

PRÉSENTATION DU CHALLENGE « GRAINES DE SONDEUR »

Camelia GOGA¹ et Pascal ARDILLY²

TITLE

Presentation of the challenge "Graines de sondeur"

RÉSUMÉ

Afin de faire partager aux lycéens les attraits de la statistique, la Société Française de Statistique (SFdS) et l'Institut de Recherche sur l'Enseignement des Mathématiques (IREM) de Dijon ont mis en place pour la première fois en 2014 le challenge de statistique « Graines de sondeur ». Ce challenge était consacré au traitement de données issues d'une enquête par sondage ainsi qu'à diverses problématiques liées à l'échantillonnage, et il a été organisé dans le cadre du 8^e Colloque francophone sur les sondages (Université de Bourgogne, 18-20 novembre 2014). Il a été proposé à tous les lycéens de la voie générale, technologique ou professionnelle de l'académie de Bourgogne. La première édition a mis en évidence certaines difficultés tenant aux élèves... ou aux professeurs de mathématiques. Le challenge a été étendu en 2015 et 2016 à cinq autres académies, et en 2017 il se déroule dans huit académies. Nous présentons dans cet article la mise en place et le déroulement du challenge, ainsi qu'un premier bilan de ces premières éditions.

Mots-clés : algorithmes, enquête par sondage, IREM, lycéens.

ABSTRACT

In order to make statistics appealing to high-school pupils, the French Society of Statistics (SFdS) and the Research Institute of Teaching Mathematics (IREM) from Dijon organized for the first time in 2014 the challenge in statistics intitled "Graines de sondeur". This challenge was concerned by the treatment of survey data as well as by various sampling issues, and it was organized during the 8th French Survey Sampling Colloquium (University of Burgundy, november 18-20, 2014). It was proposed to all general, technological and professional high-schools from the region of Burgundy. The first edition of this challenge showed that several difficulties were encountered by pupils as well as by their mathematics teachers. The challenge was carried on in five French regions on 2015, 2016 and in eight regions on 2017. In this article, we present its organisation and development. A first assessment of these editions is also drawn.

Keywords: algorithms, survey sampling, IREM, high-school pupils.

1 Bref historique et contexte

De nos jours, la statistique occupe une place de plus en plus importante dans la vie courante et de nombreux domaines tels que l'économie, la médecine, la biologie, l'industrie, l'agriculture ont besoin des études statistiques. Pour répondre aux nouveaux enjeux de la société, la statistique a pris une place importante dans les programmes de mathématiques de lycée. Néanmoins, les notions enseignées sont essentiellement consacrées à la statistique descriptive (moyenne, variance, histogramme, médiane) et elles sont testées par les élèves sur des jeux de données de tailles relativement faibles. La notion d'intervalle de fluctuations est

¹ LMB, Université de Bourgogne Franche-Comté, Besançon, camelia.goga@univ-fcomte.fr

² Institut national de la statistique et des études économiques, Lyon, pascal.ardilly@insee.fr

Présentation du challenge « Graines de sondeur »

enseignée à partir de la classe de seconde, de façon très schématique et sans fondement probabiliste. La loi normale est pour sa part introduite en classe de terminale. A partir de la seconde, les élèves sont donc confrontés au caractère aléatoire des expériences ou aux théorèmes limites, notions nouvelles qui s'avèrent abstraites et plus difficiles à comprendre que les concepts utilisés en géométrie ou en algèbre. En particulier, les simulations sont souvent effectuées sur des jeux de données de trop petite taille.

