



« Analyse de données et cartographie – Devoir sur table »
Etude de la ségrégation urbaine au Cap (Afrique du Sud)

Hadrien Commenges, Université Paris 1 Panthéon-Sorbonne,
 UMR 8504 Géographie Cités

Pour citer cette feuille : Commenges, H., 2017, « Analyse de données et cartographie – Devoir sur table. Etude de la ségrégation urbaine au Cap (Afrique du Sud) », *Feuilles de géographie*, Feuille 2017-2, 6 pages.

Type de Feuille	Feuille d'évaluation (devoir sur table)
Niveau	Licence 3 (contrôle de connaissance des méthodes de modélisation statistique bivariée et multivariée niveau Licence 3)
Durée	2h – 2h30
Objectifs	<ul style="list-style-type: none"> - Sujet : étude de la ségrégation urbaine au Cap (Afrique du Sud). Plusieurs analyses statistiques sont effectuées sur des variables issues du recensement sud-africain de 2010. Le tableau initial compte 851 observations correspondant aux quartiers de la ville du Cap (sub-places), quartiers caractérisés par une plusieurs variables socio-éco-démographique. S'ajoutent à ces variables classiques, des variables qui disent si le quartier considéré est un ancien

township et s'il s'agit d'un township « black » ou d'un township « coloured ».

- Type d'analyses : régression linéaire simple (bivarié), analyse de la variance (bivarié), analyse de la covariance (multivarié)
- Matériel nécessaire : le document présente des résultats d'analyses faites au préalable. L'étudiant peut faire l'exercice avec une calculatrice simple (voire à la main).

Mots-clés

Modélisation statistique ; régression linéaire ; ANOVA, ANCOVA ;
Le Cap ; Afrique du Sud

Analyse de données et cartographie – Devoir sur table

Sujet : étude de la ségrégation urbaine au Cap (Afrique du Sud)

Éléments de contexte

Catégories raciales (extrait de Giraut F. et Vacchiani-Marcuzzo C.) : la dénomination de groupes « raciaux » en Afrique du Sud, et plus particulièrement dans l’Afrique du Sud post-apartheid, est un problème ardu doublé d’un sujet politiquement sensible. La classification raciale en vigueur sous l’apartheid comprenait : les « Blancs » (Sud-africains ayant des origines européennes) ; les « Indiens », Sud-africains d’origine indienne ; les « Africains » ou « Noirs » ou « Natifs », Sud-africains d’origine africaine ; les « Métis » ou « Colorés » (Coloureds), groupe hétérogène incluant les personnes ayant des origines mélangées.

Cette classification a joué un rôle important dans la construction identitaire. Tout en condamnant le système qui l’a conçue, les chercheurs l’emploient toujours abondamment. Nous suivrons les conventions usuelles et emploieront les catégories ci-dessus telles qu’elles sont comprises dans le contexte sud-africain.

Townships : avant et pendant l’apartheid (1948-1994) désigne les quartiers urbains planifiés destinés à recevoir la population noire (*black*) et métisse (*coloured*). Le plus connu des townships est Soweto à Johannesburg (**South Western Township**). Dans la période post-apartheid, les townships continuent d’exister en tant que quartiers mais il n’y a plus, comme sous l’apartheid d’assignation à résidence. Les townships ayant des densités de population très élevées : dans la période post-apartheid leur population s’est installée dans les interstices de la ville, souvent à proximité dans les zones tampon entourant les townships.

