

# Análisis de Datos Categóricos

Tarea 02

*Rivera Torres Francisco de Jesús*

*Rodríguez Maya Jorge Daniel*

*Samayoa Donado Víctor Augusto*

*Trujillo Bariños Georgina*

*Mayo 23, 2019*

## Ejercicio 1

Los datos `bacalao.csv` corresponden a número de bacalaos capturados en diferentes estaciones de pesca. Para cada estación, se provee del ID de la estación, la temperatura promedio del agua ( $^{\circ}F$ ), la latitud, la longitud y la profundidad máxima.

Utilizando como variable dependiente el número de peces capturados (conteos) y como variables explicativas el resto de las variables, se han generado dos modelos: modelo 1 y modelo 2. Se muestran a continuación las salidas de R para ambos modelos.

Tabla 1: Datos de la encuesta

ID estación	Latitud	Longitud	Profundidad	Temperatura	Peces capturados
356	71.10	22.43	349	3.95	44
357	71.32	23.68	382	3.75	138
358	71.60	24.90	294	3.45	675
359	71.27	25.88	304	3.65	357
363	71.52	28.12	384	3.35	76
364	71.48	29.10	344	3.65	368

```
modelo2<-glm(peces_capturados ~ Latitude + Longitude + Temperature + Depth,
             family = poisson,
             data = Bacalao)
summary(modelo2)

##
## Call:
## glm(formula = peces_capturados ~ Latitude + Longitude + Temperature +
##      Depth, family = poisson, data = Bacalao)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -24.280   -9.521   -3.173    2.118   49.442
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept) -0.6384251  0.6400393  -0.997    0.319
## Latitude     0.0681683  0.0082264   8.286   <2e-16 ***
## Longitude    0.0926485  0.0022925  40.414   <2e-16 ***
## Temperature  0.1047362  0.0115462   9.071   <2e-16 ***
## Depth        -0.0056388  0.0001372 -41.107   <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: 21533  on 88  degrees of freedom
## Residual deviance: 15468  on 84  degrees of freedom
## AIC: 16068
##
## Number of Fisher Scoring iterations: 5

modelo1<-glm(peces_capturados ~ Latitude + Longitude + Temperature,
             family = poisson,
             data = Bacalao)
summary(modelo1)

##
## Call:
## glm(formula = peces_capturados ~ Latitude + Longitude + Temperature,
##      family = poisson, data = Bacalao)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -24.691  -11.437   -4.987    3.259   50.310
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.765474   0.627919   4.404 1.06e-05 ***
## Latitude    -0.013039   0.007901  -1.650  0.0989 .
## Longitude    0.122820   0.002334  52.630 < 2e-16 ***
## Temperature  0.047880   0.011449   4.182 2.89e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 21533  on 88  degrees of freedom
## Residual deviance: 17095  on 85  degrees of freedom
## AIC: 17693
##
## Number of Fisher Scoring iterations: 5

anova(modelo1, modelo2, test = "Chisq")
```

Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
85	17095.08	NA	NA	NA
84	15467.93	1	1627.155	0

**Inciso 1.a)**

Escribe los modelos 1 y 2 con los valores de los coeficientes, indicando claramente los componentes aleatorio, sistemático y la función liga.

**Inciso 1.b)**

Escribe las hipótesis nula ( $H_0$ ) y alternativa ( $H_A$ ) del Análisis de Devianza que se muestra en la salida de R.

**Inciso 1.c)**

Escribe la conclusión de la prueba de Análisis de Devianza para la que escribiste las hipótesis, incluyendo en tu conclusión si vale o no la pena la inclusión de “*Depth*” (profundidad máxima) en el modelo.

**Inciso 1.d)**

Interpreta los coeficientes del modelo 2.

**Ejercicio 2**

En el archivo `encuesta.csv` se encuentran los conteos de una encuesta realizada en EU en donde se les preguntó a adultos si estaban de acuerdo con la distribución de condones a adolescentes (si condón/no condón) , si estaban de acuerdo con el sexo premarital (desacuerdo / acuerdo) y su posición política (liberal/demócrata/ republicano).

