

A cluster of white numbers (2, 3, 4, 5, 6, 7, 8, 9) arranged in a semi-circular pattern above the title.

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.

The screenshot shows the PowerStats interface. On the left is the 'GROUP' sidebar with a 'QuickSelect by dataset name' dropdown and a list of categories: PRE-ELEMENTARY, SCHOOLS AND STAFFING SURVEY, SCHOOL SURVEY ON CRIME AND SAFETY, EDUCATION LONGITUDINAL STUDY, POSTSECONDARY, and Students. Under 'Students', 'All undergraduates' is selected and highlighted with a red box. On the right is the 'WORK SPACE' area. At the top, a 'Group' box contains 'All undergraduates', also highlighted with a red box. Below is a table with three columns representing different interview years: 2012, 2008, and 2004. The 2008 column is highlighted with a red box. Each column contains details about the dataset, including 'Includes' (General demographics, Types of aid and amounts received, Cost of attending college, Combinations of work, study, and borrowing, Enrollment patterns), 'Approximate number of respondents', and 'Study name'. At the bottom of each column is a 'SELECT' button, with the one in the 2008 column highlighted by a red box.

Group	Group	Group
Older years	Older years	Older years
were undergraduate students when interviewed in 2012.	were undergraduate students when interviewed in 2008.	were undergraduate students when interviewed in 2004.
Includes: <ul style="list-style-type: none">General demographicsTypes of aid and amounts receivedCost of attending collegeCombinations of work, study, and borrowingEnrollment patterns	Includes: <ul style="list-style-type: none">General demographicsTypes of aid and amounts receivedCost of attending collegeCombinations of work, study, and borrowingEnrollment patterns	Includes: <ul style="list-style-type: none">General demographicsTypes of aid and amounts receivedCost of attending collegeCombinations of work, study, and borrowingEnrollment patterns
Approximate number of respondents: 95,000	Approximate number of respondents: 113,500	Approximate number of respondents: 79,900
Study name: National Postsecondary Student Aid Study: 2012 Undergraduates Visit study website View technical information View all variable information, by subject View all variable information, by variable name	Study name: National Postsecondary Student Aid Study: 2008 Undergraduates Visit study website View technical information View all variable information, by subject View all variable information, by variable name	Study name: National Postsecondary Student Aid Study: 2004 Undergraduates Visit study website View technical information View methodology report View all variable information, by subject View all variable information, by variable name
View example tables	View example tables	View example tables
SELECT	SELECT	SELECT

2. CHOOSE A TYPE OF REGRESSION

INSTRUCTIONS

1. Click the *Logistic Regression* icon.

TIP

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

CHOOSE TYPE OF REGRESSION

 <p>LINEAR REGRESSION</p>	 <p>LOGISTIC REGRESSION</p>	 <p>CORRELATION MATRIX</p>
<p>Appropriate for <u>continuous or ordered categorical dependent variables</u>. Linear regression estimates the linear relationship, or best-fit line, between each independent variable and the dependent variable while controlling for all other independent variables.</p>	<p>Appropriate for dependent variables that take on two values, typically zero and one, or that are recoded into two values. Logistic regression estimates the relationship between each independent variable and the probability that the dependent variable equals one while controlling for all other independent variables.</p>	<p>Measures the linear association between a pair of variables. For a user-defined list of variables, PowerStats produces the correlation for each pair and presents them in a matrix.</p>
<p>See examples  (66 KB)</p>	<p>See examples  (72 KB)</p>	<p>See examples  (66 KB)</p>

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.

