

BIOMETRÍA II

CLASE 9

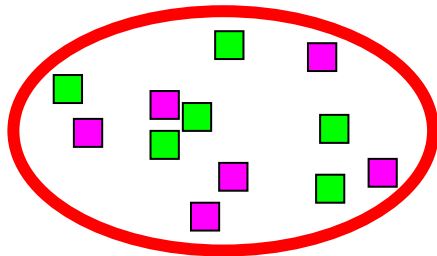
MODELOS LINEALES CON EFECTOS ALEATORIOS

Adriana Pérez
Depto de Ecología, Genética y Evolución
FECN, UBA

Efecto de la disponibilidad lumínica sobre la herbivoría en *Berberis buxifolia* (calafate)

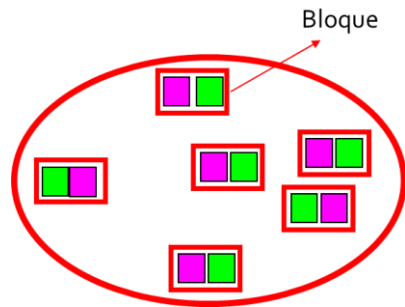


2

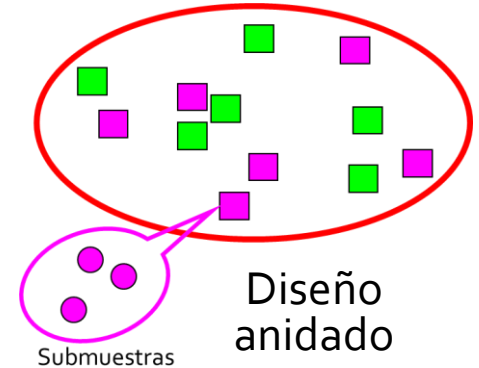


Diseño completamente aleatorizado

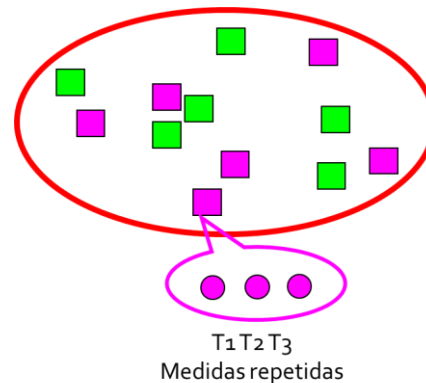
Observaciones independientes



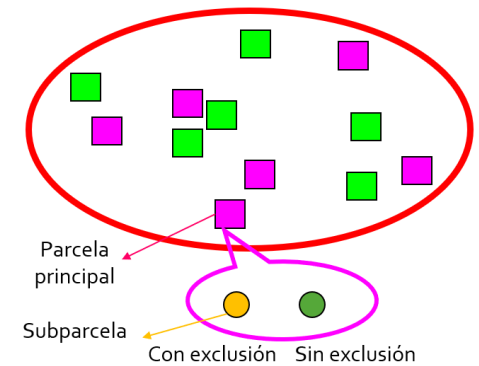
Diseño de Bloques al azar (DBA)



Diseño anidado



Diseño de medidas repetidas (DMR)



Diseño de parcela dividida (Split-plot)

Observaciones no independientes (datos agrupados)

Datos agrupados

3

- Observaciones no independientes. Existe algún nivel de agregación de los datos, que induce una estructura de correlación entre las observaciones
- Ignorar esa falta de independencia afecta las estimaciones de los EE de los coeficientes y por lo tanto la inferencia no es correcta
- Se soluciona declarando en el modelo esa correlación entre las observaciones

¿Cómo modelamos datos correlacionados?

4

- En forma implícita, mediante la adición de VE de efecto aleatorio. Se induce la correlación entre observaciones. [Modelos condicionales o mixtos](#)
- En forma explícita, mediante la imposición de una estructura de correlación entre las observaciones (matriz de covarianzas). [Modelos marginales](#)
- Con errores con distribución normal ambas aproximaciones (formas explícita e implícita) son equivalentes
- Sin embargo, con datos no normales o cuando el modelo es no lineal, los parámetros bajo estas dos estrategias son intrínsecamente diferentes y se interpretan como:
 - “parámetros promedios poblacionales” (modelos marginales): ajustan un modelo general para la estructura promedio de la población de individuos
 - “parámetros sujeto-específicos” (modelos condicionales): proporcionan un modelo para cada individuo, pero donde la forma general del modelo es la misma para cada sujeto

VE de efectos fijos

VE de efectos aleatorios

5

- los niveles de la VE **son elegidos deliberadamente** por el investigador, porque le interesa estudiar esos efectos en particular
- La inferencia se efectúa sobre **esos** niveles
- Los tratamientos asignados por el investigador a las UE son de efectos fijos, aunque existen factores que no son asignables aleatoriamente pero son fijos (sexo, variedad, edad, precipitaciones, etc)
- $H_0: \mu_i = \mu \text{ o } \beta_i = 0$

- los niveles **son elegidos al azar** por el investigador de una población mayor de niveles de interés
- Las conclusiones se aplican sobre **la población de niveles**, no solo los estudiados
- El objetivo es modelar la falta de independencia y/o estudiar la variabilidad entre los niveles (espacial, temporal, genética, de la técnica experimental)
- Las VE de efectos aleatorios son cuali (bloques, individuos, días, líneas, etc); niveles no informativos
- Si se repite el experimento, no necesariamente se repiten los mismos niveles
- $H_0: \sigma^2_i = 0$

Volviendo a los ejemplos

6

VR: nivel de herbivoría

| Diseño | VE de efectos fijos | VE de efectos aleatorios |
|----------------------------|---------------------|--------------------------|
| Completamente aleatorizado | | |
| bloques al azar | | |
| anidado | | |
| medidas repetidas | | |

Nivel de herbivoría:

1. biomasa consumida
2. Porcentaje de hojas con signos de herbivoría
3. Ramoneada/no ramoneada
4. Cantidad de herbívoros capturados en un día

Distribución de probabilidades?

