15 Homework Summary Reports 2

1. **Specifying Multiple Titles**

The data set order\_fact is in the STA5066 directory.

Use the following starter program:

proc means data=orion.order\_fact;

where Order\_Type=2;

var Total\_Retail\_Price;

run;

proc means data=orion.order\_fact;

where Order\_Type=3;

var Total\_Retail\_Price;

run;

Follow the steps to produce the reports on the next page.

**a.** Specify the following title to appear in both reports: **Orion Star Sales Analysis**

**b.** Specify a secondary title to appear in the first report with a blank line between the titles:

**Catalog Sales Only**

**c.** Specify the following footnote for the first report:

**Based on the previous day's posted data**

The text specified for a title or footnote may be enclosed in single quotes or double quotes. Use double quotes when the text contains an apostrophe.

**d.** Specify a secondary title to appear only in the second report with a blank line between the titles:

**Internet Sales Only**

**e.** Cancel all footnotes for the second report.

**f.** Submit the program to create the following PROC MEANS reports:

PROC MEANS Output

**Orion Star Sales Analysis**

**Catalog Sales Only**

**The MEANS Procedure**

| **Analysis Variable : Total\_Retail\_Price Total Retail Price for This Product** | | | | |
| --- | --- | --- | --- | --- |
| **N** | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| 170 | 199.5961765 | 282.9680817 | 2.6000000 | 1937.20 |

Based on the previous day's posted data

**Orion Star Sales Analysis**

**Internet Sales Only**

**The MEANS Procedure**

| **Analysis Variable : Total\_Retail\_Price Total Retail Price for This Product** | | | | |
| --- | --- | --- | --- | --- |
| **N** | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| 123 | 174.7280488 | 214.3528338 | 2.7000000 | 1542.60 |

1. **Inserting Dates and Times into Titles**

The data set order\_fact is in the STA5066 directory.

Use the following starter program:

proc means data=orion.order\_fact;

var Total\_Retail\_Price;

run;

1. Add a title with the following text, substituting the current date and time (The time and date should correspond to when you run the program):

Sales Report as of ***4:57 PM*** on ***Monday, January 28, 2008***

1. Submit the program to create the following report (the date for your report will be different):

PROC MEANS Output

**Sales Report as of 5:19:37 PM on November 7, 2020**

**The MEANS Procedure**

| **Analysis Variable : Total\_Retail\_Price Total Retail Price for This Product** | | | | |
| --- | --- | --- | --- | --- |
| **N** | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| 617 | 162.2001053 | 233.8530183 | 2.6000000 | 1937.20 |

**3. Overriding Existing Labels and Formats**

The data set customer is in the STA5066 directory

Use the following starter program:

proc print data=orion.customer;

where Country='TR';

title 'Customers from Turkey';

var Customer\_ID Customer\_FirstName Customer\_LastName

Birth\_Date;

run;

1. Display only the year portion of birth dates.
2. Modify the column heading for each variable as shown in the sample output that follows. Be sure that the column headers for the customer's last name and first name are also split into two lines.
3. Submit the program to produce the following report:

**Customers from Turkey**

| **Obs** | **Customer ID** | **First Name** | **Last Name** | **Birth Year** |
| --- | --- | --- | --- | --- |
| **47** | 544 | Avni | Argac | 1964 |
| **48** | 908 | Avni | Umran | 1979 |
| **49** | 928 | Bulent | Urfalioglu | 1969 |
| **50** | 1033 | Selim | Okay | 1979 |
| **51** | 1100 | Ahmet | Canko | 1964 |
| **52** | 1684 | Carglar | Aydemir | 1974 |
| **55** | 2788 | Serdar | Yucel | 1944 |

1. **Creating User Defined Formats**

The data set employee\_payroll is in the STA5066 directory.

