

Analyse descriptive

1 Rappels

1.1 Variables quantitatives et qualitatives

Donner un exemple de variable :

- quantitative discrète;
- quantitative continue;
- qualitative nominale;
- qualitative ordinale.

Pour chaque type de variable, quel type de représentation graphique pouvez utiliser si vous souhaitez comprendre la distribution des effectifs?

1.2 Exemple

53 étudiants lancent chacun 2 dés simultanément. On enregistre à chaque essai la somme des 2 dés obtenue (12, 6, 5, 2, 11...). Les résultats sont résumés dans le tableau ci-dessous.

Quel est le type de la variable ‘somme des dés’? Représenter graphiquement ces données.

Somme des dés	2	3	4	5	6	7	8	9	10	11	12
nb étudiants	3	3	4	6	9	7	8	6	4	1	2

2 Statistiques descriptives sous Matlab

Une entreprise vend un produit P dans diverses villes de France classées en 3 régions. On s'intéresse aux volume des ventes (en euros).

2.1 Chargement des données

Importer le fichier *vente.csv* sous Matlab. Les données doivent être organisées sous le format Table.

Quels sont les types de la variable *Region* et *Vente*?

2.2 Indicateurs par région

Décrire les données par région (voir la fonction *grpstats*): donner l'effectif, le min, le max, l'étendue, la moyenne, écart type, médiane, l'intervalle interquartile. Résumer les résultats sous forme de tableau.

2.3 Représentation graphique

Représenter les ventes dans chaque région à l'aide d'une boîte à moustache (Boxplot). Changer l'échelle sur l'axe des ordonnées (voir *ylim*). Que représente les éléments de la boîte?

2.4 Interprétation

Commenter les données et les représentations graphiques. Quel indicateur de tendance centrale vous paraît le plus adapté pour étudier ces données?

2.5 Histogramme et distribution cumulée

1. Compter le nombre de villes dans chaque région. Représenter graphiquement ces données à l'aide d'un diagramme en secteur (*pie*).
2. Représenter l'histogramme des ventes dans la région 1 (*histogram*). On regroupera les ventes par intervalles de 500 euros. Quel est l'intervalle modal?

3. Représenter la fonction de répartition empirique (*ecdf*).
4. Les deux graphiques précédents sont-ils adaptés pour représenter des données quantitatives discrètes? Justifiez.