GRAU INTERUNIVERSITARI D'ESTADÍSTICA I INVESTIGACIÓ OPERATIVA (UB- UPC)

CURS 2016-2017 Q1 – EXAMEN FINAL : MODEL LINEAL GENERALITZAT

(Data: 11 de Gener a les 15:00h Aula –S04-FME)

Nom de l´alumne:	**DNI:**
Professors:	Lídia Montero – Josep Anton Sànchez
Localització:	Edifici C5 D217 o H6-67
Normativa de l'examen:	ÉS PERMÉS DUR APUNTS TEORIA *SENSE ANOTACIONS*, CALCULADORA I TAULES ESTADÍSTIQUES
Durada de l'examen:	2h 30 min
Sortida de notes:	Abans del 18 de Gener al Web Docent de MLGz
Revisió de l'examen:	18 de Gener a 13h a C5-217-C Nord o H- P6-67

PROBLEMA 1 (5 Punts)

Aquest problema es refereix a les dades d'un estudi de nidificació dels crancs de ferradura (J. Brockmann, Etologia 1996); vegeu també https://onlinecourses.science.psu.edu/stat504, Agresti (1996) Sec. 4.3 i Agresti (2002) Sec. 4.3. Cada cranc de ferradura femella en l'estudi tenia un cranc mascle unit a ella en el seu niu. L'estudi va investigar **els factors que afecten a que el cranc femella tingui altres mascles que resideixen a prop seu**, aquests altres mascles s'anomenen **satèl·lits**. Les variables explicatives que es creu que afecten al nombre de satel.lits són el color del cranc femella (color - nivells 1 a 4 sense etiquetes disponibles), condició de la columna vertebral de la femella (spine, nivells 1 a 3 sense etiquetes disponibles), el pes de la femella (pes), i l'amplada de la closca (amplada). **El target a analitzar per a cada cranc femella és el seu número de satèl·lits (satel)**. Hi ha 173 femelles en aquest estudi. La resposta es suposa distribuïda segons un model Poisson.

```
> summary(df)
       id      color  spine       pes          amplada         satel
 Min.   :  1   1:12   1: 37   Min.   :21.0   Min.   :1.200   Min.   : 0.000
 1st Qu.: 44   2:95   2: 15   1st Qu.:24.9   1st Qu.:2.000   1st Qu.: 0.000
 Median : 87   3:44   3:121   Median :26.1   Median :2.350   Median : 2.000
 Mean   : 87   4:22           Mean   :26.3   Mean   :2.437   Mean   : 2.919
 3rd Qu.:130                  3rd Qu.:27.7   3rd Qu.:2.850   3rd Qu.: 5.000
 Max.   :173                  Max.   :33.5   Max.   :5.200   Max.   :15.000
```

1. Primer es vol estudiar si l'efecte brut de l'ample de l'esquena de la femella pot explicar el nombre de satèl·lits. És una variable estadísticament significativa l'amplada de l'esquena?

La formulació de la pregunta indica que s'ha de buscar algunes sortides que permetin fer un test de la deviança per la H0: L'efecte brut d'amplada no és significatiu. El summary de m1 Y-Amplada permet veure que segons el test de Wald la covariant amplada és significativa (té un coeficient en el model de Poisson pel logaritme del nombre esperat de satel.lits igual a 0.5892, significativament diferent de zero). Les deviances del model nul i del model amb la covariant amplada només, es poden comparar D(m0)-D(m1) = 71.95 que segons la distribució asimptòtica de referència una shi quadrat de 1 grau de llibertat, dona un p valor de 0 pràcticament amb la qual cosa hi ha evidència per rebutjar la H0 que l'efecte brut no és significatiu.

```
> anova(m0,m1,test="Chisq")
Analysis of Deviance Table

Model 1: satel ~ 1
Model 2: satel ~ amplada
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       172     632.79
2       171     560.84  1   71.949 < 2.2e-16 ***
```

2. Escriviu l'equació del model de Poisson que explica el nombre de satèl·lits en funció de l'amplada de l'esquena de la femella i interpreteu els seus coeficients.

$$\rightarrow \log(\mu_i) = \eta + \theta \cdot amplada = -0.4282 + 0.5892 amplada$$

Segons el summary del model m1 es pot escriure l'equació anterior. Per cada unitat d'amplada d'increment el logaritme del nombre de satèl.lits s'incrementa en 0.59 i en l'escala de la resposta, el nombre de satel.lits per femella es multiplica per exp(0.59)=1.80 per cada unitat d'increment de l'amplada, o dit d'una altra manera cada unitat addicional de l'amplada fa incrementar en un 80% el nombre de satèl.lits per femella.

