Chapitre 1 : Les distributions statistiques à une dimension

Objectif général du chapitre :

Comprendre et maîtriser les distributions statistiques à une dimension.

I. Présentation générale d'un tableau statistique

Objectifs spécifiques :

- Comprendre la présentation générale d'un tableau statistique.

Durée : 0,5 H.

Contenu :

On suppose une population « P » avec un nombre total des individus « n » et décrite selon le caractère « C » dont les « k » modalités sont : {M_1, M_2,… M_i,… M_k}.

On pose, aussi, n_i : le nombre des individus qui présentent la modalité M_i, avec $\sum_{i=1}^{k} n_i = n$.

Et f_i : La fréquence de la modalité M_i, avec $f_i = \frac{n_i}{n}$: C'est la proportion des individus qui présentent la modalité M_i, avec $\sum_{i=1}^{k} f_i = 1$.

Le tableau statistique des effectifs et des fréquences se présente de la manière suivante :

(1) Caractère	(2) Effectif n_i	(3) Fréquence f_i
M_1	n_1	f_1
M_2	n_2	f_2
…	…	…
M_i	n_i	f_i
…	…	…
M_k	n_k	f_k
Total	n	1

Titre : Tableau 1 : La répartition de la population P selon le caractère C.

II. Distribution à caractère qualitatif

Objectifs spécifiques :

- Maîtriser les principes d'une distribution à caractère qualitatif.

Durée : 1 H.

Contenu :

a/ Tableau statistique :

On va présenter le tableau statistique d'un caractère qualitatif à travers l'exemple n° 1 : Une entreprise veut étudier la qualification de ses employés présente les données suivantes :

(1) Qualification	(2) Nombre d'employés : n_i	(3) Fréquence: f_i
Apprentis	5	0,05
Stagiaires	15	0,15
Ouvriers	56	0,56
Ouvriers qualifiés	24	0,24
Total	100	1

Titre : Tableau 2 : la distribution des effectifs et des fréquences de 100 employés suivant la qualification.

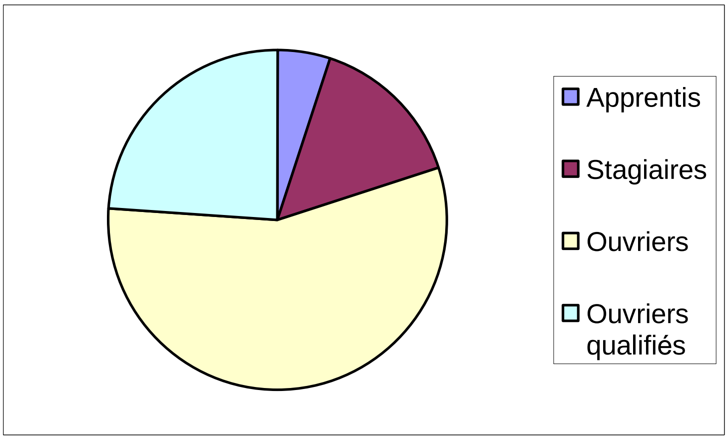
b/ Représentation graphique :

- Diagramme en secteur circulaire : La population est représentée par un cercle divisé en k secteurs. Chaque secteur possède un angle noté $\theta_i = 360 \times f_i$.

Application à l'exemple 1 : On ajoute au tableau 2 au-dessus la colonne (4) suivante :

Qualification	(4) θ_i °
Apprentis	18
Stagiaires	54
Ouvriers	201,6
Ouvriers qualifiés	86,4
Total	360

Titre : Tableau 3 : La distribution des écarts (θ_i) de 100 employés suivant la qualification.

