

Renforcement des capacités en recherche et évaluation
Glossaire de la capsule 4 – Analyse de données quantitatives

Table des matières

Analyse comparative entre les sexes plus, ACS Plus (en Anglais Gender-based Analysis, GBA+) :	2
Distribution :	2
Distribution normale:	2
Données aberrantes :	2
Données manquantes :	2
Écart interquartile:	3
Écart type :	3
Étendue :	3
Figures :	3
Fréquence :	5
Maximum :	5
Médiane :	5
Minimum:	5
Mode :	6
Moyenne :	6
Nettoyage des données :	6
Statistique descriptive:	6
Statistique inférentielle :	6
Tableau de fréquence croisé :	7
Taux de réponse :	7
Test du Chi Carré (χ^2) :	7
Test-t de Student :	7
Valeur-p:	7
Variable :	7
Variable continue :	7
Variable nominale :	8
Variable ordinale :	8

Analyse comparative entre les sexes plus, ACS Plus (en Anglais Gender-based Analysis, GBA+)¹ :

L'ACS Plus est un processus analytique qui fournit une méthode rigoureuse pour évaluer les inégalités systémiques, ainsi qu'un moyen de déterminer comment différents groupes de femmes, d'hommes et de personnes de diverses identités de genre peuvent vivre les politiques, programmes et initiatives. Le signe « plus » dans ACS Plus reconnaît que l'ACS Plus ne se limite pas aux différences biologiques (sexe) et socioculturelles (genre). Nous avons tous de multiples facteurs identitaires qui se recoupent et contribuent à faire de nous ce que nous sommes. L'ACS Plus examine de nombreux autres facteurs identitaires comme la race, l'origine ethnique, la religion, l'âge, et les handicaps de nature physique ou mentale, et la façon dont leur interaction influence la manière dont nous pouvons ressentir les politiques et les initiatives gouvernementales.

Distribution² :

La distribution d'une variable est le profil des valeurs, c'est-à-dire l'ensemble formé de toutes les valeurs possibles et des fréquences associées à ces valeurs. Les distributions de fréquences sont représentées sous forme de tableaux ou de graphiques.

Distribution normale³:

La distribution normale ou de Gauss est une distribution avec les caractéristiques suivantes:

- la distribution prend la forme d'une courbe en cloche
- la distribution est symétrique
- la moyenne, la mode et la médiane sont identiques (noté μ)
- 68.3% des observations est à l'intérieur de l'intervalle $\mu \pm \sigma$ où σ représente l'écart-type de la variable

Données aberrantes⁴ :

Une donnée aberrante (outlier) est une valeur ou une observation qui est distante des autres observations effectuées sur le même phénomène, c'est-à-dire qu'elle contraste grandement avec les valeurs normalement mesurées. Une donnée aberrante peut être due à la variabilité inhérente au phénomène observé ou bien elle peut aussi indiquer une erreur expérimentale. Ces données aberrantes sont généralement exclues de la série de données au moment de l'analyse des résultats.

Données manquantes⁵ :

Les données manquantes ou les valeurs manquantes se produisent lorsqu'aucune valeur de données n'est représentée pour une variable pour une observation donnée. Les données manquantes sont courantes et peuvent avoir un effet significatif sur l'inférence, les performances de prédiction ou toute autre utilisation faite avec les données. Des données manquantes peuvent exister dans les données en raison d'une « omission de réponse » pour l'observation donnée. Dans cette situation, aucune information n'est fournie pour un élément, pour plusieurs éléments ou pour une unité entière, c'est-à-dire un sujet.

Écart interquartile⁶:

L'écart interquartile donne une idée juste de la dispersion des données. Pour calculer cette mesure, il faut d'abord identifier les quartiles. Le quartile inférieur, ou premier quartile (Q1), est la valeur au-dessous de laquelle se trouvent 25 % des données lorsqu'elles sont arrangées en ordre croissant. Le quartile supérieur, ou troisième quartile (Q3), est la valeur au-dessous de laquelle se trouvent 75 % des données arrangées en ordre croissant. La médiane est considérée comme le second quartile (Q2). L'écart interquartile est la différence entre le quartile supérieur et le quartile inférieur.

