PSTAT 130: SAS Summer 2017

Homework 4

Due: Thursday September 7th
Instructions: For the General Knowledge Questions, answer as succinctly as possible. For the Programming Assignment, show your code as well as a partial screenshot of your output.

Programming Assignment

1. Creating Frequency Reports

a. Use PROC FREQ to create a report using the **sanfran** data set that displays the frequency count for each **Destinatation**, and a separate frequency count for each **DepartDay**. Add an appropriate title to each frequency table. Your output should look like this:

SAS Output

Flights from San Francisco

The FREQ Procedure

Destination	Frequency	Percent	Cumulative Frequency	Cumulative Percent
ANC	10	19.23	10	19.23
HND	8	15.38	18	34.62
HNL	3	5.77	21	40.38
RDU	6	11.54	27	51.92
SEA	25	48.08	52	100.00

Flights from San Francisco by Day of Week

The FREQ Procedure

DepartDay	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	6	11.54	6	11.54
2	13	25.00	19	36.54
3	5	9.62	24	46.15
4	7	13.46	31	59.62
5	7	13.46	38	73.08
6	8	15.38	46	88.46
7	6	11.54	52	100.00

b. You can specify many options in the TABLES statement to control the calculations and appearance of a frequency table. The NOCUM option suppresses the printing of the cumulative frequencies and cumulative percentages. Modify the program from Exercise 1.a above to repeat the frequency tables, but without the cumulative frequencies. Part of your output will look like this:

Partial SAS Output

Flights from San Francisco

The FREQ Procedure

Destination	Frequency	Percent
ANC	10	19.23
HND	8	15.38
HNL	3	5.77
RDU	6	11.54
SEA	25	48.08

c. Use PROC FREQ to create a two-way frequency table using the **sanfran** data set that displays the frequency count for each **Destination** by **DepartDay** combination. SAS Output looks like bleow:

The FREQ Procedure

Table of DepartDay by Destination

DepartDay / Frequency / Percent / Row Pct / Col Pct	ANC	HND	HNL	RDU	SEA	Total
1	0 0.00 0.00 0.00	1 1.92 16.67 12.50	0 0.00 0.00 0.00	2 3.85 33.33 33.33	3 5.77 50.00 12.00	6 11.54
2	3 5.77 23.08 30.00	2 3.85 15.38 25.00	0 0.00 0.00 0.00	1 1.92 7.69 16.67	7 13.46 53.85 28.00	13 25.00
3	1 1.92 20.00 10.00	1 1.92 20.00 12.50	0 0.00 0.00 0.00	1 1.92 20.00 16.67	2 3.85 40.00 8.00	5 9.62
4	1 1.92 14.29 10.00	3 5.77 42.86 37.50	0 0.00 0.00 0.00	0 0.00 0.00 0.00	3 5.77 42.86 12.00	7 13.46
5	3 5.77 42.86 30.00	1 1.92 14.29 12.50	1 1.92 14.29 33.33	1 1.92 14.29 16.67	1 1.92 14.29 4.00	7 13.46
6	2 3.85 25.00 20.00	0 0.00 0.00 0.00	2 3.85 25.00 66.67	0 0.00 0.00 0.00	4 7.69 50.00 16.00	8 15.38
7	0 0.00 0.00 0.00	0 0.00 0.00 0.00	0 0.00 0.00 0.00	1 1.92 16.67 16.67	5 9.62 83.33 20.00	6 11.54
Total	10 19.23	8 15.38	3 5.77	6 11.54	25 48.08	52 100.00

2. Validating Data with PROC FREQ

a. PROC FREQ is useful in checking the validity and completeness of data (i.e., invalid values stand out). Use PROC FREQ to check the validity of the variables **Gender** and **JobCode** in the **mechanics** data set.

1) What do you notice about the values of the variable **Gender**?

2) What do you notice about the values of the variable **JobCode**?

b. Modify the previous report to display the frequency count for each **Gender** by **JobCode**. What are the **JobCode** values for the invalid values of **Gender**? (The output is not shown because it provides the answer.)

