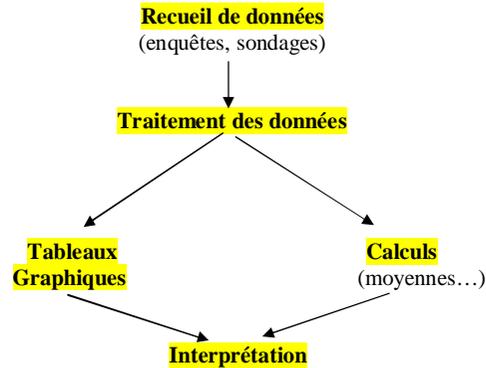
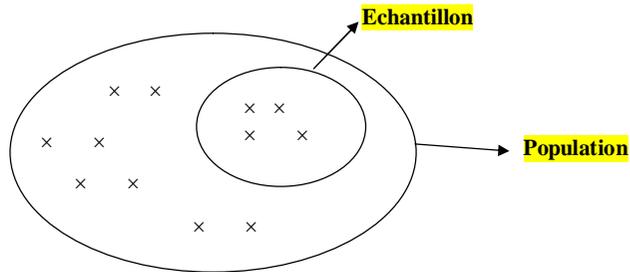


I. Présentation générale



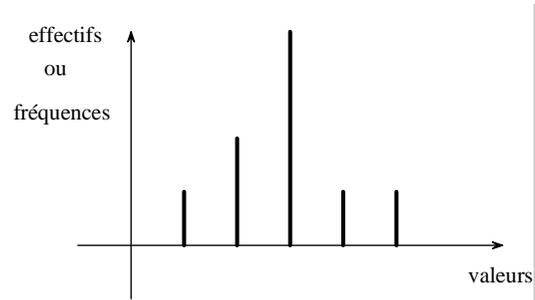
Exemple d'étude statistique : travail pour des sociétés d'assurance (durée de vie moyenne suivant les pays). Une autre chose est de déterminer les causes.

II. Vocabulaire usuel des statistiques

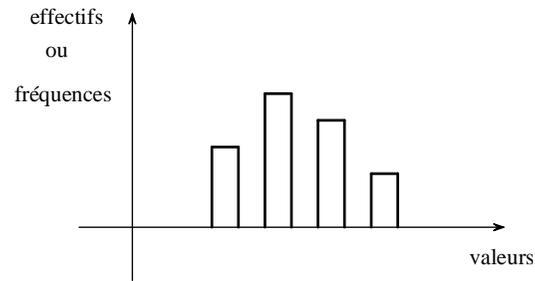
Notion	Exemples
Population : ensemble sur lequel porte l'étude.	Etude de la répartition des cylindrées pour les voitures en France Etude sur l'ensemble des voitures en France
Individus : éléments qui composent la population	Les voitures
Caractère étudié : aspect que l'on observe pour les individus	Les cylindrées
Echantillon : partie de la population Il faut prendre un échantillon représentatif sinon les résultats seront faussés.	Il faudrait prendre les voitures de toute la France.
Série statistique : ensemble des valeurs collectées	
Série qualitative : quand le caractère n'est pas mesurable par des nombres.	
Série quantitative : quand le caractère étudié fait l'objet d'une mesure, est mesurable par un nombre. On parle alors plutôt de variable quantitative .	Etude des pointures
Une variable quantitative peut être - discrète c'est-à-dire qu'elle ne peut prendre que des valeurs isolées ; - continue c'est-à-dire que le caractère étudié peut prendre toutes les valeurs dans un intervalle.	Nombre d'enfants d'une famille Température en France
Effectif total : nombre total d'individus dans la population	
Effectif d'une valeur : nombre d'individus qui présentent cette valeur. Parfois, on regroupe les valeurs en classes ; c'est le cas lorsque l'on a une variable quantitative continue.	
Fréquence d'une valeur : quotient de l'effectif de la valeur sur l'effectif total. Si une série statistique est connue par ses fréquences, on parle de distribution de fréquences .	

III. Présentation des résultats en graphiques (pour mémoire)

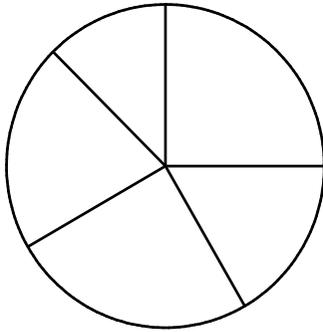
1°) Diagrammes en bâtons



2°) Diagrammes en barres



3°) Diagrammes circulaires ou semi-circulaires (« camemberts »)



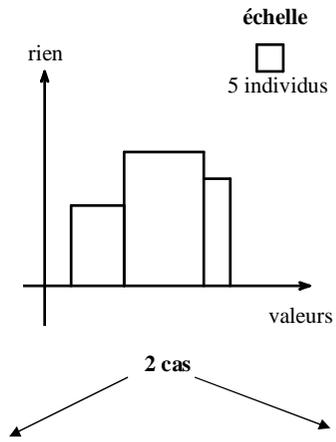
Méthode pour déterminer les angles par rapport aux effectifs ou aux fréquences :

On sait que l'effectif total ou 100 % correspond à la totalité du cercle c'est-à-dire 360 °.

4°) Histogrammes

Série statistique dont les données sont regroupées en classes.

Représentation graphique par des rectangles accolés dont **les aires sont proportionnelles aux effectifs**.



Histogramme à « pas » constant
(toutes les classes ont la même amplitude)

Histogramme à « pas » non constant

Utilisation d'un tableur très utile (permet de visualiser les résultats très rapidement).

On verra aussi les diagrammes en boîtes.

IV. Tableaux

1°) Compétences

- Savoir organiser les données dans un tableau
- Savoir lire un tableau (interpréter les chiffres)

2°) Organiser les valeurs dans un tableau

Exemple :

On considère la liste de données statistiques rangées dans l'ordre croissant :

2,3 ; 2,3 ; 2,3 ; 3,1 ; 3,1 ; 3,1 ; 3,1 ; 3,1 ; 5 ; 5 ; 7,1 ; 7,1 ; 7,1 ; 7,1 ; 7,1 ; 7,1 ; 9,2 ; 9,2 ; 10 ; 10.

Il y a 20 valeurs ; on dit que l'**effectif total** est égal à 20.

La valeur 2,3 apparaît 3 fois ; on dit que l'effectif de la valeur 2,3 est égal à 3.
La valeur 3,1 apparaît 5 fois ; on dit que l'effectif de la valeur 3,1 est égal à 5.
Etc.

