

# Recuperativa Unidad II

J. Patricio Parada G.

25/08/2020

## Índice

<b>1. Paro cardíaco</b>	<b>1</b>
<b>2. El Dataset</b>	<b>1</b>
2.1. Columnas . . . . .	1
2.2. Estructura del dataset . . . . .	2
2.3. Tipo de datos . . . . .	3
2.4. Resumen estadístico . . . . .	4

## 1. Paro cardíaco

Comúnmente llamado ataque cardíaco, el paro cardíaco es una condición riesgosa y virtualmente mortífera que pone fin a millones de vidas al año. Es una de las causas de muerte más frecuentes en humanos y se debe a variados factores; puede ser a consecuencia del estilo de vida llevado o debido a otras afecciones o enfermedades.

El conjunto de datos anexo presenta 12 factores que eventualmente proporcionan información respecto a si un paciente es candidato a sufrir un ataque cardíaco.

## 2. El Dataset

El conjunto de datos adjunto corresponde a

```
data <- read.csv("CRP_dataset.csv")
```

### 2.1. Columnas

Las columnas (variables) que conforman el conjunto de datos corresponden a

```
colnames(data)
```

```
## [1] "Age"                "Gender"
## [3] "Chain_smoker"       "Consumes_other_tobacco_products"
## [5] "HighBP"            "Obese"
## [7] "Diabetes"           "Metabolic_syndrome"
## [9] "Use_of_stimulant_drugs" "Family_history"
## [11] "History_of_preeclampsia" "CABG_history"
## [13] "Respiratory_illness" "UnderRisk"
```

donde:

- Age: edad

- **Gender**: sexo del paciente. 1 para masculino, 2 para femenino.
- **Chain\_smoker**: fumador. 0 no fumador, 1 fumador.
- **Consumes\_other\_tobacco\_products**: consumidor de otros productos derivados del tabaco. 0 no consumidor, 1 consumidor.
- **HighBP**: hipertensión. 0 no hipertenso, 1 hipertenso.
- **Obese**: obesidad. 0 sin obesidad, 1 obeso.
- **Diabetes**: diabetes, 0 sin diabetes, 1 con diabetes.
- **Metabolic\_syndrome**: síndrome metabólico. 0 no tiene, 1 paciente con síndrome.
- **Use\_of\_stimulant\_drugs**: uso de drogas estimulantes. 0 no consumidor, 1 consumidor.
- **Family\_history**: historial familiar de paro cardíaco. 0 no tiene historial, 1 tiene historial.
- **History\_of\_preeclampsia**: historial de preeclampsia. 0 sin historial, 1 con historial.
- **CABG\_history**: historial de cirugía de bypass de arteria coronaria. 0 sin historial, 1 con historial.
- **Respiratory\_illness**: enfermedad respiratoria. 0 sin enfermedades respiratorias, 1 posee enfermedades respiratorias.
- **UnderRisk**: paciente bajo riesgo. **yes**: sí, **no**: no.

## 2.2. Estructura del dataset

La estructura del conjunto corresponde a

```
str(data)
```

```
## 'data.frame':   889 obs. of  14 variables:
## $ Age           : int  48 69 53 52 48 58 42 43 41 54 ...
## $ Gender        : int  1 1 1 1 1 2 1 2 1 1 ...
## $ Chain_smoker   : int  1 0 3 0 0 0 0 0 0 0 ...
## $ Consumes_other_tobacco_products: int  1 1 1 1 0 1 1 1 1 1 ...
## $ HighBP        : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Obese         : int  1 1 1 1 0 1 0 1 1 0 ...
## $ Diabetes      : int  0 0 3 0 0 0 0 0 0 0 ...
## $ Metabolic_syndrome : int  0 0 3 0 1 0 0 0 0 0 ...
## $ Use_of_stimulant_drugs : int  0 0 0 0 1 0 1 0 0 1 ...
## $ Family_history : int  1 1 1 1 0 1 1 1 1 1 ...
## $ History_of_preeclampsia : int  0 0 0 0 0 0 0 3 0 0 ...
## $ CABG_history   : int  0 0 0 0 0 0 3 0 0 0 ...
## $ Respiratory_illness : int  0 0 0 0 0 0 0 0 0 0 ...
## $ UnderRisk      : chr  "no" "no" "no" "no" ...
```

de donde se puede observar que son 14 parámetros y 889 observaciones.

