

Feuille de TD 2 : Le modèle linéaire généralisé

Exercice 1

Vérifiez, dans chacun des cas suivants, que la loi de Y a une structure exponentielle (en précisant le paramètre naturel, le paramètre de dispersion, les fonctions b , γ et c , et la forme canonique si pertinent) et déduisez-en $\mathbb{E}[Y]$ et $\text{Var}(Y)$:

1. (loi binomiale) si $Y \sim \mathcal{B}(n, p)$,
2. (loi Gamma) si $Y \sim \Gamma(k, \lambda)$, i.e. Y a pour densité par rapport à la mesure de Lebesgue

$$f_{(k,\lambda)}(y) = \frac{y^{k-1} \lambda^k e^{(-\lambda y)}}{(k-1)!} \mathbb{1}_{y>0}.$$

Exercice 2

On s'intéresse dans cet exercice à la durée de vie de 17 patients atteints d'une leucémie. Le tableau suivant met en relation la durée de vie (variable Y_i exprimée en semaine depuis le dépistage de la maladie) et la variable x_i correspondant au \log_{10} du nombre initial de globules blancs.

Y_i	65	156	100	134	16	108	121	4	39
x_i	3.36	2.88	3.63	3.41	3.78	4.02	4	4.23	3.73

Y_i	143	56	26	22	1	1	5	65
x_i	3.36	2.88	3.63	3.41	3.78	4.02	4	4.23

1. On choisit de modéliser l'influence de x_i sur Y_i à travers la relation

$$\mathbb{E}[Y_i|x_i] = \exp(\beta_1 + \beta_2 x_i).$$

Quelle est la fonction de lien associée ?

2. On suppose que les Y_i suivent une loi exponentielle. Donnez l'expression du score et de la matrice d'information de Fisher.
3. Construisez un intervalle de confiance pour β_1 au niveau de confiance de 95%.

Exercice 3

On s'intéresse à la prédiction de présence de gel dans différents aéroports. Le jeu de données `gel` contient les variables suivantes :

- `gelobs` : vaut 1 en cas de gel, et 0 sinon,
- `nom` : nom de l'aéroport, à savoir "Ajaccio", "Marseille" ou "Rennes",
- `fllat` : flux de chaleur latente (en J/m²),
- `hu1000` : humidité relative à 1000m de hauteur,
- `hu2` : humidité relative à 2m de hauteur,
- `nh` : nébulosité des nuages hauts (à plus de 5000m de hauteur),
- `pmer` : pression au niveau de la mer (en hPa),
- `gradt` : gradient vertical de température entre 2m et 1000m,
- `tprev` : température prévue (en degrés C), par un modèle déterministe,
- `ff` : force du vent (en m/s),
- `ddf` : direction du vent.

On souhaite expliquer la présence/absence de gel en fonction des autres variables. Dues à de grandes différences d'échelles, les variables `fllat` et `pmer` ont été centrées et réduites. Voici un tableau récapitulatif des données :

```
> summary(gel)
```

gelobs	nom	fllat	hu1000	hu2	nh
0:1514	AJACCIO :545	Min. : -9.935614	Min. : 2.814	Min. : 14.70	Min. : 0.000
1: 120	RENNES :545	1st Qu.: -0.029809	1st Qu.: 41.205	1st Qu.: 76.94	1st Qu.: 0.000
	MARSEILLE:544	Median : 0.325996	Median : 63.662	Median : 88.99	Median : 1.468
		Mean : 0.000000	Mean : 62.033	Mean : 84.38	Mean : 2.894
		3rd Qu.: 0.460250	3rd Qu.: 83.619	3rd Qu.: 94.93	3rd Qu.: 5.960
		Max. : 1.517546	Max. : 100.008	Max. : 100.01	Max. : 8.000

pmer	gradt	tprev	ff	ddf
Min. : -3.642992	Min. : -9.502	Min. : -8.217	Min. : 0.06634	C:216
1st Qu.: -0.596667	1st Qu.: -4.066	1st Qu.: 3.494	1st Qu.: 1.58732	E:589
Median : 0.103044	Median : -1.665	Median : 6.727	Median : 2.60929	N:389
Mean : 0.000000	Mean : -1.117	Mean : 6.730	Mean : 3.27396	S:227
3rd Qu.: 0.687446	3rd Qu.: 1.332	3rd Qu.: 9.927	3rd Qu.: 4.11367	W:213
Max. : 2.616971	Max. : 12.389	Max. : 20.848	Max. : 18.26777	

