Simplifions la statistique

Par Louis Blais, stat., ASSQ

MESURES DE DISPERSION

Les mesures de tendance centrale sont très importantes en sta t, employées seules, elles ne perme ent pas de bien décrire les données. Les mesures qui les complètent bien sont les mesures de dispersion. Une mesure de dispersion est une mesure qui quan étalement des observa aleur minimum à la valeur maximum.

En e et, deux échan érents peuvent avoir la même moyenne arithmé des dispersions tr érentes. Ou encore, l'étalement des observa t tellement vaste que la moyenne seule ne veut pra t rien dire. Il existe plusieurs mesures de dispersion. Nous allons nous a arder aux principales, soit l'étendue, la variance, l'écart type et le coe t de varia verrons aussi les cen derniers ne cons t pas une mesure proprement dite, mais un ensemble de valeurs qui décrit la dispersion des données observées.

Pour illustr ons les sta une pr ts d'épine e noire livrés, prêts à être mis en terre et dont la qualité doit être v our l'exercice, on a mesuré la hauteur de 360 plants en cen tres. Le tableau 1 présente les sta t inventaire.

TABLEAU 1 Sta
hauteur des plants produits en pépinière

Hauteur moyenne :	14,3 cm
Variance :	11,33 cm²
Écart type :	3,37 cm
Minimum :	8 cm
Maximum :	24 cm

Forêts, Faune et Parcs
Québec

L'ÉTENDUE

L'étendue, ou *range* en anglais, est érence entre la valeur maximale et la valeur minimale des observa . Ce e mesure était tr vant l'avènement des ordinateurs, car elle est très facile à calculer. Cependant, c'es e le moins de garan ontre les valeurs extrêmes, puisque le c t pas compte des autres observa Elle peut donc donner une idée erronée de la dispersion des observa Dans l'exemple présenté en intr étendue est égale à 16 cm (24-8).

LA VARIANCE

La variance est la moyenne arithmé écarts, mis au carré, entre les observa t leur moyenne arithmé . Elle se calcule ainsi :

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

La moyenne de la popula étant pas connue, on divise par « n-1 » plutôt que par « n » pour éviter un biais dans le calcul de la variance de l'échan

Plus les observa t près de la moyenne, moins elles sont dispersées, plus la variance sera pe e. Plus les observa t dispersées, plus la variance sera grande. À noter que les unités de la variance sont toujours des « unités » au carré. Par exemple, si les unités des observa sont en millimètres, on ob t une variance en millimètres carrés. La variance est tr as d' séquentes, par exemple dans les tests sta tervalles de con déterminer la taille d'un échan Dans l'exemple précédent, la variance est de 11,33 cm². Les unités étant au carré, il es savoir si ce e variance est élevée ou non.

La variance sert au calcul de l'écart type, de l'erreur type et du coe t de varia également à calculer l'intervalle de con t à déterminer le nombre d'unités d'échan e nécessaire selon le niveau de précision et de con t dans les tests sta amment lorsque l'on compare deux moyennes de deux échan

Dans Excel 2013, quatre f alculent la variance. Ce sont « VAR.P.N », « VAR.P », « VAR.S » et « VAR ». Les f AR.P.N » et « VAR.P » sont iden t calculent la variance de la popula en supposant que la moyenne de la popula t connue. Dans les faits, le diviseur de la f t « n » plutôt que « n-1 » et on ne connaît jamais, ou très rarement, la moyenne de la popula aut AR.S » ou « VAR », qui sont équivalentes.

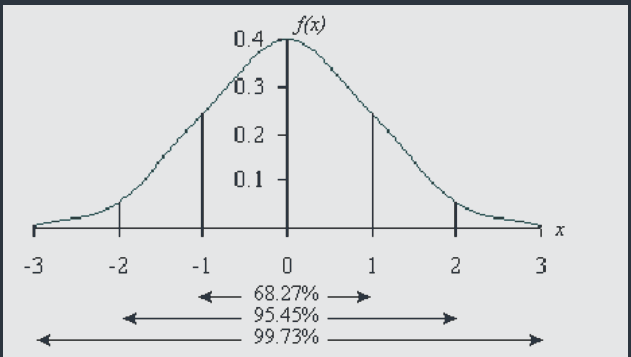
L'ÉCART TYPE

Mathéma t, **l'écart type** n'est rien d'autre que **la racine carrée de la variance**. Il se calcule ainsi :

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{(n-1)}}$$

C'est la moyenne quadra arts par rapport à la moyenne arithmé serva par « *n-1* » plutôt que par « *n* »). L'écart type sert à exprimer l'écart moyen des observa apport à la moyenne arithmé 'avantage ici est que ses unités (p.ex. : cm) sont les mêmes que celles des observa

Pour désigner l'écart type, cert t à tort le terme « dévia tandard » qui est une tr érale de l'anglais « *standard deviation* ». Il ne faut pas non plus le confondre avec « l'erreur type ».



Si les observa ent une loi normale, 68,27 % des observa t à ± 1 écart type, 95,45 %, à ± 2 écarts types et 99,73 %, à ± 3 écarts types. Ce e connaissance permet de bien évaluer la dispersion des observa e ci-dessus).

Dans l'exemple des plants produits en pépinière, l'écart type est de 3,37 cm. Ce e sta qu'en moyenne chaque plant s'écarte de 3,37 cm de la moyenne, mais cet énoncé est incomplet. En fait, 68,27 % des observa t un écart de ± 3,37 cm, 95,45 % de ± 6,74 cm (2 x 3,37 cm) et 99,73 % de ± 10,1 cm (3 x 3,37 cm).

