One-way ANOVA / Completely randomized design

Animal Science Lab

ANOVA is used to test null hypotheses that three or more treatments are equally effective. The experimental design requires that the subjects be allocated at random to each treatment group. Assumptions in ANOVA: the observed data constitutes independent random samples, the population from which the samples came are normally distributed and each of the populations has the same variance.

1. Concept of ANOVA

Subjects are allocated at random to each treatment group.

Treatments

A	**B**	**C**
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
.	.	.
.	.	.
r	r	r

Variability Within Treatments is referred as the Error.

Mean of each treatment: $\bar{Y}_A$ $\bar{Y}_B$ $\bar{Y}_C$

$$\underbrace{\bar{Y}_A \quad \bar{Y}_B \quad \bar{Y}_C}_{}$$

Variability Between Treatment (k=3)

H_o: $\mu_A = \mu_B = \mu_C$
H_1: not all equal

- If H_o is **true** then, $\bar{Y}_A = \bar{Y}_B = \bar{Y}_C$ → then the variability between treatment means should be small because they are all the same or very similar.

- If H_o is **not** true, the means are not all equal → meaning that at least one mean is different, then the variability between treatment means should be large.

So to complete the test of hypothesis, the question is how to test if the variability between treatment means is small or large. Small or large relative to what? To accomplish this we move to the next step where in ANOVA the total variation of a data set is partition in to two or more components according to a **model**.

2. Concept of Variance and Model

a) Simple model – one group of observations

	Observations	$(Y_i - \bar{Y})$	$(Y_i - \bar{Y})^2$
	$Y_1 =$ 6	1	1
	$Y_2 =$ 9	4	16
	$Y_3 =$ 2	-3	9
	$Y_4 =$ 3	-2	4
Sum	20	0	30
Mean	5		

Model for each observation or data point

Obs		= mean + deviation from mean (error)		
Y_1 6	=	5	+	(6 – 5)
Y_1 6	=	5	+	1
Y_2 9	=	5	+	4
Y_3 2	=	5	+	-3
Y_4 3	=	5	+	-2

b) Three treatments – Example of ANOVA model

	Columns (i)			
Rows (j)	**1**	**2**	**3**	total
1	6	10	2	
2	9	4	6	
3	3	9	4	
4		12		
5		8		
6		11		
Sum	18	54	12	84
Mean	6	9	4	7
Obs	3	6	3	12

Variables are arranged in a two-way table such that i represent a column and j represents a row. For example, the variable in column 2 and row 4 is denoted as $Y_{24} = 12$

$\mathbf{Y_{i.}}$ The dot subscript represents the total sum of the variables in a column. For example, $Y_{1.} = 18$, $Y_{2.} = 54$, $Y_{3.} = 12$

$\mathbf{Y_{..}}$ The two dot subscript represents the total sum of all the variables in the data set, such as $\mathbf{Y_{..}} = 84$

Model:

Individual Obs = overall mean + deviation of the treatment mean from the overall mean + deviation of an individual obs from its treatment mean

$$Y_{ij} = \bar{Y}_{..} + (\bar{Y}_{i.} - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_{i.})$$

$$\text{Ex: } 12 = 7 + (9 - 7) + (12 - 9)$$

Moving the overall mean to the left side of the equation:

$$(Y_{ij} - \bar{Y}_{..}) = (\bar{Y}_{i.} - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_{i.})$$

$$\text{Ex: } (12 - 7) = (9 - 7) + (12 - 9)$$

One-way ANOVA / Completely randomized design

Animal Science Lab

5 = 2 + 3

3. Transforming the model into a Sum of Squares (SS):

$$\sum\sum(Y_{ij} - \bar{Y}_{..})^2 = k\sum(\bar{Y}_{i.} - \bar{Y}_{..})^2 + r\sum(Y_{ij} - \bar{Y}_{i.})^2$$

Total SS (TSS) — **SS treatments (SST)** — **SS error (SSE)**

Computational formula: (k = # treatments, r= # observations)

$$\left(\sum\sum Y_{ij}^2 - \frac{Y_{..}^2}{kr}\right) = \left(\frac{\sum Y_{i.}^2}{r} - \frac{Y_{..}^2}{kr}\right) + \left(\sum\sum Y_{ij}^2 - \frac{\sum Y_{i.}^2}{r}\right)$$

(TSS) — **(SST)** — **(SSE)**

4. Calculations:

1) Correction term (C):

$$C = \frac{Y_{..}^2}{kr} = \frac{(84)^2}{12} = 588$$

2) Each individual observation squared:

$$\sum\sum Y_{ij}^2 = 6^2 + 9^2 + 3^2 + 10^2 + 4^2 + 9^2 + 12^2 + 8^2 + 11^2 + 2^2 + 6^2 + 4^2 = 708$$

3) Sum of obs in each treatment squared, divided by the number of obs in each the treatment:

$$\frac{\sum Y_{i.}^2}{r} = \frac{18^2}{3} + \frac{54^2}{6} + \frac{12^2}{3} = 642$$

4) Calculate SS:

Total SS: $TSS = \sum\sum Y_{ij}^2 - C = 708 - 588 = 120$

SS Treatments: $SST = \frac{\sum Y_{i.}^2}{r} - C = 642 - 588 = 54$

SS error: $TSS - SST = 120 - 54 = 66$

5. Summarize Analysis in ANOVA Table:

Source	df	SS	MS	F
Between Treatments	k-1	SST	SST/k-1	MST/MSE
Within Treatments (error)	k(r-1)	SSE	SSE/k(r-1)	
Total	kr – 1	TSS		

Source	df	SS	MS	F
Between Treatments	3-1	54	54/2=27	27/7.33= 3.68
Within Treatments (error)	9	66	66/9=7.33	
Total	12–1	120		

In a situation when there are equal number of observations in each treatment group then the df for the error can be calculated as df = k(r-1). However, if the number of observations in each treatment varies, like in the current example, then df = $(r_1\text{-}1)+(r_2\text{-}1)+(r_3\text{-}1)$= (3-1)+(6-1)+(3-1) = 9.

Note that when the SS are divided by the df they become variances and we call them Mean Squares. For example, SST/k-1 = 54/3-1 = 27, this is the Mean Square Treatments (MST).

To compare the equality of the MST and the MSE we use an F test with the df associated with each one of the Mean Squares.

6. Decision and Conclusion

H_o: $\mu_A = \mu_B = \mu_C$
H_1: not all equal

Going back to the concept of ANOVA shown in #1, we test the equality of variances by comparing the equality of the variance between treatments (MST) with the variance within treatments or error (MSE) using a variance ratio test (F test).

H_o: F = MST/MSE = 1
H_1: F >1

F = MST/MSE = 27 / 7.33 = 3.68 (calculated F value)
From Table A-7 we look up the critical F value with 3 and 9 df, for and alpha=0.05
$F_{(3,9)}$ = 4.26 (critical F value)

Since, (calculated F value) 3.68 < 4.26 (F critical). The calculated F value is in the acceptance region, meaning that the variances are equal.

Since the variances are equal we conclude that the treatments had no effect on the mean of the three treatments and **$\mu_A = \mu_B = \mu_C$.**