BITS Pilani
Pilani Campus

Hypothesis Testing

Akanksha Bharadwaj
Asst. Professor, CS/IS Department

The Difference In Two Means Using The Z Statistic

- In some research designs, the sampling plan calls for selecting two independent samples, calculating the sample means and using the difference in the two sample means to estimate or test the **difference** in the two **population means**.

- The object might be to determine whether the two samples come from the same population or, if they come from different populations, to determine the amount of difference in the populations.

- This type of analysis can be used to determine, for example, whether the **effectiveness of two brands** of toothpaste differs or whether two brands of tires wear differently.

Central Limit Theorem

- The central limit theorem states that the difference in two sample means, $\bar{x}_1 - \bar{x}_2$, is normally distributed for large sample sizes (both $n1$ and $n2$ >= 30) regardless of the shape of the populations.

$$\mu_{\bar{x}_1-\bar{x}_2} = \mu_1 - \mu_2$$

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- These expressions lead to a z formula for the difference in two sample means.

$$\frac{\bar{x}-\mu}{\boxed{\frac{\sigma}{\sqrt{n}}}}$$

z FORMULA FOR THE DIFFERENCE IN TWO SAMPLE MEANS (INDEPENDENT SAMPLES AND POPULATION VARIANCES KNOWN) (10.1)

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\boxed{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}}$$

where

μ_1 = the mean of population 1
μ_2 = the mean of population 2
n_1 = size of sample 1
n_2 = size of sample 2

Hypothesis Testing

- As a specific example, suppose we want to conduct a hypothesis test to determine whether the average annual wage for an advertising manager is different from the average annual wage of an auditing manager.

- Because we are testing to determine whether the means are different, it might seem logical that the null and alternative hypotheses would be

$$H_0: \mu_1 = \mu_2 \longrightarrow H_0: \mu_1 - \mu_2 = 0$$

$$H_a: \mu_1 \neq \mu_2 \qquad H_a: \mu_1 - \mu_2 \neq 0$$

Exercise

A sample of 87 professional working women showed that the average amount paid annually into a private pension fund per person was \$3352. The population standard deviation is \$1100. A sample of 76 professional working men showed that the average amount paid annually into a private pension fund per person was \$5727, with a population standard deviation of \$1700. A women's activist group wants to "prove" that women do not pay as much per year as men into private pension funds. If they use α = .001 and these sample data, will they be able to reject a null hypothesis that women annually pay the same as or more than men into private pension funds? Use the eight-step hypothesis-testing process.

1 $H_0: \mu_w - \mu_m = 0$

$H_a/H_1: \mu_w - \mu_m < 0$

2 z-test

Solution

(3) $\alpha = 0.001$, one-tailed test

(4) critical value ?

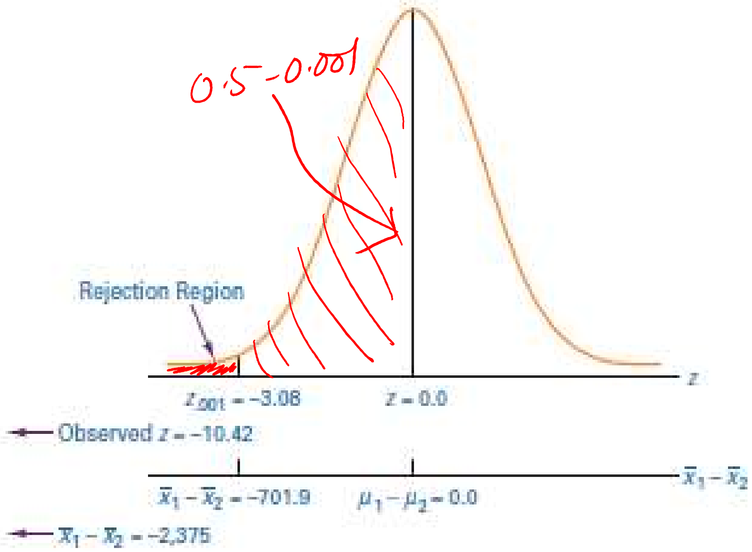
z value for 0.499 is -3.08

(5) observed value = ?

$$Z = \frac{(3352 - 5727) - 0}{\sqrt{\frac{1100^2}{87} + \frac{1700^2}{76}}}$$

$$= \frac{-2375}{227.9} = -10.42$$

0.5 - 0.001



Reject the NULL hypo.

If this problem were worked by the critical value method, what critical value of the difference in the two means would have to be surpassed to reject the null hypothesis for a table z value of -3.08? The answer is

$$(\bar{x}_1 - \bar{x}_2)_c = (\mu_1 - \mu_2) - z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$= 0 - 3.08(227.9) = -701.9$$

The difference in sample means would need to be at least 701.9 to reject the null hypothesis. The actual sample difference in this problem was -2375 ($3352 - 5727$), which is considerably larger than the critical value of difference. Thus, with the critical value method also, the null hypothesis is rejected.

