

Correction Exercices supplémentaires TD2

Questions de réflexion

1/ Pour une variable nominale, seul le mode peut être utilisé, les modalités de la variable n'étant pas ordonnées.

2/ Pour les variables quantitatives, la médiane ou la moyenne sont davantage utilisées que le mode, car elles donnent en général une meilleure estimation de la tendance centrale. Cependant, n'oubliez pas que lorsque le choix est possible, l'utilisation des paramètres de tendance centrale dépend de la forme des données.

Par exemple, si votre distribution comprend des données se situant majoritairement aux extrêmes (voir l'exercice 2 du TD2), l'utilisation du mode est préférable car la moyenne et la médiane ne renverront qu'une valeur correspondant au centre de la distribution des données, ce qui ne sera pas représentatif de la distribution réelle des données, alors que l'utilisation du mode vous permettra de mettre en évidence que votre set de données est bimodal.

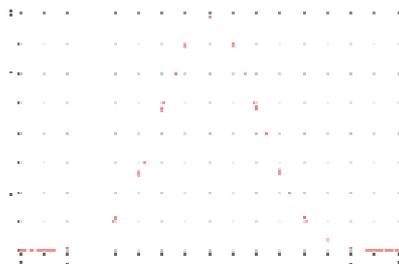
Autre exemple, si la distribution de vos données est relativement équilibrée entre les différentes modalités, ou si les données se situent majoritairement au centre des modalités, l'utilisation de la moyenne est souvent préférable, celle-ci prenant en compte l'ensemble des données, étant plus stable et manipulable algébriquement, contrairement à la médiane et au mode. Attention cependant à prendre en compte les valeurs aberrantes (voir l'exercice 3 du TD2) qui impacteront fortement votre moyenne et la rendront potentiellement non représentative des données ! Si vos données présentent des valeurs aberrantes extrêmes, il sera alors préférable d'utiliser la médiane, qui n'est pas influencée par les scores extrêmes.

3/ Même chose que pour la question 2.

4/ Si le mode, la médiane et la moyenne sont équivalents, cela signifie que la distribution des données est symétrique et unimodale. Dans ce cas, on va préférer utiliser la moyenne comme paramètre de tendance centrale, étant donné qu'elle prend en compte l'ensemble des données, qu'elle est plus stable et manipulable algébriquement.

D'autre part, le fait que la distribution des données soit symétrique signifie que même si la distribution comprend des valeurs aberrantes aux extrémités, ces valeurs seront présentes des deux côtés de la distribution, et n'impacteront donc pas le calcul de la moyenne puisqu'elles seront compensées.

La représentation la plus commune d'une distribution dont le mode, la médiane et la moyenne sont équivalentes est celle d'une loi normale, dont la représentation graphique est une courbe de Gauss ou « courbe en cloche » (la loi normale sera approfondie au second semestre).



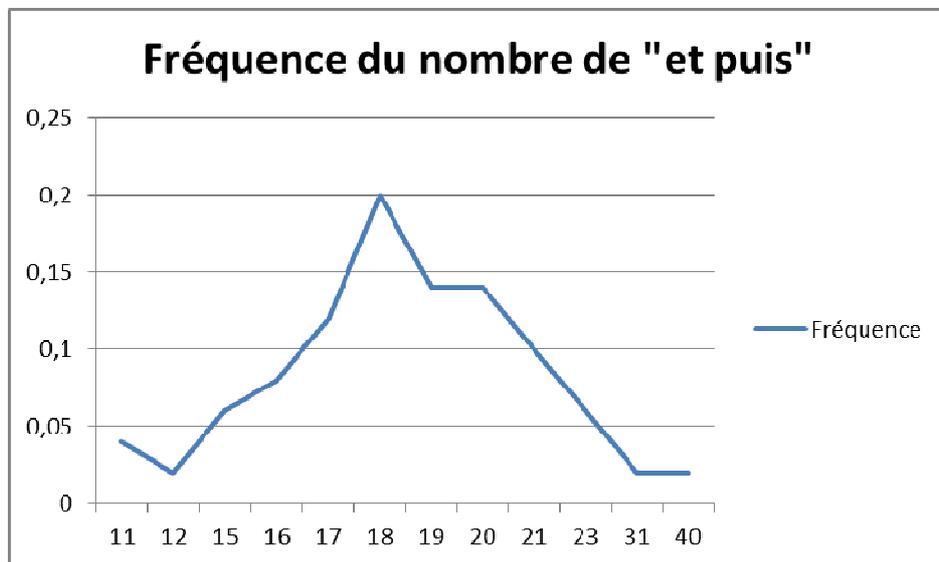
Exercices

A) 1/

Modalités	11	12	15	16	17	18	19	20	21	23	31	40
Effectif	2	1	3	4	6	10	7	7	5	3	1	1

2/ Pour calculer les fréquences, on divise l'effectif de chaque modalité par l'effectif total.

