Chapter 6

Introduction to Inference

Introduction to the Practice of

STATISTICS EIGHTH EDITION

Moore / McCabe / Craig

Lecture Presentation Slides

Chapter 6
Introduction to Inference

6.1 Estimating with Confidence

6.2 Tests of Significance

6.3 Use and Abuse of Tests (On Your Own)

6.4 Power and Inference as a Decision

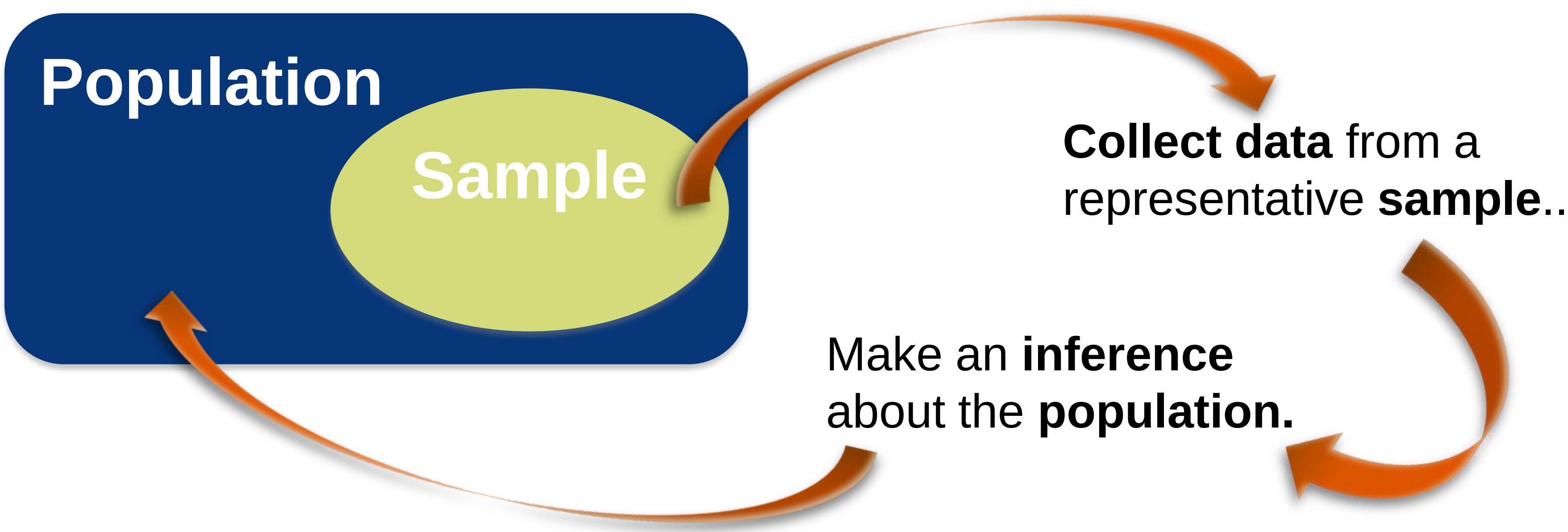
6.1 Estimating with Confidence

- Inference
- Statistical confidence
- Confidence intervals
- Confidence interval for a population mean
- Choosing the sample size

Statistical Inference

After we have selected a sample, we know the responses of the individuals in the sample. However, the reason for taking the sample is to infer from that data some conclusion about the wider population represented by the sample.

Statistical inference provides methods for drawing conclusions about a population from sample data.



Simple Conditions for Inference About a Mean

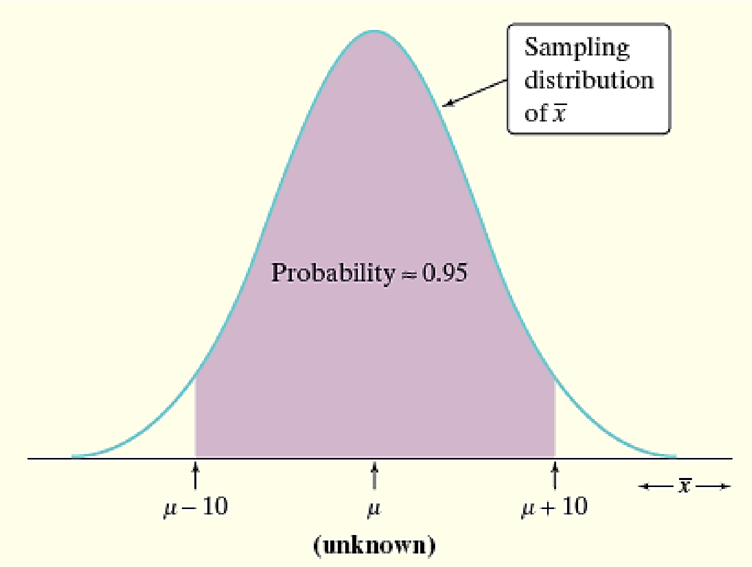
This chapter presents the basic reasoning of statistical inference. We start with a setting that is too simple to be realistic.

