Using the Online SDA Tabulator with IHIS

Before you begin to use the Survey Data Analyzer (SDA), an online data tabulator, we recommend that you open a separate browser and pull up the variables page on the Integrated Health Interview Series (IHIS) website (www.ihis.us) to access the appropriate documentation for each variable of interest. Most years of the survey have at least several hundred variables, and the website has tools to help you locate the variables you want. Variable descriptions on the website report codes and frequencies (useful when recoding data), specify the appropriate weight in each year, and discuss changes that limit comparability of the variable over time. For example, the variable HEALTH, which we will use below, has gone from a 4-point scale to a 5-point scale over time. Changes like this are described in detail in the variable descriptions on the IHIS website.

There are two main tasks for which SDA is useful: 1) calculating frequencies and cross-tabulations and 2) comparing means. These tasks are discussed, in turn, below. Researchers who wish to use IHIS data for more advanced analyses (e.g., regressions) should download a data extract from the IHIS website and conduct their analysis with a statistical package such as SAS, SPSS, or Stata, rather than with SDA.

Frequencies / Cross-tabulation Program

First, select the year(s) in which you are interested from the SDA-IHIS interface (http://www.ihis.us/ihis/sda.shtml). Let's use the 2009 sample.

Next, choose the variables you would like to analyze. To generate frequencies or cross-tabulations, you can either type the variable names into the Row and Column boxes pictured below ...

SDA Frequencies/Crosstabulation Program
Help: **General** / **Recoding Variables**

REQUIRED Variable names to specify
Row:

OPTIONAL Variable names to specify
Column:
Control:
Selection Filter(s):   *Example: age(18-50)*
Weight: None

TABLE OPTIONS

Percentaging:
☑ Column ☐ Row ☐ Total
☐ **Confidence intervals** Level: 95 percent
☐ **Standard error of each percent**
Sample design: ⦿ Complex ○ SRS

N of cases to display:
☐ Unweighted ☑ Weighted

☐ **Summary statistics**
☐ **Question text** ☐ **Suppress table**
☑ **Color coding** ☐ **Show Z-statistic**
☐ **Include missing-data values**

CHART OPTIONS

Type of chart: Stacked Bar Chart
Bar chart options:
Orientation: ⦿ Vertical ○ Horizontal
Visual Effects: ⦿ 2-D ○ 3-D

Show percents: ☐ Yes
Palette: ⦿ Color ○ Grayscale
Size - width: 600 ▾ height: 400 ▾

Title:

[Run the Table] [Clear Fields]

... or use the variable dictionary on the left side of the screen. (Given the number of variables in most years of IHIS, this second option is probably more time consuming and less helpful.)

The variable dictionary is generated based on the sample(s) you have selected. The variables available in each sample are organized into thematic groups by record type (household, person). To select a variable to analyze using the dictionary, click on the variable name.

After you click on the variable name, the "selected box" above will be populated with the variable you have chosen. Now use the "Copy to" buttons below your variable name to enter your variable into the field of your choice. (If you entered the variable names directly, after finding the variables you want to use on the IHIS website, you will go directly to the step directly below this text.)

The two most important fields in the frequencies/cross-tabulation program are "Row" and "Column." With variables entered into these fields, the system will produce a basic cross tabulation, or table. By default, the SDA will calculate column percentages so that the values in each column sum to 100.

We will walk through a brief example of some considerations that should be made to obtain a table that presents data in a meaningful way. First we will show you some points to consider in selecting the row, column, and other characteristics. Let's say you want a table that presents the relationship between age and health status. To create this table, enter "age" as the row variable and "health" as the column variable. The input is pictured below.

To obtain output, click "Run the Table." The beginning of the output looks like this:

Variables					
Role	Name	Label	Range	MD	Dataset
Row	**age**	Age	0-85		1
Column	**health**	Health status	1-9		1

Frequency Distribution								
Cells contain: -Column percent -N of cases	health							
	1 Excellent	2 Very Good	3 Good	4 Fair	5 Poor	7 Unknown- refused	9 Unknown-don't know	*ROW* *TOTAL*
0: 0	2.4 745	1.1 291	.8 170	.1 7	.0 0	1.8 1	2.1 1	*1.4* *1,215*
1: 1	2.5 785	1.1 299	.9 198	.2 17	.1 2	.0 0	.0 0	*1.5* *1,301*
2: 2	2.4 744	1.3 334	.9 189	.2 15	.1 3	.0 0	.0 0	*1.5* *1,285*
3: 3	2.4 753	1.3 344	.9 195	.3 18	.2 4	.0 0	.0 0	*1.5* *1,314*
4: 4	2.2 698	1.3 336	1.0 213	.3 21	.1 3	.0 0	.0 0	*1.4* *1,271*
5: 5	2.5 774	1.4 359	1.0 222	.3 23	.2 5	.0 0	.0 0	*1.6* *1,383*
6: 6	2.2 692	1.3 347	1.1 231	.2 11	.0 1	.0 0	.0 0	*1.4* *1,282*
7: 7	2.3 718	1.4 374	1.0 228	.4 26	.2 5	.0 0	.0 0	*1.5* *1,351*
8: 8	2.2 679	1.4 365	1.0 222	.4 28	.1 3	.0 0	2.1 1	*1.5* *1,298*
9: 9	2.4 749	1.5 392	1.2 257	.4 26	.1 3	.0 0	.0 0	*1.6* *1,427*
10: 10	2.2 681	1.3 346	.9 207	.4 29	.1 3	.0 0	.0 0	*1.4* *1,266*

The resulting table is potentially problematic for four reasons. First, the table is large and therefore does not provide an overall sense of the relationship (if any) between age and health status. This is because the age variable has 86 valid values, resulting in 86 different rows in the table. The partial output above pictures only the first 11 rows of the age variable.

Second, the column percentages (rather than the row percentages) sum to 100. Whether you want percents for rows or columns depends on your research question. In our example, the column percentages largely reflect the size of the age category, rather than the relationship between age and health. It would be much more useful if the row percentages (showing the distribution of health status within each age group) were present and summed to 100.

Third, the table includes data from individuals whose health status was coded as "unknown--refused" and "unknown--don't know." You will probably want to exclude these "unknown" cases, to get meaningful row percentages of health status by age that sum to 100 percent.

Fourth, the table is unweighted. The National Health Interview Survey (NHIS) has a complex sample design that oversamples some demographic groups in some years, so weighted results should be used. The solution to the first problem (of too many values) is to use an SDA function known as recoding. To recode a variable, hold your cursor over the "Create Variables" option at the top of the SDA web page.

