

Analysis: Creating Statistics

S U M M A R Y

This lesson is an introduction to creating statistics using commands available in the Analysis module. You will create results using the FREQUENCIES, MEANS, and TABLES commands. Record your results on the Tear Out Answer Sheet provided. Verify your results with the answers listed in *Appendix E Lessons 5 Answer Key*.

In Lesson 4, you used the READ command to view records entered in the survey during Lesson 3. In Lesson 5, you will merge the 12 records from the Asthma Survey 2005.MDB into records gathered and recorded in an MS Excel spreadsheet format. Many times information needed for analyses will come from a variety of sources. In this example, you have student records in an .MDB and in an MS Excel spreadsheet. You will use the MERGE command to place all the student records in one project and then create statistics to see what these records can tell you about asthma among the sampled student body.

Some of the things you are interested in discovering from the data is the frequency of asthma among the students, the worst month for asthma among the students, and what is the average number of missed days of school due to asthma. These types of data will be used to create the report that illustrates the need for asthma initiative funding.

Length of time to complete: 2 hours Beginner

Commands in this Lesson

Brief explanation of the commands used in this lesson and the results produced by using those commands.

Use the Display Command

Use the DISPLAY command to view information about the variables in the project.

Use the Set Command

Use the SET command to generate viewing options for records and statistics.

Use the Merge Command

Use the MERGE command to join information from an MS Excel spreadsheet to an Epi Info™ project file.

Use the Frequency Command

Use the FREQUENCY command to produce results based on a number of selected occurrences. Includes a Practice Lesson.

Skills Review

Use the SELECT command to create a Frequency for a subset of records.

Use the Means Command

Use the MEANS command to define an average.

Skills Review

Use the ASSIGN command to populate a field. Compute the MEANS of the new values.

Use the Tables Command

Use the TABLES command to produce results based on cross-tabulation of variables. Includes a Practice Lesson.

Save a Program

Use the Program Editor to save and run the program created in this lesson.

BEFORE YOU BEGIN

Complete Lesson Four Analysis Basics

WHAT YOU NEED

Asthma Survey 2005.MDB

AsthmaTable.xls

Tear Out Q/A Sheet- Record your answers on this page.

Appendix E Lesson 5 Answer Key- Compare your answers to this page.

FIVE GOALS

- Merge records from an Epi Info™ project file with an MS Excel file.
- Use the FREQUENCY Command to discover which condition has the highest number of affected students.
- Use the MEANS Command to discover which school in the survey had the highest average of missed school days due to breathing problems.

- Use the TABLES Command to discover the odds and risk ratios for having wheezing and asthma.
- Save a program (.PGM) for use in a later lesson.

Tear Out Q/A Sheet

Use this sheet to record your answers in the Answers column as you progress through the Lesson 5. Answers are listed in Appendix E Lesson 5 Answer Key.

A. Questions in Lesson 5 are marked with a letter.

Frequency Sections	Answers
A. Are there more males or females in the survey population?	
B. Which condition has the highest frequency?	
C. Which month had the highest number of students with breathing difficulties?	
D. Which school had the highest number of students with bronchitis?	
E. Which gender has the highest frequency of Asthma?	
F. How many students answered Yes to Reactive Airway Disease?	
G. What is the percentage of students with Reactive Airway Disease that have been prescribed medication?	
H. Which zip code has the most surveyed students living in it?	
I. Of that zip code, which condition had the highest frequency?	
J. How many students answered Yes to all four conditions: Asthma, Reactive Airway Disease, Bronchitis, or Wheezing.	
K. Based on all four conditions and zip code, are the conditions evenly distributed among the survey population?	
L. Does one school have more students in affected zip codes and with all four conditions than the others do?	

Means Sections	Answers
M. What was the most frequent number of missed days?	
N. What is the mean (average) number of days missed due to asthmatic conditions?	
O. Which school had the highest average of missed days?	
P. Was the school with the highest average of missed days for	

males the same as the school with the highest average of missed days for females?	
Q. Is the average number of missed days higher among those who answered Yes to asthma than the overall average? Refer to Question N.	
R. What is the average age of the student population?	

Tables Section	Answers
S. Does one school have a higher number of students with asthma?	
T. Which school has the highest number of students with wheezing?	
U. How many students had both asthma and wheezing?	
V. What are the odds that you have the condition wheezing if you have the condition asthma?	
W. Is the Odds Ratio for Bronchitis and Reactive Airway Disease lower or higher than Asthma and Wheezing? Refer to Question V for the Odds Ratio for Asthma and Wheezing.	
X. How many children with the condition wheezing were prescribed medication? From those results, what is the Risk Ratio?	
Y. Is the Risk Ratio for being prescribed medication for the condition asthma higher or lower than being prescribed medication for wheezing? Compare the results to Question X.	
Z. Are the risk ratios for Asthma and Wheezing higher for children under 10 than they are for the survey population as a whole? Compare to the results from Question Y.	

Commands in this Lesson

The following commands are used in this lesson.

DISPLAY

Use this command to display table, view, and database information. Use the display option Variables Currently Available to see all the variables in the dataset, including names, field types, and format information. Use Display, prior to merging or creating statistics, to ensure that field types and variables names have been coded as needed. Located in the Variables folder.

SET

Use this command to view the current values of Analysis option settings and generate commands to change them. Statistical and graphic viewing options can be selected. Yes, No, and Missing values can be viewed in alternate forms. Allows the inclusion or exclusion of missing records in statistical computations. Located in the Options folder.

MERGE

Use this command to merge records in one dataset with those in another, using one or more defined identifiers or build keys to establish the match between records. Merged records can be appended or updated to the new dataset. Merge is only supported when the READ data source is an Epi Info™ or MS Access table. The MERGE data source can be of any type. Located in the Data folder.

