

Manuel

de

probabilités et statistique

Cours + QCM

Françoise Couty-Fredon

Professeure certifiée hors classe

Jean Debord

Praticien hospitalier attaché au CHU de Limoges
Chargé de cours à la faculté des sciences de Limoges

Daniel Fredon

Maître de conférences de mathématiques appliquées

3^e édition

DUNOD

Directeur d'ouvrage
Daniel FREDON

Le pictogramme qui figure ci-contre mérite une explication. Son objet est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, particulièrement dans le domaine de l'édition technique et universitaire, le développement massif du photocopillage.

Le Code de la propriété intellectuelle du 1^{er} juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique s'est généralisée dans les établissements

d'enseignement supérieur, provoquant une baisse brutale des achats de livres et de revues, au point que la possibilité même pour

les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée. Nous rappelons donc que toute reproduction, partielle ou totale, de la présente publication est interdite sans autorisation de l'auteur, de son éditeur ou du

Centre français d'exploitation du droit de copie (CFC, 20, rue des Grands-Augustins, 75006 Paris).



© Dunod, 2007, 2014, 2018

11, rue Paul Bert 92240 Malakoff
www.dunod.com

ISBN 978-2-10-078094-5

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2° et 3° a), d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

Comment utiliser ce Mini-Manuel ?

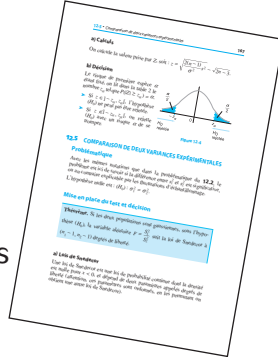
La page d'entrée de chapitre



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



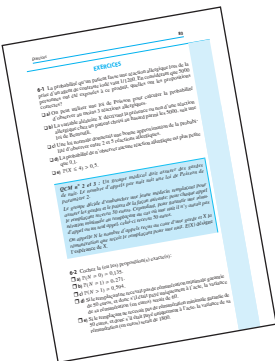
Un peu de méthode



Un exemple pour comprendre



Les points clés à retenir



Les exercices, QCM ou QROC

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

Table des matières

Avant-propos	1
---------------------	----------

Statistique descriptive

1	Statistique à une dimension	3
2	Statistique à deux dimensions	17

Probabilités

3	Probabilités (généralités)	33
4	Probabilité conditionnelle	45
5	Variabes aléatoires discrètes (cas fini)	59
6	Variabes aléatoires discrètes (cas infini)	77
7	Variabes aléatoires continues	91

Statistique inférentielle

8	Échantillonnage – Estimation d'un paramètre	107
9	Introduction aux tests statistiques	123
10	Test du khi-deux (χ^2)	129
11	Comparaison de deux proportions	145
12	Comparaison de deux moyennes, de deux variances	159
13	Analyse de la variance	181

14	Régression linéaire	197
15	Corrélation	211
16	Tests non paramétriques	225

Tables

1	– Fonction de répartition de la loi normale réduite	241
2	– Loi normale réduite (table de l'écart réduit)	242
3	– Lois de Student	243
4	– Lois de Pearson ou lois du χ^2	244
5	– Lois de Snedecor ($\alpha = 0,025$)	245
6	– Lois de Snedecor ($\alpha = 0,05$)	246
7	– Test de Mann et Whitney ($\alpha = 0,05$)	247
8	– Test de Mann et Whitney ($\alpha = 0,01$)	248
9	– Test de Wilcoxon	248
10	– Table du coefficient de corrélation linéaire	249
11	– Coefficient de corrélation de rang de Spearman	250
12	– Test de Kruskal et Wallis	250

Glossaire	251
------------------	------------

Index	255
--------------	------------

Avant-propos

Ce livre est destiné à tous les étudiants en sciences de la vie, de la Terre et de la santé : licences, pharmacie, médecine, IUT et BTS à dominante biologique ou agricole. Mais il concerne aussi tous les utilisateurs de statistiques en laboratoire.

