

J. P. BENZÉCRI

## **Les sondages d'opinion : point de vue d'un statisticien**

*Les cahiers de l'analyse des données*, tome 5, n° 4 (1980),  
p. 475-480

[http://www.numdam.org/item?id=CAD\\_1980\\_\\_5\\_4\\_475\\_0](http://www.numdam.org/item?id=CAD_1980__5_4_475_0)

© Les cahiers de l'analyse des données, Dunod, 1980, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

LES SONDAGES D'OPINION :  
POINT DE VUE D'UN STATISTICIEN  
[SONDAGES]

par J. P. Benzécri (1)

Peut-être beaucoup croient-ils encore aujourd'hui que la statistique est la forme la mieux élaborée du mensonge ? En butte depuis longtemps aux soupçons, les statisticiens sont eux mêmes devenus des spécialistes raffinés de la critique. Alors que semble passé l'âge d'or des sondages (l'expression d'Alain Lancelot, renvoie à la décennie 64-74 de la 5-ème république) les rédacteurs de ce volume 1981 d'Universalialia, ont permis à un statisticien de montrer quels outils nouveaux le calcul électronique offre à qui veut décrire l'opinion.

En termes statistiques, le modèle sous-jacent aux sondages est simple. Une grande urne contient 50 millions de boules diversement colorées ; en observant la couleur de 1000 de ces boules, on a une approximation satisfaisante de la composition du contenu de l'urne : si par exemple on a trouvé 300 boules bleues, la part relative du bleu dans l'urne est 300/1000, l'erreur sur le numérateur 300 étant de l'ordre de la racine carrée de ce nombre. Du point de vue des conclusions pratiques une telle erreur ne cesse d'être négligeable que si l'on doit comparer des proportions presque égales : ou, pour user de termes concrets, si l'on doit prédire lequel aura la victoire de deux partis de même force.

La réalité, ici comme ailleurs, défie les modèles simplistes : toutes les boules ne se laissent pas également saisir ; certaines n'ont pas de teinte franche ; beaucoup s'harmonisent au fond, à l'instar du caméléon. De celui qu'il interroge l'enquêteur ne reçoit pas toujours de réponse ; et l'on peut douter que ceux qui acceptent de répondre le fassent avec sincérité. Il ne suffit pas de donner la parole à des sujets pour qu'ils en usent en toute responsabilité et liberté. On imagine sans peine que l'enquêteur peut solliciter l'enquêté, l'influencer : on prend sans doute moins garde à ce que comme nous l'a démontré L. Lebart, enquêteurs et enquêtés se cooptent en quelque sorte : car ce ne sont pas les mêmes portes qui s'ouvriront à un gendarme en retraite ou à un étudiant impécunieux ; et l'enquêteur qui doit pour satisfaire aux quotas, interroger trois représentants d'une catégorie professionnelle donnée, sait en général où les trouver : il reviendra souvent frapper à la porte qui s'est ouverte à lui une fois.

On nous a tant dit qu'il était laid de mentir, que nous avons fini par croire que celui qui ne répond pas la vérité est le seul coupable. Mais un proverbe anglais dit :

*Ask no questions, You will be talked no lies.*

---

(1) J.P. Benzécri, professeur de statistique. Université P. & M. Curie

(\*) Le présent article paraît simultanément dans Les Cahiers et dans UNIVERSALIA 1981.

En effet celui qui n'interroge pas s'expose le moins à entendre des mensonges. Il y a des questions indiscreètes qui visent une vérité objective que l'on dissimule : en qui demande le contenu d'une valise, comment ne soupçonnerait-on pas un douanier ? Mais surtout il y a la vérité subjective qui n'existe peut-être pas. Tout homme a un âge, tout employé un salaire ; il n'est pas vrai que nous ayons tous une opinion formée sur l'organisation économique idéale ou sur les hommes susceptibles de l'instaurer et de la diriger. Nous croyons au contraire qu'il faut de fortes suggestions pour que nous ayons en la matière au moins l'illusion d'une opinion. Toutefois sans avoir à proprement parler d'opinion on peut être contraint d'adhérer à un parti...

