mini **manuel**

Statistiques et probabilités en économie-gestion

2e édition

- **L'essentiel du cours**
- **Exercices corrigés**
- **Étude de cas**

Benjamin Legros

DUNOD

DANS LA MÊME COLLECTION

Collain B., Déjean F., Le Theule M.-A., *Mini Manuel de Comptabilité générale*, 2e ed., 2014
Legros B., *Mini Manuel de Finance d'entreprise*, 2e éd., 2014
Legros B., *Mini Manuel de Mathématiques financières*, 2e éd., 2016
Kruger A, Carpentier L., Ferrandi J.-M., Ingarao A., et al., *Mini Manuel de Marketing*, 2e éd., 2015

ISBN 978-2-10-074530-2

Table des matières

Partie 1
Statistiques

La page d'entrée de chapitre


CHAPITRE 1 Statistiques à une variable

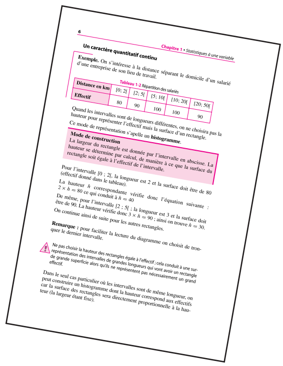
OBJECTIFS
➤ Savoir construire un diagramme adapté à une série statistique
➤ Maîtriser les paramètres de base de l'évaluation statistique (moyenne, médiane, écart-type et quartiles)
➤ Savoir interpréter un résultat à l'aide d'intervalles représentatifs ou du coefficient de variation

PLAN
1.1 Étude statistique et représentation
1.2 Indicateurs de tendance centrale
1.3 Indicateurs de dispersion
1.4 Interprétation des résultats

1.1 ÉTUDE STATISTIQUE ET REPRÉSENTATION

a) Vocabulaire de l'étude

Elle donne le plan du cours, ainsi qu'un rappel des objectifs pédagogiques du chapitre.



Le cours

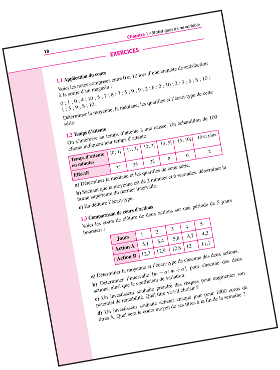
Le cours, concis et structuré, expose les notions importantes du programme.

Les rubriques

 Une erreur à éviter

 Un peu de méthode

 Les points clés à retenir



Les exercices

Ils sont proposés en fin de chapitre, avec leur solution, pour se tester tout au long de l'année.

PARTIE 1

Statistiques

Prenons l'exemple des ventes d'un magasin : quels paramètres permettent de comprendre un résultat annuel ? Il y a des facteurs externes comme la situation économique, le niveau de vie des clients ou la présence de concurrents ; et des facteurs internes : la publicité, les prix ou la qualité de service. Cette liste est loin d'être exhaustive et l'influence réelle de chacun des paramètres est difficile à maîtriser. Pourtant, il est fondamental de comprendre au mieux les éléments essentiels influant sur ces ventes.

En économie et en finance, les problèmes ont en commun la multiplicité des éléments influant. L'outil des statistiques permet d'appréhender ce type de problèmes et de réaliser une synthèse des grandes dynamiques en présence.

La première partie de l'ouvrage présente les outils statistiques les plus utiles pour le gestionnaire.

Le premier chapitre, « Statistiques à une variable », sert à construire les outils d'observation d'un phénomène. Le second chapitre, « Statistiques à deux variables », permet d'évaluer le lien entre deux grandeurs pour ensuite réaliser des prévisions. Le troisième chapitre, « Séries chronologiques », est particulièrement utile à la finance de marché et dans les phénomènes variant dans le temps.

CHAPITRE 1

Statistiques à une variable

OBJECTIFS

- ➤ Savoir construire un diagramme adapté à une série statistique.
- ➤ Maîtriser les paramètres de base de l’évaluation statistique (moyenne, médiane, écart-type et quartiles).
- ➤ Savoir interpréter un résultat à l’aide d’intervalles représentatifs ou du coefficient de variation.
- ➤ Savoir calculer des moyennes arithmétiques, géométriques et harmoniques.
- ➤ Savoir calculer des pourcentages d'augmentation ou de diminution.
- ➤ Utiliser la moyenne géométrique pour des variations.
- ➤ Connaître les indices de Laspeyres, Paasche et Fisher pour un ensemble de produits.

