

“Let’s Get SAS[®] To Do It!”

Getting Data from SUDAAN[®] to SAS[®] to EXCEL[®]

Anna Vincent,
Research Specialist,
Behavioral Risk Factor Surveillance System (BRFSS),
Center for Health Statistics (CHS),
Texas Department of State Health Services (DSHS)

Abstract

This paper will show that a person can be new to SAS and create beneficial reports. Armed with the SAS Basic Programming Essentials Training Manual and using the ODS and ExcelXP Tagset, data results will be moved from SUDAAN to SAS and finally into EXCEL without touching the data, using the names of the risk factor variables, or how many values those variables had. The jumbled SUDAAN output will be transformed into something that is informative and easy to read. And when my supervisor wanted charts with confidence intervals, the simple solution was incorporating a chart template into the syntax!

Introduction

“The Texas Behavioral Risk Factor Surveillance System (BRFSS), initiated in 1987, is a federally funded (CDC) telephone survey conducted on a monthly basis of randomly selected adult Texans to collect data on lifestyle risk factors contributing to the leading causes of death and chronic diseases. As a primary source for comprehensive statewide data on preventive health practices and health risk behaviors, BRFSS is an important tool for decision-making throughout DSHS and the public health community. Public and private health authorities at the federal, state, and local levels rely on BRFSS to identify public health problems, design policies and interventions, set goals, and measure progress toward those goals.”

BRFSS webpage: <http://www.dshs.state.tx.us/chs/brfss/default.shtm>

Since the BRFSS data touches almost every program in the state health department, data needs will always be different based on the program requesting it. Programs may want one or all of the standard risk factors and health indicators for one or more counties, or they might want different aggregations of age groups, education groups, or income groups. In addition, sometimes the way a question is asked will change from time to time, or the answers to those questions will change. For example, the question on diabetes changed from two answers to four answers. The BRFSS is very fluid and dynamic, and the results produced from it need to be as well.

Typically, data results are obtained in either SUDAAN or SPSS and then hand typed or cut and pasted into pre-designed tables. This is time consuming and prone to errors. A new solution was needed that was fast and readable for clients, to make reports that had titles, footnotes, confidence intervals, demographics and current dates for sometimes up to 130 tables!

Getting Started

We start in SUDAAN as they have the “gold standard” for confidence intervals (CI) that the Centers for Disease Control (CDC) desire. SAS 9.2 currently gives only symmetrical CI so we

must start in SUDAAN which produces asymmetrical CIs and is easier to export into SAS than SPSS.

This is the code that we use:

```
proc crosstab data = "H:\brfss_data\brfss09\state\state_09_Working_av.sav" design = wr filetype= spss
  nowgt nocol;
  nest STSTR seqno/ missunit;
  weight cdcfinwt;
  setenv linesize = 106 pagesize = 45 colwidth = 11 topmgn = 1;
  subgroup agegr4 raceeth sex educat4 inccat3b admin2 fairpoor keptusal c06q01 diabetes rfsmok
c03q04 chckup1yr ;
  levels 4 4 2 4 3 8 2 2 4 2 2 8
  2 ;
  tables admin2*(fairpoor keptusal c06q01 diabetes rfsmok c03q04 chckup1yr)*(sex raceeth agegr4
  educat4 inccat3b);
  print nsum rowper lowrow uprow;
  output nsum rowper lowrow uprow /
  filename="H:\Brfss\avincent\Myrepresentations\master"
  filetype=SASXPORT replace
  Levfile="H:\Brfss\avincent\Myrepresentations\mastrlev";
```

Texas BRFSS data files are currently managed in SPSS. With the **subgroup** it is important that known demographics need to come first, the geographic area second (if necessary) and then all risk factor variables to follow. Near the bottom are the few lines of code that will export the tables into SAS. The Levfile is what your format file is named.

From SUDAAN: "LEVFILE=filename: Optional parameter, standalone SUDAAN versions only. Use LEVFILE=filename with SASXPORT input files to specify the name of an optional file containing records with SAS-style format information. If this parameter is not supplied, SUDAAN will look for a file named LEVFILE.STX. If SUDAAN cannot find a level file, whether the name is supplied explicitly or by default, then SUDAAN will not be able to use formats named on FORMAT statements within your SUDAAN program. This option is available in all SUDAAN Procedures."

After the SUDAAN program has been run, the data file in SAS looks like this (Picture 1):

Picture 1:

The SAS System												10:00 Thursday, August 25, 2011				3
Obs	PROCNUM	TABLENO	AGEGR4	RACEETH	SEX	EDUCAT4	INCCAT3B	ADMIN2	FAIRPOOR							
21	1	1	*	*	0	*	*	2	2							
22	1	1	*	*	1	*	*	2	0							
23	1	1	*	*	1	*	*	2	1							
24	1	1	*	*	1	*	*	2	2							
25	1	1	*	*	2	*	*	2	0							
26	1	1	*	*	2	*	*	2	1							
27	1	1	*	*	2	*	*	2	2							
28	1	1	*	*	0	*	*	3	0							
29	1	1	*	*	0	*	*	3	1							
30	1	1	*	*	0	*	*	3	2							
Obs	KEPTUSAL	C06Q01	DIABETES	RFSMOK	C03Q04	CHECKUP	_C1	_C2	_C3	_C4	_C5					
21	*	*	*	*	*	-2	-2	-2	0	-2	-2					
22	*	*	*	*	*	-2	-2	-2	1	-2	-2					
23	*	*	*	*	*	-2	-2	-2	1	-2	-2					
24	*	*	*	*	*	-2	-2	-2	1	-2	-2					
25	*	*	*	*	*	-2	-2	-2	1	-2	-2					
26	*	*	*	*	*	-2	-2	-2	1	-2	-2					
27	*	*	*	*	*	-2	-2	-2	1	-2	-2					
28	*	*	*	*	*	-2	-2	-2	0	-2	-2					
29	*	*	*	*	*	-2	-2	-2	0	-2	-2					
30	*	*	*	*	*	-2	-2	-2	0	-2	-2					
Obs	_C6	_C7	_C8	_C9	_C10	_C11	_C12	_C13	NSUM	ROWPER	LOWROW	UPROW				
21	1	1	-2	-2	-2	-2	-2	-2	2082	86.632	84.5483	88.4724				
22	1	0	-2	-2	-2	-2	-2	-2	958	100.000	.	.				
23	1	1	-2	-2	-2	-2	-2	-2	192	12.116	9.5248	15.2935				
24	1	1	-2	-2	-2	-2	-2	-2	766	87.884	84.7065	90.4752				
25	1	0	-2	-2	-2	-2	-2	-2	1669	100.000	.	.				
26	1	1	-2	-2	-2	-2	-2	-2	353	14.558	12.0982	17.4183				
27	1	1	-2	-2	-2	-2	-2	-2	1316	85.442	82.5817	87.9018				
28	1	0	-2	-2	-2	-2	-2	-2	1768	100.000	.	.				
29	1	1	-2	-2	-2	-2	-2	-2	447	22.031	18.8035	25.6384				
30	1	1	-2	-2	-2	-2	-2	-2	1321	77.969	74.3616	81.1965				

The columns for the demographics are in the same order as they are in the **subgroup** for SUDAAN. Then SUDAAN runs the data as you have your **tables** set up in SUDAAN. So the values that are created for each column are created in the order that the variables are in the **tables** section. In this example the values that are created first are admin2 by sex by fairpoor.

Also notice the **_C** columns; these will be used for arrays. I do not know if SUDAAN or SAS creates these, but they can be very useful.

To help decipher the output here is an example of what is wanted:

Picture 2:

The SAS System										10:00 Thursday, August 25, 2011			3	
Obs	PROCNUM	TABLENO	AGEGR4	RACEETH	SEX	EDUCAT4	INCCAT3B	ADMIN2	FAIRPOOR					
21														
22					1			2	0					
23					1			2	1					
24					1			2	2					
25														
26														
27														
28														
29														
30														
Obs	KEPTUSAL	C06Q01	DIABETES	RFSMOK	C03Q04	CHECKUP	_C1	_C2	_C3	_C4	_C5			
21														
22														
23														
24														
25														
26														
27														
28														
29														
30														
Obs	_C6	_C7	_C8	_C9	_C10	_C11	_C12	_C13	NSUM	ROWPER	LOWROW	UPROW		
21														
22									958					
23										12.116	9.5248	15.2935		
24										87.884	84.7065	90.4752		
25														
26														
27														
28														
29														
30														

These are the variables that are needed to get the grand totals for the “male” demographic of the health risk variables. In this case FAIRPOOR was created first because it was first in the **table** listing on SUDAAN. It is also the first question that is asked on the survey. “Would you say that in general your health is...Excellent, Very good, Good, Fair, or Poor?”.