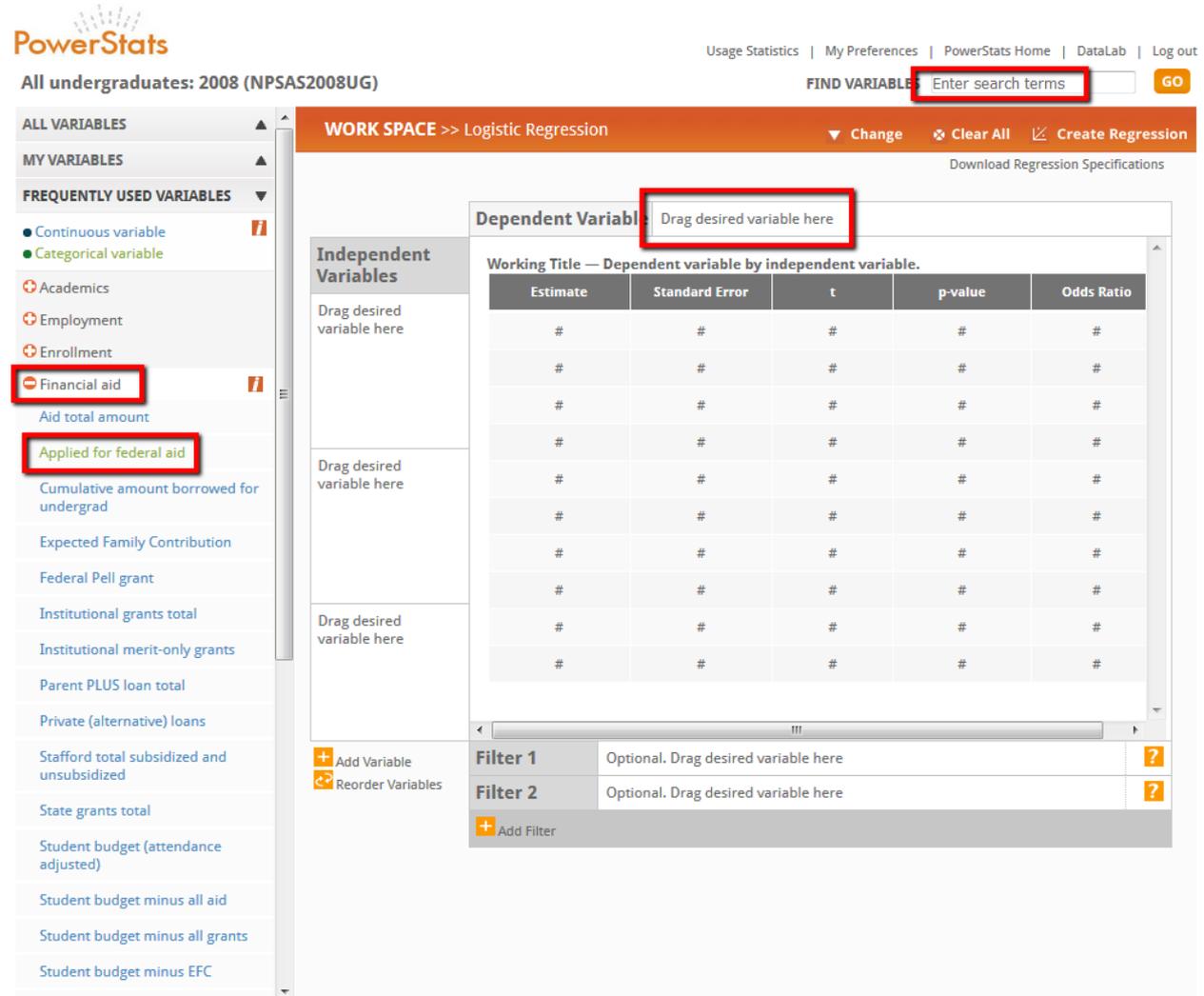
3. Click *View descriptive statistics*. Note that *Applied for federal aid* is a categorical variable with two values, No (0) and Yes (1).

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.



The screenshot shows the PowerStats interface for the dataset "All undergraduates: 2008 (NPSAS2008UG)". The "FIND VARIABLE" search bar contains "Enter search terms". The "WORK SPACE" is set to "Logistic Regression".

The left sidebar lists variables under "ALL VARIABLES", "MY VARIABLES", and "FREQUENTLY USED VARIABLES". The "Financial aid" variable is selected, and its sub-variables are listed: "Aid total amount", "Applied for federal aid", "Cumulative amount borrowed for undergrad", "Expected Family Contribution", "Federal Pell grant", "Institutional grants total", "Institutional merit-only grants", "Parent PLUS loan total", "Private (alternative) loans", "Stafford total subsidized and unsubsidized", "State grants total", "Student budget (attendance adjusted)", "Student budget minus all aid", "Student budget minus all grants", and "Student budget minus EFC".

