Chapter(10)
Tow-Samples Tests of Hypothesis
(Examples)

Test of hypothesis about the difference between Two population means

Independent samples

Step (1): State the Null (H_0) and alternate (H_1) hypothesis

Case1:

$H_0: \mu_1 \le \mu_2$ Or $H_0: \mu_1 - \mu_2 \le 0$

$H_1: \mu_1 > \mu_2$ $H_1: \mu_1 - \mu_2 > 0$

Case2:

$H_0: \mu_1 \ge \mu_2$ Or $H_0: \mu_1 - \mu_2 \ge 0$

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

Case3:

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

$H_1: \mu_1 \ne \mu_2$ $H_1: \mu_1 - \mu_2 \ne 0$

Step (2): Select a level of significance.

Step (3): Select the Test Statistic (computed value)

I. with known population standard deviation σ.

$$\text{Test statistic} = Z_c = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

The confidence limits are:

$$(\bar{X}_1 - \bar{X}_2) \pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

II.: with unknown population standard deviation,

($n_1 + n_2 - 2 \ge 30 \; \& \; \sigma_1 = \sigma_2$).

$$\text{Test statistic } Z_c = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

The confidence limits are:

$$(\bar{X}_1 - \bar{X}_2) \pm Z_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

III. with unknown population standard deviation, $(n_1 + n_2 - 2 < 30 \,\&\, \sigma_1 = \sigma_2)$.

$$\text{Test statistic} = t_c = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

The confidence limits are:

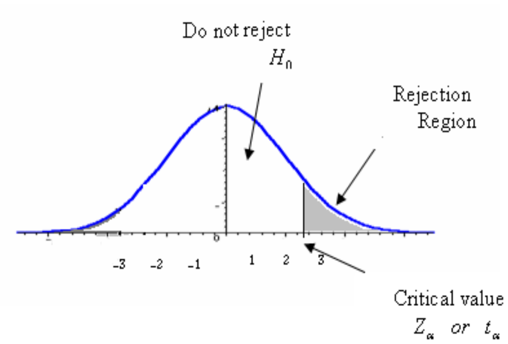
$$(\bar{X}_1 - \bar{X}_2) \pm t_{\left(\frac{\alpha}{2}, n_1 + n_2 - 2\right)}, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Step (4): Selected the Critical value

Types of test	Z	t
The one – tailed test (Right)	Z_α	$t_{(\alpha, n_1+n_2-2)}$
The one – tailed test (left)	$-Z_\alpha$	$-t_{(\alpha, n_1+n_2-2)}$
The two – tailed test	$\pm Z_{\alpha/2}$	$\pm t_{\left(\frac{\alpha}{2}, n_1+n_2-2\right)}$

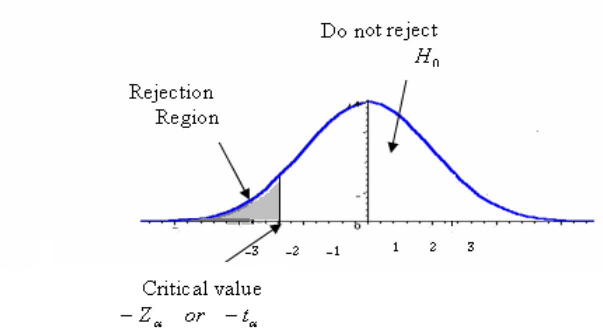
Step (5): Formulate the Decision Rule and Make a Decision

Case1: Reject H_0 if $Z_c > Z_\alpha$, $t_c > t_{v,\alpha}$



Case2: Reject H_0 if $|Z_c| > Z_\alpha$, $|t_c| > t_{v,\alpha}$

This means $Z_c < -Z_\alpha$, $t_c < -t_{v,\alpha}$

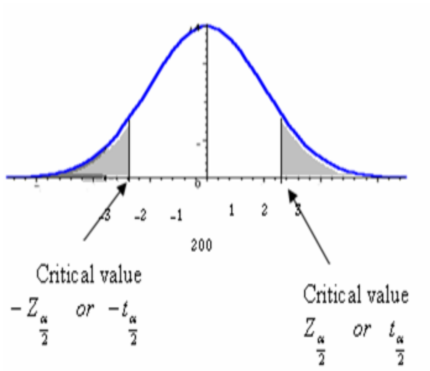


Case3: Reject H_0 if $|Z_c| > Z_{\frac{\alpha}{2}}$, $|t_c| > t_{v,\frac{\alpha}{2}}$ which means:

$$Z_c > Z_{\frac{\alpha}{2}} \quad , t_c > t_{v,\frac{\alpha}{2}}$$

or

$$Z_c < -Z_{\frac{\alpha}{2}} \quad , \ t_c < -t_{v,\frac{\alpha}{2}}$$



Example (1)

Customers at Food Town Super Markets have a choice when paying for their groceries. They may check out and pay using the standard cashier – assisted checkout, or they may use the new Fast Lane procedure. In the standard procedure a Food-Town employee puts it on a short conveyor where another employee puts it in a bag and then into the grocery cart. In the Fast Lane procedure the customer scans each item, bags it, and places the bags in the cart themselves.

The Fast Lane procedure is designed to reduce the time a customer spends in the checkout line.

The Fast Lane facility was recently installed at the Byrne Road Food Town location. The store manager would like to know if the mean checkout time using the standard checkout method is longer than using the Fast Lane. She gathered the following sample information. The time is measured from when the customer enters the line until their bags are in the cart. Hence the time includes both waiting in line and checking out.

Customer type	Sample mean	Population σ	Sample size
Standard	5.50 minutes	0.40 minutes	50
Fast Lane	5.30 minutes	0.30 minutes	100

-With 0.01 significant level test the hypothesis for difference between two means for the following:

- Compute the p-value?

.-Find a 99% Confidence Interval for the difference in the mean**.**

Solution:

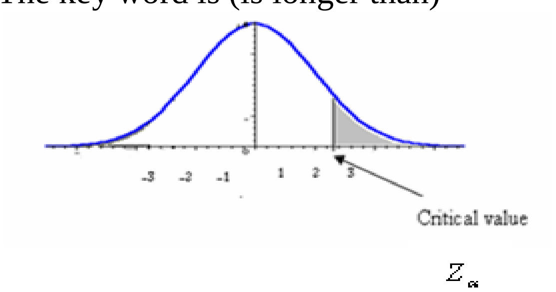
Step 1-state the null hypothesis

$H_0 : \mu_1 \leq \mu_2$

$H_1 : \mu_1 > \mu_2$

This one-tailed test (right)

The key word is (is longer than)



Z_α

Step 2: Select the level of significance.

α = 0.01 as stated in the problem

Step 3: Select the test statistic.

Use Z-distribution since the assumptions are met

$\bar{X}_1 = 5.50 \quad , \quad \sigma_1 = 0.40 \quad , \quad n_1 = 50$

$\bar{X}_2 = 5.30 \quad , \quad \sigma_2 = 0.30 \quad , \quad n_2 = 100$

Then $Z_c = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \dfrac{5.5 - 5.3}{\sqrt{\dfrac{0.40^2}{50} + \dfrac{0.30^2}{100}}} = \dfrac{0.2}{0.064} = 3.13$

Step 4: Formulate the decision rule. (Critical value)

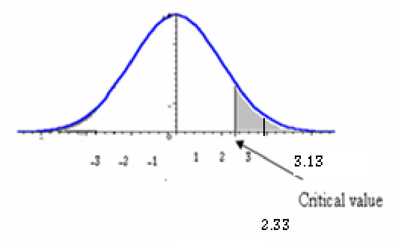
$Z_{0.01} = 2.33$

$0.5 - 0.01 = 0.4900$

Reject H_0 *if* $Z_c > 2.33$

Step 5: Make a decision and interpret the result.

