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

Helen Jang
Andrea Livingston
Shirley He
MPR Associates, Inc.

February 2005

TABLE OF CONTENTS

INTRODUCTION

The Data Analysis System (DAS) is a software application that allows you to produce tables and correlation matrices from NCES data sets, mainly postsecondary data. There is a separate DAS for each data set, but all have a consistent interface and command structure. Each DAS runs in two modes—Tables and Correlations. The Integrated Postsecondary Education Data System (IPEDS) DAS runs in the tables mode only.

In the Tables mode, the DAS will create a table of estimates, corresponding standard errors that are calculated by taking into account the complex sampling designs used in NCES surveys, and weighted sample sizes for the estimates.

In the Correlations mode, the DAS will produce a correlation matrix that can be used for linear regression models. The design effects (DEFTs) for each variable are included in the output. Since statistical procedures generally compute regression coefficients based on a simple random sample assumptions, the standard errors must be adjusted with the design effect to take into account the stratified sampling method used in NCES surveys.

It is helpful to have a basic knowledge of SPSS or SAS and a general understanding of multiple regressions before you use the correlations mode. 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 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 3 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 for Table 3 from *Debt Burden Four Years After College*

Programming Instructions

Before you create a Correlation Parameter 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.

Programming Instructions for this Regression Model

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

Weight: WTD00 B&B panel weight

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

As mentioned previously, the DAS can be run in two different modes: Tables which allows you to create TPFs for tables and Correlations which allows you to create CPFs for correlation matrices. For this tutorial we will use the Correlations mode.

To use the DAS Online, you will need a web browser. You can use Microsoft Internet Explorer (5.0 or higher) or Netscape (4.7 or higher).

➡ To Access the DAS Correlation Applications

1. Go to the DAS website (http://nces.ed.gov/das).
2. Click *DAS Online Application*.
3. From the Left Menu, click *To Create Correlations*.
   Information on available DAS applications appears.
4. Under College Graduates (Baccalureate and Beyond), click *Graduated in 1992-93*.
   The NCES Data Usage Agreement appears. Review the agreement and click "I agree to the terms above."
   The DAS Online application appears.

Step 2: Tagging the Variables

To generate a correlation matrix, you must select the variables and specify how you want them to appear. This process is called "tagging," and a tagged variable with specifications is called a "parameter." In this step, you will learn how to tag variables based on the programming instructions (see page 4). (For more information on tags and how to use them, click the Help button in the DAS Online application.) Before you can tag a variable, you must first locate it in the Main window.

Search for Variables

1. From the Toolbar, click the *Search for Variable* button.
   The Search window appears.
2. Type the search term in the *Search for* field.
3. Select one of the four search methods.
4. Click *Search*.
   The search results appear.

Note You can also find variables by expanding the subjects in the Variables List. For this tutorial, we will use the Search function.

Tag the Variables

To tag a variable, you must do the following:

1. In the Search Results window, click the appropriate variable.
   The Variable Information window for that variable appears.
2. Click the Select a Tag tab.
3. Click the appropriate tag.
   The tag-specific Variable Information window appears.

Tip If you expand the subject categories that contain the tagged variables in the Variables List you will see a red notation indicating the tag assigned to each variable. You can also view the TPF at any time by clicking Show on the main toolbar.

Note Instructions for tagging variables start in the tag-specific window, which appears after you have located and clicked the appropriate variable.

Weight Variable

The DAS will not produce unweighted output. If a dataset contains only one weight, no weight variables will be displayed in the Variables List and the DAS will automatically apply it to your table.

WTA00 (B&B base weight): Weight tag

➡ **To create a Weight (W) tag**

1. In the Weight window, click Weight.
2. Click Save.

Filter Variable

B2ETHNIC (Respondent Ethnicity): And_Filter tag

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

➡ **To create a Filter (F) tag**

1. In the Select a Tag window, click And_Filter.
2. Select Integer Filter.
3. Click the boxes for categories 1–5.
4. Click Save.

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 (Highest degree program enrolled after BA): 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 become the comparison group, "Did Not Enroll." The percentages for these two groups will add up to 100.

You should see the following default entries.

- In the Variable Label field: **Highest degree program enrolled after BA**
- In the Missing Values pull-down menu: **–1**

➡ **To create a Lump (L) tag**

1. Check the boxes for Master's degree to Doctoral degree (categories 4–8).
2. In the Lump Title field, type:
   Graduate degree program
3. Click Save.

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 will be tagged so that it can be compared with the comparison group. In other words, 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.

