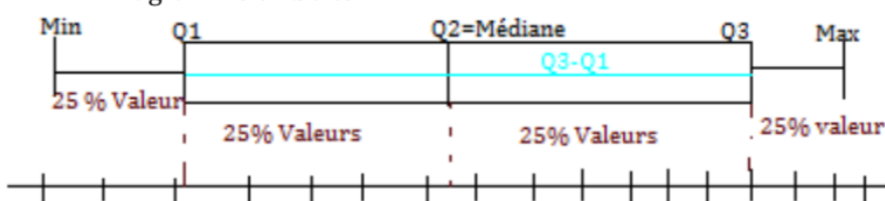


B_i =Borne inférieure, \bar{x} = moyenne, F = \sum fréquences absolues classe précédant médianes, a =amplitude classe médiane, n =nbr observation série (frq absolue de tout), f_{m_e} = fréquence classe médiane, E =Étendue, s =Écart-type, s^2 = variance, f_{q_n} = Fréquence absolue classe contenant premier quartile ou troisième				
Tendance centrale : Indiquer l'ordre de grandeur de la série et la position autour de laquelle se distribuent les données d'une série.	\bar{x} = moyenne arithmétique	$\bar{x} = \frac{x_1 + \dots + x_n}{n} = \frac{\sum x_i}{n}$	$\bar{x} = \frac{\sum_{i=1}^k f_i x_i}{n}$ x_i : p. milieu de classe	Variable quantitative, Affecté par valeurs aberrantes, Unique Pas de classes ouvertes
	M_e = Médiane	Impair: $m_e = \text{valeur de la } \left(\frac{n+1}{2}\right)^{\text{ième}} \text{ observation}$ Pair: $m_e = \frac{\left(\frac{n}{2}\right)^{\text{ième}} \text{ observation} + \left(\frac{n}{2}+1\right)^{\text{ième}} \text{ observation}}{2}$	Classe médiane = $n/2$ ème observation $M_e \approx B_i + \frac{(n-F)}{f_{M_e}} \cdot a$	Plus utilisée, centre de position, mieux évaluer asymétrie, unique, Aucune Val Aberrantes, avec classes ouvertes 50% supérieur, 50% inférieur
	M_o = Mode (Dépouillement)	Valeur la plus fréquente, 1 uni modale, 2 bimodale, 3+ multimodale	Valeur = Milieu de la classe modale	Variables qualitatives. Pas unique, compte pas toutes les valeurs, dépend des fréquences
Mesure de dispersion : détermine le niveau de fluctuations des données autour de la valeur centrale. Apprécié l'étalement de la série.	Variance s^2 Écart-type $s = \sqrt{s^2}$	$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}{n-1}$ $= \frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{n-1}$	$s^2 = \frac{\sum_{i=1}^k f_i (x_i - \bar{x})^2}{n-1}$ VAR & ECARTYPE	Plus utilisé, variance même valeur que moyenne et x_i , affecté par valeur aberrantes, comme se répartie autour moyenne, Val. Élevé = forte dispersion, val. Faible = homogène, se calcul pas si classe sont ouvertes
	Étendue	$E = X_{max} - X_{min}$. Ne tient compte que des valeurs extrêmes de la série, indépendant nbr série, peu utilisée lorsque plus de 10 val. Unique		
	Coefficient de variation (CV)	<i>Mesure de dispersion relative</i> $CV(\%) = \frac{s}{\bar{x}} \times 100$	La moyenne - on utilise la valeur absolue	Plus faible (<15) => homogène. Pas même unité de mesure. Représentatif ou volatile = CV plus faible
	Règle de Tchebycheff	$P(\bar{x} - ks < X < \bar{x} + ks) > 1 - \frac{1}{k^2}$	Pour tout ensemble de données et un nombre k , $k \geq 1$, la proportion de données comprise dans un intervalle est supérieure ou égale à $(1 - (1/k^2))$	
Mesures de positions : Permettent de situer une donnée par rapport à aux autres	Quartiles	<ol style="list-style-type: none"> Répartir les obs. Trier en ordre et indiquer quel groupe appartient. Si pas entier, Quartile correspond à la valeur du rang de nbr entier supérieur. Si entier, Q = Moyenne des données de rang ce nombre et le nombre entier immédiatement supérieur 	$C_p \approx B_i + \frac{(p \cdot n - F)}{f_{C_p}} \cdot a$ (p : pourcentage désiré)	Quartile (3) / Déciles(10) / Centile(100) $Q_1 \leq \frac{n}{4} \leq \frac{3n}{4}$ $Q_3 = \text{inverse } Q_1$ $Q_2 = \text{Médiane}$
	Intervalle interquartile	$IQ = (Q_3 - Q_1)$	Mesure l'étendue entre le premier et troisième quartile. Contient 50% des obs situé au milieu de la série	Données Aberrantes : Si $x < Q_1 - 1,5(Q_3 - Q_1) = \text{Abb.}$ Si $x > Q_3 + 1,5(Q_3 - Q_1) = \text{Abb.}$
Mesure de forme : Donnent une idée de la symétrie et de l'aplatissement d'une série.	Coefficient asymétrie (Skewness)	$S_k \approx \frac{3(\bar{x} - M_e)}{s} = \frac{\bar{x} - M_o}{s}$	$S_k = 0$: symétrique $\bar{x} = M_e = M_o$ $S_k < 0$ symétrie - $\bar{x} < M_e < M_o$ $S_k > 0$ symétrie + $M_o < M_e < \bar{x}$	Indique de degré de symétrie Coefficient de Pearson S_k est utilisé habituellement en 1 et -1
	Coefficient aplatissement (Kurtosis)	$a_4 = \frac{\sum (x_i - \bar{x})^4 / n - 1}{s^4}$	Courbe Si $\alpha_4 < 3$: aplatie Si $\alpha_4 = 3$: normale Si $\alpha_4 > 3$: aigüe	*Excel soustrait 3 distributions plus ou moins aplatie selon si fréquence. Proche des valeurs centrale
Mesures stats appropriés	Nominale (variable qualitative) : Mode Ordinale (variable quantitative ou qualitative) : Mode/Médiane Intervalle /rapport (variable quantitative) : Médiane, Moyenne, Mode, Étendue, Écart- type et Quartiles			

