

Vicky Dingler

November 3, 2016

Use SAS to Compare Text Strings



Search



Surveys and Data Collection

Proven methods and technological innovations in survey science that deliver answers for an increasingly data-driven world

Cutting-edge survey research has been a cornerstone of our work since our founding. Federal agencies, commercial clients, and



Find the Right College for You

Use College Navigator to search and compare information about postsecondary institutions across the country.

The National Center for Education Statistics (NCES) is the primary federal entity for collecting and analyzing data related to education.

About Us

Surveys and Programs

Publications and Products

Data and Tools

[Learn more about NCES](#)

WHAT'S NEW AT NCES

OCT
27

[NAEP Science 2015 State Snapshot Reports](#)

Each state and jurisdiction that participated in the NAEP 2015 Science assessment receives a one-page snapshot report that presents key findings and trends in a condensed format. [» More info](#)

OCT
25

[Trends in High School Dropout and Completion Rates in the United States: 2013](#)

This report draws on an array of nationally representative surveys and administrative datasets to present statistics on high school dropout and completion rates. [» More info](#)

THE CONDITION OF EDUCATION



DATALAB *Fast, flexible, powerful.*

Questions? Contact NCES
nces.info@rti.org



KEY FEATURES OF POWERSTATS

1 2 3



PERCENTAGE DISTRIBUTION

General percentage of a population in each category of a variable, displayed in columns.



AVERAGES, MEDIANS, & PERCENTS

Computes any of three statistics (averages, medians, or percentages) for your selected variables, displayed in columns.



CENTILES

Produces values of a continuous variables at centiles, displayed in columns.

GO TO:

[PRE-ELEMENTARY EDUCATION](#)

[SCHOOLS AND STAFFING SURVEY](#)

[SCHOOL SURVEY ON CRIME AND SAFETY](#)

[HIGH SCHOOL LONGITUDINAL STUDY](#)

[EDUCATION LONGITUDINAL STUDY](#)

[POSTSECONDARY EDUCATION](#)



National Center for Education Statistics

Table 1. Percentage distribution of cumulative persistence and attainment at any institution of first-time beginning students in 2003–04, by initial degree program, receipt of Pell Grants, and dependency status: Spring 2009

Student characteristic	Total students (in thousands)	Cumulative persistence and attainment as of spring 2009		Total
		Attained or persisting as of spring 2009	No degree, left without return, as of spring 2009	
Degree-seeking¹ students				
All students	3,007	65.9	34.1	100.0
Dependent students	2,333	71.7	28.3	100.0
Independent students	673	45.8	54.2	100.0
Cumulative Pell through 2009				
No Pell through 2009	1,690	67.5	32.5	100.0
Dependent students	1,403	72.9	27.1	100.0
Independent students	286	41.5	58.5	100.0
Any Pell through 2009	1,317	63.9	36.1	100.0
Dependent students	930	70.1	29.9	100.0
Independent students	387	49.1	50.9	100.0
Certificate seeking students				
All students	422	64.4	35.6	100.0
Dependent students	155	70.4	29.7	100.0
Independent students	267	60.9	39.1	100.0
Cumulative Pell through 2009				
No Pell through 2009	91	65.0	35.0	100.0
Dependent students	34	69.1	30.9	100.0
Independent students	57	62.4	37.6	100.0
Any Pell through 2009	331	64.2	35.8	100.0
Dependent students	120	70.7	29.3	100.0
Independent students	211	60.5	39.5	100.0
Non-degree-seeking students				
All students	318	51.3	48.7	100.0
Dependent students	190	62.3	37.8	100.0

Varinfo, valcodes, itemword

pell_varinfo.xlsx - Excel

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW Dingler, Victoria

Clipboard Font Alignment Number Styles Cells Editing

F2 : X ✓ fx Analysis ID

	A	B	C	D	E	F	G	H	I
1	var_name	var_type	format	recfmt	var_label	prefix		file_id	file_r
2	NP04ID	X		6	Analysis ID	Survey sample		F9PELL	
3	PLSCHL2	X		6	Pell Attending institution code (OPEID)	Financial aid: Pell grant		F9PELL	
4	PLBR2	X		2	Pell institution branch	Financial aid: Pell grant		F9PELL	
5	PGM	D		2 A	Pell Grant type	Financial aid: Pell grant		F9PELL	
6	PLYEAR	X		4	Pell grant year	Financial aid: Pell grant		F9PELL	
7	PLSCHL1	X		6	Pell institution code (OPEID)	Financial aid: Pell grant		F9PELL	
8	PLBR1	X		2	Pell institution branch	Financial aid: Pell grant		F9PELL	
9	PLAMTSCH	C		4	Pell grant scheduled amount	Financial aid: Pell grant		F9PELL	
10	PLAMTP1	C		4	Pell grant paid amount	Financial aid: Pell grant		F9PELL	
11	PLAMTP2	C		4	Pell grant remaining	Financial aid: Pell grant		F9PELL	

