# Combining Multiple Health and Demographic Variables

## Overview

In this activity, students use the *Health Indicators* database on the Statistics Canada website (www.statcan.ca) and on E-STAT to look for possible relationships between demographic factors and health outcomes, or to look for patterns in health data by sub-provincial health regions. Health-related data are provided for Canada, the provinces and territories, and the 130 health regions across Canada. By correlating variables across these areas, students can assess the likelihood of a relationship existing among the variables.

Contributors: Joel Yan and Jennifer Hall, Dissemination Division, and Brenda Wannell, Health Statistics Division, Statistics Canada

---

## Objectives

- Demonstrate the wealth of data available for student analysis on the Health Indicators database on E-STAT and how the data can be analyzed using software available to students
- Learn how to extract health-related datasets from E-STAT and import it into spreadsheet software (e.g. Excel, Quattro Pro) or statistical software (e.g. Fathom)
- Solve problems involving complex relationships with the aid of diagrams
- Locate data to answer questions of significance of personal interest by searching well-organized databases
- Describe the relationship between two variables by use the use of scatter graphs and interpreting the correlation coefficient

---

## Suggested Grade Level and Subject Areas

Senior, Mathematics, Health

---

## Duration

- 2-3 75 minute class periods

---

## Materials

- E-STAT Student Instructions Sheet
- Fathom Student Instructions Sheet
- Student Worksheet

- Computers with internet access, a word processing program, a spreadsheet program, and statistical software
- Computer projector

---

**Classroom Instructions**

1) Book a computer lab.
2) Give the students a brief overview of Statistics Canada's website via overhead computer projector.
3) Distribute the E-STAT Student Instructions Sheet and the Fathom Student Instructions Sheet. Have students work through the instructions (E-STAT first, then Fathom) and complete the associated questions on the Student Worksheet.

**Note:** To save time, you can use the provided datasets in Excel and Fathom formats. Then, students can start directly on the Fathom Student Instructions Sheet.

4) Fathom Analysis #3 can be altered to the teacher's discretion (e.g. The students could write a report or do a presentation on their findings).

---

**Evaluation**

Handing in and subsequent marking of the Student Worksheet is at the teacher's discretion. Students can be informally evaluated on their computer skills and work habits.

---

**Statistics Canada and Fathom**

This lesson plan was prepared by Statistics Canada to facilitate the use of Statistics Canada data by teachers of the Ontario mathematics curriculum. This lesson requires the use of the *Fathom* software. *Fathom* is licensed by the Ontario Ministry of Education and used by schools across Ontario. Use of *Fathom* in this lesson is in no way an endorsement or recommendation of the *Fathom* software by Statistics Canada.

Please send comments or examples of how you used this exercise in your class to Learning resources.

# Using E-STAT and *Health Indicators* to Research Health Issues
# Student Instructions

## Accessing *Health Indicators* Data

1) On the Statistics Canada home page (www.statcan.ca), select your language of choice, and then click on **Publications** on the left side-bar.

2) Under **Find Statistics Canada publications**, select **Free internet publications**.

3) Click on **Health** and then choose **Health indicators.**

4) Click on the **View** link beside **most recent release**.

5) Now you have several choices for selecting data. Data tables are organized here according to the following headings and split by source to identify whether they come from Statistics Canada (STC) or the Canadian Institute for Health Information (CIHI), as follows:
   Health status
   Non-medical determinants of health
   Health system performance
   Community and health system characteristics

   In this case we are looking for smokers by health region and want to see if there is any correlation with average educational attainment.

6) Click on **Non-medical determinants of health** to go to the large menu of tables available. Click on **Smoking** to see the list of tables for this topic.
   ** Go to your worksheet and answer E-STAT Question #1. **

## Accessing E-STAT

1) Return to the Statistics Canada home page (www.statcan.ca), select your language of choice, and then scroll down and click on **Learning Resources** on the left side-bar.

2) Click on the **E-STAT** button on the right side of the screen.

3) Click on **Accept and enter** to gain entry to E-STAT.

4) Click on **Search CANSIM** on the left side-bar.

5) Type **smoking** (the first health determinant that we will analyze) in the Search box and click on the **Search** button.
   ** Go to your worksheet and answer E-STAT Question #2. **

6) Scroll down and select **Table 105-0227 - Smoking status, by age group and sex, household population aged 12 and over, Canada, provinces, territories, health regions and peer groups**.

