PowerStats

LEARN BY DOING:

Running a logistic regression and interpreting results

Need help?
nces.info@rti.org

http://nces.ed.gov/datalab

THE REGRESSION YOU'LL CREATE

This tutorial will guide you through the steps taken to create the logistic regression shown below.

Logistic Regression Analysis of Applied for federal aid based on Citizenship and Total income: Parents and independent

EXPAND ALL

ODDS RATIO RESULTS

REGRESSION MODEL INFORMATION

ESTIMATED FULL SAMPLE REGRESSION COEFFICIENTS

	Std.B	S.E.	t	p-value
Intercept				
Citizenship				
Resident alien	-0.009	0.006	-1.608	0.110
Foreign or international student	-0.139	0.004	-38.379	0.000
Total income: Parents and independent	-0.201	0.005	-43.571	0.000

Dependent variable: Applied for federal aid, reference category includes: No.
For Citizenship, reference category includes: US citizen.

The names of the variables used in this regression are: CINCOME, FEDAPP and CITIZEN2. The variable names are unique identifiers. To locate these variables, enter the variable name in the search box.

Source: U.S. Department of Education, National Center for Education Statistics, 2007-08 National Postsecondary Student Aid Study (NPSAS:08).

MEASURES OF FIT

HYPOTHESIS TESTING RESULTS

BASIC STEPS REQUIRED TO CREATE A LOGISTIC REGRESSION

1. Choose a group of students (dataset).
2. Choose a type of regression.
3. Select a dependent variable.
4. Select one or more independent variables.
5. Run your regression.

1. CHOOSE A GROUP OF STUDENTS (DATASET)

INSTRUCTIONS

1. Drag *All undergraduates* to the Group box.

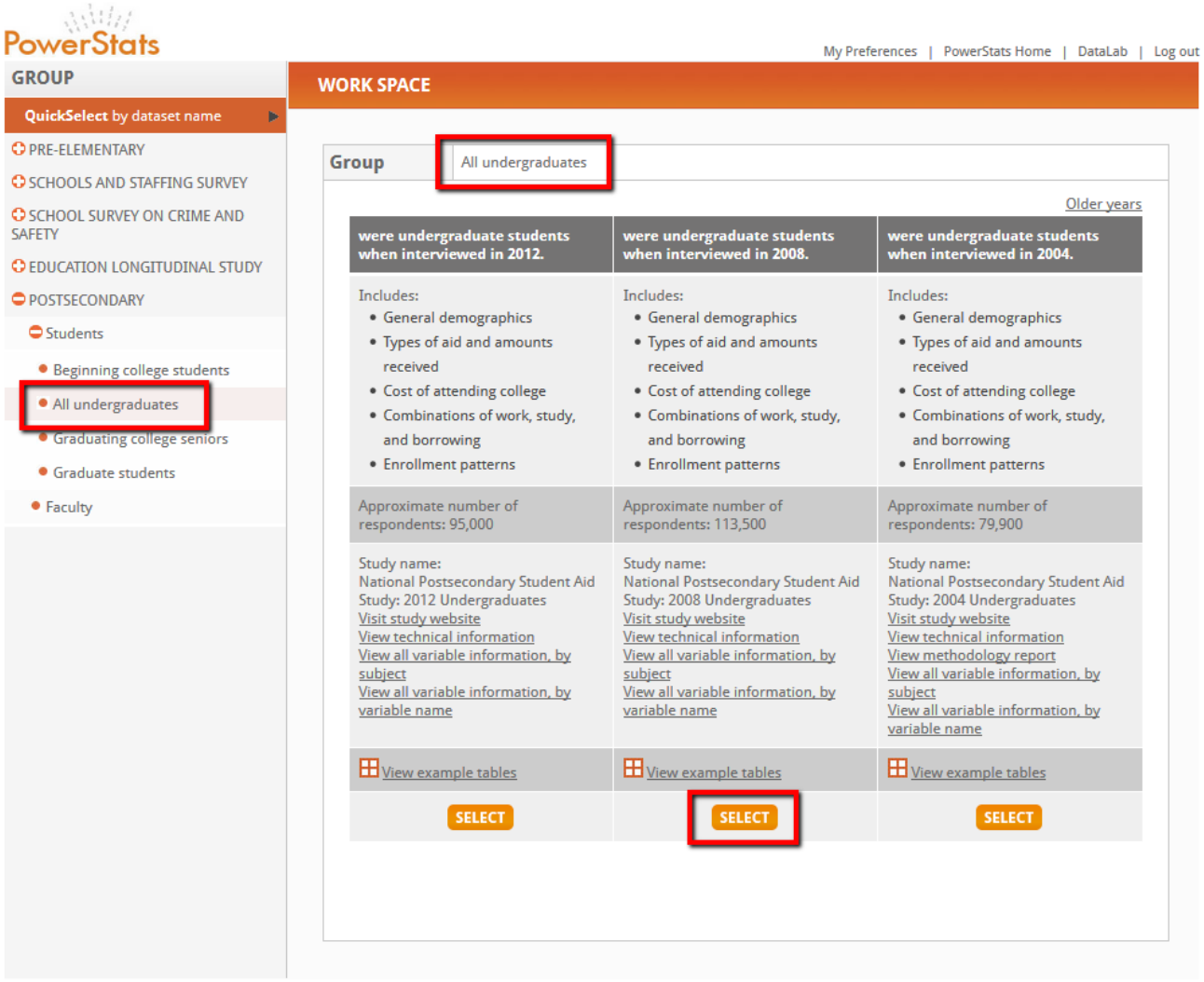
   The Group box is updated with your selection.

   Information about the dataset and the option to select it appear in the Work Space.

2. Click *Select* in the box labeled "were undergraduate students when interviewed in 2008."

TIP
Use the *QuickSelect by dataset name* option if you are familiar with NCES postsecondary studies.

NOTE
If a dataset contains more than one weight variable, you will be prompted to select one.

