

7ème Colloque francophone sur les sondages
Rennes, 5-7 novembre 2012

QUELQUES SUGGESTIONS D'APPLICATIONS AUX ENQUÊTES ÉLECTORALES DE MÉTHODES UTILISÉES DANS LES ENQUÊTES MARKETING - EXEMPLES DE LA PRÉSIDENTIELLE 2012 -

Daniel Bachelet

Conseil en marketing
et expert auprès de la Commission des Sondages

Les opinions et propositions de cet article sont de la seule responsabilité de
l'auteur et n'engagent en rien la Commission des Sondages.

Introduction

Cet exposé a pour objectif de démontrer que si la spécificité des enquêtes électorales ne doit pas être sous-estimée, l'utilisation de beaucoup de méthodes mises au point pour l'élaboration et l'exploitation des enquêtes marketing, notamment pour les biens durables, permettrait de faire des progrès importants pour la qualité des intentions de vote et pour leur compréhension. Le sujet étant vaste, trois domaines d'application seront présentés, en commençant par résumer les méthodes utilisées dans les enquêtes marketing, puis en décrivant les possibilités pour les enquêtes électorales qui en résultent, avec pour chaque domaine un exemple d'application à la présidentielle France 2012 n'utilisant que des données publiées :

- **La mesure de la concurrence, le modèle « cannibal » et l'estimation des transferts d'intentions de vote**
- **La gestion des « non-répondants partiels »**
- **Le choix des critères de redressement**

1 La mesure de la concurrence, le modèle « cannibal » et l'estimation des transferts d'intentions de vote

1.1 Les mesures d'intensité de la concurrence dans les enquêtes marketing

Dès les années 1980, des travaux ont permis de définir « l'ensemble de considération » (« evoked set », voir la référence RAM dans le résumé) comme « le sous ensemble des marques qu'un consommateur envisage d'acheter parmi l'ensemble des marques qu'il connaît dans une classe de produit donné ». Des mesures de l'intensité de la concurrence entre marques par des indices de type «Odd ratio» indépendants du volume de vente et de la notoriété des marques ont alors été proposées. Rappelons la définition de l'Odd ratio :

Odd = KN/LM variant entre 0 et l'infini.

Nous préférons :

ODB = $KN/(KN+LM)$

Indice d'intensité de concurrence de B sur A	Marque considérée B	Autres marques
Marque achetée A	K	L
Autres marques achetées	M	N

ODB = 0 : aucune concurrence ($K=0$) ; ODB = 1 : concurrence totale (L ou $M=0$).

ODB = 0.5 : concurrence au prorata du marché ($K / K+M = K+L / K+M+L+N$).

Dans l'automobile, la « considération » est mesurée depuis longtemps par le modèle qui aurait été acheté si celui acheté n'avait pas existé (modèle de substitution). Les indices de concurrence servent à PSA à évaluer si les modèles Peugeot et Citroën ne se concurrencent pas trop et à mettre au point les segmentations produit (ex récents Brésil et Chine).

1.2 Les mesures du cannibalisme dans les enquêtes marketing

Pour les FMCG (Fast Moving Consumer Good), un tri sur des données de panel entre deux périodes suffit à déterminer l'origine des gains et des pertes de parts de marché d'une marque. Pour tous les autres produits, dont l'automobile, ce n'est pas possible, les durées de conservation du produit étant trop longues et surtout trop variables d'un utilisateur à l'autre.

La mesure de la concurrence en terme de volume, notamment entre les modèles d'une même marque ou d'un même groupe (cannibalisme) est cependant indispensable pour définir les politiques de gamme et d'investissement, les ventes en cannibalisme ne permettant pas de rentabiliser l'investissement, quelque soit le volume de vente.

C'est pourquoi la méthode « cannibal » a été mise au point et est utilisée depuis plus de 20 ans par PSA Peugeot Citroën (1^{ère} note interne en 1986). Elle consiste, en se basant sur les modèles de substitution dans les enquêtes disponibles, à en déduire des proximités entre modèles et à effectuer un redressement classique par itération sur les parts de marché en t et t+1 pour calculer des pseudo transferts entre les 2 périodes aussi proches que possibles des substitutions citées, mais respectant les parts de marché.

1.3 L'estimation des transferts d'intentions de vote en cours de campagne électorale et entre le dernier sondage et le scrutin

Dans les enquêtes électorales, les transferts d'intentions de vote présentent beaucoup d'intérêt. Mais la détermination des transferts en utilisant un panel, comme pour les achats des FMCG, n'est pas à recommander à notre avis. En effet, le risque d'un «biais de panel», dit de « conditionnement », est important. Il résulte d'un changement d'attitude et de comportement induit par l'interview répétée des individus sur le même sujet (voir Market - Etudes et Recherches en marketing, page 191- Nathan – Y. Evrard, B. Pras, E. Roux). Dans les panels d'achat, les nouveaux panélistes ne sont d'ailleurs pris en compte qu'après quelques mois. Pour la même raison, à notre connaissance, des panels ne sont pas utilisés pour des baromètres d'image ou d'opinion.

