

Exercice 1**Shanon****8 points**

Un professeur souhaite étudier statistiquement les notes sur 10 obtenues par ses élèves à un examen. Les résultats sont donnés dans le tableau ci-dessous.

nom de l'élève	note x_i
mike	6
bob	5
evan	9
ruppert	3
aaron	3
steve	1
zachariy	9
logan	6
roger	5
don	6
bill	5
ron	6
karl	5
stefen	7
tom	5
neil	10
john	2
george	1
dave	1
tim	1

1. Quelles sont les différentes modalités possibles et celles observées ?
2. Construisez un tableau avec les fréquences et les fréquences cumulées des notes.
3. Tracez la fonction de répartition empirique des notes.
4. Déterminez graphiquement la médiane et les quartiles des notes
5. En déduire la DIQ
6. Dessinez la boîte à moustache associée **sur une feuille**. Selon vos calculs, y a-t-il une variable aberrante parmi ces notes ?
7. On souhaite connaître la valeur centrale de l'ensemble de nos observations $x_i, i = 1, n$. On modélise ce problème à travers une fonction coût J :

$$J(c) = \sum_{i=1}^n c \left(\log \frac{x_i}{c} - 1 \right)$$

- a) Donnez l'expression de c en fonction des $x_i, i = 1, n$ qui minimise le coût J .
- b) Application numérique : calculez c pour les données observées.

Exercice 2**l'ACP du décathlon****12 points**

Le but de cet exercice est d'analyser globalement le tableau de données X qui contient les performances réalisées par des athlètes lors de deux compétitions de décathlon. Les noms des athlètes sont disponibles dans une variable 'Athlete' et les noms des sports dans une variable 'Events'

1. Trouvez graphiquement la médiane de Tuckey du couple de variables 100m et 1500m.
2. Prétraitement
 - a) Pourquoi faut-il centrer et réduire les observations avant de faire une ACP ?
 - b) Centrez et réduisez les données en Matlab sans utiliser les instructions Matlab `mean` et `std`. et en prenant soin d'éliminer les variables inutiles.
3. Corrélations
 - a) Qu'est-ce que la matrice des corrélations ?

