Significance Tests: The Basics

- ✓ STATE the null and alternative hypotheses for a significance test about a population parameter.
- ✓ INTERPRET a *P*-value in context.
- ✓ DETERMINE whether the results of a study are statistically significant and MAKE an appropriate conclusion using a significance level.
- ✓ INTERPRET a Type I and a Type II error in context and GIVE a consequence of each.

Introduction

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 significance tests, 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.

Activity: I'm a Great Free-Throw Shooter!

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 *The Reasoning of a Statistical Test* applet at the book's Web site.

1. Launch the applet.
2. Set the applet to take 25 shots. Click "Shoot." Record how many of the 25 shots the player makes.
3. Click "Shoot" again for 25 more shots. Repeat until you are convinced *either* that the player makes less than 80% of his shots *or* that the player's claim is true.
4. Click "Show true probability." Were you correct?

Stating a hypothesis

- A statistical test starts with a careful statement of the claims we want to compare. Because the reasoning of tests look for evidence *against* a claim, we start with the claim we seek evidence AGAINST.
  - This is called our **NULL HYPOTHESIS. (H_0)**
  - The claim about the population we are trying to find evidence FOR is our **ALTERNATIVE hypothesis (H_A)**

In our basketball example:

You are field-testing a new-flavor soft-drink. The company only plans to market the drink if more than 60% of consumers like the flavor.

- *We seek evidence AGAINST the claim that μ = 115.*
- Null: H_0:
- Alternate: H_A:

You are field-testing a new recipe for a soft-drink. Based on past studies, you know that the current recipe is liked by 70% of consumers. The company would like to know if consumer opinions regarding this new recipe are different.

- H_0:
- H_a:

Stating Hypotheses

In any significance test, the null hypothesis has the form

H_0 : parameter = value

The alternative hypothesis has one of the forms

H_a : parameter < value
H_a : parameter > value
H_a : parameter ≠ value

To determine the correct form of H_a, read the problem carefully.

The alternative hypothesis is **one-sided** if it states that a parameter is *larger than* the null hypothesis value or if it states that the parameter is *smaller than* the null value.

It is **two-sided** if it states that the parameter is *different from* the null hypothesis value (it could be either larger or smaller).

Stating Hypotheses

- ✓ The hypotheses should express the hopes or suspicions we have **before** we see the data. It is cheating to look at the data first and then frame hypotheses to fit what the data show.
- ✓ Hypotheses always refer to a *population*, not to a sample. Be sure to state H_0 and H_a in terms of *population parameters*.
- ✓ It is *never* correct to write a hypothesis about a sample statistic, such as $\hat{p} = 0.64$ or $\bar{x} = 85$.

So let's say that I tell you guys that I make 80% of my basketball freethrows.

To test my claim, you ask me to shoot 20 free throws. I only make 8 ☹ and you say "aha! Someone who makes 80% of their free throws would NEVER only make 10/20!"

•But I say, 'Hey, what if I'm having a bad day! 80% is a long-run percentage. Didn't you listen to me when I taught sampling variablity?" Since that's true, you say "OK, well **I'll decide how likely your claim is based on the probability that someone who genuinely makes 80% would shoot 10/20 on one trial run**"

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 I made only 50% 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.

Interpreting *P*-Values

The null hypothesis H_0 states the claim that we are seeking evidence against. 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 is called the ***P*-value** of the test.

✓ 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.

Example:

A driving school company claimed that 90% of their students passed the driving test on their first try. Data from a reporter's survey of randomly selected local teens who used this driving school produces a p-value of 0.185

(1) Write a pair of hypotheses for the reporter

(2) Interpret the p-value.

(3) Based on the p-value, what conclusion can you make?

Interpretation: Statistical Significance

- We set a "maximum" p-value before calculating our observed test statistic and we call this our **significance level.**
  - α is the symbol for our chosen significance level we need to beat. Most commonly we choose α =.05
    - meaning we need our calculated p value to be less than .05,
- If our P-value is less than our significant level – Reject H_0 (We have **statistically significant results**)
- If our P-value is less than our significant level – Fail to Reject H_0 (We have **do not have** statistically significant results)

Statistically Significant? Reject H_0?	P-value = 0.047
$\alpha = 0.01$	
$\alpha = 0.05$	
$\alpha = 0.1$	

Type I Errors

- The Environmental Protection Agency has determined that safe drinking water should contain no more than 1.3 mg/liter of copper. You have been hired to test water from a new source, which would provide needed drinking water during a serious drought, in order to determine if the water is safe to drink.

- μ = the true mean level of copper (mg/liter) in this particular water source.
- Ho: $\mu = 1.3$
- Ha: $\mu < 1.3$

- You run a test and find a p-value of 0.049. What does this mean?

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**

- μ = the true mean level of copper (mg/liter) in this particular water source.
- Ho: $\mu = 1.3$
- Ha: $\mu < 1.3$

		TRUTH ABOUT Ho	
		Ho is true	Ho is False (Ha is true)
Decision about Ho	Reject Ho	TYPE I ERROR	**Correct Decision**
	Fail to reject Ho	**Correct Decision**	TYPE II ERROR

TYPE I ERROR

Inform the EPA that the water is safe to drink when it is actually UNSAFE!

Consequence?

People drink unsafe water and could get sick ☹

TYPE II ERROR

Inform the EPA that the water is not safe to drink even though it really WAS safe.

Consequence?

People are unable to use this valuable water source and the government imposes water restrictions ☹