Heureusement, les questionnaires des baromètres d'intentions de vote comportent souvent une question sur la sûreté du choix et l'Ifop a demandé aux indécis un choix en second tout à fait utilisable comme mesure de la « considération » et qui permet de mettre en œuvre la méthode « cannibal » s'il y a suffisamment d'indécis. Il y en a 30 % dans le dernier baromètre de l'Ifop avant le scrutin, soit 540 interviews d'indécis seulement, ce qui nous a conduit à utiliser aussi le baromètre précédent.

A titre d'exemple, nous avons estimé les transferts entre la moyenne des dernières intentions de vote des instituts et le résultat du 1^{er} tour en utilisant la méthode « cannibal » basée sur les sûrs de leur choix et les choix en second des indécis.

1.4 L'estimation des transferts d'intentions de vote entre le scrutin et le dernier sondage d'intentions de vote : dernier sondage publié

Dernier sondage publié	Ipsos	CSA	TNS	LH2	BVA	Ifop	OW	HI	Moy E	Moy téléph	Moy internet	résultats déf R	Ecart R - E
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	18-19 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	0,1
P. Poutou	1,5	1,5	1	1	1,5	1	2	1,5	1,4	1,3	1,5	1,2	-0,2
JL Mélenchon	14	14,5	13	15	14	13,5	13	12	13,6	14,1	12,8	11,1	-2,5
F. Hollande	29	28	27	27	30	27	27,5	27,5	27,9	28,2	27,3	28,6	0,8
E. Joly	2	2	3	2,5	2	3	2	3	2,4	2,3	2,7	2,3	-0,1
F. Bayrou	10	10,5	10	10	10	10,5	10	11	10,3	10,1	10,5	9,1	-1,1
N. Sarkozy	25,5	25	27	26,5	26,5	27	27,5	26,5	26,4	26,1	27,0	27,2	0,7
N. Dupont-Aignan	1,5	1,5	2	1,5	2	1,5	1,5	2	1,7	1,7	1,7	1,8	0,1
M. Le Pen	16	16	17	15,5	14	16	16	16	15,8	15,7	16,0	17,9	2,1
J. Cheminade	0,5	0	0	0	0	0	0	0	0,1	0,1	0,0	0,3	0,2
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	0,0
Précision (moy des écarts)	0,95	1,11	0,73	1,08	1,14	0,89	0,79	0,80	0,80	0,82	0,76		
	Téléph.	Téléph.	Téléph.	Téléph.	Téléph + Int	Internet	Internet	Internet					

1.4 L'estimation des transferts d'intentions de vote entre le scrutin et le dernier sondage d'intentions de vote : les sûrs et les indécis

Les choix en second	N. Arthaud	P. Poutou	JL Mélenchon	F. Hollande	E. Joly	F. Bayrou	N. Sarkozy	N. Dupont-Aignan	M. Le Pen	J. Cheminade	ABS-BN	T	% dernier Sondage
N. Arthaud	0,52	0,12	0,25	0,05	0,02	0,00	0,00	0,00	0,00	0,00	0,04	1,00	0,3
P. Poutou	0,12	0,55	0,20	0,05	0,02	0,00	0,02	0,00	0,00	0,00	0,04	1,00	1,1
JL Mélenchon	0,02	0,03	0,65	0,16	0,03	0,04	0	0	0,03	0	0,04	1,00	10,6
F. Hollande	0,01	0,01	0,08	0,78	0,02	0,05	0,01	0,005	0	0,005	0,03	1,00	21,7
E. Joly	0,005	0,005	0,08	0,15	0,55	0,13	0	0,005	0,02	0,005	0,05	1,00	1,9
F. Bayrou	0	0	0,04	0,15	0,01	0,54	0,15	0,01	0,02	0,01	0,07	1,00	8,0
N. Sarkozy	0	0	0	0,02	0	0,05	0,87	0,005	0,02	0,005	0,03	1,00	20,6
N. Dupont-Aignan	0,00	0,00	0,00	0,10	0,00	0,05	0,10	0,67	0,01	0,01	0,06	1,00	1,3
M. Le Pen	0	0	0,03	0,03	0,02	0,01	0,08	0,02	0,77	0,01	0,03	1,00	12,3
J. Cheminade	0,00	0,00	0,01	0,12	0,01	0,05	0,12	0,05	0,03	0,51	0,10	1,00	0,0
Abstention-nuls	0,00	0,01	0,03	0,08	0,01	0,03	0,08	0,01	0,05	0,00	0,70	1,00	22,25*
Total résultats déf	0,4	0,9	8,7	22,3	1,8	7,1	21,2	1,4	14,0	0,2	22,0	100,0	100,0
Total dernier sondage	0,3	1,1	10,6	21,7	1,9	8,0	20,6	1,3	12,3	0,0	22,3	100,0	
Ecart	0,1	- 0,2	- 1,9	0,7	- 0,1	- 0,8	0,6	0,1	1,7	0,1	- 0,3	0,00	
Source Ifop baromètre 12-15 avril vague 8 - 1808 interviews -							Prise en compte du baromètre Ifop 5-7 avril vague 7 - 2022 interviews -						
Répartitions estimées du poste "autres candidats" de l'ifop							Estimations empiriques			*ABS : moy des estimations Ipsos-TNS-OW-BVA			

