

Chapitre 1

Analyse de l'information chiffrée

Version professeur



Fil directeur du chapitre

Lire, organiser et interpréter des données portant sur deux caractères, puis utiliser un ajustement affine pour estimer une valeur inconnue.

Sommaire du chapitre

I	Des données pour comprendre une situation	2
II	Tableaux croisés d'effectifs	4
III	Nuage de points	7
IV	Point moyen d'un nuage	9
V	Ajustement affine	11
VI	Interpolation, extrapolation et esprit critique	13
VII	Méthodes essentielles	15
VIII	Utiliser un tableur	16
IX	Bilan du chapitre	17



Objectifs du chapitre

Dans ce chapitre, on apprend à :

- identifier une population, des individus et des caractères statistiques ;
- distinguer un caractère qualitatif et un caractère quantitatif ;
- lire, compléter et interpréter un tableau croisé d'effectifs ;
- représenter deux caractères quantitatifs par un nuage de points ;
- calculer les coordonnées du point moyen d'un nuage ;
- déterminer et utiliser un ajustement affine pour interpoler ou extrapoler ;
- garder un regard critique sur la validité d'un modèle.

CONNAÎTRE

REPRÉSENTER

CALCULER

MODÉLISER

RAISONNER

COMMUNIQUER



Automatismes transversaux

Les automatismes de l'épreuve anticipée ne sont pas tous liés au chapitre en cours. Ils seront travaillés régulièrement sous forme de questions flash : pourcentages, calcul numérique, lecture graphique, équations simples, probabilités élémentaires et interprétation d'une information chiffrée.

I Des données pour comprendre une situation



Pourquoi analyser des données ?

Les données chiffrées sont omniprésentes : évolution de la population, santé publique, climat, économie, sport, sondages, résultats d'enquêtes. Les mathématiques permettent d'organiser ces données pour éviter les impressions trompeuses et prendre des décisions argumentées.

Dans ce chapitre, on étudie surtout des données portant sur deux caractères. L'objectif n'est pas seulement de calculer, mais aussi de comprendre si deux informations semblent liées, et si un modèle simple peut aider à prévoir ou estimer une valeur.

I.1 Vocabulaire de base

Définition 1 (Population, individu et caractère)

Une population statistique est l'ensemble des objets ou personnes étudiés. Un élément de cette population est appelé un individu. Un caractère est une information observée sur chaque individu.

Définition 2 (Caractère qualitatif et caractère quantitatif)

Un caractère est qualitatif lorsque ses valeurs ne sont pas numériques ou lorsqu'elles correspondent à des catégories.

Un caractère est quantitatif lorsque ses valeurs sont des nombres sur lesquels on peut effectuer des calculs utiles.



Exemple

On étudie un fichier contenant des informations sur des élèves d'un lycée : âge, classe, moyen de transport, temps de trajet, nombre d'heures de sommeil.

- La population est l'ensemble des élèves du fichier.
- Chaque élève est un individu.
- La classe et le moyen de transport sont des caractères qualitatifs.
- L'âge, le temps de trajet et le nombre d'heures de sommeil sont des caractères quantitatifs.



Attention

Un numéro, un code postal ou un numéro de téléphone est écrit avec des chiffres, mais ce n'est généralement pas un caractère quantitatif : calculer une moyenne de numéros de téléphone n'a pas de sens.

I.2 Deux caractères étudiés simultanément



Idée essentielle

L'analyse bivariée consiste à étudier simultanément deux caractères sur une même population.

On peut chercher à répondre à des questions comme :

- les deux caractères semblent-ils liés ?
- peut-on organiser les données dans un tableau ?
- peut-on représenter les données par un graphique ?
- peut-on proposer un modèle pour estimer une valeur ?



Situations possibles

Deux caractères étudiés	Question possible
Age et temps d'écran	Le temps d'écran augmente-t-il avec l'âge ?
Température moyenne et consommation électrique	Les jours les plus froids conduisent-ils à une consommation plus forte ?
Taille et masse	Peut-on observer une tendance générale ?
Type de transport et ponctualité	Certains modes de transport sont-ils associés à davantage de retards ?

