UNIVERSITI TEKNOLOGI PETRONAS

Chapter 10
One Way ANOVA

Learning Outcomes:

At the end of the lesson student should be able to:

- Understand the concept of one-way ANOVA.
- Know how to construct ANOVA table.
- Understand how to use ANOVA to analyze data from these experiments.

The idea of ANOVA

- A **factor** is a variable that can take one of several **levels** used to differentiate one group from another.
- An experiment has a **one-way,** or **completely randomized design** if several levels of one factor are being studied and the individuals are randomly assigned to its levels. (There is only one way to group the data.)

The idea of ANOVA

- **Analysis of variance (ANOVA)** is the technique used to determine whether more than two population means are equal.

- **One-way ANOVA** is used for completely randomized, one-way designs.

One-Way ANOVA

- The *response* variable is the variable you're comparing
- The *factor* variable is the categorical variable being used to define the groups
  - We will assume *k* samples (groups)
- The *one-way* is because each value is classified in exactly one way
  - Examples include comparisons by gender, race, color, etc.

One-Way ANOVA

- Conditions or Assumptions:
  - The data are randomly sampled
  - The variances of each sample are assumed equal
  - The residuals are normally distributed

One-Way ANOVA

- The null hypothesis is that the means are all equal

- The alternative hypothesis is that at least one of the means is different

One-Way ANOVA

- The ANOVA doesn't test that one mean is less than another
- Its only test whether all means are equal or at least one is different.
- The hypothesis is:

$$H_0 : \mu_F = \mu_M = \mu_B$$

$$vs$$

$$H_1 : \text{at least one mean is not equal}$$

One-Way ANOVA

- Variation

  - Variation is the sum of the squares of the deviations between a value and the mean of the value

  - Sum of Squares is abbreviated by SS and often followed by a variable in parentheses such as $SS_{Treatment}$ or SS_{Error} so we know which sum of squares we're talking about

One-Way ANOVA

- Are all of the values identical?
  - No, so there is some variation in the data
  - This is called the total variation
  - Denoted SS_{Total} for the total Sum of Squares (variation)
  - Sum of Squares is another name for variation

One-Way ANOVA

- Are all of the sample means identical?
    - No, so there is some variation between the groups or treatments
    - This is called the between group or treatment variation
    - Denoted $SS_{Treatment}$ for Sum of Squares (variation) between the groups

One-Way ANOVA

- Are each of the values within each group identical?
  - No, there is some variation within the groups
  - This is called the within group variation
  - Sometimes called the error variation
  - Denoted SS_{Error} for Sum of Squares (variation) within the groups

One-Way ANOVA

- One-way ANOVA table:

Source of variation	SS	df	MS	F	p
Between group (Treatment)	$SS_{Treatment}$	k - 1	$MS_T = SS_T / k-1$	MS_T / MS_E	
Within group (Error)	SS_{Error}	N - k	$MS_E = SS_E / N-k$		
Total	SS_{Total}	N - 1			

One-Way ANOVA Calculation of SS:

- Grand Mean
  - The grand mean is the average of all the values when the factor is ignored
  - It is a weighted average of the individual sample means

$$\bar{\bar{x}} = \frac{\left(\sum_{i=1}^{k} x_i\right)^2}{N = \sum_{i=1}^{k} n_i} = \frac{\left(\sum_{i=1}^{k} n_i \overline{x_i}\right)^2}{N = \sum_{i=1}^{k} n_i}$$

$$\bar{\bar{x}} = \frac{\left(n_1 \overline{x_1} + n_2 \overline{x_2} + \ldots\ldots + n_k \overline{x_k}\right)^2}{n_1 + n_2 + \ldots\ldots + n_k}$$

One-Way ANOVA Calculation of SS:

- Total variation: SS_{Total}
  - the variation between observations and the grand mean

$$SS_{Total} = \sum_{i=1}^{k}\sum_{j=1}^{n}\left(x_{ij} - \overline{\overline{x}}\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}^2 - \overline{\overline{x}}$$

- Between Group Variation: $SS_{Treatment}$
  - the variation between each sample mean and the grand mean

$$SS_{Treatment} = \sum_{i=1}^{k} n_i (\overline{x_i} - \overline{\overline{x}})^2 = \sum_{i=1}^{k} n_i \left(\overline{x_i}\right)^2 - \overline{\overline{x}}$$

One-Way ANOVA Calculation of SS:

- Within Group Variation, SS_{Error}
  - the weighted total of the individual variations

$$SS_{Error} = \sum_{i=1}^{k} (n_i - 1) s_i^2$$

where s_i^2 is the variance within a group

OR

$$SS_{Error} = SS_{Total} - SS_{Treatment}$$

One-Way ANOVA - df

- The between group/treatment df is one less than the number of groups
  - If we have three groups/Treatments, so df for treatment is 2

- The within group df is the sum of the individual df's of each group
  - The sample sizes are 7, 9, and 8
  - df(W) = 6 + 8 + 7 = 21