Écart type⁷ :

L'écart-type sert à mesurer la dispersion, ou l'étalement, d'un ensemble de valeurs autour de leur moyenne. Plus l'écart-type est faible, plus la population est homogène.

Étendue⁸ :

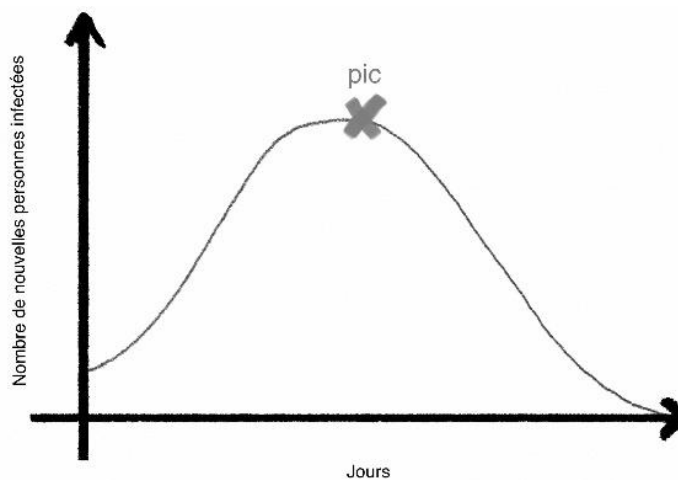
L'étendue est la différence entre la plus grande et la plus petite donnée d'un échantillon.

Figures⁹ :

Les figures sont des présentations visuelles des résultats. Ils se présentent sous la forme de graphiques, de tableaux, de dessins, de photos ou de cartes. Les chiffres fournissent un impact visuel et peuvent communiquer efficacement votre principale découverte.

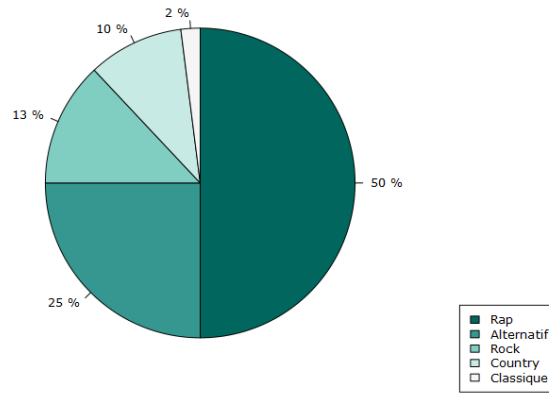
Quelques exemples :