II Tableaux croisés d'effectifs

II.1 Cas de deux caractères qualitatifs

Définition 3 (Tableau croisé d'effectifs)

Un tableau croisé d'effectifs permet d'organiser les données lorsque l'on étudie deux caractères sur une même population. Chaque case du tableau contient le nombre d'individus correspondant simultanément à deux modalités.



Exemple guidé

Dans une enquête auprès de 120 élèves, on croise le moyen de transport principal et le fait d'être arrivé ou non en retard au moins une fois dans la semaine.

	Retard	Aucun retard	Total
Transport en commun	18	42	60
Marche ou vélo	6	34	40
Voiture	8	12	20
Total	32	88	120

On lit par exemple que 18 élèves utilisent les transports en commun et ont eu au moins un retard. On lit aussi que 60 élèves utilisent les transports en commun, et que 32 élèves ont eu au moins un retard.

II.2 Fréquences et interprétation

Définition 4 (Fréquence)

La fréquence d'une catégorie est le quotient :

$$\text{fréquence} = \frac{\text{effectif de la catégorie}}{\text{effectif total}}.$$

On peut l'écrire sous forme décimale, fractionnaire ou en pourcentage.



Exemple rédigé

Dans le tableau précédent, la fréquence des élèves ayant eu au moins un retard est :

$$\frac{32}{120} = \frac{4}{15} \approx 0,267.$$

Ainsi, environ 26,7% des élèves interrogés ont eu au moins un retard.



Comparer des groupes

Pour comparer deux groupes de tailles différentes, il faut souvent comparer des fréquences plutôt que des effectifs. Par exemple, 18 retards parmi 60 élèves ne s'interprète pas de la même manière que 8 retards parmi 20 élèves.



Comparer deux fréquences

On compare la proportion de retards selon le mode de transport.

$$\text{Transport en commun : } \frac{18}{60} = 0,30 = 30\%.$$

$$\text{Voiture : } \frac{8}{20} = 0,40 = 40\%.$$

Même s'il y a moins de retards en voiture en effectif absolu, la fréquence de retards est plus grande dans le groupe voiture.



Esprit critique

Un tableau croisé peut suggérer une relation entre deux caractères, mais il ne prouve pas forcément une relation de cause à effet.



Exemple type épreuve – tableau croisé et fréquences

Un vendeur possède un stock de 1 000 voitures. Les caractéristiques sont résumées dans le tableau ci-dessous.

	Blanche	Noire	Rouge	Total
Française	150	x	400	750
Étrangère	100	50	100	250
Total	250	250	500	1 000

Questions.

1. Indiquer ce que représente x et déterminer sa valeur.
2. Calculer le pourcentage de voitures noires dans le stock.
3. Calculer le pourcentage de voitures blanches parmi les voitures françaises.
4. Comparer le pourcentage de voitures noires dans tout le stock et le pourcentage de voitures noires parmi les voitures françaises.

Correction.

1. Le nombre x représente le nombre de voitures françaises noires. Dans la ligne Française, on a :

$$150 + x + 400 = 750.$$

Donc :

$$\begin{aligned} x &= 750 - 150 - 400 \\ &= 200. \end{aligned}$$

2. Il y a 250 voitures noires sur 1 000 voitures, donc :

$$\frac{250}{1\,000} = 0,25 = 25\%.$$

3. Il y a 150 voitures blanches parmi les 750 voitures françaises, donc :

$$\frac{150}{750} = 0,20 = 20\%.$$

4. Dans tout le stock, la proportion de voitures noires est 25%. Parmi les voitures françaises, elle vaut :

$$\frac{200}{750} \approx 0,267 = 26,7\%.$$

La proportion de voitures noires est donc légèrement plus élevée parmi les voitures françaises que dans l'ensemble du stock.

**Exemple type épreuve – lire puis justifier**

Dans un lycée comptant 2 000 élèves, on donne la répartition des effectifs suivant le sexe et le choix de la LV1.

