

# Statistique et société

octobre 2013

Volume 1, Numéro 2





# Sommaire

## Statistique et Société

Volume 1, Numéro 2

---

**7 Édito**  
Emmanuel Didier  
Rédacteur en chef

### DOSSIER : Sondages électoraux

---

**9 Présentation du dossier : Sondages électoraux, comment faire pour qu'ils servent la démocratie ?**  
Par la rédaction

**11 Droit international et comparé des sondages électoraux : avantages et inconvénients du modèle français**  
Romain Rambaud  
Maître de conférences en droit public  
à l'Université de Bourgogne

**17 Élection présidentielle : la régulation paisible de la Commission des sondages**  
Mattias Guyomar  
Secrétaire général de la Commission des sondages

**23 Pour une plus grande transparence sur la méthodologie des sondages électoraux**  
Jeanne Fine  
Statisticienne, Toulouse

**29 Sondages d'intentions de vote: de la transparence à la confiance**  
Jean Chiche  
Centre d'études de la vie politique française

**33 Un sondage en pleine lumière**  
Bruno Jeanbart  
OpinionWay

**39 Hommage à Jacques Antoine (19/08/1928 - 25/12/2012)**  
Daniel Bachelet  
Statisticien

# Sommaire

## Statistique et Société

Volume 1, Numéro 2

---

**43 Nature et déterminants de l'erreur d'échantillonnage dans les enquêtes par sondage**

Pascal Ardilly  
Insee

**51 Climat et statistiques en débat**

Liliane Bel  
AgroParisTech





## Statistique et société

---

Magazine trimestriel publié par la Société Française de Statistique.

Le but de Statistique et société est de montrer d'une manière attrayante et qui invite à la réflexion l'utilisation pratique de la statistique dans tous les domaines de la vie, et de montrer comment l'usage de la statistique intervient dans la société pour y jouer un rôle souvent inaperçu de transformation, et est en retour influencé par elle. Un autre dessein de Statistique et société est d'informer ses lecteurs avec un souci pédagogique à propos d'applications innovantes, de développements théoriques importants, de problèmes actuels affectant les statisticiens, et d'évolutions dans les rôles joués par les statisticiens et l'usage de statistiques dans la vie de la société.

### Rédaction

Rédacteur en chef : **Emmanuel Didier**, CNRS, France

Rédacteurs en chef adjoints :

**Jean-Jacques Droesbeke**, Université Libre de Bruxelles, Belgique

**François Husson**, Agrocampus Ouest, France

**Jean-François Royer**, SFdS - groupe Statistique et enjeux publics, France

**Jean-Christophe Thalabard**, Université Paris 5, France

### Comité éditorial

Représentants des groupes spécialisés de la SFdS :

**Ahmadou Alioum**, groupe Biopharmacie et santé

**Christophe Biernacki**, groupe Data mining et apprentissage

**Alain Godinot**, groupe Statistique et enjeux publics

**Delphine Grancher**, groupe Environnement

**Marthe-Aline Jutand**, groupe Enseignement

**Elisabeth Morand**, groupe Enquêtes

**Alberto Pasanisi**, groupe Industrie

Autres membres :

**Jean Pierre Beaud**, Département de Science politique, UQAM, Canada

**Corine Eyraud**, Département de sociologie, Université d'Aix en Provence, France

**Michael Greenacre**, Department of Economics and Business, Pompeu Fabra  
Université de Barcelone, Espagne

**François Heinderyckx**, Département des sciences de l'information, Université  
Libre de Bruxelles, Belgique

**Dirk Jacobs**, Département de sociologie, Université Libre de Bruxelles, Belgique

**Gaël de Peretti**, INSEE, France

**Theodore Porter**, Département d'histoire, UCLA, États-Unis

**Carla Saglietti**, INSEE, France

**Patrick Simon**, INED, France

Design graphique

fastboil.net

ISSN 2269-0271



# Éditorial

Emmanuel Didier

Rédacteur en chef

---

Chères lectrices, chers lecteurs,

La première livraison de *Statistique et Société* a été accueillie avec beaucoup de chaleur. Nombreux sont ceux qui ont exprimé leur intérêt pour le dossier sur les risques OGM, qui présentait des opinions différentes et enrichissantes. Ils se sont aussi réjouis de la variété des thèmes traités dans le numéro. La statistique mène à tout, et de façon intéressante.

Dans cette deuxième livraison, le dossier porte sur les sondages pré-électorales. Cet outil est celui des méthodes statistiques le plus évidemment lié à la démocratie et, en même temps, celui qui suscite le plus de controverses. Depuis l'élection du Président de la République au suffrage universel direct en 1965, il n'est plus d'élection sans sondage, ni d'élection sans controverse sur les sondages. En un sens, nous préparons donc ici les élections municipales qui se profilent....

Ce numéro se termine avec deux articles inscrits dans des rubriques vouées à se pérenniser : un sur les méthodes, un autre sur l'actualité statistique.

Bonne lecture ! N'hésitez pas à nous faire part de vos réactions, et rendez-vous au prochain numéro dont le dossier portera sur les spécificités de la statistique européenne.

Emmanuel Didier



# Sondages électoraux : comment faire pour qu'ils servent la démocratie ?

Par la rédaction

---

En démocratie, les électeurs doivent pouvoir exprimer leur vote librement. Depuis que des enquêtes par sondage sont utilisées pour connaître les intentions des électeurs, c'est-à-dire en France essentiellement depuis 1965, leur existence a soulevé des craintes. S'il est légitime que les candidats et les commentateurs cherchent à influencer le corps électoral, ils ne doivent pas le faire en diffusant des informations fausses, présentées comme des observations scientifiques. Pour que le scrutin reste sincère, il faut donc que l'objectivité des mesures d'intentions de vote puisse être garantie. Statisticiens et politologues, qui concourent à la conception, la réalisation et la publication des enquêtes pré-électorales au sein des instituts de sondage, sont interpellés.

Tous les pays démocratiques sont confrontés à ce problème : les réponses institutionnelles diffèrent selon l'histoire et la culture des différents pays. Romain Rambaud en donne un aperçu dans son article « Droit international et comparé des sondages électoraux ». Il montre les spécificités de la voie française, et décline ce qu'en sont, selon lui, les avantages et les inconvénients.

Cette voie française met en jeu la loi, plus que la régulation professionnelle chère aux pays anglo-saxons. La loi du 19 juillet 1977 porte sur la publication des sondages préélectoraux : elle impose la publication de quelques informations relatives à chaque sondage, notamment son commanditaire et le nombre des personnes interrogées ; et elle instaure une autorité administrative indépendante, la Commission des sondages, chargée de veiller au respect de ces dispositions et « de proposer des règles tendant à assurer dans le domaine de la prévision électorale l'objectivité et la qualité des sondages publiés ou diffusés » (article 5).

Depuis 36 ans, la régulation des sondages est exercée en France sous l'égide de cette loi. Cette période a été scandée par six élections présidentielles : chacune d'elles a vu le nombre des sondages pré-électoraux s'accroître, pour atteindre 409 en 2012. Désormais ces sondages sont omniprésents pendant toutes les campagnes électorales. Pour autant, les insatisfactions et les doutes n'ont pas disparu : ils ont été avivés par l'écart entre les résultats des enquêtes et le résultat du scrutin lors du premier tour de l'élection présidentielle de 2002.

Même si cette situation ne s'est pas reproduite depuis, une méfiance persiste à l'égard des enquêtes pré-électorales. Cette méfiance est vivace chez certains statisticiens, qui craignent que les méthodes des instituts de sondage fassent place à des modifications discrétionnaires des résultats. De cette méfiance on trouvera le témoignage dans ce dossier sous la plume de Jeanne Fine, dans son article « Pour une plus grande transparence sur la méthodologie des sondages électoraux ». Ce sentiment a trouvé sa traduction politique dans une proposition

de loi adoptée par le Sénat en février 2011, par laquelle les obligations légales des instituts de sondage en matière de publication se trouvaient considérablement étendues. Cette proposition de loi a été fortement restreinte par la Commission des Lois de l'Assemblée Nationale en juin 2011, et n'a finalement pas été soumise au vote de cette assemblée.

Si l'évolution législative n'a donc pas jusqu'à présent abouti, l'évolution réelle a été considérable au cours des dix dernières années. Le bilan d'activité pour 2011-2012 de la Commission des sondages, dressé par son secrétaire général Mattias Guyomar sous le titre « Élection présidentielle : la régulation paisible de la Commission des sondages » montre à quel point les méthodes des instituts ont évolué dans le sens souhaité par les statisticiens. La désinvolture dans la manipulation des résultats, qui prévalait encore dans certains instituts au début des années 2000, ne pourrait plus « passer » aujourd'hui, et un plus grand professionnalisme statistique est désormais de rigueur. Pour le secrétaire général de la Commission, ces progrès sont obtenus grâce à la coopération que le cadre juridique actuel permet entre les instituts et la Commission, aidée par ses experts-statisticiens, sans nuire aux libertés fondamentales d'expression politique dans ce pays.

Reste que la « transparence » n'est alors entière qu'au bénéfice des membres de cette Commission. Beaucoup pensent qu'il faudrait aller plus loin en direction du public. Jean Chiche dans son article « Sondages d'intentions de vote : de la transparence à la confiance » crédite l'ensemble des instituts de sondage de bons résultats en 2012, mais appelle à traiter les enquêtes en question selon les règles en usage dans le monde scientifique. Bruno Jeanbart, responsable d'un institut de sondage, estime excessives les craintes que ressentent la majorité de ses confrères vis-à-vis de la publication de résultats bruts des enquêtes : dans son article « Un sondage en pleine lumière », il relate une expérience qu'il a lui-même réalisée en ce sens. Pour ces deux auteurs, progresser vers une transparence accrue est nécessaire pour augmenter la confiance dans l'outil.

Alors, faut-il renforcer la loi ? Le débat est justifié, les « pour » et les « contre » ont de bons arguments à avancer : ce dossier le montre amplement. Le lecteur formera son opinion. Statistique et Société a souhaité pouvoir associer à ce débat la figure de Jacques Antoine, disparu fin 2012. Comme le rappelle Daniel Bachelet dans son « Hommage », Jacques Antoine, qui fut statisticien public puis dirigeant d'un institut de sondages, a épaulé en tant qu'expert la Commission des sondages pendant plus de 25 ans. Les progrès encore à venir de la régulation des sondages auront besoin de compétences et de dévouements comme les siens.

# Droit international et comparé des sondages électoraux : avantages et inconvénients du modèle français

Romain Rambaud

Maître de conférences en droit public à l'Université de Bourgogne, CREDESPO<sup>1</sup>

---

**La forme de l'encadrement juridique des sondages électoraux** est propre à chaque pays. La France se spécifie par son modèle administratif : le droit des sondages y est mis en œuvre par une autorité administrative indépendante, la Commission des sondages. Si ce modèle présente des avantages indéniables, notamment du point de vue de la qualité des sondages publiés, il souffre aujourd'hui de difficultés à s'adapter aux nouveaux impératifs de la démocratie : transparence, pédagogie et efficacité.

Si, aujourd'hui, le *droit des sondages* n'est pas encore une discipline mature du droit, une tendance émerge, partout dans le monde mais à des degrés très divers pour renforcer l'encadrement des sondages, notamment lorsqu'ils sont de nature électorale. Le système français de régulation des sondages présente à ce titre un visage tout à fait particulier, peut-être unique au monde. La comparaison avec d'autres systèmes permet de tirer des enseignements qui devraient être pris en compte dans le cadre du débat actuel sur l'éventuelle réforme de la loi n° 77-808 du 19 juillet 1977 *relative à la publication et à la diffusion de certains sondages d'opinion*, qui régit en France la matière.

Au niveau international, la réglementation des sondages est assurée par des instances d'autorégulation. La plus importante est l'ESOMAR (*European society for Opinion and Marketing Research*, à sa création transformée en *World Association for Social, Opinion and Market Research* en conservant le même sigle), créée en 1948, qui a pour objet de défendre les intérêts de la profession et qui regroupe 4900 membres dans 130 pays. Toutes les entreprises adhérant à l'ESOMAR doivent adhérer à ses codes, et notamment le code international ICC/ESOMAR de 1977 mis à jour en décembre 2007 [4]. Quant à la WAPOR (*World association of public opinion research*), créée en 1947, il s'agit d'une organisation tournée vers la recherche qui participe à l'édiction des principes communs. Il existe un guide ESOMAR/WAPOR, dédié spécifiquement aux sondages d'opinion et aux sondages électoraux, mis à jour en 2009.

Dans les pays anglo-saxons, l'autorégulation des instituts de sondage domine. Elle est assurée, aux États-Unis, par deux institutions : d'une part l'AAPOR (*American association of public opinion research*), créée en 1947, institution se situant entre une organisation professionnelle et une association de recherche, qui produit des standards ; d'autre part le NCPP (*National Council on published polls*), créé en 1969, qui est un conseil d'autorégulation composé de 27 instituts de sondages. Au Royaume-Uni, l'autorégulation est prise en charge par le BPC (*British Polling Council*), association composée de 15 instituts de sondage et dirigée par un Professeur de science politique.

En Europe, les systèmes sont très diversifiés, et le droit européen ne pèse guère. En Allemagne, la loi ne régit pas les sondages à l'exception de l'interdiction de publier des sondages de

sortie des urnes le jour du scrutin. Pour le reste, le droit des sondages est pris en charge par un système d'autorégulation mal adapté aux sondages politiques, réalisé par le DRDMS (*Der Rat der Deutschen Markt- und Sozialforschung*), créé en 2001. En Italie, au contraire, le droit des sondages est mis en œuvre selon un modèle de régulation publique non spécialisée<sup>2</sup> exercée par l'AGCOM<sup>3</sup>, équivalent italien du Conseil Supérieur de l'Audiovisuel français, en vertu de l'article 8 de la loi du 22 février 2000 relative au pluralisme politique et aux campagnes électorales.

La France se distingue parce qu'elle dispose d'une régulation publique spécialisée prise en charge par la Commission des sondages en vertu de la loi du 19 juillet 1977. Son modèle présente un certain nombre d'avantages, mais aussi d'inconvénients. Le lecteur trouvera ci-dessous page 16 un tableau de comparaison entre le modèle français et différents modèles étrangers, au point de vue de ces caractéristiques.

Par prudence, il faut garder à l'esprit les limites de la comparaison : le droit des sondages de chaque pays s'est construit en réponse à ses problèmes propres et en fonction du contexte institutionnel et politique. Il n'y a donc pas de modèle idéal de droit des sondages. Toutefois, ces limites n'empêchent pas de reconnaître les points forts du modèle français, ni de pouvoir s'inspirer des autres modèles pour améliorer le nôtre sur ses points faibles.

## Les avantages comparatifs du modèle français

On en compte au moins trois :

### Une structure indépendante et spécialisée

Le modèle français est particulier d'abord parce qu'il est pris en charge par une autorité administrative indépendante, ce qui lui donne de nombreux atouts. D'une part, la régulation est meilleure que l'autorégulation en termes d'indépendance, notamment vis-à-vis des instituts de sondage, même si le risque de « capture »<sup>4</sup> n'est pas inexistant. D'autre part, la régulation spécialisée est meilleure que la régulation non spécialisée du point de vue du contrôle. On le voit notamment en Italie avec l'AGCOM, où la non spécialisation conduit à un contrôle du respect de la transparence mais à un faible contrôle des méthodes des instituts (contrôle « au fond »), ou en Allemagne, où la régulation est celle des études de marché en général et n'est pas très approfondie pour les sondages électoraux.

### Un contrôle effectif des sondages

C'est selon nous la France qui présente le meilleur système du point de vue du contrôle au fond, car il existe, à la fois, un contrôle de la rigueur des méthodologies et, sur les points les plus sensibles comme les critères de redressement, un contrôle restreint à l'erreur manifeste d'appréciation et à la cohérence méthodologique. Ce système permet un juste équilibre entre le nécessaire contrôle et la liberté d'appréciation donnée aux instituts. Il se rapproche, dans une certaine mesure, du code ESOMAR/WAPOR, lequel pose de nombreux principes de fond, tout en reconnaissant la liberté des instituts.

Cela évite à la France de connaître les défauts des systèmes dans lesquels il n'existe guère de contrôle au fond, comme les systèmes anglo-saxons et dans une moindre mesure italien.

Notamment, aux États-Unis et au Royaume-Uni, les instances d'autorégulation refusent de procéder à un contrôle des mérites respectifs des méthodes des instituts pour se contenter d'un contrôle des modalités de publication, ce qui peut poser des problèmes de fiabilité. Les dernières élections présidentielles nous en ont fourni des exemples. Certains sites internet américains pro-républicains n'ont pas hésité à publier leurs sondages donnant Mitt Romney

vainqueur contre toute évidence<sup>5</sup>. La complexité du système américain a conduit un statisticien, Nate Silver, à élaborer un modèle statistique synthétisant tous les sondages, modèle qui comprend un mécanisme de redressement politique des instituts de sondages eux-mêmes, c'est-à-dire selon que ceux-ci surestiment systématiquement la droite ou la gauche<sup>6</sup>. Le modèle anglo-saxon se caractérise donc par un très grand libéralisme. Il est évident qu'un tel libéralisme serait sans doute culturellement inenvisageable en France.

