

**Epreuve de modélisation (Option Proba/Stats)**  
**Tests du  $\chi^2$  d'indépendance, régression linéaire**

Le répertoire stixbox contient des fichiers de données (`datas1.m, \dots, datas23.m`) qui peuvent être utiles pour mettre en oeuvre sur des exemples les techniques statistiques du programme. La fonction `x=getdata(i)` affiche une description succincte des données du fichier `datasi.m` et retourne les données dans le tableau `x`.

**Expérience 1.** Charger les données du fichier `datas17.m` concernant un problème de contrôle de qualité. Tester l'uniformité de la distribution du nombre de pièces produites en fonction du jour de la semaine puis imaginer une façon de tester si le jour de fabrication d'une pièce a une influence sur la qualité de sa fabrication.

**Expérience 2.** Charger les données du fichier `datas5.m` qui liste pour un certain nombre d'animaux le poids du corps (en kg) (noté  $P$ ) et le poids du cerveau en grammes (noté  $p$ ).

- (1) Afficher les paires  $(\log(P), \log(p))$  par une commande `plot` adéquate et remarquer l'existence de quelques données atypiques ? De quels animaux s'agit-il ?
- (2) Retirer les données atypiques. On note  $X_i = \log(P_i)$  et  $Y_i = \log(p_i)$  où  $i$  est le  $i$ ème animal. On considère le modèle

$$Y_i = a + bX_i + \epsilon_i$$

où les  $(\epsilon_i)$  sont des variables aléatoires gaussiennes indépendantes centrées et de variance  $\sigma^2$ .

- (a) Déterminer les estimateurs des moindres carrés  $\hat{a}$  et  $\hat{b}$  de  $a$  et  $b$ . Afficher les données  $(X_i, Y_i)$  et la droite d'équation  $y = \hat{a} + \hat{b}x$ .
  - (b) Exprimer la matrice de variance-covariance de  $\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix}$  en fonction de  $\sigma^2$ .
  - (c) Donner la forme de l'estimateur sans biais  $S^2$  de  $\sigma^2$  et calculer sa valeur.
  - (d) Construire une région de confiance pour  $\hat{\theta} = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix}$  de niveau de confiance 95%.
- (3) On se pose la question de savoir si le poids du cerveau est proportionnel au poids du corps. A partir des résultats précédents, proposer une méthode pour tester cette hypothèse et faites le test.
  - (4) Quel est le poids du cerveau prédit par le modèle pour les animaux atypiques que vous avez retirés au début. Construire à chaque fois un intervalle de confiance à 95%. Cela vous conforte-t-il dans cette élimination ?