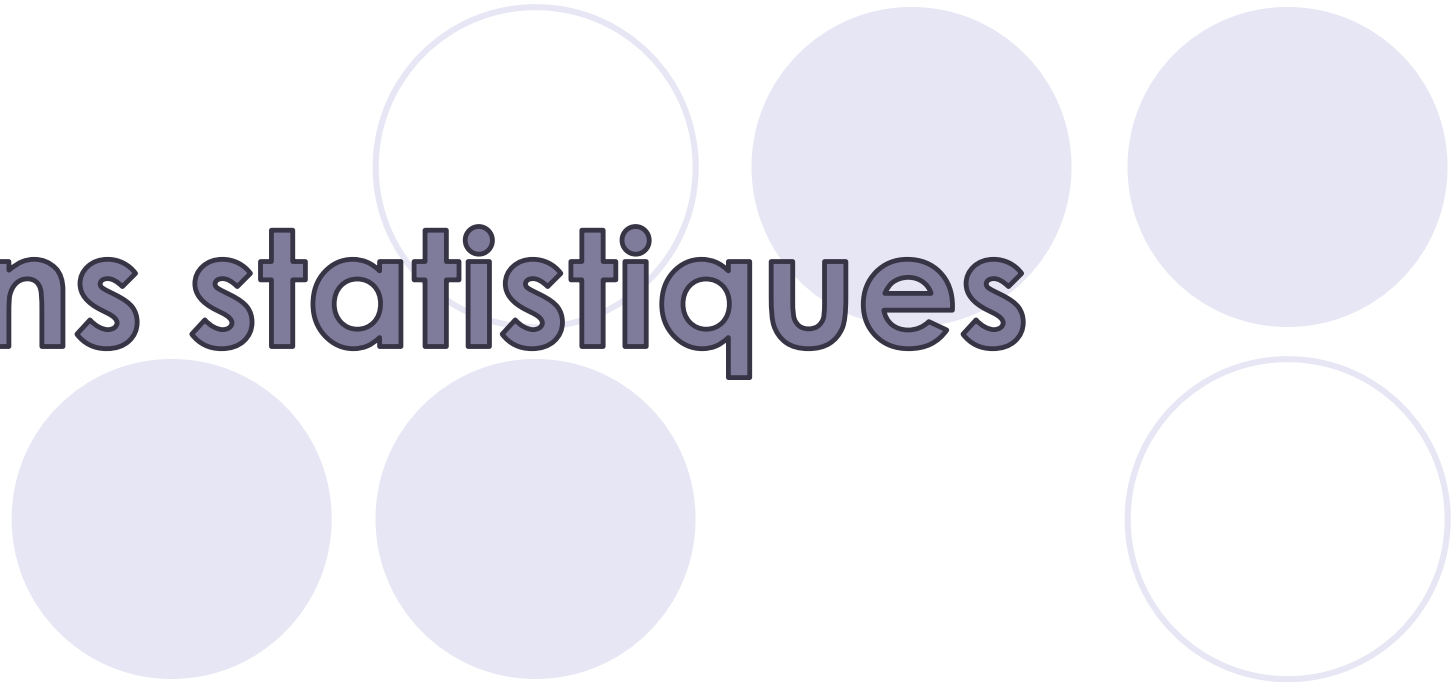


Notions statistiques

Cours 2



Plan du cours



- ◉ Retour sur le cours précédent
- ◉ Statistiques descriptives

Retour cours précédent

- ◉ Définition de statistique
- ◉ Etapes d'une étude statistique
- ◉ Statistique descriptive
 - ◉ Unité statistique
 - ◉ Population
 - ◉ Variable
 - ◉ Variable qualitative
- ◉ Variables quantitatives
 - ◉ Distribution d'une variable
 - ◉ La fréquence

Exemple

Distributions des habitants dans les villes du Québec:

Lieu de vie	Quantité	Fréquence
Montréal-Laval	42	0,49
Québec	27	0.32
Gatineau	12	0.14
Autres	4	0.05