13_specs

pell_valcodes.xlsx [Last saved by user] - Excel

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW Dingler, Victoria

Clipboard Font Alignment Number Styles Cells Editing

A20 : X ✓ fx 13

	A	B	C	D	E	F	G	H	I
1	file_no	var_name	val_code	val_label					
2	13	NP04ID		1 {Continuous}					
3	13	PLSCHL2		1 {Continuous}					
4	13	PLBR2		0 {zero}					
5	13	PLBR2		1 {Continuous}					

TextPad - M:\RTI\NPSAS 16\Team Building Meeting\pell_itemword.txt

File Edit Search View Full Screen Macros Configure Window Help

pell_itemword.txt X

```

Variable name: NP04ID
Description: Analysis ID.
Notes:
Code:
Question Wording:
Item Word:
Applies to:
Source: Survey sample

Variable name: PLSCHL2
Description: Attending Campus School Code.
Notes:
Code:
Question Wording:
Item Word:
Applies to:
Source: National Student Loan Data System (August 2009)

Variable name: PLBR2
Description: Attending Campus School Branch Code.
Notes:
Code:
Question Wording:
Item Word:
Applies to:
Source: National Student Loan Data System (August 2009)

Variable name: PGM
Description: Grant Category Type.
Notes:
Code:
Question Wording:
Item Word:
Applies to:
Source: National Student Loan Data System (August 2009)

Variable name: PLYEAR
Description: Pell paid year (FY).
Notes:
Code:
Question Wording:
Item Word:
Applies to:
Source: National Student Loan Data System (August 2009)
    
```

Compare vars between xls and ds

```
data one;
input vname $ 1-15;
file1='xls';
cards;
S15DFRECN_12Y
S15DFRENR_12Y
S15DFRFAM_12Y
;
proc sort; by vname; run;
```

```
data two;
input vname $ 1-15;
file2='ds';
cards;
S15EVRDFR_12Y
S15DFRENR_12Y
S15DFRFAM_12Y
;
proc sort; by vname; run;
```

```
data three;`
merge one two; by vname;
proc print; where file1='xls' and file2="";
proc print; where file1="" and file2='ds';
run;
```

Compare vars between xls and ds – log, listing

```
39 data three;  
40 merge one two; by vname;
```

NOTE: There were 3 observations read from the data set WORK.ONE.
NOTE: There were 3 observations read from the data set WORK.TWO.
NOTE: The data set WORK.THREE has 4 observations and 3 variables.

```
41 proc print; where file1='xls' and file2='';
```

NOTE: There were 1 observations read from the data set WORK.THREE.
WHERE (file1='xls') and (file2='');

```
42 proc print; where file1='' and file2='ds';  
43 run;
```

NOTE: There were 1 observations read from the data set WORK.THREE.
WHERE (file1='') and (file2='ds');

Obs	vname	file1	file2
-----	-------	-------	-------

2	S15DFRECN_12Y	xls	
---	---------------	-----	--

The SAS System

Obs	vname	file1	file2
-----	-------	-------	-------

10	S15EVRDFR_12Y		ds
----	---------------	--	----

Compare vars between vars and vals tab

```
data one;
input vname $ 1-15;
file1='var';
cards;
S15DFRECN_12Y
S15DFRENR_12Y
S15DFRFAM_12Y
S15DFRGOV_12Y
S15DFRMIL_12Y
;
proc sort; by vname; run;
```

```
data two;
input vname $ 1-15;
file2='val';
cards;
S15EVRDFR_12Y
S15EVRDFR_12Y
S15EVRDFR_12Y
S15EVRDFR_12Y
S15DFRECN_12Y
S15DFRECN_12Y
S15DFRECN_12Y
S15DFRECN_12Y
S15DFRENR_12Y
S15DFRENR_12Y
S15DFRENR_12Y
S15DFRENR_12Y
;
proc sort; by vname; run;
```

```
data twoa; set two; by vname; if first.vname;
data three;
merge one twoa; by vname;
proc print; where file1='var' and file2="";
proc print; where file1=" and file2='val';
run;
```

Compare vars between val and var – log, listing

NOTE: There were 12 observations read from the data set WORK.TWO.

