Exercices de statistique descriptive : indicateurs de position et de dispersion pour des séries discrètes

Seconde M. Guery

Exercice 1 : (Correction)

Voici les longueurs, en kilomètres, de chacune des vingt étapes du Tour de France 2009.

187 ; 196 ; 39 ; 196 ; 181 ; 224 ; 176 ; 160 ; 194 ; 192 ; 211 ; 200 ; 199 ; 207 ; 159 ; 169 ; 40 ; 178 ; 167 ; 164.

1. Déterminer la distance journalière moyenne parcourue par les cyclistes.
2. Ranger cette série statistique dans l'ordre croissant, puis déterminer sa médiane, ainsi que le premier et le troisième quartile.
   Interpréter ces résultats.
3. Un journaliste du journal Le Monde affirme que la moitié des épreuves mesurent moins de 181 kilomètres inclus de distance, tandis qu'un journaliste de l'Equipe écrit que la moitié des épreuves mesurent moins de 184 kilomètres. Ont-ils raison tous les deux ? Sinon, lequel a tort ? Justifier.

Exercice 2 : (Correction)

L'affluence en milliers de spectateurs, au stade Vélodrome, pour les matches de l'OM lors du championnat de football 2011/2012 est indiquée dans le tableau ci-après :

Affluence (en milliers)	45	46	48	49	50	52	55	56	58
Fréquence (en %)	11	5	5	11	11	5	21	26	5

1. Calculez la moyenne, la médiane, les quartiles et l'écart inter-quartile à partir des fréquences.
2. La capacité maximale du stade est de 60 000 spectateurs. Mais, en raison de travaux d'entretien et de mesures de sécurité, on estime qu'à partir de 55 000 spectateurs, le stade est "plein".
   L'affluence moyenne est-elle un bon indicateur pour savoir si le stade est souvent plein ? Sinon, lequel semble intéressant ?

Exercice 3 : (Correction)

L'enneigement de la station de sport d'hiver de l'Alpe d'Huez durant la saison de ski 2008 est indiqué par la hauteur de neige moyenne, exprimée en cm, relevée chaque semaine.

Hauteur	50	100	120	130	140	160	180	200	240	260
Nombre de semaines	1	2	1	1	1	6	1	3	3	3

1. Recopier et compléter le tableau avec les effectifs cumulés croissants.
2. Calculer la moyenne, la médiane ainsi que les quartiles de cette série (arrondir au cm près).
3. Pour la pratique du ski dans les meilleures conditions, la hauteur de neige doit dépasser 140 cm. Quel indicateur permet de préciser la durée favorable ?
4. Est-il exact que durant le quart de la saison, l'enneigement a été exceptionnel (hauteur de plus de 2 m) ?

Exercice 4 : (Correction)

Un professeur de mathématiques à la retraite note quotidiennement le nombre de lettres qu'il reçoit chez lui (il n'a vraiment rien d'autre à faire....). Au cours du mois de février d'une année non bissextile (donc le mois est composé uniquement de 28 jours), voici ce qu'il a noté :

1 0 0 2 3 1 0 5 0 2 0 2 1 1 0 0 0 4 2 1 1 2 0 0 1 2 0 1

1. Ranger cette série dans l'ordre croissant.
2. Recopier et compléter le tableau suivant :

Nombre de lettres	0	1	2	3	4	5
Effectif						
E.c.c						

3. Représenter cette série par un nuage de points, puis par un diagramme en bâtons.
4. Déterminer le nombre moyen de lettres qu'a reçus le retraité. Arrondir à $0,1$ près.
5. Déterminer la médiane de cette série. Que peut-on interpréter ?
6. Déterminer le premier et le troisième quartile, puis compléter cette phrase : "Les trois quarts du temps, le retraité reçoit plus de .... lettres par jour."
7. Question de logique : l'enseignant à la retraite ne se souvient plus si le mois de février a commencé un vendredi ou un samedi (il perd un peu la tête). Pouvez-vous l'aider ?

