Lycée Lucie Aubrac - 2^{nde}

Chapitre 3
Statistiques

Table des matières

Dans l'ensemble de ce cours, on considère l'exemple de la série statistiques suivante, donnant les distances des planètes du système solaire au Soleil :

Planète	Distance au soleil en millions de km
Mercure	58
Vénus	108
Terre	150
Mars	228
Jupiter	778
Saturne	1429
Uranus	2871
Neptune	4503

1 Indicateurs de position

1.1 Moyenne

Définition 1

La **moyenne** d'une série statistique $x_1, x_2, \ldots, x_n$ se note $\overline{x}$ et est donné par la formule suivante :

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Exemple.
En moyenne, la distance d'une planète du système solaire au Soleil est d'environ 1266 millions de km.

$$\overline{x} = \frac{58 + 108 + 150 + 228 + 778 + 1429 + 2871 + 4503}{8} \simeq 1266$$

1.2 Médiane

Définition 2

La **médiane** d'une série statistique $x_1, x_2, \ldots, x_n$ est une valeur qui permet de couper l'ensemble des valeurs en deux parties égales : la moitié des valeurs étant plus grande que la médiane et l'autre moitié étant moins grande.

Remarque.

- Lorsque la série comprend un nombre impair de valeurs, la médiane est simplement une des valeurs de la série.
- Lorsque la série comprend un nombre pair de valeurs, la médiane est la moyenne entre les deux valeurs du « milieu ».

Exemple.
Dans le cas des distances planète-Soleil, la médiane est la moyenne entre la 4^{e} et la 5^{e} plus grande valeur étant donné que la série comporte 8 valeurs.

$$m = \frac{228 + 778}{2} = 503 \text{ millions de km}$$

1.3 Quartiles

Définition 3

- Le **premier quartile** d'une série statistique est une valeur, notée Q_1 telle qu'au moins 25% des valeurs soient plus petites que Q_1.
- Le **troisième quartile** d'une série statistique est une valeur, notée Q_3 telle qu'au moins 75% des valeurs soient plus petites que Q_3.

Exemple.
Dans le cas des distances planète-Soleil,
Q_1 correspond à la deuxième valeur (car $8 \times \frac{1}{4} = 2$) et Q_3 correspond à la sixième valeur (car $8 \times \frac{3}{4} = 6$).

$$Q_1 = 108 \qquad \text{et } Q_3 = 1429$$

Remarque.
L'intervalle interquartile est l'intervale $[Q_1\,;Q_3]$. Il contient donc 50% des valeurs de la série statistique.

1.4 Valeur minimale et maximale

Définition 4

La **valeur minimale** d'une série statistique est la plus petite valeur de la série. la **valeur maximale** est la plus grande valeur de la série.

Exemple.
Dans le cas des distances planète-Soleil, la valeur minimale est 58 et la valeur maximale est 4503.

2 Indicateurs de dispersion

2.1 Étendue

Définition 5

L'**étendue** d'une série statistique est la différence entre la valeur maximale et la valeur minimale.

Exemple.
Dans le cas des distances planète-Soleil, l'étendue est de $4503 - 58 = 4445$ millions de km.

2.2 Écart interquartile

Définition 6

L' **écart interquartile** est $Q_3 - Q_1$.

Exemple.
Dans le cas des distances planète-Soleil, l'écart interquartile est $1429 - 108 = 1321$ millions de km.

2.3 Écart-type

Définition 7

- La **variance** V d'une série statistique x_1, x_2, ..., x_n est définie par :
$$V = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2}{n}$$
- L'**écart type** σ d'une série statistique est définie par :
$$\sigma = \sqrt{V}$$

Remarque.
On retiendra que la variance est « la moyenne des carrés des écarts à la moyenne »

Exemple.
Dans le cas des distances planète-Soleil, on a $\overline{x} = 1266$. Ainsi,

$$V = \frac{(58 - 1266)^2 + (108 - 1266)^2 + (150 - 1266)^2 + \ldots + (4503 - 1266)^2}{8} \simeq 2304324$$

$\sigma = \sqrt{V} \simeq 1518$

Remarque.
La donnée de la médiane et de l'écart interquartile n'est pas sensible aux valeurs extrêmes, contrairement à la donnée de la moyenne et de l'écart-type.

3 Échantillonnage

Proposition 1 – (admise)

Si l'effectif d'une série statistique est suffisamment grand, et si le phénomène étudié suit une loi de Gauss (la répartition des valeurs se fait selon une courbe en cloche) :

- L'intervalle $\left[\overline{x} - \sigma\,;\overline{x} + \sigma\right]$ contient environ 68% des valeurs.
- L'intervalle $\left[\overline{x} - 2\sigma\,;\overline{x} + 2\sigma\right]$ contient environ 95% des valeurs.
- L'intervalle $\left[\overline{x} - 3\sigma\,;\overline{x} + 3\sigma\right]$ contient environ 99% des valeurs.

Exemples.
Les phénomènes suivants suivent une loi de Gauss :

- Pourcentage de Pile obtenus lors du lancer de 50 pièces ;
- La taille des femmes dans la population française ;
- Le nombre de votants par bureau de vote lors d'une élection ;
- La taille des tiges des coquelicots dans un champ ;
- Les notes d'une évaluation.

Ainsi, si la moyenne a une évaluation est $\overline{x} = 10$ et l'écart-type est $\sigma = 3$, cela signifie qu'environ 68% des notes sont comprises entre 7 et 13.

Proposition 2 – Loi des Grands Nombres (admise)

On répète n fois une même expérience aléatoire de façon indépendante et dont la probabilité de succès est p. Expérimentalement, plus n sera grand, plus la fréquence du succès se rapprochera de p. De plus, on peut considérer que la fréquence de succès suit une loi de Gauss et on a :

$$\sigma \simeq \sqrt{\frac{p(1-p)}{n}}.$$

Remarque.
En général, lorsque $n \geqslant 25$, on considère que n est suffisament grand pour utiliser la loi des Grands Nombres.

Exemple.
On lance 200 fois un dé équilibré à six faces et on compte le nombre de 6 obtenus. Déterminer avec une marge d'erreur de 5%, un intervalle contenant la fréquence de 6 obtenus.

Solution :
Les lancers étant indépendants les uns des autres, on peut utiliser la loi des Grands Nombres. On a $p = \dfrac{1}{6}$ donc :

$$\sigma \simeq \sqrt{\frac{\frac{1}{6}(1-\frac{1}{6})}{200}} \simeq 0{,}0264.$$

Par conséquent, cela signifie que l'intervalle suivant contient 95% des fréquences expérimentales :

$$\left[p - 2\sigma\,;p + 2\sigma\right] \simeq \left[\frac{1}{6} - 2 \times 0{,}0264\,;\frac{1}{6} + 2 \times 0{,}0264\right] \simeq \left[0{,}114\,;0{,}219\right].$$

Autrement dit, avec une marge d'erreur de 5%, l'intervalle $\left[0{,}114\,;0{,}219\right]$ contient la fréquence de 6 obtenus.

Savoir-faire du chapitre

- Connaître et savoir calculer les différents indicateurs de position (moyenne, médiane, quartiles, valeur minimale et maximale).
- Connaître et savoir calculer les différents indicateurs de dispersion (étendue, écart interquartile, écart-type).
- Comparer des séries statistiques en utilisant des indicateurs pertinent.