

Sondages d'intentions de vote : exemples d'erreurs possibles et étude des performances des instituts de sondages lors des élections de 2024 en France

Antoine Auberger
Université Paris-Panthéon-Assas (Paris 2), CRED, France
7 rue de la grande chaumière
75006 - Paris
antoine.auberger@u-paris2.fr

6 novembre 2024

Version préliminaire (1^{ère} version, 31 mai 2024)
(merci de ne pas citer sans l'accord de l'auteur s'il-vous-plaît)

Résumé

On étudie les erreurs des instituts de sondages pour les sondages pré-électorales d'intentions de vote et on essaie de les expliquer ? On fait cette étude principalement trois pays : Etats-Unis (élections présidentielles), Royaume Uni (élections législatives) et France (principalement les élections présidentielles et législatives mais les élections régionales de 2021 sont abordées).

La méthodologie et le mode peuvent avoir une influence sur les performances des sondages pré-électorales d'intentions de vote et aux Etats-Unis, la méthode des quotas avait été abandonnée après l'élection présidentielle de 1948 malgré le succès de cette méthode pour l'élection présidentielle de 1936. Au Royaume-Uni partiellement et en France, la méthode des quotas est toujours utilisée. Il y a aussi d'autres explications possibles pour les erreurs qui sont présentées. Certaines erreurs sont moins visibles si elles ne changent pas le résultat prévu comme pour le vote Mélenchon qui avait été sous-estimé pour l'élection présidentielle de 2022 mais sans conséquences pour le second tour annoncé.

On étudie ensuite les performances des instituts de sondages en France pour les élections de 2024. On peut dire que les performances sont (très) bonnes dans l'ensemble pour les élections européennes de 2024 et satisfaisantes pour les élections législatives de 2024 malgré les erreurs pour les sièges (la projection dépend des intentions de vote mais aussi de la méthodologie utilisée).

Mots-clés : Sondage d'opinion - histoire - élections – erreurs - Etats-Unis, Royaume-Uni, France – élections européennes 2024 – élections législatives 2024

Polls of voting intentions: examples of possible errors and study of the performance of polling institutes during the 2024 elections in France

Abstract

We study the errors of polling institutes in pre-election polls of voting intentions and try to explain them. We do this study mainly in three countries: the United States (presidential election), the United Kingdom (parliamentary elections), and France (mainly presidential and parliamentary elections, but the regional elections of 2021 are discussed).

The methodology can influence the performance of pre-electoral polls of voting intentions. In the United States, quota sampling was abandoned after the 1948 American presidential election despite the success of this method in the 1936 American presidential election. Quota sampling is still used partly in the UK and partly in France. Other possible explanations for the errors are presented. Some errors are less visible if they do not change the expected result, as for the

Mélenchon vote, which had been underestimated for the 2022 presidential election but without consequences for the announced second round.

We then study the performance of polling institutes in France for the 2024 elections. The performance is (very) good overall for the 2024 French European elections and satisfactory for the 2024 French legislative elections despite the errors in seats (the projection depends on the voting intentions and the methodology used).

Key-Words: opinion polls - history - elections - errors - United States - United Kingdom – France – 2024 French European elections – 2024 French legislative elections

Introduction

On étudie les erreurs des instituts de sondages pour les sondages pré-électorales d'intentions de vote et on essaie de les expliquer ? On s'appuie principalement sur trois pays : Etats-Unis, Royaume Uni et France. Dans ces trois pays, les instituts de sondages se comportent différemment quant à la méthode utilisée : les instituts de sondages américains ont arrêté la méthode des quotas après l'élection américaine de 1948 en la rendant responsable de l'échec des sondages pré-électorales (avec cette méthode, ils trouvaient plus facilement des électeurs républicains et ont surestimé le vote républicain pour la période 1936-1948). Ces dernières années, avec les sondages par internet, il y a même aux Etats-Unis un abandon partiel des sondages aléatoires par téléphone et on constate qu'il y a différentes possibilités de sélectionner l'échantillon. Au Royaume Uni, il y a eu des échecs et la question du choix de la méthodologie s'est posée. En France, les instituts de sondages sont restés fidèles à la méthode des quotas et réalisent leurs sondages pré-électorales d'intentions de vote en ligne (par internet) et on pense que les instituts de sondages en France ont un savoir-faire pour les redressements politiques. Deville (1991) présente les fondements théoriques de la méthode des quotas et Ardilly (2023) résume le débat sur le plan statistique entre la méthode des quotas avec un échantillon empirique (biais) et la méthode aléatoire avec un échantillon probabiliste (variance d'échantillonnage plus grande). On sait aussi qu'en France les instituts de sondages ne disposent pas de base de sondages et que la méthode des quotas est moins coûteuse et plus rapide que la méthode aléatoire avec laquelle les instituts de sondages doivent essayer de recontacter les personnes tirées au hasard et ne répondant pas. Pour une taille d'échantillon de 1000 à 5000 personnes, l'erreur d'échantillonnage est moins élevée avec la méthode des quotas qu'avec la méthode aléatoire et cela compense le biais de cette méthode.

