### A picture containing text Description automatically generatedUniversidad de Granada

### Escuela Internacional de Posgrado

### Máster en Estadística Aplicada

### Materia: Técnicas Estadísticas Multivariantes.

### Alumno: Francisco Javier Márquez Rosales

# **Análisis Análisis multivariante de la varianza (MANOVA):**

# **Ejercicios de MANOVA:**

Noviembre, 2022

Ejercicios de Manova

Realiza las siguientes actividades y súbelas en un pdf:

1- Hacer un resumen esquemático del tema.

2- Estudiar el libro de R MANOVA (help(manova)) y los libros relacionados y la página web https://gaopinghuang0.github.io/2017/11/20/MANOVA-notes-and-R-code y hacer un resumen de ellos.

3- Buscar un ejemplo donde se puedan aplicar un MANOVA y aplicarlo.

**3- Buscar un ejemplo donde se puedan aplicar un MANOVA y aplicarlo.**

**Solución:**

A continuación aplicaremos un MANOVA a los datos extraidos de la revista estadounidense Motor Trend de 1974 y comprenden el consumo de combustible y 10 aspectos del diseño y rendimiento del automóvil para 32 automóviles (modelos 1973–74).

Las variables incluidas en el conjunto original son las siguientes:

El diccionario de datos del archivo:

[, 1] mpg Miles/(US) gallon [, 2] cyl Number of cylinders [, 3] disp Displacement (cu.in.) [, 4] hp Gross horsepower [, 5] drat Rear axle ratio [, 6] wt Weight (1000 lbs) [, 7] qsec 1/4 mile time [, 8] vs Engine (0 = V-shaped, 1 = straight) [, 9] am Transmission (0 = automatic, 1 = manual) [,10] gear Number of forward gears [,11] carb Number of carburetors

Inicialemnte hacemos una rápida verificación del conjunto, tanto de la estructura como del contenido de las variables:

str(mtcars)

## 'data.frame': 32 obs. of 11 variables:

## $ mpg : num 21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...

## $ cyl : num 6 6 4 6 8 6 8 4 4 6 ...

## $ disp: num 160 160 108 258 360 ...

## $ hp : num 110 110 93 110 175 105 245 62 95 123 ...

## $ drat: num 3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...

## $ wt : num 2.62 2.88 2.32 3.21 3.44 ...

## $ qsec: num 16.5 17 18.6 19.4 17 ...

## $ vs : num 0 0 1 1 0 1 0 1 1 1 ...

## $ am : num 1 1 1 0 0 0 0 0 0 0 ...

## $ gear: num 4 4 4 3 3 3 3 4 4 4 ...

## $ carb: num 4 4 1 1 2 1 4 2 2 4 ...

summary(mtcars)

## mpg cyl disp hp

## Min. :10.40 Min. :4.000 Min. : 71.1 Min. : 52.0

## 1st Qu.:15.43 1st Qu.:4.000 1st Qu.:120.8 1st Qu.: 96.5

## Median :19.20 Median :6.000 Median :196.3 Median :123.0

## Mean :20.09 Mean :6.188 Mean :230.7 Mean :146.7

## 3rd Qu.:22.80 3rd Qu.:8.000 3rd Qu.:326.0 3rd Qu.:180.0

## Max. :33.90 Max. :8.000 Max. :472.0 Max. :335.0

## drat wt qsec vs

## Min. :2.760 Min. :1.513 Min. :14.50 Min. :0.0000

## 1st Qu.:3.080 1st Qu.:2.581 1st Qu.:16.89 1st Qu.:0.0000

## Median :3.695 Median :3.325 Median :17.71 Median :0.0000

## Mean :3.597 Mean :3.217 Mean :17.85 Mean :0.4375

## 3rd Qu.:3.920 3rd Qu.:3.610 3rd Qu.:18.90 3rd Qu.:1.0000

## Max. :4.930 Max. :5.424 Max. :22.90 Max. :1.0000

## am gear carb

## Min. :0.0000 Min. :3.000 Min. :1.000

## 1st Qu.:0.0000 1st Qu.:3.000 1st Qu.:2.000

## Median :0.0000 Median :4.000 Median :2.000

## Mean :0.4062 Mean :3.688 Mean :2.812

## 3rd Qu.:1.0000 3rd Qu.:4.000 3rd Qu.:4.000

## Max. :1.0000 Max. :5.000 Max. :8.000

Disponemos de 11 variables, algunas son variables categoricas aún cuando están expresadas con numeros, este es el caso de:

vs Engine (0 = V-shaped, 1 = straight) am Transmission (0 = automatic, 1 = manual)

