Lecture 8

Large-Sample Tests of Hypotheses

Introduction

- Suppose that a pharmaceutical company is concerned that the mean potency μ of an antibiotic meet the minimum government potency standards. They need to decide between two possibilities:
  - **The mean potency μ does not exceed the mean allowable potency.**
  - **The mean potency μ exceeds the mean allowable potency.**
- This is an example of a **test of hypothesis.**

Introduction

- Similar to a courtroom trial. In trying a person for a crime, the jury needs to decide between one of two possibilities:
  - **The person is guilty.**
  - **The person is innocent.**
- To begin with, the person is assumed innocent.
- The prosecutor presents evidence, trying to convince the jury to reject the original assumption of innocence, and conclude that the person is guilty.

Parts of a Statistical Test

1. **The null hypothesis, H_0:**
   – Assumed to be true until we can prove otherwise.
2. **The alternative hypothesis, H_a:**
   – Will be accepted as true if we can disprove H_0

Court trial:	**Pharmaceuticals:**
H_0: innocent	**H_0: μ does not exceeds allowed amount**
H_a: guilty	**H_a: μ exceeds allowed amount**

Parts of a Statistical Test

3. **The test statistic and its *p*-value:**

- A single statistic calculated from the sample which will allow us to reject or not reject H_0, and
- A probability, calculated from the test statistic that measures whether the test statistic is **likely** or **unlikely**, assuming H_0 is true.

4. **The rejection region:**
   - A rule that tells us for which values of the test statistic, or for which *p*-values, the null hypothesis should be rejected.

Parts of a Statistical Test

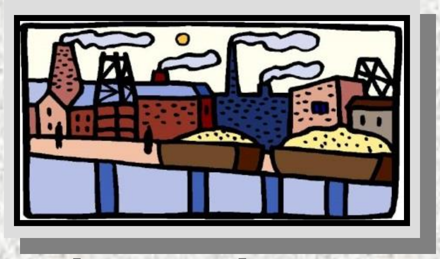
5. Conclusion:

- Either "Reject H_0" or "Do not reject H_0", along with a statement about the reliability of your conclusion.

How do you decide when to reject H_0?

- Depends on the **significance level**, α, the maximum tolerable risk you want to have of making a mistake, if you decide to reject H_0.
- Usually, the significance level is $\alpha = .01$ or $\alpha = .05$.

Example



- The mayor of a small city claims that the average income in his city is $35,000 with a standard deviation of $5000. We take a sample of 64 families, and find that their average income is $30,000. Is his claim correct?

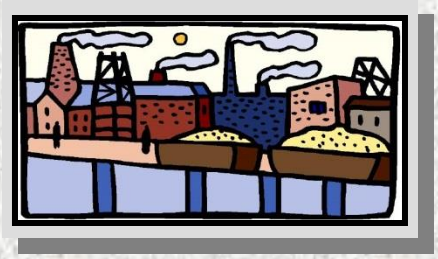
1-2. We want to test the hypothesis:

H_0: $\mu = 35{,}000$ (mayor is correct) versus

H_a: $\mu \neq 35{,}000$ (mayor is wrong)

Start by assuming that H_0 is true and $\mu = 35{,}000$.

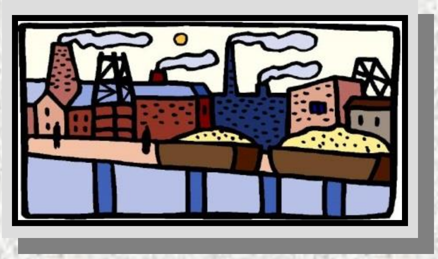
Example



3. **The best estimate of the population mean μ is the sample mean, $30,000**:

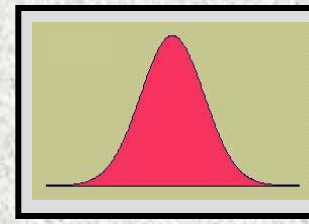
   - From the Central Limit Theorem the sample mean has an approximate normal distribution with mean μ = 35,000 and standard error SE = 5000/8 = 625.
   - The sample mean, $30,000 lies z = (30,000 – 35,000)/625 = -8 standard deviations below the mean.
   - The probability of observing a sample mean this far from μ = 35,000 (assuming H_0 is true) is *nearly zero*.