Pascal Ardilly assiste en 2012, lors des Journées de la Sociedade Portuguesa de Estatística (SPE), à la remise des prix du concours national de statistique adressé aux collégiens et lycéens portugais. Ce concours se présente sous la forme d'un projet en statistique réalisé sur des données réelles fournies par différents groupes de la SPE. Les équipes lauréates présentent leurs travaux en poster durant le congrès. Le président de la SPE décerne les prix et les cadeaux (par niveau et section) en séance plénière. Ce concours est très apprécié par les élèves participants et bénéficie d'une large implication et d'un soutien actif de la part des membres de la SPE. D'autres pays européens (Italie, Espagne, ...) organisent des concours nationaux de statistique pour les lycéens et/ou les collégiens sous la forme d'un projet statistique ou des QCM (questionnaire à choix multiples). En France, il existe plusieurs compétitions en mathématiques (rallyes de mathématiques, olympiades, concours Kangourou), mais les sujets de statistique sont peu représentés. Cela est probablement dû au fait que la statistique n'est présente dans les programmes de mathématiques de lycée que depuis peu de temps et que les enseignants ne possèdent pas forcément une formation suffisante dans cette discipline.

En s'appuyant sur l'expérience portugaise, Pascal Ardilly propose alors à Benoît Riandey, président du groupe « Enquêtes, Modélisations et Applications » de la SFdS, que cette dernière organise une manifestation du même type en France. L'idée est reçue avec enthousiasme par les groupes « Enquêtes, Modélisations et Applications » et « Enseignement de la statistique » mais, dans un premier temps, elle reste sans suite. En 2013, Camelia Goga, présidente du comité d'organisation du 8^e Colloque francophone sur les sondages, et Pascal Ardilly prennent l'initiative d'organiser une manifestation statistique adressée aux lycéens en lien avec les sondages en s'appuyant sur le modèle portugais. Avec le soutien d'Anne Gégout-Petit, présidente de la SFdS, cette proposition a été soumise au conseil de la SFdS et approuvée. Néanmoins, cette mise en place pour la première fois d'un concours en statistique a duré plusieurs mois car de nombreuses questions pratiques se sont posées, en commençant par la constitution de l'équipe organisatrice et du jury d'évaluation, la recherche d'un nom évocateur pour le challenge, le choix de la forme du concours (QCM, projet...), l'élaboration des premiers sujets, la politique de promotion à appliquer auprès des établissements d'enseignement, etc. Les membres du groupe « Enquêtes, Modélisations et Applications » ont été sollicités et, après de nombreux échanges par courriel, le nom « Challenge Graines de sondeur » a été choisi. Il a été décidé qu'il prendrait la forme d'un projet statistique proposé à des équipes de quatre élèves maximum, encadrées par des professeurs de mathématiques.

Le déroulement du challenge a été piloté par un comité formé par Pascal Ardilly (Insee), Guillaume Chauvet (Ensaï/Crest), Jérôme Cubillé (Edf Palaiseau), Camelia Goga et Catherine Labruère-Chazal (Université de Bourgogne). Pour une élaboration des sujets conforme avec les programmes de mathématiques et les intérêts des jeunes participants, le comité de pilotage a fait appel à Yannick Prot, professeur de mathématiques au lycée Marey à Beaune.

2 Déroutement de la première édition du challenge en 2014

L'IREM de Dijon a lancé, en février 2014, le premier appel à participation au challenge après avoir obtenu l'accord et le soutien du rectorat et des inspecteurs de mathématiques de l'académie de Dijon. Plus précisément, une lettre contenant tous les renseignements pratiques concernant le challenge a été envoyée aux proviseurs de tous les lycées de l'académie de Dijon. Une page internet (<http://sondages2014.sfds.asso.fr/graines-de-sondeur>) avait été spécialement créée sur le site du Colloque Sondages pour les besoins du challenge. Les équipes intéressées trouvaient sur cette page tous les renseignements pratiques importants : les objectifs du challenge, l'affiche à télécharger, les sujets, le calendrier, le cahier des charges et les modalités d'évaluation des travaux, ainsi que des documents de vulgarisation de la théorie des sondages, des références bibliographiques et une « hot line » (gds.sondages2014@sfds.asso.fr) destinée à les aider en cas de difficulté.

