

Muestreo

Hugo Villegas

2025-10-14

Carga de librerías.

```
library(dplyr)
library(sampling)
```

R Muestreo Aleatorio Simple.

Utilizaremos la función `slice_sample` que es de la librería **Sampling** para el muestreo aleatorio simple (muestreo sin reemplazo).

A continuación cargamos los datos..

```
data <- rio::import("Base.xlsx")
dim(data)
```

```
## [1] 70  4
```

```
head(data,5)
```

```
##   id      sexo nivel_estudios estatura_centimetros
## 1  1 Masculino      Superior              179
## 2  2 Femenino      Superior              172
## 3  3 Masculino      Básico               149
## 4  4 Masculino Medio Superior              186
## 5  5 Femenino      Básico               186
```

```
set.seed(109)
muestra_aleat <- slice_sample(.data = data, n=8, replace = FALSE)
muestra_aleat
```

```
##   id      sexo nivel_estudios estatura_centimetros
## 1  8 Femenino      Básico               151
## 2 22 Femenino      Superior              190
## 3 25 Femenino      Básico              159
## 4 10 Masculino Medio Superior              151
## 5 55 Masculino      Superior              190
## 6 67 Masculino      Básico              180
## 7 41 Masculino      Superior              165
## 8 31 Femenino Medio Superior              160
```

Muestreo aleatorio estratificado.

El muestreo aleatorio estratificado es una técnica que se utiliza cuando la población que vas a estudiar no es homogénea, es decir, tiene grupos internos (subpoblaciones) muy diferentes entre sí.

El muestreo aleatorio estratificado es una técnica que se utiliza cuando la población que vas a estudiar no es homogénea, es decir, tiene grupos internos (subpoblaciones) muy diferentes entre sí.

En términos sencillos:

Divides la población en “estratos” (subgrupos). Los elementos dentro de cada estrato deben ser lo más parecidos posible entre sí, y los estratos deben ser lo más diferentes posible entre ellos.

Ejemplo: Si estudias el gasto en tecnología, puedes dividir a la gente por Nivel de Ingreso (Bajo, Medio, Alto).

Tomas una muestra aleatoria simple de CADA estrato. Haces un sorteo justo dentro del grupo de “Ingreso Bajo”, otro sorteo en “Ingreso Medio”, y así sucesivamente.

Aseguras la representación. Esto garantiza que todos los grupos importantes (estratos) estén representados en la muestra final, lo que da resultados más precisos que un muestreo aleatorio simple sobre toda la población.

El muestreo aleatorio estratificado es una técnica que se utiliza cuando la población que vas a estudiar no es homogénea, es decir, tiene grupos internos (subpoblaciones) muy diferentes entre sí.

Aplicación en R

Para aplicar el muestreo estratificado de forma sencilla en R, la mejor herramienta es la librería dplyr (parte del ecosistema tidyverse) con las funciones group_by() y slice_sample().

```
# Instalar y cargar dplyr (si no lo has hecho)
# install.packages("dplyr")
library(dplyr)

# Usaremos el data frame 'iris' como ejemplo
data(iris)

# La columna 'Species' (Especie) será nuestro ESTRATO
# (tiene 3 especies: setosa, versicolor, virginica)
head(table(iris$Species),5)
```

```
##
##      setosa versicolor  virginica
##          50          50          50
```

```
# setosa versicolor virginica
#      50          50          50
```

Muestreo estratificado por cantidad.

Si se quiere seleccionar la misma cantidad en cada estrato.

```
# Queremos 10 filas de CADA especie
n_por_estrato <- 10

set.seed(42) # Para reproducibilidad
```

```
muestra_uniforme <- iris %>%
  group_by(Species) %>%
  slice_sample(n = n_por_estrato) %>%
  ungroup() # Desagrupar al final es una buena práctica

# Verificamos que se seleccionaron 10 de cada uno:
table(muestra_uniforme$Species)
```

```
##
##      setosa versicolor  virginica
##         10          10          10
```

```
# setosa versicolor virginica
#      10          10          10
```

Muestreo estratificado por proporción

Si quieres seleccionar un porcentaje o fracción de cada estrato, de modo que la muestra mantenga la misma proporción que la población (afijación proporcional).

En el caso del dataset iris, las proporciones son iguales (50/50/50), pero el código es útil para cualquier proporción.

```
# Queremos el 20% (frac = 0.20) de las filas de CADA especie
fraccion_por_estrato <- 0.20

set.seed(42) # Para reproducibilidad

muestra_proporcional <- iris %>%
  group_by(Species) %>%
  slice_sample(prop = fraccion_por_estrato) %>%
  ungroup()

# Verificamos los tamaños de la muestra (50 * 0.20 = 10 para cada uno):
table(muestra_proporcional$Species)
```

```
##
##      setosa versicolor  virginica
##         10          10          10
```

```
# setosa versicolor virginica
#      10          10          10
```