

Análisis de la varianza



Introducción.

El análisis de la varianza (**ANOVA**), puede considerarse una generalización del contraste de hipótesis de medias poblacionales iguales para el caso de poblaciones normales y con varianzas desconocidas, pero iguales. La generalización consiste en poder considerar más de dos poblaciones. La **hipótesis nula** será la que afirma que **las medias poblacionales de la variable métrica en estudio, para todas las poblaciones, son iguales**. La hipótesis alternativa, por su lado, afirmará que existe al menos dos poblaciones con medias diferentes. Como todo contraste, para llevarlo a cabo hemos de tener una muestra representativa de cada población, y fijar un nivel de significación (usualmente 0,05).

También se puede considerar el ANOVA como un **tipo especial de análisis de regresión**, en la que la variable dependiente es una variable métrica, y las variables explicativas son atributos o factores (en escala nominal u ordinal). La misión de los factores es clasificar a los casos que constituyen nuestra muestra en distintas submuestras, cada una representativa de una de las subpoblaciones cuyas medias en la variable en estudio se quiere comparar.

Aunque se pueden realizar ANOVAs con más de un atributo o factor, en esta práctica nos ceñiremos al caso más simple, en el que solo hay un atributo o factor que se ocupa de distribuir los casos de la muestra entre los distintos grupos o submuestras (a partir de la categoría del atributo o factor que toma cada caso).

En concreto, en esta práctica, comprobaremos **si la dimensión del grupo empresarial al que pertenecen las empresas eólicas** (medida en función

del número de empresas integradas en el grupo empresarial, y concretada en el factor DIMENSION) **tiene una influencia significativa sobre la rentabilidad económica** (variable RENEKO), **en términos medios**. Para ello se ha seleccionado una muestra constituida por 100 empresas productoras de electricidad mediante tecnología eólica. Así, la población, constituida por todas las empresas de generación eléctrica eólica de España, queda dividida en tres subpoblaciones: la subpoblación de empresas que pertenecen a grupos empresariales de DIMENSION o tamaño “GRANDE”, la subpoblación de empresas que pertenecen a grupos empresariales de DIMENSION o tamaño “MEDIO”, y la subpoblación de empresas que pertenecen a grupos empresariales de DIMENSION o tamaño “PEQUEÑO”. Cada una de estas subpoblaciones tendrán sus respectivas rentabilidades económicas medias, que desconocemos (ya que no tenemos los datos de la población, es decir, de todas las empresas eólicas del país; sino solo de una muestra de 100 empresas). Lo que si tenemos para cada subpoblación es una submuestra que la representa (parte de las 100 empresas de la muestra, que queda fraccionada en tres según el factor DIMENSION). Y de cada submuestra, tenemos la correspondiente rentabilidad media muestral. Lo que comprobaremos con el contraste de ANOVA, en definitiva, es si las diferencias observadas entre las rentabilidades medias de cada submuestra son lo suficientemente amplias como para pensar que, a nivel de toda la población, puede considerarse que existen diferencias importantes (significativas) entre las rentabilidades medias.

Preparando los datos.

Abriremos R-Studio pulsando en el icono correspondiente. A continuación, crearemos nuestro **proyecto** siguiendo la instrucción **File → New Project**. R-Studio nos preguntará si crea el proyecto en una nueva carpeta o en una ya existente. Vamos a crearlo, por ejemplo, en el disco extraíble D, carpeta R, subcarpeta “anova”, que ya está creada. Nos saldrá una ventana para buscar la carpeta y, cuando la encontremos, pulsamos **Open** y **Create Project**.

Vamos a ir a la carpeta del proyecto y vamos a guardar en ella los dos archivos de esta práctica: un archivo de Microsoft® Excel® llamado “eolica_100_mv.xlsx” y un *script* denominado “anova_eolica.R”. Si abrimos el primer archivo, comprobaremos que la tercera hoja que contiene (hoja “Datos”) guarda los datos que debemos importar desde R-Studio. Estos

datos se corresponden con diferentes variables económico-financieras de las 100 empresas productoras de electricidad mediante generación eólica con mayor volumen de activo.

