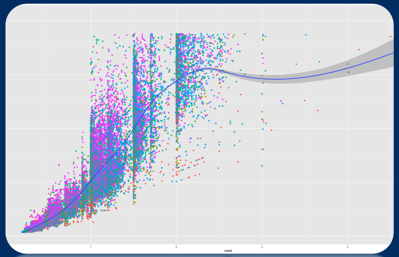
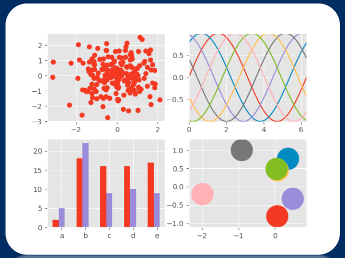
Curso: Introducción a R – Día 5
Modelos de Efectos Mixtos y Aplicaciones





Presentan:

Dr. Ulises Olivares Pinto

Dr. Alberto Prado Farías

Escuela Nacional de Estudios Superiores Unidad Juriquilla

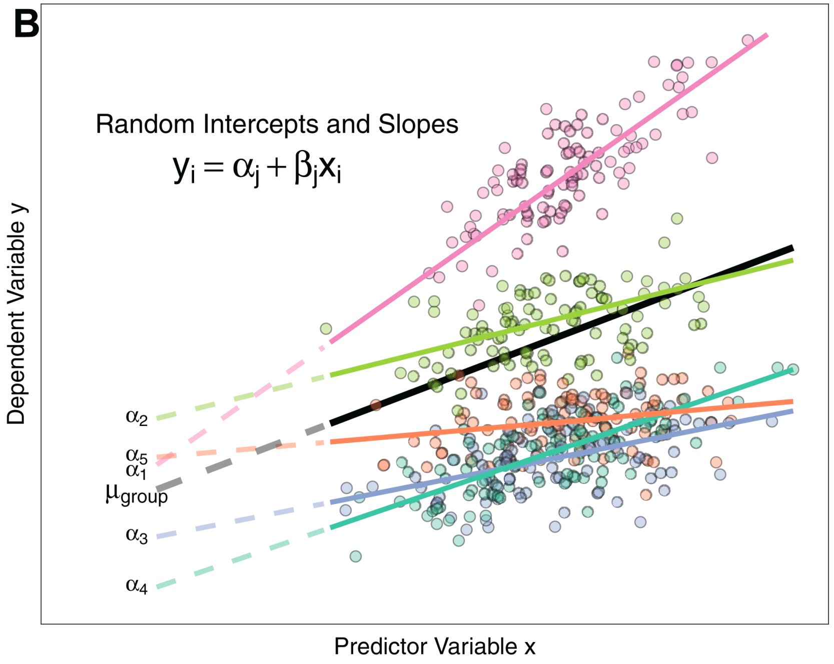
Contenido

1. Introducción.
   - ✓ Efectos fijos vs aleatorios
   - ✓ Efectos anidados vs cruzados.
2. Selección de modelos.
3. Visualización y validación de modelos.

1. Introducción – Modelos de efectos mixtos

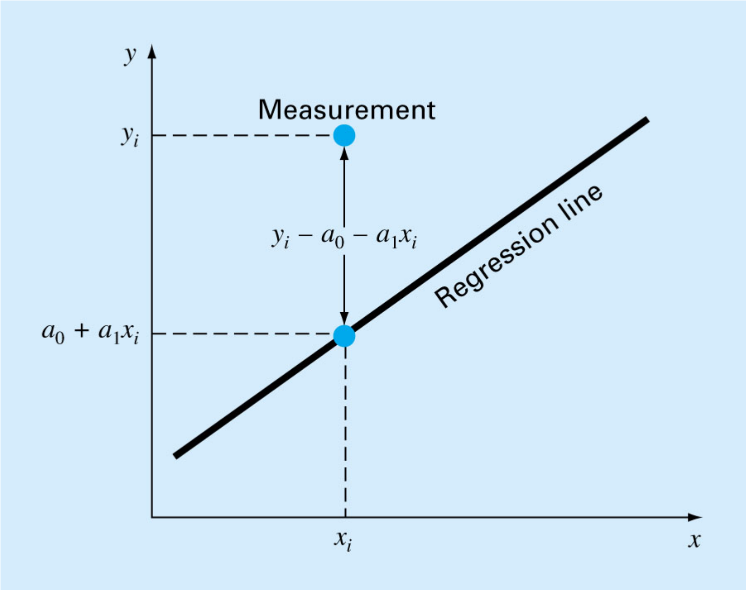
Son modelos estadísticos capaces de representar efectos **fijos** y efectos **aleatorios.** Dichos modelos son muy útiles en múltiples disciplinas por su capacidad de representar comportamientos y/o procesos complejos.

- Los modelos de efectos mixtos son muy extensos y complejos, en esta sesión se brindará una breve introducción a los modelos lineales mixtos (LMM).



1. Introducción – Efectos fijos vs aleatorios

- **Efectos fijos:** Son aquellos parámetros que son considerados constantes y se atribuyen a un **conjunto finito de niveles o valores**, en los cuales se tiene un interés particular. **No cambian** durante todo el proceso experimental. (modelos lineales).
  - ✓ Ejemplo:
    - Sexo: Masculino / femenino
    - Dificultad: Fácil / media / difícil
    - Dosis: Alta / baja

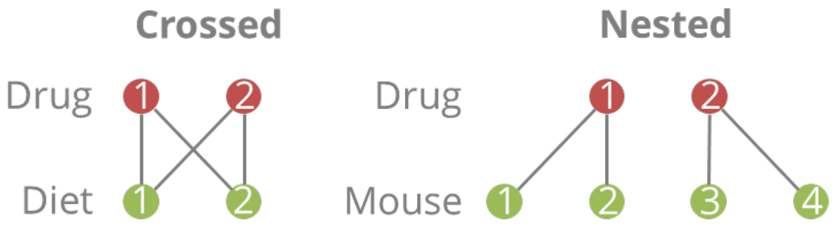


- **Efectos aleatorios:** Son aquellos parámetros que provienen de una muestra de una población que usualmente se selecciona de forma aleatoria. (Generalmente se asocian con un error en un modelo).
  - ✓ Ejemplo:
    - Sujetos de prueba
    - Estímulos
    - Escenarios

