# Chapter 11: Creating Summary Reports

# Chapter 11: Creating Summary Reports

# Objectives

- Produce one-way and two-way frequency tables with the FREQ procedure.
- Enhance frequency tables with options.
- Use PROC FREQ to validate data in a SAS data set.

# Business Scenario

Orion Star management wants to know the number of male and female sales employees in Australia.

# Considerations

Use the FREQ procedure to analyze the **Gender** variable in a subset of **orion.sales**.

**The FREQ Procedure**

| Gender | Frequency | Percent |
|---|---|---|
| F | XX | XX.XX |
| M | XX | XX.XX |

# FREQ Procedure

The FREQ procedure produces a one-way frequency table for each variable named in the TABLES statement.

```
proc freq data=orion.sales;
   tables Gender;
   where Country='AU';
run;
```

**PROC FREQ DATA=***SAS-data-set*;
    <**TABLES** *variable(s) </ options>>* ;
**RUN**;

If the TABLES statement is omitted, a one-way frequency table is produced for ***every*** variable in the data set. This can produce voluminous output and is seldom desired.

# Viewing the Output

A one-way frequency table was created for **Gender**. It lists the discrete values found in the data set and the number of observations in which the variable has that value.

**The FREQ Procedure**

| Gender | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---|---|---|---|---|
| F | 27 | 42.86 | 27 | 42.86 |
| M | 36 | 57.14 | 63 | 100.00 |

The default output includes frequency and percentage values, including cumulative statistics.

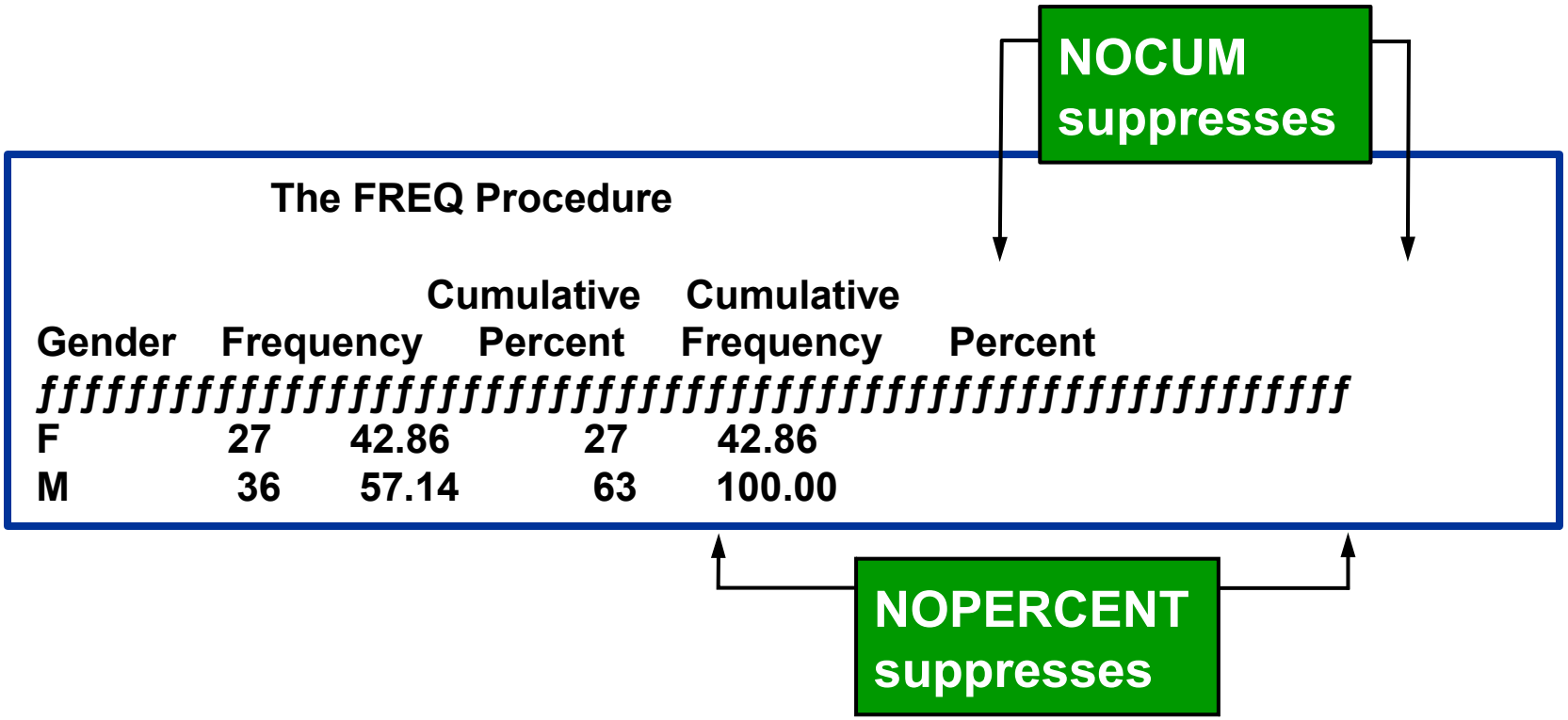
# Options to Suppress Statistics

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

**TABLES** *variable(s) / options* **;**

| Option | Description |
|---|---|
| NOCUM | Suppresses the cumulative statistics. |
| NOPERCENT | Suppresses the percentage display. |

# Options to Suppress Statistics



# 11.01 Quiz

Open and submit **p111a01**. Review the log to determine the cause of the error. Correct the program and resubmit.

What change was needed?

```
proc freq data=orion.sales;
   tables country nocum nopercent;
run;
```

# 11.01 Quiz – Correct Answer

What change was needed? **A slash is required before the options in the TABLES statement.**

```
31   proc freq data=orion.sales;
32      tables country nocum nopercent;
ERROR: Variable NOCUM not found.
ERROR: Variable NOPERCENT not found.
33   run;
```

```
proc freq data=orion.sales;
   tables country / nocum nopercent;
run;
```

```
 The FREQ Procedure
Country    Frequency
ffffffffffffffffff
AU              63
US             102
```

p111a01s

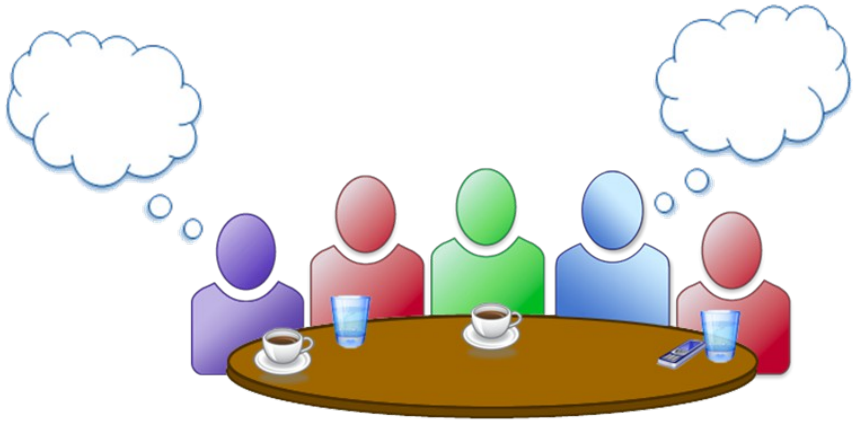
# Idea Exchange

This step creates a table for every variable in the data set:

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

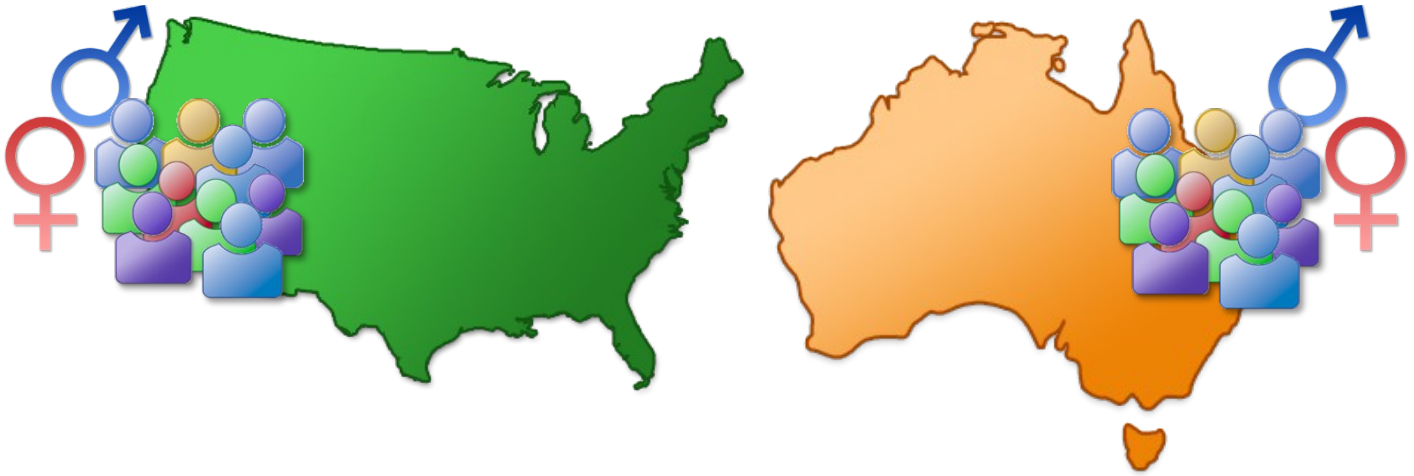
- **Employee_ID**
- **First_Name**
- **Last_Name**
- **Gender**
- **Salary**
- **Job_Title**
- **Country**
- **Birth_Date**
- **Hire_Date**

Which variables are most appropriate for a frequency analysis? Why?



# Business Scenario

Orion Star management wants to know how many sales employees are in each country, as well as the count of males and females.



# TABLES Statement

You can list multiple variables in a TABLES statement. A separate table is produced for each variable.

```
proc freq data=orion.sales;
    tables Gender Country;
run;
```

PROC FREQ Output

The FREQ Procedure

| Gender | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---|---|---|---|---|
| F | 68 | 41.21 | 68 | 41.21 |
| M | 97 | 58.79 | 165 | 100.00 |

| Country | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---|---|---|---|---|
| AU | 63 | 38.18 | 63 | 38.18 |
| US | 102 | 61.82 | 165 | 100.00 |

p111d02

# BY Statement

The BY statement is used to request separate analyses for each BY group.

```
proc sort data=orion.sales out=sorted;
   by Country;
run;

proc freq data=sorted;
   tables Gender;
   by Country;
run;
```

The data set must be sorted or indexed by the variable (or variables) named in the BY statement.

