Programme d'appui budgétaire à la mise en œuvre du cadre stratégique de lutte contre la pauvreté en Mauritanie-SBC CSLP IIIN° MR/FED/22576

FORMATION DES CADRES DU SYSTEME STATISTIQUE NATIONAL DE LA REPUBLIQUE ISLAMIQUE DE MAURITANIE

Statistique Descriptive et Analyse des Données Démographiques

MANUEL DE FORMATION

Table des matières

Module1 : Statistique Descriptive

Chapitre 1. Définitions fondamentales

La science statistique est un ensemble des méthodes scientifiques de la collecte de données, leur analyse, leur traitement, l'interprétation des résultats et leur présentation afin de rendre les données compréhensibles par tous. L'objectif de la Statistique Descriptive est de décrire de façon synthétique et parlante des données observées pour mieux les analyser. Le terme «statistique» est issu du latin «statisticum», c'est-à-dire qui a trait à l'État. Ce terme a été utilisé, semble-t-il pour la première fois, à l'époque de Colbert, par Claude Bouchu, intendant de Bourgogne, dans une «Déclaration des biens, charges, dettes et statistiques des communautés de la généralité de Bourgogne de 1666 à 1669 ».

Par contre, l'apparition du besoin « statistique » de posséder des données chiffrées et précises, précède sa dénomination de plusieurs millénaires. À son origine, il est le fait de chefs d'États (ou de ce qui en tient lieu à l'époque) désireux de connaître des éléments de leur puissance : population, potentiel militaire, richesse, . . .

Avec le temps l'importance de la statistique s'augmente au jour le jour, il s'applique à la plupart des disciplines : agronomie, biologie, démographie, économie, sociologie, linguistique, psychologie, . . .

1.1 Population

En statistique, on travaille sur des populations. Ce terme vient du fait que la démographie, étude des populations humaines, a occupé une place centrale aux débuts de la statistique, notamment au travers des recensements de population. Mais, en statistique, le terme de population s'applique à tout objet statistique étudié, qu'il s'agisse d'étudiants (d'une université ou d'un pays), de ménages ou de n'importe quel autre ensemble sur lequel on fait des observations statistiques. Nous définissons la notion de population.

Définition 1 :

On appelle population l'ensemble sur lequel porte notre étude statistique. Cet ensemble est noté Ω.

Exemple 2
– On considère l'ensemble des étudiants de. Dans ce cas

$$\Omega = \textit{ensemble des étudiants.}$$

– Si l'on s'intéresse maintenant à la circulation automobile dans une ville, la population est alors constituée de l'ensemble des véhicules susceptibles de circuler dans cette ville à une date donnée. Dans ce cas

$$\Omega = \textit{ensemble des véhicules.}$$

Une population est composée d'individus. Les individus qui composent une population
statistique sont appelés unités statistiques.

Définition 3
On appelle individu tout élément de la population, il est noté ! (! dans).

Définition 2:

On appelle individu tout élément de la population Ω, il est noté ω (ω dans Ω).

1.2. Mesure et variable

- On s'intéresse à des unités statistiques ou unités d'observation : par exemple des individus, des entreprises, des ménages. En sciences humaines, on s'intéresse dans la plupart des cas à un nombre fini d'unités.
- Sur ces unités, on mesure un caractère ou une variable : le chiffre d'affaires de l'entreprise, le revenu du ménage, l'âge de la personne, la catégorie socioprofessionnelle d'une personne. On suppose que la variable prend toujours une seule valeur sur chaque unité. Les variables sont désignées par simplicité par une lettre (X, Y, Z).
- Les valeurs possibles de la variable, sont appelées modalités.
- L'ensemble des valeurs possibles ou des modalités est appelé le domaine de la variable.

1.3 Typologie des variables

- Variable qualitative : La variable est dite qualitative quand les modalités sont des catégories.
- Variable qualitative nominale : La variable est dite qualitative nominale quand les modalités ne peuvent pas être ordonnées.
- Variable qualitative ordinale : La variable est dite qualitative ordinale quand les modalités peuvent être ordonnées. Le fait de pouvoir ou non ordonner les modalités est parfois discutable. Par exemple : dans les catégories socioprofessionnelles, on admet d'ordonner les modalités : 'ouvriers', 'employés', 'cadres'. Si on ajoute les modalités 'sans profession', 'enseignant', 'artisan', l'ordre devient beaucoup plus discutable.
- Variable quantitative : Une variable est dite quantitative si toutes ses valeurs possibles sont numériques.
- Variable quantitative discrète : Une variable est dite discrète, si l'ensemble des valeurs possibles est dénombrable.
- Variable quantitative continue : Une variable est dite continue, si l'ensemble des valeurs possibles est continu.

Exemple 1. Les modalités de la variable nombre d'enfants par ménage sont : 0,1,2,3,4,5,. C'est une variable quantitative discrète.

1.4 Série statistique

On appelle série statistique la suite des valeurs prises par une variable X sur les unités d'observation.
Le nombre d'unités d'observation est noté **n**.
Les valeurs de la variable X sont notées **x1, . . . , xi, . . . , xn.**

Exemple 1.3 On s'intéresse à la variable 'état-civil' notée *X* et à la série statistique des valeurs prises par *X* sur 20 personnes. La codification est :

C : célibataire,
M : marié(e),
V : veuf (ve),
D : divorcée.

Le domaine de la variable *X* est *{C,M, V,D}*. Considérons la série statistique suivante:

M M D C C M C C C M
C M V M V D C C C M

Ici, $n = 20$.

$x1 = M, x2 = M, x3 = D, x4 = C, x5 = C, \dots, x20 = M.$

Une variable qualitative nominale a des valeurs distinctes qui ne peuvent pas être ordonnées. On note J le nombre de valeurs distinctes ou modalités.

Les valeurs distinctes sont notées $x1, \dots, xj, \dots, xJ$. On appelle ***effectif*** d'une modalité ou d'une valeur distincte, le nombre de fois que cette modalité (ou valeur distincte) apparaît. On note nj l'effectif de la modalité xj.

La fréquence d'une modalité est l'effectif divisé par le nombre d'unités d'observation.

$$f_j = \frac{n_j}{n} \quad , \text{ ou } j=1,\dots\dots\dots\dots J$$

Exemple 1.4 Avec la série de l'exemple précédent, on obtient le tableau statistique

x_j	n_j	f_j
C	9	0.45
M	7	0.35
V	2	0.10
D	2	0.10

1.5 Représentation des données

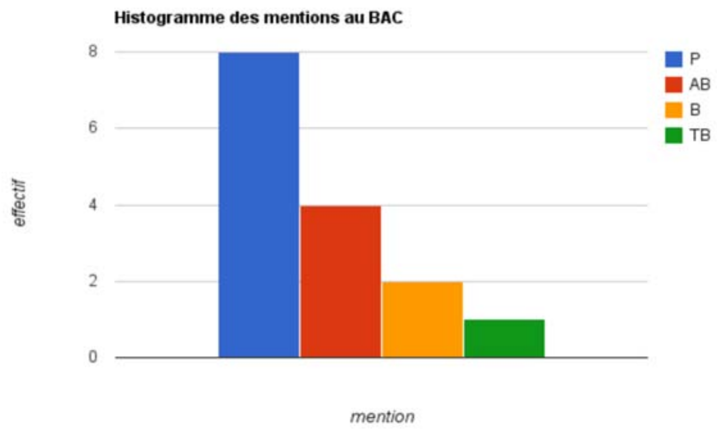
Pour un groupe de 15 étudiants, on a observé les valeurs des variables: Couleur des Yeux, Mention au Bac et Note à l'Examen de Statistique; ainsi le tableau de données suivant a été obtenu. Ces données seront souvent utilisées dans ce chapitre.

