

# **SUDAAN Example Manual Addendum**

## **Release 9.0.3**

**September 28, 2007**

An electronic copy of this addendum to the SUDAAN Example Manual is available at our website at [www.rti.org/SUDAAN](http://www.rti.org/SUDAAN).

Copyright 2007 by RTI International  
P.O. Box 12194  
Research Triangle Park, NC 27709

All rights reserved. No part of this publication may be reproduced or transmitted by any means without permission from the publisher.

# Table of Contents

	<u>Page</u>
1. Overview.....	3
2. Examples Illustrating SUDAAN's New Virtual Memory Management .....	4
3. Examples Illustrating the SUDWORK Parameter .....	5
4. Examples that Produce Output in RTF Format.....	5
5. Examples Illustrating the Revised CLASS Statement .....	14
6. Examples Illustrating the Revised PERCENTILE and HISTPCT Statements in DESCRIPT .....	15
7. Examples Illustrating the Revised SORTBY Statement in RECORDS .....	19
8. Examples Illustrating the New CROSSTAB Procedure Enhancements.....	22

# 1. Overview

This document provides examples associated with some of the enhancements that will be introduced in *SUDAAN Release 10.0*, scheduled for shipment in the latter part of 2008. In order to provide annual license holders and new SUDAAN users with the most up-to-date SUDAAN software, the enhancements illustrated in this addendum have been introduced in beta-form with the current maintenance release, *SUDAAN Release 9.0.3*. Any questions or comments about these new enhancements are most welcomed and should be directed via email to [SUDAAN@rti.org](mailto:SUDAAN@rti.org). *SUDAAN Release 9.0.3* also incorporates several bug fixes, as documented on the SUDAAN website (see the Technical Assistance page on [www.rti.org/SUDAAN](http://www.rti.org/SUDAAN).)

In summary, the new enhancements that are being introduced in beta-form in *Release 9.0.3* include:

- One of the most exciting enhancements introduced in *Release 9.0.3* is SUDAAN's use of a new, internal **virtual memory manager**. This is available in most procedures and will allow users to process very large datasets in SUDAAN. Examples addressing the new virtual memory manager are presented in **Section 2** of this Addendum.
- In conjunction with the new virtual memory manager, a new SUDWORK parameter option has been introduced for all PROC statements. This option will allow users to redirect their SUDAAN work space to an alternate location (*e.g.*, another hard drive, an external drive, or another location on a network drive). Examples illustrating this new parameter option are presented in **Section 3** of this Addendum.
- A second equally exciting enhancement introduced in *Release 9.0.3* is that SUDAAN can now provide output in both **ASCII** and **RTF** format. Examples that utilize the new RTF format are presented in **Section 4**.
- A few new enhancements have been introduced with the CLASS statement (examples are presented in **Section 5** of this Addendum).
- Enhancements to the PERCENTILES statement and a change in the default number of bins used to compute percentiles have been introduced in the DESCRIPT procedure. Examples are presented in **Section 6** of this Addendum.
- The SORTBY statement has been revised in the RECORDS procedure (examples are presented in **Section 7** of this Addendum)
- Four major analytical enhancements have been incorporated in the **CROSSTAB** procedure:
  1. Additional hypothesis tests for single 2-way tables and stratified 2-way tables,
  2. A goodness-of-fit test against known proportions,
  3. Additional test statistics for all hypothesis tests (analogous to regression procedures), and
  4. A new method for computing confidence intervals for extreme proportions

Examples utilizing the new CROSSTAB enhancements are presented in **Section 8** of this Addendum.

The syntax changes associated with these enhancements have been documented in the ***SUDAAN Language Manual Addendum, Release 9.0.3*** which is posted on the Technical Assistance page on the SUDAAN website, located at [www.rti.org/SUDAAN](http://www.rti.org/SUDAAN).

## 2. Examples Illustrating SUDAAN's New Virtual Memory Management

### Example 2.1

The following example illustrates the use of the USEVMEM option. Since USEVMEM is set equal to zero, SUDAAN will perform all necessary computations using internal RAM only. In other words, there will be no work space files created by SUDAAN.

```
PROC LOGISTIC Data="c:\Sudt7a" FILETYPE=SUDAAN DESIGN=JACKKNIFE USEVMEM=0;
NEST STRATUM NFSU;
WEIGHT finalwt;
CLASS gender paygroup family educat race_eth;
MODEL esmoke30=gender paygroup family educat race_eth agec;
OUTPUT beta sebeta / filename="c:\T7jNDOS3.xpt" filetype=sasxport replace;
```

### Example 2.2

The following example also illustrates the use of the USEVMEM option. Since USEVMEM is set equal to one, SUDAAN will store intermediate results in external work space files. Setting USEVMEM=1 will minimize the amount of internal RAM that SUDAAN needs for the procedure.

```
PROC LOGISTIC Data="c:\Sudt7a" FILETYPE=SUDAAN DESIGN=JACKKNIFE USEVMEM=1;
NEST STRATUM NFSU;
WEIGHT finalwt;
CLASS gender paygroup family educat race_eth;
MODEL esmoke30=gender paygroup family educat race_eth agec;
OUTPUT beta sebeta / filename="c:\T7jNDOS3.xpt" filetype=sasxport replace;
```

### Example 2.3

The following example also illustrates the use of the USEVMEM option. Since USEVMEM is set equal to 150, SUDAAN will store intermediate results in external work space files if holding them in core would require more than 150 MB of memory, otherwise, SUDAAN will use RAM to hold intermediate results. Setting USEVMEM=150 will ensure that SUDAAN will use no more than 150 MB RAM for the intermediate results.

```
PROC LOGISTIC Data="c:\Sudt7a" FILETYPE=SUDAAN DESIGN=JACKKNIFE USEVMEM=150;
NEST STRATUM NFSU;
WEIGHT finalwt;
CLASS gender paygroup family educat race_eth;
MODEL esmoke30=gender paygroup family educat race_eth agec;
OUTPUT beta sebeta / filename="c:\T7jNDOS3.xpt" filetype=sasxport replace;
```

### 3. Examples Illustrating the SUDWORK Parameter

#### Example 3.1

The following example illustrates the use of the SUDWORK parameter. For this procedure call, SUDAAN is forced to store work files on an external location (since USEVMEM=1) and the SUDWORK parameter tells SUDAAN to create these work files in a subdirectory under c:\temp\SUDAAN.

```
PROC LOGISTIC DATA="c:\Sudt7a" FILETYPE=SUDAAN
              DESIGN=JACKKNIFE USEVMEM=1 SUDWORK="c:\temp\sudaan";
NEST STRATUM NFSU;
WEIGHT finalwt;
CLASS gender race_eth;
MODEL esmoke30=gender race_eth agec;
PRINT beta sebeta;
```

### 4. Examples that Produce Output in RTF Format

#### Example 4.1

The following example illustrates table output from SUDAAN in both text format and RTF format. The first PRINT statement causes SUDAAN to create a text file and place it in the file c:\Example\_4-1.TXT. The second PRINT statement causes SUDAAN to create the same table in RTF format and place it in c:\Example\_4-1.RTF. Since the font name and size was not specified, the RTF file will be created with a Times New Roman font and a point size of 12.