Efecto de la exposición postnatal a etanol sobre el volumen del cerebro en ratones



7

- La exposición temprana al etanol causa alteraciones cognitivas y conductuales persistentes
- Se desea estudiar los efectos neuroestructurales asociados a esta exposición en ratones
- Para ello se seleccionaron 6 camadas de ratones de 7 días. De cada camada se eligieron 3 ratones al azar, que fueron asignados a uno de los siguientes tratamientos:

a) Solución salina, b) Etanol 1 g/kg, c) Etanol 2 g/kg

A los 82 días se determinó el volumen cerebral por resonancia magnética

Experimento o estudio observacional?

VR:

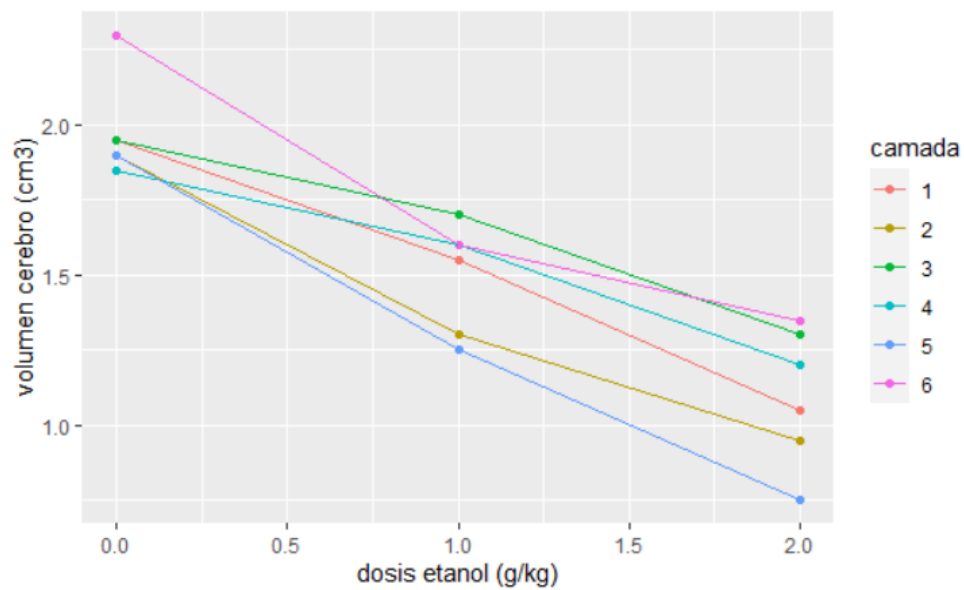
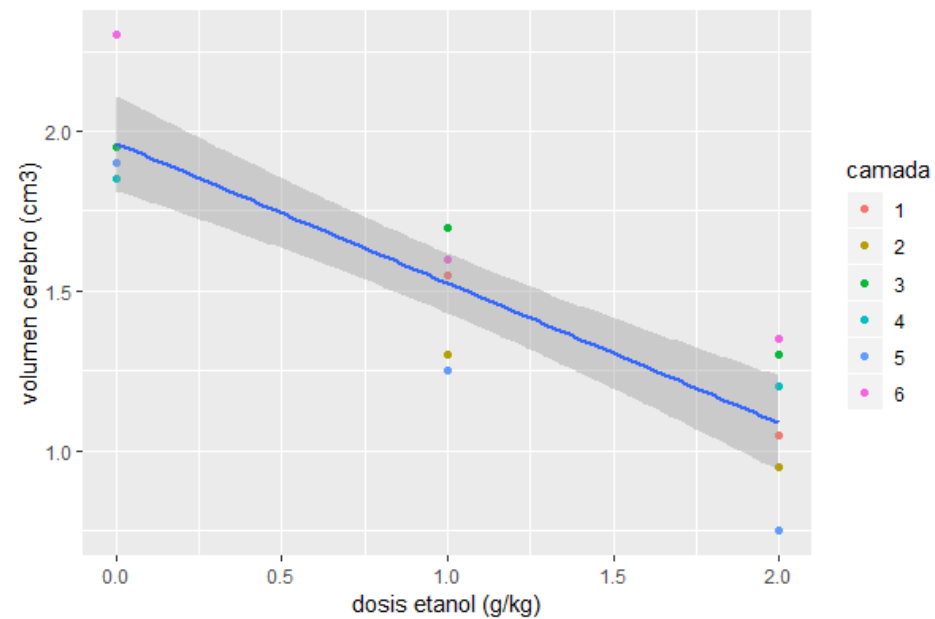
Tipo? Potencial distribución de probabilidades?

VE:

Tipo? De efectos fijos o aleatorios?

| | camada | etanol | vol |
|----|--------|--------|------|
| 1 | 1 | 0 | 1.95 |
| 2 | 2 | 0 | 1.90 |
| 3 | 3 | 0 | 1.95 |
| 4 | 4 | 0 | 1.85 |
| 5 | 5 | 0 | 1.90 |
| 6 | 6 | 0 | 2.30 |
| 7 | 1 | 1 | 1.55 |
| 8 | 2 | 1 | 1.30 |
| 9 | 3 | 1 | 1.70 |
| 10 | 4 | 1 | 1.60 |
| 11 | 5 | 1 | 1.25 |
| 12 | 6 | 1 | 1.60 |
| 13 | 1 | 2 | 1.05 |
| 14 | 2 | 2 | 0.95 |
| 15 | 3 | 2 | 1.30 |
| 16 | 4 | 2 | 1.20 |
| 17 | 5 | 2 | 0.75 |
| 18 | 6 | 2 | 1.35 |

Ratones_camada.txt



Sin considerar las camadas

9

$$Y_i = \beta_0 + \beta_1 \text{etanol}_i + \varepsilon_i \quad i = 1, \dots, 18$$

```
m1<-lm(vol~etanol, bd)
```

