

Detección de Casos Atípicos Multivariantes en el Conjunto de Datos mtcars

Nazaret Basaldella

2024-11-22

Introducción

“En este informe, se lleva a cabo un análisis multivariante para la detección de casos atípicos en el conjunto de datos `mtcars`. Este dataset incluye características de diferentes modelos de automóviles, como consumo de combustible, peso y potencia, entre otros. El objetivo es identificar observaciones que se alejan significativamente del comportamiento general de los datos, utilizando la distancia de Mahalanobis como herramienta principal. Los casos atípicos son importantes porque pueden representar errores en los datos o vehículos con características únicas que requieren un análisis adicional.” —

Código para la detección de casos atípicos

El siguiente código realiza el análisis paso a paso para detectar los casos atípicos en el conjunto de datos `mtcars`:

```
# Cargar Los datos mtcars
data(mtcars)

# Instalar y cargar Las librerías necesarias

if(!require(MASS)) install.packages("MASS") # Para cálculo de distancias de Mahalanobis
```

```
## Loading required package: MASS
```

```
if(!require(car)) install.packages("car") # Para detección de valores atípicos
```

```
## Loading required package: car
```

```
## Loading required package: carData
```

```
# Inspección inicial de Los datos
```

```
head(mtcars)
```

```
##           mpg cyl disp  hp drat   wt  qsec vs am gear carb
## Mazda RX4      21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag  21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710      22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive  21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant         18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```
# Calcular las distancias de Mahalanobis
# Calcular las distancias utilizando todas las variables
mahalanobis_distances <- mahalanobis(
  mtcars,
  center = colMeans(mtcars),
  cov = cov(mtcars)
)

# Determinar el umbral de detección de atípicos
# Nivel de significancia para detectar valores extremos
alpha <- 0.05
# Grados de libertad (número de variables en mtcars)
df <- ncol(mtcars)
# Valor crítico del chi-cuadrado
chi_squared_threshold <- qchisq(1 - alpha, df)

# Paso 5: Identificar valores atípicos
# Comparar las distancias de Mahalanobis con el valor crítico
outliers <- mahalanobis_distances > chi_squared_threshold

# Mostrar los nombres de los casos atípicos
cat("Casos atípicos detectados:\n")
```

```
## Casos atípicos detectados:
```

```
rownames(mtcars)[outliers]
```

```
## [1] "Merc 230"      "Ford Pantera L"
```

```
# Paso 6: Visualización de los casos atípicos
# Agregar las distancias de Mahalanobis al dataset original
mtcars$mahalanobis <- mahalanobis_distances
# Agregar una columna para marcar los casos atípicos
mtcars$outlier <- ifelse(outliers, "Sí", "No")

# Mostrar el conjunto de datos con las nuevas columnas
print(mtcars)
```

