

# Le sondage : un outil de formation efficace ?

Jean-Jacques Droesbeke<sup>1</sup> & Catherine Vermandele<sup>2</sup>

<sup>1</sup> Université libre de Bruxelles, LMTD, CP 139, avenue F.D. Roosevelt, 50, 1050 Bruxelles, Belgique, [jjdroesb@ulb.ac.be](mailto:jjdroesb@ulb.ac.be)

<sup>2</sup> Université libre de Bruxelles, LMTD, CP 139, avenue F.D. Roosevelt, 50, 1050 Bruxelles, Belgique, [vermande@ulb.ac.be](mailto:vermande@ulb.ac.be)

**Résumé.** La théorie des sondages est enseignée dans l'enseignement supérieur à différents niveaux. Elle fait généralement l'objet d'un cours à part entière, soit destiné à des statisticiens, soit construit de façon moins théorique pour des utilisateurs potentiels. Elle apparaît aussi parfois comme chapitre d'un cours de statistique générale dispensé à des étudiants de disciplines variées de sciences humaines. Au stade actuel, elle n'est pas du tout abordée dans l'enseignement secondaire. Dans l'« approche sondage », l'aléatoire est induit par la procédure aléatoire de génération de l'échantillon et ne fait donc pas référence à un modèle probabiliste donné. Cette approche permet ainsi de sortir du paradigme de l'« approche modèle » dans des contextes simples et de (re)donner du sens au vocabulaire de la statistique. Il apparaît ainsi que la théorie des sondages est probablement l'une des manières les plus efficaces d'introduire les concepts fondamentaux de la statistique inférentielle. L'objectif de notre communication est de le montrer par des exemples réels qui concernent aussi bien l'enseignement supérieur que le secondaire ou la formation continue.

**Mots-clés.** Enseignement de la statistique, approche sondage

## 1 La place actuelle de la théorie des sondages dans l'enseignement de la statistique et des probabilités

A ce jour, la théorie des sondages n'est pas du tout abordée dans l'enseignement secondaire. Dans l'enseignement supérieur, dans les filières qui ont pris le parti de l'enseigner, elle fait généralement l'objet d'un cours à part entière, soit destiné à des statisticiens, soit construit de façon moins théorique pour des utilisateurs potentiels (politologues, sociologues, économistes...). Ce cours apparaît dans le cursus d'un étudiant après qu'il ait déjà suivi l'un ou l'autre cours de statistique dans le(s)quel(s) lui ont été enseignés la statistique descriptive et les éléments classiques de la statistique inférentielle (estimation ponctuelle et par intervalle de confiance, tests d'hypothèses...).

C'est bien souvent dans le cours explicitement consacré à la théorie des sondages que les étudiants entendent pour la première fois parler des méthodes de sondages aléatoires simples et complexes les plus fréquemment mises en œuvre, des différents estimateurs utilisés (ils sont tout étonnés de découvrir que l'on n'estime pas toujours la moyenne-population par la moyenne arithmétique des observations réalisées dans l'échantillon !) et de leurs caractéristiques. Ce cours donne généralement aussi l'occasion d'aborder la problématique du calage et du redressement. Il permet enfin de mettre en avant les différentes erreurs qui peuvent être liées à la procédure d'enquête par sondage et, dans ce cadre, étudie quelques aspects du traitement de la non-réponse.

## 2 Le sondage : un outil de formation efficace ?

La statistique enseignée aujourd'hui dans le secondaire ou à l'université (en France comme en Belgique) traite essentiellement de variables réelles et repose sur la modélisation probabiliste et l'échantillonnage i.i.d. (variables aléatoires indépendantes et identiquement distribuées) (*approche modèle*), alors que la statistique des sciences humaines et sociales s'intéresse très souvent à des

variables catégorielles définies sur des populations finies, desquelles il est possible d'extraire des échantillons en utilisant des procédures aléatoires (*approche sondage*). Dans cette « approche sondage », l'aléatoire est induit par la procédure aléatoire de génération de l'échantillon et ne fait donc pas référence à un modèle probabiliste donné. L'approche « sondage » permet ainsi de sortir du paradigme de l'approche « modèle » dans des contextes simples et, par là-même, de (re)donner du sens au vocabulaire de la statistique (population, individus, échantillons indépendants et échantillons appariés, variables aléatoires, estimation, fluctuation et erreur d'échantillonnage, ...). Une fois l'approche « sondage » bien comprise, les élèves ou étudiants sont beaucoup plus à même d'appréhender l'approche « modèle » et apparaissent moins décontenancés par son caractère plus abstrait.

