

# Intermediate Analysis

## S U M M A R Y

---

This lesson pulls together many of the skills you learned in Lessons 1-10 while adding more advanced data management concepts. Many skills learned in the Basic tutorial will be reinforced through the Intermediate Analysis Lesson including, DEFINE, ASSIGN, LIST, DISPLAY, READ, WRITE, SELECT, GRAPH, and MAP. You will also be introduced to a new command called RELATE.

This lesson introduces you to ways you can use Epi Info™ to create everyday statistics needed in epidemiologic work. Use the Tear Out Answer Sheet and To Do List to record your results as you progress through the lesson. Many of the sections heavily emphasize data management skills needed to combine a variety of data sources for analysis.

Length of time to complete: 3 hours Intermediate/Advanced

### **Scenario**

Presentation of the scenario that drives the tasks learned in the Intermediate Analysis Lesson. In this lesson, you will be preparing and analyzing data that could be used to write a proposal requesting legislative funds for an asthma initiative. Introduction to Advanced Statistics and the RELATE command.

### **Use BRFSS Data**

Focus on using data downloaded from the BRFSS website to compute lifetime and current prevalence rates. Introduction to Complex Sample Frequencies.

### **Use Primary Data**

Use the primary data source (800 survey records) to complete a Frequency stratified by school.

### **Use Census Data**

Focus on using data downloaded from the United States Census website in conjunction with the original 800 record dataset to compute the county child asthma prevalence.

### **Mortality Rates**

Create a RELATE between a denominator and numerator table to compute the asthma mortality rate for the county and the state. You will also Graph the rate.

## **10-Year Trends**

Create a graphic representation of 10-year hospitalization discharge rates based on age groups.

## **3-Year Rates**

Create comparison graphs for 3-year hospital discharge rates for state and county by age.

## **Mapping Rates**

Map asthma hospitalization rates by zip code to determine high risk areas.

## **Create a Scatter Plot**

Create a scatter plot to determine if income affects the rate of asthma hospitalization.

## **Risk Ratios**

Create a program (.PGM) and routeout file to compute asthma risk ratios for a series of zip codes. Use the DIALOG command to create a prompt for users.

## **Use StatCalc**

Use the Utility program StatCalc to create an estimate of sample size.

BEFORE YOU BEGIN

---

Complete all Ten Beginner and Intermediate Lessons.

WHAT YOU NEED

---

Asthma Final.MDB

Tear Out Answer Sheet and To Do List

Appendix F Intermediate Analysis Answer Key

BRFSS2003 Data Table

CensusAge.HTM

Population Data Table

Numerators Data Table

NYDisAge10Yr Data Table

Age Proxy Data Table

DisCtyAge Data Table

NY3Yr Data Table

Dis0103byZip Data Table

IncomeData Data Table

RatebyZip Data Table

ALBZCTA\_region.SHP

ALBZCTA\_region.DBF

ALBZCTA\_region.SHX

ALBZCTA\_region.PRJ

DischargesbyZip Data Table

#### FIVE GOALS

---

- Understand data management fundamentals involved in using Epi Info™ with a variety of data sources.
- Understand the use of the RELATE command and the ways tables can be linked to create rates.
- Be able to use the DEFINE and ASSIGN commands to create new variables and assign values to them.
- Be able to locate prevalence rates and confidence limits from table results.
- Be able to use the SELECT and IF/THEN commands to specify subsets of data that needs to be managed or analyzed.

## Scenario

The local health department (LHD) director of County X gets notice from the county administrator/executive that there are discretionary legislative funds for a health initiative. There is a good chance of getting the funding for asthma as the LHD has already done some work in the area, and will be able to show some results with added funds. The director sits down with the health education coordinator/planner to identify data that will be needed to write a strong proposal.

A series of questions and datasets have been developed to aid in the creation of the proposal. Answers to the questions in the form of data tables, graphs, and maps can be used to illustrate the need for asthma funding and for writing the proposal. This lesson focuses on preparing the data for the proposal, but does not cover creating the actual report as Lessons 1-10 did.

Datasets used in this lesson come from a range of sources and will help to show you the many ways Epi Info™ can be used to manage data sources. Questions deal with the prevalence of asthma across subgroups and regions as well as hospital discharge and mortality rates. For this lesson, you will use data downloaded from the BRFSS website, the United States Census website, CDC WONDER, and SPARCS.

See Appendix G: Preparing Data for Use in the Epi Info™ Tutorial for more information on gathering data and preparing to work with Epi Info™ and Epi Map.

See Appendix I: Data Sources for more information on the provided data sources and web links to guide you.

## To Do List for Intermediate Analysis Lesson

Expected results and sample .PGMs are listed in *Appendix F Intermediate Analysis Answer Key*.

- Familiarize yourself with the scenario. You are developing data to aid in the writing of a proposal to request asthma funding.
- Read through the Advanced Statistics section and the information on using the RELATE command.
- Read the What is BRFSS information.
- Read the How to Get BRFSS information.
- Use BRFSS data to compute lifetime and current asthma prevalence for the state.
- Use the 800 records in your primary data source to compute prevalence rates for each of the schools in the survey.
- Familiarize yourself with the Sampling Scheme.
- Use data downloaded from the US Census Bureau to create a table with the total population per age for the county (denominator table).
- Use the primary data source to create a table that contains information about the survey respondents age and sex (numerator table).
- Create a weight variable for the county child data.
- Create a PSU variable for the county child data.
- Compute the county child asthma prevalence using the Complex Sample Frequencies command.
- Relate data from CDC WONDER, the primary data source, and New York hospital discharge data from SPARCS
- Use DEFINE and ASSIGN to calculate the asthma Mortality Rate.
- Create a bar graph of the mortality rates.
- Use GRAPH to create a line graph that illustrates the 10-year trends in hospitalization for asthma by age group.
- Use GRAPH to create a line graph for the state and a line graph for the county to illustrate 3-three asthma hospitalization discharge rates.
- Use MAP to plot hospitalization rates by zip code.

## INTERMEDIATE ANALYSIS

- Use GRAPH to create a scatter plot that illustrates how income affects the rate of asthma hospitalization.
- Use the TABLES command to create asthma hospitalization risk ratios between zip codes in the state.
- Use the StatCalc program to estimate a sample size for the number of records needed from a hospital sample to achieve a 95% confidence interval for your statistics.

### Tear Out Answer Sheet

Use this sheet to record your answers in the *Answers* column as you progress through the Intermediate Analysis Lesson. Answers are listed in *Appendix F Intermediate Analysis Answer Key*.

A. Questions in the lesson are marked with a letter

<b>Computing Prevalence- Using BRFSS Data</b>	<b>Results</b>
A. What is the lifetime asthma prevalence among adults in the state?	
B. What is the current asthma prevalence among adults in the state?	
C. Is the state adult asthma prevalence rate higher in the subgroup gender?	Males- Females-
D. Is the state adult asthma prevalence rate higher in the subgroup ethnicity?	Hispanics-

<b>Primary Data- 800 Records</b>	<b>Results</b>
E. From the school asthma survey data, what is the difference in prevalence rates?	School A- School B- School C- School D-

<b>Census Data</b>	<b>Results</b>
F. What is the county childhood asthma prevalence?	
G. What is the county childhood asthma prevalence for the subgroup gender?	Males- Females-

<b>Mortality Data</b>	<b>Results</b>
H. Is the asthma mortality rate for the state and county	State 2000-

INTERMEDIATE ANALYSIS

significantly different?	State 2001- State 2002- County 2000- County 2001- County 2002-
--------------------------	--

<b>10 Year Trends- Hospital Discharge Data</b>	<b>Results</b>
I. Is the rate of hospital discharge for asthma increasing or decreasing in any age groups?	

<b>3 Year Rates- Hospital Discharge Data</b>	<b>Results</b>
J. Are there significant differences in the 3-year asthma discharge rates for the state versus the county by age?	

<b>Mapping Rates</b>	<b>Results</b>
K. Based on the map with hospitalization rates by zip code, which areas have the highest hospitalization rates for asthma?	

<b>Scatter Plots</b>	<b>Results</b>
L. Does income appear to affect the rate of hospitalization for asthma?	

<b>Risk Ratios- Hospitalization Data</b>	<b>Results</b>
M. What is the risk ratio of being hospitalized for asthma if you live in the zip code 12202 versus if you lives in the zip code	



INTERMEDIATE ANALYSIS

12211?	
<b>StatCalc</b>	<b>Results</b>
N. From the New York hospital discharge data, how many records do you need to sample to get a 95% confidence interval from a sample size of 38998?	

## Advanced Statistics

The Complex Sample Frequencies (FREQ), Complex Sample Tables (TABLES), and Complex Sample Means (MEANS) commands in the Analysis program perform statistical calculations that assume data comes from simple random (or unbiased systematic) samples. In many survey applications, more complicated sampling strategies are used. These may involve sampling features like stratification, cluster sampling, and the use of unequal sampling fractions. Surveys that include some form of complex sampling include the coverage surveys of the WHO Expanded Program on Immunization (EPI) (Lemeshow and Robinson, 1985) and CDC's Behavioral Risk Factor Surveillance System (Marks et al., 1985).

The CSAMPLE functions compute proportions or means with standard errors and confidence limits for studies in which the data did not come from a simple random sample. If tables with two dimensions are requested, the odds ratio, risk ratio, and risk difference are also calculated. Data from complex sample designs should be analyzed with methods that account for the sampling design.

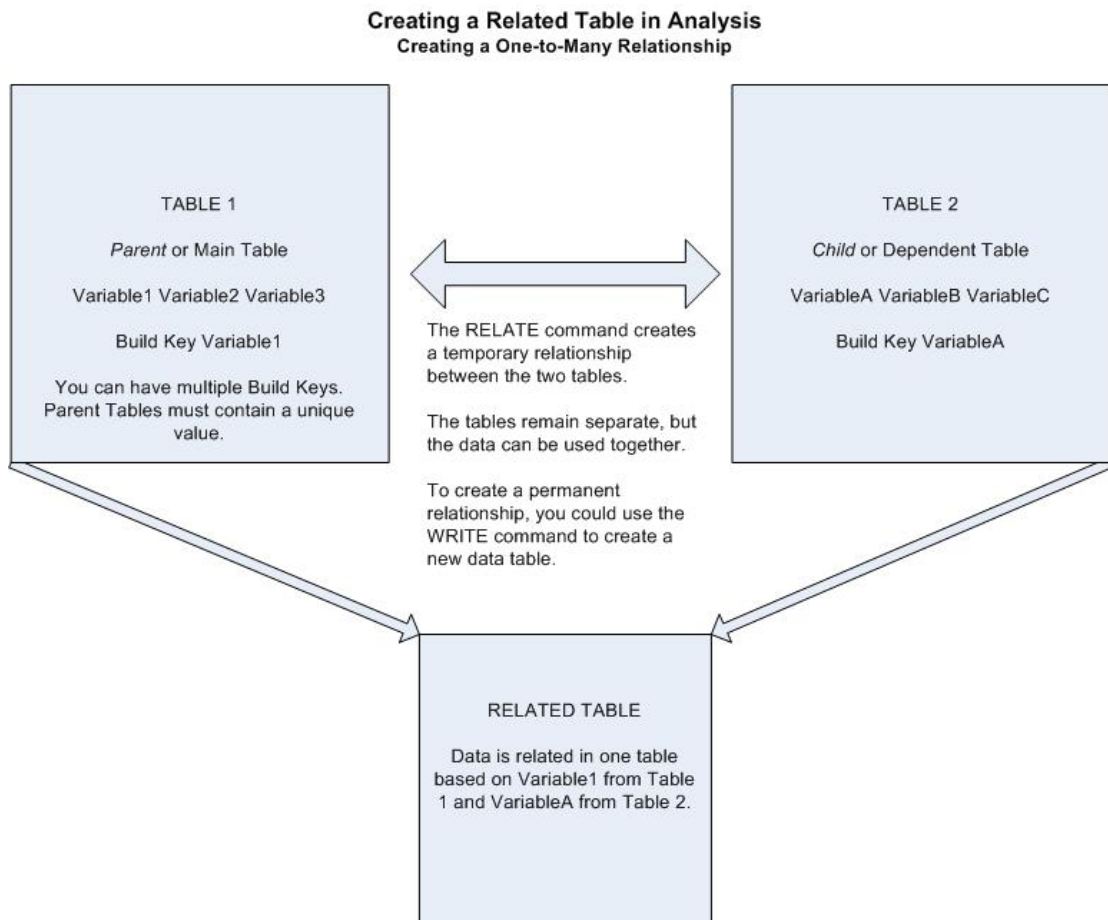
The following sections rely heavily on a new command called RELATE. The RELATE command allows you to link one or more tables to the current dataset using a common identifier (key) to find matching records. The key can include several matching fields. The linked tables remain separate.

## RELATE Command

To use RELATE, at least one table must have been made active with the READ command. The table to be linked must have a key field that has identified related records in the two tables. The keys in the main and related tables or files do not have to have the same name.

After issuing the RELATE command, the variables in the related table may be used as though they were part of the main table. Where variable names are duplicated in the related tables, the variable names will be suffixed with a sequence number.

Frequencies, cross-tabulations, and other operations involving data in both the main and related tables can be performed. You can also use the WRITE command to create a new table. More than one table can be related to the main table by using a series of RELATE commands.



## What is BRFSS?

The Behavioral Risk Factor Surveillance System (BRFSS) is a state-based system of health surveys that generates information about health risk behaviors, clinical preventive practices, and health care access and use primarily related to chronic diseases and injury. The BRFSS is a cross-sectional telephone survey conducted by state health departments with technical and methodological assistance provided by the CDC. Every year, states conduct monthly telephone surveillance using a standardized questionnaire to determine the distribution of risk behaviors and health practices among non-institutionalized adults. The states forward the responses to the CDC, where the monthly data are aggregated for each state. The data are returned to the states, and then published on the BRFSS Web site. BRFSS interviewers ask questions related to behaviors that are associated with preventable chronic diseases, injuries, and infectious diseases. States use standard procedures to collect data through monthly telephone interviews. (National Center for Chronic Disease Prevention and Health Promotion <http://www.cdc.gov/brfss/faqs.htm#1>)

## How to Get BRFSS Data

BRFSS data is readily available and can be downloaded from the BRFSS website at <http://www.cdc.gov/brfss/index.htm>.

This section of the tutorial deals with year 2003 data downloaded from the BRFSS website. Data can be downloaded into a variety of formats. For this lesson, you will be working with data derived from a .DBF file. The original .DBF was extracted to a folder on the local drive.

The file was Read into Analysis. Analysis accepted the first 255 variables included in the dataset. The .DBF file contained over 264,000 records. This encompasses all the records for surveyed areas.

For this example, you are working with records from New York State. You would select only the records for New York and then write those records to a new file.

To obtain only the information dealing with a specific topic, in this case asthma, you would visit the BRFSS website and obtain detailed information on which variables pertain to asthma and also the codes used to signify race, gender, ethnicity, etc. Once you know the variable names, you can select only the questions/variables related to asthma and write those to a new table. This brings the record count down to 5,544.

Because of how the BRFSS is conducted (multi-stage sample process) with variables that are not equally weighted you will use the C-Sample (complex sample routines) instead of normal tables and means commands. The BRFSS dataset contains variables for Weight, Primary Sampling Unit (PSU), and Stratification, already computed as part of the table.

For this example, due to the large size of the dataset, the BRFSS data has already been downloaded and written out to a new file specifically dealing with asthma data. You can use the data table contained in the Asthma Final project to compute the prevalence of lifetime and current asthma in the state of New York. The Asthma Final project contains a table called BRFSS2003 for use in the following lesson.

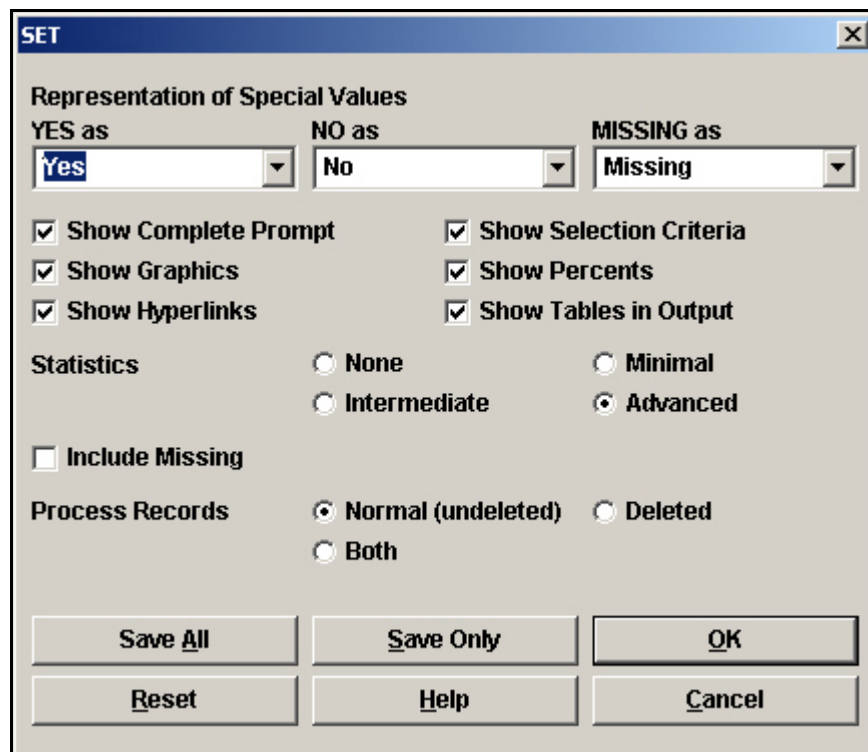
## Read/Import the BRFSS Data

To compute the prevalence using BRFSS data, you can use the Analysis command Complex Sample Frequencies. Complex Sample Frequencies can be used to determine the following prevalence rates needed for your proposal.

A. What is the lifetime asthma prevalence among adults in the state?

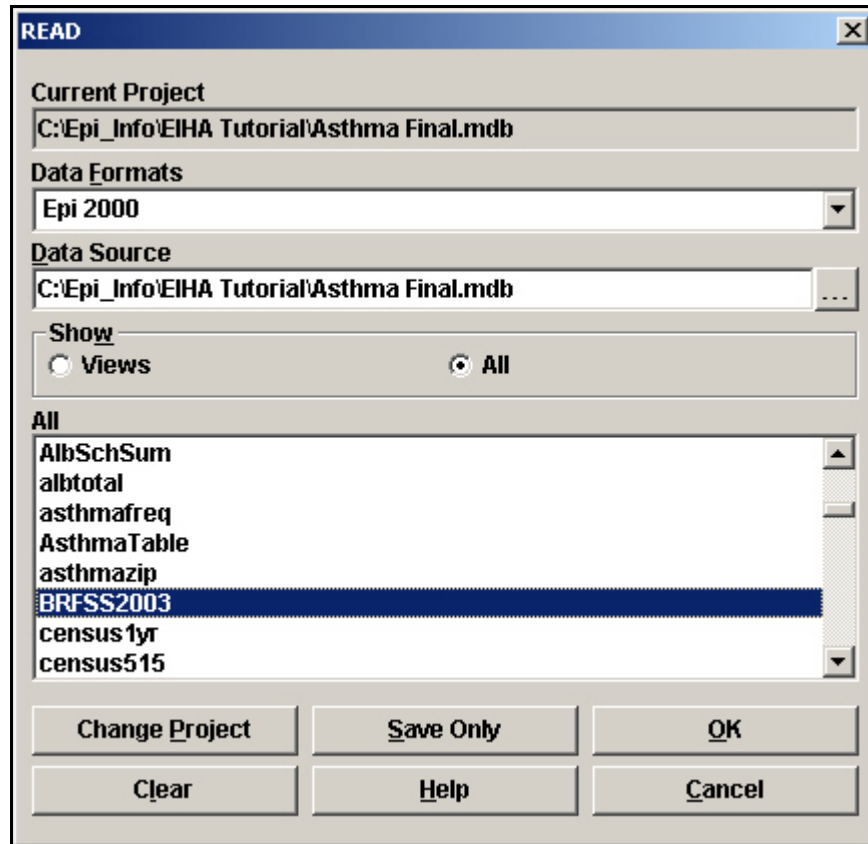
To get the prevalence for the county, you could use Expanded Behavioral Risk Factor Surveillance System (EBRFSS) data, if it is available to you. You can determine increase or decrease by following the same steps outlined below, but use prior year or later year data. You would need to download, select, and format data for the requested year from the BRFSS website.

1. Click **Analyze Data**. The Analysis module opens.
2. From Command Tree Options folder, click **Set**. The SET dialog box opens.



3. From the Statistics section, select the Advanced radio button.
4. Click **OK**.
5. From the Command Tree Data folder, click **Read/Import**. The READ dialog box opens.

6. Click **Change Project**.
7. Locate the EIHA\_Tutorial folder and select the project file called Asthma Final.MDB.
8. Click **Open**.
9. Click the **Show All** radio button. The All field populates with all the tables in the project.



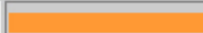


10. Select the BRFSS2003 table.
11. Click **OK**. You should have a record count of 5,544. This is the number of records for the state of NY.
12. From the Command Tree Variables folder, click **Display**. The DISPLAY dialog box opens.
13. Click **OK** to see all the variables in the table.
  - These variables were selected out of the larger dataset and written to this new table called BRFSS2003. Information in this table pertains only to New York for the year 2003.
  - By consulting the BRFSS website, you know that the calculated variable that contains information on the lifetime prevalence of asthma is called `_LTASTHM`. Since you

imported the file as a .DBF, the variable name contains a prefix of A to compensate for the space field. The variable you need to figure the prevalence is called A\_LTASTHM. This is a calculated variable based on the BRFSS question, "Have you ever been told by a doctor, nurse, or other health professional that you had asthma?"

- By using the DISPLAY command, you can also see that the majority of variables have been coded as Numbers.

