Table des matières

Principe:

Soit une expérience aléatoire observée n fois avec trois issues (résultats) possibles, notées i_1, i_2, i_3 d'effectifs respectifs n_1, n_2, n_3 (On a donc: $n_1+n_2+n_3=n$)
On obtient le tableau des ***effectifs observés***.

Résultats	i_1	i_2	i_3	Total
Effectifs observés	n_1	n_2	n_3	n

On veut savoir si l'expérience aléatoire est conforme (en adéquation) à un modèle où les résultats sont équiprobables, on a alors le tableau suivant des effectifs théoriques

Résultats	i_1	i_2	i_3	Total
Effectifs théoriques	$\frac{n}{3}$	$\frac{n}{3}$	$\frac{n}{3}$	n

pour cela, on évalue l'écart (la distance) entre les deux répartitions
(*de même qu'on évalue une distance en géométrie analytique entre deux points* $A(x_A; y_A; z_A)$ *et* $B(x_B; y_B; z_B)$ *par* $d^2=(x_B-x_A)^2+(y_B-y_A)^2+(z_B-z_A)^2$ *)*
en calculant: $D^2_{obs}=(n_1-\frac{n}{3})^2+(n_2-\frac{n}{3})^2+(n_3-\frac{n}{3})^2$

On ***simule*** ensuite n expériences conformes au modèle équiprobable.
On répète ***un grand nombre de fois*** cette simulation et on calcule à chaque fois D^2_{sim} .
On approche ainsi la loi de probabilité de la variable aléatoire D^2 .

En général, on travaille avec les fréquences: $f_i=\frac{n_i}{n}$

$$d^2_{obs}=(\frac{n_1}{n}-\frac{1}{3})^2+(\frac{n_2}{n}-\frac{1}{3})^2+(\frac{n_3}{n}-\frac{1}{3})^2=(f_1-\frac{1}{3})^2+(f_2-\frac{1}{3})^2+(f_3-\frac{1}{3})^2$$

$$d^2_{obs}=\frac{1}{n^2}D^2_{obs} .$$

Selon le tableau obtenu, on pourra accepter ou non le modèle équiprobable.

Le raisonnement est identique pour une expérience aléatoire ayant k issues possibles):

$$d^2_{obs}=(f_1-\frac{1}{k})^2+(f_2-\frac{1}{k})^2+...+(f_k-\frac{1}{k})^2$$

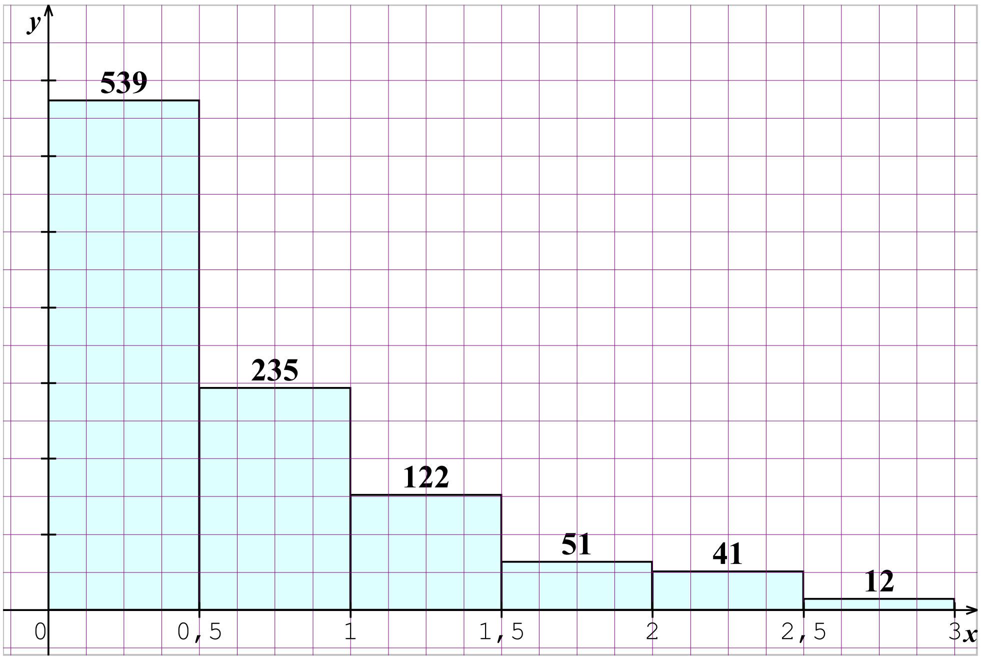
Exemple: (à partir du Bac ES Pondichéry 2003)