Dans cette série de valeurs rangées dans l'ordre croissant, la 15^e valeur est 7,1.

On dresse un tableau statistique en mettant les valeurs sur la 1^{ère} ligne et les effectifs sur la 2^e ligne.

x_i	2,3	3,1	5	7,1	9,2	10	
n_i	3	5	2	6	2	2	N = 20

Retenir

• Les notions

- effectifs de chaque valeur
- effectif total

- Les valeurs sont toujours sur la 1^{ère} ligne ; les effectifs sont toujours sur la 2^e ligne.
- Les effectifs sont toujours des entiers naturels.

Attention à ne pas confondre les 2 notamment dans le cas où les valeurs sont aussi des entiers naturels.

3°) Effectifs cumulés croissants

Exemple :

On range les valeurs dans l'ordre croissant sur la 1^{ère} ligne.

x_i	2,3	3,1	5	7,1	9,2	10	
n_i	3	5	2	6	2	2	N = 20
Effectif cumulé croissant	3	8	10	16	18	20	

Les effectifs cumulés s'interprètent ainsi :

- Il y a 3 valeurs inférieures ou égales à 2,3.
- Il y a 8 valeurs inférieures ou égales à 3,1.
- Il y a 10 valeurs inférieures ou égales à 5.
- Il y a 16 valeurs inférieures ou égales à 7,1.
- Il y a 18 valeurs inférieures ou égales à 9,2.
- Il y a 20 valeurs inférieures ou égales à 10.

4°) Retrouver une valeur de rang donné à partir d'un tableau

Exemple :

Quelle est la 15^e valeur de la série lorsque les valeurs sont rangées dans l'ordre croissant ?

On dit que 15 est plus grand que 10 et plus petit que 16 donc la 15^e valeur est ... (on remonte à la 1^{ère} ligne).

C'est l'intérêt principal des effectifs cumulés croissants.

Retenir

Il ne faut pas oublier de remonter sur la 1^{ère} ligne.

5°) Compétences

- Savoir interpréter une valeur correspondant à un effectif.
- Savoir interpréter une valeur correspondant à un effectif cumulé croissant.
- Savoir donner une valeur de rang donné lorsque les valeurs sont rangées dans l'ordre croissant à l'aide des effectifs cumulés croissants.
- Savoir passer d'un tableau à une liste

V. Définitions des paramètres d'une série statistique (lexique)

Valeurs	x_1	x_2	...	x_p	
Effectifs	n_1	n_2	...	n_p	N

Effectif total

N.B. :

- toujours les **valeurs** sur la 1^{ère} ligne (ce sont de nombres quelconques)
- toujours les **effectifs** sur la 2^e ligne (ce sont des entiers naturels)

Fréquence en %	$\frac{\text{effectif de la valeur}}{\text{effectif total (N)}} \times 100$
Moyenne	$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{N}$ Lorsque l'on ne tient pas compte des valeurs extrêmes ou aberrantes, on parle de « moyenne élaguée ».
Médiane	Un nombre tel qu'au moins 50 % des valeurs soient inférieures ou égales à ce nombre et 50 % des valeurs soient supérieures ou égales à ce nombre.
Etendue	Valeur maximale – valeur minimale
Mode(s)	La (ou les) valeur(s) de la série de plus grand effectif Il peut y en avoir plusieurs ; on parle alors de série « unimodale » ou « multimodale »
Quartiles	• 1^{er} quartile : la plus petite valeur de la série telle qu'au moins 25 % des valeurs soient inférieures ou égales à cette valeur • 3^e quartile : la plus petite valeur de la série telle qu'au moins 75 % des valeurs soient inférieures ou égales à cette valeur
Intervalle interquartile	$[Q_1 ; Q_3]$
Ecart interquartile	$I = Q_3 - Q_1$
Variance	Formule de définition $V = \frac{n_1 (x_1 - \bar{x})^2 + n_2 (x_2 - \bar{x})^2 + \dots + n_p (x_p - \bar{x})^2}{N}$ On a donc $V \geq 0$ Formule de calcul (propriété de Kœnig-Huyghens) $V = \frac{n_1 (x_1)^2 + n_2 (x_2)^2 + \dots + n_p (x_p)^2}{N} - (\bar{x})^2$
Ecart-type	$\sigma = \sqrt{V}$ (σ : sigma minuscule)

VI. Détermination de la médiane (cas d'une série quantitative discrète)

1°) Principe

- Il faut toujours ranger les valeurs dans l'ordre croissant.
- Le principe de la médiane est de séparer en 2 groupes de même effectif.

« Il faut acquérir la logique des statistiques ».

Exemple :

x_i	5,3	7,1	7,4	
n_i	4	3	2	N = 9

5,3 ; 5,3 ; 5,3 ; 5,3 ; 7,1 ; 7,1 ; 7,1 ; 7,4 ; 7,4

La médiane est 7,1.

Donner du sens à la médiane.

Interpréter

Au moins 50 % des valeurs sont supérieures ou égales à 7,1.

Au moins 50 % des valeurs sont inférieures ou égales à 7,1.

2°) Exemple

Valeurs rangées dans l'ordre croissant

2 CAS

Effectif total pair

Exemple

N = 24

24 = 2 × 12

12 ; 12

Med = $\frac{12^{\text{e}} \text{ val} + 13^{\text{e}} \text{ val}}{2}$

Effectif total impair

Exemple

N = 23

23 = 2 × 11 + 1

11 ; 1 ; 11

Med = 12^e val

Quand le résultat ne tombe pas juste, on prend toujours l'entier immédiatement supérieur.
On regarde ensuite les valeurs correspondantes sur la 1^{ère} ligne grâce aux effectifs cumulés croissants.

4°) Bêtise à ne pas faire

Ne pas écrire $Q_1 = \frac{N}{4}$ ou $Q_3 = \frac{3N}{4}$.

5°) Remarque

Les quartiles sont toujours dans la même unité que les valeurs.

Environ la moitié des valeurs appartiennent à l'intervalle interquartile.

6°) Déciles, centiles (quantiles)

Même principe de calcul que pour les quartiles.

$D_1, D_2, D_3, \dots, D_9 \left(\frac{N}{10}, \frac{2N}{10}, \dots \right) \rightarrow$ **déciles**

On définit de même l'**intervalle interdécile** $[D_1, D_9]$ et l'**écart interdécile** $D_9 - D_1$.
Environ 80 % des valeurs appartiennent à l'intervalle interdécile.

