6 Regressió generalitzada I

Mètodes empírics 2

13/05/2024

Avui

- Cas d'estudis
- Límits de models lineals
- Generalització

Gestos a través de contextos i cultures

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

Vols saber quins factors (no) afecten el nombre de gestos que fa una persona de mitjana en una conversa. Un col·lega t'ha donat les dades següents:

https://tinyurl.com/gestures-data

1. Descriu les dades

2. Descriu com penses que les variables es podrien relacionar

3. Quins valors poden tenir les variables del teu interès?

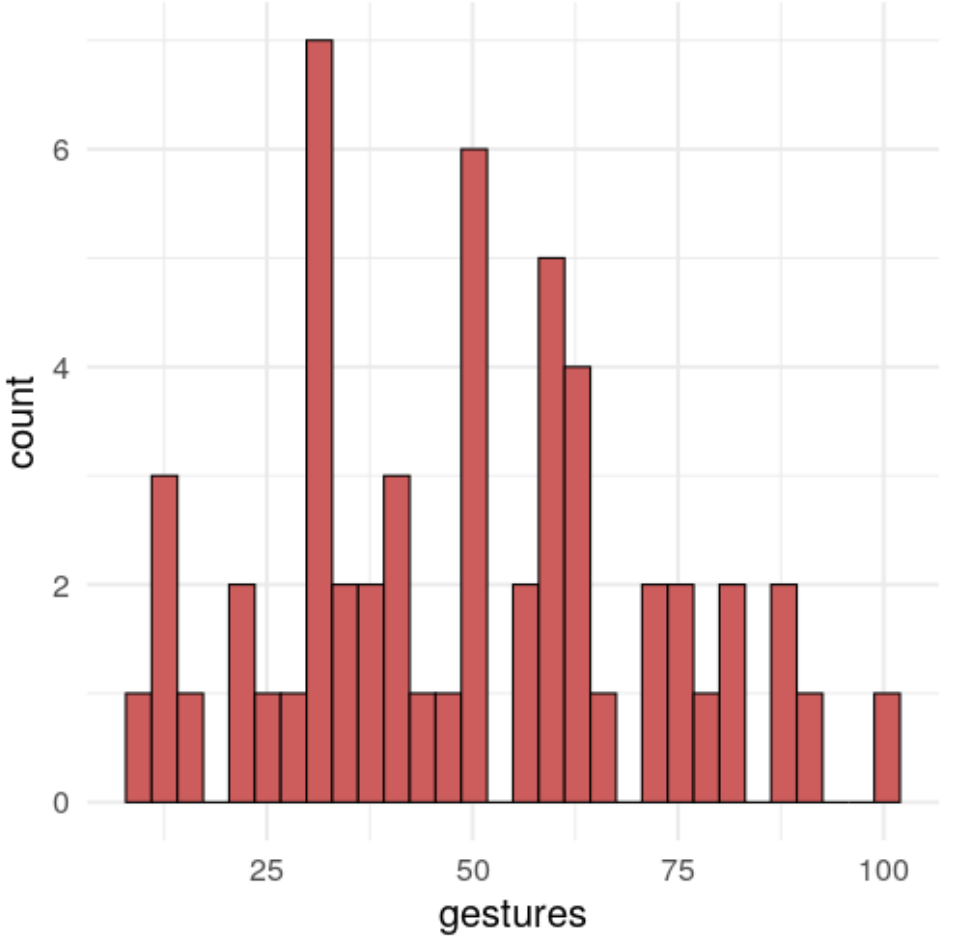
Dades

```
##          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 independents (possibles predictors)

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

Variable dependent (resultat)



Variable dependent (resultat)

