Using the DAS for Windows: How to Produce a Correlation Matrix for a Linear Regression Model

Helen Jang
Andrea Livingston
Shirley He
MPR Associates, Inc.

December 2002

TABLE OF CONTENTS

INTRODUCTION

The Data Analysis System (DAS) is a Windows software application that provides users with access to Department of Education survey data. By using the DAS, you can create programming instruction files (DAS files) that specify the information you want to display in a table. This table will contain the estimates (usually percentages of students) and corresponding standard errors that are calculated taking into account the complex sampling designs used in NCES surveys. In addition, the DAS can create correlation matrices that can be used as input for most popular statistical programs to conduct multivariate analysis.

A basic knowledge of SPSS or SAS and a general understanding of multiple regressions are recommended prior to using the correlation matrix option of the DAS. For information about SPSS or SAS, visit www.spss.com or www.sas.com.

TUTORIAL

About This Tutorial

This tutorial will lead you through the five steps necessary to create the regression model for a specific table in a NCES report, titled *Debt Burden Four Years After College* (NCES 2000–188). The next section will present some important background information about the report, providing a context for the steps that follow.

Background

Before you begin creating the regression model for table 3 in *Debt Burden Four Years After College*, it will be useful to have some background information about the report. This report examines the debt of 1992–93 bachelor's degree recipients in light of their financial circumstances in 1997, about four years after they earned their degree. One purpose of the study is to describe debt burden by examining how student loan payments are related to income and by searching for other indications of the impact of borrowing.

Table 2 of this report shows that undergraduate borrowing appears to have a minor discouraging effect on graduate degree enrollment. For example, undergraduate borrowers were slightly less likely than nonborrowers to have enrolled in a graduate degree program by 1997 (27 percent versus 30 percent). Because this finding does not take into account the various other factors that affect graduate degree enrollment or undergraduate borrowing, this linear regression model was used to describe the relationship between undergraduate borrowing and graduate degree enrollment while adjusting for the covariance of independent variables.

The analysis for the *Debt Burden* report examined one dependent variable and eight independent variables. The dependent variable was defined as students' likelihood of enrolling in a graduate degree program between the time they graduated and were interviewed in 1997. The independent variables included sex, race/ethnicity, age at degree receipt, whether they borrowed from any source for their undergraduate education, type of institution from which they graduated, undergraduate major, cumulative undergraduate grade-point average (GPA), and parents' education.

Note You can obtain a copy of *Debt Burden Four Years After College* by visiting the NCES Web site at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2000188.

See Appendix A-1 for Table 3 from *Debt Burden Four Years After College*

Programming Instructions

Before you create a Correlation Matrix File (CPF), it is helpful to prepare programming instructions. These instructions should provide answers to the following questions:

- Which data set will you use?
- Will you use a weight variable? If so, which one?
- Will you use a filter?
- What is the dependent variable?
- What are the independent variables and comparison group for each variable?

Note See next page for the programming instructions used for this regression model.

See Appendix B-1 for information about tags and their usage.

Programming Instructions for this Regression Model

Data set: B&B:97 Baccalaureate and Beyond Second Followup (B&B:97)

Weight: BNBPANEL

Filter: B2ETHNIC Respondent ethnicity [1+2+3+4+5]
* [6] Others will be excluded from this model

Dependent (column) variable:
B2HENPRG Highest degree program enrolled after BA
[4+5+6+7+8] Graduate degree program

Independent (row) variables:

GENDER: Gender of student
[1] Male (Comparison group)
[2] Female

B2AGATBA: Respondent age when received BA
[13-24] 24 years or younger (Comparison group)
[25-29] 25-29 years
[30+] 30 years or older

B2TOTUDB: Total undergraduate debt
[0] Did not borrow (Comparison group)
[1-4,999] Less than $5,000
[5,000+] $5,000 or more

NORMGPA: Normalized GPA on a 4.0 scale
[1-299] Less than 3.0
[300+] 3.0 or higher (Comparison group)

SECTOR_B: Institutional type
[3+4] Public 4-year (Comparison group)
[6+7] Private, nfp, 4-year
[1+2+5+8+9] Others

B2ETHNIC Race
[1] American Indian
[2] Asian
[3] Black
[4] Hispanic
[5] White (Comparison group)
[6] Others (do not need this category)

B2BAMAJR Bachelors degree field recoded
[1] Business and management
[3+6+7] Engineering, mathematics, or science (Comparison group)
[8+9+10+11] Humanities / social science
[2+4+5+12] Others

PAREDUC Highest education level by either parent
[1+2+3] High school or less
[4+5+6+7+8+9] Some postsecondary
[10+11+12+13+14] Bachelor’s or advanced degree (Comparison group)

Step 1: Getting Started

The DAS can be run in two different modes: DAS-T which allows you to create TPFs for tables and DAS-C which allows you to create CPFs for correlation matrices. For this tutorial we will use DAS-C.

In the first step, you'll begin the process of creating the regression model by learning how to launch the DAS in the correlation matrix mode and choose the output mode.

➨ To Launch the B&B:97 DAS in Correlation Matrix Mode

1. Click 

and select Run.
2. In the command line, type:
   C:\Dasw\DASW.EXE C=B97
3. Click OK.

Tip You can create a shortcut for the DAS-C mode in your Start menu. See Appendix B-3 for instructions.

➨ To Choose the Output Mode

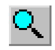
1. From the File menu, select Setup.
2. Under the Correlation Output Mode, select SPSS.
3. Click OK.

Note If you are more familiar with SAS, you can choose it instead. In this example, we use SPSS.

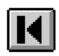
Step 2: Tagging the Variables

In this step, you will learn how to tag variables based on the programming instructions for creating a CPF (see page 4). (For more information on tags and how to use them, refer to Appendix B-1.) Before you can tag a variable, you must first locate it in the Description window.

Search for Variables

1. From the Main menu, click Search, or click  on the toolbar.
2. In the Text to Find field, type the variable name.
3. Under the Search From submenu, select Variable.
4. Select Whole Words.
5. Click Search Next.

A black outline appears around the variable that matches your search criteria.

Tip The DAS always starts its search from the cursor insertion point (designated by a black rectangle), and moves either forward or backward depending on the direction you specify. For best results, click the  button, which takes you to the top of the variables list, and then initiate a forward search.

Tag the Variables

The Tag menu will appear when you right-click a variable. Follow the instructions below to create the tags according to the programming instructions. After a variable is tagged, an **X** appears in the box next to the variable in the Variable List, and the processing instructions (parameters) for the tag appear in the Parameter window.

Weight and Filter Variables

When a data set contains more than one weight variable, the analyst will provide you with the appropriate weight to use. Filter works like "select if;" filter will keep the sample you selected.

BNBPANEL: Weight tag

In this example, we are using the B&B panel weight.

➡ **To create a weight (W) tag**

From the Tag menu, select Weight.