# Viewing the Output

Each group displays on a separate page with a BY line.

```
------------------------ Country=AU ----------------------

                The FREQ Procedure

                              Cumulative    Cumulative
 Gender    Frequency     Percent     Frequency      Percent
 ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
 F              27      42.86           27       42.86
 M              36      57.14           63      100.00
```

```
------------------------ Country=US ----------------------

                The FREQ Procedure

                              Cumulative    Cumulative
 Gender    Frequency     Percent     Frequency      Percent
 ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
 F              41      40.20           41       40.20
 M              61      59.80          102      100.00
```

# Crosstabulation Table

An asterisk between two variables generates a two-way frequency table, or *crosstabulation table*.

```
proc freq data=orion.sales;
   tables Gender*Country;
run;
```

rows

columns

A two-way frequency table generates a single table with statistics for each distinct combination of values of the selected variables.

# Viewing the Output

## PROC FREQ Output

```
      The FREQ Procedure

    Table of Gender by Country

Gender     Country

Frequency,
Percent  ,
Row Pct  ,
Col Pct  ,AU      ,US      ,  Total
ƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒˆ
F        ,    27 ,    41 ,    68
         ,  16.36 ,  24.85 ,  41.21
         ,  39.71 ,  60.29 ,
         ,  42.86 ,  40.20 ,
ƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒˆ
M        ,    36 ,    61 ,    97
         ,  21.82 ,  36.97 ,  58.79
         ,  37.11 ,  62.89 ,
         ,  57.14 ,  59.80 ,
ƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒˆ
Total         63     102     165
           38.18   61.82   100.00
```

# 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. |

# Options to Suppress Statistics

Table of Gender by Country

```
Gender     Country

Frequency,
Percent   ,
Row Pct   ,
Col Pct   ,AU       ,US       ,  Total
ƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒˆ
F         ,     27 ,     41 ,     68
          ,  16.36 ,  24.85 ,  41.21
          ,  39.71 ,  60.29 ,
          ,  42.86 ,  40.20 ,
ƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒˆ
M         ,     36 ,     61 ,     97
          ,  21.82 ,  36.97 ,  58.79
          ,  37.11 ,  62.89 ,
          ,  57.14 ,  59.80 ,
ƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒˆ
Total          63      102      165
            38.18    61.82   100.00
```

NOFREQ suppresses

NOPERCENT suppresses

# Options to Suppress Statistics

Table of Gender by Country

| Gender | Country | | |
|---|---|---|---|
| Frequency, Percent, Row Pct, Col Pct | AU | US | Total |
| F | 27 | 41 | 68 |
| | 16.36 | 24.85 | 41.21 |
| | 39.71 | 60.29 | |
| | 42.86 | 40.20 | |
| M | 36 | 61 | 97 |
| | 21.82 | 36.97 | 58.79 |
| | 37.11 | 62.89 | |
| | 57.14 | 59.80 | |
| Total | 63 | 102 | 165 |
| | 38.18 | 61.82 | 100.00 |

NOROW suppresses

NOCOL suppresses

# LIST and CROSSLIST Options

You can use the LIST and CROSSLIST options in the TABLES statement to "flatten" the output.

LIST option

| Gender | Country | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---|---|---|---|---|---|
| F | AU | 27 | 16.36 | 27 | 16.36 |
| F | US | 41 | 24.85 | 68 | 41.21 |
| M | AU | 36 | 21.82 | 104 | 63.03 |
| M | US | 61 | 3 | | |

CROSSLIST option

Table of Gender by Country

| Gender | Country | Frequency | Percent | Row Percent | Column Percent |
|---|---|---|---|---|---|
| F | AU | 27 | 16.36 | 39.71 | 42.86 |
| | US | 41 | 24.85 | 60.29 | 40.20 |
| | Total | 68 | 41.21 | 100.00 | |
| M | AU | 36 | 21.82 | 37.11 | 57.14 |
| | US | 61 | 36.97 | 62.89 | 59.80 |
| | Total | 97 | 58.79 | 100.00 | |
| Total | AU | 63 | 38.18 | 100.00 | |
| | US | 102 | 61.82 | 100.00 | |
| | Total | 165 | 100.00 | | |

# Business Scenario

A new data set, **orion.nonsales2**, must be validated. It contains information on non-sales employees and might include invalid and missing values.

Partial **orion.nonsales2**

| Employee_ ID | First | Last | Gender | Salary | Job_Title | Country |
|---|---|---|---|---|---|---|
| 120101 | Patrick | Lu | M | 163040 | Director | AU |
| 120104 | Kareen | Billington | F | 46230 | Admin Mgr | au |
| 120105 | Liz | Povey | F | 27110 | Secretary I | AU |
| 120106 | John | Hornsey | M | . | Office Asst II | AU |
| 120107 | Sherie | Sheedy | F | 30475 | Office Asst II | AU |
| 120108 | Gladys | Gromek | F | 27660 | Warehouse Asst II | AU |

# Considerations

Use the FREQ procedure to screen for invalid, missing, and duplicate data values.

Requirements of non-sales employee data:

- **Employee_ID** values must be unique and not missing.
- **Gender** must be *F* or *M*.
- **Job_Title** must not be missing.
- **Country** must have a value of *AU* or *US*.
- **Salary** values must be in the numeric range of 24000 to 500000.

# 11.02 Quiz

What problems exist with the data in this partial data set?

| Employee_ID | First | Last | Gender | Salary | Job_Title | Country |
|---|---|---|---|---|---|---|
| 120101 | Patrick | Lu | M | 163040 | Director | AU |
| 120104 | Kareen | Billington | F | 46230 | Administration Manager | au |
| 120105 | Liz | Povey | F | 27110 | Secretary I | AU |
| 120106 | John | Hornsey | M | . | Office Assistant II | AU |
| 120107 | Sherie | Sheedy | F | 30475 | Office Assistant III | AU |
| 120108 | Gladys | Gromek | F | 27660 | Warehouse Assistant II | AU |
| 120108 | Gabriele | Baker | F | 26495 | Warehouse Assistant I | AU |
| 120110 | Dennis | Entwisle | M | 28615 | Warehouse Assistant III | AU |
| 120111 | Ubaldo | Spillane | M | 26895 | Security Guard II | AU |
| 120112 | Ellis | Glattback | F | 26550 | | AU |
| 120113 | Riu | Horsey | F | 26870 | Security Guard II | AU |
| 120114 | Jeannette | Buddery | G | 31285 | Security Manager | AU |
| 120115 | Hugh | Nichollas | M | 2650 | Service Assistant I | AU |
| . | Austen | Ralston | M | 29250 | Service Assistant II | AU |
| 120117 | Bill | Mccleary | M | 31670 | Cabinet Maker III | AU |
| 120118 | Darshi | Hartshorn | M | 28090 | Cabinet Maker II | AU |

Hint: There are seven data problems.

# 11.02 Quiz – Correct Answer

What problems exist with the data in this partial data set?

| Employee_ID | First | Last | Gender | Salary | Job_Title | Country |
|---|---|---|---|---|---|---|
| 120101 | Patrick | Lu | M | 163040 | Director | AU |
| 120104 | Kareen | Billington | F | 46230 | Administration Manager | au |
| 120105 | Liz | Povey | F | 27110 | Secretary I | AU |
| 120106 | John | Hornsey | M | . | Office Assistant II | AU |
| 120107 | Sherie | Sheedy | F | 30475 | Office Assistant III | AU |
| 120108 | Gladys | Gromek | F | 27660 | Warehouse Assistant II | AU |
| 120108 | Gabriele | Baker | F | 26495 | Warehouse Assistant I | AU |
| 120110 | Dennis | Entwisle | M | 28615 | Warehouse Assistant III | AU |
| 120111 | Ubaldo | Spillane | M | 26895 | Security Guard II | AU |
| 120112 | Ellis | Glattback | F | 26550 | | AU |
| 120113 | Riu | Horsey | F | 26870 | Security Guard II | AU |
| 120114 | Jeannette | Buddery | G | 31285 | Security Manager | AU |
| 120115 | Hugh | Nichollas | M | 2650 | Service Assistant I | AU |
| . | Austen | Ralston | M | 29250 | Service Assistant II | AU |
| 120117 | Bill | Mccleary | M | 31670 | Cabinet Maker III | AU |
| 120118 | Darshi | Hartshorn | M | 28090 | Cabinet Maker II | AU |

Hint: There are seven data problems.

# FREQ Procedure for Data Validation

The FREQ procedure lists all discrete values for a variable and reports missing values.

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

p111d03

# Viewing the Output

## PROC FREQ Output

**The FREQ Procedure**

| Gender | Frequency |
|---|---|
| F | 110 |
| G | 1 |
| M | 123 |

**Frequency Missing = 1**

| Country | Frequency |
|---|---|
| AU | 33 |
| US | 196 |
| au | 3 |
| us | 3 |

# 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;**

# Viewing the Output

## PROC FREQ Output

The FREQ Procedure

Number of Variable Levels

| Variable | Missing Levels | Nonmissing Levels | 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 |

# Check for Uniqueness

The values of **Employee_ID** must be unique and not missing. PROC FREQ can be used to check for duplicate or missing values.

```
proc freq data=orion.nonsales2 order=freq;
   tables Employee_ID / nocum nopercent;
run;
```

The ORDER=FREQ option displays the results in descending frequency order.

# Viewing the Output

## Partial PROC FREQ Output

**The FREQ Procedure**

| Employee_ID | Frequency |
|---|---|
| 120108 | 2 |
| 120101 | 1 |
| 120104 | 1 |
| 120105 | 1 |
| 120106 | 1 |
| 121141 | 1 |
| 121142 | 1 |
| 121146 | 1 |
| 121147 | 1 |
| 121148 | 1 |

**Frequency Missing = 1**

p111d04

# NLEVELS Option

NLEVELS can also be used to identify duplicates, when the number of distinct values is known.

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

This example uses the NOPRINT option to suppress the frequency table. Only the Number of Variable Levels table is displayed.

# Viewing the Output

## Partial PROC FREQ Output

**The FREQ Procedure**

**Number of Variable Levels**

| Variable | Missing Levels | Nonmissing Levels | Levels |
|---|---|---|---|
| Employee_ID | 234 | 1 | 233 |

There are 235 employees, but there are only 234 distinct **Employee_ID** values. Therefore, there is one duplicate value and one missing value for **Employee_ID**.

# 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

# Viewing the Output

## PROC FREQ Output

**The FREQ Procedure**
**Number of Variable Levels**

| Variable | Missing Levels | Nonmissing Levels | 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.

# 11.03 Quiz

Modify **p111a02** to analyze **Job_Title**. Display the NLEVELS table listing the frequency counts in decreasing order.

How many unique, nonmissing job titles exist?

Which job title occurs most frequently?

What is the frequency of missing job titles?

