DAS
ONLINE
2.0

STEP-BY-STEP
TUTORIALS

Estimating Covariance Analyses:
Weighted Least Squares

ESTIMATING COVARIANCE ANALYSES: WEIGHTED LEAST SQUARES

In this example, you will run a simple Weighted Least Squares (WLS) linear regression. You will use the 2004 National Postsecondary Student Aid Study (NPSAS) dataset. The dependent variable will be the total amount of aid the student received, and the independent variables will be race/ethnicity, dependency status, and student's income percentile.

To run a regression using NPSAS:04 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: All Postsecondary Students, and click *2004 Undergraduates*.
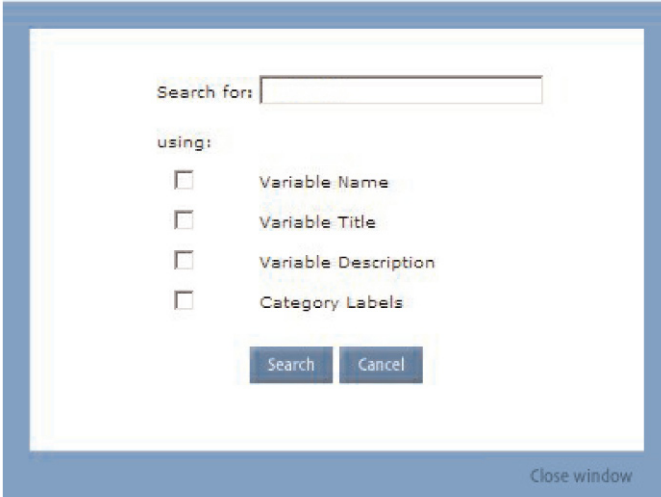
Selecting Variables

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, or description or by category label (see the Searching for Variables section below).

▶ **Selecting the Dependent Variable: Total aid**

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 looking for the total aid, so enter "Total aid" in the "Search for" field.

3. Select the fields to be searched. Total aid may be included in the variable title or the variable description.
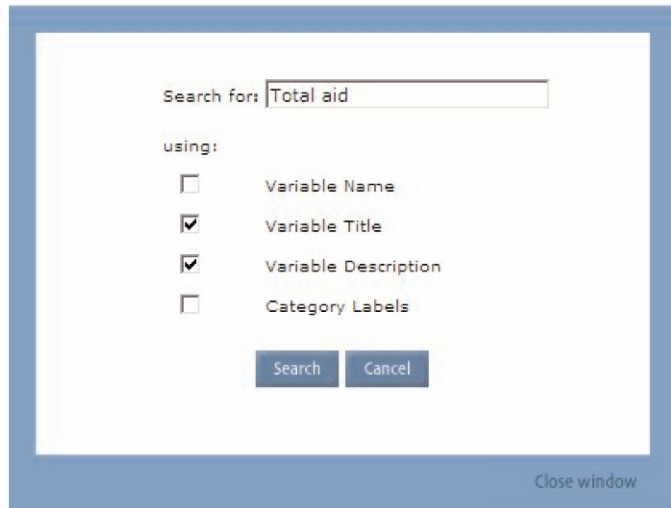
4. Click *Search*.

Viewing Variable Descriptions/Percentages

5. From the Variables list, click on Aid total amount (TOTAID). The Variable Information Window appears.

6. Scroll to the "Statistics" section to view the percentages. Note that TOTAID is a continuous variable with a minimum value of 50 and a maximum value of 56,740.

