Consideremos una población con una variable aleatoria X cuya función de densidad es $f_X(x)$, su valor esperado es $E(X)$ o μ_X y su varianza es $V(X)$ o σ_X^2. Sea $x_1 = \{x_{1,1}, x_{1,2}, ..., x_{1,n}\}$ una muestra aleatoria de tamaño n de dicha variable X, donde $x_{1,1}$ es el valor de la variable del primer individuo u objeto seleccionado, $x_{1,2}$ es el valor de la misma variable para el segundo individuo u objeto, etc. De esta muestra, se puede visualizar la tabla de frecuencias (o histograma); también se pueden calcular algunos estadísticos, tales como la suma de los elementos t_1, la media muestral $\bar{x}_1$ y la varianza muestral s_1^2, entre otros, donde:

$$t_1 = \sum_{i=1}^{n} x_{1,i}, \qquad \bar{x}_1 = \frac{1}{n}\sum_{i=1}^{n} x_{1,i} = \frac{t_1}{n}, \qquad S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_{1,i} - \bar{x}_1\right)^2.$$

$$t_1 = \sum_{i=1}^{n} x_{1,i}, \qquad \bar{x}_1 = \frac{1}{n} \sum_{i=1}^{n} x_{1,i} = \frac{t_1}{n}, \qquad S_1^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(x_{1,i} - \bar{x}_1\right)^2.$$

¿Ha notado alguna relación entre las medias muestrales, las varianzas muestrales y los histogramas con las medias poblacionales, las varianzas poblacionales y las funciones de densidad de la población? ¿Qué pasa si se aumenta el tamaño de la muestra?

Si realizamos otra muestra $x_2 = \{x_{2,1}, x_{2,2}, ..., x_{2,n}\}$, esta tendrá también un histograma, t_2, $\bar{t_2}$, s_2^2, etc. Debido a la aleatoriedad del muestreo, estos histogramas no han de ser idénticos, como tampoco las sumas, ni las medias, ni las varianzas muestrales. De hecho, $(t_1, t_2, ...)$, $(\bar{x}_1, \bar{x}_2, ...)$ y $(s_1^2, s_2^2, ...)$ se pueden considerar valores de las nuevas variables aleatorias T, $\bar{X}$ y S^2, respectivamente.

Como T, $\bar{X}$ y S^2 son variables aleatorias continuas, han de tener:

- Función de densidad: $f_T(x)$, $f_{\bar{X}}(x)$, $f_{S^2}(x)$
- Valor esperado: $E(T)$, $E(\bar{X})$, $E(S^2)$
- Varianza: $V(T)$, $V(\bar{X})$, $V(S^2)$

6.2.2. Distribución de la suma muestral

Sea $X_1, X_2, ..., X_n$ la muestra aleatoria de tamaño n de una población X cuya función de densidad es $f_X(x)$, su valor esperado es $E(X) = \mu_X$ y su varianza es $V(X) = \sigma_X^2$. De la suma de los elementos de la muestra $T = X_1 + X_2+, ..., +X_n$, se puede deducir que:

- Su valor esperado viene dado por:

$$\begin{aligned} E(T) &= E\left(X_1 + X_2+, ..., +X_n\right) = E(X_1) + E(X_2)+, ..., +E(X_n) \\ &= E(X) + E(X)+, ..., +E(X) = n\mu_X \end{aligned}$$

- Su varianza viene dada por:

$$\begin{aligned} V(T) &= V\left(X_1 + X_2+, ..., +X_n\right) = V(X_1) + V(X_2)+, ..., +V(X_n) \\ &= V(X) + V(X)+, ..., +V(X) = n\sigma_X^2 \end{aligned}$$

- Si las X_i están distribuidas normalmente, entonces T también está distribuida normalmente, es decir:

$$X \hookrightarrow N(\mu_X, \sigma_X^2) \qquad \Rightarrow \qquad T \hookrightarrow N(n\mu_X, n\sigma_X^2)$$

6.2.3. Distribución de la media muestral

Continuando con la muestra $X_1, X_2, ..., X_n$ de una población X, y teniendo en cuenta los resultados del apartado anterior, de la media muestral

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{T}{n},$$

se puede deducir que:

– Su valor esperado viene dado por:

$$E(\bar{X}) = E\left(\frac{T}{n}\right) = \frac{1}{n}E(T) = \frac{1}{n}n\mu_X = \mu_X$$

– Su varianza viene dada por:

$$V(\bar{X}) = V\left(\frac{T}{n}\right) = \frac{1}{n^2}V(T) = \frac{1}{n^2}n\sigma_X^2 = \frac{\sigma_X^2}{n}$$

– Si las X_i están distribuidas normalmente, entonces $\bar{X}$ también está distribuida normalmente, es decir:

$$X \hookrightarrow N(\mu_X, \sigma_X^2) \qquad \Rightarrow \qquad \bar{X} \hookrightarrow N\left(\mu_X, \frac{\sigma_X^2}{n}\right)$$

6.2.4. Distribución de la varianza muestral

Similarmente, considerando la muestra $X_1, X_2, ..., X_n$ de una población X, y teniendo en cuenta los resultados de los apartados anteriores, de la varianza muestral