- **Courbe¹⁰ :**



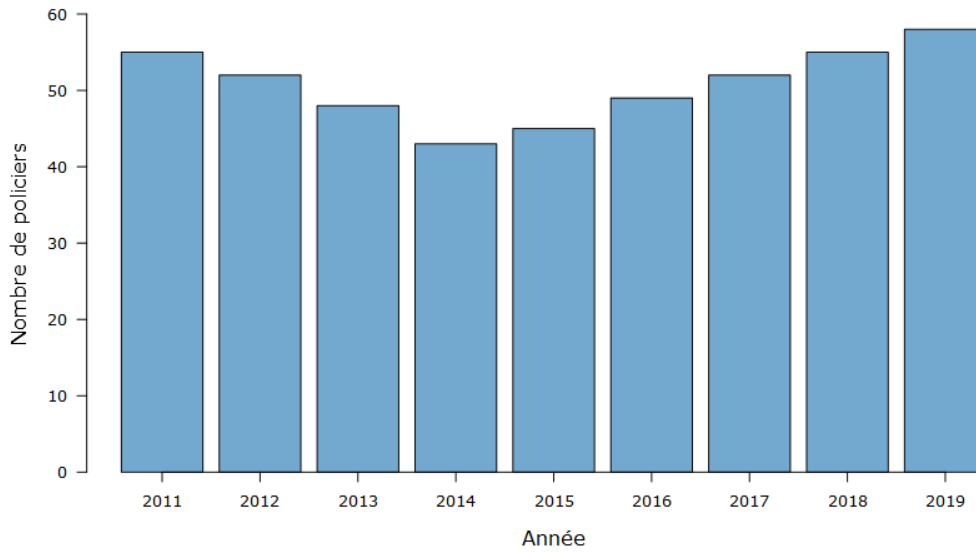
- **Diagramme circulaire ¹¹:**

Graphique 5.4.2
Genres musicaux préférés des jeunes adultes de 14 à 19 ans



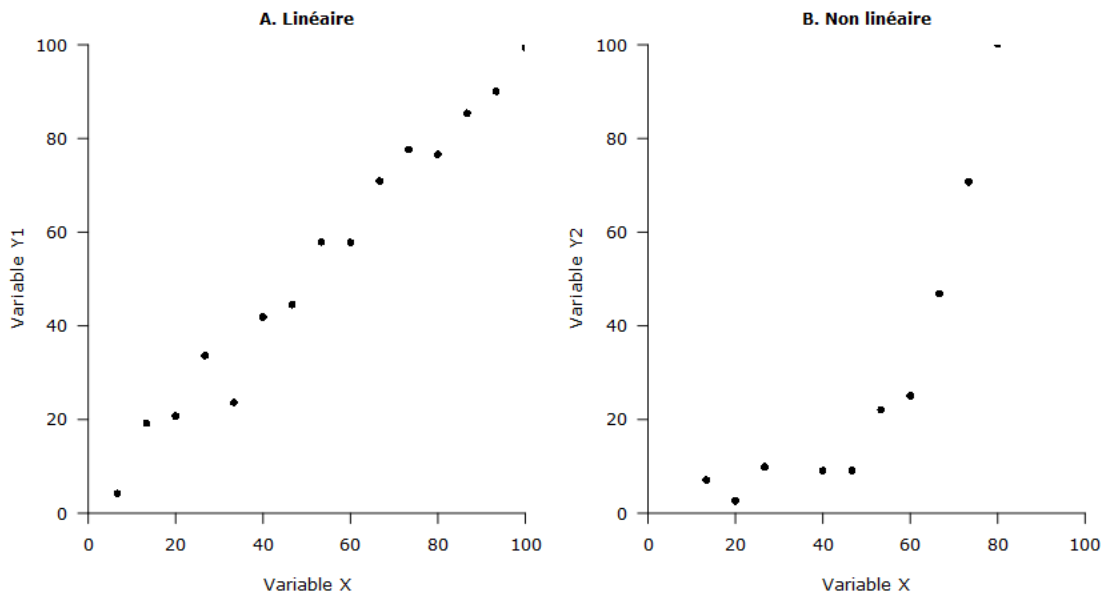
- **Diagramme en bandes ¹² :**

Graphique 5.2.1
Nombre de policiers à Crimeville, 2011 à 2019



- Nuage de points :

Graphique 5.6.2
Relation linéaire ou relation non linéaire



Fréquence¹³ :

La fréquence d'une donnée dans une série statistique correspond au quotient de l'effectif de cette donnée par l'effectif total. La fréquence d'une donnée peut s'exprimer par un nombre décimal inférieur ou égal à 1. La fréquence d'une donnée peut aussi s'exprimer en pourcentage.
Fréquence = taille de la classe/ taille de la population

Maximum¹⁴ :

La valeur « maximum » en statistique est la plus grande valeur que l'on retrouve dans une population ou échantillon.

Médiane¹⁵ :

La médiane est une valeur qui sépare la moitié inférieure et la moitié supérieure des termes d'une série statistique quantitative (si le nombre d'observations n est pair, la médiane est la demi-somme des termes de rang n et n + 1).

Minimum¹⁶ :

La valeur « minimum » en statistique est la plus petite valeur que l'on retrouve dans une population.

Mode¹⁷ :

Le mode ou valeur dominante désigne la valeur la plus représentée d'une variable quelconque dans une population.

