Exercice #1

a) $\mathcal{L} = \prod_{i=1}^{L} \frac{\mu_i^{y_i}}{y_i!} e^{-\mu_i} \qquad \mu_i = \exp(\alpha + \beta x_i)$

b) $\ln(\mathcal{L}) = \sum_{i=1}^{L} (y_i \ln(\mu_i) - \mu_i - \ln(y_i!))$

$$\ln(\mathcal{L}) = \sum_{i=1}^{L} (y_i(\alpha + \beta x_i) - \exp(\alpha + \beta x_i) - \ln(y_i!))$$

$$\frac{d \ln(\mathcal{L})}{d\alpha} = \sum_{i=1}^{L} (y_i - \exp(\alpha + \beta x_i)) = \sum_{i=1}^{L} (y_i - \mu_i)$$

$$\frac{d \ln(\mathcal{L})}{d\beta} = \sum_{i=1}^{L} (x_i y_i - x_i \exp(\alpha + \beta x_i)) = \sum_{i=1}^{L} (x_i y_i - x_i \mu_i)$$

$$s(\alpha, \beta) = \begin{bmatrix} \sum_{i=1}^{L} (y_i - \mu_i) \\ \sum_{i=1}^{L} (x_i y_i - x_i \mu_i) \end{bmatrix}$$

c) $\frac{d^2 \ln(\mathcal{L})}{d\alpha^2} = -\sum_{i=1}^{L} \mu_i \qquad \frac{d^2 \ln(\mathcal{L})}{d\beta^2} = -\sum_{i=1}^{L} (x_i^2 \mu_i)$

$$\frac{d^2 \ln(\mathcal{L})}{d\alpha d\beta} = \frac{d^2 \ln(\mathcal{L})}{d\beta d\alpha} = -\sum_{i=1}^{L} (x_i \mu_i)$$

$$I(\alpha, \beta) = \sum_{i=1}^{L} (x_i \mu_i) \begin{bmatrix} 1 & x_i \\ x_i & x_i^2 \end{bmatrix}$$

d) Voici une fonction pour estimer les paramètres avec S-Plus:

```
NewtRaph<-function (x,y) {
# les deux arguments sont le vecteur des x_i
# et celui des y_i (d\'efinis par des matrices d'une colonne)
    estims<-matrix(c(3,3),ncol=1)
    postestims<-matrix(c(1,1),ncol=1)
    mu<-rep(0,length(x))
    infoFish<-matrix(rep(0,4),ncol=2)
    score<-matrix(c(0,0),ncol=1)
    while (abs(estims[1,1]-postestims[1,1])>1e-6 | abs(estims[2,1]
            -postestims[2,1])>1e-6) {
        postestims<-estims
        mu[1:length(x)]<-exp(postestims[1,1] + postestims[2,1]
            *x[1:length(x),1])
        infoFish[1,1]<-sum(mu[1:length(x)])
        infoFish[1,2]<-sum(mu[1:length(x)]*x[1:length(x),1])
        infoFish[2,1]<-infoFish[1,2]
        infoFish[2,2]<-sum(mu[1:length(x)]*x[1:length(x),1]^2)
        score[1,1]<-sum(y[1:length(x),1]-mu[1:length(x)])
        score[2,1]<-sum(x[1:length(x),1]*(y[1:length(x),1]
            -mu[1:length(x)]))
        estims<-postestims+solve(infoFish)%*%score
```

```
    }
rownames(estims)<-c("alpha estim\'e","beta estim\'e")
estims
}
```

Exercice #2

a) Estimation de α et β avec les "ordinary least squares":

```
a<-read.table('limule.dat')
y<-rep(0,dim(a)[1])
y[a[1:dim(a)[1],4]>0]<-1
y[a[1:dim(a)[1],4]==0]<-0
x<-c(a[1:dim(a)[1],5])/1000
Sxy<-sum(x*y)-dim(a)[1]*mean(x)*mean(y)
Sxx<-sum(x^2)-dim(a)[1]*mean(x)^2
beta<-Sxy/Sxx
alpha<-mean(y)-beta*mean(x)
alpha
beta
```

$\widehat{\alpha} = -0.1448709 \qquad \widehat{\beta} = 0.3227033$
Une augmentation du poids de 3kg augmente de 1 le nombre moyen de mâles satellites.
$\widehat{\pi}(x) = \widehat{\alpha} + \widehat{\beta}x$
$\widehat{\pi}(5.2) = 1.533186$
Étant donné qu'en fait π est supposé être compris entre 0 et 1, (si Y n'est pas continue) à 5.2kg il est assez certain que la probabilité d'avoir un satellite ou plus est de un.