Les élèves désirant s'inscrire au challenge devaient envoyer avant le 10 avril 2014 un message à gds.sondages2014@sfds.asso.fr en spécifiant la composition de leur équipe, le professeur, leur lycée d'affectation ainsi que le sujet choisi. Les équipes avaient ensuite trois mois, jusqu'au 10 juillet 2014, pour élaborer leur projet. Courant septembre 2014, un courriel précisait aux équipes participantes que l'équipe lauréate serait invitée à présenter son travail lors du 8^e Colloque Francophone sur les Sondages, à Dijon, du 18 au 20 novembre 2014.

Les sujets proposés portaient sur différents aspects de la méthodologie des enquêtes par sondage et de vrais jeux de données avaient été mis à la disposition des équipes.

2.1 Objectifs du challenge

Les objectifs majeurs du challenge étaient de faire découvrir aux jeunes lycéens la statistique en « apprivoisant » le hasard et en fouillant de vrais jeux de données statistiques, ainsi que le travail de recherche en équipe des statisticiens d'enquête. Les jeunes statisticiens pouvaient ainsi devenir pour quelques mois un (ou une) Sherlock Holmes et mener leur propre enquête statistique aux résultats parfois surprenants. Voici un extrait de la présentation du challenge telle qu'elle figure sur la page d'accueil du site internet du challenge ainsi qu'une capture d'image de cette page (voir Figure 1):

Êtes-vous passionné(e)s par le mystère des chiffres ? Par le hasard et la statistique ? Vous allez pouvoir endosser le rôle de chercheur et mener votre propre enquête statistique aux résultats parfois surprenants ! Inscrivez-vous au challenge et découvrez le statisticien sondeur qui sommeille en vous.

Les « plus » du challenge :

- *découvrir le travail de recherche en équipe ainsi que celui de statisticien ;*
- *fouiller les données et apprivoiser le hasard en s'amusant ;*
- *découvrir le milieu universitaire et celui des statisticiens d'enquête.*

En 2014, le comité avait décidé d'attribuer les prix sur la base du rapport écrit des élèves ainsi qu'en jugeant la présentation orale d'une vingtaine de minutes de leur projet. Pour cela, il était prévu d'inviter les meilleures équipes à exposer leur travail au 8^e Colloque francophone sur les sondages et de décider du classement seulement après les présentations orales. Finalement, une seule équipe s'est distinguée par son rapport écrit et a pu ainsi effectuer une présentation orale. Avec l'aide du comité de programme du colloque, une

Présentation du challenge « Graines de sondeur »

session spécialement dédiée au challenge avait été prévue dans le programme scientifique du colloque pour permettre cette présentation. Le comité a considéré que le colloque était l'occasion parfaite pour faire connaître aux lycéens le milieu des chercheurs et des praticiens issus des instituts publics ou privés.



FIGURE 1 – Page d'accueil du site internet du challenge

2.2 Sujets et cahier des charges

En 2014, cinq sujets ont été proposés, couvrant différents aspects de la méthodologie des enquêtes par sondage et diverses problématiques liées à l'échantillonnage. Ces sujets sont accessibles via l'adresse <http://sondages2014.sfds.asso.fr/graines-de-sondeur/>. Un glossaire contenant quelques définitions de notions statistiques employées dans les descriptifs des sujets comme, par exemple, base de sondage, estimation, biais, ... était également fourni aux participants. Les données – pour les sujets qui s'appuyaient sur des données fournies par les organisateurs – étaient accessibles par un lien offert sur le site du challenge. Chaque sujet suggérait quelques pistes de recherche mais l'objectif était de laisser les équipes proposer une méthode originale. Les sujets proposés allaient au-delà des programmes de statistique du lycée et il était conseillé (et attendu) que les élèves consultent les documents bibliographiques disponibles ou référencés sur le site.

Pour l'édition de 2014, l'évaluation des projets a été faite par les membres du comité de pilotage, qui constituaient également le jury. Un cahier des charges avait été élaboré et donnait des instructions sur la forme attendue du projet : volumétrie, structure (introduction, contenu, conclusion). Néanmoins, toute liberté était laissée sur le choix du logiciel de traitement de texte ainsi que sur celui des outils informatiques utilisés pour réaliser le projet : logiciels ou langages de programmation (tableur, Algobox, Python, ...).