Analysis	Create Variables	Codebook	Getting Started
	Recode variables	Compute a new variable	List Created Variables
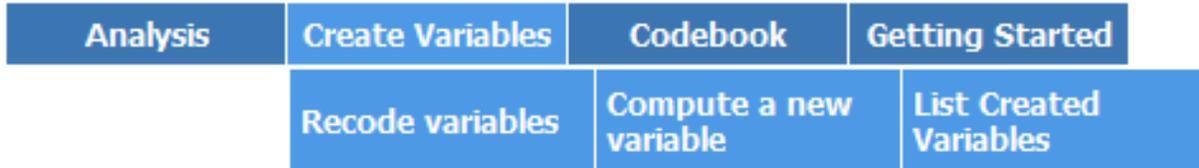
Then click on "Recode variables." The right-hand side of the screen should look like this:

SDA Recode Program
Help: General / Recoding Rules

NAMES of the variables

Name for the new variable to be created:
Replace that variable, if it already exists? ○ Yes ● No

Name(s) of existing variables to use for the recode:
(Need at least 1 input variable; can use up to 6 variables)

Var 1	Var 2	Var 3	Var 4	Var 5	Var 6

RECODING RULES (See explanation and examples)

OUTPUT Variable		VALUES of the INPUT Variables					
Value	Label	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6

Define MORE output categories (if needed)

What to do with unspecified combinations of input variables (if any):
● Convert them to MD code ○ Assign the value of input variable# 1 ▾

Our goal is to simplify our original output by collapsing the values of the age variable into more manageable categories. One strategy might be to group the responses by decade (i.e., 0-9, 10-19, etc.). To do this, we first need to name the new, recoded variable that we are creating. You may choose any name you want, as long as it is not a name already taken by another variable. Let's call our new age variable "age_r," with the "r" indicating that the variable is a recode.

The second step in recoding is to identify the already existing variable(s) that will be used as the source(s) from which we create our new, grouped age variable. In this example, the existing variable is "age." You can type "age" into the "Var 1" box, or you can select age from the variable list to the left and click the "Copy to: Var 1" button.
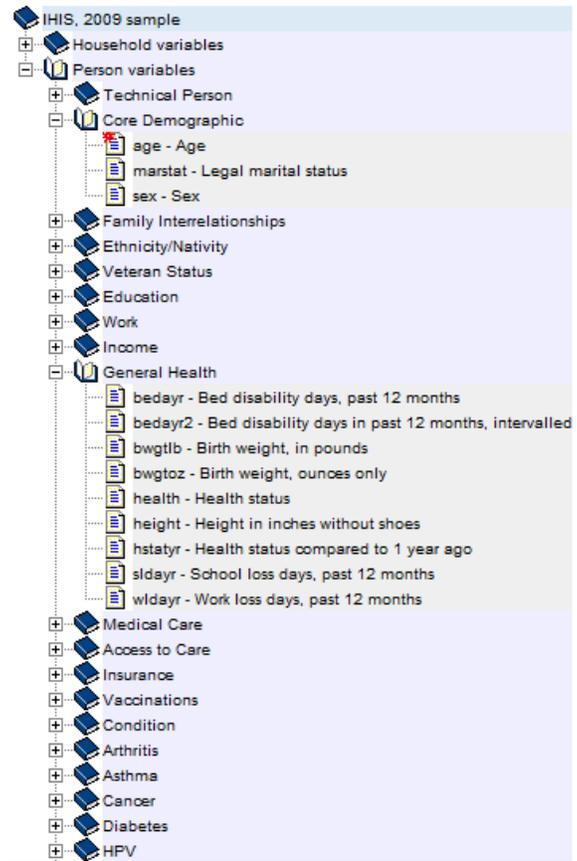
The top portion of the right-hand side of the screen should now look like this:

SDA Recode Program
Help: General / Recoding Rules

NAMES of the variables

Name for the new variable to be created: `age_r`
Replace that variable, if it already exists? ○ Yes ◉ No

Name(s) of existing variables to use for the recode:
(Need at least 1 input variable; can use up to 6 variables)

Var 1	Var 2	Var 3	Var 4	Var 5	Var 6
age					

Just below this, we need to enter values and labels for the output variable (age_r) and the values of the input variable (age) that will be used to define the values of the output variable. At this point, it is a good idea to look at the codes and frequencies for the variable you are recoding, by checking these through the variable description on the IHIS website in a separate browser window.

If we go to the variable description for the IHIS variable "age" and click on "codes," this brings up a table of codes for each year (marked by X's if the code is present in the specific year, or by unweighted numeric values if you select "case-count view"). Through the codes and fequencies page, we learn that the variable "age" is topcoded at age 85+ for some survey years, and that the number of cases in each single year age category becomes quite small at the highest ages. We might, then, decide to make the final age category be age 80 or older.

To return to the SDA interface, if we wanted people age 0-9 to be put into one category with the value "0" and the label "0's," and we repeated this pattern for each age grouping, the screen should look like this:

RECODING RULES (See explanation and examples)

OUTPUT Variable		VALUES of the INPUT Variables					
Value	Label	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6
0	0's	0-9					
1	10's	10-19					
2	20's	20-29					
3	30's	30-39					
4	40's	40-49					
5	50's	50-59					
6	60's	60-69					

Define MORE output categories (if needed)

What to do with unspecified combinations of input variables (if any):
◉ Convert them to MD code ○ Assign the value of input variable# 1 ▾

At this point, we have only assigned output values up through age 69, so we need to click on "Define MORE output categories (if needed)" to continue our recoding.

Definitions for Additional Output Categories

OUTPUT Variable			VALUES of the INPUT Variables					
Value	Label	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6	
7	70's	70-79						
8	80+	80-99						

Continue with other specifications

We have now assigned output values for all of the valid values of the input variable "age." If you want, you may assign a label to the new variable age_r, although this is not necessary. Now, click on the "Start Recoding" button, located just above the "Definitions for Additional Output Categories" menu.

OPTIONAL Specifications for the New Variable

Label:

Missing-data codes:

Minimum valid value:

Maximum valid value:

Descriptive text:

Color coding? ⦿ On ○ Off

[Start Recoding] [Clear Fields]

SDA will then create our new variable "age_r" by recoding values of the input variable "age." As you can see from the output, a variable that had 86 values has been used to create a variable that has only 9 values.

Description of the derived variable

age_r

Percent	N	Value	Label
14.8	13,127	0	0's
14.9	13,217	1	10's
13.1	11,553	2	20's
13.4	11,885	3	30's
14.3	12,652	4	40's
12.9	11,409	5	50's
8.7	7,736	6	60's
4.8	4,258	7	70's
2.9	2,609	8	80+
100.0	88,446		Total

We can now use the new variable "age_r" in any analysis, just as we would use any other variable. To return to the Frequencies/Crosstabulation menu, hold your cursor over the "Analysis" option at the top of the screen and click on "Frequencies or crosstabulation."

