

Reporte de investigación

Fátima Pérez

2024-06-25

Introducción

En este trabajo se recolectarán datos biológicos como, estatura, peso, medidas de pie derecho e izquierdo así como tipo de pie, plano, cavo y normal. Esta recopilación de datos se realizó en la escuela primaria D. Rafael Lucio en alumnos de quinto grado de primaria donde se debía recabar información para conocer que alumnos están dispuestos para participar en pruebas de atletismo.

Metodología

```
install.packages("readxl")
```

```
library(readxl)
```

```
atletismo<-read.csv("atletismo.csv")
```

Exploración de la matriz

1.- Dimension de la matriz

```
dim(atletismo)
```

```
## [1] 20 7
```

2.- Nombre de las columnas

```
colnames(atletismo)
```

```
## [1] "id"           "genero"       "peso"         "estatura"
## [5] "pie"         "pie_derecho"  "pie_izquierdo"
```

3.- Tipo de variable ()

```
str(atletismo)
```

```
## 'data.frame': 20 obs. of 7 variables:
## $ id : chr "i1" "i2" "i3" "i4" ...
## $ genero : chr "M" "F" "F" "M" ...
## $ peso : int 30 40 30 60 26 30 30 41 33 40 ...
## $ estatura : num 1.34 1.5 1.4 1.34 1.38 1.34 1.41 1.4 1.26 1.46 ...
## $ pie : chr "Normal" "Normal" "Normal" "Plano" ...
## $ pie_derecho : num 20.4 21.8 20.6 21.1 19.9 19.5 21.1 20.9 NA 21 ...
## $ pie_izquierdo: num NA NA NA 20.8 21 20.2 NA 20.6 19.1 NA ...
```

4.- Datos perdidos

```
anyNA(atletismo)
```

```
## [1] TRUE
```

5.- Descripción de una variable con su categoría

```
atletismo$genero
```

```
## [1] "M" "F" "F" "M" "F" "F" "M" "M" "F" "F" "F" "M" "F" "M" "F" "F" "F" "M" "F"
## [20] "F"
```

Metodología para remplazo de datos perdidos

```
install.packages("dplyr")
```

```
library(dplyr)
```

1.- Conteo de datos perdidos

```
sapply(atletismo,function(x) sum(is.na(x)))
```

```
##           id           genero           peso           estatura           pie
##           0             0             0             0             0
##  pie_derecho pie_izquierdo
##           6             11
```

2.- Preparación de la matriz

```
atlet <- atletismo %>%
  mutate(genero = as.factor(genero)) %>%
  mutate(pie = as.factor(pie)) %>%
  mutate(peso = as.numeric(peso)) %>%
  mutate(estatura = as.numeric(estatura)) %>%
  mutate(pie_derecho = as.numeric(pie_derecho)) %>%
  mutate(pie_izquierdo = as.numeric(pie_izquierdo))
```

```
str(atlet)
```

```
## 'data.frame':   20 obs. of  7 variables:
## $ id           : chr  "i1" "i2" "i3" "i4" ...
## $ genero       : Factor w/ 2 levels "F","M": 2 1 1 2 1 1 2 2 1 1 ...
## $ peso         : num  30 40 30 60 26 30 30 41 33 40 ...
## $ estatura     : num  1.34 1.5 1.4 1.34 1.38 1.34 1.41 1.4 1.26 1.46 ...
## $ pie          : Factor w/ 2 levels "Normal","Plano": 1 1 1 2 1 1 1 1 1 1 ...
## $ pie_derecho  : num  20.4 21.8 20.6 21.1 19.9 19.5 21.1 20.9 NA 21 ...
## $ pie_izquierdo: num  NA NA NA 20.8 21 20.2 NA 20.6 19.1 NA ...
```