Nous donnons ci-dessous une brève description de ces cinq sujets.

Sujet 1 : « Estimation du niveau en mathématique d'un lycée »

Les élèves de Première passent en janvier 2014 une évaluation nationale en mathématiques. Les résultats sont communiqués de façon individuelle en mars, et le lycée reçoit les notes de l'ensemble des élèves début juillet.

En avril, le proviseur souhaite obtenir une estimation de la réussite de son établissement en se basant sur un échantillon de 50 élèves. Il dispose également pour chaque élève de sa note moyenne en mathématiques obtenue lors des épreuves internes du lycée pour le dernier trimestre de l'année 2013.

L'objectif du projet était de proposer des stratégies au proviseur, et de mesurer la pertinence de ces stratégies *a posteriori* quand tous les résultats sont communiqués au lycée début juillet.

Sujet 2 : « Estimation de l'impôt moyen sur le revenu par foyer »

A l'été 2010, le conseil régional de Bourgogne souhaite estimer l'impôt sur le revenu payé en moyenne par foyer fiscal en 2010, correspondant aux revenus de 2009. Il a pour cela la possibilité de contacter le centre des impôts de 100 communes, pour obtenir pour chacune d'entre elles l'impôt global acquitté et le nombre total de foyers qui ont déposé une déclaration. Il dispose également pour chaque commune, au travers du site internet de l'Insee, des données communales de déclarations de revenu des années 2008 et antérieures, et des données du recensement.

L'objectif du projet était de proposer des stratégies au conseil régional, et de mesurer la pertinence de ces stratégies *a posteriori* quand les statistiques communales exhaustives sont disponibles début 2011.

Sujet 3 : « Enquête sur l'usage d'alcool, de tabac, de drogues »

Une enquête est mise en place pour mesurer la proportion et les caractéristiques des lycéens consommant ou ayant consommé de l'alcool, du tabac ou de la drogue. Un échantillon de lycées est pour cela tiré au sort. Après saisie des questionnaires, les données sont mises à la disposition des élèves sous la forme d'un tableau où chaque ligne correspond à un questionnaire, et chaque colonne à la réponse à une question.

L'objectif de ce sujet était de faire réfléchir les élèves aux traitements des données manquantes. Les élèves devaient remarquer que dans le cas de cette enquête, certains individus n'avaient pas rendu leur questionnaire (non-réponse totale) tandis que d'autres individus n'avaient répondu qu'à certaines questions (non-réponse partielle). Les élèves devaient estimer des moyennes (ou des proportions) de variables associées aux différentes questions posées sur le thème principal de l'enquête, en proposant des solutions argumentées aux différents problèmes pouvant se poser, notamment du fait des réponses manquantes.

Sujet 4 : « Algorithmes de sélection d'un échantillon »

L'étude statistique d'une population se base généralement sur l'étude d'un échantillon sélectionné de façon aléatoire dans la population. Le tirage avec remise est une méthode possible de sélection, mais en pratique on se base plutôt sur des méthodes de tirage sans remise. Parmi les méthodes de sélection sans remise les plus utilisées, deux méthodes sont

Présentation du challenge « Graines de sondeur »

considérées dans ce sujet : l'échantillonnage systématique et l'échantillonnage simple sans remise.

Les objectifs de ce sujet consistent à implémenter ces trois algorithmes (avec et sans remise) pour une population de taille N quelconque, ainsi qu'à étudier leurs propriétés et à les comparer.

Sujet 5 : « Réaliser une enquête sur un thème libre »

Dans ce sujet, les élèves sont laissés libres de concevoir une enquête à mener auprès d'un échantillon de personnes physiques, sur un thème de leur choix. Néanmoins, l'énoncé du sujet les mettait en garde sur les différentes difficultés à surmonter :

- la définition du sujet de l'enquête et de la population visée,
- la recherche préalable d'études ou d'informations sur le thème retenu,
- le choix du mode de recueil des données, la rédaction du questionnaire,
- la sélection de l'échantillon,
- la réalisation de l'enquête
- l'analyse des résultats,
- l'appréciation de la qualité des résultats.