Confidence Intervals

- Sometimes being able to estimate the difference in the means of two populations is valuable.

- Algebraically, formula 10.1 can be manipulated to produce a formula for constructing confidence intervals for the difference in two population means.

CONFIDENCE INTERVAL TO ESTIMATE $\mu_1 - \mu_2$ (10.2)

$$(\bar{x}_1 - \bar{x}_2) - z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

innovate achieve lead

BITS Pilani
Pilani Campus

SS ZG536, ADV STAT TECHNIQUES FOR ANALYTICS
Contact Session 7

Testing Hypotheses About A Proportion

Introduction

- To validly use this test, the sample size must be large enough such that n*p>= 5 and *n*q* >= 5.

z TEST OF A POPULATION PROPORTION (9.4)

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}}$$

where

$\hat{p}$ = sample proportion
p = population proportion
$q = 1 - p$

Example

- A manufacturer believes exactly 8% of its products contain at least one minor flaw. Suppose a company researcher wants to test this belief. The null and alternative hypotheses are

$$H_0: p = .08$$
$$H_a: p \neq .08$$

- This test is two-tailed because the hypothesis being tested is whether the proportion of products with at least one minor flaw is .08. Alpha is selected to be .10

z_c for 0.45 area $= \pm 1.645$

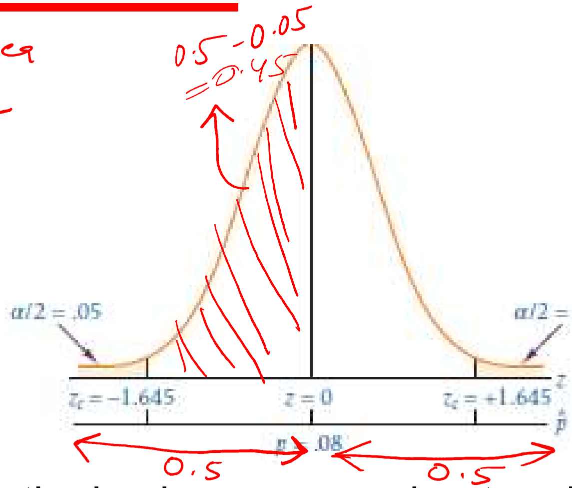
The business researcher randomly selects a sample of 200 products, inspects each item for flaws, and determines that 33 items have at least one minor flaw. Calculating the sample proportion gives:

$\hat{p} = ? = \frac{33}{200} = 0.165$

observed value of $z = \dfrac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \dfrac{0.165 - 0.08}{\sqrt{\frac{(0.08)(0.92)}{200}}}$

$= \dfrac{0.085}{0.019} = 4.43$

Reject null hypo.



For the business researcher to reject the null hypothesis, the observed z value must be greater than 1.645 or less than -1.645

Exercise

A survey of the morning beverage market shows that the primary breakfast beverage for 17% of Americans is milk. A milk producer in Wisconsin, where milk is plentiful, believes the figure is higher for Wisconsin. To test this idea, she contacts a random sample of 550 Wisconsin residents and asks which primary beverage they consumed for breakfast that day. Suppose 115 replied that milk was the primary beverage. Using a level of significance of .05, test the idea that the milk figure is higher for Wisconsin.

1. $H_0: p = 0.17$

   $H_a: p > 0.17$

2. z-test, one-tailed test

3. $\alpha = 0.05$

Solution

(4) Z value for 0.45 area

$Z_c = 1.645$

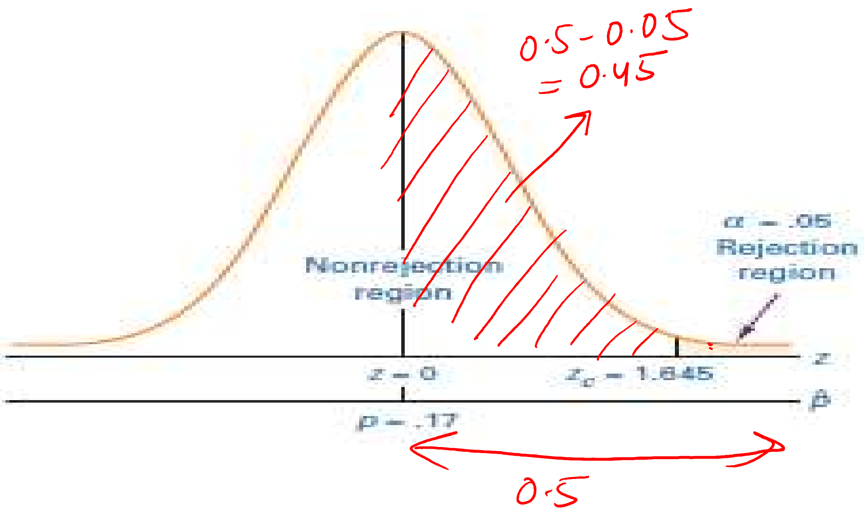
(5) observed value of Z = ?