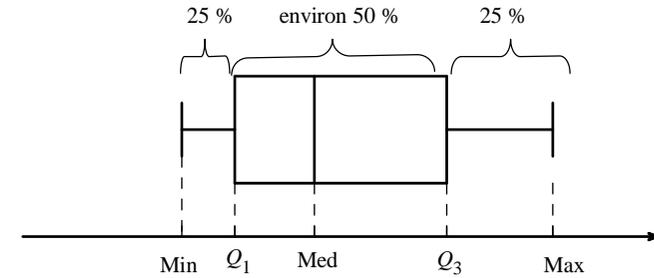
$C_1, C_2, C_3, \dots, C_{99} \left(\frac{N}{100}, \frac{2N}{100}, \dots \right) \rightarrow$ **centiles**

On définit de même l'**intervalle intercentile** $[C_1, C_{99}]$ et l'**écart intercentile** $C_{99} - C_1$.
Environ 90 % des valeurs appartiennent à l'intervalle intercentile.

VIII. Diagrammes en boîte ou à moustaches John W. Tuckey : statisticien américain (1915 – 2000)

1°) Principe

$\left. \begin{array}{l} \text{Min} \\ Q_1 \\ \text{Med} \\ Q_3 \\ \text{Max} \end{array} \right\} 5 \text{ paramètres}$



2°) Méthode



- On trace un axe d'échelle, horizontal ou vertical.
- Sur cet axe, on place les valeurs dans l'ordre croissant, à savoir : minimum, Q_1 , médiane, Q_3 , maximum.
- On trace une « boîte » rectangulaire allant de Q_1 à Q_3 séparée par une cloison au niveau de la médiane. La hauteur de la boîte est laissée au choix.
- Enfin, on trace les deux « moustaches » qui relient au maximum et au minimum. On les arrête par deux petits traits verticaux.
- Pour une meilleure lisibilité, tracer en pointillés des traits qui relient aux valeurs de l'axe d'échelle.

3°) Propriété

En général, il y a environ la moitié des valeurs dans la boîte (sauf cas exceptionnels de boîtes sans moustaches, par exemple).

4°) Intérêt

Comparaison à vue de deux (ou plus) séries statistiques en traçant les diagrammes en boîte l'un en dessous de l'autre avec un même axe d'échelle pour les deux.
(Voir exercices)

5°) Attention

- Mettre la légende sur l'axe d'échelle
- Mettre la flèche sur l'axe d'échelle
- Ne pas se tromper sur les valeurs (Q_1 et Q_3 sur la première ligne).

IX. Moyennes

Il y a deux types de moyennes :

- moyenne pondérée c'est-à-dire avec des coefficients ;
- moyenne non pondérée c'est-à-dire sans coefficients.

Pondéré vient du mot latin « **pondus** » qui signifie « **poids** ».

Lorsque dit d'une personne qu'elle a une **surcharge pondérale**, c'est qu'elle est trop grosse.

1°) Moyenne arithmétique

• Définition

La **moyenne d'une série quantitative simple**, notée \bar{x} , est la somme de toutes les valeurs du caractère, divisée par l'effectif total.

• Exemple des tailles d'un groupe de personnes.

2°) Moyenne pondérée

• Définition

La **moyenne (pondérée)** de p valeurs x_1, x_2, \dots, x_p affectées des coefficients respectifs n_1, n_2, \dots, n_p est

$$\bar{x} = \frac{1}{n} (n_1 x_1 + n_2 x_2 + \dots + n_p x_p) \text{ avec } n = n_1 + n_2 + \dots + n_p.$$

• Exemples

Exemple 1 : calcul d'une moyenne de notes avec des coefficients

15 : coefficient 3

12 : coefficient 2

9 : coefficient 3

Total des coefficient = 8

$$\text{Moyenne : } \frac{15 \times 3 + 12 \times 2 + 9 \times 3}{8} = 12$$

Exemple 2 : calcul d'une moyenne à partir d'un tableau d'effectifs

Dans une classe, on a demandé à chaque élève le nombre d'enfants vivant dans son foyer. Voici les résultats :

Nombre d'enfants	1	2	3	4	5	6	
Effectif	5	7	5	4	3	2	Total = 26

$$\text{Moyenne du nombre d'enfants par foyer} = \frac{1 \times 5 + 2 \times 7 + 3 \times 5 + 4 \times 4 + 5 \times 3 + 6 \times 2}{26} = \frac{77}{26} \approx 2,96.$$

3°) Diverses remarques

On a une moyenne pondérée lorsque :

- l'on tient compte des effectifs des valeurs ;
- l'on donne des coefficients (cas de notes affectées de certains coefficients).

Une moyenne arithmétique correspond au cas où toutes les valeurs ont le même coefficient.

La moyenne d'une série de valeurs est toujours dans la même unité que les valeurs.

Par exemple, si l'on calcule la moyenne d'une série de tailles en centimètres, le résultat sera en centimètres.

Non sens d'une expression employée fréquemment dans la présentation des prévisions météorologiques comme « normale saisonnière ». Nous sommes toujours soit au-dessus soit au-dessous des normales saisonnières. Par définition d'une moyenne, on est toujours soit au-dessus soit au-dessous. Il y a toujours une fluctuation autour de la moyenne.

4°) Moyenne élaguée, valeur aberrante

5°) Cas de valeurs regroupées en classes

Dans le cadre d'un **regroupement par classes**, la moyenne se calcule à l'aide des centres des classes, on affecte à chaque centre de classe l'effectif de la classe correspondante.

X. Variance. Ecart-type

1°) Calcul

Valeurs	x_1	x_2	...	x_p	
Effectifs	n_1	n_2	...	n_p	Total = N

$$\text{Moyenne pondérée } \bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{N}$$

En utilisant la moyenne, on peut calculer la variance V et l'écart-type σ de **deux façons**. Les deux formules donnent le même résultat.

→ par la définition

$$V = \frac{n_1 (x_1 - \bar{x})^2 + n_2 (x_2 - \bar{x})^2 + \dots + n_p (x_p - \bar{x})^2}{N}$$

→ par la formule de Kœnig-Huygens

$$V = \frac{n_1 (x_1)^2 + n_2 (x_2)^2 + \dots + n_p (x_p)^2}{N} - (\bar{x})^2$$

\bar{x} désigne la moyenne dans les deux égalités

Ecart-type

$$\sigma = \sqrt{V}$$

Commentaires

Pour calculer l'écart-type, il faut avoir calculé la moyenne.

La variance sert à calculer l'écart-type.

Pour calculer la variance, il vaut mieux utiliser la formule de Kœnig-Huygens.

