Drug and Alcohol Services Information System

The DASIS Report

March 15, 2002

Analyzing TEDS Online

In Brief

- Over 75 studies, including the Treatment Episode Data Set from 1992 to 1999, are publicly available from the Substance Abuse and Mental Health Data Archive (SAMHDA) at www.icpsr.umich.edu/SAMHDA

- Program administrators, researchers, and others with Internet access may run their own analysis without using specialized software

- Hypertext links on the input screens provide online help

The Treatment Episode Data Set (TEDS) is a compilation of data on the demographic characteristics and substance abuse problems of those admitted for substance abuse treatment. The information comes primarily from facilities that receive some public funding. Information on treatment admissions are routinely collected by State administrative systems and then submitted to SAMHSA in a standard format. Approximately 1.6 million records are included in TEDS each year.

The 1992 to 1999 TEDS files are available through the Substance Abuse and Mental Health Data Archive (SAMHDA) and the archive's online data analysis system (DAS).[1,2] In this report, we will demonstrate the DAS recode and compute procedures and the comparison of means function. For a more basic overview of the DAS, see **Accessing and Analyzing the NHSDA Online:** http://www.samhsa.gov/oas/facts/SAMHDA help.cfm.

The recode and compute procedures allow users to temporarily store recoded or created variables on SAMHDA's web site. This is useful when you are running
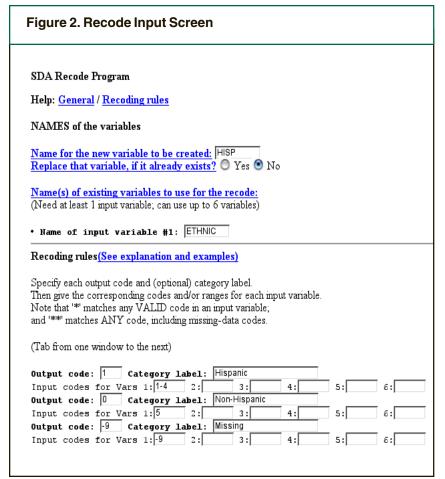
analyses over the course of several days or weeks. New variables are kept on the web site for 30 days after their last use. [3,4] Newly created variables are specific to the file for which they were created. For analyses across multiple years of a study, you must create the new variable for each year or file.

Both the recode and the compute procedures present the user with screens on which to enter the specifications for creating a new variable. The user enters the input variable name(s), the rules to convert one or more existing variables into a new variable, labels for the new variable(s), and other specifications. The recode or compute application then reads these specifications and carries out the data transformations.
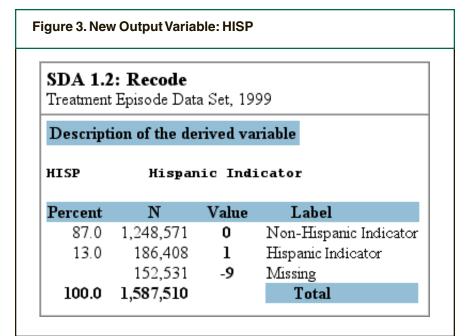
Given the detailed geographic coding in TEDS, comparisons between geographic areas among several states or within the same state are possible. In doing these comparisons, it may be useful to create new variables. To illustrate, we will examine

Figure 1. Input Variable: Ethnic

SDA 1.2: Tables
Treatment Episode Data Set, 1999

Variables					
Role	Name	Label	Range	MD	Dataset
Row	ETHNIC	ETHNICITY (HISPANIC ORIGIN)	1-5	-9	1

Frequency Distribution		
Cells contain: -N of cases		Distribution
ETHNIC	1 PUERTO RICAN	64,569
	2 MEXICAN	79,353
	3 CUBAN	4,252
	4 OTHER HISPANIC	38,234
	5 NOT OF HISPANIC ORIGIN	1,248,571
	COL TOTAL	*1,434,979*

Allocation of cases

Valid cases	1,434,979
Cases with invalid codes on row variable	152,531
Total cases	*1,587,510*

Figure 2. Recode Input Screen

SDA Recode Program

Help: **General** / **Recoding rules**

NAMES of the variables

Name for the new variable to be created: HISP
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)

• **Name of input variable #1:** ETHNIC

Recoding rules(See explanation and examples)

Specify each output code and (optional) category label.
Then give the corresponding codes and/or ranges for each input variable.
Note that '*' matches any VALID code in an input variable;
and '**' matches ANY code, including missing-data codes.

(Tab from one window to the next)

Output code: 1 **Category label:** Hispanic
Input codes for Vars 1: 1-4 2: 3: 4: 5: 6:
Output code: 0 **Category label:** Non-Hispanic
Input codes for Vars 1: 5 2: 3: 4: 5: 6:
Output code: -9 **Category label:** Missing
Input codes for Vars 1: -9 2: 3: 4: 5: 6:

the variation of prior treatment episodes between Hispanics and non-Hispanics in New York State compared to the rest of the U.S. We will accomplish this in three steps using: (1) the recode procedure to collapse the existing ethnicity variable into two categories: Hispanic and Non-Hispanic; (2) the compute procedure to create another new variable with two categories, one for New York and one for all other states; and (3) the comparison of means function to determine the average number of prior treatments based on the two new variables. Access the DAS as follows:

● Log on to the SAMHDA website: http://www.icpsr.umich.edu/ SAMHDA.