El terme independent correspondria al cas d'amplada 0, impossible, que donaria un nombre esperat de satèl.lits de exp(-0.4282)=0.652, que no val la pena interpretar.

3. Segons el model m1, quina és la probabilitat que una femella de 2.15 unitats d'amplada d'esquena tingui 0 satèl.lits? I 1 satèl.lit? I més de 1 satèl.lit?

*El nombre esperat de satèl.lits per una femella de 2.15 unitats d'amplada d'esquena és exp(-0.4282+0.5892*2.15)=2.31 satèllits, que és l'esperanca de la llei de Poisson que modela el nombre de satèl.lits. En una llei de Poisson(2.31), la probabilitat que tingui 0 satèl.lits és 0.0992, la probabilitat que tingui 1 satèl.lit és 0.2293 i la probabilitat que tingui més d'un satèl.lit és 0.6715. Els detalls es mostren a continuació:*

$$\mu_i = 2.31 \rightarrow Poisson(\mu_i = 2.31) \rightarrow P([Y = 0]) = \frac{\mu_i^0}{0!}\exp(-\mu_i) = \frac{2.31^0}{0!}\exp(-2.31) = 0.0993$$

$$P([Y = 1]) = \frac{\mu_i^1}{1!}\exp(-\mu_i) = \frac{2.31^1}{1!}\exp(-2.31) = 0.2293$$

$$P([Y > 1]) = 1 - P([Y \leq 1]) = 1 - P([Y = 1]) - P([Y = 0]) = 0.6715$$

```
> dpois(0,2.31)
[1] 0.09926125
> dpois(1,2.31)
[1] 0.2292935
> 1-ppois(1,2.31)
[1] 0.6714453
```

4. Es considera el model additiu, quin efectes nets són estadísticament significatius?

El model additiu considera totes les covariants pes, amplada i els factors spine i color (3 i 4 nivells, respectivament, que no es descriu en l'enunciat què volen dir). Els resultats facilitats inclouen el test d'efectes nets sobre el model additiu produït per Anova(m2) on al nivell de confiança habitual només el color i l'amplada són efectes nets significatius.

```
Analysis of Deviance Table (Type II tests)

Response: satel
        LR Chisq Df Pr(>Chisq)
color     9.2463  3   0.026190 *
spine     1.7984  2   0.406896
pes       0.1135  1   0.736183
amplada   9.0654  1   0.002605 **
```

5. Escriviu les equacions del millor model per resposta Poisson disponible i interpreteu l'efecte dels nivells del factor spine.

El millor model un cop que es consideren al model GRAN els efectes nets de totes les variables i les interaccions amb els 2 factors color i spine de les covariants, però no interaccions entre els factors i posteriorment s'usa el criteri d'Schwartz (BIC) per seleccionar el model amb menor BIC encaixat en el model GRAN, considera els efectes principals de l'amplada i spine i les interaccions. (m4) S'hauria d'escriure

$$\text{SPINE nivell i} = 1 \rightarrow \log(\mu_i) = \eta + \theta \cdot \text{amplada} = 0.447 + 0.3064\,\text{amplada}$$

$$\text{SPINE nivell i} = 2 \rightarrow \log(\mu_i) = (\eta + \alpha_2) + (\theta + \gamma_2) \cdot \text{amplada} =$$
$$= (0.447 - 1.5224) + (0.3064 + 0.4814)\,\text{amplada} = -1.075 + 0.788\,\text{amplada}$$

$$\text{SPINE nivell i} = 3 \rightarrow \log(\mu_i) = (\eta + \alpha_3) + (\theta + \gamma_3) \cdot \text{amplada} =$$
$$= (0.447 - 1.587) + (0.3064 + 0.560)\,\text{amplada} = -1.14 + 0.866\,\text{amplada}$$

És un model en l'escala del logaritme del valor esperat del nombre de satél.lits de rectes en diferents pendents per l'amplada.
Per l'spine 1 per cada unitat d'increment d'amplada el nombre esperat de satèl.lits es multiplica per exp(0.3064)=1.3585.
Per l'spine 2 per cada unitat d'increment d'amplada el nombre esperat de satèl.lits es multiplica per exp(0.788)=2.198.
Per l'spine 3 per cada unitat d'increment d'amplada el nombre esperat de satèl.lits es multiplica per exp(0. 886)=2.377.

