STAT 425 and STAT 625- Statistical Software

Dr. William Howell, Professor

Homework #2-Producing Detail Reports; Formatting Data Values

Directions: Please submit one program file, one output file, and one log file for the entire assignment. Use comment statements to separate your answers. For questions that do not require a SAS program use comment statements. For example:

/*

Question #1d: my answer

Question #2a: my answer

*/

/*Question #4b: */

--SAS program—

/*Question #5*/

Please make sure the log and output file contain only one run. For example, clear the screen for the log and output file and submit your program one last time before you upload your solutions to ***Blackboard***. See lab 1 for the instructions on how to clear your output and log files.

Part I:Producing Detail Reports

1. **Displaying orion.customer_dim with the PRINT Procedure**
   - **a.** Write a PRINT step to display **orion.customer_dim**.
   - **b.** Modify the program to display a subset of **orion.customer_dim** by selecting only the observations for customers between the ages of 30 and 40. Also, suppress the Obs column. The resulting report should contain 17 observations.
   - **c.** Add a statement to use **Customer_ID** instead of Obs as the identifying column. Submit the program and verify the results.
   - **d.** Add a statement to limit the variables to those shown in the report below.

Customer_ID	Customer_Name	Customer_ Age	Customer_Type
4	James Kvarniq	33	Orion Club members low activity
9	Cornelia Krahl	33	Orion Club Gold members medium activity
11	Elke Wallstab	33	Orion Club members high activity
...	...	...	...
54655	Lauren Marx	38	Internet/Catalog Customers
70201	Angel Borwick	38	Orion Club Gold members low activity

2. **Sorting orion.employee_payroll and Displaying a Subset of the New Data Set**
   - **a.** Sort **orion.employee_payroll** by **Employee_Gender**, and by descending **Salary** within gender. Place the sorted observations into a temporary data set named **sort_sal**.

b. Print a subset of the **sort_sal** data set. Select only the observations for active employees (those without a value for **Employee_Term_Date**) who earn more than $65,000. Group the report by **Employee_Gender**, and include a total and subtotals for **Salary**. Suppress the Obs column. Display only **Employee_ID**, **Salary**, and **Marital_Status**. The results contain 18 observations.

```
---------------------------------- Employee_Gender=F ---------------------------------

                                                 Marital_
                          Employee_ID    Salary    Status

                            120260       207885      M
                            120719        87420      M
                            ...
                            120677        65555      M
                         ---------------  ------
                         Employee_Gender  605190


---------------------------------- Employee_Gender=M ---------------------------------

                                                 Marital_
                          Employee_ID    Salary    Status

                            120259       433800      M
                            120262       268455      M
                            ...
                            120268        76105      S
                         ---------------  -------
                         Employee_Gender  2072410
                                          =======
                                          2677600
```

3. Writing an Enhanced Detail Report

a. Write a program to display a subset of **orion.employee_addresses** as shown below. The program should sort the observations by **State**, **City**, and **Employee_Name** and then display the sorted observations grouped by **State**. The resulting report should contain 311 observations.

```
                                US Employees by State

------------------------------ State=CA ---------------------------

      Employee                                                  Zip
         ID       Name                              City       Code

       120656     Amos, Salley                    San Diego    92116
       120759     Apr, Nishan                     San Diego    92071
       121017     Arizmendi, Gilbert              San Diego    91950
       121062     Armant, Debra                   San Diego    92025
       121049     Bataineh, Perrior               San Diego    92126
        ...
```

Part II: Formatting Data Values

1. Displaying Formatted Values in a Detail Report

Write a PROC PRINT step to display the report below using **orion.sales** as input. Subset the observations and variables to produce the report. Include titles, labels, and formats. The results contain 13 observations.