# 11.03 Quiz – Correct Answer

How many unique, nonmissing job titles exist? **124**

Which job title occurs most frequently? **Trainee**

What is the frequency of missing job titles? **1**

```
proc freq data=orion.nonsales2 nlevels
          order=freq;
   tables Job_Title /nocum nopercent;
run;
```

p111a02s

# Identifying Observations with Invalid Data

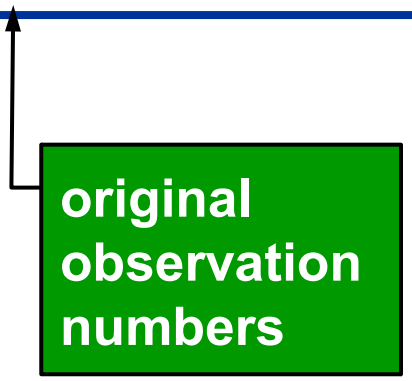
PROC FREQ has uncovered the existence of invalid data values for **Gender**, **Country**, and **Employee_ID**. Use PROC PRINT to display the observations with invalid values.

```
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;
```

# 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 |



# Business Scenario

The manager of Human Resources has requested a report showing the number and percent of sales employees hired each year.

# 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

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

p111d05

# 11.04 Quiz

Open and submit **p111a03** and view the output. Add a statement to apply the TIERS format to **Salary** and resubmit.

Can user-defined formats be used to group data?

# 11.04 Quiz – Correct Answer

Can user-defined formats be used to group data? **yes**

The FREQ Procedure

| Salary | Frequency | Cumulative Percent | Cumulative Frequency | Percent |
|---|---|---|---|---|
| Tier1 | 1 | 0.61 | 1 | 0.61 |
| Tier2 | 158 | 95.76 | 159 | 96.36 |
| Tier3 | 4 | 2.42 | 163 | 98.79 |
| Tier4 | 2 | 1.21 | 165 | 100.00 |

p111a03s

# FORMAT Statement

User-defined formats can also be used to display levels with alternate text in a frequency table.

```
proc freq data=orion.sales;
   tables Gender*Country;
   format Country $ctryfmt.
          Gender $gender.;
run;
```

p111d06

# Viewing the Output

## Partial PROC FREQ Output

**Table of Gender by Country**

```
Gender     Country

Frequency,
Percent  ,
Row Pct  ,
Col Pct  ,Australi,United S,  Total
         ,a       ,tates    ,
ƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒˆ
Female   ,     27 ,     41 ,     68
         ,  16.36 ,  24.85 ,  41.21
         ,  39.71 ,  60.29 ,
         ,  42.86 ,  40.20 ,
ƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒˆ
Male     ,     36 ,     61 ,     97
         ,  21.82 ,  36.97 ,  58.79
         ,  37.11 ,  62.89 ,
         ,  57.14 ,  59.80 ,
ƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒˆ
Total          63      102      165
            38.18    61.82   100.00
```

**Labels are wrapped.**

# FORMAT= Option

Use the *FORMAT= option* in the TABLES statement to format the frequency value and to change the width of the column.

```
proc freq data=orion.sales;
   tables Gender*Country / format=12.;
   format Country $ctryfmt.
          Gender $gender.;
run;
```

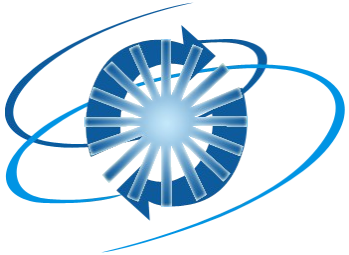
p111d06

# Viewing the Output

## PROC FREQ Output

```
          The FREQ Procedure
        Table of Gender by Country

Gender     Country

Frequency,
Percent  ,
Row Pct  ,
Col Pct  ,Australia   ,United States,      Total
ƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆ
Female   ,       27 ,       41 ,       68
         ,    16.36 ,    24.85 ,    41.21
         ,    39.71 ,    60.29 ,
         ,    42.86 ,    40.20 ,
ƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆ
Male     ,       36 ,       61 ,       97
         ,    21.82 ,    36.97 ,    58.79
         ,    37.11 ,    62.89 ,
         ,    57.14 ,    59.80 ,
ƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒˆ
Total          63         102         165
               38.18      61.82      100.00
```

Columns are 12 characters wide.



# Exercise

This exercise reinforces the concepts discussed previously.

# Chapter 11: Creating Summary Reports

11.1 The FREQ Procedure

11.2 The MEANS and UNIVARIATE Procedures

11.3 Using the Output Delivery System

# Objectives

- Calculate summary statistics and multilevel summaries with the MEANS procedure.
- Enhance summary tables with options.
- Identify extreme and missing values with the UNIVARIATE procedure.

# Business Scenario

The payroll manager would like to see the average salary for all employees.

# 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.

# Viewing the Output

## PROC MEANS Output

The MEANS Procedure

| Variable | N | Mean | Std Dev | Minimum | Maximum |
|---|---|---|---|---|---|
| Employee_ID | 165 | 120713.90 | 450.0866939 | 120102.00 | 121145.00 |
| Salary | 165 | 31160.12 | 20082.67 | 22710.00 | 243190.00 |
| Birth_Date | 165 | 3622.58 | 5456.29 | -5842.00 | 10490.00 |
| Hire_Date | 165 | 12054.28 | 4619.94 | 5114.00 | 17167.00 |

Default statistics are displayed for all numeric variables.

# 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**

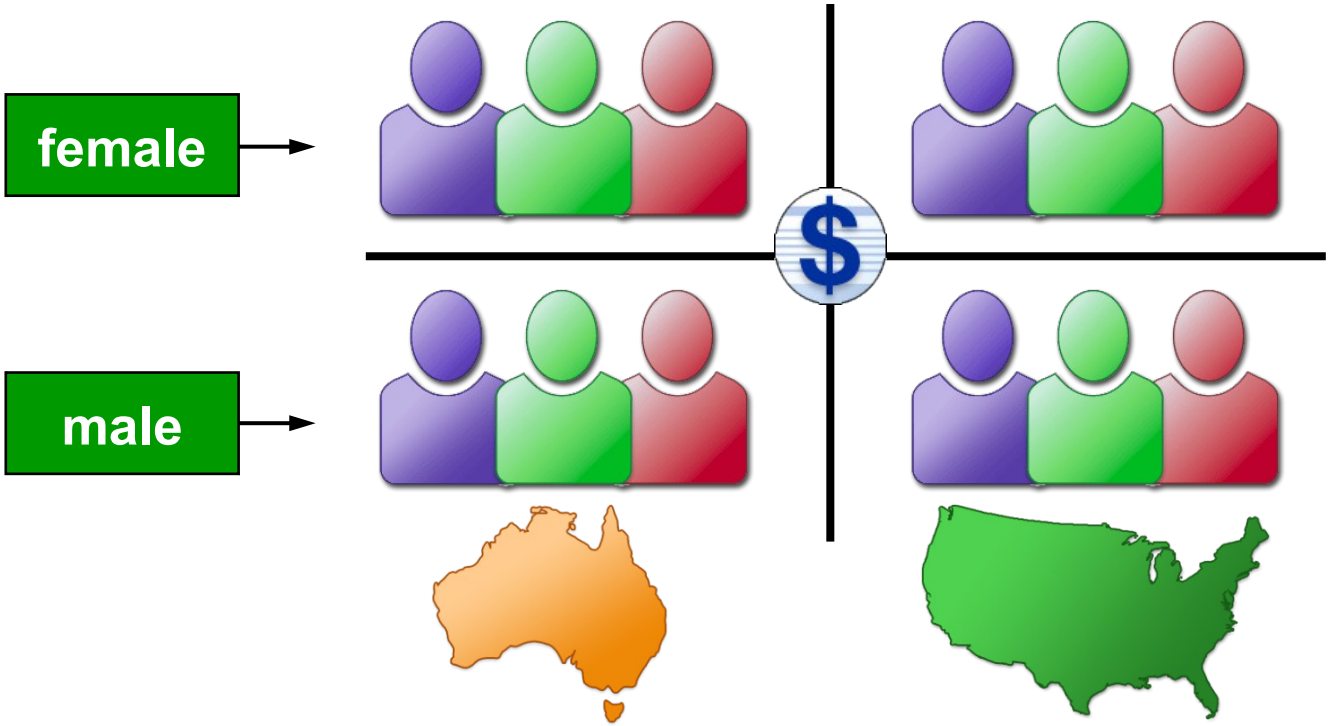
| N | Mean | Std Dev | Minimum | Maximum |
|---|---|---|---|---|
| 165 | 31160.12 | 20082.67 | 22710.00 | 243190.00 |



p111d07

# Business Scenario

Analyze **Salary** by **Country** within **Gender**.



# 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.

# 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)

# 11.05 Quiz

For a given data set, there are 63 observations with a **Country** value of *AU*. Of those 63 observations, only 61 observations have a value for **Salary**.

Which output is correct?

a.

```
Analysis Variable : Salary

                 N
   Country     Obs        N
   ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
   AU           63       61
   ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
```

b.

```
Analysis Variable : Salary

                 N
   Country     Obs        N
   ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
   AU           61       63
   ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
```

# 11.05 Quiz – Correct Answer

For a given data set, there are 63 observations with a **Country** value of *AU*. Of those 63 observations, only 61 observations have a value for **Salary**.

Which output is correct?

a.