Diagramme en boîte



Tiges	Feuilles	Nombre de feuilles
Nbr sur les tiges	XXXX	Total feuille ligne
	XX	
	X	
		Grand total feuilles

Diagramme à tige et à feuille

Les étapes d'une étude stat

- 1) Analyse de la situation
- 2) Identification des informations à colliger
- 3) Choix du type de recherche et de la méthode de collecte de données
 - Recherche exploratoire
 Découvrir la nature de la situation à analyser afin de formuler ultérieurement des hypothèses (à l'aide d'entrevues, des fichiers sur les clients, ...)
 - Recherche descriptive
 Représenter graphiquement des données et les résumer par des mesures de tendance centrale et de dispersion
 - Recherche causale ou explicative
 Tester des hypothèses et déterminer la relation entre deux ou plusieurs variables
- 4) Identification des outils d'investigation et plan de sondage
- 5) Traitement statistique et codage des données
- 6) Rédaction du rapport de l'étude

Une étude de marché / sondage / échantillonnage consiste à prélever un sous-ensemble d'unités statistiques d'une population afin de mieux en connaître les caractéristiques (préférences, revenus, intention d'achat, ...) ou de faire des prévisions

- Moins coûteux qu'un recensement qui considère toutes les unités statistiques de la population
- Bien que plus précis, les recensements sont plus chers et prennent en général plus de temps

Question ouverte : Formulation laisse au répondant toute latitude réponse.
Question fermée : Formulation comporte liste de réponse

Stat. Descriptive : rend intelligible une série d'obs. et permet de dégager caractéristique essentielles (moyenne, médiane, écart-type, proportion) qui se dissimule dans une masse de donnée. Représent. : tableau, diagramme

La statistique : Science qui comporte un ensemble de méthodes et de théories appliquée à l'analyse de donnée. Comporte :

- + faire la collecte de données
- + organiser et résumer des données, la statistique descriptive
- + tirer des conclusions à partir de données, l'inférence statistique
- + prendre des décisions rationnelles, la théorie de la décision statistique

Échantillon : groupe restreint (ss-ensemble) unités stat. Nbr unité détermine taille éch.
Échantillon Aléatoire : ss-ensem d'individus tirés par hasard.