## Une durée d'interdiction de publication pertinente

Enfin, le droit comparé démontre l'utilité de mettre en place une interdiction courte des sondages avant l'élection et de ne jamais diffuser les estimations de résultats avant la fermeture des bureaux de vote.

Tous les systèmes interdisent au moins la publication de sondages de sortie des urnes le jour du vote : ESOMAR/WAPOR, la recommandation du 7 novembre 2007 du Conseil de l'Europe, le droit allemand, et même les États-Unis où a été mise en place en 2013 une chambre de quarantaine.

Le droit français, qui prévoit une interdiction de publier des sondages la veille et le jour des scrutins jusqu'à la fermeture du dernier bureau de vote, apparaît donc conforme aux principes reconnus comme solides au plan international.

A contrario, la comparaison démontre le peu de pertinence des systèmes d'interdiction trop longue, comme cela existe en Italie ou Grèce, avec un délai de 15 jours qui a été largement méconnu, notamment sur internet. Deux requêtes sont d'ailleurs actuellement pendantes devant la CEDH<sup>7</sup> s'agissant de la conventionnalité du délai grec d'interdiction de 15 jours.

Le droit français a donc de nombreux avantages. Toutefois, il est affecté d'inconvénients qu'il est aujourd'hui impératif de corriger.

## Les inconvénients comparatifs du modèle français

On en compte également trois :

### Un système peu transparent

En France, la transparence est assez faible et cela est contesté. Certes, les instituts ont fait des efforts importants, mais ils ne publient toujours pas de notice comportant les résultats bruts et/ou les redressements. Par ailleurs, la Commission des sondages maintient un système de double notice, c'est-à-dire que les notices publiques, communiquées aux individus sur demande, sont différentes des notices données à la Commission par les instituts et sont très incomplètes. Cette situation n'est pas acceptable : seule l'Allemagne paraît aussi peu ouverte à la transparence.

Le code ESOMAR/WAPOR prévoit en effet que l'institut de sondages devrait mettre ses résultats bruts à disposition quand les résultats rendus publics diffèrent substantiellement des données issues du terrain, notamment lorsque sont appliqués des redressements non standards. Néanmoins, ce code prévoit également que les instituts sont légitimes à conserver le secret de leur méthode de redressement pour des raisons concurrentielles. C'est donc une transparence limitée.

Le modèle anglo-saxon est aussi transparent et il est même entièrement articulé autour de la logique de la transparence. Aux États-Unis, le système est censé donner aux tiers la possibilité de comprendre eux-mêmes la manière dont les sondages sont réalisés et les instituts doivent donner aux citoyens qui le demandent des informations complètes. Au Royaume-Uni, le

système est très transparent notamment du point de vue des notices. Ainsi les principes du BPC prévoient que l'institut de sondage doit publier sur son site internet l'ordre des questions posées, des tables avec les bases redressées et non redressées, une description des procédures de redressement, etc. Et effectivement, les notices sont très complètes, on trouve les bases brutes et redressées ainsi que les critères de redressement... mais on ne dispose toutefois pas des pourcentages associés aux chiffres bruts.

En Italie, le système est très transparent et d'une manière tout à fait originale. En effet, la loi prévoit que les notices des sondages sont publiées sur un site internet de la présidence du Conseil sur lequel on trouve l'ensemble des sondages publiés sur une période récente<sup>8</sup>. On trouve aussi de très nombreuses informations, bien au-delà de celles que l'on trouve habituellement : des méthodes détaillées, mais aussi le commanditaire, la marge d'erreur, la pondération, le taux de non-répondants, de déchets et de remplacements, etc. C'est donc un système très transparent même si l'on peut nourrir des doutes du fait que le site en question est un site de la présidence du Conseil... et donc un site de l'exécutif ! En revanche, cette transparence n'est pas totale puisqu'elle trouve sa limite dans le secret des affaires des instituts, comme l'a reconnu l'AGCOM. C'est ainsi que, lorsqu'un sondage est réalisé avec redressement politique en Italie, les données détaillées, par exemple les résultats bruts et les méthodes de redressement, ne sont pas fournies.

### Un manque d'accompagnement médiatique

En France, il n'existe aucun dispositif pour accompagner la publication du sondage. La comparaison montre que le droit français est en retrait par rapport aux autres systèmes et qu'il existe des solutions à mettre en place.

Ainsi, le code ESOMAR/WAPOR s'inscrit clairement dans une logique d'accompagnement médiatique. Il propose la contractualisation en amont des modalités de publication. Il va jusqu'à dire que l'institut doit approuver en avance la forme et le contenu exact de l'article qui reprend le sondage, et que cela doit être prévu par le contrat. Également, l'institut peut intervenir a posteriori pour apporter des précisions quand une publication trop courte altère l'analyse des résultats, quand la publication n'est pas suffisante ou quand la publication méconnaît les stipulations contractuelles.

Le modèle anglo-saxon est également focalisé sur l'accompagnement de la publication. Le NCPP a ainsi récemment créé un Polling Review Board (PRB), afin de pouvoir répondre aux questions des journalistes sur les sondages. Il a par ailleurs publié un guide à l'intention des journalistes, calqué en fait sur celui de l'ESOMAR. La logique du BPC est la même : il est là pour répondre aux journalistes et il a également édicté un guide à leur destination.

### Un système de sanction inadapté

Enfin, la comparaison démontre qu'en ce qui concerne le dispositif de sanctions, la France est dans une situation très particulière. Elle a peut-être le plus mauvais des systèmes, car c'est le seul qui fait intervenir le juge pénal, alors qu'une telle intervention pose de véritables problèmes en matière de liberté d'expression et qu'elle est largement ineffective.

Deux systèmes de sanction sont performants dans tous les modèles : la publication (en France, il s'agit de la mise au point, décidée par la Commission des sondages et qui consiste en la publication d'un rectificatif dans le média concerné) et d'autres sanctions, administratives, c'est-à-dire des amendes quand la régulation est publique comme en Italie, et des avertissements pouvant aller jusqu'à l'exclusion dans les systèmes autorégulés, comme aux États-Unis, au Royaume-Uni et en Allemagne. C'est en ce sens qu'il faudrait aller en France.

## Conclusion

En conclusion, la comparaison indique que le droit français doit être renforcé dans trois directions :

1. Un renforcement de la transparence, qui peut passer par la négociation d'un système spécial d'accès aux documents administratifs comme il en existe d'autres en droit électoral
2. La mise en place de dispositifs, contractuels et non contractuels, d'accompagnement de la publication de sondages
3. La réforme du système de sanction pour qu'il repose moins sur des sanctions pénales et plus sur un régime de sanction administrative

## Références

1. Centre de recherche et d'étude en droit et science politique
2. En cela que l'autorité de régulation ne se consacre pas uniquement aux sondages
3. Autorita per le garanzie nelle Comunicazioni
4. Au sens où une autorité de régulation peut dans certains cas tomber sous la dépendance des organismes qu'elle est censée contrôler
5. Le site [www.UnSkewedPolls.com](http://www.UnSkewedPolls.com) par des repondérations transformait un retard compris entre 3 et 11 points en une avance de 8 points
6. Le site de Nate Silver : [fivethirtyeight.blogs.nytimes.com](http://fivethirtyeight.blogs.nytimes.com)
7. Cour européenne des droits de l'homme
8. Le site : <http://www.sondaggipoliticoelettorali.it/>

[1] RAMBAUD Romain, « Droit comparé des sondages électoraux », Revue politique et parlementaire, Avril/Juin 2013, n° 1067, p. 230.

[2] RAMBAUD Romain, Le droit des sondages électoraux, Paris, 2012, L'Harmattan, 306 p.

[3] CHUNG Robert, The Freedom to Publish Opinion Poll Results A Worldwide Update of 2012, WAPOR.

[4] [http://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ICCESOMAR\\_Code\\_English\\_.pdf](http://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ICCESOMAR_Code_English_.pdf)

	<b>Modèles de régulation</b>				
	<b>International (ESOMAR)</b>	<b>Anglo-Saxon USA - UK</b>	<b>Allemand</b>	<b>Italien</b>	<b>Français</b>
<b>Qui régule ?</b>	Autorégulation professionnelle	Id.	Id.	Autorité publique non spécialisée	Autorité publique spécialisée
<b>Y a-t-il un contrôle au fond ?</b>	Des principes sont énoncés, tout en reconnaissant la liberté des instituts	Non	Non	Faible	Oui
<b>Quelle durée d'interdiction des sondages au voisinage de l'élection ?</b>	Sortie des urnes	Le jour du vote seulement (USA : sortie des urnes seulement)	Sortie des urnes seulement	Quinze jours avant	Deux jours avant
<b>Quelle transparence sur les méthodes ?</b>	Forte, sous réserve du secret des affaires	Id.	Faible	Forte	Faible
<b>Y a-t-il un accompagnement des médias pour publier les sondages ?</b>	Le code recommande que les instituts prévoient dans leurs contrats les modalités de publication	Oui	Faible	Faible	Non
<b>Quel système de sanctions ?</b>	De l'avertissement à l'exclusion professionnelle	Id.	Id.	Publication d'avertissements. Amendes	Publication d'avertissements. Possibilité de sanctions pénales.

**Tableau :** Comparaison de cinq modèles de régulation des sondages

# Élection présidentielle : la régulation paisible de la Commission des sondages

Mattias Guyomar

Conseiller d'État, secrétaire général de la Commission des sondages

---

Les lignes qui suivent dressent le bilan de l'activité de la Commission des sondages lors de la dernière campagne présidentielle. On peut en tirer les conclusions suivantes : la forte hausse de l'activité de la Commission n'a pas nui à l'efficacité de son contrôle. La « régulation paisible » qu'elle est parvenue, au fil des années, à mettre en place a permis d'approfondir le contrôle de la qualité des sondages électoraux et de veiller, dans toute la mesure du possible, au respect des interdictions de diffusion de sondages, la veille et le jour du scrutin.

## Forte hausse de l'activité en 2011-2012

La loi du 19 juillet 1977 donne compétence, dans sa version aujourd'hui en vigueur, à la Commission des sondages pour contrôler d'office tous les sondages qui ont un lien direct ou indirect avec un scrutin dès lors qu'ils ont fait l'objet d'une publication - que celle-ci soit délibérée ou que le sondage ait été rendu public de quelque manière que ce soit (par exemple à l'occasion d'une fuite). S'agissant de la dernière campagne présidentielle, la Commission a contrôlé **409 sondages électoraux**. C'est un volume d'activité qui n'a rien à voir avec le passé : en 2007, il y avait eu un peu moins de 300 sondages et en 2002 seulement 193.

Cette progression s'explique par un certain nombre de facteurs.

Le premier facteur d'explication tient à **l'effet des primaires**. L'organisation de primaires fait en effet remonter plus en amont encore qu'auparavant la demande de sondages relatifs à l'élection présidentielle. 118 sondages ont été réalisés et publiés avant l'organisation des primaires. Ils ont été contrôlés compte tenu du lien qu'ils possédaient avec l'élection présidentielle puisqu'il s'agissait de désigner le candidat socialiste à cette élection. Sur ces 109 sondages, 24 portaient spécifiquement sur les primaires organisées par le parti socialiste. Les autres étaient relatifs au 1er tour du scrutin présidentiel en présentant aux personnes sondées plusieurs hypothèses de candidature socialiste ce qui visait à mesurer le potentiel électoral relatif de chacun des candidats testés. Fin 2011, l'effet « primaires » a donc joué à plein.

Le second facteur sur lequel il convient d'insister tient à la **publication de sondages quotidiens**. Ipsos avait eu recours en 2007 à ce qu'il est convenu d'appeler la technique du « *rolling* ». En 2012 c'est l'Ifop qui a pris l'initiative de cette vague quotidienne. Cette technique consiste à interroger un échantillon de personnes dont un tiers seulement est renouvelé chaque jour. Les résultats se fondent sur l'ensemble de cet échantillon dont un tiers est renouvelé quotidiennement et de manière glissante, ce qui est assez intéressant en termes de lissage des évolutions. Mais cela fait un sondage par jour, ce qui contribue à augmenter l'offre globale soumise à notre contrôle.

Dans ce contexte, la mission de la Commission n'a pas changé. Il lui incombe de veiller, dans

son champ d'intervention, à la sincérité du scrutin. La Commission y puise sa légitimité pour intervenir, en qualité de régulateur, dans un domaine régi par de grandes libertés. Côté instituts de sondage, la liberté de commerce et d'industrie est en jeu ; côté organes de presse qui les publient, c'est la liberté d'expression et de communication. La Commission ne peut intervenir, et, le cas échéant, faire ingérence dans l'exercice de ces libertés que dans la mesure où le respect de la sincérité du scrutin l'exige. En bref, sa mission consiste à éviter la diffusion, pendant la campagne électorale, de sondages de mauvaise qualité ou incohérents, voire délibérément biaisés, qui pourraient indûment influencer les électeurs.

## Un contrôle économe et efficace

Un deuxième constat s'impose : cette année encore, le cadre juridique de l'intervention de la Commission et les moyens limités dont elle dispose n'ont pas entravé **l'efficacité de son contrôle**.

La Commission des sondages est une autorité administrative indépendante qui comporte très peu de moyens. Ses onze membres, y compris sa présidente, exercent leurs fonctions à titre accessoire, de même que le secrétaire général. La seule personne qui est affectée à temps plein à la Commission est son secrétaire permanent, cadre A de la fonction publique. La Commission a par ailleurs recours à des statisticiens, actuellement au nombre de trois, qui lui sont attachés et fournissent des vacations en tant que de besoin. Cette très modeste structure correspond à la mission qui est confiée à cet organe de contrôle, mission qui est variable dans le temps : elle travaille à temps plein en période électorale, beaucoup moins en temps normal. Une organisation plus importante serait une source inutile de dépenses publiques.

Les instruments juridiques dont la Commission est dotée sont également modestes puisque elle possède uniquement la possibilité d'ordonner, à titre de sanction, une « mise au point » (art. 9 de la loi) qui consiste à imposer à l'organe qui a publié le sondage de publier le correctif qu'il a appelé de sa part. Pour limitée qu'elle soit, cette possibilité est suffisante pour asseoir l'autorité dans le domaine qu'elle régule.

La dématérialisation des procédures, qu'il s'agisse des relations avec les instituts ou du travail interne de la Commission, a permis que le contrôle s'effectue dans des délais très brefs compatibles avec une intervention publique de la Commission en temps utile. La Commission contrôle, avec l'aide précieuse de ses experts statisticiens, les sondages et leur confection au vu d'une notice qui comporte des éléments confidentiels. On y trouve tout ce qui se rattache aux méthodes d'échantillonnage, à l'administration de l'enquête et aux redressements qui sont appliqués aux résultats bruts qui ressortent de l'enquête. Ces informations ne sont pas publiques afin de protéger le savoir-faire des instituts de sondage qui relève, dans une certaine mesure, du secret industriel. Au vu de cette masse d'informations que livrent les instituts, avec une transparence qui mérite d'être soulignée, la Commission vérifie que le sondage répond à une cohérence méthodologique et que les résultats établis sont fiables par rapport à l'enquête qui a été menée.

**Le contrôle a posteriori des sondages rendus publics**, qui apparaît le mieux à même de concilier l'intervention de l'organe de régulation avec le nécessaire respect des libertés d'expression, de communication et du commerce et de l'industrie, s'avère donc suffisamment efficace pour qu'une modification des compétences de la Commission ne s'impose pas.

En sus de ce contrôle systématique, **7 réclamations** ont été adressées par des particuliers. Celle présentée par M. Mélenchon a fait l'objet d'un rejet qu'il a contesté devant le Conseil d'Etat. Ce recours a été rejeté par une décision du 8 février 2012 qui conforte un certain nombre des positions adoptées par la Commission des sondages. Que ce soit à la suite de son contrôle

systématique ou en réponse à une réclamation, la Commission a prononcé **7 mises au point**.

Ce petit nombre de mises au point, très faible si on le rapporte au nombre total de sondages contrôlés, témoigne d'abord de la qualité globale du travail effectué par les instituts de sondages. Mais il ne doit masquer ni l'ampleur ni l'efficacité des interventions de la Commission qui se déclinent sous des formes bien plus variées que celles prévues par la loi. Cette autorité a en effet tissé au fil du temps avec les instituts de sondage comme avec les organes de presse des relations qui nous permettent d'exercer ce que le rapport d'activité pour 2012 qualifie de « **régulation paisible** ». La Commission organise ainsi, de manière régulière, des réunions de travail avec les différents acteurs du domaine. Elles permettent d'échanger sur les évolutions techniques, les difficultés rencontrées et conduisent la Commission à dégager de grandes lignes de méthodes voire des recommandations de bonne pratique. C'est à la Commission qu'il est revenu d'édicter, de cette manière, un certain nombre de règles qui sont venues opportunément compléter la loi.

Ces « guidelines » sont généralement respectées par les instituts. Lorsqu'elles ne le sont pas, la Commission dispose d'une **riposte graduée** qui peut conduire, indépendamment de toute mise au point, à un échange informel, une lettre d'observations, une convocation en séance plénière. Dans ces conditions, et à la condition que la frontière de l'incohérence méthodologique n'ait pas été franchie, la Commission obtient des instituts qu'ils modifient leur comportement et leurs méthodes dans le sens de ses observations.