$Z_c = 3.13 > 2.33$



The decision is to Reject H_0 at significant level 0.01 which means that we got significant result that the difference of 0.20 minutes between the mean checkout time using the standard method and using the Fast Lane it is not due to chance but it is significantly different , we conclude that the Fast Lane methods is faster.

The P-value for the test statistic is:

$p - value = P(Z > 3.13) = 0.5 - 0.4991 = 0.0009$

$p - value = 0.0009 < \alpha = 0.01$

$\therefore$ Reject the H_0 .

The confidence limits are:

Since we rejected H_0 can we be 99% confident that $\mu_1 > \mu_2$?

$$(\bar{X}_1 - \bar{X}_2) \pm Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$= (5.5 - 5.30) \pm 2.58 \sqrt{\dfrac{(0.40)^2}{50} + \dfrac{(0.30)^2}{100}} = 0.2 \pm 2.58\sqrt{0.0032 + 0.0009} = 0.2 \pm 2.58\sqrt{0.0041}$

$= 0.2 \pm 2.58(0.0640) = 0.2 \pm 0.1652$

$0.0348 \le \hat{\mu}_1 - \hat{\mu}_2 \le 0.3652$

$(0.0.0348 \ , \ 0.3652)$

Because the interval does not include zero, we reject $H_0 : \mu_1 \le \mu_2$, and accept

$H_0 : \mu_1 > \mu_2$

Example (2)

A manufacturer suspects a difference in the quality of the spare parts he receives from two suppliers. He obtains the following data on service life of random samples of parts from two suppliers,

For suppliers A: $n_1 = 50$, $\bar{X}_1 = 150$, $S_1 = 10$

For suppliers B: $n_2 = 100$, $\bar{X}_2 = 153$, $S_2 = 5$

- Test whether the difference between the 2 samples is statistically significant at the 1% level of significance. ($\sigma_1 = \sigma_2$)
- Compute the p-value?

.-Find a 99% Confidence Interval for the difference in the mean**.**

-Find a 99% Confidence Interval for the difference in the mean**.** **(**$\sigma_1^2 = \sigma_2^2$**)**

Solution:

Step 1: state the hypothesis:

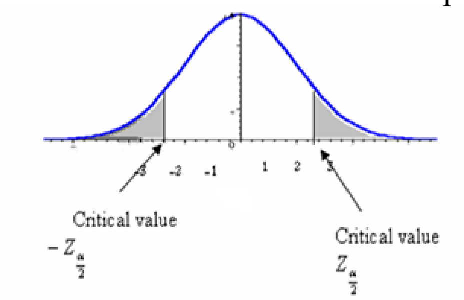
$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

Test is two-tailed test (key word is difference between 2 samples)

Step 2- Select the level of significance.

α = 0.01 as stated in the problem



Step 3: Select the test statistic.

Use Z-distribution since the assumptions are met

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{49 \times 100 + 99 \times 25}{50 + 100 - 2} = \frac{4900 + 2475}{148} = \frac{7375}{148} = 49.8311$$

$S_p = 7.0591$

$$Z_c = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{150 - 153}{7.0591\sqrt{\frac{1}{50} + \frac{1}{100}}} = \frac{-3}{7.0591\sqrt{0.02 + 0.01}} = \frac{-3}{7.0591\sqrt{0.03}}$$

$$= \frac{-3}{(7.0591)(0.1732)}$$

$$= \frac{-3}{1.2227} = -2.4536$$

Step 4: Formulate the decision rule. (Critical value)

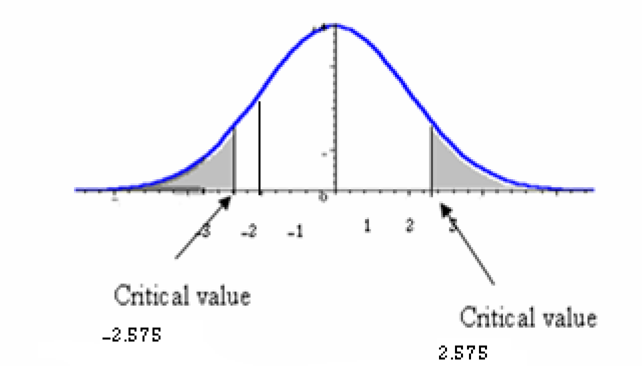
$Z_{\frac{\alpha}{2}} = Z_{\frac{0.01}{2}} = Z_{0.005} = \pm 2.58$

$0.5 - 0.005 = 0.495$

Reject H_0 if $Z_c > 2.575$ Or $Z_c < -2.575$

Step 5: Make a decision and interpret the result.

Reject H_0 if $Z_c > Z_{\frac{\alpha}{2}}$ Or $Z_c < -Z_{\frac{\alpha}{2}}$



Since -2.575 < - 2.4536 < 2.575 so we don't reject H_0 at significant level 0.01 and conclude that there is no significant difference between the two samples.

$P-value = 2P(Z > 2.4536) = 2(0.5 - 0.49286) = 2(0.00714) = 0.01428$

$P-value = 0.01428 > \alpha = 0.01$

$\therefore$ Don't reject H_0.

The confidence limits are:

$$(\bar{X}_1 - \bar{X}_2) \pm Z_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

.

$$(\bar{X}_1 - \bar{X}_2) \pm Z_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

S_p^2 Is the pooled (the common) variance is...

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{50(10)^2 + 100(5)^2}{50 + 100 - 2} = \frac{7500}{148} = 50.6757$$

$S_p = 7.1187$

$$(\bar{X}_1 - \bar{X}_2) \pm Z_{\frac{\alpha}{2}} S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = (150 - 153) \pm 2.58(7.1187)\sqrt{\frac{1}{50} + \frac{1}{100}}$$

$= -3 \pm 10.60\sqrt{0.02 + 0.01} = -3 \pm 10.60\sqrt{0.03} = -3 \pm 18.3662(0.2583)$

$= -3 \pm 4.744$

$-7.744 \le \hat{\mu}_1 - \hat{\mu}_2 \le 1.744$

$(-7.744\ ,\ 1.744)$

Since this interval contains 0 at 99% does not reject $H_0 : \mu_1 = \mu_2$.

Example (3) Owens Lawn Care, Inc., manufactures and assembles lawnmowers that are shipped to dealers throughout the United States and Canada. Two different procedures have been proposed for mounting the engine on the frame of the lawnmower. The question is: Is there a difference in the mean time to mount the engines on the frames of the lawnmowers? The first procedure was developed by longtime Owens employee Herb Welles (designated as procedure 1), and the other procedure was developed by Owens Vice President of Engineering William Atkins (designated as procedure 2). To evaluate the two methods, it was decided to conduct a time and motion study.

A sample of five employees was timed using the Welles method and six using the Atkins method. The results, in minutes, are shown below.

Methods	Sample mean	Sample S	Sample size
Welles Method	4 minutes	2.92 minutes	**5**
Atkins Method2	5 minutes	2.09 minutes	**6**

-Is there a difference in the mean mounting times? Use the 0.10 significance level. ($\sigma_1 = \sigma_2$)

-Find a 90% Confidence Interval for the difference in the mean. ($\sigma_1^2 = \sigma_2^2$)

Solution:

Step 1–state the null hypothesis

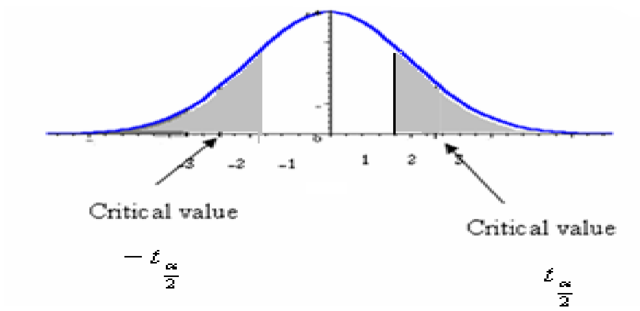
$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$ This two–tailed test

The key word is (Is there a difference)

Step 2: Select the level of significance.