Estas variables al igual que las siguientes serán descartadas para efectos del análisis:

gear Number of forward gears carb Number of carburetors

así nuestro conjunto se reduce a 7 variables en donde la variable cilindros (cyl) sería, para este ejemplo, la variable respuesta del modelo.

Preparamos los datos

*#preparamos los datos para su análisis, nos vamos a quedar con las variables numericas y como variable respuesta 'cilindors'*

yvars<- cbind(mtcars$mpg, mtcars$disp, mtcars$hp, mtcars$drat,mtcars$wt, mtcars$qsec)

colnames(yvars) <-c("mpg","disp","hp","drat","wt","qsec")

*#creo un conjunto solo con las variables a analizar, esto para facilitar análisis, graficas y validacion de supuestos posteriores. Convertimos a caracter la variable cilindros*

mtcars1<-mtcars

mtcars1$cil<-as.character(mtcars$cyl)

mtcars1$cyl <- NULL

mtcars1$vs <- NULL

mtcars1$am <- NULL

mtcars1$gear <- NULL

mtcars1$carb <- NULL

str(mtcars1)

## 'data.frame': 32 obs. of 7 variables:

## $ mpg : num 21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...

## $ disp: num 160 160 108 258 360 ...

## $ hp : num 110 110 93 110 175 105 245 62 95 123 ...

## $ drat: num 3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...

## $ wt : num 2.62 2.88 2.32 3.21 3.44 ...

## $ qsec: num 16.5 17 18.6 19.4 17 ...

## $ cil : chr "6" "6" "4" "6" ...

Realizaremos en primer lugar un análisis MANOVA para comprobar la igualdad de los grupos según los diferentes tipos (números) de cilindros:

mma<-manova(yvars ~ cil, data=mtcars1)

mma

## Call:

## manova(yvars ~ cil, data = mtcars1)

##

## Terms:

## cil Residuals

## mpg 824.8 301.3

## disp 398891.0 77293.8

## hp 104030.5 41696.3

## drat 4.4 4.5

## wt 18.2 11.5

## qsec 34.6 64.4

## Deg. of Freedom 2 29

##

## Residual standard errors: 3.223099 51.62659 37.91839 0.3938298 0.6298047 1.489994

## Estimated effects may be unbalanced

parece haber diferencias entre las variables según el número de cilindros. Si obtenemos los contrastes mas aplicados en este caso:

summary(mma, test="Wilks")

## Df Wilks approx F num Df den Df Pr(>F)

## cil 2 0.064564 11.742 12 48 1.542e-10 \*\*\*

## Residuals 29

## ---

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

summary(mma, test="Roy")

## Df Roy approx F num Df den Df Pr(>F)

## cil 2 9.8951 41.23 6 25 8.875e-12 \*\*\*

## Residuals 29

## ---

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

summary(mma, test="Hotelling-Lawley")

## Df Hotelling-Lawley approx F num Df den Df Pr(>F)

## cil 2 10.317 19.774 12 46 2.936e-14 \*\*\*

## Residuals 29

## ---

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

summary(mma, test="Pillai")

## Df Pillai approx F num Df den Df Pr(>F)

## cil 2 1.2048 6.3127 12 50 1.249e-06 \*\*\*

## Residuals 29

## ---

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

El resultado obtenido para todos los contrastes, todos con p-valor menor a 0.05, indica que existen diferencias entre los vectores medias de los tres tipos de cilindros (números de cilindros en motores).

Contrastes marginales de medias Hemos contrastado la igualdad de medias conjuntas, en este caso haremos los contrastes marginales.

