

CLASE 02 - VARIABLES ALEATORIAS CONTINUAS

Diplomado en Análisis de datos con R para la Acuicultura.

Dra. Angélica Rueda Calderón

Pontificia Universidad Católica de Valparaíso

09 April 2022

PLAN DE LA CLASE

1. Introducción

- ▶ Diferencia entre variable, variable aleatoria, datos y factores.
- ▶ Clasificación de variables aleatorias.
- ▶ Observar y predecir variables cuantitativas continuas.
- ▶ Formato correcto para importar datos a R.

2. Práctica con R y Rstudio cloud

- ▶ Elaborar un script de R e importar datos desde excel.
- ▶ Observar y predecir variable aleatoria con distribución Normal.

CONCEPTOS Y DEFINICIONES

1. **Variable**

Características que se pueden medir u observar en un individuo o en un ambiente: peso, temperatura, sexo, crecimiento, madurez, flotabilidad, rendimiento, sobrevivencia, biomasa cosechada.

2. **Variable aleatoria**

Es un número que representa un resultado de un experimento aleatorio. Depende entonces de función matemática o distribución de probabilidad.

CONCEPTOS Y DEFINICIONES

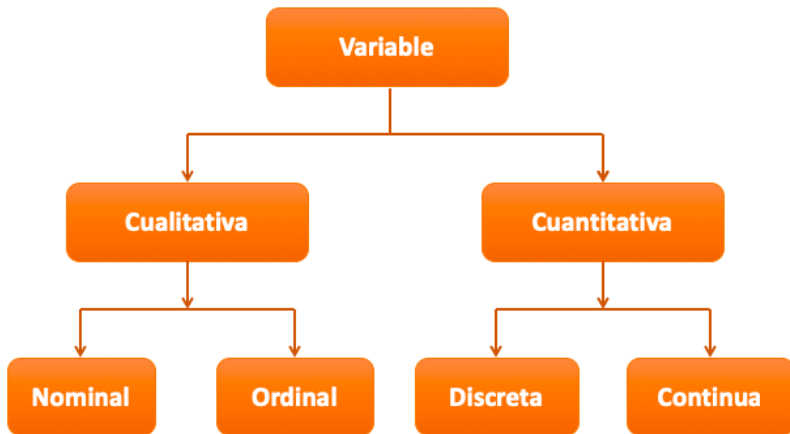
3. Datos u observaciones

Son los valores que puede tomar una variable aleatoria. 6078 gramos, 55 mm, células por mililitro, macho / hembra, 13°C, Maduro /No maduro, Kg wfe/N, vivo/muerto.

4. Factor

Usado para identificar tratamientos de un experimento o variables de clasificación. Se usan como *variables independientes o predictoras*, es decir, tienen un efecto sobre una *variable dependiente o respuesta*. Ej. Sexo (niveles: macho o hembra) tiene un efecto sobre nivel de hormonas.

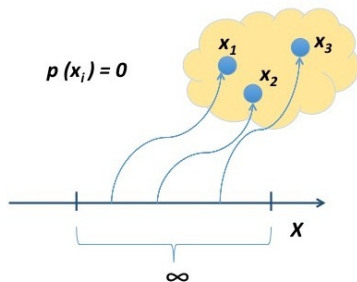
CLASIFICACIÓN DE VARIABLES



VARIABLE ALEATORIA CUANTITATIVA CONTINUA

Definición: Puede tomar cualquier valor dentro de un intervalo (a,b) , (a,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.

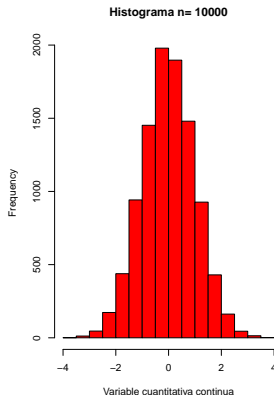
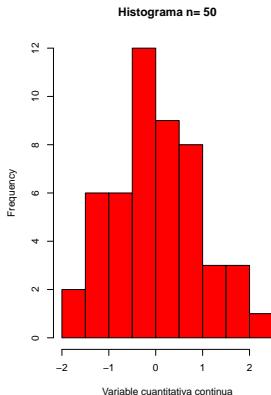
- ▶ Peso del cuerpo (g).
- ▶ Talla (cm).



