

# STATISTIQUES ET CARTOGRAPHIE : QUELLE(S) METHODE(S) PRENDRE EN COMPTE ?

Jean-Paul BORD - Guy LANDURE\*  
Université de Tours - Atelier de Cartographie

A partir d'un exemple de carte à un caractère en zonal ordonné (1) portant sur l'accroissement naturel en 1982, dans les pays du Maghreb-Machrek (2), l'article se propose de voir les analogies et les différences visuelles résultant de 4 méthodes de traitement des données.

Avant d'arriver à la carte, les données (tableau ci-dessous) seront réparties ou ordonnées selon 4 méthodes :

1. d'après une courbe de répartition des valeurs,
2. idem, mais en tenant compte de la moyenne arithmétique,
3. d'après la médiane et les quartiles,
4. enfin, à partir de la variance et de l'écart-type, variables centrées réduites.

Pourquoi une telle étude ?

Ici, la carte n'est que le support visuel d'une analyse statistique ; comme les méthodes statistiques, à notre disposition, sont variées, il est utile d'étudier si elles aboutissent à des cartes différentes ou non, pourquoi et comment ?

Dans un premier temps, chacune des méthodes sera expliquée et donnera lieu à une carte.

Dans un deuxième temps, les résultats seront analysés par chacun des 2 intervenants.

Enfin, en conclusion, nous essaierons de noter les différences ou analogies entre ces différentes cartes, mais, d'en voir aussi les limites.

## Tableau des données

Accroissement naturel - en % - 1982.

Algérie	34	Jordanie	37	Qatar	31 (1)
Arabie-Saoudite	31	Koweït	32	Somalie	23
Bahrein	35 (1)	Liban	20	Soudan	27
Djibouti	25 (2)	Libye	34	Syrie	41
Egypte	24	Maroc	25	Tunisie	25
E.A.U. (3)	25	Mauritanie	24	Yemen (4)	26
Irak	34	Oman	32	Sud-Yemen (5)	29

## Sources :

L'expansion N° 266/266 bis 109 Juillet/5 Septembre 1985.

(1) Chiffres vers 1975 d'après Péninsule arabique d'aujourd'hui CEROAC - Tome I - sous la direction de P. Bonenfant Editions du CNRS 1982.

(2) Annuaire de l'Afrique et du Moyen-Orient 1981 - 82 publié par le groupe Jeune Afrique.

(3) Emirats Arabes Unis.

(4) République Arabe du Yemen ou Nord-Yemen.

(5) République Démocratique du Sud-Yemen.

(1) zonal : type d'implantation (un pays = une zone)

ordonné : un caractère ordonné exprime une relation d'ordre ; ici, les chiffres exprimés en % sont ramenés à des groupes et ordonnés.

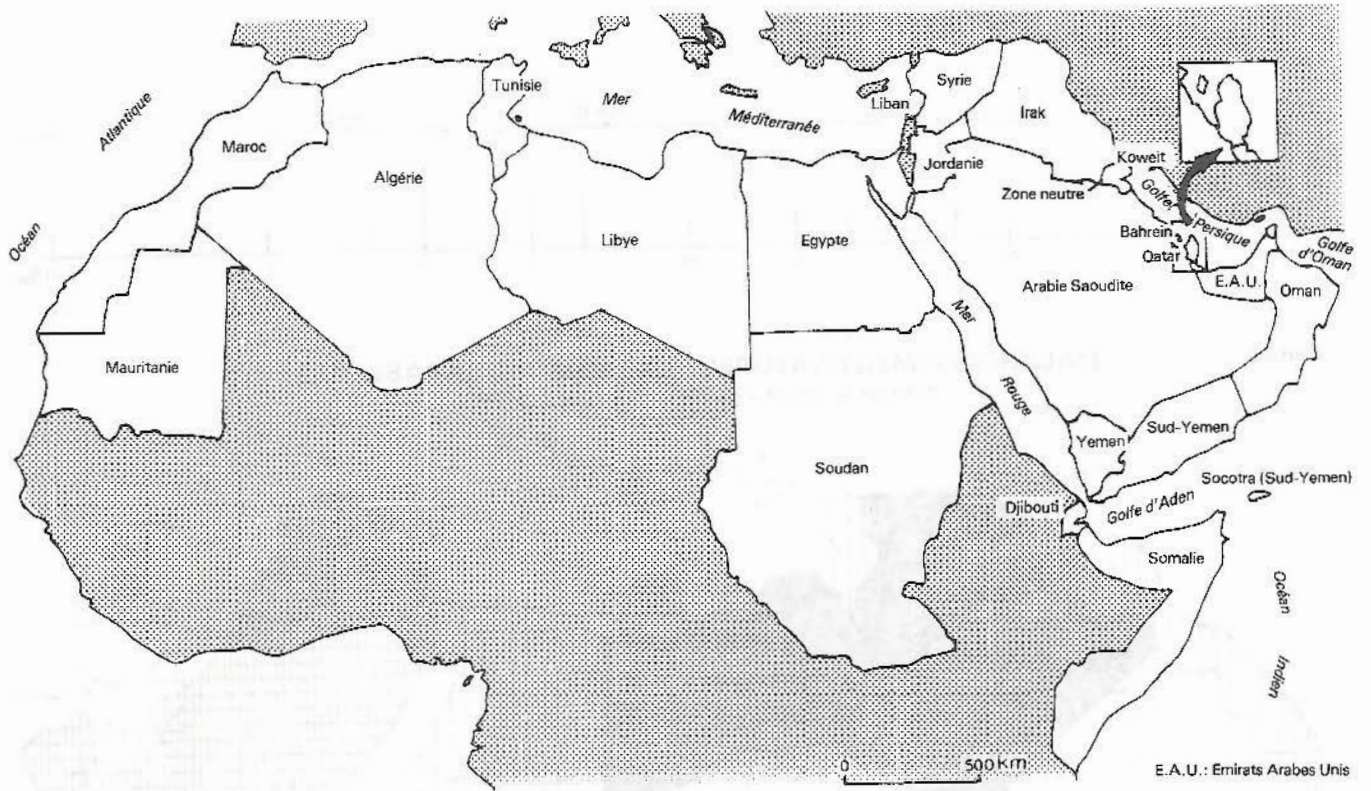
(2) Maghreb-Machrek : les pays pris en compte sont ceux de la ligue arabe, voir carte 1.

