Module : Techniques quantitatives de gestion

Chapitre 2 : Les graphiques des séries statistiques

I) Les avantages de la représentation graphique

Par la représentation graphique, sous la forme de graphes ou diagrammes construits géométriquement, le statisticien cherche à donner une image plus ou moins fidèle de l'ensemble des données consignées dans les tableaux statistiques. Cette représentation graphique constitue d'autre part une étape préliminaire, obligatoire et riche d'enseignements, de l'analyse statistique.

Les avantages essentiels d'un graphique sont au nombre de 4 :

- Une traduction concrète d'éléments abstraits constituant la série statistique. L'allure générale du diagramme en donne une vue synthétique ;
- Un moyen de contrôle des données initiales. En effet, les erreurs, omissions ou anomalies sont très facilement détectées sur le graphe ;
- Un instrument de comparaison entre différentes distributions similaires. Les graphes permettent la mise en évidence du type de liaison, entre les phénomènes observés simultanément sur un même graphique ;
- Permet l'orientation du statisticien quant au choix de la méthode d'analyse (à appliquer) et des caractéristiques (à déterminer).

Encore faut-il que le graphique soit simple (facile à construire), clair (nombre de renseignements limités), équilibré et précis (par un choix judicieux des échelles), enfin bien adapté au phénomène représenté.

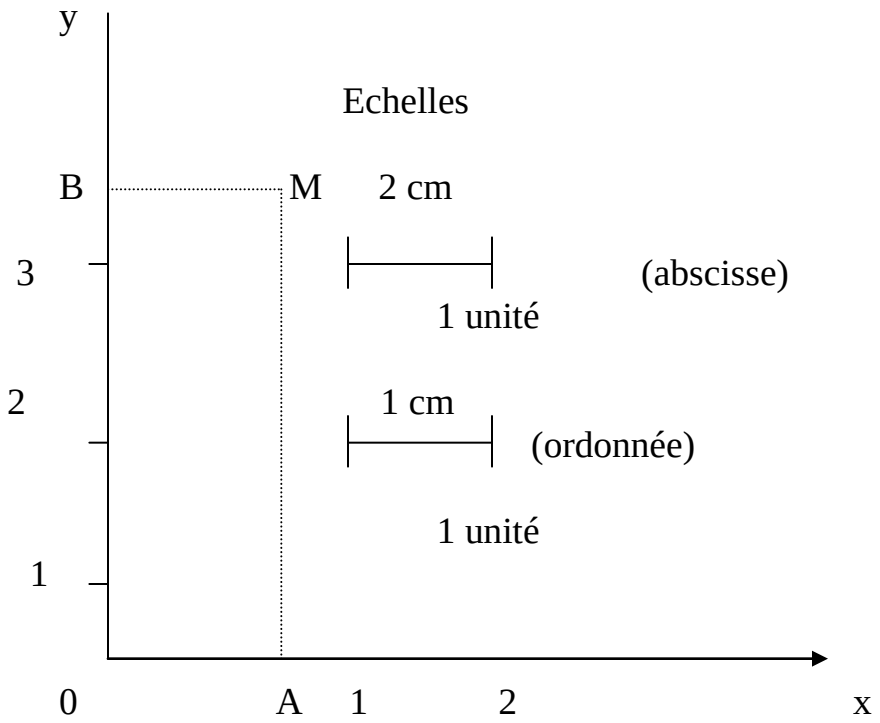
De fait, les graphes sont très nombreux ; la méthode statistique concerne essentiellement les séries de fréquences.

II) Les diagrammes de fréquences

1) La construction générale des graphes

Le graphique est construit dans un repère orthonormé, c'est-à-dire en utilisant de coordonnées cartésiennes (le plus souvent rectangulaires lorsque les échelles sont différentes sur les deux axes) relatives au premier quadrant (abscisses et ordonnées positives)

Graphe cartésien rectangulaire



Dans le système xOy, et en projetant orthogonalement (c'est-à-dire à angle droit) un point quelconque M, on obtient :

$\overline{OA}$ = xm (abscisse du point M)

$\overline{OB}$ = ym (ordonnée du point M)

Ces coordonnées étant mesurées en fonction de l'échelle et donc de l'unité choisie les 2 axes (ici 2 cm correspondent à une unité d'abscisse et 1 cm correspond à une unité d'ordonnée).

Pour les séries de fréquences, le caractère statistique xi figure sur l'axe horizontal Ox et les fréquences absolues (ni) ou relatives (fi) sont portées sur l'axe vertical Oy.

2) Les variantes des diagrammes

Selon le type de variable, les graphes représentant les distributions de fréquences sont de deux types :

- Diagramme en bâtons (lorsque le caractère est discontinu),
- Histogramme (si la variable est continue).

Les diagrammes sont simples lorsque les fréquences (portées en ordonnées) sont simples.

Les graphes sont cumulatifs dans le cas contraire (ni ou fi cumulés).

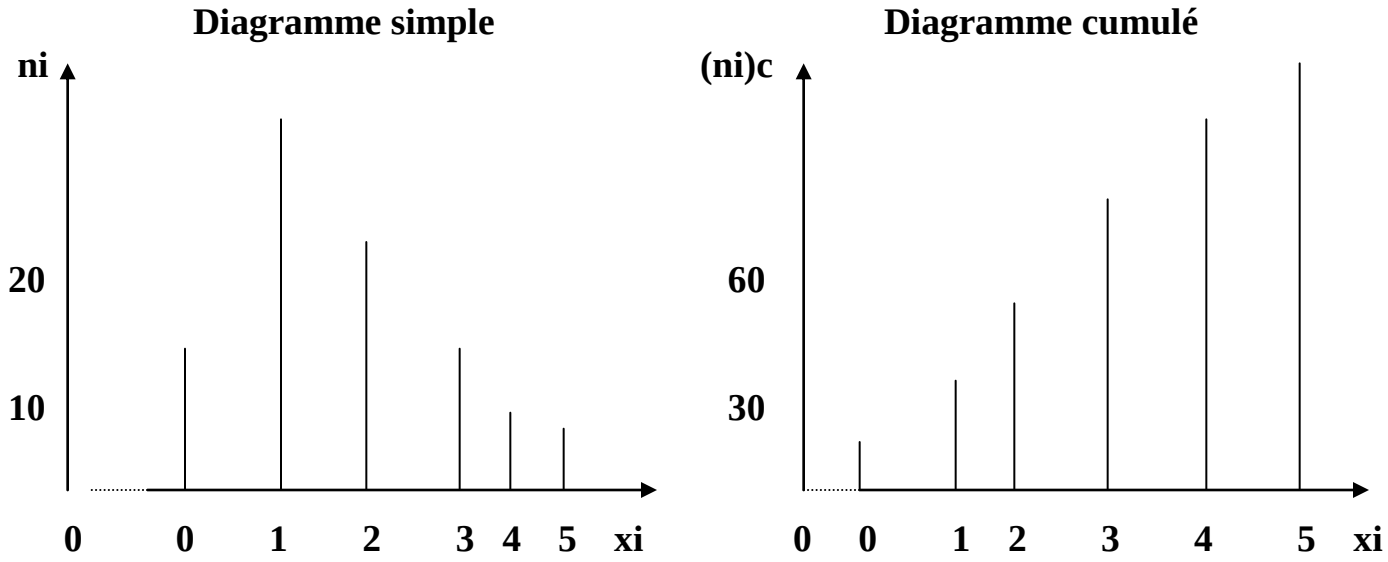
3) Les méthodes de construction des graphes

- Les diagrammes en bâtons

Sur l'axe des abscisses, on positionne régulièrement les différentes valeurs isolées de la variable x. la différence entre deux valeurs successives représente l'unité d'abscisse (relative à l'échelle choisie). Les fréquences figurent sur l'axe des ordonnées avec une échelle différente (ou non).