Il était demandé aux élèves de préciser dans leur rapport (en les justifiant) les choix qu'ils avaient effectués pour chacun des points énumérés ci-dessus.

2.3 Ressources

Les compétences requises dans la résolution des sujets proposés dépassaient très clairement les connaissances en statistique exigées par les programmes de lycée. Pour venir en aide aux élèves et aux professeurs de mathématiques, plusieurs articles et exposés de vulgarisation de la théorie des sondages ont été proposés aux participants sous forme numérique :

1. un extrait de l'article « *Les sondages : délaissés par les statisticiens et malmenés par les politologues* » de Jeanne Fine et Jean-Louis Piednoir³, Bulletin de l'APMEP n° 474, 2008, accessible aux élèves sans aide supplémentaire ;
2. des ressources nécessitant un niveau technique plus élevé pour les professeurs de mathématiques qui pouvaient ensuite guider leurs élèves dans leurs démarches statistiques :
 - le livre « *Les sondages : principes et méthodes* » de A.-M. Dussaix et J.-M. Grosbras, 1993, collection *Que sais-je ?* ;
 - l'article « *Peut-on croire aux sondages ?* » de Jean-Claude Deville paru en juin 2006 dans la revue *Pour la science*, n° 344 ;
3. une présentation filmée de Guillaume Chauvet donnée dans le cadre des formations de l'IREM de Rennes en 2007.

Les élèves et leurs professeurs de mathématiques pouvaient contacter à tout moment l'équipe organisatrice via l'adresse email gds.sondages2014@sfds.asso.fr pour toute question théorique liée aux sujets.

³ L'article présent sur le site, paru dans la revue N°63 de la Mission laïque française et signé par Jean-Louis Piednoir, est en réalité adapté de cet article original.

3 Poursuite du challenge, après 2014

Ce challenge a été reconduit en septembre 2015 dans cinq autres académies : Bordeaux, Dijon, Lyon, Montpellier et Nancy. L'académie de Besançon a rejoint le groupe en septembre 2016, suivie des académies de Grenoble et de Strasbourg pour la 4^e édition, en 2017. Le nouveau site du challenge a désormais l'adresse <http://graines-de-sondeur.sfds.asso.fr>.

La mise en place dans plusieurs académies a occasionné des changements par rapport à la première édition. L'équipe organisatrice est formée actuellement des directeurs de chaque IREM participant et s'occupe de la mise en place pratique du challenge : demande d'accord pour le lancement du challenge auprès des recteurs des académies, relation avec les inspecteurs de mathématiques, envoi des lettres dans les établissements et, éventuellement, prise de contact avec des enseignants. Une deuxième équipe, formée par des experts en statistique (actuellement P. Brion, SFdS ; G. Chauvet, Ensai ; M. Christine, INSEE ; G. Cubille, EDF ; C. Goga, UFC ; P. Joly, INSEE) a été mise en place pour répondre aux besoins méthodologiques du challenge : élaboration des sujets, réponses aux questions plus techniques posées par les équipes (via le courriel gds2015@sfds.asso.fr) et correction des projets. Les deux équipes participent à la lecture des projets et un classement est réalisé dans chaque académie. Le challenge a été globalement coordonné par Camelia Goga et Pascal Ardilly pour les trois premières éditions et, à partir de cette 4^e édition, Philippe Brion remplace Pascal Ardilly.

Le challenge se déroule maintenant sur une année scolaire : il est lancé en septembre et les équipes désirant y participer doivent envoyer un message à l'IREM de leur académie avant le 15 novembre. La date limite de l'envoi des travaux au correspondant IREM de leur académie est fixée début mars (de la même année scolaire). Les présentations des meilleurs projets sont programmées lors de la *Semaine des mathématiques* (14-20 mars 2016, 13-19 mars 2017, 12-18 mars 2018) dans chaque académie pour permettre aux jeunes élèves de visiter l'université et d'assister à quelques conférences.

