

ENSIIE : Projet MST 2

2015-2016

Le projet est à rendre pour le 25 Mai 2016.

Consignes

1. Les étudiants travaillent par groupes de 1 ou 2.
2. La rédaction d'un rapport écrit en word ou en latex est obligatoire.
3. Les étudiants doivent utiliser le logiciel R et rendront leurs codes en même temps que leur rapport dans un fichier **à part** portant l'extension `.R`.
4. 1 point par jour de retard sera retiré de la note finale.
5. La notation tiendra compte de la qualité de rédaction du rapport ainsi que de la qualité des graphiques demandés.

Vous pouvez contacter vos chargés de TD par mail :

morgane.pierrejean@genopole.cnrs.fr

nicolas.brunel@ensiie.fr

Nous étudions ici le nombre de jours d'absence dans une école maternelle et primaire. Le jeu de donnée comprend 150 élèves et 3 variables (le nombre de jour d'absence, le sexe de l'élève et son âge). Le but de ce projet est de réaliser une petite étude en utilisant ce qui a été vu durant le module de modélisation statistique 2. Le projet comprend plusieurs parties: statistiques descriptives, inférence, intervalle de confiance, tests et prédiction.

PARTIE I [statistiques descriptives]

- 1.1 Charger le jeu de données disponible à cette adresse http://www.math-evry.cnrs.fr/members/mpierrejean/welcome#teaching_in_french
- 1.2 Tracer les boîtes à moustaches du nombre de jours d'absence pour les filles et les garçons sur un même graphique. Que remarquez-vous?
- 1.3 Tracer les boîtes à moustache du nombre de jours d'absence pour chaque Age. Que remarquez-vous?
- 1.4 Tracer l'histogramme du nombre de jour d'absence. Quelles distributions le nombre de jours d'absence pourrait-il suivre (justifiez)?

PARTIE II [statistique inferentielle]

On considère que le nombre de jour d'absence suit une loi de Poisson $\mathcal{P}(\lambda)$. Dans cette partie nous cherchons à estimer λ .

- 2.1 Vérifier que les observation suivent une loi de poisson.
- 2.2 Ecrire la vraisemblance du modèle.
- 2.3 Déterminer le maximum de vraisemblance du paramètre λ . Tracer la log-vraisemblance de l'échantillon en fonction de la valeur du paramètre λ .
- 2.4 Donner une estimation du paramètre λ et la faire apparaître sur le graphique.

PARTIE III [intervalle de confiance]

3.1 Ecrire mathématiquement l'intervalle de confiance asymptotique pour le paramètre λ en justifiant.

3.2 Calculer avec l'aide de `R` les bornes de l'intervalle de confiance asymptotique.

PARTIE IV [tests]

4.1 Nous souhaitons comparer les jours d'absence entre les filles et les garçons. Proposez un test et l'implémenter sous `R`. Que concluez-vous?

4.2 Nous souhaitons comparer les jours d'absence entre les moins de 9 ans et les plus de 9 ans inclus. Proposez un test et l'implémenter sous `R`. Que concluez-vous?

PARTIE V [prédiction]

Dans cette partie, nous nous intéressons à l'influence de l'âge sur le nombre de jours d'absence. Le nombre de jours d'absence suit un modèle de Poisson.

5.1 Ecrire le modèle de Poisson dans notre cas.

5.2 Déterminer les coefficients α et β à l'aide de la fonction `glm`.

5.3 Predire le nombre de jours d'absence pour un élève masculin âgé de 10 ans.