

How to use Time Series Data from E-STAT with Fathom "Beer Consumption Demographics"


by Joel Yan, Statistics Canada, (joel.yan@statcan.ca), John Kezys, Mohawk College, based on a grade 10 project by Mario Rizzo, St. Pius H.S., Ottawa
(Student activity updated by Shirley Dalrymple, York Region DSB, shirleydalrymple@hotmail.com)
[Latest version revised July 15, 2005 by Mylène Abi-Zeid, OCDSB, mylene.abi-zeid@sympatico.ca]

The greatest consumers of domestic beer are sports-minded young men and women.

During the eighties the 20 to 24 age cohort experienced the greatest growth.

Foot D., 1996, *Boom Bust & Echo*, Macfarlane Walter & Ross, ISBN 0-921912-97-8

Your Task

Use E-STAT to investigate Dr. Foot's hypothesis by creating time series plots for the 20 to 24 age male cohort and beer consumption in Canada from 1971 to 2004. **Questions marked with  must be answered.**

Procedure

To Access Beer Consumption data


1. Startup E-STAT at <http://estat.statcan.ca>. Select your language of choice. Click 'Accept and enter' to enter E-STAT.
2. Click **Search CANSIM** on the left side bar.
3. Under Search by: select **Text search** and click **Continue**
4. In the 4-Search for: box enter **beer consumption** then click **Find tables**
5. Under Active tables, select **002-0011 – Apparent per capita food consumption in Canada, annual**
6. On the subset selection page, choose as follows:
 - Under Commodity select **Ale, beer, stout and porter, retail weight (Litres per year)**
 - Under From select **1971**
 - Under To select **2004**
7. Click the **Retrieve as individual Time Series** button.
8. Click the **Add more series** button at the bottom of the page.

To Access 20 to 24 age cohort data


9. Under Search by: select **Subject** and click **Continue**
10. Click on the yellow folder for **Population and demography**
11. Click on **Population characteristics**
12. Under Active tables, select **051-0001 Estimates of population, by age group and sex, Canada, provinces and territories, annual**
13. On the subset selection page, choose as follows:
 - Under Geography select **Canada**
 - Under Sex select **Both sexes**
 - Under Age group select the first series '**All ages**' and also scroll down and select '**20 to 24 years**' (**while holding down the Ctrl key on a Windows platform or the Command key on a Macintosh computer)
 - Under From select **1971**
 - Under To select the most recent year
14. Click the **Retrieve as individual Time Series** button.
15. In the Output specification screen under output format selection, scroll down and select **Plain text table, time as rows**
16. Click on **Go**. [Note: Make sure that you have 3 series selected and that there are 4 columns of data that appear on your screen before you go on to the next step.]

Fathom Analysis

Copying and Pasting the Data



17. Click and drag in the table so that you have selected all the data, but not the legend at the top of the page, or the source line at the bottom of the page
18. Choose **Copy** from the **Edit** menu.
19. Switch to *Fathom*. (If *Fathom* isn't already running, you will need to launch it)
20. Create a new collection by dragging down the open box icon .
21. With the collection selected, chose **Paste Cases** from the **Edit** pull-down menu.
22. Create a case table by choosing **Case table** from the **Insert** menu.
23. Rename the collection 'Beer Consumption'.
24. Rename the attributes (i.e., Year, Total_Population, Pop20_24, Beer_Consumption).
25. Save your *Fathom* document by choosing **Save** from the **File** pull-down menu.

Graphing the Data

26. Create a "Beer_Consumption vs. Year" scatter plot.
 Describe the pattern in this graph. Provide an explanation for the shape of this distribution.

Manipulating the Data - Converting to Percentages

To study Dr. Foot's claim, you need to check the variation in the 20-24 year old cohort to see if they are the largest beer drinkers on average. Since beer consumption is measured in average number of litres consumed per year over the entire population, we need to convert the 20-24 year old cohort, to a percentage of the total population.