14. Create a Frequency of the A\_LTASTHM.

A_LTASTHM	Frequency	Percent	Cum Percent	
1	4875	87.9%	87.9%	
2	660	11.9%	99.8%	
9	9	0.2%	100.0%	
<b>Total</b>	5544	100.0%	100.0%	

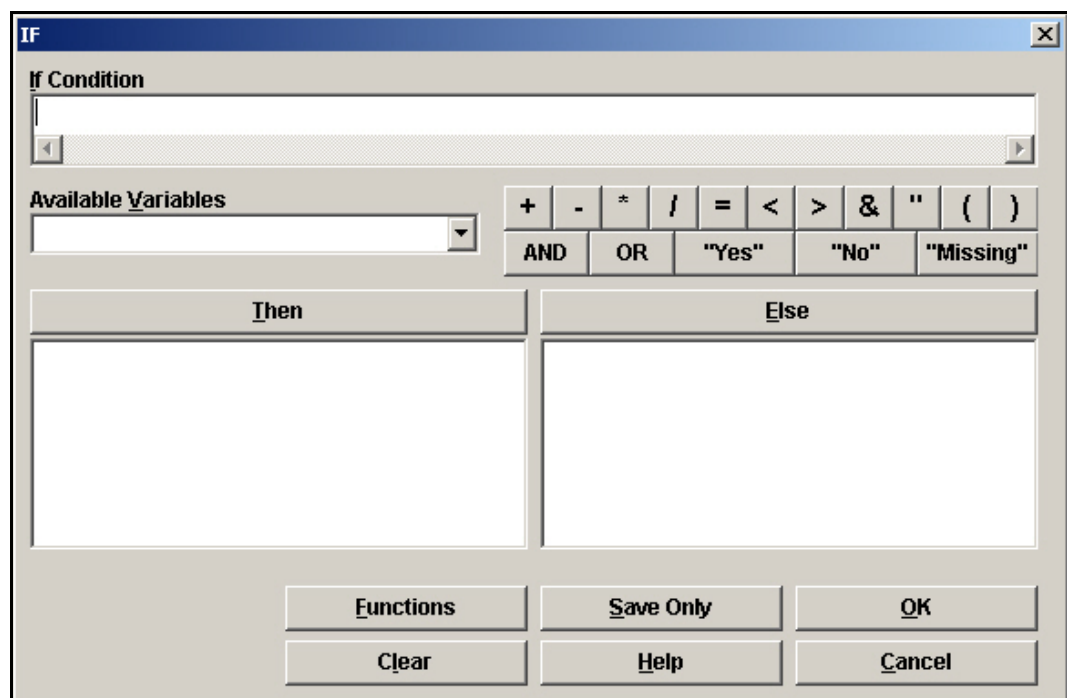
- Notice the frequency produces results for NO answers (the number 1) and YES answers (the number 2). There is also a number 9 that indicates missing values. To enhance your statistics, you want to remove the missing values from the calculations.

## Create an If/Then Statement

In Lesson 2 Check Code, you created an IF/THEN statement to create a skip pattern. Now you will use Analysis to create an IF/THEN statement that removes missing values from your statistics. The IF dialog box in Make View and Analysis are the same; however, the navigation from the THEN is slightly different, rather than selecting from the tabs as you do in the Make View Program Editor, you will select from the Command Tree.

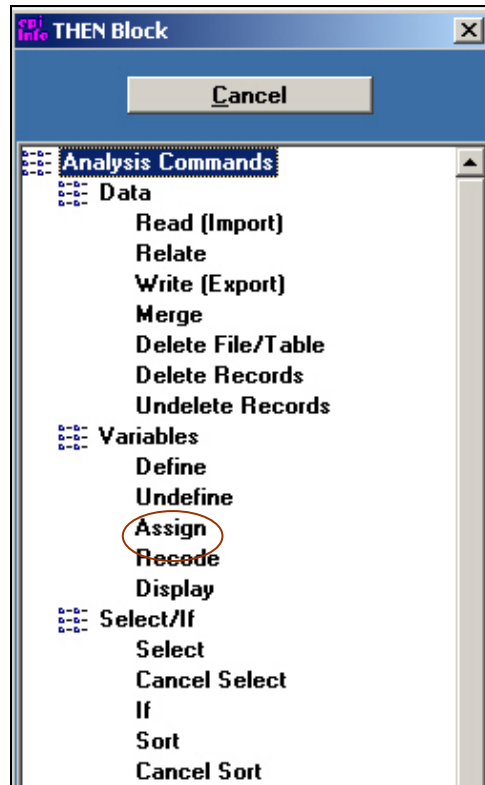
You are making the statement that **IF** the value of A\_LTASTHM is 9 **THEN** it will be considered Missing. Missing values will not be included in the statistics because you did not select to include them when you opened the SET dialog box in the previous section.

1. From the Command Tree Select/If folder, click **IF**. The IF dialog box opens.

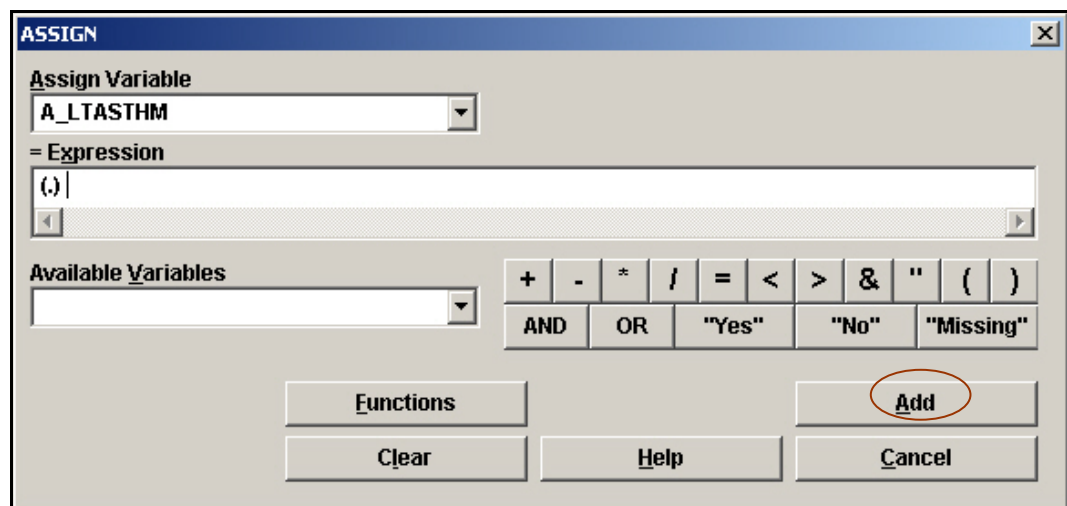


2. From the Available Variables drop-down, select A\_LTASTHM. The If Condition field populates.
3. Click =.
4. Type the number 9.
5. Click **Then**. The Command Tree window activates.





6. From the Command Tree Variables folder, click **Assign**. The ASSIGN dialog box opens.
7. From the Assign Variable drop-down, select A\_LTASTHM.
8. Place your cursor in the =Expression field.
9. Click **Missing**.



10. Click **Add**. The IF dialog box opens.

The screenshot shows the 'IF' dialog box with the following details:

- If Condition:** A\_LTASTHM=9
- Available Variables:** A\_LTASTHM
- Then:** ASSIGN A\_LTASTHM= (.)
- Else:** (Empty)
- Buttons:** Functions, Save Only, OK, Clear, Help, Cancel

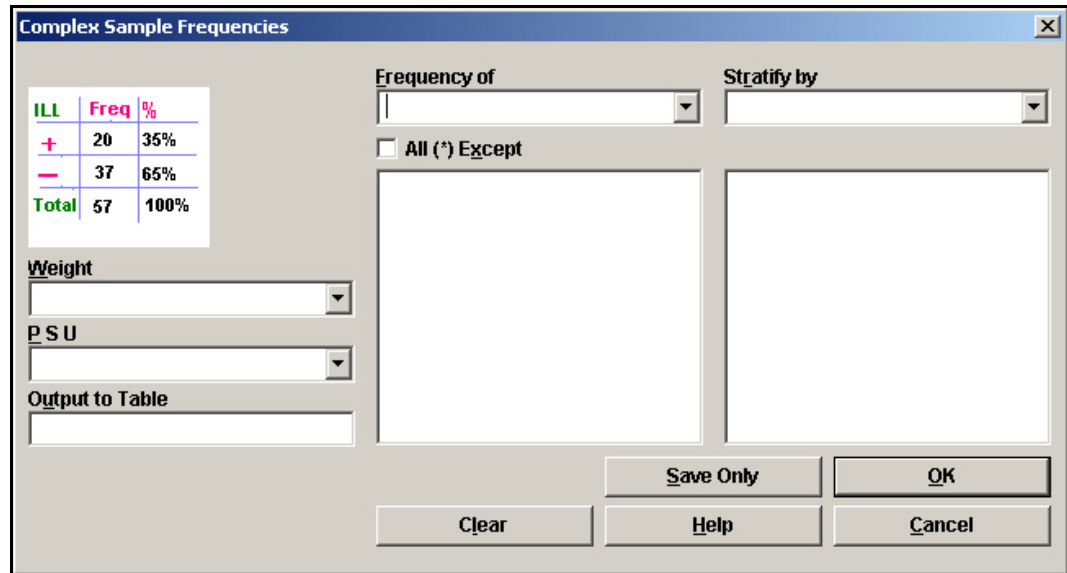
11. Click **OK**. Your code appears in the Program Editor:

```
IF A_LTASTHM=9 THEN  
  ASSIGN A_LTASTHM= (.)  
END
```

## Complex Sample Frequency

Now that the Missing values have been excluded, you can compute the prevalence.

1. From the Command Tree Advanced Statistics folder, click **Complex Sample Frequencies**. The Complex Sample Frequencies dialog box opens.



2. From the Frequency Of drop-down menu, select A\_LTASTHM.
3. From the Weight drop-down, select A\_FINALWT. The Weight variable has been computed for you, and is part of the BRFSS data. To find out how this variables is calculated refer to the BRFSS website.
  - When variables are not weighted, each record counts as the same, one record represents one person. Weighting attempts to compensate for circumstances where variables must represent more than one individual due to the sampling process, one record may count as 100 people, depending on the sample size. Weighting may also be needed when you do not have the original data, but only summarized data showing frequency of particular responses.
4. From the PSU drop-down menu, select A\_PSU. The PSU is the Primary Sampling Unit. The PSU variable has been computed for you and is part of the BRFSS data. To find out how this variables is calculated refer to the BRFSS website.
  - In cluster sampling, the PSU represents the clusters from which the samples were drawn randomly; the clusters themselves are selected from a set of similar clusters, which together include all of the population to be sampled. The random sample of the population from the random sample of clusters is taken as a proxy for a random sample from the population as a whole. Complex sample analysis attempts to adjust for the reduction of variance due to the fact that each member of the population did

not have a uniform probability of selection. Clusters are rarely chosen on the basis of variables such as age, gender, or race.

- Click **OK**. Results appear in the Output window.


The code in the Program Editor looks like this:

```
FREQ A_LTASTHM WEIGHTVAR=A_FINALWT PSUVAR=A_PSU
```

Your Output looks like the following.

A_LTASTHM	TOTAL	
<b>1</b>	4875	
Row %	100.000	
Col %	88.306	
SE %	0.507	
LCL %	87.313	
UCL %	89.300	
<b>2</b>	660	
Row %	100.000	
Col %	11.694	Prevalence
SE %	0.507	
LCL %	10.700	Confidence Intervals
UCL %	12.687	
<b>TOTAL</b>	5535	
Design Effect	1.378	

- The confidence intervals give you the range of values where the prevalence can fall based on the existing variables and conditions. For this example, the lifetime prevalence is 11.7; however, the confidence intervals tell you it could be as low as 10.7 or as high as 12.7.
- If you wished to determine if the prevalence is increasing or decreasing, you would need to download several years of data from the BRFSS website and complete similar steps to determine the variable names and values. Using the WRITE command, you could create a table to hold data for each year in one location.

 Practice

**B.** What is the current asthma prevalence among adults in the state?

Compute the current adult asthma prevalence, using the same series of steps for the A\_CASTHMA variable.

By consulting the BRFSS website, you know that the calculated variable that contains information on the current prevalence of asthma is called \_CASTHMA. Since you imported the file as a .DBF, the variable name contains a prefix of A to compensate for the space field. The variable you need to figure for prevalence is called A\_CASTHMA. This is a calculated variable based on the BRFSS question, "Do you still have asthma?"

First, exclude the missing variables using the If/Then statement.

```
IF A_CASTHMA=9 THEN
ASSIGN A_CASTHMA= (.)
END
```

Click **Complex Sample Frequencies**.

Create a Frequency of A\_CASTHMA using the Weight of A\_FINALWT and the PSU of A\_PSU.

Results appear in the Output Window:

INTERMEDIATE ANALYSIS

A_CASTHMA	TOTAL
<b>1</b>	5079
Row %	100.000
Col %	92.362
SE %	0.413
LCL %	91.553
UCL %	93.170
<b>2</b>	439
Row %	100.000
Col %	7.638
SE %	0.413
LCL %	6.830
UCL %	8.447
<b>TOTAL</b>	5518
Design Effect	1.331

Prevalence

Confidence Intervals

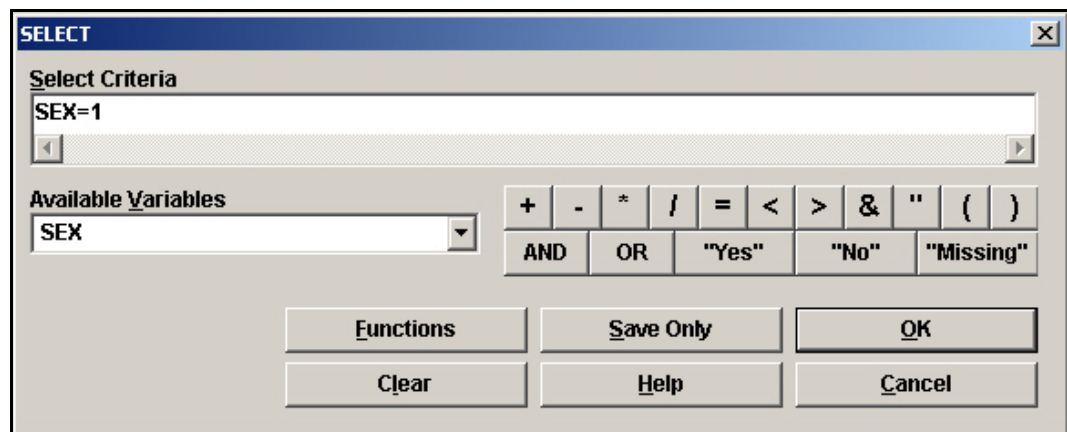
## Selecting Subgroups

To strengthen your proposal, you would like to compare asthma prevalence rates among subgroups in the state and see if there are differences.

C. Is the state adult asthma prevalence rate higher in the subgroup gender?

Subgroups available in the BRFSS data include age, sex, ethnicity, and race. Gender in the BRFSS dataset is coded as the number 1 for males, and the number 2 for females. To compute the prevalence for each group, you will use the SELECT command.

1. From the Command Tree Select/If folder, click **Select**. The SELECT dialog box opens.
2. From the Available Variables drop-down, select Sex.
3. Click =.
4. Type 1.



5. Click **OK**. There are 2,172 males in the survey.
6. Click **Complex Sample Frequencies**. The Complex Sample Frequencies dialog box opens.
7. Select A\_LTASTHM from the Frequency of drop-down box.
8. Select a Weight of A\_FinalWT.
9. Select a PSU of A\_PSU.
10. Click **OK**.
11. Click **Cancel Select**.

 **Practice**

Compute the prevalence for the subgroup women.

Select Sex=2. You should have 3,372 records.

Click **Complex Sample Frequencies**. Use the BRFSS Weight and PSU variables to compute the prevalence of the variable A\_LTASTHM. Record the results on your answer sheet.

Click **Cancel Select**.



### Complex Sample Tables

To strengthen your proposal, you would like to compare asthma prevalence rates among subgroups in the state and see if there are differences.

D. Is the state adult asthma prevalence rate higher in the subgroup ethnicity?

Use the BRFSS data to compute the lifetime prevalence rate for the subgroup ethnicity. You will use the Complex Sample Tables command to compute for prevalence and the odds ratio.

In the BRFSS data table, Ethnicity is categorized in the variable HISPANC2 with Yes answers coded as 1 and No answers coded as 2. The lifetime prevalence of asthma is categorized in the variable A\_LTASTHM with No answers coded as 1 and Yes answers coded as 2. To compute with accuracy using Complex Sample Tables the first value in order must represent YES and the second value in order must represent NO. You will need to do some data management to set up your tables in the proper sequence.

Your current data table looks like this:

A_LTASTHM	1= NO 2= YES
HISPANC2	1=YES 2=NO

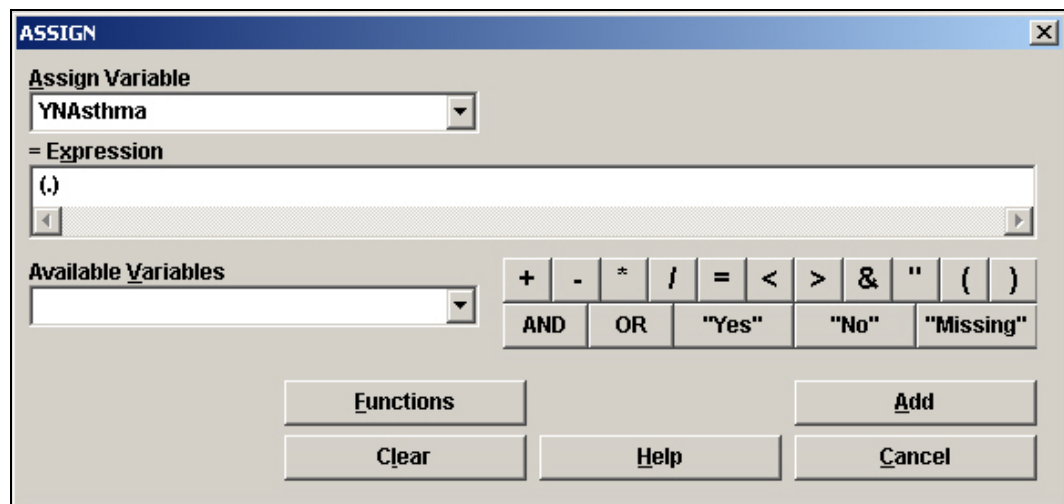
You want it to look like this:

A_LTASTHM	1= YES 2= NO
HISPANC2	1=YES 2=NO

In order to get the YES/NO values to match in the proper order, you will create an IF /THEN to statement to ASSIGN newly defined variables. The IF statement will remove the Missing variables and reverse the order of the coded 1s and 2s using a mathematical statement.

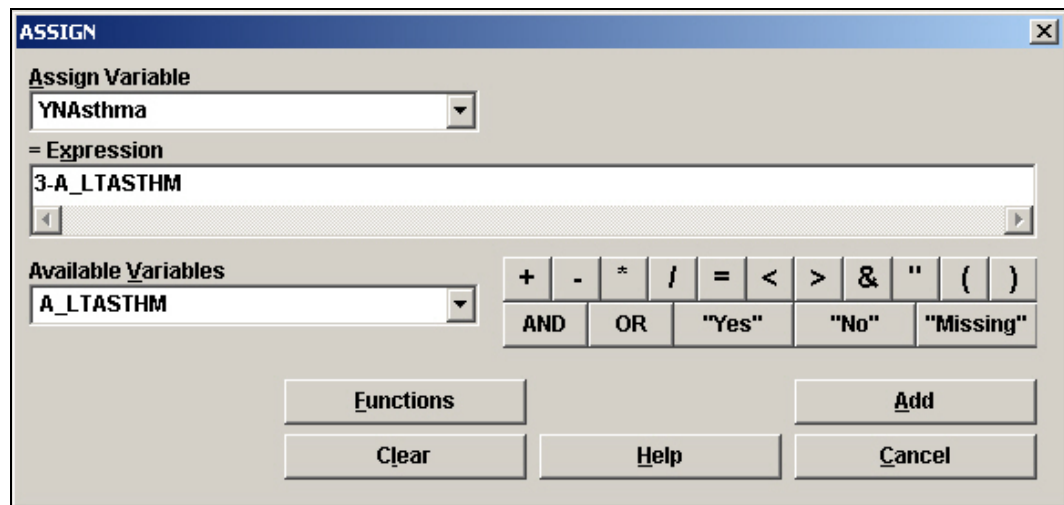
**DEFINE** a new lifetime asthma prevalence variable and **ASSIGN** it using an **IF/THEN** statement.

1. Read in the data table BRFSS2003.
2. From the Command Tree Variables folder, click **Define**. The DEFINE dialog box opens.
3. In the Variable Field Name field, type *YNAsthma*.
4. Click **OK**.
5. Click **IF**. The IF dialog box opens.
6. From the Available Variables drop-down, select A\_LTASTHM.
7. Click the greater than > operator.
8. Type in the number 2.
  - The IF Condition dialog box should read A\_LTASTHM>2.
9. Click **THEN**. The Command Tree highlights.
10. Click **Assign**. The ASSIGN dialog box opens.
11. From the Assign Variable drop-down, select YNAsthma.
12. Move the cursor to the =Expression field and click **Missing**.

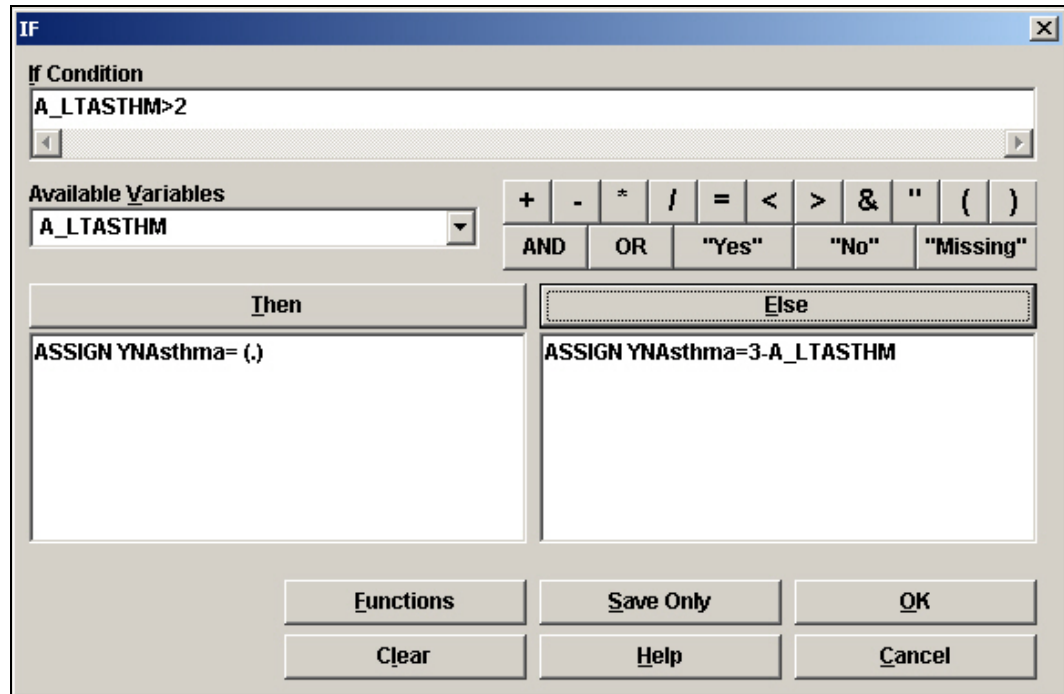


13. Click **Add**. The IF dialog box opens.
  - You have removed the values of 7 and 9 by stating that **IF** the value of A\_LTASTHM is greater than 2, **THEN** the values are assigned as Missing.