\* Jean-Paul BORD - Maître Assistant, Responsable de l'Atelier de cartographie de l'Université de TOURS. Membre du L.A. 365. URBAMA (Urbanisation du Monde Arabe). TOURS.

Guy LANDURE - Maître de conférence, Géographie urbaine et statistique. Université de TOURS.

Carte 1

MONDE ARABE



# 1. Présentation des 4 méthodes et résultats cartographiques

## A) D'APRÈS LA COURBE DE RÉPARTITION DES VALEURS

### 1) Sans tenir compte de la moyenne (carte 2)

Pour représenter une série quantitative (quantités relatives, ici, en ‰) le nombre de paliers ou groupes n'est pas déterminé par avance. Il dépend de l'homogénéité plus ou moins grande des groupes de données identifiés sur la courbe de répartition ou de distribution des valeurs.

La répartition des données sur la courbe peut donner lieu à des groupes divers selon les auteurs (3) ; si l'on procède de manière logique, les valeurs sont réparties en 7 groupes non homogènes :

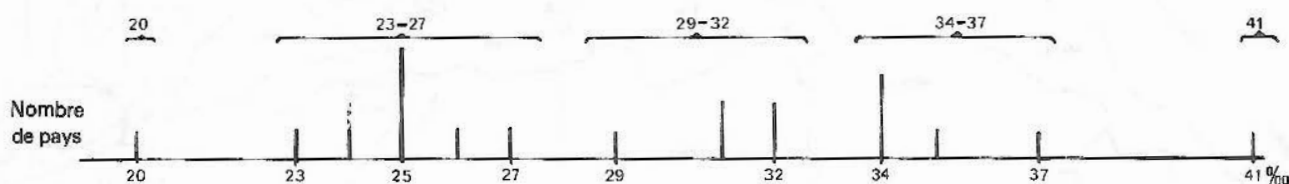
20, 23 - 27, 29, 31 - 32, 34 - 35, 37 et 41

ce premier découpage peut, par regroupement des données proches, donner une deuxième répartition avec moins de groupes, donc plus facile à mémoriser visuellement :

20, 23 - 27, 29 - 32, 34 - 37, 41 (comme sur la courbe ci-dessous)

