

L'estimation par intervalle de confiance dans les sondages

Florent Arnal

Dans cette capsule, je vais vous parler d'estimation par intervalle de confiance.

Dans un sondage de l'institut BVA, publié le 4 mai 2019, lié aux intentions de vote en France aux élections européennes de 2019 à 4 jours du scrutin, on pouvait lire

« **Une participation qui pourrait se situer légèrement au-dessus du niveau de 2014.** À quelques jours de l'échéance, entre **46% et 52% des Français déclarent avoir l'intention de se rendre aux urnes dimanche prochain**, une fourchette de participation qui situerait la mobilisation un léger cran au-dessus de la participation observée en mai 2014 (44% pour la France métropolitaine). »

Lors de la publication du sondage, une fiche technique explicitant, notamment, la méthodologie ainsi qu'une note sur les marges d'erreur était présentée.

On pouvait y lire que l'échantillon avait été constitué suivant les critères suivants : sexe, âge, profession de la personne de référence et de l'interviewé, région, catégorie d'agglomération.

L'institut de sondage considère alors que l'échantillon est représentatif de la population française en capacité de voter.

FICHE TECHNIQUE

 Recueil	Etude réalisée par l'institut BVA auprès d'un échantillon interrogé par Internet du 20 au 21 mai 2019.
 Echantillon	Echantillon de 1347 inscrits sur les listes électorales , identifiés au sein d'un échantillon de 1500 Français , représentatif de la population nationale âgée de 18 ans et plus. Critères de redressement : échantillon brut redressé sur critères sociodémographiques (sexe, âge, profession de la personne de référence et de l'interviewé, région, catégorie d'agglomération), sur le vote au 1 ^{er} tour de l'élection présidentielle de 2017 et aux élections européennes de 2014.
 Note intentions de vote	Les intentions de vote qui figurent dans ce rapport reposent sur la base des personnes inscrites sur les listes électorales, certaines d'aller voter et ayant exprimé une intention de vote, soit 678 individus . Pour cet effectif, pour un pourcentage obtenu par enquête de 20%, la marge d'erreur est égale à 3,1 points. Le pourcentage a donc 95% de chance d'être compris entre 16,9% et 23,1%. Les résultats d'intentions de vote ne constituent pas un élément de prévision du résultat électoral. Ils donnent une indication significative de l'état du rapport de forces à moins d'une semaine du scrutin européen.

La note sur les marges d'erreur nous alerte sur le fait que cette étude, comme toute étude statistique à partir d'un échantillon, même représentatif, présente des résultats soumis aux marges d'erreur inhérentes aux lois statistiques. Par exemple, à la lecture de ce tableau, on constate que lorsque l'on interroge environ 700 personnes, pour un pourcentage obtenu de 20%, la marge d'erreur est d'environ 3%. Ainsi, pour un pourcentage observé de 20%, on pourra considérer que le pourcentage exact dans la population est vraisemblablement compris entre 17% et 23%.

NOTE SUR LES MARGES D'ERREUR

Cette étude présente des résultats soumis aux marges d'erreur inhérentes aux lois statistiques. Le tableau ci-dessous présente les valeurs des marges d'erreur suivant le résultat obtenu et la taille de l'échantillon considéré.

TAILLE DE L'ÉCHANTILLON	INTERVALLE DE CONFIANCE À 95% SELON LE SCORE					
	5 OU 95%	10 OU 90%	20 OU 80%	30 OU 70%	40 OU 60%	50%
100	4,4	6,0	8,0	9,2	9,8	10,0
200	3,1	4,2	5,7	6,5	6,9	7,1
300	2,5	3,5	4,6	5,3	5,7	5,8
400	2,2	3,0	4,0	4,6	4,9	5,0
500	1,9	2,7	3,6	4,1	4,4	4,5
600	1,8	2,4	3,3	3,7	4,0	4,1
700	1,6	2,3	3,0	3,5	3,7	3,8
800	1,5	2,1	2,8	3,2	3,5	3,5
900	1,4	2,0	2,6	3,0	3,2	3,3
1 000	1,4	1,8	2,5	2,8	3,0	3,1
2 000	1,0	1,3	1,8	2,1	2,2	2,2
4 000	0,7	0,9	1,3	1,5	1,6	1,6
6000	0,6	0,8	1,1	1,3	1,4	1,4
10 000	0,4	0,6	0,8	0,9	0,9	1,0

Pour un pourcentage attendu voisin de 50% et un échantillon de 700 personnes et la marge d'erreur est d'environ 3,8%.

NOTE SUR LES MARGES D'ERREUR

Cette étude présente des résultats soumis aux marges d'erreur inhérentes aux lois statistiques. Le tableau ci-dessous présente les valeurs des marges d'erreur suivant le résultat obtenu et la taille de l'échantillon considéré.

