Exercice 1

chap. 3 p.14 :

Calcul de la vraisemblance.

Poser le calcul de la vraisemblance dans le cas suivant : 4 lots de 20 insectes sont soumis à un insecticide (dose 1, 2, 4 et 8 respectivement). On dénombre les insectes morts dans chaque lot. On trouve 1, 3, 9, 10 respectivement.

Dose	Décès	Total
1	1	20
2	3	20
4	9	20
8	10	20

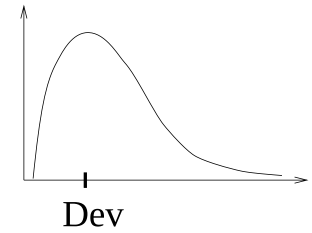
Exercice 2

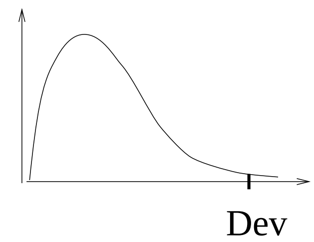
chap. 3 p.17 :

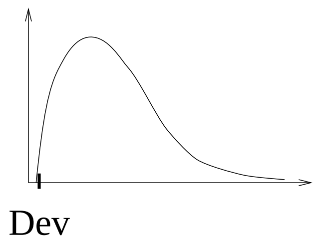
Loi de la déviance.

On étudie l'effet d'un insecticide sur des guêpes. On constitue 10 groupes de 12 guêpes. On applique une dose différente à chacun des groupes, et on observe le nombre de guêpes survivantes dans chacun des groupes. On choisit le lien logit, le logarithme de la dose comme régresseur, et on ajuste un modèle de régression logistique, à deux paramètres, la pente et l'ordonnée à l'origine.

Quel est le nombre de degrés de liberté résiduels ? La déviance obtenue est reportée sur le graphe présentant la distribution d'un χ^2 dont le nombre de degrés de liberté est égal au nombre de degrés de liberté résiduels. Trois cas de figure caricaturaux sont proposés. Dans chacun des cas de figure, que concluez-vous ?







Exercice 3

chap. 3. p.20 :

Test du rapport de vraisemblance.

Exemple “Aircraft”. L’exemple est traité avec SAS et Splus, dans le document “Collett, mise en œuvre avec SAS et Splus”, chapitre 3, Models for binary and binomial data.
A partir des sorties numériques de SAS ou de Splus, représenter sur un schéma le modèle nul, le modèle saturé, le modèle estimé, les déviances, et les degrés de liberté associés. Réaliser un test d’adéquation du modèle, puis un test d’explication. Conclusion ?

```
> summary(glm1)
Call: glm(formula = cbind(k, n - k) ~ load, family = binomial, data = aircraft)
Deviance Residuals:
       Min         1Q     Median         3Q       Max
 -0.294752 -0.1112901 0.04161701 0.08846694 0.3501585

Coefficients:
                   Value   Std. Error   t value
(Intercept) -5.339711512 0.5456932048 -9.785190
       load  0.001548434 0.0001575379  9.828958

(Dispersion Parameter for Binomial family taken to be 1 )

    Null Deviance: 112.8321 on 9 degrees of freedom

Residual Deviance: 0.3719169 on 8 degrees of freedom

Number of Fisher Scoring Iterations: 3

Correlation of Coefficients:
     (Intercept)
load -0.9884009
```

Exercice 4

chap. 3 p.22 :

Données groupées et non groupées.

exemple “cancer de la prostate”. (Collett, mise en œuvre avec SAS et Splus, chap. 3, Models for binary and binomial data, a futher example on model selection.)

“The data are from the observational study on nodal involvement in patients presenting with prostatic cancer. In the data base there are observations on six variables for each of 53 patients, the variables where as follows.

Age :	age of patient in years,
Acid :	level of serum acid phosphatase,
X-ray :	result of an X-ray examination (0=negative, 1=positive), ´’
Size :	size of tumour (0 = small, 1= large),
Grade :	grade of tumour (0=less serious, 1= more serious),
Nodes :	nodal involvement (0 = no involvement, 1= involvement).