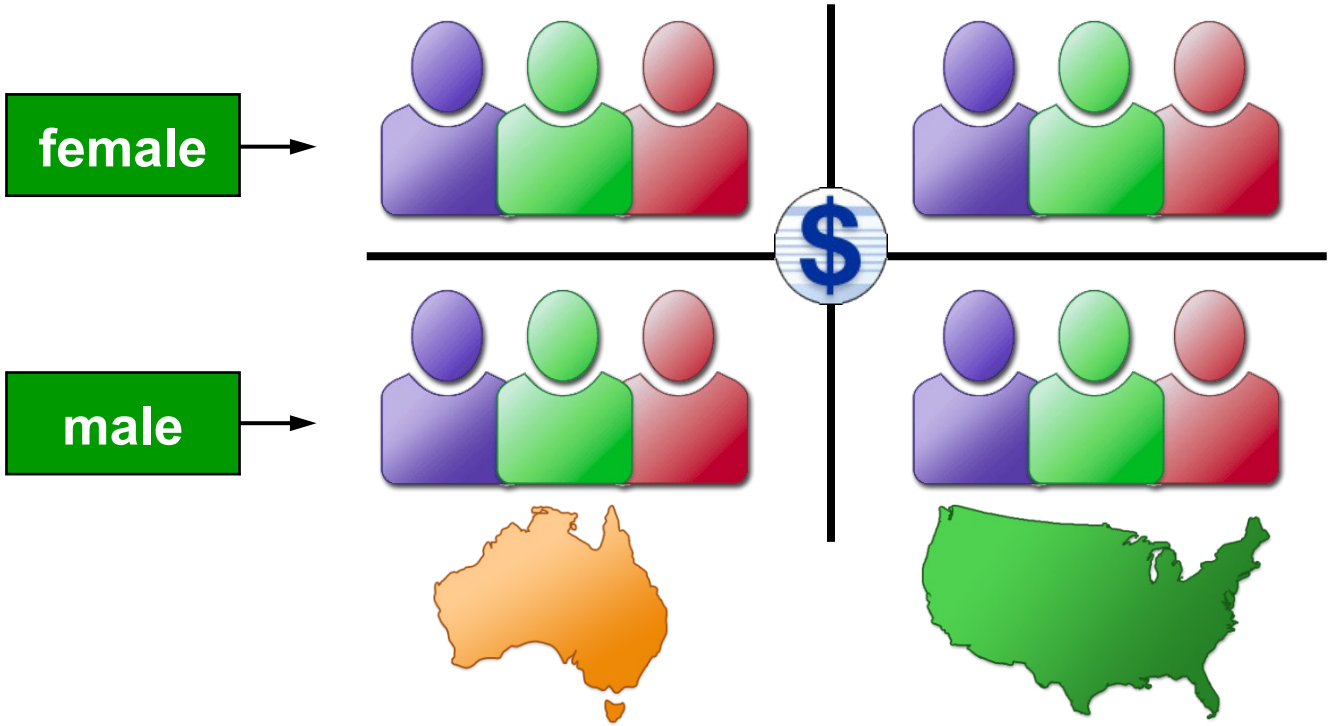
```
Analysis Variable : Salary

                N
Country      Obs        N
ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
AU            63       61
ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
```

b.

```
Analysis Variable : Salary

                N
Country      Obs        N
ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
AU            61       63
ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
```

# Business Scenario

Analyze **Salary** by **Country** within **Gender**. Generate a report that includes the number of missing **Salary** values, as well as the minimum, maximum, and sum of salaries.



# 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.

# PROC MEANS Statistics

The statistics are displayed in the order in which they are requested.

The MEANS Procedure

Analysis Variable : Salary

| Gender | Country | N Obs | N Miss | Minimum | Maximum | Sum |
|---|---|---|---|---|---|---|
| F | AU | 27 | 0 | 25185.00 | 30890.00 | 747965.00 |
| | US | 41 | 0 | 25390.00 | 83505.00 | 1207900.00 |
| M | AU | 36 | 0 | 25745.00 | 108255.00 | 1152050.00 |
| | US | 61 | 0 | 22710.00 | 243190.00 | 2033505.00 |

# 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. |

# MAXDEC= Option

MAXDEC=0

```
                          The MEANS Procedure
                        Analysis Variable : Salary

                N
Country      Obs      N            Mean         Std Dev         Minimum         Maximum
ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff
AU           63      63           30159           12699           25185          108255

US          102     102           31778           23556           22710          243190
```

MAXDEC=1

```
                          The MEANS Procedure
                        Analysis Variable : Salary

                N
Country      Obs      N            Mean         Std Dev         Minimum         Maximum
ffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffffff
AU           63      63         30159.0         12699.1         25185.0        108255.0

US          102     102         31778.5         23555.8         22710.0        243190.0
```

p111d10

# NONOBS Option

**N Obs included by default**

```
                          The MEANS Procedure
                        Analysis Variable : Salary

              N
Country    Obs      N            Mean         Std Dev         Minimum         Maximum
ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
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
 ƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒƒ
 AU            63        30158.97        12699.14        25185.00       108255.00

 US           102        31778.48        23555.84        22710.00       243190.00
```



p111d10

# Other PROC MEANS Statistics

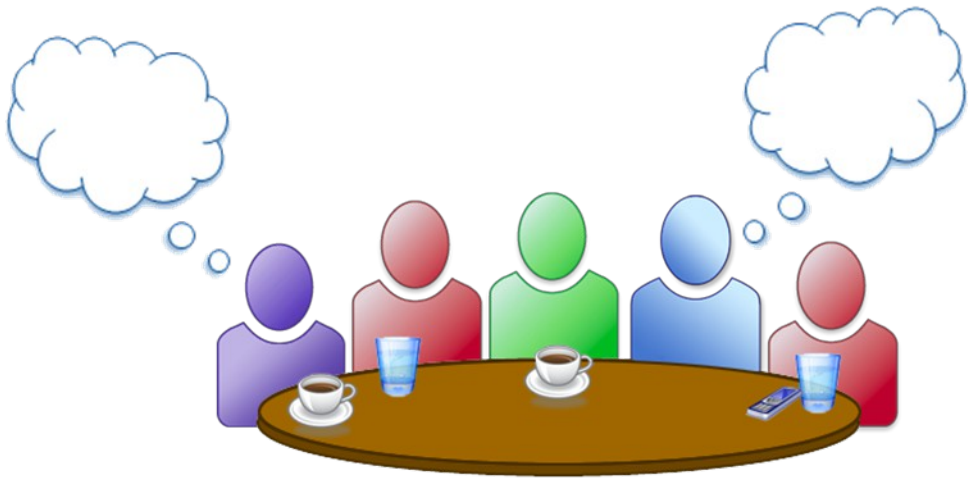
| Descriptive Statistic Keywords | | | | |
|---|---|---|---|---|
| CLM | CSS | CV | LCLM | MAX |
| MEAN | MIN | MODE | N | NMISS |
| KURTOSIS | RANGE | SKEWNESS | STDDEV | STDERR |
| SUM | SUMWGT | UCLM | USS | VAR |

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

| Hypothesis Testing Keywords | | | | |
|---|---|---|---|---|
| PROBT | T | | | |

# Idea Exchange

Which PROC MEANS statistics would you request when validating numeric variables?



# Business Scenario

Validate salary data in **orion.nonsales2.** **Salary** must be in the numeric range of 24000 to 500000.

Partial **orion.nonsales2**

| Employee_ID | First | Last | Gender | Salary | Job_Title | Country |
|---|---|---|---|---|---|---|
| 120101 | Patrick | Lu | M | 163040 | Director | AU |
| 120104 | Kareen | Billington | F | 46230 | Admin Mgr | au |
| 120105 | Liz | Povey | F | 27110 | Secretary I | AU |
| 120106 | John | Hornsey | M | . | Office Asst II | AU |
| 120107 | Sherie | Sheedy | F | 30475 | Office Asst II | AU |
| 120108 | Gladys | Gromek | F | 27660 | Warehouse Asst II | AU |

# 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.

# 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.

# 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** |

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)*;

# Viewing the Output

## Partial PROC UNIVARIATE Output

**The UNIVARIATE Procedure**
**Variable: Salary**

**Extreme Observations**

| ----------Lowest---------- | | | ----------Highest---------- | | |
|---|---|---|---|---|---|
| Value | Employee_ID | Obs | Value | Employee_ID | Obs |
| 2401 | 120191 | 20 | 163040 | 120101 | 1 |
| 2650 | 120115 | 13 | 194885 | 121141 | 231 |
| 24025 | 120196 | 25 | 207885 | 120260 | 28 |
| 24100 | 120190 | 19 | 268455 | 120262 | 29 |
| 24390 | 121132 | 228 | 433800 | 120259 | 27 |

# Viewing the Output: Missing Values Section

The *Missing Values* section displays the number and percentage of observations with missing values of the analysis variable.

Partial PROC UNIVARIATE Output

**Missing Values**

| Missing Value | Count | Percent Of All Obs | Percent Of Missing Obs |
|---|---|---|---|
| . | 1 | 0.43 | 100.00 |

# 11.06 Quiz

PROC UNIVARIATE identified two observations with **Salary** values less than 24,000.

What procedure can be used to display the observations containing the invalid values?

# 11.06 Quiz – Correct Answer

PROC UNIVARIATE identified two observations with **Salary** values less than 24,000.

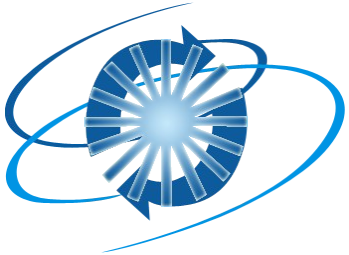
What procedure can be used to display the observations containing the invalid values? **PROC PRINT**

```
proc print data=orion.nonsales2;
   where Salary<24000;
run;
```

PROC PRINT Output

| Obs | Employee_ID | First | Last | Gender | Salary | Job_Title | Country |
|---|---|---|---|---|---|---|---|
| 4 | 120106 | John | Hornsey | M | . | Office Assistant II | AU |
| 13 | 120115 | Hugh | Nicholas | M | 2650 | Service Assistant I | AU |
| 20 | 120191 | Jannene | Graham-Rowe | F | 2401 | Trainee | AU |

p111a04s



# Exercise

This exercise reinforces the concepts discussed previously.

# Chapter 11: Creating Summary Reports

# Objectives

- Define the Output Delivery System and ODS destinations.
- Create report output in PDF, RTF, and HTML format.
- Specify a style definition using the STYLE= option.
- Create report output that can be viewed in Microsoft Excel.

# Business Scenario

Generate reports in various formats for distribution within Orion Star.

# SAS Procedure Output

Historically, SAS procedures generate plain text output using monospace fonts.

| Obs | First_Name | Last_Name | Job_Title | Salary |
|---|---|---|---|---|
| 1 | Satyakam | Denny | Sales Rep. II | 26780 |
| 2 | Monica | Kletschkus | Sales Rep. IV | 30890 |
| 3 | Kevin | Lyon | Sales Rep. I | 26955 |
| 4 | Petrea | Soltau | Sales Rep. II | 27440 |
| 5 | Marina | Iyengar | Sales Rep. III | 29715 |
| 6 | Shani | Duckett | Sales Rep. I | 25795 |
| 7 | Fang | Wilson | Sales Rep. II | 26810 |
| 8 | Michael | Minas | Sales Rep. I | 26970 |
| 9 | Amanda | Liebman | Sales Rep. II | 27465 |
| 10 | Vincent | Eastley | Sales Rep. III | 29695 |

# Output in SAS Enterprise Guide

SAS Enterprise Guide produces SAS Report output by default.

- The output from each program is written to a specially formatted file.
- You can request the following output styles by selecting **Tools** ▸ **Options** ▸ **Results General**:
  - PDF
  - RTF
  - HTML
  - LISTING (plain text)
- The PDF, RTF, and HTML files can be saved and opened outside of Enterprise Guide.

