

CLASE 03 - VARIABLES DISCRETAS Y CUALITATIVAS

Diplomado en Análisis de Datos y Modelamiento Predictivo con Aprendizaje Automático para la Acuicultura.

Dr. José Gallardo Matus

Pontificia Universidad Católica de Valparaíso

15 April 2023

PLAN DE LA CLASE

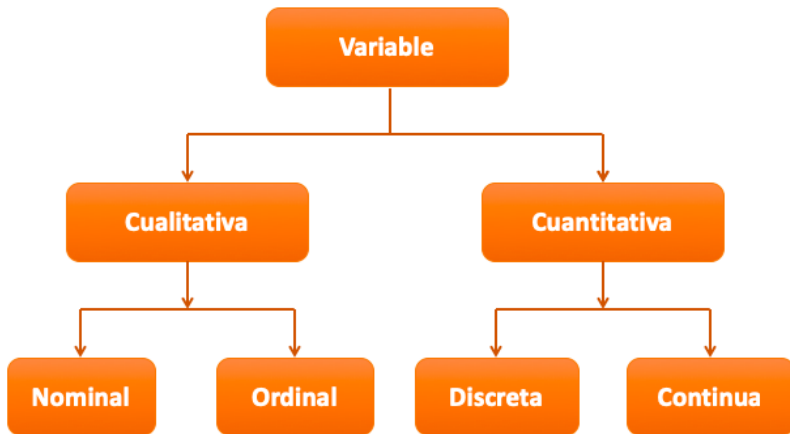
1. Introducción

- ▶ Variables aleatorias discretas: Observar y predecir.
- ▶ Variables aleatorias cualitativas: Observar y predecir.
- ▶ ¿Cómo exportar datos de R?.

2. Práctica con R y Rstudio cloud

- ▶ Observar y predecir variables aleatorias discretas con distribución Bernoulli o Binomial.
- ▶ Exportar datos de R a excel, txt y csv.

CLASIFICACIÓN DE VARIABLES



VARIABLES ALEATORIAS DISCRETAS

- ▶ Son aquellas que presentan un número contable de valores discretos; por ejemplo:
- ▶ **Fecundidad medida como número de huevos:** (1, 15, 26, 50, etc.).
- ▶ **Número de parásitos** (1, 3, 5, 6, etc.).
- ▶ **Número de días a la muerte durante desafío contra patógenos** (1, 2, 3, ..., 40).
- ▶ **Número de larvas:** (1, 15, 26, 50, etc.).

Regla de oro: Si lo puede contar se puede tratar como una variable discreta.

VARIABLES ALEATORIAS CUALITATIVAS

- ▶ Las variables cualitativas son aquellas que expresan una cualidad.

ORDINALES * Color del salmón: Escala salmofan 20-34.

- ▶ Nivel de cataratas: Alto, Medio, Bajo.
- ▶ Infestación por caligus: Centro de Alta Diseminación (CAD) =
o > 3 hembras Ovigeras; Centro normal < 3 hembras ovigeras.

NOMINALES

- ▶ Sexo: Macho, hembra.
- ▶ Estado de madurez: Maduro, Inmaduro.

ESTUDIO DE CASO: DAÑO BRANQUIAL

- ▶ Estudio de daño branquial por amebiasis, Taylor et al.~2009.
- ▶ Infection level : ¿Nominal u ordinal?.
- ▶ ¿Por qué Gill Score es una buena idea?

Table 1. Gross gill score system to estimate the severity of AGD.

Infection level	Gill score	Gross description
Clear	0	No sign of infection and healthy red colour
Very light	1	1 white spot, light scarring or undefined necrotic streaking
Light	2	2-3 spots/small mucus patch
Moderate	3	Established thickened mucus patch or spot groupings up to 20% of gill area
Advanced	4	Established lesions covering up to 50% of gill area
Heavy	5	Extensive lesions covering most of the gill surface

Adapted from Tasmanian Atlantic salmon farming company, Tassal Operations Pty Ltd.

