The George Washington University
Department of Statistics

STAT 6197 – Spring 2020

Week 7 – February 28, 2020

Major Topic: Summarizing Data, Creating Reports, and Formatting Output

Detailed Topics:

1) Summarizing Data Using Selected Base SAS Procedures (PROC Steps)
    - PROC MEANS
    - PROC SUMMARY
    - PROC FREQ
    - PROC SURVEYMEANS
    - PROC SURVEYFREQ
    - PROC TABULATE
    - PROC UNIVARIATE
    - PROC REPORT

2) Formatting Output Using Output Delivery System (ODS)
    - Create Output in Variety of Forms
    - Alter Output Appearance
    - Select Results Objects from PROC Steps
    - Create New Data Tables from PROC Step Output

Readings:

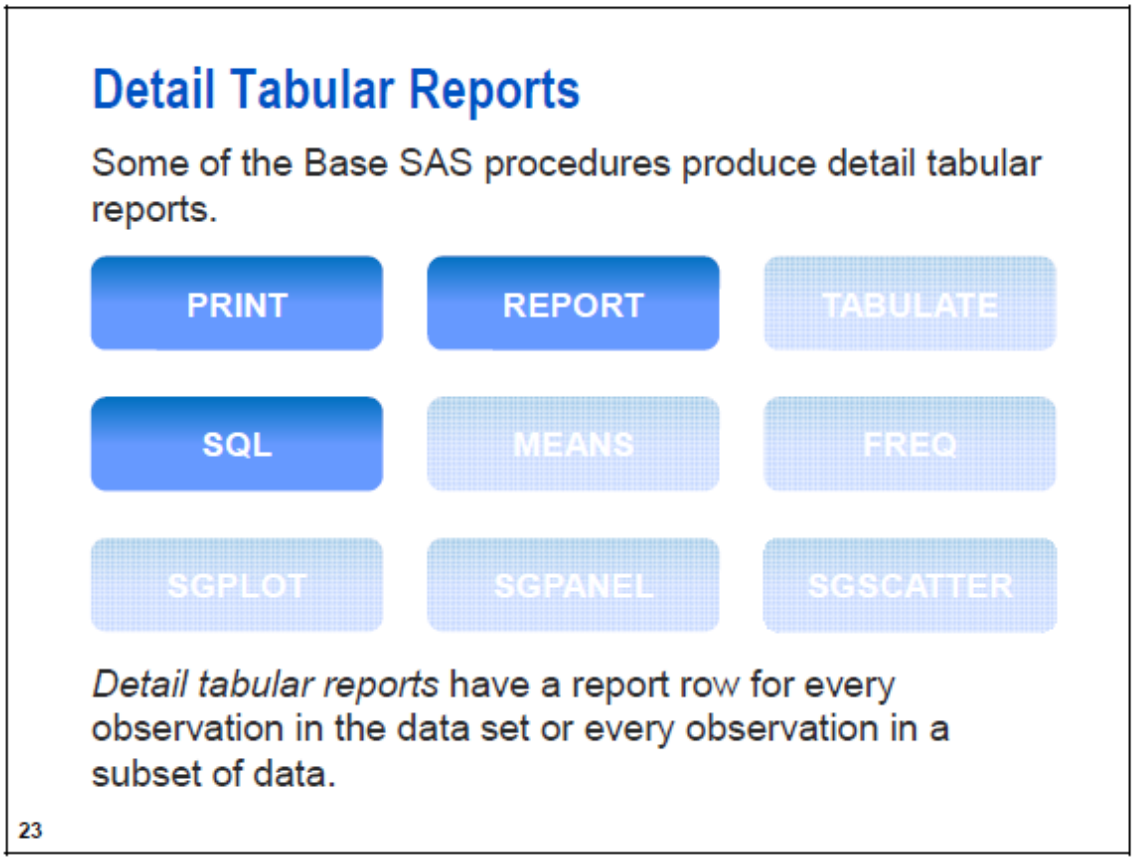
1. Relevant Chapters/Sections from Delwiche L, and Slaughter S. *The Little SAS Book: A Primer*, Fifth Edition Paperback – November 7, 2012

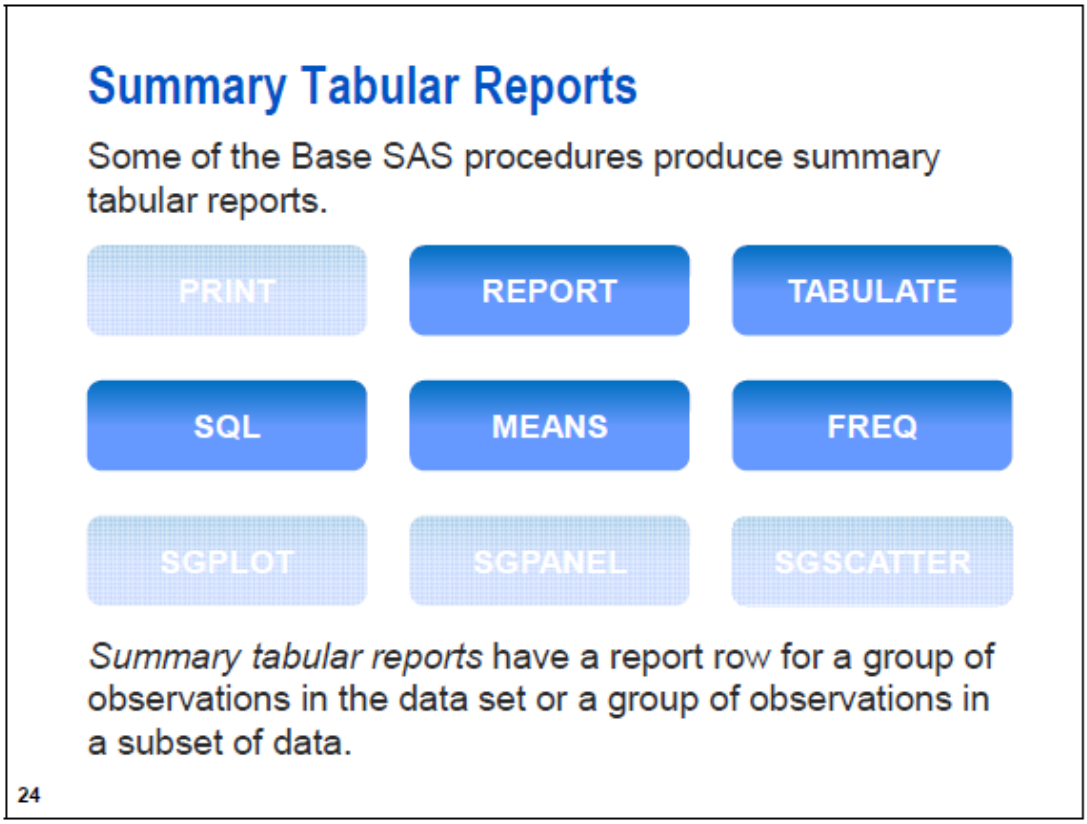
2. Exercises from Relevant Chapters/Sections - Ottesen RA, Delwiche LD, and Slaughter SJ. *Exercises and Projects for The Little SAS Book*, Fifth Edition Paperback – July 1, 2015

3. Carpenter, A.L. The MEANS/SUMMARY Procedure: Getting Started. SAS Global Forum 2010.

4. SAS/STAT 14.3 User's Guide. The SURVEYMEANS Procedure
5. SAS/STAT 14.3 User's Guide. The SURVEYFREQ Procedure
6. SAS(R) 9.4 Output Delivery System: User's Guide, Fifth Edition

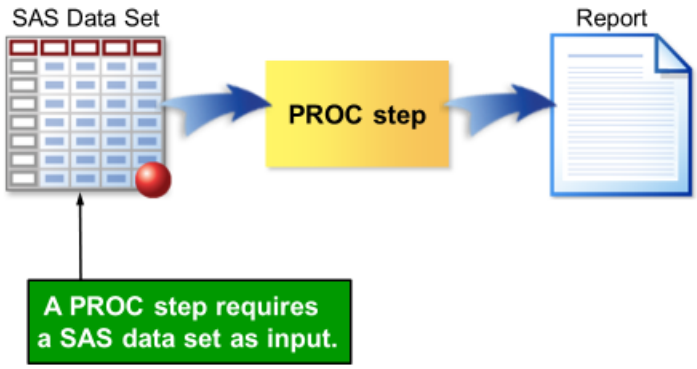
7. Using Procedure Based ODS Data Components in Statistical Reporting by VJ Faber





Planning to Create a Summary Report

To create a summary report in SAS, a PROC step is used.



9

Planning to Create a Summary Report

These steps are required for creating a summary report.

Step 1	Name the SAS data set to be referenced.
Step 2	Determine the PROC step to use.
Step 3	Determine the variables to analyze.
Step 4	Determine the grouping variables.

10

MEANS Procedure

The MEANS procedure produces summary reports with descriptive statistics.

```
proc means data=orion.sales;
run;
```

PROC MEANS DATA=*input-data-set <options statistics>*;
  <**VAR** *analysis-variable(s)*;>
  <**CLASS** *classification-variable(s)*;>
RUN;

- *Analysis variables* are the ***numeric*** variables for which statistics are to be computed.
- *Classification variables* are variables whose values define subgroups for the analysis. They can be character or numeric.

33 p111d07

VAR Statement

The VAR statement identifies the analysis variable (or variables) and their order in the output.

```
proc means data=orion.sales;
   var Salary;
run;
```

VAR *variable(s)*;

The MEANS Procedure

Analysis Variable : Salary

N	Mean	Std Dev	Minimum	Maximum
165	31160.12	20082.67	22710.00	243190.00

35 p111d07

CLASS Statement

The *CLASS statement* specifies one or more variables that the procedure uses to group the data.

Country	Year	N Obs	Variable	Label	N	Mean
CANADA	1993	240	PREDICT	Predicted Sales	240	497.20
			ACTUAL	Actual Sales	240	504.25
	1994	240	PREDICT	Predicted Sales	240	473.71
			ACTUAL	Actual Sales	240	524.88
GERMANY	1993	240	PREDICT	Predicted Sales	240	488.00
			ACTUAL	Actual Sales	240	530.85
	1994	240	PREDICT	Predicted Sales	240	476.81
			ACTUAL	Actual Sales	240	494.14
U.S.A.	1993	2				
	1994	2				

```
proc means data=sashelp.prdsale
           maxdec=2 n mean;
  var predict actual;
  class country year;
run;
```

104

CLASS Statement

The *CLASS statement* identifies variables whose values define subgroups for the analysis.

```
proc means data=orion.sales;
   var Salary;
   class Gender Country;
run;
```

CLASS *classification-variable(s)*;

- Classification variables are character or numeric.
- They typically have few discrete values.
- The data set does ***not*** need to be sorted or indexed by the classification variables.

36

p111d08

Viewing the Output

Statistics are produced for each combination of values of the classification variables.

The MEANS Procedure
Analysis Variable : Salary

Gender	Country	N Obs	N	Mean	Std Dev	Minimum	Maximum
F	AU	27	27	27702.41	1728.23	25185.00	30890.00
	US	41	41	29460.98	8847.03	25390.00	83505.00
M	AU	36	36	32001.39	16592.45	25745.00	108255.00
	US	61	61	33336.15	29592.69	22710.00	243190.00

- *N Obs* – the number of observations with each unique combination of class variables
- *N* – the number of observations with nonmissing values of the analysis variable (or variables)

37

PROC MEANS Statistics

Use options in the PROC MEANS statement to request specific statistics.

```
proc means data=orion.sales nmiss min max sum;
   var Salary;
   class Gender Country;
run;
```

The requested statistics override the default statistics.

39

p111d09

PROC MEANS Statement Options

Options can also be placed in the PROC MEANS statement.

Option	Description
MAXDEC=	Specifies the number of decimal places to display.
NONOBS	Suppresses the N Obs column.

41

MAXDEC= Option

MAXDEC=0

The MEANS Procedure
Analysis Variable : Salary

Country	N Obs	N	Mean	Std Dev	Minimum	Maximum
AU	63	63	30159	12699	25185	108255
US	102	102	31778	23556	22710	243190

MAXDEC=1

The MEANS Procedure
Analysis Variable : Salary

Country	N Obs	N	Mean	Std Dev	Minimum	Maximum
AU	63	63	30159.0	12699.1	25185.0	108255.0
US	102	102	31778.5	23555.8	22710.0	243190.0

p111d10

42

NONOBS Option

N Obs included by default

The MEANS Procedure
Analysis Variable : Salary

Country	N Obs	N	Mean	Std Dev	Minimum	Maximum
AU	63	63	30158.97	12699.14	25185.00	108255.00
US	102	102	31778.48	23555.84	22710.00	243190.00

NONOBS option

The MEANS Procedure
Analysis Variable : Salary

Country	N	Mean	Std Dev	Minimum	Maximum
AU	63	30158.97	12699.14	25185.00	108255.00
US	102	31778.48	23555.84	22710.00	243190.00

43 p111d10

Other PROC MEANS Statistics

Descriptive Statistic Keywords				
CLM	CSS	CV	LCLM	MAX
MEAN	MIN	MODE	N	NMISS
KURTOSIS	RANGE	SKEWNESS	STDDEV	STDERR
SUM	SUMWGT	UCLM	USS	VAR

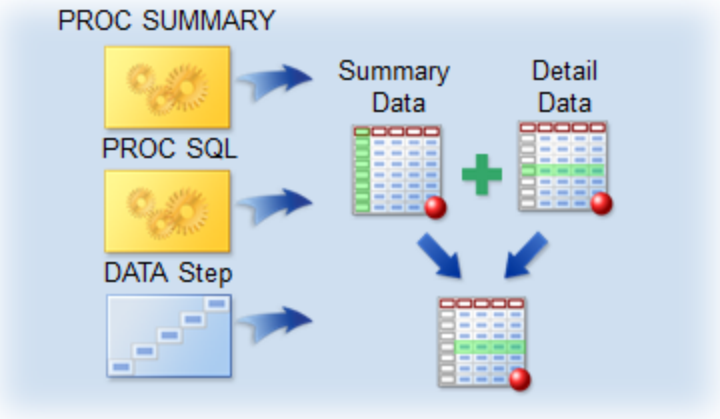
Quantile Statistic Keywords				
MEDIAN | P50	P1	P5	P10	Q1 | P25
Q3 | P75	P90	P95	P99	QRANGE

Hypothesis Testing Keywords				
PROBT	T			

44

Introduction

This section explores how to combine summary and detail data using PROC SUMMARY, PROC SQL, and the DATA step.