Il faut connaître les 2 formules de la variance (formules littérales).

2°) Exemple

Valeurs	1	2	3	4	
Effectifs	5	7	1	8	Total N=21

• **Moyenne :** $\bar{x} = \frac{5 + 7 \times 2 + 3 + 4 \times 8}{21} \approx 2,57$

• **Variance :** $V = \frac{5 \times (1)^2 + 7 \times (2)^2 + 1 \times (3)^2 + 8 \times (4)^2}{21} - (2,57)^2$ (formule de Kœnig-Huygens)
 $V \approx 1,49$

Attention, on met la valeur au carré pas l'effectif !

• **Ecart-type :** $\sigma = \sqrt{V}$
 $\sigma = \sqrt{1,49}$
 $\sigma \approx 1,22$

3°) Cas où les valeurs sont regroupées en classes

Pour calculer l'écart-type, on prend alors, comme pour la moyenne, le **centre de chaque classe**.

4°) Unité

L'écart-type est dans la **même unité que les valeurs** (comme la moyenne et la médiane).

Par exemple, s'il s'agit de tailles en cm, la moyenne, la médiane, les quartiles et l'écart-type seront aussi en cm.

La variance est dans l'unité au carré ; mais on ne le met pas car la variance ne sert qu'à calculer l'écart-type.

5°) A quoi servent la variance et l'écart-type ?

Donner la moyenne d'un contrôle est insuffisant. On devrait donner en plus pour avoir une bonne idée :

- les notes extrêmes (l'étendue)
- l'écart-type

Exemple :

C'est souvent dans les matières scientifiques (maths-physique) que l'écart-type des notes est le plus grand contrairement aux matières littéraires où l'on a souvent un faible écart-type

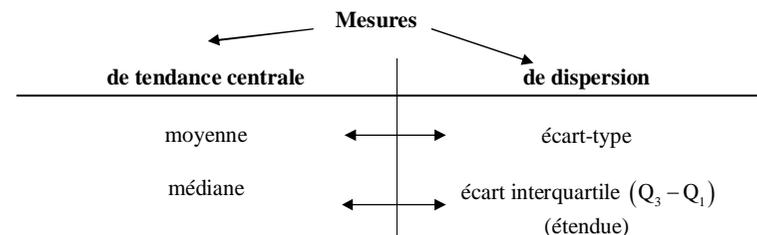
En effet, souvent les notes en français sont concentrées autour de moyenne, entre 8 et 12 disons. Il est assez difficile d'avoir au-dessus de 14 ; il est également rare d'avoir des notes très faibles.

Alors qu'en maths, on a souvent une très grande amplitude de notes.

XI. Comparaisons des différents paramètres

1°) Couples de paramètres

On distingue 2 types de mesures (paramètres) :



2°) Interprétation

- Plus l'écart-type est faible, plus les valeurs sont concentrées autour de la moyenne.
- Plus l'écart interquartile est faible, plus les valeurs sont concentrées autour de la médiane.

(Un écart-type nul signifie que toutes les valeurs sont égales).

3°) Application : comparaison de deux séries statistiques

Méthode

On compare les mesures de tendance centrale puis les mesures de dispersion associées.

(On a besoin des 2 pour comparer).

La comparaison peut s'effectuer « graphiquement » grâce aux diagrammes en boîte (on compare alors les médianes et les écarts-types).

Voir exercices.

4°) Avantages-inconvénients

Le couple (médiane ; écart interquartile) est moins sensible que le couple (moyenne ; écart-type) aux variations des valeurs extrêmes.

XII. Utilisation de la calculatrice

1°) Rentrer une liste

TI 82-83	Casio
STAT Edit	Menu 2 STAT

L1
List 1

L2
List 2

L3
List 3

On saisit les valeurs dans la 1^{ère} colonne, les effectifs dans la 2^e.

2°) Calcul des paramètres

STAT CALC

\bar{x} = ... **moyenne**

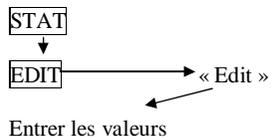
$\sum x$ = ... on s'en fiche

$\sum x^2$ = ... on s'en fiche

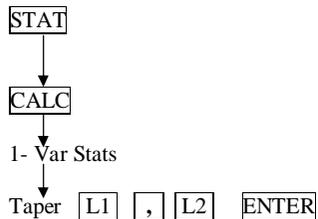
S_x = ...

σ_x = ... **écart-type**

Attention, les conventions de calcul de la calculatrice ne sont pas les mêmes que celles données en cours pour la médiane et les quartiles.



Puis



XIII. Calculs de la moyenne d'une série à l'aide des moyennes de deux sous-groupes

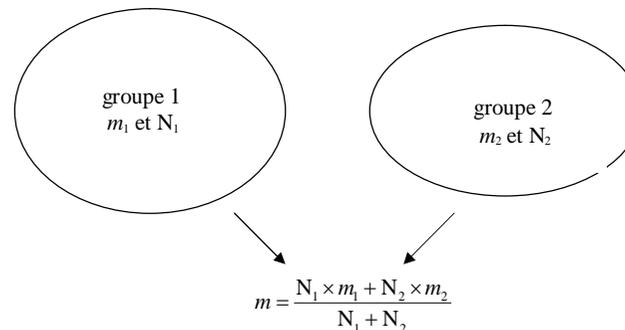
1°) Propriété

On considère une série statistique S_1 de moyenne m_1 et d'effectif total N_1 .

On considère une série statistique S_2 de moyenne m_2 et d'effectif total N_2 .

Alors la moyenne de la série statistique obtenue en réunissant les séries S_1 et S_2 est égale à

$$m = \frac{N_1 \times m_1 + N_2 \times m_2}{N_1 + N_2}$$

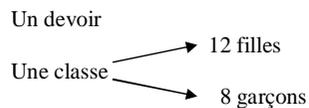


2°) Démonstration

Facile.

Voir cours de 2^e.

3°) Exemple



Moyenne des filles = 14

Moyenne des garçons = 16

$$\text{Moyenne de la classe} = \frac{4 \times 12 + 8 \times 16}{20} = 14,8$$

4°) Remarque

Il n'existe pas de formule analogue pour les autres indicateurs tels que la variance, l'écart-type, la médiane etc.

XIV. Formules de calculs de quelques indicateurs à l'aide des fréquences

1°) Moyenne

Valeurs	x_1	x_2	...	x_p	
Effectifs	n_1	n_2	...	n_p	Effectif total=N
Fréquences	$f_1 = \frac{n_1}{N}$	$f_2 = \frac{n_2}{N}$		$f_p = \frac{n_p}{N}$	1

On a : $\bar{x} = f_1 \times x_1 + f_2 \times x_2 + \dots + f_p \times x_p$.