Luego, cerraremos el archivo de Microsoft® Excel®, “eolica_100_mv.xlsx”, y volveremos a R-Studio. Después, abriremos nuestro script “anova_eolica.R” con **File → Open File...** Este *script* contiene el programa que vamos a ir ejecutando en la práctica.

La primera línea / instrucción en el script es:

```
rm(list = ls())
```

La instrucción tiene como objeto limpiar el *Environment* (memoria) de objetos de anteriores sesiones de trabajo.

Para importar los datos que hay en la hoja “Datos” del archivo “eolica_100_mv.xlsx”, ejecutaremos el código:

```
library(readxl)
datos <- read_excel("eolica_100_mv.xlsx", sheet = "Datos")
summary (datos)
```

R ha considerado la primera columna como una variable de tipo cualitativo, atributo, o factor. En realidad, esta columna no es una variable; sino que está formada por los nombres de los diferentes casos u observaciones. Para evitar que R tome la columna de los nombres de los casos como una variable más, podemos redefinir nuestro *data frame* diciéndole que tome esa primera columna como el conjunto de los nombres de los individuos (filas del *data frame*):

```
datos <- data.frame(datos, row.names = 1)
```

Si hacemos ahora un **summary()**:

```
summary (datos)
```

Observaremos que ya no aparece NOMBRE, puesto que la columna correspondiente ya no es considerada como una variable.

Antes de proceder al análisis, hemos de preparar nuestros datos mediante la localización de *missing values* y *outliers*. Los *outliers*, en el ANOVA,

pueden tener una gran influencia (ya que se trabaja en *términos medios*) sobre los resultados, por lo que deben ser tratados convenientemente.

Para localizar los casos concretos de *missing values*, puede recurrirse a utilizar las herramientas de manejo de *data frames* del paquete `dplyr`. Previamente, realizaremos una **copia** del *data frame* original, "datos", a la que llamaremos "muestra", que es con la que trabajaremos (para mantener la integridad del *data frame* original); y filtraremos los casos para **detectar aquellos que carecen de valor** en la variable "RENECO":

```
library (dplyr)
muestra<- select(datos, everything())
muestra %>% filter(is.na(RENECO)) %>% select(RENECO)
```

La función `is.na()` comprueba si, en la posición correspondiente a una fila o caso, para la variable escrita en el argumento; hay o no un dato o valor. Como resultado, se obtienen dos empresas sin dato para la variable RENECO:

	RENECO
Viesgo Renovables SL.	NA
Sargon Energias SLU	NA

Ante la existencia de *missing values*, se puede actuar de varios modos. Por ejemplo, **se puede intentar obtener por otro canal de información el conjunto de valores** de RENECO que no están disponibles, **o recurrir a alguna estimación** para los mismos y asignarlos. En caso de que esto sea difícil, se puede optar, simplemente, por **eliminar** estos casos, en especial cuando representan un porcentaje muy reducido respecto al total de casos. En nuestro ejemplo, vamos a suponer que hemos optado por esta última vía, al no conseguir unos valores más o menos verosímiles de RENECO para las empresas de las que se carece de dato. Esta eliminación de casos se podrá realizar mediante el código:

```
muestra <- muestra %>% filter(! is.na(RENECO))
```

El operador "!" significa "no".

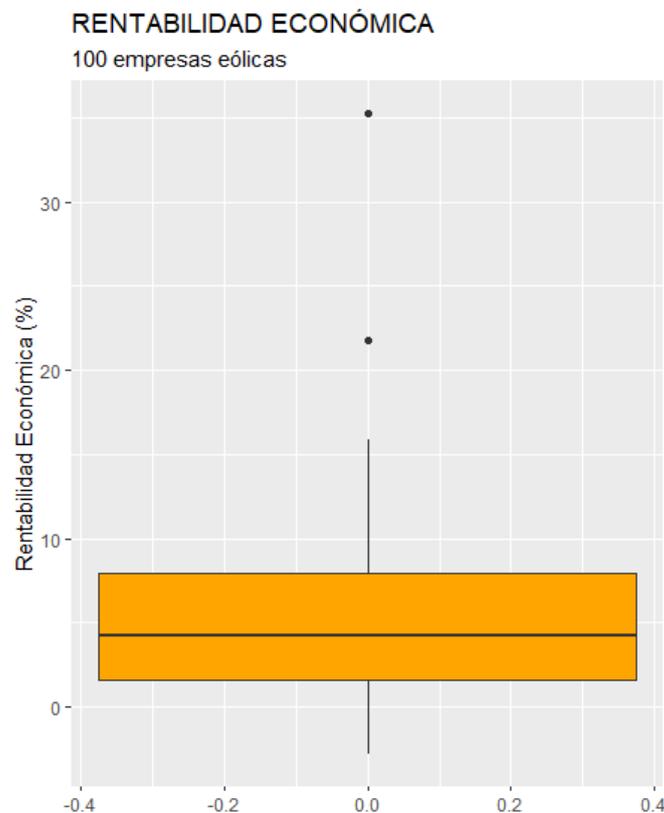
Podemos comprobar cómo en el *Global Environment* aparece el *data frame* "muestra" con dos casos menos (98).

