

<https://www.datos.gov.co/Inclusi-n-Social-y-Reconciliaci-n/Beneficiarios-M-s-Familias-en-Acci-n/xfif-my2>

Modelo condicional:

Modelo con condiciones sobre una variables

Ejemplo:

$\log(\text{uij}) = \beta_0 + \beta_1 + \beta_2 \text{Niveleducativo}$

```
library(HSAUR2)
```

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

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

```
library(lme4)
```

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

```
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(ggplot2)
```

```
data("epilepsy", package = "HSAUR2")
epilepsy$x <- as.numeric(epilepsy$period)
epilepsy$y <- epilepsy$seizure.rate
```

A data frame with 236 observations on the following 6 variables.

treatment the treatment group, a factor with levels placebo and Progabide.

base the number of seizures before the trial.

age the age of the patient.

seizure.rate the number of seizures (response variable).

period treatment period, an ordered factor with levels 1 to 4.

subject the patient ID, a factor with levels 1 to 59.

Details In this clinical trial, 59 patients suffering from epilepsy were randomized to groups receiving either the anti-epileptic drug Progabide or a placebo in addition to standard chemotherapy. The numbers of seizures suffered in each of four, two-week periods were recorded for each patient along with a baseline seizure count for the 8 weeks prior to being randomized to treatment and age. The main question of interest is whether taking progabide reduced the number of epileptic seizures compared with placebo.

2. ¿Cuántos pacientes tiene la base de datos?

```
valores_ordenados <- epilepsy %>% count(subject)
valores_ordenados[order(valores_ordenados$n), ]
```

```
##      subject n
## 1          1 4
## 2          2 4
## 3          3 4
## 4          4 4
## 5          5 4
## 6          6 4
## 7          7 4
## 8          8 4
## 9          9 4
## 10         10 4
## 11         11 4
## 12         12 4
## 13         13 4
## 14         14 4
## 15         15 4
## 16         16 4
## 17         17 4
## 18         18 4
## 19         19 4
## 20         20 4
## 21         21 4
## 22         22 4
## 23         23 4
## 24         24 4
## 25         25 4
## 26         26 4
## 27         27 4
## 28         28 4
## 29         29 4
## 30         30 4
## 31         31 4
## 32         32 4
## 33         33 4
## 34         34 4
## 35         35 4
## 36         36 4
## 37         37 4
## 38         38 4
## 39         39 4
## 40         40 4
## 41         41 4
## 42         42 4
## 43         43 4
## 44         44 4
## 45         45 4
## 46         46 4
## 47         47 4
```

```
## 48      48 4
## 49      49 4
## 50      50 4
## 51      51 4
## 52      52 4
## 53      53 4
## 54      54 4
## 55      55 4
## 56      56 4
## 57      57 4
## 58      58 4
## 59      59 4
```

Se tienen 59 pacientes

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

Todos tienen 4 seguimientos

4.¿Cuántos pacientes recibieron Progabide? ¿Cuántos pacientes recibieron el placebo?