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|--------------|
| (Intercept) | 1.96250 | 0.06999 | 28.038 | 4.96e-15 *** |
| etanol | -0.43750 | 0.05422 | -8.069 | 4.96e-07 *** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1878 on 16 degrees of freedom
Multiple R-squared: 0.8028, Adjusted R-squared: 0.7904
F-statistic: 65.12 on 1 and 16 DF, p-value: 4.956e-07

$$\sqrt{\hat{\sigma}^2}$$

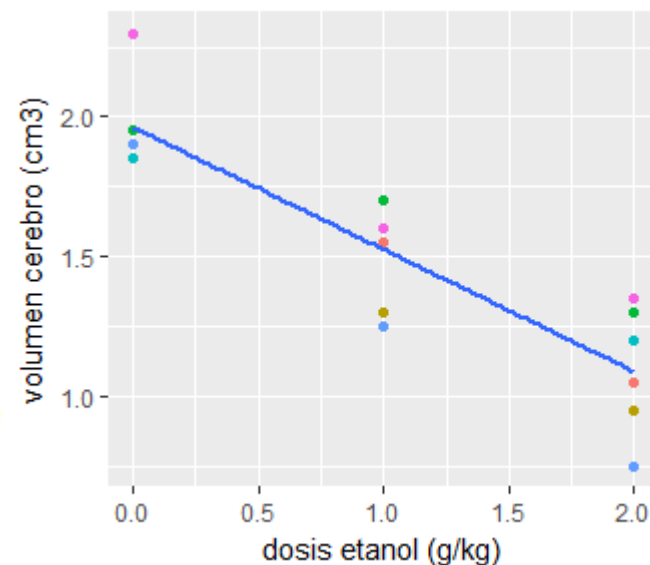
```
> anova(m0)
```

Analysis of Variance Table

Response: vol

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|---------|---------|---------|---------------|
| etanol | 1 | 2.29688 | 2.29688 | 65.116 | 4.956e-07 *** |
| Residuals | 16 | 0.56437 | 0.03527 | | |

$\hat{\sigma}^2$



La variabilidad explicada por las camadas va a parar al error!



Incorporando las camadas como VE de efectos fijos

BLUE: best linear unbiased estimate.
Efecto fijo de la camada

10

$$Y_i = \beta_0 + \beta_1 etanol_i + \beta_2 C_{2i} + \beta_3 C_{3i} + \dots + \varepsilon_i$$

```
m1<-lm(vol~etanol+factor(camada), ratones)
```

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) | |
|---------|----------|------------|---------|----------|-----|
| etanol | -0.43750 | 0.03387 | -12.92 | 5.44e-08 | *** |
| camada1 | 1.95417 | 0.07574 | 25.80 | 3.43e-11 | *** |
| camada2 | 1.82083 | 0.07574 | 24.04 | 7.37e-11 | *** |
| camada3 | 2.08750 | 0.07574 | 27.56 | 1.68e-11 | *** |
| camada4 | 1.98750 | 0.07574 | 26.24 | 2.85e-11 | *** |
| camada5 | 1.73750 | 0.07574 | 22.94 | 1.22e-10 | *** |
| camada6 | 2.18750 | 0.07574 | 28.88 | 1.01e-11 | *** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1173 on 11 degrees of freedom
Multiple R-squared: 0.9966, Adjusted R-squared: 0.9945
F-statistic: 462.4 on 7 and 11 DF, p-value: 1.085e-12

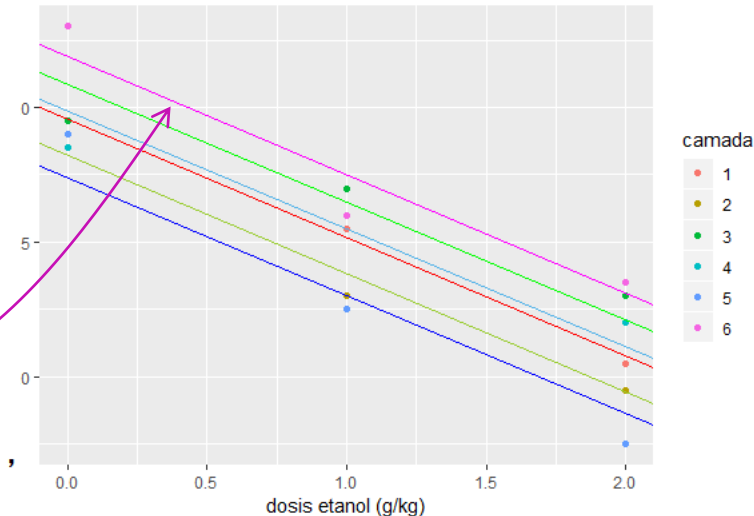
```
> anova(m2)
```

Analysis of Variance Table

Response: vol

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) | |
|-----------|----|---------|---------|----------|-----------|-----|
| etanol | 1 | 2.29688 | 2.29688 | 166.8157 | 5.444e-08 | *** |
| camada | 5 | 0.41292 | 0.08258 | 5.9978 | 0.006458 | ** |
| Residuals | 11 | 0.15146 | 0.01377 | | | |

Es la forma de evaluar si hay efecto de la camada



- ✓ Mejora la precisión ✓
- ✓ Muchos parámetros para estimar ✗
- ✓ Inferencia solo para los bloques medidos ✗