Deux nouveaux sujets ont été conçus en 2016 : un premier sujet propose d'estimer les intentions de vote pour le futur délégué dans une classe (en lien avec les élections présidentielles de l'année dernière) et un second sujet traite de l'estimation du nombre de Poks lancés dans une ville par une agence de jeux vidéo (en lien avec les enquêtes spatiales). En 2017, le comité a finalement retenu quatre sujets : un sujet nouveau concernant l'estimation des proportions des élèves d'un lycée qui aiment certaines séries de télévision et trois sujets anciens (sujets n° 3, 4 et 5 de la première édition).

Des prix très attractifs (tablettes numériques, chèques cadeaux) ont été proposés pour la première édition du challenge grâce aux subventions du colloque et de l'IREM de Dijon. Pour les éditions suivantes, les prix ont pu être accordés aux équipes lauréates grâce à la participation des IREM et aux subventions du groupe « Enquêtes, Applications et Modèles » de la SFdS. Une remise d'un prix national est prévue chaque année lors des Journées de la Statistique de la SFdS, mais elle n'a encore jamais été organisée car elle est difficile à mettre en œuvre et le nombre de très bons projets a été faible jusqu'à maintenant.

Enfin, trois formations destinées aux enseignants de mathématiques ont été données respectivement en 2015 à Bordeaux par Marthe-Aline Juthand, à Lyon par Pascal Ardilly avec l'aide de l'APMEP (Association des Professeurs de Mathématiques de l'Enseignement Public) et à Dijon par Camelia Goga avec l'aide de l'IREM de Dijon. Cette dernière, filmée, est accessible en ligne sur le site du challenge.

4 Bilan des quatre premières éditions du challenge, après 2014

Malheureusement, peu d'équipes ont participé à la première édition du challenge et seulement une équipe a été invitée au 8^e Colloque francophone sur les sondages (Dijon, 2014) pour présenter son projet. L'équipe, formée par quatre élèves de seconde, avait choisi le sujet 3, un sujet assez difficile qui demandait le traitement des données manquantes. Les élèves n'ont pas vraiment traité la problématique des données manquantes mais néanmoins, ils ont réalisé une bonne analyse descriptive des données et ils ont fait une très bonne présentation orale lors du colloque. D'ailleurs, lors de leur présentation orale, donc six mois après la remise de leur projet écrit, les élèves avaient beaucoup gagné en maturité et montraient qu'ils commençaient à entrevoir par eux-mêmes les bonnes solutions aux problèmes techniques posés par le sujet.

Le comité de pilotage pense que la faible participation en 2014 a été due en grande partie au retard du lancement du challenge (février 2014 au lieu de septembre 2013) et au fait qu'il se déroulait sur deux années scolaires.

En 2015, une cinquantaine d'équipes se sont inscrites au challenge et une vingtaine d'équipes (toutes localisées dans les académies de Bordeaux et de Nancy) ont finalement rendu leurs travaux. Les présentations ont eu lieu dans les locaux des universités de Bordeaux et de Nancy, en présence de la presse locale. A Bordeaux, deux classes de seconde ont présenté leurs travaux, et lors de cette journée, les élèves ont pu assister à la conférence « *Notions et intuitions sur les sondages* » donnée par Anne Ruiz-Gazen, professeur en statistique à l'Université Toulouse 1. Il est à noter que l'équipe gagnante de l'académie de Bordeaux avait réalisé une enquête sur la « Cop 21 » et avait écrit à ce sujet au président de la république, François Hollande.

Globalement, la participation au challenge a été plus faible en 2016 mais il faut mentionner la très bonne participation d'une classe de seconde du lycée Lamartine de Macon en Bourgogne. Cette classe a bénéficié d'un encadrement fort de la part de trois enseignantes de mathématiques et de sciences humaines. Lors de la remise de prix à l'université de Bourgogne, A. Ruiz-Gazen a été invitée de nouveau pour donner sa conférence sur les sondages aux lycéens.

