6 Regresión generalizada I

Métodos empíricos 2

17/05/2022

Hoy

- Caso de estudios
- Límites de modelos lineales
- Generalización

Gestos a travéz de contextos y culturas

Brown et al. (to appear): Iconic gestures are modulated by social context: A study of multimodal politeness across two cultures. *Gesture*.

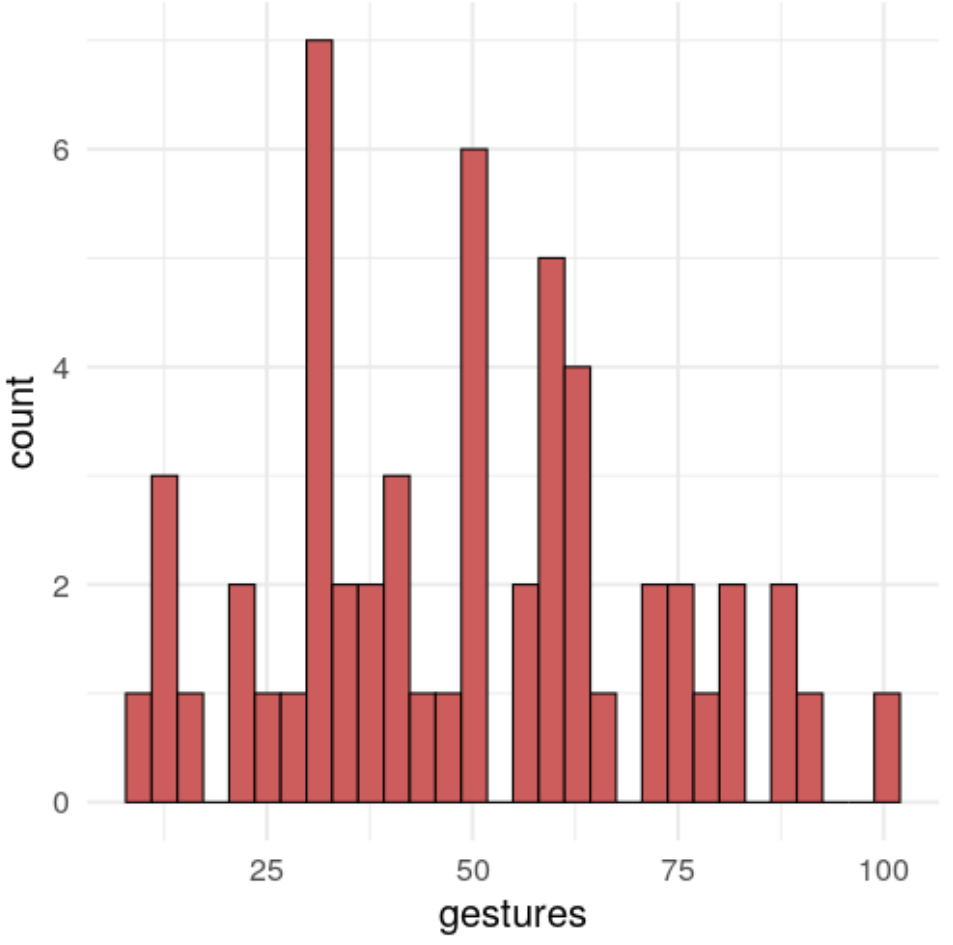
Datos

```
##          ID context dur language gender gestures
## 1 Catalan_1  friend 137  Catalan      M       61
## 2 Catalan_1    prof 136  Catalan      M       78
## 3 Catalan_2  friend  90  Catalan      F       31
## 4 Catalan_2    prof 107  Catalan      F       40
## 5 Catalan_3  friend 181  Catalan      M       81
## 6 Catalan_3    prof 165  Catalan      M       49
```

Variables independientes (possibles predictores)

- `gender`: M o F
- `context`: friend o prof
- `language`: Catalan o Korean

Variable dependiente (resultado)