Tabla 3: Datos de la encuesta

Conteos	Posición	Condomes	Premarital
172	republicano	no c	sexo no
160	democrata	no c	sexo no
102	liberal	no c	sexo no
40	republicano	si c	sexo no
57	democrata	si c	sexo no
55	liberal	si c	sexo no
32	republicano	no c	sexo si
70	democrata	no c	sexo si

Tabla 3: Datos de la encuesta (*continued*)

Conteos	Posición	Condomes	Premarital
39	liberal	no c	sexo si
26	republicano	si c	sexo si
73	democrata	si c	sexo si
100	liberal	si c	sexo si

```

modeloA <- glm(conteos ~ posicion + condones + premarital,
               family = poisson,
               data = Encuesta)
summary(modeloA)

##
## Call:
## glm(formula = conteos ~ posicion + condones + premarital, family = poisson,
##      data = Encuesta)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.153  -3.328  -1.673   1.901   7.730
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      4.95205    0.06375  77.684 < 2e-16 ***
## posicionliberal  -0.19574    0.07846  -2.495 0.012603 *
## posicionrepublicano -0.28768    0.08051  -3.573 0.000352 ***
## condonessi c      -0.49358    0.06774  -7.287 3.17e-13 ***
## premaritalsexo si  -0.54437    0.06817  -7.985 1.40e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 305.60  on 11  degrees of freedom
## Residual deviance: 171.04  on  7  degrees of freedom
## AIC: 253.25
##
## Number of Fisher Scoring iterations: 5
modeloB <- glm(conteos ~ posicion * condones * premarital,
               family = poisson,
               data = Encuesta)
summary(modeloB)

##
## Call:

```

```
## glm(formula = conteos ~ posicion * condones * premarital, family = poisson,
##      data = Encuesta)
##
## Deviance Residuals:
##  [1]  0  0  0  0  0  0  0  0  0  0  0  0  0
##
## Coefficients:
##                                     Estimate Std. Error
## (Intercept)                        5.07517    0.07906
## posicionliberal                    -0.45020    0.12670
## posicionrepublicano                0.07232    0.10984
## condonessi c                      -1.03212    0.15425
## premaritalsexos i                  -0.82668    0.14330
## posicionliberal:condonessi c        0.41448    0.22755
## posicionrepublicano:condonessi c    -0.42649    0.23368
## posicionliberal:premaritalsexos i   -0.13473    0.23660
## posicionrepublicano:premaritalsexos i -0.85508    0.24000
## condonessi c:premaritalsexos i      1.07409    0.22755
## posicionliberal:condonessi c:premaritalsexos i 0.48516    0.33971
## posicionrepublicano:condonessi c:premaritalsexos i 0.17689    0.39026
##                                     z value Pr(>|z|)
## (Intercept)                        64.196 < 2e-16 ***
## posicionliberal                    -3.553 0.000381 ***
## posicionrepublicano                0.658 0.510254
## condonessi c                      -6.691 2.21e-11 ***
## premaritalsexos i                  -5.769 7.99e-09 ***
## posicionliberal:condonessi c        1.821 0.068532 .
## posicionrepublicano:condonessi c    -1.825 0.067988 .
## posicionliberal:premaritalsexos i   -0.569 0.569053
## posicionrepublicano:premaritalsexos i -3.563 0.000367 ***
## condonessi c:premaritalsexos i      4.720 2.36e-06 ***
## posicionliberal:condonessi c:premaritalsexos i 1.428 0.153249
## posicionrepublicano:condonessi c:premaritalsexos i 0.453 0.650363
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance:  3.0560e+02  on 11  degrees of freedom
## Residual deviance: -1.8652e-14  on  0  degrees of freedom
## AIC: 96.209
##
## Number of Fisher Scoring iterations: 3
modeloC <- glm(conteos ~ premarital + condones*posicion,
               family = poisson,
               data = Encuesta)
summary(modeloC)
```

```
##
## Call:
## glm(formula = conteos ~ premarital + condones * posicion, family = poisson,
##      data = Encuesta)
##
## Deviance Residuals:
##      1      2      3      4      5      6      7      8
##  3.5914  1.1786  1.3215 -0.2753 -2.9506 -4.7474 -5.6015 -1.6207
##      9     10     11     12
## -1.8566  0.3547  3.3901  5.1534
##
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   4.98052    0.07053   70.616 < 2e-16 ***
## premaritalsexo si             -0.54437    0.06817   -7.985 1.40e-15 ***
## condonessi c                  -0.57054    0.10973   -5.200 2.00e-07 ***
## posicionliberal               -0.48932    0.10696   -4.575 4.76e-06 ***
## posicionrepublicano           -0.11996    0.09617   -1.247  0.21229
## condonessi c:posicionliberal    0.66521    0.15995    4.159 3.20e-05 ***
## condonessi c:posicionrepublicano -0.55792    0.17915   -3.114  0.00184 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 305.60  on 11  degrees of freedom
## Residual deviance: 122.98  on  5  degrees of freedom
## AIC: 209.19
##
## Number of Fisher Scoring iterations: 4
modeloD <- glm(conteos ~ condones + posicion*premarital,
               family = poisson,
               data = Encuesta)
summary(modeloD)

##
## Call:
## glm(formula = conteos ~ condones + posicion * premarital, family = poisson,
##      data = Encuesta)
##
## Deviance Residuals:
##      1      2      3      4      5      6      7      8
##  3.3575  2.1124  0.4534 -4.9907 -2.9491 -0.5924 -0.6821 -2.0721
##      9     10     11     12
## -5.7149  0.8321  2.4234  5.7907
##
## Coefficients:
```

```

##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   4.90339    0.07258  67.561 < 2e-16
## condonessi c                  -0.49358    0.06774  -7.287 3.17e-13
## posicionliberal               -0.32365    0.10477  -3.089 0.002008
## posicionrepublicano          -0.02331    0.09657  -0.241 0.809248
## premaritalsexos si            -0.41705    0.10771  -3.872 0.000108
## posicionliberal:premaritalsexos si  0.29528    0.15863   1.861 0.062689
## posicionrepublicano:premaritalsexos si -0.87909    0.18319  -4.799 1.60e-06
##
## (Intercept)                   ***
## condonessi c                  ***
## posicionliberal               **
## posicionrepublicano
## premaritalsexos si            ***
## posicionliberal:premaritalsexos si  .
## posicionrepublicano:premaritalsexos si ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##    Null deviance: 305.60  on 11  degrees of freedom
## Residual deviance: 127.41  on  5  degrees of freedom
## AIC: 213.62
##
## Number of Fisher Scoring iterations: 5
modeloE <- glm(conteos ~ posicion + condones*premarital,
               family = poisson,
               data = Encuesta)
summary(modeloE)

##
## Call:
## glm(formula = conteos ~ posicion + condones * premarital, family = poisson,
##      data = Encuesta)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.7227  -1.0642  -0.5804   1.1670   4.2070
##
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   5.12827    0.06326  81.065 < 2e-16 ***
## posicionliberal               -0.19574    0.07846  -2.495 0.012603 *
## posicionrepublicano          -0.28768    0.08051  -3.573 0.000352 ***
## condonessi c                  -1.04916    0.09425 -11.132 < 2e-16 ***
## premaritalsexos si            -1.12428    0.09693 -11.598 < 2e-16 ***

