Analyse des tableaux de fréquences **STT-16569/STT-65623**
21 avril 2008 **9h30-11h30**

Épreuve partielle #2

(20 points)
1) Considérons le modèle de régression logistique M où la probabilité de succès π ne dépend que d'une constante α :

$$\log\left(\frac{\pi}{1-\pi}\right) = \alpha,$$

où $0 < \pi < 1$.

a) Quelles sont les valeurs possibles du paramètre α ?

b) Étant donné n variables de Bernoulli indépendantes $Y_1, \ldots, Y_n$, posons $\pi_i = E(Y_i)$, $i = 1, \ldots, n$. Que peut-on dire de plus sur les Y_i si le modèle M s'applique ?

c) En termes de π, la fonction de vraisemblance pour l'ajustement de M s'écrit

$$L(\pi) = \prod_{i=1}^{n} \pi^{y_i}(1-\pi)^{1-y_i}, \quad 0 < \pi < 1.$$

Déterminer l'estimateur de vraisemblance maximale de π en distinguant le cas où $0 < \sum_i y_i < n$, et les deux cas extrêmes $\sum_i y_i = 0$ ou n.

d) Déterminer l'estimateur de vraisemblance maximale de α en faisant la même distinction de cas qu'en c).

(25 points)
2) Les données suivantes proviennent d'une étude visant à déterminer la toxicité d'un gaz. Après avoir exposé les insectes à différentes concentrations de gaz pendant 5 heures, on a dénombré ceux qui en sont morts.

Concentration x_i	**N. d'insectes exposés n_i**	**N. d'insectes morts**
1.6907	59	6
1.7242	60	13
1.7552	62	18
1.7842	56	28
1.8113	63	52
1.8369	59	53
1.8610	62	61
1.8839	60	60

a) Décrire brièvement le type de données.

↪ **suite à la page suivante**

b) On décide de modéliser par une régression logistique la probabilité de décès en fonction de la concentration. Décrire brièvement le modèle appliqué à cet exemple.

c) L'ajustement a produit la sortie suivante.

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -60.717      5.181  -11.72   <2e-16 ***
dose          34.270      2.912   11.77   <2e-16 ***

    Null deviance: 284.202  on 7  degrees of freedom
Residual deviance:  11.232  on 6  degrees of freedom
```

Dans cette sortie, la déviance résiduelle est la déviance de notre modèle, tandis que la déviance nulle est celle du modèle ne contenant qu'une ordonnée à l'origine et aucune variable explicative (*comme dans le numéro 1 plus haut*).

En vous inspirant des exemples vus en classe, interpréter brièvement ces résultats.

d) Les moyennes prédites des 8 binomiales sont 3.46, 9.84, 22.45, 33.90, 50.10, 53.29, 59.22, 58.74. Expliquer comment ces nombres sont obtenus et donner un exemple de calcul.

e) Selon le modèle ajusté, quelle est la proportion estimée d'insectes tués lorsque la dose de gaz est 1.8 ?

f) Sur un même graphique, représenter en ordonnée la proportion d'insectes tués ainsi que cette proportion estimée par le modèle avec en abscisse la concentration de gaz. Commenter brièvement.

(30 points)

3) Lors d'une expérience, on a comparé deux procédés de fabrication d'une puce électronique. Le procédé A appliqué à 10 puces a entraîné sur celles-ci respectivement $8, 7, 6, 6, 3, 4, 7, 2, 3, 4$ imperfections alorsque le procédé B appliqué lui aussi à 10 puces a entraîné $9, 9, 8, 14, 8, 13, 11, 5, 7, 6$ imperfections. On considère que ces fréquences sont des variables de Poisson indépendantes.

a) On s'intéresse d'abord au modèle $\log \mu = \alpha + \beta x$, où $x = 1$ pour le procédé B et $x = 0$ sinon. Brièvement, interpréter α et β.

b) On ajuste le modèle présenté en a) pour obtenir le résultat suivant.

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.6094     0.1414  11.380  < 2e-16 ***
x1            0.5878     0.1764   3.332 0.000861 ***

    Null deviance: 27.857  on 19  degrees of freedom
Residual deviance: 16.268  on 18  degrees of freedom
```

