

Statistique descriptive unidimensionnelle

Un serveur a enregistré des informations sur chaque pourboire qu'il a reçu sur une période de quelques mois en travaillant dans un restaurant. Les données collectées sont disponible [ici](#) et une description de celles-ci est disponible [ici](#). L'objectif de ce TP est de faire un analyse Statistique descriptive sur ces données avec Python. À cette fin, nous allons utiliser les modules Pandas, Matplotlib et numpy, que vous pouvez installer avec l'instruction suivante :

i | `pip3 install Nom du module`

Partie 1

1. Importer le jeu de données sous forme de DataFrame à l'aide de l'instruction

i | `pandas.read_csv("url_des_données")`

2. Vérifier qu'il n'y a pas de valeurs manquantes dans le DataFrame.
3. Quel est le nombre de variables et le nombre d'observations (dimension du DataFrame)?
4. Donner le type de chaque variable.

Partie 2

1. Décrire toutes les variables quantitatives de ce jeu de données sous forme de tableau (effectif, min, max, moyenne, médiane,...).
2. Quel est le montant maximal du pourboire donné au serveur?
3. Construire le tableau Statistique de la variable Smoker qui contient en colonnes les : effectifs (n_i) et les fréquences (f_i).
4. Entre les fumeurs est les non-fumeur, qui donnent le plus de pourboires (en terme de nombre pas de montant laissés)?
5. Quel est le nombre des pourboires donnés par les femmes non fumeuses?

Partie 3

1. Tracer la courbe de densité de la variable total_bill.
2. Pensez-vous que cette variable suis une loi normale? Justifier votre réponse.
3. Le graphique précédent est-il adapté pour représenter la variable time? Justifier votre réponse.

4. Donner une représentation graphique, de votre choix, de cette dernière variable.
5. Tracer le diagramme en secteur de la variable sex.
6. Tracer et commentez la boîte à moustache de la variable tip.