$\hat{p} = \frac{115}{550} = 0.209$

$$Z = \frac{0.209 - 0.17}{\sqrt{\frac{0.17 \times 0.83}{550}}} = \frac{0.039}{0.016}$$

$\simeq 2.44$

Reject Null hypo.



Solving For Type II Errors

- A researcher reaches the statistical conclusion to fail to reject the null hypothesis
- If the null hypothesis is true, the researcher makes a correct decision.
- If the null hypothesis is false, then the result is a Type II error.

- Determining the probability of committing a Type II error is more complex than finding the probability of committing a Type I error.

- The probability of committing a Type I error either is given in a problem or is stated by the researcher before proceeding with the study.

- A Type II error, , varies with possible values of the alternative parameter

Example

- Suppose a researcher is conducting a statistical test on the following hypotheses.

$$H_0: \mu = 12 \text{ ounces}$$
$$H_1: \mu < 12 \text{ ounces}$$

- Often, when the null hypothesis is false, the value of the alternative mean is unknown, so the researcher will compute the probability of committing Type II errors for several possible values.

- Suppose that, in testing the preceding hypotheses, a sample of 60 cans of beverage yields a sample mean of 11.985 ounces. Assume that the population standard deviation is 0.10 ounces. From α=0.05 and a one-tailed test, the table $z_{.05}$ value is -1.645. The observed z value from sample data is

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{11.985 - 12}{0.1/\sqrt{60}} = -1.16$$

Acept NULL hypo.

- What is the probability of committing a Type II error in this problem if the population mean actually is 11.99?

$Z_c = -1.645$

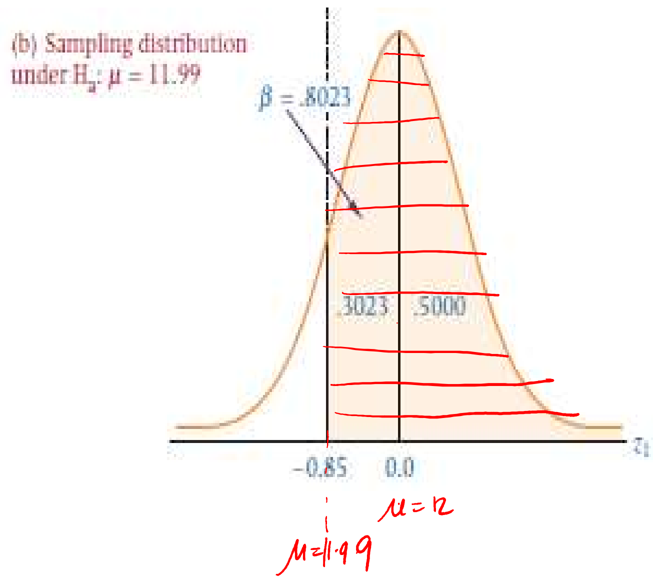
$$\bar{x}_c = -Z_c \times \frac{\sigma}{\sqrt{n}} + \mu = -1.6425 \times \frac{0.1}{\sqrt{60}} + 12$$

$\bar{x}_c = 11.979$

If μ actually equals 11.99ounces, what is the probability of failing to reject μ=12ounces when 11.979 ounces is the critical value?

$$Z_1 = \frac{11.979 - 11.99}{0.10/\sqrt{60}} = -0.85$$

prob. of type 2 error $= 0.5 + 0.3023$
$= 0.8023$



Exercise

Re-compute the probability of committing a Type II error for the soft drink example if the alternative mean is 11.96 ounces.

$Z_c = -1.645$ $\quad\quad$ $\bar{x}_c = 11.979$

$Z_1 = ?$

$$= \frac{11.979 - 11.96}{\frac{0.10}{\sqrt{60}}} = 1.47$$

Solution

The null hypothesized mean is still 12 ounces, the critical value is still 11.979 ounces, and n = 60.

