STATISTIQUES

Ph DEPRESLE

26 juin 2015

Table des matières

1 Vocabulaire

Document de base : Une entreprise Z emploie 52 personnes. On souhaite étudier le montant, en euros, de leur salaire mensuel net. Le service comptable fournit le tableau suivant :

Salaire mensuel net (en €)	1000	1200	1500	2500	3000
Nombre de personnes	5	8	24	13	2

- ◇ On veut étudier un ensemble appelé population dont les éléments sont appelés individus.
- ◇ On s'intéresse à certaines propriétés de ces individus, les caractères.
- ◇ L'observation du caractère étudié se traduit par une liste appelée série statistique.

Exemple : la population est l'ensemble des 52 personnes et le caractère est le montant du salaire mensuel net de chacune d'elles.

- ◇ L'effectif d'une valeur du caractère est le nombre d'individus dont le caractère prend cette valeur.
- ◇ L'effectif total est le nombre d'individus de la population étudiée.
- ◇ La fréquence d'une valeur du caractère est le quotient de l'effectif de la valeur par l'effectif total ; c'est un nombre compris entre 0 et 1.

Exemple : les valeurs du caractère sont 1000, 1200, 1500,2500 et 3 000.
L'effectif total de l'entreprise est égal à 52 (= 5 + 8 + 24 + 13 + 2).

L'effectif de la valeur 1200 est 8 et la fréquence de la valeur 1200 est $\dfrac{8}{52} \approx 0,154$
Cela signifie que environ 15,4 % des personnes gagnent 1200 €.

Définition 1. *Soit x une valeur du caractère. L'effectif cumulé croissant associé à x est le nombre de valeurs de la série inférieures ou égales à x.*

Exemple : Reprenons le tableau précédent et complétons le :

Salaire mensuel net (en €)	1000	1200	1500	2500	3000
Nombre de personnes	5	8	24	13	2
Effectif cumulé croissant	5	13	37	50	52
Fréquence cumulée croissante	9,62%	25%	71,15%	96,15%	100%

La ligne des effectifs cumulés croissants indique que 37 personnes ont un salaire mensuel inférieur ou égal à 1500€. La ligne des fréquences cumulées croissantes indique que 25% des personnes ont un salaire mensuel inférieur ou égal à 1200€.

2 Moyenne

Définition 2. *On note (x_i) ($1 \leqslant i \leqslant p$) les valeurs d'une série statistique, et n_i l'effectif relatif à la valeur x_i, N l'effectif total ($N = \sum_{i=1}^{p} n_i$). La moyenne, notée $\overline{x}$ est le nombre réel défini par :*

$$\overline{x} = \frac{\sum_{i=1}^{p} x_i n_i}{N}$$

.

Exemple 1 : Voici un relevé des 5 notes d'un élève : 12 ; 14 ; 15 ; 11 ; 18.
Moyenne des 5 notes :
$$m = \frac{12 + 14 + 15 + 11 + 18}{5} = 14$$

Exemple 2 : Calcul de la moyenne à partir d'un tableau d'effectifs.
Soit la série statistique suivante :

Rythme cardiaque	59	63	70	71	73	77	79	80	84	86	88	90
Effectif	2	1	5	2	5	4	2	2	4	5	1	1

L'effectif total est 34.

Moyenne du rythme cardiaque :
$$\overline{x} = \frac{59 \times 2 + 63 \times 1 + 70 \times 5 + \cdots + 86 \times 5 + 88 \times 1 + 90 \times 1}{34} = 76,71$$

3 Médiane

Définition 3. *La valeur médiane partage les valeurs d'une série statistique en deux groupes de même effectif :*

- *les valeurs inférieures ou égales à la valeur médiane,*
- *les valeurs supérieures ou égales à la valeur médiane.*

Exemple 1 : Un professeur a classé par ordre croissant les notes des 13 garçons et des 14 filles d'une classe.
$$\underbrace{7 \ldots 8 \ldots 9 \ldots 9 \ldots 10 \ldots 10}_{\text{6 notes}} \quad \underset{\uparrow}{11} \quad \underbrace{12 \ldots 13 \ldots 14 \ldots 14 \ldots 15 \ldots 17}_{\text{6 notes}}$$
11 (7e note) est la note médiane

Il a fait de même pour les filles :
$$\underbrace{6 \ldots 7 \ldots 9 \ldots 9 \ldots 10 \ldots 11 \ldots 12}_{\text{7 notes}} \quad \underbrace{13 \ldots 13 \ldots 13 \ldots 14 \ldots 14 \ldots 15 \ldots 15}_{\text{7 notes}}$$
12,5 (moyenne de la 7e et de la 8e note) est une note médiane

Exemple 2 : Détermination de la médiane des notes des 27 élèves à partir d'un tableau d'effectifs.

