



École nationale  
de la statistique  
et de l'analyse  
de l'information

**Dossier suivi par :**

Khaled Larbi

Tel. : +33 (0)2 99 05 32 42

khaled.larbi@ensai.fr

**Objet**    **PROJETS**  
———    **STATISTIQUES**

Bruz, le 1 octobre 2019

## Encadrer des projets statistiques de 2<sup>e</sup> année à l'Ensaï

L'Ensaï fait appel à des praticiens de la statistique pour encadrer des projets statistiques de 2<sup>e</sup> année. Cette activité pédagogique a pour objectif l'application, voire l'approfondissement, des connaissances statistiques acquises.

### Travail des élèves

Les élèves travaillent par groupes de trois ou quatre sur un sujet qu'ils ont choisi, sous la direction du praticien qui l'a proposé. Ils ont à circonscrire le problème soumis et à le traiter durant le temps imparti (de début janvier à fin avril), en utilisant les concepts, méthodes et outils appropriés.

Chaque groupe rend compte de son travail sous forme d'un rapport écrit et d'une soutenance.

La quantité de travail que l'on peut attendre des élèves est de l'ordre de 4 heures hebdomadaires par élève, compte tenu de leur charge de travail extérieure au projet (ils continuent parallèlement à suivre les cours de l'école).

### Contenu statistique du projet

Pour répondre à la problématique, les élèves doivent utiliser leurs connaissances en séries temporelles, théorie des sondages, analyse de données, régression linéaire, régression sur variables catégorielles, économétrie mais également en data science. D'autres outils peuvent être utilisés. Le tuteur s'engage alors à les présenter aux étudiants.

Les élèves devront donc *a minima* utiliser un modèle.

Campus de Ker Lann  
rue Blaise Pascal BP37203  
35172 BRUZ CEDEX  
Tél : 33 (0)2 99 05 32 32  
communication@ensai.fr

## Rôle de l'encadrant

Le tuteur amène un sujet composé de :

- une problématique,
- un ou plusieurs jeu(x) de données qui doit être structuré, de volume raisonnable et documenté si besoin,
- des éléments de bibliographie nécessaires aux étudiants pour contextualiser l'analyse et discuter les résultats obtenus lors de la rédaction du rapport final.

L'encadrement d'une telle activité nécessite de la part de l'encadrant une bonne connaissance des données qu'il fournit aux élèves et une bonne maîtrise des techniques statistiques de base utilisées pour leur traitement.

Pour des raisons d'organisation, chaque tuteur est incité à proposer plusieurs problématiques différentes, donc à encadrer les travaux de plusieurs groupes d'élèves. Les sujets peuvent s'appuyer sur le même jeu de données.

## Calendrier

- D'octobre 2019 à mi-novembre 2019, les futurs animateurs proposent des sujets complets (avec les jeux de données). L'Ensaï valide les sujets.
- En décembre, le recueil des sujets proposés est diffusé par l'Ensaï aux élèves, qui forment librement leur groupe. Ils émettent une liste de choix. Puis chaque sujet est affecté à un groupe.
- Lors de la première séance de suivi (ayant lieu un lundi matin au début du mois de janvier), le tuteur rencontre le ou les groupes qu'il encadrera au cours d'une réunion de présentation.
- Pour la suite du projet, chaque groupe d'élèves rencontre le tuteur de son projet quatre fois, dans le cadre de séances de « suivi de projet ». Deux fois au plus, ces suivis peuvent être organisés en visioconférence. Des créneaux sont réservés au projet le lundi matin dans l'emploi du temps des élèves.
- Mi-avril, les élèves remettent leur rapport.
- Les soutenances des projets statistiques ont lieu mi-mai. Chaque groupe présente son rapport devant un jury composé d'un président, d'un permanent de l'Ensaï, d'un expert en communication et du tuteur.

## Si vous souhaitez encadrer un ou plusieurs projets

Pour chaque proposition de sujet, vous devez présenter, en une page, le thème de l'étude, la problématique, le type de données, ainsi que quelques éléments de bibliographie. Un exemple de proposition est joint en annexe.

