

IPEDS Analytics: Delta Cost Project Database 1987-2012

Data File Documentation

IPEDS Analytics: Delta Cost Project Database 1987-2012

Data File Documentation

APRIL 2015

Donna M. Desrochers

Jie Sun

American Institutes for Research

U.S. Department of Education

Arne Duncan

*Secretary***Institute of Education Sciences**

Sue Betka

*Acting Director***National Center for Education Statistics**

Peggy G. Carr

Acting Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high-priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high-quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public. Unless specifically noted, all information contained herein is in the public domain.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

NCES, IES, U.S. Department of Education
1990 K Street NW
Washington, DC 20006-5651

April 2015

The NCES Home Page address is <http://nces.ed.gov>.The NCES Publications and Products address is <http://nces.ed.gov/pubsearch>.

This publication is only available online. To download, view, and print the report as a PDF file, go to the NCES Publications and Products address shown above.

This report was prepared for the National Center for Education Statistics under Contract No. ED-IES-10-C-0016 with American Institutes for Research. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

Suggested Citation

Desrochers, D.M., and Sun, J. (2015). *IPEDS Analytics: Delta Cost Project Database 1987-2012* (NCES 2015-091). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved [date] from <http://nces.ed.gov/pubsearch>.

Content Contact

Richard Reeves

(202) 502-7436

richard.reeves@ed.gov

Contents

Overview	1
Design	2
File Updates for the IPEDS Analytics: Delta Cost Project Database, 1987-2012	8

Overview

The IPEDS Analytics: Delta Cost Project Database was created to make data from the Integrated Postsecondary Education Data System (IPEDS) more readily usable for longitudinal analyses. Currently spanning the period from 1987 through 2012, it has a total of 215,613 observations and 974 variables derived from the institutional characteristics, finance, enrollment, completions, graduation rates, student financial aid, and human resources IPEDS survey components and a limited number of outside sources.

The maintenance and hosting of the IPEDS Analytics: Delta Cost Project Database was taken over by the National Center for Education Statistics (NCES) in 2012. The database was originally created by the Delta Cost Project (an independent, nonprofit organization) in 2007. For a detailed history of the development of the database under the Delta Cost Project, which covers the 1987-2009 database, please refer to its location on the NCES website, http://nces.ed.gov/ipeds/deltacostproject/download/DCP_History_Documentation.pdf.

The database is posted online in two parts for easier downloading; the first part contains the file for the 1987-1999 academic years and the second for the 2000-2012 academic years. These files are intended to be merged together to create the full 1987-2012 database.

Design

The IPEDS Analytics: Delta Cost Project Database was created to make IPEDS data more readily usable for longitudinal analyses. The database contains one observation per institution for each year of data that is available; it includes all institutions that reported institutional characteristic data to IPEDS in the fall of each academic year. Some of the data have been adjusted to harmonize changes in financial reporting standards that occurred over time, by employing industry-accepted manipulations of the data. When possible, missing data were replaced via imputation. The database further improves the capacity for longitudinal analyses by creating consistent institutional groupings and matched sets to account for changes to the IPEDS universe of institutions over the time period. Additionally, variables to adjust the financial information to constant dollars have been included for the Consumer Price Index-Urban Consumers (CPI-U), the Higher Education Price Index (HEPI), and the Higher Education Cost Adjustment (HECA).

Institutional Groupings

NCES allows certain institutions (“parent institutions”) to report data for branch campuses or other affiliated institutions (“child institutions”) for various IPEDS surveys. Parent institutions may have one or more child institutions and these parent/child relationships may differ over time and/or by survey. Participation in combined reporting often depends on the type of survey—child institutions may report their own data on some surveys (e.g., enrollment or completions) while the parent institution may report their combined data on other surveys (e.g., finance). These reporting relationships can also change when affiliated institutions are opened or closed, and as a result parent/child reporting structures may change over time and/or cease to exist.

