

Test 1 sur le 'Data mining'

Considérer les notes d'un groupe de 10 étudiants en 'datamining' et 'méta-heuristiques' :

	Datamining				Méta-heuristiques			
	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 mode
 Que peut-on en conclure?
 - d. Calculer approximativement le premier et le troisième quartiles
 - e. Donner le résumé des cinq nombres pour ces données.
 - f. Donner une *boxplot* de ces données.
- 4) Dessiner un *scatter plot* et un *q-q plot* pour les deux variables Final1 et Final2.
- 5) Calculer le *coefficient de corrélation* (Pearson's product moment coefficient) pour les deux variables. Que peut-on en conclure?

Bon courage !