Example



4. From the Empirical Rule, values more than three standard deviations away from the mean are considered **extremely unlikely.** Such a value would be extremely unlikely to occur if indeed H_0 is true, and would give reason to reject H_0.
5. Since the observed sample mean, $30,000 is so unlikely, we choose to **reject H_0: μ = 35,000** and conclude that the mayor's claim is incorrect.
6. The probability that μ = 35,000 and that we have observed such a small sample mean just by chance is *nearly zero*.

Large Sample Test of a Population Mean, μ

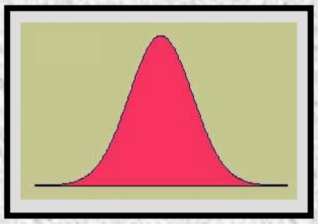


- Take a random sample of size $n \geq 30$ from a population with mean m and standard deviation s.
- We assume that either

1. **s is known or**
2. ***s*** **$\approx$ *s* since *n* is large**

- The hypothesis to be tested is
  – **$H_0: \mu = \mu_0$ versus $H_a: \mu \neq \mu_0$**

Test Statistic

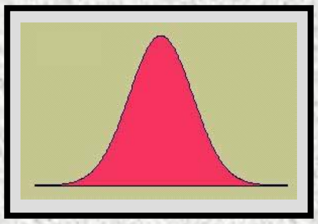


- Assume to begin with that H_0 is true. The sample mean $\bar{x}$ is our best estimate of μ, and we use it in a standardized form as the **test statistic:**

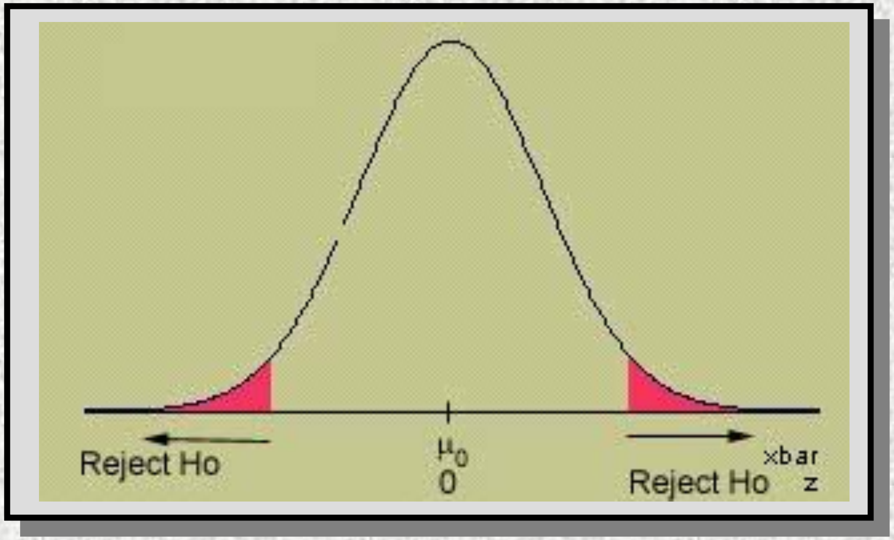
$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \approx \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

since $\bar{x}$ has an approximate normal distribution with mean μ_0 and standard deviation $\sigma / \sqrt{n}$.

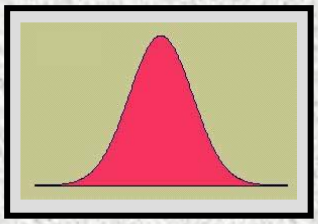
Test Statistic



- If H_0 is true the value of $\bar{x}$ should be close to μ_0, and z will be close to 0. If H_0 is false, $\bar{x}$ will be much larger or smaller than μ_0, and z will be much larger or smaller than 0, indicating that we should reject H_0.



Likely or Unlikely?



- Once you've calculated the observed value of the test statistic, calculate its ***p*-value**:

***p*-value:** The probability of observing, *just by chance*, a test statistic as extreme or even more extreme than what we've actually observed. If H_0 is rejected this is the actual probability that we have made an incorrect decision.

- If this probability is very small, less than some **preassigned significance level, α,** H_0 can be rejected.

