

# Biometría II

## TP N° 5

### Diseño Experimental e Introducción a Modelos Mixtos: Diseños en Bloque, de Factores Anidados y Medidas Repetidas

Hasta ahora hemos trabajado en modelos que provenían de estudios observacionales o experimentales. El denominador común de estos estudios es que los niveles de los factores bajo estudio siempre eran elegidos por el investigador. Incluso factores con niveles que no eran fijados por el investigador (por ejemplo, “sexo”) eran agregados a los modelos porque interesaba conocer específicamente los efectos de los niveles de dicho factor. En estas situaciones, el objetivo siempre es comparar los promedios de los niveles (cuando se trata de VE cualitativas) o los coeficientes con los que se relacionan con la VR (en la VE cuantitativas). Ahora bien... si tenemos una variable explicatoria cualitativa (factor) y si los niveles elegidos no nos interesan particularmente sino que provienen de una población de posibles niveles del factor? ¿Y si lo que nos interesa es estudiar la variabilidad entre los niveles? En estos casos, estamos bajo el estudio de lo que llamaremos VE de efectos *aleatorios*). ¿Y si en un modelo hay VE de efectos aleatorios y VE de efectos fijos? En ese caso estamos trabajando con *modelos mixtos*. Si la variable respuesta sigue una distribución normal y tenemos un modelo mixto, entonces estamos ante un *modelo lineal general* mixto. Si la variable respuesta sigue alguna distribución distinta (por ejemplo Binomial, Poisson) y tenemos un modelo mixto, entonces estamos bajo un *modelo lineal generalizado mixto*.

## Parte A: Diseño Experimental e Introducción a Modelos Mixtos

### Problema 1. Entrada en calor

Para entrar en calor, te proponemos que respondas para el siguiente enunciado:

Se ha sugerido que la diferenciación genética de especies emparentadas de crustáceos del género *Artemia* (habitantes de cuerpos de agua con elevada salinidad) podría relacionarse con diferencias en el contenido y la distribución de la heterocromatina. Como primera aproximación se procedió a estudiar la variabilidad intraespecífica en poblaciones de la región pampeana, para lo cual se eligieron al azar y analizaron 15 lagunas dentro de esta región. En cada una se seleccionaron aleatoriamente 10 individuos pertenecientes a la especie *A. franciscana*. Se fotografiaron 5 células interfásicas de cada individuo y a partir de las células ampliadas se cuantificó el porcentaje de heterocromatina por triplicado. El porcentaje de heterocromatina se obtiene dividiendo la longitud de las bandas de heterocromatina por la longitud total del cromosoma correspondiente.

- Cómo modelarías la variable respuesta a partir del diseño, antes de ver los datos.

$$\%Heterocromatina = Long\ Bandas\ Heterocromatina / LongTotal$$

$$\%Heterocromatina \sim N(\mu, \sigma)$$

- Qué tipo de VEs involucra (efectos fijos, efectos aleatorios).

Cuerpo de Agua; a=1:15; cualitativa aleatoria

Individuos; b=1:10; cualitativa aleatoria

Células; c=1:5; cualitativa aleatoria

Determinaciones; d=1:3

- Qué tipo de modelo es.

Modelo Aleatorio, Diseño completamente anidado.

$$Y_{ijkl} = \mu + A_i + B_{j(i)} + C_{k(j(i))} + \epsilon_{ijkl}$$

$$A_i \sim NID(0, \sigma_a^2)$$

$$B_i \sim NID(0, \sigma_b^2)$$

$$C_i \sim NID(0, \sigma_c^2)$$

$$\epsilon_{ijkl} \sim NID(0, \sigma^2)$$

## Problema 2. Practiquemos un poco, hagamos un cuadro

Independientemente del ensayo, ¿estas VE serían incluidas en un modelo explicatorio como de efectos fijos o aleatorios? (pensemos en generalidades y no en situaciones particulares):

Droga suministrada: Fijo  
 Fertilizante aplicado: Fijo  
 Camada: Aleatorio  
 Sujetos: Aleatorio  
 Día del experimento: Aleatorio  
 Temperatura: Fijo  
 Muestra: Aleatorio  
 Placa de petri: Aleatorio  
 Tubo de ensayo: Aleatorio  
 Parcela: Aleatorio  
 Sexo: Fijo  
 Edad: Fijo

## Problema 3. Diseños experimentales

Para cada uno de los casos de estudio planteados, responda las preguntas correspondientes.

**Caso 1.** En el marco del proyecto “Asistencia técnica para la recuperación y revitalización de los bosques templados de Chile, con énfasis en los *Nothofagus caducifolios*”, ejecutado por la Corporación Nacional Forestal de Chile (CONAF), se quiere controlar al coleóptero *Brachysternus prasinus* que causa intensas defoliaciones sobre ejemplares de *Nothofagus antarctica* (ñire) en el Centro y Sur de Chile. Para ello se llevó a cabo un experimento a fin de determinar la eficacia de dos insecticidas para el control de esta plaga, en un bosque natural de ñire. El gráfico muestra distintos diseños experimentales, donde las x representan a los árboles y los símbolos corresponden a los distintos tratamientos (dos insecticidas y un control). En cada árbol se registró la superficie defoliada en 10 hojas seleccionadas al azar. Tengan en cuenta que:

Diseño a): Cada rectángulo representa una parcela de una hectárea. A cada parcela se le asigna al azar un tratamiento y de cada parcela se seleccionan al azar 5 árboles para efectuar las mediciones.

Diseño b): Similar a a) pero por duplicado.

Diseño c): Se seleccionan al azar 5 árboles para cada tratamiento.

¿Se trata de un estudio observacional o experimental? Para cada diseño identifique la unidad experimental, la variable respuesta, la cantidad de replicas y el/los tipo/s de VE y sus niveles. ¿Cómo es la relación entre VE (cruzados, anidados)? ¿Cómo especificaría la estructura de agrupamiento de los datos en R?

Diseño a)

*No se puede discernir entre el efecto de parcela e insecticida. El diseño está mal.*

Diseño b)

DBA.

VR: área defoliada (cuantitativa continua)

Insecticida; 3 niveles; cualitativa fija

Bloque; 2 niveles; cualitativa aleatoria

Árboles; 5 niveles; cualitativa aleatoria; anidada en Bloque x Insecticida

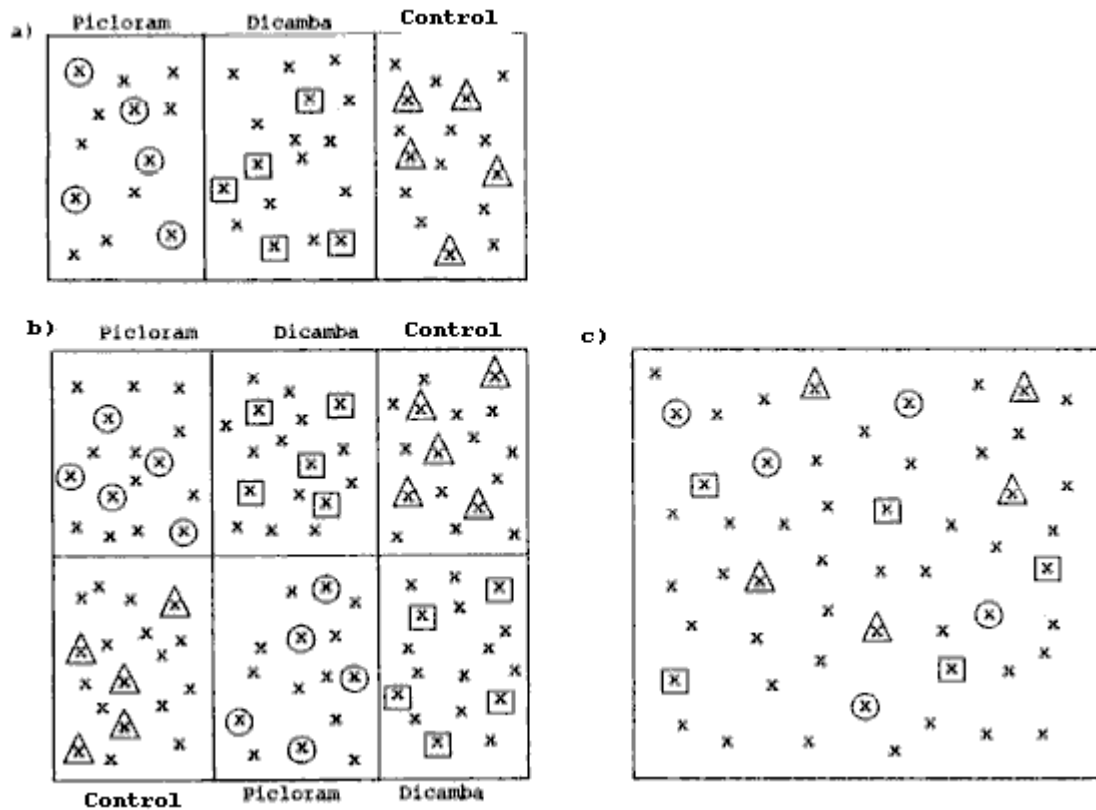


Figure 1: Diseños Experimentales

*U.E: Parcela.*

*Replicas: 2.*

*en R: lmer(area ~ insecticida + (1 | Bloque) + (1 | arbol), datos)*

Diseño c)

*Disenio anidado.*

*VR: area defoliada (cuantitativa continua)*

*Insecticida; 3 niveles; cualitativa fija*

*Arboles; 5 niveles; cualitativa aleatoria anidada en Insecticida*

*U.E: arbol.*

*Replicas: 5.*

*en R: lmer(area ~ insecticida + (1 | arbol), datos)*

**Caso 2.** En la provincia de Chubut existen areas de bosque nativo que han sufrido un gran proceso de degradacion como consecuencia de incendios. Estas perturbaciones han afectado a los bosques de *Austrocedrus chilensis* (Cipres de las Cordilleras) presentes en los alrededores del Lago Puelo. Con el fin de restaurar estos bosques con plantines de cipres se llevo a cabo un experimento a campo a fin de evaluar distintas proporciones de biosolidos en el suelo. Se espera encontrar la proporcion que acelere el crecimiento y mejore la calidad de las plantas una vez que estas sean transplantadas en la zona a restaurar. El ensayo se efectuó en una ladera de características homogéneas al este del lago Puelo. Los plantines se dispusieron en hileras norte-sur y se definieron 20 parcelas de 10 plantas cada una. Cada parcela fue asignada al azar a uno de los siguientes tratamientos, de modo tal de obtener un diseño balanceado: I. Sin biosolidos; II. 25% de biosolidos; III. 50% de biosolidos; IV. 75% de biosolidos. Al año de establecida la plantacion se midió la altura de cada una de las plantas.

- Identifique la unidad experimental, la variable respuesta, el tipo de VE (fijos/aleatorios) y sus niveles, la cantidad de replicas y el tamaño total de la muestra. ¿Cómo especificaría el agrupamiento de los datos en R?.

*VR: Altura de cada planta al año de establecida la plantacion*

*VE: -Proporcion de biosolidos en el suelo. (Cualitativa. Fija. Niveles: 4. Sin biosolidos; 25% de biosolidos; 50% de biosolidos; 75% de biosolidos).*

-Parcelas (Cualitativa. Aleatoria. Niveles: 1:20)

UE: 1 parcela con 10 plantas de cipres en este del Lago Puelo.

replicas: 5

Disenio anidado

en R: `lmer(altura~biosolidos+(1/parcela), Datos)`

- El siguiente es un plano del terreno con las 20 parcelas. Asigne de manera conveniente los tratamientos a las parcelas.

	1	2	3	4	5	
norte	6	7	8	9	10	Sur
	11	12	13	14	15	
	16	17	18	19	20	

Figure 2: Disenios Experimentales2

Se asignan de manera aleatoria entre todas las parcelas.

- ¿Qué cambiaría en el diseño si existiese en el campo experimental un gradiente de humedad norte-sur?.

En ese caso hay que asegurarse que transversalmente al gradiente se ubiquen al azar una replica de cada tratamiento

**Caso 3.** En Pergamino se desea evaluar el efecto de dos dosis del herbicida Round-up sobre la comunidad de arcnidos en cultivos de soja. Para ello, se divide un lote de 90 ha en 9 parcelas de igual tamano. Como de norte a sur existe una pendiente pronunciada, los investigadores deciden aplicar un disenio de bloques (ver esquema): un bloque corresponde a la loma, otro a la media loma y el tercero al bajo. En cuanto a los tratamientos, las parcelas que limitan al oeste corresponden al bloque control (50 l/m2 de agua sin Rond-up), las siguientes tres al bloque dosis 1 (5l de herbicida en 50 l/m2) y las ultimas tres, que limitan al este, corresponden al bloque dosis 2 (15 l de herbicida en 50 l/m2). En cada parcela se registra la abundancia de arcnidos.

	Norte			
	loma	Loma	loma	
	control	dosis 1	dosis 2	
Oeste	m.lom a	m.lom a	m.lom a	Este
	control	dosis 1	dosis 2	
	bajo	Bajo	bajo	
	control	dosis 1	dosis 2	
	Sur			

Figure 3: Disenios Experimentales3

Identifique la unidad experimental, la unidad de observacion, la variable respuesta, el tipo de VE y sus niveles y la cantidad de replicas. ¿Como especificaria el agrupamiento de los datos en R? Discuta el disenio experimental. Si considera que el disenio es incorrecto plantee un disenio apropiado.

VR: Abundancia de arcnidos, cuantitativa continua

VE: -Bloque (cualitativa, efecto aleatorio, 3 niveles: Loma, media loma y bajo)

-herbicida (cualitativa, efecto fijo, 3 niveles: control, dosis 1 y dosis 2)

UE: Cada parcela condicionada por el bloque

UO: Cada parcela

replicas: 3

Disenio bloques al azar

en R: `modelo<- lmer(abundancia~dosis + (1| bloque), datos)`

**Caso 4.** Un investigador esta interesado en comparar la densidad de oligoquetos en suelos de sistemas naturales (SN) y de sistemas agricolas convencionales (SAC) en el sudeste bonaerense. Supongamos tres abordajes metodologicos posibles, definidos por los recursos de los que dispone:

En cada caso, identifique la unidad experimental, la unidad de observacion, la variable respuesta, el tipo de VE y sus niveles y la cantidad de replicas, así como la poblacion sobre la que efectua la inferencia.

- Elige al azar 10 lotes pertenecientes a SN y 10 lotes con SAC ubicados en distintas localidades del sudeste bonaerense.

VR: densidad de oligoquetos

VE: tipo de suelo (cualitativa fija, 2 niveles: SN y SAC)

UE: Lote

UO: Lote, tengo una obs por lote

replicas: 10

Disenio completamente aleatorizado

poblacion sobre la que efectua la inferencia: La densidad de oligoquetos de todos los lotes SN y SA de las localidades del sudeste bonaerense

en R: `lm(densidad~tipo de suelo, bd)`

- En Balcarce selecciona 10 lotes con SN de todos los lotes con SN existentes en la localidad y 10 lotes con SAC de todos los lotes con SAC.

VR: densidad de oligoquetos

VE: tipo de suelo (cualitativa fija, 2 niveles: SN y SAC)

UE: Lote

UO: Lote, tengo una obs por lote

replicas: 10

Disenio completamente aleatorizado

poblacion sobre la que efectua la inferencia: La densidad de oligoquetos de todos los lotes SN y SAC de la localidad de Balcarce

en R: `lm(densidad~tipo de suelo, bd)`

- En cierta localidad selecciona un lote con SN de todos los lotes con SN existentes en la localidad y un lote con SAC de todos los lotes con SAC. De cada lote extrae 10 muestras de suelo.

VR: densidad de oligoquetos

VE: tipo de suelo (cualitativa fija, 2 niveles: SN y SAC), lote (cualitativa aleatoria)

UE: Lote

UO: Muestras de suelo (no son independientes entre si, vienen de la misma U.E)

replicas: 1

Disenio anidado

poblacion sobre la que efectua la inferencia: La densidad de oligoquetos del único lote elegido de SN y SAC de cierta localidad

en R: `lmer(densidad~tipo de suelo +(1|lote, bd)`

## Parte B: Diseño en Bloques y Modelos con Factores Anidados

### Problema 4. Modelos condicionales y marginales

Con el objetivo de estudiar el efecto del suministro de una dieta mixta compuesta por alfalfa y avena sobre la absorcion de glucosa y resistencia a la insulina en equinos, se desarrollo una experiencia en un grupo de 16 equinos, todos clinicamente sanos. Se separaron al azar en dos grupos, y los animales de cada grupo fueron sometidos a un regimen alimenticio y de manejo distinto:

Lote A (A campo): 9 equinos fueron alimentados exclusivamente con pasturas naturales.

Lote B (Estabulados): 7 equinos fueron alimentados en corral con heno de alfalfa y grano de avena.

Luego de dos meses de tratamiento se efectuo en cada animal la prueba de tolerancia oral a la glucosa (PTOG). La prueba consiste en administrar un concentrado de glucosa y tomar muestras de sangre cada 30 minutos luego del suministro durante 8 horas. Se compararon las curvas de glucemia de ambos grupos. Base de datos en *PTOG.txt*.

- Identifique la unidad experimental, la variable respuesta, factores y niveles y condicion de fijos o aleatorios. ¿Dónde interviene el azar en este ensayo?.

VR: *glucemia medida por PTOG*

VE: *alimentacion (cualitativa fija, 2 niveles: estabulado, pastura, cruzada con tiempo), tiempo (cualitativa fija, 4 niveles: 0,30, 60, 90, cruzada con tiempo), equino (cualitativa aleatoria, 16 niveles, anidada en tratamiento)*

UE: *equino*

- Plantee el modelo condicional y marginal en terminos estadisticos y del problema. Compare ambas formas de modelar la estructura de agrupamiento de los datos.

## modelo condicional

$$Y_{ijk} = \mu + \alpha_i + B_{j(i)} + \gamma_k + \alpha\gamma_{ik} + \epsilon_{ijk}$$

$$B_{j(i)} \sim NID(0, \sigma_{equinos}^2)$$

$$\epsilon_{ijk} \sim NID(0, \sigma^2)$$

$$i : 1 : 2$$

$$j|i_1 : 1 : 9$$

$$j|i_2 : 1 : 7$$

$$k : 1 : 4$$

## modelo marginal

$$Y_{ijk} = \mu + \alpha_i + \beta_k + \alpha\beta_{ik} + \epsilon_{ijk}$$

$$B_{j(i)} \sim NID(0, \sigma_{equinos}^2)$$

$$\epsilon_{ijk} \sim NID(0, \sum_j)$$

$$i : 1 : 2$$

$$j|i_{=1} : 1 : 9$$

$$j|i_{=2} : 1 : 7$$

$$k : 1 : 4$$

$$\text{matriz simetria compuesta} = \begin{pmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{pmatrix} * \sigma^2$$

En el modelo condicional se explicitan terminos aleatorios y nos permite estimar componentes de varianza mientras que en los modelos marginales no se incluyen factores aleatorios y la estructura de los datos se especifica en la matriz de varianza-covarianza. Sin embargo, un modelo condicional implica una matriz de varianza-covarianza de simetria compuesta y no siempre es la adecuada para describir un disenio experimental.

- Resuelva ambos modelos y evalúe los resultados obtenidos con ambos. ¿Con cual de los dos modelos se quedaria?

```
library(readxl)
PTOG <- read_excel("/home/pangolin/Documents/materias/biome2/2020/tps/tp5/nuevo/PTOG.xlsx")
str(PTOG)
```

```
## tibble [64 x 4] (S3: tbl_df/tbl/data.frame)
## $ Tratamiento: chr [1:64] "A campo" "A campo" "A campo" "A campo" ...
## $ Equino      : chr [1:64] "AL" "BA" "MA" "PR" ...
## $ Tiempo      : num [1:64] 0 0 0 0 0 0 0 0 0 0 ...
## $ Glucemia    : num [1:64] 85.1 72.3 72.3 60.8 77.4 ...
```

```
# Equino y Tratamiento como factor
PTOG$Equino <- factor(PTOG$Equino)
PTOG$Tratamiento <- factor(PTOG$Tratamiento)
PTOG$Tiempo <- factor(PTOG$Tiempo)
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

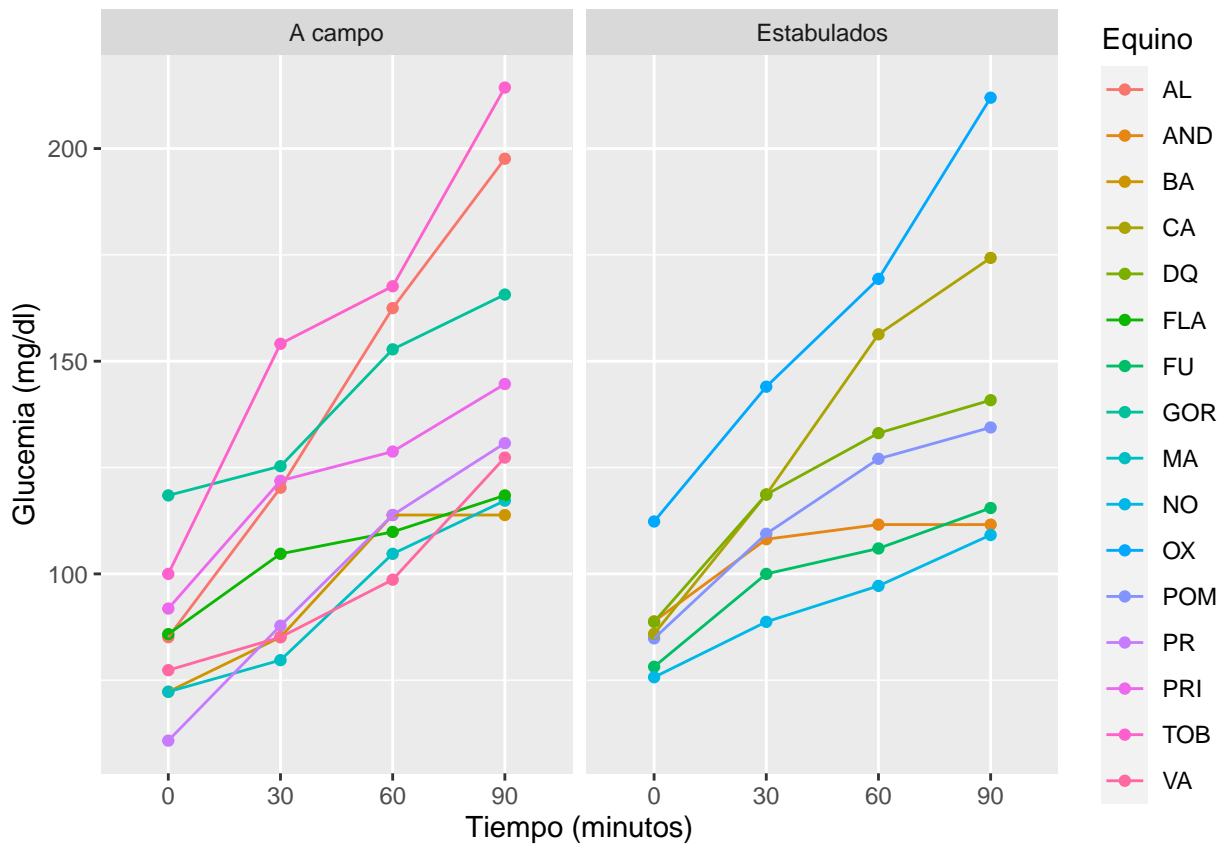
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
# lo ordeno para ver todas las observaciones de los equinos
ordenado <- arrange(PTOG, Equino, Tiempo)
head(ordenado, 8)
```

```
## # A tibble: 8 x 4
##   Tratamiento Equino Tiempo Glucemia
##   <fct>      <fct>   <fct>    <dbl>
## 1 A campo    AL      0      85.1
## 2 A campo    AL     30     120.
## 3 A campo    AL     60     162.
## 4 A campo    AL     90     198.
## 5 Estabulados AND    0      88.8
## 6 Estabulados AND    30     108.
## 7 Estabulados AND    60     112.
## 8 Estabulados AND    90     112.
```

```
# Medidas repetidas Spagueti plots
```

```
library(ggplot2)
p <- ggplot(PTOG, aes(Tiempo, Glucemia, group = Equino))
p + geom_point(aes(color = Equino)) + geom_line(aes(color = Equino)) +
  facet_grid(. ~ Tratamiento) + labs(x = "Tiempo (minutos)", y = "Glucemia (mg/dl)")
```