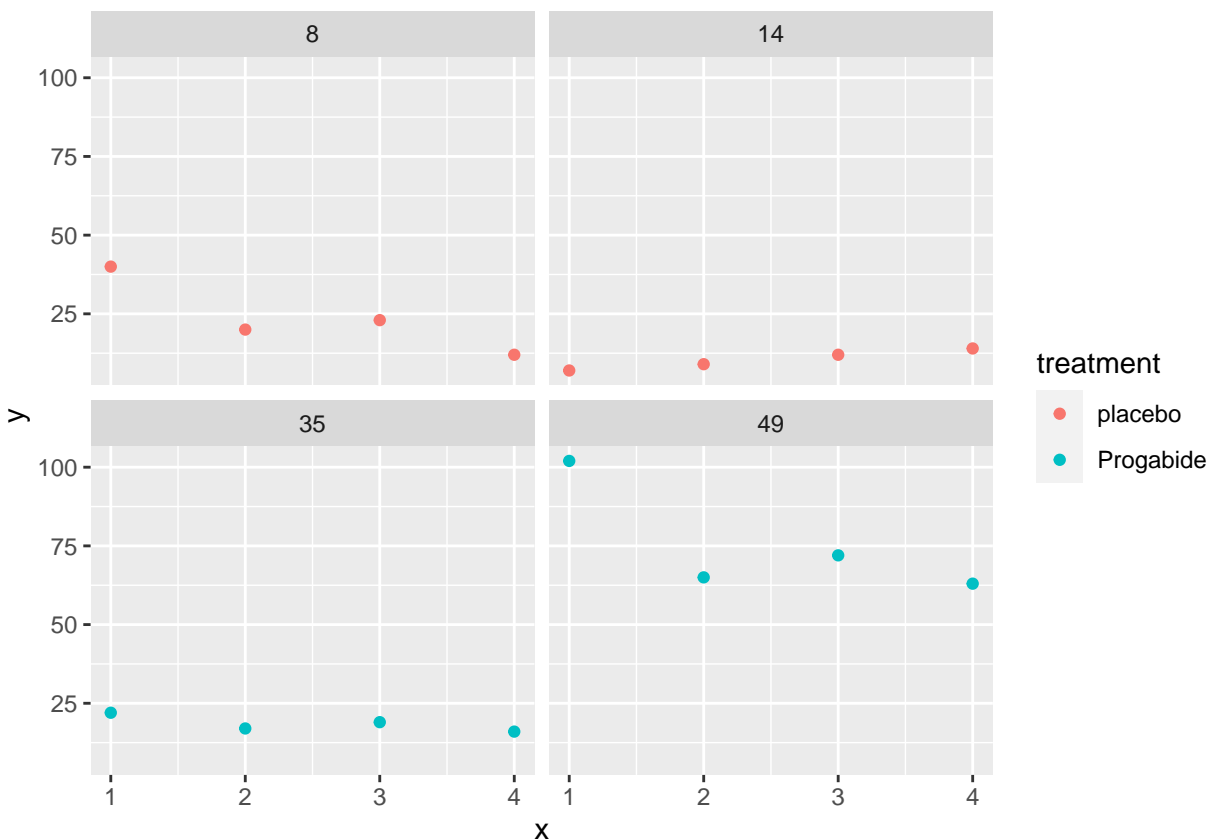
```
epilepsy[,c('subject','treatment')] |> unique() |> count(treatment)
```

```
##   treatment  n
## 1  placebo 28
## 2  Progabide 31
```

Se tienen 28 pacientes con placebo y 31 con progabide

5.Haga un dibujo igualito al mostrado abajo para monitorear la evolución de la variable Y en función del periodo X para los pacientes 8, 14, 35, 49. Use colores para diferenciar los pacientes que recibieron Progabide y el placebo.

```
pacientes <- c(8,14,35,49)
datos_grafica <- epilepsy |> filter(subject %in% pacientes)
ggplot(data=datos_grafica, aes(x=x, y=y, color=treatment)) + geom_point() + facet_wrap(~subject)
```



6. Ajuste el siguiente modelo en el objeto mod1.

$$y_{ij}|b_0 \sim \text{Poisson}(\lambda_{ij}) \quad (1)$$

$$\log(\lambda_{ij}) = \beta_0 + \beta_1 x_{ij} + \beta_2 \text{treatment}_{\text{Progabide}i} + b_{0i} \quad (2)$$

$$b_0 \sim N(0, \sigma_{b_0}^2), \quad (3)$$

donde i corresponde al paciente y j a la visita. Note que el tratamiento de referencia es placebo. Use nAGQ=1 para la aproximación por cuadratura de la integral interna.

```
mod1 <- glmer(y~x + treatment + (1|subject), data=epilepsy, nAGQ = 1, family=poisson(link='log'))
summary(mod1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: poisson ( log )
## Formula: y ~ x + treatment + (1 | subject)
## Data: epilepsy
##
##      AIC      BIC    logLik deviance df.resid
##  1400.8   1414.7   -696.4   1392.8     232
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -3.4124 -0.7893 -0.1553  0.5511  7.1627
##
## Random effects:
##   Groups   Name      Variance Std.Dev.
##  subject (Intercept) 0.8751   0.9355
## Number of obs: 236, groups:  subject, 59
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.91718    0.18855  10.168 < 2e-16 ***
## x                -0.05872    0.02015  -2.914  0.00357 **
## treatmentProgabide -0.28898    0.25271  -1.144  0.25283
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) x
## x              -0.259
## trtmntPrghd -0.694  0.000
```