14. Click **ELSE**. The Command Tree highlights.
15. Click **Assign**. The ASSIGN dialog box opens.
16. From the Available Variable drop-down, select YNAsthma.
17. Move your cursor to the =Expression field and type, *3-A\_LTASTHM*.
  - You are assigning the variable a mathematical statement that says  $YNAsthma=3-A\_LTASTHM$ , in other words, the variable YNAsthma=3-2 or 3-1. This reformats the Yes/No values into the proper order with 1 now being Yes and 2 being No and also places them in the proper sequence with the Yes value coming first in the series.



18. Click **Add**. The IF dialog box opens.



19. Click **OK**. Now you need to define and assign values for a new HISPANC2 variable.

**DEFINE** a new Ethnicity variable and **ASSIGN** it using an **IF/THEN** statement.

1. Click **Define** and create a new variable called *Hisp*.
2. Click **OK**.
3. Click **IF**. The IF dialog box opens.
4. From the Available Variables drop-down, select HISPANC2.
5. In the IF Condition field, type >2.
6. Click **THEN**. The Command Tree highlights.
7. Click **Assign**. The ASSIGN dialog box opens.
8. From the Assign Variable drop-down, select Hisp.
9. Move your cursor to the =Expression field and click **Missing**.
10. Click **Add**. The IF dialog box opens.
11. Click **ELSE**. The Command Tree highlights.

12. Click **Assign**. The ASSIGN dialog box opens.
13. From the Assign Variable drop-down, select Hisp.
14. Move the cursor to the =Expression field and type HISPANC2.
  - Since the coded values for the variable Yes and No are in the correct order inside the variable HISPANC2, you do not have to create the mathematical functions that changes their values.
15. Click **Add**. The IF dialog box opens.
16. Click **OK**.

**View the results of your IF/THEN statements and create a Complex Sample Table.**

1. From the Command Tree Statistics folder, click **Frequencies**. The FREQ dialog box opens.
2. From the Frequency Of drop-down, select Hisp, HISPANC2, YNAsthma, and A\_LTASTHM.
3. Click **OK**. View the results and how the IF/THEN statements you created altered the results and the table structure.
4. From the Command Tree Advanced Statistics folder, click **Complex Sample Tables**. The TABLES dialog box opens.
5. From the Exposure Variable drop-down, select Hisp.
6. From the Outcome Variable drop-down, select YNAsthma.
7. From the Weight drop-down select A\_FINALWT.
8. From the PSU drop-down, select A\_PSU.
9. Click **OK**. View the results.

Hispanic	YNAsthma		
	1	2	TOTAL
<b>1</b>	82	494	576
Row %	11.197	88.803	
Col %	15.674	16.126	
SE %	1.393	1.393	
LCL %	8.468	86.073	
UCL %	13.927	91.532	
Design Effect	1.124	1.124	
<b>2</b>	576	4838	5496
Row %	11.852	88.148	
Col %	84.326	83.459	83.561
SE %	0.545	0.545	
LCL %	10.784	87.079	
UCL %	12.921	89.216	
Design Effect	1.399	1.399	
<b>TOTAL</b>	658	4838	5496
Row %	11.745	88.255	
Col %	100.000	100.000	
SE %	0.510	0.510	
LCL %	10.745	87.256	
UCL %	12.744	89.255	
Design Effect	1.379	1.379	

The lifetime prevalence of asthma among Hispanics.

The lifetime prevalence of asthma among non-Hispanics.

The lifetime prevalence of asthma for the state population.

Scroll down to view the Odds and Risk Ratios and the Standard Error and Confidence Limits.

**CTABLES COMPLEX SAMPLE DESIGN ANALYSIS OF 2 X 2 TABLE**

Odds Ratio (OR) 0.938  
 Standard Error (SE) 0.147  
 95% Conf. Limits (0.70, 1.257 )

Risk Ratio (RR) 0.945  
 Standard Error (SE) 0.125  
 95% Conf. Limits (0.73, 1.225 )

$RR = (\text{Risk of YNAsthma}=1 \text{ if Hisp}=1) / (\text{Risk of YNAsthma}=1 \text{ if Hisp}=2)$

Risk Difference (RD%) - 0.655  
 Standard Error (SE) 1.496  
 95% Conf. Limits (-3.59, 2.276 )

$RD = (\text{Risk of YNAsthma}=1 \text{ if Hisp}=1) - (\text{Risk of YNAsthma}=1 \text{ if Hisp}=2)$

Notice that your tables are set up in the proper order and sequence with Yes to Asthma and Yes to Hispanic over Yes to Asthma and No to Hispanic.

You could use county data from the Expanded BRFSS to answer other questions in the same manner. To answer questions and comparisons related to county or EBRFSS data you would follow the same process as above. You could use this process to answer these types of questions:

- Is the county adult asthma prevalence higher or lower than the state?
- Is the county adult asthma prevalence higher in some population subgroups?
- How do county population subgroup patterns relate to the state sub-group patterns for asthma prevalence?

## Create a Frequency

You can use the 800 records used as your primary data source to find prevalence rates for the individual schools in the survey. For your proposal, this data represents the child population in the county.

**E.** From the school asthma survey data, what is the difference in prevalence rates?

To find the difference in prevalence rates between the four schools in the survey, you will create a frequency of asthma and stratify by School Number.

1. Click **Read/Import**. The READ dialog box opens.
2. From the project Asthma Final, select viewPreInterventionSurvey.
3. Click **OK**. You should have the 800 records from the survey.
4. Click **Frequencies**. The FREQ dialog box opens.
5. From the Frequency Of drop-down menu, select Asthma.
6. From the Stratify By drop-down menu, select SchoolNum.
7. Click **OK**. The Output window populates with one table for each school.



## Sampling Scheme

The dataset Asthma Final.MDB contains the 800 records you worked with throughout the Basic tutorial. You are going to use this data to represent all the schools for the county and all the children surveyed for the county in order to compute the county childhood asthma prevalence.

To do this, you need to think in terms of the sampling scheme of the schools in relation to the county. For this example, you will be following the sampling scheme below, and use it to make your assumptions about the results you find, and how your data is selected and organized:

- All four elementary schools were chosen.
- A number of classrooms for each school were chosen at random until the number of students in the chosen classes was about 400 at each school.
- Half of the kids from each of the classrooms were chosen at random to be surveyed.
- The cluster is the classroom within each school; however, classroom information was not recorded, so age was used as a proxy for classroom.
- The Primary Sampling Unit (PSU) is SchoolNum and Age (Use cut-off for school year September 1, 2005).

You need to create a Weight for the schools to the county population by age and sex in order to get a county estimate of prevalence and be able to compute for a subgroup. In the following lesson, you will create a table that contains the sex and age data for the 800 records.

You will use the RELATE command to join the sex and age table created from the 800 records to the population table you are going to create, based on data from the United States Census. The Weight Variable will be the county population divided by the survey population first for males, and then for females. In this data, there is a number in the general population, which each respondent of a particular age and sex represents, so you will compute a Complex Sample Frequency where the PSU is SchoolNum and Age (this is a grouping field, not a numeric field).

From this data, you will get a county estimate of prevalence by sex for the selected age group (based on the survey age groups).

In the following sections you will use your data management skills to

- Import and format HTML data
- Create a numerator table using the HTML data
- Relate tables
- Create a Weight Variable

- Create a PSU Variable
- Compute the prevalence

The ultimate goal is to obtain the asthma prevalence for children in the county. The formula used will be the number of cases of asthma in the survey divided by the population of the age group for the county (from the HTML Census data).

## Import Census Data/HTML Format

You want to include some information on childhood asthma in your proposal. Use the 800 records from your primary data source and data from the Census Bureau to compute the childhood asthma prevalence rate for the county.

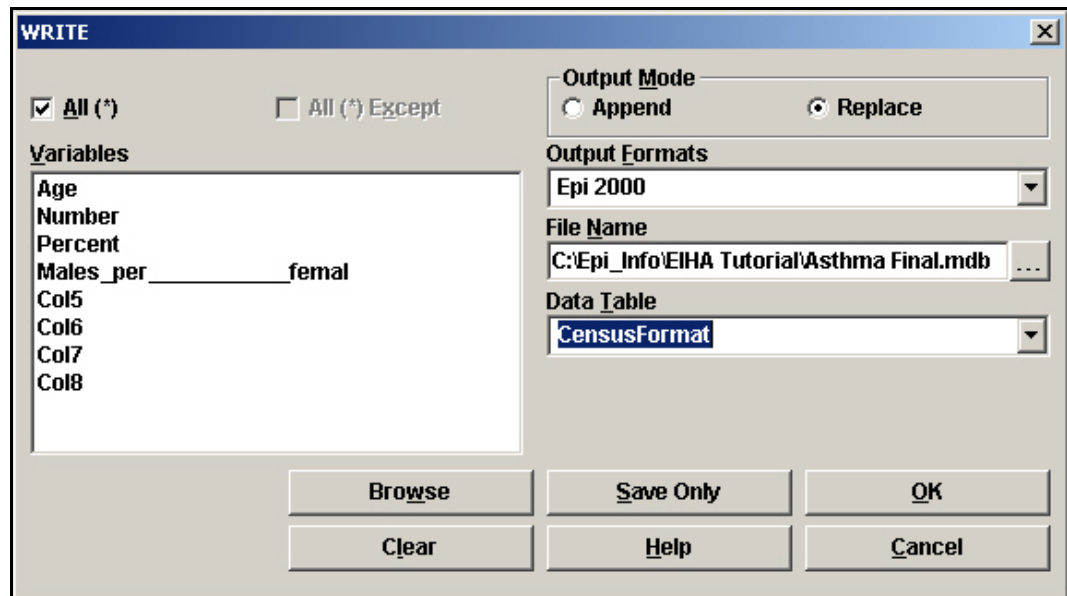
- F. What is the county childhood asthma prevalence?
- G. What is the county childhood asthma prevalence for children in the subgroup Gender?

To compute the prevalence rate, you need population data for the age group in the survey. Population data for this example was downloaded from the Census website. [www.census.gov](http://www.census.gov). Data was downloaded for Albany County by single years for age and sex.

For this lesson, you will be importing an .HTML file that was downloaded from the United States Census Bureau website. You will then format the data to use as your county child population. The prevalence will be determined by dividing the number of cases of asthma in the survey by the population of the age group for the county.

1. Click **Read/Import**. The READ dialog box opens.
  - Be sure your current project is Asthma Final. If it is not, click **Change Project** and select Asthma Final.
2. From the Data Format drop-down, select HTML.
3. Click the **Browse** button in the Data Source field. The READ window opens.
4. Locate and select the HTML file called censusage.htm.
5. Click **Open**.
6. Select TABLE 284:1. The HTML window appears so you can see how the information looks.
7. Close the HTML window.
8. Click **OK**. The READ Temp Link dialog box opens.
9. Type the link name *CensusLink*.
10. Click **OK**. You should have 42 records.
11. From the Command Tree Variables folder, click **Display**. Notice that numeric fields were imported as text and that the variable names reverted to the column number.
12. Click **List** to view the data. Notice that there are commas in the records.

- You will need to reformat the fields containing commas, use the text to number format command, and define/assign new variable names to use this data.
13. From the Command Tree Data folder, click **Write/Export**. The WRITE dialog box opens.
  14. From the Output Mode section, click the Replace radio button.
  15. Click the **Browse** button from the File Name field.
  16. Locate and select the Asthma Final project.
  17. Click **Save**.
  18. In the Data Table field, type *CensusFormat*.

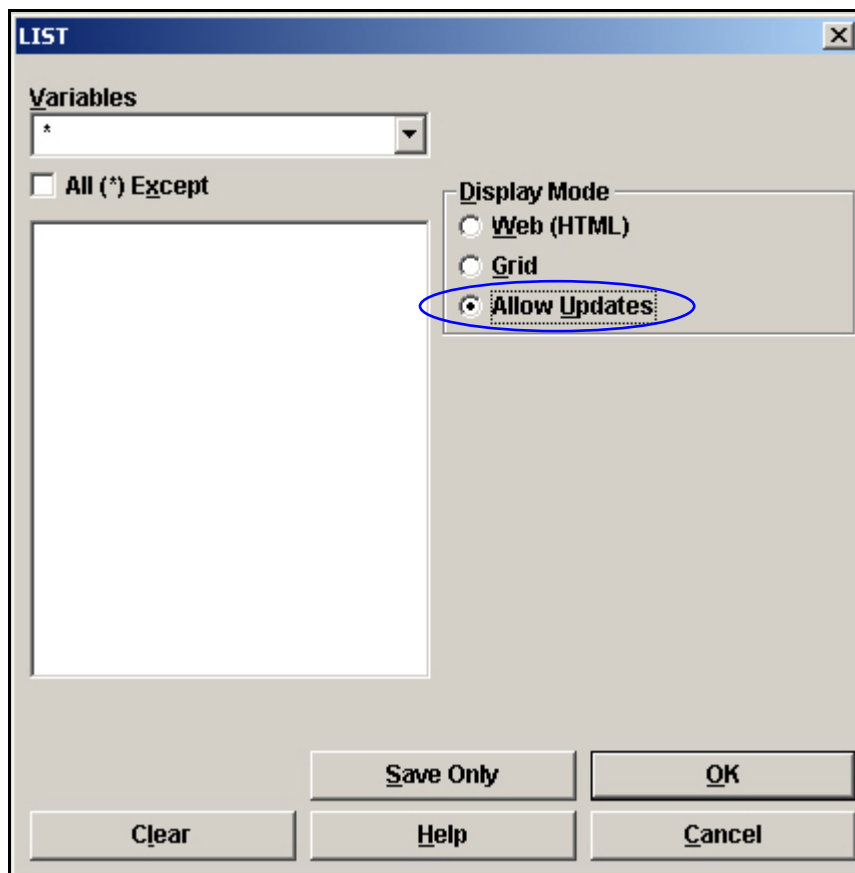


19. Click **OK**.
- You now have a new data table called CensusFormat that contains all the information imported from the Census HTML file.

## Formatting with List Update

The following formatting instructions can be used when working with any files imported into Analysis, including HTML or MS Excel files.

1. Click **Read/Import**. The READ dialog box opens.
2. Click the **Show All** radio button.
3. Select the CensusFormat table you created in the previous lesson.
4. Click **OK**. You should have the same 42 records.
5. Click **List**. The LIST dialog box opens.
6. From the Display Mode section, select the Allow Updates radio button.



7. Click **OK**. A grid table appears with all your variables listed.
  - From this grid table, you can make changes to the data in your records. Keep in mind that changes made in this mode are saved and cannot be undone.

- For this example, you are interested in the 5-12 year age groups that correspond to the survey ages. You are going to alter only those records, then select them, and place them in a new table.

Analysis Output						
Age	Number	Percent	Males_per	fema	Col5	Col6
Both sexes	Male	Female	Both sexes	Male	Female	
Total population	294,565	140,885	153,680	100.0	100.0	
Under 5 years	16,649	8,610	8,039	5.7	6.1	
Under 1 year	3,192	1,661	1,531	1.1	1.2	
1 year	3,311	1,738	1,573	1.1	1.2	
2 years	3,303	1,696	1,607	1.1	1.2	
3 years	3,392	1,727	1,665	1.2	1.2	
4 years	3,451	1,788	1,663	1.2	1.3	
5 to 9 years	18,944	9,768	9,176	6.4	6.9	
5 years	3,537	1,821	1,716	1.2	1.3	
6 years	3,638	1,838	1,800	1.2	1.3	
7 years	3,777	1,930	1,847	1.3	1.4	
8 years	3,948	2,076	1,872	1.3	1.5	
9 years	4,044	2,103	1,941	1.4	1.5	
10 to 14 years	19,432	9,967	9,465	6.6	7.1	
10 years	3,958	2,051	1,907	1.3	1.5	
11 years	3,825	1,964	1,861	1.3	1.4	
12 years	3,929	2,033	1,896	1.3	1.4	
13 years	3,837	1,937	1,900	1.3	1.4	
14 years	3,883	1,982	1,901	1.3	1.4	
15 to 19 years	22,227	11,149	11,078	7.5	7.9	
15 years	3,893	1,979	1,914	1.3	1.4	
16 years	3,836	2,011	1,825	1.3	1.4	
17 years	3,723	1,893	1,830	1.3	1.3	
18 years	5,116	2,506	2,610	1.7	1.8	
19 years	5,659	2,760	2,899	1.9	2.0	
20 to 24 years	22,609	11,123	11,486	7.7	7.9	
20 years	5,460	2,710	2,750	1.9	1.9	
21 years	5,146	2,547	2,599	1.7	1.8	

- Place your cursor in the Age column on the record for 5 years.
- Highlight the text and type the number 5.
- Repeat this process for the years 6, 7, 8, 9, 10, 11, and 12. Do not alter the total fields.

## INTERMEDIATE ANALYSIS

Analysis Output						
Age	Number	Percent	Males per	fema	Col5	Col6
Both sexes	Male	Female	Both sexes	Male	Female	
Total population	294,565	140,885	153,680	100.0	100.0	
Under 5 years	16,649	8,610	8,039	5.7	6.1	
Under 1 year	3,192	1,661	1,531	1.1	1.2	
1 year	3,311	1,738	1,573	1.1	1.2	
2 years	3,303	1,696	1,607	1.1	1.2	
3 years	3,392	1,727	1,665	1.2	1.2	
4 years	3,451	1,788	1,663	1.2	1.3	
5 to 9 years	18,944	9,768	9,176	6.4	6.9	
5	3,537	1,821	1,716	1.2	1.3	
6	3,638	1,838	1,800	1.2	1.3	
7	3,777	1,930	1,847	1.3	1.4	
8	3,948	2,076	1,872	1.3	1.5	
9	4,044	2,103	1,941	1.4	1.5	
10 to 14 years	19,432	9,967	9,465	6.6	7.1	
10	3,958	2,051	1,907	1.3	1.5	
11	3,825	1,964	1,861	1.3	1.4	
12	3,929	2,033	1,896	1.3	1.4	
13 years	3,837	1,937	1,900	1.3	1.4	
14 years	3,883	1,982	1,901	1.3	1.4	
15 to 19 years	22,227	11,149	11,078	7.5	7.9	
15 years	3,893	1,979	1,914	1.3	1.4	
16 years	3,836	2,011	1,825	1.3	1.4	
17 years	3,723	1,893	1,830	1.3	1.3	
18 years	5,116	2,506	2,610	1.7	1.8	
19 years	5,659	2,760	2,899	1.9	2.0	
20 to 24 years	22,609	11,123	11,486	7.7	7.9	
20 years	5,460	2,710	2,750	1.9	1.9	
21 years	5,146	2,547	2,599	1.7	1.8	

11. Place your cursor in the Number column next to the Age field 5.
12. Retype the number in the field without the commas. For example, replace the number 3,537 with 3537.
13. Do this for each of the age fields 5-12 that you updated in the previous steps.
14. Repeat this step for the age fields in the Percent column.
15. Repeat this step for the age fields in the Males per female column.

## INTERMEDIATE ANALYSIS

Analysis Output						
Age	Number	Percent	Males per	fema	Col5	Col6
Both sexes	Male	Female	Both sexes	Male	Female	
Total population	294,565	140,885	153,680	100.0	100.0	
Under 5 years	16,649	8,610	8,039	5.7	6.1	
Under 1 year	3,192	1,661	1,531	1.1	1.2	
1 year	3,311	1,738	1,573	1.1	1.2	
2 years	3,303	1,696	1,607	1.1	1.2	
3 years	3,392	1,727	1,665	1.2	1.2	
4 years	3,451	1,788	1,663	1.2	1.3	
5 to 9 years	18,944	9,768	9,176	6.4	6.9	
5	3537	1821	1716	1.2	1.3	
6	3638	1838	1800	1.2	1.3	
7	3777	1930	1847	1.3	1.4	
8	3948	2076	1872	1.3	1.5	
9	4044	2103	1941	1.4	1.5	
10 to 14 years	19,432	9,967	9,465	6.6	7.1	
10	3958	2051	1907	1.3	1.5	
11	3825	1964	1861	1.3	1.4	
12	3929	2033	1896	1.3	1.4	
13 years	3,837	1,937	1,900	1.3	1.4	
14 years	3,883	1,982	1,901	1.3	1.4	
15 to 19 years	22,227	11,149	11,078	7.5	7.9	
15 years	3,893	1,979	1,914	1.3	1.4	
16 years	3,836	2,011	1,825	1.3	1.4	
17 years	3,723	1,893	1,830	1.3	1.3	
18 years	5,116	2,506	2,610	1.7	1.8	
19 years	5,659	2,760	2,899	1.9	2.0	
20 to 24 years	22,609	11,123	11,486	7.7	7.9	
20 years	5,460	2,710	2,750	1.9	1.9	
21 years	5,146	2,547	2,599	1.7	1.8	

16. Click the Close **X** to exit the List Update window and return to the Output window.

- Do not close Analysis.



## Selecting with String Length

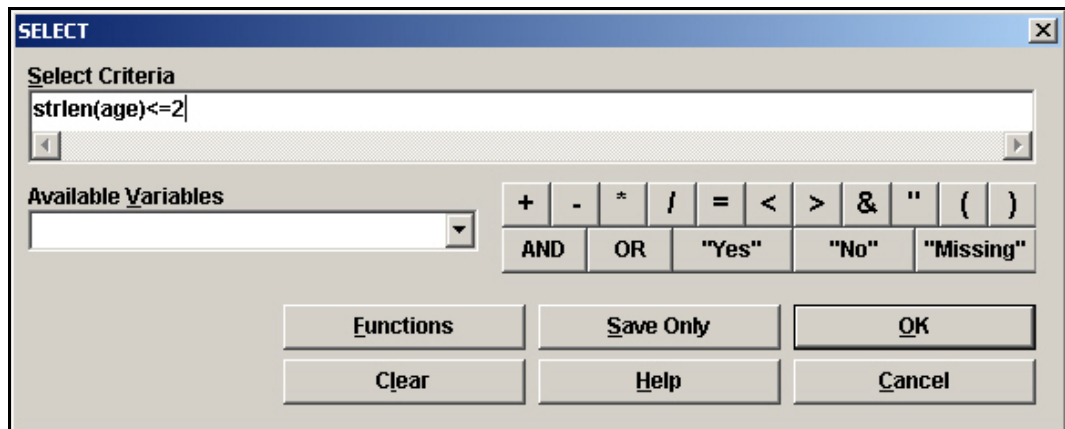
Now you have a set of Age variables that differ from the series. You can use this difference to select only those age groups you have edited.