The SEX column will also be used for our grand totals or sample size, NSUM, for the area variable (ADMIN2), in this case Public Health Region 2/3. The reason why the SEX column is used for totals is that gender is logged by the interviewer and there are no missing or refused answers.

Lines 22 and 23 in Picture 2 must be on the same line on the final report. It must have titles and footnotes and reformatted to look like this (Picture 3):

Picture 3:

Prevalence of General Health Fair to Poor PHR 2/3 Adults Ages 18 Years and Over Texas BRFSS, 2010				
Groups	Demographics	Sample Size	Percent Yes (95% CI)	Percent No (95% CI)
Totals	Totals	2,627	13.4 (11.5-15.5)	86.6 (84.5-88.5)
Gender	Male	958	12.1 (9.5-15.3)	87.9 (84.7-90.5)
	Female	1,669	14.6 (12.1-17.4)	85.4 (82.6-87.9)
Race/Ethnicity	White	2,080	10.6 (8.9-12.6)	89.4 (87.4-91.1)
	Black	202	20.3 (14.1-28.3)	79.7 (71.7-85.9)
	Hispanic	237	21.2 (14.7-29.7)	78.8 (70.3-85.3)
	Other	79	10.9 (4.7-23.0)	89.1 (77.0-95.3)
Age Group	18-29 Years	126	7.8 (3.7-15.6)	92.2 (84.4-96.3)
	30-44 Years	444	7.6 (4.9-11.5)	92.4 (88.5-95.1)
	45-64 Years	1,052	17.8 (14.7-21.4)	82.2 (78.6-85.3)
	65+ Years	976	25.6 (21.9-29.7)	74.4 (70.3-78.1)
Education	< High School	233	30.3 (21.9-40.4)	69.7 (59.6-78.1)
	High School Grad	731	20.4 (15.7-26.1)	79.6 (73.9-84.3)
	Some College	703	14.6 (11.0-19.2)	85.4 (80.8-89.0)
	College Grad	956	5.7 (4.1-8.0)	94.3 (92.0-95.9)
Income	< \$25,000	628	33.8 (27.2-41.0)	66.2 (59.0-72.8)
	\$25,000-\$49,999	576	14.2 (10.2-19.4)	85.8 (80.6-89.8)
	\$50,000 +	1,050	5.2 (3.6-7.4)	94.8 (92.6-96.4)

Survey Question: "Would you say that in general, your health is Excellent, Very Good, Good, Fair, or Poor?"

Note: All reported rates are weighted for Texas Demographics and the probability of
Note: Prepared using complex sample design from SUDAAN
Prepared by:
Texas Behavioral Surveillance System, Center for Health Statistics
Texas Department of State Health Services
August 25, 2011

So how do you get here?

Well first, one must have the ExcelXP Tagset which can be downloaded from the Support-SAS website:

- <http://support.sas.com/rnd/base/topics/odsmarkup/>
- Scroll for the ExcelXP downloads.
- There are some very good papers that detail how to create Excel files so be sure to have a look at them!
- Once downloaded, open and copy/paste into SAS editor and run.
- One must also have the SAS/ACCESS module/add-on.

Getting into SAS

The opening statements are how to get SUDAAN output into SAS:

```
libname present 'H:\Brfss\avincent\MyPresentations';

libname show xport "H:\Brfss\avincent\MyPresentations\mastrlev.stx" ;
options nofmterr;

data masterF;
set show.mastrlev;
run;

proc format cntlin=show.mastrlev;
run;

libname showmf xport "H:\Brfss\avincent\MyPresentations\master.stx";
options nofmterr;

data present.mainfile;
```

```
set showmf.master;
run;
```

The top library opens a pathway where all the data sets and files will go to. The other two are export libraries that will be around long enough to give the data table. Then we can either clear the libraries to shut them down or just leave them be.

We then run some proc statements to see the contents of the data file.

```
proc print data=present.mainfile (obs=50) ;
run;
```

```
proc contents data=present.mainfile position;
run;
```

The proc print you have seen in the pictures above. This is the important part of the proc contents:

Picture 4:

		The SAS System		09:26 Friday, July 29, 2011	
The CONTENTS Procedure					
Variables in Creation Order					
#	Variable	Type	Len	Format	Label
1	PROCNUM	Num	8		Procedure Number
2	TABLNO	Num	8		Table Number
3	AGEGR4	Num	8	1.	Age Group
4	RACEETH	Num	8	1.	Race/Ethnicity
5	SEX	Num	8	11.	Sex of respondent
6	EDUCAT4	Num	8	1.	Education Categories
7	INCCAT3B	Num	8	1.	Income Categories
8	ADMIN2	Num	8	1.	Administrative Regions
9	FAIRPOOR	Num	8	1.	General Health Fair to Poor
10	KEPTUSAL	Num	8	1.	Days poor physical or mental health kept
11	C06Q01	Num	8	11.	Have you ever been told by a doctor that
12	DIABETES	Num	8	1.	Doctor Diagnosed Diabetes
13	RFSMOK	Num	8	1.	Current Smoker
14	C03Q04	Num	8	11.	About how long has it been since you las
15	CHCKUP1Y	Num	8	8.	had a routine check up within the past 1
16	_C1	Num	8		Codes for Categorical Variable AGEGR4
17	_C2	Num	8		Codes for Categorical Variable RACEETH
18	_C3	Num	8		Codes for Categorical Variable SEX
19	_C4	Num	8		Codes for Categorical Variable EDUCAT4
20	_C5	Num	8		Codes for Categorical Variable INCCAT3B
21	_C6	Num	8		Codes for Categorical Variable ADMIN2
22	_C7	Num	8		Codes for Categorical Variable FAIRPOOR
23	_C8	Num	8		Codes for Categorical Variable KEPTUSAL
24	_C9	Num	8		Codes for Categorical Variable C06Q01
25	_C10	Num	8		Codes for Categorical Variable DIABETES
26	_C11	Num	8		Codes for Categorical Variable RFSMOK
27	_C12	Num	8		Codes for Categorical Variable C03Q04
28	_C13	Num	8		Codes for Categorical Variable CHCKUP1YR
29	NSUM	Num	8		Sample Size
30	ROWPER	Num	8		Row Percent
31	LOWROW	Num	8		Lower 95% Limit ROWPER
32	UPROW	Num	8		Upper 95% Limit ROWPER

This shows you if either SUDAAN or SAS truncated your variable names and how many _C columns you have.

Next, we set up our Macros!

```
%macro fdate(fmt);
  %global fdate;
  data _null_;
    call symput("fdate",left(put("&sysdate"d,&fmt)));
  run;
%mend fdate;
%fdate(worddate.)

%let title1 = 'Prevalence of #byval(riskv)';
%let title2 = '#byval(area)';
%let title3 = 'Adults Ages 18 Years and Over';
%let title4 = 'Texas BRFS, 2010';
```

```

%let footnote1 = 'Survey Question: ';
%let footnote3 = 'Note: All reported rates are weighted for Texas
  Demographics and the probability of selection.';
%let footnote4 = 'Note: Prepared using complex sample design from
  SUDAAN';
%let footnote5 = 'Prepared by: ';
%let footnote6 = 'Texas Behavioral Surveillance System, Center for
  Health Statistics';
%let footnote7 = 'Texas Department of State Health Services';

```

The date macro will date the tables when they are run and put the date into a word format like "January 19, 2011". The other macros are for the titles and footnotes. These are easier to have at the beginning of the syntax so that they can be modified for each new run.

The "#byval()" macro (which can be found on title1 and title2 lines) is utilized in the proc report option "by". This allows for variable data to be added into the report titles or footnotes. This is especially nice when 130 tables need to be run!

The next part of the syntax is formatting the variables. Again, this is added near the beginning of the program where it can be easily changed based on the risk factors or area of interest. It is also a good idea to keep a list from SUDAAN of what risk factors were used and in what order they were in the **subgroup** heading.

```

proc format;
  value riskf
    1 = 'General Health Fair to Poor'
    2 = 'Kept from Usual Activities due to Poor
      Mental or Physical Health'
    3 = 'Diabetes - All Answers'
    4 = 'Diabetes - Yes/No'
    5 = 'Current Smoker'
    6 = 'Routine Checkup - All Answers'
    7 = 'Routine Checkup Within the Last Year'
    ;

  value areaf
    1 = 'PHR 1'
    2 = 'PHR 2/3'
    3 = 'PHR 4/5N'
    4 = 'PHR 6/5S'
    5 = 'PHR 7'
    6 = 'PHR 8'
    7 = 'PHR 9/10'
    8 = 'PHR 11'
    ;
run;

```

The Real Work

The program that was created is broken down by each section with an explanation of why it was done.