7. Ajuste el siguiente modelo en el objeto mod2.

$$y_{ij}|b_0 \sim \text{Poisson}(\lambda_{ij}) \quad (4)$$

$$\log(\lambda_{ij}) = \beta_0 + \beta_1 x_{ij} + \beta_2 \text{treatment}_{Progabidei} + \beta_3 x_{ij} \text{treatment}_{Progabidei} + b_{0i} \quad (5)$$

$$b_0 \sim N(0, \sigma_{b_0}^2), \quad (6)$$

```
mod2 <- glmer(y~x + treatment + x*treatment + (1|subject), data=epilepsy, nAGQ = 1, family=poisson(link="log"))
summary(mod2)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: poisson ( log )
## Formula: y ~ x + treatment + x * treatment + (1 | subject)
## Data: epilepsy
##
##      AIC      BIC    logLik deviance df.resid
##  1402.2   1419.5   -696.1   1392.2     231
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.4562 -0.7899 -0.1697  0.5720  7.0782
##
## Random effects:
##   Groups   Name      Variance Std.Dev.
##  subject (Intercept) 0.8751   0.9355
## Number of obs: 236, groups:  subject, 59
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.87839    0.19513   9.626 <2e-16 ***
## x                -0.04281    0.02866  -1.494   0.135
```

```
## treatmentProgabide -0.21259 0.27099 -0.785 0.433
## x:treatmentProgabide -0.03147 0.04031 -0.781 0.435
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) x          trtmnP
## x          -0.359
## trtmntPrghd -0.719 0.259
## x:trtmntPrg 0.255 -0.711 -0.361
```

8. Ajuste el siguiente modelo en el objeto mod3.

$$y_{ij}|b_0, b_1 \sim \text{Poisson}(\lambda_{ij}) \quad (7)$$

$$\log(\lambda_{ij}) = \beta_0 + \beta_1 x_{ij} + \beta_2 \text{treatment}_{\text{Progabide}i} + \beta_3 x_{ij} \text{treatment}_{\text{Progabide}i} + b_{0i} + b_{1i} x_{ij} \quad (8)$$

$$\begin{pmatrix} b_0 \\ b_1 \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{b0}^2 & \sigma_{b01} \\ \sigma_{b01} & \sigma_{b1}^2 \end{pmatrix} \right) \quad (9)$$

$$(10)$$

```
mod3 <- glmer(y~x + treatment + x*treatment + (x|subject), data=epilepsy, nAGQ = 1, family=poisson(link=log))
summary(mod3)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: poisson ( log )
## Formula: y ~ x + treatment + x * treatment + (x | subject)
## Data: epilepsy
##
##          AIC          BIC    logLik deviance df.resid
##   1387.1    1411.4    -686.6   1373.1      229
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.1004 -0.7023 -0.1339  0.5576  6.2948
##
## Random effects:
## Groups Name          Variance Std.Dev. Corr
## subject (Intercept) 1.02745  1.0136
## x                  0.02095  0.1447  -0.39
## Number of obs: 236, groups: subject, 59
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.86491    0.21584   8.640 <2e-16 ***
## x              -0.04349    0.04540  -0.958  0.338
## treatmentProgabide -0.25098    0.29777  -0.843  0.399
## x:treatmentProgabide -0.01383    0.06209  -0.223  0.824
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
```

```
##           (Intr) x      trtmnP
## x          -0.538
## trtmntPrghd -0.700  0.344
## x:trtmntPrg  0.346 -0.641 -0.530
```

9. 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: epilepsy
## Models:
## mod1: y ~ x + treatment + (1 | subject)
## mod2: y ~ x + treatment + x * treatment + (1 | subject)
## mod3: y ~ x + treatment + x * treatment + (x | subject)
##      npar    AIC    BIC logLik deviance   Chisq Df Pr(>Chisq)
## mod1     4 1400.8 1414.7 -696.40   1392.8
## mod2     5 1402.2 1419.5 -696.10   1392.2  0.6018  1    0.4379
## mod3     7 1387.1 1411.4 -686.56   1373.1 19.0897  2 7.157e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

¿Cuál de los dos modelos es mejor?