Pour chacun des modèles ci-dessous, dire s’il s’agit de données groupées ou pas. (on suppose que l’on regroupe les données chaque fois que le modèle le permet).

Modèle :	Xray
Modèle :	Xray + Size
Modèle :	Xray + Size + Xray . Size
Modèle :	Xray + Size + Xray . Size + Age
Modèle :	Age
Modèle :	cut(Age,4)

Remarque : cut(Age,4) signifie que l’on découpe la variable age en 4 classes de même effectif.

Exercice 5

chap. 3 p.23 :

Interprétation du terme constant

Exemple “Aircraft”. L’exemple est traité avec SAS et Splus, dans le document “Collett, mise en œuvre avec SAS et Splus”, chapitre 3, Models for binary and binomial data.
Dans le cas des deux paramétrisations suivantes du même modèle, quelle est l’hypothèse nulle correspondant à la nullité du terme constant ?

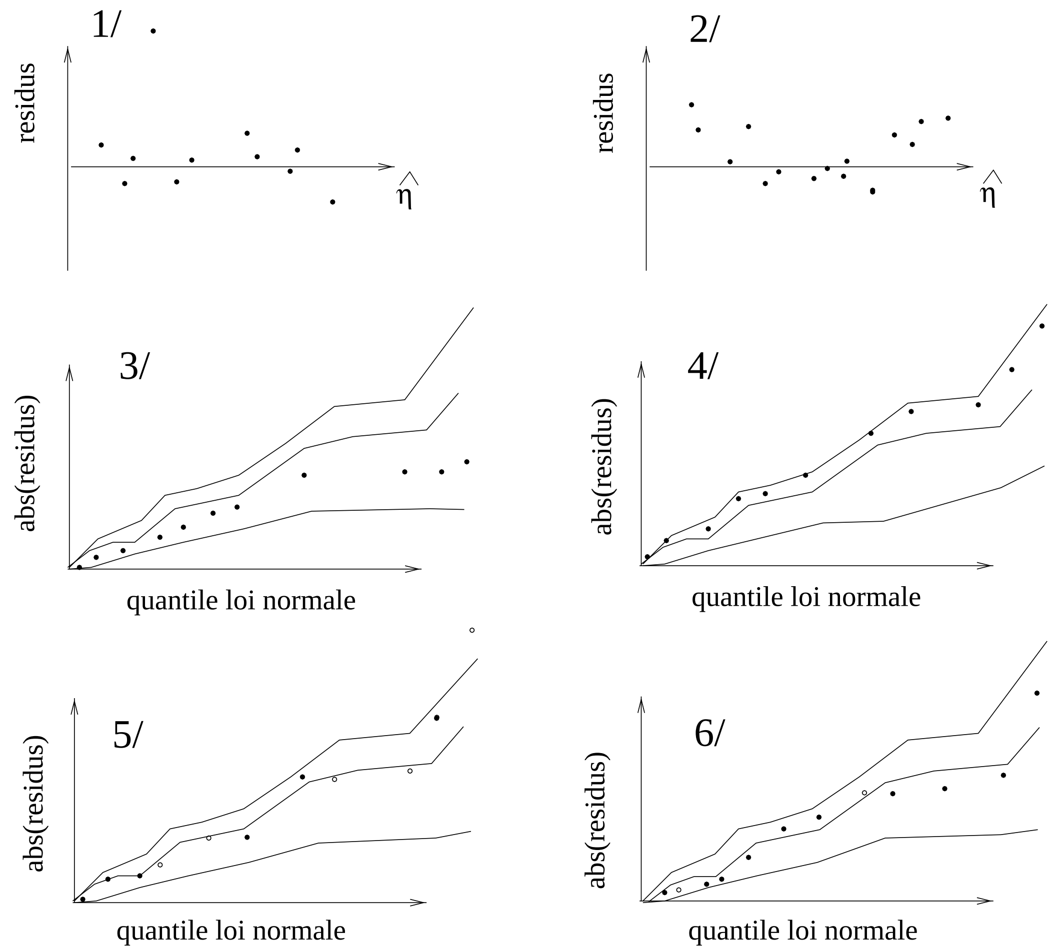
1. $\text{logit}(p) = \alpha + \beta \, \text{load}$
2. $\text{logit}(p) = \alpha' + \beta \, (\text{load} - \overline{\text{load}})$

Exercice 6

chap. 4 p.13 :

graphe des résidus

Comment interpréter chacun des graphes suivants ? proposer une solution, une amélioration du modèle le cas échéant.