Whenever SUDAAN creates an output file with FILETYPE=RTF, a page break will appear at the end of the printout. Consequently, when the table is printed, a blank page will appear at the end of the file. This page break has been included to facilitate inclusion of the RTF file inside an existing report. This extra page break can be manually omitted if desired.

#### Example 4.1: SUDAAN Code

```
PROC CROSSTAB DATA="c:\DODHB05B.XPT" FILETYPE=SASXPORT
              LEVFILE="c:\DODHB05F.XPT";
WEIGHT FINALWT;
NEST STRATUM NFSU;
SUBPOPN (ESMOKE30 > 0);
CLASS RACE_ETH ESMOKE30;
TABLES ESMOKE30*RACE_ETH;
PRINT NSUM ROWPER COLPER / Filetype=TEXT
                              Filename="c:\Example_4-1.TXT" replace;
PRINT NSUM ROWPER COLPER / Filetype=RTF
                              Filename="c:\Example_4-1.RTF" replace;
TITLE "Any Cigarette Smoking in Past 30 Days by Race/Ethnicity";
```

**Example 4.1: Page 1 from c:\Example\_4-1.TXT**

Date: 09-17-2007  
 Time: 11:47:00

SUDAAN

Page: 1  
 Table: 1

Variance Estimation Method: Taylor Series (WR)  
 For Subpopulation: ( ESMOKE30 > 0 )  
 Any Cigarette Smoking in Past 30 Days by Race/Ethnicity  
 by: Any Smoking in Past 30 Days, Race/Ethnicity: Based on IR\_Race, IR\_His.

Any Smoking in Past 30 Days		Race/Ethnicity: Based on IR_Race, IR_His		
		Total	Nonhispanic White	Nonhispanic Black
Total	Sample Size	15947	9750	2595
	Row Percent	100.00	64.47	17.60
	Col Percent	100.00	100.00	100.00
Smoked in Past 30 Days	Sample Size	4259	2820	428
	Row Percent	100.00	72.20	10.79
	Col Percent	32.17	36.03	19.73
Not Smoked in Past 30 Days	Sample Size	11688	6930	2167
	Row Percent	100.00	60.81	20.83
	Col Percent	67.83	63.97	80.27

**Example 4.1: Page 2 from c:\Example\_4-1.TXT**

Date: 09-17-2007  
 Time: 11:47:00

SUDAAN

Page: 2  
 Table: 1

Variance Estimation Method: Taylor Series (WR)  
 For Subpopulation: ( ESMOKE30 > 0 )  
 Any Cigarette Smoking in Past 30 Days by Race/Ethnicity  
 by: Any Smoking in Past 30 Days, Race/Ethnicity: Based on IR\_Race, IR\_His.

Any Smoking in Past 30 Days		Race/Ethnicity: Based on IR_Race, IR_His	
		Hispanic	Other
Total	Sample Size	1978	1624
	Row Percent	8.81	9.12
	Col Percent	100.00	100.00
Smoked in Past 30 Days	Sample Size	521	490
	Row Percent	7.58	9.43
	Col Percent	27.70	33.25
Not Smoked in Past 30 Days	Sample Size	1457	1134
	Row Percent	9.39	8.98
	Col Percent	72.30	66.75

**Example 4.1: Page 1 from c:\Example\_4-1.RTF**

Date: 09-17-2007  
Time: 11:47:00

SUDAAN

Page: 1  
Table: 1

Variance Estimation Method: Taylor Series (WR)  
For Subpopulation: ( ESMOKE30 > 0 )  
Any Cigarette Smoking in Past 30 Days by Race/Ethnicity  
by: Any Smoking in Past 30 Days, Race/Ethnicity: Based on IR\_Race, IR\_His.

		<b>Race/Ethnicity: Based on IR_Race, IR_His</b>		
		<b>Total</b>	<b>Nonhispanic White</b>	<b>Nonhispanic Black</b>
<b>Any Smoking in Past 30 Days</b>	Sample Size	15947	9750	2595
	Row Percent	100.00	64.47	17.60
	Col Percent	100.00	100.00	100.00
Smoked in Past 30 Days	Sample Size	4259	2820	428
	Row Percent	100.00	72.20	10.79
	Col Percent	32.17	36.03	19.73
Not Smoked in Past 30 Days	Sample Size	11688	6930	2167
	Row Percent	100.00	60.81	20.83
	Col Percent	67.83	63.97	80.27

**Example 4.1: Page 2 from c:\Example\_4-1.RTF**

Date: 09-17-2007  
Time: 11:47:00

SUDAAN

Page: 2  
Table:1

Variance Estimation Method: Taylor Series (WR)  
For Subpopulation: ( ESMOKE30 > 0 )  
Any Cigarette Smoking in Past 30 Days by Race/Ethnicity  
by: Any Smoking in Past 30 Days, Race/Ethnicity: Based on IR\_Race, IR\_His.

Any Smoking in Past 30 Days		Race/Ethnicity: Based on IR_Race, IR_His	
		Hispanic	Other
Total	Sample Size	1978	1624
	Row Percent	8.81	9.12
	Col Percent	100.00	100.00
Smoked in Past 30 Days	Sample Size	521	490
	Row Percent	7.58	9.43
	Col Percent	27.70	33.25
Not Smoked in Past 30 Days	Sample Size	1457	1134
	Row Percent	9.39	8.98
	Col Percent	72.30	66.75

**Example 4.2**

The following example is similar to **Example 4.1**. In this example, the RTF file is printed with half inch margins on all sides, the font is Arial and the font size is 8. Also note that the SUBPOPN restricts this output to just the first two levels of race/ethnicity (*i.e.*, RACE\_ETH<3).

