LARGE SAMPLE TEST -2

Prof. Nandini Rai
KJSCE

TESTING THE DIFFERENCE BETWEEN MEANS

Sometimes we may have two distinct populations and we may want to test whether they have equal means.

If we actually take samples from two populations, it is unlikely that the two sample means would be identical.

Even if they are equal, how are we to know whether the two samples came from population having equal means or that the two samples came from the same population?

In other words, how can we test the hypothesis that the two population have equal means?

The procedure to test this hypothesis is discussed below.

DISTRIBUTION OF THE DIFFERENCE BETWEEN MEANS

If $\bar{x}_1$ *and* $\bar{x}_2$ denote the means of the samples drawn from the first and the second population respectively having means μ_1, μ_2 and standard deviations σ_1 *and* σ_2 and if the sizes of the samples are n_1 and n_2.

Then the distribution of the difference between the means $\bar{x}_1 - \bar{x}_2$ is normally distributed with mean $\mu_1 - \mu_2$ and

standard deviation given by $s = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

i.e. $z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ is a S.N.V.

PROCEDURE TO TEST THE HYPOTHESIS $\mu_1 = \mu_2$

Step 1: We calculate $\bar{x}_1 - \bar{x}_2$

Step 2: We calculate the standard error,

$$S.E. = s = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Step 3: We calculate the test statistics

$$|z| = \left|\frac{\bar{x}_1 - \bar{x}_2}{S.E.}\right|$$

Step 4: We take the decision as: If $|z|$ is less than 1.96, 2.58, or 3 we accept the hypothesis that $\mu_1 = \mu_2$ at 5%, 1% or 0.27% level of significance otherwise we reject it.

PLEASE REMEMBER

Under the hypothesis $\mu_1 = \mu_2$ the test statistics is given by

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \qquad \text{(if } \sigma_1 \; and \; \sigma_2 \text{ are known)}$$

If the samples are drawn from the same population, so that $\sigma_1 = \sigma_2 = \sigma$, then the test statistics is given by

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

PLEASE REMEMBER

If $\sigma_1 \; and \; \sigma_2$ are not known and $\sigma_1 \neq \sigma_2$ then

$S.E. = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ and test statistics is

$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ where $s_1 \; and \; s_2$ are standard deviation of the two samples

CONTINUED.......

If $\sigma_1\ and\ \sigma_2$ are not known and $\sigma_1 = \sigma_2 = \sigma$ then (i.e. both the population have same unknown standard deviation)

$$\therefore S.E. = \sqrt{\frac{s_1^2}{n_2} + \frac{s_2^2}{n_1}}$$

and test statistics is $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_2}+\frac{s_2^2}{n_1}}}$

INTERVAL ESTIMATION FOR THE DIFFERENCE BETWEEN TWO SAMPLE MEANS

The confidence limits for $\mu_1 - \mu_2$ at various levels of confidence are

$(\bar{x}_1 - \bar{x}_2) \pm 1.96SE \quad at$ 95% level of confidence,

$(\bar{x}_1 - \bar{x}_2) \pm 2.58SE \quad at$ 99% level of confidence and

$(\bar{x}_1 - \bar{x}_2) \pm 3SE \qquad at$ 99.73% level of confidence.

EXAMPLE-1

Two samples drawn from two different populations gave the following results. Find 95% confidence limits for the difference between the population means

	Size	Mean	S.D
Sample I	400	124	14
Sample II	250	120	12

SOLUTION

Step 1: For 95% confidence interval the critical value $z_\alpha = 1.96$

Step 2: $\overline{x_1} - \overline{x_2} = 124 - 120 = 4$

$$SE = \sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}$$

$$= \sqrt{\frac{14^2}{400} + \frac{12^2}{250}}$$

Step 3: Confidence interval =

$(\bar{x}_1 - \bar{x}_2) \pm 1.96SE$

Hence, the difference between the means lies between $---$ and $---$.

EXAMPLE-2

The means of two samples of sizes 1000 and 2000 respectively are 67.50 and 68.0 inches. Can the samples be regarded as drawn from the same population of standard deviation 2.5 inches?

SOLUTION

Null hypothesis $H_0: \mu_1 = \mu_2$

Alternative Hypothesis $H_a: \ \mu_1 \neq \mu_2$

Calculation of statistic:

$$\bar{x}_1 - \bar{x}_2 = 67.5 - 68.0 = -0.5$$

Since S.D. of the population is known,

$$S.E. = \sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}} \qquad = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$\therefore z = \frac{\bar{x}_1 - \bar{x}_2}{S.E.}$$

$$\therefore |z| = 5.15$$

Level of significance: $\alpha = 0.27\%$

Critical value: The table value of z_α at 0.27% level of significance from the table is 3

Decision: Since the computed value of $|z| = 5.15$ is greater than the critical value $z_\alpha = 3$, the hypothesis is rejected

$\therefore$ The samples cannot be regarded as drawn from the same population

EXAMPLE-3

- Test the significance of the difference between the means of two normal population with the same standard deviation from the following data.

	Size	Mean	S.D.
Sample I	100	64	6
Sample II	200	67	8

SOLUTION

Null hypothesis $H_0: \mu_1 = \mu_2$

Alternative Hypothesis $H_a:\ \mu_1 \neq \mu_2$

Calculation of statistic: $\bar{x}_1 - \bar{x}_2 = 64 - 67 = -3$

Since the standard deviations of two populations are equal but unknown

$$S.E. = s = \sqrt{\frac{{s_1}^2}{n_2} + \frac{{s_2}^2}{n_1}}$$

- $z = \frac{\bar{x}_1 - \bar{x}_2}{S.E.}$

- Level of significance: $\alpha =$

Critical value: The table value of z_α =

- Decision:

EXAMPLE-4

The average of marks scored by 32 boys is 72 with standard deviation 8 while that of 36 girls is 70 with standard deviation 6. Test at 1% level of significance whether the boys perform better than the girls.