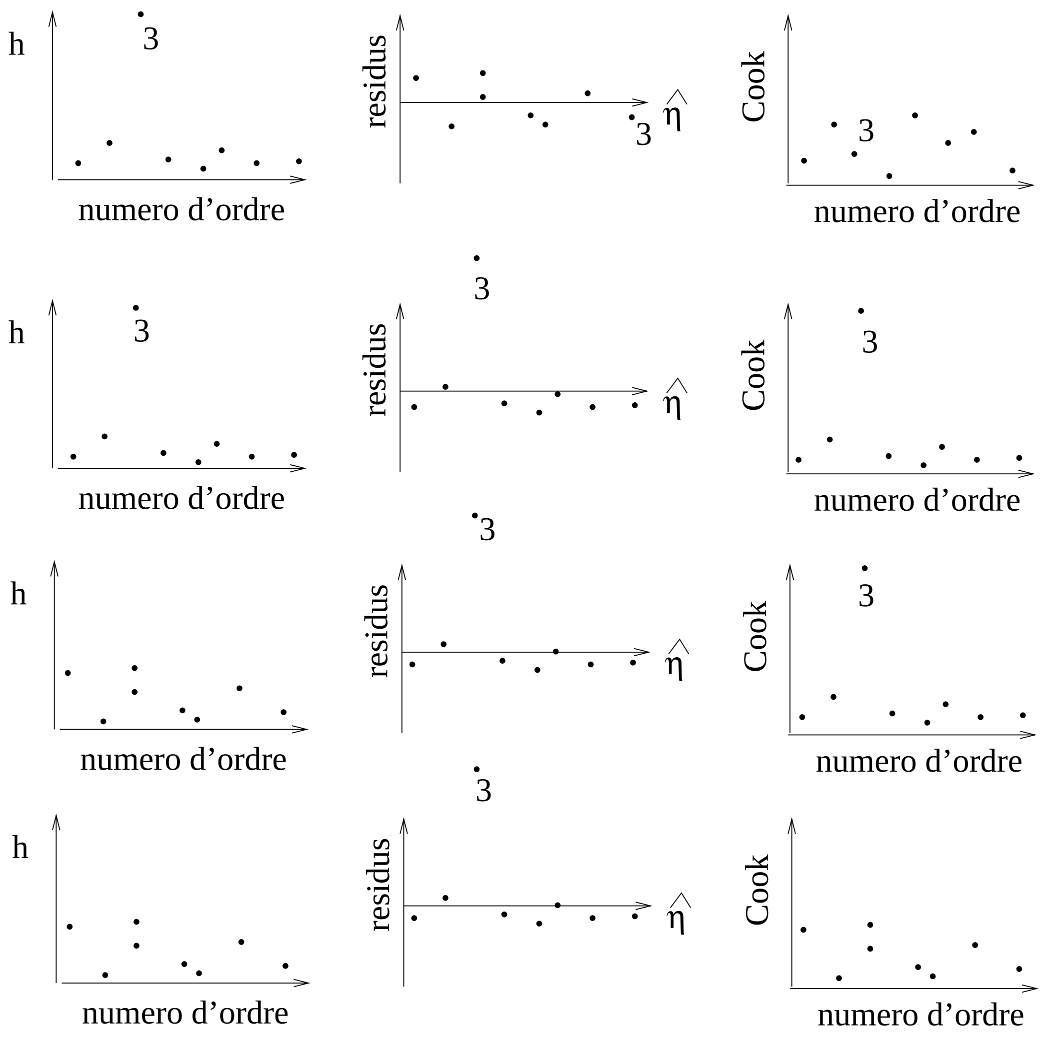


Exercice 7

à la fin du chapitre 4 :

Points influents.