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

```
predicciones_mod1 <- predict(mod1, type='response')
predicciones_mod2 <- predict(mod2, type='response')
predicciones_mod3 <- predict(mod3, type='response')
```

10. Para cada uno de los modelos ajustados calcule el coeficiente de correlación de Pearson entre

¿Cuál modelo tiene el mayor ρ ?

```
epilepsy$predicciones1 <- predicciones_mod1
epilepsy$predicciones2 <- predicciones_mod2
epilepsy$predicciones3 <- predicciones_mod3
```

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

```
summary(mod3)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: poisson ( log )
## Formula: y ~ x + treatment + x * treatment + (x | subject)
## Data: epilepsy
##
##      AIC      BIC   logLik deviance df.resid
##  1387.1   1411.4   -686.6   1373.1     229
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.1004 -0.7023 -0.1339  0.5576  6.2948
##
## Random effects:
## Groups Name Variance Std.Dev. Corr
## subject (Intercept) 1.02745  1.0136
## x 0.02095  0.1447 -0.39
## Number of obs: 236, groups: subject, 59
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.86491    0.21584   8.640 <2e-16 ***
## x             -0.04349    0.04540  -0.958   0.338
## treatmentProgabide -0.25098    0.29777  -0.843   0.399
## x:treatmentProgabide -0.01383    0.06209  -0.223   0.824
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) x      trtmnP
## x             -0.538
## trtmntPrghd -0.700  0.344
## x:trtmntPrg  0.346 -0.641 -0.530
```

$$y_{ij}|b_0, b_1 \sim \text{Poisson}(\lambda_{ij}) \quad (11)$$

$$\log(\lambda_{ij}) = 1.86491 - 0.04349 x_{ij} - 0.25098 \text{treatment}_{Progabidei} - 0.01383 x_{ij} \text{treatment}_{Progabidei} + b_{0i} + b_{1i} x_{ij} \quad (12)$$

$$\begin{pmatrix} b_0 \\ b_1 \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1.02745 & -0.05720049 \\ -0.05720049 & 0.02095 \end{pmatrix} \right) \quad (13)$$

$$(14)$$

12. Usando el modelo anterior, escriba el modelo ajustado para el paciente #8.

```
ranef(mod3)
```

```
## $subject
##      (Intercept)      x
## 1 -0.389423959 -0.031728396
## 2 -0.463612766 -0.001404272
```



```

## 3 -0.809949236 0.052958823
## 4 -0.509014950 -0.009664573
## 5 0.419325490 0.167056179
## 6 -0.290856662 0.095724402
## 7 -0.332201630 -0.112542373
## 8 1.904188026 -0.220740992
## 9 -0.089687923 0.016072013
## 10 0.940109888 -0.260465040
## 11 1.136095201 -0.044615356
## 12 0.514500813 -0.111538872
## 13 -0.309811687 -0.014964078
## 14 0.217632063 0.141960900
## 15 1.205683724 -0.121911516
## 16 -0.064792229 -0.117630034
## 17 -1.452101670 0.117519640
## 18 1.730412558 -0.031873098
## 19 -0.495563774 0.036385178
## 20 -0.719413777 0.146283942
## 21 -0.538855416 0.028915859
## 22 -0.538855416 0.028915859
## 23 -0.741380748 0.082942263
## 24 0.346804597 -0.040626926
## 25 1.301116515 0.195984344
## 26 -1.139282239 -0.005478817
## 27 -0.789338927 0.012727778
## 28 0.806249944 0.001116734
## 29 0.959397631 -0.043276778
## 30 0.547190218 -0.038362245
## 31 -0.802582474 0.004597232
## 32 -0.214465634 -0.022960825
## 33 -0.087576973 0.067680319
## 34 -0.356621996 -0.025313444
## 35 1.478524417 -0.023030196
## 36 0.082352738 0.019130543
## 37 -0.576955314 0.030862976
## 38 0.058146040 0.098498532
## 39 0.621514629 -0.055133226
## 40 -1.076571238 -0.042557028
## 41 -0.983447676 0.030608402
## 42 -0.188064869 -0.063661728
## 43 0.981923267 0.123712060
## 44 0.451093962 -0.028445366
## 45 1.175234355 -0.168619530
## 46 -0.976556040 0.073960931
## 47 0.469342333 0.047114514
## 48 -1.215099816 -0.053021193
## 49 3.035922849 -0.082102132
## 50 -0.290312053 0.008086106
## 51 0.390287961 -0.003339249
## 52 -0.742815909 0.096455793
## 53 1.264387243 0.041691176
## 54 0.006779225 -0.116233254
## 55 -0.163081509 0.009300864
## 56 0.780696092 0.107304276