Creating Groups and Formatting

Section 1:

```
Data present.data1;
      set present.mainfile;

      rename admin2=area;
      if (agegr4=0 or raceeth=0 or educat4=0 or inccat3b=0 or admin2=0)
          then delete;
      length grp 3 grp2 $3;
      if sex=0 then grp=1;
      else if sex > 0 then grp=2;
      else if raceeth > 0 then grp=3;
      else if agegr4 > 0 then grp=4;
      else if educat4 > 0 then grp=5;
      else if inccat3b > 0 then grp=6;
```

In the above, the variable admin2 is renamed into something that will be used for the rest of the syntax. That way only one area variable needs to be modified when a new data request comes in (e.g., admin2 would be replaced with region). All demographic totals are removed except for “sex” because gender has no missing values. A new variable called “grp” was created so that the groups can be resorted in the order desired after the sorting and transposing of data results.

Section 2:

```
length alltitle $15 subtitle $16 ;
if grp=1 then alltitle="Totals";
else if grp=2 then alltitle="Gender";
else if grp=4 then alltitle="Age Group";
else if grp=3 then alltitle="Race/Ethnicity";
else if grp=5 then alltitle="Education";
else if grp=6 then alltitle="Income";

if agegr4=1 then subtitle='18-29 Years';
else if agegr4=2 then subtitle='30-44 Years';
else if agegr4=3 then subtitle='45-64 Years';
else if agegr4=4 then subtitle='65+ Years';
else if sex=1 then subtitle='Male';
else if sex=2 then subtitle='Female';
else if raceeth=1 then subtitle='White';
else if raceeth=2 then subtitle='Black';
else if raceeth=3 then subtitle='Hispanic';
else if raceeth=4 then subtitle='Other';
else if educat4=1 then subtitle='< High School';
else if educat4=2 then subtitle='High School Grad';
else if educat4=3 then subtitle='Some College';
else if educat4=4 then subtitle='College Grad';
else if inccat3b=1 then subtitle='< $25,000';
else if inccat3b=2 then subtitle='$25,000-$49,999';
```

```

else if inccat3b=3 then subtitle='$50,000 +';
else if alltitle='Totals' then subtitle='Totals';

```

In section 2, the variables and demographic groups are reformatted the way they should look like in the final report.

Section 3:

```

if subtitle = 'Totals' then add1='a';
if subtitle = 'Male' then add1='b';
if subtitle = 'Female' then add1='c';
if subtitle = 'White' then add1='d';
if subtitle = 'Black' then add1='e';
if subtitle = 'Hispanic' then add1='f';
if subtitle = 'Other' then add1='g';
if subtitle = '18-29 Years' then add1='h';
if subtitle = '30-44 Years' then add1='i';
if subtitle = '45-64 Years' then add1='j';
if subtitle = '65+ Years' then add1='k';
if subtitle = '< High School' then add1='l';
if subtitle = 'High School Grad' then add1='m';
if subtitle = 'Some College' then add1='n';
if subtitle = 'College Grad' then add1='o';
if subtitle = '< $25,000' then add1='p';
if subtitle = '$25,000-$49,999' then add1='q';
if subtitle = '$50,000 +' then add1='r';
grp2=trim(add1);

```

Section 3 shows a way of keeping the demographics in the order desired. These are alphabetic because the first variable “grp” is numeric and will make it easier to tell them apart.

Section 4:

```

LCL = Round(lowrow, .1);
if LCL < 0 then LCL = 0;
UCL = Round(uprow, .1);
if UCL < 0 then UCL = 0;

Length CI $15 Value $20;
CI = "(" || Compress(Put(LCL, 5.1)) || "-" || Compress(Put(UCL, 5.1)) || ")";
Value = Put(Round(rowper, .1), 5.1) || " " || CI;
Run;

```

In Section 4 of the program, new names are given to the upper and lower CIs and are rounded to one decimal place. In the second part, parentheses are added to the “ci” variable as well as a dash between them, and then the “ci” is placed behind the estimated percentage which is also rounded to one decimal place. In the end it will look like this: 12.1 (9.5-15.3).

Section 5:

```

data present.arrayrf ;
set present.data1;
array rfactor(*) _c7 -_c50 ;
do i=1 to dim(rfactor);
    if rfactor(i)>= 0 then rfactor(i)=(i);
    else rfactor(i)=.;
end;

```

```
rf = SUM(of _c7 - _c50) ;
end;
```

```
run;
```

An array is setup for section 5 of the programming. This is where the _C columns are utilized. The risk factor data results start in column _c7; this is why all of the demographics (in this case five) must be first in the **subgroup** section of SUDAAN and the area second (making six which is why we start in _c7).

This array is numbering the risk factors and putting them into one column. So, in the variable 'rf', _c7 will be 1, _c8 is 2, etc. down to however many risk factors there are. If the _c column is empty then the counting will stop.

This is what the data set looks like:

Picture 5:

	PROCNUM	TABLENO	AGEGR4	RACEETH	SEX	EDUCAT4	INCCAT3B	area	FAIRPOOR	KEPTUSAL	C06Q01	DIABETES	RFSMDK	C03Q04	CHKKUP1Y	_C1	_C2
1	1	1	*	*	0	*	*	1	0	*	-2	*	*	-2	-2	-2	-2
2	1	1	*	*	0	*	*	1	1	*	-2	*	*	-2	-2	-2	-2
3	1	1	*	*	0	*	*	1	2	*	-2	*	*	-2	-2	-2	-2
4	1	1	*	*	1	*	*	1	0	*	-2	*	*	-2	-2	-2	-2
5	1	1	*	*	1	*	*	1	1	*	-2	*	*	-2	-2	-2	-2
6	1	1	*	*	1	*	*	1	2	*	-2	*	*	-2	-2	-2	-2
7																	
	NSUM	ROWPER	LOWROW	UPROW	grp	grp2	alltile	subtile	add1	LCL	UCL	CI	Value				
1	850	100	.	.	1 a	Totals	Totals	a	0	0	0	(0.0-0.0)	100.0 (0.0-0.0)				
2	165	17.46790882	11.80652385	25.072246874	1 a	Totals	Totals	a	11.8	25.1	(11.8-25.1)	17.5 (11.8-25.1)					
3	685	82.53209118	74.927753126	88.19347615	1 a	Totals	Totals	a	74.9	88.2	(74.9-88.2)	82.5 (74.9-88.2)					
4	281	100	.	.	2 b	Gender	Male	b	0	0	(0.0-0.0)	100.0 (0.0-0.0)					
5	53	19.896419539	10.482924635	34.504767857	2 b	Gender	Male	b	10.5	34.5	(10.5-34.5)	19.9 (10.5-34.5)					
6	228	80.103580461	65.495232143	89.517075365	2 b	Gender	Male	b	65.5	89.5	(65.5-89.5)	80.1 (65.5-89.5)					
7																	
	_c41	_c42	_c43	_c44	_c45	_c46	_c47	_c48	_c49	_c50	i	rf					
1	45	1					
2	45	1					
3	45	1					
4	45	1					
5	45	1					
6	45	1					
7	45	1					

Section 6:

```
data present.data2;
  set present.arrayrf;
  if rowper = 100 then value = .;
  if rowper = 0 then nsum = .;
  if lowrow > 0 then nsum = .;
  if uprow > 0 then nsum = .;

drop _c1-_c50      procnum lcl ucl ci i;
run;
```

The grand total sample size of the demographics is needed. Section 6 of the program sets the "value" variable as missing if it is the total sample size. Only the NSUM value associated with the ROWPER = 100 is needed and the other NSUM values should be deleted.

Variables that are no longer needed are dropped as well.

Here is that data set cleaned up.

Picture 6:

	TABLNO	AGEGR4	RACEETH	SEX	EDUCAT4	INCCAT3B	area	FAIRPOOR	KEPTUSAL	C06Q01	DIABETES	RFSMOK	C03Q04	CHCKUP1Y	NSUM	ROWPER	LOWROW
1	1	*	*	0	*	*	1	0	*	-2	*	*	-2	-2	850	100	.
2	1	*	*	0	*	*	1	1	*	-2	*	*	-2	-2	17.46790882	11.80652385	*
3	1	*	*	0	*	*	1	2	*	-2	*	*	-2	-2	82.53209118	74.927753126	*
4	1	*	*	1	*	*	1	0	*	-2	*	*	-2	-2	281	100	.
5	1	*	*	1	*	*	1	1	*	-2	*	*	-2	-2	19.896419539	10.482324635	*
6	1	*	*	1	*	*	1	2	*	-2	*	*	-2	-2	80.103580461	65.495232143	*
7	1	*	*	2	*	*	1	0	*	-2	*	*	-2	-2	569	100	.

