## **ONE FACTOR ANOVA**

By Alessandro Maurizio Idrovo Gavilanes

## Introduction

One-factor ANOVA is a statistical method for testing the null hypothesis that three or more population means are equal, against the alternative hypothesis that at least one of the populations differs from the others in terms of its expected value. This contrast is fundamental in the analysis of experimental results, in which it is of interest to compare the results of K 'treatments' or 'factors' with respect to the dependent variable or variable of interest.

## Analysis of variance assumptions

ANOVA requires the following assumptions to be met:

- The populations follow a normal distribution.
- The K samples on which the treatments are applied are independent.
- The populations all have equal variance (homoscedasticity).

## **One-factor ANOVA**

When you have k different levels corresponding to the k population means, which are also the treatments, an ANOVA attempts to answer the question: Are the k population means all the same, or is at least one mean different from the others?

Then if there are k independent random samples of size  $n_i$  (i=1, 2, ....., k), coming from k different populations (concerning k treatments, k groups, k production methods, etc.) they can be ordered as follows:

The sample means of the groups are denoted as  $\bar{y}1$ ,  $\bar{y}2$ , ...,  $\bar{y}K$  and  $\bar{y}$  is the mean of all  $N = \sum_{i=1}^{k} n_i$  observations.

## - Variability

The responses generated in an experimental situation always show some variability. In an analysis of variance, the total variability of the responses is divided into two parts: one due to **between-treatment variability** (measured by the sum of the squares of the treatments,  $SC_{trat}$ ) and another part due to **within-treatment variability** or error (measured by the sum of the squares of the error,  $SC_{error}$ ):

Total variability= between-treatment variability + within-treatment variability

$$SC_{total} = SC_{trat} + SC_{error}$$

- Between-treatment variability

$$SC_{trat} = \sum_{i=1}^{k} n_i (\bar{y}_i - \bar{y})^2$$

It measures the discrepancy between the groups and the global mean, so that if there are no differences between them (the null hypothesis is not rejected) we will obtain small variabilities. If, on the other hand, the null hypothesis is rejected, the variability between groups can be expected to be large.

- Within-treatment variability or error variability

$$SC_{error} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$$

It measures the intrinsic variability of the observations, i.e., if the experiment is well designed and no variation factors other than the one studied are included, it should be purely random error produced as a result of the biological variability of the experimental material.

# ANOVA table

The analysis of variance contrast is based on the comparison of the variability between and the variability within. Thus, the null hypothesis is rejected whenever the "**between**" variability is large with respect to the "**within**" variability.

The complete information is summarized in the following table, known as the ANOVA table, which compiles all the information necessary to perform the corresponding test.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F Value
Between Groups	$SSB = \Sigma n_j (\overline{X_j} - \overline{X})^2$	df <sub>1</sub> = k - 1	MSB = SSB / (k - 1)	f = MSB / MSE
Error	$SSE = \Sigma\Sigma(X - \overline{X_j})^2$	$df_2 = N - k$	MSE = SSE / (N - k)	
Total	SST = SSB + SSE	df <sub>3</sub> = N - 1		

# Example

A study is done on the effectiveness of three brands of spray to kill flies. To do this, each product is applied to a group of 100 flies, and the number of dead flies is counted, expressed in percentages. Six replicates are made and the results obtained are shown below (Gutiérrez Pulido & De la Vara Salazar, 2008).

Atomizer brand	Number of replicas					
	1	2	3	4	5	6
1	72	65	67	75	62	73
2	55	59	68	70	53	50
3	64	74	61	58	51	69

# Assumptions verification (Minitab software)

# Verification of the normality assumption

1. Set the hypotheses.

Ho= The residues follow a normal distribution

H1= The residues do not follow a normal distribution

## 2. Set the significance level.

α= 0.05

3. Set the test statistic.

Anderson- Darling.

4. Set the decision.

The null hypothesis is rejected if p-value <  $\alpha$ .

- 5. Resolution
- 1) The data is placed in Minitab, entering them in two columns, one corresponding to the treatments and the other corresponding to the responses.



2) Select the ANOVA option to fit general linear model.



3) Select the responses and treatments in the open window in addition to selecting the waste in storage option.