Parler n'est-ce pas toujours trop parler ? Il est facile de dire Oui quand on pense Non ; de travestir une valeur. Mais dans la suite des réponses à une batterie de questions est-il possible de ne pas laisser une image de soi-même. Tout comportement complexe n'est-il pas une signature inimitable ? Ce que vous avez dit n'est pas vrai ; mais il est vrai que vous l'avez dit. Un faux témoignage suggère à l'instruction perspicace la piste au bout de laquelle est l'irrécusable *scelus delicti*. Considérée isolément la fréquence d'une réponse (nous avons dit "d'une couleur") peut bien être dépourvue de sens, mais il vaut la peine d'accepter l'ensemble des réponses à un questionnaire comme un document verbal qui mérite d'être analysé en lui-même quelles que puissent être d'autre part les opinions et les influences qui l'ont fait naître.

Aujourd'hui, une telle étude ne se conçoit pas sans le secours de l'ordinateur : c'est cet outil électronique qui permet à l'IFOP ou à la SOFRES d'élaborer des tableaux croisant par exemple âges et opinions, ou seulement de dénombrer les fréquences des attitudes élémentaires.

Selon nous il convient d'aller bien au-delà : c'est l'ensemble des affinités entre conditions sociales individuelles et attitudes de réponse qu'il faut mettre en évidence par une analyse appropriée de si abondantes données que recueillent les enquêtes. Ce n'est pas le lieu d'exposer même succinctement les principes géométriques de la statistique multidimensionnelle : l'auteur de ces lignes l'a d'ailleurs fait pour le volume Organon de l'Encyclopaedia Universalis. Mais sans remonter aux principes, nous croyons pouvoir ici commenter des graphiques assez suggestifs pour retenir l'attention même si la genèse en est voilée.

Nous choisissons le cas d'un questionnaire clos. L'exploration statistique des réponses libres soit par l'analyse du contenu, soit par simple dénombrement des mots employés est d'un grand intérêt, mais plus complexe et plus éloignée du format usuel des sondages. Nous supposons ici qu'à chaque question il correspond un ensemble fixé de modalités de réponse, dont chaque sujet enquêté doit nécessairement adopter une modalité et une seule. Parfois l'ensemble des réponses doit être adapté à la question particulière posée. Par exemple à la question :

Avez-vous le permis de conduire et l'utilisez-vous ?

on prévoira les 5 modalités :

sans permis ; l'utilise quotidiennement ; l'utilise chaque semaine ; l'utilise occasionnellement ; ne l'a pas utilisé depuis un an.

Mais souvent les ensembles de modalités se répètent identiques : quelle que soit la question le choix est laissé entre {OUI, NON, ABSENTION} avec éventuellement des modalités intermédiaires.

L'exemple que nous proposons est une enquête réalisée par la RATP (Régie Autonome des Transports Parisiens) et analysée (après son

dépouillement conventionnel) par Mesdames F. Pétapermal et M. Rigaud. (Thèses Paris 1980). Les quelque 1800 sujets enquêtés ont été recrutés dans 10 zones de la ville de Saint-Denis (au nord de Paris) et de ses abords immédiats. L'intérêt principal, pour la RATP, était de faire remplir à chaque sujet un carnet consignait tous ses déplacements au cours d'une semaine : du contenu de ces carnets on ne dira rien ici. Mais au préalable, chaque enquêté répondait à un questionnaire comprenant deux parties

*une fiche sociodémographique* : âge, sexe, profession, revenus, loisirs ;

*une batterie d'opinions* : relatives pour la plupart aux transports et à la RATP ; avec face à chaque opinion - par exemple "le métro c'est facile" - la possibilité d'adopter l'une des 5 attitudes suivantes

1 tout à fait d'accord ; 4 pas du tout d'accord ;

2 plutôt d'accord ; 3 plutôt pas d'accord ;

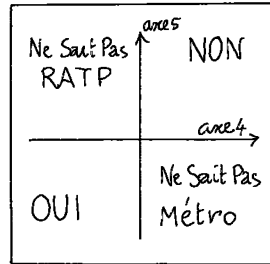
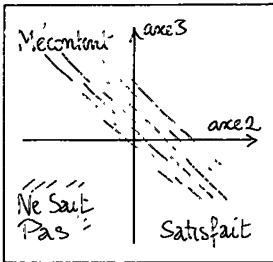
5 ne sait pas ; sans opinion.

C'est de l'analyse de la batterie d'opinion qu'on parlera, en l'illustrant de quelques informations issues de la fiche sociodémographique.

