Chapitre 4 : Statistique -28-11-12-
Seconde 5, 2012-2013, Y. Angeli

1. Introduction et vocabulaire

La *statistique* est le domaine des mathématiques qui vise à interpréter des données et les interpréter, souvent dans le but de prendre une décision. On étudie un *caractère* x mesuré sur un ensemble d'*individus*. Pour résumer une série de nombres $x_1, ..., x_n$ (résultats d'une expérience, d'un sondage...), on donne souvent un couple de nombres : une *valeur centrale* et un *paramètre de dispersion*, qui permet d'évaluer à quel points les valeurs sont éloignées de la valeur centrale.

1.1. Vocabulaire

Définition 1. Le vocabulaire de base de la statistique est résumé ici :
• L'ensemble des éléments étudiés est appelé la *population* (même s'il ne s'agit pas de gens !).
• Les différents éléments de cet ensemble sont appelés les *individus*.
• L'*effectif total* est le nombre N d'individus de la population.
• Le *caractère* ou la *variable statistique* x désignent la grandeur étudiée dans la population. Les valeurs du caractères pour les individus $1, 2, ..., N$ sont notées $x_1, x_2, \dots, x_N$. Plusieurs des x_i peuvent être identiques. (plusieurs individus peuvent avoir le même caractère)
• Les *modalités* $a_1, \dots, a_k$ sont différentes valeurs prises par le caractère x (donc pas de répétition).
• L'*effectif* n_j d'une modalité a_j représente le nombre d'individus de la population dont le caractère vaut a_j. (c'est-à-dire le nombre de fois que a_j apparaît parmi les x_i).
En particulier : $N = n_1 + \dots + n_k$.
• La *fréquence* f_j d'une modalité a_j est le quotient $\frac{n_j}{N}$ (effectif divisé par effectif total). En multipliant f_j par 100 on obtient le pourcentage de la population dont le caractère vaut a_j.

1.2. Le mode et l'étendue : une valeur centrale et un paramètre de dispersion

Définition 2. Le mode et l'étendue sont définis par :
• Le *mode* d'une série est la modalité la plus représentée, ou la plus fréquente.
• L'*étendue* est la différence entre la plus grande et la plus petite modalité.

Remarque 1. Le seul intérêt de ces données est d'être simple à comprendre et à calculer. Elles sont cependant trop naïves pour permettre de conclure la plupart des études.

Exemple 1. Soit la série de notes de 10 élèves :

Élève n ^{o}	1	2	3	4	5	6	7	8	9	10
Note x_i	4	7	7	9	9	11	11	11	15	18

La population étudiée est ....

Un individu de cette population est ....

Cette série statistique a un effectif total de .... . Remplir le tableau des modalités :

Note a_j						
Effectif n_j						
Fréquence f_j						

Le mode de cette série est ...., son étendue est ....

2. La moyenne : un autre exemple de valeur centrale.

Définition 3. La *moyenne* $\bar{x}$ des caractères x_i d'une série statistique est :

$$\bar{x} = \frac{x_1 + ... + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{n_1 a_1 + ... + n_k a_k}{N} = \frac{1}{N}\sum_{j=1}^{k} n_j a_j$$

Exemple 2. Dans l'exemple 1, $\bar{x} =$ ....

L'avantage de la moyenne réside dans le fait qu'elle est linéaire : on peut calculer la moyenne de deux groupes dont on connaît les effectifs et les moyennes :

Propriété 1.

La moyenne des images des valeurs du caractère par une fonction affine du type $f(x) = ax + b$ est l'image de la moyenne $f(\bar{x}) = a\bar{x} + b$.
Si la série z est composée de deux séries x et y d'effectifs respectifs p et q, on a : $\bar{z} = \dfrac{p\bar{x} + q\bar{y}}{p + q}$

Preuve. On a : $\dfrac{(ax_1 + b) + ... + (ax_N + b)}{N} = \dfrac{a(x_1 + ... + x_N) + Nb}{N} = a\bar{x} + b.$

• $\bar{z} = \dfrac{x_1 + ... + x_p + y_1 + ... + y_q}{p + q} = \dfrac{p\frac{x_1 + ... + x_p}{p} + q\frac{y_1 + ... + y_q}{q}}{p + q} = \dfrac{p\bar{x} + q\bar{y}}{p + q}$ □

Exemple 3. On suppose qu'au même contrôle, un groupe de 20 élèves a obtenue une moyenne de $\bar{y} = 11,4$. Calculer la moyenne des 30 élèves (ces 20 plus les 10 de l'exemple 1) :

$\bar{z} =$ ....

