

Effectifs cumulés croissants

Ce document présente la notion d'effectif cumulé croissant à travers un exemple abordé sous deux points de vue : à partir des données brutes, puis à partir de ces mêmes données groupées en classes. Nous identifierons les difficultés liées à la seconde présentation, la seule utilisée dans nos classes. Nous nous attacherons à donner du sens à la notion d'effectif cumulé croissant et nous précisons le rôle des bornes des classes dans la construction du polygone des effectifs cumulés croissants.

Données brutes

Le tableau suivant donne les observations C_i ($1 \leq i \leq 500$) d'un caractère quantitatif C sur 500 individus statistiques.

117,3	121,7	104,1	109,2	120,4	111,0	123,6	115,3	120,2	115,9	116,9	113,1	119,9	122,1	115,1	109,5	114,5	115,0
116,2	112,3	115,3	110,1	117,6	116,9	114,5	112,0	112,1	115,0	110,9	107,7	123,3	111,2	123,2	115,7	117,2	113,2
114,9	120,9	106,0	113,7	123,3	114,0	109,3	109,5	117,4	123,3	111,6	124,3	114,2	117,4	110,8	118,5	124,0	112,7
118,5	119,9	111,6	118,6	117,5	112,0	112,8	115,4	117,5	110,6	114,8	119,1	112,8	112,5	118,0	115,8	120,3	117,0
118,9	120,2	120,3	114,4	119,0	120,3	108,1	110,7	110,4	106,2	112,3	111,2	128,5	116,9	111,5	121,3	114,5	119,6
112,2	113,4	119,9	111,5	114,7	119,5	118,8	120,9	112,6	110,1	116,1	113,0	122,7	108,9	111,2	121,1	107,5	119,0
123,2	112,6	113,4	108,2	113,3	114,4	111,0	120,6	117,5	111,6	114,9	121,9	110,7	123,8	120,8	127,0	118,5	121,2
118,9	112,4	104,8	116,1	118,3	117,3	115,2	112,2	121,0	117,7	127,4	112,0	107,8	108,8	116,0	115,0	121,4	112,2
117,1	118,7	118,9	115,5	113,5	117,3	117,7	122,4	111,0	117,4	115,2	116,0	113,7	116,3	117,4	115,7	116,4	106,1
123,5	118,9	119,3	118,4	116,5	113,3	120,6	123,0	120,0	114,9	124,3	109,4	117,3	105,2	112,8	124,9	123,8	124,4
114,2	119,9	118,4	113,4	116,9	115,4	121,4	109,4	122,6	114,3	109,2	116,8	118,7	113,1	120,7	119,7	109,4	115,8
