Unit 1-7 Exercises

1. Reading a Comma-Delimited Raw Data File

a. Retrieve the starter program **p107e01**.

b. Add the appropriate LENGTH, INFILE, and INPUT statements to read the comma-delimited raw data file named **`newemps.csv`**.

Partial Raw Data File

```
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
```

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

Name	Type	Length
`First`	Character	12
`Last`	Character	18
`Title`	Character	25
`Salary`	Numeric	8

c. Submit the program to create the following PROC PRINT report:

Partial PROC PRINT Output (First 5 of 71 Observations)

Obs	First	Last	Title	Salary
1	Satyakam	Denny	Sales Rep. II	26780
2	Monica	Kletschkus	Sales Rep. IV	30890
3	Kevin	Lyon	Sales Rep. I	26955
4	Petrea	Soltau	Sales Rep. II	27440
5	Marina	Iyengar	Sales Rep. III	29715

2. Reading a Space-Delimited Raw Data File

a. Write a DATA step to create a new data set named **Work.QtrDonation** by reading the space-delimited raw data file named **donation.dat**.

Partial Raw Data File

```
120265 . . . 25
120267 15 15 15 15
120269 20 20 20 20
120270 20 10 5 .
120271 20 20 20 20
```

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

Name	Type	Length
IDNum	Character	6
Qtr1	Numeric	8
Qtr2	Numeric	8
Qtr3	Numeric	8
Qtr4	Numeric	8

b. Write a PROC PRINT step to create the following report:

Partial PROC PRINT Output (First 10 of 124 Observations)

Obs	IDNum	Qtr1	Qtr2	Qtr3	Qtr4
1	120265	.	.	.	25
2	120267	15	15	15	15
3	120269	20	20	20	20
4	120270	20	10	5	.
5	120271	20	20	20	20
6	120272	10	10	10	10
7	120275	15	15	15	15
8	120660	25	25	25	25
9	120662	10	.	5	5
10	120663	.	.	5	.

3. Using Column Input to Read a Fixed Column Raw Data File

a. Write a DATA step to create a new data set named **`Work.supplier_info`** by reading the fixed column raw data file named **`supplier.dat`**.

Use column input in the INPUT statement to read the fixed column data.

Documentation on column input can be found in the SAS Help and Documentation from the Contents tab (**SAS Products** ⇨ **Base SAS** ⇨ **SAS 9.3 Language Reference: Dictionary** ⇨ **Dictionary of Language Elements** ⇨ **Statements** ⇨ **INPUT Statement, Column**).

Partial Raw Data File

```
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

b. Write a PROC PRINT step to create the following report:

Partial PROC PRINT Output (First 10 of 52 Observations)

Obs	ID	Name	Country
1	50	Scandinavian Clothing A/S	NO
2	109	Petterson AB	SE
3	316	Prime Sports Ltd	GB
4	755	Top Sports	DK
5	772	AllSeasons Outdoor Clothing	US
6	798	Sportico	ES
7	1280	British Sports Ltd	GB
8	1303	Eclipse Inc	US
9	1684	Magnifico Sports	PT
10	1747	Pro Sportswear Inc	US

4. Reading a Comma-Delimited Raw Data File

a. Retrieve the starter program **p107e04**.

b. Add the appropriate LENGTH, INFILE, and INPUT statements to read the comma-delimited raw data file named **custca.csv**.

Partial Raw Data File

```
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
Lauren,Marx,54655,F,18/08/1969,38,31-45 years
```

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

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

c. Add a FORMAT statement and a DROP statement in the DATA step to create a data set that resembles the following when used in the PROC PRINT step:

Partial PROC PRINT Output (First 5 of 15 Observations)

Obs	First	Last	Gender	AgeGroup	Birth Date
1	Bill	Cuddy	M	15-30 years	OCT1986
2	Susan	Krasowski	F	46-60 years	JUL1959
3	Andreas	Rennie	M	61-75 years	JUL1934
4	Lauren	Krasowski	F	15-30 years	OCT1986
5	Lauren	Marx	F	31-45 years	AUG1969

5. Reading a Space-Delimited Raw Data File with Spaces in Data Values

a. Write a DATA step to create a new data set named **`Work.us_customers`** by reading the space-delimited raw data named the following:

Windows or UNIX	**`custus.dat`**
z/OS (OS/390)	**`.workshop.rawdata(custus)`**

Some of the data values contain spaces. Use an option in the INFILE statement to specify that when data values are enclosed in quotation marks, delimiters within the value are treated as part of the data value.

Partial Raw Data File

