

## Analyse et prise de décision impliquant des distributions à un caractère à l'aide d'outils statistiques

La statistique descriptive offre à l'élève une diversité de concepts lui permettant de s'initier aux inférences. Au secondaire, l'élève s'approprie des outils pour traiter des données recueillies, en tirer des informations et exercer son jugement critique afin de découvrir d'éventuelles sources de biais. Les activités qui lui sont proposées l'amèneront à représenter des données à l'aide de tableaux ou de diagrammes, selon le type de données. Il doit également être appelé à les interpréter, notamment en observant leur distribution (ex. : forme, étendue, centre, regroupements) ou en comparant des distributions. À la fin du secondaire, il est conscient de la variabilité de l'échantillon ainsi que des limites et des contraintes associées à l'échantillonnage d'une population.

L'analyse d'une distribution, ainsi que la prise de décision qui s'ensuit, demande à l'élève de choisir des mesures statistiques qui caractérisent cette distribution. Dans le cas d'une distribution à un caractère, on utilise habituellement deux mesures statistiques : une mesure de tendance centrale et une mesure de dispersion. En 3<sup>e</sup> secondaire l'élève apprend à utiliser le mode, la médiane et la moyenne pondérée ainsi que l'étendue interquartile. En 4<sup>e</sup> secondaire, l'élève inscrit dans la séquence *Culture, société et technique* apprend à utiliser l'écart moyen; celui inscrit dans la séquence *Technico-sciences*, l'écart moyen et l'écart type. Selon la distribution qu'il doit analyser, l'élève doit choisir le couple le plus approprié. Le couple le plus utilisé est celui formé de la moyenne et de l'écart type, mais on utilise aussi la médiane et l'étendue interquartile.

Par exemple, on présente à l'élève la situation suivante :

Différentes études ont mis en évidence l'impact des traumatismes crâniens sur les réflexes.

Pour déterminer si un joueur de hockey est apte à revenir au jeu à la suite d'une blessure, il doit, entre autres, passer un test mesurant la qualité de ses réflexes. Par la suite, on compare son résultat avec ceux obtenus par l'ensemble de l'équipe lors du camp d'entraînement.

Au cours de la dernière partie, Félix a subi un traumatisme crânien à la suite d'un violent plaqué. Quelques jours plus tard, il a obtenu un résultat de 34 au test sur les réflexes. Pour décider si Félix est apte à revenir au jeu, l'entraîneur doit comparer son résultat avec ceux de l'équipe :

36	36	37	38	39	39	40	40	40
41	41	41	42	42	42	42	42	42
43	43	43	43	44	44	44		

Quelle peut être la décision de l'entraîneur?

Sachant que le résultat de Félix au moment du camp d'entraînement était de 38, est-ce que la décision de l'entraîneur sera modifiée?

## Commentaires

Dans cette situation, pour faire l'analyse de la distribution et apporter par la suite des réponses, l'élève devra choisir un couple formé d'une mesure de tendance centrale et d'une mesure de dispersion. Le premier réflexe de l'élève, pour déterminer la tendance centrale d'un ensemble de données, est souvent d'en calculer la moyenne. Cependant, la moyenne n'est pas toujours adéquate ou suffisante pour caractériser les données. Sa sensibilité aux valeurs extrêmes, notamment dans le cas où la distribution est asymétrique, fait qu'on lui préférera parfois la médiane, puisque celle-ci n'est pas influencée par les valeurs extrêmes et les cas aberrants. D'autre part, le calcul de la médiane permet de donner une assurance quant à la validité de la moyenne lorsque ces deux mesures ont des valeurs semblables.

En 4<sup>e</sup> secondaire, l'élève pourrait choisir un couple formé de la moyenne et de l'écart moyen. Bien que moins fréquemment utilisé que l'écart type, l'écart moyen permet de caractériser également la dispersion d'une distribution à un caractère. La façon plus naturelle de le calculer permet à l'élève de comprendre le lien étroit entre

la moyenne et l'écart moyen. Par rapport à l'écart moyen, l'écart type est moins intuitif. Par contre, il possède l'avantage, en vertu de ses propriétés mathématiques, d'être utilisable dans un grand nombre de traitements statistiques. De plus, en se basant sur le carré des écarts, l'écart type amplifie beaucoup plus le poids des valeurs extrêmes de la distribution que ne le fait l'écart moyen. Cependant, en raison de leurs liens étroits avec la moyenne, l'écart type et l'écart moyen peuvent être grandement influencés si ladite moyenne ne représente pas de façon appropriée la tendance centrale.

D'autre part, l'élève, même en 4<sup>e</sup> secondaire, pourrait choisir un couple formé de la médiane et de l'étendue interquartile. Ce dernier est un bon estimateur de la dispersion, et il est beaucoup plus simple à comprendre et à calculer que l'écart type ou l'écart moyen. De plus, comme il concerne la dispersion des valeurs autour de la médiane, il n'est pas sensible aux valeurs extrêmes, le rendant ainsi particulièrement intéressant dans le cas de distributions asymétriques.

Revenons à l'exemple de la page précédente. Après avoir effectué divers calculs, l'élève pourra obtenir les mesures suivantes :

moyenne : 40,9	écart moyen : 1,9	écart type : 2,3
$Q_1$ : 39,5	médiane : 42	$Q_3$ : 43

L'analyse préliminaire de ces résultats, qui montre que la moyenne et la médiane ont des valeurs significativement différentes, ainsi que l'observation de la distribution, en particulier sa forme asymétrique, pourront permettre à l'élève de mettre en doute la validité de la moyenne pour représenter la tendance centrale de cette distribution. Ainsi, qu'il soit en 3<sup>e</sup> ou en 4<sup>e</sup> secondaire, l'élève devrait porter son choix sur le couple formé de la médiane et de l'étendue interquartile.

Pour établir quelle sera la décision de l'entraîneur, l'élève devra déterminer la position du résultat de 34 obtenu par Félix au regard de la distribution. Généralement, une donnée est considérée aberrante si sa valeur est inférieure à  $Q_1 - 1,5(Q_3 - Q_1)$ . Puisque 34 est inférieur à 34,25, l'élève pourra conclure que le résultat obtenu par Félix est une donnée aberrante. En conséquence, Félix n'est pas apte à revenir au jeu.

D'autre part, lors du camp d'entraînement, le résultat de Félix était inférieur à  $Q_1$ . L'élève pourrait arguer que la position de l'entraîneur s'en trouve modifiée puisque 34 est supérieur à  $38 - 1,5(Q_3 - Q_1)$ .