```

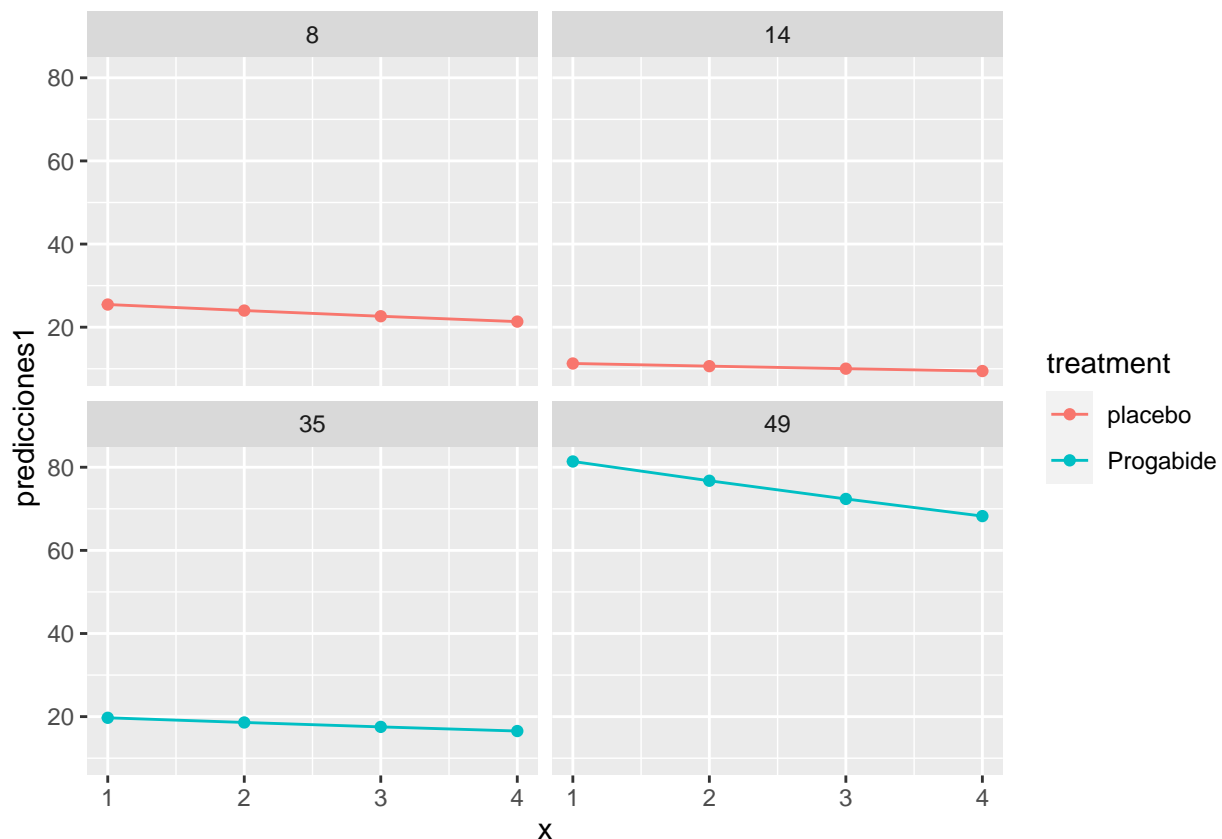
```
## 57 -0.809307448 -0.040061957
## 58 -2.053318464 0.013296077
## 59 -0.576955314 0.030862976
##
## with conditional variances for "subject"
```