Marca de spray	Respuestas:	Modelo lineal general: Almacenamiento
. Kespuesia	Perpuesta	Agustes     Cogficientes     Activity     Activity
	Aleat <u>o</u> rio/anidar <u>M</u> odelo Opcio <u>n</u> es Co <u>d</u> ific	ar
Seleccionar	Pago a paso <u>G</u> ráficas <u>R</u> esultados Almacena	an Ayuda Aceptar Cancelar

4) Once the residuals have been calculated, we go to the graph section and select probability graph.

Minitab - Sin titulo																- 0	
Archivo Editar Datos Calc Estadist	icas Gráfica Vista Ayuda Asistente Modulo de A	nalitica predictiva	Herramien	as adicionales													
第四部 2 ほうくに	3 a Creador de grancas.	2															
Navegador *	A Gráfica de dispersión																
Modelo lineal general: Respuesta	Gráfica de dispersión dividida en clases      Gráfica de motiva      Correlograma      Gráfica de burbujes      Gráfica de burbujes      Gráfica marginal	vs. Marca	de spra	у													•
	Histograma     Gráfica de puntos     Tallo y hoja																
	Corrent de probabilidad     Corrent de probabilidad     Corrent de probabilidad     Corrent de probabilidad     fordinc ade attent     fordinc ade valores individuales.     Sofficia de initiaa.     Sofficia de tenteret	a distribution partie	uler y														
	+	C5	C6	67	CB	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	
	<ul> <li>Mapa de calor</li> <li>Gráfica circular</li> </ul>																
	1 2 M Gráfica de serie de tiempo 3 M Area Gráfico																
	4 Gráfica de contorno 5 23 Gráfica de dispersión 3D 6 🗃 Gráfica de superfície 3D																
	7 2 55 4.1667																

5) Select residuals as the response variable and a normal distribution.

1	Gráfica	de probabilidad:	Individual		Gráfica de probabilidad: Distribución	×	]
ju 18 13	C1 M C2 Ri C3 Ri C3 Ri S	tarca de spray espuesta ESI Seleccionar ruda	Variables de gráficas: RESI Distribución Multiples gráficas	Egcolo Qpciones de datos. Aceptar	Distribución Presentación de datos Distribución Vormal Parámetros hjstáricas Modía Desv.Est. Un parámetro será estimado a partir de los datos si la celda se deja blanco. Puede dejar ambas celdas en blanco, ingresar valores para celdas o ingresar un valor para Desv.Est. solomente. Si ingresa un v para Desv.Est., éste debe ser mayor que cero.	en ambas alor	
					AyudaCeptarCar	icelar	

6) Results according to Minitab



Since the p-value is greater than the significance level, the null hypothesis is not rejected, therefore, we cannot reject with a 95% confidence level that the residuals follow a normal distribution (the assumption of normality is met).

# Verification of the homoscedasticity assumption (Minitab software)

# 1. Set the hypotheses.

Ho= The variances are equal.

H1= At least one of the variances is different.

## 2. Set the significance level.

α= 0.05

## 3. Set the test statistic.

Bartlett.

## 4. Set the decision.

The null hypothesis is rejected if p-value <  $\alpha$ .

# 5. Resolution

1) Select ANOVA test of equality of variances.

Estadísticas básicas	I III 12 4 2					
Regresión	• <u> </u>					
ANOVA	Un solo factor					
DOE	Análisis de medias					
Gráficas de control	AB ANOVA balanceado					
Herramientas de calidad	<ul> <li>Modelo lineal general</li> </ul>	•				
Confiabilidad/supervivencia	<ul> <li>Modelo de efectos mixtos</li> </ul>	Desv.Est. 6.841	Desv.Est. 6.841			
Analítica predictiva	ANOVA totalmente anidado	N 18 AD 0.202				
Análisis multivariado	MANOVA general	Valor p 0.857				
Series de tiempo	•					
Tablas	<ul> <li>O<sup>*</sup> Prueba de igualdad de varianzas</li> </ul>	•				
No paramétricos	Gráfica de intervalos	Prueba de igualdad de varianzas				
Pruebas de equivalencia	Gráfica de efectos principales	- Determinar si las varianzas o las desviaciones estándar				
Potencia y tamaño de la muestra	Gráfica de interacciones	dos o más grupos difieren.				

2) In the open window we select the response and corresponding treatments, in addition to indicating that the test is based on a normal distribution.

