Série 1

Exercice 1. On a observé la couleur des cheveux et des yeux de certains étudiants de l'Université du Delaware. Les résultats [1] sont présentés dans le tableau suivant :

Cheveux / Yeux	Brun	Bleu	"Hazel" [2]	Vert
Noir	68	20	15	5
Brun	119	84	54	29
Roux	26	17	14	14
Blond	7	94	10	16

(a) Combien d'étudiants ont participé à cette enquête ?

(b) Est-ce que la couleur des cheveux est une variable quantitative, qualitative nominale ou qualitative ordinale ?

(c) Résumez ce que vous pouvez dire sur la couleur des cheveux des étudiants à l'aide de quelques chiffres et graphiques adéquats.

(d) Pensez-vous que la couleur des cheveux et la couleur des yeux des étudiants ayant participé à l'enquête sont liées ? Justifiez votre réponse.

Exercice 2. On a observé la taille et le poids de quelques femmes. Les résultats [3] sont présentés dans le tableau suivant (la taille en centimètres, le poids en kilogrammes) :

Taille	168	147	163	157	175	150	180	183	170	165	155	160	152	178	173
Poids	63	52	60	57	68	53	72	74	64	61	56	59	54	70	66

(a) Est-ce que la taille est une variable quantitative, qualitative nominale ou qualitative ordinale ?

(b) Calculez la moyenne et l'écart-type des tailles observées.

(c) Donnez les valeurs ordonnées des tailles observées (i.e. classez-les en ordre croissant).

(d) Pour les tailles observées, calculez le minimum, le maximum, la médiane, le quartile inférieur, le quartile supérieur et le quantile d'ordre 30 %. Qu'est-ce que représentent ces quantités ?

(e) Calculez l'écart inter-quartile des tailles observées.

(f) Construisez le boxplot des tailles observées. Quels aspects des données cette représentation graphique nous permet-elle de visualiser ?

(g) Construisez l'histogramme des tailles observées en utilisant $a = 150$cm et $h = 10$cm. Quels aspects des données cette représentation graphique nous permet-elle de visualiser ?

(h) Pensez-vous qu'il existe une relation entre la taille et le poids des femmes ayant participé à l'enquête ? Justifiez votre réponse à l'aide d'un graphique adéquat.

1. Les données sont tirées de Snee, R. D. (1974). Graphical display of two-way contingency tables. *The American Statistician*, **28**, 9–12.

2. "Hazel eyes" sont des yeux gris-verts avec le centre marron clair.

3. Les données sont tirées de *The World Almanac and Book of Facts*, 1975. En fait, ces données représentent la taille et le poids moyens de femmes américaines âgées entre 30 et 39 ans.

Exercice 3. Les tailles et les poids dans l'exercice précédent ont été initialement mesurés en pouces (1 pouce = 2.54 cm) et en livres (1 livre $\approx$ 454 g), respectivement. Quand on veut calculer certaines caractéristiques des distributions des données initiales (c.-à-d. en pouces et en livres), il n'est pas nécessaire de refaire tous les calculs ! On peut utiliser les propriétés suivantes :

Considérons des données $x_1, \ldots, x_n \in \mathbb{R}$, et $y_1, \ldots, y_n \in \mathbb{R}$ tels que $y_i = ax_i + b$ pour tout $i = 1, \ldots, n$. On a

(i) $\bar{y} = a\,\bar{x} + b$,

(ii) $med(y) = a\, med(x) + b$,

(iii) $s_y = |a|\, s_x$.

(a) Démontrez ces propriétés.

(b) Calculez la moyenne, la médiane, et l'écart-type des tailles de l'exercice précédent mesurées en pouces.

(c) On dit que la moyenne et la médiane sont des indicateurs de la tendance centrale de la distribution des données (des statistiques de lieu). Expliquez intuitivement la notion de la statistique de lieu (notez qu'on utilise aucune définition exacte). Vous pouvez soutenir votre explication avec les propriétés *(i)* et *(ii)*.

(d) On dit que l'écart-type est un indicateur de la dispersion de la distribution des données (statistique d'échelle). Expliquez intuitivement la notion de la statistique d'échelle (notez qu'on utilise aucune définition exacte). Vous pouvez soutenir l'explication avec la propriété *(iii)*.

(e) Est-ce que l'écart inter-quartile, les quantiles et l'étendue vous semblent être des statistiques de lieu ou d'échelle ?

Formulaire

Considérons des données $x_1, \ldots, x_n \in \mathbb{R}$.
Dénotons les données ordonnées $x_{(1)} \leq x_{(2)} \ldots \leq x_{(n)}$.
On va utiliser la définition suivante de la médiane :

$$med(x) = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{si } n \text{ est impair,} \\ \\ \frac{1}{2} \cdot \left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)} \right) & \text{si } n \text{ est pair.} \end{cases}$$