Institutions that reported data together because of a parent/child reporting relationship on any of the IPEDS surveys for any year between 1987 and 2012 have been grouped together for all years in order to maintain the consistency of the data for the entire time period. This means that all of the data for these parent/child institutions has been combined to make one observation per year for the related set of institutions. The exact number of groupings in the database fluctuates from year to year; for the 2012 academic year, there were 615 institutional groupings in the dataset. Of these institutional groupings present in the 2012 academic year, 179 are public, 126 are private nonprofit, and 310 are private for-profit.

Longitudinal Institution Panels

To allow for trends analyses that are not affected by institutions entering or leaving the dataset, the database includes variables to identify panels of institutions that report data consistently over specified time periods. These institutional panels are referred to as “matched sets.” To be included in the matched set, an institution must have data on three measures—fall full-time equivalent (FTE) student enrollment, instructional expenditures, and student completions—for every year in the panel time period. There are three different matched sets that cover different time periods: 1987-2012, 2002-2012, and 2007-2012.

Any institution that meets the matched set criteria is included in the matched set variables. However, institutions in the seven major Carnegie/sector classifications representing U.S. public and private nonprofit 4-year and 2-year institutions classified as Associate's, Baccalaureate, Master's, and Research institutions (according to the Carnegie 2010 Classifications) have received additional review. Institutions in the seven major Carnegie/sector classifications that have extreme outlier data, or have changed sectors or Carnegie Classification during the matched set time periods, are removed from the pertinent matched set.

The table below shows the institution counts for the three matched set panels for institutions in the seven major Carnegie/sector classifications.

Carnegie Classification 2010 by Sector	2007-2012 6-year matched set (matched_n_07_12_6)	2002-2012 11-year matched set (matched_n_02_12_11)	1987-2012 26-year matched set (matched_n_87_12_26)
Public Research	159	159	158
Public Master's	237	236	233
Public Bachelor's	81	80	75
Public Associate's	833	822	698
Private Nonprofit Research	101	100	98
Private Nonprofit Master's	339	333	323
Private Nonprofit Bachelor's	464	456	434

Matched set indicators are available for all types of institutions, including for-profit institutions. However, data users should evaluate the suitability of the matched set for alternate organizational groupings not shown in the table above. For example, only a small proportion of 2- and 4-year for-profit institutions meet the criteria for inclusion in the 26-year matched set. In addition, some institutions have changed sectors, which can result in an inconsistent number of institutions within the panel time periods when organized by alternate classifications. For-profit institutions in parent/child groups that are matched-set eligible may still show uneven data over time as a result of rapid change (e.g., openings and closings of child institutions) in this sector.

Data Harmonization

The Delta Cost Project harmonized the IPEDS finance data to provide comparable revenue and expenditure data over time and across different financial reporting standards, to the extent possible. These adjustments ensure reasonable consistency in the patterns over time and allow broad comparisons between public and private institutions. In the standard IPEDS data, many of the finance variables are not consistent over time because of changes introduced in the conversion of the Common Form reporting format to separate Governmental Accounting Standards Board (GASB) and Financial Accounting Standards Board (FASB) reporting formats. The IPEDS Analytics: Delta Cost Project Database includes original data reported to IPEDS as well as adjusted data used by the Delta Cost Project in its trend analyses. The large amount of information collected from IPEDS surveys precludes incorporating all IPEDS variables into the

Delta Database; priority was given to those variables expected to hold widespread interest among data users, and for which multiple years of data were available. Documentation describing how the adjusted finance variables were constructed is available in the IPEDS/Delta Cost Project Database Mapping File; please refer to its location on the NCES website, http://nces.ed.gov/ipeds/deltacostproject/download/IPEDS_DCP_Database_Mapping_File_87_12.xls.

The most notable revenue adjustments were to net tuition, federal grants and contracts, and auxiliary enterprise revenues. These adjustments were made to account for the inconsistencies caused by reporting revenue amounts net of “applied discounts and allowances” under FASB, and later, GASB reporting standards. Over the entire 1987-2012 period, the net tuition amounts in the Delta Database were standardized to reflect gross tuition revenue net of only institutional grant aid. Federal grant revenues were adjusted to exclude Pell grants (where applicable), as these are captured in the net tuition revenue amounts. Sales and service of auxiliary enterprise revenues are provided in gross amounts only.