	Fille	Garçon
Anglais	712	728
Autre LV1	288	272

Un élève affirme : Dans ce lycée, il y a autant de filles que de garçons. A-t-il raison ? Justifier.
On calcule séparément le nombre de filles et le nombre de garçons :

$$\begin{aligned}\text{Filles} &= 712 + 288 = 1\,000, \\ \text{Garçons} &= 728 + 272 = 1\,000.\end{aligned}$$

Il y a donc 1 000 filles et 1 000 garçons. L'affirmation est vraie.

III Nuage de points

III.1 Deux caractères quantitatifs

Définition 5 (Nuage de points)

Lorsque l'on étudie deux caractères quantitatifs x et y sur une même population, chaque individu peut être représenté par un point de coordonnées $(x ; y)$ dans un repère. L'ensemble de ces points s'appelle un nuage de points.

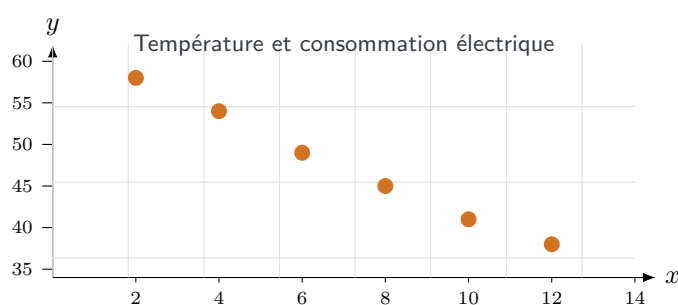
Exemple

On relève, pendant six jours, la température moyenne extérieure x en degrés Celsius et la consommation électrique y en MWh d'un bâtiment.

Jour	1	2	3	4	5	6
x : température	2	4	6	8	10	12
y : consommation	58	54	49	45	41	38

Les points du nuage sont :

$(2 ; 58)$, $(4 ; 54)$, $(6 ; 49)$, $(8 ; 45)$, $(10 ; 41)$, $(12 ; 38)$.



III.2 Lire une tendance

Tendance d'un nuage

Un nuage de points peut montrer une tendance générale :

- tendance croissante : quand x augmente, y a tendance à augmenter ;
- tendance décroissante : quand x augmente, y a tendance à diminuer ;
- absence de tendance claire : les points sont trop dispersés.

Interprétation

Dans le nuage précédent, lorsque la température augmente, la consommation électrique diminue globalement. On peut donc dire que le nuage présente une tendance décroissante. Dans ce contexte, cela peut s'expliquer par un besoin de chauffage moins important lorsque la température extérieure augmente.

**Point important**

Une tendance graphique donne une indication. Elle doit être interprétée avec prudence, surtout lorsque le nombre de données est faible ou lorsque les points sont très dispersés.

**Exemple à faire – du tableau au nuage de points**

On étudie l'évolution d'une grandeur au cours du temps. Le tableau donne la valeur y observée pour différentes valeurs de x .

x	0	2	4	6	8
y	12	17	21	26	31

1. Écrire les coordonnées des points du nuage.
2. Dire si le nuage semble présenter une tendance croissante, décroissante ou aucune tendance claire.
3. Interpréter cette tendance par une phrase.

Les points sont :

$(0 ; 12)$, $(2 ; 17)$, $(4 ; 21)$, $(6 ; 26)$, $(8 ; 31)$.

Lorsque x augmente, y augmente globalement. Le nuage présente donc une tendance croissante. Dans ce contexte, cela signifie que la grandeur observée augmente au cours du temps.

