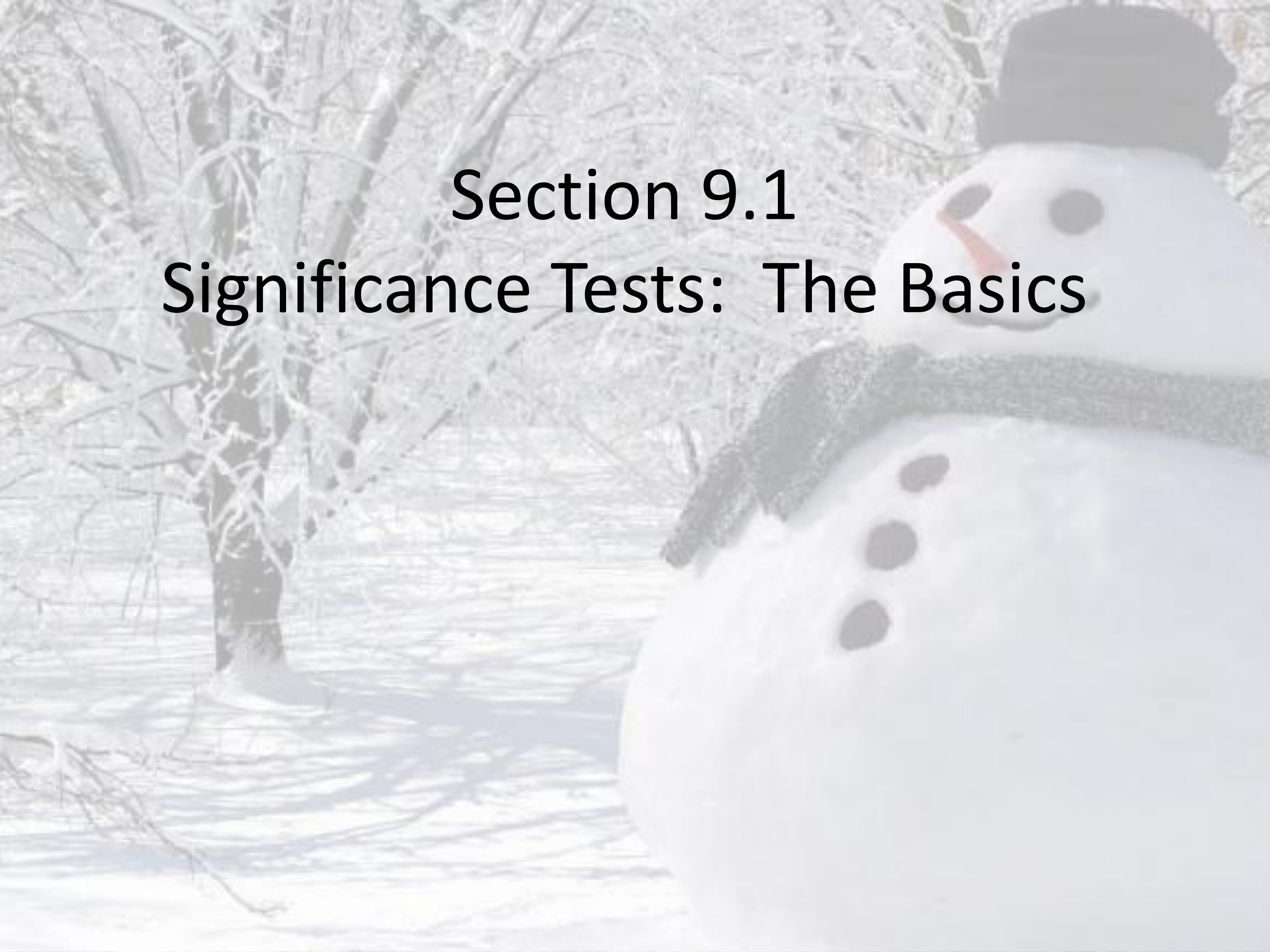


Section 9.1
Significance Tests: The Basics

- Confidence intervals are one of the two most common types of statistical inference.
- Use a confidence interval when your goal is to estimate a population parameter.
- The second common type of inference, called tests of significance, has a different goal: to assess the evidence provided by data about some claim concerning a population.

A **significance test** is a formal procedure for comparing observed data with a claim (also called a hypothesis) whose truth we want to assess. The claim is a statement about a parameter, like the population proportion p or the population mean μ.

We express the results of a significance test in terms of a probability that measures how well the data and the claim agree.

