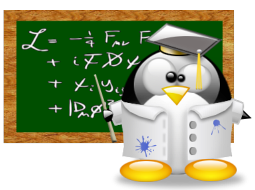
Stéphane PASQUET, 9 septembre 2014

Sommaire



Prérequis

- Statistiques de 3^{e}

Définitions *Types de séries statistiques*

Il existe deux principaux types de séries statistiques :

- Les séries statistiques dont les caractères (ce que l'on observe) sont `qualitatifs` (traduits par des mots) ;
- Les séries statistiques dont les caractères sont `quantitatifs` (traduits par des nombres).

Exemples

- Si nous observons la couleur des voitures dans un parking, les caractères sont qualitatifs (bleu, rouge, vert, ...).
- Si nous observons le salaire des employés d'une société ou le nombre de stylos à encre bleue dans la trousse des élèves d'une classe de Seconde, les caractères sont quantitatifs.

Méthode *Représentation d'une série statistique à caractères qualitatifs*

Si une série statistique est à caractères qualitatifs, on peut la représenter par :

- un diagramme circulaire ou semi-circulaire ;
- un diagramme en bâtons.

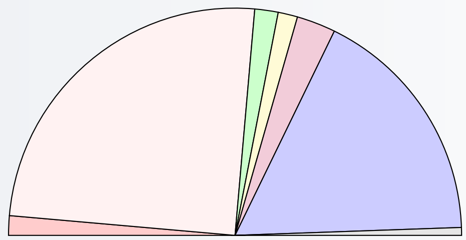
Exemples

- Au 08/09/2014, la composition de l'Assemblée Nationale était donnée par le tableau suivant :

Groupe	Effectif	Angles
Socialiste, républicain et citoyen	290	$\dfrac{290}{577} \times 180^\circ \approx 90^\circ$
Union pour un Mouvement Populaire	199	$\dfrac{199}{577} \times 180^\circ \approx 62^\circ$
Union des démocrates et indépendants	30	$\dfrac{30}{577} \times 180^\circ \approx 10^\circ$
Écologiste	18	$\dfrac{18}{577} \times 180^\circ \approx 6^\circ$
Radical, républicain, démocrate et progressiste	17	$\dfrac{17}{577} \times 180^\circ \approx 5^\circ$
Gauche démocrate et républicaine	15	$\dfrac{15}{577} \times 180^\circ \approx 5^\circ$
Députés non inscrits	8	$\dfrac{8}{577} \times 180^\circ \approx 2^\circ$
Total	**577**	**180°**

Exemples (suite)

On peut alors représenter cette répartition à l'aide d'un diagramme semi-circulaire (dont les angles ont été calculés et mis dans la dernière colonne du tableau précédent) :