You are going to use a format option called String Length to specify the length of the variable you are selecting.

1. Click **Select**. The SELECT dialog box opens.
2. In the Select Criteria field, use your keyboard and the Operator buttons to create the following code:

```
strlen(age)<=2
```

- This states that you want to see only the records from the Age field that have a length less than or equal to 2 characters.



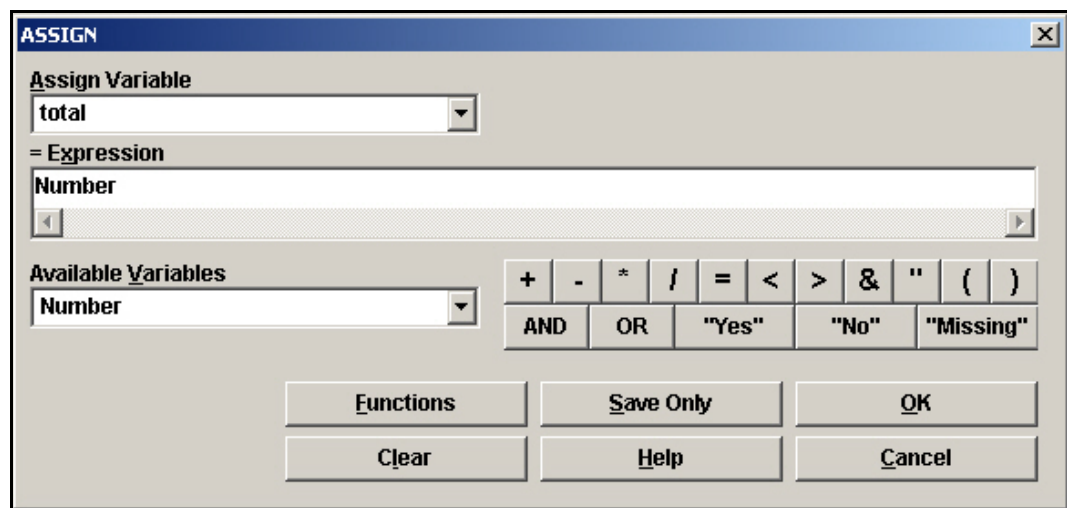
3. Click **OK**. You should have 8 records.
4. Click **List**. The selected list appears.

Age	Number	Percent	Males_per	Col5	Col6	Col7	Col8
5	3537	1821	1716	1.2	1.3	1.1	106.1
6	3638	1838	1800	1.2	1.3	1.2	102.1
7	3777	1930	1847	1.3	1.4	1.2	104.5
8	3948	2076	1872	1.3	1.5	1.2	110.9
9	4044	2103	1941	1.4	1.5	1.3	108.3
10	3958	2051	1907	1.3	1.5	1.2	107.6
11	3825	1964	1861	1.3	1.4	1.2	105.5
12	3929	2033	1896	1.3	1.4	1.2	107.2

## Define and Assign New Variables

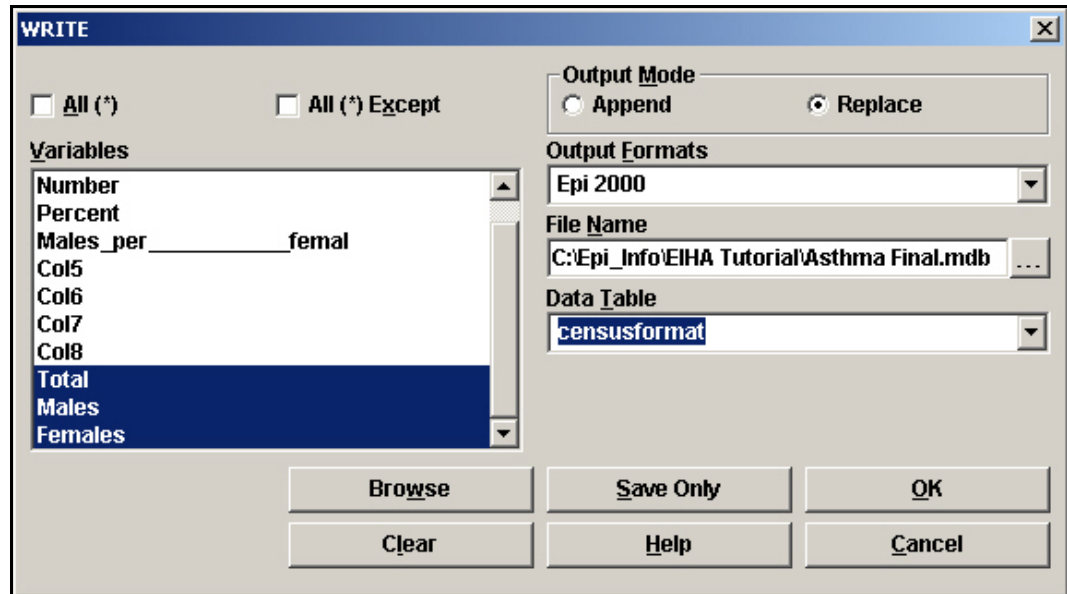
The headings brought in from the .HTML file do not match up with the data in the columns. This can happen with .HTML or MS Excel files. The column titled Number is the Total. The column titled Percent is the number of Males, and the column titled Males per Female is the number of Females. You are going to use the DEFINE command to create new variables, and use the ASSIGN command to give them current table values.

1. Click **Define** and create a new variable called *Total*.
2. Click **OK**.
3. Click **Define** and create a new variable called *Males*.
4. Click **Define** and create a new variable called *Females*.
5. Click **Assign**. The ASSIGN dialog box opens.
6. From the Assign Variable drop-down, select Total.
7. Move your cursor to the =Expression field.
8. From the Available Variables drop-down, select Number.



9. Click **OK**.
10. Use the ASSIGN command to assign the variable Males=Percent.
11. Use the ASSIGN command to assign the variable Females=Males\_per\_\_\_\_\_femal.
12. Click **List** to view the records and verify that the values were copied to the new variables.
13. Click **Write/Export**. The WRITE dialog box opens. You need to write only the new variables to a new table for use in computing the prevalence.

14. From the Variables list, select Age, Total, Males, and Females.
15. From the Output Mode, select the Replace radio button.
16. From the File Name field, click the **Browse** button.
17. Locate and select the Asthma Final.MDB.
18. From the Data Table drop-down, select CensusFormat.



19. Click **OK**.
20. Click **Read/Import**.
21. Click the **Show All** radio button.
22. Select the CensusFormat table.
23. Click **OK**. You should have eight records.
24. Click **List** to see the records.
  - You have created a table with the total population per age for the county. This table will be used as your denominator table when you compute the rates for the county.

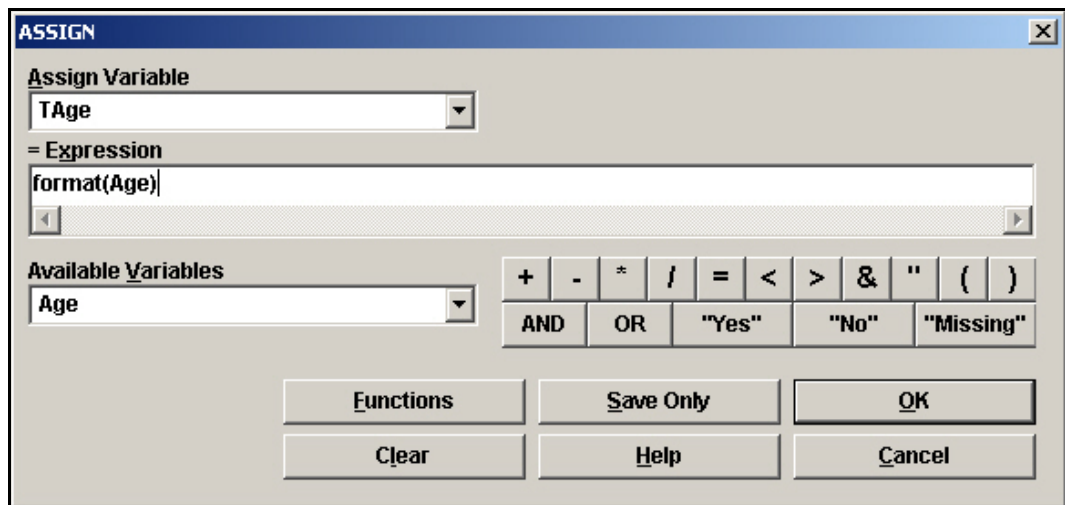
## INTERMEDIATE ANALYSIS

<b>Age</b>	<b>Total</b>	<b>Males</b>	<b>Females</b>
5	3537	1821	1716
6	3638	1838	1800
7	3777	1930	1847
8	3948	2076	1872
9	4044	2103	1941
10	3958	2051	1907
11	3825	1964	1861
12	3929	2033	1896

## Create a Numerator Table

The CensusFormat table is now your denominator table. You need to create the numerator table that contains information about the survey respondents age and sex, and then use the RELATE command to join the two tables.

1. Click **Read/Import**.
2. Open the view preInterventionSurvey. You should have 800 records.
3. Click **Assign**. The ASSIGN dialog box opens.
4. Create the following code `Age=Years(DOB,10/25/2005)`.
5. Click **OK**.
6. **List** the Age variable to make sure it calculated properly.
7. Click **Define** and create a new variable called *TAge*.
8. Click **Assign**. The ASSIGN dialog box opens.
9. Create the following code `TAge=format(age)`. Age is a number and TAge is Text.



10. From the Command Tree Statistics folder, click **Tables**. The TABLES dialog box opens.
11. From the Exposure Variable drop-down, select TAge.
12. From the Outcome Variable drop-down, select Gender.
13. In the Output to Table field, type *SurveyAgeSex*.

**TABLES**

Outcome Variable: Gender

Stratify by: [ ]

Exposure Variable: tage

Weight: [ ]

Output to Table: surveyagesex

Matched Analysis

Optional Page Settings

Columns per Page

No Line Wrap

Settings Save Only OK

Clear Help Cancel

HEIGHT	+	-
62.0	20	40
62.5	37	59

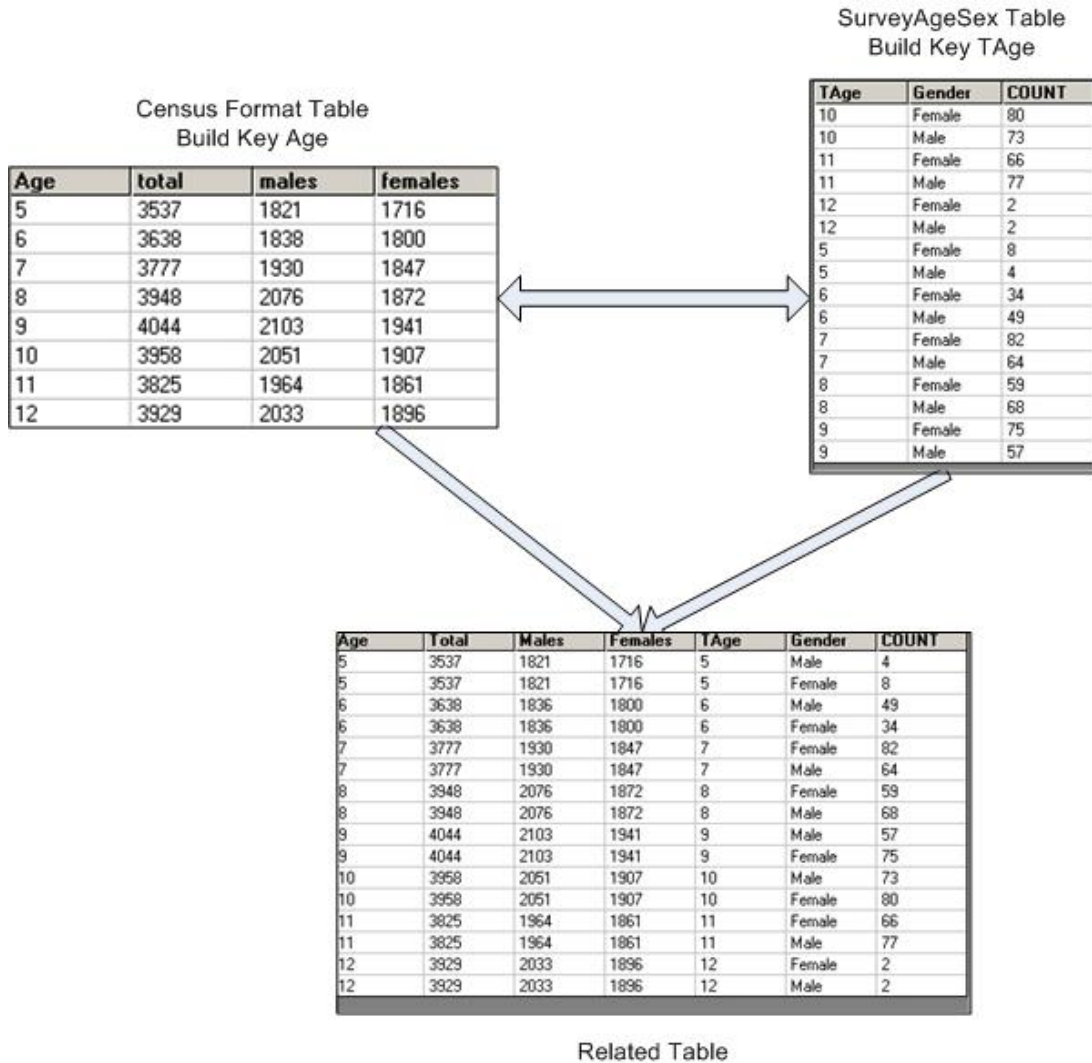
14. Click **OK**. The Output window shows the results. Now you have the number of males and the number of females per age.

<b>GENDER</b>			
<b>TAge</b>	<b>Female</b>	<b>Male</b>	<b>TOTAL</b>
<b>10</b>	80	73	153
Row %	52.3	47.7	100.0
Col %	19.7	18.5	19.1
<b>11</b>	66	77	143
Row %	46.2	53.8	100.0
Col %	16.3	19.5	17.9
<b>12</b>	2	2	4
Row %	50.0	50.0	100.0
Col %	0.5	0.5	0.5
<b>5</b>	8	4	12
Row %	66.7	33.3	100.0
Col %	2.0	1.0	1.5
<b>6</b>	34	49	83
Row %	41.0	59.0	100.0
Col %	8.4	12.4	10.4
<b>7</b>	82	64	146
Row %	56.2	43.8	100.0
Col %	20.2	16.2	18.3
<b>8</b>	59	68	127
Row %	46.5	53.5	100.0
Col %	14.5	17.3	15.9
<b>9</b>	75	57	132
Row %	56.8	43.2	100.0
Col %	18.5	14.5	16.5
<b>TOTAL</b>	406	394	800
Row %	50.8	49.3	100.0
Col %	100.0	100.0	100.0

Create a Relate

Now you are going to use the RELATE command to connect the two tables for the calculations.

To create a Weight variable, you need to relate the census population numbers to the survey count numbers, then you can divide the total numbers by the survey count to get the weight. By relating the CensusFormat table and the SurveyAgeSex table, you can define and format new variables and then WRITE a new table containing data from the two original sources.

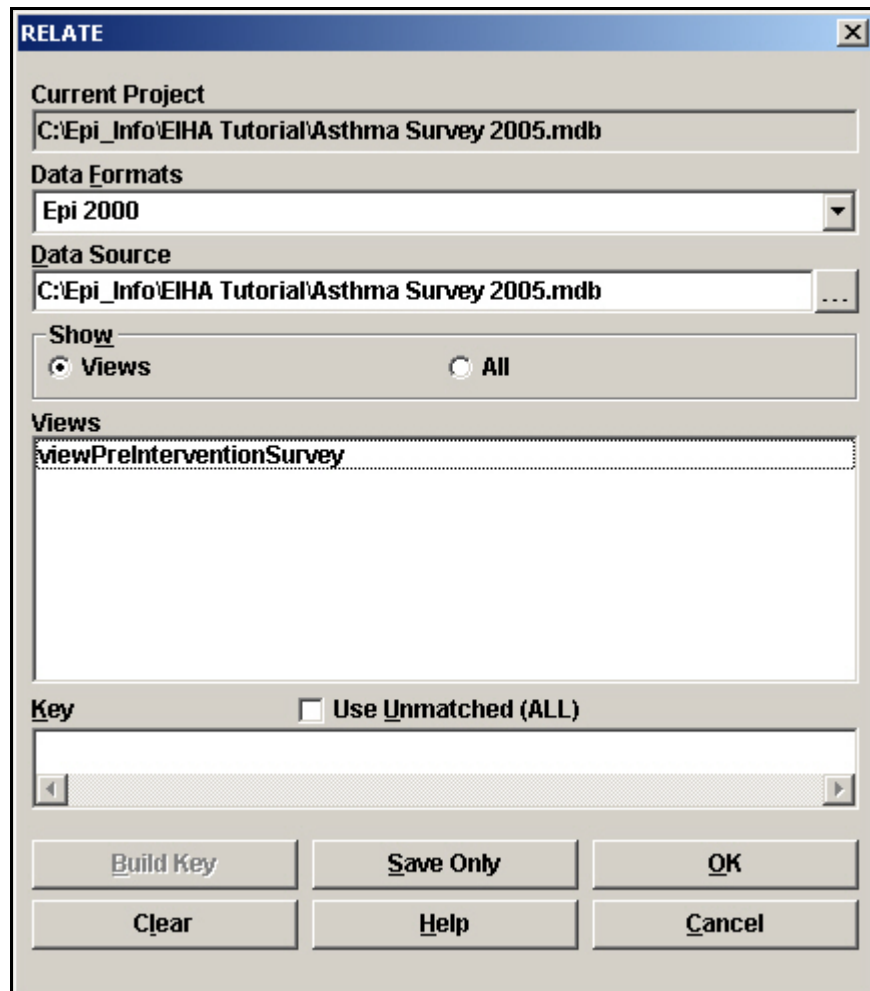


Data in the new related table can be used to create the new Weight Variables. You can use the WRITE command to create a new permanent table that holds these values.

1. Click **Read/Import**. The READ dialog box opens.
2. Click the **Show All** radio button.



3. Select the table called CensusFormat.
4. Click **OK**. You should have 8 records.
5. From the Command Tree Data folder, click **Relate**. The RELATE dialog box opens.



6. Click the **Show All** radio button.
7. Select the SurveyAgeSex table.
8. Click **Build Key**. The RELATE-BUILD KEY dialog box opens.
  - Both the tables contain variables for age so you can relate the tables on that variable.

**RELATE - BUILD KEY** [X]

**Key Component**

Available Variables

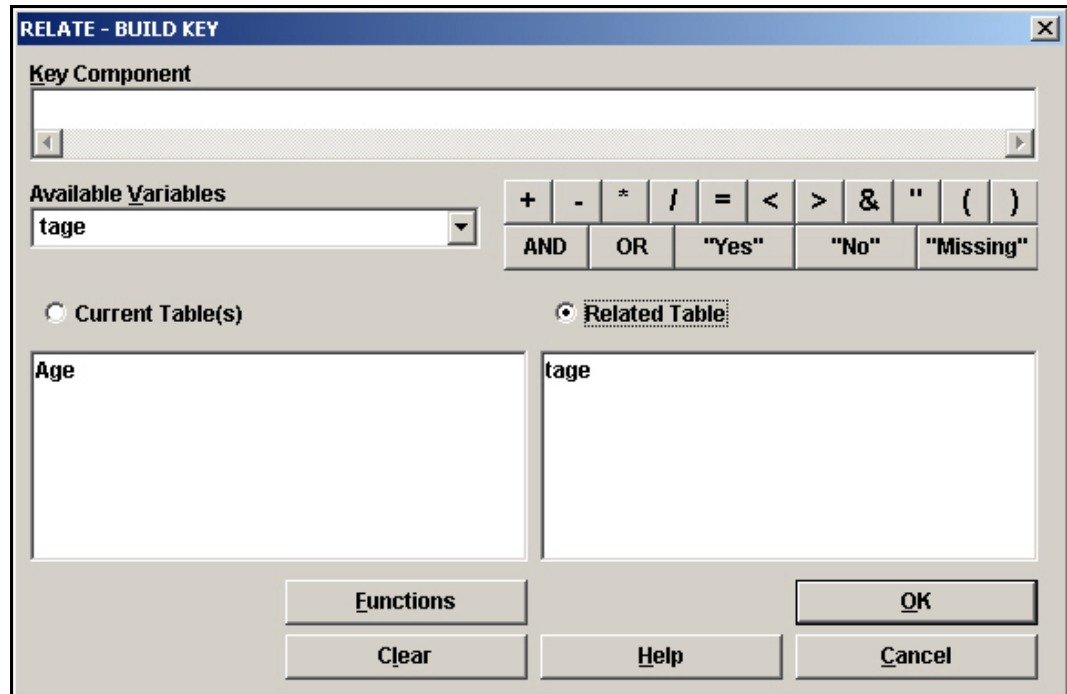
+ - \* / = < > & " ( )

AND OR "Yes" "No" "Missing"

Current Table(s)  Related Table

Functions Clear Help OK Cancel

9. Select the Current Table radio button.
10. From the Available Variables drop-down, select Age.
11. Select the Related Table radio button.
12. From the Available Variables drop-down, select TAge.
13. Click **OK**. The Related Tables field populates.



12. Click **OK**. The RELATE dialog box opens.
13. Click **OK**. You should have 16 records.
14. Click **Display**. Notice that the fields TAge and Age are Text.
15. Click **Define** and create a new variable called *NAge*. This new variable will hold the age value in a numeric form.
16. Click **Assign**. You are going to format the new age field as a number and assign a value to the variable.
17. From the Assign Variable drop-down, select *NAge*.
18. In the =Expression field type, *TXTTONUM(Age)*.