Analysis	Create Variables	Codebook	Getting Started				
Frequencies or crosstabulation	Comparison of means	Correlation matrix	Comparison of correlations	Multiple regression	Logit/Probit regression	List values of individual cases

You can now repeat the same steps as before to create a table of age and health status. However, this time you will use "age_r" as the row variable (instead of "age"). The column variable will still be "health." However, at this stage, we can also take care of the "unknown--refused" and "unknown -- don't know" categories from the health status variable. As you saw in the original table of age and health status, the valid values of health status are 1-5, and 7 and 9 are the valuse of the "unknown" groups we do not want to include. We can limit cases by typing "health(1-5)" as our column variable. This means we only want the table calculated for those having a health status value of 1-5.

To generate more meaningful percentage figures, we can check the "Row" box under Table Options for Percentaging. This allows us to see the distribution of health status within age categories. (Alternatively, we could make "age_r" the column variable, and make "health(1-5)" the row variable. In that case, column percentages, which are already set as the default percentaging option, would be more helpful.)

Neither row nor column percentages are intrinsically better than the other. The default SDA setting produces column percentages, but you should choose between row and column depending on which is appropriate for your research question.

Finally, as explained above, we need to run weighted data. The variable descriptions available on the IHIS website tell us which weights to use with each variable. That documentation tells us that for the IHIS variables HEALTH and AGE (which is the source variable for our SDA variable "age_r") we need to use PERWEIGHT. We can select perweight from the Weight drop-menu. Thus, the input incorporating all these changes should look like this:

SDA Frequencies/Crosstabulation Program
Help: General / Recoding Variables

REQUIRED Variable names to specify
Row: age_r
OPTIONAL Variable names to specify
Column: health(1-5)
Control:
Selection Filter(s):   *Example: age(18-50)*
Weight: perweight - Person weight

TABLE OPTIONS

Percentaging:
  ☑ Column ☑ Row ☐ Total
  ☐ Confidence intervals Level: 95 percent
  ☐ Standard error of each percent
  Sample design: ◉ Complex ○ SRS

N of cases to display:
  ☐ Unweighted ☑ Weighted

☐ Summary statistics
☐ Question text ☐ Suppress table
☑ Color coding ☐ Show Z-statistic
☐ Include missing-data values

CHART OPTIONS

Type of chart: Stacked Bar Chart
Bar chart options:
  Orientation: ◉ Vertical ○ Horizontal
  Visual Effects: ◉ 2-D ○ 3-D

Show percents: ☐ Yes
Palette: ◉ Color ○ Grayscale
Size - width: 600   height: 400

Title:
[Run the Table] [Clear Fields]

Note: If all you want to do is look at the distribution of cases within the sample, it is not necessary to run the table using weights. However, if you want to make any inferences about the population from which this sample is drawn, weights are necessary.

By clicking "Run the Table," we get new output:

Frequency Distribution						
Cells contain: -Column percent -Row percent -Weighted N	health					
	1 Excellent	2 Very Good	3 Good	4 Fair	5 Poor	ROW TOTAL
age_r 0: 0's	22.6 58.7 24,432,420.0	11.7 25.7 10,684,973.0	8.2 14.1 5,883,193.0	2.2 1.2 500,551.0	1.3 .2 91,424.0	13.8 100.0 41,592,561.0
1: 10's	20.4 53.7 22,027,021.0	12.6 28.2 11,568,835.0	9.0 15.8 6,493,787.0	3.8 2.1 850,104.0	1.4 .3 104,617.0	13.6 100.0 41,044,364.0
2: 20's	15.9 41.6 17,201,073.0	14.8 32.8 13,586,649.0	12.2 21.2 8,790,728.0	7.1 3.8 1,582,622.0	3.0 .5 218,869.0	13.7 100.0 41,379,941.0
3: 30's	13.0 35.6 14,030,350.0	14.5 33.7 13,267,033.0	12.8 23.4 9,221,413.0	10.7 6.1 2,405,784.0	6.2 1.2 454,925.0	13.1 100.0 39,379,505.0
4: 40's	11.5 28.6 12,427,041.0	15.9 33.5 14,557,328.0	16.3 27.0 11,719,229.0	16.1 8.3 3,613,435.0	15.4 2.6 1,119,081.0	14.4 100.0 43,436,114.0
5: 50's	8.7 23.3 9,359,499.0	14.0 31.9 12,824,209.0	16.3 29.1 11,718,532.0	19.9 11.1 4,463,663.0	25.4 4.6 1,850,909.0	13.4 100.0 40,216,812.0
6: 60's	5.0 19.2 5,365,160.0	8.9 29.1 8,129,639.0	12.3 31.5 8,796,816.0	18.3 14.7 4,096,685.0	20.7 5.4 1,504,048.0	9.3 100.0 27,892,348.0
7: 70's	2.0 13.5 2,149,063.0	5.0 28.9 4,619,502.0	7.6 34.3 5,475,211.0	12.1 17.0 2,710,570.0	13.8 6.3 1,006,168.0	5.3 100.0 15,960,514.0
8: 80+	.9 9.5 956,467.0	2.5 22.8 2,300,863.0	5.1 36.6 3,696,049.0	9.9 21.9 2,213,376.0	12.8 9.2 930,267.0	3.4 100.0 10,097,022.0
COL TOTAL	100.0 35.9 107,948,094.0	100.0 30.4 91,539,031.0	100.0 23.9 71,794,958.0	100.0 7.5 22,436,790.0	100.0 2.4 7,280,308.0	100.0 100.0 300,999,181.0

This is a much more manageable table, and it includes only respondents whose health status was known and reported.

Each cell of the table contains 3 numbers. If you look at the upper-left corner of the above graphic, you will find a key that explains what each of these numbers means. Users should bear in mind that the NHIS represents the US civillian, non-institutionalized population (hereafter, "the population" for brevity) in a given year.

If you look, for example, at the cell at the intersection of "30's" and "Excellent," you see that 13.0 is the column percent. This means that 13% of the population with excellent health are in their 30's. The second number is 35.6, which is the row percent. This indicates that 35.6% of the population in their 30's are in excellent health. This is a more meaningful number than the column percent. You might

12

contrast it with, say, the row percent in the 80+ Excellent health cell. The results there indicate that only 9.5% of people at least 80 are in excellent health. Finally, back in the cell at the intersection of "30's" and "Excellent" you see a "weighted N" of just over 14 million. This indicates that there were about 14 million people in the population in 2009 who were in their 30's and had excellent health.

Now let's imagine that you want to know whether the relationship between age and health status varies according to some third variable. Perhaps the relationship is different for males than it is for females. You can take a third variable into account by using the "Control" field, to produce a separate table for each category of the variable you enter there. If you wanted to see the relationship between age and health status for males and females separately, you would enter the variable "sex" as the control variable.