Quel diagnostic, pour chacun des groupes de 3 graphes ?



Exercice 8

à la fin du chapitre 4 :

Choix du régresseur.

Dans l'exemple "tournesol"[2] qui suit, choisir entre dose et son logarithme comme régresseur.

On étudie l' effet d'un traitement fongicide sur deux races de mildiou, champignon parasite du tournesol. La race R est nouvelle, mutante, elle présente une résistance au traitement fongicide.
Le fongicide est déposé sur les graines de tournesol, à différentes doses. Des groupes de 7 plantules sont inoculées avec l'une des races de mildiou. On compte le nombre de plantules présentant des symptômes de mildiou, c'est le nombre de plantules pour lesquelles le champignon a sporulé. Il sporule ou ne sporule pas, selon la plantule, sa race, et la dose de fongicide (metalaxyl).

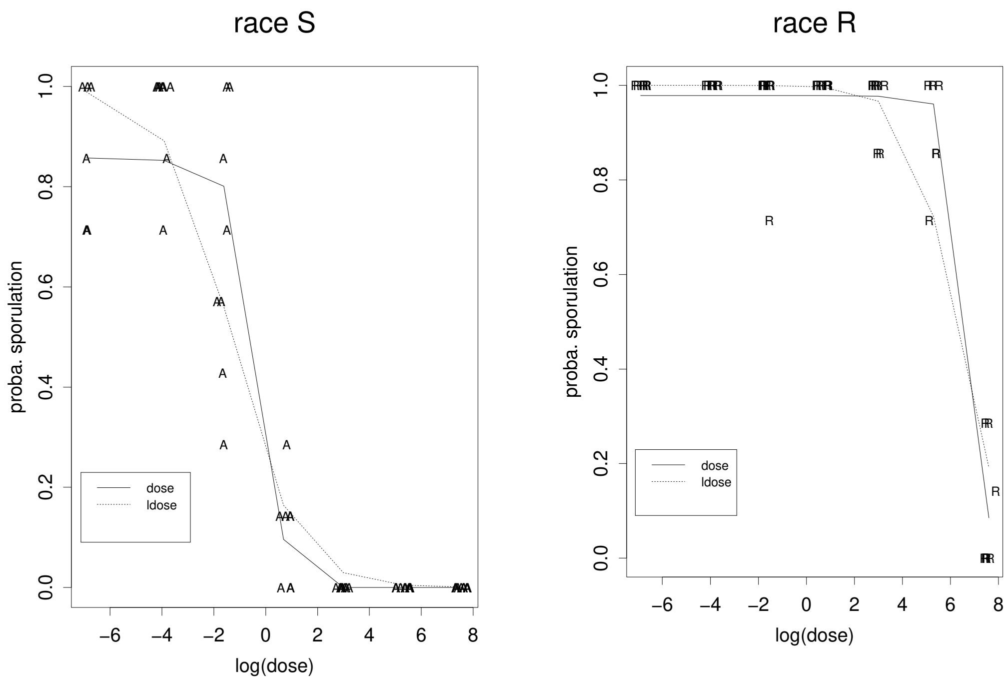
```
> glmS.logit.dose_glm(cbind(k,n-k)~dose,family=binomial,data=don.S,maxit=50)
> glmS.logit.ldose_glm(cbind(k,n-k)~ldose,family=binomial,data=don.S,maxit=50)
> glmR.logit.dose_glm(cbind(k,n-k)~dose,family=binomial,data=don.R,maxit=50)
> glmR.logit.ldose_glm(cbind(k,n-k)~ldose,family=binomial,data=don.R,maxit=50)

> summary(glmS.logit.dose)
Call: glm(formula = cbind(k, n - k) ~ dose, family = binomial,
    data = don.S, maxit = 50)
Deviance Residuals:
       Min            1Q        Median        3Q      Max
 -2.942648 -5.575504e-08 -5.575504e-08 0.3968735 1.762218

Coefficients:
                Value Std. Error   t value
(Intercept)  1.797096  0.2288690  7.852072
       dose -2.020711  0.2616394 -7.723266

(Dispersion Parameter for Binomial family taken to be 1 )
```