**Example 4.2: SUDAAN Code**

```
PROC CROSSTAB DATA="C:\DODHB05B.XPT" FILETYPE=SASXPORT
    LEVFILE="C:\DODHB05F.XPT";
WEIGHT FINALWT;
NEST STRATUM NFSU;
SUBPOPN (ESMOKE30 > 0 and RACE_ETH < 3);
CLASS RACE_ETH ESMOKE30;
TABLES ESMOKE30*RACE_ETH;
PRINT NSUM ROWPER COLPER / Filetype=TEXT
    Filename="C:\Example_4-2.TXT" replace;
PRINT NSUM ROWPER COLPER / Filetype=RTF FONTNAME="ARIAL" FONTSIZE=8 LEFTINCH=.5
    RIGHTINCH=.5 TOPINCH=.5 BOTTOMINCH=.5
    Filename="C:\Example_4-2.RTF" replace;
TITLE "Any Cigarette Smoking in Past 30 Days by Race/Ethnicity";
```

**Example 4.2: Page 1 from Example 4-2.TXT**

Date: 09-17-2007  
 Time: 12:00:12

SUDAAN

Page: 1  
 Table: 1

Variance Estimation Method: Taylor Series (WR)  
 For Subpopulation: ( ESMOKE30 > 0 AND RACE\_ETH < 3 )  
 Any Cigarette Smoking in Past 30 Days by Race/Ethnicity  
 by: Any Smoking in Past 30 Days, Race/Ethnicity: Based on IR\_Race, IR\_His.

Any Smoking in Past 30 Days		Race/Ethnicity: Based on IR_Race, IR_His		
		Total	Nonhispanic White	Nonhispanic Black
Total	Sample Size	12345	9750	2595
	Row Percent	100.00	78.56	21.44
	Col Percent	100.00	100.00	100.00
Smoked in Past 30 Days	Sample Size	3248	2820	428
	Row Percent	100.00	87.00	13.00
	Col Percent	32.53	36.03	19.73
Not Smoked in Past 30 Days	Sample Size	9097	6930	2167
	Row Percent	100.00	74.49	25.51
	Col Percent	67.47	63.97	80.27

**Example 4.2: Page 1 from Example 4-2.RTF**

Date: 09-17-2007		SUDAAN			Page: 1
Time: 12:00:12					Table:1
Variance Estimation Method: Taylor Series (WR)					
For Subpopulation: ( ESMOKE30 > 0 AND RACE_ETH < 3 )					
Any Cigarette Smoking in Past 30 Days by Race/Ethnicity					
by: Any Smoking in Past 30 Days, Race/Ethnicity: Based on IR_Race, IR_His.					
Any Smoking in Past 30 Days		Race/Ethnicity: Based on IR_Race, IR_His			
		Total	Nonhispanic White	Nonhispanic Black	
Total	Sample Size	12345	9750	2595	
	Row Percent	100.00	78.56	21.44	
	Col Percent	100.00	100.00	100.00	
Smoked in Past 30 Days	Sample Size	3248	2820	428	
	Row Percent	100.00	87.00	13.00	
	Col Percent	32.53	36.03	19.73	
Not Smoked in Past 30 Days	Sample Size	9097	6930	2167	
	Row Percent	100.00	74.49	25.51	
	Col Percent	67.47	63.97	80.27	

**Example 4.3**

The following example illustrates RTF style output with STYLE=NCHS. This example also illustrates the SMCONF option on the PROC statement. This option is new in *SUDAAN 9.0.3*.

**Example 4.3: SUDAAN Code**

```

PROC CROSSTAB DATA="c:\DODHB05B.XPT" FILETYPE=SASXPORT
      LEVFILE="c:\DODHB05F.XPT"
      SMCONF=25;
WEIGHT FINALWT;
NEST STRATUM NFSU;
SUBPOPN (ESMOKE30 > 0);
CLASS RACE_ETH ESMOKE30;
TABLES ESMOKE30*RACE_ETH;
PRINT ROWPER SEROW LOWROW UPROW ROWSPCI / Filetype=TEXT STYLE=NCHS
      Filename="c:\Example_4-3.TXT" replace;
PRINT ROWPER SEROW LOWROW UPROW ROWSPCI / Filetype=RTF STYLE=NCHS FONTSIZE=10
      Filename="c:\Example_4-3.RTF" replace;
TITLE "Any Cigarette Smoking in Past 30 Days by Race/Ethnicity";

```

**Example 4.3: Page 1 from Example 4-2.TXT**

Date: 09-17-2007  
 Time: 11:44:23

SUDAAN

Page: 1  
 Table: 1

Variance Estimation Method: Taylor Series (WR)  
 For Subpopulation: ( ESMOKE30 > 0 )  
 Any Cigarette Smoking in Past 30 Days by Race/Ethnicity  
 by: Any Smoking in Past 30 Days, Race/Ethnicity: Based on IR\_Race, IR\_His.

```

-----
Any Smoking in Past                               S
30 Days                                           P
  Race/Ethnicity:                                C
    Based on                                     I
    IR_Race, IR_His                             R
-----
          Row      SE Row      Lower      Upper      o
          Percent  Percent  Limit    Limit    w
          ROWPER   ROWPER
-----
Total
  Total          100.00      0.00      .          .
  Nonhispanic White  64.47      1.25      61.92      66.94
  Nonhispanic Black  17.60      1.01      15.60      19.74  ^
  Hispanic          8.81      0.46      7.90      9.78  ^
  Other            9.12      0.60      7.95      10.41  ^
Smoked in Past 30
  Days
  Total          100.00      0.00      .          .
  Nonhispanic White  72.20      1.19      69.73      74.54
  Nonhispanic Black  10.79      0.75      9.32      12.40  ^
  Hispanic          7.58      0.39      6.81      8.42  ^
  Other            9.43      0.73      8.00      11.01  ^
Not Smoked in Past
  30 Days
  Total          100.00      0.00      .          .
  Nonhispanic White  60.81      1.46      57.83      63.70
  Nonhispanic Black  20.83      1.25      18.35      23.48  ^
  Hispanic          9.39      0.60      8.22      10.66  ^
  Other            8.98      0.67      7.67      10.43  ^
-----
    
```

**Example 4.3: Page 1 from Example 4-2.RTF**

Date: 09-17-2007		SUDAAN			Page: 1	
Time: 12:27:07					Table: 1	
Variance Estimation Method: Taylor Series (WR)						
For Subpopulation: ( ESMOKE30 > 0 )						
Any Cigarette Smoking in Past 30 Days by Race/Ethnicity						
by: Any Smoking in Past 30 Days, Race/Ethnicity: Based on IR_Race, IR_His.						
Any Smoking in Past 30 Days	Race/Ethnicity: Based on IR_Race, IR_His	Row Percent	SE Row Percent	Lower 95% Limit ROWPER	Upper 95% Limit ROWPER	S P C I R o w
Total	Total	100.00	0.00	.	.	
	Nonhispanic White	64.47	1.25	61.92	66.94	
	Nonhispanic Black	17.60	1.01	15.60	19.74	^
	Hispanic	8.81	0.46	7.90	9.78	^
	Other	9.12	0.60	7.95	10.41	^
Smoked in Past 30 Days	Total	100.00	0.00	.	.	
	Nonhispanic White	72.20	1.19	69.73	74.54	
	Nonhispanic Black	10.79	0.75	9.32	12.40	^
	Hispanic	7.58	0.39	6.81	8.42	^
	Other	9.43	0.73	8.00	11.01	^
Not Smoked in Past 30 Days	Total	100.00	0.00	.	.	
	Nonhispanic White	60.81	1.46	57.83	63.70	
	Nonhispanic Black	20.83	1.25	18.35	23.48	^
	Hispanic	9.39	0.60	8.22	10.66	^
	Other	8.98	0.67	7.67	10.43	^

