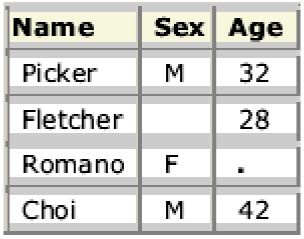
SAS Certificate Base Practice Questions and Detailed Answers

Chapter 1: Basic Concepts Answer Key

1. How many observations and variables does the data set below contain?



Name	Sex	Age
Picker	M	32
Fletcher		28
Romano	F	.
Choi	M	42

a. 3 observations, 4 variables

b. 3 observations, 3 variables

c. 4 observations, 3 variables

d. can't tell because some values are missing

Correct answer: c

Rows in the data set are called observations, and columns are called variables. Missing values don't affect the structure of the data set.

2. How many program steps are executed when the program below is processed?

```
data user.tables;
   infile jobs;
   input date name $ job $;
run;
proc sort data=user.tables;
   by name;
run;
proc print data=user.tables;
run;
```

a. three

b. four

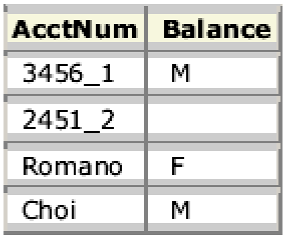
c. five

d. six

Correct answer: a

When it encounters a DATA, PROC, or RUN statement, SAS stops reading statements and executes the previous step in the program. The program above contains one DATA step and two PROC steps, for a total of three program steps.

3. What type of variable is the variable AcctNum in the data set below?



AcctNum	Balance
3456_1	M
2451_2	
Romano	F
Choi	M

a. numeric

b. character

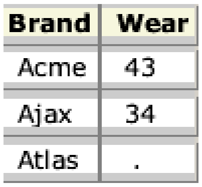
c. can be either character or numeric

d. can't tell from the data shown

Correct answer: b

It must be a **character variable**, because the values contain letters and underscores, which are not valid characters for numeric values.

4. What type of variable is the variable Wear in the data set below?



Brand	Wear
Acme	43
Ajax	34
Atlas	.

a. numeric

b. character

c. can be either character or numeric

d. can't tell from the data shown

Correct answer: a

It must be a **numeric variable**, because the missing value is indicated by a period rather than by a blank.

5. Which of the following variable names is valid?

a. 4BirthDate

b. $Cost

c. _Items_

d. Tax-Rate

Correct answer: c

Variable names follow the same rules as SAS data set names. They can be 1 to 32 characters long, must begin with a letter (A–Z, either uppercase or lowercase) or an underscore, and can continue with any combination of numbers, letters, or underscores.

6. Which of the following files is a permanent SAS file?

a. Sashelp.PrdSale

b. Sasuser.MySales

c. Profits.Quarter1

d. all of the above

Correct answer: d

To store a file permanently in a SAS data library, you assign it a libref other than the default Work. For example, by assigning the libref Profits to a SAS data library, you specify that files within the library are to be stored until you delete them. Therefore, SAS files in the Sashelp and Sasuser libraries are permanent files.

7. In a DATA step, how can you reference a temporary SAS data set named Forecast?

a. Forecast

b. Work.Forecast

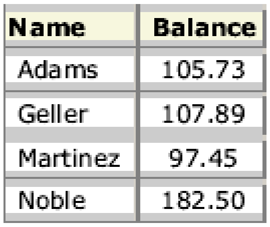
c. Sales.Forecast (after assigning the libref Sales)

d. only a and b above

Correct answer: d

To reference a temporary SAS file in a DATA step or PROC step, you can specify the one-level name of the file (for example, Forecast) or the two-level name using the libref Work (for example, Work.Forecast).

8. What is the default length for the numeric variable `Balance`?



Name	Balance
Adams	105.73
Geller	107.89
Martinez	97.45
Noble	182.50

a. 5

b. 6

c. 7

d. 8

Correct answer: d

The numeric variable `Balance` has a default length of 8. Numeric values (no matter how many digits they contain) are stored in 8 bytes of storage unless you specify a different length.

9. How many statements does the following SAS program contain?

```
proc print data=new.prodsale
                label double;
  var state day price1 price2; where state='NC';
  label state='Name of State';
run;
```

a. three

b. four

c. five

d. six

Correct answer: c

The five statements are

- PROC PRINT statement (two lines long)

- VAR statement
- WHERE statement (on the same line as the VAR statement)
- LABEL statement
- RUN statement (on the same line as the LABEL statement).

10. What is a SAS data library?

a. a collection of SAS files, such as SAS data sets and catalogs

b. in some operating environments, a physical collection of SAS files

c. in some operating environments, a logically related collection of SAS files

d. all of the above

Correct answer: d

Every SAS file is stored in a **SAS data library**, which is a collection of SAS files, such as SAS data sets and catalogs. In some operating environments, a SAS data library is a physical collection of files. In others, the files are only logically related. In the Windows and UNIX environments, a SAS data library is typically a group of SAS files in the same folder or directory.

Chapter 2: Referencing Files and Setting Options

1. If you submit the following program, how does the output look?

```
options pagesize=55 nonumber;
proc tabulate data=clinic.admit;
   class actlevel;
   var age height weight;
   table actlevel,(age height weight)*mean;
```

```
run;
options linesize=80;
  proc means data=clinic.heart min max maxdec=1;
   var arterial heart cardiac urinary;
   class survive sex;
run;
```

a. The PROC MEANS output has a print line width of 80 characters, but the PROC TABULATE output has no print line width.

b. The PROC TABULATE output has no page numbers, but the PROC MEANS output has page numbers.

c. Each page of output from both PROC steps is 55 lines long and has no page numbers, and the PROC MEANS output has a print line width of 80 characters.

d. The date does not appear on output from either PROC step.

Correct: answer: c

When you specify a system option, it remains in effect until you change the option or end your SAS session, so both PROC steps generate output that is printed 55 lines per page with no page numbers. If you don't specify a system option, SAS uses the default value for that system option.

2. In order for the date values 05May1955 and 04Mar2046 to be read correctly, what value must the YEARCUTOFF= option have?

a. a value between 1947 and 1954, inclusive

b. 1955 or higher

c. 1946 or higher

d. any value

Correct answer: d

As long as you specify an informat with the correct field width for reading the entire date value, the YEARCUTOFF= option doesn't affect date values that have four-digit years.

3. When you specify an engine for a library, you are always specifying

a. the file format for files that are stored in the library.

b. the version of SAS that you are using.

c. access to other software vendors' files.

d. instructions for creating temporary SAS files.

Correct answer: a

A SAS engine is a set of internal instructions that SAS uses for writing to and reading from files in a SAS library. Each engine specifies the file format for files that are stored in the library, which in turn enables SAS to access files with a particular format. Some engines access SAS files, and other engines support access to other vendors' files.

4. Which statement prints a summary of all the files stored in the library named Area51?

a. `proc contents data=area51._all_ nods;`

b. `proc contents data=area51 _all_ nods;`

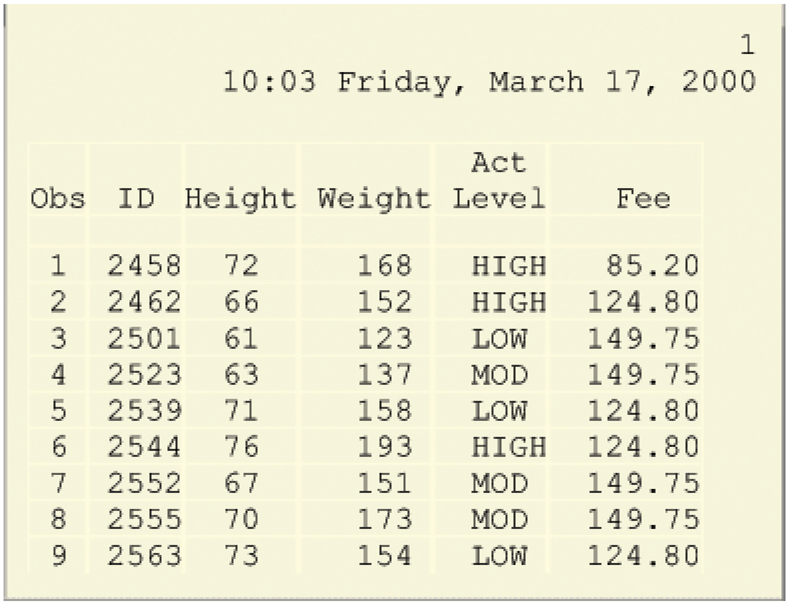
c. `proc contents data=area51 _all_ noobs;`

d. `proc contents data=area51 _all_.nods;`

Correct answer: a

To print a summary of library contents with the CONTENTS procedure, use a period to append the _ALL_ option to the libref. Adding the NODS option suppresses detailed information about the files.

5. The following PROC PRINT output was created immediately after PROC TABULATE output. Which SAS system options were specified when the report was created?



```
                                          1
         10:03 Friday, March 17, 2000

                               Act
Obs   ID  Height Weight Level     Fee

 1   2458   72     168   HIGH   85.20
 2   2462   66     152   HIGH  124.80
 3   2501   61     123   LOW   149.75
 4   2523   63     137   MOD   149.75
 5   2539   71     158   LOW   124.80
 6   2544   76     193   HIGH  124.80
 7   2552   67     151   MOD   149.75
 8   2555   70     173   MOD   149.75
 9   2563   73     154   LOW   124.80
```

a. OBS=, DATE, and NONUMBER

b. PAGENO=1, and DATE

c. NUMBER and DATE only

d. none of the above

Correct answer: b

Clearly, the DATE and PAGENO= options are specified. Because the page number on the output is 1, even though PROC TABULATE output was just produced. If you don't specify PAGENO=, all output in the Output window is numbered sequentially throughout your SAS session.

6. Which of the following programs correctly references a SAS data set named SalesAnalysis that is stored in a permanent SAS library?

```
a.  data saleslibrary.salesanalysis;
     set mydata.quarter1sales;
      if sales>100000;
   run;
b.  data mysales.totals;
      set sales_99.salesanalysis;
      if totalsales>50000;
   run;
c.  proc print data=salesanalysis.quarter1;
      var sales salesrep month;
   run;
d.  proc freq data=1999data.salesanalysis;
```

```
      tables quarter*sales;
   run;
```

Correct answer: b

Librefs must be 1 to 8 characters long, must begin with a letter or underscore, and can contain only letters, numbers, or underscores. After you assign a libref, you specify it as the first element in the two-level name for a SAS file.

7. Which time span is used to interpret two-digit year values if the YEARCUTOFF= option is set to 1950?

a. 1950-2049

b. 1950-2050

c. 1949-2050

d. 1950-2000

Correct answer: a

The YEARCUTOFF= option specifies which 100-year span is used to interpret two-digit year values. The default value of YEARCUTOFF= is 1920. However, you can override the default and change the value of YEARCUTOFF= to the first year of another 100-year span. If you specify YEARCUTOFF=1950, then the 100-year span will be from 1950 to 2049.

8. Asssuming you are using SAS code and not special SAS windows, which one of the following statements is false?

a. LIBNAME statements can be stored with a SAS program to reference the SAS library automatically when you submit the program.

b. When you delete a libref, SAS no longer has access to the files in the library. However, the contents of the library still exist on your operating system.

c. Librefs can last from one SAS session to another.

d. You can access files that were created with other vendors' software by submitting a LIBNAME statement.

Correct answer: c

The LIBNAME statement is global, which means that librefs remain in effect until you modify them, cancel them, or end your SAS session. Therefore, the LIBNAME statement assigns the libref for the current SAS session only. You must assign a libref before accessing

SAS files that are stored in a permanent SAS data library.

9. What does the following statement do?

```
libname osiris spss 'c:\myfiles\sasdata\data';
```

a. defines a library called Spss using the OSIRIS engine

b. defines a library called Osiris using the SPSS engine

c. defines two libraries called Osiris and Spss using the default engine

d. defines the default library using the OSIRIS and SPSS engines

Correct answer: b

In the LIBNAME statement, you specify the library name before the engine name. Both are followed by the path.

10. What does the following OPTIONS statement do?

```
options pagesize=15 nodate;
```

a. suppresses the date and limits the page size of the log

b. suppresses the date and limits the vertical page size for text output

c. suppresses the date and limits the vertical page size for text and HTML output

d. suppresses the date and limits the horizontal page size for text output

Correct answer: b

These options affect the format of listing output only. NODATE suppresses the date and PAGESIZE= determines the number of rows to print on the page.

Chapter 3: Editing and Debugging SAS Programs Answer Key

1. As you write and edit SAS programs it's a good idea to

a. begin DATA and PROC steps in column one.

b. indent statements within a step.

c. begin RUN statements in column one.

d. all of the above

Correct answer: d

Although you can write SAS statements in almost any format, a consistent layout enhances readability and enables you to understand the program's purpose. It's a good idea to begin DATA and PROC steps in column one, to indent statements within a step, to begin RUN statements in column one, and to include a RUN statement after every DATA step or PROC step.

2. What usually happens when an error is detected?

a. SAS continues processing the step.

b. SAS continues to process the step, and the log displays messages about the error.

c. SAS stops processing the step in which the error occurred, and the log displays messages about the error.

d. SAS stops processing the step in which the error occurred, and the program output displays messages about the error.

Correct answer: c

Syntax errors generally cause SAS to stop processing the step in which the error occurred. When a program that contains an error is submitted, messages regarding the problem also appear in the SAS log. When a syntax error is detected, the SAS log displays the word ERROR, identifies the possible location of the error, and gives an explanation of the error.

3. A syntax error occurs when

a. some data values are not appropriate for the SAS statements that are specified in a program.

b. the form of the elements in a SAS statement is correct, but the elements are not valid for that usage.

c. program statements do not conform to the rules of the SAS language.

d. none of the above

Correct c

answer:

Syntax errors are common types of errors. Some SAS system options, features of the Editor window, and the DATA step debugger can help you identify syntax errors. Other types of errors include data errors, semantic errors, and execution-time errors.

4. How can you tell whether you have specified an invalid option in a SAS program?

a. A log message indicates an error in a statement that seems to be valid.

b. A log message indicates that an option is not valid or not recognized.

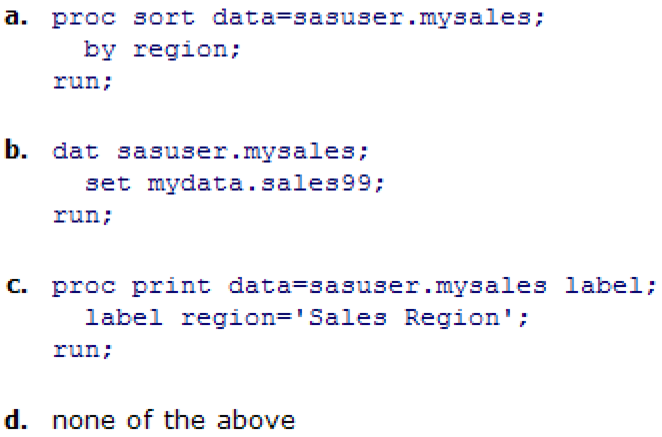
c. The message "PROC running" or "DATA step running" appears at the top of the active window.

d. You can't tell until you view the output from the program.

Correct answer: b

When you submit a SAS statement that contains an invalid option, a log message notifies you that the option is not valid or not recognized. You should recall the program, remove or replace the invalid option, check your statement syntax as needed, and resubmit the corrected program.

5. Which of the following programs contains a syntax error?



a.
```
proc sort data=sasuser.mysales;
  by region;
run;
```

b.
```
dat sasuser.mysales;
  set mydata.sales99;
run;
```

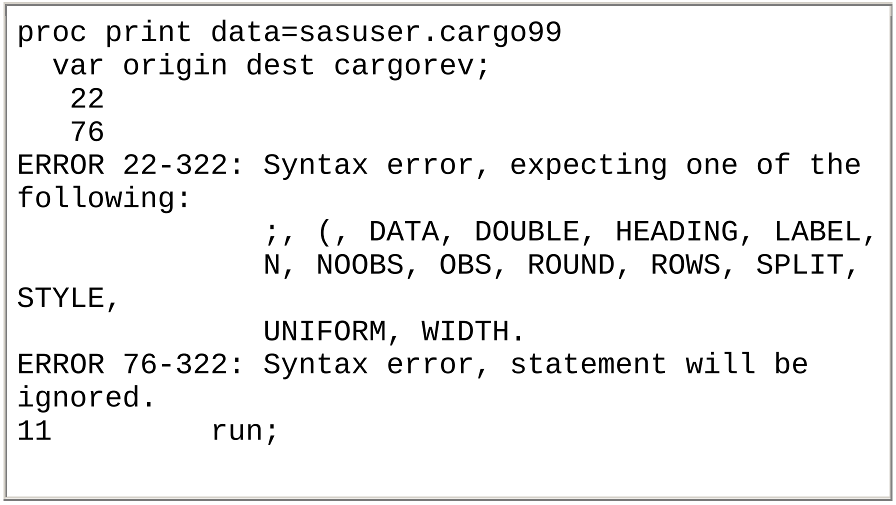
c.
```
proc print data=sasuser.mysales label;
  label region='Sales Region';
run;
```

d. none of the above

Correct answer: b

The DATA step contains a misspelled keyword (dat instead of data). However, this is such a common (and easily interpretable) error that SAS produces only a warning message, not an error.

6. What does the following log indicate about your program?



```
proc print data=sasuser.cargo99
  var origin dest cargorev;
   22
   76
ERROR 22-322: Syntax error, expecting one of the
following:
              ;, (, DATA, DOUBLE, HEADING, LABEL,
              N, NOOBS, OBS, ROUND, ROWS, SPLIT,
STYLE,
              UNIFORM, WIDTH.
ERROR 76-322: Syntax error, statement will be
ignored.
11         run;
```

a. SAS identifies a syntax error at the position of the VAR statement.

b. SAS is reading VAR as an option in the PROC PRINT statement.

c. SAS has stopped processing the program because of errors.

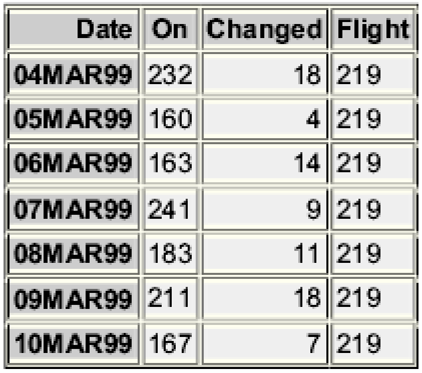
d. all of the above

Correct answer: d

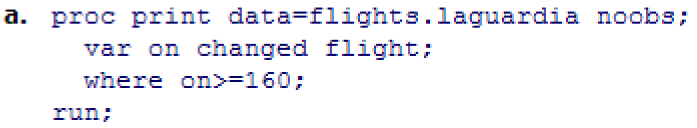
Because there is a missing semicolon at the end of the PROC PRINT statement, SAS interprets VAR as an option in PROC PRINT and finds a syntax error at that location. SAS stops processing programs when it encounters a syntax error.

Chapter 4: Creating List Reports Answer Key

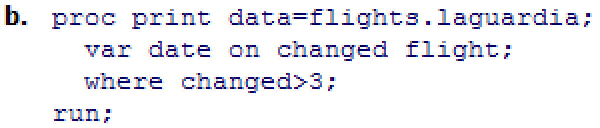
1. Which PROC PRINT step below creates the following output?



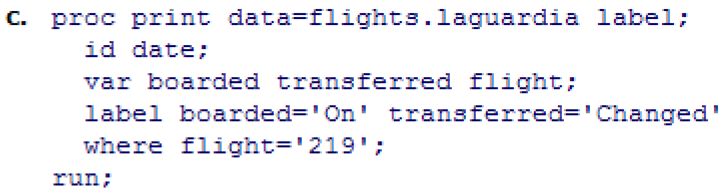
Date	On	Changed	Flight
04MAR99	232	18	219
05MAR99	160	4	219
06MAR99	163	14	219
07MAR99	241	9	219
08MAR99	183	11	219
09MAR99	211	18	219
10MAR99	167	7	219



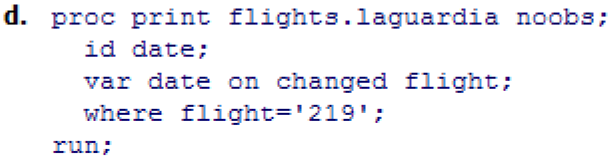
```
a. proc print data=flights.laguardia noobs;
     var on changed flight;
     where on>=160;
   run;
```



```
b. proc print data=flights.laguardia;
     var date on changed flight;
     where changed>3;
   run;
```



```
c. proc print data=flights.laguardia label;
     id date;
     var boarded transferred flight;
     label boarded='On' transferred='Changed'
     where flight='219';
   run;
```

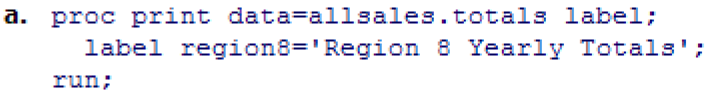


```
d. proc print flights.laguardia noobs;
     id date;
     var date on changed flight;
     where flight='219';
   run;
```

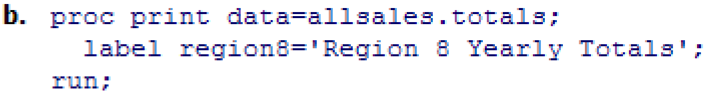
Correct answer: c

The DATA= option specifies the data set that you are listing, and the ID statement replaces the `Obs` column with the specified variable. The VAR statement specifies variables and controls the order in which they appear, and the WHERE statement selects rows based on a condition. The LABEL option in the PROC PRINT statement causes the labels that are specified in the LABEL statement to be displayed.