7) Under **Geography,** select all the health regions in one province or region of interest. To do this more easily, click on **View checklist and footnotes** and click in the desired boxes. Return to the main screen by clicking on **Return to picklist**. For example, in the sample graphs shown later, we selected the 37 Health Units in Ontario.
** Go to your worksheet and answer E-STAT Question #3. **

8) Under **Age group,** select **Total, 12 years and over**.

9) Under **Sex,** select **Both sexes**.

10) Under **Smoking status**, click the **Select all** button.

11) Under **Characteristics**, select **Percent**.

12) Click on the **Retrieve as a Table** button.

13) Under **output format,** under **HTML Table**, select **Geography as rows**. Click the **Retrieve now** button.
** Go to your worksheet and answer E-STAT Questions #4-5. **

14) Using your mouse, highlight all the contents of the HTML table, including the column labels, but not the footnotes. Copy the table by right-clicking and selecting **Copy**.

15) Next open your spreadsheet program (i.e. Excel). Position your mouse at the first cell, right-click, and select **Paste**.

16) Clean up the attribute names by deleting the footnote references, shortening the health region names, and adjusting the row and column sizes (using the Auto-fit selection feature), if you wish.

17) Save the spreadsheet with the file name **Smokers**. We will return to this file later to append possible explanatory or related factors.

## Extracting the smoking data and demographic data into a spreadsheet

1) Return to your E-STAT window. Now, we will select demographic attributes for the same areas so that we can research potential relationships with health outcomes. Click **Search CANSIM** in the left side-bar.

2) Type **109-0200** in the box (This is the table number for Census indicator profile, Canada, provinces, territories and health regions, a table with several demographic variables, such as education levels, for the same health regions) and click **Search**.

3) Under **Geography,** select exactly the same regions that you selected above, using the same technique as before. The exact same regions are required since we want to be able to compare and correlate some census variables with the smoking data already extracted.

** Go to your worksheet and answer E-STAT Questions #6-7. **

4) Under **Census profile,** select the following characteristics (using the checklist),:
   - **High school graduates aged 25 to 29, proportion of population aged 25 to 29 (Percent)**
   - **Post-secondary graduates aged 25 to 54, proportion of population aged 25 to 54 (Percent)**
   - **Long-term unemployment rate, labour force aged 15 and over**
   - **Aboriginal population, proportion of total population (Percent)**
   - **Rural population, proportion of total population (Percent)**

5) Click on the **Retrieve as a Table** button.

6) Under **output format,** under **HTML Table**, select **Geography as rows**. Click the **Retrieve now** button.

7) Using your mouse highlight all the contents of the HTML table, including the labels. You do not need to include the footnotes and source notes in the domain to be copied. Copy the table by right-clicking and selecting **Copy**.

8) Open the **Smokers** spreadsheet file that you previously created for the smoking data. In the column after the last data from the previous table, paste this data. Check that the geographic areas from the two files match and then delete the extra column with the region names by highlighting it, right-clicking, and selecting **Delete**.

## Extracting other Health Variables of Interest

1) Return to your E-STAT window. Now we will select other health characteristics for the same areas. Click on **Search CANSIM** in the left side-bar.

2) Enter 105-0200 in the box and then click **Search** to indicate you want to open this CANSIM table. This opens up the selected table containing the Canadian Community Health Survey (CCHS) indicator profile, by sex, Canada, provinces, territories and health regions.

3) Under **Geography** scroll down and select exactly the same regions that you selected before. This is required since we want to be able to compare and correlate some census variables with the smoking data already extracted.

4) Under **Sex**, select **Both sexes**.
** Go to your worksheet and answer E-STAT Question #8. **

5) Under **Health profile**, select characteristics that you want to research.

6) Under **Characteristics**, select **Percent**.

7) Click on the **Retrieve as a Table** button.

8) Under **output format,** Under **HTML Table**, select **Geography as rows**. Click the **Retrieve now** button.

9) Using your mouse, again highlight all the contents of the HTML table, including the labels, and select **Copy**.