For expenses, adjustments to the functional expenditure categories account for changes in the reporting of operations and maintenance (O&M) and interest across different reporting standards. Under the previous Common Form and GASB reporting formats, O&M and interest were separate expenditure categories; under the current FASB and New Aligned form reporting formats these amounts are embedded in the other functional expenditure categories. The main expense variables in the database were calculated to maintain consistency over time by subtracting O&M and interest from the functional expenditure categories and then summing those O&M and interest amounts separately to create variables representing total amounts.

In addition to adjusting the data to improve comparability across accounting standards, the database also includes numerous derived variables that were developed to translate accounting information into more useful concepts for institutions and policy audiences. For example, revenue variables were created to distinguish the amount of money coming from students, public sources, and private sources. Organizing funding streams into these categories allows for differentiation between revenues that are generally used at the institution’s discretion, or restricted for certain purposes (such as sponsored research, or hospitals and independent operations). Additional derived variables were created to put revenues in the context of expenditures, showing the portion of educational expenses that come from students compared to those expenses that are subsidized by the institution.

Derived expenditure variables also put spending in context for different institutional and policy purposes. For example, derived variables were constructed to isolate spending related to the academic, public service, and research missions at different types of institutions. These derivations allow for more accurate spending comparisons across different types of institutions by reorganizing spending around similar activities. Additional derived variables also put expenditures into the context of outcomes, showing how much an institution spends per degree or completion awarded in a given year. The Delta Cost Project Data Dictionary contains a full

listing of all variables and includes formulas used to construct the derived variables, please refer to http://nces.ed.gov/ipeds/deltacostproject/download/Delta_Data_Dictionary_1987_2012.xls.

Imputations

The Delta Cost Project IPEDS Database involves two different imputation procedures. The first imputation procedure is conducted yearly and utilizes a conservative methodology to address missing data wherever they may appear in the dataset. The second imputation procedure was implemented only once and was developed to account for changes in reporting standards over time for institutions following FASB accounting standards.

On-going, annual imputations. To maximize the analytic utility of the dataset, regression imputation is used as needed to replace missing data in eligible variables. Delta uses a relatively conservative method that imputes data only when there is a 1-year gap between two data values (e.g., missing 2003 data for a series would be imputed for if there were data for 2002 and 2004). If the gap between values is 2 years or more, the gap is not filled in. Furthermore, values are not imputed when data are missing at the beginning or end of the data series for an institution. There are imputation flags in the database to denote any instance where a value was imputed.

One-time FASB imputation. A second imputation procedure was developed and applied to improve the comparability between Common Form, FASB, and GASB expenditure data.¹ In this methodology, data were imputed for FASB-reporting institutions when institutional data on O&M and interest were unavailable from 1997 to 2003. Interest and O&M expense data were not reported by functional expense category for any FASB institution between 1997 and 2001, therefore each was separately imputed. This imputation process was also employed for institutions that did not report interest or O&M data (or reported partial data) for 2002 and 2003. Once O&M and interest data were available within each of the functional categories, they were subtracted from the total expenditures reported in each of the functional category, therefore improving comparability with other data reporting standards.

The specific methodology for imputing the missing interest and O&M data from 1997 to 2003 used data that were reported from 2002 to 2008. First, the reported interest and O&M in each functional expense category were computed separately as a share of total institutional expenditures in each year. Then, for each institution, an institutional median share was also determined for interest and O&M for each expense category across the 2002-2008 period; the institutional median was used in years when there was no reported share. For those institutions with no reported data for a particular expense category over the 2002-2008 period, a “peer group median share” was constructed using the median share from a set of institutions with the same Carnegie Classification and similar FTE and core expenditures (instruction, student

¹ Between 1997 and 2001, FASB institutions did not report interest or O&M as stand-alone expenditure categories in IPEDS. Thus, among FASB institutions, expenditure data within the functional categories were significantly higher during this period than in the prior and subsequent years when the interest and O&M had not been included, or was reported but could be removed.

services, academic support, and institutional support). The shares for interest and O&M (institutional shares, institutional median shares, or peer group median shares) were then applied to the total expenditures for all years, 1997-2003; imputed values were assigned where interest and O&M data were missing.

