

Comment utiliser les données du recensement d'E-STAT avec Fathom pour un projet

Les études universitaires en valent-elles la peine?

Ce que le recensement révèle sur la corrélation entre le niveau de scolarité et le revenu

par Joel Yan, Statistique Canada, joel.yan@statcan.ca, 1-800-465-1222

Exercice

Est-il dans votre intérêt de terminer vos études universitaires? Trouvez la corrélation qui existe entre la proportion des personnes titulaires d'un diplôme universitaire et le revenu moyen à l'aide des données de recensement d'une grande région urbaine du Canada, selon les secteurs de recensement. Il est possible de le faire à partir d'E-STAT en créant un graphique de dispersion dans lequel on peut voir la relation qui existe entre le revenu moyen et la pourcentage de la population détenant un diplôme universitaire, selon les secteurs de recensement. On importe ensuite les données dans Fathom afin de calculer l'équation de régression et les statistiques de corrélation.

E-STAT contient plus de mille caractéristiques de recensement pour chaque secteur de recensement (et municipalité) du Canada que l'on peut manipuler à l'aide d'E-STAT et importer dans Fathom pour en étudier les relations. La présente activité donne un exemple de la marche à suivre.

Attentes liées au cours de mathématiques de la gestion de données (12^e année) de l'Ontario :

- Résolution de problèmes impliquant des relations complexes à l'aide de diagrammes (*Gestion des données- Attentes - page 51*)
- Organisation et synthèse de données de sources secondaires (p. ex., Internet) à l'aide de la technologie (*Collecte de données- page 54*)
- Description de la relation entre deux variables en interprétant le coefficient de corrélation . (*Statistiques- Attentes- page 54*)
- Calcul du coefficient de corrélation d'un ensemble de données à l'aide d'un logiciel statistique (*Distributions de données à deux variables- page 54*)

Nota : La partie E-STAT de cette activité se trouve également en français (et en anglais) sur le site Web de Statistique Canada, à l'adresse http://www.statcan.ca/francais/Estat/guide/income_f.htm.

Marche à suivre

Comment trouver le bon tableau à l'aide de l'outil Recherche dans le Recensement

- À la table des matières E-STAT, cliquez sur **Recherche dans le Recensement** dans la barre d'outils à la gauche de l'écran.
- Sous **Recensement**, choisir Recensement de 2001 (la sélection par défaut), puis cliquez sur Afficher!
- Sous **Recensement de 2001**, de la liste déroulante sélectionnez Recensement de la population de 2001 (46 grands centres urbains, [secteurs de recensement](#) (quartiers)) dans la liste déroulante, puis cliquez sur Afficher!
- Sous **Choix d'un profil**, sélectionnez **Fréquentation scolaire, scolarité, domaine d'études, plus haut niveau de scolarité atteint et gains, 2001** dans la liste déroulante, puis cliquez sur Choisir!

Sélection de la région géographique et des caractéristiques

- Sous **Région géographique**, sélectionnez un centre urbain qui ne compte pas un grand nombre de secteurs de recensement. Pour cet exemple, nous avons choisi Grands centres urbains pour l'Ontario-2001- Oshawa (68 lieux) à partir de la liste déroulante.
- Sous **Caractéristiques désirées**, sélectionnez les variables suivantes en faisant défiler la liste :
 - Population totale de 20 ans et plus selon le plus haut niveau de scolarité atteint - Données-échantillon (20 %)
 - Avec baccalauréat ou diplôme supérieur, études universitaires, population de 20 ans et plus selon le plus haut niveau de scolarité atteint
 - Revenu moyen d'emploi \$, ayant travaillé toute l'année à plein temps, population de 15 ans et plus ayant un revenu d'emploi

TRUC : Si vous utilisez un fureteur pour Windows, vous pouvez sélectionner plusieurs articles non contigus en appuyant sur la touche « Ctrl » lorsque vous cliquez sur les caractéristiques sélectionnées. Si vous vous servez d'un fureteur pour Macintosh, vous pouvez procéder de la même façon en appuyant sur la touche de commande au moment de cliquer sur les caractéristiques sélectionnées.

Comment exprimer les données en pourcentage

E-STAT comporte une fonction permettant d'effectuer des calculs de base et de reproduire les données sous forme de graphiques ou de cartes.

