E-STAT: Scatter Plot & Line of Best Fit

In this activity you will decide if you think there is a relationship between the number of Canadians 65 years and over and retail sales of drugs, vitamins and other health supplements in Canada.

***#1*** First you will look at the number of Canadians 65 years and over.

Go to **http://estat.statcan.ca**

**→ English**

**→Table of Contents**

**→ Population and demography** (in People)

**→ CANSIM - Population estimates and projections**

→ **Table 051-0001**: Estimates of population, by age group and sex, Canada, provinces and territories, annual (persons), 1997 - 2010

Select: **→Canada** for Geography

**Both sexes** for Sex

Age Group: View Checklist and footnotes

**65 and over (at the way bottom of the list)**

***1997 - 2010***

**→ Retrieve as individual time series**

Select **Downloadable File:** CSV file: Time as rows

Save as **population.csv**

Open **OPEN CALC** and open file population.csv

Delete the top two rows

Re-label the columns A and B with appropriate headings.

Make a scatter plot of the number of Canadians 65 years and over.

Determine the equation of the line of best fit.

For this scatter graph of the number of Canadians 65 years and over:

Independent variable (x-axis):

Dependent variable (y-axis):

***#2*** Now you will look at retail sales of drugs, vitamins and supplements.

Go back to the main **E-STAT Table** of Contents

**→ Retail and wholesale** (in Economy)

**→ Retail sales by type of product**

**→** Table **080-0009:** Survey of large retailers, monthly (dollars), Jan 1997 to December 2010

Select: **→Canada** for Geography

**Drugs (prescription and over-the-counter), vitamins and other health supplements** for Retail Commodity Classification

**Unadjusted** for Adjustments

Select the earliest date to the latest date. \*\* *Notice the data is monthly. \*\**

**→ Retrieve as individual time series**

Select **Screen output:** CSV file : Time as rows

Scroll down to the bottom of the page and press **Retrieve now**

**SAVE AS drugs.csv**

Open **OPEN CALC** and open file drugs.csv

Delete the top two rows

Re-label the columns A and B with appropriate headings.

Make a scatter plot of amount spent on drug, vitamins, and other health supplements.

Determine the equation of the line of best fit.

For this scatter graph of retail sales of drugs, vitamins and supplements:

Independent variable (x-axis) :

Dependent variable (y-axis) :

***#3* Creating a Scatter Graph & Line of Best Fit**

You now need to make a scatter graph with:

Independent variable (x-axis): number of Canadians 65 years and over

Dependent variable (y-axis): retail sales of drugs, vitamins and supplements

Go back one screen to the page where you select the type of graph.

Scroll down to the bottom of the page.

**→ Add more series.** E-STAT has several methods to find other data sets of interest.

**→ Browse by subject.**

**→ Population and demography**

**→ Population estimates and projections**

**→** Table **051-0001:** Estimates of population, by age group and sex, Canada, provinces and territories, annual (persons), 1997 to 2010

Select: **→Canada** for Geography

**Both sexes** for Sex

**65 years and over** for Age group (scroll down to the bottom of the list)

Select the earliest date to the latest date. \*\* *Notice the data is annual. \*\**

**→ Retrieve as individual Time Series.** You can see both of your selections at the top of the next page.

Select: **Screen output: CSV – time as rows**

You will see the following message: **ERROR**

Series must all be of the same frequency.

The data for the number of Canadians 65 years and over is **annual** but the data for retail sales of drugs, vitamins and supplements is **monthly**. They don’t match!

→ Press **OK.**

→ Scroll down. Click on **Manipulate data**.

→ Under frequency of data, select **Annual (sum).** This adds together all of the monthly data for one year of retail sales to make an annual value. Now the data will match!

→ Scroll down. Select **Retrieve now**.

Select **Screen output:** CSV file : Time as rows

Scroll down to the bottom of the page and press **Retrieve now**

**SAVE AS combined.csv**

Open **OPEN CALC** and open file combined.csv

Delete the top three rows

Re-label the columns A and B and C with appropriate headings.

Do you need the data starting from 1981??? Which rows should you delete?

Make a scatter plot comparing population over 65 with amount spent on drug, vitamins, and other health supplements.

Determine the equation of the line of best fit.

**TO HAND IN**

**Your 3 graphs AND answers to these questions:**

1. Based on your graph and the line of best fit, are the number of Canadians 65 years and over and retail sales of drugs, vitamins and health supplements correlated? Justify your answer.
2. Do you feel that Canadians 65 years and over and retail sales of drugs, vitamins and supplements have a causal relationship? This means that the number 65 years and over is directly causes the retails sales. Justify your opinion.
3. What other factors could affect the retails sales of drugs, vitamins and supplements reported in the data?