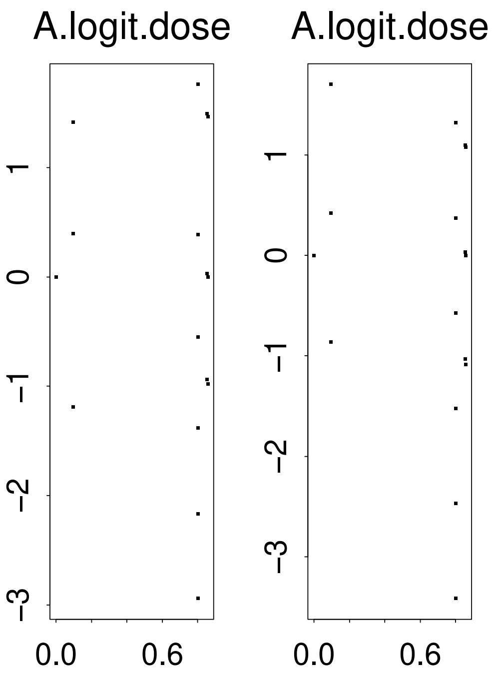
2. exemple INRA, déjà utilisé comme exemple introductif, au chapitre 1.

```
    Null Deviance: 372.4127 on 53 degrees of freedom

Residual Deviance: 53.31165 on 52 degrees of freedom

Number of Fisher Scoring Iterations: 13

Correlation of Coefficients:
     (Intercept)
dose -0.5144496

> summary(glmS.logit.ldose)
Call: glm(formula = cbind(k, n - k) ~ ldose, family = binomial,
    data = don.S, maxit = 50)
Deviance Residuals:
       Min         1Q     Median        3Q      Max
 -3.152369 -0.6464926 -0.1548669 0.3042517 2.849211

Coefficients:
                 Value Std. Error   t value
(Intercept) -1.0651452 0.21268919 -5.007990
      ldose -0.8115996 0.08891351 -9.127967

(Dispersion Parameter for Binomial family taken to be 1 )

    Null Deviance: 372.4127 on 53 degrees of freedom

Residual Deviance: 69.83558 on 52 degrees of freedom

Number of Fisher Scoring Iterations: 5

Correlation of Coefficients:
      (Intercept)
ldose 0.4942942

> summary(glmR.logit.dose)
Call: glm(formula = cbind(k, n - k) ~ dose, family = binomial,
  data = don.R, maxit =  50)
Deviance Residuals:
```

```
       Min        1Q    Median        3Q      Max
 -2.682102 0.5519452 0.5519613 0.5536376 1.540682

Coefficients:
                   Value   Std. Error   t value
(Intercept)  3.816773588 0.3697203164  10.32341
       dose -0.003096656 0.0003063746 -10.10742

(Dispersion Parameter for Binomial family taken to be 1 )

    Null Deviance: 265.2274 on 53 degrees of freedom

Residual Deviance: 43.08499 on 52 degrees of freedom

Number of Fisher Scoring Iterations: 4

Correlation of Coefficients:
     (Intercept)
dose -0.6251844

> summary(glmR.logit.ldose)
Call: glm(formula = cbind(k, n - k) ~ ldose, family = binomial,
   data = don.R, maxit = 50)
Deviance Residuals:
       Min          1Q     Median        3Q      Max
 -4.926244 0.003997515 0.06326019 0.2099192 2.131445

Coefficients:
                Value Std. Error   t value
(Intercept)  6.482130  0.8940284  7.250474
      ldose -1.042465  0.1424316 -7.319054

(Dispersion Parameter for Binomial family taken to be 1 )

    Null Deviance: 265.2274 on 53 degrees of freedom

Residual Deviance: 61.20601 on 52 degrees of freedom
```

```
Number of Fisher Scoring Iterations: 6

Correlation of Coefficients:
      (Intercept)
ldose -0.9668424

> sum(resid(glmS.logit.dose,"pearson")^2)
[1] 46.32682
> sum(resid(glmS.logit.ldose,"pearson")^2)
[1] 144.1321
> sum(resid(glmR.logit.dose,"pearson")^2)
[1] 64.23786
> sum(resid(glmR.logit.ldose,"pearson")^2)
[1] 1933.039
```