	NSUM	ROWPER	LOWROW	UPROW	grp	grp2	alltitle	subtile	add1	Value	rf
1	850	100	.	.	1	a	Totals	Totals	a	.	1
2	.	17.46790882	11.80652385	25.072246874	1	a	Totals	Totals	a	17.5 (11.8-25.1)	1
3	.	82.53209118	74.927753126	88.19347615	1	a	Totals	Totals	a	82.5 (74.9-88.2)	1
4	281	100	.	.	2	b	Gender	Male	b	.	1
5	.	19.896419539	10.482324635	34.504767857	2	b	Gender	Male	b	19.9 (10.5-34.5)	1
6	.	80.103580461	65.495232143	89.517075365	2	b	Gender	Male	b	80.1 (65.5-89.5)	1
7	569	100	.	.	2	c	Gender	Female	c	.	1
8	.	14.834844927	11.128872022	19.503934209	2	c	Gender	Female	c	14.8 (11.1-19.5)	1

Transposing

Two data sets will need to be created, one for total sample size and one for the estimates or values.

Section 7:

```
proc sort data=present.data2;
    by area rf grp grp2 alltitle subtile;
run;

proc transpose data=present.data2
    out=present.datatran let;
    var nsum value ;
    by area rf grp grp2 alltitle subtile ;
run;
```

Data must be sorted by the variables that will be used in the proc transpose step. In this case six variables need to be used.

This is what this data set looks like:

Picture 7:

	area	rf	grp	grp2	alltitle	subtile	_NAME_	_LABEL_	COL1	COL2	COL
1	1	1	1	a	Totals	Totals	NSUM	Sample Size	850	.	
2	1	1	1	a	Totals	Totals	Value	.	17.5 (11.8-25.1)	82.5	
3	1	1	2	b	Gender	Male	NSUM	Sample Size	281	.	
4	1	1	2	b	Gender	Male	Value	.	19.9 (10.5-34.5)	80.1	
5	1	1	2	c	Gender	Female	NSUM	Sample Size	569	.	
6	1	1	2	c	Gender	Female	Value	.	14.8 (11.1-19.5)	85.2	
7	1	1	3	d	Race/Ethnicity	White	NSUM	Sample Size	647	.	
8	1	1	3	d	Race/Ethnicity	White	Value	.	16.4 (12.0-22.2)	83.6	

	subtile	_NAME_	_LABEL_	COL1	COL2	COL3	COL4
1	Totals	NSUM	Sample Size	850	.	.	.
2	Totals	Value	.	17.5 (11.8-25.1)	82.5 (74.9-88.2)	.	.
3	Male	NSUM	Sample Size	281	.	.	.
4	Male	Value	.	19.9 (10.5-34.5)	80.1 (65.5-89.5)	.	.
5	Female	NSUM	Sample Size	569	.	.	.
6	Female	Value	.	14.8 (11.1-19.5)	85.2 (80.5-88.9)	.	.
7	White	NSUM	Sample Size	647	.	.	.
8	White	Value	.	16.4 (12.0-22.2)	83.6 (77.8-88.0)	.	.

	NAME	_LABEL_	COL1	COL2	COL3	COL4	COL5
72	Value	.	.	95.6 (90.5-98.0)	4.4 (2.0-9.5)	.	.
73	NSUM	Sample Size	853
74	Value	.	.	6.8 (4.6-9.9)	1.4 (0.4-4.8)	91.5 (87.7-94.1)	0.3 (0.1-1.1)
75	NSUM	Sample Size	281
76	Value	.	.	7.6 (4.2-13.2)	0.0 (0.0-0.0)	92.3 (86.7-95.7)	0.1 (0.0-0.4)
77	NSUM	Sample Size	572
78	Value	.	.	6.0 (3.7-9.6)	2.9 (0.9-9.6)	90.6 (85.0-94.2)	0.5 (0.1-2.3)
79	NSUM	Sample Size	648
80	Value	.	.	5.9 (3.9-8.9)	0.4 (0.1-2.8)	93.2 (90.1-95.4)	0.5 (0.1-1.9)

All the values are on the same row. From lines 74 on, the risk factor variables with more than two answers are all on the same row and the order that the values are in, is the order they are supposed to be in.

The file will now be split into two separate files.

Sample Size Data File

The next data step has to do with the Texas BRFSS only reporting items that have a sample size of 50 or more. This is due to statistical reliability and confidentiality purposes when the respondent numbers are low.

```
data present.subdata1(rename=(coll=ssize));
  length ssize 8 ;
  set present.datatran;
  ssize=put(coll,20.);
  if (ssize=. or ssize=0) then delete;
  if (ssize < 50) then coll='< 50';
  keep area ssize coll grp2 rf alltitle subtile ;
run;
```

A trick about renaming variables. RENAME applies to all *output* data sets only. Use the old variable name in the programming statements for the current data steps. In other words, the rename only applies to the final data set, not the *input* data set that the variables come from. So don't apply the new name in the data step.

The 'put' statement transforms a character variable, (coll), to a numeric variable, (ssize), which will be needed for statistical computations.

The 'value' columns are removed and a data set with only the total sample size is established.

Here's that data set.

Picture 8:

	ssize	Administrative Regions	rf	grp2	alltitle	subtitle	ssizec
1	850	1	1	a	Totals	Totals	850
2	281	1	1	b	Gender	Male	281
3	569	1	1	c	Gender	Female	569
4	647	1	1	d	Race/Ethnicity	White	647
5	35	1	1	e	Race/Ethnicity	Black	< 50
6	133	1	1	f	Race/Ethnicity	Hispanic	133
7	31	1	1	g	Race/Ethnicity	Other	< 50
8	67	1	1	h	Age Group	18-29 Years	67
9	159	1	1	i	Age Group	30-44 Years	159
10	302	1	1	j	Age Group	45-64 Years	302

We see the variable renamed from “col1” to “ssizec” and now has ‘< 50’ value from “ssize” being under 50 count.

Values (Estimates) Data File

Section 9:

The second data file will have the sample sizes removed and only the values left.

```
data present.subdata2 ;
  set present.datatran;
  if (_NAME_='NSUM') then delete;
keep area grp2 rf alltitle subtitle col2-col150;
run;
```

Picture 9:

	Administrative Regions	rf	grp2	alltitle	subtitle	COL2	COL3	COL4	COL5
1	1	1	a	Totals	Totals	17.5 (11.8-25.1)	82.5 (74.9-88.2)		
2	1	1	b	Gender	Male	19.9 (10.5-34.5)	80.1 (65.5-89.5)		
3	1	1	c	Gender	Female	14.8 (11.1-19.5)	85.2 (80.5-88.9)		
4	1	1	d	Race/Ethnicity	White	16.4 (12.0-22.2)	83.6 (77.8-88.0)		
5	1	1	e	Race/Ethnicity	Black	49.7 (21.8-77.7)	50.3 (22.3-78.2)		
6	1	1	f	Race/Ethnicity	Hispanic	17.8 (6.2-41.4)	82.2 (58.6-93.8)		
7	1	1	g	Race/Ethnicity	Other	7.5 (1.3-33.1)	92.5 (66.9-98.7)		
8	1	1	h	Age Group	18-29 Years	14.6 (3.1-47.4)	85.4 (52.6-96.9)		
9	1	1	i	Age Group	30-44 Years	12.2 (6.1-22.7)	87.8 (77.3-93.9)		
10	1	1	j	Age Group	45-64 Years	21.8 (14.7-31.2)	78.2 (68.8-85.3)		

Section 10:

The two data files now have to be merged back together again so the data are all on one line.

```
proc sort data=present.subdata1;
  by area rf grp2;
run;

proc sort data=present.subdata2;
  by area rf grp2;
run;

data present.data3;
  merge present.subdata1 present.subdata2 ;
  by area rf grp2 ;
run;

proc sort data=present.data3;
  by area rf grp2;
```

```
run;
```

First the subdata sets need to be sorted by the variables that they will be matched on; in this case those variables are “area,” the risk factor variable “rf,” and the demographic variable “grp2.” The sample size data set needs to come first followed by the values data set because we want the sample size on the left hand side of the final table. The data are sorted one last time, but this step could probably be skipped.