Pour convertir les données sur la scolarité en pourcentage de la population totale (la première variable de cet exemple), suivez les instructions indiquées ci-dessous :

- Cliquez sur l'icône Tableau régions en lignes, située dans la partie inférieure de l'écran.
 - Cela permet de produire un tableau dans lequel chaque ligne représente les valeurs des données d'un secteur de recensement à l'intérieur de la région urbaine sélectionnée. La première colonne indique les numéros des secteurs de recensement. Les colonnes sont ordonnées par secteur de recensement, par défaut.
- Au bas du tableau, cliquez sur le bouton radio Données exprimées comme % de la 1ère caractéristique.
- Cliquez ensuite sur le bouton Réafficher.
 - Cela permet de produire un tableau qui montre le niveau de scolarité exprimé en pourcentage de la population totale de 20 ans et plus.
Nota : : Lorsqu'on calcule le pourcentage de la population totale selon le niveau de scolarité, les valeurs relatives aux revenus demeurent inchangées. Cela s'explique par le fait qu'E-STAT fait la distinction entre les variables dont les comptes sont absolus (p. ex. le niveau de scolarité) et les variables relatives (p.ex. le revenu moyen) et calcule les pourcentages en fonction du genre de données utilisées

Tri des données

On peut aussi trier le tableau par ordre ascendant, selon n'importe quelle caractéristique du tableau. Par exemple, pour trier selon le revenu d'emploi moyen, suivez les instructions indiquées ci-dessous.

- Cliquez sur le bouton radio Tri selon les données de la 2e caractéristique
- Cliquez sur le bouton Réafficher pour visualiser les résultats.

Analyse 1

Quelles sont les valeurs minimales et maximales du revenu d'emploi moyen des secteurs de recensement dans le centre urbain choisi?

Dans quelle mesure l'éducation influe-t-elle sur le revenu d'emploi moyen?

Conseil : Dans E-STAT, vous pouvez accéder aux articles portant sur ce sujet à partir de la Tables des matières, en choisissant Articles. Sous La société, sélectionnez Éducation.

Conseil : Des détails additionnels concernant l'éducation et le revenu sont disponibles dans le Quotidien du 11 mars 2003. Ce document « [Recensement de la population: gains, niveaux de scolarité, domaine d'études et fréquentation scolaire](http://www.statcan.ca/Daily/Francais/030311/q030311a.htm) » est disponible à <http://www.statcan.ca/Daily/Francais/030311/q030311a.htm>.

Production d'une carte à partir des données par secteurs de recensement

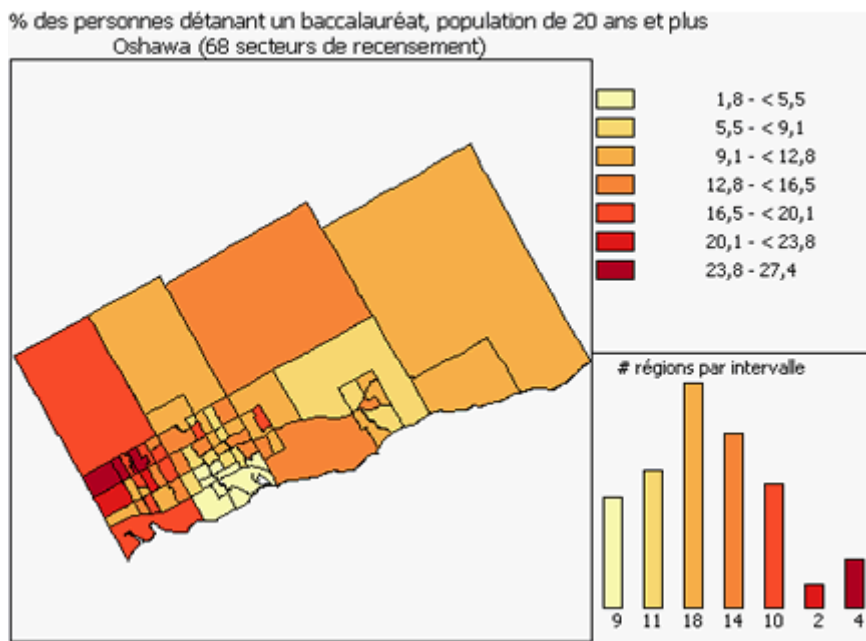
Les cartes nous permettent de visualiser et de mieux comprendre les régions pour lesquelles on examine les données. Pour tracer une carte statistique,

- cliquez sur l'icône de la carte située au-dessous du graphique.
Par défaut, cela permet de produire une carte par ombre portée en couleur de la première caractéristique (dans cet exercice, il s'agit d'une carte du pourcentage des personnes détenant un baccalauréat) comportant 7 intervalles de classe, comme celle que l'on voit ci-dessous. Ce type de carte peut être utile pour identifier les secteurs de recensement se rapportant aux caractéristiques choisies. On peut aussi utiliser l'histogramme des fréquences situé dans la partie inférieure droite de la carte pour évaluer la distribution des données dans les régions sélectionnées (p.ex. distribution normale, désaxée ou bimodale).

