Chapter 10
STUDY GUIDE
----
1) Reading
2) Practice MC
3) Practice Short Answer
4) Answer Key

HYPOTHESIS TESTING USING A SINGLE SAMPLE

In this section, we will look at the basics of setting up and carrying out a hypothesis test using a univariate data set. Then we will use this information to draw conclusions about some unknown population parameter. Finally, anytime we make a decision based on sample data, there is a risk of error, so we will discuss what types of errors might be made when testing hypotheses.

OBJECTIVES

- Correctly set up and carry out a hypothesis test about a population mean.
- Correctly set up and carry out a hypothesis test about a population proportion.
- Describe Type I and Type II errors in context.
- Understand the factors that affect the power of a test.

HYPOTHESES AND TEST PROCEDURES

(*Introduction to Statistics & Data Analysis* 3rd ed. pages 525–529/4th ed. pages 578–581)

Making decisions based on sample data helps us evaluate claims about a population. Researchers and analysts use hypothesis testing methods in a variety of settings to choose between two competing claims about a population characteristic.

The first step in carrying out a hypothesis test is developing the null and alternative hypotheses. These are statements that will be used in the decision-making process. First, the researcher forms a hypothesis based on some initial claim. For example, suppose an auto manufacturer purchases off-road tires that are supposed to have a mean tread thickness of 0.3125 in. The auto manufacturer will assume the new tires have been manufactured as specified. After all, the tire company wouldn't stay in business for long if they didn't provide what the initial claim about the mean tread thickness that the auto company believes to be fact. This initial assumption is called the null hypothesis. We write the null hypothesis as:

$H_0: \mu = 0.3125$

where,

H_0 stands for "the null hypothesis"

μ is the population mean tread thickness for all tires of this type.

The auto manufacturer may suspect that there has been a change in the mean thickness of the tire tread, so they decide to check several of the tires. This leads the auto company to develop what is called an alternative hypothesis. The alternative hypothesis is a competing hypothesis and could be written in one of the following three ways:

$H_a: \mu \neq 0.3125$ in. *or*

$\mu < 0.3125$ in. *or*

$\mu > 0.3125$ in.

here,

H_a stands for "the alternative hypothesis"

Because the auto manufacturer suspects that the mean tread thickness has changed, but does not have a specific direction in mind, they would use $\mu \neq 0.3125$ in. as the alternative hypothesis.

No matter which alternative hypothesis the company uses, the hypothesis testing procedure only allows us to favor this alternative if there is strong evidence against the null hypothesis. This evidence would come from sample data. We would evaluate what we see in the the sample to determine if the sample mean tire tread is just too far from what the null hypothesis specifies to be explained by just chance differences from sample to sample. This same reasoning is used in all hypothesis tests considered in the AP Statistics course.

The null hypothesis is usually written as

H_0: *some population characteristic* = *the hypothesized value*

and the alternative hypothesis is written as one of the following:

H_a: *some population characteristic* $\neq$ *the hypothesized value*

H_a: *some population characteristic* < *the hypothesized value*

H_a: *some population characteristic* > *the hypothesized value*

EXAMPLE The marketing manager for an online computer game store targets the company advertising toward males because he believes that 75% of the company's purchases are made by men. The sales manager claims that the proportion of purchases made by females has increased. He believes that the proportion of purchases made by men is now less than what the marketing manager believes. What null and alternative hypotheses would be used to test the sales manager's claim?

The appropriate hypotheses are shown below. (Notice that we will now use p to represent the parameter, since the hypotheses are about a proportion.)

$$H_o : p_{males} = 0.75$$
$$H_a : p_{males} < 0.75$$

The null and alternative hypotheses are written in terms of population characteristics. In this example, the alternative is written as "less than" since the sales manager's claim is that the proportion of purchases made by males is less than what the marketing manager believes, 0.75.

ERRORS IN HYPOTHESIS TESTING

(*Introduction to Statistics & Data Analysis* 3rd ed. pages 531–534/4th ed. pages 582–586)

Now that we have an understanding of how to generate the null and alternative hypotheses, a *test procedure* will be used to decide if we should reject the null hypothesis. Test procedures are considered in the next section. Once a decision is made after the test procedure is performed, there is a chance that the final decision is wrong. In other words, an error could have been made. There are two possible types of errors and they are called a *Type I* error and a *Type II* error.

Either types of error may occur when making a decision either to reject or to fail to reject the null hypothesis. For example, in the tire tread problem, if a decision is made to reject the null hypothesis, this could be a wrong decision that would cause the auto manufacturer to conclude that the tires did not meet specifications. However, if the decision was to *fail* to reject the null hypothesis, this could also be wrong and the auto manufacturer could end up using tires that do not meet specifications. In either case, there is a possible error exists that is potentially damaging in some way.

A *Type I* error is made if we reject the null hypothesis and the null hypothesis is actually true. Although the hypothesis test, based on probability, supports the decision, we are led to an incorrect inference about the population. This would amount to having strong enough evidence to conclude that the tires do not have a mean tread thickness of 0.3125 in. The company would decide to return the tires, causing the tire manufacturer to lose money. If the tires actually meet specifications, the tire manufacturer lost money due to the decision error.