Picture 10:

	ssize	area	if	grp2	alltitle	subtitle	ssizec	COL2	COL3	COL4
1	850	1	1	a	Totals	Totals	850	17.5(11.8-25.1)	82.5(74.9-88.2)	
2	281	1	1	b	Gender	Male	281	19.9(10.5-34.5)	80.1(65.5-89.5)	
3	569	1	1	c	Gender	Female	569	14.8(11.1-19.5)	85.2(80.5-88.9)	
4	647	1	1	d	Race/Ethnicity	White	647	16.4(12.0-22.2)	83.6(77.8-88.0)	
5	35	1	1	e	Race/Ethnicity	Black	< 50	49.7(21.8-77.7)	50.3(22.3-78.2)	
6	133	1	1	f	Race/Ethnicity	Hispanic	133	17.8(6.2-41.4)	82.2(58.6-93.8)	
7	31	1	1	g	Race/Ethnicity	Other	< 50	7.5(1.3-33.1)	92.5(66.9-98.7)	
8	67	1	1	h	Age Group	18-29 Years	67	14.6(3.1-47.4)	85.4(52.6-96.9)	
9	159	1	1	i	Age Group	30-44 Years	159	12.2(6.1-22.7)	87.8(77.3-93.3)	
10	302	1	1	j	Age Group	45-64 Years	302	21.8(14.7-31.2)	78.2(68.8-85.3)	
11	317	1	1	k	Age Group	65+ Years	317	24.3(17.2-33.1)	75.7(66.9-82.8)	

Every item is on the same row, and it all lines up. Since the data results are not reported for those areas with < 50 respondents, these “values” need to be suppressed and reformatted to ‘(-)’.

Section 11:

```
data present.final;
  set present.data3;
  length riskv 8;
  riskv=rf;
  format riskv riskf. area areaf.;
  array minusvar(*) col2-col9;
  do i=1 to dim(minusvar);
  if ssize < 50 then minusvar(i)='(-)';
  end;
run;
```

Another array is used to incorporate the data that are to be suppressed. For this data step the names of all the “_c” columns need to be known and written in. A look at the actual data table will be in order.

Section 12:

```
option NOBYLINE;

Proc report data = present.final split = "\" nowd spacing=3
  style(column) = [font_size = 8pt just = c]
  headline headskip ;
  by area riskv;

Columns area riskv alltitle subtitle grp2 ssizec col2-col9 ;

define area / order order =data 'Area' Noprint;
break after area / ol skip ;
define riskv / order order =data 'RF';
break after riskv / ol skip;
define alltitle / order order=data 'Groups';
break after alltitle/ skip;
define subtitle / 'Demographics';
```

```

Define grp2 / order order = data Noprint;
break after grp2 / skip ;
define ssizec / 'Sample' 'Size' ;
define col2-col9 / right 'Percent' '(95% CI)';
title1 &Title1 j=c;
title2 &title2 j=c;
footnotel &footnote5 j=1;;
run;

```

The 'option nobyline' is a global option to state that the 'by' group should not be repeated in the first line after the titles. For us, this is repetitive and annoying, but it might be useful to you.

The proc report is pretty straight forward. We have skips in our headlines. We have a line break after the area, the risk factor, and the demographic groups. This report is made just to make sure that all the data works and is in the correct place.

This is the output:

Picture 11:

Prevalence of General Health Fair to Poor				66
PHR 2/3				10:00 Thursday, August 25, 2011
Demographics	Sample Size	Percent (95% CI)	Percent (95% CI)	
Totals	2627	13.4 (11.5-15.5)	86.6 (84.5-88.5)	
Male	958	12.1 (9.5-15.3)	87.9 (84.7-90.5)	
Female	1669	14.6 (12.1-17.4)	85.4 (82.6-87.9)	
White	2080	10.6 (8.9-12.6)	89.4 (87.4-91.1)	
Black	202	20.3 (14.1-28.3)	79.7 (71.7-85.9)	
Hispanic	237	21.2 (14.7-29.7)	78.8 (70.3-85.3)	
Other	79	10.9 (4.7-23.0)	89.1 (77.0-95.3)	
18-29 Years	126	7.8 (3.7-15.6)	92.2 (84.4-96.3)	
30-44 Years	444	7.6 (4.9-11.5)	92.4 (88.5-95.1)	
45-64 Years	1052	17.8 (14.7-21.4)	82.2 (78.6-85.3)	
65+ Years	976	25.6 (21.9-29.7)	74.4 (70.3-78.1)	
< High School	233	30.3 (21.9-40.4)	69.7 (59.6-78.1)	
High School Grad	731	20.4 (15.7-26.1)	79.6 (73.9-84.3)	
Some College	703	14.6 (11.0-19.2)	85.4 (80.8-89.0)	
College Grad	955	5.7 (4.1-8.0)	94.3 (92.0-95.9)	

Prepared by:

Notice that the titles have their risk factor and area listed correctly. This is just a quick report; the formatting comes later when the data results and reports are sent to Excel.

Section 13:

```

ods listing close;
ods tagsets.excelxp file =
    "H:\Brfss\avincent\MyPresentations\showoff.xml"
style=statistical
options(sheet_label = 'Risk Factor '
        embedded_titles = 'yes'
        embedded_footnotes='yes'
        sheet_interval='table'
        autofit_height='yes'
        contents='yes');

```

ODS is versatile and easy to use. The tagset excelxp needs to be access to give the output a place to go. There are many styles to choose from, I use 'statistical'. The options for ODS are

many as well. For example, if the titles or footnotes are not embedded, they only appear when something is printed out. Clients typically need to have them so that they know what they are looking at. The 'sheet_interval' option tells SAS when to split the output on the EXCEL spreadsheet. Having the 'contents' option will make a table of contents spreadsheet with a link to every table, which is very handy when there are 130 tables!

A small warning concerning options, when the "option nobyline" was used on the global statement, it interfered with the "sheet_interval" option. When the "option byline" was used instead for this report the "sheet_interval=bygroup" option worked just fine. So that is something to watch out for.

```
Proc report data = present.final split = "\" nowd spacing=3
  style(column) = [font_size = 10pt]
  headline headskip ;
  format riskv riskf. ;
  by area riskv notsorted;
```

In the above, the proc report is much the same as the previous one. The risk factor variable is formatted and the 'by' option has two variables. Note that the procedure has the further option of 'notsorted', which means that the data set does not need to be sorted by those variables and it tells SAS to read the data as is.

```
Columns area riskv alltitle subtitle grp2 ssizec col2-col9 ;

define area / order order =data noprint;
break after area / ol skip;
define riskv / order order =data noprint ;
break after riskv / ol skip;
define alltitle / order order=data 'Groups' left width=11;
break after alltitle/ skip;
define subtitle / 'Demographics' left width=11;
Define grp2 / order order = data Noprint;
break after grp2 / skip ;
define ssizec / style = {tagattr='format:##,###'} 'Sample\Size' right;
define col2-col9 / center 'Percent\ (95% CI)' width=11;

title1 &Title1 j=c;
title2 &Title2 j=c;
title3 &Title3 j=c;
title4 &Title4 j=c;

footnote1 underline=1 &footnote1 j=1;
footnote3 &footnote3 j=1;
footnote4 &footnote4 j=1;
footnote5 &footnote5 j=1;
footnote6 &footnote6 j=1;
footnote7 &footnote7 j=1;
footnote8 "&fdate";
run;

ods tagsets.excelxp close;
ods listing;
```

Most of the rest of this is pretty straight forward. The columns are listed. The area and risk factor variables are 'noprint' as they will be listed in the titles. The ssizec has a style attribute that will get EXCEL to change its appearance by putting in the comma in the usual numeric places.

With the footnotes footnote2 is skipped so that there is a line space. The end of the program calls up the date macro that was created at the beginning of the file.

The near final product in EXCEL is below:

Picture 12:

Prevalence of General Health Fair to Poor PHR 2/3 Adults Ages 18 Years and Over Texas BRFSS, 2010										
Groups	Demographics	Sample Size	Percent (95% CI)							
Totals	Totals	2,627	13.4 (11.5-15.5)	86.6 (84.5-88.5)						
Gender	Male	958	12.1 (9.5-15.3)	87.9 (84.7-90.5)						
	Female	1,669	14.6 (12.1-17.4)	85.4 (82.6-87.9)						
Race/Ethnicity	White	2,080	10.6 (8.9-12.6)	89.4 (87.4-91.1)						
	Black	202	20.3 (14.1-28.3)	79.7 (71.7-85.9)						
	Hispanic	237	21.2 (14.7-29.7)	78.8 (70.3-85.3)						
	Other	79	10.9 (4.7-23.0)	89.1 (77.0-95.3)						
Age Group	18-29 Years	126	7.8 (3.7-15.6)	92.2 (84.4-96.3)						
	30-44 Years	444	7.6 (4.9-11.5)	92.4 (88.5-95.1)						
	45-64 Years	1,052	17.8 (14.7-21.4)	82.2 (78.6-85.3)						
	65+ Years	976	25.6 (21.9-29.7)	74.4 (70.3-78.1)						
Education	< High School	233	30.3 (21.9-40.4)	69.7 (59.6-78.1)						
	High School Grad	731	20.4 (15.7-26.1)	79.6 (73.9-84.3)						
	Some College	703	14.6 (11.0-19.2)	85.4 (80.8-89.0)						
	College Grad	955	5.7 (4.1-8.0)	94.3 (92.0-95.9)						
Income	< \$25,000	628	33.8 (27.2-41.0)	66.2 (59.0-72.8)						
	\$25,000-\$49,999	576	14.2 (10.2-19.4)	85.8 (80.6-89.6)						
	\$50,000 +	1,090	5.2 (3.6-7.4)	94.8 (92.6-96.4)						

Survey Question:
Note: All reported rates are weighted for Texas Demographics and the probability of selection.
Note: Prepared using complex sample design from SUDAAN
Prepared by:
Texas Behavioral Surveillance System, Center for Health Statistics
Texas Department of State Health Services
August 25, 2011

The table is not perfect, but the cleanup is easy. In addition, all we have to add is the 'Percent' answers ("Fair or Poor" or "Excellent, Very Good, Good" in this example) and the 'Survey Question:' in the footnotes.