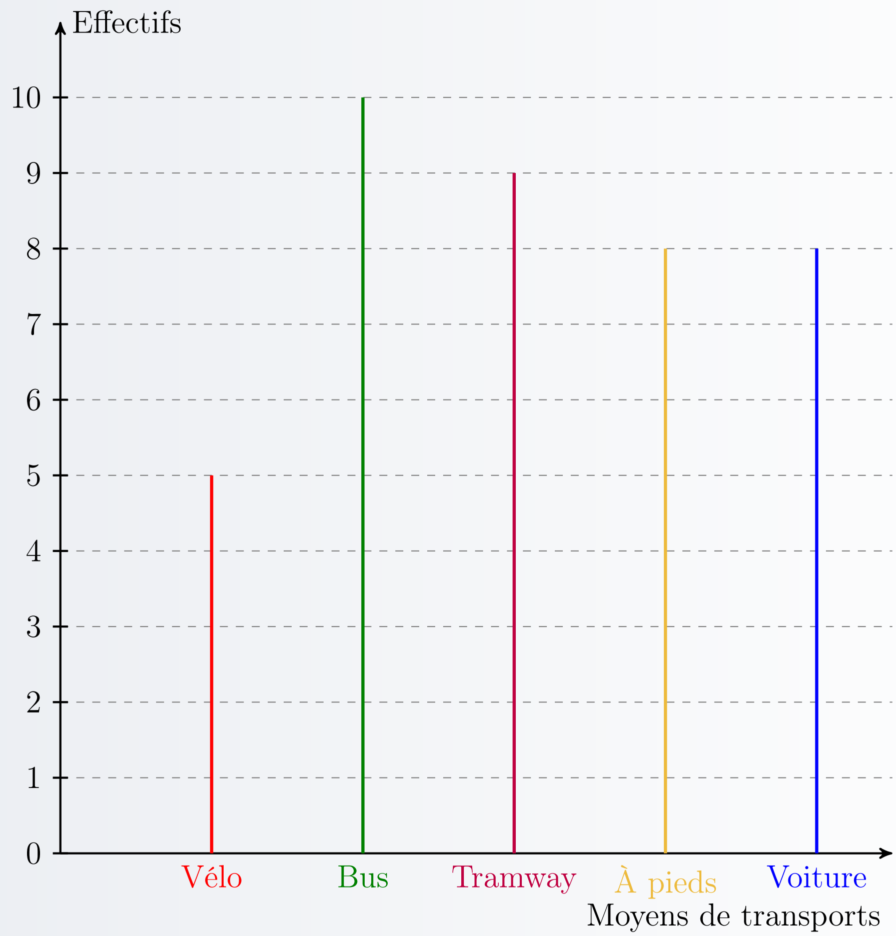


Répartition des députés (selon leur groupe) à l'Assemblée Nationale au 08/09/2014

- Nous avons interrogé 40 élèves d'une classe de Seconde pour savoir le moyen de locomotion qu'ils utilisaient le matin pour venir au lycée. Les résultats sont présentés dans le tableau suivant :

Moyens de locomotion	Vélo	Bus	Tramway	À pieds	Voiture
Effectifs	5	10	9	8	8

Pour ce genre de tableau, on peut utiliser un diagramme en bâtons :



Remarque

Notez que pour le premier exemple, afin de calculer les angles correspondants à chaque groupe, nous avons multiplié 180° (l'angle qui représente le total) par une fraction : cette fraction est la *fréquence* de l'effectif considéré.

Définition *Fréquence (rappel)*

La `fréquence` d'un caractère est donnée par la formule :

$$f = \frac{\text{effectif du caractère}}{\text{effectif total}}.$$

Définitions *Types de caractères quantitatifs*

Parmi les caractères quantitatifs, on peut (et on doit) dissocier :

- les caractères quantitatifs `discrets` : ceux que l'on peut énumérer un à un ;
- les caractères quantitatifs `continus` : ceux qu'il est impossible d'énumérer un à un. Ces caractères sont appelés des `classes`.

Exemples

- Lorsque l'on regarde le nombre de stylos à encre bleue dans la trousse des élèves d'une classe de Seconde, les caractères sont discrets (0, 1, 2, ...).
- Quand on regarde le salaire des salariés d'une grande société, il est impossible de tous les énumérer ; ce qui nous intéresse est de les regrouper par classes (par exemple, les salaires inférieurs à 1 200 €, ceux compris entre 1 200 € et 1 500 €, ...). Ces classes seront donc des intervalles : $[0\,;1\,200[$, $[1200\,;1250[$, ...

Méthode *Représentation d'une série statistique à caractères quantitatifs*

- On peut représenter les séries statistiques à caractères quantitatifs discrets par un diagramme en bâtons ou un diagramme circulaire ou semi-circulaire ;
- On peut représenter les séries statistiques à caractères quantitatifs continus par un histogramme. Dans ce cas, c'est l'aire des rectangles qui doit être proportionnelle à l'effectif (et non leur hauteur). Si l'amplitude des classes (la base des rectangles) est la même, la hauteur est aussi proportionnelle à l'effectif.