α = 0.10 as stated in the problem



Step 3: Select the test statistic.

Use t–distribution since the assumptions are met

$\overline{X}_1 = 4$, $S_1 = 2.92$, $n_1 = 5$

$\overline{X}_2 = 5$, $S_2 = 2.09$, $n_2 = 6$

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{4\times 8.53 + 5\times 4.37}{9} = \frac{34.12+21.85}{9} = \frac{55.97}{9} = 6.22$$

$S_p = 2.49$

$$t_c = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{4-5}{(2.49)\sqrt{\frac{1}{5}+\frac{1}{6}}} = \frac{-1}{(2.49)(0.609)} = \frac{-1}{1.5146} = -0.662$$

Step 4: Formulate the decision rule.

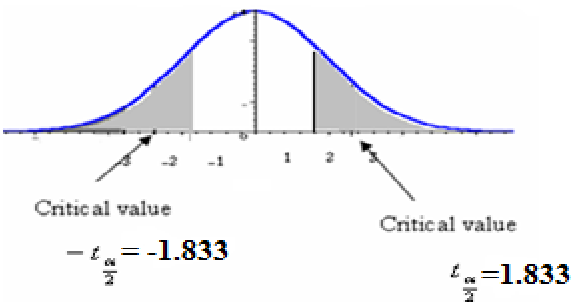
$$\pm t\left(\frac{\alpha}{2}, n_1+n_2-2\right) = \pm t\left(\frac{0.10}{2}, 5+6-2\right) = \pm t(0.05, 9) = \pm 1.833$$

Step 5: Make a decision and interpret the result.

Reject H_0 **if** $t_c > t_{v,\frac{\alpha}{2}}$ $t_c > 1.833$

Or

$t_c < -t_{v,\frac{\alpha}{2}}$ $t_c < -1.833$



The decision is not to reject the null hypothesis, because 0.66 falls in **the region between -1.833 and 1.833.**

The confidence limits are:

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\left(\frac{\alpha}{2}, n_1+n_2-2\right)}, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

$$(\bar{X}_1 - \bar{X}_2) \pm t_{n_1+n_2-2;\frac{\alpha}{2}}\, S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

S_p^2 Is the pooled (the common) variance is...

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{(15-1)2.6^2 + (15-1)1.9^2}{15+15-2}$$

$$\frac{14(2.6)^2 + 14(1.9)^2}{15+15-2} = \frac{94.64+50.54}{28} = \frac{145.18}{28} = 5.185$$

$S_p = 2.277$

$$\left(\overline{X}_1 - \overline{X}_2\right) \pm t_{n_1+n_2-2;\frac{\alpha}{2}}\, S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$= (4-5) \pm t_{9;0.05}(2.277)\sqrt{\frac{1}{5} + \frac{1}{6}}$$

$$= -1 \pm 1{,}833(2.277)\sqrt{0.2 + 0.1667}$$

$$= -1 \pm 4.1737\sqrt{0.3667}$$

$$= -1 \pm 4.1737(0.6056)$$

$$= -1 \pm 2.5274$$

$$-3.5274 \le \widehat{\mu}_1 - \widehat{\mu}_2 \le 1.5274$$

$$(-3.5274\ ,\ 1.5274)$$

Because the interval does include zero, we do not reject $H_0 : \mu_1 = \mu_2$

Test of hypothesis about the difference between Two population means

(Dependent or paired samples)

Purpose is to compare means of two non- independent samples. Two ways to obtain related samples:

1- Repeated measures designs: for, measurements on the same individuals before and after a treatment.

e.g.,Memory retention in quiet and noisy environment (on same observations)

Measure heart rate before & after exercise (on same observations)

2- **Matched samples** : each Population in one sample matched, on specific variable(s), with a Population in another sample

We will use the five steps Hypotheses – testing procedure:

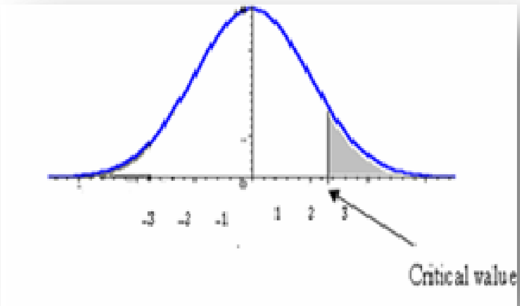
Step (1): State the null and alternate hypotheses:

Case1: One-tailed test (Right)

Upper tail critical (when $X_1 > X_2$)

H_0: $\mu_d \leq 0$

H_1: $\mu_d > 0$

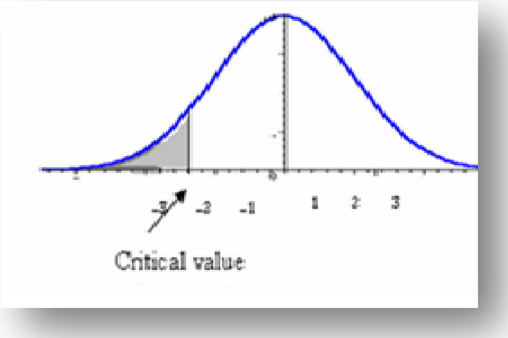


Case2: One-tailed test (Left)

Lower tail critical (when $X_1 < X_2$)

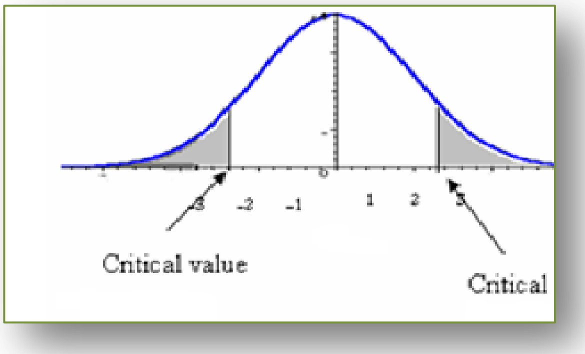
$H_{0:}$ $\mu_d \geq 0$

$H_{1:}$ $\mu_d < 0$



Case3: (Two-tailed test) (when $X_1 \neq X_2$)

H_0: $\mu_d = 0$
H_1: $\mu_d \neq 0$



Step (2): Select the level of significance (α)
Step (3): **The test statistic**

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

d_i : The **difference** between the paired or related observations.
$d_i = (X_{i1} - X_{i2})$
$\bar{d}$: The mean of the difference between the paired or related observations.
S_d : The standard deviation of the paired or related observations.

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

n : Number of paired differences.

Step (4): The critical value:

Types of test	critical value (t)
The one – tailed test (Right)	$t_{(\alpha\,,\mathrm{n}-1)}$
The one – tailed test (left)	$-t_{(\alpha\,,\mathrm{n}-1)}$
The two – tailed test	$\pm t_{\left(\frac{\alpha}{2},\mathrm{n}-1\right)}$

Step (5): Formulate the Decision Rule

Case1: Reject H_o if $t_c > t_{(\alpha\,,n-1)}$

Case2: Reject H_o if $t_c > -t_{(\alpha\,,n-1)}$

Case3: Reject H_o if

$t_c > t_{(\alpha/2\,,n-1)}$ or $t_c < -t_{(\alpha/2\,,n-1)}$

The Paired Difference Confidence Interval μ_D is:

$$\hat{\mu}_D = \overline{D} \pm t_{\alpha/2}\frac{S_D}{\sqrt{n}}$$

Example (4)

Advertisements by Sylph Fitness Center claim that completing its course will result in losing weight. A random sample of eight recent participants showed the following weights before and after completing the course.