A *Type II* error is made if we fail to reject the null and in reality the null hypothesis is not true and should have been rejected. This type of error would amount to not having enough evidence to say the tires did not meet specifications. In this instance, the company would unknowingly use these tires. This error could mean that customers receive cars with faulty tires and this could cause a lawsuit for the company and potentially even the risk of loss of life for the customer.

Avoiding these errors is, of course, desirable. However, it isn't always possible to avoid making an error because we make decisions based on looking only at a sample. What we can do is to try to keep the chance of these errors as small as possible.

We use the symbol α to denote the probability of a Type I error and the symbol β to denote the probability of a Type II error. We can control the value of α by the significance level we select for the test.

Type II error is more problematic as it is something that we can't control easily. The values of α and β are related—the smaller we make α, the larger β becomes, all other things being equal. For this reason, we generally choose a significance level α that is the largest value that is considered an acceptable risk of Type I error. This will help control for the errors by keeping α small as well as controlling a bit for β.

LARGE-SAMPLE HYPOTHESIS TESTS FOR A PROPORTION

(*Introduction to Statistics & Data Analysis* 3rd ed. pages 537–548/4th ed. pages 589–599)

Next, we take the hypotheses we developed and systematically test them to decide whether or not to reject the null hypothesis. This process is known as a test procedure and the same basic procedure is used in the many different hypothesis tests. However, depending on the type of data that we have and the question of interest, there are different hypothesis tests. The first test we consider is a large-sample hypothesis test for a population proportion.

In this case, we are looking at categorical data that come from a single sample, such as the data on the proportion of customers who are male. In this situation, the data consists of observations on a categorical variable with two possible values—male or female.

Just as was the case with a confidence interval, the hypothesis test is based on the properties of the sampling distribution of $\hat{p}$, the sample proportion. Recall that $\hat{p}$ is the sample proportion based on a random sample,

1. $\mu_{\hat{p}} = p$
2. $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$
3. When n is large enough, the sampling distribution of $\hat{p}$ is approximately normal.

It is important to check to make sure the sample size is large enough before carrying out a one-proportion hypothesis test. To verify that the sample size is large enough, check to make sure that

$$np \geq 10$$
$$n(1-p) \geq 10$$

Once we have verified that the sample size is large enough and that the sample is a random sample, we can proceed with the test.

Using the properties of the sampling distribution of $\hat{p}$, we can form a z test statistic:

$$z = \frac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}},$$

where $\hat{p}$ = population proportion.

In the example where we wanted to test $H_0 : p = 0.75$ versus $H_a : p < 0.75$, we consider whether the sample proportion is enough smaller than the hypothesized proportion of 0.75 that the difference can't be explained just by sampling variability. To do this, we calculate the value of the z statistic using 0.75 (from the null hypothesis) as the value for p: $z = \dfrac{\hat{p} - 0.75}{\sqrt{\dfrac{0.75(1-0.75)}{n}}}$

If the null hypothesis is true, then this z statistic will have a standard normal distribution. If the value of the z statistic is something that would be "unexpected" for a standard normal variable, we regard this as evidence that the null hypothesis should be rejected.

EXAMPLE Suppose that the sales manager in the online computer game customer example selects a random sample of 423 previous customers and finds that 298 were males. The sample proportion is then $\frac{298}{423} = 0.70$. We can see that 0.70 is smaller than 0.75, but is it small enough to convince us that chance differences from sample to sample could not account for this difference? This is the question that is answered by a hypothesis test.

First, let's check the assumptions needed. The sample was a random sample of customers, so that condition is met. The second condition is that the sample size is large enough, so we check

$423(0.75) = 317.25 \geq 10$

$423(1-0.75) = 105.75 \geq 10$

Next, we calculate the value of the z test statistic.

$\hat{p} = \frac{298}{423} = 0.704$ so we can now substitute into our test statistic

$$z = \frac{0.704 - 0.75}{\sqrt{\frac{0.75(1-0.75)}{423}}} = -2.1615$$

Because the conditions were met, we know that if H_0 were true, z has an approximately normal distribution. Using what we know about the normal distributions, we know that getting a z score more than +2 or less than -2 does not occur very often. In fact, we can compute the probability of observing a z value as small as -2.16 given that the distribution is standard normal:

$$\begin{aligned} &P(z \leq -2.16 \text{ if } H_o \text{ is true}) \\ &= \text{area to the left of } -2.16 \\ &= 0.0154 \end{aligned}$$

In other words, a sample proportion as small as or smaller than 0.704 would happen only about 1.5% of the time if the population proportion is really 0.75. Based on this sample, it does seem that the null hypothesis is unlikely to be correct, so we would *reject* H_0. The probability just computed is called a *P*-value.

Two key parts of a hypothesis test are the test statistic and the *P*-value.

> *Test statistic*—a value computed from the sample data that is used to make a decision to either reject H_0 or fail to reject H_0
>
> *P-value*—the probability of getting a test statistic value at least as extreme as what was observed for the sample, if the null hypothesis is actually true

If the *P*-value is small, we reject the null hypothesis. How small does the *P*-value have to be in order to reject H_0? This will depend on how large a risk of a Type I error we are willing to assume. This level of risk is preset by the researcher and is known as the significance level of the test. It is denoted by α. For example, if we set $\alpha = 0.05$, this means we know that a result as extreme as what we saw in the sample could happen as often as 5% of the time if the null is true, but we can live with a Type I error occurring for about 5 out of 100 of all possible random samples.