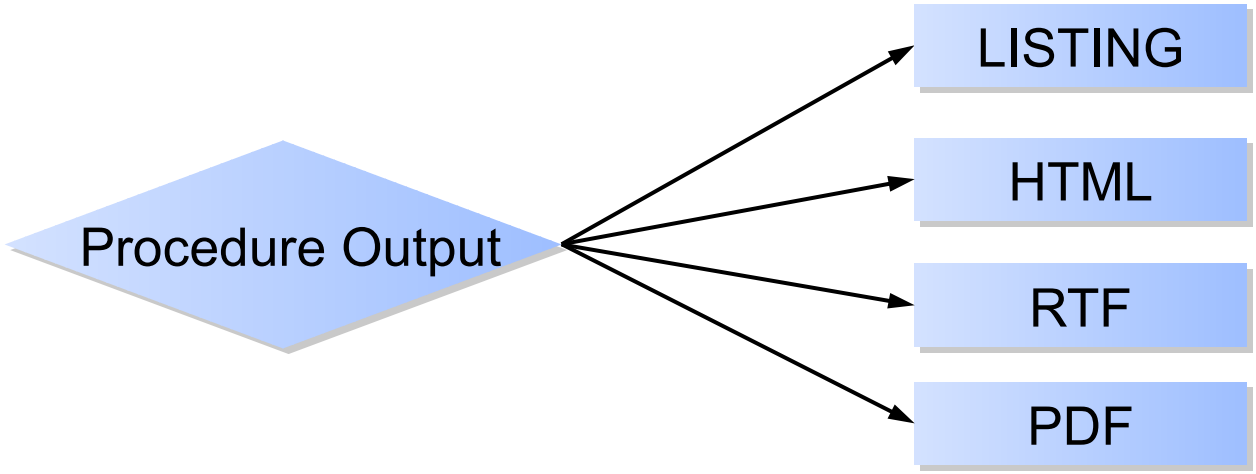
# Output in the SAS Windowing Environment

SAS 9.3 produces HTML output by default.

- All procedure output accumulates in a single file.
- LISTING output can be requested by selecting **Tools** ▸ **Options** ▸ **Preferences** ▸ **Results**.
- LISTING output is the default for prior releases and for SAS 9.3 in batch mode.

# Output Delivery System (ODS)

Use the Output Delivery System to create different output formats by directing output to various ODS destinations.




ODS statements and destinations can be used regardless of the mode, operating environment, or platform.

# ODS Statements

Open an ODS destination, submit one or more procedures that generate output, and then close the destination.

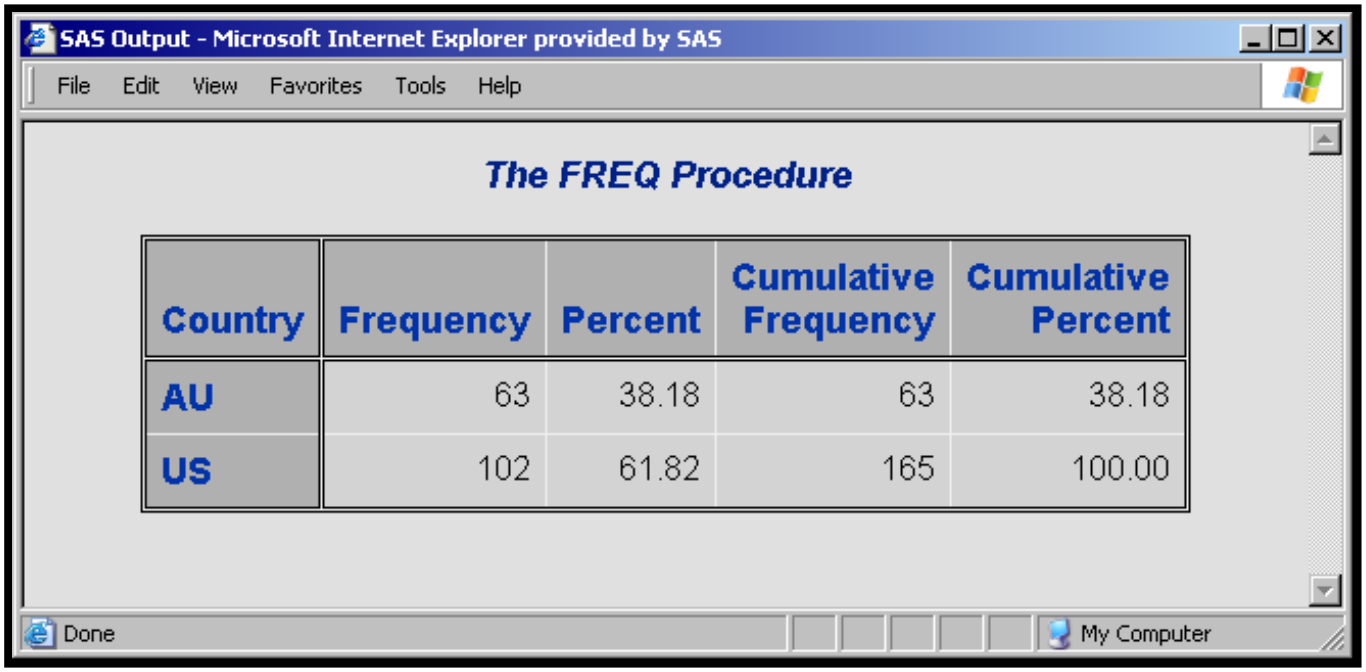
```
ods html file="&path\myreport.html";
proc freq data=orion.sales;
   tables Country;
run;
ods html close;
```

**ODS** *destination* **FILE="***filename***"** *<options>*;
*<SAS code to generate the report>*
**ODS** *destination* **CLOSE;**

p111d12

# Viewing the Output

HTML output can be viewed in the interactive environment or using a browser.




| Country | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---|---|---|---|---|
| AU | 63 | 38.18 | 63 | 38.18 |
| US | 102 | 61.82 | 165 | 100.00 |

# ODS Destinations

ODS creates various types of output based on the specified destinations and file types.

| Destination | Type of File | Extension | Viewed In |
|---|---|---|---|
| LISTING | Plain text | | SAS Output window |
| HTML | Hypertext Markup Language | html | Web Browsers such as Internet Explorer |
| PDF | Portable Document Format | pdf | Adobe products such as Acrobat Reader |
| RTF | Rich Text Format | rtf | Word processors such as Microsoft Word |

# Multiple Destinations and Procedures

Output from multiple procedures can be sent to multiple ODS destinations.

```
ods pdf file="&path\example.pdf";
ods rtf file="&path\example.rtf";

proc freq data=orion.sales;
   tables Country;
run;

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

ods pdf close;
ods rtf close;
```



p111d14

# Closing All Destinations

Use _ALL_ in the ODS CLOSE statement to close all open destinations. This includes the LISTING destination.

```
ods pdf file="&path\example.pdf";
ods rtf file="&path\example.rtf";

proc freq data=orion.sales;
   tables Country;
run;

ods _all_ close;
```

# No Open Destinations

Be sure to have at least one destination open.

```
2723  ods _all_ close;
2724  proc freq data=orion.sales;
2725     tables Country;
2726  run;

WARNING: No output destinations active.
NOTE: There were 165 observations read from the data set ORION.SALES.

2727
2728  ods listing;
2729
2730  proc freq data=orion.sales;
2731     tables Country;
2732  run;

NOTE: There were 165 observations read from the data set ORION.SALES.
```

Output is generated but not displayed.

Output is displayed.

p111d15

# 11.07 Quiz

What is the problem with this program?

```
ods html file="&path\myreport.html";

proc print data=orion.sales;
run;

ods close;
```

p111a05

# 11.07 Quiz – Correct Answer

What is the problem with this program?

```
ods html file="&path\myreport.html";

proc print data=orion.sales;
run;

ods html close;
```

p111a01s

# STYLE= Option

Use a STYLE= option in the ODS statement to specify a style definition.

**ODS** *destination* **FILE="***filename***"**
**STYLE=***style-definition*;

- A *style definition* sets presentation aspects including colors and fonts.

- STYLE= has no effect in the LISTING destination.

# HTML Examples

STYLE=DEFAULT

*The FREQ Procedure*

| Country | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---|---|---|---|---|
| AU | 63 | 38.18 | 63 | 38.18 |
| US | 102 | 61.82 | 165 | 100.00 |

STYLE=SASWEB

The FREQ Procedure

| Country | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---|---|---|---|---|
| AU | 63 | 38.18 | 63 | 38.18 |
| US | 102 | 61.82 | 165 | 100.00 |



p111d12

# PDF and RTF Examples

STYLE=PRINTER

*The FREQ Procedure*

| Country | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---|---|---|---|---|
| AU | 63 | 38.18 | | |
| US | 102 | 61.82 | | |

STYLE=JOURNAL

*The FREQ Procedure*

| *Country* | *Frequency* | *Percent* | *Cumulative Frequency* | *Cumulative Percent* |
|---|---|---|---|---|
| *AU* | 63 | 38.18 | 63 | 38.18 |
| *US* | 102 | 61.82 | 165 | 100.00 |

STYLE=RTF

*The FREQ Procedure*

| Country | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---|---|---|---|---|
| AU | 63 | 38.18 | 63 | 38.18 |
| US | 102 | 61.82 | 165 | |

STYLE=OCEAN

The FREQ Procedure

| *Country* | *Frequency* | *Percent* | *Cumulative Frequency* | *Cumulative Percent* |
|---|---|---|---|---|
| *AU* | 63 | 38.18 | 63 | 38.18 |
| *US* | 102 | 61.82 | 165 | 100.00 |



p111d17

# Available Styles

Use the TEMPLATE procedure to see the available styles.

