

ST555 Homework 3

When you have completed this HW, submit via Moodle the following:

- 1) Your SAS program (.sas) that contains the answers to the fill in the blanks as comments. Be sure to include the question number and letter for each comment.
- 2) Your SAS log
- 3) Your Results, generate your results using ods as a .pdf or .rtf file

QUESTION 1:

The file 'ffprez.sas7bdat' contains survey data about the first female president of the US. Respondents were asked from which party would the first female president belong, and what year would she be elected. The variables are candidate_party (female president party), resp_party (respondent's party) year (year female president will be elected), gender (respondent's gender), and age (respondent's age).

(a) Save the dataset in the directory of your choice. Use a LIBNAME statement to create a permanent library pointing to the same directory where the dataset is saved. For simplicity, use the libref 'HW'.

(b) Use the PRINT procedure to print the dataset without the observation numbers

(c) Now write a PRINT procedure that only prints the females respondents. From the log, how many observations are in this subset? 144

(d) Write a PRINT procedure that only prints male Republican respondents. How many male republicans are in this survey? 26

(e) Write a PRINT procedure that only prints responses from respondents that claim to be either democrat or republican, and do not think a woman will be president until after 2032. There should only be 14 observations printed. Depending on how you write this where statement, you may need to use parentheses to get the proper order of operations.

(f) A Missing value for the year variable means that the respondent does not think there will ever be a female president. Print all the respondents that think there will never be a female president.

QUESTION 2:

For question 2, you will be using the 'sales.sas7bdat' data set.

The following code has syntax warnings and errors:

```
DTA Sales;  
set orion.Sales;  
where Country='US' AND Salry <= 26000  
RUN;
```

```
PROC PRINT data=Salse;  
RUN;
```

(a) Troubleshoot the log and as a comment, identify all the warnings and errors. Correct all the errors. How many observations were printed after the code was corrected? 13

(b) Add an ID statement to the PRINT Procedure so that the Employee_ID is the identifier instead of the Obs number and only print Employees who were hired after January 1, 1998.

QUESTION 3:

For question 3, you will be using the 'employee_payroll.sas7bdat' data set.

(a) Sort the data set 'Payroll' by Employee_Gender, and within gender by Salary in descending order. Save the sorted observations into a temporary data set named Salary.

(b) Write a PRINT procedure to display only the observations for former employees (those without missing values for Employee_Term_Date) who earn more than \$50,000. Group the report by Employee_Gender, and include a total and subtotals for Salary. Display only the variables Employee_ID, Salary, and Marital_Status. How many observations are there? 9

QUESTION 4: Selecting Only the Extreme Observations Output from the UNIVARIATE Procedure

a) Write a PROC UNIVARIATE step to validate Product_ID in orion.shoes_tracker.

b) Before the PROC UNIVARIATE step, add the following ODS statement:

```
ods trace on;
```

c) After the PROC UNIVARIATE step add, the following ODS statement:

```
ods trace off;
```

d) Submit the program and notice the trace information in the SAS log.

What is the name of the last output added in the SAS log? MissingValues

e) Add an ODS SELECT statement immediately before the PROC UNIVARIATE step to select only the Extreme Observation output object. Documentation about the ODS TRACE and ODS SELECT statements can be found in the SAS Help facility and in the online documentation.

f) Submit the program to create the following PROC UNIVARIATE report:

The UNIVARIATE Procedure					
Variable: Product_ID (Product ID)					
Extreme Observations					
-----Lowest-----			-----Highest-----		
Value	Obs		Value	Obs	
2.20200E+10	4		2.2020E+11	6	
2.20200E+11	1		2.2020E+11	7	
2.20200E+11	2		2.2020E+11	9	
2.20200E+11	3		2.2020E+11	10	
2.20200E+11	5		2.2020E+12	8	

