

ANÁLISIS DE LA VARIANZA (ANOVA) CON STATS EN R

**Brenda V. Canizo
2017**

Stats

> library(help = "stats")

SSlogis	Self-Starting Nls Logistic Model
SSminimax	Self-Starting Nls Michaelis-Menten Model
SSwblull	Self-Starting Nls Weibull Growth Curve Model
SignRank	Distribution of the Wilcoxon Signed Rank Statistic
StructTS	Fit Structural Time Series
TDist	The Student's t Distribution
Tukey	The Studentized Range Distribution
TukeyPRD	Compute Tukey Honest Significant Differences
Uniform	The Uniform Distribution
Weibull	The Weibull Distribution
Wilcoxon	Distribution of the Wilcoxon Rank Sum Statistic
acf	Auto- and Cross- Covariance and -Correlation Function Evaluation
acf2AR	Compute an AR Process Exactly Fitting an ACF
add1	Add or Drop All Possible Single Terms to a Model
addmargins	Make Arbitrary Marginal on Multidimensional Tables or Arrays
aggregate	Compute Summary Statistics of Data Subsets
alias	Find Aliases (Dependencies) in a Model
anova	Anova Tables
anova.glm	Analysis of Deviance for Generalized Linear Model Fits
anova.lm	ANOVA for Linear Model Fits
anova.mlm	Comparisons between Multivariate Linear Models
ansari.test	Ansari-Bradley Test
avf	Fit an Analysis of Variance Model
approxfun	Interpolation Functions
ar	Fit Autoregressive Models to Time Series
ar.ols	Fit Autoregressive Models to Time Series by OLS
arima	ARIMA Modelling of Time Series
arima.sim	Simulate from an ARIMA Model

Conceptos básicos de ANOVA

ANOVA de un factor: analiza y compara el comportamiento de una variable continua de interés (dependiente) “Y” en distintos niveles (poblaciones o grupos o tratamientos) de un factor (variable explicativa).

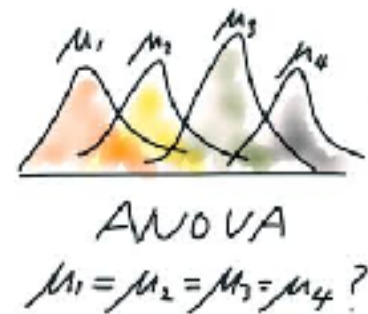
Partimos de unidades experimentales homogéneas y queremos estudiar la influencia de un solo factor midiendo la variable respuesta.

Permite contrastar la hipótesis nula de que la media de más de dos poblaciones son iguales, frente a la hipótesis alternativa de que por lo menos una es diferente.

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

H_1 : al menos una μ_i es diferente

Tiene supuestos o requisitos



Ejemplo. Un Ingeniero Agrónomo que realiza estudios en producción de trigo, extrae los valores de producción (en quintales/hectárea) en 4 zonas de la provincia. Necesita saber si hay diferencias significativas en la producción de estas 4 zonas.

Factor: zona, con 4 niveles: 1, 2, 3 y 4.

Variable respuesta: producción (quintales por hectárea)

Los datos obtenidos fueron:



	produccion	Zona
1	27	Zona 1
2	28	Zona 1
3	31	Zona 1
4	32	Zona 1
5	33	Zona 1
6	31	Zona 2
7	34	Zona 2
8	35	Zona 2
9	36	Zona 2
10	39	Zona 2
11	40	Zona 2
12	30	Zona 3
13	38	Zona 3
14	42	Zona 3
15	43	Zona 3
16	16	Zona 4
17	20	Zona 4
18	21	Zona 4
19	26	Zona 4
20	27	Zona 4
21	29	Zona 4
22	35	Zona 4

- **Cargo librerías:**

```
library(stats)    # ANOVA
library(agricolae) # Tukey
library(gplots)   # Gráfico de Medias y
Desvíos
```

- **Aislamiento de las variables a utilizar:**

```
NUM_COL_VR <- SEMANA02_BASE01_PRODUCCION[,1]
NUM_COL_FACTOR <- SEMANA02_BASE01_PRODUCCION [,2]
```

Creamos un vector con estos dos objetos.

El nuevo vector contiene el Numero de las Columnas Seleccionadas que ingresarán en el análisis ANOVA.

```
NUM_COL_SEL <- c(NUM_COL_VR , NUM_COL_FACTOR)
```

```
$produccion
[1] 27 28 31 32 33 31 34 35 36 39 40 30 38 42 43 16 20 21 26 27 29 35
```

```
$Zona
[1] "Zona 1" "Zona 1" "Zona 1" "Zona 1" "Zona 1" "Zona 2" "Zona 2" "Zona 2" "Zona 2"
[10] "Zona 2" "Zona 2" "Zona 3" "Zona 3" "Zona 3" "Zona 3" "Zona 4" "Zona 4" "Zona 4"
[19] "Zona 4" "Zona 4" "Zona 4" "Zona 4"
```