Variable information   Select a tag

Name: TOTAID
Label: Aid total amount
Type: Continuous
Description:
Total amount of all financial aid received during 2003-2004 academic year. Equal to the sum of aid of all types or the sum of aid from all sources. The types of aid and sources of aid are categorized according to the aid matrix below:
++ ---
AID CATEGORY MATRIX

[Aid Sources]
Federal " State " Institutional " Outside " Totals

[Aid Types]
Grants " TFEDGRT + STGTAMT + INGRTAMT + OTHGTAMT = TOTGRT
Loans " TFEDLN + STLNAMT + INLNAMT + PRIVLOAN = TOTLOAN
Work-study " TFEDWRK + STWKAMT + INSTWRK = TOTWKST
Other " PLUSAMT + VOCHELP + GRASTAMT + VADODAMT = OTHTYPE

Totals " TFEDAID + STATEAMT + INSTAMT + OTHRSCR = TOTAID

++ Variable descriptions --
Federal Grants: TFEDGRT: Total federal grants
Federal Loans: TFEDLN: Total federal loans (excl PLUS)
Federal Work: TFEDWRK: Total federal Work-study
Federal Other: PLUSAMT: Total PLUS loan
Federal Total: TFEDAID: Total federal aid (excl veterans/DOD)

State Grants: STGTAMT: Total state grants
State Loans: STLNAMT: Total state loans
State Work: STWKAMT: Total state work-study
State Other: VOCHELP: Vocational rehab and training (WIA)
State Total: STATEAMT: Total state aid

Institutional Grants:INGRTAMT: Total institutional grants
Institutional Loans: INLNAMT: Total institutional loans
Institutional Work: INSTWRK: Total institutional work-study
Institutional Other: GRASTAMT: Total graduate assistantships amount
Institutional Total: INSTAMT: Total institutional aid

Outside Grants:OTHGTAMT: Total outside grants (private & employer)
Outside Loans: PRIVLOAN: Total private (alternative) loans
Outside Other: VADODAMT: Veteran's benefits and DOD
Outside Total: OTHRSCR: Total outside sources

Total Grants: TOTGRT: Total grants
Total Loans: TOTLOAN: Total loans (excluding PLUS)
Total Work: TOTWKST: Total work-study
Total Other: OTHTYPE: Total other type of aid
Total Total: TOTAID: Total aid amount
++ SAS code ---
TOTAID=SUM(OF TOTGRT TOTLOAN TOTWKST OTHTYPE);
++ Equivalent formula ---
TOTAID=SUM(OF TFEDAID STATEAMT INSTAMT OTHRSCR)

Derived from: TOTGRT TOTLOAN TOTWKST OTHTYPE
Applies to: All respondents.
Perturbation procedures were applied to this and other variables to protect against disclosure of individual information.

Statistics

Value	Percentage	Label
Continuous	63.2	[50-56740;7351.51/6798.75]
0	36.8	{zero}

| Minimum	Maximum	Mean	Standard Deviation
50	56740	7351.51	6798.75

Source: Derived

Close window

Tagging Variables

7. You will use TOTAID as a dependent variable. Click *Select a tag*.

The Tag-Specific Window appears. Click *Dependent*. The Regression Window appears.

8. You can select a weighted least squares linear regression (WLS) or logistic (logit) regression. For this example, we will create a linear regression using a continuous variable.[1] Select WLS and click *Continuous*. A new window appears. Edit the variable label so that it reads "Total aid received." In this example, you want to include zero values, so do not click "Consider zero as missing value."
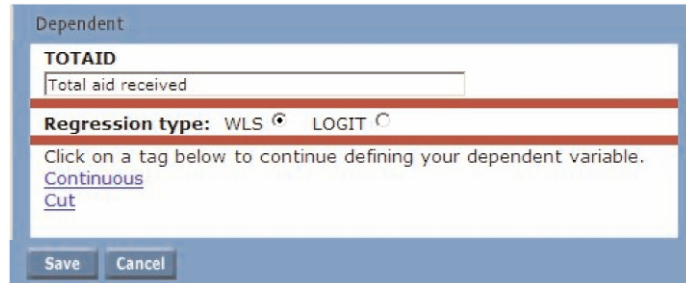
9. Click *Save*.

---

[1]If you wanted to run a logistic regression, you would select "Logit" and the Cut tag would appear. Click on *Cut*, and specify the cut points for both categories, then click *Save*.

