

Mahalanobis

Medel Colorado Yoselin Merari

2022-05-22

Distancia de Mahalanobis

Diseñar un ejercicio utilizando la distancia de Mahalanobis.

Incluye:

- 1.- Planteamiento del problema.
- 2.- Simular los datos o utilizar una matriz precargada en R.
- 3.- Dar tu interpretacion.

La Distancia de Mahalanobis es una medida de distancia, su utilidad radica en que es una forma de determinar la similitud entre dos variables aleatorias multidimensionales.

Para hacer este ejercicio se utilizó una matriz precargada en R, usando la base de datos fiel que trata sobre las erupciones y espera de volcanes.

Se instala el paquete

```
install.packages("datos")  
library(datos)
```

Se crea una data.frame

```
Datos<- data.frame(datos :: fiel)
```

Se analizan los datos

```
dim(fiel)  
## [1] 272  2
```

```
str(fiel)
```

```
## 'data.frame': 272 obs. of 2 variables:
## $ erupciones: num 3.6 1.8 3.33 2.28 4.53 ...
## $ espera : num 79 54 74 62 85 55 88 85 51 85 ...
```

```
summary(fiel)
```

```
## erupciones espera
## Min. :1.600 Min. :43.0
## 1st Qu.:2.163 1st Qu.:58.0
## Median :4.000 Median :76.0
## Mean :3.488 Mean :70.9
## 3rd Qu.:4.454 3rd Qu.:82.0
## Max. :5.100 Max. :96.0
```

Cálculo de distancia

Determinar el número de outlier que queremos encontrar

```
num.outliers <-2
```

Ordenar los datos de mayor a menor distancia, según la métrica de Mahalanobis.

```
mah.ordenacion <- order(mahalanobis(fiel , colMeans(fiel), cov(fiel)), decreasing=TRUE)
mah.ordenacion
```

```
## [1] 158 197 58 76 265 46 161 203 17 211 160 151 242 95 8 249 269 70
## [19] 66 127 51 131 69 115 170 267 149 218 111 193 135 188 65 271 89 206
## [37] 92 178 26 80 47 144 44 75 106 45 255 119 14 22 39 270 117 177
## [55] 254 235 134 103 37 90 63 94 6 19 234 148 25 263 121 209 171 213
## [73] 208 192 261 42 55 184 199 223 93 221 77 179 272 130 102 21 146 108
## [91] 2 12 38 54 99 166 150 159 137 50 162 96 52 185 9 122 68 11
## [109] 53 237 40 233 181 204 217 100 224 169 133 16 236 163 200 36 201 240
## [127] 83 110 153 173 182 72 124 231 113 27 168 191 139 31 59 120 48 259
## [145] 56 232 125 219 86 246 250 205 23 7 138 243 4 64 187 212 172 247
## [163] 251 91 194 126 78 167 109 190 142 1 156 129 18 266 147 32 61 10
## [181] 230 15 116 107 112 215 118 245 5 97 49 154 74 71 132 229 73 145
## [199] 256 43 186 262 62 88 207 198 140 183 264 101 84 180 143 104 210 30
## [217] 157 252 258 189 128 114 85 165 257 105 222 136 248 3 202 82 216 238
## [235] 268 241 175 176 41 123 60 141 196 81 34 228 260 220 239 20 164 13
## [253] 29 33 226 67 244 24 28 227 225 152 195 57 87 79 35 98 214 174
## [271] 253 155
```

Generar un vector booleano los dos valores más alejados segun la distancia Mahalanobis.

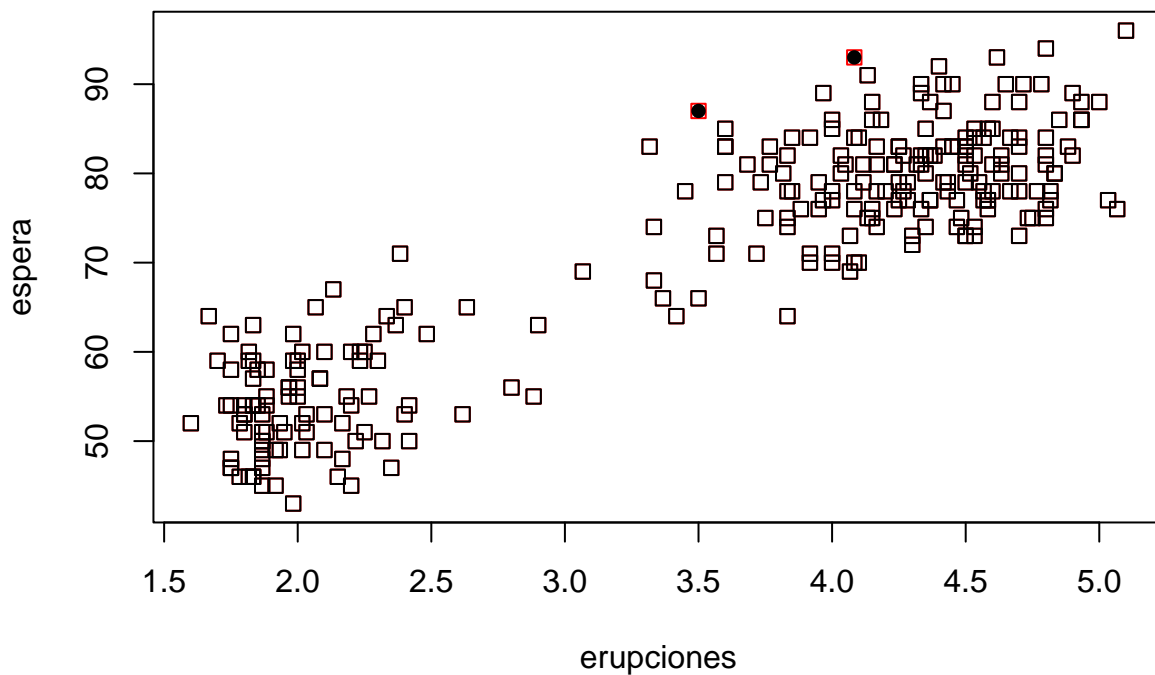
```
outlier2 <- rep(FALSE , nrow(fiel))  
outlier2[mah.ordenacion[1:num.outliers]] <- TRUE
```

Resaltar con un punto relleno los 2 valores outliers.

```
colorear.outlier <- outlier2 * 16
```

Visualizar el gráfico con los datos destacando sus outlier.

```
plot(fiel , pch=0, col="red")  
points(fiel, pch=colorear.outlier)
```



Entre la variable erupciones y espera no existen valores atípicos, existe una buena relación entre las variables.