Description des données

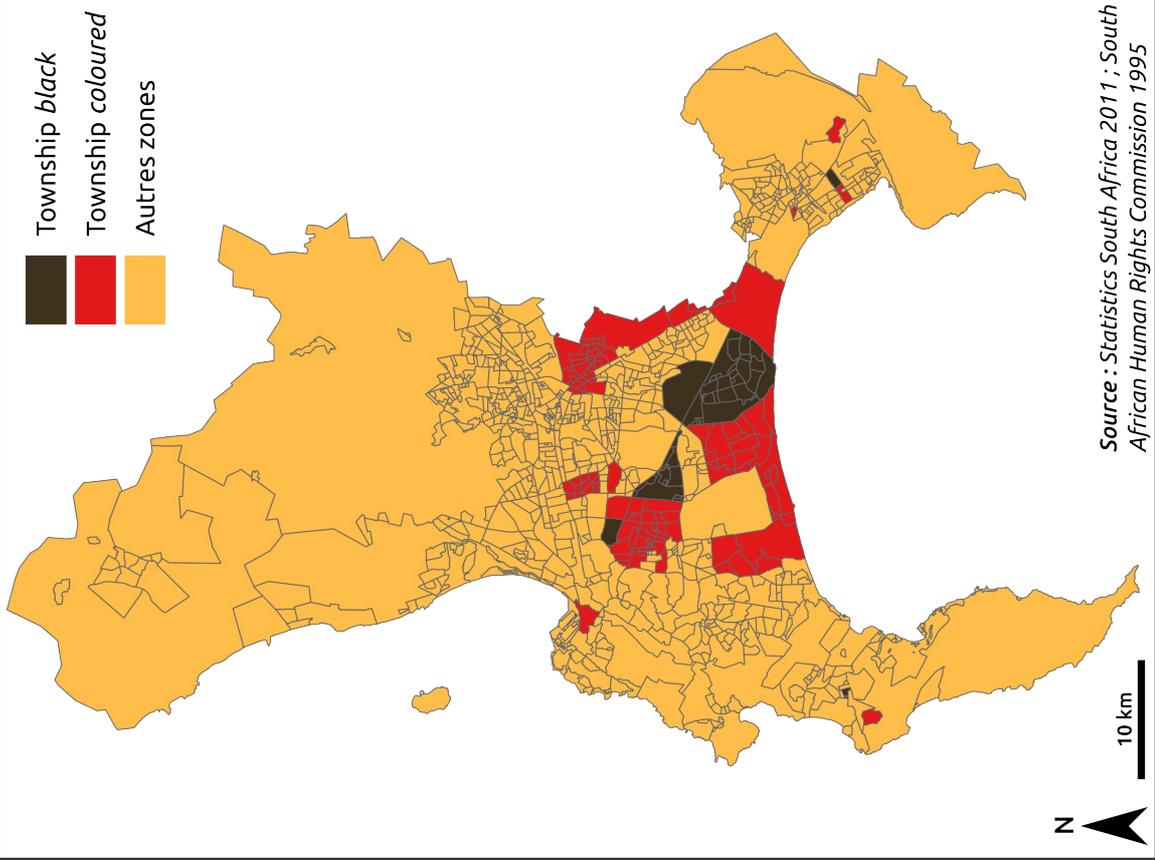
On dispose d’un tableau issu du recensement sud-africain de 2010. Ce tableau compte 851 observations correspondant aux quartiers de la ville du Cap (sub-places), quartiers caractérisés par une trentaine de variables socio-éco-démographique. Les variables peuvent être exprimées en absolu (nombre de chômeurs) ou en relatif (pourcentage de chômeurs). S’ajoutent à ces variables classiques, des variables qui disent si le quartier considéré est un ancien township et s’il s’agit d’un township « black » ou d’un township « coloured ».

Références

Sur l’Afrique du Sud : Giraut F, Vacchiani-Marcuzzo C. (2009) *Territoires et urbanisation en Afrique du Sud. Atlas et base de données géo-historique*, Institut de Recherche pour le Développement.