b) Voir le programme à la fin.

c) $\widehat{\alpha} = -3.6947 \qquad \widehat{\beta} = 1.8151$
$\ln\left(\frac{\widehat{\pi}(5.2)}{1-\widehat{\pi}(5.2)}\right) = \widehat{\alpha} + \widehat{\beta} \times 5.2 = 5.7438 \quad \Rightarrow \quad \widehat{\pi}(5.2) = 0.9968$

d) $\widehat{\alpha} = -2.2383 \qquad \widehat{\beta} = 1.0990$
Ici l'ajustement du modèle probit correspond à une fonction de répartition d'une normale de moyenne 2.03667 et d'écart-type 0.90992.
$\Phi^{-1}(\widehat{\pi}(x)) = \widehat{\alpha} + \widehat{\beta}x$
$\widehat{\pi}(5.2) = \Phi(\widehat{\alpha} + \widehat{\beta} \times 5.2) = \Phi(3.4765) = 0.999746$

Exercice #3

a) $\widehat{\alpha} = -0.4284 \qquad \widehat{\beta} = 0.5893$
$\ln(\widehat{\mu}(x)) = \widehat{\alpha} + \widehat{\beta}x$
$\widehat{\mu}(2.44) = \exp(\widehat{\alpha} + \widehat{\beta} \times 2.44) = 2.744$

b) $\exp(\widehat{\beta}) = 1.8027$
Une augmentation de 1kg du poids provoque une augmentation du nombre moyen de satellites de 80%.
$\widehat{\beta} \in [0.4619, 0.7167]$ au niveau de 95%
Donc, $\exp(\widehat{\beta}) - 1 \in [58.7\%, 105.8\%]$
Au seuil de 5%, une augmentation de 1kg du poids provoque une augmentation du nombre moyen de satellites d'au moins 58.7%.

c) $H_0 : \beta = 0 \qquad H_1 : \beta \neq 0$

$z = \dfrac{\widehat{\beta}}{ASE(\widehat{\beta})} = \dfrac{0.5893}{0.065} = 9.066$

$Z \sim N(0,1)$

$\alpha_0 = 2(1 - \Phi(z)) \approx 0 \ll 0.05$

$\Rightarrow$ On rejette H_0

$\Rightarrow$ Le nombre de satellites n'est pas indépendant du poids.

d) $dev_0 - dev_1 = 632.7917 - 560.8664 = 71.9253$

$\alpha_0 = P(\chi^2_1 \geq 71.9253) \approx 0 \ll 0.05$

$\Rightarrow$ On rejette H_0

$\Rightarrow$ Le nombre de satellites n'est pas indépendant du poids. Ce résultat confirme la conclusion du test de Wald.

Exercice #4

a) $X^2 = 390.1308$

$P(\chi^2_{117} \geq 390.1308) \approx 0 \ll 0.05$

$\Rightarrow$ On rejette H_0

$dev_1 = 632.7917 \qquad dev_s = 415.0989$

$G^2 = dev_1 - dev_s = 217.6928$

$P(\chi^2_{55} \geq 217.6928) \approx 0 \ll 0.05$

$\Rightarrow$ On rejette H_0

Selon les tests du chi-deux de Pearson et du rapport de vraisemblances, le modèle ne décrit pas bien le lien entre les variables.

Juste en regardant la droite de Henri à la page 29 dans les soties SAS, il est évident que les résidus ne sont pas du tout normaux. Sur le graphique des résidus en fonction du poids, il ne semble pas y avoir de schéma particulier, ce qui pourrait indiquer l'indépendance entre les résidus.

b) Les test conduits en a) mènent à conclure à une dispersion extra-poissonnienne.

$\widehat{\alpha} = -0.4284 \qquad \widehat{\beta} = 0.5893$

Comme dans l'exercice #3, sauf qu'ici, $ASE_{cor}(\widehat{\alpha}) = 0.3168$ et $ASE_{cor}(\widehat{\beta}) = 0.1151$ au lieu de respectivement 0.1789 et 0.0650. Ceci modifie entre autres la valeur de la statistique de Wald (maintenant 5.12) sans modifier la conclusion à laquelle elle mène.

Juste un mot sur le nouveau X^2 de 117.0000 qui maintenant fait dire que le modèle décrit bien le lien entre les variables car $P(\chi^2_{117} \geq 117.0000) = 0.482612$.