

CLASE 02 - VARIABLES ALEATORIAS CUANTITATIVAS

Curso Introducción al Análisis de datos con R para la
Acuicultura.

Dra. Angélica Rueda Calderón

Pontificia Universidad Católica de Valparaíso

09 July 2022

PLAN DE LA CLASE

1. Introducción

- ▶ Diferencia entre variable, variable aleatoria, datos y factores.
- ▶ Clasificación de variables aleatorias.
- ▶ Observar variables cuantitativas continuas y discretas.
- ▶ Distribución de variables aleatorias: normal, binomial, bernoulli.
- ▶ 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 variable aleatoria continua y discreta.

CONCEPTOS Y DEFINICIONES

1. **Variable**

Características que se pueden medir u observar en un individuo o en un ambiente: peso, temperatura, oxígeno, 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: normal, binomial, bernoulli.

CONCEPTOS Y DEFINICIONES

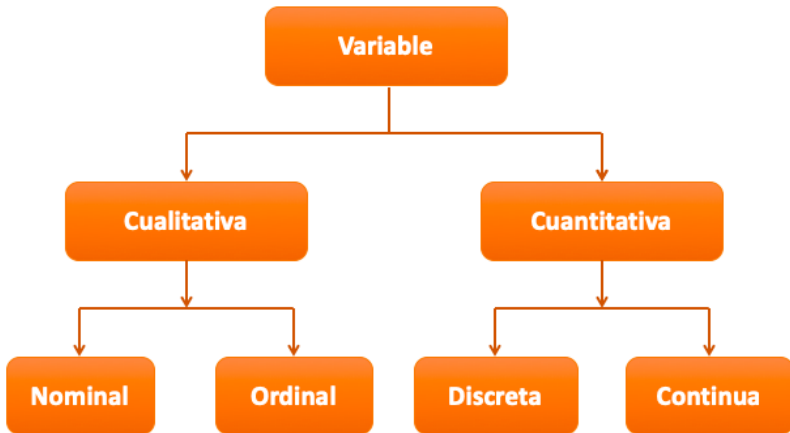
3. Datos u observaciones

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



VARIABLES CUANTITATIVAS

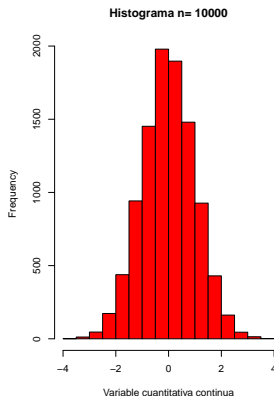
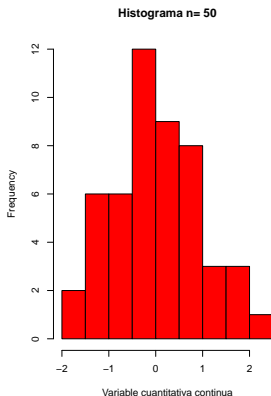
- ▶ **Continua:** Puede tomar cualquier valor dentro de un intervalo (a,b) , (a, Inf) , $(-\text{Inf}, b)$, $(-\text{Inf}, \text{Inf})$ y la probabilidad que toma cualquier punto es 0, debido a que existe un número infinito de posibilidades: **Peso del cuerpo (g), Talla (cm), Temperatura °C**
- ▶ **Discreta:** Son aquellas que presentan un número contable de valores; por ejemplo:
 - a) **Fecundidad medida como número de huevos:** (1, 15, 26, 50, etc.).
 - b) **Número de parásitos** (1, 3, 5, 6, etc.).
 - c) **Número de días a la muerte durante desafío contra patógenos** (1, 2, 3, ..., 40).
 - d) **Número de larvas:** (1, 15, 26, 50, etc.).

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**, pero también **t de Student**, **Exponencial**, entre otras
- ▶ Según sea la variable aleatoria discreta, ésta tendrá una función de distribución de probabilidad asociada que **NO** es normal. Por ejemplo: **Bernoulli**, **Binomial**, **Binomial Negativa**, **Poisson**, entre otras.
- ▶ Las variables aleatorias que son expresadas como proporciones, podrían ser analizadas con la distribución binomial.

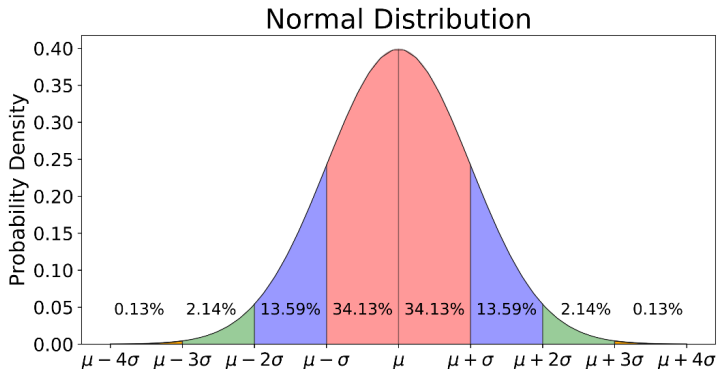
OBSERVAR VARIABLE CONTINUA CON HISTOGRAMA