Les sujets préférés des élèves s'avèrent être ceux qui traitent respectivement de l'enquête sur la consommation des drogues (sujet n° 3) et de la mise en place d'une enquête sur un thème libre (sujet n° 5). En général, les quelques professeurs qui ont adhéré au projet ont présenté plusieurs équipes de leur classe. Il y a donc eu un fort « effet de grappe » touchant la participation. Les équipes inscrites étaient en seconde ou en première, jamais en terminale (interférence avec la période d'examens). Néanmoins, les classes de première inscrites n'ont jamais rendu leur travail, faute de temps probablement car mars est aussi le moment de la remise de projets pour les travaux personnels encadrés (TPE).

Selon le comité de pilotage, voici quelques difficultés rencontrées lors du déroulement du challenge :

- des sujets (probablement trop) difficiles et ne touchant pas les domaines d'intérêt des jeunes (sont essentiellement concernés les deux sujets sur les résultats scolaires et sur les impôts) ;
- la forme des sujets est inhabituelle pour les élèves et les professeurs (projet libre, donc non « guidé ») et différente de la plupart des autres concours qui se présentent sous la

C. Goga et P. Ardilly

forme question/réponse ; en général, les élèves ont répondu aux questions posées dans les sujets – questions qui prétendaient seulement fournir des pistes d’investigation – mais sans aller plus loin ;

- une faible connaissance des outils statistiques et informatiques de la part des élèves, situation amplifiée par un manque accru, parfois, de dotations informatiques ; un fort usage des calculettes ou d’un tableur mais pas davantage compliquait la mise en œuvre ;
- peu d’implication dans l’encadrement des projets de la part des enseignants ; manque de formation des enseignants à la statistique ;
- la notion d’aléatoire apparaît (très) difficile à comprendre, encore plus le tirage d’un échantillon possédant de bonnes propriétés statistiques.

De façon générale, il est apparu que le principe de l’inférence – même en population finie – n’est pas naturel et n’est pas perçu correctement par les élèves. Ces derniers ont de fait une forte tendance à confondre la population cible et l’échantillon, ce qui rend abstraites les questions explicitement posées lors de la présentation des sujets. C’est pourquoi les sujets n° 3 et n° 5 sont de loin les plus prisés, avec un effet pervers pénalisant : les difficultés techniques n’étant pas ou mal perçues, l’énergie est consacrée à produire des commentaires (souvent sans intérêt) sur les phénomènes que prétend mesurer l’enquête, ce qui n’est pas ce que l’on attend.

Nous rappelons que les participants au challenge, élèves comme enseignants, pouvaient contacter les experts en statistique pour poser toute question technique liée aux sujets. Néanmoins, il y a eu très peu de contacts durant ces trois premières éditions. Les questions posées visaient principalement les tirages d’échantillons et leurs tailles, et en général les élèves n’ont pas tenu compte des suggestions faites.

5 Conclusion

Le challenge en est actuellement à sa 4^e édition, relative à l’année scolaire 2017-2018. Il est donc encore très jeune. Depuis son lancement, il a néanmoins gagné en notoriété et son succès a grandi auprès des établissements. C’est donc encourageant pour l’avenir.

Néanmoins, on voit bien que l’implication des professeurs de mathématiques est essentielle et constitue la clé de la réussite de l’opération. Sur ce point, on touche au problème de formation des professeurs à la statistique inférentielle. Un des enjeux des statisticiens professionnels est de contribuer – modestement – à cette formation, et des opérations comme le challenge devraient aller en ce sens.

Il faut aussi organiser la pérennité du challenge, qui a vocation à prendre encore de l’ampleur et qui ne peut plus reposer seulement sur quelques personnes pionnières, certes très motivées mais qui réclament désormais un cadre plus organisé.