

# Clase 9 Regresión logística

Curso Introducción al análisis de datos con R para la Acuicultura

Dra. María Angélica Rueda Calderón

Pontificia Universidad Católica de Valparaíso

30 July 2022

# PLAN DE LA CLASE

## 1.- Introducción

- ▶ Regresión logística ¿Qué es y para que sirve?
- ▶ Ejemplo de modelo de regresión logística.
- ▶ Interpretación de regresión logística con R.

## 2.- Práctica con R y Rstudio cloud

- ▶ Ajustar modelos de regresión logística.
- ▶ Realizar gráficas avanzadas con ggplot2.

# REGRESIÓN LOGÍSTICA

La regresión logística es una técnica de modelamiento predictivo en la cual la probabilidad de un resultado o variable dicotómica se relaciona con una o más variables predictoras.

- ▶ Predecir la ocurrencia de una enfermedad en camarón. Leung and Tran, 2000
- ▶ Percepción del beneficio o impacto de la acuicultura. Alexander et al., 2016
- ▶ Riesgo de Co-infection ISA - Caligus en salmon. Valdes-Donoso et al., 2013
- ▶ Predicción de canibalismo en bacalao de estuario. Hseu and Huang, 2012

# ECUACIÓN DE LA REGRESIÓN LOGÍSTICA

- ▶ Modelo de regresión logística en formato lineal:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1$$

- ▶ Forma de calcular la probabilidad de que el evento éxito ocurra dada la variable predictora considerada:

$$p(Y = 1|X_1) = \frac{e^{\beta_0 + \beta_1 X_1}}{1 + e^{\beta_0 + \beta_1 X_1}}$$

- ▶ Modelo de regresión logística múltiple en formato lineal:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

# SUPUESTOS REGRESIÓN LOGÍSTICA

Las principales condiciones que este modelo requiere son:

- ▶ Respuesta binaria: La variable respuesta debe ser binaria.
- ▶ Independencia: las observaciones deben ser independientes.
- ▶ Multicolinealidad: se requiere de muy poca a ninguna multicolinealidad entre los predictores (para regresión logística múltiple).
- ▶ Linealidad: entre la variable independiente y el logaritmo natural de odds (**Cociente de chances**).

# COCIENTE DE CHANCES

- ▶  $\text{COCIENTE DE CHANCES} = \text{ODDS RATIO}$
- ▶ Supongamos que tenemos 10 salmones machos (7 maduros y 3 inmaduros) y 10 salmones hembras (3 maduras y 7 inmaduras).
- ▶ ¿Qué sexo tiene mayor chance de madurar?
- ▶ ¿Cuál es la chance de madurar de los machos respecto de las hembras?

		Madurez	
		1	0
Sexo	Macho	7	3
	Hembra	3	7

# CALCULO DE CHANCES

		Madurez	
		1	0
Sexo	Macho	7	3
	Hembra	3	7

► Primero calculamos la chance en los machos.

a)  $P_M(Maduro) = 7/10 = 0.7$

b)  $P_M(Inmaduro) = 1 - P_M(Maduro) = 1 - 0.7 = 0.3$

c)  $odds_M = P_M(Maduro)/P_M(Inmaduro) = 0.7/0.3 = 2.333$

► Calcule usted la chance en las hembras.

# CALCULO DE CHANCES

► Calculamos la chance en las hembras.

a)  $P_H(Maduro) = 3/10 = 0.3$

b)  $P_H(Inmaduro) = 1 - P_H(Maduro) = 1 - 0.3 = 0.7$

c)  $odds_H = P_H(Maduro)/P_H(Inmaduro) = 0.3/0.7 = 0.429$

*Calcular el cociente de chances (Odds Ratio)*

► ¿Cuál es la chance de madurar de los machos respecto de las hembras?

$$OR = odds_M / odds_H = 2.333 / 0.429 = 5.44$$



# REGRESIÓN LOGÍSTICA CON R

- ▶ glm: Ajusta modelos lineales generalizados.
- ▶ family= gaussian, binomial, otras.

```
modelo_logit <- glm(Madurez ~ Sexo,  
                    family= binomial, data = datos)  
  
summary(modelo_logit)$coef %>% kable(digits = 3)
```

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.847	0.690	-1.228	0.220
SexoMacho	1.695	0.976	1.736	0.082

# COCIENTE DE CHANCES (ODDS RATIO) CON R

- ▶ Se calcula con el exponencial de los coeficientes del modelo de regresión logística.

```
logit_or <- exp(cbind(OR = coef(modelo_logit),  
                      confint(modelo_logit)))
```

```
logit_or %>% kable(digits = 3)
```