Partial Output

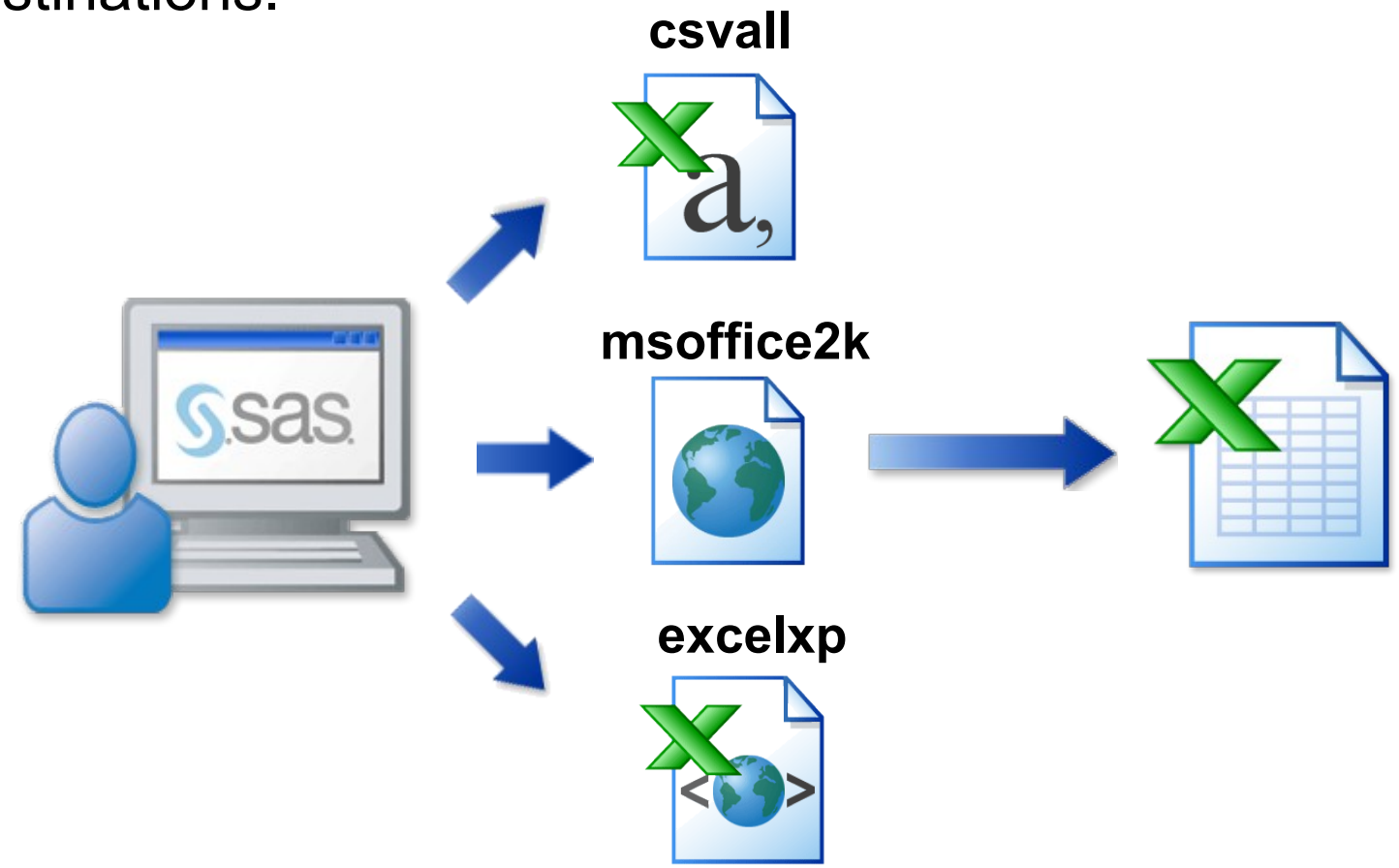
```
proc template;
   list styles;
run;
```

```
Listing of: SASHELP.TMPLMST
 Path Filter is: Styles
 Sort by: PATH/ASCENDING

 Obs   Path                  Type
 ffffffffffffffffffffffffffffff
  1    Styles                Dir
  2    Styles.Analysis       Style
  3    Styles.Astronomy      Style
  4    Styles.Banker         Style
  5    Styles.BarrettsBlue   Style
  6    Styles.Curve          Style
  7    Styles.Default        Style
  8    Styles.Dtree          Style
  9    Styles.EGDefault      Style
 10    Styles.Education      Style
```

p111d13

# Business Scenario

Create SAS reports that can be opened in Microsoft Excel using the CSVALL, MSOFFICE2K, and EXCELXP destinations.



# 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 |

# 11.08 Quiz

Complete the ODS statements below to send the output to a CSVALL destination.

```
ods          file="&path\myexcel.      ";

proc freq data=orion.sales;
   tables Country;
run;

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

ods          close;
```

# 11.08 Quiz – Correct Answer

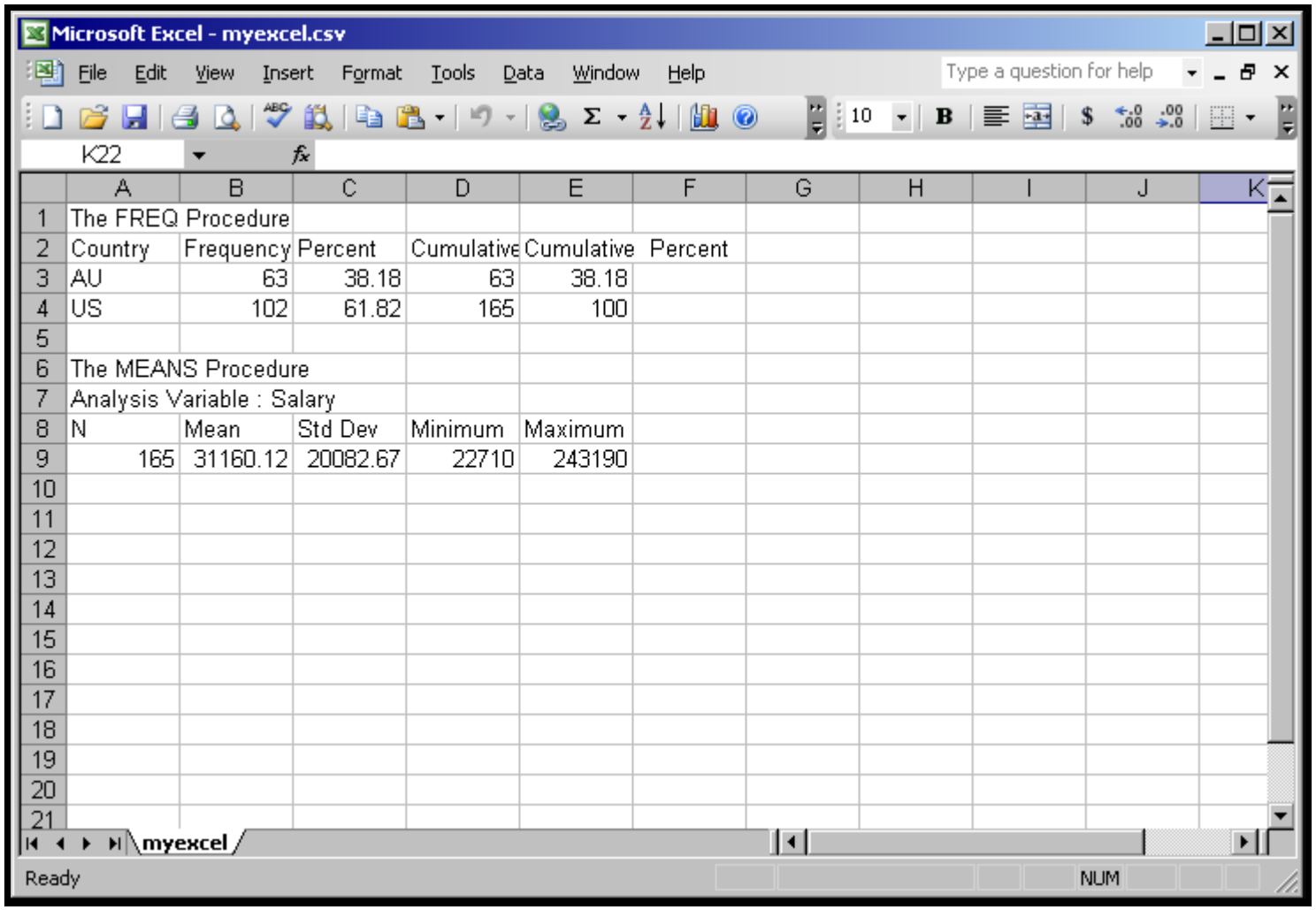
Complete the ODS statements below to send the output to a CSVALL destination.

```
ods csvall file="&path\myexcel.csv";

proc freq data=orion.sales;
   tables Country;
run;

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

ods csvall close;
```

p111a06s

# CSVALL Destination

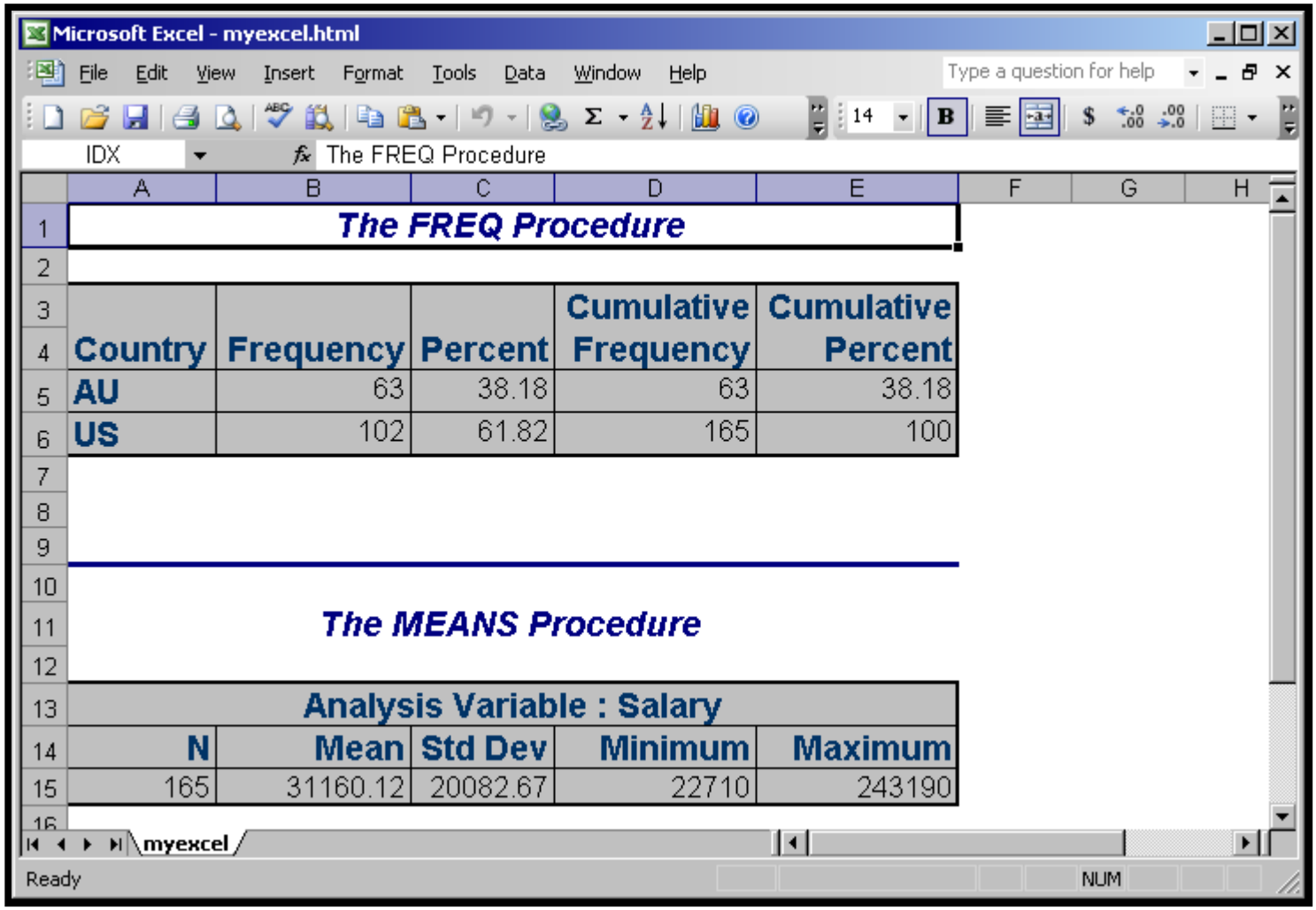
CSVALL does not include any style information.




| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | The FREQ Procedure | | | | | | | | | | |
| 2 | Country | Frequency | Percent | Cumulative | Cumulative | Percent | | | | | |
| 3 | AU | 63 | 38.18 | 63 | 38.18 | | | | | | |
| 4 | US | 102 | 61.82 | 165 | 100 | | | | | | |
| 5 | | | | | | | | | | | |
| 6 | The MEANS Procedure | | | | | | | | | | |
| 7 | Analysis Variable : Salary | | | | | | | | | | |
| 8 | N | Mean | Std Dev | Minimum | Maximum | | | | | | |
| 9 | 165 | 31160.12 | 20082.67 | 22710 | 243190 | | | | | | |
| 10 | | | | | | | | | | | |
| 11 | | | | | | | | | | | |
| 12 | | | | | | | | | | | |
| 13 | | | | | | | | | | | |
| 14 | | | | | | | | | | | |
| 15 | | | | | | | | | | | |
| 16 | | | | | | | | | | | |
| 17 | | | | | | | | | | | |
| 18 | | | | | | | | | | | |
| 19 | | | | | | | | | | | |
| 20 | | | | | | | | | | | |
| 21 | | | | | | | | | | | |

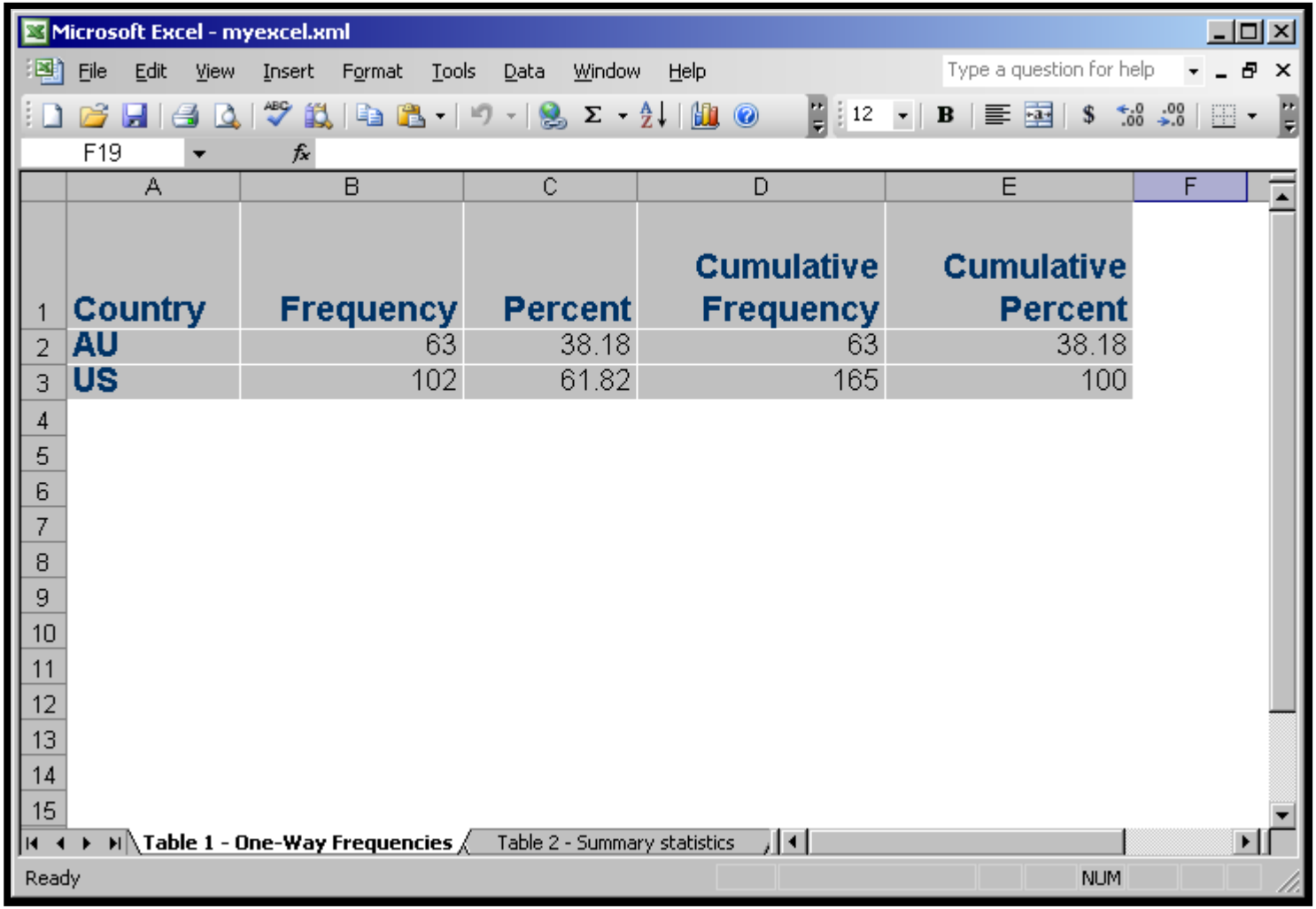
# MSOFFICE2K Destination

MSOFFICE2K keeps the style information, including spanning headers.



# EXCELXP Destination

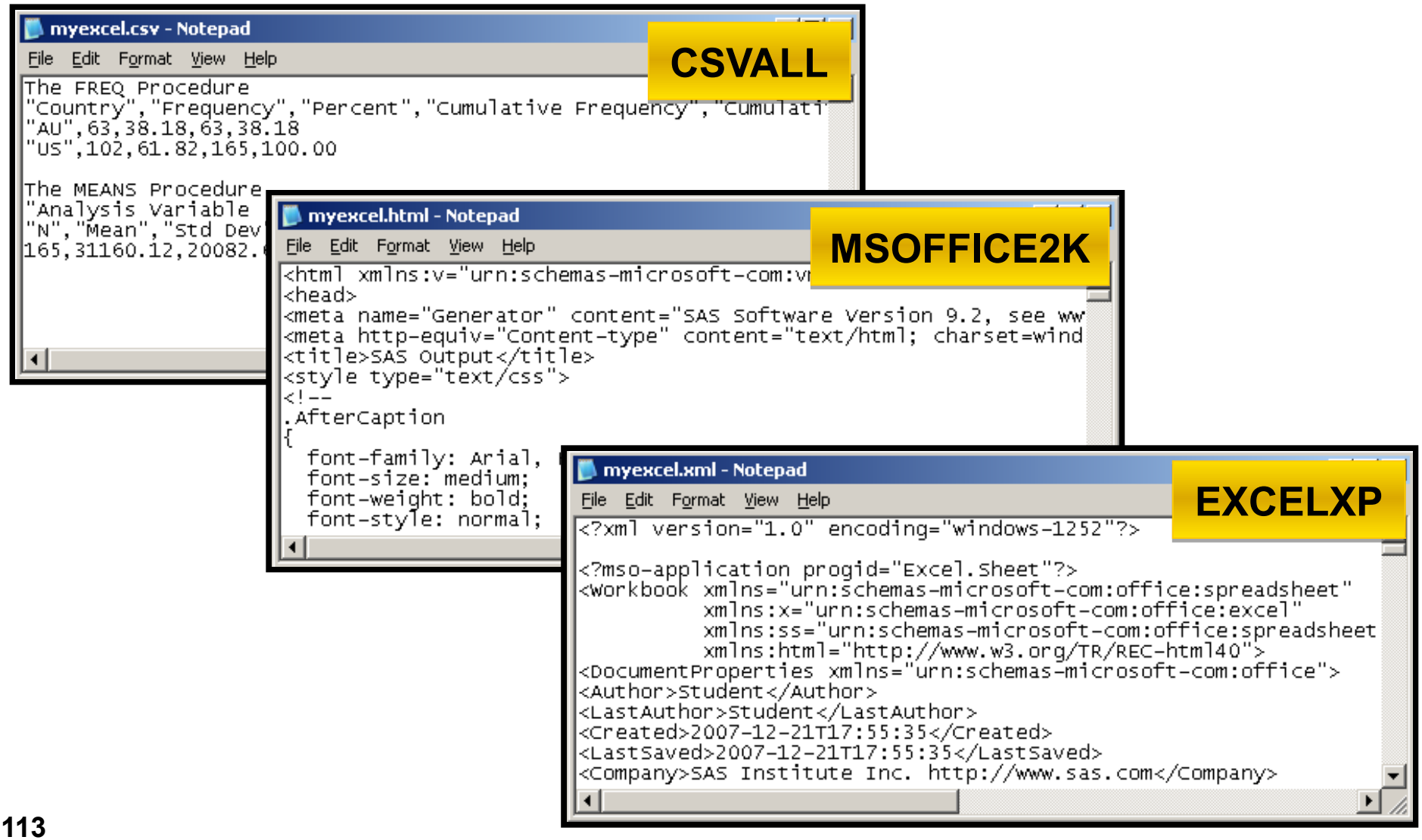
EXCELXP keeps the style information. Output from each procedure is on a separate sheet.




| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | Country | Frequency | Percent | Cumulative Frequency | Cumulative Percent | |
| 2 | AU | 63 | 38.18 | 63 | 38.18 | |
| 3 | US | 102 | 61.82 | 165 | 100 | |
| 4 | | | | | | |
| 5 | | | | | | |
| 6 | | | | | | |
| 7 | | | | | | |
| 8 | | | | | | |
| 9 | | | | | | |
| 10 | | | | | | |
| 11 | | | | | | |
| 12 | | | | | | |
| 13 | | | | | | |
| 14 | | | | | | |
| 15 | | | | | | |

# Keep in Mind

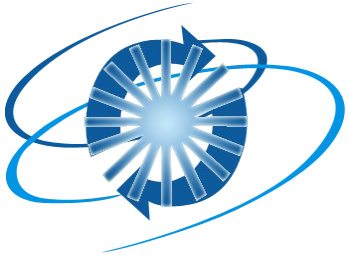
The file you are creating is not an Excel file.



