Clase 16 Introducción a Modelos Lineales Mixtos

Diplomado en Análisis de datos con R para la Acuicultura

Dra. María Angélica Rueda

Pontificia Universidad Católica de Valparaíso

31 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 pdf.

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 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 solo estiman medias de las variables predictoras.
- En un modelo lineal mixto las variables cuantitativas continuas (*e.g.*, **Peso**) o factores (*e.g.*,**Dieta**) pueden ser usadas como efectos fijos.

¿QUÉ SON 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 (**7 Familias**).
- 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 (hay repeticiones).

ii) Respuestas observadas en grupos de unidades experimentales homogéneas (bloques), pueden ser piscinas o estanques.

iii) Mediciones de los animales (individuos) de una misma familia.

¿CÓMO SE PODRÍA DECIDIR SI ES 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.*, **Familia** con 10 niveles **F1, F2**. . . **F10**).

2). ¿Son los niveles repetibles?

- Sí (Fijo) (*e.g.*, **Dieta** podrías aplicarlas en diferentes lugares).
- No (Aleatorio) (*e.g.*, **Familia** no podrías repetir las familias).

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.*, **Familias** F11, F12).

ESTUDIO DE CASO: ANALISIS DE PRODUCCIÓN Y CALIDAD

En este estudio de caso trabajaremos con un set de datos de salmón (n=354) publicado en la revista JOURNAL OF EVOLUTIONARY BIOLOGY

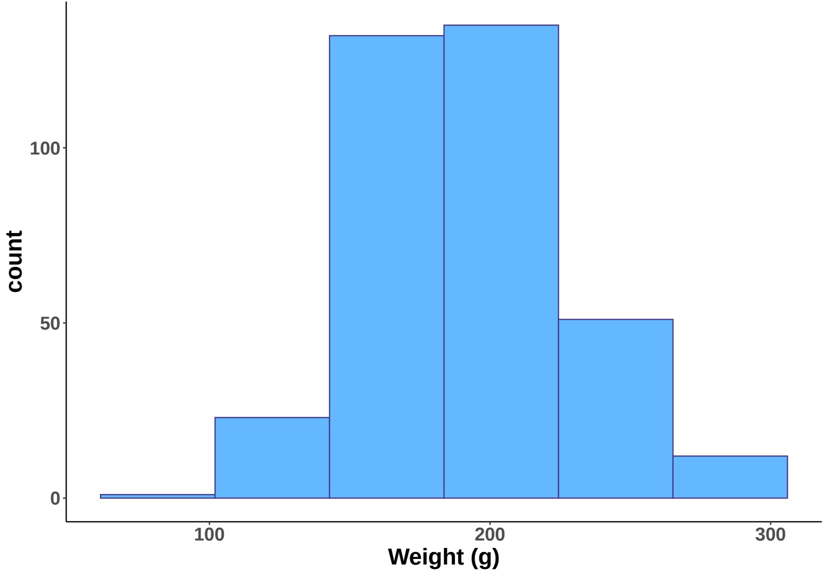
Las variables de estudio se describen a continuación:

Variable	Descripción
Tank	Estanque **(58)**
Days PI	Días de post-infección
Nbfemales	Cantidad de piojos presentes en el pez a la fecha de medición
Weight	Peso en la cosecha (g)
Status	Estado (Cultivado, Silvestre)

BASE DE DATOS

Tank	Weight	Nbfemales	DaysPI	Status
1	169	6	60	Farmed
1	183	6	70	Farmed
1	187	6	77	Farmed
1	194	6	84	Farmed
1	208	5	94	Farmed
1	219	5	103	Farmed
1	237	4	115	Farmed
2	190	3	56	Farmed
2	196	3	66	Farmed
2	204	2	75	Farmed

DISTRIBUCIÓN DE LA VARIABLE RESPUESTA (WEIGHT)



MODELO LINEAL

```
lm.datos <- lm(Weight ~ Status+ Nbfemales+ DaysPI,
               data = datos)
```

	Weight			
Predictors	*Estimates*	*std. Error*	*CI*	*p*
(Intercept)	148.64	3.83	141.09 – 156.18	**<0.001**
Status [Wild]	-14.24	2.54	-19.25 – -9.24	**<0.001**
Nbfemales	-5.74	0.56	-6.84 – -4.64	**<0.001**
DaysPI	0.81	0.04	0.73 – 0.89	**<0.001**
Observations	354			
R^2 / R^2 adjusted	0.563 / 0.559			
AIC	3247.074			

CUMPLIMIENTO DE SUPUESTOS

Robustness of MLM

Our simulation analysis shows that the effect of violations of distributional assumptions of random effect variances and residuals is surprisingly small.