Envoyez le tout à **Khaled Larbi**, enseignant en statistique ([khaled.larbi@ensai.fr](mailto:khaled.larbi@ensai.fr))

## Aspects informatiques (données et logiciels)

La totalité des données à traiter doit être disponible avant le début du projet. La taille des bases de données est limitée à 100 Mo. Si vous jugez cette taille insuffisante pour votre projet, vous devez me le signaler afin que l'on adapte le dispositif.



Les fichiers de données devront parvenir à l'Ensaï sur support informatique avant la fin de l'année 2019. Pour éviter tout problème de confidentialité, les données seront stockées dans un répertoire dont l'accès est limité aux seuls élèves du groupe concerné et au responsable des projets. A ce propos, tous les élèves signeront une charte de confidentialité dès le début des projets.

Pour le traitement informatique des données, l'utilisation de Sas ou R est obligatoire.

### **Rémunération des intervenants**

Les projets sont rémunérés au forfait. Les éventuels frais de déplacement sont pris en charge par l'Ensaï.

## Exemple de projet de 2<sup>e</sup> année proposé l'an dernier

<b>Sujet – Emergence des effets secondaires au cours des essais cliniques</b>
Nom du tuteur : Sandrine KATSAHIAN
Entreprise ou organisme : URC HEGP CIC-EC Assistance Publique – Hôpitaux de Paris AP-HP

### 1. Contexte

L'Hôpital Européen Georges Pompidou (HEGP) fait partie des 37 centres hospitaliers gérés par l'Assistance Publique des Hôpitaux de Paris. Situé au sud-ouest de Paris, il répond aux besoins de santé de près de 600 000 habitants. L'activité de l'hôpital est reconnue notamment dans les domaines cardiovasculaire et rénal, ainsi que dans la prise en charge des cancers et des urgences. L'Unité de recherche Clinique de l'HEGP a été créée pour aider les médecins investigateurs à monter, financer, réaliser et soutenir leurs projets de recherche reposant sur des patients. Ces projets sont mis en place dans l'hôpital mais peuvent également inclure la participation d'autres hôpitaux (à et en dehors de l'AP-HP). Plus particulièrement, l'URC a pour mission de mettre en œuvre le monitoring des projets (vérification des données recueillies dans le service), d'assurer la gestion puis l'analyse des données collectées durant ces projets (data management et analyse statistique) et de participer à la coordination des activités de recherche du groupe hospitalier.

### 2. Sujet proposé

Un grand nombre d'évènements indésirables est recueilli de manière la plus systématique possible au cours des essais cliniques et il est nécessaire d'en avoir une présentation synthétique.

L'objet de ce projet est de mettre en œuvre des méthodes implémentées dans les bases de données internationales (FDA, EMA, OMS), sur des données de tolérance provenant d'essais cliniques. L'objectif est de faire émerger des effets secondaires dont la fréquence apparaîtrait anormalement élevée.

### 3. Méthodologie envisagée

Programmation, application et comparaison de méthodes fréquentistes et bayésiennes, implémentées dans les bases de données internationales, au moyen des logiciels SAS et R.

Présentation visuelle des résultats (data visualization).

### 4. Résultats attendus

Le rapport du projet devra comporter une comparaison des résultats des différentes méthodes implémentées, une étude de l'effet des différents paramètres des méthodes, accompagnés d'éléments de réflexions sur la recommandation de l'utilisation de l'une ou de l'autre des méthodes aux différentes situations rencontrées.

### 5. Bibliographie

Candore G. Comparison of Statistical Signal Detection Methods Within and Across Spontaneous Reporting Databases Drug Saf (2015) 38:577–587

Caster, O., Norén, G.N., Madigan, D., Bate, A. Largescale regression-based pattern discovery: the example of screening the WHO global drug safety database. Stat. Anal. Data Min. (2010) 3 : 197–208.

Poluzzi E. Data Mining Techniques in Pharmacovigilance: Analysis of the Publicly Accessible FDA Adverse Event Reporting System (AERS). Data Mining Applications in Engineering and Medicine. Chapter 12

Wisniewski AFZ Good Signal Detection Practices: Evidence from IMI PROTECT. Drug Saf (2016) 39:469–490