Simple Conditions for Inference About a Mean

1. We have an SRS from the population of interest. There is no nonresponse or other practical difficulty.
2. The variable we measure has an exactly Normal distribution $N(\mu,\sigma)$ in the population.
3. We don't know the population mean μ, but we do know the population standard deviation σ.

Note: The conditions that we have a perfect SRS, that the population is exactly Normal, and that we know the population standard deviation are all unrealistic.

Statistical Estimation

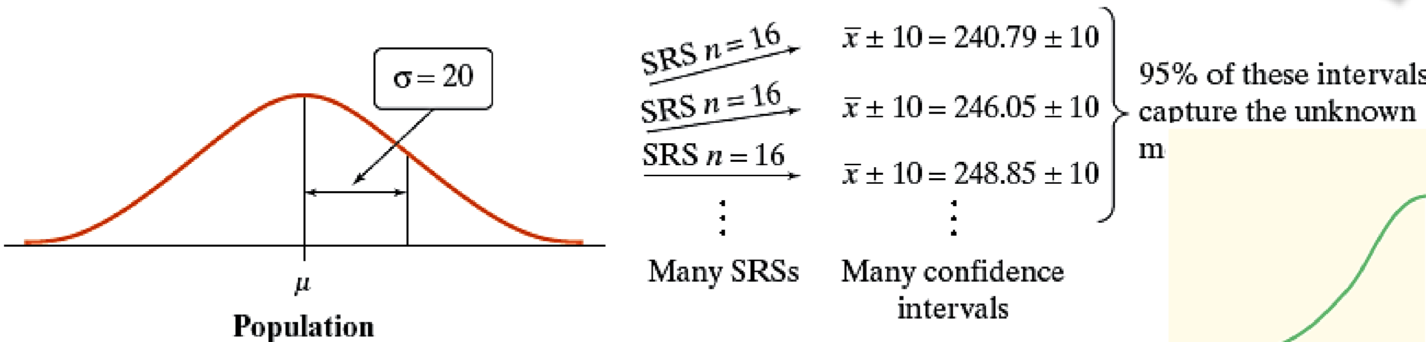


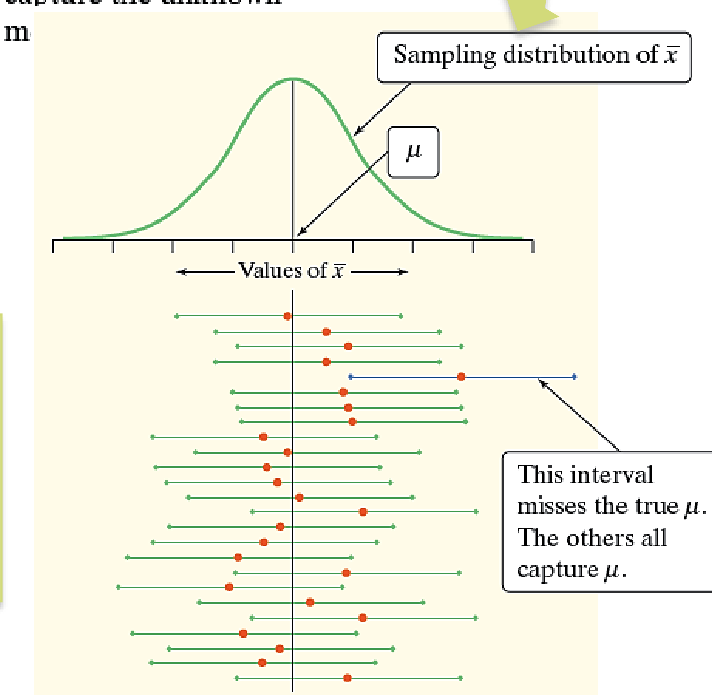
- ✓ In repeated samples, the values of the sample mean will follow a Normal distribution with unknown mean μ and known standard deviation $\sigma_{\bar{x}} = 5$.
- ✓ The 68-95-99.7 rule tells us that in roughly 95% of all samples of size 16, the sample mean will be within 10 (two standard deviations) of μ.
- ✓ If the sample mean is within 10 points of μ, then μ is within 10 points of the sample mean.
- ✓ Therefore, the interval from 10 points below to 10 points above the sample mean will "capture" μ in about 95% of all samples of size 16.

