

TP2: Statistiques Descriptives

Sonia Tiao

19/10/2020

1.Chargement de données

```
data(mtcars)
```

2.Infos sur le jeu de données

Q: Décrivez le jeu de données (nombre de ligne, colonne). A quoi correspondent les observations et les variables ?

```
str(mtcars)
```

3.Transformation de données

Q: A quoi correspond la variable “cyl” ? Convertir les éléments de cette colonne en type “factor”. Commentez.

4.Analyse univariée - Variable quantitative

Q: Pour la variable 'mpg', calculer:

1. min, max, l'intervalle des valeurs
2. la moyenne, la variance et l'écart type
3. la médiane
4. les quantiles
5. le nombre d'observations

Q: Tester la fonction summary sur:

- mtcars
- mpg

Commentez:

4-1.Représentation graphique

Q: Commentez:

```
boxplot(mtcars$mpg, col = c("pink"),main = "Boxplot pour la variable mpg (miles per gallon)", ylab = "Q
```

Q: Faire un boxplot pour les variables suivantes : mpg, qsec, wt

Mettre les 3 boxplots sur un même graphes, une légende, et 3 couleurs différentes(pour chaque boxplot)

Q: Faire un histogramme pour la variable mpg avec des légendes. Regardez les options de cette fonction hist, commentez:

4-2. La distribution des données mpg suit-elle une loi normale ?

Q: Comparer la distribution de l'échantillon avec la distribution théorique d'une loi normale à l'aide du QQ-plot. Pour cela, exécutez:

```
qqnorm(mtcars$mpg)
qqline(mtcars$mpg) #la droite sample quantiles = theoretical quantiles

shapiro.test(mtcars$mpg)
#Du résultat ci-dessus, la p-value > 0,05 indiquant que la distribution des données n'est pas significativement différente d'une loi normale
```

Q: Faites la même chose, en centrant et réduisant la variable mpg.
Indice pour réduire et centrer :

$$\text{NOUV_VAR} = (\text{VARIABLE} - \text{MOYENNE}(\text{VARIABLE})) / \text{ECARTTYPE}(\text{VARIABLE})$$

Conseil: stocker dans une variable “mpg.stand”.
Qu'observez-vous ?

5. Analyse univariée - Variable qualitative

Q: Quelle est la différence entre les deux tableaux?

```
table.cyl <- table(mtcars$cyl)
tableprop.cyl <- prop.table(table(mtcars$cyl))
```

Q: Exécuter et commenter le code suivant:

Q: Ajoutez-y des légendes au barplot et pie plot.

Q: Faire pareil avec tableprop.cyl

```
barplot(table.cyl)
pie(table.cyl)
```

6. Analyse bivariée - 2 Variables quantitatives

On choisit les variables : mpg (miles per gallon) et wt (poids des véhicules)

Q: Donner la covariance (cov) et la corrélation de pearson et de spearman (cor) des 2 variables quantitatives:

Q: Représentez le nuage de plot: mpg vs weight(wt).

Indice(étape):

1. plot(x=wt, y=mpg) avec x,y = coordonnées des points *OU* plot(wt ~ mpg) où le ~ indique “wt en fonction de mpg”.
2. Trouver la droite de régression avec la méthode des moindres carrés(modèle linéaire simple): lm(wt ~ mpg), commentez la commande summary(lm(wt ~ mpg)).
3. Ajouter la droite de régression au plot avec la fonction abline.

Q: Faire le test de corrélation entre les variables, en supposant H0 l'hypothèse de corrélation de pearson entre les variables:

Utilisez cor.test()

Commentez

7. Analyse bivariée - Variables qualitatives

Q: Faire la table de contingence pour les variables: cyl et am avec table() **Q:** Table compte les modalités, comment obtenir les proportions ?

Q: Obtenir le mosaïque plot cyl vs am et am vs cyl avec `mosaicplot()`

Q: Pour aller plus loin, transformer les données de la variable am:

Si 0, mettre “automatique”

Si 1, mettre “manuelle”

8. Analyse bivariable - Variable quantitative vs variable qualitative

Q: Calculer la moyenne de “miles per gallon” en fonction des 3 catégories de volumes de cylindres en exécutant:

```
tapply(mtcars$mpg, mtcars$cyl, mean)
```

Q: Faire de même en remplaçant la fonction `mean` par la fonction `summary`:

Q: Faire de même pour les variables mpg VS am:

Q: Réaliser un boxplot pour les variables mpg en fonction de cyl: