**Homework Reading Raw Data**

1. The raw data file newemps.csv is in the Data Sets subdirectory. The first 7 records on this file are:

Satyakam,Denny,Sales Rep. II,26780

Monica,Kletschkus,Sales Rep. IV,30890

Kevin,Lyon,Sales Rep. I,26955

Petrea,Soltau,Sales Rep. II,27440

Marina,Iyengar,Sales Rep. III,29715

Shani,Duckett,Sales Rep. I,25795

Fang,Wilson,Sales Rep. II,26810

The variables on the file are (in the order they appear):

|  |  |  |
| --- | --- | --- |
| Variable Name | Variable Type | Variable Length |
| LastName | Character | 18 |
| FirstName | Character | 12 |
| Title | Character | 25 |
| Salary | Numeric | 8 |

1. Write a DATA step that reads the raw data file and creates a SAS data set work.newemps.
2. Include a PROC CONTENTS step in your program to verify that there are 71 observations in the data set work.newemps
3. Include a PROC PRINT step in your program to print the first 7 records on the SAS data set work.emps and verify that the values are correct.
4. The raw data file donation.dat is in the Data Sets subdirectory on the SAS cloud. The first 10 records are:

120265 . . . 25

120267 15 15 15 15

120269 20 20 20 20

120270 20 10 5 .

120271 20 20 20 20

120272 10 10 10 10

120275 15 15 15 15

120660 25 25 25 25

120662 10 . 5 5

120663 . . 5 .

The following variables should be read into the program data vector:

|  |  |  |
| --- | --- | --- |
| Name | Type | Length |
| IDNum | Character | 6 |
| Q1 | Numeric | 8 |
| Q2 | Numeric | 8 |
| Q3 | Numeric | 8 |
| Q4 | Numeric | 8 |

1. Write a DATA step to read this file and create a SAS data set named work.Donation by reading the space delimited raw data file.
2. Include a PROC CONTENTS step to verify that the data set work.Donation has 124 observations.
3. Include a PROC PRINT step to print the first 10 observations on the data set work.Donation to verify that the values are correct.
4. The raw data file supplier.dat is in the Data Sets subdirectory. The first 5 lines of the data file are:

50 Scandinavian Clothing A/S NO

109 Petterson AB SE

316 Prime Sports Ltd GB

755 Top Sports DK

772 AllSeasons Outdoor Clothing US

The following is the layout of the raw data file:

|  |  |  |
| --- | --- | --- |
| Name | Starting Column | Ending Column |
| ID | 1 | 5 |
| Name | 8 | 37 |
| Country | 40 | 41 |

1. Write a DATA step to read this file and create a SAS data set named work.supplier
2. Include a PROC CONTENTS step to verify that the SAS data set work.supplier has 52 Observations
3. Include a PROC PRINT step to print the first five observations of the SAS data set work.supplier and verify that the values are correct.
4. The raw data file custca.csv is in the Data Sets subdirectory. The first 4 records are:

Bill,Cuddy,11171,M,16/10/1986,21,15-30 years

Susan,Krasowski,17023,F,09/07/1959,48,46-60 years

Andreas,Rennie,26148,M,18/07/1934,73,61-75 years

Lauren,Krasowski,46966,F,24/10/1986,21,15-30 years

The file layout (some are in nonstandard format, you should use the data above to find the correct informat):

|  |  |  |
| --- | --- | --- |
| Name | Type | Length |
| First | Character | 20 |
| Last | Character | 20 |
| ID | Numeric | 8 |
| Gender | Character | 1 |
| BirthDate | Numeric | 8 |
| Age | Numeric | 8 |
| AgeGroup | Character | 12 |

1. Write a DATA step to read the raw data file and create a SAS data set work.canada. The data step should translate the birth dates to SAS dates.
2. Write a PROC CONTENTS step to verify that the SAS data set work.canada has 15 observations.
3. Write a PROC PRINT to print the first three observations and verify that the birth dates were translated to SAS dates.
4. Write a second PROC PRINT step to print the first three observations. The step should include a format statement to display the birth dates in the same format they were on the raw data file. Use these results to verify that the correct values are in the SAS data set work.canada.
5. The raw data file sales1.dat is in the Data Sets subdirectory and has employee information for the Australian and U.S. sales staff.

