DAS
ONLINE
2.0

STEP-BY-STEP

TUTORIALS

Estimating Covariance Analyses:
Logistic Regression

ESTIMATING COVARIANCE ANALYSES: LOGISTIC REGRESSION

In this example, you will run a simple logistic (logit) regression. You will use the 2003 Baccalaureate and Beyond Longitudinal Study (B&B:93/03) dataset. The dependent variable will be whether the student earned a graduate degree, and the independent variables will be gender, race/ethnicity, age, the type of program for highest graduate enrollment, the major field of study for highest graduate enrollment, and the time between earning a bachelor's degree and enrolling in graduate school.

To run a regression using the B&B:93/03 data, go to the DAS Website (http://nces.ed.gov/das). Click *DAS Online Application*, then *To Estimate Covariance Analyses* on the left menu. In the Accessing the DAS to Estimate Covariance Analyses page, scroll down to Postsecondary Students: College graduates, and click *Graduated in 1992-93*.
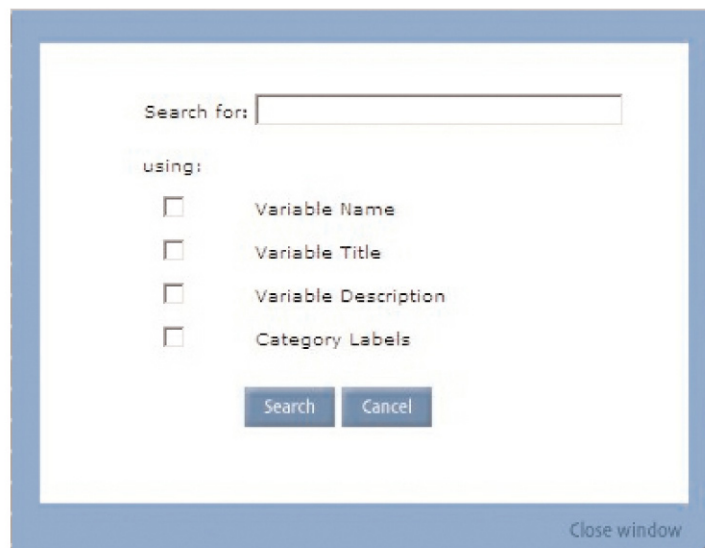
Selecting Variables

You have two options when selecting variables. You can browse for variables (see the Viewing Variables section on page 2), or you can search for variables by variable name, title, description, or by category label (see the Searching for Variables section below).

▶ **Selecting the Dependent Variable: Graduate enrollment and attainment status by 2003**

Searching for Variables

1. From the Toolbar, click *Search for Variable*. The Search Window appears.

2. In the "Search for" field, type the desired search term. You are look-
    ing for graduate enrollment and attainment status by 2003, so enter
    "Graduate enrollment and attainment status by 2003" in the "Search
    for" field.

3. Select the fields to be searched. "Graduate enrollment and attainment
    status by 2003" may be included in the variable title or the variable
    description.

