2 SÉRIES STATISTIQUES À UNE VARIABLE

Dans la suite, les différentes phases de présentations des données et de calculs de paramètres statistiques seront présentées sur deux exemples types sous les références « *exemple 1* » et « *exemple 2* ».

Exemple 1

Des enquêteurs ont rencontré 280 propriétaires de véhicules et leur ont demandé de renseigner l'âge de celui-ci, en terme d'années complètes.

3	2	1	2	2	3	4	0	1	0	2	0	3	2
0	9	1	0	4	2	1	0	6	2	2	2	3	2
3	3	3	2	4	0	2	2	2	5	0	1	1	4
2	1	1	2	2	1	0	0	4	2	0	4	2	0
0	1	2	5	4	0	2	1	2	2	0	4	0	1
0	1	1	0	0	0	0	1	3	1	1	2	6	1
3	1	4	4	2	0	2	0	0	7	3	0	1	0
6	0	4	2	2	4	1	0	1	1	1	1	1	2
3	3	2	2	0	4	3	0	3	0	4	2	0	0
0	3	1	1	1	5	2	2	2	2	1	2	4	4
0	1	1	3	3	1	1	1	0	3	5	3	1	1
2	2	1	0	0	1	1	1	4	0	2	0	0	6
0	1	1	5	4	0	2	2	4	1	3	1	0	1
4	2	1	3	2	2	3	1	5	1	1	2	2	2
1	2	4	1	2	5	1	1	2	0	1	0	0	4
5	2	1	5	1	0	2	2	4	2	1	0	1	2
6	0	0	1	2	2	2	4	0	0	4	0	1	1
2	1	3	0	0	2	0	2	2	1	2	1	4	2
2	0	1	0	4	1	1	3	3	1	1	2	0	3
2	4	1	2	2	1	2	1	3	4	5	2	1	4

Tableau 2.1

Exemple 2

Il convient d'étudier le prix payé pour les entretiens annuels d'un véhicule de moins de 4 ans d'âge. Les enquêteurs rendent visite à 75 propriétaires de voitures. Ils notent les sommes déclarées (en euros).

800	675	425	400	650	350	775	600	300	475	550	375	350	275	200
500	275	450	450	325	475	425	300	550	550	200	275	400	675	450
200	600	675	325	325	425	225	400	200	725	450	500	425	475	475
775	500	225	400	400	625	325	425	375	600	450	200	425	250	450
725	225	275	325	300	525	300	525	500	550	550	675	575	525	600

Tableau 2.2

1. Tableau brut

Les tableaux présentés ci-dessus pour les deux exemples sont des *tableaux bruts*, ce sont les listes des données ou des valeurs du caractère fournies par l'enquête.

2. La série est ordonnée

$\{x_1, x_2, x_3, \dots, x_n\}$ si les valeurs ont été ordonnées par ordre croissant par exemple et numérotées.

3. Le tableau recensé

Il reprend les valeurs ou les données du tableau brut, le plus souvent présentées en colonnes.

- La première colonne reprend les modalités ou les valeurs prises par le caractère étudié.
- La deuxième colonne donne le nombre de fois (effectif n_i) que chaque modalité ou chaque valeur (x_i) apparaît dans la liste.

La distribution ordonnée peut donc être présentée dans le tableau des effectifs. Dans la suite, n_i désigne l'effectif de la modalité x_i de même numéro. Il est évident que, si n désigne l'effectif total de la série, alors :

$$n = \sum_i n_i \ .$$

Pour l'exemple 1

L'âge des véhicules recensés varie de 0 à $9\,ans$. Ce sont des nombres entiers et leur liste présente moins de 20 possibilités. Le caractère discret est donc indiscutable. Le recensement (nombre de véhicules de moins de 1 an, de 1 an, de 2 ans, etc.) permet de construire le tableau des effectifs de 10 modalités ($x_i : i = 1, \dots, 10$).