AP Tip

Once you have calculated the value of the test statistic and the associated *P*-value, compare the *P*-value to the value of α.

If

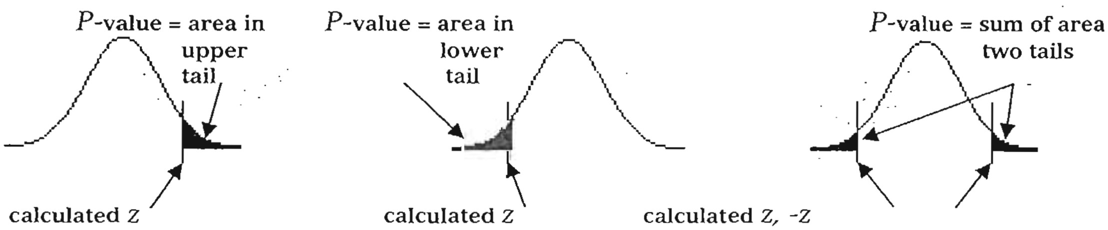
$$P\text{-value} \le \alpha, \quad H_0 \text{ should be rejected.}$$

$$P\text{-value} > \alpha, \quad H_0 \text{ should be not be rejected.}$$

When writing a solution to a hypothesis test problem, make sure to explicitly compare the *P*-value to α. This shows you know how to link the *P*-value and α in making your decision. Then *always* state your conclusion in the context of the problem.

Finally, we note that there are three different computations of the *P*-value to consider. The computation chosen will depend on which inequality (<, >, or ≠) appears in the alternative hypothesis. Here are the three possible situations that can occur.

Upper-tailed test:	**Lower-tailed test:**	**Two-tailed test:**
$H_a: p >$ hypothesized value	$H_a: p <$ hypothesized value	$H_a: p \neq$ hypothesized value



AP Tip

Using incorrect notation can lower your score on free-response questions. Remember to keep the different "p" notations straight.

P-**value** This is a probability computed from the value of the test statistic. *All* hypothesis tests involve the use of a P-value to make a decision.

p This is the notation used to denote the population proportion when we have categorical data.

$\hat{p}$ Read as p-hat, is used to denote the sample proportion.

HYPOTHESIS TESTS FOR A MEAN

(*Introduction to Statistics & Data Analysis* 3rd ed. pages 550–558/4th ed. pages 602–610)

Now that we have a handle on a hypothesis test for proportions, let's consider tests based on numerical data. With numerical data, we are usually interested in making inferences about a population mean. In Chapter 8 on confidence intervals, there were two types of intervals for estimating means, a z *interval* and a t *interval*. Which of these intervals is used in a particular situation is determined by whether we know the population standard deviation (σ) or whether we have only the sample standard deviation (s_x). In either case, we found that if n is large enough or if we know the population is roughly normally distributed, then either

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}, \text{ when we know } \sigma$$

has approximately a standard normal distribution or

$$t = \frac{\bar{x} - \mu}{\frac{s_x}{\sqrt{n}}}, \text{ when we DON'T know } \sigma$$

has a t distribution with $df = n - 1$

In either case, the forms of the hypotheses of interest look very much like those of the proportions test. The difference is that now the hypotheses are in terms of the population mean, so the null hypothesis is

$H_0: \mu =$ hypothesized value

and the alternative hypothesis is one of the following:

$H_a: \mu \neq$ hypothesized value

$H_a: \mu <$ hypothesized value

$H_a: \mu >$ hypothesized value

The test statistic is based on either the z or the t statistic shown above, depending on whether or not we know σ.

If σ is known

$$z = \frac{\bar{x} - \textit{hypothesized value}}{\frac{\sigma}{\sqrt{n}}}$$

P-value: computed as area under the z curve

If σ is NOT known

$$t = \frac{\bar{x} - \textit{hypothesized value}}{\frac{s_x}{\sqrt{n}}}$$

P-value: computed as area under the t curve with $df = n - 1$

It is very rare that the population standard deviation is known, so we will focus mainly on the t statistic. Revisiting the earlier tire tread scenario, we can write the hypotheses, check assumptions, calculate the test statistic and P-value, and give a conclusion in context.

EXAMPLE Trustworthy Tires sells their leading high performance tire to a car manufacturer. The car manufacturer requires that the tires have a mean tread thickness of 0.3125. The car manufacturer thinks tires received from Trustworthy may not be meeting this requirement and that the mean tread may be greater than 0.3125, because they have been finding that in some cars the tires are hitting parts of the wheel well area on bumps. In a random sample of 32 tires, they found the tire treads to have a mean of 0.3625 in. and a standard deviation of 0.094 in. Since both these values came from the sample, we use the notation for sample statistics: $\bar{x} = 0.3360$ and $s_x = 0.094$. For this situation, let's look at the components that are common to all hypothesis tests.

HYPOTHESES

H_0: $\mu = 0.3125$

H_a: $\mu > 0.3125$ (since the company suspects the mean tread is greater than the requirement of 0.3125)

set the significance level (α) at 0.05.