FREQUENCIES

Use this command to produce a table showing how data is distributed, counts of any given interval, how many records have each value of a selected variable, or the number of occurrences for a specified variable. If more than one variable is given, a separate table is created for each variable. Confidence limits for each proportion are included. Records may be included or excluded from the count by using SELECT statements. Located in the Statistics folder.

MEANS

Use this command to compute descriptive statistics for a continuous numeric variable. The mean of a set of data is equal to the sum of the data divided by the number of items in the data set. When used with a cross-tabulation variable, it also computes statistics showing the likelihood that the means of the groups are equal. The mean of a yes/no variable is the proportion of respondents answering yes. Means can be created for continuous or categorical values. Continuous variables fall along a continuum with one variable selected, for example, age. Categorical variables are groupings or categories with two variables selected, for example, gender. If one variable is selected, the program produces a table similar to the Frequencies table plus descriptive statistics. If two variables are selected, the first is a numeric variable containing data to be analyzed and the second is a variable that indicates how the group will

be distinguished. The output of this format is a table similar to that produced by Tables, plus descriptive statistics of the numeric variable for each value of the group variable. Means produces the following statistical tests: parametric tests, Analysis of Variance or ANOVA (for two or more samples), student's t-test (for two samples), non-parametric tests, Bartlett's Test for Inequality of Population Variances, Kruskal-Wallis one-way analysis of variance (for two or more samples), and Mann-Whitney/Wilcoxon Two Sample Test. Located in the Statistics folder.

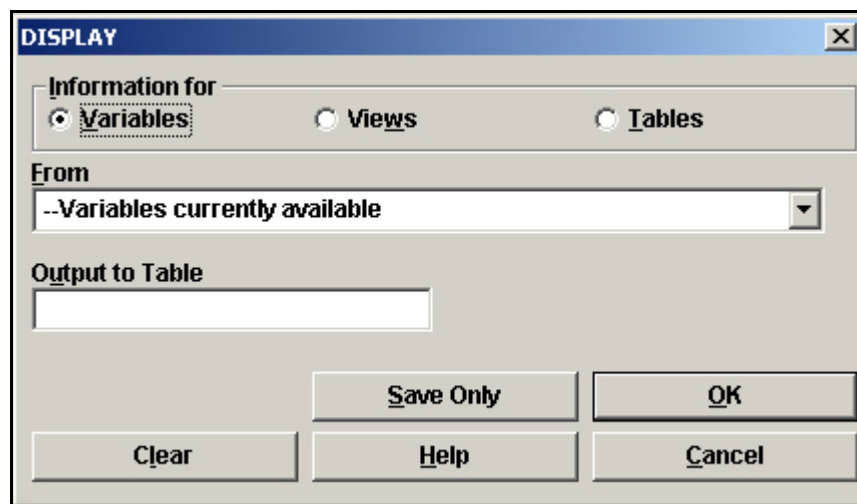
TABLES

Use this command to create cross-tabulations of categorical variables. Tables can help determine the probability that a risk factor is linked to an outcome. For these values to have their accepted epidemiological meanings, the value representing presence of the exposure (independent value) and outcome conditions (dependent variable) must appear in the first row and column of the table. Epi Info yes/no variables are automatically sorted. Values of the first selected variable will appear across the top of the table, and those of the second selected variable will be on the left margin of the table. Normally cells contain counts of records matching the values in the corresponding marginal labels. For 2x2 tables, the command produces odds ratios and risk ratios. For tables other than 2x2, Chi-Squared statistics are computed. Chi square for trend tests for the presence of a trend in dose response or other case control studies where a series of increasing or decreasing exposures is being studied. The p-value is the probability that the association between two variables may be due to chance. A low p-value of $<.05$ means the risk factor may be closely associated with the illness. Located in the Statistics folder.

Display Files

This command allows the display of table, view, and database information. You are going to use it to view information about the variables in your current dataset and to determine which variable to use as the Build Key when you merge your student records with the student records from MS Excel.

1. Read/Import the Asthma Survey 2005 file located in your EIHA Tutorial folder. You should still have 12 records. Remember that your Project and Data Source need to match when you read in a file.
2. From the Command Tree window, click **Display** from the Variables folder. The Display dialog box opens. You want to see Information for Variables From the Variables Currently Available.

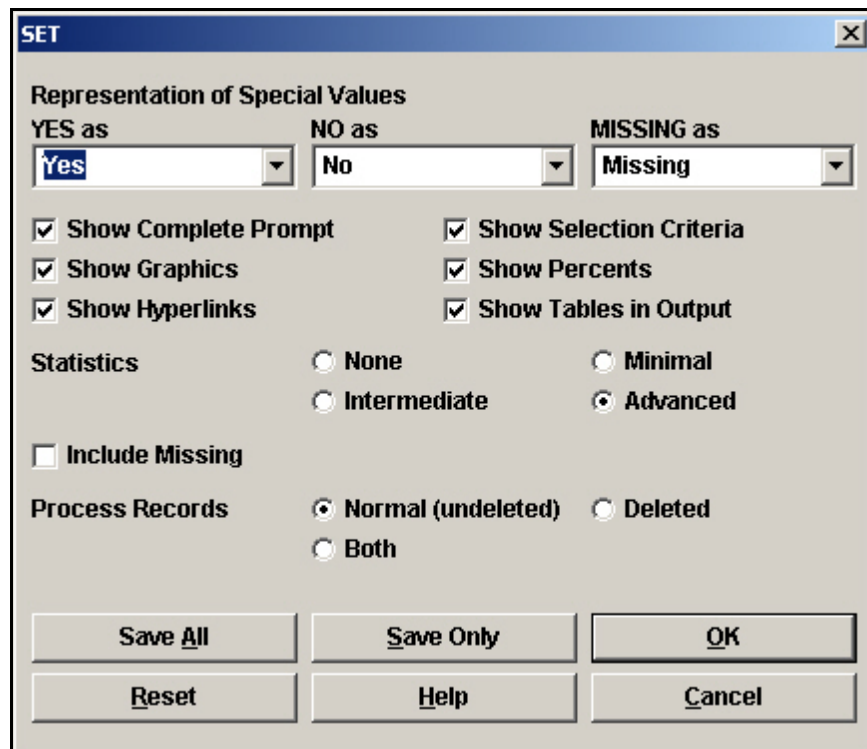


3. Click **OK** to accept the default settings. The Variables in the Asthma Survey 2005 project appear in the Output window.
4. Scroll through the variables listed.
 - Display is an important command to use before importing data or completing analyses on unfamiliar data. If needed, you could use this information to format the cells in your spreadsheet prior to merging or importing files. Display can alert you to potential errors in Analysis, for example, assuring that variables are coded as the same field type.
 - To see how the MS Excel file you are merging into the project is set up, you can open the AsthmaTable file in MS Excel and view the data. Do not make any changes to the file. Notice that the MS Excel and the Epi Info™ project file both contain variables called StudentID. This field corresponds to both sets of data and will be used to join the datasets. The spreadsheet does not contain all the personal information collected in your Epi Info™ project. Values that do not correspond will appear as missing in Analysis. Be sure to close MS Excel before beginning the merge.

Set Options

The SET command provides various options that affect the performance and output of data in Analysis. These settings are used whenever the module runs. You want to be sure that the Statistics setting is Advanced or the Risk and Odds Ratios may not appear in your results.

1. From the Command Tree Options folder, click **Set**. The SET dialog box opens.

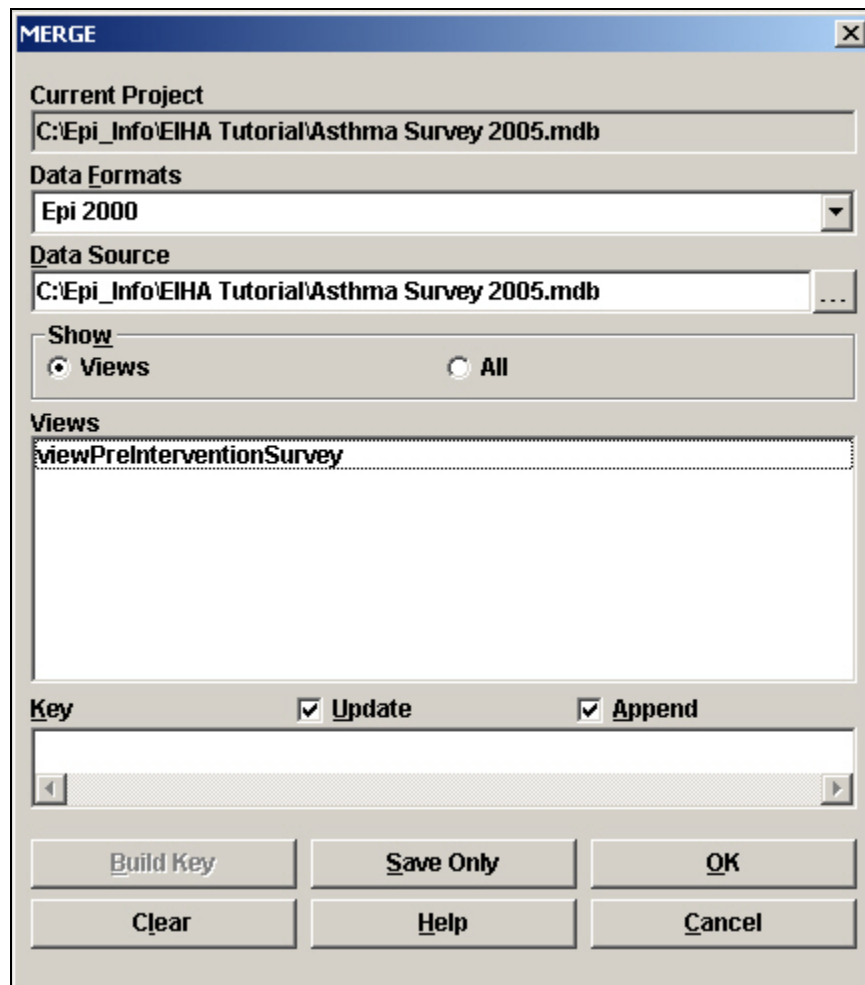



2. Set your options as follows.
 - Yes as Yes
 - No as No
 - Missing as Missing
 - All Show boxes selected
 - Statistics Advanced
 - Include Missing **not** selected
 - Process Records as Normal (undeleted)
3. Click **OK**.