## Progrès méthodologiques

Pour terminer, il convient de s'attarder sur trois éléments saillants de l'activité de la Commission pendant la campagne présidentielle 2012.

Le premier concerne le développement de la publication de sondages qui correspondent à l'extraction d'un sondage mené auprès d'un échantillon représentatif de l'ensemble de la population française. Les résultats publiés sont issus de **sous-échantillons**, concernant par exemple les jeunes, les ouvriers ou encore les catholiques. Le problème vient de ce que de tels sondages sont fréquemment présentés, dans l'article qui en fait état, comme issus d'une enquête menée auprès d'un échantillon dont le nombre correspond à l'échantillon global et non à celui du sous-ensemble pertinent. La Commission est intervenue à quatre reprises pour faire corriger la mention relative à la taille de l'échantillon. Il s'agit d'un élément très important pour apprécier la fiabilité des résultats même si les instituts français ne recourent pas à la méthode aléatoire qui permet de calculer des marges d'incertitudes scientifiquement parlant mais à la méthode des quotas. Ce n'est pas la même chose de dire qu'un sondage a été effectué auprès de 800 personnes et de préciser que sur les 800 personnes interrogées, les résultats publiés correspondent en réalité à un sous-échantillon de 90, 105 ou 150 personnes.

Le deuxième élément concerne la « **colonne de référence** ». Les résultats bruts issus de l'enquête de terrain sont affectés d'un plus ou moins grand nombre de redressements. C'est le savoir-faire de chaque institut : redressement sur souvenir de vote (qui permet de corriger les résultats en tenant compte de l'écart qu'il y a entre la réalité des scores obtenus lors des précédents scrutins et les souvenirs de vote que les gens interrogés ont eu de ces mêmes scrutins), redressements en fonction de la certitude d'aller voter ou de la sûreté du choix. On peut arriver à des panachages assez subtils. Lorsqu'on additionne la première colonne qui correspond aux résultats bruts et les colonnes correspondant à chaque redressement effectué, on obtient une somme de colonnes qui créent un intervalle entre le score le plus faible issu d'un redressement et le score le plus élevé : c'est ce que la Commission dénomme la fourchette des redressements intermédiaires. Dans les années 2000, les scores étaient bien trop souvent publiés hors fourchette. L'acquis des précédentes campagnes électorales a été d'obtenir des

instituts de sondage qu'ils ne publient pas de scores qui se situent en dehors de cette fourchette. L'acquis de cette dernière campagne dont la Commission est particulièrement fière est d'avoir obtenu des instituts qu'en principe, les résultats publiés soient issus de la même colonne de référence pour tous les candidats. Le redressement privilégié pour établir le score doit être le même pour tous les candidats testés. Ce n'était pas le cas auparavant, et cela conduisait à des corrections, pour arriver à 100 % des suffrages exprimés, qui pouvaient s'avérer quelque peu artificielles. Cette exigence méthodologique ne constitue pas une règle absolue : on peut concevoir en effet que le score brut d'un candidat qui n'était pas présent ou dont le parti n'était pas présent lors du scrutin de référence soit corrigé en fonction d'une autre élection où son parti était représenté. Mais il faut que cet écart par rapport à la colonne de référence ne soit pas excessif et que la traçabilité de l'établissement des résultats soit fournie par l'institut et justifiée. Sur ce point, l'approfondissement du contrôle méthodologique de la Commission garantit une meilleure cohérence interne des sondages.

## Le problème du dimanche soir

Enfin, troisième élément saillant, doit être soulignée l'importance d'une autre mission confiée à la Commission qui consiste à veiller au respect de **l'interdiction de publier des résultats la veille et le jour du scrutin** (art. 11 de la loi).

Sur ce point, le bilan est contrasté comme le dit le rapport de la Commission nationale de contrôle de l'élection présidentielle avec laquelle la Commission des sondages a étroitement travaillé pour assurer le respect de ces interdictions. Il est très positif en ce qui concerne les interdictions qui portent sur la veille du scrutin. Il n'y a pas eu de sondage qui ait été rendu public de manière massive le samedi.

En ce qui concerne les résultats partiels, c'est-à-dire les fuites des scores, le jour du scrutin avant 20 heures, le bilan est plus mitigé. La Commission avait déjà été confrontée à cette situation en 2007 mais l'explosion des réseaux sociaux depuis lors a modifié l'ampleur du phénomène. De nombreuses personnes ont en effet revendiqué le droit de violer la loi en se drapant derrière la liberté de communication pour publier des résultats dès 18 heures ou 19 heures à partir de bureaux de vote tests. Ces promesses de violation émanaient de personnes très différentes, médias français et étrangers (suisses et belges notamment), blogueurs. Or, l'interdiction posée par la loi conserve, en dépit des progrès techniques, sa raison d'être, à savoir la préservation de l'unité du corps électoral jusqu'au vote du dernier électeur. Elle a pour fondement solide un objectif : éviter les manœuvres susceptibles d'influencer les électeurs qui n'ont pas encore voté et qui peuvent se rendre aux urnes alors qu'ils n'y pensaient pas, ou changer leur vote. Avec la Commission nationale de contrôle, la Commission des sondages a pris des dispositions communes qui relevaient pour beaucoup de la pédagogie (publication de communiqués communs, invitations à la radio et la télévision pour expliquer les raisons des interdictions et avertir des risques encourus, à savoir le prononcé d'une amende allant jusqu'à 75000 euros, conférence de presse à l'avant-veille du second tour).

Un dispositif a été mis en place avec le concours du ministère de l'Intérieur sous le timbre de la Commission nationale de contrôle mais avec le renfort de la Commission des sondages, qui permettait le contrôle sur tout ce qui est accessible publiquement sur Internet (sur ce point, il faut faire la différence entre un tweet et un SMS, entre le message qui a vocation à être accessible et celui qui ne relève que de la communication privée). Il y a eu une surveillance du net et la photographie d'un certain nombre de violations puisqu'effectivement des fuites ont été faites le jour du scrutin. Cette entreprise n'a pas été totalement vaine puisqu'un certain nombre d'organes de presse qui avaient prévu de violer l'embargo ne l'ont pas fait. Le parquet de Paris a toutefois été saisi des infractions commises par l'AFP, et certains médias belges et suisses.

Par ailleurs, la Commission a obtenu des 9 principaux instituts de sondage qu'ils ne réalisent pas de sondage « sorties des urnes ». En effet, les fuites peuvent surgir de deux sources : soit des premiers résultats issus de bureaux tests, soit d'enquêtes qui peuvent être réalisées à la sortie des urnes et qui consistent à interroger l'électeur sortant de l'isoloir et à lui demander pour qui il vient de voter. Les neuf instituts de sondage se sont engagés à ne pas faire d'enquêtes de sortie des urnes ni au premier ni au second tour, ce qui a asséché la source d'éventuelles fuites mais n'a pas empêché que surgissent sur des sites, des blogs, des tweets, des résultats qui se prétendaient le fruit de sondages de sortie des urnes, tout-à-fait faussement. Un système de permanence a conduit à se retourner à chaque fois vers l'institut de sondage crédité de ce sondage de sortie des urnes, et les instituts de sondage ont fait des démentis le dimanche après-midi.

### **Un bilan positif**

Confrontée à un nombre élevé de sondages concernant l'élection présidentielle de 2012, la Commission des sondages en a assuré le contrôle avec efficacité. Pour garantir la sincérité du scrutin, dans la limite des prérogatives que la loi lui confère, sans ingérence excessive dans les libertés fondamentales qui sont en jeu, la Commission a établi avec les instituts de sondages des relations de travail régulières qui lui permettent non seulement de détecter les situations qui méritent son intervention, mais aussi de promouvoir un progrès d'ensemble.

### **Références**

Site Internet de la Commission des sondages : [www.commission-des-sondages.fr](http://www.commission-des-sondages.fr)

Loi du 19 juillet 1977 modifiée : <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000522846>



# Pour une plus grande transparence sur la méthodologie des sondages électoraux

Jeanne Fine

Statisticienne, Toulouse

Si les sondages électoraux pratiqués par les Instituts de sondage sont scientifiquement fondés, il est légitime de demander une plus grande transparence sur la méthodologie adoptée. Les deux points de méthodologie sur lesquels nous souhaitons insister ici sont la « marge d'erreur » et le « redressement d'échantillon ». Pour mieux comprendre les débats les concernant, nous les insérons dans une présentation des écarts entre la théorie des sondages aléatoires et la pratique des sondages électoraux par quotas.

## Une proposition de loi pour une plus grande transparence

A l'initiative d'Hugues Portelli (sénateur UMP) et de Jean-Pierre Sueur (sénateur PS), une Proposition de loi sur les sondages visant à mieux garantir la sincérité du débat politique et électoral, a été adoptée le 14 février 2011 à l'unanimité des sénateurs, ce qui est suffisamment rare pour être souligné (Portelli et Sueur, 2011). Un résumé des dispositions relatives à la transparence est donné en annexe. Il s'agit des derniers paragraphes de la présentation des quinze recommandations du rapport d'information au Sénat présenté en octobre 2010 (Portelli et Sueur, 2010). La Commission des lois de l'Assemblée Nationale a fortement amendé cette proposition de loi, en particulier pour la restreindre aux sondages électoraux. Le rapport de la discussion à l'Assemblée Nationale d'Étienne Blanc (enregistré le 1er juin 2011) est donné en référence (Blanc, 2011). Deux ans plus tard, la proposition de loi ainsi amendée n'a toujours pas été présentée à l'Assemblée !

La législation actuelle sur la publication et diffusion des sondages électoraux date de 1977 et une « commission des sondages », autorité administrative indépendante, a été créée pour contrôler le respect de la loi. On pourra consulter cette loi, modifiée en 2002, sur le site de la commission des sondages donné en référence. Pour davantage d'informations sur le droit des sondages, en France et à l'étranger, on renvoie à l'article de Romain Rambaud dans ce numéro et à son site ; voir en particulier le texte de son intervention à l'IFOP le 2 juillet 2013 (Rambaud, 2013).

Pour résumer les positions, la demande d'une plus grande transparence de la proposition de loi est soutenue par les statisticiens mais combattue par la plupart des Instituts de sondage qui préfèrent une autorégulation à une nouvelle loi ! Pour les statisticiens, si la méthode utilisée est scientifique, la méthodologie doit être clairement publiée et les résultats doivent pouvoir être vérifiés par la communauté scientifique. Si ce n'est pas le cas, il faut tout aussi clairement annoncer qu'il s'agit d'avis de politologues fondés sur des sources diverses et personnelles et que les résultats n'ont pas de fondement statistique.

La proposition de loi votée au Sénat intervenait dans le contexte d'un rapprochement entre

statisticiens et sondeurs. On pourra consulter, sur le site de la Société Française de Statistique (SFdS), la page du groupe « Enquêtes et modèles » concernant les débats méthodologiques des sondages électoraux ; cf. en particulier le compte-rendu du « séminaire sur le bon usage des sondages d'intentions de vote » organisé avec le CEVIPOF en mars 2012 (SFdS, Groupe Enquêtes et Modèles, 2012). Nous n'en sommes plus aux débats houleux qui se sont tenus autour d'une table ronde sur les sondages électoraux lors du « Colloque francophone sur les sondages » en novembre 2007 à Marseille !

La réglementation des sondages électoraux concerne les sondages ayant un rapport direct ou indirect avec une élection réglementée par le code électoral ainsi qu'avec l'élection des représentants au Parlement européen. Notons que ces sondages d'intention de vote, lorsqu'ils sont réalisés quelques jours avant une élection et bien qu'ils se présentent comme un « état de l'opinion », donnent une « prévision » qui pourra être comparée au résultat officiel. Ce n'est pas le cas des sondages de notoriété politique ou des sondages d'opinion dont il n'est jamais possible d'évaluer la qualité !

Suite au vote de la proposition de loi au Sénat, le journal *Le Monde* publie le 9 mars 2011 un article de Roland Cayrol intitulé *Il est inutile de légiférer sur la question des sondages, la suspicion contre les professionnels est injuste* dans lequel le sondeur appelle à une « réflexion adulte sur la publication des sondages ». Dans son édition du 17 mars 2011, *Le Monde* publie un article d'Hugues Portelli et Jean-Pierre Sueur répondant point par point aux arguments de Roland Cayrol et, en réponse à sa proposition d'autorégulation, *Le Monde* propose de publier un sondage d'intentions de vote et d'en faire un décryptage aussi complet que possible, article intitulé *Les secrets de fabrication d'un sondage politique*, réalisé par Ipsos et commenté par Brice Teinturier.

Pour mieux comprendre notre demande d'une plus grande transparence sur la méthodologie des sondages électoraux, en particulier sur la « marge d'erreur » et le « redressement d'échantillon », nous insérons les éléments des débats les concernant dans une présentation des écarts entre la théorie des sondages aléatoires et la pratique des sondages par quotas. Nous nous référons au sondage du *Monde* du 17 mars 2011 pour illustrer notre propos. Pour ceux que les développements plus techniques ne rebutent pas, nous renvoyons à Fine (2007) dans lequel nous décryptons deux articles publiés dans *Le Monde* suite au premier tour des présidentielles de 2002 et, directement en lien avec le sujet qui nous intéresse ici, à Riandey (2012) et à Ardilly (2010). Cette dernière référence est une annexe au rapport d'information sur les sondages au Sénat intitulée *des éléments sur la précision des sondages* et rédigée par un statisticien de l'INSEE spécialiste de la théorie et de la pratique des enquêtes par sondage.

## Méthode aléatoire et méthode des quotas

L'association Pénombre propose réflexions et échanges sur l'usage du nombre dans le débat public, par des articles accessibles sur leur site et par des rencontres que les adhérents organisent. Dans le document préparatoire à la 12<sup>ème</sup> rencontre « Nocturne », du 5 juin 2009, sur le thème : *Les sondages d'opinion, outils de mesure à utiliser avec précaution ?* (Pénombre, 2009), Alain Tripier, praticien des sondages, écrit (p.14) :

La théorie des probabilités nous enseigne que pour connaître les événements qui peuvent survenir dans une population donnée, il est possible de n'interviewer qu'une petite partie de celle-ci, à condition de respecter des règles de sélection rigoureuses garantant sa représentativité.

Un premier choix cornélien se pose alors au praticien entre :

- la méthode aléatoire (ou probabiliste) préférée par les anglo-saxons ;
- la méthode dite des « quotas » quasi hégémonique en France.

Nous sommes en désaccord avec cette présentation car la théorie des probabilités ne traite que de la méthode aléatoire. La méthode des quotas est une méthode empirique qui n'a pas de fondement théorique. Si elle est utilisée en France, alors que la méthode aléatoire est préférée par les anglo-saxons, c'est peut-être parce que la formation en probabilités et statistique a été trop longtemps négligée en France au contraire des pays anglo-saxons !

Précisons auparavant le vocabulaire. Nous n'avons en français qu'un seul mot « sondages » pour désigner les « sondages d'opinion » (*poll* en anglais, *sondeo* en espagnol) et la « théorie des sondages » (*sampling* en anglais, *muestreo* en espagnol). Les sondages d'opinion sont généralement réalisés par la « méthode des quotas » qui consiste à construire un « échantillon représentatif » de la population selon quelques variables de quotas : sexe, tranches d'âge..., c'est-à-dire, un échantillon respectant les « proportions » des différentes catégories des variables de quotas de la population. La « théorie des sondages » repose sur *l'échantillonnage aléatoire* et l'estimation en population finie ; elle a des fondements probabilistes. Le recueil de données peut être réalisé avec ou sans questionnaire, auprès d'une population humaine, d'une population de ménages, d'entreprises, d'objets... Pour le statisticien, tout échantillon aléatoire dûment contrôlé est « représentatif » de la population dans la mesure où il permet d'estimer les paramètres inconnus de la population et d'en estimer la précision. Le *sondage aléatoire simple à probabilités égales* est le plus simple et sert de référence pour mesurer les améliorations apportées par des échantillons aléatoires plus complexes.

## Présentation de la méthode aléatoire

### Sondage aléatoire simple

Dans le cadre d'un sondage d'intention de votes, un sondage aléatoire à probabilités égales de taille 1000 consisterait à établir la *base de sondage*, c'est-à-dire, la liste de la population adulte française inscrite sur les listes électorales, et à utiliser un générateur de nombres aléatoires pour sélectionner 1000 personnes de cette base de sondage. La sélection doit être réalisée *selon une procédure aléatoire dûment contrôlée* qui garantit que tous les échantillons de taille 1000 ont la même probabilité d'être tirés. Il faut arriver à contacter ces 1000 personnes sélectionnées, il faut qu'elles acceptent de répondre au questionnaire, qu'elles comprennent les questions posées, qu'elles aient une opinion sur ces questions, qu'elles répondent de façon sincère.

### Marge d'incertitude

Lorsque toutes ces conditions sont réunies, au niveau de confiance de 95%, c'est-à-dire, avec une probabilité de se tromper de 5%, on peut considérer que la proportion qui nous intéresse sur la population (proportion d'intention de vote pour tel candidat par exemple) peut être estimée à partir de la proportion  $p$  observée sur l'échantillon à 3% près (c'est-à-dire,  $p$  plus ou moins 3%).