Individus	Mention au Bac	Taille de l'individu (en cm)	Note à l'Examen de Statistique
1	P	165	12
2	AB	178	8
3	P	190	13
4	P	172	11
5	AB	157	10
6	P	180	9
7	B	170	16
8	P	150	9
9	AB	163	12
10	B	157	13
11	P	173	18
12	AB	145	15

13	TB	182	12
14	P	158	17
15	P	139	19

Etudions l'exemple de la variable **Mention au Bac**. On commence d'abord par compter le nombre d'individus appartenant à chacune des modalités de cette variable : nP = 4 individus, nB = 3 individus, nAB = 3 individus.

Mention	P	AB	B	TB
Effectif	8	4	2	1



On constate que les étudiants sont répartis inégalement entre les différentes modalités de la variable Mention au Bac. Une première façon d'apprécier la répartition d'une variable est de construire un tableau de répartition des effectifs et des fréquences entre les différentes valeurs possibles de la variable. De façon générale, la fréquence d'une modalité « M » d'une variable qualitative se calcule au moyen de la formule suivante :

fM = (fréquence de la modalité «M» d'une variable qualitative)=(effectif correspondant à « M »)/
(effectif total):

On a de plus,
pM = (pourcentage des individus correspondant à la modalité « M ») = fM X 100.

On a enfin :

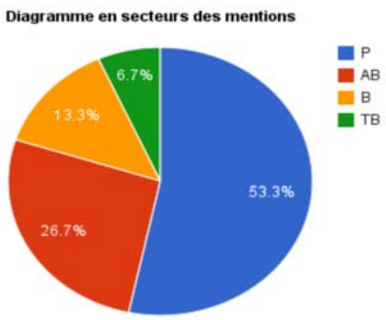
(somme des fréquences de toutes les modalités d'une variable qualitative) = 1

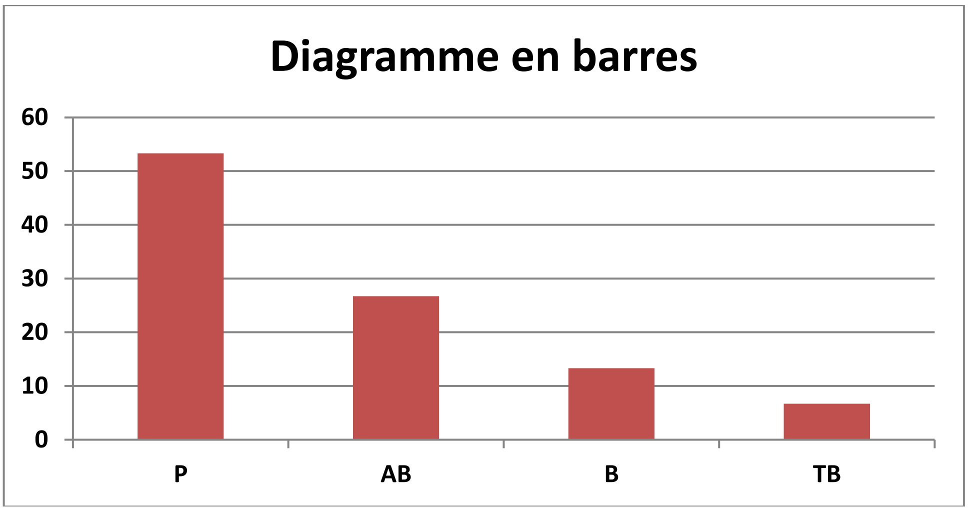
(somme de tous les pourcentages correspondant aux modalités d'une variable qualitative) = 100.

Tableau de Répartition de la variable Mention au Bac

Mention	Effectif	Fréquences	Pourcentage
P	np=8	Fp=8/15=0,533	53,3%
AB	nAB=4	FAB=4/15=0,267	26,7%
B	nB=2	FB=2/15=0,133	13,3%
TB	nTB=1	FTB=1/15=0,067	6,7%
	effectif total N = 15		

On peut représenter les pourcentages ou les fréquences à l'aide d'un diagramme en secteurs (piechart) ou un diagramme en barres des fréquences.





Chapitre 2 : Indicateurs de tendance centrale

2.1. Le mode

Le mode est la valeur distincte correspondant à l'effectif le plus élevé ; il est noté *xM.*

Si on reprend la variable 'Etat civil' , dont le tableau statistique est le suivant :

x_j	n_j	f_j
C	9	0.45
M	7	0.35
V	2	0.10
D	2	0.10
n=1		1

Le mode est *C* : célibataire.

Remarque 2.1

– Le mode peut être calculé pour tous les types de variable, quantitative et qualitative.
– Le mode n'est pas nécessairement unique.
– Quand une variable continue est découpée en classes, on peut définir une classe modale (classe correspondant à l'effectif le plus élevé).

2.2. La moyenne

La ***moyenne*** ne peut être définie que sur une variable ***quantitative***. Plusieurs types des moyennes sont connus parmi ceux la moyenne arithmétique qui est le plus utilisée.

a. La moyenne arithmétique

La moyenne arithmétique est la somme des valeurs observées divisée par leur nombre, elle est notée Xbar :

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + \dots\dots\dots + x_J}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

La moyenne peut être calculée à partir des valeurs distinctes et des effectifs :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} n_i . x_i$$

Exemple 2.1 Les nombres d'enfants de 8 familles sont les suivants 0, 0, 1, 1, 1, 2, 3, 4.

La moyenne est :

$$\overline{\boldsymbol{x}}=\frac{0+0+1+1+1+2+3+4}{8}=\frac{12}{8}=1,5$$

On peut aussi faire les calculs avec les valeurs distinctes et les effectifs. On considère le tableau :

x_i	n_i
0	2
1	3
2	1
3	1
4	1

$$\overline{\boldsymbol{x}}=\frac{2*0+3*1+1*2+1*3+1*4}{8}=\frac{12}{8}=1,5$$

Il existe d'autres moyennes moins utilisées car elles ne disposent pas des propriétés algébriques valables pour la moyenne arithmétique :

La moyenne **géomètrique** : $g = \sqrt[n]{\prod f_i x_i}$

La moyenne **harmonique** : $\frac{1}{h} = \frac{1}{n}\sum \frac{f_i}{x_i}$

La moyenne **quadratique** : $q = \sqrt[n]{f_i x_i^2}$

On a la relation suivante entre ces moyennes : $h < g < \bar{x} < q$.

b. La moyenne quadratique

Elle est notée par m2 et elle est définie de la manière suivante :

$$m_2=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{2}}}=\sqrt{\sum_{i=1}^{K} f_i \boldsymbol{x}_{\boldsymbol{i}.}^{\boldsymbol{2}}}$$

Ainsi, la moyenne quadratique de la variable « Nombre d'Enfants par Ménage », dont il est question dans l'exemple ci-dessus, vaut :

$$m_2=\sqrt{0,25 * 0^2 + 0,375 * 1^2 + 0,125 * 2^2 + 0,125 * 3^2 + 0,125 * 4^2}=\sqrt{4}=2$$

c. Moyenne harmonique

Elle est notée par m-1 et elle est définie de la manière suivante :

$$m_{-1}=\frac{N}{\sum_{i=1}^{N}\frac{1}{x_i}}=\frac{N}{\sum_{i=1}^{K}\frac{f_i}{x_i}}$$

La moyenne harmonique peut être utilisée chaque fois qu'il est possible d'attribuer un sens réel aux inverses des données (taux d'équipement, pouvoir d'achat, calcul d'indice, etc.).