Modelo condicional: camadas como VE de efectos aleatorios

BLUP: best linear unbiased prediction.
Efecto aleatorio de la camada

11

Parte fija

Parte aleatoria

$$Y_{ij} = \beta_0 + \beta_1 \text{etanol}_i + B_j + \varepsilon_{ij}$$

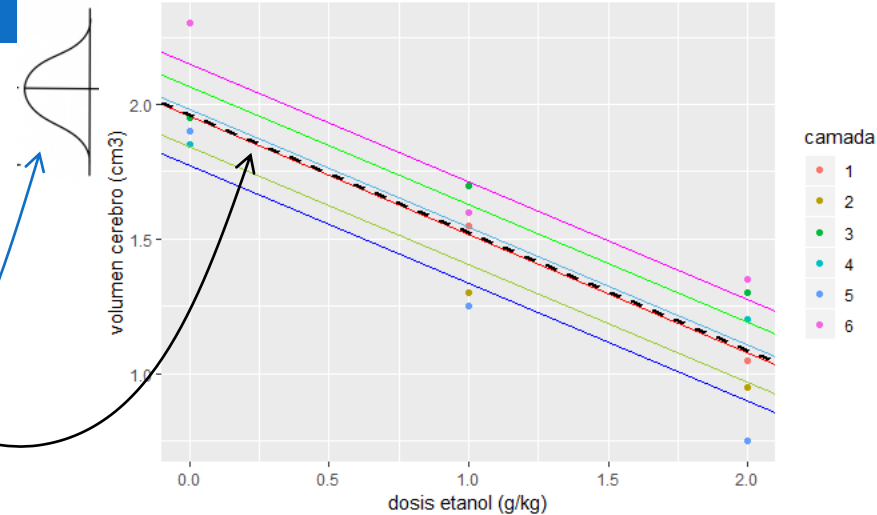
$$i = 1 a 3, j = 1 a 6$$

$$\varepsilon_{ij} \approx NID(0, \sigma^2)$$

$$B_j \approx NID(0, \sigma^2_{camadas})$$

$$\varepsilon_{ij}, B_j \text{ indep}$$

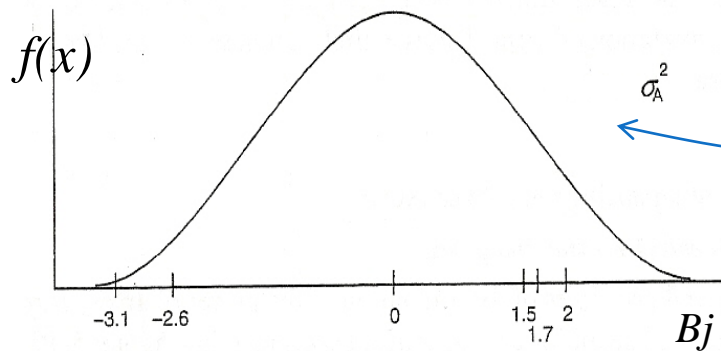
$$Var(Y) = \sigma^2_{camadas} + \sigma^2$$



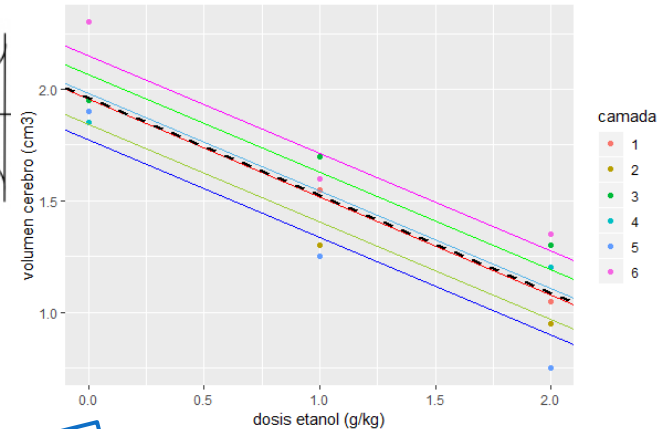
- **Modelo condicional o sujeto-específico:** proporciona un modelo para cada individuo, pero donde la forma general del modelo es la misma para cada sujeto. En este caso es un modelo de intercepto aleatorio, ya que la ordenada al origen es $\beta_0 + B_j$, que cambia aleatoriamente según B_j
- La inclusión de B_j (efecto aleatorio de la camada) induce una estructura de correlación entre las observaciones de la misma camada ✓
- La inferencia no está restringida a las 6 camadas estudiadas sino a toda la población de camadas ✓
- Un único parámetro $\sigma^2_{camadas}$ para modelar la variabilidad entre camadas ✓

B_j : Efecto aleatorio del nivel j de la VE Camada (BLUP: best linear unbiased prediction)

$$B_j \approx N(0, \sigma_{camada}^2)$$



Cuanto sube o baja la ordenada al origen de la camada j -ésima



- ✓ Suponemos que existe un gran número de niveles para el factor camada y por tanto una población de efectos B_j
- ✓ B_j es una variable aleatoria, que se considera normal e independientemente distribuida, con media 0 y varianza σ_{camada}^2
- ✓ Se seleccionan al azar niveles con el propósito de tratarlos como una representación de la población de efectos hacia la cual se pretende inferir.
- ✓ LOS B_j se predicen (no se estiman ya que no son parámetros sino VA)
- ✓ Se estima la variabilidad aportada por las camadas σ_{camada}^2

```
library(lme4)
m2 <- lmer(vol ~ etanol + (1 | camada), data = ratones)
```

- intercepto aleatorio por camada
- camada como factor

REML criterion at convergence: -8.8

Scaled residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|---------|--------|--------|--------|
| -1.2784 | -0.8701 | 0.1124 | 0.6304 | 1.2784 |

Random effects:

| Groups | Name | Variance | Std.Dev. |
|----------|-------------|----------|----------|
| camada | (Intercept) | 0.02294 | 0.1515 |
| Residual | | 0.01377 | 0.1173 |

Number of obs: 18, groups: camada, 6

Fixed effects:

| | Estimate | Std. Error | df | t value | Pr(> t) | |
|-------------|----------|------------|----------|---------|--------------|-----------------|
| (Intercept) | 1.96250 | 0.07573 | 7.59765 | 25.91 | 1.09e-08 *** | $\hat{\beta}_0$ |
| etanol | -0.43750 | 0.03387 | 11.00001 | -12.92 | 5.44e-08 *** | $\hat{\beta}_1$ |