B2ETHNIC: And_Filter tag

This filter is used to exclude category 6, "Others," from the model. To do this, we tag the first 5 categories but not the 6th one.

➡ **To create a Filter (F) tag**

1. From the Tag menu, select And_Filter.
2. In the first row of the Value field, type:
   1
3. Press the Tab key to go to the label (or click the Label field).
   The label for this category will appear in the Label field.
4. Repeat steps 2 and 3 to select categories 2–5.
5. Click OK.

Dependent Variable

The dependent variable must be either "continuous" (such as family income), or if it is "categorical," it must be coded into two groups (Yes / No). In this example, the dependent variable for the regression is B2HENPRG: Enrolled in graduate education (Enrolled / Did Not Enroll).

B2HENPRG: Lump tag

A Lump tag is used to combine categories 4–8 into one group, Graduate Degree Program (those who enrolled). The remaining untagged categories, 1–3, become the comparison group, Did Not Enroll. The percentages for these two groups will add up to 100. For more information about the Lump tag, see Appendix B-1.

➡ **To create a Lump (L) tag**

1. In the Title field, type:
   Highest degree program enrolled after BA
2. Click the Edit Lump button.
3. Click Master's degree to Doctoral degree to select categories 4–8.
4. In the Lump Title field, type:
   Graduate degree program
5. Click the Save Lump button.
6. Click the Save and Close button.

Independent Variables

Independent variables are defined as either “dummy” (categorical) variables or continuous variables. Dummy variables represent each defined categorical group of a variable. (Such as a specified age group, age 25-29.) In most cases, the biggest group is not defined as a dummy and is used as the comparison group. Each defined group is tagged so that it can be compared with the comparison group. The untagged group becomes the comparison group by default. Values that are not needed for comparison should be defined as a separate group so they will not be considered as part of the comparison group. For the Cut and Lump tags, missing values must also be identified to exclude them from the comparison. The default missing values code is –1.

Eight independent variables are used in this example. One detailed example is provided for each of the following tags: Each, Cut, and Lump. Please use these examples to tag the remaining variables.

GENDER: Each tag

This variable has only two categories, male and female. Since the categories will be compared with each other and all categories will be used, use an Each tag here.

➡ **To create an Each (E) tag**

1. Select 1 from the Base Category Code.
2. Click OK.

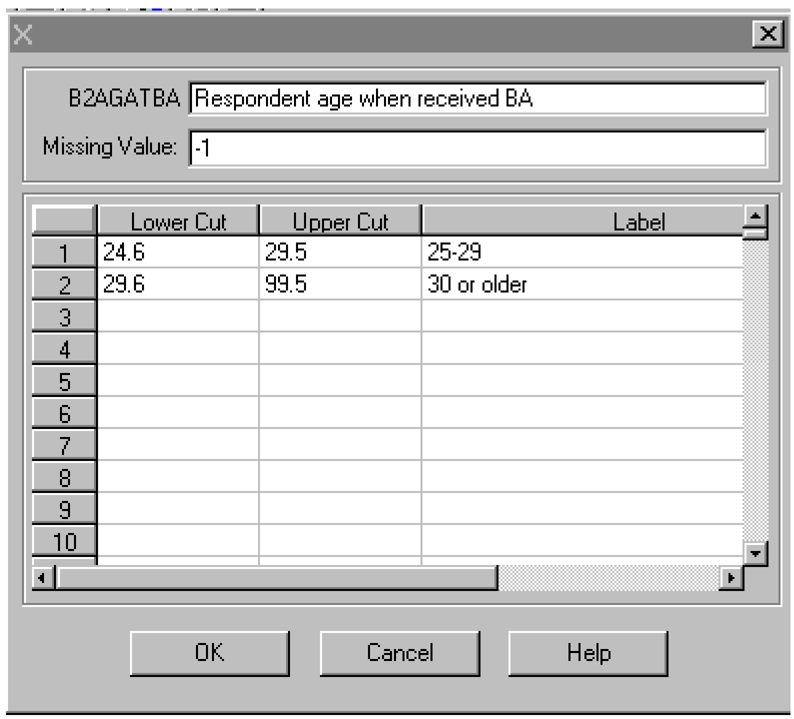
Note The Base Category Code is the value of the comparison group; in this case, 1 is male.

B2AGATBA: Cut tag

This is a continuous variable and therefore must use the Cut tag. The decimal cut point can range from 0.1 to 0.9. The upper cut for the last group can be any value above the largest value for that variable. All cut-point values must be a numeral with a decimal point.

➡ **To create a Cut (X) tag**

1. In the Missing Value field, type:
   –1
2. Enter the Lower Cut and Upper Cut values, and the information for Label.
   See the dialog box below.
3. Click OK.



More Cut Tags

You will also need to create Cut tags for the following variables using the values in the programming instructions on page 4.

B2TOTUDB: Cut tag

NORMGPA: Cut tag

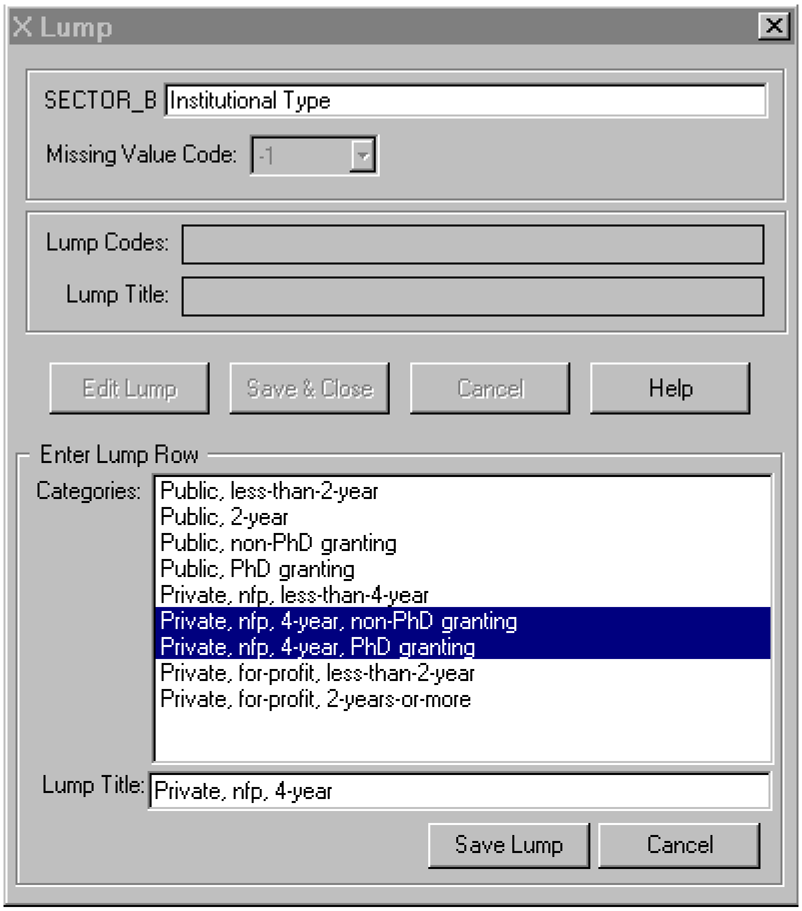
TIP You can check the values for a variable by double-clicking the variable in the Variable List and selecting View Percentages from the Description / Percentages pop-up window.