# Using the Output Delivery System

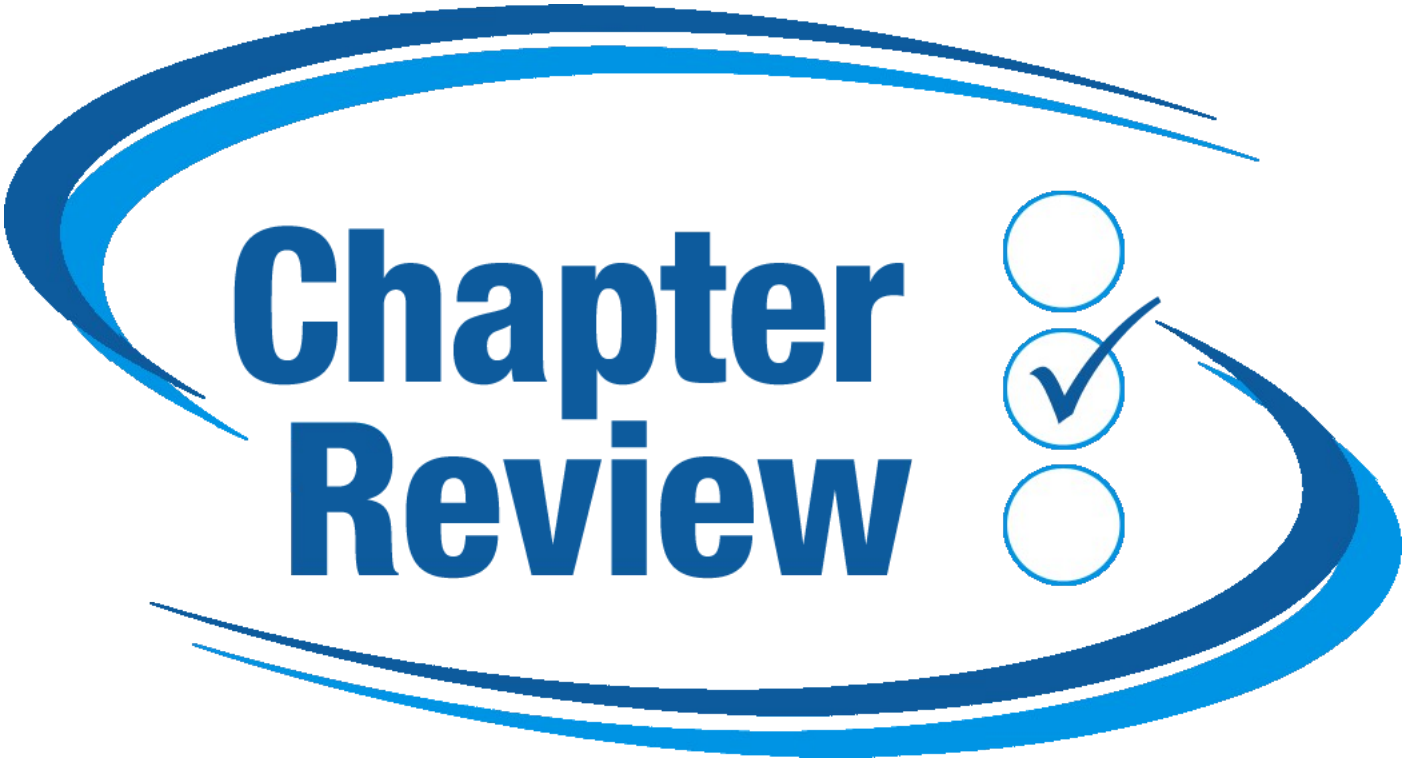
This demonstration illustrates using ODS statements to direct output to various destinations.

p111d18
p111d19



# Exercise

This exercise reinforces the concepts discussed previously.



1. Which of these procedures produces output that is most useful for detecting duplicate values?
   - PROC PRINT
   - PROC FREQ
   - PROC MEANS
   - PROC UNIVARIATE

1. Which of these procedures produces output that is most useful for detecting duplicate values?

   - PROC PRINT
   - PROC FREQ
   - PROC MEANS
   - PROC UNIVARIATE

2. Which of these programs is most useful for determining the exact observation that contains a numeric variable with an extreme value?

a.
```
proc print data=sales.totals;
     var ProdNum Sales Region;
run;
```

b.
```
proc freq data=sales.totals;
     tables ProdNum Sales Region;
run;
```

c.
```
proc univariate data=sales.totals;
run;
```

2. Which of these programs is most useful for determining the exact observation that contains a numeric variable with an extreme value?

a.
```
proc print data=sales.totals;
     var ProdNum Sales Region;
run;
```

b.
```
proc freq data=sales.totals;
     tables ProdNum Sales Region;
run;
```

c.
```
proc univariate data=sales.totals;
run;
```

3. A PROC FREQ analysis identified invalid and missing values in a data set. Which of these procedures will display the observations that contain invalid or missing values?

   - PROC PRINT
   - PROC FREQ
   - PROC MEANS
   - PROC UNIVARIATE

3. A PROC FREQ analysis identified invalid and missing values in a data set. Which of these procedures will display the observations that contain invalid or missing values?

   - PROC PRINT
   - PROC FREQ
   - PROC MEANS
   - PROC UNIVARIATE

4. Which PROC FREQ step creates the output shown here?

a.
```
proc freq data=orion.qtr1_2007;
      tables Order_Type;
run;
```

b.
```
proc freq data=orion.qtr1_2007
            nlevels;
      tables Order_Type / nocum;
run;
```

c.
```
proc freq data=orion.qtr1_2007
            nlevels;
      tables Order_Type / noprint;
run;
```

d.
```
proc freq data=otion.qtr1_2007
            nlevels;
      tables Order_Type nocum;
run;
```

Number of Variable Levels

| Variable | Label | Levels |
|---|---|---|
| Order_Type | Order Type | 3 |

Order Type

| Order_Type | Frequency | Percent |
|---|---|---|
| 1 | 13 | 59.09 |
| 2 | 2 | 9.09 |
| 3 | 7 | 31.82 |

# 4. Which PROC FREQ step creates the output shown here?

a.
```
proc freq data=orion.qtr1_2007;
      tables Order_Type;
run;
```

b.
```
proc freq data=orion.qtr1_2007
            nlevels;
      tables Order_Type / nocum;
run;
```

c.
```
proc freq data=orion.qtr1_2007
            nlevels;
      tables Order_Type / noprint;
run;
```

d.
```
proc freq data=otion.qtr1_2007
            nlevels;
      tables Order_Type nocum;
run;
```

| Number of Variable Levels | | |
|---|---|---|
| Variable | Label | Levels |
| Order_Type | Order Type | 3 |

| Order Type | | |
|---|---|---|
| Order_Type | Frequency | Percent |
| 1 | 13 | 59.09 |
| 2 | 2 | 9.09 |
| 3 | 7 | 31.82 |

5. This PROC MEANS step creates all of the statistics listed below.

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

   - minimum and maximum
   - the total number of observations that PROC MEANS processes for each subgroup (**N Obs**)
   - mean and standard deviation
   - the number of nonmissing values (**N**)

   - True
   - False

5. This PROC MEANS step creates all of the statistics listed below.

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

   - minimum and maximum
   - the total number of observations that PROC MEANS processes for each subgroup (**N Obs**)
   - mean and standard deviation
   - the number of nonmissing values (**N**)

—| True
—| False

6. What must be added to the PROC MEANS statement to produce this output?

```
proc means data=orion.customer_dim
                 _____________________________;
   var Customer_Age;
   class Customer_Gender;
   where Customer_Country ne 'US';
run;
```

The MEANS Procedure

| Analysis Variable : Customer_Age Customer Age | | |
|---|---|---|
| Customer Gender | Range | Mean |
| F | 54.0 | 35.1 |
| M | 54.0 | 47.0 |

- nonobs
- range mean
- range mean nonobs bestw.
- range mean nonobs maxdec=1

6. What must be added to the PROC MEANS statement to produce this output?

```
proc means data=orion.customer_dim
                 _____________________________;
   var Customer_Age;
   class Customer_Gender;
   where Customer_Country ne 'US';
run;
```

The MEANS Procedure

Analysis Variable : Customer_Age
Customer Age

| Customer Gender | Range | Mean |
|---|---|---|
| F | 54.0 | 35.1 |
| M | 54.0 | 47.0 |

- nonobs
- range mean
- range mean nonobs bestw.
- range mean nonobs maxdec=1

7. Which option lets you specify the number of extreme observations displayed by PROC UNIVARIATE?

   - NEXTROBS=
   - NLEVELS
   - NOPRINT
   - _ALL_

7. Which option lets you specify the number of extreme observations displayed by PROC UNIVARIATE?

   - NEXTROBS=
   - NLEVELS
   - NOPRINT
   - _ALL_

8. Which destination creates a file that keeps the style information and opens in multiple worksheets in an Excel workbook?

   - CSVALL
   - EXCELXP
   - MSOFFICE2K
   - none of the above

8. Which destination creates a file that keeps the style information and opens in multiple worksheets in an Excel workbook?

   - CSVALL
   - EXCELXP
   - MSOFFICE2K
   - none of the above

9. Which statement about style definitions is true?

   - The STYLE= option affects the display in all destinations.
   - You can use the STYLE= option in an ODS statement or in a PROC statement.
   - A style definition specifies colors, fonts, and a file format for an external file.
   - If you do not specify a style definition, SAS uses a default style definition that varies by the destination.

9. Which statement about style definitions is true?

   - The STYLE= option affects the display in all destinations.
   - You can use the STYLE= option in an ODS statement or in a PROC statement.
   - A style definition specifies colors, fonts, and a file format for an external file.
   - If you do not specify a style definition, SAS uses a default style definition that varies by the destination.

10. Suppose you submit the program shown below. What happens if you then submit a PROC PRINT step?

```
ods _all_ close;
ods csvall file='c:\ctry.csv';

proc freq data=orion.sales;
   tables Country;
run;

ods csvall close;
```

- The PROC PRINT output is displayed in the default window.
- The PROC PRINT output is not displayed and a warning is written to the log indicating that there are no active destinations.
- The PROC PRINT output is appended to the PROC FREQ output in the file **ctry.csv**.

10. Suppose you submit the program shown below. What happens if you then submit a PROC PRINT step?

```
ods _all_ close;
ods csvall file='c:\ctry.csv';

proc freq data=orion.sales;
   tables Country;
run;

ods csvall close;
```

- The PROC PRINT output is displayed in the default window.
- The PROC PRINT output is not displayed and a warning is written to the log indicating that there are no active destinations.
- The PROC PRINT output is appended to the PROC FREQ output in the file **ctry.csv**.