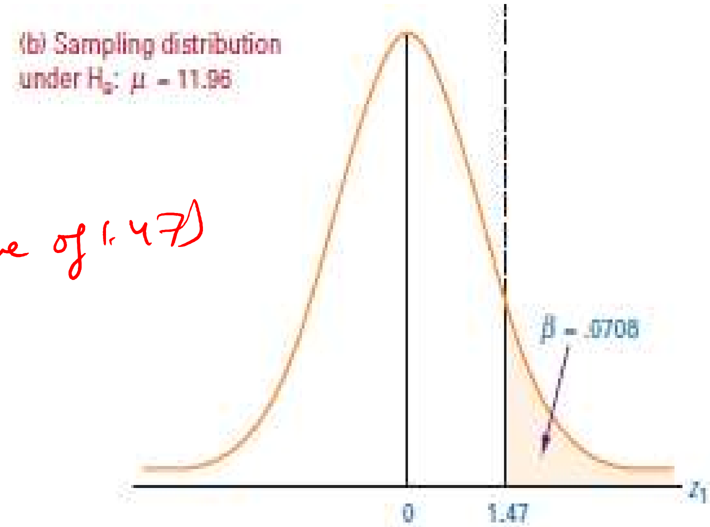
$Z_1 = 1.47$

area to the right of Z_1
is the rejection region

$= 0.5 - (\text{area for } Z_1 \text{ value of } 1.47)$

$= 0.5 - 0.4292$

$\beta = 0.0708$



0.5

Exercise (HW)

Suppose you are conducting a two-tailed hypothesis test of proportions. The null hypothesis is that the population proportion is .40. The alternative hypothesis is that the population proportion is not .40. A random sample of 250 produces a sample proportion of .44. With alpha of .05, the table z value for $\alpha/2$ is ± 1.96. The observed z from the sample information is

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} = \frac{.44 - .40}{.031} = 1.29$$

Thus the null hypothesis is not rejected. Either a correct decision is made or a Type II error is committed. Suppose the alternative population proportion really is .36. What is the probability of committing a Type II error?

Solution

Solve for the critical value of the proportion.

$$z_c = \frac{\hat{p}_c - p}{\sqrt{\frac{p \cdot q}{n}}}$$

$$\pm 1.96 = \frac{\hat{p}_c - .40}{\sqrt{\frac{(.40)(.60)}{250}}}$$

$$\hat{p}_c = .40 \pm .06$$

The critical values are .34 on the lower end and .46 on the upper end. The alternative population proportion is .36. The following diagram illustrates these results and the remainder of the solution to this problem.

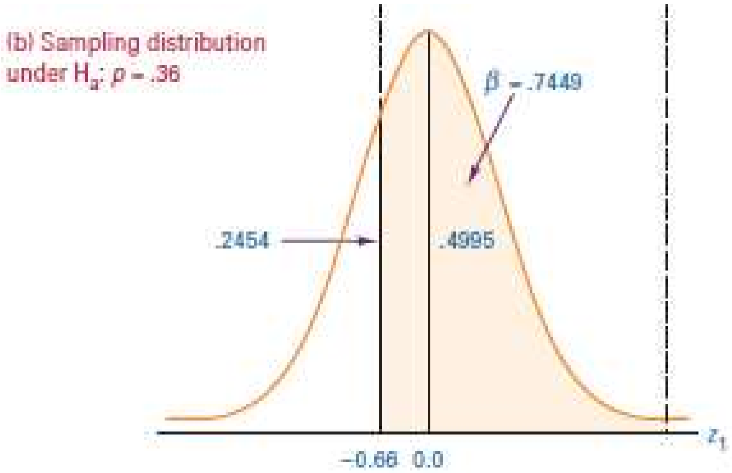
Solving for the area between $p_c = .34$ and $p_1 = .36$ yields

$$z_1 = \frac{.34 - .36}{\sqrt{\frac{(.36)(.64)}{250}}} = -0.66$$

The area associated with $z_1 = -0.66$ is .2454.

The area between .36 and .46 of the sampling distribution under H_a: $p = .36$ (graph (b)) can be solved for by using the following z value.

$$z = \frac{.46 - .36}{\sqrt{\frac{(.36)(.64)}{250}}} = 3.29$$



The area from Table A.5 associated with $z = 3.29$ is .4995. Combining this value with the .2454 obtained from the left side of the distribution in graph (b) yields the total probability of committing a Type II error:

$$.2454 + .4995 = .7449$$

innovate achieve lead

BITS Pilani

Statistical Inferences For Two Related Populations

Introduction

- In this section, a method is presented to analyze **dependent samples** or related samples.

- Some researchers refer to this test as the **matched-pairs** test. Others call it the ***t test for related measures*** or the ***correlated t test***.

- **Example**: Sometimes as an experimental control mechanism, the same person or object is measured both before and after a treatment. Certainly, the after measurement is not independent of the before measurement because the measurements are taken on the same person or object in both case

Hypothesis Testing

- The approach to analyzing two related samples is different from the techniques used to analyze independent samples.

- The matched-pairs test for related samples requires that the **two samples** be the **same size** and that the individual related scores be matched.

t FORMULA TO TEST THE DIFFERENCE IN TWO DEPENDENT POPULATIONS (10.5)

$$t = \frac{\bar{d} - D}{\frac{s_d}{\sqrt{n}}}$$

$$df = n - 1$$

where

n = number of pairs
d = sample difference in pairs
D = mean population difference
s_d = standard deviation of sample difference
$\bar{d}$ = mean sample difference

- This *t* test for dependent measures uses the sample difference, *d*, between individual matched sample values as the basic measurement of analysis instead of individual sample values.