Interpréter brièvement. ↪ **suite à la page suivante**

c) Soient μ_A, μ_B les nombres moyens d'imperfections pour les deux procédés. Au seuil 5%, tester l'hypothèse $H_0 : \mu_A = \mu_B$ au moyen d'un test bilatéral.

d) Estimer les moyennes des deux procédés.

e) Construire un intervalle de confiance de 95% pour le rapport μ_A/μ_B.

f) On classifie les puces selon l'épaisseur du revêtement de silicium ($z = 0$, mince ; $z = 1$, épais). Pour chacun des procédés, les 5 premières puces ont un revêtement mince et les 5 dernières un revêtement épais. On s'intéresse au modèle de régression de Poisson

$$\log \mu = \alpha + \beta x + \gamma z.$$

Interpréter brièvement les 3 paramètres de ce modèle.

g) L'ajustement du modèle à 2 variables explicatives a donné le résultat suivant.

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.7177     0.1602  10.719  < 2e-16 ***
x1            0.5878     0.1764   3.332 0.000861 ***
x2           -0.2296     0.1701  -1.349 0.177246

    Null deviance: 27.857  on 19  degrees of freedom
Residual deviance: 14.435  on 17  degrees of freedom
```

Commenter et tirer une conclusion générale.

(25 points)
4)Y a-t-il un lien entre un type de cancer de la peau et sa localisation ? Pour répondre à cette question, on dispose des données suivantes recueillies chez 400 patients.

	Localisation			
Type histologique	Tête et cou	Tronc	Extrémités	Total
Rousseur de Hutchinson	22	2	10	34
Mélanome superficiel	16	54	115	185
Nodulaire	19	33	73	125
Indéterminé	11	17	28	56
Total	68	106	226	400

a) Faire une brève analyse descriptive du tableau pour en faire ressortir des points saillants.

b) On décide d'ajuster le modèle log-linéaire d'indépendance. Décrire brièvement celui-ci.

c) L'ajustement du modèle d'indépendance donne le résultat qui suit. Commenter.

↪ **résultat à la page suivante**

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.4544     0.1406  24.561  < 2e-16 ***
x1           -0.4990     0.2174  -2.295   0.0217 *
x2            1.1950     0.1525   7.835 4.69e-15 ***
x3            0.8030     0.1608   4.993 5.93e-07 ***
y1           -1.2010     0.1383  -8.683  < 2e-16 ***
y2           -0.7571     0.1177  -6.431 1.27e-10 ***

    Null deviance: 295.203  on 11  degrees of freedom
Residual deviance:  51.795  on  6  degrees of freedom
```

d) On ajuste ensuite le modèle log-linéaire saturé pour obtenir le résultat suivant. Commenter.

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   3.3322     0.1890  17.632  < 2e-16 ***
x1           -1.0296     0.3684  -2.795 0.005192 **
x2            1.4127     0.2107   6.704 2.03e-11 ***
x3            0.9583     0.2223   4.311 1.63e-05 ***
y1           -0.9343     0.3558  -2.626 0.008649 **
y2           -0.4990     0.3075  -1.623 0.104612
x1:y1         1.7228     0.5216   3.303 0.000957 ***
x1:y2        -1.1104     0.8334  -1.332 0.182713
x2:y1        -1.0380     0.4448  -2.334 0.019602 *
x2:y2        -0.2570     0.3489  -0.736 0.461479
x3:y1        -0.4117     0.4393  -0.937 0.348618
x3:y2        -0.2950     0.3722  -0.792 0.428092

    Null deviance: 2.9520e+02  on 11  degrees of freedom
Residual deviance: 1.2657e-14  on  0  degrees of freedom
```

e) Sous les contraintes habituelles, on sait que

$$\lambda_{ij}^{XY} = \log\left(\frac{\mu_{ij}\mu_{IJ}}{\mu_{Ij}\mu_{iJ}}\right) = \log\left(\frac{\mu_{ij}/\mu_{iJ}}{\mu_{Ij}/\mu_{IJ}}\right).$$

Après avoir observé que certains effets d'interaction sont significativement différents de 0, donner une interprétation de ce phénomène pour les variables analysées.

Jean-Claude Massé
Professeur