US Sales Employees
Earning Under $26,000

Employee_ID	First Name	Last Name	Title	Salary	Date Hired
121036	Teresa	Mesley	Sales Rep. I	$25,965	OCT2007
121038	David	Anstey	Sales Rep. I	$25,285	AUG2010
121044	Ray	Abbott	Sales Rep. I	$25,660	AUG1979
...					
121106	James	Hilburger	Sales Rep. I	$25,880	FEB2000
121108	Libby	Levi	Sales Rep. I	$25,930	NOV2010

2. Defining Ranges in User-Defined Formats

a. Retrieve the starter program **p105e05**.

b. Create a character format named $GENDER that displays gender codes as follows:

F	Female
M	Male
Any other value	**Invalid code**

c. 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

d. In the PROC PRINT step, apply these two user-defined formats to the **Gender** and **Salary** variables, respectively. Submit the program to produce the following report:

Partial PROC PRINT Output

Salary and Gender Values
for Non-Sales Employees

Obs	Employee_ID	Job_Title	Salary	Gender
1	120101	Director	$100,000 or more	Male
2	120104	Administration Manager	Below $100,000	Female
3	120105	Secretary I	Below $100,000	Female
4	120106	Office Assistant II	Missing salary	Male
5	120107	Office Assistant III	Below $100,000	Female
6	120108	Warehouse Assistant II	Below $100,000	Female
7	120108	Warehouse Assistant I	Below $100,000	Female

8	120110	Warehouse Assistant III	Below $100,000	Male
9	120111	Security Guard II	Below $100,000	Male
10	120112		Below $100,000	Female
11	120113	Security Guard II	Below $100,000	Female
12	120114	Security Manager	Below $100,000	Invalid code
13	120115	Service Assistant I	Invalid salary	Male

Supplemental exercises for STAT 625 and Honors credit

Part III:

1. Producing a Default Listing Report of orion.order_fact

This exercise assumes that you are creating LISTING output in the SAS windowing environment.

a. Produce a default listing report of **orion.order_fact**. The output might wrap onto a second line.

b. Investigate the use of the LINESIZE= SAS system option to adjust the width of the lines. What are the minimum and maximum values for the LINESIZE= option? _____

Submit an OPTIONS statement with LINESIZE= set to the highest allowed value. Resubmit the step, and observe the horizontal scroll bar, if it is displayed.

Reset the line size to 96 when you are finished.

c. Another way to create compact output is to request vertical headings. Investigate the HEADING= option in the PROC PRINT statement, and then experiment with it to generate vertical headings and then horizontal headings.

How do you specify vertical headings? _____

How do you specify horizontal headings? _____

2. Retaining the First Observation of Each BY Group

a. Sort **orion.orders** by **Customer_ID**. Place the sorted observations in a temporary data set.

b. Display the sorted data set. The resulting report should contain 490 observations. **Customer_ID** is listed multiple times for customers that placed more than one order.

c. Investigate an option that causes PROC SORT to retain only the first observation in each BY group.

d. Add the appropriate option to the PROC SORT step to retain only the first observation in each BY group. The results contain 75 observations with no duplicate values for **Customer_ID**.

e. Explore the DUPOUT= option to write duplicate observations to a separate output data set.

3. Exploring Formats by Category

Display **orion.sales** as shown in the report below. Refer to SAS Help or product documentation to explore the **Dictionary of Formats** and investigate **Formats by Category**. Identify and use the character format that displays values in uppercase and a format that displays a character value in quotation marks. The results contain 165 observations.

```
              First_
Employee_ID   Name         Last_Name       Job_Title

     120102   TOM          ZHOU            "Sales Manager"
     120103   WILSON       DAWES           "Sales Manager"
     120121   IRENIE       ELVISH          "Sales Rep. II"
     ...
```

```
121144    RENEE        CAPACHIETTI     "Sales Manager"
121145    DENNIS       LANSBERRY       "Sales Manager"
```

4. Exploring Format Storage Options

User-defined formats are stored in the **formats** catalog in the **Work** library, **work.formats**. Use the SAS Help Facility or product documentation to explore permanent format catalogs in PROC FORMAT.

What option enables you to store the formats in a permanent library? _____

What option causes SAS to look for formats in permanent libraries? _____