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



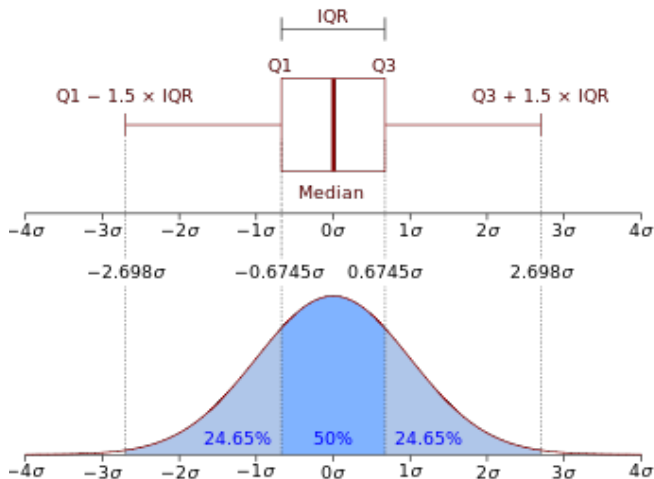
LA DISTRIBUCIÓN NORMAL

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



BOXPLOT: CAJAS Y BIGOTES

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



VARIABLE DISCRETA: DISTRIBUCIÓN BERNOLLI

- ▶ La distribución de Bernoulli, describe un experimento aleatorio que sólo admite dos resultados excluyentes (éxito y fracaso).
- ▶ El éxito es usualmente el evento de interés, a veces considerado como lo “malo”: Muerte, enfermo, patógeno +.
- ▶ La variable aleatoria discreta X asociada a este experimento toma el valor 1 cuando ocurre el suceso éxito con probabilidad $P(X = 1) = p$ y el valor 0 cuando ocurre el suceso fracaso con probabilidad $P(X = 0) = 1 - p$.

EJEMPLO VARIABLE DISTRIBUCIÓN BERNOLLI

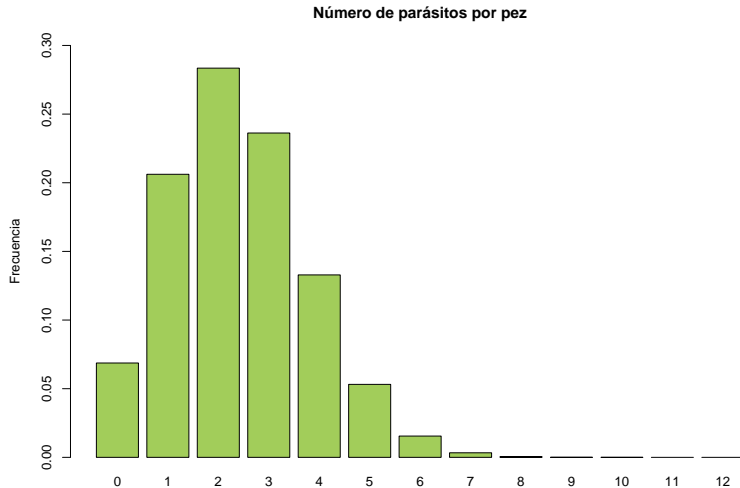
- ▶ Se saca un camarón al azar de una piscina, la probabilidad de que tenga síndrome de la mancha blanca es de 0.65.
- ▶ Sea $X=1$ si el camarón tiene síndrome de la mancha blanca y $X=0$ en el caso de que no tenga síndrome de la mancha blanca.
- ▶ ¿Cuál es la distribución de X ?

	Fracaso	Éxito
x	0	1
$f(x)=P(X=x)$	$1-p$ 0.35	p 0.65

VARIABLE DISCRETA: DISTRIBUCIÓN BINOMIAL

- ▶ Cuando se realizan n pruebas de Bernoulli sucesivas e independientes, la variable aleatoria discreta X se denomina variable binomial.
- ▶ $X =$ “número de veces que ocurre el suceso éxito en n pruebas”.
- ▶ La mayoría de las variables discretas muestran distribución binomial.
- ▶ La mejor manera de representarlas una variable discreta es con gráficas de barra **barplot()**.

EJEMPLO VARIABLE DISTRIBUCIÓN BINOMIAL



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")
```

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 EXPORTAR DATOS A R

Exportar objeto `datos_all` en formato `.txt`, `.csv` y `.xlsx`

```
library(utils)
write.table(x = datos_all, file = "datos_all.txt",
           sep = "\t", row.names = FALSE,
           col.names = TRUE)
```

```
library(utils)
write.csv(x = datos_all, file = "datos_all.csv",
         row.names = FALSE)
```

```
library(xlsx)
write.xlsx(datos_all, "datos_all.xlsx",
          sheetName = "Base_datos",
          col.names = TRUE, row.names = FALSE)
```

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.
- ▶ Observamos la distribución de una variable cuantitativa discreta usando barplot.
- ▶ Reconocemos diferentes distribuciones de probabilidad: normal, bernoulli, binomial.