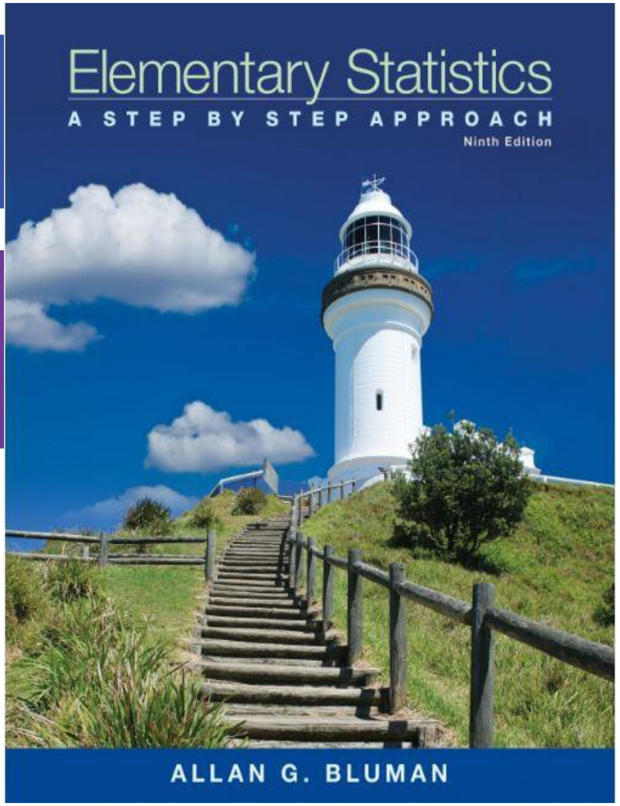
Unleashing Potentials
Shaping the Future

Chapter 4: Hypothesis Testing

Unit 4
One-way ANOVA

Prepared by:
Kamarul Ariffin Mansor
Department of Mathematical Sciences
UiTM Kedah Branch Campus



Learning objective:

Use the one-way ANOVA technique to determine if there is a significant difference among three or more means.

Introduction

- The F test, used to compare two variances, can also be used to compare three of more means.
- This technique is called **analysis of variance** or **ANOVA**.
- For three groups, the F test can only show whether or not a difference exists among the three means, not where the difference lies.

12-1 One-Way Analysis of Variance

- When an *F* test is used to test a hypothesis concerning the means of three or more populations, the technique is called **analysis of variance** (commonly abbreviated as **ANOVA**).
- The **one-way analysis of variance** test is used to test the equality of three or more means using sample variances

Assumptions for the *F* Test

The following assumptions apply when using the *F* test to compare three or more means.

1. The populations from which the samples were obtained must be normally or approximately normally distributed.
2. The samples must be independent of each other.
3. The variances of the populations must be equal.

The *F* Test

- In the *F* test, two different estimates of the population variance are made.
- The first estimate is called the **between-group variance**, and it involves finding the variance of the means.
- The second estimate, the **within-group variance**, is made by computing the variance using all the data and is not affected by differences in the means.

The *F* Test

- If there is no difference in the means, the between- group variance will be approximately equal to the within-group variance, and the *F* test value will be close to 1 - ***do not reject null hypothesis***.
- However, when the means differ significantly, the between-group variance will be much larger than the within-group variance; the *F* test will be significantly greater than 1 - ***reject null hypothesis***.

Example 12-1: Miles per Gallon

A researcher wishes to see if there is a difference in the fuel economy for city driving for three different types of automobiles: small automobiles, sedans, and luxury automobiles. He randomly samples four small automobiles, five sedans, and three luxury automobiles. The miles per gallon for each is shown. At $\alpha = 0.05$, test the claim that there is no difference among the means. The data are shown.

Small	Sedans	Luxury
36	43	29
44	35	25
34	30	24
35	29	
	40	

Using Traditional Method

Step 1: State the hypotheses and identify the claim.

H_0: $\mu_1 = \mu_2 = \mu_3$ (claim)

H_1: At least one mean is different from the others.

Step 2: Find the critical value.

Since $k = 3$, $N = 12$

V1 = d.f.N. = $k - 1 = 3 - 1 = 2$

V2= d.f.D. = $N - k = 12 - 3 = 9$

The critical value is 4.26, obtained from Table 9 – F Distribution.