112,2	118,8	122,2	116,7	115,9	123,3	117,5	112,1	120,3	111,6	113,9	105,9	112,3	105,0	117,5	114,0	106,3	115,8
110,0	118,0	115,9	109,2	117,6	110,1	111,5	113,9	110,5	122,2	109,9	117,8	118,1	116,6	111,3	119,4	109,5	116,0
118,1	120,9	116,6	112,6	113,0	120,0	108,5	119,3	113,6	109,9	112,7	116,1	115,5	112,7	107,3	111,5	117,8	114,6
122,6	112,0	118,5	107,9	114,5	113,5	121,6	105,7	124,1	115,3	121,2	109,7	122,2	111,5	114,6	114,5	109,5	114,2
114,9	123,1	112,1	116,8	121,3	121,2	99,5	119,4	122,1	109,4	111,8	125,2	113,2	119,4	115,4	116,1	120,2	120,1
107,3	125,1	118,1	114,7	118,9	113,8	123,4	111,9	102,5	118,6	116,8	113,7	129,0	127,0	116,5	118,5	124,7	113,6
118,2	121,3	108,5	123,6	119,1	120,7	121,9	116,0	110,9	112,0	112,7	114,4	119,4	123,5	110,6	117,6	121,6	110,6
122,2	108,8	119,8	111,9	105,4	111,4	111,0	112,4	119,5	114,2	112,9	115,2	118,9	113,4	110,3	114,4	124,1	115,3
116,1	114,5	123,3	112,6	113,2	112,4	114,1	120,5	117,6	112,5	110,4	112,7	110,5	118,5	108,7	108,8	122,0	123,7
114,6	110,2	101,6	116,1	115,0	114,8	123,2	113,1	122,4	120,6	122,5	112,7	113,1	109,4	112,2	122,5	112,7	116,5
111,3	117,5	117,4	111,0	118,0	112,9	115,4	111,9	111,1	113,7	121,8	107,9	104,8	115,7	112,5	121,7	123,7	107,2
111,1	113,8	117,0	119,5	107,0	124,9	116,7	112,4	114,8	123,8	115,1	115,2	119,3	111,4	124,0	116,0	124,4	114,9
123,2	110,1	121,5	114,8	112,3	114,6	116,4	120,7	113,7	120,6	120,4	123,8	107,9	111,4	116,6	119,9	122,9	120,1
118,7	122,0	111,5	113,7	115,7	126,1	111,9	111,5	124,2	112,3	118,0	122,5	112,5	113,8	112,8	116,9	121,5	112,4
114,1	118,8	122,5	120,7	111,2	115,5	111,2	110,4	126,1	121,6	115,6	117,3	116,4	119,6	116,1	107,1	115,4	123,4
122,0	113,1	111,4	125,6	114,7	116,3	124,2	117,3	114,5	122,3	103,3	116,6	114,8	116,9	117,7	124,8	116,4	114,9
116,8	112,5	118,0	122,7	117,1	108,6	109,9	118,3	116,0	117,9	115,9	109,6	113,5	120,5				

