EJEMPLO DE SOBREDISPERSIÓN

Se ajustan 4 modelos

A data set giving the number of publications by doctoral candidates in biochemistry in relation to various predictors, originally from Long (1997).

A data frame with 915 observations on the following 6 variables.

articles *number of articles published in the final three years of PhD studies*

female *dummy variable for gender, coded 1 for female*

married *dummy variable for marital status, coded 1 for married*

kid5 *number of young children, age 5 and under*

phdprestige *prestige of the PhD department*

mentor *number of publications by the mentor in the preceeding three years*

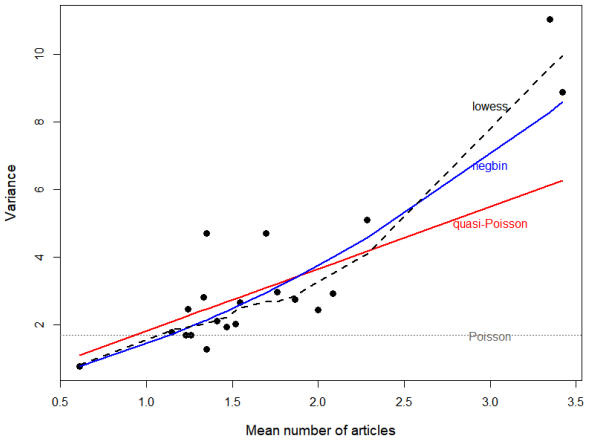
```
> art.fac <- factor(PhdPubs$articles, levels = 0 : 19)  # include zero frequencies
> art.tab <- table(art.fac)
> art.tab
art.fac
  0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19
275 246 178  84  67  27  17  12   1   2   1   1   2   0   0   0   1   0   0   1
```

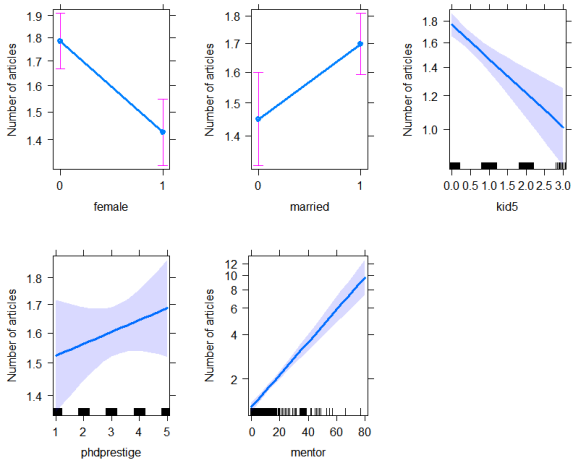
Hay muchos ceros.

```
> phd.pois <- glm(articles ~ ., data = PhdPubs, family = poisson)
> summary(phd.pois)

Call:
glm(formula = articles ~ ., family = poisson, data = PhdPubs)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.4879  -1.5380  -0.3652   0.5771   5.4832

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.26562    0.09962   2.666  0.00767 **
female1     -0.22443    0.05458  -4.112 3.93e-05 ***
married1     0.15732    0.06125   2.569  0.01021 *
kid5        -0.18491    0.04011  -4.610 4.03e-06 ***
phdprestige  0.02538    0.02527   1.004  0.31531
mentor       0.02523    0.00203  12.430  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1817.4  on 914  degrees of freedom
Residual deviance: 1633.6  on 909  degrees of freedom
AIC: 3313.3

Number of Fisher Scoring iterations: 5
```

Modelo1

Regresión Poisson




```
> sum(residuals(phd.pois, type = "pearson")^2) / phd.pois$df.residual
[1] 1.830368
```