Step 3: Compute the test value. (see the formula from past year question appendix)

a) Find the mean and variance of each sample

b) Find the **grand mean**, the mean of all values in the samples

$$\overline{X}_{GM} = \frac{\Sigma X}{N} = \frac{36 + 44 + 34 + \cdots + 24}{12} = \frac{404}{12} = 33.667$$

Step 3: Compute the test value. (continue)

c) Find between-group variance, S_B^2

$$S_B^2 = \frac{\sum(\bar{X}_i - \bar{X}_{GM})^2}{k-1} \text{ (between group)}$$

$$= \frac{4(37.25 - 33.667)^2 + 5(35.4 - 33.667)^2 + 3(26 - 33.667)^2}{3 - 1}$$

$$= \frac{242.717}{2} = 121.359$$

d) Find the within-group variance, S_W^2

$$s_W^2 = \frac{\Sigma(n_i - 1)s_i^2}{\Sigma(n_i - 1)} = \frac{(4-1)(20.917) + (5-1)(37.3) + (3-1)7}{(4-1) + (5-1) + (3-1)}$$

$$= \frac{225.951}{9} = 25.106$$

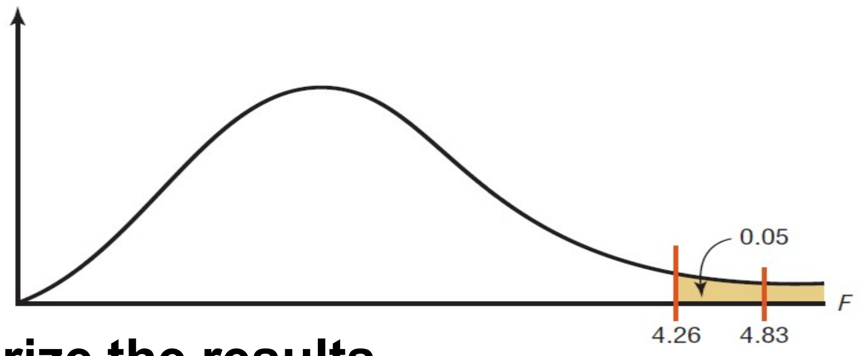
Step 3: Compute the test value. (continued)

e) Compute the F value.

$$F = \frac{s_B^2}{s_W^2} = \frac{121.359}{25.106} = 4.83$$

Step 4: Make the decision.

Reject the null hypothesis, since 4.83 > 4.26.



Step 5: Summarize the results.

There is enough evidence to reject the claim and conclude that at least one mean is different from the others.

ANOVA

- The between-group variance is sometimes called the **mean square, MS_B**.
- The numerator of the formula to compute MS_B is called the **sum of squares between groups, SS_B**.
- The within-group variance is sometimes called the **mean square, MS_W**.
- The numerator of the formula to compute MS_W is called the **sum of squares within groups, SS_W**.

ANOVA Summary Table

TABLE 12–1 Analysis of Variance Summary Table

Source	Sum of squares	d.f.	Mean square	F
Between	SS_B	$k - 1$	MS_B	
Within (error)	SS_W	$N - k$	MS_W	
Total				

TABLE 12–2 Analysis of Variance Summary Table for Example 12–1

Source	Sum of squares	d.f.	Mean square	F
Between	242.717	2	121.359	4.83
Within (error)	225.954	9	25.106	
Total	468.671	11		

SPSS OUTPUT for Example 12-1, ANOVA Table

ANOVA					
fuel					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	242.717	2	121.358	4.834	.038
Within Groups	225.950	9	25.106		
Total	468.667	11			

$\alpha = 0.05$

Based on the above table, using *p*-value method:

H_0: $\mu_1 = \mu_2 = \mu_3$ (claim)
H_1: At least one mean is different from the others.

Since *p*-value = 0.038 < $\alpha = 0.05$, Reject H_0.
Conclusion: At least one mean is different from the others.

TRY THIS YOURSELF:

At the 0.05 level of significance, is there sufficient evidence to conclude that a difference in mean sodium amounts exists among condiments, cereals and desserts. Using p-value method.

Condiments	Cereals	Desserts
270	260	100
130	220	180
230	290	250
180	290	250
80	200	300
70	320	360
200	140	300
		160

ANOVA

Sodium in milligrams

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	27543.506	2	13771.753	2.399	.118
Within Groups	109092.857	19	5741.729		
Total	136636.364	21			

SPSS PROCEDURES
ONE-WAY ANOVA

One-Way ANOVA - F Test

SPSS Procedures Example

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, is there a difference in mean distance?