C'est à tort que cette marge d'incertitude de 3% est appelée « marge d'erreur ». Il ne s'agit pas d'une erreur de calcul mais de tenir compte de la fluctuation naturelle d'une proportion calculée sur un échantillon de taille 1000 : elle fluctue d'un échantillon de taille 1000 à un autre échantillon de taille 1000. Pour un échantillon de taille  $n$ , la marge d'incertitude (au niveau de confiance de 95%) est égale à l'inverse de la racine carrée de  $n$  (10% pour un échantillon de taille 100, 5% pour un échantillon de taille 400, 3% pour un échantillon de taille 1000 et 1% pour un échantillon de taille 10 000). Cette règle simple, enseignée aujourd'hui en classe de seconde, est utilisée pour des proportions comprises entre 20% et 80%. Il y a une diminution de l'incertitude lorsque les proportions sont plus proches de 0 ou de 1 mais cette marge de 3% est un bon repère pour les résultats du sondage se rapportant à *l'ensemble de la population*.

Lorsque les résultats concernent *une sous-population*, il faut consulter la taille de l'échantillon correspondant afin de réévaluer la marge d'incertitude (5% pour un échantillon de taille 400, 10% pour un échantillon de taille 100).

Il est important de noter que, lorsque le taux de sondage (égal au rapport de la taille de l'échantillon sur la taille de la population) est faible, ce qui est le cas ici, la marge d'incertitude ne dépend que de la taille  $n$  de l'échantillon et non du taux de sondage.

Avec un tel échantillon aléatoire à probabilités égales de taille 1000, on obtient également les proportions d'hommes, de femmes, des différentes tranches d'âge... à 3% près. Il s'agit donc d'un échantillon à peu près « proportionnel » à la population selon toutes les catégories, connues ou non, c'est-à-dire d'un échantillon « représentatif » de la population selon toutes les variables, connues ou non. C'est l'immense efficacité du sondage aléatoire : **le hasard assure une représentativité auquel aucun expert ne peut prétendre en s'affranchissant d'une sélection aléatoire.**

## Redressement

Si l'on connaît par ailleurs avec certitude la répartition de la population selon les tranches d'âge et que ces proportions ne sont pas exactement respectées dans l'échantillon, on peut « redresser l'échantillon » en majorant le poids des individus des tranches sous-représentées au détriment de celles surreprésentées. Ce *redressement* améliore les estimations des proportions d'intention de vote sans pour autant changer fondamentalement la marge d'incertitude. Notons que c'est l'échantillon qui est redressé avec un système de pondérations, ce qui influe évidemment sur les estimations. Mais le redressement n'est justifié théoriquement que si les pondérations s'appuient sur une connaissance « précise » des informations utilisées concernant la population.

## Sondage aléatoire stratifié

Si l'information concernant ces variables (sexe, âge...) est indiquée dans la base de sondage *pour chaque individu de la population*, il est possible de réaliser un *sondage aléatoire stratifié proportionnel* selon ces variables. Cela consiste à tirer un échantillon aléatoire simple dans chaque catégorie avec des tailles d'échantillons proportionnelles aux tailles des catégories dans la population. Là encore, on pourra noter une amélioration des estimations des proportions d'intention de vote sans que les marges d'incertitude soient vraiment réduites.

## Biais

Nous n'avons explicitement parlé jusqu'ici que de l'incertitude due à l'échantillonnage. Plus sérieux est le problème d'estimations « biaisées » dues, par exemple, à une base de sondage inadéquate ou à un grand nombre de « non-réponses » : aussi grande que soit la taille de l'échantillon, les estimations se concentreront autour de valeurs éloignées des proportions de la population de référence que l'on cherche à estimer. Par exemple, un échantillon de personnes répondant à une enquête sur Internet n'est représentatif que de la population d'utilisateurs d'Internet prenant le temps de répondre à ce type d'enquête ! Le problème de non-réponses fait l'objet de nombreuses recherches en théorie statistique des sondages aléatoires afin d'en atténuer les conséquences négatives.

Ce problème de non-réponses ne dépend pas de l'échantillonnage, il concerne aussi les enquêtes exhaustives. Prenons un exemple fictif : dans un lycée, 40% des élèves de terminales ont déjà consommé de la marijuana mais, lors d'une enquête à ce sujet, les non-consommateurs répondent bien ne pas en avoir consommé mais seulement 10% des consommateurs répondent

en avoir consommé, les autres refusant de répondre. La proportion de consommateurs issue de l'enquête est alors de 6.25% alors que la proportion réelle est de 40%. Le problème de réponses « non sincères » ne dépend pas non plus de l'échantillonnage. Si, dans notre exemple, les 90% de consommateurs qui n'ont pas répondu, répondent en fait qu'ils n'ont pas consommé de marijuana, la proportion de consommateurs issue de l'enquête sera de 4%.

## Présentation de la méthode des quotas

La méthode des quotas est une méthode empirique qui n'a pas de fondement théorique. La méthode des quotas est un ersatz de *sondage aléatoire stratifié proportionnel*. La seule contrainte de la méthode des quotas est de construire un échantillon "proportionnel" pour quelques variables de quotas ; pour le sondage du Monde du 17 mars : sexe, tranches d'âge, catégories socioprofessionnelles, tailles d'agglomération de résidence et régions. L'échantillon n'est donc « représentatif » de la population que pour ces quelques variables. Il est facile de trouver un échantillon de taille 1000 représentatif de la population selon ces variables de quotas dans, par exemple, la sous-population des adhérents de n'importe quel parti politique.

Il est clair qu'une contrainte supplémentaire, peu discutée par les sondeurs, est de *se rapprocher le plus possible d'un sondage aléatoire*. Il faut donc trouver ce qui va jouer le rôle de base de sondage et faire des tirages aléatoires dans cette base. La base de sondage des électeurs a été remplacée par l'annuaire téléphonique dans le sondage du Monde du 17 mars et des numéros de téléphone ont été appelés « au hasard ». Des questions filtres sont alors posées afin de n'interroger que des personnes inscrites sur les listes électorales (voire même seulement des personnes inscrites et ayant l'intention de voter).

Dans le meilleur des cas, c'est-à-dire, sans compter les biais éventuels, la « marge d'incertitude » (à 95% de confiance) d'un sondage par quotas de taille 1000 est de l'ordre de 3%, marge d'incertitude d'un échantillon aléatoire simple (ou d'un sondage aléatoire stratifié proportionnel) de même taille. Certains sondeurs refusent que soit mentionnée la marge d'incertitude d'un échantillon aléatoire qu'ils cherchent à imiter alors que leur « marge d'erreur » est supérieure : « Évoquer de telles marges d'erreur conduit », selon Brice Teinturier, à « donner l'illusion de scientificité » et à « se rassurer à bon compte ». Pourtant, lors des présidentielles de 2002, les électeurs auraient été mieux informés si les sondeurs avaient annoncé les intentions de vote à 3% près ! Il aurait été plus facile de faire entendre des électeurs que l'ordre des trois premiers candidats n'était pas garanti ! De plus, si les résultats des sondages par quotas étaient accompagnés de marges d'incertitude (de sondages aléatoires de même taille) nous aurions beaucoup moins de publications erronées annonçant que tel candidat est passé devant tel autre alors que la différence entre les proportions d'intentions de vote pour les deux candidats est trop faible pour conclure. Il serait souhaitable que les Instituts de sondage fournissent une présentation des résultats de leurs sondages afin d'en limiter un traitement médiatique abusif ; certains le font déjà sur leur site.

Sans pouvoir les mesurer, de nombreuses autres sources d'erreurs s'ajoutent à celle due à l'échantillonnage : le développement des téléphones mobiles affaiblit considérablement la qualité de l'annuaire téléphonique comme base de sondage (d'où le recours à des appels de mobiles) ; il faut réaliser 7000 appels pour obtenir 847 personnes qui acceptent de répondre et qui ont exprimé une intention de vote (mais les non-réponses ne sont pas considérées comme un problème pour les sondeurs puisque les individus qui ne répondent pas peuvent être aisément remplacés) ; l'intention de vote n'est définitive que pour la moitié des répondants, ce qui affaiblit encore un peu plus la précision des résultats publiés.

Les sondeurs utilisent des redressements. Rien de choquant s'il s'agit de *redressement d'échantillon* selon des critères affichés *a priori* (permettant de retrouver la composition sociologique de

l'électorat ou même, bien que ce soit moins fiable, permettant de retrouver la composition électorale d'élections précédentes). Cf. dans ce numéro la contribution de Jean Chiche à ce sujet. En revanche, si la méthode de redressement est déterminée *a posteriori* et dépend des résultats eux-mêmes, on ne peut plus dire que l'enquête dispose d'un fondement scientifique.

## Pour une plus grande transparence sur la méthodologie des sondages électoraux

Les Instituts de sondage doivent remettre une notice méthodologique détaillée à la Commission des sondages (cf. le document en annexe sur ce que devrait, selon le projet de loi, contenir cette notice). Il serait souhaitable que cette notice soit en libre accès sur Internet ; nombreux sont les citoyens capables de les lire. Comme nous l'avons proposé plus haut, il serait souhaitable de disposer, en plus de cette notice, d'une présentation succincte des résultats par les Instituts de sondage ; cela pourrait éviter une utilisation abusive des résultats par les médias.

Nous n'avons évoqué ici que les aspects statistiques des sondages préélectoraux en supposant que les questions sont rédigées sans ambiguïté. La construction des questionnaires d'un sondage est une opération difficile : la rédaction des questions influe fortement sur les réponses. Plus généralement, le fait de demander leur opinion sur un sujet, à des personnes qui n'ont pas d'opinion a priori sur le sujet, pose un problème de fond que les sociologues sont nombreux à souligner.

Depuis plusieurs années, les sondages politiques et électoraux, au lieu d'apporter une information utile aux citoyens, perturbent trop souvent le débat démocratique. Une nouvelle loi sur une plus grande transparence pourrait limiter l'inflation sondagière et améliorer la qualité de ce débat.

### Références

Ardilly, P. (2010). Éléments sur la précision des sondages. Annexe 2 du Rapport d'information au Sénat sur les sondages n° 54 enregistré le 20 octobre 2010. <http://www.senat.fr/rap/r10-054/r10-05421.html#toc216>

Blanc, É. (2011). Rapport de la discussion à l'Assemblée Nationale de la proposition de loi sur les sondages, enregistré le 1er juin 2011. <http://www.assemblee-nationale.fr/13/rapports/r3502.asp>

Commission des sondages : <http://www.commission-des-sondages.fr/>

Fine, J. (2007). Les sondages : délaissés par les statisticiens et malmenés par les politologues [http://jeannefine.free.fr/Sondages-Toulouse2007/documents/doc\\_JFine.pdf](http://jeannefine.free.fr/Sondages-Toulouse2007/documents/doc_JFine.pdf)

Pénombre (2009). Les sondages d'opinion, outils de mesure à utiliser avec précaution ?, Document préparatoire à la 12ème Nocturne de Pénombre 5 juin 2009 <http://www.penombre.org/IMG/File/document-preparatoire.pdf>

Portelli, H. et Sueur, J.-P. (2010). Rapport d'information au Sénat n° 54 enregistré le 20 octobre 2010. Sondages et démocratie. Pour une législation plus respectueuse de la sincérité du débat politique <http://www.senat.fr/rap/r10-054/r10-0540.html#toc0>

Portelli, H. et Sueur, J.-P. (2011). Proposition de loi sur les sondages visant à mieux garantir la sincérité du débat politique et électoral, adoptée par le Sénat le 14 février 2011. <http://www.senat.fr/leg/tas10-063.html>

Rambaud, R. (2013). Le droit des sondages, conférence à l'IFOP <http://droitdesondages.blog.lemonde.fr/2013/07/02/02072013-le-droit-des-sondages-fait-son-entree-a-lifop/>

Riandey, B. (2012) Pédagogie statistique des sondages électoraux. <http://www.sfds.asso.fr/ressource.php?fct=ddoc&i=1321>

SFdS - Groupe Enquêtes et Modèles (2012). Sondages électoraux : débats méthodologiques [http://www.sfds.asso.fr/286-Sondages\\_Electoraux#AncreSem](http://www.sfds.asso.fr/286-Sondages_Electoraux#AncreSem)

# Sondages d'intentions de vote : de la transparence à la confiance

Jean Chiche  
CEVIPOF<sup>1</sup>

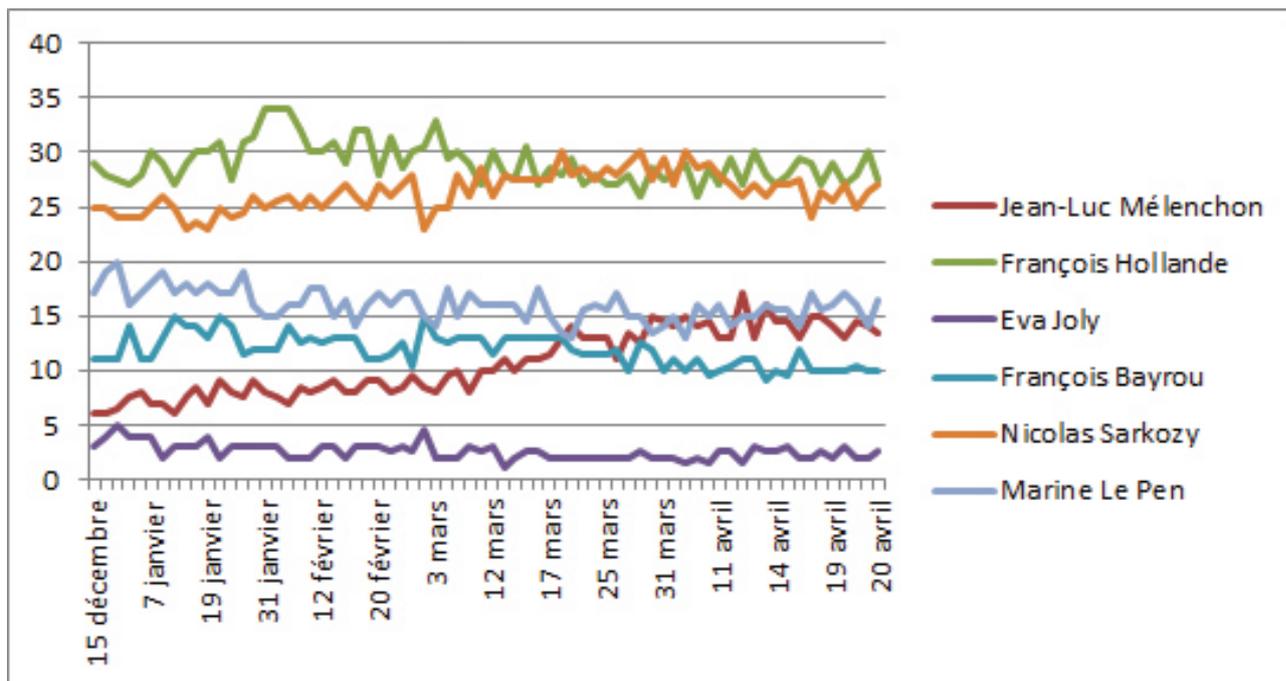
---

Les sondages préélectoraux déconcertent les citoyens par leur profusion, et leurs échecs passés ont laissé un sentiment de défiance. Pourtant, ils donnent de bons résultats, dans des conditions d'exécution difficiles. Pour qu'ils retrouvent la confiance du public, la profession et l'autorité de régulation doivent s'engager toujours plus avant vers la transparence des méthodes.

La dernière élection présidentielle a vu le nombre de sondages augmenter pour atteindre un record. 409 sondages ont été contrôlés par la commission pour l'élection de 2012, contre 293 en 2007<sup>2</sup>. L'émergence de nouveaux acteurs, petites structures proposant essentiellement des enquêtes en ligne a joué certainement un rôle important dans cet accroissement impressionnant intervenu malgré la période de crise du marché des études. Cette abondance d'enquêtes démontre à quel point les media ont eu recours aux sondages pendant la campagne électorale. Et si les media traditionnels, presse, audiovisuel, ont publié autant de résultats d'enquêtes que lors de la campagne de 2007, les nouveaux media sur Internet et les nouvelles chaînes d'information continue n'ont pas été en reste et ont contribué au nombre impressionnant de sondages. Ces enquêtes, souvent plusieurs par jour, permettent de suivre en temps réel l'évolution de la campagne électorale : mais il faut rappeler qu'elles ne permettent pas de prédire les résultats réels plusieurs mois ou plusieurs semaines à l'avance. Seuls les derniers sondages publiés permettent d'inférer sur les probables résultats finaux, à condition qu'il n'y ait pas de grand mouvement de l'opinion dans les tout derniers jours avant le scrutin.

## Saturation ?

Pour autant, plus de 50% des citoyens déclarent lire les sondages publiés, en progression de 8 points par rapport à la campagne de 2007<sup>3</sup>. Mais la saturation n'est pas loin : 63% estiment que le nombre des sondages est trop grand, ils n'étaient que 42% en 2007. Toutes ces études, tous ces chiffres assésés chaque jour donnent une impression de confusion au citoyen électeur qui n'a pas toujours la possibilité de les analyser ou de les mettre en perspective.