1.4 L'estimation des transferts d'intentions de vote entre le scrutin et le dernier sondage d'intentions de vote : résultats

Transferts estimés entre S et R	N. Arthaud	P. Poutou	JL Mélenchon	F. Hollande	E. Joly	F. Bayrou	N. Sarkozy	N. Dupont-Aignan	M. Le Pen	J. Cheminade	Abst-nuls	Total estim S	Ecart estim / dernier sondage
N. Arthaud	0,1	0,0	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,3	0,0
P. Poutou	0,1	0,5	0,3	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,1	1,1	0,0
JL Mélenchon	0,1	0,2	6,1	1,9	0,3	0,5	0,0	0,0	0,7	0,0	0,8	10,6	0,0
F. Hollande	0,1	0,1	1,4	17,1	0,3	1,1	0,2	0,1	0,0	0,1	1,1	21,6	-0,1
E. Joly	0,0	0,0	0,1	0,3	0,9	0,3	0,0	0,0	0,1	0,0	0,2	1,9	0,0
F. Bayrou	0,0	0,0	0,2	1,1	0,1	3,9	1,3	0,1	0,3	0,0	0,9	8,0	0,0
N. Sarkozy	0,0	0,0	0,0	0,4	0,0	0,9	17,6	0,1	0,7	0,0	0,9	20,6	0,0
N. Dupont-Aignan	0,0	0,0	0,0	0,1	0,0	0,1	0,1	0,8	0,0	0,0	0,1	1,3	0,0
M. Le Pen	0,0	0,0	0,2	0,2	0,1	0,1	0,6	0,1	10,7	0,0	0,4	12,3	0,0
J. Cheminade	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Abstention-nuls	0,0	0,1	0,3	1,1	0,1	0,4	1,3	0,1	1,4	0,0	17,3	22,3	0,0
Résultats déf R	0,4	0,9	8,7	22,3	1,8	7,1	21,2	1,4	14,0	0,2	22,0	100,0	E-T 0,028

1.4 L'estimation des transferts d'intentions de vote entre le scrutin et le dernier sondage d'intentions de vote : les soldes des transferts

Soldes des transferts estimés entre S et R	N. Arthaud	P. Poutou	JL Mélenchon	F. Hollande	E. Joly	F. Bayrou	N. Sarkozy	N. Dupont-Aignan	M. Le Pen	J. Cheminade	ABS-BN	Total solde calculé S - R	Total solde exact S - R
N. Arthaud	0,0	-0,1	0,0	-0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-0,1	-0,1
P. Poutou	0,1	0,0	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,2	0,2
JL Mélenchon	0,0	-0,1	0,0	0,5	0,1	0,2	0,0	0,0	0,6	0,0	0,5	1,9	1,9
F. Hollande	0,1	0,0	-0,5	0,0	0,0	-0,1	-0,1	0,0	-0,2	0,0	0,0	-0,7	-0,7
E. Joly	0,0	0,0	-0,1	0,0	0,0	0,2	0,0	0,0	0,0	0,0	0,1	0,1	0,1
F. Bayrou	0,0	0,0	-0,2	0,1	-0,2	0,0	0,4	0,0	0,2	0,0	0,5	0,8	0,8
N. Sarkozy	0,0	0,0	0,0	0,1	0,0	-0,4	0,0	-0,1	0,1	0,0	-0,4	-0,6	-0,6
N. Dupont-Aignan	0,0	0,0	0,0	0,0	0,0	0,0	0,1	0,0	-0,1	0,0	0,0	-0,1	-0,1
M. Le Pen	0,0	0,0	-0,6	0,2	0,0	-0,2	-0,1	0,1	0,0	0,0	-1,1	-1,6	-1,7
J. Cheminade	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-0,1	-0,1
Abstention-nuls	0,0	0,0	-0,5	0,0	-0,1	-0,5	0,4	0,0	1,1	0,0	0,0	0,3	0,3
Total solde estimé R - S	0,1	-0,2	-1,9	0,7	-0,1	-0,8	0,6	0,1	1,6	0,1	-0,3	0,0	0,0
Total solde exact R - S	0,1	-0,2	-1,9	0,7	-0,1	-0,8	0,6	0,1	1,7	0,1	-0,3	0,0	

1.4 L'estimation des transferts d'intentions de vote entre le scrutin et le dernier sondage d'intentions de vote : conclusions

- Pour J.L. Mélenchon, recul de 1.9% des inscrits résultant en très grande partie par des pertes de 0.5% en faveur de F. Hollande (vote utile au dernier moment), de 0.6% pour M. Le Pen et de 0.5% pour les abstentionnistes.
- Pour M. Le Pen (écart de + 1.7%), gain de 1.1% sur les abstentionnistes et, plus surprenant, de 0.6% sur J.L. Mélenchon. Ce sont les nouveaux venus au candidat du FN qui ont été sous estimés (refus de répondre et / ou fausse réponse), les fidèles étant « redressés » par le souvenir de vote. Ce redressement suffit en cas de stabilité ou de décroissance du FN (cas en 2007), mais le sous-estime en cas de croissance sensible (2002 et 2012).
- Pour F. Bayrou (écart de - 0.8%), perte de 0.4% en faveur de N. Sarkozy (vote utile au dernier moment) et de 0.5% sur les abstentionnistes.
- Pour F. Hollande (écart de + 0.7%), gains de 0.5% sur J.L. Mélenchon (vote utile) et de 0.2% sur M. Le Pen.
- Pour N. Sarkozy (écart de + 0.6%), gains de 0.4% sur F. Bayrou (vote utile) et de 0.4% sur les abstentionnistes (mobilisation ou biais).

