*Statistique descriptive*

1. Types de variables
2. Variable qualitative

* Nominale : variable à plusieurs modalités **non mesurables** qui s’excluent mutuellement

Les modalités s’expriment par des noms et ne sont pas hiérarchisées.  
Les individus sont caractérisés par leur appartenance à 1 SEULE modalité.

|  |  |
| --- | --- |
| Variable qualitative | Modalités |
| Couleur cheveux  Fumeur ?  Type habitation  Sexe | Blond – Brun – Roux – Châtain – …  Oui – Non  Maison – Loft – Studio – Villa – …  ♀ – ♂ |

* Ordinale : variable où les modalités sont les degrés d’un état, hiérarchisés sans qu’ils ne résultent d’une mesure.

|  |  |
| --- | --- |
| Variable qualitative | Modalités |
| Douleur | 0 – 2 – 4 – 6 – 8 – 10  Aucune – Inconfort – … – Intense – Insupportable |
| Echelle de Lickert | -2 / -1 / 0 / 1 / 2  Désaccord – pas d’accord – neutre – d’accord – totalement d’accord |

1. Variable quantitative

Une variable quantitative est le résultat d’une mesure ou d’un comptage. On distingue 2 catégories secondaires :

* Discrète : elles ne peuvent prendre qu’un nombre fini de valeurs, elles sont souvent issues d’un comptage. (exp : nombre d’enfants)
* Continue : elles peuvent prendre un nombre infini de valeur, issues d’une mesure effectuée avec un instrument. (exp : taille d’une personne)

En réalité, le nombre de valeurs possibles pour un caractère donné dépend de la précision de mesure. On peut considérer continu un caractère discret qui peut prendre un grand nombre de valeurs. (exp : nbr de globules rouges par mL de sang).

1. Représentation des données

Une série statistique σ correspond aux différentes modalités d’un caractère sur un échantillon d’individus appartenant à une population donnée.   
(exp : σgroupe sanguin : A A B O O AB AB A B B O A AB …)

1. Tableaux de distribution
2. Variables qualitative nominale

Soit une modalité d’une série statistique.

Pour tout , on détermine l’effectif  : le nombre de sujets présentant la modalité.   
Les modalités doivent être mutuellement exclusives, donc l’effectif total du groupe étudié est égal à la somme des effectifs de chaque modalité.  
On appelle fréquence de la modalité est la valeur .

Avec le nombre de modalité de la variable et l’effectif total. On rassemble les résultats d’une distribution des fréquences sous forme d’un tableau.

|  |  |  |
| --- | --- | --- |
| Modalités | Effectif | Fréquence |
|  |  |  |
| Total |  | 1 |

Tableau de distribution des fréquences Tableau de répartition des groupes sanguins dans un hôpital

|  |  |  |
| --- | --- | --- |
| Groupes | Effectif | Fréquence |
| O  A  B  AB | 45  40  10  5 | 0.45  0.4  0.1  0.05 |
| Total | 100 | 1 |

On définit l’angle de la modalité par .

Les variables qualitatives ne permettent pas d’effectuer les calculs usuels (moyenne, variance, …).   
2 variables P et T peuvent être mesurées sur le même individu. Les valeurs obtenues sont placées dans un tableau à double entrée dit « tableau de contingence ».

Tableau de contingence

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | T | | | | | Somme |
| P |  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| Somme |  |  |  |  |  |  |

 : effectifs des individus possédant à la fois la modalité de la ligne et colonne .

1. Variable qualitative ordinale et variable quantitative discrète

Soit , on appelle :

* Fréquences cumulées croissantes .
* Fréquences cumulées décroissantes .

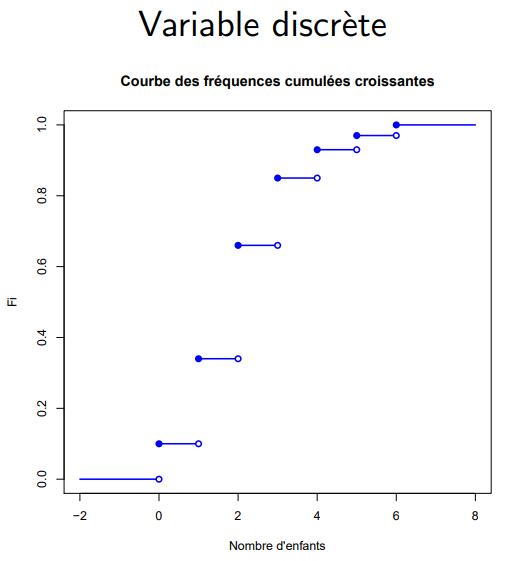
Tableau de distribution des effectifs et fréquences cumulés

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Valeur des modalité | Effectif | Fréquence | Effectif cumulé | Fréquence  cumulée | Fréquence cumulée |
|  |  |  |  |  | 1 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| 0  1  2  3  4  5  6 | 10  24  32  19  8  4  3 | 0.1  0.24  0.32  0.19  0.08  0.04  0.03 | 10  34  66  85  93  97  100 | 0.1  0.34  0.66  0.85  0.93  0.97  1 | 1  0.9  0.66  0.34  0.15  0.07  0.03 |