4. Click *Search*.

**Viewing Variable
Descriptions/Percentages**
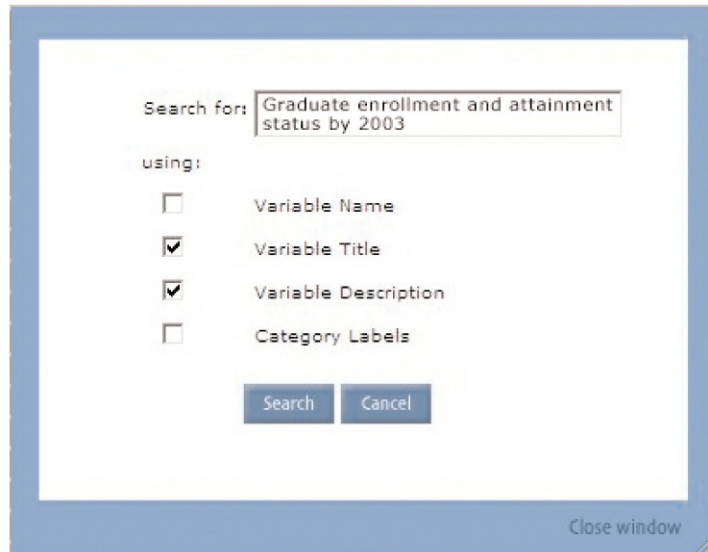
5. From the Variables list, click on *Graduate enrollment and attainment
    status by 2003 (B3ENRAT)*.
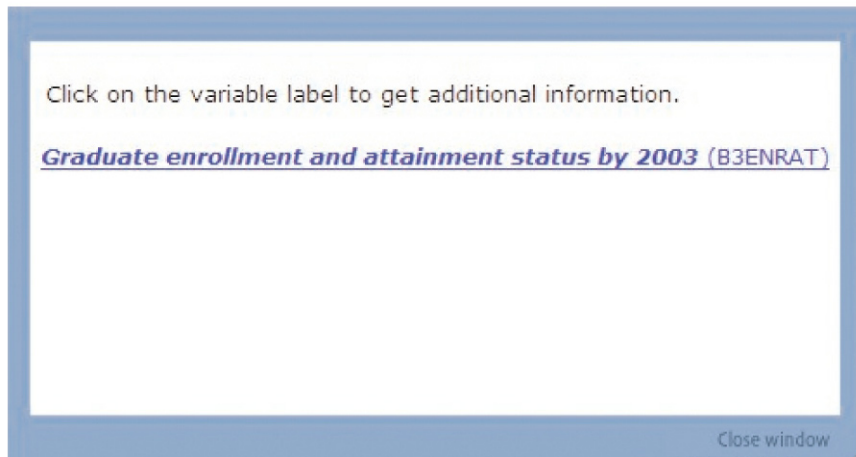
6. The Variable Information Window appears. Scroll to the "Statistics" section to view the percentages. Note that B3ENRAT is a categorical variable with eight categories and one category for missing cases.

Variable information	Select a tag		Back to search results

Name: B3ENRAT

Label: Graduate enrollment and attainment status by 2003

Type: Categorical

Description:
This variable shows the respondent's graduate degree attainment status and graduate enrollment status with the following priority:
1. Attained doctoral degree
2. Attained first-professional degree
3. Currently enrolled doctorate
4. Currently enrolled first professional
5. Attained terminal masters degree
6. Currently enrolled masters
7. No attainment, previously enrolled
8. No graduate enrollment
Respondents who have attained a master's degree are identified as having a terminal master's degree if no higher degree was attained and the respondent is currently not enrolled in a doctoral or first-professional program.
Applies to: All respondents.

Statistics

Value	Percentage	Label
0	59.2	No graduate enrollment
1	19.4	Attained terminal master's degree
2	3.9	Attained first-professional degree
3	2	Attained doctoral degree
4	4.2	Currently enrolled master's
5	0.5	Currently enrolled first professional
6	1.2	Currently enrolled doctorate
7	9.5	No attainment, previously enrolled
-9	0.2	{Missing, blank}

**Source:** Derived from B&B:93/2003, B&B:93/1997, and B&B:93/1994.

Close window

Tagging Variables

7. You will use B3ENRAT as a dependent variable. Click *Select a tag*. The Tag-Specific Window appears. Click *Dependent*.

7. The Regression Window appears. You can select a weighted least squares linear regression (WLS) or logistic (logit) regression. For this example, we will create a logistic regression using a categorical dependent variable.[1] Select *Logit* and click *Lump*. A new window appears.