Exemple : On achète des Dollars une première fois pour 100 Euros au taux de 0,87 Euro le Dollar, puis on en achète une seconde fois pour 100 Euros également mais au taux de 0,71 Euro le Dollars ; ainsi le montant total des Dollars achetés lors de ces deux opérations est :

$$\frac{100}{0,87}+\frac{100}{0,71}=255,79 \text{ dollars}$$

Le taux moyen de change du Dollar pour l'ensemble de ces opérations est, par définition, le cours de cm Euro le Dollar, qui aurait permis l'achat, en une seule fois, de 255,79 Dollars pour 200 Euros ;

$$\frac{200}{\text{cm}}=\frac{100}{0,87}+\frac{100}{0,71}=255,79,$$

Donc,

$$\text{Cm}=\frac{200}{\frac{100}{0,87}+\frac{100}{0,71}}=0,78$$

Il apparait donc que cm est la moyenne harmonique des deux cours correspondant aux deux opérations ; aussi, il est important de noter que cm est différent (strictement plus petit) de la moyenne arithmétique de ces deux cours, en effet cette dernière moyenne vaut (0,87+0,71)/2 =0,79.

d. Moyenne géométrique

On ne peut définir cette moyenne que lorsque les observations x1,….., xn sont tous des nombres réels positifs. Si tel est le cas, la moyenne géométrique de ces observations est notée par Mg, et elle est définie par :

$$\text{Mg} =\sqrt[N]{x_1 x_2 \ldots\ldots\ldots .. x_N}$$

La moyenne géométrique s'utilise, par exemple, quand on veut calculer la moyenne de taux d'intérêt.

Exemple 1 : Supposons que les taux d'intérêt pour 4 années consécutives soient respectivement de 5, 10, 15, et 10%. Que va-t-on obtenir après 4 ans si je place 100 ouguiyas ?
– Après 1 an on a, 100 × 1.05 = 105 um.
– Après 2 ans on a, 100 × 1.05 × 1.1 = 115.5 um.
– Après 3 ans on a, 100 × 1.05 × 1.1 × 1.15 = 132.825 um.

– Après 4 ans on a, 100 × 1.05 × 1.1 × 1.15 × 1.1 = 146.1075 um.

Si on calcule la moyenne arithmétique des taux on obtient

$$\overline{x}=\frac{1,05+1.10+1,15+1,10}{4}=1,1$$

Si on calcule la moyenne géométrique des taux, on obtient

$$\text{Mg} =\sqrt[4]{1.05 \times 1.10 \times 1.15 \times 1.10}= 1.099431377$$

Exemple 2: Supposons que pendant une décennie, les salaires aient été multipliés par 2 et que pendant la décennie suivante ils aient été multipliés par 4 ; alors pour la période de l'ensemble de ces deux décennies le coefficient multiplicateur est 2 * 4 = 8. Le coefficient multiplicateur moyen par décennie pour cette période de vingt ans est, par définition, le coefficient ω qui ne change pas d'une décennie à l'autre, et qui permet une multiplication par 8 des salaires entre le début et la fin de la période. 0n a donc ω^2 = 8 = 2 * 4, d'où ω =$\sqrt{2 * 4}$= 2,83.

Ainsi, il apparait que ω est la moyenne géométrique des deux coefficients multiplicateurs correspondant aux deux décennies ; aussi, il est important de noter que ω est différent (strictement plus petit) de la moyenne arithmétique de ces deux coefficients, en effet cette dernière moyenne vaut (2+4)/2 = 3.

2.3. La médiane

Définition : c'est la valeur *observée ou possible* de la série ordonnée en ordre croissant ou décroissant, qui partage cette série en deux sous-séries, chacune comprenant le même nombre d'observations (elle est dénommée par la signe M_e.

- Si n impair $M_e = x_{(n+1)/2}$
- Si n pair $M_e = \frac{x_{(n/2)}+x_{(n+1)/2}}{2}$

Remarque1 :
Si la variable est discrète et n pair, il se peut qu'il n'y ait pas de valeur médiane car ***Me*** doit correspondre à une valeur possible de la série.

Remarque 2:
La médiane correspond à la valeur telle que la fréquence cumulée est égale à ½.

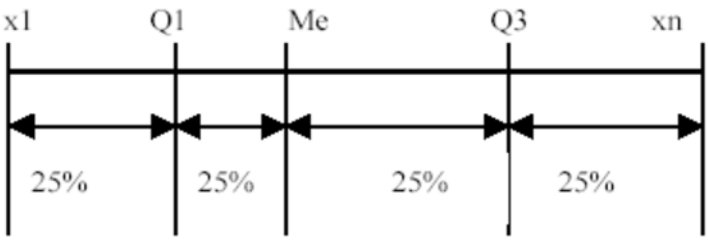
Exemple : si on considère un nombre d'enfants par femme : 1,2,3,3,3,5,5,5,5,6, ***Me***=4.
Si on considère une série de la superficie : 8,8.5,10,11,12.5,13,15,20,25,33, ***Me***=12,75.

Limites : La médiane est plus robuste que la moyenne (pas influencée par les valeurs extrêmes) mais elle est influencée par le nombre d'observations.

Chapitre 3 : Indicateurs de dispersion et de position

3.1. Les quantiles

Définition: ils correspondent à des valeurs de la variable statistique qui partagent la série ordonnée en l parties égales. Si l=4, les quantiles sont appelés quartiles. Il y a 3 quartiles, appelés Q1, Q2=Me et Q3. Si l=5, les quantiles sont appelés quintiles. Il y a 5 quintiles. Si l=10, les quantiles sont appelés déciles. Il y a 10 quintiles.



$\boxed{I=Q3-Q1}$ est appelé *l'intervalle interquartile* et comporte 50\% de la série. C'est un indicateur de dispersion de la série.

Les quantiles résument la série à partir de ses valeurs extrêmes, ses quartiles et sa médiane.

Permet une comparaison visuelle immédiate de plusieurs séries.

La médiane est le quantile d'ordre p = 1⁄2.

– On utilise souvent :

*x*1=4   le premier quartile,
*x*3=4   le troisième quartile,
*x*1=10 le premier décile,
*x*1=5   le premier quintile,

$x4=5$ le quatrième quintile,
$x9=10$ le neuvième décile,
$x0{:}05$ le cinquième percentile,
$x0{:}95$ le nonante-cinquième percentile.

– Si $F(x)$ est la fonction de répartition, alors $F(xp) \geq p$.

Exemple 1 : Soit la série statistique 12, 13, 15, 16, 18, 19, 22, 24, 25, 27, 28, 34 contenant 12 observations (n = 12).

– Le premier quartile : Comme $np = 0.25 \times 12 = 3$ est un nombre entier, on a

$x1=4 = (x(3) + x(4))/2=(15 + 16)/2= 15.5.$

– La médiane : Comme $np = 0.5 \times 12 = 6$ est un nombre entier, on a
$x1=2 =1/2*\{x(6) + x(7)\}= (19 + 22)/2 = 20.5.$

– Le troisième quartile : Comme $np = 0.75 \times 12 = 9$ est un nombre entier, on a
$x3=4 =(x(9) + x(10))/2=(25 + 27)/2= 26.$

3.2. L'étendue

L'*étendue* est simplement la différence entre la plus grande et la plus petite valeur observée.

$$\mathbf{E = x(n) - x(1).}$$

3.3. La variance et l'écart-type de la série

Définition : *La variance est la somme pondérée des carrés des écarts des valeurs de la série à la moyenne.*

L'écart type est la racine carrée de la variance.

La variance (ou écart-type) est toujours positive ou nulle.