Énoncé

Un pisciculteur possède un bassin qui contient trois variétés de truites: communes, saumonées et arc-en ciel. Il voudrait savoir s'il peut considérer que son bassin contient autant de truites de chaque variété.
Pour cela, il effectue au hasard, 400 prélèvements d'une truite avec remise et obtient les résultats suivants:

Variétés	**Commune**	**Saumonée**	**Arc-en-ciel**
Effectifs	146	118	136

1a) Calculer les fréquences de prélèvement f_c , f_s et f_a .
b) Calculer $400d^2_{obs}$
2) A l'aide d'un ordinateur le pisciculteur simule le prélèvement au hasard de 400 truites suivant la loi équirépartie. Il répète 1 000 fois cette opération et calcule à chaque fois la valeur de 400 d^2. Le diagramme à bandes représente la série des 1 000 valeurs de 400 d^2 obtenues par simulation.



Déterminer une valeur approchée à 0,5 près par défaut du neuvième décile D_9 de cette série.
3) Peut-on affirmer avec un risque d'erreur inférieur à 10% que le bassin contient autant de truites de chaque variété?

Compte-rendu

1) a) b)

Variétés	**Commune**	**Saumonée**	**Arc-en-ciel**	**Total**
Effectifs	146	118	136	400
Fréquences	$f_c=0{,}365$	$f_s=0{,}295$	$f_a=0{,}340$	1
Écarts	$0{,}365-\frac{1}{3}=\frac{19}{600}$	$0{,}295-\frac{1}{3}=\frac{-23}{600}$	$0{,}340-\frac{1}{3}=\frac{4}{600}$	XXXXXXXX
Carrés	$\frac{361}{360000}$	$\frac{529}{360000}$	$\frac{16}{360000}$	$d^2=\frac{906}{360000}$

$400d_{obs}^2=\frac{151}{150}$

Une valeur approchée à 10^{-2} près par défaut de $400d_{obs}^2$ *est* 1

Le fait de multiplier par 400 permet d'avoir des valeurs plus lisibles.

2) Tableau des effectifs cumulés et fréquences cumulées des valeurs de $400d^2$ pour les 1000 simulations.

$400d^2$	**[0;0,5]**	**[0,5;1]**	**[1;1,5]**	**[1,5;2]**	**[2;2,5]**	**[2,5;3]**
Effectifs cumulés	539	774	896	947	988	1000
fréquences cumulées	0,54	0,77	0,9	0,95	0,99	1

Le neuvième décile D_9 est donné par la valeur du $900^{ième}$ résultat (on peut décompter par la fin et prendre le $100^{ième}$)

D_9 est donc dans la classe [1,5;2] et une valeur à 0,5 par défaut est alors 1,5.

Pour déterminer une valeur théorique de D_9 *plus précise, on fait une interpolation linéaire dans la classe [1,5;2] du* $900^{ième}$ *résultat: On a (proportionnalité)* $\frac{D_9-1,5}{2-1,5}=\frac{900-896}{947-896}$

$D_9=0.5\times\frac{900-896}{947-896}+1,5\approx 1,539$

3) Dans un peu plus de 10% des cas, les valeurs simulées de $400d^2$ dépassent 1,5.

Comme la valeur observée de $400d_{obs}^2$ est 1, rien ne permet de rejeter l'hypothèse d'équiprobabilité.

Compléments:

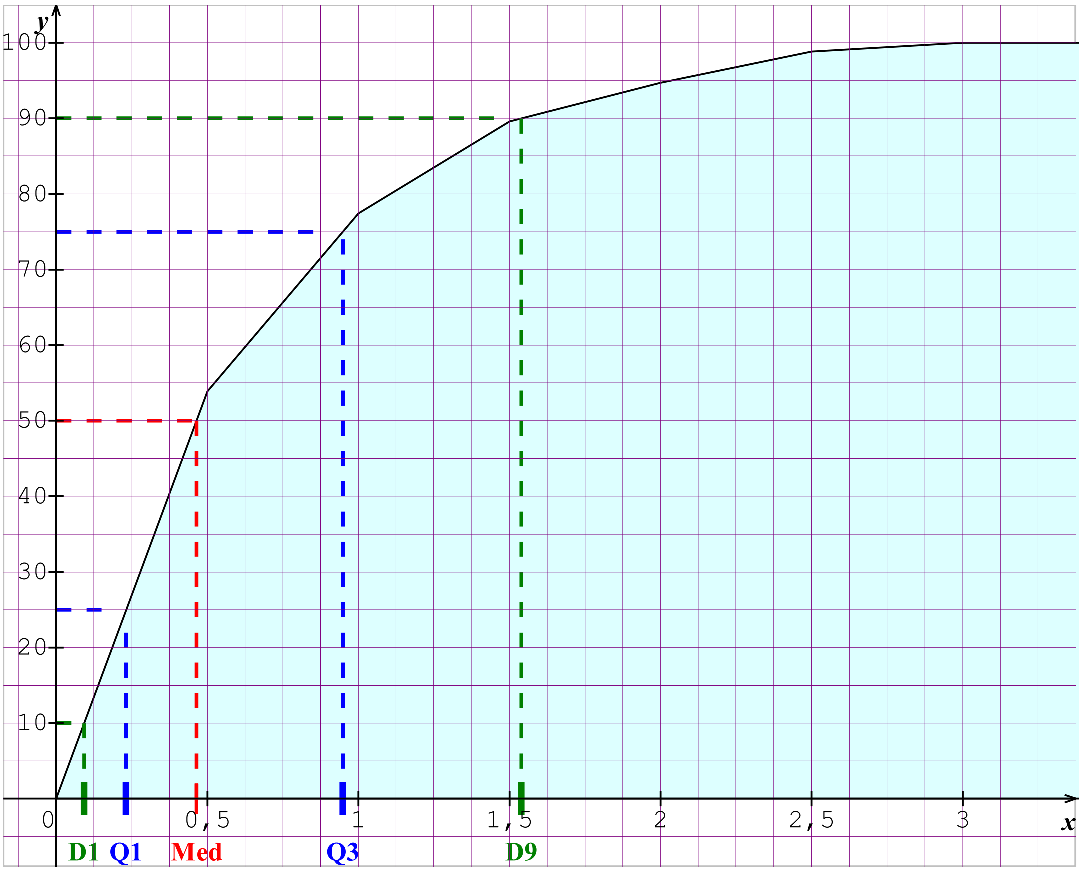
Le premier décile D_1 *est la* $100^{ième}$ *valeur.* D_1 *est dans la classe [0;0,5], une approximation linéaire donne:* $D_1=0+0,5\times\frac{100}{539}\approx 0,0927$

La médiane M_e *est la valeur du* $500^{ième}$ *valeur.* M_e *est dans la classe [0;0,5], une approximation linéaire donne:* $M_e=0+0,5\times\frac{500}{539}\approx 0,46$

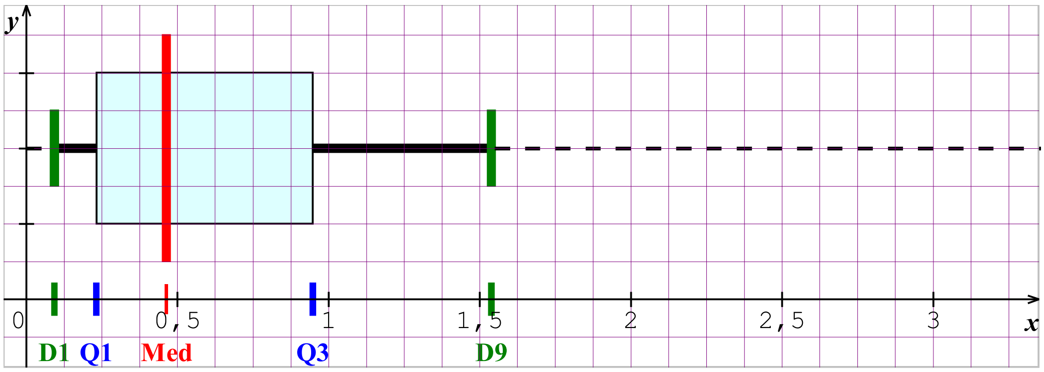
Le premier quartile Q_1 *est la* $250^{ième}$ *valeur .* Q_1 *est dans la classe [0;0,5], une approximation linéaire donne:* $Q_1=0+0,5\times\frac{250}{539}\approx 0,23$