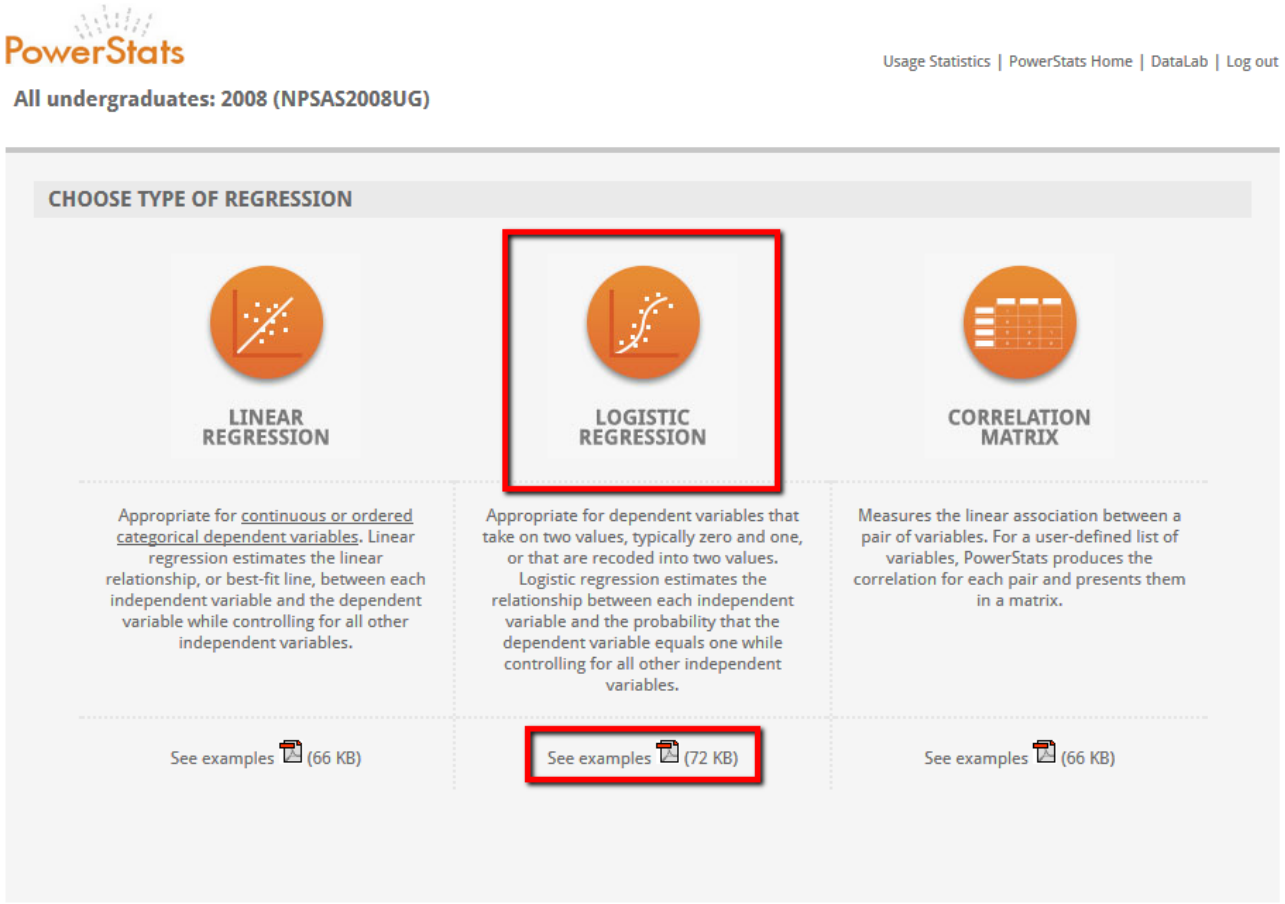


2. CHOOSE A TYPE OF REGRESSION

INSTRUCTIONS

1. Click the *Logistic Regression* icon.

TIP
Click *See examples* under each regression type to learn more.



3. SELECT A DEPENDENT VARIABLE

INSTRUCTIONS

1. Click  next to *Financial aid* to view available variables.

2. Click *Applied for federal aid*.

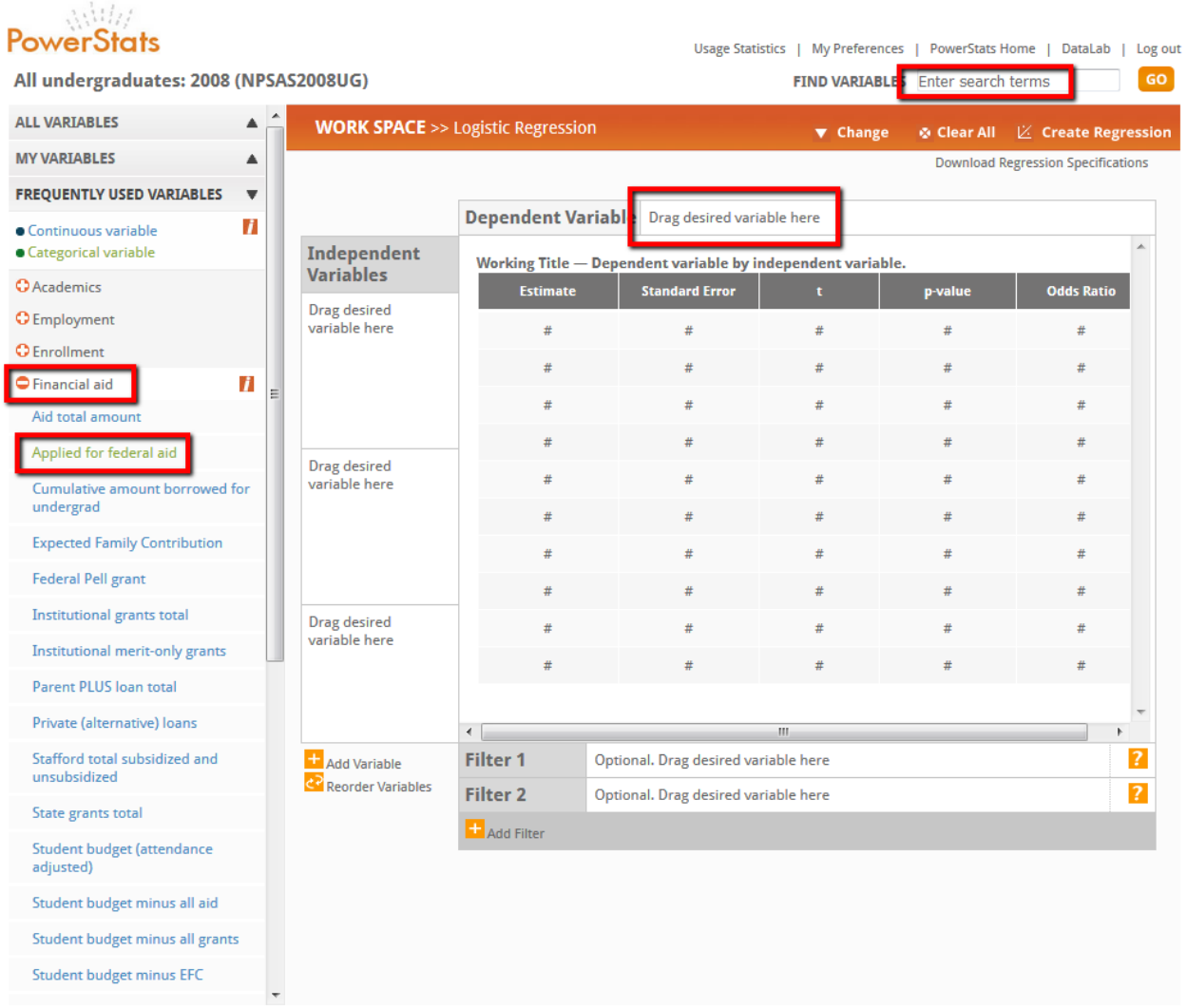
   The Variable Description screen appears.

3. Click *View descriptive statistics* . Note that *Applied for federal aid* is a categorical variable with two values, No (0) and Yes (1) .