Si on ordonne ces valeurs par ordre croissant, on obtient le tableau suivant :

99,5	101,6	102,5	103,3	104,1	104,8	104,8	105,0	105,2	105,4	105,7	105,9	106,0	106,1	106,2	106,3	107,0	107,1
107,2	107,3	107,3	107,5	107,7	107,8	107,9	107,9	107,9	108,1	108,2	108,5	108,5	108,6	108,7	108,8	108,8	108,8
108,9	109,2	109,2	109,2	109,3	109,4	109,4	109,4	109,4	109,4	109,5	109,5	109,5	109,5	109,6	109,7	109,9	109,9
109,9	110,0	110,1	110,1	110,1	110,1	110,2	110,3	110,4	110,4	110,4	110,5	110,5	110,6	110,6	110,6	110,7	110,7
110,8	110,9	110,9	111,0	111,0	111,0	111,0	111,0	111,1	111,1	111,2	111,2	111,2	111,2	111,2	111,3	111,3	111,4
111,4	111,4	111,4	111,5	111,5	111,5	111,5	111,5	111,5	111,5	111,6	111,6	111,6	111,6	111,8	111,9	111,9	111,9
111,9	112,0	112,0	112,0	112,0	112,0	112,1	112,1	112,1	112,2	112,2	112,2	112,2	112,2	112,3	112,3	112,3	112,3
112,3	112,4	112,4	112,4	112,4	112,4	112,5	112,5	112,5	112,5	112,5	112,6	112,6	112,6	112,6	112,7	112,7	112,7
112,7	112,7	112,7	112,7	112,8	112,8	112,8	112,8	112,9	112,9	113,0	113,0	113,1	113,1	113,1	113,1	113,2	113,2
113,2	113,2	113,3	113,3	113,4	113,4	113,4	113,4	113,5	113,5	113,5	113,6	113,6	113,7	113,7	113,7	113,7	113,7
113,7	113,8	113,8	113,8	113,9	113,9	114,0	114,0	114,1	114,1	114,2	114,2	114,2	114,3	114,4	114,4	114,4	114,4
114,4	114,5	114,5	114,5	114,5	114,5	114,5	114,5	114,6	114,6	114,6	114,6	114,7	114,7	114,7	114,8	114,8	114,8
114,8	114,8	114,9	114,9	114,9	114,9	114,9	114,9	115,0	115,0	115,0	115,0	115,1	115,1	115,2	115,2	115,2	115,2
115,3	115,3	115,3	115,3	115,4	115,4	115,4	115,4	115,4	115,5	115,5	115,5	115,6	115,7	115,7	115,7	115,8	115,8
115,8	115,8	115,9	115,9	115,9	115,9	116,0	116,0	116,0	116,0	116,0	116,0	116,1	116,1	116,1	116,1	116,1	116,1
116,1	116,2	116,3	116,3	116,4	116,4	116,4	116,4	116,5	116,5	116,5	116,6	116,6	116,6	116,6	116,7	116,7	116,8
116,8	116,8	116,8	116,9	116,9	116,9	116,9	116,9	117,0	117,0	117,0	117,1	117,1	117,2	117,3	117,3	117,3	117,3
117,3	117,3	117,4	117,4	117,4	117,4	117,4	117,5	117,5	117,5	117,5	117,5	117,5	117,6	117,6	117,6	117,7	117,7
117,7	117,7	117,8	117,8	117,9	118,0	118,0	118,0	118,0	118,1	118,1	118,1	118,2	118,3	118,3	118,4	118,4	118,4
118,5	118,5	118,5	118,5	118,5	118,5	118,6	118,6	118,7	118,7	118,8	118,8	118,8	118,9	118,9	118,9	118,9	118,9
118,9	118,9	119,0	119,0	119,1	119,1	119,3	119,3	119,3	119,4	119,4	119,4	119,5	119,5	119,5	119,6	119,6	119,6
119,7	119,8	119,9	119,9	119,9	119,9	119,9	120,0	120,0	120,1	120,1	120,2	120,2	120,2	120,3	120,3	120,3	120,3
120,4	120,4	120,5	120,5	120,6	120,6	120,6	120,6	120,7	120,7	120,7	120,7	120,8	120,9	120,9	121,0	121,1	121,1
121,2	121,2	121,2	121,3	121,3	121,3	121,4	121,4	121,5	121,5	121,6	121,6	121,6	121,7	121,7	121,8	121,9	121,9
122,0	122,0	122,0	122,1	122,1	122,2	122,2	122,2	122,2	122,3	122,4	122,4	122,5	122,5	122,5	122,6	122,6	122,6
122,7	122,7	122,9	123,0	123,1	123,2	123,2	123,2	123,2	123,3	123,3	123,3	123,3	123,3	123,4	123,5	123,5	123,5
123,6	123,6	123,7	123,7	123,8	123,8	123,8	123,8	124,0	124,0	124,1	124,1	124,2	124,2	124,3	124,3	124,4	124,4
124,7	124,8	124,9	124,9	125,1	125,2	125,6	126,1	126,1	127,0	127,0	127,4	128,5	129,0				

On définit l'effectif cumulé croissant d'un réel x comme le nombre d'observations inférieures ou égales à x .