SOLUTION:

Null hypothesis $H_0: \mu_1 = \mu_2$

Alternative Hypothesis $H_a: \ \mu_1 > \mu_2$

Calculation of statistic:

$$\bar{x}_1 - \bar{x}_2 = 72 - 70 = 2$$

$$S.E. = \sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}$$

(We assume that the standard deviations σ_1 and σ_2 of the two populations are not equal)

$$\therefore z = \frac{\bar{x}_1 - \bar{x}_2}{S.E.} = \frac{2}{\sqrt{3}} = 1.15$$

Level of significance: $\alpha = 1\%$

Critical value: The value of z_α at 1% level of significance from the table is 2.33

Decision: Since $z = 1.15 < z_\alpha = 2.33$,

the hypothesis is accepted

$\therefore$ Boys do not perform better than the girls

EXAMPLE-5

Two samples drawn from two different populations gave the following results.

	Size	Mean	S.D.
Sample I	125	340	25
Sample II	150	380	30

Test the hypothesis at 5% LOS that the difference of the means of the two populations is 45

SOLUTION

Null hypothesis $H_0: |\mu_1 - \mu_2| = 45$

Alternative Hypothesis H_a: $|\mu_1 - \mu_2| \neq 45$

Calculation of statistic: $\bar{x}_1 - \bar{x}_2 = -40,$

$$|\mu_1 - \mu_2| = 45$$

Since standard deviation of two populations are unknown and unequal

$$S.E. = s = \sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}$$

$$= \sqrt{\frac{25^2}{125} + \frac{30^2}{150}}$$

$$= 3.32$$

$$z = \frac{(|\bar{x}_1 - \bar{x}_2|) - (|\mu_1 - \mu_2|)}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}}$$

FEW MORE EXAMPLES

- The mean life of a sample of 100 electric light bulbs was found to be 1456 hours with S.D. 400. A second sample of 225 bulbs chosen from a different batch showed a mean life of 1400 hours with standard deviation of 144 hours. Assuming that the two populations have same standard deviation find if there is any significant difference between the mean of two batches?
- ANSWER: $|z| = 1.84$, No

P-VALUE APPROACH

A researcher wants to compare the **average exam scores** of students from **School A and School B** to determine if there is a significant difference in their performances.

Given Data:

Sample size of School A: $n1 = 50$

Sample mean of School A: $\overline{x_1} = 78$

Sample standard deviation of School A: $s1 = 10$

Sample size of School B: $n2 = 60$

Sample mean of School B: $\overline{x_2} = 74$

Sample standard deviation of School B: $s2 = 8$

Solution

Null Hypothesis $\boldsymbol{H_0}$: $\boldsymbol{\mu_1 = \mu_2}$

Alternative Hypothesis $\boldsymbol{H_a}$: $\boldsymbol{\mu_1 \neq \mu_2}$

This is a **two-tailed test.**

Significance Level We set $\alpha = 0.05$

P-VALUE APPROACH

Compute the Z-Statistic

$\bar{x}_1 - \bar{x}_2 = 78 - 74 = 4$

$S.E. = \sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}} = 1.7511$

(We assume that the standard deviations σ_1 and σ_2 of the two populations are not equal)

$$\therefore z = \frac{\bar{x}_1 - \bar{x}_2}{S.E.} = \frac{4}{1.7511} = 2.28$$

Find the p-value

$$P(Z > 2.28) \approx 0.0113$$

Since it's a **two-tailed test**, we multiply by 2:

$$p = 2 \times 0.0113 = 0.0226$$

Compare p-value with α\alpha

p-value = 0.0226 $\boldsymbol{\alpha = 0.05}$

Since **p-value < 0.05**, we **reject H0.**

Conclusion

There is sufficient statistical evidence to conclude that the **mean exam scores of School A and School B are significantly different.**

ONE MORE EXAMPLE

A researcher wants to test whether the **average salary at Company A is lower than at Company B.**

Given Data:

Company A:

- Sample size: 45
- Sample mean: 48,000
- Sample standard deviation: 5,000

Company B:

- Sample size: 50
- Sample mean: 50,000
- Sample standard deviation: 6,000

Solution:

Null Hypothesis $\boldsymbol{H_0: \mu_1 = \mu_2}$

Alternative Hypothesis $\boldsymbol{H_a: \mu_1 < \mu_2}$

This is a **left-tailed test** because we are testing if the mean salary at Company A is **lower** than at Company B.

Significance Level We set $\alpha = 0.05$

ONE MORE EXAMPLE

Compute the Test-Statistic

$$\bar{x}_1 - \bar{x}_2 = 48000 - 50000 = -2000$$

$$S.E. = \sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}} = 1129.404956$$

(We assume that the standard deviations σ_1 and σ_2 of the two populations are not equal)

$$\therefore z = \frac{\bar{x}_1 - \bar{x}_2}{S.E.} = \frac{-2000}{1129.404956} = -1.77$$

Find the p-value

Since this is a **left-tailed test**, we find $P(Z < -1.77)$ using a **Z-table:**

$$P(Z < -1.77) \approx 0.0384$$

Compare p-value with α

p-value = $\mathbf{0.0384}$ $\quad \boldsymbol{\alpha} = \mathbf{0.05}$

Since **p-value < 0.05**, we **reject H0.**

Conclusion

There is **statistically significant evidence** to conclude that the **average salary at Company A is lower than at Company B.**