

Introduction to SEER*Stat: Frequencies

Diane Nishri

Senior Research Associate, Surveillance

February, 2011

SEER

- “The Surveillance, Epidemiology, and End Results (SEER) Program is the most authoritative source of information on cancer incidence and survival in the United States”
- Case ascertainment for SEER began on January 1, 1973, in the states of Connecticut, Iowa, New Mexico, Utah, and Hawaii and the metropolitan areas of Detroit and San Francisco-Oakland
- “Considered the gold standard for cancer registries around the world”

SEER*Stat

- Software created and maintained by the US National Cancer Institute
- Produces statistics useful for studying the impact of cancer on a population. Sessions available are:
 - Frequencies
 - Rates
 - Survival
 - Limited-duration prevalence
- Designed to work with SEER data, but adaptable to any cancer registry anywhere

SEER*Stat

- Three basic steps:
 - Define the parameters of your analysis in a SEER*Stat **session**
 - Execute the session, and review the results in the **results matrix** window
 - Use the results by printing, exporting or copying them

SEER*Stat Sessions

- Specify the statistics to calculate
 - Select the appropriate session
 - Choose statistics & set parameters (*Statistic*)
- Define the records to be used
 - Select a database (*Data*)
 - Make selections if a subset of records is desired (*Selection*)
- Define the layout for output
 - Stratify results by one or more variables (*Table*)
 - Define groupings of variables in the Dictionary
 - Provide titles & other output options (*Output*)

Frequencies: Start at the beginning

- One-way frequencies for all categorical variables
 - Invalid or missing values?
 - Unusual distribution?
- Cross tabulations for 2 or more variables to check consistency
 - Site recode by sex
 - Site recode by morphology by age group

Frequency example

- Create a table showing the frequency of Renfrew County and District HU cases by age group and sex for 2003-2007
- Age groups:
 - 0-14, 15-44, 45-54, 55-64, 65-74, 75+

Creating user-defined variables

- Open a session in SEER*Stat
- Select the incidence database
- Go to **File -> Dictionary**
- Choose the SEER*Stat variable needed to create the user-defined variable
- On the right-hand menu, click **Create**
- Define the groupings needed
- Name the new variable appropriately
- Variables are in the **User-Defined** folder

Frequency Session & Data Tab

SEER*Stat 6.6.2

File Edit Session Window Help

Local Data: H:\Surveillance\SEERStat\SEERStat Fe

Frequency Session-4

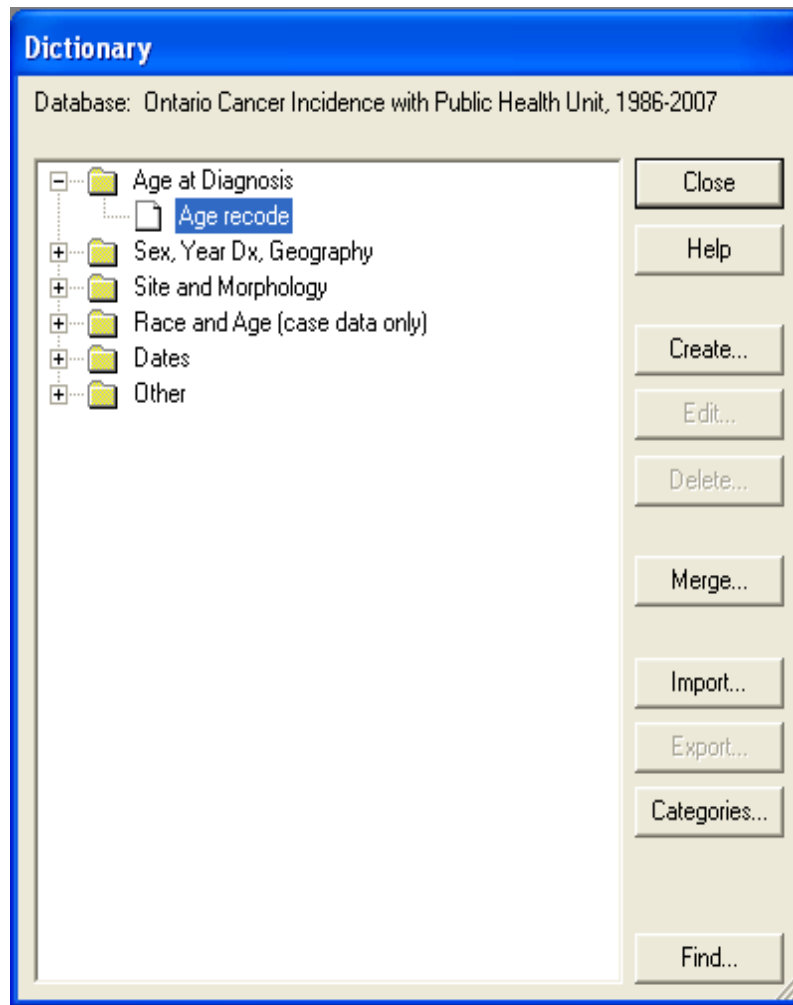
Data | Statistic | Selection | Table | Output