Variance de la série : $\boldsymbol{S_x^2 = S^2(x) = 1/n\sum_{i=1}^{n}(x_i - \overline{x})^2}$

Lorsque la série est un échantillon issu d'une population et que l'on s'intéresse aux caractéristiques de cette population via l'échantillon (inférence), on utilise plutôt $\boldsymbol{S_x^{*2}}$ qui est un meilleur estimateur de la variance théorique de la population. Dès lors que la taille n de la série est assez grande, ces deux quantités sont pratiquement égales.

Variance d'échantillonnage : $\boldsymbol{S_x^{*2} = S^{*2}(x) = 1/n\sum_{i=1}^{n}(x_i - \overline{x})^2}$

L'écart type est la racine carrée de la variance : S_x=S(x)=$\sqrt{\boldsymbol{S_x^2}}$=$\sqrt{\boldsymbol{S_x^{*2}}}$

Quand on veut estimer l'écart-type d'une variable X partir d'un échantillon de taille n, utilise la variance "corrigée" pour définir l'écart type.

$$\boldsymbol{S_x}=\sqrt{\boldsymbol{S_x^2}}= \boldsymbol{S_x}\sqrt{\frac{\boldsymbol{n}}{\boldsymbol{n-1}}}$$

<u>Exemple :</u>

i	x_i	$\overline{x}$	$x_i-\overline{x}$	$(x_i-\overline{x})^2$
1	15	27,13	-12,13	147,1369
2	15	27,13	-12,13	147,1369
3	16	27,13	-11,13	123,8769
4	17	27,13	-10,13	102,6169
5	18	27,13	-9,13	83,3569
6	20	27,13	-7,13	50,8369
7	21	27,13	-6,13	37,5769
8	22	27,13	-5,13	26,3169
9	23	27,13	-4,13	17,0569
10	23	27,13	-4,13	17,0569
11	23	27,13	-4,13	17,0569
12	24	27,13	-3,13	9,7969
13	25	27,13	-2,13	4,5369
14	25	27,13	-2,13	4,5369
15	26	27,13	-1,13	1,2769
16	26	27,13	-1,13	1,2769
17	27	27,13	-0,13	0,0169
18	28	27,13	0,87	0,7569
19	28	27,13	0,87	0,7569
20	29	27,13	1,87	3,4969
21	30	27,13	2,87	8,2369
22	30	27,13	2,87	8,2369
23	32	27,13	4,87	23,7169
24	34	27,13	6,87	47,1969
25	35	27,13	7,87	61,9369
26	36	27,13	8,87	78,6769
27	39	27,13	11,87	140,8969

28	40	27,13	12,87	165,6369
29	43	27,13	15,87	251,8569
30	44	27,13	16,87	284,5969
n=30			$\sum_{i=1}^{n}(x_i - \overline{x})^2$	**1867,467**
somme=	**814**			
Moyenne	$\overline{x} =$	**27,1333333**		
Variance	S_x^2**=**	**62,2489**		
Ecart-type	S_x**=**	**7,889797209**		

3. 4 Le Coefficient de Variation (CV):

Est une mesure de la dispersion relative exprimé en pourcentage. Le CV permet d'apprécier la représentativité de la moyenne par rapport à l'ensemble des observations. Il donne une bonne idée du degré d'homogénéité d'une série. Il faut qu'il soit le plus faible possible (<15% en pratique).

Formule de calcul : **CV=**$\frac{S_x}{\overline{x}}$***100**

Dans l'exemple précédent : CV=$\frac{7{,}889797209}{27{,}1333333}$***100=29,1%**

Chapitre 4 : Analyse bivariée

L'objectif de l'analyse bivariée est d'étudier les éventuelles relations entre deux variables statistiques.

Il s'agit d'étudier de manière conjointe deux variables quantitatives, que l'on note X et Y. Pour chaque individu de la population étudiée, on mesure simultanément les valeurs prises par X et Y. Ainsi, on obtient une série composée de N couples :

((x1,y1),(x2,y2),…………....,(xN,yN)).

4.1. Covariance et corrélation linéaire

Pour mesurer la *dépendance linéaire* entre les deux variables, on calcule la *covariance*, qui découle de la notion de variance (cf. 1 - Statistique descriptive univariée). La covariance quantifie la dispersion conjointe des deux variables autour de leur moyenne respective.
La formule correspond à la moyenne du produit des écarts à la moyenne (où x¯ est la moyenne des valeurs prises par X, idem pour y¯ par rapport à Y) :

$$\text{cov(X,Y)}= \frac{\sum_{i=1}^{N}(x_i-\overline{x})(y_i-\overline{y})}{N}$$

La covariance est une valeur réelle, qui peut être positive comme négative. Par ailleurs, elle s'exprime relativement aux échelles utilisées par les variables, ce qui rend son interprétation difficile. On lui préfère donc une version normalisée appelée coefficient de corrélation linéaire, noté r. On l'obtient en divisant la covariance par le produit de l'écart-type pour X et de l'écart-type pour Y :

$$\text{r(X,Y)}= \frac{\text{cov(X,Y)}}{\sigma(\text{X})\sigma(\text{Y})}$$

Ce coefficient et indépendant de l'échelle utilisée pour les mesures et prend sa valeur dans l'intervalle [−1;1]. Cette mesure est symétrique, c'est-à-dire que r(X,Y)=r(Y,X).

Par ailleurs, la corrélation entre une série statistique et elle-même vaut nécessairement 1, c'est-à-dire que r(X,X)=1et r(Y,Y)=1. Si les variables X et Y sont *non corrélées*, alors r(X,Y)=0. Au contraire, si r(X,Y) est égal à 1 ou -1, alors il y a une *liaison linéaire exacte* entre les deux variables. Un coefficient de corrélation r(X,Y)≠0 suggère une *liaison relative*.

Un coefficient positive indique que les deux variables varient, de manière exacte ou relative, dans le même sens ; un coefficient de corrélation négatif suggère au contraire que les deux variables varient dans des sens opposés.

4.2. Régression linéaire

Il s'agit ici d'expliquer une des deux variables en fonction de l'autre. On choisit donc une variable comme *variable expliquée* et l'autre comme *variable explicative*. Si l'on choisit Y comme variable expliquée, on l'exprime comme une fonction linéaire de X :

Y=ax+b

Afin de déterminer les valeurs de a et b, on résoudre un problème d'optimisation basé sur le critère des moindres carrés :

$$min_{ab}=\sum_{i=1}^{N}(y_i - ax_i + b)^2$$

Autrement dit, on cherche à déterminer l'équation de la droite qui minimise la somme du carré des écarts pour tous les individus.

Pour mesurer la qualité de l'ajustement de la droite de régression aux données, on mesure le coefficient de détermination, noté R2 :

$$R^2 = \frac{\sum_{i=1}^{N}(\widehat{y_i} - \bar{y})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

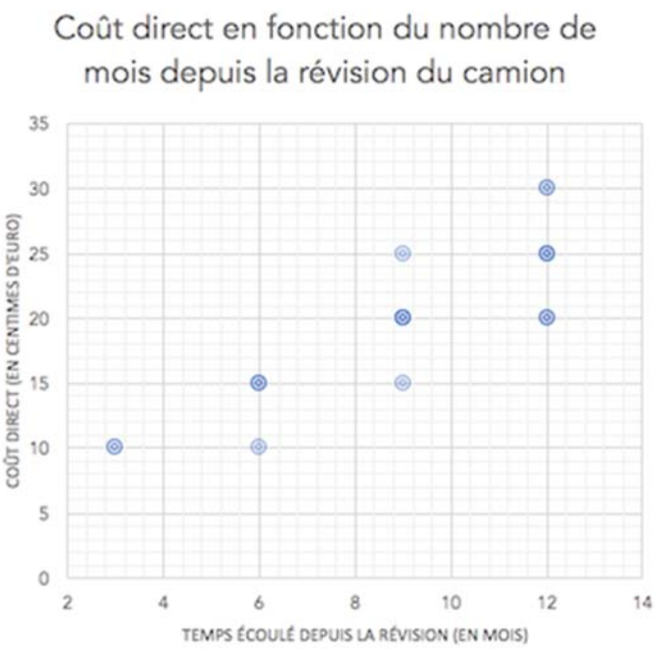
Ce coefficient varie entre 0 et 1 (la part de variance expliquée par la régression) ; plus il s'approche de 1, meilleur est l'ajustement de la droite aux données.