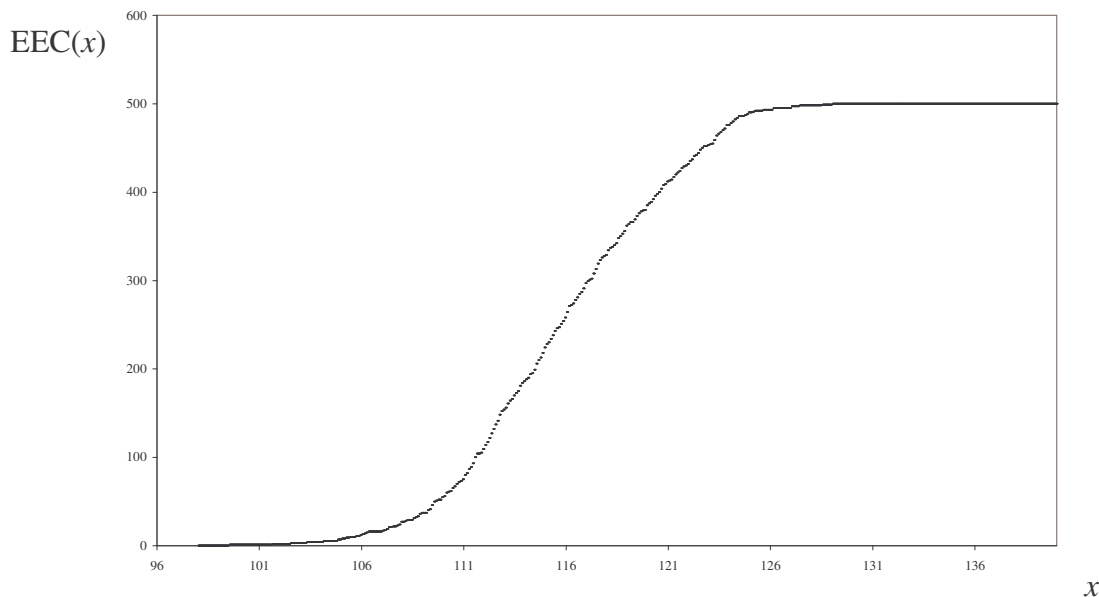
La fonction ECC des effectifs cumulés croissants est donc définie sur \mathbb{R} par $ECC(x) = \text{card}(\{C_i, 1 \leq i \leq 500 \text{ et } C_i \leq x\})$, ainsi, par exemple, pour notre exemple $ECC(109) = 37$ car il y a 37 valeurs observées inférieures ou égales à 109.

D'autre part, 109, qui n'est pas une valeur observée, est compris entre les deux valeurs observées 108,9 et 109,2 ainsi les nombres de l'intervalle $[108,9 ; 109,2[$ ont tous le même effectif cumulé croissant égal à 37.

La fonction EEC est donc la fonction en escalier définie sur \mathbb{R} de la façon suivante :

$$ECC(x) = \begin{cases} 0 & \text{si } x < 99,5 \\ 1 & \text{si } 99,5 \leq x < 101,6 \\ 2 & \text{si } 101,6 \leq x < 102,5 \\ \dots & \\ 7 & \text{si } 104,8 \leq x < 105 \\ 8 & \text{si } 105 \leq x < 105,2 \\ \dots & \\ 500 & \text{si } x \geq 129 \end{cases}$$

Elle change de valeurs à chaque observation et le "saut" est égal à un nombre d'observations de même valeur. Sa représentation graphique¹ est la suivante :



Graphique 1

¹ On trouvera sur le site r2math : <http://www.enfa.fr/r2math> les fichiers **ecc** aux formats OPENOFFICE et EXCEL où sont traitées les données.

Données groupées en classes

Supposons maintenant que l'on ne dispose sur cette série statistique que des données regroupées en classes d'amplitude 5. Cela correspond au tableau des effectifs suivant :

classes	effectifs
[95 ; 100[1
[100 ; 105[6
[105 ; 110[48
[110 ; 115[169
[115 ; 120[161
[120 ; 125[105
[125 ; 130[10

Tableau des effectifs 1

C'est souvent ainsi que sont présentées les données, on ne peut alors rien dire sur les effectifs cumulés croissants sauf pour les réels inférieurs strictement à 95 (dont les effectifs cumulés croissants sont égaux à 0) ou supérieurs ou égaux à 130 (dont les effectifs cumulés croissants sont égaux à 500). Le sens des crochets des classes autorise seulement le calcul des nombres d'observations inférieures *strictement* à chacune des bornes des classes.

Remarque :

Des crochets tournés dans l'autre sens, permettraient le calcul des effectifs cumulés croissants pour les bornes des classes, mais ce n'est pas une convention usuellement utilisée.