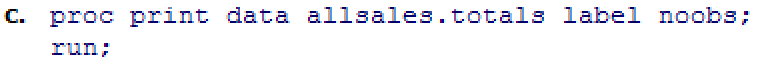
2. Which of the following PROC PRINT steps is correct if labels are not stored with the data set?



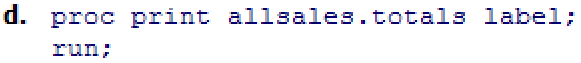
```
a. proc print data=allsales.totals label;
     label region8='Region 8 Yearly Totals';
   run;
```



```
b. proc print data=allsales.totals;
     label region8='Region 8 Yearly Totals';
   run;
```



```
c. proc print data allsales.totals label noobs;
   run;
```



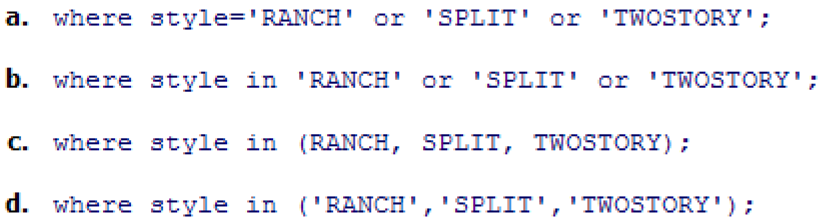
```
d. proc print allsales.totals label;
   run;
```

Correct a

answer:

You use the DATA= option to specify the data set to be printed. The LABEL option specifies that variable labels appear in output instead of variable names.

3. Which of the following statements selects from a data set only those observations for which the value of the variable `Style` is RANCH, SPLIT, or TWOSTORY?



a. `where style='RANCH' or 'SPLIT' or 'TWOSTORY';`

b. `where style in 'RANCH' or 'SPLIT' or 'TWOSTORY';`

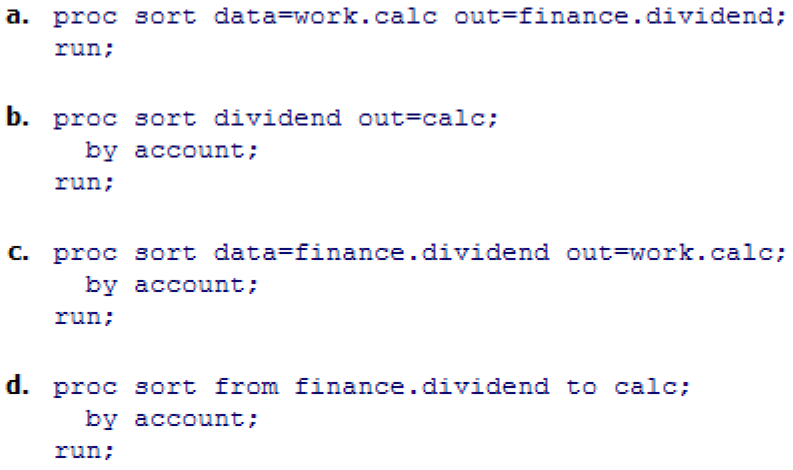
c. `where style in (RANCH, SPLIT, TWOSTORY);`

d. `where style in ('RANCH','SPLIT','TWOSTORY');`

Correct answer: d

In the WHERE statement, the IN operator enables you to select observations based on several values. You specify values in parentheses and separate them by spaces or commas. Character values must be enclosed in quotation marks and must be in the same case as in the data set.

4. If you want to sort your data and create a temporary data set named Calc to store the sorted data, which of the following steps should you submit?



a.
```
proc sort data=work.calc out=finance.dividend;
run;
```

b.
```
proc sort dividend out=calc;
  by account;
run;
```

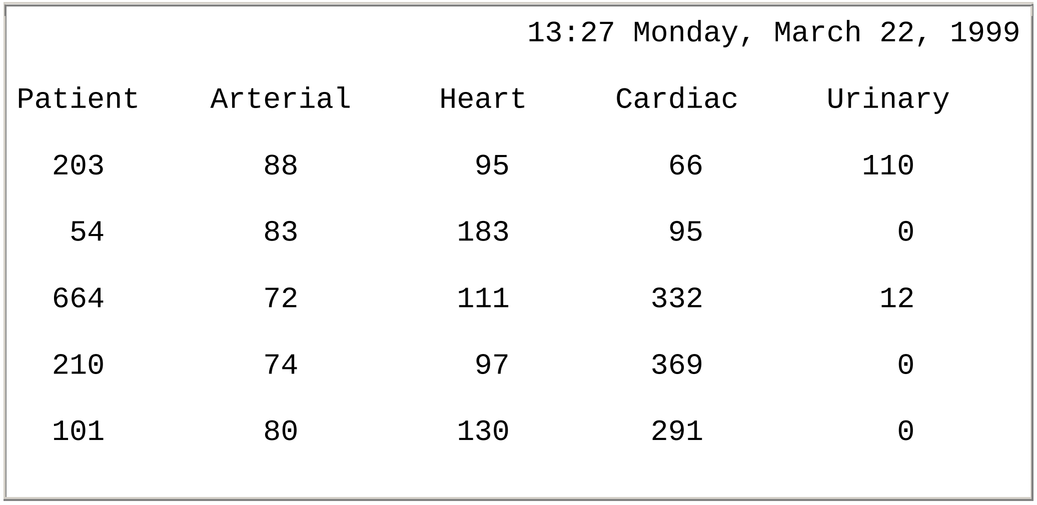
c.
```
proc sort data=finance.dividend out=work.calc;
  by account;
run;
```

d.
```
proc sort from finance.dividend to calc;
  by account;
run;
```

Correct answer: c

In a PROC SORT step, you specify the DATA= option to specify the data set to sort. The OUT= option specifies an output data set. The required BY statement specifies the variable(s) to use in sorting the data.

5. Which options are used to create the following PROC PRINT output?



```
                                  13:27 Monday, March 22, 1999

Patient     Arterial     Heart     Cardiac     Urinary

  203          88          95         66          110

   54          83         183         95            0

  664          72         111        332           12

  210          74          97        369            0

  101          80         130        291            0
```

a. the DATE system option and the LABEL option in PROC PRINT

b. the DATE and NONUMBER system options and the DOUBLE and NOOBS options in PROC PRINT

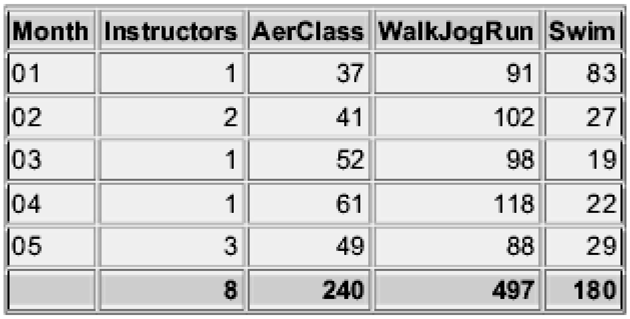
c. the DATE and NONUMBER system options and the DOUBLE option in PROC PRINT

d. the DATE and NONUMBER system options and the NOOBS option in PROC PRINT

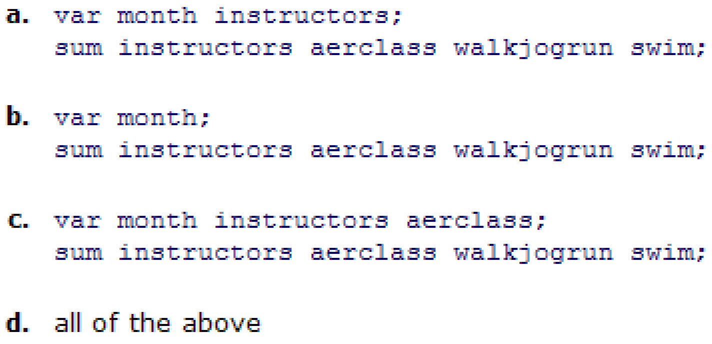
Correct answer: b

The DATE and NONUMBER system options cause the output to appear with the date but without page numbers. In the PROC PRINT step, the DOUBLE option specifies double spacing, and the NOOBS option removes the default `Obs` column.

6. Which of the following statements can you use in a PROC PRINT step to create this output?



Month	Instructors	AerClass	WalkJogRun	Swim
01	1	37	91	83
02	2	41	102	27
03	1	52	98	19
04	1	61	118	22
05	3	49	88	29
	8	240	497	180



a.
```
var month instructors;
sum instructors aerclass walkjogrun swim;
```

b.
```
var month;
sum instructors aerclass walkjogrun swim;
```

c.
```
var month instructors aerclass;
sum instructors aerclass walkjogrun swim;
```

d. all of the above

Correct answer: d

You do not need to name the variables in a VAR statement if you specify them in the SUM statement, but you can. If you choose not to name the variables in the VAR statement as well, then the SUM statement determines the order of the variables in the output.

7. What happens if you submit the following program?

```
proc sort data=clinic.diabetes;
run;
proc print data=clinic.diabetes;
  var age height weight pulse;
  where sex='F';
run;
```

a. The PROC PRINT step runs successfully, printing observations in their sorted order.

b. The PROC SORT step permanently sorts the input data set.

c. The PROC SORT step generates errors and stops processing, but the PROC PRINT step runs successfully, printing observations in their original (unsorted) order.

d. The PROC SORT step runs successfully, but the PROC PRINT step generates errors and stops processing.

Correct answer: c

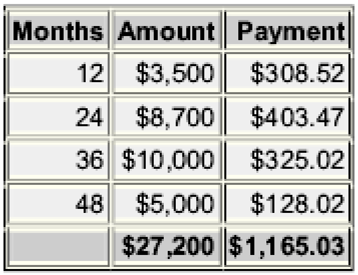
The BY statement is required in PROC SORT. Without it, the PROC SORT step fails. However, the PROC PRINT step prints the original data set as requested.

8. If you submit the following program, which output does it create?

```
proc sort data=finance.loans out=work.loans;
```

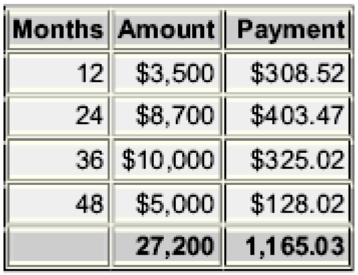
```
  by months amount;
run;
proc print data=work.loans noobs;
  var months;
  sum amount payment;
  where months<360;
run;
```

a.



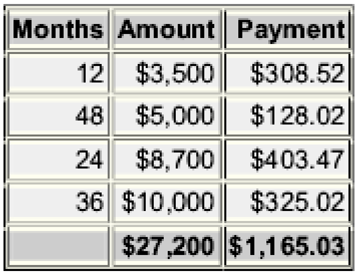
Months	Amount	Payment
12	$3,500	$308.52
24	$8,700	$403.47
36	$10,000	$325.02
48	$5,000	$128.02
	$27,200	**$1,165.03**

b.



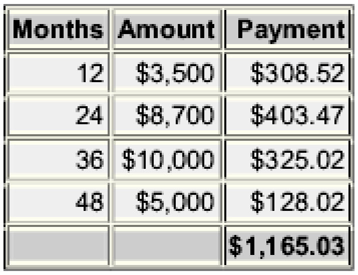
Months	Amount	Payment
12	$3,500	$308.52
24	$8,700	$403.47
36	$10,000	$325.02
48	$5,000	$128.02
	27,200	**1,165.03**

c.



Months	Amount	Payment
12	$3,500	$308.52
48	$5,000	$128.02
24	$8,700	$403.47
36	$10,000	$325.02
	$27,200	**$1,165.03**

d.



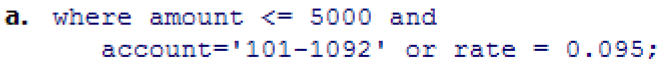
Months	Amount	Payment
12	$3,500	$308.52
24	$8,700	$403.47
36	$10,000	$325.02
48	$5,000	$128.02
		$1,165.03

Correct answer: a

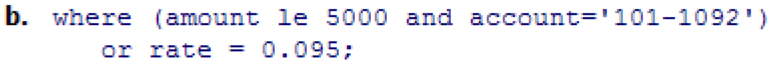
Column totals appear at the end of the report in the same format as the values of the variables, so b is incorrect. Work.Loans is sorted by Month and Amount, so c is incorrect. The program sums both Amount and Payment, so d is incorrect.

9. Choose the statement below that selects rows which

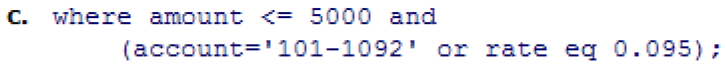
- the amount is less than or equal to $5000
- the account is 101-1092 or the rate equals 0.095.



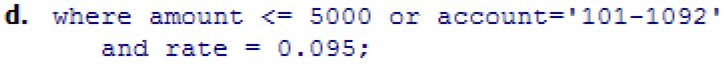
```
a. where amount <= 5000 and
      account='101-1092' or rate = 0.095;
```



```
b. where (amount le 5000 and account='101-1092')
      or rate = 0.095;
```



```
c. where amount <= 5000 and
       (account='101-1092' or rate eq 0.095);
```



```
d. where amount <= 5000 or account='101-1092'
      and rate = 0.095;
```

Correct answer: c

To ensure that the compound expression is evaluated correctly, you can use parentheses to group

```
account='101-1092' or rate eq 0.095
```

```
OBS Account   Amount  Rate  MonthsPayment

1   101-1092  $22,000 10.00%60    $467.43
2   101-1731  $114,0009.50% 360   $958.57
3   101-1289  $10,000 10.50%36    $325.02
4   101-3144  $3,500  10.50%12    $308.52
5   103-1135  $8,700  10.50%24    $403.47
6   103-1994  $18,500 10.00%60    $393.07
7   103-2335  $5,000  10.50%48    $128.02
8   103-3864  $87,500 9.50% 360   $735.75
9   103-3891  $30,000 9.75% 360   $257.75
```

For example, from the data set above, a and b above select observations 2 and 8 (those that have a rate of 0.095); c selects no observations; and d selects observations 4 and 7 (those that have an amount less than or equal to 5000).

10. What does PROC PRINT display by default?

a. PROC PRINT does not create a default report; you must specify the rows and columns to be displayed.

b. PROC PRINT displays all observations and variables in the data set. If you want an additional column for observation numbers, you can request it.

c. PROC PRINT displays columns in the following order: a column for observation numbers, all character variables, and all numeric variables.

d. PROC PRINT displays all observations and variables in the data set, a column for observation numbers on the far left, and variables in the order in which they occur in the data set.

Correct answer: d

You can remove the column for observation numbers. You can also specify the variables you want, and you can select observations according to conditions.

Chapter 5: Creating SAS Data Sets from Raw Data Answer Key

1. Which SAS statement associates the fileref Crime with the raw data file C:\States\Data\Crime?

a. `filename crime 'c:\states\data\crime';`

b. `filename crime c:\states\data\crime;`

c. `fileref crime 'c:\states\data\crime';`

d. `filename 'c:\states\data\crime' crime;`

Correct a

answer:

Before you can read your raw data, you must reference the raw data file by creating a fileref. You assign a fileref by using a FILENAME statement in the same way that you assign a libref by using a LIBNAME statement.

2. Filerefs remain in effect until

a. you change them.

b. you cancel them.

c. you end your SAS session.

d. all of the above

Correct answer: d

Like LIBNAME statements, FILENAME statements are global; they remain in effect until you change them, cancel them, or end your SAS session.

3. Which statement identifies the name of a raw data file to be read with the fileref Products and specifies that the DATA step read only records 1-15?

a. `infile products obs 15;`

b. `infile products obs=15;`

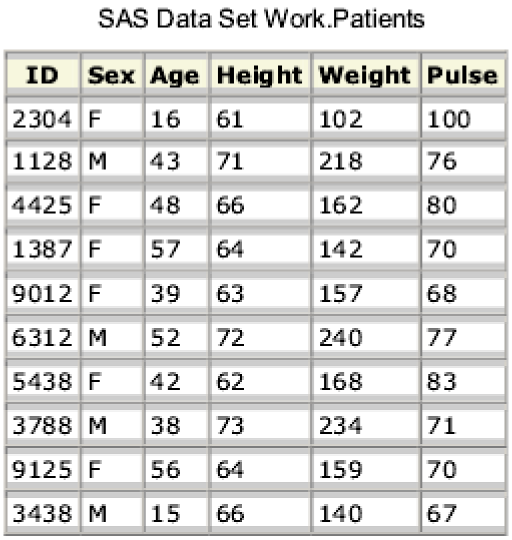
c. `input products obs=15;`

d. `input products 1-15;`

Correct answer: b

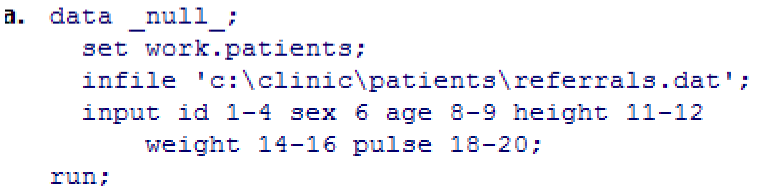
You use an INFILE statement to specify the raw data file to be read. You can specify a fileref or an actual filename (in quotation marks). The OBS= option in the INFILE statement enables you to process only records 1 through n.

4. Which of the following programs correctly writes the observations from the data set below to a raw data file?

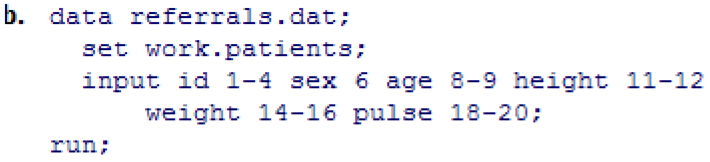


SAS Data Set Work.Patients

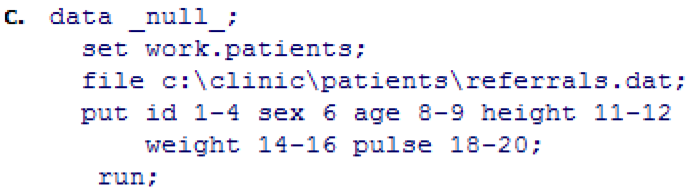
ID	Sex	Age	Height	Weight	Pulse
2304	F	16	61	102	100
1128	M	43	71	218	76
4425	F	48	66	162	80
1387	F	57	64	142	70
9012	F	39	63	157	68
6312	M	52	72	240	77
5438	F	42	62	168	83
3788	M	38	73	234	71
9125	F	56	64	159	70
3438	M	15	66	140	67



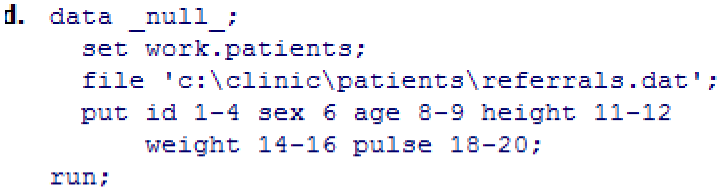
a.
```
data _null_;
  set work.patients;
  infile 'c:\clinic\patients\referrals.dat';
  input id 1-4 sex 6 age 8-9 height 11-12
      weight 14-16 pulse 18-20;
run;
```



b.
```
data referrals.dat;
  set work.patients;
  input id 1-4 sex 6 age 8-9 height 11-12
      weight 14-16 pulse 18-20;
run;
```



c.
```
data _null_;
  set work.patients;
  file c:\clinic\patients\referrals.dat;
  put id 1-4 sex 6 age 8-9 height 11-12
      weight 14-16 pulse 18-20;
   run;
```



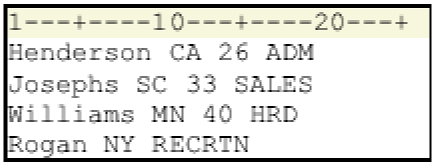
d.
```
data _null_;
  set work.patients;
  file 'c:\clinic\patients\referrals.dat';
  put id 1-4 sex 6 age 8-9 height 11-12
      weight 14-16 pulse 18-20;
run;
```

Correct answer: d

The keyword _NULL_ in the DATA statement enables you to use the power of the DATA step without actually creating a SAS data set. You use the FILE and PUT statements to write out the observations from a SAS data set to a raw data file. The FILE statement specifies the raw data file and the PUT statement describes the lines to write to the raw data file. The filename and location that are specified in the FILE statement must be enclosed in quotation marks.

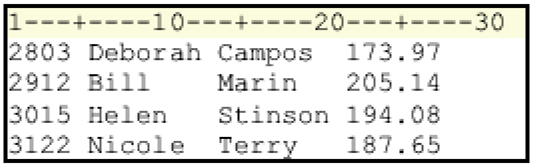
5. Which raw data file can be read using column input?

a.