---
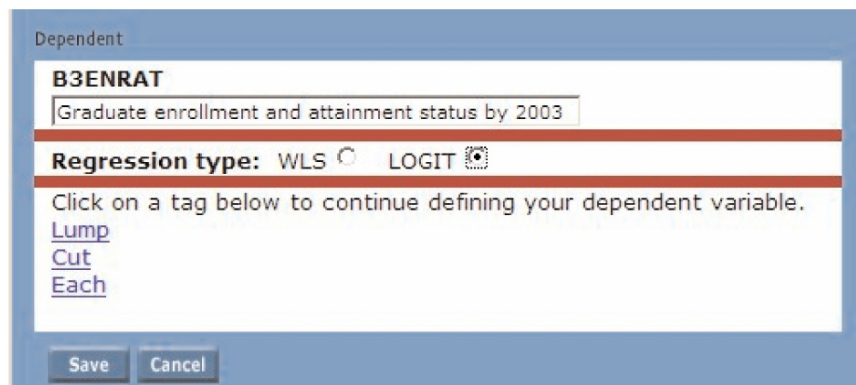
[1]If you wanted to run a Weighted Least Squares (WLS) regression, you would select "WLS" and the Lump, Cut, or Each tag, specify the categories, then click *Save*.

8. Edit the variable label so that it reads "Graduate attainment status by 2003." The reference category will include students who attained a graduate degree. Check the "Use as reference" box next to Attained terminal master's degree, Attained first-professional degree, and Attained doctoral degree.

   The next category will include students who did not earn a graduate degree by 2003 or were currently enrolled in 2003. Check the "use in lump" box next to No graduate enrollment, Currently enrolled master's, Currently enrolled first professional, Currently enrolled doctorate, and No attainment, previously enrolled. Enter "No graduate enrollment or currently enrolled" in the Lump title box. Click *Add Lump*. Enter "Attained graduate degree" in the reference category label box.

Lump

B3ENRAT

Graduate attainment status by 2003

Categories
Select one or more lumps by clicking one or more categories and then Add lump Button.

	Use in lump	Use as reference
0 No graduate enrollment	☑	☐
1 Attained terminal master's degree	☐	☑
2 Attained first-professional degree	☐	☑
3 Attained doctoral degree	☐	☑
4 Currently enrolled master's	☑	☐
5 Currently enrolled first professional	☑	☐
6 Currently enrolled doctorate	☑	☐
7 No attainment, previously enrolled	☑	☐
-9 {Missing, blank}	☐	☐