SECTOR_B: Lump tag

Each lumped group must be tagged as a separately. For this variable, you will create two Lump Tags: Private, nfp, 4-year; and Others. The comparison group is Public 4-year PhD granting and non-PhD granting institutions, which are the untagged categories. Follow the instructions below to make the first lump tag.

➡ To create the Lump (L) tags

1. In the Title field, type:
   Institutional type
2. Click the Edit Lump button.
3. Select categories.
   As shown in the Lump dialog box below.
4. In the Lump Title field, type:
   Private, nfp, 4-year
5. Click the Save Lump button.
6. Click the Save and Close button.



Repeat the steps above to create the second Lump tag for categories 1, 2, 5, 8, and 9.

More Lump Tags

You will also need to create Lump tags for the following variables using the values in the programming instructions on page 4.

B2ETHNIC: Lump tag

For this variable, you will need to create four Lump tags; each tag will contain only one category. Categories 5 and 6 should be left untagged. Because category 6 is excluded by the Filter tag, category 5 becomes the comparison group.

Tip You can also use the Each tag for this variable and ignore category 6.

B2BAMAJR: Lump tag

PAREDUC: Lump tag

Note See Appendix A-2 for the completed CPF file.

Step 3: Saving and Running the CPF File

In this step, you'll learn the procedures for saving the CPF file and then running it.

➡ To save the CPF file

1. From the File menu, select Save As.
2. Type in the file name:
   Example.cpf
3. Click Save.
4. Enter a title:
   Example
5. Click Save.

Note By default, the CPF file will be saved in C:\Dasw\B97. The CPF file name and location now appear at top of the Parameter window.

➡ To run the CPF file

1. Compress your CPF file into a .ZIP file.
2. Submit (drop off) the ZIP file to the DAS FTP site for processing.
3. Retrieve (pick up) the ZIP file containing the processed file.

For detailed information about creating, submitting, and retrieving these files, visit the NCES DAS Web site (http://nces.ed.gov/das/das_windows/).

Step 4: Running the Correlation Matrix in SPSS

Before running the SPSS file, you must rename and modify it.

➡ To rename the file

1. Unzip the processed file.
2. Rename Rmatrix.prn to **Example.sps**.
   This is the SPSS syntax (program) file that contains the correlation matrix for generating the regression model.
3. Save the file in C:\Dasw\B97.

Note See Appendix A-3 to view **Example.sps**.

➡ To modify the file

1. Open example.sps in SPSS.
2. Find the section that begins with the command:
   REGRESSION MATRIX
3. Delete the blank line above /MISSING=PAIRWISE.
4. Delete the blank line above /DEP=.
5. Copy the dependent variable B2HENP1 from the /VARIABLES= section and paste it into the /DEP= section.
6. On the line /METHOD=, type:
   enter.

Note This is the default syntax for including all the independent variables in the model. If you want only certain variables, you must type them in after the word "enter." There must be a period at the end of this line.

---

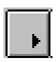
7. The modified REGRESSION MATRIX section should now look like this:

```
REGRESSION MATRIX=IN(*)
 /VARIABLES=
   B2HENP1 GENDER1 B2AGAT1 B2AGAT2 B2TOTU1 B2TOTU2 NORMGP1 SECTOR1
   B2ETHN1 B2ETHN2 B2ETHN3 B2ETHN4 B2BAMA1 B2BAMA2 B2BAMA3 PAREDU1
   PAREDU2
 /MISSING=PAIRWISE
  /DEP=B2HENP1
 /METHOD=enter.
```

Note The variables in Example.sps have slightly different names from those in the CPF file because they are dummy variables representing the different tags. The variable names contain only the first six characters of the original name and a number at the end. For example, B2AGATBA (Respondent age when received BA) was divided into two groups: B2AGAT1 and B2AGAT2.

A possible problem may arise if the two variables in the model have the same first six characters in their names. If this occurs, you must change the name of one of the variables throughout Rmatrix.prn.

➡ To run the file in SPSS

1. Highlight the program from the line
   MATRIX DATA VARIABLES=ROWTYPE_
   to the line
   /METHOD=enter.
2. Click  on the toolbar or press CTRL-R.

➡ To save the file

1. Go to the Output window.
2. From the File menu, select Save As.
3. In the file name field, type:
   Example.lst
4. Save the file in C:\Dasw\B97.

Note Go to Appendix A-4 to see what the file **Example.lst** looks like.

The following files should now be in the directory C:\Dasw\B97:

- **Example.cpf**
- **Example.sps**
- **Example.lst**

Step 5: Adjusting Means and Standard Errors

This section contains a simplified example of the procedures that were used to estimate adjusted means and standard errors for table 3 in the report *Debt Burden Four Years After College*.

For this example, the following simplified CPF file was created.

```
VCBB:97 Baccalaureate and Beyond Second Followup 06/01/00
W BNBPANEL Panel weight for NPSAS and B&B
F B2ETHNIC Respondent ethnicity
 1 American Indian/Alaskan native
 2 Asian or Pacific Islander
 3 Black, non-Hispanic
 4 Hispanic
 5 White, non-Hispanic
L B2HENPRG Highest degree program enrolled after BA
 -1 4+5+6+7+8 Grad degree program
E GENDER Gender of student
 2 Female
X NORMGPA Normalized GPA on a 4.0 scale
 -1 0.5 299.5 Less than 3.0
O SPSS_WIN
Title Example 2
```

Consider a hypothetical case in which two variables—Gender and Normalized GPA—are used to describe an outcome, $\hat{Y}$ (such as graduate school enrollment). The variables gender and Normalized GPA are recoded into dummy variables representing gender, *G*, and a dummy variable representing Normalized GPA, *N*:

Gender	G
Female	1
Male	0
and	
Normalized GPA	N
Less than 3.0	1
3.0 or above	0

The following regression equation is then estimated using the results from the SPSS Regression Procedure and the means from the correlation matrix output from the DAS:

$$\hat{Y} = \mathrm{a} + \mathrm{b}_1 G + \mathrm{b}_2 N \qquad (1)$$

Where

- $\hat{Y}$ is the adjusted percentage or mean (in this case the percentage enrolling in graduate school);
- a is the intercept from the regression model;
- b_1 is the regression coefficient for the dummy variable representing students who are female;
- G is the proportion of students who are female;
- b_2 is the regression coefficient for graduates with a GPA of less than 3.0; and
- N is the proportion of graduates with a GPA of less than 3.0.

