Statistics 6430 (SAS)
Homework 1
Due: Friday, July 20

Complete the following problems in a commented SAS program. Problems should be clearly marked with comments. Clearly indicate and explain your answer as comments within the code. Include any requested output as a single PDF. The program name should include your name or computing ID. When reading data in, the external files should not be modified in any way.

Problem 1: For each of the following, determine which are valid variable names and explain your reasoning.

A. `Hgt-Lbs`

B. `Xyz123abc987`

C. `789cba321zyx`

D. `FieldData`

E. `field-data`

F. `fIeLdDaTa`

G. `Data_2006`

H. `2006Data`

I. `Data2005-2006`

J. `Data06`

K. `LastWeek'sData`

L. `DataFromLastWeek`

M. `Income`

N. `Average_Income`

O. `Average_Income_In_Adjusted_Dollars`

Problem 2: Using the valid variable names from the problem above, which, if any, could not be used together in a dataset? Explain.

Problem 3: Explain what makes the following statements true or false.

A. A variable is one of three types: character, numeric, or mixed.

B. A statement is ended with a semi-colon or at the end of a line.

Problem 4: A code that will be shared among several users reads in a data set with the INPUT statement shown below. For each of the following blocks of code, determine whether it will execute and read the data in correctly. Explain.

```
input ID $ height weight SBP DBP;
```

A.
```
input ID $ height weight SBP /*systolic*/ DBP /*diastolic*/;
```

B.
```
input ID $ height weight SBP */systolic*/; DBP */diastolic*/;
```

C.
```
input ID $ height weight SBP *systolic DBP *diastolic;
```

D.
```
input ID $ height weight SBP *systolic; DBP *diastolic;
```

E.
```
input ID $ height weight SBP DBP;
*SBP: systolic, DBP: diastolic;
```

Problem 5: Consider the DATA step below that reads in data containing two numerical variables (this data set is not provided). Identify and explain the errors in this DATA step that will either keep the code from executing or cause the data to read in incorrectly.

```
data orig-data;
infile OriginalData.txt;
input x1 $ x2 $;
run;
```

Problem 6: An experiment is conducted with two factors – pause time and node speed. Use a DATA step to create a dataset named `comp_exp` that contains two variables – `x1` and `x2`. The dataset should contain all possible pairs of the values for these two factors. The pause times used are 5, 20, and 40 and the node speeds used are 10, 30, and 50.

Problem 7: The files *DataIa.csv*, *DataIb.txt*, and *DataIc.txt* contains the same employee information with different file formatting. For each file, use a FILENAME statement and DATA step to read in the corresponding data to datasets named `empA`, `empB`, and `empC`, all with variables named `ID`, `name`, `dept`, `hire_date`, and `salary`.

Problem 8: Use a FILENAME statement and DATA step to read in the political information contained in the file *DataII.csv* to a dataset named `vote` with variables named `state`, `party`, and `age`.

Problem 9: The file *DataIII.txt* contains account information for a bank – name in columns 1-15, account number in columns 16-20, account balance in columns 21-26, and interest rate in columns 27-30. Use a FILENAME statement and DATA step to read in the data to a dataset named `bank` with variables named `name`, `acct`, `balance`, and `rate`.

Problem 10: The file *DataIV.txt* contains stock price information. The variable information is given in the table below. Use a FILENAME statement and DATA step to read in the data to a dataset named `stocks` with variables named `symbol`, `pdate`, `pprice`, `nshares`, `sdate`, and `sprice`.

Field	Description	Starting column
1	Stock symbol	1
2	Purchase date	5
3	Purchase price	15
4	Number of shares	21
5	Selling date	25
6	Selling price	35