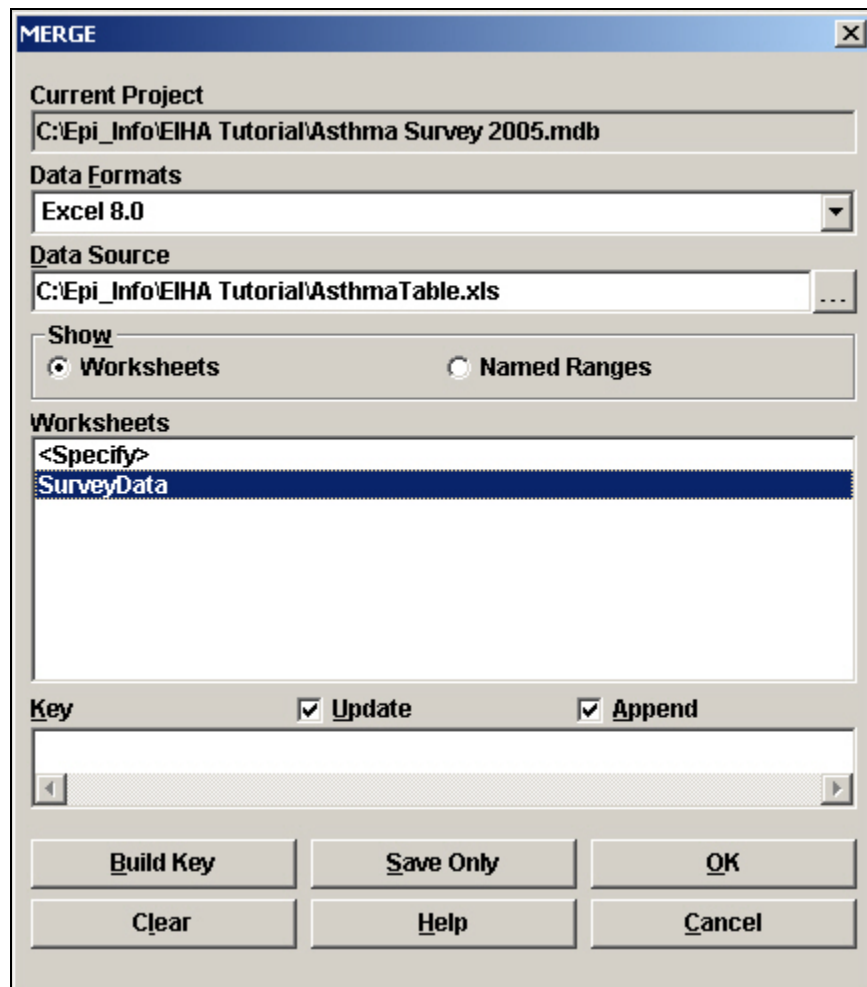
Merge Files

The current project contains 12 student records. In order to view all the available records for the county, use the MERGE command to join the 12 records you created in Enter Data with a larger data table gathered using MS Excel. Merge is only supported when the READ data source is an Epi Info™ or MS Access table. The Merge data source can be of any type, for this example, you will use the MS Excel spreadsheet called AsthmaTable.xls located in the EIHA Tutorial folder. A merge requires a Build Key that represents a matching variable inside both sets of data. The MS Excel file used for this lesson has been formatted so that the column headings and data types match the field names you created in Make View. You are going to merge the files based on the Build Key StudentID. When you created the View, you made StudentID a Required field. In our example, there are no duplicate values on the field StudentID. Each record will have a unique value.

- From the Command Tree window Data folder, click **Merge**. The MERGE dialog box opens. Notice that the MERGE dialog box is very similar to the READ dialog box.
 - Be sure your current project is C:\Epi_Info\EIHA Tutorial\Asthma Survey 2005.mdb.

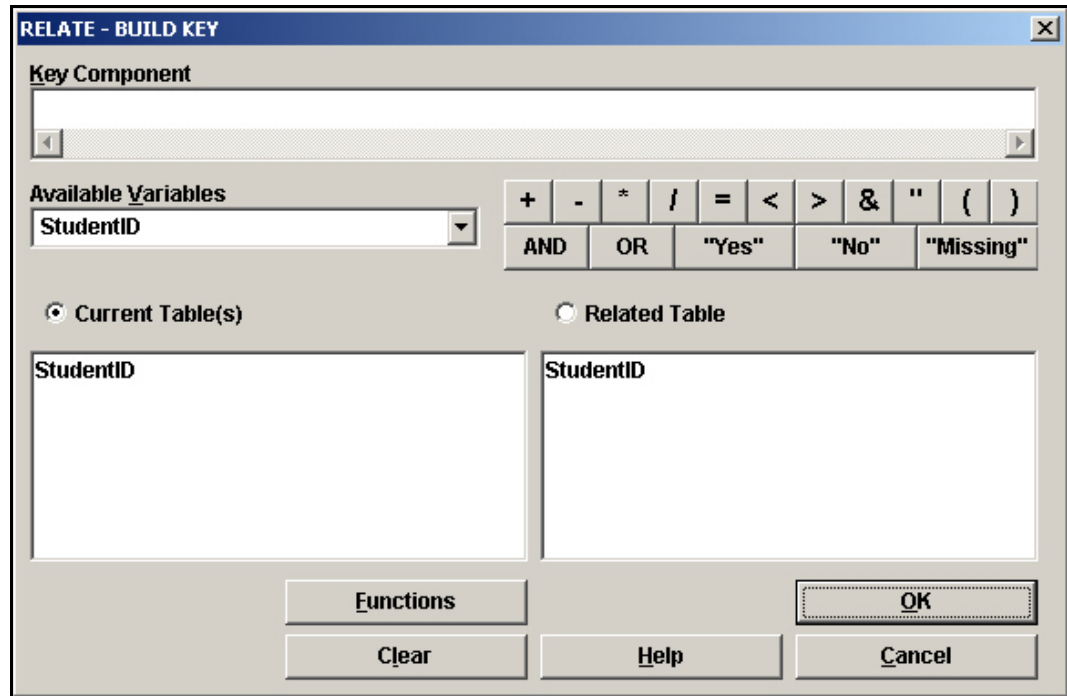


2. From the Data Formats drop-down, select Excel 8.0. The Data Source field clears.
3. Click the Data Source **Browse**  button. The Merge File Name window opens.
4. Select the file called AsthmaTable.xls located in the EIHA Tutorial folder.
5. Click **Open**. The MERGE dialog box populates with information about the spreadsheet.
6. Click the Worksheet called **SurveyData**. The Build Key button activates.
 - SurveyData is the named worksheet in the MS Excel file.



7. Click **Build Key**. The RELATE-BUILD KEY dialog box opens.
 - Keep in mind that you need a variable that can be joined from the Current Table that corresponds to the Related Table.

8. From the Available Variables drop-down, select StudentID.
9. Click **OK**. The Current Table field shows StudentID.
10. From the Available Variables drop-down select, StudentID.
11. Click **OK**. The Related Tables field shows StudentID.



