Lab 10

Stat 426

Spring 2021

Instructions

Complete all questions. To prepare for the randomly collected lab, follow the instructions on the class website to prepare the work for submission. These submission rules will apply to all labs throughout the semester.

Combining datasets

Concatenation

(1) Create a library, name of your choice (up to 8 characters max), to `S:\Courses\stat-renaes\Stat426\data1` or your own folder if you downloaded the zipped folder of datasets called `sasdatafiles`

(2) Create a new dataset called `thirdqtr` by concatenating (like-structured datasets) the following datasets from your library: `mnth7_2007`, `mnth8_2007`, and `mnth9_2007`

Look at the log to answer the following questions:
How many observations in `thirdqtr` are from mnth7_2007?
How many observations in `thirdqtr` are from mnth8_2007?
How many observations in `thirdqtr` are from mnth9_2007?

(3) Use `PROC PRINT` to verify it looks like the following bit of output

The SAS System

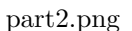| Obs	Order_ID	Order_Type	Employee_ID	Customer_ID	Order_Date	Delivery_Date
1	1242691897	2	99999999	90	02JUL2007	04JUL2007
2	1242736731	1	121107	10	07JUL2007	07JUL2007
3	1242773202	3	99999999	24	11JUL2007	14JUL2007
4	1242782701	3	99999999	27	12JUL2007	17JUL2007
5	1242827683	1	121105	10	17JUL2007	17JUL2007
6	1242836878	1	121027	10	18JUL2007	18JUL2007
7	1242838815	1	120195	41	19JUL2007	19JUL2007
8	1242848557	2	99999999	2806	19JUL2007	23JUL2007
9	1242923327	3	99999999	70165	28JUL2007	29JUL2007
10	1242938120	1	120124	171	30JUL2007	30JUL2007

(4) Look at the datasets in your library called `sales` and `nonsales` using `PROC CONTENTS`. Example to help is:

```
PROC CONTENTS data=herc.sales;
run;
PROC CONTENTS data=herc.nonsales;
run;
```

What are the names of the two variables that are different between the two datasets?

(5) Create a new data set called `allemps` by concatenating these un-like structures and using the RENAME option so that the variables that are different in `nonsales` are renamed to match the ones in `sales`

(6) Use `PROC PRINT` to view the new dataset to verify it has 400 observations

Interleaving (which is a presorted concatenation)

(7) Use PROC SORT on the following data sets in your library: `shoes_eclipse` and `shoes_tracker` (use PROC CONTENTS to see the descriptor portion of the datasets). The BY variable for both to sort by is `Product_Name`. You have to sort before interleaving these files or you will get an error and the log will state that you have to sort first. Interleave these together with a DATA step, using SET and a BY statement. (Example on class website)

(8) Use PROC PRINT to view the data and verify it looks like the following: part2.png

Obs	Product_Group	Product_Name	Supplier_ID
1	Eclipse Shoes	Atmosphere Imara Women's Running Shoes	1303
2	Eclipse Shoes	Atmosphere Shatter Mid Shoes	1303
3	Eclipse Shoes	Big Guy Men's Air Deschutz Viii Shoes	1303
4	Eclipse Shoes	Big Guy Men's Air Terra Reach Shoes	1303
5	Eclipse Shoes	Big Guy Men's Air Terra Sebec Shoes	1303
6	Eclipse Shoes	Big Guy Men's International Triax Shoes	1303
7	Eclipse Shoes	Big Guy Men's Multicourt Ii Shoes	1303
8	Eclipse Shoes	Cnv Plus Men's Off Court Tennis	1303
9	Tracker Shoes	Hardcore Junior/Women's Street Shoes Large	14682
10	Tracker Shoes	Hardcore Men's Street Shoes Large	14682

Merging

(9) From your library, use PROC CONTENTS to look at and determine the common variable of the following files: `orders` and `order_item`

(10) MERGE the two listed files in #9 to create a new data set called `masterorders`, and MERGE them BY the common variable you find in PROC CONTENTS

(11) Check the log to make sure that your newly merged data set has 732 observations and 12 variables