On commencera par étudier les explications possibles des erreurs pour les sondages d'intentions de vote et ensuite on étudiera les performances des sondages d'intentions de vote pour les élections de 2024 en France.

Explications possibles des erreurs pour les sondages d'intentions de vote

On trouve des explications connues pour les erreurs des sondages pré-électorales d'intentions de vote : voir notamment, Mellon et Prosser (2017), Prosser et Mellon, (2018), Sturgis et al. (2018) : articles rédigés en lien avec l'erreur faite pour les élections législatives britanniques de 2015 et Gelman (2021) en lien avec les élections présidentielles américaines de 2016 et 2020.

Méthodologie

Les erreurs des sondages d'intentions de vote peuvent être liées à la méthodologie : quotas / méthode aléatoire (sélection de l'échantillon), sondages faits par internet (électeurs mieux informés et s'intéressant davantage à la politique pour ceux / celles qui sont volontaires pour répondre à des sondages pré-électorales d'intentions de vote et essayer d'influencer l'opinion publique).

Sélection de l'échantillon et analyse des répondants : on sait que l'échantillonnage est important ainsi que le traitement des réponses / non réponses avec l'exemple de l'élection présidentielle américaine de 1936 ; l'échantillon était biaisé ainsi que le sous-échantillon de personnes qui avaient répondu. Le sondage réalisé par le Literary Digest n'avait pas fait de redressement par rapport au vote lors de l'élection présidentielle américaine de 1932 et cela aurait permis une correction de l'erreur très importante (annonce de la victoire de Roosevelt).

Choix de la méthode et du mode : la méthode des quotas a eu des succès lors des élections présidentielles américaines de 1936, 1940 et 1944 mais elle a été jugée au moins en partie responsable de l'erreur pour l'élection présidentielle américaine de 1948. Depuis aux Etats-Unis, la méthode aléatoire est utilisée avec des listes d'électeurs à la disposition des instituts de sondages. On est passé

de sondages réalisés par courrier et en face à face à des sondages par téléphone. Depuis 10 à 20 ans, le développement d'internet permet l'utilisation d'internet pour les sondages pré-électorales (sondages probabilistes ou non probabilistes).

En France, les instituts de sondages utilisent internet depuis les sondages pré-électorales réalisés pour l'élection présidentielle de 2012 : sondages faits à l'aide de panels d'électeurs volontaires pour répondre à des sondages.

Les sondages pré-électorales réalisés par téléphone à l'aide de la méthode aléatoire (Etats-Unis) sont plus longs à mettre en œuvre et plus coûteux même s'ils peuvent être faits sans personne qui interviewe.

Biais des sondages probabilistes lié aux non réponses : voir Bethlehem (1988, 2017a) : $B(p) = \frac{R_{\rho,Y} S_{\rho} S_Y}{\bar{p}}$

Les sondages probabilistes réalisés par internet ont un biais lié à l'utilisation d'internet (internet / pas internet) et pour les sondages avec un panel de personnes volontaires, il y a un biais lié au panel (panel / pas panel), d'auto sélection (personnes volontaires pour faire partie d'un panel). Pour les sondages réalisés par la méthode des quotas, il y a un biais supplémentaire lié à l'utilisation de cette méthode qui a un avantage pour la variance d'échantillonnage pour les sondages avec une taille d'échantillon jusqu'à 5000 personnes.

Bethlehem (2010) étudie la variance d'un sondage fait par internet avec auto sélection¹ (les deux biais s'ajoutent) avec une comparaison du biais lié à la non réponse avec la méthode aléatoire ainsi qu'une expression de la variance pour les sondages réalisés par internet (somme de deux termes pour la variance de l'estimateur stratifié). On peut essayer d'adapter son calcul aux sondages pré-électorales d'intentions de vote en France (quatre biais et quatre composantes pour la variance).

Si des sondages pré-électorales peuvent ne pas bien prendre en compte des catégories d'électeurs (par internet avec un panel ?), cela peut être une source d'erreur des sondages d'intentions de vote (exemple des Etats-Unis pour 2016 ?)

L'erreur de 2015 au Royaume-Uni a été attribuée (en partie) à l'utilisation de la méthode des quotas et à la non représentativité des échantillons utilisés.²

Durand et Johnsson (2021) s'intéressent à l'influence du choix de la méthodologie sur les performances des instituts de sondages pour quatre pays : Canada, France, Royaume-Uni, Etats-Unis.

Aux Etats-Unis, la méthode des quotas avait été rejetée après l'élection présidentielle de 1948 mais il y a maintenant différentes façons de sélectionner un échantillon (téléphone, internet) alors que les sondages par téléphone avec la méthode aléatoire a longtemps été la plus utilisée. Au Royaume-Uni, il y a eu et il y a débat entre la méthode aléatoire et la méthode des quotas. En France, tous les instituts de sondages réalisent leurs sondages d'intentions de vote par internet avec la méthode des quotas.³

- Abstention différentielle

Cela semble expliquer les erreurs pour les sondages d'intentions de vote lors des élections régionales de 2021 en France. Les électeurs du RN avaient moins participé que prévu (élections favorables aux majorités régionales sortantes). Difficultés liées à une faible participation électorale et aussi une participation électorale plus faible que celle qui avait été estimée.