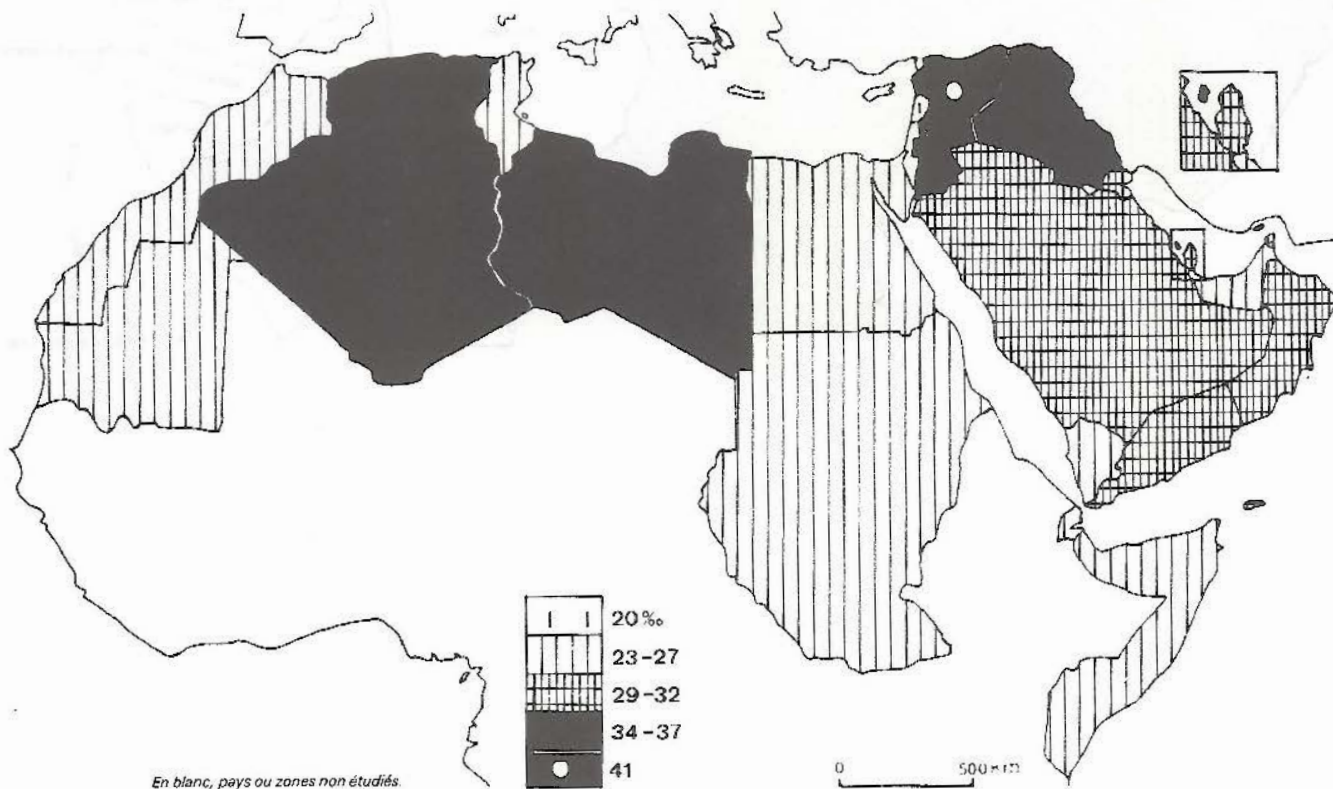
A chacun des groupes est affecté une valeur, reportée ensuite sur la carte 2 ; il en sera d'ailleurs de même pour toutes les autres cartes.

Courbe 1



Carte 2

## L'ACCROISSEMENT NATUREL DANS LE MONDE ARABE. 1982 d'après la courbe de répartition des valeurs



En blanc, pays ou zones non étudiés.

(3) Le nombre de classes peut par exemple se calculer par les formules  
 $K = 1 + 3,3 \log_{10} n$  (Huntsberger)

$K = 5 \log 10^n$  (Brooks-Carruthers)

**2) En tenant compte de la moyenne (carte 3)**

Il faut tout d'abord indiquer que la moyenne, ici, 28,4 ‰, est obtenue par les chiffres absolus et non en faisant la somme des ‰, divisés par le nombre de pays (dans ce 2<sup>e</sup> cas, le chiffre obtenu serait un peu différent, 29,2 ‰ pour le cas présent).

Si l'on tient compte de la moyenne, la répartition des groupes est modifiée.

Le chiffre moyen ( $m = 28,4$ ) représenté, ici, par le Sud-Yemen (29 ‰) répartit les valeurs en 5 groupes :

2 groupes extrêmes : 20 et 41

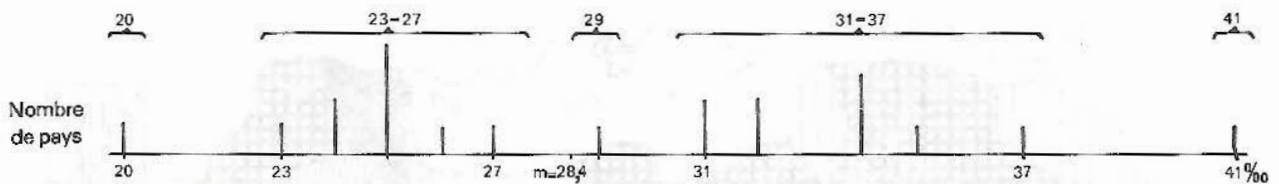
2 groupes plus proches de la moyenne : 23-27 et 31-37 et la moyenne.

Le seul problème est posé, ici, par le chiffre 37 (Jordanie) qu'il est plus logique (par sa proximité) d'intégrer au groupe 31-35 qu'à 41.

Dans ce cas, 2 types de figurés sont possibles :