- Analysis of the *d* values effectively converts the problem from a two-sample problem to a single sample of differences, which is an adaptation of the single-sample means formula.

FORMULAS FOR $\bar{d}$ AND s_d
(10.6 AND 10.7)

$$\bar{d} = \frac{\Sigma d}{n}$$

$$s_d = \sqrt{\frac{\Sigma(d-\bar{d})^2}{n-1}} = \sqrt{\frac{\Sigma d^2 - \frac{(\Sigma d)^2}{n}}{n-1}}$$

Example

- Suppose a stock market investor is interested in determining whether there is a significant difference in the P/E (price to earnings) ratio for companies from one year to the next. In an effort to study this question, the investor randomly samples nine companies from the *Handbook of Common Stocks* and records the P/E ratios for each of these companies at the end of year 1 and at the end of year 2. The data are shown in Table 10.5. These data are related data because each P/E value for year 1 has a corresponding year 2 measurement on the same company. Because no prior information indicates whether P/E ratios have gone up or down, the hypothesis tested is two tailed. Assume α=.01. Assume that differences in P/E ratios are normally distributed in the population.

Company	Year 1 P/E Ratio	Year 2 P/E Ratio
1	8.9	12.7
2	38.1	45.4
3	43.0	10.0
4	34.0	27.2
5	34.5	22.8
6	15.2	24.1
7	20.3	32.3
8	19.9	40.1
9	61.9	106.5

Solution

① $H_0: D = 0$
$H_1: D \neq 0$

② t-test , 2 tailed test

③ $\alpha = 0.01$
$\alpha/2 = 0.005$ $n = 9$
$\therefore df = 9-1 = 8$

④ $t_{0.005,8} = \pm 3.355$

Company	Year 1 P/E	Year 2 P/E	d
1	8.9	12.7	−3.8
2	38.1	45.4	−7.3
3	43.0	10.0	33.0
4	34.0	27.2	6.8
5	34.5	22.8	11.7
6	15.2	24.1	−8.9
7	20.3	32.3	−12.0
8	19.9	40.1	−20.2
9	61.9	106.5	−44.6

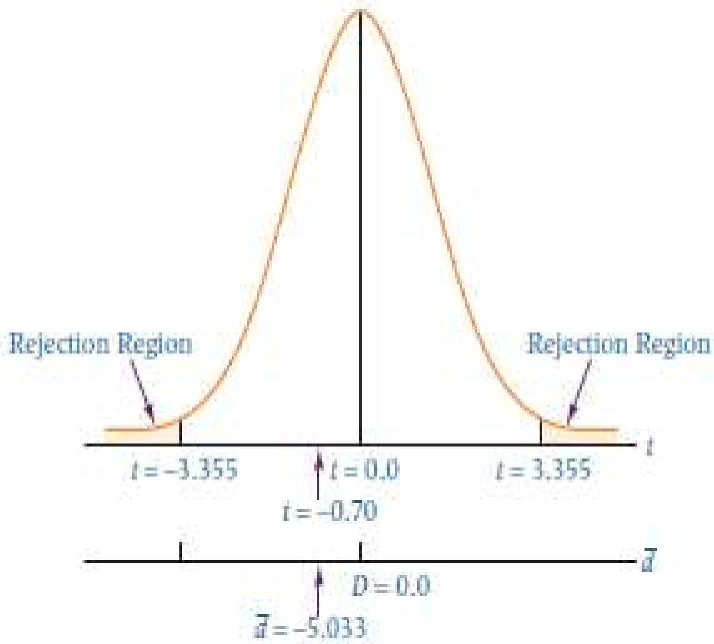
$\bar{d} = -5.033, \quad s_d = 21.599, \quad n = 9$

$$\text{Observed } t = \frac{-5.033 - 0}{\frac{21.599}{\sqrt{9}}} = -0.70$$

(5) observed value of t

$$t = \frac{-5.033 - 0}{\frac{21.599}{\sqrt{9}}} = -0.70$$

fail to reject NULL hypo.



Exercise (HW)

- Let us use this hypothetical study in which consumers are asked to rate a company both before and after viewing a video on the company twice a day for a week. Use an alpha of .05 to test to determine whether there is a significant increase in the ratings of the company after the one-week video treatment. Assume that differences in ratings are normally distributed in the population.