Modalités	11	12	15	16	17	18	19	20	21	23	31	40
Effectif	2	1	3	4	6	10	7	7	5	3	1	1
Fréquence	$\frac{2}{50} = 0.04$	0.02	0.06	0.08	0.12	0.2	0.14	0.14	0.1	0.06	0.02	0.02



3/ Mode = 18 (valeur ayant l'effectif ou la fréquence le plus élevé)

Médiane = 18. Nous avons 50 données. Donc la médiane correspond à la moyenne entre la $50/2 = 25^{\text{ème}}$ et la $(50/2) + 1 = 26^{\text{ème}}$ donnée. Les 25 et $26^{\text{ème}}$ données ont toutes les deux la valeur 18, la médiane est donc 18.

Moyenne =

$$[(2 \cdot 11) + (1 \cdot 12) + (3 \cdot 15) + (4 \cdot 16) + (6 \cdot 17) + (10 \cdot 18) + (7 \cdot 19) + (7 \cdot 20) + (5 \cdot 21) + (3 \cdot 23) + (1 \cdot 31) + (1 \cdot 40)] / 50 = 18.86$$

On constate que les trois paramètres de tendance centrale sont quasiment égaux, et que la distribution ressemble à une loi normale. Comme dit dans la question de réflexion 4, dans ce cas-là nous préférons la moyenne.

B) 1/

Modalités	1	2	3	4	5	6	7	8	9
Effectif	5	4	3	3	4	3	4	4	5

2/ Médiane = 5. Nous avons 35 données, donc la médiane correspond à la valeur de la $(35+1)/2 = 18^{\text{ème}}$ donnée, qui est de 5.

$$\text{Moyenne} = [(5*1)+(4*2)+(3*3)+(3*4)+(4*5)+(3*6)+(4*7)+(4*8)+(5*9)] / 35 = 5.06$$

3/

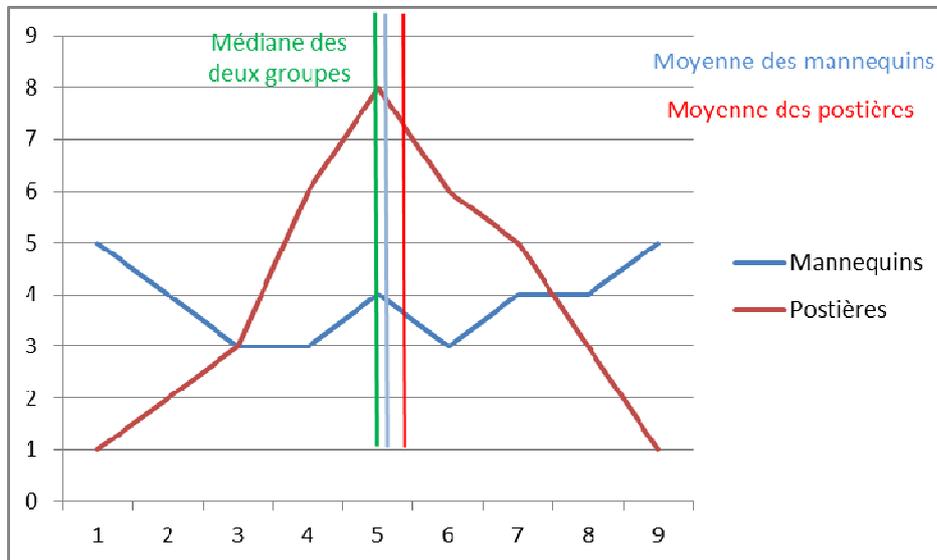
Modalités	1	2	3	4	5	6	7	8	9
Effectif	1	2	3	6	8	6	5	3	1

Médiane = 5. Nous avons 35 données, donc la médiane correspond à la valeur de la $(35+1)/2 = 18^{\text{ème}}$ donnée, qui est de 5.

$$\text{Moyenne} = [(1*1)+(2*2)+(3*3)+(6*4)+(8*5)+(6*6)+(5*7)+(3*8)+(1*9)] / 35 = 5.4$$

4/ On constate que les médianes des deux groupes sont égales, et que les moyennes sont très proches. Pour autant, on ne peut pas conclure à l'équivalence de ces deux groupes, car il faut tenir compte de la dispersion des données.

5/



6/ Il aurait mieux valu choisir le mode, ce qui nous aurait permis de mieux nous rendre des différences de distribution entre les deux groupes.

L
a

d
i
s
t
r