1.5 Les indices d'intensité de la concurrence entre les candidats : résultats (hors sûrs ou fidèles)

Indice de concurrence basé sur le choix en second	JL Mélenchon	F. Hollande	E. Joly	F. Bayrou	N. Sarkozy	M. Le Pen
JL Mélenchon	1	0,86	0,57	0,36	0	0,49
F. Hollande	0,78	1	0,64	0,65	0,41	0,32
E. Joly	0,67	0,70	1	0,00	0	0,66
F. Bayrou	0,42	0,69	0,71	1	0,83	0,20
N. Sarkozy	0	0,17	0	0,75	1	0,76
M. Le Pen	0,53	0,00	0,35	0,33	0,71	1
Indice de concurrence avec les transferts	JL Mélenchon	F. Hollande	E. Joly	F. Bayrou	N. Sarkozy	M. Le Pen
JL Mélenchon	1	0,87	0,55	0,33	0	0,48
F. Hollande	0,79	1	0,64	0,61	0,36	0,33
E. Joly	0,67	0,73	1	0,00	0	0,65
F. Bayrou	0,46	0,72	0,70	1	0,77	0,21
N. Sarkozy	0	0,23	0	0,79	1	0,80
M. Le Pen	0,59	0,00	0,38	0,34	0,70	1,00

2 La gestion des « non-répondants partiels »

2.1 Les non-répondants partiels dans les enquêtes marketing

Pour les tableaux de résultats, l'usage le plus répandu est de les présenter sur les répondants, malgré le risque de biais. Mais, dès qu'il y a des analyses de données à faire, les données manquantes sont nécessairement remplacées. Des méthodes automatiques bien connues des statisticiens sont alors utilisées lorsque les taux de non-répondants (NR) sont inférieurs à 10 %. Citons à titre d'exemples pour une enquête européenne de PSA sur les attentes automobiles la méthode du plus proche voisin d'Eole à l'intérieur des acheteurs du même modèle et pour les enquêtes internationales inter constructeurs NCBS (New Car Buyer Survey) la méthode des chaînes de Markov dans SAS mise en œuvre par Socio logiciel pour PSA.

Mais lorsqu'une question a un taux de NR de plus de 10 % sur un sujet important, nous utilisons des méthodes explicatives comme l'analyse discriminante puisqu'il s'agit le plus souvent de variables discrètes. Par exemple, la question NCBS au Brésil sur la cylindrée de l'autre voiture du foyer des couples multi-équipés avec 24 % de NR sur les 12000 individus concernés, question utilisée pour distinguer la voiture principale de la seconde voiture pour laquelle les attentes sont très différentes. Avec un taux de « bien classés » de 66 %, la proportion de seconde voiture a été estimée à 43 %, pour 27 % par une chaîne de Markov dont les résultats étaient moins discriminants des segments produit.

Des méthodes spécifiques doivent parfois être mises au point. C'est le cas du prix payé à l'achat d'un véhicule neuf pour lequel il faut non seulement estimer une valeur pour chaque NR, mais aussi corriger les valeurs aberrantes. Une procédure sophistiquée et automatisée a été mise au point en utilisant les médianes et quartiles au niveau de la version motorisation-carrosserie du modèle lorsque l'effectif le permettait (8 répondants ou plus), ou sinon au niveau modèle-carrosserie, ou sinon au niveau du segment produit.

2.2 Les non-répondants partiels dans les enquêtes électorales

Dès que les taux de non-répondants pour les « certains d'aller voter » (« NR ») ne sont pas négligeables (> à 5 ou 10 %), une analyse discriminante (« AD ») bayésienne des répondants (autant de modalités que de candidats, plus ABS et BN) en fonction des autres questions de l'enquête posées régulièrement et susceptibles d'être corrélées pourrait permettre de déterminer leurs affectations de manière plus pertinente que de les répartir au prorata des répondants.

- Pour les intentions de vote au 1^{er} tour de la présidentielle (plus de 10 % de NR dans les enquêtes téléphoniques), les questions de souvenirs de vote disponibles, de proximité partisane et sociodémographiques détaillées sont à tester comme variables explicatives de l'AD (y compris les modalités NR de ces questions).
- De même, pour les souvenirs de vote susceptibles d'être utilisés pour les redressements du 1^{er} tour de la présidentielle, comme le 1^{er} tour de la présidentielle 2007 (environ 10 % de NR) et comme les régionales 2010 et les européennes 2009 (beaucoup plus de NR), les mêmes questions sont à utiliser.
- Pour les intentions de vote au 2^{ème} tour de la présidentielle (environ 15 % de NR selon Ipsos, LH2 et TNS), le souvenir de vote au 1^{er} tour est de loin la question la plus explicative du choix au second tour et il est alors possible de tenir compte de cette information pour estimer les intentions de vote des NR, sans avoir besoin des données individuelles .