- A basketball player claims to make 80% of the free throws that he attempts. We think he might be exaggerating. To test this claim, we'll ask him to shoot some free throws- virtually- using a simulation.

http://bcs.whfreeman.com/tps4e/#628644__666398__

- When do you have enough data to decide whether the player's claim is valid? How large a sample of shots do you need to make your decision?

Hypotheses

A significance test starts with a statement of the claims we want to compare

- They are the null, H_o, and alternative, H_a, hypotheses.
- The null hypothesis, H_o, is the claim tested by a statistical test. Often the null hypothesis is a statement of “no difference”.
- The alternative hypothesis, H_a, is the claim about the population that we are trying to find evidence for.

In the free-throw shooter example, our hypotheses are

$$H_0 : p = 0.80$$
$$H_a : p < 0.80$$

where p is the long-run proportion of made free throws.

In any significance test, the null hypothesis has the form

$$\boldsymbol{H_0 : \text{parameter} = \text{value}}$$

The alternative hypothesis has one of the forms

$$\boldsymbol{H_a : \text{parameter} < \text{value}}$$
$$\boldsymbol{H_a : \text{parameter} > \text{value}}$$

One-sided alternative hypothesis

$$\boldsymbol{H_a : \text{parameter} \neq \text{value}}$$

Two-sided alternative hypothesis

To determine the correct form of H_a, read the problem carefully. ****Always state H_0 and H_a in terms of population parameters**

Anemia

Hemoglobin is a protein in red blood cells that carries oxygen from the lungs to body tissues. People with less than 12 grams of hemoglobin per deciliter of blood (g/dl) are anemic. A public health official in Jordan suspects that Jordanian children are at risk of anemia. He measures a random sample of 50 children.

a) Describe the parameter of interest in this setting.

The parameter of interest is the true mean μ amount of hemoglobin in Jordanian children.

b) State appropriate hypotheses for performing a significance test.

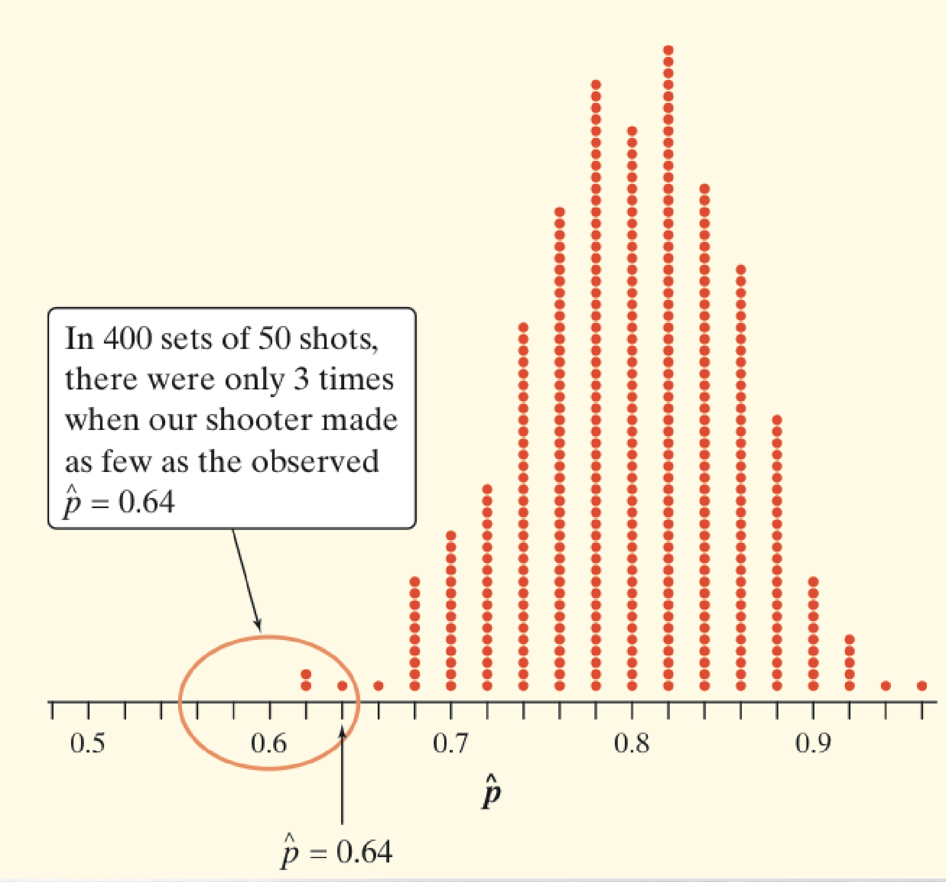
Because we are only concerned if Jordanian children have lower than 12 g/dl of hemoglobin, this will be one-sided. That is,

$$H_0: \mu = 12$$
$$H_a: \mu < 12$$

- Looking at the basketball player example, he attempts 50 free-throws. He makes 32 of them. His sample proportion is 32/50 = 0.64
- We can simulate 400 sets of 50 shots assuming that the player is really an 80% shooter.