(La démonstration est très facile).

2°) Variance

Avec les mêmes notations qu'au 1°), on obtient les deux formules de la variance suivantes à partir de la définition et de la formule de Kœnig-Huygens

$$V = f_1 \times (x_1 - \bar{x})^2 + f_2 \times (x_2 - \bar{x})^2 + \dots + f_p \times (x_p - \bar{x})^2$$

$$V = [f_1 \times (x_1)^2 + f_2 \times (x_2)^2 + \dots + f_p \times (x_p)^2] - (\bar{x})^2$$

XV. Détermination des indicateurs dans le cas d'une série de valeurs regroupées en classes

1°) Calcul de la moyenne et de la variance

On prend pour valeurs le centre de chaque classe.

2°) Détermination de la médiane et des quartiles

Méthode

On utilise une méthode graphique.

On utilise le **polygone des effectifs cumulés croissants** ou des fréquences cumulées croissantes. Pour cela, on calcule d'abord les effectifs cumulés croissants (tableau) puis on trace le polygone des effectifs cumulés croissants : il s'agit d'un ligne brisée ouverte constituée de **segments de droites**.

- **Pour lire la médiane**, on place la valeur $\frac{N}{2}$ sur l'axe des ordonnées ; on lit sur l'abscisses du point du polygone qui a pour ordonnée $\frac{N}{2}$.
- **Pour lire le 1^{er} quartile**, on place la valeur $\frac{N}{4}$ sur l'axe des ordonnées ; on lit sur l'abscisses du point du polygone qui a pour ordonnée $\frac{N}{4}$.

- **Pour lire le 3^e quartile**, on place la valeur $\frac{3N}{4}$ sur l'axe des ordonnées ; on lit sur l'abscisses du point du polygone qui a pour ordonnée $\frac{3N}{4}$.

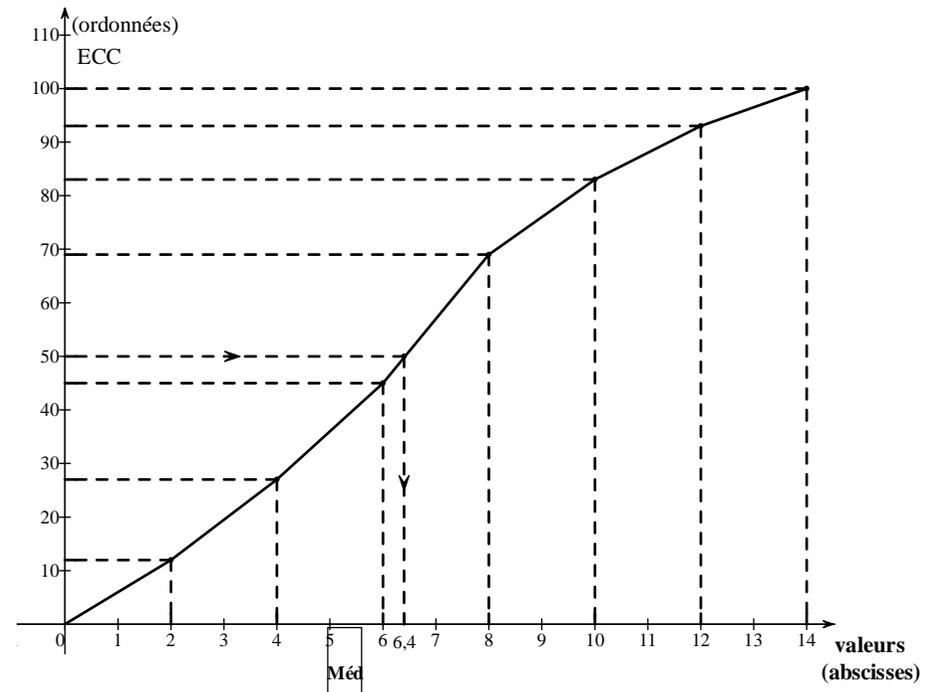
Cette méthode s'adapte évidemment sans problème aux quantiles (déciles, centiles ...)

Exemple :

Le tableau ci-dessous donne la distribution des effectifs d'une série.

Classe	[0 ; 2[[2 ; 4[[4 ; 6[[6 ; 8[[8 ; 10[[10 ; 12[[12 ; 14[
Effectif	12	15	18	24	14	10	7
Effectif cumulé croissant	12	27	45	69	83	93	100
Borne supérieure	2	4	6	8	10	12	14

On détermine la médiane et les quartiles graphiquement grâce au polygone des effectifs cumulés croissants.



Détermination graphique d'une valeur approchée de la médiane.

On utilise le polygone des effectifs cumulés croissants.

L'effectif total est égal à 100.

On le divise par 2 : $\frac{100}{2} = 50$.

On se place à 50 d'ordonnée.

On va jusqu'au polygone.

On lit la valeur en abscisse.

On lit graphiquement $Med \approx 6,5$.

Détermination graphique d'une valeur approchée du 1^{er} quartile.

On utilise le polygone des effectifs cumulés croissants.

L'effectif total est égal à 100.

On le divise par 4 : $\frac{100}{4} = 25$.

On se place à 25 d'ordonnée.

On va jusqu'au polygone.

On lit la valeur en abscisse.

On lit graphiquement $Q_1 \approx 3,5$.

Même méthode pour le 3^e quartile.

On lit graphiquement : $Q_3 \approx 8,9$

N.B. : on peut aussi calculer la médiane et les quartiles par interpolation linéaire (méthode d'interpolation linéaire qui sera vue en enseignement obligatoire).

XVI. Tableaux croisés

1°) Tableau croisé d'effectifs

La répartition des élèves d'un lycée est donnée dans le tableau suivant.

Niveau \ Sexe	2 ^e	1 ^{ère}	T ^{ale}	Total
Garçons	179	152	184	515
Filles	161	113	131	405
Total	340	265	315	920

« Marges par niveaux »

Effectifs marginaux des niveaux

2°) Tableau des pourcentages par rapport à l'effectif total

Niveau \ Sexe	2 ^e	1 ^{ère}	Tale	Total
Garçons	19,5 % *	16,5 %	20 %	56 %
Filles	17,5 %	12,3 %	14,2 %	44 %
Total	37 %	28,8 %	34,2 %	100 %

Fréquences marginales

$$* \frac{179}{920} \times 100$$

3°) Tableaux des pourcentages par lignes

Niveau \ Sexe	2 ^e	1 ^{ère}	T ^{ale}	Total
Garçons	34,8 % *	29,5 %	35,7 %	100 %
Filles	39,8 %	27,9 %	32,3 %	100%

$$* \frac{179}{515} \times 100 = 34,8$$

34,8 % des garçons du lycée sont en 2^e.