Database Name	Linked To	Linked B
Ontario Cancer Incidence 1981-2007		
Ontario Cancer Incidence with Census Division, 1986-2007		
Ontario Cancer Incidence with Public Health Unit, 1986-2007		
Ontario Cancer Incidence with LHIN, 1986-2007		
Ontario Cancer Mortality, 1981-2007		
Ontario Cancer Mortality with Census Division, 1986-2007		
Ontario Cancer Mortality with Public Health Unit, 1986-2007		
Ontario Cancer Mortality with LHIN, 1986-2007		

Original Sort Order * Change Linked Database...

Suggested citation for the selected database:
Cancer Care Ontario - SEER*Stat Release 8 - OCRIS (May 2010) released February 2011. Population Data Source: Pop Est Summary (Statistics Canada, Ontario Ministry Finance), Ontario Ministry of Health and Long-Term Care: IntelliHEALTH ONTARIO, extracted October 2010 (1986-2007).

File -> Dictionary



Original variable definition

Edit Variable - Age recode

Name: Category:

Description:

Groupings

- 00-04 years
- 05-09 years
- 10-14 years
- 15-19 years
- 20-24 years
- 25-29 years
- 30-34 years
- 35-39 years
- 40-44 years
- 45-49 years
- 50-54 years
- 55-59 years
- 60-64 years
- 65-69 years

Values (0 Selected)

- 00-04 years
- 05-09 years
- 10-14 years
- 15-19 years
- 20-24 years
- 25-29 years
- 30-34 years
- 35-39 years
- 40-44 years
- 45-49 years
- 50-54 years
- 55-59 years
- 60-64 years
- 65-69 years
- 70-74 years

Save to Dictionary

Create new groupings and labels

Edit Variable - Age recode

Name: Category:

Description:

Groupings

- 00-14 years
- 15-44 years
- 45-54 years
- 55-64 years
- 65-74 years
-

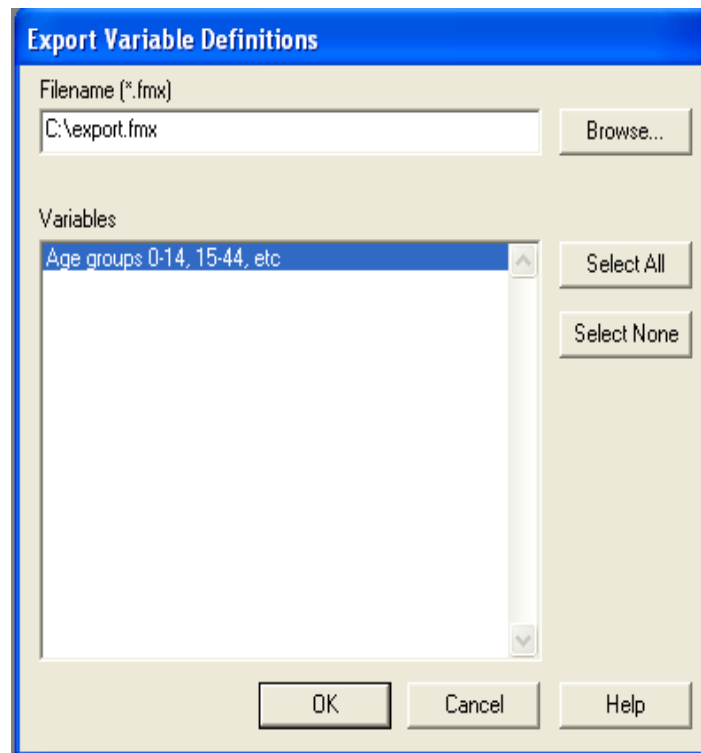
Values (3 Selected)

