Clase 4.0
Análisis

Marcos Rosetti y Luis Pacheco-Cobos

Estadística y Manejo de Datos con R (EMDR) — Virtual

Modelos Lineales Generalizados (GLM)

Modelos Lineales Generalizados

- Combinan variables predictoras discretas y continuas.
- Podemos especifcar la naturaleza de la variable de respuesta mediante una función liga (por ejemplo: Gausiana, Poisson o Binomial).
- Cuando la variable de respuesta es binomial, también se conoce como *Regresión Logística*.

Modelos Lineales Generalizados

- 1 variable continua.

```
?Puromycin
head(Puromycin)
```

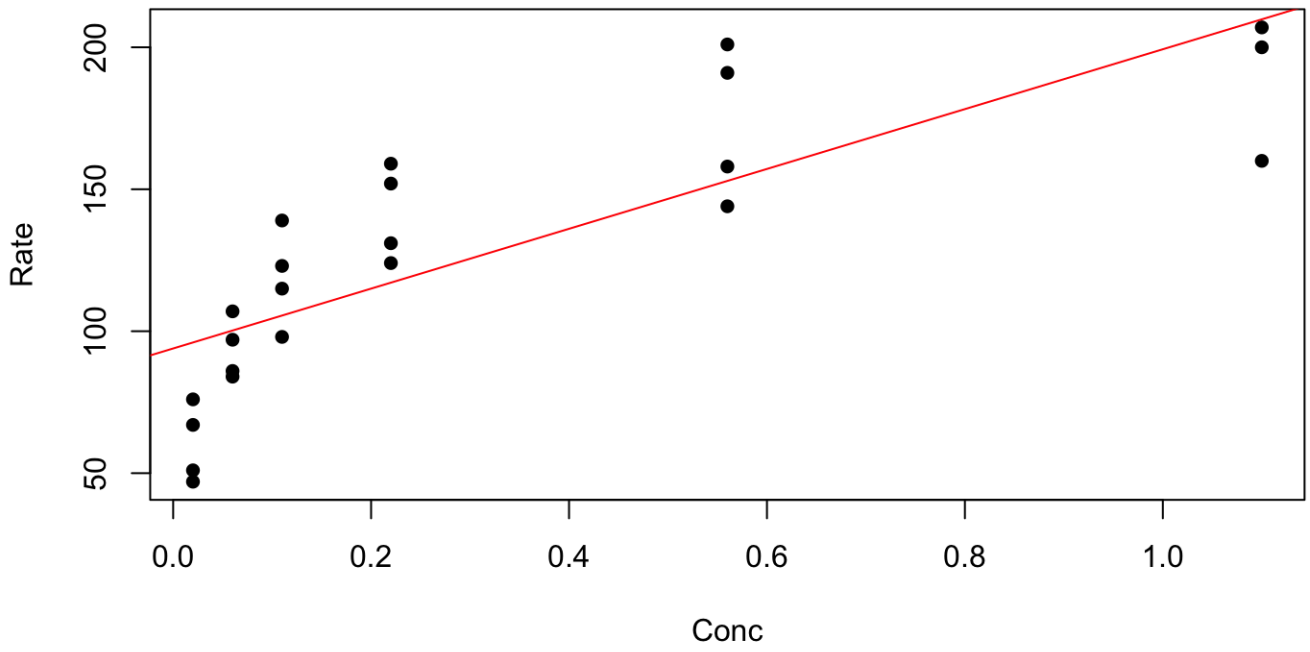
```
pmycin_glm <- glm(rate ~ conc, data = Puromycin)
summary(pmycin_glm)
```

```
##
## Call:
## glm(formula = rate ~ conc, data = Puromycin)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -49.861  -15.247   -2.861   15.686   48.054
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    93.92       8.00   11.74 1.09e-10 ***
## conc          105.40      16.92    6.23 3.53e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 830.4261)
##
##     Null deviance: 49665  on 22  degrees of freedom
## Residual deviance: 17439  on 21  degrees of freedom
## AIC: 223.78
##
## Number of Fisher Scoring iterations: 2
```

Modelos Lineales Generalizados

- 1 variable continua.

```
plot(Puromycin$conc, Puromycin$rate, pch = 16, xlab = "Conc", ylab = "Rate")
abline(lm(rate ~ conc, data = Puromycin), col = "red")
```



Modelos Lineales Generalizados

- 1 una variable discreta.

