

Homework #8- Data Transformaton Part II; Debugging Techniques

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: Data Transformations II

1. Changing a Variable's Data Type

The data set **orion.US_newhire** contains information about newly hired employees.

Partial **orion.US_newhire**

	ID	Telephone	Birthday
	120-012-40-4928	5467887	05DEC1972
	120-012-83-3816	6888321	03MAY1969
	120-341-44-0781	9418123	23NOV1976
	120-423-01-7721	7839191	28JUN1971

Partial PROC CONTENTS Output of **orion.US_newhire**

Variables in Creation Order			
#	Variable	Type	Len
1	ID	Char	15
2	Telephone	Num	8
3	Birthday	Char	9

a. Create a new data set from **orion.US_newhire**.

- Name the new data set **US_converted**.
- Remove the embedded hyphens in **ID**.
- Convert **ID** to a numeric value.
- Convert **Telephone** to character and place a – (hyphen or dash) between the third and fourth digits.
- Convert **Birthday** to a SAS date value.

b. Print **US_converted** with an appropriate title and use PROC CONTENTS to check the variables types.

Partial **US_converted** (10 Total Observations)

US New Hires		
ID	Telephone	Birthday
120012404928	546-7887	4722
120012833816	688-8321	3410
120341440781	941-8123	6171
120423017721	783-9191	4196

Partial PROC CONTENTS Output

Variables in Creation Order			
#	Variable	Type	Len
1	ID	Num	8
2	Telephone	Char	8
3	Birthday	Num	8

Part II- Supplemental exercises for STAT 625 and Honors credit

2. Calculating Statistics for Missing, Median, and Highest Values

The data set **orion.orders_midyear** contains an observation for each customer, with the total retail value of the customer's monthly orders for the first half of the year.

Partial **orion.orders_midyear** (24 Total Observations)

Customer_ID	month1	month2	month3	month4	month5	month6
5	213.1	.	478.0	525.80	394.35	191.79
10	188.1	414.09	2876.9	3164.59	2373.44	169.29
11	78.2	70.38
12	135.6	.	117.6	129.36	97.02	122.04
18	.	.	29.4	32.34	24.26	.

- Orion Star wants to look at information about the median order and the top two months' orders, but only for frequent customers.
 - Create a data set named **work.freqcustomers** that contains the requested statistics.
 - Frequent customers are defined to be those who placed an order in at least five of the six months.
- Print your results with an appropriate title.



The *median* of a set of values is the middle or central value. For example, the median of {1, 200, 3} is the value 3. If the set has an even number of values, the median is the midpoint between the two middle values.



Consult the SAS documentation as needed to learn more about functions that can generate the desired results. It might be particularly useful to look at "Functions and CALL Routines by Category" in *SAS Language Dictionary*.

Partial PROC PRINT Output (9 Total Observations)

Month Statistics on Frequent Customers									
Customer_ID	month1	month2	month3	month4	month5	month6	Month_ Median	Month_ Highest	Month_ 2nd Highest
5	213.10	.	478.0	525.80	394.35	191.790	394.35	525.80	478.00
10	188.10	414.09	2876.9	3164.59	2373.44	169.290	1393.77	3164.59	2876.90
12	135.60	.	117.6	129.36	97.02	122.040	122.04	135.60	129.36
34	642.50	.	86.3	94.93	71.20	578.250	94.93	642.50	578.25
41	134.00	119.20	313.0	344.30	258.23	120.600	196.11	344.30	313.00