Indices simples et quantiles (variable quantitative discrètes)

Introduction

- La construction d'indices statistiques vise à **résumer la distribution** de la variable, i.e. la **répartition des valeurs observées**.
- Les valeurs étant numériques, on peut alors s'intéresser à résumer la distribution de 2 façons :
  - la **localisation** : reflète un endroit de l'échelle autour duquel se situent les valeurs observées → **une « position » sur l'échelle**
  - la **dispersion** : renseigne sur l'éloignement des valeurs observées les unes par rapport aux autres. → **un « écartement » sur l'échelle**
- **Attention :** ces 2 types d'indices renseignent sur des notions très différentes.

Indices simples

Indice de localisation

- le **minimum** (min) : la valeur minimale observée → à partir d'où se situent les valeurs sur l'échelle
- le **maximum** (max) : la valeur maximale observée → jusqu'où se situent les valeurs sur l'échelle
- le **mode** (mode) : la valeur la plus fréquemment observée
  - valeur la plus répétée dans l'échantillon, correspondant à l'effectif (ou fréquence) le plus élevé dans le tableau de la distribution (bâton le plus haut dans le diagramme en bâtons)

- **Exemple :** Nombre de chèques à montant élevé traités par la banque au cours des 25 1ers jours de janvier 2021
  - min = 0
  - max = 5
  - mode = 2

Nb chèques	0	1	2	3	5
Effectifs	5	6	7	4	3

Indice de dispersion

- L'**étendue** (ét) : max - min.
  - éloignement entre la valeur **min** et la valeur **max**

- **Exemple :** Nombre de chèques à montant élevé traités par la banque au cours des 25 1ers jours de janvier 2014
  - ét = 5

Nb chèques	0	1	2	3	5
Effectifs	5	6	7	4	3

Indices basés sur les rangs : les quantiles

Introduction

- Les **quantiles** sont des **indices de localisation** basés sur les rangs.
- On commence par **ranger** les valeurs de la variable (et donc les individus associés) **par ordre croissant**.
- On note alors l'échantillon ordonné : $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$
  - le min est alors la première valeur dans ce classement : min = $x_{(1)}$
  - le max est alors la dernière valeur dans ce classement : max = $x_{(n)}$
- On peut alors proposer comme **indice de localisation** une valeur qui **sépare** l'échantillon avec une certaine proportion d'individus ayant des **valeurs plus petites**, et les autres individus ayant des **valeurs plus grandes**.

La médiane

- **Intuition :** la **médiane** est une valeur observable qui **partage en deux effectifs égaux** l'échantillon rangé par ordre croissant de la variable

- Exemple
  - Données brutes : 4, 18, 12, 9, 7, 22, 10, 3, 6, 17, 14
  - Données triées : 3, 4, 6, 7, 9, **10**, 12, 14, 17, 18, 22

La médiane

- La médiane (Med) doit vérifier simultanément les 2 propriétés suivantes :
  - Au moins 1 individu sur 2 dans l'échantillon a une valeur inférieure ou égale à Med
  - Au moins 1 individu sur 2 dans l'échantillon a une valeur supérieure ou égale à Med

Exemple : Nombre de chèques à *montant élevé* traités par la banque au cours des 25 premiers jours de janvier 2014 :

Nb chèques	0	1	2	3	5
Effectifs	5	6	7	4	3

$$Med = 2$$

Alors

$$\text{fréq}(observations \leq 2) = \frac{18}{25} = 0.72 \geq \frac{1}{2}$$

$$\text{fréq}(observations \geq 2) = \frac{14}{25} = 0.56 \geq \frac{1}{2}$$

La médiane

- **Nombre d'observations n impair :** la médiane est la ((n+1)/2)^e observation de la série ordonnée

$$\text{Méd} = x_{\left(\frac{n+1}{2}\right)}$$

- Exemple : 3, 5, 6, **8**, 10, 12, 14

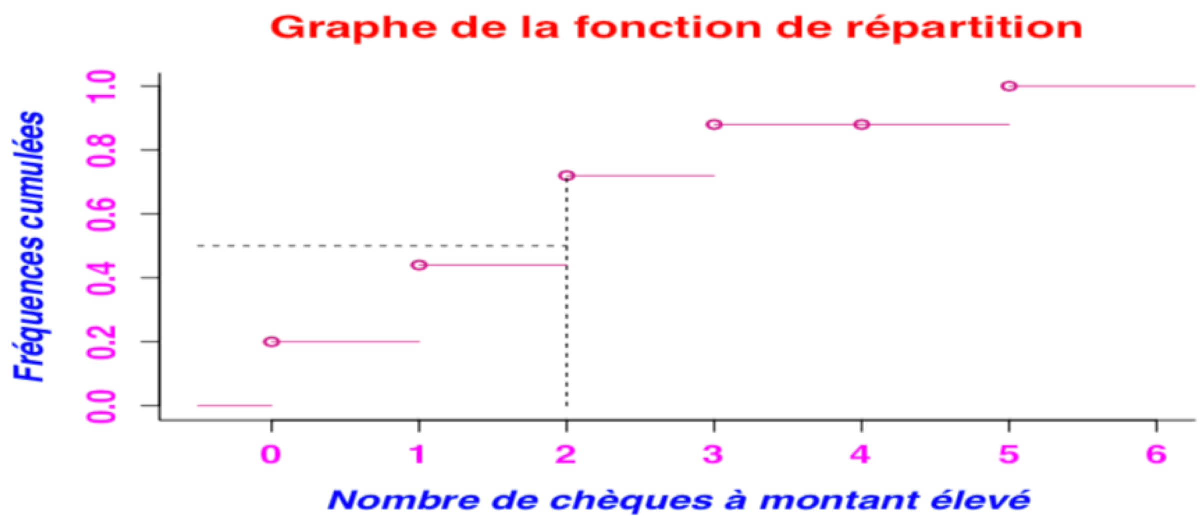
- **Nombre d'observations n pair :** la médiane est la valeur centrale entre la (n/2)^e et la (n/2+1)^e observation de la série ordonnée
- Exemple : 5, 9, 16, 18, **19, 21**, 24, 25, 26, 43

$$Med = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$$

La médiane

- on cherche l'abscisse du ou des points d'ordonnée 0.5 (ou dépassant pour la première fois 0.5)

Exemple : Graphe de la fonction de répartition en janvier 2014



Les quartiles

- Il y en a 3. Ils sont notés : Q_1, Med, Q_3

- **Intuition :** les 3 valeurs Q1, Med, Q3 sont des **valeurs observables** qui partagent en **quatre effectifs égaux** l'échantillon ordonné.