Exemple

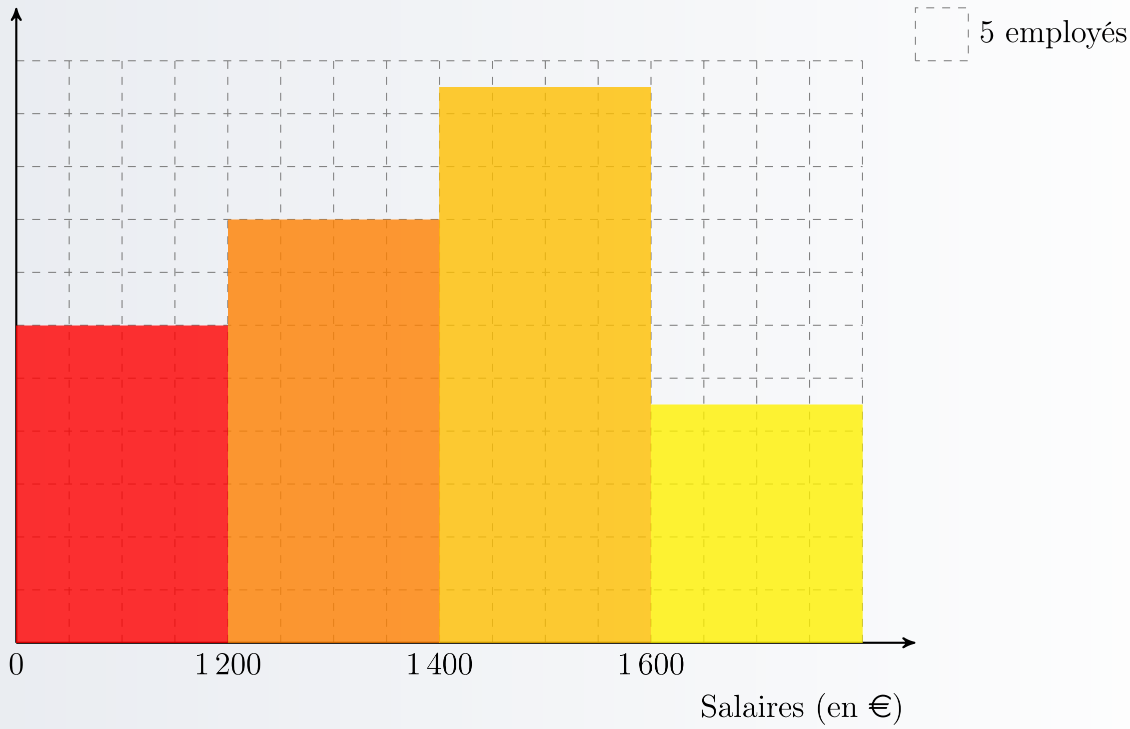
Dans une société quelconque, on considère la répartition des salaires suivante :

Salaires (en €)	$\leqslant 1\,200$	$]1\,200\,;1\,400]$	$]1\,400\,;1\,600]$	$> 1\,600$
Effectifs	120	150	210	90

...

Exemple (suite)

On peut alors représenter cette répartition de la façon suivante :



Définition

Moyenne

- Dans le cas d'une série statistique à caractères quantitatifs discrets donnés par un tableau comme le suivant :

Caractères (x_i)	x_1	x_2	$\cdots$	x_p
Effectifs (n_i)	n_1	n_2	$\cdots$	n_p

  la moyenne de la série est donnée par la formule :

$$\overline{x} = \frac{n_1 \times x_1 + n_2 \times x_2 + \cdots + n_p \times x_p}{n_1 + n_2 + \cdots + n_p}.$$

- Dans le cas d'une série statistique à caractères quantitatifs continus donnés par un tableau comme le suivant :

Classes	$[a_1\,;a_2[$	$[a_2\,;a_3[$	$\cdots$	$[a_p\,;a_{p+1}]$
Centres des classes	c_1	c_2	$\cdots$	c_p
Effectifs (n_i)	n_1	n_2	$\cdots$	n_p

  la moyenne de la série est donnée par la formule :

$$\overline{x} = \frac{n_1 \times c_1 + n_2 \times c_2 + \cdots + n_p \times c_p}{n_1 + n_2 + \cdots + n_p}.$$

Exemples

- On considère la série statistique représentée par le tableau suivant :

Caractères (x_i)	0	1	2	3
Effectifs (n_i)	5	7	6	9

Sa moyenne est alors :

$$\overline{x} = \frac{5 \times 0 + 7 \times 1 + 6 \times 2 + 9 \times 3}{5+7+6+9} = \frac{46}{27} \approx 1,7.$$

