Les statistiques expliquées par les chats

Les statistiques et l'analyse de données sont un point essentiel de la culture scientifique nécessaire aux jeunes pour devenir des citoyens éclairés. Le développement de leur esprit critique implique le développement de compétences en lien avec l'analyse et la prise de recul face à des données et leurs interprétations.
Au travers d'un exemple simple, appliqué à un petit groupe d'individus jusqu'à la généralisation à une population beaucoup plus grande, l'objectif est de sensibiliser les jeunes aux problèmes liés à la représentativité des échantillons ainsi qu'aux limites de l'interprétation de résultats en ne se fiant qu'à la moyenne

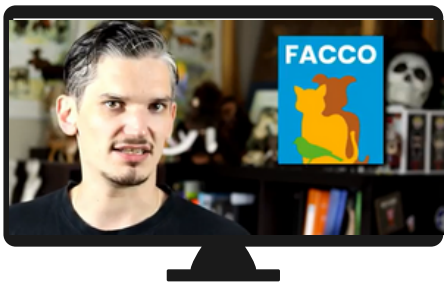
Groupes visés	15 - 16 ans
Disciplines principales	Mathématiques SVT ou Physique pour l'apport théorique sur la méta-analyse
Lien spécifique avec le programme scolaire (si pertinent)	Traitement de l'information chiffrée et des données statistiques (moyenne-médiane-écart type-étendu-quartiles)
Durée	55 minutes
Temps de Préparation estimé	30 min à 1h

Objectifs pédagogiques

- Pendant la séquence, les jeunes utilisent différents indicateurs statistiques
- Après la séquence, les jeunes ont compris qu'un petit nombre n'est pas forcément représentatif pour valider une hypothèse, il faut un plus grand nombre pour avoir une conclusion réelle.
- Après la séquence, les jeunes sont d'accord pour dire que les chiffres doivent être associés à leur contexte pour bien les comprendre
- Après la séquence, les jeunes sont d'accord pour dire qu'il faut se méfier des "vérités générales" tirées de chiffres

Cliquez sur la vidéo pour accéder à l'introduction de la séquence

ou copiez ce lien dans votre navigateur :
https://bit.ly/DIScERNE9



Eléments de préparation de la séquence



Ressources et supports à prévoir pour l'activité

- Vidéo d'intro
- Feuille de route encadrant/encadrante
- Feuille d'activité jeunes

Etapes	Déroulé synthétique de la séquence	Temps estimé
1	Préparation d'un sondage et récupération de données sur les animaux de compagnie dans les foyers des jeunes	En amont
2	Visionnage de la vidéo d'introduction et questions aux jeunes	5 min
3	Mise en place de l'échantillon fallacieux et réflexion collective	5 min
4	Travaux de groupe sur les échantillonages	10 min
5	Analyse des données et questionnement de la représentativité	10 min
6	Extension de l'étude : échantillonnage plus large	10 min
7	Analyse de la représentativité et comparaison avec les chiffres nationaux	5 min
8	Ouverture de l'analyse statistique vers d'autres domaines	10 min

Déroulé détaillé de l'activité

Temps estimé : 55 minutes

1 Étape 1 : en amont

Faire un sondage rapide auprès des jeunes : sur un papier, demandez leur d'écrire leur nom, leur prénom et s'ils ont un ou plusieurs animaux domestiques dans leur foyer, et si oui lesquels. Conservez les papiers. Reportez les dans un tableau que vous pourrez projeter au tableau et/ou imprimer sur une feuille à distribuer aux jeunes. Ces données vous seront utiles à l'**étape 6** de la séance.

Demandez à d'autres encadrant, encadrante d'autres groupes de faire ce même exercice avec les jeunes de leur groupe sans indiquer leur nom et leur prénom. Reportez les données dans un tableau que vous pourrez projeter au tableau et/ou imprimer sur une feuille à distribuer aux jeunes. Ces données vous seront utiles pour l'**étape 7** de la séance. **(voir en annexes les fiches sondage à imprimer)**

Nous recommandons également la lecture de cet article qui explicite ce qu'est la méta-analyse : "La méta-analyse : bien plus que le simple calcul d'un effet combiné !"
https://www.inspq.qc.ca/bise/la-meta-analyse-bien-plus-que-le-simple-calcul-d-un-effet-combine

2 Étape 2 : Introduction *(temps recommandé : 5 min)*



Avec les jeunes, visionnez la vidéo d'introduction sur le nombre de chats en France et le nombre de foyers en France.

