

Modelos de estadística clásica: muestreos aleatorios

Armando Ocampo

Dudas de la clase previa

Recodificar NA's

Los NA's son valores faltantes o no disponibles, los cuales suelen representar un problema durante la transformación y procesamiento de los datos. Por lo cual, previo a realizar cualquier análisis o función sobre el dataset es necesario conocer el comportamiento de nuestros datos. Para esto utilizaremos la función `summary()` de la paquetería *base* (Esta paquetería se encuentra instalada por defecto en el lenguaje de programación R, por lo cual no es necesario instalarla).

Generaremos un data frame con valores NA.

```
dummy_data <- data.frame(id=c(2,1,3,4,NA),
                          horas_estudio=c(2, 5, 4, 2, 1),
                          horas_recre=c(4, 2, 4, 3, 1),
                          edad = c(20, 15, 16, 22, 20))

summary(dummy_data)
```

##	id	horas_estudio	horas_recre	edad
## Min.	:1.00	Min. :1.0	Min. :1.0	Min. :15.0
## 1st Qu.	:1.75	1st Qu.:2.0	1st Qu.:2.0	1st Qu.:16.0
## Median	:2.50	Median :2.0	Median :3.0	Median :20.0
## Mean	:2.50	Mean :2.8	Mean :2.8	Mean :18.6
## 3rd Qu.	:3.25	3rd Qu.:4.0	3rd Qu.:4.0	3rd Qu.:20.0
## Max.	:4.00	Max. :5.0	Max. :4.0	Max. :22.0
## NA's	:1			

Además de algunas medidas de estadística descriptiva, esta función detalla si existen NA's en nuestro dataset. Otra manera de conocer si existen o no NA's es mediante la función `is.na()`. Esta generará como resultado TRUE, en el sitio donde encuentre un resultado faltante.

```
is.na(dummy_data)
```

##	id	horas_estudio	horas_recre	edad
## [1,]	FALSE	FALSE	FALSE	FALSE
## [2,]	FALSE	FALSE	FALSE	FALSE
## [3,]	FALSE	FALSE	FALSE	FALSE
## [4,]	FALSE	FALSE	FALSE	FALSE
## [5,]	TRUE	FALSE	FALSE	FALSE

```
# podemos acompañarlo con la función sum(), para saber el total de NA's
sum(is.na(dummy_data))
```

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

*# También, es posible utilizar la función which() para determinar el sitio
en la columna donde se encuentra el NA. En este caso, conocemos que el NA*

```
# se encuentra en la columna id, ahora identificaremos su posición
```

```
which(is.na(dummy_data$id))
```

```
## [1] 5
```