Unit 5: Hypothesis Testing

The Practice of Statistics, 4th edition – For AP*

STARNES, YATES, MOORE

Unit 5: Hypothesis Testing

- 9.1 Significance Tests: The Basics
- 9.2 Tests about a Population Proportion
- 9.3 Tests about a Population Mean
- **9.1 & 9.2 Errors and the Power of a Test**

+

Section 9.1 & 9.2
Errors and the Power of a Test

Learning Objectives

After this section, you should be able to…

- ✓ INTERPRET a Type I error and a Type II error in context, and give the consequences of each.
- ✓ DESCRIBE the relationship between the significance level of a test, *P*(Type II error), and power.

Type I and Type II Errors

When we draw a conclusion from a significance test, we hope our conclusion will be correct. But sometimes it will be wrong. There are two types of mistakes we can make. We can reject the null hypothesis when it's actually true, known as a **Type I error**, or we can fail to reject a false null hypothesis, which is a **Type II error**.

Definition:

If we reject H_0 when H_0 is true, we have committed a **Type I error**.

If we fail to reject H_0 when H_0 is false, we have committed a **Type II error**.

		Truth about the population	
		H_0 true	H_0 false (H_a true)
Conclusion based on sample	Reject H_0	**Type I error**	*Correct conclusion* ***(Power of Test)***
	Fail to reject H_0	*Correct conclusion*	**Type II error**

Example: Perfect Potatoes

A potato chip producer and its main supplier agree that each shipment of potatoes must meet certain quality standards. If the producer determines that more than 8% of the potatoes in the shipment have "blemishes," the truck will be sent away to get another load of potatoes from the supplier. Otherwise, the entire truckload will be used to make potato chips. To make the decision, a supervisor will inspect a random sample of potatoes from the shipment. The producer will then perform a significance test using the hypotheses

$$H_0 : p = 0.08$$
$$H_a : p > 0.08$$

where p is the actual proportion of potatoes with blemishes in a given truckload.

Describe a Type I and a Type II error in this setting, and explain the consequences of each.

- A Type I error would occur if the producer concludes that the proportion of potatoes with blemishes is greater than 0.08 when the actual proportion is 0.08 (or less). *Consequence*: The potato-chip producer sends the truckload of acceptable potatoes away, which may result in lost revenue for the supplier.
- A Type II error would occur if the producer does not send the truck away when more than 8% of the potatoes in the shipment have blemishes. *Consequence*: The producer uses the truckload of potatoes to make potato chips. More chips will be made with blemished potatoes, which may upset consumers.

Error Probabilities

We can assess the performance of a significance test by looking at the probabilities of the two types of error. That's because statistical inference is based on asking, "What would happen if I did this many times?"

For the truckload of potatoes in the previous example, we were testing

$$H_0 : p = 0.08$$
$$H_a : p > 0.08$$

where p is the actual proportion of potatoes with blemishes. Suppose that the potato-chip producer decides to carry out this test based on a random sample of 500 potatoes using a 5% significance level ($\alpha = 0.05$).