Demander aux jeunes, d'après la vidéo combien de leurs camarades devraient avoir un poisson chez eux ? S'ils ont bien suivi la vidéo, ils répondront tous. Sont-ils d'accord ? Certains jeunes devraient signaler qu'ils n'ont pas de poisson chez eux. Cela signifie que tous les foyers n'ont pas de poisson chez eux contrairement à ce qui est dit dans la vidéo. D'après eux, pourquoi se sont-ils trompés dans la vidéo ?

3 Étape 3 : *(temps recommandé : 5 min)*



Pour poursuivre, demandez à 4 jeunes - prétendument sélectionnés au hasard - de se lever. Demandez-leur de se lever et demandez-leur s'ils ont un chat à la maison. La réponse des 4 élèves sera "Oui".
Explicitez aux jeunes que d'après votre échantillon, 100% des jeunes du groupe ont un chat à la maison. Sont-ils d'accord ? De nouveau, certains jeunes signaleront qu'ils n'ont pas de chat à la maison. Par ailleurs, ce résultat n'est pas non plus celui rapporté dans la vidéo ("50% de foyers ont un chat").
De nouveau, interrogez les jeunes: où est l'erreur ? Normalement, ils devraient comprendre que le problème réside dans le choix des jeunes qui constituent l'échantillon dit "représentatif".
Pour être certain du nombre de chats dans le foyer des jeunes du groupe, il faut s'intéresser à l'ensemble du groupe.

4 Étape 4 : *(temps recommandé : 10 min)*

Demandez aux jeunes de se mettre en groupe de 4 voire 5.
Demandez à chaque groupe de calculer le nombre moyen de chats des jeunes qui constituent leur groupe. Quel résultat ont-ils trouvé ? Notez les différents résultats de chaque groupe au tableau.
Normalement, le nombre de chats moyen par groupe ne sera pas le même pour tous les groupes. Qu'en concluent-ils ? Quel résultat est le bon ? Y-a-t-il un bon résultat parmi ceux qu'ils ont trouvés ?

En interrogeant les jeunes, orientez la discussion pour faire émerger l'idée qu'il faut refaire ce calcul mais à l'échelle du groupe entier.

5 Étape 5 : *(temps recommandé : 10 min)*

Projeter alors au tableau et/ou distribuer aux jeunes une feuille reprenant les données que vous avez collecté grâce aux papiers donnés par les jeunes de votre groupe avant la séance spécifiant le ou les animaux domestiques présents dans leur foyer. N'oubliez pas d'enlever les noms et les prénoms des jeunes.

En reprenant ces données, demandez à chaque groupe de calculer le nombre moyen de chats par foyer des jeunes du groupe. Quel résultat ont-ils trouvé? Notez les différents résultats de chaque groupe au tableau pour vous assurer qu'ils ont bien tous trouvé le même chiffre. Qu'en concluent-ils?

Les calculs qu'ils viennent de réaliser portent sur les jeunes de leur groupe. Mais comment savoir si le groupe est représentatif des autres? Pour cela, il faudrait comparer leur résultat avec le nombre moyen de chats par foyer d'autres groupes. Ça tombe bien! Vous avez justement des jeux de données qui permettent de faire ce calcul!

6 Étape 6 : *(temps recommandé : 10 min)*

Proposer maintenant aux jeunes d'étendre l'étude à d'autres personnes présentes dans l'établissement. De nouveau par groupe, demandez leur de calculer le nombre moyen de chats par foyer des autre groupes. Pour ne pas que cette étape prenne trop de temps, nous vous conseillons soit :

- de vous limiter à seulement une voire deux autres groupes
- de prendre des données de plus de groupe mais de répartir le calcule des moyens dans les groupes

Que conclue-t-il ? Est-ce que tous les résultats sont les mêmes ? Sont-ils différents ? Si oui, qu'est-ce que cela signifie ? Comment peuvent-ils expliquer ce résultat ?

Normalement, les moyennes entre les différentes groupes ne seront pas parfaitement identiques mais similaires. Demandez aux jeunes de préciser pour les autres groupes, combien de jeunes n'ont pas de chat, combien en ont un, deux voire plus. Faites remarquer que ces éléments sont gommés par le calcul des moyennes. Ainsi deux moyennes similaires peuvent refléter des populations assez différentes.