Pour chaque valeur xi, on trace un bâton, perpendiculairement à l'axe des abscisses et dont la longueur est proportionnelle à la fréquence correspondante : le sommet de chaque bâton a donc pour coordonnées (xi, ni) ou (xi , fi).

a) Diagramme en bâton



Si la 1ère valeur de xi se confondait avec l'origine 0, le 1er bâton 0 serait peu visible (sur l'axe des ordonnées)

Nombre d'enfants xi	Fréquences simple des ménages ni	fi	Fréquences cumulées des ménages (ni)c	(fi)c
0	24	0,15	24	0,15
1	48	0,30	72	0,45
2	40	0,25	112	0,70
3	24	0,15	136	0,85
4	16	0,10	152	0,95
5	8	0,05	160	1,00
	160	1,00		

Les fréquences cumulées de la dernière ligne du tableau statistique sont égales à la somme des fréquences :

(n6) = $\sum_{i}$ ni = 160 et (f6)c = $\sum_{i}$ fi = 1,00.

Concernant les diagrammes, on interrompt le début de l'axe des abscisses, afin de bien dissocier l'origine 0 des fréquences absolues et la première valeur de xi (0 enfant). De plus, on notera que l'échelle des ordonnées est modifiée pour les fréquences cumulées par rapport à celle des fréquences simples.

b) L'histogramme

Lorsque l'amplitude de classe (a) est constate, la différence entre deux bornes successives (également constate) sera prise comme unité d'abscisse (relative à l'échelle choisie).

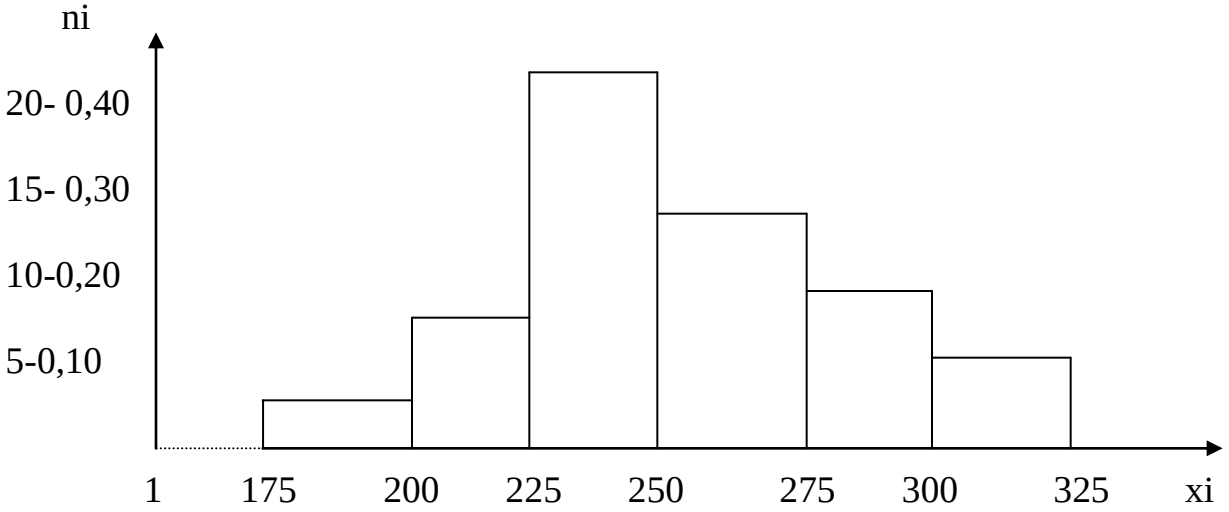
Pour la série n° 3, on choisit pour unité d'abscisse a = 25 et l'échelle sera cm = 1 unité d'abscisse.

L'histogramme est constitué par une succession de rectangles « accolés » ou adjacents, dont les bases (B) sont précisément les intervalles de classe unitaires et dont les hauteurs respectives (H) sont mesurées par les fréquences (absolues ou relatives) correspondant aux centres de classe xi (ou moyennes arithmétiques des bornes).

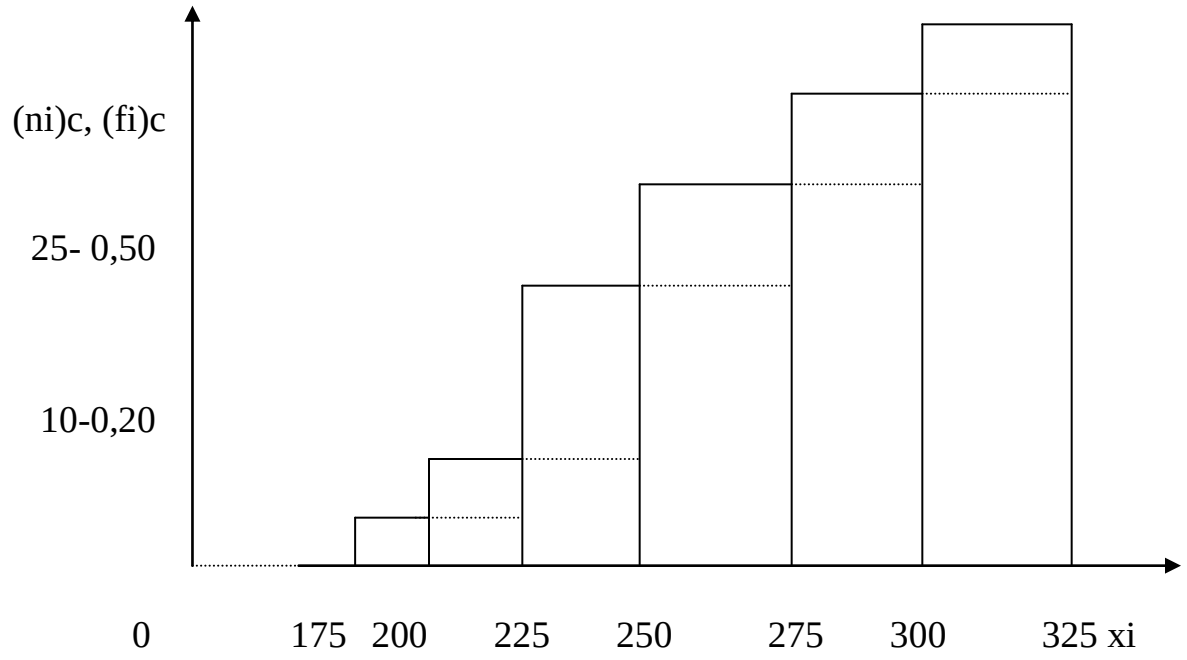
Si l'axe des ordonnées porte des effectifs ni, l'aire de l'histogramme sera mesurée par N = $\sum_{1}$ ni. Si l'axe des ordonnées porte des fréquences relatives fi, l'aire de l'histogramme sera mesurée par $\sum_{1}$ fi = 1.

Comme fi = ni / N, l'histogramme (avec ni) devient l'histogramme (avec fi) par simple changement d'échelle (sur l'axe des ordonnées).

- Histogramme simple



- Histogramme cumulé



III) De l'histogramme a la courbe de fréquences (série n° :3)

Tranche De Revenu	Fréquence absolue		Fréquence cumulée (jusqu'à)		Fréquence cumulée (plus de)	
i	ni	fi	(ni)c	(fi)c	(ni)c	(fi)c
175 à 200	4	0,04	4	0,04	100	1 ,00
200 à 225	16	0,16	20	0,20	96	0,96
225 à 250	36	0,36	56	0,56	80	0,80
250 à 275	22	0,22	78	0,78	44	0,44
275 à 300	14	0,14	92	0,92	22	0,22
300 à 325	8	0,08	100	1,00	8	0,08

La lecture du tableau relatif à la troisième série :

Il existe deux façons de cumuler les fréquences absolues ou relatives :

- De haut en bas du tableau : cumul descendant « jusqu'à ». dans ce cas, les limites supérieurs des classes sont concernées ;
- De bas en haut du tableau : cumul ascendant « plus de ». dans ce cas, les limites inférieures de classes sont concernées.

En réalité, les bornes supérieures ne sont pas exactement atteintes.

Ainsi, pour la série no 3, doit-on lire les différentes classes de la façon suivante : 175 à moins de 200, 200 à moins de 225… Ce qui-signifie que, pour chaque classe, le revenu le plus élevé pourra être aussi voisin que possible de la borne supérieure sans jamais l'atteindre : il en différera toujours d'une certaine quantité si minime soit-elle (1 centime par exemple) qui sera notée ε (« epsilon »). On dira que la borne supérieur est atteinte à « ε » près.