Valeurs **Ages en années** x_i	**Effectifs** **Nombre de véhicules** n_i
0	60
1	74
2	72
3	27
4	30
5	10
6	5
7	1
8	0
9	1
Total	$n = 280$

Tableau 2.3

Dans ce tableau :

- x_i représente les valeurs du caractère étudié (âges en années complètes des véhicules),
- n_i le nombre de fois qu'apparaît la valeur x_i ou effectif de x_i (nombre de véhicules de x_i *ans*).
- n est l'effectif total de la population $n = \sum_{i=1}^{10} n_i$.

Pour l'exemple 2

Les modalités sont entières mais le nombre de réponses possibles est supérieur à 20. Trop peu de mesures sont identiques. Une statistique sur caractère discret n'aurait que peu d'intérêt car elle ne pourrait mettre en évidence que des constatations valables pour une ou deux voitures et aurait donc peu de signification.

Il faut donc effectuer des groupements préalables en classes, c'est à dire en catégories regroupant des valeurs relativement proches les unes des autres. La distribution est alors une « *distribution groupée* ».

Il s'agit d'abord de choisir un nombre J de classes, chaque classe est identifiée par un nombre entier i ($1 \le i \le J$) et caractérisée par :

- sa limite inférieure et sa limite supérieure (bornes) ;
- son centre noté x_i, calculé en faisant la moyenne arithmétique des deux bornes de la classe ;
- sa longueur ou son étendue, égale à la différence entre ses valeurs extrêmes ;
- son effectif n_i.

Suivent donc, pour l'exemple 2, trois essais de groupement, basés sur le fait que les valeurs oscillent entre 200 € et 800 €.

1e essai

(*) de 0 à moins de 250 €

Classes €	x_i €	n_i Nombre de véhicules
[0;250[(*)	125	8
[250;500[	375	39
[500;750[	625	25
[750;1000[	875	3
		n = 75

Tableau 2.4

Les informations sont trop concentrées, elles présentent peu d'intérêt : la plupart des voitures se retrouvent dans deux classes contiguës. Il y a trop peu de classes.

2e essai

Classes €	x_i €	n_i Nombre de véhicules
[125;250[	187,5	8
[250;375[	312,5	16
[375;500[	437,5	23
[500;625[	562,5	17
[625;750[	687,5	8
[750;875[	812,5	3
		n = 75

Tableau 2.5

Les centres de classes ne sont pas des valeurs entières, cela ne facilite pas les calculs ultérieurs, surtout s'ils doivent être réalisés *à la main*, mentalement ou avec une calculatrice simple du type *4 opérations*.

D'autre part, il y a une amélioration sur la dispersion des résultats mais pas suffisamment. Habituellement, il est possible de constater une dispersion suffisante sans devenir excessive, lorsque le nombre de classes est compris entre 10 et 20.

3e essai

Classes €	Centres des classes x_i	Effectifs n_i
[175 ; 225[	200	5
[225 ; 275[	250	4
[275 ; 325[	300	8
[325 ; 375[	350	7
[375 ; 425[	400	7
[425 ; 475[	450	12
[475 ; 525[	500	8
[525 ; 575[	550	8
[575 ; 625[	600	5
[625 ; 675[	650	2
[675 ; 725[	700	4
[725 ; 775[	750	2
[775 ; 825[	800	3
		$n = 75$

Tableau 2.6

Les valeurs centrales sont entières et les informations bien réparties. Ce tableau à 13 *classes* est donc conservé. Il faut remarquer que le nombre de classes dépend du but poursuivi.

En général, pour construire les classes, il est pratique de procéder comme suit :

- Calculer l'étendue de la série (la plus grande moins la plus petite valeur).
- Diviser cette étendue par un nombre compris entre 10 et 20 (et donc calculer l'étendue de chaque classe).
- Choisir une valeur initiale proche du plus petit nombre de la série et construire les classes successives en tenant compte de la longueur de classe choisie.

Il est possible que certaines classes (souvent la première et/ou la dernière) puissent ne pas être de même largeur que les autres classes.

4- Différentes expressions des données et leurs représentations graphiques

4-1 Fréquence

La fréquence relative d'une valeur x_i est le rapport de l'effectif n_i à l'effectif total n

$$f_i = \frac{n_i}{n}$$

et exprime donc l'importance relative de la valeur x_i dans la série observée.

