STAT 480 Final Exam (PART II)

(Aug 8th 0:00 AM - Aug 11th 11:59 PM)

Read the following instructions carefully before get started.

Instructions:

- The exam and dataset will be made available to you at 0:00AM, Aug 8th and it is due electronically before 11:59PM, Aug 11th. LATE EXAMS WILL BE PENALIZED 10% FOR EACH HOUR LATE!

- This exam includes 2 problems. You should submit a .sas file in the drop box on CANVAS before deadline. Make you codes readable and include some comments in your program if necessary. Missing any part of the exam or including redundant codes will result in losing some points.

- You may NOT consult with any other person in any way, electronically or in person, about this exam. If you have questions about the exam, you can email me or come to my office hour.

- Turning in this exam indicates that you have read and agree with the following statement:
  I affirm that the report I have submitted is my work, and my work only, and that I have neither given nor received any help or information from any other person. I have cited any written sources I used to complete my report. I recognize that giving or receiving any help on this exam constitutes cheating and will result in my receiving a zero grade on this exam, as well as potentially other consequences under PSUs Policy on Academic Integrity.

Problem 1. (40 pts) Refer to Table 1 to complete the following questions.

Name	Class	Homework	Midterm1	Midterm2	Project
Justin	1	.	88	90	82
Christal	1	100	85	80	88
Rebecca	1	80	77	.	65
Easton	2	90	.	76	73
Jacob	2	60	67	53	70
Cindy	2	85	83	66	90

Table 1: Grades

1. Create a new folder named `Final` in your computer. Specify a library reference name for this library and read the dataset in Table 1 into the library.

2. Create a new variable *Avgmid*, which is the average of the two midterms. If one of the two midterms is a missing value, I expect the average to be the grade of non-missing one. For example, if John misses midterm1 and his grade for midterm2 is 90, then *Avgmid* for John should be 90. And then calculate the final score (denoted by the variable *Final*) according to the following rule:

$$\text{Final} = 30\%\text{Homework} + 40\%\text{Avgmid} + 30\%\text{Project}$$

If *Homework* is a missing value, then the final score is defined to be

$$\text{Final} = 50\%\text{Avgmid} + 50\%\text{Project}$$

3. Use *Means* Procedure to generate a summary dataset, which should includes the mean, median and standard deviation for the final score of students in Class 1, students in Class 2 and all of the students.

4. Use *Report* Procedure to generate a report for the summary dataset, use *Define* Statement to define labels for the variables.

5. From the report you generated, what are the mean, median and standard deviation for the final score of students in Class 1, students in Class 2 and all of the students? Please write answers in the comments of your .sas file.

Problem 2. (40 pts) We have the following small dataset.

Item	Shop	Year	Date	Price
Cellphone	A	2013	02/25/2013	$259.99
Vacuum Cleaner	W	2014	09/30/2014	$78.29
Toaster	A	2014	07/12/2014	$32.35
Microwave	E	2013	12/05/2013	$87.40
Headphone	E	2014	01/03/2014	$96.72
Monitor	A	2013	03/21/2013	$275.99
Clothes Iron	W	2013	10/04/2013	$39.73

Table 2: Shopping List

The columns, from left to right, represent the following variables (see Table 3):

Variable Name	Description	Variable Type
Item	Item name	Character
Shop	Code for shop	Character
Year	Year when bought	Numeric
Date	Date when bought	Numeric
Price	Price of the item	Numeric

Table 3: Description

1. Use proper method to read the dataset in Table 2 into the library and generate two reports using *Report* Procedure and one two-way frequency table using *Freq* Procedure. I want you to create reports and table looking similar to Figure 1. (See next page)

   Pay attention to:

   (a) the group of items, (**hint**: the items are grouped by *Shop* and *Year*);
   (b) the label of variables;
   (c) the format of variables, (**hint**: you might use *Format* Procedure to define your own informat and format);
   (d) the title;
   (e) there is one variable with label "Number of Items" in the summary report.

2. What is the meaning of the four numbers in the cell *Amazon× 2014*? Please write answers in the comments in your .sas file.

Detailed Report of Items Bought in 2013 and 2014

Shop	Year	Item	Date when bought	Price
Amazon	2013	Cellphone	02/25/2013	$259.99
		Monitor	03/21/2013	$275.99
	2014	Toaster	07/12/2014	*$32.35
Ebay	2013	Microwave	12/05/2013	*$87.40
	2014	Headphone	01/03/2014	*$96.72
Walmart	2013	Clothes Iron	10/04/2013	*$39.73
	2014	Vacuum Cleaner	09/30/2014	*$78.29

Summary Report of Items Bought in 2013 and 2014

Shop	Year	Number of Items	Sum of Price
Amazon	2013	2	$535.98
	2014	1	*$32.35
Ebay	2013	1	*$87.40
	2014	1	*$96.72
Walmart	2013	1	*$39.73
	2014	1	*$78.29

Table of Shop by Year

Frequency Percent Row Pct Col Pct		Year		
	Shop	2013	2014	Total
	Amazon	2 28.57 66.67 50.00	1 14.29 33.33 33.33	3 42.86
	Ebay	1 14.29 50.00 25.00	1 14.29 50.00 33.33	2 28.57
	Walmart	1 14.29 50.00 25.00	1 14.29 50.00 33.33	2 28.57
	Total	4 57.14	3 42.86	7 100.00

4

Figure 1: Expected output of Problem 2