ASSUMPTIONS The assumptions that must be satisfied are the same as those with confidence intervals.

We must have a random sample and the sample size must be large or the population distribution of tread thickness must be approximately normal. The problem states that the sample was a random sample, so that condition is met. The sample size is large (n is greater than 30). Because both conditions are met, it is reasonable to proceed with the test.

TEST STATISTIC

$$t = \frac{\bar{x} - \text{hypothesized value}}{\frac{s_x}{\sqrt{n}}} = \frac{0.3360 - 0.3125}{\frac{0.094}{\sqrt{32}}} = 1.414$$

with $df = n - 1$
$= 32 - 1$
$= 31$

$P - value = 0.08$

Note that because the inequality in the null hypothesis was >, the P-value is the area to the right of 1.41 under the t curve with df = 31.

CONCLUSION In this example, with $\alpha = 0.05$ we see that the P-value is not less than α. We fail to reject the null hypothesis and conclude that there is not convincing evidence that the mean tread thickness of tires is greater than 0.3125 in.

AP Tip

When carrying out a test for means, whether or not you know the standard deviation of the population is what determines if you should use a z *test* or t test. Only use the z test when you are sure the standard deviation given is from the population.

Note that the steps in a test about a population mean are the same as for the test for a population proportion. What distinguishes the two tests is the type of data (numerical for means and categorical for proportions), the specific assumptions that must be checked, and the test statistic used for the test.

SAMPLE PROBLEM 1 A well-known brand of pain relief tablets is advertised to begin relief within 24 minutes. To test this claim, a random sample of 18 subjects suffering from the same types of headache pain record when they first notice relief after taking the pain relief tablet. The data gathered from this study are shown.

Time to Pain Relief (in minutes)

25	24	27	23	25	22	25	25	24
27	24	23	26	28	25	24	26	23

Does the sample suggest the mean pain relief time is longer than the advertised time? Test the appropriate hypotheses using a 0.05 significance level.

SOLUTION TO PROBLEM 1 To solve this problem, we carry out a hypothesis test. All four parts of a hypothesis should be complete as shown.

HYPOTHESIS

With μ representing the population mean time to relief,

H_o: $\mu = 24$

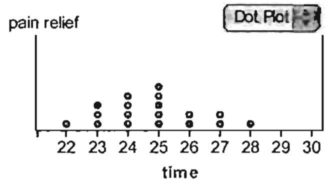
H_a: $\mu > 24$

$\alpha = 0.05$ (*given*)

ASSUMPTIONS

Random sample: The problem states the 18 subjects are a random sample.

Large sample or normal population distribution: Since there were only 18 subjects, we need to be willing to assume that the distribution of relief times in the population is approximately normal. A dotplot of the sample relief times is shown below. Because the dotplot is approximately symmetric and there are no outliers, it is reasonable to think that the population distribution is approximately normal.



TEST STATISTIC A *t* test will be used since we don't know the population standard deviation.

$$\bar{x} = 24.78, \quad s_x = 1.59, \quad df = 17$$

$$t = \frac{24.78 - 23}{\frac{1.59}{\sqrt{18}}} = 2.072$$

$$P\text{-value} = 0.027$$

CONCLUSION Since 0.027 < 0.05, we reject the null hypothesis in favor of the alternative hypothesis. There is convincing evidence that the mean time to pain relief is greater than 24 minutes.

As a final note, should you be in the unusual situation where the population standard deviation (σ), is given, you would use the *z* test statistic and the association *P*-value would be determined using the standard normal distribution. Otherwise, the process would be the same as the process illustrated in the example above.

AP Tip

For a standard hypothesis test, you will be required to provide complete answers to four key parts in every test. Get in the habit of thinking in terms of these four pieces. If you chunk the information under each of these categories, you will have a better chance of remembering all aspects of any hypothesis test. The four pieces are

State the Hypotheses and define any symbols used in the hypotheses.

H_0 and H_a (correctly written)

α level

Identify the test procedure by name or by formula. Check All Assumptions (or Conditions)

The assumptions vary by test, but you should always state AND CHECK the assumption appropriate for the test you are performing.

Test Statistic

Show your z or t calculation

Give the P-value associated with the value of the test statistic

Conclusion *(be sure you state it in context too)*

$P \le \alpha,\ H_0$ should be rejected.

$P > \alpha,\ H_0$ should not be rejected.

POWER AND PROBABILITY OF TYPE II ERROR

(*Introduction to Statistics & Data Analysis* 3rd ed. pages 562–567/4th ed. pages 613–621)

While the AP curriculum does not require you to calculate power, you are expected to know the factors that affect the power of a test. There are three factors that are generally considered when thinking about power. First, increasing α will raise the power. Although this seems like a fast fix, it is dangerous because the probability of a Type I error will also increase. Another way to raise power is simply to increase the sample size, although this isn't always practical. The other things that affect power aren't really things that we can control, but it is helpful to know that they do affect power. The variability in the population affects power, with power being greater when the variability in the population is small. Also, the difference between the actual value of the population characteristic and the hypothesized value affects power. The larger the difference, the greater the power of the test.