PLAN

1.1 ÉTUDE STATISTIQUE ET REPRÉSENTATION

a) Vocabulaire de l’étude

L’analyse statistique consiste à extraire une information utile et synthétique d’un ensemble d’observations. L’étude doit se limiter à une **population** qui, pour le gestionnaire, sera par exemple la production d’une usine, les salariés d'une entreprise ou encore un ensemble de consommateurs d'un produit. L'analyse de cette population se fait au travers d'une grille de critères encore appelés **caractères**.

Ces caractères peuvent être **qualitatifs**. On entend par qualitatif un caractère qui ne peut pas être évalué par un chiffre ; par exemple, la couleur d'une voiture, le poste dans une entreprise ou la nationalité.

Ces caractères peuvent être aussi **quantitatifs**. Selon le besoin de l'étude, on peut choisir de considérer chaque résultat individuellement ; on parle alors de **caractère discret**, par exemple, le résultat à une épreuve ou le nombre de voitures dont dispose un individu.

Si le détail des résultats n'apporte pas de grand intérêt, on regroupe les résultats par intervalles. On parle alors de **caractères continus** ; par exemple, la distance du domicile d'un salarié à son lieu de travail. Il sera plus intéressant de savoir combien habitent entre 5 km et 10 km du lieu de travail plutôt que combien habitent précisément à 7 km.

b) La fréquence

Les statistiques sont un vecteur majeur de communication. L'actualité des entreprises regorge de données statistiques diverses qui par leurs présentations informent (ou désinforment).

Le premier élément de communication est la **fréquence** : on exprime l'importance d'une donnée sous forme de pourcentage.

Exemple. L'usine A a produit 457 objets dont 34 sont défectueux et l'usine B a produit 537 objets dont 42 sont défectueux.

Il n'est pas évident à première vue de dire quelle usine a la meilleure qualité de production. Si on exprime les choses ainsi : l'usine A a 7,44 % de produits défectueux et l'usine B en a 7,82 %, la comparaison est facile.

Pour cela, on utilise **la formule de la fréquence** :

$$\text{Fréquence} = \frac{\text{Effectif considéré}}{\text{Effectif Total}} \times 100$$

Le résultat est ainsi exprimé sous forme de pourcentage.

Exemple. Pour l'usine A, l'effectif considéré est 34, l'effectif total est 457. Avec la formule précédente on retrouve donc 7,44 %.

c) Modes de représentation

Comment représenter au mieux une série statistique ? La réponse dépend du type de la série étudiée ainsi que de l'information que l'on souhaite rendre visible. Il n'y a pas de règles absolues pour représenter une série : dans la littérature, on constate l'utilisation de tous types de diagrammes pour tous types de séries. Cependant, pour éviter d'induire de fausses informations, certains diagrammes semblent plus adaptés que d'autres.

Un caractère quantitatif discret

Exemple. Une population de 100 clients évaluent un centre d'appels téléphoniques par une note de 0 à 5 :

Tableau 1-1 Résultat de l'évaluation

Notes	0	1	2	3	4	5
Effectif	10	15	10	35	25	5

Il s'agit dans cet exemple de représenter la gradation des notes – 0 moins bon que 1 lui même moins bon que 2 ... – et l'importance de la représentation de chaque note donnée par l'effectif.

Pour figurer une série discrète, le mode de représentation le plus simple à produire et à interpréter est le **diagramme en bâton**.

Mode de construction

Dans un diagramme en bâton, on représente les notes par des bâtons dont la hauteur est égale à l'effectif.

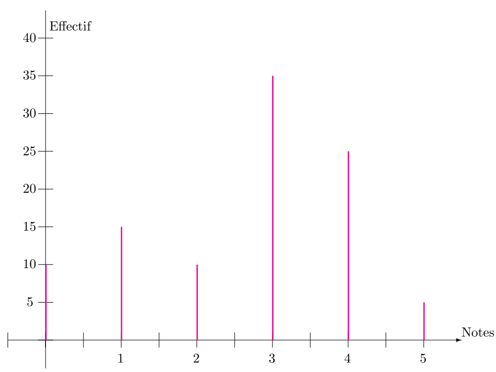


Figure 1-1 Diagramme en bâton

Un caractère quantitatif continu

Exemple. On s'intéresse à la distance séparant le domicile d'un salarié d'une entreprise de son lieu de travail.