▶ Selecting the Independent Variable: Race/ethnicity

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 12/31/03
  Age at start of postsecondary education
  Age groups as of 12/31/03
  Citizenship
  Claimed as a dependent
  Delayed enrollment into PSE- number of years
  Dependency and marital status (separated=married)
  Dependency and marital status (separated=unmarried)
  Dependency status
  Dependency status (3 categories)
  Dependency status (4 categories)
  English is the primary language
  Gender
  Immigrant status
  Index of risk
  Military service type
  Orphan or ward of court
  Race---American Indian or Alaska Native recognized tribe
  Race--American Indian or Alaska Native
  Race--Asian
  Race--Black or African-American
  Race--Native Hawaiian/other Pacific Islander
  Race--Other
  Race--White
  Race-ethnicity (with multiple)
  Race-ethnicity: Hispanic or Latino origin
  Race-ethnicity: Type of Hispanic origin
  Race/ethnicity (with multiple) and gender
  Race: census categories
  Single-parent independent students
  Student was born in the United States
  Student's marital status
  Student's parents were born in the United States
  Veteran status

Tagging Variables   1. From the Variables list, select "Race-ethnicity (with multiple)". Note that there are eight categories. Click *Select a tag*. You will want to collapse some of the smaller categories into one category using a lump tag. Click *Lump*. A new window appears.

Variable information   Select a tag

Name: RACE
Label: Race-ethnicity (with multiple)
Type: Categorical
Description:
Student's race-ethnicity with Hispanic or Latino origin (HISPANIC=1)
as a separate category. Based on the census race categories (RACECEN),
but the race categories exclude Hispanic origin unless specified.
See individual race variables for their number of responses:
[RAWHITE] White
[RABLACK] Black or African American
[RAASIAN] Asian
[RAINDIAN] American Indian or Alaska Native
[RAISLAND] Native Hawaiian or other Pacific Islander
++ Categories ---
<1> White
<2> Black or African American
<3> Hispanic or Latino
<4> Asian
<5> American Indian or Alaska Native
<6> Native Hawaiian or other Pacific Islander
<7> Other
<8> More than one race
++ SAS code ---
RACE=RACECEN;
if RACECEN>2 then RACE=RACECEN+1;
if HISPANIC=1 then RACE=3;

Derived from: RACECEN HISPANIC
Applies to: All respondents.
Perturbation procedures were applied to this and other variables to
protect against disclosure of individual information.

Statistics

Value	Percentage	Label
1	63.1	White
2	14	Black or African American
3	12.7	Hispanic or Latino
4	5.4	Asian
5	0.9	American Indian or Alaska Native
6	0.5	Native Hawaiian / other Pacific Islander
7	1.3	Other
8	2	More than one race

Source: Derived

Close window

2. Edit the variable label so that it reads "Race/ethnicity." The reference
    category will be White students (category 1). Check the "Use as refer-
    ence" box next to White.

    The second category will be Black or African American. Check the
    "Use in lump" box next to Black or African American. Click *Add
    Lump*.

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

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

The last lump will include all remaining categories. Check the boxes marked "Use in lump" for the remaining valid values. Edit the lump title to read "Other." Click *Add Lump*.

3. Click *Save*.

▶ **Selecting the Independent Variable: Dependency status**

1. Search for "Dependency status" in the Variable title and description by clicking on Search for Variable in the toolbar. Select "Dependency status (DEPEND)." Note that this is a categorical variable with two categories. Click *Select a tag*.

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

3. The reference category will be Dependent students (category 1). Check the "Use as reference" box next to Dependent.

   The second category is Independent students. Check the "Use as dummy" box next to Independent. Click *Save*.

▶ Selecting the Independent Variable: Income percentile

1. Search for "Income percentile" in the Variable title and description by clicking on *Search for Variable* in the toolbar. Select "Income percentile rank for all students (PCTALL)." Note that this is a continuous variable. Click *Select a tag*.

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

3. Change the variable title to, "Income percentile rank." 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*.

Continuous

PCTALL

Income percentile rank

☐ Consider zero as missing value

Save   Cancel

▶ Choosing a weight

1. Because there is only one weight for the NPSAS:04 Undergraduate dataset, the DAS automatically assigns a weight. Other datasets have more than one weight and require you to select one. When using other datasets with more than one weight, click *Search for Variable* from the main toolbar. From the toolbar, click *Search for Variable*. In the "search for" field, type "Weight." Check the "Variable title" field. A list of weights appears. Choose the appropriate weight for your analysis.