- 20-24 years
- 25-29 years
- 30-34 years
- 35-39 years
- 40-44 years
- 45-49 years
- 50-54 years
- 55-59 years
- 60-64 years
- 65-69 years
- 70-74 years
-
-
-

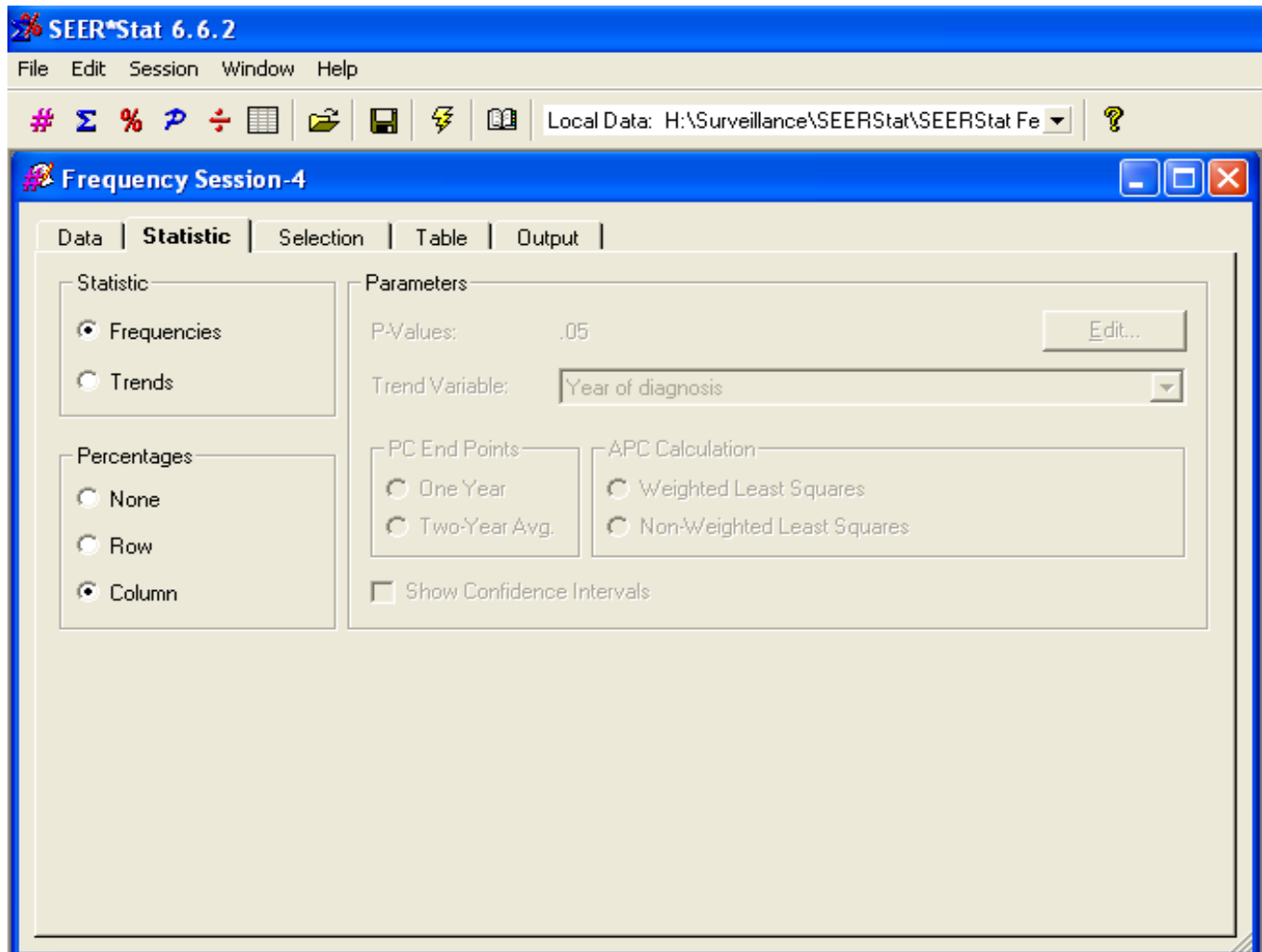
Save to Dictionary

How to share User-Defined variables

- File -> Dictionary... -> Export... or Import...
- A dictionary is specific to a database



