Statistique à une variable

1. Vocabulaire

L’ensemble sur lequel porte l’étude est la **population**, chaque élément de cet ensemble est un **individu statistique** ou une **unité statistique**.
Ce qu’on étudie est le **caractère**.
Un caractère est dit **qualitatif** lorsqu’il n’est pas mesurable (couleur, lieu,…).
Un caractère est dit **quantitatif** lorsqu’il s’agit d’une grandeur mesurable (prix, poids,...).
Nous ne nous intéresserons qu’à des caractères quantitatifs.
Un caractère quantitatif peut être **discret** (les nombres entiers) ou **continu** (les nombres réels).
La plupart des séries statistiques auront les valeurs de leur caractère regroupées en **classes** (il s’agit en général d’intervalles fermés à gauche et ouverts à droite).
Les nombres $\mathbf{n_i}$ d’individus correspondants à les valeurs $\mathbf{x_i}$ du caractère sont appelés **effectifs**.
Il est fréquente de noter $N=\sum n_i$ l’effectif total de la population étudiée.
Les réels $f_i=\dfrac{n_i}{\sum N}$ sont les **fréquences**, celles-ci peuvent être exprimées en pourcentages.

Exemple :

Une unité de production fabrique des axes de moteurs électriques, une étude statistique portant sur leurs longueurs en millimètres a donné les résultats suivants :

Longueurs des axes	[89,7 ; 89,8[	[89,8 ; 89,9[	[89,9 ; 90[	[90 ; 90,1[	[90,1 ; 90,2[	[90,2 ; 90,3[
x_i	89,75	89,85	89,95	90,05	90,15	90,25
n_i	3	14	36	33	13	1
f_i	0,03	0,14	0,36	0,33	0,13	0,01
Effectifs cumulés croissants	3	17	53	86	99	100
Effectifs cumulés décroissants	100	97	83	47	14	1

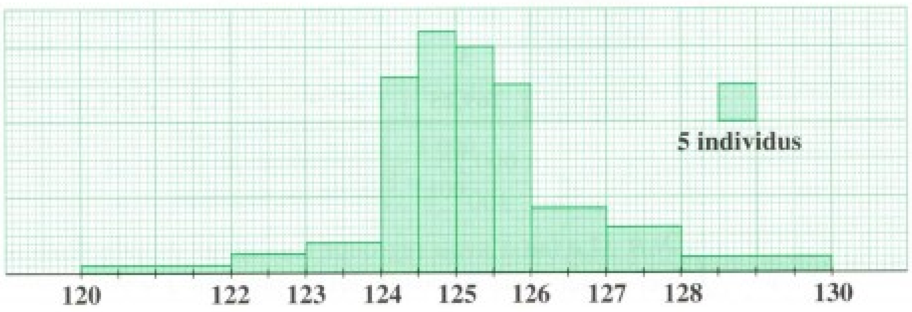
Dans cet exemple $N=100$.
Pour chacune des classes $[a_i , a_{i+1}[$ on peut calculer :
$n_1+n_2+...+n_i$ il s’agit des **effectifs cumulés croissants** ou,
$n_i+n_{i+1}+...+n_p$ (s’il y a p classes), il s’agit des **effectifs cumulés décroissants**.
Même travail avec les fréquences : **fréquences cumulées croissantes et décroissantes**.
Le nombre d’axes de longueur au plus égale à 90 mm est 53.
Le nombre d’axes de longueur supérieur à 89,9 mm est 83.

2. Représentation graphique : histogramme

On représente pour chaque classe un rectangle dont l'**aire est proportionnelle à son effectif**.
Si les classes ont toutes la même amplitude, il suffit de représenter des rectangles dont les hauteurs sont proportionnelles aux effectifs. Si non, il faut revenir à une amplitude de base :

Classes	[120, 122[	[122, 123[	[123, 124[	[124 ; 124,5[	[124,5 ; 125[	[125 ; 125,5[	[125,5 ; 126[	[126, 127[	[127, 128[	[128, 130[
Effectifs	4	5	8	26	32	30	25	17	12	8

On constate que les amplitudes des classes sont 0,5 ; 1 et 2. On peut considérer une amplitude de base égale à 0,5 , les classes d'amplitudes 1 étant considérées comme deux classes d'amplitude 0,5 et celles d'amplitude 2 comme quatre classes de base. On peut alors représenter l'histogramme.