Variable dependiente (resultado)

```
##  [1]  61  78  31  40  81  49  32  49  39  32  35  30  88  73  59  47  22
## [43]  51  27  67  49  61  61  31  16  63  42  22  10
```

Límites de modelos lineales

$$y_i \sim \text{Normal}(\mu_i, \sigma)$$

$$y_i \sim \text{Poisson}(\lambda_i)$$

$$y_i \sim \text{Bernoulli}(p_i)$$

Modelo lineal

$$y_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = \beta_0 + \beta_1 x_1 + \ldots \beta_n x_n$$

Modelo lineal generalizado: Normal

$$y_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\mu_i = f(\beta_0 + \beta_1 x_1 + \ldots \beta_n x_n)$$

$$f(x) = x$$

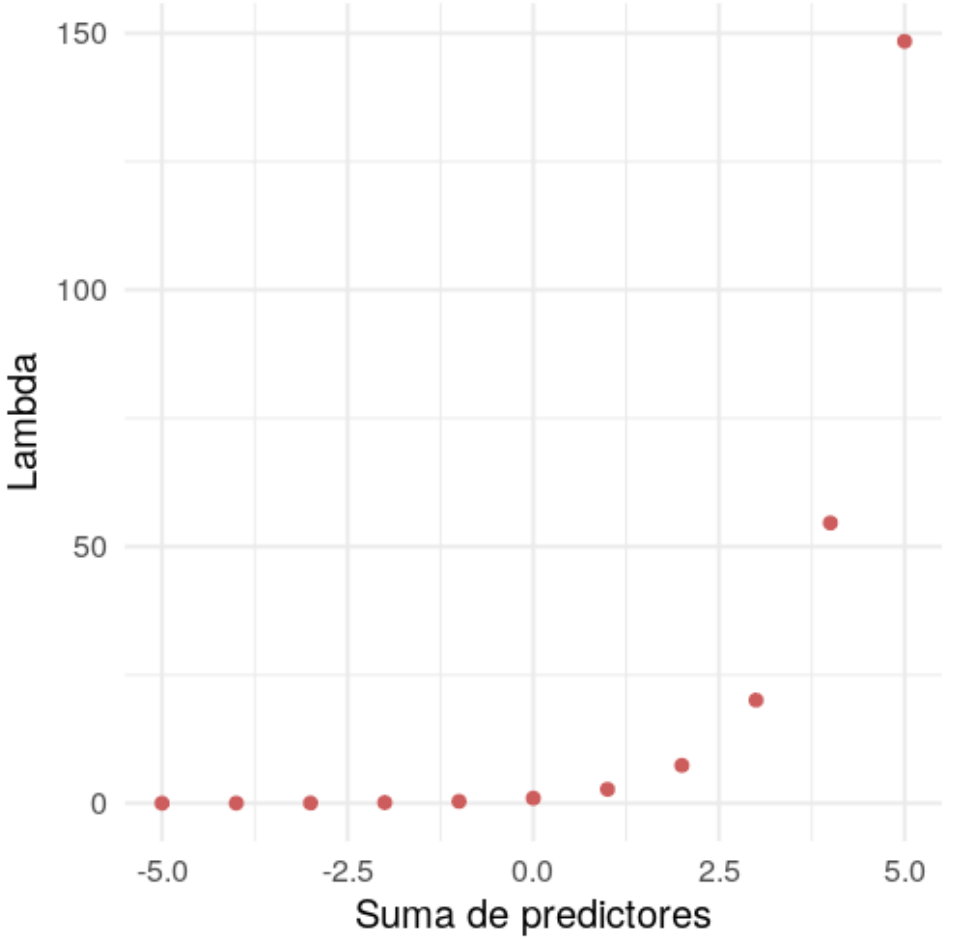
Modelo lineal generalizado: Poisson

$$y_i \sim \text{Poisson}(\lambda_i)$$

$$\lambda_i = f(\beta_0 + \beta_1 x_1 + \ldots \beta_n x_n)$$

$$f(x) = \exp(x)$$