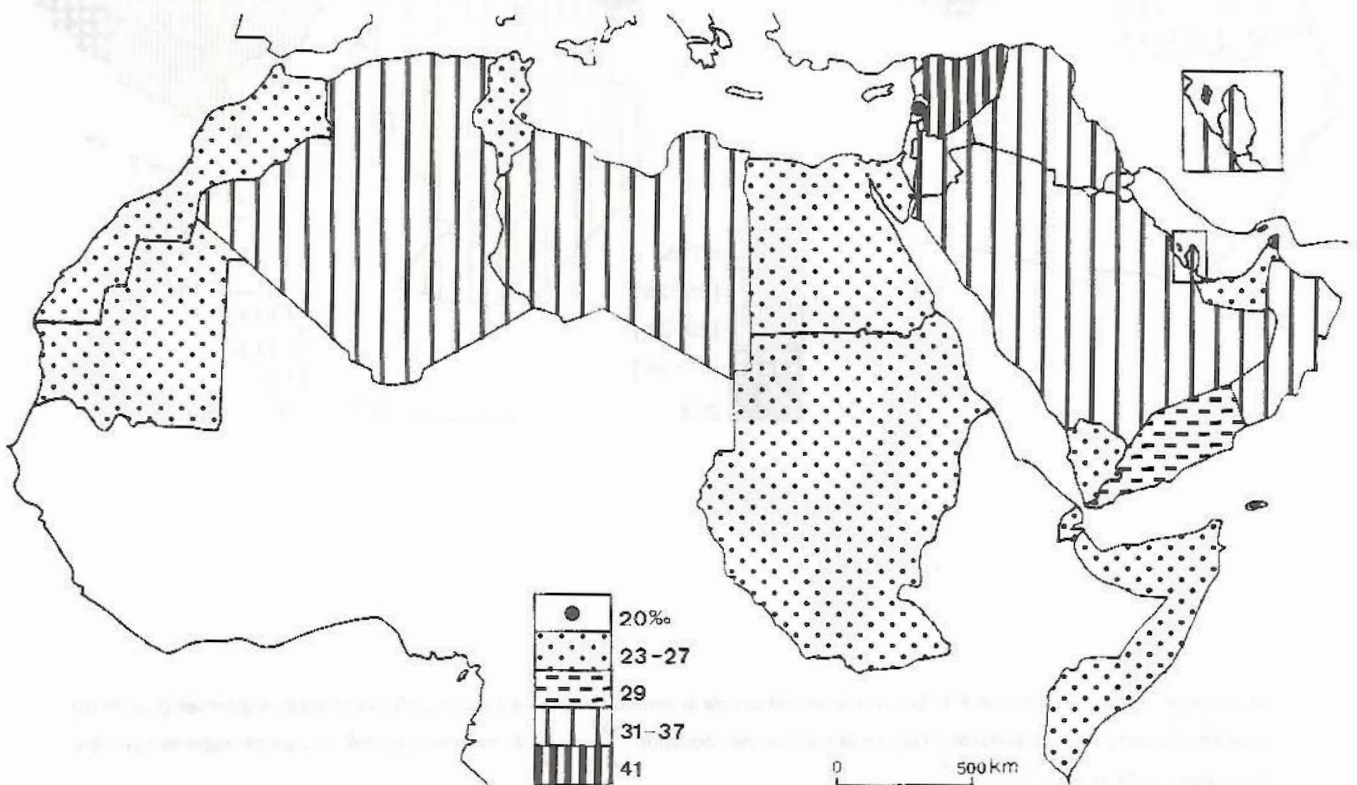
- soit ceux adoptés carte 2 (progression régulière du blanc au noir).
- soit, une double variation ordonnée rendue par l'utilisation conjointe de trames-traits et de trames-points de part et d'autre de la moyenne représentée en tiretés (carte 3)

Courbe 2



Carte 3

**L'ACCROISSEMENT NATUREL DANS LE MONDE ARABE. 1982**  
en tenant compte de la moyenne arithmétique



## B) D'APRÈS 2 MÉTHODES STATISTIQUES

### 1) Selon les quartiles (4) (carte 4)

— Les quartiles indiquent, approximativement, comment l'ensemble des données est distribué sur l'étendue de la série.

Lorsque nous connaissons la plus haute et la plus petite valeur de la variable, ainsi que les 3 quartiles (Q1 quartile inférieur ; Q2 médiane ; Q3 quartile supérieur) nous