79

Business Scenario

Use **orion.totalsalaries** to calculate percentages of the total company payroll for each manager. This requires a few steps.

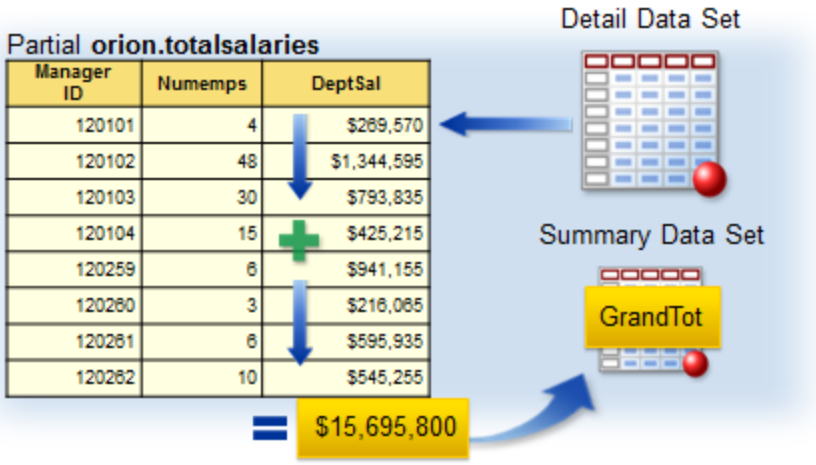
Partial **orion.totalsalaries**

Manager ID	Numemps	Dept Sal	GrandTot	Percent
120101	4	$269,570	$15,695,800	1.72%
120102	48	$1,344,595	$15,695,800	8.57%
120103	30	$793,835	$15,695,800	5.06%
120104	15	$425,215	$15,695,800	2.71%
120259	6	$941,155	$15,695,800	6.00%
120260	3	$216,065	$15,695,800	1.38%
120261	6	$595,935	$15,695,800	3.80%
120262	10	$545,255	$15,695,800	3.47%

81

Business Scenario

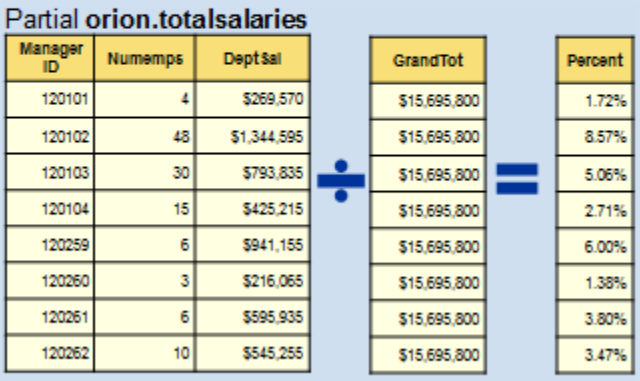
Calculate the grand total of **DeptSal** and store it in a summary data set.



82

Business Scenario

Combine the summary data with each row of the detailed data. Divide each detail amount by the grand total to calculate the percent of payroll by manager.



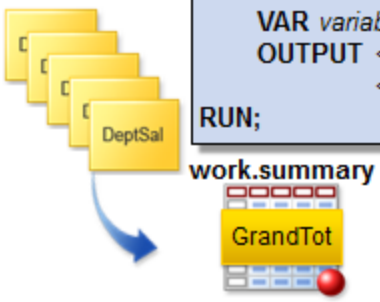
83

Creating a Summary Data Set

PROC SUMMARY generates descriptive statistics. The OUTPUT statement with the OUT= option creates a SAS data set with the summary statistics.

```
proc summary data=orion.totalsalaries;
   var DeptSal;
   output out=summary sum=GrandTot;
run;
```

```
PROC SUMMARY <option(s)><statistic-keyword(s)>;
    VAR variable(s);
    OUTPUT <OUT=SAS-data-set>
           <output-statistic-specification(s)>;
RUN;
```

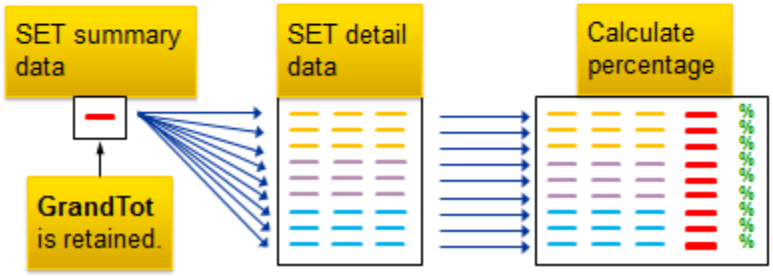


85 p307d10

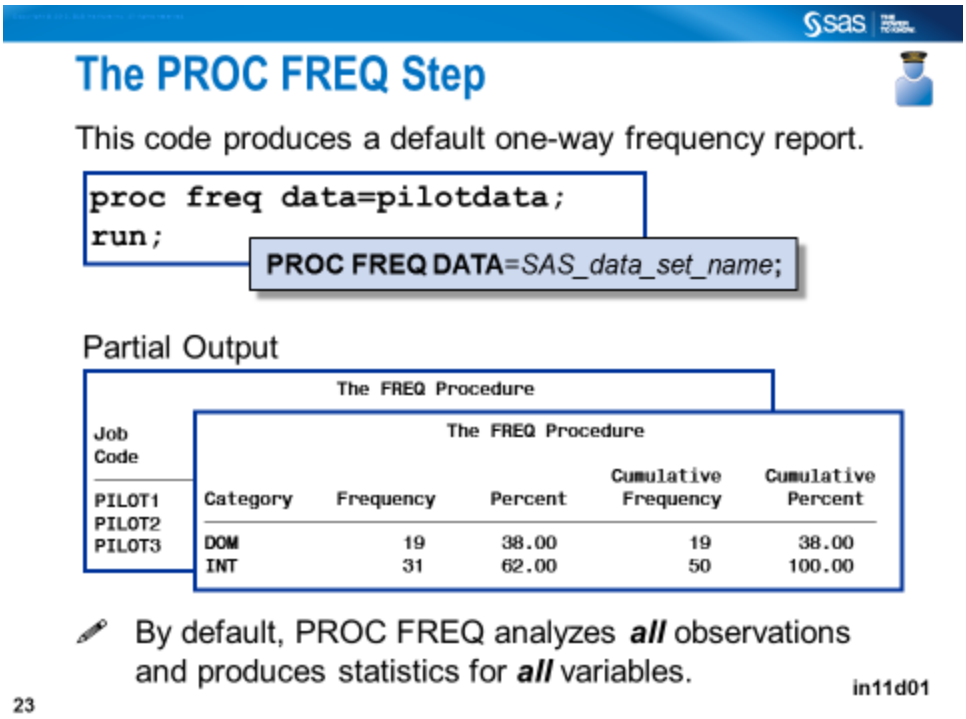
Combining the Summary and Detail Data

Use two SET statements in the DATA step to combine the summary and detailed data.

```
data percent;
   if _N_=1 then
      set summary(keep=GrandTot);
   set orion.totalsalaries;
   Percent=DeptSal/GrandTot;
   format Percent percent8.2;
run;
```



86 p307d10



The TABLES statement names the variables on which to calculate the frequency statistics.

Two-Way Frequency Report

A *two-way frequency table* is generated by specifying two variable names in the TABLES statement. The two names are joined with an asterisk (*).

```
proc freq data=pilotdata;
   tables Category*JobCode;
run;
```

column variable

row variable

TABLES *variable*variable*;

- The first variable refers to the row variable. The variable after the asterisk refers to the column variable.

in11d03

30

Two-Way Frequency Report

A **Category** by **JobCode** two-way table is produced.

The FREQ Procedure

Table of Category by JobCode

Category	JobCode			
Frequency Percent Row Pct Col Pct	PILOT1	PILOT2	PILOT3	Total
DOM	10 20.00 52.63 100.00	9 18.00 47.37 36.00	0 0.00 0.00 0.00	19 38.00
INT	0 0.00 0.00 0.00	16 32.00 51.61 64.00	15 30.00 48.39 100.00	31 62.00
Total	10 20.00	25 50.00	15 30.00	50 100.00

31

Options to Suppress Statistics

Use options in the TABLES statement to suppress the display of selected default statistics.

TABLES *variable(s) / options* ;

Option	Description
NOROW	Suppresses the display of the row percentage.
NOCOL	Suppresses the display of the column percentage.
NOPERCENT	Suppresses the percentage display.
NOFREQ	Suppresses the frequency display.

9

CROSSLIST Option

The *CROSSLIST option* displays two-way tables in column format, instead of cell format.

```
proc freq data=sashelp.orsales ;
   where product_line in ('Outdoors','Sports');
   tables product_line*year / crosslist ;
run;
```

Options in the TABLES statement must come after a forward slash.

88

The FREQ Procedure

Table of Product_Line by Year

Product_ Line	Year	Frequency	Percent	Row Percent	Column Percent
Outdoors	1999	28	6.25	25.00	25.00
	2000	28	6.25	25.00	25.00
	2001	28	6.25	25.00	25.00
	2002	28	6.25	25.00	25.00
	Total	112	25.00	100.00	
Sports	1999	84	18.75	25.00	75.00
	2000	84	18.75	25.00	75.00
	2001	84	18.75	25.00	75.00
	2002	84	18.75	25.00	75.00
	Total	336	75.00	100.00	
Total	1999	112	25.00		100.00
	2000	112	25.00		100.00
	2001	112	25.00		100.00
	2002	112	25.00		100.00
	Total	448	100.00		

89

NLEVELS Option

The *NLEVELS option* displays a table that provides the number of distinct values for each analysis variable.

```
proc freq data=orion.nonsales2 nlevels;
   tables Gender Country / nocum nopercent;
run;
```

```
PROC FREQ DATA=SAS-data-set NLEVELS;
    TABLES variable(s) ;
RUN;
```

16 p111d03

Viewing the Output

PROC FREQ Output

The FREQ Procedure

Number of Variable Levels

Variable	Levels	Missing Levels	Nonmissing Levels
Gender	4	1	3
Country	4	0	4

Gender	Frequency
F	110
G	1
M	123

Frequency Missing = 1

Country	Frequency
AU	33
US	196
au	3
us	3

17

NLEVELS Option

The _ALL_ keyword with the NOPRINT option displays the number of levels for all variables without displaying frequency counts.

```
proc freq data=orion.nonsales2 nlevels;
   tables _all_ / noprint;
run;
```

p111d04

22

Viewing the Output

PROC FREQ Output

The FREQ Procedure
Number of Variable Levels

Variable	Levels	Missing Levels	Nonmissing Levels
Employee_ID	234	1	233
First	204	0	204
Last	228	0	228
Gender	4	1	3
Salary	230	1	229
Job_Title	125	1	124
Country	4	0	4

No frequency tables were displayed.