OBSERVAR VARIABLE CONTINUA CON HISTOGRAMA

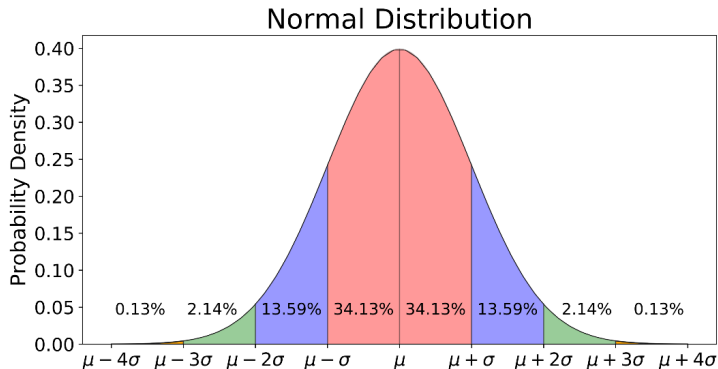
Al observar con un histograma **hist()** notamos que:

1. La frecuencia o probabilidad en un intervalo es distinta de cero.
2. Cuando aumenta el **n** muestral se perfila una distribución llamada **normal**.



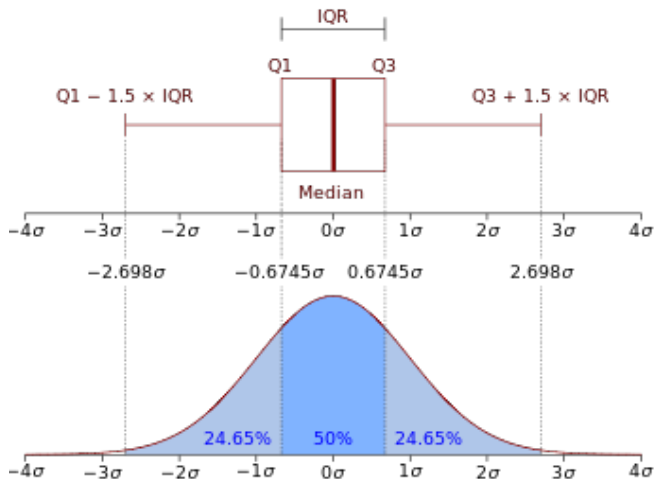
LA DISTRIBUCIÓN NORMAL

$$X \sim \text{Normal}(\mu, \sigma)$$



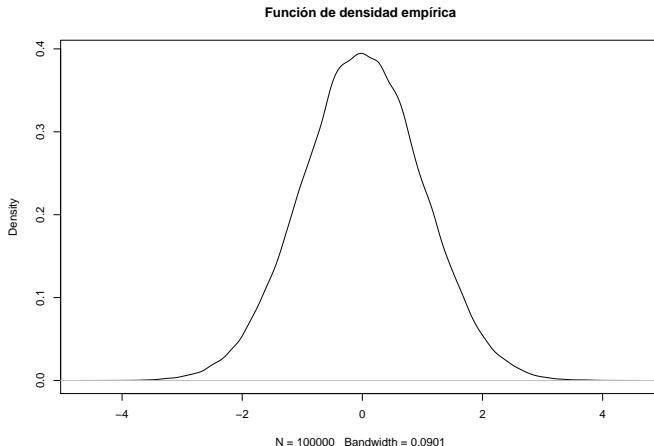
OBSERVAR CON BOXPLOT

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



PREDECIR VARIABLE CONTINUA (V.C.)

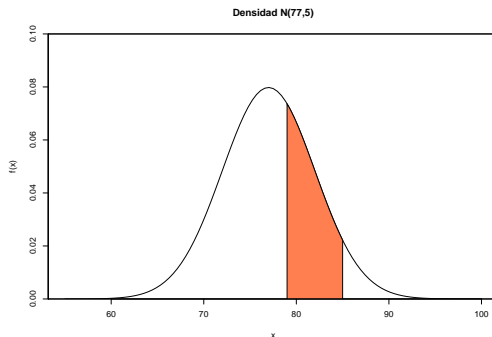
Podemos predecir la probabilidad de que la variable aleatoria tome un determinado valor usando la función de densidad empírica **density()**.