Use the following starter program:

data Q1Birthdays;

set orion.employee\_payroll;

BirthMonth=month(Birth\_Date);

if BirthMonth le 3;

run;

proc freq data=Q1Birthdays;

tables BirthMonth Employee\_Gender;

title 'Employees with Birthdays in Q1';

run;

1. Create a character format named **$gender** that displays gender codes as follows:

|  |  |
| --- | --- |
| F | Female |
| M | Male |

1. Create a numeric format named **moname** that displays month numbers as follows:

|  |  |
| --- | --- |
| 1 | January |
| 2 | February |
| 3 | March |

1. In the PROC FREQ step, apply these two user-defined formats to the **Employee\_Gender**   
   and **BirthMonth** variables, respectively.
2. Submit the program to produce the report on the next page:

**Employees with Birthdays in Q1**

**The FREQ Procedure**

| **BirthMonth** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| --- | --- | --- | --- | --- |
| **January** | 44 | 38.94 | 44 | 38.94 |
| **February** | 34 | 30.09 | 78 | 69.03 |
| **March** | 35 | 30.97 | 113 | 100.00 |

| **Employee\_Gender** | **Frequency** | **Percent** | **Cumulative Frequency** | **Cumulative Percent** |
| --- | --- | --- | --- | --- |
| **Female** | 52 | 46.02 | 52 | 46.02 |
| **Male** | 61 | 53.98 | 113 | 100.00 |

1. **Defining Ranges in User defined Formats**

The data set nonsales is in the STA5066

Use the following starter program:

proc print data=orion.nonsales (obs=10);

var Employee\_ID Job\_Title Salary Gender;

title1 'Distribution of Salary and Gender Values';

title2 'for Non-Sales Employees';

run;

* 1. Create a character format named **$gender** that displays gender codes as follows:

|  |  |
| --- | --- |
| F | Female |
| M | Male |
| Any other value | Invalid code |

* 1. Create a numeric format named **salrange** that displays salary ranges as follows:

|  |  |
| --- | --- |
| At least 20,000 but less than 100,000 | Below $100,000 |
| At least 100,000 and up to 500,000 | $100,000 or more |
| missing | Missing salary |
| Any other value | Invalid salary |

* 1. In the PROC PRINT step, apply these two user-defined formats to the **Gender** and **Salary** variables, respectively.
  2. Submit the program to produce the report on the next page:

**Distribution of Salary and Gender Values**

**for Non-Sales Employees**

| **Obs** | **Employee\_ID** | **Job\_Title** | **Salary** | **Gender** |
| --- | --- | --- | --- | --- |
| **1** | 120101 | Director | $100,000 or more | M |
| **2** | 120104 | Administration Manager | Below $100,000 | F |
| **3** | 120105 | Secretary I | Below $100,000 | F |
| **4** | 120106 | Office Assistant II | Missing salary | M |
| **5** | 120107 | Office Assistant III | Below $100,000 | F |
| **6** | 120108 | Warehouse Assistant II | Below $100,000 | F |
| **7** | 120108 | Warehouse Assistant I | Below $100,000 | F |
| **8** | 120110 | Warehouse Assistant III | Below $100,000 | M |
| **9** | 120111 | Security Guard II | Below $100,000 | M |
| **10** | 120112 |  | Below $100,000 | F |

1. **Subsetting and Grouping Observations**

The data set order\_fact is in the STA5066 subdirectory.

Use the following starter program:

proc means data=orion.order\_fact;

var Total\_Retail\_Price;

title 'Orion Star Sales Summary';

run;

* 1. Add a PROC SORT step to sort the observations in **orion.order\_fact** based   
     on the **Order\_Type** variable.

To avoid overwriting the **orion.order\_fact** data set, be sure to use   
the OUT= option to create a new data set containing the sorted observations.   
Remember to use the new data set in the PROC MEANS step.

* 1. Restrict the PROC MEANS analysis to two **Order\_Type** values: 2 and 3.
  2. Modify the PROC MEANS step to generate the summary analysis separately for each   
     selected **Order\_Type** value in the sorted data set.
  3. Submit the program to produce the following output:

**Orion Star Sales Summary**

**The MEANS Procedure**

**Order Type=2**

| **Analysis Variable : Total\_Retail\_Price Total Retail Price for This Product** | | | | |
| --- | --- | --- | --- | --- |
| **N** | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| 170 | 199.5961765 | 282.9680817 | 2.6000000 | 1937.20 |

**Order Type=3**

| **Analysis Variable : Total\_Retail\_Price Total Retail Price for This Product** | | | | |
| --- | --- | --- | --- | --- |
| **N** | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| 123 | 174.7280488 | 214.3528338 | 2.7000000 | 1542.60 |

1. **Subsetting and Grouping by Multiple Variables**

The data set order\_fact is in the STA5066 subdirectory.