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 (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.

➡ **To create an Each (E) tag**

1. The Base Category pull-down menu should show the default value of 1.
2. Click Save.

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

B2AGATBA (Respondent age when received BA): Cut tag

This is a continuous variable and therefore you 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.

➡ **To create a Cut (X) tag**

1. In the Missing Value field, type:
   –1
2. Enter the information shown below in the corresponding fields.

Lower cut	Upper Cut	Label
24.5	29.5	25-29
29.5	99.5	30 or older

3. Click Save.

More Cut Tags

You will also need to create Cut tags for the following variables using the values below.

B2TOTUDB (Total Undergraduate debt): Cut tag

Lower cut	Upper Cut	Label
0.5	4999.5	Less than $5,000
4999.6	999999.5	$5,000 or more

NORMGPA (Normalized GPA): Cut tag

Lower cut	Upper Cut	Label
0.5	299.5	Less than 3.0

TIP You can check the values for a variable by scrolling to the Statistics section in the Variable Information window.

SECTOR_B (Institution type): Lump tag

Each lumped group must be tagged as a lump separately. For this variable, you will create two Lump Tags: Private, nfp, 4-year; and Other. 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 group.

You should see the following default entries.

- In the Variable Label field: **Institutional type**
- In the Missing Values pull-down menu: **–1**

➡ **To create the Lump (L) tags**

1. Check the boxes for Private, nfp, 4-year, non-PhD granting; and Private, nfp, 4-year, PhD granting.
2. In the Label field, type:
   Private, nfp, 4-year
3. Click Save.

Repeat the steps above to create the second Lump tag for the following categories. The label for this lump is **Other**.

1 Public, less-than-2-year
2 Public, 2-year
5 Private, nfp, less-than-4-year
8 Private, for-profit, less-than-2-year
9 Private, for-profit, 2-years-or-more

More Lump Tags

You will also need to create Lump tags for the following variables.

B2ETHNIC (Race): Lump tag

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

- Lump tag 1: American Indian
- Lump tag 2: Asian
- Lump tag 3: Black
- Lump tag 4: Hispanic

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

B2BAMAJR (Bachelors degree field): Lump tag

The comparison group is Engineering, mathematics, or science. The following categories should be left untagged: engineering, biological sciences, and other sciences. Categories for each lump are shown in the brackets.

- Lump tag 1: Business Management [business management]
- Lump tag 2: Humanities/social sciences [social sciences, history, humanities, and psychology]

- Lump tag 3: Other [education, health professions, public affairs/social services, and other]

PAREDUC (Highest education level by either parent): Lump tag

The comparison group is Bachelor's or advanced degree. The following categories should be left untagged: bachelor's degree, master's degree or equivalent, first-professional degree, other advanced professional degree, and doctorate. Categories for each lump are shown in the brackets.

- Lump tag 1: High school or less [Less than high school, GED, and high school graduation]
- Lump tag 2: Some postsecondary [Less than 1 year, 1 year but less than 2 years, 2 years or more, less than 2 years of college, associate's degree, and 2 or more years of college]

Note See Appendix B 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 Toolbar, click Save.
   The Save CPF window appears.
2. Select SPSS output.
3. Type a title for your table.
4. Click Save.
   The standard Windows File Download dialog box appears.
5. Click Save
   The standard Windows Save As dialog box appears.
6. Indicate where you want to save the file.
7. Click Save

➡ To run the CPF file

1. From the Toolbar, click Run.
   The Run CPF window appears. Your selections from the previous step is displayed (SPSS output and title)
2. Click the Run button.
   The Confirm Input window appears.
3. Review the parameters.

4. Click the Run button.
   A message appears that says: Processing your correlation. When your correlation is processed, a note appears informing you that the output file is ready for download.

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 proccessed file.
2. Rename myFile.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 C 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 myFile.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 D 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 WTD00 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:

$$\hat{Y} = a + b_1 G + b_2 N \qquad (1)$$

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 C) while (a, b_1 and b_2) are the regression output produced by the SPSS program (see Appendix D).

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:

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

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) under 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.

[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).

[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. 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[5]	*31.9*	*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 B. Correlation Parameter File: Example.cpf

```
VCBB:97 Baccalaureate and Beyond Second Followup 06/01/00

W WTA00 B&B base weight
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 24.5 29.5 25-29
 -1 29.5 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 C. 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
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 D. 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

Multiple R .26761 Analysis of Variance
R Square .07162 DF Sum of Squares Mean Square
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.