- At the 0.01sigenificance level, can we conclude the students lost weight (in pounds)

-Find the confidence interval for μ_D

Note: 1 kg = 2.20 pounds

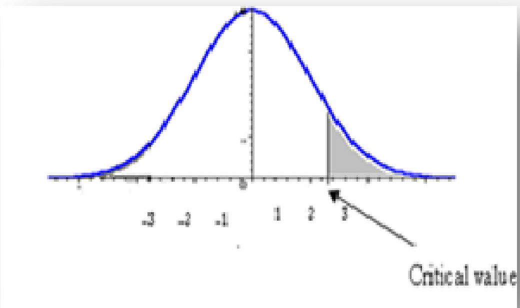
No	Before	After
1	155	154
2	228	207
3	141	147
4	162	157
5	211	196
6	164	150
7	184	170
8	172	165

Solution:

Step (1): State the null and alternate hypotheses

$X_1 > X_2$

$H_0 : \mu_d \leq 0 \; H_1 : \mu_d > 0$



Step (2): Select the level of significance (α=0.01)
Step (3): The test statistic

No	Before B	After A	D (B-A)	$\left(D_i - \overline{D}\right)^2$
1	155	154	1	62.02
2	228	207	21	147.02
3	141	147	-6	221.27
4	162	157	5	15.02
5	211	196	15	37.52
6	164	150	14	26.27
7	184	170	14	26.27
8	172	165	7	3.52
Total			71	538.66

$$\bar{d} = \frac{\sum d}{n} = \frac{71}{8} = 8.875$$

$$S_d = \sqrt{\frac{\sum\left(d - \bar{d}\right)^2}{n-1}} = \sqrt{\frac{538.66}{7}} = 8.7722$$

$$t_c = \frac{\bar{d}}{S_d/\sqrt{n}} = \frac{8.875}{8.7722/\sqrt{8}} = \frac{8.875}{8.7722/2.8284} = \frac{8.875}{3.1015} = 2.8615$$

Step (4): The critical value

In one –tailed test (Right)

$t_{(\alpha,n-1)} = t_{(0.01,7)}$ =2.998

Reject H_0 **if**

$t_c > t_{0.01,7}$

Step (5): Formulate the Decision Rule

Do not reject H_0. We cannot conclude that the students lost weight

The Paired Difference Confidence Interval μ_D is:

$$\hat{\mu}_D = \overline{D} \pm t_{\alpha/2}\frac{S_D}{\sqrt{n}}$$

$$= 8.875 \pm 3.499\frac{8.7722}{\sqrt{8}}$$

$$= 8.875 \pm 10.8519$$

$$-1.9769 < \hat{\mu}_D < 19.7269$$

Since this interval contains 0 you are 99% confident that $H_0 : \mu_d \leq 0$

Do not reject H_0

Example (5)

Assume you send your salespeople to a "customer service" training workshop.

–Has the training made a difference in the number of complaints (at the 0.01 level)?

-Find the confidence interval for μ_D

You collect the following data:

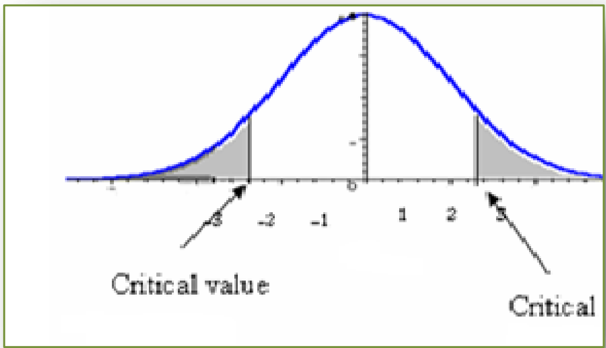
Salesperson	Number of Complaints Before	Number of Complaints After
C.B.	6	4
T.F.	20	6
M.H.	3	2
R.K.	0	0
M.O.	4	0

Solution:

Step (1): State the null and alternate hypotheses

(when $X_1 \neq X_2$)

$H_0 : \mu_d = 0 \quad H_1 : \mu_d \neq 0$



Step (2): Select the level of significance (α=0.01)
Step (3): The test statistic

Salesperson	Number of Complaints Before	Number of Complaints After	D (B-A)	$\left(D_i - \overline{D}\right)^2$
C.B.	6	4	2	4.84
T.F.	20	6	14	96.04
M.H.	3	2	1	10.24
R.K.	0	0	0	17.64
M.O.	4	0	4	0.04
Total			21	128.8

$$\bar{d} = \frac{\sum d}{n} = \frac{21}{5} = 4.2$$

$$S_d = \sqrt{\frac{\sum\left(d - \bar{d}\right)^2}{n-1}} = \sqrt{\frac{128.8}{4}} = 5.6745$$

$$t_c = \frac{\bar{d}}{S_d / \sqrt{n}} = \frac{4.2}{5.6745 / 2.2361} = \frac{4.2}{2.5377} = 1.66$$

Step (4): The critical value
(Two-tailed test)

$$t_{\left(\frac{\alpha}{2}, n-1\right)} = t_{(0.005, 5)} = 4.604$$

Reject H_0 if

$t_c > 4.604$

or

$t_c < -4.604$

Step (5): Formulate the Decision Rule
Do not reject H_0 (t_{stat} is not in the rejection region)
There is insufficient of a change in the number of complaints.

−The Paired Difference Confidence Interval μ_D is:

$$\hat{\mu}_D = \overline{D} \pm t_{\alpha/2} \frac{S_D}{\sqrt{n}}$$

$$= 4.2 \pm 4.604 \frac{5.6745}{\sqrt{5}}$$

$$= 4.2 \pm 11.6836$$

$$-7.48 < \hat{\mu}_D < 15.87$$

Since this interval contains 0 you are 99% confident that $\mu_D = 0$

Do not reject H_0

Tow-Samples Tests of Hypothesis
Test of hypothesis about the difference between Two population Proportions

We investigate whether two samples came from populations with an equal proportion of successes.

Step (1): State the Null (H_0) and alternate (H_1) hypothesis

Case 1: $\begin{array}{l} H_0 : \pi_1 \geq \pi_2 \\ H_1 : \pi_1 < \pi_2 \end{array}$ **OR** $\begin{array}{l} H_0 : \pi_1 - \pi_2 \geq 0 \\ H_1 : \pi_1 - \pi_2 < 0 \end{array}$

Case 2 : $\begin{array}{l} H_0 : \pi_1 \leq \pi_2 \\ H_1 : \pi_1 > \pi_2 \end{array}$ **OR** $\begin{array}{l} H_0 : \pi_1 - \pi_2 \leq 0 \\ H_1 : \pi_1 - \pi_2 > 0 \end{array}$

Case 3: $\begin{array}{l} H_0 : \pi_1 = \pi_2 \\ H_1 : \pi_1 \neq \pi_2 \end{array}$ **OR** $\begin{array}{l} H_0 : \pi_1 - \pi_2 = 0 \\ H_1 : \pi_1 - \pi_2 \neq 0 \end{array}$

Step (2): Select a level of significance.