4. Drag *Aid total amount* to the Dependent Variable box.

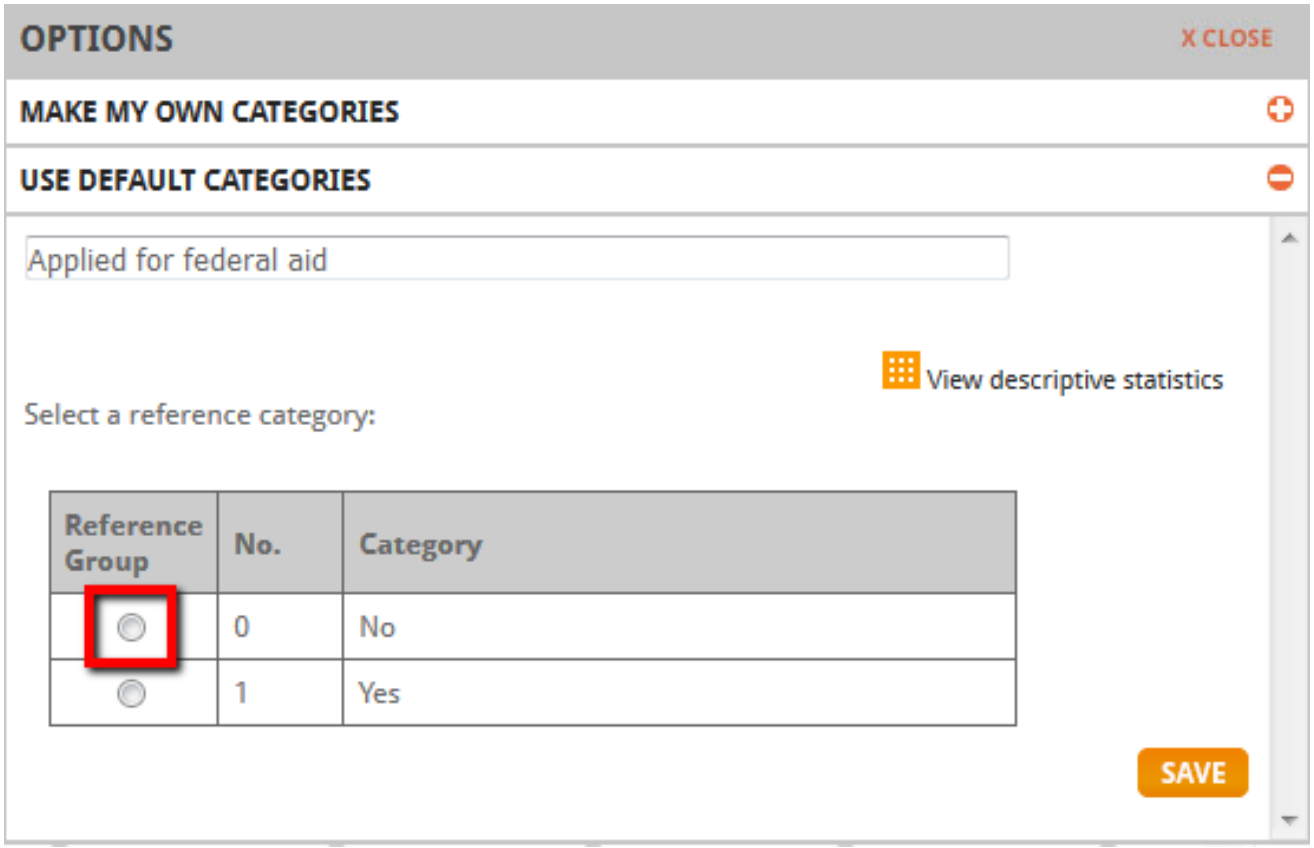
   The Usage Option screen appears.

TIP
Use the *Find Variables* option to search for variables by name and keyword.



3. SELECT A DEPENDENT VARIABLE, cont'd.

5. Click *No* for the reference or base group. Usually it is most logical to code dummy variables like this, with *Yes* representing the presence of the phenomenon (having applied for federal aid) and *No* representing the absence of the phenomenon.



4. SELECT INDEPENDENT VARIABLE 1

INSTRUCTIONS

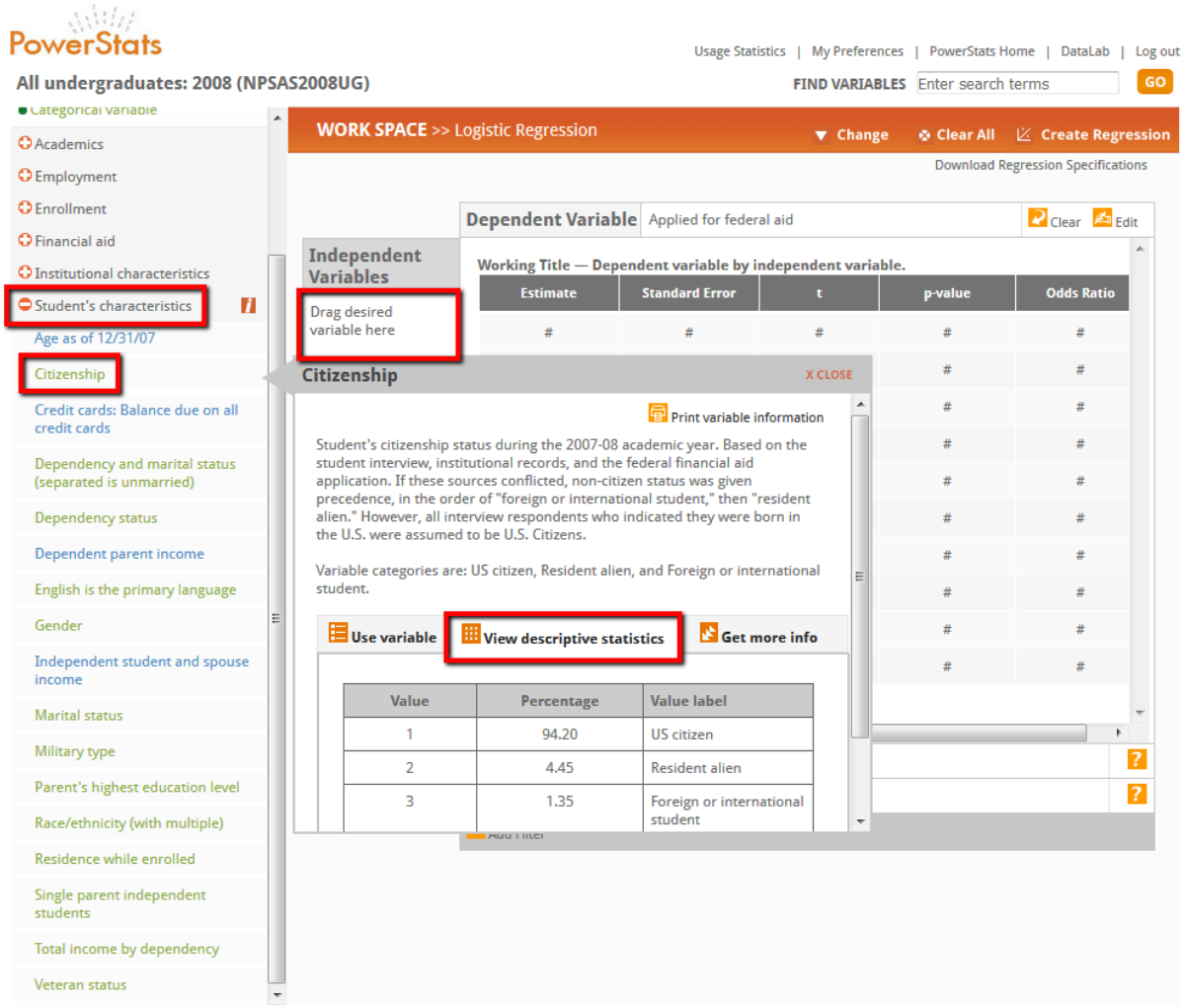
1. Click ⊕ next to *Student's characteristics* to view available variables.

2. Click on *Citizenship.*

   The Variable Description screen appears.