Les fréquences sont des nombres compris entre 0 et 1. Ils peuvent donc facilement être convertis en termes de pourcentage. Dans le groupement des valeurs en classes, le nombre f_i fournit le pourcentage d'éléments étudiés se trouvant dans la classe liée à la valeur centrale x_i.

Dans l'exemple 1

$$f_1 = \frac{60}{280} = 0{,}2142...$$

ce qui signifie qu'environ 21 % des véhicules recensés ont moins de 1 an.

Dans l'exemple 2

$$f_1 = \frac{5}{75} = 0{,}0666...$$

ce qui signifie qu'environ 6,7 % des véhicules de moins de 4 ans ont un coût d'entretien compris annuellement entre 175 et 225 €.

Evidemment, $\sum_{i=1}^{J} f_i = 1$ (ou $\sum_{i=1}^{J} f_i = 100\%$).

Les tableaux montrent les valeurs correspondant aux deux exemples.

Exemple 1

Valeurs x_i	Effectifs n_i	Fréquences f_i	Fréquences f_i %
0	60	0,2143	21,43
1	74	0,2643	26,43
2	72	0,2571	25,71
3	27	0,0964	9,64
4	30	0,1071	10,71
5	10	0,0357	3,57
6	5	0,0179	1,79
7	1	0,0036	0,36
8	0	0,0000	0,00
9	1	0,0036	0,36
	280	1,0000	100,00

Tableau 2.7

Exemple 2

Classes €	Centres des classes x_i	Effectifs n_i	Fréquences f_i	Fréquences f_i %
[175 ; 225[	200	5	0,0667	6,67
[225 ; 275[	250	4	0,0533	5,33
[275 ; 325[	300	8	0,1067	10,67
[325 ; 375[	350	7	0,0933	9,33
[375 ; 425[	400	7	0,0933	9,33
[425 ; 475[	450	12	0,1600	16,00
[475 ; 525[	500	8	0,1067	10,67
[525 ; 575[	550	8	0,1067	10,67
[575 ; 625[	600	5	0,0667	6,67
[625 ; 675[	650	2	0,0267	2,67
[675 ; 725[	700	4	0,0533	5,33
[725 ; 775[	750	2	0,0267	2,67
[775 ; 825[	800	3	0,0400	4,00
		75	1	100,00

Tableau 2.8

Représentation graphique des effectifs et des fréquences

a) Les effectifs (ou les fréquences), pour une série à **caractère discret** (exemple 1), se représentent graphiquement par un diagramme en ***bâtons*** : dans un système d'axes orthogonaux, à partir des points de l'axe des abscisses définis par les valeurs x_i, il faut élever les segments de droites parallèles à l'axe des ordonnées de hauteurs égales aux effectifs n_i (ou aux fréquences f_i) correspondants. La *figure* 2.1 reprend ce diagramme en bâtons, pour l'exemple 1.

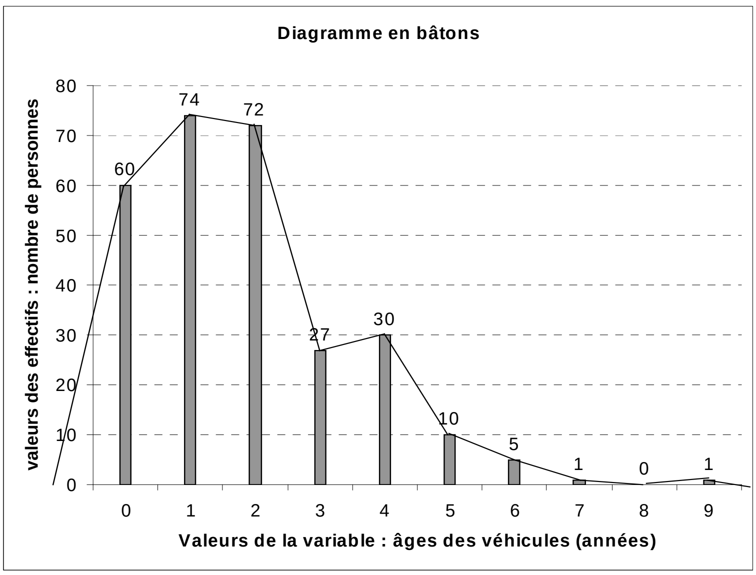

Figure 2.1

En reliant les sommets des bâtons figurant dans cette dernière représentation, une ligne brisée représentant le polygone appelé ***polygone des effectifs*** (ou des fréquences) est obtenue.