Lecture de la 3ème ligne

36 m »nages (représentant 36 % de la population) gagnent entre 225 et 250 ; 56 ménages (représentant 56 % de la population) gagnent jusqu'à 250 (borne non comprise) et 80 ménages (représentant 80 % de la population) gagnant plus de 225 (borne comprise).

- L'histogramme est une surface

La surface de chaque rectangle élémentaire, ayant la même base (B = 25 = 1 unité d'abscisse) et une hauteur correspondant à une fréquence déterminée (ni ou fi), est mesurée par la valeur du produit B x H :

B x H = 1 x ni (pour les fréquences absolues) : la surface du premier rectangle est donc 4.

B x H = 1 x fi (pour les fréquences relatives) : la surface du deuxième rectangle est donc 0,16.

La surface totale de l'histogramme est mesurée par la somme des fréquences :

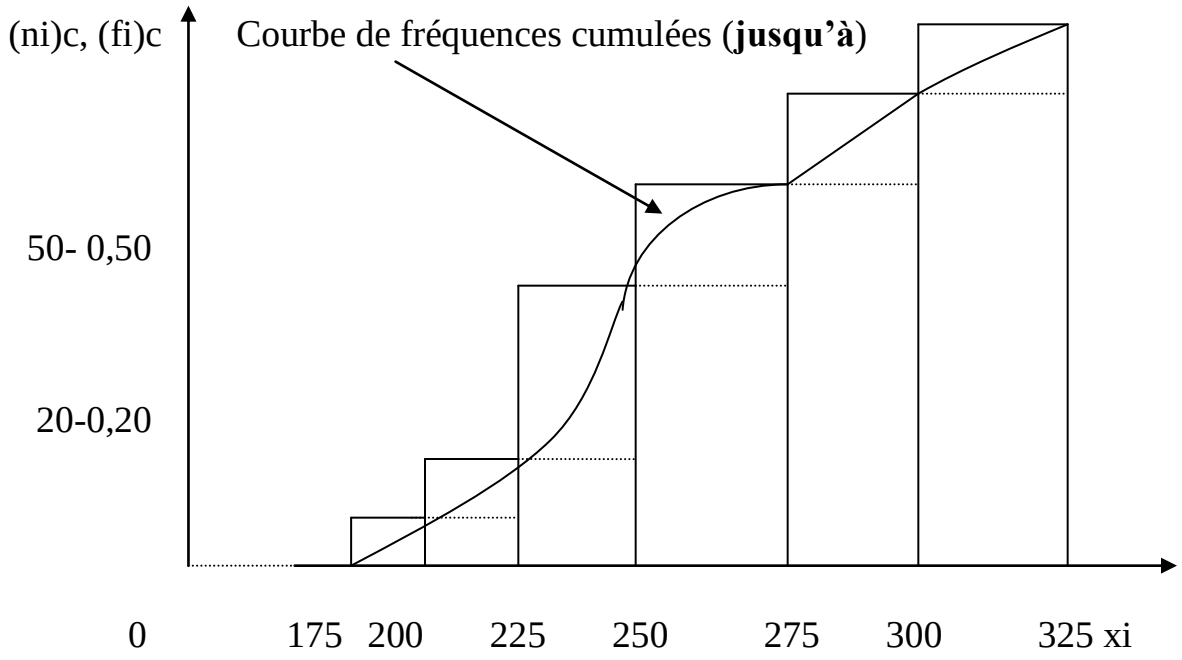
$$\sum_i ni = 100 \quad \text{ou} \quad \sum_i fi = 1$$

($\sum_i fi = 100$ si les valeurs de fi sont exprimées en %)

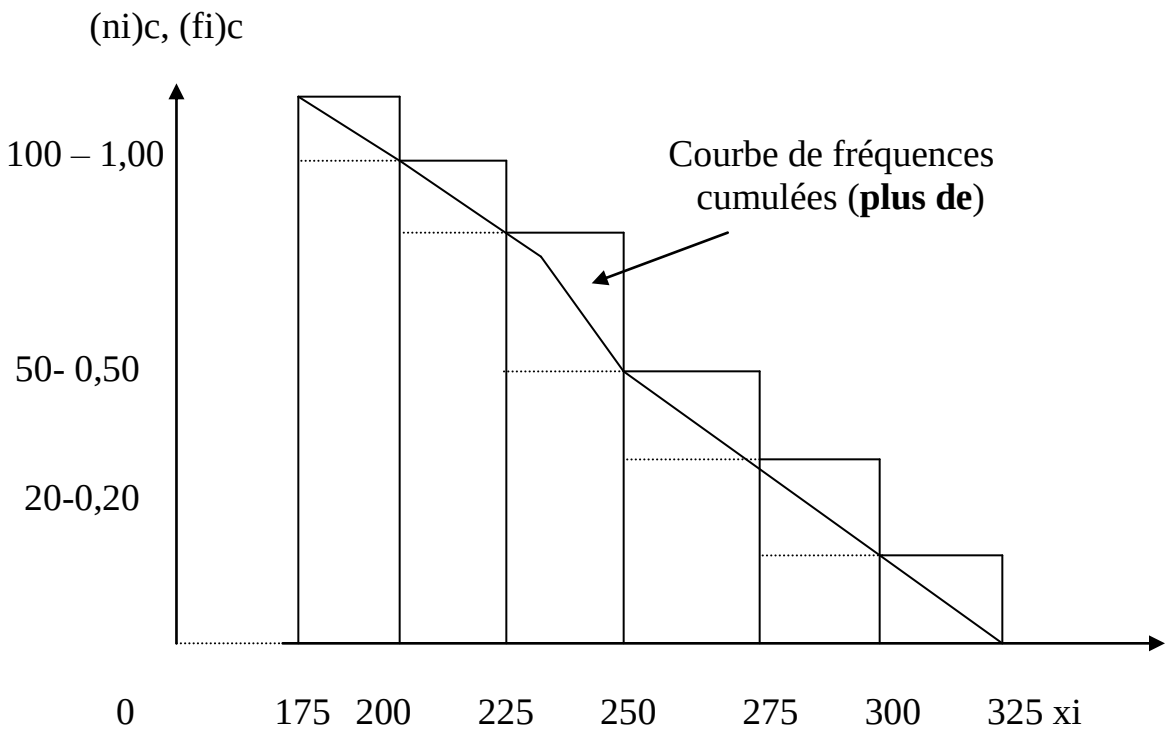
IV) Courbe de fréquences

Théoriquement, lorsque la taille de la population croît indéfiniment (effectif tend vers l'infini) alors que l'intervalle de classe tend cers 0 (le nombre de classes tend également vers l'infini), les surfaces des rectangles élémentaires se « réduisent à une infinité de bâtons accolés » ; la courbe enveloppante reliant leurs extrémités devient alors la courbe de fréquences.

- Courbe de fréquences cumulées



Le début de l'axe des abscisses est interrompu, sinon les distances des différentes bornes à l'origine seraient erronées.



1) L'ajustement peut être analytique (recherche de la fonction qui ajuste le mieux le polygone).
2) Pour l'histogramme cumulatif, on assimilera le polygone à la courbe de fréquences.