23

Identifying Observations with Invalid Data

PROC FREQ has uncovered the existence of invalid data values for **Gender**, **Country**, and **Employee_ID**. Use PROC PRINT to display the observations with invalid values.

```
proc print data=orion.nonsales2;
   where Gender not in ('F','M') or
         Country not in ('AU','US') or
         Job_Title is null or
         Employee_ID is missing or
         Employee_ID=120108;
run;
```

24

Viewing the Output

PROC PRINT Output

Obs	Employee_ID	First	Last	Gender	Salary	Job_Title	Country
2	120104	Kareen	Billington	F	46230	Administration Manager	au
6	120108	Gladys	Gromek	F	27660	Warehouse Assistant II	AU
7	120108	Gabriele	Baker	F	26495	Warehouse Assistant I	AU
10	120112	Ellis	Glattback	F	26550		AU
12	120114	Jeannette	Buddery	G	31285	Security Manager	AU
14	.	Austen	Ralston	M	29250	Service Assistant II	AU
84	120695	Trent	Moffat	M	28180	Warehouse Assistant II	au
87	120698	Geoff	Kistanna	M	26160	Warehouse Assistant I	au
101	120723	Deanna	Olsen		33950	Corp. Comm. Specialist II	US
125	120747	Zashia	Farthing	F	43590	Financial Controller I	us
197	120994	Danelle	Sergeant	F	31645	Office Administrator I	us
200	120997	Mary	Donathan	F	27420	Shipping Administrator I	us

original observation numbers

25

Using Formats in PROC FREQ

A FORMAT statement can be used in PROC FREQ to format data values.

```
proc freq data=orion.sales;
   tables Hire_Date / nocum;
   format Hire_Date date9.;
run;
```

Partial PROC FREQ Output

The FREQ Procedure

Hire_Date	Frequency	Percent
01JAN1978	17	10.30
01FEB1978	2	1.21
01APR1978	1	0.61
01JUL1978	1	0.61
01AUG1978	1	0.61

many discrete values, and not what the manager requested

26 p111d05

Using Formats in PROC FREQ

A FORMAT statement can also be used in PROC FREQ to group the data.

```
proc freq data=orion.sales;
   tables Hire_Date / nocum;
   format Hire_Date year4.;
run;
```

Partial PROC FREQ Output

The FREQ Procedure

Hire_Date	Frequency	Percent
1978	23	13.94
1979	2	1.21
1980	4	2.42
1981	3	1.82
1982	7	4.24

fewer discrete values

27 p111d05

PROC SURVEYFREQ - Example

```
1  *Example_surveyfreq.sas;
2  options nonumber nodate ls=132 ps=58;
3  %LET Path=C:\Users\Pradip Muhuri\SASClassGWU\TopicsByWeek\Week9;
4  LIBNAME xnew "&Path";
5  options fmtsearch=(xnew.formats);
6  proc surveyfreq data=xnew.xh163;
7        tables  age_group *AMTOTV_cat13 / cv deff nowt ;
8        strata  varstr;
9        cluster varpsu;
10       weight  perwt13f;
11      title1 "MEPS 2013 Public Use Files";
12 run;
```

PROC SURVEYMEANS

The SURVEYMEANS procedure is used when the analysis involves data that are collected using complex multi-stage probability survey design (US Surveys, such as NHIS, MEPS, and NSDUH). The conventional PROC MEANS is not appropriate to analyze such data. The design characteristics and the survey weights must be specified as shown in the following example.

```
1  *Ex9_Surveymeans_ODS_RTF.sas;
2  OPTIONS nocenter nodate nonumber ps=58 ls=132  ;
3  libname out "D:\A_Data";
4  libname library "D:\A_Data";
5  ods graphics off;
6  proc surveymeans data=out.MEPS_FYC_2014 nobs mean ;
7              stratum varstr;
8              cluster varpsu;
9              weight perwt14f ;
10             var TOTEXP14;
11             domain age_grp sex age_grp*sex('Male');
12             ods output Statistics=out.ALL_FYC14
13                    domain=out.Stat_FY14;
14 run;
```

Line 7: The VAR statement identifies the variables to be analyzed. The CLASS statement (not added to the above code) identifies those numeric variables that are to be analyzed as categorical variables.

Line 8: The DOMAIN statement lists the variables that define domains for subpopulation analysis.

Line 9: The STRATA statement lists the variables that form the strata in a stratified sample design.

Like 10: The CLUSTER statement specifies cluster identification variables in a clustered sample design.

Like 11: The WEIGHT statement names the sampling weight variable.

PROC TABULATE

The TABULATE can be used to

- display descriptive statistics (e.g., N, SUM, and Mean) in tabular format
- produce tables in up to three dimensions
- report multiple variables one after another hierarchically within each dimension
- label and format the variables as well as the statistics

This procedure is appropriate for summary reports, for example, when the row collapses or summarizes data based on the group or category variables (Zender, 2008). In contrast, the PROC PRINT is appropriate for detail reports in which every observation in the data is listed.

One-Dimensional Table

Multiple TABLE statements can be in one step.

Country	
AU	US
N	N
63	102

```
proc tabulate data=orion.sales;
   class Country Gender;
   table Country;
   table Gender;
run;
```

Gender	
F	M
N	N
68	97

- N is the default statistic when there is only a CLASS statement and no VAR statement.

r102d02

27

Two-Dimensional Table

A two-dimensional table contains a row and a column dimension.

	Country	
	AU	US
	N	N
Gender		
F	27	41
M	36	61

```
proc tabulate data=orion.sales;
   class Country Gender;
   table Gender, Country;
run;
```

r102d02

28

Three-Dimensional Table

A three-dimensional table contains a page, a row, and a column dimension.

Job_Title Sales Rep. I

	Country	
	AU	US
	N	N
Gender		
F	8	13
M		

Job_Title Sales Rep. II

	Country	
	AU	US
	N	N
Gender		
F	10	14

Job_Title Sales Rep. III

	Country	
	AU	US
	N	N
Gender		
F	7	8

```
proc tabulate data=orion.sales;
   class Country Gender Job_Title;
   table Job_Title, Gender, Country;
run;
```

r102d02

29

Analysis Variable

SUM is the default statistic for an analysis variable that is used in the TABLE statement.

Job_Title Sales Rep. I

	Gender	
	F	M
	Salary	Salary
	Sum	Sum
Country		
AU	209390.00	346355.00
US	343730.00	769920.00

- Analysis variables are defined in the VAR statement.

```
proc tabulate data=orion.sales;
   class Country Gender Job_Title;
   var Salary;
   table Job_Title, Country, Gender*Salary;
run;
```

r102d02

36

ALL Keyword

The keyword ALL can be included in the TABLE statement to summarize all of the categories for class variables in the given dimension.

	Country		All
	AU	US	
	N	N	N
Gender			
F	27	41	68
M	36	61	97
All	63	102	165

- The keyword ALL does not belong in the CLASS or VAR statement.

```
proc tabulate data=orion.sales;
   class Country Gender;
   table Gender All, Country All;
run;
```

r102d03

45

ALL Keyword

The keyword ALL can be specified in any of the three dimensions.

```
table Job_Title All,
      Gender All,
      Country All;
```

Job_Title Sales Rep. IV

	Country		All
	AU	US	
	N	N	
Gender			
F	2	5	
M	3	6	
All	5	11	

All

	Country		All
	AU	US	
	N	N	N
Gender			
F	27	41	68
M	36	61	97
All	63	102	165

r102d03

46

Concatenated Variables

The keyword ALL can be included with concatenated variables.

	Gender		All	Country		All
	F	M		AU	US	
	N	N	N	N	N	N
Job_Title						
Sales Rep. I	21	42	63	21	42	63
Sales Rep. II	24	22	46	18	28	46

```
proc tabulate data=orion.sales;
   class Country Gender Job_Title;
   table Job_Title, Gender All Country All;
run;
```

r102d03

47

Default Statistics

N is the default statistic when there is only a CLASS statement and no VAR statement.

	Country		All
	AU	US	
	N	N	N
Gender			
F	27	41	68
M	36	61	97
All	63	102	165

```
proc tabulate data=orion.sales;
   class Country Gender;
   table Gender All, Country All;
run;
```

r102d04

57

Default Statistics

SUM is the default statistic when there is a VAR statement with or without a CLASS statement.

		Country		All
		AU	US	
		Salary	Salary	Salary
		Sum	Sum	Sum
Job_Title	Gender			
Sales Rep. I	F	209390.00	343730.00	553120.00
	M	346355.00	769920.00	1116275.00
Sales Rep. II	F	270645.00	387120.00	657765.00
	M	215375.00	388965.00	604340.00
Sales Rep. III	F			
	M			
Sales Rep. IV	F			
	M			
All				

```
proc tabulate data=orion.sales;
   class Country Gender Job_Title;
   var Salary;
   table Job_Title*Gender All,
         Country*Salary All*Salary;
run;
```

r102d04

58

Default Statistics

The default statistics can be specified in the TABLE statement. The asterisk (*) is used to associate a statistic keyword with a variable or the keyword ALL.

```
table Gender All,
      Country*N All*N;
```

```
table Job_Title*Gender All,
      Country*Salary*Sum All*Salary*Sum;
```

r102d04

59

Statistic Keywords

The following descriptive statistics and quantile statistics can be specified in the TABLE statement:

CSS	CV	KURTOSIS	LCLM	MAX
MEAN	MIN	MODE	N	NMISS
RANGE	SKEWNESS	STDDEV	STDERR	SUM
SUMWGT	UCLM	USS	VAR	
PCTN	PCTSUM	REPPCTN	REPPCTSUM	PAGEPCTN
PAGEPCTSUM	ROWPCTN	ROWPCTSUM	COLPCTN	COLPCTSUM
MEDIAN	P1	P5	P10	P25
P75	P90	P95	P99	QRANGE

60

Statistic Keywords

	Country		Salary	Salary
	AU	US		
	N	N	Mean	Max
Job_Title				
Sales Rep. I	21	42	26498.33	32235.00
Sales Rep. II	18	28	27437.07	35990.00
Sales Rep. III	17	17	29446.32	36605.00
Sales Rep. IV	5	11		
All	61	98		

```
proc tabulate data=orion.sales;
   class Country Job_Title;
   var Salary;
   table Job_Title All,
         Country*N
         Salary*Mean Salary*Max;
run;
```

r102d04

61

Parentheses

Multiple statistics can be specified in parentheses.

	Gender		Gender	
	F	M	F	M
	Salary	Salary	Salary	Salary
	Min	Min	Max	Max
Country				
AU	25185.00	25745.00	30890.00	108255.00
US	25390.00	22710.00	83505.00	243190.00
All	25185.00	22710.00	83505.00	243190.00

```
table Country All,
      Gender*Salary*Min
      Gender*Salary*Max;
```

	Gender			
	F		M	
	Salary		Salary	
	Min	Max	Min	Max
Country				
AU	25185.00	30890.00	25745.00	1082
US	25390.00	83505.00	22710.00	2431
All	25185.00	83505.00	22710.00	2431

```
table Country All,
      Gender*Salary*(Min Max);
```

r102d04

64

REPPCTN, ROWPCTN, and COLPCTN

	Country		All	Country		All
	AU	US		AU	US	
	N	N	N	RepPctN	RepPctN	RepPctN
Gender						
F	27	41	68	16.36	24.85	41.21
M	36	61	97	21.82	36.97	58.79
All	63	102	165	38.18	61.82	100.00

```
table Gender All,
      Country*N All*N
      Country*Reppctn
      All*Reppctn;
```

	Country		All	Country		All
	AU	US		AU	US	
	N	N	N	RowPctN	RowPctN	RowPctN
Gender						
F	27	41	68	39.71	60.29	100.00
M	36	61	97	37.11	62.89	100.00
All	63	102	165	38.18	61.82	100.00

```
table Gender All,
      Country*N All*N
      Country*Rowpctn
      All*Rowpctn;
```

	Country		All	Country		All
	AU	US		AU	US	
	N	N	N	ColPctN	ColPctN	ColPctN
Gender						
F	27	41	68	42.86	40.20	41.21
M	36	61	97	57.14	59.80	58.79
All	63	102	165	100.00	100.00	100.00

```
table Gender All,
      Country*N All*N
      Country*Colpctn
      All*Colpctn;
```

71

PCTN and PCTSUM

The denominator within angle brackets (<>) specifies which categories to sum for the denominator.

```
table Gender*(Country All) All,
      Job_Title*(N Pctn<Country All>)
      All*(N Pctn<Country All>);
```

		Job_Title				All	
		Sales Rep. I		Sales Rep. II			
		N	PctN	N	PctN	N	PctN
Gender	Country						
F	AU	8	38.10	10	41.67	18	40.00
	US	13	61.90	14	58.33	27	60.00
	All	21	100.00	24	100.00	45	100.00
M	Country						
	AU	13	30.95	8	36.36	21	32.81
	US	29	69.05	14	63.64	43	67.19
	All	42	100.00	22	100.00	64	100.00
All		63	100.00	46	100.00	109	100.00

Adding the denominator with PCTN and PCTSUM offers a great deal of flexibility.

r102d05

75

KEYLABEL Statement

The KEYLABEL statement, a statement specific to the TABULATE procedure, can be added to assign descriptive labels to statistic keywords and the ALL keyword.

	Location		Total
	AU	US	
	Annual Salary	Annual Salary	Annual Salary
	Average	Average	Average
Employee Gender			
F	27702.41	29460.98	28762.72
M	32001.39	33336.15	32840.77
Total	30158.97	31778.48	31160.12

```
keylabel Mean='Average'
         All='Total';
```

KEYLABEL *keyword-1* **='***description-1* **'**
<keyword-n **='***description-n* **'>;**

r103d01

6

KEYLABEL Statement

If the KEYLABEL statement specifies a blank value for a statistic, the cell is removed from the table.

	Location		Total
	AU	US	
	Annual Salary	Annual Salary	Annual Salary
	Average	Average	Average
Employee Gender			
F	27702.41	29460.98	2876
M	32001.39	33336.15	3284
Total	30158.97	31778.48	3116

```
keylabel Mean='Average'
         All='Total';
```

	Location		Total
	AU	US	
	Annual Salary	Annual Salary	Annual Salary
Employee Gender			
F	27702.41	29460.98	2876
M	32001.39	33336.15	3284
Total	30158.97	31778.48	3116

```
keylabel Mean=' '
         All='Total';
```

r103d01

9

Labels in the TABLE Statement

If the cell containing a classification or analysis variable is not desired, a blank label for the cell must be specified in the TABLE statement.

	Location		Total
	AU	US	
Employee Gender	Annual Salary	Annual Salary	Annual Salary
F	27702.41	2946	
M	32001.39	3333	
Total	30158.97	3177	

```
table Gender All,
      Country*Salary*Mean
      All*Salary*Mean;
```

	Location		Total
	AU	US	
F	27702.41	29460.98	28762.72
M	32001.39	33336.15	32840.77
Total	30158.97	31778.48	31160.12

```
table Gender=' ' All,
      Country*Salary=' '*Mean
      All*Salary=' '*Mean;
```

variable='label '

r103d01

10

Labels in the TABLE Statement

A label can also be specified in the TABLE statement to change the heading of a keyword.