3. Click on the *View descriptive statistics* tab. Note that 94% of the survey respondents are US citizens.

4. Drag the variable *Citizenship* to the first Independent Variables box.



4. SELECT INDEPENDENT VARIABLE 1, cont'd.

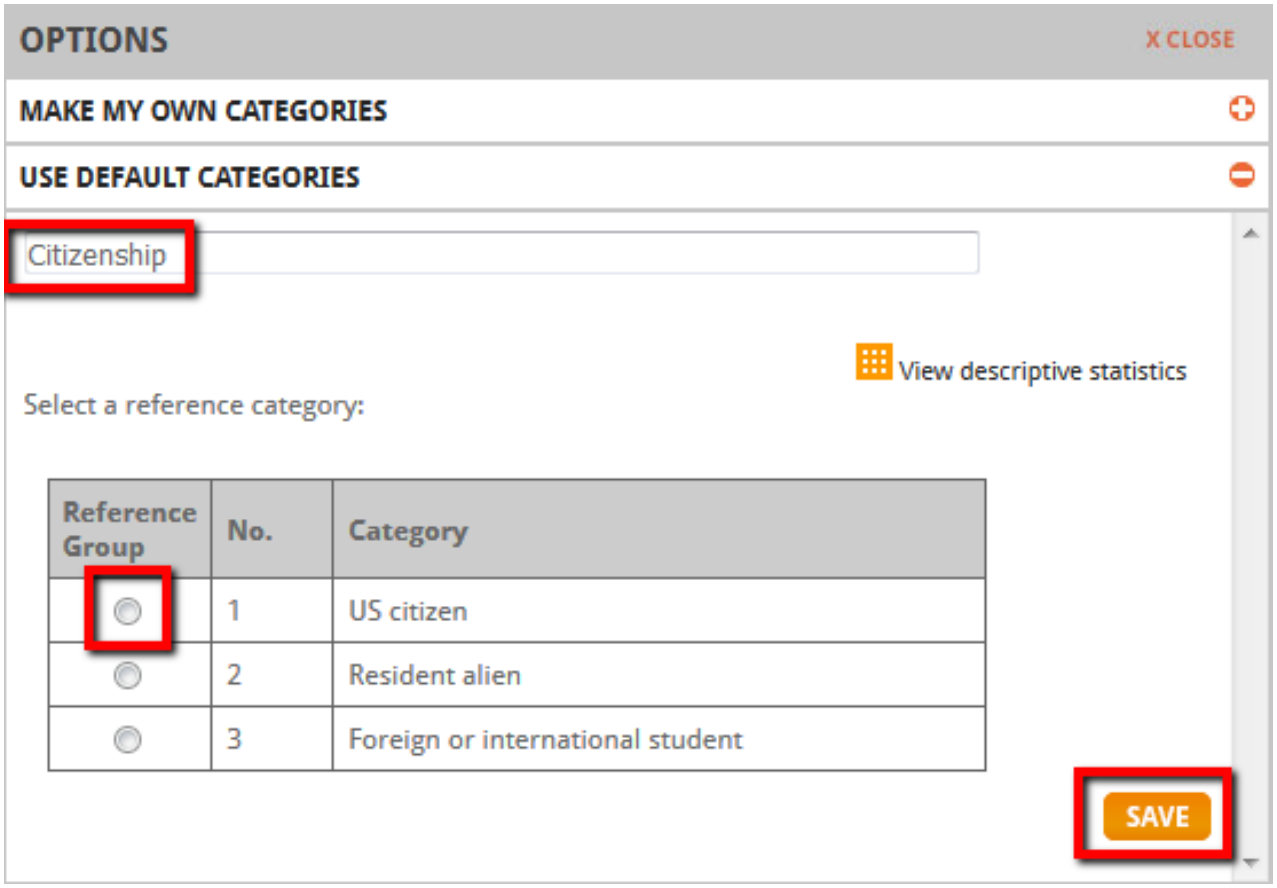
5. Click *US citizen* to use it as the reference group.

   Generally, you want to select the category with the largest number of responses as the reference group. Recall from the last page that 94% of survey respondents are US citizens.

   PowerStats will assign a value of zero to all US citizens and a value of one to all non-US citizens.

6. To make this distinction clear in the regression output, change the default variable label from *Citizenship* to *Non-citizen*.

7. Click *Save*.



4. SELECT INDEPENDENT VARIABLE 2

INSTRUCTIONS

1. In the Find Variables box, enter “income.”

   The matching variables appears in Search Results.

2. Click *Total income: Parents and independent.*

   In the Variable Information screen appears.

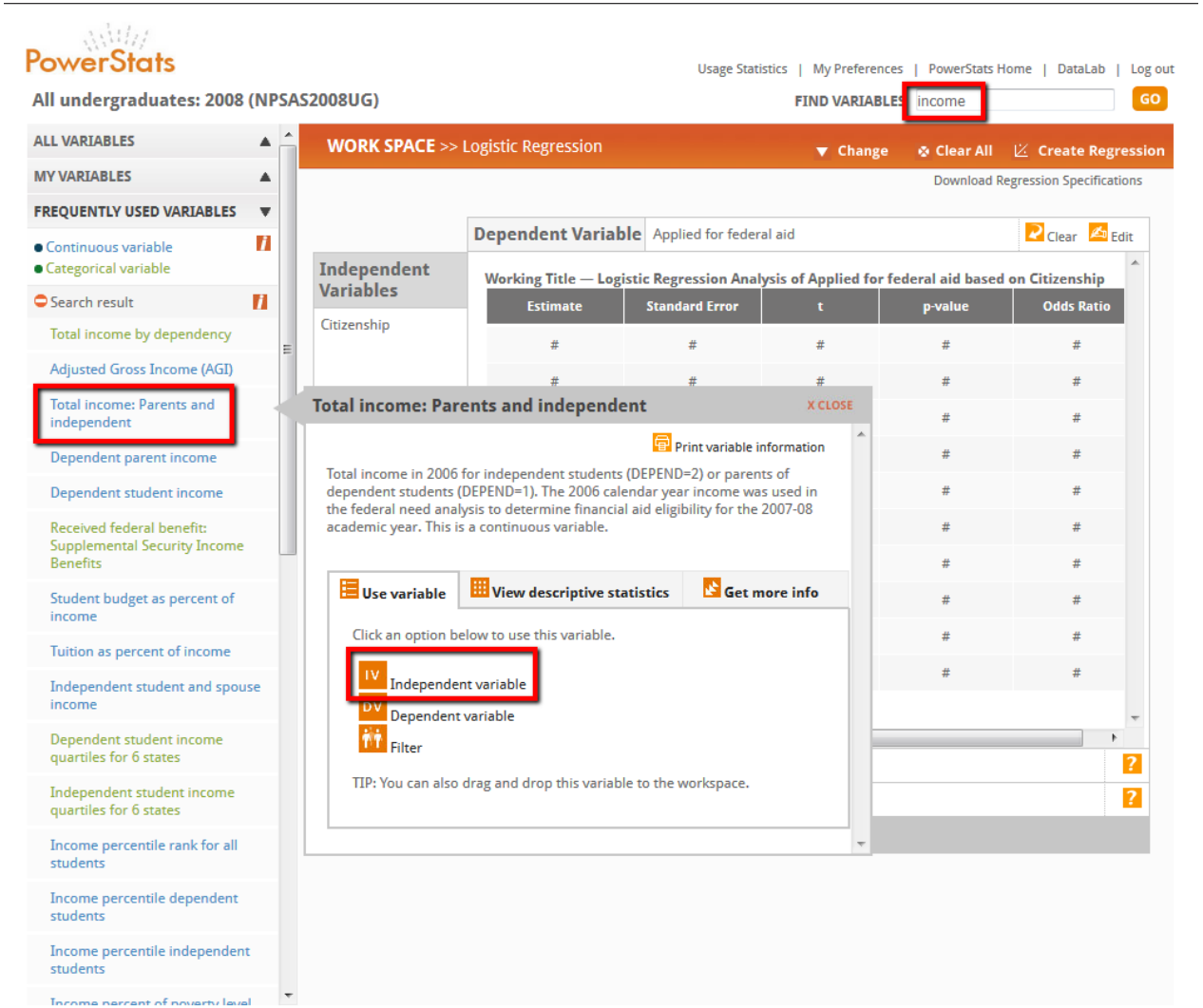
3. Click *Independent variable*.

   The Continuous Variable screen appears.

4. Click *As a continuous variable*.

   The Variable Options screen appears.