## Merging the Health Data into the existing Spreadsheet File

10) Open the **Smokers** spreadsheet file that you previously created for the smoking data. In the column after the last data from the previous table, paste this data. Check that the geographic areas from the two files match and then delete the extra column with the region names by highlighting it, right-clicking, and selecting **Delete**.

11) Highlight the entire spreadsheet by clicking on the box in the top left corner. Right-click anywhere on the data and select **Copy**.

## Importing the Combined Data Variables into Fathom

1) Open **Fathom**.
2) Drag a new **collection** to the workspace.
3) With the collection selected, right-click and select **Paste Cases**. This will import the health data from the spreadsheet. Rename **Smoking Status** to **Geography** by highlighting the text.

4) Double click on the collection box. Change the name of the collection to a meaningful name, such as Health Data for Ontario Health Units.

5) Make a case table for the collection (for example by choosing **Case table** from the **Insert** menu)

6) If the first case does not have numeric values for the two numeric attributes, delete this case by right-clicking where the case number is listed (i.e. 1/38) and selecting **Delete Inspected Case**.

7) Save the Fathom document as **Smokers** by choosing **Save** from the **File** menu.

***Fathom Analysis #1 – Does smoking correlate to education levels?***

1) Drag a graph into the workspace. Drag and drop the Daily_smoker attribute on the x-axis. Using the pull-down menu above the graph, change the graph type to a histogram.
** Go to your worksheet and answer Fathom Questions #1-2.**

2) Drag another graph into the workspace. Drag the Postsecondary_graduates_aged_25_to_54 attribute to the x-axis and the Daily_smoker attribute to the y-axis.

3) Right-click on the graph and select Least-Squares Line. This will overlay in red the least-squares line, as well as provide its equation and R squared value.
** Go to your worksheet and answer Fathom Questions #3-4.**

4) Remove the furthest outlier by selecting **Delete Inspected Case** as before. Replace the outlier after you answer the questions by selecting **Undo** from the **Edit** menu.
** Go to your worksheet and answer Fathom Questions #5-7.**

---

***Fathom Analysis #2 – Find another relationship using this Fathom collection***

Repeat the steps shown in Analysis #1 to explore graphically some possible relationships using other variables contained in the Fathom file.

Some relationships you may consider exploring:

- chronic conditions such as diabetes, asthma, and high blood pressure versus any of the non- medical determinants of health
- level of depression versus long-term employment or education level

---

***Fathom Analysis #3 - Find another relationship using other data from Health Indicators***

Select other variables related to health status and non-medical determinants of health. The *Health Indicators* product contains many such variables. E-STAT actually contains a number of other interesting CCHS profiles (e.g. urban, rural, aboriginal, immigrant) for provinces and territories. Data can be combined from many of these profiles with data from the CCHS Health profiles as indicated above as long as the geographic areas match. Again, the variables can be combined and imported into Fathom, a spreadsheet, or other analytical software for further analysis.

**Some relationships you may consider exploring:**

- life expectancy against smoking, drinking, or life stress
- lung cancer mortality against smoking and exposure to second hand smoke
- percent aboriginal population against infant mortality
- circulatory disease death against level of physical activity or BMI (obesity)
- cancer deaths against dietary practices
- smoking initiation, smoking status, or injuries against percentage rural or aboriginal population
- deaths due to specific conditions against disease prevention measures such as flu shots, mammograms, and pap smears
- obesity against chronic conditions
- chronic conditions against any of the non-medical determinants of health (e.g. level of depression, physical activity, income)

**Caution:** For whichever of the more than 80 health indicators you extract for your analysis, remember to select exactly the same health regions (under Geography) and similar time periods. Keep in mind that some tables are not available at the health region level and data from the Canadian Institute for Health Information are limited to regions with populations greater than 75 000. Once the data have been merged in a single spreadsheet, check that all the geographic areas match.