SDA Frequencies/Crosstabulation Program
Help: General / Recoding Variables

REQUIRED Variable names to specify
Row: age_r
OPTIONAL Variable names to specify
Column: health(1-5)
Control: sex
Selection Filter(s): _____   Example: age(18-50)
Weight: perweight - Person weight

Clicking on "Run the Table" yields this output:

				Statistics for sex = 1(Male)			
Cells contain: -Column percent -Row percent -Weighted N				health			
		1 Excellent	2 Very Good	3 Good	4 Fair	5 Poor	ROW TOTAL
age_r	0: 0's	22.3 57.4 12,186,206.0	12.7 26.6 5,639,515.0	8.9 14.6 3,101,571.0	2.4 1.2 245,055.0	1.7 .3 55,419.0	14.4 100.0 21,227,766.0
	1: 10's	20.8 54.0 11,337,387.0	13.4 28.3 5,943,984.0	9.2 15.3 3,206,079.0	4.5 2.2 469,591.0	1.6 .2 51,655.0	14.2 100.0 21,008,696.0
	2: 20's	16.3 43.1 8,907,776.0	14.7 31.5 6,519,442.0	12.9 21.7 4,482,877.0	6.6 3.3 682,207.0	3.0 .5 97,515.0	14.0 100.0 20,689,817.0
	3: 30's	13.3 37.0 7,248,292.0	14.8 33.6 6,581,844.0	12.8 22.7 4,451,966.0	11.0 5.8 1,135,220.0	6.0 1.0 195,942.0	13.3 100.0 19,613,264.0
	4: 40's	11.5 29.4 6,271,213.0	15.8 32.8 6,999,389.0	16.7 27.4 5,838,438.0	16.5 8.0 1,706,010.0	15.5 2.4 506,572.0	14.5 100.0 21,321,622.0
	5: 50's	8.4 23.8 4,585,213.0	14.0 32.3 6,218,611.0	16.2 29.4 5,657,920.0	19.1 10.3 1,977,460.0	25.4 4.3 826,382.0	13.1 100.0 19,265,586.0
	6: 60's	5.0 20.4 2,725,301.0	8.4 28.0 3,742,448.0	12.1 31.5 4,211,853.0	19.1 14.7 1,969,942.0	22.6 5.5 735,468.0	9.1 100.0 13,385,012.0
	7: 70's	1.9 14.8 1,060,294.0	4.3 26.4 1,895,083.0	7.3 35.4 2,541,166.0	12.1 17.5 1,252,003.0	12.9 5.9 420,142.0	4.9 100.0 7,168,688.0
	8: 80+	.6 8.0 302,137.0	1.9 22.2 841,183.0	3.9 36.3 1,371,678.0	8.7 23.8 898,840.0	11.3 9.7 368,769.0	2.6 100.0 3,782,607.0
	COL TOTAL	100.0 37.0 54,623,819.0	100.0 30.1 44,381,499.0	100.0 23.6 34,863,548.0	100.0 7.0 10,336,328.0	100.0 2.2 3,257,864.0	100.0 100.0 147,463,058.0

Statistics for sex = 2(Female)						
Cells contain: -Column percent -Row percent -Weighted N	health					
	1 Excellent	2 Very Good	3 Good	4 Fair	5 Poor	ROW TOTAL
age_r 0: 0's	23.0 60.1 12,246,214.0	10.7 24.8 5,045,458.0	7.5 13.7 2,781,622.0	2.1 1.3 255,496.0	.9 .2 36,005.0	13.3 100.0 20,364,795.0
1: 10's	20.0 53.4 10,689,634.0	11.9 28.1 5,624,851.0	8.9 16.4 3,287,708.0	3.1 1.9 380,513.0	1.3 .3 52,962.0	13.0 100.0 20,035,668.0
2: 20's	15.6 40.1 8,293,297.0	15.0 34.2 7,067,207.0	11.7 20.8 4,307,851.0	7.4 4.4 900,415.0	3.0 .6 121,354.0	13.5 100.0 20,690,124.0
3: 30's	12.7 34.3 6,782,058.0	14.2 33.8 6,685,189.0	12.9 24.1 4,769,447.0	10.5 6.4 1,270,564.0	6.4 1.3 258,983.0	12.9 100.0 19,766,241.0
4: 40's	11.5 27.8 6,155,828.0	16.0 34.2 7,557,939.0	15.9 26.6 5,880,791.0	15.8 8.6 1,907,425.0	15.2 2.8 612,509.0	14.4 100.0 22,114,492.0
5: 50's	9.0 22.8 4,774,286.0	14.0 31.5 6,605,598.0	16.4 28.9 6,060,612.0	20.5 11.9 2,486,203.0	25.5 4.9 1,024,527.0	13.6 100.0 20,951,226.0
6: 60's	5.0 18.2 2,639,859.0	9.3 30.2 4,387,191.0	12.4 31.6 4,584,963.0	17.6 14.7 2,126,743.0	19.1 5.3 768,580.0	9.4 100.0 14,507,336.0
7: 70's	2.0 12.4 1,088,769.0	5.8 31.0 2,724,419.0	7.9 33.4 2,934,045.0	12.1 16.6 1,458,567.0	14.6 6.7 586,026.0	5.7 100.0 8,791,826.0
8: 80+	1.2 10.4 654,330.0	3.1 23.1 1,459,680.0	6.3 36.8 2,324,371.0	10.9 20.8 1,314,536.0	14.0 8.9 561,498.0	4.1 100.0 6,314,415.0
COL TOTAL	100.0 34.7 53,324,275.0	100.0 30.7 47,157,532.0	100.0 24.1 36,931,410.0	100.0 7.9 12,100,462.0	100.0 2.6 4,022,444.0	100.0 100.0 153,536,123.0

Entering "sex" as a control variable allows a quick visual comparison of whether the relationship between age and health status is different for males than for females. For both groups the tables indicate a negative relationship between age and health. That is, older people tend to have worse health status than youger people.

Users will notice that the default output includes color coding, which is representative of Z-statistics.

Color coding:	<-2.0	<-1.0	<0.0	>0.0	>1.0	>2.0	Z
N in each cell:	Smaller than expected			Larger than expected			

To see explanations of the meaning of Z-statistics and the color coding, click on the "Color coding" and "Show Z-statistic" links in the interface of Table Options.

You can also opt to turn the color coding off by unchecking the "color coding" box shown above.

Another SDA option, "selection filter(s)," allows you to analyze only a specific subset of respondents for a given variable. This option works best when using the IHIS data file for 1997 forward, when you want to select only certain years (e.g., 2002 and 2007) to use in your analysis.

When used with other, non-year variables, however, the selection filter is only useful for getting a sense of data distribution. The selection filter is not useful for variance estimation or significance testing, because of the complex sample design of the NHIS.

To continue the example begun above, let's say we wanted to check the relationship between age and health by sex, but only for people who have never been married. In other words, is the relationship between health and age different for never married females than for never married males? To do this, enter the variable "marstat" (legal marital status) as a selection filter, followed by the code for "never married" in parentheses.