Le troisième quartile Q_3 *est la* $750^{ième}$ *valeur.* Q_3 *est dans la classe [0,5;1], une approximation linéaire donne:* $Q_3=0,5+0,5\times\frac{750-539}{235}\approx 0,95$ *par excès*

Diagramme des fréquences cumulées



Diagrammes en boîte



Exercice commun S et ES

1)

Variété	**1**	**2**	**3**	**4**	**Total**
Nombre de grains radioactifs	18	27	35	20	100
Fréquences	$f_1=0{,}18$	$f_2=0{,}27$	$f_3=0{,}35$	$f_4=0{,}20$	1
Écarts	$0{,}18-\frac{1}{4}=$ $-0{,}07$	$0{,}27-\frac{1}{4}=$ $0{,}02$	$0{,}35-\frac{1}{4}=$ $0{,}1$	$0{,}20-\frac{1}{4}=$ $-0{,}05$	XXXXXXXX
Carrés	0,004 9	0,000 4	0,01	0,002 5	0,017 8

$d^2=400\times0{,}017\,8=7{,}12$

Le fait de multiplier par 400 permet d'avoir des valeurs plus lisibles.

2) 10 000 simulations de 100 tirages de grains de blé.

j	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Nombre de séries pour lesquelles $d^2>j$	3915	2618	1715	1114	728	467	306	180

Ce tableau donne les effectifs cumulés décroissants, d'où, le neuvième décile D_9 arrondi à l'entier le plus proche est 6

(90% des effectifs prennent une valeur inférieure à D_9 ou encore 10% des effectifs prennent une valeur supérieure à D_9)

Le $95^{ième}$ centile C_{95} arrondi à l'entier le plus proche est 8

(95% des effectifs prennent une valeur inférieure à C_{95} ou encore 5% des effectifs prennent une valeur supérieure à C_{95})

3 a) Comme $d^2>D_9$, un tel mélange est considéré comme non homogène avec une marge d'erreur de 10%)

3 b) Avec une marge d'erreur de 5%, on raisonne alors par rapport au 95ème centile

Comme $d^2\leq C_{95}$, nous pouvons affirmer avec une marge d'erreur de 5% que le mélange est homogène.

Cela peut paraître paradoxal mais s'explique ainsi : se donner une marge d'erreur de 5% signifie que l'on considère que seulement 5% des résultats de la simulation sont marginaux. Il est donc moins probable que notre observation tombe dans ces 5% que dans une tranche de 10%. Une marge d'erreur de 5% a donc tendance à valider notre observation comme étant presque toujours équirépartie !

3 c) Et si, cas extrême, on se donne une marge d'erreur de 0%, cela signifiera que toute observation tombant dans la plage de valeurs obtenues par simulation sera considérée comme suivant un modèle d'équirépartition (autrement dit, une marge d'erreur de 0% signifie qu'on ne prend pas le risque de dire que le mélange n'est pas homogène.

Compléments:

Tableau des effectifs par classe

j	[0;3]	]3;4]	]4;5]	]5;6]	]6;7]	]7;8]	]8;9]	]9;10]	]10;11]
Effectifs	6085	1297	903	601	386	261	161	126	180