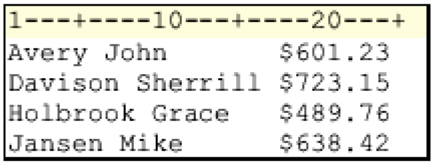
```
1---+----10---+----20---+
Henderson CA 26 ADM
Josephs SC 33 SALES
Williams MN 40 HRD
Rogan NY RECRTN
```

b.



```
1---+----10---+----20---+----30
2803 Deborah Campos  173.97
2912 Bill    Marin   205.14
3015 Helen   Stinson 194.08
3122 Nicole  Terry   187.65
```

c.



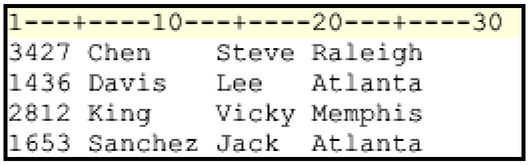
```
1---+----10---+----20---+
Avery John       $601.23
Davison Sherrill $723.15
Holbrook Grace   $489.76
Jansen Mike      $638.42
```

d. all of the above

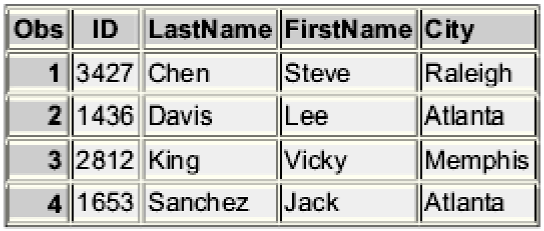
Correct answer: b

Column input is appropriate only in some situations. When you use column input, your data must be standard character or numeric values, and they must be in fixed fields. That is, values for a particular variable must be in the same location in all records.

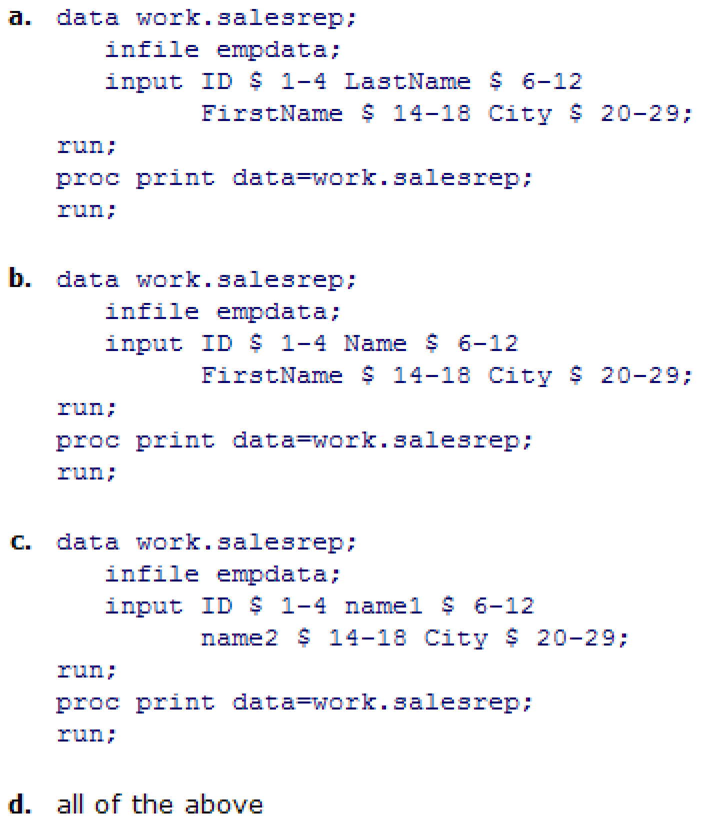
6. Which program creates the output shown below?



```
1---+----10---+----20---+----30
3427 Chen    Steve Raleigh
1436 Davis   Lee   Atlanta
2812 King    Vicky Memphis
1653 Sanchez Jack  Atlanta
```



Obs	ID	LastName	FirstName	City
1	3427	Chen	Steve	Raleigh
2	1436	Davis	Lee	Atlanta
3	2812	King	Vicky	Memphis
4	1653	Sanchez	Jack	Atlanta



a.
```
data work.salesrep;
   infile empdata;
   input ID $ 1-4 LastName $ 6-12
         FirstName $ 14-18 City $ 20-29;
run;
proc print data=work.salesrep;
run;
```

b.
```
data work.salesrep;
   infile empdata;
   input ID $ 1-4 Name $ 6-12
         FirstName $ 14-18 City $ 20-29;
run;
proc print data=work.salesrep;
run;
```

c.
```
data work.salesrep;
   infile empdata;
   input ID $ 1-4 name1 $ 6-12
         name2 $ 14-18 City $ 20-29;
run;
proc print data=work.salesrep;
run;
```

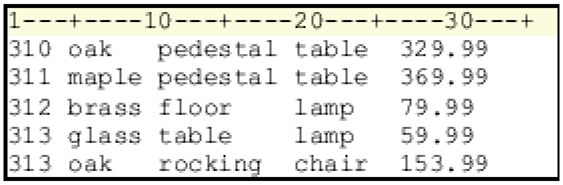
d. all of the above

Correct answer: a

The INPUT statement creates a variable using the name that you assign to each field. Therefore, when you write an INPUT statement, you need to specify the variable names exactly as you want them to appear in the SAS data set.

7. Which statement correctly reads the fields in the following order: `StockNumber`, `Price`, `Item`, `Finish`, `Style`?

Field Name	Start Column	End Column	Data Type
StockNumber	1	3	character
Finish	5	9	character
Style	11	18	character
Item	20	24	character
Price	27	32	numeric



```
1---+----10---+----20---+----30---+
310 oak   pedestal table  329.99
311 maple pedestal table  369.99
312 brass floor    lamp   79.99
313 glass table    lamp   59.99
313 oak   rocking  chair  153.99
```

a.
```
input StockNumber $ 1-3 Finish $ 5-9 Style $ 11-18
      Item $ 20-24 Price 27-32;
```

b.
```
input StockNumber $ 1-3 Price 27-32
      Item $ 20-24 Finish $ 5-9 Style $ 11-18;
```

c.
```
input $ StockNumber 1-3 Price 27-32
      $ Item  20-24 $ Finish 5-9 $ Style 11-18;
```

d.
```
input StockNumber $ 1-3 Price $ 27-32
      Item $ 20-24 Finish $ 5-9 Style $ 11-18;
```

Correc t answer: b

You can use column input to read fields in any order. You must specify the variable name to be created, identify character values with a $, and name the correct starting column and ending column for each field.

8. Which statement correctly re-defines the values of the variable `Income` as 100 percent higher?

a. `income=income*1.00;`

b. `income=income+(income*2.00);`

c. `income=income*2;`

d. `income=*2;`

Correct answer: c

To re-define the values of the variable `Income` in an Assignment statement, you specify the variable name on the left side of the equal sign and an appropriate expression including the variable name on the right side of the equal sign.

9. Which program correctly reads instream data?

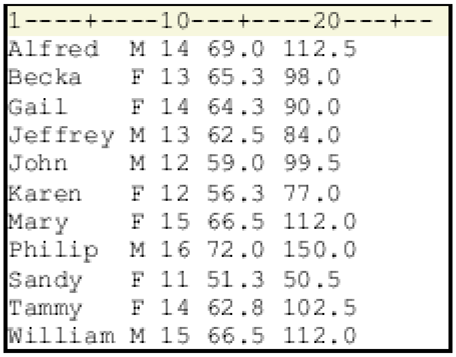
```
a.  data finance.newloan;
  input datalines;
     if country='JAPAN';
```

```
        MonthAvg=amount/12;
     1998 US     CARS   194324.12
     1998 US     TRUCKS 142290.30
     1998 CANADA CARS    10483.44
     1998 CANADA TRUCKS  93543.64
     1998 MEXICO CARS    22500.57
     1998 MEXICO TRUCKS  10098.88
     1998 JAPAN  CARS    15066.43
     1998 JAPAN  TRUCKS  40700.34
;

  b.   data finance.newloan;
        input Year 1-4 Country $ 6-11
              Vehicle $ 13-18 Amount 20-28;
        if country='JAPAN';
        MonthAvg=amount/12;
        datalines;
     run;

  c.   data finance.newloan;
        input Year 1-4 Country 6-11
              Vehicle 13-18 Amount 20-28;
        if country='JAPAN';
        MonthAvg=amount/12;
        datalines;
     1998 US     CARS   194324.12
     1998 US     TRUCKS 142290.30
     1998 CANADA CARS    10483.44
     1998 CANADA TRUCKS  93543.64
     1998 MEXICO CARS    22500.57
     1998 MEXICO TRUCKS  10098.88
     1998 JAPAN  CARS    15066.43
     1998 JAPAN  TRUCKS  40700.34
     ;
  d.   data finance.newloan;
        input Year 1-4 Country $ 6-11
              Vehicle $ 13-18 Amount 20-28;
        if country='JAPAN';
        MonthAvg=amount/12;
        datalines;
     1998 US     CARS   194324.12
     1998 US     TRUCKS 142290.30
     1998 CANADA CARS    10483.44
     1998 CANADA TRUCKS  93543.64
     1998 MEXICO CARS    22500.57
     1998 MEXICO TRUCKS  10098.88
     1998 JAPAN  CARS    15066.43
     1998 JAPAN  TRUCKS  40700.34
     ;
```

Correct answer: d

To read instream data, you specify a DATALINES statement and data lines, followed by a null statement (single semicolon) to indicate the end of the input data. Program a contains no DATALINES statement, and the INPUT statement doesn't specify the fields to read. Program b contains no data lines, and the INPUT statement in program c doesn't specify the necessary dollar signs for the character variables `Country` and `Vehicle`.

10. Which SAS statement subsets the raw data shown below so that only the observations in which Sex (in the second field) has a value of F are processed?



```
1----+----10---+----20---+--
Alfred  M 14 69.0 112.5
Becka   F 13 65.3 98.0
Gail    F 14 64.3 90.0
Jeffrey M 13 62.5 84.0
John    M 12 59.0 99.5
Karen   F 12 56.3 77.0
Mary    F 15 66.5 112.0
Philip  M 16 72.0 150.0
Sandy   F 11 51.3 50.5
Tammy   F 14 62.8 102.5
William M 15 66.5 112.0
```

a. if sex=f;

b. if sex=F;

c. if sex='F';

d. a or b

Correct answer: c

To subset data, you can use a subsetting IF statement in any DATA step to process only those observations that meet a specified condition. Because `Sex` is a character variable, the value F must be enclosed in quotation marks and must be in the same case as in the data set.

Chapter 6: Understanding DATA Step Processing Answer Key

1. Which of the following is not created during the compilation phase?

a. the data set descriptor

b. the first observation

c. the program data vector

d. the `_N_` and `_ERROR_` automatic variables

Correct answer: b

At the beginning of the compilation phase, the program data vector is created. The program data vector includes the two automatic variables `_N_` and `_ERROR_.` The descriptor portion of the new SAS data set is created at the end of the compilation phase. The descriptor portion includes the name of the data set, the number of observations and variables, and the names and attributes of the variables. Observations are not written until the execution phase.

2. During the compilation phase, SAS scans each statement in the DATA step, looking for syntax errors. Which of the following is not considered a syntax error?

a. incorrect values and formats

b. invalid options or variable names

c. missing or invalid punctuation

d. missing or misspelled keywords

Correct answer: a

Syntax checking can detect many common errors, but it cannot verify the values of variables or the correctness of formats.

3. Unless otherwise directed, the DATA step executes

a. once for each compilation phase.

b. once for each DATA step statement.

c. once for each record in the input file.

d. once for each variable in the input file.

Correct answer: c

The DATA step executes once for each record in the input file, unless otherwise directed.

4. At the beginning of the execution phase, the value of _N_ is 1, the value of _ERROR_ is 0, and the values of the remaining variables are set to

a. 0

b. 1

c. undefined

d. missing

Correct answer: d

The remaining variables are initialized to missing. Missing numeric values are represented by periods, and missing character values are represented by blanks.

5. Suppose you run a program that causes three DATA step errors. What is the value of the automatic variable `_ERROR_` when the observation that contains the third error is processed?

a. 0

b. 1

c. 2

d. 3

Correct answer: b

The default value of _ERROR_ is 0, which means there is no error. When an error occurs, whether it is one error or multiple errors, the value is set to 1.

6. Which of the following actions occurs at the end of the DATA step?

a. The automatic variables _N_ and _ERROR_ are incremented by one.

b. The DATA step stops execution.

c. The descriptor portion of the data set is written.

d. The values of variables created in programming statements are re-set to missing in the program data vector.

Correct answer: d

By default, at the end of the DATA step, the values in the program data vector are written to the data set as an observation, the value of the automatic variable _N_ is incremented by one, control returns to the top of the DATA step, and the values of variables created in programming statements are set to missing. The automatic variable _ERROR_ retains its value.

7. Look carefully at the DATA step shown below. Based on the INPUT statement, in what order will the variables be stored in the new data set?

```
data perm.update;
   infile invent;
   input IDnum $ 15-19 Item $ 1-13 Instock 21-22
         BackOrd 24-25;
   Total=instock+backord;
run;
```

a. `IDnum Item InStock BackOrd Total`

b. `Item IDnum InStock BackOrd Total`

c. `Total IDnum Item InStock BackOrd`

d. `Total Item IDnum InStock BackOrd`

Correct answer: a

The order in which variables are defined in the DATA step determines the order in which the variables are stored in the data set.

8. If SAS cannot interpret syntax errors, then

a. data set variables will contain missing values.

b. the DATA step does not compile.

c. the DATA step still compiles, but it does not execute.

d. the DATA step still compiles and executes.

Correct answer: c

When SAS can't interpret syntax errors, the DATA step compiles, but it does not execute.

9. What is wrong with this program?

```
data perm.update;
   infile invent
   input Item $ 1-13 IDnum $ 15-19 Instock 21-22
         BackOrd 24-25;
   total=instock+backord;
run;
```

a. missing semicolon on second line

b. missing semicolon on third line

c. incorrect order of variables

d. incorrect variable type

Correct answer: a

A semicolon is missing from the second line. It will cause an error because the INPUT statement will be interpreted as invalid INFILE statement options.

10. Look carefully at this section of a SAS session log. Based on the note, what was the most likely problem with the DATA step?

```
NOTE: Invalid data for IDnum in line 7 15-19.
RULE: ----+----1----+----2----+----3----+----4
7       Bird Feeder LG088 3 20
Item=Bird Feeder IDnum=. InStock=3 BackOrd=20
Total=23 _ERROR_=1 _N_=1
```

a. A keyword was misspelled in the DATA step.

b. A semicolon was missing from the INFILE statement.

c. A variable was misspelled in the INPUT statement.

d. A dollar sign was missing in the INPUT statement.

Correct answer: d

The third line of the log displays the values for `IDnum`, which are clearly character values. The fourth line displays the values in the program data vector and shows that the values for `IDnum` are missing, even though the other values are correctly assigned. Thus, it appears that numeric values were expected for `IDnum`. A dollar sign, to indicate character values, must be missing from the INPUT statement.

Chapter 7: Creating and Applying User-Defined Formats Answer Key

1. If you don't specify the LIBRARY= option, your formats are stored in Work.Formats, and they exist

a. only for the current procedure.

b. only for the current DATA step.

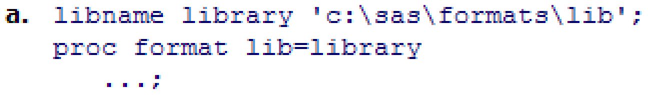
c. only for the current SAS session.

d. permanently.

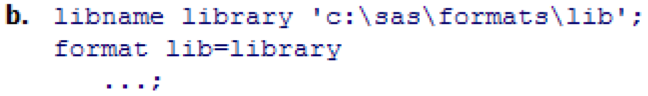
Correct answer: c

If you do not specify the LIBRARY= option, formats are stored in a default format catalog named Work.Formats. As the libref Work implies, any format that is stored in Work.Formats is a temporary format that exists only for the current SAS session.

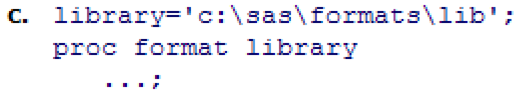
2. Which of the following statements will store your formats in a permanent catalog?



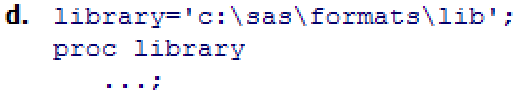
```
a. libname library 'c:\sas\formats\lib';
   proc format lib=library
      ...;
```



```
b. libname library 'c:\sas\formats\lib';
   format lib=library
      ...;
```



```
c. library='c:\sas\formats\lib';
   proc format library
      ...;
```



```
d. library='c:\sas\formats\lib';
   proc library
      ...;
```

Correct answer: a

To store formats in a permanent catalog, you first write a LIBNAME statement to associate the libref with the SAS data library in which the catalog will be stored. Then add the LIB= (or LIBRARY=) option to the PROC FORMAT statement, specifying the name of the catalog.

3. When creating a format with the VALUE statement, the new format's name

- cannot end with a number
- cannot end with a period
- cannot be the name of a SAS format, and

a. cannot be the name of a data set variable.

b. must be at least two characters long.

c. must be at least eight characters long.

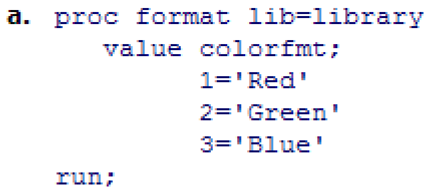
d. must begin with a dollar sign ($) if used with a character variable.

Correct answer: d

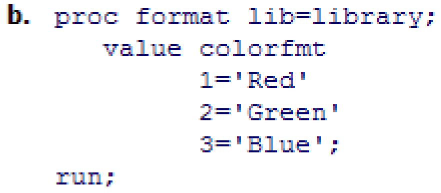
The name of a format that is created with a VALUE statement must begin with a dollar sign

($) if it applies to a character variable.

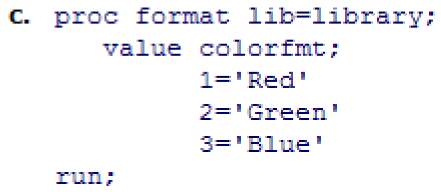
4. Which of the following FORMAT procedures is written correctly?



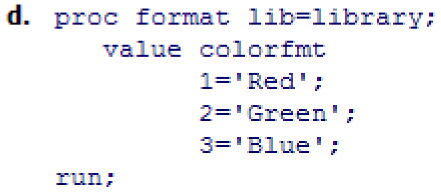
```
a. proc format lib=library
      value colorfmt;
            1='Red'
            2='Green'
            3='Blue'
   run;
```



```
b. proc format lib=library;
      value colorfmt
            1='Red'
            2='Green'
            3='Blue';
   run;
```



```
c. proc format lib=library;
      value colorfmt;
            1='Red'
            2='Green'
            3='Blue'
   run;
```



```
d. proc format lib=library;
      value colorfmt
            1='Red';
            2='Green';
            3='Blue';
   run;
```

Correct answer: b

A semicolon is needed after the PROC FORMAT statement. The VALUE statement begins with the keyword VALUE and ends with a semicolon after all the labels have been defined.

5. Which of these is false? Ranges in the VALUE statement can specify

a. a single value, such as 24 or 'S'.

b. a range of numeric values, such as 0–1500.

c. a range of character values, such as 'A–'M'.

d. a list of numeric and character values separated by commas, such as 90,'B',180,'D',270.

Correct answer: d

You can list values separated by commas, but the list must contain either all numeric values or all character values. Data set variables are either numeric or character.

6. How many characters can be used in a label?

a. 40

b. 96

c. 200

d. 256

Correct answer: d

When specifying a label, enclose it in quotation marks and limit the label to 256 characters.

7. Which keyword can be used to label missing values as well as any values that are not specified in a range?

a. LOW

b. MISS

c. MISSING

d. OTHER

Correct answer: d

MISS and MISSING are invalid keywords, and LOW does not include missing values. The keyword OTHER can be used in the VALUE statement to label missing values as well as any values that are not specifically included in a range.

8. You can place the FORMAT statement in either a DATA step or a PROC step. What happens when you place the FORMAT statement in a DATA step?

a. You temporarily associate the formats with variables.

b. You permanently associate the formats with variables.

c. You replace the original data with the format labels.

d. You make the formats available to other data sets

Correct answer: b

By placing the FORMAT statement in a DATA step, you permanently associate the defined formats with variables.

9. The format JOBFMT was created in a FORMAT procedure. Which FORMAT statement will apply it to the variable `JobTitle` in the program output?

a. `format jobtitle jobfmt;`

b. `format jobtitle jobfmt.;`

c. `format jobtitle=jobfmt;`

d. `format jobtitle='jobfmt';`

Correct answer: b

To associate a user-defined format with a variable, place a period at the end of the format name when it is used in the FORMAT statement.