Dans le cas de la régression linéaire, le coefficient de détermination équivaut au carré du coefficient de corrélation linéaire, c'est-à-dire R^2=r^2.

Représentation graphique d'une série statistique bivariée

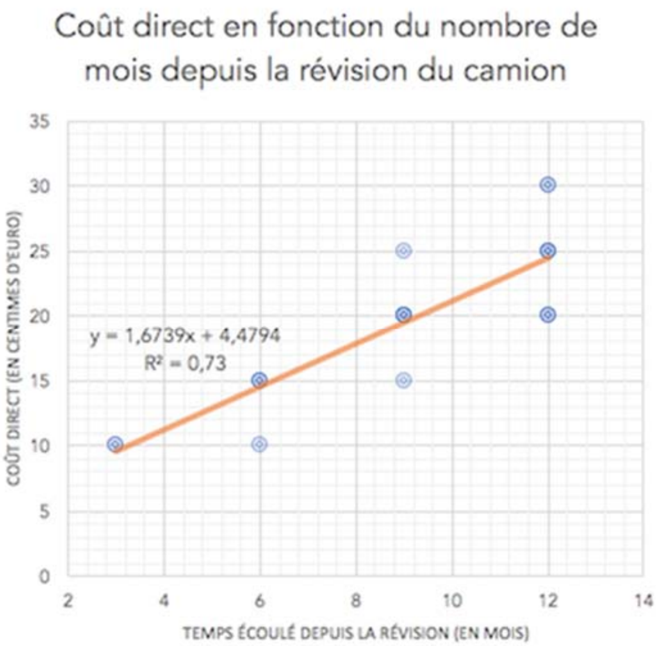
Nuage de points

Le nuage de points permet de décrire la population selon les deux variables étudiées. Les points pouvant se superposer (dans le cas où plusieurs individus sont identiques), on utilise parfois une représentation semi-transparente des points.



Droite de régression linéaire

On peut superposer la droite de régression linéaire, et faire figurer son équation ainsi que le coefficient de détermination.



Exercices :

Exercice 1 : On pèse les 50 élèves d'une classe et nous obtenons les résultats Résumés dans le tableau suivant :

43 43 43 47 48
48 48 48 49 49
49 50 50 51 51
52 53 53 53 54
54 56 56 56 57
59 59 59 62 62
63 63 65 65 67
67 68 70 70 70
72 72 73 77 77
81 83 86 92 93

1. De quel type est la variable poids ?
2. Construisez le tableau statistique en adoptant les classes suivantes : [40 ;45]]45 ;50]]50 ;55]]55 ;60]]60 ;65]]65 ;70]]70 ;80]]80 ;100]
3. Calculez tous les paramètres (de la tendance centrale, de position et de dispersion) sans prendre en compte les classes.

Exercice 2 : On a étudié la taille de la population des 20 enfants de troisième primaire de l'école de Kiffa et on a obtenu les résultats suivants :

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
127	128	136	139	128	133	127	126	131	130	120	129	126	133	138	132	122

A. Quel est le type de cette variable
B. Déterminer le mode, la médiane, la moyenne, l'écart-type, les trois quartiles ainsi que l'écart interquartile de la série
C. Dessiner un histogramme de classes d'amplitude 4 dont la première est [120,124[
D. Dessiner la boîte à moustache correspondante

Exercice 3 : Lors de l'analyse des données issues du RGPH, on s'intéresse au nombre d'enfants de moins de 18 ans par famille. Un échantillon aléatoire de 500 familles révèle la distribution empirique des fréquences suivantes :

Nombre d'enfants	0	1	2	3	4	5	6	7	8
Nombre de familles	91	146	104	63	47	33	10	4	2

1. Construire le diagramme en barre
2. Déterminer la fréquence empirique des fréquences relatives
3. Calculer le nombre moyen, la médiane et le mode du nombre d'enfants par famille
4. Calculer la variance, l'écart-type et l'étendue du nombre d'enfants par famille

Exercice 4 : Voici la répartition de la population par grande groupe d'âge lors du RGPH 2013

Milieu/Sexe	Effectifs			
	0-14 ans	15-59 ans	60 ans et plus	Total
Masculin	792 318	849 811	100 945	1 743 074
Féminin	771 981	924 227	98 086	1 794 294
Ensemble	**1 564 299**	**1 774 038**	**199 031**	**3 537 368**

Calculer les proportions par Grande groupe d'âge et comparer les hommes aux femmes en utilisant deux diagrammes en secteur.

Module2 : Analyse démographique

Chapitre 1 : Définition des concepts et notions de base

1.1 Démographie

De l'ancien grec, (Demos : peuple et graphein : écrire), la démographie est définie par le dictionnaire multilingue des Nations Unies comme une science ayant pour objet l'étude des populations humaines en ce qui concerne leur dimension, leurs structures, leur évolution ainsi que de certains de leurs caractères généraux.

Comme toute science, la démographie se base sur l'observation des faits en utilisant les statistiques. C'est le point de départ de toute recherche démographique. Celle-ci passe en fait par trois niveaux d'analyses :

- Analyse descriptive des faits démographiques observés
- Analyse explicative de type interne où il s'agit d'expliquer certains faits démographiques par d'autres faits démographiques
- Analyse explicative de type externe ou on cherche des explications aux faits démographiques par des faits d'ordre économique, sociologique, etc.

Plus concrètement, la démographie, c'est **l'étude des phénomènes démographiques** sachant qu'un phénomène démographique se manifeste par un **événement démographique**.

Exemples :

<u>Evénement démographique</u>	<u>Phénomène démographique</u>
naissance	Fécondité, Natalité
décès	Mortalité
mariage	Nuptialité
divorce	Divortialité

Dans le cadre de ce cours, on se limite à l'analyse descriptive des faits démographiques observés.

1.2 Population

Une population est un ensemble d'individus coexistant à un moment donné et délimité selon des critères variés d'appartenance. Plus communément, le mot population désigne l'ensemble des habitants d'un territoire (état, province, département, ville, village...) à une date donnée (*population totale*). Il peut aussi désigner des fractions variées de cet ensemble (population urbaine, rurale, population scolaire, population active, etc.).