6. El model triat en el punt 5 explica bé les dades? Indiqueu quina proposta de modelització us sembla adient amb els resultats disponibles. Estimeu el paràmetre de sobredispersió si penseu que les dades mostren aquesta característica.

La deviança residual del model (m4) és 545.12 unitats amb 167 graus de llibertat. El test de goodness of fit basat en la deviança residual i la seva distribució asimptòtica com una shi quadrat amb 167 g.ll. diria H0 "El model ajusta bé les dades" i el pvalor = P(Shi(167)>545.12)=0 per tant hi ha evidència per rebutjar la H0.

```
1-pchisq(m4$deviance,m4$df.residual)
[1] 0
```

De tota manera la distribució assimptòtica no és addient doncs les dades són individuals, amb la máxima desagregació. La regla pràctica en casos de dades desagregades a nivel d'individu que els graus de llibertat de la deviança residual i la magnitud de la deviança residual han de ser del mateix ordre de magnitud. En aquest cas 545.12 unitats a comparar amb 167 g.ll clarament no són del mateix ordre de magnitud.

```
> sum(resid(m4,"pearson")^2)
[1] 518.4218
> sum(resid(m4,"pearson")^2)/m4$df.residual
[1] 3.104322
```

Les dades de l'estadístic de Pearson incloses al llistat final, 518.42, que dividit entre 167 g.ll. dóna un estimador del factor de sobredispersió de 3.10.

$$\hat{\varphi} = \frac{X^2}{n-p} = 3.10$$

Els resultats del test de sobredispersió del paquet AER indica que l'alpha és més gran que 0. La sortida dona trafo=2 que vol dir la parametrització per la funció variança corresponent a la binomial negativa amb alpha 0.549951

$$V\left[Y_i \middle| X_i\right] = \mu_i + \alpha\mu_i^2 = \mu + \frac{1}{\theta}\mu^2 \qquad \alpha > 0$$

Clarament hi ha sobredispersió en les dades, per tant cal proposar una alternativa a la modelització Poisson, bé amb una versió no paramètrica que captura la sobredispersió i permet estimar els paràmetres del model per quasi-versemblança o bé la formulació com a resposta binomial negativa de la qual es disposa de resultats al final del problema.

7. Determineu el millor model segons la proposta binomial negativa i interpreteu quin és el nombre esperat de satèl.lits per una femella de 2.15 unitats d'amplada d'esquena, en els grups de referència dels factors ?

El millor model amb els resultats disponibles per la resposta binomial negativa és el que relaciona el logaritme del nombre esperat de satèl.lits amb l'amplada de l'esquena, és el model (m8) té un AIC de 752.64 un cop fixat el paràmetre theta de la binomial negativa usant el mètode ad-hoc glm.nb(). Amb aquest model el nombre esperat de satèl.lits per una femella d'amplada 2.15 seria:

$$\log(\mu_i) = \eta + \theta \cdot amplada = -0.8637 + 0.7599 amplada = -0.8637 + 0.7599 \cdot 2.15 = 0.77$$
$$\rightarrow \mu = \exp(0.77) = 2.16 \quad satèl.lits$$

La proposta de modelització de la resposta de Poisson conté el factor spine i la covariant amplada amb interaccions, mostra sobredispersió. En canviar a la modelització binomial negativa, ja no es fan necessàries ni les interaccions ni l'efecte principal del factor spine. De tota manera la deviança residual té una magnitud superior, 196.15, lleument als graus de llibertat (171).

8. Amb el llistat de resultats disponibles, quin contrast de models està ben efectuat? Per què?

   a. anova(m8,m7,test="F")
   b. anova(m8,m7,test="Chisq")

Clarament el test adient requereix considerar el paràmetre de dispersió de la modelització binomial negativa i per tant emprar un contrast basat en Fisher. Els dos models, m7 i m8, són encaixats i ambdós tenen resposta modelitzada binomial negativa. Per tant, cal emprar l'opció a.