Moyenne¹⁸ :

La moyenne est l'indicateur le plus simple pour résumer l'information fournie par un ensemble de données statistiques. La moyenne est calculable pour les variables numériques, qu'elles soient discrètes ou continues. On l'obtient simplement en additionnant l'ensemble des valeurs et en divisant cette somme par le nombre de valeurs. Ce calcul peut être fait à partir des données brutes ou d'un tableau de fréquences.

Nettoyage des données¹⁹ :

Il faut explorer les données pour comprendre le format et les variables et les vérifier pour voir s'il y a des erreurs et des valeurs manquantes. Il peut être nécessaire de nettoyer les données avant de les utiliser à des fins d'analyse, ce qui comprend notamment la correction du formatage, le retrait ou la correction de données erronées, ou tout simplement le retrait d'espace supplémentaire. Il est important de documenter ce que vous avez trouvé et ce que vous avez fait pour nettoyer les données.

Statistique descriptive²⁰:

Les statistiques descriptives, comme leur nom l'indique, sont utilisées pour résumer ou décrire l'ensemble de données. En ce qui concerne les ensembles de données, ce sont des observations ou de réponses recueillies auprès d'une population ou d'un échantillon d'une population. Alors qu'une population est un groupe entier auprès duquel des données sont recueillies, un échantillon est une sous-ensemble d'une population. La population n'est pas nécessairement composée de personnes, mais peut être constituée de toute autre entité, comme des entreprises, des villes, des pays, etc.

Statistique inférentielle²¹ :

Les statistiques inférentielles utilisent un échantillon aléatoire de données d'une population afin de décrire cette dernière et de faire des déductions à son sujet. Les statistiques inférentielles sont pertinentes lorsqu'il est difficile ou impossible d'examiner chaque membre d'une population entière. Par exemple, il n'est pas pratique de mesurer le diamètre de chaque clou fabriqué dans une aciérie, mais vous pouvez mesurer les diamètres d'un échantillon de clous aléatoire représentatif et utiliser ces informations pour tirer des conclusions générales sur les diamètres de tous les clous fabriqués²². Les statistiques inférentielles permettent donc à un chercheur de faire des hypothèses basées sur des populations réduites, pour des populations plus larges²³.

Tableau de fréquence croisé²⁴ :

Un tableau croisé de fréquences est un tableau de fréquences qui porte sur deux caractères d'une même population. Il décrit donc la ventilation de chaque catégorie d'une variable en fonction d'une autre variable catégorielle.

Taux de réponse²⁵ :

Le taux de réponse désigne la proportion de personnes qui ont participé à un sondage en comparaison au nombre réel de personnes composant l'échantillon à partir de la population cible. Le taux de réponse fournit des informations exploitables sur la précision des données collectées. Il correspond au nombre de personnes qui ont terminé votre sondage divisé par le nombre total de personnes inclus dans l'échantillon de groupe.

Test du Chi Carré (χ^2)²⁶ :

Le test du chi-carré est un test statistique appliqué à des ensembles de données catégorielles afin d'évaluer la probabilité qu'une différence observée entre les ensembles soit due au hasard. Il teste une hypothèse nulle selon laquelle la distribution de fréquence de certains événements observés dans un échantillon est conforme à une distribution théorique particulière.

Test-t de Student²⁷ :

Le test t, également appelé test de Student, est un test statistique populaire servant à mesurer les différences entre les moyennes de deux groupes ou d'un groupe par rapport à une valeur standard. Il est basé sur une distribution Student. L'exécution de ce test sert à comprendre si les différences sont statistiquement significatives, c'est-à-dire si elles ne sont pas le fruit du hasard.

Valeur-p²⁸ :

La valeur-p, parfois aussi appelée p-valeur est la probabilité, calculée en faisant l'hypothèse que H_0 (l'hypothèse nulle) soit vraie, d'obtenir un résultat aussi extrême ou encore plus extrême que celui observé. Plus la valeur-p est faible, plus les données semblent incompatibles avec l'hypothèse nulle.

Ou La plus petite valeur de probabilité pour laquelle l'hypothèse nulle peut encore être rejetée²⁹. Ou

Le niveau de signification en dessous duquel l'hypothèse nulle ne peut plus être rejetée³⁰.