IV Point moyen d'un nuage

IV.1 Définition

Définition 6 (Point moyen)

On considère un nuage de n points :

$$M_1(x_1 ; y_1), M_2(x_2 ; y_2), \dots, M_n(x_n ; y_n).$$

Le point moyen G du nuage est le point de coordonnées :

$$G(\bar{x} ; \bar{y})$$

où

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad \text{et} \quad \bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n}.$$

Calcul du point moyen

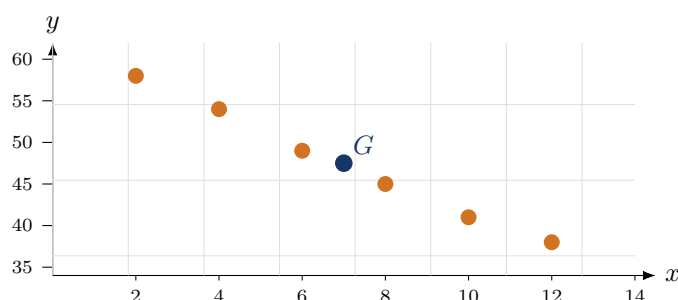
Avec les données précédentes :

$$\bar{x} = \frac{2 + 4 + 6 + 8 + 10 + 12}{6} = 7.$$

$$\bar{y} = \frac{58 + 54 + 49 + 45 + 41 + 38}{6} = 47,5.$$

Le point moyen du nuage est donc :

$$G(7 ; 47,5).$$



IV.2 Rôle du point moyen

Pourquoi calculer le point moyen ?

Le point moyen résume la position globale du nuage. Il est souvent utilisé pour construire une droite d'ajustement. Une droite d'ajustement pertinente passe généralement près du point moyen.

À retenir

Le point moyen n'est pas forcément l'un des points du nuage. C'est un point calculé à partir de toutes les données.

Exemple à faire – calculer un point moyen

On considère le nuage de points :

$$A(1 ; 9), \quad B(3 ; 12), \quad C(5 ; 16), \quad D(7 ; 19).$$

Calculer les coordonnées du point moyen G .

On calcule la moyenne des abscisses et la moyenne des ordonnées :

$$\begin{aligned}\bar{x} &= \frac{1 + 3 + 5 + 7}{4} = \frac{16}{4} = 4, \\ \bar{y} &= \frac{9 + 12 + 16 + 19}{4} = \frac{56}{4} = 14.\end{aligned}$$

Donc le point moyen est :

$$G(4 ; 14).$$

v Ajustement affine

V.1 Idée d'un ajustement affine

Définition 7 (Ajustement affine)

Lorsqu'un nuage de points est approximativement aligné, on peut le modéliser par une droite d'équation :

$$y = ax + b.$$

On parle alors d'ajustement affine.



Interprétation des coefficients

Dans le modèle affine $y = ax + b$:

- le coefficient a est le coefficient directeur de la droite ;
- il indique la variation approximative de y lorsque x augmente de 1 ;
- le nombre b est l'ordonnée à l'origine du modèle.

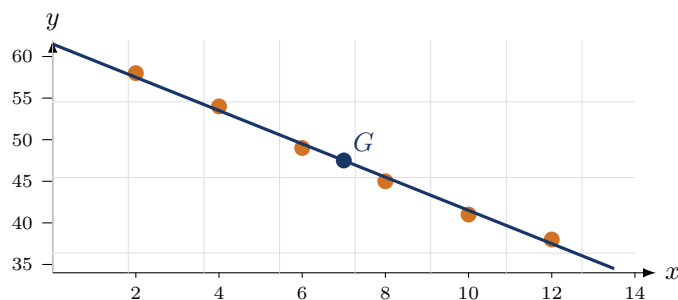


Modèle proposé

Pour le nuage précédent, on propose l'ajustement affine :

$$y = -2x + 61,5.$$

Ce modèle signifie qu'à chaque degré Celsius supplémentaire, la consommation électrique diminue en moyenne d'environ 2 MWh.



V.2 Déterminer une droite d'ajustement



Programme

Plusieurs méthodes d'ajustement peuvent être rencontrées : ajustement au jugé, droite de Mayer, droite des moindres carrés obtenue avec un outil numérique. Aucune connaissance théorique détaillée sur ces méthodes n'est exigible dans ce chapitre.



Méthode pratique

Pour exploiter un ajustement affine donné ou obtenu :

1. on écrit l'équation de la droite sous la forme $y = ax + b$;
2. on remplace x par la valeur connue pour estimer y ;

3. ou bien on résout une équation si l'on cherche la valeur de x correspondant à une valeur de y ;
4. on conclut en rappelant le contexte.