Statistic Tab



Selection Tab

The screenshot displays the SEER*Stat 6.6.2 interface. The main window is titled "Frequency Session-4" and has tabs for "Data", "Statistic", "Selection", "Table", and "Output". The "Selection" tab is active, showing a "Select Only" section with the following options:

- Malignant Behavior
- Known Age
- Male or Female Sex
- Cases in Limited-Use Database

An "Edit..." button is located below these options. A "Case Selection" dialog box is open in the foreground, showing a tree view of variables on the left and a list of values on the right. The "Year of diagnosis" variable is selected, and the operator "is not = to" is chosen. The values list includes years from 1999 to 2007, with 2003-2007 highlighted. The "Conjunction" is set to "AND".

Variable	Operator	Values
Age at Diagnosis	is = to	1999
Sex, Year Dx, Geography	is not = to	2000
Sex		2001
Year of diagnosis		2002
Public Health Unit		2003
Site and Morphology		2004
Race and Age (case data only)		2005
Dates		2006
		2007

Additional options in the dialog include "Modify", "Conjunction: AND", "Find...", "New Line", and "NAACCR Item #=390". The "Selection Statement" field contains the following text:

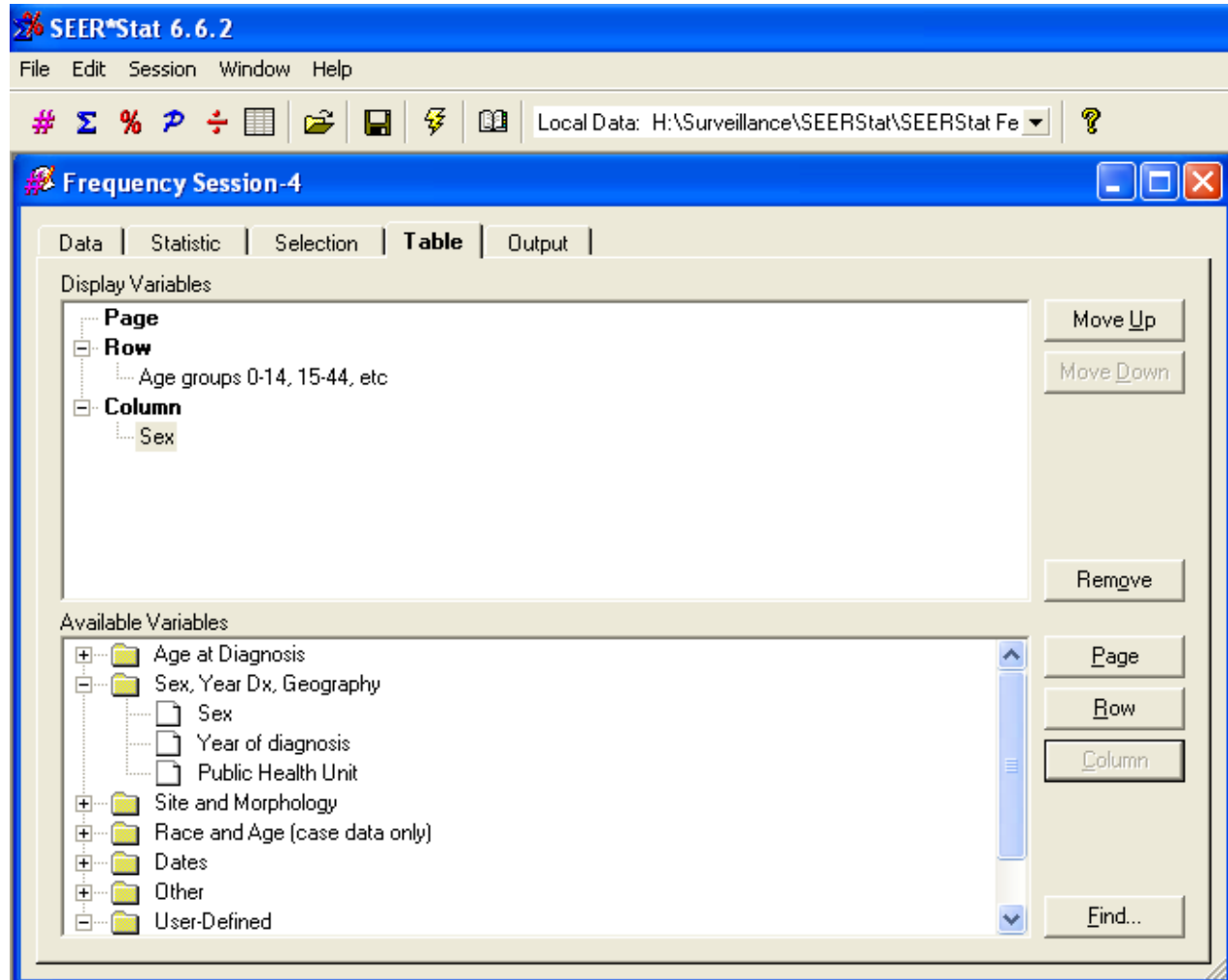
```
{Sex, Year Dx, Geography, Public Health Unit} = '3557 Renfrew County and District HU'  
AND {Sex, Year Dx, Geography, Year of diagnosis} = '2003', '2004', '2005', '2006', '2007'
```