Variable³¹ :

Une variable est une caractéristique mesurable qui peut prendre différentes valeurs. La taille, l'âge, le revenu, la province ou le pays de naissance, les années d'études et le type de logement sont tous des exemples de variables.

Variable continue³² :

On dit qu'une variable est continue si elle prend un nombre infini de valeurs réelles possibles à l'intérieur d'un intervalle donné.

Variable nominale³³ :

Une variable nominale décrit un nom, une étiquette ou une catégorie sans ordre naturel. Le sexe et le genre de logement en sont des exemples. La variable « Mode de transport pour se rendre au travail » est également une variable nominale.

Variable ordinale³⁴ :

Une variable ordinale est une variable dont les valeurs sont définies par une relation d'ordre entre les catégories possibles. Par exemple, la variable « comportement » est ordinale parce que la catégorie « Excellent » est meilleure que la catégorie « Très bon », qui est elle-même meilleure que la catégorie « Bon » et ainsi de suite.

¹ Gouvernement du Canada, (2022). Analyse comparative entre sexes plus. Femmes et Égalité des genres Canada. <https://femmes-egalite-genres.canada.ca/fr/analyse-comparative-entre-sexes-plus/est-analyse-comparative-entre-sexes-plus.html>

² Government of Canada, Statistics Canada. (2021a). Statistics: Power from data! . Government of Canada, Statistics Canada. <https://www150.statcan.gc.ca/n1/edu/power-pouvoir/toc-tdm/5214718-eng.htm>

³ Gouvernement du Québec. (n.d.). Distribution normale. GDT. <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/17027528/distribution-normale>

⁴ Members of the Research Methods Seminar (E600) taught by Mike Palmquist in the 1990s and 2000s. (1994-2022). *Glossary of Key Terms*. Writing@CSU. Colorado State University. <https://writing.colostate.edu/guides/guide.cfm?guideid=90>

Donnée Statistique Aberrante: Lexique de Mathématique. Lexique de mathématique | pour l'enseignement primaire et secondaire. (2019). <https://lexique.netmath.ca/donnee-statistique-aberrante/>

⁵ Cnam – UE RCP208. (2023). Cours - Données manquantes <https://cedric.cnam.fr/vertigo/Cours/ml/coursDonneesManquantes.html>

⁶ Gouvernement du Canada, Statistique Canada (2021a). Calculer l'étendue et l'écart interquartile. 4.5.1 Calculer l'étendue et l'écart interquartile. <https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch12/5214890-fra.htm>

⁷ Gouvernement du Canada, Statistique Canada (2021b). Calculer l'étendue et l'écart interquartile. 4.5.1 Calculer l'étendue et l'écart interquartile. <https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch12/5214890-fra.htm>

⁸ Gouvernement du Canada, Statistique Canada. (2021c). Calculer la variance et l'écart-type. 4.5.3 Calculer la variance et l'écart-type. <https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch12/5214891-fra.htm>

⁹ Howell D.c., (2008). Méthodes statistiques en sciences humaines. 2^e édition.

¹⁰ Binaire, P. (2020). Lire une courbe épidémique. binaire. <https://www.lemonde.fr/blog/binaire/2020/04/15/lire-une-courbe-epidemie/>

¹¹ Government of Canada, Statistics Canada. (2021b). Data visualization, pie chart. 5.4 Pie chart. <https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch9/pie-secteurs/5214826-eng.htm>

¹² Government of Canada, Statistics Canada. (2021c). Data Visualization, Bar Chart. 5.2 Bar chart. <https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch9/bargraph-diagrammeabarres/5214818-eng.htm>

¹³ Gouvernement du Canada, Statistique Canada. (2021d). Distribution de fréquences. 4.3 Distribution de fréquences. <https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch8/5214814-fra.htm>

¹⁴ Gouvernement du Canada, Statistique Canada. (2021e). Calculer la variance et l'écart-type. 4.5.3 Calculer la variance et l'écart-type. <https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch12/5214891-fra.htm>

