

Clase 18 Introducción a Modelos Lineales Mixtos

Diplomado en Análisis de datos con R para la Acuicultura.

Dra. María Angélica Rueda

Pontificia Universidad Católica de Valparaíso

28 May 2022

PLAN DE LA CLASE

1.- Introducción

- ▶ Modelos lineales mixtos (MLM).
- ▶ Efectos fijos y efectos aleatorios.
- ▶ Ecuación del modelo lineal mixto (MLM).
- ▶ Interpretación de MLM con R.

2.- Práctica con R y Rstudio cloud

- ▶ Ajustar modelos lineales mixtos.
- ▶ Realizar gráficas avanzadas con ggplot2.
- ▶ Elaborar un reporte dinámico en formato html.

MODELOS LINEALES MIXTOS

Los modelos lineales mixtos (MLM) son una generalización del modelo lineal de regresión clásico, contemplando la posible existencia de observaciones correlacionadas (ej. Medidas repetidas en el mismo individuo) o con variabilidad heterogénea, vinculadas a la presencia de factores aleatorios.

$$Y = X\beta + Zu + \epsilon$$

Efectos fijos ($X\beta$)

Efectos aleatorios ($Zu + \epsilon$)

Los modelos lineales mixtos surgen cuando no se cumplen los siguientes supuestos:

- ▶ Que hayan observaciones correlacionadas.
- ▶ Que NO haya homogeneidad de varianzas.

¿QUÉ SON LOS EFECTOS FIJOS?

- ▶ Los efectos fijos se asumen que son determinados a propósito por el analista de los datos, eso dependerá de las variables a las que se les desea estimar **efectos promedios**.
- ▶ Los efectos fijos estiman **las medias** de las variables predictoras.
- ▶ En un modelo lineal mixto las variables cuantitativas continuas (e.g., **Edad**) o factores (e.g., **Dieta**) son considerados como efectos fijos.

¿QUÉ SON LOS EFECTOS ALEATORIOS?

- ▶ Los efectos aleatorios están asociados a grupos de observaciones. Los efectos aleatorios estiman **varianzas**.
- ▶ Para considerar una variable predictora cualitativa como un efecto aleatorio del modelo lineal mixto, dicha variable debe tener al menos 5 niveles.
- ▶ Una variable predictora categórica con dos niveles (binaria) **NO** puede ser un efecto aleatorio.
- ▶ Una variable aleatoria continua **NO** puede ser un efecto aleatorio.

ALGUNOS EJEMPLOS DE EFECTOS ALEATORIOS

- i) Medidas repetidas sobre un mismo individuo, los individuos se pueden incluir como efecto aleatorio.
- ii) Respuestas observadas en grupos de unidades experimentales homogéneas, ej. jaulas o estanque se puede considerar como efecto aleatorio.
- iii) Mediciones de los animales de diferentes familias, familia se puede incluir como efecto aleatorio.

¿EFECTO FIJO O ALEATORIO?

1). ¿Cuál es el número de niveles?

- ▶ Pequeño (Fijo) (e.g., **Dieta** con tres niveles **D1**, **D2** y **D3**).
- ▶ Grande o infinito (Posiblemente aleatorio) (e.g., **Jaulas** con 10 niveles por 10 centros de cultivo **j1**, **j2**...**j00**).

2). ¿Son los niveles repetibles?

- ▶ Sí (Fijo) (e.g., **Dieta** podrías aplicarlas en diferentes lugares).
- ▶ No (Aleatorio) (e.g., **Jaula** no podría repetirse en distintos centros de cultivo).

3). ¿Se necesitan realizar inferencias para niveles no incluidos en el muestreo?

- ▶ No (Posiblemente fijo) (e.g., **Dieta** D4 y D5).
- ▶ Sí (Posiblemente aleatorio) (e.g., **Jaula** j200, j300).

ESTUDIO DE CASO: ANALISIS DE PRODUCCIÓN Y CALIDAD

En este estudio de caso trabajaremos con un set de datos de producción y calidad de salmón Chinook ($n=204$) publicado en la revista aquaculture.

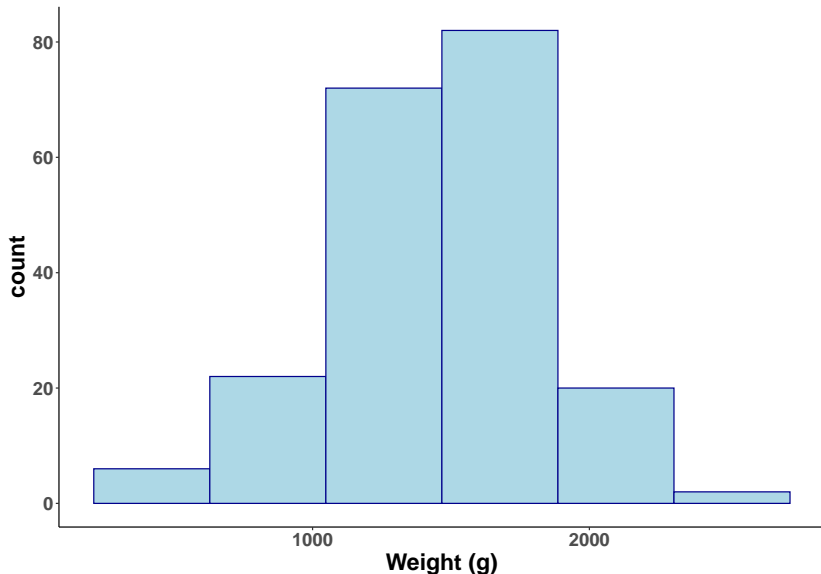
Las variables de estudio se describen a continuación:

Variable	Descripción
ID	Identificación del individuo
Family	Identificación familiar
Weight	Peso en la cosecha (g)
Fat_Meter	Concentración de lípidos medido con sensor
Dry_Lipid	Concentración de lípidos medido por gravimetría

BASE DE DATOS

ID	Family	Weight	Fat_Meter	Dry_Lipid
1	369512	2164	8.9	33.35
2	91524	1838	8	25.51
3	40	1739	7.4	21.89
4	CH13	1670	9	28.73
5	101829	1883	10.4	27.57
6	55	1354	8.7	19.66
7	44	803	3	11.93
8	51	1890	11.3	29.25
9	369512	1485	6.7	13.48
10	95776	1877	9.4	17.74

DISTRIBUCIÓN DE LA VARIABLE RESPUESTA (WEIGHT)



MODELO LINEAL

```
lm.quality <- lm(Weight ~ Dry_Lipid + Fat_Meter,  
  data = quality)
```

<i>Predictors</i>	Weight			
	<i>Estimates</i>	<i>std. Error</i>	<i>CI</i>	<i>p</i>
(Intercept)	595.99	63.25	471.26 – 720.71	<0.001
Dry Lipid	10.71	2.75	5.28 – 16.14	<0.001
Fat Meter	81.00	8.14	64.95 – 97.05	<0.001
Observations	204			
R ² / R ² adjusted	0.504 / 0.499			
AIC	2890.477			

Independencia

```
##  
## Durbin-Watson test  
##  
## data: Weight ~ Dry_Lipid + Fat_Meter  
## DW = 1.502, p-value = 0.0003308  
## alternative hypothesis: true autocorrelation is not 0
```

CUMPLIMIENTO DE SUPUESTOS

Homogeneidad de varianzas

```
##  
## studentized Breusch-Pagan test  
##  
## data: lm.quality  
## BP = 15.685, df = 2, p-value = 0.0003926
```