Spécification des titres de la carte

- Cliquez sur le bouton Spécifier les titres, situé au-dessus de la carte.
- Inscrivez un titre dans la boîte prévue à cet effet. Par exemple, " % des personnes détenant un baccalauréat, population de 20 ans et plus ".
- Dans la boîte Sous-titre, inscrivez " Oshawa (68 secteurs de recensement) ".
- Cliquez sur la rubrique Retracer en utilisant ces titres.

Voici la nouvelle carte affichant les titres modifiés.



Analyse 2

L'histogramme des fréquences situé à la droite de la carte indique la distribution répartie sur sept classes de la caractéristique choisie (% des personnes détenant un baccalauréat, population de 20 ans et plus, Oshawa). Le nombre sous chaque barre verticale représente le nombre des secteurs de recensement dans cette classe. Dans cet exemple, la distribution des valeurs est normale. La troisième classe (9,1 - <12,8) est le mode ou la plus haute fréquence.

Création d'un graphique de dispersion

E-STAT permet aussi de produire des graphiques de dispersion dans lesquelles on peut voir la relation qui existe entre deux variables. Dans cet exercice, nous cherchons à savoir s'il existe une relation entre le niveau de scolarité et le revenu moyen d'emploi selon les secteurs de recensement. Pour créer un graphique de dispersion comportant la ligne de régression linéaire :

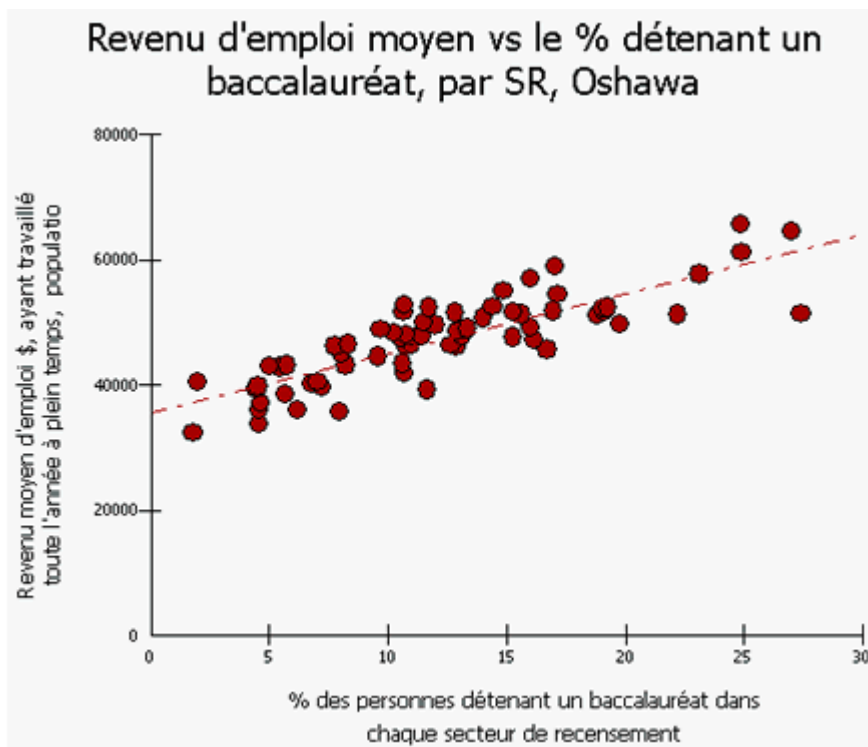
- Cliquez sur l'icône Graphique dispersion, située dans la partie inférieure de l'écran.

Spécification des titres

Vous pouvez modifier les titres des graphiques, si nécessaire.

- Cliquez sur le bouton Spécifier les titres, situé en dessous du graphique
- Inscrivez un titre dans la boîte prévue à cet effet. Par exemple, inscrivez Revenu d'emploi moyen vs le % détenant un baccalauréat, par SR, Oshawa.
- Dans la boîte Sous-titre, inscrivez % des personnes détenant un baccalauréat dans chaque secteur de recensement.
- Cliquez sur la rubrique Retracer en utilisant ces titres.

Voici le nouveau graphique à dispersion affichant les titres modifiés.