En attribuant aux bornes des classes, les effectifs cumulés croissants du tableau suivant, on fait l'hypothèse implicite que les données observées sont à l'intérieur strict des classes, hypothèse qui n'a pas de raison d'être vérifiée en général.

x	ECC(x)
dans $]-\infty ; 95[$	0
95	0
100	1
105	7
110	55
115	224
120	385
125	490
dans $[130 ; +\infty[$	500

Tableau 2

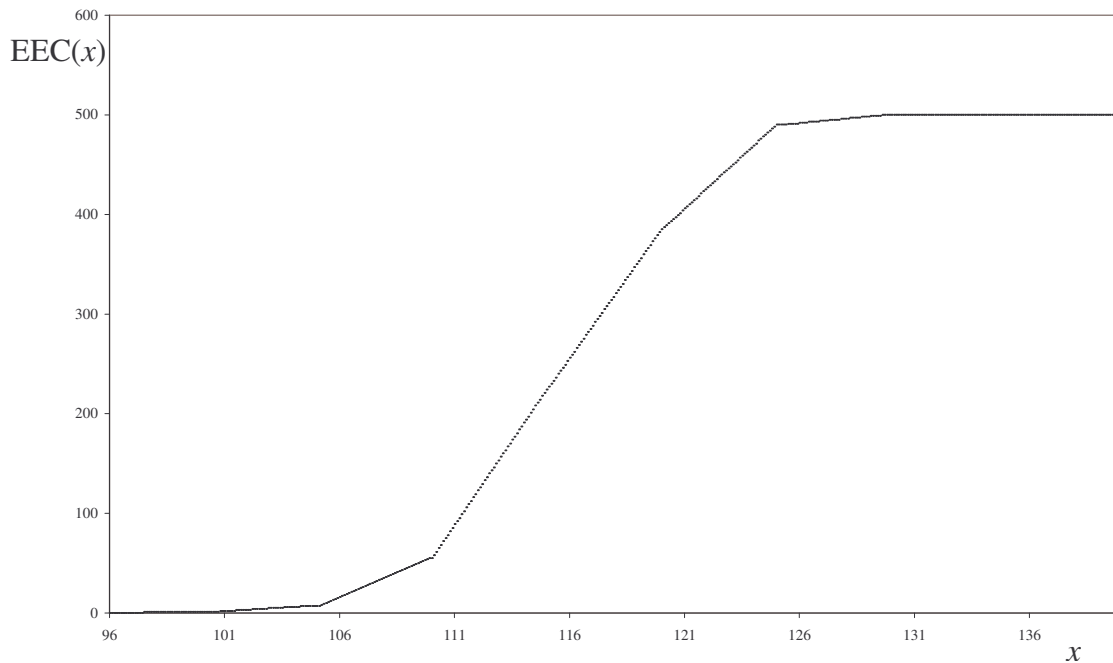
Sans rien supposer d'autre, on est encore loin de pouvoir représenter la fonction ECC (ou même une approximation de ECC) sur \mathbb{R} !

Si on suppose de plus que les données sont régulièrement réparties à l'intérieur strict des classes, on aurait alors :

- une donnée égale à 97,5 ;
- six données égales respectivement à $100 + \frac{5}{7}$, à $100 + 2 \times \frac{5}{7}$, à $100 + 3 \times \frac{5}{7}$, ..., à $100 + 6 \times \frac{5}{7}$, donc régulièrement réparties strictement entre 100 et 105 ;
- quarante-huit données égales respectivement à $105 + \frac{5}{49}$, à $105 + 2 \times \frac{5}{49}$, ..., à $105 + 48 \times \frac{5}{49}$, donc régulièrement réparties strictement entre 105 et 110 ;
- ...