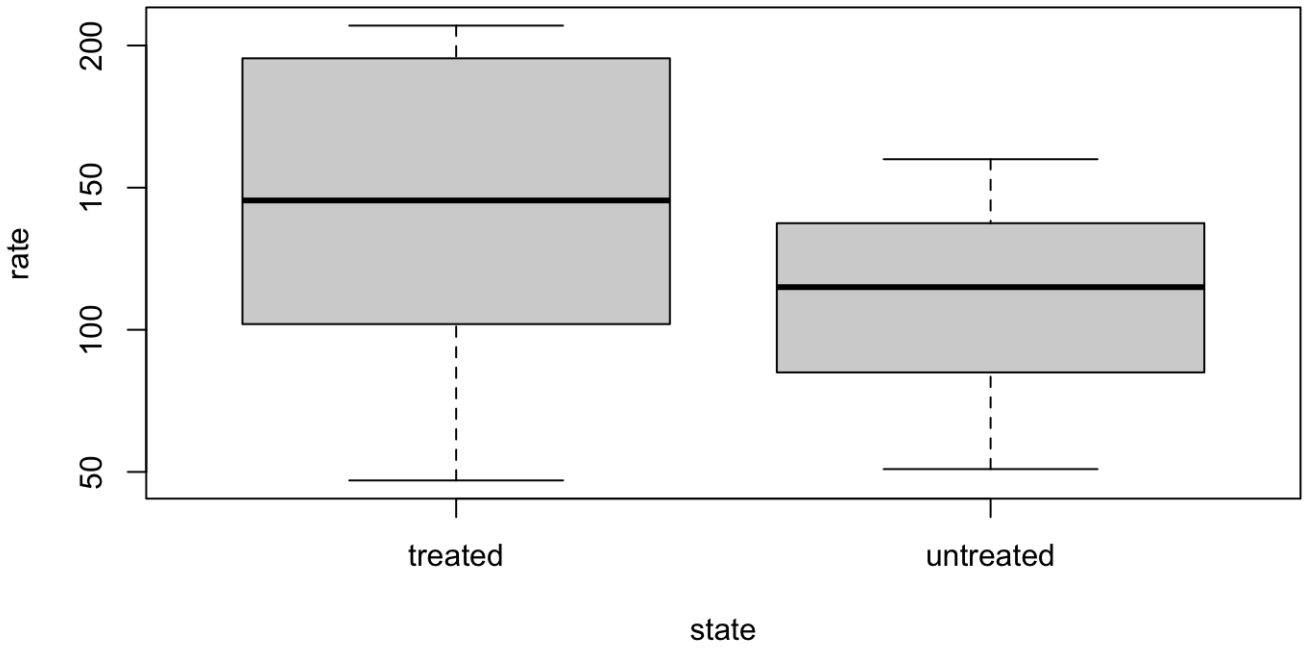
```
pmycin_glm <- glm(rate ~ state, data = Puromycin)
summary(pmycin_glm)
```

```
##
## Call:
## glm(formula = rate ~ state, data = Puromycin)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -94.583  -30.655    4.273   40.273   65.417
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      141.58      13.24  10.690 5.94e-10 ***
## stateuntreated   -30.86      19.15  -1.611    0.122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 2104.814)
##
##     Null deviance: 49665  on 22  degrees of freedom
## Residual deviance: 44201  on 21  degrees of freedom
## AIC: 245.17
##
## Number of Fisher Scoring iterations: 2
```

Modelos Lineales Generalizados

- 1 una variable discreta.

```
boxplot(rate ~ state, data = Puromycin)
```



Modelos Lineales Generalizados

- 2 variables e interacción.

```
pmycin_glm <- glm(conc ~ rate*state, data = Puromycin)
summary(pmycin_glm)
```