27. Create a "Pop20_24 vs. Year" scatter plot.
 Compare this graph to the "Beer_Consumption vs. Year" graph. Describe any connections.
28. Define a new attribute to determine the percentage of the 20-24 year old cohort to the total population  open the Case table and click on the attribute labeled <new> at the top right of the case table. Give this new attribute a name (Pop20_24asPercent)
29. To assign values for this new attribute, click on Pop20_24asPercent. From the **Edit** menu, select **Edit formula**. In the formula box, expand the Attribute list (click on the ⊕ beside Attributes). Type in the following formula: **100×(Pop20_24 ÷ Total_Population)** and click **Apply**.

Looking for Possible Relationships using Graphing techniques

30. Make a graph showing Beer_Consumption vs. Pop20_24asPercent.
31. From the **Graph** pull-down menu, select **Least-Squares Line** and **Make Residual Plot**.

Analysis

- Describe the correlation of the data using the least-squares line and the residual plot (i.e. comment on strength, direction, trend).
- Explain the pattern of beer consumption in relation to the population percentage of 20-24 year olds.
- Does this pattern tend to support or refute Dr. Foot's statement? Explain.
- If you were a marketing specialist, how would you use this information to market beer?

Enrichment Report on one of the following:

- ✓ Examine the relationship of average beer consumption to another age cohort, such as 15-19 or 25-29.
- ✓ Investigate consumption of another beverage (e.g. wine, milk) using the same E-STAT data table and identify which, if any, age cohorts are most highly related to high consumption of the selected product.

	Level 1	Level 2	Level 3	Level 4
TIPS	- Follows the investigation instructions with considerable assistance -Draws conclusions and inferences with little reference to the data	- Follows the investigation instructions with some assistance -Draws conclusions and inferences with some reference to the data	- Follows the investigation instructions with very little assistance -Draws conclusions and inferences with reference to the data	- Follows the investigation instructions with no assistance -Draws insightful conclusions and inferences with reference to the data

Notes, Graphs and Follow-up activities for use by the Teacher

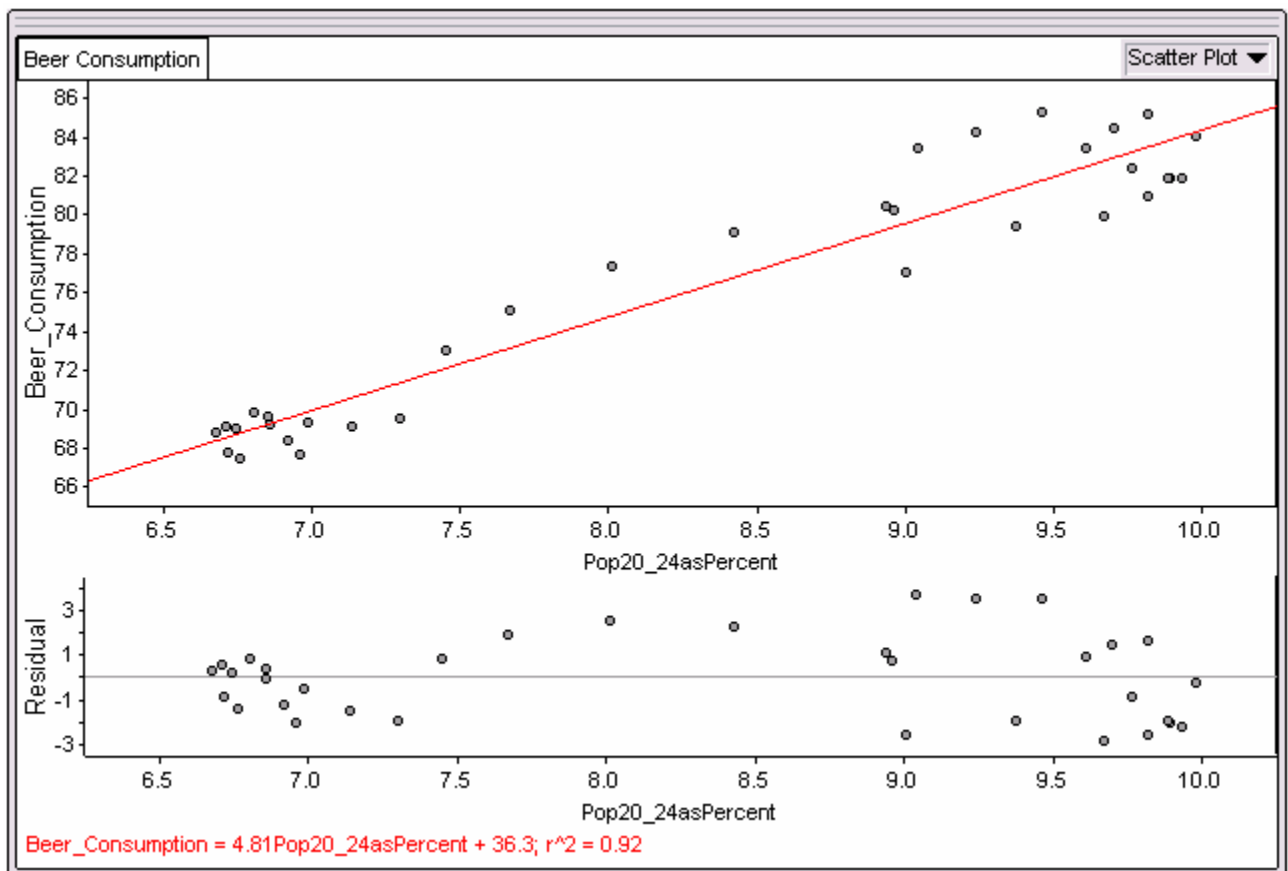
Purpose of this activity:

This activity is designed to introduce a math class to the use of E-STAT's CANSIM database with Fathom! This activity provides an example of how different time series from more than 2,400 large multi-dimensional tables on CANSIM can be combined to analyze possible relationships between the data sets. Any of the more than 26 million CANSIM time series from about 200 different surveys can be brought together to explore possible relationships, as long as the time frames for the data overlap.

Related Expectations for the Ontario Grade 12 Mathematics of Data Management Course:

- Solve problems involving complex relationships with the aid of diagrams. (*ODV.02, Organization of Data for Analysis – Overall Expectations – page 49 in the Ontario curriculum*)
- Locate data to answer questions of significance of personal interest by searching well-organized databases. (*OD1.01, Organization of Data for Analysis – Organizing Data – page 49*)
- Describe relationship between two variables by use of scatter graphs and interpreting the correlation coefficient (*ST4.01 and ST4.02, Statistics – Overall Expectations – page 52*)

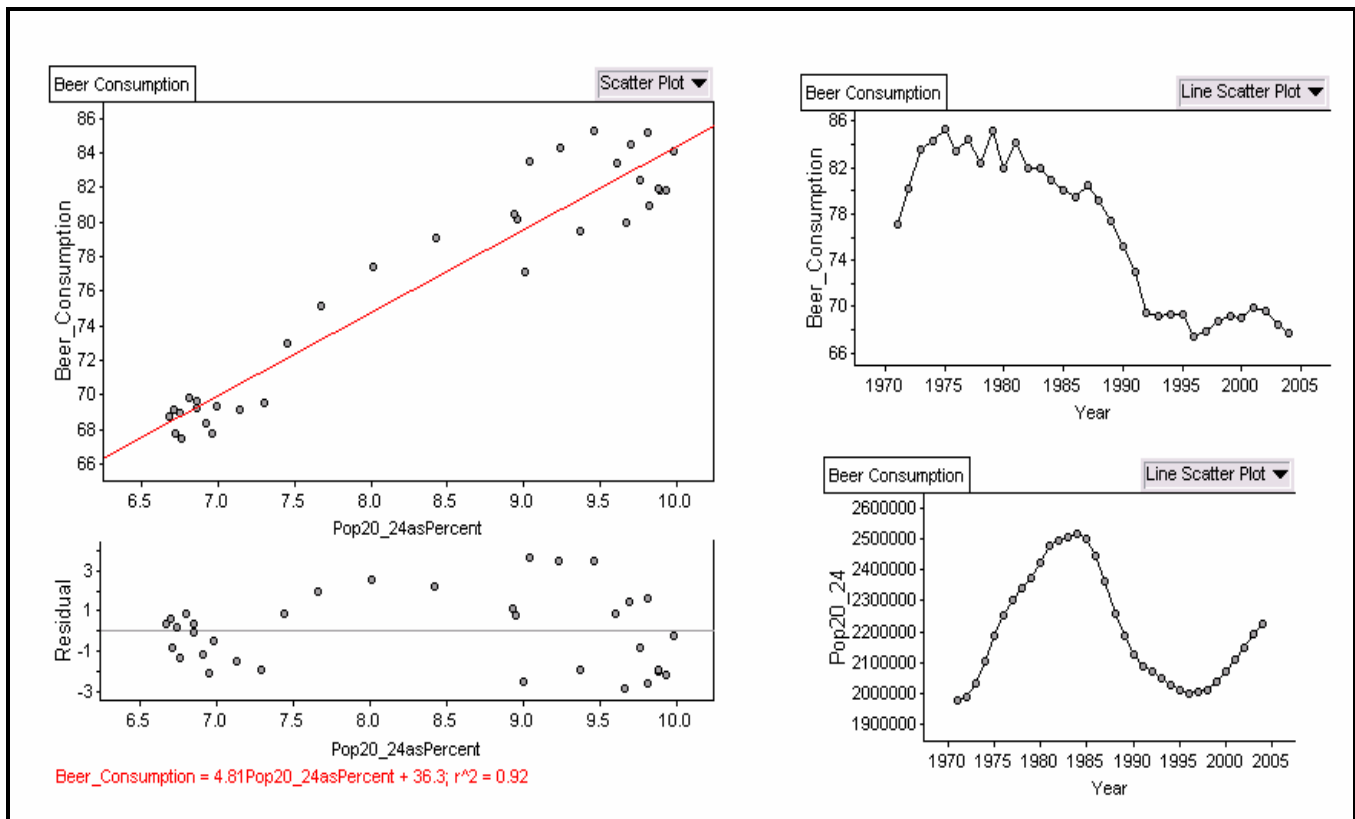
Figure 1: Scatter graph and residual plot showing the relationship between Average beer consumption of Canadians in litres per year (on the y-axis) by the percentage of the total Canadian population who are 20 to 24 years old (on the x-axis)



Note: There is a very high correlation (r^2 of 0.92) between average beer consumption (shown on the y-axis above) and the 20-24 year old cohort as a percent of the total population (shown on the x-axis above).

Question: Explain what the residual plot under the scatter graph shows us.

Figure 2: Graphs to be produced during this activity



Further Questions for Analysis:

- How does the population of the 20 to 24 year old cohort vary over time?
- What function would appear to model this population well over time?
- *Fathom* calculates the correlation coefficient for the scatter graph above as 0.92. What does this mean?

Optional enrichment activity:

- Find another interesting relationship using 2 different tables from the over 2,400 multidimensional tables and 26 million time series on CANSIM on E-STAT.

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.

File: Lessons/ Lesson2a Beer Consumption Demographics.doc
 Updated: December 12, 2003 to identify page and unit for the expectations
 Updated: Feb. 5, 2004 to add the official Ont. codes for the expectations and change title
 Updated: October 27, 2004 for E-stat entry instructions and to clarify the outputs

Revised May 2005: added note about STC & Fathom & credit to Mario for original student project. See Mario's project in file 'Lesson 2a – student project'
 Revised July 15, 2005 by Mylène Abi-Zeid: reference period to 2004; graphs (screen captures); number of tables & series