L'inconvéniant de la moyenne est qu'elle est sensible aux valeurs extrêmes :

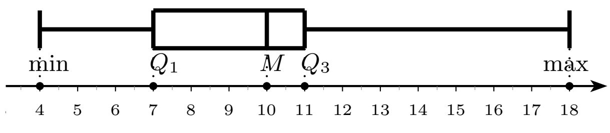
Exemple 4. Une entreprise compte 99 salariés gagnant 1 200 € par mois et et un dirigeant gagnant 81 200 € par mois. Quel est le salaire moyen ? Est-il représentatif de la situation ?

....

3. La médiane : un troisième exemple de valeur centrale

Définition 4. On suppose les x_i rangés dans l'ordre croissant :
• Une *médiane* M est telle que 50% des caractères sont au dessus de M et 50% en dessous.
Si N est impair, la médiane est la valeur du milieu.
Si N est pair, une médiane est la moyenne des deux valeurs du milieu.
• Le *premier quartile* Q_1 est une médiane de la série des x_i où $i > \frac{N+1}{2}$.
• Le *troisième quartile* Q_3 est une médiane de la série des x_i où $i < \frac{N+1}{2}$.
• L'*intervalle interquartile* est $]Q_1, Q_3[$.
• L'*écart interquartile* est $Q_3 - Q_1$. C'est le paramètre de dispersion associé à la médiane.

Exemple 5. Dans l'exemple 1 :
4 7 7 9 9 11 11 11 14 18
Med= ... ; $Q_3 =$ ... ; $Q_1 =$ ... ; $Q_3 - Q_1 =$ ...



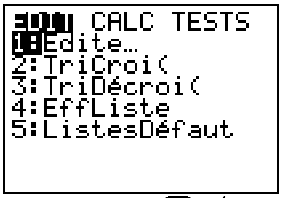
L'inconvéniant de la médiane est qu'elle n'est pas linéaire. Son intérêt est d'être peu sensible aux valeurs extrêmes :

Exemple 6. Dans l'exemple 4, le salaire médian (la médiane des salaires) est ....

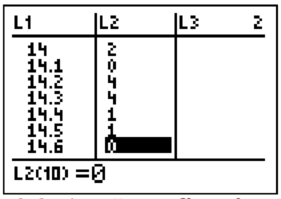
4. Utiliser la calculatrice

Exemple 7. Durant quatre cours de mathématiques, un professeur a fait tracer à chacun de ses élèves un carré de 10 cm de côté et mesurer la diagonale. Ces mesures sont consignées ici :

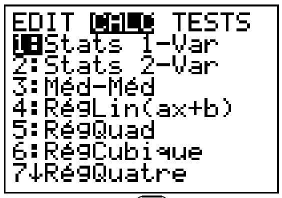
Modalité	13,7	13,8	13,9	14	14,1	14,2	14,3	14,4	14,5	14,6	N	Mode	Éte.	Moy.	Méd.
2nde1	1	1	2	2	0	4	4	1	1	0	16		0,8	14,14	14,2
2nde2	2	2	5	2	4	0	1	1	1	0					
2nde3	0	2	5	0	2	4	3	6	2	2					
2nde4	0	0	3	1	3	2	4	3	0	0					



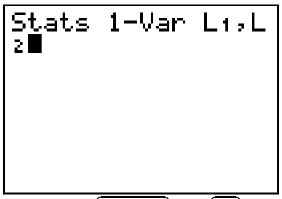
STAT [edit] et 1 (Éditer)



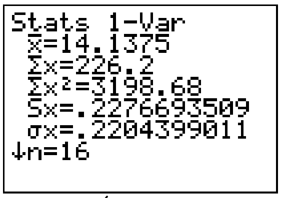
modalités→L1 ; effectifs→L2



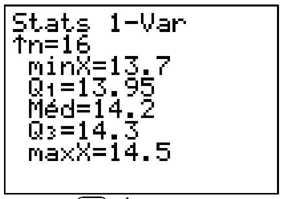
STAT [calc] et 1 (Stats 1-var)