Las estimaciones son por máxima verosimilitud restringida en lugar de máxima verosimilitud

$\hat{\sigma}^2_{camadas}$ Varianza entre camadas
 $\hat{\sigma}^2$ Varianza dentro de camadas

de parámetros?
 GL residuales?

library(lmerTest)
 para obtener p-valores

Para ver significación de los coeficientes por test de Wald puede usarse la función Anova(modelo) del paquete car

```
#También puede utilizarse la librería nlme
library(nlme)
m2b <- lme(vol ~ etanol, random= ~1 | camada, data = ratones)
```

Linear mixed-effects model fit by REML

Data: ratones

| | AIC | BIC | logLik |
|--|------------|----------|----------|
| | -0.8271654 | 2.263189 | 4.413583 |

Random effects:

Formula: ~1 | camada

| | (Intercept) | Residual |
|--|-------------|----------|
|--|-------------|----------|

| | | |
|---------|-----------|-----------|
| StdDev: | 0.1514534 | 0.1173411 |
|---------|-----------|-----------|

$\hat{\sigma}_{camadas}$ y $\hat{\sigma}$

Fixed effects: vol ~ etanol

| | Value | Std. Error | DF | t-value | p-value |
|-------------|---------|------------|----|-----------|---------|
| (Intercept) | 1.9625 | 0.07573227 | 11 | 25.91366 | 0 |
| etanol | -0.4375 | 0.03387346 | 11 | -12.91572 | 0 |

$\hat{\beta}_0$

$\hat{\beta}_1$

Correlation:

(Intr)

etanol -0.447

Standardized Within-Group Residuals:

| | Min | Q1 | Med | Q3 | Max |
|--|------------|------------|-----------|-----------|-----------|
| | -1.2784414 | -0.8700553 | 0.1124409 | 0.6303889 | 1.2784414 |

Number of Observations: 18

Number of Groups: 6

Según Pinheiro y Bates (2000) si los GL residuales son pequeños (<50?) los p no son confiables

Estimación de los parámetros en modelos lineales mixtos

Los modelos con distribución normal y VE con efectos fijos y aleatorios se conocen como **modelos lineales generales y mixtos**. Dos métodos de estimación:

- **Mínimos cuadrados**: Si se verifica distribución normal de los errores, homocedasticidad y diseño sin desbalanceo grosero permiten modelar la falta de independencia. Pero debe indicarse cuál es el término de error para la construcción de las F, lo que lo torna un método tedioso en diseños complicados
- **Máxima verosimilitud**: Más versátil. Permite modelar distintas estructuras de correlación, heterocedasticidad, se bancan desbalanceo. Si n es grande, las estimaciones por MV proveen estimadores insesgados y consistentes. Sin embargo, las estimaciones de las varianzas están subestimadas (porque no corrigen por GL), por lo que en la práctica se aplica una corrección a los GL: **MV restringida o REML** (opción por defecto en lme y lmer)

Supuestos

16

- Muestra aleatoria
- Linealidad para predictoras continuas
- $\varepsilon_{ij} \approx NID(0, \sigma^2)$; $B_j \approx NID(0, \sigma^2_{camadas})$; ε_{ij} y B_j independientes
- Además: ausencia de patrones en los residuos; chequeo de outliers

QQ plot

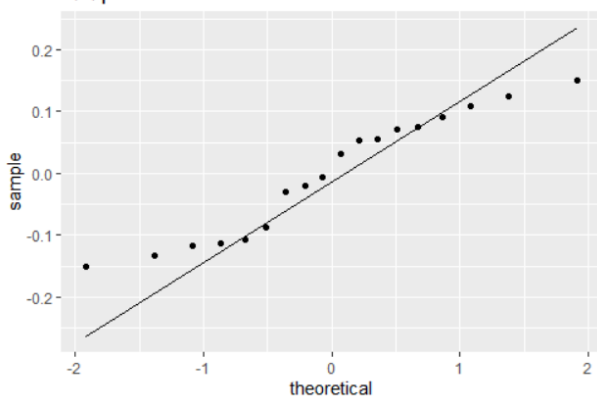
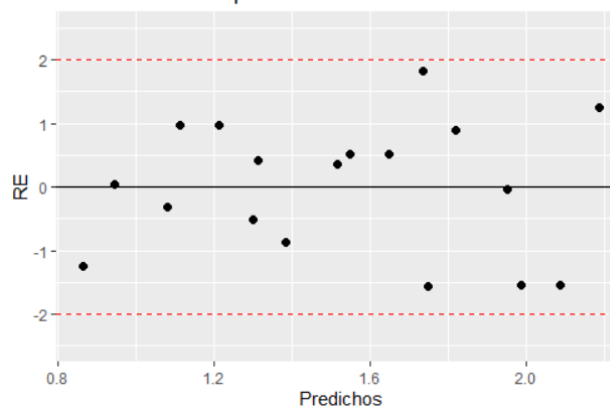
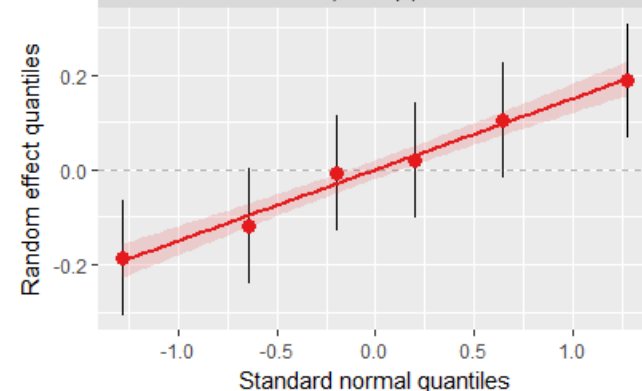


Grafico de RE vs predichos



(Intercept)



shapiro-wilk normality test
data: e
 $w = 0.92747$, $p\text{-value} = 0.1752$

Debería probarse también
normalidad con los B_j
aunque son pocos valores...