V) Propriétés des courbes de fréquences

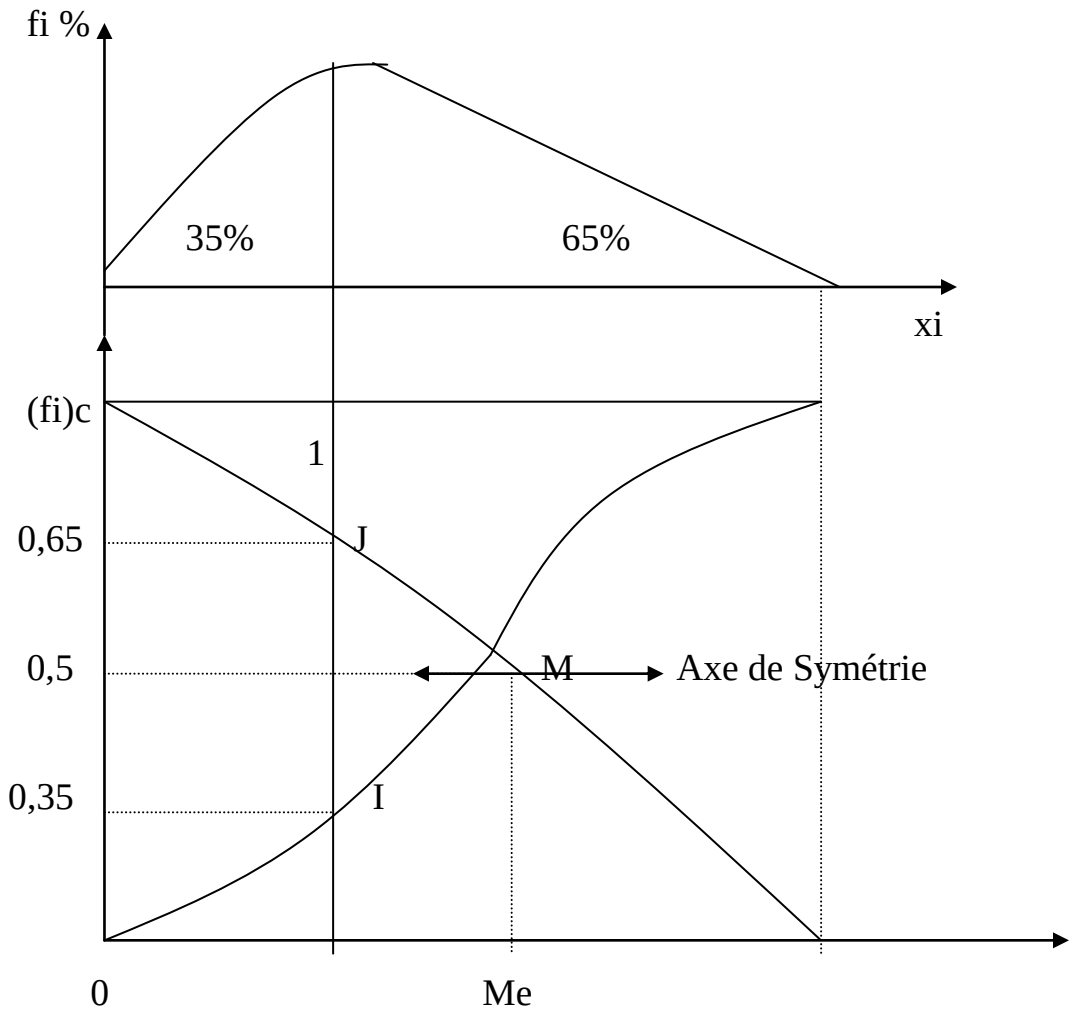
Lorsqu'on fait correspondre une courbe de fréquences cumulée (jusqu'à) avec une courbe de fréquence simples, toute droite parallèle à l'axe des ordonnées :

- partage l'histogramme simple en deux aires (gauche et droite) ;
- coupe la courbe de fréquences cumulées en un point dont l'ordonnée mesure l'aire gauche.

Si l'ordonnée du point d'intersection I est 0,35, cela signifie que l'aire gauche représente 35% de la surface totale de l'histogramme.

Lorsqu'on fait intervenir la courbe de fréquences cumulées (plus de), la même droite coupe la courbe en un point dont l'ordonnée mesure l'aire droite.

L'ordonnée du point d'intersection J (pour la même droite) sera 0,65 (l'aire droite représente 65% de la surface totale). On vérifie bien que l'aire de l'histogramme est égale à 100 (si les fréquences sont exprimées en %).



Les deux courbes de fréquences cumulées (« jusqu'à » et « plus de ») se coupent en un point M d'ordonnée 0,50 et ayant pour abscisse la valeur de la médiane. L'axe de symétrie est la parallèle à l'axe des abscisses passant par le point M.

Les histogrammes et leurs courbes de fréquences simples que nous venons d'observer sont de type asymétrique.

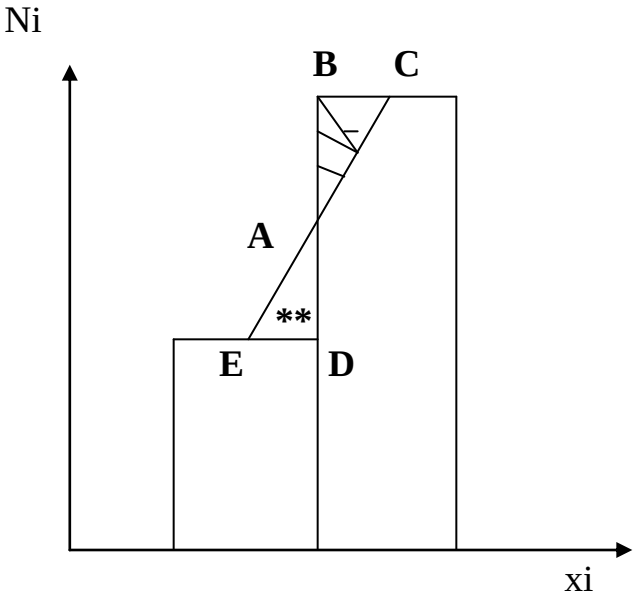
Par contre, la plupart des distributions économiques (répartition des entreprises selon leur taille, mesurée par le chiffre d'affaires ou le nombre d'ouvriers …) sont représentées par des histogrammes et des courbes de fréquences de type hyperbolique (plus communément appelés « J renversé ». Il en est de même pour les statistiques fiscales (il existe peu de contribuables titulaires de revenus élevés et beaucoup d'assujettis des revenus moyens ou modestes).

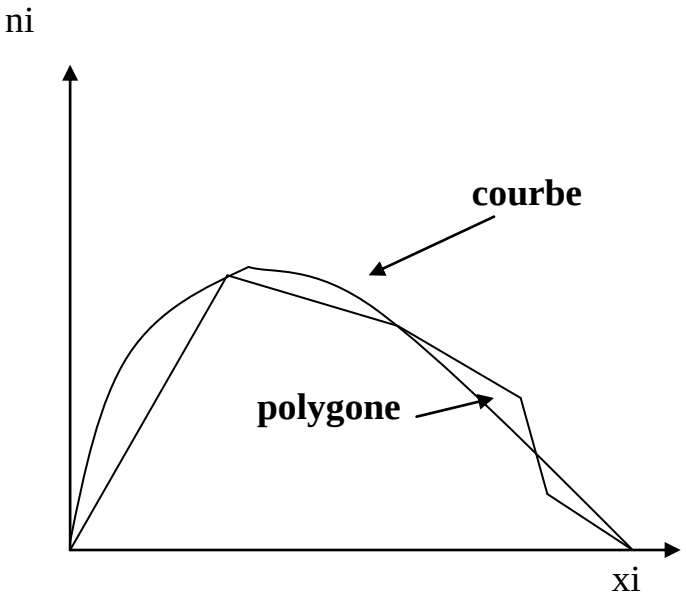
Le statisticien est donc nécessairement amené, lors de l'établissement du tableau statistique, à choisir des intervalles de classe inégaux et croissants pour « saisir une fréquence minimale » pour chaque classe sinon certaines classes relatives aux hauts revenus ne comporteraient pas de fréquences et il y aurait « rupture » dans la construction de l'histogramme.