Il apparaît ainsi que la théorie des sondages est probablement l'une des manières les plus efficaces d'introduire les concepts fondamentaux de la statistique inférentielle. L'objectif de notre communication est de le montrer par des exemples réels qui concernent aussi bien l'enseignement supérieur que le secondaire ou la formation continue.

### **3 Quelle place donner à la théorie des sondages dans un cours de statistique générale de base dans l'enseignement supérieur ?**

Depuis de très nombreuses années déjà, le cours de statistique générale (« Eléments de statistique ») dispensé à l'Université libre de Bruxelles aux étudiants de disciplines variées de sciences humaines (sciences sociales, sciences humaines, sciences politiques, droit, criminologie, information et communication...) contient un chapitre intitulé « Les méthodes de sondage ». Celui-ci permet à l'enseignant de faire le pont, de manière naturelle et concrète, entre les chapitres précédents consacrés aux éléments de la théorie des probabilités et aux variables aléatoires, et les chapitres suivants portant sur l'estimation ponctuelle et par intervalle de confiance. Ce chapitre propose notamment une introduction à la procédure d'estimation fondée sur un échantillon obtenu par  $n$  tirages aléatoires, tantôt *avec* remise, tantôt *sans* remise, dans une population finie. Les notions d'échantillonnage aléatoire simple, de plan de sondage et de probabilités d'inclusion peuvent être aisément illustrées (et donnent l'occasion de mettre en pratique des raisonnements probabilistes simples et des règles de dénombrement étudiées auparavant). Le principe de l'estimation est d'autant plus facilement compris par les étudiants que les paramètres à estimer ont un sens concret, puisqu'ils correspondent à des caractéristiques de la distribution d'une certaine variable d'intérêt dans la population. Le caractère aléatoire de l'estimateur peut être facilement expliqué, puisque directement induit par le caractère aléatoire de la procédure de prélèvement de l'échantillon. On peut aisément montrer, à l'aide d'exemples numériques de petite taille, à quoi correspond la distribution d'échantillonnage d'un estimateur et, au travers de l'étude de cette dernière, introduire les notions d'erreur d'échantillonnage, de fluctuation d'échantillonnage, de biais d'un estimateur et de mesure de la précision de celui-ci par l'intermédiaire de sa variance ou de son erreur quadratique moyenne. La comparaison des variances des estimateurs dans le cas du sondage à probabilités égales avec et sans remise permet de faire réfléchir les étudiants sur les caractéristiques de la population, de l'échantillon et de la procédure d'échantillonnage qui ont un impact sur la précision de ces estimateurs. Ces réflexions initient la présentation (beaucoup plus succincte) d'autres méthodes de sondages aléatoires classiques. Une fois les notions de fluctuation d'échantillonnage et de distribution d'échantillonnage d'un estimateur bien assimilées par les étudiants, il est beaucoup plus aisé de leur faire comprendre ce qu'est un intervalle de confiance et comment interpréter le niveau de confiance qui lui est associé.

Il nous semble que cette première approche de l'estimation au travers de la théorie des sondages, de par son caractère très concret, facilite la prise de conscience par les étudiants que toute décision fondée sur les résultats d'une procédure d'inférence statistique s'accompagne d'un risque, mais que ce risque peut être évalué et, dans une certaine mesure, contrôlé ou limité.

## 4 La théorie des sondages a-t-elle aussi sa place dans l'enseignement des probabilités et de la statistique dans le secondaire ?

Convaincus de l'efficacité de la théorie des sondages comme outil d'initiation des jeunes (et moins jeunes...) à l'analyse statistique et au raisonnement probabiliste, nous avons soumis une proposition concrète de projet à la Ville de Bruxelles (l'un des pouvoirs organisateurs de l'enseignement secondaire en Fédération Wallonie-Bruxelles). Ce projet a été mis en œuvre en 2012-2013 et en 2013-2014 dans des classes de sixième (dernière année) du secondaire (en sciences humaines et en maths fortes) de deux lycées différents de la Ville de Bruxelles. Les différentes phases de ce projet sont les suivantes :