Correction exercice 1 : (Énoncé)

1. On calcule la moyenne :

$$\begin{aligned}\overline{x} &= \frac{187+196+39+196+181+224+176+160+194+192+211+200+199+207+159+169+40+178+\ldots}{20}\\ &= \frac{3439}{20}\\ &= 171,95.\end{aligned}$$

   La distance moyenne parcourue quotidiennement par les concurrents est de $171,95$ kilomètres.

2. Rangeons cette série dans l'ordre croissant :

   $$39\,;\,40\,;\,159\,;\,160\,;\,164\,;\,167\,;\,169\,;\,176\,;\,178\,;\,181\,;\,187\,;\,192\,;\,194\,;\,196\,;\,196\,;\,199\,;\,200\,;\,207\,;\,211\,;\,224.$$

   Déterminons la médiane : $\frac{N}{2}=\frac{20}{2}=10$, donc la médiane est la moyenne de la 10ème et de la 11ème valeur de la série, soit $Me=\frac{181+187}{2}=184$.
   Déterminons le premier quartile : $\frac{N}{4}=\frac{20}{4}=5$, donc le premier quartile est la cinquième valeur de la série, soit $Q_1=164$.
   Déterminons le troisième quartile : $\frac{3N}{4}=3\times\frac{N}{4}=3\times 5=15$, donc le troisième quartile est la quinzième donnée de la série, soit $Q_3=196$.
   On en déduit que la moitié des épreuves mesurent moins de 184 kilomètres, le quart des épreuves mesurent moins de 164 kilomètres, et que le quart des épreuves mesurent plus de 196 kilomètres.

3. Les deux journalistes ont raison : le second a déterminé la médiane, quant au premier, il a regardé la dixième épreuve, qu'il a incluse dans son raisonnement. En fait, le premier journaliste a déterminé le second quartile !
   En effet, pour déterminer le second quartile, on calcule $\frac{N}{2}$, et on raisonne comme les deux autres quartiles : si on obtient un entier, on prend la donnée dont la position correspond à cet entier. On ne fait pas la moyenne avec la donnée suivante, comme pour la médiane.

Correction exercice 2 : (Énoncé) *Pour des raisons évidentes de santé, je ne me suis pas permis de travailler avec le stade de Gerland. Le médecin m'a interdit de parler de l'OL à cause d'un eczéma persistant.*

1. Calculons la moyenne :

$$\begin{aligned}m &= 45\times\frac{11}{100}+46\times\frac{5}{100}+48\times\frac{5}{100}+49\times\frac{11}{100}+50\times\frac{11}{100}+52\times\frac{5}{100}+55\times\frac{21}{100}+56\times\frac{26}{100}+58\times\frac{5}{100}\\ &= 45\times 0,11+46\times 0,05+48\times 0,05+49\times 0,11+50\times 0,11+52\times 0,05+55\times 0,21+56\times 0,26+58\times 0,05\\ m &= 52,15.\end{aligned}$$

   L'affluence moyenne dans le stade est de 52 150 spectateurs.
   Calculons la médiane : pour cela, déterminons les fréquences cumulées croissantes (ce qui nous sera utile par ailleurs pour trouver les quartiles) :

Affluence (en milliers)	45	46	48	49	50	52	55	56	58
Fréquence (en %)	11	5	5	11	11	5	21	26	5
F.c.c (en %)	11	16	21	32	43	48	69	95	100

   Ainsi, la médiane est égale à 55. De plus, le premier quartile est égal à 49 et le troisième est égal à 56. L'écart inter-quartile est donc égal à 7.

2. La moyenne, en milliers, est inférieure à 55. On pourrait donc avoir tendance, si l'on va trop vite, à penser que le stade n'est pas souvent plein. Or, la médiane étant de 55 justement, on peut conclure que pour plus de la moitié des matches, l'affluence est supérieure ou égale à 55 000 spectateurs. Donc le stade est souvent "plein" (plus d'une fois sur deux !). Donc la moyenne n'est pas un bon indicateur. Il vaut mieux se fier à la médiane.
   Pour être plus précis, pour 52% des matches l'affluence était supérieure ou égale à 55 000 spectateurs.

