Feuille d'Exercices n° 1 : STATISTIQUES SIMPLES

Exercice 1.1 : *Compléments de cours*

Pour définir la variance, on avait choisi de calculer la moyenne des carrées des écarts par rapport à la moyenne ; le résultat suivant donne une bonne raison de faire ce choix.

Prouver que la fonction $g\ : t \longmapsto \frac{1}{n}\sum_{i=1}^{p} n_i(x_i - t)^2$ admet un minimum atteint en $t = \bar{x}$ (la moyenne de la série) et ce minimum vaut V (la variance de la série).

Exercice 1.2 : *Questions d'application*

1) Donner deux séries listées de même médiane mais de moyennes "très" différentes.
2) Donner deux séries listées de même moyenne mais de médianes "très" différentes.
3) Comparer la moyenne et la médiane des séries suivantes :

$(x_i) = 9,9,9,10,19,19,19$ et $(y_i) = 0,0,0,10,11,11,11$ d'un côté ;
$(z_i) = 0,0,0,0,20,20,20$ et $(t_i) = 0,0,0,15,15,15,15$ d'un autre côté.

4) Compléter les tableaux suivants et déterminer la médiane de chacune des trois séries suivantes :

x_i	2	3	4	5
n_i	20	30	35	15
N_i				

x_i	2	3	4	5
n_i	52	28	15	5
N_i				

x_i	2	3	4	5
n_i	12	13	22	53
N_i				

Que peut-on remarquer ?

5) Utiliser un changement de variable convenable pour donner "à la main" la variance de la série

$$(x) = (2000, 2001, 2002, 2003, 2004, 2005, 2006).$$

Exercice 1.3 : *Caractère quantitatif discret*

Les 35 étudiants d'une classe ont obtenu les notes suivantes à un contrôle donné :

9	11	8	5	15	9	14
7	11	5	6	8	6	10
4	11	8	8	7	13	4
13	5	5	6	4	10	5
7	11	4	6	9	7	4

1) Dresser et compléter le tableau de groupement par valeurs suivant :

	i=1	i=2	i=3	i=4	i=5	i=6	i=7	i=8	i=9	i=10	i=11	Sommes
x_i												*
n_i												
N_i												*
$n_i x_i$												
$n_i x_i^2$												

2) Donner la médiane de la série.

3)

a) Si la plus haute note passe à 18, la médiane change-t-elle ?

b) On a oublié de noter une question aux 5 étudiants qui ont 4. Si leur note passe à 6, la médiane change-t-elle ?

c) On relève toutes les notes de 3 points(on augmente chaque note de trois points). Quelle est alors la médiane ?

d) L'un des étudiants qui a obtenu 7 est exclu de la série. Que devient la médiane ?

4) Déterminer l'écart interquartile de la série initiale.

5) Déterminer la moyenne et l'écart-type de cette série.

Exercice 1.4 : *Taux de cholestérol : caractère quantitatif continu*

Le taux de cholestérol X est observé chez 220 personnes. L'unité de X est le en g/l (en gramme par litre de sang). On relève les résultats suivants :

Taux	$[1,6;1,8[$	$[1,8;2[$	$[2;2,2[$	$[2,2;2,4[$	$[2,4;2,6[$
Effectif	68	59	45	30	18

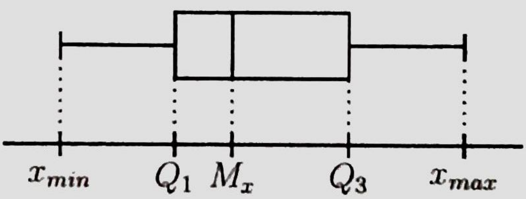
1) Faire un changement de variable convenable puis compléter le tableau statistique précédent(en rajoutant 6 lignes associées respectivement aux x_i, y_i, $n_i y_i$, $n_i y_i^2$, fréquences f_i, fréquences cumulées F_i, où x_i est le centre de la i ème classe et où $y_i = \frac{x_i - a}{b}$ avec a et b à choisir).

2) Représenter l'histogramme des fréquences cumulées croissantes de cette série statistique puis le polygone correspondant.

3) Déterminer le mode, la moyenne $\bar{x}$ et l'écart-type σ de cette série. Déterminer la médiane, les quartiles, l'écart interquartile et la dispersion à l'intérieur de l'intervalle interquartile.

4) Déterminer approximativement le pourcentage de personnes dont le taux de cholestérol est supérieur à 2.3 g/l.

5) On appelle **diagramme de Tukey** ou **diagramme en boite** ou **boîte à moustaches** ou encore **boîte à pattes** un diagramme comme ci-dessous sur lequel sont indiquées les caractéristiques suivantes : minimum, premier quartile, médiane, troisième quartile et maximum.



La boîte à moustaches, une traduction de « Box & Whiskers Plot », fut inventée en 1977 par le statisticien américain John Wilder Tukey (1915-2000) pour représenter schématiquement une distribution ou d'en comparer plusieurs entre elles.

Tracer la boite à moustache associée aux données de la statistique du taux de cholestérol.

Exercice 1.5 : *Sous-populations*

Soit X un caractère défini sur une population d'effectif N tel que $\forall w \in \Omega \; 0 \leq X(w) \leq 1$.
Soient Ω_1 et Ω_2 2 sous-populations d'effectifs N_1 et N_2 telles que $\Omega = \Omega_1 \cup \Omega_2$ et $\Omega_1 \cap \Omega_2 = \emptyset$.
Soient $X_1 = X_{|\Omega_1}$ et $X_2 = X_{|\Omega_2}$.

1) Montrer que : $\bar{X} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1+N_2}$ et $V(X) = \frac{N_1 V(X_1) + N_2 V(X_2)}{N_1+N_2} + \frac{N_1}{N}(\bar{X}_1 - \bar{X})^2 + \frac{N_2}{N}(\bar{X}_2 - \bar{X})^2$.

2) Montrer que $0 \leq \bar{X}_1 \leq 1$ et $0 \leq \bar{X}_2 \leq 1$, $|\bar{X}_2 - \bar{X}_1| \leq 1$ et $|\bar{X} - \bar{X}_1| \leq \frac{N_2}{N}$.

3) Montrer que $0 \leq V(X_1) \leq 1$ et $0 \leq V(X_2) \leq 1$, $|V(X_2) - V(X_1)| \leq 1$.

4) Montrer que $V(X) - V(X_1) = \frac{N_2(V(X_2) - V(X_1))}{N} + \frac{N_1}{N}(\bar{X}_1 - \bar{X})^2 + \frac{N_2}{N}(\bar{X}_2 - \bar{X})^2$.