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2,$$

$$E(S^2) = \frac{1}{n-1}\left[n\sigma_X^2 + n\mu_X^2 - n\mu_X^2 - \sigma_X^2\right] = \sigma_X^2$$

– Cuando el tamaño de la muestra tiende a infinito, su varianza tiende a cero:

$$\lim_{n\to\infty} V(S^2) = 0$$

– Si las X_i están distribuidas normalmente, entonces $(n-1)\dfrac{S^2}{\sigma_X^2}$ está distribuida según la *función chi-cuadrado* (χ^2) con $(n-1)$ grados de libertad, es decir:

$$X \hookrightarrow N(\mu_X, \sigma_X^2) \qquad \Rightarrow \qquad (n-1)\frac{S^2}{\sigma_X^2} \hookrightarrow \chi^2_{n-1},$$

Esta distribución, también denominada *distribución de Pearson* o *ji-cuadrada*, es una distribución de probabilidad continua con un parámetro k que representa los grados de libertad de la variable aleatoria y su función de densidad es:

$$f(x;k) = \begin{cases} \dfrac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2} & x > 0, \\ 0 & x \leq 0 \end{cases}$$

donde Γ es la función gamma.

Apply

La función apply nos permite aplicar una función a una matriz, lista o vector que se le pase cómo parámetro.

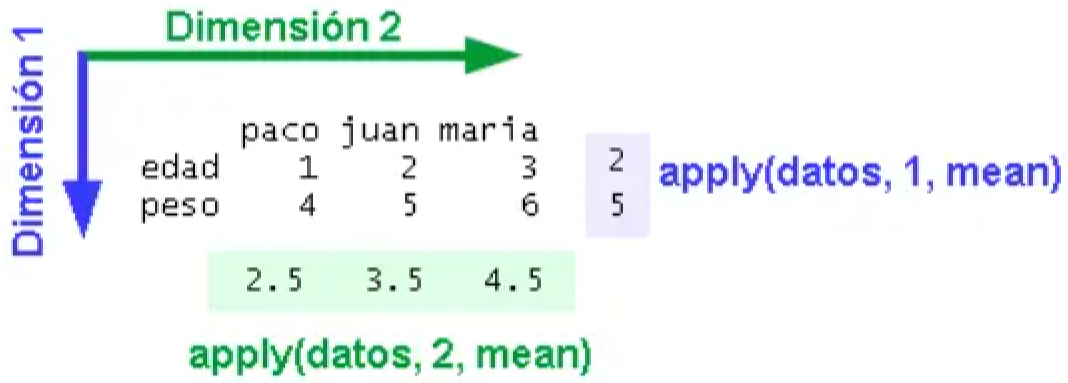
Argumento 1: matriz, lista o vector

Argumento 2: Puede tomar valores 1 o 2

- 1 para operar sobre las filas
- 2 para operar sobre las columnas

Argumento 3: Operador que se aplica sobre filas o columnas, según indique el argumento 2.

Ejemplos de funciones: sum, mean, count,…



Consideramos la variable aleatoria X que tiene la siguiente función de densidad:

$$f(x) = \begin{cases} \dfrac{4}{\pi(1+x^2)} & 0 < x < 1 \\ 0 & resto \end{cases}$$

Su valor esperado (μ) es 0.4412712 y su varianza (σ^2) es 0.07851927. Represente gráficamente dicha función de densidad.

```
x1 <- seq(0,1,0.001)
f <- function(x) {4/(pi*(1+x^2))} f1 <- f(x1)
curve(f,0,1,col="red",lwd=5, main="Función de densidad", xlab="X", ylab="f(x)")
```

```
mean.X <- 0.4412712
var.X <- 0.07851927
```

Simula 50 muestras de tamaño n igual a 4 (utiliza la semilla 321).

```
N <- 50
n <- 4
set.seed(321)
samples <- sample(x1,n*N,replace=TRUE,prob=f1)
samples.X = as.data.frame(matrix(samples, ncol=n))
```

Calcula la suma de las observaciones de cada muestra

```
sum.samples.X = apply(samples.X,1,sum)
```

¿Cuál es el valor de la media de la suma muestral?

```
mean(sum.samples.X)
```

¿Qué relación tiene (aproximada) con la media de la población?

```
Es n veces mayor ##
```

¿Cuál es el valor de la varianza de la suma muestral?

```
var(sum.samples.X)
```

¿Qué relación tiene con la varianza de la población?

```
# Es n veces mayor ##
```

- ¿La distribución de la suma muestral es exactamente a una $N(n\mu, n\sigma^2)$?

```
# Si
# No ##
```

Calcula la media de cada muestra

```
med.samples.X = apply(samples.X,1,mean)
```

¿Cuál es el valor de la media de la media muestral?

```
mean(med.samples.X)
```

¿Qué relación tiene (aproximada) con la media de la población?

```
Son iguales ##
```

¿Cuál es el valor de la varianza de la media muestral?

```
var(med.samples.X)
```

¿Qué relación tiene con la varianza de la población?

```
Es n veces menor ##
```