Muy diferente de 1

```
> summary(phd.qpois)

Call:
glm(formula = articles ~ ., family = quasipoisson, data = PhdPubs)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.4879  -1.5380  -0.3652   0.5771   5.4832

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.265625   0.134783   1.971 0.049055 *
female1     -0.224425   0.073843  -3.039 0.002440 **
married1     0.157324   0.082867   1.899 0.057947 .
kid5        -0.184914   0.054273  -3.407 0.000685 ***
phdprestige  0.025378   0.034192   0.742 0.458154
mentor       0.025231   0.002746   9.188  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for quasipoisson family taken to be 1.830389)

    Null deviance: 1817.4  on 914  degrees of freedom
Residual deviance: 1633.6  on 909  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 5
> phd.nbin  <- glm.nb(articles ~ ., data = PhdPubs)
> summary(phd.nbin)

Call:
glm.nb(formula = articles ~ ., data = PhdPubs, init.theta = 2.266961466,
    link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1359  -1.3654  -0.2807   0.4431   3.4765

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.212949   0.132741   1.604 0.108661
female1     -0.216253   0.072594  -2.979 0.002893 **
married1     0.152791   0.081944   1.865 0.062242 .
kid5        -0.176339   0.052794  -3.340 0.000837 ***
phdprestige  0.029342   0.034273   0.856 0.391921
mentor       0.028677   0.003236   8.861  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(2.267) family taken to be 1)

    Null deviance: 1109.4  on 914  degrees of freedom
Residual deviance: 1004.1  on 909  degrees of freedom
```

Modelo 2.

Quasipoisson

Modelo 3

Binomial Negativa

El otro parámetro de la BN (Θ=2.267)

```
AIC: 3135.4

Number of Fisher Scoring iterations: 1


              Theta:  2.267
          Std. Err.:  0.272

 2 x log-likelihood:  -3121.367

###D=alfa=1/theta=1/2.267=0.4411116
###para hacer la prueba de si D es chiquito
 > library(pscl)
 > odTest(phd.nbin )
 Likelihood ratio test of H0: Poisson, as restricted NB model:
 n.b., the distribution of the test-statistic under H0 is non-sta    rd
 e.g., see help(odTest) for details/references

 Critical value of test statistic at the alpha= 0.05 level: 2.7055
 Chi-Square Test Statistic =  179.972 p-value = < 2.2e-16
```

Var(y)=mu+(1/theta)mu$^2$

Se rechaza que +(1/theta) sea chiquito.

Modelo 4

Zero Inflated Poisson

```
> summary(modelozero<-zeroinfl(articles ~ ., data = PhdPubs))


Call:
zeroinfl(formula = articles ~ ., data = PhdPubs)

Pearson residuals:
    Min      1Q  Median      3Q     Max
-2.2751 -0.8656 -0.2839  0.5403  7.4055

Count model coefficients (poisson with log link):
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.599179   0.118614   5.051 4.38e-07 ***
female1     -0.208790   0.063526  -3.287  0.00101 **
married1     0.106230   0.070966   1.497  0.13442
kid5        -0.142706   0.047436  -3.008  0.00263 **
phdprestige  0.006998   0.029812   0.235  0.81441
mentor       0.017846   0.002334   7.647 2.06e-14 ***

Zero-inflation model coefficients (binomial with logit link):
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.563316   0.494050  -1.140  0.25420
female1      0.108162   0.281731   0.384  0.70104
married1    -0.355580   0.317962  -1.118  0.26343
kid5         0.219738   0.196580   1.118  0.26365
phdprestige -0.005371   0.141183  -0.038  0.96965
mentor      -0.133126   0.046431  -2.867  0.00414 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of iterations in BFGS optimization: 21
Log-likelihood: -1605 on 12 Df
```

Modelamos conteos Poisson. (incluye posibles ceros)

Modelamos las observaciones que SIEMPRE son cero. Coeficientes no significativos

```
 > round(phd.coefi, 4)
                pois   qpois    nbin    zero
 (Intercept)  0.2656  0.2656  0.2129  0.5992
 female1     -0.2244 -0.2244 -0.2163 -0.2088
 married1     0.1573  0.1573  0.1528  0.1062
 kid5        -0.1849 -0.1849 -0.1763 -0.1427
 phdprestige  0.0254  0.0254  0.0293  0.0070
 mentor       0.0252  0.0252  0.0287  0.0178
```