Individual	Before	After
1	32	39
2	11	15
3	21	35
4	17	13
5	30	41
6	38	39
7	14	22

① $H_0: D = 0$

$H_a: D < 0$

Solution

Step 1

$H_0: D = 0$

$H_a: D < 0$

Step 2

$$t = \frac{\bar{d} - D}{\frac{S_d}{\sqrt{n}}}$$

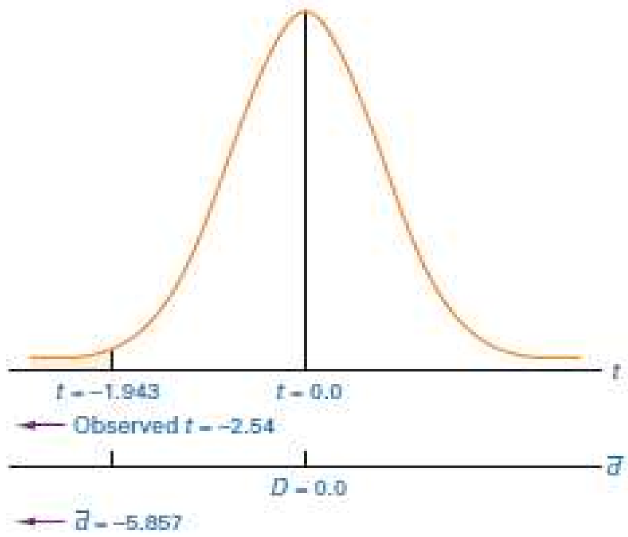
Step 3

$\alpha = 0.05$

Step 4

$df = n-1 = 7-1 = 6$ [one tail test]

$t_{0.05, 6} = -1.943$



Step 5 $\bar{d} = -5.857$ $S_d = 6.0945$

Individual	Before	After	d
1	32	39	−7
2	11	15	−4
3	21	35	−14
4	17	13	+4
5	30	41	−11
6	38	39	−1
7	14	22	−8
$\bar{d} = -5.857$		$s_d = 6.0945$	

Step 6 $$t = \frac{-5.857 - 0}{\frac{6.0945}{\sqrt{7}}} = -2.54$$

Step 7 Action:- $\because$ -2.547 is less than the critical value of -1.943

REJECT the NULL hypothesis.

Confidence Intervals

- Sometimes a researcher is interested in estimating the mean difference in two populations for related samples.
- A confidence interval for *D,* the mean population difference of two related samples, can be constructed by algebraically rearranging formula (10.5), which was used to test hypotheses about *D.*

CONFIDENCE INTERVAL FORMULA TO ESTIMATE THE DIFFERENCE IN RELATED POPULATIONS, *D* (10.8)

$$\bar{d} - t\frac{s_d}{\sqrt{n}} \leq D \leq \bar{d} + t\frac{s_d}{\sqrt{n}}$$

$$df = n - 1$$

Example

- The sale of new houses apparently fluctuates seasonally. Superimposed on the seasonality are economic and business cycles that also influence the sale of new houses. In certain parts of the country, new-house sales increase in the spring and early summer and drop off in the fall. Suppose a national real estate association wants to estimate the average difference in the number of new-house sales per company in Indianapolis between 2008 and 2009. To do so, the association randomly selects 18 real estate firms in the Indianapolis area and obtains their new-house sales figures for May 2008 and May 2009. The numbers of sales per company are shown in Table 10.7. Using these data, the association's analyst estimates the average difference in the number of sales per real estate company in Indianapolis for May 2008 and May 2009 and constructs a 99% confidence interval. The analyst assumes that differences in sales are normally distributed in the population.

Realtor	May 2008	May 2009
1	8	11
2	19	30
3	5	6
4	9	13
5	3	5
6	0	4
7	13	15
8	11	17
9	9	12
10	5	12
11	8	6
12	2	5
13	11	10
14	14	22
15	7	8
16	12	15
17	6	12
18	10	10

Solution

99% confidence interval

$$\alpha = 1 - 0.99 = 0.01$$

$$\alpha/2 = 0.005$$

$$n = 18 \Rightarrow \therefore df = 17$$

$$\bar{d} - t \times \frac{S_d}{\sqrt{n}} \le D \le \bar{d} + t \times \frac{S_d}{\sqrt{n}}$$

$$t_{0.005,17} = 2.898$$

$$-5.625 \le D \le -1.153$$

Realtor	May 2008	May 2009	d
1	8	11	−3
2	19	30	−11
3	5	6	−1
4	9	13	−4
5	3	5	−2
6	0	4	−4
7	13	15	−2
8	11	17	−6
9	9	12	−3
10	5	12	−7
11	8	6	+2
12	2	5	−3
13	11	10	+1
14	14	22	−8
15	7	8	−1
16	12	15	−3
17	6	12	−6
18	10	10	0

$\bar{d} = -3.389$ and $s_d = 3.274$

innovate achieve lead

BITS Pilani

Statistical Inferences About Two Population Proportions, P1 - P2

Introduction

- Sometimes a researcher wishes to make inferences about the difference in two population proportions.