- On considère la série statistique représentée par le tableau suivant :

Classes	$[0\,;5[$	$[5\,;10[$	$[10\,;15[$	$[15\,;20]$
Centres des classes	2,5	7,5	12,5	17,5
Effectifs (n_i)	5	7	6	9

Sa moyenne est alors :

$$\overline{x} = \frac{5 \times 2,5 + 7 \times 7,5 + 6 \times 12,5 + 9 \times 17,5}{5+7+6+9} = \frac{297,5}{27} \approx 11.$$

Définition

Médiane

La `médiane` d'une série statistique est la valeur m_e du caractère qui sépare la série en deux séries de mêmes effectifs.

Exemples

- **Cas d'une série statistique à caractères quantitatifs discrets.**
On considère la série donnée par le tableau suivant :

Caractères (x_i)	0	1	2	3
Effectifs (n_i)	5	7	6	9

Nous allons ajouter une ligne où l'on mettra les `effectifs cumulés croissants` (e.c.c.) :

Caractères (x_i)	0	1	2	3
Effectifs (n_i)	5	7	6	9
E.c.c.	**5**	$5+7=\mathbf{12}$	$12+6=\mathbf{18}$	$18+9=\mathbf{27}$

L'effectif total de cette série est donc 27 (le dernier e.c.c.). Or, $27 \div 2 = 13,5$, donc la 14e valeur des caractères sera la médiane : il s'agit ici de « 2 » car sur la ligne des e.c.c., on voit que l'on atteint 14 pour « 2 ». Donc $m_e = 2$.

Exemples (suite)

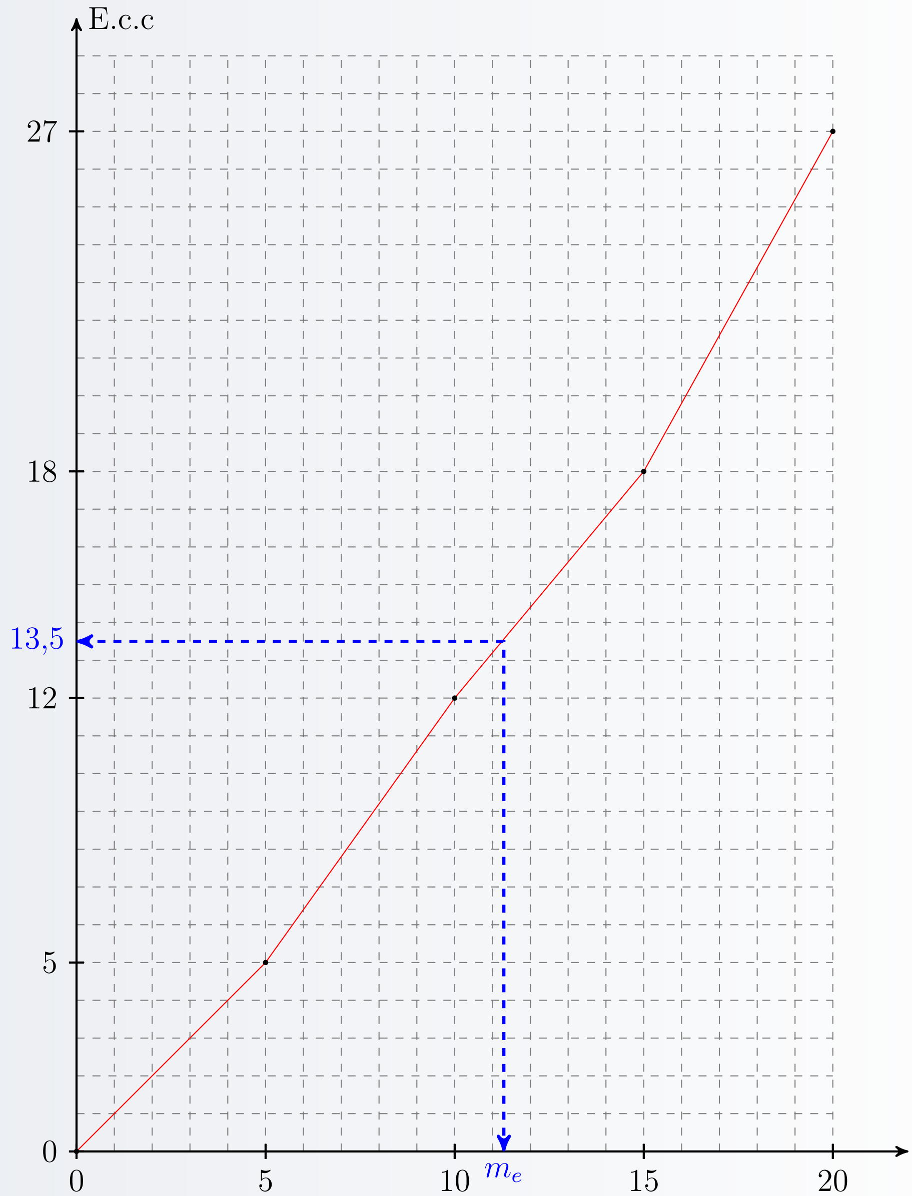
- **Cas d'une série statistique à caractères quantitatifs continus.**
  On considère la série donnée par le tableau suivant :