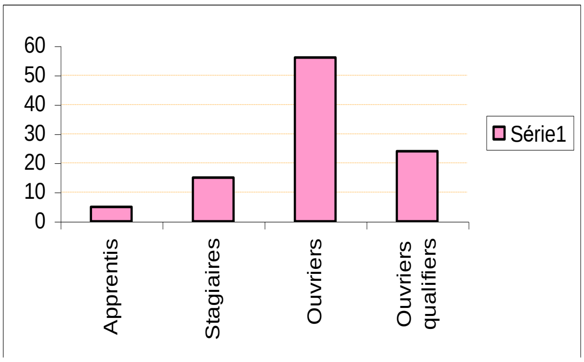


Titre : Graphique 1 : Diagramme en secteurs circulaire des employés suivant la qualification.

- Graphique en tuyaux d'orgue :

Les tuyaux ont la même largeur, leurs hauteurs sont proportionnelles aux valeurs de n_i ou de f_i. On choisit aussi des espaces égaux entre les tuyaux.

Application à l'exemple 1 : On va représenter le graphique des qualifications en fonction des effectifs des employés.



Titre : Graphique 2 : Graphique en tuyaux d'orgue des employés suivant leurs qualifications

III. Distribution à caractère quantitatif

Objectifs spécifiques :

- Comprendre le mécanisme d'une distribution à caractère quantitatif.

Durée : 3 H.

Contenu :

A/ <u>Variable statistique discrète :</u>

a/ Présentation d'un tableau statistique :

On a : X est une variable discrète, x_i est la valeur numérique de la modalité « M_i » (ou une variable particulière de la variable statistique X).

Exemple n° 2 : La répartition des ouvriers suivant le caractère « le nombre de retards pour le mois de mars » :

(1) Nombre de retards	(2) Nombre d'ouvriers n_i	(3) Fréquence f_i
0	8	0,16
1	6	0,12
2	22	0,44
3	10	0,2
4	4	0,08
Total	50	1

Titre:Tableau 4 : La distribution des effectifs et des fréquences des ouvriers suivant le nombre de retards

Dans ce cas, il est nécessaire de définir quelques éléments qu'on aille utiliser dans les représentations graphiques :

❖ Définition 1 : On appelle une fonction de répartition l'application notée :

$F : \Re \rightarrow [0, 1]$

$x_i \rightarrow F(x_i) = P[X < x_i]$

$P[X < x_i]$ est la proportion des individus pour lesquels la variable statistique est < à une valeur particulière x_i. La fonction de répartition présente plusieurs propriétés :

♠ Quel que soit $x \leq x_1$, $F(x) = 0 \Rightarrow F(x_1) = 0$.

♠ $F(x_2) = P[X < x_2] = P[X = x_1] = f_1$.

$F(x_3) = P[X < x_3] = P[X = x_1 \text{ ou } X = x_2] = P[X = x_1] + P[X = x_2] = f_1 + f_2$.

Et ainsi de suite, jusqu'à : $F(x_i) = P[X < x_i] = P[X = x_1 \text{ ou } X = x_2 \text{ ou} \ldots X = x_i] = P[X = x_1] + P[X = x_2] + \ldots + P[X = x_{i-1}] = f_1 + f_2 + \ldots + f_{i-1}$. D'où, $F(x_i) = \sum_{j=1}^{i-1} f_j$, quel que soit $i \geq 2$. Et $F(x_k) = \sum_{j=1}^{k-1} f_j < 1$.

♠ Quel que soit $x > x_k$, $F(x) = 1$.

♠ Quel que soit $x \in]x_{i-1}, x_i[$, c'est à dire que $x = x_{i-1} + \varepsilon$, avec $\varepsilon \in]0, x_i - x_{i-1}[$.

Alors, $F(x) = F(x_{i-1} + \varepsilon) \Leftrightarrow [X < x_{i-1} + \varepsilon] = P[X < x_i] = F(x_i)$. D'où, la courbe de F sera discontinue à droite de $x_{(i-1)}$.

❖ Définition 2 : $F(x_i) = \sum_{j=1}^{i-1} f_j$: est appelée la fréquence cumulée croissante notée FC ↑.