- **Generación de resultados**

Crearemos un objeto llamado 'LM' que contendrá todos los cálculos realizados por R para un Modelo Lineal General.

```
LM <- lm(VR ~ FACTOR)
```

Le pedimos a R que tome el Modelo Lineal que recién calculó, y los muestre el formato ANOVA

```
ANOVA_COMPLETO <- aov(LM)
```

Pediremos a R la tabla típica de ANOVA y la guardaremos en un nuevo objeto que llamamos 'TABLA_ANOVA'

```
TABLA_ANOVA <- summary(ANOVA_COMPLETO)[[1]]
```

Solicitamos a R el test de comparación múltiple de **TUKEY** (que solo se lee en el caso de que ANOVA rechace la H_0).

```
TUKEY_COMPLETO <- HSD.test(LM,"FACTOR")
```

Pedimos a R la tabla típica de TUKEY y la guardamos en un nuevo objeto que llamaremos 'TABLA_TUKEY'

```
TABLA_TUKEY <- TUKEY_COMPLETO$groups
```

Toda interpretación que hagamos solo es válida si los errores son estadísticamente NORMALES y HOMOGENEOS.

El objeto 'ANOVA_COMPLETO' tiene dentro los RESIDUOS del ANOVA realizado.

Se crea un nuevo objeto llamado "RESIDUOS"

```
RESIDUOS <- ANOVA_COMPLETO$residuals
```

NORMALIDAD DE LOS ERRORES (Test de Shapiro-Wilks)

```
SHAPIRO <- shapiro.test(RESIDUOS)
```

HOMOGENEIDAD DE VARIANZAS DE LOS ERRORES (Test de Bartlett)

```
BARTLETT <- bartlett.test(RESIDUOS, FACTOR)
```

1) ANOVA

```
> ANOVA_COMPLETO <- aov(LM)
```

```
> TABLA_ANOVA <- summary(ANOVA_COMPLETO)[[1]]
```

```
> TABLA_ANOVA
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
FACTOR	3	612.26	204.087	8.5583	0.0009585 ***
Residuals	18	429.24	23.847		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

2) TUKEY

```
> TUKEY_COMPLETO <- HSD.test(LM,"FACTOR")
```

```
> TABLA_TUKEY <- TUKEY_COMPLETO$groups
```

```
> TABLA_TUKEY
```

	trt	means	M
--	-----	-------	---

1	Zona 3	38.25000	a
---	--------	----------	---

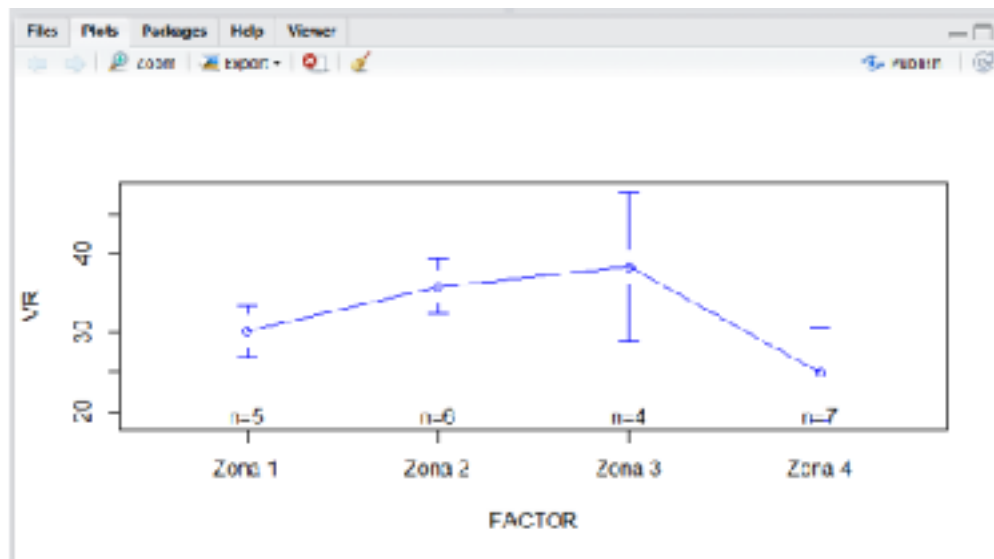
2	Zona 2	35.83333	a
---	--------	----------	---

3	Zona 1	30.20000	ab
---	--------	----------	----

4	Zona 4	24.85714	b
---	--------	----------	---

3) GRAFICO MEDIAS y DESVIOS

```
> plotmeans(VR ~ FACTOR, col="blue")
```



- Conclusiones

Debido a que se cumplen los requisitos (supuestos) de este Modelo Lineal (Normalidad y Homogeneidad de Varianza de los Errores), el valor p del ANOVA es válido para sacar conclusiones. Es decir **existe al menos una media diferente** entre los niveles del factor.

Con la prueba Tukey, observamos cuáles medias son diferentes entre sí, y a partir de eso realizamos las recomendaciones del caso: como se está buscando mayor producción, desde el punto de vista estrictamente estadístico se recomienda por igual las zonas 1, 2 y 3 ya que entre ellas no hay diferencias significativas