2.2 Les non-répondants partiels dans les enquêtes électorales : exemple du 2^{ème} tour de la présidentielle 2012

Les non exprimés ayant déclaré avoir l'intention de voter (NR, 15 % des interviewés) ont de fortes chances d'arbitrer entre les 2 candidats comme les interviewés ayant voté pour le même candidat au 1^{er} tour, sauf à voter blanc. Or, F. Bayrou et M. Le Pen constituent environ 70 % des NR en dehors des 22 % d'ABS-BN du 1^{er} tour et les interviewés ayant choisi ces candidats ont des reports de voix plus élevés en faveur de N. Sarkozy que pour F. Hollande.

Avec un redressement classique sur le 1^{er} tour, les NR sont répartis comme les ABS-BN au prorata du total des intentions de vote exprimées pour les 2 candidats, moins nombreuses pour N. Sarkozy, ce qui conduit tous les instituts à un biais en faveur de F. Hollande.

En utilisant les moyennes des données de report entre le 1^{er} et le 2^{ème} tour publiées par les instituts, les tableaux ci-après présentent :

- le résultat obtenu par le redressement classique sur le 1^{er} tour, avec les « SIV » (SIV : n'a pas exprimé d'intention de vote pour F. Hollande ou N. Sarkozy) exclus.
- celui que nous avons calculé la veille du scrutin, par une « affectation des SIV » en précisant que si le principe de la réaffectation des SIV était pertinent, il n'était pas « justifiable » de les réaffecter tous, mais seulement ceux certains d'aller voter (les NR), données que nous n'avions pas.
- celui obtenu par une « affectation des NR », les NR étant estimés faute de mieux par les SIV déflatés du taux d'abstentionnistes (« ABS ») du candidat.

2.2 Les non-répondants partiels dans les enquêtes électorales : derniers sondages et résultats du 2^{ème} tour de la présidentielle

Rappel des taux d'abstention : Ipsos 18 %, TNS 20 %

Résultat : ABS 19,6 % - BN 4,7 % - ABS+BN 24,3 %

2ème tour dernier sondage	Ipsos	CSA	TNS	BVA	LH2	Ifop	OW	HI	Moy 8 instituts	Estim. aff. des SIV	Estim. aff. des NR	Résultat
Date publication	04-mai	04-mai	04-mai	04-mai	03-mai	04-mai	03-mai	04-mai	04 mai	05-mai	05-oct	06-mai
Date terrain	03-mai	03-mai	03-mai	03-mai	27av-2mai	1-4mai	2-3 mai	2-3 mai				
F. Hollande	52,5	53,0	53,5	52,5	53,0	52,0	52,5	53,0	52,8	51,9	51,5	51,6
N. Sarkozy	47,5	47,0	46,5	47,5	47,0	48,0	47,5	47,0	47,2	48,1	48,5	48,4
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
Ecarts aux résultats	Ipsos	CSA	TNS	BVA	LH2	Ifop	OW	HI	Moy 8 instituts	Estim. aff. des SIV	Estim. aff. des NR	Résultat
F. Hollande	0,9	1,4	1,9	0,9	1,4	0,4	0,9	1,4	1,12	0,25	-0,10	0,00
N. Sarkozy	-0,9	-1,4	-1,9	-0,9	-1,4	-0,4	-0,9	-1,4	-1,12	-0,25	+0,10	0,00

2.2 Les non-répondants partiels dans les enquêtes électorales

En vert, moyenne des reports publiés par les instituts : 8 pour JL M, F B, M LP, 6 pour FH, NS, 1 pour E J et 3 pour les ABS. Pour les petits candidats en gris, reports estimées empiriquement.

- SIV : n'a pas exprimé d'intention de vote, ABS : abstentionnistes, BN : blancs et nuls -

Moy. 8 instituts	1er tour	FH	NS	SIV	T	ABS-BN	FH	NS	SIV
N. Arthaud	0,4	72	7	21	100	26	0,3	0,0	0,1
P. Poutou	0,9	72	7	21	100	26	0,6	0,1	0,2
J-L. Mélenchon	8,7	84	5	11	100	26	7,3	0,4	1,0
F. Hollande	22,3	98	1	2	100	11	21,8	0,2	0,3
E. Joly	1,8	60	9	31	100	16	1,1	0,2	0,6
F. Bayrou	7,1	34	37	29	100	16	2,4	2,6	2,0
N. Sarkozy	21,2	1	97	3	100	8	0,2	20,5	0,5
N. Dupont-Aignan	1,4	35	35	30	100	18	0,5	0,5	0,4
M. Le Pen	14,0	18	54	28	100	26	2,5	7,6	3,9
J. Cheminade	0,2	35	35	30	100	18	0,1	0,1	0,1
Abstention-BN	22,0	29	30	41	100	27	6,4	6,6	9,0
Total	100,0					18	43,1	38,7	18,1
Résultats redr classique							52,8	47,2	

2.2 Les non-répondants partiels dans les enquêtes électorales

Pour chaque candidat du 1^{er} tour, tous les SIV du 2^{ème} tour sont affectés à FH ou NS au prorata des exprimés, ce qui suppose implicitement que les reports indiqués par les instituts ne concernent que les « certains d'aller voter », alors que ce n'est pas le cas.