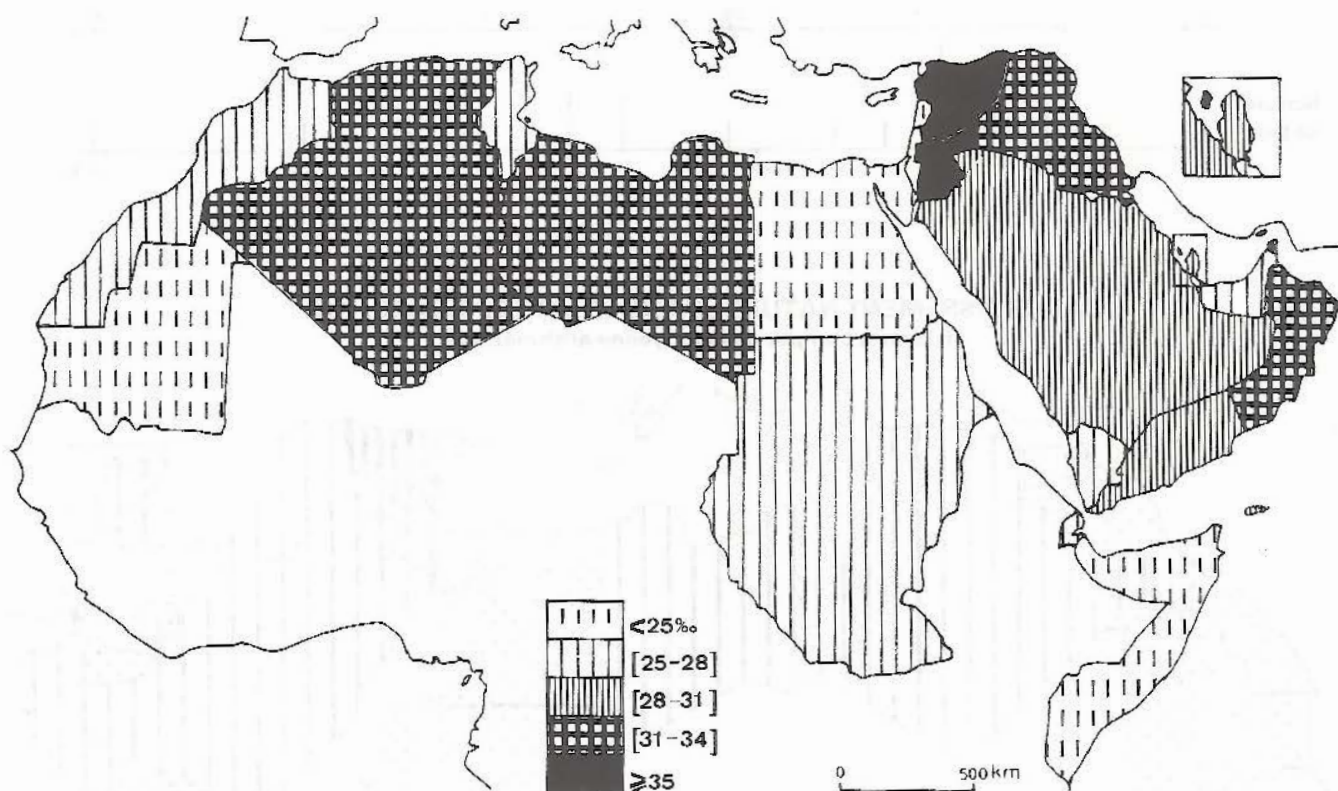
avons déjà une première réduction des données, celles-ci se trouvant "condensées" en cinq nombres.

— L'intérêt de leur utilisation réside dans le fait que nous pouvons faire abstraction des valeurs extrêmes de la distribution (50 % de l'effectif), valeur qui, généralement, "influencent" fortement la distribution et que l'on peut parfaitement visualiser, en traçant la courbe des effectifs/fréquences (oblique gauche/droite).

En ce sens, l'interquartiles a l'avantage de ne pas contenir de données anormalement grandes ou petites.

Carte 4

### L'ACCROISSEMENT NATUREL DANS LE MONDE ARABE. 1982 d'après les quartiles



(4) quartile : si l'on désigne par  $F(x)$  le nombre de valeurs de la série inférieures à  $x$  et par  $n$  l'effectif total, le premier quartile ou quartile inférieur peut être considéré comme la solution de l'équation  $\frac{F(x)}{n} = \frac{1}{4}$ , le troisième quartile ou quartile supérieur comme la solution de l'équation  $\frac{F(x)}{n} = \frac{3}{4}$

Médiane (quantile d'ordre  $\frac{1}{2}$ ). On appelle médiane d'une série de données tout nombre tel que l'effectif des valeurs inférieures soit égal à l'effectif des valeurs supérieures à ce nombre. La médiane tient compte du rang de tous les individus et non des valeurs du caractère.

## 2) Selon la variable centrée réduite (5) carte 5

L'opération consiste, ici, à exprimer toutes les valeurs de la variable indépendamment de l'unité de mesure. En conséquence, la variable centrée réduite ou "standardisée" a pour moyenne 0 et écart-type 1.

L'avantage principal d'une telle opération, est de pouvoir comparer directement la position d'un individu dans