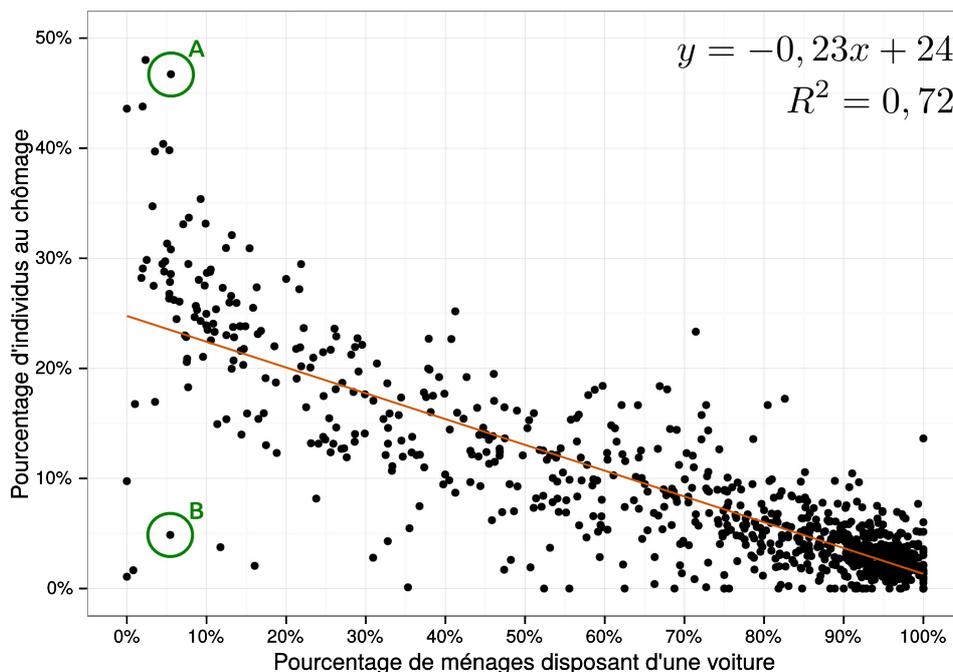
Sur l’analyse de la variance : Weiss N.A. (2012) *Introductory statistics*, 9th edition, Addison-Wesley.

Le Cap : localisation des townships



1. Exercice 1

1.1. Examiner le graphique suivant et expliquer l'analyse à laquelle il correspond : quel type de variables ; quel type d'analyse ; quelle est l'hypothèse explicative sous-jacente dans cette relation ; quelle serait l'hypothèse inverse ($y =$ voiture et $x =$ chômage) ; comment interpréter l'équation de la droite et le R^2 .



Attention les pourcentages prennent des valeurs entre 0 et 100 (et non entre 0 et 1)

1.2. On s'intéresse aux quartiers A et B indiqués sur le graphique. Tous deux ont une valeur de 7 % en abscisse (x) mais des valeurs très différentes en ordonnée (y) : 5 % pour B, 47 % pour A. Calculer, pour ces deux quartiers, la valeur espérée de y sachant la valeur de x , qualifier (en termes statistiques) et quantifier l'écart entre la valeur observée et la valeur espérée. Proposer des hypothèses pour expliquer la situation particulière de ces deux quartiers.

Exercice 2

On cherche à mettre en relation les deux variables suivantes : pourcentage de population noire (PCTBLACK) et township noir (TWN_BLACK). Cette seconde variable a deux modalités : le quartier est un ancien township noir (TRUE), le quartier n'est pas un ancien township noir (FALSE).

2.1. Expliquer l'analyse à effectuer : quel type de variables ; quel type d'analyse ; quelle est l'hypothèse explicative sous-jacente dans cette relation.

2.2. À partir du tableau suivant, faire des hypothèses simples sur la répartition de la population noire au Cap.

TWN_BLACK	Moyenne de PCTBLACK	Écart-type de PCTBLACK	Nbr. Observations
TRUE	98,1	2,7	50
FALSE	15,6	20,2	801
ENSEMBLE	20,4	27,6	851

2.3. Analyser le tableau suivant (extrait des 6 premières lignes du tableau complet de 851 observations) :

- Nommer les variables C1, C2, C3 et C4 et dire à quoi elles correspondent.
- Expliquer la décomposition de la variance.
- Calculer le R^2 .
- Conclure sur la relation entre le pourcentage de noirs et la catégorie township noir.