**Estimer une valeur**

Avec le modèle $y = -2x + 61,5$, on estime la consommation pour une température de 9°C :

$$\begin{aligned}y &= -2 \times 9 + 61,5 \\ &= 43,5.\end{aligned}$$

Pour une température de 9°C , le modèle prévoit une consommation d'environ 43,5 MWh.

**Exemple type épreuve – exploiter un ajustement donné**

On modélise la valeur y d'une grandeur en fonction du rang x par la droite d'ajustement :

$$y = 1,8x + 12.$$

1. Estimer la valeur de y pour $x = 10$.
 2. Pour quelle valeur de x le modèle donne-t-il $y = 39$?
 3. Expliquer ce que représente le coefficient 1,8 dans ce modèle.
1. Pour $x = 10$:

$$\begin{aligned}y &= 1,8 \times 10 + 12 \\ &= 18 + 12 \\ &= 30.\end{aligned}$$

Le modèle donne donc $y = 30$.

2. On résout l'équation :

$$\begin{aligned}1,8x + 12 &= 39 \\ 1,8x &= 27 \\ x &= \frac{27}{1,8} \\ x &= 15.\end{aligned}$$

Le modèle donne $y = 39$ pour $x = 15$.

3. Le coefficient 1,8 signifie que, selon ce modèle, lorsque x augmente de 1, la grandeur y augmente en moyenne d'environ 1,8 unité.

VI Interpolation, extrapolation et esprit critique

VI.1 Interpolation et extrapolation

Définition 8 (Interpolation)

Faire une interpolation, c'est estimer une valeur située à l'intérieur de l'intervalle des valeurs observées.

Définition 9 (Extrapolation)

Faire une extrapolation, c'est estimer une valeur située à l'extérieur de l'intervalle des valeurs observées.



Exemple

Dans le tableau des températures, les valeurs observées de x vont de 2 à 12.

- Estimer la consommation pour $x = 9$ est une interpolation, car 9 est entre 2 et 12.
- Estimer la consommation pour $x = 18$ est une extrapolation, car 18 est en dehors de l'intervalle observé.

VI.2 Validité d'un modèle



Un modèle n'est pas la réalité

Un ajustement affine est un modèle simplifié. Il peut être utile pour estimer, mais il ne doit pas être utilisé sans réflexion. Avant de conclure, il faut se demander :

- les points sont-ils assez proches d'une droite ?
- le nombre de données est-il suffisant ?
- l'estimation est-elle une interpolation ou une extrapolation ?
- le résultat a-t-il un sens dans le contexte ?



Extrapolation dangereuse

Avec le modèle $y = -2x + 61,5$, pour $x = 40$, on obtiendrait :

$$y = -2 \times 40 + 61,5 = -18,5.$$

Une consommation électrique négative n'a pas de sens. Le modèle ne peut donc pas être utilisé pour une température aussi éloignée des données observées.



Exemple à faire – choisir entre interpolation et extrapolation

Un ajustement a été construit à partir de données observées pour x compris entre 5 et 25. On utilise ce modèle pour estimer une valeur lorsque $x = 18$, puis lorsque $x = 40$.

Dire, dans chaque cas, s'il s'agit d'une interpolation ou d'une extrapolation, puis indiquer quelle estimation est a priori la plus prudente.

Pour $x = 18$, la valeur appartient à l'intervalle $[5 ; 25]$. Il s'agit donc d'une interpolation.

Pour $x = 40$, la valeur est en dehors de l'intervalle $[5 ; 25]$. Il s'agit donc d'une extrapolation.

L'estimation pour $x = 18$ est a priori plus prudente, car elle reste dans l'intervalle des valeurs observées.

**Phrase de conclusion utile**

On peut écrire : Le modèle permet une estimation raisonnable dans l'intervalle des valeurs observées, mais son utilisation en dehors de cet intervalle doit être justifiée avec prudence.