12. Click **OK** to accept the Build Key designations. The MERGE window opens.
 - The Update and the Append checkboxes should be selected. For matching records, Update replaces the value of any field in the READ table whose build key matches the MERGE table. For unmatched records, Append creates a new record in the READ table with values only for those fields that exist in the MERGE table.
 - For this merge, you want to Update and Append records, meaning that records containing the same StudentIDs will be updated (overwritten) if there is new information, and all other records will be appended (added) to the end of the data table.
13. Click **OK** from the MERGE dialog box. The FILESPEC dialog box opens.
 - Be sure the First Row Contains Field Names checkbox is selected.
14. Click **OK**. The READ dialog box opens.
 - Because you are bringing a new data source into the project, you must decide if you want to create a temporary or permanent link between the two sets of

data. For this exercise, you will be working with the 800 records for several exercises so make the link permanent.

15. In the READ dialog box, type the link name *AsthmaTable*.
16. Click **OK**. The Record Count in the Output window reads 800.
 - **LIST** the variables in a Grid.
 - Notice the first twelve records contain information from your survey. The StudentIDs are the same, so the records were updated to reflect any changes.
 - Use the Close X to exit the grid listing and return to the Output window.
 - Notice that the Program Editor shows the commands used to create the Merge and the List.

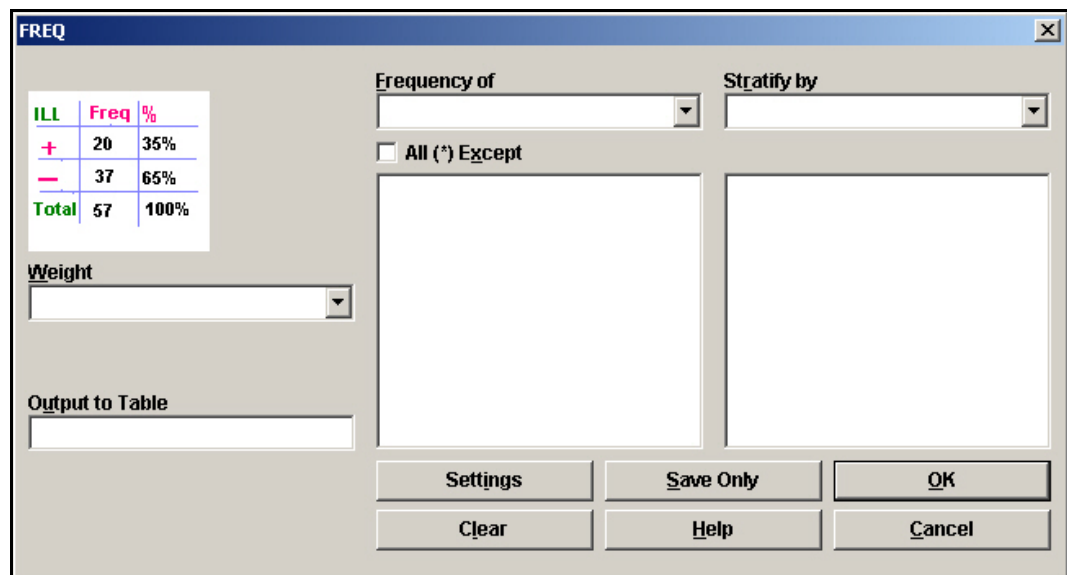
Create a Frequency

You are going to use the FREQUENCY command to get a count and view the 95% Confidence Limits on selected variables. The FREQ command produces a frequency table that shows how many records have a value for each variable, the percentage of the total, and a cumulative percentage.

The command FREQ * creates a table for each variable in the current view other than unique identifiers. This command is often used to begin analyses on a new data set. To exclude variables, use the All (*) Except checkbox and select variables to be excluded.

Since you just merged in new data, you want to develop some basic information and then build up to more complex statistics.


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



2. From the Frequency of drop-down, select SchoolNum.
3. Click **OK**. The Frequency distribution of SchoolNum appears in the Output window.
 - The Frequency column gives you the count of students per school. The Percent column indicates the school's percentage of the total number.
 - The 95% Confidence Limits are a range of values for a variable that indicates the likely location of the true value of a measure. In this frequency, the Confidence Limits for School A are 21.6%-27.7% and the actual percentage is 24.5%, which falls within the range.

[Forward](#)

School Number	Frequency	Percent	Cum Percent	
A	196	24.5%	24.5%	■
B	205	25.6%	50.1%	■
C	208	26.0%	76.1%	■
D	191	23.9%	100.0%	■
Total	800	100.0%	100.0%	■

 **Practice**

Answers to questions throughout Lesson 5 can be entered on the Tear Out Q/A Sheet provided at the beginning of Lesson 5. Each of the variables used in the following frequencies were created in the original survey in Make View.

A. Are there more males or females in the survey population?

Hint: Create a Frequency of Gender to find out.

B. Which condition has the highest frequency?

Hint: Create a Frequency of Asthmatic Conditions. Notice you get four tables, one for each condition.

C. Which month had the highest number of students with breathing difficulties?

Hint: Create a Frequency of WMonth.

Stratify a Frequency

Frequency tables can be stratified or grouped by a selected variable.

D. Which school had the highest frequency of students with bronchitis?

1. Click **Frequencies**. The **FREQ** dialog box opens.
2. From the Frequency of drop-down menu, select Bronchitis.
3. From the Stratify By drop-down menu, select SchoolNum.
4. Click **OK**. You will get four tables containing the frequency of bronchitis per school.

Practice

E. Which gender has the highest frequency of Asthma?

Hint: Create a Frequency of Asthma stratified by Gender.

Select Variables for a Frequency

You can use the SELECT Command to specify a subset of students and then create a frequency for only those records. You are interested in discovering how many students have Reactive Airway Disease and if they are currently taking medication for the condition.

- F. How many students answered Yes to having Reactive Airway Disease?
- G. What is the percentage of students with Reactive Airway Disease that have been prescribed medication?

1. Use the SELECT command to view those who answered Yes to having Reactive Airway Disease. (Hint: RAD="Yes")
2. Create a Frequency of students who have been prescribed medication (Hint: Frequency of Medication)
3. Click **Cancel Select** to return to the original 800 records.