Lump title: [   ] **Add Lump**

Selected categories:
✕ 0+4+5+6+7 No graduate attainment or currently enrolled
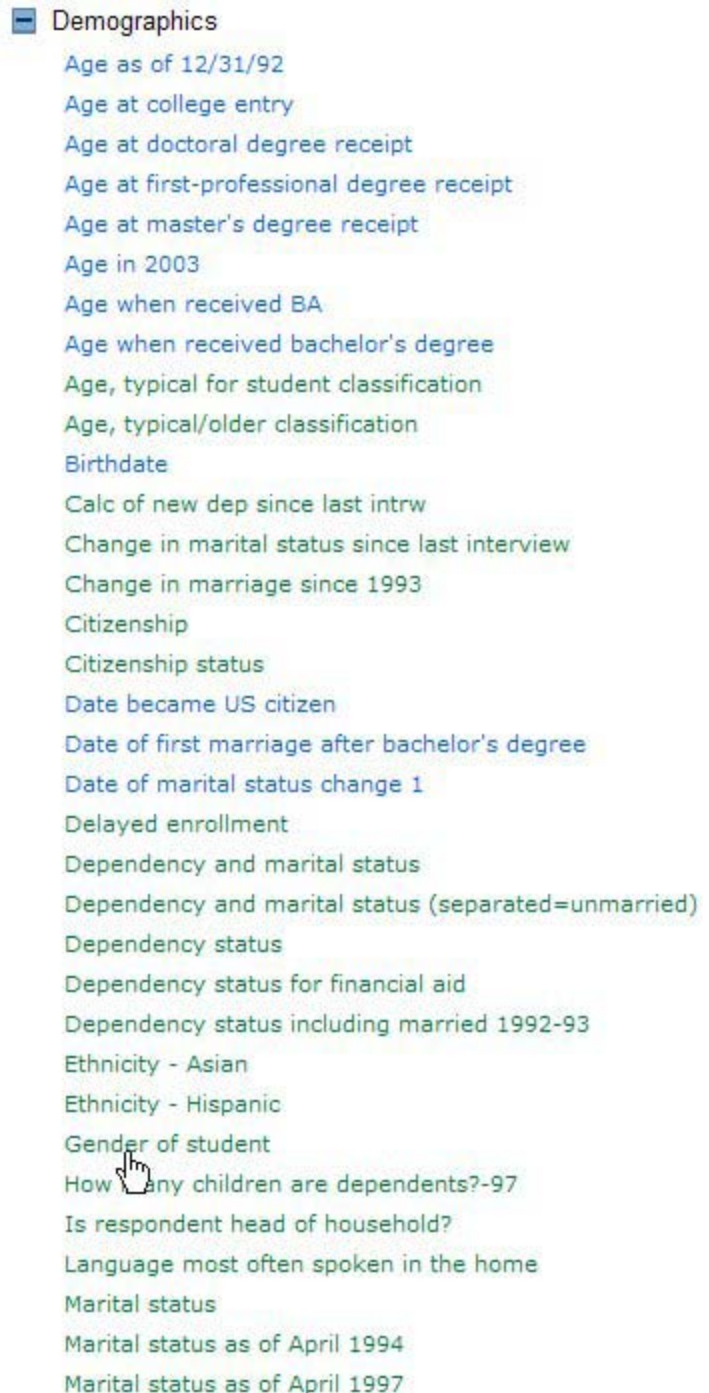Reference category label: Attained graduate degree

Save **Cancel**

9. Click *Save*.

▶ Selecting the Independent Variable: Gender

Viewing Variables

1. In the Variables list, scroll down to the Background section and click ⊞ next to Demographics to expand the subject listing. Variables displayed in green are categorical variables. Variables displayed in blue are continuous variables.

⊟ Demographics

    Age as of 12/31/92
    Age at college entry
    Age at doctoral degree receipt
    Age at first-professional degree receipt
    Age at master's degree receipt
    Age in 2003
    Age when received BA
    Age when received bachelor's degree
    Age, typical for student classification
    Age, typical/older classification
    Birthdate
    Calc of new dep since last intrw
    Change in marital status since last interview
    Change in marriage since 1993
    Citizenship
    Citizenship status
    Date became US citizen
    Date of first marriage after bachelor's degree
    Date of marital status change 1
    Delayed enrollment
    Dependency and marital status
    Dependency and marital status (separated=unmarried)
    Dependency status
    Dependency status for financial aid
    Dependency status including married 1992-93
    Ethnicity - Asian
    Ethnicity - Hispanic
    Gender of student
    How many children are dependents?-97
    Is respondent head of household?
    Language most often spoken in the home
    Marital status
    Marital status as of April 1994
    Marital status as of April 1997

Tagging Variables

1. From the Variables list, select "Gender of student." Note that there are four categories, including one category for missing cases, and one category for skipped cases. Click *Select a tag*. You will want to keep all categories using the each tag. Click *Each*. A new window appears.

Variable information   Select a tag

Name: GENDER
Label: Gender of student
Type: Categorical
Description:
Gender of student
What is the student's gender?
Case [CADE->M_STGEN = 1 or 2]
CADE->M_STGEN = DERIVED->GENDER
Case [CADE->M_STGEN<>1 or 2] and [CATI->G001=1 or 2 or -7 or -8]
CATI->G001 = DERIVED->GENDER
Otherwise GENDER = -9
Key to variable(s) used in the construction of derived variable:
G001 - Student: Sex of the respondent
M_STGEN - Student: Gender

Statistics

Value	Percentage	Label
1	45.2	Male
2	54.4	Female
-1	0.2	{Missing}
-3	0.2	{Skipped}

Source: N93 CADE, N93 Student CATI

Close window

2. Edit the variable label so that it reads "Gender." The reference category will be Male students (category 1). Check the "Use as reference" box next to Male.