RESULTATS PEL PROBLEMA 3

```
> m0<-glm(satel~1,family=poisson, data=df)
> m1<-glm(satel~amplada,family=poisson, data=df)
> summary(m1)

Call:
glm(formula = satel ~ amplada, family = poisson, data = df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.9306  -1.9981  -0.5627   0.9299   4.9992

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept)  -0.4282     0.1789  -2.394   0.0167 *
amplada       0.5892     0.0650   9.065   <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: 632.79  on 172  degrees of freedom
Residual deviance: 560.84  on 171  degrees of freedom
AIC: 920.14

Number of Fisher Scoring iterations: 5

> head(cbind(df$amplada,df$satel,m1$fitted))
  [,1] [,2]     [,3]
1 3.05    8 3.930859
2 2.60    4 3.015367
3 2.15    0 2.313093
4 1.85    0 1.938332
5 3.00    1 3.816746
6 2.30    3 2.526827
> anova(m0,m1,test="Chisq")
Analysis of Deviance Table

Model 1: satel ~ 1
Model 2: satel ~ amplada
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       172     632.79
2       171     560.84  1   71.949 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> m2<-glm(satel~color+spine+pes+amplada,family=poisson, data=df)
> Anova(m2)
Analysis of Deviance Table (Type II tests)

Response: satel
        LR Chisq Df Pr(>Chisq)
color     9.2463  3   0.026190 *
spine     1.7984  2   0.406896
pes       0.1135  1   0.736183
amplada   9.0654  1   0.002605 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> m3<-glm(satel~(color+spine)*(pes+amplada),family=poisson, data=df)
> m4<-step(glm(satel~(color+spine)*(pes+amplada),family=poisson,
data=df),k=log(nrow(df)))
Start:  AIC=974.14
….
Step:  AIC=931.33
satel ~ spine + amplada + spine:amplada

                Df Deviance    AIC
<none>               545.12 931.33
- spine:amplada  2   559.46 935.38
> summary(m4)

Call:
glm(formula = satel ~ spine + amplada + spine:amplada, family = poisson,
    data = df)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)     0.4470     0.3139   1.424 0.154477
spine2         -1.5224     0.8252  -1.845 0.065057 .
spine3         -1.5868     0.4325  -3.669 0.000244 ***
amplada         0.3064     0.1056   2.902 0.003712 **
spine2:amplada  0.4813     0.3317   1.451 0.146724
spine3:amplada  0.5596     0.1538   3.639 0.000274 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)
```

```
    Null deviance: 632.79  on 172  degrees of freedom
Residual deviance: 545.12  on 167  degrees of freedom
AIC: 912.41

> Anova(m4)
Analysis of Deviance Table (Type II tests)

Response: satel
              LR Chisq Df Pr(>Chisq)
spine            1.378  2  0.5021015
amplada         61.696  1  4.009e-15 ***
spine:amplada   14.349  2  0.0007659 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> AIC(m2,m3,m4)
   df      AIC
m2  8 920.8618
m3 18 917.3851
m4  6 912.4136
> BIC(m2,m3,m4)
   df      BIC
m2  8 946.0881
m3 18 974.1444
m4  6 931.3333

> sum(resid(m4,"pearson")^2)
[1] 518.4218
>library(AER)
> dispersiontest(m4,trafo=2)

        Overdispersion test

data:  m3
z = 3.9466, p-value = 3.964e-05
alternative hypothesis: true alpha is greater than 0 sample estimates:
    alpha
0.5139603

> library(MASS)
> summary(m5)

Call:
glm.nb(formula = satel ~ spine + amplada + spine:amplada, data = df,