Notice in the above printout:

1. The Race/Ethnicity column label is wrapping before the “y” in “Ethnicity”. This can happen on occasion. In other words, depending on the font and point size chosen, the RTF file may exhibit an undesirable wrap in a column label. These can be easily fixed manually in any type of word processing software package that reads table-formatted RTF files. This can also be fixed by changing the point size in SUDAAN.
2. Also notice the “SPCI Row” column label has spaces between all the letters. To maintain consistency with SUDAAN 9.0.1 and 9.0.2, the RTF-style output in SUDAAN 9.0.3 has been designed to mimic what one would get in the SUDAAN text-style output. Notice that “SPCI Row” appears as a column of length 1 in the text-style output. Within any cell, SUDAAN has been programmed to take each line in the text-style output, add a character to the end and omit

the line-feed/carriage-return from the end of the line. In doing this, the column label for SPCI Row becomes “S P C I R o w”.

## 5. Examples Illustrating the Revised CLASS Statement

### **Example 5.1: Revised CLASS Statement**

This example illustrates the modified functionality of the CLASS statement in SUDAAN. The data is from the 1988-1994 NHANES III (see *Section 1.8.5* of the *SUDAAN 9 Example Manual* for more detail on NHANES III). In this example, the mean body mass index (BMI) is estimated for each selected level of self-reported health status.

### **Example 5.1: SUDAAN Code Using NHANES III data**

```
proc descript data=in.hanes3s5 design=wr;
nest sdpstra6 sdpps6;
weight wtpfqx6;
subpopn HAB1 = 1 or HAB1 = 5; /* Health status is excellent or poor */
class HAB1;
table HAB1; /* means by health status */
print mean semean lowmean upmean / style=nchs;
RFORMAT HAB1 HS.;
```

```
Number of observations read      : 20050      Weighted count :187647206
Observations in subpopulation   : 4111       Weighted count: 45012255
Denominator degrees of freedom  : 49
```

The frequency distribution for each CLASS variable is presented first with the formatted values shown. These formatted values were defined by the HS format on the RFORMAT statement.

### Example 5.1: Frequency Distribution for CLASS Variables

Frequencies and Values for CLASS Variables  
by: Is health in general excellent, ..., poor.

```
-----  
Is health in  
  general  
  excellent-  
  , ..., poor      Frequency      Value  
-----  
Ordered  
  Position:  
  1                3070      Excellent  
Ordered  
  Position:  
  2                1041        Poor  
-----
```

### Example 5.1: Mean BMI by Health Status

Variance Estimation Method: Taylor Series (WR)  
For Subpopulation: HAB1 = 1 OR HAB1 = 5  
by: Variable, Is health in general excellent, ..., poor.

```
-----  
Variable  
  Is health in  
  general  
  excellent, ..., -  
  poor  
-----  
                          Mean      SE Mean      Lower 95%      Upper 95%  
                          Mean      Mean      Limit          Limit  
                          Cntrst   Cntrst   Cntrst         Cntrst  
-----  
Body Mass Index  
  Total                25.28      0.14          25.00          25.56  
  Excellent            24.95      0.14          24.67          25.23  
  Poor                 27.59      0.34          26.92          28.27  
-----
```

For each level of health status (where status = excellent or status = poor) the mean BMI is shown. Notice that the other 3 levels of health status (very good, good, and fair) are omitted from the output table.

## 6. Examples Illustrating the Revised PERCENTILE and HISTPCT Statements in DESCRIPT

The following examples illustrate the changes made to the PERCENTILE statement in PROC DESCRIPT beginning with SUDAAN *Release 9.0.3*. The changes include increasing the default number of bins used to

calculate the percentiles from 20 to 100, changing the method used to calculate the lower and upper bound of the confidence intervals and adding an option that adds a specified amount of random noise to each observation.

The main reason for these changes was to correct an error in SUDAAN that caused the point estimates to lie outside of the confidence interval. This situation tended to occur when there are a large number of observations in the data set with identical values (tied values). If there are no ties in the data then results from *Release 9.0.3* and previous releases of SUDAAN will be identical provided the number of bins used is identical. The number of bins used can be controlled by using the HISTPCT / NPCT=*integer* option in DESCRIPT.

The change in the method used to produce the confidence interval cannot be controlled by the user. In situations where there are numerous ties in the data it will not be possible to reproduce the results from SUDAAN *Release 9.0.1* or *Release 9.0.2* with the newer SUDAAN *Release 9.0.3*. In this situation, the results from SUDAAN *Release 9.0.3* are preferred to the results from SUDAAN *Release 9.0.1/Release 9.0.2*.

The first example below shows an output that was generated with SUDAAN *Release 9.0.1* to show the results using the old confidence interval method. **Example 6.1** uses a data set that has numerous observations tied at 40. The remaining observations are continuous between 0 and 20 and between 60 and 80. No observations, except the ones tied at 40, fall between 20 and 60. While data of this type may be unusual, it illustrates the problems that have been corrected in SUDAAN *Release 9.0.3*.

#### **Example 6.1: SUDAAN Code**

```
PROC DESCRIPT DATA=test DESIGN=WR;
NEST stratum psu;
WEIGHT analwgt1;
VAR SALES;
PERCENTILES 5 10 25 50 75 90 95;
SETENV COLWIDTH=6 DECWIDTH=2;
PRINT QTILE SEQTILE LOWQTILE UPQTILE/ SEQTILEFMT=F6.3;
```

**Example 6.1: Results using SUDAAN Release 9.0.1**

Date: 09-19-2007	Research Triangle Institute	Page : 1		
Time: 18:43:31	The DESCRIPT Procedure	Table : 1		
Variance Estimation Method: Taylor Series (WR)				
by: Variable, One, Percentiles.				
for: Variable = SALES.				
-----				
One				
Percentiles	Quant-	SE	Lower	Upper
	ile	Quant-	95%	95%
		ile	Limit	Limit
-----				
1				
5.00	12.94	1.838	9.52	17.16
10.00	19.97	5.086	17.86	39.01
25.00	23.79	0.018	39.14	39.21
50.00	30.17	0.019	39.46	39.54
75.00	36.54	0.021	39.78	39.87
90.00	62.84	6.475	39.97	66.90
95.00	71.00	1.742	66.75	73.99
-----				

Note from the printout above that the percentiles for 25%, 50% and 75% fall outside of the confidence intervals. Recall that there are no observations between 20 and 40. However, several of the point estimates are in this range.