If we estimate that μ lies somewhere with sample mean $\pm$ 10**,** we'd be using a method that captures the true μ in about 95% of all possible samples of this size.

Confidence Level

The confidence level, or *C*, is the overall capture rate if the method is used many times. The sample mean will vary from sample to sample, but when we use the method "*estimate* ± *margin of error*" to get each interval, *C* × 100% of all intervals capture the unknown population mean μ. (**Note:** By convention, *C* will be expressed as a decimal, so it will always be between 0 and 1.)





Interpreting a Confidence Level

To say that we are 95% *confident* is shorthand for "95% of all possible samples of a given size from this population will yield intervals that capture the unknown parameter."

Confidence Interval

The Big Idea: The sampling distribution of $\bar{X}$ tells us how close to μ the sample mean $\bar{X}$ is likely to be. All confidence intervals we construct will have a form similar to this:

$$\textbf{estimate} \pm \textbf{margin of error}$$

A **level C confidence interval** for a parameter has two parts:

- An **interval** calculated from the data, which has the form:

  estimate $\pm$ margin of error

- A ***confidence level C***, where *C* is the probability that the interval will capture the true parameter value in repeated samples. In other words, the confidence level is the success rate for the method.

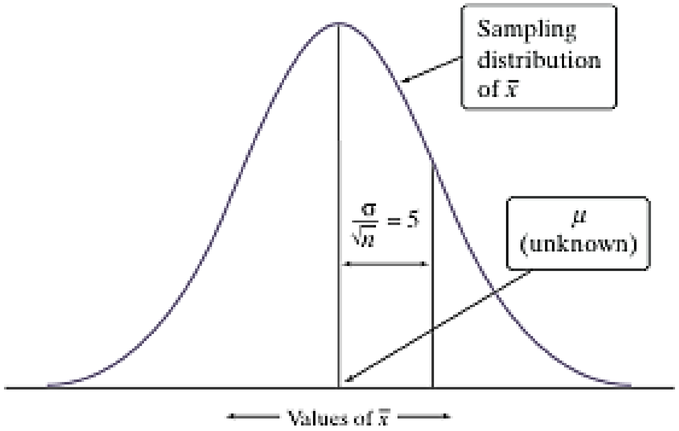
We usually choose a confidence level of 90% or higher because we want to be quite sure of our conclusions. The most common confidence level is 95%.

Confidence Interval for a Population Mean

Previously, we estimated the unknown mean μ by constructing a confidence interval using the sample mean = 240.79.

To calculate a 95% confidence interval for μ, we use the formula:

estimate $\pm$ (critical value) • (standard deviation of statistic)



$$\bar{x} \pm z^* \times \frac{\sigma}{\sqrt{n}} = 240.79 \pm 1.96 \times \frac{20}{\sqrt{16}}$$
$$= 240.79 \pm 9.8$$
$$= (230.99, 250.59)$$

Confidence Interval for the Mean of a Normal Population

Choose an SRS of size *n* from a population having unknown mean μ and known standard deviation σ. A level **C confidence interval for μ** is:

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

The critical value z^* is found from the standard Normal distribution.

Finding Specific z^* Values

We can use a table of *z/t* values (Table D). For a particular confidence level, *C*, the appropriate z^* value is just above it.

z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level *C*											

<u>Example</u>: For a 98% confidence level, $z^* = 2.326$.

Note that using $Z = 2$ for a 95% confidence level is just a quick approximation. The exact z^* is actually __.

The Margin of Error

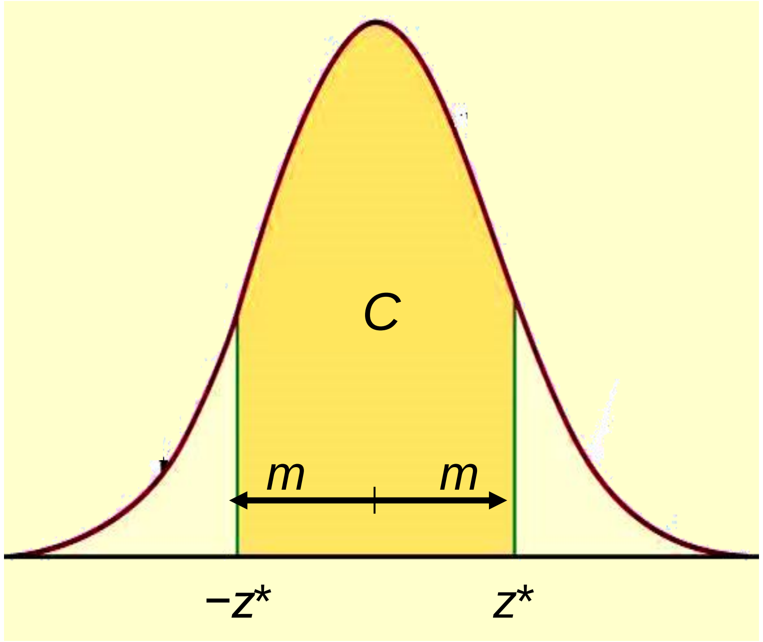
The confidence level *C* determines the value of *z** (in Table D).