To find out the code for "never married," you can review the codes and frequencies by clicking on "codes" in the MARSTAT variable description on the IHIS website. Alternatively, you can enter "marstat" in the variable dictionary to the left side of the screen and click "View." Doing so will open another window:

marstat		Legal marital status	
Percent	**N**	**Value**	**Label**
20.8	18,359	0	NIU
40.4	35,716	10	Married
0.0	0	11	Married - Spouse present
0.0	0	12	Married - Spouse not in household
0.0	0	13	Married - Spouse in household unknown
4.2	3,731	20	Widowed
7.5	6,607	30	Divorced
2.0	1,745	40	Separated
24.9	21,988	50	Never married
0.3	300	99	Unknown marital status
100.0	**88,446**		**Total**

Properties	
Data type:	numeric
Mean:	20.68
Std Dev:	19.58
Record/columns:	1/50-51

Selected Study: IHIS, 2009 sample

As you can see, the code for "Never married" is 50. Thus, your input to produce tables for only never-married people will look like this:

SDA Frequencies/Crosstabulation Program
Help: General / Recoding Variables

REQUIRED Variable names to specify
Row: age_r
OPTIONAL Variable names to specify
Column: health(1-5)
Control: sex
Selection Filter(s): marstat(50)   *Example: age(18-50)*
Weight: perweight - Person weight

Clicking "Run the Table" produces separate tables for males and females again, since we have kept "sex" as a control variable.

Statistics for sex = 1(Male)							
Cells contain: -Column percent -Row percent -Weighted N		health					
		1 Excellent	2 Very Good	3 Good	4 Fair	5 Poor	ROW TOTAL
age_r	1: 10's	42.0 53.1 6,760,434.0	29.9 28.1 3,575,431.0	23.8 16.3 2,074,413.0	13.5 2.3 290,171.0	5.5 .2 27,288.0	32.3 100.0 12,727,737.0
	2: 20's	40.4 42.6 6,502,657.0	40.8 31.9 4,873,152.0	37.5 21.4 3,272,244.0	25.9 3.6 555,986.0	12.1 .4 59,477.0	38.7 100.0 15,263,516.0
	3: 30's	9.9 31.8 1,589,526.0	13.3 31.7 1,584,073.0	15.2 26.6 1,327,626.0	18.2 7.8 390,726.0	21.1 2.1 104,083.0	12.7 100.0 4,996,034.0
	4: 40's	4.1 19.4 657,484.0	9.3 32.8 1,110,766.0	11.4 29.3 993,787.0	22.6 14.3 485,408.0	29.1 4.2 143,404.0	8.6 100.0 3,390,849.0
	5: 50's	2.3 19.3 371,914.0	5.2 32.2 621,010.0	7.1 32.3 622,341.0	10.3 11.5 221,984.0	18.5 4.7 91,422.0	4.9 100.0 1,928,671.0
	6: 60's	.9 20.0 151,758.0	1.2 18.2 138,070.0	3.8 43.3 327,641.0	4.3 12.3 92,950.0	9.5 6.2 47,077.0	1.9 100.0 757,496.0
	7: 70's	.5 26.0 73,225.0	.2 9.9 27,922.0	.9 28.5 80,145.0	3.9 29.6 83,434.0	3.4 6.0 16,839.0	.7 100.0 281,565.0
	8: 80+	.0 7.1 5,782.0	.1 21.9 17,841.0	.3 36.9 29,999.0	1.1 29.7 24,159.0	.7 4.4 3,540.0	.2 100.0 81,321.0
	COL TOTAL	100.0 40.9 16,112,780.0	100.0 30.3 11,948,265.0	100.0 22.1 8,728,196.0	100.0 5.4 2,144,818.0	100.0 1.3 493,130.0	100.0 100.0 39,427,189.0

Statistics for sex = 2(Female)						
Cells contain: -Column percent -Row percent -Weighted N	health					
	1 Excellent	2 Very Good	3 Good	4 Fair	5 Poor	ROW TOTAL
age_r 1: 10's	45.9 52.0 6,291,035.0	31.4 28.2 3,412,717.0	26.9 17.3 2,090,431.0	13.6 2.2 269,064.0	7.0 .3 39,917.0	34.7 100.0 12,103,164.0
2: 20's	36.9 39.8 5,057,659.0	40.6 34.8 4,419,696.0	34.1 20.9 2,650,703.0	24.9 3.9 491,782.0	15.5 .7 88,829.0	36.4 100.0 12,708,669.0
3: 30's	9.4 29.7 1,291,083.0	13.0 32.6 1,415,538.0	15.4 27.5 1,196,029.0	18.5 8.4 366,241.0	13.5 1.8 77,595.0	12.5 100.0 4,346,486.0
4: 40's	4.4 21.9 603,646.0	7.8 30.7 848,002.0	11.3 31.8 877,899.0	15.9 11.4 313,678.0	20.4 4.2 117,033.0	7.9 100.0 2,760,258.0
5: 50's	2.3 17.4 311,394.0	4.0 24.5 439,168.0	7.3 31.8 569,585.0	16.0 17.6 315,322.0	27.3 8.7 156,521.0	5.1 100.0 1,791,990.0
6: 60's	.6 11.7 82,894.0	1.8 27.9 197,146.0	3.1 34.6 243,900.0	6.5 18.2 128,102.0	9.3 7.6 53,568.0	2.0 100.0 705,610.0
7: 70's	.3 14.5 40,512.0	.9 36.2 101,031.0	.9 24.3 67,741.0	2.8 19.7 54,898.0	2.6 5.4 15,110.0	.8 100.0 279,292.0
8: 80+	.2 12.6 25,389.0	.4 20.2 40,762.0	1.0 36.8 74,288.0	1.8 17.8 35,889.0	4.5 12.7 25,602.0	.6 100.0 201,930.0
COL TOTAL	100.0 39.3 13,703,612.0	100.0 31.2 10,874,060.0	100.0 22.3 7,770,576.0	100.0 5.7 1,974,976.0	100.0 1.6 574,175.0	100.0 100.0 34,897,399.0

This time, however, only respondents who have never been married have been included. As you see, the number of cases drops off dramatically, beginning with people in their 30's compared to those in their 20's.

Note: While use of a selection filter to produce a table for a subset population may give you a general sense of the relationship between variables, users should be cautious about drawing statistical inferences from such results. Due to the complex sample design of the National Health Interview Survey, restricting analysis to a subpopulation as described above may yield incorrectly computed standard errors. See the IHIS user note on Variance Estimation for further discussion of this problem and for examples of correct practice in subpopulation analysis using a statistical package like SAS, Stata, or SAS-callable SUDAAN.