SAMPLE PROBLEM 2 Consider the earlier scenario of the online computer game customers. The sales manager was interested in deciding if the proportion of males was less than the 0.75 claimed by the marketing manager. The appropriate hypotheses for this situation were:

$$H_0: p = 0.75$$
$$H_a: p < 0.75$$

(a) Identify the Type I error in this scenario *and* provide a possible consequence of this error.

(b) Identify the Type II error in this scenario and provide a possible consequence of this error.

(c) What can be done to increase the power of this test?

SOLUTION TO PROBLEM 2

(a) A Type I error would result if it were concluded that the proportion of customers who are male is less than 0.75, when in fact this proportion is 0.75. A possible consequence of this error would be that the company might change its strategy of targeting males in its advertising, which might result in a decrease in sales.

(b) A Type II error in this situation would occur by concluding there wasn't enough evidence to say that the proportion of customers who are male is less than 0.75 when this proportion really is less than 0.75. In this case, the company would probably continue to target males in its advertising, which might result in a loss of potential sales to female customers.

(c) One way to increase power would be to increase the significance level of the test. However, this will also increase the chance of a Type I error. Also, increasing sample size will increase power.

INTERPRETATION OF RESULTS IN HYPOTHESIS TESTING

(*Introduction to Statistics & Data Analysis* 3rd ed. pages 571–574/4th ed. pages 623–625)

Once data has been gathered and an appropriate hypothesis test carried out, the findings are typically shared with others interested in the outcome. In communicating results in journals and newspapers, it is not common to provide the same level of detail that you would want to provide in a solution to a hypothesis testing question on the AP exam. Some of the important things to conclude when reporting results are:

- **Hypotheses:** In either symbols or words, you need to clearly state both the null hypothesis and the alternative hypothesis.
- **Test Procedures:** Clearly state what test you used (large sample z test for proportions, and so on) and mention any assumptions that are necessary in order for this test to be appropriate.
- **Test Statistic:** Be sure to report the value of the test statistic as well as the associated P-value. This will allow the readers to know if they would draw the same conclusion given the sample data.
- **Conclusion in Context:** Be certain you have provided a conclusion in terms of the originally posed research question. This needs to

include a comparison of the P-value to α. Stating that you reject the H_0 is *not* sufficient.

In many cases, the reported results only include a statement such as P-value < 0.05. This is common and tells the reader that the results of the test yielded a P-value smaller than 0.05, hence statistically significant. Journals may also use a standard method of coding. * = significant, would mean their P-value was <0.05, ** = very significant, means P-value < 0.01.

As you review published reports, be sure to look for the four key components you would report and ask yourself some questions about these pieces. What were the hypotheses they tested? Did they use an appropriate test for these? What was the associated P-value and what significance level was used? Also, were the conclusions reached consistent with the results of the test?

AP Tip

In writing a conclusion for a hypothesis test, remember you are always either rejecting or failing to reject the null hypothesis. This means you either have convincing evidence in favor of the alternate hypothesis or that there is not enough evidence. You *never say that you accept the null hypothesis* because that implies strong evidence for the null hypothesis.

Hypothesis Testing Using a Single Sample: Student Objectives for the AP Exam

- You will be able to write the null and alternative hypothesis for a test about a population mean or a population proportion.
- You will be able to describe Type I and Type II errors in context.
- You will be able to describe a possible consequence of each type error in context.
- You will be able to carry out a test of hypotheses about a population mean.
- You will be able to carry out a test of hypotheses about a population proportion.
- You will be able to interpret the result of a hypothesis test in context.

MULTIPLE-CHOICE QUESTIONS

1. A psychologist reports that the result of a hypothesis test was statistically significant at the 0.05 level. Which of the following is consistent with this statement?
   (A) The P-value calculated was smaller than the significance level of 0.05.
   (B) The P-value calculated was larger than the significance level of 0.05.
   (C) The significance level calculated was larger than 0.05.
   (D) The significance level calculated was smaller than 0.05.
   (E) There was not enough information to make a decision.

2. A concrete learner is a student who learns best when various types of hands-on or manipulative activities are used to illustrate abstract concepts. Researchers have long believed that 60% of all students remain concrete learners until they are between 16 and 21 years of age. Each student in a random sample of 32 students age 17 to 19 was evaluated, and it was found that 24 of the 32 were concrete learners. Would it be appropriate to use the z test for a population proportion to test to determine if the proportion of concrete learners in this age group is less than 0.60?
   (A) Yes. Since $32(0.6) = 19.2$ and $32(1 - 0.6) = 12.8$, and we can proceed with the test.
   (B) Yes. Since 32 is larger than 30, the sample is sufficiently large and we can proceed with the test.
   (C) Yes. Since we know from the sample was a random sample, we can proceed with the test.
   (D) No. Since $32(0.05) = 1.6$, we do not have a large enough sample to proceed with test.
   (E) No. While 32 is larger than 30, it is so close to 30 and we don't know if the population distribution is normal.

3. A Type I error occurs in which of the following situations?
   (A) H_a is rejected and the null hypotheses is true.
   (B) H_0 is not rejected and the null hypotheses is false.
   (C) H_0 is rejected and the null hypotheses is true.
   (D) The P-value is too small to reject the null hypothesis.
   (E) The α level is too small and so the null hypothesis is rejected.