A.logit.dose
A.logit.dose

Exercice 9

à la fin du chapitre 8 :

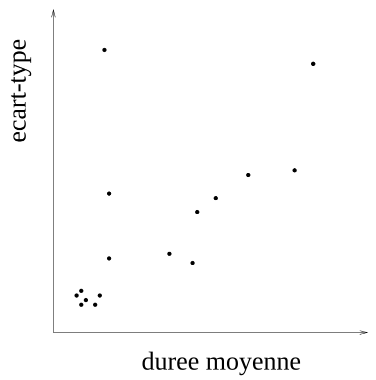
Choix de la loi.

A partir de la description des problèmes ci-dessous[3], choisir une loi succeptible de convenir.

1/ Le nombre total de nouveaux cas de SIDA, déclarés chaque mois, en Angleterre, à partir de novembre 85 est reporté. On souhaite modéliser l'évolution du nombre de nouveaux cas.
0 0 3 0 1 1 1 2 2 4 2 8 0 3 4 5 2 2 2 5 4 3 15 12 7 14 6 10 14 8 19 10 7 20 10 19

2/ L'effet de la température sur la durée du développement embryonnaire de la drosophile est étudié. Voici un extrait des données, et une représentation graphique.

numéro d'expérience	Temp.	Durée moyenne	Taille du lot	Ecart-type
1	14.95	67.5	54	2.41
2	16.16	57.10	182	2.28
...	...	...	...	...
21	30.05	16.81	148	0.39
22	30.80	16.97	195	0.39
23	32.00	18.20	58	2.23



3. les exemples 1 et 4 sont extraits de fiches d'exercice écrites par P.M.E. Altham, mises à disposition par l'auteur, à l'adresse //www.statslab.cam.ac.uk/ pat, dans le document INTRODUCTION TO S-PLUS FOR GENERALIZED LINEAR MODELLING, P.M.E. Altham, Statistical Laboratory, University of Cambridge.
Les exemples 3 et 5 sont empruntés à Collett.

3/ On étudie le cancer de la prostate : "The data are from the observational study on nodal involvement in patients presenting with prostatic cancer. In the data base there are observations on six variables for each of 53 patients, the variables where as follows.

Age :	age of patient in years,
Acid :	level of serum acid phosphatase,
X-ray :	result of an X-ray examination (0=negative, 1=positive),
Size :	size of tumour (0 = small, 1= large),
Grade :	grade of tumour (0=less serious, 1= more serious),
Nodes :	nodal involvement (0 = no involvement, 1= involvement).

On cherche à savoir quels sont les variables et facteurs explicatifs de la variable Nodes.

4/ On étudie la liaison entre le nombre d'accidents de la route, le volume du trafic, le type de rue, et l'heure de la journée. Les nombres d'accidents reportés sont de 11 9 4 4 20 4. Les 3 premières valeurs correspondent à une première rue, les 3 suivantes à une autre rue. Les heures de la journée sont 7h-9h30, 9h30-15h, 15h-18h30 pour les deux rues. Le volume du trafic associé à chaque observation est 2206 3276 19999 1399 2276 1417.

5/ On cherche à comparer la proportion de plants de pruniers qui "prennent" ou pas selon qu'ils sont plantés de façon précoce, ou au printemps, et selon que les racines sont coupées court ou long.

	precoce.court	precoce.long	tardif.court	tardif.long
vivant	107	156	31	84
mort	133	84	209	156