The first 6 observations of the data set:

120102 Tom Zhou M Sales Manager $108,255 AU 08/11/1969 06/01/1989

120103 Wilson Dawes M Sales Manager $87,975 AU 01/22/1949 01/01/1974

120121 Irenie Elvish F Sales Rep. II $26,600 AU 08/02/1944 01/01/1974

120122 Christina Ngan F Sales Rep. II $27,475 AU 07/27/1954 07/01/1978

120123 Kimiko Hotstone F Sales Rep. I $26,190 AU 09/28/1964 10/01/2007

120124 Lucian Daymond M Sales Rep. I $26,480 AU 05/13/1959 03/01/2007

The record layout is shown in the table below (some are in nonstandard format, you should use the data above to determine the correct informat):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Description | Beginning Column | Ending Column | Type |
| id | Employee ID | 1 | 6 | Numeric |
| FirstName | First Name | 8 | 20 | Character |
| LastName | Last Name | 21 | 38 | Character |
| Gender | Gender | 40 | 40 | Character |
| JobTitle | Job Title | 43 | 62 | Character |
| salary | Salary | 64 | 71 | Numeric |
| country | Country | 73 | 74 | Character |
| BirthDate | Birth Date | 76 | 85 | Numeric |
| HireDate | Hire Date | 87 | 96 | Numeric |

1. Write a DATA step to read this data and create a SAS data set work.sales that contains the data.
2. Include a PROC CONTENTS step to verify that the SAS data set work.sales has 165 observations.
3. Include a PROC PRINT step to examine the first six records on the SAS data set work.sales.
4. Include a second PROC PRINT step to examine the first six records on the SAS data set work.sales. Include a format statement to display the variables in the same format they were on the raw data file to verify that the first six observations are correct.
5. The raw data file sales2.dat is in the Data Sets subdirectory. The file contains information for the Australian and U.S. sales staff and information for each employee is in three lines of raw data.

The first 6 lines of the file are:

120102 Tom Zhou

Sales Manager 06/01/1989 $108,255

M 08/11/1969 AU

120103 Wilson Dawes

Sales Manager 01/01/1974 $87,975

M 01/22/1949 AU

The file layout:

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Start Column | length | type/format |
| Line 1 | | | |
| Employee ID | 1 | 6 | Numeric |
| First Name | 8 | 12 | Character |
| Last Name | 21 | 18 | Character |
| Line 2 | | | |
| Job Title | 1 | 20 | Character |
| Hire Date | 22 | 10 | Numeric mm/dd/yyyy |
| Salary | 33 | 8 | Numeric e.g. $100,000 |
| Line 3 | | | |
| Gender | 1 | 1 | Character |
| Birth Date | 3 | 10 | Numeric mm/dd/yyyy |
| Country | 14 | 2 | Character |

1. Write a DATA step that uses the raw data file, to create a new SAS data set work.staff2 that contains the variables on the raw data file.
2. Include a PROC PRINT step to display the SAS dataset work.staff2.
3. Include a second PROC PRINT step to display the first two records on the SAS dataset work.staff2. Include a format statement to assure the variable display as they were on the raw data file and verify that these two records contain the correct values.
4. The raw data file sales3.dat is in the Data Sets subdirectory. The file contains employee information for the Australian and U.S. Sales staff. The information for each employee is on two lines of the raw data file. The first 6 lines of the raw data are:

120102 Tom Zhou M Sales Manager

$108.255 AU 11/08/1969 01/06/1989

120103 Wilson Dawes M Sales Manager

$87.975 AU 22/01/1949 01/01/1974

120121 Irenie Elvish F Sales Rep. II

$26.600 AU 02/08/1944 01/01/1974

The record layouts for the first line is the same for both US and Australian employees, but the contents of the second line depends on the country the employee is from:

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Starting Column | Length | Type |
| Line 1 layout (all employees) | | | |
| Employee ID | 1 | 6 | Numeric |
| First Name | 8 | 12 | Character |
| Last Name | 21 | 18 | Character |
| Gender | 40 | 1 | Character |
| Job Title | 43 | 20 | Character |
| Line 2 layout for Australian employees | | | |
| Salary | 1 | 8 | Numeric e.g. $100.000 |
| Country | 10 | 2 | Character |
| Birth Date | 13 | 10 | Numeric dd/mm/yyyy |
| Hire Date | 24 | 10 | Numeric dd/mm/yyyy |
| Line 2 layout for U.S. employees | | | |
| Salary | 1 | 8 | Numeric e.g. $100.000 |
| Country | 10 | 2 | Character |
| Birth Date | 13 | 10 | Numeric mm/dd/yyyy |
| Hire Date | 24 | 10 | Numeric mm/dd/yyyy |

1. Write a DATA step to create a file work.sales that contains the information in the raw data file. Each employees data should be in a single observation.
2. Include a PROC CONTENTS step to examine the descriptor portion on the work.sales data set.
3. Include a PROC PRINT step that prints the first 10 Australian Employees.
4. Include a second PROC PRINT step that prints the first 15 U.S. Employees.