Example

- The daily yield for a chemical plant has averaged 880 tons for several years. The quality control manager wants to know if this average has changed. She randomly selects 50 days and records an average yield of 871 tons with a standard deviation of 21 tons.

$$\mathrm{H}_0 : \mu = 880$$

$$\mathrm{H}_a : \mu \neq 880$$

Test statistic :

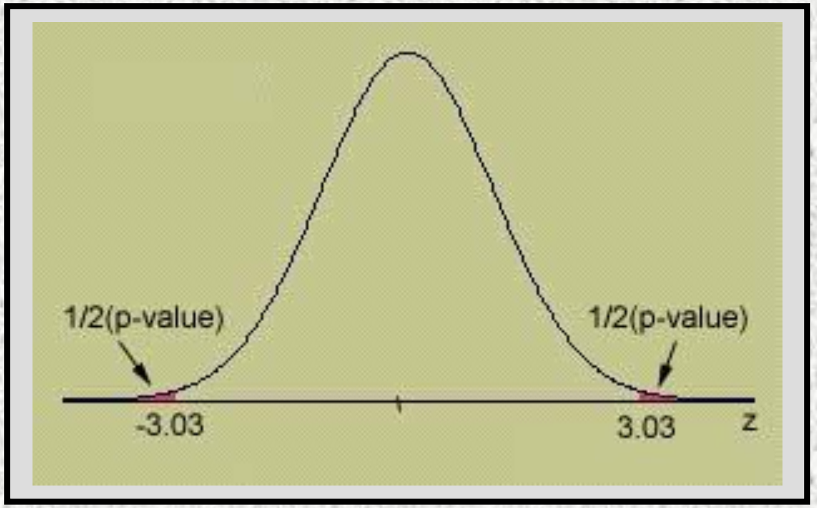
$$z \approx \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{871 - 880}{21 / \sqrt{50}} = -3.03$$

Example

What is the probability that this test statistic or something even more extreme (far from what is expected if H_0 is true) could have happened *just by chance*?

$$p\text{-value}: P(z > 3.03) + P(z < -3.03)$$

$$= 2P(z < -3.03) = 2(.0012) = .0024$$



This is an unlikely occurrence, which happens about 2 times in 1000, assuming μ = 880!

Example

- To make our decision clear, we choose a significance level, say $\alpha = .01$.

If the *p*-value is less than α, H_0 is rejected as false. You report that the results are statistically significant at level α.

If the *p*-value is greater than α, H_0 is not rejected. You report that the results are not significant at level α.

Since our *p*-value =.0024 is less than, we reject H_0 and conclude that the average yield has changed.

Using a Rejection Region

If α = .01, what would be the **critical value** that marks the “dividing line” between “not rejecting” and “rejecting” H_0?

If *p*-value < α, H_0 is rejected.

If *p*-value > α, H_0 is not rejected.

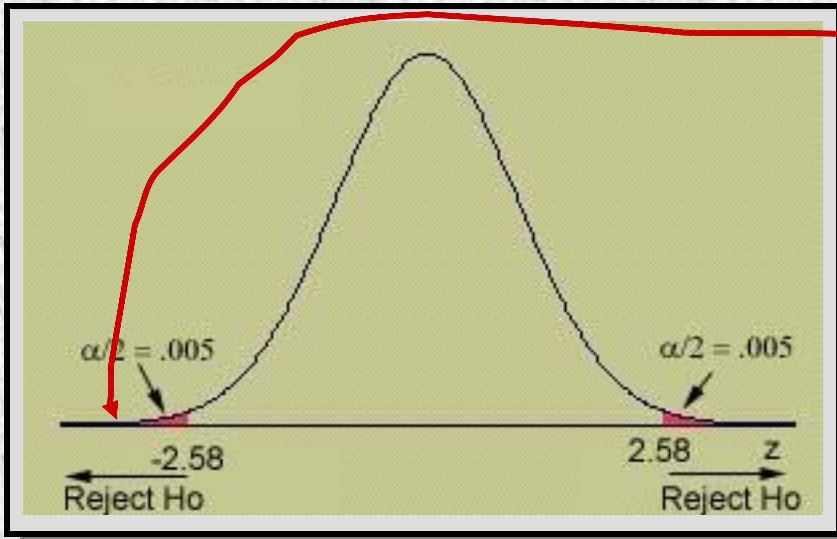
The dividing line occurs when *p*-value = α. This is called the **critical value** of the test

Test statistic > critical value implies *p*-value < α, H_0 is rejected.