Buttons for "Up", "Down", "Delete", "Copy", and "Paste" are on the right side of the dialog. At the bottom, there are "Add (...)", "Delete (...)", "Del All (...)", "OK", "Cancel", and "Help" buttons.

Select Only the First Matching Record...

- If you check this option, your selection may result in fewer cases than expected
- SEER*Stat will select the first cancer case, regardless of the site, and then select the cases which meet the other selection criteria
- If your database contains all incident cases, and a person has a non-colorectal case before their colorectal case, they would be excluded from a colorectal analysis if this option is checked

Table Tab



Selection Tab vs. Table Tab

- Use the Selection tab to reduce the number of records analyzed based on specific variables
- Use the Table tab to choose variables to include in the output table
- A case is excluded from your analysis if it does not pass the selection criteria of the Selection tab and match at least one table definition of the Table tab. In general, it is best to select your case subset on the Selection tab rather than relying on the table definitions to exclude cases.

Output tab

The screenshot displays the SEER*Stat 6.6.2 software interface. The main window is titled "SEER*Stat 6.6.2" and includes a menu bar (File, Edit, Session, Window, Help) and a toolbar with various icons. A secondary window, "Frequency Session-4", is open, showing the "Output" tab. The "Output" tab contains a text area for a title, a dropdown for "Number of Decimal Places for Trends" (set to 0.1), a checkbox for "Hide Statistics When Fewer Than" (set to 6 Cases), and a checkbox for "Include Individual Year Counts".

SEER*Stat 6.6.2

File Edit Session Window Help

Σ % \mathcal{P} \div Local Data: H:\Surveillance\SEERStat\SEERStat Fe

Frequency Session-4

Data | Statistic | Selection | Table | **Output**

Title (Up to 5 lines)

Cancer Incidence by Age Group and Sex
Renfrew County & District Health Unit, 2003-2007

Number of Decimal Places for Trends: 0.1 Set Default

Hide Statistics When Fewer Than 6 Cases

Include Individual Year Counts

Results Matrix

Frequency Matrix-4 _ □ ×

Renfrew County & District Health Unit cancer incidence by age group and sex, 2003-2007

	Male and female			Male			Female		
	Count	Column %	Cum %	Count	Column %	Cum %	Count	Column %	Cum %
00-14 years	20	0.66%	0.66%	10	0.65%	0.65%	10	0.67%	0.67%
15-44 years	164	5.44%	6.11%	58	3.79%	4.45%	106	7.15%	7.82%
45-54 years	313	10.39%	16.50%	124	8.11%	12.56%	189	12.74%	20.57%
55-64 years	638	21.18%	37.68%	332	21.71%	34.27%	306	20.63%	41.20%
65-74 years	869	28.85%	66.53%	506	33.09%	67.36%	363	24.48%	65.68%
75+ years	1,008	33.47%	100.00%	499	32.64%	100.00%	509	34.32%	100.00%
Total	3,012	100.00%	100.00%	1,529	100.00%	100.00%	1,483	100.00%	100.00%

^ Statistic not displayed due to fewer than 6 cases.