Step (3): Select the Test Statistic (computed value)

The value of the test statistic is computed from the following formula.

$$Z_c = \frac{P_1 - P_2}{\sqrt{\frac{\overline{P}\left(1-\overline{P}\right)}{n_1} + \frac{\overline{P}\left(1-\overline{P}\right)}{n_2}}}$$

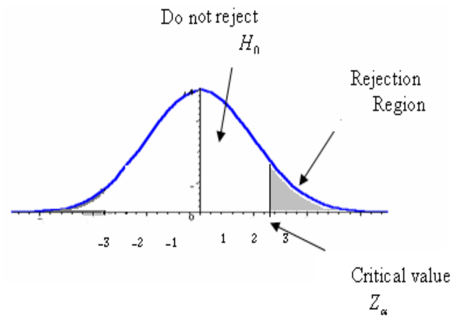
The two samples are pooled using the following formula.

$$\overline{P} = \frac{X_1 + X_2}{n_1 + n_2}$$

Step (4): **Selected the Critical value**

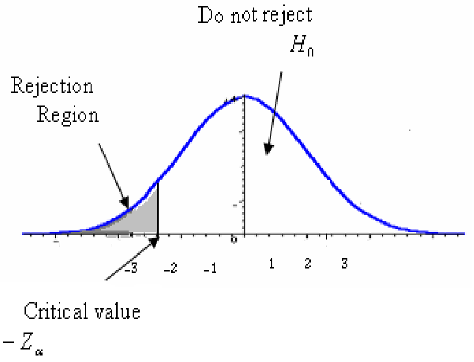
The one – tailed test (Right)	$\mathbf{Z_\alpha}$
The one – tailed test (left)	$-Z_\alpha$
The two – tailed test	$\pm Z_{\alpha/2}$

Step (5): **Formulate the Decision Rule and Make a Decision**



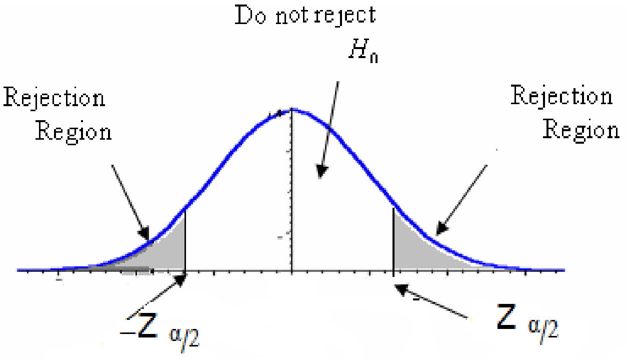
Case1: Reject H_0 if $Z_c > Z_\alpha$

Case2: Reject H_0 if $|Z_c| > Z_\alpha$ or $Z_c < -Z_\alpha$



Case3: Reject H_0 if $|Z_c| > Z_{\frac{\alpha}{2}}$ which mean

$$Z_c > Z_{\frac{\alpha}{2}} \quad or Z_c < -Z_{\frac{\alpha}{2}}$$



Confidence Interval for Two Population Proportions

$$\hat{\pi}_1 - \hat{\pi}_2 = (p_1 - p_2) \pm Z_{\alpha/2} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Example (6)

Manila Perfume Company recently developed a new fragrance that it plans to market under the name Heavenly. A number of market studies indicate that Heavenly has very good market potential. The Sales Department at Minnelli is particularly interested in whether there is a difference in the proportions of younger and older women who would purchase Heavenly if it were marketed. There are two independent populations; a population consisting of the100 younger women, 19 of them liked the Heavenly and a population consisting of the 200 older women. 62 of them liked Heavenly. Each sampled woman will be asked to smell Heavenly and indicate whether she likes the fragrance well enough to purchase a bottle. (With significant level 0.05)

-Find a 95% Confidence Interval for the difference in the proportion**.**

Solution:

Step 1: State the null and alternate hypotheses.

H_0: π1= π2

H_1: π1 ≠ π2

Step 2: State the level of significance.

The .05 significance level is stated in the problem.

Step 3: Find the appropriate test statistic.

We will use the z-distribution

$let\ \ P_1 = young\ \ women$

$$P_1 = \frac{X_1}{n_1} = \frac{19}{100} = 0.19$$

$let\ \ P_2 = Older\ \ women$

$$P_2 = \frac{X_2}{n_2} = \frac{62}{200} = 0.31$$

The pooled estimate for the overall proportion is:

$$\overline{P} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{19+62}{100+200} = \frac{81}{300} = 0.27$$

$$Z_c = \frac{P_1 - P_2}{\sqrt{\frac{\overline{P}\left(1-\overline{P}\right)}{n_1} + \frac{\overline{P}\left(1-\overline{P}\right)}{n_2}}} = \frac{0.19-0.31}{\sqrt{\frac{0.27(1-0.27)}{100} + \frac{0.27(1-0.27)}{200}}} = -2.21$$

Step 4: Selected the Critical value.

$$\pm Z_{\frac{\alpha}{2}} = \pm Z_{\frac{0.05}{2}} = \pm Z_{0.025}$$

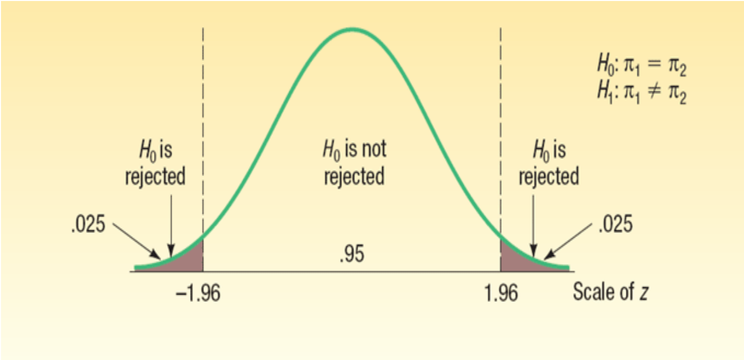
0.5 – 0.025 = 0.475

$$\pm Z_{0.025} = \pm 1.96$$

Step 5: State the decision rule

Reject H_0 if $Z_c > Z_{\frac{\alpha}{2}}$ Or $Z_c < -Z_{\frac{\alpha}{2}}$

$Z_c > 1.96$ Or $Z_c < -1.96$



Whereas $Z_c < -1.96$, -2.21<-1.96

Then Reject H0 at significant level 0.05

The computed value of 2.21 is in the area of rejection. Therefore, the null hypothesis is rejected at the .05 significance level. To put it another way, we reject the null hypothesis that the proportion of young women who would purchase Heavenly is equal to the proportion of older women who would purchase Heavenly.

Confidence Interval for Two Population Proportions

$$\hat{\pi}_1 - \hat{\pi}_2 = (p_1 - p_2) \pm Z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

$$= (0.19 - 0.31) \pm 1.96\sqrt{\frac{0.19(1-0.19)}{100} + \frac{0.31(1-0.31)}{200}}$$

$$= -0.12 \pm 1.96 \times 0.0511$$

$$= -0.12 \pm 0.1001$$

$$-0.2201 \le \hat{\pi}_1 - \hat{\pi}_2 \le -0.0199$$

Since this interval does not contain 0 can be 95% confident the two proportions are different.

We reject that $H_0 : \pi_1 = \pi_2$

Example (7)

Is there a significant difference between the proportion of men and the proportion of women who will vote Yes on Proposition A? In a random sample, 36 of 72 men and 35 of 50 women indicated they would vote Yes (With significant level 0.05)

-Find a 95% Confidence Interval for the difference in the proportion**.**

Solution:

Step 1: State the null and alternate hypotheses.

$$H_0: \pi1 = \pi2$$

$$H_1: \pi1 \neq \pi2$$

Step 2: State the level of significance.