Test statistic < critical value implies *p*-value > α, H_0 is not rejected.

Example

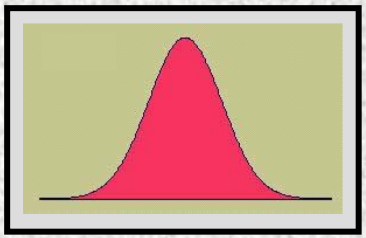
What is the critical value of z that cuts off exactly $\alpha/2 = .01/2 = .005$ in the tail of the z distribution?



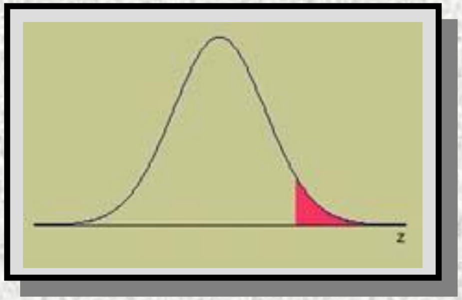
For our example, $z = -3.03$ falls in the rejection region and H_0 is rejected at the 1% significance level.

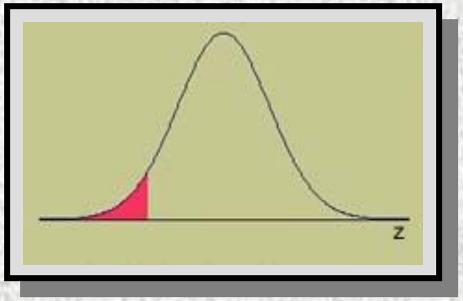
Rejection Region: Reject H_0 if $z > 2.58$ or $z < -2.58$. If the test statistic falls in the rejection region, its p-value will be less than $\alpha = .01$.

One Tailed Tests



- Sometimes we are interested in a detecting a specific directional difference in the value of μ.
- The alternative hypothesis to be tested is **one tailed:**
  - $H_a: \mu > \mu_0$ **or** $H_a: \mu < \mu_0$
- Rejection regions and *p*-values are calculated using only one tail of the sampling distribution.





Example

- A homeowner randomly samples 64 homes similar to her own and finds that the average selling price is \$252,000 with a standard deviation of \$15,000. Is this sufficient evidence to conclude that the average selling price is greater than \$250,000? Use α = .01.

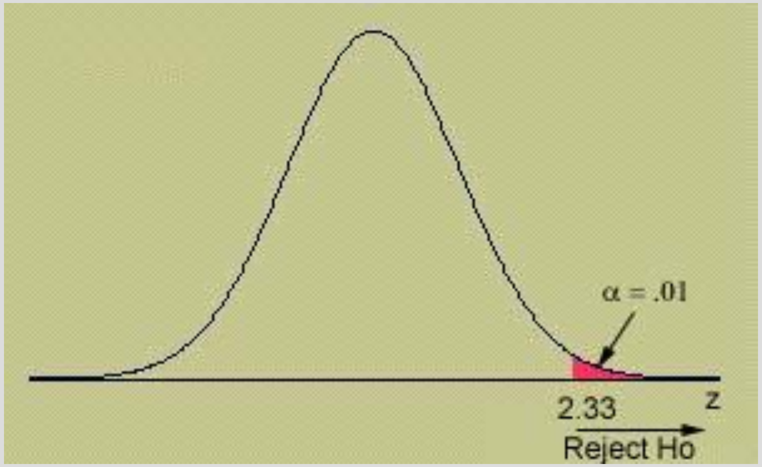
$$H_0 : \mu = 250{,}000$$
$$H_a : \mu > 250{,}000$$

Test statistic :

$$z \approx \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{252{,}000 - 250{,}000}{15{,}000 / \sqrt{64}} = 1.07$$

Critical Value Approach

What is the critical value of z that cuts off exactly $\alpha = .01$ in the right-tail of the z distribution?