	OR	2.5 %	97.5 %
(Intercept)	0.429	0.092	1.542
SexoMacho	5.444	0.876	43.418

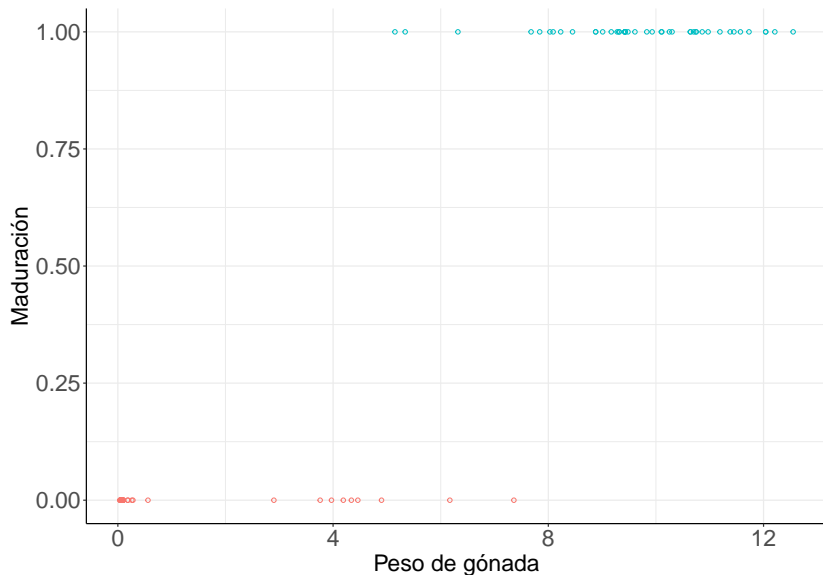
# ESTUDIO DE CASO: MADURACIÓN EN SALMÓN DEL ATLÁNTICO

- ▶ En este estudio de caso trabajaremos con un subconjunto de la base de datos relacionada a la maduración en salmones machos ( $n=90$ ).

Variable	Descripción
<b>Genotype</b>	Genotipo Vgll3: EE-EL-LL
<b>Gonad</b>	Peso de gónada
<b>Maturation</b>	Estado de maduración (1: maduro) o (0: inmaduro)

- ▶ Determinaremos si la probabilidad de maduración está relacionada con el peso de la gónada.
- ▶ Como tarea usted evaluará probabilidad de maduración en función del genotipo del pez (gen Vgll3 de maduración).

# RELACIÓN MADURACIÓN VS PESO DE GÓNADA

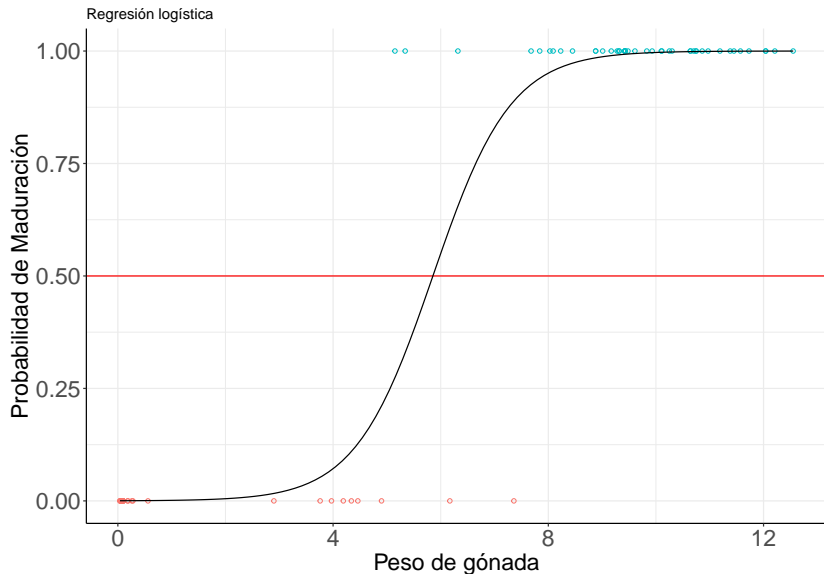


# REGRESIÓN LOGÍSTICA SIMPLE

```
mod_logit <- glm(Maturation ~ Gonad,  
                 family= binomial, data = maduracion)  
summary(mod_logit)$coef %>% kable()
```

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-8.089844	2.6425566	-3.06137	0.0022033
Gonad	1.381678	0.4255612	3.24672	0.0011674

# PREDICCIÓN REGRESIÓN LOGÍSTICA



# PREDICCIÓN MADURACIÓN

- Usando la ecuación podemos predecir probabilidad de madurar a un peso de gónada determinado 4g. v/s 8g.

Table 2: Predicciones de maduración según el peso de la gónada.

Peso de Gónada	Predicción
4	0.0715492
8	0.9509014

# COCIENTE DE CHANCES (ODDS RATIO) CON R

## ► El cociente de chances (OR)

```
OR_def <-exp(cbind(OR = coef(mod_logit),  
                  confint(mod_logit)))
```

```
OR_def %>% kable(digits = 3)
```

	OR	2.5 %	97.5 %
(Intercept)	0.000	0.000	0.014
Gonad	3.982	2.198	13.282



# RESUMEN DE LA CLASE

- 1). Revisión de conceptos: regresión logística.
- 2). Cociente de chances - Odd Ratio.
- 3). Construir y ajustar modelos de regresión logística con R.
- 4). Gráficas avanzadas con ggplot2.