The margin of error also depends on *z**.

$$m = z^* \sigma / \sqrt{n}$$

Higher confidence ***C*** implies a larger margin of error ***m*** (thus less precision in our estimates).

A lower confidence level ***C*** produces a smaller margin of error ***m*** (thus better precision in our estimates).



How Confidence Intervals Behave

The z confidence interval for the mean of a Normal population illustrates several important properties that are shared by all confidence intervals in common use.

- The user chooses the confidence level and the margin of error follows.
- We would like high confidence and a small margin of error.
  - High confidence suggests our method almost always gives correct answers.
  - A small margin of error suggests we have pinned down the parameter precisely.

The margin of error for the z confidence interval is:

$$z * \times \frac{\sigma}{\sqrt{n}}$$

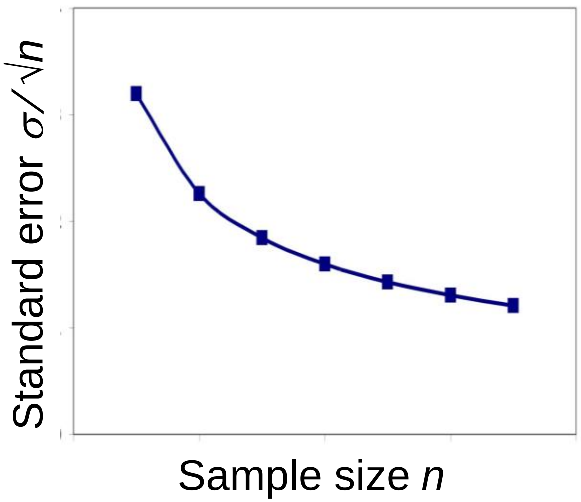
The margin of error gets smaller when:

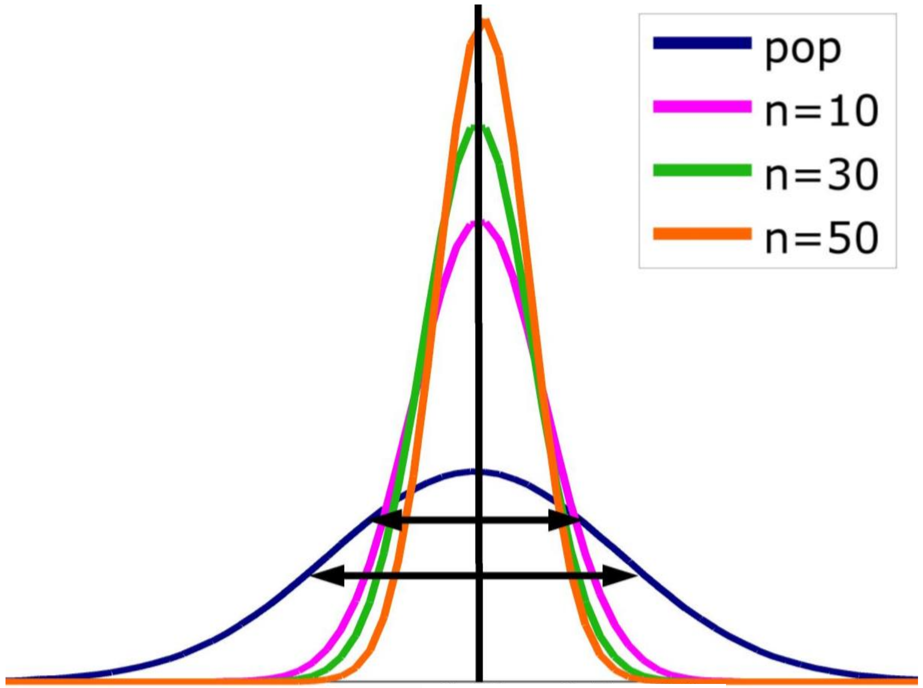
- z^* gets smaller (the same as a lower confidence level C).
- σ is smaller. It is easier to pin down μ when σ is smaller.
- n gets larger. Since n is under the square root sign, we must take four times as many observations to cut the margin of error in half.

Impact of Sample Size

The spread in the sampling distribution of the mean is a function of the number of individuals per sample.