Unité statistique : Élément (Individu, objet) qui fait le sujet d'une ou plusieurs mesures.

Ensemble statistique (population stat.) : Ensemble de toutes les unités stats.

Variable statistique (caractère) : particularité étudiée (Étudiée ou dénombrable)

a) **-Qualitative** : Non mesurable. On ne peut qu'identifier un état (une nature).

-Quantitative : Soit mesuré par un nombre. Discrète; nbr limité de valeurs/ Continue ; toutes les valeurs finies ou non d'un intervalle.

b) **-Indépendante** (Facteur/ Variable explicative) : variable qui peut avoir un effet ou influencer une autre variable.

-Dépendante (Réponse / Variable): variable dont les variations peuvent être expliquées ou dépendent des variations d'une autre variable.

Modalité : Spécificité, état ou valeur qui lui sont propre peut avoir 1,2,3+

Fréquence absolue (effectif) : Nombre d'obs apparait à une modalité donnée (fi = n)

Fréquence relative (%) : = fréquence absolu ÷ nombre total des données

Fréquence cumulée : nombre d'observations dont la valeur est égale ou inférieure à cette modalité

Variation en % :
$$= \frac{y-x}{x} \times 100$$

Stat inférentielle : tirer des conclusions sur tout phénomène ou effectuer analyse statistique plus approfondie (estimation intervalle de confiance, analyse corrélation, etc.) Comporte marge d'erreur calculable.

Échelles de mesure :

Nominale : valeur numérique assignée arbitrairement à une modalité sans idée de distance entre les modalités

Ordinale : Permettent identifier modalité, mais aussi relation d'ordre entre les modalités. Permet d'apprécier le degré d'importance.

D'intervalle : informer distance (écart) entre deux modalités.

De rapport : zéro introduit dans l'échelle. Le plus utilisé ?

Fixer les classes :

1. Déterminer le nombre de classes; $5 < \text{nb cls} < 20$, souvent entre **6 et 12**

2. Règle de Sturge

$$k = 1 + 3,322 * \log_{10} n$$
 n = Nbr de données à dépouiller
 k = Nbr de classe à utiliser

3. Déterminer l'amplitude

$$a = E/k = \text{Étendue}/k$$

Tri : Mettre variable dans un certain ordre

Dépouillement des données : regroupement données en classe dans lequel on met un trait pour chaque données appartenant à sa classe respective. La somme du trait donne la fréquence absolue.

- Par valeur (≤ 12 modalités)
- Par classes
- Un diagramme en feuilles (les tiges : un élément commun à un groupe d'obs. et les feuilles : chacune des obs.)

Distribution fréquence absolue (distribution des effectifs) : Répartition des données accompagnée dans fréquences absolue respective

Dépouillement selon types de données			Représentation graphique selon type de données				
Qualitative	Nominale	Répartition des données sous forme de tableau de fréquence abs, relative et %	Quali	Nominale	Diagramme à barre (rectangle horizontaux), Diagramme à secteurs, Diagramme à barre groupé (Si 2 catégories)		
	Ordinale			Ordinale			
Quantitative	Ordinale	Fréquences	Quanti	Ordinale	Histogramme		
	Intervalle / Rapport	Valeurs Diagramme en feuille Distribution des fréquences absolues en classes		Intervalle / Rapport	Diagramme à bâton, diagramme en boîte, histogramme, polygone de fréquence, courbe cumulative		
Deux var. qualitatives	Nominale / Ordinale	Tableau croisé	Quali (2 var)	Nominale / Ordinale	Histogramme effet 3D	Quanti évoluant temps rapport	Courbe de tendance. Graphique linéaire
			Quanti (2 var)	Intervalle / rapport			