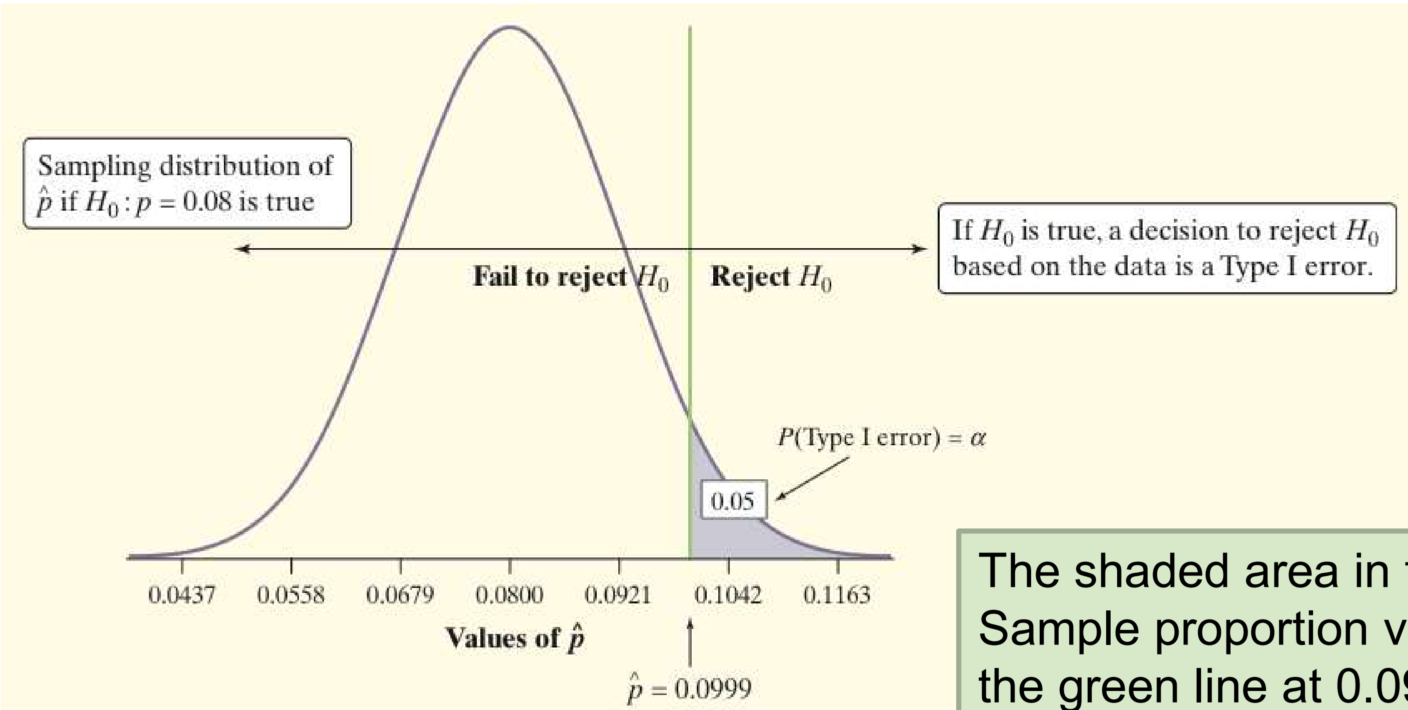
Assuming $H_0 : p = 0.08$ is true, the sampling distribution of $\hat{p}$ will have:

Shape: Approximately Normal because 500(0.08) = 40 and 500(0.92) = 460 are both at least 10.

Center: $\mu_{\hat{p}} = p = 0.08$

Spread: $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.08(0.92)}{500}} = 0.0121$

Error Probabilities



The shaded area in the right tail is 5%. Sample proportion values to the right of the green line at 0.0999 will cause us to reject H_0 even though H_0 is true. This will happen in 5% of all possible samples. That is, *P*(making a Type I error) = 0.05.

Error Probabilities

The probability of a Type I error is the probability of rejecting H_0 when it is really true. As we can see from the previous example, this is exactly the significance level of the test.

Significance and Type I Error

The significance level α of any fixed level test is the probability of a Type I error. That is, α is the probability that the test will reject the null hypothesis H_0 when H_0 is in fact true. Consider the consequences of a Type I error before choosing a significance level.