4°) Tableaux des pourcentages par colonnes

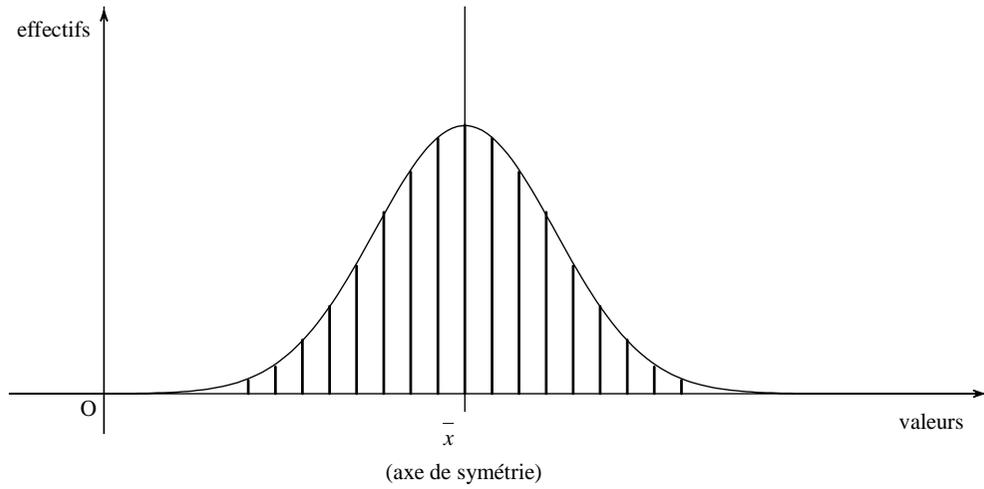
Niveau \ Sexe	2 ^e	1 ^{ère}	T ^{ale}
Garçons	52,6 %	57,4 %	58,4 %
Filles	47,4 %	42,6 %	41,6 %
Total	100 %	100 %	100 %

XVII. Données gaussiennes

Gauss : mathématicien allemand (1777-1855)

1°) Définition

Dans de nombreux cas, on observe un diagramme en bâtons ayant l'allure suivante :



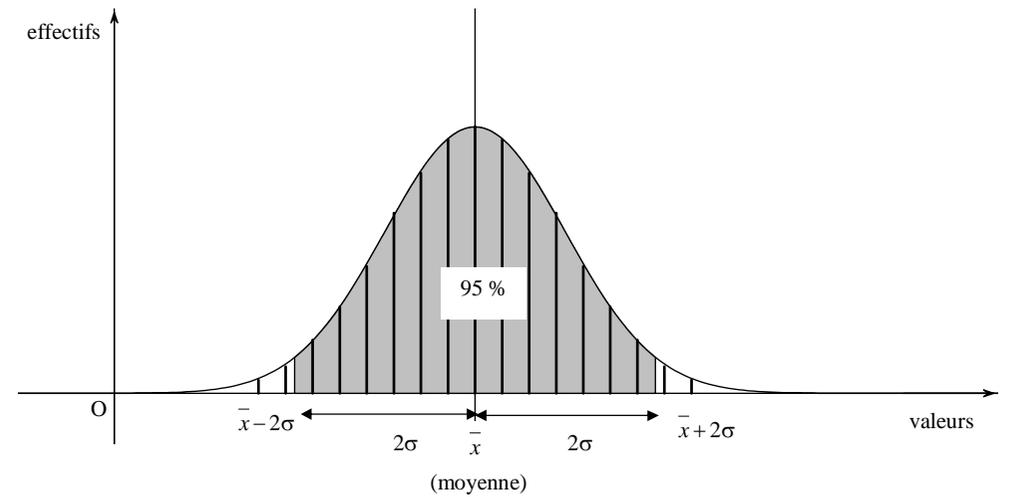
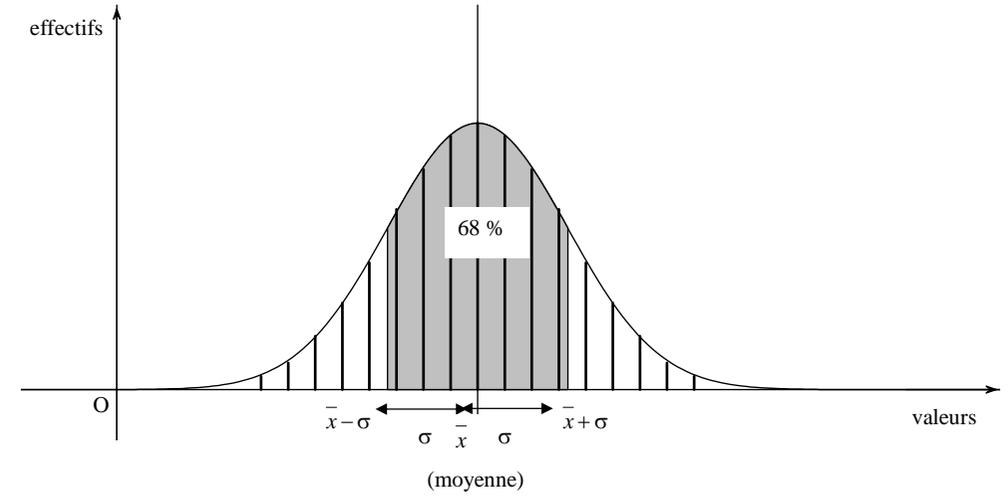
Courbe en cloche ou courbe de Gauss.

Par exemple, taille des élèves dans un groupe.

Sur le graphique, on voit que le mode est confondu avec la médiane et avec la moyenne (distribution symétrique par rapport à la moyenne).

Dans ce cas, on dit qu'il s'agit de **données gaussiennes** (on parle aussi de **répartition normale**).

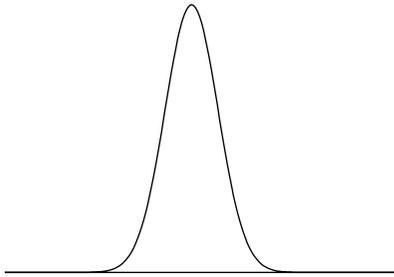
2°) Propriété



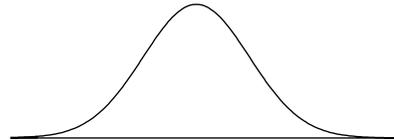
- Pour des données gaussiennes, environ 68 % des valeurs sont situées dans l'intervalle $[\bar{x}-\sigma; \bar{x}+\sigma]$.