b) Dans le cas de la représentation des effectifs (ou fréquences) d'une série à **caractère continu**, il faut construire le diagramme appelé ***histogramme***, il est visible sur la *figure* 2.2, pour l'exemple 2.

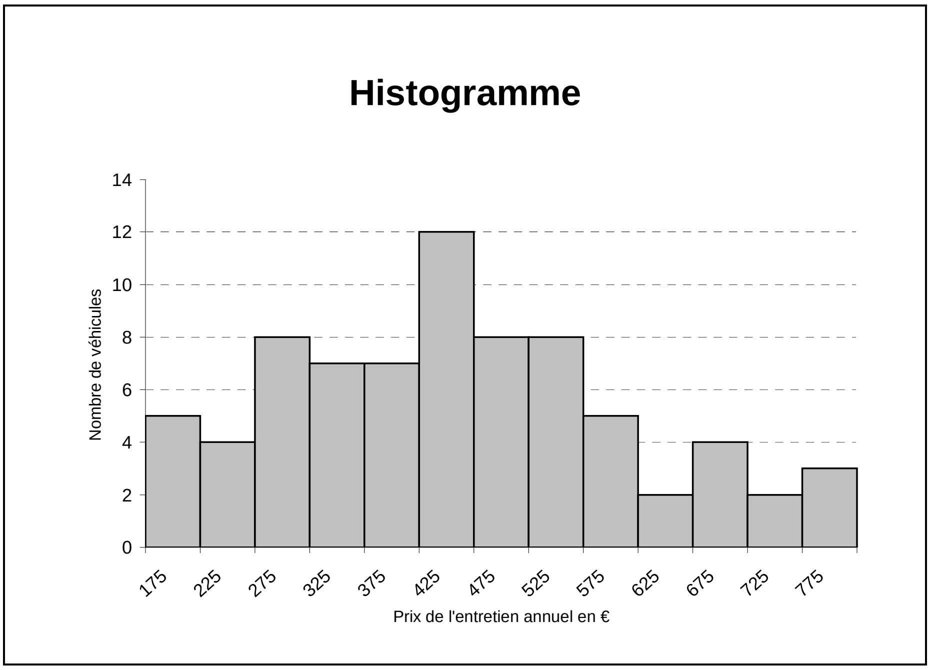

Figure 2.2

- Si les classes sont de même longueur, les bornes des intervalles de classes sont portées en abscisse et les effectifs (ou les fréquences) en ordonnée. Chaque classe est alors représentée par un rectangle ayant pour base le segment compris entre les bornes de la classe et comme hauteur l'effectif (ou la fréquence) de la classe. Les bases des rectangles ayant été supposées égales, les aires de ces rectangles sont proportionnelles aux effectifs (ou aux fréquences) des classes.
- Dans le cas où les intervalles de classes ne sont pas de même longueur, il faut ajuster les hauteurs des rectangles pour que les aires des rectangles restent proportionnelles aux effectifs des classes.

Il s'agit de choisir d'abord une classe de référence, soit l sa longueur. Si la longueur d'une autre classe est αl , la hauteur du rectangle représentatif de cette classe doit être $\frac{n_i}{\alpha}$. En effet, ainsi, l'aire du rectangle correspondant est égale à $\alpha l.\frac{n_i}{\alpha} = l.n_i$ est donc bien proportionnelle à l'effectif de la classe.

Remarques

1- Il est recommandé d'éviter de construire des représentations trop effilées ou trop aplaties de crainte d'en tirer des impressions fausses.

2- La représentation des fréquences est analogue à celle des effectifs. Il suffit, en effet, de multiplier par n la longueur de l' « unité » choisie en ordonnée pour qu'un diagramme des effectifs soit identique à un diagramme des fréquences.