Note	6	7	8	9	10	11	12	13	14	15	17
Effectifs	1	2	1	4	3	2	2	4	4	3	1
Effectifs cumulés croissants	1	3	4	8	11	13	15	19	23	26	27

12 (14e note de la série classée en ordre croissant) est la note médiane.

3.1 Quartiles

3.1.1 Définition

Définition 4. *La médiane Me et les quartiles Q_1 et Q_3 partagent la population d'une série statistique suivant le principe suivant :*
25% des valeurs d'une série statistique sont inférieures ou égale au 1^{er} quartile
25% des valeurs d'une série statistique sont supérieures ou égale au 3^{eme} quartile.

Exemple : Soit la série statistique suivante :

Rythme cardiaque	59	63	70	71	73	77	79	80	84	86	88	90
Effectif	2	1	5	2	5	4	2	2	4	5	1	1

L'effectif total est 34.

Si on appelle $x_1, x_2, \ldots, x_{34}$ les valeurs de cette série rangées en ordre croissant, on a (l'effectif étant pair)

Me $= \dfrac{x_{17} + x_{18}}{2} = 77$

Les quartiles sont $Q_1 = 71$ et $Q_3 = 84$

3.1.2 Écart interquartile et étendue

Définition 5. *Soit une série statistique . On appelle écart interquartile la différence entre les quartiles Q_3 et Q_1 :*

$$E = Q_3 - Q_1$$

Il mesure la dispersion des 50% des valeurs qui entourent la médiane.

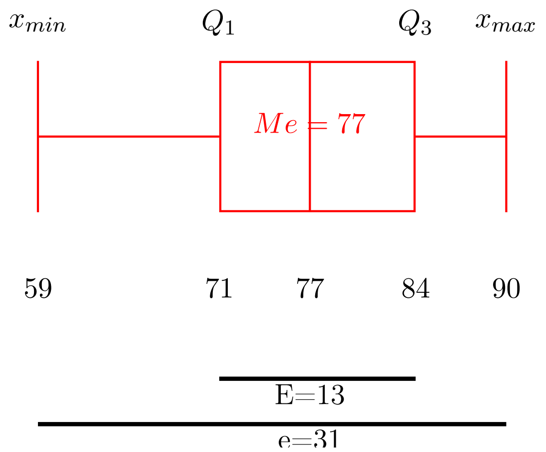
Définition 6. *x_{min} et x_{max} étant respectivement la plus petite et la plus grande valeur d'une série statistique, l'étendue est la différence des valeurs extrêmes :*

$$e = x_{max} - x_{min}$$

3.1.3 Représentation par un diagramme en boîte

Exemple : La série statistique précédente peut ainsi se représenter par le diagramme en boîte suivant :

Rythme cardiaque	59	63	70	71	73	77	79	80	84	86	88	90
Effectif	2	1	5	2	5	4	2	2	4	5	1	1



50% des valeurs sont situées dans la boîte

4 Échantillonnage

4.1 Échantillon

Définition 7. *Un échantillon de taille n est constitué des résultats de n répétitions indépendantes de la même expérience.*

Remarque :
Très souvent l'expérience consiste à choisir un individu de la population. Pour que les n expériences soient indépendantes il faudrait que les n tirages soient effectués avec remise.
Bien souvent, pour des raisons pratiques, il n'y a pas de remise lors du prélèvement. On peut néanmoins prouver que les résultats suivants restent vrais lorsqu'il n'y a pas remise après chaque prélèvement, pour peu que l'effectif total soit très grand par rapport au nombre d'éléments prélevés.
Exemples :

1. Dans une entreprise qui fabrique des lecteurs MP4, pour effectuer un contrôle dans la fabrication, on peut prélever un échantillon de taille 100. C'est-à-dire que l'on prend au hasard un des lecteurs pour vérifier si le logo est défectueux, puis on le remet. Et ceci 100 fois de suite.
2. Dans un établissement scolaire, pour étudier ce que les élèves pensent du restaurant, on peut adresser un questionnaire de satisfaction à 50 élèves. Pour cela, on prélève au hasard une fiche du fichier des élèves de l'établissement, on lui fait remplir le questionnaire de satisfaction, on note son nom et on remet la fiche. Et ceci 50 fois de suite.