- The larger the sample size, the smaller the standard deviation (spread) of the sample mean distribution.
- The spread decreases at a rate equal to $\sqrt{n}$.





Choosing the Sample Size

You may need a certain margin of error (e.g., in drug trials or manufacturing specs). In most cases, we have no control over the population variability (σ), but we can choose the number of measurements (*n*).

The confidence interval for a population mean will have a specified margin of error *m* when the sample size is:

$$m = z * \frac{\sigma}{\sqrt{n}} \quad \Leftrightarrow \quad n = \left(\frac{z * \sigma}{m}\right)^2$$

Remember, though, that sample size is not always stretchable at will. There are typically costs and constraints associated with large samples. The best approach is to use the smallest sample size that can give you useful results.

Example

Density of bacteria in solution:

Measurement equipment has standard deviation $\sigma = 1 * 10^6$ bacteria/ml fluid.

How many measurements should you make to obtain a margin of error of at most $0.5 * 10^6$ bacteria/ml with a confidence level of 95%?

For a 95% confidence interval, $z^* = 1.96$.

$$n = \left(\frac{z \times \sigma}{m}\right)^2 \Rightarrow n = \left(\frac{1.96 \times 1}{0.5}\right)^2 = 3.92^2 = 15.3664$$

Using only 15 measurements will not be enough to ensure that *m* is no more than $0.5 * 10^6$. Therefore, we need at least 16 measurements.

z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Some Cautions

- The data should be an **SRS** from the population.
- The confidence interval and sample size formulas are **not** correct for other sampling methods.
- Inference **cannot** rescue badly produced data.
- Confidence intervals are **not resistant** to outliers.
- If $\boldsymbol{n}$ is small **(<15)** and the population is not Normal, the true confidence level will be **different** from ***C*.**
- The standard deviation σ of the population must be known.
- The margin of error in a confidence interval covers only random sampling errors!

6.2 Tests of Significance

- The reasoning of tests of significance
- Stating hypotheses
- Test statistics
- *P*-values
- Statistical significance
- Test for a population mean
- Two-sided significance tests and confidence intervals

Statistical Inference

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 evidence in the data about some claim concerning a population.

A **test of significance** 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, called the P-value, that measures how well the data and the claim agree.

The Reasoning of Tests of Significance

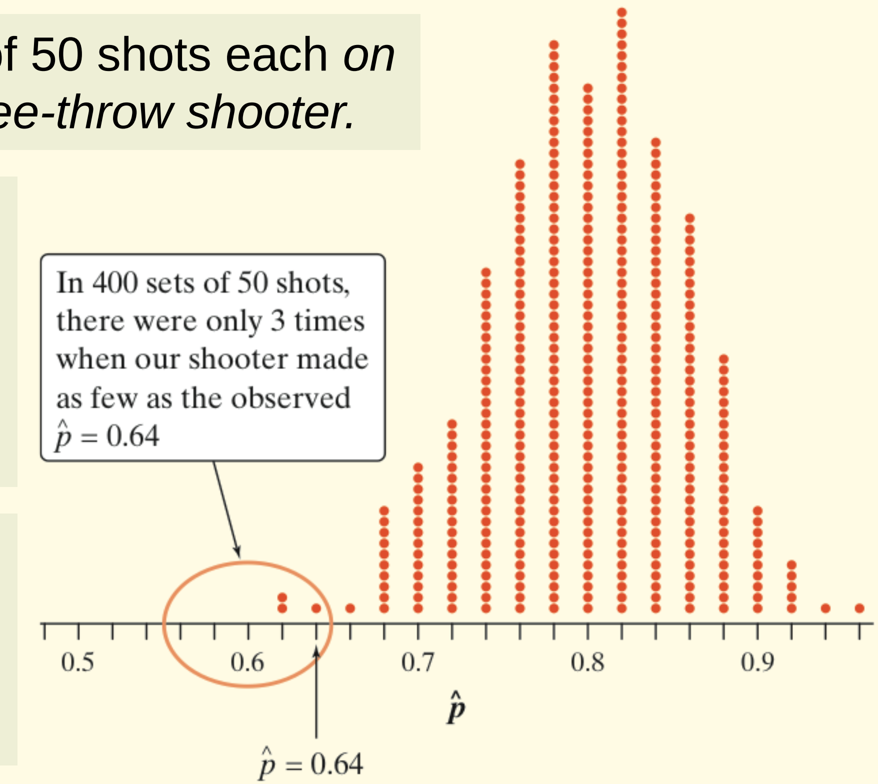
Suppose a basketball player claimed to be an 80% free-throw shooter. To test this claim, we have him attempt 50 free throws. He makes 32 of them. His sample proportion of made shots is 32/50 = 0.64.

What can we conclude about the claim based on this sample data?

We can use software to simulate 400 sets of 50 shots each *on the assumption that the player is an 80% free-throw 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.

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



Stating Hypotheses

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

The claim tested by a statistical test is called the **null hypothesis (H_0).** The test is designed to assess the strength of the evidence against the null hypothesis. Often, the null hypothesis is a statement of "no effect" or "no difference in the true means."

The claim about the population for which we're trying to find evidence is the **alternative hypothesis (H_a).** The alternative is **one-sided** if it states either that a parameter is (1) *larger* than the null hypothesis value, or (2) *smaller than* the null hypothesis value. It is **two-sided** if it states that the parameter is *different from* the null value.

In the free-throw shooter example, let p be the true long-run proportion of made free throws, and suppose our hypotheses are

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

Here, H_a is a *one-sided* alternative.

Example

Does the job satisfaction of assembly-line workers differ when their work is machine-paced rather than self-paced? One study chose 18 subjects at random from a company with over 200 workers who assembled electronic devices. Half of the workers were assigned at random to each of two groups. Both groups did similar assembly work, but one group was allowed to pace themselves while the other group used an assembly line that moved at a fixed pace. After two weeks, all the workers took a test of job satisfaction. Then they switched work set-ups and took the test again after two more weeks. The response variable is the difference in satisfaction scores, self-paced minus machine-paced.

The parameter of interest is the mean μ of the differences (*self-paced* minus *machine-paced*) in job satisfaction scores in the population of all assembly-line workers at this company.

State appropriate hypotheses for performing a significance test.

Because the initial question asked whether job satisfaction differs, the alternative hypothesis is two-sided; that is, either $\mu < 0$ or $\mu > 0$. For simplicity, we write this as $\mu \neq 0$. That is:

$$H_0: \mu = 0$$
$$H_a: \mu \neq 0$$

Test Statistic

A test of significance is based on a statistic that estimates the parameter that appears in the hypotheses. When H_0 is true, we expect the estimate to be near the parameter value specified in H_0.

Values of the estimate far from the parameter value specified by H_0 give evidence against H_0.

A **test statistic** calculated from the sample data measures how far the data diverge from what we would expect if the null hypothesis H_0 were true.

$$z = \frac{\text{estimate - hypothesized value}}{\text{standard deviation of the estimate}}$$

Large values of the statistic show that the data are not consistent with H_0.

P-Value

The null hypothesis H_0 states the claim that we seek to *disprove*. 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 or more extreme* than the one actually observed is called the ***P*-value** of the test. The smaller the *P*-value, the stronger the evidence against H_0.

- 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 make one of two decisions based on the strength of the evidence against the null hypothesis (and in favor of the alternative hypothesis)—**reject H_0 or fail to reject H_0**.

- If our sample result is too unlikely to have happened by chance assuming H_0 is true, then we will reject H_0.
- Otherwise, we will fail to reject H_0.

Note: A fail-to-reject H_0 decision in a significance test does not 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.

In a nutshell, our conclusion in a significance test comes down to:

P-value small → reject H_0 → conclude H_a (in context)

P-value large → fail to reject H_0 → cannot conclude H_a (in context)

Statistical Significance

There is no ironclad rule for how small a *P*-value should be in order to reject H_0—it's a matter of judgment and depends on the specific circumstances. But we can compare the *P*-value with a fixed value that we regard as decisive, called the **significance level.** We write it as α, the Greek letter alpha. When our *P*-value is less than the chosen α, we say that the result is **statistically significant.**

If the *P*-value is smaller than α, we say that the data are **statistically significant at level α.** The quantity α is called the **significance level** or the **level of significance.**

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

P-value $< \alpha \rightarrow$ reject $H_0 \rightarrow$ conclude H_a (in context)

P-value $\geq \alpha \rightarrow$ fail to reject $H_0 \rightarrow$ cannot conclude H_a (in context)

Four Steps of Tests of Significance

Tests of Significance: Four Steps

1. State the null and alternative **hypotheses.**
2. Calculate the value of the **test statistic.**
3. Find the ***P*-value** for the observed data**.**
4. State a **conclusion.**

We will learn the details of many tests of significance in the following chapters. The proper test statistic is determined by the hypotheses and the data collection design.