Tableau 1-2 Répartition des salariés

Distance en km	[0; 2[	[2; 5[	[5; 10[	[10; 20[	[20; 50[
Effectif	80	90	100	100	90

Quand les intervalles sont de longueurs différentes, on ne choisira pas la hauteur pour représenter l'effectif mais la surface d'un rectangle.

Ce mode de représentation s'apelle un **histogramme**.

Mode de construction
La largeur du rectangle est donnée par l'intervalle en abscisse. La hauteur se détermine par calcul, de manière à ce que la surface du rectangle soit égale à l'effectif de l'intervalle.

Pour l'intervalle [0 ; 2[, la longueur est 2 et la surface doit être de 80 (effectif donné dans le tableau).

La hauteur h correspondante vérifie donc l'équation suivante : $2 \times h = 80$ ce qui conduit à $h = 40$

De même, pour l'intervalle [2 ; 5[; la longueur est 3 et la surface doit être de 90. La hauteur vérifie donc $3 \times h = 90$; ainsi on trouve $h = 30$.

On continue ainsi de suite pour les autres rectangles.

Remarque : pour faciliter la lecture du diagramme on choisit de tronquer le dernier intervalle.

Ne pas choisir la hauteur des rectangles égale à l'effectif ; cela conduit à une surreprésentation des intervalles de grandes longueurs qui vont avoir un rectangle de grande superficie alors qu'ils ne représentent pas nécessairement un grand effectif.

Dans le seul cas particulier où les intervalles sont de même longueur, on peut construire un histogramme dont la hauteur correspond aux effectifs car la surface des rectangles sera directement proportionnelle à la hauteur (la largeur étant fixe).

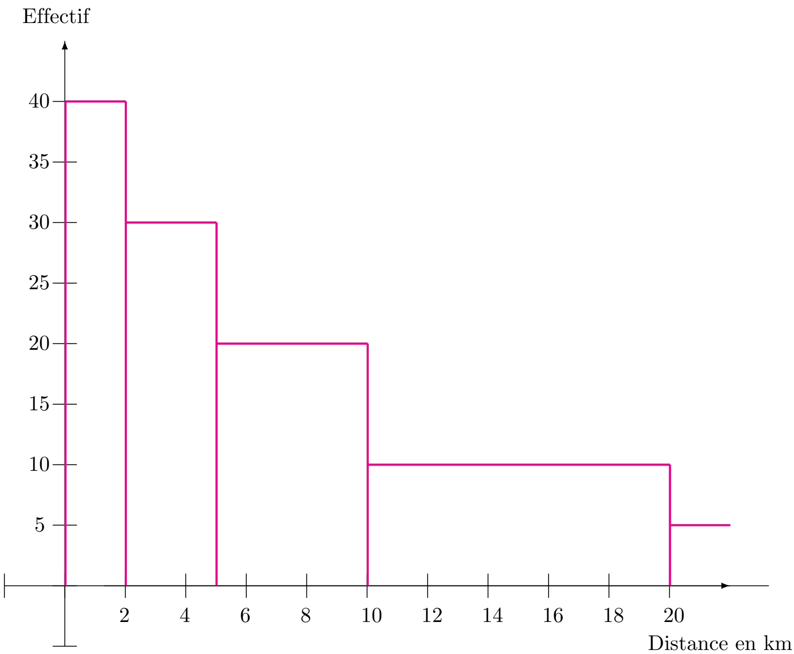


Figure 1-2 Histogramme

Un caractère qualitatif

Exemple. On s'intéresse aux couleurs des voitures d'une sortie de production pendant une période donnée. Voici les résultats constatés :

Tableau 1-3 Sortie de production

Couleurs	Noir	Rouge	Jaune	Vert
Production	10 000	5 000	2 000	3 000

L'utilisation d'un histogramme ou d'un diagramme en bâton induirait, dans le cas de données qualitatives, l'idée d'une progression du type « telle couleur meilleure que telle autre ». Ce qui n'est pas l'intention ici.

Il est plus judicieux de représenter cette série par un **diagramme circulaire** où l'angle considéré est proportionnel à l'effectif.

