

## Clase 3 Variables Aleatorias Continuas

OCE 386 - Introducción al análisis de datos con R

Dr. José Gallardo Matus. | Pontificia Universidad Católica de  
Valparaíso | [jose.gallardo@pucv.cl](mailto:jose.gallardo@pucv.cl)

24 August 2021

# PLAN DE LA CLASE

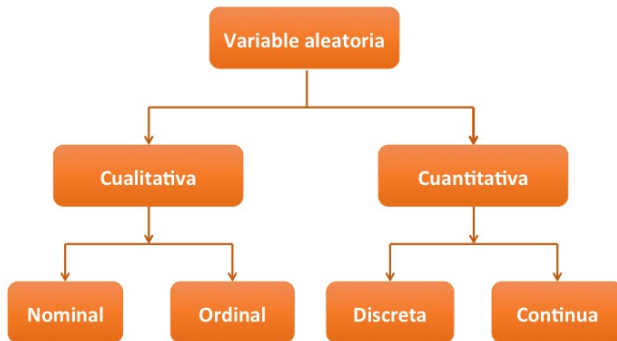
## 1.- Introducción

- ▶ Clasificación de variables aleatorias.
- ▶ Observar una variable cuantitativa continua.
- ▶ Predecir una variable cuantitativa continua.

## 2.- Práctica con R y Rstudio cloud

- ▶ Observa y predice una variable aleatoria continua.
- ▶ Elabora un reporte dinámico en formato pdf.

# TIPOS DE VARIABLES ALEATORIAS



# CASOS ESPECIALES

**1.- Variable aleatoria binaria:** Posee dos resultados posibles; por ejemplo, éxito o fracaso, macho o hembra, sano o enfermo,  $(0,1)$ .

**2.- Variable aleatoria dependiente del tiempo:**

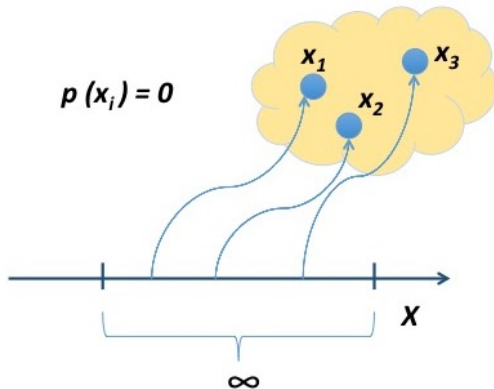
**a) Discreta:** Días a la muerte de un organismo o fallo de un componente en un sistema en un tiempo  $t$ .

**b) Continua:** Señales de sensores ambientales o señales biométricas.

Algunas de estas variables se conocen como **series de tiempo** y en términos estrictos son más bien una *sucesión de variables aleatorias* a través del tiempo.

# VAR. ALEATORIA CUANTITATIVA CONTINUA

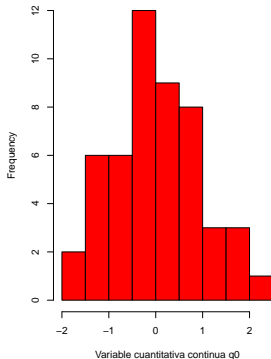
**Definición:** Puede tomar cualquier valor dentro de un intervalo  $(a,b)$ ,  $(a,\text{Inf})$ ,  $(-\text{Inf},b)$ ,  $(-\text{Inf},\text{Inf})$  y la probabilidad que toma cualquier punto es 0, debido a que existe un número infinito de posibilidades.



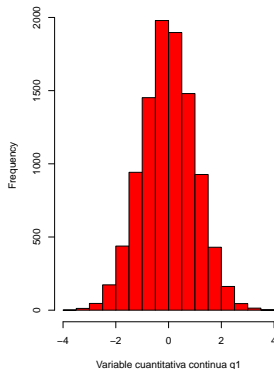
# OBSERVAR UNA VAR. CUANTITATIVA CONTINUA

Podemos observar en un histograma **hist()** que cuando aumenta el **n** muestral se perfila una distribución llamada **normal**.

Histograma n= 50



Histograma n= 10000



# ESTADÍSTICOS DE UNA MUESTRA NORMAL

Los estadísticos más comunes son la media **mean()**, mediana **median()** y la moda **modeest::mlv()**.

