

TP-2 : Représentation graphique des séries univariées

La seconde séance est dédiée à la représentation graphique des données : à l'utilisation des diagrammes en bâtons, des histogrammes, des boîtes à moustaches et des fonctions des fréquences cumulés sous R.

1 Diagramme en bâtons

On rappelle que les diagrammes en bâtons conviennent aux séries qualitatives (ou non numériques) et aux variables quantitatives discrètes.

On a relevé pour vingt individus, certaines caractéristiques biologiques et sociologiques. Le tableau ci-dessous résume les résultats obtenus :

Individu	Groupe sanguin	Âge	Nombre d'enfants
1	B	38	3
2	AB	42	4
3	A	16	0
4	A	51	2
5	O	28	1
6	A	34	1
7	O	21	1
8	A	54	4
9	AB	35	0
10	A	22	3

Individu	Groupe sanguin	Âge	Nombre d'enfants
11	A	12	0
12	B	40	2
13	O	41	1
14	O	31	2
15	A	25	1
16	AB	64	6
17	A	60	1
18	O	55	0
19	B	49	2
20	A	42	2

1. Quel est l'effectif total ?
2. Quel est la nature de la variable "Nombre d'enfants" ?
3. Affectez à la variable `NbEnf` le vecteur correspondant au nombre d'enfants.
4. À l'aide de la commande `table`, obtenez les effectifs correspondant à chaque modalité.
5. En utilisant la commande `barplot` tracez un diagramme en bâtons des effectifs pour la variable "Nombre d'enfants".
6. En utilisant les options de la commande `barplot` ajoutez un nom aux axes et un titre.
7. Quel est le mode de la variable "Nombre d'enfants" ?
8. Quel est la nature de la variable "Groupe sanguin" ?
9. En utilisant la commande `barplot` tracez un diagramme en bâtons des effectifs pour la variable "Groupe sanguin".
10. En utilisant la commande `barplot` tracez un diagramme en bâtons des fréquences pour la variable "Groupe sanguin".

11. À l'aide des options de la fonction `barplot`, ajoutez un titre, les modalités et le nom des axes sur votre diagramme en bâtons.
12. Quel est le mode de la variable "Groupe sanguin" ?

2 Fonction des fréquences cumulées

1. En utilisant les commandes `plot` et `ecdf`, tracez la fonction des fréquences cumulées pour la variable "Nombre d'enfants" ?
2. Peut-on tracer une fonction des fréquences cumulées pour la variable "Groupe sanguin" ?

3 Histogrammes

Les histogrammes permettent de représenter des données associées à une variable continue.

3.1 Réserves de pétrole

On reprend l'exercice du TD1 sur les 20 plus grandes réserves de pétrole. On rappelle que les valeurs des 20 plus grandes réserves de pétrole (en milliards de barils) sont les suivants :

297.7, 268.35, 174.4, 157.3, 140.3, 104, 97.8, 80, 48, 37.2, 30.5, 30, 25.5, 25.3, 13.9, 13.2, 10.4, 10.4, 9.9, 7.0

1. En utilisant la fonction `hist`, tracez un histogramme des effectifs pour les 20 plus grandes réserves de pétrole. Comparez les résultats avec ceux obtenus au TD1.
2. Obtenez-vous un histogramme des effectifs ou un histogramme des densités de fréquence ? Ajoutez une option vous permettant de choisir entre un histogramme des effectifs ou des densités de fréquence.
3. Quelle est la classe modale ?
4. Commentez la forme des données.
5. En utilisant une autre option de la fonction `hist`, prenez pour classes les intervalles suivants $[0; 30[$, $[30; 60[$, $[60; 100[$, $[100; 160[$, $[160; 300[$.
6. Quelle est la classe modale ?

3.2 Données géologiques

On reprend maintenant les données de la librairie `geoR` de R.

1. Charger la librairie `geoR`.
2. En utilisant la fonction `hist`, tracez l'histogramme des effectifs correspondant à la colonne `mg020` de `camg`.
3. Quelle est la classe modale ?
4. Commentez la forme des données.

5. En utilisant les options de la fonction `hist`, tracez l'histogramme des densités, puis imposez pour limite inférieure 10 et pour limite supérieure 40, puis remplissez l'histogramme de la couleur grise et enfin mettez y le titre "Histogramme de mg020".

4 Boîtes à moustaches

Nous souhaitons tracer des boîtes à moustaches.

1. En utilisant la fonction `boxplot`, tracez la boîte à moustaches correspondant aux résultats à l'examen du groupe 2 $d = (2, 4, 11, 16, 18)$. Comparez le résultat avec celui obtenu en cours. Affichez les valeurs permettant de tracer la boîte à moustache et comparez les à ceux du cours.
2. On considère à nouveau les 20 plus grandes réserves pétrolières. Tracez la boîte à moustaches et comparez le résultat à celui obtenu au TD1. Affichez les valeurs permettant de tracer la boîte à moustache et comparez les à ceux du cours.
3. Étudiez la variable `mg020`. Vous pouvez utiliser la commande `summary`.
4. Tracez la boîte à moustaches correspondant à la colonne `mg020` de `camg`. Commentez la forme des données.
5. À l'aide des différentes options de la fonction `boxplot`, dans premier temps tracez la boîte à moustaches à l'horizontale, puis imposez pour limite inférieure 10 et pour limite supérieure 50, puis mettez de la couleur grise à l'intérieur du rectangle et enfin mettez y le titre "Boîte à moustaches de mg020".
6. Que donne les commandes `b1$stats` (où `b1` est la boîte à moustaches que vous aviez obtenue à la question 3) et `boxplot.stats(camg$mg020)` ?
7. En utilisant la commande `par(mfrow=c(2,1))`, tracez la boîte à moustaches et l'histogramme de la colonne `mg020` de `camg` sur la même figure.
8. Affichez toutes les lignes de `camg` telles que la colonne `region` vaut respectivement 1, 2 puis 3. Calculez le nombre de lignes de `camg` telles que la colonne `region` vaut respectivement 1, 2 puis 3
9. Effectuez les boîtes à moustaches correspondant à la quantité de `mg020` associées à la région 1, 2 et 3. Mettez ces trois boîtes à moustaches côte à côte et mettez y le titre "Boîtes à moustaches de mg020 associées à chacune des trois `region` (14, 48 et 116 observations)".