

TP1 : Statistiques descriptives univariées

Le Chenadec Gilles

27 août 2018

1 Identité (1 point :-)

NOM :

PRENOM :

Si binôme, il y a (ni trinôme ni plus-nôme)

NOM :

PRENOM :

2 Informations préliminaires (mais néanmoins importantes)

- Le TP est probablement noté.
- 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 Word (**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 matlab (ctrl-C dans l'éditeur matlab, ctrl-V dans Word),
 - figures (menu Edit d'une figure Matlab puis copy figure puis ctrl-V dans Word) 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 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).
- **Pour vous aider :**
 - Les enseignants sont là à votre disposition (en cours ou par mail!)
 - Cours Matlab sur Moodle :
 - [cours](#)
 - [fichier](#)

3 Statistiques Descriptives univariées sur des données d'Iris

Les techniques de statistiques descriptives vont être illustrées via l'analyse des données célèbres, collectées par Edgar Anderson. Il s'agit des mesures en centimètres des variables suivantes : longueur du sépale (sepalLength), largeur du sépale (sepalWidth), longueur du pétale (petalLength) et largeur du pétale (petalWidth) pour trois espèces d'iris : setosa, versicolor et virginica.



FIGURE 1 – Iris Setosa, Versicolor, Virginica

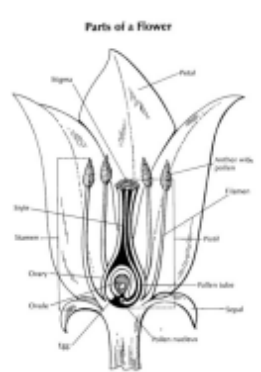


FIGURE 2 – Nomenclature des parties d'une fleur

3.1 Analyse préalable

Chargement des données

Question 1: Taper et comprendre les commandes suivantes :

```
1 clear
2 load('fisheriris.mat')
3 whos
```

Remarque : la variable meas est composé de quatre variables correspondant aux quatre colonnes : sepal length, sepal width, petal length et petal width.

Comprendre les données (signification des individus et des variables)

Question 2: Quelle commande vous permet de donner le nombre d'individus statistiques ?

Question 3: Trouver les variables qualitatives et leurs modalités associées (utiliser la commande unique). Sont-elles nominales ou ordinales ?

Question 4: Trouver les variables quantitatives. Sont-elles continues ou discrètes ?

Question 5: Séparer la matrice en quatre variables sepalLength, sepalWidth, petalLength, petalWidth

3.2 Etude de la variable species

Question 6: Quels sont les effectifs de chaque modalité ? (indication : il faut recoder les variables textes en numérique avec la commande strmatch par exemple pour éviter une boucle)

Question 7: Les représentations graphiques classiques liées aux variables qualitatives sont la représentation en secteurs ou camembert (pie), la représentation en bâtons (hist). Représenter ces graphiques.

3.3 Etude de la variable petalLength

Deux approches permettent d'étudier cette variable ; il s'agit de l'approche graphique et celle créant des résumés numériques. Se rapporter aux cours pour les définitions des outils proposés.

Première approche : graphique

Question 8: Tracer l'histogramme en fréquences relatives et l'histogramme des fréquences relatives cumulées (*histc*). Faire varier le nombre de classes de l'histogramme (cf. aide).

Question 9: Décrire les caractéristiques de l'histogramme et analyser ces caractéristiques en fonction du nombre de classes.

Question 10: Tracer la boîte à moustaches (*boxplot*) et rappeler les différents éléments la constituant.

Deuxième approche : résumés numériques

Question 11: Calculer les résumés numériques de localisation (moyenne et médiane) et ceux de dispersion : (écart-type, variance et quartiles). Retrouver en particulier, les valeurs des éléments de la boîte à moustache.