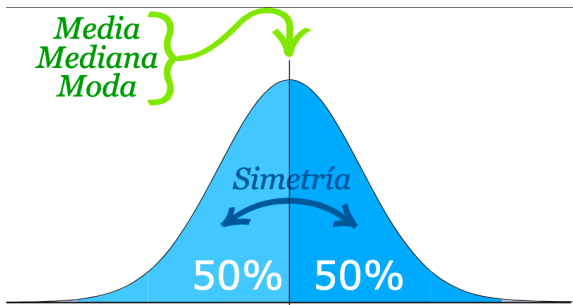


Imagen con Copyright©2020 DisfrutaLasMatematicas.com

# ESTADÍSTICOS DE UNA MUESTRA NORMAL CON R

```
mean(q1)
```

```
## [1] -0.002019208
```

```
median(q1)
```

```
## [1] -0.01122962
```

```
modeest::mlv(q1, method = "meanshift")
```

```
## Registered S3 method overwritten by 'rmutil':
```

```
##   method          from
```

```
##   plot.residuals psych
```

```
## [1] -0.1019375
```

```
## attr(,"iterations")
```

```
## [1] 290
```



# ESTADÍSTICOS DE DISPERSIÓN Y POSICIÓN

Entre los estadísticos de dispersión destacan la desviación estándar **sd()**, el rango **range()** y la varianza **var()**.

Los cuantiles y sus derivados (cuartiles, quintiles, deciles, percentiles) son puntos tomados a intervalos regulares de una función de distribución de una variable aleatoria.

```
sd(q1)
```

```
## [1] 0.9992802
```

```
range(q1)
```

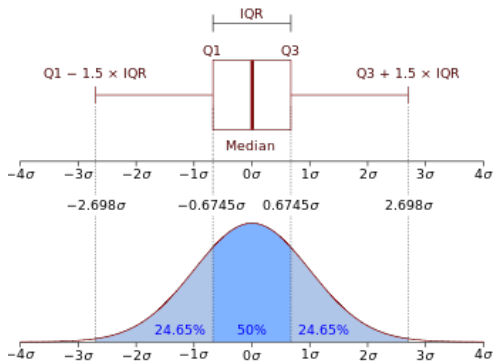
```
## [1] -3.845320 3.847768
```

```
quantile(q1, probs=c(0.05, 0.5, 0.8))
```

```
##           5%           50%           80%  
## -1.64043322 -0.01122962  0.84310278
```

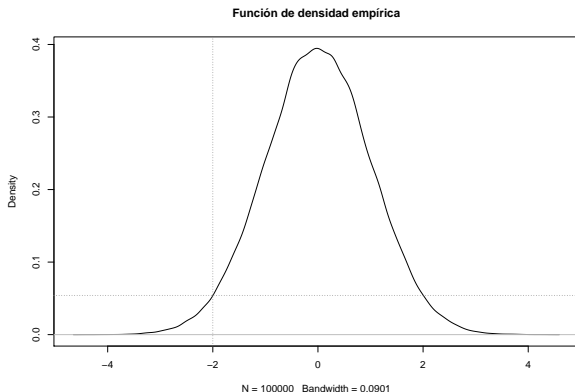
# OBSERVAR UNA VARIABLE CON BOXPLOT

Las gráficas de cajas y bigotes **boxplot()** son muy adecuadas para observar variables aleatorias cuantitativas continuas.



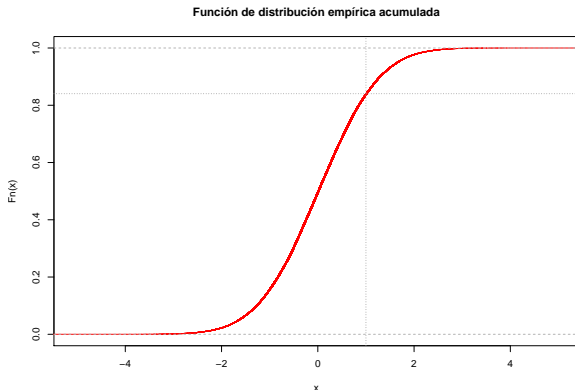
# PREDECIR UNA VAR. CUANTITATIVA CONTINUA

La función de densidad empírica **density()** permite predecir la probabilidad de que la variable aleatoria tome un determinado valor. Ej. Probabilidad de que tome -2 = **dnorm(-2, 0, 1)** = 0.054



## PREDECIR VARIABLES CONTINUAS 2

La función de distribución empírica acumulada **ecdf()** permite predecir la probabilidad de que la variable aleatoria tome un valor menor o igual a un determinado valor. Ej. Probabilidad menor o igual a 1 = **Fn <- ecdf(q2); Fn(1)=0.84035**.



# PRÁCTICA ANÁLISIS DE DATOS

- 1). **Guía de trabajo programación con Rmarkdown disponible en drive.** Ingresa al siguiente link de la *clase 3*
- 2). **La tarea se realiza en Rstudio.cloud.** Ingresa al siguiente proyecto de *Rstudio.Cloud*

# RESUMEN DE LA CLASE

- ▶ Identificamos y clasificamos variables aleatorias.
- ▶ Observamos una variable cuantitativa continua usando histogramas y boxplot.
- ▶ Predecimos el comportamiento de una variable cuantitativa continua usando funciones de densidad y de distribución acumulada.