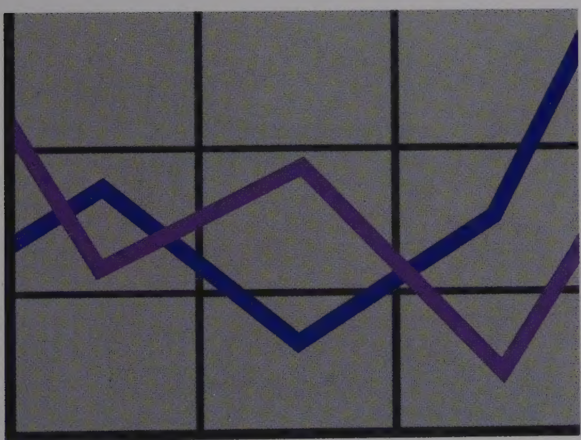
LIBRES COURS

Économie

Pierre Bailly
Christine Carrère

Statistiques descriptives
Exercices



PUG

Statistiques descriptives

Exercices

BP47 – 38040 Grenoble cedex 9

Tél. : 04 76 82 56 52 – Fax : 04 76 82 78 35

email : pug@pug.fr /www.pug.fr

ISBN 978 2 7061 14137

PIERRE BAILLY

CHRISTINE CARRÈRE

Statistiques descriptives

Exercices

Presses universitaires de Grenoble

Avant-Propos

Cet ouvrage est destiné à compléter le cours de statistique descriptive par des exercices pratiques d'illustration des principes statistiques exposés dans celui-ci. Il en est la suite logique et devrait permettre de mieux appréhender les méthodes de la statistique descriptive qui ne peuvent se concevoir exclusivement sur un plan purement théorique. Seule la confrontation à des problèmes pratiques permet d'acquérir une véritable maîtrise et donc une meilleure intelligence de ces techniques.

Ces exercices ont été classés en cinq chapitres : les outils, les distributions à une dimension, les distributions à deux dimensions, les séries chronologiques, les nombres indices. Chaque chapitre débute par une brève présentation du contenu.

L'objet de cet ouvrage est l'acquisition des techniques de la statistique descriptive par la répétition des principes de base permettant la maîtrise des concepts utilisés. Un corrigé détaillé est proposé pour chacun des exercices, il constitue un moyen d'auto vérification. En effet, la manière la plus efficace de travailler consiste à chercher à résoudre par soi même chaque exercice. Ensuite, l'utilisateur pourra confronter sa solution à celle fournie, donc être mieux à même de comprendre les difficultés et problèmes soulevés.

Chapitre 1

Les outils

Nous présentons dans ce chapitre des exercices sur quatre thèmes : les nomenclatures et les types de variable, les tableaux statistiques, les représentations graphiques, l'utilisation des pourcentages et des taux.

1. Types de variable, variable ou caractère

Mots-clefs

Variable discrète, variable continue, caractère qualitatif

Énoncé

Quelle est la nature des caractères ci-dessous ?

Nombre d'actions vendues chaque jour à la bourse

Rémunérations des enseignants d'un lycée

Indicateur du moral des ménages

Écart de rémunération entre hommes et femmes

Les pays de l'Union européenne

Les niveaux de formation des salariés

Les formes de contrat de travail

Taux de croissance du PIB

Prix à la consommation

Solde commercial

Nombre de personnes par ménages

Corrigé

Nombre d'actions vendues chaque jour à la bourse	variable discrète
Rémunérations des enseignants d'un lycée	variable quantitative continue
Indicateur du moral des ménages	variable qualitative ordonnée
Écart de rémunération entre hommes et femmes	variable continue

Les pays de l'Union européenne	caractère qualitatif
Les niveaux de formation des salariés	variable ordonnée
Les formes de contrat de travail	caractère qualitatif
Taux de croissance du PIB	variable quantitative
Prix à la consommation	variable quantitative
Solde commercial	variable quantitative
Nombre de personnes par ménage	variable statistique discrète.

2. UTILISATION DE LA NAF

Mots-clefs

Caractère qualitatif, nomenclature

Énoncé

1. Quel est le code de la nomenclature NAF 31 correspondant à la « Fabrication d'équipements électriques et électroniques » ?

2. Quelle est l'activité codée LL ?

Vous trouverez la nomenclature nécessaire p. 125 du livre de cours.

Corrigé

1. Le code de la nomenclature NAF correspondant à la « Fabrication d'équipements électriques et électroniques » est obtenu par lecture de la nomenclature. Le code est DL.

2. Le code LL correspond à l'activité « Administration publique ».

3. UTILISATION D'UNE NOMENCLATURE DE L'UNION EUROPÉENNE

Mots-clefs

Caractère qualitatif

Énoncé

Quelle est l'activité correspondant au E de la NACE-CE ?

Vous trouverez la nomenclature nécessaire p. 126 du livre de cours.

Corrigé

L'activité correspondant au code E est la « Production et distribution d'électricité, de gaz et d'eau ».

4. NIVEAUX DE FORMATION

Mots-clefs

Caractère qualitatif ordonné

Énoncé

Quel est le niveau de formation d'un étudiant qui, ayant suivi les cours de première année du DEUG de sociologie, n'a pas obtenu son passage en seconde année et quitte l'université ?

Vous trouverez la nomenclature nécessaire page 126 du livre de cours.

Corrigé

Le niveau de formation de cet étudiant sera IV, plus précisément IV sup. La nomenclature des formations est un caractère qualitatif ordonné.

5. NOMBRE DE PERSONNES DANS LES MÉNAGES

Mots-clefs

Variable discrète, tableau statistique

Énoncé

Nous disposons de la distribution des ménages selon leur composition.

Ménages suivant le nombre de personnes du ménage en France en 1995

Ensemble (en milliers)	23 126
Soit en pourcentage, suivant le nombre de personnes dans le ménage	
1 personne	29,2
2 personnes	31,8
3 personnes	16,8
4 personnes	14,2
5 personnes et plus	8,0

Source : TEF 1998/99

1. Construisez le tableau statistique en calculant les effectifs pour chacune des catégories de ménages.
2. Combien de personnes ont été comptées dans cette étude ?

Corrigé

Un ménage est constitué des personnes occupant une même unité d'habitation.

1. La construction du tableau statistique nécessite de calculer l'effectif de chaque catégorie de ménages. Il est obtenu en multipliant le nombre total des ménages par son importance relative. Les résultats ont été arrondis au millier d'unité près.

Par exemple, le calcul du nombre de ménages comprenant trois personnes est le suivant :

$$\text{Effectifs des ménages de trois personnes} = 23\,126 \cdot 0,168 = 3\,885,168 \cong 3\,885$$

Effectifs des ménages suivant le nombre de personnes dans le ménage

Valeurs de la variable	Fréquences	Pourcentages	Effectifs (milliers)
x_i	f_i	p_i	n_i
1 personne	0,292	29,2	6 753
2 personnes	0,318	31,8	7 354
3 personnes	0,168	16,8	3 885
4 personnes	0,142	14,2	3 284
5 personnes et plus	0,080	8,0	1 850
Total	1,000	100,0	23 126

2. Pour calculer le nombre de personnes concernées par l'étude, nous devons faire une hypothèse sur la taille des ménages de la classe « 5 personnes et plus ». Dans le cas où nous considérerions que la taille moyenne de cette catégorie de ménages est de 6, nous obtiendrions une population de 57 352 milliers de personnes. Si nous retenons l'hypothèse de 7 personnes par ménage dans cette classe, nous avons une population de 59 202 milliers de personnes.

Si nous retenons la première hypothèse, le détail du calcul est :

$$1 \cdot 6\,753 + 2 \cdot 7\,354 + 3 \cdot 3\,885 + 4 \cdot 3\,284 + 6 \cdot 1\,850 = 57\,352$$

Si nous retenons la seconde hypothèse, le détail du calcul est :

$$1 \cdot 6\,753 + 2 \cdot 7\,354 + 3 \cdot 3\,885 + 4 \cdot 3\,284 + 7 \cdot 1\,850 = 59\,202\,.$$

6. APPELS TÉLÉPHONIQUES

Mots-clefs

Tableau statistique, variable continue

Énoncé

Vous disposez d'une facture détaillée des appels d'un abonné à France Télécom sur la période 09/03/05-05/05/05.

Effectuez le regroupement en classes de ces données selon les trois variables continues. Vous veillerez à ce que les classes que vous avez choisies respectent les conditions qu'on attend d'un regroupement en classes.

1. Plage horaire de l'appel
2. Durée des appels
3. Montant des appels.

Détail des appels

Date	Heure	Durée	Montant
jj.mm		mm:ss	Hors taxe en centime d'euro
09.03	11:12	11:25	2,570
11.03	21:16	06:38	1,040
14.03	09:40	01:29	3,070
15.03	17:04	00:19	0,610
15.03	17:28	02:17	1,970
15.03	18:31	02:24	2,070
15.03	20:27	07:10	1,100
15.03	20:53	06:47	3,160
16.03	16:15	14:38	3,310
16.03	20:05	04:27	2,190
17.03	10:41	05:25	1,180
17.03	14:36	00:34	2,460
17.03	15:17	10:21	2,320
17.03	16:06	07:39	1,690
17.03	21:05	12:17	1,690
23.03	11:31	00:35	2,460
24.03	13:30	21:14	4,850
24.03	21:22	27:02	3,400
25.03	16:34	00:10	2,460
25.03	21:19	00:37	2,460
26.03	18:58	01:44	4,280
27.03	16:31	03:44	0,700
27.03	22:51	03:47	0,710
30.03	15:34	00:37	2,460
01.04	18:48	05:45	1,250
01.04	21:55	10:52	1,530
03.04	10:09	00:44	3,070
04.04	20:55	03:18	0,650
05.04	20:47	06:35	3,080
06.04	20:56	08:00	1,190
07.04	20:20	07:41	1,160
10.04	17:14	00:42	1,840
21.04	08:58	05:23	1,170
21.04	09:28	04:04	0,860
21.04	09:34	06:35	1,450

Date	Heure	Durée	Montant
jj.mm		mm:ss	Hors taxe en centime d'euro
24.04	14:21	00:26	2,460
26.04	09:05	00:31	3,070
26.04	20:50	14:01	1,890
26.04	21:40	03:33	0,680
26.04	21:44	10:29	1,480
27.04	20:42	03:03	0,620
27.04	20:48	08:19	1,230
27.04	20:57	08:49	1,290
27.04	21:11	04:49	0,830
27.04	21:17	05:10	0,870
27.04	21:22	03:16	0,650
29.04	20:29	05:46	0,940
01.05	14:01	01:33	3,070
01.05	17:11	03:17	0,650
03.05	20:38	09:40	1,390
03.05	20:57	12:22	1,700
04.05	16:13	20:31	4,680
04.05	20:28	03:04	0,620
05.05	14:19	08:36	1,910
05.05	14:49	26:28	6,060
05.05	19:54	05:48	14,400

Corrigé

1. Pour effectuer le dépouillement sur la plage horaire de l'appel, nous constatons qu'aucun appel n'a lieu avant 8h et aucun après 23h. Nous choisissons des plages de deux heures démarrant à 8h, la dernière plage sera une plage d'une heure.

Plage horaire de l'appel

Classes	Centres des classes	Effectifs	Fréquences	Pourcentages	Fréquences cumulées
	c_i	n_i	f_i	$p_i\%$	F_i
[8 ; 10[		5			
[10 ; 12[		4			
[12 ; 14[		1			
[14 ; 16[		7			
[16 ; 18[		9			
[18 ; 20[		4			
[20 ; 22[		25			
[22 ; 23]		1			
Total		56			

Avant d'aller plus loin dans le tableau, nous constatons un déséquilibre au niveau de la plage entre 20h et 22h. La perte d'information est importante et peut facilement être réduite en utilisant des plages d'une heure pour ce créneau horaire.

Plage horaire de l'appel

Classes	Centres des classes	Effectifs	Fréquences	Pourcentages	Fréquences cumulées
	c_i	n_i	f_i	$p_i\%$	F_i
[8 ; 10[	9	5	0,089	8,9	0,089
[10 ; 12[	11	4	0,071	7,1	0,161
[12 ; 14[	13	1	0,018	1,8	0,179
[14 ; 16[	15	7	0,125	12,5	0,304
[16 ; 18[	17	9	0,161	16,1	0,464
[18 ; 20[	19	4	0,071	7,1	0,536
[20 ; 21[	20,5	15	0,268	26,8	0,804
[21 ; 22]	21,5	10	0,179	17,9	0,982
[22 ; 23]	22,5	1	0,018	1,8	1,000
Total		56	1,000	100,0	

2.

Durée des appels

Classes	Centres des classes	Effectifs	Fréquences	Pourcentages	Fréquences cumulées
	c_i	n_i	f_i	$p_i\%$	F_i
[0 ; 5[		26			
[5 ; 10[		16			
[10 ; 15[		10			
[15 ; 20[		0			
[20 ; 25[		2			
[25 ; 30]		2			
Total		56			

Avant d'aller plus loin, constatons que les premières classes sont disproportionnées par rapport aux autres et qu'il y a une classe vide. Un autre découpage possible est explicité dans le tableau suivant.

Durée des appels (regroupement plus judicieux)

Classes	Centres des classes	Effectifs	Fréquences	Pourcentages	Fréquences cumulées
	c_i	n_i	f_i	$p_i\%$	F_i
[0 ; 2[	1	13	0,232	23,2	0,232
[2 ; 4[	3	10	0,179	17,9	0,411
[4 ; 6[	5	9	0,161	16,1	0,571
[6 ; 8[	7	7	0,125	12,5	0,696
[8 ; 10[	9	5	0,089	8,9	0,786
[10 ; 12[	11	4	0,071	7,1	0,857
[12 ; 14[	13	2	0,036	3,6	0,893
[14 ; 16[	15	2	0,036	3,6	0,929
[16 ; 30[	23	4	0,071	7,1	1,000
Total		56	1,000	100,0	

3.

Montants des appels

Classes	Centres des classes	Effectifs	Fréquences	Pourcentages	Fréquences cumulées
	c_i	n_i	f_i	$p_i\%$	F_i
[0 ; 1[	0,5	13	0,232	23,2	0,232
[1 ; 2[	1,5	20	0,357	35,7	0,589
[2 ; 3[	2,5	10	0,179	17,9	0,768
[3 ; 4[	3,5	8	0,143	14,3	0,911
[4 ; 5[	4,5	3	0,054	5,4	0,964
[5 ; 10[	7,5	1	0,018	1,8	0,982
[10 ; 15[	12,5	1	0,018	1,8	1,000
Total		56	1	100,0	

7. TABLE DE MORTALITÉ

Mots-clefs

Variable continue, tableau statistique

Énoncé

Extrait de la table de mortalité de la génération féminine française de 1899

Age exact	Survivants à l'âge exact
0	100 000
1	84 883
2	82 247
3	80 843
4	79 995
5	79 186
6	78 763
7	78 411

Source : « La mortalité par génération en France depuis 1899 »,
Travaux et documents, Cahier INED n°63, 1973

1. Présentez le tableau statistique de la variable « âge du décès » sous sa forme habituelle.
2. Donnez la signification concrète de chacune des colonnes du tableau statistique obtenu.

Corrigé

1. L'étude porte sur 100 000 filles nées en 1899 dont le décès est survenu avant l'âge de 7 ans.

Comment obtient-on les effectifs du tableau ? Nous allons prendre comme exemple le cas de la première classe. Nous savons qu'il y a eu 100 000 naissances ; un an plus tard seules 84 883 femmes sont encore vivantes, le nombre de décès est donc de $100\,000 - 84\,883 = 15\,117$. Nous répétons le raisonnement pour tous les âges, ce qui nous permet de construire le tableau statistique.

Répartition de la génération féminine française de 1899 suivant l'âge du décès

Classes	Centres des classes	Effectifs	Fréquences	Pourcentages	Fréquences cumulées
	c_i	n_i	f_i	$p_i\%$	F_i
[0 ; 1[	0,5	15117	0,700	70,0	0,700
[1 ; 2[	1,5	2636	0,122	12,2	0,822
[2 ; 3[	2,5	1404	0,065	6,5	0,887
[3 ; 4[	3,5	848	0,039	3,9	0,927
[4 ; 5[	4,5	809	0,037	3,7	0,964
[5 ; 6[	5,5	423	0,020	2,0	0,984
[6 ; 7]	6,5	352	0,016	1,6	1,000
		21589	1,000	100,0	

2. La première colonne reprend la tranche d'âge des décès. La deuxième représente l'âge moyen du décès par classe annuelle. La colonne des effectifs donne le nombre de femmes décédées dans la tranche d'âge considérée. Au total 21589 filles sont décédées avant l'âge de 7 ans. La colonne suivante donne la fréquence des femmes décédées dans la tranche d'âge, ainsi que la troisième qui exprime la même chose en pourcentage. La dernière colonne donne la fréquence des femmes décédées avant la borne supérieure de la tranche d'âge. La fréquence signifie que 92,7 % des femmes mortes avant l'âge de 7 ans sont mortes avant l'âge de 4 ans. Le tableau montre que la mortalité féminine est très forte au cours de la première année (70 % des décès).

8. Bilan des apports et des usages de l'eau

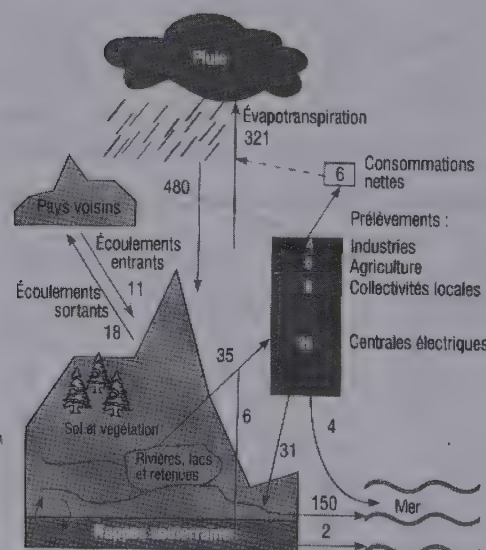
Mots-clefs

Analyse de graphique

Énoncé

L'exercice consiste à commenter un pictogramme

Bilan des apports et des usages des eaux continentales
(en milliards de m^3/an)



Source : TEF 1998/1999, Paris : INSEE

1. Quelle est la quantité nette d'eau consommée par les centrales électriques ?
2. Quel est le bilan des échanges d'eau avec les pays voisins ?
3. Quelle est l'équation d'équilibre des usages humains de l'eau ?
4. Quelle est l'équation d'équilibre des eaux continentales ?

Corrigé

1. Pour calculer la part des centrales électriques dans la quantité nette d'eau consommée, nous supposons que la « consommation » nette, en fait l'évaporation, est proportionnelle à la quantité utilisée. Cette hypothèse est sans doute une sous-estimation dans le cas des centrales électriques. Pour un usage total de 41 milliards de m^3 par an, 35 provenant des précipitations et 6 des nappes phréatiques, nous obtenons une consommation de :

$$6 \cdot \frac{26}{41} = 3{,}8 \text{ milliards de m}^3$$

2. La France reçoit 11 milliards de m^3 par écoulement, les écoulements vers les pays voisins s'élèvent à 18 milliards de m^3, donc le solde des échanges extérieurs est de – 7 milliards de m^3.

3. Les activités humaines prélèvent 35 milliards de m^3 dans « les rivières, lacs et retenues » et 6 milliards de m^3 dans les nappes phréatiques soit 41 milliards de m^3.

Les usages humains se traduisent par une évapotranspiration de 6 milliards de m^3, un écoulement de 4 milliards de m^3 vers la mer et de 31 milliards de m^3 dans les « rivières, lacs et retenues », soit également 41 milliards de m^3.

Emplois – ressources des usages humains (milliards de m^3)

Emplois		Ressources	
Consommations nettes (évaporation)	6	Prélèvements dans les « rivières, lacs et retenues »	35
Écoulement vers la mer	4	Ponctions dans les nappes phréatiques	6
Rejets vers les « rivières, lacs et retenues »	31		
Total	41	Total	41

4. Les apports sont de 480 milliards de m^3 de précipitations (pluie, neige) moins 7 milliards de m^3 provenant des échanges avec les pays voisins. Les apports sont de 473 milliards de m^3.

Les usages comprennent 321 milliards de m^3 sous forme d'évapotranspiration, 150 milliards de m^3 sous forme de ruissellement et 2 milliards de m^3 d'écoulement des nappes phréatiques vers la mer. Nous obtenons bien 473 milliards de m^3.

Emplois – ressources des eaux continentales (milliards de m³)

Emplois		Ressources	
Évapotranspiration	321	Précipitations	480
Ruissellement	150	Échanges avec les pays voisins	– 7
Écoulement vers la mer	2		
Total	473	Total	473

9. Analyse de la médicalisation

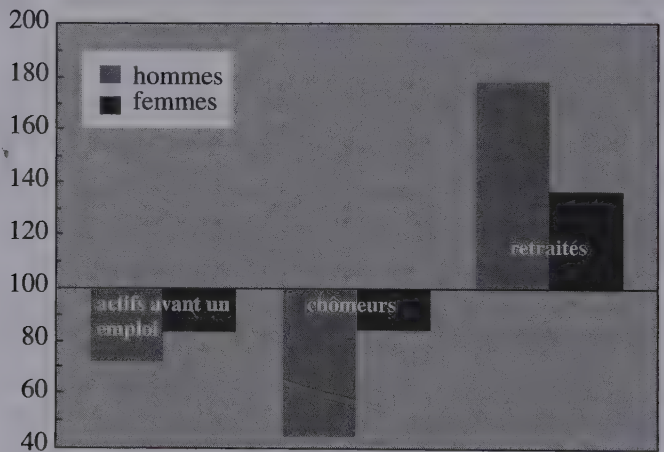
Mots-clefs

Commentaire de graphique, indice

Énoncé

Commentez le graphique ci-dessous selon les catégories et selon les genres.

Disparité de la médicalisation



Source : TEF 98/99

Note : 100 correspond à l'indice du nombre moyen de séances de médecins par sexe soit 5,28 séances pour les hommes et 8,36 séances pour les femmes.

Corrigé

La disparité de la médicalisation par catégorie apparaît dans ce graphique. Les actifs et les chômeurs ont moins recours au médecin que les retraités comme on pouvait s'y attendre. Les actifs et les chômeurs ont globalement une consommation inférieure à la moyenne.

Globalement, les femmes consultent beaucoup plus souvent que les hommes soit 5,28 séances pour les hommes et 8,36 séances pour les femme dans un rapport de $\dfrac{8{,}36}{5{,}28} = 1{,}583 \cong 1{,}6$. Cette situation s'inverse pour la catégorie « retraités ».

Les femmes « actives » ont un recours au médecin aussi fréquent que les « chômeuses ». Elles ont une consommation médicale plus importante que les hommes de même catégorie. Pour la population active (actifs ayant un emploi et chômeurs), les hommes s'occupent moins de leur santé que les femmes. L'écart entre les hommes et les femmes est beaucoup plus important dans la catégorie « chômeurs ». Les hommes au chômage réduisent très sensiblement leur consommation médicale auprès des médecins alors que les femmes en profitent « pour mieux se soigner ». Par contre les hommes retraités ont une consommation médicale bien supérieure à celle des femmes. Faut-il y voir les effets d'une moindre consommation antérieure ?

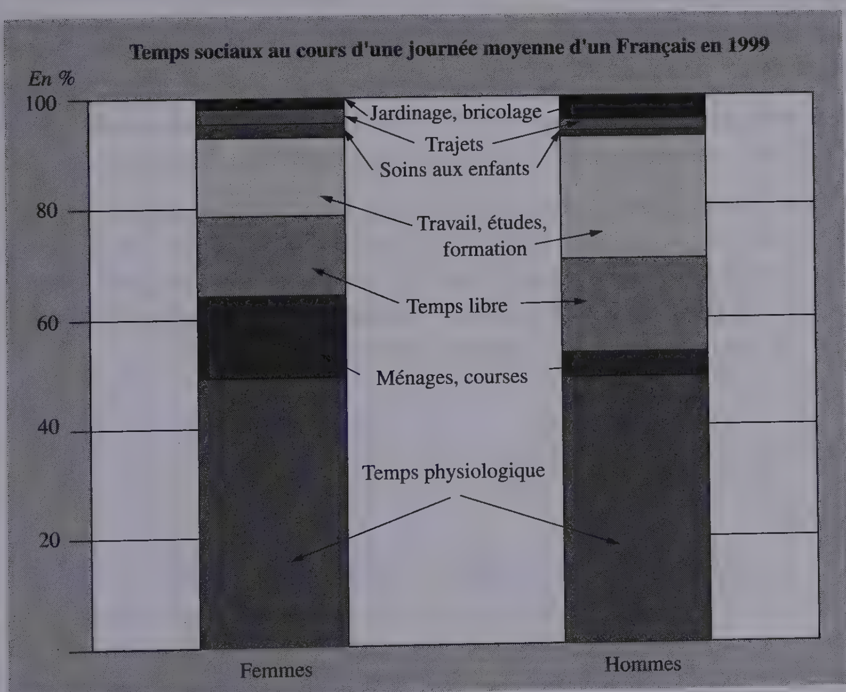
10. Structure des temps sociaux féminins et masculins

Mots-clefs

Diagramme à cumul interne, analyse de diagramme

Énoncé

La parité et le temps libre



Source : TEF 98/99

Quelles informations vous donnent ce graphique sur les différences dans les temps sociaux entre les hommes et les femmes ?

Corrigé

Les temps sociaux se subdivisent en quatre périodes : le temps « physiologique » consacré à dormir, se laver, manger, etc., le temps de travail professionnel ou d'études, celui consacré aux travaux domestiques tels que le ménage, la lessive, les courses, etc., et le temps des loisirs qui comprend les promenades, la télévision, la pratique d'un sport, la lecture, etc.

Les temps physiologiques sont proches pour les hommes et les femmes avec un très léger avantage aux femmes.

Les temps de « travail, études, formation » sont dans un rapport du simple au double entre les hommes et les femmes. Cela tient d'une part au fait que les femmes inactives sont plus nombreuses que les hommes inactifs et d'autre part que les femmes ont plus souvent que les hommes des emplois à temps partiel. Les emplois à temps partiel sont en majorité occupés par les femmes. Cette dernière information provient d'une connaissance du domaine de l'économie du travail ce qui permet une analyse plus complète du graphique.

Le travail domestique est essentiellement un temps féminin en particulier pour ce qui concerne le ménage, les courses et les soins aux enfants. Le temps féminin est plus que le double du temps masculin dans ces domaines. L'activité domestique est une activité féminine.

En définitive, les femmes disposent, en moyenne de légèrement moins de temps libre que les hommes, la différence reste néanmoins peu significative.

D'un point de vue global, ce graphique donne l'impression qu'entre les hommes et les femmes, il y a échange entre temps de travail (masculin) et activité domestique (féminine). Le modèle traditionnel de partage des tâches entre les hommes et les femmes subsiste pour l'essentiel. Il faudrait, pour préciser les évolutions, disposer de données pour des périodes antérieures.

11. APPELS TÉLÉPHONIQUES

Mots-clefs

Variable discrète, diagramme en bâtons, diagramme cumulatif

Énoncé

Fréquences des appels téléphoniques

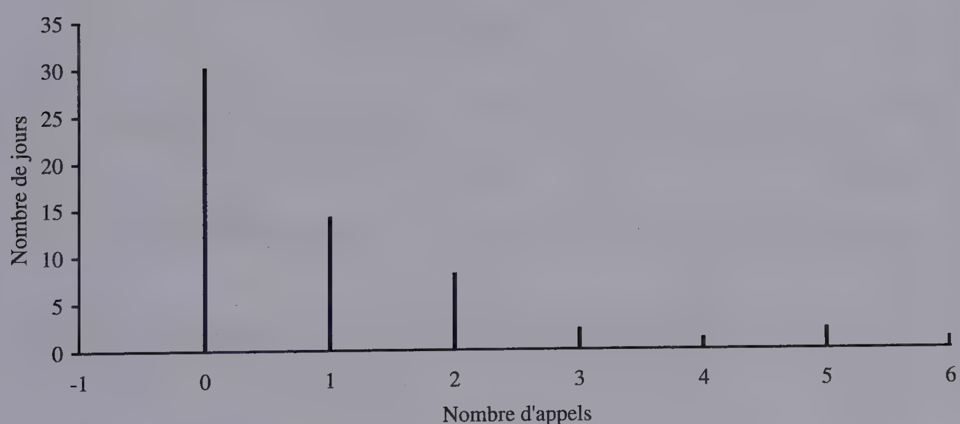
Nombre d'appels	Nombre de jours	Fréquences	Fréquences cumulées
x_i	n_i	f_i	F_i
0	30	0,517	0,517
1	14	0,241	0,759
2	8	0,138	0,897
3	2	0,034	0,931
4	1	0,017	0,948
5	2	0,034	0,983
6	1	0,017	1,000
	58	1,000	

1. Construisez le graphique des effectifs des appels.
2. Construisez le diagramme cumulatif.

Corrigé

1. Nous construisons un diagramme en bâtons appelé diagramme différentiel car il représente les différentes modalités de la variable. Ce diagramme s'impose puisque nous avons une variable discrète.

Nombre d'appels par jour

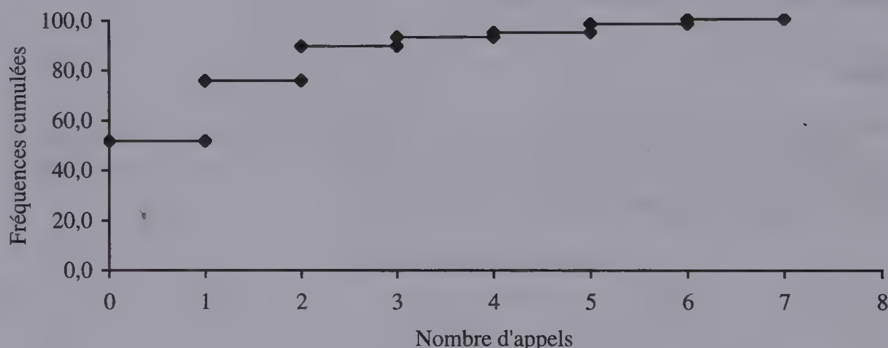


2. Le diagramme cumulatif est également appelé diagramme intégral au sens de l'intégration mathématique. Il représente le graphe des fréquences cumulées. C'est un graphique « en escalier », les valeurs ne sont connues que sur des intervalles en raison même de la nature de la variable.

Tableau des appels cumulés

Nombre d'appels	Nombre de jours	Fréquences cumulées
x_i	n_i	F_i
0	30	0,517
1	14	0,759
2	8	0,897
3	2	0,931
4	1	0,948
5	2	0,983
6	1	1,000
	58	

Diagramme cumulatif



12. ÉVOLUTIONS DES PROFESSIONS ET CATÉGORIES SOCIALES

Mots-clefs

Représentation en secteurs, caractère qualitatif, représentation en barre

Énoncé

Les tableaux suivants fournissent des informations sur l'importance des professions et catégories sociales dans la population active occupée en 1982, en 1990 et en 2005.

1. Définissez en quelques lignes ces différentes catégories.
2. Donnez des représentations graphiques qui fassent apparaître l'importance relative des différentes catégories sociales pour l'année 1982, pour l'année 1990 et pour l'année 2005.

3. Donnez une représentation graphique qui fasse apparaître les évolutions entre les différentes années.

Les professions et catégories sociales

Catégories	1982	1990	2005
Agriculteurs exploitants	1 475	1 162	651
Artisans, commerçants, chefs d'entreprises	1 835	1 820	1 505
Cadres et professions intellectuelles supérieures	1 895	2 435	3 660
Professions intermédiaires	3 971	4 308	5 745
Employés	6 247	5 925	7 232
Ouvriers	7 749	6 265	5 972

Source : INSEE

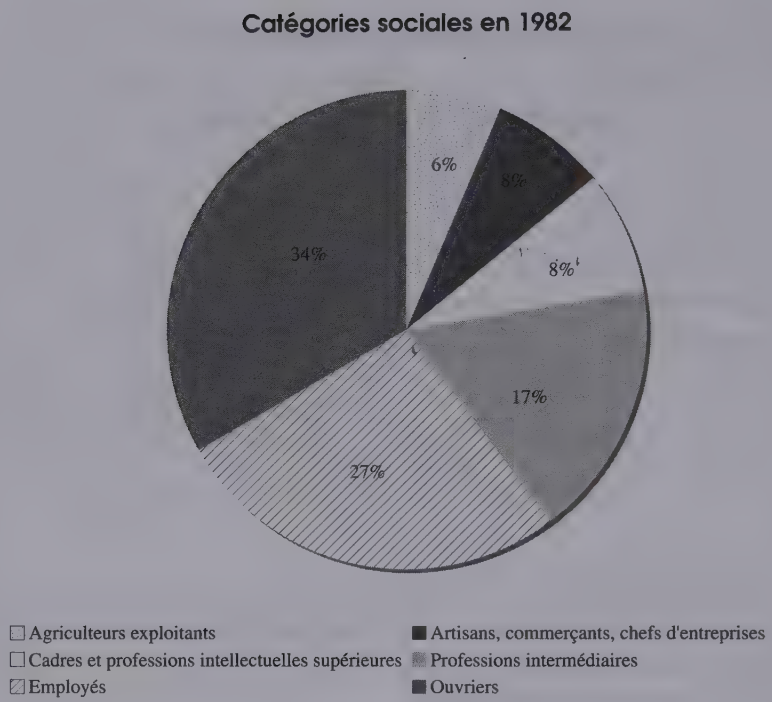
Corrigé

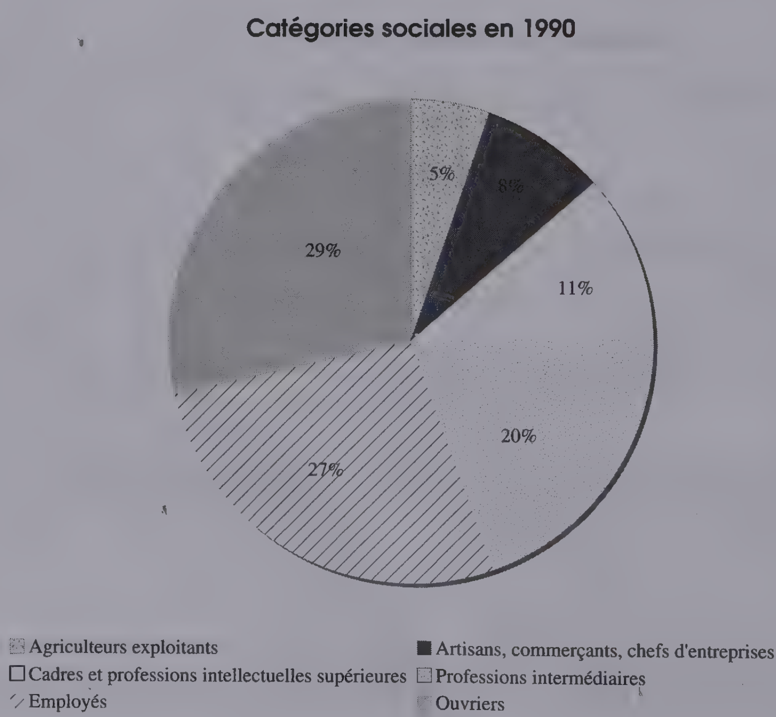
1. Voir correction de l'exercice 25 page 56.

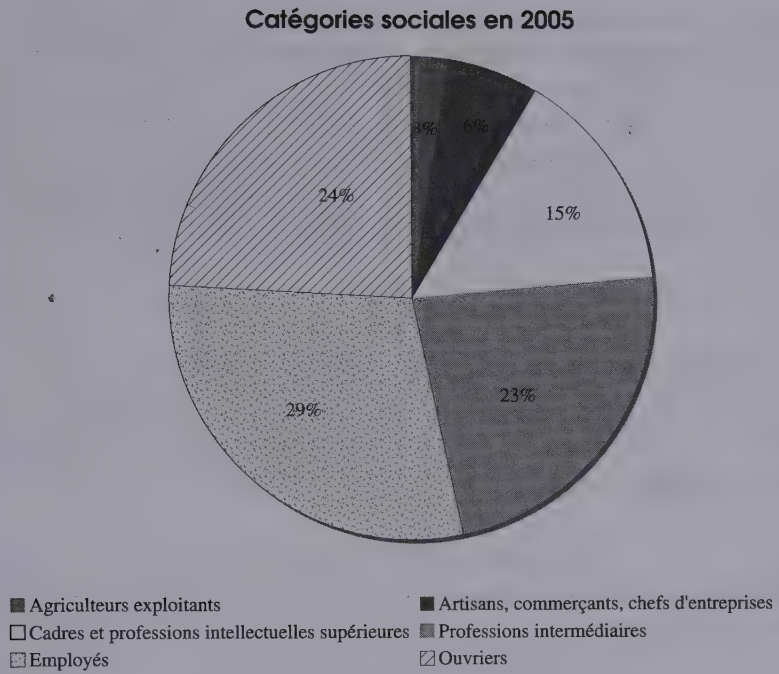
2. Nous utilisons une représentation en secteur qui fait bien apparaître l'importance relative de chaque catégorie. Le tableau ci-dessous fournit les données pour la construction des graphiques.

Importance relative des différentes catégories

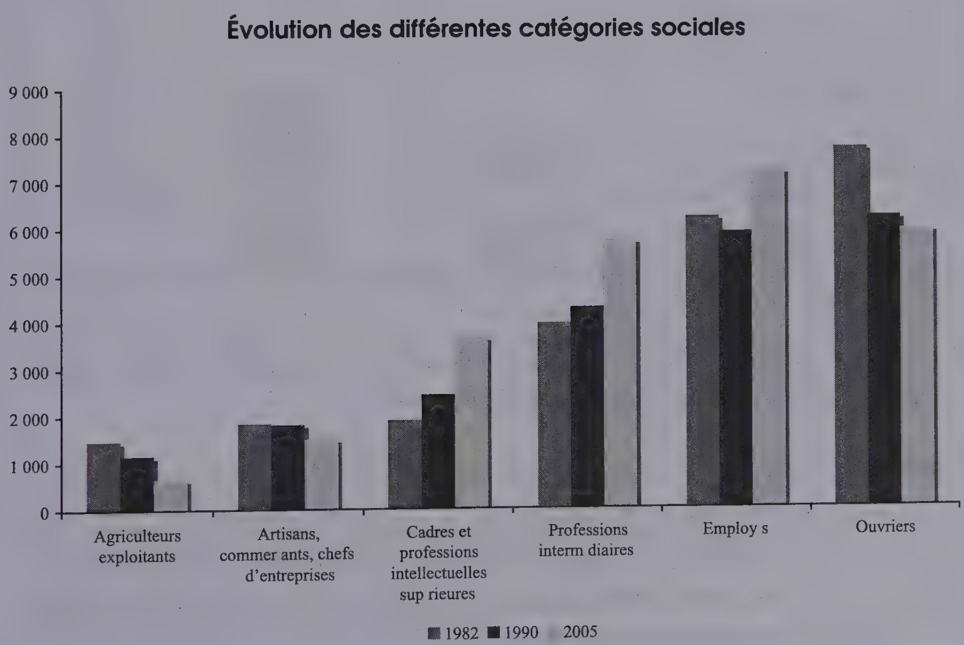
Catégorie	Effectifs n_i (milliers)			Fréquences f_i (%)		
	1982	1990	2005	1982	1990	2005
Agriculteurs exploitants	1 475	1 162	651	6,4	5,3	2,6
Artisans, commerçants, chefs d'entreprises	1 835	1 820	1 505	7,9	8,3	6,1
Cadres et professions intellectuelles supérieures	1 895	2 435	3 660	8,2	11,1	14,8
Professions intermédiaires	3 971	4 308	5 745	17,1	19,7	23,2
Employés	6 247	5 925	7 232	27,0	27,0	29,2
Ouvriers	7 749	6 265	5 972	33,4	28,6	24,1
Ensemble	23 172	21 915	24 765	100,0	100,0	100,0


Catégories sociales en 1982
6%
8%
8%
17%
27%
34%
Agriculteurs exploitants
Artisans, commerçants, chefs d'entreprises
Cadres et professions intellectuelles supérieures
Professions intermédiaires
Employés
Ouvriers


Catégories sociales en 1990
5%
8%
11%
20%
27%
29%
Agriculteurs exploitants
Artisans, commerçants, chefs d'entreprises
Cadres et professions intellectuelles supérieures
Professions intermédiaires
Employés
Ouvriers



3. Nous utilisons un diagramme qui fait apparaître les évolutions des différentes catégories sociales pour les années étudiées.



Ce graphique montre bien la diminution absolue du nombre des indépendants « Agriculteurs exploitants », « Artisans, commerçants et chefs d'entreprise », en particulier des agriculteurs, ainsi que de la catégorie « Ouvriers ». La catégorie sociale modale en France en 2005 est celle des employés, dont nous savons par ailleurs que le taux de féminisation est élevé. Ce graphique illustre également l'importance croissante de la catégorie des « Professions intermédiaires » presque aussi nombreuse que celle des « Ouvriers » ainsi que la place croissante des « Cadres ».

13. ÉVOLUTION DES REJETS DE DIOXYDE DE SOUFRE

Mots-clefs

Pourcentage, graphique en secteurs

Énoncé

L'évolution de la pollution en dioxyde de soufre en France est donnée dans le tableau suivant par secteur.

La structure des rejets par source

	1980 (%)	1990 (%)	1994 (%)
Résidences et bureaux	12,6	15,1	14,5
Industrie	31,8	21,6	22,0
Centrales électrothermiques	36,5	26,1	17,3
Transformations d'énergie	6,3	10,2	12,6
Procédés industriels	9,0	14,9	16,5
Transports	3,8	12,1	17,2
Ensemble	100,0	100,0	100,0

Source : TEF 93/94.

Cette pollution représentait 3 348 milliers de tonnes en 1980 et 1 200 milliers de tonnes en 1990, 961 milliers de tonnes en 1994.

Représentez graphiquement ces trois distributions ; le graphique devra rendre compte de la décroissance relative du phénomène.

Corrigé

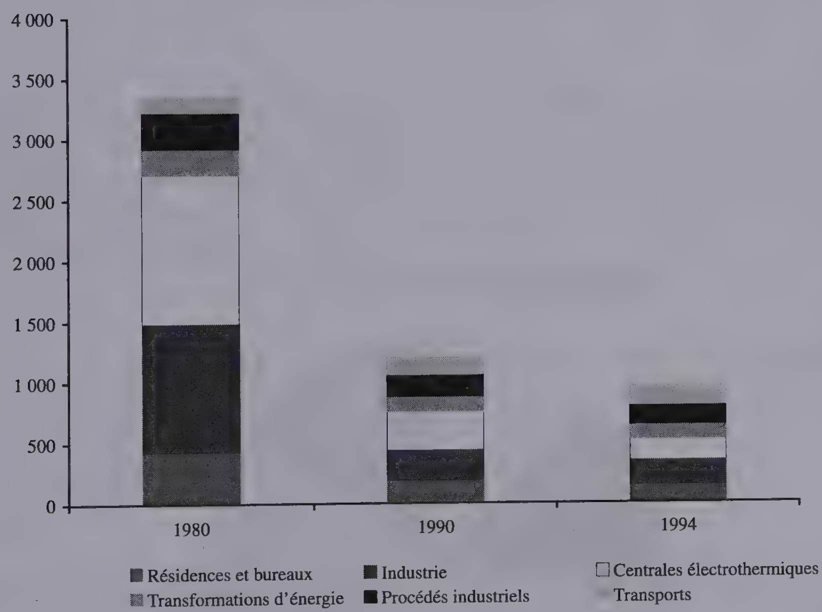
Parmi les possibilités de représentations graphiques, une solution est de faire apparaître d'une part les modifications relatives des sources de pollution au dioxyde, et d'autre part les évolutions de la pollution au cours du temps. Ce qui nous conduit à construire le tableau des quantités.

Les quantités de rejet

	1980	1990	1994
Résidences et bureaux	421,8	181,2	139,3
Industrie	1064,7	259,2	211,4
Centrales électrothermiques	1222,0	313,2	166,3
Transformations d'énergie	210,9	122,4	121,1
Procédés industriels	301,3	178,8	158,6
Transports	127,2	145,2	165,3
Ensemble	3348,0	1200,0	961,0

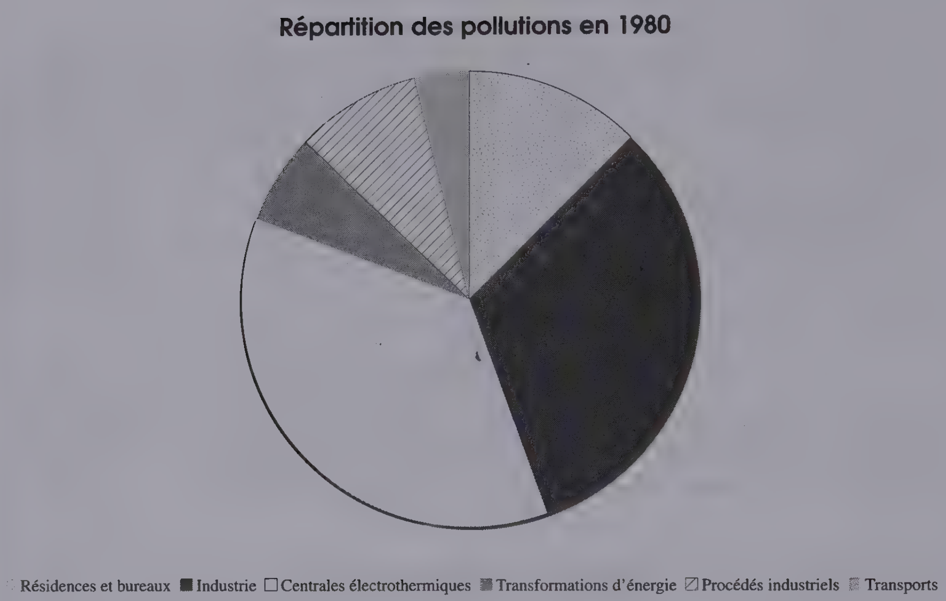
Le premier graphique retenu est en « tuyaux d'orgue », il montre la forte décroissance du volume de la pollution en dioxyde de soufre. Les modifications de l'importance relative des sources de pollution est peu lisible.

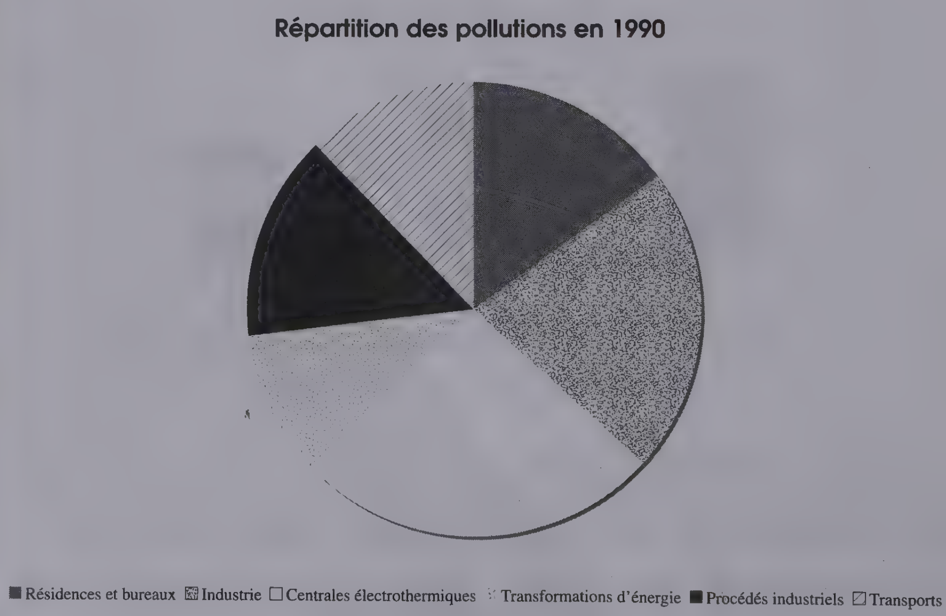
Évolution du volume des rejets



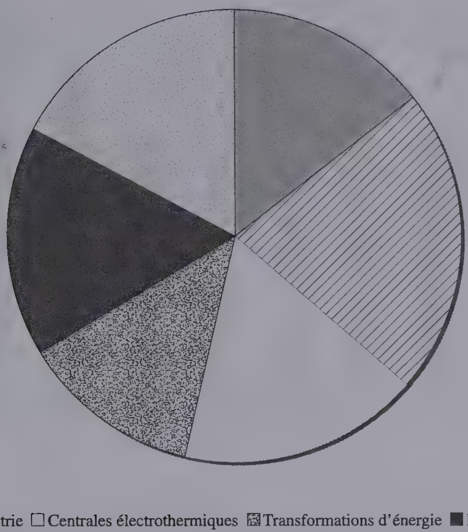
Le volume de la pollution s'est considérablement réduit entre 1980 et 1994. Parmi les raisons expliquant cette évolution, nous pouvons évoquer la lutte contre la pollution et la diminution des activités industrielles sur le territoire de la France.

Pour mettre en lumière les modifications de la structure des diverses sources de rejets de dioxyde de soufre nous utiliserons des graphiques en secteurs.


Répartition des pollutions en 1980
Résidences et bureaux
Industrie
Centrales électrothermiques
Transformations d'énergie
Procédés industriels
Transports


Répartition des pollutions en 1990
Résidences et bureaux
Industrie
Centrales électrothermiques
Transformations d'énergie
Procédés industriels
Transports

Répartition des pollutions en 1994



Cet ensemble de graphiques permet de voir que la réduction de la pollution n'est pas homothétique. Très globalement, les activités de consommation (transports, résidences) voient leur importance augmenter relativement aux rejets industriels. La réduction importante de la pollution se traduit par une répartition différente des principales sources de production.

14. HISTOGRAMME D'UN TABLEUR ET HISTOGRAMME STATISTIQUE

Mots-clefs

Histogramme, variable continue

Énoncé

Les tableurs disponibles proposent une représentation graphique baptisée « histogramme ». Cet exercice vise à expliciter la différence entre les deux représentations, il suppose que le lecteur dispose d'un tableur.

Soit la répartition des revenus d'une population présentée dans le tableau suivant en milliers d'euros.

1. Représentez cette distribution à l'aide de la fonction « histogramme » d'un tableur.
2. Construisez l'histogramme statistique de la série.

Répartition des revenus

Classes	Effectifs (n_i)
[0 ; 40[	4 500
[40 ; 50[	8 500
[50 ; 60[	9 000
[60 ; 70[	8 500
[70 ; 80[	6 500
[80 ; 90[	6 000
[90 ; 100[	6 000
[100 ; 120[	6 000
[120 ; 150[	7 000
[150 ; 180[	6 000
[180 ; 250[	3 000
[300 ; 400[	3 500
	74 500

Corrigé

1. Nous allons tout d'abord construire l' « histogramme » avec cette fonction d'un tableur.

Données pour « l'histogramme » du tableur

Classes	Effectifs
1	4 500
2	8 500
3	9 000
4	8 500
5	6 500
6	6 000
7	6 000
8	6 000
9	7 000
10	6 000
11	3 000
12	3 500

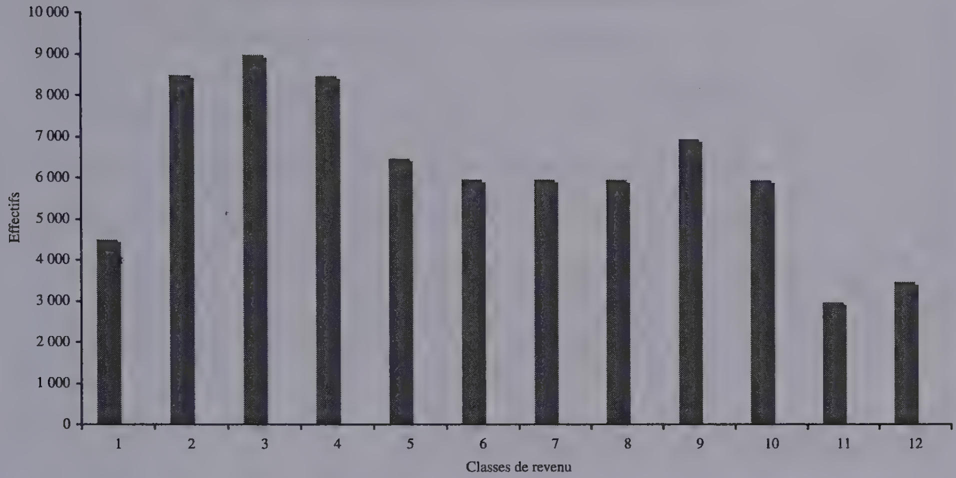
Nous obtenons le graphique ci-contre.

En quoi cette représentation est-elle insatisfaisante ?

La première raison tient au fait que la représentation donne une image d'une variable discrète, or nous avons une variable continue. La contradiction entre le caractère continu de la variable et sa représentation graphique sous forme discrète rend sans intérêt une telle « représentation » qui justement ne représente pas la nature de la variable.

La seconde raison du caractère insatisfaisant de cette représentation s'explique par l'absence de prise en compte des différences dans l'amplitude des classes.

« Histogramme » selon un tableur



Cette fois encore le graphique est infidèle à ce qu'il veut représenter.

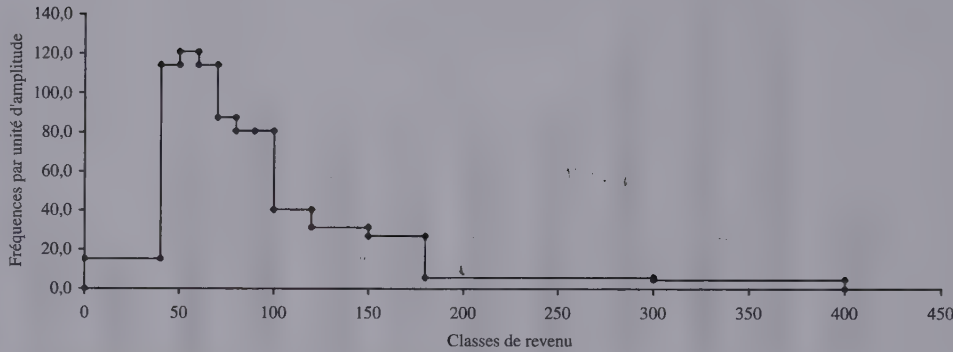
2. L'histogramme statistique va rendre compte du caractère continu de la variable et des amplitudes différentes.

Tableau statistique pour la construction de l'histogramme statistique

Classes	n_i	c_i	f_i	$\frac{f_i}{a_i} \times 100$
[0 ; 40[	4 500	20,0	6,0	15,1
[40 ; 50[	8 500	45,0	11,4	114,1
[50 ; 60[	9 000	55,0	12,1	120,8
[60 ; 70[	8 500	65,0	11,4	114,1
[70 ; 80[	6 500	75,0	8,7	87,2
[80 ; 90[	6 000	85,0	8,1	80,5
[90 ; 100[	6 000	95,0	8,1	80,5
[100 ; 120[	6 000	110,0	8,1	40,3
[120 ; 150[	7 000	135,0	9,4	31,3
[150 ; 180[	6 000	165,0	8,1	26,8
[180 ; 250[	3 000	215,0	4,0	5,8
[300 ; 400[	3 500	350,0	4,7	4,7
	74 500		100,0	

Le graphique suivant rend compte du caractère continu de la variable et du fait que les classes ont des amplitudes différentes. Cet histogramme est également construit à l'aide d un tableur mais pas en utilisant la fonction histogramme.

Histogramme de la distribution des revenus



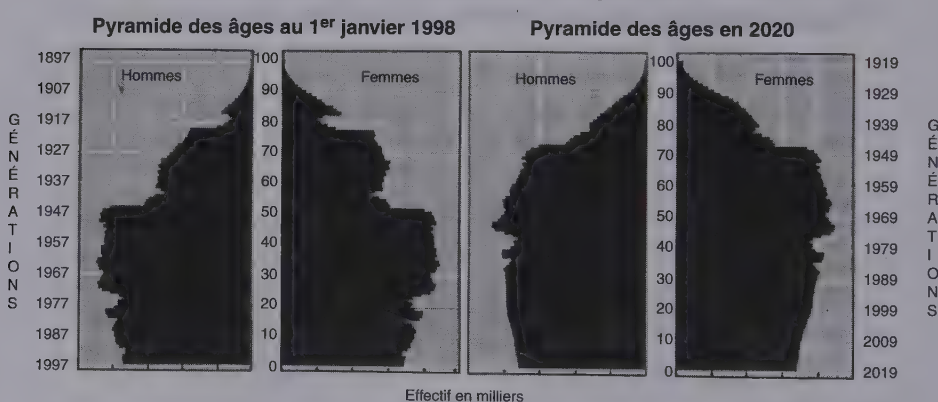
15. ANALYSE DE LA PYRAMIDE DES ÂGES

Mots-clefs

Histogramme, analyse de graphique, représentation graphique

Énoncé

Pyramides des âges de la population française en 1998 et 2020



Population par groupe d'âge
Âge moyen au 1er janvier

Années	Moins de 20 ans %	20 ans à 59 ans %	60 ans et plus %	Âge moyen années
1946	29,5	54,5	16,0	35,6
1970	33,2	48,8	18,0	34,8
1980	30,6	52,4	17,0	35,7
1990	27,8	53,2	19,0	36,9
1996	26,0	54,0	20,0	37,9
1997 *p*	25,9	53,8	20,3	38,1
1998 *p*	25,8	51,8	20,4	38,3

Projection* de population par groupe d'âge à l'horizon 2020

Années	Moins de 20 ans %	20 ans à 59 ans %	60 ans et plus %	Population totale milliers
2000	25,9	53,6	20,5	59 412
2005	25,0	54,2	20,8	60 642
2010	24,2	53,0	22,8	61 721
2015	23,4	51,7	24,9	62 648
2020	22,7	50,5	26,8	63 453

* Hypothèse de fécondité = 1,8 enfant par femme.

Source : TEF 98/99

1. À quel type de représentation graphique fait appel la pyramide des âges ?
2. Quelles réflexions vous inspirent la comparaison entre les deux pyramides ?
3. Comparez la distribution des hommes à celles des femmes en 1998 ? Vos conclusions sont-elles les mêmes en 2020 ?
4. La population est-elle en croissance ? Justifiez votre réponse.
5. La population est-elle vieillissante ? Justifiez votre réponse.
6. Représentez l'évolution des tranches d'âge sur la période 1970/2020.

Corrigé

1. Les variables représentées sont les âges des hommes et des femmes, vivant à une date donnée. Il s'agit donc de variables continues. Les pyramides sont des histogrammes horizontaux.

2. La comparaison entre la pyramide des âges au 1er janvier 1998 et celle qui est prévue en 2020 montre un épaississement de la courbe des plus de 50 ans accompagné de la disparition du creux de la génération 1915, le creux de la génération 1940 étant encore perceptible. Les jeunes générations sont beaucoup plus lisses dans la pyramide de 2020 car les chiffres utilisés sont des prévisions de population. On ne projette pas une augmentation du nombre de naissances en 2020.

3. En janvier 1998, on constate qu'il naît légèrement plus de garçons que de filles mais que, plus on avance en âge, et plus la pyramide des femmes s'épaissit par rapport à celle des hommes. Les creux des deux guerres sont aussi perceptibles chez les hommes que chez les femmes.

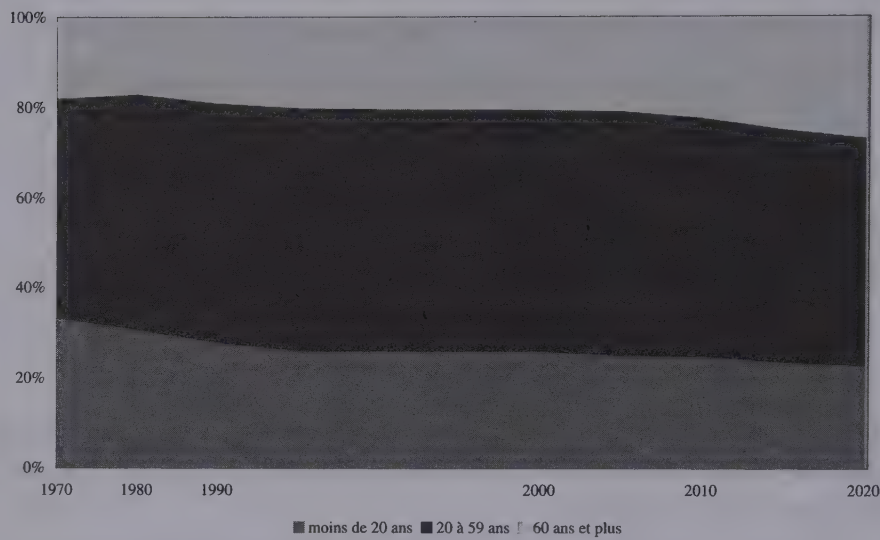
Ces tendances se maintiennent en 2020. On voit, toutefois, d'une façon plus nette l'épaississement de l'histogramme des femmes pour les tranches âgées de la population.

4. Le tableau présentant la projection de la population indique une croissance prévue de la population à l'horizon 2020.

5. Nous voyons sur la pyramide des âges que globalement la population est en croissance du fait des ses tranches vieillissantes. Ce qui est confirmé par les chiffres des tableaux. Nous constatons une proportion croissante des personnes de plus de 60 ans, une réduction de la part de la population jeune (moins de 20 ans). Ce vieillissement s'apprécie également par l'élévation continue de l'âge moyen de la population.

6. Le graphique ci-contre, en bandes superposées, illustre l'analyse numérique.

Tranches d'âges sur la période 1970-2020



16. Distribution de salariés

Mots-clefs

Histogramme, polygone des fréquences

Énoncé

La répartition d'une population de salariés par établissement au sein d'un bassin d'emploi est donnée par le tableau ci-dessous.

1. Construisez l'histogramme.
2. Construisez le polygone des fréquences.

Répartition des salariés par établissement

Classes	Effectifs
[0 ; 20[	925
[20 ; 40[	1 665
[40 ; 50[	2 590
[50 ; 60[	3 145
[60 ; 80[	2 775
[80 ; 110[	2 405
[110 ; 160[	1 850
[160 ; 240[	1 480
[240 ; 360[	1 110
[360 ; 500]	555
	18 500

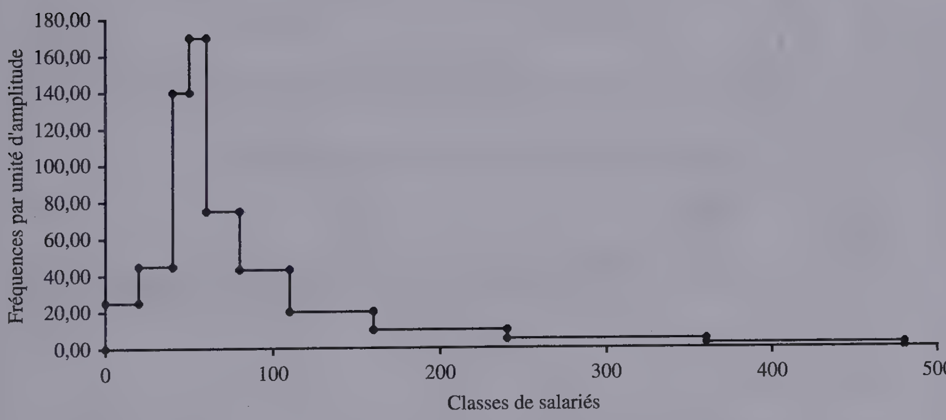
Corrigé

1. Pour construire l'histogramme, nous élaborons le tableau statistique nécessaire.

Calculs pour l'histogramme

Classes	n_i	a_i	c_i	f_i	$\frac{f_i}{a_i} \times 100$
[0 ; 20[	925	20	10	5,0	25,0
[20 ; 40[	1 665	20	30	9,0	45,0
[40 ; 50[	2 590	10	45	14,0	140,0
[50 ; 60[	3 145	10	55	17,0	170,0
[60 ; 80[	2 775	20	70	15,0	75,0
[80 ; 110[	2 405	30	95	13,0	43,3
[110 ; 160[	1 850	50	135	10,0	20,0
[160 ; 240[	1 480	80	200	8,0	10,0
[240 ; 360[	1 110	120	300	6,0	5,0
[360 ; 500]	555	140	430	3,0	2,1
	18 500			100,0	

Histogramme de la distribution

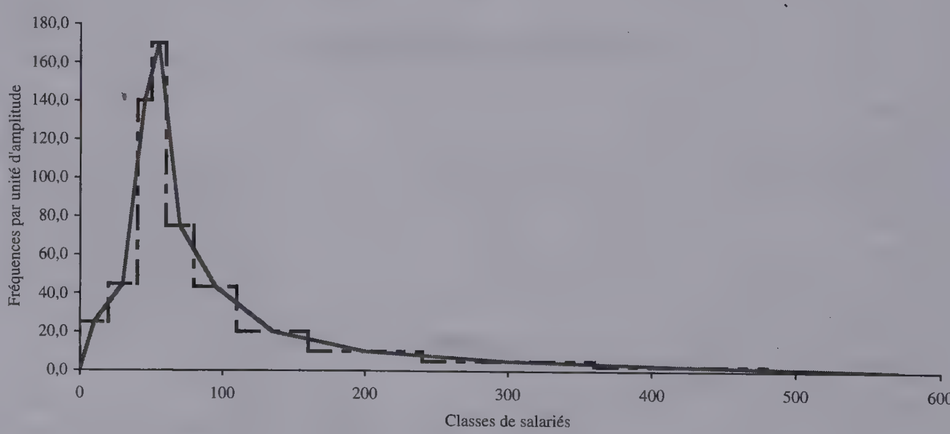


2. Nous construisons le polygone des fréquences en joignant les milieux de chaque sommet des rectangles. Les ordonnées sont les fréquences par unité d'amplitude, les abscisses sont des centres de classe si les classes sont d'amplitude égale. Pour fermer la courbe nous avons pris le point origine et un point qui se trouve à une distance égale à la moitié de l'amplitude de la dernière classe. Un tableau explicite les valeurs prises.

Valeurs pour le polygone des fréquences

c_i	$\dfrac{f_i}{a_i} \times 100$
0	0
10	25,0
30	45,0
45	140,0
55	170,0
70	75,0
95	43,3
135	20,0
200	10,0
300	5,0
430	2,1
570	0

Polygone des fréquences



Le polygone des fréquences donne une vision plus réaliste de la distribution en éliminant les ruptures entre les classes. Il permet également de percevoir la dissymétrie de la distribution.

Ce polygone de fréquences ne respecte pas le principe de la conservation des aires. Il n'est donc pas rigoureusement satisfaisant du point de vue statistique. Néanmoins, compte tenu de la diversité de l'amplitude des classes et de leur étendue, il permet de fournir une première image de la distribution des fréquences.

17. Répartition des subventions agricoles

Mots-clefs

Polygone des fréquences

Énoncé

Nous connaissons la valeur des subventions versées à une population d'agriculteurs.

Répartition des subventions par exploitation

Classes	Effectifs
[10 ; 20[	12
[20 ; 30[	24
[30 ; 40[	36
[40 ; 50[	20
[50 ; 70]	30
	122

1. Construisez l'histogramme.

2. Construisez le polygone des fréquences.

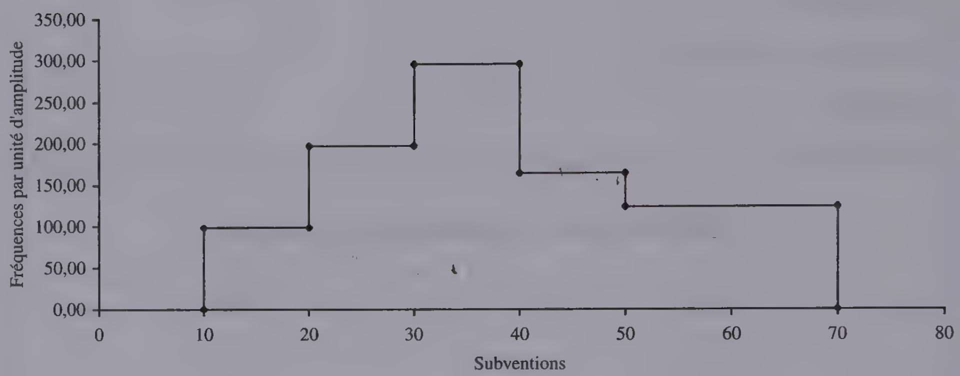
Corrigé

1. Pour construire l'histogramme, nous élaborons le tableau statistique nécessaire.

Calculs pour la construction de l'histogramme

Classes	b_i	b_{i+1}	a_i	c_i	n_i	f_i	$\frac{f_i}{a_i}$ x100
[10 ; 20[	10	20	10	15	12	9,84	98,4
[20 ; 30[	20	30	10	25	24	19,67	196,7
[30 ; 40[	30	40	10	35	36	29,51	295,1
[40 ; 50[	40	50	10	45	20	16,39	163,9
[50 ; 70]	50	70	20	60	30	24,59	123,0
					122	100,00	

Histogramme de la distribution



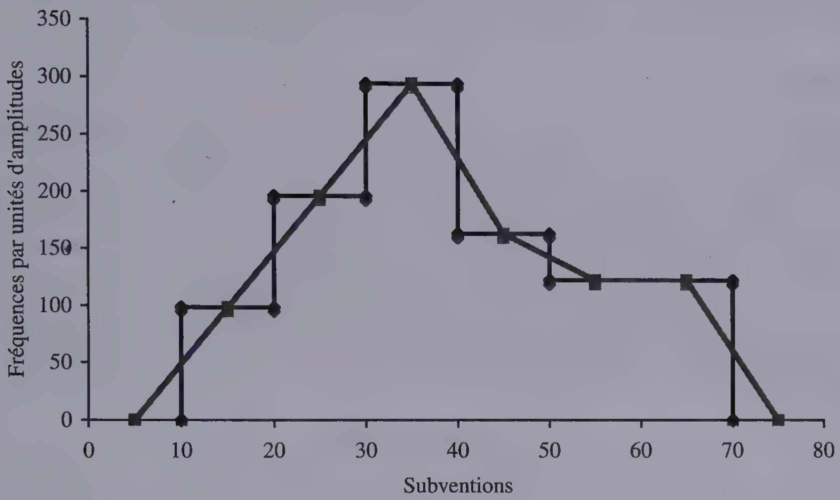
2. Le polygone des fréquences représente la distribution des fréquences. Nous le construisons en joignant les points de chaque milieu des segments de l'histogramme. Les ordonnées sont les fréquences par unité d'amplitude, les abscisses sont des centres de classe si les classes sont d'amplitude égale. Pour fermer la courbe nous avons pris le point origine et un point qui se trouve à une distance égale à la moitié de l'amplitude de la dernière classe. Un tableau explicite les valeurs prises.

Fréquences par unité d'amplitude

c_i	$\frac{f_i}{a_i}$ x100
5	0
15	98,4
25	196,7
35	295,1
45	163,9
55	123,0
65	123,0
75	0

L'unité d'amplitude est de 10. Pour la dernière classe dont l'amplitude est de deux unités, nous n'avons pas retenu le centre de classe, nous avons subdivisé cette classe en classes ayant l'amplitude unitaire, ce qui nous conduit à créer deux classes. Nous considérons qu'entre 65 et 75, la répartition est stable. L'objectif est de faire en sorte que la surface du polygone des fréquences soit identique à celle de l'histogramme.

Polygone des fréquences



Le polygone des fréquences donne une vision plus réaliste de la distribution en éliminant les ruptures entre les classes. Il permet également de percevoir la dissymétrie de la distribution.

18. REPRÉSENTER DES INDICATEURS ENTRE MEMBRES DE L'UE

Mots-clefs

Diagramme polaire

Énoncé

Nous disposons pour quelques pays de l'Union européenne de trois indicateurs importants, ce sont des données relatives au quatrième trimestre 2005.

Les indicateurs

	Croissance annuelle du PIB en volume	Hausse annuelle des prix à la consommation	$\frac{\text{Dette publique}}{\text{PIB}}$
Allemagne	1,6	2,1	66,4
France	1,2	2,0	65,1
Royaume-Uni	1,8	1,9	41,5
Espagne	3,5	4,1	46,9
Italie	0,1	2,2	106,5
Estonie	10,4	4,5	5,5

Réalisez une représentation graphique de ces séries permettant de visualiser le profil de chacun des pays relativement aux trois indicateurs.

Corrigé

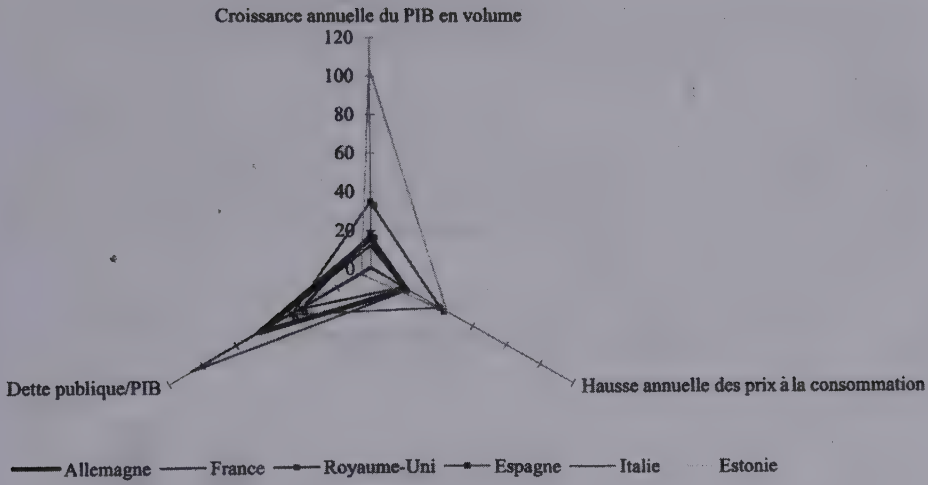
Il apparaît logique de vouloir représenter cette distribution avec un graphique ayant trois axes. Si nous représentons directement les données, nous n'aurons pas un graphique très satisfaisant puisque les valeurs prises par la part de la dette publique dans le PIB sont d'un ordre de grandeur différent des deux autres variables. Pour rendre les valeurs plus comparables nous avons multiplié par 10, pour les besoins du graphique, les valeurs du PIB en volume et des indices de prix à la consommation. Nous obtenons alors le tableau pour la représentation graphique.

Valeurs initiales

	Croissance annuelle du PIB en volume	Hausse annuelle des prix à la consommation	$\frac{\text{Dette publique}}{\text{PIB}}$
Allemagne	1,6	2,1	66,4
France	1,2	2,0	65,1
Royaume-Uni	1,8	1,9	41,5
Espagne	3,5	4,1	46,9
Italie	0,1	2,2	106,5
Estonie	10,4	4,5	5,5

Valeurs utilisées pour la représentation graphique

	Croissance annuelle du PIB en volume	Hausse annuelle des prix à la consommation	$\frac{\text{Dette publique}}{\text{PIB}}$
Allemagne	16	21	66,4
France	12	20	65,1
Royaume-Uni	18	19	41,5
Espagne	35	41	46,9
Italie	1	22	106,5
Estonie	104	45	5,5



Ce tableau montre l'opposition entre les pays ayant une dette publique forte (Italie, France, Allemagne) au pays ayant une dette moyenne (Royaume-Uni, Espagne) à un pays très peu endetté l'Estonie, qui est également un nouvel adhérent de l'UE. L'Estonie se caractérise également par un taux de croissance du pays, PIB en volume, très supérieur aux autres pays de l'échantillon.

19. Répartition d'une population selon les PCS et le genre

Mots-clefs

Double distribution, graphique en barres

Énoncé

Le tableau suivant donne la répartition de la population active par catégorie socioprofessionnelle (C.S.P.) des individus et par sexe pour une population donnée en 1989.

Les catégories socioprofessionnelles en 1989 (en milliers)

Catégories socioprofessionnelles	Hommes	Femmes
Agriculteurs exploitants	799	469
Artisans commerçants et chefs d'entreprises	1 160	597
Cadres et professions intellectuelles supérieures	1 644	671
Professions intermédiaires	2 618	1 975
Employés	1 625	5 146
Ouvriers	5 626	1 495

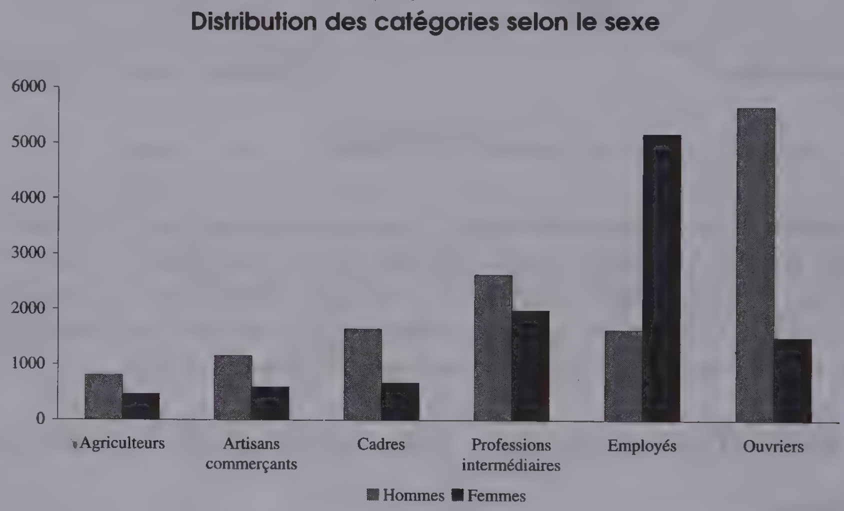
Source : INSEE

1. Représentez graphiquement cette double distribution.
2. Représentez les distributions conditionnelles selon diverses possibilités.

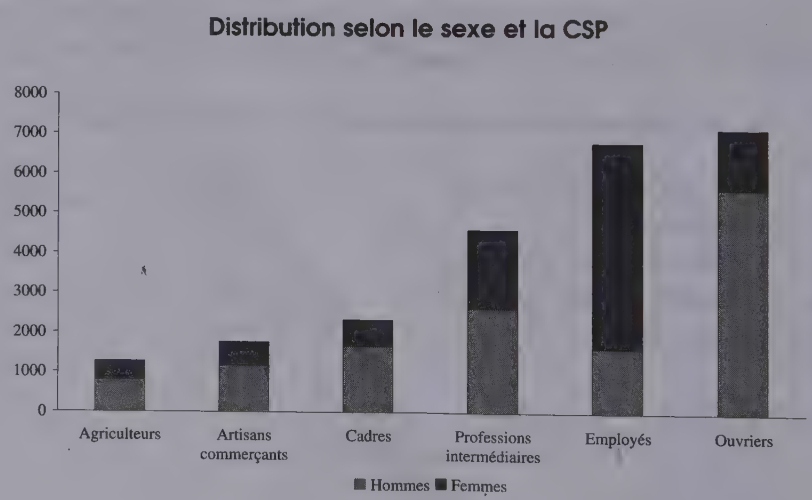
Corrigé

La représentation en tuyaux d'orgue est dans le cas de variables qualitatives la meilleure représentation graphique. Plusieurs solutions sont possibles.

1. Nous représentons côte à côte les deux distributions en tenant compte des effectifs de chaque catégorie.



2. Dans la deuxième représentation, nous utilisons des cartouches superposées rendant mieux compte de l'importance relative des actifs masculins et féminins, en gardant une représentation des valeurs absolues.



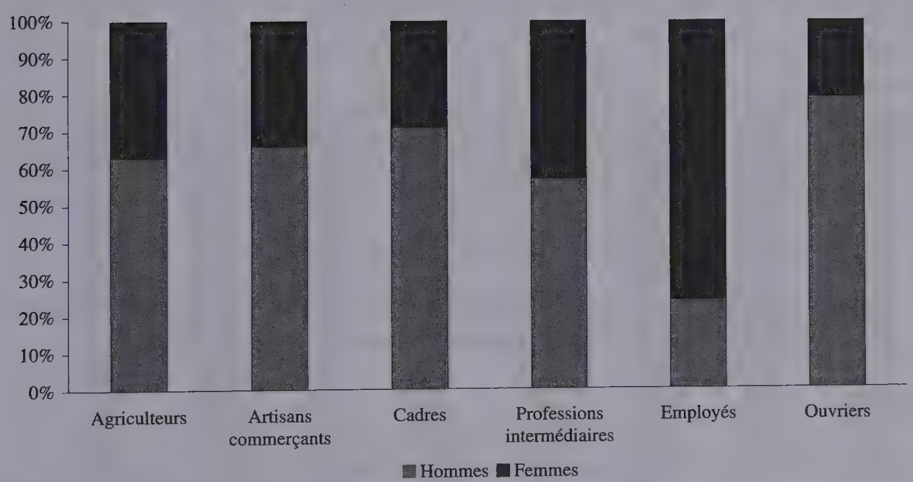
Par rapport au graphique précédent, nous constatons, entre autres, qu'il y a plus d'ouvriers que d'employés en 1989 sans distinction de genre.

3. Nous utilisons un graphique superposé pour représenter les distributions conditionnelles suivant la catégorie sociale et le genre dans un diagramme à cumul interne.

Tableau des fréquences relatives

Catégories	Hommes	f_{ih}	Femmes	f_{if}	Ensemble	f_i
Agriculteurs	799	0,63	469	0,37	1 268	1,00
Artisans	1 160	0,66	597	0,34	1 757	1,00
Cadres	1 644	0,71	671	0,29	2 315	1,00
Professions intermédiaires	2 618	0,57	1 975	0,43	4 593	1,00
Employés	1 625	0,24	5 146	0,76	6 771	1,00
Ouvriers	5 626	0,79	1 495	0,21	7 121	1,00

Part des femmes dans chacune des CSP



Cette représentation rend sensible le fait que la très grande majorité des employés sont des employées, alors que dans les autres catégories les hommes sont majoritaires.

20. RÉPARTITION DES ENTREPRISES PAR SECTEUR D'ACTIVITÉ ET NOMBRE DE SALARIÉS

Mots-clefs

Représentation graphique en double barres

Énoncé

Répartition des entreprises par nombre de salariés et activités au 1/1/1998 (en milliers)

	Petites entreprises		Moyennes entreprises				Grandes entreprises	Total
	0	1 à 9	10 à 49	50 à 199	200 à 499	Total	⩾ 500	
IAA	17,47	44,59	5,63	1,09	0,24	6,96	0,12	69,13
Industrie hors IAA	65,95	81,71	29,55	6,97	1,55	38,07	0,82	186,55
Construction	126,02	149,77	18,61	1,77	0,25	20,62	0,11	296,52
Commerce	275,31	293,14	34,86	4,44	0,61	39,92	0,28	608,64
Transports	47,10	28,55	7,52	1,51	0,29	9,32	0,10	85,07
Hôtellerie	78,65	109,38	8,05	0,59	0,07	8,72	0,06	196,81
Services aux entreprises	214,73	145,39	19,34	2,78	0,62	22,75	0,33	383,20
Services aux ménages	288,86	141,93	7,61	1,19	0,19	8,98	0,04	439,81
Total	1114,08	994,45	131,18	20,33	3,83	155,34	1,86	2265,72

1. Quelles sont les variables étudiées ? Quel est leur type ?
2. Proposez une représentation graphique permettant de visualiser la répartition des activités au sein de chaque groupe de moyennes entreprises.

Corrigé

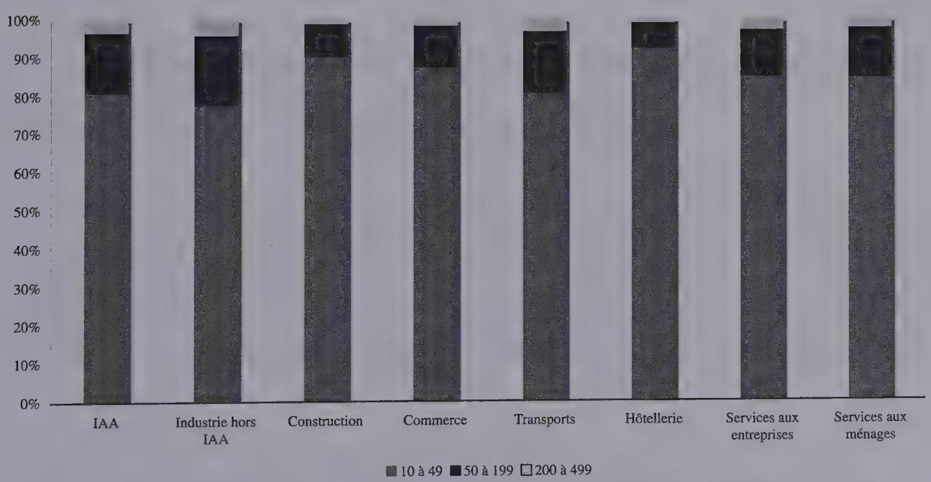
1. Nous avons un tableau croisant une variable qualitative (une nomenclature des activités) et une variable quantitative continue (le nombre de salariés). Cette variable est continue car, en équivalent temps plein, toutes les valeurs de la variable peuvent être atteintes.

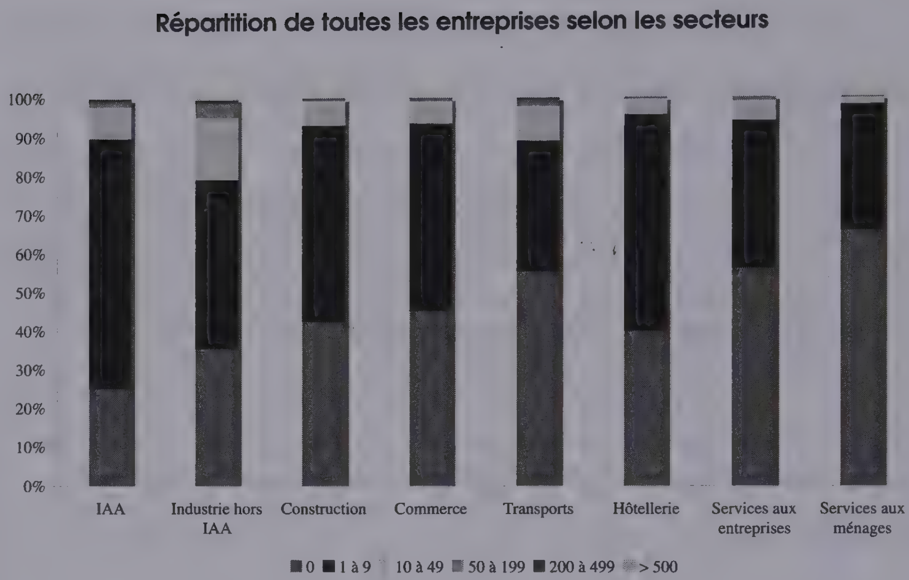
2.

Tableau de la répartition des entreprises pour construire le graphique (en %)

	Petites entreprises		Moyennes entreprises			Grandes entreprises	Total
	0	1 à 9	10 à 49	50 à 199	200 à 499	⩾ 500	
IAA	25,3	65,4	8,1	1,6	0,3	0,2	100,0
Industrie hors IAA	35,4	43,8	15,8	3,7	0,8	0,4	100,0
Construction	42,5	50,5	6,3	0,6	0,1	0,0	100,0
Commerce	45,2	48,2	5,7	0,7	0,1	0,0	100,0
Transports	55,4	33,6	8,8	1,8	0,3	0,1	100,0
Hôtellerie	40,0	55,6	4,1	0,3	0,0	0,0	100,0
Services aux entreprises	56,0	37,9	5,0	0,7	0,2	0,1	100,0
Services aux ménages	65,7	32,3	1,7	0,3	0,0	0,0	100,0
Total	49,2	43,9	5,8	0,9	0,2	0,1	100,0

Répartition des moyennes entreprises par taille en fonction du secteur d'activités





Il est clair que les très grandes entreprises n'apparaissent pas sur le graphique.

21. GRAPHIQUE DES RECETTES ET DES DÉPENSES

Mots-clefs

Double graphique

Énoncé

Soit les éléments de l'exécution du budget d'une université grenobloise.

Taux d'exécution budgétaire

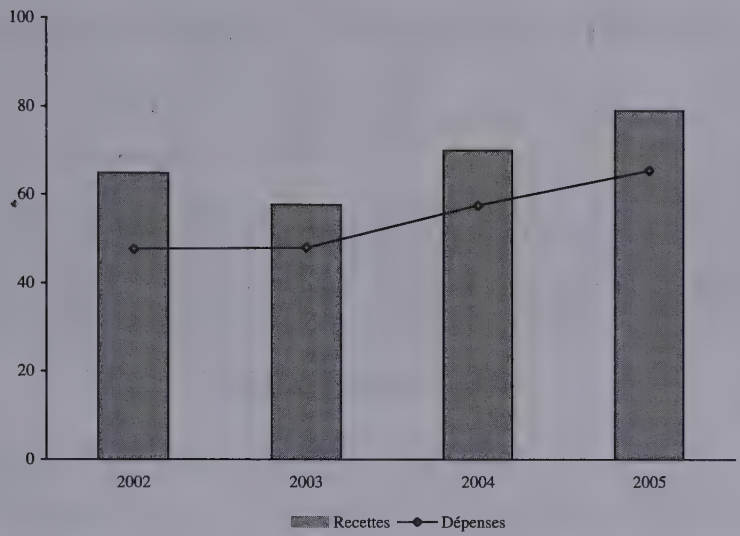
	2002	2003	2004	2005
Recettes	65,0	58,0	70,5	79,6
Dépenses	47,7	48,2	57,9	66,0

Donnez une représentation graphique de ces deux distributions.

Corrigé

Pour représenter cette double distribution nous utilisons deux formes de représentation : les barres pour les recettes, et une courbe pour les dépenses.

Les taux d'exécution budgétaires



22. SOUTENANCE DE THÈSES

Mots-clefs

Pourcentage, multiplicateur

Énoncé

L'évolution des soutenances de thèse dans une université de Grenoble a évolué comme ci-dessous.

	1999	2000	2001	2002
Effectifs	425	448	431	406
% de femmes	34,5	31,5	30	25,5
% de post-doctorants	20	36,5	38,5	36

Source : publication de l'UJF

1. Calculez le taux d'évolution du nombre de thèses entre 1999 et 2002.
2. Le pourcentage des femmes décroît-il plus rapidement que celui des effectifs ?
3. Peut-on constater une réduction du nombre de thèses soutenues par les hommes ?

Corrigé

1. Nous constatons sur la période une réduction du nombre de thèses, pour quantifier cette diminution, nous calculons les multiplicateurs selon la formule usuelle :

$$(1 + r_1) = \frac{X_1}{X_0}$$

Par exemple entre 2000 et 1999 le multiplicateur est : $1 + r = \frac{448}{425} = 1{,}054$.

L'habitude est de multiplier le résultat par 100 pour faciliter la lecture. Dans les calculs, il est nécessaire de ne pas oublier cette convention.

Multiplicateurs des thèses

	2000/1999	2001/2000	2002/2001
Multiplicateurs annuels	105,4	96,2	94,2

Entre 1999 et 2001 le multiplicateur est de :

$$(1 + r) = 1{,}054 \cdot 0{,}962 \cdot 0{,}942 \cong 0{,}955$$

Le multiplicateur entre 1999 et 2002 est de 95,5. Le taux de réduction sur la période est de 4,5 %.

Nous pouvons également le calculer directement $(1 + r) = \frac{406}{425} = 0{,}955$.

2. Pour obtenir le taux de réduction du nombre de thèses soutenues par les femmes, nous calculons le nombre de thèses soutenues par les femmes puis le multiplicateur entre 1999 et 2002. Il est aussi possible d'utiliser les calculs précédents. Nous allons présenter les deux calculs.

Calcul du nombre de thèses soutenues par des femmes :

Nombre de thèses × pourcentage de thèse soutenues par des femmes

Par exemple pour 2000 : $448 \times 0{,}315 = 141{,}12$ soit 141 thèses.

Évolution du nombre de thèses soutenues par des femmes

Années	1999	2000	2001	2002	Produit des multiplicateurs annuels
Nombre de thèses soutenues	147	141	129	104	
Multiplicateurs		96,24	91,62	80,1	70,60

Soit $(1 + r) = 0{,}9624 \cdot 0{,}9162 \cdot 0{,}801 \cong 0{,}706$,

Le multiplicateur est de 70,6.

Le multiplicateur global est de 95,5 ; pour les femmes, il est de 70,6. Le taux de réduction de thèses « féminines » atteint 29,6 %.

Nous pouvons obtenir le même résultat sans avoir à calculer le nombre de thèses soutenues par les femmes.

Soit n_i le nombre total de thèses soutenues l'année i, n_i' le nombre de thèses soutenues par des femmes et f_i la proportion de ces thèses, $n_i' = f_i \cdot n_i$.

Le multiplicateur pour les thèses « féminines » entre $i - 1$ et i est :

$$\frac{n_i'}{n_{i-1}'} = \frac{f_i \cdot n_i}{f_{i-1} \cdot n_{i-1}} = \frac{f_i}{f_{i-1}} \cdot (1 + r)_i$$

Exemple numérique entre 2001 et 2002, le multiplicateur global est 94,2 ; le multiplicateur pour les femmes est de $94{,}2 \cdot \dfrac{25{,}5}{30} = 80{,}07$.

Nous pouvons également le calculer directement $(1 + r)_f = \dfrac{406}{425} \cdot \dfrac{25{,}5}{34{,}5}$ $= 0{,}706$. Le multiplicateur est de 70,6.

3. Nous calculons directement le multiplicateur des thèses « masculines ».

$$(1 + r)_h = \frac{406}{425} \cdot \frac{(100 - 25{,}5)}{(100 - 34{,}5)} = 1{,}0866$$

Le multiplicateur est de 108,7.

Le nombre de thèses soutenues par des hommes est plus important en 2002 qu'en 1999, il a augmenté de plus de 8 %. La réduction du nombre de thèses soutenues s'explique donc par la seule réduction du nombre des thèses « féminines ».

La diminution de 4,5 % du nombre de thèses soutenues s'explique par une réduction de 29,4 % des thèses « féminines » et une augmentation de 8,7 % des thèses « masculines ».

23. TAUX DE SURVIE DES ENTREPRISES

Mots-clefs

Pourcentage

Énoncé

Les créations d'entreprises en 1995

Ensemble	295 416
Soit en pourcentages, suivant la taille	
0 salarié	74,5
1 ou 2 salariés	17,1
3 à 5 salariés	5,5
6 à 9 salariés	1,5
10 salariés et plus	1,4

Taux de survie à trois et cinq ans des entreprises créées en 1995 (en %)

	Survie à trois ans	Survie à cinq ans
0 salarié	67	51
1 ou 2 salariés	75	61
3 à 5 salariés	83	70
6 à 9 salariés	85	74
10 salariés et plus	87	77

1. Définissez le taux de survie.
2. Calculez le taux de survie à trois ans des entreprises de moins de 5 salariés.
3. Déterminez le taux de survie à cinq ans des entreprises de plus de 6 salariés.

Corrigé

1. La durée de vie des entreprises est obtenue par la différence du nombre d'entreprises entre la date de cessation économique et celle de la création. Le taux de survie à 1, 3, 5 ou 7 ans est donc égal au nombre des entreprises de la génération N encore actives lors de leur 12^e, 36^e, 60^e ou 84^e mois, divisé par le nombre d'entreprises créées l'année N.
2. Pour calculer le taux de survie à trois ans des entreprises de moins de 5 salariés nous devons construire les tableaux statistiques suivants.

Les créations d'entreprises en 1995

Classe d'entreprises	Effectifs n_i	Fréquences f_i	Pourcentages p_i	Fréquences cumulées F_i
0 salarié	220 085	0,745	74,5	0,745
1 ou 2 salariés	50 516	0,171	17,1	0,916
3 à 5 salariés	16 248	0,055	5,5	0,971
6 à 9 salariés	4 431	0,015	1,5	0,986
10 salariés et plus	4 136	0,014	1,4	1,000
	295 416	1,000	100,0	

La survie à trois ans

Classe d'entreprises	Effectifs n_i	Fréquences f_i	Pourcentages p_i	Fréquences cumulées F_i
0 salarié	147 457	0,718	71,8	0,718
1 ou 2 salariés	37 887	0,184	18,4	0,903
3 à 5 salariés	12 656	0,062	6,2	0,964
6 à 9 salariés	3 766	0,018	1,8	0,982
10 salariés et plus	3 598	0,018	1,8	1,000
	205 364	1,000	100,0	

La survie à cinq ans

Classe d'entreprises	Effectifs n_i	Fréquences f_i	Pourcentages p_i	Fréquences cumulées F_i
0 salarié	112 243	0,697	69,7	0,698
1 ou 2 salariés	30 815	0,192	19,2	0,889
3 à 5 salariés	11 374	0,071	7,1	0,960
6 à 9 salariés	3 279	0,020	2,0	0,980
10 salariés et plus	3 185	0,020	2,0	1,000
	160 896	1,000	100,0	

Le taux de survie à trois ans des entreprises d'au moins de 5 salariés est le rapport entre le nombre d'entreprise de 5 salariés au moins existant trois ans après leur création et le nombre d'entreprise créées en 1995.

Les entreprises existant après trois ans sont de 147 457 avec 0 salariés, 37 887 avec 1 ou 2 salariés, 12 656 de 3 à 5 salariés soit 198 000. Les entreprises créées étaient pour les mêmes catégories 220 805 avec 0 salariés, 50 516 avec 1 ou 2 salariés, 16 248 de 3 à 5 salariés soit 286 849.

Le taux de survie à trois ans est donc $\dfrac{198\,000}{220\,805} \cdot 100 = 69{,}03\ \%$.

3. À partir du tableau de la survie des entreprises à cinq ans et celui des créations d'entreprise, nous pouvons calculer le taux de survie à cinq ans des entreprises de plus de 6 salariés comme le rapport du nombre d'entreprises de plus de 6 salariés existant cinq ans après leur création et le nombre d'entreprises de plus de 6 salariés créées en 1995.

Le nombre d'entreprises de six salariés cinq ans après leur création est égal au nombre total des entreprises multiplié par le pourcentage des entreprises de plus de 6 salariés : $160\,896 \cdot (1 - 0{,}96)$.

Le nombre d'entreprises de plus de six salariés créées en 1995 : $295\,416 \cdot (1 - 0{,}971)$.

Le taux cherché est donc $\dfrac{160\,896 \cdot (1 - 0{,}96)}{295\,416 \cdot (1 - 0{,}971)} \cdot 100 = 75{,}12\ \%$

Chapitre 2

Les distributions à une dimension

Les valeurs de tendance centrale synthétisent une distribution statistique par un seul chiffre.

Un seul nombre pour résumer toute une distribution est tout à fait insuffisant. Il est nécessaire de disposer d'informations sur la dispersion des valeurs autour de la caractéristique retenue. Les caractéristiques de dispersion sont absolues ou relatives. La dissymétrie est évaluée par divers indicateurs. Les coefficients d'asymétrie mesurent la répartition des valeurs de part et d'autre d'une valeur centrale. Les mesures de la concentration sont multiples, les plus utilisées sont l'indice C_x, l'indice d'Herfindahl et l'indice de Gini associé à la courbe de Lorenz.

24. Revenus mensuels

Mots-clefs

Variable continue, histogramme

Énoncé

Répartition des salariés par tranches de revenus mensuels pour une région

Revenus en euros	
de 50 à moins de 150	10 000
de 150 à moins de 200	22 000
de 200 à moins de 250	47 000
de 250 à moins de 300	29 000
de 300 à moins de 400	54 500
de 400 à moins de 600	36 000
de 600 à moins de 1000	8 000
de 1000 à moins de 2500	500

Certaines tranches de revenu peuvent paraître très faibles, il s'agit de salariés à temps partiel.

1. Construisez l'histogramme de cette distribution.

2. Déterminez-en la classe modale et le mode.

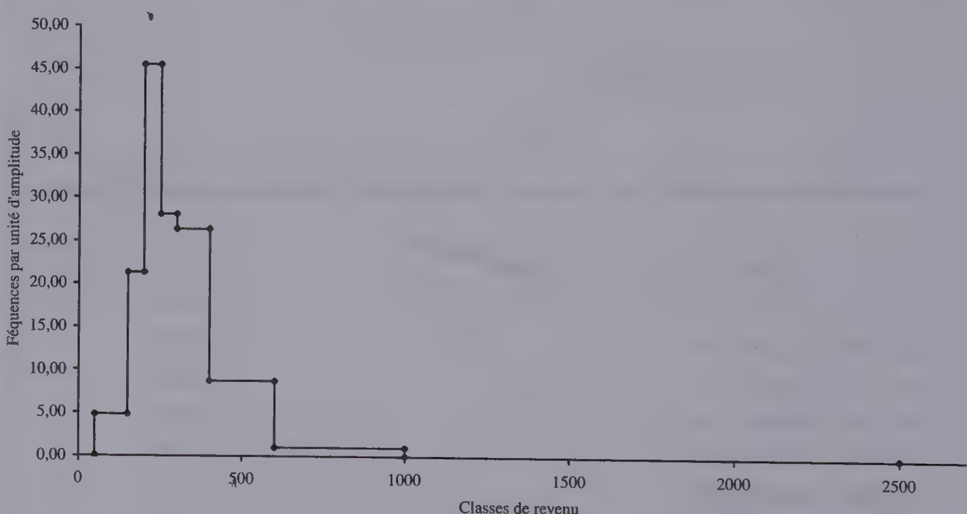
Corrigé

1. Cela nécessite de construire le tableau statistiques ci-dessous.

Calculs pour la construction de l'histogramme et la détermination du mode

	a_i	c_i	n_i	f_i	$\frac{f_i}{a_i}$ x100
[50 ; 150[	100	100	10000	4,8	4,8
[150 ; 200[	50	175	22000	10,6	21,3
[200 ; 250[	50	225	47000	22,7	45,4
[250 ; 300[	50	275	29000	14,0	28,0
[300 ; 400[	100	350	54500	26,3	26,3
[400 ; 600[	200	500	36000	17,4	8,7
[600 ; 1000[	400	800	8000	3,9	1,0
[1000 ; 2500]	1500	1750	500	0,2	0,0
			207000	100,0	

Histogramme de la distribution



2. La classe modale de la distribution est [200 ; 250[, c'est la classe pour laquelle la fréquence par unité d'amplitude $\left(\frac{f_i}{a_i}\right)$ est la plus élevée.

Le mode est calculé par la formule suivante :

$$M_o = b_i + \frac{\Delta_i}{\Delta_i + \Delta_{i+1}} \cdot a_i$$

Avec M_o le mode, b_i limite inférieure de la classe modale (i), a_i l'amplitude de la classe modale, $\Delta_i = \frac{f_i}{a_i} - \frac{f_{i-1}}{a_{i-1}}$ différence entre la fréquence de la classe modale et la fréquence de la classe précédente dans la distribution, $\Delta_i = \frac{f_i}{a_i} - \frac{f_{i+1}}{a_{i+1}}$ différence entre la fréquence de la classe modale et la fréquence de la classe suivante dans la distribution.

Le mode est donc :

$$M_o = b_i + \frac{\Delta_i}{\Delta_i + \Delta_{i+1}} \cdot a_i = 200 + \frac{45{,}41 - 21{,}26}{(45{,}41 - 21{,}26) + (45{,}41 - 28{,}02)} \cdot 50 \cong 229\,€$$

25. PCS

Mots-clefs

Caractère qualitatif, diagramme en barres

Énoncé

L'enquête « Emploi » de l'INSEE donne une répartition selon les PCS de la population active occupée.

Actifs occupés en 2004

Catégorie	Effectifs (en milliers)
Agriculteurs exploitants	671
Artisans, commerçants, chefs d'entreprise	1 464
Cadres et professions intellectuelles supérieures	3 556
Professions intermédiaires	5 771
Employés	7 114
Ouvriers	6 127
Ensemble	24 713

Source : INSEE

1. Définissez chacune des catégories des PCS.
2. Représentez graphiquement cette distribution.
3. Déterminez le mode de cette distribution.

Corrigé

1. Les « Agriculteurs exploitants » sont des indépendants dont l'activité est agricole, les salariés agricoles sont classés dans les ouvriers.

La catégorie « Artisans, commerçants, chefs d'entreprise » regroupe les indépendants non agricoles à l'exception des professions libérales. Elle inclut donc les artisans (chefs d'entreprise produisant des biens avec moins de 10 salariés), les commerçants (chefs d'entreprise dont l'activité principale est la vente et dont le nombre de salariés est inférieur à 10), les « chefs d'entreprise » de 10 salariés et plus, quelle que soit leur activité (non agricole).

Les « Cadres » sont des salariés auxquels les conventions collectives reconnaissent une fonction d'encadrement, de direction d'autres personnes (d'où des avantages salariaux mais aussi des obligations horaires plus importantes que les autres salariés). Les « professions intellectuelles supérieures » sont inclues dans cette catégorie. Ce sont les professions libérales, les professeurs, les professions scientifiques, les professions de l'information, des arts et du spectacle, que les personnes soient salariées ou indépendantes.

Les « Professions intermédiaires » sont entre les cadres et le personnel d'exécution. La frontière ouvrier qualifié – technicien n'est pas toujours évidente notamment parce que l'appellation « technicien » (classé dans les professions intermédiaires) est plus valorisante qu' « ouvrier qualifié » (classé dans les ouvriers).

Les « Employés » sont des personnels d'exécution « non manuel », travaillant dans un bureau ou un commerce (quelle que soit l'activité de l'entreprise agricole, industrielle ou de service) dont l'essentiel de l'activité réside dans le traitement de l'information.

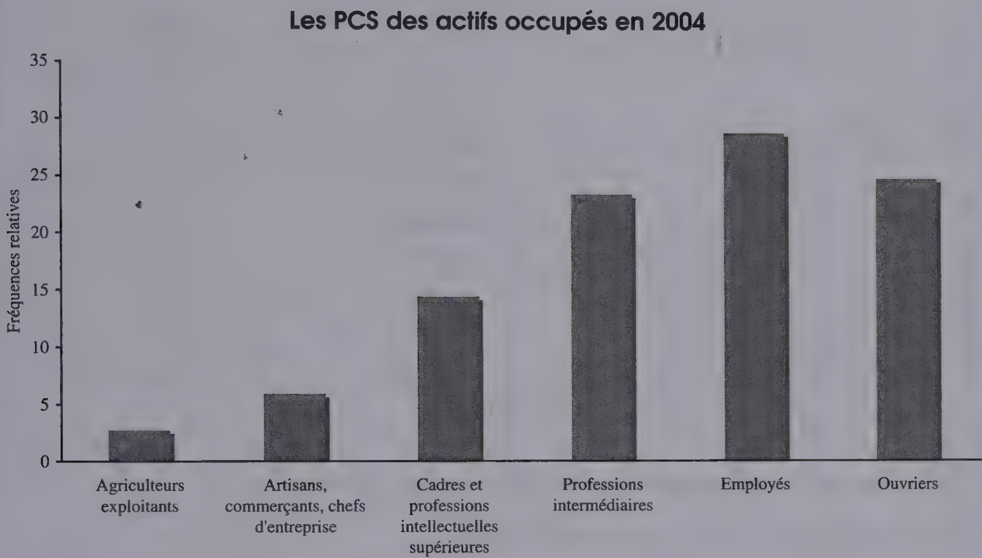
Les « Ouvriers » sont des salariés d'exécution. Ils peuvent travailler dans l'industrie mais aussi dans l'artisanat ou dans l'agriculture. Ils produisent ou transforment souvent un bien matériel.

Le tableau facilite la construction du graphique.

Structure des PCS des actifs occupés

Catégories	f_i (en %)
Agriculteurs exploitants	2,7
Artisans, commerçants, chefs d'entreprise	5,9
Cadres et professions intellectuelles supérieures	14,4
Professions intermédiaires	23,4
Employés	28,8
Ouvriers	24,8
Ensemble	100,0

2. Puisque que nous recherchons le mode d'un caractère qualitatif, la représentation graphique en colonnes est une des plus adaptée.



3. La variable étant qualitative, le mode est facile à déterminer. Le graphique permet de déterminer le mode, c'est la catégorie dont la colonne est la plus élevée, c'est-à-dire la catégorie « employés ». Le tableau statistique nous indiquait que cette catégorie était la plus fréquente (28,8 %).

26. NOMBRES DE PERSONNES PAR MÉNAGE

Mots-clefs

Variable quantitative discrète

Énoncé

Les enquêtes démographiques de l'INSEE permettent de disposer de la distribution des ménages selon le nombre de personnes.

Ménages selon le nombre de personnes du ménage en 1999

Nombre de personnes du ménage	1999	
	Nombre de ménages	%
Ensemble	23 808 072	100
1	7 380 109	31,0
2	7 414 180	31,1
3	3 848 586	16,2
4	3 276 615	13,8
5	1 322 024	5,6
6	363 422	1,5
7	121 569	0,5
8	46 921	0,2
9 ou plus	34 646	0,1

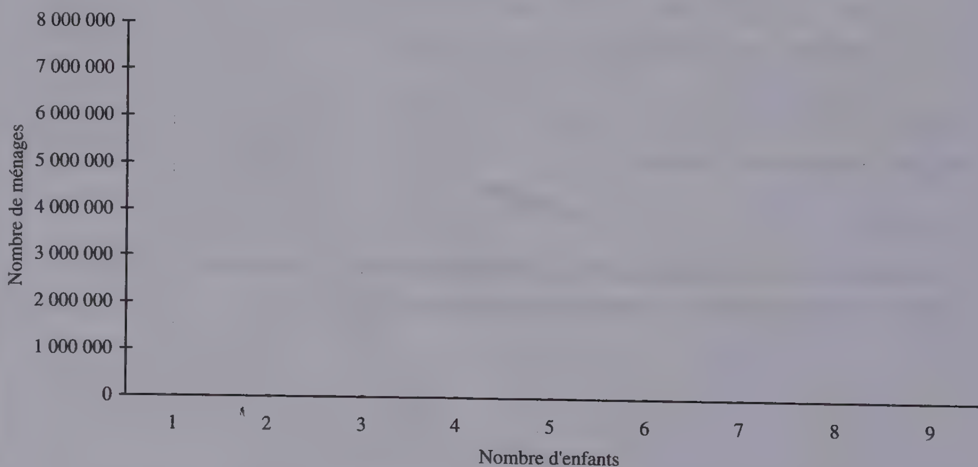
Source INSEE

1. Donnez une représentation graphique de cette distribution.
2. Déterminez le mode de cette distribution.

Corrigé

1. La représentation graphique usuelle pour une variable discrète est une représentation en bâton.

Distribution des personnes par ménage



2. Le mode est la valeur la plus fréquente, autrement dit dont la fréquence relative est la plus élevée. C'est aussi la valeur de la variable qui, sur le graphique, est la plus haute. Ici le mode est deux personnes.

27. Chiffre d'affaires

Mots-clefs

Courbe cumulative, médiane, médiale

Énoncé

Le tableau suivant donne la distribution d'entreprises d'un secteur quelconque en fonction de leur chiffre d'affaires :

Répartition des C.A.

C.A. (en millions d'euros)	Effectifs
[10 ; 20[	80
[20 ; 40[	240
[40 ; 60[	320
[60 ; 70[	200
[70 ; 120[	500
[120 ; 180[	280
[180 ; 300]	380
Total	2000

1. Représentez la courbe cumulative des fréquences de cette distribution.
2. Déterminez graphiquement le C.A. de l'entreprise médiane.
3. Calculez le C.A. de l'entreprise médiane.
4. Calculez le C.A. de l'entreprise médiale.

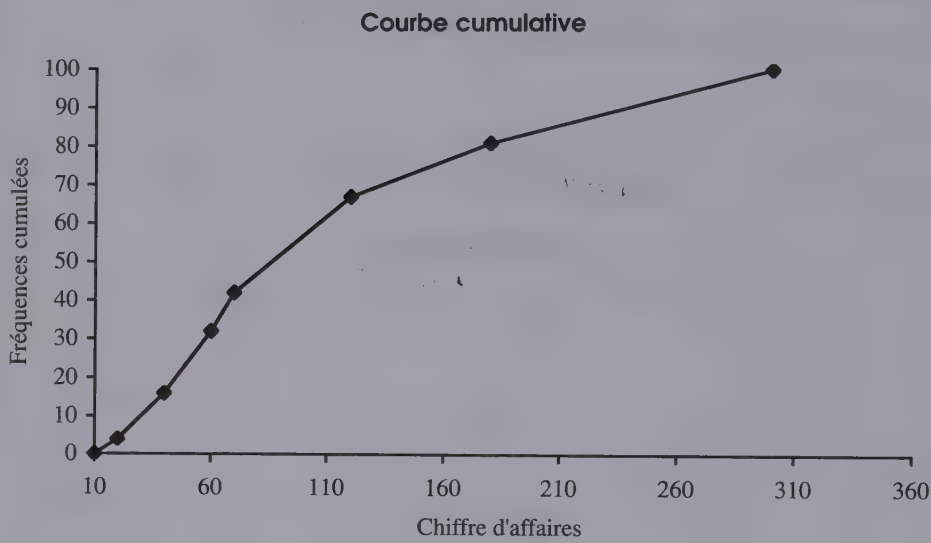
Corrigé

Pour répondre aux questions posées, nous avons besoin de construire le tableau statistique suivant.

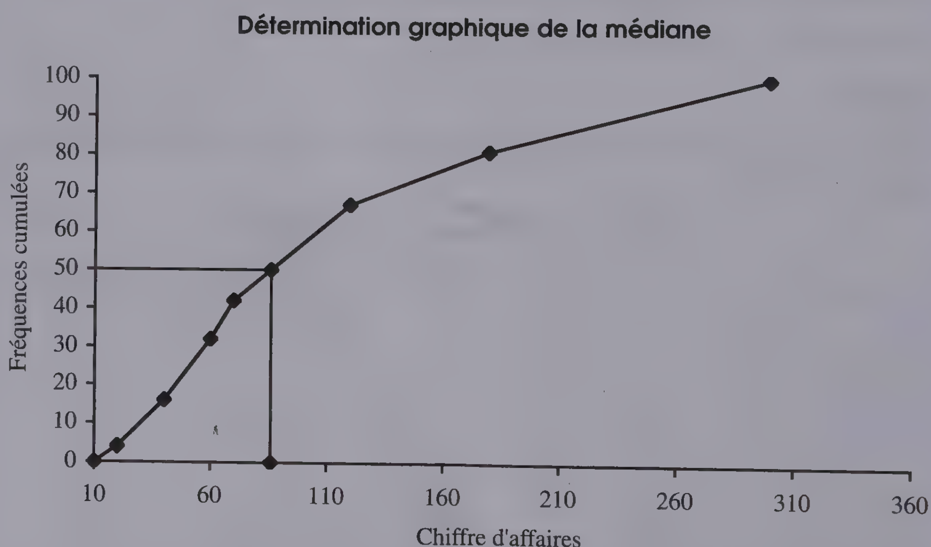
Tableau pour le calcul de la médiane

Classe de C.A.	a_i	n_i	f_i (en %)	F_i (en %)
[10 ; 20[	10	80	4,0	4,0
[20 ; 40[	20	240	12,0	16,0
[40 ; 60[	20	320	16,0	32,0
[60 ; 70[	10	200	10,0	42,0
[70 ; 120[	50	500	25,0	67,0
[120 ; 180[	60	280	14,0	81,0
[180 ; 300]	120	380	19,0	100,0
Total		2000	100,0	

1. La courbe cumulative des fréquences correspond au graphe des fréquences cumulées.



2. La médiane est la valeur de la variable, ici le chiffre d'affaires, qui divise en deux l'effectif de la distribution. La médiane est la valeur M_e telle que $F(M_e) = 0{,}5$ ou 50 %. Pour déterminer graphiquement la valeur de la médiane, il suffit de tracer une parallèle à l'axe des abscisses pour une ordonnée de 50 %.



La valeur de la médiane est proche de 90 millions d'euros.

3. Nous pouvons calculer la médiane par interpolation linéaire. La formule générale du calcul de la médiane est

$$M_e = b_i + a_i \cdot \frac{50 - F_i}{F_{i+1} - F_i}$$

$$M_e = 70 + 50 \cdot \frac{50 - 42}{67 - 42} = 86$$

Le chiffre d'affaires médian de la distribution est 86 millions d'euros.

4. La médiale se calcule à partir du tableau statistique suivant.

Tableau pour le calcul de la médiale

Classe de CA	a_i	c_i	n_i	f_i	$f_i c_i$	f'_i	F'_i
[10 ; 20[	10	15,0	80	4,0	60,0	0,6	0,6
[20 ; 40[	20	30,0	240	12,0	360,0	3,3	3,9
[40 ; 60[	20	50,0	320	16,0	800,0	7,3	11,2
[60 ; 70[	10	65,0	200	10,0	650,0	6,0	17,1
[70 ; 120[	50	95,0	500	25,0	2375,0	21,8	38,9
[120 ; 180[	60	150,0	280	14,0	2100,0	19,3	58,2
[180 ; 300]	120	240,0	380	19,0	4560,0	41,8	100,0
Total			2000	100,0	10905,0	100,0	

La médiale se calcule par interpolation linéaire :

$$M_l = b_i + a_i \cdot \frac{50 - F'_i}{F'_{i+1} - F'_i}$$

$$M_l = 120 + 60 \cdot \frac{50 - 38{,}9}{58{,}2 - 38{,}9} = 154{,}5$$

La masse des CA des entreprises inférieure à 154,5 est égale à la masse des CA des entreprises au-dessus de ce chiffre.

28. SURFACES AGRICOLES UTILES

Mots-clefs

Histogramme, médiane, médiale

Énoncé

Le tableau ci-dessous donne pour une région quelconque, la distribution des exploitations agricoles en fonction de la surface agricole utile (S.A.U.).

Distribution des exploitations

Classes de SAU (en ha)	moins de 5	de 5 à 10	de 10 à 20	de 20 à 50	de 50 à 100	plus de 100
Nombre d'exploitations	7 800	9 600	13 200	20 400	7 200	1 800

1. Construisez l'histogramme de cette série.
2. Calculez la surface médiane.
3. Calculez la surface médiale.

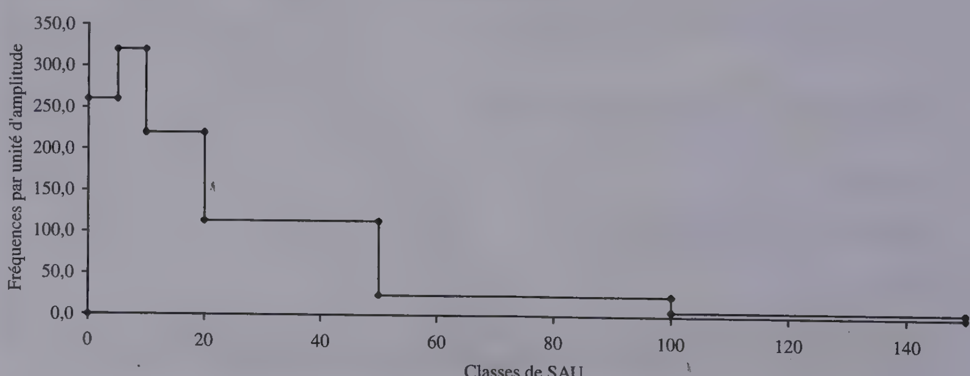
Corrigé

1. Pour construire l'histogramme, nous devons disposer des fréquences par unité d'amplitude.

Tableau pour la construction de l'histogramme

Classes de SAU (ha)	a_i	n_i	f_i (en %)	$\frac{f_i}{a_i}$ x100
[0 ; 5[	5	7800	13,0	260,00
[5 ; 10[	5	9600	16,0	320,00
[10 ; 20[	10	13200	22,0	220,00
[20 ; 50[	30	20400	34,0	113,33
[50 ; 100[	50	7200	12,0	24,00
[100 ; 150]	50	1800	3,0	6,00
Total		60000	100,0	

Histogramme



2. Pour la détermination de la médiane, un nouveau tableau est nécessaire.

Tableau des calculs pour la médiane

Classes de SAU (ha)	b_i	b_{i+1}	a_i	f_i	F_i
[0 ; 5[	0	5	5	0,13	0,13
[5 ; 10[	5	10	5	0,16	0,29
[10 ; 20[	10	20	10	0,22	0,51
[20 ; 50[	20	50	30	0,34	0,85
[50 ; 100[	50	100	50	0,12	0,97
[100 ; 150]	100	150	50	0,03	1,00
Total				1,00	

La classe médiane est la classe de [10 ; 20[ha qui comprend la valeur $F_i = 0{,}5$.

$$M_e = b_i + a_i \cdot \frac{50 - F_i}{F_{i+1} - F_i}$$

$$M_e = 10 + 10 \cdot \frac{0{,}50 - 0{,}29}{0{,}22} = 19{,}55 \text{ ha}$$

Nous avons autant d'exploitations qui ont une surface de plus de 19,55 ha que d'exploitations ayant une surface inférieure.

3. La médiale de la série est la médiane de la variable ; nous devons calculer les fréquences relatives f_i' de la variable.

Tableau pour le calcul de la médiale

Classes de SAU (ha)	a_i	c_i	n_i	f_i	f_ic_i	f_i'	F_i'
[0 ; 5[	5	2,5	7800	13,0	32,5	1,0	1,0
[5 ; 10[	5	7,5	9600	16,0	120,0	3,0	4,0
[10 ; 20[	10	15,0	13200	22,0	330,0	12,0	16,0
[20 ; 50[	30	35,0	20400	34,0	1190,0	39,0	55,0
[50 ; 100[	50	75,0	7200	12,0	900,0	29,0	84,0
[100 ; 150]	50	125,0	1800	3,0	375,0	16,0	100,0
Total			60000	100,0	2947,5	100,0	

Avec

$$f_i' = \frac{f_ic_i}{\sum_{i=1}^{k} f_ic_i}$$

nous pouvons calculer la médiale selon la formule habituelle :

$$M_l = b_i + a_i \cdot \frac{50 - F_i^{'}}{F_{i+1}^{'} - F_i^{'}}$$

$$M_l = 20 + 30 \cdot \frac{50 - 16}{55 - 16} = 46{,}2 \text{ ha}$$

L'exploitation médiale est de 46,2 ha, cela signifie que la surface cumulée des exploitations ayant moins de 46,2 ha est égale à la surface cumulée des exploitations ayant plus de 46,2 ha.

29. NOMBRES DE PERSONNES PAR MÉNAGE

Mots-clefs

Variable discrète, moyenne arithmétique, diagramme en bâtons

Énoncé

Connaissant la répartition des personnes par ménages en 1999, nous désirons connaître le nombre moyen de personnes par ménage. Nous savons par ailleurs que la population totale est de 57 220 124.

Ménages selon le nombre de personnes du ménage en 1999

Nombre de personnes du ménage	1999	
	Nombre de ménages	Proportion en %
Ensemble	23 808 072	100
1	7 380 109	31,0
2	7 414 180	31,1
3	3 848 586	16,2
4	3 276 615	13,8
5	1 322 024	5,6
6	363 422	1,5
7	121 569	0,5
8	46 921	0,2
9 ou plus	34 646	0,1

Source : INSEE

1. Calculez le nombre moyen de personnes par ménages.
2. Comparez ce résultat à celui obtenu directement.
3. Réalisez une représentation graphique de cette distribution.

Corrigé

1. Pour calculer le nombre moyen de personnes par ménage nous construisons le tableau statistique suivant.

Calcul du nombre moyen de personnes

x_i	n_i	f_i (%)	$f_i x_i$
1	7 380 109	31	31
2	7 414 180	31,1	62,2
3	3 848 586	16,2	48,6
4	3 276 615	13,8	55,2
5	1 322 024	5,6	28
6	363 422	1,5	9
7	121 569	0,5	3,5
8	46 921	0,2	1,6
9,5	34 646	0,1	0,9
Ensemble		100	240

Nous avons choisi pour la dernière catégorie la valeur de 9,5 pour tenir compte du fait qu'il y a des ménages à plus de 9 personnes.

Le nombre moyen de personnes par ménage est de 2,4 personnes.

2. La population de l'enquête est de 57 220 124 personnes.

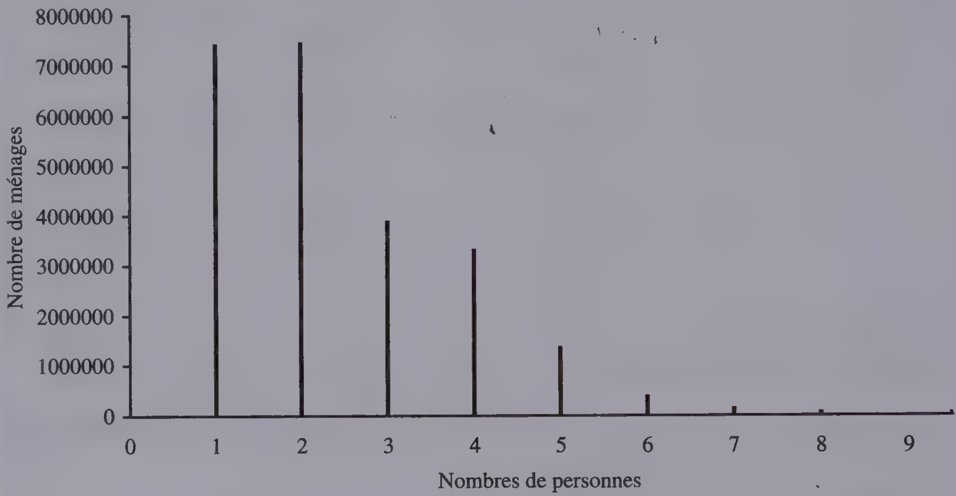
Le tableau suivant nous permet de calculer le nombre de personnes que nous utilisons dans nos calculs.

Calcul de la population de distribution

Nombre de personnes par ménages	Nombre ménages	Nombre de personnes
x_i	n_i	$n_i x_i$
1	7 380 109	7 380 109
2	7 414 180	14 828 360
3	3 848 586	11 545 758
4	3 276 615	13 106 460
5	1 322 024	6 610 120
6	363 422	2 180 532
7	121 569	850 983
8	46 921	375 368
9,5	34 646	329 137
		57 206 827

L'écart entre la population réelle et la population des calculs est de 13 297 soit une erreur de 0,023 %. Pour obtenir une population statistique identique à la population réelle, nous aurions dû prendre pour la dernière catégorie un effectif moyen par ménage de 9,88, ce qui ne change pas la moyenne.

3. La représentation graphique la plus adaptée est un diagramme en bâtons, la variable étant discrète.



30. EFFET DE LA BORNE SUPÉRIEURE SUR LES RÉSULTATS

Mots-clefs

Variable continue, moyenne arithmétique

Énoncé

Soit la répartition suivante des exploitations vinicoles d'une appellation d'un vignoble français :

1. Déterminez la borne supérieure de la dernière classe.

2. Calculez la surface moyenne des exploitations de la distribution.

3. Calculez la surface moyenne de cette distribution en utilisant les données brutes.

4. Expliquez cet écart.

Distribution des exploitations

Classes des exploitations	Superficies (ha)	Nombre d'exploitations
0 à moins de 0,5	24	87
0,5 à moins de 1	32	47
1 à moins de 2	111	74
2 à moins de 5	586	166
5 à moins de 10	722	109
10 à moins de 15	278	24
15 à moins de 20	225	13
20 et plus	421	13
	2399	533

Corrigé

1. La construction du tableau statistique nécessite de déterminer la borne supérieure de la dernière classe. Pour cela nous avons le choix entre deux solutions.

La solution usuelle consiste à prendre l'amplitude de la dernière classe égale à celle de l'avant-dernière, soit une amplitude de 5 ha et un centre de classe de 22,5 ha. Le tableau obtenu sera dit tableau « standard ».

Ici, une seconde solution s'offre à nous. Nous pouvons utiliser des informations supplémentaires pour évaluer l'amplitude de cette dernière classe. Nous savons que la superficie des 13 exploitations regroupées dans la dernière classe est de 421 ha, soit une superficie moyenne de 32,4 ha. Nous choisissons alors un centre de classe proche de cette moyenne soit 32,5 ha. L'amplitude de classe est de 25 et la valeur de la borne supérieure de 45 ha. Nous construirons alors un tableau dit « réaliste ».

Nous disposons de deux solutions différentes, la seconde est plus satisfaisante car plus proche de la réalité des informations disponibles. La surface correspondant à la dernière classe est de 292,5 ha dans le cas « standard » et de 422,5 ha pour la seconde solution.

2. Nous obtenons deux tableaux statistiques pour le calcul de la moyenne : d'une part un tableau « standard », d'autre par un tableau « réaliste. »

Tableau « standard »

Classe	Nombre d'exploitations	Centre de classes	
$[b_i ; b_{i+1}[$	n_i	c_i	$n_i c_i$
[0 ; 0,5[	87	0,25	21,75
[0,5 ; 1[	47	0,75	35,25
[1 ; 2[	74	1,5	111,00
[2 ; 5[	166	3,5	581,00
[5 ; 10[	109	7,5	817,50
[10 ; 15[	24	12,5	300,00
[15 ; 20[	13	17,5	227,50
[20 ; 25]	13	22,5	292,5
	533		2 386,5

Nous pouvons calculer la surface obtenue selon notre hypothèse usuelle.

$$\bar{c} = \frac{1}{n} \sum_{i=1}^{k} n_i c_i = \frac{2\,386{,}5}{533} \cong 4{,}45 \text{ ha}$$

Tableau « réaliste »

Classe	Nombre d'exploitations	Centre de classes	
$[b_i ; b_{i+1}[$	n_i	c_i	$n_i c_i$
[0 ; 0,5[	87	0,25	21,75
[0,5 ; 1[	47	0,75	35,25
[1 ; 2[	74	1,5	111,00
[2 ; 5[	166	3,5	581,00
[5 ; 10[	109	7,5	817,50
[10 ; 15[	24	12,5	300,00
[15 ; 20[	13	17,5	227,50
[20 ; 45]	13	32,5	422,50
	533		2516,50

Nous utilisons les résultats du tableau « réaliste «. La moyenne est alors :

$$\bar{c} = \frac{1}{n} \sum_{i=1}^{k} n_i c_i = \frac{2\,516{,}5}{533} \cong 4{,}72 \text{ ha}$$

L'écart est de 0,27 ha entre les deux moyennes, ce qui nous donne une différence de surface de 143,91 ha pour l'ensemble de la distribution $(0{,}27 \cdot 533 \doteq 143{,}91)$.

3. La superficie moyenne, calculée à l'aide des données brutes, est de :

$$\overline{c} = \frac{2\,399}{533} = 4{,}5 \text{ hectares}$$

L'écart avec le résultat le plus « réaliste » est de 0,22 ha, soit une surface de 117,26 ha. Il est de 0,05 ha avec la détermination « standard » de la dernière classe, soit un écart de 26,65 ha.

4. Nous allons construire le tableau qui permet de comprendre l'écart entre les deux résultats : celui de l'évaluation réaliste et le calcul direct. Nous pouvons comparer les superficies réelles et les superficies calculées. Les superficies calculées sont le produit du centre d'une classe par l'effectif de celle-ci :

$$s' = n_i c_i$$

Tableau des comparaisons

Borne inférieure	Borne supérieure	Superficies	Nombre d'exploitations	Centre de classes	Superficies estimées	Différences
b_i	b_{i+1}	s_i	n_i	c_i	$s'_i = n_i c_i$	$d_i = (s_i - s'_i)$
0	0,5	24	87	0,25	21,75	2,25
0,5	1	32	47	0,75	35,25	– 3,25
1	2	111	74	1,5	111,00	0,00
2	5	586	166	3,5	581,00	5,00
5	10	722	109	7,5	817,50	– 95,50
10	15	278	24	12,5	300,00	– 22,00
15	20	225	13	17,5	227,50	– 2,50
20	45	421	13	32,5	422,5	–1,5
		2399	533		2516,50	– 117,5

La moyenne des écarts est donc $\dfrac{-117{,}5}{533} = -0{,}22$. La différence entre les deux résultats tient à la méthode. L'erreur relative est de l'ordre de 5 % $\left(\dfrac{0{,}22}{4{,}5} = 0{,}0489\right)$.

Nous voyons que l'écart ne s'explique que peu par les valeurs des bornes de la dernière classe. L'écart le plus important apparaît pour la cinquième classe dont la superficie est surestimée. Nous avons retenu comme centre de classe, donc comme surface moyenne de cette classe, 7,5 ha alors que la surface moyenne est de $\dfrac{722}{109} \cong 6{,}6$ ha ; cet écart de 0,9 ha multiplié par 109 rend compte de la différence. La distribution au sein de cette classe n'est pas symétrique avec une majorité d'exploitations proche de la borne inférieure.

Avec la première hypothèse sur l'amplitude de la dernière classe, seules les deux dernières du tableau sont modifiées.

Hypothèse d'une borne de la dernière classe à 25

Borne inférieure	Borne supérieure	Superficies	Nombre d'exploitations	Centre de classes	Superficies estimées	Différences
b_i	b_{i+1}	s_i	n_i	c_i	$s'_i = n_i c_i$	$d_i = (s_i - s'_i)$
0	0,5	24	87	0,25	21,75	2,25
0,5	1	32	47	0,75	35,25	– 3,25
1	2	111	74	1,5	111,00	0,00
2	5	586	166	3,5	581,00	5,00
5	10	722	109	7,5	817,50	– 95,50
10	15	278	24	12,5	300,00	– 22,00
15	20	225	13	17,5	227,50	– 2,50
20	25	421	13	22,5	292,5	128,55
		2399	533		2386,50	12,5

La sous-estimation de l'importance de la dernière classe inverse le signe de la différence. Ce résultat plus proche n'est que l'effet du hasard. Il n'est pas explicable. Cet exemple montre l'importance des hypothèses dans l'obtention des résultats.

31. Effets de structure

Mots-clefs

Moyenne arithmétique, effets de structure

Énoncé

Nous disposons de la distribution des salaires annuels nets en euro par catégories de salariés pour deux régions françaises l'Alsace et la région Rhône-Alpes.

Région Alsace

Catégorie de salariés	Cadres	Professions intermédiaires	Employés	Ouvriers
Salaires	38505	21473	14855	16412
Effectifs	78 957	156 616	189 640	215 310

Région Rhône-Alpes

Catégorie de salariés	Cadres	Professions intermédiaires	Employés	Ouvriers
Salaires	40298	21427	15150	15676
Effectifs	273 846	548 469	617 548	607 185

1. Calculez les salaires moyens pour chacune des régions.
2. Comment s'explique la différence ?

Corrigé

1.

Moyenne des salaires pour la région Alsace

		Cadres	Professions intermédiaires	Employés	Ouvriers	Ensemble
Salaires	x_i	38 505	21 473	14 855	16 412	
Effectifs	n_i	78 957	156 616	189 640	215 310	640 523
Fréquences	f_i	0,123	0,245	0,296	0,336	1
	$f_i x_i$	4736,115	5260,885	4397,08	5514,432	19908,51

La moyenne s'obtient immédiatement

$$\overline{x} = \sum_{i=1}^{k} f_i x_i = 19\,908{,}51 \text{ euros}$$

Le salaire moyen dans la région Alsace est de 19 909 euros.

Moyenne des salaires pour la région Rhône-Alpes

		Cadres	Professions intermédiaires	Employés	Ouvriers	Ensemble
Salaires	x_i	40 298	21 427	15 150	15 676	
Effectifs	n_i	273 846	548 469	617 548	607 185	2 047 048
Fréquences	f_i	0,134	0,268	0,302	0,297	1
	$f_i x_i$	5399,932	5742,436	4575,3	4655,772	20 373

Le salaire moyen dans la région Rhône-Alpes est de 20 373 euros.

L'écart est de 464 euros.

2. Les salaires des différentes catégories de salariés ne sont pas systématiquement supérieurs pour la région Rhône-Alpes. Nous avons deux effets : un effet salaire et un effet de structure.

Nous allons tout d'abord calculer ce que serait le salaire moyen avec les salaires de la région Rhône-Alpes et la structure des salariés pour la région Alsace.

Moyenne des salaires pour la région Rhône-Alpes avec la structure de l'Alsace

		Cadres	Professions intermédiaires	Employés	Ouvriers	Ensemble
Salaires	x_i	40 298	21 427	15 150	15 676	
Fréquences	f_i	0,123	0,245	0,296	0,336	1
	$f_i x_i$	4956,654	5249,615	4484,4	5267,136	19 958

La moyenne est de 19 958 euros alors que la moyenne des salaires rhône-alpins est de 20 373 euros. L'écart correspondant aux salaires par catégories est de 415 euros. Cet écart est dû à la différence des structures. L'écart avec la moyenne des salaires alsaciens est de 50 euros qui est relatif aux différences de salaires.

Nous pouvons calculer ce que serait le salaire moyen en Rhône-Alpes avec la structure de l'Alsace.

Moyenne des salaires pour la région Alsace avec la structure Rhône-Alpes

		Cadres	Professions intermédiaires	Employés	Ouvriers	Ensemble
Salaires	x_i	38 505	21 473	148 55	164 12	
Fréquences	f_i	0,134	0,268	0,302	0,297	1
	$f_i x_i$	5159,67	5754,764	4486,21	4874,364	20 275

Le salaire moyen de la région Rhône-Alpes avec la répartition par catégorie de la région Alsace est de 20 575 euros. L'écart dû aux différences des répartitions est de 202 euros (20 575-20 373).

Les salaires ouvriers et professions intermédiaires sont plus élevés en Alsace, alors que les salaires des cadres et des employés sont plus favorables en Rhône-Alpes.

32. Redressement d'échantillon

Mots-clefs

Moyenne arithmétique

Énoncé

Lors d'une enquête effectuée auprès d'un échantillon de 6 800 répondants, sur leur opinion concernant un nouvel aménagement sportif, les résultats ont été les suivants :

Tableau des résultats

Classe d'âge	Échantillon	Opinion favorable
15 - 30 ans	1 000	0,85
30 - 50 ans	2 000	0,54
50 ans et +	3 800	0,35
Total	6 800	

Répartition par âges au sein de la population de référence

Classe d'âge	Pourcentage
15 - 30 ans	0,25
30 - 50 ans	0,35
50 ans et +	0,40

1. Quel est le pourcentage d'opinions favorables dans l'échantillon ?
2. Quel est le pourcentage d'opinions favorables dans la population ?
3. L'échantillon est-il représentatif de la population de référence ? Expliquez pourquoi.

Corrigé

1. Ce pourcentage est la moyenne arithmétique des pourcentages d'opinions favorables.

Classe d'âge	Échantillon	Fréquences	Opinions favorable	
	n_i	f_i	x_i	$f_i x_i$
[15 ; 30[	1 000	0,15	0,85	0,1250
[30 ; 50[	2 000	0,29	0,54	0,1588
[50 ans et +	3 800	0,56	0,35	0,1956
Total	6 800	1		0,4794

Dans l'échantillon, le taux d'opinions favorable est de 47,94 % soit 48 %.

La répartition par âge n'est pas identique dans l'échantillon et dans la population de référence. Il est nécessaire de redresser l'échantillon pour avoir une évaluation de la proportion d'opinions favorables dans la population de référence.

2. Nous supposons que le pourcentage d'opinions favorables est une caractéristique de chaque classe d'âge.

Écart entre échantillon théorique et échantillon empirique

Classe d'âge	Structure de la population de référence	Échantillon théorique	Échantillon empirique	Coefficient de redressement
[15 ; 30[	0,25	1 700	1 000	1,70
[30 ; 50[	0,35	2 380	2 000	1,19
[50 ans et +	0,4	2 720	3 800	0,72
Total	1	6 800	6 800	

Le coefficient de redressement est obtenu en divisant les effectifs de l'échantillon théorique par ceux de l'échantillon empirique.

Nous disposons de plus de répondants dans la dernière classe que dans les deux autres, or les réponses sont plus favorables pour ces deux classes que pour la troisième. Nous devons donc supposer un biais qui minore le pourcentage d'opinions favorables dans la population de référence.

Nous allons détailler la méthode retenue en calculant ce que serait la répartition des effectifs.

Classe d'âge	Effectif empirique favorable	Effectif théorique favorable
[15 ; 30[	850	1 445
[30 ; 50[	1 080	1 285
[50 ans et +	1 330	952
Total	3 260	3 682

L'échantillon théorique des opinons favorables est obtenu en multipliant le nombre d'opinions favorables dans l'échantillon empirique par le coefficient de redressement, nous obtenons alors :

$$\text{moyenne dans l'échantillon redressé} = \frac{3\,682}{6\,800} = 0{,}5414$$

Le pourcentage d'opinions favorables dans la population est donc de 54,14 % au lieu de 47,94 % comme dans l'échantillon empirique. Nous ne disposons pas des outils pour estimer si la différence entre les deux pourcentages est significative.

3. L'échantillon empirique diffère de l'échantillon théorique, il n'est pas fidèle à la structure de la population dans son ensemble. Sous réserve de régularité des opinions dans chaque catégorie de la population, il est possible de redresser l'échantillon et d'obtenir un résultat plus satisfaisant. Nous voyons que les résultats bruts conduisent à une majorité d'opinions défavorables alors que nous devons supposer qu'il y a une majorité d'opinions favorables.

33. DES TAUX ET DES IMPÔTS

Mots-clefs

Changement de variable, moyenne arithmétique, variance et écart type, sous populations

Énoncé

Soit un groupe de personnes dont le revenu annuel moyen avant impôt est de 12 500 € avec un écart type de 3250 €.

1. On prélève un impôt uniforme de 1 250 € par personne.
   a. Quel est le revenu moyen après impôt ?
   b. Quel est l'écart type du revenu après impôt ?
2. L'impôt est cette fois prélevé selon un taux uniforme de 12,5 %.
   a. Quel est le revenu moyen après impôt ?
   b. Quel est l'écart type du revenu après impôt ?
3. Un second groupe de personnes dispose d'un revenu moyen de 15 050 € (avant impôt) avec un écart type de 4 500 €. L'effectif de ce groupe est deux fois plus important que celui du groupe précédent.
   a. Quel est le revenu moyen pour l'ensemble de ces deux groupes ?
   b. Quel est l'écart type pour l'ensemble de ces deux groupes ?

Corrigé

1. a. La moyenne du revenu annuel est obtenue par la formule suivante où r_i représente le centre de la tranche de revenu de la personne i ; n_i le nombre de personnes appartenant à la même tranche revenu r_i ; f_i la fréquence relative de la tranche de revenu i.

$$\overline{r} = \frac{1}{n}\sum_{i=1}^{m} n_i r_i = \sum_{i=1}^{m} f_i r_i \text{ avec } \sum_{i=1}^{m} f_i = 1$$

Avec le prélèvement uniforme de 1 250 € les revenus de chaque tranche diminuent et deviennent :

$$r' = r_i - 1250 = r_i - a$$

Donc

$$\overline{r} = \sum_{i=1}^{m} f_i r_i' = \sum_{i=1}^{m} f_i (r_i - a) = \sum_{i=1}^{m} f_i r_i - a = \overline{r} - a$$

Le revenu moyen après impôt devient :

12 500 – 1 250 = 11 250 €

b. La variance et l'écart type de la distribution ne sont pas modifiées. En effet, la variance de la distribution est donnée par :

$$V(r) = \sum_{i=1}^{m} f_i(r_i - \overline{r})^2$$

La variance de la nouvelle distribution des r' est :

$$V(r') = \sum_{i=1}^{m} f_i(r_i' - \overline{r}')^2 = \sum_{i=1}^{m} f_i\,[(r_i - a) - (\overline{r} - a)]^2 = \sum_{i=1}^{m} f_i(r_i - \overline{r})^2 = V(r)$$

En d'autres termes, un changement d'origine ne modifie pas la dispersion d'une distribution.

2. a. Nous examinons maintenant la modification de la moyenne sous l'effet d'un taux d'imposition uniforme.

La nouvelle variable r_i' s'obtient comme le produit de l'ancienne par un multiplicateur :

$$r' = r_i(1 - t)$$

La nouvelle moyenne s'obtient par application de la définition de la moyenne

$$\overline{r}_i' = \sum_{i=1}^{m} f_i r_i(1 - t) = (1 - t)\sum_{i=1}^{m} f_i r_i = (1 - t)\overline{r}$$

La moyenne des revenus après impôt est cette fois de :

$12\,500 \cdot (1 - 0{,}125) = 10\,937{,}5$ €

b. Nous examinons maintenant la modification de la moyenne et de la variance sous l'effet d'un taux d'imposition uniforme. La nouvelle variable r_i' s'obtient comme le produit de l'ancienne par un multiplicateur :

$$r_i' = r_i(1 - t)$$

La nouvelle moyenne s'obtient par application de la définition de la moyenne :

$$\overline{r}' = \sum_{i=1}^{m} f_i r_i(1 - t) = (1 - t)\sum_{i=1}^{m} f_i r_i = (1 - t)\overline{r}$$

La moyenne des revenus après impôt est cette fois de :

$12\,500 \cdot (1 - 0{,}125) = 10\,937{,}5$ €

La nouvelle variance est :

$$V(r') = \sum_{i=1}^{m} f_i(r' - \overline{r}')^2 = \sum_{i=1}^{m} f_i[(1-t)r_i - (1-t)\overline{r}]^2$$
$$= (1-t)^2 \sum_{i=1}^{m} f_i(r_i - \overline{r})^2 = (1-t)^2 V(r)$$

L'écart type est multiplié par $1 - t$.

Le nouvel écart type devient : $\sigma_r = 3\,250 \cdot 0{,}875 = 2\,843{,}75$. Un prélèvement proportionnel réduit la dispersion des revenus. La distribution des revenus après impôt est moins dispersée que celle des revenus avant impôt.

3. a. Nous appliquons le théorème sur la moyenne d'une population composée de h sous populations :

$$\overline{x} = \sum_{j=1}^{h} f_j \overline{x}_j$$

La moyenne de l'ensemble des deux groupes est :

$$\overline{x} = 15\,050 \cdot \frac{2}{3} + 12\,500 \cdot \frac{1}{3} = 14\,200 \text{ €}$$

b. Nous appliquons les théorèmes sur la moyenne et la variance d'une population composée de h sous populations :

$$\overline{x} = \sum_{j=1}^{h} f_j \overline{x}_j$$

$$V(x) = \sum_{j=1}^{h} f_j V_j(x) + \sum_{j=1}^{h} f_j (\overline{x}_j - \overline{x})^2$$

La moyenne de l'ensemble des deux groupes est :

$$\overline{x} = 15\,050 \cdot \frac{2}{3} + 12\,500 \cdot \frac{1}{3} = 14\,200 \text{ €}$$

La variance de distribution de la population totale est :

$$V(x) = \left[\frac{1}{3}(3\,250)^2 + \frac{2}{3}(4\,500)^2\right]$$
$$+ \left[\left(\frac{1}{3}(12\,500)^2 + \frac{2}{3}(15\,050)^2\right) - (14\,200)^2\right] = 352\,641\,667$$

L'écart type est alors : $\sigma_x \cong 18\,778{,}76$.

34. Calcul d'une variance et d'un écart type

Mots-clefs

Variable continue, moyenne arithmétique, coefficient de variation

Énoncé

Nous disposons de la distribution, classée, des gains mensuels bruts des salariés d'une entreprise quelconque.

Tranche de gains mensuels (100€)

	– de 20	[20 ; 25[	[25 ; 30[	[30 ; 35[	[35 ; 40[	[40 ; 45[	[45 ; 50[	[50 ; 60[	[60 ; 70[	[70 ; 80[	[80 ; 100]
Nombre de salariés	118	272	491	588	575	464	343	353	114	30	13

1. Calculez la moyenne arithmétique de cette série.

2. Calculez la variance et l'écart type. En déduire le coefficient de variation.

Corrigé

Les réponses aux questions seront facilitées par le tableau suivant :

Tableau des calculs

	c_i	n_i	f_i	F_i	$f_i c_i$	$f_i c_i^2$
[15 ; 20[	17,5	118	3,5	3,5	61,25	1 071,875
[20 ; 25[	22,5	272	8,1	11,6	182,25	4 100,625
[25 ; 30[	27,5	491	14,6	26,2	401,50	11 041,25
[30 ; 35[	32,5	588	17,5	43,7	568,75	18 484,375
[35 ; 40[	37,5	575	17,1	60,8	641,25	24 046,875
[40 ; 45[	42,5	464	13,8	74,6	586,50	24 926,25
[45 ; 50[	47,5	343	10,2	84,8	484,50	23 013,75
[50 ; 60[	55,0	353	10,5	95,3	577,50	31 762,50
[60 ; 70[	65,0	114	3,4	98,7	221,00	14 365,00
[70 ; 80[	75,0	30	0,9	99,6	67,50	5 062,50
[80 ; 100]	90,0	13	0,4	100,0	36,00	3 240,00
Total			100,0		3 828,00	161 115,00

1. La valeur de la moyenne des centres de classe $\overline{c}$ s'obtient par lecture de la somme de la colonne $f_i c_i$ qu'il faut diviser par 100 puisque les fréquences sont exprimées en pourcentage.

$$\bar{c} = \frac{1}{100}\sum_{i=1}^{k} f_i c_i = 38{,}28 \cong 38{,}3$$

2. La variance s'obtient à partir du tableau :

$$V(c) = \frac{1}{100}\sum_{i=1}^{k} f_i c_i^2 - \bar{c}^2 = 1\,611{,}15 - (38{,}28)^2 = 145{,}7916$$

d'où l'écart type :

$$\sigma_c = \sqrt{V(c)} = \sqrt{145{,}7916} = 12{,}07$$

Le coefficient de variation est alors :

$$CV = \frac{12{,}07}{38{,}28} = 0{,}315$$

Nous avons une série faiblement dispersée.

35. Les enfants par famille

Mots-clefs

Variable discrète, moyenne, écart type, médiane, écart absolu moyen

Énoncé

Les familles selon le nombre d'enfants de 0 à 16 ans se répartissent de la manière suivante.

Nombre d'enfants par famille

Nombre d'enfants	0	1	2	3	4	5 et plus	Total
Nombre de familles	7 130	3 201	2 498	919	241	130	14 119

1. Pour l'ensemble des familles :
   a. Calculez la moyenne.
   b. Calculez l'écart type.
2. Pour les familles ayant des enfants :
   a. Déterminez la médiane.
   b. Calculez la moyenne et l'écart type.
   c. Calculez l'écart absolu moyen par rapport à la médiane.

Corrigé

1. a. Le nombre moyen d'enfants, par famille est :

$$\overline{x} = \frac{\sum_{i=1}^{k} f_i x_i}{100} = \frac{89{,}5}{100} = 0{,}895$$

Soit environ 0,9 enfant par famille. Nous constatons, encore une fois, que ce nombre est inférieur à 1. Il y a plus de familles que d'enfants dans cette population. Si toutes les familles avaient le même nombre d'enfant, elles auraient moins d'un enfant.

b. Nous calculons tout d'abord la variance

$$V(x) = \sum_{i=1}^{m} f_i x_i^2 - \left(\sum_{i=1}^{m} f_i x_i\right)^2 = 2{,}072 - (0{,}895)^2 \cong 1{,}2712$$

soit

$$\sigma_x = \sqrt{V(x)} = \sqrt{1{,}27} \cong 1{,}1 \text{ enfants}$$

Ce qui est une assez forte dispersion dans cette série, d'autant plus que la valeur maximale a été fixée à 5,5 mais qu'elle lui est peut-être supérieure.

2. Pour les calculs, nous utiliserons le tableau statistique habituel.

Familles ayant des enfants

Nombre d'enfants	x_i	n_i	f_i	$f_i x_i$	$f_i x_i^2$
1	1	3 201	45,8	45,8	45,8
2	2	2 498	35,7	71,4	142,8
3	3	919	13,1	39,3	117,9
4	4	241	3,5	14,0	56,0
5 et plus	5,5	130	1,9	10,5	57,5
Total		6 989	100,0	181,0	420,0

a. La médiane de cette distribution est de deux enfants.

b. Le nombre moyen d'enfants par famille ayant des enfants

$$\overline{x} = \frac{\sum_{i=1}^{k} f_i x_i}{100} = \frac{181}{100} = 1{,}81$$

Soit environ 2 enfants par famille. Les familles ayant des enfants ont en moyenne deux enfants.

Pour calculer l'écart-type, nous calculons tout d'abord la variance

$$V(x) = \sum_{i=1}^{m} f_i x_i^2 - \left(\sum_{i=1}^{m} f_i x_i\right)^2 = 4{,}2 - (1{,}81)^2 \cong 0{,}9239$$

$$\sigma_x = \sqrt{V(x)} = \sqrt{0{,}9239} \cong 0{,}961 \text{ enfants}$$

La distribution des familles ayant des enfants est moins étalée que l'ensemble des familles.

c. L'écart absolu médian par rapport à la médiane est défini par :

$$e_M = \frac{1}{n}\sum_{i=1}^{m} n_i |x_i - M_e| = \sum_{i=1}^{m} f_i |x_i - M_e|$$

Calcul de l'écart absolu moyen

x_i	f_i	$\|x_i - M_e\|$	$f_i \|x_i - M_e\|$
1	0,458	1	0,458
2	0,357	0	0,000
3	0,131	1	0,131
4	0,035	2	0,070
5,5	0,019	3,5	0,0665
	1,000		0,7255

L'écart absolu moyen est de 0,72 enfant. Nous constatons que cette évaluation de la dispersion est plus faible que l'évaluation avec l'écart type. Il constitue une évaluation maximale de la dispersion.

36. Vitesse moyenne

Mots-clefs

Moyenne harmonique

Énoncé

Sur le trajet de Paris à Grenoble, le TGV roule à 240 km/h sur 60 % du trajet, 150 km/h sur 25 % et à 120 km/h sur 15 % du trajet. À quelle vitesse moyenne parcourt-il l'ensemble du trajet ?

Corrigé

La vitesse moyenne entre Paris et Grenoble est obtenue par le rapport entre la distance parcourue et le temps mis pour la parcourir. La moyenne est donc une moyenne harmonique des vitesses. Nous pouvons le montrer simplement.

Soit x la distance entre Paris et Grenoble

$$x = \nu \cdot t$$

$$t = t_1 + t_2 + t_3$$

$$\nu = \frac{x}{t}$$

avec :

$$t = \frac{0,6x}{240} + \frac{0,25x}{150} + \frac{0,15x}{120}$$

donc :

$$\frac{1}{\nu} = \frac{0,6}{240} + \frac{0,25}{150} + \frac{0,15}{120} + \frac{0,15}{120} = 0{,}005\,416\,67$$

Soit une vitesse moyenne de $v = 184{,}6$ km/h.

La vitesse moyenne est la moyenne harmonique des vitesses pondérées par la part de la distance parcourue. La vitesse moyenne est obtenue par application de la formule de définition de la moyenne harmonique :

$$\frac{1}{\nu} = \sum_{i=1}^{k} \frac{f_i}{\nu_i}$$

37. Densité moyenne

Mots-clefs

Moyenne harmonique

Énoncé

Un pays est divisé en plusieurs régions qui contiennent respectivement : 15 %, 18 %, 40 % et 27 % de la population totale du pays. La densité moyenne d'habitants par km^2 est pour chacune des régions respectivement de 120, 150, 200, 350. Quelle est la densité moyenne de la population pour l'ensemble du pays ?

Corrigé

La densité moyenne pour l'ensemble du pays est obtenue en calculant la moyenne harmonique des densités régionales.

Si l'on appelle P la population totale, T le territoire complet inconnu, la densité moyenne de la population sera :

$$d = \frac{P}{T}.$$

Si $P = \sum_{i=1}^{k} P_i$. est la population présente sur le territoire, la densité du territoire sera :

$$d = \frac{\sum_{i=1}^{k} P_i}{\sum_{i=1}^{k} T_i} \text{ avec } T_i = \frac{P_i}{d_i}$$

$$d = \frac{\sum_{i=1}^{k} P_i}{\sum_{i=1}^{k} \frac{P_i}{d_i}} = \frac{P}{\sum_{i=1}^{k} \frac{P_i}{d_i}}$$

$$\frac{1}{d} = \frac{\sum_{i=1}^{k} \frac{P_i}{d_i}}{P} = \sum_{i=1}^{k} \frac{P_i}{P} \cdot \frac{1}{d_i}$$

$$\frac{1}{d} = \sum_{i=1}^{k} \frac{f_i}{d_i}.$$

D'où

$$\frac{1}{d} = \sum_{i=1}^{k} \frac{f_i}{d_i}$$

$$\frac{1}{d} = \frac{0,15}{120} + \frac{0,18}{150} + \frac{0,40}{200} + \frac{0,27}{350} = 0,052\,214$$

La densité moyenne est $d = 191,5$ habitants au km^2.

38. Taux de croissance des prix

Mots-clefs

Taux moyens, indice, multiplicateur

Énoncé

Les taux annuels de croissance de l'indice général des prix à la consommation pour un pays X ont évolué de la manière suivante.

Taux de croissance des prix à la consommation
(janvier n par rapport à janvier n-1)

	2001	2002	2003	2004	2005	2006
En %	1,1	2,2	2,0	2,0	1,6	2,0

1. Quel a été le taux de croissance moyen du niveau général des prix sur la période 2001 – 2006 ?

2. Si le montant horaire du SMIC était de 6,67 € en janvier 2001, quel devrait être son niveau en janvier 2006 :

 a. pour qu'il y ait maintien du pouvoir d'achat du montant horaire du SMIC ;

 b. pour que le taux horaire du SMIC ait gagné 10 % de pouvoir d'achat.

Corrigé

1. Pour déterminer le taux de croissance moyen de l'indice des prix à la consommation nous devons calculer le multiplicateur moyen des prix sur la période. Le multiplicateur moyen est la moyenne géométrique des multiplicateurs annuels.

$$1 + r = \sqrt[5]{\prod_{i=1}^{i=5}(1 + r_i)}$$

	2001	2002	2003	2004	2005	2006
En %	1,1	2,2	2,0	2,0	1,6	2,0
Multiplicateur		1,022	1,02	1,02	1,016	1,02

$$1 + r = \sqrt[5]{1{,}022 \cdot 1{,}02 \cdot 1{,}02 \cdot 1{,}016 \cdot 1{,}02} = 1{,}020$$

Le taux de croissance moyen des prix est donc de 2,0 %.

2. L'évolution du pouvoir d'achat du SMIC est donnée par le rapport :

$$\frac{\text{taux horaire du SMIC en janvier 2001}}{\text{taux horaire du SMIC en janvier 2006}}$$

a. Pour qu'il y ait strict maintien du pouvoir d'achat du SMIC, le rapport ci-dessus doit être égal à la hausse des prix à la consommation sur la période considérée. Le taux horaire doit être multiplié par le multiplicateur des prix, soit 1,02. Le taux horaire du SMIC devrait être : $6{,}67 \cdot 1{,}02 = 6{,}8034 \cong 6{,}8$ €

b. Dans le cas où le pouvoir d'achat du SMIC aurait crû de 10 %, le multiplicateur serait différent et multiplié par 1,1 :

$$1{,}02 \cdot 1{,}10 = 1{,}122$$

Le taux horaire du SMIC devrait être de $6{,}67 \cdot 1{,}122 = 7{,}4837 \cong 7{,}5$ €.

Nous pouvons comparer ce taux de 7,5 au taux réel en janvier 2006 qui était de 8,03.

39. VARIATION D'UN INDICE BOURSIER

Mots-clefs

Taux moyen, indice, multiplicateur

Énoncé

L'indice boursier d'un actif financier coté a connu au cours des six derniers mois les évolutions suivantes : 9,2 % ; 7,8 % ; – 3 % ; – 16 % ; 11 % ; – 2,8 %

1. Quel est le taux mensuel moyen de variation de cet indice boursier ?
2. Un opérateur avait acquis pour 1400 € de cet actif au début de la période. S'il vend à la fin de la période, aura-t-il réalisé un bénéfice ?
3. L'opérateur désirait réaliser un bénéfice supérieur ou égal à 15 % de son investissement. Quand aurait-il du vendre ?

Corrigé

1. Le calcul du taux de variation est impossible directement ; il est nécessaire d'utiliser les multiplicateurs et de calculer le multiplicateur moyen.

Tableau des multiplicateurs

Taux d'évolution	Multiplicateurs
9,2	1,092
7,8	1,078
–3	0,97
–16	0,84
11	1,11
–2,8	0,972
Ensemble	1,03486

Le multiplicateur moyen est la moyenne géométrique des multiplicateurs.

$$(1+r) = \sqrt[6]{1{,}034\,860\,149} = 1{,}005\,72$$

$$\text{Taux moyen} = 1{,}0057 - 1 = 0{,}0057 \cong 0{,}6\ \%$$

Le taux de croissance moyen est donc environ de 0,6 %.

2. Le produit des multiplicateurs étant supérieur à un, l'opérateur réalisera un bénéfice. Ce bénéfice serait égal à la différence entre les 1400 euros avancés et la valeur de l'actif en fin de période.

La valeur de l'actif est obtenue par le produit de la valeur initiale par le produit des multiplicateurs. Soit :

$$1\,400 \cdot 1{,}0348 \cong 1\,449\ €.$$

Le bénéfice est donc de 49 €.

3. Quelle aurait été la meilleure période pour vendre ? Nous calculons pour chaque période la valeur de l'actif compte tenu de l'évolution du cours.

Calcul des bénéfices

Multiplicateur	Produit des multiplicateurs	Valeur du patrimoine	Bénéfice (1)
1,092	1,092	1529	129
1,078	1,177	1648	248
0,97	1,142	1599	199
0,84	0,959	1343	–57
1,11	1,065	1491	91
0,972	1,035	1449	49

(1) calculés en fonction de la valeur des 1400 de départ

Un bénéfice égal ou supérieur à 15 % de la mise de fonds initiale, devait être égal ou supérieur à 210 €.

C'est à la fin du deuxième mois que le bénéfice répondait au critère posé.

40. ÉVOLUTION DU TAUX DE DIOXYDE

Mots-clefs

Taux moyen, multiplicateurs

Énoncé

L'évolution de la pollution en dioxyde de soufre (SO2) est donnée dans le tableau suivant par secteur.

Pollution au dioxyde de soufre

	1980 (%)	1990 (%)	1994 (%)
Résidences et bureaux	12,6	15,1	14,5
Industrie	31,8	21,6	22,0
Centrales électrothermiques	36,5	26,1	17,3
Transformations d'énergie	6,3	10,2	12,6
Procédés industriels	9,0	14,9	16,5
Transports	3,8	12,1	17,2
	100,0	100,0	100,0

Source : TEF 93/94

Cette pollution représentait 3348 milliers de tonnes en 1980 et 1200 en 1990, 961 en 1994.

Quel est le taux de décroissance moyen entre 1980 et 1990 ?

Corrigé

Le taux de décroissance moyen ne peut être calculé directement. Nous calculons le multiplicateur sur les dix ans soit :

$$(1+r)^{10} = \frac{1\,200}{3\,348} = 0{,}3584$$

Le multiplicateur moyen annuel est la moyenne géométrique du multiplicateur précédent soit :

$$1+r = \sqrt[10]{0{,}3584} = 0{,}975$$

Le taux de décroissance annuel moyen donc de :

$$r = 1 - 0{,}9025 = 0{,}0975$$

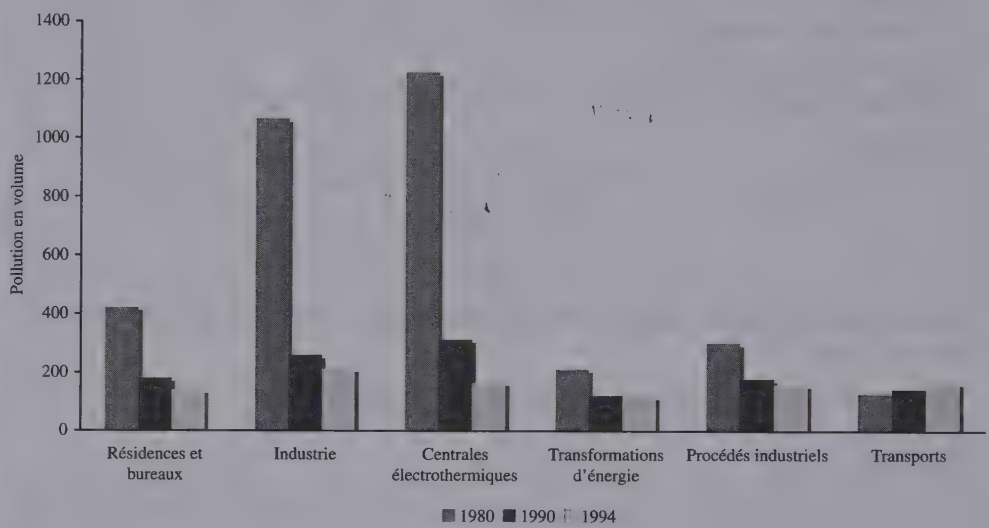
soit 9,75 %

Tableau des pollutions en volume

Sources de pollution	1980	1990	1994
Résidences et bureaux	422	181	139
Industrie	1 065	259	211
Centrales électrothermiques	1 222	313	166
Transformations d'énergie	211	122	121
Procédés industriels	301	179	159
Transports	127	145	165
Ensemble	3 348	1 200	961

Le graphique suivant montre de façon éclairante la réduction massive de cette forme de pollution.

Représentation en volume.



41. TAILLE DU CÔTÉ MOYEN DE LA PARCELLE MOYENNE

Mots-clefs

Moyenne arithmétique

Énoncé

Nous disposons d'un certain nombre de parcelles constructibles de forme carrée. Dans le cadre d'une meilleure structuration de l'espace, les responsables veulent pouvoir disposer de la valeur du côté moyen de la parcelle moyenne.

Distribution des parcelles

Nombres de parcelles	Côté de chaque parcelle
45	9
25	11
15	12
10	17

Quel est le côté moyen de cet ensemble de parcelles ?

Corrigé

Nous ne pouvons pas calculer directement la longueur du côté moyen. Nous pouvons calculer la moyenne arithmétique des côtés.

Calcul direct du côté moyen des parcelles

Nombre de parcelles	Côté des parcelles	
n_i	x_i	$n_i x_i$
45	9	405
25	11	275
15	12	180
10	17	170
95		1030

La moyenne arithmétique des côtés des parcelles est donc :

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{k} n_{i.} x_i = \frac{1030}{95} = 10,842$$

La surface obtenue est de $10,842^2 \cdot 95 = 11\ 167,37$

Or la surface réelle est de 11 720 comme le montre le calcul dans le tableau suivant.

Calcul de la surface totale

Côté des parcelles	Nombre de parcelles	Surfaces
9	45	3 645
11	25	3 025
12	15	2 160
17	10	2 890
		11 720

Puisque nous disposons des surfaces, nous pouvons calculer la moyenne quadratique des côtés des parcelles donc la surface moyenne des parcelles.

Calcul de la moyenne quadratique

Côté des parcelles	Nombre de parcelles	
x_i	n_i	$n_i x_i^2$
9	45	3645
11	25	3025
12	15	2160
17	10	2890
	95	11720

La moyenne quadratique est donc

$$Q^2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 = \frac{11\,720}{95} = 123{,}368$$

La moyenne quadratique est celle de la surface moyenne des parcelles, la surface totale est : $123{,}368 \cdot 95 = 11\,719{,}96$. L'écart avec la surface d'ensemble est de 0,04 m^2.

Le côté moyen est donc $Q = \sqrt{123{,}368421} \cong 11{,}1071$, soit environ 11,11 m.

42. CONCENTRATION DES SALARIÉS

Mots-clefs

Courbe de Lorenz, indice de Gini

Énoncé

Nous disposons, dans le tableau suivant, de la répartition d'entreprises par classe d'effectifs de salariés (chiffres fictifs).

Répartition des salariés par tranche d'effectifs

Tranche d'effectifs	Nombre d'entreprises (en milliers)
0 - 4 salariés	160
5 à 19 salariés	250
20 à 49 salariés	50
50 à 99 salariés	18
100 à 199 salariés	15
200 à 499 salariés	5
500 salariés et plus	2

1. Quelle est la variable étudiée ? Quel en est le type ?
2. Tracez la courbe de Lorenz.
3. Calculez l'indice de Gini de la série.
4. Que pensez-vous de la concentration de cette série ?

Corrigé

1. La série statistique porte sur les effectifs de salariés. Nous considérerons que cette variable est continue. S'il s'agit de personnes, variable discrète, nous traiterons cette variable discrète comme une variable continue, s'il s'agit du nombre de salariés calculés en équivalent temps plein, nous avons effectivement une variable continue.

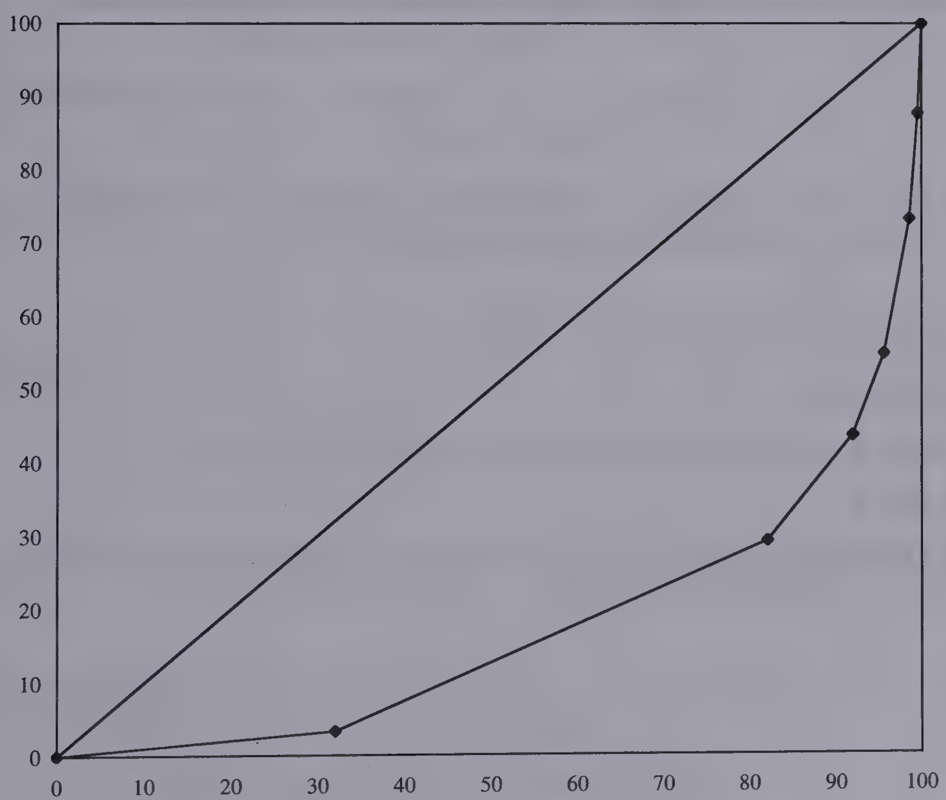
Un tableau permet de répondre aux questions posées.

Tableau des calculs

Classes de salariés	c_i	n_i	f_i (%)	F_i	$f_i c_i$	f'_i (%)	F'_i	$f_i(F'_{i-1} + F'_i)$
[0 ; 5[	2,5	160	32	32	80,0	3,3	3,3	105,6
[5 ;20[	12,5	250	50	82	625,0	25,8	29,1	1618,6
[20 ;50[	35	50	10	92	350,0	14,4	43,5	725,8
[50 ;100[	75	18	3,6	95,6	270,0	11,1	54,6	353,3
[100 ;200[	150	15	3	98,6	450,0	18,6	73,2	383,5
[200 ; 500[	350	5	1	99,6	350,0	14,4	87,6	160,8
[500 ;1000]	750	2	0,4	100	300,0	12,4	100,0	75,1
		500	100		2425,0	100,0		3422,6

2. La courbe de Lorenz est la représentation graphique des fréquences cumulées donc des points (F_i, F'_i).

Courbe de Lorenz



3. L'indice de Gini I_G est le double de la surface comprise entre la courbe de Lorenz et la première diagonale. Le double de la surface complémentaire est obtenu en calculant

$$\sum_{i=1}^{k} f_i(F'_i + F'_{i-1}).$$

L'indice de Gini est alors

$$I_G = 1 - \sum_{i=1}^{k} f_i(F'_i + F'_{i-1})$$

$$\sum_{i=1}^{k} f_i(F'_i + F'_{i-1}) = 3\,422{,}6$$

Ce chiffre est obtenu en faisant le produit de fréquences en pourcentage ; nous devons diviser ce résultat par 100 000 pour obtenir la surface complémentaire.

$$I_G = 1 - \sum_{i=1}^{k} f_i(F'_i + F'_{i-1}) = 1 - 0{,}34226 = 0{,}65774 \cong 0{,}66$$

4. Les résultats montrent une concentration importante de la distribution des salariés au profit des plus grandes entreprises.

43. CONCENTRATION DE L'ÉDITION

Mots-clefs

Indice de Gini, courbe de Lorenz, médiane, moyenne, écart-type

Énoncé

La répartition des maisons d'édition (livres) selon le chiffre d'affaires est la suivante pour les années 2000 et 2004 :

Répartition des maisons d'édition par chiffre d'affaires

C.A. (en millions d'euros)	Nombre de maisons d'édition 2000	2004
100 et plus	15	21
de 50 à moins de 100	13	25
de 20 à moins de 50	41	36
de 10 à moins de 20	36	44
de 5 à moins de 10	51	49
de 2 à moins de 5	72	71
de 1 à moins de 2	57	55
de 0,5 à moins de 1	65	46
de 0,2 à moins de 0,5	55	40

Pour chacune des années 2000 et 2004 :

1. Calculez la médiane.
2. Calculez la moyenne et l'écart type.
3. Estimez l'évolution de la concentration de la branche, en construisant les courbes de Lorenz et en calculant les indices de Gini pour 2000 et 2004.

Corrigé

A - Pour l'année 2000, nous construisons le tableau suivant :

Classes	c_i	n_i	f_i (%)	F_i	f_ic_i	$f_ic_i^2$	f_i'	F_i'
[0,2 ; 0,5[	0,35	55	13,58	13,58	4,8	1,7	0,34	0,34
[0,5 ; 1[	0,75	65	16,05	29,63	12,0	9,0	0,87	1,21
[1 ; 2[	1,5	57	14,07	43,70	21,1	31,7	1,52	2,73
[2 ; 5[	3,5	72	17,78	61,48	62,2	217,8	4,49	7,22
[5 ; 10[	7,5	51	12,59	74,07	94,4	708,2	6,82	14,04
[10 ; 20[	15,0	36	8,89	82,96	133,4	2000,2	9,62	23,66
[20 ; 50[	35,0	41	10,13	93,09	354,6	12409,2	25,57	49,23
[50 ; 100[	75,0	13	3,21	96,30	240,8	18056,3	17,37	66,60
[100 ; 150]	125,0	15	3,70	100,00	462,5	57812,5	33,40	100,00
Total		405	100,00		1385,9	91246,6	100,00	

1. La médiane appartient à la classe médiane [2 ; 5[millions d'euros de C.A. La valeur de la médiane est calculée par interpolation linéaire.

$$M_e = 2 + 3 \cdot \frac{50 - 43,70}{17,78} = 3,06 \text{ millions d'euros}$$

2. La moyenne est de :

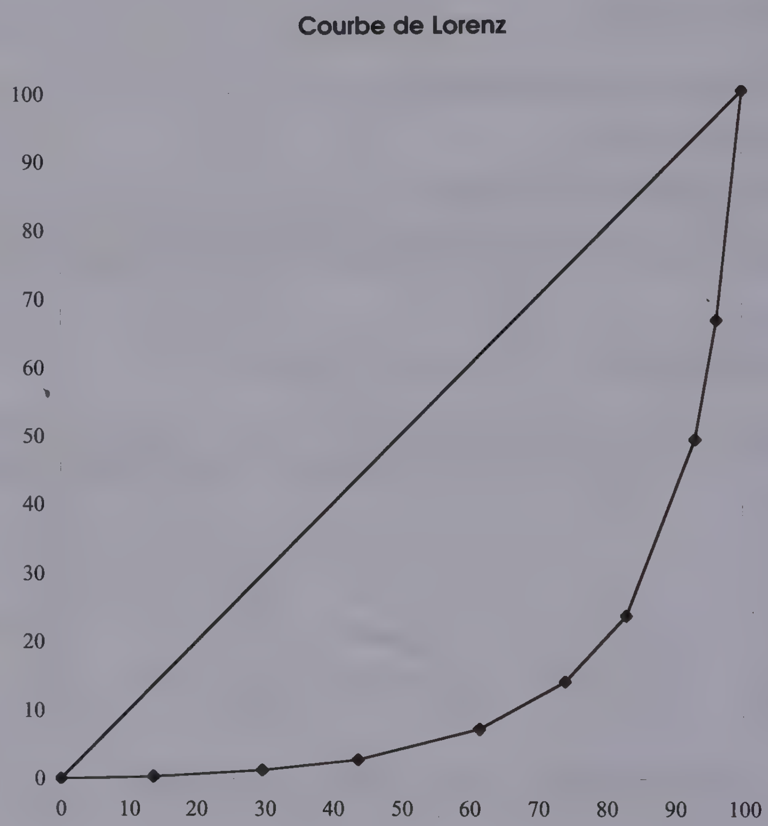
$$\bar{c} = \sum_{i=1}^{k} f_i c_i \cong 13{,}86 \text{ millions d'euros}$$

La variance :

$$V(c) = \sum_{i=1}^{k} f_i c_i^2 - \bar{c}^2 = 912{,}466 - (13{,}8588)^2 \cong 720{,}3966$$

L'écart type : $\sigma_c = \sqrt{V(c)} \cong 26{,}84$ millions d'euros

3.



L'indice de Gini est :

$$I_{G_{2000}} = 1 - \sum_{i=1}^{k} f_i (F'_{i-1} + F'_i) = 0{,}7041$$

B - Pour l'année 2004, le tableau devient :

Classes	c_i	n_i	f_i (%)	F_i	$f_i c_i$	$f_i c_i^2$	f_i'	F_i'
[0,2 ; 0,5[	0,35	40	10,34	10,34	3,6	1,3	0,20	0,20
[0,5 ; 1[	0,75	46	11,89	22,23	8,9	6,7	0,48	0,68
[1 ; 2[	1,50	55	14,21	36,44	21,3	32,0	1,15	1,83
[2 ; 5[	3,50	71	18,34	54,78	64,2	224,7	3,47	5,30
[5 ; 10[	7,50	49	12,66	67,44	95,0	712,2	5,13	10,43
[10 ; 20[	15,00	44	11,37	78,81	170,5	2558,1	9,21	19,64
[20 ; 50[	35,00	36	9,30	88,11	325,6	11395,3	17,58	37,22
[50 ; 100[	75,00	25	6,46	94,57	484,5	36337,5	26,15	63,37
[100 ; 150]	125,00	21	5,43	100,00	678,3	84786,3	36,63	100,00
Total		387	100,00		1851,9	136054,1	100,0	

1. La médiane appartient à la classe médiane [2 ; 5[. La valeur de la médiane est calculée par interpolation linéaire :

$$M_e = b_i + a_i \cdot \frac{50 - F_i}{F_{i+1} - F_i} = 2 + 3 \cdot \frac{50 - 36,44}{54,78 - 36,44} = 4,22$$

2. La moyenne est de :

$$\bar{c} = \sum_{i=1}^{k} f_i c_i \cong 18,5 \text{ millions d'euros}$$

La variance est de :

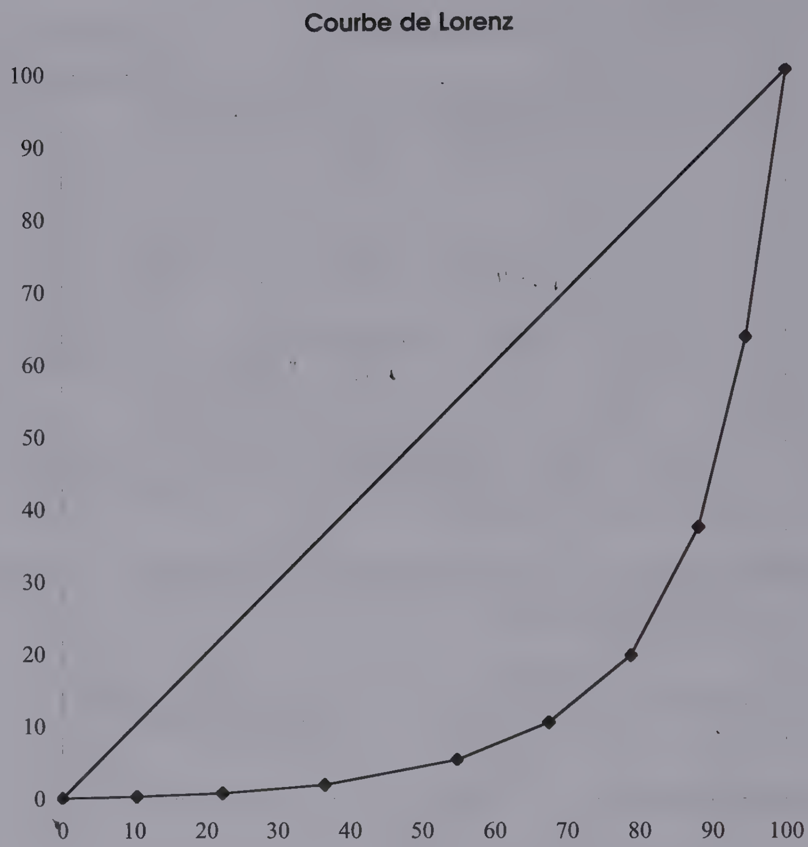
$$V(c) = \sum_{i=1}^{k} f_i c_i^2 - \bar{c}^2 = 1\,360,541 - (18,5193)^2 \cong 1\,017,58$$

L'écart type est de :

$$\sigma_c = \sqrt{V(c)} = 31,9 \text{ millions d'euros}$$

3. Indice de Gini et courbe de Lorenz :

$$I_{G_{2004}} = 1 - \sum_{i=1}^{k} f_i(F'_{i+1} + F'_i) = 0,72155$$



4. Évolution de la concentration

Entre 2000 et 2004, le C.A. des entreprises de l'édition de livres a crû sensiblement ; les valeurs prises par la médiane et la moyenne le montrent. Cette croissance du C.A. s'accompagne d'une plus grande dispersion, comme l'indique l'écart type, et se traduit par un accroissement de la concentration (mesuré par l'indice de Gini).

44. CONCENTRATION DES CA

Mots-clefs

C_x, indice d'Herfindahl

Énoncé

Le tableau ci-dessous donne la distribution des C.A. HT des entreprises d'une branche donnée.

Distribution des C.A. H.T.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
250	200	180	150	140	120	95	80	75	70	60	50	40	35	25	10

1. Calculez le C_4 de cette branche.
2. Calculez l'indice et le nombre équivalent d'Herfindahl ?
3. Si les pouvoirs publics imposent de ne pas augmenter l'indice de concentration de plus de 200, la fusion des entreprises 1 et 2 serait-elle possible ?

Corrigé

Pour calculer les coefficients demandés, il est nécessaire de connaître les parts de marché. Elles seront estimées par le C.A. H.T.

Le tableau statistique suivant permet de répondre aux questions.

Tableau des calculs préliminaires

	CA HT	m_i	$\sum_{i=1}^{k} m_i$	m_i^2
A	250	15,82	15,82	250,4
B	200	12,7	28,5	160,2
C	180	11,4	39,9	129,8
D	150	9,5	49,4	90,1
E	140	8,9	58,2	78,5
F	120	7,6	65,8	57,7
G	95	6,0	71,8	36,2
H	80	5,1	76,9	25,6
I	75	4,7	81,6	22,5
J	70	4,4	86,1	19,6
K	60	3,8	89,9	14,4
L	50	3,2	93,0	10,0
M	40	2,5	95,6	6,4
N	35	2,2	97,8	4,9
O	25	1,6	99,4	2,5
P	10	0,6	100,0	0,4
Total	1580	100,0		909,3

1. Le C_4 de cette branche correspond à la part de marché détenue par les quatre plus importantes entreprises de la branche.

$$C_4 = \sum_{i=A}^{D} m_i = 49{,}4$$

Il vaut donc 49,4 %.

2. L'indice d'Herfindahl est :

$$C_H = \sum_{i=A}^{P} m_i^2 = 900{,}309$$

Le nombre équivalent d'Herfindahl est :

$$\eta_H = \frac{1}{C_H} = \frac{1}{0{,}0909309} = 10{,}997 \cong 11$$

La distribution des parts de marché est équivalente à une structure de 11 entreprises équivalentes.

3. La fusion des deux premières entreprises A et B en une entreprise A' modifient l'indice d'Herfindahl de la manière suivante :

Tableau après fusion de A et B

	CA HT	m_i	m_i^2
A'	450	28,48	811,2
C	180	11,4	129,8
D	150	9,5	90,1
E	140	8,9	78,5
F	120	7,6	57,7
G	95	6	36,2
H	80	5,1	25,6
I	75	4,7	22,5
J	70	4,4	19,6
K	60	3,8	14,4
L	50	3,2	10
M	40	2,5	6,4
N	35	2,2	4,9
O	25	1,6	2,5
P	10	0,6	0,4
Total	1580	100	1309,9

L'indice d'Herfindahl augmente de 909,3 à 1309,9. L'augmentation est de 400,6 ; compte tenu du critère énoncé, cette fusion ne serait pas autorisée par les autorités de tutelle.

45. LE MARCHÉ DE L'AUTOMOBILE

Mots-clefs

Indice d'Herfindahl, structure de marché

Énoncé

Nous disposons des informations suivantes sur les plus grands groupes industriels de l'automobile pour l'année.

Distribution des chiffres d'affaires

	C.A. HT (milliards de F)
Général Motors	695
Ford Motor	502
Toyota Motor	427
Daemler-Benz	323
Nissan Motor	269
Volkswagen	259
Fiat	257
Honda Motor	184
Renault	166
Chrysler	166
P.S.A. Peugeot Citroën	160
Mitsubishi	130
Mazda	114
Robert Bosh	111

Source : TEF 93/94

Dans la suite nous supposerons que nous avons, dans le tableau, les chiffres de toutes les entreprises du secteur.

1. Calculez le C_4 selon les C.A.

2. Calculez le nombre équivalent d'Herfindahl pour le C.A. ; nous considérons que les C.A. sont représentatifs des parts de marché.

3. Caractérisez la structure du marché.

4. Si nous supposons une fusion entre Renault et Nissan d'une part et Daemler-Benz et Chrysler d'autre part, qu'elle serait l'indice d'Herfindahl et le nombre équivalent ?

Corrigé

Pour répondre aux questions, le plus simple est de construire le tableau statistique suivant :

	C.A. HT	m_i (%)	$\sum_{i=1}^{k} m_i$	m_i^2
Général Motors	695	18,5	18,5	341,1
Ford Motor	502	13,3	31,8	178,0
Toyota Motor	427	11,3	43,2	128,8
Daemler Benz	323	8,6	51,7	73,7
Nissan Motor	269	7,1	58,9	51,1
Volkswagen	259	6,9	65,8	47,4
Fiat	257	6,8	72,6	46,6
Honda Motor	184	4,9	77,5	23,9
Renault	166	4,4	81,9	19,5
Chrysler	166	4,4	86,3	19,5
PSA Peugeot-Citroën	160	4,3	90,6	18,1
Mistsubishi	130	3,5	94,0	11,9
Mazda	114	3,0	97,1	9,2
Robert Bosh	111	2,9	100,0	8,7
	3763	100,0		977,4

1. Le C_4 est la somme des parts de C.A. ou des effectifs des quatre premières entreprises. Le C_4 atteint 51,7 % pour le C.A. Les quatre premières entreprises assurent plus de la moitié du chiffre d'affaires du secteur.

2. Le nombre équivalent d'Herfindahl est obtenu de la manière suivante. On calcule la somme du carré des fréquences soit :

$$\sum_{i=1}^{n} m_i^2 = C_H$$

$$C_H = 997,4$$

Le nombre équivalent d'Herfindahl η_H (la comparaison ce fait par rapport à 10000) est l'inverse de C_H d'où :

$$\eta_H = \frac{10000}{977,4} = 10,2$$

Ce nombre décrit la structure du marché comme si environ η_H entreprises se partageaient également l'ensemble du chiffre d'affaires. Dans notre cas, le marché de l'automobile est structuré comme si 10 entreprises se partageaient le marché.

3. Ce marché est en situation de concurrence asymétrique. En effet, les quatre premières entreprises n'assurent qu'environ 50 % du C.A. et des effectifs. Il faut 9 et 8 entreprises sur 14 pour arriver à 80 % de ces variables.

4. La répartition des CA est différente nous obtenons le tableau suivant.

Nouvelle distribution

	C.A. HT	m_i	$\sum_{i=1}^{k} m_i$	$m_i^2 \cdot 10^4$
Général Motors	695	0,185	0,185	341,115
Ford Motor	502	0,133	0,318	177,967
Daemler-Benz-Chysler	489	0,130	0,448	168,869
Renault-Nissan	435	0,116	0,564	133,632
Toyota Motor	427	0,1133	0,677	128,762
Volkswagen	259	0,069	0,746	47,373
Fiat	257	0,068	0,814	46,644
Honda Motor	184	0,049	0,863	23,909
P.S.A. Peugeot Citroën	160	0,0437	0,906	18,079
Mitsubishi	130	0,035	0,940	11,935
Mazda	114	0,030	0,971	9,178
Robert Bosh	111	0,029	1,000	8,701
	3763	1		1116,164

Le classement des groupes se modifie, l'indice est de 116,164, il augmente de presque 139 points. Le nombre équivalent est cette fois de 9,0.

46. REVENUS IMPOSABLES

Mots-clefs

Mode, médiane, moyenne arithmétique, variance, écart type, courbe de Lorenz, indice de Gini, dissymétrie, variable continue

Énoncé

Vous disposez, pour l'année X, de la distribution des personnes ayant rempli une déclaration de revenus (IRPP) par tranche de revenu imposable.

Les tranches de revenus imposables

Tranche de revenu imposable (annuel en euros)	Nombre de déclarations (milliers)
Inférieur à 20 000 €	925
20 000 à moins de 40 000 €	1 665
40 000 à moins de 50 000 €	2 590
50 000 à moins de 60 000 €	3 145
60 000 à moins de 80 000 €	2 775
80 000 à moins de 110 000 €	2 405
110 000 à moins de 160 000 €	1 850
160 000 à moins de 240 000 €	1 480
240 000 à moins de 360 000 €	1 110
360 000 € et plus	555
Total	18 500

1. Représentez graphiquement cette distribution.
2. Déterminez le mode.
3. Calculez la médiane et le rapport inter décile ainsi que le coefficient de dispersion.
4. Calculez la médiale.
5. Calculez la moyenne arithmétique.
6. Calculez la variance et l'écart type.
7. Construisez la courbe de Lorenz et calculer l'indice de Gini.
8. Évaluez la symétrie de la distribution.

Corrigé

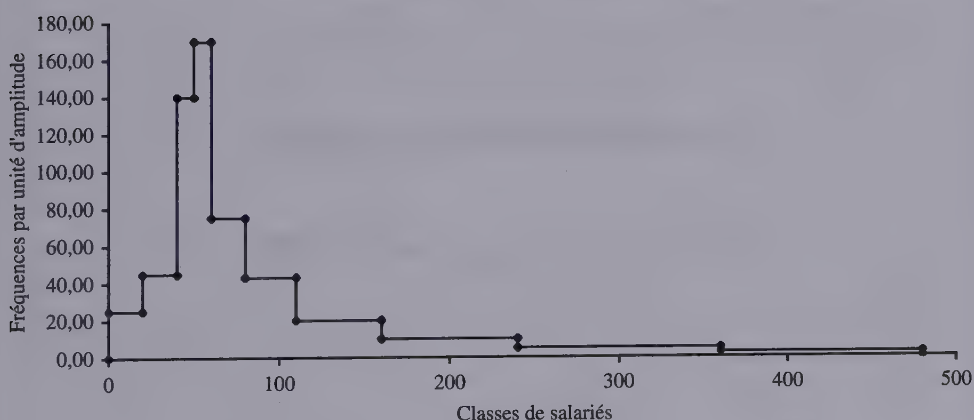
1. Pour construire l'histogramme nous devons disposer du tableau des fréquences par unité d'amplitude.

Tableau des calculs

Classes	Unité d'amplitude	n_i	f_i	$\frac{f_i}{a_i}$	F_i
[0 ; 20 000[	2	925	5,00	2,50	5,0
[20 000 ; 40 000[	2	1 665	9,00	4,50	14,0
[40 000 ; 50 000[	1	2 590	14,00	14,00	28,0
[50 000 ; 60 000[	1	3 145	17,00	17,00	45,0
[60 000 ; 80 000[	2	2 775	15,00	7,50	60,0
[80 000 ; 110 000[	3	2 405	13,00	4,33	73,0
[110 000 ; 160 000[	5	1 850	10,00	2,00	83,0
[160 000 ; 240 000[	8	1 480	8,00	1,00	91,0
[240 000 ; 360 000[	12	1 110	6,00	0,50	97,0
[360 000 ; 480 000]	12	555	3,00	0,25	100,0
Total		18 500	100,00		

Nous avons choisi une unité d'amplitude de 10 000. Les amplitudes de la première et de la dernière classe ont été choisies, faute d'informations complémentaires, conformément à la règle de la classe la plus proche (l'amplitude est identique à celle de la classe suivante respectivement de la classe précédente), la borne inférieure est zéro, la borne supérieure est 480 000.

Histogramme de la distribution



2. Ici la détermination directe du mode est impossible faute d'information sur la distribution des revenus imposables à l'intérieur de chaque classe. L'histogramme nous permet de déterminer la classe modale comme celle ayant la fréquence par unité d'amplitude la plus élevée. La classe modale est la classe des revenus imposables compris entre 50 000 € et 60 000 €.

Nous déterminons une estimation du mode selon la formule :

$$M_o = b_i + \frac{\Delta_i}{\Delta_i + \Delta_{i+1}} \cdot a_i$$

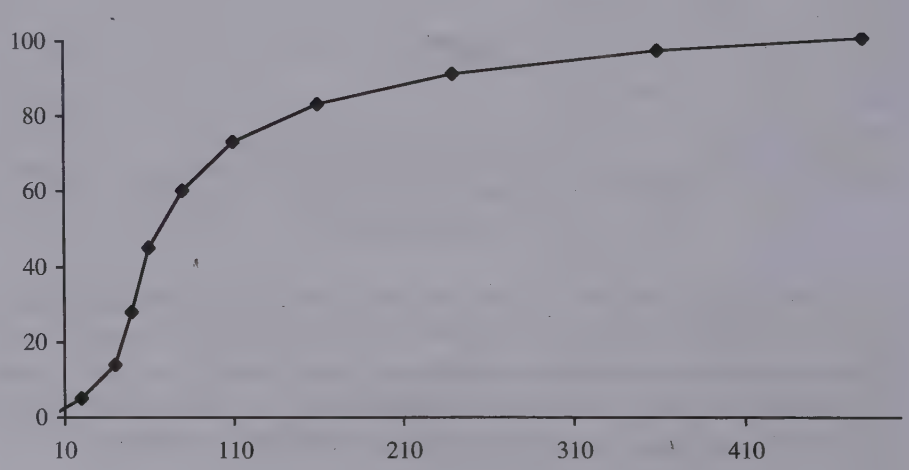
$$M_o = 50\,000 + 10\,000 \cdot \frac{3}{3+9{,}5} = 52\,400 \text{ €}$$

3. Les valeurs de la médiane et des quartiles s'obtiennent par interpolation linéaire à partir de la distribution des fréquences cumulées, donnée par le tableau ci-dessous.

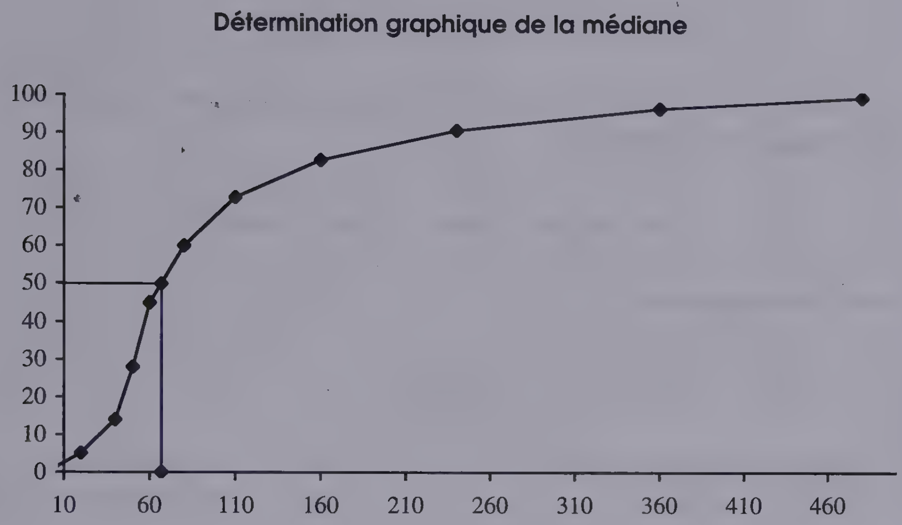
Tableau des fréquences

Classes	f_i	F_i
[0 ; 20 000[	5,0	5,0
[20 000 ; 40 000[	9,0	14,0
[40 000 ; 50 000[	14,0	28,0
[50 000 ; 60 000[	17,0	45,0
[60 000 ; 80 000[	15,0	60,0
[80 000 ; 110 000[	13,0	73,0
[110 000 ; 160 000[	10,0	83,0
[160 000 ; 240 000[	8,0	91,0
[240 000 ; 360 000[	6,00	97,0
[360 000 ; 480 000]	3,0	100,0
Total	100,0	

Courbe cumulative de la distribution



La détermination graphique permet de trouver un ordre de grandeur de la valeur de la médiane.



La forme générale du calcul de la médiane est :

$$M_e = b_i + a_i \cdot \frac{50 - F_{i-1}}{F_i + F_{i-1}}$$

La médiane (M_e) appartient à la tranche de revenus imposables de 60 000 € à moins de 80 000 €.

$$M_e = 60\,000 + 20\,000 \cdot \frac{5}{15} = 66\,666{,}66 \cong 66\,667 \text{ €}$$

La forme générale du calcul du premier quartile est :

$$Q_1 = b_i + a_i \cdot \frac{25 - F_{i-1}}{F_i + F_{i-1}}$$

Le premier quartile (Q_1) appartient à la classe de revenus imposables de 40 000 à moins de 50 000 €.

$$Q_1 = 40\,000 + 10\,000 \cdot \frac{11}{14} = 47\,857{,}14 \cong 47\,857 \text{ €}$$

La forme générale du calcul du troisième quartile est :

$$Q_3 = b_i + a_i \cdot \frac{75 - F_{i-1}}{F_i + F_{i-1}}$$

Le troisième quartile (Q_3) appartient à la classe des revenus imposables de 110 000 à moins de 160 000 €.

$$Q_3 = 110\,000 + 50\,000 \cdot \frac{2}{10} = 120\,000 \text{ €}$$

Le rapport interquartile est :

$$\frac{Q_3}{Q_1} = 2{,}51$$

Le coefficient de dispersion est :

$$CD_{is} = \frac{Q_3 - Q_1}{M_e} = 1{,}082$$

Le tableau des fréquences permet de calculer les déciles en recourrant à une méthode d'interpolation linéaire.

Le premier décile appartient à la classe 2.

$$D_1 = b_i + a_i \cdot \frac{10 - F_{i-1}}{F_i + F_{i-1}} = 20\,000 + 20\,000 \cdot \frac{10 - 5}{9} = 31\,111$$

Le neuvième décile appartient à la classe 8.

$$D_9 = b_i + a_i \cdot \frac{90 - F_{i-1}}{F_i + F_{i-1}} = 160\,000 + 80\,000 \cdot \frac{90 - 83}{8} = 230\,000$$

Le rapport inter décile est donc :

$$\frac{D_9}{D_1} = \frac{230\,000}{31\,111} \cong 7{,}4$$

4. Nous devons construire un tableau faisant apparaître les fréquences de la variable, ce tableau nous permettra également de calculer la moyenne.

Tableau du calcul de la médiale

Classes	a_i	c_i	n_i	f_i	f_ic_i	f'_i	F'_i
[0 ; 20 000[	20	10,0	925	5,0	50,0	0,5	0,5
[20 000 ; 40 000[	20	30,0	1 665	9,0	270,0	2,7	3,1
[40 000 ; 50 000[	10	45,0	2 590	14,0	630,0	6,2	9,3
[50 000 ; 60 000[	10	55,0	3 145	17,0	935,0	9,2	18,5
[60 000 ; 80 000[	20	70,0	2 775	15,0	1050,0	10,3	28,8
[80 000 ; 110 000[	30	95,0	2 405	13,0	1235,0	12,1	41,0
[110 000 ; 160 000[	50	135,0	1 850	10,0	1350,0	13,3	54,2
[160 000 ; 240 000[	80	200,0	1 480	8,0	1600,0	15,7	69,9
[240 000 ; 360 000[	120	300,0	1 110	6,0	1800,0	17,7	87,6
[360 000 ; 480 000]	120	420,0	555	3,0	1260,0	12,4	100,0
Total			18500	100,0	10180,0	100,0	

La médiale appartient à la classe [110 000 ; 160 000[, le calcul de la médiale est formellement identique à celui de la médiane, il ne s'agit que de la médiane des valeurs.

$$M_l = b_i + a_i \cdot \frac{50 - F'_i}{F'_{i+1} - F'_i}$$

$$M_l = 110\,000 + 50\,000 \cdot \frac{50 - 40{,}8}{13{,}2} = 144\,630$$

5. Nous pouvons utiliser les résultats du tableau précédent. La moyenne arithmétique s'obtient immédiatement à partir du tableau.

$$\overline{c} = \sum_{i=1}^{k} f_i c_i = \frac{10180}{100} = 101{,}8$$

Le revenu imposable moyen est de 101 800 euros.

6. Nous devons construire un nouveau tableau.

Nous pouvons calculer la variance

$$V(c) = \sum_{i=1}^{k} f_i c_i^2 - \overline{c}^2 = \sum_{i=1}^{k} f_i c_i^2 - \left(\sum_{i=1}^{k} f_i c_i\right)^2$$

$$V(c) = 18506{,}5 - (101{,}80)^2 = 8143{,}26$$

Tableau du calcul de la variance

Classes	c_i	f_i (en %)	$f_i c_i$	$f_i c_i^2$
[0 ; 20[	10,0	5,0	50,0	500,0
[20 ; 40[	30,0	9,0	270,0	8100,0
[40 ; 50[	45,0	14,0	630,0	28350,0
[50 ; 60[	55,0	17,0	935,0	51425,0
[60 ; 80[	70,0	15,0	1050,0	73500,0
[80 ; 110[	95,0	13,0	1235,0	117325,0
[110 ; 160[	135,0	10,0	1350,0	182250,0
[160 ; 240[	200,0	8,0	1600,0	320000,0
[240 ; 360[	300,0	6,0	1800,0	540000,0
[360 ; 480]	420,0	3,0	1260,0	529200,0
Total		100,0	10180,0	1850650,0

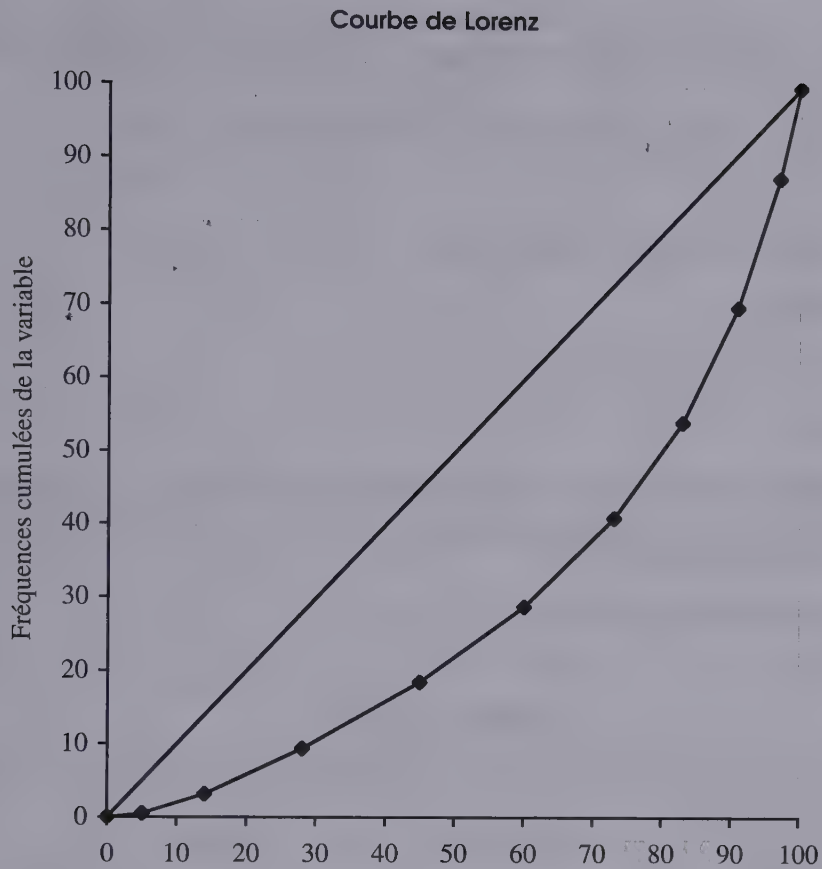
L'écart type est la racine de la variance $\sigma_c = \sqrt{V(c)}$.

$$\sigma_c = 90\,240 \text{ €}$$

7. La courbe de Lorenz est une représentation graphique des fréquences cumulées des effectifs (F_i) et des fréquences cumulées de la variable (F_i').

Tableau des fréquences cumulées

Classes	F_i (en %)	F_i' (en %)	f_i (en %)	$f_i \cdot (F_{i-1}' + F_i')$
[0 ; 20[	5,0	0,5	5,0	2,5
[20 ; 40[	14,0	3,1	9,0	32,7
[40 ; 50[	28,0	9,3	14,0	174,7
[50 ; 60[	45,0	18,5	17,0	473,4
[60 ; 80[	60,0	28,8	15,0	710,2
[80 ; 110[	73,0	41,0	13,0	907,3
[110 ; 160[	83,0	54,2	10,0	951,9
[160 ; 240[	91,0	69,9	8,0	993,3
[240 ; 360[	97,0	87,6	6,0	945,4
[360 ; 480]	100,0	100,0	3,0	562,9
Total			100,0	5754,2



L'indice de Gini se calcule à partir des fréquences et des fréquences cumulées selon la formule suivante :

$$I_G = 1 - \sum_{i=1}^{k} f_i \cdot (F'_{i-1} + F'_i).$$

Dans le tableau, les calculs ont été réalisés en pourcentage, il est nécessaire de diviser le résultat obtenu par 10000.

$$I_G = 1 - 0{,}5742 = 0{,}4258 \cong 0{,}43$$

8. Il existe plusieurs méthodes pour évaluer la symétrie d'une distribution. Ici nous utiliserons les indicateurs calculés antérieurement, le mode, la médiane et la moyenne arithmétique. Le mode est inférieur à la médiane, elle-même inférieure à la moyenne arithmétique : $M_o < M_e < \overline{c}$. La distribution est donc étalée vers la droite.

Le coefficient de Yule donne la même indication.

$$s=\frac{(Q_3-M_e)-(M_e-Q_1)}{(Q_3-M_e)+(M_e-Q_1)}=\frac{(120\,000-66\,667)-(66\,667-47\,857)}{(120\,000-66\,667)+(66\,667-47\,857)}\cong 0{,}48$$

Le résultat est positif, il indique que la distribution est étalée à droite.

47. MESURE DE LA SATISFACTION

Mots-clefs

Variable qualitative ordonnée

Énoncé

Lors d'une enquête par questionnaire, les personnes interrogées ont eu à répondre à la question : « Quel est votre degré de satisfaction par rapport au professionnalisme du stagiaire ? »

Le dépouillement conduit au tableau suivant.

Tableau des résultats

Professionnalisme	Nombre de citations	Fréquences
Pas du tout satisfait	0	0.0
Pas satisfait	10	28.57
Passablement satisfait	12	34.28
Correctement satisfait	10	28.57
Satisfait	3	8.57
Très satisfait	0	0.00
Total	35	100.00

Pour le traitement des données, les catégories sont post-codées sur une notation de 1 « pas du tout satisfait » à 6 « très satisfait ».

1. Quels sont le mode, la médiane, la moyenne et l'écart type de cette série ?

2. Donnez l'interprétation de chacune des caractéristiques.

Corrigé

Un tableau des calculs est préalable à la réponse aux questions

Tableau des calculs

Professionnalisme	Code	Nombre de citations			Pourcentages
	x_i	n_i	$n_i x_i$	$n_i x_i^2$	p_i
Pas du tout satisfait	1	0	0	0	0.0
Pas satisfait	2	10	20	40	28.6
Passablement satisfait	3	12	36	108	34.3
Correctement satisfait	4	10	40	160	28.6
Satisfait	5	3	15	75	8.5
Très satisfait	6	0	0	0	0.0
Total		35	111	383	100.0

1. Le mode est la modalité « passablement satisfait », c'est-à-dire le code 3.

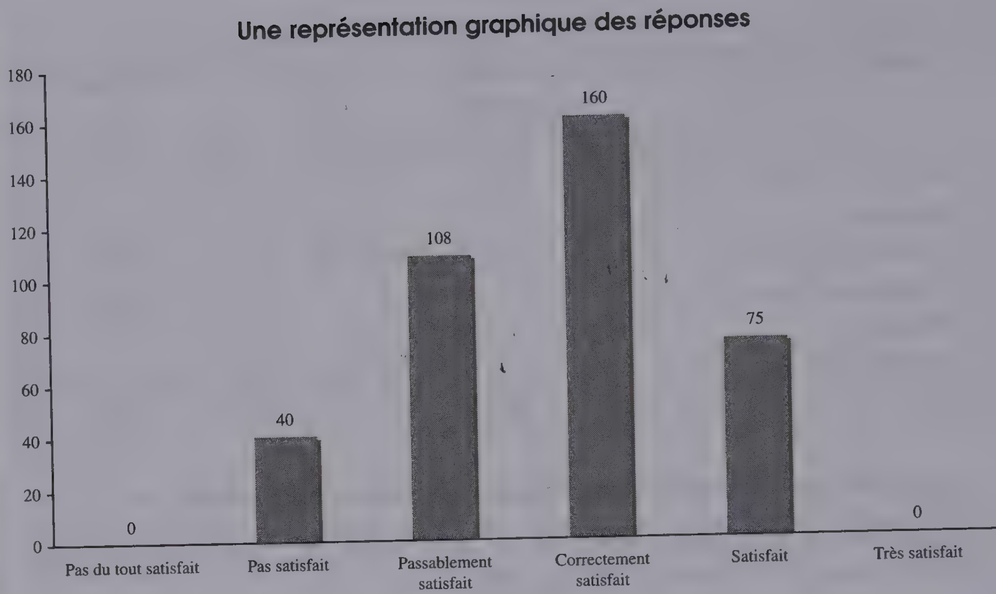
La médiane au sens strict n'existe pas ; au sens large disons qu'elle se situe entre « pas satisfait » et « passablement satisfait ».

La moyenne est de 3,17. Bien sûr, c'est un code qui n'existe pas. Ce chiffre signifie que si les 35 personnes avaient donné le même avis, elles auraient été un peu plus que « passablement satisfaites ».

L'écart type est de 0,94. Il nous permet de dire, qu'en moyenne, on s'écarte de moins d'une modalité de la moyenne.

2. Les caractéristiques d'une telle série ne peuvent être qu'analysées sur un mode qualitatif.

Ces dernières caractéristiques sont parfois utilisées pour des calculs ultérieurs. Il faut néanmoins remarquer qu'un écart de 1 point entre « passablement satisfait » et « correctement satisfait » n'est pas équivalent à un écart de 1 point entre « satisfait » et « très satisfait » et donc que les calculs de la moyenne et de l'écart type ne sont théoriquement pas valides.



48. EFFET DE LA RÉPARTITION EN CLASSES

Mots-clefs

Moyenne arithmétique, effets de structure, variable continue, mode, médiane, quartile, dissymétrie

Énoncé

L'enquête sur l'emploi de mars 2002 donne des informations sur la structure de la population active par classe d'âge.

1. Pour la population active de 15 à 75 ans et plus, regroupée en 13 classes.
   a. Construisez l'histogramme de la distribution.
   b. Déterminez le mode.
   c. Calculez la médiane et les quartiles de la distribution de la population active et donner une indication sur la dissymétrie de cette distribution.
   d. Calculez l'âge moyen des actifs.
2. Nous retenons maintenant une répartition en 5 classes et nous calculons les mêmes caractéristiques que pour la distribution en 13 classes.
   a. Construisez l'histogramme de la distribution.
   b. Déterminez le mode.
   c. Calculez la médiane et les quartiles de la distribution de la population active et donnez une indication sur la dissymétrie de cette distribution.

d. Calculez l'âge moyen des actifs.

3. Interprétez les différences entre les deux séries de résultats.

Répartition de la population active selon l'âge

Groupes d'âge	Ensemble (en milliers)
15 -19 ans	330
20 -24 ans	1 971
25 -29 ans	3 185
30 -34 ans	3 642
35 -39 ans	3 763
40 - 44 ans	3 744
45 - 49 ans	3 622
50 - 54 ans	3 429
55 -59 ans	2 058
60 - 64 ans	413
65 - 69 ans	73
70 - 74 ans	27
75 ans et plus	25

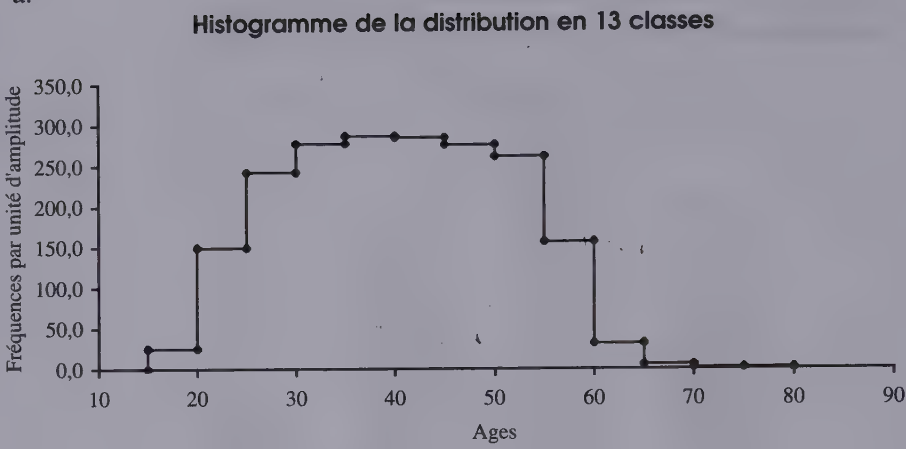
Corrigé

1. Nous construisons le tableau statistique dans le cas d'une distribution à 13 classes.

Tableau des calculs

Groupes d'âge	a_i	c_i	n_i	f_i	$\frac{f_i}{a_i}$ x100	F_i	$f_i c_i$
[15 ; 20[	5	17,5	330	1,3	25,1	1,3	22,0
[20 ; 25[	5	22,5	1971	7,5	149,9	8,8	168,7
[25 ; 30[	5	27,5	3185	12,1	242,4	20,9	333,3
[30 ; 35[	5	32,5	3642	13,9	277,2	34,7	450,4
[35 ; 40[	5	37,5	3763	14,3	286,4	49,0	536,9
[40 ; 45[	5	42,5	3744	14,2	284,9	63,3	605,5
[45 ; 50[	5	47,5	3622	13,8	275,6	77,1	654,6
[50 ; 55[	5	52,5	3429	13,0	260,9	90,1	685,0
[55 ; 60[	5	57,5	2058	7,8	156,6	98,0	450,3
[60 ; 65[	5	62,5	413	1,6	31,4	99,5	98,2
[65 ; 70[	5	67,5	73	0,3	5,6	99,8	18,7
[70 ; 75[	5	72,5	27	0,1	2,1	99,9	7,4
[75 ; 80[	5	77,5	25	0,1	1,9	100,0	7,4
Total			26282	100,0			4038,4

a.



b. L'histogramme nous permet de déterminer la classe modale, il s'agit de la classe [35 ; 39].

Nous pouvons calculer le mode.

$$M_o = b_i + \frac{\Delta_i}{\Delta_i + \Delta_{i+1}} \cdot a_i = 35 + 5 \cdot \frac{9{,}21}{9{,}21 + 1{,}45} = 39{,}3 \text{ ans}$$

c. L'âge médian se détermine comme les quartiles par interpolation linéaire.

$$M_e = b_i + a_i \cdot \frac{50 - f_{i-1}}{F_i - F_{i-1}} = 40 + 5 \cdot \frac{50 - 49{,}0}{14{,}2} = 40{,}3 \text{ ans}$$

Calcul du premier quartile :

$$Q_1 = b_i + a_i \cdot \frac{25 - F_{i-1}}{F_i + F_{i-1}} = 30 + 5 \cdot \frac{25 - 20{,}9}{13{,}9} = 31{,}5 \text{ ans}$$

Calcul du troisième quartile :

$$Q_3 = b_i + a_i \cdot \frac{75 - F_{i-1}}{F_i + F_{i-1}} = 45 + 5 \cdot \frac{75 - 63{,}3}{13{,}8} = 49{,}2 \text{ ans}$$

L'indicateur s de dissymétrie indique une distribution presque symétrique, avec un faible étalement vers la droite.

$$s = \frac{(Q_3 - M_e) - (M_e - Q_1)}{(Q_3 - M_e) + (M_e - Q_1)} = \frac{(49{,}2 - 40{,}3) - (40{,}3 - 31{,}5)}{(49{,}2 - 40{,}3) + (40{,}3 - 31{,}5)} = 0{,}004$$

d. L'âge moyen de la distribution de la population active est obtenu par lecture directe du tableau, l'âge moyen de la population active est de :

$$\overline{c}_{13} = \sum_{i=1}^{m} f_i c_i \cong 40{,}38 \text{ ans}$$

2. Nous réduisons le nombre de classes de la population à cinq. Nous obtenons une nouvelle distribution et le tableau associé.

Nouvelle distribution statistique en 5 classes

Classes d'âge	n_i
[15 ; 25[	2 301
[25 ; 40[	10 590
[40 ; 50[	7 366
[50 ; 60[	5 487
[60 ; 80]	538
Total	26 282

a.

Tableau statistique de la distribution en 5 classes

Classes d'âge	a_i	c_i	n_i	f_i	$\frac{f_i}{a_i}$x100	F_i	$f_i c_i$
[15 ; 25[	10	20,0	2301	8,8	87,6	8,8	175,1
[25 ; 40[	15	32,5	10590	40,3	268,6	49,0	1309,5
[40 ; 50[	10	45,0	7366	28,0	280,3	77,1	1261,2
[50 ; 60[	10	55,0	5487	20,9	208,8	98,0	1148,3
[60 ; 80]	20	70,0	538	2,0	10,2	100,0	143,3
Total			26282	100,0			4037,4

b. L'histogramme nous permet de déterminer la classe modale, il s'agit de la classe [40 ; 50[.

Nous pouvons calculer le mode.

$$M_o = b_i + \frac{\Delta_i}{\Delta_i + \Delta_{i+1}} \cdot a_i = 40 + 10 \cdot \frac{11{,}64}{11{,}64 + 71{,}49} = 41{,}4 \text{ ans}$$

c. La médiane se détermine comme les quartiles par interpolation linéaire.

$$M_e = b_i + a_i \cdot \frac{50 - F_{i-1}}{F_i - F_{i-1}} = 40 + 10 \cdot \frac{50 - 49{,}0}{28{,}0} = 40{,}3 \text{ ans}$$

$$Q_1 = b_i + a_i \cdot \frac{25 - F_{i-1}}{F_i + F_{i-1}} = 25 + 15 \cdot \frac{25 - 8,8}{40,3} = 31,0 \text{ ans}$$

$$Q_3 = b_i + a_i \cdot \frac{75 - F_{i-1}}{F_i + F_{i-1}} = 40 + 10 \cdot \frac{75 - 49}{28} = 49,3 \text{ ans}$$

L'indicateur s de dissymétrie indique une distribution presque symétrique, avec un faible étalement vers la gauche.

$$s = \frac{(Q_3 - M_e) - (M_e - Q_1)}{(Q_3 - M_e) + (M_e - Q_1)} = \frac{(49,3 - 10,3) - (40,3 - 31,0)}{(49,3 - 10,3) + (40,3 - 31,0)} = -0,02$$

d. L'âge moyen de la distribution de la population active est obtenu par lecture directe du tableau, l'âge moyen de la population active est de :

$$\bar{c}_5 = \sum_{i=1}^{k} f_i.c_i = 40,37 \text{ ans}$$

3.

Un tableau pour comparer les résultats des deux distributions

	M_o	M_e	Q_1	Q_3	$\bar{c}$	s
13 classes	39,3	40,3	31,5	49,2	40,38	0,004
5 classes	41,4	40,3	31,0	49,3	40,37	– 0,02
Écart	– 2,1	0	0,5	– 0,1	0,01	0,024

Il est clair que le seul écart qui dépasse l'unité est entre les modes. Les autres indicateurs sont très proches, souvent de l'ordre de l'arrondi. L'effet d'une modification des classes garde, pour cette distribution, les valeurs des caractéristiques sauf au plan de la représentation graphique, les histogrammes donnent deux images distinctes et les modes ont des évaluations sensiblement différentes.

49. LES POPULATIONS PÉNALES

Mots-clefs

Histogramme, moyenne arithmétique, mode, médiane, quartiles, variance, écart type

Énoncé

La répartition par âge et par sexe de la population pénale en France est la suivante au premier janvier 1981.

Population pénale

Tranches d'âge	Hommes	Femmes
de 15 à 18 ans	897	37
de 18 à 21 ans	5 359	155
de 21 à 25 ans	8 646	267
de 25 à 30 ans	8 623	241
de 30 à 40 ans	8 394	317
de 40 à 50 ans	3 959	138
de 50 à 60 ans	1 404	74
de 60 à 70 ans	429	17

1. Pour chacune des deux distributions :
   a. Construisez les histogrammes.
   b. Déterminez les classes modales ou les modes.
   c. Calculez les médianes et les quartiles.
   d. Calculez l'âge moyen et les écarts-types.
2. À partir de tous les résultats que vous avez calculés, comparez et commentez les caractéristiques des deux populations.
3. En utilisant les résultats précédents, déterminez les caractéristiques de la population totale.

Corrigé

1. a. Pour construire l'histogramme, nous devons calculer les fréquences compte tenu de l'amplitude des classes ; les deux tableaux suivants fournissent les informations nécessaires.

Pour les hommes

Tranches d'âges	Effectifs	Fréquences	Fréquences par unité d'amplitude
	n_i	f_i	$\frac{f_i}{a_i}$
[15 ; 18[	897	2,4	79,3
[18 ; 21[	5359	14,2	473,7
[21 ; 25[	8646	22,9	573,2
[25 ; 30[	8623	22,9	457,3
[30 ; 40[	8394	22,3	222,6
[40 ; 50[	3959	10,5	105,0
[50 ; 60[	1404	3,7	37,2
[60 ; 70]	429	1,1	11,4
	37711	100,0	

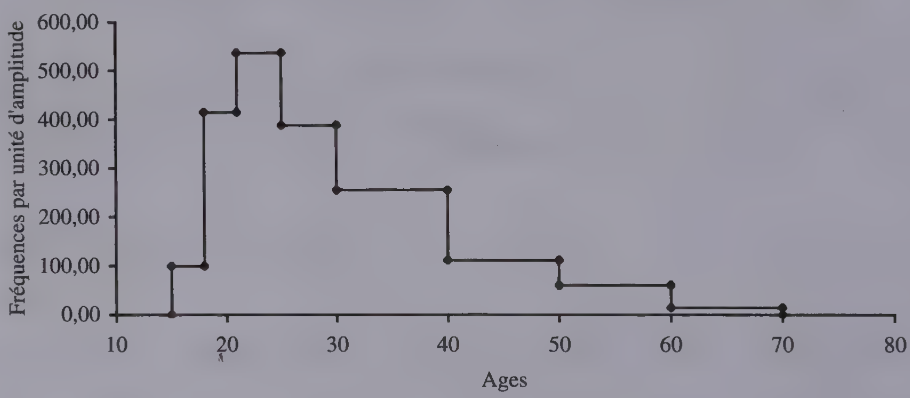
Pour les femmes

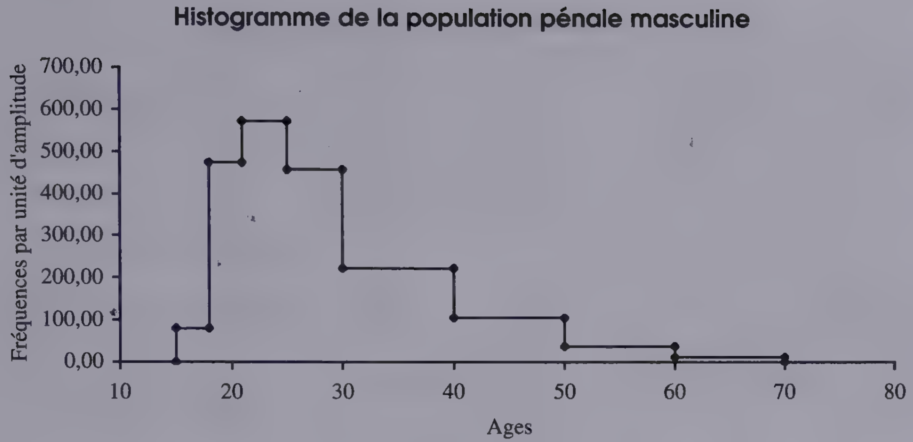
Tranches d'âges	Effectifs	Fréquences	Fréquences par unité d'amplitude
	n_i	f_i	$\frac{f_i}{a_i}$
[15 ; 18[	37	3,0	1,0
[18 ; 21[	155	12,4	4,1
[21 ; 25[	267	21,4	5,4
[25 ; 30[	241	19,4	3,9
[30 ; 40[	317	25,4	2,5
[40 ; 50[	138	11,1	1,1
[50 ; 60[	74	5,9	0,6
[60 ; 70]	17	1,4	0,1
	1246	100,0	

La colonne des fréquences par unité d'amplitude est obtenue de la manière suivante : $100 \cdot \dfrac{f_i}{a_i}$ ou a_i est l'amplitude de la classe.

Les histogrammes de la distribution sont représentés ci-dessous.

Histogramme de la population pénale féminine





b. L'histogramme et la colonne $\frac{f_i}{a_i}$ du tableau des hommes indiquent la classe modale [21 ; 25[.

Le mode est :

$$M_{oh} = b_i + \frac{\Delta_i}{\Delta_i + \Delta_{i+1}} \cdot a_i = 21 + 4 \cdot \frac{99{,}48}{99{,}48 + 115{,}85} = 22{,}8 \text{ ans}$$

L'histogramme et la colonne $\frac{f_i}{a_i}$ du tableau des femmes, indiquent la classe modale qui est la classe [21 ; 25[.

$$M_{of} = b_i + \frac{\Delta_i}{\Delta_i + \Delta_{i+1}} \cdot a_i = 1 + 4 \cdot \frac{121{,}05}{121{,}05 + 148{,}88} = 22{,}8 \text{ ans}$$

c. Les autres caractéristiques seront calculées à l'aide des tableaux statistiques suivants :

Population pénale masculine

Tranches d'âges	c_i	f_i	F_i	$f_i c_i$	$f_i c_i^2$
[15 ; 18[	16,5	2,4	2,4	39,2	647,6
[18 ; 21[	19,5	14,2	16,6	277,1	5403,6
[21 ; 25[	23,0	22,9	39,5	527,3	12128,4
[25 ; 30[	27,5	22,9	62,4	628,8	17292,4
[30 ; 40[	35,0	22,3	84,6	779,1	27267,0
[40 ; 50[	45,0	10,5	95,1	472,4	21259,0
[50 ; 60[	55,0	3,7	98,9	204,8	11262,2
[60 ; 70]	65,0	1,1	100,0	73,9	4806,4
Total		100,0		3002,7	100066,5

Population pénale féminine

Tranches d'âges	c_i	f_i	F_i	$f_i c_i$	$f_i c_i^2$
[15 ; 18[	16,5	3,0	3,0	49,0	808,4
[18 ; 21[	19,5	12,4	15,4	242,6	4730,2
[21 ; 25[	23,0	21,4	36,8	492,9	11335,7
[25 ; 30[	27,5	19,3	56,2	531,9	14627,3
[30 ; 40[	35,0	25,4	81,6	890,4	31165,7
[40 ; 50[	45,0	11,1	92,7	498,4	22427,8
[50 ; 60[	55,0	5,9	98,6	326,6	17965,5
[60 ; 70]	65,0	1,4	100,0	88,7	5764,4
Total		100,0		3120,5	108825,1

La colonne des fréquences cumulées (F_i) permet de déterminer la classe médiane ; soit pour les hommes la classe 25 à 30 ans, la médiane se calcule par interpolation linéaire.

$$M_{eh} = b_i + a_i \cdot \frac{50 - F_{i-1}}{F_i - F_{i-1}} = 25 + 5 \cdot \frac{50 - 39{,}5}{22{,}9} \cong 27{,}3 \text{ ans}$$

Pour les femmes la classe médiane est aussi 25 à 30 ans, mais conduit à une médiane de :

$$M_{ef} = b_i + a_i \cdot \frac{50 - F_{i-1}}{F_i - F_{i-1}} = 25 + 5 \cdot \frac{50 - 36{,}8}{19{,}3} = 28{,}4 \text{ ans}$$

Les quartiles se déterminent de manière analogue aux médianes :

– pour la population pénale masculine :

$$Q_{1h} = b_i + a_i \cdot \frac{25 - F_{i-1}}{F_i + F_{i-1}} = 21 + 4 \cdot \frac{25 - 16{,}6}{22{,}9} = 22{,}5 \text{ ans}$$

$$Q_{3h} = b_i + a_i \cdot \frac{75 - F_{i-1}}{F_i + F_{i-1}} = 30 + 10 \cdot \frac{75 - 62{,}4}{22{,}3} = 35{,}7 \text{ ans}$$

– pour la population pénale féminine :

$$Q_{1f} = b_i + a_i \cdot \frac{25 - F_{i-1}}{F_i + F_{i-1}} = 21 + 4 \cdot \frac{25 - 15{,}4}{21{,}4} = 22{,}8 \text{ ans}$$

$$Q_{3f} = b_i + a_i \cdot \frac{75 - F_{i-1}}{F_i + F_{i-1}} = 30 + 10 \cdot \frac{75 - 56{,}2}{25{,}4} = 37{,}4 \text{ ans}$$

d. L'âge moyen de la population pénale masculine, $\overline{c}_h$, s'obtient à partir du total de la colonne $f_i c_i$ d'après la formule :

$$\overline{c}_h = \sum_{i=1}^{k} f_i.c_i = \frac{3002,7}{100} \cong 30 \text{ ans}$$

$$\overline{c}_f = \sum_{i=1}^{k} f_i.c_i = \frac{3120,5}{100} \cong 31,2 \text{ ans}$$

e. Calculs des écarts-types

– pour la population masculine

$$V(c_h) = \sum_{i=1}^{k} f_i.c_i^2 - \overline{c}_h^2 = 1000,665 - (30,027)^2 = 99,0554$$

$$\sigma_{c_h} = \sqrt{V(c_h)} = 9,95 \text{ ans}$$

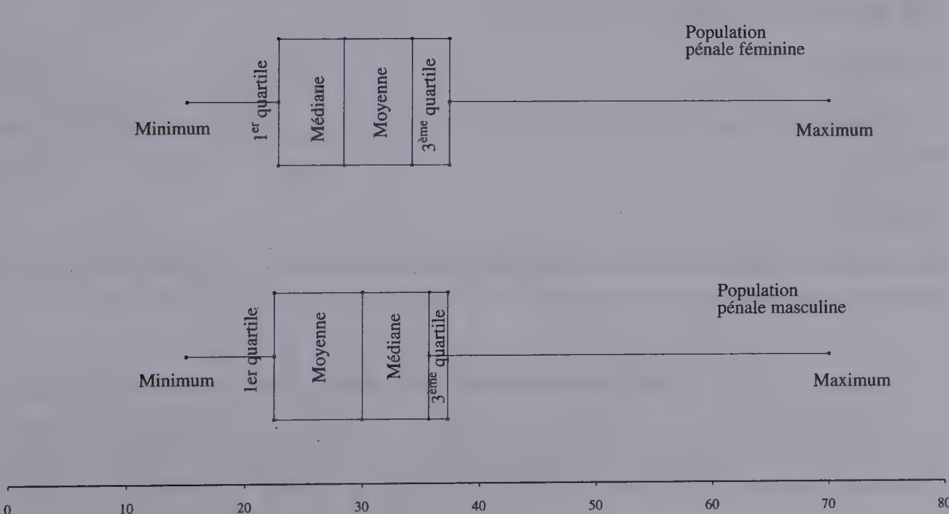
– pour la population féminine

$$V(c_f) = \sum_{i=1}^{k} f_i.c_i^2 - \overline{c}_f^2 = 1088,251 - (31,205)^2 = 114,496$$

$$\sigma_{c_f} = \sqrt{V(c_f)} = 10,7 \text{ ans}$$

2. La population pénale féminine est tout à fait marginale dans la population pénale totale. Cette population est plus âgée en moyenne que la population masculine, avec une dispersion de la distribution plus importante.

Graphique de comparaison des deux distributions



3. L'âge moyen de la population pénale totale s'obtient en pondérant les moyennes des populations des hommes et des femmes par leur importance relative dans la population pénale totale.

Soit N la population pénale totale :

$$N = 1\,246 + 37\,711 = 38\,957$$

Soit f_h la proportion relative des hommes :

$$f_h = \frac{37711 \cdot 100}{38957} = 96{,}8$$

Soit f_f la proportion relative des femmes :

$$f_f = 100 - 96{,}8 = 3{,}2$$

Alors la moyenne d'âge de la population pénale est :

$$\bar{c} = \sum_{j=1}^{2} f_j \bar{c}_j = 0{,}968 \cdot 30{,}0 + 0{,}032 \cdot 31{,}2 \cong 30{,}04 \text{ ans.}$$

La variance et l'écart type :

$$V(c) = \left(f_h V(c_h) + f_f V(c_f)\right) + \left(f_h \bar{c}_h^2 + f_f \bar{c}_f^2 - \bar{c}^2\right).$$

$$V(c) = (0{,}968 \cdot 99{,}0554 + 0{,}032 \cdot 114{,}496)$$
$$+ \left(0{,}968 \cdot 30{,}027^2 + 0{,}032 \cdot 31{,}205 - 30{,}01^2\right) = 101{,}07683$$

$$\sigma_c = \sqrt{V(c)} = 10{,}05.$$

50. SURFACES AGRICOLES UTILES

Mots-clefs

Fréquences, histogramme, médiane, mode, moyenne, courbe de Lorenz, indice de Gini

Énoncé

Le tableau ci-dessous donne, pour une région quelconque, la distribution des exploitations agricoles en fonction de la surface agricole utile (S.A.U.).

Distribution des exploitations (en ha)

Classes de SAU	moins de 5	de 5 à 10	de 10 à 20	de 20 à 50	de 50 à 100	plus de 100
Nombre d'exploitations	7 800	9 600	13 200	20 400	7 200	1 800

Nous savons d'autre part que les exploitations de plus de 100 ha recouvrent une superficie totale de 243 000 ha.

1. Construisez le tableau des fréquences et l'histogramme de cette distribution.
2. Calculez la surface médiane.
3. Déterminez la classe modale et le mode.
4. Calculez la surface moyenne.
5. La SAU totale de la région A est de 1 680 000 ha. En déduire la surface moyenne des exploitations. Comparez ce résultat avec celui trouvé précédemment.
6. Le tableau ci-dessous décrit la répartition de la SAU en fonction des différentes classes.
   a. Calculez la médiane de cette distribution.
   b. Construisez la courbe de Lorenz et calculez le coefficient de Gini.

Distribution des surfaces (en ha)

Classes de S.A.U.	moins de 5	de 5 à 10	de 10 à 20	de 20 à 50	de 50 à 100	plus de 100
% de la S.A.U.	1	3	12	39	29	16

Corrigé

1. Le tableau des fréquences de la distribution se présente comme suit.

Classes de SAU (ha)	n_i	f_i	a_i	$100 \cdot \frac{f_i}{a_i}$
[0 ; 5[	7800	13,0	5	260,00
[5 ; 10[	9600	16,0	5	320,00
[10 ; 20[	13200	22,0	10	220,00
[20 ; 50[	20400	34,0	30	113,33
[50 ; 100[	7200	12,0	50	24,00
[100 ; 170]	1800	3,0	100	3,00
Total	60000	100,0		

Pour déterminer la borne inférieure de la première classe nous avons retenu le principe dit de la classe la plus proche, qui indique que l'amplitude inconnue est égale à l'amplitude de la classe suivante, ici la deuxième. Pour déterminer la borne supérieure de la dernière classe nous disposons d'une information complémentaire qui nous permet de déterminer son amplitude et sa borne supérieure.

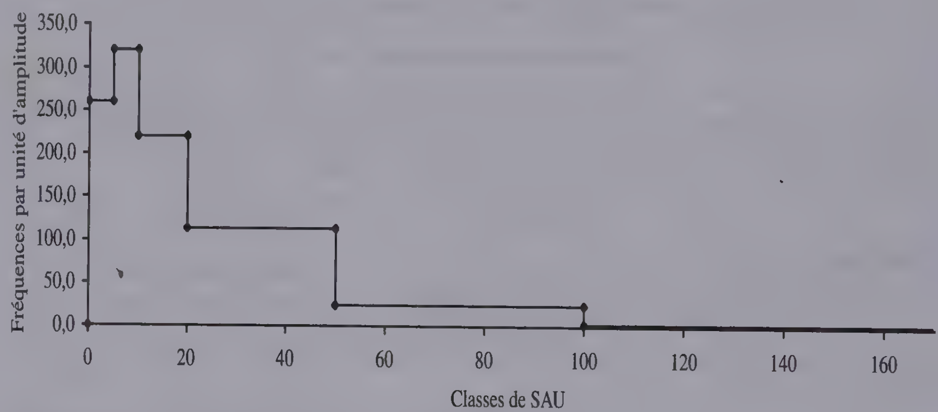
Nous savons que la surface totale pour les 1800 exploitations de plus de 100 ha est de 243 000 ha. La surface moyenne des exploitations de plus de 100 ha est donc de 135 ha (243000/1800), ce qui constitue le centre de la classe. La borne supérieure est alors de 170.

L'histogramme est une représentation où les surfaces sont proportionnelles à l'importance des différentes classes. Avec a_i amplitude de classe et f_i fréquence de la classe, la surface représentative est :

$$s_i = a_i \cdot h_i \quad \text{où} \quad h_i = k \cdot \frac{f_i}{a_i} = k \cdot f_{i/a_i}$$

où k est un coefficient quelconque par exemple 100.

Histogramme de la distribution



2. Détermination de la médiane

Calcul de la médiane

Classes de SAU	a_i	f_i	$\frac{f_i}{a_i}$	F_i
[0 ; 5[	5	0,13	0,026	0,13
[5 ; 10[	5	0,16	0,032	0,29
[10 ; 20[	10	0,22	0,022	0,51
[20 ; 50[	30	0,34	0,011	0,85
[50 ; 100[	50	0,12	0,002	0,97
[100 ; 170]	50	0,03	ε	1,00
Total		1,00		

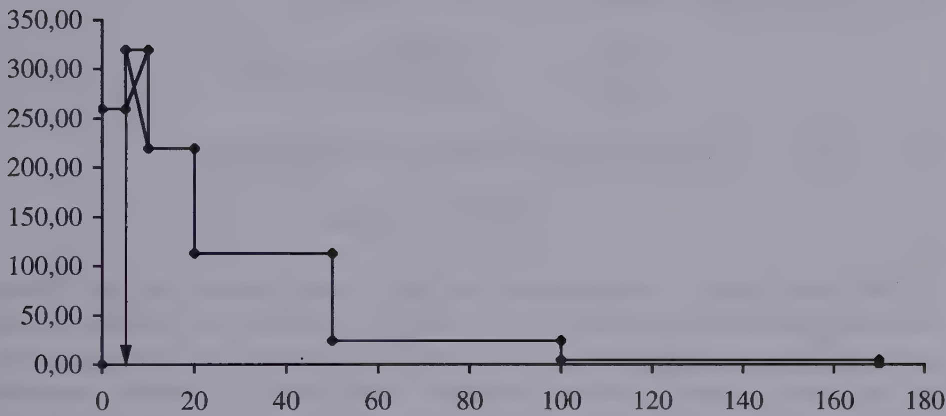
La classe médiane est donc la classe de 10 à 20 ha :

$$M_e = 10 + 10 \cdot \frac{0{,}50 - 0{,}29}{0{,}22} = 19{,}55 \text{ ha}$$

3. La classe modale est la classe de 0 à 5 ha. Elle nous est donnée par la plus grande fréquence $\frac{f_i}{a_i}$. Nous pouvons estimer le mode :

$$M_o = b_i + a_i \times \frac{\Delta_i}{\Delta_i + \Delta_{i+1}} = 0 + 5 \frac{0{,}032}{0{,}032 + (0{,}032 - 0{,}026)} \cong 4{,}24 \text{ ha}$$

Détermination graphique du mode



4. Le calcul de la surface moyenne nécessite un nouveau tableau :

Utilisation des fréquences

Classes de SAU	c_i	f_i	$f_i c_i$
[0 ; 5[	2,5	0,16	0,4
[5 ; 10[	7,5	0,13	0,975
[10 ; 20[	15,0	0,22	3,3
[20 ; 50[	35,0	0,34	11,9
[50 ; 100[	75,0	0,12	9,0
[100 ; 150]	125,0	0,03	3,75
Total		1,00	29,325

Nous obtenons immédiatement la moyenne $\overline{c} = 29{,}325$ ha.

Un second calcul est possible avec les effectifs.

Utilisation des effectifs

Classes de SAU	c_i	f_i	$f_i c_i$
[0 ; 5[	2,5	9 600	24 000
[5 ; 10[	7,5	7 800	58 500
[10 ; 20[	15,0	13 200	198 000
[20 ; 50[	35,0	20 400	714 000
[50 ; 100[	75,0	7 200	540 000
[100 ; 150]	125,0	1 800	225 000
Total		60 000	1 759 500

La surface moyenne obtenue par un second calcul est :

$$\bar{c} = \frac{1}{n}\sum_{i=1}^{k} n_i c_i = \frac{1\,759\,500}{60\,000} \cong 29{,}325 \text{ ha}$$

5. La surface moyenne obtenue par les données disponibles est :

$$\bar{c} = \frac{1\,860\,000}{60\,000} \cong 28 \text{ ha}$$

La différence entre le résultat obtenu avec les données globales et celui obtenu par le regroupement en classes des exploitations s'explique par la répartition non uniforme des exploitations au sein de chaque classe ainsi que par les hypothèses sur les classes extrêmes. Dans cet exemple, l'estimation de la surface moyenne en ayant recours aux centres de classes surestime la moyenne. La différence entre la valeur directe et la valeur calculée est d'environ 5 %.

$$\textit{écart} = \frac{29{,}325 - 28}{28} = 0{,}047 \cong 0{,}05$$

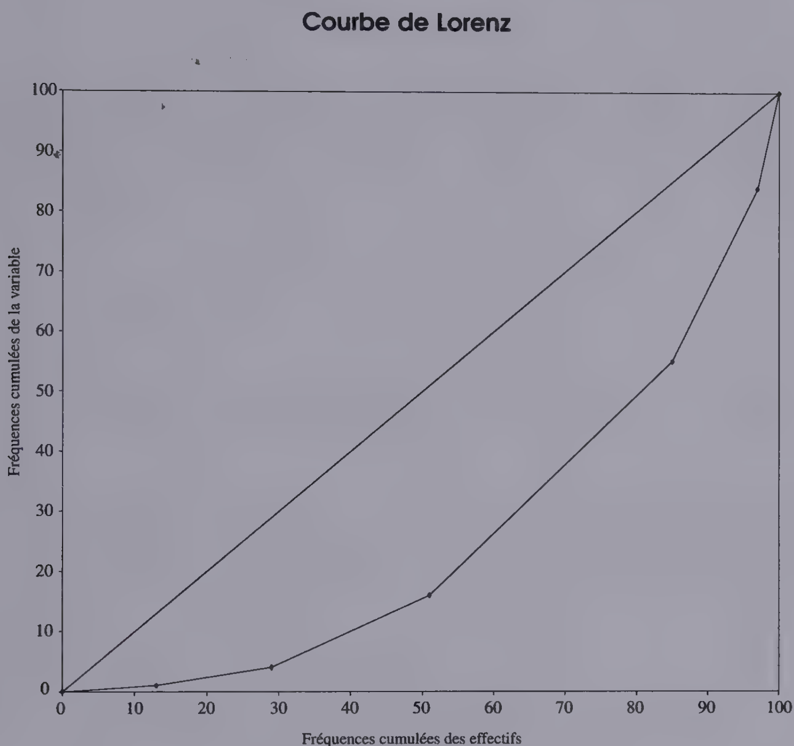
6. a. La médiane de la distribution des surfaces, appelée la médiale, s'obtient simplement à partir du tableau suivant :

Tableau du calcul de la médiale

Classes de SAU	f_i' (en %)	F_i'	f_i	$f_i \cdot (F_{i-1}' + F_i')$
[0 ; 5[	16	16	13,0	13,0
[5 ; 10[	13	29	16,0	80,0
[10 ; 20[	22	51	22,0	440,0
[20 ; 50[	34	85	34,0	2414,0
[50 ; 100[	12	97	12,0	1668,0
[100 ; 150]	3	100	3,0	552,0
	100		100,0	5167,0

La classe médiale de la répartition des S.A.U. est la classe [10 ; 20[.

$$M_l = 10 + 10 \cdot \frac{50 - 29}{22} = 19{,}55 \text{ ha}$$



L'indice de Gini est donc :

$$I_G = 1 - 0{,}5167 = 0{,}4833 \cong 0{,}48$$

Chapitre 3

Les distributions à deux dimensions

Ce chapitre présente des exercices sur les distributions pour lesquelles nous disposons d'observations concernant simultanément deux caractères (x_i, y_i) pour chaque individu de la population. Les aspects suivants font l'objet d'exercices : le tableau de contingence, la recherche et l'estimation des liaisons, l'ajustement linéaire, la corrélation.

51. Patrimoine

Mots-clefs

Caractéristiques marginales et conditionnelles

Énoncé

Le tableau ci-dessous donne la répartition du patrimoine mobilier en fonction du patrimoine immobilier (évalué en unités d'habitation) pour l'ensemble des souscripteurs à une société d'assurances.

1. Calculez les caractéristiques marginales suivantes :
   a. la moyenne et la variance du patrimoine mobilier ;
   b. la moyenne et la variance du patrimoine immobilier.
2. Calculez des caractéristiques conditionnelles suivantes :
   a. la moyenne et la variance du patrimoine mobilier pour les sociétaires possédant une unité d'habitation ;
   b. la moyenne et la variance des unités d'habitation pour les sociétaires entre 200 et 300 milliers d'euros de patrimoine mobilier.

Répartition des sociétaires assurés

	Répartition du patrimoine mobilier (en milliers d'euros)						
UH	[0 ; 50[	[50 ; 100[	[100 ; 200[	[200 ; 300[	[300 ; 400[	[400 ; 500[	[500 ; 600]
0	867	969	656	167	49	12	6
1	178	952	1825	841	328	168	45
2	26	148	464	339	173	110	35
3	4	28	80	79	51	38	13
4	2	12	44	40	27	21	10
5	1	6	25	22	15	15	9

Corrigé

1. a. Pour répondre à cette question nous devons construire le tableau statistique relatif à la répartition du patrimoine mobilier.

Tableau des calculs

Centres de classe	n_i	$n_i c_i$	$n_i c_i^2$
25	1078	26950	373750
75	2115	158625	11896875
150	3094	464100	69615000
250	1488	372000	93000000
350	643	225050	78767500
450	364	163800	73710000
550	118	64900	35695000
Total	8900	1475425	363358125

La moyenne est :

$$\overline{c} = \frac{1}{n}\sum_{i=1}^{k} n_i c_i = \frac{1\,475\,425}{8\,900} = 165{,}8 \text{ milliers d'euros}$$

La variance s'obtient par :

$$V(c) = \frac{1}{n}\sum_{i=1}^{k} n_i c_i^2 - \overline{c}^2 = \frac{363\,358\,125}{8\,900} - (165{,}8)^2 = 13\,344{,}4$$

L'écart type est :

$$\sigma_c = \sqrt{V(c)} = \sqrt{13\,344{,}4} = 115{,}5$$

b.

Tableau pour le calcul des moyenne et variance des unités d'habitation

Unité d'habitation	n_i	$n_i u_i$	$n_i u_i^2$
0	2726	0	0
1	4337	4337	4337
2	1295	2590	5180
3	293	879	2637
4	156	624	2496
5	93	465	2325
Total	8900	8895	16975

La moyenne est :

$$\overline{u} = \frac{1}{n}\sum_{i=1}^{k} n_i u_i = \frac{8\,895}{8\,900} = 1{,}0 \text{ unité d'habitation}$$

La variance s'obtient par :

$$V(u) = \frac{1}{n}\sum_{i=1}^{k} n_i u_i^2 - \overline{u}^2 = \frac{16\,975}{8\,900} - (1)^2 = 0{,}907$$

L'écart type est :

$$\sigma_u = \sqrt{V(u)} = \sqrt{0{,}907} = 0{,}95 \text{ unité d'habitation}$$

2. Pour les caractéristiques conditionnelles :

a. Les caractéristiques statistiques pour les sociétaires possédant une unité d'habitation constituent la distribution conditionnelle pour $i = 2$.

Tableau de la distribution pour une unité d'habitation

Centres de classe	n_i	$n_i c_i$	$n_i c_i^2$
25	178	4450	111250
75	952	71400	5355000
150	1825	273750	410662500
250	841	210250	5262500
350	328	114800	40180000
450	168	75600	34020000
550	45	24750	1361500
Total	4337	775000	186903750

La moyenne est :

$$\overline{c}_2 = \frac{1}{n}\sum_{i=1}^{k} n_i c_i = \frac{775\,000}{4\,337} = 178{,}7 \text{ milliers d'euros}$$

La variance s'obtient par :

$$V(c_2) = \frac{1}{n}\sum_{i=1}^{k} n_i c_i^2 - \overline{c}_2^2 = \frac{186\,903\,750}{4\,337} - (178{,}7)^2 = 11\,163{,}3$$

L'écart type est alors :

$$\sigma_{c_2} = \sqrt{V(c_2)} = \sqrt{11\,163{,}3} = 105{,}7 \text{ milliers d'euros}$$

b. Pour les sociétaires ayant entre 200 et 300 milliers de francs de patrimoine mobilier, les données correspondent à la distribution conditionnelle pour $j = 4$.

Tableau statistique de la quatrième distribution

Unité d'habitation (u_4)	n_i	$n_i u_i$	$n_i u_i^2$
0	167	0	0
1	841	841	841
2	339	678	1356
3	79	237	711
4	40	160	640
5	22	110	550
Total	1488	2026	16975

La moyenne est :

$$\overline{u}_4 = \frac{1}{n}\sum_{i=1}^{k} n_i u_i = \frac{2\,026}{1\,448} = 1{,}4 \text{ unité d'habitation}$$

La variance s'obtient par :

$$V(u_4) = \frac{1}{n}\sum_{i=1}^{k} n_i u_i^2 - \overline{u}_4^2 = \frac{4\,098}{1\,448} - (1{,}4)^2 = 0{,}900$$

L'écart type est :

$$\sigma_{u_4} = \sqrt{V(u_4)} = \sqrt{0{,}900} = 0{,}95 \text{ unité d'habitation.}$$

52. RELATION ENTRE HAUSSE DES PRIX ET TAUX DE CROISSANCE

Mots-clefs

Coefficient de Spearman, coefficient de corrélation des rangs

Énoncé

Un argument parfois invoqué pour permettre une adhésion des nouveaux États membres de l'UE à l'euro est que leur taux de hausse des prix plus élevé s'explique par un taux de croissance du PIB également plus élevé. Donc que nous aurions un lien fort entre ces deux indicateurs. Le calcul d'un coefficient de corrélation de rang de Spearman permet de tester cette hypothèse.

À partir des données disponibles, calculez le coefficient de rang de Spearman entre le taux de croissance du PIB et le taux de hausse des prix :

1. pour les 25 États membres ;

2. pour les 10 nouveaux adhérents.

Tableau des données année 2005

États membres	Évolution du PIB en volume	Indice des prix (en %)
Belgique	1,2	1,9
République tchèque	6	2,6
Danemark	3,1	0,9
Allemagne	1	1,8
Estonie	9,8	3
Grèce	3,7	3
Espagne	3,4	3,1
France	1,2	2,3
Irlande	4,7	2,3
Italie	0	2,3
Chypre	3,8	1,9
Lettonie	10,2	6,2
Lituanie	7,5	1,2
Luxembourg	4	3,2
Hongrie	4,1	6,8
Malte	2,4	2,7
Pays-Bas	1,1	1,4
Autriche	1,8	2
Pologne	3,2	3,6
Portugal	0,4	2,5
Slovénie	3,9	3,7
Slovaquie	6,1	7,5
Finlande	2,1	0,1
Suède	2,7	1
Royaume-Uni	1,8	1,3

Source : Eurostat

Corrigé

Nous calculons le coefficient de corrélation de rang de Spearman. Il mesure la relation qui existe entre deux classements. On calcule le rang de chaque élément dans la série croissante des deux puis la différence de classement d_i pour chaque couple de valeur puis le coefficient selon la formule :

$$R_s = 1 - \frac{6\sum_{i=1}^{i=n} d_i^2}{n(n^2-1)}$$

1.

Tableau des calculs

États membres	Évolution du PIB en volume	Rang	Indice des prix	Rang	d_i^2
Belgique	1,2	5	1,9	8	9
République tchèque	6	21	2,6	15	36
Danemark	3,1	12	0,9	2	100
Allemagne	1	3	1,8	7	16
Estonie	9,8	24	3	17	49
Grèce	3,7	15	3	17	4
Espagne	3,4	14	3,1	19	25
France	1,2	5	2,3	11	36
Irlande	4,7	20	2,3	11	81
Italie	0	1	2,3	11	100
Chypre	3,8	16	1,9	8	64
Lettonie	10,2	25	6,2	23	4
Lituanie	7,5	23	1,2	4	361
Luxembourg	4	18	3,2	20	4
Hongrie	4,1	19	6,8	24	25
Malte	2,4	10	2,7	16	36
Pays-Bas	1,1	4	1,4	6	4
Autriche	1,8	7	2	10	9
Pologne	3,2	13	3,6	21	64
Portugal	0,4	2	2,5	14	144
Slovénie	3,9	17	3,7	22	25
Slovaquie	6,1	22	7,5	25	9
Finlande	2,1	9	0,1	1	64
Suède	2,7	11	1	3	64
Royaume-Uni	1,8	7	1,3	5	4
Total					1337

Le coefficient de rang de Spearman est donc :

$$R_s = 1 - \frac{6\sum_{i=1}^{i=m} d_i^2}{n(n^2-1)} = 1 - \frac{6 \cdot 1\,337}{25(25^2-1)} = 0{,}4858$$

La corrélation de rang est peu significative, l'hypothèse d'un lien fort entre les deux variables n'est pas vérifiée en fonction des données disponibles.

2.

Tableau des calculs

États membres	Évolution du PIB en volume	Rang	Indice des prix	Rang	d_i^2
Chypre	3,6	3	2,6	5	4
Estonie	11,5	10	4	8	4
Hongrie	4,2	4	2,4	3	1
Lettonie	10,5	9	6,6	10	1
Lituanie	8,7	8	3,1	7	1
Malte	2,8	1	2,9	6	25
Pologne	2,8	1	0,9	1	0
République Tchèque	6,7	6	2,4	3	9
Slovaquie	7,6	7	4,3	9	4
Slovénie	5,1	5	2	2	9
					58

Le coefficient de rang de Spearman est dans ce cas :

$$R_s = 1 - \frac{6\sum_{i=1}^{i=m} d_i^2}{n(n^2-1)} = 1 - \frac{6 \cdot 58}{10(10^2-1)} = 0{,}65$$

La relation postulée est plus significative que pour les 25 États membres sans que pour autant que la corrélation de rang permette de postuler un lien fort.

53. UN SIMPLE AJUSTEMENT LINÉAIRE

Énoncé

En utilisant les données du tableau ci-dessous, répondez aux questions suivantes :

1. Construisez une droite qui ajuste les données du tableau suivant.
2. Calculez les coefficients de la droite d'ajustement par des méthodes empiriques.
3. Trouvez les équations des droites d'ajustement par la méthode des moindres carrés.
4. Comparez les différentes estimations avec les valeurs réelles.

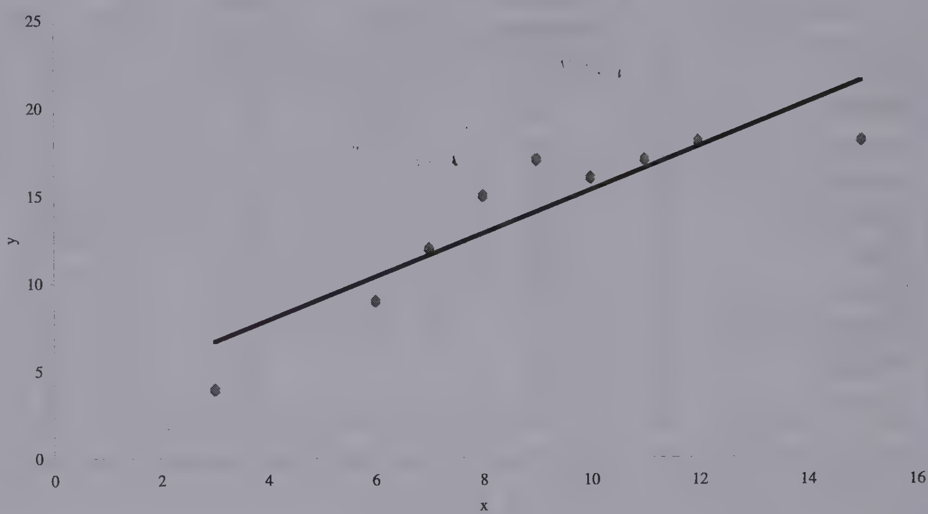
Tableau des données

x_i	3	6	7	8	9	10	11	12	15
y_i	4	9	12	15	17	16	17	18	18

Corrigé

1. La première méthode consiste à tracer les points sur un graphique, puis à main levée, à tracer une droite qui représente au mieux l'ensemble des points.

Droite d'ajustement



2. Avec la méthode de Mayer, nous avons deux possibilités, choisir deux points quelconques proches de la droite déterminée graphiquement, soit regrouper les points en deux sous ensembles.
   a. Une méthode plus analytique consiste à choisir deux points quelconques de la droite tracée puis de calculer l'équation de cette droite.

Soit un point $A(x_1, y_1)$ et un point $B(x_2, y_2)$, l'équation de la droite d'ajustement : $\widehat{y} = ax + b$; alors nous avons les deux équations à deux inconnues suivantes qui permettent de calculer a et b :

$$\begin{cases} y_1 = ax_1 + b \\ y_2 = ax_2 + b \end{cases}$$

Nous choisissons le point (6,12) et le point (11,17) qui répondent à notre critère. Le système à résoudre est le suivant :

$$\begin{cases} 12 = 6a + b \\ 17 = 11a + b \end{cases}$$

Nous trouvons : $a = 1{,}25$ et $b = 3{,}25$

$$\widehat{y} = 1{,}25x + 3{,}25$$

b. Nous regroupons, mais un autre choix était possible, en deux sous ensembles, puis nous prenons le point moyen de ces deux groupes (le barycentre).

Premier groupe

						Moyenne
x_i	3	6	7	8	9	6,6
y_i	4	9	12	15	17	11,4

Deuxième groupe

					Moyenne
x_i	12	10	11	15	12
y_i	18	16	17	18	17,25

Le nouveau système est :

$$\begin{cases} 11{,}4 = 6{,}6a + b \\ 17{,}25 = 12a + b \end{cases}$$

La résolution du système donne $a = 1{,}09$ et $b = 4{,}78$.

$$\widehat{y} = 1{,}09x + 4{,}78$$

Nous obtenons deux équations différentes mais proches.

3. La troisième méthode élimine toute subjectivité, en utilisant la méthode des moindres carrés. Pour cela il est nécessaire de construire le tableau statistique suivant :

Tableau des calculs

	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	3	4	9	16	12
2	6	9	36	81	54
3	7	12	49	144	84
4	8	15	64	225	120
5	9	17	81	289	153
6	10	16	144	324	216
7	11	17	100	256	160
8	12	18	121	289	187
9	15	18	225	324	270
Total	81	126	829	1 948	1 256

Le coefficient a est donné par la relation :

$$a = \frac{Cov(x,y)}{V(x)} = \frac{\sum_{i=1}^{k} x_i y_i - n \cdot \overline{x} \cdot \overline{y}}{\sum_{i=1}^{k} x_i^2 - n\overline{x}^2} = \frac{\sum_{i=1}^{k} x_i y_i - \frac{\sum_{i=1}^{k} x_i \sum_{i=1}^{k} y_i}{n}}{\sum_{i=1}^{k} x_i^2 - \frac{\left(\sum_{i=1}^{k} x_i\right)^2}{n}}$$

$$a = \frac{1\,256 - \frac{81 \cdot 126}{9}}{829 - \frac{81^2}{9}} = 1{,}22$$

$$b = \overline{y} - a\overline{x} = \frac{126}{9} - 1{,}22 \cdot \frac{81}{9} = 3{,}0$$

L'équation de la droite d'ajustement de x en y est donc :

$$\widehat{y} = 1{,}22x + 3$$

L'équation de la droite d'ajustement de y en x $\widehat{x} = a'y + b'$ est analogue

$$a' = \frac{Cov(x,y)}{V(y)} = \frac{\sum_{i=1}^{k} x_i y_i - n \cdot \overline{x} \cdot \overline{y}}{\sum_{i=1}^{k} y_i^2 - n\overline{y}^2} = \frac{\sum_{i=1}^{k} x_i y_i - \frac{\sum_{i=1}^{k} x_i \sum_{i=1}^{k} y_i}{n}}{\sum_{i=1}^{k} y_i^2 - \frac{\left(\sum_{i=1}^{k} y_i\right)^2}{n}}$$

$$a' = \frac{1\,256 - \frac{81 \cdot 126}{9}}{1\,948 - \frac{126^2}{9}} = 0{,}66$$

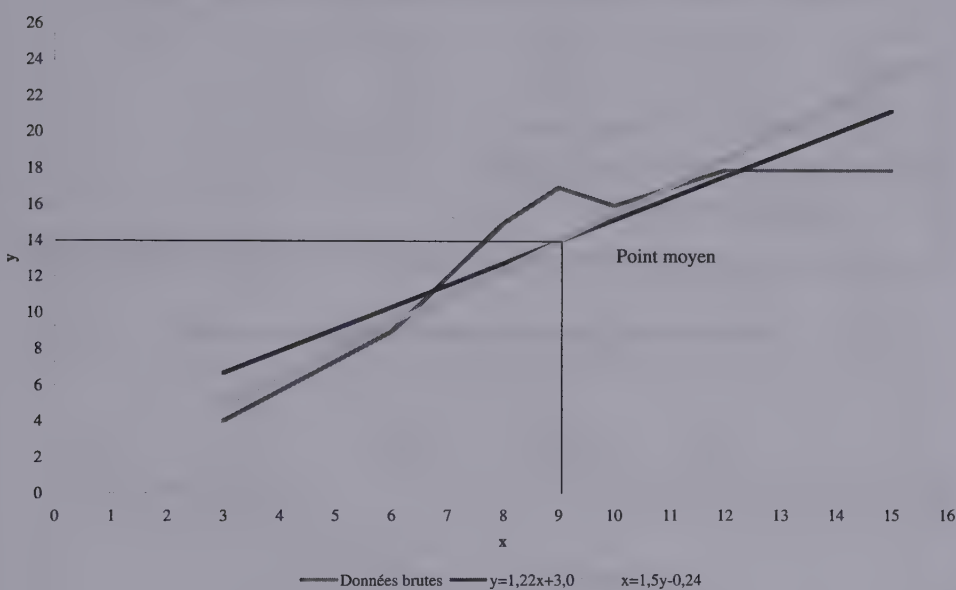
d'où :

$$b' = \overline{x} - a'\overline{y} = \frac{81}{9} - 0{,}66 \cdot \frac{126}{9} = -0{,}24$$

$$\widehat{x} = 0{,}66y - 0{,}24$$

4. Pour comparer les résultats nous avons regroupé sur une même représentation les graphiques des droites d'ajustement. Nous remarquons que les écarts ne sont pas très importants. La droite permet de réaliser des extrapolations. Le graphique montre cependant que les écarts augmentent dès que les valeurs s'éloignent du barycentre, ce qui rend l'extrapolation de plus en plus incertaine. La qualité des extrapolations se réduit au fur et à mesure de l'éloignement des valeurs moyennes.

Représentations des différentes équations et du nuage de points



54. Quelle relation entre le profit brut des entreprises et leur investissement ?

Mots-clefs

Corrélation et ajustement

Énoncé

Nous disposons pour quelques années des données relatives à l'EBE à la FBCF. Les outils statistiques permettent de donner des éléments sur l'éventuelle liaison entre les deux agrégats.

Tableau des données initiales

Année	1	2	3	4	5	6	7	8	9	10
E.B.E.	307	351	374	423	475	537	617	699	815	867
F.B.C.F	213	235	282	306	342	357	372	408	438	472

Source : INSEE

En utilisant les données du tableau initial, il est demandé de répondre aux questions suivantes :

1. Expliquez la signification des termes E.B.E, F.B.C.F., S.N.F. .
2. Représentez le nuage de points de la distribution (E.B.E. en abscisses).
3. Estimez le coefficient de détermination linéaire entre l'E.B.E. et la F.B.C.F. .
4. Déterminer les équations des droites d'ajustement et les construire.
5. Commenter en quelques lignes (5 maximum) les résultats obtenus.

Corrigé

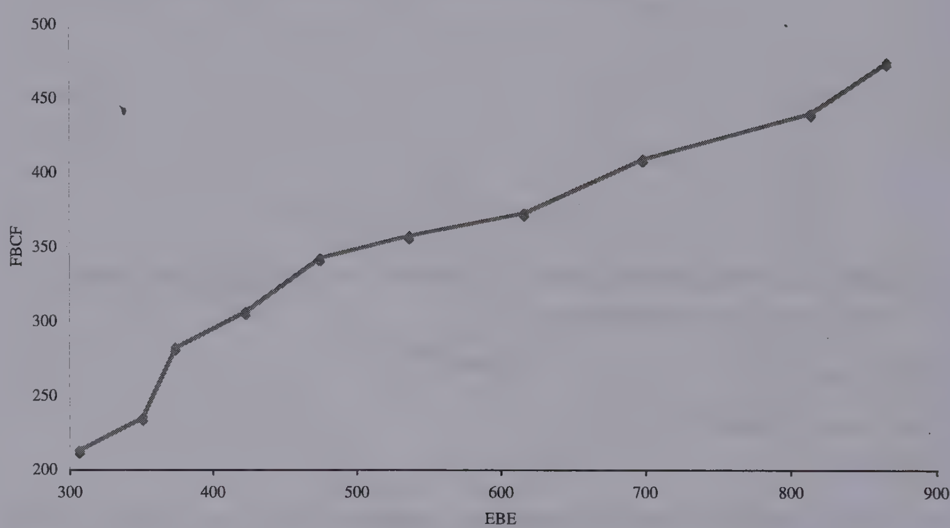
1. EBE : excédent brut d'exploitation

FBCF : formation brute de capital fixe

SNF : sociétés non financières.

2. Le nuage de points de la distribution est représenté par le graphique ci-dessous :

Représentation de l'évolution de l'EBE et de la FBCF



Ce graphique montre que les points représentatifs des couples de variables sont grossièrement alignés. Un ajustement linéaire apparaît pertinent.

3. L'examen du graphique précédent indique que les calculs du coefficient de corrélation linéaire et des droites d'ajustement sont pertinents pour formaliser la relation entre l'E.B.E. et la F.B.C.F. Pour cela nous construisons le tableau statistique ci-après.

Tableau des calculs intermédiaires

Année	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	307	213	94 249	45 369	65 391
2	351	235	123 201	55 225	82 485
3	374	282	139 876	79 524	105 468
4	423	306	178 929	93 636	129 438
5	475	342	225 625	116 964	162 450
6	537	357	288 369	127 449	191 709
7	617	372	380 689	138 384	229 524
8	699	408	488 601	166 464	285 192
9	815	438	664 225	191 844	356 970
10	867	472	751 689	222 784	409 224
Total	5465	3425	3 335 453	1 237 643	2 017 851

Le coefficient de corrélation est :

$$r = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} \cdot \sqrt{\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2}}$$

$$r = \frac{2\,017\,851 - \dfrac{5\,465 \cdot 3\,425}{10}}{\sqrt{3\,335\,453 - \dfrac{5\,465^2}{10}} \cdot \sqrt{1\,237\,643 - \dfrac{3\,425^2}{10}}} = 0{,}973$$

$$r \cong 0{,}973$$

d'où le coefficient de détermination

$$r^2 = 0{,}947$$

La valeur du coefficient de détermination ne permet pas de rejeter l'hypothèse d'une liaison entre les deux variables ; nous calculerons donc les équations des droites de régression de x en y et de y en x.

Les équations des deux droites d'ajustement sont respectivement :

$$\widehat{y} = ax + b$$

$$\widehat{x} = a'y + b'$$

$$a = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} = \frac{2\,017\,851 - \dfrac{5\,465 \cdot 3\,425}{10}}{3\,335\,453 - \dfrac{(5\,465)^2}{10}} \cong 0{,}42$$

$$a' = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2} = \frac{2\,017\,851 - \dfrac{5\,465 \cdot 3\,425}{10}}{1\,237\,643 - \dfrac{(3\,425^2)}{10}} \cong 2{,}26$$

$$b = \overline{y} - a\overline{x} = \frac{3\,245}{10} - 0{,}42 \cdot \frac{5\,465}{10} = 113{,}63$$

$$b' = \overline{x} - a'\overline{y} = \frac{5\,465}{10} - 2{,}26 \cdot \frac{3\,425}{10} = -228{,}3$$

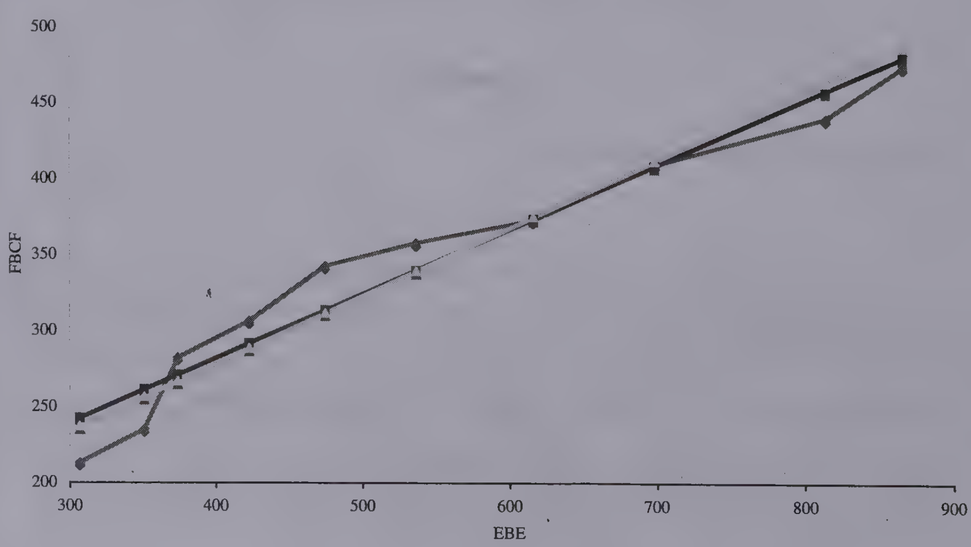
d'où les équations des deux droites de régression :

$$\widehat{y} = ax + b = 0{,}42x + 113{,}63$$

$$\widehat{x} = a'y + b' = 2{,}26y - 228{,}3$$

4. Le graphique suivant visualise les résultats obtenus.

Graphique des deux droites d'ajustement



Pour les dix années considérées, les investissements apparaissent fortement corrélés au profit brut. Néanmoins, ce n'est qu'en tendance car on remarque que, pour les dernières années, l'investissement réel est inférieur à l'investissement estimé.

Tableau des résultats

	1	2	3	4	5	6	7	8	9	10
F.B.C.F	213	235	282	306	342	357	372	408	438	472
$FBCF_{est}$	242	261	270	291	313	339	372	406	455	477
Écarts	– 29	– 26	12	15	29	18	0	2	– 17	– 5
E.B.E.	307	351	374	423	475	537	617	699	815	867
EBE_{est}	254	303	410	464	545	579	613	695	763	839
Écarts	53	48	– 36	– 41	– 70	– 42	4	4	52	28

55. Relations entre les niveaux de revenu et le taux d'épargne

Mots-clefs

Coefficient de corrélation, droites d'ajustement, variable continue (classes)

Énoncé

Est-il pertinent de lier le niveau de revenu et l'épargne ? Une autre formulation de la même question serait la suivante : le niveau de revenu permet-il une évaluation du taux d'épargne ?

Tableau des revenus et de taux d'épargne

Classe de revenu des ménages (en 1000 €)	Taux de détention d'actifs (1) (en %)	
	Plan d'épargne logement	SICAV /FCP
– 30	5	2
30 – 50	9	2
50 – 75	15	5
75 – 100	19	9
100 – 130	23	12
130 – 200	28	19
200 – 300	38	37
300 000 et plus	41	52

(1) L'usage est d'utiliser le terme de taux pour ces proportions.

1. Calculez le coefficient de corrélation linéaire entre les classes de revenu et les taux de détention d'actifs des Sociétés d'investissement à capital variable (SICAV) et Fonds commun de placement (FCP).

2. Calculez les équations des droites d'ajustement correspondantes.
3. Calculez le coefficient de corrélation linéaire entre les classes de revenu et les plans d'épargne logement (P.E.L.).
4. Répondez à la question initialement posée en utilisant les résultats des calculs précédents.

Corrigé

1. Pour calculer le coefficient de corrélation linéaire entre revenus et SICAV, il est nécessaire de construire le tableau des calculs intermédiaires.

Tableau des calculs intermédiaires pour les SICAV

Classes de revenus (1000 euros)	c_i	y_i	c_i^2	y_i^2	$c_i \cdot y_i$
[0 ; 30[	15,0	2	225,00	4	30,0
[30 ; 50[	40,0	2	1 600,00	4	80,0
[50 ; 75[	62,5	5	3 906,25	25	312,5
[75 ; 100[	87,5	9	7 656,25	81	787,5
[100 ; 130[	115,0	12	13 225,00	144	1 380,0
[130 ; 200[	165,0	19	27 225,00	361	3 135,0
[200 ; 300[	250,0	37	62 500,00	1 369	9 250,0
[300 ; 400[	350,0	52	122 500,00	2 704	18 200,0
	1 085,0	138	238 837,50	4 692	33 175,0

Le coefficient de corrélation est le rapport de la covariance aux écarts types.

$$r = \frac{\sum_{i=1}^{k} c_i y_i - \frac{1}{k}\sum_{i=1}^{k} c_i \cdot \sum_{i=1}^{k} y_i}{\sqrt{\sum_{i=1}^{k} c_i^2 - \frac{1}{k}\left(\sum_{i=1}^{k} c_i\right)^2} \cdot \sqrt{\sum_{i=1}^{k} y_i^2 - \frac{1}{k}\left(\sum_{i=1}^{k} y_i\right)^2}}$$

$$r = \frac{33\,175 - \dfrac{1\,085 \cdot 138}{8}}{\sqrt{238\,837{,}5 - \dfrac{1\,085^2}{8}} \cdot \sqrt{4\,692 - \dfrac{138^2}{8}}} = 0{,}9932$$

Le coefficient de corrélation linéaire r est de 0,99. Le coefficient de détermination est r^2 et vaut de 0,986. La valeur du coefficient est suffisante pour que le calcul des équations des droites d'ajustement soit pertinent.

2. Les équations des droites d'ajustement sont obtenues à partir du même tableau. Le coefficient a de la droite de régression $\widehat{y} = ax + b$ est :

$$a = \frac{\sum_{i=1}^{k} c_i y_i - \frac{1}{k}\sum_{i=1}^{k} c_i \cdot \sum_{i=1}^{k} y_i}{\sum_{i=1}^{k} c_i^2 - \frac{1}{k}\left(\sum_{i=1}^{k} c_i\right)^2} = \frac{33\,175 - \frac{1\,085 \cdot 138}{8}}{238\,837{,}5 - \frac{1\,085^2}{8}} = 0{,}1577$$

Le coefficient a' de la droite d'ajustement $\widehat{c} = a'x + b'$ est :

$$a' = \frac{\sum_{i=1}^{k} c_i y_i - \frac{1}{k}\sum_{i=1}^{k} c_i \cdot \sum_{i=1}^{k} y_i}{\sum_{i=1}^{k} y_i^2 - \frac{1}{k}\left(\sum_{i=1}^{k} y_i\right)^2} = \frac{33\,175 - \frac{1\,085 \cdot 138}{8}}{4\,692 - \frac{138^2}{8}} = 6{,}255$$

Les coefficients b et b' sont alors :

$$b = \overline{y} - a\overline{c} = \frac{138}{8} - 0{,}1577 \cdot \frac{1\,085}{8} = -4{,}138$$

$$b' = \overline{c} - a'\overline{y} = \frac{1\,085}{8} - 6{,}255 \cdot \frac{138}{8} = 27{,}726$$

Les deux équations sont donc :

$$\widehat{y} = a \cdot c + b = 0{,}1577c - 4{,}138$$

$$\widehat{c} = a'y + b' = 6{,}255y - 27{,}726$$

3.

Tableau des calculs pour l'épargne logement

Classes de revenus (1000 euros)	c_i	y_i	c_i^2	y_i^2	$c_i \cdot y_i$
[0 ; 30[	15,0	5	225,00	25	75
[30 ; 50[	40,0	9	1 600,00	81	360
[50 ; 75[	62,5	15	3 906,25	225	937,5
[75 ; 100[	87,5	19	7 656,25	361	1662,5
[100 ; 130[	115,0	23	13 225,00	529	2645,0
[130 ; 200[	165,0	28	27 225,00	784	4620
[200 ; 300[	250,0	38	62 500,00	1444	9500
[300 ; 400[	350,0	41	122 500,00	1681	14350
	1 085,0	178	238 837,50	5130	34150

$$r = \frac{\sum_{i=1}^{k} c_i y_i - \frac{1}{k}\sum_{i=1}^{k} c_i \cdot \sum_{i=1}^{k} y_i}{\sqrt{\sum_{i=1}^{k} c_i^2 - \frac{1}{k}\left(\sum_{i=1}^{k} c_i\right)^2} \cdot \sqrt{\sum_{i=1}^{k} y_i^2 - \frac{1}{k}\left(\sum_{i=1}^{k} y_i\right)^2}}$$

$$r = \frac{34\,150 - \frac{1}{8}1\,085 \cdot 178}{\sqrt{238\,837{,}5 - \frac{1}{8}(1\,085)^2} \cdot \sqrt{5\,130 - \frac{1}{8}(138)^2}} = 0{,}967$$

Le coefficient de corrélation linéaire entre les niveaux de revenu et la détention d'un PEL est de 0,96, celui de détermination est de 0,934.

4. Le coefficient de détermination entre le revenu et l'épargne sous forme de SICAV ($r^2 = 0{,}986$) est proche de un, l'estimation d'une liaison linéaire est acceptable. Le taux de détention des SICAV-FCP est bien en relation linéaire avec le niveau des revenus, la connaissance du niveau de revenu permet d'estimer correctement le taux de détention de cet actif financier. La relation entre le revenu et l'épargne logement est moins forte tout en restant très significative, les hauts revenus s'orientent moins vers l'épargne logement.

56. Ajustement linéaire dans un tableau de contingence

Mots clefs

Coefficient de corrélation, coefficient de détermination, croisement de variables continues (classes)

Énoncé

Le tableau suivant fournit pour quelques années les valeurs de deux grandeurs dont on suppose que les variations sont liées.

Tableau des distributions

X \ Y	4	6	8	10	12	14
5	30					
6	20	30				
7	10	30	35	10		
8			20	25	30	5
9				15	15	25
10						5
12						5

1. Calculez le coefficient de détermination linéaire entre X et Y.
2. Calculez l'équation des droites d'ajustement et les représenter graphiquement.

Corrigé

Pour répondre aux questions posées, nous construisons le tableau statistique standard pour une distribution à deux dimensions.

Tableau statistique des calculs intermédiaires

	4	6	8	10	12	14	$n_{i.}$	B_i	D_i	$\overline{y}_i$	$V_i(y)$	E_i
5	30						30	120	480	4	0,00	600
6	20	30					50	260	1400	5,2	0,96	1560
7	10	30	35	10			85	600	4480	7,1	2,88	4200
8			20	25	30	5	80	840	9080	10,5	3,25	6720
9				15	15	25	55	680	8560	12,4	2,78	6120
10						5	5	70	980	14	0,00	700
12						5	5	70	980	14	0,00	840
$n_{.j}$	60	60	55	50	45	40	310	2640	25960			20740
A_j	340	390	405	405	375	375	2290					
C_j	1960	2550	2995	3305	3135	3565	17510					
$\overline{x}_j$	5,67	6,50	7,36	8,10	8,33	9,38						
$V_j(x)$	0,56	0,25	0,23	0,49	0,22	1,23						
E_j	1360	2340	3240	4050	4500	5250	20740					

1. Nous savons que le coefficient de détermination peut se calculer à partir des résultats intermédiaires :

$$r = \frac{E - \dfrac{A \cdot B}{n}}{\sqrt{C - \dfrac{A^2}{n}} \cdot \sqrt{D - \dfrac{B^2}{n}}}$$

$$r = \frac{20\,740 - \dfrac{2\,290 \cdot 2\,640}{310}}{\sqrt{17\,510 - \dfrac{2\,290^2}{310}} \cdot \sqrt{25\,960 - \dfrac{2\,640^2}{310}}} = 0{,}86$$

Le coefficient de détermination est alors :

$$r^2 = 0{,}74$$

2. Nous calculons les coefficients des équations des droites d'ajustement en utilisant les résultats intermédiaires.

Nous calculons tout d'abord la droite d'ajustement $\widehat{Y} = aX + b$

$$a = \frac{E - \dfrac{A \cdot B}{n}}{C - \dfrac{A^2}{n}}$$

$$a = \frac{20\,740 - \dfrac{2\,290 \cdot 2\,640}{310}}{17\,510 - \dfrac{2\,290^2}{310}} = 2{,}086$$

d'où

$$b = \frac{B}{n} - a \cdot \frac{A}{n} = \frac{2\,640}{310} - 2{,}086 \cdot \frac{2\,290}{310} = -6{,}892$$

L'équation de la droite d'ajustement est donc :

$$\widehat{Y} = 2{,}1X - 6{,}9$$

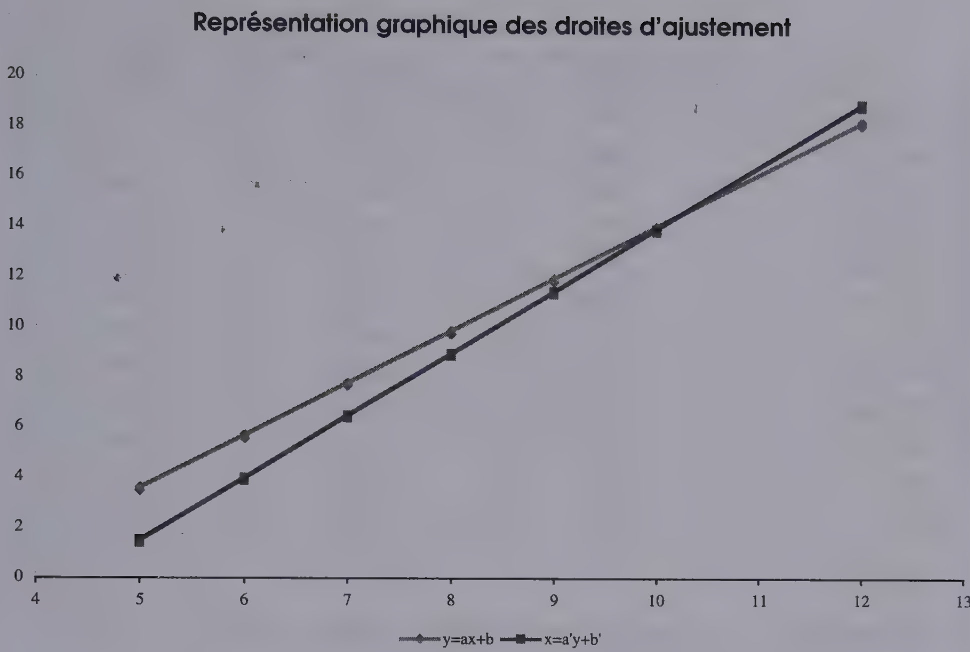
Nous calculons maintenant l'équation de la droite d'ajustement $X = a'Y + b'$

$$a' = \frac{E - \dfrac{A \cdot B}{n}}{D - \dfrac{B^2}{n}}$$

$$a' = \frac{20\,740 - \dfrac{2\,290 \cdot 2\,640}{310}}{25\,960 - \dfrac{2\,640^2}{310}} = 0{,}356$$

$$b' = \frac{A}{n} - a' \cdot \frac{B}{n} = \frac{2\,290}{310} - 0{,}356 \cdot \frac{2\,640}{310} = 4{,}36$$

D'où l'équation de la droite : $\widehat{X} = 0{,}4Y + 4{,}4$



57. RÉPARTITION DES ÉTUDIANTS

Mots-clefs

Indépendance, force de la liaison, Khi2, caractères qualitatifs

Énoncé

Soit la répartition des 17 557 étudiants de l'Université Joseph Fourrier de Grenoble suivant leur genre et leur UFR.

1. Construisez une représentation graphique de cette répartition.
2. La répartition des étudiants entre hommes et femmes est-elle identique pour toutes les UFR ?
3. Que vaut le *V* de Cramer ?

Effectifs étudiants en 2005-2006

UFR	Hommes	Femmes	Total
APS	820	494	1314
Biologie	347	492	839
UJF Valence	520	492	885
Chimie	226	212	438
DSU	1521	971	2492

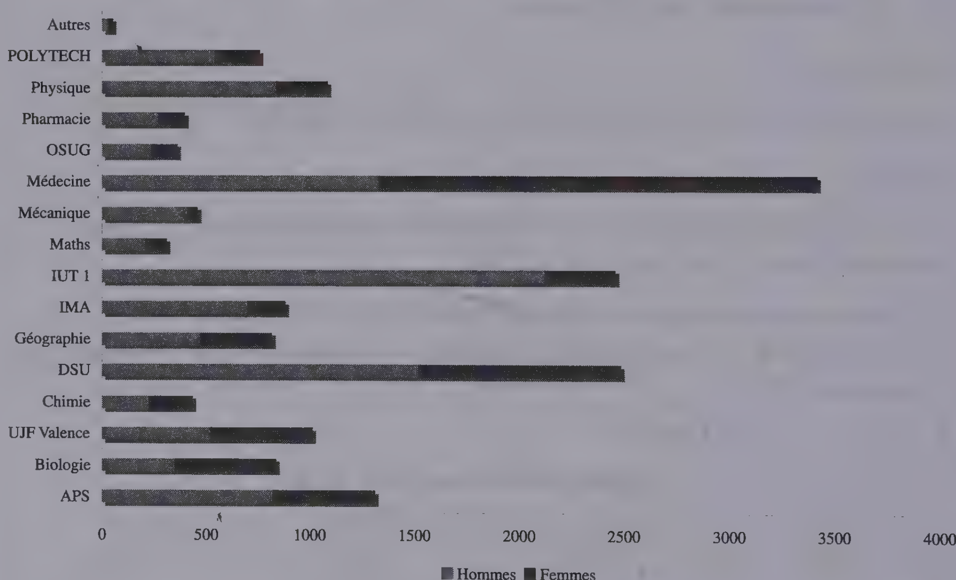
Effectifs étudiants en 2005-2006 (suite)

UFR	Hommes	Femmes	Total
Géographie	470	348	818
IMA	698	185	883
IUT 1	2127	339	2466
Maths	208	104	312
Mécanique	409	53	462
Médecine	1329	2097	3426
OSUG	236	129	365
Pharmacie	271	129	365
Physique	835	252	1087
POLYTECH	542	218	760
Autres	28	25	53
Total	10587	6540	17127

Corrigé

1. Nous retenons une représentation graphique en barres doubles.

Répartition des étudiants à l'université Joseph Fourrier



2. Nous allons tester l'indépendance des variables.

Nous avons dans cette université 61,8 % d'hommes et 38,2 % de femmes, si cette répartition était respectée pour toutes les UFR nous devrions avoir la répartition suivante dite répartition théorique.

Distribution théorique

UFR	Hommes	Femmes	Total
APS	812,2	501,8	1314
Biologie	518,6	320,4	839
UJF Valence	625,6	386,4	1012
Chimie	270,7	167,3	438
DSU	1540,4	951,6	2492
Géographie	505,6	312,4	818
IMA	545,8	337,2	883
IUT 1	1524,3	941,7	2466
Maths	192,9	119,1	312
Mécanique	285,6	176,4	462
Médecine	2117,8	1308,2	3426
OSUG	225,6	139,4	365
Pharmacie	247,3	152,7	400
Physique	671,9	415,1	1087
POLYTECH	469,8	290,2	760
Autres	32,8	20,2	53
Total	10587	6540	17127

Pour effectuer un calcul du χ^2 nous calculons les carrés relatifs des écarts

$$\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{m}\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Avec O_{ij} la valeur observée et E_{ij} la valeur théorique

Carrés relatifs des écarts

UFR	Hommes	Femmes	Total
APS	0,07	0,12	0,19
Biologie	56,79	91,94	148,73
UJF Valence	17,81	28,84	46,65
Chimie	7,40	11,97	19,37
DSU	0,24	0,40	0,64
Géographie	2,51	4,07	6,58
IMA	42,43	68,68	111,11
IUT 1	238,26	385,69	623,95
Maths	1,19	1,92	3,11
Mécanique	53,33	86,34	139,67
Médecine	293,78	475,57	769,35

Carrés relatifs des écarts (suite)

UFR	Hommes	Femmes	Total
OSUG	0,48	0,77	1,25
Pharmacie	2,28	3,69	5,97
Physique	39,58	64,07	103,65
POLYTECH	11,10	17,97	29,07
Autres	0,69	1,12	1,81
Total	767,95	1243,16	2011,11

La somme des carrés relatifs est suffisamment importante pour que nous puissions retenir l'hypothèse que la répartition entre les hommes et les femmes n'est pas indépendante des UFR.

Nous pouvons calculer la contribution de chacune des cases du tableau au χ^2 global.

Contributions

UFR	Hommes	Femmes	Total
APS	0,0	0,0	0,0
Biologie	2,8	4,6	7,4
UJF Valence	0,9	1,4	2,3
Chimie	0,4	0,6	1,0
DSU	0,0	0,0	0,0
Géographie	0,1	0,2	0,3
IMA	2,1	3,4	5,5
IUT 1	11,8	19,2	31,0
Maths	0,1	0,1	0,2
Mécanique	2,7	4,3	6,9
Médecine	14,6	23,6	38,3
OSUG	0,0	0,0	0,1
Pharmacie	0,1	0,2	0,3
Physique	2,0	3,2	5,2
POLYTECH	0,6	0,9	1,4
Autres	0,0	0,1	0,1
Total	38,2	61,8	100,0

Le tableau représente les fréquences

$$f_{ij} = \frac{n_{ij}}{n} \quad .$$

Soit, par exemple, pour la modalité UFR = Médecine et la modalité sexe = femmes, nous obtenons

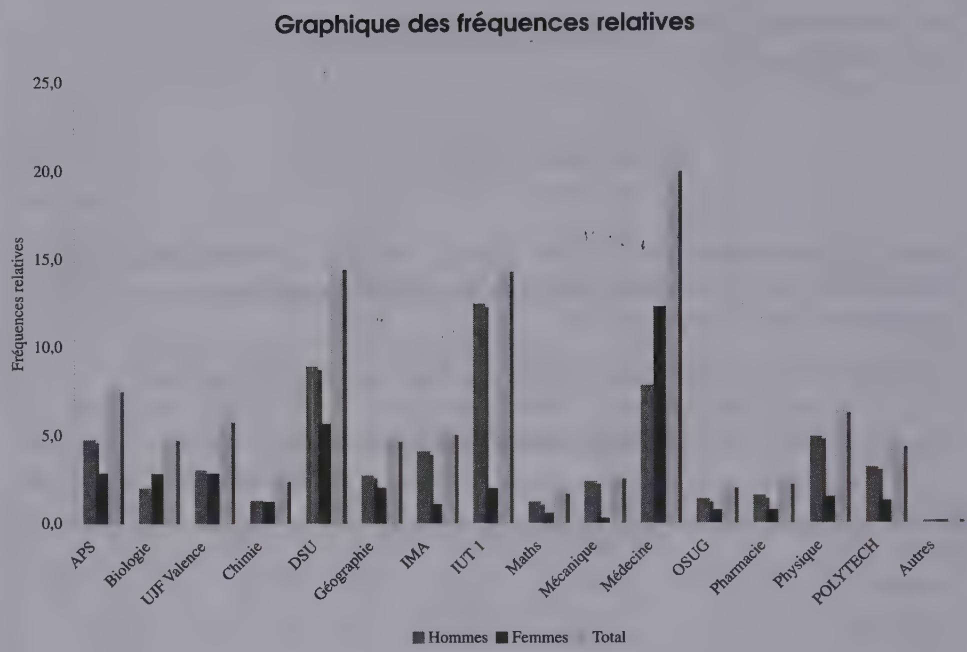
$$f_{\text{médecine/femmes}} = \frac{475{,}57}{2\,011{,}11} \cong 23{,}6$$

Deux UFR expliquent près de 70 % du χ^2 (69,3 %), ce sont respectivement médecine et IUT 1, en médecine les femmes sont surreprésentées alors que les hommes le sont à l'IUT1.

Le graphique des fréquences relatives confirme les résultats du calcul.

Tableau des fréquences relatives

UFR	Hommes	Femmes	Total
APS	4,8	2,9	7,7
Biologie	2,0	2,9	4,9
UJF Valence	3,0	2,9	5,9
Chimie	1,3	1,2	2,6
DSU	8,9	5,7	14,6
Géographie	2,7	2,0	4,8
IMA	4,1	1,1	5,2
IUT 1	12,4	2,0	14,4
Maths	1,2	0,6	1,8
Mécanique	2,4	0,3	2,7
Médecine	7,8	12,2	20,0
OSUG	1,4	0,8	2,1
Pharmacie	1,6	0,8	2,3
Physique	4,9	1,5	6,3
POLYTECH	3,2	1,3	4,4
Autres	0,2	0,1	0,3
Total	61,8	38,2	100,0



3. Puisque nous supposons une absence d'indépendance entre le caractère sexe et le caractère UFR, nous pouvons calculer un V de Cramer

$$V_C = \sqrt{\frac{\chi^2}{N(k-1)}}$$

avec N l'effectif de la distribution, soit 17 127 et k le minimum entre le nombre de ligne et de colonne, dans notre exemple 16 lignes et 2 colonnes donc $k = 2$. D'où :

$$V_C = \sqrt{\frac{2\,011{,}11}{17\,127 \cdot 1}} \cong 0{,}34$$

La force de la relation n'est pas très considérable, elle « explique » un peu plus du tiers des liens entre les deux caractères. Cela tient au rôle déterminant joué par deux UFR dans l'importance des différences.

Chapitre 4

Les séries chronologiques

Une chronique ou série chronologique est une distribution à deux dimensions dont l'une est le temps.

L'analyse d'une chronique consiste à mettre en évidence et à classer les facteurs qui influent sur la grandeur considérée. Ces mouvements peuvent être de plus ou moins long terme. La périodicité d'une chronique est la durée qui sépare deux observations. L'étude d'une série chronologique exige que les relevés successifs soient comparables.

On distingue quatre composantes du mouvement d'une chronique : la tendance longue, les fluctuations cycliques (C), les variations saisonnières (S), les variations aléatoires ou accidentelles (R).

Il est possible de fournir un modèle des séries chronologiques reposant sur l'hypothèse d'une relation entre les composantes. Cette relation peut s'exprimer par une composition additive des mouvements ou une composition multiplicative.

La détermination de la tendance ou trend permet d'atteindre plusieurs objectifs. Tout d'abord élaborer des modèles descriptifs des tendances antérieures. Ensuite, si les raisons existent d'un maintien de ces orientations, permettre de faire des prévisions à structure constante. Enfin éliminer le trend pour étudier les autres mouvements, en particulier les fluctuations cycliques et saisonnières. La détermination se fait généralement à partir des données annuelles qui éliminent la composante saisonnière et les fluctuations aléatoires.

58. Lissage d'une série et représentation graphique

Mots-clefs

Moyennes mobiles, série chronologique

Énoncé

Appliquez à la série chronologique suivante une moyenne mobile centrée d'ordre 3 puis représentez sur le même graphique la série brute et la série des moyennes mobiles.

Série initiale

t_i	1	2	3	4	5	6	7	8	9
x_t	78	60	66	84	96	75	102	87	72
t_i	10	11	12	13	14	15	16	17	18
x_t	81	87	99	104	95	92	98	105	96

Corrigé

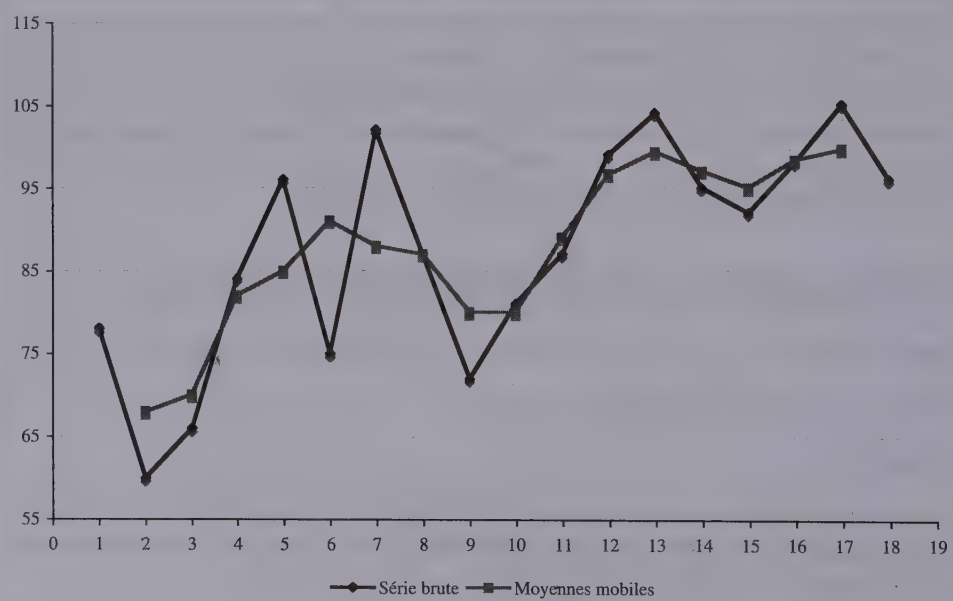
La série des moyennes mobiles s'obtient à partir de la série brute. Nous calculons la moyenne sur trois périodes et nous affectons le résultat au temps médian.

$$x'_t = \frac{1}{3}(x_{t-1} + x_t + x_{t+1})$$

Moyennes mobiles

t_i	1	2	3	4	5	6	7	8	9
x'_t		68,0	70,0	82,0	85,0	91,0	88,0	87,0	80,0
t_i	10	11	12	13	14	15	16	17	18
x'_t	80,0	89,0	96,7	99,3	97,0	95,0	98,3	99,7	

Représentation graphique



59. ÉVOLUTION D'UN INDICE DE PRODUCTION INDUSTRIELLE

Mots-clefs

Moyennes mobiles

Énoncé

La série suivante retrace les fluctuations de l'indice trimestriel pour la production industrielle au cours de 6 années.

Les indices de production industrielle

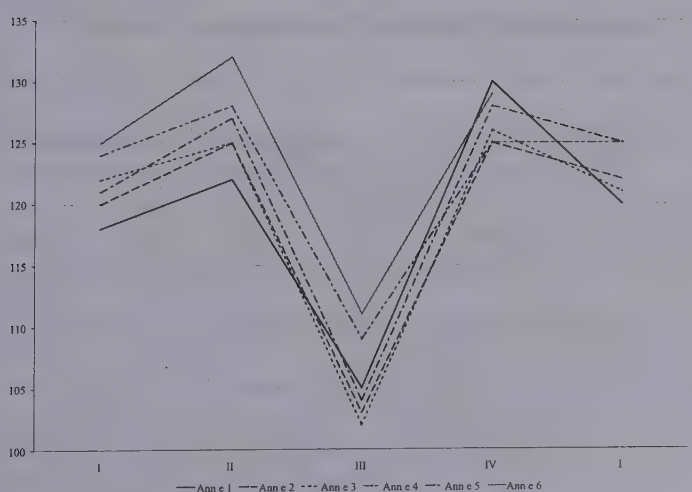
	I	II	III	IV
1	118	122	105	130
2	120	125	103	125
3	122	125	102	126
4	121	127	104	128
5	124	128	109	125
6	125	132	111	129

1. Montrez le caractère saisonnier de cette chronique.
2. Calculez le *trend* selon la méthode des moyennes mobiles.
3. Représentez graphiquement la nouvelle distribution.

Corrigé

1. Pour montrer le caractère saisonnier de cette chronique, une solution consiste à construire le graphique pour chaque année, avec, en abscisses, les quatre trimestres de chaque année et le premier trimestre de l'année suivante.

Graphique des données brutes



On voit clairement que la série est saisonnière, c'est-à-dire que les variations se produisent identiques à elles-mêmes d'une année sur l'autre (périodique de période = 1 an).

2. Le calcul du *trend* utilise les moyennes mobiles d'ordre 4 afin de lisser les variations sur une année.

$$\begin{cases} S_1 = y_1 + y_2 + y_3 + y_4 \\ S_2 = y_2 + y_3 + y_4 + y_5 \end{cases}$$

$$y'_3 = \frac{1}{8}(S_1 + S_2) = \frac{1}{4}\left(\frac{1}{2}y_1 + y_2 + y_3 + y_4 + \frac{1}{2}y_5\right)$$

Par exemple, pour calculer la moyenne mobile du quatrième trimestre de l'année 1, soit $y'_{1/IV}$:

$$S_{1/IV} = 122 + 105 + 130 + 120 = 477$$

$$S_{2/I} = 105 + 130 + 120 + 125 = 480$$

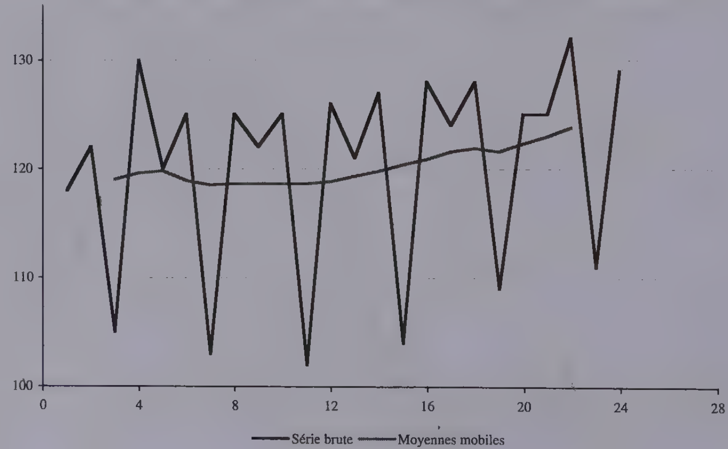
$$y'_{1/IV} = \frac{1}{8}(S_{1/II} + S_{1/III}) = \frac{477 + 480}{8} = 119{,}6$$

Moyennes mobiles

	I	II	III	IV
1			119,0	119,6
2	119,8	118,9	118,5	118,6
3	118,6	118,6	118,6	118,8
4	119,3	119,8	120,4	120,9
5	121,6	121,9	121,6	122,3
6	123,0	123,8		

3.

Représentation graphique des moyennes mobiles



Le lissage est visible sur ce deuxième graphique, les tendances sont plus nettes, les fluctuations conjoncturelles ou accidentelles ont été en quelque sorte gommées par l'utilisation des moyennes mobiles.

60. ESPÉRANCE DE VIE

Mots-clefs

Ajustement linéaire

Énoncé

L'évolution de l'espérance de vie à la naissance pour les femmes est donnée par le tableau suivant :

Espérance de vie des femmes

Année	1896	1902	1906	1910	1926	1930	1936	1950	1986
Espérance de vie	47,8	48,3	50,1	52,3	56,8	59,4	61,3	67,5	79,5

1. Dites rapidement ce que signifie la notion d'espérance de vie à la naissance.
2. Calculez l'équation de la droite d'ajustement de l'espérance de vie par rapport au temps.
3. Expliquez la pertinence de cet ajustement en calculant le coefficient de corrélation linéaire.
4. Si rien ne devait se modifier, qu'elle devrait être l'espérance de vie des femmes en 1990 ?

Corrigé

1. L'espérance de vie à la naissance est la moyenne, à une date donnée, de la distribution des âges au moment de la mort.
2. Pour calculer l'équation de la droite de l'espérance de vie en fonction du temps, nous devons construire le tableau suivant :

Tableau des calculs

Année	Temps	Espérance de vie				Espérance estimée
	t_i	e_v	t_i^2	e_v^2	$e_v \cdot t_i$	e_v^{est}
1896	0	47,8	0	2284,84	0	46,8
1902	6	48,3	36	2332,89	289,8	49,0
1906	10	50,1	100	2510,01	501,0	50,5
1910	14	52,3	196	2735,29	732,2	51,9
1926	30	56,8	900	3226,24	1704,0	57,8
1930	34	59,4	1 156	3528,36	2019,6	59,2
1936	40	61,3	1 600	3757,69	2452,0	61,4
1950	54	67,5	2 916	4556,25	3645,0	66,6
1986	90	79,5	8 100	6320,25	7155,0	79,7
1990	94					81,2
Total	278	523,0	15 004	31251,82	18498,6	

L'année 1896 est le temps $t = 0$. La droite d'ajustement de l'espérance de vie par rapport au temps a pour équation (attention il n'y a que 9 valeurs) :

$$a = \frac{\sum_{t=1}^{9} e_v \cdot t - 9 \cdot \overline{e}_v \cdot \overline{t}}{\sum_{t=1}^{9} t^2 - 9 \cdot \overline{t}^2} = 0{,}37$$

$$b = \overline{e}_v - a\overline{t} = 58{,}11 - 0{,}37 \cdot 30{,}89 = 46{,}83$$

$$\widehat{e_v} = 0{,}37t + 46{,}83$$

3. Le coefficient de corrélation linéaire est donné par la formule suivante :

$$r = \frac{\sum_{t=1}^{9} e_v \cdot t - 9 \cdot \overline{e}_v \cdot \overline{t}}{\sqrt{\sum_{t=1}^{9} t^2 - 9 \cdot \overline{t}^2} \cdot \sqrt{\sum_{t=1}^{9} e_v^2 - 9 \cdot \overline{e}_v^2}} = 0{,}998$$

Le coefficient de détermination r^2 est proche de 0,996. L'ajustement apparaît pertinent puisque 99,6 % de la variance est expliquée par l'ajustement linéaire.

4. L'estimation de l'espérance de vie en 1990 s'obtient par application de notre équation. Les espérances de vie estimées apparaissent dans la dernière colonne du tableau ci-dessus. Cette espérance en 1990 serait de 81,2 ans.

Les données réelles donnent en 1990 une espérance de vie pour les femmes de 80,9 ans, l'ajustement se révèle pertinent.

61. ÉVALUATION DE LA RELATION ENTRE LES EXPORTATIONS ET LES IMPORTATIONS

Mots-clefs

Ajustement linéaire

Énoncé

Pour l'analyse du commerce extérieur de la branche des « Matériels électriques et électroniques professionnels », les instituts de statistiques utilisent deux indicateurs : le taux de pénétration du marché intérieur et le taux d'effort à l'exportation.

$$\text{taux de pénétration} = \frac{\text{Importations}}{\text{Marché intérieur}}$$

$$\text{effort de l'exportation} = \frac{\text{Exportations}}{\text{Production distribuée}}$$

Le tableau ci-dessous retrace l'évolution de ces deux indicateurs.

Tableau des indices

Année	Taux de pénétration	Effort à l'exportation
1978	26,9	29,8
1979	27,4	30,1
1980	28,4	29,6
1981	29,5	30,8
1982	32,0	31,0
1983	34,5	34,8
1984	35,9	36,0
1985	36,9	37,0
1986	37,8	36,9
1987	41,1	39,3
1988	42,5	38,0

Source : Insee, 1989

1. Calculez le coefficient de corrélation linéaire entre ces deux indicateurs.
2. Calculez les équations des droites d'ajustement en fonction du temps.
3. Qu'en pensez-vous ?

Corrigé

1. Le tableau ci-dessous nous permet de disposer des éléments nécessaires au calcul du coefficient de corrélation.

Tableau des calculs

Taux de pénétration	Effort à l'exportation			
x_i	y_i	x_i^2	y_i^2	$x_i y_i$
26,9	29,8	723,61	888,04	801,6
27,4	30,1	750,76	906,01	824,7
28,4	29,6	806,56	876,16	840,6
29,5	30,8	870,25	948,64	908,6
32,0	31,0	1024,00	961,00	992,0
34,5	34,8	1190,25	1211,04	1200,6
35,9	36,0	1288,81	1296,00	1292,4
36,9	37,0	1361,61	1369,00	1365,3
37,8	36,9	1428,84	1361,61	1394,8
41,1	39,3	1689,21	1544,49	1615,2
42,5	38,0	1806,25	1444,00	1615,0
372,9	373,3	12940,15	12805,99	12850,8

Le coefficient de corrélation est le rapport de la covariance aux écart type.

$$r = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2} \cdot \sqrt{\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2}}$$

$$r = \frac{12\,850,8 - \dfrac{372,9 \cdot 373,3}{11}}{\sqrt{12\,940,15 - \dfrac{372,9^2}{11}} \cdot \sqrt{12\,805,99 - \dfrac{373,3^2}{11}}} = 0,966$$

Le coefficient de corrélation linéaire est de 0,9883, le coefficient de détermination est de 0,934.

2. Les coefficients de l'équation de la droite d'ajustement du taux de pénétration en fonction du temps sont obtenus à partir de ce tableau. Nous avons effectué

un changement d'origine pour le temps pour simplifier les calculs, l'année 1978 devient le temps $t = 1$.

Calcul pour la droite d'ajustement

t_i	x_i	t_i^2	$x_i t_i$
1	26,9	1	29,6
2	27,4	4	54,8
3	28,4	9	85,2
4	29,5	16	118,0
5	32,0	25	160,0
6	34,5	36	207,0
7	35,9	49	251,3
8	36,9	64	295,2
9	37,8	81	340,2
10	41,1	100	411,0
11	42,5	121	467,5
66	372,9	506	2417,1

Le coefficient a de la droite d'ajustement $\hat{x} = at + b$ est :

$$a = \frac{Cov(x,t)}{V(t)} = \frac{\sum_{i=1}^{n} x_i t_i - n\overline{x}\overline{t}}{\sum_{i=1}^{n} t_i^2 - n\overline{t}^2} = \frac{2417{,}1 - 11 \cdot 33{,}9 \cdot 6}{506 - 11 \cdot 6^2} = 1{,}63$$

Le coefficient b :

$$b = \overline{x} - a\overline{t} = 33{,}9 - 1{,}63 \cdot 6 = 24{,}12$$

L'équation de l'évolution du taux de pénétration en fonction du temps est alors :

$$\hat{x} = 1{,}63t + 24{,}12$$

Tableau des calculs pour l'ajustement du taux de pénétration

t_i	y_i	t_i^2	$y_i t_i$
1	29,8	1	29,8
2	30,1	4	60,2
3	29,6	9	88,8
4	30,8	16	123,2
5	31,0	25	155,0
6	34,8	36	208,8
7	36,0	49	252,0
8	37,0	64	296,0
9	36,9	81	332,1
10	39,3	100	393,0
11	38,0	121	418,0
66	373,3	506	2356,9

Le coefficient a' de la droite d'ajustement $\hat{y} = a't + b'$ est :

$$a' = \frac{Cov(y,t)}{V(t)} = \frac{\sum_{i=1}^{n} y_i t_i - n\overline{y}\overline{t}}{\sum_{i=1}^{n} t_i^2 - n\overline{t}^2} = \frac{2356,9 - 11 \cdot 33,94 \cdot 6}{506 - 11 \cdot 6^2} = 1,06$$

Le coefficient b' est alors :

$$b' = \overline{y} - a'\overline{t} = 33,94 - 1,065 \cdot 6 = 27,55$$

L'équation de la droite d'ajustement de l'effort à l'exportation en fonction du temps est :

$$\hat{y} = 1,06t + 27,55$$

Nous pourrions vérifier que les coefficients de détermination entre le taux de pénétration et l'effort à l'exportation avec le temps sont proches de 1, donc que les régressions sont satisfaisantes.

3. L'analyse statistique précédente montre que, sur la période considérée, les indicateurs du commerce extérieur évoluent identiquement. L'effort à l'exportation permet de déterminer avec un bon niveau de probabilités le taux de pénétration, ce qui signifie que les deux variables sont liées. Dans la situation de l'économie française, l'effort à l'exportation n'apparaît pas comme la solution pour réduire les taux de pénétration dont la couverture des importations. La croissance simultanée des deux taux traduit quantitativement le processus de mondialisation en cours.

62. Droite d'ajustement et prévision

Mots-clefs

Ajustement linéaire, coefficient de détermination, prévision

Énoncé

Les effectifs de la branche « Textile - habillement »

Années	1981	1982	1983	1984	1985	1986	1987	1988
Effectif salarié (en milliers)	495	482	468	447	428	411	391	370

Source : Insee, 1989

1. Calculez l'équation de la droite d'ajustement du nombre de salariés en fonction du temps.
2. Cette droite constitue-t-elle une bonne estimation de la tendance de l'emploi salarié dans la branche considérée ?
3. Si rien ne change, quel devrait être l'emploi salarié dans la branche « Textile - habillement » en 1990.

Corrigé

1. Les coefficients de l'équation de la droite d'ajustement du taux de pénétration en fonction du temps sont obtenus à partir de ce tableau. Nous avons effectué un changement d'origine pour le temps pour simplifier les calculs, l'année 1981 devient le temps $t = 1$.

t_i	x_i	t_i^2	$x_i t_i$
1	495	1	495
2	482	4	964
3	468	9	1 404
4	447	16	1 788
5	428	25	2 140
6	411	36	2 466
7	391	4	2 737
8	370	64	2 960
36	3 492	204	14 954

Le coefficient a de la droite d'ajustement $\hat{x} = at + b$ est :

$$a = \frac{\sum_{i=1}^{n} x_i t_i - n\bar{x}\bar{t}}{\sum_{i=1}^{n} t_i^2 - n\bar{t}^2} = \frac{14954 - \frac{36 \cdot 3492}{8}}{204 - \frac{36^2}{8}} = -18,095$$

Le coefficient b :

$$b = \bar{y} - a \cdot \bar{t} = \frac{3492}{8} + 18,095 \cdot \frac{36}{8} = 517,9$$

L'équation de l'évolution du taux de pénétration en fonction du temps est alors :

$$\hat{x} = -18,1t + 517,9$$

2. Nous considérerons que la droite d'ajustement constitue une bonne estimation de la tendance de l'emploi dans la branche considérée si le coefficient de détermination est supérieur à 0,9.

Années	1	2	3	4	5	6	7	8
x_i	495	482	468	447	428	411	391	370
x_i^2	245 025	232 324	219 024	199 809	183 184	168 921	152 881	136 900

$$r = \frac{\sum_{i=1}^{n} x_i t_i - n\bar{x}\bar{t}}{\sqrt{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}\sqrt{\sum_{i=1}^{n} t_i^2 - n\bar{t}^2}}$$

$$r = \frac{14954 - \frac{36 \cdot 3492}{8}}{\sqrt{204 - \frac{36^2}{8}}\sqrt{1538068 - \frac{3492^2}{8}}} = -0,998$$

Le coefficient de corrélation linéaire est de – 0,9979. Le coefficient de détermination est de 0,996. La régression est satisfaisante.

3. La valeur du coefficient de détermination permet de calculer une estimation correcte de l'emploi en 1990, de plus cette date n'est pas trop éloignée de la période d'estimation ; donc nous pouvons considérer que les choses n'ont pas fondamentalement changé. En appliquant les résultats précédents, l'année 1990 est l'année $t = 10$ donc l'emploi estimé par la régression est de :

$$\hat{x}_{90} = -18,1 \cdot 10 + 517,9 = 336,9 \cong 337 \text{ milliers}$$

L'emploi dans la branche textile était de 361 milliers, les tendances se sont donc modifiées.

63. CORRECTIONS SAISONNIÈRES DE VENTES DE MARCHANDISES

Mots-clefs

Coefficients saisonniers, composition additive, moyennes mobiles, tendance

Énoncé

La série des indices trimestriels des ventes de marchandises d'une entreprise est fournie pour trois années par le tableau suivant.

Ventes trimestrielles de marchandises

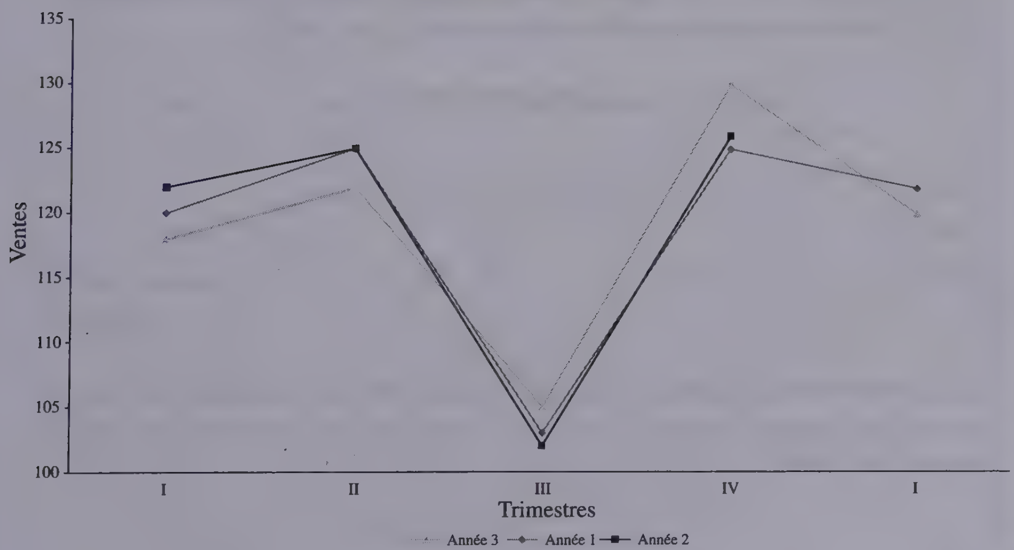
Trimestre \ Année	Année 1	Année 2	Année 3
I	118,2	148,6	163,3
II	129,0	154,5	175,3
III	138,9	163,0	189,1
IV	157,1	184,0	217,9

1. Mettez en lumière graphiquement le caractère saisonnier de cette série.
2. En supposant que le modèle de composition des mouvements soit additif, calculez la série corrigée des variations saisonnières. La tendance sera estimée par le calcul des moyennes mobiles.

Corrigé

1. Le graphique est construit sur cinq trimestres afin de faire apparaître la continuité entre les années.

Graphique de la série



On constate que les variations annuelles sont périodiques, de période une année, ce qui traduit la saisonnalité de la série.

2. La méthode consiste donc à calculer les moyennes mobiles pour chaque trimestre, en supposant que la composition des mouvements est additive.

Les moyennes mobiles sur quatre trimestres se définissent de la manière suivante :

$$y'_t = \frac{1}{4}\left(\frac{y_{t-2}}{2} + y_{t-1} + y_t + y_{t+1} + \frac{y_{t+2}}{2}\right)$$

La composition des mouvements étant additive le mouvement résultant s'écrit :

$$y_{ik} = E_{ik} + S_k + Z_{ik}$$

Données brutes y_{ik}

	I	II	III	IV
Année 1	118,2	129,0	138,9	157,1
Année 2	148,6	154,5	163,0	184,0
Année 3	163,3	175,3	189,1	217,9

Les moyennes mobiles sur quatre trimestres y'_{ik}

	I	II	III	IV
Année 1			139,6	146,6
Année 2	152,8	159,2	164,4	168,8
Année 3	174,7	182,2		

Les coefficients bruts sont obtenus en faisant la différence entre la valeur observée et la valeur obtenue par la méthode de la moyenne mobile.

Les coefficients bruts S'_{ik}

	I	II	III	IV
Année 1			– 0,7	10,5
Année 2	– 4,2	– 4,7	– 1,4	15,2
Année 3	– 11,4	– 7,0		
S'_k	– 7,8	– 5,9	– 1,1	12,9

La somme des coefficients bruts S'_k est –1,9. Pour être satisfaisant, le total devrait être 0, donc il est nécessaire de rectifier les coefficients de manière que leur somme soit nulle.

$$S_k = S'_k - \overline{S}', \text{ avec } \overline{S}' = -0{,}475$$

Les coefficients définitifs

	I	II	III	IV
S_k	– 7,3	– 5,4	– 0,6	13,3

La série CVS se calcule à partir de la série brute en retranchant à chaque terme de la série brute le coefficient saisonnier correspondant.

Série CVS y_{ik}^{CVS}

	I	II	III	IV
Année 1	125,5	134,4	139,5	143,8
Année 2	155,9	159,9	163,6	170,7
Année 3	170,6	180,7	189,7	204,6

64. Correction des variations saisonnières du chômage

Mots-clefs

Coefficients saisonniers, modèle multiplicatif, moyennes mobiles, variance

Énoncé

1. Vérifiez le comportement saisonnier du phénomène.
2. Déterminez les coefficients saisonniers, en posant l'hypothèse d'une composition multiplicative des composantes de la chronique, par la méthode des moyennes mobiles sur quatre trimestres.
3. Calculez alors la série CVS.
4. Représentez sur le même graphique la série brute et la série CVS.

D.E.F.M. données trimestrielles (en milliers)

Trimestres \ Années	1	2	3	4	5	6
I		2 252	2 482	2 509	2 702	2 624
II		2 184	2 301	2 386	2 527	2 437
III		2 280	2 335	2 499	2 579	2 470
IV	2 227	2 522	2 479	2 677	2 680	

Source : Statistiques du travail

Corrigé

1. Un tableau permet de mettre en évidence le caractère saisonnier du mouvement des D.E.F.M. Nous calculons les rapports inter trimestriels comme indiqué ci-dessous.

Rapports trimestriels

	I/IV	II/I	III/II	IV/III
Année 2	1,01	0,97	1,04	1,11
Année 3	0,98	0,96	1,01	1,06
Année 4	1,01	0,95	1,05	1,07
Année 5	1,01	0,93	1,02	1,04
Année 6	0,98	0,93	1,01	
Moyennes	0,998	0,948	1,026	1,07
Variances	0,000 216	0,000 256	0,000 264	0,000 65

Le calcul des coefficients d'évolution relatifs et de leur moyenne montrent des mouvements nettement saisonnalisés, chaque année les coefficients inter trimestriels relatifs sont proches. Cette première impression est confirmée par le calcul des variances.

La variance des cinq coefficients moyens trimestriels (variance des moyennes 0,002 614 333) est supérieure à la variance de ces coefficients pour chaque trimestre (moyenne des variances 0,000 346 5) au cours des cinq années. Nous avons là un indicateur des variations saisonnières de la série.

2. Nous effectuerons la désaisonnalisation par la méthode des moyennes mobiles sur quatre trimestres. Nous supposerons une composition multiplicative des mouvements.

D.E.F.M. Données brutes

	I	II	III	IV
Année 1				2 227
Année 2	2 252	2 184	2 280	2 522
Année 3	2 482	2 301	2 335	2 479
Année 4	2 509	2 386	2 499	2 677
Année 5	2 702	2 525	2 579	2 680
Année 6	2 624	2 437	2 470	

Les moyennes mobiles sont calculées de la manière suivante :

$$y'_t = \frac{1}{4}\left(\frac{y_{t-2}}{2} + y_{t-1} + y_t + y_{t+1} + \frac{y_{t+2}}{2}\right)$$

Moyennes mobiles (quatre trimestres) y'_t

	I	II	III	IV
Année 2		2 272,6	2 338,3	2 381,6
Année 3	2 403,1	2 404,6	2 402,6	2 416,6
Année 4	2 447,8	2 493,0	2 541,9	2 583,6
Année 5	2 611,3	2 621,8	2 612,5	2 591,5
Année 6	2 566,6			

Les coefficients bruts sont obtenus en faisant le rapport entre la moyenne mobile et la valeur brute.

Coefficients bruts S'_{ik}

	I	II	III	IV
Année 2		0,961	0,975	1,059
Année 3	1,033	0,957	0,972	1,026
Année 4	1,025	0,957	0,983	1,036
Année 5	1,035	0,964	0,987	1,035
Année 6	1,022			

Moyennes des coefficients bruts

	I	II	III	IV
S'_k	102,9	96,0	97,9	103,9

La somme des coefficients bruts est :

$$\sum_{k=1}^{4} S'_k = 400,66$$

Il faut donc rectifier les coefficients. Soit $\overline{S}' = 100,175$ la moyenne des coefficients bruts, nous obtenons les coefficients définitifs S_k à partir des coefficients bruts par division par la moyenne des coefficients bruts soit :

$$S_k = \frac{S'_k}{\overline{S}'} \times 100$$

Coefficients définitifs

	I	II	III	IV
S_k	102,7	95,8	97,8	103,7

3. Nous obtenons alors la série corrigée des variations saisonnières.

$$y_{ik}^{CVS} = \frac{y_{ik}}{S_k}$$

Série CVS

	I	II	III	IV
Année 1				2 147
Année 2	2 193	2 279	2 332	2 432
Année 3	2 417	2 402	2 388	2 390
Année 4	2 443	2 490	2 556	2 581
Année 5	2 631	2 637	2 638	2 585
Année 6	2 555	2 544	2 526	

65. PRÉVISION DE L'ÉVOLUTION D'UN C.A.

Mots-clefs

Correction des variations saisonnières, composition multiplicative, tendance, moyennes mobiles

Énoncé

Évolution trimestrielle du C.A.

Trimestres	I	II	III	IV
Année 1	112,8	123,6	130,3	115,2
Année 2	122,1	132,4	138,8	123,4
Année 3	134,1	145,4	152,5	135,6
Année 4	138,2	150,3	157,1	140,5

1. Quel procédé rapide permet de montrer les fluctuations saisonnières ?
2. La composition des mouvements est multiplicative, le *trend* sera estimé par les moyennes mobiles sur quatre trimestres.

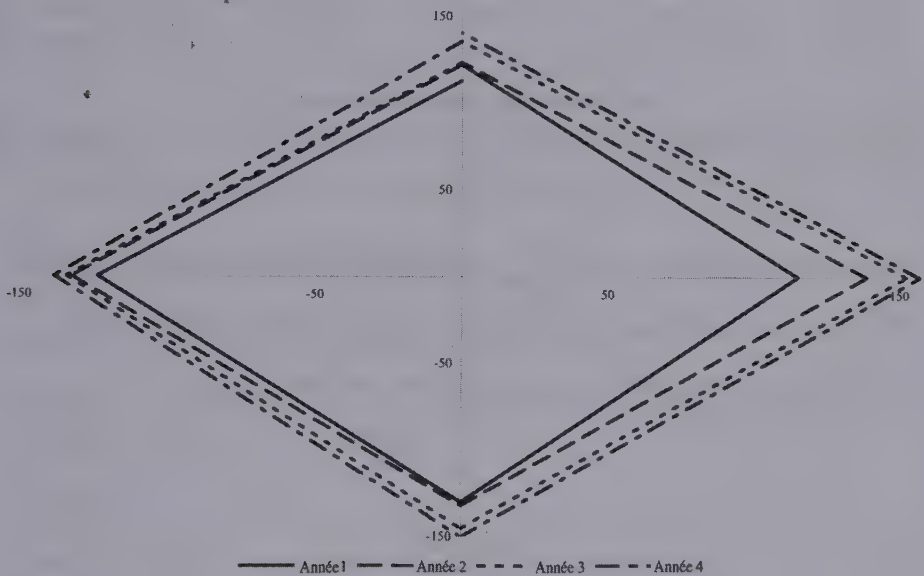
   a. Calculez les coefficients saisonniers.

   b. Calculez la série désaisonnalisée.
3. Pour une croissance de 5 %, par rapport à l'année précédente pour la même période, quelle doit être la valeur réelle du C.A. pour le premier trimestre de l'année 5.

Corrigé

1. Un graphique permet de visualiser de manière satisfaisante le caractère saisonnier du mouvement des C.A., nous utiliserons un diagramme.

Graphique des évolutions du CA



2. La désaisonnalisation s'effectuera en calculant les moyennes mobiles sur quatre trimestres en supposant la composition multiplicative des mouvements.

Moyennes mobiles sur quatre trimestres y'_t

Trimestres	I	II	III	IV
Année 1			121,64	123,9
Année 2	126,06	128,15	130,67	133,8
Année 3	137,14	140,37	142,41	143,54
Année 4	144,72	145,91		

a. Les moyennes mobiles sont calculées de la manière suivante :

$$y'_t = \frac{1}{4}\left(\frac{y_{t-2}}{2} + y_{t-1} + y_t + y_{t+1} + \frac{y_{t+2}}{2}\right)$$

Les coefficients bruts sont obtenus en faisant le rapport entre la moyenne mobile et la valeur brute.

Coefficients bruts S'_{ik}

	I	II	III	IV
Année 1			1,07	0,93
Année 2	0,97	1,03	1,06	0,92
Année 3	0,98	1,04	1,07	0,94
Année 4	0,95	1,03		

Moyennes des coefficients bruts

S'_k	0,97	1,03	1,07	0,93

La somme des coefficients bruts $\sum_{k=1}^{4} S'_k$ est de 400, ce qui ne nécessite aucune rectification des coefficients S'_k qui donc sont les coefficients définitifs S_k.

b. Nous obtenons alors la série corrigée des variations saisonnières.

$$y_{ik}^{CVS} = \frac{y_{ik}}{s_k}$$

Série CVS

	I	II	III	IV
Année 1	116,3	120,0	121,8	123,9
Année 2	125,9	128,5	129,7	132,7
Année 3	138,6	141,2	142,5	145,8
Année 4	142,5	145,9	146,8	151,1

3. Pour une croissance de 5 % sur l'année, le calcul 142,5 .1,05 = 149,63 donne la valeur en terme corrigé des variations saisonnières pour le trimestre considéré. En terme réel, nous devons le multiplier par le coefficient saisonnier soit 0,97. Le C.A. réel du premier trimestre de l'année 5 sera 145,1.

66. Évaluation d'un indice de production

Mots-clefs

Correction des variations saisonnières, moyennes mobiles, ajustement, prévision

Énoncé

La série suivante retrace les fluctuations de l'indice trimestriel pour la production industrielle au cours de 6 années.

	I	II	III	IV
Année 1	110	115	113	122
Année 2	118	129	123	133
Année 3	127	141	133	147
Année 4	139	152	142	159
Année 5	149	163	150	172
Année 6	157	176	159	183

1. Montrez le caractère saisonnier de cette série et son modèle de composition.

2. Calculez le *trend* selon la méthode des moyennes mobiles.

3. Calculez les coefficients saisonniers.

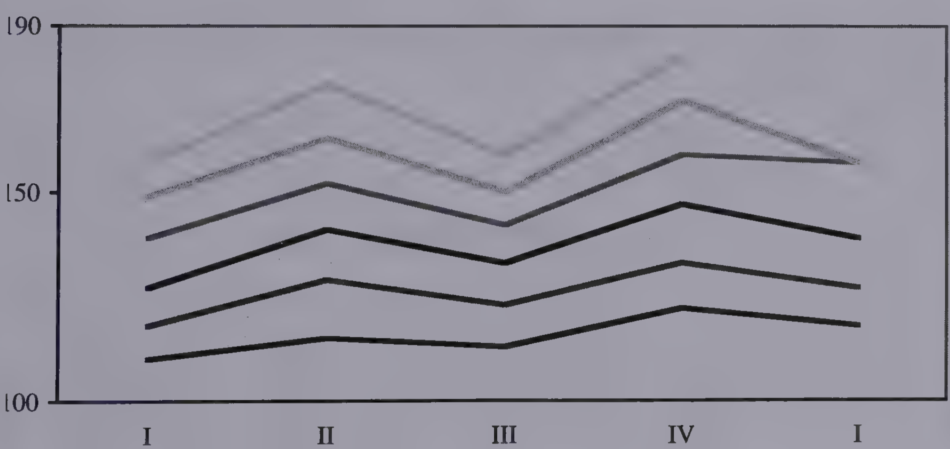
4. Calculez la chronique corrigée des variations saisonnières et déterminez l'équation de la droite d'ajustement de la série CVS.

5. Évaluez la valeur brute de l'indice de production industriel pour le second semestre de la huitième année.

Corrigé

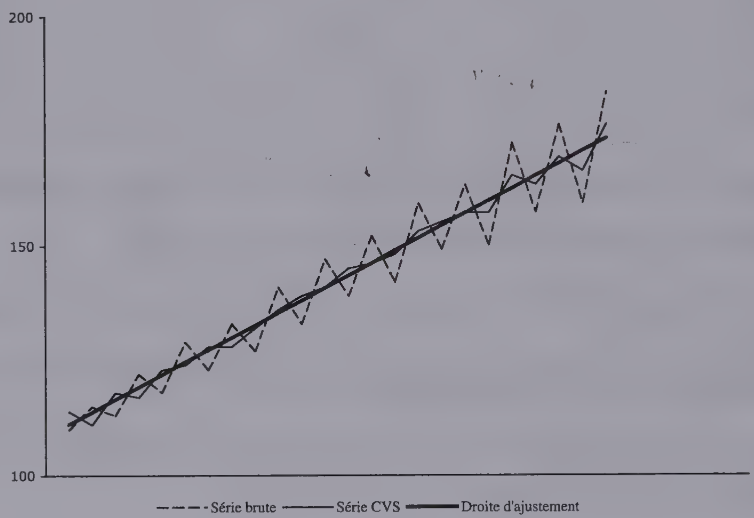
1. Mise en lumière du caractère saisonnier de la chronique

Graphique trimestriel



La représentation graphique de la série permet de choisir le modèle de composition.

Évolution de la série



Il apparaît que les fluctuations ont tendance à s'amplifier au cours du temps, ce qui nous oriente vers une composition multiplicative des mouvements.

$$y_{ik} = E_{ik} \cdot S_{ik} + Z_{ik}$$

2. Les moyennes sont obtenues par l'application de la formule des moyennes mobiles d'ordre 4 :

$$y'_t = \frac{1}{4}\left(\frac{y_{t-2}}{2} + y_{t-1} + y_t + y_{t+1} + \frac{y_{t+2}}{2}\right)$$

Moyennes mobiles (quatre trimestres)

	I	II	III	IV
1			116,00	118,75
2	121,75	124,38	126,88	129,50
3	132,25	135,25	138,50	141,38
4	143,88	146,50	149,25	151,88
5	154,25	156,88	159,50	162,13
6	164,88	167,38		

Nous déterminons ainsi les E_{ik}.

3. Les coefficients saisonniers sont obtenus en faisant le rapport de $S'_{ik} = \dfrac{y_{ik}}{E_{ik}}$

Coefficients bruts

	I	II	III	IV
1			97,41	102,74
2	96,92	103,72	96,95	102,70
3	96,03	104,25	96,03	103,98
4	96,61	103,75	95,14	104,69
5	96,60	103,90	94,04	106,09
6	95,22	105,15		

Nous calculons les coefficients bruts moyens S'_k

Coefficients bruts moyens

	I	II	III	IV
S'_k	96,28	104,16	95,91	104,04

Somme des coefficients bruts moyens : $\sum_{k=1}^{4} S'_k = 400{,}39$

Nous sommes donc conduits à rectifier les coefficients bruts pour obtenir les coefficients définitifs.

$$S_k = \frac{S'_k}{\overline{S}} = \frac{S'_k}{100{,}1}$$

Coefficients rectifiés

	I	II	III	IV
S_k	96,18	104,06	95,82	103,94

4. Nous pouvons calculer la série corrigée des variations saisonnières.

$$y_{ik}^{CVS} = \frac{y_{ik}}{S_k}$$

Série CVS

	I	II	III	IV
1	114	111	118	117
2	123	124	128	128
3	132	136	139	141
4	145	146	148	153
5	155	157	157	165
6	163	169	166	176

Ayant éliminées les fluctuations annuelles nous pouvons calculer l'équation de la droite d'ajustement de la série CVS. Nous choisissons pour temps 1 le premier trimestre de l'année 1. Nous obtenons l'équation de la droite d'ajustement :

$$\hat{y}_{ik}^{CVS} = 2{,}69t + 108{,}5$$

La représentation graphique de la droite d'ajustement et de la série CVS indique une bonne relation entre la droite et la série (voir graphique précédent).

La valeur du coefficient de détermination linéaire, $r^2 = 0{,}9945$, confirme la validité de cet ajustement

5. L'évaluation de la valeur brute de l'indice du second trimestre de la huitième année se réalise en deux temps.

Tout d'abord nous évaluons sa valeur CVS en utilisant l'équation de l'ajustement, puis, en appliquant le coefficient saisonnier, nous obtenons une évaluation de la valeur cherchée.

Le second trimestre de l'année 8 correspond au temps $t = 30$ dont la valeur CVS évaluée est :

$$\hat{y}_{II/8}^{CVS} = 2{,}71 \cdot 30 + 113{,}72 = 195{,}07$$

Le coefficient du second trimestre est de 1,0406 donc :

$$\hat{y}_{II/8}^{brut} = 195{,}07 \cdot 1{,}0406 \cong 203$$

Ce résultat ne constitue pas la vraie valeur qui sera constatée, il constitue une prévision sur la valeur probable si les tendances antérieures se poursuivent.

Chapitre 5

Les indices

Les indices ont été conçus pour effectuer des comparaisons sur des variables économiques mesurables. Ils synthétisent en un seul nombre les modifications affectant un ensemble de variables. C'est le cas de l'indice des prix à la consommation. Les changements s'exprimant en grandeurs relatives, les comparaisons en sont facilitées.

Un indice simple est le rapport des valeurs prises par une grandeur entre deux dates, deux lieux, etc. Un indice synthétique, ou indice composé, est un indicateur de tendance centrale d'une distribution d'indices simples. Les indices synthétiques sont souvent des moyennes d'indices simples, moyenne arithmétique pour l'indice de Laspeyres, moyenne harmonique pour l'indice de Paasche.

67. Indice simple de valeur

Mots-clefs

Indice simple, indice de valeur

Énoncé

Calculez les indices de valeur pour les années 1 et 2 des différents produits et de la consommation totale en base 100 année 0.

Structure des dépenses par fonction de consommation (milliers d'euros)

	Année 0	Année 1	Année 2
A	450	500	550
B	360	400	450
C	150	250	300
D	240	350	500
Total	1 200	1 500	1 800

Corrigé

Nous avons ici à calculer des indices simples de valeur selon la forme générale des indices simples. Si nous appelons G_0^h la valeur de la consommation en produit h pour l'année 0, respectivement G_1^h et G_2^h pour les années 1 et 2, les indices pour le produit h seront les suivants :

$$i(v)_{1/0}^h = \frac{G_1^h}{G_0^h} \cdot 100$$

$$i(v)_{2/0}^h = \frac{G_2^h}{G_0^h} \cdot 100$$

Nous obtenons alors le tableau suivant :

Indices de valeur base 100 année 0

	Année 0	Année 1	Année 2
	$i(v)_{0/0}^h$	$i(v)_{1/0}^h$	$i(v)_{2/0}^h$
A	100,0	111,1	122,2
B	100,0	111,1	125,0
C	100,0	166,7	200,0
D	100,0	145,8	208,6
Total	100,0	125,0	150,0

Nous voyons que la consommation a augmenté de 25 % entre l'année 0 et l'année 1, qu'elle est de 50 % entre l'année 0 et l'année 2, ce qui se traduit par une augmentation de 20 % de la valeur de la consommation entre l'année 1 et l'année 2 $\left(\frac{150}{125} \cdot 100 = 120\right)$. Ces augmentations globales recouvrent des situations différentes selon les produits.

68. INDICES DE VOYAGEURS

Mots-clefs

Indice de quantité

Énoncé

Soit la distribution suivante du nombre de voyageurs (en millions) transportés en métro par trimestre au cours de quatre années consécutives.

Données de base

Années \ Trimestres	I	II	III	IV
1	48	44	40	44
2	56	52	44	48
3	64	52	44	56
4	72	68	48	72

1. Calculez en base 100 le premier trimestre de l'année 1 l'indice du nombre de passagers.
2. Même question pour un indice base 100 année 2.
3. Même question en base 100 pour l'ensemble des trimestres.

Corrigé

1. Ici, il s'agit bien de quantité et non de volume, car nous disposons du nombre de voyageurs quel que soit le prix qu'ils ont acquitté pour utiliser le métro, alors que souvent les chiffres disponibles reprennent une valeur dont on ne connaît que les prix.

Nous avons à calculer les indices du nombre de passagers en prenant pour période de référence le premier trimestre de l'année 1.

La définition générale d'un indice est :

$$i_{i,k/1,I} = \frac{G_{i,k}}{G_{1,I}} \cdot 100.$$

Ici $G_{1,I}$ est le nombre de passagers pour le premier trimestre de l'année 1 soit 48. $G_{i,k}$ correspond au nombre de passagers transportés l'année i pour le trimestre k.

L'indice pour le second trimestre de l'année 1 est donc :

$$i_{1,II/1,I} = \frac{44}{48} \cdot 100 = 91{,}7$$

ou encore :

$$i_{3,II/1,I} = \frac{G_{3,II}}{G_{1,I}} \cdot 100 = \frac{52}{48} \cdot 100 = 108{,}3$$

Nous obtenons alors le tableau des indices simples, il s'agit d'indices de quantité.

Indices base 100 =1,I

Années \ Trimestres	I	II	III	IV
1	100,0	91,7	83,3	91,7
2	116,7	108,3	91,7	100,0
3	133,3	108,3	91,7	116,7
4	150,0	141,7	100,0	150,0

2. Nous prenons comme base de référence le nombre moyen de passagers par trimestre au cours de l'année 2. La moyenne arithmétique du nombre de passagers de l'année est de 50. Nous allons maintenant calculer les indices de chaque trimestre en prenant pour base cette moyenne, selon la formule suivante :

$$i_{i,k} = \frac{G_{i,k}}{50} \cdot 100$$

L'indice du troisième trimestre de la deuxième année sera donc :

$$i_{2,III} = \frac{44}{50} \cdot 100 = 88$$

Indices 100 année 2

Années \ Trimestres	I	II	III	IV
1	96,0	88,0	80,0	88,0
2	112,0	104,0	88,0	96,0
3	128,0	104,0	88,0	112,0
4	144,0	136,0	96,0	144,0

3. Pour déterminer les indices en base 100 pour l'ensemble des trimestres, la démarche est identique à la celle de la question précédente. Il suffit de calculer le nombre moyen de passagers par trimestre soit 53,25 puis de calculer les différents indices. Ce chiffre est la moyenne arithmétique du nombre trimestriel de passagers au cours des quatre années.

La forme générale de l'indice sera donc :

$$i_{i,k} = \frac{G_{i,k}}{53,25} \cdot 100$$

Nous calculons comme exemple l'indice relatif au quatrième trimestre de la quatrième année.

$$i_{4,IV} = \frac{72}{53,25} \cdot 100 \cong 135,2$$

Indices 100 moyenne générale

Années \ Trimestres	I	II	III	IV
1	90,1	82,6	75,1	82,6
2	105,2	97,7	82,6	90,1
3	120,2	97,7	82,6	10,5,2
4	135,2	127,7	90,1	135,2

69. Indices de prix

Mots-clef

Indices simple, circularité

Énoncé

Nous disposons du tableau suivant des indices élémentaires de prix en base 100 l'année précédente. Nous voulons obtenir les indices de prix en base 100 l'année 0.

Indices des prix en base 100 année précédente

	Année 0	Année 1	Année 2	Année 3
A	101	103	101	102
B	102	104	100	101
C	103	102	102	103
D	102	102	104	99

Corrigé

La propriété de circularité des indices simples nous permet de répondre à la question. Nous savons que dans le cas des indices simples, l'indice de l'année 3 en base 100 l'année 0 peut s'écrire :

$$i(p)^h_{3/0} = i(p)^h_{3/2} \cdot i(p)^h_{2/1} \cdot i(p)^h_{1/0}.$$

Pour faciliter la compréhension des calculs à venir, nous explicitons dans le tableau la signification des données.

Indices de prix en base 100 année précédente

	Année 0	Année 1	Année 2	Année 3
Indices	$i(p)^h_{0/-1}$	$i(p)^h_{1/0}$	$i(p)^h_{2/1}$	$i(p)^h_{3/2}$
A	101	103	101	102
B	102	104	100	101
C	103	102	102	103
D	102	102	104	99

Pour le produit A, l'application de cette propriété permet d'obtenir les indices suivants : $i(p)^A_{0/0} = 100$ (propriété d'identité), l'indice $i(p)^A_{1/0}$ est connu, il est fourni, en l'occurrence 103. L'indice $i(p)^A_{2/0}$ est obtenu en appliquant la propriété de circularité :

$$i(p)^A_{2/0} = i(p)^A_{1/0} \cdot i(p)^A_{2/1} = 1{,}03 \cdot 1{,}01 \cdot 100 = 104{,}0 .$$

De même pour l'indice $i(p)^A_{3/0}$:

$$i(p)^A_{3/0} = i(p)^A_{1/0} \cdot i(p)^A_{2/1} \cdot i(p)^A_{3/2} = 1{,}03 \cdot 1{,}01 \cdot 1{,}02 \cdot 100 = 106{,}1 .$$

Nous obtenons alors le tableau suivant :

Indices base 100 année 0

$i(p)^h_{i/0}$	Année 0	Année 1	Année 2	Année 3
A	100,0	103,0	104,0	106,1
B	100,0	104,0	104,0	105,0
C	100,0	102,0	104,0	107,2
D	100,0	102,0	106,1	105,0

70. INDICES DE DÉPENSE

Mots-clefs

Indice de volume, indice de prix, indice de quantité, indice simple

Énoncé

Un amateur de prunes « Reine Claude » suit l'évolution du prix de son fruit préféré au cours des mois de disponibilité.

Indices des prix en base 100 année précédente

	Juin	Juillet	Août	Septembre
Prix (en francs)	25	20	15	19
Quantités achetées (en kg)	2	6	10	5

Quelle est l'évolution de l'indice de la dépense de consommation de cette variété de prune en base 100 en juin ?

Corrigé

Pour calculer l'indice de la dépense nous allons la décomposer en ces deux composantes (prix et quantités). Nous sommes donc conduits à calculer les indices de prix et les indices de quantités. Nous obtiendrons ensuite l'indice de la dépense en multipliant ces deux indices, par application de la propriété des indices simples. Pour simplifier les écritures nous affectons un code aux différents mois.

Codes mensuels

	Juin	Juillet	Août	Septembre
Code	0	1	2	3

Pour calculer les indices simples nous faisons le rapport des différentes grandeurs pour les périodes considérées. Nous donnons en exemple les calculs pour le mois d'août codé 2.

$$i(p)_{2/0} = \frac{p_2}{p_0} \cdot 100 = \frac{15}{25} \cdot 100 = 60$$

$$i(q)_{2/0} = \frac{q_2}{q_0} \cdot 100 = \frac{10}{2} \cdot 100 = 500$$

$$i(v)_{2/0} = i(p)_{2/0} \cdot i(q)_{2/0} = 0{,}6 \cdot 5 \cdot 100 = 300$$

Indices base 100 en juin

	Juin	Juillet	Août	Septembre
Indices des prix	100	80	60	76
Indices des quantités	100	300	500	250
Indices de la dépense	100	240	300	190

71. Indices de prix de la VAB

Mots-clefs

Indice de prix, circularité, réversibilité, indice simple

Énoncé

Le tableau fournit la hausse annuelle des prix de la valeur ajoutée.

Croissance des prix de la VAB en %

	1983/1982	1984/1983	1985/1984	1986/1985	1987/1986
Agriculture	6,9	– 2,7	1,2	3,4	– 3,4
I.A.A.	8,6	9,3	3,6	– 0,4	– 0,6
Énergie	11,1	8,6	16,2	11,4	– 1,1
Biens intermédiaires	8,5	7,3	8,3	6,4	2,2
Biens d'équipement	9,6	9,1	8,2	2,5	1,9
Biens de consommation	8,3	5,6	9,2	12,5	5,1
B.G.C.A.	5,5	5,6	4,7	4,7	4,6
Commerce	11,4	8,7	5,2	5,2	4,9
Transports	8,0	7,0	6,4	3,9	– 9,9
Services marchands	12,3	9,5	6,4	5,5	4,4

Calculez les indices des prix de la VAB, que vous considérez comme des indices simples, en base 100 en 1985.

Corrigé

La solution la plus simple est de calculer les indices de prix en base 100 l'année précédente, puis d'utiliser la propriété de circularité et de réversibilité des indices simples. La transformation consiste à passer des taux aux multiplicateurs (les indices sont une manière de présenter des multiplicateurs).

Indices en base 100 année précédente

	1983	1984	1985	1986	1987
	$i(p)^h_{83/82}$	$i(p)^h_{84/83}$	$i(p)^h_{85/84}$	$i(p)^h_{86/85}$	$i(p)^h_{87/86}$
Agriculture	106,9	97,3	101,2	103,4	96,6
I.A.A.	108,6	109,3	103,6	99,6	99,94
Énergie	111,1	108,6	116,2	111,4	98,9
Biens intermédiaires	108,5	107,3	108,3	106,4	102,2
Biens d'équipement	109,6	109,1	108,2	102,5	101,9
Biens de consommation	108,3	105,6	109,2	112,5	105,1
B.G.C.A.	105,5	105,6	104,7	104,7	104,6
Commerce	111,4	108,7	105,2	105,2	104,9
Transports	108,0	107,0	106,4	103,9	90,1
Services marchands	112,3	109,5	106,4	105,5	104,4

Disposant des indices en base 100 l'année précédente, il est facile de calculer les indices en base 100 1985. Nous explicitons les calculs pour la branche « Commerce », codée C :

– l'indice $i(p)^C_{86/85}$ est obtenu directement, soit 105,2 ;

– l'indice $i(p)^C_{87/85}$ est obtenu par application de la propriété de circularité :

$$i(p)^C_{87/85} = i(p)^C_{86/85} \cdot i(p)^C_{87/86}$$

soit pour la branche « Commerce »

$$i(p)^C_{87/85} = 1{,}052 \cdot 1{,}049 \cdot 100 = 110{,}35$$

– l'indice $i(p)^h_{84/85}$ est obtenu par application de la propreté de réversibilité :

$$i(p)^h_{84/85} = \frac{1}{i(p)^h_{85/84}}$$

soit dans le cas de la branche « Commerce »

$$i(p)^C_{84/85} = \frac{1}{1{,}052} \cdot 100 = 95{,}1$$

– l'indice $i(p)^h_{83/85}$ est obtenu par application des propriétés de circularité et de réversibilité. Il vaut :

$$i(p)^h_{83/85} = i(p)^h_{84/85} \cdot i(p)^h_{83/84} = \frac{1}{i(p)^h_{85/84}} \cdot \frac{1}{i(p)^h_{84/83}}$$

soit dans le cas de la branche « Commerce »

$$i(p)^C_{84/85} = \frac{1}{1{,}052} \cdot \frac{1}{1{,}087} \cdot 100 = 0{,}951 \cdot 0{,}92 \cdot 100 = 87{,}5$$

Indices base 100 1985

	$i(p)^h_{83/85}$	$i(p)^h_{84/85}$	$i(p)^h_{85/85}$	$i(p)^h_{86/85}$	$i(p)^h_{87/85}$
Agriculture	101,6	98,8	100,0	103,4	99,9
I.A.A.	88,3	96,5	100,0	99,6	99,5
Énergie	79,2	86,1	100,0	111,4	110,2
Biens intermédiaires	86,1	92,3	100,0	106,4	108,7
Biens d'équipement	84,7	92,4	100,0	102,5	104,4
Biens de consommation	86,7	91,6	100,0	112,5	118,2
B.G.C.A.	90,4	95,5	100,0	104,7	109,5
Commerce	87,4	95,1	100,0	105,2	110,4
Transports	87,8	94,0	100,0	103,9	93,6
Services marchands	85,8	94,0	100,0	105,5	110,1

72. Indice des importations

Mots-clefs

Indice de Laspeyres, indice de volume, indice de prix, circularité, indice de valeur

Énoncé

À partir des informations ci-dessous, nous voulons apprécier l'évolution du volume général des importations.

Importations par produits (valeurs en millions d'euros, indice 100 année précédente)

Produits	Année 0		Année 1		Année 2		Année 3	
	Valeurs	Indice de prix	Valeurs	Indice de prix	Valeurs	Indice de prix	Valeurs	Indice de prix
	V_0^h	$i(p)_{0/-1}^h$	V_1^h	$i(p)_{1/0}^h$	V_2^h	$i(p)_{2/1}^h$	V_3^h	$i(p)_{3/2}^h$
Agriculture	44 000	106	56 000	105	62 000	104	75 000	103
IAA	56 000	107	68 000	106	75 000	105	80 000	104
Énergie	25 000	106	30 000	104	37 000	103	38 000	102
Biens intermédiaires	170 000	105	190 000	103	210 000	102	240 000	103
Biens d'équipement	160 000	109	185 000	103	210 000	106	250 000	108
Biens de consommation	81 000	110	95 000	105	105 000	104	118 000	103
Transports	62 000	102	88 000	103	100 000	105	105 000	105
Services	50 000	108	65 000	103	80 000	100	88 000	98

1. Calculez l'indice des prix des importations pour l'année 3 en base 100 l'année 0 selon une formule de Laspeyres.

2. Calculez l'indice de volume correspondant et donnez un commentaire des résultats.

Corrigé

1. Un indice de Laspeyres est une moyenne arithmétique des indices élémentaires de prix pondéré par les coefficients de l'année de base. Ici nous cherchons un indice de prix, donc nous avons à calculer une moyenne d'indices élémentaires des prix.

$$L(p)_{1/0} = \sum_{h=1}^{n} k_0^h \cdot \frac{p_1^h}{p_0^h} = \sum_{h=1}^{n} k_0^h \cdot i(p)_{1/0}^h$$

Nous devons calculer les coefficients de pondération

$$k_0^h = \frac{V_0^h}{V_0}$$

où V_0^h est la valeur des importations en produit h pour l'année 0 et V_0 la valeur totale des importations pour l'année 0. Les pondérations représentent les fréquences relatives de chaque catégorie de produit dans le total des importations.

Structure des importations

Année 0	Valeurs des importations	Pondérations
	V_0^h	k_0^h
Agriculture	44 000	6,8
IAA	56 000	8,6
Énergie	25 000	3,9
Biens intermédiaires	170 000	26,2
Biens d'équipement	160 000	24,7
Biens de consommation	81 000	12,5
Transports	62 000	9,6
Services	50 000	7,7
Total	648 000	100,0

Nous disposons du tableau des indices de prix en base 100 l'année précédente

Indices des prix base 100 année précédente

	Année 0	Année 1	Année 2	Année 3
	$i(p)_{0/-1}^h$	$i(p)_{1/0}^h$	$i(p)_{2/1}^h$	$i(p)_{3/2}^h$
Agriculture	106	105	104	103
IAA	107	106	105	104
Énergie	106	104	103	102
Biens intermédiaires	105	103	102	103
Biens d'équipement	109	103	106	108
Biens de consommation	110	105	104	103
Transports	102	103	105	105
Services	108	103	100	98

Nous devons également calculer les indices de prix de l'année 3 pour chaque produit en base 100 l'année 0 par application de la propriété de circularité des indices simples.

La première colonne correspond aux indices $i(p)^h_{0/-1}$

$$i(p)^h_{3/0} = i(p)^h_{3/2} \cdot i(p)^k_{2/1} \cdot i(p)^k_{1/0}$$

Par exemple pour les biens intermédiaires, notés BI :

$$i(p)^{BI}_{1/0} = \frac{p^{BI}_1}{p^{BI}_0} \cdot 100 = 103 \ ; \ i(p)^{BI}_{2/1} = \frac{p^{BI}_2}{p^{BI}_1} \cdot 100 = 102 \ ;$$

$$i(p)^{BI}_{3/2} = \frac{p^{BI}_3}{p^{BI}_2} \cdot 100 = 103.$$

Le premier indice correspond à l'indice :

$$i(p)^{BI}_{0/-1} = \frac{p^{BI}_0}{p^{BI}_{-1}} \cdot 100 = 105$$

L'indice de l'année 3 en base 100 l'année 0 est donc :

$$i(p)^{BI}_{3/0} = i(p)^{BI}_{3/2} \cdot i(p)^{BI}_{2/1} \cdot i(p)^{BI}_{1/0} = 1{,}03 \cdot 1{,}02 \cdot 1{,}03 \cdot 100 = 108{,}2.$$

Nous appliquons la même méthode à chacun des produits et nous obtenons le tableau suivant :

Indices élémentaires des prix base 100 année 0

	Année 0	Année 3	Pondérations
	$i(p)^h_{0/0}$	$i(p)^h_{3/0}$	k^h_0
Agriculture	100,0	112,5	6,8
IAA	100,0	115,8	8,6
Énergie	100,0	109,3	3,9
Biens intermédiaires	100,0	108,2	26,2
Biens d'équipement	100,0	117,9	24,7
Biens de consommation	100,0	112,5	12,5
Transports	100,0	113,6	9,6
Services	100,0	100,9	7,7
			100,0

Nous obtenons donc facilement l'indice global des importations selon une formule de Laspeyres.

$$L(p)_{3/0} = \sum_{h=1}^{n} k^h_0 \cdot \frac{p^h_3}{p^h_0} = \sum_{h=1}^{n} k^h_0 \cdot i(p)^h_{3/0} = 112{,}1$$

2. Un indice de valeur est le produit d'un indice de prix et d'un indice de volume :

$$V_{1/0} = I(p)_{1/0} \cdot I(q)_{1/0}$$

Nous avons calculé l'indice de prix par une méthode de Laspeyres. Nous pouvons calculer l'indice de valeur pour l'ensemble des importations.

$$V_{3/0} = \frac{V_3}{V_0}$$

La valeur des importations pour l'année 0 est de 648000 et de 994000 pour l'année 3.

$$V_{3/0} = \frac{V_3}{V_0} = \frac{994\,000}{648\,000} \cdot 100 = 153{,}4$$

L'indice de volume est alors :

$$I(q)_{3/0} = \frac{V_{3/0}}{L(p)_{3/0}} = \frac{153{,}4}{112{,}1} \cdot 100 = 136{,}9\,.$$

Il s'agit d'un indice de Paasche.

L'augmentation de 53,4 % de la valeur des importations s'explique par une hausse des prix de 12,1 % et d'une augmentation des volumes de 36,9 %.

73. ÉVALUATION DE LA CONSOMMATION EN VOLUME

Mots-clefs

Indices synthétiques, indice de Paasche, indice de prix, indice de volume

Énoncé

Nous disposons d'informations sur le total de la consommation des ménages pour l'année 1 et la structure de cette consommation pour l'année 4. Nous disposons également des indices élémentaires de prix des produits pour quatre années. Nous voulons évaluer l'évolution de la consommation des ménages.

Consommation par type de produits (millions d'euros)

Produits	Année 1	Année 2	Année 3	Année 4
Produits alimentaires	820	850	890	920
Habillement	250	260	270	280
Logement	730	740	800	800
Meubles	300	310	320	300
Santé	320	380	420	550
Transports	620	650	660	690
Loisirs	280	300	320	330
Autres	480	510	520	530
Total	3800	4000	4200	4400

Indices élémentaires de prix (base 100 année précédente)

Produits	Année 1	Année 2	Année 3	Année 4
	$i(p)^h_{1/0}$	$i(p)^h_{2/1}$	$i(p)^h_{3/2}$	$i(p)^h_{4/3}$
Produits alimentaires	102	101	99	101
Habillement	101	102	102	101
Logement	102	101	102	104
Meubles	103	101	98	105
Santé	106	102	99	103
Transports	104	102	101	102
Loisirs	102	102	102	101
Autres	103	103	101	102

1. Calculez les indices élémentaires en base 100 l'année 1.
2. Calculez l'indice des prix pour l'ensemble des secteurs pour l'année 4 en base 100 l'année 1.
3. Calculez l'indice volume de la consommation des ménages base 100 l'année 1.

Corrigé

1. Nous disposons des indices élémentaires de prix en base 100 année précédente, le tableau ci-dessous explicite les données de base qui permettront de calculer les indices élémentaires en base 100 année 1.

Indices élémentaires de prix (base 100 année précédente)

Produits	Année 1	Année 2	Année 3	Année 4
	$i(p)^h_{1/0}$	$i(p)^h_{2/1}$	$i(p)^h_{3/2}$	$i(p)^h_{4/3}$
Produits alimentaires	102	101	99	101
Habillement	101	102	102	101
Logement	102	101	102	104
Meubles	103	101	98	105
Santé	106	102	99	103
Transports	104	102	101	102
Loisirs	102	102	102	101
Autres	103	103	101	102

Nous devons pour chaque type de produits calculer les indices en base 100 année 1.

– Pour l'année 1 $i(p)^h_{1/1} = 100$;

– Pour l'année 2 $i(p)^h_{2/1}$ est donné par le tableau ;

– Pour l'année 3 $i(p)^h_{3/1} = i(p)^h_{3/2} \cdot i(p)^h_{2/1}$;

– Pour l'année 4 $i(p)^h_{4/1} = i(p)^h_{4/3} \cdot i(p)^h_{3/1} = i(p)^h_{4/3} \cdot i(p)^h_{3/2} \cdot i(p)^h_{2/1}$.

Pour le produit logement, noté L, nous obtenons par exemple :

– $i(p)^L_{1/1} = 100$;

– $i(p)^L_{2/1} = 101$ est donné dans le tableau initial ;

– $i(p)^L_{3/1} = i(p)^L_{3/2} \cdot i(p)^L_{2/1} = 1{,}02 \cdot 1{,}01 \cdot 100 = 103{,}0$;

– $i(p)^L_{4/1} = i(p)^L_{4/3} \cdot i(p)^L_{3/1} = i(p)^L_{4/3} \cdot i(p)^L_{3/2} \cdot i(p)^L_{2/1}$

$$= 1{,}04 \cdot 1{,}03 \cdot 100 = 1{,}04 \cdot 1{,}02 \cdot 1{,}01 \cdot 100 = 107{,}1$$

Nous obtenons alors le tableau demandé.

Indice élémentaires de prix (base 100 année 1)

Produits	Année 1	Année 2	Année 3	Année 4
	$i(p)^h_{1/1}$	$i(p)^h_{2/1}$	$i(p)^h_{3/1}$	$i(p)^h_{4/1}$
Produits alimentaires	100,0	101,0	100,0	101,0
Habillement	100,0	102,0	104,0	105,1
Logement	100,0	101,0	103,0	107,1
Meubles	100,0	101,0	99,0	103,9
Santé	100,0	102,0	101,0	104,0
Transports	100,0	102,0	103,0	105,1
Loisirs	100,0	102,0	104,0	105,1
Autres	100,0	103,0	104,0	106,1

2. Nous pouvons calculer l'indice des prix uniquement selon une formule de Paasche ; en effet l'indice de Paasche des prix est une moyenne harmonique des indices élémentaires des prix pondérés par les coefficients budgétaires de l'année terminale. Soit :

$$\frac{1}{P(p)_{4/1}} = \sum_{h=1}^{n} \frac{k_4^h}{i(p)^h_{4/1}}$$

Nous disposons des indices élémentaires des prix, nous devons calculer les coefficients budgétaires de l'année 4.

	Année 4	
	Dépenses	Coefficients budgétaires
Produits	V_4^h	k_4^h
Produits alimentaires	920	20,9
Habillement	280	6,4
Logement	800	18,2
Meubles	300	6,8
Santé	550	12,5
Transports	690	15,7
Loisirs	330	7,5
Autres	530	12,0
Total	4 400	100,0

L'indice de Paasche est donc

$$P(p)_{4/1} = 104,5$$

3. Nous pouvons calculer l'indice du volume de la consommation en base 100 année 1 et, en disposant de ce résultat, calculer le volume de la consommation de l'année 4 en prix de l'année 1.

L'indice de valeur est le rapport entre la consommation l'année 4 et l'année 1

$$V_{4/1} = \frac{\text{Total des consommations année 4}}{\text{Total des consommations année 1}} = 100 \cdot \frac{4\,400}{3\,800} = 115{,}79$$

Nous pouvons calculer l'indice de volume de la consommation l'année 4 en base 100 l'année 1. C'est un indice de Laspeyres des volumes.

$$L(q)_{4/1} = \frac{V_{4/1}}{P(p)_{4/1}} = 100 \cdot \frac{115{,}8}{104{,}5} = 110{,}8\,.$$

La consommation des ménages a augmenté de 15,8 % dont 4,5 s'explique par la hausse des prix et 10,8 % par l'augmentation du volume de la consommation.

Le volume de la consommation pour l'année 4 en prix de l'année 1 est obtenu en divisant la valeur de la consommation de l'année 4 par l'indice de Paasche des prix.

$$\text{Volume de la consommation année 4} = \frac{4\,400}{1{,}045} \cong 4\,210{,}53\,.$$

74. Indice de la consommation

Mots-clefs

Indice chaîne, indice de prix, indice de volume, indice de Laspeyres

Énoncé

Nous disposons d'informations sur les consommations d'un groupe de ménages ainsi que sur les indices élémentaires de prix des produits pour quatre années.

Consommation par produits (en millions d'euros)

Produits	Année 1	Année 2	Année 3	Année 4
Produits alimentaires	820	850	890	920
Habillement	250	260	270	300
Logement	730	740	800	830
Meubles	300	310	320	300
Santé	320	380	420	570
Transports	620	650	660	690
Loisirs	280	300	320	350
Autres	480	510	520	540
Total	3800	4000	4200	4500

Indices élémentaires (base 100 année précédente)

Produits	Année 1	Année 2	Année 3	Année 4
Produits alimentaires	102	101	99	101
Habillement	101	102	102	101
Logement	102	101	102	104
Meubles	103	101	98	105
Santé	106	102	99	103
Transports	104	102	101	102
Loisirs	102	102	102	101
Autres	103	103	101	102

1. Calculez l'indice des prix pour l'ensemble des secteurs pour l'année 4 en base 100 l'année 1 par la méthode de l'indice chaîne.
2. Calculez l'indice de volume de la consommation des ménages base 100 l'année 1.

Corrigé

1. L'indice chaîne des prix est le produit des indices de Laspeyres pour chacune des années.

$$C(p)_{4/1} = L(p)_{4/3} \cdot L(p)_{3/2} \cdot L(p)_{2/1}$$

avec

$$L(p)_{i/i-1} = \sum_{h=1}^{n} k_{i-1}^{h} \cdot i(p)_{i/i-1}^{h}$$

Pour calculer les différents chaînons, nous avons besoin des indices en base 100 l'année précédente et des pondérations de cette même année.

Indices simples (base 100 année précédente)

Produits	$i(p)_{1/1}^h$	$i(p)_{2/1}^h$	$i(p)_{3/2}^h$	$i(p)_{4/3}^h$
Produits alimentaires	100,0	101	99	101
Habillement	100,0	102	102	101
Logement	100,0	101	102	104
Meubles	100,0	101	98	105
Santé	100,0	102	99	103
Transports	100,0	102	101	102
Loisirs	100,0	102	102	101
Autres	100,0	103	101	102

Les pondérations sont les fréquences relatives des postes de dépenses

$$k_1^h = \frac{p_1^h \cdot q_1^h}{\sum_{h=1}^{m} p_1^h \cdot q_1^h}$$

Tableau des pondérations

Produits	k_1^h	k_2^h	k_3^h	k_4^h
Produits alimentaires	21,6	21,3	21,2	20,9
Habillement	6,6	6,5	6,4	6,4
Logement	18,4	18,5	19,0	18,2
Meubles	7,9	7,8	7,6	6,8
Santé	9,2	9,5	10,0	12,5
Transports	16,3	16,3	15,7	15,7
Loisirs	7,4	7,5	7,6	7,5
Autres	12,6	12,8	12,4	12,0
Total	100,0	100,0	100,0	100,0

Les indices chaînons sont les suivants, multipliés par 100.

$$L(p)_{2/1} = \sum_{h=1}^{m} k_1^h \cdot i(p)_{2/1}^h = 101,65$$

$$L(p)_{3/2} = \sum_{h=1}^{m} k_2^h \cdot i(p)_{3/2}^h = 100,48$$

$$L(p)_{4/3} = \sum_{h=1}^{m} k_3^h \cdot i(p)_{4/3}^h = 102,36$$

L'indice chaîne est alors :

$$C(p)_{4/1} = L(p)_{4/3} \cdot L(p)_{3/2} \cdot L(p)_{2/1}$$

$$C(p)_{4/1} = 1,0165 \cdot 1,0048 \cdot 1,0236 \cdot 100 = 104,54$$

Les prix ont augmenté de 4,54 % sur les quatre années.

2. L'indice chaîne de volume est le rapport de l'indice de valeur sur l'indice chaîne des prix soit :

$$C(q)_{4/1} = \frac{V_{4/1}}{C(p)_{4/1}}$$

avec

$$V_{4/1} = \frac{4\,500}{3\,800} \cdot 100 = 115{,}79$$

d'où

$$C(q)_{4/1} = \frac{115{,}79}{104{,}54} \cdot 100 = 110{,}76$$

Le volume de la consommation a augmenté de 10,76 %.

75. INDICE DE VALEUR

Mots-clefs

Indice de Paasche, indice de Laspeyres

Énoncé

Un consommateur considère les prix et les quantités consommées de trois articles A, B et C qui constituent ses dépenses en 2005 et 2006 :

Données de base

	2005		2006	
	Prix	Quantité	Prix	Quantité
Article A	170	250	190	280
Article B	270	120	300	140
Article C	340	90	340	90

1. Calculez l'indice de Laspeyres des prix $L(p)$ pour l'année 2006 en base 100 l'année 2005.

2. Calculez l'indice de Paasche des prix $P(p)$ pour l'année 2006 en base 100 en 2005.

3. Calculez l'indice de valeur de la consommation $I(v)$.

4. Quelle est la nature exacte des indices obtenus en divisant $I(v)$ par $L(p)$ et en divisant $I(v)$ par $P(p)$?

Corrigé

1. La première idée qui vient à l'esprit pour résoudre cet exercice est d'appliquer la formule de définition de l'indice des prix de Laspeyres

$$L(p)_{1/0} = \sum_{h=1}^{n} \frac{q_0^h \cdot p_1^h}{q_0^h \cdot p_0^h}$$

ou celle de Paasche

$$P(p)_{1/0} = \sum_{h=1}^{n} \frac{q_1^h \cdot p_1^h}{q_1^h \cdot p_0^h}$$

Puisque nous disposons des quantités et des prix pour chacune des années, cela ne soulèverait aucune difficulté. Cependant la connaissance des quantités est fort rare et il vaut mieux, même dans ce cas, appliquer la formule de Laspeyres (resp. de Paasche) comme moyenne arithmétique des indices élémentaires de prix pondérés par les fréquences de l'année de base (resp. moyenne harmonique des indices élémentaires de prix pondérés par les fréquences de l'année terminale).

Pour mener à bien ce calcul nous devons déterminer les pondérations et les indices élémentaires.

Les pondérations dans l'indice de Laspeyres sont les fréquences relatives de chaque produit dans le total de la consommation, c'est l'importance relative de chaque poste budgétaire dans la dépense totale de consommation. L'indice de Laspeyres est une moyenne arithmétique des indices élémentaires avec les pondérations de l'année de base. Nous calculons donc les pondérations pour l'année 1983.

Calcul des pondérations pour 2005

	Prix	Quantité	Prix.Quantité	Pondérations
Article A	170	250	42 500	0,40
Article B	270	120	32 400	0,31
Article C	340	90	30 600	0,29
Total			105 500	1,00

$$L(p)_{06/05} = \sum_{h=1}^{n} k_{05}^h \cdot i(p)_{06/05}^h$$

Nous devons aussi disposer des indices élémentaires de prix (i) :

$$\text{Article A : } i(p)_{06/05}^A = 100 \cdot \frac{190}{170} = 111{,}8$$

$$\text{Article B : } i(p)_{06/05}^B = 100 \cdot \frac{300}{270} = 111{,}1$$

$$\text{Article C : } i(p)_{06/05}^C = 100 \cdot \frac{340}{340} = 100$$

d'où l'indice de Laspeyres des prix :

$$L(p)_{06/05} = 0{,}4 \cdot 111{,}8 + 0{,}31 \cdot 111{,}1 + 0{,}29 \cdot 100 = 108{,}2$$

2. L'indice de Paasche des prix est la moyenne harmonique des indices élémentaires de prix pondérés par les coefficients de l'année terminale. Nous calculons aisément les pondérations pour l'année 2006.

Calcul des pondérations pour 2006

	Prix	Quantité	Prix.Quantité	Pondérations
Article A	190	280	53 200	0,423
Article B	300	140	42 000	0,334
Article C	340	90	30 600	0,243
Total			125 800	1,000

$$\frac{1}{P(p)_{06/05}} = \frac{0,423}{111,8} + \frac{0,334}{111,1} + \frac{0,243}{100} = 0,0092198$$

$$P(p)_{06/05} = \frac{1}{0,0092198} = 108,46 \cong 108,5$$

3. L'indice de la consommation $Ic_{06/05}$ est obtenu en faisant le rapport entre la consommation totale de l'année 2006 sur la consommation totale de l'année 2005.

$$Ic_{06/05} = 100 \cdot \frac{125\ 800}{105\ 500} = 119,2$$

4. L'indice de valeur $Ic_{06/05}$ peut être considéré comme le produit d'un indice de prix et d'un indice de volume, ce qui permet de distinguer ce qui dans l'évolution de la valeur totale ressort du mouvement des prix de ce qui ressort de l'évolution des volumes.

La division de $Ic_{06/05}$ par $L(p)_{06/05}$ est un indice de volume $Iv^1_{06/05}$ et plus précisément un indice de Paasche de volume.

$$Iv^1_{06/05} = \frac{Ic_{06/05}}{L(p)_{06/05}} = P(q)_{06/05}$$

$$\text{donc } Iv^1_{06/05} = P(q)_{06/05} = 100 \cdot \frac{119,2}{108,2} = 110,2$$

Entre 2006 et 2005 pour le consommateur concerné, les volumes ont augmenté de 10,2 % alors que sa dépense augmentait de 19,2 %, la différence s'expliquant par la hausse des prix de 8,2 % pour les produits entrant dans sa consommation.

La division de $Ic_{06/05}$ par $P(p)_{06/05}$ est un indice de volume du type $Iv^2_{06/05}$ et plus précisément un indice de Laspeyres de volume.

$$Iv^2_{06/05} = \frac{Ic_{06/05}}{P(p)_{06/05}} = L(q)_{06/05}$$

$$\text{donc } Iv^2_{06/05} = L(q)_{06/05} = 100 \cdot \frac{119{,}2}{108{,}5} = 109{,}9$$

Entre 2005 et 2006 pour le consommateur concerné, les volumes ont augmenté de 9,9 % alors que sa dépense augmentait de 19,2 %, la différence s'expliquant par la hausse des prix de 8,5 % pour les produits entrant dans sa consommation.

Un indice de valeur peut donc s'analyser de deux manières différentes comme le produit d'un indice de prix par un indice de volume sans qu'il soit possible, sur un plan purement statistique, de choisir la meilleure décomposition.

Le lecteur s'étonnera, sans doute, de la complexification apportée aux calculs. Effectivement, dans cet exercice, la méthode allonge sensiblement les calculs. Cependant, dans la plupart des cas concrets, nous ne disposons que des pondérations, ou des moyens de calculer les pondérations ou les indices de prix. Le concept de quantité n'a de sens que pour un produit parfaitement homogène, les économistes raisonnent rarement à ce niveau de détail. Pour prendre un exemple dans l'automobile, les quantités n'ont de sens que pour un modèle dans une gamme avec des options bien spécifiées et une même disponibilité. La définition permet de comprendre la signification des indices, les calculs pratiques n'utilisent que les formules des moyennes.

76. Calcul des indices de prix à la consommation

Mots-clefs

Indice de Laspeyres, indice de Paasche, Indices chaînes de Laspeyres

Énoncé

Les indices élémentaires de prix pour différents produits sont fournis ci-dessous en base 100 l'année précédente. On demande de calculer pour chacun des produits les indices en base 100 l'année 1.

Indices de prix

Produit	Année 1	Année 2	Année 3	Année 4
A	102,4	104,3	101,3	102,4
B	105,6	108,5	102,4	101,9
C	101,5	105,6	103,5	104,2
D	102,8	104,9	102,8	105,8
E	106,8	107,4	105,9	108,4
F	100,1	102,8	101,4	102,1

La structure de la consommation de ces produits est connue à l'aide du tableau suivant :

Structure de la consommation

Produit	Année 1	Année 2	Année 3	Année 4
A	25,4	24,2	23,1	22,5
B	16,3	16,4	16,5	16,6
C	12,8	13,5	14,2	15,0
D	18,8	20,0	22,4	24,5
E	14,3	15,6	16,7	17,5
F	12,4	10,3	7,1	3,9

1. Calculez l'indice synthétique des prix en utilisant une formule de Laspeyres pour l'année 4 sur base 100 l'année 1.

2. Calculez l'indice synthétique des prix en utilisant une formule de Paasche pour l'année 4 sur base 100 l'année 1.

3. Calculez l'indice synthétique des prix sur une base 100 l'année 1 en employant la méthode des indices chaînes pour l'année 4 sur base 100 l'année 1.

Corrigé

1. Les pondérations sont connues pour chacune des années, il reste à calculer les indices élémentaires de prix pour l'année en base 100 l'année 1.

Indices de prix en écriture formelle

Produit	Année 1	Année 2	Année 3	Année 4
	$i(p)^h_{1/-1}$	$i(p)^h_{2/1}$	$i(p)^h_{3/2}$	$i(p)^h_{4/3}$
A	102,4	104,3	101,3	102,4
B	105,6	108,5	102,4	101,9
C	101,5	105,6	103,5	104,2
D	102,8	104,9	102,8	105,8
E	106,8	107,4	105,9	108,4
F	100,1	102,8	101,4	102,1

Indices élémentaires année 4
(base 100 année 1)

A	B	C	D	E	F
$i(p)^A_{4/1}$	$i(p)^B_{4/1}$	$i(p)^C_{4/1}$	$i(p)^D_{4/1}$	$i(p)^E_{4/1}$	$i(p)^F_{4/1}$
108,1	113,2	104,5	114,1	123,3	106,4

$$i(p)^A_{4/1} = 1{,}043 \cdot 1{,}013 \cdot 1{,}024 \cdot 100 = 108{,}1\,.$$

Les indices des autres produits se calculent de façon analogue.

L'indice de Laspeyres est une moyenne arithmétique des indices élémentaires des prix.

$$L(p)_{4/1} = \sum_{h=1}^{n} k_1^h \cdot \frac{p_4^h}{p_1^h} = \sum_{h=1}^{n} k_1^h \cdot i(p)_{4/1}^h$$

Pondérations pour l'année 1

Produit	k_1^h
A	25,4
B	16,3
C	12,8
D	18,8
E	14,3
F	12,4

$$L(p)_{4/1} = \frac{1}{100}(108{,}1 \cdot 25{,}4 + 113{,}2 \cdot 16{,}3 + 104{,}5 \cdot 12{,}8$$
$$+114{,}1 \cdot 18{,}8 + 123{,}3 \cdot 14{,}3 + 106{,}4 \cdot 12{,}4)$$
$$L(p)_{4/1} = 111{,}6.$$

2. L'indice de Paasche est une moyenne harmonique des indices élémentaires des prix.

$$\frac{1}{P(p)_{4/1}} = \sum_{h=1}^{n} \frac{k_4^h}{i(p)_{4/1}^h}$$

Pondérations pour l'année 4

Produit	k_4^h	$i(p)_{4/1}^h$
A	22,5	108,1
B	16,6	113,2
C	15,0	104,5
D	24,5	114,1
E	17,5	106,4
F	3,9	102,8

$$\frac{1}{P(p)_{4/1}} = \frac{1}{100}\left(\frac{22{,}5}{108{,}1} + \frac{16{,}6}{113{,}2} + \frac{15{,}0}{104{,}5} + \frac{24{,}5}{114{,}1} + \frac{17{,}5}{123{,}3} + \frac{3{,}9}{106{,}4}\right)$$

$$P(p)_{4/1} \cong 112{,}2$$

3. Le calcul de l'indice chaîne nécessite le calcul des indices de Laspeyres pour chaque année en base 100 l'année précédente.

$$L(p)_{2/1} = \sum_{h=1}^{n} k_1^h \cdot i(p)_{2/1}^h$$

Produit	k_1^h	$i^h(p)_{2/1}$	$k_1^h \cdot i(p)_{2/1}^h$
A	25,4	104,3	2649,22
B	16,3	108,5	1768,55
C	12,8	105,6	1351,68
D	18,8	104,9	1972,12
E	14,3	107,4	1535,82
F	12,4	102,8	1274,72
Total	100,0		10552,11

$$L(p)_{2/1} = \frac{10552,11}{100} \cong 105,5$$

$$L(p)_{3/2} = \sum_{h=1}^{n} k_2^h \cdot i(p)_{3/2}^h$$

Produit	k_2^h	$i(p)_{3/2}^h$	$k_2^h \cdot i(p)_{3/2}^h$
A	24,2	101,3	2451,46
B	16,4	102,4	1679,36
C	13,5	103,5	1397,25
D	20,0	102,8	2056
E	15,6	105,9	1652,04
F	10,3	101,4	1044,42
Total	100,0		10280,53

$$L(p)_{3/2} = \frac{10280,53}{100} \cong 102,8$$

$$L(p)_{4/3} = \sum_{h=1}^{n} k_3^h \cdot i(p)_{4/3}^h$$

Produit	k_3^h	$i(p)_{4/3}^h$	$k_3^h \cdot i(p)_{4/3}^h$
A	23,1	102,4	2365,44
B	16,5	101,9	1681,35
C	14,2	104,2	1479,64
D	22,4	105,8	2369,92
E	16,7	108,4	1810,28
F	7,1	102,1	724,91
Total	100,0		10431,54

$$L(p)_{4/3} = \frac{10431,54}{100} \cong 104,3$$

Nous pouvons calculer l'indice chaîne :

$$C(p)_{4/1} = L(p)_{4/3} \cdot L(p)_{3/2} \cdot L(p)_{2/1}$$

$$C(p)_{4/1} = 1,043 \cdot 1,028 \cdot 1,055 \cdot 100 \cong 113,1$$

Nous obtenons trois solutions pour mesurer l'indice de prix :

$$L(p)_{4/1} = 111,6$$

$$P(p)_{4/1} \cong 112,2$$

$$C(p)_{4/1} \cong 113,1$$

Les écarts s'expliquent par les modifications importantes des coefficients.

77. ÉVOLUTION DU PRIX DE LA VAB

Mots-clefs

Indice de Laspeyres, indice de Paasche, indice chaîne, indice de valeur

Énoncé

Nous disposons de la valeur ajoutée brute par grandes branche ainsi que des hausses des prix de la VAB.

Tableau de la VAB par branches (en milliards d'euros prix courants)

Année	3	4	5	6	7
Agriculture	170	175	180	188	185
I.A.A.	124	135	146	148	147
Énergie	173	194	231	246	246
Biens intermédiaires	242	256	274	297	310
Biens d'équipement	287	308	321	334	349
Biens de consommation	196	203	221	246	254
B.G.C.A.	235	242	250	265	287
Commerce	422	466	497	546	568
Transports	236	260	284	305	316
Services marchands	1 044	1 163	1 267	1 398	1 520

Hausse des prix de la valeur ajoutée (en % de l'année précédente)

Année	3	4	5	6	7
Agriculture	6,9	– 2,7	1,2	3,4	– 3,4
I.A.A.	8,6	9,3	3,6	– 0,4	– 0,6
Énergie	11,1	8,6	16,2	11,4	– 1,1
Biens intermédiaires	8,5	7,3	8,3	6,4	2,2
Biens d'équipement	9,6	9,1	8,2	2,5	1,9
Biens de consommation	8,3	5,6	9,2	12,5	5,1
B.G.C.A.	5,5	5,6	4,7	4,7	4,6
Commerce	11,4	8,7	5,2	5,2	4,9
Transports	8,0	7,0	6,4	3,9	– 9,9
Services marchands	12,3	9,5	6,4	5,5	4,4

1. Calculez les indices de prix de la VAB en base 100 l'année précédente pour les années 4, 5, 6, 7, en utilisant une formule de Laspeyres.
2. Calculez les indices de prix de la VAB en base 100 en année 3 pour l'année 7 selon une formule de Laspeyres et selon une formule de Paasche.
3. Calculez, selon la méthode des indices chaînes, les indices de prix pour les années 4, 5, 6, 7 en base 100 année 3.
4. Calculez les indices de volume de la VAB pour chacune des années.

Corrigé

1. Le calcul de l'indice des prix de Laspeyres nécessite de disposer des pondérations de l'année de base. Ici nous devons disposer des pondérations pour les années 3, 4, 5, 6, c'est-à-dire de l'importance relative de la VAB de chaque branche dans la VAB de l'ensemble des branches.

Le tableau suivant synthétise l'ensemble des pondérations.

Pondérations

	k_3^h	k_4^h	k_5^h	k_6^h
Agriculture	5,4	5,1	4,9	4,7
I.A.A.	4,0	4,0	4,0	3,7
Énergie	5,5	5,7	6,3	6,2
Biens intermédiaires	7,7	7,5	7,5	7,5
Biens d'équipement	9,2	9,1	8,8	8,4
Biens de consommation	6,3	6,0	6,0	6,2
B.G.C.A.	7,5	7,1	6,8	6,7
Commerce	13,5	13,7	13,5	13,7
Transports	7,5	7,6	7,7	7,7
Services marchands	33,4	34,2	34,5	35,2
	100,0	100,0	100,0	100,0

Les indices élémentaires de prix sont obtenus en transformant les taux de croissance en multiplicateur. Par exemple :

Indices élémentaires des prix en base 100 l'année précédente

	$i(p)_{3/2}^h$	$i(p)_{4/3}^h$	$i(p)_{5/4}^h$	$i(p)_{6/5}^h$	$i(p)_{7/6}^h$
Agriculture	106,9	97,3	101,2	103,4	96,6
I.A.A.	108,6	109,3	103,6	99,6	99,4
Énergie	111,1	108,6	116,2	111,4	98,9
Biens intermédiaires	108,5	107,3	108,3	106,4	102,2
Biens d'équipement	109,6	109,1	108,2	102,5	101,9
Biens de consommation	108,3	105,6	109,2	112,5	105,1
B.G.C.A.	105,5	105,6	104,7	104,7	104,6
Commerce	111,4	108,7	105,2	105,2	104,9
Transports	108,0	107,0	106,4	103,9	90,1
Services marchands	112,3	109,5	106,4	105,5	104,4

L'indice de la VAB selon une formule de Laspeyres est la moyenne arithmétique des indices élémentaires. Soit $L(p)_{1/0}$ cet indice pour l'année 1 en base 100 l'année 0 :

$$L(p)_{1/0} = \sum_{h=1}^{n} k_0^h \cdot i(p)_{1/0}^h$$

Nous obtenons les indices demandés :

$$L(p)_{4/3} = \sum_{h=1}^{n} k_3^h \cdot i(p)_{4/3}^h \cong 107{,}74$$

$$L(p)_{5/4} = \sum_{h=1}^{n} k_4^h \cdot i(p)_{5/4}^h \cong 106{,}77$$

$$L(p)_{6/5} = \sum_{h=1}^{n} k_5^h \cdot i(p)_{6/5}^h \cong 105{,}55$$

$$L(p)_{7/6} = \sum_{h=1}^{n} k_6^h \cdot i(p)_{7/6}^h \cong 102{,}16$$

Par exemple :

$$\begin{aligned} L(p)_{4/3} = {} & [(170 \cdot 0{,}973) + (124 \cdot 1{,}093) + (173 \cdot 1{,}086) - (242 \cdot 1{,}073) \\ & +(287 \cdot 1{,}091) + (196 \cdot 1{,}056) + (235 \cdot 1{,}056) + (422 \cdot 1{,}087) \\ & +(236 \cdot 1{,}07) + (1044 \cdot 1{,}095)] = 107{,}74 \end{aligned}$$

2. Pour obtenir l'indice des prix de la VAB en base 100 année 3 pour l'année 7, il est nécessaire de calculer les indices élémentaires pour l'année 7 en base 100 année 3.

Indices élémentaires des prix (en base 100 année 3)

	$i(p)^h_{3/3}$	$i(p)^h_{4/3}$	$i(p)^h_{5/3}$	$i(p)^h_{6/3}$	$i(p)^h_{7/3}$
Agriculture	100	97,3	98,5	101,8	98,4
I.A.A.	100	109,3	113,2	112,8	112,1
Énergie	100	108,6	126,2	140,6	139,0
Biens intermédiaires	100	107,3	116,2	123,6	126,4
Biens d'équipement	100	109,1	128,8	132,0	128,8
Biens de consommation	100	105,6	115,3	129,7	136,3
B.G.C.A.	100	105,6	110,6	115,8	121,1
Commerce	100	108,7	114,4	120,3	126,2
Transports	100	107,0	113,8	118,3	106,6
Services marchands	100	109,5	116,5	122,9	128,3

$$i(p)^h_{7/3} = i(p)^h_{7/6} \cdot i(p)^h_{6/5} \cdot i(p)^h_{5/4} \cdot i(p)^h_{4/3}$$

Par exemple pour la branche « Agriculture » :

$$i(p)^{agri}_{7/3} = 100 \cdot 0{,}973 \cdot 1{,}012 \cdot 1{,}034 \cdot 0{,}966 \cong 0{,}984$$

$$L(p)_{7/3} = \sum_{h=1}^{n} k^h_3 \cdot i(p)^h_{7/3} = 124{,}1$$

L'indice de Laspeyres des prix pour l'année 7 en base 100 en année 3 est de 124,1.

$$\frac{1}{P(p)_{3/7}} = \sum_{h=1}^{n} k^h_7 \cdot \frac{1}{i(p)^h_{3/7}}$$

L'indice de Paasche pour les mêmes références est de 123,8.

3. Les indices chaînes sont les produits d'indices de Laspeyres selon la formule générale :

$$C(p)_{3/0} = L(p)_{3/2} \cdot L(p)_{2/1} \cdot L(p)_{1/0}$$

Puisque nous connaissons les indices de Laspeyres d'une année sur l'autre, il est facile d'obtenir les indices chaînes correspondants.

$$C(p)_{4/3} = L(p)_{4/3} = 107{,}74$$

$$C(p)_{5/3} = L(p)_{5/4} \cdot L(p)_{4/3} = 115{,}03$$

$$C(p)_{6/3} = L(p)_{6/5} \cdot L(p)_{5/4} \cdot L(p)_{4/3} = 121{,}41$$

$$C(p)_{7/3} = L(p)_{7/6} \cdot L(p)_{6/5} \cdot L(p)_{5/4} \cdot L(p)_{4/3} = 124{,}04$$

4. Les indices de volume correspondants sont alors obtenus par le rapport entre l'indice de valeur global et l'indice des prix.

Tableau des indices de valeur

Année	3	4	5	6	7
Valeurs	3129	3402	3671	3973	4182
Indices de valeur (base 100 année 3)	100,00	108,72	117,32	126,97	133,65
Indices de valeur (base 100 année précédente)	100,0	108,72	107,91	108,23	105,26

Un indice de volume est obtenu en divisant un indice de valeur par un indice de prix selon la formule générale :

$$I(vol)_{1/0} = \frac{I(val)_{1/0}}{I(p)_{1/0}}$$

De l'analyse d'un indice de valeur nous savons que :

$$I(val)_{1/0} = L(p)_{1/0} \cdot P(q)_{1/0} = P(p)_{1/0} \cdot L(q)_{1/0}$$

Nous obtenons ainsi les indices de Paasche de volume pour chacune des années en utilisant les indices de Laspeyres des prix.

$$P(q)_{4/3} = \frac{V_{4/3}}{L(p)_{4/3}} = 100 \cdot \frac{108{,}72}{107{,}74} = 100{,}9$$

$$P(q)_{5/4} = \frac{V_{5/4}}{L(p)_{5/4}} = 100 \cdot \frac{107{,}91}{106{,}77} = 101{,}07$$

$$P(q)_{6/5} = \frac{V_{6/5}}{L(p)_{6/5}} = 100 \cdot \frac{108{,}23}{105{,}55} = 102{,}54$$

$$P(q)_{7/6} = \frac{V_{7/6}}{L(p)_{7/6}} = 100 \cdot \frac{105{,}26}{102{,}16} = 103{,}03$$

En utilisant les indices chaînes de prix nous obtenons une nouvelle série d'indices de volume qui sont des indices chaînes.

$$C(q)_{4/3} = \frac{V_{4/3}}{C(p)_{4/3}} = 100 \cdot \frac{108{,}72}{107{,}74} = 100{,}91$$

$$C(q)_{5/3} = \frac{V_{5/3}}{C(p)_{5/3}} = 100 \cdot \frac{117{,}3}{115{,}03} = 101{,}97$$

$$C(q)_{6/3} = \frac{V_{6/3}}{C(p)_{6/3}} = 100 \cdot \frac{126{,}97}{121{,}41} = 104{,}58$$

$$C(q)_{7/3} = \frac{V_{7/3}}{C(p)_{7/3}} = 100 \cdot \frac{133{,}65}{124{,}04} = 104{,}75$$

Pour l'année 7 en base 100 l'année 3 :

$$P(q)_{7/3} = \frac{V_{7/3}}{L(p)_{7/3}} = 100 \cdot \frac{133{,}65}{124{,}1} = 107{,}70$$

$$L(q)_{7/3} = \frac{V_{7/3}}{P(p)_{7/3}} = 100 \cdot \frac{133{,}65}{123{,}8} = 107{,}96$$

78. Comparaison d'un indice de Laspeyres et d'un indice chaîne

Mots-clefs

Indice de Laspeyres, indice chaîne, indie de Paasche

Énoncé

Nous disposons des indices de production pour différentes branches d'activité. Nous désirons calculer un indice de production pour l'ensemble activités.

Indices de production (base 100 année -3)

	1	2	3	4
Énergie	124	131	130	129
Biens intermédiaires	119	125	124	119
Biens de consommation	143	149	149	144
Biens d'équipement	144	140	141	140

Nous considérerons les indices ci-dessus comme des indices simples, relatifs aux niveaux de production annuels.

Par ailleurs nous disposons des parts de la valeur de la production de chaque branche dans la valeur de la production totale des branches considérées pour les mêmes années.

Répartition de la production selon les branches (en pourcentage)

	1	2	3	4
Énergie	22	21	20	19
Biens intermédiaires	36	37	38	38
Biens de consommation	25	26	27	28
Biens d'équipement	17	16	15	15

1. Calculez, selon les formules de Laspeyres et Paasche, l'indice synthétique de production de l'année 4 en base 100 année 1.
2. Calculez l'indice chaîne pour l'année 4 en base 100 l'année 1. Comparer le résultat à celui obtenu précédemment avec la formule de Laspeyres, expliquer la différence.

Corrigé

1. Nous devons tout d'abord calculer les indices de production en base 100 année 1. Puisqu'il s'agit d'indices simples qui possèdent la propriété de circularité des indices, il suffit de faire le rapport entre les indices. Nous disposons des indices de forme générale $i(q)^h_{i/-3}$. Nous devons calculer les indices en base 100 l'année 1.

$$i(q)^h_{2/1} = \frac{i(q)^h_{2/-3}}{i(q)^h_{1/-3}} \; ; \; i(q)^h_{3/1} = \frac{i(q)^h_{3/-3}}{i(q)^h_{1/-3}} \; ; \; i(q)^h_{4/1} = \frac{i(q)^h_{4/-3}}{i(q)^h_{1/-3}}$$

D'où le tableau des indices de production en base 100 année 1.

Indices simples de production

	1	2	3	4
	$i(q)^h_{1/1}$	$i(q)^h_{2/1}$	$i(q)^h_{3/1}$	$i(q)^h_{4/1}$
Énergie	100	105,6	104,8	104,0
Biens intermédiaires	100	105,0	104,2	100,0
Biens de consommation	100	104,2	104,2	100,7
Biens d'équipement	100	97,2	97,9	97,2

L'indice synthétique de Laspeyres est une moyenne arithmétique des indices élémentaires de production pondérés par la structure de la production pour l'année 1.

$$L(q)_{t/1} = \sum_{h=1}^{n} k_1^h \cdot i(q)^h_{t/1}$$

L'indice synthétique de Paasche des prix est une moyenne harmonique des indices élémentaires de production pondérés par la structure de la production de l'année 4.

Nous pouvons présenter les calculs sous forme de tableau :

Calculs pour les indices

Indice élémentaire de production	Pondération année 1	Pondération année 4		
$i(q)_{4/1}^h$	k_1^h	k_4^h	$k_1^h \cdot i(q)_{4/1}^h$	$\frac{k_4^h}{i(q)_{4/1}^h}$
104,0	22	19	2288,0	0,18269
100,0	36	38	3600,0	0,38000
100,7	25	28	2517,5	0,27805
97,2	17	15	1652,4	0,15431
	100	100	10057,9	0,99505

L'indice de Laspeyres de la production est de 100,579 soit 100,58 ; l'indice de Paasche de la production est de 100,497 peu différent de 100,50.

2. L'indice chaîne est le produit d'indices de Laspeyres pour chaque année.

Calcul des indices de Laspeyres

	$i(q)_{2/1}^h$	k_1^h	$k_1^h \cdot i(q)_{2/1}^h$	$i(q)_{3/2}^h$	k_2^h	$k_2^h \cdot i(q)_{3/2}^h$	$i(q)_{4/3}^h$	k_3^h	$k_3^h \cdot i(q)_{4/3}^h$
E	105,6	22	2 323,2	99,2	21	2 083,2	99,2	20	1 984,0
B I	105,0	36	3 780,0	99,2	37	3 670,4	96,0	38	3 648,0
B C	104,2	25	2 605,0	100,0	26	2 600,0	96,6	27	2 608,2
B E	97,2	17	1 652,4	100,7	16	1 611,2	99,3	15	1 489,5
			10 360,6			9 964,8			9 729,7

Les indices de Laspeyres pour la production sont de 103,6 année 2 en base 100 année 1, 99,65 année 3 base 100 année 2, 97,3 année 4 base 100 année 3.

L'indice chaîne pour l'année 4 en base 100 année 1 est le produit des indices de chacune des années soit 100,45.

Pour une même évolution, nous obtenons trois estimations de la croissance de la production entre 1 et 4. Selon la formule de Laspeyres, la croissance serait de 0,58 % ; selon la formule de Paasche elle serait de 0,50 % ; et selon la formule de l'indice chaîne de 0,45 %. L'indice de Fischer de la production moyenne géométrique des indices chaînes de Paasche et de Laspeyres est de 100,54, soit une croissance de 0,54 %. Ces différences ne s'expliquent que par l'utilisation de méthodes de calculs différentes. Le choix de la méthode la plus appropriée dépend des objectifs poursuivis.

79. UTILISATION DES INDICES POUR ÉVALUER L'ÉVOLUTION DU POUVOIR D'ACHAT

Mots-clefs

Raccords d'indices

Énoncé

La structure des budgets des ménages diffère selon le niveau de revenus comme l'indique le tableau ci-dessous pour l'année 1.

Structures des consommations

Rubriques	Structure du budget hauts revenus	Structure du budget bas revenus
Habitation	30,1	32,6
Alimentation	17,9	27,5
Transports	17,9	12,6
Habillement	9,7	7,6
Culture Loisirs	8,8	7,5
Santé	1,6	6,8
Vacances	5,4	1,5
Divers	8,6	3,9

Les indices de prix pour quelques années sont donnés par le tableau suivant.

Indices de prix (base 100 année précédente)

Rubriques	Année 0	Année 1	Année 2
	$i(p)^h_{0/-1}$	$i(p)^h_{1/0}$	$i(p)^h_{2/1}$
Habitation	106,6	105,5	104,4
Alimentation	101,4	102,4	103,2
Transports	101,8	100,3	101,4
Habillement	105,0	103,2	103,0
Culture Loisirs	104,1	101,5	101,2
Santé	102,5	103,6	104,3
Vacances	103,9	104,4	102,3
Divers	101,9	102,1	101,0

Le problème est de déterminer si l'évolution du coût de la vie, mesuré par un indice synthétique, est identique pour les deux catégories de ménages ?

1. Calculez l'indice du coût de la vie pour l'année 1 en base 100 l'année 0. Vous expliquerez le choix de la formule d'indice des prix utilisée.

2. Calculez l'indice du coût de la vie pour l'année 2 en base 100 l'année 1. Ici encore le choix de la formule d'indice utilisé devra être explicité.

3. Sachant, que pour l'année 1, le budget par unité de consommation est de 88 300 €/an et par personne pour les hauts revenus, et de 41 000 €/an et par personne pour les bas revenus, quelles devraient être les dépenses pour l'année 2 :

   a. pour que les deux catégories de ménages maintiennent leur pouvoir d'achat ?

   b. pour qu'il y ait une croissance de 2 % du pouvoir d'achat pour les deux catégories de ménages ?

Corrigé

1. Nous disposons de la structure des dépenses pour l'année 1. Pour estimer la croissance des prix en base 100 l'année 0, nous sommes contraints d'utiliser un indice de Paasche des prix. Les pondérations sont celles de l'année courante.

Tableau de calcul des indices de Paasche

Rubriques	Indices élémentaires	Budget hauts revenus	Budget bas revenus		
	$i(p)^h_{1/0}$	k_0^{hr}	k_0^{br}	$\frac{k_0^{hr}}{i(p)^h_{1/0}}$	$\frac{k_0^{br}}{i(p)^h_{1/0}}$
Habitation	105,5	30,1	32,6	0,28530806	0,30900474
Alimentation	102,4	17,9	27,5	0,17480469	0,26855469
Transports	100,3	17,9	12,6	0,17846461	0,12562313
Habillement	103,2	9,7	7,6	0,09399225	0,07364341
Culture Loisirs	101,5	8,8	7,5	0,08669951	0,07389163
Santé	103,6	1,6	6,8	0,01544402	0,06563707
Vacances	104,4	5,4	1,5	0,05172414	0,01436782
Divers	102,1	8,6	3,9	0,08423115	0,03819785
		100,0	100,0	0,00970668	0,0096892

L'indice de Paasche pour les hauts revenus est :

$$\frac{1}{P(p)_{1/0}} = \sum_{h=1}^{n} \frac{k_0^{hr}}{i(p)^h_{1/0}}$$

L'indice des prix à la consommation pour les hauts revenus est de 103,02 et de 103,21 pour les bas revenus.

2. Pour estimer l'évolution des prix entre l'année 1 et l'année 2 nous disposons des pondérations pour l'année 1, nous calculerons un indice des prix selon une formule de Laspeyres.

Tableau pour le calcul du coût de la vie

Rubriques	Indice élémentaire	Budget hauts revenus	Budget bas revenus		
	$i(p)^h_{2/1}$	k^{hr}_0	k^{br}_0	$k^{hr}_0 \cdot i(p)^h_{2/1}$	$k^{br}_0 \cdot i(p)^h_{2/1}$
Habitation	104,4	30,1	32,6	3142,44	3403,44
Alimentation	103,2	17,9	27,5	1847,28	2838
Transports	101,4	17,9	12,6	1815,06	1277,64
Habillement	103,0	9,7	7,6	999,1	782,8
Culture Loisirs	101,2	8,8	7,5	890,56	759
Santé	104,3	1,6	6,8	166,88	709,24
Vacances	102,3	5,4	1,5	552,42	153,45
Divers	101,0	8,6	3,9	868,6	393,9
		100,0	100,0	10282,34	10317,47

Le tableau donne immédiatement l'indice de Laspeyres pour les hauts revenus : 102,8 ainsi que pour les bas revenus : 103,2. L'écart est de 0,4 %.

3. a. Le revenu par personne pour les hauts revenus était de 88 300 € l'année 1. Pour qu'il y ait maintien du pouvoir d'achat, le revenu doit être augmenté de 2,8 % soit une multiplication par 1,028. Le revenu pour l'année 2 devrait être de 90 772 €.

Pour les bas revenus, le multiplicateur est de 103,2 donc le revenu maintenant le pouvoir d'achat devrait être de 42 312 € par personne.

b. Si le pouvoir d'achat doit augmenter de 2 % pour chaque catégorie les multiplicateurs se modifient :

– pour les hauts revenus, il devient 1,028.1,02 = 1,049, donc le revenu nominal pour une croissance de 2 % est de 92 627 € par an et par personne

– pour les bas revenus, le multiplicateur devient 1,032 .1,02 = 1,053, donc le revenu nominal pour une croissance de 2 % est de 43 173 € par an et par personne.

80. ÉVOLUTION DU POUVOIR D'ACHAT AVEC PLUSIEURS INDICES

Mots-clefs

Indice de Laspeyres

Énoncé

Un ménage veut apprécier l'évolution de son pouvoir d'achat. Il estime la ventilation de ses consommations hors logement selon la répartition suivante pour l'année 1984.

Structure des consommations

Rubriques	Pondérations (en %)
Produits à base de céréales	12
Viandes	27
Lait et produits laitiers	16
Fruits et légumes	10
Boissons	5
Loisirs-culture	7
Transports	15
Divers	8

Les informations disponibles permettent de connaître les hausses des prix des différentes catégories de produits depuis 1984, sur la période 1984-1989, ainsi que pour l'année 1989.

Hausses de prix

Rubriques	Depuis 1984 (en %)	En 1989 en (%)
Produits à base de céréales	14,5	4,1
Viandes	18,4	3,5
Lait et produits laitiers	12,3	3,0
Fruits et légumes	21,5	4,2
Boissons	11,2	2,5
Loisirs-culture	15,0	2,9
Transports	17,4	3,4
Divers	16,5	2,1

1. Quel est l'indice, en 1988, des prix à la consommation pour l'ensemble des produits consommés sur une base 100 en 1984 ?

2. Quelle est la hausse globale des prix ? Quelle est la hausse annuelle moyenne des prix ?

3. Quel sera l'indice des prix des produits en fin 1989, sur une base 100 en 1988 ? Quels sont les problèmes qui se posent pour calculer cet indice ?

4. Si le revenu annuel du ménage était de 120 000 F en fin 1988, quel devrait être son revenu fin 1989 pour qu'il y ait :

   a. maintien du pouvoir d'achat de revenu ?

   b. gain de 1,2 % de pouvoir d'achat du revenu ?

5. Y aurait-il maintien du pouvoir d'achat si le revenu mensuel moyen augmentait régulièrement au cours de l'année 1989 jusqu'à 10 340 F pour le mois de décembre 1989 ?

Corrigé

1. Nous disposons des pondérations de l'année de base, l'indice des prix que nous calculerons sera un indice de Laspeyres.

Nous devons déterminer les indices élémentaires de prix. Puisque nous connaissons les hausses de prix pour 1984-1989 et les hausses de prix pour 1989, nous pouvons calculer les indices élémentaires de prix pour 1988 en base 100 en 1984 en divisant le multiplicateur des prix entre 1984 et 1989 par le multiplicateur de 1989 et cê pour chaque rubrique.

Indices des prix base 100 1984

Rubriques	Indice 1988 $i(p)^h_{88/84}$
Produits à base de céréales	110,0
Viandes	114,4
Lait et produits laitiers	109,0
Fruits et légumes	116,6
Boissons	108,5
Loisirs-culture	111,8
Transports	113,5
Divers	114,1

L'indice global des prix de la consommation du ménage en 1988 en base 100 en 1984 est :

$$L(p)_{88/84} = \sum_{h=1}^{n} k^h_{84} \cdot i(p)^h_{88/84}$$

$$L(p) = 110{,}0 \cdot 0{,}12 + 114{,}4 \cdot 0{,}27 + 109 \cdot 0{,}16 + 116{,}6 \cdot 0{,}10 + 108{,}5 \cdot 0{,}05 \\ +111{,}8 \cdot 0{,}07 + 113{,}5 \cdot 0{,}15 + 114{,}1 \cdot 0{,}0{,}8$$

$$L(p) \cong 112{,}59$$

2. La hausse globale entre 1984 et 1989 est donc de 12,6 %. Le multiplicateur moyen annuel des prix est la moyenne géométrique du multiplicateur (1,1259) des prix entre 1984 et 1988 soit 1,030. La hausse moyenne des prix est d'environ 3,0 %.

3. Nous ne pouvons calculer directement la hausse des prix en 1989 sur une base 100 en 1988 puisque nous ne disposons pas des pondérations pour l'année 1988. Nous serons donc conduits à calculer l'indice global des prix pour 1989 en base 100 en 1984. Puis faute d'informations suffisantes nous ferons le rapport de l'indice de Laspeyres des prix de 1989 sur l'indice de Laspeyres des prix de 1988. L'indice trouvé ne sera pas un indice de Laspeyres et nous

savons que les pondérations implicites utilisées ne correspondent ni aux pondérations de 1984 ni aux pondérations de 1988.

Puisque l'indice des prix 1989, en base 100 en 1984, est de 116,4, l'indice 1989, en base 100 en 1988, est :

$$I(p)_{89/88} = 103{,}37 \cong 103{,}4.$$

La hausse moyenne des prix au cours de l'année 1989 est d'environ 3,4 %.

4. a. Pour qu'il y ait maintien du pouvoir d'achat annuel, il faut faire la somme des revenus sur l'année et non se contenter de comparer les revenus mensuels des mois de décembre. Le maintien du pouvoir d'achat du revenu à la fin de l'année nécessite que le revenu annuel moyen soit de $120\,000 \cdot 1{,}034 = 124\,080$

   b. Pour qu'il y ait augmentation du pouvoir d'achat, le revenu annuel devrait être de :

$$124\,080 \cdot 1{,}012 = 125\,569$$

5. Le revenu mensuel moyen en 1988 est de 10 000 F, pour qu'il y ait maintien du pouvoir d'achat le revenu mensuel devrait être de 10 340. Si le revenu de décembre 1988 était de 10 000 F, le fait que le revenu du dernier mois de l'année 1989 soit multiplié par 1,034 n'indique pas un maintien du pouvoir d'achat annuel.

81. AMBIGUÏTÉ DES ÉVOLUTIONS ÉVALUÉES PAR PLUSIEURS INDICES

Mots-clefs

Raccord d'indices

Énoncé

Nous disposons de la série de l'indice I (base 100 année 0) de l'année 1950 (année 0) à l'année 1964 (année 14) et de l'indice I' (base 100 année 10) de l'année 1960 (année 10) à l'année 1974 (année 24) relatifs à une grandeur G.

Indice I

$I_{0/0}$	$I_{1/0}$	$I_{2/0}$	$I_{3/0}$	$I_{4/0}$	$I_{5/0}$	$I_{6/0}$	$I_{7/0}$	$I_{8/0}$	$I_{9/0}$	$I_{10/0}$	$I_{11/0}$	$I_{12/0}$	$I_{13/0}$	$I_{14/0}$
100	102	105	106	108	112	115	118	120	121	124	127	132	135	138

Indice I'

$I'_{10/10}$	$I'_{11/10}$	$I'_{12/10}$	$I'_{13/10}$	$I'_{14/10}$	$I'_{15/10}$	$I'_{16/10}$	$I'_{17/10}$	$I'_{18/10}$	$I'_{19/10}$	$I'_{20/10}$	$I'_{21/10}$	$I'_{22/10}$	$I'_{23/10}$	$I'_{24/10}$
100	102	103	104	106	108	109	110	112	116	115	116	123	125	127

1. Quel est de l'indice la grandeur G année 24 en base 100 année 0 ?
2. Quel est l'indice de la grandeur G année 20 en base 100 année 6 ?
3. Vous expliquerez les problèmes que posent ces calculs.

Corrigé

1. Nous disposons des valeurs des indices I et I' pour cinq années 10, 11, 12, 13, 14.

$$I^*_{24/0} = I'_{24/10} \cdot C_r$$

$$C_r = \frac{I_{10/0}}{I_{10/10}} = 124$$

Il existe plusieurs coefficients de raccordement (C_r) possibles.

Coefficients de raccordement

0	1	2	3	4
124,0	124,5	128,2	129,8	130,2

Nous obtenons différentes valeurs possibles : 157,5 ; 158,1 ; 162,8 ; 164,9 ; 165,3.

Pour obtenir un résultat, il est nécessaire de calculer une valeur de tendance centrale de cette distribution, soit la moyenne arithmétique 161,7 soit, pour rester au niveau de précision des données de base, 162.

2. Le principe est le même que dans la question précédente, les coefficients de raccordement sont identiques, seule l'année de base est modifiée.

$$I^*_{20/6} = \frac{I_{20/10}}{I_{6/0}} \cdot C_r$$

D'où les différents résultats possibles : 124 ; 124,51 ; 128,16 ; 129,81 ; 130,19. Nous prenons la moyenne donc : 127,3 soit, pour conserver la même précision, 127.

3. Nous comparons des indices dont les structures peuvent être différentes ainsi que les méthodes de calcul (indice de Laspeyres puis indice chaîne par exemple).

Table des matières

Achevé d'imprimer par

CPI

Imprimerie France Quercy
46090 Mercuès

N° d'impression : 71961
Dépôt légal : août 2007

Imprimé en France

Statistiques descriptives. Exercices

Ce livre d'exercices de statistique descriptive vise non seulement à acquérir les techniques de calcul des outils de description et de synthèse des séries statistiques mais également à donner des éléments de pertinence pour l'analyse et la compréhension des résultats trouvés. Il constitue le complément du cours de statistique descriptive qui présente les concepts.

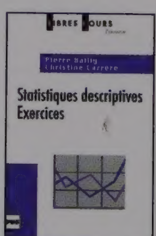
Ce ouvrage, organisé en cinq parties, propose des exercices sur les différents aspects de la statistique descriptive :

- les outils de base de la statistique descriptive (les variables, les nomenclatures, les représentations graphiques) ;
- les caractéristiques décrivant les distributions à une dimension (mode, médiane, moyennes, indicateurs de dispersion et de concentration) ;
- les techniques permettant de mettre en lumière les relations entre variables pour des distributions à deux dimensions (corrélation, ajustement, coefficient de détermination) ;
- les méthodes les plus simples du traitement des séries chronologiques (tendance longue, correction des variations saisonnières) ;
- les calculs des indices, une des informations les plus utilisées dans l'information économique et sociale.

Cet ouvrage s'adresse aux étudiants de première année de licence d'économie-gestion, d'AES (administration économique et sociale) et plus généralement en sciences sociales.

Ce livre accompagne le cours de statistiques descriptives des mêmes auteurs.

Christine Carrère et Pierre Bailly enseignent les statistiques à l'université Pierre-Mendès-France de Grenoble.



Presses universitaires de Grenoble
BP 47 – 38040 GRENOBLE CEDEX 9
www.pug.fr

18 €

ISBN 978-2-7061-1413-7 – ISSN 0298-1882
Code Sofedis-Sodis S368949

9 782706 114137