Projet Master 1 DSMS: Modélisation de la mortalité des scarabées

Contexte de l'étude : On étudie la mortalité de scarabées en fonction de la dose d'un insecticide. Les données proviennent du Tableau 7.2 (non reproduit ici). Pour chaque dose X_i, on observe Y_i scarabées morts parmi n_i individus exposés.

1 Partie théorique

1.1 Modèle logistique (lien logit)

1. **Fonction de vraisemblance** :
   i) Exprimer la vraisemblance $L(\beta)$ pour le modèle logistique.
   ii) Donner l'expression de la log-vraisemblance $\ell(\beta)$.
2. **Score et information de Fisher** :
   i) Donner l'expression du vecteur score $U(\beta)$.
   ii) Donner l'expression de la matrice d'information de Fisher $I(\beta)$.
3. **Algorithme de Newton-Raphson** :
   i) Écrire la relation de récurrence pour l'estimation de β par la méthode du score de Fisher.
   ii) Préciser la valeur initiale de β.
4. **Tests d'ajustement** :
   i) Donner l'expression de la deviance D pour ce modèle.
   ii) Donner l'expression de la statistique du χ^2 de Pearson pour ce modèle.
   iii) Donner l'expression des résidus de deviance.
   iv) Donner l'expression des résidus de Pearson.

1.2 Modèle probit

On considère maintenant le modèle avec fonction lien *probit* définie par :

$$g(p) = \Phi^{-1}(p), \quad \text{où} \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

est la fonction de répartition de la loi normale centrée réduite.
Consigne : Répondre aux questions 1 à 4 du modèle logistique pour ce modèle probit.

1.3 Modèle log-log complémentaire (c-log-log)

On considère le modèle avec fonction lien inverse :

$$g^{-1}(x) = 1 - \exp\left(-\mathrm{e}^{x}\right), \quad x \in \mathbb{R}.$$

Cette fonction correspond à la fonction de répartition de la loi de Gumbel (loi des valeurs extrêmes type I).
Consigne : Répondre aux questions 1 à 4 du modèle logistique pour ce modèle c-log-log.

2 Partie pratique

2.1 Modèle logistique (lien logit)

a) **Visualisation des données** :

  i) Tracer le graphique de la proportion observée $p_i = Y_i/n_i$ en fonction de la dose X_i.

  ii) Commenter l'allure de la relation dose-réponse.

b) **Estimation des paramètres** :

  i) Implémenter l'algorithme de Newton-Raphson jusqu'à convergence.

  ii) Donner les estimateurs du maximum de vraisemblance obtenus.

c) **Prédictions** :

  i) Calculer les valeurs prédites linéaires $\widehat{\eta}_i$.

  ii) Calculer les proportions prédites $\widehat{p}_i$.

  iii) Calculer les effectifs prédits $\widehat{Y}_i$ pour chaque dose.

d) **Visualisation des résultats** :

  i) Tracer sur un même graphique les proportions observées et les proportions prédites $\widehat{p}_i$ en fonction de la dose X_i.

  ii) Commenter la qualité de l'ajustement.

e) **Diagnostics** :

  i) Calculer numériquement la deviance D pour ce modèle.

  ii) Calculer la statistique du χ^2 de Pearson.

2.2 Modèle probit

Consigne : Reprendre l'analyse pratique (questions a à e) avec le modèle probit.

2.3 Modèle c-log-log

Consigne : Reprendre l'analyse pratique (questions a à e) avec le modèle c-log-log.

3 Comparaison des modèles

a) **Synthèse des résultats** :

   i) Récapituler dans un tableau les valeurs de la deviance D et du χ^2 de Pearson pour les trois modèles.

   ii) Comparer les estimateurs des paramètres obtenus pour chaque modèle.

b) **Sélection du modèle** :

   i) Proposer un critère de sélection basé sur la deviance.

   ii) Discuter des avantages et inconvénients de chaque modèle.

c) **Conclusion** :

   i) Quel modèle semble le plus adapté aux données ?

   ii) Justifier le choix du modèle retenu en s'appuyant sur les résultats numériques et graphiques.