   The second category is Female students. Check the "Use as dummy" box next to Female.

3. Click *Save*.

▶ Selecting the Independent Variable: Race/ethnicity

1. Search for "Race" in the Variable title and description by clicking
    on Search for Variable in the toolbar. Select "Respondent ethnicity
    (B2ETHNIC)." Note that this is a categorical variable with five catego-
    ries and one category for not applicable cases. Click *Select a tag*.

2. Click *Lump*. A new window appears.

3. Edit the variable label so that it reads "Race/ethnicity." The reference
    category will be White, non-Hispanic students (category 5). Check the
    "Use as reference" box next to White, non-Hispanic.

    The second category will be Asian or Pacific Islander. Check the "Use
    in lump" box next to Asian or Pacific Islander. Click *Add Lump*.

Check the "Use in lump" box next to Black, non-Hispanic. Click *Add Lump*.

Check the "Use in lump" box next to Hispanic. Click *Add Lump*.

4.   Click *Save*.

▶ Selecting the Independent Variable: Age at bachelor's degree

1. Search for "Age" in the Variable title and description by clicking on *Search for Variable* in the toolbar. Select "Respondent age when received BA (B2AGATBA)." Note that this is a continuous variable. Click *Select a tag*.

2. Click *Continuous*. A new window appears.

3. Change the variable title to, "Age at bachelor's degree completion." Because there are no zero values in this variable, it is not necessary to check the "Consider zero as missing value" box.

4. Click *Save*.

▶ Selecting the Independent Variable: Highest graduate enrollment: degree program type

1. Search for "Highest graduate enrollment" in the Variable title and description by clicking on S*earch for Variable* in the toolbar. Select "Highest graduate enrollment: degree program type (B3HENPRG)." Note that this is a categorical variable with five categories, including one category for missing cases. Click *Select a tag*.

2. Click *Lump*. A new window appears.

3. Edit the variable label so that it reads "Highest graduate enrollment." The reference category will be Doctoral degree (category 5). Check the "Use as reference" box next to Doctoral.

The second category will be Master's degrees. Check the "Use in lump" box next to Master's degree and Post-master's certificate. Enter "Master's degree" in the Lump title box. Click *Add Lump*.

Check the "Use in lump" box next to First professional. Click *Add Lump*.

4. Click *Save*.

▶ Selecting the Independent Variable: Highest graduate enrollment: major field of study

1. In the Variables list, scroll down to the Education section and click ⊞ next to Program to expand the subject listing. Variables are displayed alphabetically by variable title. Click *Highest graduate enrollment: major field of study*. Note that this is a categorical variable with nine categories, one category for skipped cases, and one category for missing cases. Click *Select a tag*.

2. Click *Lump*. A new window appears.

3. Edit the variable label so that it reads "Major for highest graduate enrollment." The reference category will be Arts and Humanities (category 1). Check the "Use as reference" box next to Arts and Humanities.

   The second category will be Social and behavioral sciences. Check the "Use in lump" box next to Social and behavioral sciences. Click *Add Lump*.

   Check the "Use in lump" box next to Life and physical sciences. Click *Add Lump*.

   Check the "Use in lump" box next to Engineering/mathematics/computer science. Click *Add Lump*.

   Check the "Use in lump" box next to Education. Click *Add Lump*.

   Check the "Use in lump" box next to Business and management. Click *Add Lump*.

   Check the "Use in lump" box next to Medicine/health. Click *Add Lump*.

   Check the "Use in lump" box next to Law. Click *Add Lump*.

   Check the "Use in lump" box next to Other. Click *Add Lump*.

Lump

B3HENMAJ

Major for highest graduate enrollment

Categories

Select one or more lumps by clicking one or more categories and then Add
lump Button.