La forme circulaire évite de donner l'illusion d'une gradation. Pour retrouver un sens de lecture on utilise un diagramme **semi-circulaire**. L'exemple le plus classique est la composition politique d'un parlement.

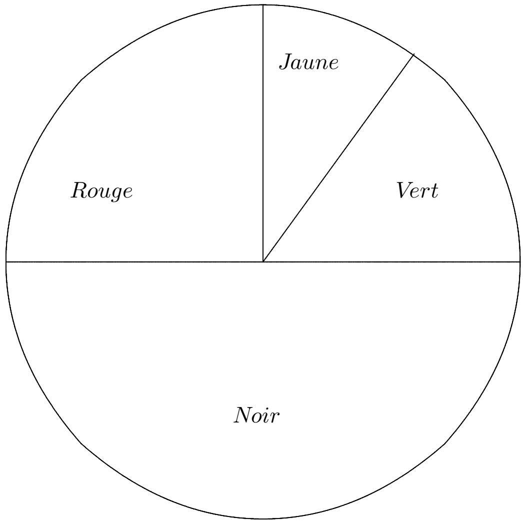


Figure 1-3 Diagramme circulaire

1.2 INDICATEURS DE TENDANCE CENTRALE

a) Moyennes

Une moyenne est une valeur qui se trouve au milieu de toutes les autres. C'est un indicateur de **tendance centrale** qui permet d'appréhender une population de manière globale. La moyenne, comme les autres éléments du calcul statistique, ne s'utilise que pour des caractères quantitatifs. On pourrait essayer de calculer la couleur moyenne des voitures de l'exemple précédent, mais cela ne serait d'aucune utilité.

Moyenne arithmétique simple

Exemple. Dans l'exemple suivant, on s'intéresse à la production mensuelle d'une usine sur une période de 6 mois :

Tableau 1-4 Production mensuelle

Mois	Janvier	Février	Mars	Avril	Mai	Juin
Production en unités	10 000	50 000	20 000	30 000	30 000	40 000

Chaque mois a la même importance ; si on cherche à calculer une moyenne de la production mensuelle, on ajoute les différentes productions pour les diviser par le nombre de mois :

$$\text{Moyenne} = \frac{10\,000 + 50\,000 + 20\,000 + 30\,000 + 30\,000 + 40\,000}{6}$$
$$= 30\,000$$

En notant $x_1, x_2, \cdots, x_n$ les valeurs de la série et $\overline{x}$ la moyenne, on a la formule générale d'une **moyenne simple** :

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Moyenne arithmétique pondérée

Dans le cas du *Tableau 1-1*, on va chercher à déterminer une note moyenne de satisfaction. Les effectifs associés à chaque note sont à prendre en considération.

En notant $e_1, e_2, \cdots, e_n$ les effectifs associés aux valeurs $x_1, x_2, \cdots, x_n$ et $\overline{x}$ la moyenne, on a la formule générale d'une **moyenne pondérée** :

$$\overline{x} = \frac{e_1x_1 + e_2x_2 + \cdots + e_nx_n}{e_1 + e_2 + \cdots + e_n} = \frac{\sum_{i=1}^{n} e_i x_i}{\sum_{i=1}^{n} e_i}$$

Ce qui, appliqué à la série du *Tableau 1-1,* conduit à :

$$\text{Moyenne} = \frac{0 \times 10 + 1 \times 15 + 2 \times 10 + 3 \times 35 + 4 \times 25 + 5 \times 5}{100}$$
$$= 2{,}65$$

Moyenne pour un caractère continu

Il se pose le problème du choix de la valeur à considérer pour appliquer la formule précédente. La solution la plus pratique est de retenir les milieux de chacune des classes comme valeurs. Dans l'exemple du *Tableau 1-2,* on trouve :

$$\text{Moyenne} = \frac{1 \times 80 + 3{,}5 \times 90 + 7{,}5 \times 100 + 15 \times 100 + 35 \times 90}{80 + 90 + 100 + 100 + 90}$$
$$\approx 12{,}60 \text{ km}$$

Remarque : cette valeur est en réalité le résultat d'une approximation ; on a supposé que les effectifs étaient répartis de manière homogène dans les intervalles. Cela justifie le choix du milieu des intervalles. On procèdera de même pour les calculs d'écart-moyen et d'écart-type.