- The total df is one less than the sample size
  - df(Total) = 24 – 1 = 23

EXAMPLE : One-Way ANOVA

- The statistics classroom is divided into three rows: front, middle, and back
- The Professor noticed that the further the students were from him, the more likely they were to miss class or sleep in the class
- He wanted to see if the students sit in front and near to him will did better on the exams

EXAMPLE : One-Way ANOVA

- A random sample of the students in each row was taken
- The score for those students on the exam was recorded
  - Front: 82, 83, 97, 93, 55, 67, 53
  - Middle: 83, 78, 68, 61, 77, 54, 69, 51, 63
  - Back: 38, 59, 55, 66, 45, 52, 52, 61

One-Way ANOVA

		$\sum x$	$\sum x^2$	n
FRONT	**82, 83, 97, 93, 55, 67, 53**	**530**	**41,994**	**7**
MIDDLE	**83, 78, 68, 61, 77, 54, 69, 51, 63**	**604**	**41,494**	**9**
BACK	**38, 59, 55, 66, 45, 52, 52, 61**	**428**	**23,460**	**8**
TOTAL		**1,562**	**106,948**	**24**

$N = n_1+n_2+n_3 = 7+9+8 = 24$ (number of observations)
$k = 3$ (number of groups

One-Way ANOVA

Calculation of GRAND MEAN:

$$\bar{\bar{x}} = \frac{\left(\sum_{i=1}^{k} x_i\right)^2}{N = \sum_{i=1}^{k} n_i} = \frac{(1562)^2}{24} = 101{,}660.2$$

One-Way ANOVA Calculation of SS:

- Total variation: SS_{Total}
  - the variation between observations and the grand mean

$$SS_{Total} = \sum_{i=1}^{k}\sum_{j=1}^{n}\left(x_{ij} - \overline{\overline{x}}\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}^2 - \overline{\overline{x}}$$

$$= 106{,}948 - 101{,}660.2 = 5{,}287.8$$

One-Way ANOVA Calculation of SS:

- Between Group Variation: $SS_{Treatment}$

  - the variation between each sample mean and the grand mean

$$SS_{Treatment} = \sum_{i=1}^{k} \frac{(\sum_{j=1}^{n_i} x_{ij})^2}{n_i} - \bar{\bar{x}}$$

$$= \frac{(530)^2}{7} + \frac{(604)^2}{9} + \frac{(428)^2}{8} - 101,660.2$$

$$= 40,128.57 + 40,535.11 + 22,898 - 101,660.2$$

$$= 1,901.48$$

One-Way ANOVA Calculation of SS:

- Within Group Variation, SS_{Error}
  - the weighted total of the individual variations

$$SS_{Error} = SS_{Total} - SS_{Treatment}$$

$$= 5{,}287.8 \; - 1{,}901.48$$

$$= 3{,}386.32$$

One-Way ANOVA

- Filling in SS and the degrees of freedom:

Source	SS	df	MS	F	p
Between	1901.48	2	950.74	5.90	0.009
Within	3386.32	21	161.25		
Total	5287.80	23			

$F_{Critical} = f_{0.05,2,21} = 3.47$

From Excel:

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Front:	7	530	75.71429	310.9048
Middle:	9	604	67.11111	119.8611
Back:	8	428	53.5	80.28571

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1901.516	2	950.7579	5.896056	0.009284	3.46680011
Within Groups	3386.317	21	161.2532			
Total	5287.833	23				

One-Way ANOVA

- Since $F > F_{Critical} = 3.47$, then we reject the null hypothesis that the means score of the three rows in class were the same. So we can conclude that at least one row has a different mean.

- We can also use the p-value to make a decision. From ANOVA table, P-value = 0.009, which is less than the significance level of 0.05, so we reject the null hypothesis.

One-Way ANOVA

- There is enough evidence to support the claim that there is a difference in the mean scores of the front, middle, and back rows in class.

- The ANOVA doesn’t tell which row is different.

One-Way ANOVA

Exercise:

UTP wishes to compare four programs for training staff to perform a certain task. Twenty new staffs are randomly selected to the training programs, with 5 in each program. At the end of the training, a test is conducted to see how quickly trainees can perform the task. The number of times the task is performed per minute is recorded for each trainee and the results as in table below:

Program	Time in minutes
Program 1	9, 12, 14, 11, 13
Program 2	12, 14, 11, 13, 11
Program 3	9, 8, 7, 8, 11
Program 4	10, 6, 9, 9, 10

i. Construct the ANOVA table.

ii. Are there any differences between the four programs?

Solution From Excel:

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Program 1	5	59	11.8	3.7
Program 2	5	61	12.2	1.7
Program 3	5	43	8.6	2.3
Program 4	5	44	8.8	2.7

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	54.95	3	18.31667	7.044872	0.003113	3.238871522
Within Groups	41.6	16	2.6			
Total	96.55	19				