Una vez tratados los casos con valores perdidos o *missing values*, **es necesario detectar la presencia de outliers** o casos atípicos en la muestra,

que pudieran desvirtuar los resultados derivados del ANOVA. Para ello, realizaremos un **boxplot** o gráfico de caja. Aplicaremos, por ejemplo, el código siguiente, que utiliza la gramática del paquete **ggplot2**:

```
library (ggplot2)
ggplot(data = muestra, map = (aes(y = RENECA))) +
  geom_boxplot(fill = "orange") +
  ggtitle("RENTABILIDAD ECONÓMICA", subtitle = "100 empresas eólicas")
+
  ylab("Rentabilidad Económica (%)")
```

Obteniéndose el gráfico:



La *caja* contiene el 50% de los casos centrales (los que van del primer cuartil al tercero, cuya diferencia se llama *rango intercuartílico*), y contiene una línea horizontal que es la *mediana*. Por arriba sobresale una línea que llega al valor más grande que no llega a ser atípico o *outlier*; y por debajo de la caja lo mismo pero mínimo. Un valor atípico o *outlier* es el que se aleja más de 1.5 veces el rango intercuartílico (altura de la caja) del tercer cuartil, por arriba; o del primer cuartil, por abajo. Se registran mediante puntos.

En nuestro caso, el *boxplot* ratifica la existencia de dos casos atípicos. Para identificar esos dos casos concretos, recurrimos al paquete `dplyr`, y **establecemos un filtro** con el siguiente código:

```
Q1 <- quantile (muestra$RENECO, c(0.25))
Q3 <- quantile (muestra$RENECO, c(0.75))
Q1; Q3
muestra %>% filter(RENECO > Q3 + 1.5*IQR(RENECO) | RENECO < Q1 -
1.5*IQR(RENECO)) %>% select(RENECO)
```

En el código anterior, las dos primeras filas calculan los cuartiles primero (Q1) y tercero (Q3) mediante la función `quantile()`. Es preciso tener en cuenta que esta función calcula los percentiles. Luego se filtran, mediante la función de `dplyr filter()`, los *outliers*, calculados como aquellos casos con valores de RENECO mayores que Q3 más 1,5 veces el rango intercuartílico de la variable; o menores que Q1 menos 1,5 veces dicho rango intercuartílico. Para calcular el rango intercuartílico se recurre a la función `IQR()`. Finalmente, con `select()`, se muestran los casos en la consola de `R-Studio`.

Al ejecutar el código anterior, se visualizarán esas dos empresas atípicas:

```
RENECO
Molinos Del Ebro SA 35.262
Sierra De Selva SL 21.761
```

Como ocurría con los *missing values*, el tratamiento de los *outliers* depende de la información que se tenga, existiendo varias alternativas (corrección del dato, estimación, etc.) Si no se tiene información fiable, y los *outliers* no representan una gran proporción respecto al total de casos, puede optarse por su eliminación de la muestra, como haremos en este ejemplo. Podemos hacerlo creando un nuevo *data frame* a partir de “muestra”; pero sin esos dos casos. Ese nuevo *data frame* se llamará, por ejemplo, “**muestra_so**”:

```
muestra_so <- muestra %>% filter(RENECO <= Q3 + 1.5*IQR(RENECO) & RENECO
>= Q1 - 1.5*IQR(RENECO))
```

Es importante observar que, en el código de `filter()`, las desigualdades deben cambiar, así como el operador “|” por el operador “&”.

En el *Global Environment* podemos comprobar cómo el *data frame* “muestra_so” posee el mismo número de variables que el *data frame* “muestra”; pero con dos observaciones o casos menos (96).

Análisis de la Varianza.

Vamos a proceder ahora a realizar el análisis la varianza o ANOVA que, como ya hemos dicho, puede interpretarse como una generalización del contraste estadístico de diferencia entre medias de dos poblaciones normales. Aquí, hablaremos de tres (sub)poblaciones (la población de empresas eólicas la hemos dividido en tres grupos dependiendo de los tamaños de los grupos empresariales de pertenencia de las empresas). Las submuestras que representan a cada (sub)población se forman en R a través del valor del factor DIMENSION, que puede adoptar los niveles "GRANDE", "MEDIO" o "PEQUEÑO".

Para conocer el tamaño de cada submuestra, y el valor de la rentabilidad económica muestral, basta con ejecutar el código:

```
# Visualizando número de frecuencias y medias de los grupos con dplyr:
muestra_so %>% group_by(DIMENSION) %>% summarise (observaciones =
table(DIMENSION), media = mean(RENECO))
```

En el código anterior se ha empleado la gramática del paquete **dplyr** en combinación con las funciones **table()**, que cuenta el número de casos en cada grupo, y **mean()**, que calcula las medias de los grupos.

El resultado obtenido en la consola es:

```
  DIMENSION observaciones media
  <chr>      <table>        <dbl>
1 GRANDE    51                5.25
2 MEDIO     34                3.73
3 PEQUEÑO   11                7.21
```

Además, estas submuestras pueden ser visualizadas a través de un gráfico generado con el paquete **ggplot2**. En concreto, el gráfico presenta un histograma para cada (sub)muestra o nivel de DIMENSION, añadiendo los casos individuales y la media de cada una de ellas:

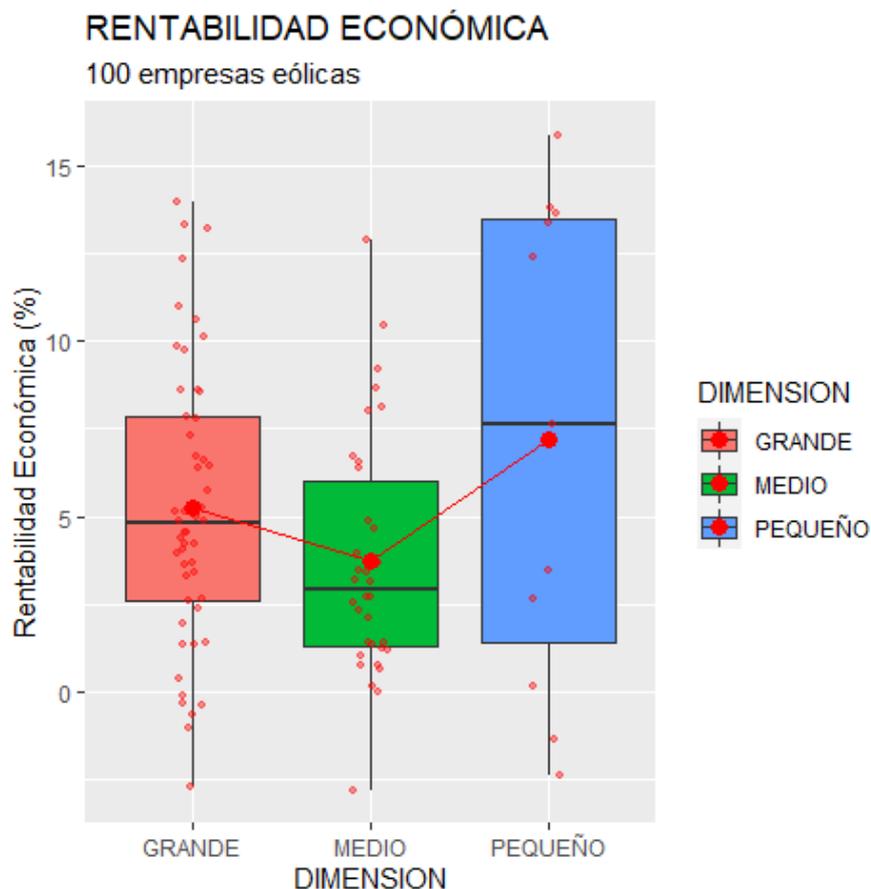
```
# Graficando casos y medias.
ggplot(data = muestra_so, map = (aes(x = DIMENSION, y = RENECO, fill =
DIMENSION))) +
  geom_boxplot(outlier.shape = NA) +
  stat_summary(fun = "mean",
              geom = "point",
              size = 3,
              col = "red") +
```

```

stat_summary(fun = "mean",
             geom = "line",
             col = "red",
             map = (aes(group = TRUE))) +
geom_jitter(width = 0.1,
            size = 1,
            col = "red",
            alpha = 0.40) +
ggtitle("RENTABILIDAD ECONÓMICA", subtitle = "100 empresas eólicas")
+
ylab("Rentabilidad Económica (%)")