**ASSIGN** [X]

Assign Variable  
nage

= Expression  
{xttonum(age)}

Available Variables

+ | - | \* | / | = | < | > | & | " | ( | )  
AND | OR | "Yes" | "No" | "Missing"

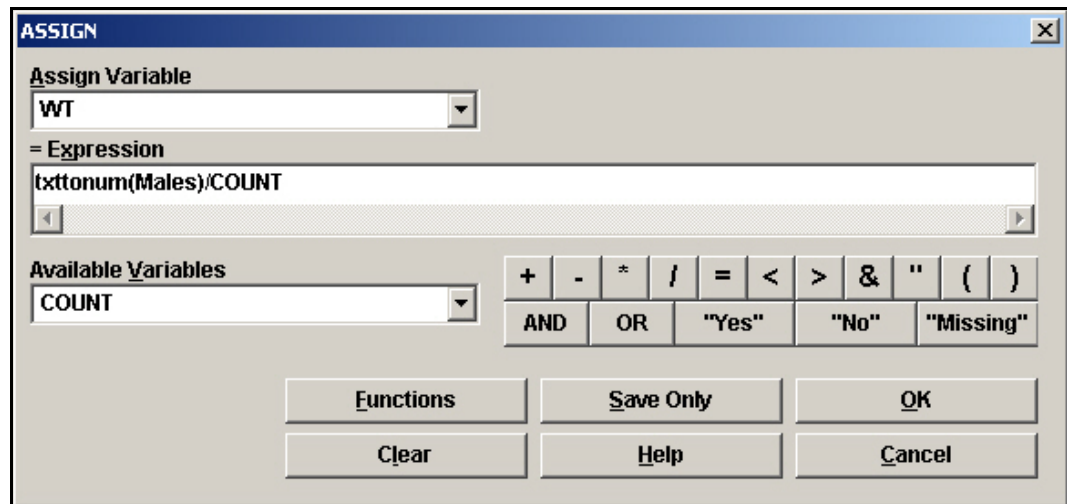
Functions | Save Only | OK  
Clear | Help | Cancel

19. Click **OK**.

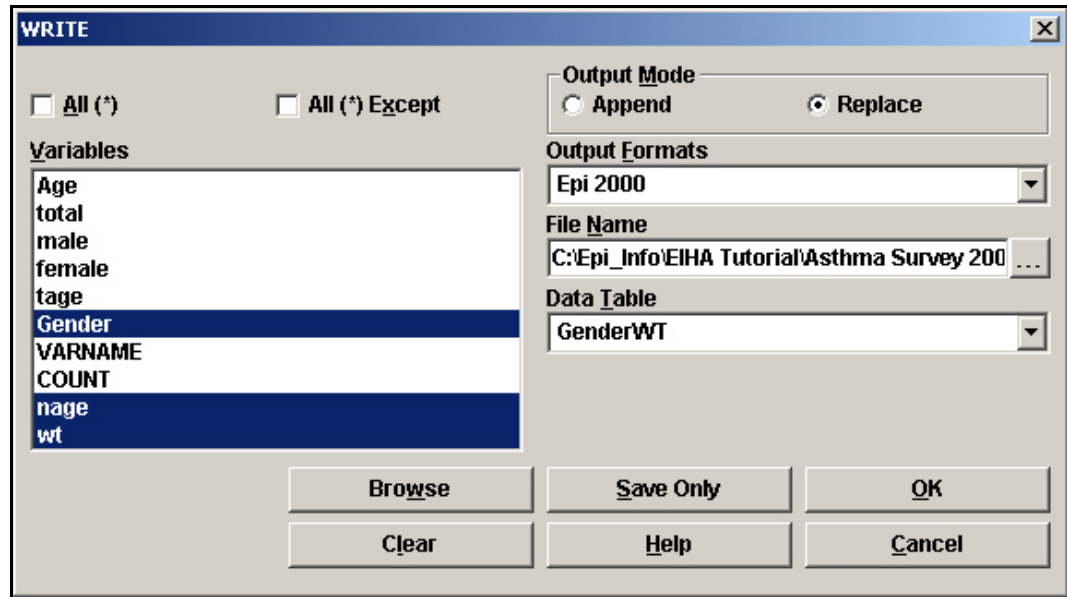
## Create a Weight Variable

Now you will create a Weight variable to use in your calculations. First, select the males in the dataset, create the Weight calculation, and write the variables out to a new table. Repeat the process for the females in the dataset.

1. Click **Select**. The SELECT dialog box opens.
2. Select Gender="Male".
3. Click **OK**. You should have 8 records.
4. Click **Define** and create a new variable called *WT*.
5. Click **Assign**. The ASSIGN dialog box opens.
6. From the Assign Variable drop-down, select WT.
7. In the =Expression field create the code: *TXTTONUM(Males)/Count*.
  - You are changing the variable Male to a number and you are dividing it by the variable Count, which is the number of students in the survey who are male.



8. Click **OK**.
9. Click **Write/Export**. The WRITE dialog box opens.
10. Select the Variables Gender, NAge, and WT.
11. From the Output Mode section, select Replace.
12. From the File Name Browse window, select the Asthma Final project.
13. In the Data Table field, type *GenderWT*.



14. Click **OK**.

15. Click **Cancel Select**.

## Practice

Repeat steps 1-15 above for the Females in the survey.

Instead of Replacing the file GenderWT, however, you will APPEND the female information.

Create a PSU Variable

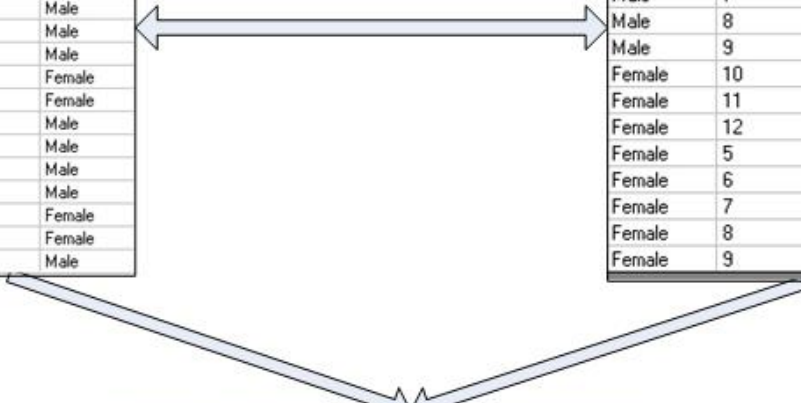
To create the PSU variable, you are going to open the 800 records from the viewPreInterventionSurvey table and relate them to the GenderWT table you created in the previous section. There are some formatting and data management steps you must follow prior to completing the RELATE.

viewPreInterventionSurvey  
Build Key Age and Gender

Age	Gender
11	Male
7	Female
10	Male
9	Male
11	Male
10	Male
7	Male
9	Male
8	Male
9	Male
10	Female
9	Female
8	Male
10	Male
6	Male
6	Male
11	Female
10	Female
8	Male

GenderWT Table  
Build Key NAge and Gender

Gender	NAge	WT
Male	10	28.0958904109589
Male	11	25.5064935064935
Male	12	1016.5
Male	5	455.25
Male	6	37.469387755102
Male	7	30.15625
Male	8	30.5294117647059
Male	9	36.8947368421053
Female	10	23.8375
Female	11	28.1969696969697
Female	12	948
Female	5	214.5
Female	6	52.9411764705882
Female	7	22.5243902439024
Female	8	31.728813559322
Female	9	25.88



Age	Gender	GenderI	NAge	WT
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589
10	Male	Male	10	28.0958904109589

Related Table

- 1. Click **Read/Import**.
- 2. Click the **Show All** radio button.

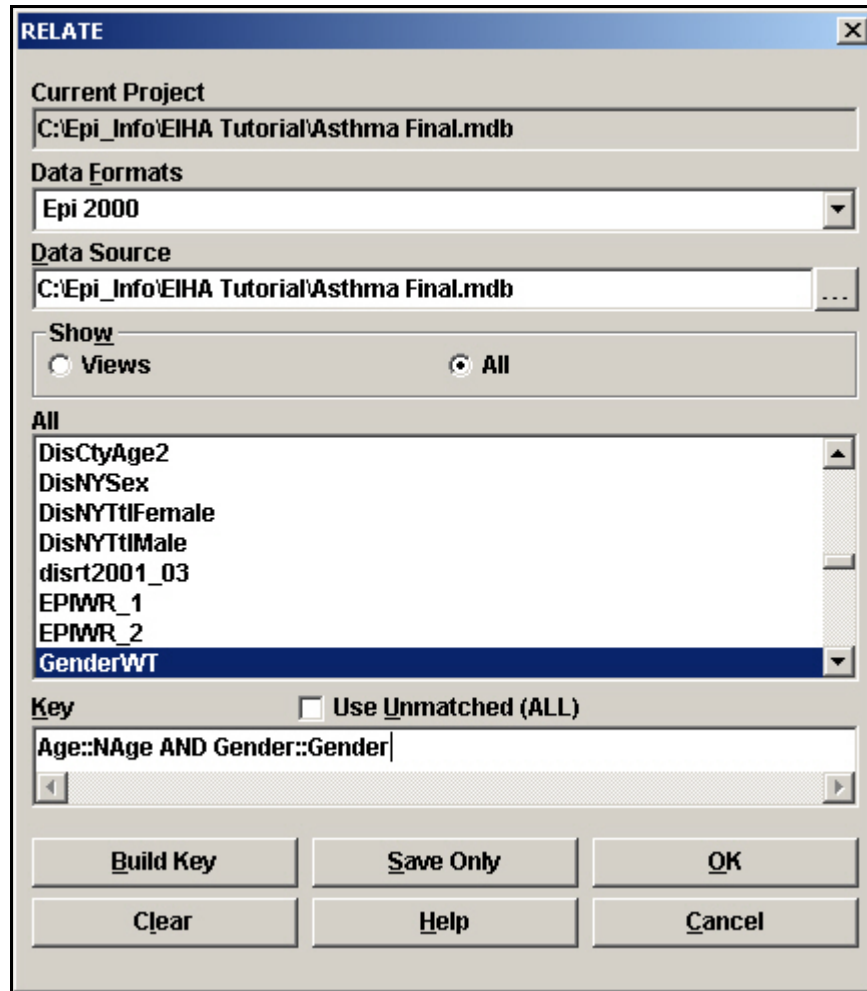
3. Select the new table GenderWT.
4. Click **List** to see the computed weight values.
  - In the Gender column, make sure the words Male and Female are written with capitol letters. If not, modify the records using the List Update command. Complex Sample routines are case-sensitive.
5. Click **Display** to see your format changes.
6. Click **Read/Import**.
7. Read in the Asthma Final project with your 800 records.
  - Click **List** to verify that the words Male and Female are written with capitol letters. If not, modify the records using the List Update command. Complex Sample routines are case-sensitive.
8. Click **Assign** and compute the Age variable **Hint:** Age=YEARSDOB, 10/25/2005).
9. From the Command Tree Statistics folder, click **Tables**. The TABLES dialog box opens.
10. For the Exposure Variable drop-down, select Age.
11. From the Outcome Variable drop-down, select Gender.
12. Click **OK**.
  - Look at the table and see that all the ages and genders of the data are represented.
  - Now you know that you can use the RELATE command to relate your GenderWT table to the survey records.



GENDER			
Age	Female	Male	TOTAL
<b>5</b>	8	4	12
Row %	66.7	33.3	100.0
Col %	2.0	1.0	1.5
<b>6</b>	34	49	83
Row %	41.0	59.0	100.0
Col %	8.4	12.4	10.4
<b>7</b>	82	64	146
Row %	56.2	43.8	100.0
Col %	20.2	16.2	18.3
<b>8</b>	59	68	127
Row %	46.5	53.5	100.0
Col %	14.5	17.3	15.9
<b>9</b>	75	57	132
Row %	56.8	43.2	100.0
Col %	18.5	14.5	16.5
<b>10</b>	80	73	153
Row %	52.3	47.7	100.0
Col %	19.7	18.5	19.1
<b>11</b>	66	77	143
Row %	46.2	53.8	100.0
Col %	16.3	19.5	17.9
<b>12</b>	2	2	4
Row %	50.0	50.0	100.0
Col %	0.5	0.5	0.5
<b>TOTAL</b>	406	394	800
Row %	50.8	49.3	100.0
Col %	100.0	100.0	100.0

13. Click **Relate**. The RELATE dialog box opens.

14. Select the table GenderWT. For this Relate, you will type the code into the Key field.



15. Move the cursor into the Key field.

16. Type the following Relate: *Age::NAge AND Gender::Gender*.

- Relates for redefined or new variables must be typed into the Program Editor, or the Key field.

17. Click **OK**.

- Code will appear in the Program Editor as `RELATE GenderWT Age::NAge AND Gender::Gender`.
- Now you need to define a primary sampling unit or PSU. In this example, PSU will be the SchoolNum and Age.

18. Click **Define** and create a new variable called *PSU*.

19. Click **Assign**. The ASSIGN dialog box opens.

20. From the Assign Variable drop-down, select PSU.

21. In the =Expression field type with the keyboard or use the drop-down menus and operator buttons to create the following expression:

SchoolNum & format(nage,"00")

**ASSIGN** [X]

**Assign Variable**  
psu

**= Expression**  
SchoolNum & format(Nage,"00")

**Available Variables**  
Nage

+ - \* / = < > & " ( )  
AND OR "Yes" "No" "Missing"

Functions Save Only OK  
Clear Help Cancel

22. Click **OK**.

## Compute the County Child Prevalence

You now have all the variables and data needed to compute the prevalence for the children in the county.

1. From the Command Tree Advanced Statistics folder, click **Complex Sample Frequencies**.
2. From the Frequency Of drop-down, select Asthma.
3. From the Weight drop-down, select WT.
4. From the PSU drop-down, select PSU.

ILL	Freq	%
+	20	35%
-	37	65%
<b>Total</b>	<b>57</b>	<b>100%</b>

Complex Sample Frequencies

Frequency of: [Asthma]      Stratify by: [ ]

All (\*) Except

Weight: [wt]

PSU: [psu]

Output to Table: [ ]

Buttons: Clear, Save Only, Help, OK, Cancel

5. Click **OK**. Results appear in the Output window:

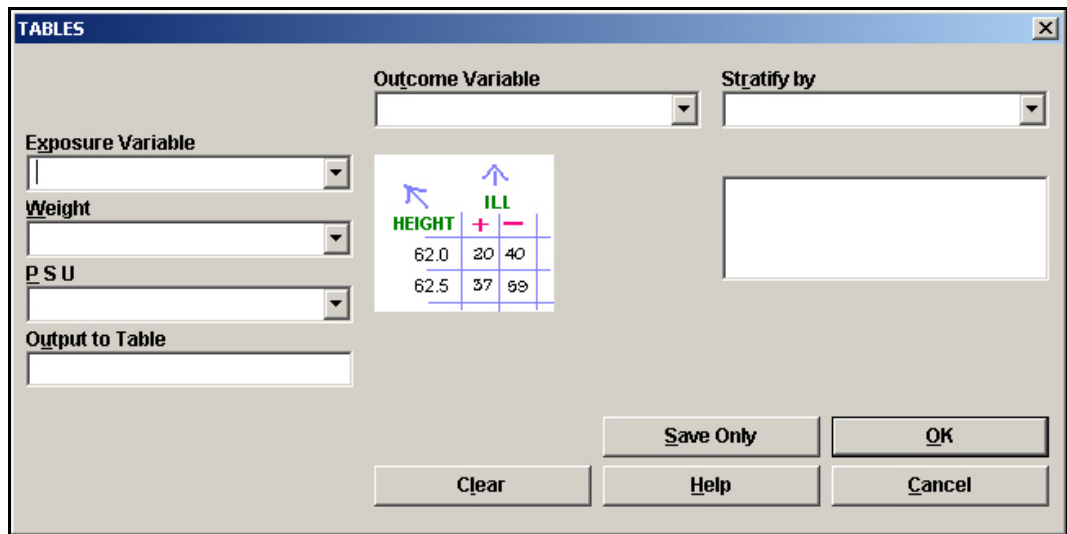
Asthma	TOTAL
<b>Yes</b>	91
Row %	100.000
Col %	9.978
SE %	1.611
LCL %	6.692
UCL %	13.264
<b>No</b>	709
Row %	100.000
Col %	90.022
SE %	1.611
LCL %	86.736
UCL %	93.308
<b>TOTAL</b>	800
Design Effect	2.312

Prevalence

Confidence Interval

**Compute for the Subgroup Gender**

- From the Command Tree Advanced Statistics folder, click **Complex Sample Tables**. The TABLES dialog box opens.



- From the Exposure Variable drop-down, select Gender.
- From the Outcome Variable drop-down, select Asthma.

4. From the Weight drop-down, select WT.
5. From the PSU Variable drop-down, select PSU.
6. Click **OK**. The results appear in the Output window.

Gender	Asthma		
	Yes	No	TOTAL
<b>Female</b>	43	363	406
Row %	10.599		
Col %	51.418	48.582	100.000
SE %	3.049	3.049	
LCL %	4.381	83.184	
UCL %	16.816	95.619	
Design Effect	3.982	3.982	
<b>Male</b>	48	346	394
Row %	9.396		
Col %	48.582	51.418	100.000
SE %	1.481	1.481	
LCL %	6.375	87.583	
UCL %	12.417	93.625	
Design Effect	1.016	1.016	
<b>TOTAL</b>	91	709	800
Row %	9.978	90.022	100.000
Col %	100.000	100.000	100.000
SE %	1.611	1.611	
LCL %	6.692	86.736	
UCL %	13.264	93.308	
Design Effect	2.312	2.312	

Statistical Table Results for Child County Prevalence by Gender

**CTABLES COMPLEX SAMPLE DESIGN ANALYSIS OF 2 X 2 TABLE**

Odds Ratio (OR) 1.143  
 Standard Error (SE) 0.400  
 95% Conf. Limits (0.53, 2.467 )

Risk Ratio (RR) 1.128  
 Standard Error (SE) 0.382  
 95% Conf. Limits (0.57, 2.249 )  
 $RR = (\text{Risk of Asthma}=-1 \text{ if Gender}=\text{Female}) / (\text{Risk of Asthma}=-1 \text{ if Gender}=\text{Male})$

Risk Difference (RD%) 1.203  
 Standard Error (SE) 3.489  
 95% Conf. Limits (-5.91, 8.318 )  
 $RD = (\text{Risk of Asthma}=-1 \text{ if Gender}=\text{Female}) - (\text{Risk of Asthma}=-1 \text{ if Gender}=\text{Male})$

Other questions you could answer using Analysis with similar steps and data:

- What is the current asthma prevalence among children in the state?
- Is the state childhood asthma prevalence higher in any population subgroups?
- How is the county childhood asthma prevalence different from the state?

## Mortality Rates

As an illustration to provide with your proposal, you would like to graph the mortality rates for the county and the state.

H. Are the asthma mortality rates for the state and county significantly different?

There are two pre-made tables inside the Asthma Final project. One is a table called Population and the other is called Numerators. Each table was created using the Summarize and Write commands with information from CDC WONDER and MS Excel tables with NY specific data. New variables were defined to hold data and reformat it for use in this lesson.

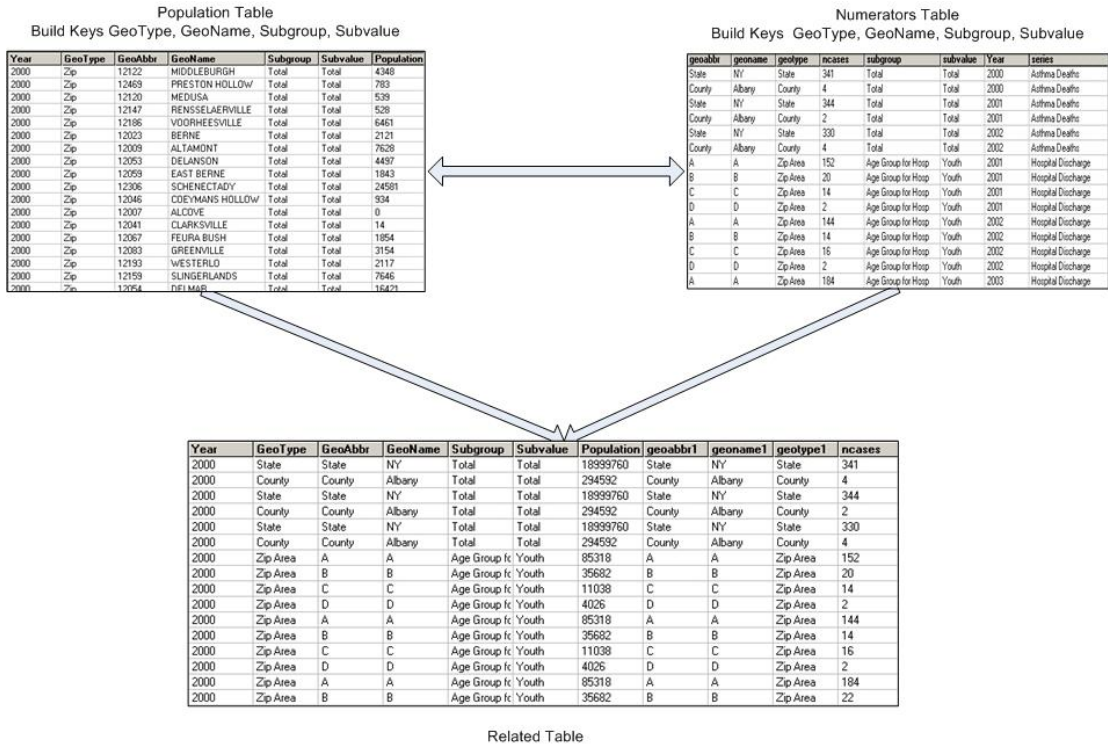
The Wide-ranging OnLine Data for Epidemiologic Research or WONDER, is an easy-to-use internet system that makes the information resources of the CDC available to public health professionals and the public at large. It provides access to a wide array of public health information. (Source: <http://wonder.cdc.gov/wonder/help/faq.html#1>)

You want to create a bar graph of the mortality rates for the county and state. To compute the results you need to use the RELATE command and join two tables so you can divide the total asthma deaths in the county and the state by the total population of the county and the state. You will create a related table to join the population data from WONDER to the data table that contains the number of deaths in the county and state. Once the rates are computed, you can easily plot them onto a graph.