Parte fija

17

- Predicciones (sin incluir efectos aleatorios)

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 \text{etanol}_i$$

Fixed effects:

| | Estimate | Std. Error | t value |
|-------------|----------|------------|---------|
| (Intercept) | 1.96250 | 0.07573 | 25.91 |
| etanol | -0.43750 | 0.03387 | -12.92 |

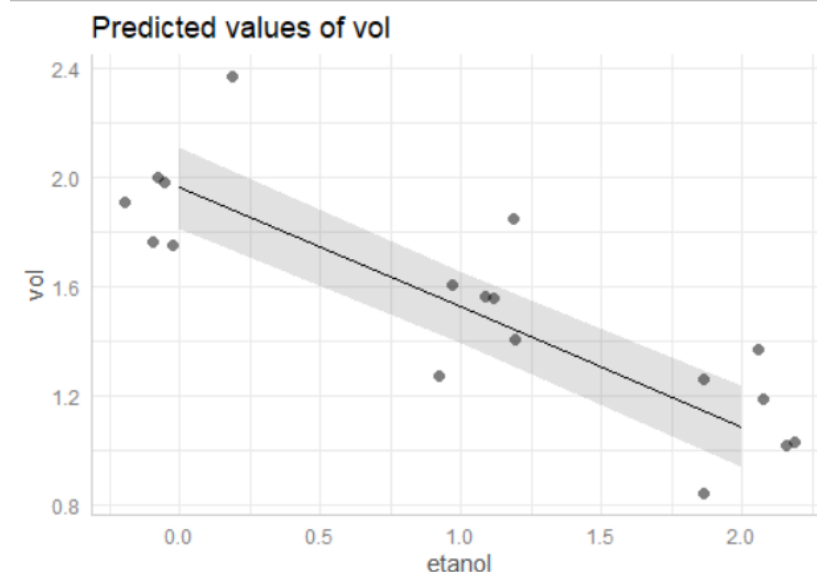
- Interpretación de los coeficientes?
- Estimación de la respuesta media para las tres dosis de etanol?

```
> intervals(m2b)
```

Approximate 95% confidence intervals

Fixed effects:

| | lower | est. | upper |
|-------------|-----------|---------|-----------|
| (Intercept) | 1.795814 | 1.9625 | 2.129186 |
| etanol | -0.512055 | -0.4375 | -0.362945 |



Parte aleatoria

Efectos aleatorios

18

```
(Intercept) etanol
1 1.955556 -0.4375
2 1.844453 -0.4375
3 2.066659 -0.4375
4 1.983332 -0.4375
5 1.775014 -0.4375
6 2.149986 -0.4375
```

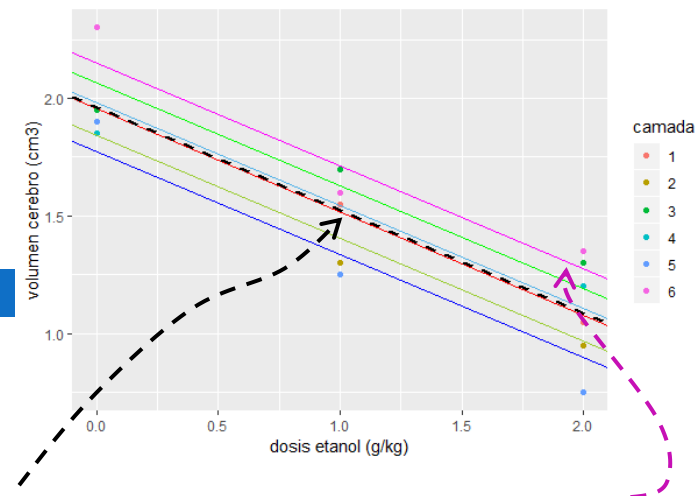
1.96250

```
ranef(m2)  $B_j$ 
$camada
1 -0.0069
2 -0.1180
3 0.1042
4 0.0208
5 -0.1875
6 0.1875
```

| | etanol | camada | vol | pred fija | efecto aleat | pred fija + aleat |
|----|--------|--------|------|-----------|--------------|-------------------|
| 1 | 0 | 1 | 1.95 | 1.9625 | -0.0069 | 1.9556 |
| 2 | 0 | 2 | 1.90 | 1.9625 | -0.1180 | 1.8445 |
| 3 | 0 | 3 | 1.95 | 1.9625 | 0.1042 | 2.0667 |
| 4 | 0 | 4 | 1.85 | 1.9625 | 0.0208 | 1.9833 |
| 5 | 0 | 5 | 1.90 | 1.9625 | -0.1875 | 1.7750 |
| 6 | 0 | 6 | 2.30 | 1.9625 | 0.1875 | 2.1500 |
| 7 | 1 | 1 | 1.55 | 1.5250 | -0.0069 | 1.5181 |
| 8 | 1 | 2 | 1.30 | 1.5250 | -0.1180 | 1.4070 |
| 9 | 1 | 3 | 1.70 | 1.5250 | 0.1042 | 1.6292 |
| 10 | 1 | 4 | 1.60 | 1.5250 | 0.0208 | 1.5458 |
| 11 | 1 | 5 | 1.25 | 1.5250 | -0.1875 | 1.3375 |
| 12 | 1 | 6 | 1.60 | 1.5250 | 0.1875 | 1.7125 |
| 13 | 2 | 1 | 1.05 | 1.0875 | -0.0069 | 1.0806 |
| 14 | 2 | 2 | 0.95 | 1.0875 | -0.1180 | 0.9695 |
| 15 | 2 | 3 | 1.30 | 1.0875 | 0.1042 | 1.1917 |
| 16 | 2 | 4 | 1.20 | 1.0875 | 0.0208 | 1.1083 |
| 17 | 2 | 5 | 0.75 | 1.0875 | -0.1875 | 0.9000 |
| 18 | 2 | 6 | 1.35 | 1.0875 | 0.1875 | 1.2750 |

Predicciones
marginales

Predicciones
sujeto específicas



Parte aleatoria

Efectos aleatorios (BLUPs)

$$BLUP_j = BLUE_j \left(\frac{\sigma_{camadas}^2}{\sigma_{camadas}^2 + \sigma^2 / n_i} \right)$$

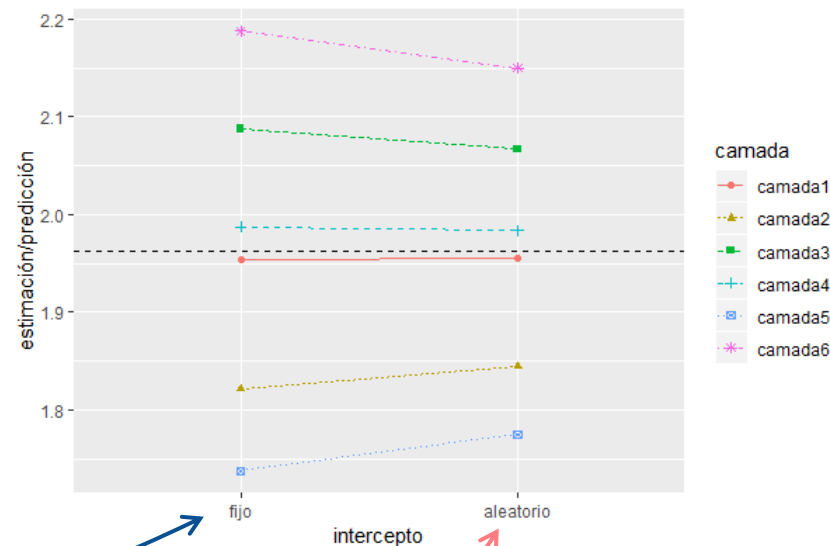
19

Fixed effects:

| | Estimate | Std. Error | t value |
|-------------|----------|------------|---------|
| (Intercept) | 1.96250 | 0.07573 | 25.91 |
| etanol | -0.43750 | 0.03387 | -12.92 |

Los interceptos fijos y aleatorios difieren:

| camada | interc_fijo | interc_aleat |
|---------|-------------|--------------|
| camada1 | 1.954167 | 1.955556 |
| camada2 | 1.820833 | 1.844453 |
| camada3 | 2.087500 | 2.066659 |
| camada4 | 1.987500 | 1.983332 |
| camada5 | 1.737500 | 1.775014 |
| camada6 | 2.187500 | 2.149986 |



y por lo tanto los efectos de las camadas:

| | camada | BLUE | BLUP |
|-------|---------|--------------|--------------|
| B_j | camada1 | -0.008333333 | -0.006943935 |
| | camada2 | -0.141666667 | -0.118046892 |
| | camada3 | 0.125000000 | 0.104159022 |
| | camada4 | 0.025000000 | 0.020831804 |
| | camada5 | -0.225000000 | -0.187486240 |
| | camada6 | 0.225000000 | 0.187486240 |

Los interceptos aleatorios están “encogidos” (shrinkage), es decir son más parecidos a la media. Es el costo de ampliar la inferencia

Parte aleatoria

Componentes de varianza

- Se denomina así a cada varianza que contribuye a la varianza aleatoria de Y y mide la variabilidad aportada por cada VE con efectos aleatorios- En este caso

$$\sigma_{Y_{ij}}^2 = \sigma_{camadas}^2 + \sigma^2$$

Random effects:

| Groups | Name | Variance | Std.Dev. |
|----------|-------------|----------|----------|
| camada | (Intercept) | 0.02294 | 0.1515 |
| Residual | | 0.01377 | 0.1173 |

$$\hat{\sigma}_{Y_{ij}}^2 = 0,02294 + 0,01377 = 0,03671$$

$$\hat{\sigma}_{Y_{ij}} = 0,19cm^3$$

> intervals(m2b)

Approximate 95% confidence intervals

Random Effects:

Level: camada

| | lower | est. | upper |
|-----------------|------------|-----------|-----------|
| sd((Intercept)) | 0.07164942 | 0.1514534 | 0.3201442 |

Within-group standard error:

| | lower | est. | upper |
|--|------------|------------|------------|
| | 0.07726336 | 0.11734111 | 0.17820783 |

- La estructura de correlación inducida por el factor aleatorio se denomina **coeficiente de correlación intraclase**. Se calcula como

$$CCI = \frac{\sigma_{camada}^2}{\sigma_{camada}^2 + \sigma^2}$$

- 62% de la variación aleatoria (no impuesta por el tratamiento) en el volumen cerebral de los ratones está explicado por la variación entre camadas

Comparación de modelos

21

Prueba de devianza o cociente de verosimilitudes (LRT)

- Solo para modelos anidados, con distinta estructura de la parte fija o aleatoria

$$D = -2 \ln \frac{\mathcal{L}(\text{modelo } 1)}{\mathcal{L}(\text{modelo } 2)} \sim \chi^2_{p2-p1} \quad \text{donde } m1 \text{ anidado en } m2$$

$p1, p2 = \text{número de parámetros de cada modelo}$

- Para comparar modelos con distinta parte fija: `anova(modelo1, modelo2)`.
También puede usarse el comando `drop1(modelo2)`
- Para determinar la significancia del factor aleatorio `ranova(modelo2)`

```
#library(lmerTest)
ranova(modelo2)
```