Moyenne géométrique

La moyenne arithmétique ne s'applique pas aux évolutions en pourcentage car ceux-ci ne peuvent s'additionner. Prenons l'exemple d'une variation de +10 % suivie d'une variation de –10 %. Il ne s'agit pas globalement d'une absence de variation mais au contraire d'une variation de –1 %.

En effet, augmenter de 10 % revient à multiplier une grandeur par 1,1 et baisser de 10% revient à multiplier par 0,9. Ainsi, au total il y aura une multiplication par $1{,}1 \times 0{,}9 = 0{,}99$. Ceci correspond à une baisse de 1%.

Exemple. Pour une variation en pourcentage par exemple +5 %, – 6 %, +8 %, +14 % , on multiplie les valeurs correspondantes pour obtenir la variation globale. En assumant un taux de variation moyen équivalent i sur 4 variations, on trouve alors $(1+i)^4 = 1{,}05 \times 0{,}94 \times 1{,}08 \times 1{,}14$, soit $i = (1{,}05 \times 0{,}94 \times 1{,}08 \times 1{,}14)^{1/4} - 1 = 4{,}99\%$. Il s'agit d'un calcul de moyenne géométrique.

En notant $x_1, x_2, \ldots, x_n$ les valeurs d'une série et la moyenne, on a la formule générale d'une **moyenne géométrique** :

$$\overline{x} = (x_1 \times x_2 \times \ldots \times x_n)^{1/n} = \left(\prod_{i=1}^{n} x_i\right)^{1/n}.$$

Moyenne harmonique

Exemple. Considérons un investissement de 1 000 € pour acheter des actions un premier jour lorsque leur cours est à 5. Supposons qu'un second investissement de 1 000 € est réalisé 20 jours plus tard lorsque le cours de cette action est passé à 4 €. On peut alors calculer le cours moyen du portefeuille de cet individu pour ce titre. Le premier jour l'individu a acheté $\dfrac{1\,000}{5} = 200$ actions et le 20$^{\text{ième}}$ jour il a acheté $\dfrac{1\,000}{4} = 250$ actions. Il a réalisé en tout un investissement de 2000 €.

Ainsi le cours moyen de son portefeuille est de $\dfrac{1\,000 + 1\,000}{\frac{1\,000}{5} + \frac{1\,000}{4}}$

$= \dfrac{2}{\frac{1}{5}+\frac{1}{4}} = 4{,}44$. Il s'agit de la moyenne harmonique des valeurs 4 et 5.

En notant les valeurs d'une série et la moyenne, on a la formule générale d'une moyenne harmonique :

$$\overline{x} = \frac{n}{\dfrac{1}{x_1} + \dfrac{1}{x_2} + \ldots + \dfrac{1}{x_n}} = \frac{n}{\displaystyle\sum_{i=1}^{n} \frac{1}{x_i}}.$$

b) Médiane

On considère la suite de notes suivantes : 7 ; 8 ; 9 ; 20. La moyenne de cette série est 11. Il s'agit bien du milieu de ces 4 notes ; pour autant 75 % de ces notes sont en dessous de 10 et la moyenne semble donner une indication positive à savoir 11. La moyenne est en effet très influencée par les valeurs extrêmes, ici le 20.

Il est donc utile de s'intéresser à un autre indicateur de tendance centrale qui est **la médiane**.

Définition : la médiane est la valeur de la série pour laquelle 50 % de la population a ses valeurs en dessous et 50 % a ses valeurs au dessus.

Ici, la médiane est entre 8 et 9, elle est de 8,5. Cette valeur de 8,5 indique bien le fait que 3 notes sur 4 sont en dessous de 10 et que l'évaluation n'est pas aussi positive que semblait l'indiquer la moyenne.

Un autre exemple de différence sensible entre moyenne et médiane est le salaire. On note un salaire moyen net en France autour de 1 800 € et un salaire médian net autour de 1 500 €. L'explication est la même que pour la série précédente ; des salaires élevés mais peu nombreux augmentent la moyenne des salaires mais n'ont que peu d'effet sur la médiane.

Calcul pour un caractère discret

Commençons par des exemples simples :

Exemple.

Série *a* : – 1 ; 3 ; 6 ; 8 ; 9

Les valeurs sont classées par ordre croissant, il y a 5 valeurs. La valeur qui sépare la série en deux sous-séries de même taille est 6. Ainsi la médiane est 6.