

TP2: Statistiques descriptives bivariées

Le Chenadec Gilles

August 30, 2022

1 Informations préliminaires (mais néanmoins importantes)

- Le TP est probablement noté. 1 ou 2 TPs de la séquence seront entièrement notés. La note finale de TP comprendra des points :
 - sur la réponse aux questions et l'argumentation des TPs notés;
 - sur la forme et la qualité de la rédaction des TPs notés;
 - sur le rendu en qualité et en ponctualité de tous les TPs.
- Il vous est demandé de commencer le compte-rendu de ce TP en séance et de le déposer dans Moodle, 1 semaine jour pour jour après la séance (si la séance a lieu, le lundi 7 de 8h10 à 12h15, la date limite est le lundi 14 à 12h15, un rendu après est à ranger dans la catégorie "risque").
- Vous devez déposer dans Moodle un **fichier PDF (Mettre dans le nom du fichier, les deux noms en cas de binome)** dans lequel vous reporterez tous vos résultats:
 - réponses textuelles aux questions,
 - si vous avez besoin de mettre des équations :
 - * word est équipé d'un éditeur d'équations
 - * utilisez l'appareil photo de votre smartphone si vous préférez le papier crayon
 - codes python,
 - figures avec indication des abscisses et ordonnées,
 - valeurs numériques des résultats des calculs, etc.
- Il est très fortement conseillé :
 - de **prendre le clavier** à tour de rôle pour progresser
 - de **justifier** la réponse aux questions posées
 - de **ne pas plagier** le code, les phrases ni du cours, ni des camarades (l'école est dotée du logiciel anti-plagiat compilatio qui compare les documents rendus par les étudiants et les documents disponibles sur internet).

2 Statistiques descriptives bivariées sur des données d'Iris

Une fois réalisés les graphiques pour chaque variable prise séparément, l'étude peut porter sur la relation entre deux variables. On parle de croisement de deux variables ou d'étude bivariée. L'analyse bivariée a pour but de comprendre la relation entre deux variables. Les moyens d'analyse diffèrent selon la nature des variables considérées.

2.1 Étude de la largeur du pétale en fonction de la longueur du pétale

Représentation graphique La représentation graphique liant deux variables quantitatives est le nuage de points ou l'histogramme 2D.

Question 2.1: Tracer le nuage de points de la longueur du pétale en fonction de la largeur du pétale pour les 150 iris contenus dans les données ((numpy.)plot, (numpy.)scatter). Ne pas oublier de mettre des titres sur les axes. Décrire le nuage de points.

Coefficient de corrélation et droite de régression

Question 2.2: Rappeler la définition du coefficient de corrélation et le calculer par la fonction `((numpy.)corrcoef)`.

Question 2.3: Donner l'équation de la droite de régression linéaire, créer une fonction permettant de la calculer à partir de deux variables X et Y et tracer la sur le même graphique

Question 2.4: Analyser le lien entre les deux variables.

2.2 Étude de la longueur de pétale selon les différentes espèces

Nous abordons ici l'analyse des liens entre une variable qualitative et une variable quantitative. Les représentations graphiques permettant de lier une variable qualitative et une variable quantitative sont l'histogramme et la boîte à moustaches.

Représentations graphiques

Question 2.5: Représenter sur une même figure, les trois histogrammes de la longueur de pétale: un histogramme pour chaque espèce avec une couleur. Commenter cette figure.

Question 2.6: Représenter sur une même figure, une boîte à moustache par espèce. Commenter cette figure.

Mesures de corrélation

Question 2.7: Calculer le rapport de corrélation lié à la décomposition de la variance en variance intra-classe et interclasse. Qu'en concluez-vous?