Simplifions la statistique

Par Louis Blais, stat., ASSQ

MESURES DE TENDANCE CENTRALE

Les valeurs des unités d'échan e observées ont souvent tendance à se concentrer autour d'une même valeur. Ce sont des mesures de tendance centrale. Les mesures les plus fréquemmen sont la moyenne arithmé la médiane et le mode d'un échan yenne géométrique et la moyenne quadra t aussi parfois xiste également une moyenne dite harmonique, disponible dans Excel 2013. La moyenne arithmé t ce que plusieurs tests sta seraient tr oire impossibles à réaliser avec les autres mesures, ce qui ne veut pas dire qu'elles sont dénuées d'intérêt, au contraire. Toutes ces mesures — y compris les mesures de dispersion[1] — servent à caractériser l'ensemble des valeurs d'une variable.

1. Voir la fiche *Simplifions la statistique – Mesures de dispersion* (Blais, 2016)

LA MOYENNE ARITHMÉTIQUE

C'est la mesure la plus fréquemmen la trouve aussi bien dans les rapports sta

Plusieurs décisions sont prises e mesure qui se calcule comme suit :

$$\overline{x} = \sum {}^{x_i}/_{n}$$

Pour des données qui sont mesurées, mais notées par classe (par exemple, le recouvrement), on suggère d' pour calculer la moyenne.

Facile à calculer, ce e moyenne a comme principal inconvénient d'être sensible aux données extrêmes, surtout si le nombre d'unités d'échan e est pe si les données extrêmes sont très nombreuses.

LA MÉDIANE

La médiane est la valeur du

se situant au-dessus d'elle, l'autr correspond au cen [1] 50. Si la dis est symétrique par rapport à la moyenne, la médiane et la moyenne, la médiane et la

Forêts, Faune et Parcs
Québec

moyenne auront des valeurs pra t égales. Dans le cas contraire, ces valeurs seron érentes, parfois même tr érentes. L'avantage de la médiane par rapport à la moyenne est que sa valeur est complètement indépendante des valeurs extrêmes qui auront été observées. La médiane peut êtr te que la moyenne lorsque le nombre d'observa t faible. En e et, dans ce cas, une ou deux observa xtrêmes peuvent grandement in yenne.

Un ex érence entre la moyenne et la médiane est le revenu annuel des Québécois. En juin 2013, le revenu annuel moyen avant impôt était de 51 000 $ alors que la médiane était de 42 400 $ (Sta e érence vient du fait que les salaires très élevés font augmenter la moyenne. Si on veut connaître le revenu total de tous les Québécois, la moyenne est la meilleure mesure. Par contre, pour obtenir le portrait réel du revenu des Québécois, la médiane est plus appropriée.

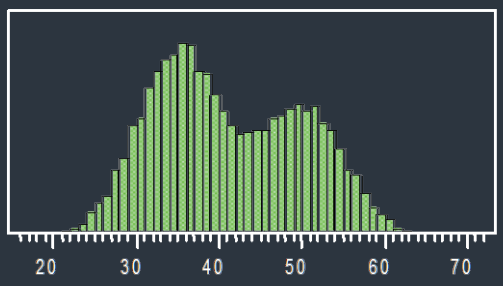
Dans Excel 2013, il existe deux f nomment « Cen e » et « Cen xclure ». La inclusion ou d'exclusion concerne l'inclusion de la valeur du 50e cen e) ou son exclusion (exclure). Si le nombre d'observa t très grand, l'inclure ou l'exclure n'aura pas d'in aura une s'il y a peu d'observa généralement moins de 30).

LE MODE D'UN ÉCHANTILLON

Pour des variables discrètes, dont les valeurs sont un code (par exemple, le drainage) ou une classe (par exemple, le recouvrement), le mode d'un échan est la valeur la plus fréquemment observée. Il peut ne pas être unique. Par exemple, lorsque le drainage a été noté et que les drainages « *rapide* » et « *modéré* » sont les plus souvent observés et reviennent un même nombre de fois, on aura alors deux modes pour le drainage. Ce n'est pas une sta yée fréquemment, mais, pour une variable discrète, elle permet d'avoir rapidement un aperçu de la valeur « centrale », c'est-à-dire celle qu'on trouve le plus fréquemment dans un échan

Pour les variables con le mode d'un échan t peu intéressant. Avec ce type de variable, il es t de déterminer la région du ou des pics de la dis ariable. Un histogramme de la dis serva permet de dire si on a a aire à une dis vec un seul pic (p. ex. : en forme de cloche). On dit alors que la dis t unimodale. L'histogramme peut également nous dire aussi s'il y a plusieurs pics dans la dis s que la dis es e ci-dessous). C'est une indica e variable peut avoir été mesurée sur plusieurs popula es ou selon plusieurs strat érentes. Une a en e doit être apportée si aucune autre variable notée ne permet de décrire les groupes que l'histogr e ci-contre, il y aurait deux groupes, par exemple, deux strates). Certains logiciels donnent toujours le mode, mais celui-ci ne cons oma t une sta alable.

Le problème, c'est qu' une même mesure soit un peu plus fréquente pour qu'elle devienne le mode, peu importe si elle est près du centre de la dis eut rien dire dans ce cas. Pour des variables con e vaut mille mots!