1. Quel est le modèle préconisé pour faire cette étude ? Justifiez votre réponse.
2. Dans cette question, on s'intéresse au modèle `mod.tprev` ne prenant en compte que la température prévue grâce au modèle déterministe (à savoir, la variable `tprev`).
 - (a) Écrivez l'équation de ce modèle `mod.tprev`.
 - (b) L'effet de la température prévue est-il significatif ?
 - (c) Que signifie "Number of Fisher Scoring iterations: 9" dans la sortie R ?
 - (d) Comment peut-on interpréter $e^{-1.2149} \simeq 0.3$?
3. Dans cette question, on souhaite étudier l'influence de la direction de vent sur la présence/absence de gel.
 - (a) Ecrivez l'équation traduisant le modèle `mod.ddf`
 - (b) Testez si l'effet de la direction du vent est significative.
 - (c) Dans la sortie `exp(mod.ddf$coefficients)`, que représentent les quantités 0.04347826 et 4.52923077 ?

4. On va maintenant considérer le modèle `mod.complet` considérant toutes les variables explicatives, ainsi qu'un terme d'interaction entre l'humidité à 1000m, et la nébulosité haute.

- (a) Écrivez l'équation du modèle `mod.complet`. Quels sont les résultats de significativité des paramètres ?
- (b) À partir de ce modèle, on utilise une procédure pas-à-pas pour sélectionner les variables, grâce à la commande R

```
mod.AIC <- stepAIC(mod.complet,direction="backward")
```

- i. Expliquez en quoi consiste cette procédure de sélection de variables.
- ii. Est-il possible de comparer les modèles `mod.complet` et `mod.AIC` ?
Au vu des résultats ci-dessous, qu'en déduisez-vous ?

5. Dans cette question, on considère le modèle `mod.AIC` de la question précédente.

- (a) Estimez les odds-ratios entre les aéroports d'Ajaccio et de Marseille, puis entre Marseille et Rennes. Qu'en déduisez-vous ?
- (b) On souhaite prédire, pour chaque aéroport, la probabilité qu'il y ait du gel les jours où l'humidité à 2m est de 80, où le vent souffle à 3m/s, que la température prévue est de -1 degrés C, que la nébulosité des nuages hauts vaut 2 et que la pression au niveau de la mer est nulle.
 - i. Comment prédiriez-vous ces probabilités ?
 - ii. En vous basant sur les sorties R, que concluez-vous ?

```
summary(gel)
##      gelobs      nom      fillat      hu1000      hu2
## 0:1514  AJMCC10:1545  Min.: -3698947  Min.:  2.814  Min.: 14.70
## 1: 120  REMRES :1545  1st Qu.: -132366  1st Qu.: 41.285  1st Qu.: 76.94
##      Mean: -124639  Mean:  62.033  Mean:  84.38
##      3rd Qu.: -5342  3rd Qu.: 83.619  3rd Qu.: 94.83
##      Max.:  268709  Max.: 100.008  Max.: 100.01
##
##      ch      pmax      g1      g2
## Min.: 0.000  Min.: -3.6430  Min.: -9.502  Min.: -9.217
## 1st Qu.: 0.000  1st Qu.: -0.5967  1st Qu.: -4.066  1st Qu.: 3.494
## Median: 0.000  Median: -1.0000  Median: -1.067  Median:  6.727
## Mean: 0.000  Mean: -0.6900  Mean: -0.677  Mean:  0.727
## 3rd Qu.: 0.960  3rd Qu.: 0.6674  3rd Qu.: 1.332  3rd Qu.: 9.927
## Max.: 18.000  Max.:  2.6170  Max.: 12.389  Max.: 20.848
##
##      if      of      of2
## Min.: 0.06634  0.216  0.000
## 1st Qu.: 1.89732  0.589  0.000
## Median: 2.60929  0.389  0.000
## Mean: 3.27367  0.277  0.000
## 3rd Qu.: 4.27367  0.277  0.000
## Max.: 18.28777  0.277  0.000
```

