Chapter 8-2: Basics of Hypothesis Testing

Two main activities in statistical inference are using sample data to:

1. **estimate a population parameter—forming confidence intervals**

2. **test a hypothesis or claim about a population parameter**

Hypothesis

A claim or statement about a property of a population

Hypothesis Test

A standard procedure for testing a claim about a property of a population

Example Hypotheses

- A newspaper headline makes the claim: "Most workers get their jobs through networking"

- Medical researchers claim: "Mean body temperature of healthy adults is not equal to 98.6°F"

- The FAA claims: "The mean weight of an airline passenger with carry on baggage is greater than the 185 lb that it was 20 years ago

Testing will be done based on the:

Rare event rule for inferential statistics

If, **under a given assumption**, the probability of a particular observed event is exceptionally small, we conclude that the assumption is probably not correct

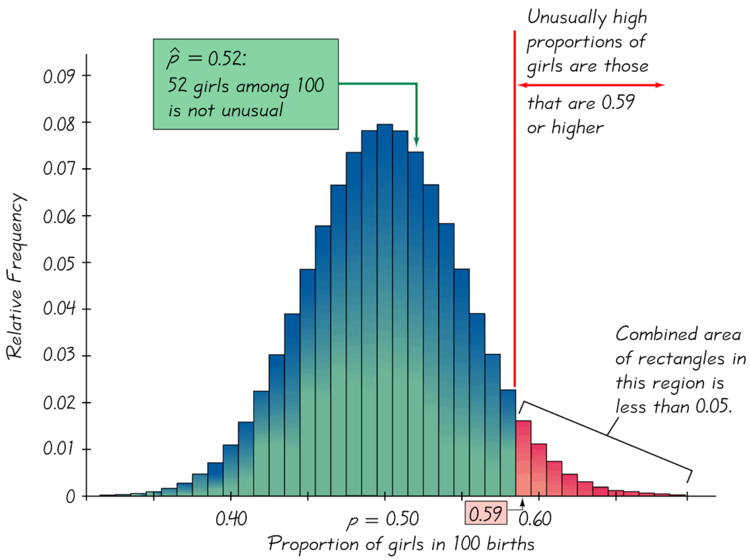
We analyze sample data in an attempt to distinguish between results that can ***easily occur by chance*** and results that are ***highly unlikely to occur by chance***

If we observe a highly unlikely result one of two things has happened—either

1. a rare event has indeed occurred, or

2. the underlying assumption is not true

For the methods we will learn in this class, we will be using the normal distribution (z) or t-distribution to make decisions about something being a rare event.



Components of a Formal Hypothesis Test

1.Hypotheses

Null hypothesis (H_0) is a statement that the value of a population parameter is ***equal to*** some claimed value

Examples include:

$$H_0: p = 0.5$$
$$H_0: \mu = 98.6$$
$$H_0: p_1 = p_2 \Rightarrow p_1 - p_2 = 0$$
$$H_0: \mu_1 = \mu_2 \Rightarrow \mu_1 - \mu_2 = 0$$

We test the null hypothesis directly ➔ assume it is true and reach a conclusion to either reject H_0 or fail to reject H_0

Alternative hypothesis (H_1 or H_a) is a statement that the population parameter has a value that somehow differs from the null hypothesis.

Statements for alternative hypotheses will use one of these symbols: $<$ or $>$ or $\neq$.

Examples include:

$$H_1\!:p \neq 0.5;\ \ H_1\!:p > 0.5$$

$$H_1\!:\mu < 98.6$$

$$H_1\!:p_1 \neq p_2 \Rightarrow p_1 - p_2 \neq 0$$

$$H_1\!:\mu_1 < \mu_2 \Rightarrow \mu_1 - \mu_2 < 0$$

2. Test Statistic

✸ Calculated from sample data

✸ Convert a sample proportion $\hat{p}$, or a sample mean $\bar{x}$ to a z- or t-score **with the assumption that the null hypothesis is true**

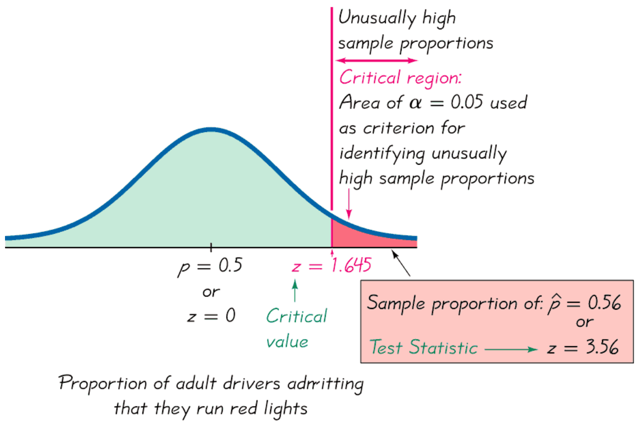
Calculation of a z- or t-score uses:

- the value of the population parameter proposed under the null hypothesis
- a measure of the sampling variability for the particular statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{pq}{n}}}; \qquad z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}; \qquad t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

3.Critical Region

The set of all values of the test statistic that cause us to reject the null hypothesis.