Analyse 3

La ligne de régression linéaire constitue-t-elle le meilleur modèle pour prévoir le comportement de la variable du revenu moyen en fonction du % des diplômés universitaires dans une région donnée? _____

Que représente la pente de la ligne de régression linéaire obtenue?

Quelle est la pente approximative de la ligne de régression linéaire pour le centre urbain que vous avez choisi? _____

Conseil : Des détails additionnels concernant les graphiques de dispersion ou Nuages de points sont disponibles dans [Les statistiques : le pouvoir des données!](http://www.statcan.ca/francais/edu/power/toc/contents_f.htm), disponible à http://www.statcan.ca/francais/edu/power/toc/contents_f.htm

1. E-STAT nous permet de tracer facilement des graphiques de dispersion comportant une ligne de régression linéaire pour que nous puissions évaluer les relations possibles entre n'importe quelles deux variables. Cependant, à l'heure actuelle, E-STAT ne donne pas l'équation de la ligne de régression linéaire ni la valeur de r^2 . Pour obtenir ces statistiques, il est possible d'exporter les données d'E-STAT dans un autre logiciel, tel qu'un tableur, une calculatrice graphique ou un logiciel d'analyse comme Fathom.
2. Descendez jusqu'à **Choisissez un format** et sélectionnez « Fichier.WK1 Géo=lignes ».
3. Dans la fenêtre de **téléchargement** sélectionnez « Ouvrir » et cliquez sur OK.

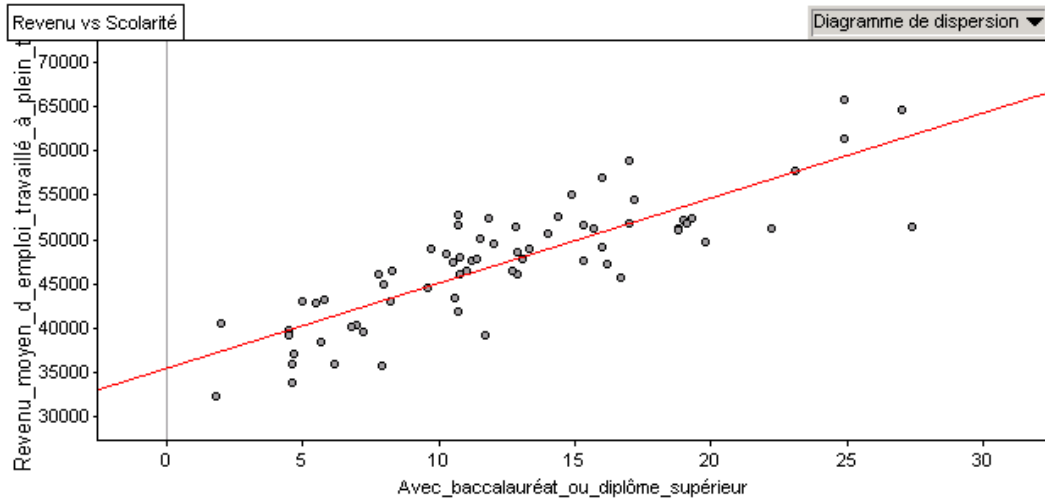
4. Lorsque les données s'ouvrent dans votre tableur, sélectionnez et copiez les colonnes qui vous intéressent. Ensuite, sélectionnez toutes les lignes de données en appuyant sur « Ctrl-A », puis sur « Ctrl-C ».