## modelo condicional

```
#### Modelo condicional usando un factor aleatorio para inducir la
#### estructura de correlación
library(nlme)
```

```
##
## Attaching package: 'nlme'
## The following object is masked from 'package:dplyr':
##
## collapse
m1a <- lme(Glucemia ~ Tratamiento * Tiempo, random = ~1 | Equino,
  PTOG)
summary(m1a)
```

```
## Linear mixed-effects model fit by REML
## Data: PTOG
##      AIC      BIC    logLik
## 522.2882 542.5418 -251.1441
##
## Random effects:
## Formula: ~1 | Equino
## (Intercept) Residual
## StdDev:    22.83379 13.49951
##
## Fixed effects: Glucemia ~ Tratamiento * Tiempo
##
##              Value Std.Error DF   t-value p-value
## (Intercept)  84.89000   8.841939 42   9.600835  0.0000
## TratamientoEstabulados  2.89714  13.367755 14   0.216726  0.8315
## Tiempo30        22.23222   6.363730 42   3.493583  0.0011
```



```

## Tiempo60          43.17222  6.363730 42  6.784106  0.0000
## Tiempo90          62.87111  6.363730 42  9.879600  0.0000
## TratamientoEstabulados:Tiempo30  2.50063  9.621056 42  0.259913  0.7962
## TratamientoEstabulados:Tiempo60 -2.30508  9.621056 42 -0.239587  0.8118
## TratamientoEstabulados:Tiempo90 -8.12254  9.621056 42 -0.844246  0.4033
## Correlation:
##              (Intr) TrtmnE Timp30 Timp60 Timp90 TE:T30
## TratamientoEstabulados      -0.661
## Tiempo30                    -0.360  0.238
## Tiempo60                    -0.360  0.238  0.500
## Tiempo90                    -0.360  0.238  0.500  0.500
## TratamientoEstabulados:Tiempo30  0.238 -0.360 -0.661 -0.331 -0.331
## TratamientoEstabulados:Tiempo60  0.238 -0.360 -0.331 -0.661 -0.331  0.500
## TratamientoEstabulados:Tiempo90  0.238 -0.360 -0.331 -0.331 -0.661  0.500
##              TE:T60
## TratamientoEstabulados
## Tiempo30
## Tiempo60
## Tiempo90
## TratamientoEstabulados:Tiempo30
## TratamientoEstabulados:Tiempo60
## TratamientoEstabulados:Tiempo90  0.500
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -1.74572983 -0.51350831 -0.07204079  0.52115311  2.31297771
##
## Number of Observations: 64
## Number of Groups: 16

# devuelve ef fijos
a = fitted(m1a, level = 0)
a[names(a) == "AL"]

##           AL           AL           AL           AL
## 84.8900 107.1222 128.0622 147.7611

# devuelve ef fijos + aleatorios
b = fitted(m1a, level = 1)
b[names(b) == "AL"]

##           AL           AL           AL           AL
## 107.3394 129.5717 150.5117 170.2106

# miramos el efecto aleatorio solo del primer equino
ranef(m1a)[[1]][1]

## [1] 22.44944

# vemos que la diferencia es la misma
b[names(b) == "AL"] - a[names(a) == "AL"]

##           AL           AL           AL           AL
## 22.44944 22.44944 22.44944 22.44944

library(mgcv)

## This is mgcv 1.8-31. For overview type 'help("mgcv-package")'.

# extraemos mariz var-covar para el primer equino
extract.lme.cov2(m1a, PTOG, start.level = 1)[[1]][1]

## [[1]]
##           [,1]           [,2]           [,3]           [,4]

```

```
## [1,] 703.6190 521.3822 521.3822 521.3822
## [2,] 521.3822 703.6190 521.3822 521.3822
## [3,] 521.3822 521.3822 703.6190 521.3822
## [4,] 521.3822 521.3822 521.3822 703.6190
```

```
# obtenemos matriz var-covar de otra manera
VarCorr(m1a)
```

```
## Equino = pdLogChol(1)
##          Variance StdDev
## (Intercept) 521.3822 22.83379
## Residual    182.2368 13.49951
```

```
# la correlacion entre 2 observaciones del mismo equino
(icc <- 22.83^2/(22.83^2 + 13.5^2))
```

```
## [1] 0.740923
```

## modelo marginal

```
#### Modelo marginal usando simetría compuesta
```

```
m2 <- gls(Glucemia ~ Tratamiento * Tiempo, correlation = corCompSymm(form = ~1 |
  Equino), PTOG)
summary(m2)
```

```
## Generalized least squares fit by REML
## Model: Glucemia ~ Tratamiento * Tiempo
## Data: PTOG
##      AIC      BIC    logLik
## 522.2882 542.5418 -251.1441
##
```

```
## Correlation Structure: Compound symmetry
## Formula: ~1 | Equino
## Parameter estimate(s):
##      Rho
## 0.7410007
##
```

```
## Coefficients:
```

	Value	Std.Error	t-value	p-value
## (Intercept)	84.89000	8.841939	9.600835	0.0000
## TratamientoEstabulados	2.89714	13.367755	0.216726	0.8292
## Tiempo30	22.23222	6.363730	3.493583	0.0009
## Tiempo60	43.17222	6.363730	6.784106	0.0000
## Tiempo90	62.87111	6.363730	9.879600	0.0000
## TratamientoEstabulados:Tiempo30	2.50063	9.621056	0.259913	0.7959
## TratamientoEstabulados:Tiempo60	-2.30508	9.621056	-0.239587	0.8115
## TratamientoEstabulados:Tiempo90	-8.12254	9.621056	-0.844246	0.4021

```
##
## Correlation:
##      (Intr) TrtmnE Timp30 Timp60 Timp90 TE:T30
## TratamientoEstabulados -0.661
## Tiempo30 -0.360 0.238
## Tiempo60 -0.360 0.238 0.500
## Tiempo90 -0.360 0.238 0.500 0.500
## TratamientoEstabulados:Tiempo30 0.238 -0.360 -0.661 -0.331 -0.331
## TratamientoEstabulados:Tiempo60 0.238 -0.360 -0.331 -0.661 -0.331 0.500
## TratamientoEstabulados:Tiempo90 0.238 -0.360 -0.331 -0.331 -0.661 0.500
##      TE:T60
## TratamientoEstabulados
## Tiempo30
## Tiempo60
```

```
## Tiempo90
## TratamientoEstabulados:Tiempo30
## TratamientoEstabulados:Tiempo60
## TratamientoEstabulados:Tiempo90 0.500
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -1.2787961 -0.7378451 -0.1137975  0.5599576  2.6172346
##
## Residual standard error: 26.52582
## Degrees of freedom: 64 total; 56 residual
```

- compare el modelo marginal con simetria compuesta y el condicional. ¿Que puede decir acerca de los coeficientes de los modelos?.

*ambos cuando la distribucion propuesta es normal, estiman de igual manera los coeficientes fijos. el condicional induce una matriz simetria compuesta pero explicita factores aleatorios. el marginal no incluye factores aleatorios pero la estructura esta dada en la matriz.*

-Y acerca de las varianzas? Calcule la varianza total no explicada por cada modelo.

*el condicional permite calcular componentes de varianza, pero la suma de las varianzas hace a la varianza estimada por el modelo marginal.*

```
(var_tot_condicional <- (22.83^2 + 13.5^2))
```

```
## [1] 703.4589
```

```
(var_tot_marginal <- 26.52^2)
```

```
## [1] 703.3104
```

-¿Cuál es el equivalente al ICC en el modelo marginal?

*el equivalente es el  $\rho$  (Rho, se ve en el summary).*

**tambien podriamos modelar varianzas, dejamos un ejemplo**

```
#### BONUS ROMPE-CABEZAS con lme() podemos incluso combinar
#### estructuras marginales y condicionales. Por ej. podriamos
#### combinar un efecto aleatorio de Equino con una estructura de
#### modelado de varianza segun el Tiempo
m3 <- lme(Glucemia ~ Tratamiento * Tiempo, random = ~1 | Equino, weights = varIdent(form = ~1 |
  Tiempo), PTOG)
summary(m3)
```

```
## Linear mixed-effects model fit by REML
## Data: PTOG
##      AIC      BIC    logLik
## 514.4681 540.7977 -244.2341
##
## Random effects:
## Formula: ~1 | Equino
##      (Intercept) Residual
## StdDev:    21.63363 12.60337
##
## Variance function:
## Structure: Different standard deviations per stratum
## Formula: ~1 | Tiempo
## Parameter estimates:
##      0      30      60      90
## 1.0000000 0.2906156 0.9016041 1.6390943
## Fixed effects: Glucemia ~ Tratamiento * Tiempo
```

```
##                               Value Std.Error DF   t-value p-value
## (Intercept)                 84.89000   8.345715 42 10.171687  0.0000
## TratamientoEstabulados       2.89714  12.617535 14  0.229612  0.8217
## Tiempo30                     22.23222   4.374935 42  5.081727  0.0000
## Tiempo60                     43.17222   5.656542 42  7.632264  0.0000
## Tiempo90                     62.87111   8.066406 42  7.794191  0.0000
## TratamientoEstabulados:Tiempo30  2.50063   6.614280 42  0.378066  0.7073
## TratamientoEstabulados:Tiempo60 -2.30508   8.551888 42 -0.269540  0.7888
## TratamientoEstabulados:Tiempo90 -8.12254  12.195260 42 -0.666041  0.5090
## Correlation:
##                               (Intr) TrtmnE Timp30 Timp60 Timp90 TE:T30
## TratamientoEstabulados       -0.661
## Tiempo30                     -0.483   0.320
## Tiempo60                     -0.374   0.247   0.713
## Tiempo90                     -0.262   0.173   0.500   0.387
## TratamientoEstabulados:Tiempo30  0.320 -0.483 -0.661 -0.472 -0.331
## TratamientoEstabulados:Tiempo60  0.247 -0.374 -0.472 -0.661 -0.256   0.713
## TratamientoEstabulados:Tiempo90  0.173 -0.262 -0.331 -0.256 -0.661   0.500
##                               TE:T60
## TratamientoEstabulados
## Tiempo30
## Tiempo60
## Tiempo90
## TratamientoEstabulados:Tiempo30
## TratamientoEstabulados:Tiempo60
## TratamientoEstabulados:Tiempo90  0.387
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -2.2620439 -0.4939192 -0.1007537  0.4871216  1.8107094
##
## Number of Observations: 64
## Number of Groups: 16
```

## Problema 5. Habito de fumar

El habito de fumar suele provocar desordenes en muchos aspectos de la vida del ser humano. Una de las consecuencias reportadas es la alteracion en los patrones de descanso. En un estudio se investigo la asociacion entre el habito de fumar y las horas de suenio diarias (descanso), con el proposito de evaluar si el desorden en el descanso difiere segun el sexo. Para ello se seleccionaron al azar hombres y mujeres adultos fumadores y no fumadores (tres de cada combinacion). A cada uno de ellos se le determino el numero de horas de suenio en cuatro dias cualesquiera elegidos al azar. Base de datos en *HF.txt*.

- Identifique: tipo de estudio (observacional/experimental), variable respuesta, unidad experimental o unidad de observacion, factores y niveles, fuentes de variacion, tipo de factores (fijo/aleatorios, cruzados/anidados), disenio empleado, fuentes de variacion que aportan a la varianza de Y.

VR: Suenio diario (horas)

VE: Sexo, cualitativa fija (2 niveles, cruzada con habito), habito fumar, cualitativa fija (2 niveles, cruzada con sexo), Individuo, cualitativa aleatoria (12 niveles, anidado en la interaccion sexo y habito).

Tipo de estudio: Observacional

UE: Individuo

Replicas: 3

- Indique el modelo estadistico utilizado, en términos teóricos y aplicado a este experimento.

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + B_{k(ij)} + \epsilon_{ijkl}$$

i=1:2, j=1:2, k=1:3, l=1:4

$$B_k \sim NID(0, \sigma^2 \text{individuos})$$

$$\epsilon_{ijkl} \sim NID(0, \sigma^2)$$

- Escribir el modelo y plantear las hipótesis correspondientes en términos del problema. Resolver verificando los supuestos y concluir. Realice comparaciones de ser necesario y cuando corresponda. Calcule los componentes de varianza y concluya.

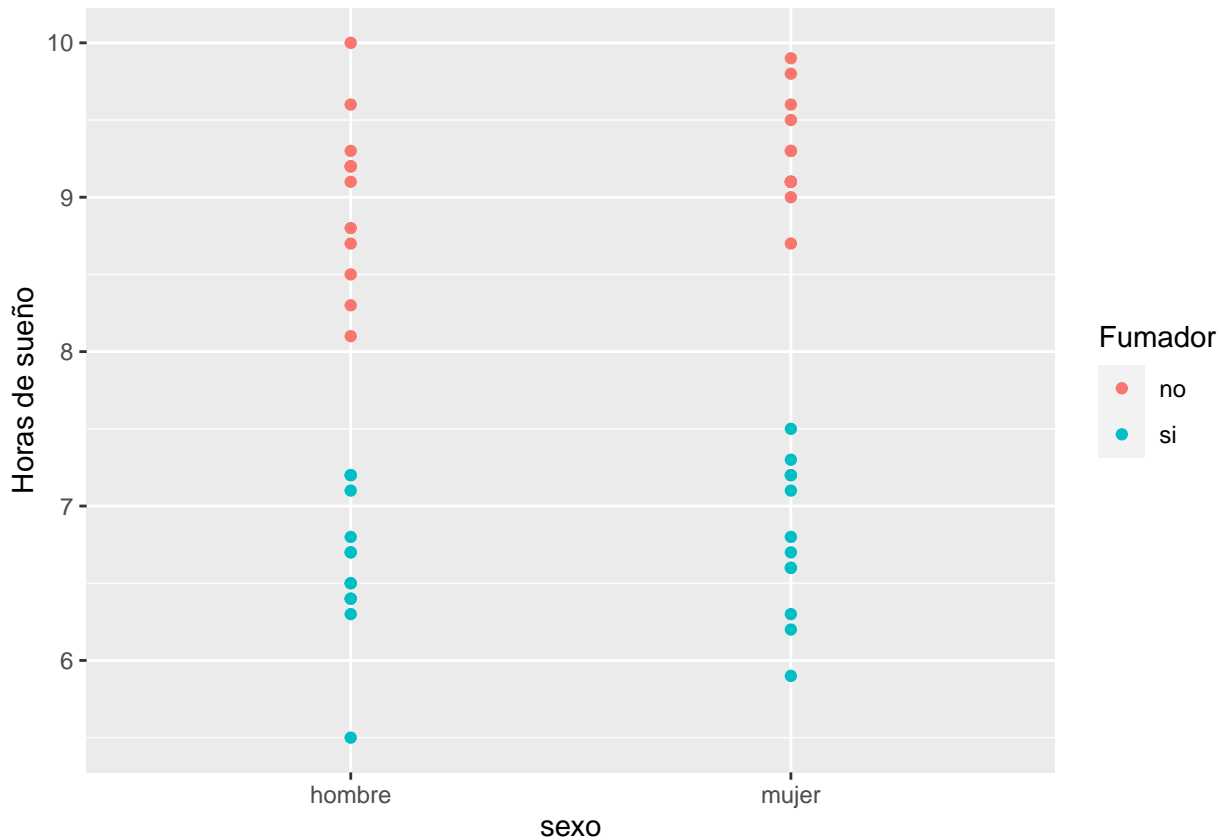
```
Datos <- read.delim("/home/pangolin/Documents/materias/biome2/2018/tps/tp5/parteB/HF.txt")
names(Datos)
```

```
## [1] "sexo"          "Individuo"      "Fumador"        "horas.suenio"
```

```
# Explorando...
```

```
library(ggplot2)
```

```
ggplot(Datos, aes(sexo, horas.suenio, color = Fumador)) + geom_point() +
  labs(y = "Horas de sueño")
```



```
# Para calcular las medias de horas sueño para cada combinación de
# niveles de los dos factores
(medias.int <- aggregate(horas.suenio ~ sexo * Fumador, Datos, mean))
```

```
##      sexo Fumador horas.suenio
## 1 hombre      no      9.000000
## 2  mujer      no      9.291667
## 3 hombre      si      6.608333
## 4  mujer      si      6.783333
```

```
# Planteando el modelo... Incorporaremos individuo como factor ¿De
# que tipo es? ¿Como se relaciona con los otros factores?
```

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
##
```

```
## Attaching package: 'lme4'

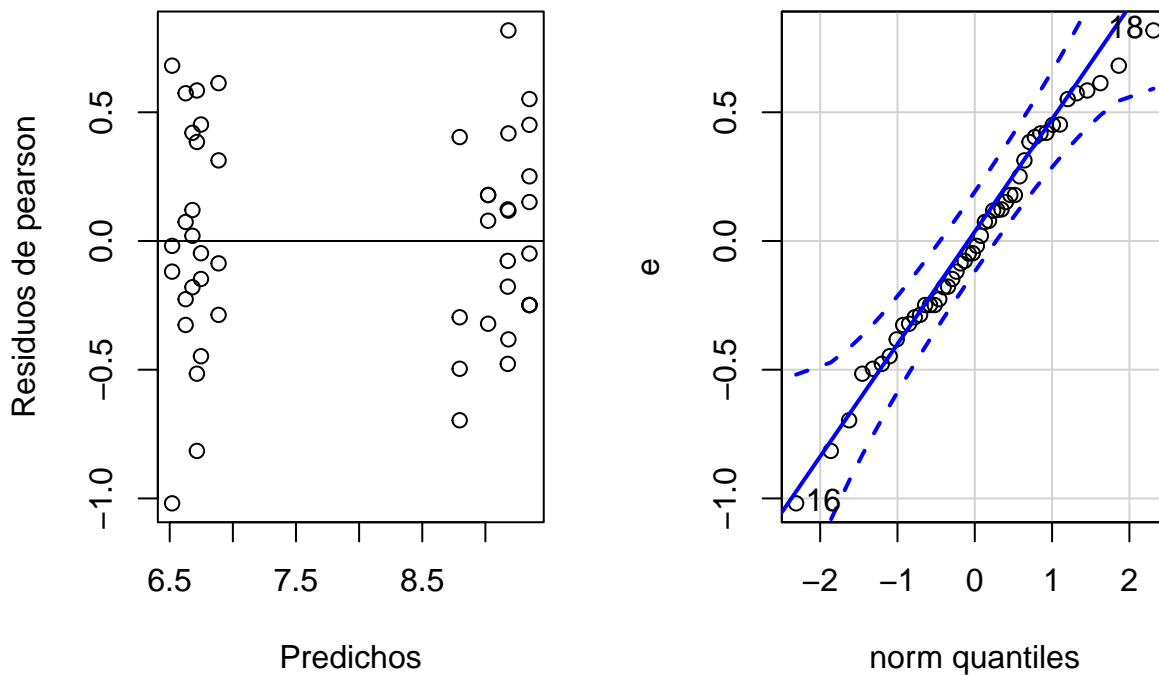
## The following object is masked from 'package:nlme':
##
##      lmList

modelo <- lmer(horas.suenio ~ sexo * Fumador + (1 | Individuo), Datos)

# Supuestos
e <- resid(modelo) # residuos de pearson
pre <- predict(modelo) #predichos
par(mfrow = c(1, 2))
plot(pre, e, xlab = "Predichos", ylab = "Residuos de pearson", main = "Gráfico de dispersión de RE vs PRED",
     cex.main = 0.8)
abline(0, 0)
car::qqPlot(e)

## Registered S3 methods overwritten by 'car':
##      method                                from
##      influence.merMod                      lme4
##      cooks.distance.influence.merMod      lme4
##      dfbeta.influence.merMod              lme4
##      dfbetas.influence.merMod            lme4
```

**Gráfico de dispersión de RE vs PRED**



```
## [1] 16 18
shapiro.test(e)

##
## Shapiro-Wilk normality test
##
## data: e
## W = 0.98679, p-value = 0.8602
```

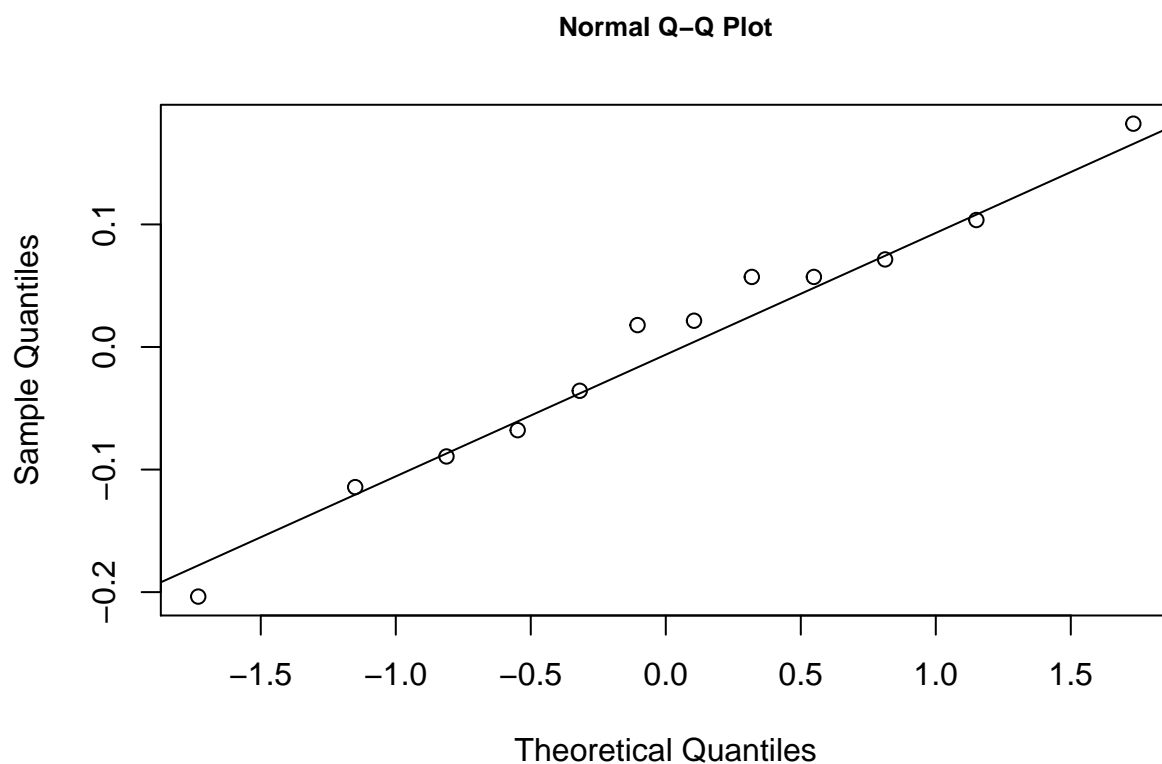
```
# ¿Que supuesto quedaria probar?
(alfai <- ranef(modelo)$Individuo$(Intercept))

## [1] -0.20362698  0.01786202  0.02143442 -0.08931008  0.18219256  0.07144806
## [7] -0.11431690 -0.06787566  0.05715845  0.10359969  0.05715845 -0.03572403

shapiro.test(alfai)

##
## Shapiro-Wilk normality test
##
## data:  alfai
## W = 0.98176, p-value = 0.9897

qqnorm(alfai, cex.main = 0.8)
qqline(alfai)
```



```
anova(modelo)

## Analysis of Variance Table
##          npar Sum Sq Mean Sq F value
## sexo          1  0.373    0.373   1.9251
## Fumador        1 41.152   41.152 212.2431
## sexo:Fumador    1  0.023    0.023   0.1203

summary(modelo)

## Linear mixed model fit by REML ['lmerMod']
## Formula: horas.suenio ~ sexo * Fumador + (1 | Individuo)
##   Data: Datos
##
## REML criterion at convergence: 67.1
##
## Scaled residuals:
```