Independencia

```
##
##  Durbin-Watson test
##
## data:  Weight ~ Status + Nbfemales + DaysPI
## DW = 0.89858, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is not 0
```

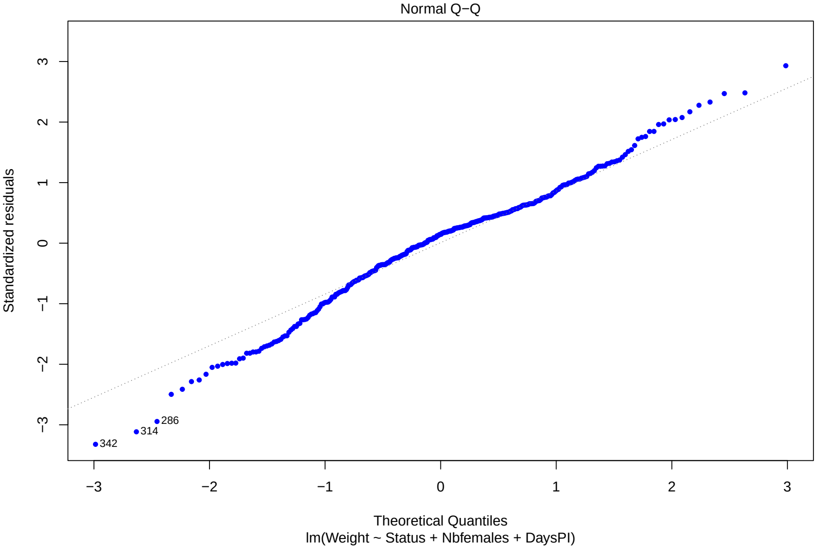
CUMPLIMIENTO DE SUPUESTOS

Homogeneidad de varianzas

```
##
##  studentized Breusch-Pagan test
##
## data:  lm.datos
## BP = 33.753, df = 3, p-value = 2.234e-07
```

CUMPLIMIENTO DE SUPUESTOS

Normalidad



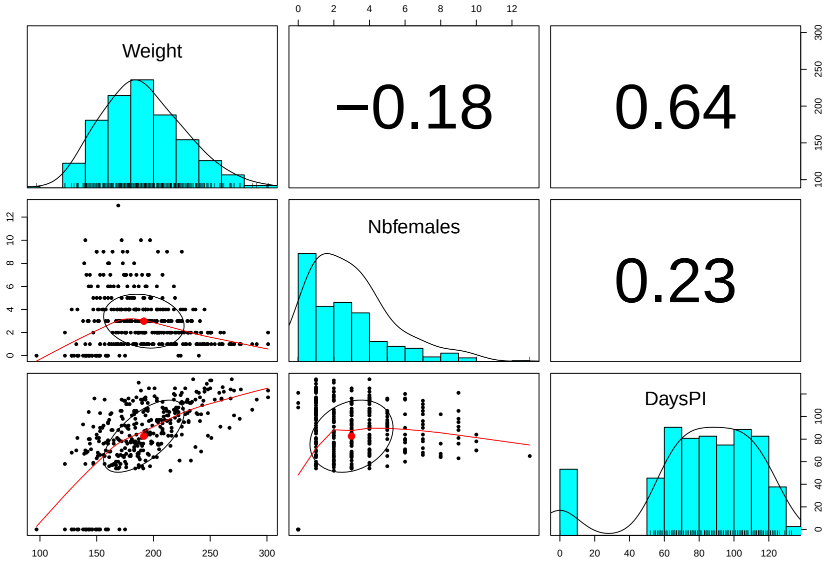
CUMPLIMIENTO DE SUPUESTOS

Normalidad

```
##
##  Shapiro-Wilk normality test
##
## data:  lm_residuals
## W = 0.98434, p-value = 0.0006903
```