Pour l'analyse statistique multidimensionnelle, la donnée de base est le tableau de nombres : en bref d'un grand tableau illisible même pour un spécialiste exercé qui n'y décèle ni contrastes ni similitudes, l'analyse révèle la structure par des graphiques plans, construits en disposant des points dont les coordonnées successives (ou facteurs) sur des axes, appelés *axes factoriels* ont été calculés à partir du tableau initial. Dans le cas présent d'un questionnaire, ce tableau est le tableau de BURT défini comme suit. Le tableau de BURT a autant de lignes et de colonnes qu'il y a de modalités de réponse sur l'ensemble des questions : donc avec 30 questions admettant chacune 5 modalités de réponse, cela fait un tableau carré à 150 lignes et 150 colonnes ; et à l'intersection d'une ligne telle que "tout à fait d'accord avec la phrase : le métro c'est triste" et d'une colonne comme "sans opinion sur le thème : c'est mieux de voyager dans le métro parce qu'on a le temps de lire" on inscrit le nombre de sujets enquêtés qui ont adopté à la fois ces deux attitudes de réponse. Sans être versé dans la statistique, on conçoit que ce tableau qui est le bilan des associations observées entre toutes les attitudes élémentaires possibles permette de déployer dans une représentation spatiale un réseau complexe d'affinités implicites.

Le fait massif qui vient d'une telle analyse est l'extrême stéréotypie des comportements : ainsi que l'a d'abord remarqué N. Tabard la forme des réponses permises l'emporte sur le contenu des questions posées. Dans le cas présent, on a sur le premier axe une opposition entre les modalités 1 (tout à fait d'accord) et 4 (pas du tout d'accord) d'une part et d'autre part les modalités 2 (plutôt d'accord) 3 (plutôt pas d'accord) et 5 (ne sait pas) ; c'est-à-dire une opposition entre l'engagement catégorique et la réserve.

Ensuite, le graphique du plan 2 x 3 (plan où les coordonnées sont les facteurs 2 et 3) fait apparaître les diverses attitudes globales possibles vis-à-vis de la RATP et des transports en communs. L'amas des abstentions (NSP = 5) s'oppose à une bande qui va de *Mécontent* à *Satisfait*. Sur cette bande, la plupart des opinions voient leurs modalités de réponse rangées dans l'ordre 1, 2, 3, 4 de *Satisfait* à *Mécontent*, s'il s'agit d'une opinion favorable (e.g. "Le métro c'est facile") ou dans l'ordre opposé (4, 3, 2, 1 : de *Satisfait* à *Mécontent*) pour une opinion défavorable (e.g. "Le métro c'est triste").



Dans le plan rapporté aux axes 4 x 5 une interprétation globale apparaît clairement. D'une part les réponses NON (4) s'opposent aux réponses OUI (1) et cela sans distinction de sens pour l'opinion en question (qu'elle manifeste satisfaction ou mécontentement). D'autre part les réponses 5 (Ne Sait Pas) sont partagées en deux classes : il faut en effet préciser que la batterie d'opinions proposée aux enquêtés concerne soit les moyens de transport eux-mêmes (Métro, voire Autobus, jugés tristes, laids, faciles, confortables, etc.) soit l'institution même de la RATP (e.g. "La RATP aujourd'hui c'est avant tout des ingénieurs et des techniciens") : on conçoit que s'abstenir de juger la Régie (RATP) ne relève pas de la même attitude que de s'abstenir de juger les transports.

Il est hasardeux de proposer au lecteur une interprétation générale telle quelle sans preuve donc sans nuance ; et cela d'autant plus que les exceptions ne sont pas moins instructives que la règle... Sur le graphique du plan 2 x 3 publié ici on constatera d'abord que les groupements de points 5 et les chapelets 1, 2, 3, 4 ou 4, 3, 2, 1 ont bien la place annoncée. Mais expliquons au moins les sigles de quelques questions qui offrent des chapelets parfaits.

MT = "Le métro c'est triste" ; donc MT1 est l'approbation franche de cette opinion ; MT2 ou approbation mitigée ; MT3 la négation mitigée ; MT4 la négation catégorique ; MT5 l'abstention ;

ML = "Le métro c'est laid" ; (ML1 = Oui etc.) ;

MF = "Le métro c'est facile" ;

RES = "Quelque chose a changé à la RATP : on sait qu'il y a maintenant des hommes responsables".