The proportions G and N are found in the table at the bottom of the DAS output under the column label "mean" (see Appendix A-3) while (a, b_1 and b_2) are the regression output produced by the SPSS program (see Appendix A-4).

To estimate the adjusted mean for any subgroup evaluated at the mean of all other variables, one substitutes the appropriate values for that subgroup's dummy variables (1 or 0) and the mean for the dummy variable(s) representing all other subgroups. For example, suppose $\hat{Y}$ represents graduate school enrollment and is being described by gender (*G*) and normalized GPA (*N*), coded as shown above, with means as follows:

Variable	Mean
G	0.546
N	0.439

Next, suppose the regression equation results in:

$$\hat{Y} = 0.376 + (-0.019)G + (-0.153)N \quad (2)$$

To estimate the adjusted value for older students, one substitutes the appropriate parameter estimates and variable values into equation 2.

Variable	Parameter	Value
a	0.376	—
G	-0.019	1.000
N	-0.153	0.439

This results in:

$$\hat{Y} = 0.376 + (-0.019)(1) + (-0.153)(0.439) = 0.290$$

In this case, the adjusted mean for women is 0.290 and represents the expected outcome for the expected likelihood of attaining a degree for female students who resemble the average student with respect to the other variables in the model (in this example, normalized GPA). In other words, the adjusted percentage of who enrolled in graduate school after controlling for normalized GPA is 29 percent (0.290 x 100 for conversion to a percentage).

It is relatively straightforward to produce a multivariate model using the DAS correlation matrix, computed using pairwise missing values. In regression analysis, there are several common approaches to the problem of missing data. The two simplest are pairwise deletion of missing data and listwise deletion of missing data. In pairwise deletion, each correlation is calculated using all of the cases for the two relevant variables. For example, suppose you have a regression analysis that uses variables X1, X2, and X3. The regression is based on the correlation matrix between X1, X2, and X3. In pairwise deletion the correlation between X1 and X2 is based on the nonmissing cases for X1 and X2. Cases missing on either X1 or X2 would be excluded from the calculation of the correlation. In listwise deletion, the correlation between X1 and X2 would be based on the nonmissing values for X1, X2, and X3. That is, all of the cases with missing data on any of the three variables would be excluded from the analysis.[1]

The correlation matrix can be used by most statistical software packages as the input data for least squares regression. That is the approach used for the *Debt Burden* report, with an additional adjustment to incorporate the complex sample design into the statistical significance tests of the parameter estimates (described below). For tabular presentation, parameter estimates and standard errors were multiplied by 100 to match the scale used for reporting unadjusted and adjusted percentages.

Most statistical software packages assume simple random sampling when computing standard errors of parameter estimates. Because of the complex sampling design used for the NPSAS and B&B surveys, this assumption is incorrect. A better approximation of their standard errors is to multiply each standard error by the design effect associated with the dependent variable from the SPSS regression model (found at the bottom of the DAS output (see Appendix B-2) under

---

[1]Although the DAS simplifies the process of making regression models, it also limits the range of models. Analysts who wish to use an approach other than pairwise treatment of missing values or to estimate probit/logit models (which are the most appropriate for models with categorical dependent variables) can apply for a restricted data license from NCES. See John H. Aldrich and Forrest D. Nelson, "Linear Probability, Logit and Probit Models," *Quantitative Applications in Social Sciences*, Vol. 45 (Beverly Hills, CA: Sage, 1984).

the column labeled "DEFT"[2]) where the DEFT is the ratio of the true standard error to the standard error computed under the assumption of simple random sampling.

[2]The adjustment procedure and its limitations are described in C.J. Skinner, D. Holt, and T.M.F. Smith, eds., *Analysis of Complex Surveys* (New York: John Wiley & Sons, 1989).

APPENDICES

Appendix A-1. Table 3 from *Debt Burden Four Years After College*

Table 3—Percentage of 1992–93 bachelor's degree recipients who enrolled in a graduate degree program by 1997 and the adjusted percentage after taking into account the covariation of the variables listed in the table

	Unadjusted percentages[1]	Adjusted percentages[2]	Least squares coefficient[3]	Standard error[4]
Total	29.8	29.8	51.9	3.1
Age received bachelor's degree				
24 years or younger	*31.9*[5]	*31.5*	†	†
25–29 years	21.7*	24.6*	-6.9	3.1
30 years or older	26.5*	26.2	-5.3	2.8
Bachelor's degree major				
Business and management	16.4*	17.4*	-22.9	3.2
Engineering, mathematics, or science	*40.8*	*40.3*	†	†
Humanities/social science	35.5*	34.4	-5.9	3.2
Others	29.5*	29.7*	-10.6	3.0
Race/ethnicity				
American Indian/Alaskan Native	20.8	27.0	-2.0	12.7
Asian/Pacific Islander	31.0	29.0	0.1	4.7
Black, non-Hispanic	31.6	37.8*	8.8	4.1
Hispanic	32.5	35.3	6.4	4.5
White, non-Hispanic	*29.5*	*28.9*	†	†
Amount borrowed for undergraduate education				
Did not borrow	*31.4*	*30.9*	†	†
Borrowed				
Less than $5,000	28.6	29.0	-1.9	3.0
$5,000 or more	27.9*	28.6	-2.3	2.2
Sex				
Male	*29.7*	*30.6*	†	†
Female	29.8	29.1	-1.5	2.0
Grade point average				
Less than 3.0	21.4*	21.2*	-15.4	2.0
3.0 or higher	*36.5*	*36.6*	†	†
Bachelor's degree-granting institution				
Public 4-year	*28.5*	*29.0*	†	†
Private, not-for-profit 4-year	33.0*	32.1	3.1	2.2
Other	24.9	24.3	-4.7	5.5
Parents' highest education				
High school or less	24.6*	26.4*	-6.3	2.3
Some postsecondary	26.9*	27.9	-4.8	2.7
Bachelor's or advanced degree	*34.2*	*32.7*	†	†

*p < .05.

†Not applicable for the reference group.

[1]The estimates are from the B&B:1993/1997 Data Analysis System.

[2]The percentages are adjusted for differences associated with other variables in the table (see report appendix B).

[3]Least squares coefficient, multiplied by 100 to reflect percentage (see report appendix B).

[4]Standard error of least squares coefficient, adjusted for design effect, multiplied by 100 to reflect percentage (see report appendix B).

[5]The italicized group in each category is the reference group being compared.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 1993 Baccalaureate and Beyond Longitudinal Study, Second Follow-up (B&B:1993/1997), Data Analysis System.