```
##      Min      1Q   Median      3Q      Max
## -2.31424 -0.58673 -0.07566  0.75157  1.85727
##
## Random effects:
##   Groups      Name      Variance Std.Dev.
## Individuo (Intercept) 0.03637  0.1907
## Residual              0.19389  0.4403
## Number of obs: 48, groups: Individuo, 12
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)      9.0000      0.1682  53.517
## sexomujer         0.2917      0.2378   1.226
## Fumadorsi        -2.3917      0.2378 -10.056
## sexomujer:Fumadorsi -0.1167      0.3363  -0.347
##
## Correlation of Fixed Effects:
##              (Intr) sexmjr Fumdrs
## sexomujer    -0.707
## Fumadorsi    -0.707  0.500
## sxmjr:Fmdrs  0.500 -0.707 -0.707

# Parte fija
fitted(modelo) #predicciones parte fija + aleatoria

##      1      2      3      4      5      6      7      8
## 8.796373 8.796373 8.796373 8.796373 6.626195 6.626195 6.626195 6.626195
##      9     10     11     12     13     14     15     16
## 9.021434 9.021434 9.021434 9.021434 6.519023 6.519023 6.519023 6.519023
##     17     18     19     20     21     22     23     24
## 9.182193 9.182193 9.182193 9.182193 6.679781 6.679781 6.679781 6.679781
##     25     26     27     28     29     30     31     32
## 9.177350 9.177350 9.177350 9.177350 6.715458 6.715458 6.715458 6.715458
##     33     34     35     36     37     38     39     40
## 9.348825 9.348825 9.348825 9.348825 6.886933 6.886933 6.886933 6.886933
##     41     42     43     44     45     46     47     48
## 9.348825 9.348825 9.348825 9.348825 6.747609 6.747609 6.747609 6.747609

fixef(modelo) #estimacion parametros efectos fijos

##      (Intercept)      sexomujer      Fumadorsi sexomujer:Fumadorsi
##      9.0000000      0.2916667      -2.3916667      -0.1166667

X <- model.matrix(modelo)
pred_fija <- X %*% fixef(modelo) #predicciones parte fija
pred_fija

##      [,1]
## 1  9.000000
## 2  9.000000
## 3  9.000000
## 4  9.000000
## 5  6.608333
## 6  6.608333
## 7  6.608333
## 8  6.608333
## 9  9.000000
## 10 9.000000
## 11 9.000000
## 12 9.000000
## 13 6.608333
```



```
## 14 6.608333
## 15 6.608333
## 16 6.608333
## 17 9.000000
## 18 9.000000
## 19 9.000000
## 20 9.000000
## 21 6.608333
## 22 6.608333
## 23 6.608333
## 24 6.608333
## 25 9.291667
## 26 9.291667
## 27 9.291667
## 28 9.291667
## 29 6.783333
## 30 6.783333
## 31 6.783333
## 32 6.783333
## 33 9.291667
## 34 9.291667
## 35 9.291667
## 36 9.291667
## 37 6.783333
## 38 6.783333
## 39 6.783333
## 40 6.783333
## 41 9.291667
## 42 9.291667
## 43 9.291667
## 44 9.291667
## 45 6.783333
## 46 6.783333
## 47 6.783333
## 48 6.783333
```

```
# Parte aleatoria
```

```
alfai <- ranef(modelo)$Individuo$(Intercept)" #alfa i de cada individuo
alfai
```

```
## [1] -0.20362698 0.01786202 0.02143442 -0.08931008 0.18219256 0.07144806
## [7] -0.11431690 -0.06787566 0.05715845 0.10359969 0.05715845 -0.03572403
```

```
# significacion del modelo (comparacion con modelo nulo)
```

```
modelo0 <- lmer(horas.suenio ~ (1 | Individuo), Datos)
anova(modelo, modelo0)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: Datos
```

```
## Models:
```

```
## modelo0: horas.suenio ~ (1 | Individuo)
```

```
## modelo: horas.suenio ~ sexo * Fumador + (1 | Individuo)
```

```
##          npar      AIC      BIC logLik deviance Chisq Df Pr(>Chisq)
```

```
## modelo0      3 105.222 110.836 -49.611  99.222
```

```
## modelo       6  71.328  82.555 -29.664  59.328 39.894  3 1.122e-08 ***
```

```
## ---
```

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

```
AIC(modelo0, modelo)
```

```
##          df      AIC
```

```
## modelo0 3 105.3742
## modelo 6 79.1041

# otra posibilidad (incluye Test Wald para coeficientes)
library(nlme)
modelo_lme <- lme(horas.suenio ~ sexo * Fumador, random = ~1 | Individuo,
  data = Datos)
summary(modelo_lme)

## Linear mixed-effects model fit by REML
## Data: Datos
##      AIC      BIC    logLik
## 79.1041 89.80924 -33.55205
##
## Random effects:
## Formula: ~1 | Individuo
##      (Intercept) Residual
## StdDev: 0.1907132 0.4403282
##
## Fixed effects: horas.suenio ~ sexo * Fumador
##
##              Value Std.Error DF   t-value p-value
## (Intercept)  9.000000 0.1681703 36  53.51718  0.0000
## sexomujer    0.291667 0.2378287  8   1.22637  0.2549
## Fumadorsi   -2.391667 0.2378287  8 -10.05626  0.0000
## sexomujer:Fumadorsi -0.116667 0.3363406  8  -0.34687  0.7376
## Correlation:
##              (Intr) sexmjr Fumdrs
## sexomujer    -0.707
## Fumadorsi    -0.707  0.500
## sexomujer:Fumadorsi  0.500 -0.707 -0.707
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -2.31423594 -0.58672629 -0.07566236  0.75156630  1.85726805
##
## Number of Observations: 48
## Number of Groups: 12

intervals(modelo_lme)

## Approximate 95% confidence intervals
##
## Fixed effects:
##
##              lower      est.      upper
## (Intercept)  8.6589348  9.0000000  9.3410652
## sexomujer    -0.2567674  0.2916667  0.8401007
## Fumadorsi   -2.9401007 -2.3916667 -1.8432326
## sexomujer:Fumadorsi -0.8922695 -0.1166667  0.6589362
## attr(,"label")
## [1] "Fixed effects:"
##
## Random Effects:
## Level: Individuo
##
##              lower      est.      upper
## sd((Intercept)) 0.05837091 0.1907132 0.6231107
##
## Within-group standard error:
##      lower      est.      upper
## 0.3495075 0.4403282 0.5547489
```

```
# calculamos icc
(icc <- (0.03637/(0.03637 + 0.19389)))
```

```
## [1] 0.1579519
```

\*El ICC es muy bajo, indicando que la variabilidad entre individuos es muy baja comparada con la variabilidad dentro de cada individuo.\*\*

*Conclusiones: existe una relacion entre el habito de fumar y las disminucion en las horas de suenio. En particular la disminucion de horas es entre 1.84 y 2.94 horas. Esta disminucion en las horas de suenio no depende del sexo de la persona.*

## Problema 6. Hormona de crecimiento

Se efectua un experimento para estudiar el efecto de la hormona de crecimiento en ratas jovenes. Se prueba, en forma inyectable, una dosis baja, una dosis alta y un tercer tratamiento, que seria el testigo, consistente en la inyeccion de solucion fisiologica. Se toman 6 camadas de ratas al azar, y tambien al azar se seleccionan tres animales de cada una. Se asignan los tratamientos al azar dentro de cada camada y al cabo de 15 dias se mide el aumento de peso, en decigramos. Base de datos en *Hormona.txt*.

- Identifique la variable respuesta, los factores y sus niveles y su condicion de fijos o aleatorios, cruzados o anidados, justificando su respuesta.

VR: Ganancia en peso (dgr)

VE: Hormona, cualitativa fija (3 niveles), camada, cualitativa aleatoria (6 niveles)

Tipo de estudio: Experimental

UE: Cada camada condicionada por la camada.

Replicas: 6.

- Indique el modelo estadistico utilizado, en terminos teoricos y aplicado a este experimento.

$$Y_{i|k} = \beta_0 + \beta_1 * Hormona\ baja_i + \beta_2 * Hormona\ alta_i + B_k + \epsilon_{ik}$$

i=1:3, k=1:6

$$B_k \sim NID(0, \sigma_{camadas}^2)$$

$$\epsilon_{ik} \sim NID(0, \sigma^2)$$

- Describa grafica y estadisticamente los resultados. Realice un grafico de perfiles para estudiar el paralelismo entre bloques. ¿Que significa “paralelismo” en el contexto de este ensayo? ¿Que piensa que ocurriria si los bloques respondieran de forma “No-paralela”?.

*Paralelismo significa que el comportamiento de cada camada en los tratamientos tenga el mismo sentido biologico. Es decir que si algun tratamiento genera un aumento en la VR ese aumento se vea para la mayoria de las camadas aunque pueden tener diferente magnitud.*

*Si no se cumple el paralelismo entonces las estimaciones seran incorrectas y no se podra confiar en las predicciones*

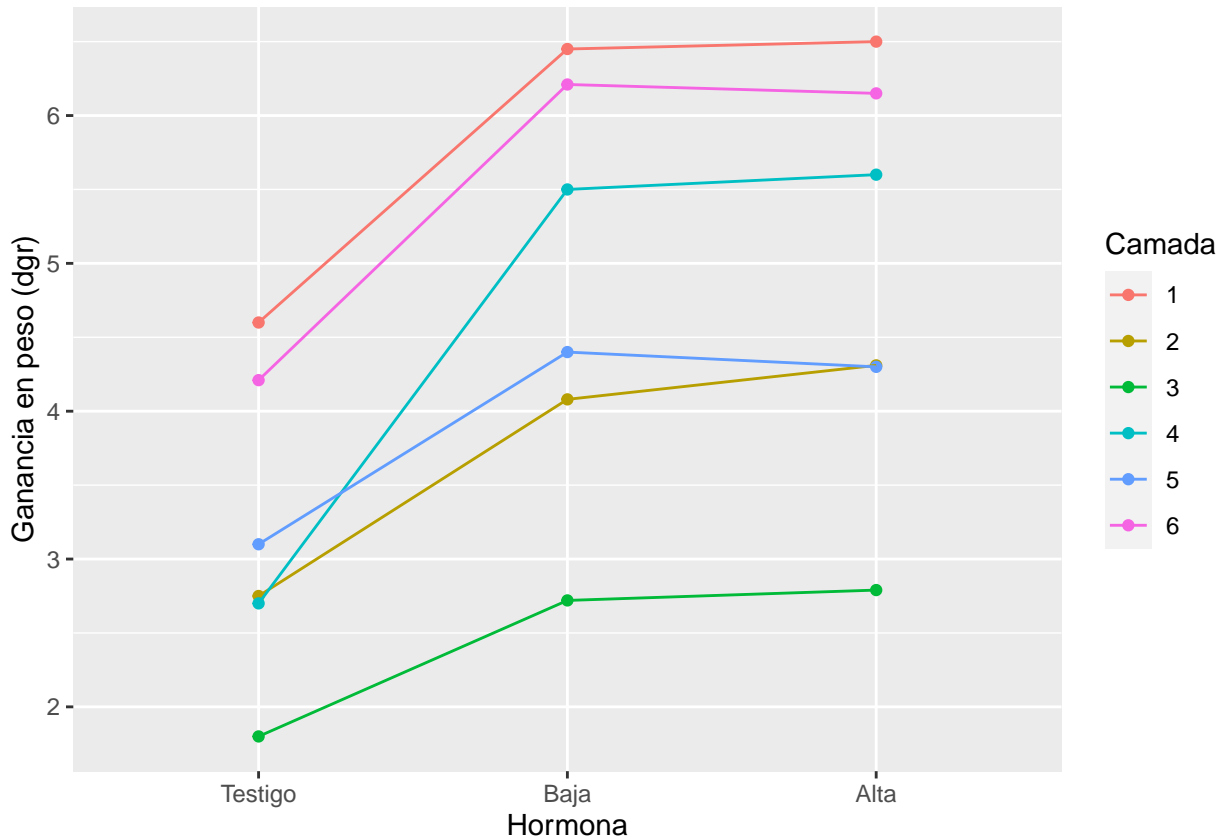
```
# Abramos el data.frame problema 'Hormonas'
Datos <- read.delim("/home/pangolin/Documents/materias/biome2/2018/tps/tp5/parteA/Hormona.txt")
# Miremos la tabla Pasamos el bloque ('camada') a variable
# cualitativa
Datos$Camada <- factor(Datos$Camada)
# ordeno los niveles (Testigo-Baja-Alta)
levels(Datos$Hormona)
```

```
## [1] "Alta" "Baja" "Testigo"
```

```
Datos$Hormona <- factor(Datos$Hormona, levels = c("Testigo", "Baja",
"Alta"))
levels(Datos$Hormona)
```

```
## [1] "Testigo" "Baja"      "Alta"
```

```
library(ggplot2)
p <- ggplot(Datos, aes(Hormona, GanPeso, color = Camada, group = Camada))
p + geom_point() + labs(x = "Hormona", y = "Ganancia en peso (dgr)") +
  geom_line()
```



- Verifique los supuestos del modelo.

*Supuestos: Homocedasticidad de la VR en cada nivel. VR normal. Independencia de las observaciones.*

*Y se suma, Normalidad e independencia de la variable de efecto aleatorio.*

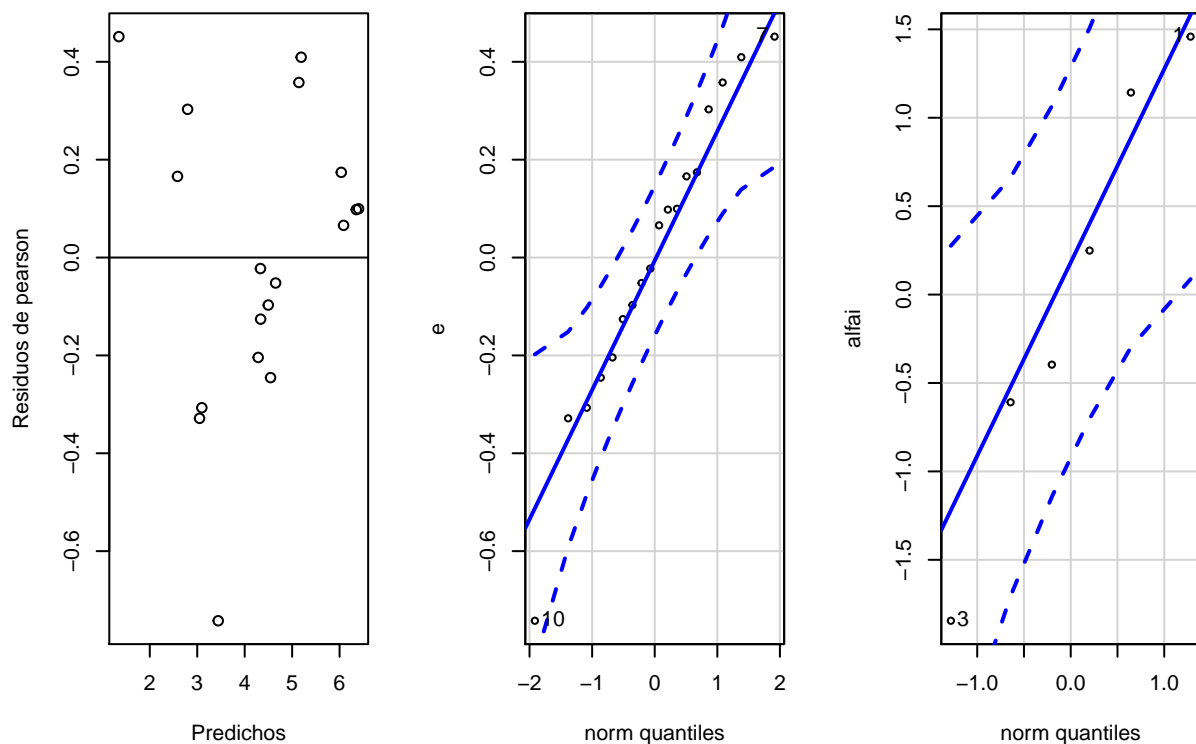
*Ademas es deseable que los “bloques” (Camadas) respondan al tratamiento de manera “paralela”.*

```
# Modelo considerando Camadas como de efectos aleatorios
library(lme4)
m2 <- lmer(GanPeso ~ Hormona + (1 | Camada), data = Datos)
# Supuestos
e <- resid(m2) # residuos de pearson
pre <- predict(m2) #predichos
alfai <- ranef(m2)$Camada$(Intercept)
par(mfrow = c(1, 3))
plot(pre, e, xlab = "Predichos", ylab = "Residuos de pearson", main = "Gráfico de dispersión de RE vs PRED",
      cex.main = 0.8)
abline(0, 0)
car::qqPlot(e)
```

```
## [1] 10 7
```

```
car::qqPlot(alfai)
```

Gráfico de dispersión de RE vs PRED



```
## [1] 3 1
```

```
par(mfrow = c(1, 1))
```

```
shapiro.test(e)
```

```
##
## Shapiro-Wilk normality test
##
## data:  e
## W = 0.9631, p-value = 0.6623
```

```
shapiro.test(alfai)
```

```
##
## Shapiro-Wilk normality test
##
## data:  alfai
## W = 0.96188, p-value = 0.8341
```

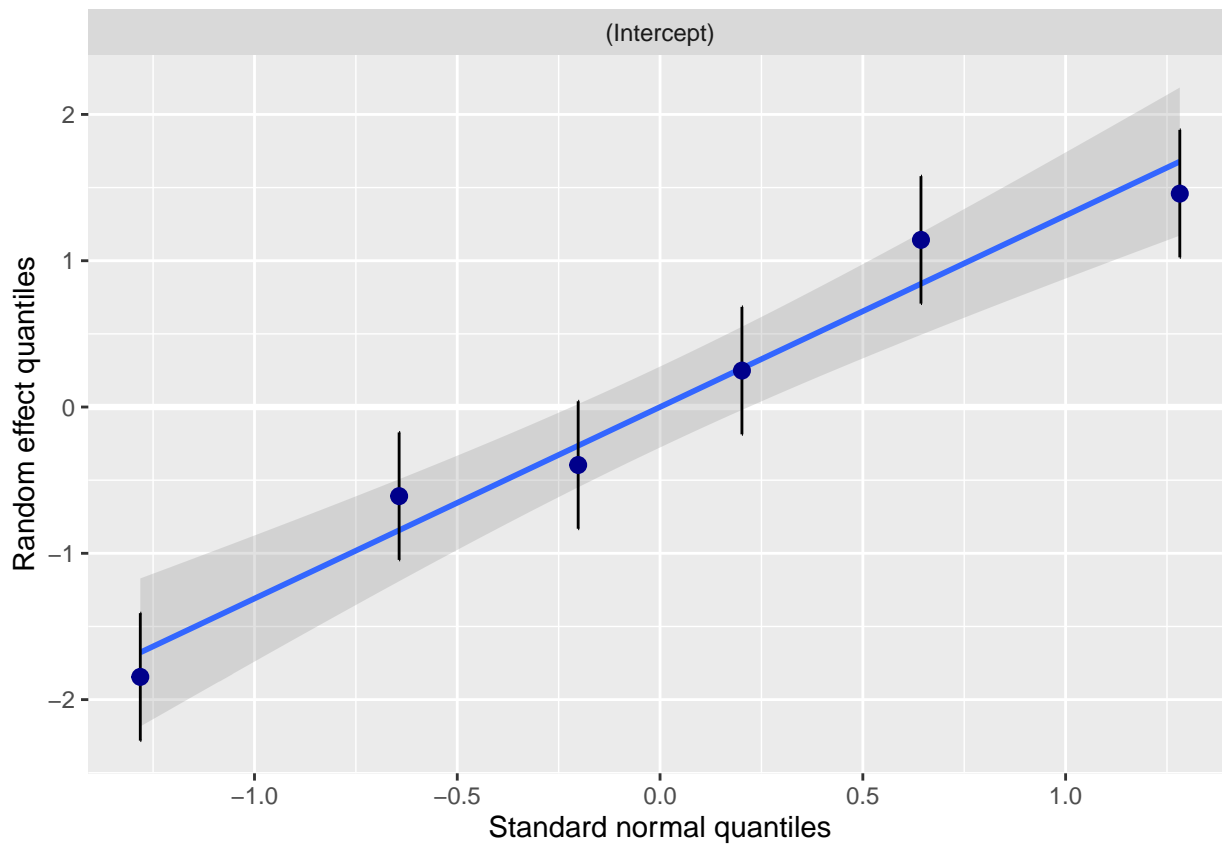
```
library(sjPlot)
```

```
## Install package "strengjacke" from GitHub (`devtools::install_github("strengjacke/strengjacke")`) to local
```

```
plot_model(m2, type = "diag")[[2]]
```

```
## $Canada
```

```
## `geom_smooth()` using formula 'y ~ x'
```



- Plantee las hipótesis. Resuelva y concluya, asumiendo un nivel de significación del 5%.

### Hipotesis

$$H_{01} - > \beta_1 = 0$$

*No hay influencia de la dosis baja de hormona sobre el aumento de peso.*

$$H_{02} - > \beta_2 = 0$$

*No hay influencia de la dosis alta de hormona sobre el aumento de peso.*

$$H_{03} - > \sigma^2[\alpha_k] = 0$$

*No hay variabilidad en la ganancia de peso entre las diferentes camadas.*

`summary(m2)`

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: GanPeso ~ Hormona + (1 | Camada)
##   Data: Datos
##
## REML criterion at convergence: 37
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.89663 -0.47165  0.05532  0.43962  1.15342
##
## Random effects:
##   Groups   Name                Variance Std.Dev.
```

```
## Canada (Intercept) 1.5347 1.2388
## Residual 0.1532 0.3914
## Number of obs: 18, groups: Canada, 6
##
## Fixed effects:
## Estimate Std. Error t value
## (Intercept) 3.1933 0.5304 6.021
## HormonaBaja 1.7000 0.2260 7.524
## HormonaAlta 1.7483 0.2260 7.738
##
## Correlation of Fixed Effects:
## (Intr) HrmnBj
## HormonaBaja -0.213
## HormonaAlta -0.213 0.500

# También se puede usar la library('nlme')
library("nlme")
m2b <- lme(GanPeso ~ Hormona, random = ~1 | Canada, data = Datos)
summary(m2b)

## Linear mixed-effects model fit by REML
## Data: Datos
## AIC BIC logLik
## 46.97939 50.51964 -18.4897
##
## Random effects:
## Formula: ~1 | Canada
## (Intercept) Residual
## StdDev: 1.238824 0.3913637
##
## Fixed effects: GanPeso ~ Hormona
## Value Std.Error DF t-value p-value
## (Intercept) 3.193333 0.5303851 10 6.020783 1e-04
## HormonaBaja 1.700000 0.2259539 10 7.523657 0e+00
## HormonaAlta 1.748333 0.2259539 10 7.737565 0e+00
## Correlation:
## (Intr) HrmnBj
## HormonaBaja -0.213
## HormonaAlta -0.213 0.500
##
## Standardized Within-Group Residuals:
## Min Q1 Med Q3 Max
## -1.89663448 -0.47164789 0.05531859 0.43962153 1.15342264
##
## Number of Observations: 18
## Number of Groups: 6

intervals(m2b) # para intervalos de confianza de los estimadores

## Approximate 95% confidence intervals
##
## Fixed effects:
## lower est. upper
## (Intercept) 2.011562 3.193333 4.375105
## HormonaBaja 1.196543 1.700000 2.203457
## HormonaAlta 1.244877 1.748333 2.251790
## attr(,"label")
## [1] "Fixed effects:"
##
## Random Effects:
```

```
## Level: Canada
##               lower      est.      upper
## sd((Intercept)) 0.6528444 1.238824 2.350766
##
## Within-group standard error:
##               lower      est.      upper
## 0.2524908 0.3913637 0.6066183
##### significacion parte fija #####
m0 <- lmer(GanPeso ~ (1 | Canada), data = Datos)
anova(m0, m2)

## refitting model(s) with ML (instead of REML)

## Data: Datos
## Models:
## m0: GanPeso ~ (1 | Canada)
## m2: GanPeso ~ Hormona + (1 | Canada)
##      npar      AIC      BIC logLik deviance Chisq Df Pr(>Chisq)
## m0      3 66.697 69.368 -30.348  60.697
## m2      5 44.643 49.095 -17.322  34.643 26.053  2 2.201e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
AIC(m0, m2)
```