```
##  [1]  61  78  31  40  81  49  32  49  39  32  35  30  88  73  59  47  22
## [27]  40  43  91  76  62  37  89  49  30  30  33  11  26  13  63  62  51
## [53]  22  10
```

Límits de models lineals

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

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

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

Model lineal

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

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

Model lineal generalitzat: Normal

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

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

$$f(x) = x$$

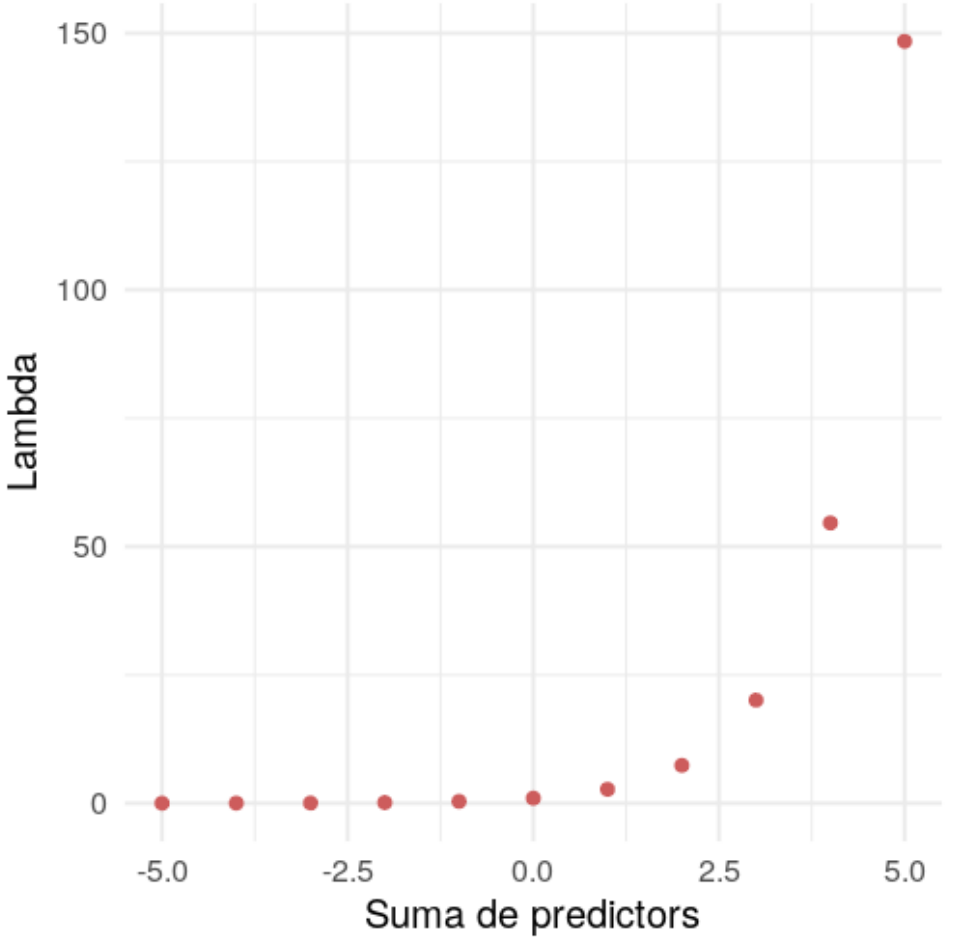
Model lineal generalitzat: Poisson

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

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

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

Espai exponencial



Model lineal generalitzat: Poisson

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

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

Model lineal generalitzat: Poisson

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

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

Link functions canòniques

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

Poisson: Logaritme $$f(x) = \exp(x)$$

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

Regressió de Poisson (R)

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

Gestos a través de contextos i cultures

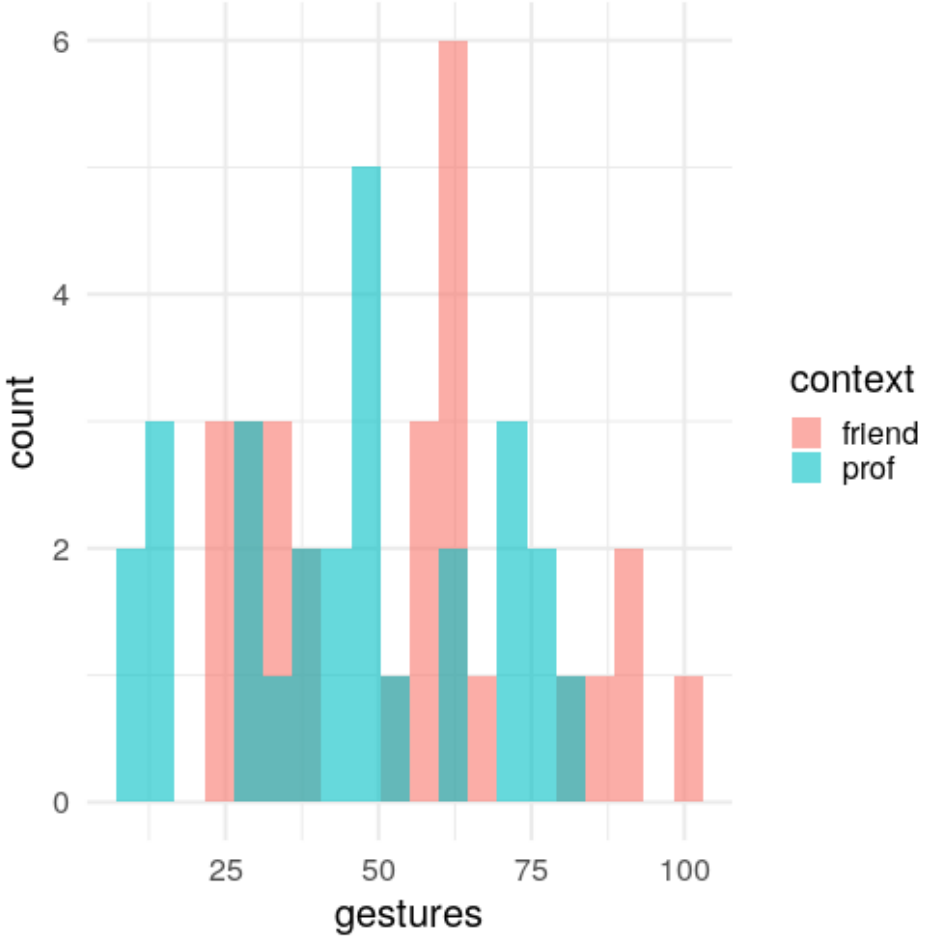
Brown et al. (to appear): Iconic gestures are modulated by social context: A study of multimodal politeness across two cultures. *Gesture*.

Dades