Espacio exponencial



Modelo lineal generalizado: Poisson

$$y_i \sim \text{Poisson}(\lambda_i)$$

$$\lambda_i = exp(\beta_0 + \beta_1 x_1 + \ldots \beta_n x_n)$$

Modelo lineal generalizado: Poisson

$$y_i \sim \text{Poisson}(\lambda_i)$$

$$log(\lambda_i) = \beta_0 + \beta_1 x_1 + \ldots \beta_n x_n$$

Link functions canónicas

Normal: Identitad $$f(x) = x$$

Poisson: Logarítmo $$f(x) = \exp(x)$$

Bernoulli/Binomial: Logit $$f(x) = \frac{exp(x)}{1 + exp(x)}$$

Regresión de Poisson (R)

```
glm(formula = gestures ~ 1 + language,
    data    = df,
    family  = poisson(link = 'log')
    )
```

Gestos a travéz de contextos y culturas

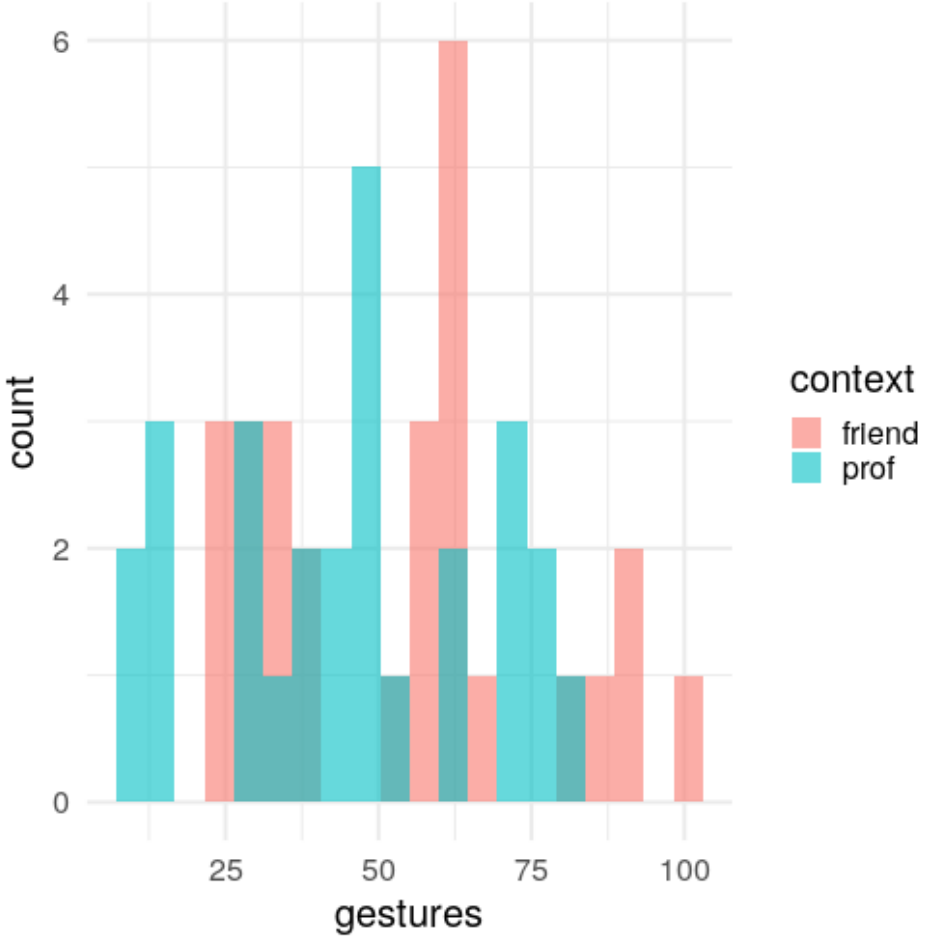
Brown et al. (to appear): Iconic gestures are modulated by social context: A study of multimodal politeness across two cultures. *Gesture*.