```

El gráfico obtenido con el código anterior es:



Los puntos pequeños recogen los casos, mientras que los grandes son las medias de los grupos.

El gráfico sugiere que existen diferencias entre las rentabilidades medias de los grupos de empresas. Con el ANOVA lo que se pretende discernir es si las diferencias **observadas** entre las rentabilidades medias en cada submuestra implican una **diferencia relevante entre las medias o esperanzas poblacionales** de las rentabilidades para cierta significación estadística (es

decir, no solo considerando las empresas que componen la muestra; sino todas las empresas eólicas que componen la población).

En el ANOVA, la **hipótesis nula** a contrastar sugiere **que no existen diferencias entre las rentabilidades económicas medias de los grupos** (subpoblaciones) de empresas eólicas (discriminadas por el tamaño del grupo empresarial al que pertenecen). De ser así, el factor DIMENSION no tendría una influencia estadísticamente significativa sobre el valor medio de la variable RENEKO.

Las conclusiones a las que lleguemos con el *contraste F de ANOVA* serán válidas en la medida en que se cumplan las **hipótesis de normalidad y homogeneidad en las varianzas** de la variable dependiente o métrica (RENEKO), en los tres grupos o subpoblaciones en que queda dividida la población atendiendo los niveles o categorías del factor (DIMENSION). Hemos de contrastar, pues, ambas hipótesis, a partir de la información de las tres submuestras que tenemos y que representan, respectivamente, a cada una de esas subpoblaciones.

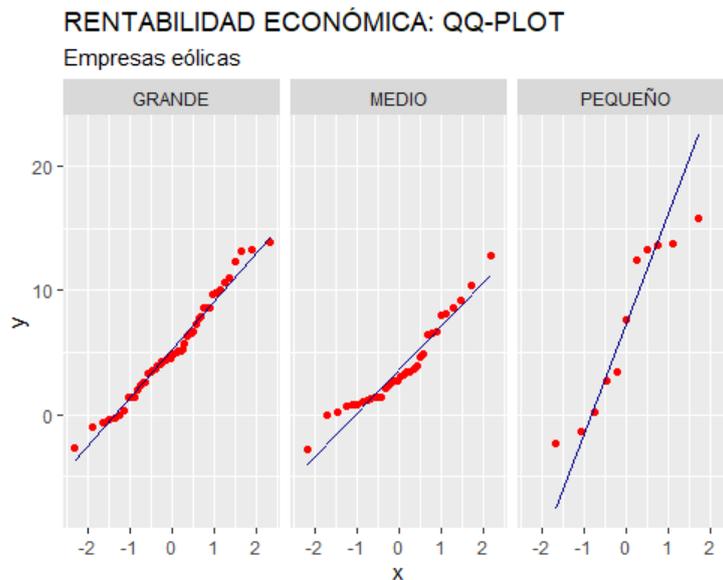
En cuanto a la hipótesis de **normalidad**, se pueden usar dos vías para verificar su cumplimiento: el análisis gráfico y la realización de contrastes estadísticos.

