STA 311 Quiz 07

Available: Thursday, 10/22/2020, 12:00 PM
Due: Sunday, 10/25/2020, 11:30 PM

This quiz focuses on defining new variables through basic arithmetic operations, mathematical functions, and SAS built-in functions. Some of the questions require you to write the SAS program and then choose the correct answer from the output.

The working data set for this quiz is Nepal trial data. The link to this data in the CSV format is on the course web page. **Please make sure you will have any problem to read this CSV data file into SAS before Friday since most of the questions in this quiz will be based on this data set. I may not be able to answer your email this weekend**.

The code that I used to create the SAS data set is given below:

```
PROC IMPORT OUT= Nepal_Trial_Data
            DATAFILE= "C:\STA311\w09-nepaltrialdata.csv"
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;
```

Problem 1
Write a SAS data step to create a SAS data set based on Nepal_trial_data with the following actions:
(1). Use SAS ROUND function to round off decimal values to one decimal place for variables WT, HT, and ARM.
(2). There are missing values in some of the records in the data set. These missing values were denoted by 99.9 in variables WT and ARM and 999.9 in HT. We want to delete all records where WT > 90.

The above two actions in the data steps will create a clean SAS data set. Let's use **Nepal_Clean_Trial**. How records are there in **Nepal_Clean_Trial**?

A. Number of observations = 887, number of variables = 14
B. Number of observations = 867, number of variables = 12
C. Number of observations = 887, number of variables = 16
D. Number of observations = 857, number of variables = 12
E. Number of observations = 877, number of variables = 12

Correct Answer: A.

Problem 2

Create a SAS data set under the name **Nepal_Trial_Dates** with the following action in the data step:

(1). Define a SAS date variable from variables DAY, MONTH, and YEAR using names VIS_DATE.

(2). Format the SAS date variable VIS_DATE using MMDDYY10.

(3). DROP variables DIED, ALIVE, BF, LIT, ID, DAY, MONTH, and YEAR;

We can make a frequency table of VIS_DATE in **Nepal_Trial_Dates**. Which date had the lowest number of visits in the data set?

A. 07/01/1946
B. 07/17/1946
C. 08/01/1946
D. 01/24/1947
E. 03/29/1947

Correct Answer: A

Problem 3

Since the data set **Nepal_Trial_Dates** has multiple records for each subject, we want to create a SAS data set that contains the record of each subject's first visit. To obtain this data set,

(1) sort data set **Nepal_Trial_Dates** by ID and VIS_DATE.
(2) Create a SAS data set using the name **Nepal_Trial_FIRST_VIS** to store the record associated with the first visit of each subject in **Nepal_Trial_Dates.**

Based on the information in the log file, how many children were in data set **Nepal_Trial_FIRST_VIS**?

A. 887
B. 197
C. 77
D. 2258
E. 1000

Correct Answer: B

Problem 4

Based on the SAS data set **Nepal_Trial_FIRST_VIS**, we want to know how many boys are in this data set. To obtain this information, you can construct a frequency table of variable SEX. Note that the numerical coding used in this data set: 1 = male and 2 = female. To get a nicer frequency table of variable SEX, you can define a format for SEX and then format variable SEX. Here are the steps:

(1) Use PROC STEP to define a format using coding: 1 = male and 2 = female.
(2) Write a procedure step PROC FREQ to tabulate SEX with a specified format.

Based on the frequency table, the number of boys in the data set is:

A. 467
B. 37
C. 1192
D. 104
E. 528

Correct Answer: D.

Problem 5
This question focuses on grouping the mother's age in the data set **Nepal_Trial_FIRST_VIS**. We first group numerical age mothers. Here is the grouping criterion: 1 = "17-24", 2 = "25-30", 3 = "31 – 37", 4 = "38-55". We define a grouped age using the name **grp_age**. Then use PROC FREQ to tabulate **grp_age**.

How many mothers are aged between 38 and 55?

A. 54
B. 80
C. 45
D. 47
E. 16

Correct Answer: E.

Problem 6.
This question calculates the duration between the first visit and the last visits of each child in the data. To do this, we create a data set **Nepal_Trial_LAST_VIS** based on **Nepal_Trial_Dates** and then merge with **Nepal_Trial_FIRST_VIS** by ID.
(1). Sort data set **Nepal_Trial_Dates** by ID and VIS_DATE
(2). Use automatic variable LAST.ID to create data set **Nepal_Trial_LAST_VIS.**
(3). Merge data sets **Nepal_Trial_LAST_VIS** and **Nepal_Trial_FIRST_VIS** by ID using name **Nepal_Trial_Duration** and using KEEP and RENAME options to retain information of ID and visit dates to indicate first and last visits.
(3) duration = last_vis – first_vis

What is the inter-quartile range (IQR = Q3 – Q1)?

A. 7
B. 1
C. 493
D. 8
E. 500

Correct Answer: D.