1.3 Résident présent – Résident absent

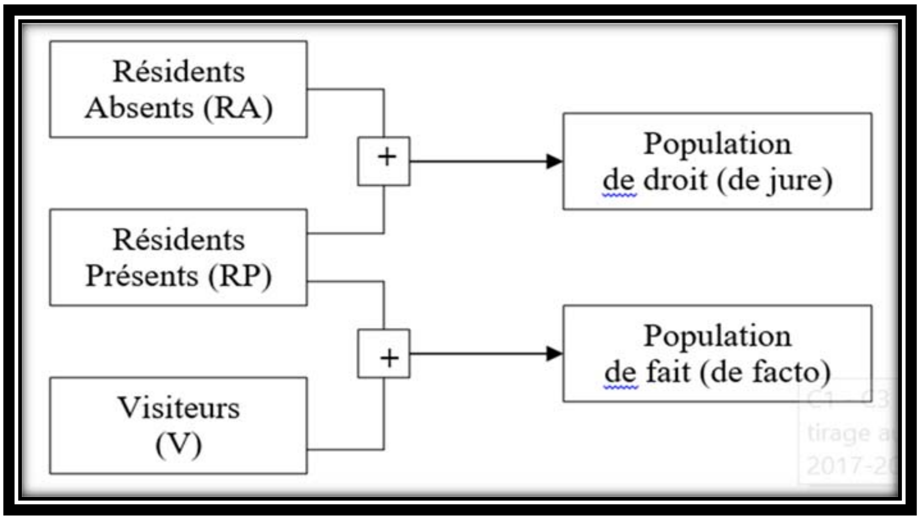
Le concept de résidence est lié à celui de domicile habituel (logement où on habite depuis au moins six mois ou on a l'intention d'habiter pendant plus de six mois). Les résidents d'une localité peuvent donc être classés en deux catégories :

- les résidents présents (R.P.) : ceux qui ont passé la nuit de référence dans le logement ;
- les résidents absents (R.A.) : ceux qui n'ont pas passé la nuit de référence dans le logement et dont l'absence remonte à moins de six mois (et dont on a toutes les raisons de penser qu'ils ont l'intention de revenir).

Les visiteurs ne sont pas des résidents mais des personnes présentes dans le logement la nuit de référence.

1.4 Population de droit – Population de fait

La population de droit (population résidente) est constituée des résidents présents et des résidents absents tandis que la population de fait (population présente) est constituée de résidents présents et des visiteurs. Les analyses sur l'état de la population portent généralement sur la population de droit constitue la population d'un pays.



1.5 Evènements démographiques

a. **Fécondité** : La fécondité est l'un des phénomènes qui intéresse le démographe.

La fécondité comme la fertilité désignent la capacité des êtres vivants à produire une descendance.
On parle de la natalité lorsqu'on s'intéresse à la fréquence des naissances au sein de la population à l'exclusion des sous-populations. De même, on parle de la fécondité lorsqu'on s'intéresse à la fréquence des naissances, non pas au sein des populations, mais plutôt parmi les femmes en âge de procréer (femmes âgées de 15-49 ans).

b. **Nuptialité**

Par nuptialité, on désigne le phénomène qui a trait à la constitution ou à la dissolution d'unions ou de mariages de personnes de sexes opposés au sein d'une population et dans un but de procréation.

c. **Mortalité**

La mortalité est, selon le dictionnaire multilingue de Louis Henry, l'action de la mort sur les populations. De façon générale, l'étude de la mortalité consiste à mesurer l'impact de la mort sur une population donnée.

d. Migration

Selon le dictionnaire démographique, on appelle migration ou mouvement migratoire un ensemble de déplacements ayant pour effet de transférer la résidence des intéressés d'un certain lieu d'origine ou lieu de départ, à un certain lieu de destination, ou lieu d'arrivée.

D'après cette définition, la migration se caractérise essentiellement par le fait qu'elle entraine un changement de lieu de résidence habituel (ou domicile) et de milieu. Cela suppose que le changement de domicile dans la ville ou dans le même quartier dans la même ville serait une migration.

Chapitre 2 : sources de données et type d'observations

2.1 Les sources de données démographiques

Les principales sources en démographie sont :

a. Recensement : C'est l'opération qui consiste à recueillir des données démographiques et sociales se rapportant, à un moment donné, à tous les habitants d'u territoire déterminé, le plus souvent un pays. Il s'agit donc d'une photographie instantanée de la population.
b. Etat civil : Historiquement, l'état civil est la plus importante des sources de renseignements sur les évènements démographiques. Il a pour rôle de consigner les évènements du mouvement démographique déclarés par les habitants (naissances, mariages, divorces, décès …). Ainsi, l'enregistrement continu des faits d'état civil permet au démographe de se faire une idée sur la dynamique (mouvement) de la population.
c. Enquêtes démographiques : A côté du recensement et de l'état civil qui constituent les deux principaux instruments pour une observation exhaustive (toute la population est concernée), les enquêtes démographiques par sondage (à passage unique ou à passage répétés) constituent une source appréciable des données démographiques. Le mauvais fonctionnement de

l’état civil et les difficultés liées à l’organisation d’un recensement (opération lourde) ont conduit les démographes à développer d’autres méthodes d’observations, en l’occurrence les enquêtes par sondage.

d. Autre sources : La démographie peut également largement profiter du travail d’un nombre d’instruments qui ont réuni des statistiques de population pour leurs besoins propres. Exemples : Registres de population, registres ou fichiers de migration, les documents établis par les compagnies d’assurances, les services sanitaires, les services scolaires, la sécurité sociale, la justice etc.

2.2 Différents type d’observations en démographie

Les différents types d’observations en démographie sont :

a. L’observation continue :

C’est quand les évènements sont enregistrés au fur et à mesure qu’ils se produisent. Par exemple : le registre de l’état civil.

b. L’observation rétrospective :

C’est quand les évènements ne sont enregistrés que relativement très longtemps après leur occurrence, en interrogeant les personnes qui les ont subis : Exemple le recensement.

c. L’observation suivie :

Cette expression indique un mode d’observation dans lequel les événements concernant un même individu sont rapprochés les uns des autres de manière à ce que l’on puisse tracer tout ou une partie de la biographie de l’individu. On peut y parvenir : a) en suivant l’individu à partir d’une certaine date et en relevant tous les événements qui les concernent ; b) en l’interrogeant à une certaine date sur son passé ; en combinant les deux.

d. L’observation instantanée :

Le mode d’observation concernant les « état » de population par opposition aux « flux » qui sont appréhendés dans les trois modes d’observations.

2.3 Evolution d'une population

Accroissement d'une population

On considère l'effectif d'une population à deux dates t_0 (P_0) et t_1 (P_1). L'accroissement de cette population (A) entre les temps t_0 et t_1 (période t) est donné par la formule :

$$A_t = P_1 - P_0$$

Exemple : L'accroissement de la population mauritanienne entre 2000 et 2013 est :

$A_{2013}=P_{2000}-P_{2013}$=3 537 368-2 508159=1 029 209.

On peut aussi calculer l'accroissement en utilisant les données sur le mouvement de la population.

On désigne par :N_t : les naissances au cours de la période t ;D_t : les décès au cours de la période t ;

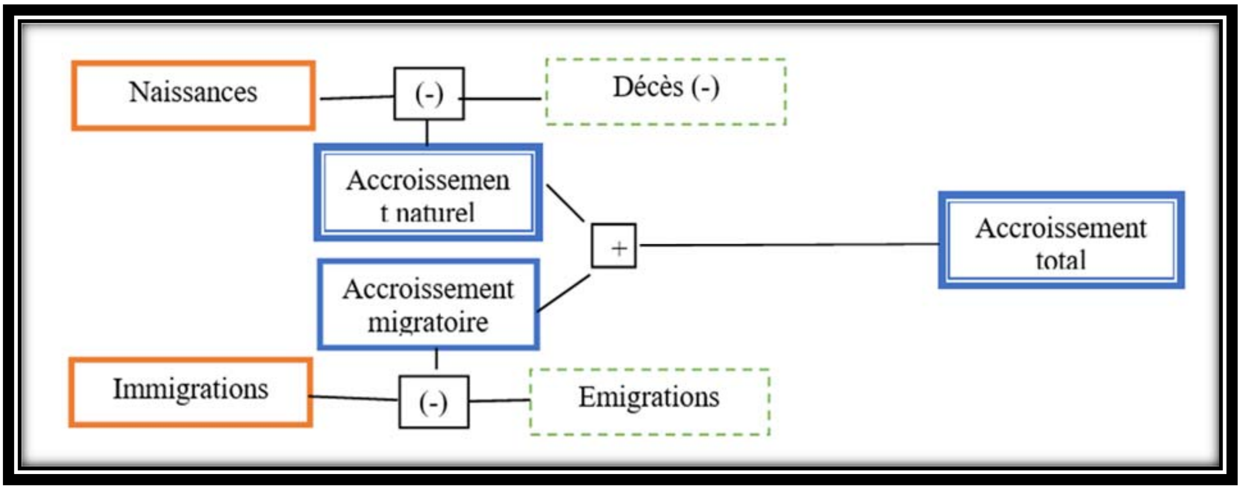
I_t : les immigrations au cours de la période t ; et E_t : les émigrations au cours de la période t ;