● Select Data Analysis System and follow the prompts to the 1999 TEDS until you reach the "SAMHDA SDA Archive" screen.

Figure 3. New Output Variable: HISP

SDA 1.2: Recode
Treatment Episode Data Set, 1999

Description of the derived variable

HISP   Hispanic Indicator

Percent	N	Value	Label
87.0	1,248,571	0	Non-Hispanic Indicator
13.0	186,408	1	Hispanic Indicator
	152,531	-9	Missing
100.0	1,587,510		**Total**

- Refer to the online codebook for the original variable names, codes, and value (code) labels.

- Click on the hypertext links found on each screen to obtain instructions and examples for any DAS procedure.

Step 1

The recode procedure combines codes from one or more existing variables into different groupings for a new variable. In our example, the codes and frequencies for the original Hispanic variable, which is named "ETHNIC," are shown in Figure 1. We see that the original variable includes four Hispanic categories and one category for non-Hispanic. Referring to these codes, we complete our first step of creating a new Hispanic variable:

From SDA Menu: Select the Recode Variables option. We will return to this menu to use "Compute a New Variable" and "Comparison of Means."

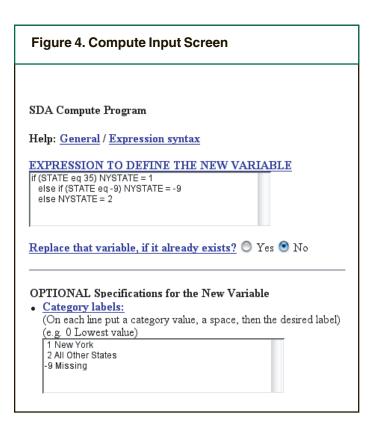*New Variable Name:* Enter the new variable name; we call the new variable "HISP."

You may choose to replace that variable if it already exists.[5]
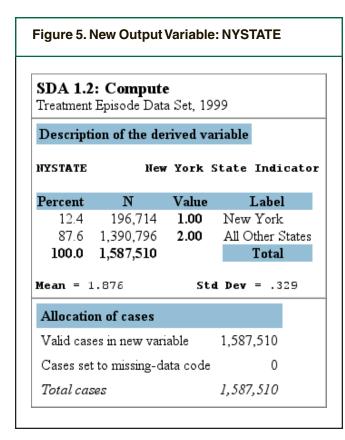
Input Variables: Enter the original variable name in "Input Variable #1" (ETHNIC). In our example, we have only one input variable. The DAS is not case-sensitive regarding variable names.
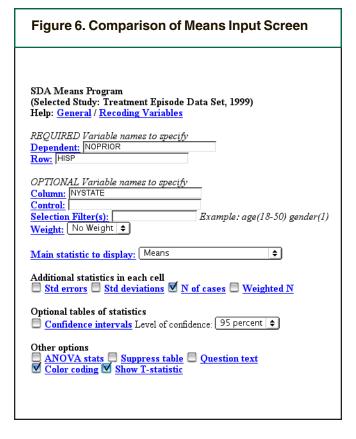
Recoding Rules: Designate the new output codes and labels in this section along with the input codes for the recode. The input codes (1-4) for ETHNIC are collapsed to form the output code (1) for HISP. The input code (5) is used for the output code (0) for non-Hispanic and -9 is used for missing data for both the original and new variables.

Start Recoding: Click on this button to begin recoding. Figure 2 shows an excerpt from the recode input screen, while Figure 3 shows the new codes and frequencies.

Figure 4. Compute Input Screen

SDA Compute Program

Help: <u>General</u> / <u>Expression syntax</u>

<u>EXPRESSION TO DEFINE THE NEW VARIABLE</u>

```
if (STATE eq 35) NYSTATE = 1
  else if (STATE eq -9) NYSTATE = -9
  else NYSTATE = 2
```

<u>Replace that variable, if it already exists?</u> ○ Yes ● No

OPTIONAL Specifications for the New Variable
- <u>Category labels:</u>
  (On each line put a category value, a space, then the desired label)
  (e.g. 0 Lowest value)

```
1 New York
2 All Other States
-9 Missing
```

Figure 5. New Output Variable: NYSTATE

SDA 1.2: Compute
Treatment Episode Data Set, 1999

Description of the derived variable

NYSTATE   New York State Indicator

Percent	N	Value	Label
12.4	196,714	1.00	New York
87.6	1,390,796	2.00	All Other States
100.0	1,587,510		Total

Mean = 1.876   Std Dev = .329

Allocation of cases

Valid cases in new variable	1,587,510
Cases set to missing-data code	0
Total cases	1,587,510