Modele mod.tprev

```
mod.tprev <- glm(gelobs ~ tprev, data=gel, family=binomial(link="logit"))
summary(mod.tprev)
##      Call:
## glm(formula = gelobs ~ tprev, family = binomial(link = "logit"),
##      data = gel)
##
## Deviance Residuals:
##      -3.09746   -0.07539   -0.01059   -0.00131    2.80709
##
## Coefficients:
## (Intercept)  -0.4279  -2.591  0.00958 **
## tprev        -1.2149   0.1115 -10.892 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 857.67 on 1633 degrees of freedom
## Residual deviance: 263.82 on 1632 degrees of freedom
## AIC: 267.82
##
## Number of Fisher Scoring iterations: 9
```

Modele mod.ddf

```
mod.ddf <- glm(gelobs ~ ddf, data=gel, family=binomial(link="logit"))
summary(mod.ddf)
##      Call:
## glm(formula = gelobs ~ ddf, family = binomial(link = "logit"),
##      data = gel)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.5996 -0.3500 -0.3500 -0.2391  2.6956
##
## Coefficients:
## (Intercept)  Estimate Std. Error z value Pr(>|z|)
## ddf         0.3737    0.3825   0.977   0.329 ***
## ddfn        1.5108    0.3669   4.117 3.84e-05 ***
## ddfS        -0.4709    0.5358  -0.879   0.380
## ddfW        -0.4055    0.5361  -0.756   0.449
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 857.67 on 1633 degrees of freedom
## Residual deviance: 798.26 on 1629 degrees of freedom
## AIC: 808.26
##
## Number of Fisher Scoring iterations: 6
anova(mod0,mod.ddf,test="Chisq")
##      Analysis of Deviance Table
##
## Model 1: gelobs ~ 1
## Model 2: gelobs ~ ddf
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      1633      857.67
## 2      1629      798.26  4  59.413 3.854e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
exp(mod.ddf$coefficients)
## (Intercept)      ddf      ddfn      ddfS      ddfW
## 0.04347826  1.45306559  4.52923077  0.62443439  0.66666667
```

Modèle mod.complet

```
mod.complet <- glm(gelobs ~ . + nh*hu1000, data = gel, family = binomial(link="logit"))
summary(mod.complet)

##
## Call:
## glm(formula = gelobs ~ . + nh * hu1000, family = binomial(link = "logit"),
##      data = gel)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.08750 -0.06587 -0.00728 -0.00071  2.84622
##
## Coefficients:
##      (Intercept)  3.138e+00  1.676e+00  1.873  0.061031 .
## nomMENDUS      1.043e+00  7.364e-01  1.416  0.167394
## nomMARSEILLE   2.446e+00  7.366e-01  3.325  0.000863 ***
## flilat         -6.620e-07  1.975e-06 -0.335  0.797520
## hu1000         2.128e-03  1.178e-02  0.181  0.856536
## hu2            -5.715e-02  1.925e-02 -2.968  0.002995 **
## nh             -3.280e-02  1.760e-01 -0.471  0.637975
## pmer           -3.996e-01  1.841e-01 -2.170  0.030001 *
## gradt1         7.488e-02  7.435e-02  1.007  0.313907
## tprev          -1.226e+00  1.226e-01 -10.004 < 2e-16 ***
## ff            -2.716e-01  1.516e-01 -1.791  0.073227 .
## ddfE           5.689e-01  7.073e-01  0.804  0.421194
## ddfn           1.243e+00  7.798e-01  1.593  0.111057
## ddfs           4.099e-01  9.418e-01  0.435  0.663374
## ddfw           2.135e-01  8.800e-01  0.243  0.808351
## hu1000.nh      -5.844e-04  2.599e-03 -0.225  0.822126
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 857.67  on 1633  degrees of freedom
## Residual deviance: 234.18  on 1618  degrees of freedom
## AIC: 266.18
##
## Number of Fisher Scoring iterations: 9
```

