

Solución taller sobre capítulo 4 parte 4

Santiago Franco

10/11/2022

```
library(HSAUR2)
```

```
## Warning: package 'HSAUR2' was built under R version 4.1.3
```

```
## Loading required package: tools
```

```
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
```

```
library(lme4)
```

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

```
library(ggplot2)
```

1. Consulte la ayuda de la base de datos para conocer los detalles de las variables recolectadas.

```
datos <- toenail  
help(toenail)
```

```
## starting httpd help server ... done
```

A data frame with 1908 observations on the following 5 variables.

patientID a unique identifier for each patient in the trial.

outcome degree of separation of the nail plate from the nail bed (onycholysis).

treatment a factor with levels itraconazole and terbinafine.

time the time in month when the visit actually took place.

visit number of visit attended.

Details: De Backer et al. (1998) describe a clinical trial to compare two competing oral antifungal treatments for toenail infection (dermatophyte onychomycosis). A total of 378 patients were randomly allocated into two treatment groups, one group receiving 250mg per day of terbinafine and the other group 200mg per day of itraconazole. Patients were evaluated at seven visits, intended to be at weeks 0, 4, 8, 12, 24, 36, and 48 for the degree of separation of the nail plate from the nail bed (onycholysis) dichotomized into moderate or severe and none or mild. But patients did not always arrive exactly at the scheduled time and the exact time in months that they did attend was recorded. The data is not balanced since not all patients attended for all seven planned visits.

Importante: Son datos asociados al tratamiento de la onychomycosis que se tomaron de 378 pacientes divididos en dos grupos de tratamiento, uno recibiendo 250mg diarios de terbinafine y el otro recibiendo 200mg diarios de itraconazole.

```
head(datos)
```

```
##   patientID      outcome treatment      time visit
## 1         1 moderate or severe terbinafine 0.0000000    1
## 2         1 moderate or severe terbinafine 0.8571429    2
## 3         1 moderate or severe terbinafine 3.5357140    3
## 4         1      none or mild terbinafine 4.5357140    4
## 5         1      none or mild terbinafine 7.5357140    5
## 6         1      none or mild terbinafine 10.0357100   6
```

2. ¿Cuáles pacientes solo asistieron UNA vez al control?

```
conteo <- datos %>% count(patientID)
conteo[conteo['n']==1, ]
```

```
##   patientID n
## 34         45 1
## 35         48 1
## 48         63 1
## 77         99 1
## 291        377 1
```

Se tienen datos de 5 pacientes que sólo asistieron una vez al control.

3. ¿Cuál fue el mayor número de visitas de un paciente?

```
max(conteo['n'])
```

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

El máximo número de visitas posible de un paciente es 7, ya que en el contexto de los datos se tiene que hubieron 7 controles.

4. ¿Cuántos pacientes recibieron terbinafine? ¿Cuántos pacientes recibieron itraconazole?

```
unicos <- unique(datos[,c('patientID', 'treatment')])  
unicos %>% count(treatment)
```

```
##      treatment      n  
## 1 itraconazole 146  
## 2 terbinafine 148
```

Se tienen 146 pacientes con itraconazole y 148 pacientes con terbinafine.