4.2 Intervalle de fluctuation

Soit p la proportion de la population possédant le caractère étudié.
A chaque échantillon de taille n de cette population on associe la proportion f de l'échantillon vérifiant ce critère, appelée fréquence observée.

Définition 8.
On appelle intervalle de fluctuation au seuil de 95%, relatif aux échantillons de taille n, l'intervalle de centre p contenant, avec une probabilité égale à 0,95 la fréquence f observée.

Propriétés 1. *(Admis)*
Si p est compris entre 0,2 et 0,8 et $n > 25$, alors la fréquence appartiendra à l'intervalle $\left[p-\dfrac{1}{\sqrt{n}};p+\dfrac{1}{\sqrt{n}}\right]$ avec une probabilité d'au moins 0,95.
Cet intervalle est un intervalle de fluctuation de la fréquence au seuil d'au moins 95%.

Exemple :
Une urne contient 400 boules rouges et 600 boules blanches.
Pour un échantillon de taille 100, on compte le nombre de boules rouges de l'échantillon et on calcule la fréquence observée f.
Le prélèvement de cet échantillon peut être simulé à l'aide d'une calculatrice ou d'un ordinateur.
Une telle simulation permet d'« effectuer» des milliers de tels prélèvements et de vérifier que, dans 95% des cas, la fréquence (proportion) des boules rouges dans l'échantillon est située dans l'intervalle [0,3 ; 0,5] ce qui correspond bien à l'intervalle $\left[p-\dfrac{1}{\sqrt{n}};p+\dfrac{1}{\sqrt{n}}\right]$ où
$p = 0,4$ et $n = 100$.

5 Les exercices

1. On considère la série suivante :

valeur	1	5	13	17
effectif	2	1	3	2

   (a) Calculer la moyenne de la série.

   (b) Calculer la médiane et l'écart interquartile de la série.

2. On a recensé le nombre d'enfants vivant dans chacun des foyers d'une petite ville. On a résumé les résultats dans le tableau suivant :

Nombre d'enfants	0	1	2	3	4	5	6	7
effectif	290	170	155	95	43	27	20	10

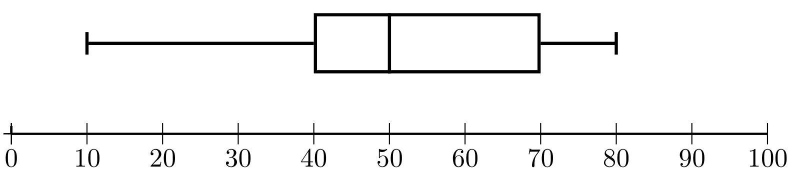
   (a) Calculer le nombre moyen d'enfants m par foyer.

   (b) Calculer l'écart type σ du tableau.

   (c) Calculer le pourcentage de foyers dont le nombre d'enfants appartiennent à l'intervalle $[m - \sigma; m + \sigma]$.

3. Dans une classe il y a 20 filles et 15 garçons. La taille moyenne de l'ensemble des élèves est de 1,7 m ; la taille moyenne des garçons est de 1,8 m.
   Quelle est la taille moyenne des filles de la classe ?

4. Le diagramme en boîtes d'une série est le suivant :



   (a) Déterminer la médiane et l'intervalle interquartille de la série.

   (b) Sachant que la population étudiée est d'un effectif égal à 72, combien d'individus ont une valeur du caractère comprise entre 50 et 70 ?

5. Un mélange de café se compose de 45 % d'arabica et de 55 % de robusta.
   L'arabica coûte 2 euros le kilogramme et le robusta coûte 1,80 euros le kilogramme.
   Calculer le prix du kilogramme du mélange arabica-robusta.