Affectation SIV	1er tour	FH	NS	SIV	T	FH	NS
N. Arthaud	0,4	72	7	21	100	0,4	0,0
P. Poutou	0,9	72	7	21	100	0,8	0,1
J-L. Mélenchon	8,7	84	5	11	100	8,2	0,5
F. Hollande	22,3	98	1	2	100	22,1	0,2
E. Joly	1,8	60	9	31	100	1,6	0,2
F. Bayrou	7,1	34	37	29	100	3,4	3,7
N. Sarkozy	21,2	1	97	3	100	0,2	21,0
N. Dupont-Aignan	1,4	35	35	30	100	0,7	0,7
M. Le Pen	14,0	18	54	28	100	3,4	10,5
J. Cheminade	0,2	35	35	30	100	0,1	0,1
Abstention-BN	22,0	29	30	41	100	10,8	11,2
Résultats aff. SIV	100,0	43,1	38,7	18,1		51,7	48,3

2.2 Les non-répondants partiels dans les enquêtes électorales

Le tri des SIV par les « certains d'aller voter » pour obtenir les NR n'étant pas disponible, ceux-ci ont été répartis au prorata des reports indiqués par les instituts, les SIV des votants au 1^{er} tour étant considérés comme NR, les ABS-NB du 1^{er} tour comme ABS-NB au 2^{ème} tour.

Affectation des NR	1er tour	FH	NS	NR	T	ABS-BN	FH	NS	ABS-BN
N. Arthaud	0,4	53	5	16	74	26	0,3	0,0	0,1
P. Poutou	0,9	53	5	16	74	26	0,6	0,1	0,2
J-L. Mélenchon	8,7	62	3	8	74	26	6,1	0,3	2,3
F. Hollande	22,3	87	1	1	89	11	19,7	0,2	2,5
E. Joly	1,8	50	8	26	84	16	1,3	0,2	0,3
F. Bayrou	7,1	29	31	24	84	16	2,9	3,1	1,1
N. Sarkozy	21,2	1	89	2	92	8	0,2	19,3	1,7
N. Dupont-Aignan	1,4	29	29	25	82	18	0,6	0,6	0,3
M. Le Pen	14,0	13	40	21	74	26	2,5	7,8	3,6
J. Cheminade	0,2	29	29	25	82	18	0,1	0,1	0,0
Abstention-BN	22,0	21	22	30	73	27	4,7	4,8	12,5
Total	100,0						38,8	36,5	24,6
Résultats aff. NR							51,5	48,5	

3 Le choix des critères de redressement

3.1 Le choix des critères de redressement dans les enquêtes marketing

Dans les enquêtes marketing par téléphone, par internet ou par la poste, malgré le tirage aléatoire des interviewés, la qualité discutable des bases de sondage et les proportions importantes de non-répondants à l'enquête rendent indispensables un redressement de l'échantillon pour en améliorer la représentativité.

Améliorer la représentativité signifie réduire les biais sur la question objectif principal de l'enquête. Redresser signifie pondérer les interviewés sur les critères disponibles dans l'enquête les plus discriminants de cette question et connus sans biais par ailleurs.

Pour choisir les critères les plus efficaces, il suffit d'effectuer une analyse discriminante pas à pas des modalités de la question objectif de l'enquête en fonction des critères discriminants des candidats.

A titre d'exemple, dans des enquêtes sur la fidélité des automobilistes pour leurs opérations après vente au garage de la marque automobile achetée, Ipsos a effectué à notre demande des analyses discriminantes dans plus de 20 pays en fonction d'une demi-douzaine de critères précédemment utilisés pour n'en retenir que 2 en optimisant leurs modalités (âge du véhicule et VN+VO marque / VO concurrence).

3.2 Le choix des critères de redressement dans les enquêtes électorales

Les instituts effectuent tous un redressement sur les 5 critères sociodémographiques utilisés dans les enquêtes « omnibus » : sexe, âge, profession du chef de famille ou de l'interviewé, région et catégorie d'agglomération. Les médias peuvent ainsi indiquer que « **l'échantillon est représentatif de la population française de 18 ans et plus** », phrase ambiguë puisque l'échantillon n'est alors représentatif que sur les 5 critères ayant fait l'objet du redressement et pas nécessairement sur d'autres critères comme l'opinion politique.