- b) Quelle est la différence entre la matrice des corrélations et la matrice de variance-covariance ?
- c) Calculez la matrice des corrélations sans utiliser l'instruction Matlab `corrcoef`.
4. Axes principaux
- a) Calculez les deux axes principaux les plus importants.
- b) Quelle est leur importance en pourcentage d'information ?
5. Individus
- a) Représentez les individus dans le plan des deux premiers facteurs principaux.
6. Variables
- a) Représentez les variables dans le plan des deux premiers facteurs principaux.
- b) Donnez une interprétation à cette représentation.
- c) Comment interprétez-vous la proximité entre Schoenbeck et Smirnov ?
- d) Comment interprétez-vous l'opposition entre Smirnov et Hardee ?

```
X = [10.85 7.63 13.64 2.09 49.46 14.54 41.43 4.85 62.38 24.79
10.64 7.88 15.19 2.15 49.05 13.99 47.21 4.75 67.23 35.06
10.31 7.76 15.43 2.18 49.02 14.02 53.22 4.60 61.59 59.43
10.69 7.93 15.26 2.03 49.74 14.56 47.85 5.15 58.92 32.49
10.70 7.41 15.32 2.14 48.91 14.27 45.81 5.10 63.45 59.74
11.19 6.83 14.33 2.03 48.39 14.55 45.26 5.20 61.55 12.14
10.68 7.63 15.57 2.07 47.78 14.33 44.53 5.00 64.16 41.60
10.84 7.37 15.09 1.91 49.54 14.62 40.92 4.55 68.16 30.12
11.02 7.12 15.05 1.98 48.78 14.75 47.34 4.57 56.51 27.62
10.82 7.37 14.80 1.97 47.25 14.24 43.24 5.00 66.34 39.91
11.07 7.53 14.20 2.00 48.56 14.56 41.00 4.90 50.70 25.90
10.69 7.78 14.14 1.97 47.18 14.37 44.16 5.55 69.10 35.59
10.54 7.90 16.78 2.04 48.08 13.73 48.33 4.90 72.32 37.20
11.15 7.26 14.12 2.10 49.98 14.25 44.81 5.10 56.16 36.68
10.82 7.59 14.14 2.01 48.99 14.20 44.24 5.06 67.63 31.76
10.64 7.24 15.72 2.18 48.13 14.87 45.64 4.65 63.57 23.13
10.60 8.07 13.60 2.04 48.34 13.47 45.04 4.80 66.86 31.41
10.99 7.59 14.75 2.06 49.83 14.16 44.56 5.15 62.27 29.93
10.69 7.48 15.14 2.12 46.72 14.35 43.78 4.60 64.03 29.31
10.66 7.42 15.67 2.03 48.65 13.97 46.55 5.20 56.55 30.27
11.19 7.26 14.90 2.03 49.07 14.39 46.38 4.70 58.05 33.14
10.66 7.48 13.96 2.01 49.52 14.14 41.38 4.60 64.20 26.56
10.50 7.26 16.05 2.11 47.63 13.82 49.70 4.90 60.32 35.09
11.09 7.35 12.70 2.06 49.86 14.56 39.66 5.00 58.80 19.62
10.64 6.96 14.55 2.01 47.99 13.95 44.04 4.60 60.20 31.99
11.25 7.24 14.42 2.09 48.82 14.59 40.25 4.80 63.94 25.44
10.60 7.24 14.39 2.01 49.86 14.25 41.82 5.00 56.93 50.77
10.61 7.34 14.64 1.96 48.20 14.48 45.84 5.00 61.82 20.12
10.98 6.71 15.26 2.11 48.22 14.87 46.26 3.90 56.40 44.46
10.84 7.59 13.34 2.13 48.25 14.20 43.99 5.10 58.95 55.65
11.14 7.52 15.49 2.06 51.64 14.54 46.50 5.00 60.43 43.10
10.83 7.16 15.16 1.96 51.21 15.37 45.13 4.90 59.31 03.03
11.05 7.52 14.22 2.00 50.13 14.48 44.65 5.50 69.01 26.81
11.63 6.89 15.10 2.10 50.31 15.18 46.85 4.80 66.20 26.70
10.93 7.57 14.11 1.90 49.23 14.88 40.81 5.00 68.50 30.82
10.72 7.45 14.57 1.92 48.10 13.83 46.12 5.00 65.37 26.81
11.00 7.32 14.19 1.88 48.80 14.78 44.60 5.05 59.44 21.06
10.77 7.58 14.34 1.98 48.25 14.97 43.07 4.70 64.68 25.71];

events = ['100m'; 'Long'; 'Shot'; 'High'; '400m'; '110m'; 'Disc'; 'Pole'; 'Jave'; '1500'];

Athlete = struct('name',{'Sebrle','Clay','Karpov','Macey','Warners','Zsivoczky','Hernu','Nool',...
'Bernard','Schwarzl','Pogorelov','Schoenbeck','Barras','Smith','Averyanov','Ojaniemi',...
'Smirnov','Qi','Drews','Parkhomenko','Terek','Gomez','Turi',...
'Lorenzo','Karlivans','Korkizoglou','Uldal','Bourguignon','Martineau',...
'O Brien','Thompson','Hingsen','Freimuth','Hardee','Eaton','Suárez','Schrader','Warner'});

>> events(1,:)

ans =

100m

>> Athlete(1).name

ans =

Sebrle
```

