

# FORMULARIO DE R

## Exploración de datos

| Medidas estadísticas  |   |  |   |
|---|---|--|---|
| # Creación del vector de datos x<br>x <- c(3, 10, 12, 20, 23) |   |  |   |
| # Media<br>mean(x)  | # Mediana<br>median(x)  | # Moda<br>library(modeest)<br>mfv(x)                       | # Percentil<br>quantile(x, prob)              |
| # Rango<br>max(x) - min(x)                                    | # Rango intercuartílico<br>IQR(x)                                     | # Varianza<br>var(x)                                       | # Desviación estándar<br>sd(x)                |
| # Coeficiente de variación<br>sd(x)/abs(mean(x))              | # Coeficiente de asimetría de Fisher<br>library(e1071)<br>skewness(x) | # Coeficiente de curtosis<br>library(e1071)<br>kurtosis(x) | # Resumen rápido<br>summary(x)                |
|   |   |  | # Resumen rápido<br>library(PASWR2)<br>eda(x) |

| Gráficos exploratorios   |  |
|--|--|
| # Gráfico de barras<br>m <- c("A","B","A","C","B")<br>barplot(table(m))  | # Gráfico circular<br>m <- c("A","B","A","C","B")<br>pie(table(m))               |
| # Histograma<br>x <- c(3, 10, 12, 20, 23)<br>hist(x, breaks = "Sturges") | # Gráfico de cajas<br>x <- c(3, 10, 12, 20, 23)<br>boxplot(x, horizontal = TRUE) |

## Asociación entre variables

|   |                                  |  |
|---|----------------------------------|--|
| <b># Diagrama de dispersión</b><br>plot(x, y) | <b># Covarianza</b><br>cov(x, y) | <b># Correlación de Pearson</b><br>cor(x, y) |
|---|----------------------------------|--|

|  |  |
|--|--|
| <b># Tabla de contingencia</b><br>table(x, y)  | <b># Gráfico de barras apilado</b><br>barplot(table(x,y))  |
| <b># Chi-Cuadrado</b><br>chisq.test(x, y)  | <b># Cramer's V</b><br>library(vcd)<br>assocstats(table(x, y))                                       |
| <b># Correlación de Spearman</b><br>cor(x, y, method = "spearman"))<br><br># Si las variables x e y son cualitativas, éstas deben ir en forma numérica:<br>cor(as.numeric(x), as.numeric(y), method = "spearman")  |  |
| <b># Correlación de Kendall</b><br>cor(x, y, method = "kendall"))<br><br># Si las variables x e y son cualitativas, éstas deben ir en forma numérica:<br>cor(as.numeric(x), as.numeric(y), method = "kendall")<br><br># Si se tiene ya la información resumida en una tabla de contingencia llamada tabla1:<br>library(DescTools)<br>KendallTauB(tabla1) |  |
| <b># Correlación Biserial Puntual</b><br>cor(continua, as.numeric(nominal))<br><br><b># Correlación Biserial Puntual (método alternativo)</b><br>library(ltm)<br>biserial.cor(continua, nominal, level = 2)  | <b># Correlación Biserial de Rango</b><br>library(rcompanion)<br>wilcoxonRG(table(nominal, ordinal)) |

## INTERVALOS DE CONFIANZA

### 1. Intervalo de Confianza para la Media Poblacional

#### 1.1 Intervalo de Confianza para la Media Poblacional con varianza conocida

| Fórmula  | Código en R  |
|--|--|
| $\left( \bar{X} - Z_{\left(1-\frac{\alpha}{2}\right)} \frac{\sigma}{\sqrt{n}} ; \bar{X} + Z_{\left(1-\frac{\alpha}{2}\right)} \frac{\sigma}{\sqrt{n}} \right)$ | # Datos resumidos<br>library(TeachingDemos)<br>z.test(promedio, stdev = , n = , conf.level = )                     |
|  | # Vector de datos<br>x <- c(3, 10, 12, 20, 23)<br><br>library(TeachingDemos)<br>z.test(x, stdev = , conf.level = ) |
| <b>Tamaño de muestra:</b> $n = \left( Z_{\left(1-\frac{\alpha}{2}\right)} \frac{\sigma}{E} \right)^2$  | library(samplingbook)<br>sample.size.mean(e = , S = , level = )  |

#### 1.2 Intervalo de Confianza para la Media Poblacional con varianza desconocida

| Fórmula  | Código en R  |
|--|--|
| $\left( \bar{X} - t_{\left(1-\frac{\alpha}{2}, n-1\right)} \frac{S}{\sqrt{n}} ; \bar{X} + t_{\left(1-\frac{\alpha}{2}, n-1\right)} \frac{S}{\sqrt{n}} \right)$ | # Vector de datos<br>x <- c(3, 10, 12, 20, 23)<br><br>t.test(x, conf.level = ) |

#### 1.3 IC para la Media Poblacional con varianza desconocida y muestra grande

| Fórmula  |
|--|
| $\left( \bar{X} - Z_{\left(1-\frac{\alpha}{2}\right)} \frac{S}{\sqrt{n}} ; \bar{X} + Z_{\left(1-\frac{\alpha}{2}\right)} \frac{S}{\sqrt{n}} \right)$ |

## 2. Intervalo de Confianza para la Proporción Poblacional

| Fórmula  | Código en R  |
|--|--|
| $\left( \hat{p} - Z_{(1-\frac{\alpha}{2})} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} ; \hat{p} + Z_{(1-\frac{\alpha}{2})} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$ | <pre>library(binom) binom.confint(x = , n = , conf.level = , methods = "asymptotic")</pre> |
| <b>Tamaño de muestra:</b> $n = \hat{p}(1-\hat{p}) \left( \frac{Z_{(1-\frac{\alpha}{2})}}{E} \right)^2$   | <pre>library(samplingbook) sample.size.prop(e = , P = , level = )</pre>                    |

## 3. Intervalo de Confianza para la Varianza Poblacional

| Fórmula  | Código en R   |
|--|---|
| $\left( \frac{(n-1)s^2}{\chi^2_{(1-\frac{\alpha}{2}, n-1)}} ; \frac{(n-1)s^2}{\chi^2_{(\frac{\alpha}{2}, n-1)}} \right)$ | <pre>library(TeachingDemos) sigma.test(x = , n = , conf.level = )</pre> |

### 3.1 Intervalo de Confianza para la Desviación Estándar Poblacional

| Fórmula  | Código en R   |
|--|---|
| $\left( \sqrt{\frac{(n-1)s^2}{\chi^2_{(1-\frac{\alpha}{2}, n-1)}}} ; \sqrt{\frac{(n-1)s^2}{\chi^2_{(\frac{\alpha}{2}, n-1)}}} \right)$ | <pre>library(TeachingDemos) sqrt(sigma.test(x = , n = , conf.level = )\$conf.int)</pre> |

## Manejo de base de datos

| Lectura de datos  |
|---|
| <b># Desde un archivo de valores separados por comas (.csv)</b><br>datos <- read.csv("NombreArchivo.csv", header = TRUE)  |
| <b># Desde un archivo de valores separados por comas (.csv) y que se pueda convertir directamente todas las variables cualitativas en factor</b><br>datos <- read.csv("NombreArchivo.csv", header = TRUE, stringAsFactors = TRUE) |
| <b># Desde un archivo excel (.xlsx)</b><br>library(readxl)<br>datos <- read_excel("NombreArchivo.xlsx")   |

| Estructura de una base de datos  |
|--|
| <b># Estructura de datos</b><br>str(datos)   |
| <b># Variable cualitativa nominal</b><br>datos\$x <- factor(datos\$x)  |
| <b># Variable cualitativa ordinal</b><br>datos\$y <- factor(datos\$y, levels = c("Bajo", "Medio", "Alto"), ordered=TRUE)         |
| <b># Variable cualitativa ordinal (método alternativo)</b><br>datos\$y <- ordered(datos\$y, levels = c("Bajo", "Medio", "Alto")) |

| Tabla de Contingencia con una base de datos   |   |
|---|---|
| table(datos\$Variable1)   | table(datos\$Variable1, datos\$Variable2)   |
| <b># Distribución condicional por filas</b><br>tabla <- table(datos\$Variable1, datos\$Variable2)<br>prop.table(tabla, margin = 1)  | <b># Distribución condicional por columnas</b><br>tabla <- table(datos\$Variable1, datos\$Variable2)<br>prop.table(tabla, margin = 2) |
| <b># Tabla cruzada general (Similar a lo obtenido en Minitab)</b><br><b># Frecuencia, Porcentaje por filas, Porcentaje por columnas, Porcentaje en base al total de datos</b><br>library(gmodels)<br>CrossTable(datos\$Variable1, datos\$Variable2, prop.chisq = FALSE) |   |