Eactores:       Marca de sprey'         Marca de sprey'       Image: Cancelar and the sprey'         United de sprey'       Image: Cancelar and the sprey'         Opcioges       Granueses         Ayuda       Aceptar         Cancelar       12         C13       C14         C15       C16	Prueba de igualdad de	Poueba de Rantiert           e varianzas           Los datos de respuesta están en una columna para todos los niveles de factores            Bespuesta:         Prueba de igualdad de varianzas: opciones
Opciones     Grainwaw       Seleccionar     Resultados       Almacenamiento     212       C13     C14       C15	-	Factories:     Nivel de gonfianza:       'Marca de spray'     Image: Space and
Seleccionar Almacenamiento 12 C13 C14 C15 C16		Opciones
Avuda Aceptar Cancelar	) Avuda	Resultados Almacenamiento 12 C13 C14 C15 C16

3) Results according to Minitab.



Since the p-value is greater than the significance level, the null hypothesis is not rejected, therefore, we cannot reject with a 95% confidence level that the variances are equal **(the assumption of homoscedasticity is met)**.

# Verification of the independence assumption

Since the populations come from different treatments, we assume that the populations are independent (the assumption of independence is met).

#### **ANOVA resolution (Minitab)**

Once the assumptions have been verified, the ANOVA table is created.

1) Select one-factor ANOVA.



2) Select the treatments and response in the corresponding boxes, in addition to checking the box to assume equal variances.

	Los datos de	e respuesta están en u	na columna para todo:	s los niveles	de factores						
	<u>R</u> espuesta:	Respuesta		Análisis	de varianza de	e un solo f	actor: Opci	ones			×
а 1.	Eactor:	'Marca de spray'		✓ Agum Nivel de g Tipo de ir Ayud	ir varianzas igua confianza: ntervalo de confi a	les 95 anza: Bilal	eral	(para la	tabla de medi <u>A</u> cep	as y la gráfica tar C	de intervale Cancelar
		Ogciones	Comparaciones		<u>G</u> ráficas	12	C13	C14	C15	C16	C17
	_		1.00	. 1							

3) Results according to Minitab.

# Método

Hipótesis nula	Todas las medias son iguales
Hipótesis alterna	No todas las medias son iguales
Nivel de significancia	α = 0.05

Se presupuso igualdad de varianzas para el análisis.

# Información del factor

Factor	Niveles	Valores	
Marca de spray	3	1; 2; 3	

# Análisis de Varianza

Fuente	GL	SC Ajust.	MC Ajust.	Valor F	Valor p
Marca de spray	2	296.3	148.17	2.79	0.093
Error	15	795.7	53.04		
Total	17	1092.0			

Since the p-value is greater than the significance level, the null hypothesis is not rejected, therefore, we cannot reject with a 95% confidence level that the population means are equal.

Г

٦

# ANOVA resolution (MegaStat)

1) Open MegaStat in Excel and select analysis of variance one-factor.

2) Select the data of treatments and number of replicas in the input range, in addition to checking the box of plot data.

Hoja1!\$B\$3:\$G\$5	_ Input range	<u>0</u> K
		<u>C</u> lear
- Post-Hoc Analysis	✓ Plot Data	Cancel
● <u>W</u> hen p < .05		
⊂ Ne <u>v</u> er	Display partitioning	<u>H</u> elp
© <u>A</u> lways		

3) Results according to MegaStat.

One factor	ANOVA				
	Mean	n	Std. Dev		
	69,0	6	5,10	A	
	59,2	6	8,18	В	
	62,8	6	8,13	С	
	63,7	18	8,01	Total	
ANOVA table					
Source	SS	df	MS	F	p-value
Treatment	296,33	2	148,167	2,79	,0931
Error	795,67	15	53,044		
Total	1.092,00	17			

Since the p-value is greater than the significance level, the null hypothesis is not rejected, therefore, we cannot reject with a 95% confidence level that the population means are equal.

## Conclusion

One-factor ANOVA is a very useful statistical method for testing the null hypothesis that three or more population means are equal, against the alternative hypothesis that at least one of the populations differs from the others in terms of its expected value.

## **Bibliography and recommended links**

Gutiérrez Pulido, H., & De la Vara Salazar, R. (2008). Análisis y Diseño de experimentos. Ciudad de México: McGraw-Hill Interamericana.

Milton, J. (2014). Estadística para Biología y Ciencias de la Salud. Ciudad de México: McGraw-Hill Interamericana.

http://www.ub.edu/aplica\_infor/spss/cap4-7.htm

https://www.youtube.com/watch?v=efSGqgvZCtw&ab\_channel=LEARN%26APPLY%3ALeanand SixSigma