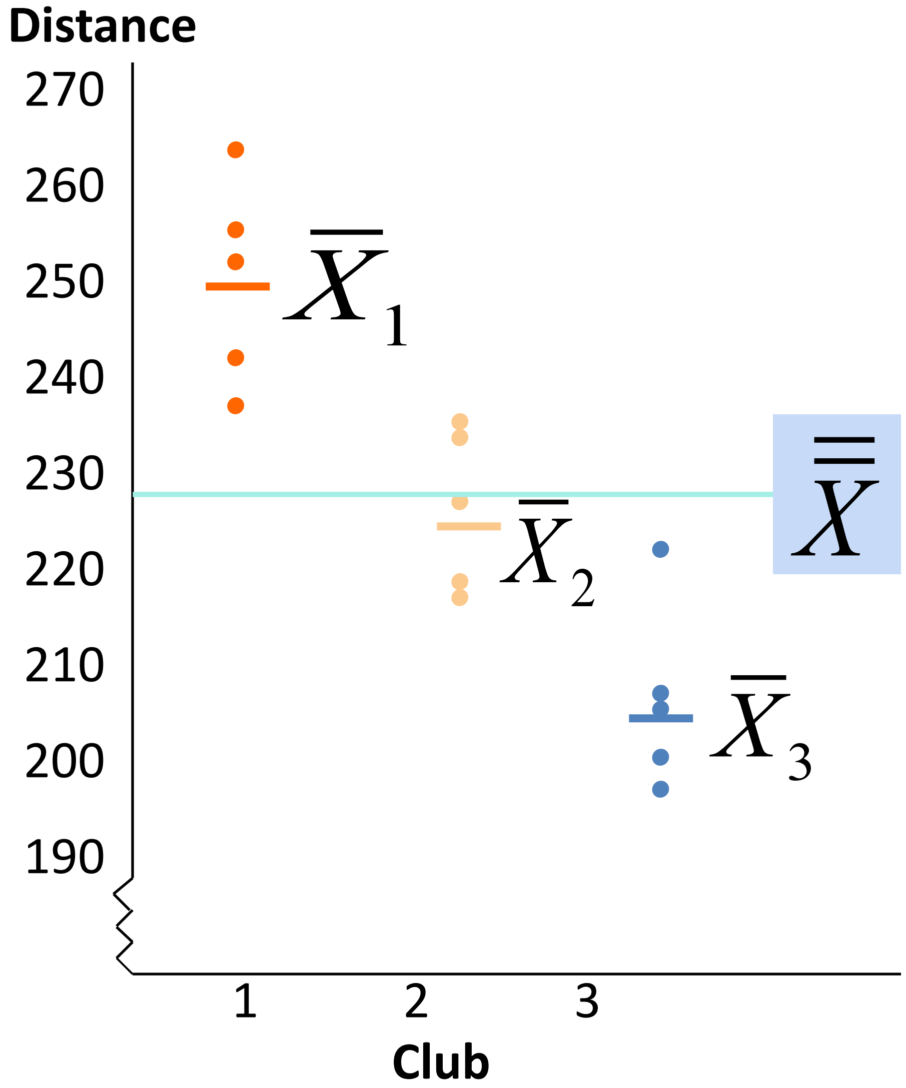
Club 1	Club 2	Club 3
254	234	200
263	218	222
241	235	197
237	227	206
251	216	204