	Use in lump	Use as reference
1 Arts and humanities	☐	☑
2 Social and behavioral sciences	☑	☐
3 Life and physical sciences	☑	☐
4 Engineering/math/computer science	☑	☐
5 Education	☑	☐
6 Business and management	☑	☐
7 Medicine/health	☑	☐
8 Law	☑	☐
9 Other	☑	☐
-3 {Skipped}	☐	☐
-9 {Missing, blank}	☐	☐

Lump title: [   ] **Add Lump**

Selected categories:

✕ 2 Social and behavioral sciences
✕ 3 Life and physical sciences
✕ 4 Engineering/math/computer science
✕ 5 Education
✕ 6 Business and management
✕ 7 Medicine/health
✕ 8 Law
✕ 9 Other

Reference category label: [Arts and humanities]

Save   **Cancel**

4. Click *Save*.

▶ Selecting the Independent Variable: Time between BA and graduate school enrollment

1. Search for "Months between BA and graduate school enrollment" in the Variable title and description by clicking on *Search for Variable* in the toolbar. Select "Months between BA and graduate school enroll-ment (2003) (B3GRDENR)." Note that this is a continuous variable. Click *Select a tag*.

2. Click *Cut*. A new window appears.

3. Edit the variable label so that it reads "Time between bachelor's degree and graduate enrollment."

4. The first category will be 13 to 36 months, inclusive. Check the box marked Inclusive, and enter "13" for the lower cut point. Check the second box marked Inclusive and enter "36" for the upper cut point. The label will change as you make these selections.

   The second category will be 37 to 60 months, inclusive. Follow the same procedure as above: the lower cut point (inclusive) is 37, and the upper cut point (inclusive) is 60

   The third category is more than 61 months, inclusive. The lower cut point (inclusive) is 61. Leave the upper cut point box blank. This cat-egory will include all values above and including 61 months.

Cut

B3GRDENR

Time between bachelor's degree and graduate enrollı

Missing and reserve codes for this variable are listed below. You may include these in your table by entering the appropriate cut points.
(-3) {Skipped}
(-9) {Missing, blank}

Enter your cut points and customize labels below

	Inclusive	Lower cut point	Inclusive	Upper cut point	Label
1	☑	13	☑	36	13 <= X <= 36
2	☑	37	☑	60	37 <= X <= 60
3	☑	61	☐		X >= 61
4	☐		☐		
5	☐		☐		
6	☐		☐		
7	☐		☐		
8	☐		☐		
9	☐		☐		
10	☐		☐		

	Inclusive	From	Inclusive	To
Reference range:	☐		☐	
Reference category label:				

Save   Cancel

5. Verify that the automated labels match your intended specifications. If not, click *Cancel* and follow the steps above again. If so, change the labels to the following:

> 1 to 3 years
> 3 to 5 years
> More than 5 years

6. The reference category will be from 0 to 12 months. Enter these values and check the inclusive box next to 12 as shown below.

Enter "1 year or less" in the reference category label box.

7.   Click *Save*.

▶ Choosing a weight

1. Search for "Weight" in the Variable title by clicking on *Search for Variable* in the toolbar.

    A list of weights appears. Select "Panel weight for resp to all surveys (WTC000)." Click *Select a tag*.

2. Weight variables have only one tag option, Weight. Select Weight.
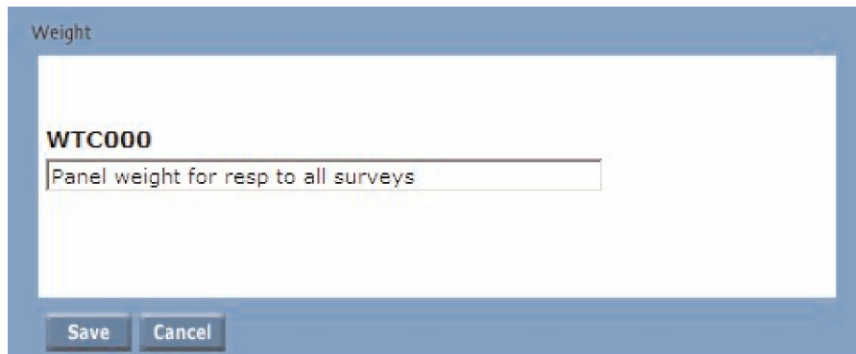
3.   Click *Save*.