On obtient alors, l'accroissement total de la population : $A_t = (N_t - D_t) + (I_t - E_t)$

- $N_t - D_t$ = Accroissement naturel de la population (Naissances-décès) ;
- $I_t - E_t$ = Accroissement migratoire de la population. Cette différence est appelée aussi solde migratoire

On peut schématiser les relations ci-dessus par le diagramme ci-après.

Composantes principales de l'accroissement total d'une population



Taux brut d'accroissement d'une population

Le taux ayant une dimension annuelle, on prend t = une année et on désigne par P_m la population moyenne au cours de l'année t, P_0 et P_1 les effectifs de la population en début et en fin d'année.

$$P_m = \frac{P_1 + P_0}{2}$$

Le taux d'accroissement r est égal à :

$$r = \frac{A_t}{P_m} = \frac{N_t - D_t + I_t - E_t}{P_m}$$

Evolution d'une population à taux d'accroissement naturel constant

On considère maintenant que la population P est à taux d'accroissement constant r, P_t et P_{t-1} les effectifs de la population au cours de l'année t et t-1, A_t l'accroissement de cette population entre t_0 et t_1.

$A_t = r.\ P_{t-1}$

$P_1 = P_0 + A_1 = P_0 + r.\ P_0 = P_0\ (1 + r)$

$P_2 = P_1 + A_2 = P_1 + r.\ P_1 = P_1\ (1 + r) = P_0\ (1 + r)^2$

.

.

.

$P_n = P_0\ (1 + r)^n$ (3)

Une population dont le taux d'accroissement naturel est constant évolue en progression géométrique. Si r est positif, la population croît ; si r est négatif la population décroît.

Evolution exponentielle :

L'évolution exponentielle est donnée par la formule suivante

$$P(n) = P(0)e^{rn}$$

Temps de doublement d'une population

Le temps de doublement d'une population est la durée requise, à partir d'une date donnée, pour que son effectif soit multiplié par 2.

$P_t = 2.\ P_0 = P_0\ (1 + r)^t$ et $t = \dfrac{Log\ 2}{Log\ (1 + r)}$

Exemple : une population qui augmente constamment au taux de 2,5 % par an doublerait en 28,1 ans.

Temps de doublement de la population pour quelques taux d'accroissement naturel annuel.

Taux d'accroissement naturel (%)	Temps de doublement (ans)

0,1	693,5
0,5	139,0
1,0	69,7
1,5	46,6
2,0	35,0
2,5	28,1
2,7	26,0
3,0	23,4
3,5	20,1
4,0	17,7

Chapitre 3 : Indicateurs de la dynamique démographiques et méthodes de calcul

3.1Fécondité

Les principaux indicateurs de la fécondité sont :

Taux de natalité

Le taux de natalité d'une année t(TN_t) s'obtient en rapportant les naissances d'une année t à la population moyenne de cette même année.

$$TN_t = \frac{N_t}{\frac{P_t + P_{t+1}}{2}} \ (‰)$$

Le taux de natalité exprime un potentiel de croissance d'une population, mais il ne suffit pas pour décrire réellement la propension des femmes à procréer, car il dépend de la structure par âge de la population. De nombreuses personnes figurant au dénominateur ne sont pas concernées par les naissances qui sont au numérateur.

Taux global de fécondité

On calcule le taux global de fécondité de l'année t en rapportant les naissances en t au nombre de femmes en âge de procréer de cette même année.

$$TGF_t = \frac{N_t}{F_{15-49,t}}$$

RQ : Le taux global de fécondité s'appelle aussi taux de fécondité aux âges. Il se rapporte à l'ensemble de la période de procréation.

- Au dénominateur figurent les femmes exposées au risque de la naissance aux âges révolus de 16 ans, 17 ans … 49 ans.
- Les taux de fécondité sont généralement exprimés en naissances pour 1000.

Taux de fécondité générale par âge

On calcule le taux de fécondité générale par âge (f_x) en rapportant les naissances issues des mères d'âge x à l'effectif des femmes d'âge x.

$$f_x = \frac{N_x}{F_x} \quad x = 15, 16, \dots, 49 \; ans$$

Somme des naissances réduites

Dans l'analyse transversale de la fécondité, l'indicateur de l'intensité est la somme des naissances réduites, encore appelé indice synthétique fécondité, encore appelé l'indice conjoncturel de fécondité.

C'est le nombre moyen d'enfants par femme et on le calcule en faisant la somme des taux de fécondité par âge d'une année donnée.

$$\boldsymbol{SNR = a \sum f_{x,t}}$$

Où $f_{x,t}$ est le taux de fécondité générale à l'âge x de l'année t.

Exemple de calcul de l'ISF :

Age de la mère	Nbre de femmes	Nombre de naissances	Taux de fécondité ((2)÷(1))	Taux de fécondité par âge ((3)x5)
15-19	2924	152	0,051988479	0,259942396
20-24	2586	570	0,220451612	1,102258062
25-29	2406	581	0,241492365	1,207461826
30-34	2118	521	0,245993831	1,229969157
35-39	1727	423	0,244970572	1,224852859
40-44	1484	100	0,067389876	0,336949382
45-49	198	51	0,257288548	1,286442742
			ISF	6,647876425

3.2 Mortalité

Taux de mortalité :

Le taux de mortalité appelé également taux brut de mortalité est le nombre de décès pour 1 000 habitants durant une année donnée

$$\boldsymbol{\frac{Nombre\ de\ décès\ 2017}{Population\ totale\ en\ 2017} * 1000}$$

Taux de mortalité maternelle

Définition :
« C’est le décès d’une femme survenue au cours de sa grossesse ou dans un délai de 42 jours après sa terminaison quelle que soit la durée ou la localisation … »

Risque associé aux accouchements

$$\text{Rapport ou ratio de mortalité maternelle} = \frac{\text{Nombre de décès maternels}}{\text{Nombre de naissances vivantes}} * 100\ 000$$

$$\text{Taux de mortalité maternelle} = \frac{\text{Nombre de décès maternels}}{\text{Nombre moyen de femmes } 15-49 \text{ ans}} * 100\ 000$$

Le rapport mesure la probabilité pour une femme de décéder dès qu’elle est enceinte.

Le taux donne la probabilité

- de concevoir
- et d’en décéder

Table de mortalité :

Elle constitue l’un des outils les plus performants du démographe, sert à simuler le taux de mortalité d’une population tout au long de son existence. On élabore à partir des taux de mortalité par âge de cette population, que l’on applique à une population hypothétique de 100.000 personnes, toutes nées sur la même période. Sur cette table, d’années en années, les plus ancies disparaissent.