$$y_{ij} \sim \text{Poisson}(\lambda_{ij}) \quad (15)$$

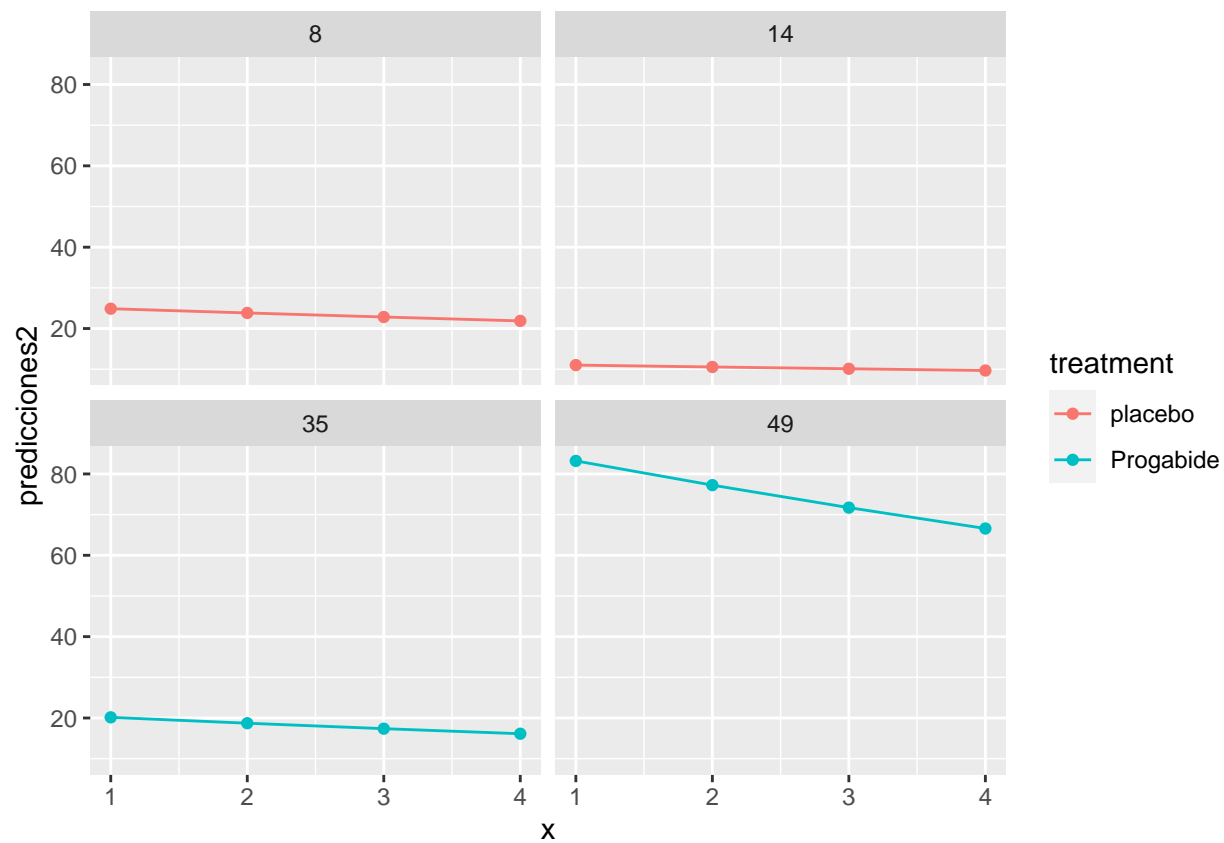
$$\log(\lambda_{8j}) = 1.86491 - 0.04349 x_{8j} - 0.25098 \text{treatment}_{\text{Progabide}i} - 0.01383 x_{8j} \text{treatment}_{\text{Progabide}i} + 1.904188026 - 0.220740 \quad (16)$$

13. Vuelva a construir la figura anterior con los pacientes 8, 14, 35, 49 pero incluyendo la curva ajustada de $\hat{\lambda}$ con cada modelo. Abajo se muestra una de las tres figuras.