**Graphique 1** : 70 résultats d'enquêtes parvenus successivement à la connaissance du public

Ces courbes d'intentions de vote ont été réalisées au Cevipof par l'auteur. Elles utilisent les sondages publiés au jour le jour, à l'exclusion des enquêtes en continu. Les données de ces sondages ont été publiées au jour le jour et sont disponibles sur le site : <http://www.sondages-en-france.fr/>

Le graphique 1 présente la succession des estimations qui ont été publiées pendant les quatre derniers mois de la campagne. Ces courbes montrent à quel point il est difficile, si on ne dispose pas d'instruments adéquats, d'interpréter les mouvements. Il arrive aussi que deux enquêtes donnent le même jour des résultats qui semblent contradictoires, quand on ne les encadre pas des marges d'erreurs classiques. L'aide à l'interprétation des résultats est alors essentielle. Et le plus souvent, même si les instituts publient des mises en garde ou des fiches techniques plus détaillées, elles ne sont pas retransmises au grand public. On doit signaler cependant le travail pédagogique que certains media ont réalisé en proposant des comparateurs de sondages (Le Nouvel Observateur, Le Monde ...). Ces comparateurs ont permis aux électeurs qui souhaitaient suivre l'évolution de la campagne électorale au prisme de cette « course de chevaux » de mieux comprendre des informations complexes comme des évolutions dans le temps de séries numériques. La meilleure compréhension des résultats publiés permet alors une plus grande confiance dans l'instrument sondage.

## Doutes sur la qualité

A la critique du trop-plein s'ajoute le reproche sur la qualité des enquêtes. L'adéquation des résultats avec la réalité fait l'objet de doutes. Seuls 49% des français estiment que les sondages reflètent bien la réalité. Ce sentiment diffus, mais de plus en plus prégnant, est lié au souvenir douloureux du 21 avril 2002. Peu importe que depuis cette date, les sondages d'intentions de vote aient correctement donné les ordres d'arrivées des principaux candidats ou partis à toutes les élections. La confiance a été altérée de façon durable et ceci malgré les efforts d'explications faits après le 22 avril 2002 par les principaux responsables des instituts. Ces explications souvent de nature politologique – volatilité, décision prise au dernier moment ... - n'ont pas convaincu une part importante de l'opinion notamment parmi les classes populaires : 52% des ouvriers et 53% des employés pensent que « les sondages se trompent toujours dans leur prévision de

vote »<sup>4</sup>. Quand la confiance est altérée, la défiance n'est jamais loin avec comme conséquence principale le retour à la théorie du complot. « S'ils ne nous disent pas la vérité, c'est qu'ils sont à la solde des puissants ». Peu importe qui sont les puissants. Nous ne sommes plus ici dans la critique d'ordre statistique ou méthodologique mais dans une critique sociale et politique.

## Malgré les difficultés, les résultats sont bons !

On sait, aujourd'hui, les difficultés de plus en plus grandes, rencontrées dans la réalisation d'enquêtes de qualité<sup>5</sup>. Du plan de sondage (quels numéros sur quels annuaires pour les enquêtes téléphone ou quels internautes interrogés suivant la façon dont ils sont recrutés pour les enquêtes en ligne) aux taux de réponses si faibles qu'il serait illusoire de penser que l'échantillon aléatoire puisse être une bonne méthode aujourd'hui en France pour les enquêtes d'opinion, la réalisation d'étude est de moins en moins aisée. Et les nouvelles sociétés d'études n'ont pas toujours les moyens de travailler dans des conditions qui ne permettraient pas le doute.

Et pourtant malgré toutes ces contraintes et ces difficultés dans la collecte les résultats sont bons.

Si on analyse le tableau des estimations produites par les principaux instituts, à la veille du scrutin du 21 avril 2012, on constatera que les écarts entre instituts sont très faibles et que les écarts aux résultats réels sont également très peu importants. Ces écarts n'ont en aucun cas, en 2007 comme en 2012, modifié l'ordre des deux candidats arrivés en tête. Et c'est là le plus important avec le mode de scrutin de l'élection présidentielle.

Dernier sondage publié et résultats définitifs	Ipsos Tel	CSA Tel	TNS Tel	LH2 Tel	BVA Tel + Web	Ifop Web	OW Web	HI Web	Moy	Moy téléph	Moy Web	Résultats définitifs
Date publication	20-avr-12	19-avr-12	20-avr-12	19-avr-12	20-avr-02	20-avr-12	18-avr-12	20-avr-12				
Date terrain	18-19 avril	18-19 avril	18-19 avril	17-18 avril	19-19 avril	17-20 avril	16-17 avril	18-19 avril				
N. Arthaud	0	1	0	1	0	0,5	0,5	0,5	0,4	0,4	0,5	0,6
P. Poutou	1,5	1,5	1	1	1,5	1	2	1,5	1,4	1,3	1,5	1,2
JL Mélenchon	14	14,5	13	15	14	13,5	13	12	13,6	14,1	12,8	11,1
F. Hollande	29	28	27	27	30	27	27,5	27,5	27,9	28,2	27,3	28,6
E. Joly	2	2	3	2,5	2	3	2	3	2,4	2,3	2,7	2,3
F. Bayrou	10	10,5	10	10	10	10,5	10	11	10,3	10,1	10,5	9,1
N. Sarkozy	25,5	25	27	26,5	26,5	27	27,5	26,5	26,4	26,1	27	27,2
N. Dupont-Aignan	1,5	1,5	2	1,5	2	1,5	1,5	2	1,7	1,7	1,7	1,8
M. Le Pen	16	16	17	15,5	14	16	16	16	15,8	15,7	16	17,9
J. Cheminade	0,5	0	0	0	0	0	0	0	0,1	0,1	0	0,3
Total	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Eff. total inscrits	1021	1005	1000	859	2000	2582	1002	1068				

**Tableau 1** : Les derniers sondages avant l'élection de 2012

Lecture : Les instituts de sondage travaillant uniquement par téléphone figurent dans les quatre premières colonnes (en bleu). Les instituts utilisant le web figurent dans les quatre colonnes suivantes (en jaune).

Source : Daniel Boy, Jean Chiche<sup>6</sup>

## Redonner confiance en accroissant la transparence

En 2002 des erreurs ont été commises, la sous-estimation de Jean Marie Le Pen et la sur estimation de Lionel Jospin ayant entraîné l'inversion que l'on sait. C'était encore une époque, que l'on espère révolue, où les instituts se permettaient d'utiliser des estimateurs différents, suivant des modèles d'hypothèses différents pour chacun des candidats. Rappelons qu'en 2002 il y avait eu 16 candidats ! Aujourd'hui la commission des sondages impose l'utilisation d'une colonne de référence, c'est-à-dire d'un estimateur construit suivant une hypothèse de redressement pour une sous population donnée. Le travail didactique a donc porté ses fruits. Les instituts ont également tenu compte, en partie, des conclusions du rapport sénatorial et de quelques articles de la loi adoptée en première lecture au Sénat<sup>7</sup>. Les rapports de résultats mentionnent tous les marges d'erreur qui doivent être utilisées pour lire les tableaux publiés. Cependant, à de très rares exceptions près, les rapports fournis aux media ne mentionnent ni les résultats bruts, ni les hypothèses de redressement ayant permis les estimations. Ces informations sont disponibles pour la commission des sondages, mais ne peuvent être consultés par le simple citoyen.

Pour redonner confiance aux utilisateurs et lecteurs de sondages, il faut principalement expliquer, être transparent à chaque étape de réalisation de l'enquête et bien sûr être didactique et complet dans la publication des résultats. Cette transparence doit permettre le contrôle complet du travail fait sur un sondage, de l'échantillonnage à l'estimateur utilisé pour établir les résultats. La méthode d'établissement des intentions de vote est maintenant bien connue<sup>8</sup>. Mais les moyens réduits de la commission des sondages ne lui permettent pas d'aller aussi loin qu'on pourrait le souhaiter. Dans le domaine des mesures d'audience, le CESP, pour réaliser ses audits, dispose de personnels très qualifiés et a des moyens de contrôle très importants des enquêtes y compris du terrain. Ses audits sont reconnus par l'ensemble des acteurs du marché et les études auditées sont lues en pleine confiance.

On pourrait imaginer un monde où la transparence méthodologique, qui est la règle en matière scientifique, serait également de mise en matière d'enquêtes électorales. Le citoyen qui désirerait comprendre la publication d'une étude pourrait aller consulter les données fournies à la commission. Et une fois l'élection faite, on peut rêver à un monde où l'universitaire aurait à disposition, en open data, les données individuelles du sondage, qui sont strictement anonymes. A lui ou elle ensuite de faire le travail d'analyse de ces données et de publier ses conclusions.

Cette transparence amènerait de la confiance. Et aujourd'hui les sondages électoraux, même s'ils donnent le plus souvent des résultats fiables, en manquent cruellement.

## Références

1. Centre de recherches politiques - CNRS UMR 7048, Sciences Po
2. Marie Eve Aubin, Mattias Guyomar : la commission des sondages face aux élections présidentielles et législatives de 2012 – Rapport de la Commission des sondages 2012 (<http://www.commission-des-sondages.fr/hist/pdf/2012.pdf>)
3. Selon une enquête de l'IFOP, publiée dans Ifop Focus n° 72 « Les Français et les sondages lors de la séquence électorale de 2012 » (13/11/2012). Disponible sur le site [www.ifop.com](http://www.ifop.com)
4. Selon la même enquête IFOP « sondages sur les sondages »
5. Voir l'excellent article d'A. M. Dussaix : «La qualité des enquêtes » revue MODULAD numéro 39-2008, pages 137-171
6. Daniel Boy et Jean Chiche - Communication au 7ème colloque francophone sur les sondages à Rennes en novembre 2012.
7. Portelli, H. et Sueur, J.-P. (2010). Rapport d'information au Sénat n° 54 enregistré le 20 octobre 2010. Sondages et démocratie. Pour une législation plus respectueuse de la sincérité du débat politique <http://www.senat.fr/rap/r10-054/r10-0540.html#toc0>
8. Portelli, H. et Sueur, J.-P. (2011). Proposition de loi sur les sondages visant à mieux garantir la sincérité du débat politique et électoral, adoptée par le Sénat le 14 février 2011. <http://www.senat.fr/leg/ta10-063.html>
8. Voir par exemple : <http://www.sfds.asso.fr/ressource.php?fct=sdoc&i=1111>

# Un sondage en pleine lumière

Bruno Jeanbart

OpinionWay<sup>1</sup>

Beaucoup de citoyens sont convaincus que les résultats publiés des enquêtes pré-électorales résultent de l'application de « recettes de cuisine » et peuvent donc être manipulés. Faut-il ignorer ces soupçons ? OpinionWay a choisi au contraire de mettre en lumière tous les résultats d'un de ses sondages, depuis les données brutes jusqu'aux chiffres publiés, en les expliquant pas à pas. Cette démarche n'a pas entraîné de polémique, preuve que le public peut comprendre les méthodes des sondeurs...si ceux-ci prennent la peine de les expliquer.

Depuis plus de 30 ans et la publication du célèbre article de Pierre Bourdieu, « L'opinion publique n'existe pas » ou du livre de Patrick Champagne, « Faire l'opinion », les sondages politiques ne cessent de défrayer la chronique, en particulier lors des campagnes électorales. Ce phénomène est exacerbé à l'approche de l'élection présidentielle, qui constitue « la mère de toutes les batailles » dans la vie politique en France. Jeune sondeur, j'entendis un jour l'un de mes aînés affirmer sans ambage que « l'élection présidentielle rendait fous les acteurs et commentateurs politiques en France ». Mon expérience professionnelle depuis n'a fait que confirmer ce propos. La tension qui caractérise les campagnes présidentielles ne cesse de croître et lorsque vous scrutez l'opinion à cette occasion, elle fait partie intégrante de votre quotidien. L'intérêt des médias et des acteurs politiques pour votre travail pendant une campagne présidentielle est décuplé et ne peut laisser insensible. Il se double depuis l'élection de 1995 de polémiques récurrentes sur les sondages publiés durant cette période, dont l'apogée fut probablement le 21 avril 2002. Il faut dire que l'histoire des sondages en France est très liée à l'élection du Président de la République au suffrage universel. C'est en effet le scrutin de 1965 qui installa et légitima les sondages, en annonçant à l'époque un événement qui paraissait incroyable, au sens strict du terme, à savoir la mise en ballottage du Général de Gaulle.

## Entendre les critiques

Tout le paradoxe de notre métier, connu principalement pour une infime partie de son activité, s'exprime dans ce moment si particulier. Il convient de rappeler en effet que les sondages politiques publiés, même s'il est difficile de mesurer précisément la part qu'ils représentent, ne dépassent pas 1 ou 2 % du marché des études. Ils sont la face émergée de l'iceberg, très visible certes, mais peu représentative de l'activité des instituts de sondage, et qui déforme par là même la réalité de notre travail quotidien. Face à la vindicte qui frappe le sondage électoral, deux attitudes sont dès lors possibles. La première consiste à évacuer la question en rejetant en bloc la ou les critiques qui nous sont adressées. C'est une tentation forte et compréhensible compte tenu des excès de la charge contre les sondeurs<sup>2</sup>. Si tout ce qui est excessif est insignifiant, alors ignorons les critiques outrées plutôt que d'y répondre. Le problème, c'est qu'en choisissant cette voie, d'un certain point de vue, nous reproduisons une attitude que nous tentons de combattre chaque jour face à certains de nos clients, en opposant la réalité

aux représentations que les acteurs exogènes s'en font. Or dans ce type de situation, seule l'image que l'on se fait de vous compte et même s'il ne s'agit que d'une perception infondée, c'est celle là qu'il faut corriger. Il semble d'autant plus indispensable de ne pas ignorer ces attaques que la montée en puissance de la valeur « transparence » est une tendance lourde des sociétés démocratiques d'aujourd'hui, la France et la sphère politique n'échappant pas à ce phénomène, comme l'ont montré les débats récents sur le patrimoine des élus. Reste donc une seconde attitude dans ce contexte : entendre la critique pour faire le tri entre ce qu'elle exprime de légitime et les fantasmes qu'elle véhicule. Parce qu'ils interviennent dans la campagne en livrant une information parmi de multiples autres, les sondages politiques doivent naturellement être questionnés, tant par les acteurs que par les observateurs de la campagne. Mais pour cela, encore faut-il qu'acteurs et observateurs disposent des éléments pour le faire. Les « sondeurs politiques » ne peuvent pas s'exonérer de toute responsabilité lorsqu'ils constatent qu'aussi bien les citoyens que les élites connaissent mal leur métier, les techniques qu'ils emploient, les avantages des sondages et leurs limites. C'est la raison pour laquelle en 2012, OpinionWay a décidé d'ouvrir sur son site dédié à la présidentielle un blog<sup>3</sup> pour mieux expliquer la manière dont nous travaillons dans ce moment si particulier et y publier le moment venu le fameux graal qui excite tout l'univers critique des sondages depuis 20 ans, à savoir nos données brutes.

## La repondération politique

Nous pensons depuis un moment qu'il était nécessaire de publier les données brutes afin d'en finir avec le fantasme de la « recette de cuisine », expression trop souvent utilisée pour décrire les « redressements politiques » appliqués aux intentions de vote. En effet, le maintien du secret tend à renforcer et légitimer la critique de l'outil : si les données d'origine demeurent inaccessibles, c'est bien la preuve que l'on nous cache la réalité. Dès lors, comment faire confiance aux résultats publiés et ne pas s'imaginer qu'ils sont calculés au « doigt mouillé », plus ou moins chanceux, du sondeur ? Or, contrairement à ce qui se dit trop souvent, la méthode utilisée pour établir des intentions de vote n'est en rien obscure et ne s'apparente en aucun cas à une quelconque « recette ». Tous les instituts publiant ce type d'enquêtes l'utilisent et elle se base sur une méthodologie statistique précise, qui, rappelons-le, est la suivante :

- dans un premier temps, demander aux répondants de l'échantillon pour qui ils ont voté lors du scrutin comparable précédent (ou de différents scrutins précédents) ;
- puis, pondérer les résultats de cette reconstitution du vote afin d'éliminer les écarts entre le résultat de l'élection passée obtenu dans l'échantillon et le résultat réel de l'élection. Ainsi, si seuls 12 % des sondés déclarent avoir voté pour Marine Le Pen à la dernière élection présidentielle alors qu'elle a obtenu 17,9 % des suffrages, chaque électeur déclaré de Marine Le Pen verra son poids dans l'échantillon augmenter pour corriger cet écart ;
- enfin, appliquer les effets de cette pondération statistique à l'intention de vote, pour tenter de corriger les biais éventuels liés à cet effet normatif qui pèse sur le déclaratif des individus.