Dans Excel 2013, plusieurs f alculent l'écart type. Ce sont « ECARTYPE », « ECARTYPE.STANDARD », « ECARTYPEP », « ECARTYPE.PEARSON » et « ECART.MOYEN ». Les f « ECARTYPE » et « ECARTYPE.STANDARD » sont iden t calculent l'écart type présenté dans ce CARTYPEP » et « ECARTYPE.PEARSON » sont iden t calculent l'écart type d'une popula t que l'on connaît la vraie moyenne. Comme pour la variance, elle divise par « n » au lieu de « n-1 », mais puisque l'on connaît rarement la vraie moyenne, il ne faut t, la f « ECART.MOYEN » calcule l'écart moyen par rapport à la moyenne en valeur absolue. Elle peut être intéressante, bien qu' énéralement l'écart type qui peut servir à faire d'autres calculs.

L'ERREUR TYPE

L'erreur type est l'écart type divisé par la racine carrée du nombre d'unités d'échan e, c'est-à-dire : $\frac{s}{\sqrt{n}}$

Ce e mesure a un lien direct avec la précision des sondages. Plus la taille de l'échan t élevée, plus l'erreur type sera pe e, et plus le sondage sera pr és que l'écart type, c'est-à-dire les mêmes unités que les observa L'écart type et l'erreur type sont deux mesur Prenons l'exemple de la pépinière qui produit des plants pour le reboisement. Les plants doivent-ils se ressembler le plus possible ou doivent-ils avoir une hauteur moyenne le plus près possible de l' de pr e veut des plants qui se ressemblent le plus possible, l'écart type est une excellente mesure. Par contre, si l' t de produire des plants avec une moyenne de « *x* » cm, indépendamment de l'écart entre les plants, c'est l'erreur type qui est importante. Il faut cependant noter que, dans le deuxième c e, si l'écart type est élevé et que l'échan e doit avoir une certaine précision, le nombre de plants à inventorier sera plus grand et les coûts de l'échan e seront plus élevés.

Dans notre exemple, l'erreur type est de 0,18 cm. Cela veut dire que s'il était possible de réaliser une in é d'échan es indépendants de 360 plants de la même popula yenne alors obtenue (si on les mesurait tous) s'écarterait, en moyenne, d'environ ± 0,18 cm de la vraie valeur.

LE COEFFICIENT DE VARIATION

Le coe t de varia t calculé en faisant le **rapport de l'éc moyenne**, c'est-à-dire : $CV = {}^{100s}/_{\bar{x}}$

C'est une mesure sans unité, exprimée en pourcentage (%), mais ce n'est pas une pr Le coe t de varia tre supérieur à 100 % si l'écart type est supérieur à la moyenne. Il est tr ela importance de l'écart type en tenant compte de sa moyenne. Un faible coe t de varia xemple, inférieur à 10 %) indique une faible dispersion des données

autour de la valeur centrale. Comme on l'a men écédemment, deux échan peuvent avoir la même moyenne, tout en ayant des v érentes. Aussi, deux échan peuvent avoir la même variance, mais une moyenne érente. Dans ce dernier cas, le coe t de varia a plus grand pour l'échan moyenne est plus pe e, ce qui indique que la variabilité est plus importante. Ce e sta es squ'on fait un échan e et qu'on recherche une certaine précision (par exemple, pour es yenne d'une pr ts). Plus le coe t de varia t élevé, plus on aura besoin d'unités d'échan e pour obtenir ce e précision.

Dans l'exemple sur la pr ts, le coef- t de varia t de 23,6 s de la pr ts est de diminuer la variabilité individuelle de la hauteur de chaque plant par rapport à la moyenne, il faudrait alors viser un coe t de varia açons de faire pour le diminuer : augmenter la hauteur des plants sans augmenter l'écart type ou diminuer l'écart type tout en conservant la même moyenne.

LES CENTILES

Un cen t chacune des 99 valeurs qui divisent les données triées, de la plus pe e valeur à la plus grande, en 100 parts égales, de sorte que eprésente 1/100 de l'échan Un c t le 50^{e} cen t la médiane. Mais d'autres cen ^{e}, 25^{e}, 75^{e} et 95^{e}. L'ensemble de ces cen çu de la dispersion des données. Ainsi, les données situées entre le 25^{e} et le 75^{e} cen omprennen serva autour de la médiane. Entre le 5^{e} et le 95^{e} cen ce sont 90 % des observa t comprises entre ces valeurs.

TABLEAU 2 Cen échan ts d'épine e noire

Cen	5^{e}	25^{e}	50^{e}	75^{e}	95^{e}
Hauteur (cm)	9	12	15	17	21

Ces données montrent que la médiane (tableau 2) et la moyenne (tableau 1) sont rela ement semblables. Elles traduisent bien la symétrie des données autour de la moyenne. Notons que 50 % des plants ont entre 12 et 17 cm de hauteur et que 90 % d'entre eux ont entre 9 et 21 cm de hauteur.

Les cen tant indépendants des valeurs extrêmes qui peuvent être présentes dans un échan risquent d'in t la moyenne, même si dans certains cas ils son prendre une décision. En e et, pour déterminer des aines décisions seront prises, les cen ent perme e de connaître l'ensemble des cas touchés par ces seuils.

D'autre part, comme il n'est pas certain que les observa ent une loi normale, ement si la dis t asymétrique, les cen ont alors servir de valeurs de référence dans la dis serva

CONCLUSION

s à a eindre, chaque mesure de disper é, qui peut aller d'une simple descrip jusqu'à la détermina un nombre d'unités d'échan e.

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 publication : DAEF2-0069-F-7