
Colles série 2 : statistiques descriptives

Les données utilisées sont celles de `mtcars`

Sujet 1:

La variable `cyl` a-t-elle un effet sur `mpg` ?

1. Représenter sur un même graphique les boxplot de `mpg` selon `cyl` avec titre et labels ("2 paires", "3 paires" et "4 paires") sous chaque boxplot et une légende de l'axe des x : "nombre de cylindres", et un titre de votre choix. Y ajouter trois droites horizontales passant par les valeurs moyennes de chaque échantillon.
2. Interprétation d'une part de chaque boxplot individuel et d'autre part de leurs positions et formes relatives.

Sujet 2:

La variable `carb` a-t-elle un effet sur `mpg` ?

1. Représenter sur un même graphique les boxplot de `mpg` selon `carb` avec un titre et une légende de l'axe des x : "nombre de carburateurs" (on pourra utiliser `boxplot(mpg~carb), ...`). Y ajouter les droites horizontales passant par les valeurs moyennes de chaque échantillon.
2. Les échantillons des véhicules à 6 ou 8 carburateurs vous semblent-ils représentatifs ? Quelle interprétation proposer de la suite des quatre premiers boxplots ? Que penser de l'écart entre les moyennes de l'échantillon des moteurs à deux carburateurs ou à quatre ?

Sujet 3:

On se demande ici s'il y a un lien entre les variables `mpg` et `wt`.

1. Quelle est la nature de ces variables ? Tracer le nuage des points ayant pour abscisse `wt` et pour ordonnée `mpg` et lui donner un titre. Y ajouter la droite des moindres carrés en rouge. Les coefficients (`intercept` et `slope`) de cette droite sont fournis par la commande :
`lm(mpg wt)$coefficients`
(à condition que les vecteurs `mpg` et `wt` aient été préalablement construits).
2. Interpréter ce graphique en essayant de donner une réponse plausible à la question posée.

Sujet 4:

Le modèle gaussien est-il adapté à la variable `qsec` ?

1. Tracer l'histogramme de l'échantillon `qsec` en veillant à produire le graphique de sorte que la surface définie par l'ensemble des éléments de l'histogramme vaille un. Y ajouter un titre et une droite verticale rouge passant par la valeur moyenne de l'échantillon et une verte passant par la valeur médiane.
2. Y ajouter le tracé de la densité d'une loi normale en rouge. Les paramètres (`mu` et `sd`) seront calculés en fonction des données. Interpréter ce graphique.

Sujet 5:

Les variables **gear** et **am** sont-elles indépendantes ?

1. Calculer le tableau en effectifs des modalités croisées de **gear** (en lignes) et **am** (en colonnes). Calculer les deux distributions conditionnelles de la variable **gear** selon la modalité prise par **am**. Commenter.
2. Représenter ces deux dernières à l'aide d'un diagramme en barres où figureront côte à côte les deux distributions. Ajouter une légende pour l'axe des abscisses. Comparer et répondre à la question de départ.

Sujet 6:

On s'intéresse à la variable **carb**.

1. Quelle est la nature de cette variable ? Partitionner la fenêtre graphique en deux parties l'une au dessus de l'autre. Tracer dans la partie supérieure un diagramme représentant la distribution observée et y ajouter un titre. Tracer dans la partie inférieure la fonction de répartition empirique (aussi appelée distribution cumulée) avec un titre également.
2. Sur lequel de ces deux graphiques peut-on lire les quantiles empiriques ? Donner par exemple la valeur du premier décile (d'ordre 10%) par une lecture graphique.