

Exercice1 :

Considérer les données suivantes de l'attribut *âge* écrites par ordre croissant :

13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.

- 1) Quelle est la *moyenne* de ces données ? Quelle est la *médiane* ?
- 2) Quel est le *mode* des données ? Commenter la modalité des données (i.e., bimodales, trimodales, etc.).
- 3) Pouvez-vous déterminer (approximativement) le premier quartile (Q_1) et le troisième quartile (Q_3) des données ?
- 4) Donner le *résumé des cinq nombres* des données.
- 5) Dessiner un *boxplot* des données.
- 6) Quelle est la différence entre un *quantile-quantile plot* et un *quantile plot* ?

Exercice2 :

Supposer que dans un hôpital, on ait effectué des tests sur le taux des graisses de 18 adultes sélectionnés au hasard et dont les résultats accompagnés de l'âge sont comme suit : age

Âge	23	23	27	27	39	41	47	49	50
%fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2
Âge	52	54	54	56	57	58	58	60	61
%fat	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7

- 1) Calculer la moyenne, la médiane et l'écart-type de l'âge et du taux de graisse %fat.
- 2) Dessiner les *boxplots* pour âge and %fat.
- 3) Dessiner un *scatter plot*.
- 4) Normaliser les deux variables avec la technique du z-score.
- 5) Calculer le *coefficient de corrélation* (Pearson's product moment coefficient). Ces deux variables sont-elles corrélées positivement, négativement ou pas du tout ?

Exercice 3

Considérer les notes suivantes d'un groupe de 10 étudiants pour les cours respectifs d'1 MODULE X et un module 'Y' :

	X				Y			
	TP1/10	Test1/10	EMD1/20	Final1/20	TP2/10	Test2/10	EMD2/20	Final2/20
1	7,75	5	13,5	13,13	7,25	6,50	16	14,88
2	6	6	11	11,50	7,75	5,50	11,5	12,38
3	6	4	7,5	8,75	7,87	6,25	10	12,06
4	6,17	4	5,5	7,84	7,75	4,00	11,5	11,44
5	8	5	11	12,00	6,62	6,50	13	13,63
6	7,75	2	10,5	10,13	7,50	7,50	12,5	13,31
7	6	4	5,5	7,75	7,37	3,50	12,5	11,75
8	6,75	5	9	10,38	7,62	7,50	5,5	10,19
9	5,67	6	6	8,84	8,37	4,00	8,5	10,06
10	6,5	6	9	10,75	7,37	5,50	17	15,44

La note 'Final' est obtenue comme suit : $\text{Final} = (\text{TP} + \text{Test} + \text{EMD})/2$

- 1) A-t-on besoin de normaliser les notes de TP, Test et EMD pour calculer la note 'Final' ? Pourquoi ? Si oui, procéder à la normalisation des données.
- 2) Que représente concrètement 'Final' par rapport aux notes de TP, Test et EMD ?
- 3) Pour chacune des notes Final1 et Final2, Calculer :
 - a. La moyenne
 - b. La médiane
 - c. Le modeQue peut-on en conclure?
 - d. Calculer approximativement le premier et le troisième quartile
 - e. Donner le résumé des cinq nombres pour ces données.
 - f. Dessiner une boxplot pour ces données.
- 4) Dessiner un q-q plot pour les deux variables Final1 et Final2. Que peut-on en conclure ?
- 5) Calculer le coefficient de corrélation pour les deux variables. Que peut-on en conclure ?

Rappel :

$$\text{Coefficient Corrélation} = (\Sigma (X * Y) - n * \bar{X} * \bar{Y}) / ((n-1) \alpha_x * \alpha_y)$$

Avec X et Y les moyennes des variables X et Y

Et aussi : α_x et α_y les écarts type des variables X et Y