ANOVA-like table for random-effects: single term deletions

```
Model:
vol ~ etanol + (1 | camada)
      npar logLik      AIC      LRT Df Pr(>Chisq)
<none>      4 4.4136 -0.8272
(1 | camada)  3 1.3663  3.2673 6.0945  1   0.01356 *
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comparación por AIC/BIC

- No se recomienda comparar AIC calculados con distintos paquetes

Modelo marginal: estructura de correlación entre observaciones de una camada

22

Parte fija

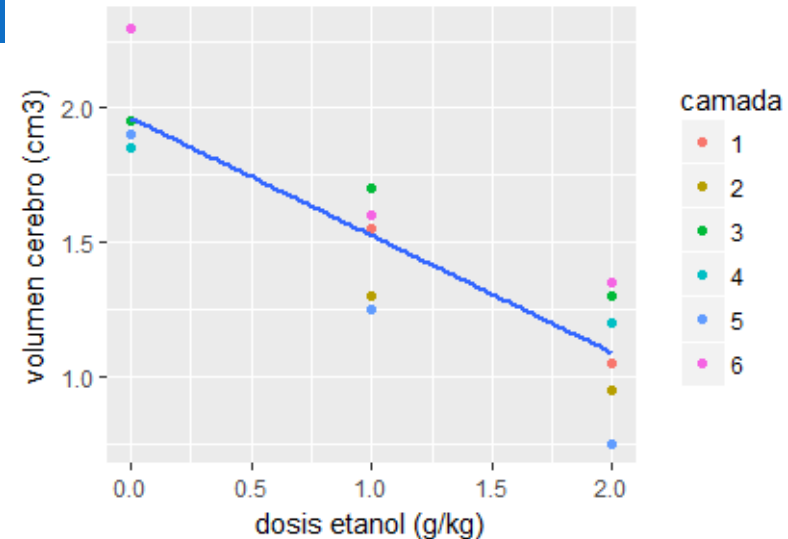
Parte aleatoria

$$Y_{ij} = \beta_0 + \beta_1 \text{etanol}_i + \varepsilon_{ij}$$

$$i = 1 \text{ a } 3, j = 1 \text{ a } 6$$

$$\varepsilon_{ij} \approx NID(0, \Sigma_j)$$

matriz de covarianza



- **Modelo marginal:** ajusta un modelo general para la estructura promedio de la población de individuos
- No incluye VE de efectos aleatorios
- La estructura de correlación entre las observaciones de la misma camada está explicitada en la matriz de covarianza

```
library(nlme)
m2c <- gls(vol ~ etanol, correlation=corCompSymm(form = ~ 1|camada),
data = ratones)
```

Estructura de correlación entre
observaciones de una camada

Generalized least squares fit by REML

Model: vol ~ etanol

Data: ratones

| | AIC | BIC | logLik |
|--|------------|----------|----------|
| | -0.8271654 | 2.263189 | 4.413583 |

Las estimaciones son por máxima
verosimilitud restringida en lugar de
máxima verosimilitud

Correlation Structure: Compound symmetry

Formula: ~1 | camada

Parameter estimate(s):

Rho

0.6248966

Coefficients:

| | Value | Std.Error | t-value | p-value |
|-------------|---------|------------|-----------|---------|
| (Intercept) | 1.9625 | 0.07573224 | 25.91367 | 0 |
| etanol | -0.4375 | 0.03387347 | -12.91571 | 0 |

Residual standard error: 0.1915909

Degrees of freedom: 18 total; 16 residual

No descompone la varianza
entre camadas y dentro de
camadas

Modelo condicional vs modelo marginal

24

- Las estimaciones de la parte fija del modelo son las mismas!
- Con errores con distribución normal ambas aproximaciones son equivalentes. Es decir que postular un modelo mixto con factor aleatorio para la ordenada al origen implica (“es equivalente a”) postular un modelo marginal con cierta estructura de correlación (simetría compuesta)
- Pero los modelos marginales no incluyen efectos aleatorios y por lo tanto no estiman componentes de varianza
- La interpretación no es la misma:
 - ▣ **Modelo marginal:** las predicciones obtenidas son para individuos promedio dadas características fijas (sexo, edad, etc)
 - ▣ **Modelo condicional:** las predicciones obtenidas son para individuos promedio dadas características fijas **y aleatorias** (sujeto-específica)

¿Cuándo tratar a una VE como de efectos aleatorios?

25

- Si no me interesa poner a prueba hipótesis entre los niveles de la variable
- Si no me interesa estimar magnitud del efecto
- Si quiero cuantificar la variabilidad entre los niveles de la variable
- Si es razonable suponer que los niveles analizados de la variable fueron aleatoriamente muestreados o son representativos de una población de niveles
- Si hay suficientes niveles del factor como para efectuar una estimación de confiable de la varianza de la población de los efectos (al menos 5)
- Si son los niveles del factor sólo etiquetas numéricas (no informativos)

Comentarios

26

- Es clave cómo se identifica a los niveles del factor aleatorio para determinar la estructura de agrupamiento. La recomendación es darle una identificación única a cada nivel
- Recordar que siempre son factores (VE categóricas)
- Es importante distinguir entre efectos aleatorios como control de la heterogeneidad del material experimental (como los bloques, incorporados al diseño experimental) y efectos aleatorios de interés (como en estudios genéticos evolutivos o en estudios ecológicos enfocados en la heterogeneidad). En el primer caso se deberían dejar en el modelo, independientemente de la cantidad de niveles y la significación. En el segundo caso podemos aplicar selección de modelos para decir si son relevantes o no
- Otra forma de lidiar con la heterocedasticidad es agregar como VE de efectos aleatorios al individuo (por más que existe una única observación)

Experimentos multiambientales

27

- Muchas veces un experimento se conduce en varios ambientes, donde los ambientes elegidos intentan representar una población relativamente mayor de ambientes. Dentro de cada ambiente se evalúan generalmente dos o más tratamientos bajo un cierto diseño experimental con o sin repeticiones. Los modelos posibles son:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad (1)$$

$$Y_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ij} \quad (2)$$

- los efectos de ambiente (β_j) podrían ser considerados como fijos o aleatorios según los supuestos que se hagan respecto a los ambientes incorporados en el experimento
- El modelo (1) corresponde a un DBA sin interacción
- El modelo (2) corresponde a un modelo con interacción tratamiento x ambiente y es necesario tener réplicas para poder estimarlo. Si los efectos del ambiente son considerados aleatorios, la interacción también lo es
- Ej: interacción genotipo-ambiente