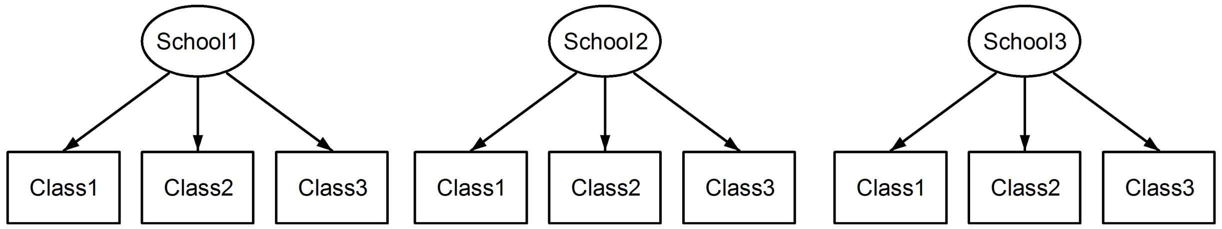
1. Introducción – Efectos anidados vs cruzados

- **Efectos anidados:** Se refiere a las variables (propiedades) de los sujetos.

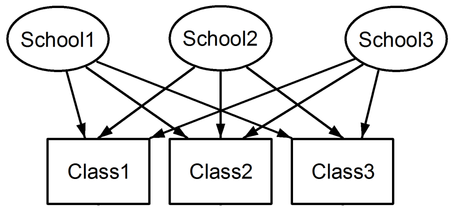
- **Efectos cruzados:** Variables entre sujetos o entre poblaciones.