Use the following starter program:

proc print data=orion.order\_fact;

var Order\_Type Order\_ID Order\_Date Delivery\_Date ;

title1 'Orion Star Sales Details';

run;

* 1. Sort the **orion.order\_fact** data set by **Order\_Type** (in ascending sequence)   
     and **Order\_Date** (in descending sequence).

Create a new data set containing the sorted observations. Do not overwrite the **orion.order\_fact** data set. Remember to use the new data set in the PROC PRINT step.

* 1. Divide the PROC PRINT report based on **Order\_Type** using a BY statement.
  2. Limit the observations in the PROC PRINT report based on the following criteria:
     1. Orders placed in the first four months of 2005 (January 1 to April 30).
     2. Orders that were delivered exactly two days after the order was placed.
  3. Add a second title to clarify that filters have been applied to the data.
  4. Submit the program to produce the report on the next page:

**Orion Star Sales Details**

**2-Day Deliveries from January to April 2005**

**Order Type=2**

| **Obs** | **Order\_Type** | **Order\_ID** | **Order\_Date** | **Delivery\_Date** |
| --- | --- | --- | --- | --- |
| **409** | 2 | 1235611754 | 27APR2005 | 29APR2005 |
| **410** | 2 | 1235611754 | 27APR2005 | 29APR2005 |
| **411** | 2 | 1235591214 | 25APR2005 | 27APR2005 |
| **412** | 2 | 1235591214 | 25APR2005 | 27APR2005 |
| **413** | 2 | 1234972570 | 24FEB2005 | 26FEB2005 |
| **415** | 2 | 1234659163 | 24JAN2005 | 26JAN2005 |
| **417** | 2 | 1234588648 | 17JAN2005 | 19JAN2005 |
| **418** | 2 | 1234588648 | 17JAN2005 | 19JAN2005 |
| **419** | 2 | 1234538390 | 12JAN2005 | 14JAN2005 |

**Order Type=3**

| **Obs** | **Order\_Type** | **Order\_ID** | **Order\_Date** | **Delivery\_Date** |
| --- | --- | --- | --- | --- |
| **568** | 3 | 1235176942 | 15MAR2005 | 17MAR2005 |
| **569** | 3 | 1235176942 | 15MAR2005 | 17MAR2005 |
| **570** | 3 | 1234891576 | 16FEB2005 | 18FEB2005 |

1. **Creating a Simple Tabular Report with PROC TABULATE**

The data set customer\_dim is in the STA5066 directory.

Use the following starter program

proc tabulate data=orion.customer\_dim;

title 'Ages of Customers by Group and Gender';

run;

* 1. Add a CLASS statement to allow **Customer\_Group** and **Customer\_Gender** to be classification variables.
  2. Add a VAR statement to permit **Customer\_Age** as an analysis variable
  3. Add a TABLE statement to create a report with the following characteristics:
     1. An extra row that combines all groups appears at the bottom of the table.
     2. The N and MEAN statistics based on **Customer\_Age** are displayed for each combination of **Customer\_Group** and **Customer\_Gender**.
  4. Submit the program to produce the following report:

**Ages of Customers by Group and Gender**

|  | | **Customer Age** | |
| --- | --- | --- | --- |
| **N** | **Mean** |
| **Customer Group Name** | **Customer Gender** | 4 | 49.25 |
| **Internet/Catalog Customers** | **F** |
| **M** | 4 | 54.25 |
| **Orion Club Gold members** | **F** | 11 | 35.36 |
| **M** | 10 | 38.90 |
| **Orion Club members** | **F** | 15 | 32.53 |
| **M** | 33 | 47.03 |
| **All** | | 77 | 41.97 |