Un redressement pour être représentatif au plan politique est indispensable, non seulement parce que les NR à l'enquête n'ont aucune raison d'avoir les mêmes opinions politiques que les répondants, mais aussi en raison de fausses réponses éventuelles sur des sujets sensibles comme le vote FN. C'est pourquoi, pour les sondages d'intentions de vote, les instituts utilisent comme critère de redressement, en plus des critères sociodémographiques, un ou plusieurs souvenirs de vote choisis a priori (élection de même type, élection plus récente, ...). A notre connaissance, le choix de ces critères n'a pas fait l'objet d'une AD pas à pas de la question d'intentions de vote pour rechercher les plus discriminants, AD qu'il n'est pas possible de réaliser sans les fichiers au niveau individuel. Les tableaux disponibles permettent en revanche de calculer le pouvoir discriminant (variance expliquée ou inter candidat / variance totale) de chaque critère utilisé, ou qu'il serait envisageable d'utiliser.

3.2 Choix des critères de redressement dans les enquêtes électorales : exemples de calculs de la variance expliquée

Ipsos - Baromètre d'intentions de vote pour l'élection présidentielle - Vague 18 - 19 et 20 avril 2012 – « Certains d'aller voter" et exprimés

	P. Poutou	JL Mélenchon	F. Hollande	E. Joly	F Bayrou	N. Sarkozy	Dupont-Aignan	M. Le Pen	autres	Total	V inter / V totale %
Int. Vote 1er tour	1,5	13,9	28,9	2,0	9,9	25,6	1,5	15,9	0,8	100,0	

Age

18 - 34 ans	61	32	22	15	23	24	8	30	0	25	2,4
35 - 44 ans	17	23	20	15	14	12	9	12	39	16	1,8
45 - 59 ans	22	22	27	57	27	21	50	38	10	27	2,9
60 ans et plus	0	23	31	12	36	43	33	20	51	31	4,3
Total	100	100	100	100	100	100	100	100	100	100	3,0

Présid. 2007 1er tour

EXG PC Verts	53	48	6	63	3	0	0	1	34	11	31,1
S. Royal	0	32	71	8	1	2	11	3	16	27	47,1
F. Bayrou	10	12	17	21	79	7	48	4	19	19	30,4
N. Sarkozy	9	4	5	8	15	89	41	35	0	32	55,2
J.M. Le Pen	28	3	2	0	3	2	0	58	31	11	40,5
Total	100	100	100	100	100	100	100	100	100	100	43,3

3.2 Choix des critères de redressement dans les enquêtes électorales

Ipsos - vague 18 - 18 et 19 avril 2012 - 1 ^{er} t.	V inter / V totale %	Ifop - vague 8 - 12 au 15 avril 2012 - 1 ^{er} t.	V inter / V totale %	Ipsos - vague 18 - 18 et 19 avril 2012 - 1 ^{er} t.	V inter / V totale %	Ifop rolling 20 avril : 1er tour	V inter / V totale %
Sexe	<u>1,2</u>	Sexe	<u>0,8</u>	Catégorie d'agglom.	<u>1,3</u>	Catégorie d'agglom.	<u>0,4</u>
Homme	1,2	Homme	0,7	Rural	1,3	Communes rurales	0,7
Femme	1,2	Femme	0,9	Moins de 20.000 hab.	1,1	Communes urbaines de province	0,2
Age	<u>3,0</u>	Age	<u>1,4</u>	20.000 - 100.000 hab.	0,5	Agglomération parisienne	0,4
18 - 34 ans	2,4	18 à 24 ans	0,7	Plus de 100.000 hab	1,3	Statut	<u>1,4</u>
35 - 44 ans	1,8	25 à 34 ans	1,2	Agglomération de Paris	2,2	Salarié du secteur public	0,2
45 - 59 ans	2,9	35 à 49 ans	1,2	Niveau d'études	<u>3,7</u>	Salarié du secteur privé	1,8
60 ans et plus	4,3	50 à 64 ans	0,8	Inférieur au Bac	4,6	Retraité	1,9
		65 ans et plus	3,1	Baccalauréat	1,2	Autre inactif	0,9
PCS Chef de Famille	<u>2,4</u>	CSP interviewé	<u>1,8</u>	Supérieur au Bac	4,7	Indépendant / Employeur	1,4
		Agriculteur	0,2	Présid. 2007 1er t.	<u>43,3</u>	Présid. 2007 1er t.	<u>36,2</u>
Artisan Com. Chef d'entr.	2,6	Artisan ou commerçant	0,7	s/t EXG PC Verts	31,1	O Besancenot	8,9
Cadre supérieur	3,2	Prof. libérale, cadre sup.	2,1	S. Royal	47,1	S Royal	39,5
Prof. intermédiaire	1,4	Prof. intermédiaire	1,1	F. Bayrou	30,4	F Bayrou	22,4
Employé	2,0	Employé	1,2	N. Sarkozy	55,2	N Sarkozy	56,4
Ouvrier	3,0	Ouvrier	4,4	J.M. Le Pen	40,5	JM Le Pen	35,3
Retraité	2,8	Retraité	1,9	Présid 2007 2ème t.	<u>63,4</u>	Autre	17,0
Inactif	0,5	Autre inactif	0,9	Région.2010 1er t.	<u>47,5</u>		
				Proximité politique	<u>49,6</u>	Proximité politique	<u>28,7</u>