5. Leave *Consider zero as missing value* unchecked and click *Save*.



5. RUN THE REGRESSION

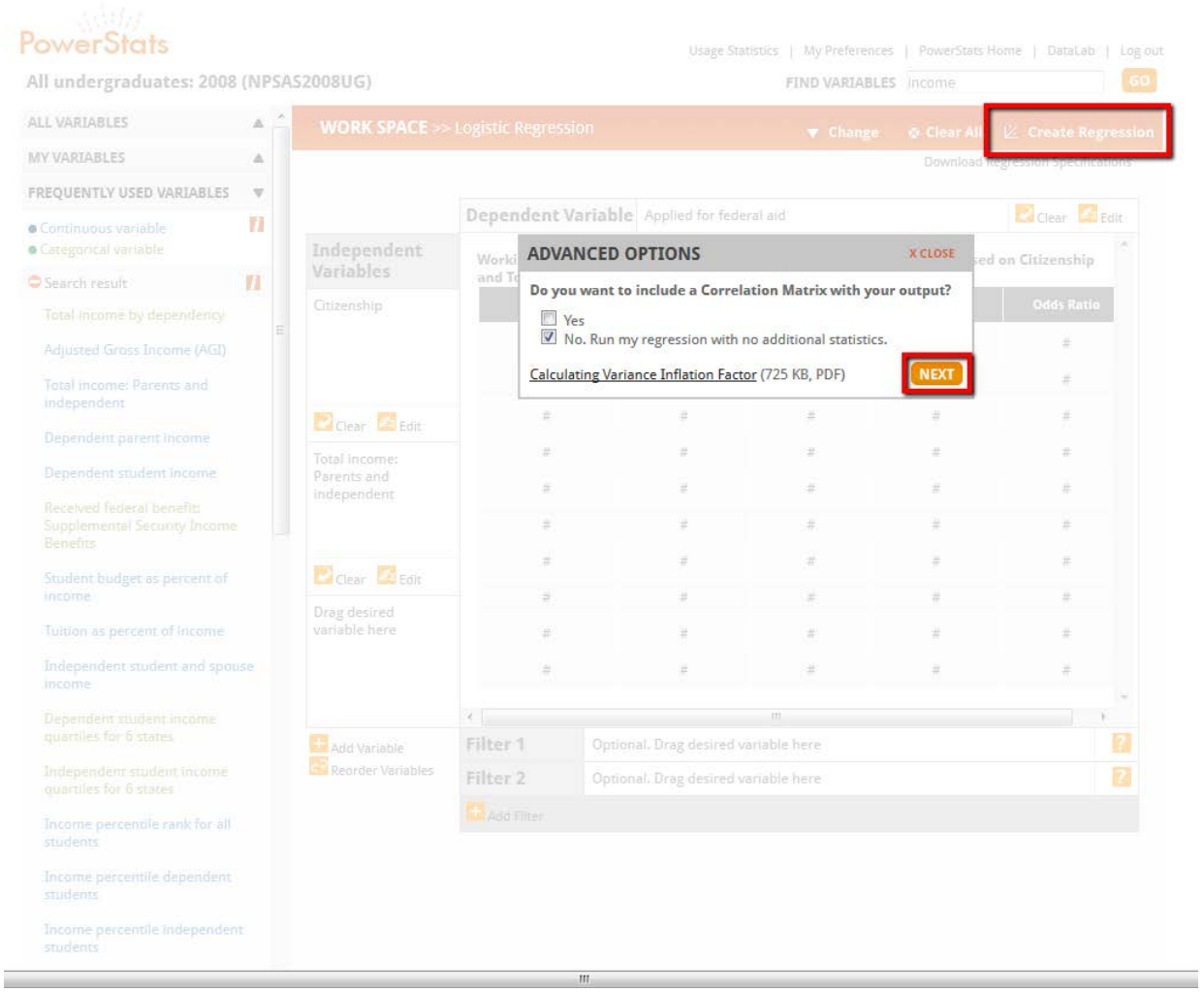
INSTRUCTIONS

1. Click *Create Regression*.

   The Advanced Options screen appears.

2. Without selecting additional statistics, click *Next*.

   PowerStats will begin to process the regression.

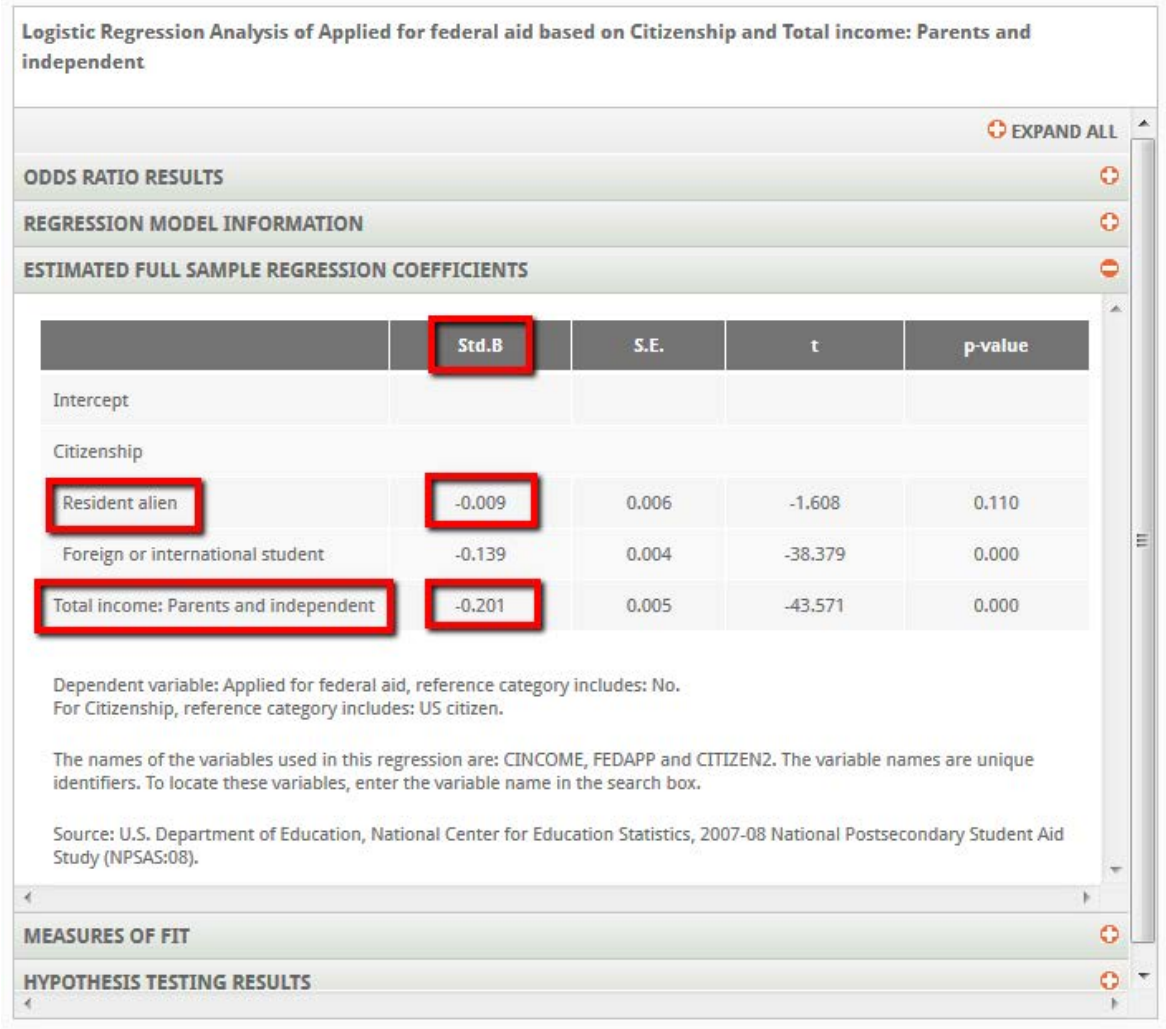


INTERPRETING LOGISTIC REGRESSION RESULTS

INTERPRETING LOGISTIC REGRESSION STANDARDIZED COFFICIENTS

PowerStats reports the results of logistic regression in terms of standardized regression coefficients (also called beta weights). Although standardized coefficients do not have any substantive interpretation, they share a single scale, and therefore can be compared with each other to assess relative magnitudes.