2. Weight variables have only one tag option, Weight. Select *Weight*, and click *Save*.

▶ **Running the Table**

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

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

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*.

---

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

4. Once the file is processed, a screen will appear allowing you to download the output file or return to the specifications.
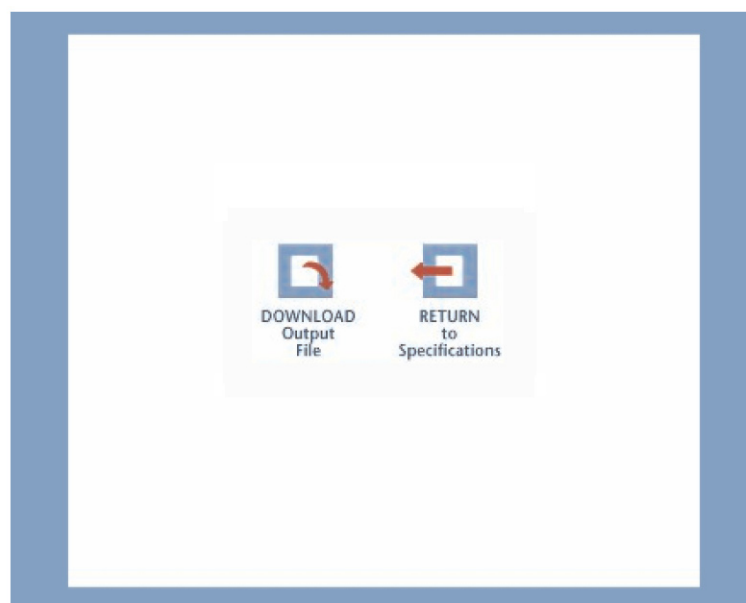
5. To download the regression output, click *Download Output File*. Save this file to your computer. The output file will be a comma-separated file that you can open in Microsoft Excel.

	A	B	C	D	E	F	G
1	WLS_regression						
2							
3	Time the job executed :	1/5/2007 12:52					
4	Full sample weight :	WTA000					
5	Variance estimation method :	BRR					
6	Type of analysis :	LINEAR					
7	Number of replicates :	64					
8	Estimated cases ratio(%):	100					
9							
10	Model includes :	6 variables					
11	No problems detected						
12							
13	The following analysis was performed:						
14	WLS						
15							
16							
17	Degrees of Freedom :	64					
18	t VALUE :	2.008					
19	R**2 :	0.040311803					
20	Adjusted F(wald) :	160.738769598302 (df=6 59)					
21	Dependent variable :	Total aid received					
22							
23	ESTIMATED FULL SAMPLE REGRESSION COEFFICIENTS						
24		b	s.e.	t		Lower 95%	Upper 95%
25	Intercept	7144.862	129.786	55.051	6884.251	7405.472	
26	Race-ethnicity						
27	"Black or African American"	670.773	209.084	3.208	250.932	1090.615	
28	"Hispanic or Latino"	-775.108	129.693	-5.976	-1035.53	-514.684	
29	"Asian"	-829.884	175.76	-4.722	-1182.81	-476.958	
30	"Other"	-77.034	162.463	-0.474	-403.26	249.191	
31	vs. "White"						
32	Dependency status						
33	"Independent"	-1708.795	105.301	-16.228	-1920.24	-1497.35	
34	vs. "Dependent"						
35	Income percentile rank	-31.637	1.349	-23.453	-34.346	-28.928	
36							
37	HYPOTHESIS TESTING RESULTS						
38		WaldF		Num. DF	Denom. D	Probability F	
39	Overall Fit	160.739		6	59	0	
40	Race-ethnicity	26.363		4	61	0	
41	Dependency status	263.341		1	64	0	
42	Income percentile rank	550.025		1	64	0	
43	---						
44							
45							
46	Source: NPSAS:2004 Undergraduate Students, Data Analysis System						
47	1/5/2007						

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.