3

Modèle mod.AIC

```
library(MASS)
mod.AIC <- stepAIC(mod.complet, direction="backward")

## Start:  AIC=266.18
## gelobs ~ nom + flilat + hu1000 + hu2 + nh + pmer + gradt1 + tprev +
## ff + ddf + nh * hu1000
##
##              Df Deviance   AIC
## - ddf          4    238.52 262.52
## - hu1000.nh    1    234.23 264.23
## - flilat       1    234.28 264.28
## - gradt1      1    235.19 265.19
## <none>         1    234.18 266.18
## - ff          1    237.46 267.46
## - pmer        1    238.93 268.93
## - hu2         1    243.06 273.06
## - nom         2    247.78 275.78
## - tprev       1    689.30 719.30
##
## Step: AIC=262.52
## gelobs ~ nom + flilat + hu1000 + hu2 + nh + pmer + gradt1 + tprev +
## ff + hu1000.nh
##
##              Df Deviance   AIC
## - flilat      1    238.53 260.53
## - hu1000.nh   1    238.83 260.83
## - gradt1      1    239.10 261.10
## <none>         1    238.52 262.52
## - ff          1    240.58 262.58
## - pmer        1    243.01 265.01
## - hu2         1    249.92 271.92
## - nom         2    255.83 275.83
## - tprev       1    724.85 746.85
##
## Step: AIC=260.53
## gelobs ~ nom + hu1000 + hu2 + nh + pmer + gradt1 + tprev + ff +
## hu1000.nh
##
##              Df Deviance   AIC
## - hu1000.nh   1    238.83 258.83
## - gradt1      1    239.10 259.10
## <none>         1    238.53 260.53
## - ff          1    242.63 262.63
## - pmer        1    243.01 263.01
## - hu2         1    249.94 269.94
## - nom         2    258.17 276.17
## - tprev       1    726.27 746.27
##
## Step: AIC=258.83
## gelobs ~ nom + hu1000 + hu2 + nh + pmer + gradt1 + tprev + ff
##
##              Df Deviance   AIC
## - hu1000      1    238.91 256.91
## - gradt1      1    239.24 257.24
## <none>         1    238.83 258.83
## - nh          1    242.43 260.43
```

