NOM DATE PÉRIODE

Matériel de soutien aux familles

Inférences statistiques

Dans cette unité, votre élève utilisera un petit *échantillon* de données pour estimer des informations sur un groupe plus large appelé *population* et utilisera la simulation pour déterminer une plage de valeurs pour l'estimation. Une population est l'ensemble des sujets d'intérêt pour une question et un échantillon est un groupe plus petit au sein de cette population.

Par exemple, nous pouvons vouloir déterminer le montant *moyen* que les familles aux États-Unis dépensent chaque mois en nourriture. La population comprend toutes les familles des États-Unis, mais effectuer une collecte d'informations auprès de tout le monde serait très difficile et coûterait beaucoup d'argent, de sorte que nous pourrions commencer à recueillir des données auprès d’un échantillon de 50 familles.

Une question importante à prendre en compte lorsque l'on commence à recueillir des informations auprès d'un échantillon est comment l'échantillon doit être sélectionné. Les données que vous recueillez peuvent être très différentes si vous interrogez des familles qui font leurs courses dans une épicerie locale par rapport à si vous interrogez des personnes qui se trouvent devant un restaurant chic. De même, le montant dépensé en nourriture à San Francisco est probablement très différent du montant dépensé dans l'Iowa rural. Il peut même y avoir des habitudes de consommation cachées d'une manière à laquelle nous n'avons pas encore pensé. Alors, comment assurer que votre échantillon soit représentatif des familles aux États-Unis sans utiliser trop de familles appartenant à des groupes qui ne sont pas représentatifs dans leurs dépenses ?

La solution est d'utiliser une sélection *aléatoire*. Nous pouvons sélectionner 50 familles à l'aide d'un processus aléatoire, par exemple en demandant à un ordinateur de sélectionner au hasard des familles dans une base de données, sans tenir compte d'autres facteurs. Cela devrait réduire les partis pris qui pourraient être appliqués par des humains essayant d'obtenir des informations sur les gens et cela inclura probablement des proportions plus précises des différents types de familles aux États-Unis. Bien que le caractère aléatoire ne permet pas d'éliminer entièrement les partis pris dans la sélection de l'échantillon, il réduira considérablement les partis pris présents par rapport à une sélection qui ne soit pas aléatoire.

Des chercheurs ont mené des études comme celles-ci et ont trouvé le montant moyen dépensé en nourriture chaque mois. Un rapport indique que le montant moyen dépensé en nourriture chaque mois est de 600 $ avec une *marge d'erreur* de 150 $. La marge d'erreur est utilisée pour dire que nous ne nous attendons pas à ce que chaque famille de l'échantillon dépense exactement 600 $.

La marge d'erreur est un élément importante à rechercher dans les résultats de statistiques. Il serait irresponsable de discuter de statistiques sans prévoir une marge d'erreur pour décrire dans quelle mesure la valeur pourrait varier. De nombreux graphiques inclus dans les rapports la présenteront en petits caractères sur le graphique. Soyez attentifs à quelque chose comme $\pm 3\%$ dans un graphique sur le taux d'approbation d'un fonctionnaire ou dans des sondages lors d'une prochaine élection. Cela signifie que les pourcentages indiqués dans le graphique peuvent en fait être jusqu'à 3 % inférieurs ou supérieurs au nombre indiqué.

Voici une tâche à essayer avec votre élève :

Un vote doit bientôt être organisé dans une ville sur une possible augmentation de 2% de l'impôt sur les sociétés pour augmenter le financement des écoles publiques. Les nouvelles locales montrent une image qui indique que 52 % de la population votante est en faveur de l'augmentation de l'impôt et dans le coin, est indiqué « Marge d'erreur de $\pm 3.5\%$ ». Le journaliste semble confiant que les impôts sur les sociétés seront augmentés parce que tout ce qui dépasse 50 % des voix en faveur de l'augmentation des impôts fera passer la loi.

1. Le journaliste qui a trouvé le chiffre de 52 % est arrivé à cela en se rendant dans 4 des 20 quartiers différents de la ville et en demandant aux résidents leur avis. Y a-t-

il quelque chose de mal dans la façon dont cela a été effectué ? Pouvez-vous penser à une meilleure façon de collecter des données ?

2. Que signifie la marge d'erreur dans cette image ?
3. Faut-il être confiant dans le fait que les impôts seront augmentés ? Expliquez votre raisonnement.

Solutions :

1. Ne se rendre que dans 4 quartiers de la ville pourrait laisser de côté les opinions de nombreux votants d'autres quartiers dans lesquels le journaliste n'est pas allé. Une meilleure façon de recueillir des informations pourrait être de sélectionner au hasard plusieurs ménages de la ville pour sonder leur opinion. Une sélection aléatoire est plus susceptible d'éviter tout parti pris du journaliste sur les quartiers à visiter.
2. La marge d'erreur signifie que le pourcentage réel en faveur d'une augmentation des impôts pourrait être 3,5 % supérieur ou inférieur aux 52 % annoncés sur la base de l'échantillon. Cela signifie que le pourcentage réel se situerait entre 48,5 % et 55,5 %.
3. Exemples de réponses :
   - Je pense qu'il y a encore de bonnes chances que les impôts soient augmentés. Bien que le pourcentage réel puisse être aussi bas que 48,5 % en fonction de la marge d'erreur, il pourrait également atteindre 55,5 %. La plupart des pourcentages possibles sont supérieurs à 50 %, je pense donc que l'augmentation sera appliquée.
   - Je pense qu'il n'est pas encore clair si l'augmentation sera appliquée. Sur la base de la marge d'erreur, le pourcentage réel pourrait être aussi bas que 48,5 %, ce qui bloquerait l'application de l'augmentation. Je ne suis pas non plus sûr des méthodes utilisées par le journaliste pour sélectionner un échantillon pour ce rapport, de sorte que le rapport peut ne pas être très précis.