SAS Data sets for Q1 to Q2: WORK.Manager and WORK.Employee

Q1. Creating a data set 'WORK.Allemployees' by using DATA STEP. It is required to stack two SAS data sets 'WORK.Manager' and 'WORK.Employee'. Observing the resulting SAS table and explain the difference between

```
(A)   SET  WORK.Manager WORK.Employee;
(B)   SET  WORK.Manager; SET WORK.Employee;
```

Q2. Using the procedure 'PROC APPEND' to stack two SAS data sets 'WORK.Manager' and 'WORK.Employee'. It is required to create the following resulting data sets based on the different conditions:

(a) 'WORK.Allemployees_1' by specifying the option 'DATA=WORK.Manager' in the 'PROC APPEND'.
(b) 'WORK.Allemployees_2' by specifying the option 'DATA=WORK.Employee' in the 'PROC APPEND'.
(c) 'WORK.Allemployees_3' by (1)first dropping the column 'tenure' in the table 'WORK.Employee' to get a new data set 'WORK.Employee_new' and (2)stacking it with the data set 'WORK.Manager' by specifying the option 'DATA= WORK.Employee_new' in the 'PROC APPEND'

Observing SAS Log Window and compare three resulting tables.

SAS Data set for Q3 to Q4: WORK.Goods_sell

Q3. Using the 'PROC SURVEYSELECT' and DATA STEP approach respectively to draw 10% random samples from the SAS data set 'WORK.Goods_sell'.

Q4. Applying the 'PROC SURVEYSELECT' with 'STRATA' option to draw stratified samples from the data set 'WORK.Goods_sell'. The stratified variable is 'Transaction_type' and sample size for each Transaction type ('AMEX', 'EMAI', 'MASTERCARD', 'PAYPAL' and 'VISA') are 100, 50, 200, 100 and 200 respectively. Also use the 'PROC FREQ' to very your result.

SAS Data set for Q5 to Q6: WORK.Coffee_data

Q5.Using the 'PROC MEANS' to check the number of missing value for the variables 'income' and 'Cups_Per_Week'. Then creating a new table 'WORK.Coffee_data_fix' which only contains the variables 'Person_ID',

'income', 'Cups_Per_Week' and two missing indicators for variables 'income' and 'Cups_Per_Week' respectively.

Q6. Using the 'PROC STDIZE' on the data set 'WORK.Coffee_data_fix' created in Q5 to solve the following questions:

(1) Replacing the missing values of the variable 'income' with its mean value.
(2) Replacing the missing values of the variable 'Cups_Per_Week' with its median value.
(3) Do 'RANGE' standardization for two variables above.