▶ Running the Table

1. On the main DAS window, click *Run* in the toolbar.

2. Enter a table title, such as "Logistic_regression." You will run a basic logistic regression in this example, so you will not select any of the advanced regression options.[2] Click *Run*.

---

[2]For more information on the advanced options, see the DAS User Guide at http://nces.ed.gov/dasol/help.

3. Your table specifications will appear in a new window. Make sure that you have a dependent variable, at least one independent variable, and a weight variable. Click *Run*.

Confirm Input

Please review list and press [Run] or [Cancel]

Table Title: Logistic_regession

Dependent LOGIT

Lump B3ENRAT {-9} Graduate attainment status by 2003
  0+4+5+6+7 No graduate attainment or currently enrolled
  0= Attained graduate degree

Each GENDER {-1,-3} Gender
  2 Female
  0= Male

Lump B2ETHNIC {1,-2} Race/ethnicity
  2 Asian or Pacific Islander
  3 Black, non-Hispanic
  4 Hispanic
  0= White, non-Hispanic

Lump B3HENPRG {0,-9} Highest graduate enrollment
  1+2 Master's degree
  3 First professional
  0= Doctoral degree

Lump B3HENMAJ {-3,-9} Major for highest graduate enrollment
  2 Social and behavioral sciences
  3 Life and physical sciences
  4 Engineering/math/computer science
  5 Education
  6 Business and management
  7 Medicine/health
  8 Law
  9 Other
  0= Arts and humanities

Cut B3GRDENR Time between bachelor's degree and graduate enrollment
  13<=X<=36 1 to 3 years
  37<=X<=60 3 to 5 years
  X>=61 More than 5 years
  0=0<=12 1 year or less

Weight WTC000 Panel weight for resp to all surveys

4.   Once the file is processed, a screen will appear allowing you to download the output file or return to the specifications.

5.   To download the regression output, click *Download Output File*. Save this file to your computer. The output file will be a comma-separated file that you can open in Microsoft Excel.