	Location		Column Total
	AU	US	
F	27702.41	29460.98	28762.72
M	32001.39	33336.15	32840.77
Row Total	30158.97	31778.48	31160.12

```
table Gender=' ' All='Row Total',
      Country*Salary=' '*Mean
      All='Column Total'*Salary=' '*Mean;
```

keyword='label '

Recommendation: Only put a label in the TABLE statement if the label cannot be accomplished in the LABEL or KEYLABEL statement.

r103d01

11

3.02 Quiz – Correct Answer

When a label for a variable or keyword is specified in the LABEL or KEYLABEL statement ***and*** the TABLE statement, which label is used?

```
proc tabulate data=orion.sales;
   class Gender Country;
   var Salary;
   table Gender='GEN' All='Row',
         Country*Salary='SAL'*Mean
         All='Column'*Salary=' '*Mean;
   label Gender='Employee Gender'
         Salary='Annual Salary';
   keylabel Mean=' '
            All='Total';
run;
```

	Country		Column
	AU	US	
	SAL	SAL	
GEN			
F	27702.41	29460.98	28762.72
M	32001.39	33336.15	32840.77
Row	30158.97	31778.48	31160.12

13

FORMAT Statement

The FORMAT statement can be added to the TABULATE procedure to assign a format to a classification variable.

	Country		All
	Australia	United States	
	Salary	Salary	Salary
	Mean	Mean	Mean
Gender			
Female	27702.41	29460.98	28762.72
Male	32001.39	33336.15	32840.77
All	30158.97	31778.48	31160.12

```
format Gender $genfmt.
       Country $ctryfmt.;
```

FORMAT *variable-1 <variable-n> format*
<variable-2 <variable-n> format>;

r103d02

14

FORMAT= Option

The FORMAT= option can be added to the PROC TABULATE statement to specify a format for the statistical cells.

	Country				All	
	Australia		United States			
	Salary		Salary		Salary	
	N	Mean	N	Mean	N	Mean
Gender						
Female	27	27,702	41	29,461	68	28,763
Male	36	32,001	61	33,336	97	32,841
All	63	30,159	102	31,778	165	31,160

```
proc tabulate data=orion.sales
              format=comma12.;
```

FORMAT=*format*

r103d02

17

TABLE Statement

In the TABLE statement, the comma is used to separate the dimension expressions.

TABLE *page-expression* , *row-expression* , *column-expression* ;

- A one-dimensional table has a column expression only (no commas).
- A two-dimensional table has a row expression and a column expression (one comma).
- A three-dimensional table has a page, a row, and a column expression (two commas).

24

Formats in the TABLE Statement

A format modifier can be added to the TABLE statement by using the asterisk (*) to associate a format modifier with an analysis variable or keyword.

	Country				All	
	Australia		United States			
	Salary		Salary		Salary	
	N	Mean	N	Mean	N	Mean
Gender						
Female	27	27,702	41	29,461	68	$28,762.72
Male	36	32,001	61	33,336	97	$32,840.77
All	63	30,159	102	31,778	165	$31,160.12

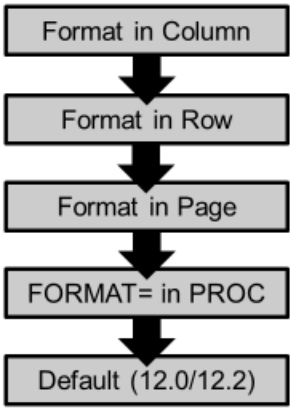
F=*format*

```
table Gender All,
      Country*Salary*(N Mean)
      All*Salary*(N Mean*f=dollar12.2);
```

18

Format Precedence

The following is the order for formats associated with the statistical cells.



By default, the format specified in the column dimension is used over any other format.

If no format is specified in any of the dimensions, the format specified with the FORMAT= option in the PROC statement is used over the default format.

19

3.03 Quiz – Correct Answer

Which format is used for the mean column?

```
proc tabulate data=orion.sales
              format=comma12.;
   class Gender Country;
   var Salary;
   table Gender All*f=dollar12.,
         Country*Salary*(N Mean*f=dollar12.2);
   format Gender $genfmt.
          Country $ctryfmt
run;
```

	Country			
	Australia		United States	
	Salary		Salary	
	N	Mean	N	Mean
Gender				
Female	27	$27,702.41	41	$29,460.98
Male	36	$32,001.39	61	$33,336.15
All	$63	$30,158.97	$102	$31,778.48

21

3.04 Quiz – Correct Answer

Why should the PERCENT*w.d* format ***not*** be used in order to create the desired report?

Initial Report

	Job_Title				All	
	Sales Rep. I		Sales Rep. II			
	N	PctN	N	PctN	N	PctN
Gender						
F	21	19.27	24	22.02	45	41.28
M	42	38.53	22	20.18	64	58.72
All	63	57.80	46	42.20	109	100.00

Desired Report

	Job_Title				All	
	Sales Rep. I		Sales Rep. II			
	N	PctN	N	PctN	N	PctN
Gender						
F	21	19.27%	24	22.02%	45	41.28%
M	42	38.53%	22	20.18%	64	58.72%
All	63	57.80%	46	42.20%	109	100.00%

The PERCENT*w.d* format multiplies values by 100 and adds a percent sign (%). The above values do not need to be multiplied by 100.

r103d02

24

Percent Format

A picture format can be created in order to add a percent sign to a percentage value.

```
proc format;
   picture pct(round) low-high='009.99%';
run;
```

PROC FORMAT;
PICTURE *name* <(*format-option(s)*)>
value-or-range-1 <..., *value-or-range-n*>='*picture*'
value-or-range-n <..., *value-or-range-n*>='*picture*';
RUN;

- The PICTURE statement creates a template for printing numbers.
- The ROUND option rounds the value before formatting.

25

continued...

Percent Format

- The nonzero digit selector (2-9) prints a digit.
- The zero-digit selector (0) prints a digit but not a zero.

```
proc format;
   picture pct(round) low-high='009.99%';
run;
```

The above step produces the following formatted values based on the stored values:

Stored value	.5	1.234	25	99.999
Formatted value	0.50%	1.23%	25.00%	100.00%

26

Percent Format

After the picture format is created, the format can be used in the TABULATE procedure.

```
proc format;
   picture pct(round) low-high='009.99%';
run;
proc tabulate data=orion.sales;
   class Gender Job_Title;
   table Gender All,
         Job_Title*(N Pctn*f=pct.)
         All*(N Pctn*f=pct.);
run;
```

	Job_Title				All	
	Sales Rep. I		Sales Rep. II			
	N	PctN	N	PctN	N	PctN
Gender						
F	21	19.27%	24	22.02%	45	41.28%
M	42	38.53%	22	20.18%	64	58.72%
All	63	57.80%	46	42.20%	109	100.00%

r103d02

27

BOX= Option

The BOX= option places text in the box area. This option is specified in the TABLE statement after a slash.

Warehouse Title	Gender	
	F	M
Warehouse Assistant I	5	
Warehouse Assistant II	3	
Warehouse Assistant III	1	
Warehouse Assistant IV	1	
Warehouse Manager		1

```
table Job_Title=' ', Gender*N=' '
      / box='Warehouse Title';
```

BOX = '*string*'

Job Title	Gender	
	F	
Warehouse Assistant I	5	
Warehouse Assistant II	3	
Warehouse Assistant III	1	
Warehouse Assistant IV	1	
Warehouse Manager		1

```
label Job_Title='Job Title';
table Job_Title=' ', Gender*N=' '
      / box=Job_Title;
```

BOX=*variable*

34

continued...

BOX= Option

The _PAGE_ value places the page dimension text in the box area.

Country AU	Gender	
Job Title	F	M
Warehouse Assistant I	2	2
Warehouse Assistant II	1	1
Warehouse Assistant III		1
Warehouse Assistant IV	1	

```
table Country, Job_Title,
      Gender*N=' '
      / box=_page_;
```

BOX=_PAGE_

Country US	Gender	
Job Title	F	M
Warehouse Assistant I	3	1
Warehouse Assistant II	2	
Warehouse Assistant III	1	
Warehouse Manager		1

r103d03

35

Missing Analysis Values

By default, missing analysis values are displayed with a period.

Job Title	Gender	
	F	M
Warehouse Assistant I	5	3
Warehouse Assistant II	3	1
Warehouse Assistant III	1	1
Warehouse Assistant IV	1	.
Warehouse Manager	.	1

36

MISSTEXT= Option

The MISSTEXT= option specifies a text string for missing analysis values. This option is specified in the TABLE statement after a slash.

Job Title	Gender	
	F	M
Warehouse Assistant I	5	3
Warehouse Assistant II	3	1
Warehouse Assistant III	1	1
Warehouse Assistant IV	1	None
Warehouse Manager	None	1

```
table Job_Title=' ', Gender*N=' '
      / box='Job Title'
        misstext='None';
```

MISSTEXT='*string*'

r103d03

37

MISSING= Option

An alternative to the MISSTEXT= option is the MISSING= option in the OPTIONS statement.

MISSING=*character*

```
options missing=0;

proc tabulate data=orion.warehouse;
   class Gender Job_Title;
   table Job_Title=' ', Gender*N=' '
         / box='Job Title';
run;

options missing=.;
```

resets back to the default

The MISSING= option specifies the character to be printed for missing numeric values but is limited to one character.

r103d03

38

MISSING Option

The MISSING option includes observations with missing classification values. If the option is placed in the PROC statement, all missing classification values are included.

```
proc tabulate data=orion.warehouse missing;
   class Gender Job_Title;
   table Job_Title=' ', Gender*N=' '
         / box='Job Title';
run;
```

Job Title	Gender		
		F	M
	.	1	.
Warehouse Assistant I	1	5	3
Warehouse Assistant II	.	3	1
Warehouse Assistant III	.	1	1
Warehouse Assistant IV	.	1	.
Warehouse Manager	.	.	1

r103d03

42

MISSING Option

The MISSING option can be placed in the CLASS statement to control which missing classification values are included.

```
proc tabulate data=orion.warehouse;
   class Gender / missing;
   class Job_Title;
   table Job_Title=' ', Gender*N=' '
         / box='Job Title';
run;
```

Job Title	Gender		
		F	M
Warehouse Assistant I	1	5	3
Warehouse Assistant II	.	3	1
Warehouse Assistant III	.	1	1
Warehouse Assistant IV	.	1	.
Warehouse Manager	.	.	1

The MISSING option is specified in the CLASS statement after a slash.

r103d03

43

PRINTMISS Option

The PRINTMISS option creates row and column headings that are the same for all pages of a table. This option is specified in the TABLE statement after a slash.

```
table Country, Job_Title, Gender*N=' '
      / box=_page_ printmiss;
```

Country AU	Gender	
	F	M
Job_Title		
Warehouse Assistant I	2	2
Warehouse Assistant II	1	1
Warehouse Assistant III	.	1
Warehouse Assistant IV	1	.
Warehouse Manager	.	

Country US	Gender	
	F	M
Job_Title		
Warehouse Assistant I	3	1
Warehouse Assistant II	2	.
Warehouse Assistant III	1	.
Warehouse Assistant IV	.	.
Warehouse Manager	.	1

r103d03

47

Trafficlighting

The STYLE= option can be added to the TABULATE procedure in order to incorporate trafficlighting.

Job Title	Gender			
	Female		Male	
	Salary		Salary	
	N	Mean	N	Mean
Chief Marketing Officer	1	$207,885	.	.
Marketing Assistant I	1	$27,380	.	.
Marketing Assistant II	2	$29,203	3	$29,772
Marketing Assistant III	.	.	1	$31,630
Marketing Assistant IV	.	.	1	$34,925
Marketing Manager	1	$63,640	2	$63,165
Senior Marketing Manager	1	$87,420	.	.

Trafficlighting is the act of highlighting individual cells based on the cell's value.

r103d04

60

```
proc format;
   value ncell 1 = 'Gold';
   value meancell 0-30000 = 'LightBlue'
                  60000-high = 'Pink'
                  . = 'Wheat';
run;

proc tabulate data=orion.marketing;
   class Gender Job_Title;
   var Salary;
   table Job_Title=' ',
         Gender*Salary*
         (N*{style={background=ncell.}}
          Mean*f=dollar12.*
          {style={background=meancell.}})
         / box=Job_Title;
   label Job_Title='Job Title';
   format Gender $genfmt.;
run;
```

64

Trafficlighting Output

The program from the previous slide creates the following tabular report:

Job Title	Gender			
	Female		Male	
	Salary		Salary	
	N	Mean	N	Mean
Chief Marketing Officer	1	$207,885	.	.
Marketing Assistant I	1	$27,380	.	.
Marketing Assistant II	2	$29,203	3	$29,772
Marketing Assistant III	.	.	1	$31,630
Marketing Assistant IV	.	.	1	$34,925
Marketing Manager	1	$63,640	2	$63,165
Senior Marketing Manager	1	$87,420	.	.

- A count of 1 gets a gold background.
- A mean value between 0 and 30,000 gets a light blue background.
- A mean value 60,000 or higher gets a pink background.
- A missing mean value gets a wheat background.

65

UNIVARIATE Procedure

PROC UNIVARIATE displays extreme observations, missing values, and other statistics for the variables included in the VAR statement.

```
proc univariate data=orion.nonsales2;
   var Salary;
run;
```

PROC UNIVARIATE DATA=*SAS-data-set*;
    <**VAR** *variable(s)*;>
RUN;

If the VAR statement is omitted, PROC UNIVARIATE analyzes all numeric variables in the data set.

46 p111d11

Viewing the Output: Extreme Observations

The *Extreme Observations* section includes the five lowest and five highest values for the analysis variable and the corresponding observation numbers.