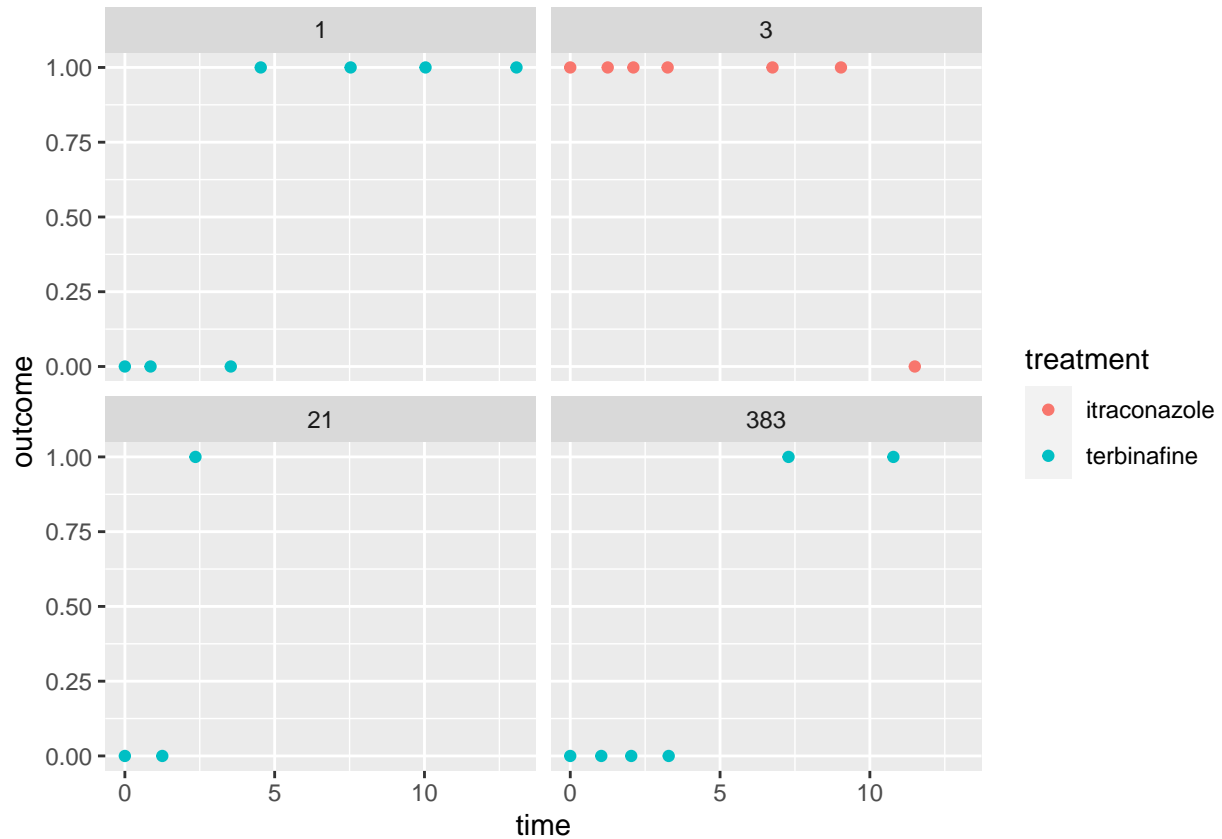
5. Haga un dibujo igualito al mostrado abajo para monitorear la evolución del estado de la enfermedad Y en función del tiempo para los pacientes 1, 3, 21 y 383.

Se realiza un arreglo de los datos de cara a su graficación y modelamiento:

```
datos$outcome = ifelse(datos$outcome=="none or mild", 1, 0)
```

```
pacientes <- c(1,3,21, 383)
```

```
datos_pacientes <- datos %>% filter(patientID %in% pacientes)  
ggplot(datos_pacientes, aes(x=time, y=outcome, color=treatment)) + geom_point() + facet_wrap(~ patientID)
```



```
datos$treatment <- factor(toenail$treatment, levels=c("terbinafine", "itraconazole"))
```

6. Ajuste el siguiente modelo en el objeto mod1.

```
mod1 <- glmer(outcome ~ time + treatment + (1|patientID), data=datos, family=binomial(link="logit"), na.rm=TRUE)
```

7. Ajuste el siguiente modelo en el objeto mod2.

```
mod2 <- glmer(outcome ~ time + treatment+ treatment*time+ (1|patientID), data=datos, family=binomial(link="logit"), na.rm=TRUE)
```

8. Ajuste el siguiente modelo en el objeto mod3.

```
mod3 <- glmer(outcome ~ time + treatment+ treatment*time+ (time|patientID), data=datos, family=binomial(link="logit"), na.rm=TRUE)
```