You can say how strong the evidence against the player's claim is by giving the probability that he would make as few as 32 out of 50 free throws if he really makes 80% in the long run.

The observed statistic is so unlikely if the actual parameter value is p = 0.80 that it gives convincing evidence that the player's claim is not true.



The Reasoning of Significance Tests

The observed statistic is so unlikely if the actual parameter value is p = 0.80 that it gives convincing evidence that the players claim is not true.

There are two possible explanations for the fact that he made only 64% of his free throws.

1) **The null hypothesis is correct.** The player's claim is correct (p = 0.8), and just by chance, a very unlikely outcome occurred.

2) **The alternative hypothesis is correct.** The population proportion is actually less than 0.8, so the sample result is not an unlikely outcome.

Basic Idea

An outcome that would rarely happen if the null hypothesis were true is good evidence that the null hypothesis is not true.

P-values

- The probability that measures the strength of the evidence against a null hypothesis is called a *P-value*. The probability, computed assuming H_0 is true, that the statistic would take a value as extreme as or more extreme than the one actually observed.

  ✓ Small *P*-values are evidence against H_0 because they say that the observed result is unlikely to occur when H_0 is true.

  ✓ Large *P*-values fail to give convincing evidence against H_0 because they say that the observed result is likely to occur by chance when H_0 is true.

Statistical Significance

- The final step in performing a significance test is to draw a conclusion about the competing claims you were testing. We will either decide to reject H_o or fail to reject H_o.

Note: A fail-to-reject H_0 decision in a significance test doesn't mean that H_0 is true. For that reason, you should never "*accept* H_0" or use language implying that you believe H_0 is true.

- There is no rule for how small the *P*- value needs to be in order to reject H_o. We can compare the *P*-value with a fixed value that we regard as decisive, called the **significance level, α**.

When we use a fixed level of significance to draw a conclusion in a significance test,

P-value < α → reject H_0 → convincing evidence for H_a

P-value ≥ α → fail to reject H_0 → not convincing evidence for H_a

- If we choose α = 0.05, we are requiring that the data give evidence against H_o so strong that it would happen no more than 5% of the time (1 time in 20 in the long run) when H_o is true.

Statistical significance

- If the P-value is as small or smaller than α, we say that the data are statistically significant at level α.
- “Significant” in the statistical sense does not mean “important”. It means “not likely to happen by chance”.
- The P-value is more informative than a statement of significance because it allows us to assess significance at any level we choose.
  - A result with $P=0.03$ is significant at the $\alpha=0.05$ level but not significant at the $\alpha=0.01$ level.

- The conclusion to a significance test should always include three components:
  - An explicit comparison of the *P*-value to a stated significance level
  - A decision about the null hypothesis: reject or fail to reject H_o.
  - An explanation of what the decision means in context.

For his second semester project in AP® Statistics, Zenon decided to investigate whether students at his school prefer name-brand potato chips to generic potato chips. He randomly selected 50 students and had each student try both types of chips, in random order. Overall, 32 of the 50 ($\hat{p}$ = 0.64) students preferred the name-brand chips. Zenon performed a significance test using the hypotheses

$$H_0 : p = 0.5$$
$$H_a : p > 0.5$$

where p = the true proportion of students at his school who prefer name-brand chips. The resulting *P*-value was 0.0239.

Problem: What conclusion would you make at each of the following significance levels?

(a) $\alpha = 0.05$

Because the *P*-value of 0.0239 is less than $\alpha = 0.05$, we reject H_0. There is convincing evidence that students at Zenon's school prefer name-brand chips.

(b) $\alpha = 0.01$

Because the *P*-value of 0.0239 is greater than $\alpha = 0.01$, we fail to reject H_0. There is not convincing evidence that students at Zenon's school prefer name-brand chips.

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.

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_a is true, 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*
	Fail to reject H_0	*Correct conclusion*	**Type II error**

Type I and Type II Errors

The probability of a Type I error is the probability of rejecting H_0 when it is really true…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 actually true.

Consider the consequences of a Type I error before choosing a significance level.

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.