Statistiques descriptives

- ◉ Méthodes graphiques
- ◉ Mesures de tendance

Méthodes graphiques

- ◉ Variables qualitatives
 - ◉ Diagramme circulaire (Pie chart)
 - ◉ Diagramme à bâtons (Bar chart)
- ◉ Variables quantitatives
 - ◉ Histogramme (Histogram)
 - ◉ Graphique « tiges et feuilles » (Stem-and-leaf)
 - ◉ Diagramme en boîte (Box plot) boîte à moustaches
 - ◉ résumé des caractéristiques de position du caractère étudié (médiane, quartiles, minimum, maximum ou déciles).
 - ◉ Comparer un même caractère dans deux populations de tailles différentes.

Nouvelles définitions

- ◉ Moyenne arithmétique
 - ◉ Somme des valeurs divisée par le nombre de valeur
- ◉ La médiane
 - ◉ la donnée centrale d'une série rangée en ordre croissant ou décroissant.
 - ◉ Si données en nombre impair
 - ◉ la médiane est la donnée centrale
 - ◉ la médiane des données 1, 3, 5, 7, 9, 10, 13 est 7.
 - ◉ Si données en nombre pair
 - ◉ la médiane est la moyenne des deux données centrales.
 - ◉ la médiane des données 1, 3, 5, 7, 9, 10, 13, 15 est $(7 + 9)/2 = 8$
 - ◉ La médiane peut ne pas être dans la liste
 - ◉ Robuste à la présence de valeurs aberrantes.

Nouvelles définitions

- ◉ Mesures de dispersion

- ◉ Rapporter une mesure de tendance centrale ne suffit pas, pour donner une idée complète d'une série de données ou d'une distribution.
- ◉ Ajouter des mesures de dispersion (ou de volatilité).

- ◉ Etendue

- ◉ Différence entre la plus grande donnée et la plus petite donnée
 - ◉ série de donnée suivante : 18 13 11 22 1 4 6 17 8
 - ◉ L'étendue est égale à $22 - 1 = 21$.

Nouvelles définitions

◉ Variance

- ◉ Une mesure servant à caractériser la dispersion d'un échantillon ou d'une distribution.
- ◉ Ecart entre la variable et sa moyenne (ou espérance)
- ◉ on prend la moyenne (ou l'espérance) de ces écarts élevés au carré.

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

- ◉ Fonction de la moyenne des observations
- ◉ Comme la moyenne, la variance n'est pas robuste à la présence de valeurs aberrantes.

Etapes pour calculer la variance

- ◉ Pour une suite donnée
 - ◉ Calculer la moyenne de la série (\bar{x})
 - ◉ Pour chaque élément de la série calculer l'écart à la moyenne ($x_i - \bar{x}$)
 - ◉ Elever chaque différence trouvée au carré $(x_i - \bar{x})^2$
 - ◉ Sommer les valeurs trouvées: $\sum_{i=1}^n (x_i - \bar{x})^2$
 - ◉ Diviser par le nombre d'éléments de série – 1: $\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n-1)}$

Nouvelles définitions

- ◉ Ecart-type
 - ◉ Racine carrée de la variance
- ◉ Ecart interquartile (interquartile range)
 - ◉ Le $p^{\text{ième}}$ percentile ou quantile d'un jeu de données présenté en ordre croissant est la valeur telle qu'au plus $p\%$ des valeurs sont en dessous d'elle et au plus $(100-p)\%$ sont au dessus.
 $IQR = Q_3 - Q_1$
 - ◉ Les percentiles les plus utilisés sont le 25^{ième}, le 50^{ième} et 75^{ième} appelés respectivement le premier quantile (Q_1), la médiane (Q_2) et le troisième quantile Q_3 .
 - ◉ L'écart interquartile
 - ◉ $IQR = Q_3 - Q_1$

Nouvelles définitions

- ◉ L'IQR est robuste aux valeurs aberrantes
- ◉ Calcul des quantiles
 - ◉ 1^{er} quantile $\frac{(n+1)}{4}$
 - ◉ 3^{eme} quantile $\frac{3(n+1)}{4}$