The same code was run using SUDAAN Release 9.0.3 with the output written to an RTF file.

**Example 6.1: SUDAAN Release 9.0.3 with RTF Output**

Date: 09-19-2007	SUDAAN	Page: 1			
Time: 18:37:58		Table: 1			
Variance Estimation Method: Taylor Series (WR)					
by: Variable, One, Percentiles.					
for: Variable = SALES.					
One	Percentiles	Quantile	SE Quantile	Lower 95% Limit	Upper 95% Limit
1	5.00	12.94	1.639	9.12	15.93
	10.00	19.97	5.681	16.35	39.97
	25.00	23.79	4.063	23.08	39.98
	50.00	30.17	2.548	29.39	39.99
	75.00	36.54	1.032	35.70	40.00
	90.00	62.84	6.793	39.49	67.74
	95.00	71.00	1.703	67.64	74.72

Notice that the problem with the quantiles falling outside of the confidence interval has been solved. However, several of the point estimates are still poor due to the large number of ties. To address this latter situation, **Example 6.2** illustrates the use of the NOISE option. The same data set that is used in **Example 6.1** is used here. Note how the point estimates for the percentiles change.

**Example 6.2 SUDAAN Code with NOISE option**

```

PROC DESCRIPT DATA=test DESIGN=WR;
NEST stratum psu;
WEIGHT analwgt1;
VAR SALES;
PERCENTILES 5 10 25 50 75 90 95 / NOISE;
SETENV COLWIDTH=6 DECWIDTH=2;
PRINT QTILE SEQTILE LOWQTILE UPQTILE/ SEQTILEFMT=F6.3;
PRINT QTILE SEQTILE LOWQTILE UPQTILE/ SEQTILEFMT=F6.3
                                filename="C:\foraddy1.rtf"

filetype=RTF replace;

```

**Results for Example 6.2**

Date: 09-19-2007 SUDAAN Page: 1  
Time: 18:37:58 Table: 1

Variance Estimation Method: Taylor Series (WR)  
Quantiles estimated using random noise (seed=56237485) with bound=1.000000e-006 to handle ties  
by: Variable, One, Percentiles.

for: Variable = SALES.

One	Percentiles	Quantile	SE Quantile	Lower 95% Limit	Upper 95% Limit
1	5.00	12.94	1.639	9.12	15.93
	10.00	40.00	5.687	16.35	40.00
	25.00	40.00	0.000	40.00	40.00
	50.00	40.00	0.000	40.00	40.00
	75.00	40.00	0.000	40.00	40.00
	90.00	62.84	6.670	40.00	67.74
	95.00	71.00	1.703	67.64	74.72

By using the NOISE option the percentiles produced by SUDAAN are more realistic and less biased.

## 7. Examples Illustrating the Revised SORTBY Statement in RECORDS

The following five examples illustrate the use of the SORTBY in PROC RECORDS. There are five calls to PROC RECORDS in these examples, each using a relatively simple dataset with two variables, X and Y. The first call simply prints the initial dataset. The other calls to PROC RECORDS illustrate what happens with different SORTBY statements. When the option "DESCENDING" appears before a variable, PROC RECORDS will sort that variable in descending order. Otherwise the variable is sorted in the default ASCENDING order.

### Example 7.1: SUDAAN Code

```
proc records data="C:\Test2.xpt" filetype=SASXPRT;  
print x y;
```

### Example 7.1: SUDAAN Results

Observation Number	X	Y
1	3.000	-1.000
2	4.000	-3.000
3	4.000	-2.000
4	1.000	-4.000
5	3.000	-2.000
6	2.000	-4.000
7	1.000	-2.000
8	3.000	-4.000
9	1.000	-1.000
10	2.000	-1.000
11	2.000	-3.000
12	3.000	-3.000
13	4.000	-4.000
14	1.000	-3.000
15	2.000	-2.000
16	4.000	-1.000

### Example 7.2: SUDAAN Code

```
proc records data="C:\Test2.xpt" filetype=sasxprt;  
sortby x y;  
print x y;
```

### Example 7.2: SUDAAN Results

Observation Number	X	Y
1	1.000	-4.000
2	1.000	-3.000
3	1.000	-2.000
4	1.000	-1.000
5	2.000	-4.000
6	2.000	-3.000
7	2.000	-2.000
8	2.000	-1.000
9	3.000	-4.000
10	3.000	-3.000
11	3.000	-2.000
12	3.000	-1.000
13	4.000	-4.000
14	4.000	-3.000
15	4.000	-2.000
16	4.000	-1.000

### Example 7.3: SUDAAN Code

```
proc records data="C:\Test2.xpt" filetype=sasxport;  
sortby x descending y;  
print x y;
```

### Example 7.3: SUDAAN Results

Observation Number	X	Y
1	1.000	-1.000
2	1.000	-2.000
3	1.000	-3.000
4	1.000	-4.000
5	2.000	-1.000
6	2.000	-2.000
7	2.000	-3.000
8	2.000	-4.000
9	3.000	-1.000
10	3.000	-2.000
11	3.000	-3.000
12	3.000	-4.000
13	4.000	-1.000
14	4.000	-2.000
15	4.000	-3.000
16	4.000	-4.000

**Example 7.4: SUDAAN Code**

```
proc records data="C:\Test2.xpt" filetype=sasxport;  
sortby descending x y;  
print x y;
```

**Example 7.4: SUDAAN Results**

```
-----  
Observation  
Number          X          Y  
-----  
1              4.000      -4.000  
2              4.000      -3.000  
3              4.000      -2.000  
4              4.000      -1.000  
5              3.000      -4.000  
6              3.000      -3.000  
7              3.000      -2.000  
8              3.000      -1.000  
9              2.000      -4.000  
10             2.000      -3.000  
11             2.000      -2.000  
12             2.000      -1.000  
13             1.000      -4.000  
14             1.000      -3.000  
15             1.000      -2.000  
16             1.000      -1.000  
-----
```

**Example 7.5: SUDAAN Code**

```
proc records data="C:\Test2.xpt" filetype=sasxport;  
sortby descending x descending y;  
print x y;
```

### Example 7.5: SUDAAN Results

Observation Number	X	Y
1	4.000	-1.000
2	4.000	-2.000
3	4.000	-3.000
4	4.000	-4.000
5	3.000	-1.000
6	3.000	-2.000
7	3.000	-3.000
8	3.000	-4.000
9	2.000	-1.000
10	2.000	-2.000
11	2.000	-3.000
12	2.000	-4.000
13	1.000	-1.000
14	1.000	-2.000
15	1.000	-3.000
16	1.000	-4.000

## 8. Examples Illustrating the New CROSSTAB Procedure Enhancements

Below are a series of examples to illustrate some of the new features in CROSSTAB, including the small percentage confidence intervals, the goodness-of-fit test, and tests for trend and related hypotheses. All examples were run using SAS-Callable SUDAAN *Release 9.0.3*.