4. A Type II error occurs in which of the following situations?
   (A) H_0 is rejected and the null hypotheses is true.
   (B) H_0 is not rejected and the null hypotheses is false.
   (C) H_a is not rejected and the null hypotheses is false.
   (D) The P-value is too small to reject the null hypothesis.
   (E) The α level is too large and so the null hypothesis is rejected.

5. A graduate student at a private university wanted to study the amount of money that students at his university carried with them. A recent study reported that the average amount of money carried by college students is $31. He decides to collect data and carry out a test to see if there is evidence that the average is higher for students at his university. Which of the following describes a Type II error in this context?
   (A) This would lead to the incorrect idea that students at his university, on average, spend more money each month than students at other universities.
   (B) This would lead to the incorrect idea that students at his university carry, on average, more than $31.
   (C) This would lead to the incorrect idea that students at his university carry, on average, less than $31.
   (D) This would lead to the incorrect idea that there was no reason to believe that students at his university carry, on average, more than $31.
   (E) This would lead to the correct idea that students at other campuses carry, on average, less than $31.

6. An animal rights group has been very supportive of a new silicon product that caps the nails on cats as an alternative to surgically declawing the pets. The company who makes the caps claims they last for an average of 69 days before needing to be replaced. Before publically endorsing the product, the animal rights group plans to collect data to see if there is convincing evidence that the mean time before replacement is needed is actually less than what the company claims. Which of the following would be an appropriate pair of hypotheses for the animal rights group to test?
   (A) H_o: $\mu = 69$ days, H_a: $\mu > 69$ days
   (B) H_o: $\mu = 69$ days, H_a: $\mu < 69$ days
   (C) H_o: $\mu = 69$ days, H_a: $\mu \neq 69$ days
   (D) H_o: $\bar{x} = 69$ days, H_a: $\bar{x} > 69$ days
   (E) H_o: $\bar{x} = 69$ days, H_a: $\bar{x} < 69$ days

7. Neutering dogs is a common surgical practice. The mean time to recover from the general anesthetic used is 28 hours. A veterinarian believes that since changing to a new anesthetic, the mean recovery time is shorter than before. To investigate, she selects a random sample of 40 surgeries done with the new anesthetic and finds that the mean recovery time was 25 hours and the standard deviation was 2.5. She plans to use this sample data to test to see if there is evidence that the recovery time is shorter with the new anesthetic. Which of the following is the correct test statistic for this study?

(A) $z = \frac{25-28}{2.5}$, with $df = 39$

(B) $t = \frac{25-28}{\frac{2.5}{\sqrt{39}}}$, with $df = 39$

(C) $t = \frac{25-28}{\frac{2.5}{\sqrt{40}}}$, with $df = 39$

(D) $t = \frac{25-28}{\frac{2.5}{\sqrt{40}}}$, with $df = 40$

(E) $t = \frac{25-28}{\frac{2.5}{\sqrt{39}}}$, with $df = 40$

8. A recently published study reported that 63% of the nation's students have some type of structured homework study time. A school surveyed each student in a random sample of 83 students who attend the school and found that only 52% reported having a structured homework time. This data was used to carry out a hypothesis test to determine if there was evidence that the proportion of students at the school who had structured homework time was less that the proportion reported in the national study. Which of the following would be the test statistic for this test?

(A) $z = \dfrac{0.52 - 0.63}{\sqrt{\dfrac{0.52(0.48)}{83}}}$

(B) $z = \dfrac{0.52 - 0.63}{\sqrt{\dfrac{0.52(0.48)}{82}}}$

(C) $z = \dfrac{0.52 - 0.50}{\sqrt{\dfrac{0.52(0.48)}{83}}}$

(D) $z = \dfrac{0.52 - 0.63}{\sqrt{\dfrac{0.63(0.37)}{83}}}$

(E) $z = \dfrac{0.52 - 0.63}{\sqrt{\dfrac{0.63(0.37)}{82}}}$

9. Bicycles purchased from a discount store come unassembled. The assembly instructions that come with the bicycle claim that the average assembly time is 30 minutes. A consumer group has received complaints from people who say that the assembly time was greater than the time claims. They decide to purchase 40 of these bikes and have asked 40 different people to assemble them. The consumer group believed that it was reasonable to regard these 40 people as representative of the population of people who might purchase this bike. For this sample, they found that the assembly times had a mean of 34.2 minutes and a standard deviation of 8.6 minutes. Is there convincing evidence that the claimed average assembly time is too low at the 0.05 significance level?

(A) No, $z = 0.49$ P-value = 0.312.
(B) Yes, $t = 3.09$, df = 39, P-value = 0.002.
(C) Yes, $t = 3.05$, df = 39, P-value = 0.004.
(D) No, $t = 0.49$, df = 39, P-value = 0.313.
(E) Yes, $t = 3.05$, df 40, P-value = 0.002.