“^ Statistic not displayed due to fewer than 6 cases”

- If we had restricted the time period to 2005-2007, there would have been fewer than 6 male cases aged 0-14
- The total would not be reported, and there would be no percentages calculated for males
- Options:
 - More years
 - Different age groups (0-44)
 - Combine sexes

Matrix Menu

The screenshot shows the SEER*Stat 6.6.2 interface. The 'Matrix' menu is open, displaying options: Retrieve Session, Options..., Font..., Order, Filter..., Lock, Page, Export, and Properties... The main window displays a table titled 'cancer incidence by age group and sex, 2005-2007'. The table is structured as follows:

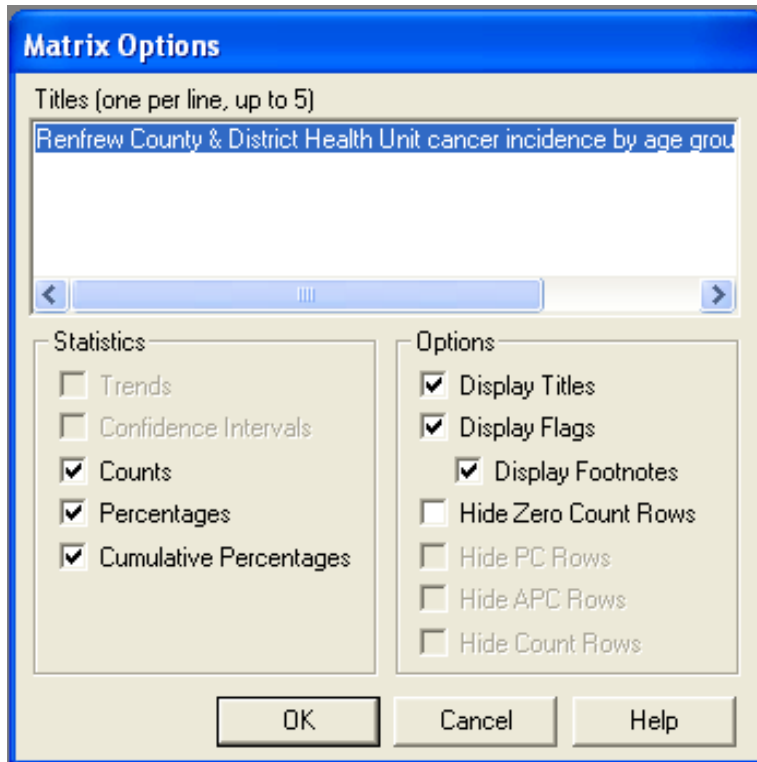
	and female			Male		
	column %	Cum %	Count	Column %	Cum %	Count
00-14 years	0.60%	0.60%	^	^	^	^
15-44 years	4.83%	5.43%	36	^	^	^
45-54 years	10.26%	15.69%	73	^	^	^
55-64 years	21.12%	36.81%	201	^	^	^
65-74 years	27.54%	64.34%	292	^	^	^
75+ years	650	35.66%	100.00%	323	^	^
Total	1,823	100.00%	100.00%	^	^	^

At the bottom of the window, a status bar indicates: ^ Statistic not displayed due to fewer than 6 cases.

Matrix Menu

Retrieve Session	Extract the session that created the matrix
Options	View or edit matrix options
Export	Export the results to a text file for use in other applications
Order	Change the order of row/column variables
Lock	Lock a row/column while scrolling

Matrix Options



- Choices that you make here will affect the display of the matrix on your screen as well as in printed, copied, or exported data.
- Calculation of statistics is not affected.
- If you change the title in Matrix Options and then retrieve the session, the session will contain the original, not the edited title.