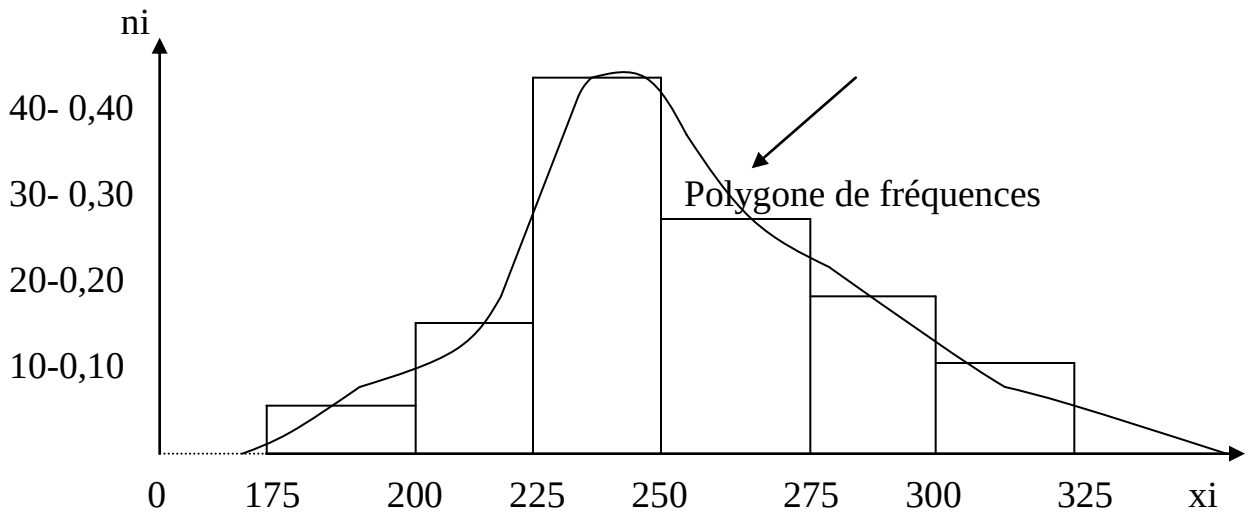
VI) Polygone des fréquences

Le polygone de fréquences simples est la ligne qui joint le milieu des paliers supérieurs. Aux extrémités de la distribution, on complète le tableau par deux classes « fictives » (1 50 à 175 et 325 à 350) de fréquence nulle, qui vont permettre de relier le polygone à l'axe des abscisses. Ainsi la surface comprise entre le polygone et l'axe des abscisses est celle de l'histogramme.

En effet, il y a conservation de l'aire puisque les triangles « retranchés » (aire hachurée) compensent exactement les triangles « ajoutés » (aire ponctuée) :
aire ABC = aire EAD.







VII) Histogramme et intervalle de classe inégal

Considérons la distribution, d'origine fiscale, suivante : La répartition d'une population de contribuables dont les fréquences relatives sont exprimées en % fonction de la tranche de revenu imposable, exprimé en milliers de Dirhams. L'intervalle de classe (ai) est croissant et suit une progression de type arithmétique (20, 40, 60, 80, 100). En prenant 20 comme intervalle de classe unitaire, les amplitudes de classe varient selon la progression :

1(= 20/20), 2(= 40/20), 3(= 60/20), 4(= 80/20) et 5(= 100/20)

Nous savons que la surface de chaque rectangle élémentaire (B x H) est proportionnelle à la fréquence. Or, les rectangles successifs ont des bases respectivement multipliées par les coefficients 1, 2, 3, 4 et 5.

Pour conserver la même surface, il faudra donc diviser les « hauteurs » correspondantes (c'est-à-dire les fréquences) par ces mêmes coefficients. La surface sera alors, seulement, proportionnelle aux fréquences fi.

Revenus imposables i	Nombre de contribuables fi (en %)	Intervalle ai	Coefficients Ci = ai/20	Hauteurs corrigés fi/Ci	Surface rectangles BxH
60 à 80	28	20	1	28	1x28 = 28
80 à 120	40	40	2	20	2x20 = 40
120 à 180	15	60	3	5	3x5 = 15
180 à 260	12	80	4	3	4x3 = 12
260 à 360	5	100	5	1	5x1 = 5

On vérifie bien dans le tableau que la surface de chaque rectangle est proportionnelle aux fréquences (%).

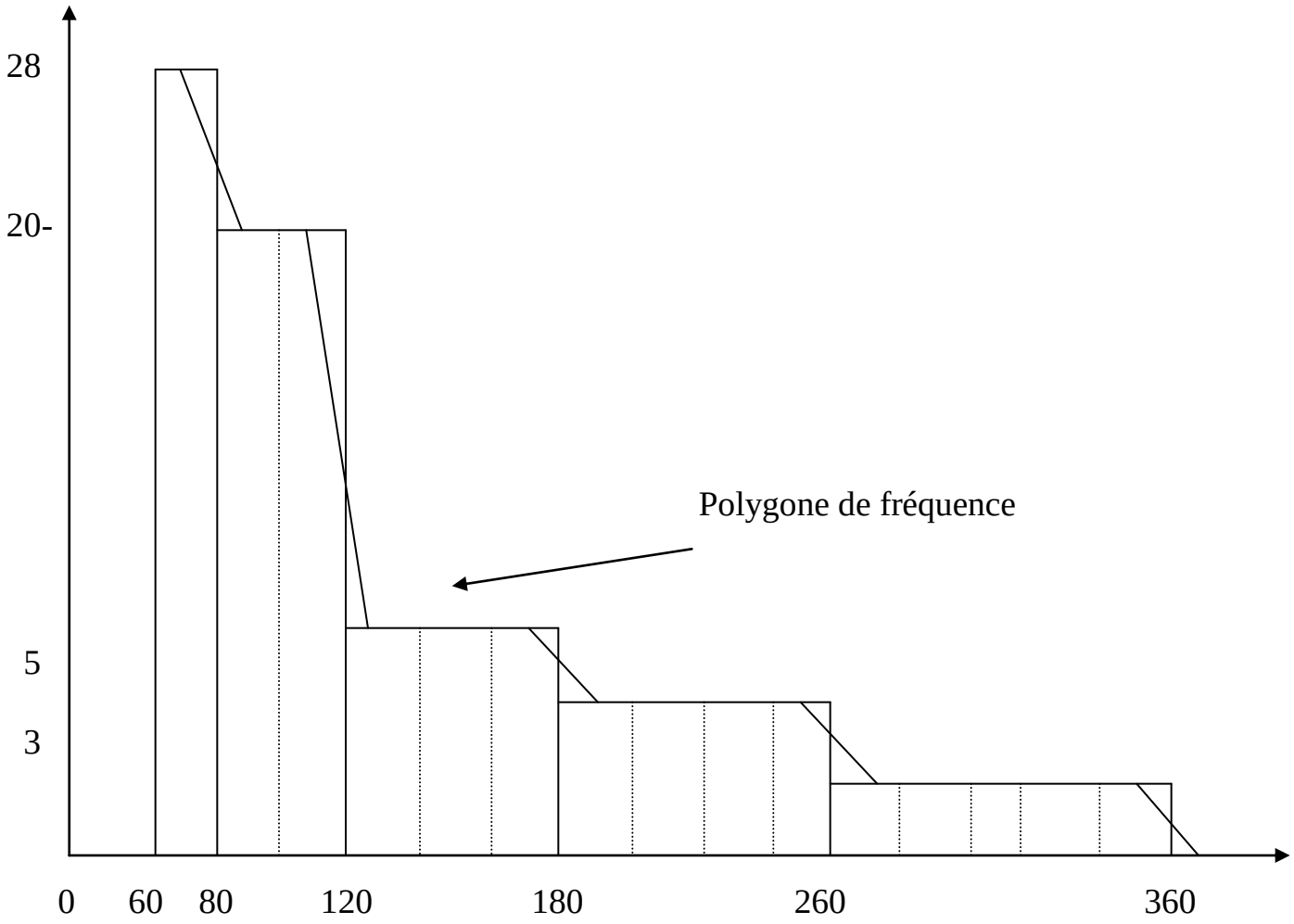
Sur le graphique, le « cloisonnement en pointillés » des rectangles met en évidence les différentes bases unitaires (à partir desquelles on détermine les valeurs de Ci).

Le polygone de fréquence comporte des paliers de plus en plus importants traduisant des intervalles de classe croissants.

On remarquera que le polygone relie les milieux des paliers d'amplitude unitaire (20) pour respecter la conservation des aires.

L'ajustement du polygone est une courbe de fréquences (en J renversé).