Appendix A-2. Correlation Parameter File: Example.cpf

```
VCBB:97 Baccalaureate and Beyond Second Followup 06/01/00
W BNBPANEL Panel weight for NPSAS and B&B
F B2ETHNIC Respondent ethnicity
 1 American Indian/Alaskan native
 2 Asian or Pacific Islander
 3 Black, non-Hispanic
 4 Hispanic
 5 White, non-Hispanic
L B2HENPRG Highest degree program enrolled after BA
 -1 4+5+6+7+8 Grad degree program
E GENDER Gender of student
 2 Female
X B2AGATBA Respondent age when received BA
 -1 24.6 29.5 25-29
 -1 29.6 99.5 30 or older
X B2TOTUDB Total undergraduate debt
 -1 0.5 4999.5 Less than $5,000
 -1 4999.5 999999.5 $5,000 or more
X NORMGPA Normalized GPA on a 4.0 scale
 -1 0.5 299.5 Less than 3.0
L SECTOR_B Institutional type
 -1 6+7 Private, nfp, 4-year
L SECTOR_B Institutional type
 -1 1+2+5+8+9 others
L B2ETHNIC Race
 -1 1 American Indian
L B2ETHNIC Race
 -1 2 Asian
L B2ETHNIC Race
 -1 3 Black
L B2ETHNIC Race
 -1 4 Hispanic
L B2BAMAJR Bachelors degree field recoded
 -1 1 Business and management
L B2BAMAJR Bachelors degree field recoded
 -1 8+9+10+11 Humanities/Soc Sci
L B2BAMAJR Bachelors degree field recoded
 -1 2+4+5+12 Others
L PAREDUC Highest education level by either parent
 -1 1+2+3 High school or less
L PAREDUC Highest education level by either parent
 -1 4+5+6+7+8+9 Some college
O SPSS_WIN
Title Example
```