Les sondages peuvent surestimer la participation électorale : à corriger éventuellement ? (critique potentielle des access panels avec des électeurs plus intéressés par la politique que la moyenne des électeurs)

C'est aussi l'une des raisons pour l'erreur de 2015 au Royaume-Uni : les électeurs favorables au parti travailliste ont moins voté que ceux favorables au parti conservateur (électeurs favorables à un parti, qui pensaient voter et qui s'abstiennent).

Changements tardifs pour le vote

- Changements de vote, choix tardif (électeurs NSP). C'était l'une des raisons pour expliquer l'erreur de l'élection présidentielle américaine de 1948. Deux instituts de sondages dont Gallup avaient arrêté les sondages deux semaines avant l'élection présidentielle américaine de 1948.

¹ D'après Bethlehem (2010), le biais lié à l'auto sélection est donné par : $B(\bar{y}_S) = E(\bar{y}_S) - \bar{Y}_I \approx \bar{Y}_I^* - \bar{Y}_I = \frac{C(\rho,Y)}{\bar{p}}$

² Une partie des sondages sont réalisés par téléphone et une partie par internet (voir Wikipedia : « Opinion polling for the 2015 United Kingdom general election »).

³ Ce qui peut être critiqué notamment, c'est que la méthode des quotas est utilisée à partir de quotas marginaux alors que les quotas croisés sont importants et qu'il n'y a qu'Ipsos qui mentionne une repondération. Le(s) redressement(s) politiques peuvent permettre d'obtenir quand même des résultats satisfaisants.

2015 au Royaume-Uni : sans doute car un sondage non publié réalisé uniquement la veille du vote donnait une avance de 6 points au parti conservateur alors que les autres sondages les donnaient « au coude à coude ».

C'est aussi une raison qui expliquait que la surprise de la présence de Le Pen au second tour de l'élection présidentielle française de 2002 car la publication des sondages était interdite après le vendredi soir une semaine avant le premier tour. Il y avait des sondages réalisés lors de la semaine précédant les élections mais ils n'étaient pas publiés. La connaissance des éventuels changements lors de la dernière semaine avant l'élection est importante pour que les électeurs soient informés. Si les électeurs de gauche avaient su que Jospin n'était pas sûr d'être au second tour, certains électeurs auraient fait le choix du vote utile pour que Jospin soit présent au second tour.

On sait qu'il est important de faire des sondages au cours des derniers jours avant une élection (ou des élections) pour prendre en compte que certains électeurs prennent leur décision très tardivement, ainsi que les éventuels changements.

- Electeurs ne se prononçant pas (NSP) qui votent plus pour un candidat (parti) : cela peut en partie expliquer l'erreur pour l'élection présidentielle américaine de 1948.

- Electeurs mentant sur leur préférence

Electeurs de Trump (2016, 2020)

Electeurs de Le Pen (J.M.) en France (élection présidentielle française de 2002)

- « Shy voters » / Electeurs refusant de révéler leur préférence

Raison qui avait expliqué le vote pour Trump sous-estimé en 2016.

Raison qui avait pu expliquer le vote sous-estimé pour le parti conservateur lors de certaines élections législatives britanniques et notamment pour 1992 et 2015.

Le vote Le Pen a longtemps été sous-estimé avec le résultat brut des sondages : nécessité de faire un redressement en France (cas de 2002 notamment avec un redressement important). Normalement, avec les redressements politiques, il y a une protection. Il peut y avoir un risque de surestimation avec des électeurs volontaires pour répondre (par internet).

- Electeurs refusant de répondre aux sondages : c'est ce qui est dit pour certains électeurs de Trump (élections présidentielles américaines de 2016 et 2020)

- Catégories de la population qui sont oubliées (ne peuvent pas être interrogées) : on sait que l'utilisation du téléphone fixe a diminué par rapport au téléphone mobile. Pour les sondages faits par internet, il y a des personnes qui n'ont pas internet ou qui ne savent pas l'utiliser.

Les sondages pré-électorales faits par internet n'ont pas le problème des non-réponses ou très peu (les électeurs sont volontaires) et normalement cela permet d'avoir moins de réponses insincères (électeurs mentant sur leurs préférences) et d'éviter la difficulté d'électeurs ne souhaitant pas révéler leurs préférences.

- Erreurs liées au système électoral

Aux Etats-Unis, deux principaux candidats à l'élection présidentielle avec un système de grands électeurs par Etat. Possibilité d'avoir la majorité en voix (Clinton) sans avoir la majorité des grands électeurs (2016). Deux principaux candidats qui peuvent être assez proches ou en tout cas il y a des électeurs indépendants et aussi la possibilité pour des électeurs démocrates de voter pour le candidat républicain et inversement. Pour l'élection présidentielle américaine de 1948, il y avait deux « petits candidats » proches du candidat démocrate (Truman) qui ont obtenu un score non négligeable et qui a pu être surestimé lors des sondages (intérêt de voter pour l'un des deux principaux candidats : vote utile).