Exemple d’une table de mortalité :

Age x	S_x	d(x,x+1)	$_1q_x$	$_1p_x$
0	S0 = 100 000	d(0,1) = 5 000	$_1q_0$ = 50 °/∞	$_1p_0$ = 950 °/∞
1	S1 = 95 000	d(1,2)	$_1p_1$	$_1q_1$
2	S2	d(2,3)	$_1q_2$	$_1p_2$
. . .	. . .	. . .	. . .	. . .
x	S_x	d(x,x+1)	$_1q_x$	$_1p_x$
. . .	. . .	. . .	. . .	. . .
99	S_{99}	d(99,100)	$_1q_{99}$	$_1p_{99}$
	$S_{100} = 0$			

Survivants : $S_{x+1} = S_x - d(x, x+1)$: le nombre de survivants à l'âge x+1

- S0 = « nombre de survivants à l'âge 0 » ou « nombre de survivants à la naissance »ou encore « nombrede naissances vivantes ».

- S1 = « nombre de survivants à l'âge de 1 an » ou « nombre d'individus ayant fêté leur premieranniversaire »

Décès de la table : $d(x, x+1) = S_x - S_{x+1}$.Décès entre l'âge x et l'âge x+1

Plus généralement, **d(x, x+a)** = décès entre un âge x et un âge x+a.

Quotient de mortalité $qx = \frac{d(x,x+1)}{Sx}$

$_aq_x$ = Probabilité de décéder entre l'âge x et l'âge x+a

Exemple : **Quotient de mortalité infanto-juvénile ($_5q_0$) qui désigne le r**isque ou probabilité pour un enfant né vivant, de décéder avant d'atteindre son cinquième anniversaire.

L'espérance de vie à la naissance (e0)

C'est le nombre moyen d'années que peut espérer vivre un enfant né au cours de l'année d'observation, si les taux spécifiques de mortalité tels qu'observés au cours de l'année d'observation s'appliquaient à lui à tous les âges autrement dit si la population concernée est stationnaire. En pratique, il se déduit de la table de mortalité.

3.3 Nuptialité

Les principaux indicateurs de la nuptialité sont :

Le taux brut de nuptialité : il mesure le nombre annuel de mariages pour 1000 personnes. Il s'obtient par la formule :

$$Taux = \frac{Nombre\ annuel\ de\ mariages}{population\ moyenne\ de\ l'année} * 1000$$

Le taux brut de nuptialité : Le taux brut de nuptialité comme tous les taux bruts, n'est pas un indice car il dépend de plusieurs facteurs tels que : la structure par âge de la population, la structure par sexe de la population, etc.

Il est influencé par les aléas de la conjoncture : les conditions qui, au cours d'une année ou d'une période, peuvent forcer les individus à différer ou à récupérer les mariages.

Le taux de nuptialité : on fait appel rarement à cet indice dans l'analyse de la nuptialité. On l'utilise seulement lorsque les données disponibles ne permettent pas de le faire mieux.

- **Fréquence du célibat définitif**

En général, le dernier terme de la table de nuptialité (C_{35} ou C_{50})n'est pas nul. Celui-ci indique le nombre de personnes qui sont restées définitivement célibataires. Le rapport de ce nombre à celui de la racine de la table est appelé fréquence du célibat définitif et notée :

F = C_{50}/C_{14} (ou C_{35}/C_{14})

- **Intensité de la nuptialité**

Le complément à 1 du célibat définitif mesure l'intensité de la nuptialité des célibataires ou encore le nombre moyen de mariages par personne.

$I = 1 - C_{50} / C_{14}$ ou $I = 1 - C_{35} / C_{14}$

- **Calendrier de la nuptialité**

Le calendrier de la nuptialité est la distribution des mariages M(x, x+a) selon l'âge des individus.

Pour comparer les calendriers de deux tables différentes il est nécessaire de comparer les fréquences de mariages entre anniversaires et non directement les séries des m(x, x+1)

- **L'âge moyen au premier mariage**

On calcule l'âge moyen au premier mariage (m), en faisant la moyenne arithmétique avec comme pondération les âges aux mariages. Pour cela, on fait l'hypothèse que les mariages se répartissent de manière uniforme entre deux anniversaires successifs.

Dans le cas où les mariages se produisent entre 14 et 50 ans, l'âge moyen, m, est égal à :

$$\boldsymbol{m} = \frac{1}{C_{14} - C_{50}} \left(\frac{14{,}5 * m(14{,}15) + 15{,}5 * m(15{,}16) + 16{,}5 * m(16{,}17) + \ldots + 49{,}5 * m(49{,}50)}{m(14{,}15) + m(15{,}16) + m(16{,}17) + \ldots + m(49{,}50)} \right)$$

3.4 Migration

Définitions :

La migration est définie comme un « ensemble de déplacements ayant pour effet de transférer la résidence des intéressés d'un certain lieu d'origine, ou lieu de départ, à un certain lieu de destination, ou lieu d'arrivée pour une certaine durée. (L.Henry, 1981 p.105). Ainsi, on peut distinguer les concepts suivants :

- **Migrant** : est migrant toute personne, quel que soit son âge, qui a changé de lieu de résidence au cours de la période de référence ;
- **Non-migrant** : est non migrant toute personne qui, au cours de la période de référence, réside dans son lieu de naissance ;
- **Emigrant** : c'est le migrant sortant d'une zone (wilaya ou pays);
- **Immigrant** : c'est le migrant entrant dans une zone (wilaya ou pays).

Trois éléments sortent dans la définition de la migration :

- Lieu d'origine, lieu de départ;
- le changement de résidence pour une durée
- La résidence : Un individu peut posséder plusieurs résidences : Quelle est la résidence d'un élève qui loge chez un tuteur ? Quelle est la résidence d'un travailleur saisonnier ?

La migration est un phénomène renouvelable. Ne pas confondre migrant (l'individu) et la migration (l'évènement).

Les indicateurs de la migration dans les recensements

Dans les recensements on pose toute une série de questions qui permettent de construire des indicateurs de migration

- -Le lieu de résidence au moment du recensement. C'est une information importante. En général il n'y a pas d'erreur de collecte sur cette variable ou du moins très peu.
- -Le lieu de naissance.
- Au moment de la collecte, il faut préciser le niveau le plus bas où cette information doit être collectée. On suit en général la division administrative du pays. En Mauritanie, le niveau le plus fin au RGPH2013 est la commune.
- Le lieu de résidence à une date ultérieure. Cette question peut être posée sous différentes formes: Lieu de résidence il y a un an, il y cinq ans, lieu de

résidence au dernier recensement, ou depuis quand habiter-vous ici, etc.Cette information permet d’identifier les migrants pour une période donnée.

- En fin une question est posée sur les émigrés du ménage au cours des cinq dernières années

A partir de ces différents indicateurs on construit des variables. Ces variables étant construites à partir de date et de lieu on se trouve souvent confronté à des problèmes de qualité de données.

Les principaux indicateurs de migrations calculés à partir des données de recensement sont :

- Migrant interne
- Migrant externe ou international
- Migrant durée de vie
- Migrant récent
- Migrant de retour

Bibliographie :

- Abdennasser Chekroun, « Statistiques descriptives et exercices », Université de Tlemcen, 2017 ;

- Antoine Ayache & Julien Hamonier « Cours de Statistique Descriptive », pub. Sur l'internet sans date de publication ;

- Yves Tillé, « Résumé du Cours de Statistique Descriptive », 2010 ;

- Dodge Y.(2003), *Premiers pas en statistique*, Springer ;

- Droesbeke J.-J. (1997), *´ Eléments de statistique*, Editions de l'Université libre de Bruxelles/Ellipses ;

- Jacob S Siegel and David A Swanson « Methods and matrials en Demography », second édition ;

- Population Reference Bureau « Guide de la Démographie » quatrième édition.

- Population Reference Bureau « Guide de la Population » sixième édition, 2011.