¹⁵ Gouvernement du Canada, Statistique Canada. (2021f). Calcul de la médiane. 4.4.2 Calcul de la médiane. <https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch11/median-mediane/5214872-fra.htm>

¹⁶ Gouvernement du Canada, Statistique Canada. (2021g). Calculer la variance et l'écart-type. 4.5.3 Calculer la variance et l'écart-type. <https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch12/5214891-fra.htm>

-
- ¹⁷ Gouvernement du Canada, Statistique Canada. (2021g). Calculer la variance et l'écart-type. 4.5.3 Calculer la variance et l'écart-type. <https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch12/5214891-fra.htm>
- ¹⁸ Mesures de la dispersion. Statistique Canada (2021). <https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch12/5214891-fra.htm>
- ¹⁹ Gouvernement du Canada, Statistique Canada (2020). Formation sur la Littératie des Données. Gouvernement du Canada, Statistique Canada. <https://www.statcan.gc.ca/fr/afc/litteratie-donnees>
- ²⁰ Walinga, J., Stangor, C., & Walters, S. (2020). Descriptive and Inferential statistics. Go to the cover page of Psychology – 1st Canadian Edition. <https://psychology.pressbooks.tru.ca/chapter/descriptive-statistics/>
- ²¹ Walinga, J., Stangor, C., & Walters, S. (2020). Descriptive and Inferential statistics. Go to the cover page of Psychology – 1st Canadian Edition. <https://psychology.pressbooks.tru.ca/chapter/descriptive-statistics>
- ²² Fox, W. et Imbeau L.M., (1999). Statistiques sociales. 3^e édition
- ²³ Walinga, J., Stangor, C., & Walters, S. (2020). Descriptive and Inferential statistics. Go to the cover page of Psychology – 1st Canadian Edition. <https://psychology.pressbooks.tru.ca/chapter/descriptive-statistics/>
- ²⁴ Yergeau, E. et Poirier, M. (2021). SPSS à l'UdeS. URL : <http://spss.espaceweb.usherbrooke.ca>
- ²⁵ Gouvernement du Canada, Travaux publics et Services gouvernementaux Canada. (2019). Introduction - Améliorer La participation des répondants aux sondages téléphoniques - recherche sur l'opinion publique au gouvernement du Canada. TPSGC. <https://www.tpsgc-pwgsc.gc.ca/rop-por/rapports-reports/telephone/introduction-fra.html>
- ²⁶ Yergeau, E. et Poirier, M. (2021). SPSS à l'UdeS. URL : <http://spss.espaceweb.usherbrooke.ca>
- ²⁷ Ycart, B. (2002). « Tests statistiques », Cahier de Mathématiques Appliquées, no 6. [cma6.pdf \(imag.fr\)](#)
- ²⁸ Léger, Christian (2016). La valeur-p sous surveillance, Association mathématique du Québec.
- ²⁹ Wooldridge, J. M. (2018) Introduction à l'économétrie: Une approche moderne, 2e édition.
- ³⁰ Wooldridge, J. M. (2018) Introduction à l'économétrie: Une approche moderne, 2e édition.
- ³¹ Gouvernement du Canada, Services publics et Approvisionnement Canada. (2023). Variable - TERMIUM Plus® https://www.btb.termiumplus.gc.ca/tpv2alpha/alpha-fra.html?lang=fra&i=1&srchtxt=variable&codom2nd_wet=1#resultrecs
- ³² Gouvernement du Canada, Services publics et Approvisionnement Canada. (2023). Variable - TERMIUM Plus® https://www.btb.termiumplus.gc.ca/tpv2alpha/alpha-fra.html?lang=fra&i=1&srchtxt=variable&codom2nd_wet=1#resultrecs
- ³³ Gouvernement du Canada, Statistique Canada. (2021h, September 2). Types de variables. 4.2 Types de variables. <https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch8/5214817-fra.htm>
- ³⁴ Gouvernement du Canada, Statistique Canada. (2021h, September 2). Types de variables. 4.2 Types de variables. <https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch8/5214817-fra.htm> .