Au Royaume-Uni, il y a deux principaux partis (Conservateur et Travailliste) avec un troisième parti plus ou moins important suivant les élections (Lib Dem) et qui se situait entre ces deux partis et aussi d'autres partis plus petits. Avec un scrutin à un tour, cela favorise les éventuels changements de vote et les décisions tardives de vote et le résultat en sièges accentue les éventuels écarts du vote (différence de vote entre les deux principaux partis).

- Traitement médiatique

Exemple des élections présidentielles américaines de 1948, 2016 : élections jouées d'avance ?

Exemple de l'élection présidentielle française de 2002 (premier tour) : les médias en commentant les sondages d'intentions de vote n'avaient jamais envisagé un second tour avec Le Pen (Chirac / Le Pen ou Jospin / Le Pen) malgré la présence de nombreux candidats à gauche (dispersion des voix) comme Chevènement (ancien ministre du gouvernement Jospin) et Taubira (le parti radical de gauche est très proche du parti socialiste).

Suite notamment aux élections législatives britanniques de 1992 et 2015, des recommandations ont été faites : on peut essayer de mieux exploiter les données d'un sondage. Etudier le vote des personnes certaines d'aller voter qui ne donnent pas d'intentions de vote, étudier le vote des personnes qui ne sont pas certaines d'aller voter (probabilités de vote).

Elections européennes de 2024 (en France)

Sondages d'intentions de vote et méthodologie

On compte dix instituts de sondages (on les trouve dans le tableau 1) qui proposent des sondages d'intentions de vote pour les élections européennes de 2024 en France. OpinionWay et l'Ipsos proposent chacun deux sondages : OpinionWay pour Les échos – radio classique et CNews - Europe 1 - Le Journal du Dimanche et l'Ipsos pour le Cevipof, le Monde, la Fondation Jean Jaurès et l'Institut Montaigne et pour Le Parisien et Radio France. Le sondage réalisé par l'Ipsos pour Le Monde a la particularité d'être réalisé avec une taille d'échantillon plus élevée (environ 10 000 personnes) que les autres sondages qui sont réalisés pour environ 1000 à 2000 personnes. L'Ifop propose un sondage rolling depuis le 8 avril 2024 publié par LCI – Le Figaro – Sud radio (le principe est d'interroger trois jours de suite 500 électeurs ou 1000 électeurs pendant la dernière semaine avant les élections) après avoir publié des sondages d'intentions de vote sur les élections européennes en France depuis décembre 2023. On s'intéresse aux personnes inscrites sur les listes électorales.

Tous les instituts de sondages sauf l'Ipsos commencent par constituer un échantillon représentatif de la population française des 18 ans et plus avec les variables sexe, âge, profession, région et catégorie d'agglomération avec utilisation de la méthode des quotas (l'Ifop et Elabe précisent après stratification par région et catégorie d'agglomération) et ensuite les instituts de sondages enlèvent de leur échantillon les personnes qui ne sont pas inscrites sur les listes électorales. Tous les instituts de sondages font un redressement socio-démographique de leur échantillon. L'Ipsos constitue un échantillon représentatif de la population française en ne prenant en compte que des personnes inscrites sur les listes électorales. La taille de l'échantillon pour les derniers sondages réalisés avant le premier tour est comprise entre 1035 personnes inscrites sur les listes électorales (YouGov) et 8923 personnes inscrites sur les listes électorales.⁴

Tous les sondages réalisés sont faits en ligne : l'Ipsos, Harris Interactive et YouGov mentionnent un access panel et Cluster17 un panel propriétaire. Les sondages d'intentions de vote réalisés par internet (access panel ou fichier d'électeurs), par questionnaire auto-administré en ligne sur système CAWI (« Computer Assisted Web Interview ») sont peu coûteux et rapides à réaliser (par rapport aux sondages faits par téléphone). Leurs inconvénients : biais lié à internet (personnes ayant internet / personnes n'ayant pas internet ou ne le maîtrisant pas suffisamment), biais lié au panel : le panel d'électeurs ne représente qu'une partie de la population des personnes qui ont internet, biais lié à l'auto-sélection (les personnes contactées sont volontaires pour répondre et il faut répondre rapidement pour voir ses réponses prises en compte donc il peut y avoir une corrélation entre la probabilité de répondre et l'intérêt pour le sondage), biais lié à la méthode des quotas mais pour un échantillon de taille égale à environ 1000 à 3000 (5000) électeurs inscrits sur les listes électorales, le biais est compensé par une faible variance d'échantillonnage (voir Deville, 1991, Ardilly, 2023).

On rappellera les électeurs pris en compte pour les intentions de vote (note 10, ...), le(s) redressement(s) politique(s) utilisé(s), le nombre de catégories par variables.