*# mtcars$mpg, mtcars$cyl, mtcars$disp, mtcars$hp, mtcars$drat,mtcars$wt, mtcars$qsec, mtcars$gear, mtcars$vs)*

modelo1<-aov(mtcars1$mpg ~ mtcars1$cil )

summary(modelo1)

## Df Sum Sq Mean Sq F value Pr(>F)

## mtcars1$cil 2 824.8 412.4 39.7 4.98e-09 \*\*\*

## Residuals 29 301.3 10.4

## ---

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

modelo2<-aov(mtcars1$disp ~ mtcars1$cil )

summary(modelo2)

## Df Sum Sq Mean Sq F value Pr(>F)

## mtcars1$cil 2 398891 199445 74.83 3.55e-12 \*\*\*

## Residuals 29 77294 2665

## ---

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

modelo3<-aov(mtcars1$hp ~ mtcars1$cil )

summary(modelo3)

## Df Sum Sq Mean Sq F value Pr(>F)

## mtcars1$cil 2 104031 52015 36.18 1.32e-08 \*\*\*

## Residuals 29 41696 1438

## ---

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

modelo4<-aov(mtcars1$drat ~ mtcars1$cil )

summary(modelo4)

## Df Sum Sq Mean Sq F value Pr(>F)

## mtcars1$cil 2 4.364 2.1822 14.07 5.36e-05 \*\*\*

## Residuals 29 4.498 0.1551

## ---

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

modelo5<-aov(mtcars1$wt ~ mtcars1$cil )

summary(modelo5)

## Df Sum Sq Mean Sq F value Pr(>F)

## mtcars1$cil 2 18.18 9.088 22.91 1.07e-06 \*\*\*

## Residuals 29 11.50 0.397

## ---

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

modelo6<-aov(mtcars1$qsec ~ mtcars1$cil )

summary(modelo6)

## Df Sum Sq Mean Sq F value Pr(>F)

## mtcars1$cil 2 34.61 17.30 7.794 0.00196 \*\*

## Residuals 29 64.38 2.22

## ---

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

Del resultado de los contrastes marginales podemos observar que individualemte para cada una de las variables se observan diferencias significativas en las medias para cada tipo de numero de cilindros

Obtengamos ahora una representación grafica del conjunto para visualizar distribución, comportamiento bivariante y segemntación por cada grupo de cilindros:

**require**(RColorBrewer)

## Loading required package: RColorBrewer

## Warning: package 'RColorBrewer' was built under R version 4.1.3

colors <- brewer.pal(nlevels(mtcars1$cil),"Dark2")

## Warning in brewer.pal(nlevels(mtcars1$cil), "Dark2"): minimal value for n is 3, returning requested palette with 3 different levels

**require**(car)

## Loading required package: car

## Loading required package: carData

scatterplotMatrix(mtcars1[1:6],groups=mtcars1[[7]],

smooth=FALSE,regLine=FALSE, legend=T, oma=c(0,0,8,0),

col=colors)

A picture containing calendar

Description automatically generated

*#legend("top",legend=levels(mtcars1[[7]],pch=1:3,col=colors,horiz=TRUE,bty="n",*

*#cex=1.5,xpd=TRUE)*

A continuación contrastaremos las hipotesis que sustentan el modelo

1. Normalidad:

Debemos validar que el conjunto de datos se comporta como una normal multivariante

**require**(MVN)

## Loading required package: MVN

## Warning: package 'MVN' was built under R version 4.1.3

mvn(data =mtcars1[1:6] , univariateTest = "SW", desc = T)

