Math 338 Name: Nhi Vu

Kafai – HW 5

We have three data sets, ONE, TWO, and THREE. Each data set contains the variables:

ID (IDNUM in data set TWO)

SEX

DOB

SALARY

In addition, data set TWO contains TAXRATE and WITHHOLD; data set THREE contains HEIGHT and WEIGHT .

1. Write a program to create a single SAS data set ALL which combines the observations from data sets ONE, TWO, and THREE and contains the variables ID, DOB, and SALARY.
2. Write a program to create a new SAS data set for those people born on or before January 1, 1960, and who earn $50,000 or more per year.

ONE

|  |  |  |  |
| --- | --- | --- | --- |
| ID | SEX | DOB | SALARY |
| 01 | M | 10/21/46 | 70000 |
| 02 | F | 11/01/55 | 68000 |
| 07 | M | 01/27/59 | 47000 |

TWO

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| IDNUM | SEX | DOB | SALARY | TAXRATE | WITHHOLD |
| 03 | F | 01/01/33 | 78000 | 0.36 | 23000 |
| 05 | M | 03/07/29 | 37000 | 0.25 | 9000 |

THREE

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | SEX | DOB | SALARY | HEIGHT | WEIGHT |
| 04 | M | 10/23/49 | 65000 | 68 | 158 |
| 06 | F | 07/04/65 | 55000 | 74 | 202 |

/\*\*\*\*READ ALL THE DATA SETS\*\*\*NHI VU\*\*\*/

options nodate nonumber LINESIZE=**70**;

**data** ONE;

infile "\\Client\C$\Users\nhivutu\Desktop\Math338\data\ONE.txt";

input ID SEX $ DOB mmddyy8. SALARY;

**run**;

**data** TWO;

infile "\\Client\C$\Users\nhivutu\Desktop\Math338\data\TWO.txt";

input IDNUM SEX $ DOB mmddyy8. SALARY TAXRATE WITHHOLD;

**run**;

**data** THREE;

infile “\\Client\C$\Users\nhivutu\Desktop\Math338\data\THREE.txt";

input ID SEX $ DOB mmddyy8. SALARY HEIGHT WEIGHT;

**run**;

**( Note**: data sets ONE and THREE are already sorted in ID order and data set TWO is already sorted in IDNUM order.)

/\*\*\*PART A\*\*\*NHI VU\*\*\*/

**data** all (keep= ID DOB SALARY);

merge ONE TWO (rename=(IDNUM=ID)) THREE;

by ID;

**run**;

**proc** **print** data=all;

title "PART A";

format ID z2. DOB mmddyy10. SALARY dollar7.;

**run**;

PART A

Obs ID DOB SALARY

1 01 10/21/1946 $70,000

2 02 11/01/1955 $68,000

3 03 01/01/1933 $78,000

4 04 10/23/1949 $65,000

5 05 03/07/1929 $37,000

6 06 07/04/1965 $55,000

7 07 01/27/1959 $47,000

/\*\*\*PART B \*\*\* NHI VU\*\*\*/

**data** newList;

set all;

if DOB <=**01**/**1**/**1960** and SALARY >=**50000** then output;

**run**;

**proc** **print** data=newList;

format ID z2. DOB mmddyy10. SALARY dollar7.;

title "PART B";

**run**;

**quit**;

PART B

Obs ID DOB SALARY

1 01 10/21/1946 $70,000

2 02 11/01/1955 $68,000

3 03 01/01/1933 $78,000

4 04 10/23/1949 $65,000