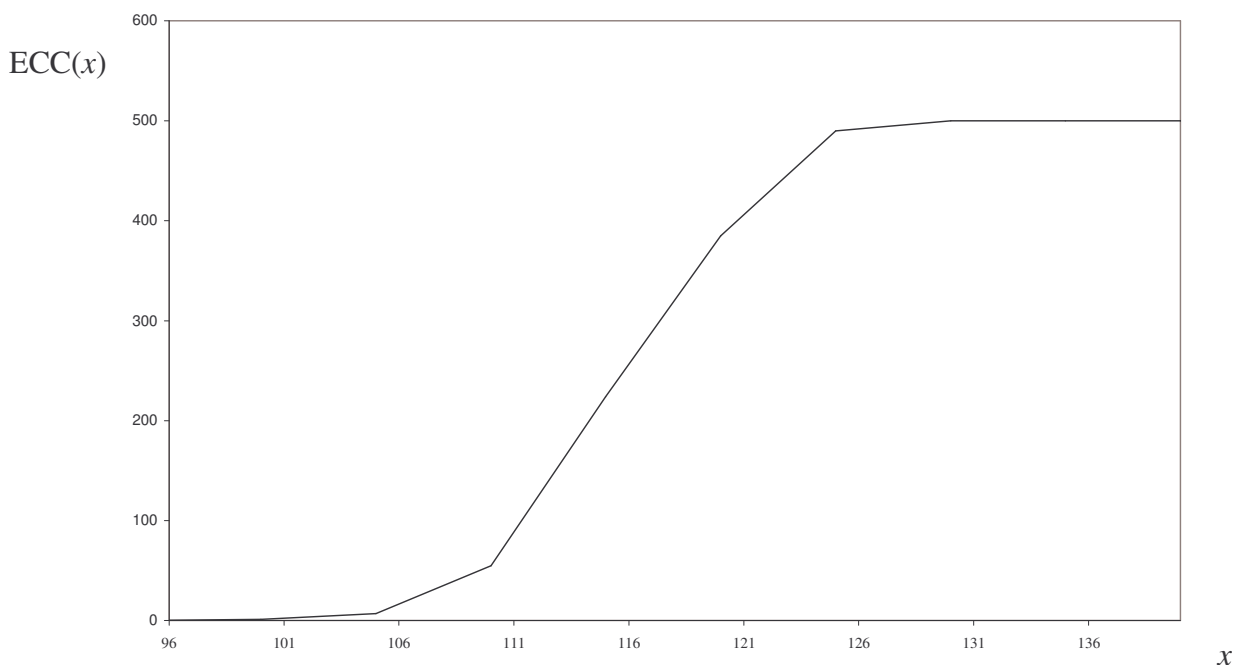
- dix données égales respectivement à $125 + \frac{5}{11}$, à $125 + 2 \times \frac{5}{11}$, ..., à $125 + 10 \times \frac{5}{11}$, donc régulièrement réparties strictement entre 125 et 130.

Dans ces conditions la représentation graphique de EEC serait la courbe en escaliers suivante :



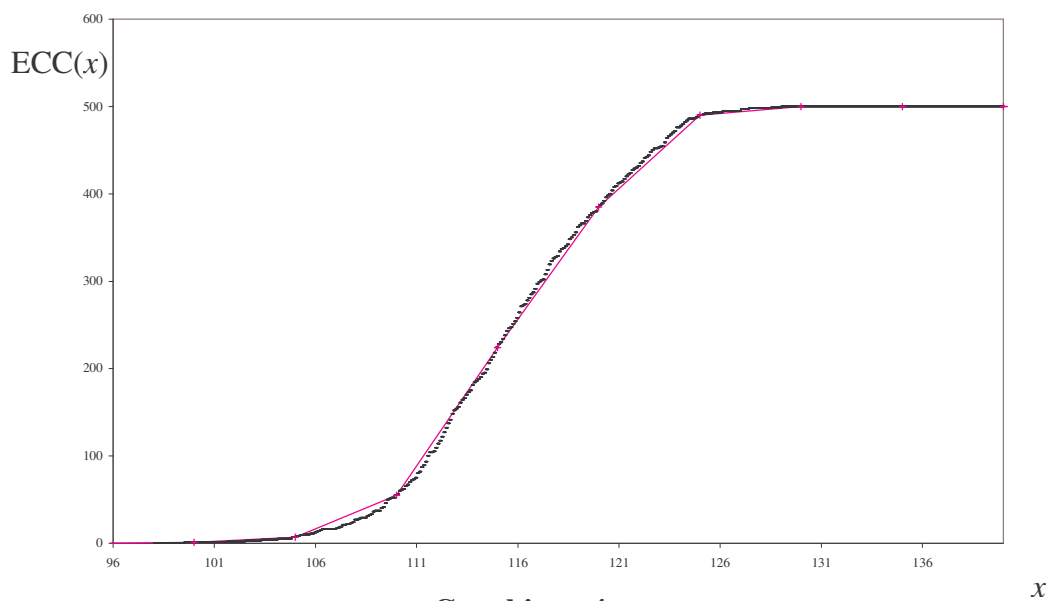
Graphique 2

On approche le tracé précédent par le polygone des effectifs cumulés croissants obtenu en joignant par des segments les points construits à partir du **tableau 2** :