Diagramme des fréquences cumulées **décroissantes**

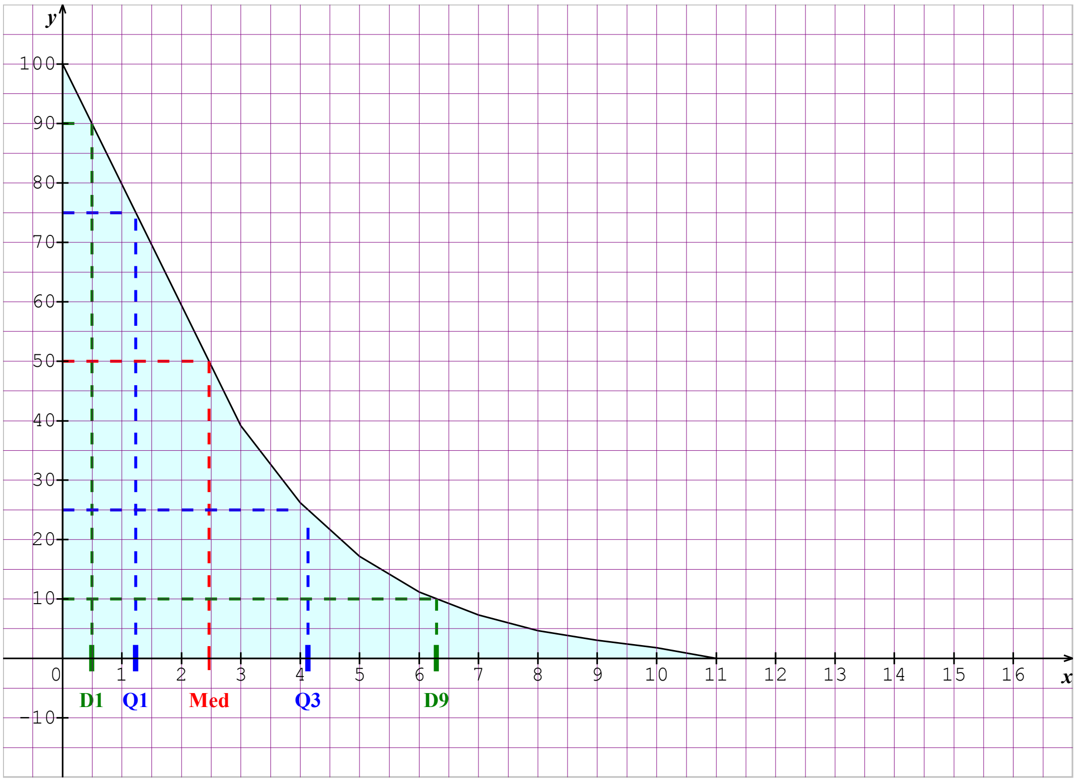
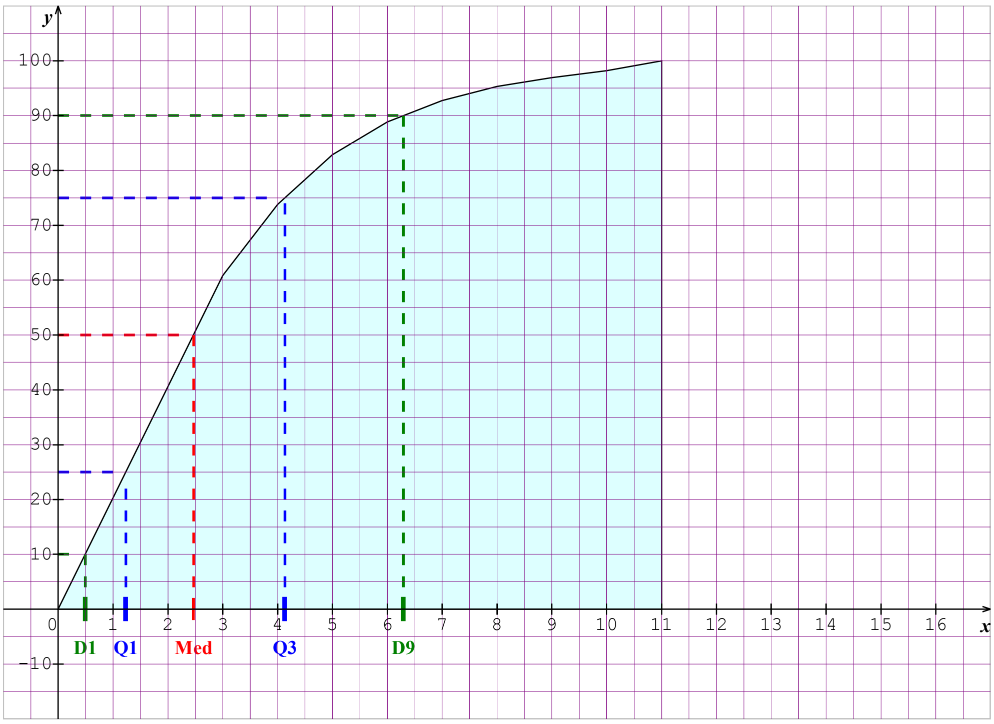


Diagramme des fréquences cumulées **croissantes**



Ce que l'on conçoit bien s'énonce clairement, Et les mots pour le dire arrivent aisément. Boileau

Regroupement par classes