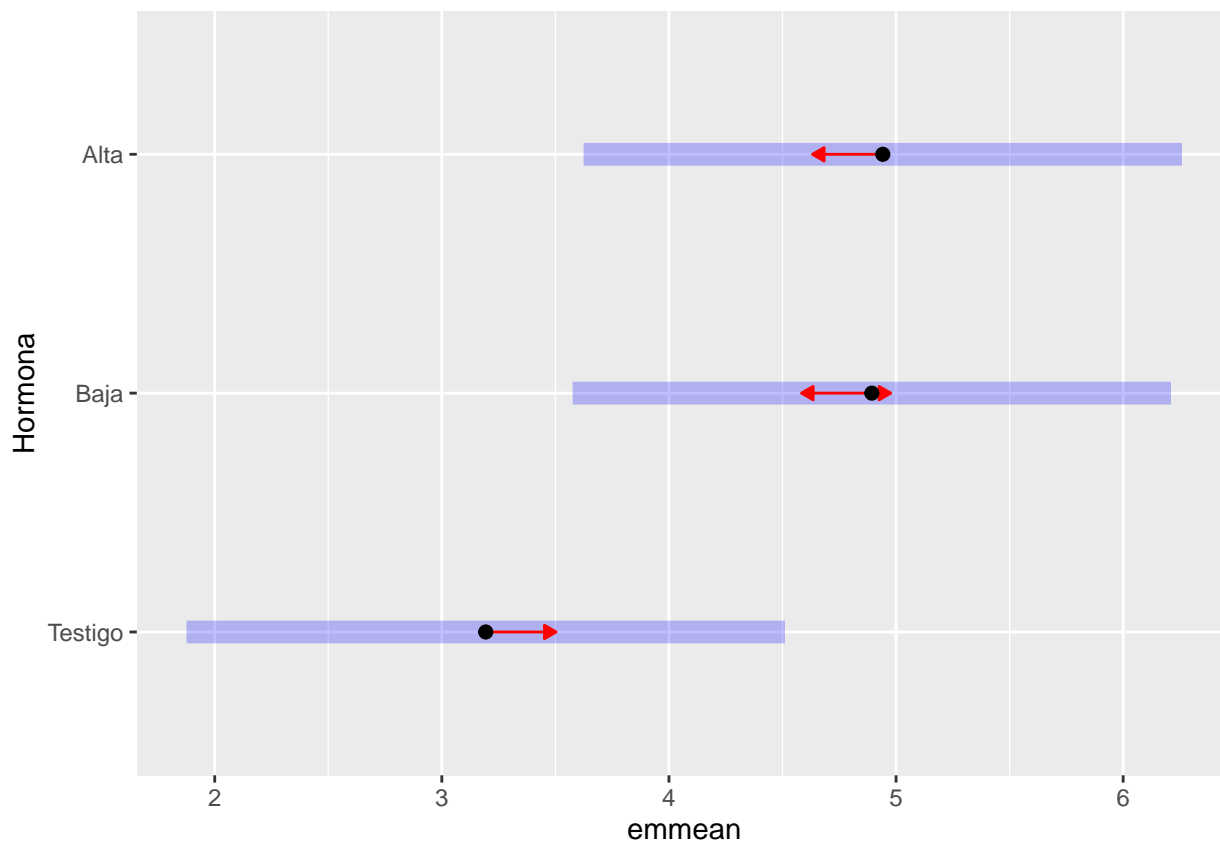
```
##      df      AIC
## m0    3 66.28331
## m2    5 46.97939
```

- Efectue las comparaciones pertinentes y concluya. Represente graficamente los resultados.

```
### comparaciones a posteriori #####
library(emmeans)
source("/home/pangolin/Documents/funciones2.R")
(comp <- emmeans(m2, pairwise ~ Hormona))

## $emmeans
## Hormona emmean SE df lower.CL upper.CL t.ratio p.value
## Testigo 3.19 0.53 5.65 1.88 4.51 6.021 0.0012
## Baja 4.89 0.53 5.65 3.58 6.21 9.226 0.0001
## Alta 4.94 0.53 5.65 3.62 6.26 9.317 0.0001
##
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95
##
## $contrasts
## contrast estimate SE df lower.CL upper.CL t.ratio p.value
## Testigo - Baja -1.7000 0.226 10 -2.319 -1.081 -7.524 0.0001
## Testigo - Alta -1.7483 0.226 10 -2.368 -1.129 -7.738 <.0001
## Baja - Alta -0.0483 0.226 10 -0.668 0.571 -0.214 0.9752
##
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 3 estimates
## P value adjustment: tukey method for comparing a family of 3 estimates
plot(comp$emmeans, comparisons = TRUE)
```

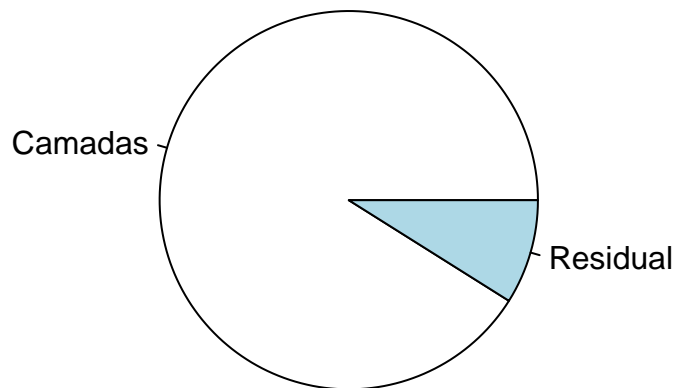




Se observa una ganancia de peso significativa en los grupos suministrados con hormona, ya sea en baja o altas dosis respecto del grupo testigo. No se encontraron diferencias entre los grupos de hormona alta y baja indicando que una mayor dosis no incrementaría la ganancia de peso.

```
### Parte Aleatoria ##### Grafico de componentes de varianza Simple
### Pie Chart
slices <- c(1.53, 0.15)
lbls <- c("Camadas", "Residual")
pie(slices, labels = lbls, main = "Variance components")
```

## Variance components



calculamos icc

```
(icc <- 1.238824^2/(1.238824^2 + 0.3913637^2))
```

```
## [1] 0.9092541
```

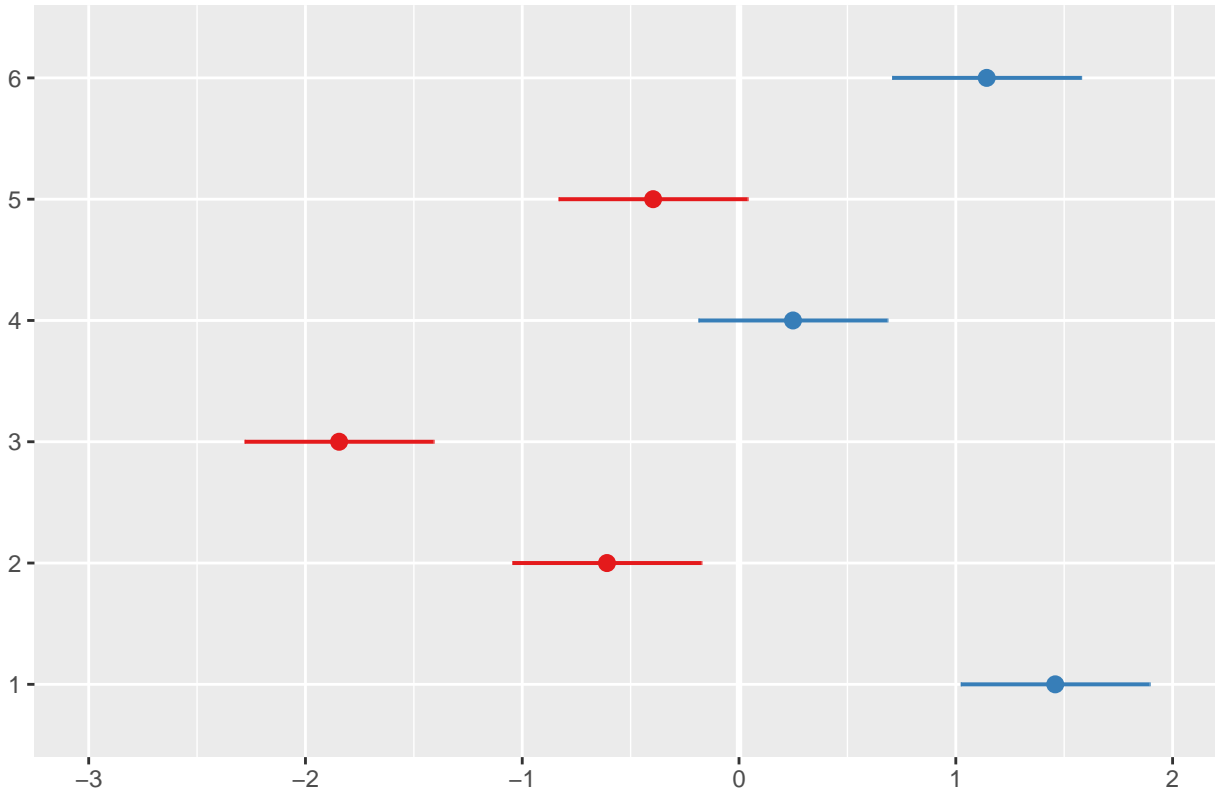
*El icc obtenido muestra que la variación entre camadas es muy grande por lo que explica la mayor parte de la varianza sin explicar. Esto significa que fue una buena decisión bloquear por camada y para hacer las predicciones es una buena idea un modelo condicional para hacer predicciones camada-específicas.*

```
ranef(m2) #alfa i (BLUP) de cada camada, graficado
```

```
## $Camada
## (Intercept)
## 1 1.4586950
## 2 -0.6091786
## 3 -1.8447411
## 4 0.2489406
## 5 -0.3962618
## 6 1.1425459
##
## with conditional variances for "Camada"
```

```
library(sjPlot)
# efectos aleatorios
plot_model(m2, type = "re")
```

## Random effects



```
#### Predicciones sujeto-especificas: parte fija + aleatoria #####
```

```
coef(m2) #muestra los coef para cada nivel aleatorio (camada)
```

```
## $Camada
## (Intercept) HormonaBaja HormonaAlta
## 1 4.652028 1.7 1.748333
## 2 2.584155 1.7 1.748333
## 3 1.348592 1.7 1.748333
## 4 3.442274 1.7 1.748333
## 5 2.797072 1.7 1.748333
## 6 4.335879 1.7 1.748333
##
## attr(,"class")
## [1] "coef.mer"
```

```
fitted(m2) #predicciones parte fija + aleatoria
```

```
## 1 2 3 4 5 6 7 8
## 4.652028 6.352028 6.400362 2.584155 4.284155 4.332488 1.348592 3.048592
## 9 10 11 12 13 14 15 16
## 3.096926 3.442274 5.142274 5.190607 2.797072 4.497072 4.545405 4.335879
## 17 18
## 6.035879 6.084213
```

```
fixef(m2) #estimacion parametros efectos fijos
```

```
## (Intercept) HormonaBaja HormonaAlta
## 3.193333 1.700000 1.748333
```

```
X <- model.matrix(m2)
pred_fija <- X %*% fixef(m2) #predicciones parte fija
alfai <- rep(round(ranef(m2)$Camada$(Intercept)", 4), each = 3)
pred <- data.frame(Datos$Hormona, Datos$Camada, Datos$GanPeso, pred_fija,
```

```
alfai, round(fitted(m2), 4))
colnames(pred) <- c("Hormona", "camada", "GanPes", "pred fija", "efecto aleat",
  "pred fija + aleat")
head(pred)
```

```
## Hormona camada GanPes pred fija efecto aleat pred fija + aleat
## 1 Testigo      1  4.60  3.193333      1.4587      4.6520
## 2 Baja        1  6.45  4.893333      1.4587      6.3520
## 3 Alta        1  6.50  4.941667      1.4587      6.4004
## 4 Testigo      2  2.75  3.193333     -0.6092      2.5842
## 5 Baja        2  4.08  4.893333     -0.6092      4.2842
## 6 Alta        2  4.31  4.941667     -0.6092      4.3325
```

- Resuelva el ejercicio incluyendo “Camadas” como factor fijo. Compare y concluya. ¿Considera que fue efectivo bloquear? Justifique.

```
# Porque no se considero a camada como de efectos fijos?
m1 <- lm(GanPeso ~ Hormona + Camada, data = Datos)
summary(m1)
```

```
##
## Call:
## lm(formula = GanPeso ~ Hormona + Camada, data = Datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.75056 -0.17889  0.01278  0.17361  0.51278
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.7006     0.2609  18.016 5.94e-09 ***
## HormonaBaja    1.7000     0.2260   7.524 2.01e-05 ***
## HormonaAlta    1.7483     0.2260   7.738 1.58e-05 ***
## Camada2       -2.1367     0.3195  -6.687 5.46e-05 ***
## Camada3       -3.4133     0.3195 -10.682 8.66e-07 ***
## Camada4       -1.2500     0.3195  -3.912 0.002905 **
## Camada5       -1.9167     0.3195  -5.998 0.000132 ***
## Camada6       -0.3267     0.3195  -1.022 0.330742
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3914 on 10 degrees of freedom
## Multiple R-squared:  0.9588, Adjusted R-squared:  0.93
## F-statistic: 33.28 on 7 and 10 DF, p-value: 3.808e-06
```

```
anova(m1)
```

```
## Analysis of Variance Table
##
## Response: GanPeso
##              Df Sum Sq Mean Sq F value    Pr(>F)
## Hormona       2 11.8980   5.9490  38.840 1.930e-05 ***
## Camada        5 23.7861   4.7572  31.059 8.796e-06 ***
## Residuals    10  1.5317   0.1532
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
AIC(m1)
```

```
## [1] 24.72938
```

La decision de poner a camadas como factor fijo o no tiene que ver con una decision metodologica y de la pregunta biologica

detras. El costo que tiene ademas es el tener que calcular una mayor cantidad de parametros. Ademas, el encontrar diferencias entre las camadas para con la respuesta a la hormona daria idea que la conclusion depende de la camada.

## Problema 7. Albumina

Se lleva a cabo un estudio a fin de determinar si existen diferencias en la albuminuria media de pacientes diabeticos con mas de cinco anios de evolucion de la enfermedad y diabeticos recientes. Para ello se seleccionan al azar siete individuos de cada grupo y se le efectuan cuatro determinaciones de albumina en orina. Base de datos en *Albumina.txt*.

- Escribir el modelo y plantear las hipótesis correspondientes en terminos del problema.

$$Y_{i|k} = \beta_0 + \beta_1 * Diabeticos_i + \beta_2 * Diabeticos\_recientes_i + B_k + \epsilon_{ik}$$

i=1:2, k=1:7

$$B_k \sim NID(0, \sigma_{individuo}^2)$$

$$\epsilon_{ik} \sim NID(0, \sigma^2)$$

### Hipotesis

$$H_{01} - > \beta_1 = 0$$

No hay influencia de ser diabetico hace mas de 5 anios sobre la albumina en orina.

$$H_{02} - > \beta_2 = 0$$

No hay influencia de ser diabetico reciente sobre la albumina en orina.

$$H_{03} - > \sigma^2[\alpha_k] = 0$$

No hay variabilidad en la albumina en orina entre individuos.

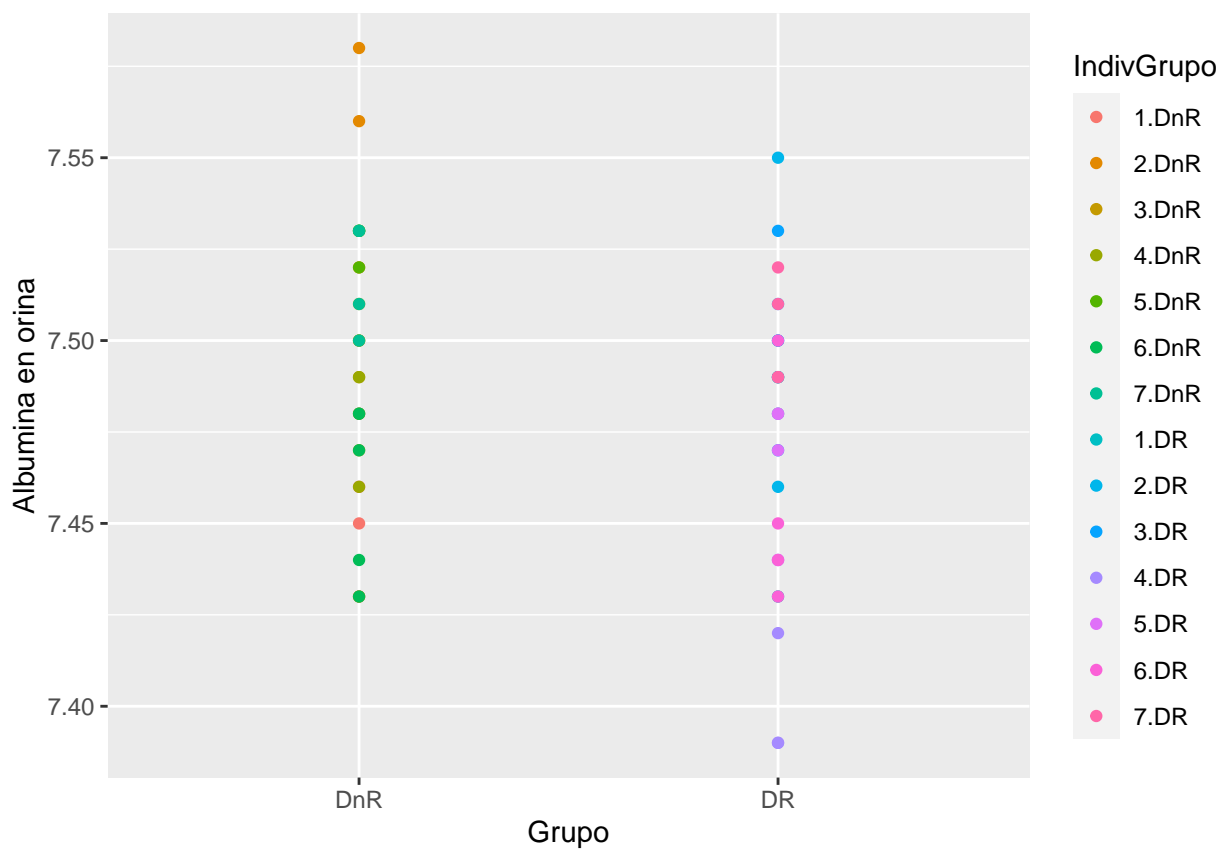
- Resolver verificando los supuestos y concluir. Realice comparaciones de ser necesario y cuando corresponda. Estime los componentes de varianza y concluya.

```
Datos <- read.delim("/home/pangolin/Documents/materias/biome2/2020/tps/tp5/nuevo/TP_5_Modelos Mixtos/Anidados
```

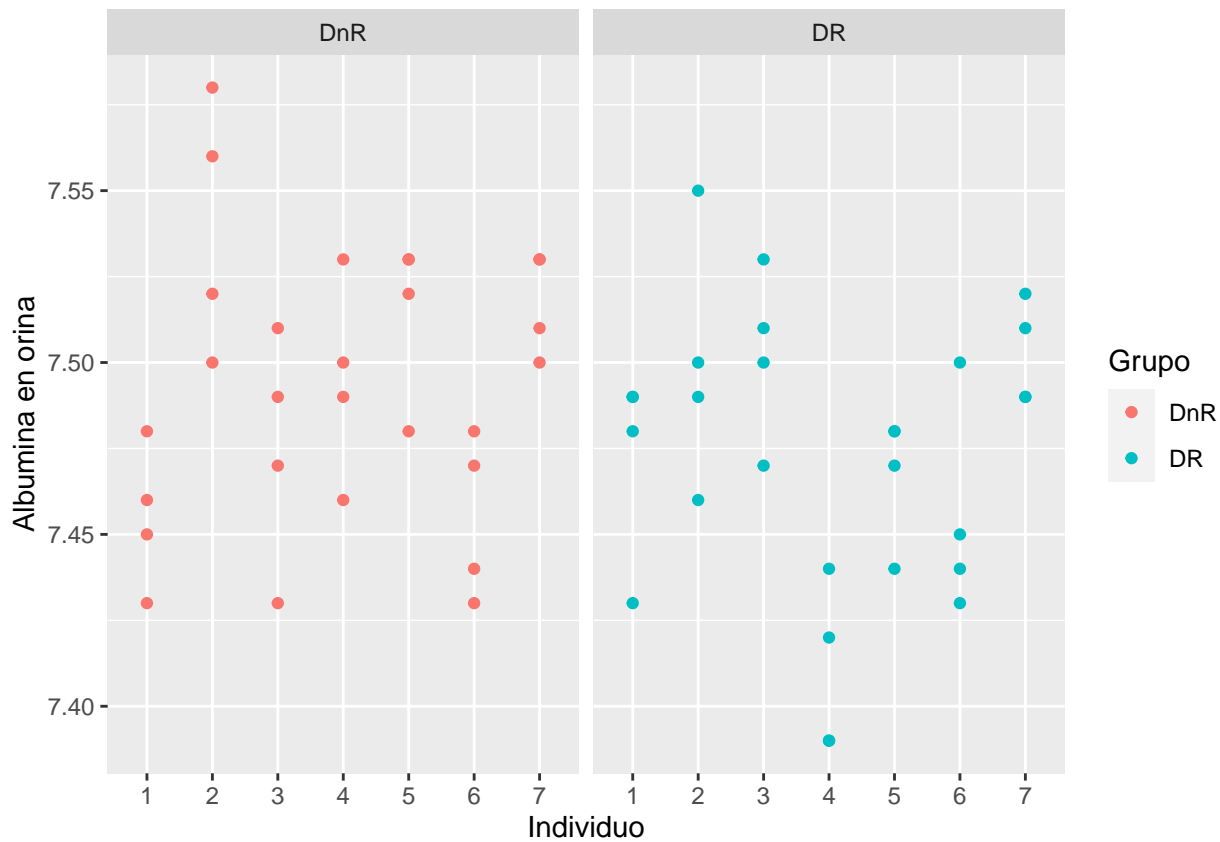
```
Datos$individuo <- factor(Datos$individuo) # el número de individuo como factor
# Hay dos individuos identificados con el mismo número(uno de cada
# grupo) Si no cambiamos esto, el programa entiende que es un
# mismo individuo que está en ambos grupos!
Datos$IndivGrupo <- with(Datos, interaction(individuo, Grupo))
# Ahora tenemos esta nueva variable que asigna un nombre distinto
# a cada individuo.
head(Datos)
```

```
## Grupo individuo albumina IndivGrupo
## 1 DR 1 7.43 1.DR
## 2 DR 1 7.48 1.DR
## 3 DR 1 7.49 1.DR
## 4 DR 1 7.49 1.DR
## 5 DR 2 7.49 2.DR
## 6 DR 2 7.46 2.DR
```

```
library(ggplot2)
p <- ggplot(Datos, aes(Grupo, albumina, color = IndivGrupo, group = IndivGrupo))
p + geom_point() + labs(x = "Grupo", y = "Albumina en orina")
```



```
p1 <- ggplot(Datos, aes(individuo, albumina, color = Grupo, group = Grupo))
p1 + geom_point() + facet_grid(. ~ Grupo) + labs(x = "Individuo",
y = "Albumina en orina")
```