    init.theta = 0.9643183766, link = log)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)     0.2730     0.7244   0.377    0.706
spine2         -2.0169     1.7221  -1.171    0.242
spine3         -1.4789     0.9079  -1.629    0.103
amplada         0.3695     0.2573   1.436    0.151
spine2:amplada  0.7178     0.7367   0.974    0.330
spine3:amplada  0.5229     0.3371   1.551    0.121

(Dispersion parameter for Negative Binomial(0.9643) family taken to be 1)

    Null deviance: 220.59  on 172  degrees of freedom
Residual deviance: 197.11  on 167  degrees of freedom
AIC: 759.98

              Theta:  0.964
          Std. Err.:  0.177
 2 x log-likelihood:  -745.985

> Anova(m5)
Analysis of Deviance Table (Type II tests)

Response: satel
              LR Chisq Df Pr(>Chisq)
spine           0.5435  2     0.7621
amplada        18.4075  1  1.784e-05 ***
spine:amplada   2.1525  2     0.3409
---
```

```
> step(m5)
Start:  AIC=757.98
satel ~ spine + amplada + spine:amplada

                Df Deviance    AIC
- spine:amplada  2   199.26 756.14
<none>               197.11 757.98

Step:  AIC=756.11
satel ~ spine + amplada

          Df Deviance    AIC
- spine    2   196.73 752.64
<none>         196.20 756.11
- amplada  1   214.23 772.14

Step:  AIC=752.64
satel ~ amplada

          Df Deviance    AIC
<none>         196.16 752.64
- amplada  1   216.44 770.92

Call:  glm.nb(formula = satel ~ amplada, data = df, init.theta = 0.9310998563,
    link = log)

Coefficients:
(Intercept)      amplada
    -0.8637       0.7599

Degrees of Freedom: 172 Total (i.e. Null);  171 Residual
Null Deviance:      216.4
Residual Deviance: 196.2       AIC: 754.6
> m6<-glm.nb(satel ~  amplada, data = df)
> summary(m6)

Call:
glm.nb(formula = satel ~ amplada, data = df, init.theta = 0.9310998551,
    link = log)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.8637     0.4046  -2.135   0.0328 *
amplada       0.7599     0.1578   4.817 1.46e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial(0.9311) family taken to be 1)

    Null deviance: 216.44  on 172  degrees of freedom
Residual deviance: 196.16  on 171  degrees of freedom
AIC: 754.64

              Theta:  0.931
          Std. Err.:  0.168

 2 x log-likelihood:  -748.643
> m7<-glm(satel ~  spine *amplada, family=neg.bin(theta=0.931),data = df)
> m8<-glm(satel ~  amplada, family=neg.bin(theta=0.931),data = df)
> summary(m8)

Call:
glm(formula = satel ~ amplada, family = neg.bin(theta = 0.931),
    data = df)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.8636     0.3764  -2.295    0.023 *
amplada       0.7598     0.1467   5.178 6.24e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Negative Binomial family taken to be 0.8651478)

    Null deviance: 216.42  on 172  degrees of freedom
```

```
Residual deviance: 196.15  on 171  degrees of freedom
AIC: 752.64
> anova(m8,m7,test="F")
Analysis of Deviance Table

Model 1: satel ~ amplada
Model 2: satel ~ spine * amplada
  Resid. Df Resid. Dev Df Deviance      F Pr(>F)
1       171     196.15
2       167     193.53  4   2.6213 0.7603 0.5525
> anova(m8,m7,test="Chisq")
Analysis of Deviance Table

Model 1: satel ~ amplada
Model 2: satel ~ spine * amplada
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       171     196.15
2       167     193.53  4   2.6213    0.551
```