Here's what it looks like with the less than 50 values:

Picture 13:

Prevalence of General Health Fair to Poor PHR 1 Adults Ages 18 Years and Over Texas BRFSS, 2010										
Groups	Demographics	Sample Size	Percent (95% CI)							
Totals	Totals	1,945	19.9 (16.0-24.6)	80.1 (75.4-84.0)						
Gender	Male	704	17.3 (11.3-25.7)	82.7 (74.3-88.7)						
	Female	1,241	22.7 (18.5-27.6)	77.3 (72.4-81.5)						
Race/Ethnicity	White	1,520	17.1 (12.4-23.2)	82.9 (76.8-87.6)						
	Black	74	34.7 (17.2-57.6)	65.3 (42.4-82.8)						
	Hispanic	286	25.1 (16.9-35.5)	74.9 (64.5-83.1)						
	Other	< 50	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
Age Group	18-29 Years	122	13.0 (3.5-38.2)	87.0 (61.8-96.5)						
	30-44 Years	338	14.2 (8.7-22.2)	85.8 (77.8-91.3)						
	45-64 Years	752	26.5 (21.2-32.4)	73.5 (67.6-78.8)						
	65+ Years	725	30.5 (25.4-36.2)	69.5 (63.8-74.6)						
Education	< High School	216	35.7 (25.4-47.5)	64.3 (52.5-74.6)						
	High School Grad	562	24.5 (16.3-35.2)	75.5 (64.8-83.7)						
	Some College	542	11.8 (8.4-16.3)	88.2 (83.7-91.6)						
	College Grad	622	14.0 (8.6-22.2)	86.0 (77.8-91.4)						
Income	< \$25,000	539	27.3 (19.7-36.5)	72.7 (63.5-80.3)						
	\$25,000-\$49,999	444	22.2 (12.2-37.1)	77.8 (62.9-87.8)						
	\$50,000 +	695	10.1 (6.0-16.4)	89.9 (83.6-94.0)						

Survey Question:
Note: All reported rates are weighted for Texas Demographics and the probability of selection.
Note: Prepared using complex sample design from SUDAAN
Prepared by:
Texas Behavioral Surveillance System, Center for Health Statistics
Texas Department of State Health Services
August 25, 2011

And here is an example of having some long answers:

Picture 14:

Prevalence of Routine Checkup - All Answers										
PHR 2/3										
Adults Ages 18 Years and Over										
Texas BRFSS, 2010										
Groups	Demographics	Sample Size	Percent (95% CI)							
Totals	Totals	2,612	62.6 (58.9-66.1)	12.7 (10.6-15.1)	11.6 (9.2-14.5)	11.6 (9.3-14.5)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	1.5 (0.7-3.0)
Gender	Male	955	57.7 (51.7-63.4)	13.2 (10.0-17.3)	12.9 (9.1-17.9)	13.9 (9.9-19.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	2.3 (1.0-5.4)
	Female	1,657	67.3 (63.1-71.2)	12.1 (9.6-15.2)	10.4 (7.8-13.7)	9.5 (7.1-12.6)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.7 (0.3-1.6)
Race/Ethnicity	White	2,068	65.4 (61.3-69.3)	12.6 (10.1-15.6)	9.9 (7.5-12.9)	11.9 (9.4-15.1)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.2 (0.1-0.5)
	Black	201	71.0 (59.7-80.1)	13.9 (8.1-22.9)	8.3 (3.7-17.5)	6.5 (2.2-17.9)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.2 (0.0-1.7)
	Hispanic	236	50.6 (40.6-60.6)	13.2 (8.2-20.6)	14.6 (8.5-24.1)	13.6 (7.3-24.1)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	7.9 (3.6-16.6)
	Other	78	44.4 (27.8-62.4)	11.6 (5.2-23.8)	30.2 (14.6-52.2)	11.7 (2.4-41.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	2.2 (0.3-12.4)
Age Group	18-29 Years	124	39.9 (26.4-52.6)	19.6 (11.4-31.5)	22.1 (12.9-35.2)	16.2 (8.3-29.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	2.3 (0.4-12.7)
	30-44 Years	443	55.3 (48.9-61.5)	13.7 (10.1-18.1)	14.2 (10.1-19.6)	14.6 (10.2-20.4)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	2.3 (0.9-5.4)
	45-64 Years	1,047	71.6 (67.1-75.7)	10.7 (8.3-13.6)	7.8 (5.5-10.8)	9.3 (7.1-12.1)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.6 (0.1-3.9)
	65+ Years	969	83.9 (80.5-86.9)	7.7 (5.6-10.5)	3.2 (2.0-5.1)	4.5 (3.0-6.5)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.7 (0.3-1.8)
Education	< High School	233	49.5 (38.0-61.0)	14.1 (7.6-24.8)	13.3 (6.4-25.5)	12.6 (6.6-22.7)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	10.6 (4.4-23.1)
	High School Grad	723	60.4 (52.4-67.8)	11.2 (7.8-15.9)	9.9 (6.0-15.8)	16.8 (10.8-25.2)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	1.7 (0.3-8.3)
	Some College	699	60.8 (53.4-67.7)	13.8 (9.3-19.9)	15.9 (10.4-23.5)	9.3 (6.0-14.1)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.3 (0.1-1.4)
Income	College Grad	953	67.3 (61.9-72.2)	12.5 (9.5-16.1)	9.7 (6.9-13.6)	10.2 (7.0-14.5)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.4 (0.1-1.5)
	< \$25,000	625	47.7 (39.9-55.6)	16.9 (12.1-23.0)	15.5 (10.2-23.3)	14.9 (9.4-23.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	5.0 (1.9-12.2)
	\$25,000-\$49,999	572	65.6 (57.4-73.0)	14.1 (8.4-22.6)	8.5 (5.1-13.8)	9.2 (5.7-14.5)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	2.7 (0.9-7.8)
	\$50,000 +	1,051	68.9 (63.9-73.5)	11.5 (8.8-14.8)	10.0 (7.1-14.1)	9.5 (6.7-13.3)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.1)

Survey Question:
 Note: All reported rates are weighted for Texas Demographics and the probability of selection.
 Note: Prepared using complex sample design from SUDAAN
 Prepared by:
 Texas Behavioral Surveillance System, Center for Health Statistics
 Texas Department of State Health Services
 August 25, 2011

Here is something that SUDAAN does that is not the fault of SAS. Notice the three columns of all zeroes (middle right in Picture 14). This question asked respondents when their last routine medical checkup was. There are only five answers that they could give, “Within the past year”, “within the past two years”, “within the past 5 years”, “5 or more years ago”, or “never”. You might be inclined to panic and say “oh no! What went wrong!?!”. Well actually nothing went wrong. The answers to the question are numbered, 1-4 for the years, 8 = never. The other answers for this question, 7=“Don’t Know/Not Sure” and 9=“Refused”, are set to missing in SPSS and are not used for calculation.

SUDAAN must have continuous numbering when it runs. So the gaps between answers 4 and 8 must be filled by something. So when the data results are cleaned, these columns of all zeroes need to be deleted. Sometimes a column does indeed have no answers, so caution must be utilized. Another way in preventing the “zero columns” is recoding the original question at the very beginning (in this case, in SPSS before even running the results in SUDAAN). An example would be to recode the answer 8 to a 5.

Charts

After all this work was done and I proudly showed it to my supervisor, she said, “That’s nice! But what about some charts?”

So back to SAS Enhance Editor! (and the Help button!)

The charts that are made in SAS are for internal use only; they do not go out to clients. We want to have a look at where the story is before we make charts for our clients in Excel.