SPNAME	PCTBLACK	TWN_BLACK	C1	C2	C3	C4
Hanover Park	2,37	FALSE	20,42	15,57	-4,85	-13,2
Manenberg	11,69	FALSE	20,42	15,57	-4,85	-3,88
New Rest	97,71	TRUE	20,42	98,05	77,64	-0,34
Kanana	99,13	TRUE	20,42	98,05	77,64	1,08
Scarborough	16,37	FALSE	20,42	15,57	-4,85	0,8
Castle Rock	49,53	FALSE	20,42	15,57	-4,85	33,95
...
MOYENNE	20,42	-	20,42	20,42	0	0
VARIANCE	762,52	-	0	376,23	376,23	386,29

Exercice 3

On propose un modèle qui met en relation PCTUNEMPLOYED avec la variable TWN_BLACK. Le R^2 est égal à 0,23 et voici les moyennes pour chacune des deux modalités de la variable et pour l'ensemble.

TWN_BLACK	Moyenne de PCTUNEMPL	Écart-type de PCTUNEMPL	Nbr. Observations
TRUE	24,0	7,5	50
FALSE	7,3	7,2	801
ENSEMBLE	8,3	8,2	851

3.1. Expliquer l'analyse effectuée : quel type de variables ; quel type d'analyse ; quelle est l'hypothèse explicative sous-jacente dans cette relation ?

3.2. Y a-t-il une corrélation entre les deux variables ?

3.3. La variation du taux de chômage est-elle plus forte dans les anciens townships noirs (TRUE) ou bien dans les autres quartiers (FALSE) ?

3.4. Si on considère qu'il y a une corrélation entre TWN_BLACK et PCTUNEMPLOYED (ce modèle) et qu'il y a par ailleurs une corrélation entre TWN_BLACK et PCTBLACK (modèle de l'exercice 2). Que peut-on en conclure sur la relation entre PCTUNEMPLOYED et PCTBLACK ?

Exercice 4

On cherche à expliquer le taux de chômage (PCTUNEMPLOYED), variable notée par la suite y .

Considérer les trois modèles déjà calculés :

- **modèle 1** (exercice 1) : taux de chômage (PCTUNEMPLOYED) et pourcentage de ménages disposant d'une voiture (PCTCAR)
- **modèle 2** (exercice 2) : pourcentage de Noirs (PCTBLACK) et township noir (TWN_BLACK)
- **modèle 3** (exercice 3) : taux de chômage (PCTUNEMPLOYED) et township noir (TWN_BLACK)

On propose un nouveau modèle avec deux variables explicatives – TWN_BLACK et PCTCAR – toujours dans le but d'expliquer le taux de chômage (PCTUNEMPLOYED). Le coefficient de la variable PCTCAR est de -0,22 et le R^2 de ce modèle est de 0,73.

4.1. Expliquer l'analyse correspondant à ce troisième modèle : quel type de variables ; quel type d'analyse ; quelle est l'hypothèse explicative sous-jacente dans cette relation ?

4.2. Faire le graphique correspondant à cette analyse (très schématiquement).

4.3. Interpréter (d'un point de vue statistique) le coefficient de la variable PCTCAR ; interpréter (d'un point de vue statistique) le coefficient R^2 ; quel est l'apport de ce modèle par rapport aux modèles 1 et 2 ; proposer des explications géographiques,

4.4. On considère ce même modèle mais on ajoute un terme d'interaction entre les deux variables explicatives (TWN_BLACK et PCTCAR). Le coefficient de la variable PCTCAR reste égal à -0,22 et le R^2 reste égal à 0,73. Interpréter (d'un point de vue statistique) ce résultat et proposer des explications géographiques.