



FACULTE DE MEDECINE DE CONSTANTINE

DEPARTEMENT DE MEDECINE

ENSEIGNEMENT GRADUE 6^{ème} Année 2012-2013

Intitulé du cours : **STATISTIQUE DESCRIPTIVE**

Présenté par **Dr LEMDAOUI MOHAMED CHERIF**
Docent en Epidémiologie

PLAN

I- INTRODUCTION

II- POPULATION ET VARIABLES

- 1- Population
- 2- Unité Statistique
- 3- Variable ou Caractère
- 4- Modalité
- 5- Caractère qualitatif
- 6- caractère quantitatif

III- REGROUPEMENT DES DONNEES

- 1- Série statistique
- 2- Tableau brut des données
- 3- Tableau de fréquence

IV- REPRESENTATION GRAPHIQUE

- 1- Histogramme
- 2- Polygone de fréquence
- 3- Diagramme en bâtonnets ou en barres
- 4- Diagramme en barres horizontales
- 5- Diagramme circulaire ou à secteurs
- 6- Pyramide

V- MESURES DE TENDANCE CENTRALE OU PARAMETRES DE POSITION ET PARAMETRES DE DISPERSION

- 1- Paramètres de position
 - Moyenne Arithmétique
 - Médiane
 - Quartiles
 - Déciles et percentiles
 - Mode
- 2- Mesures de dispersion
 - Etendue
 - Ecart type ou déviation standard
 - Variance
 - l'intervalle semi interquartile
 - Coefficient de variation

VI- CONCLUSION

Références bibliographiques

I- INTRODUCTION

Le mot statistique vient du mot latin « STATUS » qui signifie situation ou état.

Une statistique est un ensemble de nombres présentés sous forme de tableaux ou de graphiques et concernant un événement donné.

Donc la statistique est une méthode d'analyse et de raisonnement scientifique, mathématique appliquée à un ensemble de données et qui permet des généralisations afin d'établir des indicateurs, des indices, des liaisons et des prévisions.

- La statistique descriptive est un instrument de description visant à simplifier, résumer et synthétiser, tout en veillant à ne pas altérer la fidélité de l'information.

Il s'agit d'organiser et résumer des observations. On ne fait pas de comparaisons et on s'intéresse en général à un seul groupe, échantillon ou population. Exemple : Lorsqu'on calcule la fréquence d'une maladie dans une population donnée, on fait de la statistique descriptive .

- La statistique inférentielle est une partie de la statistique qui, contrairement à la statistique descriptive, ne se contente pas de décrire des observations, mais extrapole les constatations faites à un ensemble plus vaste, permet de tester des hypothèses sur cet ensemble, et de prendre des décisions le concernant. Exemple : Lorsqu'on cherche à étudier si la différence entre la fréquence d'une maladie dans deux groupes distincts (exposés ou non à un facteur) est significative ou bien si elle est imputable au hasard, on fait de la statistique inférentielle .

- POPULATION ET VARIABLES

1- Population

La population désigne un ensemble d'unités statistiques.

.Ces unités statistiques peuvent différer les uns des autres dans une même population, un même groupe, une même famille,etc, et ces différences ne se résument pas à de simples erreurs de mesures, car chacun diffère des autres par ces gènes, par son environnement, par des hasards de la vie .

2- Unité Statistique

C'est l'unité sur laquelle va porter l'observation.

3- Variable ou Caractère

Un individu donné de la population peut être étudié selon certaines propriétés. Ces propriétés sont appelées caractères ou variables statistiques.

Elles diffèrent d'un état à un autre et d'une unité d'observation à une autre.

4- Modalité

C'est une éventualité que peut prendre un caractère. Chaque caractère possède deux ou plusieurs modalités.

5- Caractère qualitatif :

Il s'exprime par des qualificatifs, il est nominal

Une variable qualitative est une variable qui ne prend pas de valeur numérique. Exemple : sexe, nationalité, groupe sanguin, profession, couleur des yeux...etc . Chaque variable qualitative a plusieurs modalités. Exemple : pour la variable état matrimonial, les modalités sont célibataire, marié, veuf, divorcé.

Une variable qualitative ordinale (échelle ordinale) prend des valeurs qui sont ordonnées, hiérarchisées. On peut classer les modalités les unes par rapport aux autres .

Exemple : évolution d'une maladie = état de santé du patient

⇒ amélioration, stabilité ou détérioration

Il existe un type particulier de variables qualitatives, ce sont les variables binaires qui peuvent prendre que deux valeurs on les appelle aussi variables dichotomiques car elles partagent la population en deux parties ou variables booléennes pouvant prendre la valeur « vrai » ou « faux » ou variables de Bernouilli codées respectivement 1 et 0

6- caractère quantitatif :

Une variable quantitative est une variable qui prend des valeurs de nature numérique.

Variable quantitative discrète ou discontinue

Une variable quantitative discrète peut prendre des valeurs dénombrables, isolées séparées les unes des autres. Exemple : le nombre d'enfants d'un ménage

Variable quantitative continue

Une variable quantitative continue peut prendre théoriquement toutes les valeurs à l'intérieur d'un intervalle entre deux modalités Exemple : la taille, l'âge, la glycémie, la pression artérielle ..etc..