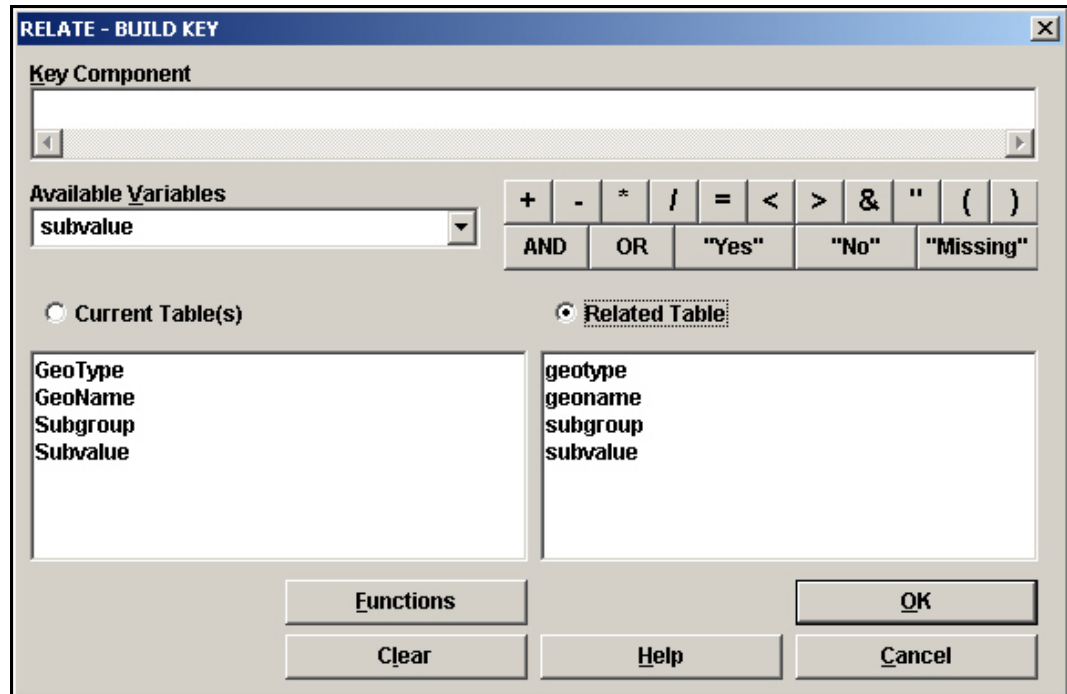
1. Click **Read/Import**. The READ dialog box opens.
2. Click the **Show All** radio button.
3. Select the Population table.
4. Click **OK**. You should have 233 records.
5. Click **List** to view the contents.
  - Now you will use the RELATE command to link the information in the population (denominator) to the numerators in order to compute a rate.



## INTERMEDIATE ANALYSIS

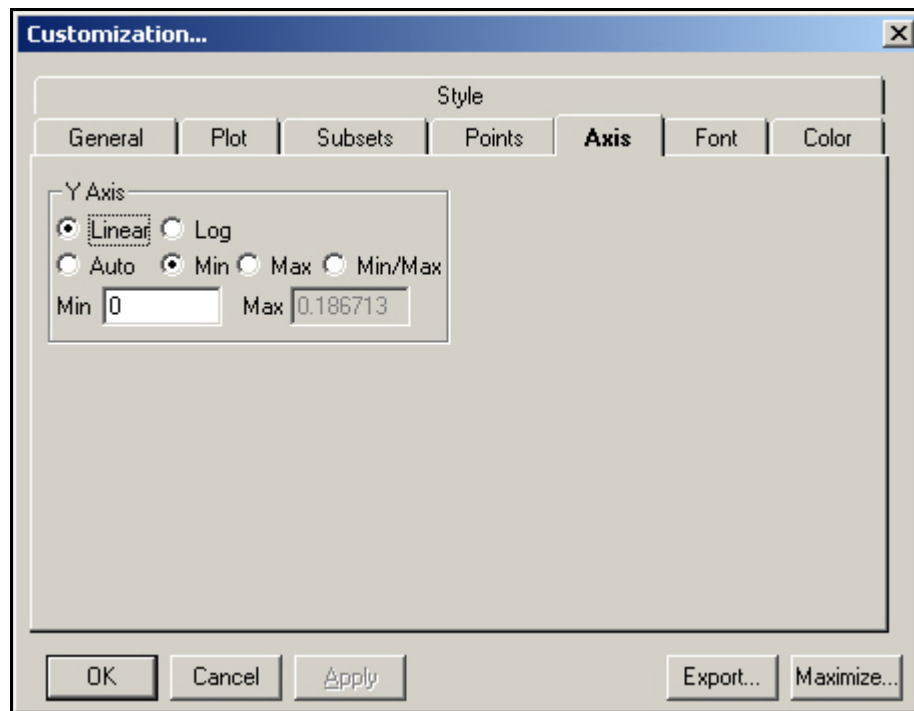


6. Click **Relate**. The RELATE dialog box opens.
7. Click the **Show All** radio button.
8. Select the Numerators table.
9. Click **Build Key**. The RELATE-BUILD KEY dialog box opens.
10. Relate the two tables on the following fields: Geotype, Geoname, Subgroup, and Subvalue.
  - The Population table contains the total numbers for the population by each of the subgroups selected. The Numerators table contains the Number of Cases by series for each of the subgroups selected.
  - Geotype contains information on how the information is divided. Most of the values in this field are Zip.
  - Geoname contains information on the specific breakdown of data by counties or state.
  - Subgroup contains information on how the information is divided, for example Age.
  - Subvalue contains the specific subgroup information, for example age 5-14.



11. Click **OK**.
12. From the READ dialog box, click **OK**. You should have 42 records.
13. Click **Select**. The SELECT dialog box opens.
14. From the Available Variables drop-down, select Series.
15. In the =Expression field, type *"Asthma Deaths"*.
16. Click **OK**. You should have 6 records.
17. Click **Define** and create a new variable called *Rate*.
18. Click **Assign**. The ASSIGN dialog box opens.
19. From the Assign Variable drop-down, select Rate.
20. In the =Expression field, type  $ncases/population*10000$ .
  - You are creating a rate by dividing the number of cases by the population and multiplying by 10,000.
  - You will have one rate that when listed will appear with an E-02 value. Analysis avoids listing values that begin with 0.0. You can reformat the values to Text to view it the 0.0 format. The value will appear as 0.0 in the graph.
21. Click **OK**.

22. From the Command Tree Statistics folder, click **Graph**. The GRAPH dialog box opens.
23. From the Graphtype drop-down, select Bar.
24. From the Main Variable drop-down, select Year1.
25. From the Weight Variable drop-down, select Rate.
26. From the Bar for Each Value of drop-down, select GeoName.
27. Click **OK**. The graph appears in the Epi Graph window.
28. Right click the graph and select Customization Dialog from the pop-up menu.
29. From the **General** tab in the Display section, select the Both radio button.
30. Click **Apply**.
31. Select the **Axis** tab.
32. From the Y Axis section, click the Min radio button.
33. Highlight the Min field text, and enter the number 0.

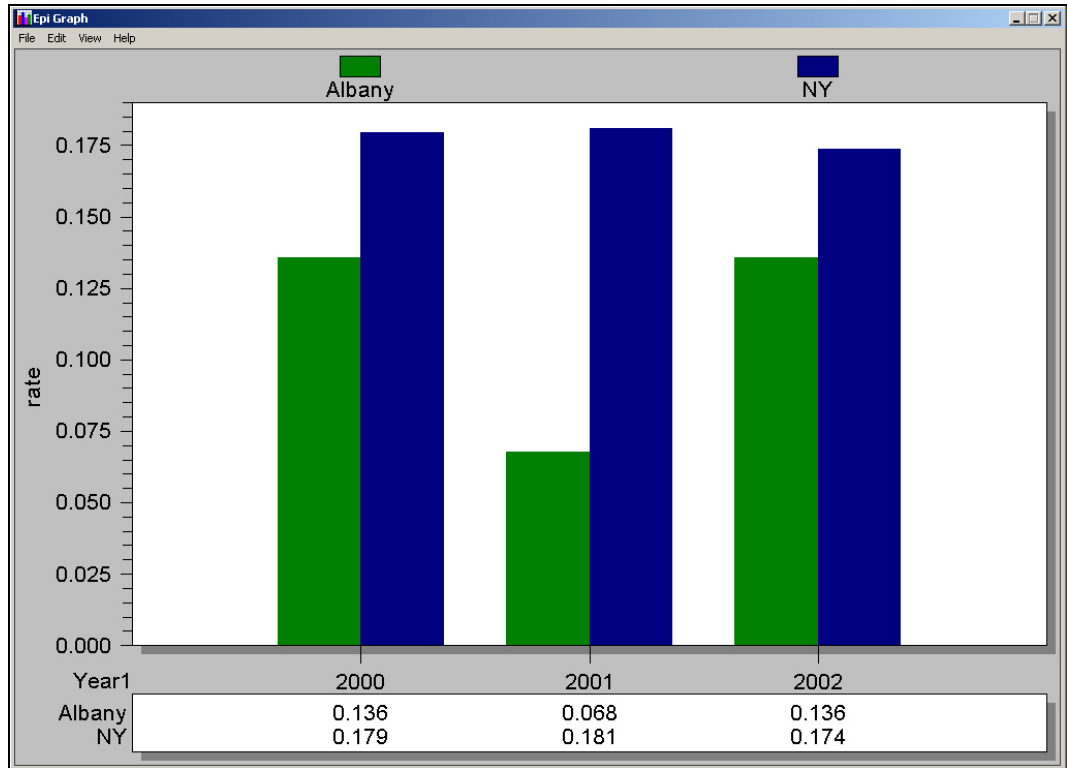


34. Click **Apply**.
35. Click **OK**.

## INTERMEDIATE ANALYSIS

- Notice the Y Axis on the graph has been reset to start at 0.
- Notice the rates for each year are visible at the bottom of the graph.

36. Close Epi Graph to return to the Analysis module.



## 10 Year Trends- Hospital Discharge Data

As an illustration to provide with your proposal, you would like to graph the hospitalization rates for ten years and across age groups.

1. Is the rate of hospital discharge for asthma increasing or decreasing in any age groups?

You have a table in your Asthma Final project that contains ten years of asthma discharge data by age group. You can use this table to plot statistics and look for trends in asthma discharges.

Hospital discharge data was obtained from the Statewide Planning and Research Cooperative System (SPARCS) and was imported into Analysis as an MS Excel file and formatted.

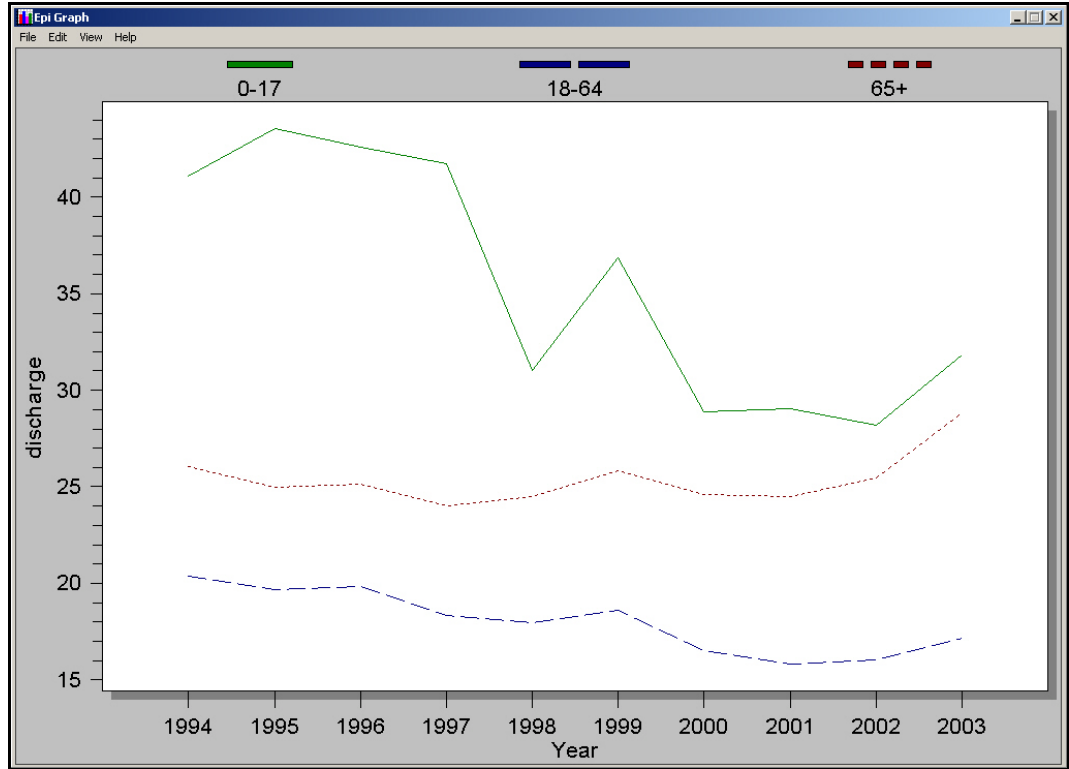
Population information for this exercise was downloaded from the CDC WONDER system.

Ages in the population denominator table were not divided into the same groups as the hospital data. To match the two groups, a third table was created to hold the age data and to match the values into three groups. The age table is called AgeProxy. Information from the AgeProxy table was selected from the population table and written out for use with the discharge data to create the ten-year trend graph.

1. Click **Read/Import**.
2. Click the **Show All** radio button.
3. Select the file NYDisAge10yr.
4. Click **OK**. You should have 30 records.
5. Click **List** to see the contents of the file. Notice the Series name is NY Discharges by Age.
6. Click **Relate**. The RELATE dialog box opens.
7. Locate and select the file called AgeProxy.
8. Click **Build Key**. The RELATE- BUILD KEY dialog box opens.
9. Relate the NYDisAge10yr table and the AgeProxy table on the following variables: Geotype, Geoname, Subgroup, and Subvalue.
  - Geotype is State.
  - Geoname is NY.

- Subgroup is Age.
  - Subvalue is the breakdown by Age Groups 0-17, 18-64, and 65+.
10. Click **OK**.
  11. From the RELATE dialog box, click **OK**. You should have 30 records.
  12. Click **Define** and create a new variable called *Discharge*.
  13. Click **Assign**.
  14. From the Assign Variable drop-down, select Discharge.
  15. In the =Expression, type  $ncases/population*10000$ .
  16. Click **OK**.
  17. Click **Graph**. The GRAPH dialog box appears.
  18. From the Graph Type drop-down, select Line.
  19. From the Main Variable drop-down, select Year.
  20. From the Weight Variable drop-down, select Discharge.
  21. From the Line for Each Value Of drop-down, select Subvalue.
  22. Click **OK**. Epi Graph opens with the results.

# INTERMEDIATE ANALYSIS



### 3 Year Rates- Hospital Discharge Data

As an illustration to provide with your proposal, you would like to graph the hospital discharge rates for three years and across age groups.

J. Are there significant differences in the three-year asthma discharge rates for the state, versus the county, by age?

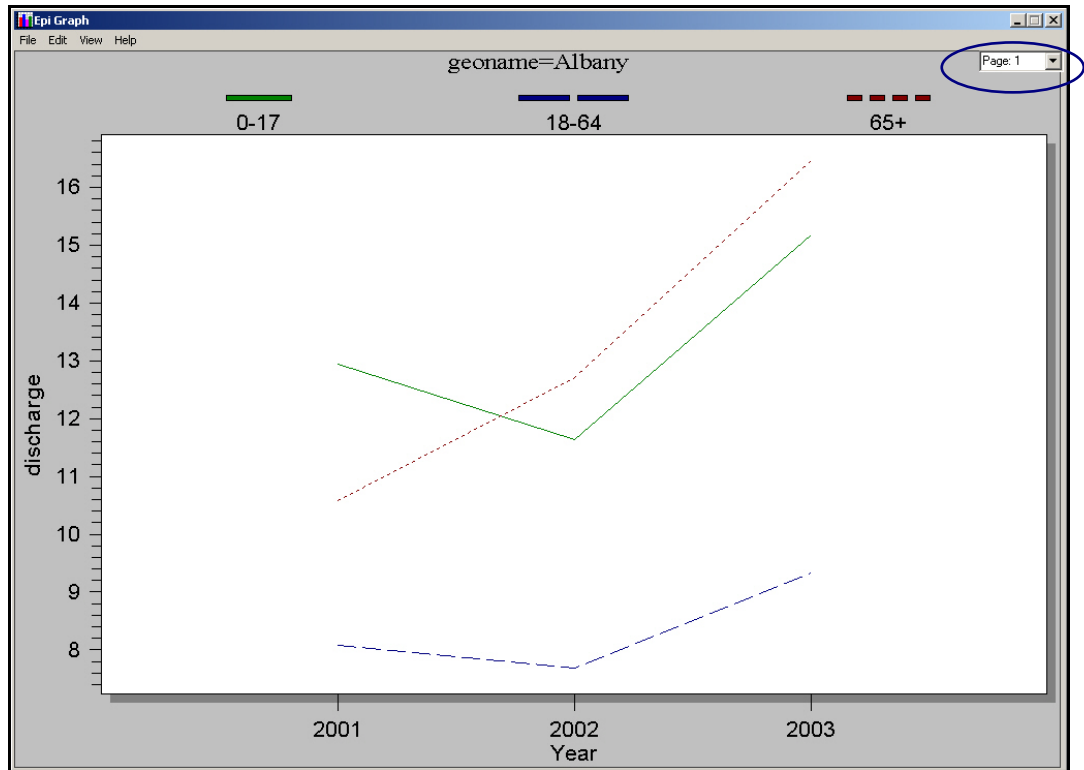
Being able to Read, Write, and Relate tables is an important and useful concept in using Analysis. For this exercise, you will read in two tables, one containing state discharge information, and one containing county discharge information. You will then use the Write command to combine the two tables in order to relate them to the AgeProxy table from the earlier lesson. The Age Proxy table contains the population numbers divided into age groups that you need to compute the rates.

1. Click **Read/Import**.
2. Click the **Show All** radio button.
3. Select the table DisCtyAge.
4. Click **List**, if you want to see how the table is arranged.
5. Click **Write/Export**.
6. From the Output Mode section, click Replace.
7. From the File Name Browse window, select Asthma Final.
8. In the Data Table field, type *Dis3Yrs*.
9. Click **OK**.
10. **Read/Import** the table NY3Yr.
11. Click **List**, if you want to see how the table is arranged.
12. Click **Write/Export**.
13. From the Output Mode section, click Append.
14. From the File Name Browse window, select Asthma Final.
15. From the Data Table drop-down, select Dis3Yrs.
16. Click **Ok**.



17. Click **Read/Import**.
18. Click the **Show All** radio button.
19. Select the Dis3Yrs table.
20. Click **OK**. You should have 18 records.
21. Click **Relate**.
22. Select the AgeProxy table.
23. Click **Build Key** and join on the following fields: Geotype, Geoname, Subgroup, and Subvalue.
  - Geotype is County or State.
  - Geoname is Albany or NY.
  - Subgroup is Age.
  - Subvalue is the age group breakdowns 0-17, 18-64, and 65+.
24. Click **OK**.
25. From the RELATE dialog box, click **OK**.
26. Click **Define** and create a new variable called *Discharge*.
27. Click **Assign** to create the following code:  $discharge = ncases / Population * 10000$ .
28. Click **Graph**. The GRAPH dialog box opens.
29. Select the Graph Type of Line.
30. From the Main Variable drop-down, select Year.
31. From the Weight Variable drop-down, select Discharge.
32. From the Line for Each Value of, select Subvalue.
33. From the One Graph for Each Value of drop-down, select GeoName.
34. Click **OK**. Epi Graph opens and contains two graphs: one for NY, and one for Albany.

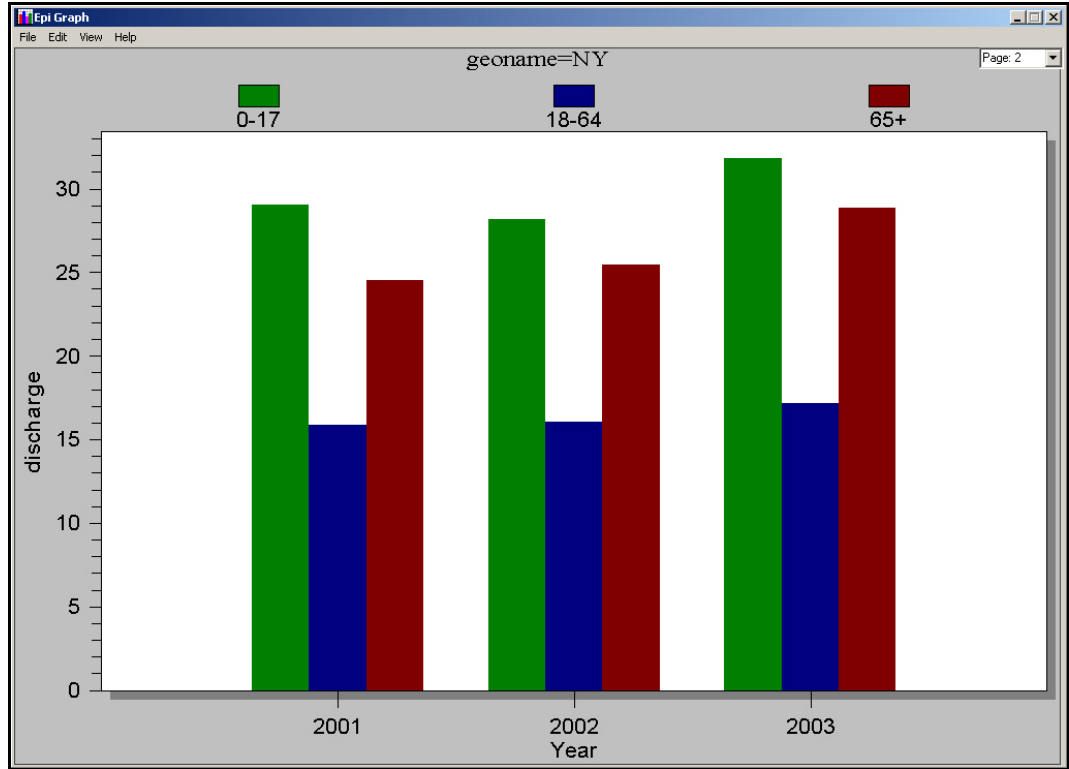
## INTERMEDIATE ANALYSIS



- Use the Page drop-down menu at the top right of the page to go between the two graphs.
- Use the Customization dialog box to change the graph types to Bar, and see the difference between the two graphs.
- Use the Customization dialog box to change the Axis Min to 0 and see the difference.