Chargement des données dans Fathom

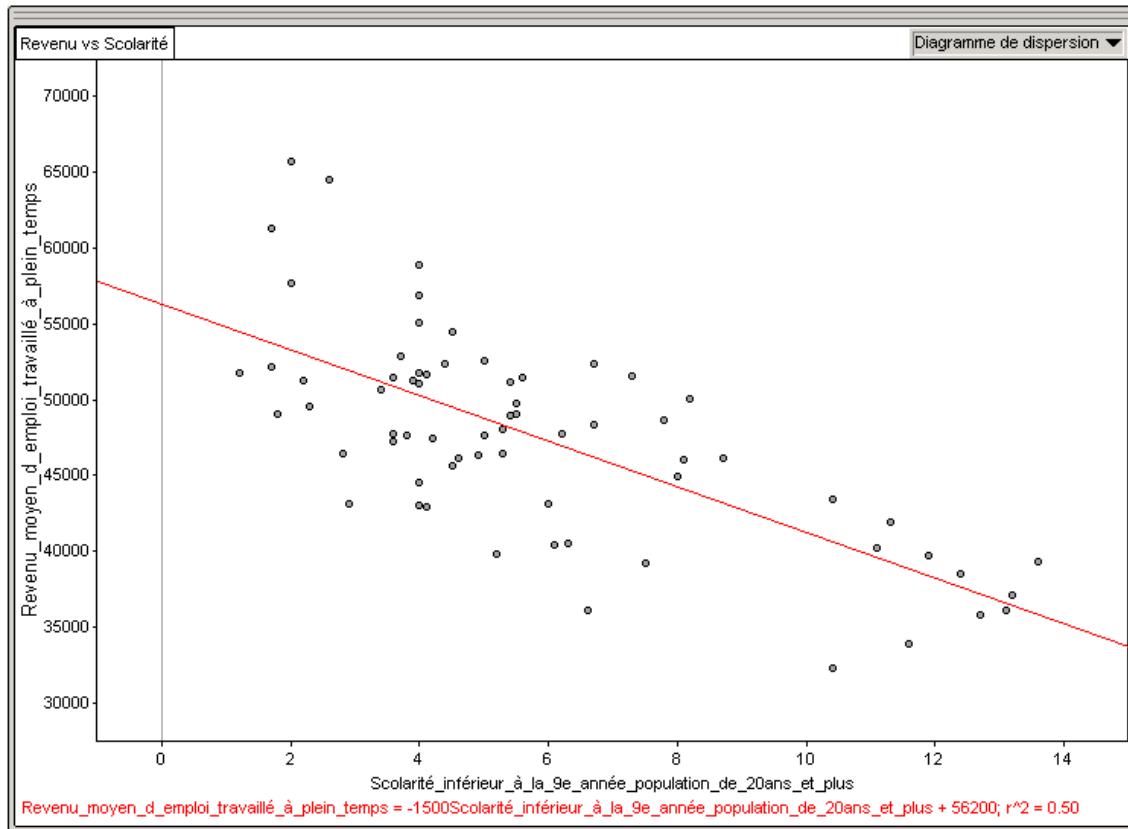
5. Basculez vers Fathom. (Si Fathom n'est pas déjà ouvert, vous devrez le lancer.)
6. Dans un nouveau document, créez une nouvelle collection vide.
7. Après avoir sélectionné la collection, choisissez Coller les cas dans le menu Édition.
8. Créez un tableau de cas pour la collection (par exemple en choisissant Tableau de cas dans le menu Insertion).
9. Vous souhaitez peut-être renommer le deuxième attribut pour secteur_de_recensement. Étant donné que les noms d'attribut sont assez longs, il serait utile de les raccourcir. C'est pourquoi nous renommons les attributs Avec_baccalauréat_ou_diplôme_supérieur et Revenu_moyen_d_emploi_travaillé_à_temps_plein.
10. Renommez la collection « Revenu vs Scolarité ».
11. Enregistrez votre document Fathom en sélectionnant Enregistrer dans le menu Fichier.

Création d'un graphique à partir des données

12. Créez un graphique illustrant la variation du revenu moyen en fonction du niveau d'éducation.
13. Dans le menu déroulant **Graphique**, sélectionnez Droite des moindres carrés pour en tracer une dans le graphique (voir ci-dessous). Vous obtiendrez ainsi l'équation de la ligne de régression linéaire, de même que le coefficient de variation (r^2).



14. Décrivez la répartition des points sur le graphique. Donnez une explication possible pour la forme de cette distribution.
15. Vous pouvez également créer la trace des résidus pour vous aider à évaluer la qualité de la régression linéaire de la ligne des moindres carrés.
16. Répétez l'ensemble du processus et de l'analyse à l'aide du pourcentage de la population ayant un niveau de scolarité inférieur à la 9^e année. Il vous faudra retourner à E-STAT pour extraire de nouvelles données. Le graphique obtenu figure ci-dessous.



Nota : Cette analyse pourrait être répétée pour n'importe quel niveau géographique. Par exemple, à l'aide d'E-STAT, vous pourriez extraire les mêmes variables pour les 49 comtés, districts et municipalités régionales de l'Ontario et répéter l'analyse. Pour approfondir l'analyse de la relation, consultez les résultats du recensement 2001 sur la relation entre l'éducation et le revenu. Vous trouverez le tout à l'adresse suivante : <http://www12.statcan.ca/english/census01/products/highlight/Earnings/Index.cfm?Lang=F>. Sélectionnez Recensement dans la barre bleue au haut de l'écran, Données dans la barre gauche, puis Tableaux thématiques.

Fichier : mdm4u/ La corrélation entre éducation et revenu. Jan17.doc

Mise à jour : janvier 19, 2004

Corresponding English File: mdm4u/ Fathom education & Income.doc