4

```
## - pmer 1 243.41 261.41
## - ff 1 243.87 261.87
## - hu2 1 250.15 268.15
## - nom 2 259.15 275.15
## - tprev 1 727.50 745.50
##
## Step: AIC=256.91
## gelobs ~ nom + hu2 + nh + pmer + gradt1 + tprev + ff
##
## Df Deviance AIC
## - gradt1 1 240.11 256.11
## <none> 238.91 256.91
## - nh 1 242.74 258.74
## - pmer 1 243.42 259.42
## - ff 1 243.94 259.94
## - hu2 1 251.36 267.36
## - nom 2 259.42 273.42
## - tprev 1 729.82 745.82
##
## Step: AIC=256.11
## gelobs ~ nom + hu2 + nh + pmer + tprev + ff
##
## Df Deviance AIC
## <none> 240.11 256.11
## - pmer 1 243.78 257.78
## - nh 1 244.05 258.05
## - ff 1 247.18 261.18
## - hu2 1 251.59 265.59
## - nom 2 259.86 271.86
## - tprev 1 767.27 781.27
summary(mod.AIC)

## Call:
## glm(formula = gelobs ~ nom + hu2 + nh + pmer + tprev + ff, family = binomial(link = "logit"),
## data = gel)
##
## Deviance Residuals:
## -3.09483 -0.06076 -0.00776 -0.00082 2.74307
##
## Coefficients:
## (Intercept) 3.85700 1.48197 2.603 0.009252 **
## nomRENNES 1.84227 0.58865 3.130 0.001750 ***
## nomMARSEILLE 2.69060 0.65375 4.116 3.86e-05 ***
## hu2 -0.06055 0.01790 -3.383 0.000718 ***
## nh -0.12901 0.06679 -1.931 0.053424 .
## pmer -0.33018 0.17330 -1.905 0.056739 .
## tprev -1.23926 0.12056 -10.279 < 2e-16 ***
## ff -0.22897 0.08975 -2.551 0.010739 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 857.67 on 1633 degrees of freedom
## Residual deviance: 240.11 on 1626 degrees of freedom
```

```
## AIC: 256.11
##
## Number of Fisher Scoring iterations: 9
anova(mod.AIC,mod.complet,test="ChiSq")
##
## Analysis of Deviance Table
##
## Model 1: gelobs ~ nom + hu2 + nh + pmer + tprev + ff
## Model 2: gelobs ~ nom + flrat + hu1000 + hu2 + nh + pmer + gradt1 + tprev +
## ff + ddf + nh + hu1000
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1 1626 240.11
## 2 1618 234.18 8 5.9377 0.6542
new.data <- data.frame(nom=c("AJACCIO", "MARSEILLE", "RENNES"), hu2=80, tprev=-1, ff=3, nh=2, pmer=0)
predict(mod.AIC,new.data,type="response")
## 1 2 3
## 0.3333717 0.8805480 0.7593815
```

Exercice 4

On s'intéresse à l'étude d'une population de limules vivant sur la côte Est des Etats-Unis. Jane Brockmann a réuni une base de données afin d'étudier la reproduction des limules. Généralement une femelle est entourée d'un mâle et autour de ce couple gravitent d'autres mâles appelés satellites. Le jeu de données est composé de cinq variables :

- "couleur" : 2=clair moyen, 3= moyen, 4=foncé moyen, 5= foncé.
- "spine" : 1= both good, 2=one worm or broken, 3=both worn or broken
- "width" correspond à la largeur de la carapace en cm
- "weight" correspond au poids en grammes
- "satellite" donne le nombre de satellites.

