Sondages d'intention de vote : l'estimation des "marges d'erreur"

Léo Gerville-Réache [1]

[1] *Université de Bordeaux 2, CNRS, UMR 5251, Bordeaux, F-33000,*
leo.gerville@u-bordeaux2.fr

Déterminer et publier les marges d'erreur associées aux résultats d'un sondage politique est, pour le Sénat une nécessité de transparence scientifique (Portelli, 2010). Le problème réside dans le fait que, pour un échantillonnage par quotas, comme pour un échantillonnage probabiliste, le défaut de couverture, le taux de non-réponse, les indécis ou encore l'honnêteté des réponses sont autant de paramètres qui sont clairement susceptibles d'introduire des biais dans les estimations des intentions de vote. L'apparente impossibilité d'estimer ces biais et les efforts réalisés (quotas, redressement, reconstitution de votes antérieurs, ...) pour essayer de les minimiser conduisent les instituts de sondage à considérer ces biais comme négligeables. En conséquence chaque marge d'erreur est alors estimée par la célèbre formule "universelle" : $1.96\sqrt{\hat{p}(1-\hat{p})/n}$. En attendant 2012, on observait en 2007 (sans évoquer 2002), sur les sondages (particulièrement stables) du mois précédant le premier tour de l'élection présidentielle, que les estimations d'intentions de votes publiées par les 6 principaux instituts de sondage français surestimaient (resp. sous-estimaient) jusqu'à 5 % le résultat de Jean-Marie Le Pen (resp. Nicolas Sarkozy).

Le calcul de la marge d'erreur (demi intervalle de confiance au niveau $(1-\alpha)$), de l'estimation $\hat{p}_n$ d'une proportion p, issue d'un échantillon de n personnes est basé sur le théorème centrale limite et la loi normale (raisonnable quand $n \times p$ dépasse quelques unités).

On montre que l'estimation de la marge d'erreur (M_E) dépend de $\hat{p}_n$, de son écart-type, de la valeur absolue de son biais (noté B) et du niveau de confiance $(1-\alpha)$, par la formule approximative :

$$M_E \simeq z\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}},$$

avec z tel que :

$$\Phi\left(|B|/\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}+z\right)-\Phi\left(|B|/\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}-z\right)=1-\alpha,$$

où Φ est la fonction de répartition de la loi normale standard.

Cette communication développera l'analyse statistique des sondages d'intention de vote de 2007 et 2012 ainsi que la démonstration du calcul de la marge d'erreur avec biais. Une application sera présentée pour 2012, sur la base de 2007. Les biais de 2012 poseront alors des bases de 2017.

Bibliographie

Portelli, H., et Sueur, J.P. (2010). Rapport d'information de Sénat n°54 sur les sondages politiques.