3. Creating Basic Summary Reports

a. Generate a PROC MEANS report using the **sanfran** data set as input to display statistics for the variables **CargoRev** and **TotPassCap** only. Remove any titles currently in effect.

b. Modify the previous report to display the data for each **Destination** (note: you do NOT need to sort the **sanfran** dataset to do this). Include the following statistics (number of observations, mean, median, mode, range and standard deviation. Limit all output to two decimal places.

Partial SAS Output (not all statistics are shown)

The MEANS Procedure

Destination	N Obs	Variable	N	Mean	Std Dev
ANC	10	CargoRev	10	35811.30	4458.74
		TotPassCap	10	257.60	11.69
HND	8	CargoRev	8	78625.50	3251.06
		TotPassCap	8	250.50	8.33
HNL	3	CargoRev	3	59684.00	3464.64
		TotPassCap	3	207.00	0.00
RDU	6	CargoRev	6	37840.00	4787.04
		TotPassCap	6	267.00	0.00
SEA	25	CargoRev	25	13813.32	2316.59
		TotPassCap	25	151.80	4.97

4. Creating a List Report with PROC REPORT

Use PROC REPORT and the **employees** data set to produce a list report with the following characteristics:

- Output should be sent to the Output window.
- The report should display only the variables **Division, City**, and **Salary**.

- Each variable displayed should have a descriptive report column heading.
- Salary should be displayed with dollar signs, commas, and no decimals.
- The columns of the report should be wide enough so that individual data values are not truncated.
- The observations on the report should be ordered by the values of **Division** and **City**.
- The report should be titled Employee Salary Data.

Partial PROC REPORT Output

Employee Salary Data

Division Name	City Based	Salary
AIRPORT OPERATIONS	AUSTIN	$22,000
		$37,000
		$35,000
	BRUSSELS	$16,000
		$38,000
	CARY	$29,000
		$41,000

5. Creating a Summary Report with Grand Total

Use PROC REPORT and the **employees** data set to produce a summary report with the following characteristics:

- The report should display only the variables **Division, City**, and **Salary**.
- Each variable displayed should have a descriptive report column heading.
- Salary should be displayed with dollar signs, commas, and no decimals.
- The columns of the report should be wide enough so that individual data values are not truncated.
- The observations on the report should be summarized by the values of **City** for each **Division**.
- A grand total should appear at the bottom with a single line above the total and a double line below the total.
- The report should be titled Employee Salary Data by Division / City.

Partial PROC REPORT Output (Top of Report)

Employee Salary Data by Division / City

Division Name	City Based	Salary
AIRPORT OPERATIONS	AUSTIN	$94,000
	BRUSSELS	$54,000
	CARY	$2,510,000
	COPENHAGEN	$254,000
	FRANKFURT	$285,000
	GENEVA	$72,000
	LONDON	$122,000
	PARIS	$147,000
	ROCKVILLE	$79,000
	ROME	$112,000
	SYDNEY	$108,000
	TOKYO	$73,000
	TORONTO	$137,000
CORPORATE OPERATIONS	ATLANTA	$105,000
	CARY	$210,000

Partial PROC REPORT Output (Bottom of Report)

	PITTSBURGH	$52,000
	ROCKVILLE	$81,000
	SAN FRANCISCO	$41,000
	SAN JOSE	$21,000
	SINGAPORE	$63,000
	TOKYO	$101,000
	TORONTO	$83,000
		$16,290,000

6. Creating a One-Dimensional Frequency Report

Use PROC TABULATE and the **employees** data set to produce a summary report that displays a frequency count for the variable **Division** with an appropriate title.

7. Creating a Two-Dimensional Frequency Report

Modify the previous report to do the following:

- subset the data to only display divisions that have the word OPERATIONS in the name
- display the variable **City** in the row dimension
- add row and column totals
- add an appropriate title

8. Creating a Report on an Analysis Variable - Modify the previous report to do the following:

- display the mean of the variable **Salary** in the column dimension
- display the overall mean of the variable **Salary** in the column dimension

- display the data with dollar signs, commas, and no digits after the decimal point
- add an appropriate title