Pour obtenir le **polygone des effectifs** (ou des fréquences), il faut joindre les milieux des bases supérieures des rectangles consécutifs de l'histogramme des effectifs. Ce polygone est complété par les segments joignant les points $(x_0,0)$ & (x_1,n_1) et les points (x_J,n_J) & $(x_{J+1},0)$.

4.2 Effectifs cumulés – fréquences cumulées

L'**effectif cumulé descendant** relatif à la $k^{ième}$ valeur x_k (ou à la $k^{ième}$ classe) est la somme de tous les effectifs depuis la valeur n_1 jusque la valeur n_k comprise :

$$
\begin{aligned}
nc_k &= n_1 + n_2 + ... + n_k \\
&= \sum_{i=1}^{k} n_i
\end{aligned}
$$

et nc_k représente donc le nombre d'éléments de la série dont la valeur est inférieure ou égale à x_k. Les tableaux 2.8 et 2.9 sont ceux des données complétées de la colonne des effectifs cumulés pour les exemples 1 et 2.

Exemple 1

Ainsi, dans l'exemple 1, $nc_3 = 206$, ce qui signifie que dans l'enquête, il y a 206 véhicules de $2\ ans$ ou moins de $2\ ans$.

Valeurs x_i	Effectifs n_i	Effectifs cumulés desc nc_i
0	60	60
1	74	134
2	72	206
3	27	233
4	30	263
5	10	273
6	5	278
7	1	279
8	0	279
9	1	280
	280	

Tableau 2.9

Exemple 2

Classes €	Centres des classes x_i	Effectifs n_i	Effectifs cumulés desc nc_i
[175 ; 225[	200	5	5
[225 ; 275[	250	4	9
[275 ; 325[	300	8	17
[325 ; 375[	350	7	24
[375 ; 425[	400	7	31
[425 ; 475[	450	12	43
[475 ; 525[	500	8	51
[525 ; 575[	550	8	59
[575 ; 625[	600	5	64
[625 ; 675[	650	2	66
[675 ; 725[	700	4	70
[725 ; 775[	750	2	72
[775 ; 825[	800	3	75
		75	

Tableau 2.10

Dans l'exemple 2, $nc_6 = 43$ et donc 43 *véhicules* ont été entretenus pour moins de 475 €. Il est intéressant également, d'exprimer ces valeurs cumulées en pourcentage, c'est à dire de calculer les **<u>fréquences cumulées descendantes</u>**.

$$fc_k = f_1 + f_2 + ... + f_k$$
$$= \sum_{i=1}^{k} f_i$$

ou

$$fc_k = \frac{nc_k}{n}$$

Dans l'exemple 1, $fc_3 \approx 25{,}7$ et 25,7 % des véhicules de l'enquête ont 2 *ans* ou moins de 2 *ans*.

Dans l'exemple 2, $fc_6 = 57{,}3\,\%$ et 57,3 % des véhicules de l'enquête ont coûté annuellement moins de 475 € en entretien. Les tableaux 2.10 et 2.11 sont tous deux complétés de la colonne des fréquences cumulées descendantes.

Exemple 1

Valeurs x_i	Effectifs n_i	Fréquences en % f_i en %	Fréquences cumulées fc_i
0	60	21,43	21,43
1	74	26,43	47,86
2	72	25,71	73,57
3	27	9,64	83,21
4	30	10,71	93,93
5	10	3,57	97,50
6	5	1,79	99,29
7	1	0,36	99,64
8	0	0,00	99,64
9	1	0,36	100,00
	280	100,00	

Tableau 2.11

Exemple 2

Classes €	Centres des classes x_i	Effectifs n_i	Fréquences en % f_i en %	Fréquences cumulées desc en % fc_i en %
[175 ; 225[	200	5	6,67	6,67
[225 ; 275[	250	4	5,33	12,00
[275 ; 325[	300	8	10,67	22,67
[325 ; 375[	350	7	9,33	32,00
[375 ; 425[	400	7	9,33	41,33
[425 ; 475[	450	12	16,00	57,33
[475 ; 525[	500	8	10,67	68,00
[525 ; 575[	550	8	10,67	78,67
[575 ; 625[	600	5	6,67	85,33
[625 ; 675[	650	2	2,67	88,00
[675 ; 725[	700	4	5,33	93,33
[725 ; 775[	750	2	2,67	96,00
[775 ; 825[	800	3	4,00	100,00
		75	100,00	