Datos

```
head(df)
```

```
##          ID context dur language gender gestures
## 1 Catalan_1  friend 137  Catalan      M       61
## 2 Catalan_1    prof 136  Catalan      M       78
## 3 Catalan_2  friend  90  Catalan      F       31
## 4 Catalan_2    prof 107  Catalan      F       40
## 5 Catalan_3  friend 181  Catalan      M       81
## 6 Catalan_3    prof 165  Catalan      M       49
```

Gestos ~ Contexto



Modelo 1: Contexto

```
m1_contexto <- glm(formula = gestures ~ 1 + context,
                   data    = df,
                   family  = poisson(link = 'log')
                   )

coef(m1_contexto)
```

```
## (Intercept) contextprof
##   3.9855488  -0.1772416
```

$$\lambda_i \approx \exp(3.99 - (\text{polite} \times 0.18))$$

Modelo 1: Contexto

$$\lambda_i \approx \exp(3.99 - (\text{polite} \times 0.18))$$

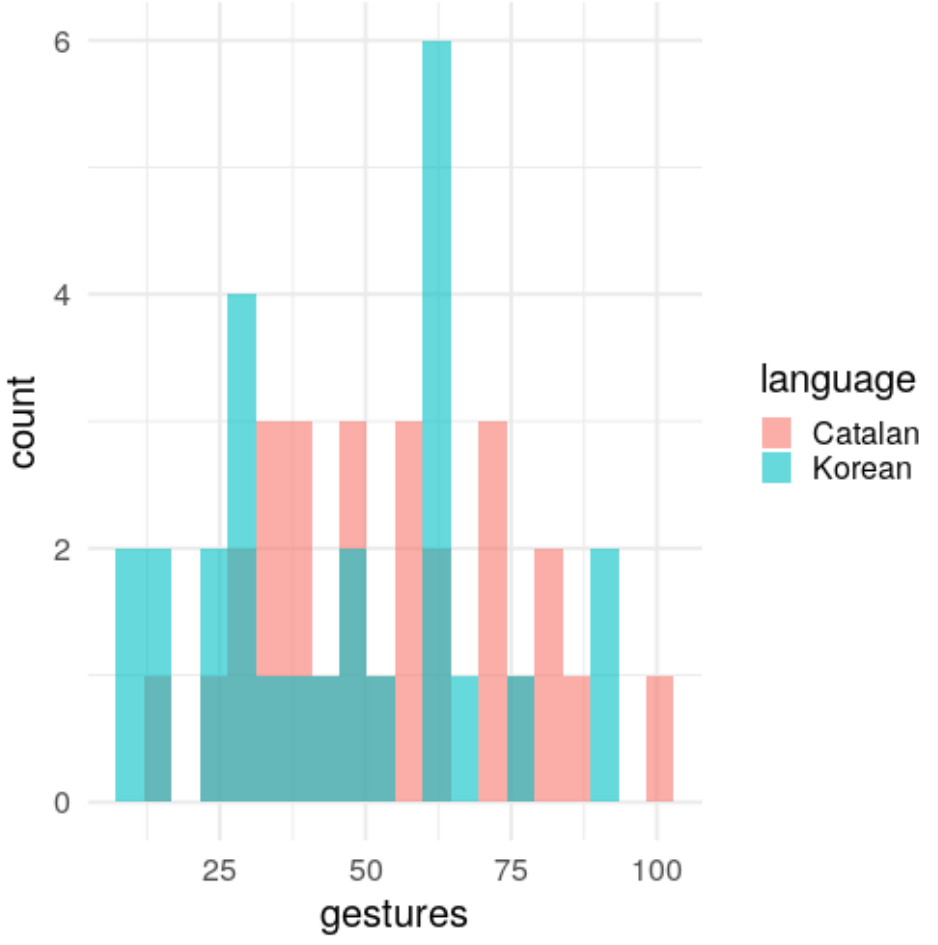
```
exp(3.99 - 0.18) #expected gestures in polite context
```

```
## [1] 45.15044
```

```
exp(3.99)        #expected gestures in informal context
```

```
## [1] 54.05489
```