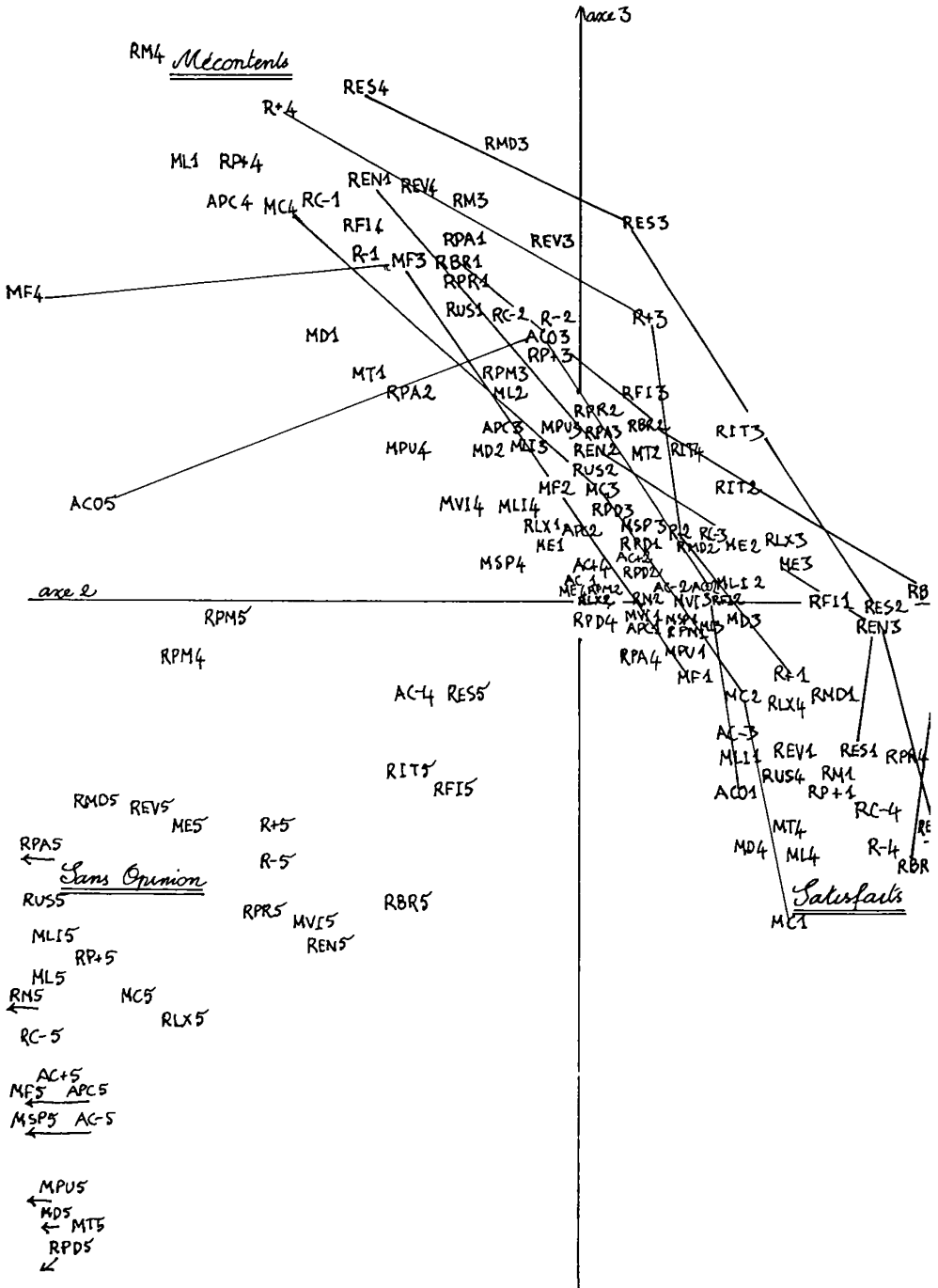
Quant aux exceptions, nous relevons la question ACO

ACO = "La RATP pense aux usagers en améliorant le confort" ;

opinion favorable à la RATP ; et à laquelle aucun sujet n'a pu répondre "non" ; par le fait, la réponse la plus boudeuse est devenue le 5 (N.S.P.) qui se place avec les opinions défavorables dans le plan 2 x 3 (ACO5 est dans le même quart de plan que MF4 : le métro facile : NON ; ML1 : le métro laid OUI ; etc.).