Tableau 2.12

Des **effectifs cumulés ascendants** et des **fréquences cumulées ascendantes** peuvent également être calculés, les sommes s'effectuant à partir de la dernière classe en remontant vers le haut du tableau. L'effectif cumulé ascendant de la $k^{ième}$ classe est noté nc_k^* et

$$\begin{aligned} nc_k^* &= n_k + n_{k+1} + \ldots + n_J \\ &= \sum_{i=k}^{J} n_i \end{aligned}$$

où J est le nombre total de classes.

La fréquence cumulée ascendante de la $k^{ième}$ classe est

$$\begin{aligned} fc_k^* &= f_k + f_{k+1} + \ldots + f_J \\ &= \sum_{i=k}^{J} f_i \end{aligned}$$

ou

$$fc_k^* = \frac{nc_k^*}{n}$$

Le tableau 2.12 reprend les valeurs ascendantes pour l'exemple 2.

Classes €	Centres des classes x_i	Effectifs n_i	Fréqu. en % f_i en %	Effectifs cumulés desc nc_i	Fréquences cumulées desc en % fc_i en %	Effectifs cumulés asc nc_i*	Fréquences cumulées asc en % fc_i* en %
[175 ; 225[	200	5	6,67	5	6,67	75	100,0000
[225 ; 275[	250	4	5,33	9	12,00	70	93,3333
[275 ; 325[	300	8	10,67	17	22,67	66	88,0000
[325 ; 375[	350	7	9,33	24	32,00	58	77,3333
[375 ; 425[	400	7	9,33	31	41,33	51	68,0000
[425 ; 475[	450	12	16,00	43	57,33	44	58,6667
[475 ; 525[	500	8	10,67	51	68,00	32	42,6667
[525 ; 575[	550	8	10,67	59	78,67	24	32,0000
[575 ; 625[	600	5	6,67	64	85,33	16	21,3333
[625 ; 675[	650	2	2,67	66	88,00	11	14,6667
[675 ; 725[	700	4	5,33	70	93,33	9	12,0000
[725 ; 775[	750	2	2,67	72	96,00	5	6,6667
[775 ; 825[	800	3	4,00	75	100,00	3	4,0000
		75	100,00				

Tableau 2.13

Il est visible que 24 *véhicules* sur 75 présentent des frais d'entretien annuels d'au moins 525 € (32 % des véhicules présentent des frais d'entretien annuel d'au moins 525 €).

Représentation graphique des effectifs cumulés (fréquences cumulées)

Dans le cas discret

Les valeurs cumulées (effectifs ou fréquences) sont portées en ordonnée pour chaque valeur de x_i portée en abscisse.

La *figure* 2.3 montre le **diagramme en escaliers** des effectifs cumulés descendants relatif à l'exemple 1.