Fréquences



- Classes ouvertes et fermées

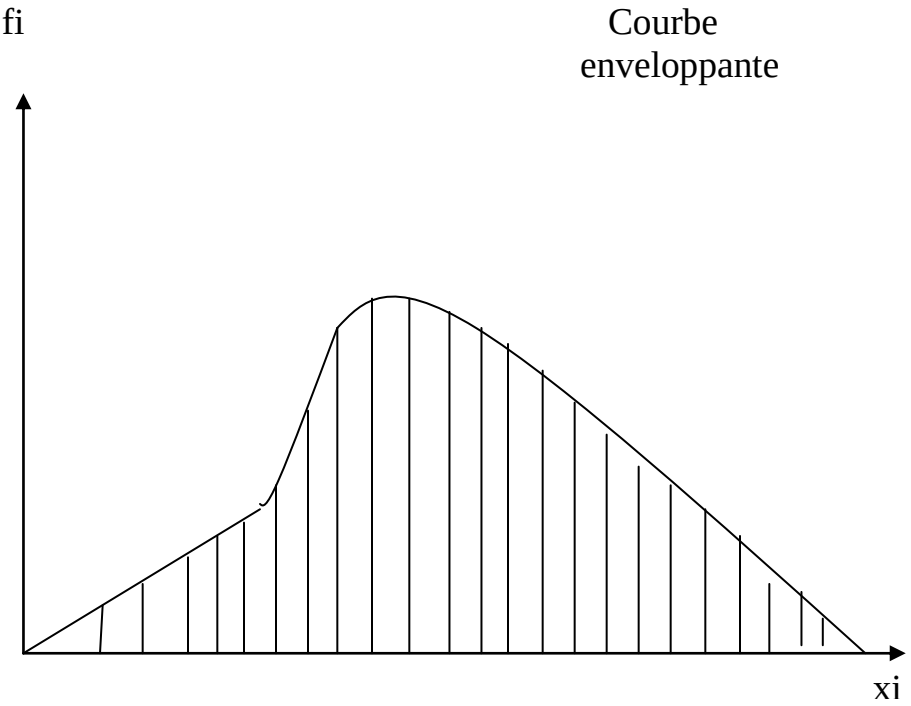
En générale, afin de ne pas multiplier le nombre de classes, les tableaux relatifs aux distributions économiques peuvent présenter aux extrémités des classes ouvertes sans la mesure où une seule borne est mentionnée, par exemple : moins de 80, 80 à 120, 120 à 180, 180 à 260, plus de 260 pour la série précédente.

Il s'agit alors de « fermer » les classes par l'estimation des « bornes » complémentaires ». C'est l'évolution de l'amplitude des classes intermédiaires fermées qui permet une estimation rationnelle qui devient facile à réaliser dans le cas d'une progression arithmétique.

A noter que :

1) le calcul statistique s'applique directement, **sans transformation**, aux distributions de fréquences dont l'intervalle de classe est inégal.
2) Schéma de la courbe de fréquences

Hypothèse : si N $\longrightarrow \infty$ et a $\longrightarrow 0$



VIII) Les diagrammes sans fréquences :

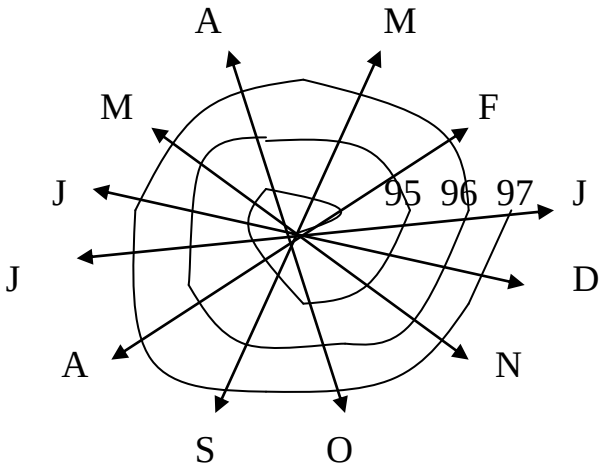
1- Graphe à coordonnées polaire :

Ces graphiques avec la représentation de séries chronologiques concernant un phénomène tendant à s'accroître ou à diminuer dans le temps et présentant des fluctuations saisonnières importantes. Le prince de cette représentation graphique repose sur l'utilisation des coordonnées polaires pour repérer les points.

A partir d'un pôle nommé O (origine), tout point M est repéré par son module p et son argument θ tels que :

$$\theta = (\overrightarrow{Ox},\overrightarrow{Ox}) \text{ et } p = \overrightarrow{OM}$$

Ainsi pour représenter des séries mensuelles relatives à l'indice des prix pour les différents mois des années 1995, 1996 et 1997, il faudra construire douze axes tels que OM. Les arguments dans ce cas seront : $\theta = \pi /6$

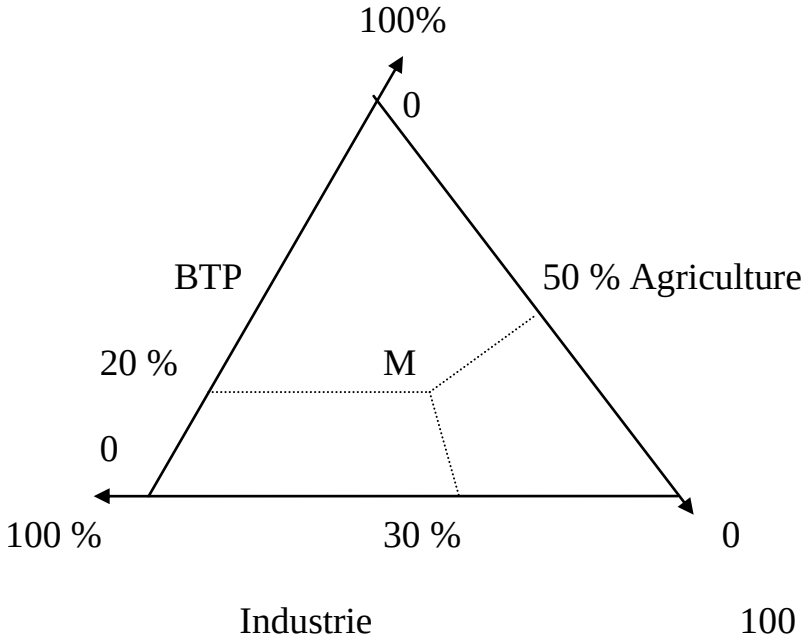


2- Graphique triangulaire :

Ce type de graphique est utilisé pour représenter une quantité constante fractionnée en trois éléments variables. C'est le cas, par exemple, de la répartition de la population active en trois branches (Agriculture, Industrie, Bâtiment et travaux publics)

On suppose que la part de chaque branche est la suivante :

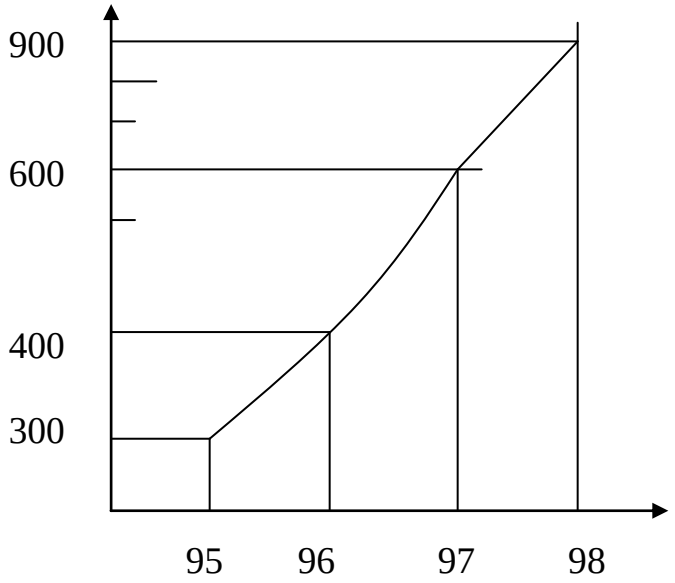
Agriculture	50 %
Industrie et commerce	30 %
Bâtiments et travaux publics	20%



3- Graphes logarithmiques et semi-logarithmiques

L'utilisation des représentations logarithmiques ou semi-logarithmique présente une utilité certaine dans le cas de comparaison des données de deux variables dont les valeurs sont très différentes (cas de la production d'une petite entreprise et celle d'une firme multinationale). En effet les repères à l'échelle arithmétique sont basés sur la proportionnalité des graduations aux valeurs de la variable. C'est le cas, par exemple, de la représentation graphique des ventes d'une entreprise sur une période donnée.