- ¿La distribución de la media muestral es exactamente una $N(\mu, \sigma^2/n)$?

```
# Si
# No ##
```

Calcula la varianza de de cada muestra

```
var.samples.X = apply(samples.X,1,var)
```

¿Cuál es el valor de la media de la varianza muestral?

```
mean(var.samples.X)
```

¿Qué relación tiene (aproximada) con la varianza de la población?

```
Son iguales ##
```

- ¿La distribución de la varianza muestral es exactamente una χ^2_{n-1}?

```
# Si
# No ##
```

- ¿La distribución del estadístico $(n-1)\dfrac{S^2}{\sigma^2}$ es exactamente una distribución χ^2_{n-1}?

```
# Si
# No ##
```

Si el tamaño de la muestra es 50, -¿Cuál de las siguientes afirmaciones es incorrecta?

- $\mu_{\bar{X}} = \mu_X$ La media de la media muestral es igual a la media de población
- $\bar{X} \hookrightarrow N(\mu, \sigma^2/n)$ La distribucion de la suma es una normal
- $T \hookrightarrow N(n\mu, n\sigma^2)$
- $(n-1)\dfrac{S^2}{\sigma^2} \hookrightarrow \chi^2_{n-1}$ ##

Calcula la probabilidad de que:

- La suma de la muestra esté entre 17 y 24

```
n2 <- 50
pnorm(24, mean=n2*mean.X, sd=sqrt(n2*var.X))-pnorm(17, mean=n2*mean.X, sd=sqrt(n2*var.X))
```

La media de la muestra esté entre 0.39 y 0.41

```
pnorm(0.41, mean=mean.X, sd=sqrt(var.X/n2))-pnorm(0.37, mean=mean.X, sd=sqrt(var.X/n2))
```

La varianza de la muestra esté entre 0.05 y 0.15

```
# No conocemos la distribución de la varianza
```

Este apartado se deja en blanco o se pone que no se conoce la distribución

Los pesos de los hombres adultos de una determinada población se distribuyen normalmente, con una media de 80 kg y una desviación estándar de 15 kg.
Calcula la probabilidad de que:
•La suma del peso de 9 hombres esté entre 700 y 800kg.

```
mean.Y <- 80
var.Y <- 15^2
n <- 9
pnorm(800, mean=n*mean.Y, sd=sqrt(n*var.Y))-pnorm(700, mean=n*mean.Y, sd=sqrt(n*var.Y))
```

La media del peso de 9 hombres esté entre 78 y 80kg

```
pnorm(80, mean=mean.Y, sd=sqrt(var.Y/n))-pnorm(78, mean=mean.Y, sd=sqrt(var.Y/n))
```

La varianza del peso de 9 hombres esté entre 200 y 250

```
pchisq(250*(n-1)/var.Y,df=n-1)-pchisq(200*(n-1)/var.Y,df=n-1)
```

Simula 100 muestras de tamaño n igual a 9 (utiliza la semila 321)

```
N <- 100 # Número de muestras
n <- 9 # tamaño de la muestra
set.seed(321)
samples <- rnorm(N*n, mean=mean.Y, sd=sqrt(var.Y)) # Simulación de N * n eventos
samples.Y = as.data.frame(matrix(samples, ncol=n)) # Organización en un data.frame
```

Haz el histograma de la suma muestral

```
sum.samples.Y = apply(samples.Y,1,sum) hist(sum.samples.Y,prob=T)
```

- ¿La distribución de la suma muestral se puede aproximar a una $N(n\mu, n\sigma^2)$? Compare el histograma y la distribución aproximada.

```
hist(sum.samples.Y,prob=T)
curve(dnorm(x,mean=n*mean.Y,sd=sqrt(n*var.Y)), add=T, lwd=2, col="red")
```

Haz el histograma de la media muestral

```
med.samples.Y = apply(samples.Y,1,mean) hist(med.samples.Y,prob=T)
```

¿La distribución de la media muestral se puede aproximar a una $N(\mu, \sigma^2/n)$? Compare el histograma y la distribución aproximada.

```
hist(med.samples.Y,prob=T)
curve(dnorm(x,mean=mean.Y,sd=sqrt(var.Y/n)), add=T, lwd=2, col="red")
```

Haz el histograma de la varianza muestral

```
var.samples.Y = apply(samples.Y,1,var)
hist(var.samples.Y,prob=T)
```

¿La distribución de la varianza muestral se puede aproximar a una χ^2_{n-1}? Compare el histograma y la distribución aproximada.

```
hist(var.samples.Y,prob=T)
curve(dchisq(x,df=n-1), add=T, lwd=2, col="red")
```

Haz el histograma del estadístico $(n-1)\dfrac{S^2}{\sigma^2}$

```
hist(var.samples.Y*(n-1)/var.Y,prob=T)
```

¿La distribución del estadístico $(n-1)\dfrac{S^2}{\sigma^2}$ se puede aproximar a una χ^2_{n-1}? Compare el histograma y la distribución aproximada.

```
hist(var.samples.Y*(n-1)/var.Y,prob=T)
curve(dchisq(x,df=n-1), add=T, lwd=2, col="red")
```