The "Dependent Variable" box contains "Working Title — Dependent variable by independent variable." and a table with the following columns: Estimate, Standard Error, t, p-value, and Odds Ratio. The table is currently empty, with placeholder values "#".

The "Independent Variables" box contains three "Drag desired variable here" prompts.

At the bottom, there are filter options: "Filter 1" and "Filter 2", both with "Optional. Drag desired variable here" and a question mark icon. An "Add Filter" button is also present.

3. SELECT A DEPENDENT VARIABLE, cont'd.

- 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.

OPTIONS X CLOSE

MAKE MY OWN CATEGORIES +

USE DEFAULT CATEGORIES -

Applied for federal aid

 [View descriptive statistics](#)

Select a reference category:

Reference Group	No.	Category
<input checked="" type="radio"/>	0	No
<input type="radio"/>	1	Yes

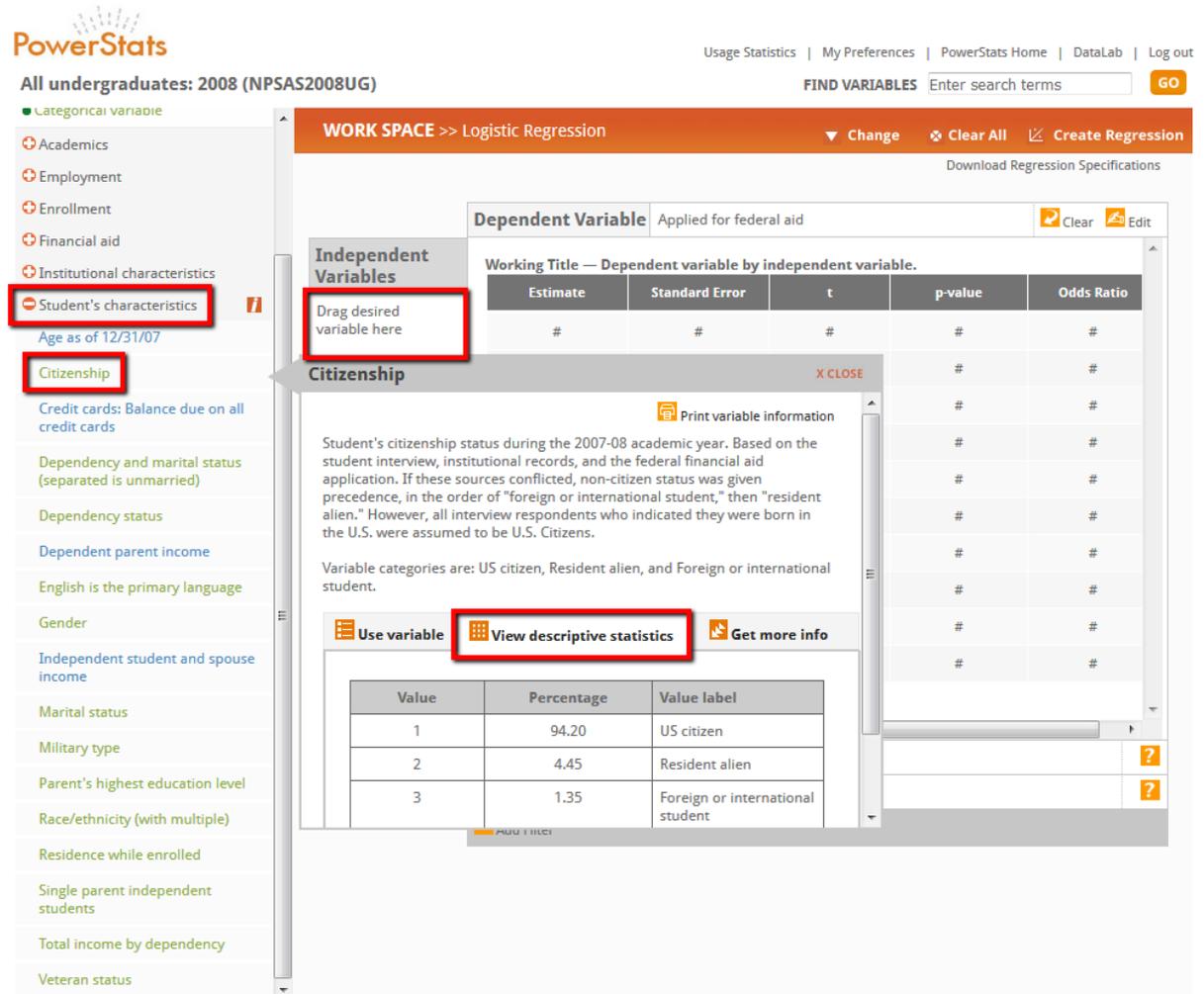
SAVE

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.



The screenshot shows the PowerStats interface for the dataset 'All undergraduates: 2008 (NPSAS2008UG)'. The 'Categorical variables' list on the left includes 'Student's characteristics' and 'Citizenship', both highlighted with red boxes. The 'WORK SPACE >> Logistic Regression' panel shows 'Applied for federal aid' as the dependent variable. The 'Independent Variables' section has a red box around the text 'Drag desired variable here'. The 'Citizenship' variable description is shown, with a red box around the 'View descriptive statistics' button. Below this, a table displays the distribution of citizenship status.

Value	Percentage	Value label
1	94.20	US citizen
2	4.45	Resident alien
3	1.35	Foreign or international student

4. SELECT INDEPENDENT VARIABLE 1, cont'd.

- 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.

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

OPTIONS X CLOSE

MAKE MY OWN CATEGORIES +

USE DEFAULT CATEGORIES -

Citizenship

View descriptive statistics

Select a reference category:

Reference Group	No.	Category
<input checked="" type="radio"/>	1	US citizen
<input type="radio"/>	2	Resident alien
<input type="radio"/>	3	Foreign or international student

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.