```
"James Kvarniq" 4 M 27JUN1974 33 "31-45 years"
"Sandrina Stephano" 5 F 09JUL1979 28 "15-30 years"
"Karen Ballinger" 10 F 18OCT1984 23 "15-30 years"
"David Black" 12 M 12APR1969 38 "31-45 years"
"Jimmie Evans" 17 M 17AUG1954 53 "46-60 years"
```

The following variables should be created in the data set **`Work.us_customers`**:

Name	Type	Length
`Name`	Character	20
`ID`	Numeric	8
`Gender`	Character	1
`BirthDate`	Numeric	8
`Age`	Numeric	8
`AgeGroup`	Character	12

b. Add a FORMAT statement in the DATA step to make the **`BirthDate`** resemble a three-character month with a four-digit year.

c. Write a PROC PRINT step with a VAR statement to create the following report:

Partial PROC PRINT Output (First 7 of 28 Observations)

Obs	Name	Gender	Birth Date	AgeGroup	Age
1	James Kvarniq	M	JUN1974	31-45 years	33
2	Sandrina Stephano	F	JUL1979	15-30 years	28
3	Karen Ballinger	F	OCT1984	15-30 years	23
4	David Black	M	APR1969	31-45 years	38
5	Jimmie Evans	M	AUG1954	46-60 years	53
6	Tonie Asmussen	M	FEB1954	46-60 years	53
7	Michael Dineley	M	APR1959	46-60 years	48

6. Reading Missing Values at the End of a Record

a. Write a DATA step to create a new data set named **`Work.prices`** by reading the asterisk-delimited raw data file named **`prices.dat`**.

Some of the records do not have a value for **`UnitSalesPrice`** and the last delimiter is missing.

Documentation on the INFILE statement options can be found in the SAS Help and Documentation from the Contents tab (**SAS Products** ⇨ **Base SAS** ⇨ **SAS 9.3 Language Reference: Dictionary** ⇨ **Dictionary of Language Elements** ⇨ **Statements** ⇨ **INFILE Statement**).

Partial Raw Data File

```
210200100009*09JUN2007*31DEC9999*$15.50*$34.70
210200100017*24JAN2007*31DEC9999*$17.80
210200200023*04JUL2007*31DEC9999*$8.25*$19.80
210200600067*27OCT2007*31DEC9999*$28.90
210200600085*28AUG2007*31DEC9999*$17.85*$39.40
```

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

Name	Type	Length
`ProductID`	Numeric	8
`StartDate`	Numeric	8
`EndDate`	Numeric	8
`UnitCostPrice`	Numeric	8
`UnitSalesPrice`	Numeric	8

b. Write a PROC PRINT step and add a LABEL and a FORMAT statement in the DATA step to create a data set that resembles the following when used in the PROC PRINT step:

Partial PROC PRINT Output (First 10 of 259 Observations)

Obs	Product ID	Start of Date Range	End of Date Range	Cost Price per Unit	Sales Price per Unit
1	210200100009	06/09/2007	12/31/9999	15.50	34.70
2	210200100017	01/24/2007	12/31/9999	17.80	.
3	210200200023	07/04/2007	12/31/9999	8.25	19.80
4	210200600067	10/27/2007	12/31/9999	28.90	.
5	210200600085	08/28/2007	12/31/9999	17.85	39.40
6	210200600112	01/04/2007	12/31/9999	9.35	21.80
7	210200900033	09/17/2007	12/31/9999	6.45	14.20
8	210200900038	02/01/2007	12/31/9999	9.30	20.30
9	210201000050	04/02/2007	12/31/9999	9.00	19.60
10	210201000126	04/22/2007	12/31/9999	2.30	6.50