# Using E-STAT and *Health Indicators* to Research Health Issues
# Student Worksheet

## E-STAT Questions

1) How many *Health Indicators* tables are listed relating to smoking? _____

2) How many tables are listed related to smoking? _____

3) As indicated in brackets after the word **Geography**, for how many different geographic areas do we have the data on smoking rates? _____

4) What does each column contain? _____

5) How many rows of data are there? _____

6) How many characteristics are included in the **Census Profile**? ____

7) Scan the list and name one characteristic on this list that you feel might be correlated to smoking rates by health region. _____

8) How many characteristics are included within the Health Profile? ____

## Fathom Questions

1) Paste your Fathom graph here by selecting **Copy as Picture** from the **Edit** menu:

2) Write a description of the pattern you see on this graph.
   _____
   _____
   _____

3) Paste your Fathom graph here by selecting **Copy as Picture** from the **Edit** menu:

4) Write a description of the pattern you see on this graph.
   _____
   _____
   _____

5) What point (area) is the furthest outlier? _____

6) Why might this area be an outlier?

_____

_____

7) Remove the outlier by **Deleting the Inspected Case** (as before). What happens to the $r^2$ value and the slope?

$r^2$: _____

Slope: _____

# Using E-STAT and *Health Indicators* to Research Health Issues
# Student Worksheet: TEACHER VERSION

**Note: All answers provided are from July 2006. The E-STAT answers may change with time.**

## E-STAT Questions

1) How many *Health Indicators* tables are listed relating to smoking?

***18***

2) How many tables are listed related to smoking?

***38***

3) As indicated in brackets after the word **Geography**, for how many different geographic areas do we have the data on smoking rates?

***167***

4) What does each column contain?

***The percentage of people for each smoking status***

5) How many rows of data are there?

***37***

6) How many characteristics are included in the **Census Profile**?

***81***

7) Scan the list and name one characteristic on this list that you feel might be correlated to smoking rates by health region.

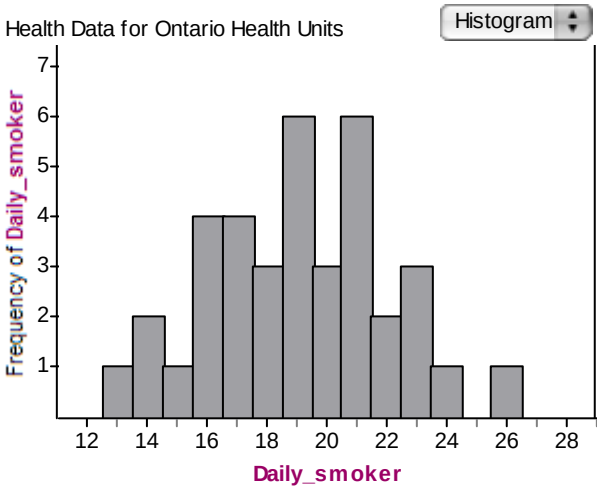
***Answers vary.***

8) How many characteristics are included within the Health Profile?

***24***

## Fathom Questions

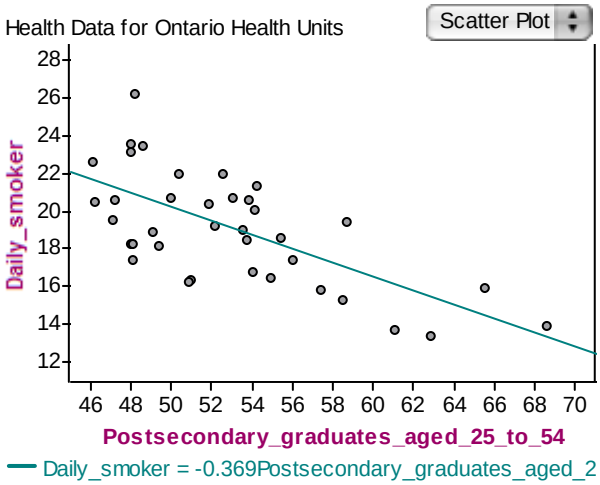
1) Paste your Fathom graph here by selecting **Copy as Picture** from the **Edit** menu:




2) Write a description of the pattern you see on this graph.

***Answers vary.***

3) Paste your Fathom graph here by selecting **Copy as Picture** from the **Edit** menu:




4) Write a description of the pattern you see on this graph.

***Answers vary.***

5) What point (area) is the furthest outlier?

***Porcupine***

6) Why might this area be an outlier?

***Answers vary.***

7) Remove the outlier by **Deleting the Inspected Case** (as before). What happens to the $r^2$ value and the slope?

**$r^2$ changes from 0.45 to 0.46**
**Slope changes from -0.369 to -0.343.**