10. The prom committee is thinking about changing the location of the prom. The new location is more expensive to rent, and for the increased cost to be reasonable, they would want to be fairly certain that more than 46% of the senior class would attend the prom. A survey of a random sample of 52 seniors found that 25 would attend if the site changed. Which of the following pairs of hypotheses should the prom committee test?
    (A) H_o: $\mu = 46\%$, H_a: $\mu > 46\%$
    (B) H_o: $\mu = 46\%$, H_a: $\mu \neq 48\%$
    (C) H_o: $p = 0.46$, H_a: $p > 0.46$
    (D) H_o: $p = 0.46$, H_a: $p \neq 0.46$
    (E) H_o: $p = 0.46$, H_a: $p < 0.46$

11. Which of the following is closest to the *P*-value associated with a two-tailed *t* test with 20 degrees of freedom if the value of the test statistic is 2.0?
    (A) 0.001
    (B) 0.01
    (C) 0.03
    (D) 0.05
    (E) 0.10

12. Which of the following statements are true?
    I. The null hypothesis for test about a population proportion written as H_0: $\hat{p}$ = hypothesized value.
    II. For the z test to be an appropriate test for a population proportion, the following condition must be met: $np \geq 10$ and $n(1-p) \geq 10$.
    III. The standard deviation of the statistic $\hat{p}$ is $\sigma_{\hat{p}} = \sqrt{np(1-p)}$.
    (A) I only
    (B) II only
    (C) III only
    (D) II and III
    (E) I, II, and III

13. A local group claims that more than 60% of the teens driving after 10 p.m. are exceeding the speed limit. They plan to collect data in hopes that a hypothesis test will provide convincing evidence in support of their claim. Which of the following is true about the hypotheses the group should test?
    (A) The null hypothesis states that less than 60% of the teens are exceeding the speed limit.
    (B) The null hypothesis states that more than 60% of the teens are exceeding the speed limit.
    (C) The alternative hypothesis states that less than 60% of the teens are exceeding the speed limit.
    (D) The alternative hypothesis states that less than or equal to 60% of the teens are exceeding the speed limit.
    (E) The alternative hypothesis states that more than 60% of the teens are exceeding the speed limit.

14. A study by a geological research team found that a new piece of equipment designed to measure the forces of an earthquake is not effective. They based this conclusion on data from a sample of 40 pieces of equipment and they carried out a test with $\alpha = 0.05$. The manufacturer of the equipment claims this study was flawed and that their equipment is good. The research team is considering carrying out a second study with the intention of increasing the power of the test. Which of the following would ensure an increase in the power of the test?
   (A) Move the equipment to three randomly chosen new locations.
   (B) Change $\alpha = 0.05$ to $\alpha = 0.02$.
   (C) Carry out a two-sided test instead of a one-sided test.
   (D) Increase the sample size to 60 pieces of equipment being tested.
   (E) Decrease the sample size to 20 pieces of equipment being tested.

15. Suppose that the mean height of women in the United States is 64.5 in. with a standard deviation of 2.5 in. A clothing designer feels that women who use her products may actually be taller on average. She selects a random sample of 70 women from all women who have previously purchased her clothing. What is the population of interest, and what test would the designer use to test her claim?
   (A)The population is all women in the United States and the appropriate test is a t test with df = 70.
   (B) The population is all women in the United States and the appropriate test is a z test with $= 2.5$.
   (C)The population is all women who have previously purchased the designer's clothing and the appropriate test is a t test with df = 70.
   (D) The population is all women who have previously purchased the designer's clothing and the appropriate test is a t test with df = 69.
   (E) The population is all women who have previously purchased the designer's clothing and the appropriate test is a z test with $= 2.5$.

FREE-RESPONSE PROBLEMS

1. A bridal gown industry publication claims that nationwide the average amount spent for a wedding gown is $1,012. A local bridal shop in an urban community has noticed their more expensive gowns are not selling well. Instead, the brides seem to be selecting only lower priced gowns or clearance gowns. The shop wonders if the average amount spent for a wedding gown is less than $1,012 for their customers. To investigate, they selected a random sample of 50 wedding gown sales. They found a sample mean of $985 and a standard deviation of $235.

   Is there convincing evidence that the average amount spent on a wedding gown at this shop is less than the national figure? Test the relevant hypotheses using a 0.05 significance level.

2. A local school district believes that the proportion of seniors who are absent from school on the last day of school may be increasing. Over the past 5 years, 39% of the seniors have missed the last day. This year, the school district is considering a new reward program sponsored by local businesses where seniors who were at school on the last day would be entered in a drawing for an iPad. To see if this program might reduce the proportion of seniors who miss school on the last day, a random sample of 398 seniors from the school district was surveyed. Each student in the sample was asked if they planned to attend on the last day of class given the possibility of winning an iPad. Only 129 of the 398 seniors indicated that they would miss the last day of school. The school district would like to know if there is convincing evidence that the new program would reduce the number of seniors absent on the last day of school.
   (a) What hypotheses should the school district test?
   (b) Identify the appropriate test and verify that any conditions needed for the test are met.
   (c) Describe Type I and Type II errors in the context of this problem.

Answers

Multiple-Choice Questions

1. **A.** When a researcher says the results were statistically significant, it means the *P*-value was less than the set significance level (*Introduction to Statistics & Data Analysis* 3rd ed. pages 571–574/4th ed. pages 623–625).