VARIABLES CUALITATIVAS: TABLAS Y MEDIAS

- ▶ Recomendamos usar tablas de frecuencia para resumir información de variables cualitativas.
- ▶ Use media, varianza y desviación estándar apropiada a su variable.

Tabla 1: Tabla de frecuencia y cálculo de la media ponderada.

Score	Measure 2	Ponderado 2	Measure 3	Ponderado 3
0	9	0.00	65	0.00
1	62	0.07	154	0.19
2	185	0.43	211	0.52
3	272	0.94	162	0.60
4	193	0.89	114	0.56
5	158	0.91	106	0.65
n / Media	879	3.23	812	2.52

OBSERVAR DISTRIBUCIÓN DE DAÑO BRANQUIAL

- ▶ Los científicos de datos, usualmente tratan las variables cualitativas como discretas.
- ▶ Para variables cualitativas o discretas prefiera diagramas de barras.

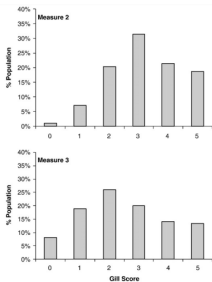


Fig. 2. Gill score distribution in the 680 survival trial fish at each ACD measurement. Sample sizes were 869 (Measure 1), 879 (Measure 2) and 812 (Measure 3).

DISTRIBUCIÓN BERNOULLI

- ▶ La distribución de Bernoulli, describe un experimento aleatorio que sólo admite dos resultados excluyentes (éxito y fracaso).
- ▶ El éxito es usualmente el evento de interes, 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 CON 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 ?

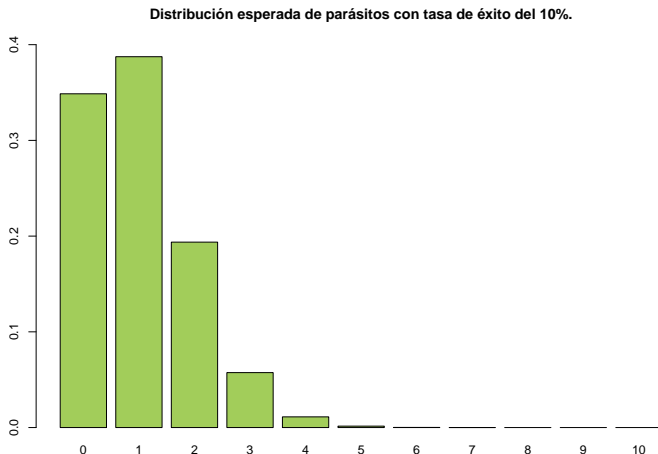
	Fracaso	Éxito
x	0	1
$f(x)=P(X=x)$	$1-p$ 0.35	p 0.65

DISTRIBUCIÓN BINOMIAL

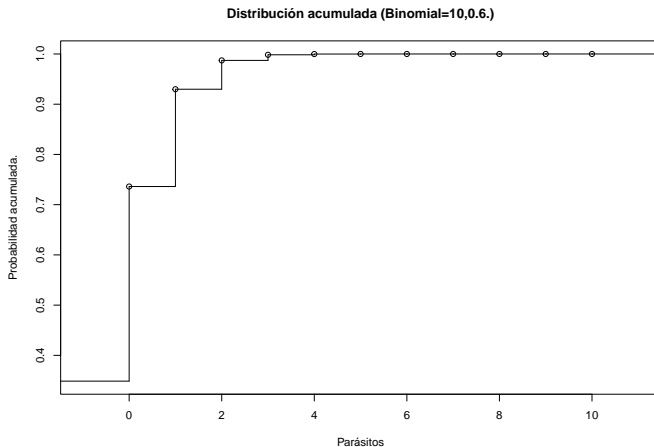
- ▶ Cuando se realizan n pruebas de Bernoulli sucesivas e independientes, la variable aleatoria discreta X toma una distribución binomial.
- ▶ $X =$ “número de veces que ocurre el suceso éxito en n pruebas”.
- ▶ Ejemplos
 - a) Número de camarones que sufren síndrome de la mancha blanca en una finca.
 - b) Número de parásitos en peces de cultivo.