On dit que l'intervalle $[\bar{x}-\sigma; \bar{x}+\sigma]$ est la **plage de normalité** au niveau de confiance 0,68.

- $[\bar{x}-2\sigma; \bar{x}+2\sigma]$ correspond à 95 %.
- $[\bar{x}-3\sigma; \bar{x}+3\sigma]$ correspond à 99 %.



σ petit



σ grand

3°) Complément sur les plages de normalité

- **Rappel sur centre et longueur d'un intervalle $[a; b]$ ($a \leq b$)**

Centre de cet intervalle : $c = \frac{a+b}{2}$

Longueur de cet intervalle : $b-a$ (différence entre la borne supérieure et la borne inférieure)

- **Exemple :** $[1; 7]$

Centre : 4

Longueur : 6

- **Applications aux plages de normalité**

Intervalle $[\bar{x}-\sigma; \bar{x}+\sigma]$

Centre de cet intervalle : $\frac{\bar{x}-\cancel{\sigma} + \bar{x} + \cancel{\sigma}}{2} = \frac{2\bar{x}}{2} = \bar{x}$

Longueur de cet intervalle : $(\bar{x}+\sigma) - (\bar{x}-\sigma) = \cancel{\bar{x}} + \sigma - \cancel{\bar{x}} + \sigma = 2\sigma$ (règle de 4° pour enlever les parenthèses)

Intervalle $[\bar{x}-2\sigma; \bar{x}+2\sigma]$

Centre de cet intervalle : $\frac{\bar{x}-\cancel{2\sigma} + \bar{x} + \cancel{2\sigma}}{2} = \frac{2\bar{x}}{2} = \bar{x}$

Longueur de cet intervalle : $(\bar{x}+2\sigma) - (\bar{x}-2\sigma) = 4\sigma$

Toutes les plages de normalité ont pour centre \bar{x} .
Leur longueur dépend uniquement de σ ($2\sigma, 4\sigma, 6\sigma$).

4°) Une application dans la vie courante

Dans la vie courante, la distribution normale est utilisée dans de nombreux domaines.

C'est le cas par exemple en médecine pour fixer la posologie d'un médicament.

On croit toujours que plus on prend d'un médicament plus celui-ci aura d'efficacité.

Or c'est faux, la posologie d'un médicament est fixée pour qu'il y ait le maximum d'efficacité.

Avant de lancer un médicament, on procède à des essais sur des groupes de personnes volontaires de même poids (cobayes).

On étudie ensuite l'efficacité en fonction de la posologie.

On observe une distribution en forme de courbe en cloche

On constate un pic d'efficacité pour une certaine posologie.

En dessous ou au-dessus de cette posologie, l'efficacité n'est pas totale.

On prescrira donc le médicament pour un poids donné avec cette posologie qui apparaît sur le graphique.

COMPETENCES

- Savoir organiser les données dans un tableau
- Savoir lire un tableau (interpréter les chiffres).

Peut-être faire un paragraphe spécial :

- 1°)
- listes
 - effectifs de chaque valeur, effectif total

Les valeurs sont toujours sur la 1^{ère} ligne ; les effectifs sont toujours sur la 2^e ligne.
Les effectifs sont toujours des entiers naturels.

Attention à ne pas confondre les 2 notamment dans le cas où les valeurs sont aussi des entiers naturels.

- 2°)
- rangement des valeurs dans l'ordre croissant
 - effectifs cumulés croissants, décroissants

Rangement des valeurs dans l'ordre croissants ou décroissant.

- Savoir interpréter une valeur correspondant à un effectif.
- Savoir interpréter une valeur correspondant à un effectif cumulé croissant ou décroissant.

Savoir donner une valeur de rang donné lorsque les valeurs son rangées dans l'ordre croissant à l'aide des effectifs cumulés.

Exemple :

Quelle est la 15^{ème} valeur de la série lorsque les valeurs sont rangées dans l'ordre croissant ?
On dit que 15 est plus grand que 11 et plus petit que 19 donc la 15^e valeur est ... (on remonte à la 1^{ère} ligne).

Passage

tableau ← → liste

dans les 2 sens

1^{ère} L : comment j'ai fait en 1^{ère} L année 2009 – 2010 2 octobre 2009

1^{ère} séance : 2 octobre 2009

Demander aux élèves le nombre de frères et sœurs en incluant les demi-frères et les demi-sœurs (même ceux qui ni vivent plus avec vous) et en se comptant soi-même.

Un élève au tableau note les résultats (c'est la phase de recueil des données).

- Demander aux élèves
- de trier les donner (les ranger dans l'ordre croissant)
 - d'organiser les données en tableaux (tableaux d'effectifs, fréquences, effectifs cumulés croissants) ;
 - de faire des graphiques (diagrammes en bâtons ou en barres).
 - de calculer moyenne, étendue, médiane
 - de déterminer le ou les mode(s) de la série

Les renseignements se trouvent dans le cours.

2^e séance : 6 octobre 2009

Présentation des résultats :

- arrondi pour les fréquences : la somme des fréquences est égale à 0,98.

Dû à quoi ?

- moyenne : notation \bar{x}
 $\bar{x} = 3,7$

On dira que la moyenne est environ égale à 3,7 enfants (même si le résultat n'a pas de signification réelle).

- diagramme en bâtons (il y a un « pic »)
- mode ; valeur modale

Anne-Claire Leconte médiane d'une série avec effectif total pair. On fait « moitié-moitié ».
Un commentaire à moi : on fait « fifty-fifty ».

Travail pour la prochaine séance :

- Etudier le cours sur les quartiles ;
- Pour la série du nombre de frères et sœurs, calculer le 1^{er} et le 3^e quartile et l'écart interquartile.

3^e séance : 9 octobre 2009

Quartile :

Convention de calcul ;

Qu'est-ce que le 2^e quartile ? C'est la médiane ; on l'a déjà calculé.

A quoi servent les quartiles ? à se positionner.

Dans une classe avec des notes, il est important de savoir si vous êtes dans le 1^{er} quart, dans le 2^e quart, dans le 3^e quart ou dans le 4^e quart.

De même pour les déciles.

A quoi sert l'écart interquartile ?

Comme l'étendue mais plus précis.