```

```
## condonessi c:premaritalsexos 1.39371 0.14491 9.617 < 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: 305.597 on 11 degrees of freedom
## Residual deviance: 74.325 on 6 degrees of freedom
## AIC: 158.53
##
## Number of Fisher Scoring iterations: 4

modeloF <- glm(conteos ~ condones*posicion + condones*premarital + posicion*premarital,
               family = poisson,
               data = Encuesta)

summary(modeloF)

##
## Call:
## glm(formula = conteos ~ condones * posicion + condones * premarital +
##      posicion * premarital, family = poisson, data = Encuesta)
##
## Deviance Residuals:
##      1       2       3       4       5       6       7
## -0.02677 -0.31740  0.44208  0.05572  0.54949 -0.57816  0.06234
##      8       9      10      11      12
##  0.49355 -0.67967 -0.06860 -0.46536  0.44661
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      5.10016    0.07541  67.630 < 2e-16
## condonessi c     -1.13079    0.13460  -8.401 < 2e-16
## posicionliberal  -0.51928    0.11821  -4.393 1.12e-05
## posicionrepublicano 0.04937    0.10457   0.472 0.63683
## premaritalsexos -0.91124    0.12763  -7.140 9.34e-13
## condonessi c:posicionliberal 0.63420    0.16799   3.775 0.00016
## condonessi c:posicionrepublicano -0.33869    0.18779  -1.804 0.07130
## condonessi c:premaritalsexos 1.28630    0.14862   8.655 < 2e-16
## posicionliberal:premaritalsexos 0.10082    0.16805   0.600 0.54854
## posicionrepublicano:premaritalsexos -0.78360    0.19042  -4.115 3.87e-05
##
## (Intercept)          ***
## condonessi c          ***
## posicionliberal       ***
## posicionrepublicano
## premaritalsexos      ***
## condonessi c:posicionliberal ***
## condonessi c:posicionrepublicano .
```



```

## condonessi c:premaritalsexos si ***
## posicionliberal:premaritalsexos si
## posicionrepublicano:premaritalsexos si ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 305.5974  on 11  degrees of freedom
## Residual deviance:   2.0664  on  2  degrees of freedom
## AIC: 94.275
##
## Number of Fisher Scoring iterations: 3
modeloG <- glm(conteos ~ condones*posicion + posicion*premarital,
               family = poisson,
               data = Encuesta)
summary(modeloG)

##
## Call:
## glm(formula = conteos ~ condones * posicion + posicion * premarital,
##      family = poisson, data = Encuesta)
##
## Deviance Residuals:
##      1       2       3       4       5       6       7       8       9
##  0.923  1.770  2.980 -1.712 -2.538 -3.195 -1.877 -2.332 -3.625
##     10     11     12
##  2.809  2.797  3.017
##
## Coefficients:
##
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      4.9319    0.0786  62.745 < 2e-16
## condonessi c      -0.5705    0.1097  -5.200  2.0e-07
## posicionliberal   -0.6172    0.1275  -4.840  1.3e-06
## posicionrepublicano  0.1444    0.1100   1.313 0.189122
## premaritalsexos si -0.4170    0.1077  -3.872 0.000108
## condonessi c:posicionliberal  0.6652    0.1600   4.159  3.2e-05
## condonessi c:posicionrepublicano -0.5579    0.1792  -3.114 0.001844
## posicionliberal:premaritalsexos si  0.2953    0.1586   1.861 0.062689
## posicionrepublicano:premaritalsexos si -0.8791    0.1832  -4.799  1.6e-06
##
## (Intercept) ***
## condonessi c ***
## posicionliberal ***
## posicionrepublicano
## premaritalsexos si ***
## condonessi c:posicionliberal ***

```

```
## condonessi c:posicionrepublicano      **
## posicionliberal:premaritalsexos      .
## posicionrepublicano:premaritalsexos  ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 305.597  on 11  degrees of freedom
## Residual deviance:  79.352  on  3  degrees of freedom
## AIC: 169.56
##
## Number of Fisher Scoring iterations: 4
anova (modeloF, modeloB, test = "Chisq")
```

Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
2	2.066	NA	NA	NA
0	0.000	2	2.066	0.356

### Inciso 2.a)

Haz una tabla con las devianzas, grados de libertad, AIC, términos incluidos en cada modelo ajustado y tipo de independencia

### Inciso 2.b)

Escribe las hipótesis nula ( $H_0$ ) y alternativa ( $H_A$ ) del Análisis de Devianza que se muestra en la salida de R.

### Inciso 2.c)

Escribe la conclusión de la prueba de Análisis de Devianza para la que escribiste las hipótesis, incluyendo en tu conclusión si vale o no la pena la inclusión de la triple interacción en el modelo.

### Inciso 2.d)

Elige al mejor modelo justificando la elección en el AIC, devianzas, grados de libertad y parsimonia.

### Inciso 2.e)

¿El *mosaicplot* que elegiste apoya tu elección? Justifica tu respuesta.