Analyzing Multiple Years of Data

You will notice on the SDA-IHIS interface that you have the choice of either using data from a single year or using data from multiple years for 1997 forward. The time period of 1997 forward was chosen because a redesign of the NHIS in 1997 resulted in more variable consistency for 1997 forward than for earlier years. If you want to analyze data from multiple years prior to 1997, you will need to analyze each year separately.

You may use the 1997 forward data file to analyze pooled data from multiple years. Alternatively, you may use the "Control" function to produce separate output for multiple years at once. The procedure for doing this is similar to the example above where we used "sex" as a control variable to produce separate tables for males and females. To produce separate output for each year when using the 1997 forward data file, enter the variable "year" as the control variable.

The above input produces the requested table for each year that both variables are available from 1997 forward, as well as a final table using pooled data from all available years for 1997 forward.

This is the table for 1997:

		health					
Cells contain: -Column percent -Row percent -Weighted N		1 Excellent	2 Very Good	3 Good	4 Fair	5 Poor	ROW TOTAL
age_r	0: 0's	22.0 56.1 22,299,129.0	13.3 26.8 10,658,574.0	10.0 15.0 5,967,176.0	3.9 1.7 683,333.0	2.1 .3 120,595.0	15.0 100.0 39,728,807.0
	1: 10's	19.8 52.0 20,051,134.0	13.7 28.4 10,977,834.0	11.1 17.3 6,677,886.0	4.3 2.0 764,170.0	2.1 .3 122,993.0	14.5 100.0 38,594,017.0
	2: 20's	15.1 42.3 15,315,600.0	15.1 33.5 12,151,397.0	12.2 20.3 7,345,237.0	7.1 3.5 1,268,104.0	2.7 .4 156,830.0	13.7 100.0 36,237,168.0
	3: 30's	16.8 39.8 17,093,006.0	17.8 33.3 14,298,494.0	15.2 21.3 9,124,813.0	11.2 4.6 1,989,381.0	6.8 .9 392,625.0	16.2 100.0 42,898,319.0
	4: 40's	13.3 33.9 13,457,197.0	16.5 33.4 13,238,280.0	15.7 23.7 9,395,182.0	15.7 7.0 2,782,809.0	14.1 2.1 818,621.0	15.0 100.0 39,692,089.0
	5: 50's	6.8 26.1 6,903,154.0	10.4 31.6 8,349,648.0	12.2 27.7 7,322,600.0	15.6 10.5 2,775,676.0	18.6 4.1 1,074,073.0	10.0 100.0 26,425,151.0
	6: 60's	3.6 18.7 3,623,101.0	6.6 27.2 5,264,017.0	10.6 32.9 6,363,342.0	16.5 15.1 2,923,496.0	20.7 6.2 1,195,680.0	7.3 100.0 19,369,636.0
	7: 70's	2.0 13.4 2,035,508.0	4.7 24.9 3,769,915.0	8.9 35.1 5,317,079.0	16.5 19.3 2,919,195.0	19.1 7.3 1,103,816.0	5.7 100.0 15,145,513.0
	8: 80+	.7 9.9 709,364.0	2.0 21.9 1,566,516.0	4.1 34.3 2,458,278.0	9.2 22.8 1,631,931.0	13.8 11.2 800,878.0	2.7 100.0 7,166,967.0
	COL TOTAL	100.0 38.3 101,487,193.0	100.0 30.3 80,274,675.0	100.0 22.6 59,971,593.0	100.0 6.7 17,738,095.0	100.0 2.2 5,786,111.0	100.0 100.0 265,257,667.0

Statistics for year = 1997(1997)

And this is the table for all years 1997-2009:

Statistics for all valid cases						
Cells contain: -Column percent -Row percent -Weighted N	health					
	1 Excellent	2 Very Good	3 Good	4 Fair	5 Poor	ROW TOTAL
age_r 0: 0's	22.0 56.7 295,152,555.0	12.4 27.1 140,896,357.0	8.9 14.6 76,244,250.0	2.9 1.4 7,480,086.0	1.2 .2 1,009,090.0	14.2 100.0 520,782,338.0
1: 10's	20.1 51.2 270,595,385.0	13.6 29.2 154,426,780.0	10.7 17.4 92,059,443.0	4.0 2.0 10,380,253.0	1.7 .3 1,413,350.0	14.4 100.0 528,875,211.0
2: 20's	15.5 41.7 207,945,004.0	14.9 33.8 168,429,265.0	11.8 20.5 101,933,246.0	6.8 3.5 17,580,133.0	2.9 .5 2,475,201.0	13.5 100.0 498,362,849.0
3: 30's	14.8 37.5 198,421,938.0	16.1 34.5 182,504,929.0	13.5 22.0 116,521,873.0	10.1 4.9 26,017,868.0	6.4 1.0 5,391,101.0	14.4 100.0 528,857,709.0
4: 40's	12.9 31.1 173,697,757.0	16.8 34.0 190,115,424.0	16.4 25.2 140,971,892.0	16.3 7.5 41,743,466.0	15.3 2.3 12,881,591.0	15.2 100.0 559,410,130.0
5: 50's	8.0 24.3 106,844,237.0	12.2 31.6 138,770,624.0	14.7 28.8 126,424,856.0	18.7 10.9 48,094,557.0	22.7 4.4 19,157,541.0	11.9 100.0 439,291,815.0
6: 60's	4.0 18.6 53,997,429.0	7.2 28.2 81,793,716.0	11.0 32.6 94,464,007.0	16.7 14.8 42,973,025.0	19.5 5.7 16,425,103.0	7.9 100.0 289,653,280.0
7: 70's	1.9 12.8 25,809,878.0	4.6 25.8 52,204,371.0	8.4 35.6 71,966,208.0	14.7 18.7 37,843,107.0	17.0 7.1 14,282,433.0	5.5 100.0 202,105,997.0
8: 80+	.8 9.7 10,790,899.0	2.2 21.9 24,379,418.0	4.7 36.2 40,371,881.0	9.6 22.2 24,737,424.0	13.3 10.0 11,174,757.0	3.0 100.0 111,454,379.0
COL TOTAL	100.0 36.5 1,343,255,082.0	100.0 30.8 1,133,520,884.0	100.0 23.4 860,957,656.0	100.0 7.0 256,849,919.0	100.0 2.3 84,210,167.0	100.0 100.0 3,678,793,708.0

A Word of Caution Regarding the Use of Household Variables

Researchers wishing to do an analysis using household variables should exercise caution. If the desired unit of analysis is the household, such an analysis should not be done using SDA. The data that SDA employs in its analyses are rectangularized, meaning that household records have been attached to each person record. The result of this data structure is that the frequencies reported by SDA for household variables, such as "region," actually reflect the number of **persons**, rather than the number of **households**.

If, however, the desired unit of analysis is persons (even when analyzing "household" variables), it is appropriate to use SDA. If you want to know the number of people living in each Census region of the United States, you can apply the **person** weight variable in the Frequencies/Crosstabulation Program to obtain frequencies for "region."