In this example, the comparison implies that the relationship between *Total income* and *Applied for federal aid* (-0.201) is over 20 times as strong as the relationship between *Resident alien* and *Applied for federal aid* (-0.009).



Logistic Regression Analysis of Applied for federal aid based on Citizenship and Total income: Parents and independent

EXPAND ALL

ODDS RATIO RESULTS

REGRESSION MODEL INFORMATION

ESTIMATED FULL SAMPLE REGRESSION COEFFICIENTS

	Std.B	S.E.	t	p-value
Intercept				
Citizenship				
Resident alien	-0.009	0.006	-1.608	0.110
Foreign or international student	-0.139	0.004	-38.379	0.000
Total income: Parents and independent	-0.201	0.005	-43.571	0.000

Dependent variable: Applied for federal aid, reference category includes: No.
For Citizenship, reference category includes: US citizen.

The names of the variables used in this regression are: CINCOME, FEDAPP and CITIZEN2. The variable names are unique identifiers. To locate these variables, enter the variable name in the search box.

Source: U.S. Department of Education, National Center for Education Statistics, 2007-08 National Postsecondary Student Aid Study (NPSAS:08).

MEASURES OF FIT

HYPOTHESIS TESTING RESULTS

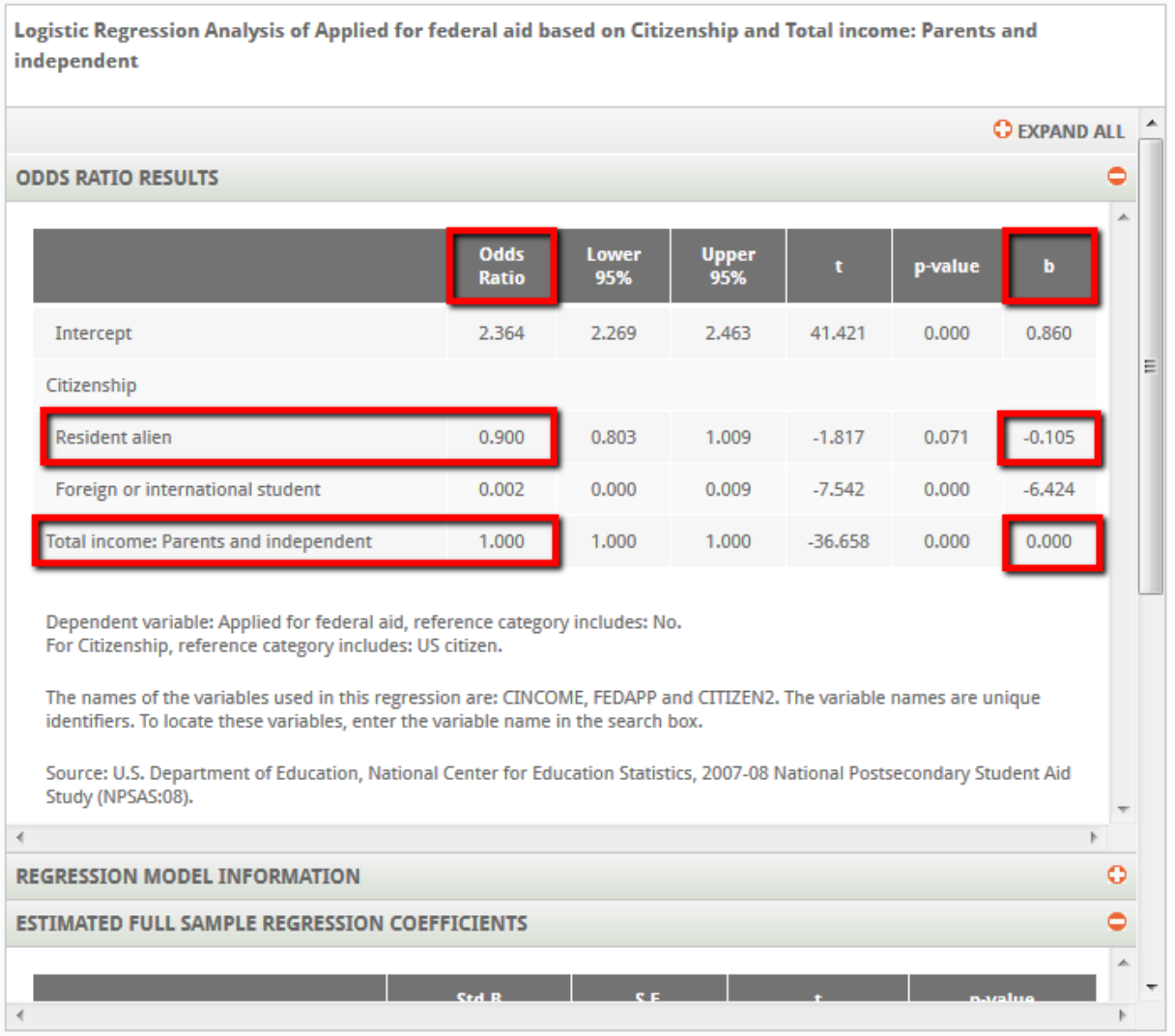
INTERPRETING LOGISTIC REGRESSION RESULTS, cont'd.

INTERPRETING LOGISTIC REGRESSION ODDS RATIOS

Odds ratios represent the proportional change in the probability that the dependent variable equals one for each additional unit of the independent variable, all else equal. For example, the odds of resident aliens applying for financial aid are only 0.900 as great as the odds of US citizens applying, holding Total income constant. In other words, resident aliens are almost 10% less likely than US citizens to apply for financial aid, all else equal.

In addition, PowerStats reports the natural logarithm, also known as the unstandardized coefficient, of each odds ratio. The log-odds ratio can be used to calculate the predicted probability that the dependent variable equals one for specific values of any independent variable.

Note
Although the odds ratio for *Total income* would seem to imply that income has no relationship with applying for financial aid, the t-statistic reveals a statistically significant negative association that is rounded to 1.000.



Logistic Regression Analysis of Applied for federal aid based on Citizenship and Total income: Parents and independent

EXPAND ALL

ODDS RATIO RESULTS

	Odds Ratio	Lower 95%	Upper 95%	t	p-value	b
Intercept	2.364	2.269	2.463	41.421	0.000	0.860
Citizenship						
Resident alien	0.900	0.803	1.009	-1.817	0.071	-0.105
Foreign or international student	0.002	0.000	0.009	-7.542	0.000	-6.424
Total income: Parents and independent	1.000	1.000	1.000	-36.658	0.000	0.000

Dependent variable: Applied for federal aid, reference category includes: No.
For Citizenship, reference category includes: US citizen.

The names of the variables used in this regression are: CINCOME, FEDAPP and CITIZEN2. The variable names are unique identifiers. To locate these variables, enter the variable name in the search box.

Source: U.S. Department of Education, National Center for Education Statistics, 2007-08 National Postsecondary Student Aid Study (NPSAS:08).