LA MOYENNE GÉOMÉTRIQUE

La moyenne géométrique est la racine n[e] du produit a aleurs prises par une variable. Elle se calcule comme suit :

$$\bar{x}_g = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{(x_1 * x_2 * \cdots * x_n)}$$

S'il y a deux données, on calcule la racine carrée de leur produit. S'il y en a dix, on calcule la racine 10[e] du produit des dix mesures. La moyenne géométrique est moins sensible aux données extrêmes et peut représenter une bonne es

Elle s'emploie occasionnellement lorsqu'on a une dis ymétrique, par exemple dans le cas des dénombrements où il y a une très grande variabilité ou dans celui des mesures environnementales sur les polluants atmosphériques. Pour pouv e moyenne, les valeurs doivent êtr es et non nulles. Très peu de logiciels calculent la moyenne géométrique. En e et, la racine n[e] d'un produit est en général fas alculer, même pour un ordinateur puissant. Excel e cependant une f alcule ce e moyenne (la f « Moyenne.geometrique »).

LA MOYENNE QUADRATIQUE

La moyenne quadra t dé omme la racine carrée de la moyenne arithmé arrés des valeurs des observa est-à-dire :

$$\bar{x}_Q = \sqrt{\left(\frac{\sum x_i^2}{n}\right)}$$

TABLEAU 1 DHP et surface terrière de 10 arbres dans une place e à ray e

Arbres	DHP (mm)	ST (m^2)	DHP (mm^2)
1	110	0,009503	12 100
2	250	0,049087	62 500
3	350	0,096211	122 500
4	140	0,015394	19 600
5	150	0,017671	22 500
6	200	0,031416	40 000
7	180	0,025447	32 400
8	400	0,125664	160 000
9	300	0,070686	90 000
10	120	0,011310	14 400
Sommes	2 200	0,452389	576 000
Moyenne	220		240

Ce type de moyenne es emen des variables reliées à des unités de surface. En foresterie, puisque la surface terrière a un lien direct avec le DHP, la moyenne quadra t plus appropriée. En e et, si vous calculez la moyenne quadra tres et que vous vous en servez pour calculer la surface terrière, vous allez obtenir exactement le même résultat que si vous calculiez la surface terrière par hectar e prise individuellement. L'exemple suivant l'illustre très bien. La moyenne arithmé diamètre est de 220 mm. La moyenne quadra de 240. La surface terrière est la somme de toutes les surfaces individuelles. La surface terrière totale des dix arbres donne 0,452389 m².

On calcule la surface terrière de la manière suivante :

$$ST = \sum \left(\pi(DHP_{mm}/2)\left[{}^{1m}/_{1000_{mm}}\right]\right)^2$$

$$= \sum \left(\pi(DHP_{mm}/2000)\right)^2$$

(On doit diviser le diamètre par deux, puisqu'une surface se calcule avec un rayon et non avec un diamètre, et conv tres en mètres en divisant par 1 000, d' e 2000).

tre quadra yen, on ob t :

$$10\pi(240/2000)^2 = 0{,}452389m^2$$

on a dix arbres, ce qui donne exactement le même résultat que dans le tableau 1.)

yenne arithmé alcul de la surface terrière, on ob t :

$$10\pi(220/2000)^2 = 0{,}38013m^2$$

Ce dernier calcul ne donne pas la bonne surface terrière.

Comme pour la moyenne géométrique, beaucoup de logiciels sta alculent pas ce e moyenne. On doit la programmer.

LA MOYENNE HARMONIQUE

La moyenne harmonique est raremen même s'il est possible de la calculer dans Excel 2013. On l'ob t de la manière suivante :

$$\bar{x}_h = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$$

Voici un exemple d'applica moyenne : si vous roulez sur une distance de 10 km à une vitesse de 40 km/h, puis que vous parcourez les 10 km suivants à 60 km/h, votre vitesse moyenne pour les 20 km sera de 48 km/h. La moyenne harmonique permet de calculer précisément la vitesse moyenne sur ce e distance. Il ne faut pas confondre avec la vitesse moyenne basée sur le temps. Si les kilomètres parcourus étaient remplacés par un temps parcouru (10 minutes plutôt que 10 km), votre vitesse moyenne serait alors de 50 km/h pendant 20 minutes, soit la moyenne arithmé esses.

CONCLUSION

En conclusion, chaque type de moyenne correspond à un besoin précis. Si la moy calculer une quan é totale, par exemple, un volume total, la moyenne arithmé t le meilleur choix. Elle est également plus appropriée lors d'un test de comparaison entre deux moyennes. Lorsqu'il y a peu d'unités d'échan e, ou que la dis n'est pas symétrique autour de la moyenne, la médiane permet alors de fournir un meilleur portrait des données que la moyenne arithmé Pour déterminer une surface terrière, le diamètre quadra yen du DHP es la place e est à ray e. Pour les autres types de moyennes, l'intérêt est de savoir qu'elles existent même si elles son s logiciels les fournissent par défaut.

POUR EN SAVOIR PLUS...

Ministère des Forêts, de la Faune et des Parcs
Direction de l'aménagement et de l'environnement forestiers
5700, 4e Avenue Ouest
Québec (Québec) G1H 6R1
daef@mffp.gouv.qc.ca

N° de publica ion : DAEF2-0069-F-8

MFFP - Juin 2016