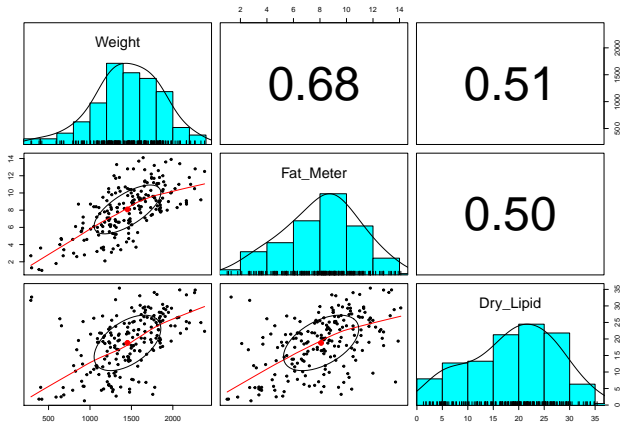
CUMPLIMIENTO DE SUPUESTOS

Normalidad

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  lm_residuals  
## W = 0.99393, p-value = 0.5743
```

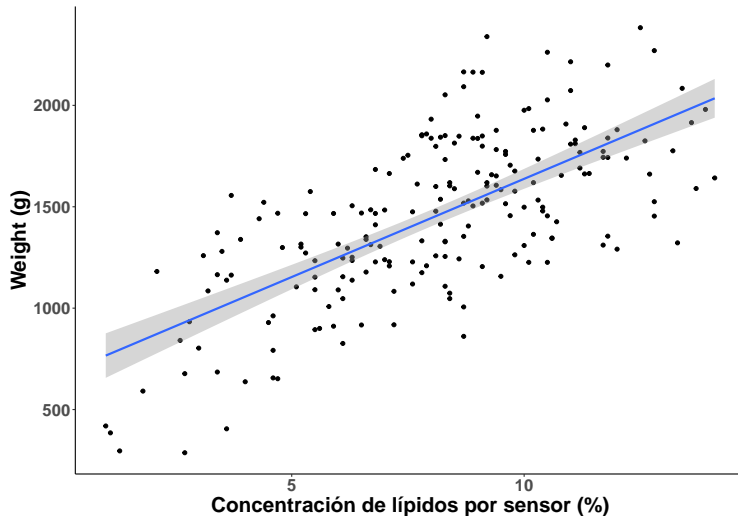
CUMPLIMIENTO DE SUPUESTOS

Multicolinealidad



CUMPLIMIENTO DE SUPUESTOS

Linealidad



CUMPLIMIENTO DE SUPUESTOS

Linealidad

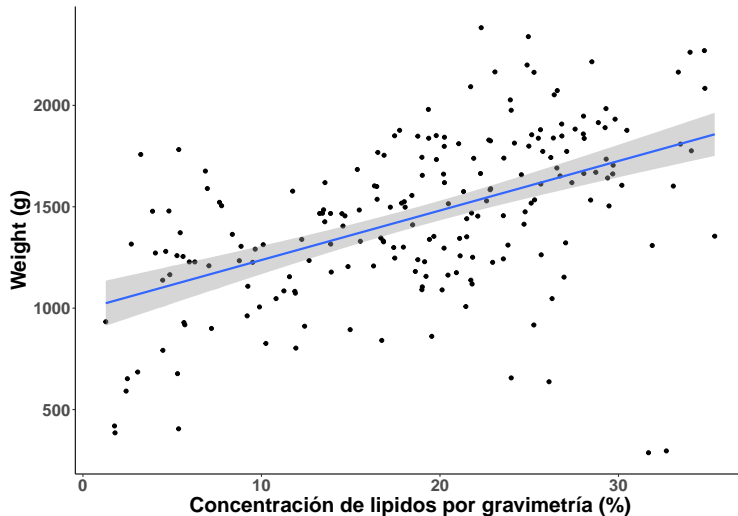
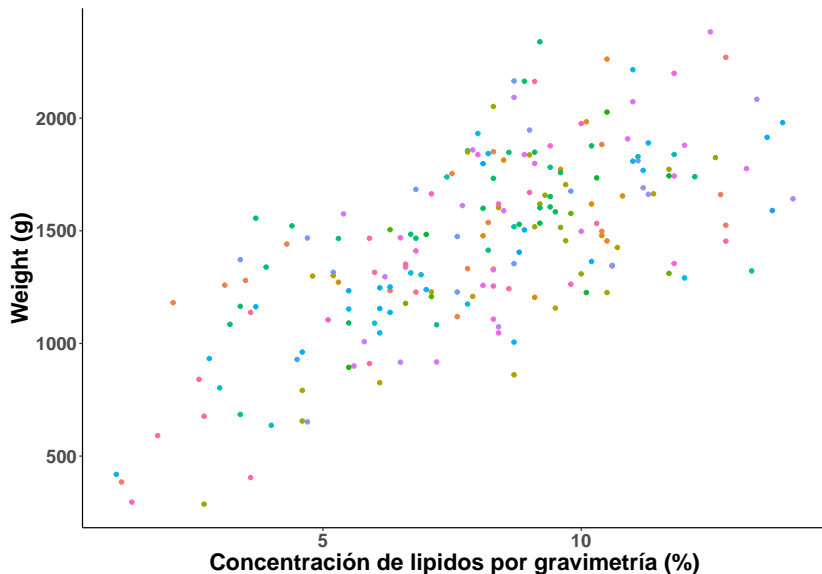


GRÁFICO DE DISPERSIÓN POR FAMILIA



MODELOS LINEALES MIXTOS

```
library(lme4)
```

Función **lmer**

Cantidad de Familias
58

- Familia se puede considerar como efecto **aleatorio**.

MODELO LINEAL MIXTO

```
MLM <- lmer(Weight ~ Dry_Lipid + Fat_Meter + (1|Family),  
            data = quality)
```

<i>Predictors</i>	Weight			
	<i>Estimates std. Error</i>		<i>CI</i>	<i>p</i>
(Intercept)	539.49	65.84	409.55 – 669.43	<0.001
Dry Lipid	11.24	2.71	5.89 – 16.58	<0.001
Fat Meter	84.85	8.10	68.88 – 100.82	<0.001
Random Effects				
σ^2	71637.80			
τ_{00} Family	8374.14			
ICC	0.10			
N_{Family}	58			
Observations	204			
Marginal R^2 / Conditional R^2	0.529 / 0.578			

R^2 Marginal y R^2 Condicional

- ▶ R^2_{Marginal} : proporción de la varianza explicada solo por los efectos fijos.
- ▶ $R^2_{\text{Condicional}}$: proporción de la varianza explicada por todo el modelo.

```
r2_nakagawa(MLM)
```

```
## # R2 for Mixed Models
##
##   Conditional R2: 0.578
##   Marginal R2: 0.529
```

SELECCIÓN DE MODELOS (AIC)

Criterios de selección de modelos AIC

```
AIC %>% kable()
```

	AIC
Modelo lineal	2890.477
Modelo lineal mixto	2867.282

RESUMEN DE LA CLASE

- 1). Modelos lineales mixtos.
- 2). Construir y ajustar modelos lineales mixtos.