The sum of the interest and O&M data for each functional category were then scaled to ensure they summed, respectively, to the total interest expenditures reported by institutions and total O&M expenditures reported or previously imputed.² Finally, the O&M and interest data were subtracted from the functional expense category totals for all years, 1997-2003, to generate the new values for the functional expenditure categories.³

For a more detailed history of the development of the database, including data harmonization, groupings, imputations, and other processing issues from the 1987-2009 database, please refer to http://nces.ed.gov/ipeds/deltacostproject/download/DCP_History_Documentation.pdf.

Cautions to Users

As a service to the higher education community, NCES assumed responsibility for the Delta Cost Project IPEDS Database in 2012. In this caretaker role, NCES committed to: (a) provide annual updates to the database to bring in new data as it becomes available, (b) update institutional groupings as necessary, and (c) provide historically consistent imputations for data missing from the prior year where possible. Importantly, NCES considers the Database an analytic, rather than statistical, product.

To promote high-quality analysis, NCES has reviewed the database for compliance with NCES Statistical Standards. In so doing, a limited number of inconsistencies were noted about which users should be aware. These include: (a) percentage or share values that do not sum to 100 percent, (b) imputed values that are outside of the expected range, and (c) negative values where a negative amount is not feasible.

The majority of these inconsistencies appear related to imputation, specifically affecting variables where both total amounts and component parts are included in the database. Delta Cost Project imputation methodology did not consistently force the reconciliation of imputed component amounts to match reported totals, or vice versa. For example, if a component amount, such as salary expenses for academic support, was imputed, then it is possible for this amount to be greater than the total amount reported for academic support expenses as a whole. While it is rare for this mismatch to happen, it is possible using the Delta Cost Project imputation methodology and can result in unreasonable values for derived or imputed variables. In addition, a calculation such as the wage and salary share of total expenditures may result in a value greater than 100 percent.

² IPEDS did not collect data on total O&M spending by FASB-reporting institutions from 1997 to 2001; total O&M spending was previously imputed in the Delta Database for these years.

³ O&M was not subtracted from auxiliaries, hospitals, independent, and other operations to maintain comparability with GASB reporting.

The Delta Cost Project imputation methodology could not reconcile imputed components to match reported totals because the required data were not consistently collected by IPEDS. For example, in some survey years (e.g., Common Form 1987 to 1999, as well as FASB 2000 and 2001), only total expenditures and wage and salary expenditures were reported within each spending category. As a result, information about spending on the other total spending components (e.g., benefit expenses, interest, depreciation, or other expenses) necessary to control the imputed wage and salary expenditures to a reported total was unavailable. Even in those years when the FASB surveys collected this information (i.e., fiscal years 1997 through 1999, and 2002 to 2003) the data were often incomplete. Although reporting of component expenditure data is more reliable in later years, the Delta Database does not include all component data for each expenditure category, making it difficult to implement historical corrections that would force the components to sum to reported totals in those years.

In some instances, negative values are reported in the database. These occurrences may reflect legitimate negative amounts (e.g., investment losses), inconsistencies in published IPEDS datasets, or derived variable calculations that result in a negative amount. The Delta Database does not include corrections or adjustments to suspected reporting inconsistencies in IPEDS source data. The derived variable formulas are applied consistently across the database and those cases where the underlying data (reported or imputed) return a negative value are not adjusted.

To maintain consistency across the dataset, NCES continues to follow the Delta Cost Project methodology for subsequent database updates, including the imputation for missing data. Users should evaluate the sensitivity of their analysis to imputation, reporting both the extent of imputed data use and the analyst's assessment of the impact on imputation on their findings.

File Updates for the IPEDS Analytics: Delta Cost Project Database, 1987-2012

This section contains a summary of the changes incorporated into the IPEDS Analytics: Delta Cost Project Database 1987-2012. The changes described include those made since the 1987-2010 file was released (in August 2012), which includes importing the 2010-11 and 2011-12 IPEDS data into the database.