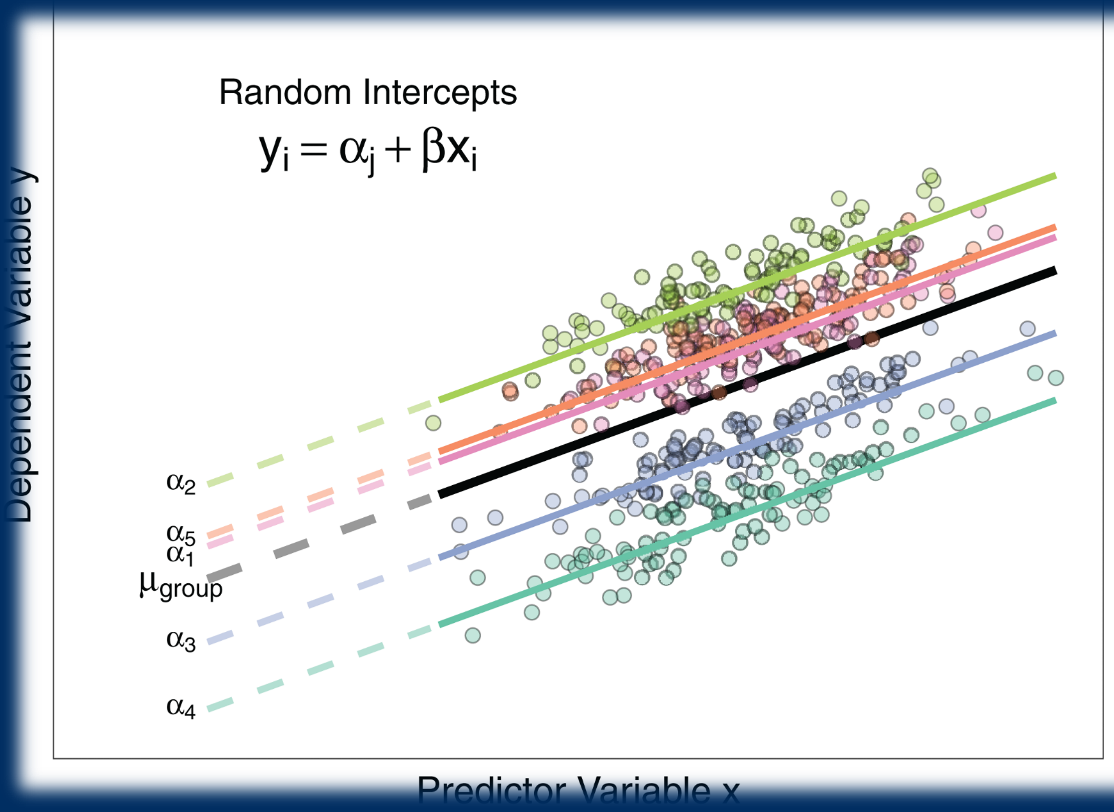
Anidado

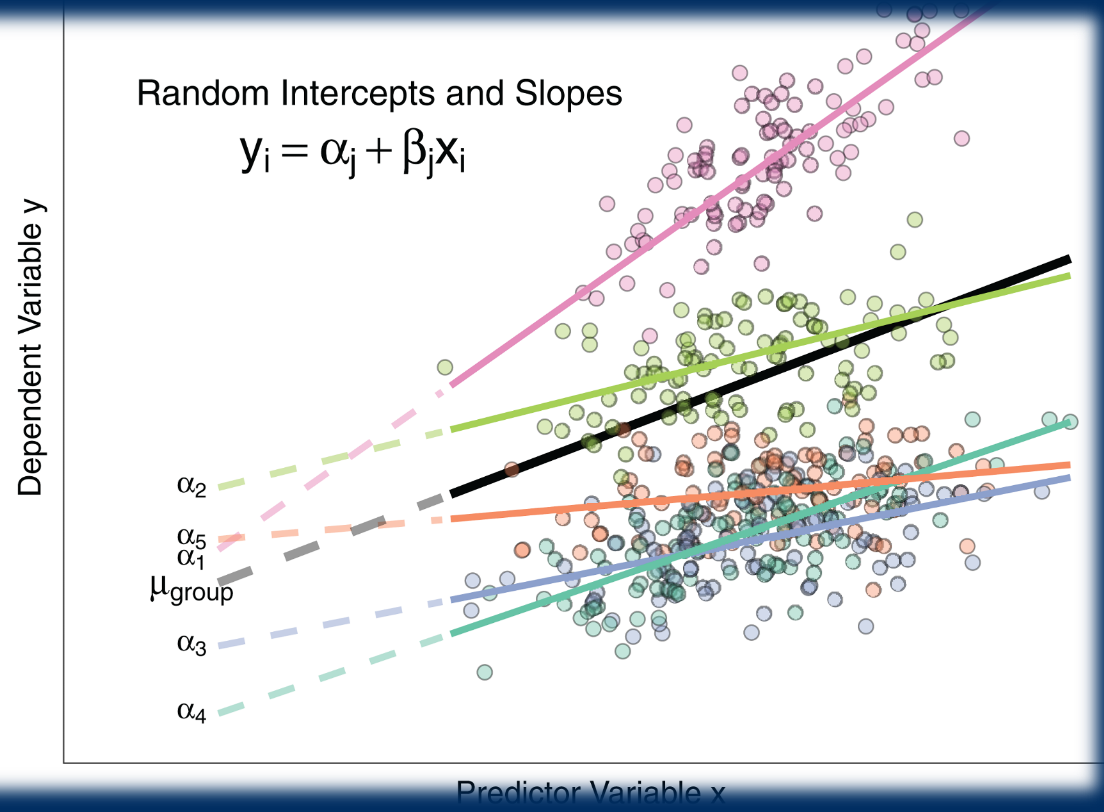


Cruzado



2. Selección de modelos





2. Selección de modelos - Caso de estudio

Caso de estudio: Estudio de sueño

Se sometió a varios sujetos de prueba a un estudio de sueño, a cada sujeto se le midió la reacción promedio en ms después de privarlos del sueñp por varios días.

2. Selección de modelos - Caso de estudio

```r
model1 <- lm(Reaction ~ Days, sleepstudy)
summary(model1)
```

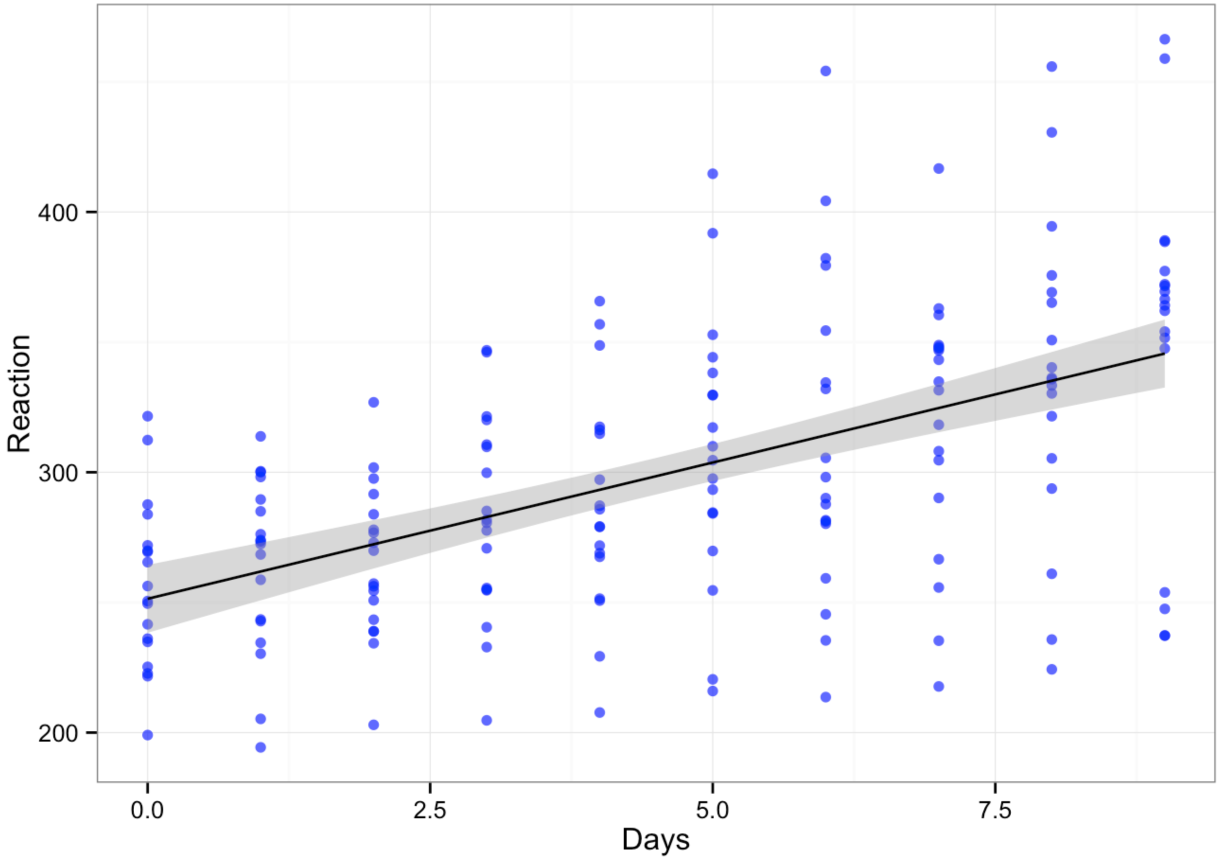
```
##
## Call:
## lm(formula = Reaction ~ Days, data = sleepstudy)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -110.848  -27.483    1.546   26.142  139.953
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  251.405      6.610  38.033  < 2e-16 ***
## Days          10.467      1.238   8.454 9.89e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 47.71 on 178 degrees of freedom
## Multiple R-squared:  0.2865, Adjusted R-squared:  0.2825
## F-statistic: 71.46 on 1 and 178 DF,  p-value: 9.894e-15
```

```r
coef(model1)
```