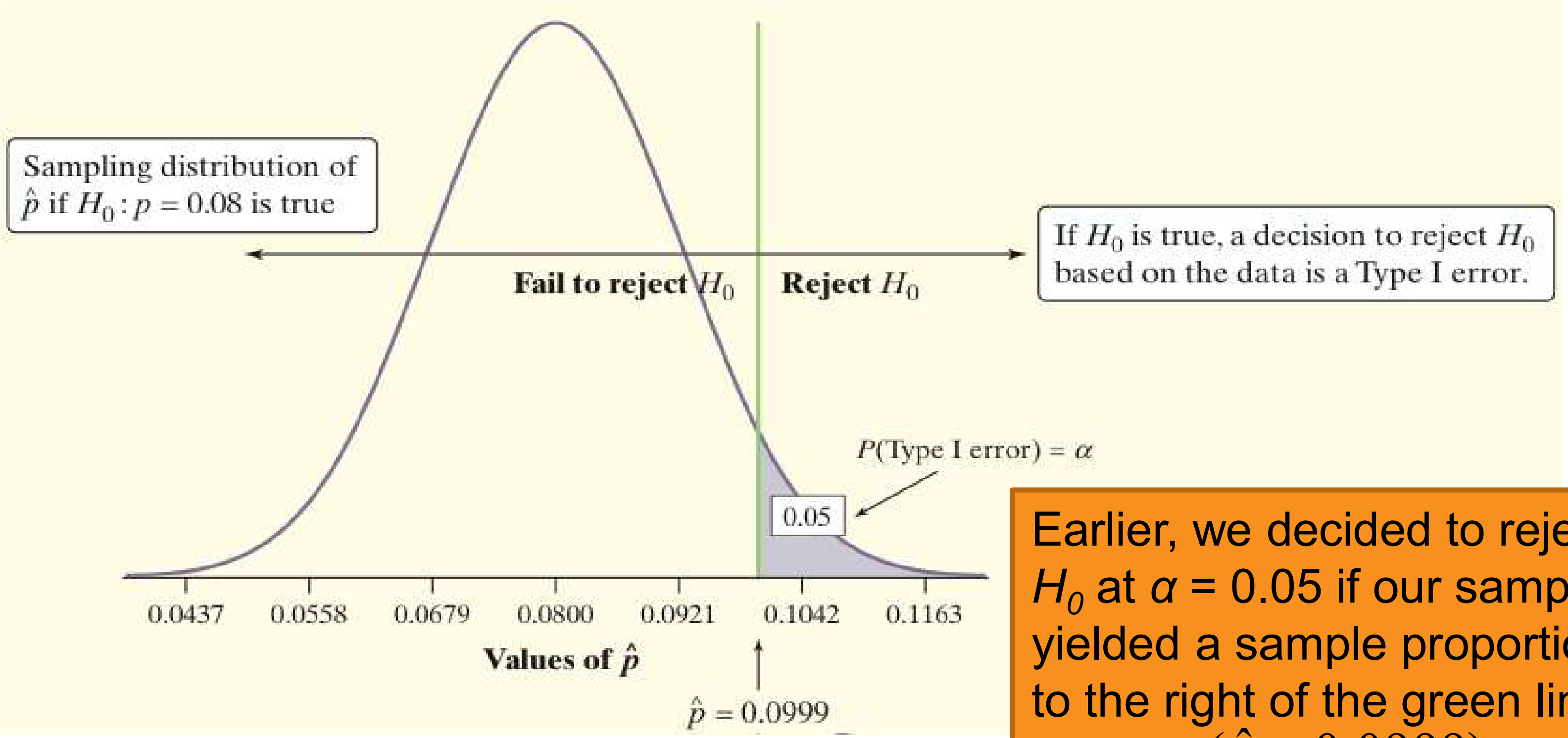
What about Type II errors? A significance test makes a Type II error when it fails to reject a null hypothesis that really is false. There are many values of the parameter that satisfy the alternative hypothesis, so we concentrate on one value. We can calculate the probability that a test *does* reject H_0 when an alternative is true. This probability is called the **power** of the test against that specific alternative.

Definition:

The **power** of a test against a specific alternative is the probability that the test will reject H_0 at a chosen significance level α when the specified alternative value of the parameter is true.

Error Probabilities

The potato-chip producer wonders whether the significance test of $H_0 : p = 0.08$ versus $H_a : p > 0.08$ based on a random sample of 500 potatoes has enough power to detect a shipment with, say, 11% blemished potatoes. In this case, a particular Type II error is to fail to reject $H_0 : p = 0.08$ when $p = 0.11$.