10. Which keyword, when added to the PROC FORMAT statement, will display all the formats in your catalog?

a. CATALOG

b. LISTFMT

c. FMTCAT

d. FMTLIB

Correct answer: d

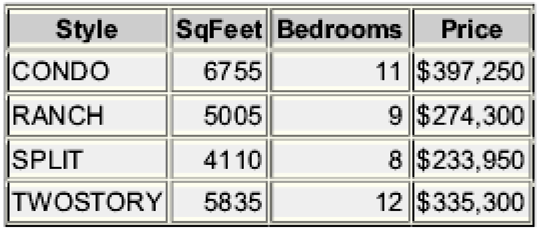
Adding the keyword FMTLIB to the PROC FORMAT statement displays a list of all the formats in your catalog, along with descriptions of their values.

Chapter 8: Creating Enhanced List and Summary Reports Answer Key

1. If `Style` has four unique values and you submit the following program, which output do you get? (Assume that all the other variables are numeric.)

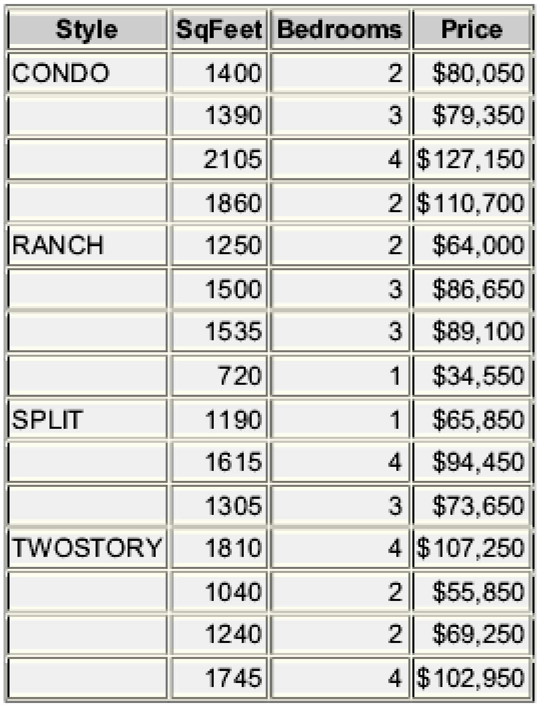
```
proc report data=sasuser.houses nowd;
   column style sqfeet bedrooms price;
   define style / group;
run;
```

a.



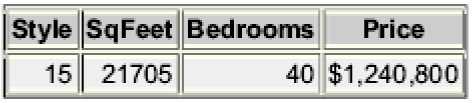
Style	SqFeet	Bedrooms	Price
CONDO	6755	11	$397,250
RANCH	5005	9	$274,300
SPLIT	4110	8	$233,950
TWOSTORY	5835	12	$335,300

b.



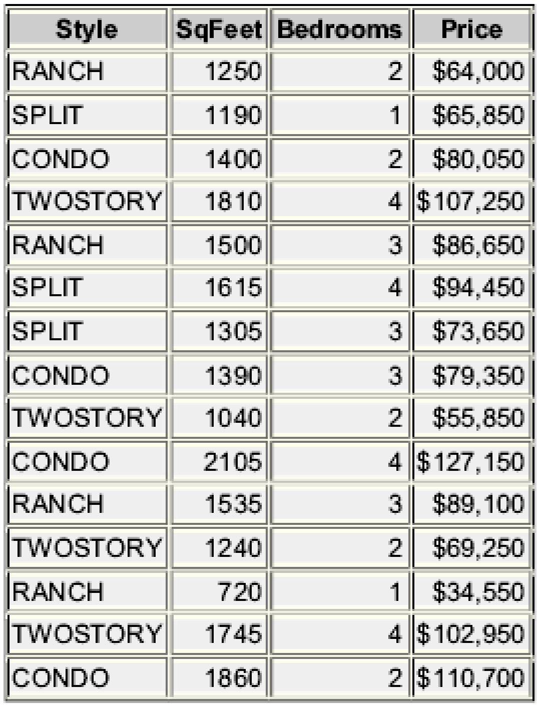
Style	SqFeet	Bedrooms	Price
CONDO	1400	2	$80,050
	1390	3	$79,350
	2105	4	$127,150
	1860	2	$110,700
RANCH	1250	2	$64,000
	1500	3	$86,650
	1535	3	$89,100
	720	1	$34,550
SPLIT	1190	1	$65,850
	1615	4	$94,450
	1305	3	$73,650
TWOSTORY	1810	4	$107,250
	1040	2	$55,850
	1240	2	$69,250
	1745	4	$102,950

c.



Style	SqFeet	Bedrooms	Price
15	21705	40	$1,240,800

d.



Style	SqFeet	Bedrooms	Price
RANCH	1250	2	$64,000
SPLIT	1190	1	$65,850
CONDO	1400	2	$80,050
TWOSTORY	1810	4	$107,250
RANCH	1500	3	$86,650
SPLIT	1615	4	$94,450
SPLIT	1305	3	$73,650
CONDO	1390	3	$79,350
TWOSTORY	1040	2	$55,850
CONDO	2105	4	$127,150
RANCH	1535	3	$89,100
TWOSTORY	1240	2	$69,250
RANCH	720	1	$34,550
TWOSTORY	1745	4	$102,950
CONDO	1860	2	$110,700

Correct answer: a

This program creates a summary report, which consolidates into one row all observations from the data set that have a unique combination of values for the variable `Style`.

2. When you define an order variable,

a. the detail rows are ordered according to their formatted values.

b. you can't create summary reports.

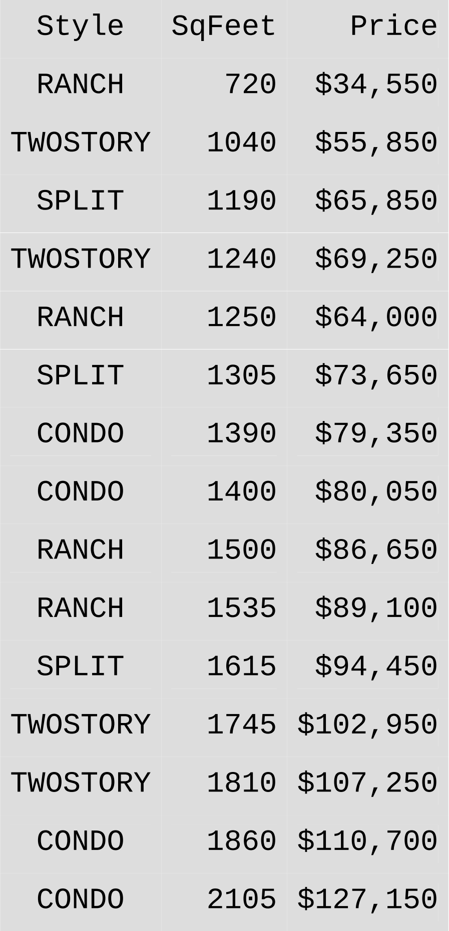
c. PROC REPORT displays only the first occurrence of each order variable value in a set of rows that have the same value for all order variables.

d. all of the above

Correct answer: d

Order variables do order rows according to the formatted values of the order variable, and PROC REPORT suppresses repetitious printing of order values. However, you can't use order variables in a summary report.

3. Which attributes or options are reflected in this PROC REPORT output?



```
   Style    SqFeet     Price
   RANCH       720   $34,550
TWOSTORY      1040   $55,850
   SPLIT      1190   $65,850
TWOSTORY      1240   $69,250
   RANCH      1250   $64,000
   SPLIT      1305   $73,650
   CONDO      1390   $79,350
   CONDO      1400   $80,050
   RANCH      1500   $86,650
   RANCH      1535   $89,100
   SPLIT      1615   $94,450
TWOSTORY      1745  $102,950
TWOSTORY      1810  $107,250
   CONDO      1860  $110,700
   CONDO      2105  $127,150
```

a. SKIPLINE and FORMAT=

b. CENTER, HEADLINE, HEADSKIP, and either WIDTH=, SPACING=, or FORMAT=

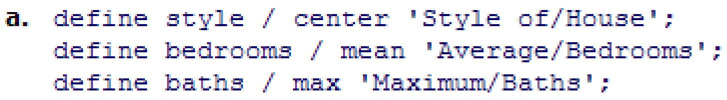
c. SPACING= only

d. CENTER, FORMAT=, and HEADLINE

Correct answer: b

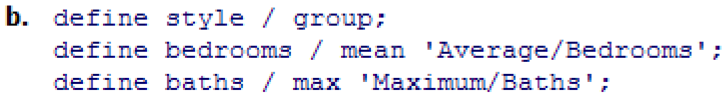
The HEADLINE option underlines the headings, and the HEADSKIP option skips a line between the headings and the rows in the report. Also, `Style` is centered, and the column for `Price` is wider than the default.

4. To create a summary report that shows the average number of bedrooms and the maximum number of baths for each style of house, which DEFINE statements do you use in your PROC REPORT step?



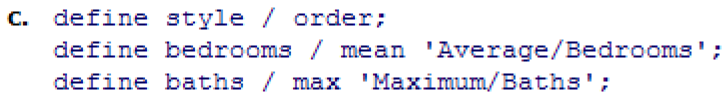
a.

```
define style / center 'Style of/House';
define bedrooms / mean 'Average/Bedrooms';
define baths / max 'Maximum/Baths';
```



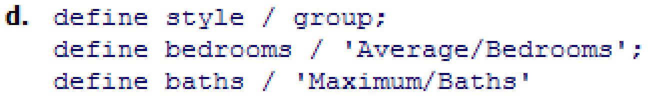
b.

```
define style / group;
define bedrooms / mean 'Average/Bedrooms';
define baths / max 'Maximum/Baths';
```



c.

```
define style / order;
define bedrooms / mean 'Average/Bedrooms';
define baths / max 'Maximum/Baths';
```



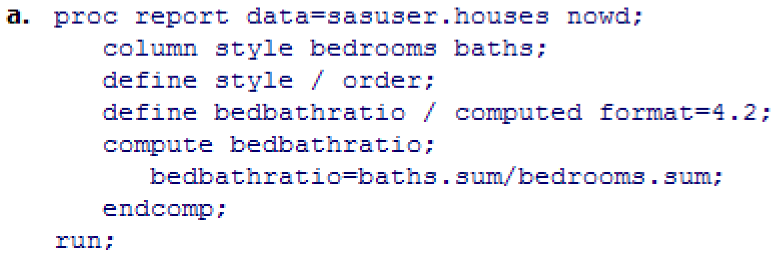
d.

```
define style / group;
define bedrooms / 'Average/Bedrooms';
define baths / 'Maximum/Baths'
```

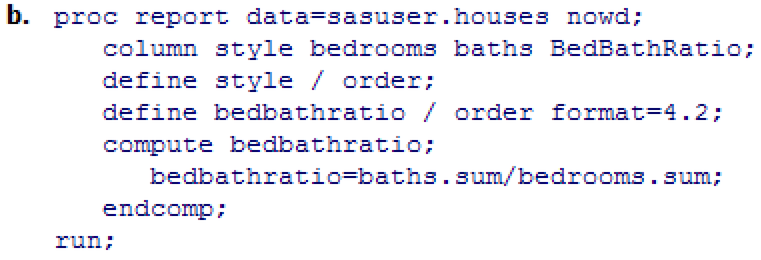
Correct answer: b

To create a summary report, you must define a group variable. To produce the statistics that you want, you must specify the MEAN and MAX statistics for `Bedrooms` and `Baths`.

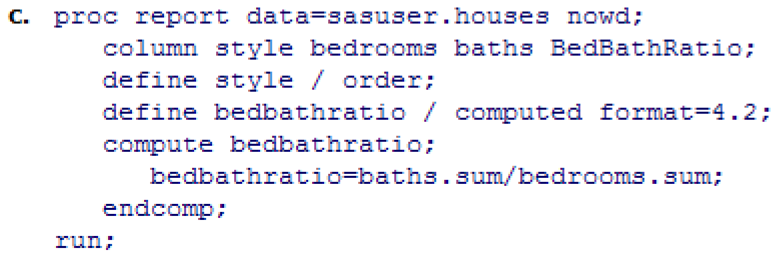
5. Which program does not contain an error?



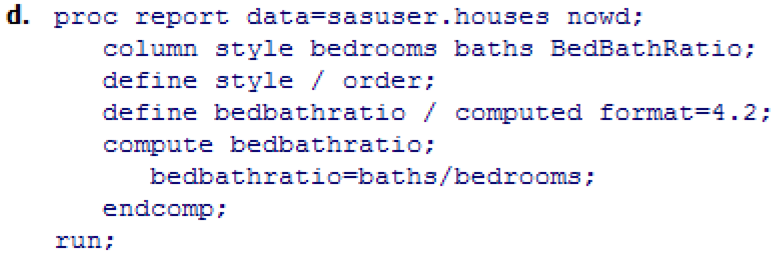
a.
```
proc report data=sasuser.houses nowd;
   column style bedrooms baths;
   define style / order;
   define bedbathratio / computed format=4.2;
   compute bedbathratio;
      bedbathratio=baths.sum/bedrooms.sum;
   endcomp;
run;
```



b.
```
proc report data=sasuser.houses nowd;
   column style bedrooms baths BedBathRatio;
   define style / order;
   define bedbathratio / order format=4.2;
   compute bedbathratio;
      bedbathratio=baths.sum/bedrooms.sum;
   endcomp;
run;
```



c.
```
proc report data=sasuser.houses nowd;
   column style bedrooms baths BedBathRatio;
   define style / order;
   define bedbathratio / computed format=4.2;
   compute bedbathratio;
      bedbathratio=baths.sum/bedrooms.sum;
   endcomp;
run;
```



d.
```
proc report data=sasuser.houses nowd;
   column style bedrooms baths BedBathRatio;
   define style / order;
   define bedbathratio / computed format=4.2;
   compute bedbathratio;
      bedbathratio=baths/bedrooms;
   endcomp;
run;
```

Correct answer: c

Program c correctly specifies a computed variable in the COLUMN statement, defines the variable in a DEFINE statement, and computes values using the form variable-name.statistic in a compute block.

6. What output does this PROC REPORT step produce?

```
proc report data=sasuser.houses nowd;
   column style sqfeet bedrooms price;
run;
```

a. a list report ordered by values of the first variable in the COLUMN statement

b. a summary report ordered by values of the first variable in the COLUMN statement

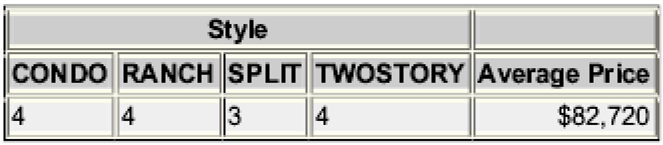
c. a list report that displays a row for each observation in the input data set and which calculates the SUM statistic for numeric variables

d. a list report that calculates the N (frequency) statistic for character variables

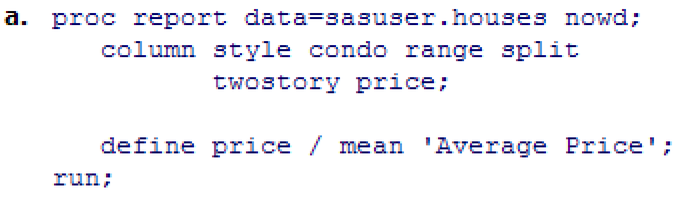
Correct answer: c

By default, PROC REPORT displays character variables as display variables. A report that contains one or more display variables has a detail row for each observation in the data set. By default, PROC REPORT displays numeric variables as analysis variables, which are used to calculate the default statistic SUM.

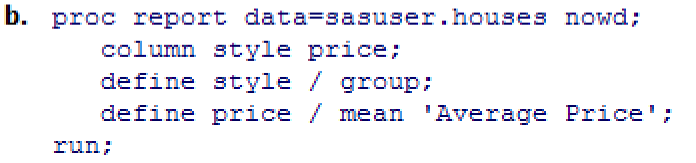
7. Which of the following programs produces this output?



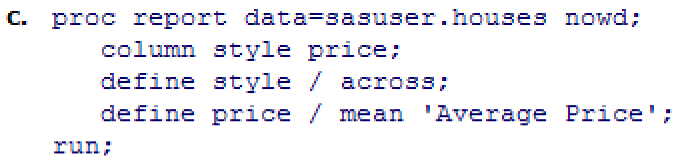
Style				
CONDO	RANCH	SPLIT	TWOSTORY	Average Price
4	4	3	4	$82,720



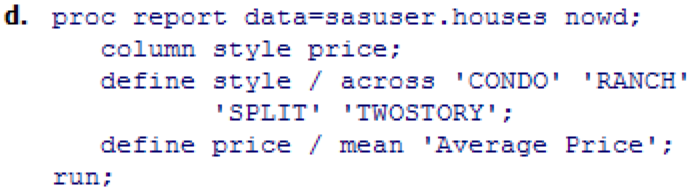
a.
```
proc report data=sasuser.houses nowd;
   column style condo range split
          twostory price;

   define price / mean 'Average Price';
run;
```



b.
```
proc report data=sasuser.houses nowd;
   column style price;
   define style / group;
   define price / mean 'Average Price';
run;
```



c.
```
proc report data=sasuser.houses nowd;
   column style price;
   define style / across;
   define price / mean 'Average Price';
run;
```



d.
```
proc report data=sasuser.houses nowd;
   column style price;
   define style / across 'CONDO' 'RANCH'
          'SPLIT' 'TWOSTORY';
   define price / mean 'Average Price';
run;
```

Correct c

answer:

In this output, the table cells contain a frequency count for each unique value of an across variable, `Style`. You don't have to specify across variable values in your PROC REPORT step.

8. If you submit this program, where does your PROC REPORT output appear?

```
proc report data=sasuser.houses nowd;
   column style sqfeet bedrooms price;
   define style / group;
run;
```

a. in the PROC REPORT window

b. as HTML and/or SAS listing output

c. both of the above

d. neither of the above

Correct answer: b

In nonwindowing mode, your PROC REPORT output appears as HTML and/or as SAS listing output, depending on your option settings.

9. How can you create output with headings that break as shown below?

```
Style of    Average     Maximum
 House     Bedrooms       Baths
 CONDO         2.75         2.5
 RANCH         2.25           3
 SPLIT     2.666666           3
TWOSTORY          3           3
```

a. You must specify the SPLIT= option in the PROC REPORT statement and use the split character in column headings in DEFINE statements.

b. You must use the default split character in column headings in DEFINE statements.

c. You must specify either the WIDTH= or the SPACING= attribute in DEFINE statements.

d. These headings split this way by default.

Correct answer: d

By default, columns for character variables are the same as the variable's length, and columns for numeric variables have a width of 9. So these headings split this way by default.

10. Suppose you want to create a report using both character and numeric variables. If you don't use any DEFINE statements in your PROC REPORT step,

a. your PROC REPORT step will not execute successfully.

b. you can produce only list reports.

c. you can order rows by specifying options in the PROC REPORT statement.

d. you can produce only summary reports.

Correct answer: b .Unless you use DEFINE statements to define order variables or group variables, you can't order rows or produce summary reports. However, DEFINE statements are not required in all PROC REPORT steps.

Chapter 9: Producing Descriptive Statistics Answer Key

1. The default statistics produced by the MEANS procedure are n-count, mean, minimum, maximum, and

a. median.

b. range.

c. standard deviation.

d. standard error of the mean.

Correct answer: c

By default, the MEANS procedure produces the n-cout, mean, minimum, and standard deviation.

2. Which statement will limit a PROC MEANS analysis to the variables `Boarded`, `Transfer`, and `Deplane`?

a. `by boarded transfer deplane;`

b. `class boarded transfer deplane;`

c. `output boarded transfer deplane;`

d. `var boarded transfer deplane;`

Correct answer: d

To specify the variables that PROC MEANS analyzes, add a VAR statement and list the variable names.

3. The data set Survey.Health includes the following variables. Which is a poor candidate for PROC MEANS analysis?

a. `IDnum`

b. `Age`

c. `Height`

d. `Weight`

Correct answer: a

Unlike `Age`, `Height`, or `Weight`, the values of IDnum are unlikely to yield any useful statistics.

4. Which of the following statements is true regarding BY-group processing?

a. BY variables must be either indexed or sorted.

b. Summary statistics are computed for BY variables.

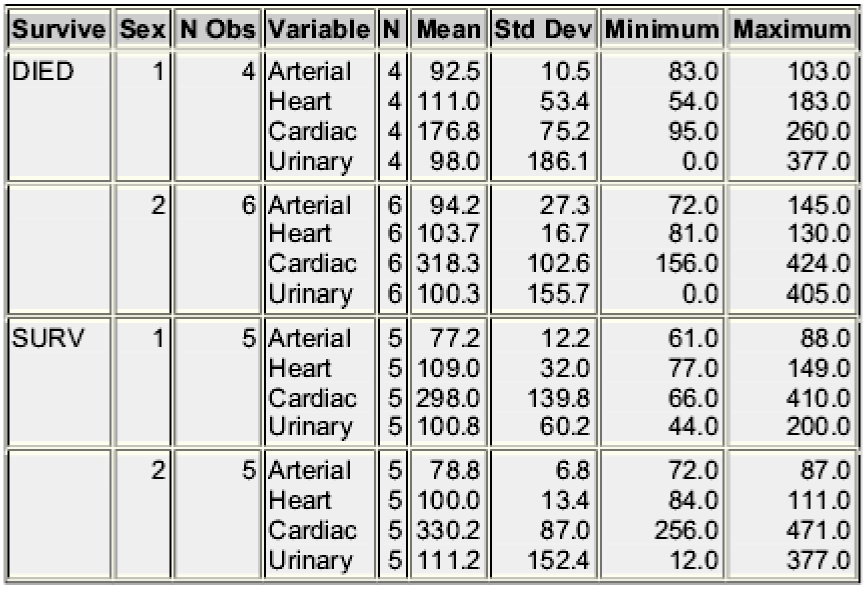
c. BY-group processing is preferred when you are categorizing data that contains few variables.

d. BY-group processing overwrites your data set with the newly grouped observations.