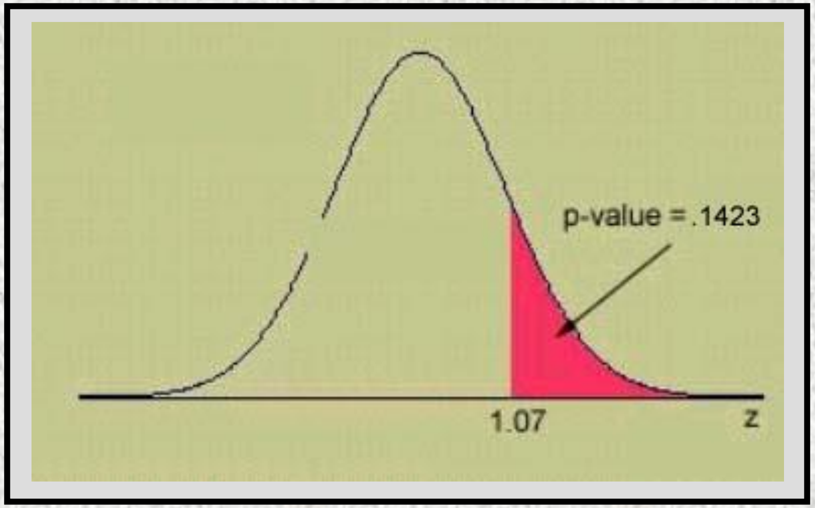
For our example, $z = 1.07$ does not fall in the rejection region and H_0 is not rejected. There is not enough evidence to indicate that μ is greater than $250,000.

Rejection Region: Reject H_0 if $z > 2.33$. If the test statistic falls in the rejection region, its p-value will be less than $\alpha = .01$.

p-Value Approach

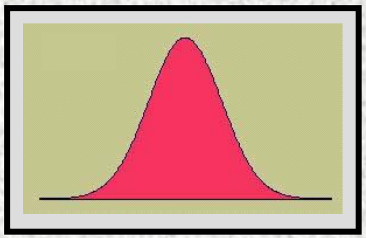
- The probability that our sample results or something even more unlikely would have occurred *just by chance*, when μ = 250,000.

$$p\text{-value}: P(z > 1.07) = 1 - .8577 = .1423$$



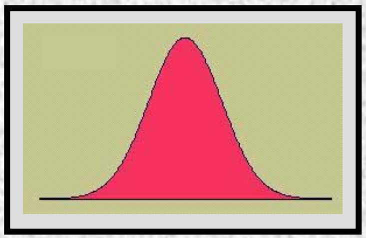
Since the *p*-value is greater than α = .01, H_0 is not rejected. There is insufficient evidence to indicate that μ is greater than \$250,000.

Statistical Significance



- The critical value approach and the *p*-value approach produce identical results.
- The *p*-value approach is often preferred because
  - Computer printouts usually calculate *p*-values
  - You can evaluate the test results at any significance level you choose.
- What should you do if you are the experimenter and no one gives you a significance level to use?

Statistical Significance



- If the *p*-value is less than **.01**, reject H_0. The results are **highly significant.**
- If the *p*-value is between **.01** and .**05**, reject H_0. The results are **statistically significant.**
- If the *p*-value is between **.05** and .**10**, do not reject H_0. But, the results are **tending towards significance.**
- If the *p*-value is greater than **.10**, do not reject H_0. The results are **not statistically significant.**

Two Types of Errors

There are two types of errors which can occur in a statistical test.

Actual Fact / Jury's Decision	Guilty	Innocent
Guilty	Correct	Error
Innocent	Error	Correct

Actual Fact / Your Decision	H_0 true (Accept H_0)	H_0 false (Reject H_0)
H_0 true (Accept H_0)	Correct	Type II Error
H_0 false (Reject H_0)	Type I Error	Correct

Define:

α = P(Type I error) = P(reject H_0 when H_0 is true)

β = P(Type II error) = P(accept H_0 when H_0 is false)

Two Types of Errors

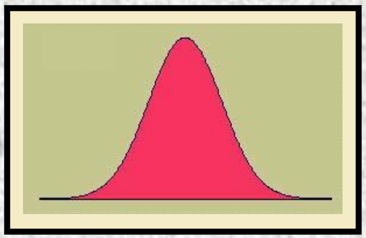
We want to keep the probabilities of error as small as possible.