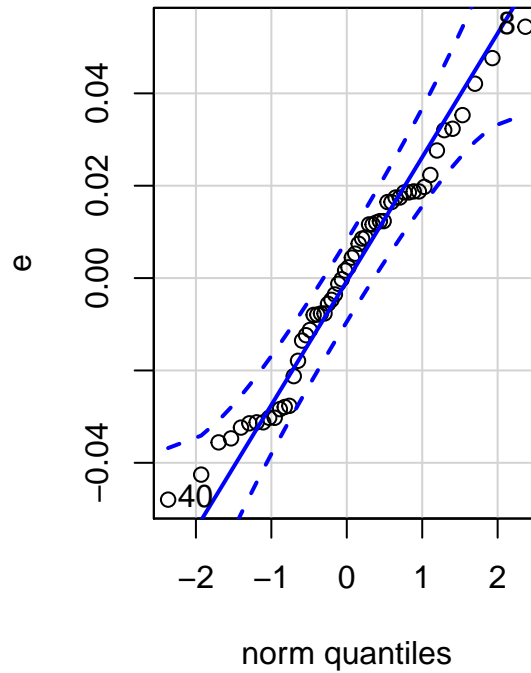
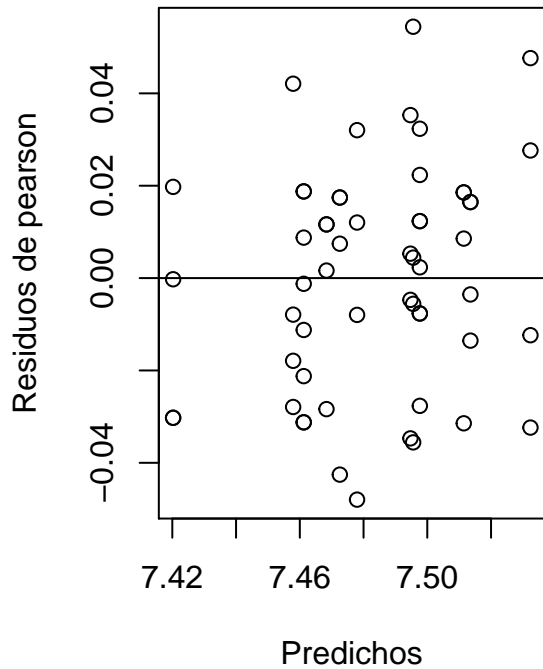
hacemos el modelo

```
library(lme4)
m1 <- lmer(albumina ~ Grupo + (1 | IndivGrupo), data = Datos)
```

supuestos e

```
e <- resid(m1) # residuos de pearson
pre <- predict(m1) #predichos
par(mfrow = c(1, 2))
plot(pre, e, xlab = "Predichos", ylab = "Residuos de pearson", main = "Gráfico de dispersión de RE vs PRED",
      cex.main = 0.8)
abline(0, 0)
car::qqPlot(e)
```

Gráfico de dispersión de RE vs PRED



```
## [1] 8 40
```

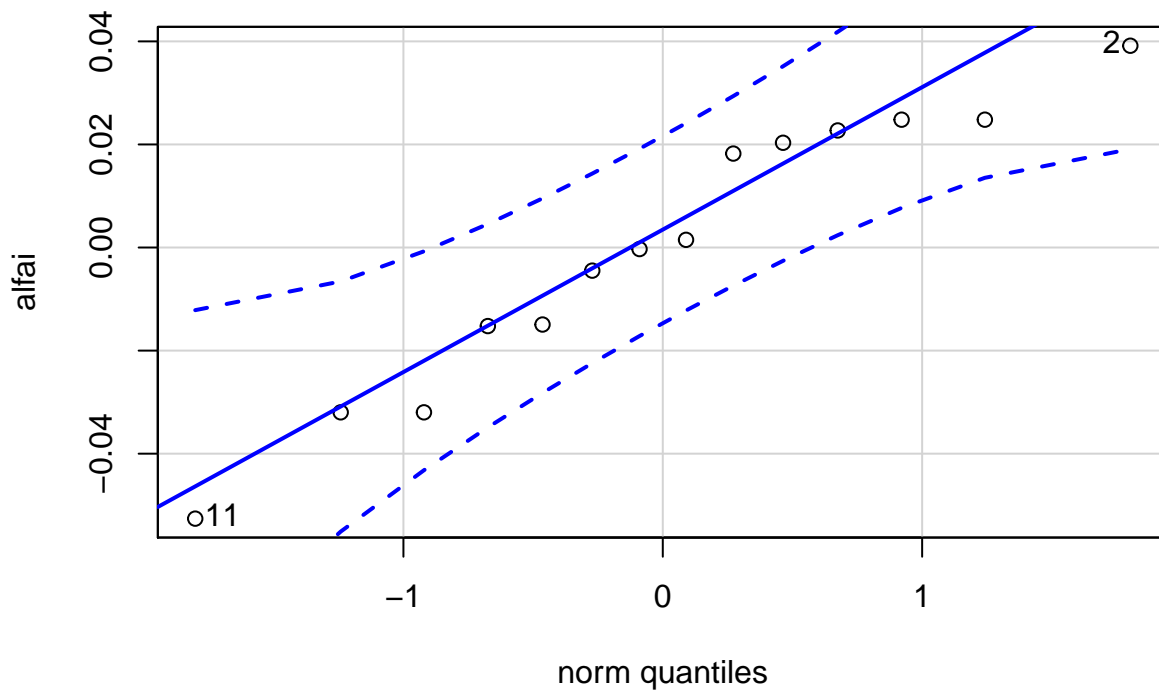
```
par(mfrow = c(1, 1))
shapiro.test(e)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  e
## W = 0.97535, p-value = 0.3052
```

### supuestos aleatorio

```
# Probamos normalidad del factor aleatorio
alfai <- ranef(m1)$IndivGrupo$(Intercept)"
car::qqPlot(alfai)
```





```
## [1] 11 2
# significacion parte fija
m0 <- lmer(albumina ~ (1 | IndivGrupo), data = Datos)
anova(m0, m1)

## refitting model(s) with ML (instead of REML)
## Data: Datos
## Models:
## m0: albumina ~ (1 | IndivGrupo)
## m1: albumina ~ Grupo + (1 | IndivGrupo)
##      npar      AIC      BIC logLik deviance Chisq Df Pr(>Chisq)
## m0      3 -215.38 -209.31 110.69 -221.38
## m1      4 -214.83 -206.73 111.42 -222.83 1.4489 1 0.2287
AIC(m0, m1)

##      df      AIC
## m0  3 -207.7624
## m1  4 -200.8477
# en este caso, la variable fija 'Grupo' no es significativa.
```

## predicciones

```
# Parte fija
fitted(m1) #predicciones parte fija + aleatoria

##      1      2      3      4      5      6      7      8
## 7.472558 7.472558 7.472558 7.472558 7.495573 7.495573 7.495573 7.495573
##      9     10     11     12     13     14     15     16
## 7.497665 7.497665 7.497665 7.497665 7.420253 7.420253 7.420253 7.420253
##     17     18     19     20     21     22     23     24
```

```
## 7.468374 7.468374 7.468374 7.468374 7.457913 7.457913 7.457913 7.457913
##      25      26      27      28      29      30      31      32
## 7.497665 7.497665 7.497665 7.497665 7.461233 7.461233 7.461233 7.461233
##      33      34      35      36      37      38      39      40
## 7.532369 7.532369 7.532369 7.532369 7.477971 7.477971 7.477971 7.477971
##      41      42      43      44      45      46      47      48
## 7.494709 7.494709 7.494709 7.494709 7.511447 7.511447 7.511447 7.511447
##      49      50      51      52      53      54      55      56
## 7.461233 7.461233 7.461233 7.461233 7.513539 7.513539 7.513539 7.513539
```

```
fixef(m1) #estimacion parametros efectos fijos
```

```
## (Intercept)      GrupoDR
## 7.49321429 -0.02035714
```

Al no encontrar una efecto de grupo, podemos concluir que no hay evidencia para decir que los niveles de albumina en sangre cambian con el tiempo de evolucion de la diabetes.

**vemos icc**

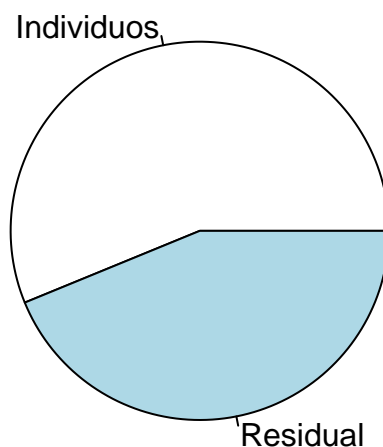
```
(icc <- 0.0009277/(0.0009277 + 0.0007232))
```

```
## [1] 0.5619359
```

Vemos que tenemos un icc de 0.56 lo cual es indicador que hay variabilidad no explicada correspondiente a los individuos. Esto nos puede indicar que hay factores no incluidos en este estudio que podrian estar relacionados con no haber encontrado un efecto del grupo.

```
# Simple Pie Chart
slices <- c(0.0009277, 0.0007232)
lbls <- c("Individuos", "Residual")
pie(slices, labels = lbls, main = "Variance components")
```

## Variance components



## Problema 8. Con lápiz y papel

Se desea estudiar la efectividad de un nuevo insecticida frente a la Deltametrina 10%, para el control de un insecto urbano. Se decide aplicar el tratamiento por manzana, para evitar la posible interferencia entre tratamientos. Para ello se seleccionan al azar 15 manzanas con niveles similares de infestación inicial y se las divide, también al azar, en 3 grupos iguales, a uno se le aplica el nuevo insecticida, a otro la deltametrina y el tercero es dejado como control (se pulveriza con líquido inerte). Se registra la abundancia del insecto a los 3 días luego de la aplicación del insecticida, en 5 viviendas de cada manzana, colocando 2 trampas para insectos en cada una de ellas.

- Plantee el modelo completo e interprete cada término en relación al ensayo.
- Indique el rango de valores de cada subíndice.

## Parte C: Diseños de Medidas Repetidas en el tiempo o en el espacio

### Problema 8. Sueño

Se ha demostrado que la privación del sueño puede afectar negativamente una amplia gama de las funciones cognitivas. Uno de los efectos más prominentes de la pérdida del sueño es el deterioro en la vigilancia y la activación. Se llevó a cabo un experimento a fin de determinar el efecto de la cafeína sobre la pérdida de atención producida por privación del sueño. Participaron 26 individuos que fueron privados totalmente de sueño durante tres días. Los individuos fueron divididos al azar en dos grupos. Uno recibió el tratamiento placebo y el otro, cafeína. Al inicio, primero, segundo y tercer día se les aplicó la prueba de de vigilancia psicomotora 10 min (PVT). Esta mide la atención sostenida o alerta al registrar los tiempos de respuesta promedio a estímulos visuales que se producen en intervalos aleatorios durante 10 min (en seg). Los datos en el archivo *Suenio.txt*.

```
Suenio <- read.delim("/home/pangolin/Documents/materias/biome2/2020/tps/tp5/nuevo/Suenio.txt")
```

- Identifique el tipo de estudio, la unidad experimental, la variable dependiente, tipo y potencial distribución de probabilidades, las variables explicativas, tipo y si son de efectos fijos o aleatorios. ¿De qué diseño se trata? ¿Con cuántas réplicas cuenta para evaluar el efecto del tratamiento con cafeína?.

Variable dependiente: tiempo de respuesta (segundos). Continua, distribución potencial normal.

Variables explicativas:

Tratamiento, 2 niveles (efectos fijos)

Tiempo, 4 niveles (efectos fijo)

Individuos, aleatorios.

Tipo de estudio: Observacional

UE: Individuo

**Modelo marginal (como comparación de medias)**

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$$

i=1:2, j=1:4, k=1:13 (con k anidado en i)

$$\epsilon_{ijk} \sim NID(0, \sum k)$$

```
summary(Suenio)
```

```
##      individuo      tratamiento      tiempo      t_reac
## Min.       : 1.0  cafeina:52  dia 1 :26  Min.       : 0.00
## 1st Qu.: 7.0  placebo:52  dia 2 :26  1st Qu.: 8.25
## Median :13.5                dia 3 :26  Median :15.65
## Mean   :13.5                inicial:26  Mean   :15.38
## 3rd Qu.:20.0                                3rd Qu.:21.93
## Max.   :26.0                                Max.   :32.30
```

```
Suenio$individuo <- factor(Suenio$individuo)
```

```
summary(Suenio)
```

```
##      individuo      tratamiento      tiempo      t_reac
## 1      : 4      cafeina:52      dia 1 :26      Min.   : 0.00
## 2      : 4      placebo:52      dia 2 :26      1st Qu.: 8.25
## 3      : 4                      dia 3 :26      Median :15.65
## 4      : 4                      inicial:26      Mean   :15.38
## 5      : 4                      3rd Qu.:21.93
## 6      : 4                      Max.    :32.30
## (Other):80
```

```
str(Suenio)
```

```
## 'data.frame': 104 obs. of 4 variables:
## $ individuo : Factor w/ 26 levels "1","2","3","4",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ tratamiento: Factor w/ 2 levels "cafeina","placebo": 1 1 1 1 1 1 1 1 1 1 ...
## $ tiempo : Factor w/ 4 levels "dia 1","dia 2",...: 4 4 4 4 4 4 4 4 4 4 ...
## $ t_reac : num 1.2 3 1.8 6.1 7.8 4.8 9.9 9.2 0 7.8 ...
```

- Describa los datos mediante un gráfico de perfiles por individuo y promedio.

```
# ordeno los niveles (para que placebo vaya a la intercept)
levels(Suenio$tratamiento) ## ¿cómo los toma?
```

```
## [1] "cafeina" "placebo"
```

```
Suenio$tratamiento <- factor(Suenio$tratamiento, levels = c("placebo",
  "cafeina"))
levels(Suenio$tratamiento) ## ¿qué cambió?
```

```
## [1] "placebo" "cafeina"
```

```
# ordeno los niveles (para que inicial vaya a la intercept)
levels(Suenio$tiempo) ##
```

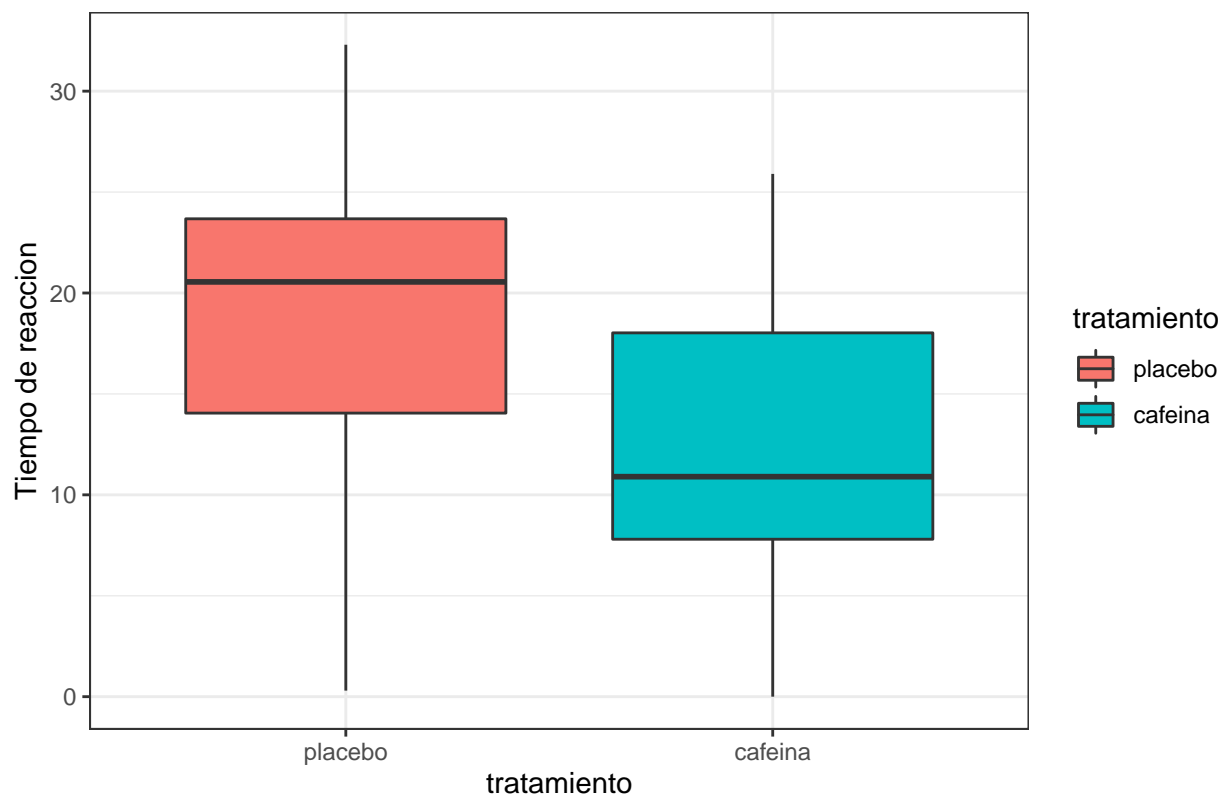
```
## [1] "dia 1" "dia 2" "dia 3" "inicial"
```

```
Suenio$tiempo <- factor(Suenio$tiempo, levels = c("inicial", "dia 1",
  "dia 2", "dia 3"))
levels(Suenio$tiempo) ## ¿qué cambió?
```

```
## [1] "inicial" "dia 1" "dia 2" "dia 3"
```

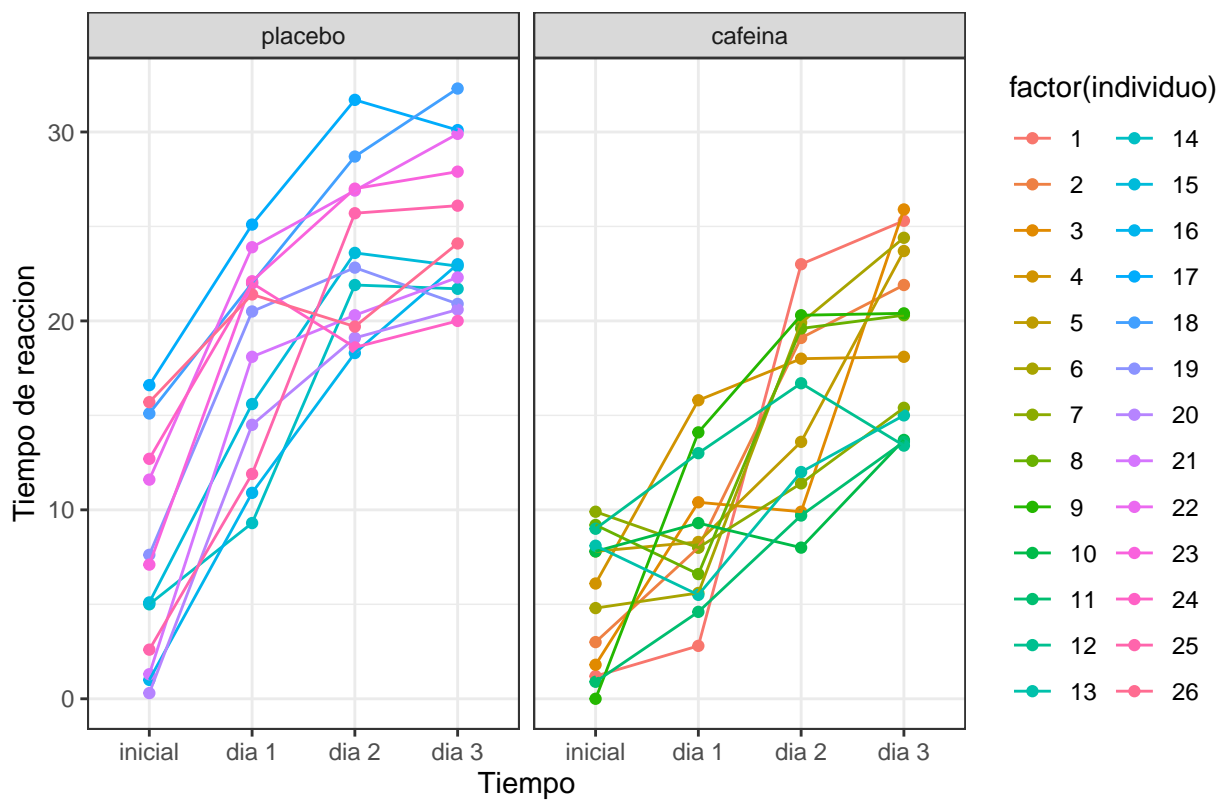
```
## Un poco de descriptiva
library(ggplot2)
ggplot(data = Suenio, aes(x = tratamiento, y = t_reac, fill = tratamiento)) +
  geom_boxplot() + labs(title = "Box-plot", x = "tratamiento", y = "Tiempo de reaccion") +
  theme_bw()
```

Box-plot



```
# spaghetti plots
ggplot(data = Suenio, aes(x = tiempo, y = t_reac, colour = factor(individuo),
  group = individuo)) + geom_point() + geom_line() + facet_grid(. ~
  tratamiento) + labs(title = "Spaghetti-plots, v. 1", x = "Tiempo",
  y = "Tiempo de reaccion") + theme_bw()
```

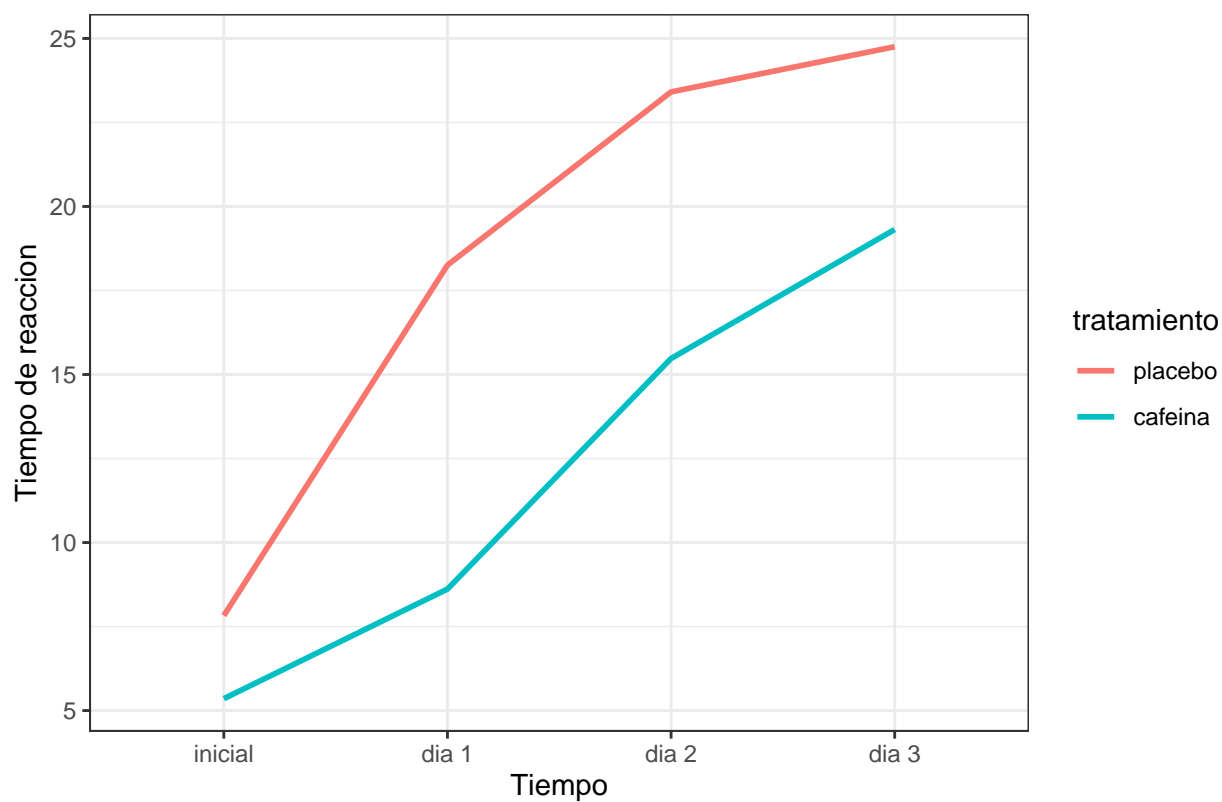
## Spaghetti-plots, v. 1



```
## tratamientos en el tiempo
ggplot(data = Suenio, aes(x = tiempo, y = t_reac, fill = tratamiento)) +
  geom_smooth(aes(group = tratamiento, colour = tratamiento), se = FALSE) +
  labs(title = "Box-plot considerando tratamiento y tiempo", x = "Tiempo",
        y = "Tiempo de reaccion") + theme_bw()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

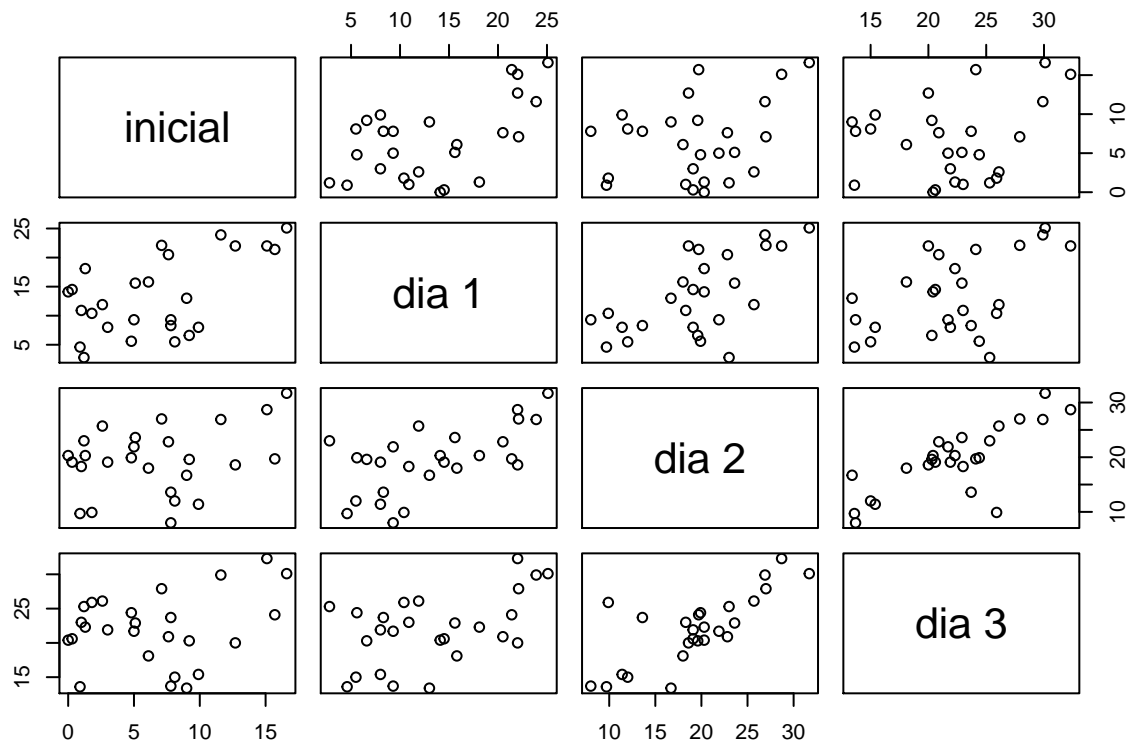
### Box-plot considerando tratamiento y tiempo



