

Sujet 1

Appréciation d'une stratégie d'estimation du niveau moyen d'un lycée en mathématique

Tous les élèves des classes de première viennent de passer en janvier 2014 une évaluation nationale en mathématique. La correction de l'épreuve est assurée au niveau national et les résultats sont communiqués en mars 2014 aux élèves par un courrier individuel et confidentiel.

Le lycée Papillon, qui abrite 10 classes de première (toutes filières confondues) regroupant au total 250 élèves, ne reçoit les notes de l'ensemble de ses élèves qu'au début juillet. En avril, ne pouvant supporter l'attente, le proviseur du lycée veut bien naturellement anticiper la réussite de son établissement afin de se positionner par rapport aux autres lycées.

Il ne peut pas raisonnablement demander leurs résultats à chacun des 250 élèves concernés, ce serait trop fastidieux à collecter. Il décide donc d'une stratégie consistant à sélectionner 50 élèves à partir des listes dont il dispose, puis à obtenir auprès de chacun d'eux les notes de l'évaluation nationale, notes qu'il lui faudra ensuite « combiner ». En tant que chef d'établissement, il a bien entendu accès à la note moyenne de mathématique obtenue au premier trimestre de l'année scolaire 2013-2014 par chaque élève, moyenne calculée à partir des épreuves internes au lycée.

Début juillet, à la lumière des résultats complets qui lui auront été (enfin !) communiqués, il pourra mesurer son erreur. Mais au-delà de cette constatation (hélas trop tardive) il pourra également apprécier avec un peu de recul la pertinence de sa stratégie, c'est-à-dire les risques qu'il a pris en avril, en se disant que cette stratégie aurait (vraisemblablement) pu conduire à sélectionner bien d'autres échantillons que celui sur lequel il s'est appuyé ...

Votre aide est requise à deux niveaux : en avril, pouvez-vous suggérer au proviseur une (ou plusieurs ...) stratégie(s) prometteuse(s) et début juillet, pouvez-vous mesurer jusqu'à quel point votre conseil a été pertinent ?

Quelques indications complémentaires :

- Il convient déjà de préciser ce qu'on calculerait si on disposait de l'intégralité de l'information, comme dans le contexte de juillet (la « vraie valeur ») ;
- Il convient ensuite de distinguer, dans le contexte d'avril, ce qui relève de l'échantillonnage de ce qui relève du traitement des données collectées (appelé « estimation ») ;
- Quand on dispose de l'intégralité de l'information (contexte de juillet), l'appréciation de la qualité d'un sondage peut se faire par une technique de simulation, en produisant un (assez) grand nombre de tirages ;
- Peut-on tirer profit d'une éventuelle relation entre moyennes de mathématique au premier trimestre et notes de mathématique obtenues à l'évaluation nationale et/ou entre la classe dans laquelle se trouve l'élève et son résultat à l'évaluation nationale ?

- L'échantillonnage se fera-il avec ou sans remise des élèves ?
- Tous les élèves du lycée doivent-ils avoir la même chance de faire partie de l'échantillon ?
- N'est-il pas profitable de passer par un tirage d'élèves au sein de sous-populations particulières, à définir ?

Les données sont disponibles dans le fichier Excel nommé *niveau_mathématiques*. Les variables fournies sont les suivantes : l'identifiant de la classe, l'identifiant de l'élève dans la classe, la note moyenne en mathématique au premier trimestre et la note de mathématique obtenue à l'évaluation nationale.