

Análisis Exploratorio de los Datos de mtcars

Melani Forsythe Matos , Ruben Martinez Rojas , Daniela Guerrero Alvarez

June 20, 2024

1 Introducción

El conjunto de datos ‘mtcars’ es un conjunto de datos incluido en el paquete ‘ggplot2’ de R. Contiene información sobre 32 modelos de automóviles, incluyendo características como el consumo de combustible, el número de cilindros, el peso y el tipo de transmisión. En este análisis, exploraremos estos datos para obtener una mejor comprensión de las características de los automóviles.

2 Descripción de los Datos

Comenzamos explorando las dimensiones del data frame ‘mtcars’ utilizando la función ‘dim()’:

```
dim(mtcars)
```

Esto nos indica que el data frame tiene 32 filas (observaciones) y 11 columnas (variables).

3 Estadísticas Descriptivas

Para obtener una visión general de las características de los automóviles, calculamos las siguientes estadísticas descriptivas:

```
# Calcular la media de cada variable  
sapply(mtcars, mean)
```

La media nos proporciona el valor promedio de cada variable en el conjunto de datos.

```
# Calcular la mediana de cada variable  
sapply(mtcars, median)
```

La mediana es el valor central de cada variable, que divide los datos en dos partes iguales.

```
# Calcular el valor mínimo de cada variable
sapply(mtcars, min)
```

El valor mínimo nos indica el valor más bajo de cada variable en el conjunto de datos.

```
# Calcular el valor máximo de cada variable
sapply(mtcars, max)
```

El valor máximo nos indica el valor más alto de cada variable en el conjunto de datos.

Estas estadísticas descriptivas nos permiten entender mejor las características de los automóviles en el conjunto de datos.

4 Varianza y Covarianza

Para obtener una visión general de la dispersión y la relación entre las variables, calculamos la varianza y la covarianza:

```
# Calcular la varianza de cada variable
variance <- sapply(mtcars, var)
```

La varianza mide la dispersión de los datos alrededor de la media. Un valor de varianza más alto indica una mayor dispersión.

```
# Calcular la covarianza de cada variable
covariance <- cov(mtcars)
```

La covarianza mide la relación lineal entre dos variables. Un valor de covarianza positivo indica una relación positiva, mientras que un valor negativo indica una relación negativa.

5 Moda

Además de las estadísticas descriptivas, también calculamos la moda de la variable 'mpg' (millas por galón):

```
# Calcular la moda
mode <- function(x) {
  freq_table <- table(x)
  mode_value <- names(freq_table)[which.max(freq_table)]
  return(as.numeric(mode_value))
}
```

La moda es el valor que aparece con mayor frecuencia en el conjunto de datos. En este caso, la moda de la variable 'mpg' nos indica el valor de consumo de combustible más común entre los automóviles.

6 Visualización de los Datos

Para complementar el análisis estadístico, también creamos varios gráficos para visualizar los datos:

```
# Crear un gráfico de barras para visualizar la moda
```

```
barplot(table(mtcars$mpg), main = "Moda de la variable mpg", xlab = "Valores de mpg", ylab =
```

Este gráfico de barras nos permite visualizar la moda de la variable ‘mpg’.

```
# Crear histogramas para visualizar la distribución
```

```
hist(mtcars$mpg, main = "Distribución de millas por galón", xlab = "Millas por galón", ylab =
```

```
hist(mtcars$cyl, main = "Distribución de número de cilindros", xlab = "Número de cilindros").
```

Estos histogramas nos permiten visualizar la distribución de las variables ‘mpg’ y ‘cyl’.

```
# Crear gráficos de dispersión y boxplot para visualizar relaciones
```

```
plot(mtcars$mpg, mtcars$cyl, main = "Relación entre millas por galón y número de cilindros").
```

```
boxplot(mtcars$mpg ~ mtcars$am, main = "Relación entre millas por galón y tipo de transmisión").
```

```
boxplot(mtcars$mpg ~ mtcars$am + mtcars$cyl, main = "Relación entre millas por galón y tipo
```

Estos gráficos nos permiten visualizar las relaciones entre las variables en el conjunto de datos.

7 Conclusión

En este análisis exploratorio, hemos utilizado una variedad de técnicas estadísticas y de visualización para obtener una mejor comprensión de los datos de ‘mtcars’. Las estadísticas descriptivas, la varianza, la covarianza y la moda nos han proporcionado información valiosa sobre las características de los automóviles. Los gráficos, por su parte, nos han permitido visualizar la distribución y las relaciones entre las variables. Este análisis exploratorio nos prepara para realizar análisis más avanzados y extraer conclusiones más profundas sobre los datos.

Las variables utilizadas en este análisis exploratorio son:

- **mpg**: Millas por galón, que mide el consumo de combustible de los automóviles.
- **cyl**: Número de cilindros, que indica la capacidad del motor de los automóviles.
- **am**: Tipo de transmisión, que indica si el automóvil tiene una transmisión manual o automática.
- **variance**: Varianza de cada variable, que mide la dispersión de los datos alrededor de la media.

- **covariance:** Covarianza entre las variables, que mide la relación lineal entre dos variables.