N'oublions pas que dans le cadre du caractère quantitatif continu et parfois discontinu, on représente généralement les variables sous forme de classe pour éviter une répartition des fréquences trop dispersées et permettre une présentation commode.

Une classe est caractérisée par son centre et son amplitude.

Les classes doivent être mutuellement exclusives et collectivement exhaustives.

Cette échelle de classification de l'information est une échelle dite par intervalle.

III- REGROUPEMENT DES DONNEES

Généralement le regroupement des données se fait sous forme d'un tableau de fréquence à partir d'une série statistique ou d'un tableau brut de données.

1- Série statistique

C'est l'ensemble des données résultant du dépouillement d'une observation d'une population ou d'un groupe

2- Tableau brut des données

C'est un tableau élémentaire de travail. Toutes les données y figurent, unité par unité et variable par variable (tableau excel...)

3- Tableau de fréquence

C'est une matrice à double entrée

Colonne type ←←← ← ← ←Entêtes des colonnes →→→→→→→
 ↓

- *****_	*****	*****	*****
*****	rubrique	Rubrique	Rubrique
*****	rubrique	Rubrique	Rubrique

Horizontalement on a les lignes

Verticalement on a les colonnes→→

Entre l'entête des colonnes et la colonne type, l'ensemble des rubriques forme le corps du tableau.

Le titre du tableau, en haut, comporte trois informations essentielles, à savoir, le quoi, le quand et le ou, en plus de la source des données.

Tableau de Fréquence

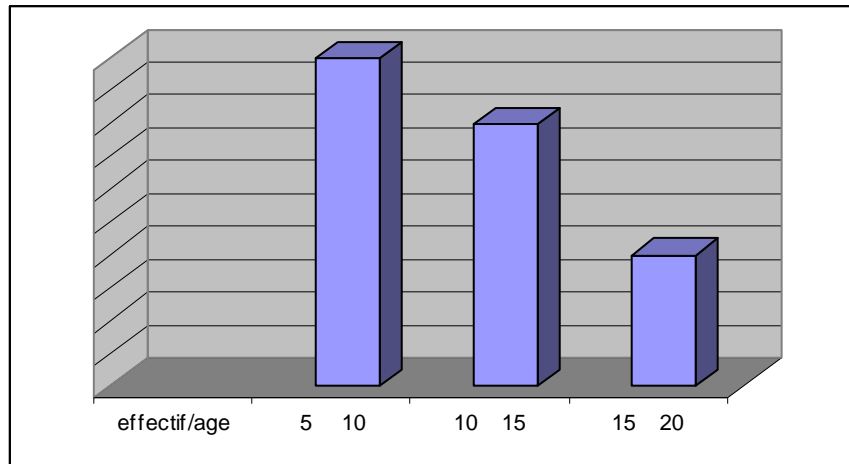
Variable/Effectif	Fréquence absolue	Fréquence relative	Pourcentage
Total	= effectif totale	= 1	= 100%

IV- REPRESENTATION GRAPHIQUE

1- Histogramme

Il provient du mot grec « HISTOS » = tissu ou voile. C'est un diagramme en surface contiguës. Ce mode de représentation graphique est utile pour les distributions de variables quantitatives continues et discontinues.

Répartition de la fréquence d'une variable donnée selon l'âge



2- Polygone de fréquence

C'est une ligne polygonale, construite à partir de l'histogramme .

Adapté à la représentation d'une variable quantitative continue mais aussi discontinue. Il convient généralement mieux que l'histogramme quand il s'agit de représenter plusieurs distributions de fréquences sur un même système d'axes

3- Diagramme en bâtonnets ou en barres

Ce sont des surfaces non contiguës. Utilisées surtout pour les caractères qualitatifs nominaux ou ordinaux.

4- Diagramme en barres horizontales .

Adapté au variables qualitatives nominales

5- Diagramme circulaire ou à secteurs

En par de gâteau, en camembert. Il est adapté à la représentation d'une seule distribution d'une variable qualitative nominale.

Le camembert peut être choisi à la place d'un diagramme en barres à condition que le nombre de classes soit faible .

6- Pyramide

Il représente un double histogramme inversé et juxtaposé, pour la distribution d'une population selon l'âge et le sexe en démographie.

V- MESURES DE TENDANCE CENTRALE OU PARAMETRES DE POSITION ET PARAMETRES DE DISPERSION

La répartition du poids à la naissance en kilogramme de trente enfants né le 27 novembre 1993 à la maternité du CHUC a donnée les résultats suivants sous forme de tableau.