```
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 ~ Context



Model 1: Context

```
m1_context <- glm(formula = gestures ~ 1 + context,
                  data    = df,
                  family  = poisson(link = 'log')
                  )

coef(m1_context)
```

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

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

Model 1: Context

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

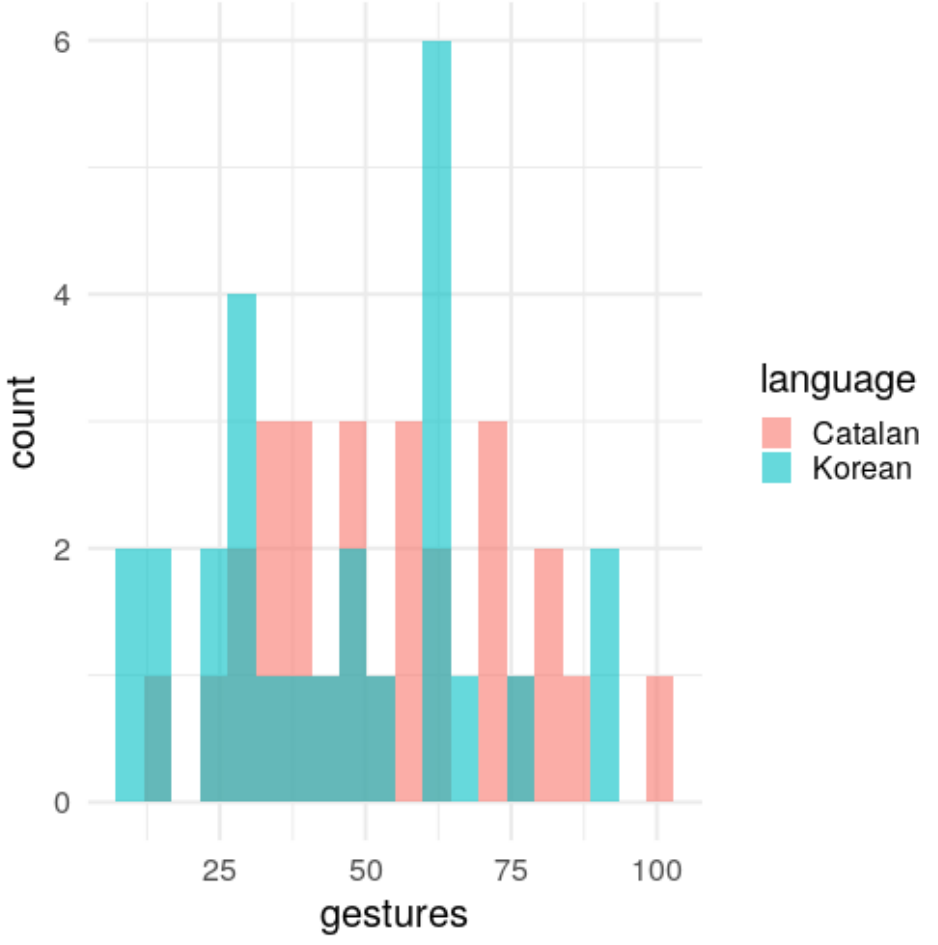
```
exp(3.99 - 0.18) #expected gestures in polite context
```