## Publier en expliquant

Avant de passer à l'acte, il convenait toutefois de déterminer sous quelle forme se ferait cette publication des données brutes. Je n'ai jamais été favorable à la diffusion des données brutes au même titre que des données pondérées, dans le ou les médias publiant l'enquête. En effet, la compréhension des résultats du sondage n'est pas un élément neutre et notre métier consiste à

donner une information claire de ce que nous estimons être l'état du rapport de force électoral. Notre objectif en publiant ces données n'est pas de rendre illisible cette information mais bien de permettre à ceux qui veulent en savoir plus et comprendre notre manière de procéder de pouvoir le faire. De plus, de mon point de vue, mesurer des intentions de vote, c'est aussi faire le choix de la pondération à retenir et assumer ce choix. L'hypothèse d'une publication à deux colonnes dans les médias (les résultats bruts et les « redressés ») n'a donc à aucun moment été envisagée. En revanche, nous pensions dans un premier temps adjoindre ces données en annexe de nos rapports. Toutefois, de nombreuses discussions avec d'anciens sondages nous ont alerté sur le risque d'une telle démarche, si elle n'était pas accompagnée d'explications de notre part. Ces remarques nous ont largement convaincus dans la mesure où nous cherchions autant à « casser » le mythe des données brutes qu'à faire preuve de pédagogie sur l'instrument lui-même. C'est pourquoi nous avons plutôt choisi, pour cette première, une publication des données sur un blog, qui offrait la possibilité de commenter les données, la manière dont elles étaient calculées et les choix que nous avons retenus. Elle permettait aussi de rappeler en amont ce qu'était une « intention de vote », ce qu'elle n'était pas, et d'en rappeler l'intérêt mais aussi les limites (en évoquant le problème de la marge d'erreur notamment). Pour des raisons de temps disponible, cette diffusion des différentes étapes de notre travail s'est concentrée sur les dernières enquêtes publiées avant le premier<sup>4</sup> et second tour<sup>5</sup> de la présidentielle, mais également avant le premier tour des législatives<sup>6</sup> (soit les vendredi 20 avril, 4 mai et 8 juin). A chaque fois, nous avons respecté le calendrier suivant : publication de l'enquête le vendredi matin dans les médias, mise en ligne sur notre blog des tableaux de redressement commentés dans la journée du vendredi. Il nous semblait en effet évident que les données devaient être diffusées avant le scrutin. Dans l'idéal, il faudrait pouvoir mettre en place la même démarche pour toutes les enquêtes publiées lors de la campagne.

## Une expérience à renouveler

Que retenir de cette expérience ? Tout d'abord, que les Cassandre nous annonçant que la publication des données brutes ne ferait qu'envenimer la situation ont eu tort. La plupart de mes confrères sont défavorables à la publication de ces données pour cette raison. Ils craignent qu'une telle démarche n'ait pour effet que de renforcer et réactiver les polémiques. Force est de constater que ce ne fut pas le cas. Il y eut bien quelques polémiques, notamment lorsque des enquêtes annoncèrent un croisement de courbe au premier tour entre Hollande et Sarkozy, mais pas sur notre démarche. Certes, la présidentielle de 2012 fut un scrutin assez simple, avec peu de rebondissements, un résultat sans surprise et un bon cru pour les intentions de vote de plusieurs instituts. Les différentes hypothèses de redressement étaient souvent très proches et sur le second tour, plaidaient toujours en faveur d'une consécutive avance de François Hollande. Par ailleurs, le remplacement de Jean-Marie Le Pen par sa fille a profondément amoindri la sous-déclaration du vote FN, rendant là encore les tableaux de redressement moins spectaculaires. Il faudrait donc vérifier que la publication des données brutes n'enflamme pas le débat dans une élection qui serait elle-même plus ouverte et sujette à débat, par exemple en plaçant alternativement en tête l'un et l'autre des candidats au second tour.

Je regrette par ailleurs que cette publication ait suscité si peu de réaction, même si je n'en suis guère surpris : en levant le voile, vous coupez en partie court aux arguments des « conspirationnistes ». Il est dommage également de ne pas avoir pu vérifier l'absence d'effet pervers de la diffusion des données brutes à l'occasion des sondages ayant fait « débat » durant cette campagne, mais ce n'est probablement que partie remise. En effet, la publication de ces données va dans le « sens de l'histoire » et je ne doute pas que cette règle finira par s'imposer. Si les instituts ne le font pas d'eux-mêmes, il y a fort à parier que le législateur leur en fera obligation demain à la première controverse venue. Pour ma part, je souhaite continuer voire amplifier cette démarche, dès les échéances électorales de 2014.

	Hypothèses de redressement						RESULTATS PUBLIES
	Base : Inscrits sur listes électorales 1002 personnes			Base : Inscrits sur listes électorales et certains d'aller voter 740 personnes			
	BRUT	Socio démographique (SD)	SD & Présidentielle 2007	BRUT	Socio démographique (SD)	SD & Présidentielle 2007	
Jacques Cheminade	-	-	-	-	-	-	0,0%
Nathalie Arthaud	1%	1%	1%	1%	-	-	0,5%
Philippe Poutou	2%	2%	2%	2%	2%	2%	2,0%
Jean-Luc Mélenchon	11%	11%	12%	11%	12%	13%	13,0%
François Hollande	28%	27%	27%	29%	28%	28%	27,5%
Eva Joly	3%	3%	2%	3%	2%	2%	2,0%
François Bayrou	14%	13%	12%	12%	12%	10%	10,0%
Nicolas Sarkozy	26%	27%	28%	26%	27%	28%	27,5%
Nicolas Dupont-Aignan	1%	2%	1%	2%	2%	1%	1,5%
Marine Le Pen	14%	14%	15%	14%	15%	16%	16,0%
N'expriment pas d'intention de vote	8%	8%	6%	4%	4%	3%	8%

**Tableau** des différentes hypothèses de redressement de notre dernière intention de vote avant le 1er tour de la présidentielle de 2012, publié sur notre blog

La colonne « Socio-démographique (SD) » contient les résultats repondérés pour éliminer les différences entre la composition socio-démographique de l'échantillon (par sexe, groupe d'âges et catégorie socio-professionnelle) et la composition socio-démographique du corps électoral, connue par ailleurs. La colonne « SD et Présidentielle 2007 » contient les résultats repondérés pour éliminer ces différences et aussi les différences entre les souvenirs de vote 2007 de l'échantillon et les résultats de la présidentielle de 2007.

	Hypothèses de redressement						RESULTATS PUBLIES
	Base : Inscrits sur listes électorales 2009 personnes			Base : Inscrits sur listes électorales et certains d'aller voter 1561 personnes			
	BRUT	Socio démographique (SD)	SD & T1 Présidentielle 2012	BRUT	Socio démographique (SD)	SD & T1 Présidentielle 2012	
François Hollande 	54%	54%	53%	53%	53%	52%	<b>52,5%</b>
Nicolas Sarkozy 	46%	46%	47%	47%	47%	48%	<b>47,5%</b>
N'expriment pas d'intention de vote	20%	19%	23%	13%	12%	11%	19%

**Tableau** des différentes hypothèses de redressement de notre dernière intention de vote avant le 2nd tour de la présidentielle de 2012, publié sur notre blog

La colonne « SD et T1 Présidentielle 2012 » contient les résultats repondérés pour éliminer les différences de composition socio-démographique entre l'échantillon et le corps électoral, et aussi les différences entre les déclarations de vote de premier tour de l'échantillon et les résultats du premier tour.

## Références

1. Directeur général adjoint de l'institut de sondages OpinionWay
2. Sans verser dans l'inventaire à la Prévert, je citerai ici juste quelques expressions entendues depuis 10 ans concernant notre métier : les sondages se tromperaient toujours, ils manipuleraient l'opinion – ces deux arguments, bien que contradictoires, étant souvent avancés de concert par les mêmes – seraient l'un des principaux « bras armés du capitalisme » et bien sûr les principaux responsables de la qualification de Jean-Marie Le Pen au second tour de la présidentielle en 2002. Tout cela est loin d'être exhaustif...
3. <http://opinionlab.opinion-way.com/blog.html>
4. résultats et commentaires disponibles ici [http://opinionlab.opinion-way.com/blog\\_entry/28/les-bruts-de-quoi-parle-t-on-au-juste.html](http://opinionlab.opinion-way.com/blog_entry/28/les-bruts-de-quoi-parle-t-on-au-juste.html)
5. résultats et commentaires disponibles ici [http://opinionlab.opinion-way.com/blog\\_entry/32/les-bruts-la-suite.html](http://opinionlab.opinion-way.com/blog_entry/32/les-bruts-la-suite.html)
6. résultats et commentaires disponibles ici [http://opinionlab.opinion-way.com/blog\\_entry/35/les-bruts-suite-mais-pas-fin.html](http://opinionlab.opinion-way.com/blog_entry/35/les-bruts-suite-mais-pas-fin.html)



# Hommage à Jacques Antoine (19/08/1928 - 25/12/2012)

Daniel Bachelet

Statisticien<sup>1</sup>

Pionnier du développement des sondages d'opinion en France, statisticien particulièrement sensible aux questions déontologiques, Jacques Antoine, longtemps le seul expert auprès de la Commission des sondages, s'est éteint à la fin de l'année dernière.

Jacques Antoine a eu une vie professionnelle particulièrement riche ; statisticien spécialiste des sondages et du marketing, économiste, sociologue et expert en prospective, il était doué et passionné par la nature de son travail. Il n'est donc pas surprenant qu'en 2010 il était toujours, depuis plus de 25 ans, expert auprès de la commission de contrôle des sondages d'intentions de vote et membre actif du comité de rédaction de la Revue Française de Marketing dont il avait été le président du comité des publications de 1978 à 1986 !

J'ai bien connu Jacques Antoine à partir de 1967 à la SEMA (Société d'Economie et de Mathématiques Appliquées) en travaillant pour lui et avec lui à la mise au point de l'échantillon de bureaux de vote et du modèle permettant d'en déduire le nombre de députés par parti dès 20 heures pour RTL. Après avoir fait l'X et l'ISUP-ENSAE, puis travaillé depuis 1951 comme administrateur de l'Insee sur les



recensements et sondages de la statistique publique, il était en effet entré à la SEMA en 1958 comme directeur du département des Etudes Economiques, puis en 1962 s'était vu confier à sa création la direction de la Sofres par Jacques Lesourne.

Il s'agissait en effet de concurrencer l'Ifop qui avait réussi en 1965 pour Europe 1 son estimation par les bureaux de vote pour la présidentielle De Gaulle / Lecanuet / Mitterrand. Les estimations du 1er tour en % des voix en 1967 ont été tout à fait correctes dès 20 heures, y compris pour une dizaine de personnalités en danger ; elles donnaient une très nette victoire de la droite qui devait lui assurer une avance de près de 100 sièges. Mais le fort succès de la droite au 1er tour a conduit une partie de ses électeurs à ne pas aller voter au second tour et en revanche a mobilisé les électeurs de gauche. La droite n'a finalement eu que 2 sièges d'avance ! Circonstance aggravante, notre modèle ne pouvait pas converger car nous avons écouté le conseil des

politologues Sofres Pierre Weill et Alain Lancelot qui malgré leur talent (A. Lancelot avait fait sa thèse sur l'abstention) ont considéré qu'il n'y avait qu'environ la moitié des circonscriptions qui étaient indécises au vu du 1er tour et qu'il ne fallait échantillonner des bureaux de vote que pour ces circonscriptions ; mais les comportements des abstentionnistes sont particulièrement difficiles à prévoir....

Jacques Antoine a décrit avec son objectivité habituelle les difficultés de ces sondages politiques dans un de ses 5 livres sur les sondages : « Histoire des Sondages »<sup>2</sup>. Ce livre, passionnant à lire, traite aussi des sondages marketing, des sondages pour la recherche économique et sociale et des sondages sur l'audience des médias. Jacques Antoine, qui a été longtemps le directeur du CESP (Centre d'Etude des Supports de Presse), organisme d'audit et de conseil indépendant, était en effet aussi un expert de ces enquêtes media.

Après avoir quitté la SEMA au début des années 70, Jacques Antoine a enseigné au Conservatoire des Arts et Métiers le marketing et « le sondage comme outil du marketing », titre du livre qu'il a fait paraître chez Dunod en 1981.

Pendant cette période, il a été très actif à ESOMAR (European Society for Opinion and Marketing Research) où il a notamment participé aux améliorations du code déontologique (anonymat, interrogation des enfants, achat mystère, etc.). Il a aussi, en 1978 avec Ludovic Lebart, mis au point l'enquête annuelle « Aspirations des Français » du CREDOC (Centre de Recherche pour l'Etude et l'Observation des Conditions de vie). Jacques Antoine était un membre actif de l'Institut International de Statistique, l'association mondiale des statisticiens, et de la SFdS (Société Française de Statistique). Il a, par exemple en 1997, présenté une communication au premier colloque francophone sur les sondages de la SFdS « Prospective des techniques et défis méthodologiques ». Au total, Jacques Antoine a diffusé plus de trente publications dont une datant de 1991 vient d'être rééditée en décembre 2010 par la Revue française de Marketing (RFM) : « Un nouveau métier pour les hommes de marketing : la veille prospective et ses applications stratégiques » (première publication au 44ème Congrès annuel de l'Esomar, Luxembourg, septembre 1991 et revue RFM 1992 N°139).

Jacques Antoine, qui se passionne de plus en plus pour la prospective, a en effet créé dès les années 80 avec sa fille Mme Paille le « Cesem Opinion » (Centre d'Etudes Socioéconomiques et de Management - Opinion) qui a pour ambition :

- d'une part, de dégager des signaux d'évolution à long terme des valeurs de société à partir de toutes les données disponibles : les sondages publiés de plus en plus nombreux (le « sondoscope » mensuel), les évolutions sociodémographiques et économiques et aussi à partir de veilles culturelle, technologique, environnementale et politique.

- d'autre part, d'aider les entreprises à élaborer des scénarios du futur pour éclairer leurs choix stratégiques.

C'est dans ce cadre qu'en 1995 PSA Peugeot Citroën a confié à Jacques Antoine et au Cesem la coordination d'une étude prospective sur le marché automobile européen avec un groupe de travail comportant des économistes et des sociologues. Quatre scénarios contrastés ont été élaborés avec leurs conséquences prévisibles sur l'évolution du marché par segment produit (segment des berlines du haut de gamme, des tous terrains, des véhicules hybrides-électriques, etc.). Nous avons publié avec Jacques Antoine la méthodologie de cette étude dans la RFM (N°179/180, 2000/4-5) et les 4 scénarios décrits ne paraissent pas 15 ans plus tard obsolètes ; nous distinguons le « nous » socioculturel (protection sociale) allant de pair avec la solidarité comme tendance économique dominante, le « nous-moi » (besoin de sens) associé à l'écologie, le « moi-nous » (hédonisme tribal, famille) à la qualité de vie et le « moi » (individualisme) associé à une économie libérale. Monsieur Calvet a beaucoup apprécié l'étude, mais nous a

demandé de préciser annuellement de quel scénario la France et l'Europe se rapprochaient le plus, ce qui n'a pas été simple !

Jacques Antoine, expert auprès de la Commission des Sondages depuis sa création suite à la loi de 1977, avait réussi progressivement à convaincre les instituts de sondage de transmettre à cette Commission une notice confidentielle très complète comportant tous les tris (par exemple sur la certitude d'aller voter) et pondérations utilisés (sociodémographiques et souvenirs de vote antérieurs) pour élaborer un ou plusieurs redressements des intentions de vote. Les instituts retenaient alors soit un seul résultat, soit des « fourchettes » entre le résultat le plus faible et le plus fort de chaque candidat à l'intérieur desquelles les politologues effectuaient un choix. Après avoir rejoint Jacques Antoine début 2002 comme expert auprès de la Commission des Sondages, nous avons constaté ensemble que les méthodes d'élaboration des scores publiés par certains instituts étaient très discutables et avaient eu une part de responsabilité dans la présence de Jean-Marie Le Pen au second tour de l'élection présidentielle. Nous avons donc cherché à convaincre les politologues des instituts de sondage d'utiliser leurs analyses pour choisir la méthode statistique de redressement par pondération la plus appropriée, appelée redressement de référence, mais pas pour modifier les résultats obtenus par cette méthode même en restant à l'intérieur de la fourchette de plusieurs redressements. La Commission des Sondages a alors décidé de recommander aux instituts de publier les résultats du redressement de référence, sauf exception devant toujours donner lieu à des justifications que la Commission se réservait le droit de remettre en cause dans un communiqué de presse.

Sondage, marketing, prévision, prospective, dans tous ces domaines connexes, Jacques Antoine a étonné « par sa rigueur, son impartialité et son honnêteté intellectuelle » comme vient de l'écrire la déléguée générale de l'ADETEM-RFM. Une prévision comporte toujours un risque d'erreur et en ce qui me concerne, il n'y a qu'une seule prévision dont je sois certain : ceux qui ont travaillé avec lui ne l'oublieront jamais.

## Références

1. Directeur du département prévision de la SEMA, puis directeur du marketing à la Direction des Plans et Produit de PSA Peugeot Citroën de 1976 à 1998, puis ensuite conseil en marketing et expert auprès de la Commission des Sondages.
2. "Histoire des sondages" Jacques Antoine Editions Odile Jacob 2005.



# Nature et déterminants de l'erreur d'échantillonnage dans les enquêtes par sondage

Pascal Ardilly

Insee, direction de la méthodologie

Lorsqu'on cherche à estimer une grandeur relative à une population à partir de données collectées seulement sur un échantillon, on s'expose à une erreur d'échantillonnage. Cette erreur a deux composantes : le biais, qui existe dès lors que la moyenne des estimations issues de tous les échantillons possibles diffère de la grandeur visée, et la variance, qui traduit une forme d'instabilité des estimations. Les sondages probabilistes permettent d'estimer l'erreur d'échantillonnage. Les sondages empiriques, par construction, ne le permettent pas. Il est entré dans l'usage de publier pour les sondages empiriques des estimations de l'erreur d'échantillonnage obtenues en « faisant comme si » il s'agissait de sondages probabilistes. C'est une pratique défendable dans certaines circonstances, mais un professionnalisme élevé est nécessaire pour la justifier.