Gestos ~ Lenguaje



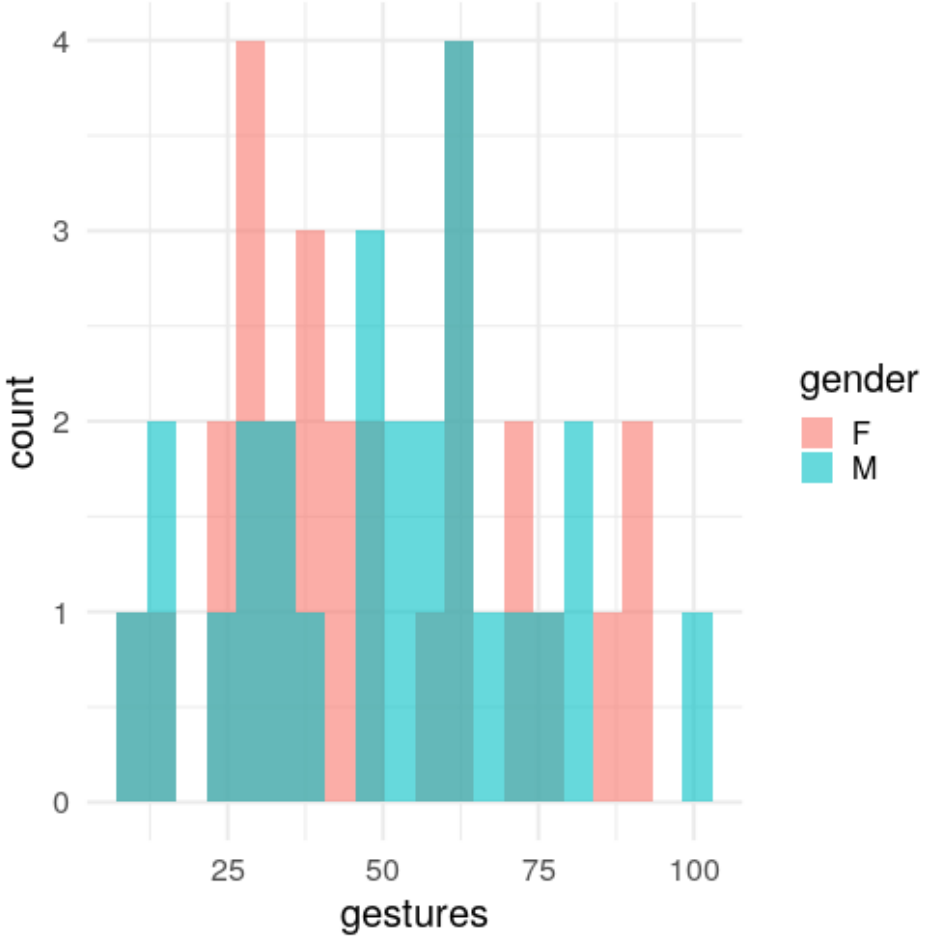
Modelo 2: Lenguaje

```
m2_lenguaje <- glm(formula = gestures ~ 1 + language,
                   data    = df,
                   family  = poisson(link = 'log')
                   )

coef(m2_lenguaje)
```

```
##    (Intercept) languageKorean
##      3.9796817     -0.1713112
```

Gestos ~ Género



Modelo 3: Género

```
m3_genero <- glm(formula   = gestures ~ 1 + gender,
                 data    = df,
                 family  = poisson(link = 'log')
                 )

coef(m3_genero)
```

```
## (Intercept)     genderM
##  3.88082715  0.04114619
```

summary(m1_contexto)

```
##
## Call:
## glm(formula = gestures ~ 1 + context, family = poisson(link = "log"),
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -6.3272  -2.8687   0.4296   1.6026   5.7274
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.98555    0.02623 151.923  < 2e-16 ***
## contextprof -0.17724    0.03886  -4.561 5.08e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 583.97  on 53  degrees of freedom
## Residual deviance: 563.08  on 52  degrees of freedom
## AIC: 870.22
##
## Number of Fisher Scoring iterations: 4
```

