

Feuille d'Exercices n° 5 : STATISTIQUES SIMPLES

Exercice 5.1 : 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 $g : t \mapsto \frac{1}{n} \sum_{i=1}^p n_i (x_i - t)^2$ admet un minimum atteint en $t = \bar{x}$ (la moyenne de la série) et ce minimum vaut V (la variance de la série).

Exercice 5.2 : Caractère quantitatif discret

Sur un échantillon de cent familles, on a relevé le nombre d'enfants par famille. Ce caractère discret est noté X , les résultats obtenus sont groupés dans le tableau suivant :

Valeurs x_i	0	1	2	3	4	5	6	7	Sommes
Effectifs n_i	5	15	23	22	16	9	5	5	

1) Dresser puis compléter le tableau précédent en rajoutant 3 lignes associées respectivement aux effectifs cumulés, les valeurs $n_i x_i$ et $n_i x_i^2$.

2) Tracer le diagramme en bâtons des effectifs et celui des effectifs cumulés.
Comment obtenir des diagrammes précédents les diagrammes en bâtons des fréquences et des fréquences cumulées ?

3) Déterminer les paramètres statistiques suivants : moyenne, médiane, mode, l'étendue, les quartiles, l'écart interquartile, la variance et l'écart-type.

Exercice 5.3 : Caractère quantitatif continu

Le taux de triglycérides X est observé chez 250 hommes de 20 à 30 ans. L'unité de X est le en g/l . On relève les résultats suivants :

Classes de X	[0.4; 0.6[[0.6; 0.8[[0.8; 1.0[[1.0; 1.2[[1.2; 1.4[[1.4; 1.6[
Effectifs n_i	5	32	86	89	32	6

1) Faire un changement de variable convenable puis compléter le tableau statistique précédent(en rajoutant 6 lignes associées respectivement aux x_i , y_i , fréquences, fréquences cumulées, les valeurs $n_i y_i$ et $n_i y_i^2$ 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.

4) Déterminer la médiane, les quartiles, l'écart interquartile et la dispersion à l'intérieur de l'intervalle interquartile.

5) Déterminer approximativement le pourcentage d'hommes dont le taux de triglycérides est supérieur à 1.3 g/l .

Exercice 5.4 : 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 5.5 : Sous-populations

Soit X un caractère défini sur une population d'effectif N tel que $\forall w \in \Omega \quad 0 \leq X(w) \leq 1$.

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 $V(X) = \frac{N_1V(X_1) + N_2V(X_2)}{N_1 + N_2} + \frac{N_1}{N}(\bar{X}_1 - \bar{X})^2 + \frac{N_2}{N}(\bar{X}_2 - \bar{X})^2$.

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

Exercice 5.6 : Comparaison de paramètres de position et de dispersion

Deux étudiants A et B ont obtenu au huit modules de la première année de la licence filière BCG les séries de notes suivantes : $A : (13, 14, 7, 13, 9, 7, 11, 12)$ et $B : (13, 13, 4, 10, 8.5, 7.5, 13, 17)$.

1) Calculer les modes, les moyennes, les médianes et les quartiles de ces deux séries, puis comparer ces résultats.

2) Calculer, pour chacune de ces deux séries, l'étendue, l'écart-interquartile, la variance et l'écart-type, commenter ces résultats.