❖ Définition 3 : la fréquence cumulée décroissante (notée FC ↓) :

La fonction cumulée décroissante est la fonction : $G(x_i) = P[X \geq x_i] = 1 - P[X < x_i]$, alors, $G(x_i) = 1 - F(x_i)$, avec $G(x_i)$ est la proportion des individus pour lesquels la valeur de la variable est ≥ à une valeur particulière x_i.

❖ Définition 4 : L'effectif cumulé croissant (noté EC ↑) :

$E(x_i) = N[X < x_i] = \sum_{j=1}^{i-1} n_j$: c'est le nombre des individus pour lesquels la valeur de la variable est < à une valeur particulière x_i. En particulier, on a : $E(x_1) = 0$ et $E(x_k) = \sum_{j=1}^{k-1} n_j < n$.

❖ Définition 5 : l'effectif cumulé décroissant (EC ↓) :

$H(x_i) = N[X \geq x_i] = n - N[X < x_i] = n - E(x_i)$: désigne le nombre des individus pour lesquels la valeur de la variable est ≥ à une valeur particulière x_i.

Le tableau statistique 4 sera, alors, complété (dans le cas général) de la manière suivante :

Variable (1)	n_i (2)	f_i (3)	EC ↑ (4)	EC ↓ (5)	FC ↑ (6)	FC ↓ (7)
x_1	n_1	f_1	0	$n - 0$	0	$1 - 0$
x_2	n_2	f_2	n_1	$n - n_1$	f_1	$1 - f_1$
…	…	…	…	…	…	…
x_k	n_k	f_k	$n_1 + n_2 + .. + n_{k-1} < n$	$n - (n_1 + n_2 + .. + n_{k-1})$	$f_1 + f_2 + .. + f_{k-1} < 1$	$1 - (f_1 + f_2 + .. + f_{k-1})$
Total	n	1				

Titre : Tableau 5 : La distribution des effectifs simple et cumulés, fréquences simples et cumulées dans le cas général.

Application à l'exemple 2 : Tableau 3 complété de la manière suivante :

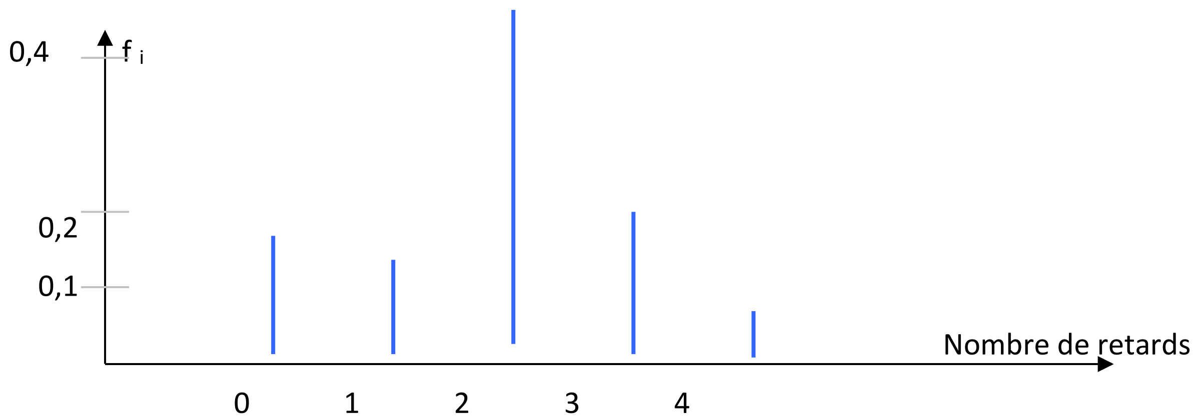
Nombre de retards (1)	n_i (2)	f_i (3)	EC ↑ (4)	EC ↓ (5)	FC ↑ (6)	FC ↓ (7)
0	8	0,16	0	50 – 0	0	1 – 0
1	6	0,12	8	42	0,16	0,84
2	22	0,44	14	36	0,28	0,72
3	10	0,2	36	14	0,72	0,28
4	4	0,08	46 < 50	4	0,92 < 1	0,08
Total	50	1				