3. Moyenne, variance et écarte type

La **moyenne** c'est de loin la **caractéristique de position** la plus utilisée, on la calcule grâce à la formule suivante :

À savoir

$$\overline{x} = \frac{\sum_{i=1}^{p} n_i x_i}{N} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{N}$$

(les x_i sont les valeurs du caractère, chaque x_i correspondant à un effectif n_i).

Exemple

Pour la série concernant les axes, on peut faire le tableau ci-dessous.

x_i	89,75	89,85	89,95	90,05	90,15	90,25	Totaux
n_i	3	14	36	33	13	1	100
n_ix_i	269,25	1 257,9	3 238,2	2 971,65	1 171,95	90,25	8 999,20

Donc $\overline{x} = \frac{8\,999,20}{100} = 89,992$.

N.B. : Le fait d'avoir regroupé les valeurs du caractère en classes ne permet pas de donner $\overline{x}$ avec cette précision. Ici on prendra $\overline{x} \approx 90$.

La variance et l'écarte type sont des **caractéristiques de dispersion**. On appel **variance** le nombre positif *V* défini par :

Définition

$$V = \frac{\sum_{i=1}^{p} n_i (x_i - \bar{x})^2}{N} \quad \text{où} \quad N = \sum_{i=1}^{n} n_i .$$

À savoir

On montre que $V = \frac{\sum_{i=1}^{p} n_i x_i^2}{N} - \bar{x}^2$.

À noter que la deuxième formule est plus pratique à utiliser, notamment avec l'outil informatique.

L'**écart type** est la racine carrée de la variance, on le note σ_x .

À savoir

$$\sigma_x = \sqrt{V}.$$

Exemple :

On fait un tableau dans lequel figurent les lignes pour x_i, n_i, $n_i x_i$ et $n_i x_i^2$. On obtient :

x_i	121	123	124,5	125,5	127	129	Totaux
n_i	2	8	15	23	8	4	60
$n_i x_i$	242	984	1 867,5	2 886,5	1 016	516	7 512
$n_i x_i^2$	29 282	121 032	232 503,75	362 255,75	129 032	66 564	940 669,5

$$\bar{x} = \frac{\sum_{i=1}^{6} n_i x_i}{N} = \frac{7\,512}{60} = 125{,}2.$$

$$V = \frac{\sum_{i=1}^{6} n_i x_i^2}{N} - \bar{x}^2 = \frac{940\,669{,}5}{60} - 125{,}2^2 = 2{,}785.$$

$\sigma_x = \sqrt{2{,}785} \approx 1{,}7$.

En réalité l'écart type de cette série est supérieur à 1,7 car le fait de regrouper les valeurs en classes nous a fait négliger la dispersion à l'intérieur de chacune des classes.

4. Comment déterminer les valeurs caractéristiques à l'aide d'une calculatrice

Exemple.

Déterminer les valeurs caractéristiques de la série « Diamètre intérieur d'un lot d'injecteurs ».

Diamètre (en mm)	[0,59 ; 0,61[	[0,61 ; 0,63[	[0,63 ; 0,65[	[0,65 ; 0,67[	[0,67 ; 0,69[
Effectif	8	16	37	25	14

Avec une calculatrice Casio Graph 35+

• *On tape* MENU STAT EXE *, on entre* ▶ *en tapant* F6 DEL-A *avec* F4 *,* YES *avec* F1 *.*
On entre les valeurs c_i (centre des classes) dans List 1 et les effectifs n_i dans List 2.

• *On sélectionne* ▶ *en tapant* F6 *,* CALC *par* F2 *puis* SET *par* F6*.*

• *On sélectionne List 1 sur la ligne* 1VarXList *avec* F1

et List 2 sur la ligne 1VarFreq *avec* F2 *puis* EXE *.*

• *On obtient les résultats en tapant* F1 *pour sélectionner* 1Var *.*

• *On lit : on lit $\overline{x}$ = 0,644 2 et σ_x = 0,022 323 97 (et aussi Q_1 = 0,64 ; Med = 0,64 ; Q_3 = 0,66).*

Avec une calculatrice TI 82 stats.fr ou 83 Plus

• *On tape* Stats *puis* 4 *.*

En face de Effliste *taper* 2nde 1 , 2nde 2 *(pour L1, L2) puis* ENTRER

• *Taper à nouveau* Stats *puis sélectionner* 1:Edite *.*

On entre les valeurs c_i (centre des classes) dans L1 et les effectifs n_i dans L2.