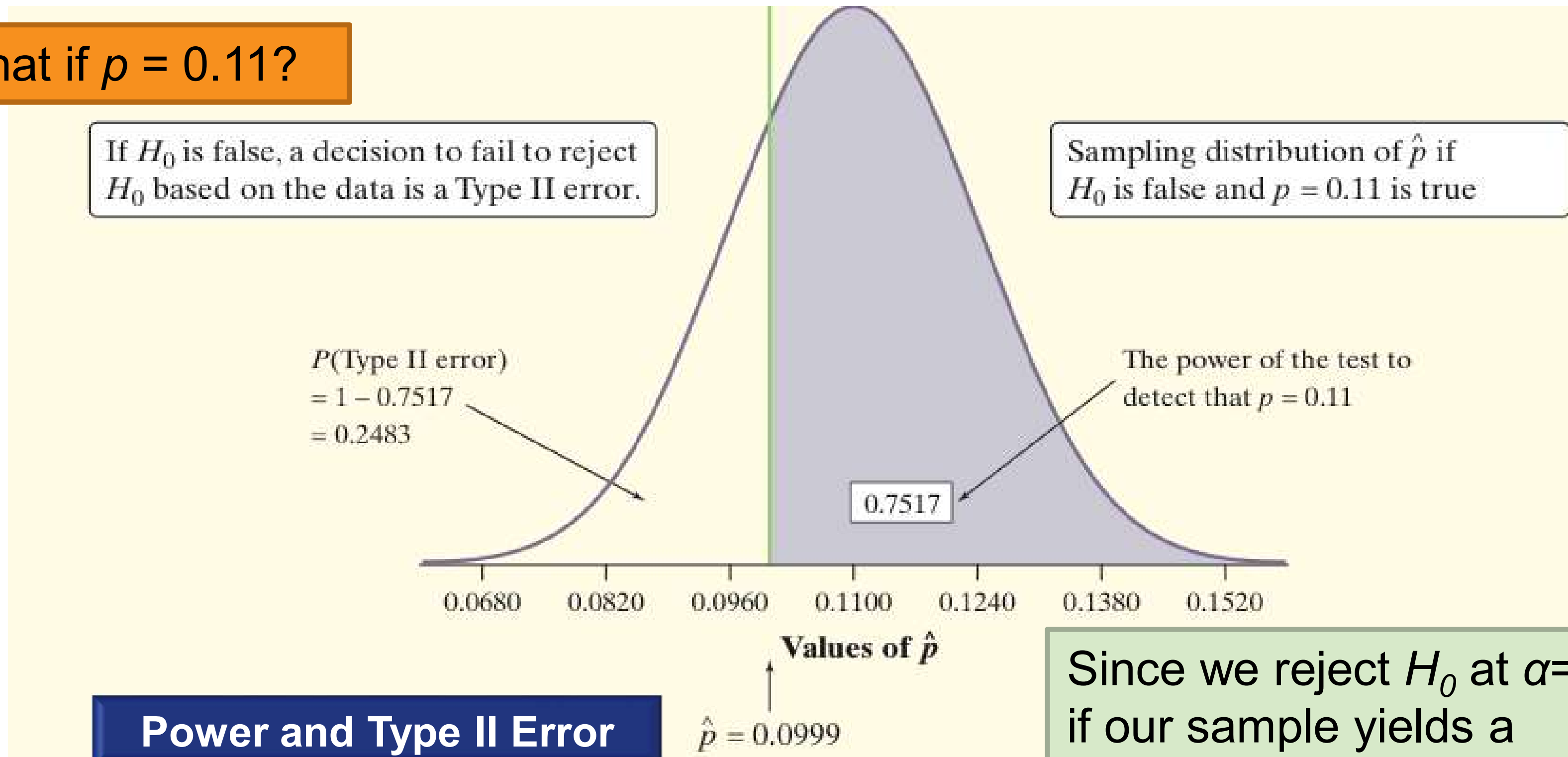


Earlier, we decided to reject H_0 at $\alpha = 0.05$ if our sample yielded a sample proportion to the right of the green line. $(\hat{p} = 0.0999)$

Error Probabilities

The potato-chip producer wonders whether the significance test of $H_0 : p = 0.08$ versus $H_a : p > 0.08$ based on a random sample of 500 potatoes has enough power to detect a shipment with, say, 11% blemished potatoes. In this case, a particular Type II error is to fail to reject $H_0 : p = 0.08$ when $p = 0.11$.

What if $p = 0.11$?



Power and Type II Error

The power of a test against any alternative is 1 minus the probability of a Type II error for that alternative; that is, power = 1 - β.

Since we reject H_0 at $\alpha = 0.05$ if our sample yields a proportion > 0.0999, we'd correctly reject the shipment about 75% of the time.

Planning Studies: The Power of a Statistical Test

How large a sample should we take when we plan to carry out a significance test? The answer depends on what alternative values of the parameter are important to detect.

Summary of influences on the question "How many observations do I need?"

- If you insist on a smaller significance level (such as 1% rather than 5%), you have to take a larger sample. A smaller significance level requires stronger evidence to reject the null hypothesis.
- If you insist on higher power (such as 99% rather than 90%), you will need a larger sample. Higher power gives a better chance of detecting a difference when it is really there.
- At any significance level and desired power, detecting a small difference requires a larger sample than detecting a large difference.

Section 10.4
Errors and the Power of a Test

Summary

- ✓ A **Type I error** occurs if we reject H_0 when it is in fact true. **A Type II error** occurs if we fail to reject H_0 when it is actually false. In a fixed level α significance test, the probability of a Type I error is the significance level α.
- ✓ The power of a significance test against a specific alternative is the probability that the test will reject H_0 when the alternative is true. **Power** measures the ability of the test to detect an alternative value of the parameter. For a specific alternative, P(Type II error) = 1 - power.