Section 14:

```
proc sort data=present.charttest1;
    by rf subtitle;
run;

data present.charttest2;
    set present.charttest1;
    by rf subtitle;
    if first.subtitle then chartvar=0;
    chartvar+1;
run;
```

The data are sorted by risk factor and the demographic titles. In the data step the X-plot line needed to be numbered in such a way that all the demographics for an area had the same number. So that, per the risk factor variable, every area and their demographics would have the same numbers throughout the data set. So if the risk factor had two answers, area one's chartvar would have values 1 and 2 while area two's chartvar would have values 3 and 4 and so forth. If the risk factor had four answers, area one's chartvar would have values 1 through 4 and area two would have values 5 through 8.

Picture 15a chartvar numbered one to sixteen for a given demographic:

	TABLENO	AGEGR4	RACEETH	SEX	EDUCAT4	INCCAT3B	area	FAIRPOOR	KEPTUSAL	CO6001	DIABETES	RFSMOK	CO3Q04	CHECKUP	NSUM	ROWPER	LOWROW	UPROW
1	29	*	*	*	*	2	1	1	*	*	*	*	*	*	-2	22.209948125	12.163104283	37.054674271
2	29	*	*	*	*	2	1	2	*	*	*	*	*	*	-2	77.790051875	62.945325729	87.836895717
3	29	*	*	*	*	2	2	1	*	*	*	*	*	*	-2	14.186090538	10.178696016	19.429322616
4	29	*	*	*	*	2	2	2	*	*	*	*	*	*	-2	85.813909462	80.570077384	89.821303984
5	29	*	*	*	*	2	3	1	*	*	*	*	*	*	-2	14.563896002	10.456634767	19.925249145
6	29	*	*	*	*	2	3	2	*	*	*	*	*	*	-2	85.436143988	80.074750855	89.543305233
7	29	*	*	*	*	2	4	1	*	*	*	*	*	*	-2	15.967873882	11.784685543	21.277992274
8	29	*	*	*	*	2	4	2	*	*	*	*	*	*	-2	84.032126118	78.722007726	88.215434457
9	29	*	*	*	*	2	5	1	*	*	*	*	*	*	-2	19.347549601	10.926060428	31.93303781
10	29	*	*	*	*	2	5	2	*	*	*	*	*	*	-2	80.652450399	68.06696219	89.073939572
11	29	*	*	*	*	2	6	1	*	*	*	*	*	*	-2	15.462774293	11.213677413	20.94217265
12	29	*	*	*	*	2	6	2	*	*	*	*	*	*	-2	84.537225707	79.05782735	88.786322587
13	29	*	*	*	*	2	7	1	*	*	*	*	*	*	-2	21.541431863	15.802150072	28.655768635
14	29	*	*	*	*	2	7	2	*	*	*	*	*	*	-2	78.458568137	71.344231365	84.197849928
15	29	*	*	*	*	2	8	1	*	*	*	*	*	*	-2	21.233941549	15.847133926	27.845943795
16	29	*	*	*	*	2	8	2	*	*	*	*	*	*	-2	78.766058451	72.154056205	84.152866074
17	29	*	*	*	*	3	1	1	*	*	*	*	*	*	-2	10.101374209	6.0417248758	16.412357903
18	29	*	*	*	*	3	1	2	*	*	*	*	*	*	-2	89.898625791	83.587642097	93.958275124

	UPROW	grp	grp2	altitle	subtile	add1	Value	if	LSE	USE	chartvar
1	37.054674271	6 q	Income	\$25,000-\$49,999	q	22.2 (12.2-37.1)	General Health Fair to Poor	10	14.8	14.8	1
2	87.836895717	6 q	Income	\$25,000-\$49,999	q	77.8 (62.9-87.8)	General Health Fair to Poor	14.8	10	10	2
3	19.429322616	6 q	Income	\$25,000-\$49,999	q	14.2 (10.2-19.4)	General Health Fair to Poor	4	5.2	5.2	3
4	89.821303984	6 q	Income	\$25,000-\$49,999	q	85.8 (80.6-89.8)	General Health Fair to Poor	5.2	4	4	4
5	19.925249145	6 q	Income	\$25,000-\$49,999	q	14.6 (10.5-19.9)	General Health Fair to Poor	4.1	5.4	5.4	5
6	89.543305233	6 q	Income	\$25,000-\$49,999	q	85.4 (80.1-89.5)	General Health Fair to Poor	5.4	4	4	6
7	21.277992274	6 q	Income	\$25,000-\$49,999	q	16.0 (11.8-21.3)	General Health Fair to Poor	4.2	5.3	5.3	7
8	88.215434457	6 q	Income	\$25,000-\$49,999	q	84.0 (78.7-88.2)	General Health Fair to Poor	5.3	4.2	4.2	8
9	31.93303781	6 q	Income	\$25,000-\$49,999	q	19.3 (10.9-31.9)	General Health Fair to Poor	8.4	12.6	12.6	9
10	89.073939572	6 q	Income	\$25,000-\$49,999	q	80.7 (68.1-89.1)	General Health Fair to Poor	12.6	8.4	8.4	10
11	20.94217265	6 q	Income	\$25,000-\$49,999	q	15.5 (11.2-20.9)	General Health Fair to Poor	4.2	5.5	5.5	11
12	88.786322587	6 q	Income	\$25,000-\$49,999	q	84.5 (79.1-88.8)	General Health Fair to Poor	5.5	4.2	4.2	12
13	28.655768635	6 q	Income	\$25,000-\$49,999	q	21.5 (15.8-28.7)	General Health Fair to Poor	5.7	7	7	13
14	84.197849928	6 q	Income	\$25,000-\$49,999	q	78.5 (71.3-84.2)	General Health Fair to Poor	7.1	5.7	5.7	14
15	27.845943795	6 q	Income	\$25,000-\$49,999	q	21.2 (15.8-27.8)	General Health Fair to Poor	5.4	6.6	6.6	15
16	84.152866074	6 q	Income	\$25,000-\$49,999	q	78.8 (72.2-84.2)	General Health Fair to Poor	6.6	5.4	5.4	16
17	16.412357903	6 r	Income	\$50,000 +	r	10.1 (6.0-16.4)	General Health Fair to Poor	4.1	6.3	6.3	1
18	93.958275124	6 r	Income	\$50,000 +	r	89.9 (83.6-94.0)	General Health Fair to Poor	6.3	4	4	2

Picture 15b data sorted back to original setting:

	TABLENO	AGEGR4	RACEETH	SEX	EDUCAT4	INCCAT3B	area	Fair	POOR	EPTUSAL	C06Q01	DIABETES	RFSMOK	C03Q04	CHECKUP	NSUM	RDWPER	LOWROW	UPROW
1	1	*	*	0	*	*	1	1	*	*	*	*	*	*	*	-2	19.931888656	15.982098188	24.572335393
2	1	*	*	0	*	*	1	2	*	*	*	*	*	*	*	-2	80.068111344	75.427664607	84.017901812
3	1	*	*	1	*	*	1	1	*	*	*	*	*	*	*	-2	17.336742092	11.307010496	25.651978
4	1	*	*	1	*	*	1	2	*	*	*	*	*	*	*	-2	82.663257908	74.348022	88.692989504
5	1	*	*	2	*	*	1	1	*	*	*	*	*	*	*	-2	22.703645587	18.478680029	27.567937888
6	1	*	*	2	*	*	1	2	*	*	*	*	*	*	*	-2	77.296354413	72.432062112	81.521319971
7	8	*	1	*	*	*	1	1	*	*	*	*	*	*	*	-2	17.148039227	12.431517669	23.18032381
8	8	*	1	*	*	*	1	2	*	*	*	*	*	*	*	-2	82.851960773	76.81967619	87.568482331
9	8	*	2	*	*	*	1	1	*	*	*	*	*	*	*	-2	34.713712643	17.231709441	57.590843588
10	8	*	2	*	*	*	1	2	*	*	*	*	*	*	*	-2	65.286287357	42.409156412	82.768290559
11	8	*	3	*	*	*	1	1	*	*	*	*	*	*	*	-2	25.087950949	16.925846187	35.503897884
12	8	*	3	*	*	*	1	2	*	*	*	*	*	*	*	-2	74.912049051	64.496102116	83.074153813
13	8	*	4	*	*	*	1	1	*	*	*	*	*	*	*	-2	7.8150663004	3.1899267276	17.905936247
14	8	*	4	*	*	*	1	2	*	*	*	*	*	*	*	-2	92.1849337	82.094063753	96.810073272
15	15	1	*	*	*	*	1	1	*	*	*	*	*	*	*	-2	13.03853688	3.5100603704	38.194379708
16	15	1	*	*	*	*	1	2	*	*	*	*	*	*	*	-2	86.96146312	61.805620232	96.48993963
17	15	2	*	*	*	*	1	1	*	*	*	*	*	*	*	-2	14.178135117	8.723687886	22.213003913
18	15	2	*	*	*	*	1	2	*	*	*	*	*	*	*	-2	85.821864883	77.786996087	91.276312114