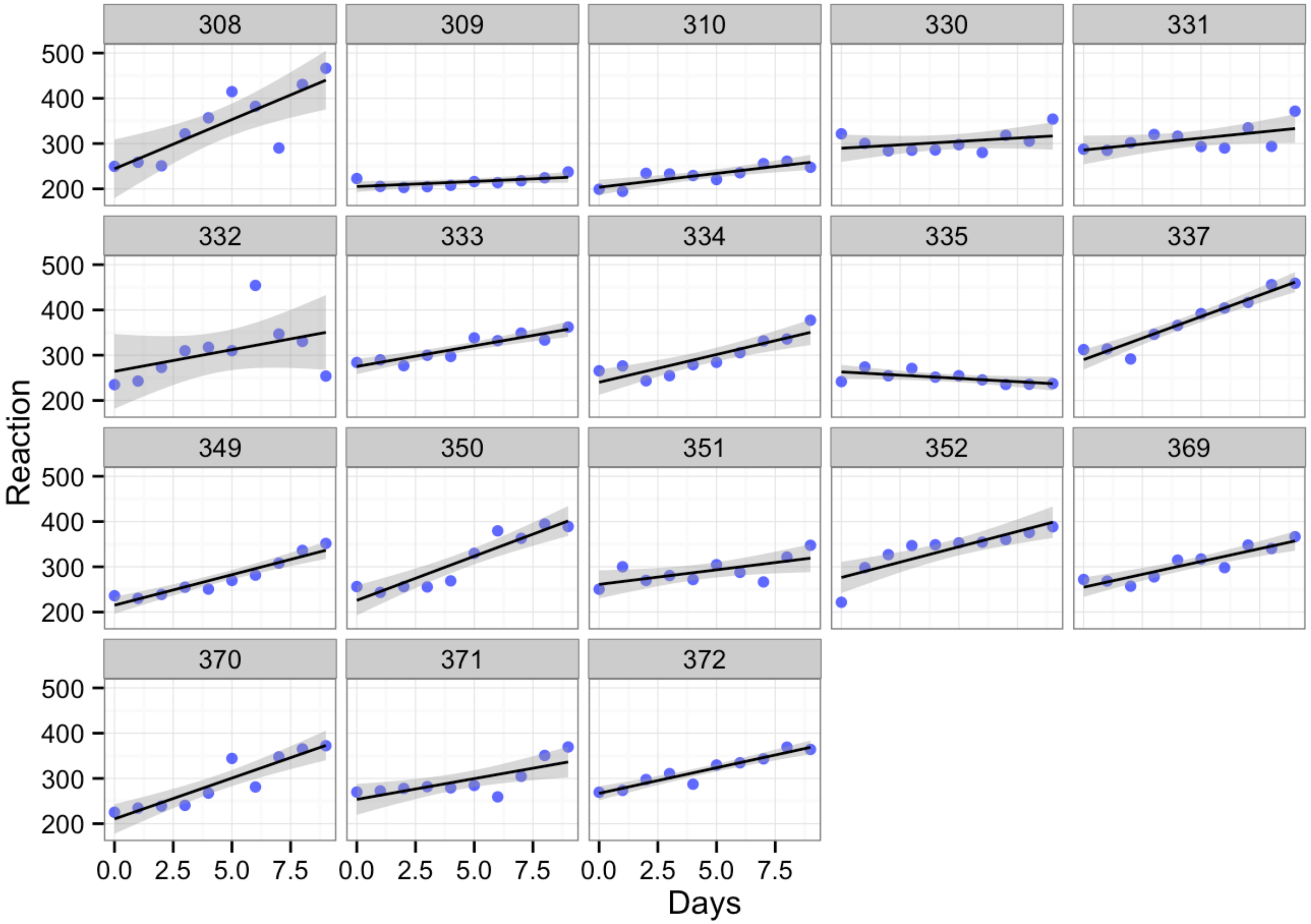
```
## (Intercept)        Days
##   251.40510    10.46729
```

```r
round(confint(model1), 2)
```

```
##              2.5 % 97.5 %
## (Intercept) 238.36 264.45
## Days          8.02  12.91
```



2. Selección de modelos - Caso de estudio



2. Selección de modelos - Caso de estudio

```r
model308 <- lm(Reaction ~ Days,
               data = subset(sleepstudy, Subject == 308))
model335 <- lm(Reaction ~ Days,
               data = subset(sleepstudy, Subject == 335))

summary(model308)
```

```
##
## Call:
## lm(formula = Reaction ~ Days, data = subset(sleepstudy, Subject ==
##     308))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -106.397   -4.098    9.688   22.269   61.674
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   244.19      28.08   8.695 2.39e-05 ***
## Days           21.77       5.26   4.137  0.00326 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 47.78 on 8 degrees of freedom
## Multiple R-squared:  0.6815, Adjusted R-squared:  0.6417
## F-statistic: 17.12 on 1 and 8 DF,  p-value: 0.003265
```

```r
summary(model335)
```

```
##
## Call:
## lm(formula = Reaction ~ Days, data = subset(sleepstudy, Subject ==
##     335))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.4264  -3.8697  -0.1774   4.5403  16.4105
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  263.035      6.694  39.296 1.93e-10 ***
## Days          -2.881      1.254  -2.298   0.0506 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.39 on 8 degrees of freedom
## Multiple R-squared:  0.3976, Adjusted R-squared:  0.3223
## F-statistic:  5.28 on 1 and 8 DF,  p-value: 0.05065
```

```r
round(confint(model308), 2)
```