Titre : Tableau 6 : La distribution des effectifs simple et cumulés, fréquences simples et cumulées des ouvriers selon le nombre de retard.

b/ Représentations graphiques :

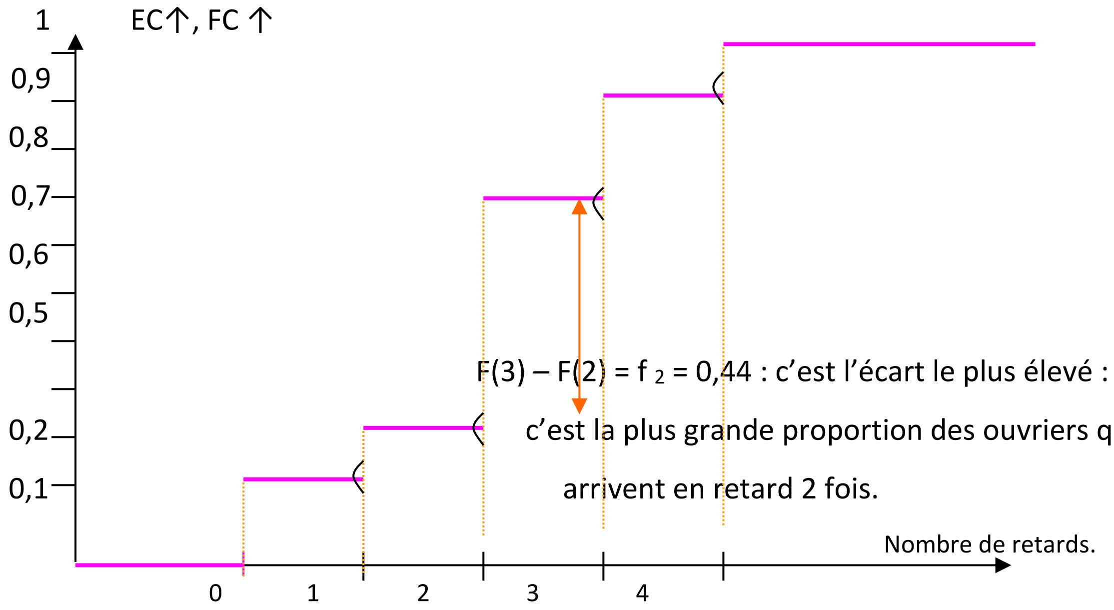
- Diagramme différentiel : diagramme en bâton. C'est une représentation graphique soit des fréquences, soit des effectifs. A chaque valeur de la variable portée en abscisse, on fait correspondre un segment vertical (ou un bâton) dont la hauteur est égale soit à la fréquence, soit l'effectif respectif.

Application à l'exemple 2 :



Titre : Graphique 3 : diagramme en bâton des ouvriers selon le nombre de retards.

- Diagramme intégral : diagramme en escalier : C'est la représentation graphique de FC croissante ou EC croissant :



Titre : Graphique 4 : Représentation des FC ↑ des ouvriers selon le nombre de retards : diagramme en escalier.

Remarque : On peut représenter, aussi, dans cette courbe FC ↓ ou EC ↓ (on obtient l'escalier opposé).

B/ <u>Variable statistique continue :</u>

a/ Présentation d'un tableau statistique 5 :

(1)	(2)	(3)
Variable X	n_i	f_i
[e_0, e_1[	n_1	f_1
[e_1, e_2[	n_2	f_2
...	...	...
[e_{i-1}, e_i[	n_i	f_i
...		
[e_{k-1}, e_k[	n_k	f_k
Total	$\sum_{i=1}^{k} n_i = n$	$\sum_{i=1}^{k} f_i = 1$

Titre : Tableau 7 : La répartition de la population P suivant le caractère continu C.

