

### Travailler les spécialités – La démarche d'échantillonnage

#### Objectif de la séquence

Sensibiliser les élèves à la démarche d'échantillonnage à travers l'intervalle de confiance dans des situations simples (utile en cas d'enquête liée au projet) :

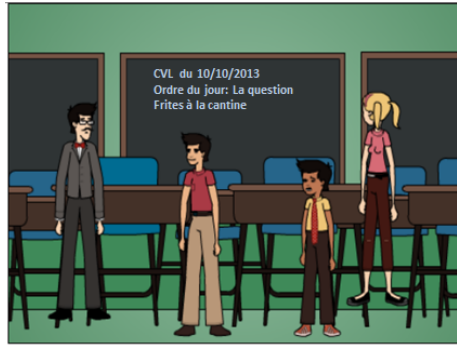
- Introduction à la notion d'inférence par une activité guidée : il s'agit d'une BD mettant en scène des élèves, membres du conseil de la vie lycéenne et qui souhaitent, à partir d'un sondage, faire valider une requête.
- Description de la méthode d'échantillonnage utilisée lors de certains sondages.
- Deux exemples d'application de la démarche.

#### L'inférence statistique

L'inférence statistique consiste à estimer (induire) les caractéristiques inconnues d'une population à partir d'un échantillon aléatoire issu de cette population. Les caractéristiques de l'échantillon, une fois connues, reflètent avec une certaine marge d'erreur possible celles de la population. L'inférence statistique est donc un ensemble de méthodes permettant de tirer des conclusions fiables à partir de données d'échantillons statistiques. « L'inférence statistique comporte toujours une marge d'incertitude au sens absolu du terme. Le rôle des mathématiques en statistique est de fournir des méthodes qui permettent d'évaluer avec précision l'étendue de cette marge d'incertitude et une fois ladite marge connue, les résultats statistiques fournissent une base véritablement scientifique aux décisions et à l'action pratique » - P.C Mahalanobis.

#### Illustration





Le lycée comportait l'an passé 1000 demi-pensionnaires parmi lesquels 56% de filles. Louis avait fait un sondage sur 400 élèves demi-pensionnaires, qu'il avait présenté au proviseur lors du CVL du mois de Mai 2012, en vue de préparer la rentrée. 190 filles avaient accepté de répondre au questionnaire sur la question. Le proviseur n'a même pas pris le temps de lire les résultats. Il a refusé d'examiner la question sous prétexte que «le groupe interrogé n'est pas représentatif de l'ensemble des demi-pensionnaires ».



« Il fallait interroger au moins 204 filles ! Vous reviendrez l'année prochaine avec votre histoire de frites ».

### Explications

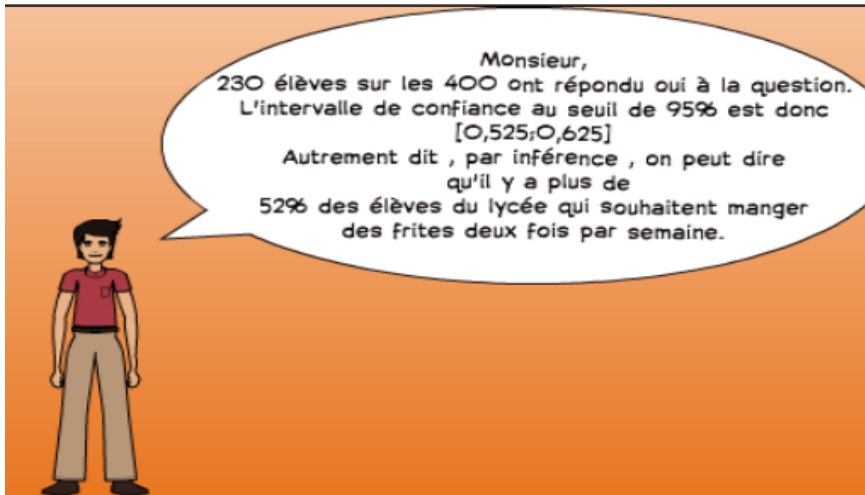
Pour que l'échantillon soit représentatif de la population des demi-pensionnaires du lycée, il faut que la fréquence de filles dans l'échantillon soit dans l'intervalle

$$\left[ 0,56 - \frac{1}{\sqrt{400}}; 0,56 + \frac{1}{\sqrt{400}} \right] = [0,51; 0,61]$$

C'est-à-dire qu'il doit y avoir entre 204 et 244 filles dans l'échantillon de personnes sondées pour que celui-ci puisse être considéré comme représentatif de l'ensemble des lycéens.

Nos amis sont réélus l'année suivante au CVL et comptent prendre leur revanche !

On suppose que cette année encore, le lycée comporte 1000 élèves demi-pensionnaires, avec la même proportion de filles.



### Applications : la méthode d'échantillonnage

1) Guillaume souhaite ouvrir une boutique de vêtements dans sa ville. Afin d'optimiser ses chances de réussite, il décide de faire un sondage auprès de sa clientèle potentielle. Il se pose alors, rapidement, toutes les questions suivantes :

- Combien de personnes doit-il interroger pour que ses résultats soient fiables ?
- Comment doit-il les choisir pour qu'ils soient représentatifs ?
- Comment interpréter les résultats de l'enquête ?

L'enquête, menée sur un échantillon, doit permettre de généraliser les résultats à l'ensemble de la population ciblée. La taille de l'échantillon est bien sûr déterminante : plus l'échantillon est important, plus la généralisation sera fiable. Toutefois, les gains de fiabilité ne sont pas proportionnels à l'augmentation de la taille de l'échantillon.

La notion de fiabilité d'échantillon est matérialisée par un seuil de confiance et une marge d'erreur. Ainsi, un échantillon défini à un seuil de confiance de 95 % et avec une marge d'erreur de 3 % (c'est la marge correspondant à un échantillon de taille 1000 au seuil de 95 %) permettra d'extrapoler chaque résultat issu d'une enquête, avec 5 % de risques de se tromper de + ou - 3 %.

Ainsi, si on obtient un taux de satisfaction de 67 % sur un produit ou service, on pourra affirmer que le taux de satisfaction réel sur l'ensemble des clients potentiel a 95 % de chances de se situer entre 64 et 70 %.

Guillaume fait finalement appel à une entreprise spécialisée dans les sondages !

2) Etude de l'Impact d'une campagne publicitaire sur les ventes

Une grande société de cosmétique vient de commercialiser une nouvelle crème pour le visage. Pour assurer le lancement de cette nouvelle gamme de produits, elle fait une grande campagne publicitaire très coûteuse. Le département mercatique est chargé de mesurer la rentabilité de cette campagne de publicité, en analysant son impact sur les ventes.

Dans un échantillon aléatoire de 1000 acheteurs de ce produit, 600 personnes déclarent avoir effectué l'achat après avoir vu la campagne publicitaire.

On note  $p$ , proportion inconnue d'acheteurs de ce produit dont l'achat a été effectué après avoir vu la campagne publicitaire.

A partir des résultats de ce sondage, estimer la proportion  $p$  inconnue à l'aide d'un intervalle de confiance au niveau de confiance de 95%. Quelle est l'amplitude de cet intervalle ?