REGRESSION MODEL INFORMATION

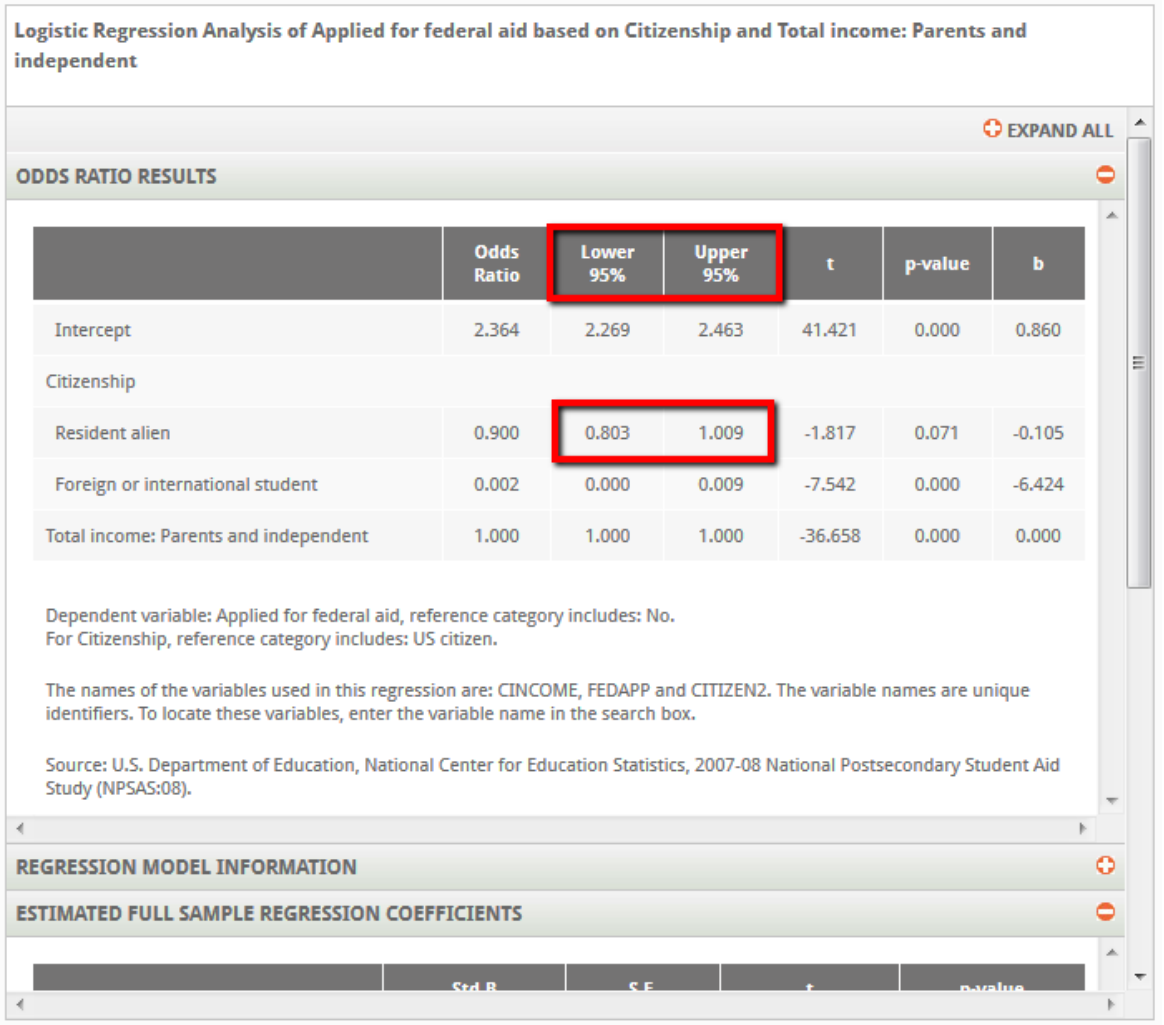
ESTIMATED FULL SAMPLE REGRESSION COEFFICIENTS

INTERPRETING LOGISTIC REGRESSION RESULTS, cont'd.

INTERPRETING CONFIDENCE INTERVALS

PowerStats reports 95% confidence intervals for odds ratios.

For illustration, the 95% confidence interval for *Resident alien* runs from 0.803 to 1.009.



Logistic Regression Analysis of Applied for federal aid based on Citizenship and Total income: Parents and independent

EXPAND ALL

ODDS RATIO RESULTS

	Odds Ratio	Lower 95%	Upper 95%	t	p-value	b
Intercept	2.364	2.269	2.463	41.421	0.000	0.860
Citizenship						
Resident alien	0.900	0.803	1.009	-1.817	0.071	-0.105
Foreign or international student	0.002	0.000	0.009	-7.542	0.000	-6.424
Total income: Parents and independent	1.000	1.000	1.000	-36.658	0.000	0.000

Dependent variable: Applied for federal aid, reference category includes: No.
For Citizenship, reference category includes: US citizen.

The names of the variables used in this regression are: CINCOME, FEDAPP and CITIZEN2. The variable names are unique identifiers. To locate these variables, enter the variable name in the search box.

Source: U.S. Department of Education, National Center for Education Statistics, 2007-08 National Postsecondary Student Aid Study (NPSAS:08).

REGRESSION MODEL INFORMATION

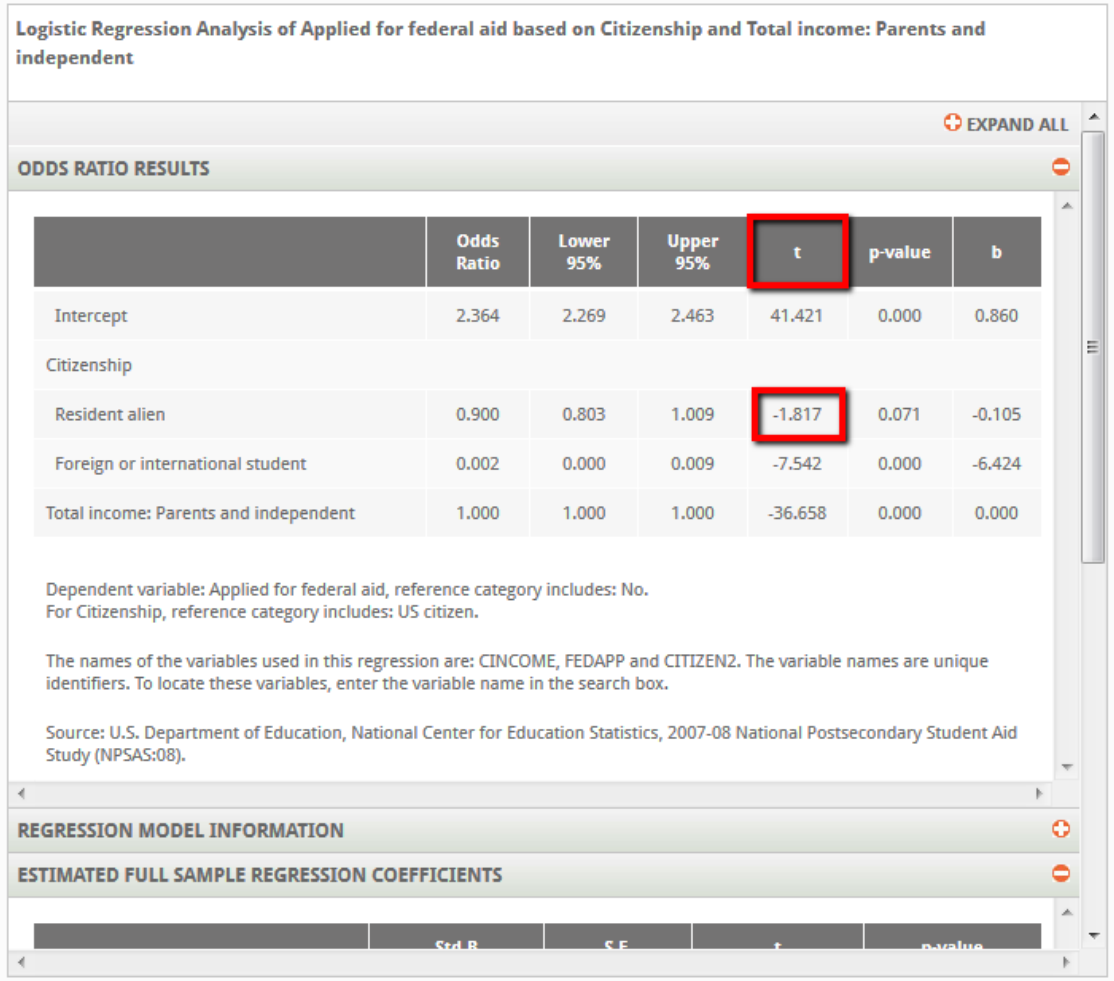
ESTIMATED FULL SAMPLE REGRESSION COEFFICIENTS

	Std B	S.E.	t	p-value

INTERPRETING LOGISTIC REGRESSION RESULTS, cont'd.