Correct answer: a

Unlike CLASS processing, BY-group processing requires that your data already be indexed or sorted in the order of the BY variables. You might need to run the SORT procedure before using PROC MEANS with a BY group.

5. Which group processing statement produced the PROC MEANS output shown below?



Survive	Sex	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum
DIED	1	4	Arterial	4	92.5	10.5	83.0	103.0
			Heart	4	111.0	53.4	54.0	183.0
			Cardiac	4	176.8	75.2	95.0	260.0
			Urinary	4	98.0	186.1	0.0	377.0
	2	6	Arterial	6	94.2	27.3	72.0	145.0
			Heart	6	103.7	16.7	81.0	130.0
			Cardiac	6	318.3	102.6	156.0	424.0
			Urinary	6	100.3	155.7	0.0	405.0
SURV	1	5	Arterial	5	77.2	12.2	61.0	88.0
			Heart	5	109.0	32.0	77.0	149.0
			Cardiac	5	298.0	139.8	66.0	410.0
			Urinary	5	100.8	60.2	44.0	200.0
	2	5	Arterial	5	78.8	6.8	72.0	87.0
			Heart	5	100.0	13.4	84.0	111.0
			Cardiac	5	330.2	87.0	256.0	471.0
			Urinary	5	111.2	152.4	12.0	377.0

a. `class sex survive;`

b. `class survive sex;`

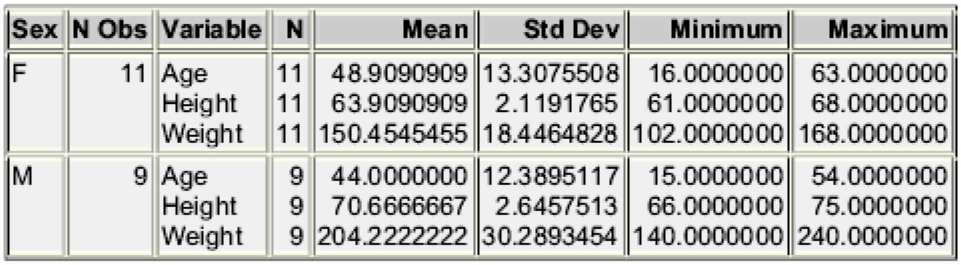
c. `by sex survive;`

d. `by survive sex;`

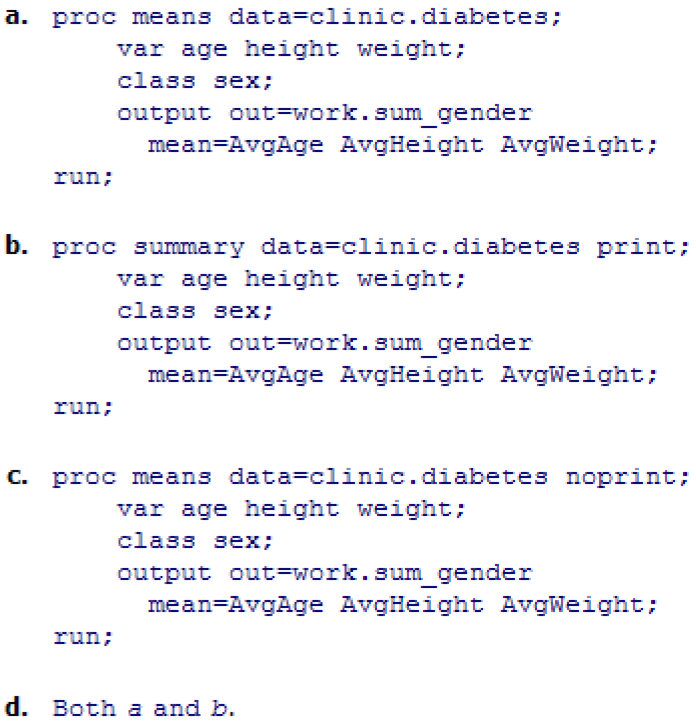
Correct answer: b

A CLASS statement produces a single large table, whereas BY-group processing creates a series of small tables. The order of the variables in the CLASS statement determines their order in the output table.

6. Which program can be used to create the following output?



Sex	N Obs	Variable	N	Mean	Std Dev	Minimum	Maximum
F	11	Age	11	48.9090909	13.3075508	16.0000000	63.0000000
		Height	11	63.9090909	2.1191765	61.0000000	68.0000000
		Weight	11	150.4545455	18.4464828	102.0000000	168.0000000
M	9	Age	9	44.0000000	12.3895117	15.0000000	54.0000000
		Height	9	70.6666667	2.6457513	66.0000000	75.0000000
		Weight	9	204.2222222	30.2893454	140.0000000	240.0000000



```
a. proc means data=clinic.diabetes;
       var age height weight;
       class sex;
       output out=work.sum_gender
         mean=AvgAge AvgHeight AvgWeight;
   run;

b. proc summary data=clinic.diabetes print;
       var age height weight;
       class sex;
       output out=work.sum_gender
         mean=AvgAge AvgHeight AvgWeight;
   run;

c. proc means data=clinic.diabetes noprint;
       var age height weight;
       class sex;
       output out=work.sum_gender
         mean=AvgAge AvgHeight AvgWeight;
   run;

d. Both a and b.
```

Correct answer: d

You can use either PROC MEANS or PROC SUMMARY to create the table. Adding a PRINT option to the PROC SUMMARY statement produces the same reports as if you used PROC MEANS.

7. By default, PROC FREQ creates a table of frequencies and percentages for which data set variables?

a. character variables

b. numeric variables

c. both character and numeric variables

d. none: variables must always be specified

Correct answer: c

By default, the PROC FREQ creates a table for all variables in a data set.

8. Frequency distributions work best with variables that contain

a. continuous values.

b. numeric values.

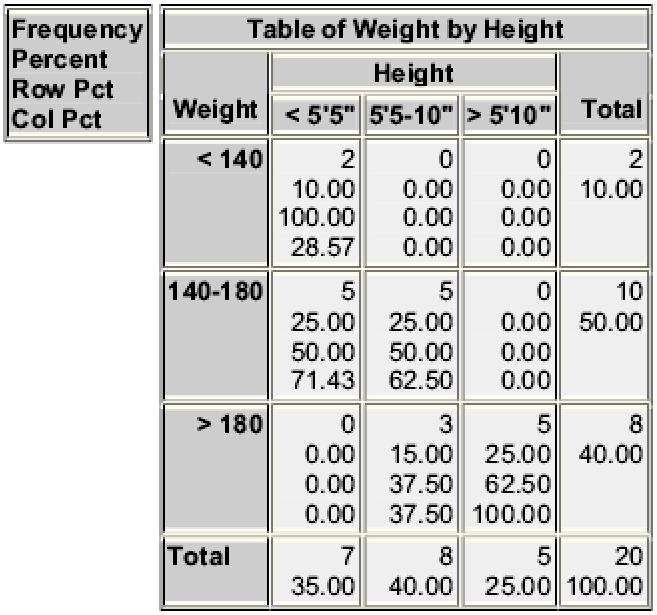
c. categorical values.

d. unique values.

Correct answer: c

Both continuous values and many unique values can result in lengthy and meaningless tables. Frequency distributors work best with categorical values.

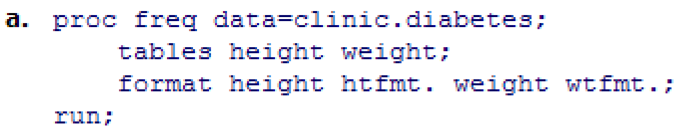
9. Which PROC FREQ step produced this two-way table?



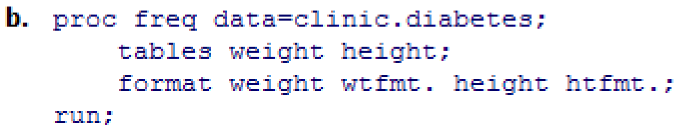
Frequency
Percent
Row Pct
Col Pct

Table of Weight by Height

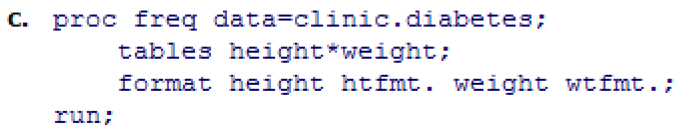
Weight	Height			Total
	< 5'5"	5'5-10"	> 5'10"	
< 140	2 10.00 100.00 28.57	0 0.00 0.00 0.00	0 0.00 0.00 0.00	2 10.00
140-180	5 25.00 50.00 71.43	5 25.00 50.00 62.50	0 0.00 0.00 0.00	10 50.00
> 180	0 0.00 0.00 0.00	3 15.00 37.50 37.50	5 25.00 62.50 100.00	8 40.00
Total	7 35.00	8 40.00	5 25.00	20 100.00



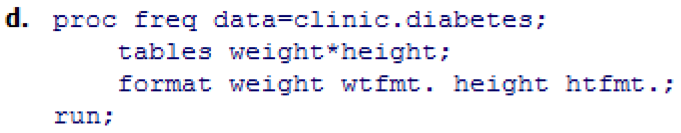
```
a. proc freq data=clinic.diabetes;
        tables height weight;
        format height htfmt. weight wtfmt.;
   run;
```



```
b. proc freq data=clinic.diabetes;
        tables weight height;
        format weight wtfmt. height htfmt.;
   run;
```



```
c. proc freq data=clinic.diabetes;
        tables height*weight;
        format height htfmt. weight wtfmt.;
   run;
```

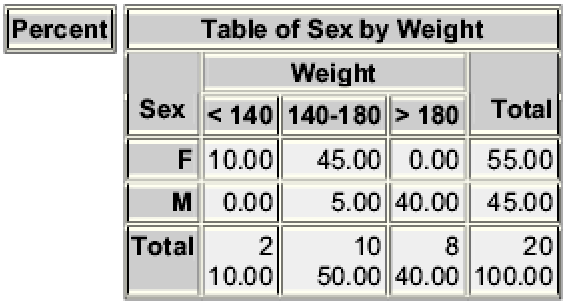


```
d. proc freq data=clinic.diabetes;
        tables weight*height;
        format weight wtfmt. height htfmt.;
   run;
```

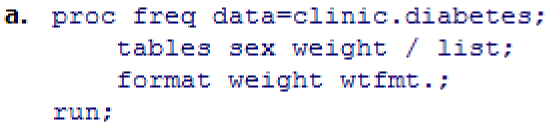
Correct answer: d

An asterisk is used to join the variables in a two-way TABLES statement. The first variable forms the table rows, and the second variable forms the table columns.

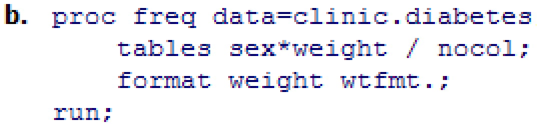
10. Which PROC FREQ step produced this table?



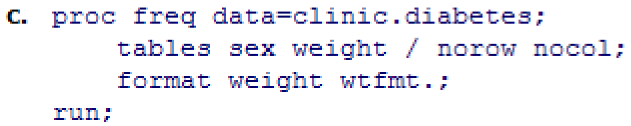
Percent	Table of Sex by Weight			
	Weight			
Sex	< 140	140-180	> 180	Total
F	10.00	45.00	0.00	55.00
M	0.00	5.00	40.00	45.00
Total	2 10.00	10 50.00	8 40.00	20 100.00



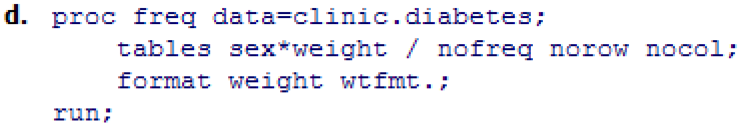
a.
```
proc freq data=clinic.diabetes;
     tables sex weight / list;
     format weight wtfmt.;
run;
```



b.
```
proc freq data=clinic.diabetes;
     tables sex*weight / nocol;
     format weight wtfmt.;
run;
```



c.
```
proc freq data=clinic.diabetes;
     tables sex weight / norow nocol;
     format weight wtfmt.;
run;
```



d.
```
proc freq data=clinic.diabetes;
     tables sex*weight / nofreq norow nocol;
     format weight wtfmt.;
run;
```

Correct answer: d

An asterisk is used to join the variables in crosstabulation tables. The only results that are shown in this table are cell percentages. The NOFREQ option suppresses cell frequencies, the NOROW option suppresses row percentages, and the NOCOL option suppresses column percentages.

Chapter 10: Producing HTML Output Answer Key

1. Using ODS statements, how many types of output can you generate concurrently?

a. 1 (only listing output)

b. 2

c. 3

d. as many as you want

Correct answer: d

You can generate any number of output types as long as you open the ODS destination for each type of output that you want to create.

2. If ODS is set to its default settings, what types of output are created by the code below?

```
ods html file='c:\myhtml.htm';
ods pdf file='c:\mypdf.pdf';
```

a. HTML and PDF

b. PDF only

c. HTML, PDF, and listing

d. No output is created because ODS is closed by default.

Correct answer: c

Listing output is created by default, so these statements create HTML, PDF, and listing output.

3. What is the purpose of closing the Listing destination in the code shown below?

```
ods listing close;
ods html ... ;
```

a. It conserves system resources.

b. It simplifies your program.

c. It makes your program compatible with other hardware platforms.

d. It makes your program compatible with previous versions of SAS software.

Correct answer: a

By default, SAS programs produce listing output. If you want only HTML output, it's a good idea to close the Listing destination before creating HTML output, because an open destination uses system resources.

4. When the code shown below is run, what will the file D:\Output\body.html contain?

```
ods html body='d:\output\body.html';
proc print data=work.alpha;
run;
proc print data=work.beta;
run;
ods html close;
```

a. The PROC PRINT output for Work.Alpha.

b. The PROC PRINT output for Work.Beta.

c. The PROC PRINT output for both Work.Alpha and Work.Beta.

d. Nothing. No output will be written to D:\Output\body.html.

Correct answer: c

When multiple procedures are run while HTML output is open, procedure output is appended to the same body file.

5. When the code shown below is run, what file will be loaded by the links in D:\Output\contents.html?

```
ods html body='d:\output\body.html'
         contents='d:\output\contents.html'
         frame='d:\output\frame.html';
```

a. D:\Output\body.html

b. D:\Output\contents.html

c. D:\Output\frame.html

d. There are no links from the file D:\Output\contents.html.

Correct answer: a

The CONTENTS= option creates a table of contents containing links to the body file, D:\Output\body.html.

6. The table of contents created by the CONTENTS= option contains a numbered heading for

a. each procedure.

b. each procedure that creates output.

c. each procedure and DATA step.

d. each HTML file created by your program.

Correct answer: b

The table of contents contains a numbered heading for each procedure that creates output.

7. When the code shown below is run, what will the file D:\Output\frame.html display?

```
ods html body='d:\output\body.html'
         contents='d:\output\contents.html'
         frame='d:\output\frame.html';
```

a. The file D:\Output\contents.html.

b. The file D:\Output\frame.html.

c. The files D:\Output\contents.html and D:\Output\body.html.

d. It displays no other files.

Correct answer: c

The FRAME= option creates an HTML file that integrates the table of contents and the body file.

8. What is the purpose of the URL= suboptions shown below?

```
ods html body='d:\output\body.html' (url='body.html')
         contents='d:\output\contents.html'
(url='contents.html')
         frame='d:\output\frame.html';
```

a. To create absolute link addresses for loading the files from a server.

b. To create relative link addresses for loading the files from a server.

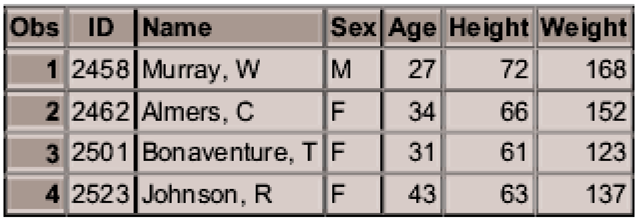
c. To allow HTML files to be loaded from a local drive.

d. To send HTML output to two locations.

Correct answer: b

Specifying the URL= suboption in the file specification provides a URL that ODS uses in the links it creates. Specifying a simple (one name) URL creates a relative link address to the file.

9. Which ODS HTML option was used in creating the following table?



Obs	ID	Name	Sex	Age	Height	Weight
1	2458	Murray, W	M	27	72	168
2	2462	Almers, C	F	34	66	152
3	2501	Bonaventure, T	F	31	61	123
4	2523	Johnson, R	F	43	63	137

a. `format=brown`

b. `format='brown'`

c. `style=brown`

d. `style='brown'`

Correct answer: c

You can change the appearance of HTML output by using the STYLE= option in the ODS HTML statement. The style name doesn't need quotation marks.

10. What is the purpose of the PATH= option?

```
ods html path='d:\output' (url=none)
         body='body.html'
         contents='contents.html'
         frame='frame.html';
```

a. It creates absolute link addresses for loading HTML files from a server.

b. It creates relative link addresses for loading HTML files from a server.

c. It allows HTML files to be loaded from a local drive.

d. It specifies the location of HTML file output.

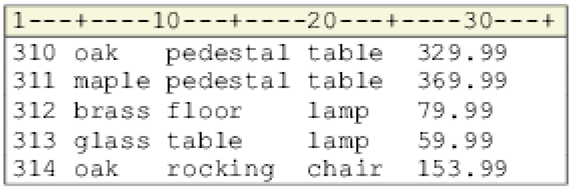
Correct answer: d

You use the PATH= option to specify the location for HTML output. When you use the PATH= option, you don't need to specify the full pathname for the body, contents, or frame files.

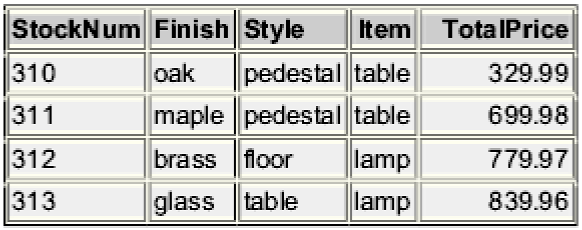
Chapter 11: Creating and Managing Variables Answer Key

1. Which program creates the output shown below?

Raw Data File Furnture



```
1---+----10---+----20---+----30---+
310 oak   pedestal table  329.99
311 maple pedestal table  369.99
312 brass floor    lamp   79.99
313 glass table    lamp   59.99
314 oak   rocking  chair  153.99
```



StockNum	Finish	Style	Item	TotalPrice
310	oak	pedestal	table	329.99
311	maple	pedestal	table	699.98
312	brass	floor	lamp	779.97
313	glass	table	lamp	839.96

```
a. data test2;
      infile furnture;
      input StockNum $ 1-3 Finish $ 5-9 Style $ 11-18
            Item $ 20-24 Price 26-31;
      if finish='oak' then delete;
      retain TotPrice 100;
      totalprice+price;
      drop price;
   run;
   proc print data=test2 noobs;
   run;

b. data test2;
      infile furnture;
      input StockNum $ 1-3 Finish $ 5-9 Style $ 11-18
            Item $ 20-24 Price 26-31;
      if finish='oak' and price<200 then delete;
      TotalPrice+price;
   run;
   proc print data=test2 noobs;
   run;

c. data test2(drop=price);
      infile furnture;
      input StockNum $ 1-3 Finish $ 5-9 Style $ 11-18
            Item $ 20-24 Price 26-31;
      if finish='oak' and price<200 then delete;
      TotalPrice+price;
   run;
   proc print data=test2 noobs;
   run;

d. data test2;
      infile furnture;
      input StockNum $ 1-3 Finish $ 5-9 Style $ 11-18
            Item $ 20-24 Price 26-31;
      if finish=oak and price<200 then delete price;
      TotalPrice+price;
   run;
   proc print data=test2 noobs;
   run;
```

Correct answer: c

Program c correctly deletes the observation in which the value of Finish is oak and the value of `Price` is less than also creates `TotalPrice` by summing the variable `Price` down observations, then drops `Price` by using the D data set option in the DATA statement.

2. How is the variable `Amount` labeled and formatted in the PROC PRINT output?

```
data credit;
   infile creddata;
   input Account $ 1-5 Name $ 7-25 Type $ 27
         Transact $ 29-35 Amount 37-50;
   label amount='Amount of Loan';
   format amount dollar12.2;

run;
proc print data=credit label;
   label amount='Total Amount Loaned';
   format amount comma10.;
run;
```

a. label Amount of Loan, format DOLLAR12.2

b. label Total Amount Loaned, format COMMA10.

c. label Amount, default format

d. The PROC PRINT step does not execute because two labels and two formats are assigned to the same vari

Correct answer: b

The PROC PRINT output displays the label Total Amount Loaned for the variable `Amount` and formats this varial the COMMA10. format. Temporary labels or formats that are assigned in a PROC step override permanent labels c that are assigned in a DATA step.