Pour satisfaire l'attente d'un tel public, nous avons choisi d'aborder une large étendue de sujets. Comme ce livre est découpé en chapitres autonomes, chacun pourra, d'une année à l'autre, retrouver les sujets qui le concernent. Un index détaillé situé en fin d'ouvrage aidera dans ce choix.

En statistique, les notations ne sont pas toutes universelles, ce qui complique la consultation d'ouvrages variés. Pour notre part, nous nous sommes efforcés de respecter les règles de cohérence suivantes :

- utiliser les lettres grecques pour des valeurs relatives à la population et des lettres latines pour des valeurs relatives à un échantillon ;
- bien distinguer une variable aléatoire (notée par une majuscule) et une valeur numérique prise par cette variable aléatoire (notée par la même lettre, mais en minuscule).

Pour des révisions express à l'approche d'un examen ou d'un concours, nous vous conseillons, chez le même éditeur :

D. Fredon ; Statistique et probabilités en 30 fiches ; collection Express Sciences.

Bien sûr, la charité bien ordonnée commence par soi-même. Mais il y a plus important : vous y trouverez des notations en cohérence avec ce livre, et des exercices différents pour compléter votre entraînement.

Toutes vos remarques, vos commentaires, vos critiques, et même vos encouragements, seront accueillis avec plaisir. Vous pouvez me les communiquer à l'adresse électronique suivante : daniel.fredon@laposte.net

Daniel Fredon

Statistique à une dimension

PLAN

- 1.1 Généralités
- 1.2 Représentations graphiques
- 1.3 Paramètres de position
- 1.4 Paramètres de dispersion
- 1.5 Paramètres de forme

OBJECTIFS

- Savoir représenter graphiquement une série statistique après avoir choisi l'aspect à mettre en évidence
- Résumer certains aspects (position, dispersion, forme) d'une série statistique en calculant un nombre adapté

1.1 GÉNÉRALITÉS

Vocabulaire général

La statistique étudie des ensembles appelés **populations**, dont les éléments sont appelés **individus**. Dans le cas d'une série statistique à une variable, à chaque individu on associe une éventualité d'un **caractère statistique**.

Si les éventualités ne sont pas des nombres, le caractère est dit **qualitatif** et les éventualités s'appellent les modalités du caractère.

Si les éventualités sont des nombres, le caractère est dit **quantitatif** et les éventualités sont les valeurs du caractère.

Un caractère quantitatif est dit **continu** s'il peut prendre toutes les valeurs d'un intervalle. Il est discontinu, ou **discret**, s'il ne peut prendre que des valeurs isolées.

Série statistique

Dans le cas d'un caractère qualitatif ou quantitatif discret, on dispose d'une série statistique quand on connaît pour chaque individu la modalité, ou la valeur, prise par le caractère. L'**effectif** d'une modalité, ou d'une valeur, est le nombre de fois où elle apparaît dans la population.

Quand le caractère quantitatif est continu, ou discret avec beaucoup de valeurs, on considère des intervalles, en général du type $]a, b]$, que l'on appelle des **classes statistiques**. La longueur $b - a$ de l'intervalle est l'**amplitude** de la classe. Sa **densité** est le quotient de l'effectif par l'amplitude.

Effectifs, fréquences

Pour une valeur (ou une modalité) d'un caractère, ou pour une classe statistique, la **fréquence** est le quotient de l'effectif concerné n_i par l'effectif total n , soit : $f_i = \frac{n_i}{n}$.

La somme des fréquences est donc égale à 1.

Si on veut obtenir la répartition en pourcentages, il suffit de multiplier les fréquences par 100.

Effectifs cumulés, fréquences cumulées

Lorsque le caractère est quantitatif, on range les valeurs (ou les classes) par ordre croissant.

- L'**effectif cumulé** jusqu'à k est la somme des effectifs associés aux valeurs du caractère qui sont inférieures ou égales à k .
- La **fréquence cumulée** jusqu'à k s'obtient en additionnant les fréquences associées aux valeurs $\leq k$, ou en divisant l'effectif cumulé par l'effectif total.