6. **QCM**

Questions	**Réponses**
1. La médiane est toujours comprise entre Q_1 et $Q-3$	☐ Oui ☐ Non ☐ Cela dépend
2. Quelle que soit la série statistique, la moyenne et la médiane sont égales.	☐ Oui ☐ Non ☐ Cela dépend
3. Il existe une série statistique de moyenne et médiane égales.	☐ Oui ☐ Non ☐ Cela dépend
4. Dans une population, on rencontre un certain caractre dans une proportion p=0,71. On prélève un échantillon de taille n =1000. L'intervalle de fluctuation de f au seuil de 95% est :	☐ [0,67 ;0 ;75] ☐ [0,68 ;0,74] ☐ [0,67 ;0,74]
5. Dans une population, on rencontre un certain caractre dans une proportion p=0,71. On prélève un échantillon de taille n =1000. L'intervalle de fluctuation de f au seuil de 95% est :	☐ [0,7 ;0,76 ☐ [0,68 ;0,74] ☐ [0,65 ;0,75]

6 Les exercices corrigés

1. (a) La moyenne est : $\dfrac{1\times 2+5\times 1+13\times 3+17\times 2}{2+1+3+2}=10.$

   (b) $M=\dfrac{13+13}{2}=13 \qquad Q_1=\dfrac{1+5}{2}=3 \qquad Q_3=\dfrac{1317}{2}=15.$
   L'écart interquartile est 12.

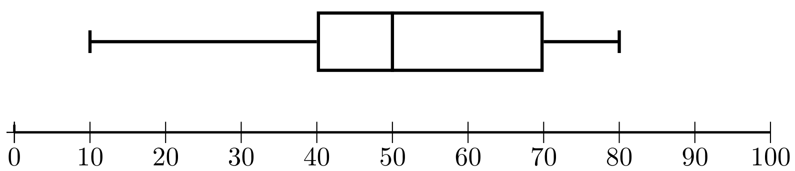
2. (a) La moyenne est :
   $$\frac{0\times 90+1\times 170+2\times 155+3\times 95+4\times 43+5\times 27+6\times 20+7\times 10}{290+170+155+95+43+27+20+10}\approx 1,56.$$

   (b) L'écart type σ est donné par $\sigma=\sqrt{V}$ où V est la variance. A l'aide de la calculatrice on obtient $\sigma\approx 0,72$.

   (c) L'intervalle $[m-\sigma; m+\sigma]$ correspond à $[1,56-0,72; 1,56+0,72)$ soit $[0,84; 2,28]$.
   on en déduit que $170+155$ soit 325 familles ont un ou deux enfants.
   325 familles ayant un ou deux enfants sur les 810 familles totales représentent $\dfrac{170+155}{810}\approx 0,40$.
   Donc 40% des foyers ont soit un, soit deux enfants.

3. Soit x la taille moyenne des filles. On doit avoir $\dfrac{20x+15\times 1,8}{35}=1,7$.
   La résolution de l'équation donne $x=1,625$.
   La taille moyenne des filles de la classe est de 1,625 m.

4. On a le diagramme suivant :



   (a) Par lecture graphique on a : $M=50; Q_1=40; Q_3=70$.

   (b) L'intervalle]40 ; 70[représente 25% de 72 donc 18 (25% de 72) individus ont une valeur du caractère comprise entre 50 et 70

5. Dans un kilo du mélange on trouve :
   450g d'arabica
   550g de robusta

   450g d'arabica coûtent : $0,450\times 2=0,90$ €
   550g de robusta coûtent : $0,550\times 1,80=0,99$ €

   Le kilogramme de mélange arabica-robusta coûte donc 1,89 € .

6. **QCM**

Questions	**Réponses**
1. La médiane est toujours comprise entre Q_1 et $Q-3$	■ Oui ☐ Non ☐ Cela dépend
2. Quelle que soit la série statistique, la moyenne et la médiane sont égales.	☐ Oui ☐ Non ■ Cela dépend
3. Il existe une série statistique de moyenne et médiane égales.	■ Oui ☐ Non ☐ Cela dépend
4. Dans une population, on rencontre un certain caractère dans une proportion p=0,71. On prélève un échantillon de taille n =1000. L'intervalle de fluctuation de f au seuil de 95% est :	■ [0,67 ;0 ;75] ☐ [0,68 ;0,74] ☐ [0,67 ;0,74]
5. Dans une population, on rencontre un certain caractère dans une proportion p=0,71. On prélève un échantillon de taille n =1000. L'intervalle de fluctuation de f au seuil de 95% est :	☐ [0,7 ;0,76 ☐ [0,68 ;0,74] ■ [0,65 ;0,75]