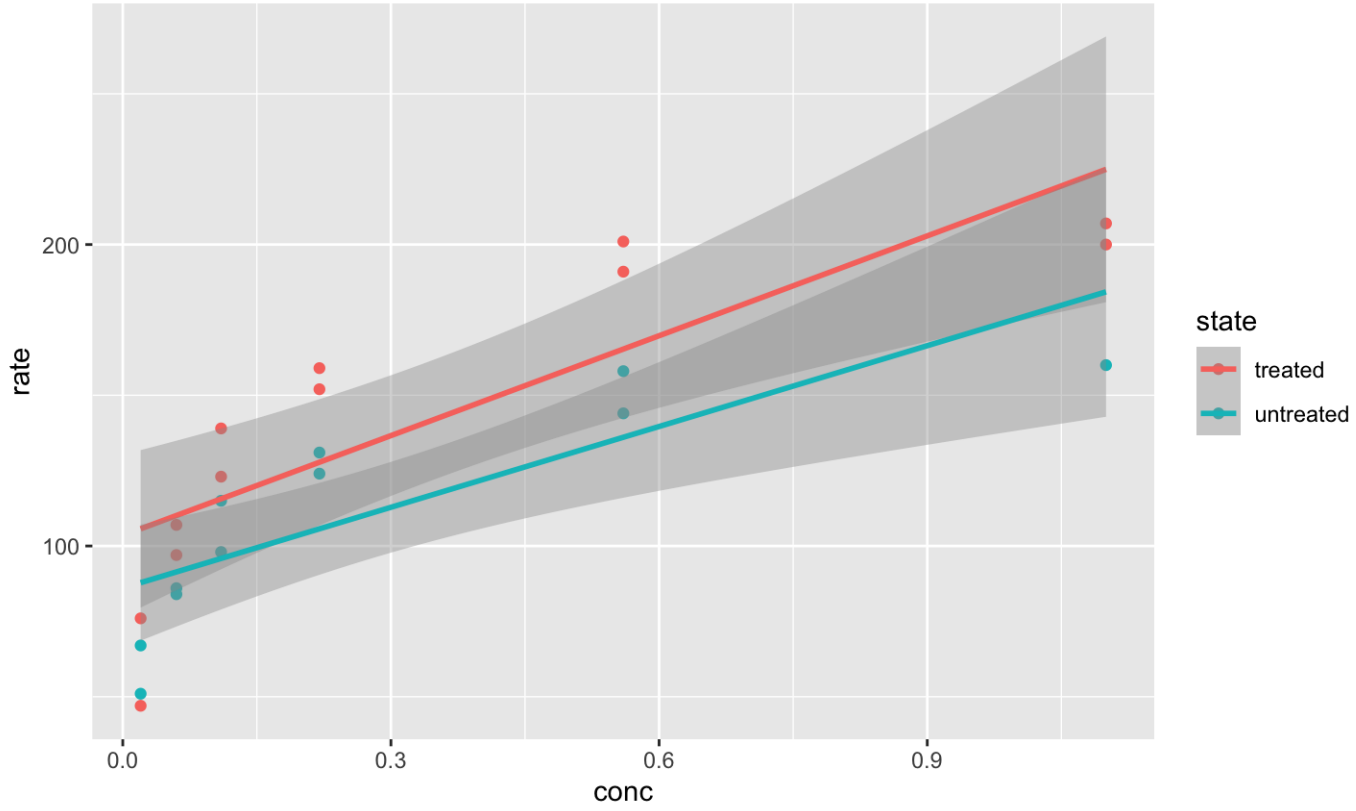
```
##
## Call:
## glm(formula = conc ~ rate * state, data = Puromycin)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.23393  -0.15653  -0.06870   0.07926   0.45212
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         -0.540523   0.187185  -2.888  0.00943 **
## rate                 0.006254   0.001245   5.025 7.52e-05 ***
## stateuntreated      -0.017989   0.288642  -0.062  0.95096
## rate:stateuntreated  0.001285   0.002265   0.567  0.57706
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.04780568)
##
##     Null deviance: 2.90099  on 22  degrees of freedom
## Residual deviance: 0.90831  on 19  degrees of freedom
## AIC: 0.94285
##
## Number of Fisher Scoring iterations: 2
```

Modelos Lineales Generalizados

- 2 variables e interacción.

```
ggplot(Puromycin, aes(x=conc, y=rate, color=state)) + geom_point() + geom_smooth(method='lm')
```

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



Modelos Lineales Generalizados

- Función liga.

```
?infert
```

```
head(infert)
```

```
##   education age parity induced case spontaneous stratum pooled.stratum
## 1    0-5yrs  26      6       1    1           2       1               3
## 2    0-5yrs  42      1       1    1           0       2               1
## 3    0-5yrs  39      6       2    1           0       3               4
## 4    0-5yrs  34      4       2    1           0       4               2
## 5   6-11yrs  35      3       1    1           1       5              32
## 6   6-11yrs  36      4       2    1           1       6              36
```

```
unique(infert$case)
```

```
## [1] 1 0
```

Modelos Lineales Generalizados

- Función liga.

```
inf.bn <- glm(case ~ age + parity + education + spontaneous,
    data = infert, family = binomial())
summary(inf.bn)
```

```
##
## Call:
## glm(formula = case ~ age + parity + education + spontaneous,
##     family = binomial(), data = infert)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7219  -0.7698  -0.6353   0.9825   2.1823
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -0.63833    1.33256  -0.479   0.6319
## age               0.01553    0.02920   0.532   0.5947
## parity           -0.30516    0.13729  -2.223   0.0262 *
## education6-11yrs -0.69194    0.76038  -0.910   0.3628
## education12+ yrs -0.84472    0.79294  -1.065   0.2867
## spontaneous       1.27783    0.22305   5.729 1.01e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 316.17  on 247  degrees of freedom
## Residual deviance: 278.26  on 242  degrees of freedom
## AIC: 290.26
```

Modelos Lineales Generalizados

- Función liga.

