduction to Statistics & Data Analysis* 3rd ed. pages 619–626/4th ed. pages 671–678).

15. **B.** (*Introduction to Statistics & Data Analysis* 3rd ed. pages 583–597/4th ed. pages 638–656).

FREE-RESPONSE PROBLEMS

1. To answer this question, we will need to carry out an appropriate hypothesis test. In this case, it would be a two-sample t test. The steps are shown here.

HYPOTHESES

$H_0 : \mu_A - \mu_B = 0$

(Note: this could have been written as $H_0 : \mu_A = \mu_B$)

$H_A : \mu_A - \mu_B < 0 \quad \alpha = 0.05$

ASSUMPTIONS

1. The rose bushes were randomly assigned to treatments.
2. Both sample sizes are greater than 30.

TEST STATISTIC Since the population standard deviations are not known, we will use a two-sample t test.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s^2_1}{n_1} + \frac{s^2_2}{n_2}}} = \frac{(12.7 - 13.3) - (0)}{\sqrt{\frac{1.5^2}{87} + \frac{2.2^2}{83}}} = \frac{(12.7 - 13.3) - (0)}{\sqrt{0.026 + 0.058}} - 2.07$$

$$**df = \frac{(V_1 + V_2)^2}{\frac{V_1^2}{n_1 - 1} + \frac{V_2^2}{n_2 - 1}} = \frac{(0.026 + 0.058)^2}{\frac{0.026^2}{86} + \frac{0.058^2}{82}} = \frac{0.007056}{0.000049} = 143$$

$p - value = 0.02$

** Again, we recommend you use the *df* from the calculator.

CONCLUSION Since 0.02 < 0.05, we reject the null hypothesis in favor of the alternative hypothesis. There is convincing evidence that the mean growth for Fertilizer B is greater than for Fertilizer A. Therefore, we would recommend Fertilizer B, even though Fertilizer A is less expensive. (*Introduction to Statistics & Data Analysis* 3rd ed. pages 583–614/4th ed. pages 638–666).

2. (a) The hypotheses would be:

$H_0 : p_w - p_m = 0$

where p_w = proportion of women changing affiliation

p_m = proportion of men changing affiliation

$H_a : p_w - p_m < 0$

(b) The appropriate test is a two-sample proportions z test. The samples were independently selected and both were random samples. Checking to see that the samples sizes are large enough.

$\hat{p}_w = \frac{58}{248} = 0.23$	$\hat{p}_m = \frac{120}{387} = 0.31$
248(0.23) >10	387(0.31) > 10
248(0.67) > 10	387(0.69) > 10

(c) $z = -2.13$, p-value = 0.016. Since $0.016 < 0.05$, we reject the null hypothesis. This means there is evidence that the proportion of men who changed political affiliation is greater than the proportion of women who did so (*Introduction to Statistics & Data Analysis* 3rd ed. pages 619–626/4th ed. pages 671–678).