Saving your work

- To save the parameters of a session or a results matrix: **File -> Save or Save As.**
- A session file may be saved at any time, even if it has not been executed.
- It is not necessary to save both the session and its results matrix. A saved matrix contains a snapshot of the session at the time of execution, or a session may be run again to recreate the matrix.
- You can also save your session's parameters as a text file in Windows' INI format: **File -> Export As Text.**

Copy selected page(s) to Excel

SEER*Stat 6.6.2

File Edit Matrix Window Help

Copy Cell Ctrl+C
Highlight All Page Ctrl+P
All Pages Ctrl+M
Session Information Ctrl+R

Frequency Matrix

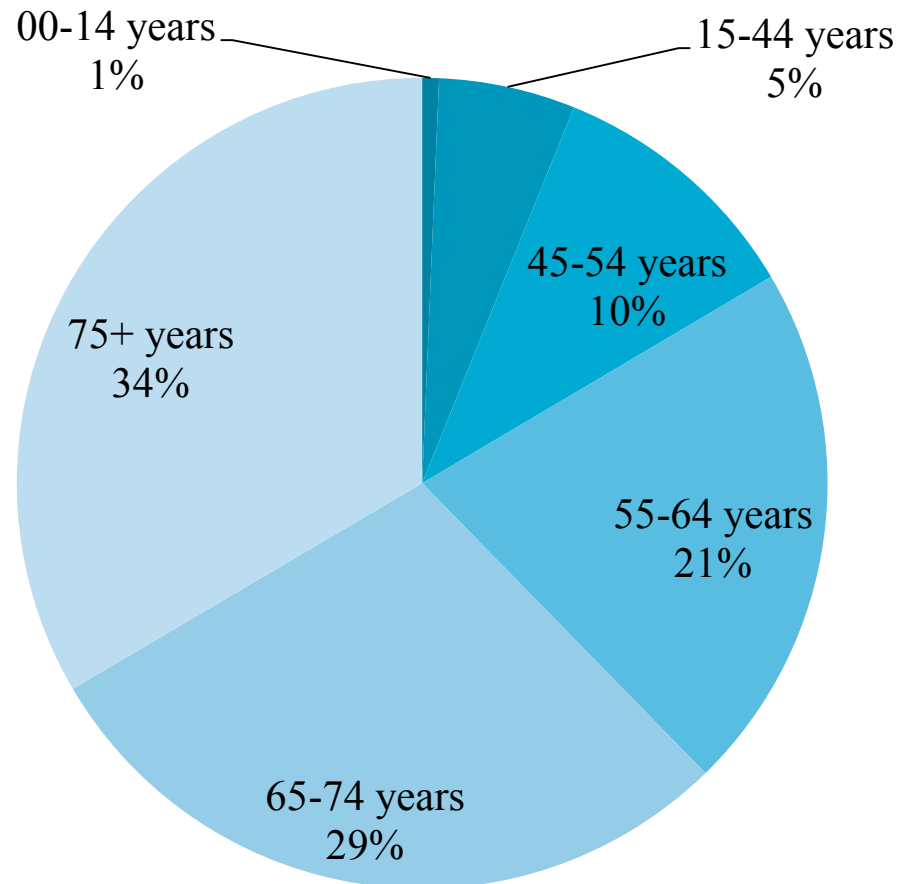
Renfrew County & District health care cancer incidence by age group and sex, 2003-2007

	Male and female			Male		
	Count	Column %	Cum %	Count	Column %	Cum %
00-14 years	20	0.66%	0.66%	10	0.65%	0.65%
15-44 years	164	5.44%	6.11%	58	3.79%	4.45%
45-54 years	313	10.39%	16.50%	124	8.11%	12.56%
55-64 years	638	21.18%	37.68%	332	21.71%	34.27%
65-74 years	869	28.85%	66.53%	506	33.09%	67.36%
75+ years	1,008	33.47%	100.00%	499	32.64%	100.00%
Total	3,012	100.00%	100.00%	1,529	100.00%	100.00%

^ Statistic not displayed due to fewer than 6 cases.

Create a graph to display your results

**New cancer cases by age group
Renfrew County & District Health Unit 2003-2007**



Exercise #1: Frequency Session

- For your Health Unit, create a frequency table of cancer incidence by age group and sex. Choose an appropriate time period. If possible, use Excel or other software to graph your results
- **BONUS:** Repeat with mortality data