

# Muestreo En R

Erika Martinez

November 29, 2015

## 1. SELECCION DE MUESTRAS SIMPLES

```
crime<-data.frame(crimtab)
dim(crime)

## [1] 924    3

#Selección de la muestra
#Tamaño de la muestra
n<-30

muestramia<- sample(1:nrow(crime),size=n,replace=FALSE)
muestramia

## [1] 543 863 617 324 394 30 130 325 667 640 753 127 913 692 289 787 662
## [18] 115 132 96 743 133 638 805 172 678 816 114 43 859

#Selección de muestras simples con dplyr
install.packages("dplyr")

## Installing package into 'C:/Users/Ery/Documents/R/win-library/3.2'
## (as 'lib' is unspecified)
## Error in contrib.url(repos, "source"): trying to use CRAN without
## setting a mirror

library(dplyr)

##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
## filter, lag
##
## The following objects are masked from 'package:base':
##
## intersect, setdiff, setequal, union

#Muestra sin reemplazo
crimemuestramia2<- crime %>%
  sample_n(size=n,replace=FALSE)

head(crimemuestramia2)

##      Var1  Var2 Freq
## 44    9.5 144.78    0
## 903  11.4 195.58    0
## 669  13.2 180.34    3
## 35   12.8 142.24    0
## 458  13.1 167.64    0
## 850  10.3 193.04    0
```

```

#Muestra con pesos
crimemuestramia3<- crime %>%
  sample_n(size=n,weight=Freq)

head(crimemuestramia3)

##      Var1   Var2 Freq
## 447    12 167.64   42
## 398   11.3 165.1   39
## 314   11.3 160.02   24
## 270   11.1 157.48   22
## 489    12 170.18   39
## 570   11.7 175.26    9

#Muestra con una proporción de casos
crimemuestramia4<- crime %>%
  sample_frac(0.05)

head(crimemuestramia4);dim(crimemuestramia4)

##      Var1   Var2 Freq
## 404   11.9 165.1   27
## 124   13.3 147.32    0
## 551    9.8 175.26    0
## 462   13.5 167.64    0
## 189   11.4 152.4    3
## 577   12.4 175.26    7
## [1] 46 3

```

EJEMPLO: Una empresa tiene 189 contables. En una muestra aleatoria de 50 de ellos, el número medio de horas trabajadas en sobretiempo en una semana fue de 9.7 horas con una desviación estándar de 6.2 horas. Halle un intervalo del 95 por ciento de confianza para el número medio de horas trabajadas en sobretiempo en una semana.

```

msa.m=function(N,n,media,desv)
{
  f=n/N
  varmed=(desv^2/n)*(1-f)
  desmed=sqrt(varmed)
  a1=media-1.96*desmed
  b1=media+1.96*desmed
  a2=N*a1
  b2=N*b1
  cat("media: IC = ",a1, "--",b1,"\n")
  cat("total: IC = ",a2, "--",b2,"\n")
}

msa.m(189,50,9.7,6.2)

## media: IC = 8.226198 -- 11.1738
## total: IC = 1554.751 -- 2111.849

```

EJEMPLO: Una agencia bancaria que cuenta con un total de 4800 clientes, los que están clasificados como clientes tipo 1 ó tipo 0. Una muestra aleatoria de 120 clientes: 89 tipo "0" y 31 tipo "1", fue usada para hallar un intervalo de confianza del 95 para la proporción de clientes que fueron denominados "tipo 1".

```
msa.p=function(N,n,exitos)
{ f=n/N
  p=exitos/n ; q=1-p
  varp=(p*q/(n-1))*(1-f)
  desp=sqrt(varp)
  a=p-1.96*desp
  b=p+1.96*desp
  cat("proporcion: IC = ",a, "--",b,"\n")
}

msa.p(4800,120,31)

## proporcion: IC = 0.1806765 -- 0.3359901
```

## SELECCION DE MUESTRAS SISTEMÁTICAS

```
install.packages("devtools")

## Installing package into 'C:/Users/Ery/Documents/R/win-library/3.2'
## (as 'lib' is unspecified)
## Error in contrib.url(repos, "source"): trying to use CRAN without
## setting a mirror

library(devtools)

## WARNING: Rtools is required to build R packages, but is not currently
## installed.
##
## Please download and install Rtools 3.3 from http://cran.r-project.org/bin/windows/Rtools
## and then run find_rtools().

install_github("DFJL/SamplingUtil")

## Downloading GitHub repo DFJL/SamplingUtil@master
## Installing SamplingUtil
## "C:/PROGRA~1/R/R-32~1.2/bin/x64/R" --no-site-file --no-environ --no-save
## --no-restore CMD INSTALL \
## "C:/Users/Ery/AppData/Local/Temp/RtmpeaJvNP/devtoolsaf85b06b72/DFJL-SamplingUtil-e0840
## --library="C:/Users/Ery/Documents/R/win-library/3.2" --install-tests

#Cargar libreria:

library(SamplingUtil)
```

```

muestrasis<- sys.sample(N=nrow(crime),n=30)
muestrasis

## [1] 17 47 77 107 137 167 197 227 257 287 317 347 377 407 437 467 497
## [18] 527 557 587 617 647 677 707 737 767 797 827 857 887

#Asignar los elementos de la muestra al data frame de datos
crimemuestrasis<- crime[muestrasis, ]

head(crimemuestrasis)

##      Var1   Var2 Freq
## 17     11 142.24    0
## 47     9.8 144.78    0
## 77    12.8 144.78    0
## 107   11.6 147.32    0
## 137   10.4 149.86    1
## 167   13.4 149.86    0

```

EJEMPLO: Desde una población de  $N = 12$  hogares, se selecciona una muestra de 4 hogares para investigar acerca de la variable "número de personas que viven en el hogar"

```

msa.m(12,4,4.50,1.290994)

## media: IC = 3.46699 -- 5.53301
## total: IC = 41.60388 -- 66.39612

```

EJEMPLO: Un auditor, examinando un total de 840 facturas pendientes de cobro, de una empresa tomó una muestra aleatoria de 120 facturas. Usando los datos del archivo "muestreo1.xls", mediante muestreo sistemático de 1 en 7

```

msa.m(840,120,131.3674,43.71545)

## media: IC = 124.1259 -- 138.6089
## total: IC = 104265.8 -- 116431.5

```