## $multivariateNormality

## Test HZ p value MVN

## 1 Henze-Zirkler 1.046162 0.001459679 NO

##

## $univariateNormality

## Test Variable Statistic p value Normality

## 1 Shapiro-Wilk mpg 0.9476 0.1229 YES

## 2 Shapiro-Wilk disp 0.9200 0.0208 NO

## 3 Shapiro-Wilk hp 0.9334 0.0488 NO

## 4 Shapiro-Wilk drat 0.9459 0.1101 YES

## 5 Shapiro-Wilk wt 0.9433 0.0927 YES

## 6 Shapiro-Wilk qsec 0.9733 0.5935 YES

##

## $Descriptives

## n Mean Std.Dev Median Min Max 25th 75th

## mpg 32 20.090625 6.0269481 19.200 10.400 33.900 15.42500 22.80

## disp 32 230.721875 123.9386938 196.300 71.100 472.000 120.82500 326.00

## hp 32 146.687500 68.5628685 123.000 52.000 335.000 96.50000 180.00

## drat 32 3.596563 0.5346787 3.695 2.760 4.930 3.08000 3.92

## wt 32 3.217250 0.9784574 3.325 1.513 5.424 2.58125 3.61

## qsec 32 17.848750 1.7869432 17.710 14.500 22.900 16.89250 18.90

## Skew Kurtosis

## mpg 0.6106550 -0.37276603

## disp 0.3816570 -1.20721195

## hp 0.7260237 -0.13555112

## drat 0.2659039 -0.71470062

## wt 0.4231465 -0.02271075

## qsec 0.3690453 0.33511422

Se puede observar que las variables de forma conjunta no son normales, ya que el p-value del test resulto en 0.001 menor a 0.05 para una prueba al 95% de confianza. De forma marginal, las que no resultaron normales fueron: desplazamiento (disp) y caballos de fuerza (hp).

1. Igualdad de las matrices de varianzas covarianzas: Ahora, contrastamos la hipótesis sobre igualdad de las matrices de varianzas covarianzas.

by(mtcars1[,-7], mtcars1$cil, var)

## mtcars1$cil: 4

## mpg disp hp drat wt qsec

## mpg 20.3385455 -97.583545 -49.424545 0.69923636 -1.83191091 -1.7903091

## disp -97.5835455 722.082545 244.484545 -4.91523636 13.10932091 14.8188091

## hp -49.4245455 244.484545 438.254545 -3.59763636 1.90629091 -6.2820909

## drat 0.6992364 -4.915236 -3.597636 0.13356909 -0.09968073 -0.1742373

## wt -1.8319109 13.109321 1.906291 -0.09968073 0.32440282 0.6113902

## qsec -1.7903091 14.818809 -6.282091 -0.17423727 0.61139018 2.8306218

## ------------------------------------------------------------

## mtcars1$cil: 6

## mpg disp hp drat wt qsec

## mpg 2.11285714 6.227619 -4.480952 0.07938095 -0.35302381 -1.0388571

## disp 6.22761905 1727.438095 -517.904762 -16.44409524 6.99838095 55.9617143

## hp -4.48095238 -517.904762 588.571429 2.50809524 -2.64738095 -26.0057143

## drat 0.07938095 -16.444095 2.508095 0.22662857 -0.06016429 -0.5063143

## wt -0.35302381 6.998381 -2.647381 -0.06016429 0.12698214 0.5267071

## qsec -1.03885714 55.961714 -26.005714 -0.50631429 0.52670714 2.9133905

## ------------------------------------------------------------

## mtcars1$cil: 8

## mpg disp hp drat wt qsec

## mpg 6.55384615 -90.178462 -37.0153846 0.04569231 -1.2643692 -0.3194615

## disp -90.17846154 4592.952308 408.5461538 -2.32561538 38.8713923 15.8160769

## hp -37.01538462 408.546154 2598.6428571 12.67708791 0.6820275 -46.0620330

## drat 0.04569231 -2.325615 12.6770879 0.13865330 -0.1017598 -0.3667445

## wt -1.26436923 38.871392 0.6820275 -0.10175984 0.5766956 0.4873249

## qsec -0.31946154 15.816077 -46.0620330 -0.36674451 0.4873249 1.4304489

contraste de igualdad de matrices de varianzas

**require**(biotools)

## Loading required package: biotools

## Warning: package 'biotools' was built under R version 4.1.3

## Loading required package: MASS

## ---

## biotools version 4.2

boxM(mtcars1[,-7], mtcars1$cil)

##

## Box's M-test for Homogeneity of Covariance Matrices

##

## data: mtcars1[, -7]

## Chi-Sq (approx.) = 88.45, df = 42, p-value = 3.708e-05

El resultado anterior nos indica rechazar la hipotesis nula, en otras palabras, las matrices de varianzas covarianzas son estadítsicamente difererentes.