The .05 significance level is stated in the problem.

Step 3: Find the appropriate test statistic.

We will use the z-distribution

$$P_1 = \frac{X_1}{n_1} = \frac{36}{70} = 0.5$$

$$P_2 = \frac{X_2}{n_2} = \frac{35}{50} = 0.70$$

The pooled estimate for the overall proportion is:

$$\bar{P} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{36+35}{72+50} = \frac{71}{122} = 0.582$$

$$Z_c = \frac{P_1 - P_2}{\sqrt{\frac{\bar{P}(1-\bar{P})}{n_1} + \frac{\bar{P}(1-\bar{P})}{n_2}}} = \frac{0.5 - 0.70}{\sqrt{\frac{0.582(1-0.57)}{72} + \frac{0.582(1-0.57)}{50}}} = -2.20$$

Step 4: Selected the Critical value.

$$\pm Z_{\frac{\alpha}{2}} = \pm Z_{\frac{0.05}{2}} = \pm Z_{0.025}$$

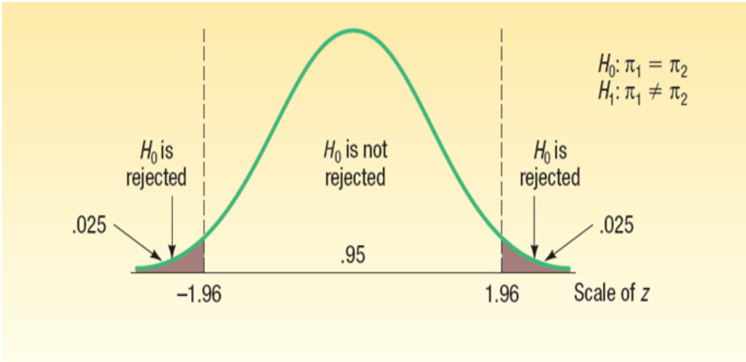
$$0.5 - 0.025 = 0.475$$

$$\pm Z_{0.025} = \pm 1.96$$

Step 5: State the decision rule

Reject H_0 if $Z_c > Z_{\frac{\alpha}{2}}$ Or $Z_c < -Z_{\frac{\alpha}{2}}$

$Z_c > 1.96$ Or $Z_c < -1.96$



Whereas $Z_c < -1.96$, -2.21<-1.96

Then Reject H0 at significant level 0.05

There is evidence of a significant difference in the proportion of men and women who will vote yes.

Confidence Interval for
Two Population Proportions

$$\hat{\pi}_1 - \hat{\pi}_2 = (p_1 - p_2) \pm Z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

$$= (0.50 - 0.70) \pm 1.96\sqrt{\frac{0.50(1-0.50)}{72} + \frac{0.70(1-0.70)}{50}}$$

$$= -0.20 \pm 1.96 \times 0.0876$$

$$= -0.20 \pm 0.1717$$

$$-0.37 \le \hat{\pi}_1 - \hat{\pi}_2 \le -0.03$$

Since this interval does not contain 0 can be 95% confident the two proportions are different.

We reject that $H_0 : \pi_1 = \pi_2$

Test of hypothesis about the difference between Two population Variances

Step (1): State the null (H_0) and alternate (H_1) hypothesis

Case 1:
$$H_0 : \sigma_1^2 \le \sigma_2^2$$
$$H_1 : \sigma_1^2 > \sigma_2^2$$

Case 2:
$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \ne \sigma_2^2$$

Step (2): Select a level of significance.

Step (3): Select the Test Statistic (computed value)

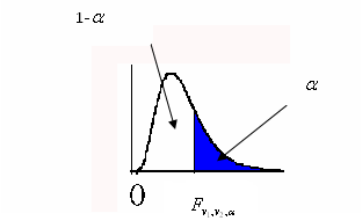
$$F_{STAT} = \frac{S_1^2}{S_2^2}$$

The properties of the F distribution (F_{v_1,v_2})

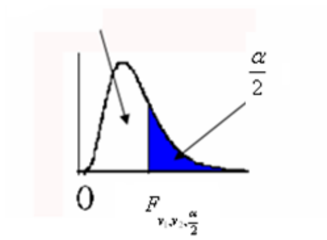
- Is continuous distribution.
- Positive skewed curve (skewed to the right curve).
- It is not symmetric curve.

Step (4): Selected the Critical value

The one – tailed test (Right)	$F_{(\nu_1 \nu_2, \alpha)}$



The two – tailed test	$F_{\left(\nu_1 \nu_2, \frac{\alpha}{2}\right)}$



Step (5): Formulate the Decision Rule and Make a Decision

Case1: Reject H_0 if $F_{STAT} > F_{(\nu_1 \nu_2, \alpha)}$

Case2: Reject H_0 if $F_{STAT} > F_{\left(\nu_1 \nu_2, \frac{\alpha}{2}\right)}$

Example (8)

Waiting time is a critical issue at fast-food chains, which not only want to minimize the mean service time but also want to minimize the variation in the service time from customer to customer. One fast-food chain out a study to measure the variability in the waiting time (defined as the time in minutes from when an order was completed to when it was delivered to the customer) at lunch and breakfast at one of the chain's stores. The results were as follows:

	Sample Standard deviations	Sample size
Lunch	4.4 minutes	25
Breakfast	1.9 minutes	21

At the 0.05 level of significance, is there evidence that there is more variability in the service time at lunch than at breakfast? Assume that the population service times are normally distributed

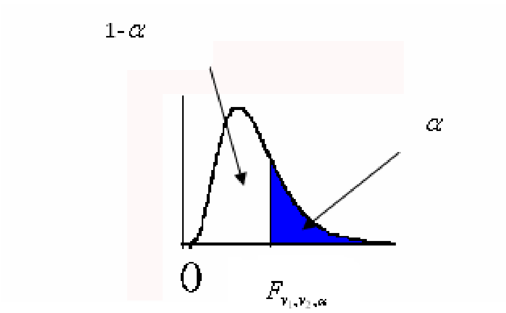
Solution:

Step 1: State the null hypothesis and the alternate hypothesis.

$H_0 : \sigma_1^{\ 2} \le \sigma_2^2$

$H_1 : \sigma_1^{\ 2} > \sigma_2^2$

The one – tailed test (Right)



Step 2: Select the level of significance.

α = 0.05

$F_{(\nu_1 \nu_2, \alpha)} = F_{(24,20\ ,0.05)} = 2.08$

Step 3: Formulate the decision rule.

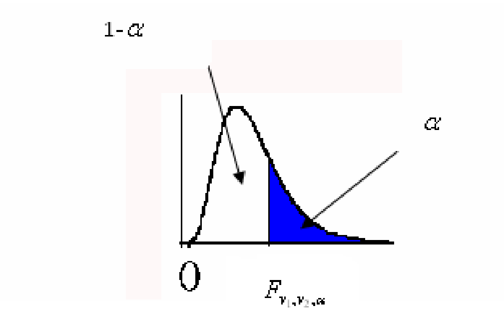
Reject H_0 if $F_{STAT} > 2.08$

Step 4: Select the test statistic.

$$F_{STAT} = \frac{S_1^2}{S_2^2} = \frac{4.4}{1.9} = 2.3158$$

V2	F Table for alpha=.05													
	V1													
	1	2	3	4	5	6	7	8	9	10	12	15	20	24
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95	248.01	249.05
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08

Step 5: Make a decision and interpret the result.