NOTE: The data set WORK.TWO has 12 observations and 2 variables.

NOTE: PROCEDURE SORT used (Total process time):

```
real time      0.00 seconds
cpu time       0.00 seconds
```

```
40 data twoa; set two; by vname; if first.vname;
```

NOTE: There were 12 observations read from the data set WORK.TWO.

NOTE: The data set WORK.TWOA has 5 observations and 2 variables.

```
42 data three;
```

```
43 merge one twoa; by vname;
```

NOTE: There were 5 observations read from the data set WORK.ONE.

NOTE: There were 5 observations read from the data set WORK.TWOA.

NOTE: The data set WORK.THREE has 6 observations and 3 variables.

```
44 proc print; where file1='var' and file2='';
```

NOTE: There were 1 observations read from the data set WORK.THREE.

```
WHERE (file1='var') and (file2=' ');
```

NOTE: PROCEDURE PRINT used (Total process time):

```
real time      0.03 seconds
cpu time       0.01 seconds
```

```
45 proc print; where file1="" and file2='val';
```

```
46 run;
```

NOTE: There were 1 observations read from the data set WORK.THREE.

```
WHERE (file1=' ') and (file2='val');
```

```
*****
```

```
Obs   vname   file1  file2
```

```
5   S15DFRMIL_12Y   var
```

```
Obs   vname   file1  file2
```

```
6   S15EVRDFR_12Y           val
```

Compare variable labels

```
data one;
input vname $ 1-8 label1 $ 10-80;
cards;
PRLVL3Y      Attainment or level at last institution enrolled through 2006
PRLVL6Y      Attainment or level at last institution enrolled through 2009
PROUT6       Cumulative persistence and attainment anywhere 2008-09
PROUTF3      Cumulative retention and attainment at first institution 2005-06
PROUTFI6     Cumulative retention and attainment at first inst 6-yr total 2009
;
proc sort; by vname;
```

```
data two;
input vname $ 1-8 label2 $ 10-80;
cards;
PRLVL3Y      Attainment or level at last institution enrolled through
PRLVL6Y      Attainment or level at last institution enrolled through 2009
PROUT6       Cumulative persistence and attainment anywhere 2008-09
PROUTF3      Cumulative retention and attainment at first institution 2005-06
PROUTFI6     Cumulative retention and attainment at first inst 6-yr total 2009
;
proc sort; by vname;
```

```
data three;
merge one two; by vname;
proc print; where label1 ne label2;
run;
```

```
*****
```

```
Obs vname label1
```

```
1 PRLVL3Y Attainment or level at last institution enrolled through 2006
label2
1 Attainment or level at last institution enrolled through
```

Compare standard deviations

```
*compare t1 output with nomoss listing from orig data;
data one;
input vname $ 1-12 stdt1 13-30;
cards;
ID                7060.17          100001.00          124766.00
SCHORDER          0.4809778         1.0000000         5.0000000
SCHINDEX          0.4809778         1.0000000         5.0000000
UNITID            105703.56         100654.00          480569.00
STATE             15.0821178         1.0000000         55.0000000
CONTROL           0.8889578         1.0000000         3.0000000
CONTROL14         0.8889578         1.0000000         3.0000000
LEVEL             0.5758684         1.0000000         3.0000000
LEVEL14           0.5762927         1.0000000         3.0000000
;
proc sort; by vname;
data two;
input vname $ 1-12 stdo 14-27;
cards;
BPS14ID           7060.17
SCHORDER          0.4809778
SCHINDEX          0.4809778
UNITID            105703.56
STATE             15.0821178
CONTROL           0.8889578
CONTROL14         0.8889578
LEVEL             0.5758684
LEVEL14           0.5762927
;
proc sort; by vname;

data three;
merge one two; by vname;
proc print; where stdt1 ne stdo; run;
```

Obs	vname	stdt1	stdo
1	BPS14ID	.	7060.17
45	ID	7060.17	.

More Information

Vicky Dingler

Sr. Data Analyst

vdingler@rti.org

nces.ed.gov

rti.org