- **The value of α is the significance level, and is controlled by the experimenter.**
- **The value of β is difficult, if not impossible to calculate.**

Rather than “accepting H_0” as true without being able to provide a measure of goodness, we choose to “not reject” H_0.

We write: There is insufficient evidence to reject H_0.

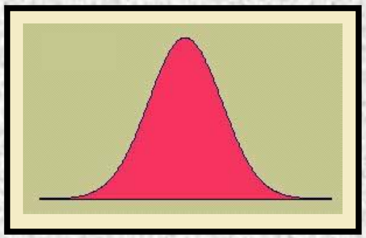
Other Large Sample Tests



- There were three other statistics in Chapter 8 that we used to estimate population parameters.
- These statistics had approximately normal distributions when the sample size(s) was large.
- These same statistics can be used to test hypotheses about those parameters, using the general test statistic:

$$z = \frac{\text{statistic} - \text{hypothesized value}}{\text{standard error of statistic}}$$

Testing the Difference between Two Means



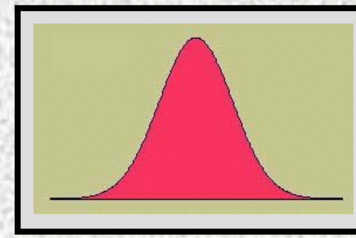
A random sample of size n_1 drawn from population 1 with mean μ_1 and variance σ_1^2.

A random sample of size n_2 drawn from population 2 with mean μ_2 and variance σ_2^2.

- The hypothesis of interest involves the difference, $\mu_1 - \mu_2$, in the form:
  - **H_0: $\mu_1 - \mu_2 = D_0$ versus H_a: one of three**

where D_0 is some hypothesized difference, usually 0.

The Sampling Distribution of $\bar{x}_1 - \bar{x}_2$

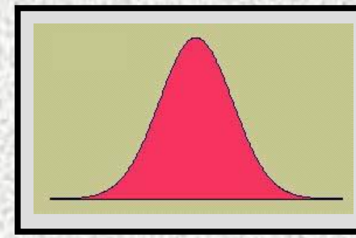


1. The mean of $\bar{x}_1 - \bar{x}_2$ is $\mu_1 - \mu_2$, the difference in the population means.

2. The standard deviation of $\bar{x}_1 - \bar{x}_2$ is $\text{SE} = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$.

3. If the sample sizes are large, the sampling distributi on of $\bar{x}_1 - \bar{x}_2$ is approximat ely normal, and SE can be estimated as $\text{SE} = \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$.

Testing the Difference between Two Means



$$\mathrm{H}_0 : \mu_1 - \mu_2 = D_0 \text{ versus}$$

$$\mathrm{H}_a : \text{one of three alternatives}$$

$$\text{Test statistic} : z \approx \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

with rejection regions and/or p-values based on the standard normal z distribution.

Example

Avg Daily Intakes	Men	Women
Sample size	50	50
Sample mean	756	762
Sample Std Dev	35	30

- Is there a difference in the average daily intakes of dairy products for men versus women? Use α = .05.

$$\mathrm{H}_0 : \mu_1 - \mu_2 = 0 \text{ (same)} \quad \mathrm{H}_a : \mu_1 - \mu_2 \neq 0 \text{ (different)}$$

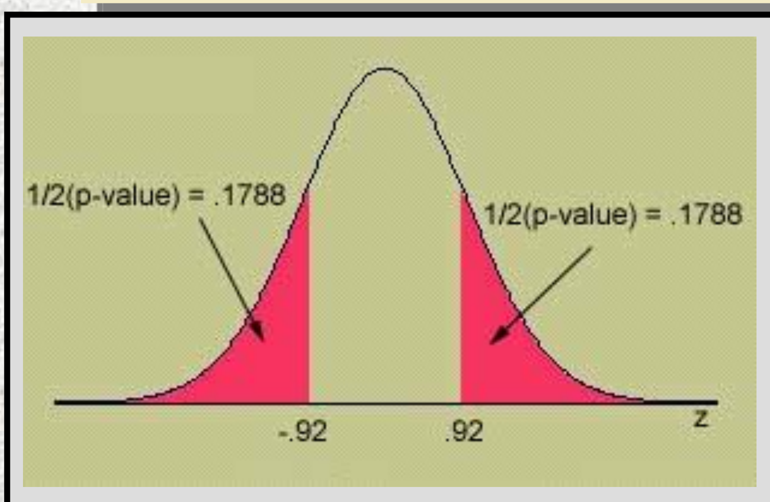
Test statistic :