Performances pour le vote (9 juin 2024)

Participation électorale

On présente dans le tableau 1 la participation électorale estimée lors du dernier sondage d'intentions de vote avant les élections européennes de 2024 en France et la valeur absolue de l'écart entre la participation électorale (France entière) et la participation électorale estimée par chaque institut de sondages et on présente aussi un résultat avec la moyenne des instituts de sondages. La participation

⁴ Tous les instituts de sondages ont une taille d'échantillon comprise entre 1000 et 3000 sauf l'Ipsos pour le sondage mentionné (pour l'autre sondage de l'Ipsos publié par Radio France - Le Parisien juste avant le premier tour, on a n = 1738).

estimée est donnée par un indice (OpinionWay) légèrement supérieur à la note 10⁵ ou par une estimation (Ipsos) ou simplement par la note 10.^{6,7}

On voit que si on prend la moyenne de la participation électorale estimée, la valeur absolue de l'écart entre la participation électorale et la participation électorale estimée est très faible (valeur absolue de l'écart entre la participation et la moyenne de la participation estimée égale à 1,3) sauf pour l'Ipsos, BVA et OpinionWay qui sous estiment assez nettement la participation électorale.

Modélisation de l'estimation de la participation : sondages faits à partir du 18 mai (publication des listes au JO). Variables explicatives envisagées : le nombre de jour avant les élections, variables statistiques (âge, profession), estimation d'un indice de participation ou note 10.

Vote

Si on s'intéresse aux sept principales listes (listes qui dépassent 5 %), l'erreur la plus faible (en valeur absolue) de la moyenne des instituts de sondages est pour les listes REC (0,08), EELV (0,20) et PS-PP (0,28) et l'erreur la plus élevée est pour la liste LFI (1,69) dont le vote est sous estimé. Le vote pour la liste RN est surestimé mais l'erreur n'est pas très importante (0,88).

Tableau : intentions de vote par institut et en moyenne (7 principales listes)

Institut	dates	LFI	EELV	PS-PP	REN-MOD-HOR	LR	RN	REC
Ipsos	6-7 juin	9,5	14,5	5	15	7	32	5,5
Harris Interactive	6-7 juin	9,5	13	5	14	7,5	32	5
Cluster 17	6-7 juin	9	13	5	15	7	31	5,5
Elabe	5-7 juin	9,5	14,5	5,5	16	6	32	5,5
Ifop	4-7 juin	9	13	5,5	14,5	7	33	6
Odoxa	5-6 juin	7	14	5	15	7	32	6
OpinionWay	4-6 juin	8	14	5	15	6	33	6
BVA	3-5 juin	7,5	14	5,5	16	6,5	33	5
YouGov	31 mai - 5 juin	6	12	4	15	6	32	5
Viavoice	20-21 mai	6,5	14	7	17	6,5	32	6
Moyenne		8,2	13,6	5,3	15,3	6,7	32,2	5,6
Vote		9,9	13,8	5,5	14,6	7,2	31,4	5,5

Si on s'intéresse à l'ensemble des listes qui ont au moins une intention de vote non nulle (25), l'erreur (en valeur absolue) de la moyenne des instituts de sondages est égale à 0,33.

En général, les instituts de sondages donnent des intentions de vote à 0,5 % près sauf OpinionWay qui n'a une précision qu'à 1 % près.⁸ Pour certaines listes, il y a : < 0,5 % et cela signifie < 0,25 car sinon il y aurait un arrondi à 0,5 % pour tous les sondages sauf OpinionWay qui n'a qu'une précision à 1 %.

On construit un modèle pour estimer la valeur absolue de la différence entre le vote pour chaque liste et l'intention de vote pour chaque sondage. On reprend le modèle de Wlezien et Jennings (2023) en l'adaptant (variables explicatives :). On souhaite tester en plus des variables statistiques liées au nombre de catégories pour l'âge, la profession, les régions, les catégories d'agglomération. On souhaite aussi tester les variables liées aux électeurs pris en compte (note 10, note 9-10, note 8 à 10, utilisation de probabilités d'aller voter).

Sièges

La prévision pour le calcul de la répartition des sièges est fait à partir des listes qui obtiennent plus de 5 % des suffrages exprimés. On constate que les sept listes ayant obtenu plus de 5 % des suffrages

⁵ Dans le sondage d'OpinionWay qui est mentionné, la note 10 correspond à 46,1 %.

⁶ Certains instituts de sondages précisent également pour certains sondages le pourcentage d'électeurs correspondant à la note 9.

⁷ On trouve dans une version plus longue un tableau sur la participation estimée.

⁸ L'institut YouGov donne dans sa notice technique des intentions de vote à 0,01 % près mais la communication s'est faite pour les six principales listes à 1 % près.

exprimés obtiennent bien plus de 5 % des intentions de vote. Il n’y a qu’Elabe qui présente des fourchettes en sièges.