Practice

- H. Which zip code has the most surveyed students living in it?

Hint: Create a frequency of zip code (variable Zip).

- I. Of that zip code, which condition had the highest frequency?

Hint: Use SELECT to view only those students living in zip code area 12046. (Zip Code is a text field. Text fields must be surrounded by quotes when used in commands Zip="12046")

Create a frequency of Asthmatic Conditions.

Remember to Cancel Select.

Skills Review

Record the following results on the Tear Out Q/A Sheet.

- J. How many students answered Yes to all four conditions: Asthma, Reactive Airway Disease, Bronchitis, or Wheezing.
- K. Based on all four conditions and zip code, are the conditions evenly distributed among the survey population?
- L. Does one school have more students in affected zip codes and with all four conditions than the others do?

You want to find out how many children from the survey responded Yes to all the asthmatic conditions. Then you want to find out if the affected students are concentrated to one zip code or one school.

Select the students who had all four conditions.

Hint: Use the SELECT command to create the code below. You need to select the students who answered Yes to Asthma and Yes to Bronchitis and Yes to RAD and Yes to Wheezing.

```
SELECT Asthma= (+) AND Bronchitis= (+) AND RAD= (+) AND Wheezing= (+)
```

Create a Frequency of Zip Code.

Create a Frequency of SchoolNum.

Cancel Select to return to the original 800 records.

Create a Means

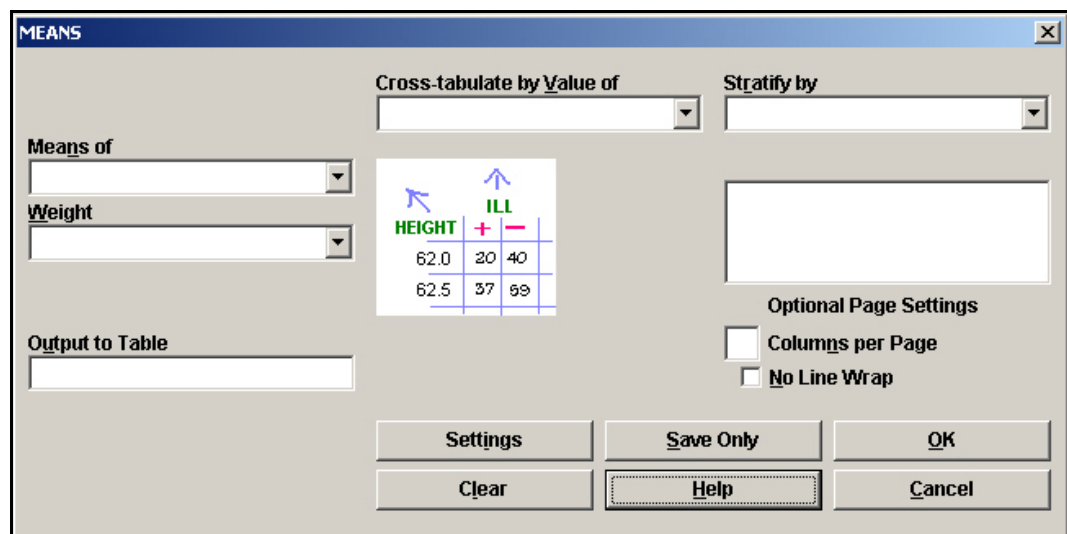
- M. What was the most frequent number of missed days?
- N. What is the Mean (average) number of days missed due to asthmatic conditions?

One of the items you want to include in the funding request is the simple means of the number of days children missed school due to asthma or an asthma related condition. The MEANS command will give you the frequency and the means results.

The MEANS command can be used to get an average for a continuous variable. The mean of a yes-no variable is the proportion of respondents answering yes. The MEANS command has two formats. If only one variable is supplied, the program produces a table like that produced by FREQUENCIES, plus descriptive statistics.

If two variables are supplied, the first is a numeric variable containing data to be analyzed and the second is a variable that indicates how groups will be distinguished. The output of this format is a table like that produced by TABLES, plus descriptive statistics of the numeric variable for each value of the group variable.

1. From the Command Tree Statistics folder, click **Means**. The MEANS dialog box opens.



2. From the Means Of drop-down menu, select the variable MissDays.
3. Click **OK**. The Output window populates with the results. The table results show a breakdown of missed days aggregated per answer.

How many days of school has your child missed due to asthma?	Frequency	Percent	Cum Percent	
0	8	1.0%	1.0%	
1	241	30.1%	31.1%	■
2	233	29.1%	60.3%	■
3	59	7.4%	67.6%	■
4	65	8.1%	75.8%	■
5	8	1.0%	76.8%	
6	17	2.1%	78.9%	
7	23	2.9%	81.8%	
8	29	3.6%	85.4%	
9	24	3.0%	88.4%	

4. Scroll down to see the Mean and Standard Deviation results.

Obs	Total	Mean	Variance	Std Dev
800	3014.0000	3.7675	13.5379	3.6794
Minimum	25%	Median	75%	Maximum Mode
0.0000	1.0000	2.0000	4.0000	15.0000 1.0000

- Variance- A measure of the dispersion shown by a set of observations.
- Standard deviation- A statistical summary of how dispersed the values of a variable are around its mean.
- Standard error (of the mean) - The standard deviation of a theoretical distribution of sample means of a variable around the true population mean of that variable.
- See Appendix H: Glossary for more definitions, or visit the source at (<http://www.cdc.gov/excite/library/glossary.htm>).

Use the Means Command with a Cross-Tabulated Variable

O. Which school had the highest average of missed days?

To determine which school had the highest average of missed days, you need to compute the means of missed days by school. Use a cross tabulation variable to compute statistics showing the likelihood that the means of the groups are equal.

1. Click **Means**. The MEANS dialog box opens.
2. Select the Means Of Variable MissDays.
3. Select the Cross-Tabulate by Value Of Variable SchoolNum.
4. Click **OK**. Results appear in the Output window.
5. Scroll to the Descriptive Statistics. Notice the Mean is computed by School.