TAILLE DE L'ÉCHANTILLON	INTERVALLE DE CONFIANCE À 95% SELON LE SCORE					
	5 OU 95%	10 OU 90%	20 OU 80%	30 OU 70%	40 OU 60%	50%
100	4,4	6,0	8,0	9,2	9,8	10,0
200	3,1	4,2	5,7	6,5	6,9	7,1
300	2,5	3,5	4,6	5,3	5,7	5,8
400	2,2	3,0	4,0	4,6	4,9	5,0
500	1,9	2,7	3,6	4,1	4,4	4,5
600	1,8	2,4	3,3	3,7	4,0	4,1
700	1,6	2,3	3,0	3,5	3,7	3,8
800	1,5	2,1	2,8	3,2	3,5	3,5
900	1,4	2,0	2,6	3,0	3,2	3,3
1 000	1,4	1,8	2,5	2,8	3,0	3,1
2 000	1,0	1,3	1,8	2,1	2,2	2,2
4 000	0,7	0,9	1,3	1,5	1,6	1,6
6000	0,6	0,8	1,1	1,3	1,4	1,4
10 000	0,4	0,6	0,8	0,9	0,9	1,0

Pour un même pourcentage attendu voisin de 50% et un échantillon de 1000 personnes, la marge d'erreur est d'environ 3,1%.

NOTE SUR LES MARGES D'ERREUR

Cette étude présente des résultats soumis aux marges d'erreur inhérentes aux lois statistiques. Le tableau ci-dessous présente les valeurs des marges d'erreur suivant le résultat obtenu et la taille de l'échantillon considéré.

TAILLE DE L'ÉCHANTILLON	INTERVALLE DE CONFIANCE À 95% SELON LE SCORE					
	5 OU 95%	10 OU 90%	20 OU 80%	30 OU 70%	40 OU 60%	50%
100	4,4	6,0	8,0	9,2	9,8	10,0
200	3,1	4,2	5,7	6,5	6,9	7,1
300	2,5	3,5	4,6	5,3	5,7	5,8
400	2,2	3,0	4,0	4,6	4,9	5,0
500	1,9	2,7	3,6	4,1	4,4	4,5
600	1,8	2,4	3,3	3,7	4,0	4,1
700	1,6	2,3	3,0	3,5	3,7	3,8
800	1,5	2,1	2,8	3,2	3,5	3,5
900	1,4	2,0	2,6	3,0	3,2	3,3
1 000	1,4	1,8	2,5	2,8	3,0	3,1
2 000	1,0	1,3	1,8	2,1	2,2	2,2
4 000	0,7	0,9	1,3	1,5	1,6	1,6
6000	0,6	0,8	1,1	1,3	1,4	1,4
10 000	0,4	0,6	0,8	0,9	0,9	1,0

On pourra noter que la marge d'erreur dépend du pourcentage observé ainsi que de la taille des échantillons.

Pour un pourcentage donné, plus l'échantillon est grand, plus cette marge est faible.

Dans l'exemple des intentions de vote, sur un échantillon de 1357 personnes, pour un pourcentage (observé) voisin de 49%, l'institut de sondage a pris une marge d'erreur d'environ 3% lui permettant d'annoncer que

« entre 46% et 52% des Français déclarent avoir l'intention de se rendre aux urnes dimanche prochain ».

Quelques jours plus tard, le taux de participation effectif était de 50,12%.

On constate donc que ce taux est bien compris entre 46% et 52%. Le pourcentage observé, voisin de 49%, certes cohérent, était quelque peu imprécis.

Nous pouvons, dès à présent, observer qu'il est préférable de donner une fourchette (intervalle avec une marge d'erreur) qu'une seule valeur (estimation ponctuelle).

Ceci est d'autant plus vrai lorsque l'on considère un sondage réalisé sur 1000 personnes qui annonce, comme c'était le cas pour ces élections européennes, 23% d'intention de vote pour une liste A et 22% pour une liste B.

En effet, si l'on se réfère au tableau précédent, la marge d'erreur est d'environ 2,5%.

Ainsi, il est vraisemblable que la liste A obtiendra entre 20,5% et 25,5% alors que la liste B obtiendra entre 19,5% et 24,5%.

Il est donc tout à fait possible, sans qu'il y ait la moindre erreur de méthodologie de l'institut de sondage, que la liste B recueille le jour de l'élection plus de voix que la liste A ...

Cet intervalle est appelé intervalle de confiance du pourcentage théorique. Dans ces exemples, la marge d'erreur (et donc l'intervalle) sont associés à un niveau de confiance de 95 %. Il s'agit du niveau de confiance utilisé couramment.

Si l'on prélevait une infinité d'échantillons, 95% des intervalles de confiance contiendraient la valeur vraie (pourcentage de français ayant l'intention d'aller voter à ce scrutin).