Les statisticiens d'enquête ont pour mission de mettre en œuvre des méthodes permettant de connaître « au mieux » des grandeurs définies sur une population. Ces grandeurs sont le plus souvent des moyennes, des totaux ou des proportions. Par exemple, juste avant une élection, on souhaite connaître la proportion d'électeurs qui vont voter pour un candidat donné. On peut aussi s'intéresser chaque mois au nombre de personnes en recherche d'emploi, ou au chiffre d'affaire annuel moyen des entreprises innovantes. En situation idéale, les valeurs exactes de ces grandeurs – que l'on appelle des paramètres – peuvent être obtenues par recensement, c'est-à-dire par une collecte exhaustive des informations individuelles formant le paramètre. Mais le plus souvent, cette option est trop onéreuse et irréalisable en pratique: on doit alors se tourner vers la solution alternative, beaucoup moins lourde, qu'est l'enquête par sondage, c'est-à-dire une collecte de données auprès d'une partie seulement de la population (l'échantillon).

Toute enquête par sondage se construit en distinguant deux étapes essentielles : la sélection de l'échantillon d'une part et le choix de la méthode d'estimation d'autre part, c'est-à-dire le processus de traitement des données collectées qui va fournir une valeur supposée proche du paramètre inconnu – cette valeur est appelée « estimation du paramètre ». L'économie de moyens qu'offre le sondage a néanmoins une contrepartie qui est l'erreur d'échantillonnage, autrement dit un inévitable écart entre le paramètre et son estimation, laquelle va d'ailleurs dépendre de l'échantillon tiré. En fin d'opération, il est souhaitable d'apprécier la qualité de l'estimation produite en proposant une estimation de l'erreur d'échantillonnage.

La théorie des sondages offre aujourd'hui les moyens de quantifier l'erreur d'échantillonnage dans toutes les circonstances créées par la combinaison d'une méthode d'échantillonnage et d'une méthode d'estimation. Curieusement, alors que les outils probabilistes ont quelques siècles d'existence, l'approche spécifique aux sondages – une forme d'extrapolation pourtant bien naturelle en population finie – a créé une discipline nouvelle qui a connu un développement

flamboyant après la dernière guerre, sous l'impulsion de pères fondateurs parmi lesquels on compte Jerzy Neyman, Morris Hansen ou William Cochran.

## Les deux composantes de l'erreur d'échantillonnage

L'erreur d'échantillonnage s'apprécie au travers de deux composantes: le biais et la variance de l'estimation (en toute rigueur, le terme « estimateur » serait plus correct, mais nous adopterons ici la terminologie la plus commune). Pour illustrer ces concepts, partons du problème posé par l'estimation de la proportion de votants pour un candidat donné à des élections présidentielles. Avant le scrutin, c'est évidemment un véritable enjeu que de prédire le nom du candidat qui l'emportera ! Il faut comprendre que la proportion exacte que l'on cherche à approcher au mieux et que l'on notera désormais  $P_{vrai}$ , est une proportion certes inconnue mais que l'on pourrait obtenir de façon exacte si on disposait de moyens suffisants pour questionner l'ensemble des électeurs ...et sous condition évidemment que ceux-ci ne changent pas d'avis entre la date du sondage et celle de l'élection proprement dite. Au demeurant, le risque causé par un éventuel changement d'opinion n'est pas à mettre au compte de l'erreur d'échantillonnage, c'est là un tout autre phénomène qui tient au fait qu'un paramètre – comme une population d'ailleurs – évolue avec le temps.

### Le biais ...

Considérons maintenant un axe sur lequel on place la valeur de  $P_{vrai}$ , et imaginons tous les échantillons possibles que l'on peut construire à partir de la population complète (le nombre de ces échantillons est gigantesque dès que la population dépasse quelques dizaines d'individus) : chaque échantillon donne lieu à une estimation de proportion qui lui est propre et qui se positionne sur l'axe au moyen, mettons, d'une petite croix. Il est clair que toutes ces petites croix se répartissent entre les valeurs zéro et un : en effet, certains échantillons sont composés uniquement de partisans du candidat-cible (estimation égale à 100%, soit 1) et certains échantillons, tout à fait à l'opposé, ne rassembleront que des opposants au candidat – l'estimation de la proportion étant alors égale à 0%. La théorie dit que la probabilité de tomber « par hasard » sur un de ces échantillons extrêmes est très faible ...mais elle n'est pas nulle ! Il est par ailleurs bien clair que plus l'échantillon est de grande taille, moins on a de chance de rencontrer une situation aussi atypique. Entre ces valeurs extrêmes, compte tenu des très grandes tailles de population auxquelles on a à faire, on trouve un quasi-continuum de situations : on se convainc facilement que, quelle que soit une estimation donnée a priori et comprise entre 0 et 100 %, il existe toujours (au moins) un échantillon qui conduit à une estimation qui en est extrêmement proche.

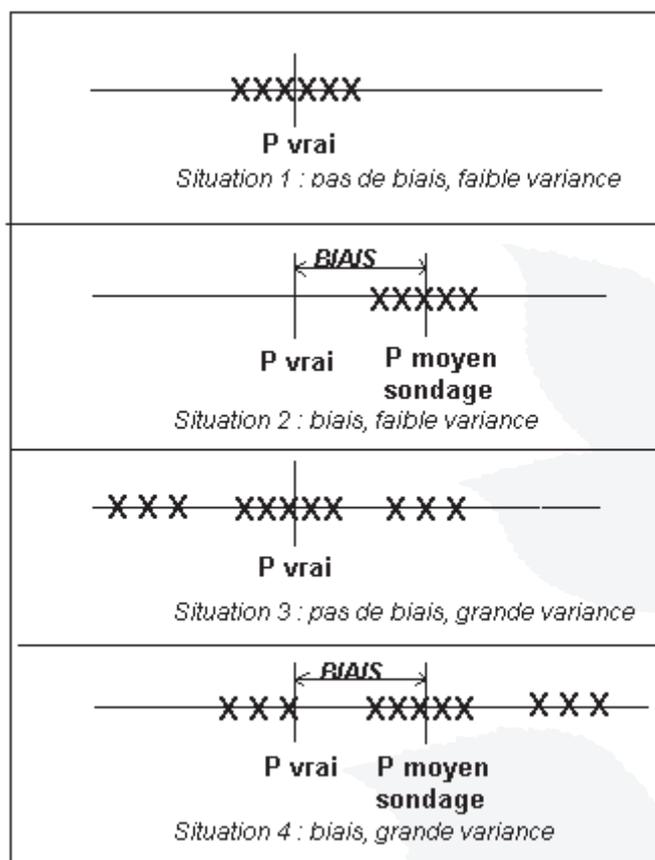
Par définition, un biais existe dès lors que la moyenne des estimations, formée à partir de l'ensemble de tous les échantillons que l'on peut constituer dans la population, diffère du paramètre: graphiquement, c'est le cas lorsque la moyenne de toutes les petites croix – on va l'appeler  $P_{sondage}^{moyen}$  – n'est pas égale à  $P_{vrai}$ . Dans le cas contraire, on a un plan de sondage sans biais. Lorsqu'on parle de moyenne des estimations, il faut comprendre que la moyenne en question est pondérée par les probabilités de sélectionner les échantillons considérés (techniquement, on parle « d'espérance mathématique »). Le schéma des petites croix est certes pratique pour comprendre la notion de biais mais cette approche purement graphique se justifie plus particulièrement lorsque tous les échantillons que l'on peut construire ont la même probabilité d'être sélectionnés. Très souvent, ce n'est pas le cas car on privilégie certains échantillons par rapport à d'autres. C'est alors comme si notre axe constituait le fléau d'une balance articulé au niveau de la vraie proportion  $P_{vrai}$  et que chaque petite croix positionnée sur ce fléau avait un poids égal à la probabilité de tirer l'échantillon auquel elle est associée. Si le plan de sondage est sans biais, la balance est équilibrée; dans le cas contraire le fléau va pencher, à droite s'il y a

surestimation, à gauche s'il y a sous-estimation. Les sondeurs ont généralement de l'aversion à mettre en place des plans de sondage biaisés – ce qui est néanmoins inévitable lorsqu'il y a de la non-réponse parce que les mécanismes probabilistes de la non-réponse ne sont ni contrôlés ni connus par le sondeur.

### ...et la variance

La variance est définie comme la moyenne des carrés des écarts entre les estimations possibles et leur espérance mathématique. Techniquement, à l'image exacte de ce qui se fait pour définir l'espérance, la moyenne des carrés est pondérée par les probabilités de tirer les différents échantillons que l'on pourrait sélectionner. La variance traduit une forme d'instabilité des estimations (ce que le biais ne permet pas du tout d'apprécier): si l'estimation est numériquement très sensible à l'échantillon tiré, il y aura une grande variance. Reprenant notre approche graphique, on est confronté à une grande variance lorsque les croix sont très étalées le long de l'axe, et à une petite variance si elles sont concentrées à un endroit quelconque de l'axe (qui n'est pas nécessairement égal à  $P_{vrai}$ , ni même proche de ce paramètre – cet aspect renvoie à la notion de biais). Mais là encore, la réalité va un peu au-delà de ce que traduit le graphique : si certaines petites croix sont très éloignées de la vraie valeur  $P_{vrai}$  mais que les échantillons qui les créent ont extrêmement peu de chance d'être sélectionnés, alors la variance va rester faible. C'est là que l'on perçoit la faiblesse de ces approches « en moyenne » : un « mauvais » hasard, certes peu probable mais pas impossible à rencontrer, peut conduire à une estimation extrême sans que la variance n'en soit vraiment affectée. Si l'estimation est grossièrement excessive, contraire au sens commun, on adoptera évidemment des méthodes de correction, éventuellement on refera l'enquête mais un échantillon simplement « un peu trop » déséquilibré ne sera pas forcément détecté par un indicateur comme la variance. Des techniques spécifiques dites de redressement permettent de réduire sensiblement ce risque, mais ceci est une autre histoire...

Le biais et la variance, considérés conjointement, constituent une information très importante qui permet d'apprécier l'erreur d'échantillonnage, constituant elle-même une composante de la qualité d'une enquête par sondage. L'ampleur du biais ne préjuge pas de celle de la variance, si bien qu'on peut trouver en pratique des plans de sondage de toutes sortes, que l'on classerait schématiquement en quatre catégories (voir graphique 1).



**Graphique 1 :** Les quatre situations dans lesquelles un plan de sondage peut se trouver

La situation 1 est la meilleure et la situation 4 est la pire. Les expressions mathématiques du biais et de la variance dépendent à la fois de la méthode d'échantillonnage et de la procédure d'estimation. S'ajoute l'impact – toujours « négatif » – du phénomène de non-réponse, qui génère systématiquement une dégradation à la fois du biais et de la variance (la non-réponse diminue la taille de l'échantillon, donc augmente la variance). Dans tous les cas, dans l'approche de la théorie classique des sondages, il est nécessaire de connaître le contexte probabiliste qui préside au tirage des individus de l'échantillon. Autrement dit, il faut être capable de dire avec quelle probabilité un individu a été échantillonné (on parle de « probabilité de sélection »). Dans le cas contraire, on ne peut pas estimer l'erreur d'échantillonnage de manière rigoureuse. Il faut également faire des hypothèses de comportement pour traiter la non-réponse et comme il s'agit justement d'un processus que l'on ne contrôle pas du point de vue probabiliste, le biais est inévitable et le calcul de variance est nécessairement entaché d'erreur.

## La problématique des intervalles de confiance

La variance n'est pas le concept le plus abouti en matière d'erreur d'échantillonnage : en effet, le plus souvent, on diffuse des intervalles de confiance pour juger de la qualité d'une estimation. Dans l'approche classique, l'intervalle de confiance à  $x\%$  (classiquement 95%) est constitué par une limite inférieure et une limite supérieure, calculées de telle sorte qu'il y ait  $x$  chances sur 100 (classiquement 95 chances sur 100) pour que le paramètre recherché ( $P_{vrai}$  dans notre exemple) soit compris entre ces limites. Les bornes de cet intervalle sont aléatoires car elles dépendent de l'échantillon tiré. La théorie de l'intervalle de confiance fait appel à quelques hypothèses techniques dont voici les deux principales : premièrement, la taille d'échantillon doit être « assez grande » (ce qui est généralement satisfait) et deuxièmement, l'estimation doit être sans biais – ce qui permet de centrer l'intervalle sur l'estimation du paramètre auquel

conduit l'échantillon sélectionné. La seconde condition nous intéresse tout particulièrement car nous verrons que les estimations des sondages dits « empiriques » sont en toute rigueur biaisées. Sans le justifier davantage, signalons que l'intervalle de confiance dit « à 95% » se construit en général ainsi: la limite inférieure est égale à l'estimation issue de l'enquête moins deux fois la racine carrée de la variance, et la limite supérieure est égale à l'estimation issue de l'enquête plus deux fois la racine carrée de la variance.

La perception du sens de l'intervalle de confiance est souvent erronée: il ne faut surtout pas imaginer que le vrai paramètre  $P_{vrai}$  se trouve a priori « quelque part » de manière uniforme au sein de cet intervalle. Derrière cette notion, il y a des éléments probabilistes qui montrent que  $P_{vrai}$  se trouve plus probablement plus près du centre de l'intervalle que de ses extrémités: plus on se rapproche du centre de l'intervalle, plus on a de chances d'y trouver  $P_{vrai}$  (techniquement, l'estimation suit une loi probabiliste bien connue appelée loi de Gauss).

Par ailleurs, on peut discuter longtemps sur le choix de la probabilité de 95%, qui est extrêmement courant en pratique, on pourrait même dire systématique, mais néanmoins largement conventionnel : en effet, on pourrait tout aussi bien opter pour des intervalles de confiance associés à une autre probabilité. C'est évidemment un moyen d'influer – de manière plus ou moins insidieuse il est vrai – sur l'amplitude de l'intervalle de confiance. Ainsi, l'intervalle de confiance à 90% sera plus étroit que l'intervalle à 95%, ce qui donnera certes une apparence de meilleure précision mais par définition on augmente le risque d'une mauvaise prévision du résultat des élections. De ce point de vue, les sondages électoraux ont la spécificité d'être confrontés ultérieurement à la vraie valeur (au – sérieux – problème près des changements d'opinion...) et cet aspect est essentiel en terme de communication. A l'opposé, pour limiter les risques de mauvaise prévision, on peut construire un intervalle de confiance couvrant la vraie valeur avec plus de 95 chances sur 100 – mais l'intervalle va s'élargir: si on opte pour un intervalle de confiance à 99%, les limites de l'intervalle sont construites en calculant plus ou moins 2,6 fois la racine carrée de la variance (au lieu de 2): la largeur de l'intervalle augmente donc de 30% par un simple effet d'affichage !

## Sondages probabilistes versus sondages empiriques et calculs d'erreur d'échantillonnage

Si on s'intéresse à la phase d'échantillonnage, on distingue traditionnellement deux classes de méthodes: d'une part les échantillonnages probabilistes, d'autre part les échantillonnages empiriques – la méthode empirique la plus connue étant la méthode des quotas.

Les sondages probabilistes ont pour caractéristique de permettre le calcul de la probabilité de sélection de chaque individu de la population couverte par l'enquête. L'échantillonnage probabiliste relève en effet de règles de sélection extrêmement précises dans une population au sein de laquelle chaque individu est clairement identifié. Une méthode de tirage totalement objective (un « algorithme » mathématique), sans aucune intervention humaine, permet d'associer à chaque individu de la population une probabilité connue d'être sélectionné. Pratiquement, c'est un programme informatique qui tire l'échantillon à partir d'un fichier informatique qui recense l'intégralité de la population (on parle de base de sondage).

A l'opposé, l'échantillonnage des sondages empiriques ne peut pas donner lieu à un calcul rigoureux des probabilités de sélection des individus enquêtés. L'échantillonnage empirique – disons de type quotas – relève en effet d'une sélection non totalement contrôlée, offrant une composante subjective parce que la sélection est concrètement réalisée par l'enquêteur en fonction des circonstances qu'il rencontre. Ce dernier dispose bien entendu de consignes pour limiter cette part d'appréciation subjective au moment de l'échantillonnage mais il conserve

– inévitablement – une liberté qui ne permet pas de maîtriser les aspects probabilistes de la sélection. La qualité d'un échantillonnage empirique est donc plus difficile à apprécier avec les outils qu'offre la théorie classique des sondages: de fait, ils n'autorisent pas de mesure formelle de la qualité à partir des notions classiques de biais et de variance d'échantillonnage, ce qui constitue une critique récurrente des méthodes empiriques (qui ont par ailleurs des atouts d'une autre nature).