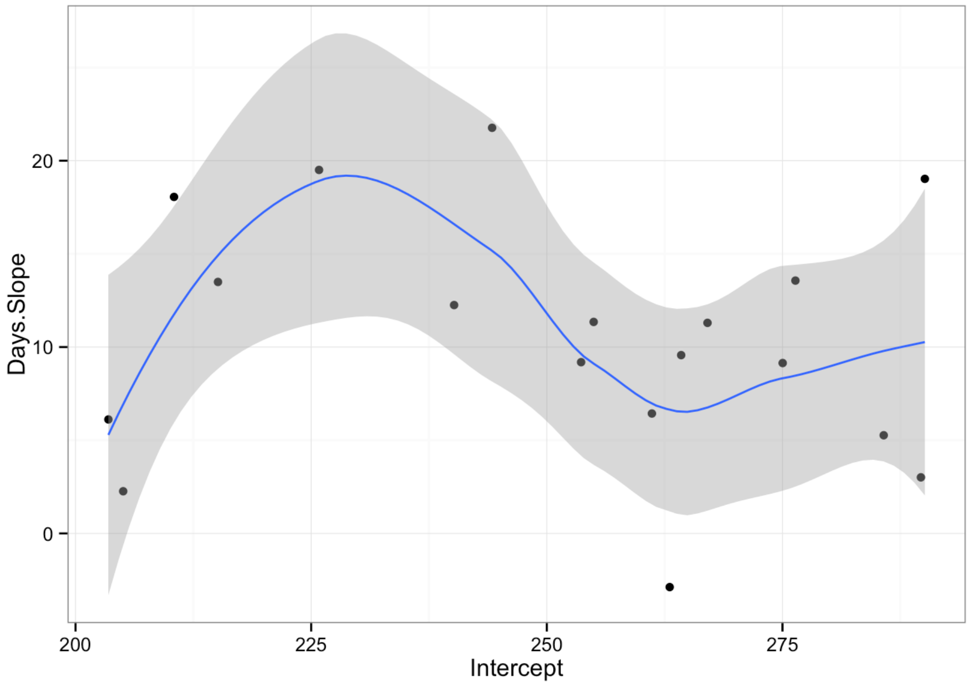
```
##              2.5 % 97.5 %
## (Intercept) 179.43 308.95
## Days          9.63  33.90
```

```r
round(confint(model335), 2)
```

```
##              2.5 % 97.5 %
## (Intercept) 247.60 278.47
## Days         -5.77   0.01
```

2. Selección de modelos - Caso de estudio

```
geom_smooth(method = "loess")
```

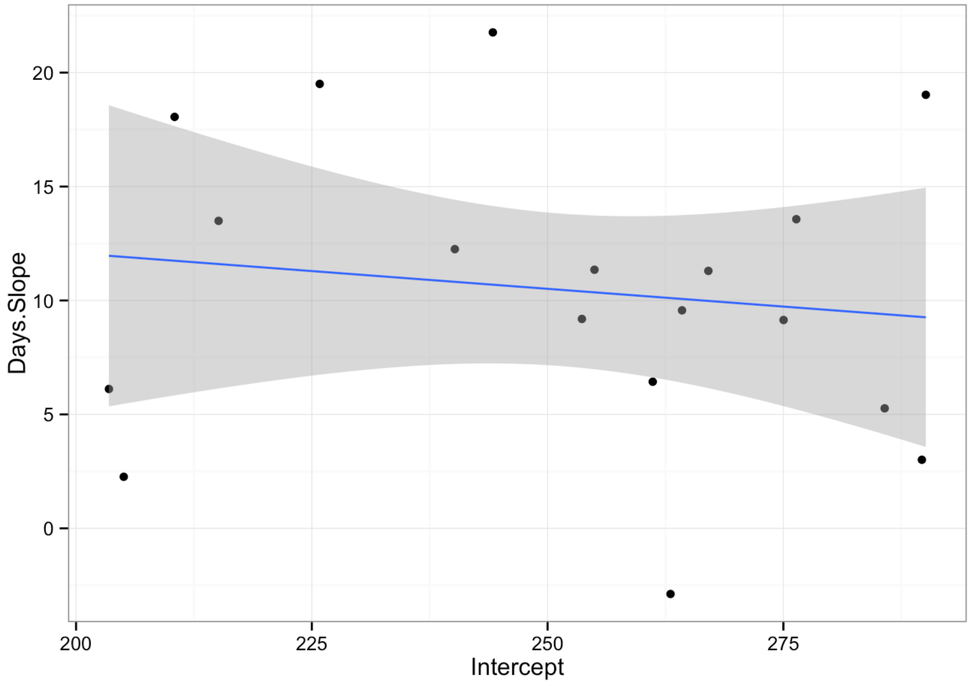


- Modelo lineal para todos los sujetos de prueba

```
model2 <- lmList(Reaction ~ Days | Subject, sleepstudy)
model2.coefficients <- coef(model2)
names(model2.coefficients) <- c("Intercept", "Days.Slope")
```

2. Selección de modelos - Caso de estudio

```
ggplot(model2.coefficients, aes(x = Intercept, y = Days.Slope))
```



Sujetos con una reacción inicial lenta tienen una pendiente menos pronunciada. Son menos afectados por una falta de sueño.

```
model3 <- lm(Days.Slope ~ Intercept, model2.coefficients)
summary(model3)
```

```
##
## Call:
## lm(formula = Days.Slope ~ Intercept, data = model2.coefficients)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.9860  -4.0312   0.2458   3.3818  11.0727
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 18.30016   14.18878   1.290    0.215
## Intercept   -0.03116    0.05609  -0.555    0.586
##
## Residual standard error: 6.696 on 16 degrees of freedom
## Multiple R-squared:  0.01892,    Adjusted R-squared:  -0.0424
## F-statistic: 0.3086 on 1 and 16 DF,  p-value: 0.5862
```

No es estadísticamente significativo $p = 0.582$

2. Selección de modelos - Caso de estudio – Modelos mixtos

Ejemplo: Suponga se examinan algunas rebanadas de manzana, las cuales están tratadas con 5 compuestos **(A, B, C, D, E)** que actúan como conservadores con la finalidad de extender la vida de anaquel.

