

# DM1 TISD - M1 Pro

## à rendre pour le Vendredi 23 octobre 2009

Les codes sont à envoyer au plus tard le 22 octobre soir à : `chi.tran@univ-lille1.fr`. La copie papier sera ramassée en TD. Il n'est pas nécessaire d'y recopier votre code **R**.

### Exercice 1 (Tremblements de terre, d'après le partiel 2008)

Nous disposons d'une base de données relative à 1000 séismes de magnitude supérieure à 4 (échelle de Richter) s'étant produit depuis 1964 dans le voisinage de Fiji. Cette base est disponible sous **R**, vous pouvez consulter sa fiche en tapant `help(quakes)`. Les séismes sont caractérisés par plusieurs variables dont la magnitude `mag`, la latitude `lat` et la longitude `long`.

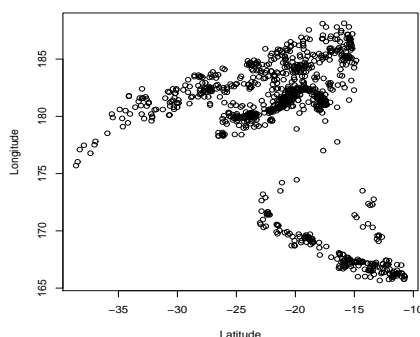


FIG. 1 – Localisation des séismes

En regardant la localisation géographique des séismes, on s'aperçoit qu'il y a clairement deux régions d'activité sismique. La première correspond à une jonction de plaque, la seconde à la tranchée de Tonga en Nouvelle Zélande. Nous nous demandons si la magnitude du séisme est liée à la région géographique où il se produit.

1. Nous considérons la variable `mag` sans distinguer la localisation géographique dans un premier temps.

1.1. Quelle est la magnitude moyenne, l'écart-type ?

1.2. Calculer l'étendue et l'intervalle inter-quartile.

1.3. Dessiner la *boîte à moustache*.

2. Créer une variable `groupe` qui vaut 1 si le séisme se produit à une longitude supérieure à 175 et 2 s'il se produit à une longitude inférieure à 175.

2.1. Donner et comparer les moyennes et les variances pour chaque groupe.

2.2. Rétablir la formule de décomposition de la variance vue en cours.

2.3. Faire numériquement l'analyse de la variance. Quelle est la part de la variance expliquée par la variable `groupe` ? Conclusion ?

3. Nous nous demandons maintenant si les magnitudes ont pour loi une loi **gamma**.

3.1. La loi Gamma de paramètre de forme  $a$  et de paramètre d'échelle  $s$  est absolument continue par rapport à la mesure de Lebesgue de densité :

$$f(x) = \frac{1}{s^a \Gamma(a)} x^{a-1} e^{-x/s} \mathbf{1}_{x>0}.$$

Lorsque  $a$  est entier, rappeler une façon rapide pour montrer que l'espérance d'une loi Gamma est  $as$  et sa variance  $as^2$ . Ce résultat reste vrai lorsque  $a \in \mathbb{R}_+^*$ .

3.2. Tracer l'histogramme de la magnitude pour chaque groupe et superposer la densité de lois Gamma bien choisies. Commenter.

3.3. Pour chaque groupe, tracer le QQ-plot pour voir si graphiquement les données pourraient être issues d'une loi Gamma. Commenter ces graphiques, en particulier les queues de distribution.

3.4. Quel est à votre avis l'intérêt de tester l'adéquation de la distribution des magnitudes à une loi paramétrique (en l'occurrence une loi Gamma) ?

### Exercice 2 (Répartition salariale sur des données groupées, d'après le partiel de 2007-2008)

Dans une entreprise, les salaires sont les suivants :

Classe de salaire	Salaires mensuels	Nombre de salariés
1	[500, 1500[	50
2	[1500, 2500[	125
3	[2500, 5500[	25

1. Pour chaque classe  $i \in \{1, 2, 3\}$  de salaires (notée  $[x_{i-1}, x_i[$  et d'effectif  $n_i$ ), calculer la fréquence empirique  $f_i$ , l'amplitude  $a_i$ , le centre  $c_i = (x_{i-1} + x_i)/2$ , la fréquence empirique cumulée  $F_i$  (proportion des salaires inférieurs à  $x_i$ ), la masse salariale approchée  $n_i c_i$  et la masse salariale cumulée approchée  $m_i$  (approximation de la somme de tous les salaires inférieurs à  $x_i$ ). On présentera les résultats dans un tableau.

2. Tracer avec **R** l'histogramme de la variable "salaire". Quelle règle faut-il respecter ?

3. Dessiner avec **R** la fonction de répartition. Comme la variable de salaire est quantitative continue, on choisira ici la version continue de la fonction de répartition empirique, obtenue par interpolation linéaire des points  $(x_i, F_i)$ .

4. Quelle est l'équation de la portion de droite représentant la fonction de répartition empirique sur l'intervalle de salaires [1500, 2500[ ? En déduire la médiane.

5. On s'intéresse à la répartition des salaires sur ces données agrégées.

5.1. Tracer la courbe de Lorenz avec **R** (on pourra utiliser la commande `segments`). Pour des données groupées, cette courbe est obtenue en reliant les points

$$\left( F_i, \frac{\sum_{j \leq i} n_j c_j}{\sum_{j=1}^3 n_j c_j} \right).$$

5.2. Calculer l'indice de Gini.

5.3. Commenter.