8. Use una prueba de razón de verosimilitud sencilla para determinar cuál de los modelos es mejor. Use el siguiente código.

```
anova(mod1,mod2, mod3)
```

```
## Data: datos
## Models:
## mod1: outcome ~ time + treatment + (1 | patientID)
## mod2: outcome ~ time + treatment + treatment * time + (1 | patientID)
## mod3: outcome ~ time + treatment + treatment * time + (time | patientID)
##      npar      AIC      BIC logLik deviance   Chisq Df Pr(>Chisq)
## mod1     4 1267.65 1289.9 -629.82  1259.65
## mod2     5 1265.63 1293.4 -627.82  1255.63   4.0163  1    0.04506 *
## mod3     7  996.63 1035.5 -491.32   982.63 273.0001  2    < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

El mejor modelo es mod3.

9. Explore la ayuda de la función predict.merMod (o predict). ¿Para qué sirve el argumento type?

```
head(predict(mod1, type = "response"))
```

```
##      1      2      3      4      5      6
## 0.1531097 0.2115974 0.4798467 0.5939419 0.8535990 0.9486057
```

```
head(predict(mod1, type = "link"))
```

```
##      1      2      3      4      5      6
## -1.71041659 -1.31532326 -0.08065683  0.38028537  1.76311195  2.91546558
```

Para obtener la respuesta a los datos utilizados o datos nuevos en forma de la variable respuesta o la función de enlace.

10. Escriba el modelo ajustado que resultó ser el mejor modelo.

```
summary(mod3)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: outcome ~ time + treatment + treatment * time + (time | patientID)
## Data: datos
##
##      AIC      BIC    logLik deviance df.resid
##   996.6   1035.5   -491.3    982.6     1901
##
## Scaled residuals:
```

```
##      Min      1Q   Median      3Q      Max
## -2.36512  0.00000  0.00242  0.00661  1.95071
##
## Random effects:
##   Groups   Name      Variance Std.Dev. Corr
## patientID (Intercept) 549.00   23.431
##          time          14.72    3.836  -0.87
## Number of obs: 1908, groups: patientID, 294
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      9.35034    0.85242  10.969  <2e-16 ***
## time              0.75546    0.35520   2.127   0.0334 *
## treatmentitraconazole 0.02474    1.02729   0.024   0.9808
## time:treatmentitraconazole -0.45703    0.34997  -1.306   0.1916
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) time   trtmnt
## time          -0.247
## trtmnttrcnz  -0.593  0.284
## tm:trtmnttr  0.258 -0.795 -0.493
```

```
VarCorr(mod3)
```

```
##   Groups   Name      Std.Dev. Corr
## patientID (Intercept) 23.4307
##          time          3.8363  -0.873
```

```
-0.873*23.4307*3.8363
```

```
## [1] -78.47152
```

$$y_{ij} \sim \text{Bernoulli}(p_{ij}) \quad (1)$$

$$\text{logit}(p_{ij}) = 9.35034 + 0.75546 \text{time}_{ij} + 0.02474 \text{treatment}_{itraconazolei} - 0.45703 \text{time}_{ij} \text{treatment}_{itraconazolei} + b_{0i} + b_{1i} \text{time}_{ij} \quad (2)$$

$$\begin{pmatrix} b_0 \\ b_1 \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 549.00 & -78.47152 \\ -78.47152 & 14.72 \end{pmatrix} \right) \quad (3)$$

$$(4)$$

11. Usando el modelo anterior, escriba el modelo ajustado para el paciente #1.

Se usa la función `ranef` para buscar los efectos aleatorios:

```
efectos <- ranef(mod3)[[1]]
head(efectos)
```

```
##      (Intercept)      time
## 1  -24.631759   3.1001113
## 2   -9.252870  -0.1025331
## 3   12.970290  -2.4729590
## 4  -10.805089   2.3814822
## 6  -18.348775   2.9263701
## 7    4.276357  -2.8925563
```

$$y_{ij} \sim \text{Bernoulli}(p_{ij}) \quad (5)$$

$$\text{logit}(p_{ij}) = 9.35034 + 0.75546 \text{time}_{ij} + 0.02474 \text{treatment}_{\text{itraconazole}i} - 0.45703 \text{time}_{ij} \text{treatment}_{\text{itraconazole}i} - 24.63175913 \quad (6)$$

12. Vuelva a construir la figura anterior con los pacientes 1, 3, 21, 383 pero incluyendo la curva ajustada con cada modelo. Abajo se muestra una de las tres figuras.

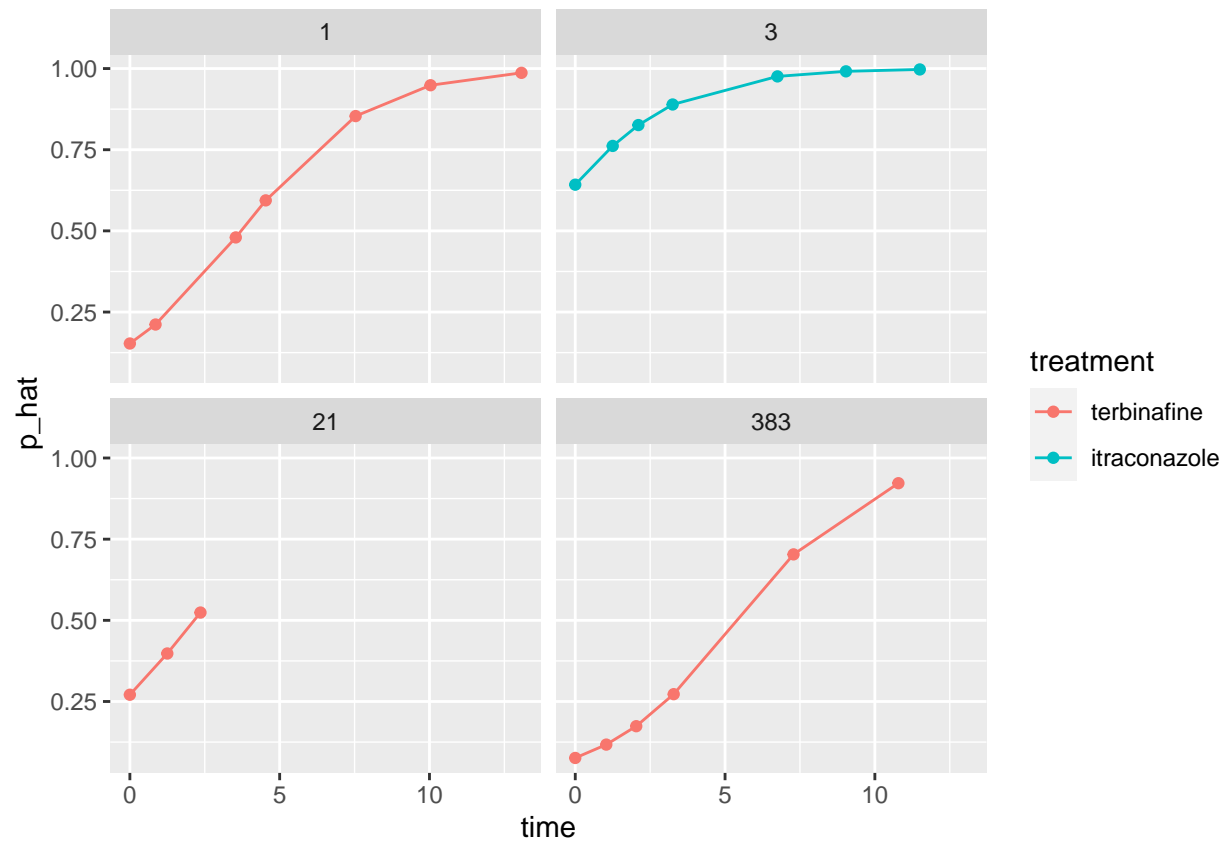
¿A cuál modelo pertenece la figura? ¿mod1, mod2 o mod3?