1. Dans un premier temps, proposez un modèle (`model`) pour étudier le nombre de satellites en fonction de la largeur de la carapace. Comment obtient-on la valeur ajustée pour le nombre de satellites pour la i ème limule étudiée ? Comment peut-on interpréter $e^{0.16405}$?
2. Construisez un intervalle de confiance asymptotique pour un des paramètres du modèle `model` au niveau de confiance 95%.
3. Ecrivez les modèles correspondant à `modelcomplet` et `model1` dans les sorties ci-dessous.
4. Au vu des différentes sorties obtenues avec la fonction `anova`, quel modèle privilégié parmi les trois modèles considérés ?
5. Même question à partir des sorties de la commande `step(modelcomplet)`.

```
> model=glm(satellite ~ width,data=limules, family=poisson(link=log))
Call:
glm(formula = satellite ~ width, family = poisson(link = log),
    data = limules)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8526  -1.9884  -0.4933   1.0970   4.9221
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.30476     0.54224  -6.095 1.1e-09 ***
width        0.16405     0.01997   8.216 < 2e-16 ***
---
Null deviance: 632.79  on 172  degrees of freedom
Residual deviance: 567.88  on 171  degrees of freedom
AIC: 927.18

> modelcomplet=glm(satellite ~ .,data=limules, family=poisson(link=log))
> model1=glm(satellite ~ width+couleur, data=limules,
             family=poisson(link=log))

> anova(model,model1,test="Chisq")
Analysis of Deviance Table
Model 1: satellite ~ width
Model 2: satellite ~ width + couleur
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       171      567.88
2       168      559.34  3   8.5338  0.03618 *

> anova(model1,modelcomplet,test="Chisq")
Analysis of Deviance Table
Model 1: satellite ~ width + couleur
Model 2: satellite ~ couleur + spine + width + weight
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       168      559.34
2       165      549.59  3   9.7592  0.02073 *

> anova(model,modelcomplet,test="Chisq")
Analysis of Deviance Table
Model 1: satellite ~ width
Model 2: satellite ~ couleur + spine + width + weight
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       171      567.88
2       165      549.59  6  18.293  0.00554 **

> step(modelcomplet)
Start: AIC=920.88
satellite ~ couleur + spine + width + weight

            Df Deviance   AIC
- spine      2   551.38 918.68
- width      1   549.70 919.00
<none>                      549.59 920.88
- couleur    3   558.83 924.12
- weight     1   558.63 927.93

Step: AIC=918.68
satellite ~ couleur + width + weight

            Df Deviance   AIC
- width      1   551.80 917.10
<none>                      551.38 918.68
- couleur    3   559.90 921.20
- weight     1   559.34 924.64

Step: AIC=917.1
satellite ~ couleur + weight

            Df Deviance   AIC
<none>                      551.80 917.10
- couleur    3   560.87 920.16
- weight     1   609.14 972.44

Call: glm(formula = satellite ~ couleur + weight, family = poisson(link = log),
    data = limules)

Coefficients:
(Intercept)    couleur3    couleur4    couleur5      weight
-0.0497806   -0.2051125   -0.4497964   -0.4520498    0.0005462

Degrees of Freedom: 172 Total (i.e. Null); 168 Residual
Null Deviance: 632.8
Residual Deviance: 551.8 AIC: 917.1
```

Exercice 5

On s'intéresse dans cet exercice aux choix alimentaires des alligators dans quatre étangs de Floride. Les scientifiques ont classé le contenu des estomacs de 219 alligators capturés en 5 catégories : poisson, invertébrés, reptile, oiseau et autres. Pour chaque alligator, on connaît sa provenance (Lake : Hancock, Oklawaha, Trafford, George), son sexe (Gender : male (m) et femelle (f)), sa taille (Size : petit (< 2.3), grand (> 2.3)). Les données sont rassemblées dans le tableau suivant :

Primary Food Choice							
Lake	Gender	Size	Fish	Invertebrate	Reptile	Bird	Other
Hancock	Male	≤2.3	7	1	0	0	5
		>2.3	4	0	0	1	2
	Female	≤2.3	16	3	2	2	3
		>2.3	3	0	1	2	3
Oklawaha	Male	≤2.3	2	2	0	0	1
		>2.3	13	7	6	0	0
	Female	≤2.3	3	9	1	0	2
		>2.3	0	1	0	1	0
Trafford	Male	≤2.3	3	7	1	0	1
		>2.3	8	6	6	3	5
	Female	≤2.3	2	4	1	1	4
		>2.3	0	1	0	0	0
George	Male	≤2.3	13	10	0	2	2
		>2.3	9	0	0	1	2
	Female	≤2.3	3	9	1	0	1
		>2.3	8	1	0	0	1

La réponse est ici le choix alimentaire prépondérant qui prend 5 modalités. Nous allons donc mettre en place un modèle de régression multinomiale (ou appelée aussi polytomique non ordonnée) qui généralise la régression logistique.

Dans la suite, on note π_m la probabilité pour un alligator de choisir de façon prépondérante la nourriture m ($m \in \{\text{"oiseau"}, \text{"invertébré"}, \text{"reptile"}, \text{"autres"}, \text{"poisson"}\} = \{1, 2, 3, 4, 5\}$), et on a $\sum_{m=1}^5 \pi_m = 1$.