Figure 6. Comparison of Means Input Screen

SDA Means Program
(Selected Study: Treatment Episode Data Set, 1999)
Help: General / Recoding Variables

REQUIRED Variable names to specify
Dependent: NOPRIOR
Row: HISP

OPTIONAL Variable names to specify
Column: NYSTATE
Control:
Selection Filter(s):   Example: age(18-50) gender(1)
Weight: No Weight

Main statistic to display: Means

Additional statistics in each cell
☐ Std errors ☐ Std deviations ☑ N of cases ☐ Weighted N

Optional tables of statistics
☐ Confidence intervals Level of confidence: 95 percent

Other options
☐ ANOVA stats ☐ Suppress table ☐ Question text
☑ Color coding ☑ Show T-statistic

Step 2

The compute procedure is designed primarily for manipulating variables mathematically. For example, you would use the compute procedure to create a new variable that is the sum of a set of existing variables by entering an algebraic expression to define the new variable. You may also use 'if-then-else' logic to collapse variables into a smaller number of categories, as with the recode procedure. This example illustrates that capability. The original variable for state (named "STATE") includes 52 codes, one for each state, plus the District of Columbia and Puerto Rico. Again, -9 indicates that data are missing. In Step 2, compute a new variable called "NYSTATE" that includes two codes, one for New York and one for all other states:

From SDA Menu: Select Compute New Variable.

Expression to Define New Variable: Enter the command syntax directly into the "Expression" box. Refer to the Help links for syntax examples such as using 'ge' for 'greater than or equal to.'

Options: Click on your preferences for replacing an existing variable and handling missing data codes.

Category Labels: Specify the new labels for each code.

Start Computing: Click on this button to begin the computations. Figure 4 shows an excerpt from the compute input screen and Figure 5 shows the new codes and frequencies.

Step 3

Our final step is to use the comparison of means function to examine differences in the average number of prior treatments among Hispanics and non-Hispanics in New York compared to the rest of the U.S.:

From SDA Menu: Select Comparison of Means.

Dependent Variable: Enter the variable name for number of prior treatment episodes, which is NOPRIOR.

Row/Column Variables: Enter HISP as the row variable and NYSTATE as the column variable.

Main Statistic: Accept the default of Means.

Other Options: Choose your preferences for other options; we have chosen color-coding and the T-statistic. Optional statistics such as analysis of variance and confidence intervals also may be selected. Figure 6 shows an excerpt of the input screen.

Results

The color-coding shown in Figure 7 is based on the T-statistic for each cell and represents the observed differences for the average number of prior treatments. The red boxes indicate that the means are larger than the overall average and the blue boxes indicate the means are smaller than the overall average. We see from the T-statistics and the color-coding that the differences between the cells is statistically significant, with the mean number of prior treatments in New York (1.82) being larger than the overall mean for the rest of the U.S. (1.42). We also see that the highest average number of prior treatment episodes observed for any cell was for non-Hispanics in New York (1.86).[6] Other TEDS variables, such as race and primary substance of abuse, could be used to further explore differences among treatment admissions.

Additional Information

For additional information about the DAS or the databases publicly available from the archive, email: samhda@icpsr.umich.edu or phone toll-free (888) 741-7242.

Figure 7. Comparison of Means Results

SDA 1.2: Means
Treatment Episode Data Set, 1999

Main Statistics		NYSTATE		
Cells contain: -Means -N of cases -T-statistic		1.00 New York	2.00 All Other States	ROW TOTAL
HISP	0 Non-Hispanic	1.86 160,111 87.4	1.41 903,636 -38.8	1.48 1,063,747 -1.5
	1 Hispanic	1.64 36,575 19.2	1.46 126,430 -5.0	1.50 163,005 3.6
	COL TOTAL	1.82 196,686 87.6	1.42 1,030,066 -37.9	1.48 1,226,752 ---

Color coding:	<-2.0	<-1.0	<0.0	>0.0	>1.0	>2.0	T
Mean in each cell:	Smaller than average			Larger than average			

Endnotes

[1] The archive is supported by the Office of Applied Studies at SAMHSA and based at the Inter-university Consortium for Political and Social Research (ICPSR) through a subcontract with the National Opinion Research Center at the University of Chicago (NORC).

[2] The DAS is based on the Survey Documentation and Analysis (SDA) software developed at the Computer-assisted Survey Methods Program (CSM) at the University of California at Berkeley.

[3] Note that recodes may also be done "on the fly" within the analytic programs. However, this option requires input for each analytic screen and is limited to recodes of a single variable.

[4] You may want to print your input screens or save your syntax in a text file to recreate the variable(s) if you do not plan to use your created variables within 30 days.

[5] You may use the recodes that others have created and you may choose to overwrite your own recodes, but please do not overwrite the recodes of other users. Select the "List user-created variables in temporary storage area" on the SDA menu to view variables already created.

[6] The variable NOPRIOR (number of prior treatments) was top-coded to group together 5 or more prior treatments as a single interval amount of 5. This has the effect of bringing the mean slightly downward.