### **Example 8.1: Small Percentage Confidence Interval**

This example illustrates the small percentage confidence interval (SPCI) enhancement in CROSSTAB. The data is from the 1988-1994 NHANES III (see *Section 1.8.5* of the *SUDAAN 9 Example Manual* for more detail on NHANES III). This example estimates the percentage of people ages 17-22 years that have been told by a doctor that they have arthritis. These estimates are computed by year of age. Since all ages in the subpopulation will likely have a very low prevalence of arthritis, we specify `SMCONF=5.0` on the PROC statement to trigger the SPCI method for all row percentages less than or equal to 5%. In the 2-way table of age-by-arthritis (`hsageir*hacla`), row percentages tell us the percentage of arthritis cases in each age group. Therefore, we request row percentages (`rowper`) on the PRINT statement.

The requested ROWSPCI will serve as a flag in the PRINT tables to indicate when the SPCI was used.

### Example 8.1: CROSSTAB Code Using NHANES III data

```
proc crosstab data=in.hanes3s5 design=wr SMCONF=5.0;
  nest sdpstra6 sdpps6;
  weight wtpfqx6;
  subpopn HSAGEIR < 22; /* 17-22 year-olds only */
  class hsageir HAC1A;
  table hsageir*HAC1A; /* age-by-arthritis table */
  print nsum rowper lowrow uprow rowspci / style=nchs;
  RFORMAT HAC1A yn.;
```

```
Number of observations read      : 20050      Weighted count :187647206
Observations in subpopulation   : 1967       Weighted count: 17314389
Denominator degrees of freedom  : 49
```

The frequency distribution for each CLASS variable is presented first. Of the 1,967 sample members in the subpopulation of 17-22 years, only 29 report having been told by a doctor that they had arthritis.

**Example 8.1: Frequency Distribution for CLASS Variables**

Frequencies and Values for CLASS Variables  
by: Age at interview (screener)

```
-----  
Age at  
  interview  
  (screener)      Frequency      Value  
-----  
Ordered  
  Position:  
  1                432          17  
Ordered  
  Position:  
  2                403          18  
Ordered  
  Position:  
  3                390          19  
Ordered  
  Position:  
  4                365          20  
Ordered  
  Position:  
  5                377          21  
-----
```

Frequencies and Values for CLASS Variables  
by: Doctor ever told you had: arthritis.

```
-----  
Doctor ever  
  told you  
  had:  
  arthritis      Frequency      Value  
-----  
Ordered  
  Position:  
  1                29          Yes  
Ordered  
  Position:  
  2               1938          No  
-----
```

**Example 8.1: Arthritis Prevalence Estimates and Confidence Intervals, by Age and Overall**

Variance Estimation Method: Taylor Series (WR)  
 For Subpopulation: HSAGEIR < 22  
 by: Age at interview (screener), Doctor ever told you had: arthritis.

```

-----
Age at interview                                     S
(screener)                                         P
  Doctor ever told                                 C
you had:                                           I
arthritis                                          R
                                                    o
                                                    w
Sample      Row      Lower      Upper
Size        Percent 95%      95%
                Limit  Limit
                ROWPER ROWPER
-----
Total
Total      1967      100.00      .      .
Yes         29        1.64        0.90     2.74  ^
No        1938      98.36        97.26     99.10  ^
17
Total      432      100.00      .      .
Yes         9         2.15        0.86     4.40  ^
No        423      97.85        95.60     99.14  ^
18
Total      403      100.00      .      .
Yes         2         0.55        0.02     2.71  ^
No        401      99.45        97.29     99.98  ^
19
Total      390      100.00      .      .
Yes         3         1.77        0.20     6.40  ^
No        387      98.23        93.60     99.80  ^
20
Total      365      100.00      .      .
Yes        10         2.14        0.51     5.77  ^
No        355      97.86        94.23     99.49  ^
21
Total      377      100.00      .      .
Yes         5         1.58        0.09     7.04  ^
No        372      98.42        92.96     99.91  ^
-----
  
```

^ LOWROW and UPROW use small percentage confidence interval calculation when ROWPER <= 5.00% or ROWPER >= 95.00%

For each year of age and overall, the percentage of people reporting arthritis (row percentage in this age\*arthritis table) is always  $\leq 5\%$  (yes) or  $\geq 95\%$  (no). SMCONF=5.0 is the criteria specified on the PROC statement to trigger the SPCI, and therefore all confidence limits for the row percentages in this example are based the SPCI approach. The last column in the table (from ROWSPCI keyword on PRINT statement) indicates when the SPCI was used. CROSSTAB also adds the table footnote to explain the meaning of the ^ symbol.

**Example 8.2: GOF Hypothesis**

This example uses the same NHANES III data set as well as external information from a more recent NHIS study to conduct a simple test using the GOFIT statement. According to the 2003-2005 NHIS survey, 21.6% of the adult population (defined as anyone ≥ 18 years of age) have been diagnosed with arthritis (see the website: <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5540a2.htm>).

In the NHANES III data, there are 19,618 sample members age 18 and over (see below). The GOFIT statement requests that the external NHIS percentages (21.6%=Yes vs. 78.4%=No) be compared to those in NHANES III to see if the two prevalence distributions are statistically different.

**Example 8.2: CROSSTAB Code for NHANES III Data**

```

proc crosstab data=hanes3s5 design=wr;
nest sdpstra6 sdpps6;
weight wtpfqx6;
subpopn hsageir >= 18; /* ages 18 and over */
class hacla;
table hacla; /* 1-way table of arthritis */
GOFIT hacla=(21.6 78.4) / waldf satadjchi;
PRINT nsum rowper serow / gof=all;
rformat HAC1A yn.;

```

Number of observations read	: 20050	Weighted count	:187647206
Observations in subpopulation	: 19618	Weighted count	:184017835
Denominator degrees of freedom	: 49		

Below is the sample frequency distribution for the NHANES III arthritis variable on the CLASS statement.

**Example 8.2: NHANES III Arthritis Frequency Distribution**

```

Frequencies and Values for CLASS Variables
by: Doctor ever told you had: arthritis.
-----
Doctor ever
  told you
  had:
  arthritis      Frequency      Value
-----
Ordered
  Position:
  1              4289          Yes
Ordered
  Position:
  2              15325         No
-----

```

As can be seen from the printout below, the arthritis percentages from the NHANES III data are estimated to be 17.7% 'Yes' and 82.3% 'No'.