	A	B	C	D	E	F	G
1	Logistic_regession						
2							
3	Time the job executed :	1/19/2007 15:59					
4	Full sample weight :	WTC000					
5	Variance estimation method :	BRR					
6	Type of analysis :	LOGISTIC					
7	Number of replicates :	44					
8	Percentage of observed over total cases :	30.35%					
9	Convergence criterion :	1.00E-06					
10	Maximum of iterations :	25					
11	Value of alpha (Confidence level %) :	0.05000 (95.00000 %)					
12							
13	Options :" Intercept"						
14	Degrees of freedom :	44					
15							
16	Dependent variable :						
17		No graduate attainm	1				
18		vs. Attained gradua	0				
19							
20	Negative log-likelihood(Pseudo R**2)	0.043					
21	-2 log-likelihood						
22	G**2 0 For intercept (-2LL 0)	-298532.51					
23	G**M 0 For model (-2LL M)	25840.246					
24	Likelihood ratio (Cox-Snell)	0.056					
25	Likelihood ratio (Cox-Snell) Maximum	0.734					
26	Likelihood ratio (Estrella)	0.057					
27	Degree of freedom for Wald	44					
28	Problems detected						
29	The standard error for the parameter estimnate is unstable						
30	Found variables with replicated weight collinearity within dummy sets						
31	GENDER : Female						
32	B2ETHNIC : Asian or Pacific Islander						
33	B2ETHNIC : Black, non-Hispanic						
34	B2ETHNIC : Hispanic						
35	B3HENPRG : Master's degree						
36	B3HENPRG : First professional						
37	B3HENMAJ : Social and behavioral sciences						
38	B3HENMAJ : Life and physical sciences						
39	B3HENMAJ : Engineering/math/computer science						
40	B3HENMAJ : Education						
41	B3HENMAJ : Business and management						
42	B3HENMAJ : Medicine/health						
43	B3HENMAJ : Law						
44	B3HENMAJ : Other						
45	B3GRDENR : 1 to 3 years						
46	B3GRDENR : 3 to 5 years						
47	B3GRDENR : More than 5 years						
48							
49	ODDS RATIO RESULTS						
50		Odds Ratio	Lower 95%	Upper 95%	t		
51	Gender						
52	"Female"	1.088	0.843	1.403	0.663		
53	vs. "Male"						
54	Race/ethnicity						
55	"Asian or Pacific Islander"	1.019	0.627	1.658	0.08		
56	"Black non-Hispanic"	1.29	0.912	1.825	1.474		
57	"Hispanic"	1.234	0.789	1.93	0.946		
58	vs. "White non-Hispanic"						
59	Highest graduate enrollment						
60	"Master's degree"	0.751	0.503	1.12	-1.44		
61	"First professional"	0.848	0.468	1.536	-0.557		
62	vs. "Doctoral degree"						
63	Major for highest graduate enrollment						
64	"Social and behavioral sciences"	0.898	0.556	1.451	-0.45		
65	"Life and physical sciences"	0.625	0.364	1.075	-1.741		
66	"Engineering/math/computer science"	0.974	0.599	1.586	-0.107		
67	"Education"	0.778	0.497	1.217	-1.127		
68	"Business and management"	0.675	0.44	1.036	-1.842		
69	"Medicine/health"	0.507	0.308	0.832	-2.751		
70	"Law"	0.248	0.12	0.512	-3.861		
71	"Other"	0.709	0.456	1.101	-1.569		
72	vs. "Arts and humanities"						
73	Time between bachelor's degree and graduate enrollment						
74	"1 to 3 years"	1.284	1.007	1.636	2.07		
75	"3 to 5 years"	1.159	0.882	1.523	1.086		
76	"More than 5 years"	2.833	2.107	3.808	7.068		
77	vs. "1 year or less"						
78							
79	HYPOTHESIS TESTING RESULTS						
80		WaldF	Num. DF	Denom. DF	Probability F		
81	Overall Fit	5.134	17	28	0		
82	Gender	0.439	1	44	0.511		
83	Race/ethnicity	1.082	3	42	0.367		
84	Highest graduate enrollment	1.148	2	43	0.327		
85	Major for highest graduate enrollment	4.161	8	37	0.001		
86	Time between bachelor's degree and grad	16.438	3	42	0		
87	--						
88							
89	WARNING: The -2log-likelihood is not adjusted for clustering and						
90	should not be used to test the overall model fit. See the overall model test						
91	above for an appropriate significance test for overall model fit.						
92							
93	Source: B&B:93/03 Baccalaureate and Beyond Longitudinal Study, Data Analysis System						

6. To download the Covariance Parameter File (CPF) that created the regression output, click *Return to Specifications*. On the main DAS page, click *Save*.[3] This file can later be imported (by clicking *Import* from the main toolbar) so that the regression can be run again.

[3]To save a file, you have three options. Click on the *Save* button in the main toolbar. The first option allows you to save your CPF and prints your job number. To choose this option, click on *Save CPF and print my Job Number*. The second option saves your CPF and sends the Job Number to your email address. To choose this option, type your email address in the textbox and click *Go*. The first two options will store your CPF for 30 days. The last option requires you to create a Power User account. Click *Login* under Power User. If you have a Power User account, log in by typing your e-mail address and password. If you do not have an account, click *Become a Power User*. Type in your e-mail address, and a password will be sent to the address you entered. Retrieve the password from e-mail and sign in as explained above. You will be able to save your CPF to your computer with a Power User account.