Partial PROC UNIVARIATE Output

Extreme Observations

-----Lowest----		-----Highest----	
Value	Obs	Value	Obs
2401	20	163040	1
2650	13	194885	231
24025	25	207885	28
24100	19	268455	29
24390	228	433800	27

Obs is the observation number, not the count of observations with that value.

47

NEXTROBS= Option

The *NEXTROBS= option* specifies the number of extreme observations to display.

```
proc univariate data=orion.nonsales2
                nextrobs=3;
   var Salary;
run;
```

Partial PROC UNIVARIATE Output

```
        The UNIVARIATE Procedure
            Variable:  Salary

          Extreme Observations

-----Lowest----        -----Highest----

  Value      Obs          Value      Obs

   2401       20         207885       28
   2650       13         268455       29
  24025       25         433800       27
```

48

p111d11

ID Statement

The ID statement displays the value of the identifying variable (or variables) in addition to the observation number.

```
proc univariate data=orion.nonsales2;
   var Salary;
   id Employee_ID;
run;
```

ID *variable(s)*;

49

p111d11

Chapter 4: Getting Started with the REPORT Procedure

- 4.1 Introduction to the REPORT Procedure
- 4.2 Creating a Detail Report
- 4.3 Creating a Summary Report with the GROUP Usage
- 4.4 Creating a Summary Report with the ACROSS Usage
- 4.5 Computing an Additional Column

1

REPORT Procedure

The REPORT procedure produces detail and summary reports.

Country	Department	Name	HireDate	Salary	Bonus
Australia	Administration	Kareen Billington	01JAN1995	$46,230	$2,312
		Liz Povey	01MAY2003	$27,110	$1,356
		John Hamsey	01JAN1978	$26,960	$1,348
		Sherie Sheedy	01FEB1978	$30,475	$1,524
		Gladys Gromek	01AUG2010	$27,660	$1,383
		Gabriele Baker	01OCT2010	$26,495	$1,325
		Dennis Entwisle	01NOV1982		
		Ubaldo Spillane	01NOV1978		
		Ellis Glattback	01JUL1994		
		Ria Horsey	01JAN1978		
		Jeannette Buddery	01JAN1978		
		Hugh Nicholles	01AUG2010		
		Austen Ralston	01FEB1982		
	Engineering	Bill Mccleary			

		Salary				
Country	Department	Count	Average	Minimum	Maximum	Total
Australia	Administration	13	$29,300	$26,495	$46,230	$380,895
	Engineering	4	$29,410	$27,645	$31,670	$117,660
	Sales	76	$27,496	$24,015	$36,605	$2,138,430
	Sales Management	3	$119,757	$87,575	$163,840	$359,270
	Stock & Shipping	10	$34,321	$26,160	$60,960	$343,210
		106	$30,921	$24,015	$163,840	$3,339,465
Belgium	Administration	11	$29,475	$25,045	$43,200	$324,220
	Engineering	2	$28,633	$27,260	$29,806	$57,066
	Sales	45	$27,429	$25,240	$32,470	$1,234,300
	Sales Management	2	$119,775	$83,305	$156,245	$239,550
		60	$30,919	$25,045	$156,245	$1,855,135
Canada				$25,780	$42,895	$268,805

r104d01

5

Detail Report

A detail report contains one row for every observation selected for the report.

Country	Department	Name	HireDate	Salary	Bonus
Australia	Administration	Kareen Billington	01JAN1985	$46,230	$2,312
		Liz Povey	01MAY2003	$27,110	$1,356
		John Hornsey	01JAN1978	$26,960	$1,348
		Sherie Sheedy	01FEB1978	$30,475	$1,524
		Gladys Gromek	01AUG2010	$27,660	$1,383
		Gabriele Baker	01OCT2010	$26,495	$1,325
		Dennis Entwisle	01NOV1983	$28,615	$1,431
		Ubaldo Spillane	01NOV1978	$26,895	$1,345
		Ellis Glattback	01JUL1994	$26,550	$1,328
		Riu Horsey	01JAN1978	$26,870	$1,344
		Jeannette Buddery	01JAN1978	$31,285	$1,564
		Hugh Nicholas	01AUG2009	$26,500	$1,325
		Austen Ralston	01FEB1984	$29,250	$1,463
				$380,895	*$19,045*
	Engineering	Bill Mccleary	[illegible]	$31,670	$1,584

6

Summary Report

A summary report consolidates data so that each report row represents multiple observations.

		Salary				
Country	Department	Count	Average	Minimum	Maximum	Total
Australia	Administration	13	$29,300	$26,495	$46,230	$380,895
	Engineering	4	$29,415	$27,645	$31,670	$117,660
	Sales	78	$27,416	$24,015	$36,605	$2,138,430
	Sales Management	3	$119,757	$87,975	$163,040	$359,270
	Stock & Shipping	10	$34,321	$26,160	$60,980	$343,210
		108	*$30,921*	*$24,015*	*$163,040*	*$3,339,465*
Belgium	Administration	11	$29,475	$25,045	$43,280	$324,220
	Engineering	2	$28,533	$27,260	$29,805	$57,065
	Sales	45	$27,429	$25,240	$32,470	$1,234,300
	Sales Management	2	$119,775	$83,305	$156,245	$239,550
		60	*$30,919*	*$25,045*	*$156,245*	*$1,855,135*
Denmark					$25,780	$42,095	$268,805

7

Summary Lines

Both detail and summary reports can contain summary lines (break lines) as well as report rows.

		Salary				
Country	Department	Count	Average	Minimum	Maximum	Total
Australia	Administration	13	$29,300	$26,495	$46,230	$380,895
	Engineering	4	$29,415	$27,645	$31,670	$117,660
	Sales	78	$27,416	$24,015	$36,605	$2,138,430
	Sales Management	3	$119,757	$87,975	$163,040	$359,270
	Stock & Shipping	10	$34,321	$26,160	$60,980	$343,210
		108	*$30,921*	*$24,015*	*$163,040*	*$3,339,465*
Belgium	Administration	11	$29,475	$25,045	$43,280	$324,220
	Engineering	2	$28,533	$27,260	$29,805	$57,065
	Sales	45	$27,429	$25,240	$32,470	$1,234,300
	Sales Management	2	$119,775	$83,305	$156,245	$239,550
		60	*$30,919*	*$25,045*	*$156,245*	*$1,855,135*
Denmark	Administration	9	$29,867	$25,780	$42,095	$268,805
	Engineering					$57,365

report rows

summary lines

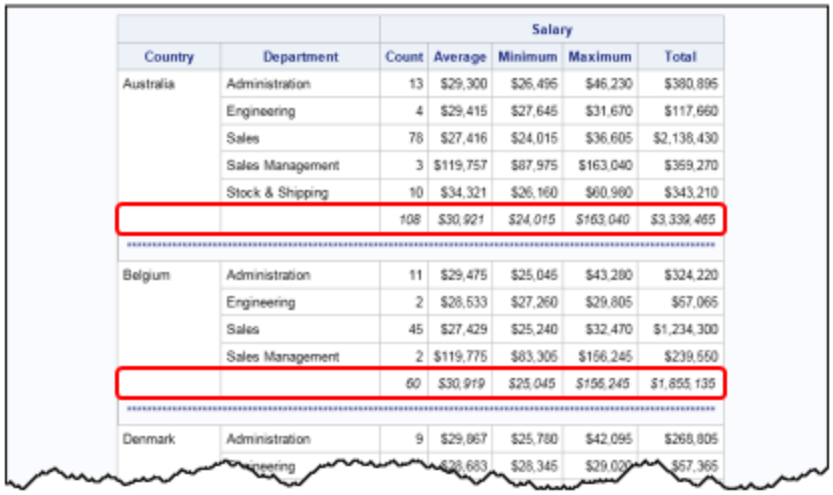
report rows

summary lines

8

Standard Summary Lines

Standard summary lines summarize numerical data for a set of report rows.



		Salary				
Country	Department	Count	Average	Minimum	Maximum	Total
Australia	Administration	13	$29,300	$26,495	$46,230	$380,895
	Engineering	4	$29,415	$27,645	$31,670	$117,660
	Sales	78	$27,416	$24,015	$36,605	$2,138,430
	Sales Management	3	$119,757	$87,975	$163,040	$359,270
	Stock & Shipping	10	$34,321	$26,160	$60,960	$343,210
		108	*$30,921*	*$24,015*	*$163,040*	*$3,339,465*
Belgium	Administration	11	$29,475	$25,045	$43,280	$324,220
	Engineering	2	$28,533	$27,260	$29,805	$57,065
	Sales	45	$27,429	$25,240	$32,470	$1,234,300
	Sales Management	2	$119,775	$83,305	$156,245	$239,550
		60	*$30,919*	*$25,045*	*$156,245*	*$1,855,135*
Denmark	Administration	9	$29,867	$25,780	$42,095	$268,805
	Engineering		$28,683	$28,345	$29,020	$57,365

9

Usage of Variables

In the REPORT procedure, the report variables are defined with usages that determine the type of report.

The following are the six usages:

Detail Report	Summary Report	Detail or Summary Report
DISPLAY	GROUP	ANALYSIS
ORDER	ACROSS	COMPUTED

13

DISPLAY Usage

Detail Report

A *display variable* has a row for every observation in the data set. Order of the rows is based on data set order.

		Anita Boudsema III	01JUL2008	$27,060	$1,353
				$55,985	$2,799
Spain	Administration	Carlos Sifre Vila	01AUG1985	$26,310	$1,316
		Horacio González Aguilera	01JAN1978	$27,780	$1,389
		María Martínez Martínez	01APR2010	$47,200	$2,360
		Pablo Martínez	01DEC2000	$43,740	$2,187
		José María Montserrat	01JUN2010	$26,960	$1,348
		María Dolores García	01SEP2010	$28,825	$1,441
		Rafael Suárez de Peón Carbó	01MAR2003	$26,645	$1,332
		Jaime Esnaola Sukía	01JAN1978	$27,520	$1,376
				$254,980	*$12,749*
	Engineering	Jorge Azemar Santiago	01JUN2005	$28,800	$1,440
		Víctor Pérez Cimarro	01AUG2010	$26,500	$1,325
				$55,300	*$2,765*
United Kingdom	Administration	David Liveing	01APR2008	$26,090	$1,305
		David Geoghegan	01JUL2008	$27,205	$1,360

14

ORDER Usage

Detail Report

An *order variable* displays a row for every observation in the data set plus orders the detail rows by the values.

		Anita Boudsema III	01JUL2008	$27,060	$1,353
				$55,985	$2,799
Spain	Administration	Carlos Sifre Vila	01AUG1985	$26,310	$1,316
		Horacio González Aguilera	01JAN1978	$27,780	$1,389
		María Martínez Martínez	01APR2010	$47,200	$2,360
		Pablo Martínez	01DEC2000	$43,740	$2,187
		José María Montserrat	01JUN2010	$26,960	$1,348
		María Dolores García	01SEP2010	$28,825	$1,441
		Rafael Suárez de Peón Carbó	01MAR2003	$26,645	$1,332
		Jaime Esnaola Sukía	01JAN1978	$27,520	$1,376
				$254,980	*$12,749*
	Engineering	Jorge Azemar			
		Víctor Pérez C			
United Kingdom	Administration	David Liveing	01APR2008	$26,090	$1,305
		David Geoghegan	01JUL2008	$27,205	$1,360

Repetitious printing is suppressed for the order variables.

15

GROUP Usage

Summary Report

A *group variable* collapses observations with the same values into summary rows plus orders the summary rows.

		Salary				
Country	Department	Count	Average	Minimum	Maximum	Total
Australia	Administration	13	$29,300	$26,496	$46,230	$380,896
	Engineering	4	$29,415	$27,645	$31,670	$117,660
	Sales	78	$27,416	$24,015	$36,605	$2,138,430
	Sales Management	3	$119,757	$87,975	$163,040	$359,270
	Stock & Shipping	10	$34,321	$26,160	$60,980	$343,210
		108	*$30,921*	*$24,015*	*$163,040*	*$3,339,465*
Belgium	Administration	11	$29,475	$25,045	$43,280	$324,220
	Engineering	2	$28,533	$27,260	$29,805	$57,065
	Sales	45	$27,429	$25,240	$32,470	$1,234,300
	Sales Manageme					
					$25,780	$42,095	$268,805

Repetitious printing is suppressed for the group variable.

16

ACROSS Usage

Summary Report

An *across variable* creates a column for each value of the variable plus orders the columns.

	Country									
Gender	Australia	Belgium	Denmark	France	Germany	Italy	Netherlands	Spain	United Kingdom	United States
Female	48	27	23	44	45	34	30	29	50	143
Male	60	33	28	54	53	43	35	37	59	173
	108	*60*	*51*	*98*	*98*	*77*	*65*	*66*	*109*	*316*

17

r104d01

ANALYSIS Usage

Detail or Summary Report

An *analysis variable* is a numeric variable that is used to calculate a specified statistic.

		Salary				
Country	Department	Count	Average	Minimum	Maximum	Total
Australia	Administration	13	$29,300	$26,495	$46,230	$380,895
	Engineering	4	$29,415	$27,645	$31,670	$117,660
	Sales	78	$27,416	$24,015	$36,605	$2,138,430
	Sales Management	3	$119,757	$87,975	$163,040	$359,270
	Stock & Shipping	10	$34,321	$26,160	$60,980	$343,210
		108	*$30,921*	*$24,015*	*$163,040*	*$3,339,465*
Belgium	Administration	11	$29,475	$25,045	$43,280	$324,220
	Engineering	2	$28,533	$27,260	$29,805	$57,065
	Sales	45	$27,429	$25,240	$32,470	$1,234,300
	Sales Management	2	$119,775	$83,305	$156,245	$239,550
		60	*$30,919*	*$25,045*	*$156,245*	*$1,855,135*
Denmark				$25,780	$42,095	$268,805

18

ANALYSIS Usage

Detail or Summary Report

The value of an analysis variable in a detail report is the value of the statistic for a single observation.