Pour aller plus loin :

Un moyen de prendre en compte ces différences est de calculer la médiane ou l'écart type des différents groupes. Si vous vous avez déjà abordé auparavant ces notions avec les jeunes, vous pouvez leur demander de calculer pour chacun des groupes la médiane et/ou l'écart type. Qu'observent-t-ils? Que signifient ces indicateurs statistiques?

7 **Étape 7** : *(temps recommandé : 10 min)*

Au tableau, projeter le résultat qu'ils auraient obtenu s'ils avaient calculé le nombre moyen de chats sur l'ensemble des groupes pour lesquelles vous avez recueilli des données. Attirez l'attention sur les différents résultats qu'ils ont obtenus en fonction des groupes qu'ils ont utilisés pour calculer le nombre de chats moyen par foyer. Lequel des ces chiffres est le plus représentatif de la population française ? Pourquoi ?

Comment pourrait-on faire pour être certain que ce chiffre est représentatif du nombre de chat moyen par foyer en France? L'idéal serait d'interroger l'ensemble des Français. Cependant cela serait compliqué et long. Aussi, certaines personnes n'auraient tout simplement pas envie de répondre à l'enquête. En tout cas, il faudrait un nombre suffisamment important de Français et qui avec des profils différents pour que l'échantillon soit représentatif des Français qui ne répondraient pas à cette enquête.

Pour apporter des éléments de réponse, vous pouvez comparer les résultats obtenus par les jeunes avec les chiffres nationaux.

Pour aller plus loin :

De nouveau, vous pouvez lancer la réflexion sur le fait que la population de référence peut présenter une importante dispersion dans le nombre de chats par foyer. La moyenne nationale peut être élevée parce que certains foyers ont beaucoup de chats. Cela peut expliquer par exemple pourquoi dans la vidéo il est dit que 100% des foyers ont un poisson. La plupart des poissons étant des animaux sociaux, il est fort possible que la plupart des foyers n'est pas un poisson mais plusieurs. A l'échelle de la moyenne cela donne l'impression que chaque foyer a un poisson alors que c'est faux.

Optionnel : Calculer la moyenne-médiane-écart type-quartiles-étendue pour la donnée étudiée à l'échelle du groupe puis de plusieurs groupes.

Étape 8 : *(temps recommandé : 10 min)*

Terminer la séance en ouvrant la discussion sur la généralisation de ce processus construction des connaissances à d'autres thématiques. Par exemple, comment procède-t-on lorsque l'on souhaite tester un nouveau médicament? Comment peut-on précéder pour vérifier si les températures augmentent réellement? Quelles sont les risques si on ne prend pas une population assez grande et/ou non représentative? Quelles sont les risques si on ne s'intéresse qu'à la moyenne? Pour cette dernière question, la moyenne peut masquer des sous populations avec des caractéristiques spécifiques pour qui un médicament ne fonctionne pas. Concernant l'augmentation des températures, cela peut conduire à sous estimer l'augmentation des températures (moins de 2 degrés sur l'année alors que le maximum de température augmente de bien plus que 2 degrés sur un territoire en été).

Nota Bene :

Il serait possible d'ouvrir (ou de remplacer) par le paradoxe des anniversaires : voir combien de personnes dans le groupe ont la même date de naissance, demander aux jeunes selon eux combien il faut de personnes dans la même pièce pour avoir 90% de chances que deux aient le même anniversaire, puis étendre avec la démonstration du nombre de personnes.

Fin de la séquence

Coordination pour l'Association Le Cercle FSER
Héloïse Dufour & Fanny Bilak

Contributeurs et contributrices à la présente séquence
Christophe le Guevouilt, Tania Louis, Magali Kintzmann, Marie Péquignot, Kévin de Checchi

Remerciements
Charlotte Barbier, Barthelemy Melanie, Lorie Trognko

Cette ressource a été produite par le Cercle FSER avec le soutien du ministère de la Culture et de Genopolys

Date de publication
Janvier 2023

Licence
Ce document a été publié par l'association Le Cercle FSER sous la licence Creative Commons suivante : Attribution-NonCommercial-NoDerivatives 4.0 International License
Le titulaire des droits autorise l'exploitation de l'œuvre originale à des fins non commerciales.

Association le Cercle FSER
90 rue de Varenne
75007 Paris
discerne@cerclefser.org
http://www.cerclefser.org/fr/