3. Consider the IF-THEN statement shown below. When the statement is executed, which expression is evaluated firs

```
if finlexam>=95
   and (research='A' or
       (project='A' and present='A'))
   then Grade='A+';
```

a. `finlexam>=95`

b. `research='A'`

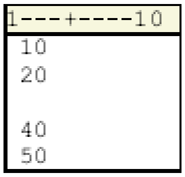
c. `project='A' and present='A'`

```
d.  research='A' or
e.  (project='A' and present='A')
```

Correct answer: c

Logical comparisons that are enclosed in parentheses are evaluated as true or false before they are compared to oth expressions. In the example above, the AND comparison within the nested parentheses is evaluated before being co the OR comparison.

4. Consider the small raw data file and program shown below. What is the value of `Count` after the fourth record is r



```
1---+----10
 10
 20

 40
 50
```

```
data work.newnums;
   infile numbers;
   input Tens 2-3;
   Count+tens;
run;
```

a. missing

b. 0

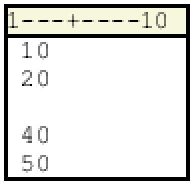
c. 30

d. 70

Correct answer: d

The Sum statement adds the result of the expression that is on the right side of the plus sign to the numeric variable the left side. The new value is then retained for subsequent observations. The Sum statement treats the missing valu so the value of `Count` in the fourth observation would be 10+20+0+40, or 70.

5. Now consider the revised program below. What is the value of `Count` after the third observation is read?



```
1---+----10
 10
 20

 40
 50
```

```
data work.newnums;
   infile numbers;
   input Tens 2-3;
   retain Count 100;
   count+tens;
run;
```

a. missing

b. 0

c. 100

d. 130

Correct answer: d

The RETAIN statement assigns an initial value of 100 to the variable `Count`, so the value of `Count` in the third o would be 100+10+20+0, or 130.

6. For the observation shown below, what is the result of the IF-THEN statement?

Status	Type	Count	Action	Control
ok	3	12	E	Go

```
if status='OK' and type=3
   then Count+1;
if status='S' or action='E'
   then Control='Stop';
```

a. `Count` = 12 `Control` = Go

b. `Count` = 13 `Control` = Stop

c. `Count` = 12 `Control` = Stop

d. `Count` = 13 `Control` = Go

Correct answer: c

You must enclose character values in quotation marks, and you must specify them in the same case in which they a the data set. The value ok is not identical to OK, so the value of `Count` is not changed by the IF-THEN statement.

7. Which of the following can determine the length of a new variable?

a. the length of the variable's first value

b. the assignment statement

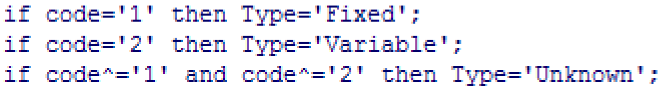
c. the LENGTH statement

d. all of the above

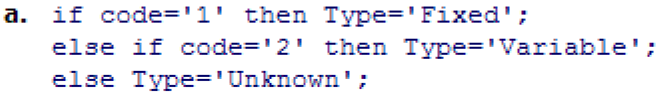
Correct answer: d

The length of a variable is determined by its first reference in the DATA step. When creating a new character varia allocates as many bytes of storage space as there are characters in the first value that it encounters for that variable. reference to a new variable can also be made with a LENGTH statement or an assignment statement. The length of variable's first value does not matter once the variable has been referenced in your program.

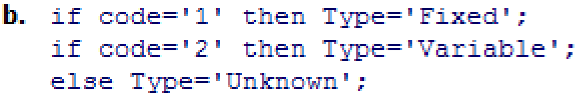
8. Which set of statements is the most efficient equivalent to the code shown below?



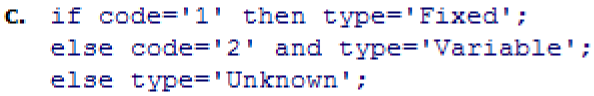
```
if code='1' then Type='Fixed';
if code='2' then Type='Variable';
if code^='1' and code^='2' then Type='Unknown';
```



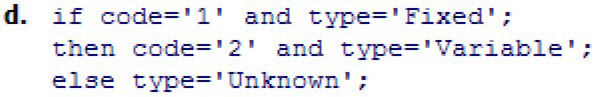
```
a. if code='1' then Type='Fixed';
   else if code='2' then Type='Variable';
   else Type='Unknown';
```



```
b. if code='1' then Type='Fixed';
   if code='2' then Type='Variable';
   else Type='Unknown';
```



```
c. if code='1' then type='Fixed';
   else code='2' and type='Variable';
   else type='Unknown';
```



```
d. if code='1' and type='Fixed';
   then code='2' and type='Variable';
   else type='Unknown';
```

Correct answer: a

Answer a is the most efficient. You can write multiple ELSE statements to specify a series of mutually exclusive co The ELSE statement must immediately follow the IF-THEN statement in your program. An ELSE statement execu the previous IF-THEN/ELSE statement is false.

9. What is the length of the variable `Type`, as created in the DATA step below?

```
data finance.newloan;
   set finance.records;
   TotLoan+payment;
   if code='1' then Type='Fixed';
   else Type='Variable';
   length type $ 10;
run;
```

a. 5

b. 8

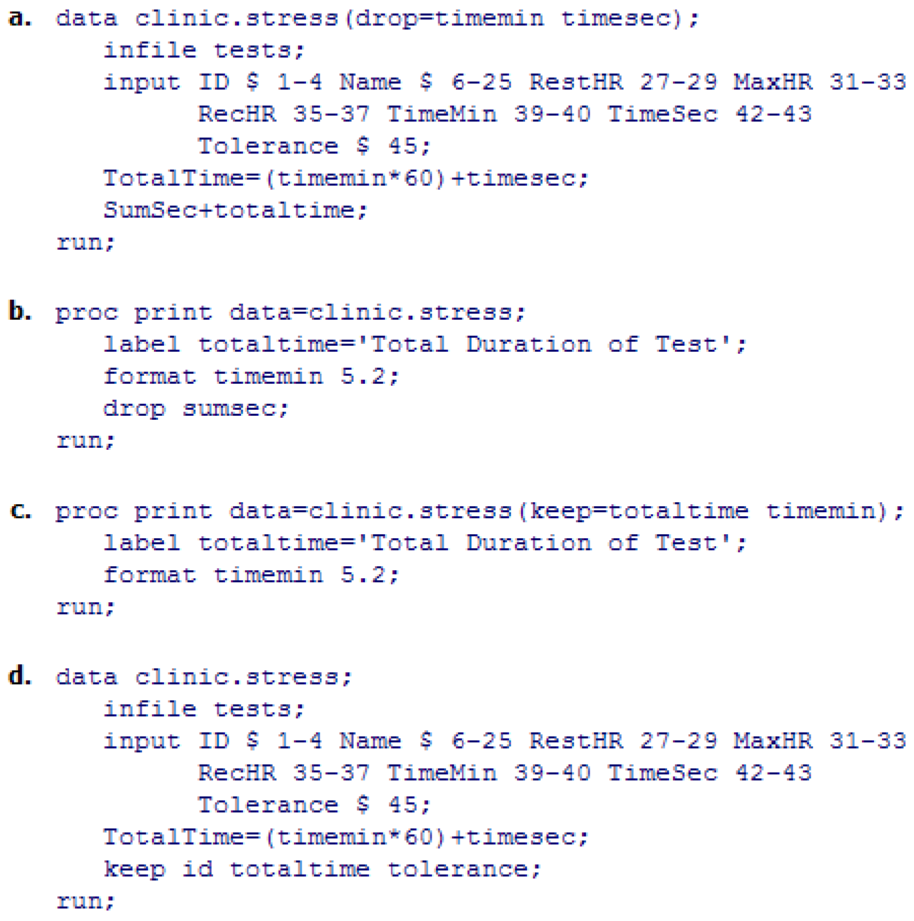
c. 10

d. It depends on the first value of `Type`.

Correct answer: a

The length of a new variable is determined by the first reference in the DATA step, not by data values. In this case, of `Type` is determined by the value Fixed. The LENGTH statement is in the wrong place; it must be read before an reference to the variable in the DATA step. The LENGTH statement cannot change the length of an existing variab

10. Which program contains an error?



```
a. data clinic.stress(drop=timemin timesec);
      infile tests;
      input ID $ 1-4 Name $ 6-25 RestHR 27-29 MaxHR 31-33
            RecHR 35-37 TimeMin 39-40 TimeSec 42-43
            Tolerance $ 45;
      TotalTime=(timemin*60)+timesec;
      SumSec+totaltime;
   run;

b. proc print data=clinic.stress;
      label totaltime='Total Duration of Test';
      format timemin 5.2;
      drop sumsec;
   run;

c. proc print data=clinic.stress(keep=totaltime timemin);
      label totaltime='Total Duration of Test';
      format timemin 5.2;
   run;

d. data clinic.stress;
      infile tests;
      input ID $ 1-4 Name $ 6-25 RestHR 27-29 MaxHR 31-33
            RecHR 35-37 TimeMin 39-40 TimeSec 42-43
            Tolerance $ 45;
      TotalTime=(timemin*60)+timesec;
      keep id totaltime tolerance;
   run;
```

Correct answer: b

To select variables, you can use a DROP or KEEP statement in any DATA step. You can also use the DR KEEP= data set options following a data set name in any DATA or PROC step. However, you cannot use KEEP statements in PROC steps.

Chapter 12: Reading SAS Data Sets Answer Key

1. If you submit the following program, which variables appear in the new data set?

```
data work.cardiac(drop=age group);
   set clinic.fitness(keep=age weight group);
   if group=2 and age>40;
run;
```

a. none

b. `Weight`

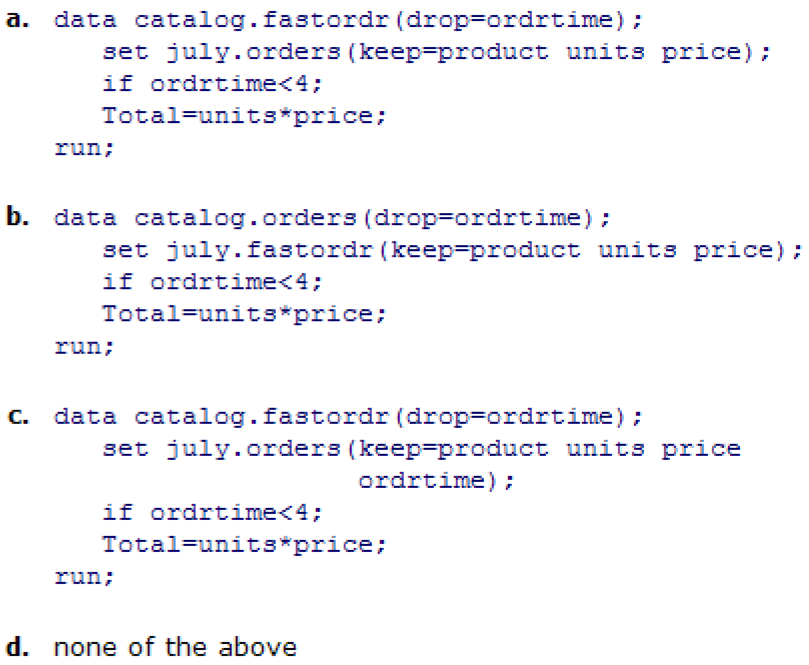
c. `Age`, `Group`

d. `Age`, `Weight`, `Group`

Correct answer: b

The variables `Age`, `Weight`, and `Group` are specified using the KEEP= option in the SET statement. After processing, `Age` and `Group` are dropped in the DATA statement.

2. Which of the following programs correctly reads the data set Orders and creates the data set FastOrdr?



```
a. data catalog.fastordr(drop=ordrtime);
      set july.orders(keep=product units price);
      if ordrtime<4;
      Total=units*price;
   run;

b. data catalog.orders(drop=ordrtime);
      set july.fastordr(keep=product units price);
      if ordrtime<4;
      Total=units*price;
   run;

c. data catalog.fastordr(drop=ordrtime);
      set july.orders(keep=product units price
                      ordrtime);
      if ordrtime<4;
      Total=units*price;
   run;

d. none of the above
```

Correct answer: c

You specify the data set to be created in the DATA statement. The DROP= data set option prevents variables from being written to the data set. Because you use the variable `OrdrTime` when processing your data, you cannot drop `OrdrTime` in the SET statement. If you use the KEEP= option in the SET statement, then you must list `OrdrTime` as one of the variables to be kept.

3. Which of the following statements is false about BY-group processing?

When you use the BY statement with the SET statement,

a. the data sets that are listed in the SET statement must be indexed or sorted by the values of the BY variable(s).

b. the DATA step automatically creates two variables, `FIRST.` and `LAST.`, for each variable in the BY statement.

c. `FIRST.` and `LAST.` identify the first and last observation in each BY group, in that order.

d. `FIRST.` and `LAST.` are stored in the data set.

Correct d

answer:

When you use the BY statement with the SET statement, the DATA step creates the temporary variables `FIRST.` and `LAST.` They are not stored in the data set.

4. There are 500 observations in the data set Company.USA. What is the result of submitting the following program?

```
data work.getobs5(drop=obsnum);
   obsnum=5;
   set company.usa(keep=manager payroll) point=obsnum;
   stop;
run;
```

a. an error

b. an empty data set

c. a continuous loop

d. a data set that contains one observation

Correct answer: b

The DATA step writes observations to output at the end of the DATA step. However, in this program, the STOP statement stops processing before the end of the DATA step. An explicit OUTPUT statement is needed in order to produce observations.

5. There is no end-of-file condition when you use direct access to read data, so how can your program prevent a continuous loop?

a. Do not use a POINT= variable.

b. Check for an invalid value of the POINT= variable.

c. Do not use an END= variable.

d. Include an OUTPUT statement.

Correct answer: b

To avoid a continuous loop when using direct access, either include a STOP statement or use programming logic that checks for an invalid value of the POINT= variable. If SAS reads an invalid value of the POINT= variable, it sets the automatic variable `_ERROR_` to 1. You can use this information to check for conditions that cause continuous processing.

6. Assuming that the data set Company.USA has five or more observations, what is the result of submitting the following program?

```
data work.getobs5(drop=obsnum);
   obsnum=5;
   set company.usa(keep=manager payroll) point=obsnum;
   output;
   stop;
run;
```

a. an error

b. an empty data set

c. a continuous loop

d. a data set that contains one observation

Correct answer: d

By combining the POINT= option with the OUTPUT and STOP statements, your program can write a single observation to output.

7. Which of the following statements is true regarding direct access of data sets?

a. You cannot specify END= with POINT=.

b. You cannot specify OUTPUT with POINT=.

c. You cannot specify STOP with END=.

d. You cannot specify `FIRST.` with `LAST.`

Correct answer: a

The END= option and POINT= option are incompatible in the same SET statement. Use one or the other in your program.

8. What is the result of submitting the following program?

```
data work.addtoend;
   set clinic.stress2 end=last;
   if last;
run;
```

a. an error

b. an empty data set

c. a continuous loop

d. a data set that contains one observation

Correct answer: d

This program uses the END= option to name a temporary variable that contains an end-of-file marker. That variable, `LAST`, is set to 1 when the SET statement reads the last observation of the data set.

9. At the start of DATA step processing, during the compilation phase, variables are created in the program data vector (PDV), and observations are set to

a. blank

b. missing

c. 0

d. there are no observations.

Correct answer: d

At the bottom of the DATA step, the compilation phase is complete, and the descriptor portion of the new SAS data set is created. There are no observations because the DATA step has not yet executed.

10. The DATA step executes

a. continuously if you use the POINT= option and the STOP statement.

b. once for each variable in the output data set.

c. once for each observation in the input data set.

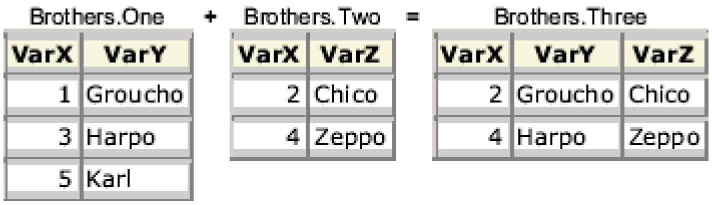
d. until it encounters an OUTPUT statement.

Correct answer: c

The DATA step executes once for each observation in the input data set. You use the POINT= option with the STOP statement to prevent continuous looping.

Chapter 13: Combining SAS Data Sets Answer Key

1. Which program will combine Brothers.One and Brothers.Two to produce Brothers.Three?



Brothers.One + Brothers.Two = Brothers.Three

Brothers.One

VarX	VarY
1	Groucho
3	Harpo
5	Karl

Brothers.Two

VarX	VarZ
2	Chico
4	Zeppo

Brothers.Three

VarX	VarY	VarZ
2	Groucho	Chico
4	Harpo	Zeppo

a.
```
data brothers.three;
   set brothers.one;
   set brothers.two;
run;
```

b.
```
data brothers.three;
   set brothers.one brothers.two;
run;
```

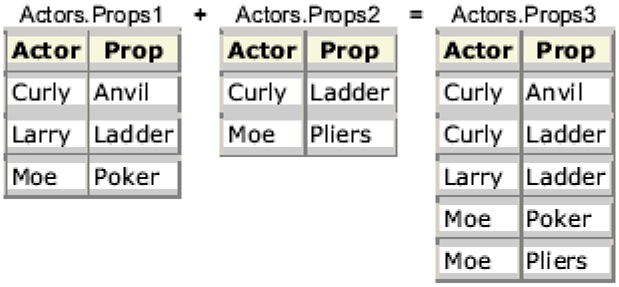
c.
```
data brothers.three;
   set brothers.one brothers.two;
   by varx;
run;
```

d.
```
data brothers.three;
   merge brothers.one brothers.two;
   by varx;
run;
```

Correct answer: a

This is a case of one-to-one reading, which requires multiple SET statements. Notice that where same-named variables occur, the values that are read in from the second data set replace those that are read in from the first one. Also, the number of observations in the new data set is the number of observations in the smallest original data set.

2. Which program will combine Actors.Props1 and Actors.Props2 to produce Actors.Props3?



Actors.Props1 + Actors.Props2 = Actors.Props3

Actor	Prop
Curly	Anvil
Larry	Ladder
Moe	Poker

Actor	Prop
Curly	Ladder
Moe	Pliers

Actor	Prop
Curly	Anvil
Curly	Ladder
Larry	Ladder
Moe	Poker
Moe	Pliers

a.
```
data actors.props3;
   set actors.props1;
   set actors.props2;
run;
```

b.
```
data actors.props3;
   set actors.props1 actors.props2;
run;
```

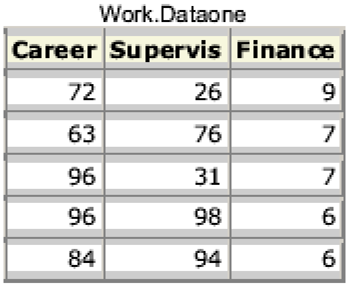
c.
```
data actors.props3;
   set actors.props1 actors.props2;
   by actor;
run;
```

d.
```
data actors.props3;
   merge actors.props1 actors.props2;
   by actor;
run;
```

Correct answer: c

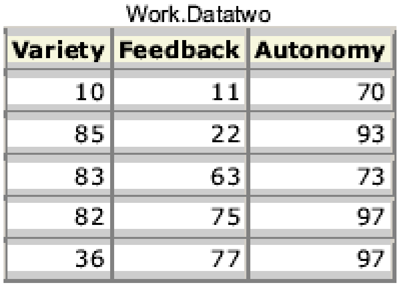
This is a case of interleaving, which requires a list of data set names in the SET statement and one or more BY variables in the BY statement. Notice that observations in each BY group are read sequentially, in the order in which the data sets and BY variables are listed. The new data set contains all the variables from all the input data sets, as well as the total number of records from all input data sets.

3. If you submit the following program, which new data set is created?



Work.Dataone

Career	Supervis	Finance
72	26	9
63	76	7
96	31	7
96	98	6
84	94	6



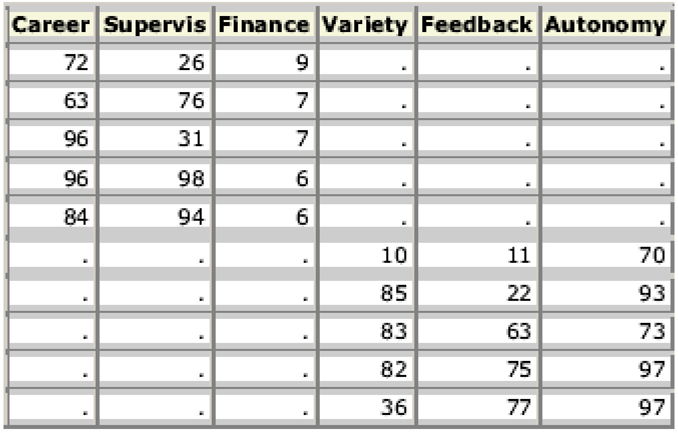
Work.Datatwo

Variety	Feedback	Autonomy
10	11	70
85	22	93
83	63	73
82	75	97
36	77	97

```
data work.jobsatis;
```

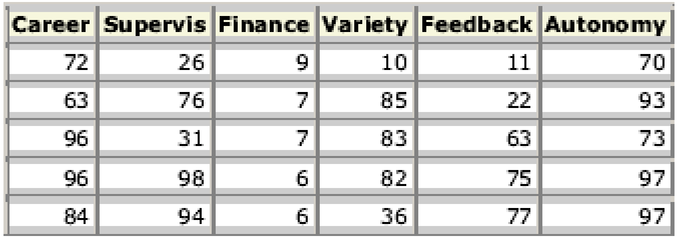
```
   set work.dataone work.datatwo;
run;
```

a.