Appendix A-3. SPSS Correlation Matrix: Example.sps

```
* Source:  NCES, BB:97 Baccalaureate and Beyond Second Followup 06/01/00
* Computation by DAS-C  Version  3.0  on 11/22/2000
*  PLEASE EDIT THIS CODE;
*  CHECK MATRIX FOR -9.9 missing values;
*  AND separate CODE from spreadsheet import values;
* Example

* Filters:
* Respondent ethnicity=American Indian/Alaskan native.    (Value=1)
*           =Asian or Pacific Islander.    (Value=2)
*           =Black, non-Hispanic.    (Value=3)
*           =Hispanic.    (Value=4)
*           =White, non-Hispanic.    (Value=5)
*           Missing Values Excluded
*----------------------SPSS for Windows --------------------.

MATRIX DATA VARIABLES=ROWTYPE_
   B2HENP1 GENDER1 B2AGAT1 B2AGAT2 B2TOTU1 B2TOTU2
   NORMGP1 SECTOR1 SECTOR2 B2ETHN1 B2ETHN2 B2ETHN3
   B2ETHN4 B2BAMA1 B2BAMA2 B2BAMA3 PAREDU1 PAREDU2.
BEGIN DATA.
MEAN  0.2979 0.5460 0.1228 0.1645 0.1423 0.3551 0.4392 0.3132 0.0339 0.0059
   0.0463 0.0612 0.0498 0.2256 0.2327 0.3771 0.3145 0.1920
STDDEV  0.4573 0.4979 0.3282 0.3707 0.3494 0.4785 0.4963 0.4638 0.1810 0.0765
   0.2101 0.2397 0.2176 0.4180 0.4225 0.4847 0.4643 0.3939
CORR  1.0000
CORR  0.0006 1.0000
CORR  -0.0660 -0.0750 1.0000
CORR  -0.0319 0.0738 -0.1660 1.0000
CORR  -0.0105 0.0051 0.0332 0.0372 1.0000
CORR  -0.0299 -0.0129 0.0922 0.0227 -0.3022 1.0000
CORR  -0.1636 -0.1276 0.0769 -0.1553 -0.0084 0.0415 1.0000
CORR  0.0472 0.0276 -0.0681 0.0667 -0.0579 0.1132 -0.0981 1.0000
CORR  -0.0199 -0.0136 0.0260 0.0155 0.0013 0.0462 -0.0510 -0.1266 1.0000
CORR  -0.0150 0.0315 0.0112 0.0518 0.0123 0.0162 0.0092 -0.0333 0.0117 1.0000
CORR  0.0060 -0.0396 -0.0153 -0.0367 -0.0177 -0.0143 0.0015 -0.0289 0.1074
   -0.0169 1.0000
CORR  0.0102 0.0605 0.0039 0.0279 0.0602 0.0335 0.1324 0.0316 -0.0054 -0.0196
   -0.0562 1.0000
CORR  0.0134 0.0242 0.0574 0.0189 0.0726 -0.0052 0.0430 -0.0025 0.0015 -0.0176
   -0.0504 -0.0584 1.0000
CORR  -0.1581 -0.0874 0.0321 0.0424 0.0066 -0.0421 0.0486 0.0481 -0.0030
   0.0031 -0.0074 0.0288 -0.0213 1.0000
CORR  0.0687 0.0215 -0.0321 -0.0488 -0.0210 -0.0292 -0.0124 0.0542 -0.0171
   0.0010 -0.0101 -0.0056 0.0204 -0.2972 1.0000
CORR  -0.0052 0.2082 -0.0080 0.0509 -0.0024 0.0448 -0.0394 -0.0642 0.0088
   0.0073 -0.0588 -0.0106 0.0034 -0.4200 -0.4285 1.0000
CORR  -0.0764 0.0432 0.0618 0.2182 0.0422 0.0888 0.0034 -0.0232 0.0033 0.0301
   -0.0110 0.0650 0.0826 0.0689 -0.0557 0.0274 1.0000
CORR  -0.0302 0.0076 0.0087 -0.0280 0.0469 0.0487 0.0143 -0.0383 0.0758
   0.0048 0.0041 0.0140 -0.0062 0.0222 -0.0231 0.0101 -0.3302 1.0000
N_MATRIX  9183.0
N_MATRIX  9173.0 9206.0
N_MATRIX  9132.0 9156.0 9165.0
N_MATRIX  9132.0 9156.0 9165.0 9165.0
N_MATRIX  9135.0 9155.0 9122.0 9122.0 9165.0
N_MATRIX  9135.0 9155.0 9122.0 9122.0 9165.0 9165.0
N_MATRIX  8925.0 8947.0 8906.0 8906.0 8908.0 8908.0 8955.0
N_MATRIX  9183.0 9206.0 9165.0 9165.0 9165.0 9165.0 8955.0 9216.0
N_MATRIX  9183.0 9206.0 9165.0 9165.0 9165.0 9165.0 8955.0 9216.0 9216.0
N_MATRIX  9183.0 9206.0 9165.0 9165.0 9165.0 9165.0 8955.0 9216.0 9216.0
   9216.0
```

```
N_MATRIX  9183.0 9206.0 9165.0 9165.0 9165.0 9165.0 8955.0 9216.0 9216.0
    9216.0 9216.0
N_MATRIX  9183.0 9206.0 9165.0 9165.0 9165.0 9165.0 8955.0 9216.0 9216.0
    9216.0 9216.0 9216.0
N_MATRIX  9183.0 9206.0 9165.0 9165.0 9165.0 9165.0 8955.0 9216.0 9216.0
    9216.0 9216.0 9216.0 9216.0
N_MATRIX  9183.0 9206.0 9165.0 9165.0 9165.0 9165.0 8955.0 9216.0 9216.0
    9216.0 9216.0 9216.0 9216.0 9216.0
N_MATRIX  9183.0 9206.0 9165.0 9165.0 9165.0 9165.0 8955.0 9216.0 9216.0
    9216.0 9216.0 9216.0 9216.0 9216.0 9216.0
N_MATRIX  9183.0 9206.0 9165.0 9165.0 9165.0 9165.0 8955.0 9216.0 9216.0
    9216.0 9216.0 9216.0 9216.0 9216.0 9216.0 9216.0
N_MATRIX  8746.0 8775.0 8730.0 8730.0 8734.0 8734.0 8533.0 8775.0 8775.0
    8775.0 8775.0 8775.0 8775.0 8775.0 8775.0 8775.0 8775.0
N_MATRIX  8746.0 8775.0 8730.0 8730.0 8734.0 8734.0 8533.0 8775.0 8775.0
    8775.0 8775.0 8775.0 8775.0 8775.0 8775.0 8775.0 8775.0 8775.0
END DATA.

VAR LABELS
  B2HENP1  'Highest degree program enrolled after BA=Grad degree program' /
  GENDER1  'Gender of student=Female' /
  B2AGAT1  'Respondent age when received BA=25-29' /
  B2AGAT2  'Respondent age when received BA=30 or older' /
  B2TOTU1  'Total undergraduate debt=Less than $5,000' /
  B2TOTU2  'Total undergraduate debt=$5,000 or more' /
  NORMGP1  'Normalized GPA on a 4.0 scale=Less than 3.0' /
  SECTOR1  'Institutional type=Private, nfp, 4-year' /
  SECTOR2  'Institutional type=others' /
  B2ETHN1  'Race=American Indian' /
  B2ETHN2  'Race=Asian' /
  B2ETHN3  'Race=Black' /
  B2ETHN4  'Race=Hispanic' /
  B2BAMA1  'Bachelors degree field recoded=Business and management' /
  B2BAMA2  'Bachelors degree field recoded=Humanities / Soc Sci' /
  B2BAMA3  'Bachelors degree field recoded=Others' /
  PAREDU1  'Highest education level by either parent=High school or less' /
  PAREDU2  'Highest education level by either parent=Some college'.

REGRESSION MATRIX=IN(*)
  / VARIABLES=
    B2HENP1 GENDER1 B2AGAT1 B2AGAT2 B2TOTU1 B2TOTU2 NORMGP1 SECTOR1
    SECTOR2 B2ETHN1 B2ETHN2 B2ETHN3 B2ETHN4 B2BAMA1 B2BAMA2 B2BAMA3
    PAREDU1 PAREDU2

  / MISSING=PAIRWISE

   / DEP=
   / METHOD=

"* Filters:"
"* Respondent ethnicity=American Indian / Alaskan native.  ","   (Value=1)"
"*          =Asian or Pacific Islander.  ","   (Value=2)"
"*          =Black, non-Hispanic.  ","   (Value=3)"
"*          =Hispanic.  ","   (Value=4)"
"*          =White, non-Hispanic.  ","   (Value=5)"
"*          Missing Values Excluded"

"MEAN","S.E.","DEFT","VAR","LABEL"
     0.2979,   0.0066,   2.0268,"B2HENP1","Highest degree program enrolled after BA=Grad degree program"
     0.5460,   0.0072,   2.0277,"GENDER1","Gender of student=Female"
     0.1228,   0.0052,   1.8621,"B2AGAT1","Respondent age when received BA=25-29"
     0.1645,   0.0075,   1.4546,"B2AGAT2","Respondent age when received BA=30 or older"
     0.1423,   0.0045,   2.2574,"B2TOTU1","Total undergraduate debt=Less than $5,000"
     0.3551,   0.0073,   1.9094,"B2TOTU2","Total undergraduate debt=$5,000 or more"
     0.4392,   0.0080,   1.8528,"NORMGP1","Normalized GPA on a 4.0 scale=Less than 3.0"
```

```
0.3132,   0.0109,   1.2410,"SECTOR1","Institutional type=Private, nfp, 4-year"
0.0339,   0.0052,   1.0179,"SECTOR2","Institutional type=others"
0.0059,   0.0000,   0.0000,"B2ETHN1","Race=American Indian"
0.0463,   0.0048,   1.2749,"B2ETHN2","Race=Asian"
0.0612,   0.0047,   1.4804,"B2ETHN3","Race=Black"
0.0498,   0.0038,   1.6679,"B2ETHN4","Race=Hispanic"
0.2256,   0.0080,   1.5170,"B2BAMA1","Bachelors degree field recoded=Business and management"
0.2327,   0.0067,   1.8503,"B2BAMA2","Bachelors degree field recoded=Humanities / Soc Sci"
0.3771,   0.0081,   1.7537,"B2BAMA3","Bachelors degree field recoded=Others"
0.3145,   0.0079,   1.7641,"PAREDU1","Highest education level by either parent=High school or less"
0.1920,   0.0060,   1.9630,"PAREDU2","Highest education level by either parent=Some college"
```