**Example 8.2: NHANES III Arthritis Prevalence Estimates**

```

Variance Estimation Method: Taylor Series (WR)
For Subpopulation: HSAGEIR >= 18
by: Doctor ever told you had: arthritis.
-----
|                                     | Doctor ever told you had: |
|                                     | arthritis                 |
|                                     | Total   | Yes      | No       | |
|---|---|---|---|---|
|                                     | Sample Size | 19614 | 4289 | 15325 |
|                                     | Row Percent | 100.00 | 17.71 | 82.29 |
|                                     | SE Row Percent | 0.00 | 0.52 | 0.52 |
|-----|-----|-----|-----|-----|

```

Below is the GOF test for NHIS (21.6% “yes”) vs. NHANES (17.71% “yes”) using the default Wald F and the Satterthwaite-adjusted chi-square. Both tests are statistically significant (and equivalent since there are only 2 categories in the percentage distribution).

**Example 8.2: Comparison of NHANES III to Known Results from NHIS (21.6%=Yes, 78.4%=No)**

```

Variance Estimation Method: Taylor Series (WR)

Test Statistics for Goodnees-of-Fit Hypotheses
For: HAC1A = (0.216 0.784)

-----
Test Statistic          DF      Test Value      P-Value
-----|-----|-----|-----|
Wald-F                  1      55.7362         0.0000
Satterthwaite-adj chi-sq 1      55.7362         0.0000
-----|-----|-----|-----|

```

### **Example 8.3: New TEST Statement Options (Hypotheses and Test Statistics)**

This example uses the NHANES III dataset to evaluate whether having been diagnosed with arthritis (yes/no) is significantly associated with the self-reported condition of someone's natural teeth (excellent/very good, good, fair, poor/no natural teeth). We also want to evaluate this hypothesis after adjusting for age group (17-49, 50+) and self-reported general health status (excellent/very good, good, fair/poor). A more complete description of the NHANES III dataset is given in *Section 1.8.5* and *Chapter 3* of the *SUDAAN 9 Example Manual*.

We use this example only to illustrate some of the new features on the TEST statement in CROSSTAB. In **Example 8.3** below, the

- TABLES statement requests: 1) a single 2-way table of arthritis-by-teeth status, and 2) a stratified (by age and health status) two-way table of arthritis-by-teeth status.
- TEST statement requests the *stratum-specific* CHISQ test of general association for each table request, plus the *stratum-adjusted* tests of general association (CMH), trend (TCMH), and the ANOVA-type test (ACMH) for each table request. There is an option to request all available test statistics for each hypothesis (ALL), as well as to display the scores used for TCMH and ACMH hypotheses (DISPLAY).
- PRINT statement requests column percentages and related statistics for each TABLES request, as well as the defaults from the STEST (statistics for CHISQ hypothesis) and ATEST (statistics for CMH, TCMH, and ACMH hypotheses) groups.

There are 20,050 sample members in this NHANES-III dataset (see below).

### **Example 8.3: CROSSTAB Code for NHANES III data input**

```
PROC CROSSTAB DATA=nhanes3 FILETYPE=SAS DESIGN=WR deft1;
NEST SDPSTRA6 SDPPSU6;
WEIGHT WTPFQX6;
CLASS HAC1A AGEGRP2 HLTHSTAT TEETHSTAT;
TABLES HAC1A*TEETHSTAT
      AGEGRP2*HLTHSTAT*HAC1A*TEETHSTAT;
TEST CHISQ CMH ACMH TCMH / display ALL;
PRINT nsum wsum colper secol / STEST=default ATEST=default;
rformat agegrp2 age2fmt.;
rformat hac1a yesnofmt.;
rformat hlthstat hlthfmt.;
rformat teethstat teethfmt.;

RTITLE "NHANES-III: Examples of TEST Statement";

Number of observations read      : 20050      Weighted count :187647206
Denominator degrees of freedom : 49
```

Below are the unweighted frequency distributions for the variables on the CLASS statement (Arthritis, Age Group, Health Status, Teeth Status).

**Example 8.3: Frequency Distributions for Variables on CLASS Statement**

Frequencies and Values for CLASS Variables  
 Doctor Ever Told You Had Arthritis.

```

-----
Doctor Ever
  Told You
  Had
  Arthritis      Frequency      Value
-----
Ordered
  Position:
  1                4298        1=Yes
Ordered
  Position:
  2                15748       2=No
-----
  
```

Frequencies and Values for CLASS Variables  
 Age Group.

```

-----
Age Group      Frequency      Value
-----
Ordered
  Position:
  1                11396       1=17-49
Ordered
  Position:
  2                8654        2=50+
-----
  
```

### Example 8.3: Frequency Distributions

Frequencies and Values for CLASS Variables  
HLTHSTAT.

HLTHSTAT	Frequency	Value
Ordered Position: 1	7813	1=Excellent/Very Good
Ordered Position: 2	7191	2=Good
Ordered Position: 3	5033	3=Fair/Poor

Frequencies and Values for CLASS Variables  
TEETHSTAT.

TEETHSTAT	Frequency	Value
Ordered Position: 1	3881	1=Excellent/Very Good
Ordered Position: 2	5666	2=Good
Ordered Position: 3	4891	3=Fair
Ordered Position: 4	5595	4=Poor/No Natural Teeth

Below is the display of scores assigned to the teeth status and arthritis variables for the TCMH and ACMH hypotheses. Since there is no SCORES statement, all scores are defined by the values of the variables on the CLASS statement.

**Example 8.3: Display of Scores for ACMH and TCMH variables**

```

Scores Used in Computation of ACMH (column only), TCMH (row and column)
Doctor Ever Told You Had Arthritis.

-----
Doctor Ever
  Told You
  Had
  Arthritis      Value      Score
-----
Ordered
  Position:
  1              1=Yes      1
Ordered
  Position:
  2              2=No       2
-----

```

```

Scores Used in Computation of ACMH (column only), TCMH (row and column)
TEETHSTAT.

-----
TEETHSTAT              Value      Score
-----
Ordered
  Position:
  1              1=Excellent/Very Good      1
Ordered
  Position:
  2              2=Good                    2
Ordered
  Position:
  3              3=Fair                    3
Ordered
  Position:
  4              4=Poor/No Natural Teeth      4
-----

```

Below are various statistics from the first TABLES request: arthritis-by-teeth status (HAC1A\*TEETHSTAT). This table contains column percentages and related statistics, as requested on the PRINT statement (NSUM WSUM COLPER SECOL). In the interest of brevity, the stratified version of this arthritis-by-teeth status table (cross-classified by levels of age group and health status) generated by the TABLES request AGEGRP2\*HLTHSTAT\*HAC1A\*TEETHSTAT is not presented here, but was included in the PRINT tables.