Variable /Eff	Centre xi	FA(ni)	FR	%	x-m	(x-m) ²	(x-m) ² .ni
2,2 2,6	2,4	4	0,133	13,3	-0,68	0,46	1,84
2,6 3,0	2,8	6	0,2	20	-0,28	0,078	0,47
3,0 3,4	3,2	15	0,5	50	0,12	0,014	0,22
3,4 3,8	3,6	5	0,166	16,6	0,52	0,27	1,35
Total		=30	=1	100%			3,88

Les paramètres de position et de dispersion représente un ensemble de valeurs caractéristiques qui permettent une représentation condensée de l'information dans la série statistique c'est-à-dire elles permettent de localiser le centre d'une distribution de fréquences et d'évaluer la variabilité dans le groupe étudié.

1- Paramètres de position

- Moyenne Arithmétique

C'est la somme de chacune des valeurs observées d'une variable donnée sur le nombre de valeurs observées (effectif).

$$M = \frac{X_1 + X_2 + \dots + X_n}{N} = \frac{\sum_{i=1}^n X_i}{N}$$

Mais souvent les valeurs X_1, X_2, \dots se répètent \Rightarrow on a alors la moyenne arithmétique pondérée.

$$M = \frac{\sum_{i=1}^k n_i X_i}{N}$$

- Médiane

C'est une valeur qui divise l'ensemble des valeurs observées disposées en ordre croissant ou décroissant, en deux parties égales. C'est-à-dire théoriquement de même fréquence (50% inférieures et 50% supérieures).

La médiane est une valeur de rang $n+1/2$

Elle n'est pas influencée par les valeurs extrêmes contrairement à la moyenne arithmétique. De ce fait on pourrait utiliser la médiane pour décrire une tendance centrale d'une distribution fortement asymétrique.

Elle est utilisée dans le cadre des variables quantitatives.

- Quartiles

Il divisent la distribution en quatre parties. Il sont au nombre de trois :

* premier quartile : valeur de la variable où 25% sont inférieures et 75% sont supérieures.

* deuxième quartile = médiane

* troisième quartile : valeur de la variable où 75% sont inférieures et 25% supérieures

- Déciles et percentiles

* Déciles : il partage la distribution en dix. Ils sont au nombre de neuf
Chaque partie comprend 10% des effectifs

* Percentiles : valeurs partageant la distribution d'une variable en 100 groupes. On a :

- percentile 10% = premier décile (10% inf et 90% sup)
- percentile 25% = premier quartile
- percentile 50% = médiane
- percentile 90% = neuvième décile (90% inf et 10% sup)

- Mode

C'est la valeur qui revient le plus souvent, c'est-à-dire la valeur de la variable qui a le plus grand effectif, c'est la valeur dominante.

On parle de distribution uni modale quand il y a un seul mode et bimodale en cas de deux modes.

2- Mesures de dispersion

- Etendue

C'est la différence entre la plus grande et la plus petite des valeurs observées.

- Ecart type ou déviation standard

Mesure la dispersion autour de la moyenne arithmétique

$$\sigma = \text{racine de carrée de } \frac{\sum_{i=1}^N (x_i - M)^2 \cdot n_i}{N}$$

- Variance

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - M)^2 \cdot n_i}{N}$$

Intervalle de normalité = $M \pm 2\sigma$

La variance apparaît comme une mesure de dispersion intéressante pour quantifier la variabilité et permettre ainsi de conclure sur l'homogénéité ou non d'une population ou d'un groupe par rapport à différentes variables.

La variance qui représente donc la dispersion des observations autour de la moyenne est aussi influencée par les valeurs extrêmes ce qui fait qu'on lui préfère parfois la dispersion autour de la médiane tel que :

- l'intervalle semi-interquartile $Q = (Q_3 - Q_1) / 2$

Plus les valeurs observées sont concentrées plus les quartiles Q_1 et Q_3 sont rapprochés, plus Q est petit, témoignant d'une faible variabilité.

- Coefficient de variation

On l'utilise pour comparer deux distributions de variables différentes.

Quand les unités sont différentes on définit une mesure relative de dispersion appelée : le Coefficient de variation qui est égal à :

$$CV = \frac{\sigma}{M} * 100$$

VI- CONCLUSION

La statistique descriptive trouve donc sa place dans le cadre de l'étude des phénomènes et des événements de santé. Elle permet le regroupement, l'organisation et la réduction des données sous forme de paramètres qui facilitent la mesure des caractères étudiés dans une population. Cette simplification est primordiale et peut aboutir à des pistes de recherche aussi bien dans le domaine épidémiologique, clinique que biologique.

Références Bibliographiques

- 1- Anahita Basirat, Initiation aux statistiques descriptives .cours
Vocabulaire de base : Statistique descriptive , Statistique inférentielle .
anahita.basirat@gipsa-lab.inpg.fr Janvier 2009
- 2- Epidémiologie ; Principes et méthodes quantitatives. Notions de base en statistique p 162, Editions INSERM 1995
- 3- Marie Bernard .P,Lapointe .C, Mesures Statistiques en Epidémiologie.
Variables et échelles de classification, p4-8 ; Mesures de base p 24-41 ;
Presses de l'université du Québec 1987
- 4- Ancelle T, Statistique Epidémiologie, chapitres : 1-2-3-45, P 4-40
Maloine, 2002

