



**INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE  
MONTERREY**

Herramientas computacionales: el arte de la analítica (Gpo 201)

**Actividad Evaluable 5: Análisis de datos**

Equipo 4:

Ian Alexei Armendariz Martinez

A01753288

Emiliano Caballero Mendoza

A01749050

Profesor:

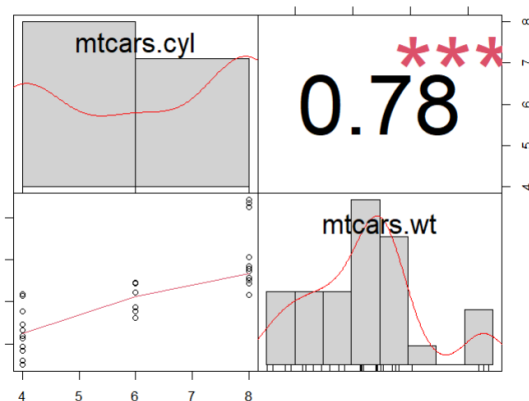
Jorge Adolfo Ramírez Uresti

## Código Fuente

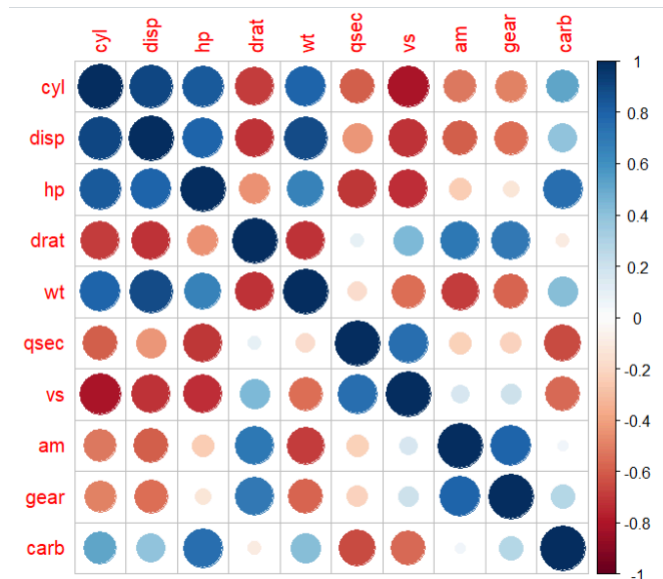
```
1 # Actividad Evaluable 5: Análisis de datos
2
3 # Dataset de mtcars
4 View(mtcars)
5
6 #Análisis de correlacion
7 # Comparacion entre potencia de motor y peso de vehiculo
8 library(PerformanceAnalytics)
9 cor(mtcars$cyl,mtcars$wt)
10 corGE <- data.frame(mtcars$cyl, mtcars$wt)
11 chart.Correlation(corGE)
12
13 library(corrplot)
14 # Grafica matriz de correlacion
15 corrcars <- cor(mtcars)
16 corrplot(corrcars,method = "circle")
17
18 # Analisis K-means
19 data("mtcars")
20 mtcars$mpg = NULL
21 km <- kmeans(mtcars,3)
22 km
23
24 #Graficacion de clusters
25 library(factoextra)
26 library(ggplot2)
27 fviz_cluster(km,mtcars)
28
```

## Análisis de Resultados

Para nuestro primer análisis, utilizamos la correlación que nos permite expresar el cambio linear de una variable a otra.



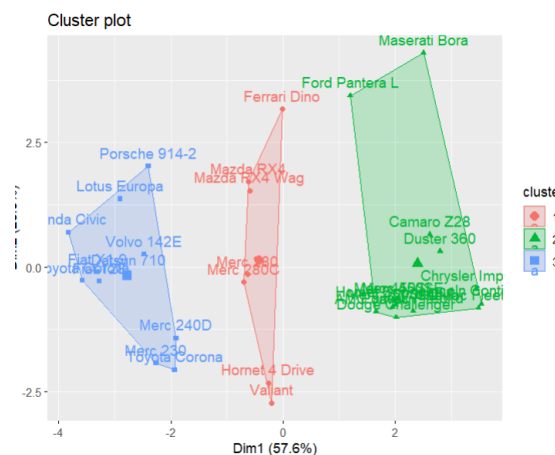
Se realizó una correlación entre el número de cilindros y el peso del motor (en toneladas) observando la gráfica, podemos observar que entre más pesado el motor, más número de cilindros tenga el auto.



Para determinar qué variables tienen una mejor correlación, utilizamos una gráfica de matriz correlación. En la cual los círculos azules grandes, tienen una correlación fuerte positiva, es decir una mejor correlación, el ejemplo de desplazamiento de motor con los cilindros de un coche. Por otra parte, círculo color rojo, tendrán una correlación negativa, como lo es el desplazamiento de la transmisión con el número de cilindros. Por último tenemos círculos pequeños con colores más bajos o “claros”.

Para la segunda parte, se utiliza el análisis de K-means, nos permitirá agrupar elementos considerando las características que tengan en común.

A nuestra consideración, utilizamos 3 grupos para representar la información.



Como se mencionó anteriormente, se escogieron 3 clusters (azul, rojo, verde) y tenemos dos porcentajes de DIM que explican un porcentaje significativo, estos porcentajes se basan en características múltiples.

El seleccionar 3 grupos, nos permite visualizar de una manera más clara, y parece una distribución más equilibrada.

## **Conclusión**

Gracias a nuestra comparación de correlación que hicimos entre cyl( el número de cilindros en un auto) y wt( el peso en toneladas del coche) está muy relacionado ya que al motor al ser más grande por la cantidad de cilindros hace una relación en que el peso del vehiculo tambien incremente.

Gracias a este tipo de análisis podemos ver que las correlaciones pueden ayudar a identificar las características que influyen más en el rendimiento de un automóvil.

Igualmente en nuestro análisis de K-means entregue sería mejor si 2, 3 o 4 clusters pudimos observar que lo mejor fue usar 3 clusters ya que al usar 4 se veía muy juntos los datos para poder visualizarlos correctamente, así que se decidió por una visualización más limpia.

Para finalizar el análisis de K-means seleccionado, se podría segmentar modelos de automóviles según características similares, lo que puede ser útil para clasificar el mercado y así tener mejores estrategias de mercadotecnia y fabricación.