Statistical Inference
ANOVA

One-Way ANOVA: Analysis of variance (ANOVA) provides methods for comparing several population means, that is, the means of a single variable for several populations. We present the simplest kind of ANOVA, **one-way analysis of variance.** This type of ANOVA is called *one-way* analysis of variance because it compares the means of a variable for populations that result from a classification by *one* other variable, called the **factor.** The possible values of the factor are referred to as the **levels** of the factor.

Assumptions (Conditions) for One-Way ANOVA

1. Simple random samples: The samples taken from the populations under consideration are simple random samples.

2. Independent samples: The samples taken from the populations under consideration are independent of one another.

3. Normal populations: For each population, the variable under consideration is normally distributed.

4. Equal standard deviations: The standard deviations of the variable under consideration are the same for all the populations.

Definition: Mean Squares and F-Statistic in One-Way ANOVA

Treatment mean square, MSTR: The variation among the sample means:
MSTR = SSTR/(k − 1),
where SSTR is the treatment sum of squares and k is the number of populations under consideration.

Error mean square, MSE: The variation within the samples:
MSE = SSE/(n − k),
where SSE is the error sum of squares and n is the total number of observations.

F-statistic, F: The ratio of the variation among the sample means to the variation within the samples: **F = MSTR/MSE**.

One-Way ANOVA Identity:
The total sum of squares equals the treatment sum of squares plus the error sum of squares: **SST = SSTR + SSE**.

ANOVA Table Format For ONE-WAY Analysis of Variance

Source	df	SS	MS = SS/df	F-statistic
Treatment	$k-1$	$SSTR$	$MSTR = \frac{SSTR}{k-1}$	$F = \frac{MSTR}{MSE}$
Error	$n-k$	SSE	$MSE = \frac{SSE}{n-k}$	
Total	$n-1$	SST		

Sums of Squares in One-Way ANOVA

For a one-way ANOVA of k population means, the defining and computing formulas for the three sums of squares are as follows.

Sum of squares	Defining formula	Computing formula
Total, SST	$\Sigma(x_i - \bar{x})^2$	$\Sigma x_i^2 - (\Sigma x_i)^2/n$
Treatment, $SSTR$	$\Sigma n_j(\bar{x}_j - \bar{x})^2$	$\Sigma(T_j^2/n_j) - (\Sigma x_i)^2/n$
Error, SSE	$\Sigma(n_j - 1)s_j^2$	$SST - SSTR$

In this table, we used the notation

n = total number of observations
$\bar{x}$ = mean of all n observations;

and, for $j = 1, 2, \ldots, k$,

n_j = size of sample from Population j
$\bar{x}_j$ = mean of sample from Population j
s_j^2 = variance of sample from Population j
T_j = sum of sample data from Population j.

Note that summations involving subscript is are over all n observations; those involving subscript js are over the k populations.

PROCEDURE: One-Way ANOVA Test

Purpose To perform a hypothesis test to compare k population means, $\mu 1, \mu 2, \ldots, \mu_k$

Assumptions
1. Simple random samples
2. Independent samples

3. Normal populations
4. Equal population standard deviations

Step 1: The null and alternative hypotheses are, respectively,

H0: $\mu 1 = \mu 2 = \cdots = \mu_k$
Ha: Not all the means are equal.

Step 2 Decide on the significance level, α.

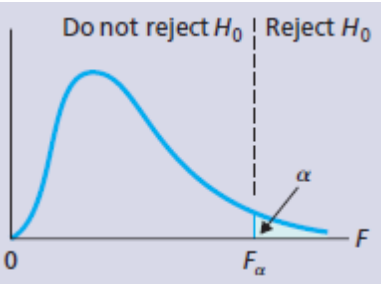
Step 3 Compute the value of the test statistic

F = MSTR / MSE

and denote that value F0. To do so, construct a one-way ANOVA table:

Source	df	*SS*	$MS = SS/df$	F-statistic
Treatment	$k-1$	*SSTR*	$MSTR = \frac{SSTR}{k-1}$	$F = \frac{MSTR}{MSE}$
Error	$n-k$	*SSE*	$MSE = \frac{SSE}{n-k}$	
Total	$n-1$	*SST*		

Step 4 The critical value is Fα with df = (k−1, n−k). Use F- Table to find the critical value.



Step 5 If the value of the test statistic falls in the rejection region, reject H0; otherwise, do not reject H0.

Step 6 Interpret the results of the hypothesis test.

Example:

Energy Consumption Recall that independent simple random samples of households in the four U.S. regions yielded the data on last year's energy consumptions shown in Table below. At the 5% significance level, do the data provide sufficient evidence to conclude that a difference exists in last year's mean energy consumption by households among the four U.S. regions?