El análisis gráfico puede llevarse a cabo mediante la realización *de gráficos QQ* de la variable RENEKO, a partir de las submuestras o grupos de empresas en que queda fragmentada la muestra a partir de la variable DIMENSION. Estos gráficos pueden generarse con la gramática del paquete **ggplot2**:

```
# Normalidad

ggplot(data = muestra_so, aes(sample = RENEKO)) +
  stat_qq(colour = "red") +
  stat_qq_line(colour = "dark blue") +
  ggtitle("RENTABILIDAD ECONÓMICA: QQ-PLOT", subtitle = "Empresas
eólicas") +
  facet_grid(. ~ DIMENSION)
```

Dando lugar al gráfico:



Los *gráficos QQ* parecen mostrar un comportamiento aproximadamente normal de la variable RENEKO en cada grupo o submuestra (y, por tanto, en cada subpoblación), al situarse los puntos relativamente cerca de la línea recta.

Para asegurarnos de la normalidad de un modo más preciso, vamos a realizar el contraste de normalidad de *Shapiro-Wilk*:

```
shapiro.test(x=muestra_so$RENECO[muestra_so$DIMENSION=="GRANDE"])
shapiro.test(x=muestra_so$RENECO[muestra_so$DIMENSION=="MEDIO"])
shapiro.test(x=muestra_so$RENECO[muestra_so$DIMENSION=="PEQUEÑO"])
```

En esta prueba, la hipótesis nula equivale al supuesto de normalidad. Para un 5% de significación estadística, un **p-valor superior a 0,05 implicará el no-rechazo de la hipótesis (nula) de normalidad.**

Los resultados obtenidos son los siguientes:

```
Shapiro-Wilk normality test
data: muestra_so$RENECO[muestra_so$DIMENSION == "GRANDE"]
W = 0.97774, p-value = 0.4479

Shapiro-Wilk normality test
data: muestra_so$RENECO[muestra_so$DIMENSION == "MEDIO"]
W = 0.93489, p-value = 0.0435

Shapiro-Wilk normality test
data: muestra_so$RENECO[muestra_so$DIMENSION == "PEQUEÑO"]
W = 0.88154, p-value = 0.1088
```

Los resultados en este caso presentan, en el caso de las submuestras correspondientes a las dimensiones “GRANDE” y “PEQUEÑO”, p-valores superiores a 0,05; por lo que **no se rechaza la hipótesis nula de normalidad** y puede considerarse que **estas dos subpoblaciones siguen un comportamiento normal** en su distribución. En cambio, en el caso grupo de dimensión “MEDIO”, el p-valor es inferior a 0,05, por lo que **se rechaza la hipótesis nula de normalidad** en la distribución de RENEKO para la población de empresas pertenecientes a grupos empresariales de tamaño (factor DIMENSION) medio.

En cuanto a la **homogeneidad de las varianzas**, contrastamos este supuesto mediante la prueba de *Bartlett*.

```
#Homogeneidad en las varianzas.  
bartlett.test(muestra_so$RENECO~muestra_so$DIMENSION)
```

El resultado obtenido es:

```
Bartlett test of homogeneity of variances  
data: muestra_so$RENECO by muestra_so$DIMENSION  
Bartlett's K-squared = 9.4336, df = 2, p-value = 0.008944
```

El p-valor es menor que 0,05; luego **se rechaza la hipótesis nula de homogeneidad de las varianzas de los grupos (subpoblaciones)**.

Puesto que en uno de los grupos (subpoblaciones) se rechaza la hipótesis de normalidad, y (sobre todo) puesto que no se puede considerar una dispersión similar en las tres subpoblaciones (varianzas homogéneas), los resultados de la prueba **F de ANOVA** pierden validez, y habría que optar por **una alternativa robusta**. No obstante, **a modo ilustrativo**, seguiremos adelante con la prueba *F de ANOVA*.