## 2.3. Tipo de datos

De acuerdo a lo observado en el conjunto de datos y lo descrito a partir de sus columnas, la totalidad de las variables serán consideradas como cualitativas. Del mismo modo, se procede a filtrar los datos para saltar las incoherencias:

```
data$Gender[!(data$Gender == 1 | data$Gender == 2)] <- NA
data$Chain_smoker[!(data$Chain_smoker == 0 | data$Chain_smoker == 1)] <- NA
data$Consumes_other_tobacco_products[
  !(data$Consumes_other_tobacco_products == 0
    | data$Consumes_other_tobacco_products == 1)] <- NA
data$HighBP[!(data$HighBP == 0 | data$HighBP == 1)] <- NA
data$Obese[!(data$Obese == 0 | data$Obese == 1)] <- NA
data$Diabetes[!(data$Diabetes == 0 | data$Diabetes == 1)] <- NA
data$Metabolic_syndrome[
  !(data$Metabolic_syndrome == 0
    | data$Metabolic_syndrome == 1)] <- NA
data$Use_of_stimulant_drugs[
  !(data$Use_of_stimulant_drugs == 0
    | data$Use_of_stimulant_drugs == 1)] <- NA
data$Family_history[
  !(data$Family_history == 0 | data$Family_history == 1)] <- NA
data$History_of_preeclampsia[
  !(data$History_of_preeclampsia == 0
    | data$History_of_preeclampsia == 1)] <- NA
data$CABG_history[!(data$CABG_history == 0 | data$CABG_history == 1)] <- NA
data$Respiratory_illness[
  !(data$Respiratory_illness == 0 | data$Respiratory_illness == 1)] <- NA
```

Para facilitar la legibilidad de los datos, se cambia el tipo de los parámetros a booleano.

```
data$Age <- as.factor(data$Age)
data$Gender <- as.factor(data$Gender)
data$Chain_smoker <- as.logical(data$Chain_smoker)
data$Consumes_other_tobacco_products <- as.logical(data$Consumes_other_tobacco_products)
data$HighBP <- as.logical(data$HighBP)
data$Obese <- as.logical(data$Obese)
data$Diabetes <- as.logical(data$Diabetes)
data$Metabolic_syndrome <- as.logical(data$Metabolic_syndrome)
data$Use_of_stimulant_drugs <- as.logical(data$Use_of_stimulant_drugs)
data$Family_history <- as.logical(data$Family_history)
data$History_of_preeclampsia <- as.logical(data$History_of_preeclampsia)
data$CABG_history <- as.logical(data$CABG_history)
data$Respiratory_illness <- as.logical(data$Respiratory_illness)
```

Para el caso particular de los valores de la columna UnderRisk, se hará uso del paquete `batman`.

```
if(!("batman" %in% rownames(installed.packages()))){
  install.packages("batman")
}
library("batman")
```

Entonces, el filtrado de los datos de UnderRisk se realiza mediante

```
data$UnderRisk <- to_logical(data$UnderRisk)
```

## 2.4. Resumen estadístico

El resumen estadístico después del filtrado corresponde a

```
summary(data)
```

```
##      Age      Gender Chain_smoker Consumes_other_tobacco_products
## 50      : 40      1      :607      Mode :logical      Mode :logical
## 54      : 40      2      :256      FALSE:779      FALSE:144
## 53      : 38      NA's: 26      TRUE :107      TRUE :740
## 46      : 36      NA's :3      NA's :5
## 57      : 36
## 41      : 35
## (Other):664
##      HighBP      Obese      Diabetes      Metabolic_syndrome
## Mode :logical      Mode :logical      Mode :logical      Mode :logical
## FALSE:812      FALSE:72      FALSE:839      FALSE:850
## TRUE :77      TRUE :812      TRUE :49      TRUE :38
##      NA's :5      NA's :1      NA's :1
##
##
##
## Use_of_stimulant_drugs Family_history History_of_preeclampsia CABG_history
## Mode :logical      Mode :logical      Mode :logical      Mode :logical
## FALSE:812      FALSE:66      FALSE:871      FALSE:868
## TRUE :73      TRUE :823      TRUE :16      TRUE :19
## NA's :4      NA's :2      NA's :2
##
##
##
## Respiratory_illness UnderRisk
## Mode :logical      Mode :logical
## FALSE:860      FALSE:698
## TRUE :29      TRUE :190
##      NA's :1
##
##
##
```