

**« Police et minorités visibles, les contrôles d'identité à Paris »  
Quelques réflexions statistiques sur une enquête.**

*Claudine Schwartz  
Claude Viterbo*

Ce texte vient en complément d'un article<sup>1</sup> sur une phrase d'E. Zemmour dont on a beaucoup parlé en mars, à savoir «*Les Français issus de l'immigration sont plus contrôlés que les autres parce que la plupart des trafiquants sont noirs et arabes... C'est un fait*». Nous y avons montré à quel point cette phrase souffre d'une carence de réflexion statistique ; on y indique aussi qu'une éventuelle sur-représentation des personnes perçues comme noires ou arabes (que nous noterons ici des PNA) dans la population de trafiquants arrêtés peut avoir de nombreuses explications, allant d'un biais auto-réalisateur évident, à celles liées par exemple à la concentration dans cette population de facteurs tels que la pauvreté et le chômage. Nous nous intéressons ici à la première partie de cette phrase, à savoir la sur-représentation des PNA dans la population contrôlée. Elle a été récemment étudiée dans l'enquête menée en 2007 et 2008 à la gare du Nord et au Châtelet, par Fabien Jobard et René Levy, chercheurs du CNRS à l'initiative de l'Open society de New-York. Un compte-rendu détaillé de cette enquête et des vidéos d'explication ont été publiés en 2009, en français, ici :

<http://www.cesdip.fr/spip.php?article428>  
<http://www.mediapart.fr/files/PoliceContrôles.pdf>

Ce compte-rendu explicite l'enjeu principal de l'enquête qui est de regarder les sur-représentations, au niveau des contrôles de police, de certaines sous-populations, notamment les jeunes, ceux qui ont de styles vestimentaires particuliers, des personnes perçues comme noires ou arabes (que nous noterons ici des PNA).

En marge de ce document, nous nous proposons de réfléchir sur des indices permettant de mesurer l'intensité de telles sur-représentations, en nous limitant à celles des PNA. Au départ d'une réflexion sur la définition de tels indices, on peut distinguer deux points de vue et définir différentes mesures qui quantifient la sur-représentation suivant ces points de vue.

### **Le point de vue de l'institution**

Si l'institution (ici la police nationale) donne mission d'interroger au hasard les gens dans un certain lieu, la proportion de PNA parmi les contrôlés doit être égale, aux fluctuations d'échantillonnage près, à celle des PNA dans la population de référence.

Un exemple de mesure de la sur-représentation des PNA parmi les contrôlés est le rapport entre les proportions de PNA parmi les contrôlés et dans la population de référence, que nous noterons  $\tau$  :

---

<sup>1</sup> <http://www.statistix.fr/spip.php?article79>

$$\tau = \frac{Prob(PNA/C)}{Prob(PNA)} \text{ où } C \text{ est l'évènement « être contrôlé ».}$$

### Le point de vue des personnes susceptibles d'être contrôlées

Ce qui importe est alors la probabilité d'être contrôlé. On cherche donc à savoir si elle est la même pour les PNA et les autres. Un exemple de mesure est le quotient de la probabilité qu'un PNA soit contrôlé par celle qu'un nonPNA le soit.

$$\kappa = \frac{Prob(C/PNA)}{Prob(C/nonPNA)}$$

Ces probabilités sont calculées à l'intérieur d'une population de référence, partitionnée en deux groupes, les PNA et les nonPNA.

Ces deux mesures étant définies, il n'y a plus qu'à les calculer ...

### L'enquête à la gare du nord de Paris

Cette gare voit passer environ 180 millions de personnes par an, c'est l'une des plus grandes gares d'Europe pour ce qui est du trafic de voyageurs. On ne peut pas prendre l'ensemble des voyageurs comme population de référence, ne serait-ce que parce que la composition vis-à-vis des critères retenus varie beaucoup selon les différents lieux de la gare. L'enquête a distingué trois lieux. Pour chacun de ces lieux, la population de référence est celle qui s'est trouvée disponible au contrôle : elle est composée de tous ceux qui sont passés dans ce lieu pendant le temps du contrôle. On sait aussi que le taux des personnes contrôlées, tant parmi les PNA que les autres, est très faible.

#### -Le hall d'entrée de la gare du Nord

L'enquête a conduit à recueillir certaines caractéristiques de personnes :

- d'un échantillon aléatoire de taille 7574 de la population de référence
- d'un échantillon aléatoire, indépendant du précédent, des personnes contrôlées dans la population de référence.

|            | PNA  | Non PNA | total |
|------------|------|---------|-------|
| contrôle   | 68   | 22      | 90    |
| population | 1920 | 5654    | 7574  |

On notera que la population de référence est ici à majorité composée de non PNA, tandis que « la plupart des contrôlés » sont des PNA.