Elections législatives françaises de 2024

Sondages d’intentions de vote et méthodologie⁹

On compte sept instituts de sondages qui proposent des sondages d’intentions de vote pour les élections législatives de 2024. OpinionWay et l’Ipsos proposent deux sondages : OpinionWay pour Les échos et Cnews, l’Ipsos pour Le Monde ou Le Monde - Fondation Jean Jaurès - Sciences Po - Institut Montaigne - France tv - Radio France ou Le Parisien - Radio France). Ifop-Fiducial propose un sondage rolling depuis le 20 juin 2024 publié par LCI, Le Figaro et Sud Radio.¹⁰

La taille de l’échantillon pour les derniers sondages réalisés avant le premier tour est comprise entre 1058 personnes inscrites sur les listes électorales (OpinionWay pour les Echos) et 10286 personnes inscrites sur les listes électorales.¹¹

On rappellera les électeurs pris en compte pour les intentions de vote (note 10, ...), le(s) redressement(s) politique(s) utilisé(s), le nombre de catégories par variables.

Performances pour le premier tour (30 juin 2024)

Participation électorale

On présente dans le tableau 1 la participation électorale estimée lors du dernier sondage d’intentions de vote avant les élections législatives de 2024 et la valeur absolue de l’écart entre la participation électorale (France entière) et la participation électorale estimée par chaque institut de sondages et on présente aussi un résultat avec la moyenne.¹²

On voit que si on prend la moyenne de la participation électorale estimée (note 10 pour OpinionWay), la valeur absolue de l’écart entre la participation électorale et la participation électorale estimée est assez faible. la participation électorale a été légèrement sous-estimée sauf par l’Ifop qui présente une très bonne estimation de la participation.

Modélisation de l’estimation de la participation : sondages faits à partir de l’annonce de la dissolution (soir du 9 juin). Variables explicatives envisagées : le nombre de jour avant les élections, variables statistiques (âge, profession), estimation d’un indice de participation ou note 10.

Vote au premier tour

On présente dans le tableau 2 les intentions de vote pour les quatre principaux regroupements par institut de sondages et en moyenne avec également le vote.¹³

Tableau 2 : intentions de vote par institut et en moyenne
(4 principaux regroupements)

Institut	dates	NFP	ENS	LR et DVD	RN ^a
Ipsos	27-28 juin	29	20	8	36
Harris Interactive	27-28 juin	28	20	8	37
Ifop	25-28 juin	29	20,5	7	36,5
Odoxa	26-27 juin	27,5	21	9	35
OpinionWay	26-27 juin	28	20	8	37
Elabe	26-27 juin	27,5	20	9	36
Cluster17	25-27 juin	29,5	20	9	35

⁹ On reprend la présentation (méthodologie) faite pour les sondages d’intentions de vote aux élections européennes de 2024 en France.

¹⁰ Le principe est rappelé lors de la présentation des sondages d’intentions de vote pour les élections européennes de 2024 en France.

¹¹ Tous les instituts de sondages ont une taille d’échantillon comprise entre 1000 et 3000 sauf l’Ipsos pour le sondage mentionné (l’Ipsos fait également des sondages avec 2000 personnes inscrites sur les listes électorales).

¹² On trouve dans une version plus longue un tableau sur la participation estimée au premier tour.

¹³ On trouve dans une version plus longue un tableau sur les performances.

moyenne	28,25	20,25	8,33	36,08
vote	28,06	21,28	10,17	33,21

^a Dans ce tableau et dans les autres tableaux, RN signifie : RN et alliés (candidats liés à Ciotti)

Les erreurs les plus élevées sont pour les intentions de vote du RN qui surestiment le vote du RN et pour les intentions de vote de LR et DVD qui sous estiment le vote pour LR et DVD. C'est l'institut Odoxa qui présente la plus faible erreur moyenne en valeur absolue pour 4 regroupements et 8 regroupements. Odoxa tient compte uniquement des personnes certaines d'aller voter (10) et exprimant une intention de vote (l'Ipsos également). Pour le redressement politique, Odoxa tient compte du vote à l'élection présidentielle de 2022, au premier tour des élections législatives de 2022 et des élections européennes de 2024. C'est l'institut Harris Interactive qui obtient les moins bonnes performances. Une question est d'essayer de relier les performances avec les choix des institutions de sondages (méthodologie pour sélectionner l'échantillon, variables statistiques, redressement politique).¹⁴

On construit un modèle pour estimer la valeur absolue de la différence entre le vote pour chaque liste et l'intention de vote pour chaque sondage. On reprend le modèle de Wlezien et Jennings (2023) en l'adaptant. On souhaite tester en plus des variables statistiques liées au nombre de catégories pour l'âge, la profession, les régions, les catégories d'agglomération. On souhaite aussi tester les variables liées aux électeurs pris en compte (note 10, note 9-10, note 8 à 10, utilisation de probabilités d'aller voter).

Performances pour le second tour (7 juillet 2024)

Il y a eu des sondages nationaux pour estimer la participation électorale au second tour.¹⁵ OpinionWay présente même des sondages pour la participation électorale suivant le type de second tour : triangulaire ou duels.