<u>Correction exercice 3</u> : (Énoncé)

1.

Hauteur	50	100	120	130	140	160	180	200	240	260
Nombre de semaines	1	2	1	1	1	6	1	3	3	3
Effectifs cumulés croissants	1	3	4	5	6	12	13	16	19	22

2. On détermine la moyenne :

$$\begin{aligned}\overline{x} &= \frac{50 \times 1 + 100 \times 2 + 120 \times 1 + 130 \times 1 + 140 \times 1 + 160 \times 6 + 180 \times 1 + ...}{22} \\ &= \frac{3880}{22} \\ &\simeq 176.\end{aligned}$$

En moyenne, la hauteur de neige était de 176 cm.
Déterminons la médiane : $\frac{N}{2} = \frac{22}{2} = 11$, donc la médiane est la moyenne de la onzième et de la douzième valeur de la série, donc en s'aidant de la ligne des effectifs cumulés croissants, on trouve $Me = \frac{160 + 160}{2} = 160$.
Déterminons le premier quartile : $\frac{N}{4} = \frac{22}{4} = 5,5$, donc le premier quartile est la sixième valeur de la série. Donc toujours en s'aidant de la ligne des e.c.c, on trouve $Q_1 = 140$.
Déterminons le troisième quartile : $\frac{3N}{4} = 3 \times 5,5 = 16,5$, donc le troisième quartile est la dix-septième valeur de la série. Donc toujours en s'aidant de la ligne des e.c.c, on trouve $Q_3 = 240$.

3. Comme $Q_1 = 140$, c'est le premier quartile qui précise la durée favorable au ski : il nous informe qu'au moins 75% du temps, la hauteur de neige dépasse les 140 cm, donc que les conditions sont favorables.

4. On peut raisonner de deux manières : d'après le tableau, on a que l'enneigement a été exceptionnel durant neuf semaines (hauteur de plus de $2m$, $2m$ inclus). Donc $\frac{9}{22} \simeq 0,41 \geq 0,25$, donc effectivement durant plus d'un quart de la saison, l'enneigement a été exceptionnel.
La seconde méthode, plus rapide, sans calcul, consiste à regarder Q_3 : on déduit que durant plus du quart de la saison, l'enneigement était d'au moins 240 cm, donc bien supérieur à $2m$. Donc durant le quart de la saison, l'enneigement a été exceptionnel.

Correction exercice 4 : (Énoncé)

1. *C'est un peu fastidieux, mais comme on est bien élevé, on va réaliser le travail demandé !*

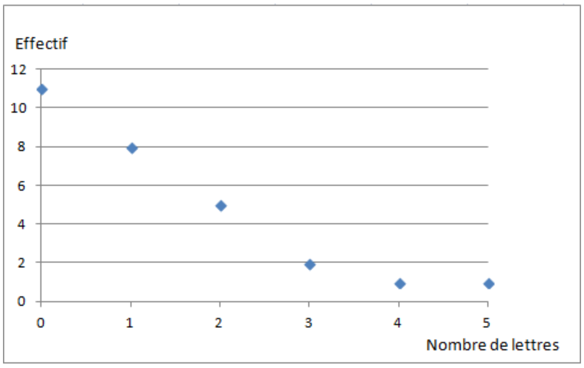
   0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 2 2 2 2 2 3 3 4 5

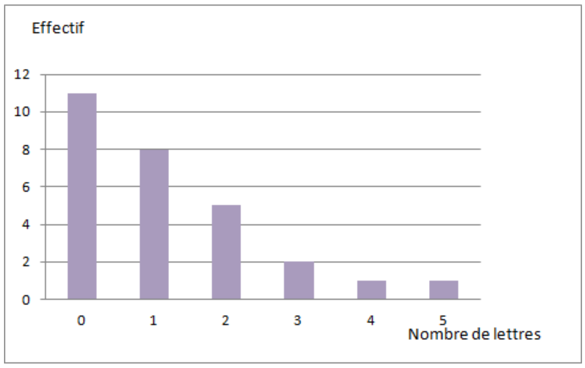
2.