Because $F_{STAT} = 2.3158 > 2.08$ you reject H_0 .Using a 0.05 level of significance, you conclude that there is evidence that there is more variability in the service time at lunch than at breakfast

Example (9)

You are a financial analyst for a brokerage firm. You want to compare dividend yields between stocks listed on the NYSE & NASDAQ. You collect the following data:

	NYSE	NASDAQ
Number	21	25
Mean	3.27	2.53
Std. dev.	1.30	1.16

<u>Is there a difference in the variances between the NYSE & NASDAQ at the</u> $\alpha = 0.05$ level?

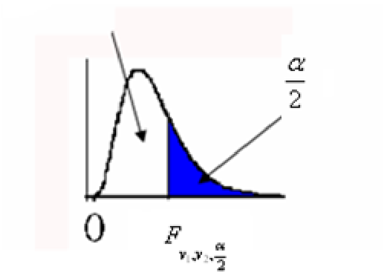
Solution:

Step 1: State the null hypothesis and the alternate hypothesis.

$H_0 : {\sigma_1}^2 = \sigma_2^2$

$H_1 : {\sigma_1}^2 \neq \sigma_2^2$

This is two-tailed test



(Note: keyword in the problem "is there a difference")

Step 2: Select the level of significance.

$\alpha = 0.05$ as stated in the problem , $\alpha/2 = 0.025$

$F_{\left(\nu_1 \nu_2, \frac{\alpha}{2}\right)} = F_{(20,24,0.025)} = 2.33$

	F Table for alpha=.025													
V2	**V1**													
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**20**	**24**
1	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28	968.63	976.71	984.87	993.10	997.25
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27

Step 3: Formulate the decision rule.

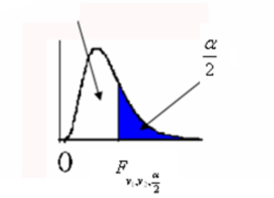
Reject H_0 if $F_{STAT} > 2.33$

Step 4: Select the test statistic.

$$F = \frac{S_1^2}{S_2^2} = \frac{(1.30)^2}{(1.16)^2} = 1.256$$

Step 5: Make a decision and interpret the result.

$F_{STAT} = 1.256$ is not in the rejection region, so we do not reject H_0



Example (10)

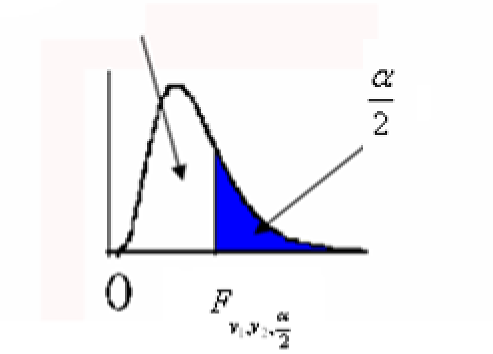
Lammers Limos offers limousine service from the city hall in Toledo, Ohio, to Metro Air port in Detroit. Sean Lammers, president of the company, is considering two routes. One is via U.S. 25 and the other via I-75. He wants to study the time it takes to drive to the airport using each route and then compare the results. He collected the following sample data, which is reported in minutes.

Route	Sample mean	Sample Standard deviations	Sample size
Route 25	58.29 minutes	8.9947 minutes	7
Route i-75	59minutes	4.3753 minutes	8

Using the 0.10 significance level, is there a difference in the variation in the driving times for the two routes?

Solution:

Step 1: State the null hypothesis and the alternate hypothesis.



$H_0 : \sigma_1{}^2 = \sigma_2^2$

$H_1 : \sigma_1{}^2 \neq \sigma_2^2$

This is two-tailed test

(Note: keyword in the problem "is there a difference")

Step 2: Select the level of significance.

α = 0.10 as stated in the problem , α/2 = 0.05

$F_{\left(v_1 v_2, \frac{\alpha}{2}\right)} = F_{(6,7,0.05)} = 3.87$

Step 3: Formulate the decision rule.

Reject H_0 if

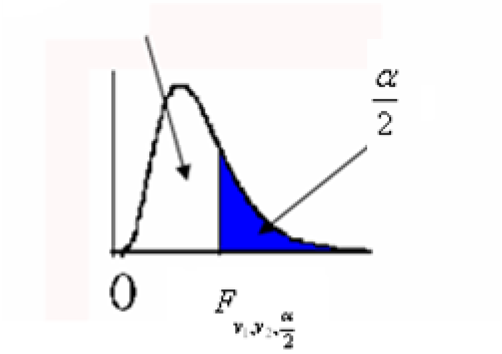
$F_{stat} > 3.87$

Step 4: Select the test statistic.

$$F_{STAT} = \frac{S_1^2}{S_2^2} = \frac{(8.9947)^2}{(4.3753)^2} = 4.23$$

F Table for alpha=.05											
V2	V1										
	1	2	3	4	5	6	7	8	9	10	12
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07

Step 5: Make a decision and interpret the result.



The decision is to reject the null hypothesis, because the computed F-Value (4.23) is larger than the critical value (3.87). We conclude that there is a difference in the variation of the travel times along the two routes.

Analysis of Variance (ANOVA)

one way ANOVA

Definition:

Analysis of variance is a general method for studying sampled-data relationships. The method enables the difference between **two or more** sample means to be analyzed, achieved by subdividing the total sum of squares.

Basic idea is to partition total variation of the data into two sources:

1. Variation within groups.
2. Variation between groups.

One way ANOVA is the simplest case.

Assumptions

- The populations from which the samples were obtained must be **normally or approximately normally** distributed.
- The populations must be **independent.**
- The **variances** of the populations must be **equal**

Data frame:

Individuals (Observations)	Groups (Populations)				
	1	2	…. I ……	K	Total
1	Y_{11}	Y_{21}	……. Y_{i1}	Y_{k1}	
2	Y_{12}	Y_{22}	……..Y_{i2}	Y_{k2}	
….	…	….	….	….	
j	Y_{1j}	Y_{2j}	……. Y_{ij}	Y_{kj}	
n	Y_{1n}	Y_{2n}	……..Y_{in}	Y_{kn}	
Sum	$Y_{1.}$	$Y_{2.}$	……..$Y_{i.}$	$Y_{k.}$	$Y_{..}$
Mean	$\bar{Y}_{1.}$	$\bar{Y}_{2.}$	$\bar{Y}_{3.}$	$\bar{Y}_{k.}$	$\bar{Y}_{...}$

Y_{ij} : The jth sample observation selected from group or population i

n_i : The number of sample observations selected from population i

n : The total sample size ($n = n_1 + n_2 + \ldots\ldots + n_k$)

$Y_{i.}$: The sum (total) of the sample measurements obtained from population i.

$Y_{..}$: The grand total.

Step (1): State the null and alternate hypotheses:

H_0: $\mu_1 = \mu_2 = \mu_3 = \dots\dots = \mu$

H_1: Not all μ_j are equal.

Step (2): Select the level of significance (α)
Step (3): The test statistic

Because we are comparing means of more than two groups, use the F statistic.