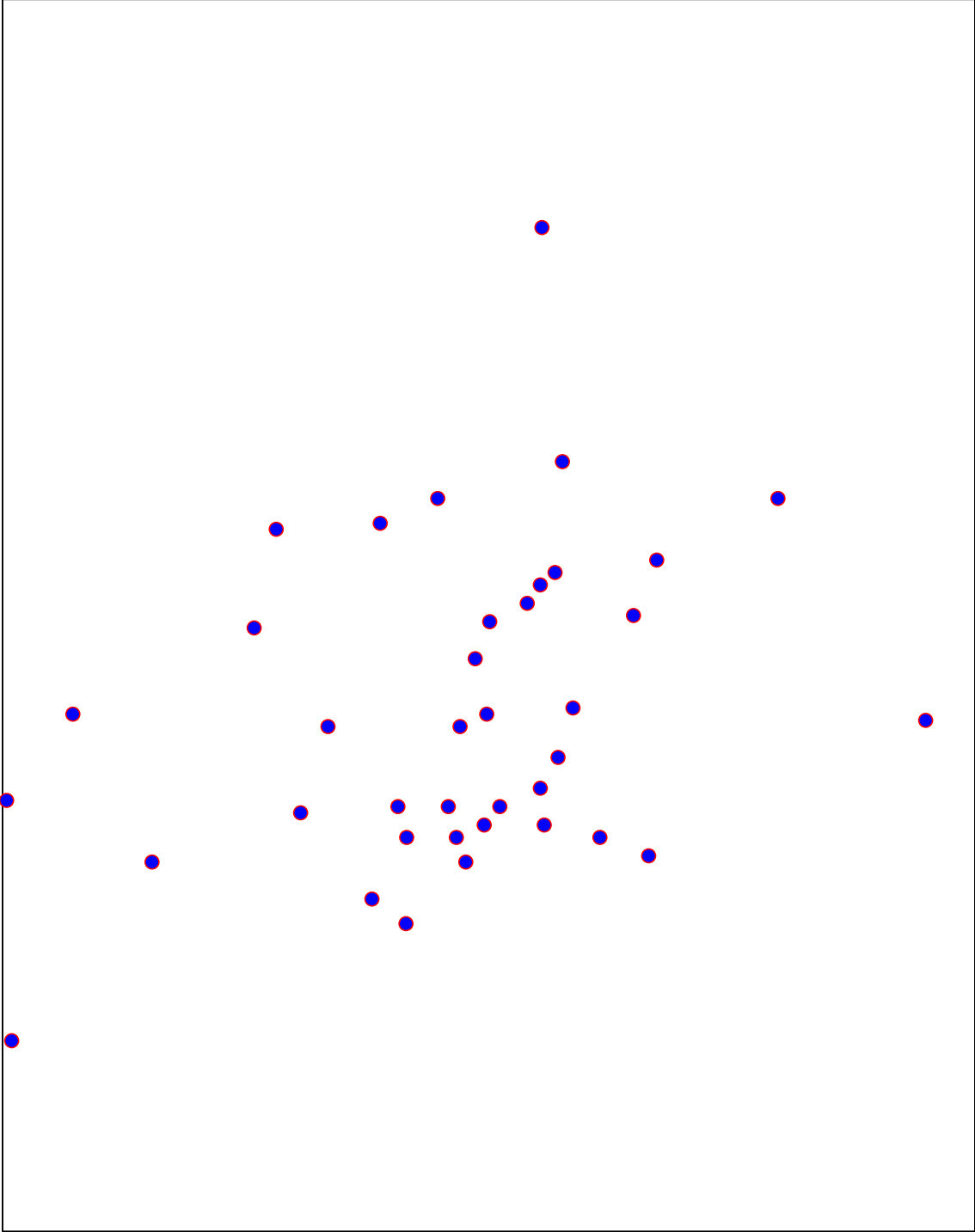


FIGURE 1 – Trouvez la médiane de Tuckey de ce nuage de points (100m vs 1500m).

Tracez ci-dessous la boîte à moustache des notes.