

Exercices de Biostatistique

Fiche 2 : Caractéristiques des distributions à un caractère

Exercice 1

19 malades atteints d'un cancer du poumon ont été traité chirurgicalement (ablation du poumon atteint) dans un même service de chirurgie et suivis jusqu'à leur décès. La série des durées de survie, mesurées en semaines à partir de la date de l'intervention chirurgicale jusqu'à celle du décès, est la suivante :

25 ; 45 ; 238 ; 194 ; 16 ; 23 ; 30 ; 16 ; 22 ; 123 ; 51 ; 412 ; 162 ; 14 ; 72 ; 35 ; 30 ; 91 ; 45.

Dans l'étude de durée de survie, on définit la fonction de survie empirique S par $S(x) = 1 - F(x)$, où F est la fonction de répartition empirique.

- 1) Déterminer la médiane et les quartiles de la série statistique. Calculer l'écart interquartile.
- 2) Calculer la moyenne \bar{x} et l'écart type σ de la série.
- 3) Construire le diagramme de Tukey (boîte à moustaches), c'est-à-dire un diagramme qui visualise sur un axe l'étendue et les positions des 3 quartiles et de la moyenne.
- 4) Représenter graphiquement $F(x)$ et $S(x)$.

Exercice 2

Le tableau suivant indique la distribution du dosage d'un métabolite (en μg par litre) parmi une population de 300 sujets sains.

- 1) Calculer : la moyenne, le mode, la médiane, l'étendue, la variance, l'écart-type, le coefficient de variation, le coefficient de dissymétrie de Pearson.
- 2) Commenter brièvement ces résultats.
- 3) Représenter graphiquement la fonction de répartition.

Classe)	effectifs
[0.65; 0.85[22
[0.85; 0.95[42
[0.95; 1.05[63
[1.05; 1.15[75
[1.15; 1.25[54
[1.25; 1.35[30
[1.35; 1.55[14

Exercice 3

On considère une série statistique de 60 taux d'hémoglobine dans le sang (gr/l) mesurés sur des adultes présumés en bonne santé. La série est rangée par valeurs croissantes, les valeurs étoilées correspondent à des mesures chez les femmes :

105* ; 110* ; 112* ; 112* ; 118* ; 119* ; 120* ; 120* ; 125* ; 126* ; 127* ; 128* ; 130* ; 132* ; 133* ; 134* ; 135* ; 138* ; 138* ; 138* ; 138* ; 141 ; 142* ; 144 ; 145* ; 146 ; 148* ; 148* ; 148 ; 149 ; 150* ; 150 ; 150 ; 151* ; 151 ; 153, 153, 153 ; 154* ; 154* ; 154 ; 155, 156 ; 156 ; 158* ; 160 ; 160 ; 160 ; 163 ; 164 ; 164 ; 165 ; 166 ; 168 ; 168 ; 170 ; 172 ; 172 ; 176 ; 179.

1) Quelles informations peut on obtenir, sans calcul, en examinant les données de l'énoncé ?

On regroupe ces données par classes d'amplitude 10 ($[104;114[$,...).

2) Faites un tableau récapitulatif des données pour les 3 distributions (hommes, femmes, ensemble).

3) Calculez les moyennes et quartiles de ces distributions.

4) Déterminez les étendus de ces distributions.

5) Calculez l'écart interquartile pour chacune des trois distributions.

6) Tracez les diagrammes de Tukey (boîtes à moustache) des trois distributions.

7) Calculez les variances et les écarts-types de ces distributions.

8) Etudier la symétrie de ces distributions.

Exercice 4

On considère une cohorte de 70 individus extrait de la population mâle de 30 ans dans une région donnée du pays. On s'intéresse à la variable poids (en kg) que l'on observe aussi 10 ans plus tard lorsque les individus ont 40 ans. Le tableau suivant résume les résultats.

	30 ans	40 ans
Moyenne	75	82
Médiane	72	75
Mode	74	79
Ecart-type	6	10

1) Calculer : le coefficient de variation et le coefficient de dissymétrie de Pearson.

2) Commenter l'évolution de cette cohorte.