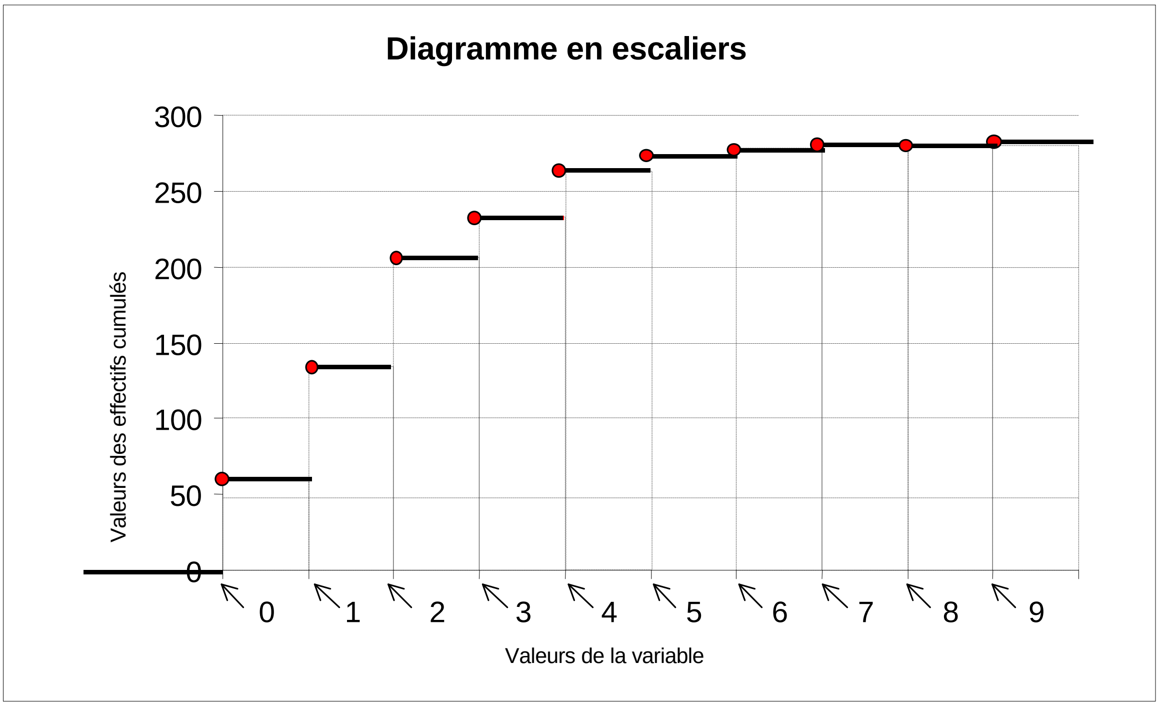

Figure 2.3

Dans le cas continu

La représentation des valeurs cumulées est un **polygone**. Les bornes des intervalles des classes sont portées sur l'axe des abscisses et les effets cumulés (fréquences cumulées) sur l'axe des ordonnées.

- Si les valeurs cumulées sont descendantes, le point initial du polygone est le point de l'axe des abscisses correspondant à la borne inférieure de la 1ère classe. Les points suivants ont pour abscisse la valeur correspondant à la borne supérieure de chaque classe et pour ordonnée l'effectif cumulé (fréquence cumulée) correspondant à cette classe.

