

# Projets Licence

---

## Sujet 1 :

*Titre* : Classification binaire supervisée

*Résumé* : L'objectif principal de l'apprentissage statistique supervisé est la prédiction. On s'intéresse ici au problème fondamental de la classification (ou discrimination) binaire : on peut par exemple penser à la détection automatique de mails indésirables ou la catégorisation de photos sur un smartphone (présence ou non d'un visage, d'un objet donné, etc...). Après avoir saisi la différence entre apprentissage supervisé et non supervisé, les objectifs de ce travail sont de comprendre ce qu'est un problème de classification supervisée, de formaliser mathématiquement le problème et de présenter les idées des principaux algorithmes utilisés pour le problème de classification binaire. On verra également comment évaluer la qualité de prédiction et la fiabilité d'une méthode : quels critères de performance ? comment évaluer un algorithme ?

## Sujet 2 :

*Titre* : Des tests statistiques simples aux tests multiples

*Résumé* : En statistique, un test d'hypothèse est une procédure de décision entre deux hypothèses qui consiste à rejeter ou à ne pas rejeter une hypothèse (appelée hypothèse nulle) à partir d'observations.

Lorsque l'on dispose d'un grand nombre de données, on peut être amené à déterminer quels paramètres ont un rôle significatif dans nos observations. Par exemple, après avoir mesuré l'expression d'un ensemble de gènes chez un individu sain et chez un individu malade, on souhaite trouver les gènes qui interviennent dans la maladie. Pour cela, on peut réaliser un test multiple qui permet de tester plusieurs hypothèses nulles simultanément.

Dans un premier temps, on rappellera ce qu'est un test simple, les risques de décision et la façon dont on construit un test (non randomisé) en pratique. Dans un second temps, on s'intéressera aux problèmes de tests multiples et on considèrera en particulier les questions suivantes : qu'est-ce qu'une procédure de test multiple ? Comment construire une procédure de test multiple ? Quelles conséquences entraîne la multiplicité des hypothèses ? Existe-t-il "une" notion d'erreur de première et de deuxième espèce pour les tests multiples ? Ces questions nous permettront ensuite d'aborder le contrôle du FWER. Enfin, une partie numérique sera menée et consistera à appliquer des procédures de tests multiples sur des données simulées et éventuellement réelles.

*Références* :

1. Tests Statistiques, Rejeter, Ne pas rejeter... Se risquer ?, Cours de Magalie Fromont, [https://perso.univ-rennes2.fr/system/files/users/fromont\\_m/PolyTests.pdf](https://perso.univ-rennes2.fr/system/files/users/fromont_m/PolyTests.pdf)
2. Graph inference by multiple testing with application to neuroimaging, Chapitre 1 : Introduction, Manuscrit de thèse de Marine Roux, <https://tel.archives-ouvertes.fr/tel-01971574/document>
3. Why is multiple testing a problem ?, Cours de Megan Goldman, <https://www.stat.berkeley.edu/~mgoldman/Section0402.pdf>
4. (Plus difficile) Type I error rate control in multiple testing : a survey with proofs, Etienne Roquain, *Journal de la Société Française de Statistique*.