Student's *t* is the ratio of the logistic regression coefficient (reported under *Ln(Odds Ratio)*) divided by the standard error. If the absolute value of *t* is greater than or equal to a critical value (typically 1.96 except for small samples) the coefficient is said to be statistically significant at the 95% confidence level. That means that even if the true coefficient were as much as two standard errors higher or lower than what is estimated by the regression, it would still be different than zero.

In the case of *Resident alien*, the t-statistic is -1.817. Its absolute value, 1.817, is less than 1.96, so the coefficient for *Resident alien* is not statistically significant at the 95% level.



Logistic Regression Analysis of Applied for federal aid based on Citizenship and Total income: Parents and independent

EXPAND ALL

ODDS RATIO RESULTS

	Odds Ratio	Lower 95%	Upper 95%	t	p-value	b
Intercept	2.364	2.269	2.463	41.421	0.000	0.860
Citizenship						
Resident alien	0.900	0.803	1.009	-1.817	0.071	-0.105
Foreign or international student	0.002	0.000	0.009	-7.542	0.000	-6.424
Total income: Parents and independent	1.000	1.000	1.000	-36.658	0.000	0.000

Dependent variable: Applied for federal aid, reference category includes: No.
For Citizenship, reference category includes: US citizen.

The names of the variables used in this regression are: CINCOME, FEDAPP and CITIZEN2. The variable names are unique identifiers. To locate these variables, enter the variable name in the search box.

Source: U.S. Department of Education, National Center for Education Statistics, 2007-08 National Postsecondary Student Aid Study (NPSAS:08).

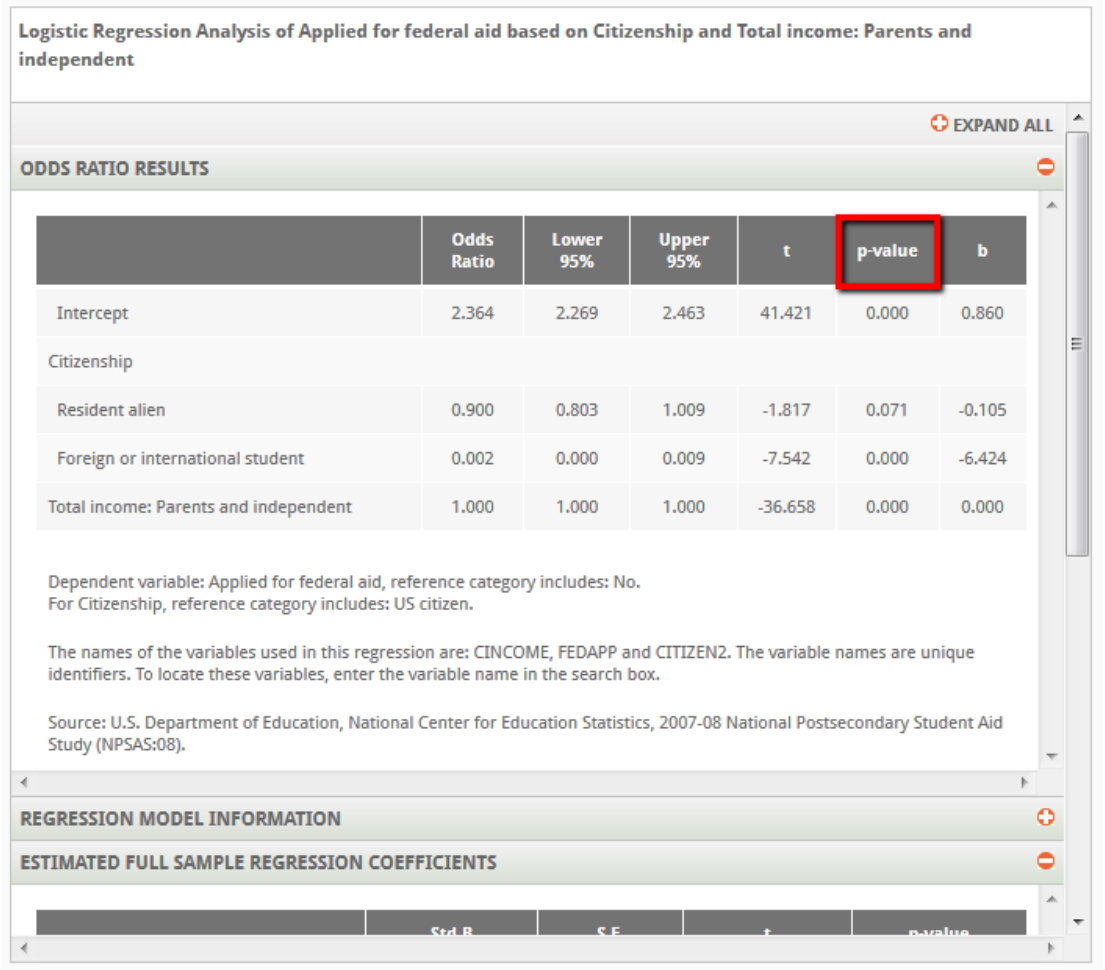
REGRESSION MODEL INFORMATION

ESTIMATED FULL SAMPLE REGRESSION COEFFICIENTS

INTERPRETING LOGISTIC REGRESSION RESULTS, cont'd.

The p-value, titled *Probability* in this window, measures the probability that a sample would have yielded a coefficient of this magnitude due to sampling error (also called sampling variation) if the true value of the coefficient were zero. Typically, a result is considered statistically significant if the p-value is less than .05.

NOTE
The p-values of .000 in this regression do not imply a zero likelihood that the coefficients were due to sampling error, but instead represent very small positive values less than .0005, and rounded to .000.



Logistic Regression Analysis of Applied for federal aid based on Citizenship and Total income: Parents and independent

EXPAND ALL

ODDS RATIO RESULTS

	Odds Ratio	Lower 95%	Upper 95%	t	p-value	b
Intercept	2.364	2.269	2.463	41.421	0.000	0.860
Citizenship						
Resident alien	0.900	0.803	1.009	-1.817	0.071	-0.105
Foreign or international student	0.002	0.000	0.009	-7.542	0.000	-6.424
Total income: Parents and independent	1.000	1.000	1.000	-36.658	0.000	0.000

Dependent variable: Applied for federal aid, reference category includes: No.
For Citizenship, reference category includes: US citizen.

The names of the variables used in this regression are: CINCOME, FEDAPP and CITIZEN2. The variable names are unique identifiers. To locate these variables, enter the variable name in the search box.

Source: U.S. Department of Education, National Center for Education Statistics, 2007-08 National Postsecondary Student Aid Study (NPSAS:08).

REGRESSION MODEL INFORMATION

ESTIMATED FULL SAMPLE REGRESSION COEFFICIENTS

OTHER FEATURES IN LOGISTIC REGRESSION

OTHER FEATURES

PowerStats reports other information from a logistic regression, organized in the following areas:

1. Model information (sample size, variable names, estimation method)
2. Statistics for hypothesis testing
3. Measures of goodness of fit

Click on the corresponding to view this information.