Classes	$[0\,;5[$	$[5\,;10[$	$[10\,;15[$	$[15\,;20]$
Effectifs (n_i)	5	7	6	9

On ajoute la ligne des e.c.c. :

Classes	$[0\,;5[$	$[5\,;10[$	$[10\,;15[$	$[15\,;20]$
Effectifs (n_i)	5	7	6	9
E.c.c.	$\mathbf{5}$	$5+7=\mathbf{12}$	$12+6=\mathbf{18}$	$18+9=\mathbf{27}$

On construit alors la « courbe brisée » suivante et on trouve m_e comme indiqué en pointillés bleus en plaçant 50 % de l'effectif total sur l'axe des ordonnées :



Méthode

Dans le dernier exemple, pour trouver numériquement la médiane, on trouve d'abord l'équation du segment dans la classe où est la médiane (ici, sur $[10\,;15[$) :

$$y = mx + p \qquad m = \frac{18-12}{15-10} = \frac{6}{5} \quad ; \quad p = 18 - \frac{6}{5} \times 15 = 18 - 18 = 0.$$

Ainsi, l'équation du segment est $y = \frac{6}{5}x$ sur $[10\,;15[$.

Maintenant, on remplace y par $13,5$:

$$13,5 = \frac{6}{5}x \Longrightarrow x = 13,5 \times \frac{5}{6} = 11,25.$$

La médiane est donc $m_e = 11,25$, ce qui correspond bien à ce que l'on peut à peu près lire sur le graphique.

Définitions

Quartiles

- Le premier quartile, noté Q_1, d'une série statistique est la valeur du caractère pour laquelle au moins 25 % des valeurs sont inférieures ou égales à Q_1 ;
- Le troisième quartile, noté Q_3, d'une série statistique est la valeur du caractère pour laquelle au moins 75 % des valeurs sont inférieures ou égales à Q_3.

Exemple

On donne la répartition des notes à un contrôle dans une classe de 27 élèves par le tableau suivant :

Notes sur 20	7	8	9	10	11	12	13	14	15
Effectifs (n_i)	2	3	5	2	1	6	3	3	2
E.c.c.	2	5	10	12	13	19	22	25	27

- 25 % de 27 est égal à $\frac{25}{100} \times 27 = 6,75$.
  La valeur « 6,75 » est atteinte pour la note « 9 » ; par conséquent, $Q_1 = 9$;
- 75 % de 27 est égal à $\frac{75}{100} \times 27 = 20,25$.
  La valeur « 20,25 » est atteinte pour la note « 13 » ; par conséquent, $Q_3 = 13$.

Remarques

- Pour trouver la valeur de Q_1 et Q_3 dans le cas d'une série statistique à caractères quantitatifs continus, on procède de la même façon que pour la médiane.
- La construction de la « courbe brisée » peut aussi se faire à partir des fréquences cumulées croissantes (f.c.c.). Cette « courbe brisée » s'appelle le `polygone des e.c.c.` ou `polygone des f.c.c.` selon ce que l'on prend pour la construire.