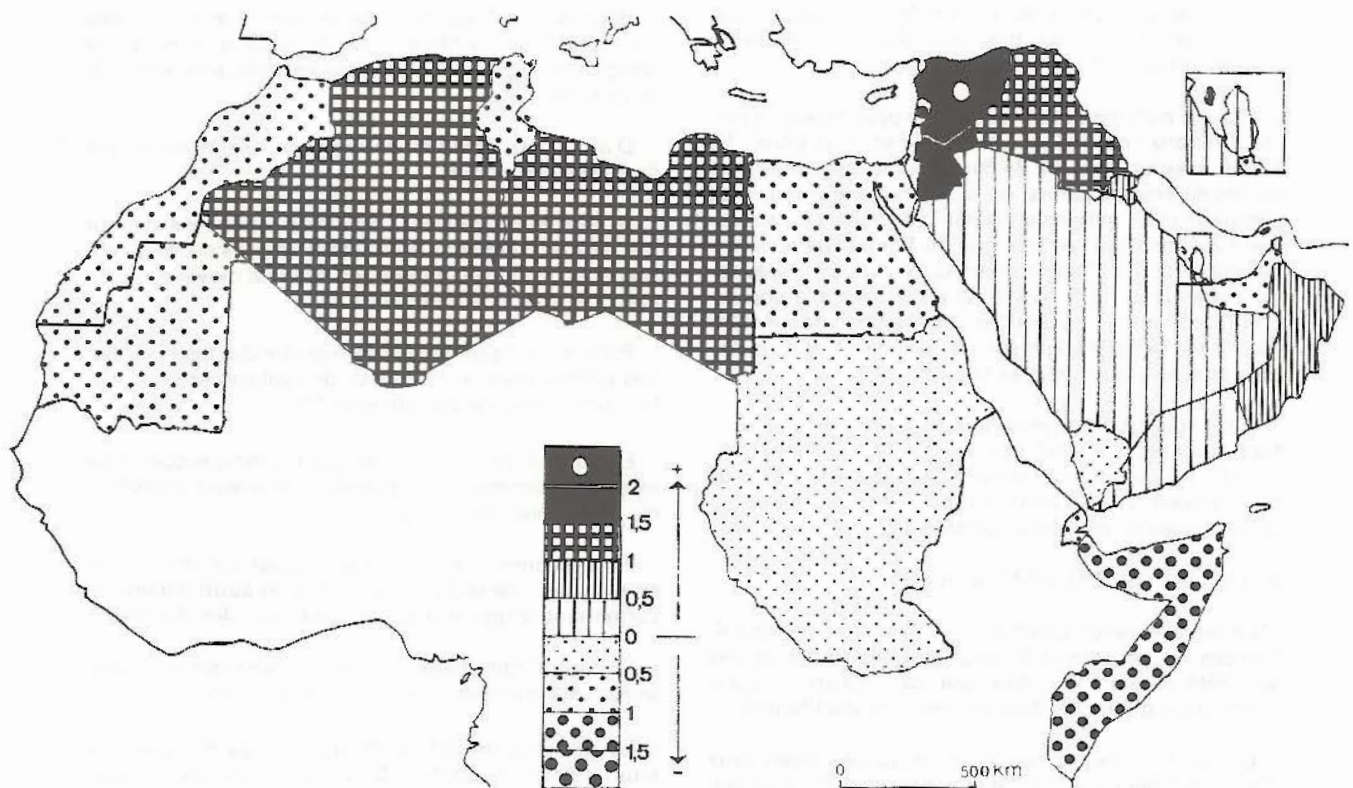
des distributions de nature et grandeurs différentes, quels que soient l'unité de mesure, la valeur des écart-types et des moyennes.

Signalons, en dernier lieu, qu'en aucune façon, la standardisation ne donne une mesure de la disposition pour l'ensemble de la variable mais pour chaque individu.

Elle rend possible la comparaison sur les mêmes bases, d'un phénomène unidimensionnel ou univarié.

Carte 5

### L'ACCROISSEMENT NATUREL DANS LE MONDE ARABE. 1982 d'après la variable centrée réduite



## II. L'analyse des résultats cartographiques par les 2 intervenants :

- le cartographe
- le statisticien

### A) LES RÉSULTATS VUS ET ANALYSÉS PAR LE CARTOGAPHE

#### 1) Quatre ensembles distincts

Au niveau supérieur de l'analyse, et quelle que soit la méthode adoptée, 4 ensembles apparaissent selon un axe méridien ; d'Ouest en Est :

- 1<sup>er</sup> ensemble : Mauritanie - Maroc (respectivement 24 et 25 ‰) où le taux d'accroissement naturel est faible, pour la zone étudiée ;

- 2<sup>e</sup> ensemble : Algérie - Libye (34 ‰) avec un fort taux d'accroissement ; 2<sup>e</sup> ensemble qui ressort fortement vu l'étendue des deux pays, mais, où la Tunisie (avec un taux de 25 ‰) reste toujours à part ;

- 3<sup>e</sup> ensemble : Egypte, Soudan, Djibouti et Somalie (taux entre 23 et 27 ‰), qui avec ses taux faibles, rappelle le premier ensemble ;

- 4<sup>e</sup> ensemble : plus hétérogène, avec un bloc où les taux sont très forts (Syrie, Jordanie, Irak, 34 à 41 ‰) qui jouxte le pays (Liban) où ils sont le plus faible

(5) Variable centrée, réduite ou standardisée : c'est l'opération qui consiste à exprimer toutes les valeurs d'une variable indépendamment de l'unité de mesure par :

centrage  $(xi - \bar{x})$ , chaque valeur est transformée en différence algébrique à la moyenne arithmétique, réduction  $\frac{(xi - \bar{x})}{\text{écart-type}}$ , chaque différence algébrique est exprimée en écart-type, nouvelle unité de mesure.

(20 %) et, un autre bloc constitué par l'Arabie Saoudite et les pays du Golfe, où les différences, bien que marquées, sont moins nettes (d'un côté, Arabie Saoudite, Sud-Yemen, Oman, Qatar et Bahrein, avec des taux compris entre 29 et 35 %, et de l'autre, E.A.U. et Yemen Nord, avec des taux plus faibles (respectivement 25 et 26 %).

Mais, cette analyse générale présente cependant des nuances suivant la méthode adoptée.