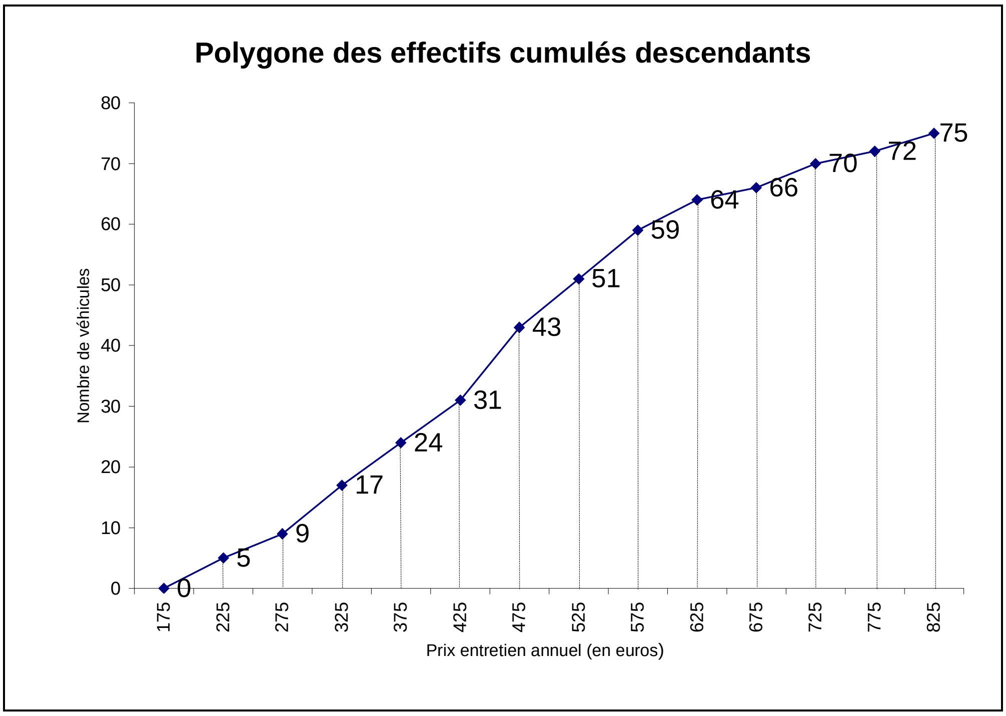

Figure 2.4

- Si les valeurs cumulées sont ascendantes, les points ont pour abscisse la valeur correspondant à la borne inférieure de chaque classe et pour ordonnée, l'effectif cumulé (fréquence cumulée) correspondant à cette classe. Le dernier point du polygone est le point de l'axe des abscisses correspondant à la borne supérieure de la dernière classe.

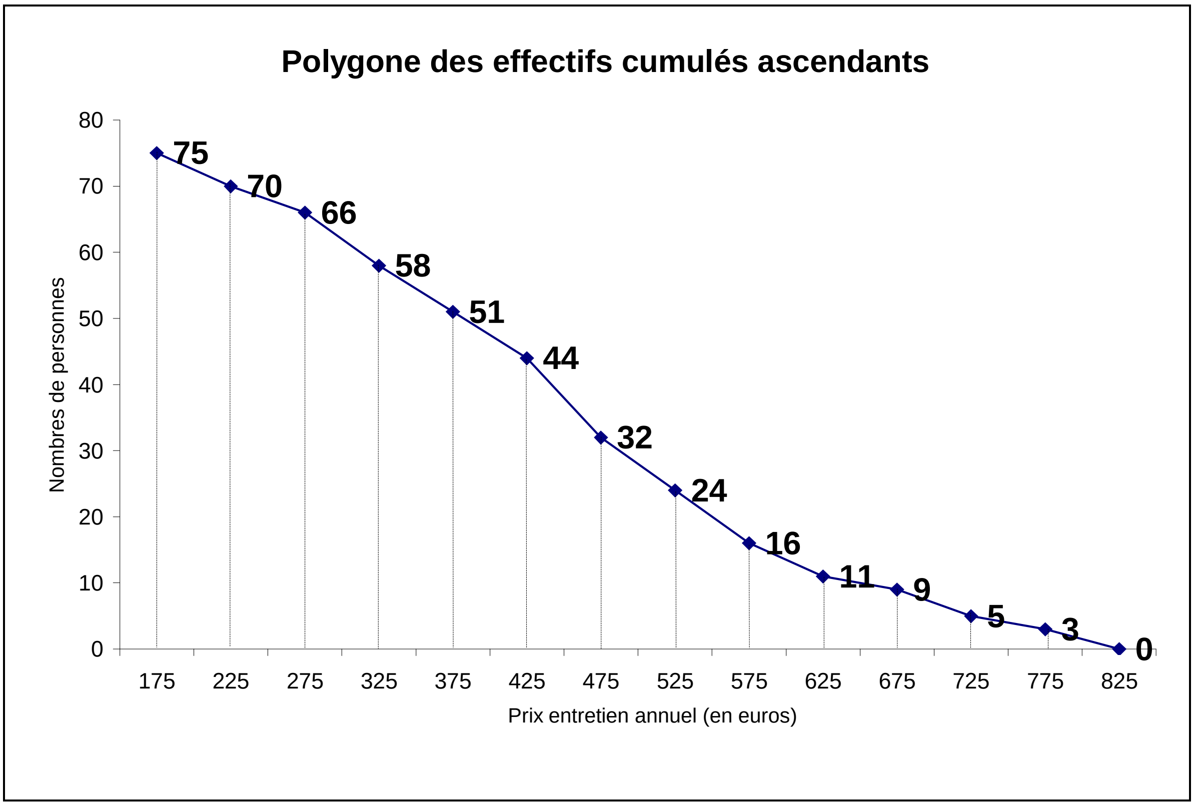

Figure 2.5