The F test statistic is found by dividing the between group variance (MSB) by the within group variance (MSW).

$$\mathbf{F_c = \frac{MSA}{MSW} = \frac{SSA/c-1}{SSW/n-c}}$$

SSA: **Sum of Squares among Groups**. (This variation due to the interaction between the samples)
SSW: Sum of Squares Within groups. (This variation due to differences within individual samples).
SST: Total Sum of Squares. (The total variation is comprised the sum of the squares of the differences of each mean with the grand mean)

$$SSA = \sum_{i=1}^{K} n_i\,(\bar{Y}_i - \bar{Y}_{..})^2 = \sum_{i=1}^{k}\left(\frac{Y_{i.}^2}{n_i}\right) - \frac{Y_{..}^2}{n}$$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Y_{ij} - \bar{Y}_{i.}\right)^2 = \sum\sum Y_{ij}^2 - \frac{Y_{..}^2}{n}$$

SSW = SST – SSA

Step (4): The critical value:

The degrees of freedom for the **numerator** are the degrees of freedom for the between group (k-1) and the degrees of freedom for the **denominator** are the degrees of freedom for the within group (n-k).

$$\mathbf{F_{(\alpha\ ,\ \ c-1\ ,n-c\ )}}$$

Step (5) : Formulate the decision Rule and make a decision

Reject H_o

If $\mathbf{F_c > F_{(\alpha\ ,c-1,n-c)}}$

It is convenient to summarize the calculation of the F statistic in (ANOVA Table)

ANOVA TABLE

Source of variation (S.V)	Sum of Squares (S.S)	Degrees of freedom	Mean Squares (MS)	F- ratio
Between groups	SSA	c-1	MSA =SSR/c-1	F = MSA / MSW
Within groups	SSW	n-c	MSW = SSE/c-k	
Total	SST	n-1		

Example (11):

Recently a group of four major carriers joined in hiring Brunner Marketing Research, Inc., to survey recent passengers regarding their level of satisfaction with a recent flight.

The survey included questions on ticketing, boarding, in-flight service, baggage handling, pilot communication, and so forth. Twenty-five questions offered a range of possible answers: excellent, good, fair, or poor. A response of excellent was given a score of 4, good a 3, fair a 2, and poor a 1. These responses were then totaled, so the total score was an indication of the satisfaction with the flight. Brunner Marketing Research, Inc., randomly selected and surveyed passengers from the four airlines. Is there a difference in the mean satisfaction level among the four airlines? (Construct the ANOVA table)
Use the .01 significance level.

Eastern	TWA	Allegheny	Ozark
94	75	70	68
90	68	73	70
85	77	76	72
80	83	78	65
	88	80	74
		68	65
		65	

Solution:

Step 1: State the null and alternate hypotheses.

H_0: $\mu_E = \mu_A = \mu_T = \mu_O$

H1: The means are not all equal

Step 2: State the level of significance.

The .01 significance level is stated in the problem.

Step (3): Find the appropriate test statistic.

$$\mathbf{F_c = \frac{MSA}{MSW} = \frac{SSA/c-1}{SSW/n-c}}$$

Individuals (Observations)	Groups (Populations)				
	E	**T**	**A**	**O**	**Total**
1	94	75	70	68	
2	90	68	73	70	
3	85	77	76	72	
4	80	83	78	65	
5		88	80	74	
6			68	65	
7			65		
Sum ($Y_{i.}$)	$Y_{1.}$= 349	$Y_{2.}$=391	$Y_{3.}$=510	$Y_{2.}$=414	ΣY_{ij} = Y..= 1664
n	4	5	7	6	22
Mean	$\bar{Y}_{1.}$**=87.25**	$\bar{Y}_{2.}$**=78.2**	$\bar{Y}_{3.}$**=72.86**	$\bar{Y}_{4.}$**=6969**	

$$SST = \sum\sum Y_{ij}^2 - \frac{Y_{..}^2}{n} = (94^2 + 90^2 + \cdots + 65^2) - \left(\frac{1664^2}{22}\right) = 1485.09$$

$$\mathrm{SSA} = \sum_{i=1}^{k}\left(\frac{Y_{i.}^2}{n_i}\right) - \frac{Y_{..}^2}{n} = \left(\frac{349^2}{4} + \frac{391^2}{5} + \frac{510^2}{7} + \frac{414^2}{6}\right) - \left(\frac{1664^2}{22}\right)$$
$$= 890.68$$

SSW = SST – SSA = 1485.09 -890.68 =594.41

$$\mathrm{F} = \frac{\mathrm{MSA}}{\mathrm{MSW}} = \frac{\mathrm{SSA/c-1}}{\mathrm{SSW/n-c}} = \frac{890.68/4-1}{594.41/22-4} = 8.99$$

Step (4): The critical value:
The degrees of freedom for the **numerator = c-1 =4-1=3**
The degrees of freedom for the **denominator = n-c =22-4 =18**

$$\mathbf{F}_{(\alpha\ ,\ c-1\ ,n-c\ )} = \mathbf{F}_{(0.01\ ,\ 3\ ,18\ )} = \mathbf{5.09}$$

V2	V1										F Table for alpha=.01
	1	2	3	4	5	6	7	8	9	10	12
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47	6055.85	6106.32
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89
6	13.75	10.93	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96
14	8.86	6.52	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.90	3.81	3.67
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46
18	8.29	6.01	5.09	4.58	4.25	4.02	3.84	3.71	3.60	3.51	3.37
19	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23

Step (5): State the decision rule.

$$\mathbf{F_c > F_{(0.01,3,18)} = 5.09}$$

Reject H_0

The computed value of F is 8.99, which is greater than the critical value of 5.09, so the null hypothesis is rejected.

Conclusion: The population means are not all equal. The mean scores are not the same for the four airlines; at this point we can only conclude there is a difference in the treatment means. We cannot determine which treatment groups differ or how many treatment groups differ.

ANOVA TABLE

(S.V)	(S.S)	DF	(MS)	F- ratio
Between groups	SSA = 890.68	c-1 = 4-1 =3	MSA =SSA/c-1= 890.86/3 = 296.95	F = MSA / MSW = 296.95/33.02 = 8.99
Within groups	SSW = 594.41	n-c =22-4 =18	MSW = SSW/n-c = 594.41/18 = 33.02	
Total	SST =1485.09	n-1 = 22-1=21		

Example (12):

You want to see if three different golf clubs yield different distances. You randomly select five measurements from trials on an automated driving machine for each club. At the 0.05 significance level, <u>is there a difference in mean distance</u>?

(Construct the ANOVA table)

<u>Club 1</u>	<u>Club 2</u>	<u>Club 3</u>
254	234	200
263	218	222
241	235	197
237	227	206
251	216	204

Solution:

Step 1: State the null and alternate hypotheses.

H_0: $\mu_1 = \mu_2 = \mu_3$

H_1: μ_j not all equal

ANOVA TABLE

(S.V)	(S.S)	DF	(MS)	F- ratio
Between groups	SSA = 4716.4	c-1 = 3-1 =2	MSA =SSA/c-1= 4716.4 /2 = 2358.2	F = MSA / MSW = 2358.2/93.3= 25.28
Within groups	SSW = 1119.6	n-c =15-3 =12	MSW = SSW/n-c = 1119.6/12 = 93.3	
Total	SST 5836.0	n-1 = 15-1=14		

Step 2: State the level of significance.

The .05 significance level is stated in the problem.

Step (3): Find the appropriate test statistic.

$$F_c = \frac{SSA/c-1}{SSW/n-c} = \frac{MSA}{MSW}$$

$$F_c = \frac{4716.4/3-1}{1119.6/15-3} \quad \frac{42358.2}{93.3} = 25.28$$

Step (4): The critical value:
The degrees of freedom for the **numerator = c-1 =3-1=2**
The degrees of freedom for the **denominator = n-c =15-3 =12**

$$F_{(\alpha\ ,\ \ c-1\ ,n-c\ )} = F_{(0.05\ ,\ \ 2\ ,12\ )} = 3.89$$

F Table for alpha=.05

V2	V1											
	1	2	3	4	5	6	7	8	9	10	12	15
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.95
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53

Step (5): State the decision rule.

$$F_c > F_{(0.05\ ,2,12)=3.89}$$

Reject H_0
The computed value of F is 25.28, which is greater than the critical value of 3.89, so the null hypothesis is rejected.
Conclusion: There is evidence that at least one μ_j differs from the rest
.