- This type of analysis has many applications in business, such as comparing the market share of a product for two different markets, studying the difference in the proportion of female customers in two different geographic regions, or comparing the proportion of defective products from one period to another

Z-test for difference between proportions

Z-test → Difference between proportion of two population (P_1-P_2)

Assumptions

- Assume that the samples are drawn from normal population
- The sample size should be more than or equal to 30
- Subjects should be selected randomly
- Two groups should be independent of each other

Z formula

z FORMULA FOR THE DIFFERENCE IN TWO POPULATION PROPORTIONS (10.9)

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\dfrac{p_1 \cdot q_1}{n_1} + \dfrac{p_2 \cdot q_2}{n_2}}}$$

where

$\hat{p}_1$ = proportion from sample 1
$\hat{p}_2$ = proportion from sample 2
n_1 = size of sample 1
n_2 = size of sample 2
p_1 = proportion from population 1
p_2 = proportion from population 2
$q_1 = 1 - p_1$
$q_2 = 1 - p_2$

Hypothesis Testing

- The sample proportions are combined by using a weighted average to produce $\bar{p}$, which, in conjunction with $\bar{q}$ and the sample sizes, produces a point estimate of the standard deviation of the difference in sample proportions.

z FORMULA TO TEST THE DIFFERENCE IN POPULATION PROPORTIONS (10.10)

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{(\bar{p} \cdot \bar{q})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2}$ and $\bar{q} = 1 - \bar{p}$

Example

- Do consumers and CEOs have different perceptions of ethics in business? A group of researchers attempted to determine whether there was a difference in the proportion of consumers and the proportion of CEOs who believe that fear of getting caught or losing one's job is a strong influence of ethical behavior. In their study, they found that 57% of consumers said that fear of getting caught or losing one's job was a strong influence on ethical behavior, but only 50% of CEOs felt the same way. Suppose these data were determined from a sample of 755 consumers and 616 CEOs. Does this result provide enough evidence to declare that a significantly higher proportion of consumers than of CEOs believe fear of getting caught or losing one's job is a strong influence on ethical behaviour?

consumers CEO's

1 $H_0: p_1 - p_2 = 0$

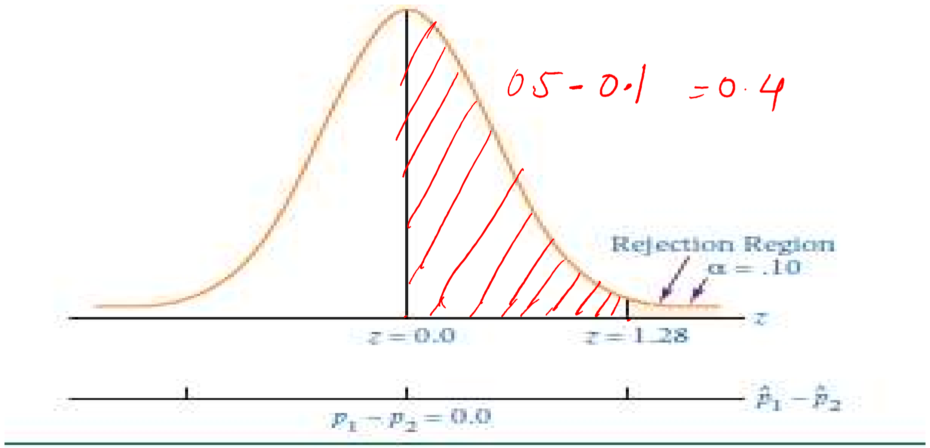
$H_a: p_1 - p_2 > 0$

Solution

2. z-test , one-tailed test

3. $\alpha = 0.1$

$\boxed{Z_c = 1.28}$ (for area 0.4)



$n_1 = 755$ $\quad n_2 = 616$

$\hat{p}_1 = 0.57$ $\quad \hat{p}_2 = 0.50$

$$\bar{p} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2} = \frac{755 \times 0.57 + 616 \times 0.50}{755 + 616}$$

$$= 0.539$$

$$\bar{q} = 1 - \bar{p} = 0.461$$

observed value of Z $= \dfrac{(0.57 - 0.50) - 0}{\sqrt{0.539 \times 0.461 \left(\frac{1}{755} + \frac{1}{616}\right)}} = 2.586$

$\boxed{\text{Reject NULL hypo.}}$

Exercise (HW)

$$\bar{p} = \frac{24+39}{100+95} = \frac{63}{195} = 0.323$$

A study of female entrepreneurs was conducted to determine their definition of success. The women were offered optional choices such as happiness/self-fulfillment, sales/profit, and achievement/challenge. The women were divided into groups according to the gross sales of their businesses. A significantly higher proportion of female entrepreneurs in the $100,000 to $500,000 category than in the less than $100,000 category seemed to rate sales/profit as a definition of success. Suppose you decide to test this result by taking a survey of your own and identify female entrepreneurs by gross sales. You interview 100 female entrepreneurs with gross sales of less than $100,000, and 24 of them define sales/profit as success. You then interview 95 female entrepreneurs with gross sales of $100,000 to $500,000, and 39 cite sales/profit as a definition of success. Use this information to test to determine whether there is a significant difference in the proportions of the two groups that define success as sales/profit. Use alpha = .01.