VII Méthodes essentielles

VII.1 Lire un tableau croisé



Méthode 1 – Lire et interpréter un tableau croisé

1. Identifier les deux caractères étudiés.
2. Repérer les modalités placées en lignes et en colonnes.
3. Lire l'effectif dans la case correspondant aux deux modalités.
4. Utiliser les totaux de ligne ou de colonne si nécessaire.
5. Comparer des fréquences lorsque les groupes n'ont pas la même taille.



Application courte

Dans le tableau des retards, la proportion d'élèves ayant eu un retard parmi ceux qui viennent à pied ou à vélo est :

$$\frac{6}{40} = 0,15 = 15\%.$$

VII.2 Calculer un point moyen



Méthode 2 – Calculer un point moyen

Pour un nuage de points $M_i(x_i ; y_i)$:

1. calculer la moyenne des abscisses ;
2. calculer la moyenne des ordonnées ;
3. écrire les coordonnées du point moyen $G(\bar{x} ; \bar{y})$.

VII.3 Utiliser un ajustement affine



Méthode 3 – Utiliser une droite d'ajustement

Si la droite d'ajustement a pour équation $y = ax + b$:

- pour estimer y , on remplace x par la valeur donnée ;
- pour estimer x , on remplace y par la valeur donnée et on résout l'équation ;
- on conclut toujours avec une phrase contextualisée.



Chercher une valeur de x

Avec le modèle $y = -2x + 61,5$, on cherche pour quelle température la consommation serait d'environ 50 MWh.

$$\begin{aligned} -2x + 61,5 &= 50 \\ -2x &= 50 - 61,5 \\ -2x &= -11,5 \\ x &= \frac{-11,5}{-2} \\ x &= 5,75. \end{aligned}$$

Selon ce modèle, une consommation de 50 MWh correspond à une température d'environ 5,75°C.

VIII Utiliser un tableur



Ce que le tableur permet de faire

Un tableur permet de :

- stocker des données dans un tableau ;
- calculer automatiquement des totaux, des fréquences ou des moyennes ;
- représenter un nuage de points ;
- ajouter une droite d'ajustement ;
- afficher une équation de droite d'ajustement.



Formules utiles

Objectif	Formule ou action possible
Calculer une moyenne	=MOYENNE(plage)
Calculer une somme	=SOMME(plage)
Obtenir une fréquence	=effectif/total puis format pourcentage
Créer un nuage de points	Insérer un graphique de type nuage de points
Afficher un ajustement affine	Ajouter une courbe de tendance linéaire et afficher son équation



Attention à l'arrondi

Une équation affichée par un tableur est souvent arrondie. Un arrondi trop grossier peut modifier les estimations. Il faut conserver suffisamment de décimales lorsque l'on utilise le modèle.

IX Bilan du chapitre

À retenir

- Un tableau croisé permet d'étudier deux caractères qualitatifs ou des caractères regroupés en classes.
- Pour comparer deux groupes de tailles différentes, on compare plutôt des fréquences.
- Un nuage de points représente deux caractères quantitatifs.
- Le point moyen d'un nuage a pour coordonnées les moyennes des abscisses et des ordonnées.
- Si un nuage est approximativement aligné, on peut utiliser un ajustement affine $y = ax + b$.
- Une interpolation est généralement plus fiable qu'une extrapolation.
- Un modèle mathématique doit toujours être interprété avec esprit critique.

Je dois savoir faire

À la fin du chapitre, je dois savoir :

- lire une information dans un tableau croisé ;
- compléter des totaux de lignes ou de colonnes ;
- calculer une fréquence et l'exprimer en pourcentage ;
- placer ou lire un point dans un nuage ;
- déterminer si un nuage présente une tendance croissante ou décroissante ;
- calculer le point moyen d'un nuage ;
- utiliser une équation de droite d'ajustement ;
- distinguer interpolation et extrapolation ;
- formuler une conclusion prudente dans un contexte réel.

Ouverture vers le TD

Le TD associé suivra exactement les parties du cours : tableaux croisés, fréquences, nuages de points, point moyen, ajustement affine, interpolation et extrapolation. Chaque partie se terminera par un exercice bilan, puis le TD se conclura par un exercice de synthèse de type épreuve anticipée.