The manager of a fast-food restaurant wants to reduce the proportion of drive-through customers who have to wait longer than two minutes to receive their food after their order is placed. Based on store records, the proportion of customers who had to wait longer than two minutes was $p = 0.63$. To reduce this proportion, the manager assigns an additional employee to assist with drive-through orders. During the next month, the manager will collect a random sample of drive-through times and test the following hypotheses:

$$H_0: \; p = 0.63$$

$$H_a: \; p < 0.63$$

where p = the true proportion of drive-through customers who have to wait longer than two minutes to receive their food. Describe a Type I and a Type II error in this setting and explain the consequences of each.

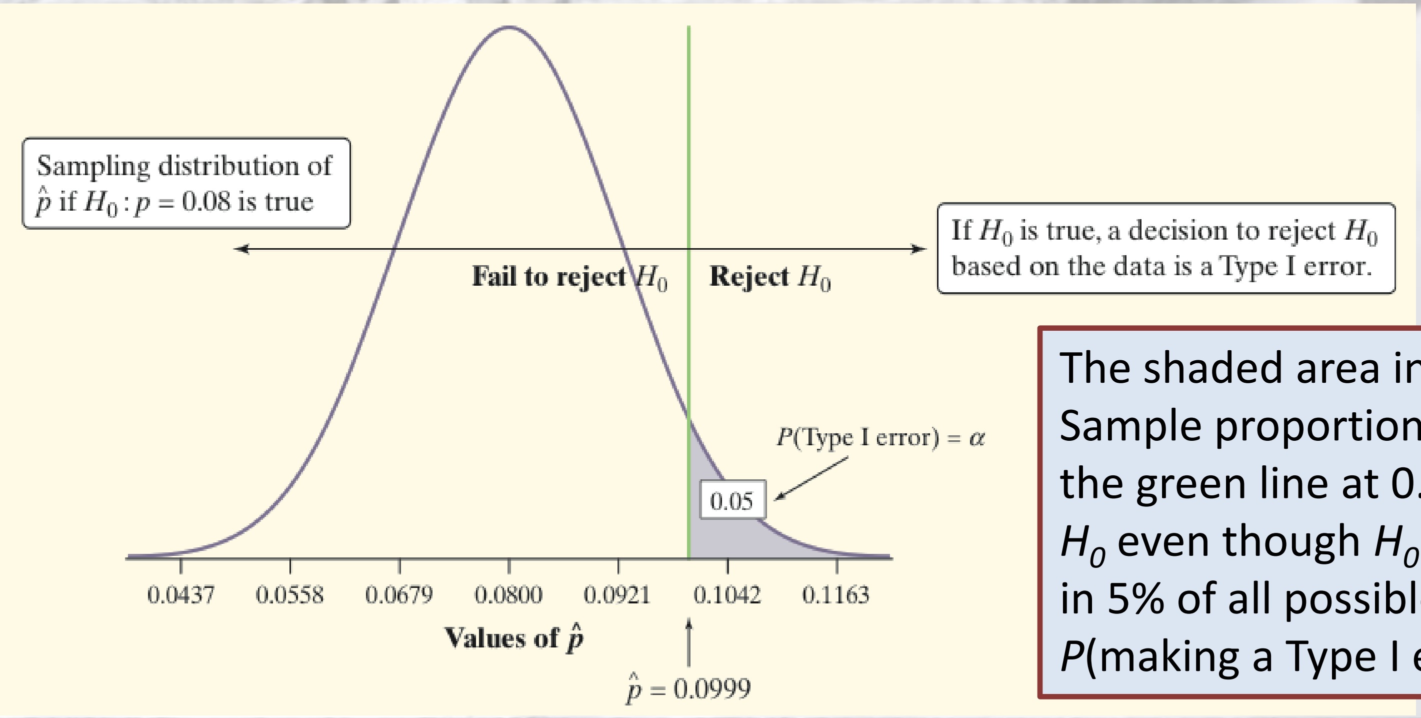
A Type I error would occur if the manager finds convincing evidence that the true proportion of drive-through customers who have to wait longer than two minutes has been reduced, when it really hasn't been reduced. A consequence is that the manager will have to pay unnecessarily for an additional employee.

A Type II error would occur if the manager didn't find convincing evidence that the true proportion of customers who had to wait longer than two minutes had been reduced, when it really had been reduced. A consequence is that the restaurant wouldn't be serving customers more quickly when it could.

For the truckload of potatoes in the other 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 ($\alpha = 0.05$). A Type I error is to reject H_0 when H_0 is actually true. If our sample results in a value of $\hat{p}$ that is much larger than 0.08, we will reject H_0. How large would $\hat{p}$ need to be?



The shaded area in the right tail is 5%. Sample proportion values to the right of the green line at 0.1 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.

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.