Descriptive Statistics for Each Value of Crosstab Variable

	Obs	Total	Mean	Variance	Std Dev	
A	196	815.0000	4.1582	14.5543	3.8150	
B	205	813.0000	3.9659	13.8371	3.7198	
C	208	786.0000	3.7788	13.6610	3.6961	
D	191	600.0000	3.1414	11.6589	3.4145	
	Minimum	25%	Median	75%	Maximum	Mode
A	0.0000	1.0000	2.0000	7.0000	15.0000	1.0000
B	0.0000	1.0000	2.0000	6.0000	15.0000	1.0000
C	0.0000	1.0000	2.0000	4.0000	15.0000	2.0000
D	1.0000	1.0000	2.0000	3.0000	15.0000	1.0000

Create a Stratified Means

P. Was the school with the highest average of missed days for males the same as the school with the highest average of missed days for females?

Use the Means command with Stratify By to group your data by gender.

1. Click **Means**. The MEANS dialog box opens.
2. Select the Means Of Variable MissDays.
3. Select the Cross-Tabulate by Value Of Variable SchoolNum.
4. Select the Stratify By Variable Gender.
5. Click **OK**. Results appear in the Output window.
 - Two tables and two sets of descriptive statistics are created, one for female and one for male.
6. Scroll through the results.

Use Select with the Means Command

Q. Is the average number of missed days higher among those who answered Yes to asthma than the overall average? Refer to Question N.

You can use the **SELECT** Command to create subsets of data and then use the **MEANS** command to create statistics.

1. Click **Select**. The **SELECT** dialog box opens. Use the **SELECT** Command to display only those students who answered Yes to Asthma.
2. From the Available Variables drop-down, select Asthma.
3. Use the Functions and Operators buttons to create the code Asthma=(+) in the Select Criteria field.
4. Click **OK**. You should have 91 records.
5. Create a Means of Missdays.
6. Click **Cancel Select** to return to the original 800 records.

 Skills Review

R. What is the average age of the student population?

The age field for the imported data has not been populated. You need to assign a code to populate this field. The variable Age did not exist in the MS Excel data that you merged. You do have a date of birth; however, so you can compute the age by using the YEARS function as you did in the Check Code lesson.

From the Command Tree Variables folder, click **Assign**. The ASSIGN dialog box opens.

From the Assign Variables drop-down, select Age.

In the =Expression field type your code. This is the same code you used to create the age field Check Code in Make View.

Use the Years function and then create your mathematical code: Survey date minus date of birth variable.

```
Age= YEARS(DOB, 10/25/2005)
```

Click **OK**. The code appears in the Program Editor.

List the Age variable in a grid.

All the once missing fields should now contain ages.

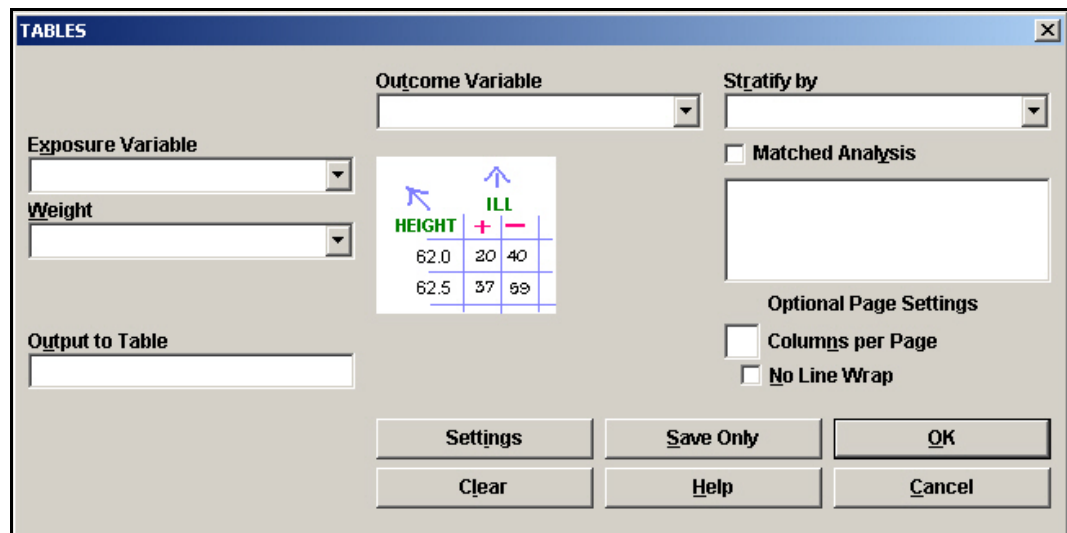
Create a Means of Age.

Create a Table

- S. Does one school have a higher number of students with asthma?
- T. Which school has the highest number of students with Wheezing?

You are going to use the Tables command to examine the relationship between two or more categorical values. You want to further explore the idea that there may be a connection between school location and asthmatic conditions. You are also interested in seeing what kinds of results you can get with the 800 student records available to you as a primary data source.

1. From the Command Tree Statistics folder, click **Tables**. The Tables dialog box opens.



2. From the Exposure Variable drop-down, select SchoolNum.
3. From the Outcome Variable drop-down, select Asthma.
4. Click **OK**. Results appear in the Output window.

ASTHMA			
School ID #	Yes	No	TOTAL
A	29	167	196
Row %	14.8	85.2	100.0
Col %	31.9	23.6	24.5
B	25	180	205
Row %	12.2	87.8	100.0
Col %	27.5	25.4	25.6
C	21	187	208
Row %	10.1	89.9	100.0
Col %	23.1	26.4	26.0
D	16	175	191
Row %	8.4	91.6	100.0
Col %	17.6	24.7	23.9
TOTAL	91	709	800
Row %	11.4	88.6	100.0
Col %	100.0	100.0	100.0

5. Create the same table, but select Wheezing as the Outcome Variable.

Create a 2x2 Table

- U. How many students had both Asthma and Wheezing?
- V. What are the odds that you have the condition Wheezing if you have the condition Asthma?