- Pour la série du nombre de frères

Commentaires

Avant de faire le paragraphe sur la variance et l'écart-type, faire un rappel sur moyenne arithmétique et sur moyenne pondérée.

Moyenne pondérée

Moyenne non pondérée

Pondéré : vient du latin pondus.

1°) Moyenne arithmétique :

Exemple des tailles d'un groupe de personnes.

2°) Moyenne pondérée

Avec des coefficients

Pondéré vient du mot latin « **pondus** » qui signifie « **poids** ».

Lorsque dit d'une personne qu'elle a une **surcharge pondérale**, c'est qu'elle est trop grosse.

Mettre deux exemples

Exemple 1 : calcul d'une moyenne de notes avec des coefficients

Exemple 2 : calcul d'une moyenne à partir d'un tableau d'effectifs

On a une moyenne pondérée lorsque :

- l'on tient compte des effectifs des valeurs
- l'on donne des coefficients (cas de notes affectées de certains coefficients).

Une moyenne arithmétique correspond au cas où toutes les valeurs ont le même coefficient.

Moyenne élaguée, valeur aberrante.

La **moyenne d'une série quantitative simple**, notée \bar{x} , est la somme de toutes les valeurs du caractère, divisée par l'effectif total.

La **moyenne (pondérée)** de p valeurs x_1, x_2, \dots, x_p affectées des coefficients respectifs $n_1, n_2, \dots,$

n_p est
$$\bar{x} = \frac{1}{n} (n_1 x_1 + n_2 x_2 + \dots + n_p x_p)$$
 avec $n = n_1 + n_2 + \dots + n_p$

Remarque : dans le cadre d'un **regroupement par classes**, la moyenne se calcule à l'aide des centres des classes, on affecte à chaque centre de classe l'effectif de la classe correspondante.

Commentaires

Comment acquérir la logique des statistiques ?

Pour la médiane, utiliser un tableau d'effectifs cumulés croissants.

Demander lorsque les valeurs sont rangées dans l'ordre croissant

- la valeur qui est en 6^e position
- la valeur qui est en 10^e position
- etc.

Pour la médiane, ça doit être à égalité de chaque côté, entre les deux parties. C'est le centre de toute la série.

Il faut toujours revenir à la ligne du dessus.

Faire un exemple sur le nombre de frères et sœurs en comptant les demi-frères et les demi-sœurs. Les élèves comprennent alors mieux tout ce qui est médiane, quartiles etc.

A propos de la variance

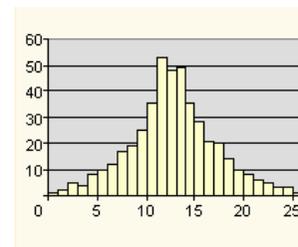
On prend la manière Koenig :

Attention, on met la valeur au carré (pas l'effectif au carré)

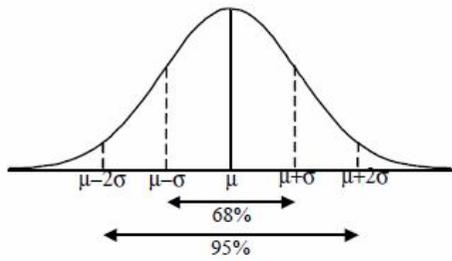
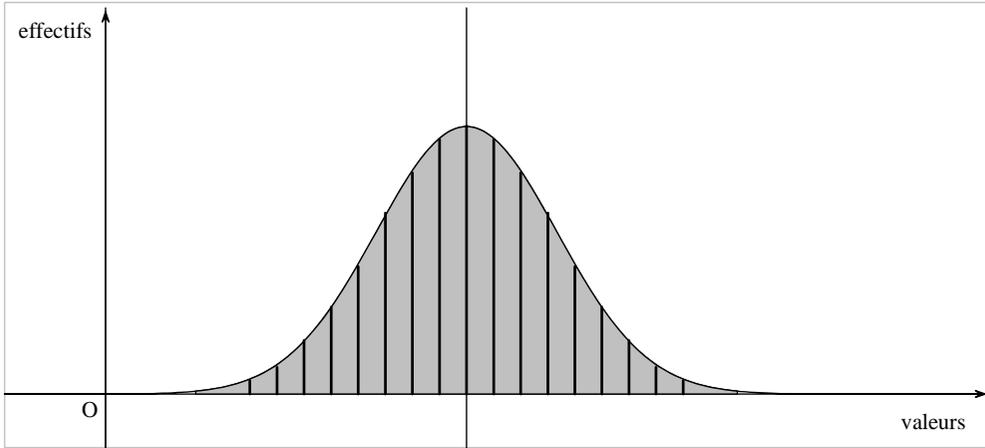
Exemple

1	2	
7	10	

$$1^2 \times 7 + 2^2 \times 10 + \dots$$



Compétence importante : interpréter les valeurs des déciles



Pour les données gaussiennes :

```

Vecteur,u,0,0,1;0;,,1,50,3,0,4,75;0
Vecteur,v,0,0,0;1;,,1,50,3,0,-4;-0,43
Segment,1,[,],-1,5;0;-1,5;0,13;,,,1,0,0,0,30
Segment,1,[,],-1,2;0;-1,2;0,19;,,,1,0,0,0,30
Segment,1,[,],-0,9;0,27;-0,9;0;,,,1,0,0,0,30
Segment,1,[,],-0,6;0;-0,6;0,33;,,,1,0,0,0,30
Segment,1,[,],-0,3;0,38;-0,3;0;,,,1,0,0,0,30
Segment,1,[,],0;0,4;0;0;,,,1,0,0,0,30
Segment,1,[,],0,3;0,38;0,3;0;,,,1,0,0,0,30
Segment,1,[,],0,6;0,33;0,6;0;,,,1,0,0,0,30
Segment,1,[,],0,9;0,27;0,9;0;,,,1,0,0,0,30
Segment,1,[,],1,2;0;1,2;0,19;,,,1,0,0,0,30
Segment,1,[,],1,5;0,13;1,5;0;,,,1,0,0,0,30
Segment,1,[,],-1,8;0,08;-1,8;0;,,,1,0,0,0,30
Segment,1,[,],-2,1;0,04;-2,1;0;,,,1,0,0,0,30
Segment,1,[,],-2,4;0,02;-2,4;0;,,,1,0,0,0,30
Segment,1,[,],1,8;0;1,8;0,08;,,,1,0,0,0,30
Segment,1,[,],2,1;0;2,1;0,04;,,,1,0,0,0,30
Segment,1,[,],2,4;0,02;2,4;0;,,,1,0,0,0,30

```