2. **A.** In a one-sample proportions test, one condition that is needed is $np \geq 10$ and $n(1 - p) \geq 10$. Notice these both use p and not $\hat{p}$ from the sample (*Introduction to Statistics & Data Analysis* 3rd ed. pages 537–548/4th ed. pages 589–599).

3. **C.** By definition, a Type I error occurs when the null hypothesis is rejected when it should not be rejected. This might happen when the *P*-value < significance level (*Introduction to Statistics & Data Analysis* 3rd ed. pages 531–534/4th ed. pages 582–586).

4. **B.** A Type II error will occur anytime you fail to reject the null when in fact the null is false. This might occur if the *P*-value is not smaller than α (*Introduction to Statistics & Data Analysis* 3rd ed. pages 531–534/4th ed. pages 582–586).

5. **D.** If a Type II error was made, this means he failed to reject the null hypothesis. In this case, that would amount to saying there was not convincing evidence that the mean for his university was greater than $31 when in fact this is incorrect and the mean really is greater than $31 (*Introduction to Statistics & Data Analysis* 3rd ed. pages 531–534/4th ed. pages 582–586).

6. **B.** Since the group is concerned only if the caps last less than 69 days on average, the alternative hypothesis would be $\mu < 69$ days (*Introduction to Statistics & Data Analysis* 3rd ed. pages 525–529/4th ed. pages 578–581).

7. **C.** Since we don't know σ, it must be a t test. Also, while $df = 39$, $n = 40$ is the value that is used in the test statistic calculation (*Introduction to Statistics & Data Analysis* 3rd ed. pages 550–558/4th ed. pages 602–610).

8. **D.** In calculating the test statistic, the denominator uses the hypothesized value and sample size of 83 (*Introduction to Statistics & Data Analysis* 3rd ed. pages 537–548/4th ed. pages 589–599).

9. **B.** A t test with 39 degrees of freedom would be used (*Introduction to Statistics & Data Analysis* 3rd ed. pages 550–558/4th ed. pages 602–610).

10. **C.** This is a test about a population proportion. The question of interest is whether the population proportion p is greater than 0.46 (*Introduction to Statistics & Data Analysis* 3rd ed. pages 525–529/4th ed. pages 578–581).

11. **D.** The P-value is approximately 0.06, which is closest to 0.05 (*Introduction to Statistics & Data Analysis* 3rd ed. pages 550–558/4th ed. pages 602–610).

12. **B.** Notice that choice I is written using $\hat{p}$ instead of p. All hypotheses are stated in terms of the population value, which would be p. Choice III is the standard deviation for a binomial random variable. The standard deviation of $\hat{p}$ is $\sqrt{\frac{p(1-p)}{n}}$ (*Introduction to Statistics & Data Analysis* 3rd ed. pages 537–548/4th ed. pages 589–599).

13. **E.** Answer choices A, B, and D can be eliminated since the null hypothesis must include the equal case. Because the group wants to show support for then claim that *more* than 60% are speeding, the alternative hypothesis would be $p > 0.60$ (*Introduction to Statistics & Data Analysis* 3rd ed. pages 525–529/4th ed. pages 578–581).

14. **D.** The easiest ways to increase power are either to increase the sample size or use a larger significance level (*Introduction to Statistics & Data Analysis* 3rd ed. pages 562–567/4th ed. pages 613–621).

15. **D.** The designer is interested in the population of women who have previously purchased the designer's clothing and the standard deviation of this population is not known (*Introduction to Statistics & Data Analysis* 3rd ed. pages 550–558/4th ed. pages 602–610).

FREE-RESPONSE PROBLEMS

1. **Hypothesis**

   H_0: $\mu = \$1012$

   H_a: $\mu < \$1012$

   $\alpha = 0.05$

   Assumptions

   The problem states this was a random sample of wedding gown sales.

   Since 50 > 30, the sample is large enough for the one sample *t* test to be appropriate.

   Test Statistic

   $\bar{x} = \$985$

   $s_x = \$235$

   $n = 50$

   $$t = \frac{985 - 1012}{\frac{235}{\sqrt{50}}} = -0.81, \quad df = 49$$

   $p = 0.21$

   Conclusion

   Since the *P*-value of 0.21 is not smaller than 0.05, there is not convincing evidence that the average amount spent on a wedding gown at this shop is less than the national average of $1,012

   (*Introduction to Statistics & Data Analysis* 3rd ed. pages 550–558/4th ed. pages 602–610).

2. (a) $H_o: p = 0.39$ where $p =$ proportion of seniors skipping

   $H_a: p < 0.39$

   (b) The appropriate test is a 1-*sample z* test where

   $$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

   The conditions of this test are (1) random sample and (2) large sample. The problem states that the sample was a random sample of seniors. Checking the conditions for large sample: (398)(0.39) = 155, (398)(0.61) = 243. Since both $np \geq 10$ and $n(1-p) \geq 10$, the sample size is large enough.

   (c) A Type I error would be that the iPad drawing would not actually reduce the proportion of seniors who miss the last day of school, but the school district thinks that that it will and implements the drawing.

A Type II error would be that the iPad drawing would in fact reduce the proportion of seniors who miss the last day of school, but the school district is not convinced of this and does not implement the drawing.

(*Introduction to Statistics & Data Analysis* 3rd ed. pages 531–534/4th ed. pages 582–586).