One-Way ANOVA Example: Scatter Diagram

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

$\bar{x}_1 = 249.2$	$\bar{x}_2 = 226.0$	$\bar{x}_3 = 205.8$

$\bar{\bar{x}} = 227.0$



UNIVERSITI TEKNOLOGI MARA

One-Way ANOVA Example Computations

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

$\bar{X}_1 = 249.2$ $\quad n_1 = 5$

$\bar{X}_2 = 226.0$ $\quad n_2 = 5$

$\bar{X}_3 = 205.8$ $\quad n_3 = 5$

$\bar{\bar{X}} = 227.0$ $\quad n = 15$

$c = 3$

$SSA = 5\,(249.2 - 227)^2 + 5\,(226 - 227)^2 + 5\,(205.8 - 227)^2 = 4716.4$

$SSW = (254 - 249.2)^2 + (263 - 249.2)^2 + \ldots + (204 - 205.8)^2 = 1119.6$

$MSA = 4716.4 \,/\, (3\text{-}1) = 2358.2$

$MSW = 1119.6 \,/\, (15\text{-}3) = 93.3$

$$F = \frac{2358.2}{93.3} = 25.275$$

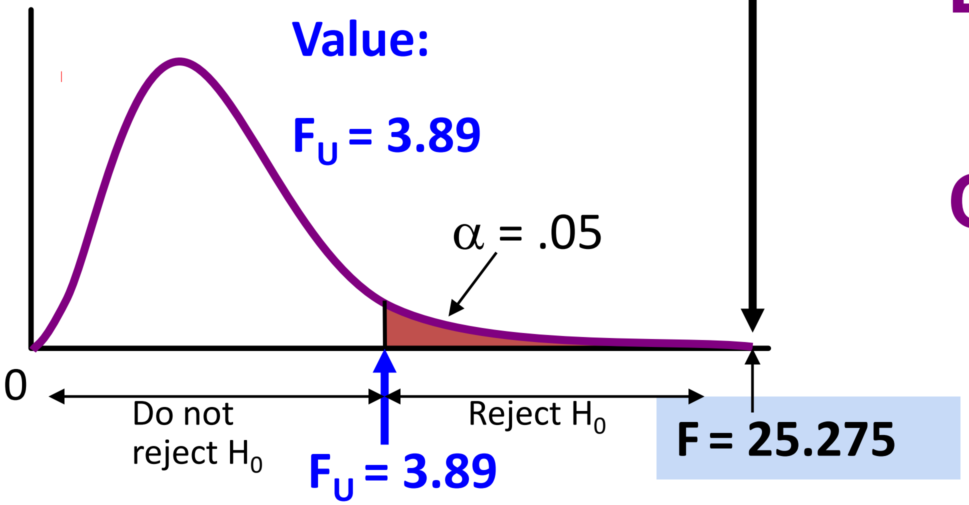
One-Way ANOVA Example Solution

H_0: $\mu_1 = \mu_2 = \mu_3$
H_1: μ_j not all equal

$\alpha = 0.05$
$df_1 = 2$ $df_2 = 12$

Critical Value:

$F_U = 3.89$



Test Statistic:

$$F = \frac{MSA}{MSW} = \frac{2358.2}{93.3} = 25.275$$

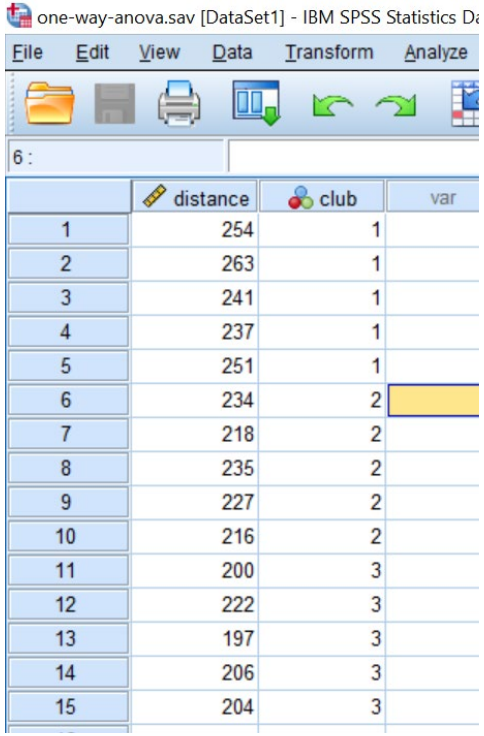
Decision:

Reject H_0 at $\alpha = 0.05$

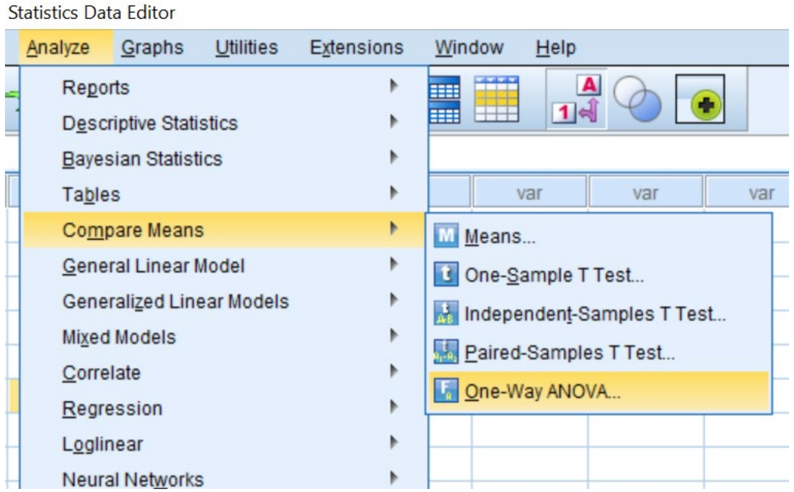
Conclusion:

There is evidence that at least one μ_j differs from the rest

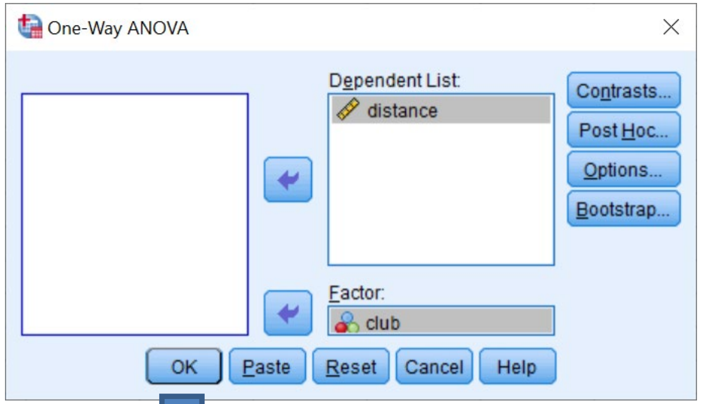
Step 1: Enter data into SPSS



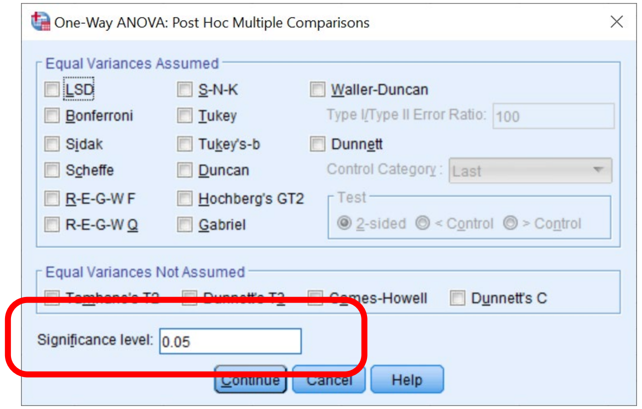
Step 2: Select analysis



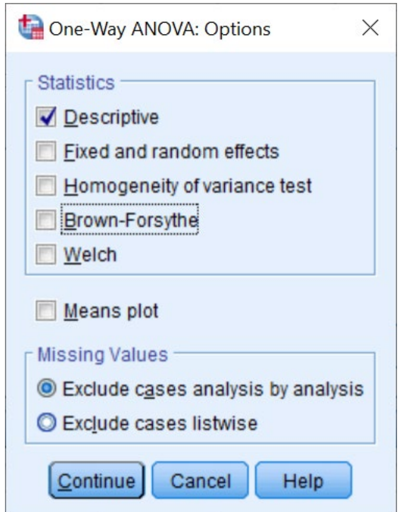
Step 3: Move variable in appropriate box



Step 4: Set significant level



Step 5: Select Options button and tick (✓) Descriptive box. Then "Continue"



Step 6: Press "OK" to run analysis.

SPSS Output

Oneway

Descriptives

distance

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
1	5	249.20	10.402	4.652	236.28	262.12	237	263
2	5	226.00	8.803	3.937	215.07	236.93	216	235
3	5	205.80	9.706	4.341	193.75	217.85	197	222
Total	15	227.00	20.417	5.272	215.69	238.31	197	263

❶ ❷

ANOVA

distance

❸ ❹

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4716.400	2	2358.200	25.275	.000
Within Groups	1119.600	12	93.300		
Total	5836.000	14			

❶ Sample mean and standard deviations
❷ grand mean and standard deviation
❸ Calculated F-statistic value from sample
❹ p-value (Sig. = .000 means that p-value < 0.001)

END OF SLIDES PRESENTATIONS

Prepared by:
Kamarul Ariffin Mansor
Department of Mathematical Sciences
UiTM Kedah Branch Campus

1 2 3 4 5 6 7 8 9 10

1 2 3 4 5 6 7 8 9 10