

# Clase 4 Variables Aleatorias Discretas

## 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)

31 August 2021

# PLAN DE LA CLASE

## 1.- Introducción

- ▶ Diferencia entre variables, datos y factores .
- ▶ Identificar variables aleatorias discretas.
- ▶ Observar y predecir variables aleatorias discretas.
- ▶ Observar y predecir otros tipos de variables.

## 2.- Práctica con R y Rstudio cloud

- ▶ Observa y predice una variable aleatoria continua con distribución Normal.
- ▶ Observa y predice variables aleatorias discretas con distribución Bernoulli o Binomial.
- ▶ Elabora un reporte dinámico en formato pdf.

# DIFERENCIA ENTRE VARIABLES, DATOS Y FACTORES

- ▶ Variables: Características que se pueden medir en un individuo o en un ambiente: peso, longitud, temperatura, oxígeno disuelto, salinidad.
- ▶ Datos u observaciones: Son los valores que puede tomar una variable (25 gramos, 55 mm, 13°C, 10 mg/l, 25psu).
- ▶ Factor: Comunmente usado para identificar a los tratamientos de un experimento o a las variables de clasificación de un estudio. Se usan como *variables independientes o predictoras*, es decir tienen un efecto sobre una *variable respuesta o dependiente*. Por ejemplo: Sexo (niveles: macho o hembra), Ambiente (niveles: Estuario, Río, Oceánico).

# PROBLEMA DATOS, VARIABLES Y FACTORES

Identifique: variables cualitativa, var. continua, var. discreta, datos, factores, niveles de un factor.

Pez	Especie	Sexo	Peso	Nº parásitos
1	A	Hembra	174	0
2	A	Hembra	155	2
3	A	Hembra	131	25
4	B	Macho	163	8
5	B	Macho	103	33
6	B	Hembra	138	15
7	C	Hembra	135	5
8	C	Macho	138	20
9	C	Hembra	135	45

# VARIABLES ALEATORIAS DISCRETAS

Las variables aleatorias discretas son aquellas que presentan un número contable de valores; por ejemplo:

- ▶ **Número de huevos o número de crías de una hembra** (10, 50, 70, etc.).
- ▶ **Número de ectoparásitos presentes en piel de un pez** (1, 3, 5, 6, etc.).
- ▶ **Número de intoxicaciones por veneno paralizante de los mariscos VPM** (1, 2, 3, ..., 40).
- ▶ **Las variables binarias pueden ser tratadas como variables discretas:** sano/enfermo; macho/hembra; vivo/muerto. Se crea una variable ficticia 0-1 que representa la probabilidad de tomar un determinado valor.

# IMPORTANCIA DE IDENTIFICAR Y ANALIZAR VARIABLES ALEATORIAS DISCRETAS

- ▶ Las variables aleatorias discretas no tienen distribución normal.
- ▶ Según sea la variable aleatoria discreta, tendrá una función de distribución de probabilidad asociada (**Bernoulli, Binomial, Binomial Negativa, Poisson, entre otras**).
- ▶ Es importante identificar la naturaleza que tiene nuestra variable en estudio, y así evitar errores en los análisis estadísticos que llevemos a cabo.

# EJEMPLO VARIABLE ALEATORIA DISCRETA BINARIA

Se toman aleatoriamente 10 peces desde una bahía contaminada y se determina que 8 de 10 están enfermos. Por lo tanto, la probabilidad de que estén enfermos es de 0.80. Sea  $X=1$  si el pez está enfermo y  $X=0$  en el caso de que este sano. Entonces,

	Sano	Enfermo
$x$	0	1
$f(x)=P(X=x)$	$1-p$	$p$
Probabilidad	0.20	0.80

# DISTRIBUCIÓN DE PROBABILIDAD DISCRETA

## TIPO BERNOULLI

Se llama distribución de Bernoulli aquella distribución donde el valor 1 (éxito) ocurre con la probabilidad  $p$  y el valor 0 (fracaso) con la probabilidad  $1-p$ .

Por lo tanto, una variable aleatoria discreta  $X$  tiene una distribución de Bernoulli si la función de probabilidad es =

$$f(x) = P(X = x) = \begin{cases} 1 - p & ; si \quad x = 0 \\ p & ; si \quad x = 1 \end{cases}$$

$$X \sim Be(p)$$



# EJEMPLO VARIABLE ALEATORIA DISCRETA

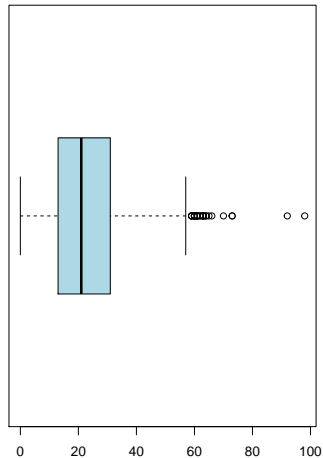
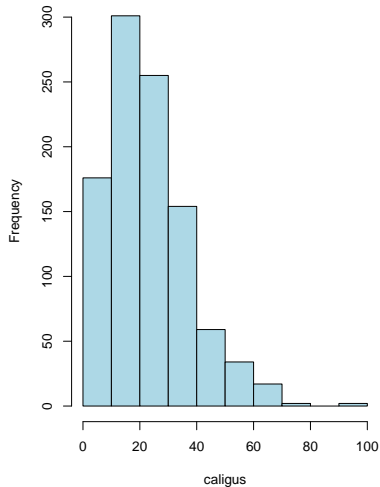
La abundancia de parásitos como el piojo de mar es una variable discreta con distribución binomial negativa, esto significa que hay muchos peces con pocos parásitos (ej= 0 o 1) y pocos con muchos parásitos.

```
# Crea muestra de parásitos  
set.seed(123)  
# n = 1000 animales  
# mu = 25 (parásitos por pez)  
# theta = 3 (número arbitrario para estimar la varianza)  
caligus <- rnegbin (1000, 25, 3)  
head (caligus, 10)
```

```
## [1] 17 42 3 24 50 25 10 4 46 26
```

# DISTRIBUCIÓN DISCRETA BINOMIAL NEGATIVA

Histogram of caligus



# PRÁCTICA ANÁLISIS DE DATOS

1.- Guía de trabajo Rmarkdown disponible en drive.

## **Clase 04-Variables aleatorias discretas**

2.- La tarea se realiza en Rstudio.cloud, proyecto (**Clase 04-Variables aleatorias discretas**).

# RESUMEN DE LA CLASE

- ▶ Identificamos y clasificamos variables aleatorias, datos y factores.
- ▶ Estudiamos sobre variables aleatorias discretas y algunas distribuciones de probabilidad asociadas (Bernoulli o Binomial).