4. Significance Level

- Probability that the test statistic will fall in the critical region when the null hypothesis is actually true

- Denoted as α

- If the test statistic falls in the critical region, we will reject the null hypothesis--α is the probability of making a mistake of rejecting the null hypothesis when it is true

- Must be set ***a priori***—prior to running your hypothesis test

5. *P*-Value

- Called the observed significance level

- Is the probability of getting a value of the test statistic that is ***at least as extreme*** as the one representing the sample data, ***assuming the null hypothesis is true***

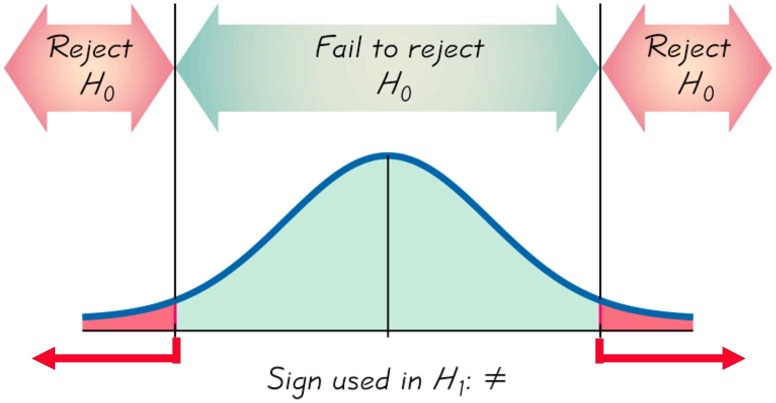
- The null hypothesis is rejected if the *P*-value is very small, such as 0.05 or less

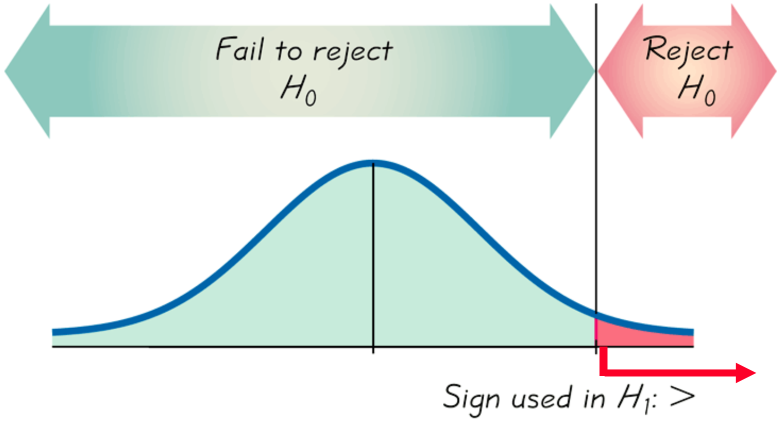
- The larger the *z*- or *t*-statistic, the smaller the *P*-value

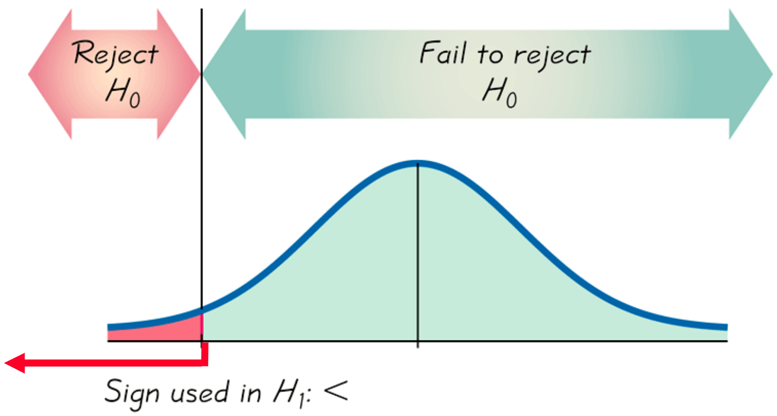
- Must differentiate between a two-tailed test and a one-tailed test

  - Two-tailed tests follow from alternative hypotheses that use $\neq$

- The significance level, α, is divided equally between the two tails that constitute the critical region







6. **<u>Conclusions</u>**--Should be stated in the context of the problem

   <u>Fail to reject</u> the claim that the population proportion of red cards in a deck is equal to 0.5 (P = 0.2507).

   Until stronger evidence is obtained, continue to assume that the population proportion of red cards in the deck is equal to 0.5 (P = 0.2507)

   There is sufficient sample evidence to warrant ***<u>rejection</u>*** of the claim that the population proportion of red cards in the deck is equal to 0.5 (P = 0.0231).

   There is sufficient sample evidence to support the claim that the population proportion of red cards in the deck is greater than 0.5 (P = 0.0231).

Identifying H_0 and H_1

1. More than 25% of Internet users pay bills online
2. Most households have telephones
3. The mean weight of women who won Miss America titles is equal to 121 lb.
4. The percentage of workers who got a job through their college is no more than 2%.
5. Plain M&M candies have a mean weight that is at least 0.8535 g.
6. The success rate with surgery is better than the success rate with splinting.
7. Unsuccessful job applicants are from a population with a greater mean age than the mean age of successful applicants.

Decision Criteria

Critical value approach:

- reject the null if the test statistic falls within the critical region
- fail to reject the null if the test statistic does not fall within the critical region

P-value approach:

- reject the null if the *P*-value $\leq \alpha$ where α is the significance level
- fail to reject the null if the *P*-value $> \alpha$

NOTE: when we cannot reject the null, we are not proving the null hypothesis true; rather, we are saying that the sample evidence is not strong enough to warrant rejection of the null hypothesis.

Potential Statistical Errors

Table 8-1 Type I and Type II Errors

		True State of Nature	
		The null hypothesis is true	The null hypothesis is false
Decision	We decide to reject the null hypothesis	**Type I error** (rejecting a true null hypothesis) α	Correct decision
	We fail to reject the null hypothesis	Correct decision	**Type II error** (failing to reject a false null hypothesis) β

Type I Error (α):

probability of rejecting the null hypothesis when the null hypothesis is true

Type II Error (β):

probability of failing to reject the null hypothesis when the null hypothesis is false

Power of a test $= 1 - \beta$;

probability of correctly rejecting a false null hypothesis