El contraste *F de ANOVA* de igualdad en las medias de la variable en estudio (rentabilidad económica, RENEKO) de las distintas (sub)poblaciones (grupos de empresas según el tamaño del grupo empresarial de pertenencia) se realiza en R mediante la función `avov()`, que guardaremos, por ejemplo, como el objeto “Datos.avov”, y que luego se mostrará en forma resumida:

```
# Test F de ANOVA  
Datos.avov <- avov(muestra_so$RENECO~muestra_so$DIMENSION)  
summary(Datos.avov)
```

Como resultado del código anterior, se obtiene la siguiente información:

```
              Df Sum Sq Mean Sq F value Pr(>F)
muestra_so$DIMENSION  2  111.5   55.73   3.131 0.0483 *
Residuals              93 1655.0   17.80
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

El valor del estadístico F de ANOVA es de 3,131; con un p-valor asociado de 0,0483. Como el p-valor es menor que 0,05, **se rechaza la hipótesis nula de medias iguales**; por lo que podremos afirmar (para una significación del 5%) que **el tamaño del grupo empresarial de pertenencia influye, en media, en la rentabilidad económica obtenida**. Aún así, hay que advertir que el resultado no es muy concluyente, ya que el p-valor está muy cercano a la significación establecida de 0,05

Comparaciones múltiples.

Cuando se rechaza la hipótesis nula de medias iguales, es interesante poder **determinar cuáles son los grupos concretos (subpoblaciones) que difieren** en la media de la variable dependiente métrica analizada (RENECO en nuestro caso). Este análisis puede llevarse a cabo mediante el test *HSD de Tuckey*. Un modo de obtener los resultados de esta prueba es ejecutarla desde las funciones del paquete **emmeans**, por lo que llamaremos a la librería correspondiente:

```
library(emmeans)
```

Los resultados de la prueba *HSD de Tuckey* se obtendrán aplicando la función **pairs()** a un objeto creado previamente con la función **emmeans()**, al que hemos llamado, por ejemplo, “medias”, y que tiene como argumentos el nombre de nuestro objeto ANOVA anterior y, entrecomillado, el nombre de la variable que actúa como factor que divide a la muestra en los tres grupos (submuestras) comparados:

```
# Comparaciones múltiples.

library(emmeans)
medias<-emmeans(Datos.aov, "DIMENSION")
pairs(medias)
```

El resultado es el siguiente:

```
contrast          estimate      SE df t.ratio p.value
```

GRANDE - MEDIO	1.52	0.934	93	1.629	0.2386
GRANDE - PEQUEÑO	-1.96	1.402	93	-1.397	0.3465
MEDIO - PEQUEÑO	-3.48	1.463	93	-2.379	0.0504

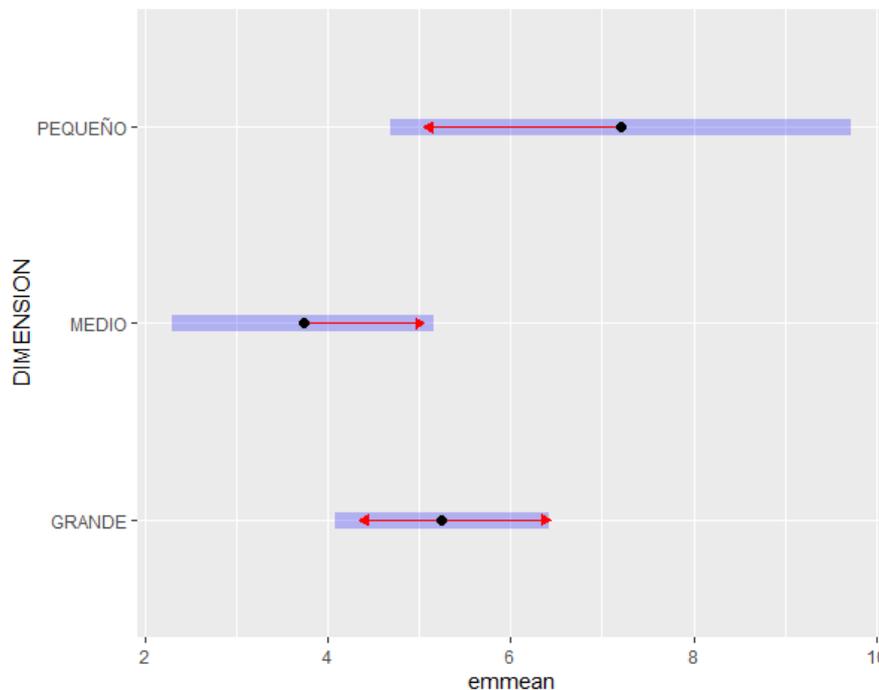
P value adjustment: tukey method for comparing a family of 3 estimates

Comprobamos cómo en la comparación entre los distintos grupos (subpoblaciones), y según los p-valores correspondientes, **no se rechaza la hipótesis nula**, por lo que la diferencia entre estos las subpoblaciones en la rentabilidad media no es significativa, para una significación de 0,05. Así pues, se concluye que la clave está en que las fincas que cultivan. Tan solo entre los grupos (subpoblaciones) “MEDIO” y “PEQUEÑO” el p-valor es muy próximo a 0,05, y podría admitirse que exista, en media, diferencias relevantes en las rentabilidades económicas. Esto es coherente con el resultado del contraste F de ANOVA en el sentido de que, aunque salió como resultado que existían diferencias significativas, el p-valor coincidía prácticamente con la significación establecida de 0,05.

Para representar el resultado, en el que se representan las medias con sus intervalos de confianza, puede ejecutarse el código:

```
plot(medias, comparisons = TRUE)
```

Obteniéndose el gráfico:



En el gráfico anterior puede advertirse cómo los intervalos de confianza de las medias de las rentabilidades de los grupos (subpoblaciones) de empresas “GRANDE” y “MEDIO”, y “GRANDE” y “PEQUEÑO”, están prácticamente solapados; al contrario que los intervalos de rentabilidad media de las empresas pertenecientes a grupos empresariales de dimensión “PEQUEÑO” y “MEDIO”, cuyo solapamiento es mucho menor.

¿Y si no se cumplen las hipótesis de ANOVA?

Cuando no se cumplen los supuestos del modelo (normalidad en la distribución de los grupos y/o homogeneidad de las varianzas de los grupos), como ocurre en este ejemplo, debemos utilizar una prueba alternativa **robusta**, como es el caso de la prueba de **Kruskal-Wallis**. Para ello usaremos la librería **pgirmess** y el código:

```
#Cuando se incumplen las hipotesis de ANOVA (test robusto)

library(pgirmess)
Datos.K<-kruskal.test(muestra_so$RENECO~muestra_so$DIMENSION)
Datos.K
```

En el código anterior, activamos en memoria las funciones del paquete **pgirmess**, asignamos al objeto llamado, por ejemplo, “Datos.K” el resultado de la función **kruskal.test()**, y lo visualizamos en pantalla. Se obtiene:

```
Kruskal-Wallis rank sum test

data: muestra_so$RENECO by muestra_so$DIMENSION
Kruskal-Wallis chi-squared = 4.5291, df = 2, p-value = 0.1039
```

El p-valor es superior a 0,05, luego, para una significación del 5%, **no se rechaza la hipótesis nula de que las medias de la variable RENECO de los grupos (subpoblaciones) pueden considerarse iguales.**

Cuando no se rechaza la hipótesis nula, no existirán diferencias significativas en las rentabilidades económicas medias, por lo que **aquí acabaría el análisis**. No obstante, si el resultado hubiera sido de rechazo de la hipótesis nula de igualdad en las medias, podría procederse a un análisis de **comparaciones múltiples** para detallar entre qué grupos (subpoblaciones) concretos se dan las diferencias relevantes. Para ello, existe otra prueba robusta de *Kruskal*, para lo que ha de ejecutarse la función **kruskalmc()**:

```
Datos.KMC<-kruskalmc(muestra_so$RENECO~muestra_so$DIMENSION)
Datos.KMC
```

En este caso, se asignan los resultados al objeto llamado, por ejemplo, "Datos.KMC" que, al invocarlo, muestra la solución:

```
Comparisons
      obs.dif critical.dif difference
GRANDE-MEDIO  11.284314    14.76510    FALSE
GRANDE-PEQUEÑO 5.188948    22.17000    FALSE
MEDIO-PEQUEÑO 16.473262    23.13246    FALSE
```

En coherencia con el resultado previo, se obtiene que no existen diferencias significativas entre las rentabilidades medias de los tres grupos (subpoblaciones), al ser comparadas dos a dos.

This work © 2022 by [Miguel Ángel Tarancón](#) and [Consolación Quintana](#) is licensed under [Attribution-NonCommercial-NoDerivatives 4.0 International](#) 

Updated: 19/10/2022