PROBLEMA 2 (5 Punts):

Glei i Goldman (2000) han estudiat les diferències en assistència prenatal a embarassades en comunitats rurals de Guatemala. La taula il·lustra la classificació de 3334 embarassos segons les característiques ètniques de les dones (Indígenes pures, indígenes de parla espanyola i mestisses), la disponibilitat de serveis sanitaris moderns a menys d'una hora de camí de la comunitat on resideixen (no o si) i si les dones han visitat o no algun metge durant l'embaràs (no o si).

Ètnia	Disponibilitat de Serveis Mèdics Moderns	Ha visitat un Doctor ?	
		Si	**No**
Indígena Pura (Ref.)	No (Ref.)	**1**	106
	Si	**9**	339
Indígena Espanyola	No (Ref.)	**0**	38
	Si	**253**	1460
Mestisses	No (Ref.)	**4**	223
	Si	**252**	649
Total		**519**	**2815**

La variable de resposta dicotòmica és si ha visitat o no un doctor durant l'embaràs.

- El model nul té una deviança residual de 245.83.
- En ajustar un model amb el factor Ètnia, la deviança es redueix fins a 113.27 unitats. Els estimadors i errors estàndard són:

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	-3,7955	0,3198	-11,87	0,000			
Etnia							
IndSpa	2,0170	0,3269	6,17	0,000	7,52	3,96	14,26
Ladino	2,5699	0,3276	7,85	0,000	13,06	6,87	24,83

Log-Likelihood = -1375,398
Test that all slopes are zero: G = 132,565; DF = 2; P-Value = 0,000

- Els odds de visitar el doctor en una comunitat amb accés a serveis mèdics moderns són 15.4 vegades els odds corresponents a les comunitats sense accés. El contrast per la hipòtesi de odds ratio és 1, duu a un estadístic de canvi en la deviança de 95.75 amb 1 g.ll.

- El model logit additiu té una deviança de 2.79. Els estimadors i errors estàndard són:

Logistic Regression Table

Predictor	**Coef**	**SE Coef**	**Z**	**P**
Constant	**-6,4899**	**0,5517**	**-11,76**	**0,000**
Etnia				
IndSpa	**1,7829**	**0,3279**	**5,44**	**0,000**

```
 Ladino       2,5886     0,3290     7,87 0,000
SSan_apr
 si           2,9526     0,4564     6,47 0,000

Log-Likelihood = -1320,158
Test that all slopes are zero: G = 243,045; DF = 3; P-Value = 0,000
```

1. Quin és l'estimador de la constant en el model nul?

L'estimador de la constant en el model logit minimal és el logit de la proporció global de visitar un doctor.
Això és, log(519/2815)=logit(519/(519+2815))=logit(0.155) = - 1.69.

2. Contrasteu la hipòtesi que la probabilitat de visitar un doctor durant l'embaràs no depèn de l'ètnia de la dona. Ho podeu fer amb la informació disponible mitjançant l'estadístic de Wald?

Directament de la sortida del paquet estadístic Minitab, la G és la diferència en deviances entre el model NULL i el model actual, en aquest cas amb 2 g.ll. i comparable amb una shi quadrat de 2 g.ll: Test that all slopes are zero: G = 132,565; DF = 2; P-Value = 0,000.

La hipòtesi nul.la es rebutja i per tant, sí que hi ha evidència estadística per acceptar diferències segons l'ètnia. No es pot fer mitjançant l'estadístic de Wald amb la informació disponible, doncs només hi ha els estadístics de Wald pel contrast individual dels coeficients de les variables mudes i com a tals, no tenen cap validesa: les variables mudes introduïdes per permetre factors en els predictors lineals només poden contrastar-se conjuntament mitjançant un test de deviança (o Wald, però no es disposa de dades pel test simultani) abans i després de la seva introducció conjunta.

3. Interpreteu el coeficient de la variable muda relativa a dones mestisses (ladino) en el model d'efecte principal i únic l'ètnia.

El coeficient 2.5699 indica que els odds d'una dona mestissa (ladina) de visitar un doctor durant l'embaraç són exp(2.5699)=13.1 vegades els odds d'una dona índia de parla no espanyola. Aproximadament en l'escala de la probabilitat ser mestissa incrementa la probabilitat de visitar un metge en 0.155(1-0.155) 2.5699=0.337 respecte les índies de parla no espanyola.

4. Les dones índies de parla espanyola i les mestisses solen viure en comunitats amb més incidència de disponibilitat de serveis sanitaris moderns. Interpreteu el coeficient de la variable muda relativa a dones mestisses (ladino) en el model additiu. Contrasteu la hipòtesi que la probabilitat de visitar un doctor durant l'embaràs no depèn de l'ètnia de la dona.

El coeficient de la variable muda relativa a dones mestisses (ladino) en el model additiu és 2.5886 i per tant, els odds de visitar un doctor per una dona mestissa és exp(2.5886)=13.311 vegades els odds per una dona índia de parla no espanyola dins del mateix grup de disponibilitat de serveis sanitaris moderns.
El contrast sol.licitat és per la variació de la deviança entre D(Disponibilitat) i D(Etnia+Disponibilitat)=2.79, en canvi no es disposa de D(Disponibilitat). La G del model Etnia respecte el model nul val 95.75 i la G del model Etnia+Disponibilitat val 245.83, per tant la deviança residual del model amb només Disponibilitat és 245.83-95.75=150.08. Ara sí, D(Disponibilitat) -D(Etnia+Disponibilitat)=150.08-2.79=147.29 amb 2 g.ll, que contrastar amb una shi quadrat de 2 g.ll dona un p valor =0 i per tant, la hipòtesi nula es rebutja i un cop introduit els efectes de la Disponibilitat de medis sanitaris moderns, les diferències ètniques continuen essent estadísticament significatives.