$\alpha/2 = 0.005$

Z_c for area $(0.5-0.005) = 0.495$

Solution

Step1 $H_0: p_1 - p_2 = 0$

$H_a: p_1 - p_2 \neq 0$

Step2 z-test for difference in proportions.

Step3&4 $\alpha = 0.01$ $\alpha/2 = 0.005$

[for area $0.5 - 0.005 = 0.495$]

critical value ← $Z_{0.005} = \pm 2.575$

Step5 $n_1 = 100$

$x_1 = 24$

$\hat{p}_1 = \frac{24}{100} = 0.24$

$n_2 = 95$

$x_2 = 39$

$\hat{p}_2 = \frac{39}{95} = 0.41$

here,

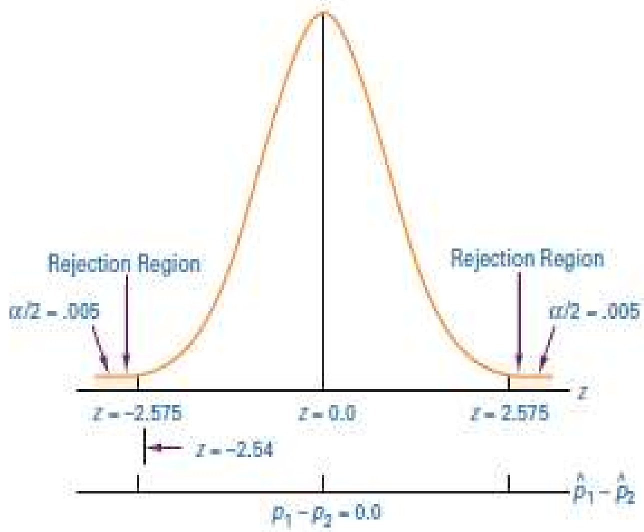
$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{24+39}{100+95}$

$\bar{p} = \frac{63}{195} = 0.323$

Step6 $z = \dfrac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \dfrac{(0.24 - 0.41) - (0)}{\sqrt{(0.323)(0.677)\left(\frac{1}{100} + \frac{1}{95}\right)}} = -2.54$

Step 7 $\because$ observed value is in the non-rejection region

NULL hypothesis is not rejected.



Confidence Intervals

- Sometimes in business research the investigator wants to estimate the difference in two population proportions.

- For example, what is the difference, if any, in the population proportions of workers in the Midwest who favor union membership and workers in the South who favor union membership?

CONFIDENCE INTERVAL TO ESTIMATE $p_1 - p_2$ (10.11)

$$(\hat{p}_1 - \hat{p}_2) - z\sqrt{\frac{\hat{p}_1 \cdot \hat{q}_1}{n_1} + \frac{\hat{p}_2 \cdot \hat{q}_2}{n_2}} \le p_1 - p_2 \le (\hat{p}_1 - \hat{p}_2) + z\sqrt{\frac{\hat{p}_1 \cdot \hat{q}_1}{n_1} + \frac{\hat{p}_2 \cdot \hat{q}_2}{n_2}}$$

Example

Suppose that in an attempt to target its clientele, managers of a supermarket chain want to determine the difference between the proportion of morning shoppers who are men and the proportion of after–5 P.M. shoppers who are men. Over a period of two weeks, the chain's researchers conduct a systematic random sample survey of 400 morning shoppers, which reveals that 352 are women and 48 are men. During this same period, a systematic random sample of 480 after–5 P.M. shoppers reveals that 293 are women and 187 are men. Construct a 98% confidence interval to estimate the difference in the population proportions of men.

for 98% CI

Morning Shoppers	After–5 P.M. Shoppers
$n_1 = 400$	$n_2 = 480$
$x_1 = 48$ men	$x_2 = 187$ men
$\hat{p}_1 = .12$	$\hat{p}_2 = .39$
$\hat{q}_1 = .88$	$\hat{q}_2 = .61$

Solution

$\alpha = ?$

$= 1 - 0.98 = 0.02$

$\alpha/2 = 0.01$

value of z for area $(0.5 - 0.01)$

$= 0.49$

$Z = 2.33$

$$(0.12 - 0.39) - 2.33\sqrt{\frac{(0.12)(0.88)}{400} + \frac{0.39 \times 0.61}{480}} \leq (p_1 - p_2) \leq (0.12 - 0.39) + 2.33\sqrt{\frac{0.12 \times 0.88}{400} + \frac{0.39 \times 0.61}{480}}$$

$0.334 \geq (p_1 - p_2) \geq -0.206$