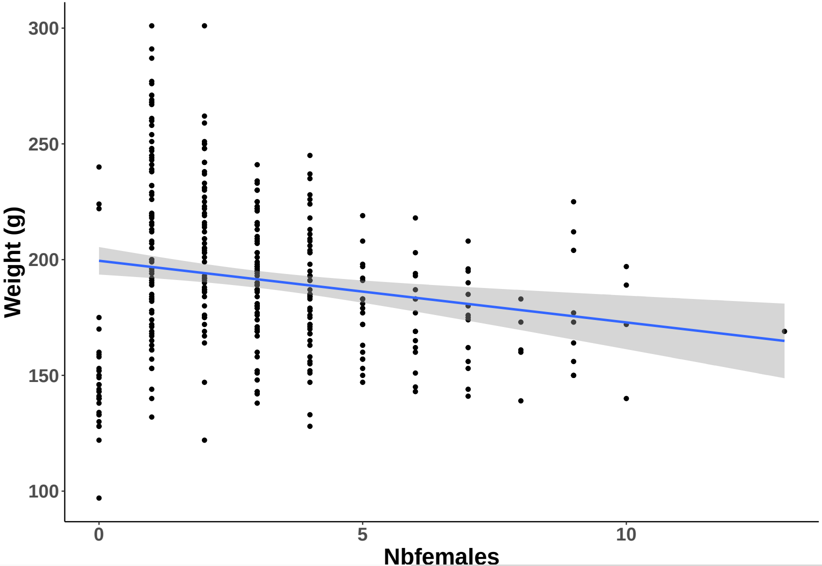
CUMPLIMIENTO DE SUPUESTOS

Multicolinealidad



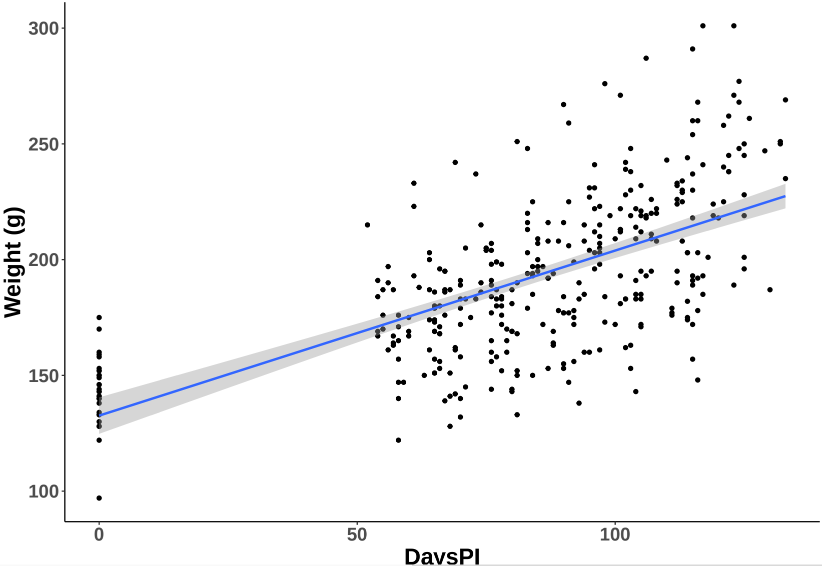
CUMPLIMIENTO DE SUPUESTOS

Linealidad



CUMPLIMIENTO DE SUPUESTOS

Linealidad



```
library(lme4)
```

Función **lmer**

Cantidad de Tanques
58

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

MODELO LINEAL MIXTO

```
MLM <- lmer(Weight ~ Status+ Nbfemales+ DaysPI + (1|Tank),
            data = datos)
```

	Weight			
Predictors	*Estimates*	*std. Error*	*CI*	*p*
(Intercept)	148.79	4.17	140.51 – 157.07	**<0.001**
Status [Wild]	-10.15	5.22	-20.53 – 0.23	0.055
Nbfemales	-5.57	0.37	-6.30 – -4.83	**<0.001**
DaysPI	0.77	0.02	0.73 – 0.81	**<0.001**
Random Effects				
σ^2	98.18			
$\tau_{00\ Tank}$	449.23			
ICC	0.82			
N_{Tank}	58			
Observations	354			

R^2 *Marginal* y R^2 *Condicional*

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

```
r2_nakagawa(MLM)
```

```
## # R2 for Mixed Models
##
##   Conditional R2: 0.916
##      Marginal R2: 0.533
```

SELECCIÓN DE MODELOS (AIC)

Criterios de selección de modelos AIC

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

	AIC
Modelo lineal	3247.074
Modelo lineal mixto	2824.480

RESUMEN DE LA CLASE

1). Modelos lineales mixtos.

2). Construir y ajustar modelos lineales mixtos.