## 2) Une analyse plus nuancée dans le détail

Ces nuances tiennent essentiellement à 2 facteurs. Tout d'abord, la visualisation elle-même montre, notamment au Moyen-Orient, des différences sensibles suivant la méthode adoptée. Ainsi, la première méthode indique une opposition Nord-Sud très nette, alors que la méthode 4 nuance quelque peu cette opposition.

D'autre part, l'étendue même des pays fausse la réalité. Prenons l'exemple de la Tunisie et de la Libye. La Libye, avec un taux de 34 % apparaît très fortement sur les différentes cartes, or, si l'on ramène le taux à sa population (3,7 millions d'habitants en 1984) l'accroissement naturel n'est, en fait, que de 125 800 personnes ; par opposition, la Tunisie, avec un taux de 25 % apparaît faiblement, or, si l'on ramène son taux à la population (7 millions d'habitants en 1984) son accroissement naturel est de 175 000 personnes, soit plus fort en quantités absolues que celui de la Libye.

A cet égard, une carte où seraient exprimées les quantités absolues donnerait des résultats sensiblement différents, ou une anamorphose tenant compte non plus de la superficie mais de la population de chaque pays, comme base de référence, serait plus près de la réalité.

## B. L'ANALYSE DU STATISTICIEN

La représentation graphique par la courbe de répartition des valeurs permet de visualiser simplement, et ceci sans référence à une quelconque caractéristique d'une distribution d'effectifs/fréquences, une distribution.

Ce que l'on peut retenir ici, c'est que les divers taux d'accroissement sont, d'une manière générale, tous élevés et que s'opposent deux ensembles :

- les pays aux taux les plus forts (Algérie, Libye, Arabie Saoudite, pays du Moyen Orient)
- aux pays au taux un peu plus en recul (Maroc, Mauritanie, Egypte, Soudan, Somalie)

Ce procédé apparaît comme le plus élémentaire, mais est sujet à caution, car fortement tributaire (par définition) des groupements effectués par l'opérateur.

Les résultats obtenus par la carte 3 est fondamentalement différent.

Tout d'abord, la visualisation permet de saisir un ordre de grandeur (référence à la moyenne arithmétique). Chaque pays est, ici, "positionné" par rapport à ce paramètre.

Cette approche peut aussi être considérée comme élémentaire, le calcul se faisant à partir des valeurs absolues et non pas relatives comme dans le premier cas.

Comme précédemment, nous voyons clairement apparaître une opposition (moins marquée cependant) entre les pays ayant les taux les plus élevés (Algérie, Libye, Arabie Saoudite, Emirats) et ceux ayant les taux les plus faibles (Maroc, Mauritanie, Tunisie, Egypte, Soudan, Somalie).

Le seul problème qui se pose — au-delà du raisonnement par les moyennes — est l'influence opérée par les valeurs extrêmes de la distribution (Jordanie 37 % ; Liban 20 % ; Syrie 41 %).

Rappelons qu'une moyenne arithmétique, n'a de sens que lorsque la distribution est homogène, c'est à dire lorsqu'elle est peu influencée par les valeurs extrêmes de la variable.

D'autre part, il est bon de ne pas perdre de vue ce que l'on veut démontrer.

Si dans le cas présent l'opérateur voulait insister sur des taux de croissance élevés dans les pays du Maghreb-Machrek, il pourrait utiliser non pas une moyenne arithmétique mais quadratique (6).

Par contre, il pourrait faire apparaître des taux peu élevés en travaillant sur l'inverse des valeurs et en calculant une moyenne harmonique (7).

En résumé, toute opération ayant comme support une moyenne, demande, au préalable, une analyse approfondie de la série statistique.

Il est recommandé de ne pas l'utiliser comme une fin en soit, mais de toujours la relier à un autre paramètre permettant d'appréhender la dispersion des données.

La carte 4 fait référence à cette dispersion par l'utilisation des quartiles inférieurs et supérieurs.

Le calcul donne  $Q1 = 25$  et  $Q3 = 34$  %, avec un taux médian à 29 %. Dans cet intervalle, quatre ensembles peuvent être distingués :

- le groupe formé par le Maroc, la Tunisie, Djibouti et E.A.U. avec un taux égal à 25 %.
- le groupe Yemen du Nord et Soudan respectivement 26 à 27 %
- le groupe Arabie Saoudite, Qatar, Koweït et Oman avec un taux variant de 31 à 32 %.
- un dernier ensemble formé par l'Irak, la Libye et l'Algérie où ce taux est égal à 34 %.

Nous faisons, ici, abstraction des valeurs extrêmes. Les quartiles — à la différence des opérations précédentes — nous indiquent sensiblement comment l'ensemble des données est distribué sur l'étendue de la série. D'une manière générale, nous voyons, ici, que pour l'ensemble des cas analysés, les taux d'accroissement varient de 25 à 34 % (valeurs observées 20 à 41 %).

(6) La moyenne quadratique de  $n$  valeurs est la racine carrée de la moyenne arithmétique des carrés des  $n$  valeurs.