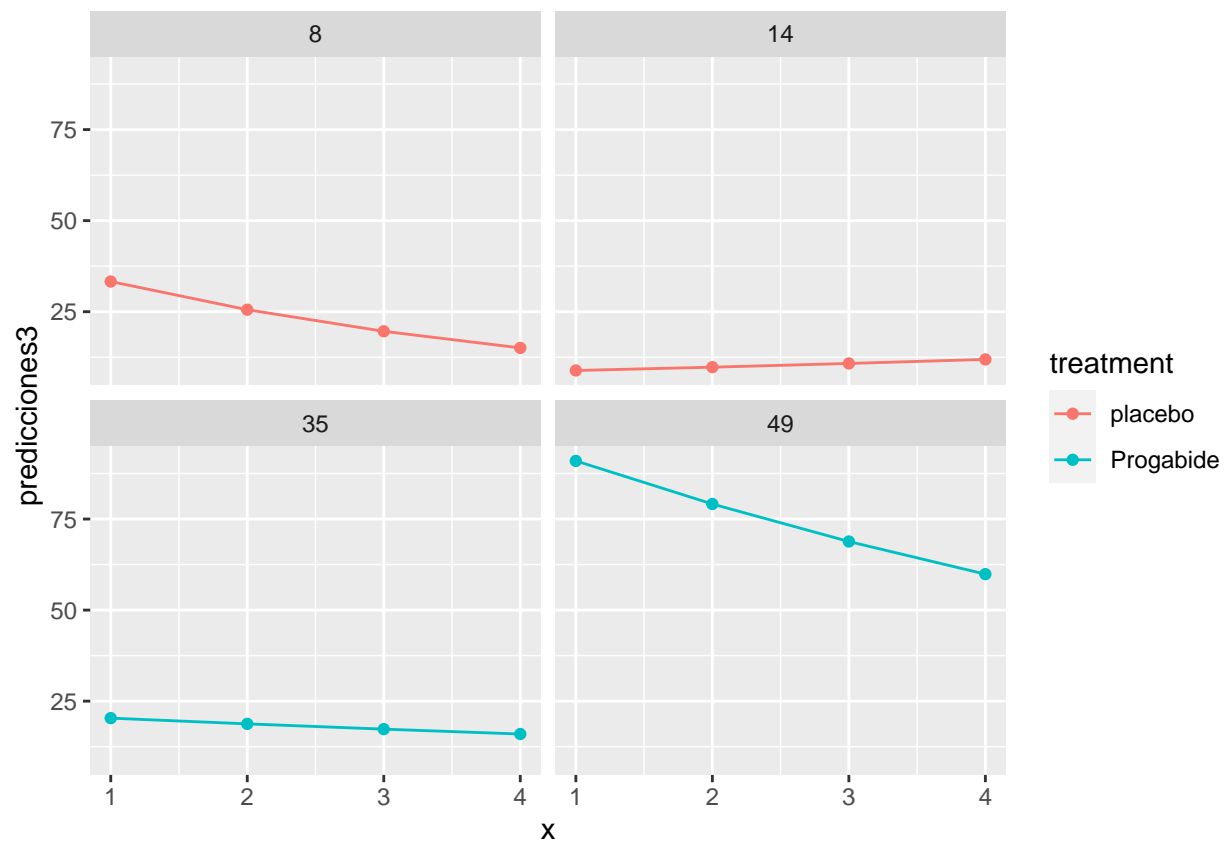
```
pacientes <- c(8,14,35,49)
datos_grafica <- epilepsy |> filter(subject %in% pacientes)
ggplot(data=datos_grafica, aes(x=x, y=predicciones1, color=treatment)) + geom_line() + geom_point() + fa
```



```
pacientes <- c(8,14,35,49)
datos_grafica <- epilepsy |> filter(subject %in% pacientes)
ggplot(data=datos_grafica, aes(x=x, y=predicciones2, color=treatment)) + geom_line() + geom_point() + fa
```



```
pacientes <- c(8,14,35,49)
datos_grafica <- epilepsy |> filter(subject %in% pacientes)
ggplot(data=datos_grafica, aes(x=x, y=predicciones3, color=treatment)) + geom_line() +geom_point() + fa
```



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