3.- Imputación de datos.

```
install.packages("mice")
```

```
library(mice)
```

```
init = mice(atlet, maxit=5)
```

```
##
## iter imp variable
## 1 1 pie_derecho pie_izquierdo
## 1 2 pie_derecho pie_izquierdo
```

```
## 1 3 pie_derecho pie_izquierdo
## 1 4 pie_derecho pie_izquierdo
## 1 5 pie_derecho pie_izquierdo
## 2 1 pie_derecho pie_izquierdo
## 2 2 pie_derecho pie_izquierdo
## 2 3 pie_derecho pie_izquierdo
## 2 4 pie_derecho pie_izquierdo
## 2 5 pie_derecho pie_izquierdo
## 3 1 pie_derecho pie_izquierdo
## 3 2 pie_derecho pie_izquierdo
## 3 3 pie_derecho pie_izquierdo
## 3 4 pie_derecho pie_izquierdo
## 3 5 pie_derecho pie_izquierdo
## 4 1 pie_derecho pie_izquierdo
## 4 2 pie_derecho pie_izquierdo
## 4 3 pie_derecho pie_izquierdo
## 4 4 pie_derecho pie_izquierdo
## 4 5 pie_derecho pie_izquierdo
## 5 1 pie_derecho pie_izquierdo
## 5 2 pie_derecho pie_izquierdo
## 5 3 pie_derecho pie_izquierdo
## 5 4 pie_derecho pie_izquierdo
## 5 5 pie_derecho pie_izquierdo
```

```
## Warning: Number of logged events: 1
```

```
meth = init$method
predM = init$predictorMatrix
```

```
anyNA(atlet)
```

```
## [1] TRUE
```

```
meth[c("pie_derecho")]="norm"
meth[c("pie_izquierdo")]="norm"
```

```
imputed = mice(atlet, method=meth, predictorMatrix=predM, m=5)
```

```
##
## iter imp variable
## 1 1 pie_derecho pie_izquierdo
## 1 2 pie_derecho pie_izquierdo
## 1 3 pie_derecho pie_izquierdo
## 1 4 pie_derecho pie_izquierdo
## 1 5 pie_derecho pie_izquierdo
## 2 1 pie_derecho pie_izquierdo
## 2 2 pie_derecho pie_izquierdo
## 2 3 pie_derecho pie_izquierdo
## 2 4 pie_derecho pie_izquierdo
## 2 5 pie_derecho pie_izquierdo
## 3 1 pie_derecho pie_izquierdo
## 3 2 pie_derecho pie_izquierdo
## 3 3 pie_derecho pie_izquierdo
## 3 4 pie_derecho pie_izquierdo
## 3 5 pie_derecho pie_izquierdo
## 4 1 pie_derecho pie_izquierdo
```

```
## 4 2 pie_derecho pie_izquierdo
## 4 3 pie_derecho pie_izquierdo
## 4 4 pie_derecho pie_izquierdo
## 4 5 pie_derecho pie_izquierdo
## 5 1 pie_derecho pie_izquierdo
## 5 2 pie_derecho pie_izquierdo
## 5 3 pie_derecho pie_izquierdo
## 5 4 pie_derecho pie_izquierdo
## 5 5 pie_derecho pie_izquierdo
```

```
imputed <- complete(imputed)
```

```
sapply(imputed, function(x) sum(is.na(x)))
```

```
##          id          genero          peso          estatura          pie
##          0           0           0           0           0
##  pie_derecho pie_izquierdo
##          0           0
```

```
str(imputed)
```

```
## 'data.frame': 20 obs. of 7 variables:
## $ id : chr "i1" "i2" "i3" "i4" ...
## $ genero : Factor w/ 2 levels "F","M": 2 1 1 2 1 1 2 2 1 1 ...
## $ peso : num 30 40 30 60 26 30 30 41 33 40 ...
## $ estatura : num 1.34 1.5 1.4 1.34 1.38 1.34 1.41 1.4 1.26 1.46 ...
## $ pie : Factor w/ 2 levels "Normal","Plano": 1 1 1 2 1 1 1 1 1 1 ...
## $ pie_derecho : num 20.4 21.8 20.6 21.1 19.9 ...
## $ pie_izquierdo: num 20.8 21.4 20.6 20.8 21 ...
```

4.- Generación de la nueva matriz con datos imputados

```
atlet_imp <- imputed
```

Estadística descriptiva

```
summary(atlet_imp)
```

Resultados

Conclusiones