The screenshot shows the PowerStats interface for running a regression. The main workspace is titled "WORK SPACE >> Logistic Regression". A "Create Regression" button is highlighted with a red box. An "ADVANCED OPTIONS" dialog box is open, asking "Do you want to include a Correlation Matrix with your output?". The "No" option is selected, and the "NEXT" button is highlighted with a red box. The dialog also includes a link for "Calculating Variance Inflation Factor (725 KB, PDF)".

PowerStats
All undergraduates: 2008 (NPSAS2008UG)

Usage Statistics | My Preferences | PowerStats Home | DataLab | Log out

FIND VARIABLES income GO

WORK SPACE >> Logistic Regression Change Clear All Create Regression Download Regression Specifications

ALL VARIABLES
MY VARIABLES
FREQUENTLY USED VARIABLES

Continuous variable
Categorical variable

Search result

Total income by dependency
Adjusted Gross Income (AGI)
Total income: Parents and independent
Dependent parent income
Dependent student income
Received federal benefit: Supplemental Security Income Benefits
Student budget as percent of income
Tuition as percent of income
Independent student and spouse income
Dependent student income: quartiles for 6 states
Independent student income: quartiles for 6 states
Income percentile rank for all students
Income percentile dependent students
Income percentile independent students

Dependent Variable Applied for federal aid Clear Edit

Independent Variables
Citizenship Clear Edit

Total Income: Parents and independent Clear Edit

Drag desired variable here

Add Variable
Reorder Variables

Filter 1 Optional. Drag desired variable here
Filter 2 Optional. Drag desired variable here
Add Filter

ADVANCED OPTIONS X CLOSE
Do you want to include a Correlation Matrix with your output?
 Yes
 No. Run my regression with no additional statistics.
Calculating Variance Inflation Factor (725 KB, PDF) NEXT

INTERPRETING LOGISTIC REGRESSION RESULTS

INTERPRETING LOGISTIC REGRESSION STANDARDIZED COEFFICIENTS

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

	Std. B	S.E.	t	p-value
--	--------	------	---	---------

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 $\ln(\text{Odds Ratio})$) 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

	Std. B	S.E.	t	p-value
--	--------	------	---	---------

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

	Std. B	S.E.	t	p-value
--	--------	------	---	---------

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.

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
- MEASURES OF FIT
- HYPOTHESIS TESTING RESULTS