- ✓ Respuesta: Vida de anaquel
- ✓ Factor: Conservador
- ✓ Tratamiento: Nivel del factor: A, B, C, D, E
- ✓ Efecto: Impacto del compuesto en la vida de anaquel

$\mu = promedio\ de\ vida\ de\ anaquel\ de\ toda\ la\ población$

$\mu_A = promedio\ de\ vida\ de\ anaquel\ de\ la\ población\ con\ el\ tratamiento\ A$

$efecto\ de\ A = \mu_A - \mu$

- 10 frutas individuales son elegidas de forma aleatoria de una población y los 5 conservadores se asignan aleatoriamente a 5 porciones de fruta.

1. Introducción – Efectos fijos vs aleatorios

¿Cuáles son los efectos fijos y aleatorios para el ejemplo anterior?

R.

Efecto fijo: Conservador

Efecto aleatorio: Fruta

Los modelos de efectos mixtos incorporan de forma natural dependencias en el modelo .

El uso de factores aleatorios da con frecuencia estimaciones más precisas.
Los modelos mixtos contienen tanto efectos fijos como aleatorios.

2. Dataframe de vías de ferrocarril

Este dataframe esta disponible en las librerías MEMSS y NLME

1. Instalar la librería **nlme**

2. Importar la misma libreria

3. Cargar los datos **data(Rail)**

4. Observar los datos **?Rail**

Rail {nlme} R Documentation

Evaluation of Stress in Railway Rails

Description

The `Rail` data frame has 18 rows and 2 columns.

Format

This data frame contains the following columns:

Rail

an ordered factor identifying the rail on which the measurement was made.

travel

a numeric vector giving the travel time for ultrasonic head-waves in the rail (nanoseconds). The value given is the original travel time minus 36,100 nanoseconds.

Details

Devore (2000, Example 10.10, p. 427) cites data from an article in *Materials Evaluation* on "a study of travel time for a certain type of wave that results from longitudinal stress of rails used for railroad track."

Source

Pinheiro, J. C. and Bates, D. M. (2000), *Mixed-Effects Models in S and S-PLUS*, Springer, New York. (Appendix A.26)

Devore, J. L. (2000), *Probability and Statistics for Engineering and the Sciences (5th ed)*, Duxbury, Boston, MA.

[Package *nlme* version 3.1-131.1 Index]

2. Dataframe de vías de ferrocarril

- Se seleccionarón aleatoriamente **seis vías** de un conjunto de vias.
- Cada vía fue muestreada en **tres ocasiones** distintas
- Se midio el tiempo que tarda una onda ultrasonica en viajar a lo largo de cada vía.
- La vía es un factor y el tiempo de viaje es la respuesta.

```
   Rail travel
1     1     55
2     1     53
3     1     54
4     2     26
5     2     37
.
.
.
15    5     50
16    6     80
17    6     85
18    6     83
```

```
> summary(Rail)
 Rail        travel
 2:3    Min.   : 26.00
 5:3    1st Qu.: 50.25
 1:3    Median : 66.50
 6:3    Mean   : 66.50
 3:3    3rd Qu.: 85.00
 4:3    Max.   :100.00
```

2. Dataframe de vías de ferrocarril

- Realizar un gráfico de dispersion x = travel y= Rail y asignar esteticas útiles utilizando ggplot.
  - ✓ ¿Como es la variabilidad entre vías con respecto a la variabilidad en la misma vía?
  - ✓ ¿La variabilidad en la misma vía es constante?
  - ✓ ¿Cómo es el promedio de viaje entre vías?

```
   Rail travel
1     1     55
2     1     53
3     1     54
4     2     26
5     2     37
.
.
.
15    5     50
16    6     80
17    6     85
18    6     83
```

```
> summary(Rail)
 Rail        travel
 2:3    Min.    : 26.00
 5:3    1st Qu.: 50.25
 1:3    Median : 66.50
 6:3    Mean   : 66.50
 3:3    3rd Qu.: 85.00
 4:3    Max.   :100.00
```

2.1 Dataframe de vías de ferrocarril – Efecto fijo

Se asume que el efecto del riel es **fijo** y entonces:

$$y_{ij} = \beta_i + e_{ij} \qquad i = 1, \ldots, 6 \quad j = 1, 2, 3$$

Donde:

Y_{ij}: *Tiempo de viaje de la observación j en la vía i*

β_i: *Tiempo promedio de viaje en la vía i*

e_{ij}: Conjunto de valores independientes que siguen una distribución norma con media 0 y varianza σ^2

2.1 Dataframe de vías de ferrocarril – Efectos fijos

Ejercicio:

1. Generar un modelo lineal (travel ~Rail).

2. Calcula el coeficiente de correlación y covarianza de ambas variables rail y travel

3. Revisar el modelo con la instrucción **summary()**

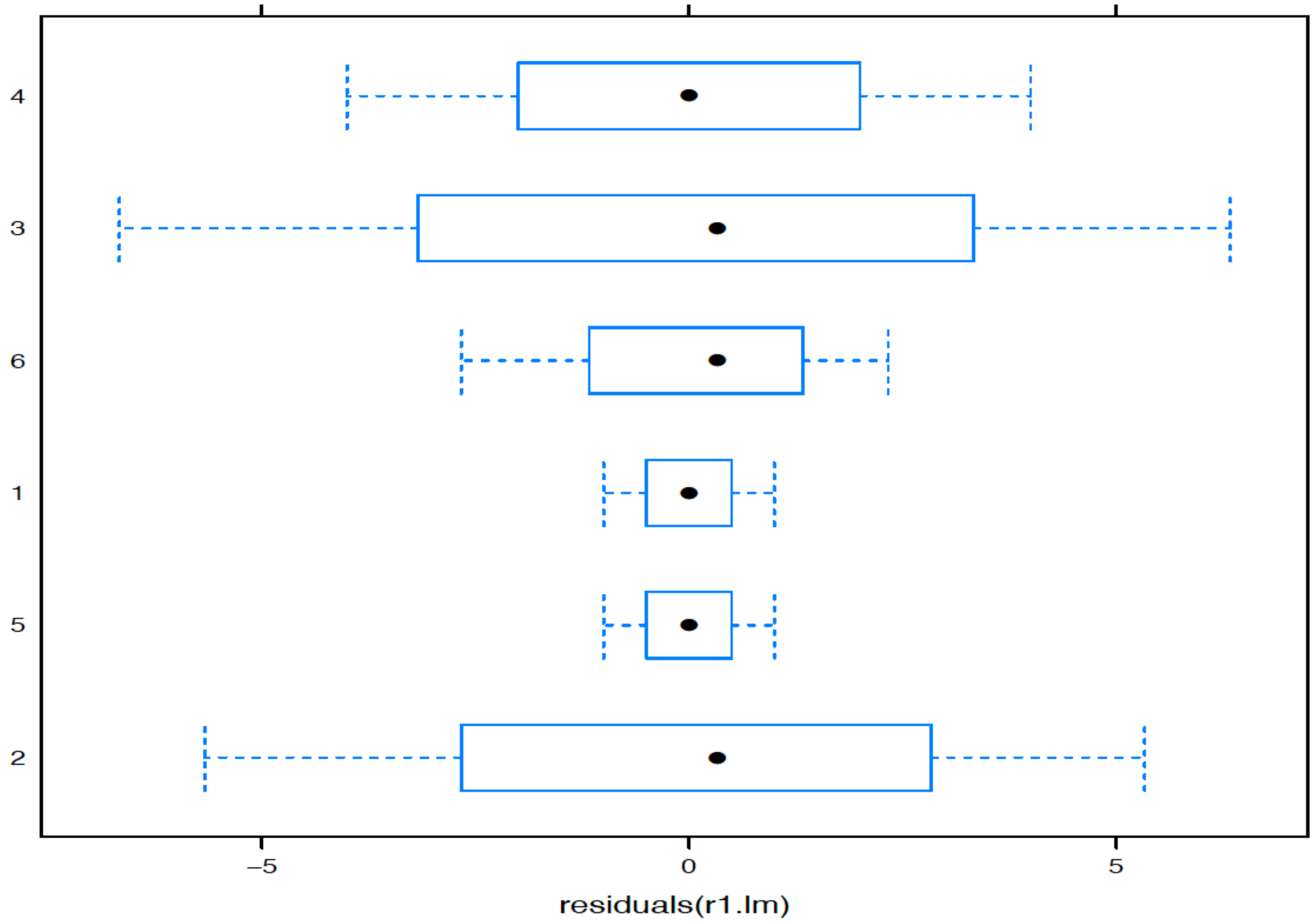
4. ¿Qúe tan bien se ajusta el modelo?

```
Call:
lm(formula = travel ~ Rail - 1, data = Rail)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.6667 -1.0000  0.1667  1.0000  6.3333 

Coefficients:
      Estimate Std. Error t value Pr(>|t|)    
Rail2   31.667      2.321   13.64 1.15e-08 ***
Rail5   50.000      2.321   21.54 5.86e-11 ***
Rail1   54.000      2.321   23.26 2.37e-11 ***
Rail6   82.667      2.321   35.61 1.54e-13 ***
Rail3   84.667      2.321   36.47 1.16e-13 ***
Rail4   96.000      2.321   41.35 2.59e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.021 on 12 degrees of freedom
Multiple R-squared:  0.9978,	Adjusted R-squared:  0.9967 
F-statistic: 916.6 on 6 and 12 DF,  p-value: 2.971e-15
```

$$\beta_2 = 31.67\,, \ldots, \beta_4 = 96.00$$

2.1 Dataframe de vías de ferrocarril – Efectos fijos

```
library(lattice)
with(rail, bwplot(Rail ~ residuals(r1.lm)))
```



2.2 Dataframe de vías de ferrocarril – Efectos fijos

- El modelo de efectos fijos dio un buen resumen de los datos, no obstante el interés se centra casi en su totalidad en la población de las vías.

- Dicho modelo no es preciso debido a que las 3 observaciones en las vías son claramente no dependientes.

- Por tal motivo, se explorarán los modelos de efectos mixtos

2.2 Dataframe de vías de ferrocarril – Efectos Mixtos

- Se assume que los rieles se representarán como efectos aleatorios:

$$y_{ij} = \beta + b_i + e_{ij}$$

- Donde:
  - ✓ $Y_{ij}: Tiempo\ de\ viaje\ de\ la\ observación\ j\ en\ la\ vía\ i$
  - ✓ $\beta\ : Tiempo\ promedio\ de\ viaje\ en\ la\ población$
  - ✓ $b_i: Es\ la\ desviación\ de \beta\ \ para\ la\ vía\ i$
  - ✓ $e_{ij}: es\ la\ desv.\, de\ la\ obs.\, j\ en\ la\ vía\ i\ con\ respecto\ al\ tiempo\ prom.\, de\ viaje\ en\ la\ vía\ i$

$$b_i \overset{iid}{\sim} N(0, \sigma_b^2) \qquad e_{ij} \overset{iid}{\sim} N(0, \sigma^2)\ .$$

Se assume que existe una correlación en las observaciones de la misma vía.

$$\text{corr} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma^2}\ .$$

2.2 Dataframe de vías de ferrocarril – Efectos Mixtos

2. Instalación de librería

```r
install.packages("lme4")
library(lme4)
```

2. Ayuda de función ?lmer

lmer {lme4} R Documentation

Fit Linear Mixed-Effects Models

Description

Fit a linear mixed-effects model (LMM) to data, via REML or maximum likelihood.