Changes to the 1987-2012 data file

1. New Variables

Variable	Label	Definition
carnegie2010	Carnegie Classification 2010	The 2010 Carnegie Classification includes all colleges and universities in the United States that are degree granting and accredited by an agency recognized by the U.S. Secretary of Education. The 2010 edition classifies institutions based on their degree-granting activities from the fall of 2008 through the spring of 2009.
carnegiegrp_2010	Carnegie Classification 2010 (Collapsed)	A collapsed version of the Carnegie Classification 2010.
carnegie_sector_2010	Carnegie Classification 2010 by Sector	The groupings created by the interaction of CarnegieGrp_2010 and Sector_revised.
ftall03pb	Full-time Age under 18 Postbaccalaureate All	Full-time postbaccalaureate (graduate and first professional) students under age 18
ftall04pb	Full-time Age 18-19 Postbaccalaureate All	Full-time postbaccalaureate (graduate and first professional) students age 18-19
ftall05pb	Full-time Age 20-21 Postbaccalaureate All	Full-time postbaccalaureate (graduate and first professional) students age 20-21
ftall06pb	Full-time Age 21-24 Postbaccalaureate All	Full-time postbaccalaureate (graduate and first professional) students age 22-24
ftall08pb	Full-time Age 25-29 Postbaccalaureate All	Full-time postbaccalaureate (graduate and first professional) students age 25-29
ftall09pb	Full-time Age 30-34 Postbaccalaureate All	Full-time postbaccalaureate (graduate and first professional) students Age 30-34
ftall10pb	Full-time Age 35-39 Postbaccalaureate All	Full-time postbaccalaureate (graduate and first professional) students age 35-39
ftall11pb	Full-time Age 40-49 Postbaccalaureate All	Full-time postbaccalaureate (graduate and first professional) students age 40-49
ftall12pb	Full-time Age 50-64 Postbaccalaureate All	Full-time postbaccalaureate (graduate and first professional) students age 50-64
ftall13pb	Full-time Age 65 and over Postbaccalaureate All	Full-time postbaccalaureate (graduate and first professional) students Age 65 and over
ftall14pb	Full-time Age unknown Postbaccalaureate All	Full-time postbaccalaureate (graduate and first professional) students age unknown
ptall03pb	Part-time Age under 18 Postbaccalaureate All	Part-time postbaccalaureate (graduate and first professional) students under age 18

ptall04pb	Part-time Age 18-19 Postbaccalaureate All	Part-time postbaccalaureate (graduate and first professional) students age 18-19
ptall05pb	Part-time Age 20-21 Postbaccalaureate All	Part-time postbaccalaureate (graduate and first professional) students age 20-21
ptall06pb	Part-time Age 21-24 Postbaccalaureate All	Part-time postbaccalaureate (graduate and first professional) students age 22-24
ptall08pb	Part-time Age 25-29 Postbaccalaureate All	Part-time postbaccalaureate (graduate and first professional) students age 25-29
ptall09pb	Part-time Age 30-34 Postbaccalaureate All	Part-time postbaccalaureate (graduate and first professional) students Age 30-34
ptall10pb	Part-time Age 35-39 Postbaccalaureate All	Part-time postbaccalaureate (graduate and first professional) students age 35-39
ptall11pb	Part-time Age 40-49 Postbaccalaureate All	Part-time postbaccalaureate (graduate and first professional) students age 40-49
ptall12pb	Part-time Age 50-64 Postbaccalaureate All	Part-time postbaccalaureate (graduate and first professional) students age 50-64
ptall13pb	Part-time Age 65 and over Postbaccalaureate All	Part-time postbaccalaureate (graduate and first professional) students Age 65 and over
ptall14pb	Part-time Age unknown Postbaccalaureate All	Part-time postbaccalaureate (graduate and first professional) students age unknown
ftallgrp1pb	Full-time Ages up to 24 Postbaccalaureate All	Full-time postbaccalaureate (graduate and first professional) students ages up to 24
ftallgrp2pb	Full-time Ages 25 to 34 Postbaccalaureate All	Full-time postbaccalaureate (graduate and first professional) students ages 25 to 34
ftallgrp3pb	Full-time Ages 35 to 49 Postbaccalaureate All	Full-time postbaccalaureate (graduate and first professional) students ages 35 to 49
ftallgrp4pb	Full-time Ages 50 and over Postbaccalaureate All	Full-time postbaccalaureate (graduate and first professional) students ages 50 and over
ptallgrp1pb	Part-time Ages up to 24 Postbaccalaureate All	Part-time postbaccalaureate (graduate and first professional) students ages up to 24
ptallgrp2pb	Part-time Ages 25 to 34 Postbaccalaureate All	Part-time postbaccalaureate (graduate and first professional) students ages 25 to 34
ptallgrp3pb	Part-time Ages 35 to 49 Postbaccalaureate All	Part-time postbaccalaureate (graduate and first professional) students ages 35 to 49
ptallgrp4pb	Part-time Ages 50 and over Postbaccalaureate All	Part-time postbaccalaureate (graduate and first professional) students ages 50 and over