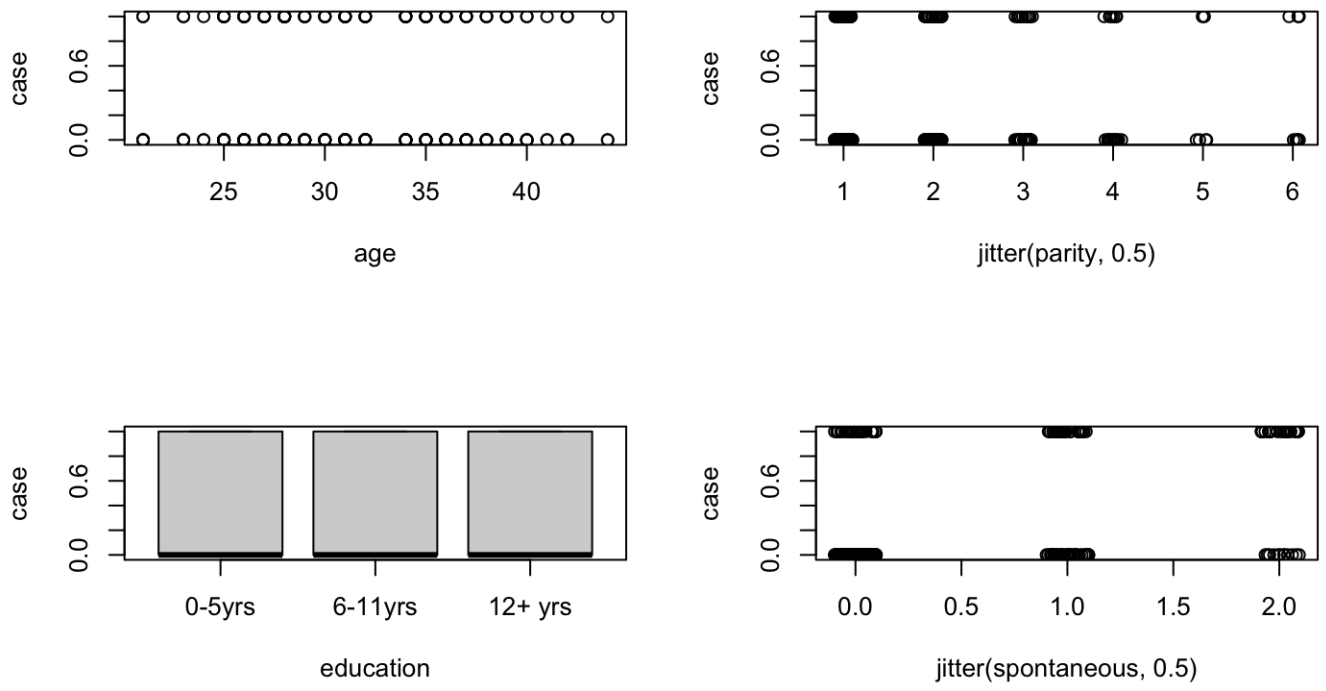
```
inf.no <- glm(case ~ age + parity + education + spontaneous,
    data = infert)
summary(inf.no)
```

```
##
## Call:
## glm(formula = case ~ age + parity + education + spontaneous,
##     data = infert)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.7752  -0.2685  -0.1795   0.3808   0.9700
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       0.371789   0.258916   1.436   0.1523
## age               0.002873   0.005664   0.507   0.6124
## parity           -0.059536   0.026263  -2.267   0.0243 *
## education6-11yrs -0.155096   0.146016  -1.062   0.2892
## education12+ yrs -0.181238   0.151961  -1.193   0.2342
## spontaneous       0.271910   0.041317   6.581 2.87e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.1935233)
##
##     Null deviance: 55.222  on 247  degrees of freedom
## Residual deviance: 46.833  on 242  degrees of freedom
## AIC: 304.42
```

Modelos Lineales Generalizados

- Comparación binomial.

```
par(mfrow=c(2,2))
with(infert, plot(case ~ age + jitter(parity,.5) + education + jitter(spontaneous,.5)))
```



Modelos Lineales Generalizados

- Comparación *post hoc*.

```
library(multcomp)
```

```
summary(glht(inf.no, mcp(education="Tukey")))
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: glm(formula = case ~ age + parity + education + spontaneous,
##     data = infert)
##
## Linear Hypotheses:
##                         Estimate Std. Error z value Pr(>|z|)
## 6-11yrs - 0-5yrs == 0  -0.15510    0.14602  -1.062    0.519
## 12+ yrs - 0-5yrs == 0  -0.18124    0.15196  -1.193    0.438
## 12+ yrs - 6-11yrs == 0 -0.02614    0.06043  -0.433    0.896
## (Adjusted p values reported -- single-step method)
```

Licencia CC BY