Appendix A-4. SPSS Output File: Example.lst

```
-> MATRIX DATA VARIABLES=ROWTYPE_
->     B2HENP1 GENDER1 B2AGAT1 B2AGAT2 B2TOTU1 B2TOTU2
->     NORMGP1 SECTOR1 SECTOR2 B2ETHN1 B2ETHN2 B2ETHN3
->     B2ETHN4 B2BAMA1 B2BAMA2 B2BAMA3 PAREDU1 PAREDU2.
MATRIX DATA has already allocated 760 bytes.
More memory will be allocated to store the data to be read.


-> BEGIN DATA.

-> MEAN  0.2979 0.5460 0.1228 0.1645 0.1423 0.3551 0.4392 0.3132 0.0339 0.0059
->     0.0463 0.0612 0.0498 0.2256 0.2327 0.3771 0.3145 0.1920
-> STDDEV  0.4573 0.4979 0.3282 0.3707 0.3494 0.4785 0.4963 0.4638 0.1810 0.0765
->     0.2101 0.2397 0.2176 0.4180 0.4225 0.4847 0.4643 0.3939
-> CORR  1.0000
-> CORR  0.0006 1.0000
-> CORR  -0.0660 -0.0750 1.0000
-> CORR  -0.0319 0.0738 -0.1660 1.0000
-> CORR  -0.0105 0.0051 0.0332 0.0372 1.0000
-> CORR  -0.0299 -0.0129 0.0922 0.0227 -0.3022 1.0000
-> CORR  -0.1636 -0.1276 0.0769 -0.1553 -0.0084 0.0415 1.0000
-> CORR  0.0472 0.0276 -0.0681 0.0667 -0.0579 0.1132 -0.0981 1.0000
-> CORR  -0.0199 -0.0136 0.0260 0.0155 0.0013 0.0462 -0.0510 -0.1266 1.0000
-> CORR  -0.0150 0.0315 0.0112 0.0518 0.0123 0.0162 0.0092 -0.0333 0.0117 1.0000
-> CORR  0.0060 -0.0396 -0.0153 -0.0367 -0.0177 -0.0143 0.0015 -0.0289 0.1074
->     -0.0169 1.0000
-> CORR  0.0102 0.0605 0.0039 0.0279 0.0602 0.0335 0.1324 0.0316 -0.0054 -0.0196
->     -0.0562 1.0000
-> CORR  0.0134 0.0242 0.0574 0.0189 0.0726 -0.0052 0.0430 -0.0025 0.0015 -0.0176
->     -0.0504 -0.0584 1.0000
-> CORR  -0.1581 -0.0874 0.0321 0.0424 0.0066 -0.0421 0.0486 0.0481 -0.0030
->     0.0031 -0.0074 0.0288 -0.0213 1.0000
-> CORR  0.0687 0.0215 -0.0321 -0.0488 -0.0210 -0.0292 -0.0124 0.0542 -0.0171
->     0.0010 -0.0101 -0.0056 0.0204 -0.2972 1.0000
-> CORR  -0.0052 0.2082 -0.0080 0.0509 -0.0024 0.0448 -0.0394 -0.0642 0.0088
->     0.0073 -0.0588 -0.0106 0.0034 -0.4200 -0.4285 1.0000
-> CORR  -0.0764 0.0432 0.0618 0.2182 0.0422 0.0888 0.0034 -0.0232 0.0033 0.0301
->     -0.0110 0.0650 0.0826 0.0689 -0.0557 0.0274 1.0000
-> CORR  -0.0302 0.0076 0.0087 -0.0280 0.0469 0.0487 0.0143 -0.0383 0.0758
->     0.0048 0.0041 0.0140 -0.0062 0.0222 -0.0231 0.0101 -0.3302 1.0000
-> N_MATRIX  9183.0
-> N_MATRIX  9173.0 9206.0
-> N_MATRIX  9132.0 9156.0 9165.0
-> N_MATRIX  9132.0 9156.0 9165.0 9165.0
-> N_MATRIX  9135.0 9155.0 9122.0 9122.0 9165.0
-> N_MATRIX  9135.0 9155.0 9122.0 9122.0 9165.0 9165.0
-> N_MATRIX  8925.0 8947.0 8906.0 8906.0 8908.0 8908.0 8955.0
-> N_MATRIX  9183.0 9206.0 9165.0 9165.0 9165.0 9165.0 8955.0 9216.0
-> N_MATRIX  9183.0 9206.0 9165.0 9165.0 9165.0 9165.0 8955.0 9216.0 9216.0
-> N_MATRIX  9183.0 9206.0 9165.0 9165.0 9165.0 9165.0 8955.0 9216.0 9216.0
->     9216.0
-> N_MATRIX  9183.0 9206.0 9165.0 9165.0 9165.0 9165.0 8955.0 9216.0 9216.0
->     9216.0 9216.0
-> N_MATRIX  9183.0 9206.0 9165.0 9165.0 9165.0 9165.0 8955.0 9216.0 9216.0
->     9216.0 9216.0 9216.0
-> N_MATRIX  9183.0 9206.0 9165.0 9165.0 9165.0 9165.0 8955.0 9216.0 9216.0
->     9216.0 9216.0 9216.0 9216.0
-> N_MATRIX  9183.0 9206.0 9165.0 9165.0 9165.0 9165.0 8955.0 9216.0 9216.0
->     9216.0 9216.0 9216.0 9216.0 9216.0
-> N_MATRIX  9183.0 9206.0 9165.0 9165.0 9165.0 9165.0 8955.0 9216.0 9216.0
->     9216.0 9216.0 9216.0 9216.0 9216.0 9216.0
-> N_MATRIX  9183.0 9206.0 9165.0 9165.0 9165.0 9165.0 8955.0 9216.0 9216.0
```

```
->      9216.0 9216.0 9216.0 9216.0 9216.0 9216.0 9216.0
-> N_MATRIX  8746.0 8775.0 8730.0 8730.0 8734.0 8734.0 8533.0 8775.0 8775.0
->      8775.0 8775.0 8775.0 8775.0 8775.0 8775.0 8775.0 8775.0
-> N_MATRIX  8746.0 8775.0 8730.0 8730.0 8734.0 8734.0 8533.0 8775.0 8775.0
->      8775.0 8775.0 8775.0 8775.0 8775.0 8775.0 8775.0 8775.0 8775.0
-> END DATA.

Preceding task required .17 seconds elapsed.


->
-> VAR LABELS
->    B2HENP1  'Highest degree program enrolled after BA=Grad degree program' /
->    GENDER1  'Gender of student=Female' /
->    B2AGAT1  'Respondent age when received BA=25-29' /
->    B2AGAT2  'Respondent age when received BA=30 or older' /
->    B2TOTU1  'Total undergraduate debt=Less than $5,000' /
->    B2TOTU2  'Total undergraduate debt=$5,000 or more' /
->    NORMGP1  'Normalized GPA on a 4.0 scale=Less than 3.0' /
->    SECTOR1  'Institutional type=Private, nfp, 4-year' /
->    SECTOR2  'Institutional type=others' /
->    B2ETHN1  'Race=American Indian' /
->    B2ETHN2  'Race=Asian' /
->    B2ETHN3  'Race=Black' /
->    B2ETHN4  'Race=Hispanic' /
->    B2BAMA1  'Bachelors degree field recoded=Business and management' /
->    B2BAMA2  'Bachelors degree field recoded=Humanities / Soc Sci' /
->    B2BAMA3  'Bachelors degree field recoded=Others' /
->    PAREDU1  'Highest education level by either parent=High school or less' /
->    PAREDU2  'Highest education level by either parent=Some college'.

->
-> REGRESSION MATRIX=IN(*)
->    / VARIABLES=
->     B2HENP1 GENDER1 B2AGAT1 B2AGAT2 B2TOTU1 B2TOTU2 NORMGP1 SECTOR1
->     SECTOR2 B2ETHN1 B2ETHN2 B2ETHN3 B2ETHN4 B2BAMA1 B2BAMA2 B2BAMA3
->     PAREDU1 PAREDU2
->    / MISSING=PAIRWISE
->    / DEP=B2HENP1
->    / METHOD=enter.

   12268 bytes of memory required for REGRESSION procedure.
      0 more bytes may be needed for Residuals plots.



                    * * * *   M U L T I P L E   R E G R E S S I O N   * * * *


Pairwise Deletion of Missing Data

Equation Number 1    Dependent Variable..   B2HENP1   Highest degree program enr

Block Number  1.  Method:  Enter


Variable(s) Entered on Step Number  1..    PAREDU2   Highest education level by either parent
                                    2..    B2ETHN2   Race=Asian
                                    3..    NORMGP1   Normalized GPA on a 4.0 scale=Less than
                                    4..    B2ETHN1   Race=American Indian
                                    5..    B2BAMA2   Bachelors degree field recoded=Humanitie
                                    6..    B2TOTU1   Total undergraduate debt=Less than $5,00
                                    7..    B2AGAT1   Respondent age when received BA=25-29
                                    8..    B2ETHN4   Race=Hispanic
```