```
# estudiamos las correlaciones entre tiempos para convertir a
# formato wide (cada tiempo en una columna distinta)
library(reshape2)
Suenio_wide <- dcast(Suenio, individuo + tratamiento ~ tiempo, value.var = "t_reac")
Suenio2 <- Suenio_wide[, 3:6]
head(Suenio2)
```

```
##   inicial dia 1 dia 2 dia 3
## 1     1.2   2.8 23.0 25.3
## 2     3.0   8.0 19.1 21.9
## 3     1.8  10.4  9.9 25.9
## 4     6.1  15.8 18.0 18.1
## 5     7.8   8.3 13.6 23.7
## 6     4.8   5.6 19.9 24.4
```

```
plot(Suenio2)
```



```
round(cov(Suenio2[1:4]), 2) #matriz de covarianza
```

```
##          inicial dia 1 dia 2 dia 3
## inicial   24.47 18.44  8.99  6.23
## dia 1     18.44 44.58 25.32 17.35
## dia 2      8.99 25.32 36.80 24.01
## dia 3      6.23 17.35 24.01 26.43
```

```
round(cor(Suenio2[1:4]), 2) #matriz de correlacion
```

```
##          inicial dia 1 dia 2 dia 3
## inicial    1.00  0.56  0.30  0.25
## dia 1      0.56  1.00  0.63  0.51
## dia 2      0.30  0.63  1.00  0.77
## dia 3      0.25  0.51  0.77  1.00
```

```
## Cómo cae esa correlación entre tiempos?
```

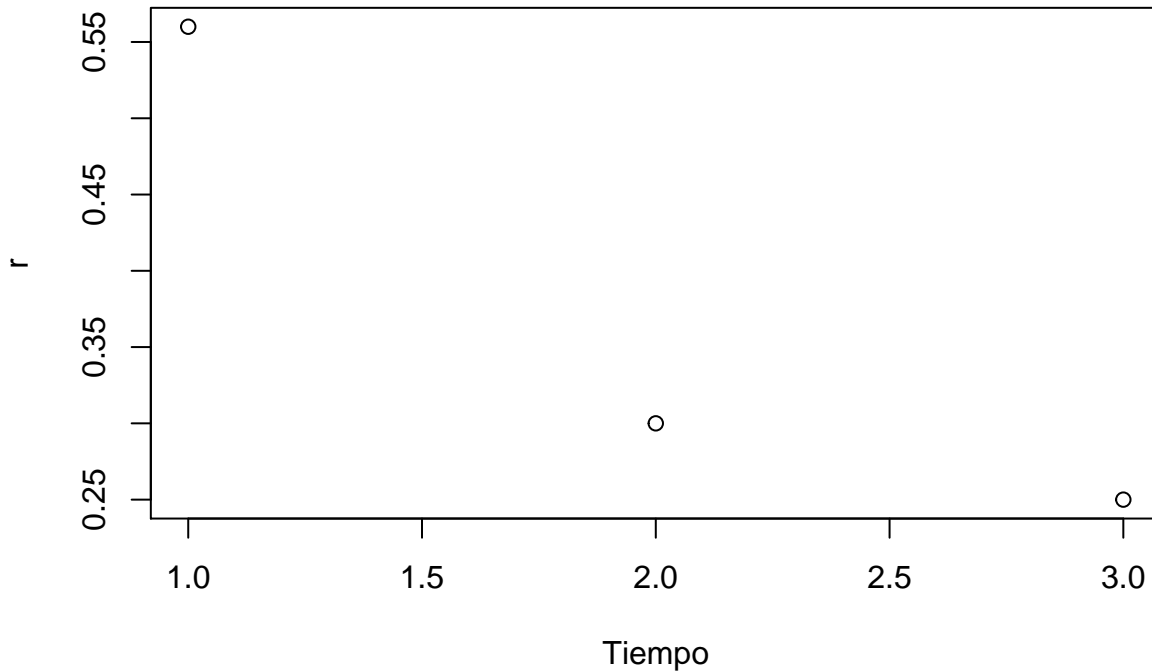
```
Tiempo <- c(1:3)
```

```
r <- c(0.56, 0.3, 0.25) ## ¿De dónde salen estos valores?
```

```
r_t <- as.data.frame(cbind(Tiempo, r))
```

```
plot(r_t)
```





- Ajuste distintas estructuras para la matriz de covarianza. Elija la que a su juicio considere la más apropiada.

## Modelado

En primer lugar, crearemos una nueva base de datos utilizando la función `melt()` del paquete `'reshape2'` en el dataframe en formato wide para tener al tiempo inicial como una covariable.

```
## Creacion de la base de datos. Argumentos de la funcion melt().
## 'id.vars' = las variables que quiero utilizar para definir cada
## fila de la matriz. 'variable.name' = el nombre del factor que
## se creara. Este contendra en sus niveles los nombres de las
## variables que se funden (= melt). En este caso nos devolvera el
## 'tiempo', porque estamos haciendo el camino inverso al dcast()
## que hicimos al comienzo.

# value.name = el nombre de la variable que almacena los valores.
# En este caso sera el tiempo de reaccion.

### Incluir tiempo inicial como co-variable.
datos_con_inicial <- melt(Suenio_wide, id.vars = c("individuo", "tratamiento",
  "inicial"), variable.name = "tiempo", value.name = "t_reac")

# exploremos las características del nuevo data frame
dim(datos_con_inicial)

## [1] 78 5
head(datos_con_inicial)

##   individuo tratamiento inicial tiempo t_reac
## 1         1    cafeina    1.2 dia 1    2.8
## 2         2    cafeina    3.0 dia 1    8.0
## 3         3    cafeina    1.8 dia 1   10.4
```

```
## 4      4      cafeina      6.1 dia 1      15.8
## 5      5      cafeina      7.8 dia 1      8.3
## 6      6      cafeina      4.8 dia 1      5.6
```

```
tail(datos_con_inicial)
```

```
##      individuo tratamiento inicial tiempo t_reac
## 73      21      placebo      1.3 dia 3      22.3
## 74      22      placebo     11.6 dia 3      29.9
## 75      23      placebo      7.1 dia 3      27.9
## 76      24      placebo     12.7 dia 3      20.0
## 77      25      placebo      2.6 dia 3      26.1
## 78      26      placebo     15.7 dia 3      24.1
```

```
str(datos_con_inicial)
```

```
## 'data.frame':    78 obs. of  5 variables:
## $ individuo  : Factor w/ 26 levels "1","2","3","4",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ tratamiento: Factor w/ 2 levels "placebo","cafeina": 2 2 2 2 2 2 2 2 2 2 ...
## $ inicial    : num  1.2 3 1.8 6.1 7.8 4.8 9.9 9.2 0 7.8 ...
## $ tiempo     : Factor w/ 3 levels "dia 1","dia 2",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ t_reac     : num  2.8 8 10.4 15.8 8.3 5.6 8 6.6 14.1 9.3 ...
```

```
# como sacamos el t_inicial del factor tiempo, revisemos como
```

```
# quedaron las correlaciones
```

```
datos_wide <- dcast(datos_con_inicial, formula = individuo + tratamiento ~
  tiempo, value.var = "t_reac")
```

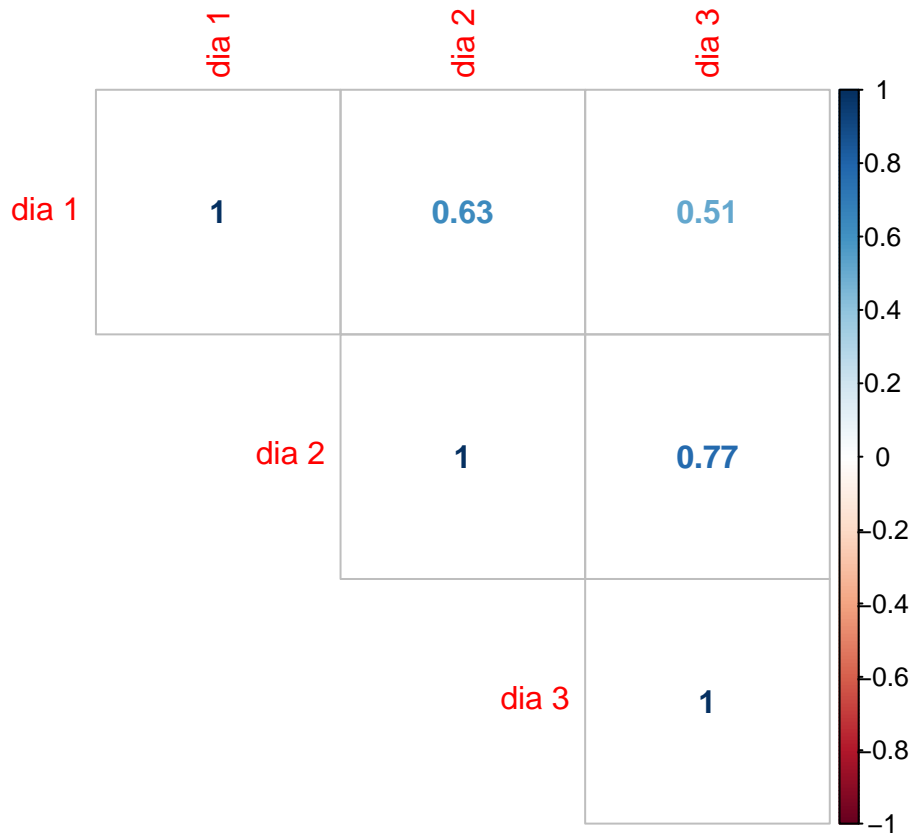
```
datos_correlaciones <- datos_wide[, 3:5]
```

```
coef_corr <- round(cor(datos_correlaciones), 2)
```

```
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```
corrplot(coef_corr, type = "upper", method = "number")
```



Generemos modelos con distintas estructuras de correlacion y chequeemos los supuestos.

```
library("nlme")
# Simetria compuesta. Varianzas iguales (recordar que la funcion
# lme() se puede utilizar aqui tambien).
modelo_corCompSymm <- gls(t_reac ~ tratamiento * tiempo + inicial,
  correlation = corCompSymm(form = ~1 | individuo), data = datos_con_inicial)
getVarCov(modelo_corCompSymm)

## Marginal variance covariance matrix
##      [,1]  [,2]  [,3]
## [1,] 19.3330  6.2631  6.2631
## [2,]  6.2631 19.3330  6.2631
## [3,]  6.2631  6.2631 19.3330
## Standard Deviations: 4.3969 4.3969 4.3969

# Autocorrelacion de orden 1. Varianzas iguales.
modelo_AR1 <- gls(t_reac ~ tratamiento * tiempo + inicial, correlation = corAR1(form = ~1 |
  individuo), data = datos_con_inicial)

getVarCov(modelo_AR1)

## Marginal variance covariance matrix
##      [,1]  [,2]  [,3]
## [1,] 19.2100  8.0496  3.3731
## [2,]  8.0496 19.2100  8.0496
## [3,]  3.3731  8.0496 19.2100
## Standard Deviations: 4.3829 4.3829 4.3829

# modelo_AR1 <- lme(t_reac ~ tratamiento * tiempo + inicial,
# (random = ~ 1 | individuo), correlation = corAR1(form = ~ 1 |
# individuo), data = datos_con_inicial) library(mgcv)
# extract.lme.cov2(modelo_AR1, datos_con_inicial, start.level=1)
```

```
# Modelo 3: Matriz de correlacion general.
```

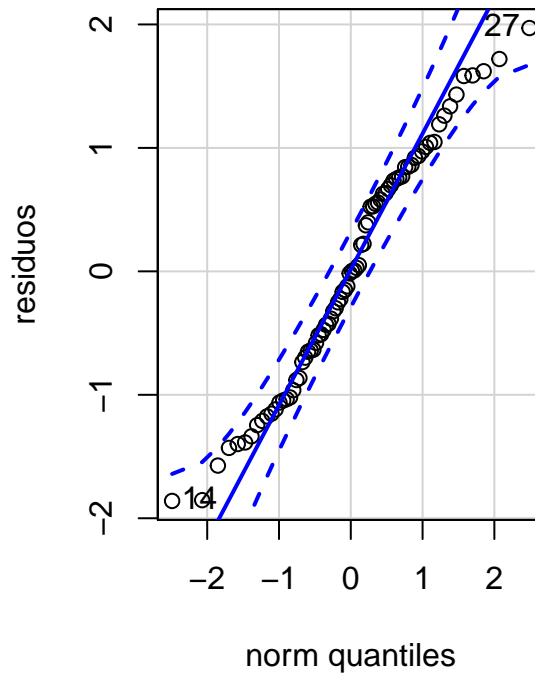
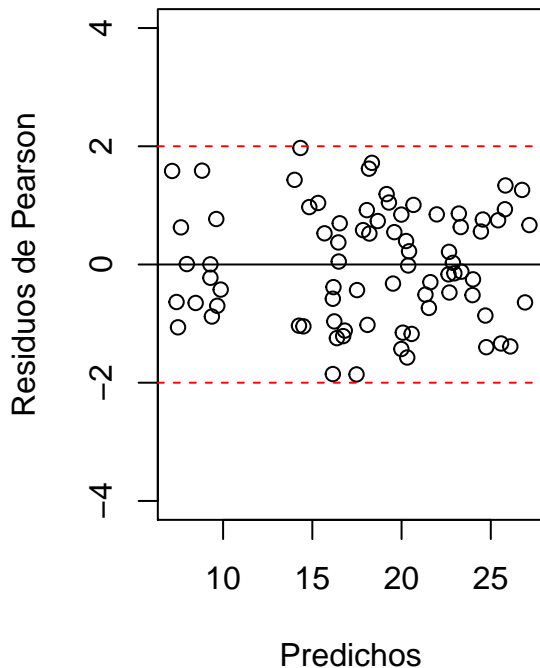
```
modelo_corSymm <- gls(t_reac ~ tratamiento * tiempo + inicial, correlation = corSymm(form = ~1 |
  individuo), data = datos_con_inicial)
getVarCov(modelo_corSymm)
```

```
## Marginal variance covariance matrix
##      [,1] [,2] [,3]
## [1,] 18.9390 3.212 1.6793
## [2,] 3.2120 18.939 12.0500
## [3,] 1.6793 12.050 18.9390
## Standard Deviations: 4.3519 4.3519 4.3519
```

```
supuestos <- function(modelo) {
  residuos <- resid(modelo, type = "pearson")
  predichos <- predict(modelo)
  par(mfrow = c(1, 2))
  plot(x = predichos, y = residuos, ylim = c(-4, 4), xlab = "Predichos",
    ylab = "Residuos de Pearson", main = "Gráfico de dispersión de residuos v. predichos",
    cex.main = 0.6)
  abline(h = c(-2, 2, 0), col = c("red", "red", "black"), lty = c(2,
    2, 1))
  car::qqPlot(residuos)
  shapiro.test(residuos)
}
```

```
supuestos(modelo_corCompSymm)
```

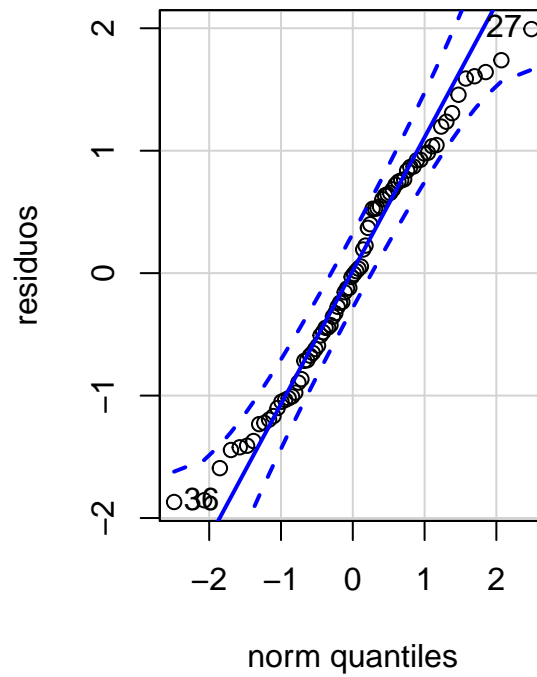
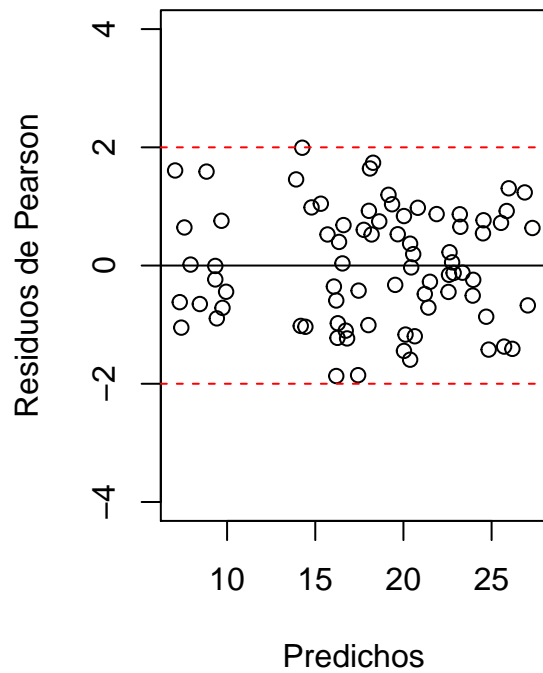
Gráfico de dispersión de residuos v. predichos



```
##
## Shapiro-Wilk normality test
##
## data:  residuos
## W = 0.97329, p-value = 0.1002
```

```
supuestos(modelo_AR1)
```

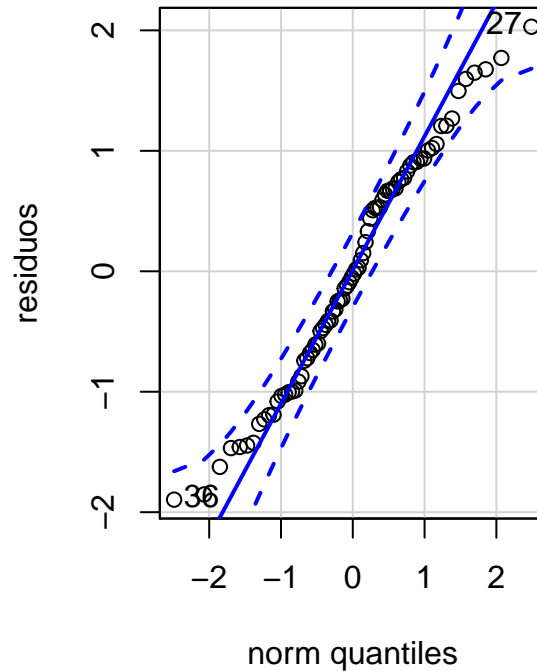
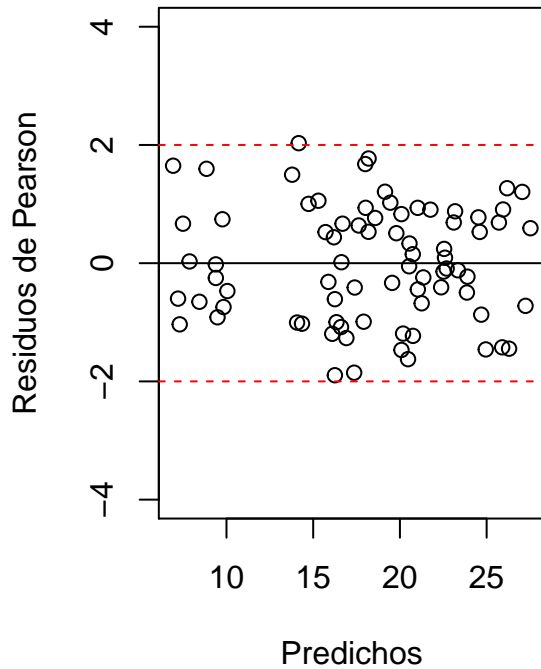
Gráfico de dispersión de residuos v. predichos



```
##  
## Shapiro-Wilk normality test  
##  
## data:  residuos  
## W = 0.97465, p-value = 0.1217
```

```
supuestos(modelo_corSymm)
```

Gráfico de dispersión de residuos v. predichos



```
##
## Shapiro-Wilk normality test
##
## data:  residuos
## W = 0.97643, p-value = 0.157
```

## Analisis de modelos

```
# solo para fines didacticos, veamos el summary de los 3 modelos
# para comparar las estructuras de correlacion obtenidas
summary(modelo_corCompSymm)
```

```
## Generalized least squares fit by REML
## Model: t_reac ~ tratamiento * tiempo + inicial
## Data: datos_con_inicial
##      AIC      BIC    logLik
## 445.3127 465.6769 -213.6564
##
## Correlation Structure: Compound symmetry
## Formula: ~1 | individuo
## Parameter estimate(s):
##      Rho
## 0.3239558
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept)  16.101774  1.619791   9.940652  0.0000
## tratamientocafeina -8.958904  1.757170  -5.098485  0.0000
## tiempodia 2      5.155385  1.418017   3.635630  0.0005
## tiempodia 3      6.500000  1.418017   4.583866  0.0000
## inicial        0.275039  0.136249   2.018647  0.0473
```

```
## tratamientocafeina:tiempodia 2 1.706154 2.005379 0.850789 0.3977
## tratamientocafeina:tiempodia 3 4.200000 2.005379 2.094367 0.0398
##
## Correlation:
## (Intr) trtmnt timpd2 timpd3 inicil trtm:2
## tratamientocafeina -0.649
## tiempodia 2 -0.438 0.403
## tiempodia 3 -0.438 0.403 0.500
## inicial -0.658 0.192 0.000 0.000
## tratamientocafeina:tiempodia 2 0.310 -0.571 -0.707 -0.354 0.000
## tratamientocafeina:tiempodia 3 0.310 -0.571 -0.354 -0.707 0.000 0.500
##
## Standardized residuals:
## Min Q1 Med Q3 Max
## -1.859695066 -0.726062769 -0.006981848 0.757692297 1.970812789
##
## Residual standard error: 4.39694
## Degrees of freedom: 78 total; 71 residual
```