Taper Stats *.Sélectionner* CALC 1 *puis* Stats 1-Var *.*

On tape 2nde 1 , 2nde 2 *pour avoir L1, L2.*

On tape ENTRER *; on lit $\overline{x}$ = 0,644 2 et σ_x = 0,022 323 9.*

(Et aussi Q_1 = 0,64 ; Med = 0,64 ; Q_3 = 0,66.)

Exercices

13 Une entreprise de céramique a des saladiers parmi ses productions. Au laboratoire, on effectue le contrôle de l'épaisseur du bord du saladier à une hauteur de 80 mm. Les résultats obtenus réalisent une série statistique regroupée dans le tableau suivant.

Épaisseur (en mm)	Effectifs
[7,0 ; 7,2[	7
[7,2 ; 7,4[	14
[7,4 ; 7,6[	18
[7,6 ; 7,8[	12
[7,8 ; 8,0[	14
[8,0 ; 8,2[	5

1. Dans cette question, on considère que les effectifs de chaque classe sont rapportés au centre de cette classe.

a) Calculer l'épaisseur moyenne $\overline{x}$ du bord des saladiers, arrondie à 10^{-2} mm.

b) Calculer l'écart type σ de cette série statistique, arrondi à 10^{-2} mm.

2. a) Calculer les fréquences, arrondies à 10^{-2}, et les fréquences cumulées croissantes.

b) Représenter graphiquement le diagramme des fréquences cumulées.
Échelle : en abscisses, 1 cm pour 0,10 ;
en ordonnées, 1 cm pour 0,05.

3. Dans cette question, on suppose une répartition uniforme des effectifs dans chaque classe.
La machine est correctement réglée si 80 % des saladiers ont une épaisseur comprise dans l'intervalle $[\overline{x}-\sigma\,;\overline{x}+\sigma]$.
Calculer $\overline{x}-\sigma$ et $\overline{x}+\sigma$ avec les valeurs trouvées à la question précédente.
Déterminer graphiquement si la machine est bien réglée.
Les traits de rappel devront figurer sur le graphique.

14 **R** Une unité de production effectue le réglage d'une machine destinée à fabriquer en grand nombre des axes de moteurs électriques.
Un échantillon de 100 axes est prélevé lors des premiers jours de production, leurs longueurs étant mesurées (en mm), on obtient le tableau suivant.

Longueur des axes (en mm)	Nombre d'axes
[89,7 ; 89,8[	3
[89,8 ; 89,9[	14
[89,9 ; 90,0[	36
[90,0 ; 90,1[	33
[90,1 ; 90,2[	13
[90,2 ; 90,3[	1

En faisant l'hypothèse que, pour chaque classe, les valeurs observées sont égales à celle du centre de la classe, calculer (à 10^{-3} mm près) une valeur approchée de la moyenne $\overline{x}$ et l'écart type s des longueurs des axes de l'échantillon.

***Pour les exercices 15 à 17**, on calculera les valeurs demandées à l'aide d'une calculatrice.*

15 La répartition des âges dans une entreprise de 50 salariés est donnée dans le tableau suivant :

Âge	24	27	30	32	35	38	41	46	59
Effectif	5	8	8	7	4	3	6	6	3

1. a) Déterminer la médiane et l'écart interquartile de cette série.

b) Donner sa moyenne et son écart type.

2. Deux des trois salariés de 59 ans partent à la retraite. Comment les quatre paramètres précédents vont-ils être modifiés ?

16 On donne la série statistique suivante :

Valeur	18	20	21	23	24	25	500
Effectif	8	12	10	12	10	8	2

1. a) Déterminer la médiane et l'écart interquartile de cette série.

b) Donner sa moyenne et son écart type.

c) Que peut-on dire de la moyenne par rapport aux valeurs de la série ? Comment peut-on l'expliquer ?

2. Lequel des couples (médiane, écart interquartile) et (moyenne, écart type) est le plus adapté pour résumer cette série statistique ? Justifier la réponse.

17 La répartition du nombre total d'essais marqués par journée du Top 14 de rugby au cours de la saison 2010/2011, est donnée dans le tableau suivant :

Nombre d'essais	14	15	16	17	18	20	21	22	23
Nombre de journées	1	2	3	1	2	1	1	1	2

Nombre d'essais	24	25	26	28	29	30	34	35
Nombre de journées	2	4	1	1	1	1	1	1

1. a) Déterminer la médiane et l'écart interquartile de cette série.

b) Donner sa moyenne et son écart type.

2. a) Sur une calculatrice, représenter la série par un nuage de points.

b) Expliquer pourquoi la médiane est supérieure à la moyenne.