Tests for a Population Mean

z TEST FOR A POPULATION MEAN

Draw an SRS of size n from a Normal population that has unknown mean μ and known standard deviation σ. To test the null hypothesis that μ has a specified value,

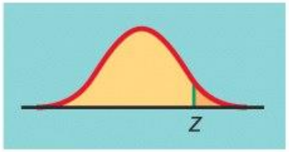
$$H_0\colon \mu = \mu_0$$

calculate the **one-sample z statistic**

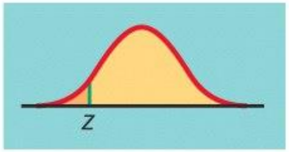
$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

In terms of a variable Z having the standard Normal distribution, the P-value for a test of H_0 against

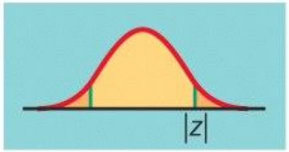
$H_a\colon \mu > \mu_0$ is $P(Z \geq z)$



$H_a\colon \mu < \mu_0$ is $P(Z \leq z)$



$H_a\colon \mu \neq \mu_0$ is $2P(Z \geq |z|)$



Example

Does the job satisfaction of assembly-line workers differ when their work is machine-paced rather than self-paced? A matched pairs study was performed on a sample of workers, and each worker's satisfaction was assessed after working in each setting. The response variable is the difference in satisfaction scores, self-paced minus machine-paced.

The null hypothesis is no average difference in scores in the population of assembly-line workers, while the alternative hypothesis (the one we'd like to prove true) is that, on average, there is a difference in scores in the population of assembly workers.

$$H_0: \mu = 0 \qquad H_a: \mu \neq 0$$

This is a two-sided alternative because we are interested in determining if a difference, positive or negative, exists.

Example

Suppose job satisfaction scores follow a Normal distribution with standard deviation σ = 60. Data from 18 workers gave a sample mean score of 17. The test statistic is:

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{17 - 0}{\frac{60}{\sqrt{18}}} \approx 1.20$$

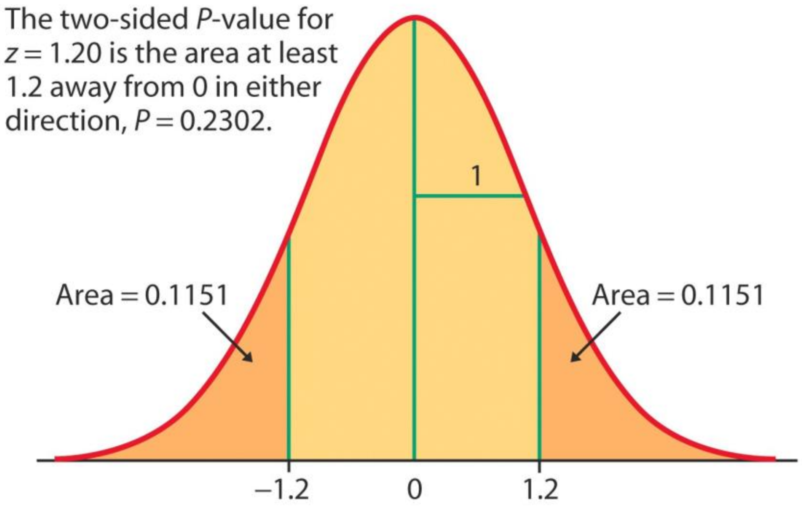
Example

For the test statistic z = 1.20 and alternative hypothesis H_a: $\mu \neq 0$, the P-value would be:

P-value = $P(Z < -1.20 \text{ or } Z > 1.20)$

$= 2\,P(Z < -1.20) = 2\,P(Z > 1.20)$

$= (2)(0.1151) = 0.2302$



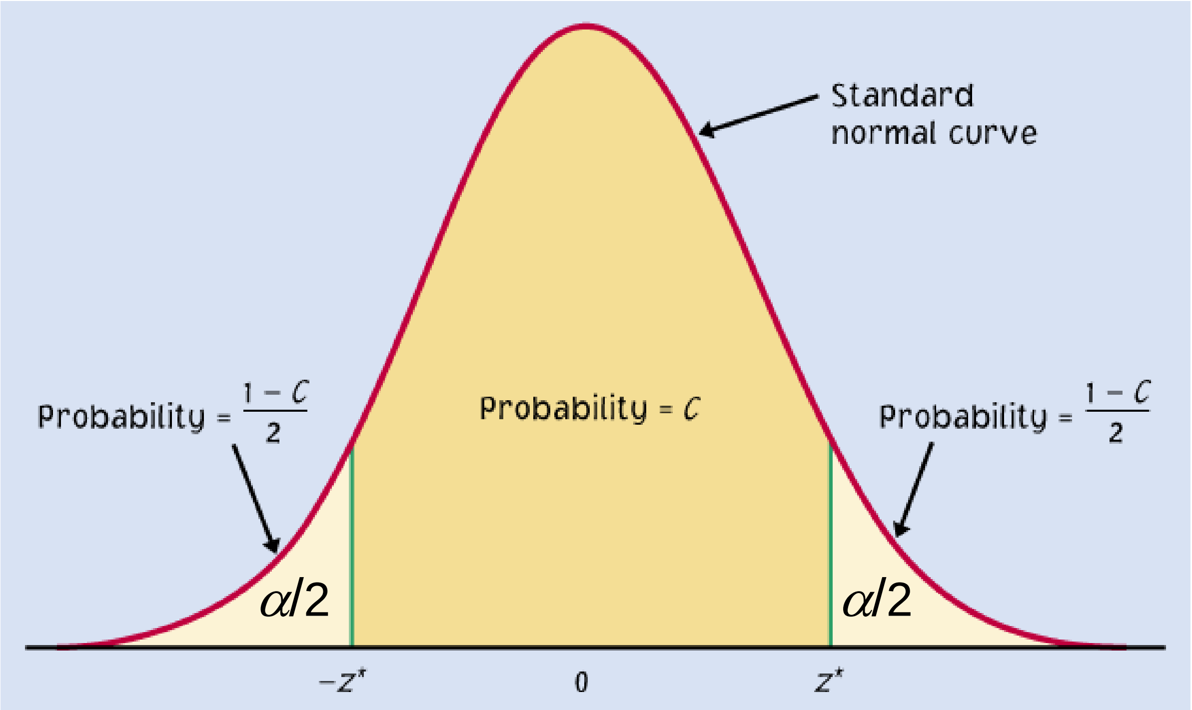
If H_0 is true, there is a chance of 0.2302 (23.02%) that we would see results at least as extreme as those in the sample. A probability of 0.2302 is not particularly small, and so the observed results are not unlikely if H_0 is true. This means there is not strong evidence in favor of H_a.

Two-Sided Significance Tests and Confidence Intervals

Because a two-sided test is symmetrical, you can also use a 1 – α confidence interval to test a two-sided hypothesis at level α.

Confidence level C and α for a two-sided test are related as follows:

$$C = 1 - \alpha$$



More About *P*-Values

- A significance test *can* be done in a black-and-white manner: We reject H_0 if $P < \alpha$, and otherwise we do not reject H_0.
- Reporting the *P*-value is a better way to summarize a test than simply stating whether or not H_0 is rejected. This is because *P* quantifies how strong the evidence is against H_0. The smaller the value of *P*, the greater the evidence.
- On the other hand, *P* does not provide specific information about the true population mean μ. **If you desire a likely range of values for the parameter, use a confidence interval**.
- Statistics in practice uses technology to get *P*-values quickly and accurately. In the absence of suitable technology, you can get approximate *P*-values by comparing your test statistic with **critical values** from a table. The example on the next slide shows how this works.

Approximating a *P*-Value

Suppose the hypotheses are:

$$H_0: \mu = 0 \qquad\qquad H_a: \mu > 0$$

and the test statistic is $z = -0.276$. Because the alternative is "μ **greater than** 0," the *P*-value is the area under the standard normal density **to the right of** –0.276.

Table A gives areas to the **left** of specified *z*, or critical, values. But using the symmetry of the normal curve, the area to the right of –0.276 is equal to the area to the left of 0.276.

From Table A, the area to the left of 0.27 is 0.6064 and the area to the left of 0.28 is 0.6103. Since 0.276 is between 0.27 and 0.28, the *P*-value satisfies:

$$0.6064 < P < 0.6103$$

There is no reason whatsoever to reject H_0 in this case as the *P*-value is so large.

6.4 Power and Inference as a Decision

- Power
- Increasing the power
- Inference as a decision
- Error probabilities

Power

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_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*
	Fail to reject H_0	*Correct conclusion*	**Type II error**

Power

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.

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.

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 satisfying the alternative hypothesis. We can calculate the probability that a test *does* reject H_0 when any *specific* alternative is true. This probability is called the **power** of the test against that specific alternative.

Power

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 between H_0 and H_a requires a larger sample than detecting a large difference.

The Common Practice of Testing Hypotheses

1. State H_0 and H_a as in a test of significance.
2. Think of the problem as a decision problem, so the probabilities of Type I and Type II errors are relevant.
3. Consider only tests in which the probability of a Type I error is no greater than a specified α.
4. Among these tests, select a test that makes the probability of a Type II error as small as possible.

Chapter 6
Introduction to Inference

6.1 Estimating with Confidence

6.2 Tests of Significance

6.3 Use and Abuse of Tests (On Your Own)

6.4 Power and Inference as a Decision