Researchers who want to obtain the number of households in each region (or use any household variable with households as the unit of analysis) should use the IHIS website to download a **hierarchical** data extract (rather than a rectangular one). To produce figures on the number of households with a given characteristic, they should then analyze household records (those with a value of "H" in the variable RECTYPE) with a statistical package such as SAS, SPSS, or Stata, rather than SDA.

Comparison of Means Program

Rather than making a table, you may wish to compute the mean value for a variable. To compute and compare means, use the comparison of means program in SDA. This program calculates the mean of a designated "dependent variable" within categories of a "row variable."

To begin, as with the frequencies/crosstabulation program, select the year(s) in which you are interested. Let's use the 2009 sample again.

[insert graphic of choosing the 2009 sample here (again, after Jason has re-designed the web page)]

Next, hold your cursor over "Analysis" and click on "Comparison of Means."

Analysis	Create Variables	Codebook	Getting Started				
Frequencies or crosstabulation	Comparison of means	Correlation matrix	Comparison of correlations	Multiple regression	Logit/Probit regression	List values of individual cases

Enter the variable for which you want to calculate the mean (average) as the "Dependent" variable. **Only enter a numerical variable here**, as those are the kind of variables for which an average can logically be calculated. If you enter a non-numerical variable (e.g., "sex"), SDA will still calculate a "mean" based on the values and frequencies of that variable, but any such "mean" will have no real meaning.

With that in mind, let's enter "bedayr," which is the number of days during the preceding 12 months that illness or injury kept a person in bed for more than half the day. By consulting either the variable description on the IHIS website, or by entering "bedayr" in the SDA variable dictionary and clicking "View," we can see that the values for "bedayr" range from 0 (none) to 999 (Unknown-don't know). However, only 1-366 actually correspond to days in bed.

bedayr		Bed disability days, past 12 months					
Percent	**N**	**Value**	**Label**				
20.0	17,710	0	None		0.0	0	361	361
2.7	2,418	1	1		0.0	0	362	362
2.7	2,353	2	2		0.0	0	363	363
1.2	1,081	3	3		0.0	0	364	364
0.6	548	4	4		0.1	77	365	365
0.7	652	5	5		0.0	0	366	366
0.2	206	6	6		68.6	60,715	996	NIU
0.4	371	7	7		0.0	14	997	Unknown-refused
0.1	71	8	8		0.0	12	998	Unknown-not ascertained
0.1	45	9	9		0.2	185	999	Unknown-don't know

(reading note: the right-hand portion of the figure)

0.0	0	**361**	361
0.0	0	**362**	362
0.0	0	**363**	363
0.0	0	**364**	364
0.1	77	**365**	365
0.0	0	**366**	366
68.6	60,715	**996**	NIU
0.0	14	**997**	Unknown-refused
0.0	12	**998**	Unknown-not ascertained
0.2	185	**999**	Unknown-don't know
100.0	**88,446**		**Total**

If you look at code 996, you see the label "NIU." NIU is the IHIS abbreviation for "not in universe"; persons with this value for a variable were not asked the survey question relating to that variable. We would normally want to exclude anyone coded as NIU from analysis, by following the procedure described above for including only the valid values of the health status variable (i.e., for excluding cases coded as "unknown" health status). For "bedayr" we want to exclude both the NIU responses and the 3 varieties of "Unknown" responses when calculating a mean value.

Thus, we enter "bedayr" as the "Dependent" variable, with the numbers 0-366 following in parentheses. This restricts the input to persons with a reported number of bed days ranging from 0 to 366. If we wanted to calculate the average number of annual bed days within each category of health status, we would enter "health" as the row variable. As in the frequencies/cross-tabulation examples, we will put (1-5) after the health variable, to include only valid responses and exclude "unknown" responses.

Earlier we learned that the "health" variable should be used in conjunction with the "perweight" weight variable, according to the IHIS website. Checking the variable description for "bedayr" on the IHIS website tells us that that "bedayr" should be used with the "sampweight" weight variable. In such cases, where two different weights are specified for two variables you wish to combine, it is appropriate to use the "narrower" of the two weight options. Everyone in the NHIS sample has a person weight, but only sample persons have a sample weight. In this case, then, we should use "sampweight" as our weight variable. (Similarly, if the choice is between "perweight" and "mortwt", you should use "mortwt." If the choice is between "sampweight" and "mortwtsa," you should use "mortwtsa." See the IHIS user note on weights for more information.)

The input to get means for "bedayr" sorted by health status looks like this:

SDA Comparison of Means Program
Help: **General** / **Recoding Variables**

REQUIRED Variable names to specify
Dependent: bedayr(0-366)
Row: health(1-5)
OPTIONAL Variable names to specify
Column:
Control:
Selection Filter(s):   *Example: age(18-50)*
Weight: sampweight - Sample Person weight ▾

Main statistic to display: Means ▾

By clicking "Run the Table," you get this output:

Main Statistics		
Cells contain: -Means -Complex Std Errs -Weighted N		
health	1: Excellent	**1.09** .059 64,661,011.0
	2: Very Good	**1.83** .160 72,280,319.0
	3: Good	**4.10** .340 60,100,911.0
	4: Fair	**14.88** 1.132 21,520,894.0
	5: Poor	**56.13** 4.627 7,098,523.0
	COL TOTAL	*5.18* *.245* *225,661,658.0*

As with the earlier example, the box in the upper-left corner explains the numbers in each cell. The bold number is the mean--the average number of bed disability days for people with a given health status. In 2009, people with excellent health had, on average, 1.09 bed disability days in the 12 months prior to being surveyed. By contrast, those with poor health had 56.13 bed disability days, on average, over the same time period. The second number in each cell is the standard error. Because IHIS uses a complex sample design, these standard errors are calculated in SDA using the Taylor series method. Standard errors are used in significance testing (see below). For more information, click on the "Complex std errs" link under Table Options.

TABLE OPTIONS
Additional statistics in each cell
☑ **Complex std errs** ☐ **SRS std errs**
☐ **Design effect (deft)** ☐ **Rho**
☐ **Std dev** ☐ **N** ☑ **Weighted N**
☐ **Z/T-statistic** ☐ **P-value**

The third number in each cell is the weighted number of cases used to calculate the mean. There are fewer total people in the population of this analysis (about 226 million) than in the cross-tabulation

done above (about 301 million) because the survey question for "bedayr" is only asked of sample adults, rather than all persons in the NHIS sample.

Significance Testing

The procedure just covered will display the means for any numeric variable. Some SDA users may wish to extend their analysis further by testing whether differences between means are statistically significant. The IHIS data are based on a sample of people included in the National Health Interview Survey. Because only a small subset of the U.S. non-institutionalized population was included in NHIS, results based on survey responses will differ somewhat from what the results would have been if everyone in the United States had answered these questions. Significance testing provides a means of evaluating how confident we are that differences (e.g., in means) observed with survey respondents would also be observed in the general population.