Revised Variables

- *Inflation Variables*

Variables	Revision
CPI_scalar_2012 HEPI_scalar_2012 HECA_Scalar_2012	The scalar variables were recalculated to inflate financial data to 2012 constant dollar amounts rather than 2010 dollar amount; the variables were renamed to reflect this change.

- *Matched Set Variables*

Variables	Revision
matched_n_87_12_26 matched_n_02_12_11 matched_n_07_12_6	The matched set variables were advanced 2 years to reflect the new FY2011 and FY2012 data that was added. The number of institutions in the matched set will vary depending on whether Carnegie_sector_2000, Carnegie_sector_2005, or Carnegie_sector_2010 is used for analysis as institutions may change categories when new Carnegie Classification are introduced. The matched set variables only include institutions in the United States (excluding U.S. territories) that have consistently reported data on fall FTE student enrollment, data on instructional spending, and completions. Some institutions with complete data were removed from the matched set because they contained extreme outlier data.

- *Fall Staff Variables*

Variables	Revision
ftall1 – ftall8 ptall1 – ptall8 derived staffing variables	Fall staff data were collected bi-annually between 1987 and 2002 with data reported in the even-numbered academic years; beginning in 2003 institutions had the option of reporting fall staff data in odd-numbered years. Imputed data for all staffing variables were removed from the database (and derived variables recalculated) because it was determined that they contained unrealistic values in some Carnegie sector categories and years. As a result, staffing data reported in even-numbered academic years is the most complete.

Revised institutional groupings

Whenever an institution serving as a “parent institution” includes a new “full child” institution(s) in its data reporting, the institutions are grouped together in the database. As long as the new “full child” institution has not previously reported data to IPEDS, the inclusion of the child institution’s data with the parent institution’s data does not change the information that was previously published in earlier data files.

Occasionally, institutions that previously reported their own data establish new parent/child relationships or merge together. When these relationships are identified, the affected institutions are grouped together for the entire span of the database, which may result in revisions to data published in earlier versions of the database. The table below lists the preexisting institutional groupings that established new parent/child relationships in 2011 or 2012 and may contain revised grouped data for either all or some of the years previously published in the 1987-2010 database.

Group ID	Institution Name
1592	Pennsylvania State University-Main Campus
1668	Oconee Fall Line Technical College
2450	Carrington College California-Sacramento
2511	Trend Setters Academy-Louisville
2515	Vanguard College of Cosmetology
2530	Salter College-West Boylston
2546	Regency Beauty Institute
2574	Harris School of Business
2588	Mildred Elley School
2617	Lincoln College of Technology-Dayton
2663	Cosmetology Career Center LLC
2676	University of Cosmetology Arts & Sciences
2711	National University College-Bayamon
2731	Platt College-Los Angeles
2760	Utah College of Massage Therapy Inc.
2796	DeVry University-Georgia
2799	DeVry University-California
2807	DeVry University-Virginia
2809	John Dewey College
2820	DeVry University-Florida
2826	DeVry University-Ohio