Your graph should look like the following:

# INTERMEDIATE ANALYSIS



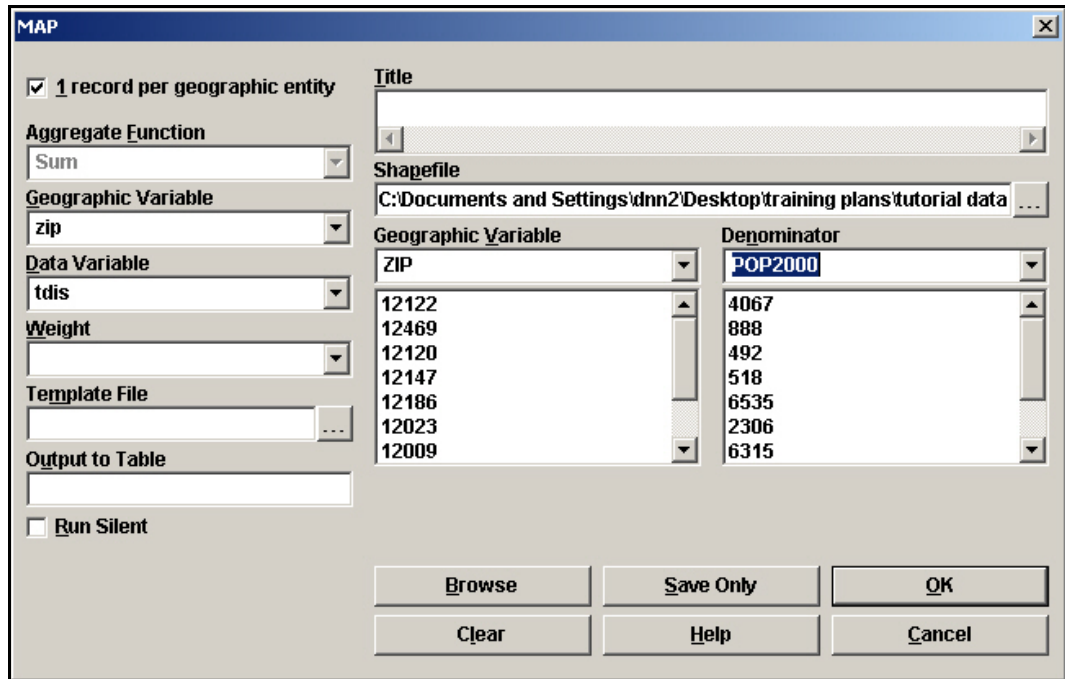
## Mapping Rates

As an illustration to provide with your proposal, you would like to map the hospitalization rates by zip code for your state.

K. Based on the map with asthma hospitalization rates by zip code, where are the high risk areas?

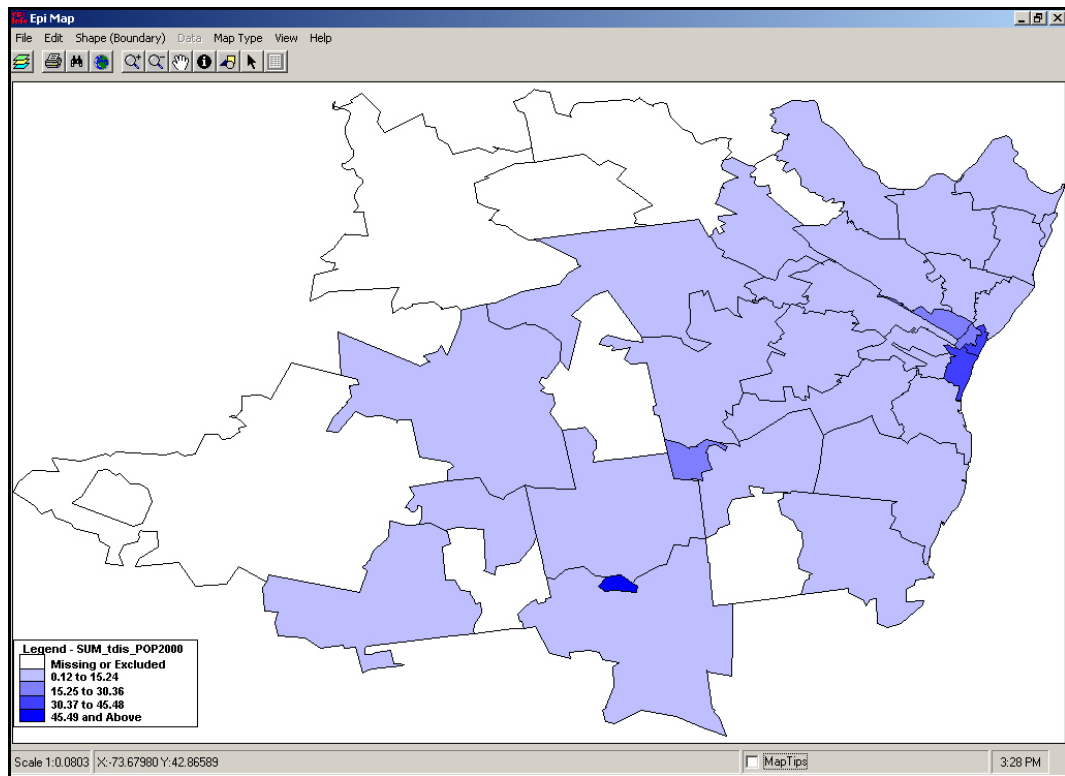
In order to map high-risk areas, you will join discharge data arranged by zip code with a map .DBF file.

1. Click **Read/Import**.
2. Click the **Show All** radio button.
3. Select the Dis0103byZip table.
4. Click **List** to view the variables.
  - Notice the table is arranged by county zip code for three years. Also, notice that the age groups have been recoded as Youth, Adult, and Senior. This is another way to work with age groups.
5. Click **Define** and create a new variable called *TDis*.
6. Click **Assign** and create the following code:  $TDis = (youth + adult + senior) * 10000 / 3$ .
  - You are adding the number of discharges for each age group, multiplying by the population factor, and dividing by the three years in the table. You are going to map this data and get your population denominator from the map data (census information for the year 2000).
7. From the Command Tree Statistics folder, click **Map**. The MAP dialog box opens.
8. Select the 1 Record per Geographic Entity checkbox.
9. From the Geographic Variable drop-down, select Zip.
10. From the Data Variable drop-down, select *TDis*.
11. From the Shapefile Browse window, select the ALBZCTA shapefile.
12. From the Geographic Variable drop-down, select Zip.
13. From the Denominator drop-down, select POP2000.



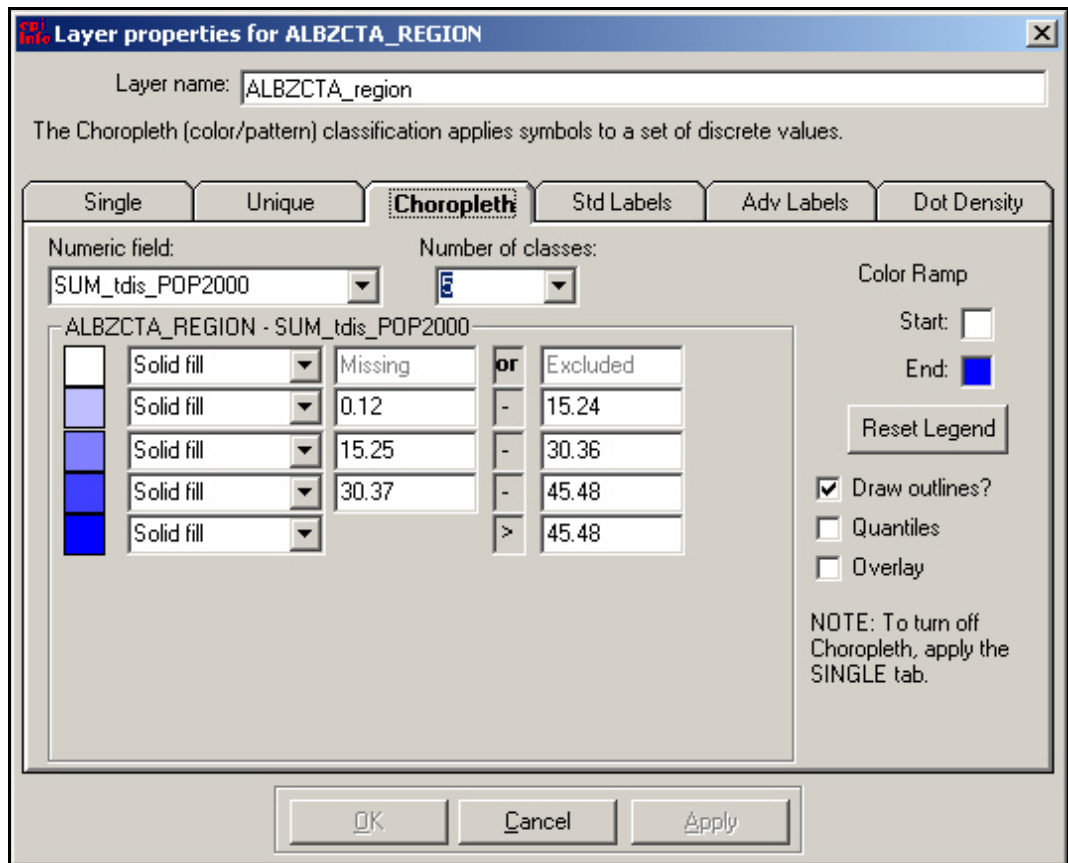
14. Click **OK**.

15. Click **Continue** at the Incomplete Join screen. Epi Map opens with the rates mapped in a choropleth.



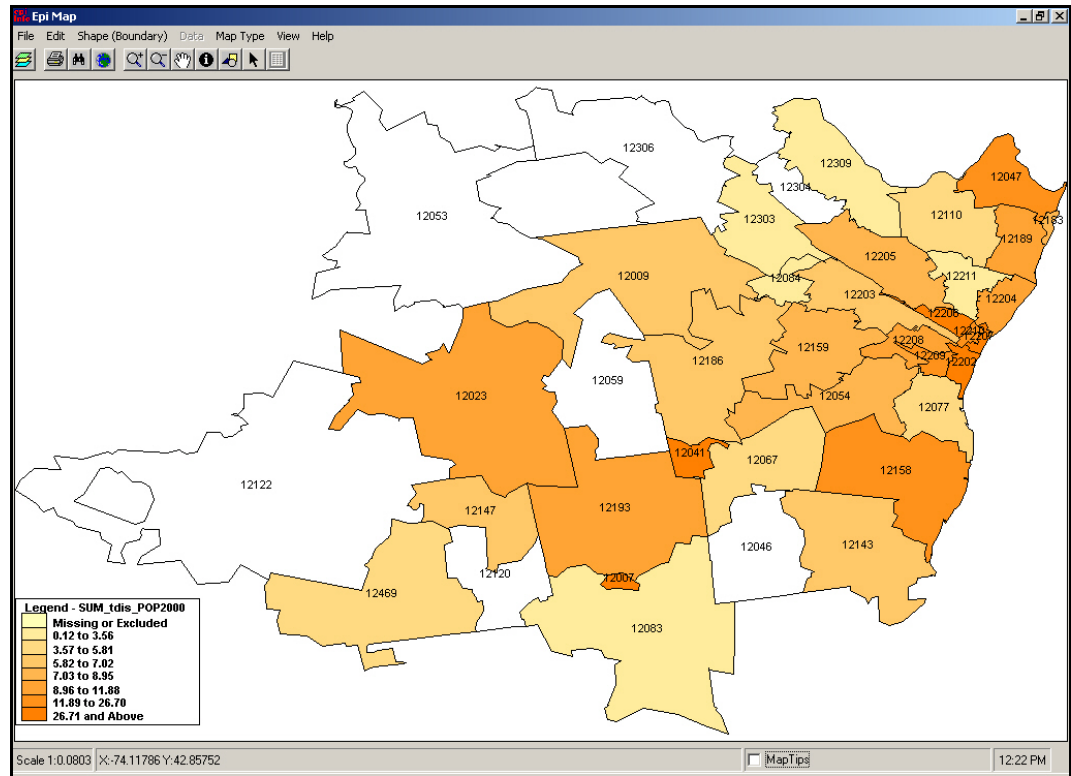
## Customize the Map

1. Open the Map Manager.
2. Click **Properties**. The Choropleth tab opens.



3. From the Number of Classes drop-down, select 8.
4. Select the Quantiles checkbox.
5. Select the Overlay checkbox.
6. Click **Reset Legend**.
7. Change the Color Ramp, if you would like.
8. Click **Reset Legend**.
9. Click **Apply**.
10. Click the **Standard Labels** tab.
11. From the Text field drop-down, select Zip.

12. Click **Apply**.
13. Close the Properties dialog box.
14. Close the Map Manager



## Income and Zip Code Scatter Plot


As an illustration to provide with your proposal, you would like to graph the hospitalization rates across median income brackets and see if there is a correlation.

L. Does income affect the rate of hospitalization for asthma?

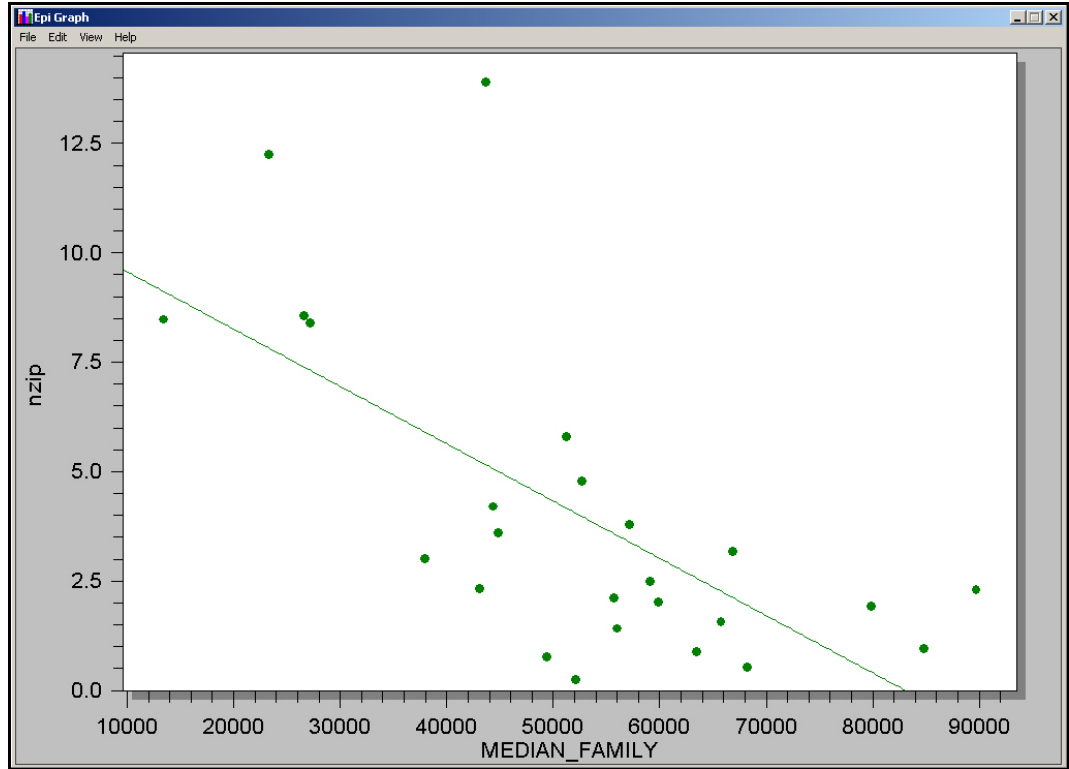
Look at median family income and per capita income by zip code. You can create a scatter plot of income by zip as a comparison with discharges by zip.

1. Click **Read/Import**. You should still be in the project Asthma Final.MDB.
2. Click the **Show All** radio button.
3. Select the IncomeData table. You should have 38 records.
4. Click **List** to view the data.
5. Click **Read/Import**.
6. Click the **Show All** radio button.
7. Select the RatebyZip table. You should have 95 records.
8. Click **List** to view the Data.
9. Click **Select**. The SELECT dialog box opens.
10. In the Select Criteria field, type *Year=2001*. You should have 35 records.
11. Click **Relate**.
12. Select the IncomeData table.
13. Click **Build Key**. The BUILD KEY dialog box opens.
14. Join the tables on the variables Zip and Zip2.
  - Zip and Zip2 are the zip code fields in the RatebyZip and IncomeData tables.
15. Click **OK**. You should have 25 records.
16. Click **Relate**. The RELATE dialog box opens.
17. From the Data Formats drop-down, select dBase IV.
  - You are going to create a relate using the map file .DBF.



18. Click the Data Source **Browse**  button.
19. From the Relate window, select the ALBZCTA\_region.DBF.
20. Click **Open**. A data format message appears.
21. Click **OK**. You need to manually assign the Build Key for this Relate.
22. Move your cursor to the Key field located in the RELATE dialog box.
23. In the Key field, type *zip::zip*.
  - You are relating a third table on the Zip Code field.
24. Click **OK**. The temporary link window opens.
25. Click **OK** to make a temporary link. You should have a record count of 25.
26. Click **Define** and create a new variable called *NZip*.
27. Click **Assign**. The ASSIGN dialog box opens.
28. From the Assign Variable drop-down, select NZip.
29. In the =Expression field type, *tdis/pop2000*.
30. Click **OK**.
31. Click **Graph**.
32. From the Graph Type drop-down menu, select Scatter XY.
33. From the Variables drop-down, select Median Family and NZip.
34. Click **OK**. Epi Graph opens.

# INTERMEDIATE ANALYSIS



### Risk Ratio

To focus on certain areas of the state for your proposal, you would like to know if there are any locations by zip code that have a higher risk for developing asthma than other areas. You would also like to develop a .PGM that you can run on each zip code separately.

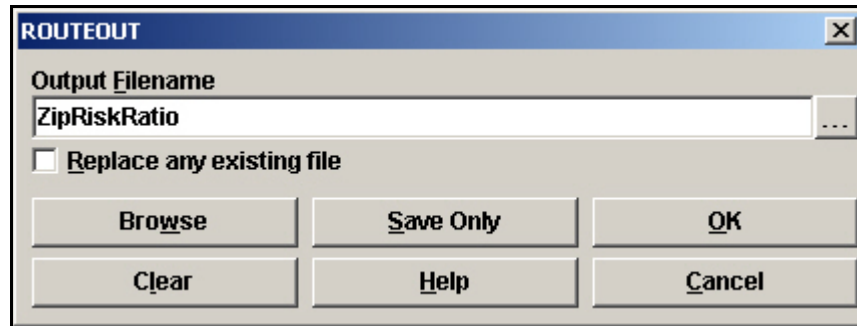
M. What is the risk ratio of being hospitalized for asthma if you live in the zip code 12202 versus if you lives in the zip code 12211?

In order to compute the risk ratios you need to create a relationship between the number of discharges in each zip code and total population for the zip code and the county. The Tables command gives you the ability to create a 2x2 table and gives you risk ratio results for dichotomous questions. To get the risk ratio we need to compute the differences between the discharge numbers in and out of the requested zip codes.

For this example, you will be using the Routeout command to save your results, and you will also need to save your PGM. Saving the PGM will allow you to run the program for each individual zip code without rewriting any code.

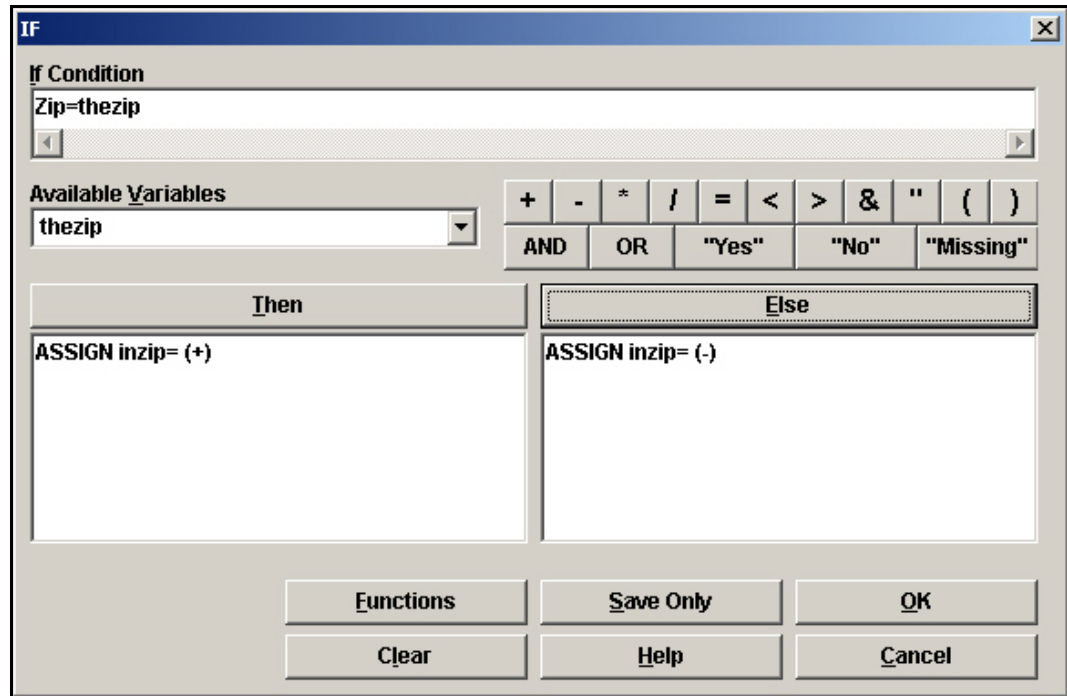
		INHOSP	
		YES	NO
INZIP	YES	YES	NO/YES
	NO	YES/NO	NO

1. From the Command Tree Output folder, click **Routeout**. The ROUTEOUT dialog box opens.
2. In the Output Filename field, type *ZipRiskRatio*.



3. Click the Replace Any Existing File checkbox.
4. Click **OK**.
5. Click **Read/Import**.
6. Select and open the table DischargesbyZip. You should have 91 records.
  - Use LIST to view how the table is organized and the type of data it contains. This table contains the hospital discharges per zip code for Albany County. The zip code field is called Zip.
7. Click **Read/Import**.
8. Select and open the Population table.
9. Click **Relate**. The RELATE dialog box opens.
10. Select the file DischargesbyZip.
11. Click **Build Key**. The RELATE dialog box opens.
12. Relate the tables on the GeoAbbr and Zip fields.
  - GeoAbbr is the zip code field in the Population table.
  - Zip is the zip code field in the DischargesbyZip table.
13. Click **OK**.
14. From the RELATE dialog box, click **OK**. You should have 185 records that are related.
  - In order to select zip codes repeatedly you need to define a new variable, which can be updated for each individual zip as needed.
15. Click **Define** and create a new standard variable called *TheZip*.
16. Click **Assign**. The ASSIGN dialog box opens.

