

**Objectifs et modes d'évaluation des projets.** L'objectif du projet est de programmer, en Python, la partie Statistique du cours de Probabilité et Statistique.

Le projet doit être fait en groupe de binôme ou de trinôme qui doit répondre de façon précise aux questions posées et fournir un *code commenté* pour chaque question.

La note du projet tiendra compte en particulier *de l'adéquation au travail demandé, de la qualité du code, de la précision des réponses aux questions posées.*

**Description du projet.** Vous devez fournir un exécutable qui présente un menu composé de deux parties: Statistique descriptive, Test d'indépendance du chi-deux.

Dans la partie *Statistique descriptive*, on doit pouvoir fournir *le nom* d'un fichier texte qui contient des valeurs numériques en colonne représentant les valeurs prises par une variable aléatoire, le *chemin d'accès* au fichier, le nombre de classes (pour l'histogramme), etc, et avoir en retour les quantités suivantes: nombre d'observations - la moyenne empirique - l'écart-type - le premier quartile - le troisième quartile - le kurtosis normalisé ou coefficient d'aplatissement - le skewness ou coefficient de dissymétrie - un histogramme.

Vous devez ensuite tester les programmes avec les données fournies: *notesproba.txt* et *notestat.txt*, qui contiennent les notes de Probabilité et de Statistique d'étudiants. Vous devez donc afficher et comparer les statistiques élémentaires précédemment citées pour les notes de Probabilité et de Statistique. En particulier, il faut: dire si le niveau des étudiants a été meilleur en Probabilité ou en Statistique - interpréter le kurtosis et le skewness en fonction des notes - interpréter l'histogramme associé à chaque variable - vérifier la cohérence des kurtosis et skewness trouvés avec l'histogramme associé à chaque variable.

Dans la partie *Statistique Inférentielle*, on considère qu'on est dans le cadre d'un modèle gaussien: les observations  $x_1, \dots, x_n$  sont des réalisations de la suite de variables aléatoires  $X_1, \dots, X_n$ , avec  $X_i \sim \mathcal{N}(\mu, \sigma^2)$  pour  $i = 1, \dots, n$ . On doit pouvoir fournir *les noms* de deux fichiers textes qui contiennent des valeurs numériques (prises par deux variables aléatoires en colonne: les fichiers *notesproba.txt* et *notestat.txt*), leur *chemin d'accès*, le niveau de confiance  $1 - \alpha$  de l'intervalle de confiance, le niveau  $\alpha_1$  du test de comparaison de la moyenne :  $H_0 = \mu = \mu_0$  contre  $H_1 : \mu > \mu_0$  ou  $H_1 : \mu < \mu_0$  pour  $\sigma$  connu, le niveau  $\alpha_2$  du test d'indépendance du chi-deux, et retourne les quantités suivantes (en interprétant les résultats): une estimation de  $\mu$  et  $\sigma$  pour chaque variable, un intervalle de confiance de niveau  $1 - \alpha$ , la décision de rejet ou non de l'hypothèse nulle aux seuils resp.  $\alpha_1$  et  $\alpha_2$  dans le test de comparaison de la moyenne et le test d'indépendance du chi-deux des notes de Probabilité et de Statistique.