3.2 Choix des critères de redressement dans les enquêtes électorales

Ifop - Analyse des votes selon les habitudes médias - 21 au 23 avril 2012 - 6210 interviews -											
Quel journal télévisé regardez-vous généralement le soir ?			Pour vous informer, quelle radio écoutez-vous le plus souvent ?			Quel journal ou quels journaux lisez vous régulièrement ?			Quel hebdomadaire ou quels hebdomadaires lisez-vous régulièrement ?		
	V inter / V totale %	Audience*		V inter / V totale %	Audience*		V inter / V totale %	Audience*		V inter / V totale %	Audience*
Le 20h sur TF1	5,9	23	RTL	3,2	16	Quotidien régional	1,4	48	Le Point	1,4	5
Le 20h sur France 2	1,2	14	Europe 1	0,9	10	Le Monde	1,2	5	Le Nouvel Observateur	3,1	7
Le 19-20sur France 3	0,6	10	France inter	5,9	14	Le Figaro	2,0	3	Le Figaro Magazine	2,6	5
Le Journal sur Canal+	0,5	3	France info	0,1	5	Le Parisien	0,2	4	Marianne	2,6	5
Le journal de 19h45 sur M6	0,7	12	RMC	0,4	9	Libération	1,3	2	L'Express	0,4	6
Les journaux ... en continu	0,5	10	France Bleue	0,5	9	Les Echos	0,5	2	Paris-Match	1,4	10
Aucun	2,7	28	NRJ	1,9	10	L'Humanité	3,0	1	Valeurs actuelles	1,6	2
ensemble	<u>2,4</u>	100	BFM radio	0,2	3	La Croix	0,8	1	Aucun	2,6	61
* En l'absence de données Ifop, approximation en fonction des données disponibles (Médiamétrie et ONE)			Aucune	0,4	24	Un journal gratuit	0,9	12	ensemble	<u>2,1</u>	100
			ensemble	<u>1,8</u>	100	Aucun	1,3	21			
						ensemble	<u>1,2</u>	100			

3.2 Choix des critères de redressement dans les enquêtes électorales : conclusions

- **Les 5 critères sociodémographiques sont très peu discriminants des intentions de vote.** Leur utilisation pour le redressement en plus des quotas ne peut cependant être que bénéfique puisqu'ils corrigent les biais même très faibles dont ils pourraient être responsables, sans accroître sensiblement le bruit d'échantillonnage, les poids étant rarement supérieurs à 1.20 en raison des quotas.
- A l'inverse, **les souvenirs de vote sont très discriminants**, même si les résultats indiqués sont surestimés puisqu'ils ne concernent que les souvenirs « exprimés » (hors NR, ABS, BN qui peuvent être nombreux). Mais le redressement par les souvenirs de vote ne corrige pas les biais éventuels de non-réponse ou de réponses volontairement erronées des nouveaux venus à un candidat, mais seulement ceux des fidèles, d'où une sous-estimation des partis sensibles comme le FN lorsqu'ils ont des intentions de vote en croissance notable (2002, 2012).
- **Pour « épauler » les souvenirs de vote, il faudrait probablement croiser les CSP par le statut**, y compris pour les retraités, et davantage détailler les CSP, en séparant par exemple les cadres sup des professions libérales et les enseignants des salariés. Des essais d'AD sur les habitudes média pourraient aussi être effectués, bien que leur pouvoir discriminant des votes soit limité, mais en partie en raison d'une proportion forte n'utilisant pas le media (21 à 61 %). Par exemple, pour les hebdomadaires, le % de variance expliquée augmente de 2.1 à 4.7 % pour les interviewés concernés.

Conclusions

- La méthode « cannibal » permet une estimation fiable des transferts d'intentions de vote en cours de campagne et entre le dernier sondage et le scrutin basée sur le choix en second des indécis s'il y a suffisamment d'indécis.
- Le remplacement des NR est indispensable, en utilisant des analyses discriminantes lorsqu'ils dépassent 10 %. En particulier, au second tour de la présidentielle, il faut affecter à un candidat les NR certains d'aller voter en fonction du choix des répondants ayant choisi le même candidat au premier tour et ne pas assimiler ces NR à des abstentionnistes.
- Le choix des critères de redressement et de leurs modalités devrait être validé par des analyses discriminantes. De plus, compte tenu des limites du redressement par les souvenirs de vote, d'autres critères sont à rechercher pour améliorer la représentativité politique de l'échantillon, peut-être un croisement de CSP plus détaillées par le statut public / privé et ou les habitudes médias, mais le succès n'est pas garanti !
- En utilisant des méthodes expérimentées pour les enquêtes marketing, des progrès sont à l'évidence possibles pour la qualité des intentions de vote et pour leur compréhension, y compris sur d'autres points que ceux abordés dans cet exposé, comme par exemple la mesure de l'importance des raisons de choix des candidats (ou enjeux décisifs, préoccupations, ...). Mais ces progrès ne pourront avoir lieu que si des chercheurs politologues et statisticiens coopèrent et peuvent avoir accès a posteriori aux fichiers des instituts.