If we continue with the previous example, our goal will be to test the difference in mean number of bed disability days between health status groups. The SDA allows users to compare differences between two groups, but not across several groups. We need to indicate a base category--the category against which the means of other categories will be compared. Let's select "Excellent" (excellent health) as the base category. We then need to tell SDA that our "main statistic to display" is "differences from row category" (because health status is the row variable. This will tell us whether there is a statistically significant difference in the mean number of bed disability days for those in excellent health versus those in very good, good, fair, or poor health. Checking the codes and frequencies tells us that the code for "Excellent" is 1, which should be entered in the appropriate box. We can also tell SDA to produce a Z/T-statistic, which tests whether the difference of means is statistically significant. To do so, check the "Z/T-statistic" box.

After incorporating all the information noted above, our input to calculate the mean of bedayr within health status categories looks like this:

SDA Comparison of Means Program
Help: General / Recoding Variables

REQUIRED Variable names to specify
Dependent: `bedayr(0-366)`
Row: `health(1-5)`

OPTIONAL Variable names to specify
Column:
Control:
Selection Filter(s):   *Example: age(18-50)*
Weight: `sampweight - Sample Person weight ▾`

Main statistic to display: `Differences from row category ▾`

If differences from a row or column, indicate base category: `1`
Optional transformation of the dependent variable: `(None) ▾`

TABLE OPTIONS

Additional statistics in each cell
☑ Complex std errs ☐ SRS std errs
☐ Design effect (deft) ☐ Rho
☐ Std dev ☐ N ☑ Weighted N
☑ Z/T-statistic ☐ P-value

Optional tables of statistics
☐ Confidence intervals Level: `95 percent ▾`
☐ Multiple classification analysis
☐ Diagnostic information (complex)

Other options
☐ ANOVA stats ☐ Suppress table
☐ Question text ☑ Color coding

CHART OPTIONS

Type of chart: `Bar Chart ▾`
Bar chart options:
  Orientation: ◉ Vertical ○ Horizontal
  Visual Effects: ◉ 2-D ○ 3-D

Show means: ☐ Yes
Palette: ◉ Color ○ Grayscale
Size - width: `600 ▾` height: `400 ▾`

Clicking on "Run the Table" yields this output:

Main Statistics		
Cells contain: -Diff from row -Complex Std Errs -Weighted N -T-statistic		
health	1: Excellent	--- --- 64,661,011.0 ---
	2: Very Good	.735 .168 72,280,319.0 4.38
	3: Good	3.001 .349 60,100,911.0 8.59
	4: Fair	13.790 1.126 21,520,894.0 12.24
	5: Poor	55.038 4.631 7,098,523.0 11.88

There are now 4 numbers in each output cell (except for the "Excellent" cell, which serves as our reference point). The bold numbers are the **difference** in means for "bedayr" relative to the "bedayr" mean for those in excellent health. They report the difference in average number of days spent in bed due to illness or injury in the past year by people with the specified health status. For example, people in good health had an average of 3 more bed disability days than those in excellent health during the 12 months prior to taking the survey. People with poor health had an average of 55 more days of bed disability than those in excellent health.

The second and third numbers are again the standard error and weighted N, respectively.

The final number in each cell is the T-statistic. This figure indicates whether a difference in means is "statistically significant." Statistical significance in this case refers to the probability that a difference of means observed in the sample would be observed in the population. Generally, a T-statistic of at least +1.96 (or less than -1.96) indicates statistical significance (i.e., highly probable that the means are different in the population). The T-statistic is greater than +1.96 for every category of health, meaning that in the population, it is likely the mean number of bed disability days is higher for people in each health category 2-5 (very good-poor) relative to the people in health category 1 (excellent).

As with the frequencies and cross-tabulation program, we can also use recoded variables and apply control variables or selection filters in conjunction with calculating means. As an example, let's re-run this comparison of means using our recoded age variable as the "Row" variable.   This time let's set our base category to 4 (people in their 40's).

SDA Comparison of Means Program
Help: General / Recoding Variables

REQUIRED Variable names to specify
Dependent: bedayr(0-366)
Row: age_r
OPTIONAL Variable names to specify
Column:
Control:
Selection Filter(s): | *Example: age(18-50)*
Weight: sampweight - Sample Person weight ▾

Main statistic to display: Differences from row category ▾

If differences from a row or column, indicate base category: 4
Optional transformation of the dependent variable: (None) ▾

TABLE OPTIONS	CHART OPTIONS
Additional statistics in each cell	**Type of chart:** Bar Chart ▾
☑ **Complex std errs** ☐ **SRS std errs**	**Bar chart options:**
☐ **Design effect (deft)** ☐ **Rho**	Orientation: ⦿ Vertical ◯ Horizontal
☐ **Std dev** ☐ **N** ☑ **Weighted N**	Visual Effects: ⦿ 2-D ◯ 3-D
☑ **Z/T-statistic** ☐ **P-value**	**Show means:** ☐ Yes
Optional tables of statistics	**Palette:** ⦿ Color ◯ Grayscale
☐ **Confidence intervals** Level: 95 percent ▾	**Size** - width: 600 ▾ height: 400 ▾
☐ **Multiple classification analysis**	
☐ **Diagnostic information (complex)**	
Other options	
☐ **ANOVA stats** ☐ **Suppress table**	
☐ **Question text** ☑ **Color coding**	

By clicking "Run the Table," we get this output:

Main Statistics		
Cells contain: -Diff from row -Complex Std Errs -Weighted N -T-statistic		
age_r	1: 10's	**-3.383** .701 7,980,707.0 -4.82
	2: 20's	**-2.336** .784 41,490,343.0 -2.98
	3: 30's	**-1.427** .729 39,415,104.0 -1.96
	4: 40's	**---** --- 43,091,853.0 ---
	5: 50's	**1.654** .823 40,104,728.0 2.01
	6: 60's	**.066** .791 27,518,576.0 .08
	7: 70's	**.052** .951 16,107,242.0 .05
	8: 80+	**3.838** 1.680 10,018,650.0 2.28

Now the difference in average number of bed disability days is reported relative to people in their 40's. Here we see that people in the oldest age group have 3.8 more bed disability days per year, on average, than people in their 40's. The T-statistic of 2.28 indicates that this difference is statistically significant (i.e., probably would be observed in the population, not just the sample). By contrast, those in their 20's have, on average, 2.3 fewer bed disability days per year than those in their 40's. The T-statistic of -2.98 indicates that this difference in means (between 20's and 40's) is also statistically significant. Notice that the T-statistics for people in their 60's and 70's are not large enough to indicate statistical significance. This means that, in the population, it is not very probable that the mean bed disability days for those age groups are different from the mean bed disability days of people in their 40's.