- (i) Les élèves de sciences humaines, dans le cadre de leur cours de sciences sociales, choisissent un thème d'étude sur lequel ils vont récolter des données par enquête auprès de tous leurs condisciples de 5<sup>e</sup> et 6<sup>e</sup> année de leur école (en 2012-2013, l'étude portait sur l'utilisation d'internet par les jeunes de fin du secondaire ; en 2013-2014, l'étude s'est penchée sur les connaissances de la vie politique belge qu'ont les jeunes de fin du secondaire et sur les sources utilisées par ces derniers pour s'informer sur cette vie politique). Ils élaborent, avec l'aide de leur professeur de sciences sociales (et sous notre œil attentif) un questionnaire qu'ils vont administrer eux-mêmes dans les différentes classes concernées.
- (ii) L'encodage des réponses au questionnaire est réalisé par les élèves de sciences humaines.
- (iii) Le professeur de mathématiques réalise avec ces élèves une analyse statistique descriptive des réponses à l'enquête. Les analyses statistiques sont ensuite rediscutées en collaboration avec le professeur de sciences sociales afin de voir dans quelle mesure elles permettent d'éclaircir la problématique de départ.
- (iv) La 4<sup>e</sup> phase concerne les élèves de maths fortes. Il leur est proposé de considérer l'ensemble des répondants (environ 200) à l'enquête réalisée par leurs condisciples de sciences humaines comme population de référence et de s'intéresser au processus d'estimation de proportions de cette population. Après une brève présentation de ce en quoi consistent le processus d'estimation et la fluctuation d'échantillonnage, on leur sélectionne, par tirages aléatoires sans remise dans la population, 100 échantillons de taille 20 (1<sup>er</sup> ensemble d'échantillons) et 100 échantillons de taille 50 (second ensemble d'échantillons) et on détermine les estimations de trois proportions-population ( $p_a < 0.25$ ,  $p_b \sim 0.5$  et  $p_c > 0.75$  ; les valeurs de ces trois proportions sont données aux élèves) dans chacun des 200 échantillons prélevés. On se retrouve ainsi avec 6 séries statistiques d'estimations de taille 100 (une série par ensemble d'échantillons et par proportion-population considérée). Les élèves sont alors divisés en 6 groupes et chaque groupe est amené à analyser l'une des 6 séries statistiques obtenues : avec l'aide de leur professeur de mathématiques, les élèves doivent déterminer la moyenne, la médiane, la variance, l'écart interquartile, les quantiles d'ordres 0.025 et 0.975 de la série, et doivent construire un histogramme. Il leur est ensuite demandé de comparer et d'analyser les résultats obtenus par les différents groupes. L'objectif de cette phase est de faire toucher du doigt par les élèves, en utilisant exclusivement des outils de statistique descriptive qu'ils sont censés maîtriser et sans faire appel au moindre modèle probabiliste, ce que signifient la fluctuation d'échantillonnage et l'absence de biais de l'estimateur ; nous souhaitons également qu'ils découvrent par eux-mêmes l'impact de la taille de l'échantillon et de la valeur de la proportion estimée sur la précision de l'estimateur.

Les phases (i) à (iii) du projet permettent d'initier les élèves de sciences humaines à la

démarche statistique et de les confronter à de nombreux aspects de la méthodologie d'enquête. Elles permettent également de positionner l'analyse statistique dans un contexte pluridisciplinaire, ce qui s'avère particulièrement motivant aussi bien pour les élèves que pour leurs professeurs.

La phase (iv) ouvre les étudiants de maths fortes à la théorie des sondages et suscite leur curiosité sur les difficultés liées aux procédures d'échantillonnage et d'estimation.

## 5 Perspectives

L'expérience a de fortes chances d'être renouvelée durant la prochaine année scolaire. Outre les nécessaires améliorations techniques et pédagogiques qui résultent des deux expériences réalisées à ce jour, nous comptons ajouter un volet supplémentaire accessible pour les élèves et de faible coût en temps. Il s'agit d'ouvrir la porte vers la stratification d'une part, y compris la post-stratification et le concept de redressement, et l'influence des non réponses d'autre part. Cette extension est actuellement à l'étude.

Nous tenterons, au cours de notre communication, d'analyser de manière objective l'intérêt que peut avoir, pour l'enseignement de la statistique, l'introduction de la théorie des sondages dans l'enseignement secondaire et dans un cours de statistique de base dans l'enseignement supérieur.

## Bibliographie

- [1] Dehon, C., Droesbeke, J.-J. et Vermandele, C. (2008), *Eléments de statistique* (5<sup>e</sup> édition revue et augmentée), Editions de l'Université de Bruxelles, Bruxelles, Editions Ellipses, Paris.
- [2] Droesbeke, J.-J., Gaspar, R. D., Salmon, C. et Vermandele, C. (2013), Réalisation d'une enquête par sondage dans le cadre de la formation scolaire à l'Athénée Emile Bockstael en 2012-2013 : synthèse et bilan, Rapport de fin de projet, Université libre de Bruxelles et Ville de Bruxelles.
- [3] Fine, J. (2010), Probabilités et statistique inférentielle : approche sondage versus approche modèle, *Statistique et Enseignement*, **1**(2), 5-21.