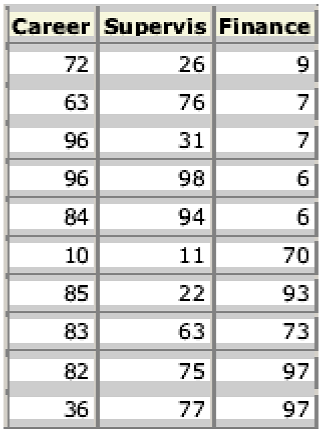
Career	Supervis	Finance	Variety	Feedback	Autonomy
72	26	9	.	.	.
63	76	7	.	.	.
96	31	7	.	.	.
96	98	6	.	.	.
84	94	6	.	.	.
.	.	.	10	11	70
.	.	.	85	22	93
.	.	.	83	63	73
.	.	.	82	75	97
.	.	.	36	77	97

b.



Career	Supervis	Finance	Variety	Feedback	Autonomy
72	26	9	10	11	70
63	76	7	85	22	93
96	31	7	83	63	73
96	98	6	82	75	97
84	94	6	36	77	97

c.



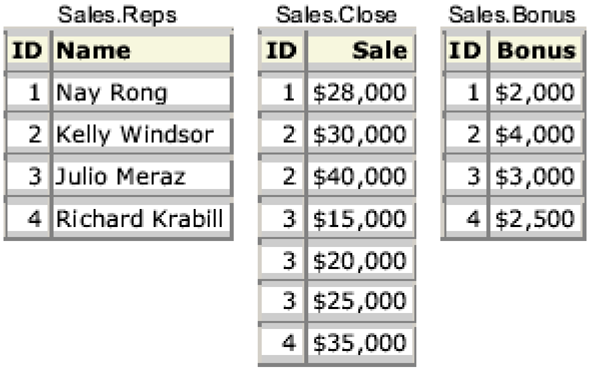
Career	Supervis	Finance
72	26	9
63	76	7
96	31	7
96	98	6
84	94	6
10	11	70
85	22	93
83	63	73
82	75	97
36	77	97

d. none of the above

Correct answer: a

Concatenating appends the observations from one data set to another data set. The new data set contains the total number of records from all input data sets, so b is incorrect. All the variables from all the input data sets appear in the new data set, so c is incorrect.

4. If you concatenate the data sets below in the order shown, what is the value of `Sale` in observation 2 of the new data set?



Sales.Reps

ID	Name
1	Nay Rong
2	Kelly Windsor
3	Julio Meraz
4	Richard Krabill

Sales.Close

ID	Sale
1	$28,000
2	$30,000
2	$40,000
3	$15,000
3	$20,000
3	$25,000
4	$35,000

Sales.Bonus

ID	Bonus
1	$2,000
2	$4,000
3	$3,000
4	$2,500

a. missing

b. $30,000

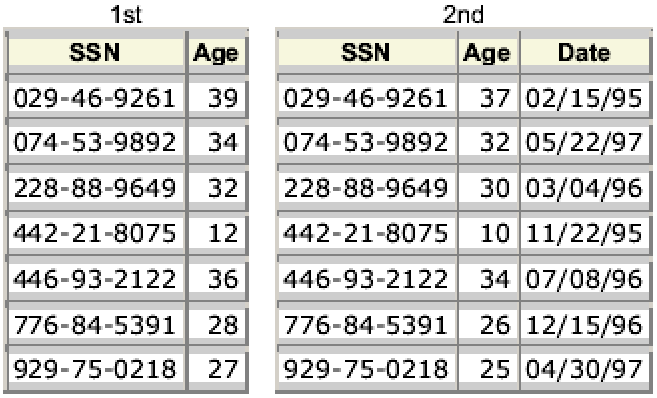
c. $40,000

d. you cannot concatenate these data sets

Correct answer: a

The concatenated data sets are read sequentially, in the order in which they are listed in the SET statement. The second observation in Sales.Reps does not contain a value for `Sale`, so a missing value appears for this variable. (Note that if you merge the data sets, the value of `Sale` for the second observation is $30,000.)

5. What happens if you merge the following data sets by variable `SSN`?



1st

SSN	Age
029-46-9261	39
074-53-9892	34
228-88-9649	32
442-21-8075	12
446-93-2122	36
776-84-5391	28
929-75-0218	27

2nd

SSN	Age	Date
029-46-9261	37	02/15/95
074-53-9892	32	05/22/97
228-88-9649	30	03/04/96
442-21-8075	10	11/22/95
446-93-2122	34	07/08/96
776-84-5391	26	12/15/96
929-75-0218	25	04/30/97

a. The values of `Age` in the 1st data set overwrite the values of `Age` in the 2nd data set.

b. The values of `Age` in the 2nd data set overwrite the values of `Age` in the 1st data set.

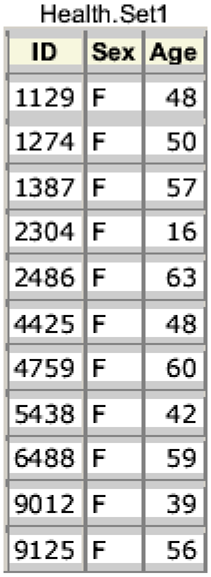
c. The DATA step fails because the two data sets contain same-named variables that have different values.

d. The values of Age in the 2nd data set are set to missing.

Correct answer: b

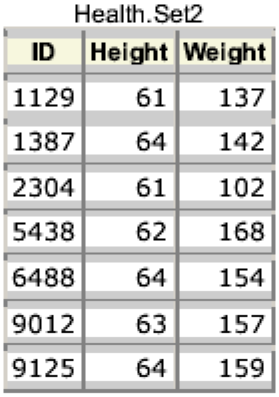
If you have variables with the same name in more than one input data set, then values of the same-named variable in the first data set in which it appears are overwritten by values of the same-named variable in subsequent data sets.

6. Suppose you merge data sets Health.Set1 and Health.Set2 below:



Health.Set1

ID	Sex	Age
1129	F	48
1274	F	50
1387	F	57
2304	F	16
2486	F	63
4425	F	48
4759	F	60
5438	F	42
6488	F	59
9012	F	39
9125	F	56



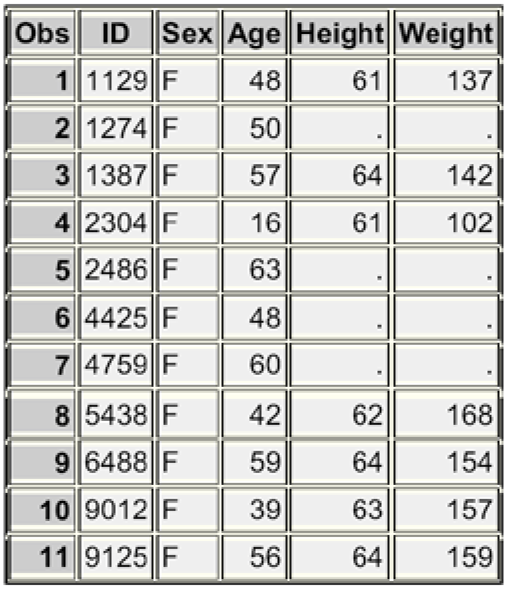
Health.Set2

ID	Height	Weight
1129	61	137
1387	64	142
2304	61	102
5438	62	168
6488	64	154
9012	63	157
9125	64	159

Which output does the following program create?

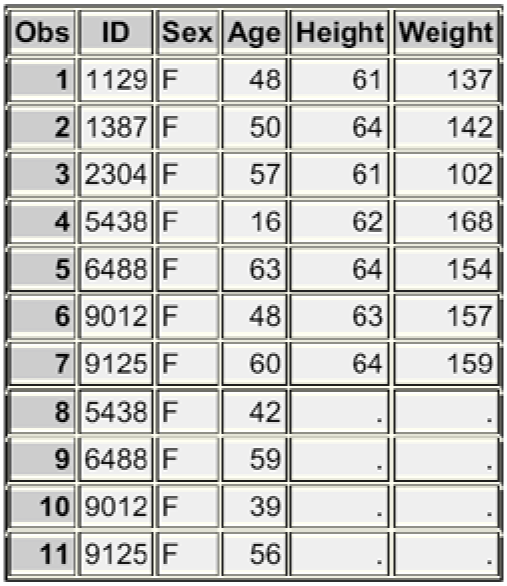
```
data work.merged;
   merge health.set1(in=in1) health.set2(in=in2);
   by id;
   if in1 and in2;
run;
proc print data=work.merged;
run;
```

a.



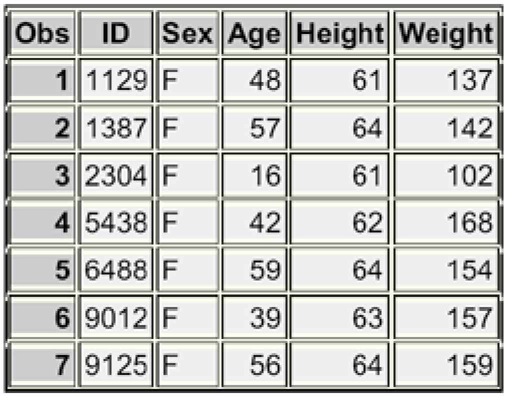
Obs	ID	Sex	Age	Height	Weight
1	1129	F	48	61	137
2	1274	F	50	.	.
3	1387	F	57	64	142
4	2304	F	16	61	102
5	2486	F	63	.	.
6	4425	F	48	.	.
7	4759	F	60	.	.
8	5438	F	42	62	168
9	6488	F	59	64	154
10	9012	F	39	63	157
11	9125	F	56	64	159

b.



Obs	ID	Sex	Age	Height	Weight
1	1129	F	48	61	137
2	1387	F	50	64	142
3	2304	F	57	61	102
4	5438	F	16	62	168
5	6488	F	63	64	154
6	9012	F	48	63	157
7	9125	F	60	64	159
8	5438	F	42	.	.
9	6488	F	59	.	.
10	9012	F	39	.	.
11	9125	F	56	.	.

c.



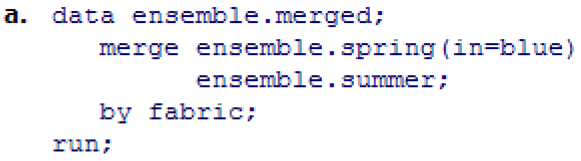
Obs	ID	Sex	Age	Height	Weight
1	1129	F	48	61	137
2	1387	F	57	64	142
3	2304	F	16	61	102
4	5438	F	42	62	168
5	6488	F	59	64	154
6	9012	F	39	63	157
7	9125	F	56	64	159

d. none of the above

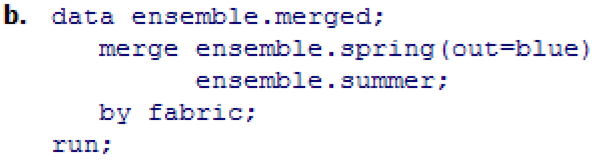
Correct answer: c

The DATA step uses the IN= data set option and the subsetting IF statement to exclude unmatched observations from the output data set. So a and b, which contain unmatched observations, are incorrect.

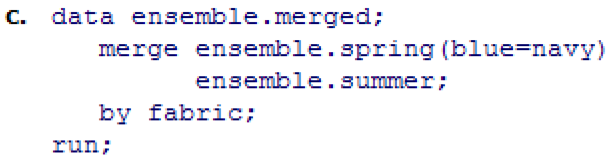
7. The data sets Ensemble.Spring and Ensemble.Summer both contain a variable named `Blue`. How do you prevent the values of the variable `Blue` from being overwritten when you merge the two data sets?



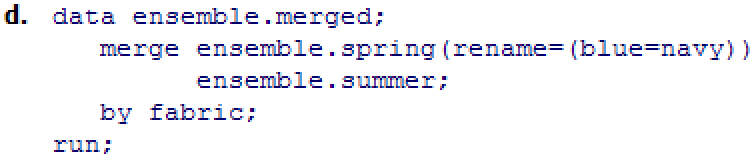
```
a. data ensemble.merged;
      merge ensemble.spring(in=blue)
            ensemble.summer;
      by fabric;
   run;
```



```
b. data ensemble.merged;
      merge ensemble.spring(out=blue)
            ensemble.summer;
      by fabric;
   run;
```



```
c. data ensemble.merged;
      merge ensemble.spring(blue=navy)
            ensemble.summer;
      by fabric;
   run;
```



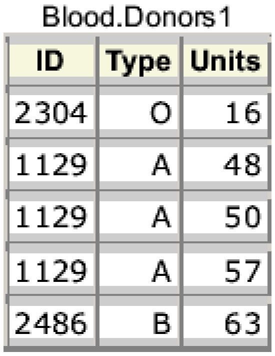
```
d. data ensemble.merged;
      merge ensemble.spring(rename=(blue=navy))
            ensemble.summer;
      by fabric;
   run;
```

Correct answer: d

Match-merging overwrites same-named variables in the first data set with same-named variables in subsequent data sets. To prevent overwriting, rename variables by using the RENAME= data set option in the MERGE statement.

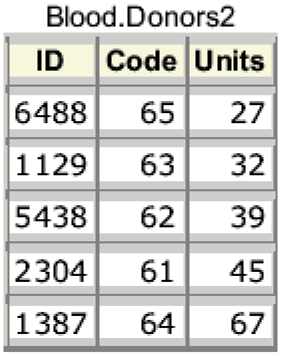
8. What happens if you submit the following program to merge Blood.Donors1 and Blood.Donors2, shown below?

```
data work.merged;
   merge blood.donors1 blood.donors2;
   by id;
run;
```



Blood.Donors1

ID	Type	Units
2304	O	16
1129	A	48
1129	A	50
1129	A	57
2486	B	63



Blood.Donors2

ID	Code	Units
6488	65	27
1129	63	32
5438	62	39
2304	61	45
1387	64	67

a. The Merged data set contains some missing values because not all observations have matching observations in the other data set.

b. The Merged data set contains 8 observations.

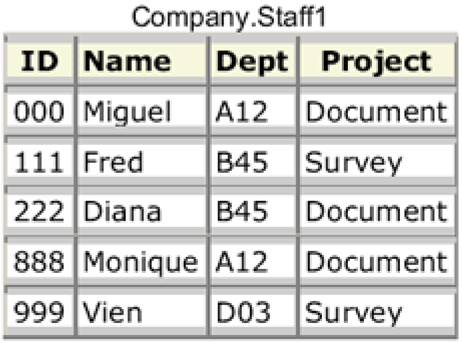
c. The DATA step produces errors.

d. Values for `Units` in Blood.Donors2 overwrite values for `Units` in Blood.Donors1.

Correct answer: c

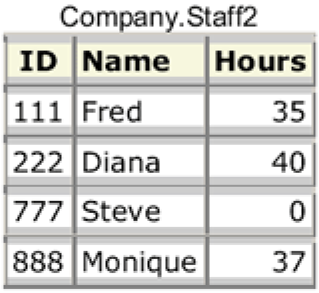
The two input data sets are not sorted by values of the BY variable, so the DATA step produces errors and stops processing.

9. If you merge Company.Staff1 and Company.Staff2 below by `ID`, how many observations does the new data set contain?



Company.Staff1

ID	Name	Dept	Project
000	Miguel	A12	Document
111	Fred	B45	Survey
222	Diana	B45	Document
888	Monique	A12	Document
999	Vien	D03	Survey



Company.Staff2

ID	Name	Hours
111	Fred	35
222	Diana	40
777	Steve	0
888	Monique	37

a. 4

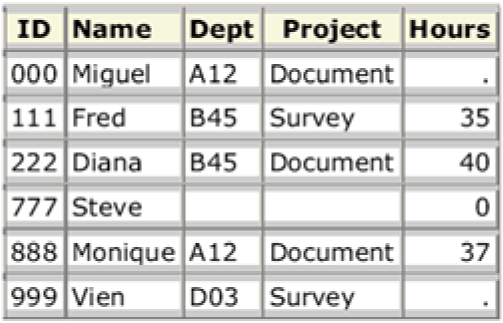
b. 5

c. 6

d. 9

Correct answer: c

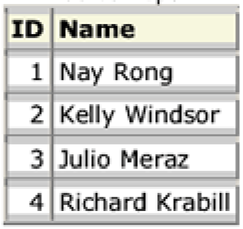
In this example, the new data set contains one observation for each unique value of `ID`. The merged data set is shown below.



ID	Name	Dept	Project	Hours
000	Miguel	A12	Document	.
111	Fred	B45	Survey	35
222	Diana	B45	Document	40
777	Steve			0
888	Monique	A12	Document	37
999	Vien	D03	Survey	.

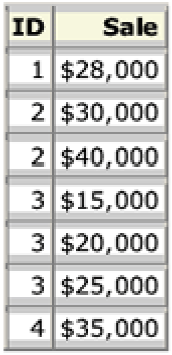
10. If you merge data sets Sales.Reps, Sales.Close, and Sales.Bonus by `ID`, what is the value of `Bonus` in the third observation in the new data set?

Sales.Reps



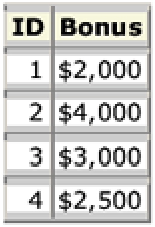
ID	Name
1	Nay Rong
2	Kelly Windsor
3	Julio Meraz
4	Richard Krabill

Sales.Close



ID	Sale
1	$28,000
2	$30,000
2	$40,000
3	$15,000
3	$20,000
3	$25,000
4	$35,000

Sales.Bonus



ID	Bonus
1	$2,000
2	$4,000
3	$3,000
4	$2,500

a. $4,000

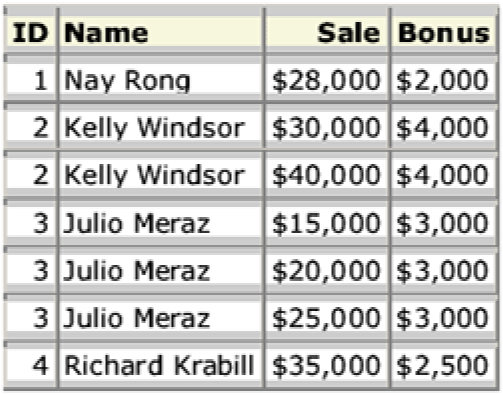
b. $3,000

c. missing

d. can't tell from the information given

Correct answer: a

In the new data set, the third observation is the second observation for ID number 2 (Kelly Windsor). The value for `Bonus` is retained from the previous observation because the BY variable value didn't change. The new data set is shown below.



ID	Name	Sale	Bonus
1	Nay Rong	$28,000	$2,000
2	Kelly Windsor	$30,000	$4,000
2	Kelly Windsor	$40,000	$4,000
3	Julio Meraz	$15,000	$3,000
3	Julio Meraz	$20,000	$3,000
3	Julio Meraz	$25,000	$3,000
4	Richard Krabill	$35,000	$2,500

Chapter 14: Transforming Data with SAS Functions Answer Key

1. Which function calculates the average of the variables `Var1`, `Var2`, `Var3`, and `Var4`?

a. `mean(var1,var4)`

b. `mean(var1-var4)`

c. `mean(of var1,var4)`

d. `mean(of var1-var4)`

Correct answer: d

Use a variable list to specify a range of variables as the function argument. When specifying a variable list, be sure to precede the list with the word OF. If you omit the word OF, the function argument might not be interpreted as expected.

2. Within the data set Hrd.Temp, `PayRate` is a character variable and `Hours` is a numeric variable. What happens when the following program is run?

```
data work.temp;
   set hrd.temp;
   Salary=payrate*hours;
run;
```

a. SAS converts the values of `PayRate` to numeric values. No message is written to the log.

b. SAS converts the values of `PayRate` to numeric values. A message is written to the log.

c. SAS converts the values of `Hours` to character values. No message is written to the log.

d. SAS converts the values of `Hours` to character values. A message is written to the log.

Correct answer: b

When this DATA step is executed, SAS automatically converts the character values of `PayRate` to numeric values so that the calculation can occur. Whenever data is automatically converted, a message is written to the SAS log stating that the conversion has occurred.

3. A typical value for the character variable `Target` is 123,456. Which statement correctly converts the values of `Target` to numeric values when creating the variable `TargetNo`?

a. `TargetNo=input(target,comma6.);`

b. `TargetNo=input(target,comma7.);`

c. `TargetNo=put(target,comma6.);`

d. `TargetNo=put(target,comma7.);`

Correct answer: b

You explicitly convert character values to numeric values by using the INPUT function. Be sure to select an informat that can read the form of the values.