Country	Department	Name	HireDate	Salary	Bonus
Australia	Administration	Kareen Billington	01JAN1985	$46,230	$2,312
		Liz Povey	01MAY2003	$27,110	$1,356
		John Hornsey	01JAN1978	$26,960	$1,348
		Sherie Sheedy	01FEB1978	$30,475	$1,524
		Gladys Gromek	01AUG2010	$27,660	$1,383
		Gabriele Baker	01OCT2010	$26,495	$1,325
		Dennis Entwisle	01NOV1983	$28,615	$1,431
		Ubaldo Spillane	01NOV1978	$26,895	$1,345
		Ellis Glattback	01JUL1994	$26,550	$1,328
		Riu Horsey	01JAN1978	$26,870	$1,344
		Jeannette Buddery	01JAN1978	$31,285	$1,564
		Hugh Nicholas	01AUG2009	$26,500	$1,325
		Austen Ralston	01FEB1984	$29,250	$1,463
				$380,895	*$19,045*
	Engineering	Bill Mccleary	[illegible]	$31,670	$1,584

19

ANALYSIS Usage

Detail or Summary Report

The value of an analysis variable in a summary line is the value of the statistic for all represented observations.

Spain	Administration	Carlos Sifre Vila	01AUG1985	$26,310	$1,316
		Horacio González Aguilera	01JAN1978	$27,780	$1,389
		María Martínez Martínez	01APR2010	$47,200	$2,360
		Pablo Martínez	01DEC2000	$43,740	$2,187
		José María Montserrat	01JUN2010	$26,960	$1,348
		María Dolores García	01SEP2010	$28,825	$1,441
		Rafael Suárez de Peón Carbó	01MAR2003	$26,645	$1,332
		Jaime Esnaola Sukía	01JAN1978	$27,520	$1,376
				$254,980	*$12,749*

		Salary				
		ount	Average	Minimum	Maximum	Total
Australia	Administration	13	$29,300	$26,495	$46,230	$380,896
	Engineering	4	$29,415	$27,645	$31,670	$117,660
	Sales	78	$27,416	$24,015	$36,605	$2,138,430
	Sales Management	3	$119,757	$87,975	$163,040	$359,270
	Stock & Shipping	10	$34,321	$26,160	$60,980	$343,210
		108	*$30,921*	*$24,015*	*$163,040*	*$3,339,465*

20

COMPUTED Usage

Detail or Summary Report

A *computed variable* is a variable that is not in the input data set but a variable defined in the procedure.

Country	Department	Name	HireDate	Salary	Bonus
Australia	Administration	Kareen Billington	01JAN1985	$46,230	$2,312
		Liz Povey	01MAY2003	$27,110	$1,356
		John Hornsey	01JAN1978	$26,960	$1,348
		Sherie Sheedy	01FEB1978	$30,475	$1,524
		Gladys Gromek	01AUG2010	$27,660	$1,383
		Gabriele Baker	01OCT2010	$26,495	$1,325
		Dennis Entwisle	01NOV1983	$28,615	$1,431
		Ubaldo Spillane	01NOV1978	$26,895	$1,345
		Ellis Glattback	01JUL1994	$26,550	$1,328
		Riu Horsey	01JAN1978	$26,870	$1,344
		Jeannette Buddery	01JAN1978	$31,285	$1,564
		Hugh Nicholas	01AUG2009	$26,500	$1,325
		Austen Ralston	01FEB1984	$29,250	$1,463
				$380,895	*$19,045*
	Engineering	Bill Mccleary	[illegible]	$31,670	$1,584

21

COLUMN Statement

The COLUMN statement specifies the report items to include in the report. It also controls the order of the items.

```
proc report data=orion.emps nowd;
   where Department='Engineering';
   column EmployeeID Name BirthDate HireDate
          JobTitle Country Gender Salary;
run;
```

COLUMN *report-item-1 < ... report-item-n>*;

Employee ID	Employee Name	Employee Birth Date	Employee Hire Date	Job Title	Employee Country	Employee Gender	Annual Salary
120117	Bill Mccleary	11SEP1968	01APR1990	Cabinet Maker III	Australia	Male	$31,670
120118	Darshi Hartshorn	03JUN1963	01JUL1988	Cabinet Maker II	Australia	Male	$28,090
120119	Lal Elleman	21DEC1973	01JAN2002	Electrician IV	Australia	Male	$30,255
120120	Krishna Peiris	05MAY1948	01JAN1978	Electrician II	Australia	Female	$27,645
120210	Nelly Baes	09MAY1953	01JAN2007	Cabinet Maker II	Belgium	Female	$29,805
120211	Johan Sarrand	16NOV1953	01JAN2007	Electrician I	Belgium	Male	$27,260
120429	Walter Szczesniak	13APR1968	01APR2009	Engineering Manager	Germany	Male	$36,240
120430	Hans-Martin de Palacios	06SEP1978	01MAR2004	Cabinet Maker I	Germany	Male	$27,250
[illegible]							

r104d02

36

Spanning Column Headings

The COLUMN statement can also specify spanning column headings.

```
column ('Engineering Employees'
        EmployeeID Name BirthDate HireDate
        JobTitle Country Gender Salary);
```

COLUMN (*'header-1' < ... 'header-n'> report-item(s)*);

Engineering Employees							
Employee ID	Employee Name	Employee Birth Date	Employee Hire Date	Job Title	Employee Country	Employee Gender	Annual Salary
120117	Bill Mccleary	11SEP1968	01APR1990	Cabinet Maker III	Australia	Male	$31,670
120118	Darshi Hartshorn	03JUN1963	01JUL1988	Cabinet Maker II	Australia	Male	$28,090
120119	Lal Ellenan	21DEC1973	01JAN2002	Electrician IV	Australia	Male	$30,255
120120	Krishna Peiris	05MAY1948	01JAN1978	Electrician II	Australia	Female	$27,645
120210	Nelly Baes	09MAY1953	01JAN2007	Cabinet Maker II	Belgium	Female	$29,805
120211	Johan Sarrand	16NOV1953	01JAN2007	Electrician I	Belgium	Male	$27,260
120429	Walter Szczesniak	13APR1988	01APR2009	Engineering Manager	Germany	Male	$36,240

r104d02

37

DEFINE Statement

The DEFINE statement describes how to use and display a report item. For example, formats and labels can be used in DEFINE statements as options.

```
proc report data=orion.emps nowd;
   where Department='Engineering';
   column ('Engineering Employees'
           EmployeeID Name BirthDate HireDate
           JobTitle Country Gender Salary;
   define HireDate / format=mmddyy10.
                     'Hire Date';
   define BirthDate / 'Birth Date' f=monyy7.;
run;
```

DEFINE *report-item-n* / <*usage*> <*option(s)*>;

r104d03

42

DEFINE Statement

Options must be placed after a forward slash. The ordering of options does not matter, and the order of the DEFINE statements does not matter.

```
define HireDate / format=mmddyy10.
                  'Hire Date';
define BirthDate / 'Birth Date' f=monyy7.;
```

'*column-header*' | **FORMAT=***format*

- The column heading must be enclosed in single or double quotation marks.
- F= is an alias for the FORMAT= option.

r104d03

43

SPLIT= Option

A split character can be specified in the column heading along with the SPLIT= option in the PROC statement.

```
proc report data=orion.emps nowd split='*';
  ...
  define HireDate / format=mmddyy10.
                    'Hire*Date';
  define BirthDate / 'Birth*Date' f=monyy7.;
run;
```

SPLIT='*character***'**

Engineering Employees

Employee ID	Employee Name	Birth Date	Hire Date	Job Title	Employee Country	Employee Gender	Annual Salary

The default split character is the forward slash.

r104d03

44

Usage

A usage can be specified in a DEFINE statement after the forward slash.

```
define Name / display 'Name';
define Salary / analysis sum
                'Yearly Salary' f=comma12.;
```

Remember, the following are the six usages:

Detail Report	Summary Report	Detail or Summary Report
DISPLAY	GROUP	ANALYSIS
ORDER	ACROSS	COMPUTED

r104d03

47

Default Usages

By default, character variables have a usage of DISPLAY.

Engineering Employees

Employee ID	Name	Birth Date	Hire Date	Job Title	Employee Country	Employee Gender	Yearly Salary
120117	Bill Mccleary	SEP1968	04/01/1990	Cabinet Maker III	Australia	Male	31,670
120118	Darshi Hartshorn	JUN1963	07/01/1988	Cabinet Maker II	Australia	Male	28,090
120119	Lal Elleman	DEC1973	01/01/2002	Electrician IV	Australia	Male	30,255

DISPLAY

Numeric variables have a usage of ANALYSIS with a default statistic of SUM.

Engineering Employees

Employee ID	Name	Birth Date	Hire Date	Job Title	Employee Country	Employee Gender	Yearly Salary
120117	Bill Mccleary	SEP1968	04/01/1990	Cabinet Maker III	Australia	Male	31,670
120118	Darshi Hartshorn	JUN1963	07/01/1988	Cabinet Maker II	Australia	Male	28,090
120119	Lal Elleman	DEC1973	01/01/2002	Electrician IV	Australia	Male	30,255

ANALYSIS SUM

48

Default Usages

A detail report is created when the usage of DISPLAY is used for one or more of the reporting variables.

Engineering Employees

Employee ID	Name	Birth Date	Hire Date	Job Title	Employee Country	Employee Gender	Yearly Salary
120117	Bill Mccleary	SEP1968	04/01/1990	Cabinet Maker III	Australia	Male	31,670
120118	Darshi Hartshorn	JUN1963	07/01/1988	Cabinet Maker II	Australia	Male	28,090
120119	Lal Elleman	DEC1973	01/01/2002	Electrician IV	Australia	Male	30,255
120120	Krishna Peiris	MAY1948	01/01/1978	Electrician II	Australia	Female	27,645
120210	Nelly Baes	MAY1953	01/01/2007	Cabinet Maker II	Belgium	Female	29,805
120211	Johan Sarrand	NOV1953	01/01/2007	Electrician I	Belgium	Male	27,260
120429	Walter Szczesniak	APR1988	04/01/2009	Engineering Manager	Germany	Male	36,240
120430	Hans-Martin de Palacios	SEP1978	03/01/2004	Cabinet Maker I	Germany	Male	27,250
120431	Susanne Kitta	MAR1983	08/01/2005	Cabinet Maker II	Germany	Female	29,430
120432	Michael Endejan	JUL1990	05/01/2010	Electrician III	Germany	Male	27,735
120433	Curt-M. Thompson	OCT1968	07/01/2010	Electrician I	Germany	Male	27,380

- The value of an analysis variable in a detail report is the value of the statistic for a single observation.

49

ORDER Usage

If the default usage is not desired, an alternative usage can be specified. Reporting variables with an ORDER usage also create a detail report but orders the rows.

```
define JobTitle / order;
define Gender / order 'Gender';
```

- When multiple order variables are used, the order of the detail rows are based on how the order variables appear in the COLUMN statement.

r104d04

50

Objectives

- Use the GROUP usage to create summary reports.
- Use options to enhance the summary report.
- Specify statistics for analysis variables.
- Use the comma operator or aliases to use multiple statistics for the same analysis variable.

61

GROUP Usage

A group variable collapses observations with the same values into summary rows. It also orders the summary rows.

```
proc report data=orion.emps nowd;
   column Country Department Salary;
   define Country / group;
   define Department / group;
   define Salary / analysis sum;
run;
```

Employee Country	Department	Annual Salary
Australia	Administration	$380,895
	Engineering	$117,660
	Sales	$2,138,430
	Sales Management	$359,270
	Stock & Shipping	$343,210
Belgium	Administration	$324,220
	Engineering	

64

r104d05

GROUP Usage

The SPANROWS option in the PROC statement and the DESCENDING and ORDER= options in the DEFINE statement can be used with summary reports.

```
proc report data=orion.emps nowd spanrows;
   column Country Department Salary;
   define Country / group order=internal;
   define Department / group descending;
   define Salary / analysis sum;
run;
```

DK
ES
FR

Denmark	Sales Management	$239,035
	Sales	$1,046,920
	Engineering	$57,365
	Administration	$268,805
Spain	Sales Management	$275,780
	Sales	$1,466,135
	Engineering	$55,300
	Administration	$254,980
France	Stock & Shipping	$457,055

r104d05

66

GROUP Usage versus ORDER Usage

The GROUP and ORDER usages have similar traits. However, the GROUP usage creates a summary report and the ORDER usage creates a detail report.

Trait	GROUP Usage	ORDER Usage
Orders the rows	Yes	Yes
Suppresses the repetitious printing of values	Yes	Yes
Collapses rows with same values	Yes	No

67

GROUP Usage

The SPANROWS option in the PROC statement and the DESCENDING and ORDER= options in the DEFINE statement can be used with summary reports.

```
proc report data=orion.emps nowd spanrows;
   column Country Department Salary;
   define Country / group order=internal;
   define Department / group descending;
   define Salary / analysis sum;
run;
```

DK
ES
FR

Denmark	Sales Management	$239,035
	Sales	$1,046,920
	Engineering	$57,365
	Administration	$268,805
Spain	Sales Management	$275,780
	Sales	$1,466,135
	Engineering	$55,300
	Administration	$254,980
France	Stock & Shipping	$457,055

66 r104d05

GROUP Usage versus ORDER Usage

The GROUP and ORDER usages have similar traits. However, the GROUP usage creates a summary report and the ORDER usage creates a detail report.