```
datos['p_hat'] <- predict(mod1, type='response')
```

```
pacientes <- c(1,3,21, 383)
```

```
datos_pacientes <- datos %>% filter(patientID %in% pacientes)
```

```
ggplot(datos_pacientes, aes(x=time, y=p_hat, color=treatment)) + geom_point() + facet_wrap(~ patientID)
```

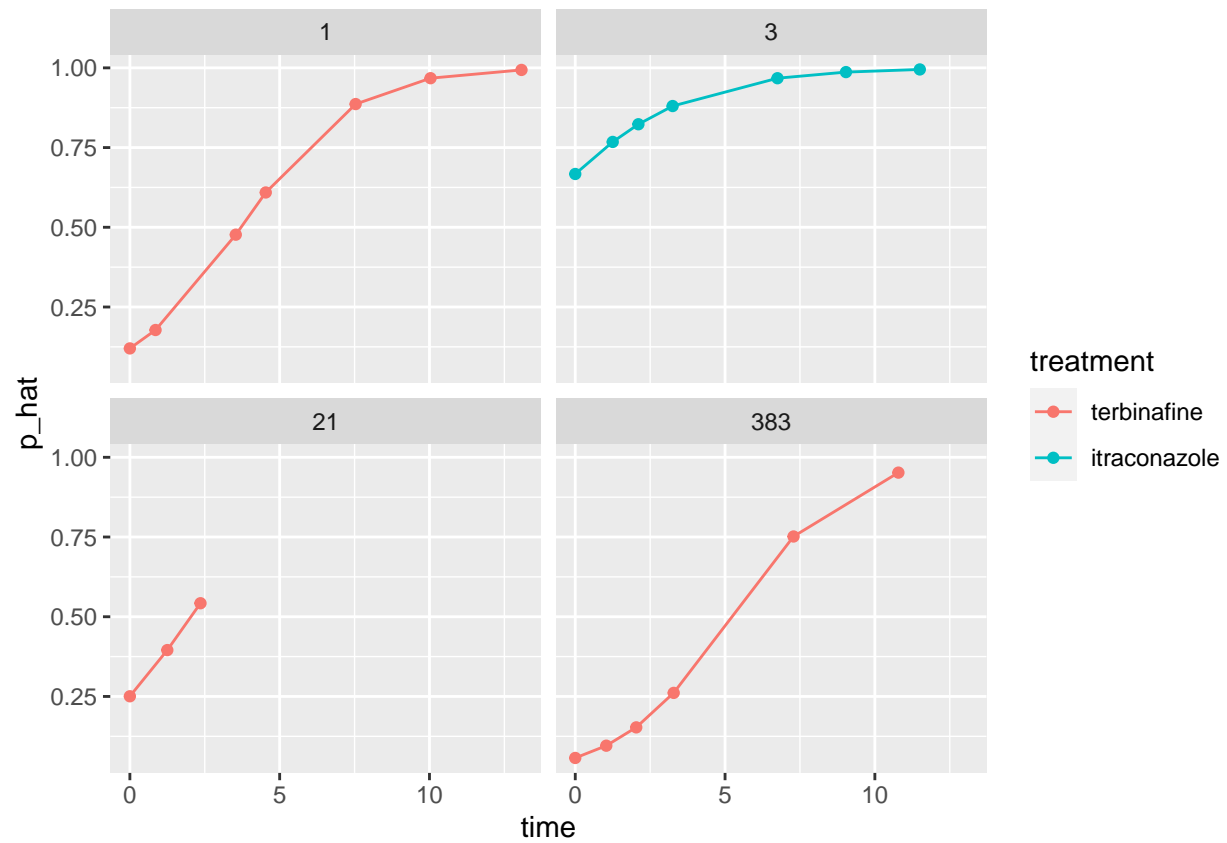


```
datos['p_hat'] <- predict(mod2, type='response')
```

```
pacientes <- c(1,3,21, 383)
```

```
datos_pacientes <- datos %>% filter(patientID %in% pacientes)
```

```
ggplot(datos_pacientes, aes(x=time, y=p_hat, color=treatment)) + geom_point() + facet_wrap(~ patientID)
```

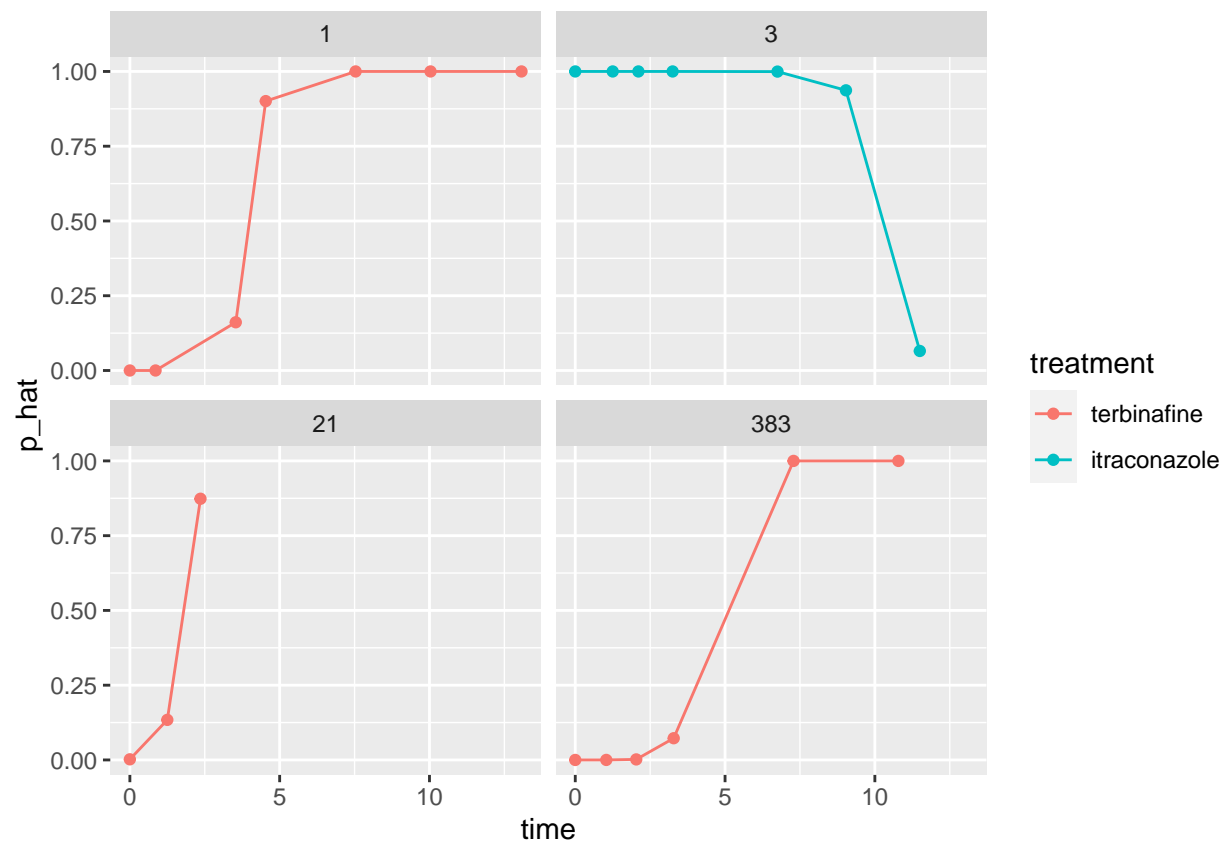



```
datos['p_hat'] <- predict(mod3, type='response')
```

```
pacientes <- c(1,3,21, 383)
```

```
datos_pacientes <- datos %>% filter(patientID %in% pacientes)
```

```
ggplot(datos_pacientes, aes(x=time, y=p_hat, color=treatment)) + geom_point() + facet_wrap(~ patientID)
```



Se observa que la figura pertenece al modelo 3.