Akaike Information Criterion (AIC)

- Estimado de predicción fuera de muestra (out of sample prediction)
- Sirve de ranking relativo a otros modelos (que vieron los mismos datos)
- AIC más bajo $\Rightarrow$ mejor

summary(m2_lenguaje)

```
##
## Call:
## glm(formula = gestures ~ 1 + language, family = poisson(link = "log"),
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -6.6496  -2.6240  -0.4041   2.4834   6.0006
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)     3.97968    0.02584 154.030  < 2e-16 ***
## languageKorean -0.17131    0.03900  -4.393 1.12e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 583.97  on 53  degrees of freedom
## Residual deviance: 564.56  on 52  degrees of freedom
## AIC: 871.7
##
## Number of Fisher Scoring iterations: 4
```

summary(m3_genero)

```
##
## Call:
## glm(formula = gestures ~ 1 + gender, family = poisson(link = "log"),
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -6.7432  -2.7667  -0.2118   1.9619   6.2463
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.88083    0.02715 142.961   <2e-16 ***
## genderM      0.04115    0.03871   1.063    0.288
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 583.97  on 53  degrees of freedom
## Residual deviance: 582.84  on 52  degrees of freedom
## AIC: 889.98
##
## Number of Fisher Scoring iterations: 4
```

Modelo 4: 3 predictores

```
m4_todos <-  glm(formula   = gestures ~ 1 + gender + context + langu
                 data    = df,
                 family  = poisson(link = 'log')
                 )
coef(m4_todos)
```

```
##     (Intercept)         genderM    contextprof languageKorean
##      4.04690799      0.03464544    -0.17724157    -0.16997795
```

Modelo 4: 3 predictores

```
##
## Call:
## glm(formula = gestures ~ 1 + gender + context + language, family = poisso
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -6.1975  -2.7275  -0.0876   2.2215   5.4705
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)     4.04691    0.03701 109.355  < 2e-16 ***
## genderM         0.03465    0.03874   0.894    0.371
## contextprof    -0.17724    0.03886  -4.561 5.08e-06 ***
## languageKorean -0.16998    0.03903  -4.356 1.33e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 583.97  on 53  degrees of freedom
## Residual deviance: 542.87  on 50  degrees of freedom
## AIC: 854.01
```

Modelo 4: 2 predictores

```
m5_casi_todos <-  glm(formula   = gestures ~ 1 + context + language
                      data    = df,
                      family  = poisson(link = 'log')
                     )
coef(m5_casi_todos)
```

```
##    (Intercept)    contextprof languageKorean
##      4.0643807     -0.1772416     -0.1713112
```

Modelo 4: 2 predictores

```
##
## Call:
## glm(formula = gestures ~ 1 + context + language, family = poisson(link =
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -6.0963  -2.7462   0.0327   2.2950   5.3442
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)     4.06438    0.03132 129.749  < 2e-16 ***
## contextprof    -0.17724    0.03886  -4.561 5.08e-06 ***
## languageKorean -0.17131    0.03900  -4.393 1.12e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 583.97  on 53  degrees of freedom
## Residual deviance: 543.67  on 51  degrees of freedom
## AIC: 852.81
##
```

AICs

```
m1_contexto$aic
```

```
## [1] 870.2175
```

```
m2_lenguaje$aic
```

```
## [1] 871.6956
```

```
m3_genero$aic
```

```
## [1] 889.9752
```

```
m4_todos$aic
```

```
## [1] 854.0088
```

```
m5_casi_todos$aic
```

```
## [1] 852.8084
```

Próxima sesión

- Entrega de "Practical exercise" (08:00 AM 24/05)

- **Modelos lineales generalizados II**

- Entrega parte II de "Revisión por pares": 24/05 - 31/05
- Informe final: 28/06