	UPROW	grp	grp2	alltitle	subtile	add1	Value	rf	LSE	USE	chartvar
1	24.572335393	1 a	Totals	Totals	a	19.9 (16.0-24.6)	General Health Fair to Poor	3.9	4.6	1	
2	84.017901812	1 a	Totals	Totals	a	80.1 (75.4-84.0)	General Health Fair to Poor	4.6	3.9	2	
3	25.651978	2 b	Gender	Male	b	17.3 (11.3-25.7)	General Health Fair to Poor	6	8.3	1	
4	88.692989504	2 b	Gender	Male	b	82.7 (74.3-88.7)	General Health Fair to Poor	8.3	6	2	
5	27.567937888	2 c	Gender	Female	c	22.7 (18.5-27.6)	General Health Fair to Poor	4.2	4.9	1	
6	81.521319971	2 c	Gender	Female	c	77.3 (72.4-81.5)	General Health Fair to Poor	4.9	4.2	2	
7	23.18032381	3 d	Race/Ethnicity	White	d	17.1 (12.4-23.2)	General Health Fair to Poor	4.7	6	1	
8	87.568482331	3 d	Race/Ethnicity	White	d	82.9 (76.8-87.6)	General Health Fair to Poor	6	4.7	2	
9	57.590843588	3 e	Race/Ethnicity	Black	e	34.7 (17.2-57.6)	General Health Fair to Poor	17.5	22.9	1	
10	82.768290559	3 e	Race/Ethnicity	Black	e	65.3 (42.4-82.8)	General Health Fair to Poor	22.9	17.5	2	
11	35.503897884	3 f	Race/Ethnicity	Hispanic	f	25.1 (16.9-35.5)	General Health Fair to Poor	8.2	10.4	1	
12	83.074153813	3 f	Race/Ethnicity	Hispanic	f	74.9 (64.5-83.1)	General Health Fair to Poor	10.4	8.2	2	
13	17.905936247	3 g	Race/Ethnicity	Other	g	7.8 (3.2-17.9)	General Health Fair to Poor	4.6	10.1	1	
14	96.810073272	3 g	Race/Ethnicity	Other	g	92.2 (82.1-96.8)	General Health Fair to Poor	10.1	4.6	2	
15	38.194379708	4 h	Age Group	18-29 Years	h	13.0 (3.5-38.2)	General Health Fair to Poor	9.5	25.2	1	
16	96.48993963	4 h	Age Group	18-29 Years	h	87.0 (61.8-96.5)	General Health Fair to Poor	25.2	9.5	2	
17	22.213003913	4 i	Age Group	30-44 Years	i	14.2 (8.7-22.2)	General Health Fair to Poor	5.5	8	1	
18	91.276312114	4 i	Age Group	30-44 Years	i	85.8 (77.8-91.3)	General Health Fair to Poor	8	5.5	2	

A chart template was found that is a bar chart that has upper and lower CI error markings. It took some time to find this and while not perfect, it works for our internal use. In SAS 9.3, much of this work will be probably be unnecessary.

Section 15:

proc template;

```

define statgraph brfsschart;
  begingraph;
  entrytitle "Prevalence of Risk Factor by Demographics,
    2010";
  entryfootnote HALIGN=left "Error Bars show 95% Confidence
    Intervals";
  layout gridded / border=false;
  layout datalattice columnvar=subtitle /
  headerlabeldisplay=value cellwidthmin=20
  columnheaders=bottom border=false columndatarange=union
  columnaxisopts=(display=(line tickvalues))
  rowaxisopts=(offsetmin=0 linearopts=(viewmax=100
    tickvaluepriority=true)
    label='Percentages' griddisplay=on);
  layout prototype / walldisplay=(fill);
  barchart x=chartvar y=rowper / group=area name='a'
  outlineattrs=(color=black);
  scatterplot x=chartvar y=rowper / yerrorlower=lowrow
  yerrorupper=uprow MARKERATTRS=(size=0)
  errorbarattrs=(thickness=1) datatransparency=0.6;
  endlayout;
endlayout;
entry ' ';
discretelegend 'a' / title='Area: ' border=true;
endlayout;

```

```

    endgraph;
end;
run;

```

Most of this should be easy to figure out. Substitute the columns needed and then call the template. We first must resort the data set to their proper order after making the chartvar.

Section 16:

```

proc sort data=present.charttest2;
    by area rf grp2;
run;

ods listing close;
ods html image_dpi=100 file='presentcharts.xls'
path='H:\Brfss\avincent\MyPresentations\';
ods graphics / reset noborder width=1000px height=500px
    imagename='BRFSSChart' imagefmt=gif noscale;

proc sgrender data=present.charttest2 template=brfsschart;
    format area areaf. rf riskf.;
    where (area=2 and rf=4 and chartvar=3);
    by area rf;
run;

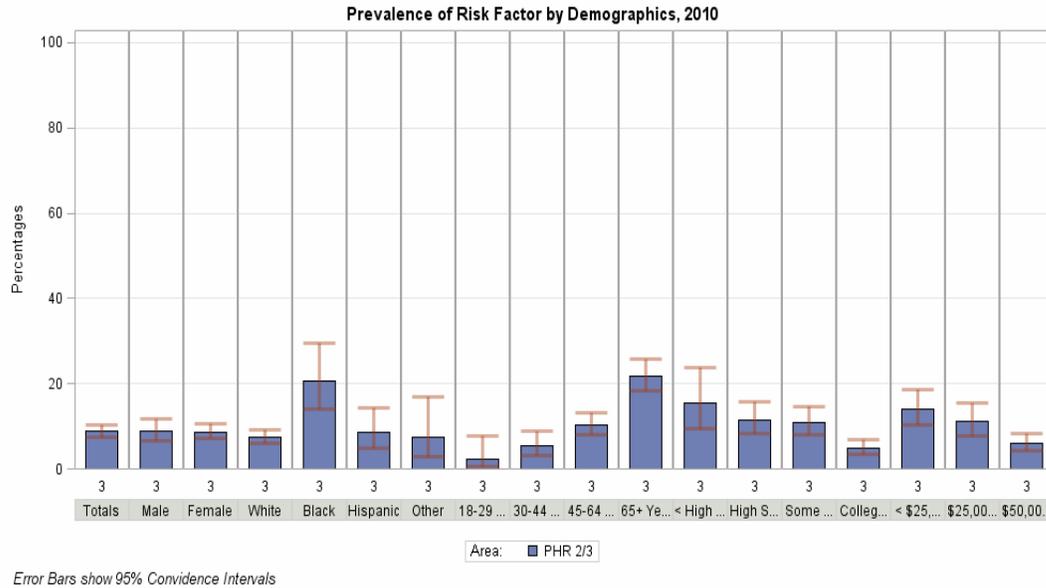
ods html close;
ods listing;

```

We use a 'where' statement in the proc sgrender (which is a proc step to identify the data set that contains the plot variables and the statgraph template) so that we don't get all the answers all at once.

And here are the results:

Picture 16:



The title didn't put what the risk factor variable was, but that's OK, it is internal use. The above table is for those who responded that they had diabetes.

A rule of thumb about CI's (technically statistical tests should be done), if the top of one bar overlaps the bottom of another bar then the two estimates are not statistically different. Compare the "male" and "female" error bars. They both overlap. Diabetes does not seem to be different between men and women.

If they don't overlap then they may be statistically significant by rule of thumb, but a good statistician would run a t-test or chi-test to be sure. In this case compare "black" with "white" or ages "18-29" to "65+" years. Both of these examples have bars that do not overlap

Also the shorter the distance between upper and lower error bars the better, since there is less variability in the sample. Typically, larger samples have less variance. The longer bars, like the "< High School", show less confidence in your estimate.

Remember that SUDAAN does asymmetric confidence intervals, but SAS has symmetrical confidence intervals. So whenever anyone states that their percentages or estimates are "plus or minus" a certain number, do the math, see if they overlap!

Conclusion

This has been a little primer to show that anyone with just the basic knowledge of SAS can create some really nice reports just by looking at SAS help and SAS support online. So don't be afraid to dive right in and play around till you get what you want. "Cut, paste, and tweak!" is my motto!

Contact Information

Comments or Questions:

Anna Vincent
Center for Health Statistics, MC-1898
Texas Department of State Health Services
PO Box 149317
Austin, TX 78714-9347
512-776-2520 work
512-458-7429 fax

Emails:

Anna.vincent@dshs.state.tx.us

brfss@dshs.state.tx.us

Website:

<http://www.dshs.state.tx.us/chs/brfss/default.shtm>