```
summary(modelo_AR1)
```

```
## Generalized least squares fit by REML
## Model: t_reac ~ tratamiento * tiempo + inicial
## Data: datos_con_inicial
## AIC BIC logLik
## 442.3927 462.7568 -212.1964
##
## Correlation Structure: AR(1)
## Formula: ~1 | individuo
## Parameter estimate(s):
## Phi
## 0.419038
##
## Coefficients:
## Value Std.Error t-value p-value
## (Intercept) 15.968615 1.6149376 9.888069 0.0000
## tratamientocafeina -8.916857 1.7515857 -5.090734 0.0000
## tiempodia 2 5.155385 1.3103218 3.934442 0.0002
## tiempodia 3 6.500000 1.5608996 4.164265 0.0001
## inicial 0.292057 0.1358763 2.149431 0.0350
## tratamientocafeina:tiempodia 2 1.706154 1.8530748 0.920715 0.3603
## tratamientocafeina:tiempodia 3 4.200000 2.2074454 1.902652 0.0611
##
## Correlation:
## (Intr) trtmnt timpd2 timpd3 inicil trtm:2
## tratamientocafeina -0.649
## tiempodia 2 -0.406 0.374
## tiempodia 3 -0.483 0.446 0.596
## inicial -0.658 0.192 0.000 0.000
## tratamientocafeina:tiempodia 2 0.287 -0.529 -0.707 -0.421 0.000
## tratamientocafeina:tiempodia 3 0.342 -0.630 -0.421 -0.707 0.000 0.596
##
## Standardized residuals:
## Min Q1 Med Q3 Max
## -1.86893457 -0.71483021 -0.01922017 0.75471096 1.99325815
##
## Residual standard error: 4.382892
## Degrees of freedom: 78 total; 71 residual
```

```
summary(modelo_corSymm)
```

```
## Generalized least squares fit by REML
## Model: t_reac ~ tratamiento * tiempo + inicial
## Data: datos_con_inicial
##      AIC      BIC    logLik
## 441.5583 466.4478 -209.7791
##
## Correlation Structure: General
## Formula: ~1 | individuo
## Parameter estimate(s):
## Correlation:
## 1      2
## 2 0.170
## 3 0.089 0.636
##
## Coefficients:
##                               Value Std.Error   t-value p-value
## (Intercept)                15.782547 1.5794715   9.992296  0.0000
## tratamientocafeina          -8.858103 1.7370203  -5.099597  0.0000
## tiempodia 2                   5.155385 1.5555092   3.314275  0.0014
## tiempodia 3                   6.500000 1.6295423   3.988850  0.0002
## inicial                      0.315836 0.1301982   2.425814  0.0178
## tratamientocafeina:tiempodia 2 1.706154 2.1998222   0.775587  0.4406
## tratamientocafeina:tiempodia 3 4.200000 2.3045208   1.822505  0.0726
##
## Correlation:
##                               (Intr) trtmnt timpd2 timpd3 inicil trtm:2
## tratamientocafeina          -0.650
## tiempodia 2                  -0.492  0.448
## tiempodia 3                  -0.516  0.469  0.792
## inicial                     -0.645  0.185  0.000  0.000
## tratamientocafeina:tiempodia 2 0.348 -0.633 -0.707 -0.560  0.000
## tratamientocafeina:tiempodia 3 0.365 -0.663 -0.560 -0.707  0.000  0.792
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -1.89559179 -0.73711904 -0.03654798  0.75926800  2.03013062
##
## Residual standard error: 4.351943
## Degrees of freedom: 78 total; 71 residual
```

```
### En base al analisis de los supuestos decida los modelos
### candidatos
```

```
AIC(modelo_corCompSymm, modelo_AR1, modelo_corSymm)
```

```
##           df      AIC
## modelo_corCompSymm  9 445.3127
## modelo_AR1          9 442.3927
## modelo_corSymm     11 441.5583
```

```
anova(modelo_AR1)
```

```
## Denom. DF: 71
##           numDF  F-value p-value
## (Intercept)      1 808.1956 <.0001
## tratamiento      1  35.8014 <.0001
## tiempo           2  32.9909 <.0001
## inicial          1   4.6201  0.0350
## tratamiento:tiempo 2   1.8450  0.1655
```



```
summary(modelo_AR1)
```

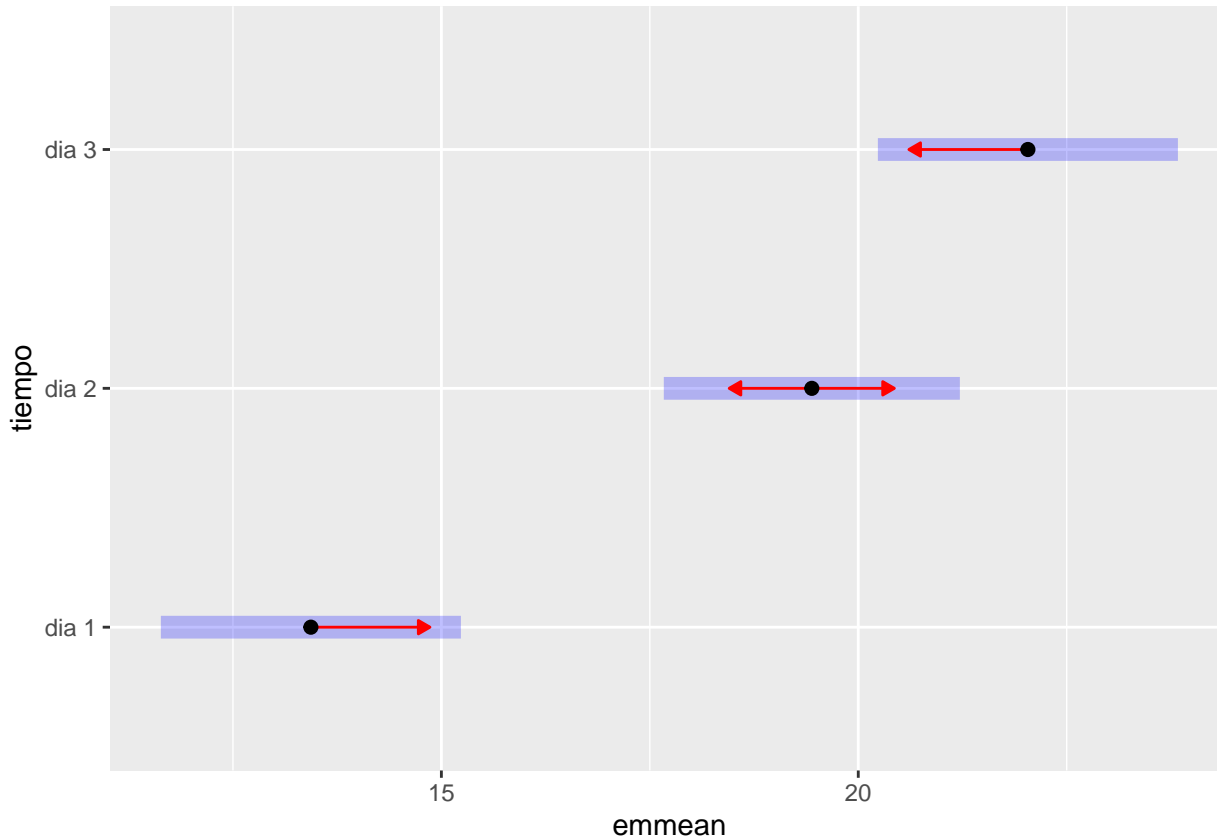
```
## Generalized least squares fit by REML
## Model: t_reac ~ tratamiento * tiempo + inicial
## Data: datos_con_inicial
##      AIC      BIC    logLik
## 442.3927 462.7568 -212.1964
##
## Correlation Structure: AR(1)
## Formula: ~1 | individuo
## Parameter estimate(s):
##      Phi
## 0.419038
##
## Coefficients:
##                                     Value Std.Error   t-value p-value
## (Intercept)                   15.968615  1.6149376   9.888069  0.0000
## tratamientocafeina             -8.916857  1.7515857  -5.090734  0.0000
## tiempodia 2                     5.155385  1.3103218   3.934442  0.0002
## tiempodia 3                     6.500000  1.5608996   4.164265  0.0001
## inicial                        0.292057  0.1358763   2.149431  0.0350
## tratamientocafeina:tiempodia 2  1.706154  1.8530748   0.920715  0.3603
## tratamientocafeina:tiempodia 3  4.200000  2.2074454   1.902652  0.0611
##
## Correlation:
##                                     (Intr) trtmnt timpd2 timpd3 inicil trtm:2
## tratamientocafeina                -0.649
## tiempodia 2                       -0.406  0.374
## tiempodia 3                       -0.483  0.446  0.596
## inicial                           -0.658  0.192  0.000  0.000
## tratamientocafeina:tiempodia 2    0.287 -0.529 -0.707 -0.421  0.000
## tratamientocafeina:tiempodia 3    0.342 -0.630 -0.421 -0.707  0.000  0.596
##
## Standardized residuals:
##      Min      Q1      Med      Q3      Max
## -1.86893457 -0.71483021 -0.01922017  0.75471096  1.99325815
##
## Residual standard error: 4.382892
## Degrees of freedom: 78 total; 71 residual
```

- Concluya acerca del efecto de la cafeína sobre la pérdida de atención producida por la privación del sueño. ¿Cambia con el tiempo?

```
library(emmeans)
(Tukey_tiempo <- emmeans(modelo_AR1, pairwise ~ tiempo))
```

```
## Warning in sweep(X, 1, sqrt(weights), "*"): STATS is longer than the extent of
## 'dim(x)[MARGIN]'
## NOTE: Results may be misleading due to involvement in interactions
## $emmeans
## tiempo emmean SE df lower.CL upper.CL t.ratio p.value
## dia 1 13.4 0.86 18.9 11.6 15.2 15.630 <.0001
## dia 2 19.4 0.86 23.6 17.7 21.2 22.620 <.0001
## dia 3 22.0 0.86 18.9 20.2 23.8 25.635 <.0001
##
## Results are averaged over the levels of: tratamiento
## Degrees-of-freedom method: satterthwaite
## Confidence level used: 0.95
##
```

```
## $contrasts
## contrast      estimate      SE    df lower.CL upper.CL t.ratio p.value
## dia 1 - dia 2    -6.01 0.927 14.5    -8.42   -3.593 -6.485 <.0001
## dia 1 - dia 3    -8.60 1.104 15.3   -11.46   -5.739 -7.792 <.0001
## dia 2 - dia 3    -2.59 0.927 14.4    -5.01   -0.175 -2.797 0.0350
##
## Results are averaged over the levels of: tratamiento
## Degrees-of-freedom method: satterthwaite
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 3 estimates
## P value adjustment: tukey method for comparing a family of 3 estimates
plot(Tukey_tiempo, comparisons = TRUE)
```

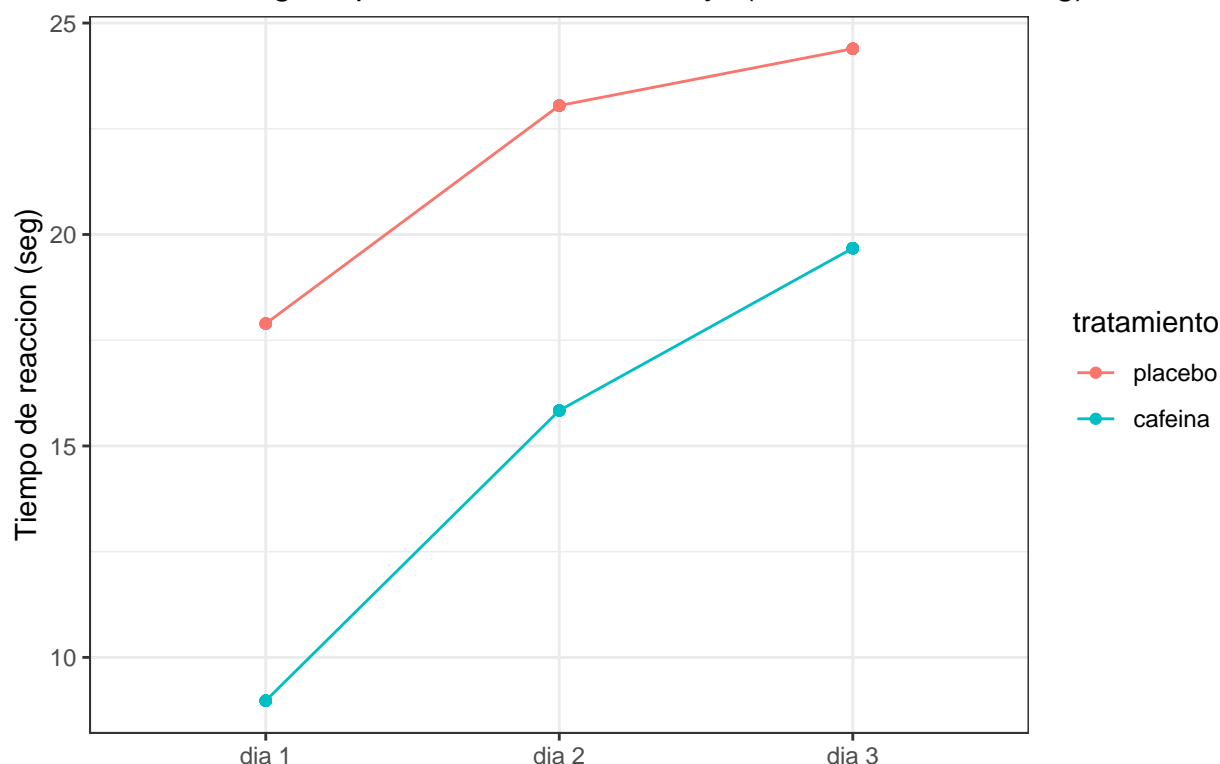


## Predicciones

```
library("margins")
model_p <- prediction(modelo_AR1, at = list(inicial = rep(mean(datos_con_inicial$inicial),
78)))

View(model_p)
ggplot(data = model_p, aes(x = tiempo, y = fitted, colour = tratamiento,
group = tratamiento)) + labs(title = "Modelo marginal predicho con t_inicial fijo (valor medio: 6.58seg)"
x = "", y = "Tiempo de reaccion (seg)") + geom_point() + geom_line() +
theme_bw()
```

### Modelo marginal predicho con $t_{\text{inicial}}$ fijo (valor medio: 6.58seg)



Vemos que la cafeína tiene un efecto sobre el tiempo de reaccion y que ese efecto no depende del tiempo (*interaccion tiempo-cafeína NS*). Sin embargo, la diferencia entre los grupos parece disminuir conforme pasa el tiempo lo cual nos indicaria que el efecto de la cafeína se pierde a medida que pasa el tiempo. Esto es logico, quizás habria que tomar mas tiempos para poner a prueba esto. Tambien se podria comparar los grupos en algun tiempo en particular (utilizando el intervalo de confianza de la diferencia de medias).

### Problema 10. Con lápiz y papel

Los líquenes son asociaciones mutualistas entre hongos y algas que pueden funcionar en la naturaleza como una unidad. Debido a que los organismos epífitos reciben la mayor parte de sus nutrientes a partir de la atmósfera, son más susceptibles a los factores atmosféricos y, por lo tanto, constituyen sustratos ideales para ser utilizados como bioindicadores. Con el objetivo de evaluar la respuesta de los líquenes ala contaminación atmosférica generada por una ruta que atraviesa un Parque Nacional, se trazaron 15 transectas desde la ruta hacia el interior del parque, 5 en áreas de bosque primario, 5 en bosque secundario y 5 en áreas de palmital. En tres puntos de cada transecta (elegidos según una hipótesis previa de gradiente decreciente de contaminación a 0, 50, y 150 metros) se eligieron 4 árboles de similares características y se contó el número de líquenes en la corteza de cada uno. ¿Existe alguna medida repetida en este estudio? Justifique.

- Identifique el tipo de estudio, la unidad experimental, la variable dependiente, los factores y sus niveles. ¿De qué diseño se trata? ¿Con cuántas réplicas?.
- Escriba el modelo, en general y en términos del problema.

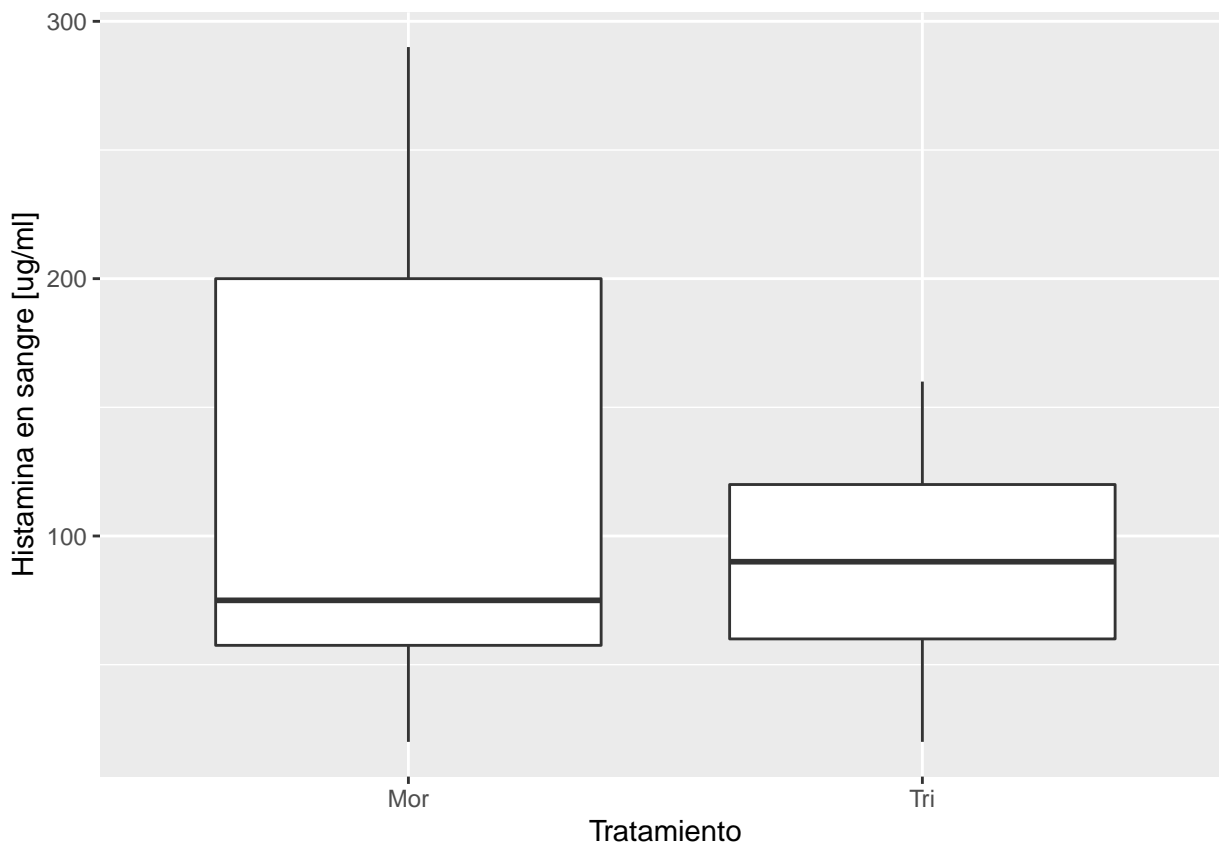
### Problema 11. Perros

La morfina es un potente analgesico que actua a traves de la interaccion con receptores opioides en el sistema nervioso central (SNC). El trimetafan es un agente bloqueante ganglionar que se utiliza como complemento de algunas anestias para inducir hipotension controlada. Existen evidencias que sugieren que ambas drogas podrian inducir la liberacion de histamina como respuesta alergica. Con el objetivo de estudiarlo, se selecciono un grupo de 16 perros mestizos que fueron separados al azar en dos grupos de 8 perros cada uno. Un grupo recibio sulfato de morfina endovenosa y el otro trimetafan endovenoso. Se registro el nivel de histamina en sangre, en  $\mu\text{g/ml}$ , al inicio de la prueba, al minuto y a los 3 minutos. Los resultados en el archivo *Perros.txt*.

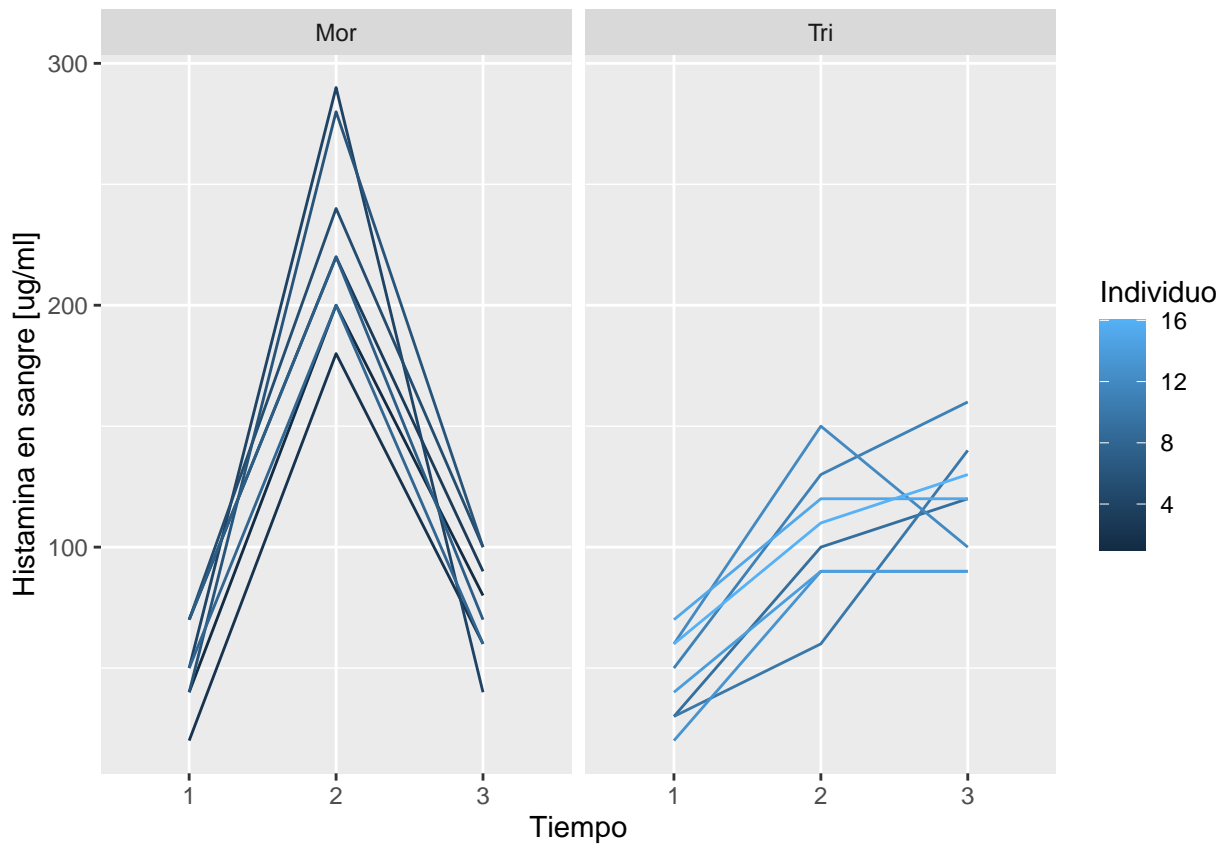
- Identifique el tipo de estudio, la unidad experimental, la variable dependiente, tipo y potencial distribución de probabilidades, las variables explicatorias, tipo y si son de efectos fijos o aleatorios. ¿De que diseño se trata? ¿Con cuantas replicas cuenta para evaluar el efecto de la droga? ¿Agregaria un tratamiento control a este ensayo? ¿En que consistiria dicho tratamiento?.
- Escriba el modelo, en general y en terminos del problema.
- Describa los datos mediante un grafico de perfiles.