Les plans de sondage probabilistes permettent des estimations en théorie sans biais mais en pratique et *in fine*, les estimations sont toujours biaisées parce qu'on est systématiquement confronté à un phénomène de non-réponse. Par ailleurs, s'il y a une seule règle statistique à énoncer ici, c'est celle qui traduit la diminution de la variance lorsque la taille de l'échantillon augmente: avec un tirage équiprobable d'individus (tirage « totalement au hasard ») et sans non-réponse, l'intervalle de confiance a une largeur qui varie (presque...) rigoureusement comme l'inverse de la racine carrée de la taille de l'échantillon: pour diviser par 2 (respectivement par 3) la largeur de l'intervalle, il faut donc un échantillon 4 fois plus gros (respectivement 9 fois plus gros). Si l'échantillonnage probabiliste est d'une autre nature (par exemple si on tire les individus avec des probabilités de sélection inégales) et si on prend en compte en sus la non-réponse, la règle doit être adaptée sur le plan mathématique mais en toute circonstance on retrouvera ce principe majeur et universel: grosso modo, quel que soit le plan de sondage, l'ordre de grandeur de la largeur de l'intervalle de confiance sera fonction de l'inverse de la racine carrée de la taille de l'échantillon répondant.

Quant aux plans de sondage empiriques visant une forme de « représentativité », c'est-à-dire en pratique correspondant aux enquêtes par quotas, le biais dépend très largement de la relation qui existe, au sein de chacune des sous-populations déterminées par les quotas, entre la variable qui permet de définir le paramètre (pour un sondage électoral, c'est une variable qui vaut 1 si l'individu déclare voter pour le candidat-cible, et 0 sinon) et la probabilité de sélection de l'individu par l'enquêteur.

L'existence de ce biais de sélection est assez intuitive et nous donnons ici un exemple un peu caricatural mais néanmoins éclairant: si on effectue une enquête sur l'emploi du temps par une méthode empirique, il y a un vrai risque de surestimation du temps moyen d'inactivité des individus. En effet, l'enquêteur qui n'a pas de consignes appropriées va plus facilement contacter des personnes qui sont plus fréquemment présentes à leur domicile, sauf à planifier de son propre chef un improbable passage sur le terrain à des horaires tardifs. L'échantillon sera probablement (plus ou moins) déséquilibré avec une sous-représentation des personnes qui travaillent beaucoup: il y a là, par nature, une corrélation positive génératrice de biais entre la probabilité de sélection et la variable « temps d'inactivité ». Ce type de biais a en particulier la très désagréable propriété de ne pas diminuer lorsque la taille de l'échantillon augmente. C'est pourquoi le sondage empirique, qui a par ailleurs des vertus indéniables dans certaines circonstances, est particulièrement sensible aux conditions de collecte et aux consignes données aux enquêteurs...c'est-à-dire au professionnalisme de la structure qui les gère !

Quant à la variance, la théorie classique des sondages ne permet pas de mener un calcul rigoureux en contexte empirique puisqu'on ne maîtrise pas les aspects probabilistes de l'échantillonnage. Très succinctement, sur le plan qualitatif, on résumera le contexte en disant que les contraintes imposées sur la structure de l'échantillon par les quotas constituent un élément qui limite à l'évidence l'ampleur de la variance – mais d'une façon que l'on ne peut pas formaliser dans l'approche classique (il y a moyen d'y échapper en modélisant le comportement des individus, mais on entre sur un terrain d'une toute autre nature). Cette question doit d'ailleurs être rapprochée des possibilités offertes par les plans de sondage probabilistes utilisant des informations externes et cela nous renvoie également à d'autres développements.



En pratique, on trouve des intervalles de confiance publiés à l'occasion de sondages empiriques, électoraux ou portant sur d'autres thèmes. Il faut être clair sur le sens de ces calculs: d'une part ils négligent le biais, d'autre part ils résultent nécessairement d'une assimilation de l'échantillonnage pratiqué à un échantillonnage probabiliste. L'auteur de ces lignes peut accepter cette façon de procéder dans certaines circonstances, mais considère qu'elle reste largement conditionnée à la maîtrise technique et opérationnelle du processus d'enquête, laquelle est par nature très dépendante du degré de professionnalisme de la structure qui en a la responsabilité.

## Référence

Ardilly, P., Les techniques de sondage, Paris, Technip, 2006.



# Climat et statistiques en débat

Liliane Bel

AgroParisTech

Le climat change-t-il vraiment ? Quels outils statistiques permettent de s'en convaincre, et d'en évaluer les conséquences ? Ce sont les questions auxquelles Julien Cattiaux<sup>2</sup> et Aurélien Ribes<sup>3</sup> du laboratoire CNRM – GAME<sup>1</sup> de Météo France ont tenté de répondre en mai dernier lors du Café de la statistique organisé à Toulouse pendant les Journées annuelles de la Société française de statistique, dans le cadre de « 2013 Mathématiques pour la planète Terre ».

Dans leur exposé initial, les deux chercheurs ont expliqué les méthodes d'étude du climat, et en quoi elles font appel à la statistique, avant d'en venir aux constats et aux prévisions.

## Méthodes

Pour comprendre et quantifier le changement climatique il faut d'abord bien comprendre le système climatique. Cela passe par la description physique de la dynamique de l'atmosphère à partir des équations de la mécanique des fluides (bilan de masse, d'énergie, de quantité de mouvement) et de la thermodynamique, ainsi que des conditions aux limites. Les équations aux dérivées partielles (Navier-Stokes entre autres) qui interviennent sont non linéaires et ne peuvent être résolues explicitement. Si la prévision déterministe du temps est inaccessible au-delà de quelques jours, sa distribution de probabilité est déterminée par de grands équilibres physiques. Ces équilibres découlent du bilan d'énergie complexe du système climatique, comprenant l'atmosphère, les océans, la cryosphère<sup>4</sup>, les surfaces continentales, la biosphère et leurs interactions.

Le climat est naturellement variable, mais des influences externes appelées forçages viennent se superposer à cette variabilité. Les forçages naturels sont dus à l'activité solaire, aux éruptions volcaniques et aux variations d'orbite (astronomique). Les forçages attribuables à l'activité humaine (« anthropiques ») sont dus entre autres aux gaz à effet de serre, aux aérosols, à l'utilisation des sols.

Pour déterminer la part de responsabilité de chaque facteur, la statistique intervient à plusieurs niveaux. L'homogénéisation des données permet de corriger les séries météorologiques des ruptures occasionnées par des changements d'instrumentation, ou de méthodologie de mesure.

La détection et l'attribution du changement résultent de l'évaluation du rôle de chaque forçage externe. La comparaison avec les climats anciens est réalisée grâce à leur reconstruction à partir de proxys<sup>5</sup>. Enfin la variabilité interne est décrite et analysée afin de la séparer de la variabilité due aux forçages externes.

## Un réchauffement atypique

Le GIEC (Groupe d'experts Intergouvernemental sur l'Évolution du Climat, ou IPCC en anglais)

a pour objectif principal d'évaluer l'information scientifique et socio-économique sur le changement climatique, ses impacts et les différentes options pour l'atténuer ou s'y adapter. Ses conclusions indiquent que le réchauffement du système climatique est sans équivoque, car il ressort désormais des observations de l'augmentation des températures moyennes mondiales de l'atmosphère et de l'océan, de la fonte généralisée des neiges et des glaces et de l'élévation du niveau moyen mondial de la mer, entre autres.

L'essentiel de l'accroissement observé sur la température moyenne globale depuis le milieu du XX<sup>ème</sup> siècle est très probablement dû à l'augmentation observée des concentrations des gaz à effet de serre anthropiques.

Les informations paléo climatiques confirment l'interprétation que le réchauffement du dernier demi-siècle est atypique sur au moins les 1300 dernières années. La dernière fois que les régions polaires ont été significativement plus chaudes qu'actuellement pendant une longue durée (il y a environ 125 000 ans), la réduction du volume des glaces polaires a conduit à une élévation du niveau des mers de 4 à 6 mètres.

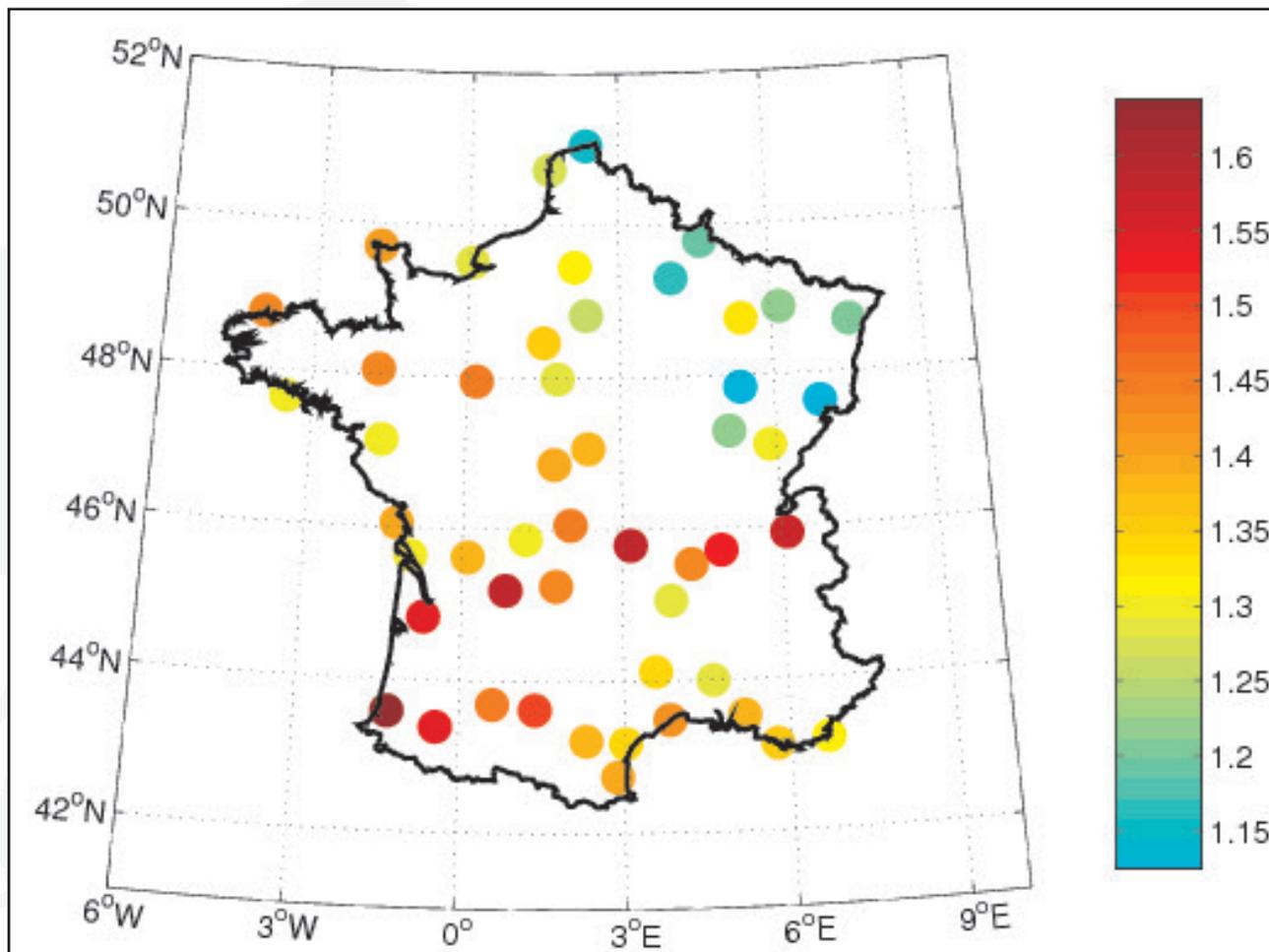
Par ailleurs de nombreux autres composants du système : glaciers de montagne, enneigement en montagne, contenu de chaleur et salinité de l'océan, cours d'eau (débits, régimes, extrêmes) montrent également des évolutions significatives. Par contre certaines caractéristiques du climat comme les tempêtes ou les épisodes de précipitations intenses en France ne montrent pas d'évolution significative pour le moment.

## **Dont la poursuite est prévisible**

Suivant le scénario envisagé pour le futur, qui dépend d'hypothèses sur le type de croissance et les énergies utilisées, les émissions de gaz à effet de serre mènent à des projections sur l'augmentation de la température moyenne globale allant de 1,8° à 4°. La répartition spatiale est inégale, la zone subissant le plus fort réchauffement étant l'hémisphère nord, et plus particulièrement l'Arctique. Des effets sur la biodiversité sont attendus avec la disparition de nombreuses espèces animales et végétales. Enfin, de nombreux impacts sur nos sociétés sont envisagés, par exemple la baisse possible des rendements agricoles.

## **Malgré ce mois de mai à Toulouse ?**

Les participants aux Journées de statistique de Toulouse se rappelleront longtemps les conditions météorologiques quasi extrêmes qui ont régné pendant la semaine du 27 au 31 mai : un froid vif et des trombes d'eau, qui ont d'ailleurs favorisé une affluence record aux exposés. Interrogés là-dessus, les intervenants ont présenté des graphiques représentant la température moyenne et le cumul des précipitations rapportés à une période de référence 1981-2010. Ces graphiques ont montré que ce mois de mai, et plus globalement le printemps 2013 à Toulouse et en France ont présenté des niveaux peu habituels. Alors qu'on observe des moyennes de températures mensuelles et annuelles généralement supérieures à celle de la période de référence (1981-2010) depuis 1986, le printemps 2013 est une des rares périodes à avoir une température nettement inférieure, température toutefois assez habituelle dans les années 1970, par exemple. Ce phénomène est purement régional, car si on regarde l'écart de température entre le printemps 2013 et la même période de référence à l'échelle du globe, on s'aperçoit qu'il est majoritairement positif. L'analyse de la température globale à l'échelle de temps annuelle et décennale montre que la tendance continue à être au réchauffement avec une variabilité marquée.



**Graphique :** augmentation de température moyenne estimée entre 1900 et 2006 en °C  
 Source : [1].

## Un débat animé

Le Café a donné lieu à de nombreux échanges avec la salle. Parmi les questions posées se trouvaient tout d'abord des questions « d'actualité » météorologique (ex : sur le moi de mai particulièrement frais), ou climatique (ex : sur la réalité de la stabilisation apparente du réchauffement moyen en surface au cours des dernières années). D'autres questions portaient sur les changements de long terme, les constantes de temps en jeu, le temps de « retour à l'équilibre » du système climatique (retour au climat de la période préindustrielle), ou les impacts sur les sociétés humaines, notamment le caractère bénéfique ou préjudiciable du réchauffement attendu. Ce changement est-il inéluctable ? Existe-t-il une ingénierie pour remédier au changement climatique ? Quelle est l'inertie du système : si on arrêta d'émettre maintenant, combien de temps faudrait-il pour perdre le 1,5 ° gagné ?

Bien que le sujet soit très empreint de sciences physiques, plusieurs questions ont abordé les interactions avec les statistiques, en particulier sur la « bonne » définition des normales climatiques, dans un climat qui n'est plus stationnaire (et sur l'opportunité de calculer des normales non stationnaires). Enfin, plusieurs questions plus polémiques ont porté sur l'évaluation du rôle de l'homme dans ces changements, et la possibilité d'une explication naturelle du réchauffement récent. La discussion s'est alors portée sur les grandes variations climatiques du dernier million d'années (alternance de périodes glaciaires et interglaciaires), et sur les différences avec les changements en cours.

## Références

1. GAME (Groupe d'étude de l'atmosphère Météorologique), laboratoire de recherche associé au CNRS (URA 1357) et faisant partie du CNRM (Centre National de Recherches Météorologiques) de Météo-France.
  2. Page Web : <http://www.cnrm-game.fr/spip.php?article629>
  3. Page Web : <http://www.cnrm.meteo.fr/gmgec/spip.php?article105>
  4. Cryosphère : terme désignant les portions de la surface de la Terre où l'eau est présente à l'état solide (banquises, glaciers, sols gelés, etc.)
  5. Dans l'étude des climats du passé, les « proxys climatiques » sont des caractéristiques du passé qui ont été conservées et qui sont utilisées pour reconstruire des variables climatiques anciennes. Par exemple, la largeur des cernes du tronc des arbres peut servir pour la reconstitution des températures anciennes.
- [1] Ribes A, J-M Azaïs, S Planton (2010) A method for regional climate change detection using smooth temporal patterns, *Climate Dynamics*, 35 (2-3), pages 391-406.
- [2] Voir aussi les articles sur le sujet sur le blog « un jour une brève » de « Mathématiques pour la planète Terre - 2013 » :
- <http://mpt2013.fr/pourquoi-corriger-les-series-climatiques/>  
<http://mpt2013.fr/le-climat-de-chez-vous-a-t-il-deja-change/>  
<http://mpt2013.fr/pourquoi-grelotte-t-on-alors-que-la-planete-se-rechauffe/>  
<http://mpt2013.fr/reconstructions-du-climat-du-dernier-millenaire/>  
<http://mpt2013.fr/les-cernes-dhier/>  
<http://mpt2013.fr/echauffement-climatique-ou-fin-dune-periode-glaciaire/>  
<http://mpt2013.fr/quelle-esperance-de-vie-pour-le-plus-grand-glacier-des-alpes/>  
<http://mpt2013.fr/dater-la-fonte-des-glaciers/>