(7) La moyenne harmonique de  $n$  valeurs est le nombre dont l'inverse est la moyenne arithmétique des inverses des  $n$  valeurs.

Nous avons opéré une réduction des données et fait une première synthèse de l'information. Lorsque la série initiale est fortement dissymétrique (forte influence des plus petites/grandes valeurs de la variable) on a tout intérêt à procéder de la sorte.

Cette première approche, dans le but de mieux cerner la dispersion des données, trouve sa quintessence par centrage et réduction de la variable.

Ce procédé est présenté sur la carte 5. Nous avons, pour cette série, moyenne arithmétique = 29,2 et écart-type = 5,3.

Première constatation : la faible valeur de l'écart-type indique une accumulation forte des données autour de la moyenne. (Un peu plus de 76 % des données dans l'intervalle  $\bar{x} \pm \delta$ , avec  $V(8) = 0,18$ ).

Tous calculs faits, nous voyons que (Z) (9) varie de -1,7 à +2,2. Ceci confirme notre premier constat ; la carte 5 va, en ce sens, à l'essentiel.

Les taux d'accroissement (maintenant exprimés en écart-type) nous montrent que cinq pays se caractérisent nettement de l'ensemble étudié : l'Algérie, la Libye, l'Irak, la Syrie et la Jordanie.

Cette façon de procéder nous permet également de calculer les taux d'accroissement "théoriques" pour chacun des cas analysés dans l'hypothèse d'une distribution normale.

Cette méthode trouve tout son intérêt lorsque l'on veut comparer des distributions dans le temps (prix, population, influence électorale d'un parti...) et détecter les causes d'un déclin ou d'un renouveau, d'une baisse ou d'une hausse.

## Conclusion

### UN ESSAI DE SYNTHÈSE

A travers cette étude, des recoupements mais aussi des différences apparaissent.

#### 1) Les points communs aux deux analyses

Quelle que soit la méthode adoptée, la répartition fait apparaître clairement des ensembles où les taux sont, soit élevés, soit faibles :

pays où les taux sont faibles : Maroc, Mauritanie, Egypte, Soudan, Somalie, Djibouti ;

pays où les taux sont forts : Algérie, Libye, pays du Moyen-Orient.

Cependant, et notamment pour le bloc Moyen-Orient, les contrastes sont plus différenciés suivant la méthode retenue.

#### 2) Des différences qui tiennent plus à l'analyse qu'aux résultats visuels

Pour le cartographe la carte reste le support essentiel de l'analyse, alors que pour le statisticien le traitement des données est fondamental, le résultat cartographique venant, pour sa part, confirmer les premières hypothèses.

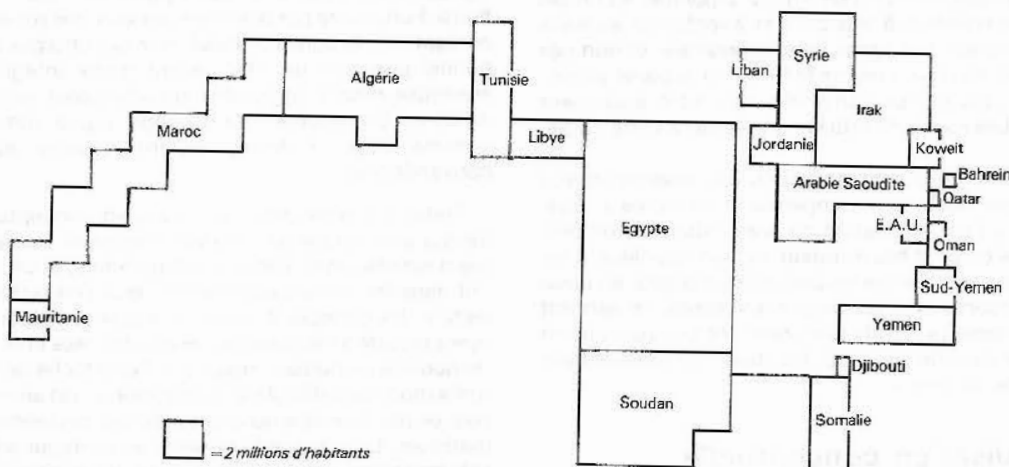
Mais quelle que soit l'information qui est support d'une quelconque analyse, il faut l'interpréter avec prudence.

Dans le cas présent, une anamorphose, où les superficies des pays seraient relatives à leur population, donnerait des résultats plus conformes à la réalité (carte n° 6, ci-dessous).

L'anamorphose géographique est une image déformée d'une carte où la surface des zones (ici les pays) est fonction (dans l'article) de la population de chaque pays ; dans la mesure du possible, sont respectées la forme d'ensemble de chaque pays, les positions relatives des pays et la forme générale de la région géographique étudiée.

Carte 6

### ANAMORPHOSE : MONDE ARABE. 1984 Superficie des pays relative à leur population



(8)  $V$  = rapport de l'écart-type à la moyenne arithmétique ; coefficient de variation ou de dispersion relative.  
(9)  $Z$  = déviation/différence algébrique de chacune des valeurs à sa moyenne que divise l'écart-type.