PREDECIR TALLA DE SALMONES

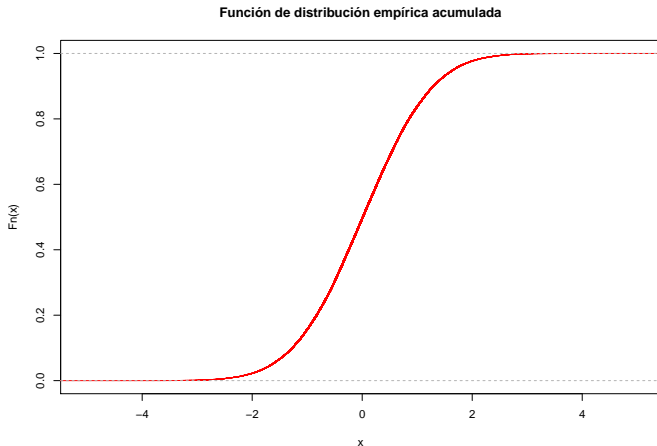
La distribución normal puede describir la distribución de las tallas de salmones a la cosecha. Por ejemplo, usted puede calcular la probabilidad de que un salmón mida entre 79 y 85 cm.

$$P(79 \leq X \leq 85) = 29\%$$



PREDECIR V.C.: DISTRIBUCIÓN ACUMULADA

Podemos predecir la probabilidad de que la variable aleatoria tome un valor menor o igual a un determinado valor, usando la función de distribución empírica acumulada **ecdf()**.



IDENTIFICA CORRECTAMENTE TU VARIABLE

- ▶ Es importante identificar la naturaleza que tiene nuestra variable en estudio, y así evitar errores en los análisis estadísticos que llevemos a cabo.
- ▶ Las variables cuantitativas continuas usualmente tienen una distribución normal.
- ▶ Otras distribuciones asociadas a variables cuantitativas continuas son: t de Student, Exponencial, entre otras.

FORMATO CORRECTO PARA IMPORTAR A R

	A	B	C	D	E	F	G	H
1	Animal	Talla	Peso	Sexo	Nombre de variables			
2	1	72,28747	2820,659	Hembra	Observaciones o datos			
3	2	79,80557	5638,815	Macho				
4	3	62,4055	5263,206	Hembra				
5	4	75,14471	7397,472	Macho				
6	5	85,8213	6742,149	Macho				
7	6	84,65912	6754,632	Hembra				
8	7	72,89857	7674,795	Hembra				
9	8	82,39114	4528,756	Macho				
10	9	74,07506	3430,988	Hembra				
11	10	77,44438	6032,77	Hembra				
12	11	72,02014	4604,848	Hembra				
13	12	80,36905	5669,301	Macho				
14	13	73,11514	4768,415	Macho				
15	14	79,27122	7350,092	Hembra				
16	15	70,60855	6325,635	Hembra				
17	16	85,10738	8526,119	Macho				
18	17	74,39336	3466,113	Hembra				
19	18	83,64237	6798,916	Hembra				

COMO IMPORTAR DATOS A R

Asuntos importantes:

1. Prefiera archivos sin formato como **txt**, **csv**. Si tiene un excel se recomienda transformarlo, particularmente cuando trabaje con miles de filas o columnas.
2. El paquete **readxl** es muy util para importar datos a R. Pero debe tener cuidado con: separador de columnas, decimales y valores faltantes.

```
library(readxl)
salmon<-read_excel("datos.xlsx",
                   sheet = 1, na = "NA")
```

PRÁCTICA VARIABLES ALEATORIAS

Guía de trabajo programación con R en Rstudio.cloud.



0. RUN



1. STUDY



3. SHARE



4. IMPROVE

RESUMEN DE LA CLASE

- ▶ Identificamos y clasificamos variables.
- ▶ Observamos la distribución de una variable cuantitativa continua usando histograma y boxplot.
- ▶ Predecimos el comportamiento de una variable cuantitativa continua con distribución normal usando funciones de densidad y de distribución acumulada.