Northeast	Midwest	South	West
15	17	11	10
10	12	7	12
13	18	9	8
14	13	13	7
13	15		9
	12		

Step 1 State the null and alternative hypotheses.
Let μ1, μ2, μ3, and μ4 denote last year's mean energy consumptions for households in the Northeast, Midwest, South, and West, respectively. Then the null and alternative hypotheses are, respectively,

H0: μ1 = μ2 = μ3 = μ4 (mean energy consumptions are equal)
Ha: Not all the means are equal.

Step 2 Decide on the significance level, α.
We are to perform the test at the 5% significance level; so, α = 0.05.

Step 3 Compute the value of the test statistic

F = MSTR / MSE

We have,

$$k = 4$$

$$n_1 = 5 \quad n_2 = 6 \quad n_3 = 4 \quad n_4 = 5$$

$$T_1 = 65 \quad T_2 = 87 \quad T_3 = 40 \quad T_4 = 46$$

and

$$n = \Sigma n_j = 5 + 6 + 4 + 5 = 20$$

$$\Sigma x_i = \Sigma T_j = 65 + 87 + 40 + 46 = 238.$$

$$\Sigma x_i^2 = (15)^2 + (10)^2 + (13)^2 + \cdots + (7)^2 + (9)^2 = 3012.$$

Consequently,

$$SST = \Sigma x_i^2 - (\Sigma x_i)^2/n = 3012 - (238)^2/20 = 3012 - 2832.2 = 179.8,$$

$$SSTR = \Sigma(T_j^2/n_j) - (\Sigma x_i)^2/n$$

$$= (65)^2/5 + (87)^2/6 + (40)^2/4 + (46)^2/5 - (238)^2/20$$

$$= 2929.7 - 2832.2 = 97.5,$$

$$SSE = SST - SSTR = 179.8 - 97.5 = 82.3.$$

F = MSTR / MSE
= 32.5 / 5.144
= 6.32

One-way ANOVA table for the energy consumption data

Source	df	*SS*	*MS* = *SS*/df	*F*-statistic
Treatment	3	97.5	32.500	6.32
Error	16	82.3	5.144	
Total	19	179.8		

Step 4 The critical value is Fα with df = (k − 1, n − k)

From Step 2, α = 0.05. Also, F-Table shows that four populations are under consideration, or k = 4, and that the number of observations total 20, or n = 20. Hence, df = (k − 1, n − k) = (4 − 1, 20 − 4) = (3, 16). From Table VIII, the critical value is F0.05 = 3.24.

Step 5 If the value of the test statistic falls in the rejection region, reject H0; otherwise, do not reject H0.

From Step 3, the value of the test statistic is F = 6.32, which falls in the rejection region. Thus, we reject H0. The test results are statistically significant at the 5% level.

Step 6
Interpretation At the 5% significance level, the data provide sufficient evidence to conclude that a difference exists in last year's mean energy consumption by households among the four U.S. regions. Evidently, at least two of the regions have different mean energy consumptions.

Exercise:

1- Copepods are tiny crustaceans tha tare an essential link in the estuarine food web. Marine scientists G. Weiss et al. at the Chesapeake Biological Laboratory in Maryland designed an experiment to determine whether dietary lipid (fat) content is important in the population growth of a Chesapeake Bay copepod. Their findings were published as the paper "Development and Lipid Composition of the arpacticoid Copepod Nitocra Spinipes Reared on Different Diets" (Marine Ecology Progress Series, Vol. 132, pp. 57–61). Independent random samples of copepods were placed in containers containing lipid-rich diatoms, bacteria, or leafy macroalgae. There were 12 containers total with four replicates per diet. Five gravid

(egg-bearing) females were placed in each container. After 14 days, the number of copepods in each container were as follows.

Diatoms	Bacteria	Macroalgae
426	303	277
467	301	324
438	293	302
497	328	272

At the 5% significance level, do the data provide sufficient evidence to conclude that a difference exists in mean number of copepods among the three different diets? (Note: T1 = 1828, T2 = 1225, T3 = 1175, and $\sum x^2_i$ = 1,561,154.).

2- Suppose the National Transportation Safety Board (NTSB) wants to examine the safety of compact cars, midsize cars, and full-size cars. It collects a sample of three for each of the treatments (cars types). Using the hypothetical data provided below, test whether the mean pressure applied to the driver's head during a crash test is equal for each types of car. Use α = 5%

	Compact cars	Midsize cars	Full-size cars
	643	469	484
	655	427	456
	702	525	402
$\overline{X}$	666.67	473.67	447.33
S	31.18	49.17	41.68