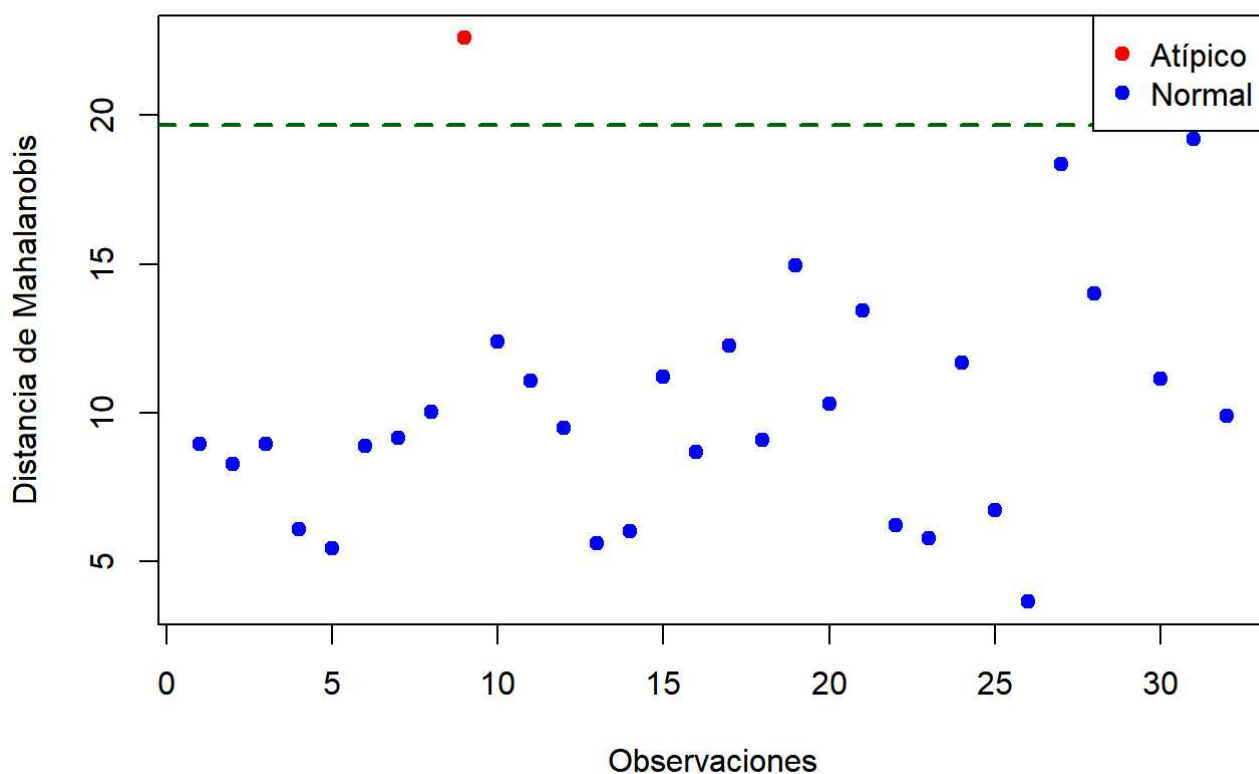
##	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
## Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
## Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
## Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
## Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
## Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
## Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
## Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
## Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
## Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
## Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
## Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
## Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
## Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
## Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
## Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
## Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
## Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
## Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
## Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
## Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
## Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
## Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
## AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
## Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
## Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
## Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
## Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
## Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
## Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
## Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
## Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
## Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2
##	mahalanobis outlier										
## Mazda RX4	8.946673	No									
## Mazda RX4 Wag	8.287933	No									
## Datsun 710	8.937150	No									
## Hornet 4 Drive	6.096726	No									
## Hornet Sportabout	5.429061	No									
## Valiant	8.877558	No									
## Duster 360	9.136276	No									
## Merc 240D	10.030345	No									
## Merc 230	22.593116	Sí									
## Merc 280	12.393107	No									
## Merc 280C	11.058878	No									
## Merc 450SE	9.476126	No									
## Merc 450SL	5.594527	No									
## Merc 450SLC	6.026462	No									
## Cadillac Fleetwood	11.201310	No									
## Lincoln Continental	8.672093	No									
## Chrysler Imperial	12.257618	No									
## Fiat 128	9.078630	No									
## Honda Civic	14.954377	No									
## Toyota Corolla	10.296463	No									
## Toyota Corona	13.432391	No									

## Dodge Challenger	6.227235	No
## AMC Javelin	5.786691	No
## Camaro Z28	11.681526	No
## Pontiac Firebird	6.718085	No
## Fiat X1-9	3.645789	No
## Porsche 914-2	18.356164	No
## Lotus Europa	14.000669	No
## Ford Pantera L	21.573003	Sí
## Ferrari Dino	11.152850	No
## Maserati Bora	19.192384	No
## Volvo 142E	9.888781	No

#Gráfico de las distancias de Mahalanobis

```
plot(mahalanobis_distances,  
     main = "Distancias de Mahalanobis",  
     xlab = "Observaciones",  
     ylab = "Distancia de Mahalanobis",  
     pch = 19, col = ifelse(outliers, "red", "blue"))  
abline(h = chi_squared_threshold, col = "darkgreen", lwd = 2, lty = 2)  
legend("topright", legend = c("Atípico", "Normal"), col = c("red", "blue"), pch = 19)
```

Distancias de Mahalanobis



#Los casos detectados como atípicos son los siguientes:

```
rownames(mtcars)[outliers]
```

```
## [1] "Merc 230"      "Ford Pantera L"
```

#Conclusión En el análisis de detección de casos atípicos multivariantes en el conjunto de datos mtcars, se identificaron dos observaciones como casos atípicos: Merc 230 y Ford Pantera L. Estas observaciones presentaron distancias de Mahalanobis significativamente mayores al umbral crítico, lo que indica que poseen combinaciones de características que las diferencian del resto de los datos.

El gráfico generado confirma esta detección, destacando los casos atípicos en color rojo por encima de la línea verde punteada que representa el valor crítico. Estas diferencias podrían estar relacionadas con características únicas en variables como el peso, la potencia o el consumo de combustible.

Se recomienda realizar un análisis detallado para validar estos casos y determinar su relevancia en función del objetivo del estudio. Dependiendo del contexto, podrían ser excluidos para evitar sesgos o considerados para identificar patrones únicos en el conjunto de datos. Este análisis refuerza la utilidad de la distancia de Mahalanobis para detectar valores atípicos multivariantes de manera eficiente.