9.. SECTOR2 Institutional type=others
10.. GENDER1 Gender of student=Female
11.. B2ETHN3 Race=Black
12.. SECTOR1 Institutional type=Private, nfp, 4-year
13.. B2AGAT2 Respondent age when received BA=30 or ol
14.. B2BAMA1 Bachelors degree field recoded=Business
15.. B2TOTU2 Total undergraduate debt=$5,000 or more
16.. PAREDU1 Highest education level by either parent
17.. B2BAMA3 Bachelors degree field recoded=Others

		Analysis of Variance			
Multiple R	.26761		DF	Sum of Squares	Mean Square
R Square	.07162				
Adjusted R Square	.06976	Regression	17	127.78032	7.51649
Standard Error	.44106	Residual	8515	1656.45959	.19453

F = 38.63838 Signif F = .0000

------------------ Variables in the Equation ------------------

Variable	B	SE B	Beta	T	Sig T
GENDER1	-.015422	.010079	-.016791	-1.530	.1260
B2AGAT1	-.069275	.015037	-.049718	-4.607	.0000
B2AGAT2	-.052577	.013705	-.042620	-3.836	.0001
B2TOTU1	-.018672	.014561	-.014267	-1.282	.1998
B2TOTU2	-.022665	.010799	-.023716	-2.099	.0359
NORMGP1	-.153928	.010033	-.167055	-15.342	.0000
SECTOR1	.031095	.010648	.031537	2.920	.0035
SECTOR2	-.047130	.026943	-.018654	-1.749	.0803
B2ETHN1	-.019465	.062670	-.003256	-.311	.7561
B2ETHN2	9.55624E-04	.023075	4.390E-04	.041	.9670
B2ETHN3	.088249	.020401	.046257	4.326	.0000
B2ETHN4	.063860	.022254	.030387	2.870	.0041
B2BAMA1	-.228880	.015725	-.209210	-14.556	.0000
B2BAMA2	-.059244	.015645	-.054736	-3.787	.0002
B2BAMA3	-.105690	.014641	-.112023	-7.219	.0000
PAREDU1	-.062782	.011479	-.063743	-5.469	.0000
PAREDU2	-.047576	.013058	-.040980	-3.643	.0003
(Constant)	.519292	.015046		34.514	.0000

End Block Number 1 All requested variables entered.

Preceding task required .11 seconds elapsed.

Appendix B-1. DAS-C Tags and How to Use Them

For more information about tags, consult the DAS User Guide (http://nces.ed.gov/das/das_windows/user.asp).

Tag	For variable type	General Usage
Weight	Weight	Must be specified when there is more than one weight variable for the survey.
Filter	Any	Selects the values of a variable that you want to include in the model. Categories not selected are excluded.
Each	Categorical	Use when comparing a single category to each of the other categories. All categories must be used.
Lump	Categorical	Use when you want to group one or more categories together or leave out categories not needed for comparison.
Cut	Continuous	Use for categorizing continuous variables.
Continuous	Continuous	If a single continuous value is desired.

Appendix B-2. Survey Dataset Abbreviations for NCES Data Sets

Baccalaureate and Beyond (B&B):

B01 BB:01 Baccalaureate and Beyond – 2000 cohort, followed up in 2001.
B97 BB:97 Baccalaureate and Beyond – 1993 cohort, followed up in 1994 and 1997.

Beginning Postsecondary Students (BPS):

F01 BPS:01 Beginning Postsecondary Students – 1996 cohort, followed up in 1996 and 2001.
F94 BPS:94 Beginning Postsecondary Students –1990 cohort, followed up in 1992 and 1994.

High School and Beyond (HS&B):

HSO High School & Beyond: Sophomores, 1980-1992
HSR High School & Beyond: 1980-86 Seniors

National Education Longitudinal Study of 1988 (NELS):

NE0 National Education Longitudinal Study: 1988/2000

National Household Education Survey of 1995 Adult Education (NHES):

NH5 NHES: National Household Education Survey of 1995 Adult Education

National Longitudinal Survey of 1972 (NLS72):

NLS National Longitudinal Study of the High School Class of 1972

National Postsecondary Student Aid Study (NPSAS):

N2U NPSAS:2000 Undergraduate Students
N2G NPSAS:2000 Graduate and First-Professional Students
N6G NPSAS:96 Graduate/First Professionals
N6U NPSAS:96 Undergraduate Students
N3G NPSAS:93 Graduate Students
N3U NPSAS:93 Undergraduate Students
N0G NPSAS:90 Graduate and First-Professional Students
N0U NPSAS:90 Undergraduate Students
N7G NPSAS:87 Graduate and First-Professional Students
N7U NPSAS:87 Undergraduates

National Survey of Postsecondary Faculty (NSOPF):

PF9 National Study of Postsecondary Faculty: 1999
PF3 National Survey of Postsecondary Faculty: 1993
PF8 National Survey of Postsecondary Faculty: 1987

Appendix B-3. How to Create a Shortcut for the DAS-C Mode

➡ Create the shortcut

1. Open Windows Explorer, and go to
   C:\WINDOWS\Start Menu\Programs\Data Analysis System.
2. Right-click the B97* icon, and select Copy.
3. Right-click the empty space in the window, and select Paste.
4. Right-click the new icon, and select Rename.
5. Rename it to:
   B97C
6. Right-click the project icon, and select Properties.
7. Click the Shortcut tab.
8. In the Target field, change C:\Dasw\DASW.EXE T=B97 to:
   C:\Dasw\DASW.EXE C=B97
9. Click OK.

***** B97 is the survey dataset abbreviation for the Baccalaureate and Beyond Second Follow-up, the dataset used in this tutorial. If you are using different dataset, substitute all "B97" with the appropriate survey dataset abbreviation. See Appendix B-2 for a complete list of abbreviations.

➡ Use the Shortcut

To open B97 in the DAS-C mode:

1. Click 

.
2. Go to Programs.
3. Go to Data Analysis System.
4. Click the B97C icon.