Usage

```r
lmer(formula, data = NULL, REML = TRUE, control = lmerControl(),
     start = NULL, verbose = 0L, subset, weights, na.action,
     offset, contrasts = NULL, devFunOnly = FALSE, ...)
```

Arguments

`formula`	a two-sided linear formula object describing both the fixed-effects and random-effects part of the model, with the response on the left of a ~ operator and the terms, separated by + operators, on the right. Random-effects terms are distinguished by vertical bars (|) separating expressions for design matrices from grouping factors. Two vertical bars (||) can be used to specify multiple uncorrelated random effects for the same grouping variable. (Because of the way it is implemented, the ||-syntax *works only for design matrices containing numeric (continuous) predictors*; to fit models with independent categorical effects, see dummy or the `lmer_alt` function from the `afex` package.)
`data`	an optional data frame containing the variables named in `formula`. By default the variables are taken from the environment from which `lmer` is called. While `data` is optional, the package authors *strongly* recommend its use, especially when later applying methods such as `update` and `drop1` to the fitted model (*such methods are not guaranteed to work properly if `data` is omitted*). If `data` is omitted, variables will be taken from the environment of `formula` (if specified as a formula) or from the parent frame (if specified as a character vector).

2.2 Dataframe de vías de ferrocarril – Efectos Mixtos

De acuerdo con la documentación.

¿Cómo podrían generar su modelo de efectos mixtos?

```
?lmer()

r2.lme <- lmer(travel ~ 1 + (1|Rail), REML=FALSE, data=Rail)
```

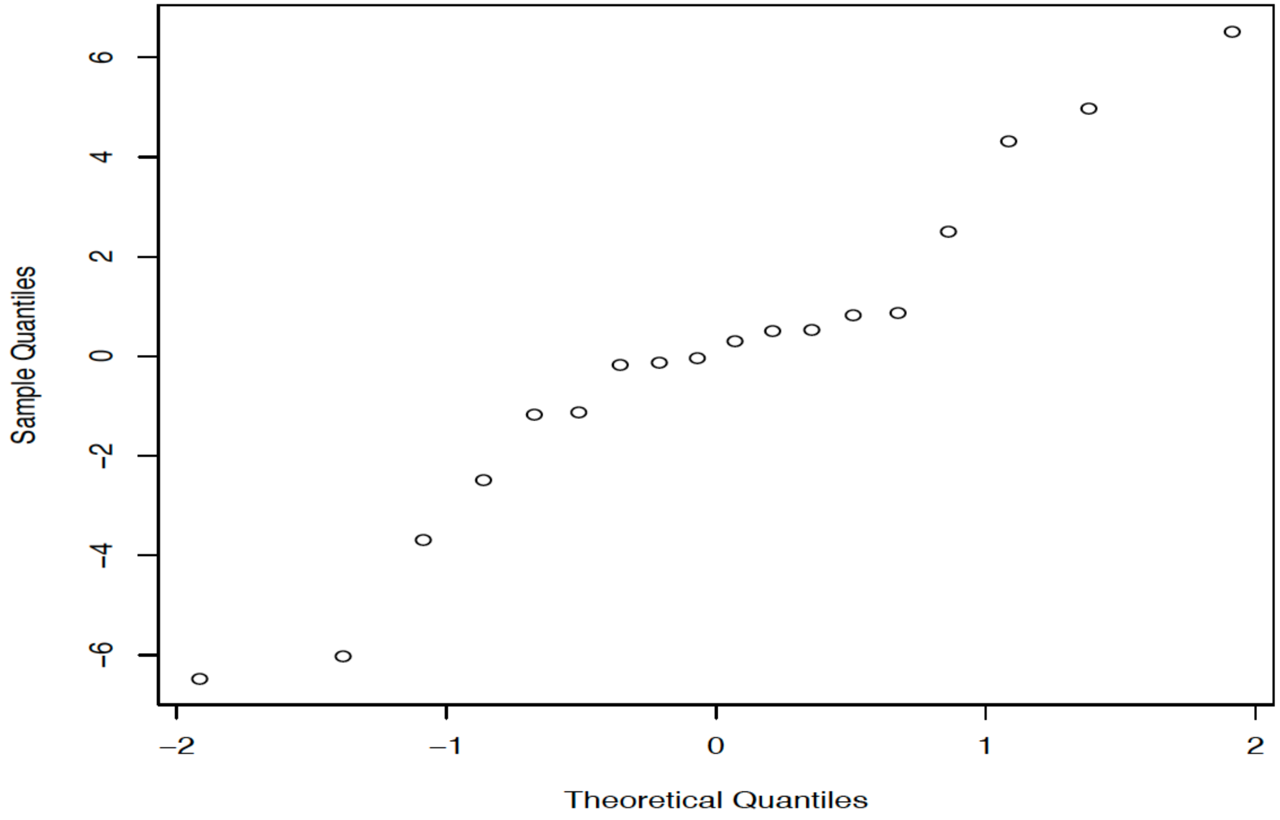
- (1 | Rail) => Existe solo un factor el cual es constante en cada nivel. Los niveles están datos por la variable Rail.

2.2 Dataframe de vías de ferrocarril – Efectos Mixtos

```
> summary(r2.lme)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: travel ~ 1 + (1 | Rail)
   Data: Rail

     AIC      BIC   logLik deviance df.resid
   134.6    137.2    -64.3    128.6       15

Scaled residuals:
     Min       1Q   Median       3Q      Max
-1.61098 -0.28887  0.03454  0.21373  1.62222

Random effects:
 Groups   Name        Variance Std.Dev.
 Rail     (Intercept) 511.86   22.624
 Residual              16.17    4.021
Number of obs: 18, groups:  Rail, 6

Fixed effects:
            Estimate Std. Error t value
(Intercept)   66.500      9.285   7.162
```

2.2 Dataframe de vías de ferrocarril – Efectos Mixtos

```
> qqnorm(resid(r2.lme), main="")
```



2.2 Dataframe de vías de ferrocarril – Efectos Mixtos

```
> plot(fitted(r2.lme), resid(r2.lme), xlab="Fitted",
  ylab="Residuals")
```