On lit la classe [e_0, e_1[: de e_0 à moins de e_1.

L'effectif « n_i » indique le nombre des individus ayant :

une valeur de la variable ≥ e_{i-1} et < à e_i.

La fréquence « f_i » indique la proportion des individus qui ont une valeur ≥ e_{i-1} et < à e_i.

L'amplitude de la classe [e_{i-1}, e_i[est définie par : « a_i » = $e_i - e_{i-1}$. D'une classe à l'autre, l'amplitude peut être constante ou variable.

Exemple n° 3 : La répartition du personnel d'une entreprise selon le salaire horaire :

(1) Salaire horaire en D	(2) n_i	(3) f_i
[0,8 – 1 [	32	0,16
[1 – 1,2 [	**56** = N [1≤X≤1,2]	0,28
[1,2 – 1,6 [	86	**0,43** = P [1,2≤X≤1,6]
[1,6 – 1,8 [	26	0,13
Total	200	1

Titre : Tableau 8 : La répartition du personnel d'une entreprise selon le salaire horaire

Puisque les modalités se présentent sous forme de classes, il faut préciser les éléments suivants :

❖ La fonction de répartition garde sa formule de définition à savoir : F (x_i) = P [X< x_i].

♦ F (e_0) = 0 et quel que soit $x \le e_0$, alors, F (x) = 0.

F (e_1) = P [X< e_1] = P [e_0 ≤X< e_1] = f_1.

F (e_2) = P [X< e_2] = P [$e_0 \le X < e_1$ ou $e_1 \le X < e_2$] = $f_1 + f_2$.

Ainsi, F (e_i) = $\sum_{j=1}^{i} f_j$, quel que soit i = 1, 2, ..., k. Et, enfin, F (e_k) = $\sum_{j=1}^{k} f_j$ = 1.

♦ Quel que soit $x \in]\, e_{i-1}, e_i\, [$, F (x) est alors déterminée par interpolation linéaire dont le principe est le suivant. On a les correspondances suivantes :

$$\begin{cases} e_{i-1} \rightarrow F(e_{i-1}), \\ x \rightarrow F(x), \\ e_i \rightarrow F(e_i). \end{cases}$$

La formule s'écrit de la manière suivante :

$\dfrac{x-e_{i-1}}{e_i-e_{i-1}} = \dfrac{F(x)-F(e_{i-1})}{F(e_i)-F(e_{i-1})}$; et puisque le seul inconnu est F (x), donc, on peut écrire :

$$F(x) = F(e_{i-1}) + \frac{F(e_i)-F(e_{i-1})}{e_i-e_{i-1}}\,(x-e_{i-1}).$$

❖ Les modalités étant des classes, donc, par convention, on a :

♦ FC↑ : la fréquence cumulée croissante : F ([e_{i-1}, e_i [ ) = P [X < e_i] = F (e_i) : on lit la borne supérieure. Alors, F (e_i) = $\sum_{j=1}^{i} f_j$. En particulier, F ([e_0, e_1 [ ) = F (e_1) = f_1 et F ([e_{k-1}, e_k [ ) = F (e_k) = $\sum_{j=1}^{k} f_j$ = 1.

♦ FC↓ : la fréquence cumulée décroissante : G ([e_{i-1}, e_i [ ) = P [X ≥ e_{i-1}] = G (e_{i-1}) : on lit la borne inférieure. Alors, G (e_{i-1}) = 1 – F (e_{i-1}).

♦ EC↑ : l'effectif cumulé croissant : N ([e $_{i-1}$, e $_i$ [ ) = N [X < e $_i$] = E (e $_i$) : on lit la borne supérieure.Alors, E (e$_i$) = $\sum_{j=1}^{i} n_j$.