Overall, the prevalence of diagnosed arthritis in the adult population increases from 11.2% for those with self-reported Excellent/Very Good Teeth, steadily up to 31.9% for those with Poor/No Natural Teeth.

**Example 8.3: Arthritis-by-Teeth Status: Column Percentages and Related Statistics**

Variance Estimation Method: Taylor Series (WR)  
 Doctor Ever Told You Had Arthritis, TEETHSTAT.

Doctor Ever Told You Had Arthritis		TEETHSTAT				
		Total	1=Excellent/Very Good	2=Good	3=Fair	4=Poor/No Natural Teeth
Total	Sample Size	20029	3881	5664	4890	5594
	Weighted Size	187505214	50781274	60305634	37175914	39242392
	Col Percent	100.00	100.00	100.00	100.00	100.00
	SE Col Percent	0.00	0.00	0.00	0.00	0.00
1=Yes	Sample Size	4291	518	899	882	1992
	Weighted Size	32625665	5703311	7964943	6442465	12514945
	Col Percent	17.40	11.23	13.21	17.33	31.89
	SE Col Percent	0.51	0.85	0.68	0.93	0.93
2=No	Sample Size	15738	3363	4765	4008	3602
	Weighted Size	154879549	45077963	52340690	30733449	26727447
	Col Percent	82.60	88.77	86.79	82.67	68.11
	SE Col Percent	0.51	0.85	0.68	0.93	0.93

Below are the results from the CHISQ hypothesis of general association between arthritis and teeth status (not stratified) from the TABLES request HAC1A\*TEETHSTAT. Using each of the 5 test statistics, there is a strong association.

In the interest of brevity, the CHISQ hypothesis tested within each level of age group by general health status (generated by the TABLES request AGEGRP2\*HLTHSTAT\*HAC1A\*TEETHSTAT) is not presented here, but was included in the PRINT tables.

**Example 8.3: Arthritis-by-Teeth Status: CHISQ test for general association (not stratified)**

```

Variance Estimation Method: Taylor Series (WR)

Test Statistics for Stratum-Specific Hypotheses
Variable HAC1A by Variable TEETHSTAT

-----
Hypothesis Test
  Test Statistic           DF      Adj DF  Test Value  P-Value
-----
CHISQ (Obs - Exp)
  Wald chi-square         3          .    238.5635   0.0000
  Wald-F                  3          .     79.5212   0.0000
  Adj Wald F              3          .     76.2754   0.0000
  Satterthwaite-adj chi-sq 3         2.55   209.7291   0.0000
  Satterthwaite-adj F      3         2.55    82.0970   0.0000
-----

```

Below are the results from the CMH, TCMH, and ACMH hypotheses for the 2-way table of arthritis by teeth status, corresponding to the TABLES request HAC1A\*TEETHSTAT. Since there is no stratification in this 2-way table, the CMH general association hypothesis is equivalent to the CHISQ hypothesis presented on the previous page. However, we now also get the test for trend and ANOVA-type test for this table. There is a significant trend in arthritis prevalence across teeth status. The ANOVA-type test (ACMH) is equivalent to the trend test (TCMH) when the row variable has only 2 levels.

**Example 8.3: Arthritis-by-Teeth Status (not stratified): General association (CMH), Trend (TCMH), and ANOVA-type Hypotheses (ACMH)**

```

Variance Estimation Method: Taylor Series (WR)

Test Statistics for Stratum-Adjusted Hypotheses
Variable HAC1A by Variable TEETHSTAT

-----
Hypothesis Test
  Test Statistic           DF      Adj DF  Test Value  P-Value
-----
CMH General Association
  Wald chi-square          3          .    238.5635   0.0000
  Wald-F                   3          .     79.5212   0.0000
  Adj Wald F               3          .     76.2754   0.0000
  Satterthwaite-adj chi-sq  3         2.55   209.7291   0.0000
  Satterthwaite-adj F      3         2.55    82.0970   0.0000
CMH Trend
  Wald chi-square          1          .    194.8541   0.0000
  Wald-F                   1          .    194.8541   0.0000
  Adj Wald F               1          .    194.8541   0.0000
  Satterthwaite-adj chi-sq  1         1.00   194.8541   0.0000
  Satterthwaite-adj F      1         1.00   194.8541   0.0000
CMH ANOVA (Row Means)
  Wald chi-square          1          .    194.8541   0.0000
  Wald-F                   1          .    194.8541   0.0000
  Adj Wald F               1          .    194.8541   0.0000
  Satterthwaite-adj chi-sq  1         1.00   194.8541   0.0000
  Satterthwaite-adj F      1         1.00   194.8541   0.0000
-----

```

Below are the results from the CMH, TCMH, and ACMH hypotheses for the *stratified* 2-way table of arthritis by teeth status, generated by the following TABLES request: AGEGRP2\*HLTHSTAT\*HAC1A\*TEETHSTAT. After adjustment for age group and general health status, there is still a significant general association and trend in arthritis prevalence across teeth status (although not nearly as significant after adjustment). The ANOVA-type test is again equivalent to the trend test for tables where the row variable has only 2 levels.

**Example 8.3: Stratified Arthritis-by-Teeth Status: General association (CMH), Trend (TCMH), and ANOVA-type Hypotheses (ACMH)**

```

Variance Estimation Method: Taylor Series (WR)

Test Statistics for Stratum-Adjusted Hypotheses
Variable HAC1A by Variable TEETHSTAT
Controlling for: Variable AGEGRP2 and HLTHSTAT

-----
Hypothesis Test
  Test Statistic           DF      Adj DF   Test Value   P-Value
-----
CMH General Association
  Wald chi-square           3         .      17.4716     0.0006
  Wald-F                    3         .       5.8239     0.0017
  Adj Wald F                3         .       5.5862     0.0023
  Satterthwaite-adj chi-sq  3        2.58    10.5241     0.0000
  Satterthwaite-adj F      3        2.58     4.0829     0.0152
CMH Trend
  Wald chi-square           1         .       6.2740     0.0123
  Wald-F                    1         .       6.2740     0.0156
  Adj Wald F                1         .       6.2740     0.0156
  Satterthwaite-adj chi-sq  1         1.00    6.2740     0.0123
  Satterthwaite-adj F      1         1.00    6.2740     0.0156
CMH ANOVA (Row Means)
  Wald chi-square           1         .       6.2740     0.0123
  Wald-F                    1         .       6.2740     0.0156
  Adj Wald F                1         .       6.2740     0.0156
  Satterthwaite-adj chi-sq  1         1.00    6.2740     0.0123
  Satterthwaite-adj F      1         1.00    6.2740     0.0156
-----

```