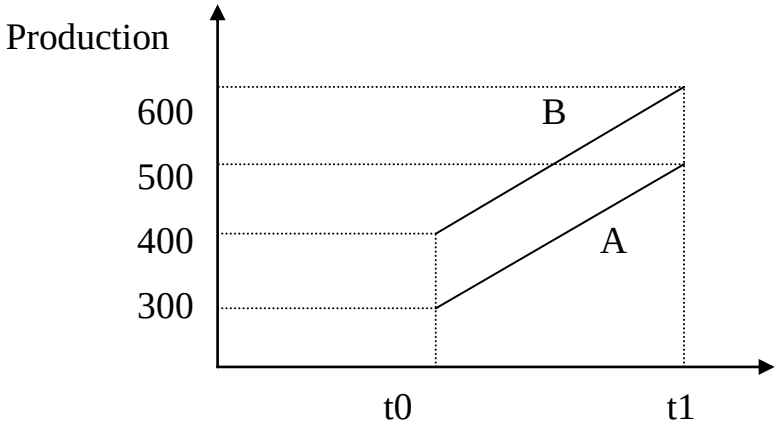
Années	Ventes
1995	300
1996	400
1997	600
1998	900



L'inconvénient de ce type de représentation se pose en cas de comparaisons de variables à valeurs très différents. Si les valeurs prises par une variable donnée sont très importantes par rapport à celles des autres variables, le graphique de cette variable sort de la feuille.

Un deuxième inconvénient de ce type de représentation est le fait que les échelles arithmétiques traduisent mal les variations relatives. C'est ce qui apparaît sur le graphique qui représente l'évolution des ventes de deux entreprises : A et B. La production de l'entreprise A passe de 300 à 500 et la production de l'entreprise B passe de 400 à 600.

Les évolutions apparaissent parallèles alors que les productions ont augmenté dans des proportions différentes.

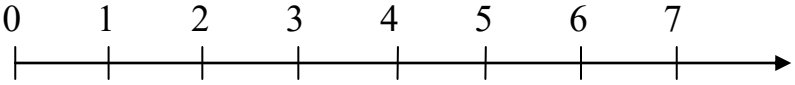


L'avantage de l'échelle semi-logarithmique est qu'il permet d'annuler ces inconvénients.

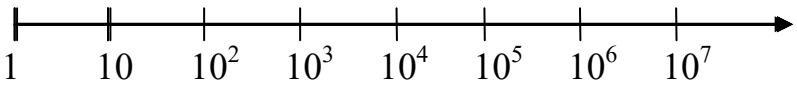
Le recours à l'échelle semi-logarithmique nécessite l'utilisation du papier semi-logarithme : ce papier comprend une échelle logarithmique décimale (en vertical) et une échelle arithmétique en horizontal. En partant des logarithmes des valeurs des variables à représenter graphiquement, l'axe logarithmitique peut être construit de la manière suivante :

$$\text{Log}1 = 0, \log 10 = 1, \log 100 = 10, \ldots, \log (10)^n = n$$

Au lieu de reporter une même distance telle que :



On reporte une distance proportionnelle aux logarithmes décimaux



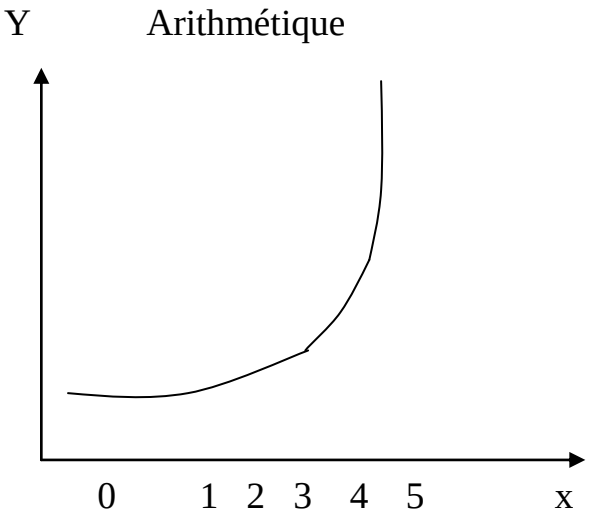
Dans ces deux cas les distances successives sont identiques : la distance entre 0 et 1 est la même que celle entre 1 et 2 sur l'axe arithmétique de même que la distance entre 10 et 100 est la même que celle entre 100 et 1000 sur l'axe logarithmique.

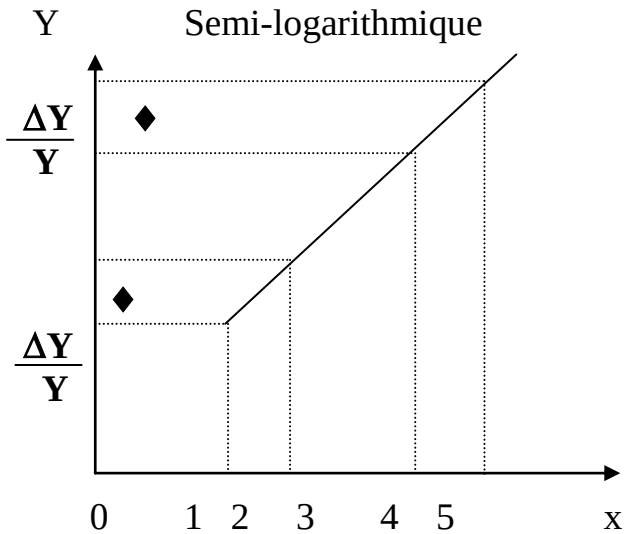
Remarque dans un diagramme semi-logarithmique, des longueurs égales représentent des rapports égaux, donc des variations relatives égales :

$$\frac{1}{10} = \frac{10}{100} = \frac{100}{1000} = 10\%$$

Alors que dans un diagramme arithmétique elles représentent des variations absolues égales.