En particulier, E ([e $_0$, e $_1$ [ ) = E (e $_1$) = n $_1$ et E ([e $_{k-1}$, e $_k$ [ ) = E (e $_k$) = $\sum_{j=1}^{k} n_j$ = n.

♦ EC↓ : l'effectif cumulé décroissant : H ([e $_{i-1}$, e $_i$ [ ) = N [X ≥ e $_{i-1}$] = H (e $_{i-1}$) : on lit la borne inférieure.Alors, H (e$_{i-1}$) = n – E (e$_{i-1}$).

Application à l'exemple 3 :

(1) Salaire horaire	(2) n_i	(3) f_i	(4) EC ↑	(5) EC ↓	(6) FC ↑	(7) FC ↓
[0,8 – 1 [	32	0,16	32	200 = 200 – 0	0,16	1 – 0 =1
[1 – 1,2 [	56	0,28	88 = n_1+ n_2	200-32=168	0,44	0,84
[1,2 – 1,6 [	**86**=N[1,2≤X≤1,6]	0,43	**174**=N[X<**1,6**]	**112**=N[X≥**1,2**]	**0,87**=F[X<**1,6**]	**0,56**=P[X≥**1,2**]
[1,6 – 1,8 [	26	0,13	200 = n	26	1	0,13
Total	200	1				

Titre: Tableau 9: La distribution des effectifs simples et cumulés, des fréquences simples et cumulées du personnel selon leur salaire horaire.

b/ Représentations graphiques :

- Diagramme différentiel : histogramme : sert à représenter les « h $_i$ ».

Pour déterminer « h $_i$ », on distingue deux cas :

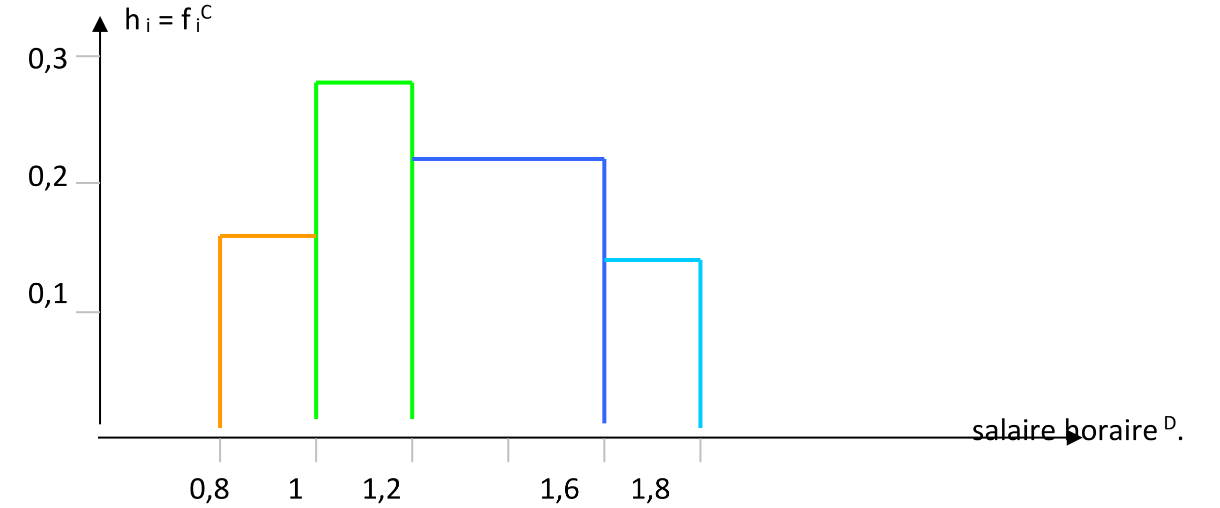
♪ Pour les amplitudes « a $_i$ » égales : $\begin{cases} h_i = f_i, \\ \text{Ou } h_i = n_i. \end{cases}$