Nombre d'enfants dans les familles

Nombres d’enfants

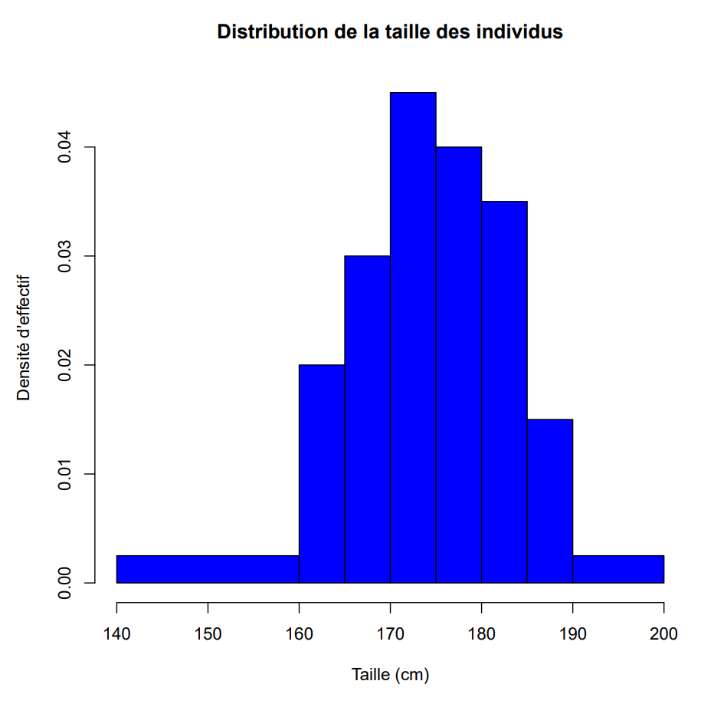
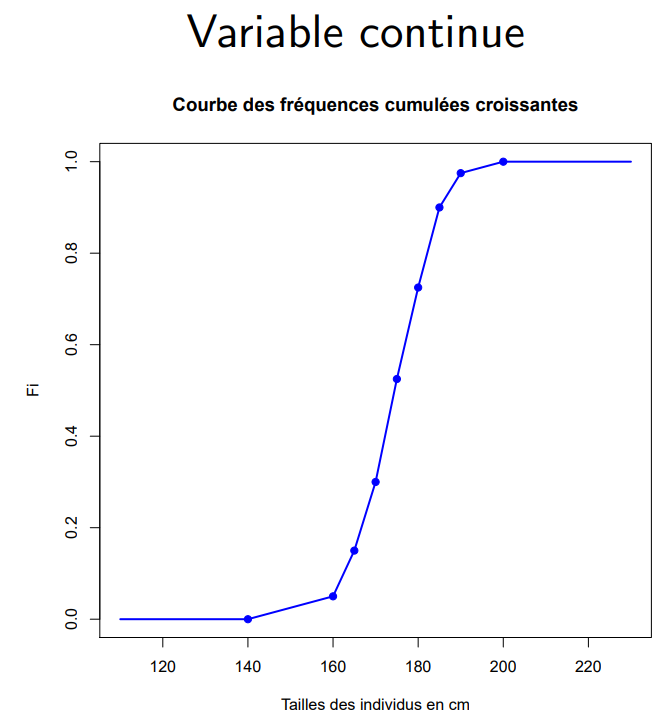


1. Variables quantitatives continues

Il est nécessaire de regrouper en classes les valeurs prises par la variable.  
(exp : taille )

L’intervalle de classe (également appelé amplitude) d’une classe vaut .  
En général on choisit des classes de même amplitude.

Si l’amplitude n’est pas constante, il faut calculer la densité .   
La densité de fréquence permet de comparer les fréquences d’une classe à l’autre.

Tailles des individus en cm

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Classes |  |  |  |  |  |
| [140,160[  [160,165[  [165,170[  [170,175[  [175,180[  [180,185[  [185,190[  [190,200[ | 10  20  30  45  40  35  15  5 | 0.05  0.1  0.15  0.225  0.2  0.175  0.075  0.025 | 0.0025  0.02  0.03  0.045  0.04  0.035  0.015  0.0025 | 10  30  60  105  145  180  195  200 | 0.05  0.15  0.3  0.525  0.725  0.9  0.975  1 |

1. Indicateurs numériques
2. De positions

* Mode d’une distribution : valeur la plus fréquente de la distribution.
* Classe modale : classe dont la densité d’effectif est la plus élevée.

On attribue au mode la valeur centrale de cette classe.

* Moyenne arithmétique : .
* Médiane : valeur qui partage la série en 2 groupes de mêmes effectifs.

Si la taille de l’échantillon est grande, on utilise un graphique ou une interpolation linéaire

* Quartiles : valeurs qui partagent la série ordonnée en 4 groupes de même effectif

Le 2ème quartile est la médiane.

* Percentiles : valeurs qui partagent la série ordonnée en 100 groupes de même effectif

|  |  |  |
| --- | --- | --- |
|  | Avantages | Inconvénient |
| Mode | * Intéressant si distributions asymétriques * Bon indicateur de la population hétérogène * Non influencé par les valeurs extrêmes | * Se prête mal aux calculs statistiques * Très sensible aux variations d’amplitude des classes * Ne tient compte que des individus proches de la classe modale |
| Moyenne | * Se prête bien aux calculs et tests * + de sens si répartition symétrique et dispersion faible | * Très sensible aux valeurs extrêmes * Ne convient pas aux valeurs ordinales * Représente mal une population hétérogène |
| Médiane | * Moins sensible aux valeurs extrêmes * Utilisable avec les variables ordinales * Peu sensible aux variations d’amplitude des classes | * Se prête mal aux calculs statistiques * Suppose une répartition équitable des valeurs * Classement long si beaucoup de valeurs |

1. De dispersion

* Valeurs extrêmes : min et max des valeurs
* Etendue : max-min
* Ecart absolu moyen :
* Distance interquartile : 3ème quartile – 1er quartile
* Variance observée :
* Ecart-type :
* Coefficient de variation :