L_1 : 2NDE et 1



Écran 1

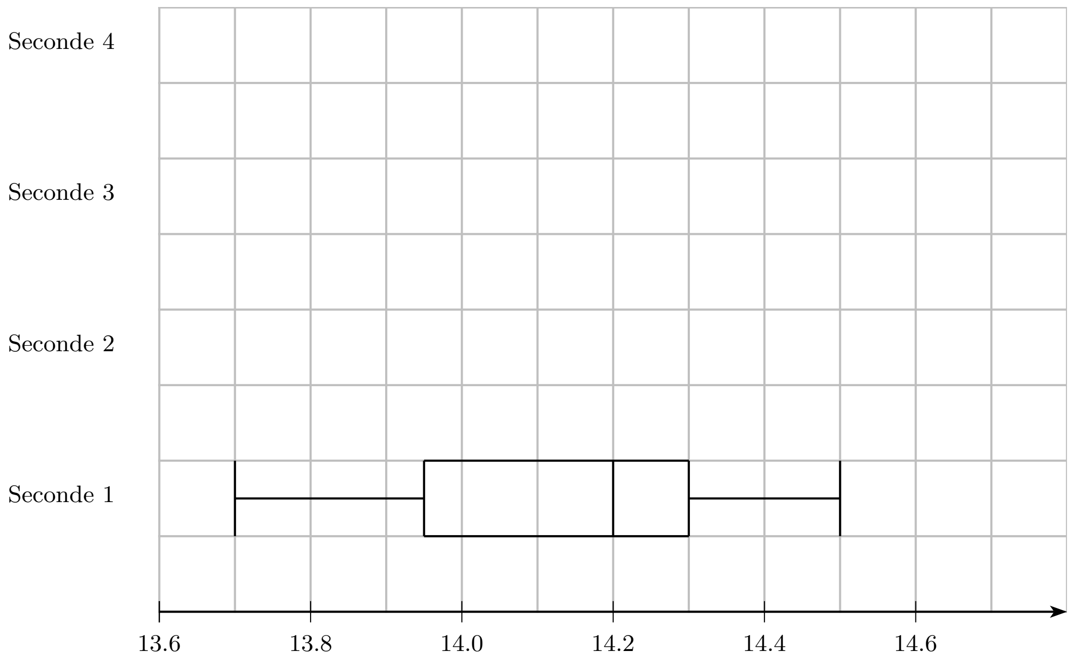


↓ Écran 2

Exemple 8. Remplir le tableau de l'exemple 7 à la calculatrice.

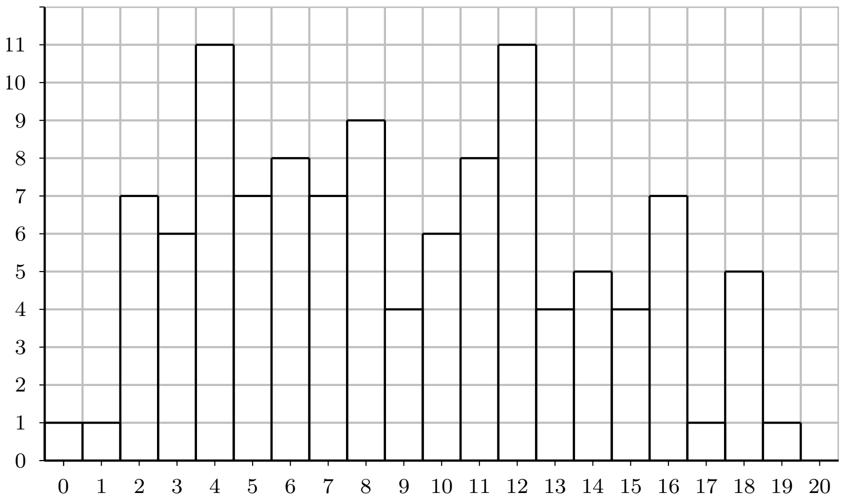
La moyenne des mesures des quatre classes est : ....

Représenter le diagramme en boîte des secondes 2,3 et 4 :



5. Exemple de série donnée par un histogramme

L'histogramme ci-dessous représente les notes obtenues en mathématiques lors du brevet par les élèves d'un collège. La hauteur de chaque barre représente le nombre d'élèves ayant obtenus la note correspondante.



Exemple 9. Remplir le tableau et en déduires les données statistique qui suivent.

Note	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Effectif	1	1	7																		
Effectif cumulé	1	2	9																		

Min= .... Q_1= .... Med= .... Q_3= .... Max= .... Mode= .... $\bar{x}$= ....

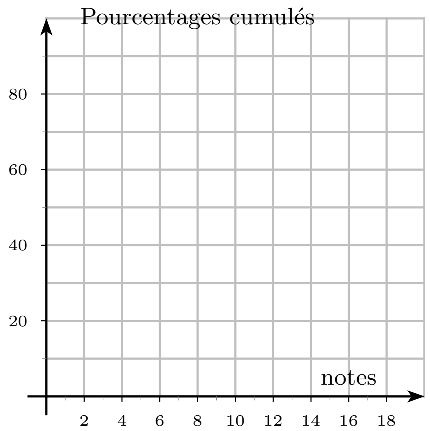
6. Exemple de série donnée par classes

Au niveau départemental, les résultats obtenus sont les suivants :

Notes	[0 ;4[	[4 ;8[	[8 ;12[	[12 ;16[	[16 ;20]
Pourcentage des candidats	25	20	35	15	5
Pourcentage cumulés					

Exemple 10. Représenter les points dont les abscisses x sont les bornes supérieures de chacun des intervalle, et les ordonnées y les pourcentage cumulés correspondants.

On suppose que les notes sont *réparties* sur chaque intervalle. Relier les points obtenus et déterminer graphiquement la médiane et les quartiles de la série départementale.



Q_1 = ...

Med= ....

Q_3 = ...

$\bar{x}$ =..

....

Comparer les notes du collège (exemple 9) à celles obtenues dans le département :