♪ Pour les amplitudes inégales : $\begin{cases} h_i = f_i^C = \dfrac{\alpha}{a_i} \times f_i \text{ : appelé fréquence corrigée,} \\ \text{ou } h_i = n_i^C = \dfrac{\alpha}{a_i} \times n_i \text{ : appelé effectif corrigé,} \end{cases}$

avec α l'amplitude la plus fréquente ou la plus petite. On ajoute alors, au tableau 6 les colonnes (8) et (9) :

(8) $a_i = e_i - e_{i-1}$	(9) $h_i = f_i^C$
0,2	0,16
0,2	0,28
0,4	0,215
0,2	0,13

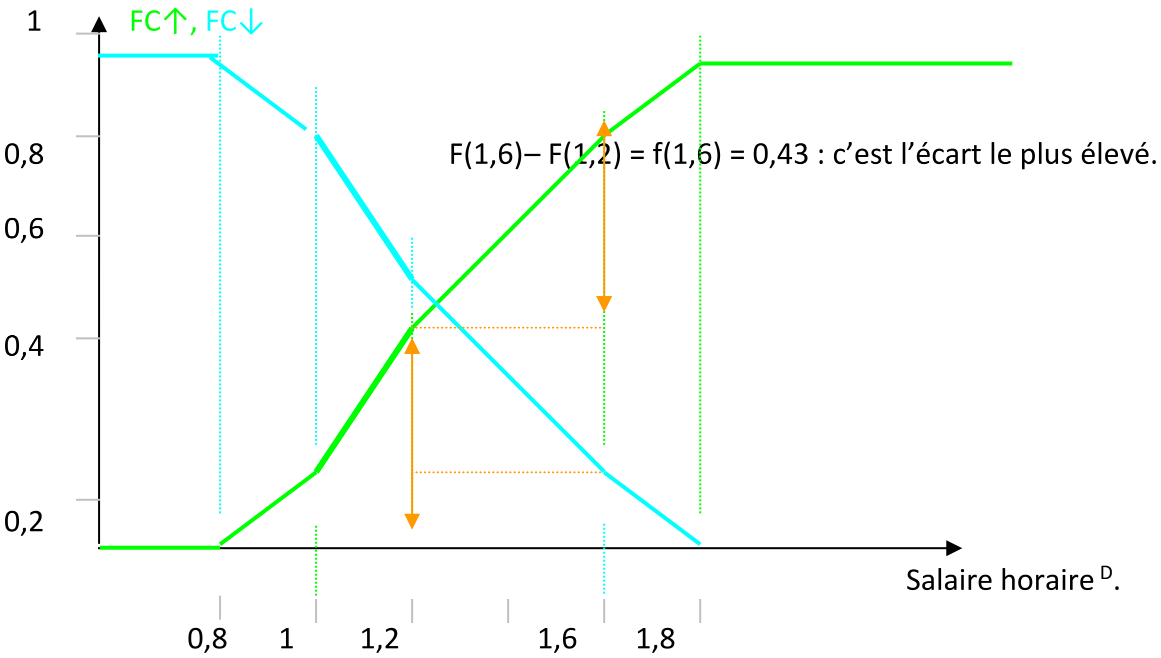
Titre: suite du tableau 9.



Titre : Graphique 5 : histogramme : Représentation des fréquences corrigées du personnel selon le salaire horaire.

Interprétation : La majorité du personnel reçoivent un salaire horaire entre 1,2 et 1,6.

- Diagramme intégral : courbe cumulative : sert à représenter les courbes respectives des EC↑, EC↓, FC↑ et FC↓. Application à l'exemple 3, on utilise les colonnes (4, 5, 6, 7) du tableau 6 :



Titre : Graphique 6 : courbe cumulative : Représentation des fréquences cumulées du personnel suivant le salaire horaire.

Interprétation : La majorité du personnel reçoivent un salaire horaire entre 1,2 et 1,6.