Elabe avait un peu sous-estimée la participation électorale (1,6 point) et l'Ifop et l'Ipsos l'avaient un peu surestimée (respectivement 2,4 et 1,4 point). OpinionWay avait sous-estimée de façon plus importante la participation électorale au second tour suivant les différentes configurations mais sur un total de 352 circonscriptions (501 circonscriptions avaient un second tour). La participation électorale estimée par OpinionWay est très proche de la participation électorale égale au pourcentage d'exprimés (63 %).

Pour les intentions de vote, il peut y avoir des sondages généraux du type : triangulaires NFP / ENS (ou HOR, UDI, DVC) / RN (ou UXD), duels NFP / RN (ou UXD), duels ENS / RN (ou UXD) mais il n'y a qu'OpinionWay qui en présente après le premier tour dans son rapport.

On constate que le vote RN a été surestimé au second tour dans les trois configurations envisagées par OpinionWay (il y a eu aussi 60 duels LR ou DVD / RN ou UXD).

En comparant avec le sondage d'Odoxa, on constate que le type de candidat du NFP (LFI, EELV, PS, ...) compte et aussi que les intentions de vote pour le RN en cas de triangulaire NFP-ENS-RN ont diminué entre les 26-27 juin et les 3-4 juillet.

Les performances en sièges sont étudiées avec attention.

On commence par présenter un tableau récapitulatif sur les projections en sièges réalisées avant le premier tour (tableau 10). Un nombre élevé de triangulaires est envisagé par des instituts de sondages : Odoxa envisage entre 120 et 170 triangulaires dans son sondage réalisé les 19 et 20 juin et plus de 170 triangulaires dans son baromètre politique (19-20 juin). Une originalité d'Odoxa avant le premier tour est de présenter un sondage d'intentions de vote suivant la configuration du second tour (voir note de bas de page 1). Avant le premier tour, la force d'un éventuel front républicain (barrage au RN et à ses alliés) a été sous-estimé. Un sondage réalisé par Elabe (26-27 juin) montre qu'environ 3 électeurs sur 4 ne suivront pas les consignes de vote au second tour et qu'une majorité d'électeurs d'Ensemble (60 %) et de NFP (55 %) ne souhaiteraient pas que si leur candidat arrive en troisième position, il se retire pour faire barrage au RN. Par contre, dans le sondage Elabe (3-4 juillet), 75 % des électeurs du NFP approuvent le retrait de leur candidat arrivé en 3^{ème}. Pour les électeurs d'Ensemble, il y a une majorité approuvant le retrait (65 %) pour soutenir un candidat socialiste ou verts mais pas pour un candidat LFI (39 % pour et 60 % contre).

¹⁴ On trouve dans une version plus longue les performances pour les autres regroupements.

¹⁵ On trouve dans une version plus longue un tableau sur la participation estimée au second tour.

Le 5 juillet au soir, on a sept projections en sièges très récentes (tableau 11) :¹⁶

Tableau 11 : projection en sièges des instituts de sondage

Institut	dates	NFP	ENS	DVD-LR	RN	Autres (avec DVG)
Cluster 17	5-juillet	165-195	130-160	30-50	170-210	15-30
Ifop	3-5 juillet	155-185	120-150	50-65	170-210	15-33
Harris Interactive	4-5 juillet	168-198	115-145	32-63	185-215	15-30
Ipsos	3-4 juillet	145-175	118-148	57-67	175-205	22-28
Elabe	3-4 juillet	165-190	120-140	35-50	200-230	10 à 12
OpinionWay ^a	3-4 juillet	147,5-177,5	127,5-158,5	38-50	205-230	8 à 12
Odoxa	3-4 juillet	140-180	115-155	40-60	210-250	10 à 20
	moyenne	171,9	135,9	49,1	204,6	18,6
	résultat	178	165	66	143	25
	v.a. écart	6,1	29,1	16,9	61,6	6,4

^a Pour OpinionWay, on calcule la moyenne entre deux sondages réalisés les 3 et 4 juillet (LegiTrack, 4^{ème} vague pour Les Echos et radio classique et celui pour CNEWS, 5^{ème} vague)

Tous les instituts de sondages ont surestimé le résultat en sièges du RN et de ses alliés (Ciotti et les candidats RN-LN) et cela s'explique par le vote RN surestimé dans les différentes configurations possibles pour le second tour.

Les meilleures performances sont réalisées par l'Ifop, Cluster 17 et l'Ipsos qui sont les instituts qui ont le moins surestimé le nombre de sièges du RN. Le résultat en sièges d'Ensemble (et des divers centre) a été sous-estimé, de même que ceux de NFP et LR et DVD mais plus ou moins suivant les instituts de sondages.