5. Hi ha alguna evidència per afirmar que l'efecte de l'ètnia en l'odds de visitar un doctor durant l'embaràs depèn de la disponibilitat de serveis mèdics moderns a la comunitat de residència?

La qüestió fa referència a l'existència d'interacció entre Etnia i Disponibilitat. L'addició de la interacció saturaria el model i per tant tindria deviança nula, aleshores el contrast per la interacció és el contrast de la deviança residual amb la seva distribució de referència, una shi quadrat amb 2 g.ll., D(Etnia+Disponibilitat)=2.79 i el p valor és molt superior a 0.05, concluint que la hipòtesi nula s'accepta i per tant no hi ha evidència que l'efecte de Etnia en els odds de visitar un doctor durant l'embaràç depengui de la disponibilitat de medis sanitaris moderns.

6. Traslladeu a probabilitats, la probabilitat estimada de visitar un doctor en dones mestisses i índies de parla no espanyola segons el tipus de comunitat on resideixen, indicant les aproximacions que efectueu.

La probabilitat global, del conjunt de dones, de visitar un doctor durant l'embaràs és p=519/3334=0.1557. L'efecte marginal del canvi en la probabilitat per cadascuna de les variables en el predictor lineal depèn tant del coeficient de la variable com de la probabilitat puntual, que depèn de la resta de les variables també, però es pot aproximar pel producte del coeficient per p(1-p), en aquest cas es defineixen tots els valors de les variables del predictor lineal i per tant, no cal recórrer a aproximacions:

Disponib.	***Ètnia***	***Predictor lineal***		***Probabilitat***
NO	IndNoSpa	-6.4899	-6.4899	0.0015
NO	Mestissa	-6.4899+2.5866	-3.9033	0.0198
SI	IndNoSpa	-6.4899+2.9526	-3.5373	0.0283
SI	Mestissa	-6.4899+2.9526+2.5866	-0.9507	0.2787

Aleshores, en termes de probabilitat de visitar un doctor les diferències ètniques, mestisses vs indígenes no espanyoles és de (0.0198-0.0015)x100=1.83% en comunitats sense disponibilitat de serveis mèdics moderns i és d'un (0.2787-0.0283)x100=25% en comunitats AMB disponibilitat. Val la pena notar que el model additiu en l'escala logit comporta implícitament una interacció en l'escala de les probabilitats: l'efecte no és el mateix per tots dos tipus de comunitats de residència.

7. Construïu un interval de confiança al 95% pel quocient dels odds de visitar un doctor en dones mestisses, comparat amb dones índies de parla no espanyola en el model additiu.

Primer cal construir un interval de confiança en l'escala logit, la del predictor lineal, i després exponenciar per traslladar l'interval a l'escala dels odds.

$$2.5866 \pm 1.96 \cdot 0.3290 \rightarrow [1.9418,\ 3.2314] \rightarrow [\exp(1.9418),\ \exp(3.2314)] \rightarrow [6.97,\ 25.31]$$

Aleshores, amb un 95% de confiança els odds que una dona mestissa visiti un doctor durant l'embaràs es troben entre 7 i 25 vegades els odds corresponents a dones indígenes de parla no espanyola dins del mateix tipus de comunitat de residència.

8. Resumiu les conclusions de l'estudi de les presents dades en un paràgraf.

Hem après que hi han diferències grans en l'atenció durant l'embaràs segons l'ètnia de les dones, de manera que les dones mestisses (i també les indígenes de parla espanyola en menor intensitat) mostren una incidència molt més alta de visitar un doctor durant l'embaràs que les dones indígenes de parla no espanyola. També hi ha diferències entre els tipus de comunitat de residència depenent de si disposen de serveis mèdics moderns a menys d'una hora de camí o no. Les diferències ètniques es mantenen després de controlar per Disponibilitat, però no es troba evidència que l'efecte de l'ètnia sigui diferencial segons la disponibilitat de serveis mèdics.