Comme le choix alimentaire prédominant est le poisson on va considérer cette catégorie comme référence. On modélise alors linéairement les quantités

$$\forall m \in \{1, 2, 3, 4\}, \ln \left(\frac{\pi_m}{\pi_5} \right) = X\theta^{(m)}$$

- Détaillez l'expression de $X\theta^{(m)}$ en fonction des variables explicatives. Vous pouvez vous aider de la sortie de R en fin d'exercice.
- Quel test est proposé dans la dernière colonne de `summary(fit5)`? Qu'en concluez-vous?
- On considère dans le tableau suivant plusieurs modèles utilisant les covariables Lake, Gender et Size. Pour chacun, on donne la valeur de la déviance. Complétez la dernière colonne donnant le nombre de degrés de liberté (ddl) pour chacun de ces modèles. Quel modèle doit-on préférer ?

Modèle	Déviance	ddl
Lake + Size + Lake x Size	35.40	
Lake + Size + Gender	50.26	
Lake + Size	52.48	
Lake	73.57	
Size	101.61	
Gender	114.66	
Null	116.76	


```

> library(VGAM)
> fit5<-vglm(cbind(Bird,Invertebrate,Reptile,Other,Fish)
~Lake+Size+Gender,data=gator,family=multinomial)
> summary(fit5)

Call:
vglm(formula = cbind(Bird, Invertebrate, Reptile, Other, Fish) ~
      Lake + Size + Gender, family = multinomial, data = gator)

Pearson residuals:
      Min      1Q   Median      3Q     Max
log(mu[,1]/mu[,5]) -1.1985 -0.5478 -0.22421 0.3678 3.478
log(mu[,2]/mu[,5]) -1.3218 -0.4611  0.01054 0.3810 1.866
log(mu[,3]/mu[,5]) -0.7033 -0.5751 -0.35511 0.2610 2.064
log(mu[,4]/mu[,5]) -1.6945 -0.2893 -0.10807 1.1236 1.367

Coefficients:
      Estimate Std. Error z value Pr(>|z|)
(Intercept):1 -2.43211    0.77066  -3.156  0.00160 **
(Intercept):2  0.16902    0.37875   0.446  0.65541
(Intercept):3 -3.41604    1.08513  -3.148  0.00164 **
(Intercept):4 -1.43073    0.53809  -2.659  0.00784 **
Lakehancock:1  0.57527    0.79522   0.723  0.46943
Lakehancock:2 -1.78051    0.62321  -2.857  0.00428 **
Lakehancock:3  1.12946    1.19280   0.947  0.34369
Lakehancock:4  0.76658    0.56855   1.348  0.17756
Lakeoklawaha:1 -0.55035    1.20980  -0.455  0.64917
Lakeoklawaha:2  0.91318    0.47612   1.918  0.05511 .
Lakeoklawaha:3  2.53026    1.12211   2.255  0.02414 *
Lakeoklawaha:4  0.02606    0.77776   0.034  0.97327
Laketrafford:1  1.23699    0.86610   1.428  0.15322
Laketrafford:2  1.15582    0.49279   2.345  0.01900 *
Laketrafford:3  3.06105    1.12972   2.710  0.00674 **
Laketrafford:4  1.55776    0.62567   2.490  0.01278 *
Size>2.3:1      0.73024    0.65228   1.120  0.26292
Size>2.3:2     -1.33626    0.41119  -3.250  0.00116 **
Size>2.3:3      0.55704    0.64661   0.861  0.38898
Size>2.3:4     -0.29058    0.45993  -0.632  0.52751
Genderm:1      -0.60643    0.68884  -0.880  0.37867
Genderm:2      -0.46296    0.39552  -1.171  0.24180
Genderm:3      -0.62756    0.68528  -0.916  0.35978
Genderm:4      -0.25257    0.46635  -0.542  0.58810
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of linear predictors: 4

Names of linear predictors:
log(mu[,1]/mu[,5]), log(mu[,2]/mu[,5]), log(mu[,3]/mu[,5]), log(mu[,4]/mu[,5])

Residual deviance: 50.2637 on 40 degrees of freedom

Log-likelihood: -73.3221 on 40 degrees of freedom

Number of iterations: 5

Reference group is level 5 of the response

```