Ces données nous permettent d'estimer  $Prob(PNA/C)$  par  $68/90$  et  $Prob(PNA)$  par  $1920/7574$  ; on peut donc calculer une estimation  $\hat{\tau}$  de  $\tau$  :

$$\hat{\tau} = (68/90)/(1920/7574) \sim 3,0$$

Il y a ainsi environ trois fois plus de PNA parmi les contrôlés que dans la population totale. Mais les données dont nous disposons ne nous permettent pas d'estimer  $Prob(C/PNA)$  et  $Prob(C/nonPNA)$  car la réunion de l'ensemble des 68 PNA contrôlés et des 1920 témoins ne constitue pas un échantillon de taille 1988 de l'ensemble des PNA de la population de référence. Nous allons donc procéder autrement pour estimer  $\kappa$ .

Introduisons l'indice appelé *rapport des cotes*, ou plus fréquemment *odds ratio*, noté  $OR$  et défini par :

$$OR = \frac{Prob(C/PNA)/Prob(nonC/PNA)}{Prob(C/nonPNA)/Prob(nonC/nonPNA)}$$

où nonC est l'évènement « ne pas être contrôlé ».

Comme on sait par ailleurs que le taux des personnes contrôlées est faible, tant parmi les PNA que les autres,  $Prob(nonC/PNA)$  et  $Prob(nonC/nonPNA)$ , sont très voisins de 1, et donc ici, l'odds ratio est peu différent de  $\kappa$ .

Par ailleurs :

$$\begin{aligned} OR &= \frac{Prob(C \text{ et } PNA) \times Prob(nonC \text{ et } non \text{ PNA})}{Prob(nonC \text{ et } PNA) \times Prob(C \text{ et } nonPNA)} \\ &= \frac{Prob(PNA/C)/Prob(nonPNA/C)}{Prob(PNA/NonC)/Prob(nonPNA/nonC)} \end{aligned}$$

Soit finalement, comme  $Prob(PNA/nonC) \sim Prob(PNA)$  :

$$OR = \tau \times \frac{Prob(nonPNA/C)}{Prob(nonPNA/nonC)}$$

On peut estimer  $Prob(nonPNA/C)$  par 22/90 et  $Prob(nonPNA/nonC)$  par 1920/7574, et prendre comme estimation  $\hat{\kappa}$  de  $\kappa$  l'estimation de  $OR$ , soit :

$$\hat{\kappa} = 9,1$$

Un PNA présent à la station de la gare du nord le jour du contrôle a ainsi environ 9 fois plus de chances de se faire contrôler qu'un non PNA.

### -La gare RER

Les données de l'enquête sont les suivantes :

|            | PNA  | Non PNA | total |
|------------|------|---------|-------|
| contrôle   | 81   | 26      | 107   |
| population | 3977 | 3630    | 7607  |

La population de référence est ici à majorité composée de PNA, « la plupart des contrôlés » étant aussi des PNA.

On a ici, en faisant les mêmes approximations que précédemment :

$$\hat{\tau} = 1,4 \text{ et } \hat{\kappa} = 2,8.$$

## Les quais des trains Thalys et eurostar

|            | PNA | Non PNA | total |
|------------|-----|---------|-------|
| contrôle   | 40  | 63      | 103   |
| population | 366 | 3218    | 3854  |

La population de référence est ici en très grande majorité composée de nonPNA (90%), une faible majorité des contrôlés étant aussi des nonPNA.

On a ici :

$$\hat{\tau} = 3,8 \text{ et } \hat{\kappa} = 5,6.$$

Comme on le voit dans ces trois exemples, les deux points de vues différents mènent à des mesures sensiblement différentes, évidemment liées ; le coefficient  $\kappa$  amplifie toujours le phénomène : si  $\tau > 1$ , alors  $\kappa > \tau$  et si  $\tau < 1$ , alors  $\kappa < \tau$ .

Afin de ne pas avoir deux indices, on pourrait accepter de compliquer un peu la première mesure définie,  $\tau$ , et de la remplacer par  $\tau'$  avec :

$$\tau' = \frac{\text{Prob}(PNA/C)/\text{Prob}(nonPNA/C)}{\text{Prob}(PNA)/\text{Prob}(nonPNA)}$$

Un calcul simple montre que  $\tau' = \kappa$ . On pourrait alors décider de considérer que  $\kappa$  est pertinent pour les deux points de vue et se limiter à ce seul indice, mais cela se discute. En effet, comme  $\kappa$  est ici une mesure « naturelle » pour le point de vue des personnes contrôlées et « peu naturelle » pour le point de vue de l'institution, on a tendance à l'interpréter de ce seul point de vue.

Par exemple, pour le hall de la gare du Nord, on dira facilement « les PNA sont environ 9 fois plus contrôlés que les autres », car tout le monde comprend cette phrase. Par contre, dire que « le rapport des proportions de PNA et des nonPNA est 9 fois plus grand dans la population contrôlée que dans la population de référence » est peu interprétable pour les non initiés. Donc le choix entre  $\tau$  et  $\kappa$  ou seulement  $\kappa$ , et de fait seulement l'odds ratio, n'est pas si simple !