Trait	GROUP Usage	ORDER Usage
Orders the rows	Yes	Yes
Suppresses the repetitious printing of values	Yes	Yes
Collapses rows with same values	Yes	No

67

MISSING Option

The MISSING option includes observations with missing group or order variable values.

Employee Country	Annual Salary
	.
Australia	$3,339,465
Belgium	$1,866,136
Denmark	$1,612,125
France	$3,112,330
Germany	$2,982,870
Italy	$2,339,736
Netherlands	$2,015,645
Spain	$2,052,196
United Kingdom	$3,329,950
United States	$12,790,136

```
proc report data=orion.emps
            nowd missing;
```

```
define Country / group
                 missing;
```

The option can be placed in the PROC statement to impact all variable values or in the DEFINE statement to impact the specified variable values.

r104d05

72

COMPLETEROWS Option

The COMPLETEROWS option displays all possible combinations of the values of the group variable.

```
proc report data=orion.emps nowd completerows;
```

With COMPLETEROWS

Employee Country	Job Title	Annual Salary
Australia	Office Assistant I	
	Office Assistant II	$26,960
	Office Assistant III	$30,475
	Office Assistant IV	.
Italy	Office Assistant I	$52,175
	Office Assistant II	.
	Office Assistant III	.
	Office Assistant IV	.
United Kingdom	Office Assistant I	$53,295
	Office Assistant II	.
	Office Assistant III	.
	Office Assistant IV	$32,240

Without COMPLETEROWS

Employee Country	Job Title	Annual Salary
Australia	Office Assistant II	$26,960
	Office Assistant III	$30,475
Italy	Office Assistant I	$52,175
United Kingdom	Office Assistant I	$53,295
	Office Assistant IV	$32,240

73

r104d05

Default Summary Statistic

SUM is the default statistic for numeric variables with the usage of ANALYSIS.

```
proc report data=orion.emps nowd;
   column Country Department Salary;
   define Country / group;
   define Department / group;
   define Salary / analysis sum;
run;
```

Employee Country	Department	Annual Salary
Australia	Administration	$380,895
	Engineering	$117,660
	Sales	$2,138,430
	Sales Management	$359,270
	Stock & Shipping	$343,210
Belgium	Administration	$324,220
	Engineering	[illegible]

76

r104d06

Default Summary Statistic

The following statements produce the same results:

```
define Salary / analysis sum 'Salary';
```

```
define Salary / analysis 'Salary';
```

```
define Salary / sum 'Salary';
```

```
define Salary / 'Salary';
```

- By default, numeric variables have a usage of ANALYSIS with a statistic of SUM.

77

Statistic Keywords

The following descriptive statistics and quantile statistics are available in the REPORT procedure:

CSS	CV	MAX	MEAN
MIN	MODE	N	NMISS
PCTN	PCTSUM	RANGE	STDDEV
STDERR	SUM	SUMWGT	USS
VAR	MEDIAN | P50	P1	P5
P10	P25 | Q1	P75 | Q3	P90
P95	P99	QRANGE	

78

N and PCTN Statistics

The N and PCTN statistics can be specified directly in the COLUMN statement.

```
proc report data=orion.emps nowd;
   column Country Department N PctN Salary;
   define Country / group;
   define Department / group;
   define N / 'Count';
   define PCTN / 'Percent' format=percent8.2;
   define Salary / analysis;
run;
```

Employee Country	Department	Count	Percent	Annual Salary
Australia	Administration	13	1.24%	$380,896
	Engineering	4	0.38%	$117,660
	Sales	78	7.44%	$2,138,430
	Sales Management	3	0.29%	$359,270
	Stock & Shipping	10	0.95%	$343,210
Belgium	Administration	11	1.05%	$324,220
	Engineering	3	0.29%	$57,065

79 r104d06

Statistic in the DEFINE Statement

A statistic can be specified in the DEFINE statement for an analysis variable.

```
proc report data=orion.emps nowd;
   column Country Department Salary;
   define Country / group;
   define Department / group;
   define Salary / analysis mean
                   'Average Salary';
run;
```

Employee Country	Department	Average Salary
Australia	Administration	$29,300
	Engineering	$29,415
	Sales	$27,416
	Sales Management	$119,757
	Stock & Shipping	$34,321
Belgium	Administration	$29,475
	Engineering	$28,533

82 r104d06

Comma Operator

If multiple statistics are required for an analysis variable, the comma operator can be used to cross the statistics with the analysis variable.

```
proc report data=orion.emps nowd;
   column Country Department Salary,(Mean Sum);
   define Country / group;
   define Department / group;
   define Salary / analysis 'Salary';
   define Mean / 'Average';
   define Sum / 'Total' format=dollar12.;
run;
```

		Salary	
Employee Country	Department	Average	Total
Australia	Administration	$29,300	$380,895
	Engineering	$29,415	$117,660
	Sales	$27,416	$2,138,430
	Sales Management	$119,757	$359,270

83 r104d06

Comma Operator

The comma is used to stack statistics under an analysis variable. The parentheses are used to group statistics to appear at the same level. The label for the analysis variable can be removed.

```
proc report data=orion.emps nowd;
   column Country Department Salary,(Mean Sum);
   define Country / group;
   define Department / group;
   define Salary / analysis ' ';
   define Mean / 'Average';
   define Sum / 'Total' format=dollar12.;
run;
```

null label

Employee Country	Department	Average	Total
Australia	Administration	$29,300	$380,895
	Engineering	$29,415	$117,660
	Sales	$27,416	$2138430
	Sales Management	$119,757	$359,270

84 r104d06

One Statistic Multiple Times

A particular statistic can be used multiple times in a summary report.

```
proc report data=orion.emps nowd;
   column Country Department N Salary,(N Mean);
   define Country / group;
   define Department / group;
   define Salary / analysis 'Salary';
   define N / format=6.;
run;
```

			Salary	
Employee Country	Department	N	N	Mean
Australia	Administration	13	13	$29,300
	Engineering	4	4	$29,415

number of observations with grouping

number of nonmissing analysis variable values

r104d07

87

Report Item Alias

An alias can be specified for a report item so that the report item can be used more than one time in a COLUMN statement.

```
proc report data=orion.emps nowd;
   column Country Department N
          ('Salary' Salary Salary=SalMean);
   define Country / group;
   define Department / group;
   define N / 'Count';
   define Salary / n 'Non-Missing' format=6.;
   define SalMean / mean 'Average';
run;
```

report-item=alias-name

			Salary	
Employee Country	Department	Count	Non-Missing	Average
Australia	Administration	13	13	$29,300
	Engineering	4	4	$29,415
	Sales	78	78	$27,416

r104d07

90

Report Item Alias

The alias has a DEFINE statement, which enables multiple statistics to be used with a variable.

```
proc report data=orion.emps nowd;
   column Country Department
          Salary Salary=SalMax
          HireDate HireDate=HDMax;
   define Country / group;
   define Department / group;
   define Salary / min 'Salary/Low';
   define SalMax / max 'Salary/High';
   define HireDate / min f=year4.
                     'Hire Year/Low';
   define HDMax / max f=year4.
                  'Hire Year/High';
run;
```

r104d07

93

ACROSS Usage

An across variable creates a column for each value of the variable plus orders the columns.

	Country									
Gender	Australia	Belgium	Denmark	France	Germany	Italy	Netherlands	Spain	United Kingdom	United States
Female	48	27	23	44	45	34	30	29	50	143
Male	60	33	28	54	53	43	35	37	59	173
	108	*60*	*51*	*98*	*98*	*77*	*65*	*66*	*109*	*316*

- The columns are ordered by the ascending, formatted values of the across variable. The default order can be changed with the DESCENDING and ORDER= options in the DEFINE statement.
- The N statistic (number of observations) is the default statistic for the across variable.

100

ACROSS Usage

```
proc report data=orion.emps nowd;
   column Country;
   define Country / across;
run;
```

Employee Country									
Australia	Belgium	Denmark	France	Germany	Italy	Netherlands	Spain	United Kingdom	United States
108	60	51	98	98	77	65	66	109	316

```
proc report data=orion.emps nowd;
   column Gender Country;
   define Gender / group;
   define Country / across;
run;
```

	Employee Country									
Employee Gender	Australia	Belgium	Denmark	France	Germany	Italy	Netherlands	Spain	United Kingdom	United States
Female	48	27	23	44	45	34	30	29	60	143
Male	60	33	28	54	53	43	35	37	59	173

r104d09

101

Comma Operator

The comma can be used to stack an analysis variable with an across variable.

```
proc report data=orion.emps nowd;
   column Gender Country,Salary;
   define Gender / group;
   define Country / across;
   define Salary / analysis sum;
run;
```

	Employee Country									
	Australia	Belgium	Denmark	France	Germany	Italy	Netherlands	Spain	United Kingdom	United States
Employee Gender	Annual Salary	Annual Salary	Annual Salary	Annual Salary	Annual Salary	Annual Salary	Annual Salary	Annual Salary	Annual Salary	Annual Salary
Female	$1,342,140	$873,240	$640,700	$1,268,696	$1,632,700	$1,007,685	$926,045	$967,240	$1,411,985	$5,456,800
Male	$1,997,325	$981,895	$971,425	$1,843,735	$1,460,170	$1,332,150	$1,050,600	$1,084,955	$1,917,965	$7,334,335

r104d09

102

Streamlined Column Headings

Specifying the analysis variable before the across variable enables streamlined column headings.

```
proc report data=orion.emps nowd;
   column Gender Salary,Country;
   define Gender / group;
   define Country / across;
   define Salary / analysis sum ' ';
run;
```

	Employee Country									
Employee Gender	Australia	Belgium	Denmark	France	Germany	Italy	Netherlands	Spain	United Kingdom	United States
Female	$1,342,140	$873,240	$640,700	$1,268,595	$1,532,700	$1,007,585	$925,045	$967,240	$1,411,985	$5,455,800
Male	$1,997,325	$981,895	$971,425	$1,843,735	$1,450,170	$1,332,160	$1,090,600	$1,084,955	$1,917,965	$7,334,335

106 r104d09

Parentheses

Parentheses can be used to group analysis variables with an across variable.

```
proc report data=orion.emps nowd;
   column Gender Country,(Salary HireDate);
   define Gender / group;
   define Country / across;
   define Salary / mean 'Annual Salary Avg';
   define HireDate / max f=year4.
                     'Hire Year Max';
run;
```

	Employee Country																	
	Australia		Belgium		Denmark		France		Germany		Italy		Netherlands		Spain		U K	
Employee Gender	Annual Salary Avg	Hire Year Max	Annual Salary Avg	Hire Year Max	Annual Salary Avg	Hire Year Max	Annual Salary Avg	Hire Year Max	Annual Salary Avg	Hire Year Max	Annual Salary Avg	Hire Year Max	Annual Salary Avg	Hire Year Max	Annual Salary Avg	Hire Year Max	Annual Sal Av	
Female	$27,961	2010	$32,342	2011	$27,857	2010	$28,832	2011	$34,060	2010	$29,635	2011	$30,835	2010	$33,353	2010	$28,2	
Male	$33,289	2010	$29,754	2010	$34,694	2011	$34,143	2011	$27,362	2011	$30,980	2011	$31,160	2010	$29,323	2010	$32,2	

107

Summary Reports

Computed variables can be included in summary reports.

```
proc report data=orion.emps nowd;
   column Country Department
          Salary Bonus;
   define Country / group;
   define Department / group;
   define Salary / mean 'Average Salary';
   define Bonus / computed f=dollar12.
                  'Average Bonus';
   compute Bonus;
      Bonus = Salary.mean * 0.05;
   endcomp;
run;
```

Employee Country	Department	Average Salary	Average Bonus
Australia	Administration	$29,300	$1,465
	Engineering	$29,416	$1,471
	Sales	$27,416	$1,371
	Sales Management	$119,757	$5,988
	Stock & Shipping	$34,321	$1,716
Belgium	Administration	$29,475	$1,474

r104d12

123

Character Computed Variable

A computed variable can be based on other variables.

```
proc report data=orion.emps nowd;
   column Country Department Name
        (1) FirstName HireDate Salary;
   define Country / order;
   define Department / order;
   define Name / display;
(2) define FirstName / computed;
   define HireDate / display;
   define Salary / analysis sum;
   compute FirstName / character length=20;
(3)    firstName = scan(Name,1,' ');
   endcomp;
run;
```

Employee Country	Department	Employee Name	FirstName
Australia	Administration	Kareen Billington	Kareen
		Liz Povey	Liz
		John Hornsey	John
		Sherie Shrestha	Sherie

117

Column Numbers

A column number needs to be used to refer to an across variable column in a compute block.

	Employee Gender				
	Female		Male		
Employee Country	Avg	Count	Avg	Count	AvgDiff
Australia	$27,961	48	$33,289	60	$-5,328
Belgium	$32,342	27	$29,754	33	$2,588
Denmark	$27,857	23	$34,694	28	$-6,837
France	$28,832	44	$34,143	54	$-5,312

_C1_ _C2_ _C3_ _C4_ _C5_ _C6_

A column number follows the form of **_C*n*_** where ***n*** is the number of the column from left to right in the report.