Nombre de lettres	0	1	2	3	4	5
Effectif	11	8	5	2	1	1
E.c.c	11	19	24	26	27	28

   On rappelle que pour déterminer la ligne des effectifs cumulés croissant, on écrit dans chacune des cases la somme des valeurs de la ligne des effectifs des colonnes précédentes. Par exemple, pour la case dédiée au chiffre 1, on effectue $11+8$ *; pour la case dédiée au chiffre 2, on effectue* $11+8+5$ *etc.*

3.





4. Notons $\overline{x}$ la moyenne. On a $\overline{x} = \dfrac{0\times 11+1\times 8+2\times 5+3\times 2+4\times 1+5\times 1}{28} = \dfrac{34}{28} = \dfrac{17\times \cancel{2}}{14\times \cancel{2}} = \dfrac{17}{14} \simeq 1,2$.
   Ainsi, en moyenne, le retraité reçoit $1,2$ lettres par jour.
5. Ici, l'effectif total est 28. Or, $28 = 2\times 14$. Ainsi, la médiane est la demi-somme de la 14eme et de la 15eme valeur de la série rangée dans l'ordre. Ainsi, en s'aidant de la ligne des effectifs cumulés croissants, on a $Me = \dfrac{1+1}{2} = 1$.
   Donc la médiane de cette série est de 1. Cela signifie que la pendant la moitié du temps, le retraité reçoit moins d'une lettre, et que l'autre moitié du temps, il reçoit plus d'une lettre par jour.
6. On a $N = 28$ et donc $\dfrac{N}{4} = 7$. Ainsi, le premier quartile de cette série est le 7eme terme de la série rangée dans l'ordre, c'est-à-dire que $Q_1 = 0$.
   De plus, $\dfrac{3N}{4} = \dfrac{3\times 28}{4} = \dfrac{3\times 7\times \cancel{4}}{4} = 21$, donc le troisième quartile de cette série est le 21eme terme, c'est-à-dire $Q_3 = 2$.
   Ainsi, "Les trois quarts du temps, le retraité reçoit moins de deux lettres par jour."
7. Cette question n'a rien à voir avec les statistiques ! C'est juste de la logique. Souvenons-nous que les facteurs ne travaillent pas le dimanche (il faut qu'ils se reposent tout de même ! Et puis le dimanche matin il y a Télé Foot...). Donc le dimanche, le retraité ne reçoit pas de courrier.
   Supposons que le mois de février ait commencé un vendredi. Le troisième jour est donc un dimanche, c'est la raison pour laquelle la troisième valeur de la série est 0. Mais sept jours plus tard, nous retombons sur un dimanche. Ainsi, la dixième valeur de la série devrait être un zéro également. Or, le dixième terme de la série est un 2, ce qui est **absurde**. Donc le mois de février **ne peut pas** commencer par un vendredi.
   D'après les hypothèses de la question, le mois a donc **nécessairement** débuté un samedi, ce qui est cohérent d'ailleurs avec tous les termes de la série : le deuxième terme est un zéro, tout comme le neuvième et le seizième.
   Remarque : pour les plus curieux, on pourrait se demander si l'hypothèse du retraité était cohérente, à savoir que le premier jour du mois était nécessairement un vendredi ou un samedi. Dans les sept premiers jours, il n'y a que trois jours où le retraité n'a pas reçu de courrier. Les deux premiers jours sont les cas étudiés précédemment. Mais si le mois commençait par un lundi, on aurait bien que le septième jour serait un dimanche, et on tomberait bien sur "0". Or, le quatorzième jour, qui devrait être également un dimanche, le retraité reçoit une lettre, ce qui n'est pas possible...
   Tout ça pour dire que le mois de février a bel et bien commencé par un samedi !

   Pour les plus téméraires, sachant que le 01/01/2013 est un mardi, ***sans utiliser de calendrier*** *(ou d'Iphone et tout le tralala !), essayez de trouver l'année la plus récente pour laquelle le mois de février ait commencé par un samedi. C'est tout de même de niveau terminale S, spécialité mathématiques... Mais vous pouvez essayer !*