For 2x2 tables, the Tables command produces odds ratios and risk ratios. You want to determine if there a correlation between the children who have asthma and those with wheezing.

1. Click **Tables**. The TABLES dialog box opens.
2. Select the Exposure Variable of Asthma.
3. Select the Outcome Variable of Wheezing.
4. Click **OK**.

Forward

WHEEZING

Asthma	Yes	No	TOTAL
Yes	63	28	91
Row %	69.2	30.8	100.0
Col %	63.6	4.0	11.4
No	36	673	709
Row %	5.1	94.9	100.0
Col %	36.4	96.0	88.6
TOTAL	99	701	800
Row %	12.4	87.6	100.0
Col %	100.0	100.0	100.0

- Notice the graphic representation of your results.
 - A 2x2 table was created because each variable had a yes or no answer.
 - The largest square represents the children who answered no to both wheezing and asthma.
5. Scroll down to view the Single Table Analysis.

Single Table Analysis				
	Point Estimate	95% Confidence Interval		
		Lower	Upper	
PARAMETERS: Odds-based				
Odds Ratio (cross product)	42.0625	24.0911	73.4400	(T)
Odds Ratio (MLE)	41.5616	23.9981	73.5508	(M)
		23.2358	76.3616	(F)
PARAMETERS: Risk-based				
Risk Ratio (RR)	13.6346	9.6419	19.2806	(T)
Risk Difference (RD%)	64.1532	54.5335	73.7729	(T)
(T=Taylor series; C=Cornfield; M=Mid-P; F=Fisher Exact)				
STATISTICAL TESTS		Chi-square 1-tailed p	2-tailed p	
Chi square - uncorrected		306.0983	0.0000000000	
Chi square - Mantel-Haenszel		305.7157	0.0000000000	
Chi square - corrected (Yates)		300.2107	0.0000000000	
Mid-p exact			0.0000000000	
Fisher exact			0.0000000000	

- Odds Ratio- A measure of association that quantifies the relationship between an exposure and health outcome from a comparative study; also known as the cross-product ratio.
- Risk Ratio- A comparison of the risk of some health-related event such as disease or death in two groups.
- Chi Square- A test of statistical significance that is used to determine how likely it is that an observed association between an exposure and a disease could have occurred due to chance alone, if the exposure was not actually related to the disease.
- See the Appendix H: Glossary or Appendix I: Data Sources for definitions and links to more information on interpreting statistics.

Skills Review

Complete the following:

W. Is the Odds Ratio for Bronchitis and Reactive Airway Disease lower or higher than Asthma and Wheezing? Refer to Question V for the Odds Ratio for Asthma and Wheezing.

Hint: Create a 2x2 table for Bronchitis Exposure and RAD Outcome.

X. How many children with the condition wheezing were prescribed medication? From those results, what is the Risk Ratio?

Hint: Create a 2x2 table for Medication Exposure and Wheezing Outcome.

Y. Is the Risk Ratio for being prescribed medication for the condition asthma higher or lower than being prescribed medication for wheezing? Compare the results to Question X.

Hint: Create a 2x2 table for Medication Exposure and Asthma Outcome.

Z. Are the risk ratios for Asthma and Wheezing higher for children under 10 than they are for the survey population as a whole? Compare to the results from Question V.

Select the children who are less than age 10. (Hint: Select Age<10)

Be sure that the Age field is calculated for all records. If you have use the READ command to bring in a new file or program or closed Analysis, then you will have to calculate the Age field as you did for the MEANS exercise.

Create a table for Asthma Exposure and Wheezing Outcome.

Click **Cancel Select**.

Saving a Program

Notice the large amount of code in the Program Editor. In order to re-create the same statistics after data has been added, you want to save the program or code. By doing so, you can run these same statistics on the new records.

Analysis automatically saves the last program created in the program editor in a file called LastPGM. This file is created and saved in the Epi_Info folder and gets overwritten each time a new PGM is created and the application is closed. You want to create your own program based on your data.

1. From the Program Editor, click **File>Save**. The Save Program dialog box opens.

- The Project File field contains the information about where you will be saving your file. You want to leave this information set to the EIHA Tutorial folder.
2. In the Program Field, type *Lesson 5*.
 3. In the Author field, type your initials.
 4. In Comments, type EIHA.
 5. Click **OK**.
 - The date created and date updated fields will be populated with new information each time you open the program.
 6. Click **Exit** to close Analysis.

Test the Saved Program

1. Open Analysis.
2. From the Program Editor, click **File>Open**. The READ dialog box opens.
3. From the Program drop-down, select Lesson 5.
4. Click **OK**. Your saved program appears in the Program Editor.
5. Select **Edit>Program Beginning**.
6. Click **Run This Command**. The first command inside your program runs. The first command was Read, so your 800 records appear. If you had updated records, the new information would be reflected in the results.
 - You can run any previous commands by placing your cursor in the command and clicking **Run This Command**.
7. Click **Run** to run the entire program.
 - The program you saved is part of your project, to keep it updated you must save it. To access the LASTPGM file, click the **Text File** button from the Save Program dialog box and locate the file in the Epi_Info folder.
 - Do not close Analysis if you plan to complete Lesson 6 at this time.

Lesson Complete!

WHAT YOU LEARNED

How to

- Use the DISPLAY command to view variable types in your project.
- Merge an MS Excel file into an Epi Info™ project.
- Use the FREQ command to see how your data is distributed.
- Use the MEANS command to see if one set of data is equal to the sum of the data divided by the number of items in the data set.
- Use the TABLES command to see if risk factors are linked to outcomes.
- Save Program (.PGM) files for use with later data tables or updated data.