```
perros_data <- read.table("/home/pangolin/Documents/materias/biome2/2020/tps/tp5/nuevo/TP_5_Modelos Mixtos/Me
header = TRUE)
perros_df <- perros_data[, c(2, 1, 3, 4)]
perros_df$Tiempo <- as.factor(perros_df$Tiempo)
```

```
library(ggplot2)
p <- ggplot(perros_df, aes(GRUPO, Histamina))
p + geom_boxplot() + labs(x = "Tratamiento", y = "Histamina en sangre [ug/ml]")
```

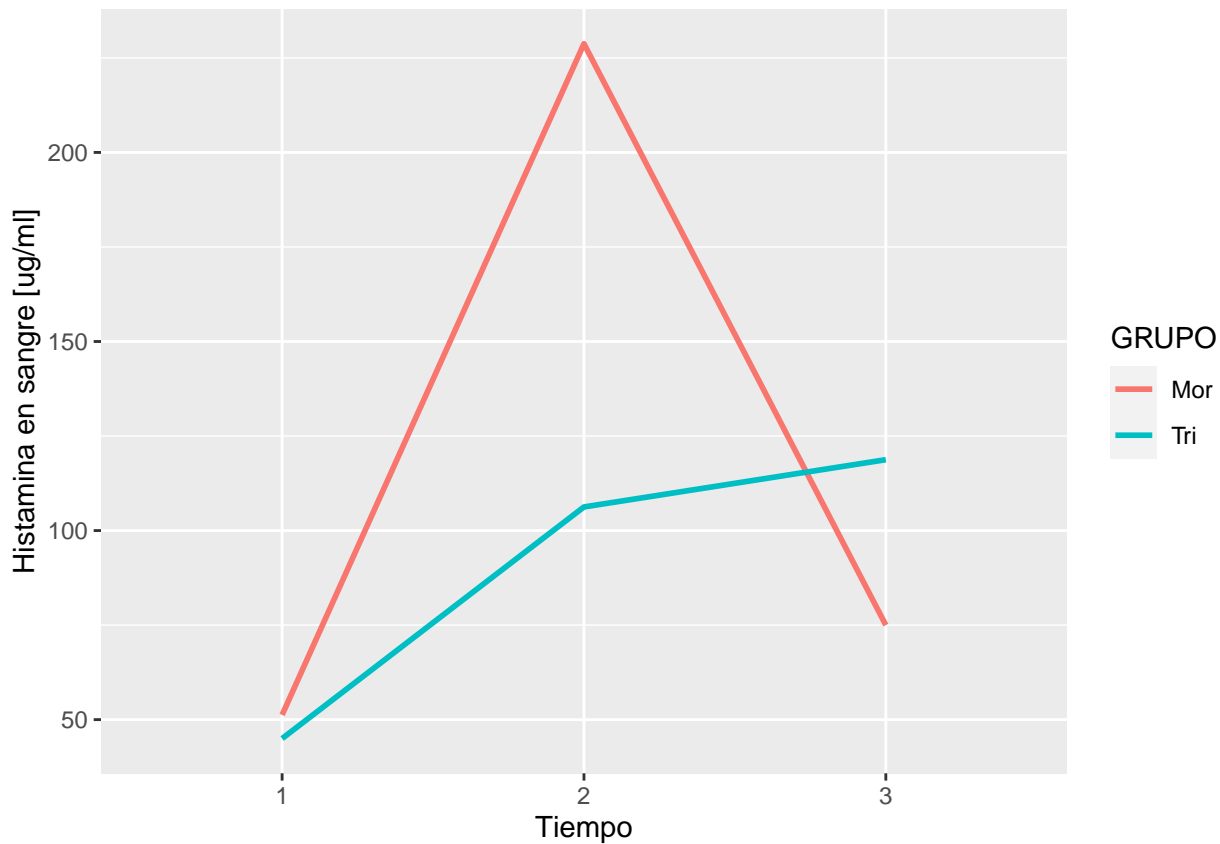


```
# spaghetti plots
p <- ggplot(data = perros_df, aes(x = Tiempo, y = Histamina, group = Individuo))
p + geom_line(aes(colour = Individuo)) + facet_grid(. ~ GRUPO) + labs(y = "Histamina en sangre [ug/ml]")
```



```
## Trataientos en el Tiempo
p <- ggplot(data = perros_df, aes(x = Tiempo, y = Histamina, colour = GRUPO,
  group = GRUPO))
p + geom_smooth(se = FALSE) + labs(y = "Histamina en sangre [ug/ml]")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 0.99
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 2.01
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 6.4961e-17
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 4.0401
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : pseudoinverse used at 0.99
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : neighborhood radius 2.01
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : reciprocal condition number 6.4961e-17
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : There are other near singularities as well. 4.0401
```

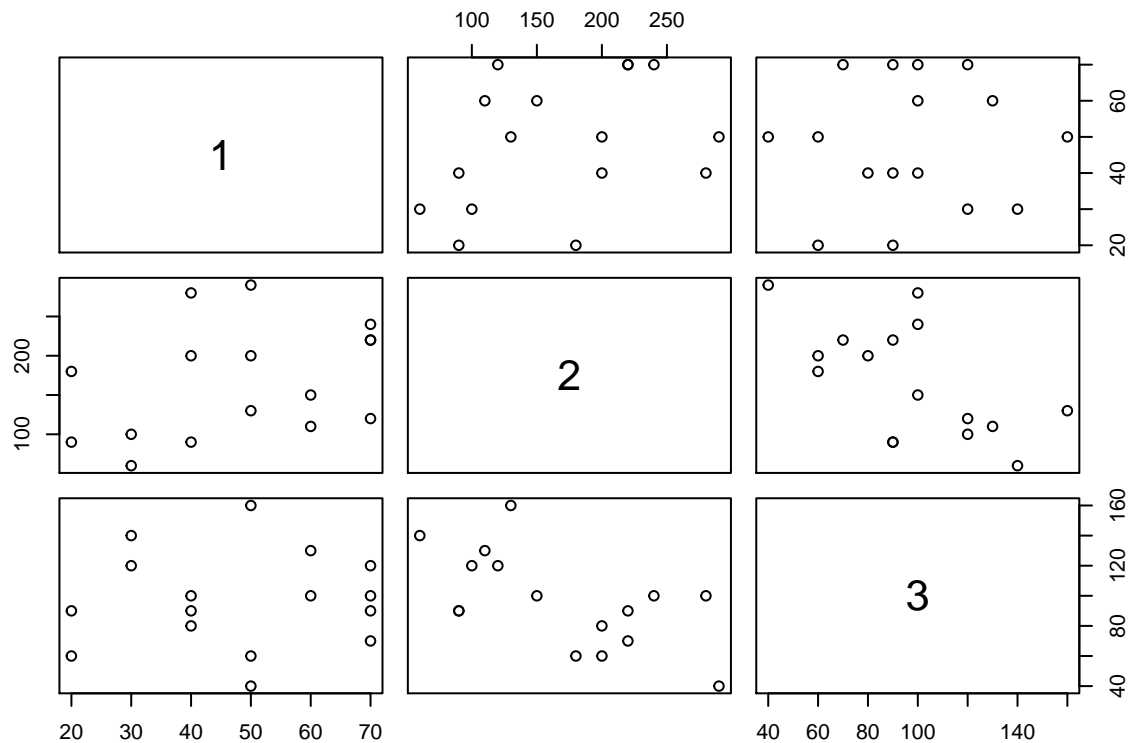


- Ajuste distintas estructuras para la matriz de covarianza. Elija la que a su juicio considere la mas apropiada.

```
# estudiamos las correlaciones entre Tiempos para convertir a
# formato wide (cada Tiempo en una columna distinta)
library(reshape2)
perros_df_wide <- dcast(perros_df, Individuo + GRUPO ~ Tiempo, value.var = "Histamina")
perros_df2 <- perros_df_wide[, 3:5]
summary(perros_df2)
```

```
##          1          2          3
## Min.   :20.00  Min.   : 60.0  Min.   : 40.00
## 1st Qu.:37.50  1st Qu.:107.5  1st Qu.: 77.50
## Median :50.00  Median :165.0  Median : 95.00
## Mean   :48.12  Mean   :167.5  Mean    : 96.88
## 3rd Qu.:62.50  3rd Qu.:220.0  3rd Qu.:120.00
## Max.   :70.00  Max.   :290.0  Max.    :160.00
```

```
plot(perros_df2)
```



```
round(cov(perros_df2[1:3]), 2) #matriz de covarianza
```

```
##      1      2      3
## 1 309.58  441.67  27.08
## 2 441.67 5073.33 -1395.00
## 3  27.08 -1395.00 1009.58
```

```
round(cor(perros_df2[1:3]), 2) #matriz de correlacion
```

```
##      1      2      3
## 1 1.00  0.35  0.05
## 2 0.35  1.00 -0.62
## 3 0.05 -0.62  1.00
```

## modelamos

```
## Ahora modelemos MR, con distintas estructuras de correlación
```

```
library(nlme)
```

```
# Los modelos 1 y 2 requieren tiempos equiespaciados que no se
```

```
# verifica en el caso de estudio Modelo 1: Simetría compuesta.
```

```
# cada individuo con id unico
```

```
m1 <- gls(Histamina ~ GRUPO * Tiempo, correlation = corCompSymm(form = ~1 |
  Individuo), perros_df)
```

```
m1b <- gls(Histamina ~ GRUPO * Tiempo, correlation = corCompSymm(form = ~1 |
  Individuo), weights = varIdent(form = ~1 | Tiempo), perros_df)
```

```
## y MODELANDO VARIANZAS POR TRATAMIENTO?
```

```
m1c <- gls(Histamina ~ GRUPO * Tiempo, correlation = corCompSymm(form = ~1 |
  Individuo), weights = varIdent(form = ~1 | GRUPO), perros_df)
```

```
# Modelo 2: AR1
```

```
m2 <- gls(Histamina ~ GRUPO * Tiempo, correlation = corAR1(form = ~1 |
  Individuo), perros_df)
```

```

# Modelo 2b: AR1, varianzas distintas entre tratamientos
m2b <- gls(Histamina ~ GRUPO * Tiempo, correlation = corAR1(form = ~1 |
  Individuo), perros_df, weights = varIdent(form = ~1 | GRUPO))
# Modelo 2c: Ar1, varianzas distintas entre Tiempos
m2c <- gls(Histamina ~ GRUPO * Tiempo, correlation = corAR1(form = ~1 |
  Individuo), perros_df, weights = varIdent(form = ~1 | Tiempo))

# Los siguientes modelos no tienen el supuesto de tiempos
# equiespaciados -> OK Modelo 3: matriz desestructurada
m3 <- gls(Histamina ~ GRUPO * Tiempo, correlation = corSymm(form = ~1 |
  Individuo), perros_df)
# Modelo 4: autoregresiva continua de orden 1
m4 <- gls(Histamina ~ GRUPO * Tiempo, correlation = corCAR1(form = ~1 |
  Individuo), perros_df)
m4b <- gls(Histamina ~ GRUPO * Tiempo, correlation = corCAR1(form = ~1 |
  Individuo), weights = varIdent(form = ~1 | Tiempo), perros_df)

AIC(m1, m1b, m1c, m2, m2b, m2c, m3, m4, m4b)

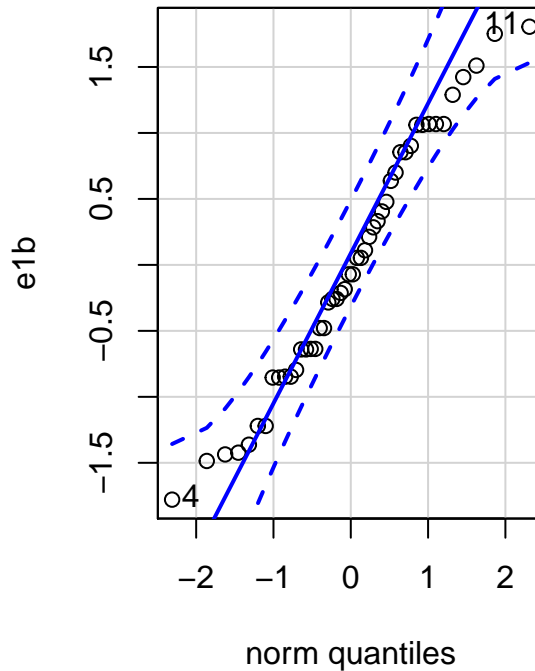
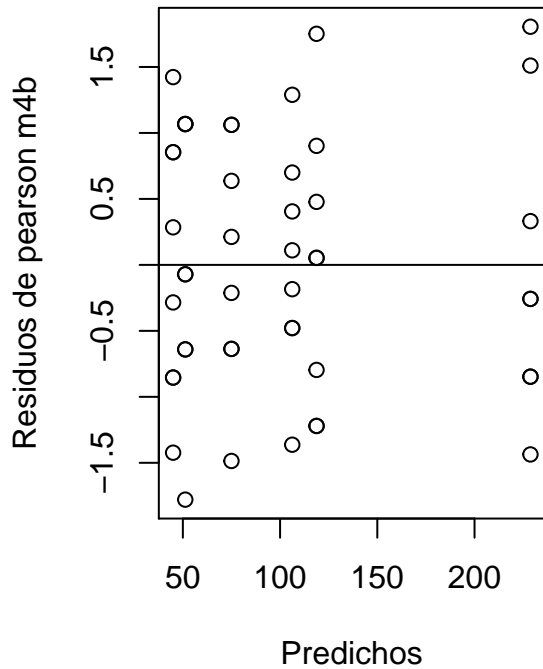
##      df      AIC
## m1    8 419.3199
## m1b  10 416.7917
## m1c   9 420.6534
## m2    8 419.8436
## m2b   9 421.1646
## m2c  10 417.4447
## m3   10 421.7333
## m4    8 419.8436
## m4b  10 417.4447

# Supuestos
e1b <- residuals(m1b, "pearson") # residuos de Pearson
pre1b <- predict(m1b) #predichos
par(mfrow = c(1, 2))
plot(pre1b, e1b, xlab = "Predichos", ylab = "Residuos de pearson m4b",
  main = "Gráfico de dispersión de RE vs PRED m4b", cex.main = 0.8)
abline(0, 0)
car::qqPlot(e1b)

```



Gráfico de dispersión de RE vs PRED m4b



```
## [1] 11 4
```

```
par(mfrow = c(1, 1))
shapiro.test(e1b)
```

```
##
## Shapiro-Wilk normality test
##
## data: e1b
## W = 0.9717, p-value = 0.2945
```

```
anova(m4b)
```

```
## Denom. DF: 42
##          numDF  F-value p-value
## (Intercept)      1 378.3695 <.0001
## GRUP0           1   0.3626 0.5503
## Tiempo          2  99.6107 <.0001
## GRUP0:Tiempo     2  42.0071 <.0001
```

```
summary(m4b)
```

```
## Generalized least squares fit by REML
## Model: Histamina ~ GRUP0 * Tiempo
## Data: perros_df
##      AIC      BIC    logLik
## 417.4447 434.8213 -198.7223
##
## Correlation Structure: Continuous AR(1)
## Formula: ~1 | Individuo
## Parameter estimate(s):
##      Phi
## 0.2372579
```

```
## Variance function:
## Structure: Different standard deviations per stratum
## Formula: ~1 | Tiempo
## Parameter estimates:
##      1      2      3
## 1.000000 1.935208 1.351836
##
## Coefficients:
##              Value Std.Error   t-value p-value
## (Intercept)    51.25  6.190688   8.278563  0.0000
## GRUPOTri      -6.25  8.754955  -0.713881  0.4792
## Tiempo2       177.50 12.110261  14.656992  0.0000
## Tiempo3        23.75 10.125642   2.345530  0.0238
## GRUPOTri:Tiempo2 -116.25 17.126495  -6.787728  0.0000
## GRUPOTri:Tiempo3   50.00 14.319820   3.491664  0.0011
##
## Correlation:
##              (Intr) GRUPOTr Tiemp2 Tiemp3 GRUPOT:T2
## GRUPOTri      -0.707
## Tiempo2       -0.276  0.196
## Tiempo3       -0.565  0.399  0.339
## GRUPOTri:Tiempo2  0.196 -0.276 -0.707 -0.240
## GRUPOTri:Tiempo3  0.399 -0.565 -0.240 -0.707  0.339
##
## Standardized residuals:
##              Min      Q1      Med      Q3      Max
## -1.78470362 -0.67990103 -0.07138814  0.85665774  1.80756793
##
## Residual standard error: 17.50991
## Degrees of freedom: 48 total; 42 residual
```

*Elegimos ese porque tiene misma cantidad de parametros estimados que el de menor AIC y no exige supuesto de equiespaciados de tiempo.*

- Concluya acerca del efecto de la morfina y el trimetafan en el nivel de histamina en sangre de perros.

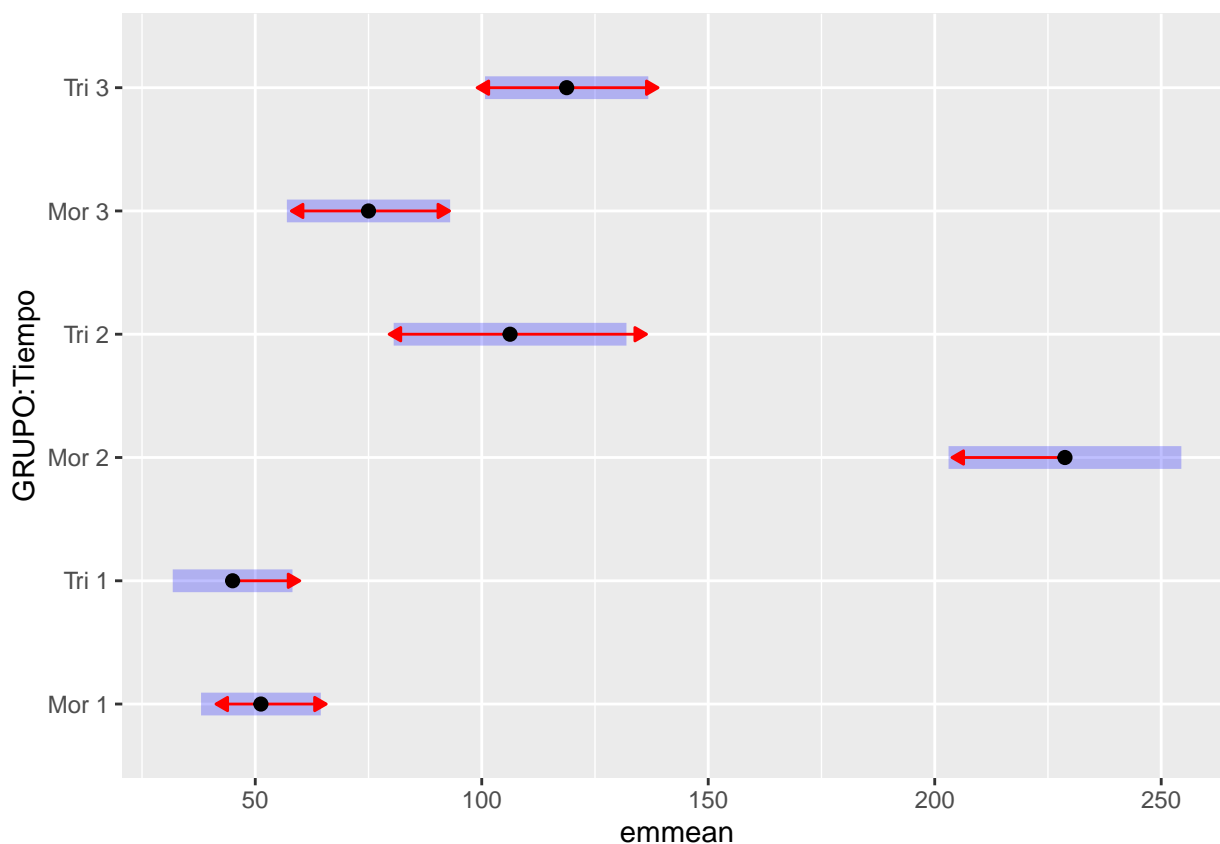
```
library(emmeans)
(Tukey_tiempo <- emmeans(m4b, pairwise ~ GRUPO * Tiempo))
```

```
## Warning in sweep(X, 1, sqrt(weights), "*"): STATS is longer than the extent of
## 'dim(x)[MARGIN]'
```

```
## $emmeans
## GRUPO Tiempo emmean SE df lower.CL upper.CL t.ratio p.value
## Mor 1 51.2 6.19 14.7 38.0 64.5 8.279 <.0001
## Tri 1 45.0 6.19 14.7 31.8 58.2 7.269 <.0001
## Mor 2 228.8 11.98 14.0 203.1 254.4 19.094 <.0001
## Tri 2 106.2 11.98 14.0 80.6 131.9 8.869 <.0001
## Mor 3 75.0 8.37 13.4 57.0 93.0 8.962 <.0001
## Tri 3 118.8 8.37 13.4 100.7 136.8 14.190 <.0001
##
## Degrees-of-freedom method: satterthwaite
## Confidence level used: 0.95
##
## $contrasts
## contrast estimate SE df lower.CL upper.CL t.ratio p.value
## Mor 1 - Tri 1 6.25 8.75 14.7 -22.3 34.77 0.714 0.9772
## Mor 1 - Mor 2 -177.50 12.11 17.3 -216.2 -138.84 -14.657 <.0001
## Mor 1 - Tri 2 -55.00 13.49 20.5 -97.3 -12.71 -4.079 0.0064
## Mor 1 - Mor 3 -23.75 10.13 25.7 -54.9 7.39 -2.346 0.2130
## Mor 1 - Tri 3 -67.50 10.41 25.1 -99.6 -35.43 -6.484 <.0001
```

```
## Tri 1 - Mor 2 -183.75 13.49 20.5 -226.0 -141.46 -13.626 <.0001
## Tri 1 - Tri 2 -61.25 12.11 17.3 -99.9 -22.59 -5.058 0.0011
## Tri 1 - Mor 3 -30.00 10.41 25.1 -62.1 2.07 -2.882 0.0764
## Tri 1 - Tri 3 -73.75 10.13 25.7 -104.9 -42.61 -7.283 <.0001
## Mor 2 - Tri 2 122.50 16.94 14.0 66.9 178.07 7.230 0.0001
## Mor 2 - Mor 3 153.75 12.88 20.1 113.3 194.23 11.934 <.0001
## Mor 2 - Tri 3 110.00 14.61 23.8 64.8 155.22 7.527 <.0001
## Tri 2 - Mor 3 31.25 14.61 23.8 -14.0 76.47 2.138 0.3024
## Tri 2 - Tri 3 -12.50 12.88 20.1 -53.0 27.98 -0.970 0.9221
## Mor 3 - Tri 3 -43.75 11.84 13.4 -82.8 -4.70 -3.697 0.0246
##
## Degrees-of-freedom method: satterthwaite
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 6 estimates
## P value adjustment: tukey method for comparing a family of 6 estimates
```

```
plot(Tukey_tiempo, comparisons = TRUE)
```



Se observa que el trimetafan aumenta la liberación de histamina por encontrarse un aumento entre el minuto 2 y el minuto 1 de trimetafan que se mantiene durante el tercer minuto. Para el caso de la morfina se ve una dinámica más rápida, incrementándose mucho la liberación de histamina entre el minuto 1 y 2, pero disminuyendo rápidamente también a valores basales en el minuto 3. Sin embargo, el valor máximo de liberación de histamina alcanzado por la morfina es mayor que el nivel alcanzado por el trimetafan.

Se deberían incluir los valores de los IC de la comparación de medias.

## Problema 12. A sacarnos la duda del millón....

Para complementar el “trabajo de hormiga” que venimos haciendo desde TPs anteriores (base de datos *Hormigas.txt*), les proponemos que respondan las preguntas que nos fuimos haciendo, pero ahora teniendo en cuenta el día en que se realizó la experiencia (algo que varios se cuestionaron si no debía considerarse desde un principio). Entonces, estudiemos:

- ¿Cómo se modifica la masa de tóxico ingerida por las hormigas en función del cebo consumido, comparando dos

cebos alternativos (SAC 30 y SAC 68) con un cebo comercial (CC)?.

- ¿Difiere el tiempo de ingesta de las hormigas entre los tratamientos?.
- ¿Difiere el numero de pausas entre los tratamientos aplicados?.
- ¿Cual es la diferencia entre el factor día de este problema y el factor tiempo de los problemas anteriores?.