17. From the Assign Variable drop-down, select *TheZip*.
18. In the =Expression field type, *"12202"*.
19. Click **OK**.
20. Click **Define** and create a new variable called *InZip*.
21. From the Command Tree Select/If folder, click **If**. The IF dialog box opens.
  - You want to create a statement that says IF the value of *zip* is 12202, then the value is Yes, else the value is No. You need Yes/No fields to create the 2x2 table.
22. In the If Condition field, type the code: *Zip=TheZip*.
23. Click **THEN**. The Command Tree activates.
24. Click **Assign**. The ASSIGN dialog box opens.
25. From the Assign Variable drop-down, select *InZip*.
26. In the = Expression field, click **Yes**.
27. Click **Add**. The IF dialog box opens.
28. Click **ELSE**.
29. Click **Assign**. The ASSIGN dialog box opens.
30. From the Assign Variable drop-down, select *InZip*.
31. In the = Expression field, click **No**.
32. Click **Add**. The IF dialog box opens.



33. Click **OK**. The Program Editor code will read

```
IF zip=thezip THEN
  ASSIGN inzip= (+)
ELSE
  ASSIGN inzip= (-)
END
```

34. Click **Select**. The SELECT dialog box opens.

35. In the SELECT CRITERIA field type, *subgroup="total"*.

36. Click **OK**. You should have 37 records.

- Now you need to summarize the records to get the population totals for the hospital data, and the population data in one table.

37. From the Command Tree Statistics folder, click **Summarize**. The SUMMARIZE dialog box opens.

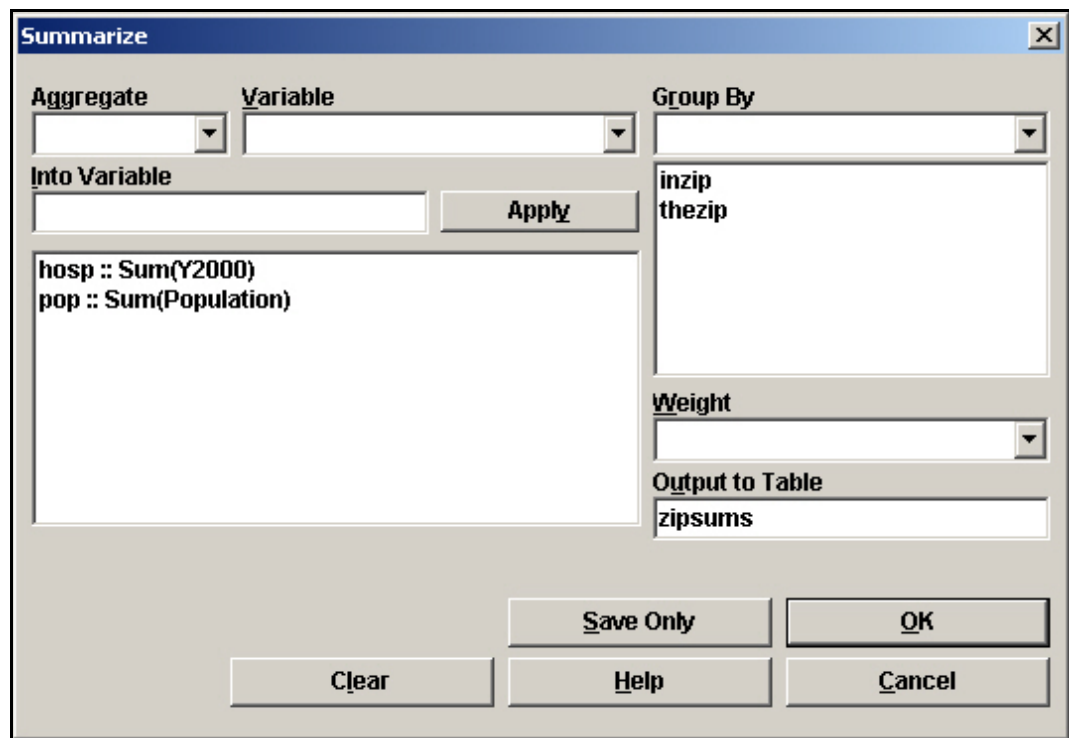
38. From the Aggregate drop-down select, Sum.

39. From the Variable drop-down, select Population.

40. In the Into Variable field, type *Pop*.

41. Click **Apply**.

42. From the *Aggregate* drop-down, select *Sum*.
43. From the *Variable* drop-down, select *Y2000*.
  - This is the number of discharges per zip for the year 2000. You are summarizing them to get the total number of discharges for the year 2000 in Albany County.
44. In the *Into Variable* field, type *Hosp*.
45. Click **Apply**.
46. From the *Group By* drop-down, select *InZip* and *TheZip*.
47. In the *Output to Table* field, type *ZipSums*.



48. Click **OK**.
49. Click **Read/Import** and select the *ZipSums* table.
50. Click **List**. The table should look like the following:

inzip	thezip	pop	hosp
No	12202	349987	369
Yes	12202	10021	48

51. Click **Define** and create a new variable called *Num*.

52. Click **Define** and create a new variable called *InHosp*.
53. Click **Assign** and create the code: InHosp=Yes.
54. Click **Assign** and create the code: Num=Hosp.
55. Click **Write/Export**.
56. In the Variables field, select InZip, TheZip, Num, and InHosp.
57. Select the Output Mode REPLACE.
58. In the Data Table field, type *ZipSums2*.
59. Click **OK**.
  - Now create the negative values.
60. Click **Assign** and create the code: InHosp=No.
61. Click **Assign** and create the code: Num=Pop.
62. Click **Write/Export**.
63. In the Variables field, select InZip, TheZip, Num, and InHosp.
64. Select the Output Mode APPEND.
65. From the Data Table drop-down, select ZipSums2.
66. Click **OK**.
67. Click **Read/Import**.
68. Locate and select the table ZipSums2.
69. Click **List** to see the values. The table should look like the following:

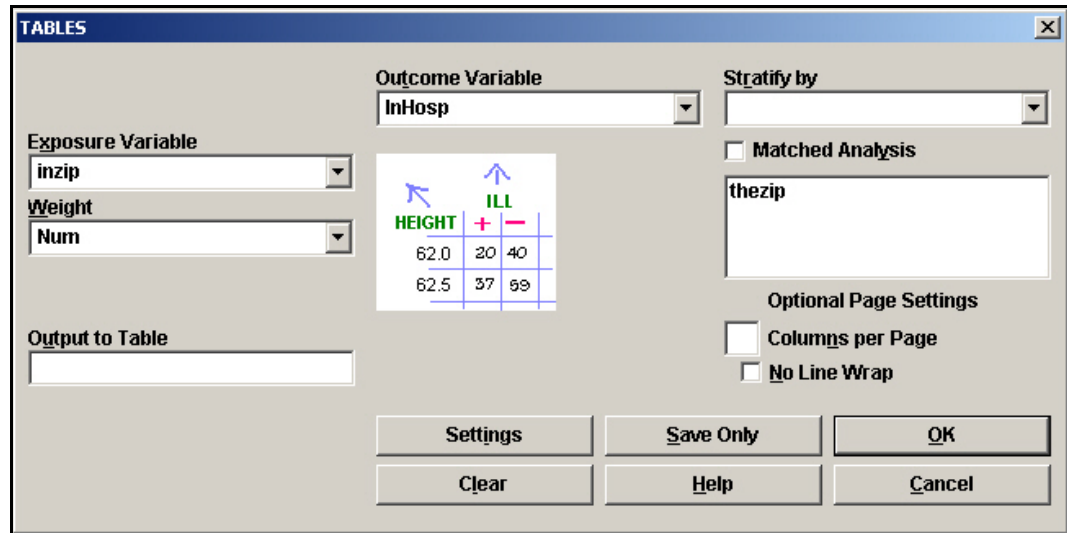
inzip	thezip	Num	InHosp
No	12202	369	Yes
Yes	12202	48	Yes
No	12202	349987	No
Yes	12202	10021	No

70. From the Command Tree Statistics folder, click **Tables**. The TABLE dialog box opens.
71. From the Exposure Variable drop-down, select InZip.
72. From the Outcome Variable drop-down, select InHosp.



73. From the Weight drop-down, select Num.

74. From the Stratify By drop-down, select TheZip.



75. Click **OK**.

**inzip : InHosp, thezip=12202**

[Forward](#)

**INHOSP**

inzip	Yes	No	TOTAL
<b>Yes</b>	48	10021	10069
Row %	0.5	99.5	100.0
Col %	11.5	2.8	2.8
<b>No</b>	369	349987	350356
Row %	0.1	99.9	100.0
Col %	88.5	97.2	97.2
<b>TOTAL</b>	417	360008	360425
Row %	0.1	99.9	100.0
Col %	100.0	100.0	100.0

- The Risk Ratios appear as follows.

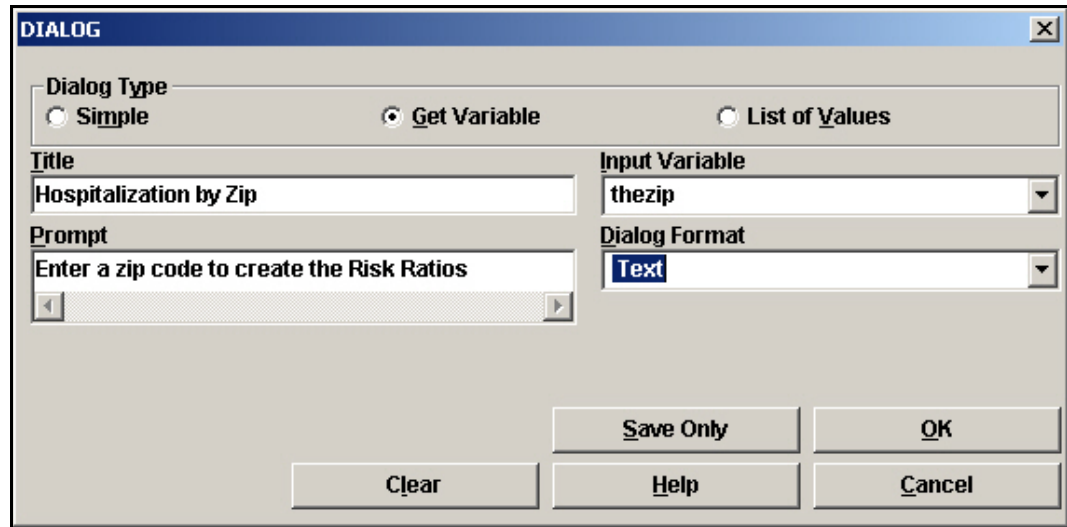
Single Table Analysis			
	Point Estimate	95% Confidence Interval	
		Lower	Upper
PARAMETERS: Odds-based			
Odds Ratio (cross product)	4.5431	3.3609	6.1411 (T)
Odds Ratio (MLE)	4.5431	3.3289	6.0939 (M)
		3.2882	6.1537 (F)
PARAMETERS: Risk-based			
Risk Ratio (RR)	4.5262		
Risk Difference (RD%)	0.3714		
(T=Taylor series; C=Cornfield; M=Mid-P; F=Fisher Exact)			
STATISTICAL TESTS			
	Chi-square	1-tailed p	2-tailed p
Chi square - uncorrected	116.8211		0.0000000000
Chi square - Mantel-Haenszel	116.8208		0.0000000000
Chi square - corrected (Yates)	113.6295		0.0000000000
Mid-p exact		0.0000000000	
Fisher exact		0.0000000000	

Four times more likely to be hospitalized for asthma if you are in this zip code.

- The PGM you created can be re-used to analyze the risk ratio for each zip code in the table. You can also create a dialog box within the .PGM that will prompt you for a zip code each time you run the program.
- Do not close the current program.

## Use the Dialog Command

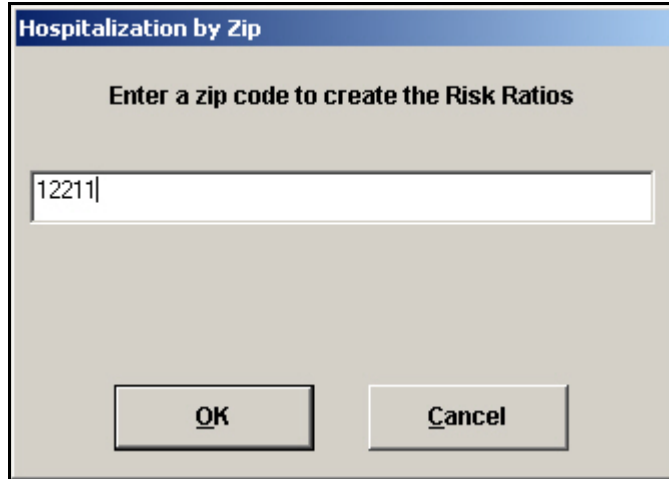
1. From the Program Editor, click **Save**. The Save Program dialog box opens.
2. In the Program field, type *ZipRatios*.
3. Click **OK**.
4. Scroll up the .PGM to the code where you assigned thezip="12202".
5. Place the cursor in front of the Assign command.
6. Type an asterisk \*.
  - This is the command to comment out a specific line of code. You want to keep the code, but you do not want it to run automatically each time you run the program.
7. Place the cursor next to the text TheZip and add a space.
8. Type the word *global* to change the variable type from standard to global.
9. Click **Run This Command** from the Program Editor to run the new DEFINE code.
10. From the Command Tree User Interaction folder, click **Dialog**. The DIALOG box opens.
11. From the Dialog Type section, select the Get Variable radio button.
12. In the Title field, type *Hospitalization by Zip*.
13. In the Prompt field, type *Enter a zip code to create the Risk Ratios*.
14. From the Input Variable drop-down, select TheZip.
15. From the Dialog Format drop-down, select Text.



16. Click **OK**. When the dialog test box opens, click **Cancel**. The code appears at the end of the program.
17. Highlight and select the code `DIALOG "Enter a zip code to create the Risk Ratios" thezip TEXTINPUT TITLETEXT="Hospitalization by Zip"`
18. Select **Edit>Cut**.
19. Paste the code under the \*Assign code.
  - Your code should look like the following example.

```
define thezip global
*assign thezip="12202"
DIALOG "Enter a zip code to create the Risk Ratios" thezip TEXTINPUT
TITLETEXT="Hospitalization by Zip"
```

20. **Save** the Program.
21. Click **Run** in the Program Editor. The code runs in the Program Editor.
22. In the Hospitalization by Zip dialog box, type 12211.
23. Click **OK**. Results appear in the Output window.



The image shows a dialog box with a blue title bar that reads "Hospitalization by Zip". Below the title bar, the text "Enter a zip code to create the Risk Ratios" is displayed. A text input field contains the number "12211". At the bottom of the dialog box, there are two buttons: "OK" and "Cancel".

- Notice that the new results were also written to the RouteOut file RiskRatio.HTM.
- You can use the RiskRatio.XML to create reports in Epi Report.

## StatCalc- Estimate Sample Size

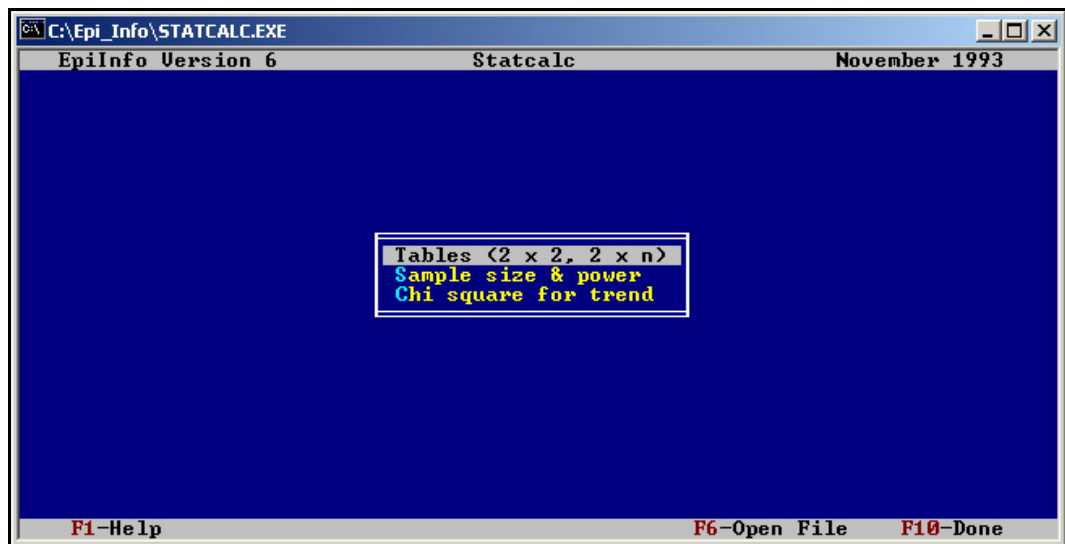
You are interested in further study of the hospitalization rates to be included in your proposal. You need to determine how many records to sample to get the most accurate statistics for the study.

N. From the New York hospital discharge data, how many records do you need to sample in order to get a 95% confidence interval from a sample size of 38,998?

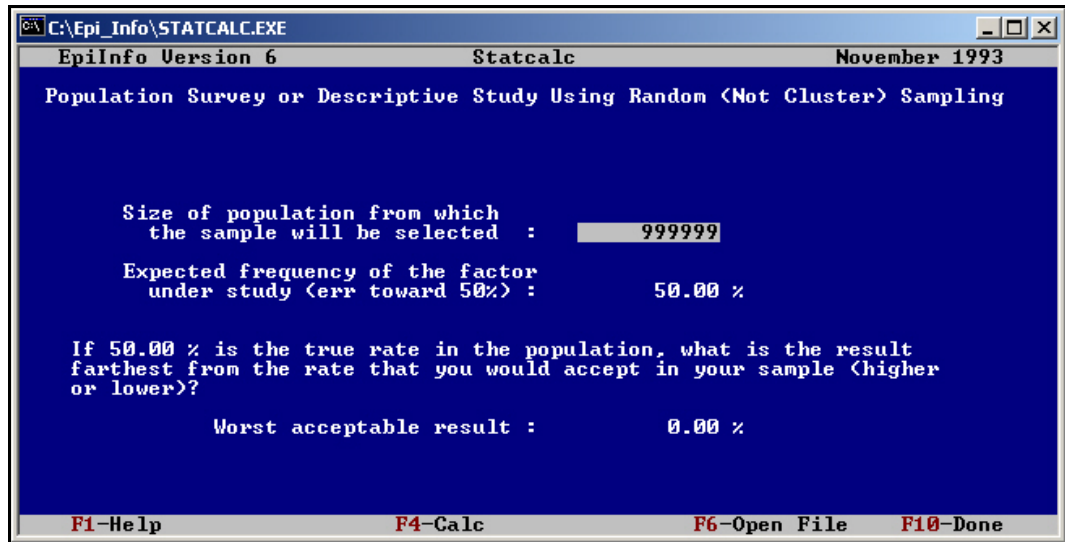
StatCalc is an epidemiologic calculator that produces statistics from summary data. You can use StatCalc to estimate how large a sample needs to be to achieve a 95% confidence interval. Population Survey samples to be taken are simple random or otherwise representative samples. The question being asked must have a yes/no, or other two-choice answer leading to a proportion of the population as the final result.

StatCalc is a DOS-based utility. You will use the keyboard rather than the mouse for navigation.

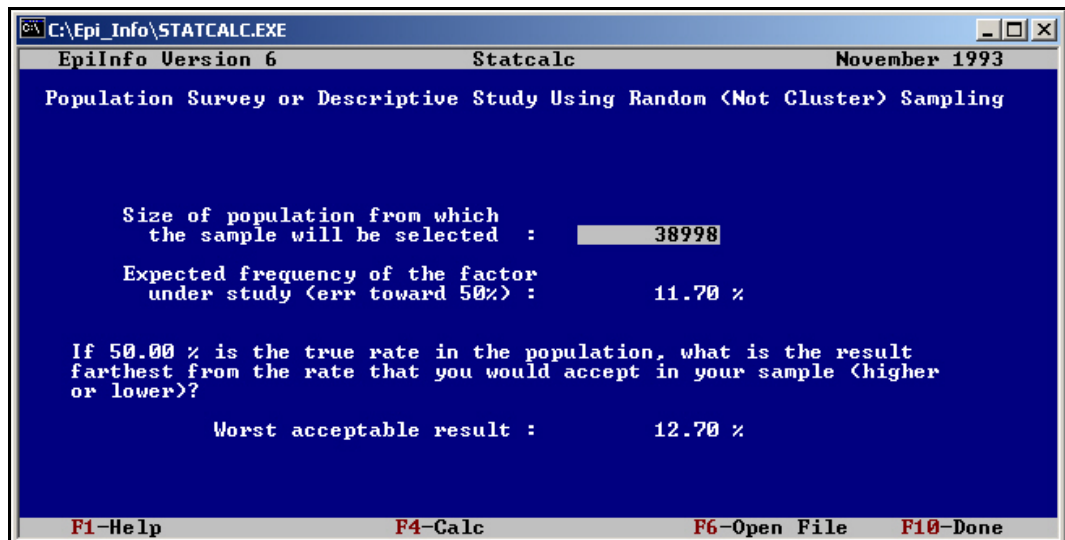
1. From the Epi Info™ main page, select **Utilities>Statcalc**. StatCalc opens.



2. Use the arrow keys on your keyboard to select Sample Size and Power.
3. Press **Enter**.
4. From the pop-up, select Population Survey.
5. Press **Enter**.



6. In the Size of Population field, type the number of NY discharges for Asthma, *38998*. This number came from the summarized hospital discharge records.
7. Press **Enter**.
8. In the Expected Frequency field, type *11.7*. This is the lifetime prevalence rate for NY from the BRFSS data.
9. Press **Enter**.
10. In the Worst Acceptable field, type *12.7*.



11. Press the **F4** key to calculate.

Confidence Level	Sample Size
80 %	1,626
90 %	2,608
95 %	3,602
99 %	5,830
99.9 %	8,693
99.99 %	11,162

12. Look under the Confidence Level row and across to the Sample Size row to get the size of the sample for a 95% confidence interval.
  - Recalculate for any sample size by entering new data and pressing F4.
13. Exit StatCalc by pressing F10 until the window closes.



Lesson Complete!

WHAT YOU LEARNED

---

How to

- Use the RELATE command to link a variety of data tables.
- Create a rate in Epi Info™ using the ASSIGN command.
- Create a graph or a map to illustrate rates.
- Use LIST UPDATE to edit records.
- Use WRITE/APPEND to create new data tables.