127

Column Numbers

```
proc report data=orion.emps nowd;
   column Country
          Gender,(Salary Salary=SalCnt)
          AvgDiff;
   define Country / group;
   define Gender / across;
   define Salary / mean 'Avg';
   define SalCnt / n 'Count' format=6.;
   define AvgDiff / computed format=dollar10.;
   compute AvgDiff;
      AvgDiff = _C2_ - _C4_;
   endcomp;
run;
```

	Employee Gender				
	Female		Male		
Employee Country	Avg	Count	Avg	Count	AvgDiff
Australia	$27,961	48	$33,289	60	$-5,328
Belgium	$32,342	27	$29,754	33	$2,588
Denmark	$27,857	23	$34,694	28	$-6,837

128

r104d12

7. The following step creates the desired report.

- ❍ True
- ❍ False

```
proc report data=orion.orders nowd;
   column CustomerAgeGroup
          Quantity=QMean Quantity=QSum;
   define CustomerAgeGroup / group 'Group';
   define QMean / mean 'Average' format=6.1;
   define QSum / 'Total' format=6.;
run;
```

Group	Average	Total
15-30 years	1.7	685
31-45 years	1.8	568
46-60 years	1.8	874
61-75 years	1.7	445

149

Output Delivery System (ODS)

Use the Output Delivery System to perform the following tasks:

- alter the appearance of the output
  - style, color, font, and so on
- change the destination of the output
- select output
- create new data tables from PROC step output

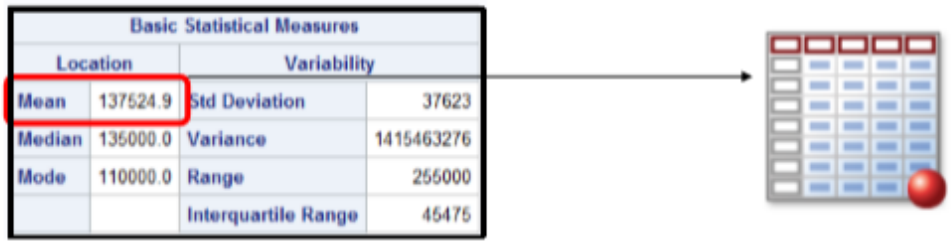
- Use ODS SELECT to request specific output.
- Use ODS OUTPUT to create a new SAS data table.
- Use the OUTPUT statement to customize the creation of a new SAS data table.

Output Delivery System (ODS)

Use the OUTPUT statement to customize the new data table.

```
proc univariate data=ameshousing;
   var saleprice;
   output out=stats mean=sp_mean;
run;
```

OUTPUT OUT=*new-data-set-name*
keyword-1 = variable-name-1 … keyword-n = variable-name-n;



Basic Statistical Measures

Location		Variability	
Mean	137524.9	Std Deviation	37623
Median	135000.0	Variance	1415463276
Mode	110000.0	Range	255000
		Interquartile Range	45475

You can specify multiple OUTPUT statements to create multiple data tables for individual values to place in the new tables [SAS Documentation].

ODS provides various formatting options that are not available when using individual procedures or the DATA step without ODS enhancements.

Destinations Used with Excel

- The CSVALL destination creates a CSV (comma-separated value) file.
- The EXCELXP destination creates an XML (Extensible Markup Language) file.
- The MSOFFICE2K destination creates an HTML (HyperText Markup Language) file.

```
ods csvall file='shoes.csv';
ods tagsets.excelxp file='shoes.xml';
ods msoffice2k file='shoes.html';

proc print data=sashelp.shoes;
 ...
run;

ods _all_ close;
```

127

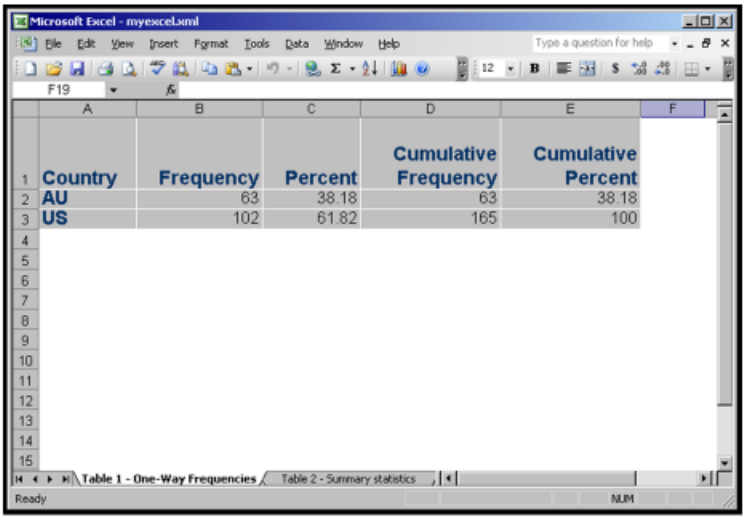
Destinations Used with Excel

Destination	Type of File	Extension	Viewed In
CSVALL	Comma-Separated Value	csv	Editor or Microsoft Excel
MSOFFICE2K	Hypertext Markup Language	html	Web browser or Microsoft Word or Microsoft Excel
TAGSETS. EXCELXP	Extensible Markup Language	xml	Microsoft Excel

70

EXCELXP Destination

EXCELXP keeps the style information. Output from each procedure is on a separate sheet.



Country	Frequency	Percent	Cumulative Frequency	Cumulative Percent
AU	63	38.18	63	38.18
US	102	61.82	165	100

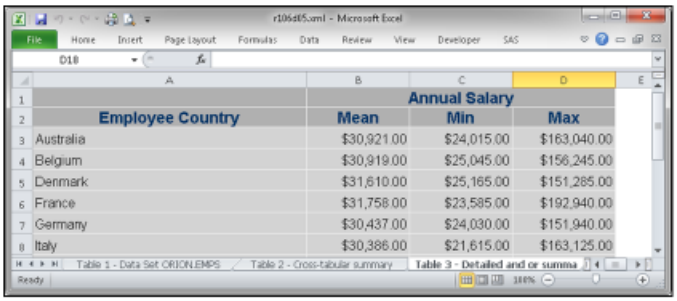
74

TAGSETS.EXCELXP Destination

The EXCELXP destination is based on SpreadsheetML, which is an XML dialect developed by Microsoft to represent information in an Excel workbook.

```
ods tagsets.excelxp file="&path2\r106d05.xml";
```

When the XML file is opened in Excel, each worksheet represents a different output table.



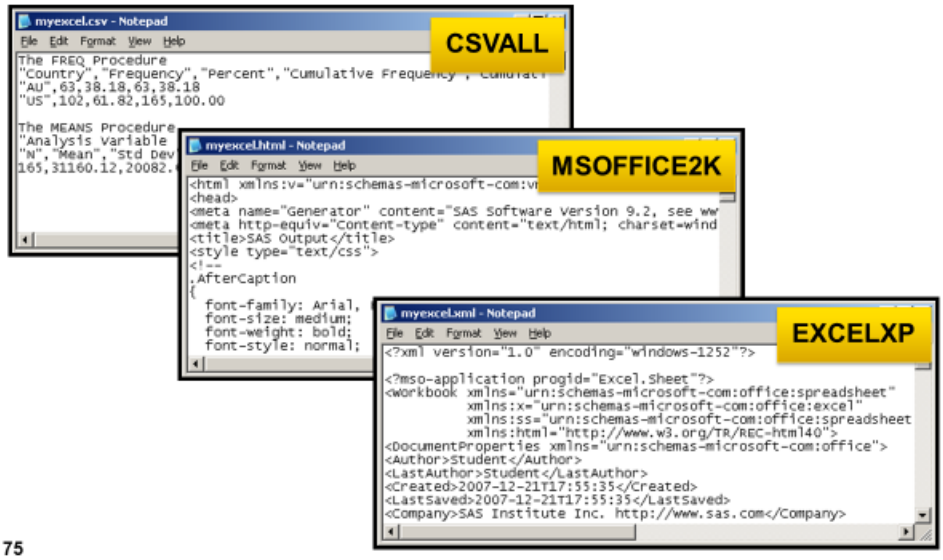
Employee Country	Annual Salary Mean	Min	Max
Australia	$30,921.00	$24,015.00	$163,040.00
Belgium	$30,919.00	$25,045.00	$156,245.00
Denmark	$31,610.00	$25,165.00	$151,285.00
France	$31,758.00	$23,585.00	$192,940.00
Germany	$30,437.00	$24,030.00	$151,940.00
Italy	$30,386.00	$21,615.00	$163,125.00

37

r106d05

Keep in Mind

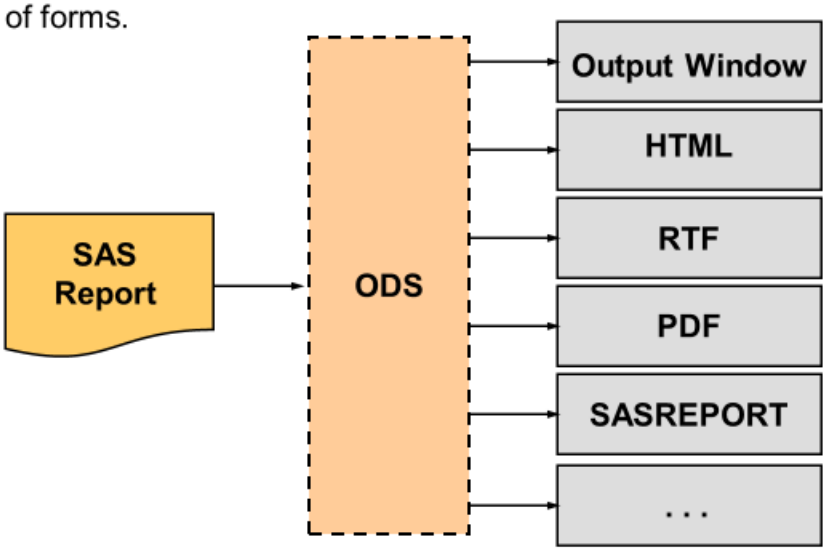
The file you are creating is not an Excel file.



75

Output Delivery System (ODS)

ODS statements enable you to create output in a variety of forms.



113

ODS RTF and PDF Statements

- The *ODS RTF statement* opens or closes the RTF destination, which produces Rich Text Format files that are viewable with a word processor.
- The *ODS PDF statement* opens or closes the PDF destination, which produces Portable Document Format files that are viewable with an Adobe product.

```
ods rtf file='shoes.rtf';
ods pdf file='shoes.pdf';

proc freq data=sashelp.shoes;
   ...
run;

ods rtf close;
ods pdf close;
```

120

Multiple ODS Destinations

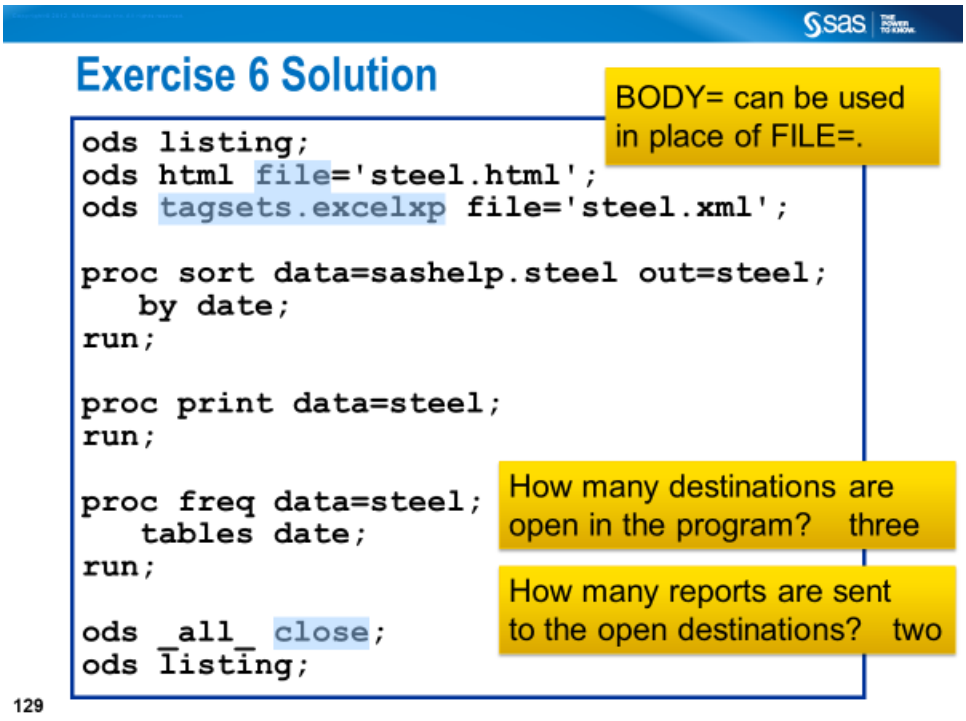
The ODS _ALL_ CLOSE statement closes ***all*** open destinations including the LISTING destination.

```
ods listing;
ods html file='shoes.html';
ods rtf file='shoes.rtf';
ods pdf file='shoes.pdf';

proc print data=sashelp.shoes;
  ...
run;

ods _all_ close;
ods listing;
```

124



Review Questions

1. What does PROC DATASETS do?
2. What is the common output that both PROC CONTENTS and PROC DATASETS can generate?
3. What are the common tasks PROC SQL can perform?
4. What do the DICTIONARY TABLES in PROC SQL contain?
5. What is the purpose of using the NOTSORTED option in PROC SORT?
6. What are the common statistics that both PROC REPORT and PROC TABULATE can generate?
7. What are limitations of PROC TABULATE?
8. What does the PRELOADFMT option, in PROC REPORT and PROC TABULATE display?
9. By default, what do PROC SUMMARY and PROC MEANS produce?
10. Why is the purpose of using ODS TRACE ON and ODS TRACE OFF as a wrapper to a PROC?
11. When would you use the following ODS statements?
    ODS SELECT …;
    ODS EXCLUDE …;
    ODS HTML (LISTING or EXCEL) …;