Montrons maintenant l'intérêt d'adjoindre aux opinions les modalités des variables sociodémographiques.

En suivant dans les plans 2 x 3 et 4 x 5 les modalités de la variable AGE, on va de AG1 (jeune, mécontent ; s'exprime par le NON) à AG6



(âgé ; satisfait s'exprime par le OUI). Parmi les mécontents on trouve ceux qui regardent peu la télévision mais vont souvent ou très souvent au cinéma (cf. AG1 !). Les sujets possédant une voiture particulière qu'ils utilisent en semaine se déclarant mécontents, mais s'abstiennent sur les opinions relatives au métro : cela est logique. Il y a quant aux opinions émises de grandes différences entre les zones de résidence : ce qui s'explique par un inégal accès à la gare SNCF, au métro, aux arrêts d'autobus. Ainsi au-delà des facteurs généraux 1 à 5 que nous avons présentés, on trouve sur l'axe 6 d'un côté de l'origine des gens âgés logés en pavillons qui ne trouvent le métro ni triste, ni laid ni désagréable... ; à l'opposé sont les personnes actives qui travaillent à Paris, lesquelles ont du métro une expérience quotidienne qui peut être pénible. Au contraire ceux qui travaillent dans la région parisienne (et donc ne se rendent pas au travail en métro) se placent sur l'axe 7 comme exprimant relativement à l'agrément de ce mode de transport des opinions peu catégoriques (modalités 2, 3) mais favorables.

Bien compris le questionnaire devient une forme d'humour.

Au lendemain d'une consultation électorale, les chroniqueurs se lamentaient de voir la pays coupé en deux : et d'évoquer ces hommes et ces femmes, engagés dans une joute d'autant plus passionnée que l'issue en était jusqu'à la fin demeurée incertaine, et qui se retrouveraient au lendemain les dents serrées sur les lèvres épanouies, pour accomplir ensemble les mêmes tâches quotidiennes.

Nous nous garderons de faire choeur à ces lamentations ! Les alternances de scrutins qui ébranlent l'assurance des gouvernements sans abolir leur pouvoir et des scrutins de confirmation, qui leur accordent une couronne d'une demi-pointure en dessous de leur tour de tête sont trop à l'image de ce que la nation dans son ensemble pense et professe, pour n'être pas l'expression collective de ce qu'au niveau individuel on appelle une opinion nuancée. L'étonnant est seulement que des électeurs qui ne peuvent se compter mesurent si justement leurs gestes pour réussir ensemble le dosage souhaité.

Comment réussissent-ils ? Réussissent-ils toujours ? Les propos échangés en passant, d'un ton de fausse indifférence, tissent comme une trame sensible, où chacun capte l'âme de tous... Cependant la télévision dont Alain Lancelot remarque opportunément qu'elle est une institution contemporaine des sondages, propose solennellement aux sens la méditation ignacienne des deux étendards. Une semaine avant l'événement, il est de rigueur qu'on voile les baromètres : ainsi le prescrit la loi n° 77-808 (du 19 juillet 1977) dont Jean Stoezel a exposé la teneur. Et la Voix parle !

Certes depuis près d'un siècle que Gustave Le Bon a écrit sa *Psychologie des foules*, les voies et centres nerveux du corps social ont changé... On pardonnera à un statisticien élevé sous les pratiques de la géométrie, de prétendre ajouter à l'inconscience collective les phantasmes de ses graphiques.

#### Bibliographie

L'exemple d'analyse de questionnaire présentée ici est publié par F. Pétapermal et M. Rigaud : Sens et valeur des réponses à un questionnaire clos ; in *Pratique de l'Analyse des Données* ; 3 : *Linguistique et Lexicologie*. DUNOD 1981.

Sur l'analyse des données : la référence élémentaire à partir de laquelle on trouvera toutes les autres est :

J.P. et F. Benzécri : *Pratique de l'Analyse des Données* ; 1 *Analyse des correspondances, exposé élémentaire* - DUNOD 1980.

Une revue est consacrée à ces questions ; c'est :

*Les Cahiers de l'Analyse des Données* ; DUNOD éditeur (à partir de 1976).

Les travaux de L. Lebart N. Tabard et collab. paraissent régulièrement dans *Consommation* (Annales du CREDOC) ; DUNOD éditeur.