$$z \approx \frac{\bar{x}_1 - \bar{x}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{756 - 762 - 0}{\sqrt{\frac{35^2}{50} + \frac{30^2}{50}}} = -.92$$

p-Value Approach

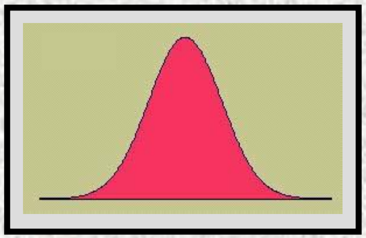
- The probability of observing values of *z* that as far away from *z* = 0 as we have, *just by chance*, if indeed $\mu_1 - \mu_2 = 0$.

$$p\text{-value}: P(z > .92) + P(z < -.92)$$
$$= 2(.1788) = .3576$$



Since the *p*-value is greater than α = .05, H_0 is not rejected. There is insufficient evidence to indicate that men and women have different average daily intakes.

Testing a Binomial Proportion *p*



A random sample of size n from a binomial population to test

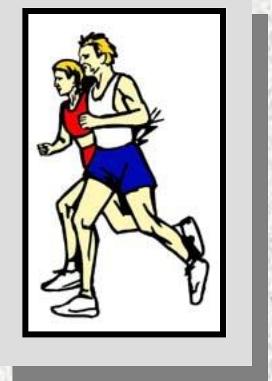
$H_0 : p = p_0$ versus

H_a : one of three alternatives

$$\text{Test statistic}: z \approx \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

with rejection regions and/or p-values based on the standard normal z distribution.

Example



- Regardless of age, about 20% of American adults participate in fitness activities at least twice a week. A random sample of 100 adults over 40 years old found 15 who exercised at least twice a week. Is this evidence of a decline in participation after age 40? Use α = .05.

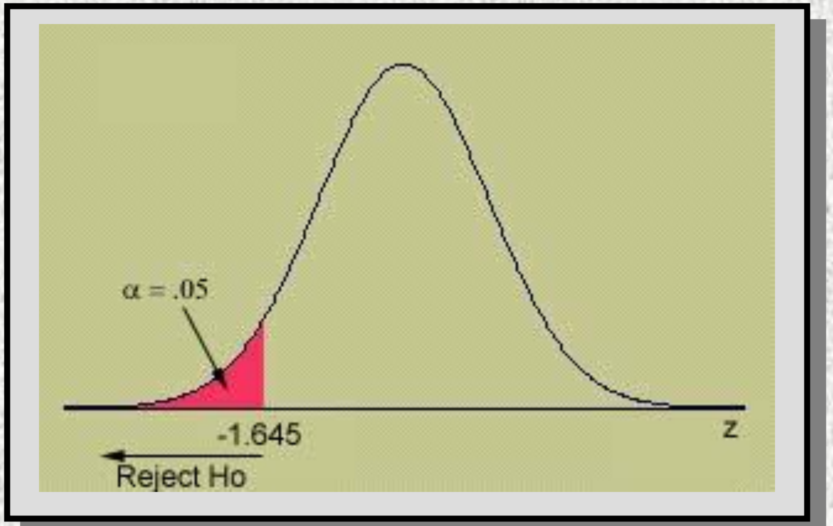
$$H_0 : p = .2$$
$$H_a : p < .2$$

Test statistic :

$$z \approx \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{.15 - .2}{\sqrt{\frac{.2(.8)}{100}}} = -1.25$$

Critical Value Approach

What is the critical value of *z* that cuts off exactly α= .05 in the left-tail of the *z* distribution?



For our example, z = -1.25 does not fall in the rejection region and H_0 is not rejected. There is not enough evidence to indicate that *p* is less than .2 for people over

Rejection Region: Reject H_0 if $z < -1.645$. If the test statistic falls in the rejection region, its *p*-value will be less than $\alpha = .05$.

Testing the Difference between Two Proportions

- To compare two binomial proportions,

A random sample of size n_1 drawn from binomial population 1 with parameter p_1.