1cm = 10 centimes ce qui entraîne 2 cm = 20 centimes et 3 cm = 30 centimes



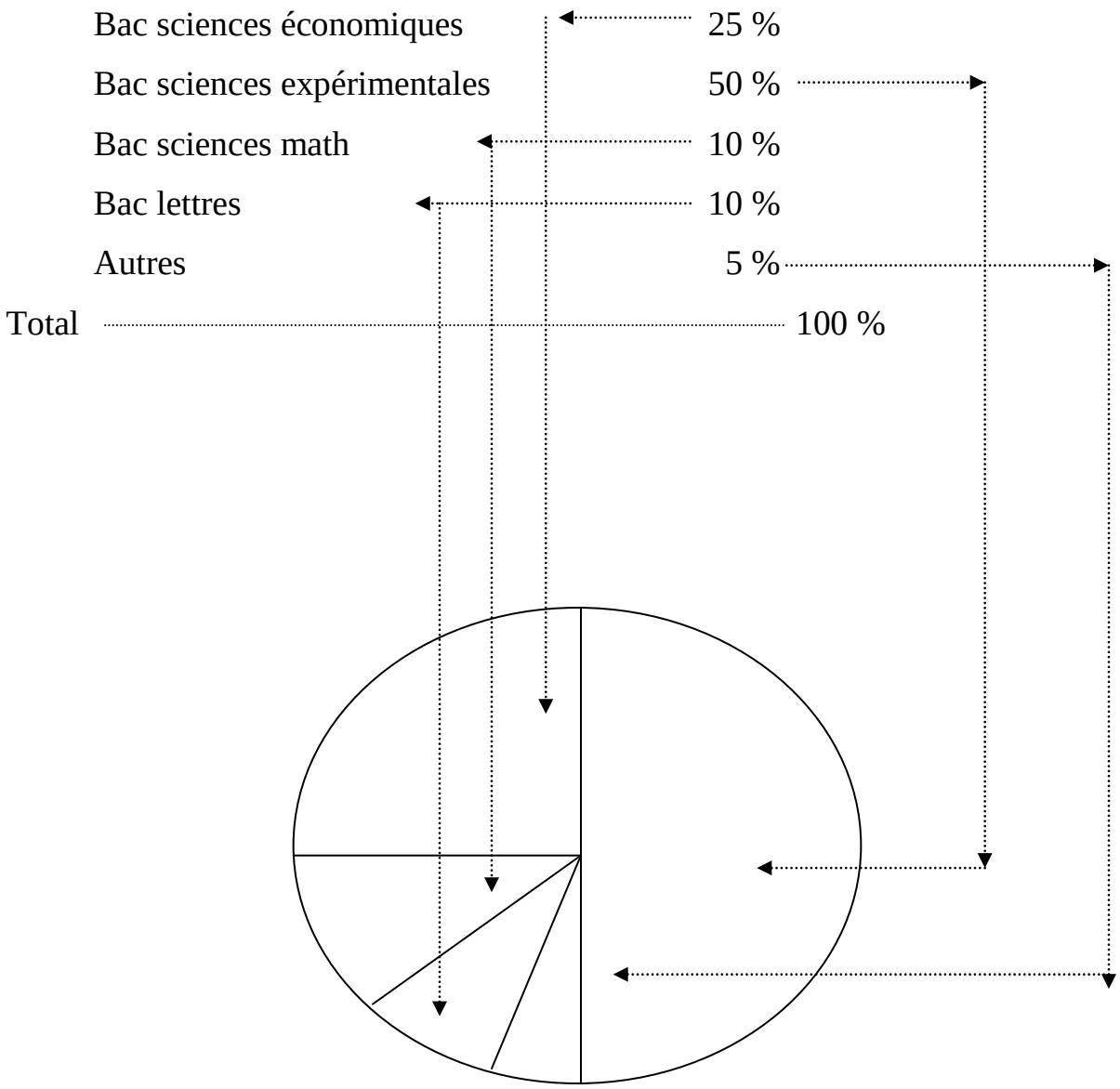


Un phénomène dont le taux de croissance annuel est constant sera représenté en semi-logarithmique par une droite.

4- Graphes circulaires à secteurs :

Ce type de graphe permet de visualiser des parts relatives, dans des surfaces ou secteurs de cercles, qui peuvent être différenciés par des couleurs ou des motifs différents. Chaque secteur correspond à une modalité et l'angle au centre est égal au produit de 360° par la fréquence fi selon la formule suivante : α = fi 360°.

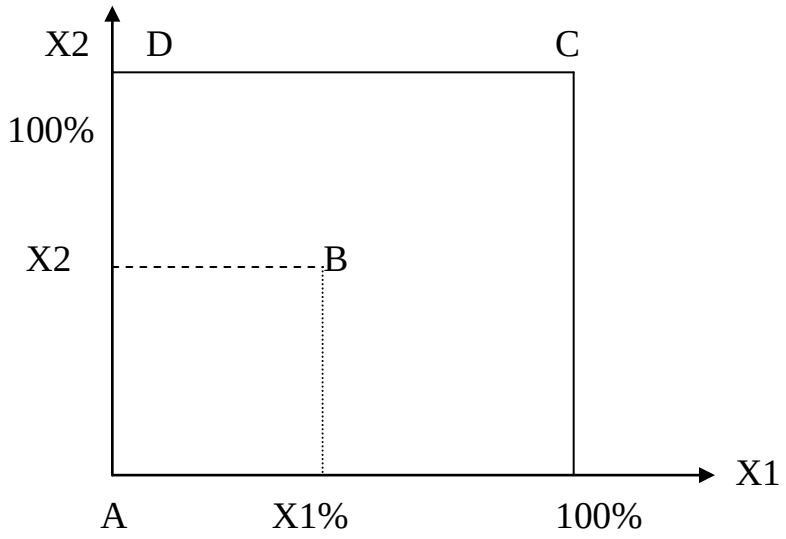
Répartition des étudiants de la première année sciences économiques selon le type du Bac obtenu (chiffre fictifs) :



5- Graphes carrés :

Ces graphes sont basés sur un principe fondamental qui consiste à utiliser ces graphes pour représenter des points d'un plan vérifiant une double condition.

Considérons un point P à l'intérieur d'un carré ABCD.



Les axes portent des pourcentages

B a pour abscisse 100%

D a pour ordonnée 100%

et P a pour coordonnées (Xi, X2)

La double condition représentée par le point P dans ce carré est

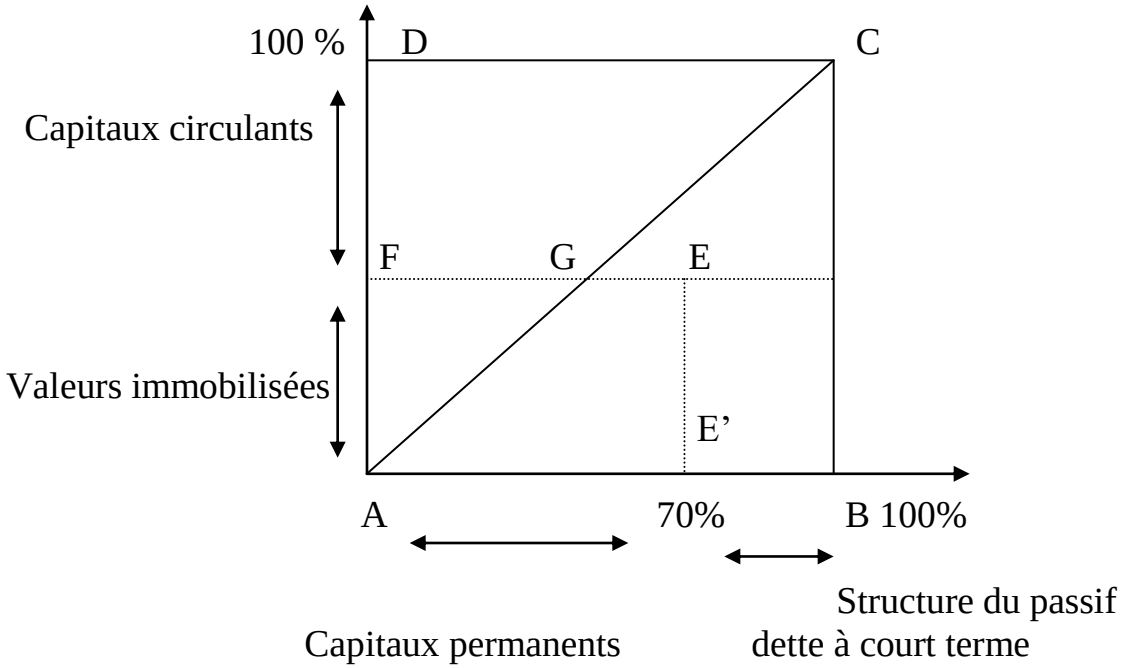
$X1 + (100-X1) = X2 + (100-X2) = 100\%$

Exemple : le bilan d'une entreprise à la date t0

Valeurs immobilisées (I)	5000	capitaux propres	7000
Stocks	2000	fournisseurs plus	
Clients + effets à recevoir	2000	effets à payer	2000
Caisse	1000	Autres dettes	
		à court	1000

- les immobilisations (I) représentent 50% de l'actif,
- les capitaux circulants (CC) (Stocks plus clients plus effets à recevoir plus caisse) représentent 50% de l'actif,
- les capitaux permanents (CP) (capitaux propres plus dettes de long terme et dettes à moyen terme) représentent 70% du passif ;
- les dettes à court terme (Dc) (Fournisseurs effets à payer plus autres créanciers) représentent 30%.

4- Structure de l'actif



Sur ce graphique Cp = AE' = FE

I = AF = FG

Donc : Le fond de roulement (FR) = Cp – I = FE – FG = GE

Toutes les entreprises représentées par des points situés à l'intérieur du triangle ABC ont un fonds de roulement positif.