VARIABLE DISCRETA CON DISTRIBUCIÓN BINOMIAL

- ▶ Simulemos 10 ensayos aleatorios de parásitos que intentan infectar peces con una probabilidad de éxito del 1 sobre 10.

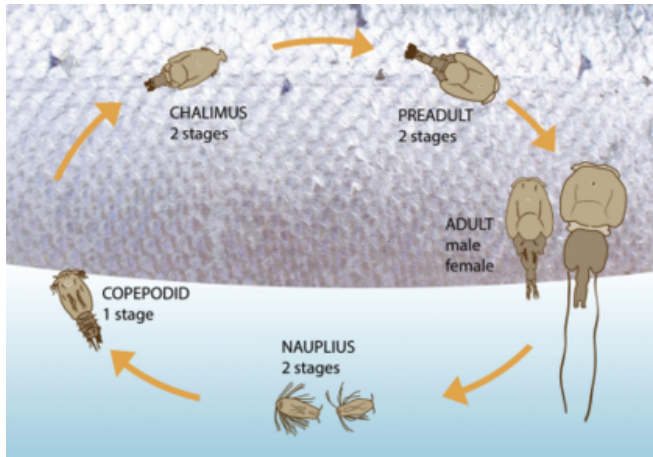


DISTRIBUCIÓN ACUMULADA DE PARÁSITOS.



ESTUDIO DE CASO: PARÁSITOS DE PECES

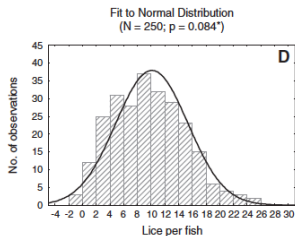
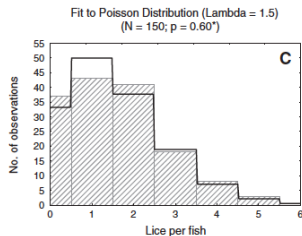
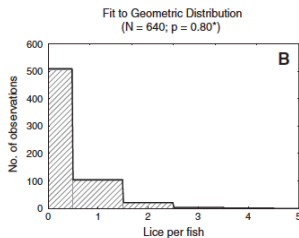
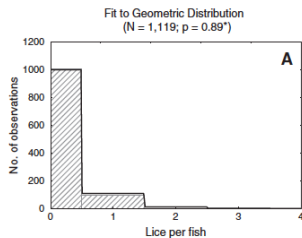
- ▶ Estudio distribución de piojos de salmon (Variable discreta)
Heuch et al.~2011.



Fuente Imagen: Thorstad et al.~2015

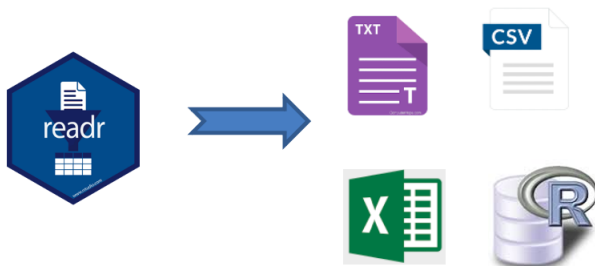
OBSERVAR DISTRIBUCIÓN DE PARÁSITOS

- La distribución de la variable depende de la prevalencia.



COMO EXPORTAR DATOS DESDE R

El paquete **Readr** es muy útil para exportar datos desde R a múltiples formatos.



PRÁCTICA OTRAS 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.
- ▶ Reconocemos variables aleatorias discretas y cualitativas.
- ▶ Simulamos y observamos distribuciones variables con probabilidad diferente de normal: Bernoulli, Binomial, Binomial negativa.