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

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

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

Gestos ~ Llenguatge



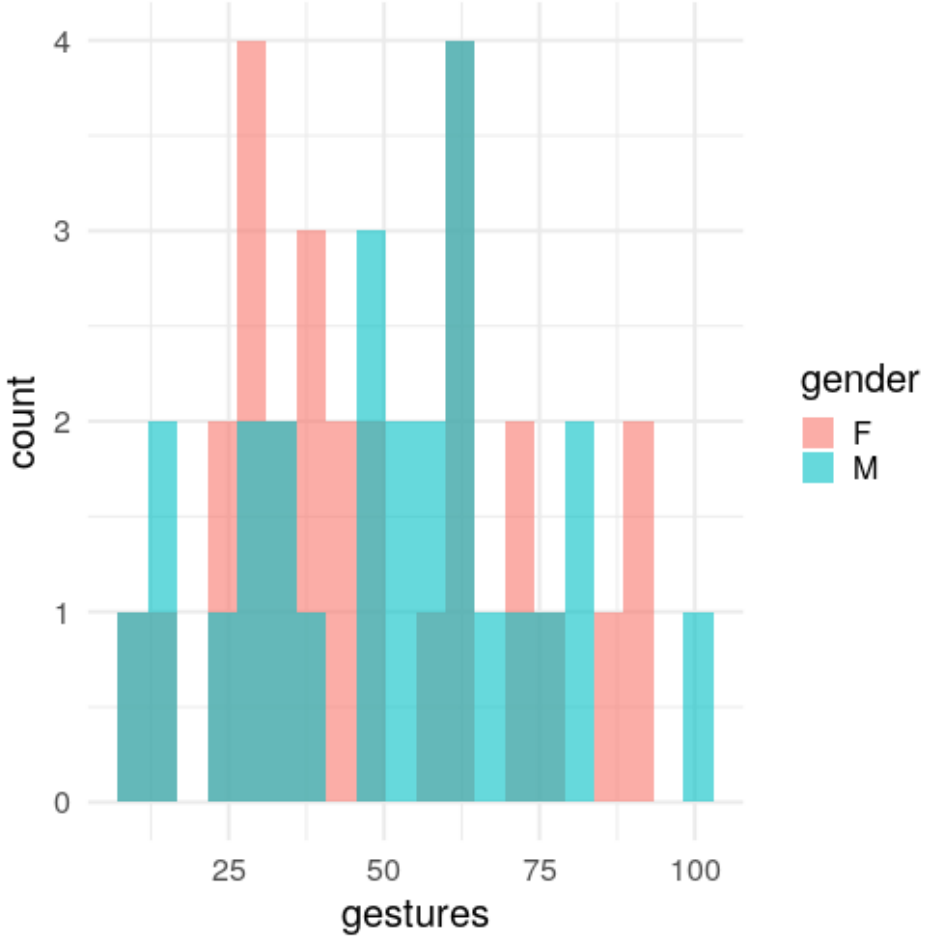
Model 2: Llenguatge

```
m2_lleng <- glm(formula = gestures ~ 1 + language,
                data    = df,
                family  = poisson(link = 'log')
                )

coef(m2_lleng)
```

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

Gestos ~ Gènere



Modelo 3: Gènere

```
m3_gen <- glm(formula   = gestures ~ 1 + gender,
                  data    = df,
                  family  = poisson(link = 'log')
                  )

coef(m3_gen)
```

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

summary(m1_context)

```
##
## 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)

- Estimat de predicció fora de mostra (out of sample prediction)
- Serveix de rànquing relatiu a altres models (que van veure les mateixes dades)
- AIC més baix $\Rightarrow$ millor

summary(m2_lleng)

```
##
## 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_gen)

```
##
## 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
```

Model 4: 3 predictors

```
m4_all <-  glm(formula   = gestures ~ 1 + gender + context + langua
                  data    = df,
                  family  = poisson(link = 'log')
                  )
coef(m4_all)
```

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

Model 4: 3 predictors

```
##
## 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
```

Model 4: 2 predictors

```
m5_gairabe_tots <-  glm(formula   = gestures ~ 1 + context + langua
                        data    = df,
                        family  = poisson(link = 'log')
                    )
coef(m5_gairabe_tots)
```

```
##    (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_context$aic
```

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

```
m2_lleng$aic
```

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

```
m3_gen$aic
```

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

```
m4_all$aic
```

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

```
m5_gairabe_tots$aic
```

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

Propera sessió

- Lliurament de Assignment 5

- **Visualització**