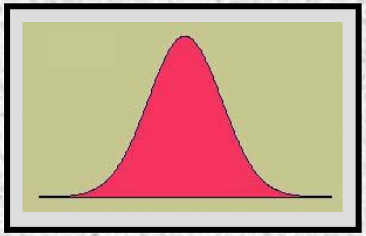
A random sample of size n_2 drawn from binomial population 2 with parameter p_2.

- The hypothesis of interest involves the difference, $p_1 - p_2$, in the form:

H_0: $p_1 - p_2 = D_0$ versus H_a: one of three

- where D_0 is some hypothesized difference, usually 0.

The Sampling Distribution of $\hat{p}_1 - \hat{p}_2$



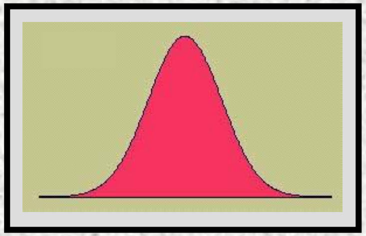
1. The mean of $\hat{p}_1 - \hat{p}_2$ is $p_1 - p_2$, the difference in the population proportions.

2. The standard deviation of $\hat{p}_1 - \hat{p}_2$ is $\text{SE} = \sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}}$.

3. If the sample sizes are large, the sampling distribution of $\hat{p}_1 - \hat{p}_2$ is approximately normal.

4. The standard error is estimated differently, depending on the hypothesized difference, D_0.

Testing the Difference between Two Proportions



$$\mathrm{H_0}: p_1 - p_2 = 0 \text{ versus}$$

$$\mathrm{H_a}: \text{one of three alternatives}$$

$$\text{Test statistic}: z \approx \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

with $\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$ to estimate the common value of p

and rejection regions or p-values

based on the standard normal z distribution.

Example

Youth Soccer	Male	Female
Sample size	80	70
Played soccer	65	39

- Compare the proportion of male and female college students who said that they had played on a soccer team during their K-12 years using a test of hypothesis.

$$\mathrm{H}_0 : p_1 - p_2 = 0 \text{ (same)} \qquad \mathrm{H}_a : p_1 - p_2 \neq 0 \text{ (different)}$$

$$\text{Calculate } \hat{p}_1 = 65/80 = .81$$

$$\hat{p}_2 = 39/70 = .56 \qquad \hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{104}{150} = .69$$

Example

Youth Soccer	Male	Female
Sample size	80	70
Played soccer	65	39

Test statistic :

$$z = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{.81 - .56}{\sqrt{.69(.31)\left(\frac{1}{80} + \frac{1}{70}\right)}} = 3.30$$

$$p\text{ - value} : P(z > 3.30) + P(z < -3.30) = 2(.0005) = .001$$

Since the *p*-value is less than α = .01, H_0 is rejected. The results are highly significant. There is evidence to indicate that the rates of participation are different for boys and girls.

Key Concepts

I. Parts of a Statistical Test

1. Null hypothesis: a contradiction of the alternative hypothesis
2. Alternative hypothesis: the hypothesis the researcher wants to support.
3. Test statistic and its *p*-value: sample evidence calculated from sample data.
4. Rejection region—critical values and significance levels: values that separate rejection and nonrejection of the null hypothesis
5. Conclusion: Reject or do not reject the null hypothesis, stating the practical significance of your conclusion.

Key Concepts

II. Errors and Statistical Significance

1. The significance level α is the probability if rejecting H_0 when it is in fact true.
2. The *p*-value is the probability of observing a test statistic as extreme as or more than the one observed; also, the smallest value of α for which H_0 can be rejected.
3. When the *p*-value is less than the significance level α , the null hypothesis is rejected. This happens when the test statistic exceeds the critical value.
4. In a Type II error, β is the probability of accepting H_0 when it is in fact false. The power of the test is $(1 - \beta)$, the probability of rejecting H_0 when it is false.

Key Concepts

III. Large-Sample Test Statistics Using the *z* Distribution

To test one of the four population parameters when the sample sizes are large, use the following test statistics:

Parameter	Test Statistic
μ	$z = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$
p	$z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}}$
$\mu_1 - \mu_2$	$z = \dfrac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$
$p_1 - p_2$	$z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$