Introduction to SAS Workshop

Assignment 2

In this exercise we will use an Excel file **bank.xls**, which is located in the **Labdata** folder. The codebook for this dataset is shown below:

This data set originally came from SPSS. There is information on 474 bank employees:

| **Variable** | **Description** | **Type** | **Codes** |
| --- | --- | --- | --- |
| ID | Employee Code | Num |  |
| SALBEG | Salary when hired | Num |  |
| SEX | Sex of employee | Num | 0=Male  1=Female |
| TIME | Year hired | Num | Ranges from 1964 to 1998 |
| AGE | Current age in years | Num |  |
| SALNOW | Current salary | Num |  |
| EDLEVEL | Education level | Num | Number of years of education |
| WORK | Number of years on the job | Num | Ranges from 0 to about 40 years |
| JOBCAT | Job classification | Num | 1= Clerical  2= Office trainee  3= Security officer  4= College trainee  5= Exempt employee  6= MBA trainee  7= Technical |
| MINORITY | Minority status | Num | 0=Non-minority  1=Minority |
| SEXRACE | Combination of Sex & Race categories | Num | 1=White male  2=Minority male  3=White female  4=Minority female |

1. Use Proc Import Commands to read in **bank.xls** and save it as a temporary dataset in the Work Library. Name your file, **bank** (or **Work.bank**). The commands to do this are contained in the Workbook, p. 19.

Now use the temporary file **Work.bank** to do the following exercises.

1. Use Proc Contents to get information about the dataset.
   * What is the name of the dataset?
   * How many observations are there in this dataset?
   * How many variables are in the dataset?
   * What type are the variables (character or numeric)?
2. Use Proc Print to get a listing of the first 15 cases in the **bank** data set.
3. Using Proc Means, obtain the mean and standard deviation of AGE and SALNOW for all employees.
4. Using Proc Means, with a class statement, obtain the mean and standard deviation of AGE and SALNOW for each job category.
5. Use Proc Freq to determine how many employees there are in each of the seven job categories in this bank.
6. Get a crosstab of sex by job category to determine how many males and females are in each job category.
   * How many women are in each job category?
   * What percent of each job category are women, and what percent of each job category are men? (Use Col Pct)
   * What percent of the women are in each job category? (Use Row Pct)
7. Using Proc Univariate, generate histograms of current salary (SALNOW) for all participants, and separately for males and females. Is current salary approximately normally distributed?
8. Obtain the mean, median, standard deviation, min, and max of current salary for males and females separately.
   * What do you notice about the means and the medians? Which one is larger, the mean or median? Why?
9. Use SAS commands to Save your temporary bank dataset to a permanent dataset in the SASDATA2 folder.
   * Check your SASDATA2 folder to be sure that the bank dataset was saved there.
   * What is the name of the file that was saved?
   * Run all of the commands in your command file at once to be sure that they run without error.

Where appropriate, answer each question in words, and provide output for the question. It is not necessary to print all of the output generated for each question.

* Make a printed copy of your completed assignment and hand it in.
* Attach your printed SAS code.
* Please be sure to include your name on the assignment.