1.2 REPRÉSENTATIONS GRAPHIQUES

Cas d'un caractère qualitatif ou quantitatif discret

- Si on veut insister sur la comparaison des effectifs, on trace un diagramme à bandes, ou un diagramme en bâtons. Les longueurs doivent être proportionnelles aux effectifs.
- Si on préfère mettre en évidence les pourcentages pour comparer visuellement les structures de plusieurs séries statistiques (c'est-à-dire les répartitions en pourcentages), on représente les données à l'aide :

- de graphiques circulaires (parfois appelés camemberts), où les angles au centre du disque, ou du demi-disque, sont proportionnels aux pourcentages ;
- ou de bandes subdivisées de longueur fixe.

Cas d'un caractère quantitatif continu

On peut utiliser les représentations précédentes. Mais on construit le plus souvent un **histogramme** :

Les intervalles des classes statistiques sont reportés sur un axe. Il servent de bases à des rectangles dont les aires sont proportionnelles aux effectifs. Pour ceci, les côtés des rectangles perpendiculaires à l'axe sont proportionnels aux densités des classes.

1.3 PARAMÈTRES DE POSITION

Moyenne

a) Définition

Notons x_1, \dots, x_p les valeurs du caractère, n_1, \dots, n_p les effectifs correspondants et $n = n_1 + \dots + n_p$ l'effectif total. La moyenne de la série statistique est le nombre :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^p n_i x_i = \sum_{i=1}^p f_i x_i.$$

Quand les informations sont fournies avec des classes statistiques, on utilise la même formule en retenant comme valeurs x_i les milieux des classes.



La définition précédente est celle de la moyenne arithmétique. On définit aussi d'autres moyennes pour $x_i > 0$ comme :

– la moyenne harmonique h telle que : $\frac{n}{h} = \sum_{i=1}^p \frac{n_i}{x_i}$

– la moyenne géométrique g telle que : $g^n = x_1^{n_1} \times \dots \times x_p^{n_p}$.

b) Propriétés

La moyenne ne change pas si on remplace les effectifs par des effectifs proportionnels.

La moyenne ne change pas si on remplace k valeurs x_1, \dots, x_k affectées de coefficients n_1, \dots, n_k par leur moyenne partielle affectée de la somme des coefficients $n_1 + \dots + n_k$.

Par exemple, si la population est subdivisée en trois sous-populations dont les moyennes partielles sont $\bar{x}_1, \bar{x}_2, \bar{x}_3$ et les effectifs N_1, N_2, N_3 , alors la moyenne de la population totale est :

$$\bar{x} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2 + N_3\bar{x}_3}{N_1 + N_2 + N_3}.$$

Médiane

a) Cas d'un caractère quantitatif discret

On ordonne les n valeurs de la série statistique par ordre croissant.

Si n est impair, la médiane est la valeur de rang $\frac{n+1}{2}$.

Si n est pair, les valeurs de rangs $\frac{n}{2}$ et $\frac{n}{2} + 1$ déterminent un intervalle médian. On retient souvent comme médiane le milieu de cet intervalle.

b) Cas d'un caractère quantitatif continu

Dans ce cas, la médiane est le nombre m tel que la fréquence cumulée jusqu'à m soit égale à 0,5.

Mode

On appelle **mode**, ou dominante, d'une série statistique toute valeur (ou modalité) correspondant à l'effectif maximal (densité maximale dans le cas de classes statistiques).

1.4 PARAMÈTRES DE DISPERSION

Variance, écart type

Avec les mêmes notations que précédemment, on appelle variance de la série statistique le nombre V :

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

On appelle écart type de la série statistique le nombre $\sigma = \sqrt{V}$.

La variance peut aussi se calculer par :

$$V = \frac{1}{n} \left(\sum_{i=1}^p n_i x_i^2 \right) - (\bar{x})^2.$$



Les calculatrices et les tableurs fournissent directement, sous des notations diverses, l'écart type σ et la variance σ^2 .

Mais ils fournissent aussi un autre nombre s , sous des notations variées, qui est tel que :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^p n_i (x_i - \bar{x})^2 = \frac{n}{n-1} \sigma^2.$$

s^2 est destiné à estimer la variance d'une population quand on ne dispose que d'un échantillon de taille n .

On l'appelle la variance estimée et elle ne doit pas être confondue avec la variance V de l'échantillon (cf. chap. 8).

Coefficient de variation

Le **coefficient de variation** d'une série statistique est le quotient $\frac{\sigma}{\bar{x}}$.

C'est un nombre sans dimension qui permet de comparer la dispersion de séries statistiques dont les moyennes sont très différentes.

Autres paramètres de dispersion

- L'**étendue** d'une série statistique associée à un caractère quantitatif est la différence entre la plus grande valeur observée et la plus petite.
- En partageant la série ordonnée des résultats en quatre parties de même effectif, on obtient les quartiles Q_1, Q_2, Q_3 . Le deuxième quartile Q_2 est la médiane. L'**écart interquartile** est le nombre $Q_3 - Q_1$.

Boîte de dispersion (ou boîte à moustaches)

C'est une représentation graphique d'un caractère quantitatif. Elle sert à comparer visuellement plusieurs séries statistiques.

Pour une série donnée, on trace un rectangle qui s'étend de Q_1 à Q_3 et on marque la médiane par un trait. On ajoute les moustaches qui sont les segments qui vont de la valeur minimale à Q_1 , et de Q_3 à la valeur maximale.

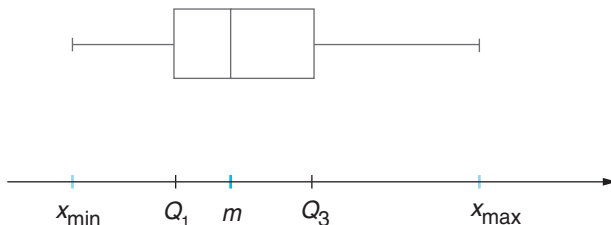


Figure 1.1

1.5 PARAMÈTRES DE FORME

Moments

Pour $r \in \mathbb{N}$, on définit le moment d'ordre r : $m_r = \frac{1}{n} \sum_{i=1}^p n_i x_i^r$,

le moment centré d'ordre r : $\mu_r = \frac{1}{n} \sum_{i=1}^p n_i (x_i - \bar{x})^r$.

Coefficient γ_1 de Fisher (dissymétrie) : $\gamma_1 = \frac{\mu_3}{\sigma^3}$

Si $\gamma_1 = 0$, la distribution est symétrique.

Si $\gamma_1 < 0$, la distribution est étalée vers la gauche.

Si $\gamma_1 > 0$, la distribution est étalée vers la droite.

Coefficient γ_2 de Fisher (aplatissement) : $\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$

Si $\gamma_2 = 0$, l'aplatissement est le même que celui de la loi de Gauss réduite (cf. chap. 7).

Si $\gamma_2 < 0$, la distribution est plus aplatie.

Si $\gamma_2 > 0$, la distribution est moins aplatie.



Effets d'un regroupement en classes

Lorsque la série statistique comporte un grand nombre de valeurs, les calculs sont simplifiés en effectuant d'abord un regroupement en classes, puis en remplaçant chaque classe par son milieu. Mais les résultats en sont légèrement modifiés.

Si la distribution des valeurs est uniforme dans chaque classe, la moyenne n'est pas changée (associativité de la moyenne).

Mais la variance, qui mesure la dispersion, est modifiée puisqu'on concentre toutes les valeurs d'une classe en un seul point. Dans le cas où toutes les classes sont de même amplitude d , on peut améliorer le résultat avec la correction de Sheppard, qui consiste à retrancher $\frac{d^2}{12}$ à la valeur de la variance obtenue à partir des valeurs groupées.



MOTS-CLÉS

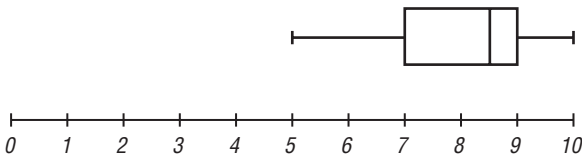
- ▶ Caractère statistique
- ▶ Représentation graphique
- ▶ Paramètres d'une série statistique

EXERCICES

1-1 Gwenaël, qui a 16 ans, souffre d'une fracture complexe du tibia et du péroné après avoir chuté à ski. Si Y est le type de fracture, quel est le type de la variable décrivant Y ?

- a) Variable discrète finie.
- b) Variable ordinale.
- c) Variable qualitative nominale.
- d) Variable entière finie.
- e) Variable quantitative non dénombrable.

1-2 Concernant le diagramme en box-plot suivant :



- a) La moyenne est égale à 8, 5.
- b) La médiane est égale à 8, 5.
- c) L'écart interquartile est égal à 2.
- d) 50 % de l'effectif est contenu dans l'intervalle $[7 ; 9]$.
- e) Il y a une valeur extrême.

1-3 Une étude sur le sommeil (en heures par jour) des enfants de 4 ans a donné les résultats suivants sur les 591 enfants de l'étude :

minimum : 7 ; premier quartile : 9 ; médiane : 10 ; troisième quartile : 11 ; maximum : 14.

- a) La variable étudiée était qualitative ordinale.
- b) 295 enfants dormaient au moins 10 heures.
- c) Le deuxième quartile était égal à 8.
- d) L'étendue était de 3.
- e) 148 enfants dormaient au moins 11 heures.

1-4 On considère une série statistique de 60 taux d'hémoglobine dans le sang (g/L) mesurés chez des adultes présumés en bonne santé. La série est rangée par valeurs non décroissantes. Les valeurs en gras indique que le taux d'hémoglobine a été mesuré sur une femme.

105 ; 110 ; 112 ; 112 ; 118 ; 119 ; 120 ; 120 ; 125 ; 126 ; 127 ; 128 ; 130 ; 132 ; 133 ; 134 ; 135 ; 138 ; 138 ; 138 ; 138 ; 141 ; 142 ; 144 ; 145 ; 146 ; 148 ; 148 ; 148 ; 149 ; 150 ; 150 ; 150 ; 151 ; 151 ; 153 ; 153 ; 153 ; 154 ; 154 ; 154 ; 155 ; 156 ; 156 ; 158 ; 160 ; 160 ; 160 ; 163 ; 164 ; 164 ; 165 ; 166 ; 168 ; 168 ; 170 ; 172 ; 172 ; 176 ; 179.

Résultats partiels

$$\text{Hommes : } \sum_{i=1}^{30} x_{ih} = 4\,766 \text{ g/L} \quad \sum_{i=1}^{30} x_{ih}^2 = 759\,954 \text{ (g/L)}^2$$

$$\text{Femmes : } \sum_{i=1}^{30} x_{if} = 3\,988 \text{ g/L} \quad \sum_{i=1}^{30} x_{if}^2 = 536\,176 \text{ (g/L)}^2$$

- a)** On considère le groupement en classes :
]104;114] ;]114;124] ;]124;134] ;]134;144] ;]144;154] ;]154;164] ;
]164;174] ;]174;184]

Pour chacune des deux séries : hommes, femmes, déterminez les effectifs et les fréquences de chaque classe.

- b)** Effectuez une représentation graphique adaptée des deux distributions groupées en classes de la question précédente.

- c)** Calculez les moyennes \bar{x} , \bar{x}_f , \bar{x}_h , des trois distributions initiales : ensemble, femmes, hommes ;

- d)** Calculez les moyennes \bar{x}' , \bar{x}'_f , \bar{x}'_h , des trois distributions (ensemble, femmes, hommes) après le regroupement en classes de la question **a**), en remplaçant chaque classe par son milieu.

- e)** Calculez les médianes m , m_f , m_h des trois distributions initiales : ensemble, femmes, hommes.

- f)** Calculez l'écart interquartile pour chacune des trois distributions initiales : ensemble, femmes, hommes.

- g)** Calculez les variances et les écarts type des trois distributions initiales : femmes, hommes, ensemble.

h) Calculez les variances et les écarts type des trois distributions après le regroupement en classes de la question **a)**, en remplaçant chaque classe par son milieu.

i) Pour la distribution des femmes, calculez les moments m_1, m_2, m_3, m_4 . Déduisez-en les valeurs des moments centrés $\mu_1, \mu_2, \mu_3, \mu_4$, puis des coefficients de forme γ_1 et γ_2 de Fisher.

1-5 Dans l'étude de la répartition de la végétation en fonction de divers facteurs écologiques, on utilise une carte au 1/200 000 sur laquelle sont représentées les séries de végétation. On superpose une grille dont la maille est de 1 cm. Des renseignements annexes fournissent, pour chaque point de la grille, la température moyenne T en °C, la pluviosité annuelle moyenne P en mm, et la nature du sol.

En étudiant la région de Limoges, on a ainsi obtenu pour la population constituée par les points étudiés :

- Pour le *chêne pédonculé*

P]700 ; 800]]800 ; 900]]900 ; 1000]]1000 ; 1100]]1100 ; 1200]
effectifs	10	85	185	122	138

P]1200 ; 1300]]1300 ; 1400]]1400 ; 1500]]1500 ; 1600]]1600 ; 1700]
effectifs	43	15	12	13	10

P]1700 ; 1800]]1800 ; 1900]]1900 ; 2000]
effectifs	6	5	1

T]7 ; 8]]8 ; 9]]9 ; 10]]10 ; 11]]11 ; 12]]12 ; 13]
effectifs	4	25	109	250	205	52

sols	acides	calcaires	montagneux
effectifs	502	49	94

- Pour le *chêne pubescent*

P]700 ; 800]]800 ; 900]]900 ; 1000]]1000 ; 1100]
effectifs	14	103	37	3

T]11 ; 12]]12 ; 13]
effectifs	34	123

sols	acides	calcaires
effectifs	23	134

- a) En assimilant chaque classe à son milieu, calculez la pluviosité moyenne \bar{P}_1 pour les zones où vit le chêne pédonculé, puis \bar{P}_2 pour le chêne pubescent. Calculez les écarts type correspondant à ces deux séries statistiques et les coefficients de variation.
- b) Calculez de même les températures moyennes \bar{T}_1 et \bar{T}_2 et les écarts type correspondants.
- c) Construisez deux graphiques pour visualiser la comparaison de la nature des sols habités par le chêne pédonculé et le chêne pubescent.
- d) Conclusions écologiques ?

SOLUTIONS

1-1 a) b) c) d) e)

Les modalités sont les types de fracture, donc non numériques. La variable est donc qualitative. Il n'y a pas d'ordre entre les types de fracture; la variable n'est donc pas ordinale.

1-2 a) b) c) d) e)

a) **faux** et b. **vrai** : Le trait dans la boîte correspond à la médiane.

c) **vrai** : L'écart interquartile est $9 - 7 = 2$.

d) **vrai** : L'intervalle interquartile contient 50 % de la population.

e) **vrai** : Les valeurs extrêmes sont 5 et 10.

1-3 a) b) c) d) e)

a) **faux** : la variable est quantitative.

b) **vrai** : la médiane est 10, il y a donc la moitié des 591 enfants, soit 295, qui dorment au moins 10 heures.

c) **faux**: le deuxième quartile est la médiane, soit 10.

d) **faux** : l'étendue est la différence entre la valeur maximale et la valeur minimale, soit $14 - 7 = 7$.

e) **vrai** : le troisième quartile est 11, il y a donc un quart des 591 enfants, soit 148, qui dorment au moins 11 heures.

1-4 a)

classes	femmes		hommes	
	effectifs	fréquences	effectifs	fréquences
]104; 114]	4	0,133	0	0
]114; 124]	4	0,133	0	0
]124; 134]	8	0,267	0	0
]134; 144]	6	0,200	2	0,067
]144; 154]	7	0,233	10	0,333
]154; 164]	1	0,033	9	0,300
]164; 174]	0	0	7	0,233
]174; 184]	0	0	2	0,067
total	30	1	30	1

b) On peut dessiner deux histogrammes en portant indifféremment en ordonnées les effectifs, les fréquences ou les densités des classes, car les classes sont de même amplitude.

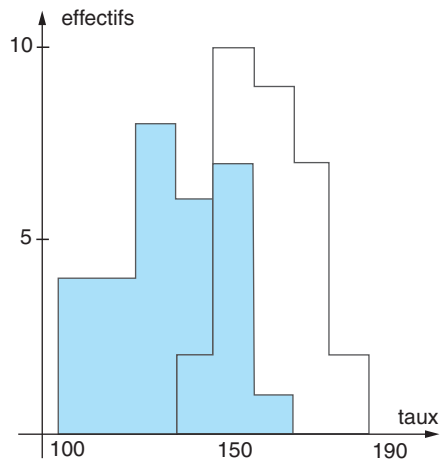


Figure 1-2

c) Pour la série totale : $\bar{x} = \frac{4\,766 + 3\,988}{60} = 145,9$

Pour la série des femmes : $\bar{x}_f = \frac{3\,988}{30} \approx 132,9$

Pour la série des hommes : $\bar{x}_h = \frac{4\,766}{30} \approx 158,9$.

d) $\bar{x}' = \frac{1}{60}[4 \times 109 + \dots + 2 \times 179] \approx 145,3$

$$\bar{x}'_f = \frac{1}{30}[4 \times 109 + \dots + 1 \times 159] \approx 132,7$$

$$\bar{x}'_h = \frac{1}{30}[2 \times 139 + \dots + 2 \times 179] = 158$$

Les différences avec les résultats de la question précédente signifient que la répartition dans chaque classe n'est pas uniforme.

e) Pour la série totale, la valeur de rang 30 est 149 et celle de rang 31 est

$$150. \text{ D'où : } m = \frac{149 + 150}{2} = 149,5.$$

Pour la série des femmes, la valeur de rang 15 est 133 et celle de rang 16 est 134. D'où : $m_f = 133,5$.

Pour la série des hommes, la valeur de rang 15 est 156 et celle de rang 16 est 160. D'où : $m_h = 158$.

f) Pour la série totale, la valeur de rang 15 est 133 ; la valeur de rang 16 est 134. D'où : $Q_1 = 133,5$.

La valeur de rang 45 est 158 ; la valeur de rang 46 est 160. D'où : $Q_3 = 159$.

Pour la série totale, l'écart interquartile est donc : $Q_3 - Q_1 = 25,5$.

Pour la série des femmes, la valeur de rang 8 est 120. D'où : $Q_1 = 120$.

La valeur de rang 23 est 114. D'où : $Q_3 - Q_1 = 25$.

Pour la série des hommes, on obtient de même : $Q_3 - Q_1 = 166 - 151 = 15$.

g) Pour la série des femmes, on a, en utilisant le théorème de Koenigs :

$$V_f = \frac{536\,176}{30} - \left(\frac{3\,988}{30}\right)^2 \approx 201,3 \text{ d'où : } \sigma_f = \sqrt{V_f} \approx 14,2.$$

Pour la série des hommes :

$$V_h = \frac{759\,954}{30} - \left(\frac{4\,766}{30}\right)^2 \approx 93,2 \text{ d'où : } \sigma_h = \sqrt{V_h} \approx 9,7.$$

Pour la série complète :

$$V = \frac{759\,954 + 536\,176}{60} - (145,9)^2 \approx 315,4 \text{ d'où : } \sigma = \sqrt{V} \approx 17,8.$$

h) On obtient :

$$\text{pour la série des femmes : } V'_f \approx 196,6 \quad \text{et} \quad \sigma'_f \approx 14,0$$

$$\text{pour la série des hommes : } V'_h = 109 \quad \text{et} \quad \sigma'_h \approx 10,4$$

$$\text{pour la série totale : } V' \approx 313,2 \quad \text{et} \quad \sigma' \approx 17,7$$

i) Pour la série des femmes :

$$m_1 = \frac{3\,988}{30} \approx 132,9; \quad m_2 = \frac{536\,176}{30} \approx 17\,872,5$$