Graphique 3

Pour le cas d'école qui nous intéresse, on peut comparer les représentations réelle et approchée de ECC et constater qu'ici, elles sont relativement proches :



Graphique 4

Pour conclure, que peut-on dire aux élèves pour donner du sens au polygone des effectifs cumulés croissants sans leur donner uniquement une *recette* de construction ? Certainement pas une explication aussi exhaustive que ci-dessus, de toute façon l'étude de la fonction des effectifs cumulés croissants dans le cas de données brutes ne figure pas dans les programmes des classes de lycées agricoles.

On peut cependant les amener à réfléchir sur la signification d'expressions du type "l'effectif cumulé croissant de 102,5 est le nombre d'observations inférieures ou égales à 102,5" et à comprendre que, dans le cas de données groupées en classes, on ne peut rien dire de précis en général sauf si on émet des hypothèses supplémentaires sur la répartition des données. En admettant que les données sont à l'intérieur strict des classes (les bornes des classes ne sont pas observées), on peut alors donner les effectifs cumulés croissants des bornes des classes, ce qui donne des points de la représentation graphique de la fonction des ECC, points que l'on joint par des segments de droite pour obtenir une approximation de la fonction des effectifs cumulés croissants, en supposant ainsi que les données sont réparties uniformément à l'intérieur strict des classes, faute de mieux !

Cela peut aider les élèves qui savent calculer mécaniquement les effectifs cumulés, mais qui ne savent quoi en faire graphiquement et les attribuent à l'une ou l'autre des bornes de la classe, quand ce n'est pas à son centre!

Il va sans dire que, dans ces conditions, la colonne de droite du tableau ci-dessous **n'est pas cohérente** avec les deux autres colonnes.

classes	effectifs	effectifs cumulés croissants
[95 ; 100[1	1
[100 ; 105[6	7
[105 ; 110[48	55
[110 ; 115[169	224
[115 ; 120[161	385
[120 ; 125[105	490
[125 ; 130[10	500

et qu'il conviendrait de le remplacer par deux tableaux distincts :

classes	effectifs
[95 ; 100[1
[100 ; 105[6
[105 ; 110[48
[110 ; 115[169
[115 ; 120[161
[120 ; 125[105
[125 ; 130[10

bornes des classes	effectifs cumulés croissants
95	0
100	1
105	7
110	55
115	224
120	385
125	490
130	500

ou par un tableau du type ci-dessous :

classes	effectifs	effectifs cumulés croissants
		0
[95 ; 100[1	1
[100 ; 105[6	7
[105 ; 110[48	55
[110 ; 115[169	224
[115 ; 120[161	385
[120 ; 125[105	490
[125 ; 130[10	500

Remarquons pour finir que, en accords avec les référentiels en vigueur, beaucoup de calculs se font à partir des données réparties en classes et cela historiquement du fait de l'absence de moyens de calcul. De nos jours, ces calculs peuvent et doivent se faire, à l'aide de tableurs par exemple, à partir de données brutes. Le regroupement en classes ne prend son sens que pour permettre de décrire la répartition (dissymétrie, aplatissement, bimodalité, etc.) des données statistiques, en particulier à l'aide de représentations graphiques.

16^e RMT - La boîte de Nelly (CAT. 8, 9)

Nelly a une boîte en forme de parallélépipède rectangle dont les trois dimensions intérieures sont des nombres entiers de centimètres. Elle peut y placer une aiguille à tricoter de 15 cm de longueur exactement, sur la grande diagonale, avec une extrémité en un sommet inférieur et l'autre extrémité au sommet supérieur opposé.

Quelles peuvent être les dimensions de la boîte de Nelly ?

Expliquez votre réponse.