L'odds ratio a des propriétés agréables, il est notamment facile de calculer des intervalles de confiance ; c'est une mesure très classiquement utilisée (mesure de l'intensité d'un facteur de risque pour une maladie par exemple) mais assez souvent, on ne s'intéresse dans ce type d'étude qu'à un seul point de vue, l'autre n'ayant éventuellement pas vraiment de sens. Le compte rendu de l'enquête sur les contrôles de police ne mentionne que l'OR, qui y est interprété en l'assimilant au coefficient  $\kappa$  défini ci-dessus. Mais comme tous les chiffres cités dans les tableaux ci-dessus sont données, chacun peut aussi calculer  $\tau$  ou interpréter de deux manières l'OR.

## La fluctuation d'échantillonnage

Si le contrôle se fait en choisissant des personnes au hasard, les rapports définis devaient être proches de 1, à la fluctuation d'échantillonnage près, fluctuation que nous avons jusqu'à présent négligée. Mais les écarts à 1 observés sont-ils significatifs ?

On peut dans chacun des trois lieux étudiés faire un test du chi-deux et l'hypothèse que les distributions des PNA sont les mêmes parmi les contrôles et parmi les non contrôlés est dans les trois cas rejetée au risque 0,01.

### *En guise de conclusion*

La population de référence dans le hall de la gare du Nord est un mélange de pauvres et de non-pauvres et le ciblage sur les PNA est de nature à augmenter sensiblement le nombre de pauvres parmi les contrôlés, et donc le nombre de trafiquants arrêtés<sup>2</sup>. Par contre en Gare RER, les voyageurs sont en très grande majorité pauvres: cibler les PNA n'a plus qu'un impact limité. Dans le cas extrême où les voyageurs sont tous pauvres et que l'on admet que la proportion de trafiquants parmi les PNA pauvres est la même que celle parmi les pauvres en général, cet impact serait nul. Mais il semble peu probable qu'une fois l'habitude prise de cibler les PNA dans certains contrôles, la police abandonne complètement cette stratégie, même lorsqu'elle n'est plus "productive".

Enfin nous n'avons pas abordé ni les aspects éthiques ni les aspects légaux de ces contrôles. Rappelons que les contrôle d'identité sont entre autres encadrés par les articles 78-1, 78-2 et 78-2-1 du code de procédure pénale<sup>3</sup>. De plus, un rapport de la cour de cassation, précise que dans les contrôles pour infraction à la loi sur le séjour des ressortissants étrangers, ni l'apparence physique, ni le fait de parler une langue étrangère ne peuvent justifier un contrôle<sup>4</sup>.

---

<sup>2</sup> D'après le site <http://www.apur.org/images/notes4pages/4P11.pdf>, à Paris, 12% des foyers sont pauvres (revenu <670 euros /mois) mais 32% si on considère les foyers dont le chef de famille est étranger non citoyen de l'UE. Bref, on a environ 3 fois plus de chances d'être pauvre si on est étranger hors UE. Donc en ne contrôlant que des étrangers hors UE, on a 3 fois plus de chances de tomber sur des pauvres. Imaginons que la probabilité d'être pauvre pour un PNA est aussi 32%. Si on suppose que parmi les pauvres il y a une proportion  $u$  de trafiquants, et  $v$  parmi les non pauvres (avec  $v < u$ ), sur 1000 PNA contrôlés, on contrôlera  $320u$  pauvres, au lieu de 120 si on contrôle 1000 personnes au hasard. Donc  $320u + 680v$  trafiquants sont arrêtés au lieu de  $120u + 880v$ . On vérifie facilement que si  $v < u$  alors  $120u + 880v < 320u + 680v$ .

<sup>3</sup> [http://www.legifrance.gouv.fr/affichCode.do;jsessionid=83720BED866C00E1C80A43C3F5B65C44.tpdjo02v\\_1?idSectionTA=LEGISCTA000006151880&cidTexte=LEGITEXT000006071154&dateTexte=20100407](http://www.legifrance.gouv.fr/affichCode.do;jsessionid=83720BED866C00E1C80A43C3F5B65C44.tpdjo02v_1?idSectionTA=LEGISCTA000006151880&cidTexte=LEGITEXT000006071154&dateTexte=20100407)

<sup>4</sup> [http://www.courdecassation.fr/publications\\_cour\\_26/rapport\\_annuel\\_36/rapport\\_2008\\_2903/etude\\_discriminations\\_2910/distinctions\\_justifiees\\_2918/controle\\_variable\\_distinctions\\_justifiees\\_2921/traitement\\_etrangere\\_situation\\_irreguliere\\_12145.html](http://www.courdecassation.fr/publications_cour_26/rapport_annuel_36/rapport_2008_2903/etude_discriminations_2910/distinctions_justifiees_2918/controle_variable_distinctions_justifiees_2921/traitement_etrangere_situation_irreguliere_12145.html)