On note que les projections en sièges se sont progressivement rapprochées du résultat final et on distingue les projections réalisées avant le premier tour qui donnaient une nette majorité relative en sièges pour le RN et ses alliés (UXD) avec même trois instituts (Harris Interactive, Odoxa et Elabe) qui envisageaient une possible majorité absolue du RN en sièges même si cela n'était pas le résultat le plus probable (maximum de la fourchette en sièges supérieur au seuil de la majorité absolue = 289 sièges). Ensuite, le désistement de nombreux candidats NFP ou ENS en cas de triangulaires avec le RN (ou UXD) a pénalisé le RN (ou UXD) car les duels (NFP / RN ou ENS / RN) sont moins favorables au RN que les triangulaires (NFP / ENS / RN).

Conclusion

Les instituts de sondages en France font tous leurs sondages d'intentions de vote par internet avec des personnes volontaires et en utilisant la méthode des quotas (coût faible, rapidité des sondages). Cette méthodologie peut être critiquée mais elle permet d'obtenir des résultats satisfaisants dans l'ensemble même si des erreurs sont possibles (sous-estimation du vote Mélenchon au premier tour de l'élection présidentielle de 2022 sans conséquences pour le duel Macron / Le Pen attendu et aussi surestimation du vote RN pour les élections régionales de 2021). Comme tous les instituts de sondages utilisent la même méthodologie (avec plus ou moins le souhait de maximiser le caractère aléatoire de l'échantillon), cette méthodologie qui est source de biais mais avec une variance d'échantillonnage réduite n'est pas remise en cause.

On peut dire que les performances des instituts de sondages sont très bonnes dans l'ensemble pour les élections européennes de 2024 en France et (assez) bonnes dans l'ensemble pour les élections législatives de 2024. Pour les élections législatives de 2024, le vote RN et alliés a été surestimé au premier tour (erreur de 2,87 pour la moyenne). La prévision en sièges avant le second tour a largement surestimé le nombre de sièges obtenu par le RN et ses alliés (erreur moyenne d'environ 62 sièges) : la

¹⁶ On trouve dans une version plus longue un tableau sur les performances (sièges).

force du front républicain (rejet du RN et de ses alliés) a été sous-estimé et il peut aussi y avoir une raison méthodologique pour la projection en sièges.

Suite aux erreurs des instituts de sondages en France (élections régionales de 2021 et élection présidentielle de 2022), il semble important de mieux faire connaître les travaux (prévisions) utilisant les sondages et des prévisions comme cela est fait notamment aux Etats-Unis (modèles de prévision de statisticiens, politologues, économistes) pour que les électeurs comprennent mieux la différence entre un sondage et une prévision.

Bibliographie

- Antoine, J., & Riandey, B. (1996). Les sondages d'intentions de vote. *Journal de la société française de statistique*, 137(4), 3-22.
- Ardilly, P. (2023). Peut-on se fier aux sondages empiriques ?, *Courrier des statistiques* N10, décembre, 113-132.
- Bethlehem, J.G. (1988). Reduction of the nonresponse bias through regression estimation, *Journal of Official Statistics*, 4, 251-260.
- Bethlehem, J. (2010). Selection bias in web surveys. *International statistical review*, 78(2), 161-188.
- Bethlehem, J. (2017a). The representativity of election polls. *Statistics, Politics and Policy*, 8(1), 1-12.
- Bethlehem, J. (2017b). Understanding public opinion polls. Chapman and Hall/CRC (sera consulté).
- Deville, J. C. (1991). Une théorie des enquêtes par quotas. *Techniques d'enquêtes, Statistique Canada*, 17(2), 177-95.
- Durand, C., & Blais, A. (2020). Quebec 2018: a failure of the polls?. *Canadian Journal of Political Science/Revue canadienne de science politique*, 53(1), 133-150.
- Durand, C., & Johnson, T. P. (2021). What About Modes? Differences Between Modes in the 21st Century's Electoral Polls Across Four Countries. *Public Opinion Quarterly*, 85(1), 183-222.
- Gelman, A. (2021). Failure and success in political polling and election forecasting. *Statistics and Public Policy*, 8(1), 67-72.
- Kenett, R. S., Pfeffermann, D., & Steinberg, D. M. (2018). Election polls—a survey, a critique, and proposals. *Annual Review of Statistics and Its Application*, 5, 1-24.
- Legleye, S. et al. (2023). Qui sont les membres des access panels? Une analyse à partir de TIC 2022. Communication présentée au colloque Sondages 2023.
- Lehingue, P. (2007). *Subunda. Coups de sonde dans l'océan des sondages*. Editions du croquant.
- Lynn, P., & Jowell, R. (1996). How might opinion polls be improved?: The case for probability sampling. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 159(1), 21-28.
- Mellon, J., & Prosser, C. (2017). Missing nonvoters and misweighted samples: explaining the 2015 great British polling miss. *Public Opinion Quarterly*, 81(3), 661-687.
- Prosser, C., & Mellon, J. (2018). The twilight of the polls? A review of trends in polling accuracy and the causes of polling misses. *Government and Opposition*, 53(4), 757-790.
- Sturgis, P., Kuha, J., Baker, N., Callegaro, M., Fisher, S., Green, J., ... & Smith, P. (2018). An assessment of the causes of the errors in the 2015 UK general election opinion polls. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(3), 757-781.