$$m_3 = \frac{1}{30} \sum_{i=1}^{30} x_{if}^3 = \frac{72\,872\,368}{30} \approx 2\,429\,079$$

$$m_4 = \frac{1}{30} \sum_{i=1}^{30} x_{if}^4 = \frac{10\,006\,377\,210}{30} \approx 333\,545\,907$$

$$\mu_2 = m_2 - m_1^2 \approx 201,3$$

$$\mu_3 = m_3 - 3m_1m_2 + 2m_1^3 \approx -285,4$$

$$\mu_4 = m_4 - 4m_1m_3 + 6m_1^2m_2 - 3m_1^4 \approx 84\,493,1$$

Coefficient γ_1 de Fisher : $\gamma_1 = \frac{\mu_3}{(\mu_2)^{3/2}} \approx -0,100$. Comme $\gamma_1 < 0$, la distribution est étalée vers la gauche.

Coefficient γ_2 de Fisher : $\gamma_2 = \frac{\mu_4}{(\mu_2)^2} - 3 \approx -0,914$. Comme $\gamma_2 < 0$, la distribution est plus aplatie que celle de la loi de Gauss réduite.

1-5 a) Paramètres des deux séries statistiques de pluviosité

Pour le chêne pédonculé :

$$\bar{P}_1 \approx 1\,073 \text{ mm}; \quad \sigma_1 \approx 200,7 \text{ mm}; \quad \frac{\sigma_1}{\bar{P}_1} \approx 0,19.$$

Pour le chêne pubescent :

$$\bar{P}_2 \approx 868,5 \text{ mm}; \quad \sigma_2 \approx 60,6 \text{ mm}; \quad \frac{\sigma_2}{\bar{P}_2} \approx 0,07.$$

b) Paramètres des deux séries statistiques de température

Pour le chêne pédonculé :

$$\bar{T}_1 \approx 10,7 \text{ }^\circ\text{C}; \quad \sigma_3 \approx 0,99 \text{ }^\circ\text{C}; \quad \frac{\sigma_3}{\bar{T}_1} \approx 0,09.$$

Pour le chêne pubescent :

$$\bar{T}_2 \approx 12,3 \text{ }^\circ\text{C}; \quad \sigma_4 \approx 0,41 \text{ }^\circ\text{C}; \quad \frac{\sigma_4}{\bar{T}_2} \approx 0,03.$$

c) Représentations graphiques de la nature des sols

Si on veut comparer la nature des sols habités par le chêne pédonculé et par le chêne pubescent, ce sont les pourcentages qui interviennent et non les effectifs. Pour ceci, on peut adopter des graphiques circulaires.

Sols	Chêne pédonculé		Chêne pubescent	
	%	angles en °	%	angles en °
acides	77,83	280	14,65	53
calcaires	7,60	27	85,35	307
montagneux	14,57	53	0	0
total	100,00	360	100,00	360

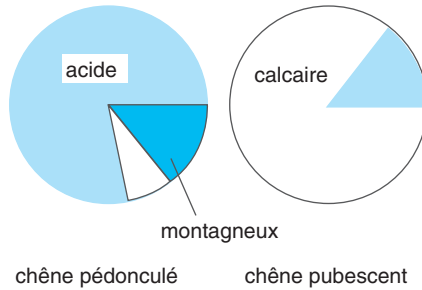


Figure 1-3

d) Conclusions écologiques

On observe que le chêne pubescent préfère un climat plus chaud ($\bar{T}_2 > \bar{T}_1$) et plus sec ($\bar{P}_2 < \bar{P}_1$) que le chêne pédonculé. Par ailleurs, les sols occupés par le chêne pédonculé sont le plus souvent acides et ceux occupés par le chêne pubescent sont souvent calcaires. On peut dire aussi que le chêne pédonculé est une espèce plus résistante car il accepte des températures et des précipitations plus variées (coefficients de variation plus élevés).

En fait, pour la série de végétation étudiée, c'est la nature du sol qui est le facteur primordial.