4. A typical value for the numeric variable `SiteNum` is 12.3. Which statement correctly converts the values of `SiteNum` to character values when creating the variable `Location`?

a. `Location=dept||'/'||input(sitenum,3.1);`

b. `Location=dept||'/'||input(sitenum,4.1);`

c. `Location=dept||'/'||put(sitenum,3.1);`

d. `Location=dept||'/'||put(sitenum,4.1);`

Correct answer: d

You explicitly convert numeric values to character values by using the PUT function. Be sure to select a format that can read the form of the values.

5. Suppose the YEARCUTOFF= system option is set to 1920. Which MDY function creates the date value for January 3, 2020?

a. `MDY(1,3,20)`

b. `MDY(3,1,20)`

c. `MDY(1,3,2020)`

d. `MDY(3,1,2020)`

Correct answer: c

Because the YEARCUTOFF= system option is set to 1920, SAS sees the two-digit year value 20 as 1920. Four-digit year values are always read correctly

6. The variable `Address2` contains values such as Piscataway, NJ. How do you assign the two-letter state abbreviations to a new variable named `State`?

a. `State=scan(address2,2);`

b. `State=scan(address2,13,2);`

c. `State=substr(address2,2);`

d. `State=substr(address2,13,2);`

Correct answer: a

The SCAN function is used to extract words from a character value when you know the order of the words, when their position varies, and when the words are marked by some delimiter. In this case, you don't need to specify delimiters, because the blank and the comma are default delimiters.

7. The variable `IDCode` contains values such as 123FA and 321MB. The fourth character identifies sex. How do you assign these character codes to a new variable named `Sex`?

a. `Sex=scan(idcode,4);`

b. `Sex=scan(idcode,4,1);`

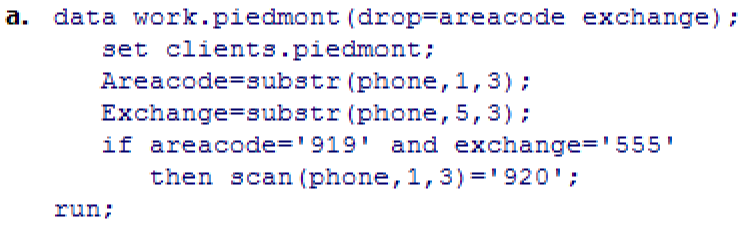
c. `Sex=substr(idcode,4);`

d. `Sex=substr(idcode,4,1);`

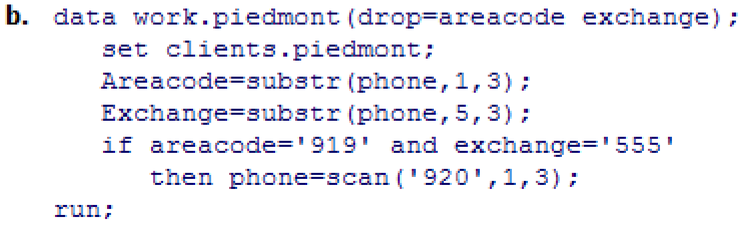
Correct answer: d

The SUBSTR function is best used when you know the exact position of the substring to extract from the character value. You specify the position to start from and the number of characters to extract.

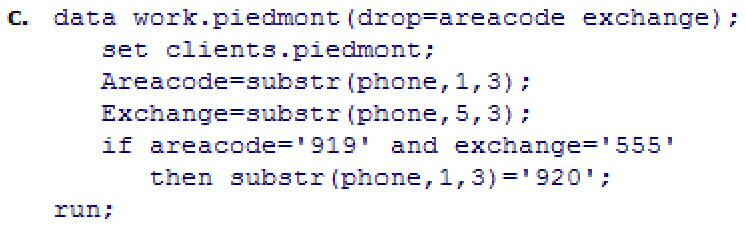
8. Due to growth within the 919 area code, the telephone exchange 555 is being reassigned to the 920 area code. The data set Clients.Piedmont includes the variable `Phone`, which contains telephone numbers in the form 919-555-1234. Which of the following programs will correctly change the values of `Phone`?



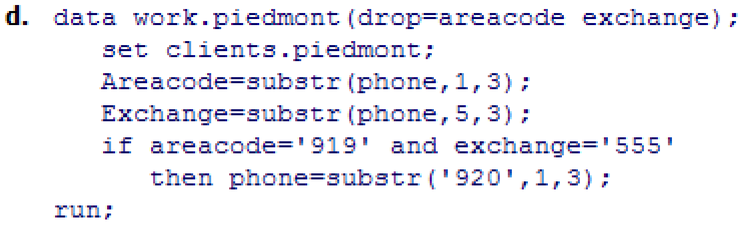
```
a. data work.piedmont(drop=areacode exchange);
      set clients.piedmont;
      Areacode=substr(phone,1,3);
      Exchange=substr(phone,5,3);
      if areacode='919' and exchange='555'
         then scan(phone,1,3)='920';
   run;
```



```
b. data work.piedmont(drop=areacode exchange);
      set clients.piedmont;
      Areacode=substr(phone,1,3);
      Exchange=substr(phone,5,3);
      if areacode='919' and exchange='555'
         then phone=scan('920',1,3);
   run;
```



```
c. data work.piedmont(drop=areacode exchange);
      set clients.piedmont;
      Areacode=substr(phone,1,3);
      Exchange=substr(phone,5,3);
      if areacode='919' and exchange='555'
         then substr(phone,1,3)='920';
   run;
```

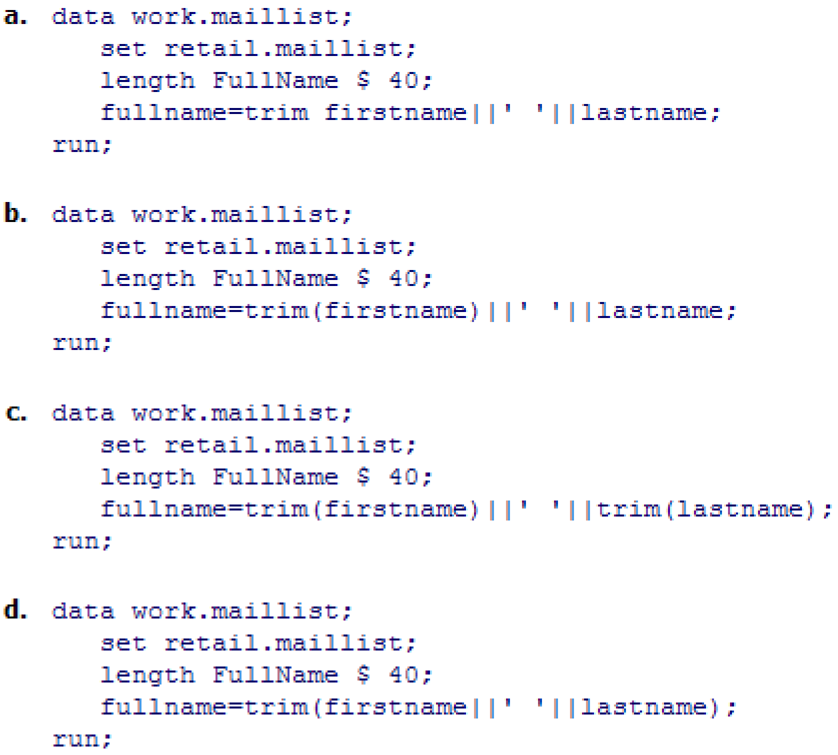


```
d. data work.piedmont(drop=areacode exchange);
      set clients.piedmont;
      Areacode=substr(phone,1,3);
      Exchange=substr(phone,5,3);
      if areacode='919' and exchange='555'
         then phone=substr('920',1,3);
   run;
```

Correct answer: c

The SUBSTR function replaces variable values if it is placed on the left side of an assignment statement. When placed on the right side (as in Question 7), the function extracts a substring.

9. Suppose you need to create the variable `FullName` by concatenating the values of `FirstName`, which contains first names, and `LastName`, which contains last names. What's the best way to remove extra blanks between first names and last names?

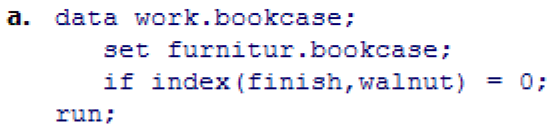


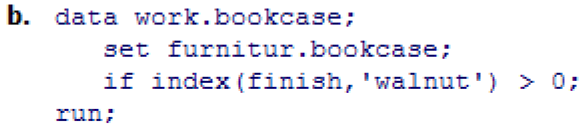
```
a. data work.maillist;
      set retail.maillist;
      length FullName $ 40;
      fullname=trim firstname||' '||lastname;
   run;

b. data work.maillist;
      set retail.maillist;
      length FullName $ 40;
      fullname=trim(firstname)||' '||lastname;
   run;

c. data work.maillist;
      set retail.maillist;
      length FullName $ 40;
      fullname=trim(firstname)||' '||trim(lastname);
   run;

d. data work.maillist;
      set retail.maillist;
      length FullName $ 40;
      fullname=trim(firstname||' '||lastname);
   run;
```

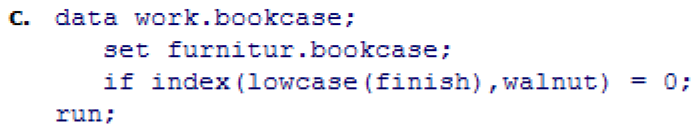
Correct answer: b

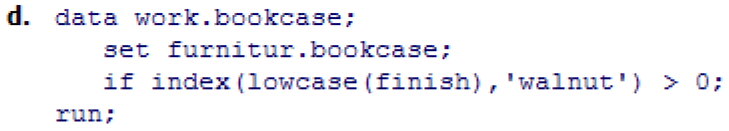
The TRIM function removes trailing blanks from character values. In this case, extra blanks must be removed from the values of `FirstName`. Although answer c also works, the extra TRIM function for the variable `LastName` is unnecessary. Because of the LENGTH statement, all values of `FullName` are padded to 40 characters.

10. Within the data set Furnitur.Bookcase, the variable `Finish` contains values such as ash/cherry/teak/matte-black. Which of the following creates a subset of the data in which the values of `Finish` contain the string walnut? Make the search for the string case-insensitive.


a.
```
data work.bookcase;
   set furnitur.bookcase;
   if index(finish,walnut) = 0;
run;
```


b.
```
data work.bookcase;
   set furnitur.bookcase;
   if index(finish,'walnut') > 0;
run;
```


c.
```
data work.bookcase;
   set furnitur.bookcase;
   if index(lowcase(finish),walnut) = 0;
run;
```


d.
```
data work.bookcase;
   set furnitur.bookcase;
   if index(lowcase(finish),'walnut') > 0;
run;
```

Chapter 15: Generating Data with DO Loops Answer Key

1. Which statement is false regarding the use of DO loops?

a. They can contain conditional clauses.

b. They can generate multiple observations.

c. They can be used to combine DATA and PROC steps.

d. They can be used to read data.

Correct answer: c

DO loops are DATA step statements and cannot be used in conjunction with PROC steps.

2. During each execution of the following DO loop, the value of `Earned` is calculated and is added to its previous value. How many times does this DO loop execute?

```
data finance.earnings;
   Amount=1000;
   Rate=.075/12;
   do month=1 to 12;
      Earned+(amount+earned)*rate;
   end;
run;
```

a. 0

b. 1

c. 12

d. 13

Correct answer: c

The number of iterations is determined by the DO statement's stop value, which in this case is 12.

3. On January 1 of each year, $5,000 is invested in an account. Complete the DATA step below to determine the value of the account after 15 years if a constant interest rate of 10% is expected.

```
data work.invest;
   ...
      Capital+5000;
      capital+(capital*.10);
   end;
run;
```

a. `do count=1 to 15;`

b. `do count=1 to 15 by 10%;`

c. `do count=1 to capital;`

d. `do count=capital to (capital*.10);`

Correct answer: a

Use a DO loop to perform repetitive calculations starting at 1 and looping 15 times.

4. In the data set Work.Invest, what would be the stored value for `Year`?

```
data work.invest;
   do year=1990 to 2004;
      Capital+5000;
      capital+(capital*.10);
   end;
run;
```

a. missing

b. 1990

c. 2004

d. 2005

Correct answer: d

At the end of the fifteenth iteration of the DO loop, the value for `Year` is incremented to 2005. Because this value exceeds the stop value, the DO loop ends. At the bottom of the DATA step, the current values are written to the data set.

5. Which of the following statements is false regarding the program shown below?

```
data work.invest;
   do year=1990 to 2004;
      Capital+5000;
      capital+(capital*.10);
      output;
   end;
run;
```

a. The OUTPUT statement writes current values to the data set immediately.

b. The stored value for Year is 2005.

c. The OUTPUT statement overrides the automatic output at the end of the DATA step.

d. The DO loop performs 15 iterations.

Correct answer: b

The OUTPUT statement overrides the automatic output at the end of the DATA step. On the last iteration of the DO loop, the value of `Year`, 2004, is written to the data set.

6. How many observations will the data set Work.Earn contain?

```
data work.earn;
   Value=2000;
   do year=1 to 20;
      Interest=value*.075;
      value+interest;
      output;
   end;
run;
```

a. 0

b. 1

c. 19

d. 20

Correct answer: d

The number of observations is based on the number of times the OUTPUT statement executes. The new data set has 20 observations, one for each iteration of the DO loop.

7. Which of the following would you use to compare the result of investing $4,000 a year for five years in three different banks that compound interest monthly? Assume a fixed rate for the five-year period.

a. DO WHILE statement

b. nested DO loops

c. DO UNTIL statement

d. a DO group

Correct answer: b

Place the monthly calculation in a DO loop within a DO loop that iterates once for each year. The DO WHILE and DO UNTIL statements are not used here because the number of required iterations is fixed. A non-iterative DO group would not be useful.

8. Which statement is false regarding DO UNTIL statements?

a. The condition is evaluated at the top of the loop, before the enclosed statements are executed.

b. The enclosed statements are always executed at least once.

c. SAS statements in the DO loop are executed until the specified condition is true.

d. The DO loop must have a closing END statement.

Correct answer: a

The DO UNTIL condition is evaluated at the bottom of the loop, so the enclosed statements are always executed at least once.

9. Select the DO WHILE statement that would generate the same result as the program below.

```
data work.invest
capital=100000
   do until(Capital gt 500000);
      Year+1;
      capital+(capital*.10);
   end;
run;
```

a. `do while(Capital ge 500000);`

b. `do while(Capital=500000);`

c. `do while(Capital le 500000);`

d. `do while(Capital>500000);`

Correct answer: c

Because the DO WHILE loop is evaluated at the top of the loop, you specify the condition that must exist in order to execute the enclosed statements.

10. In the following program, complete the statement so that the program stops generating observations when `Distance` reaches 250 miles or when 10 gallons of fuel have been used.

```
data work.go250;
   set perm.cars;
   do gallons=1 to 10 ... ;
      Distance=gallons*mpg;
      output;
   end;
run;
```

a. `while(Distance<250)`

b. `when(Distance>250)`

c. `over(Distance le 250)`

d. `until(Distance=250)`

Correct answer: a

The WHILE expression causes the DO loop to stop executing when the value of Distance becomes equal to or greater than 250.

Chapter 16: Processing Variables with Arrays Answer Key

1. Which statement is false regarding an ARRAY statement?

a. It is an executable statement.

b. It can be used to create variables.

c. It must contain either all numeric or all character elements.

d. It must be used to define an array before the array name can be referenced.

Correct answer: a

An ARRAY statement is not an executable statement; it merely defines an array.

2. What belongs within the braces of this ARRAY statement?

```
array contrib{?} qtr1-qtr4;
```

a. `quarter`

b. `quarter*`

c. `1-4`

d. `4`

Correct d

answer:

The value in braces indicates the number of elements in the array. In this case, there are four elements.

3. For the program below, select an iterative DO statement to process all elements in the `contrib` array.

```
data work.contrib;
   array contrib{4} qtr1-qtr4;
      ...
      contrib{i}=contrib{i}*1.25;
   end;
run;
```

a. `do i=4;`

b. `do i=1 to 4;`

c. `do until i=4;`

d. `do while i le 4;`

Correct answer: b

In the DO statement, you specify the index variable that represents the values of the array elements. Then specify the start and stop positions of the array elements.

4. What is the value of the index variable that references `Jul` in the statements below?

```
array quarter{4} Jan Apr Jul Oct;
do i=1 to 4;
   yeargoal=quarter{i}*1.2;
end;
```

a. 1

b. 2

c. 3

d. 4

Correct answer: c

The index value represents the position of the array element. In this case, the third element is

`Jul`.

5. Which DO statement would not process all the elements in the `factors` array shown below?

```
array factors{*} age height weight bloodpr;
```

a. `do i=1 to dim(factors);`

b. `do i=1 to dim(*);`

c. `do i=1,2,3,4;`

d. `do i=1 to 4;`

Correct answer: b

To process all the elements in an array, you can either specify the array dimension or use the DIM function with the array name as the argument.

6. Which statement below is false regarding the use of arrays to create variables?

a. The variables are added to the program data vector during the compilation of the DATA step.

b. You do not need to specify the array elements in the ARRAY statement.

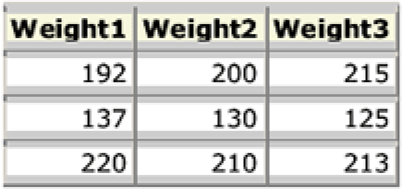
c. By default, all character variables are assigned a length of eight.

d. Only character variables can be created.

Correct answer: d

Either numeric or character variables can be created by an ARRAY statement.

7. For the first observation, what is the value of `diff{i}` at the end of the second iteration of the DO loop?



Weight1	Weight2	Weight3
192	200	215
137	130	125
220	210	213

```
array wt{*} weight1-weight10;
array diff{9};
```

```
do i=1 to 9;
   diff{i}=wt{i+1}-wt{i};
end;
```

a. 15

b. 10

c. 8

d. -7

Correct answer: a

At the end of the second iteration, `diff{i}` resolves as follows:

```
diff{2}=wt{2+1}-wt{2};
diff{2}=215-200
```

8. Finish the ARRAY statement below to create temporary array elements that have initial values of 9000, 9300, 9600, and 9900.

```
array goal{4} ... ;
```

a. `_temporary_ (9000 9300 9600 9900)`

b. `temporary (9000 9300 9600 9900)`

c. `_temporary_ 9000 9300 9600 9900`

d. `(temporary) 9000 9300 9600 9900`

Correct answer: a

To create temporary array elements, specify `_TEMPORARY_` after the array name and dimension. Specify an initial value for each element, separated by either blanks or commas, and enclose the values in parentheses.

9. Based on the ARRAY statement below, select the array reference for the array element `q50`.

```
array ques{3,25} q1-q75;
```

a. `ques{q50}`

b. `ques{1,50}`

c. `ques{2,25}`

d. `ques{3,0}`

Correct answer: c

This two-dimensional array would consist of three rows of 25 elements. The first row would contain `q1` through `q25`, the second row would start with `q26` and end with `q50`, and the third row would start with `q51` and end with `q75`.

10. Select the ARRAY statement that defines the array in the following program.

```
data rainwear.coat;
   input category high1-high3 / low1-low3;
   ...
   do i=1 to 2;
      do j=1 to 3;
         compare{i,j}=round(compare{i,j}*1.12);
      end;
   end;
run;
```

a. `array compare{1,6} high1-high3 low1-low3;`

b. `array compare{2,3} high1-high3 low1-low3;`

c. `array compare{3,2} high1-high3 low1-low3;`

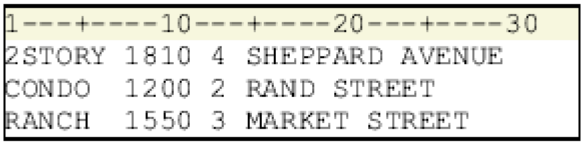
d. `array compare{3,3} high1-high3 low1-low3;`

Correct answer: b

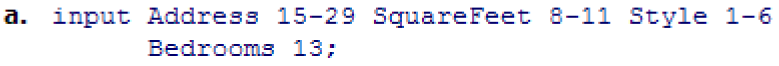
The nested DO loops indicate that the array is named `compare` and is a two-dimensional array that has two rows and three columns.

Chapter 17: Reading Raw Data in Fixed Fields Answer Key

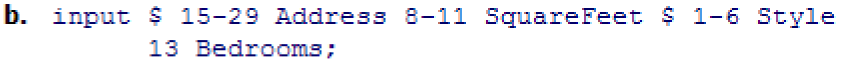
1. Which SAS statement correctly uses column input to read the values in the raw data file below in this order: `Address` (4th field), `SquareFeet` (second field), `Style` (first field), `Bedrooms` (third field)?



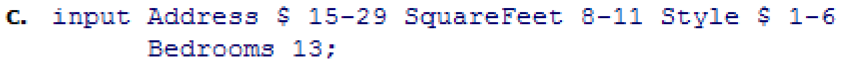
```
1---+----10---+----20---+----30
2STORY 1810 4 SHEPPARD AVENUE
CONDO  1200 2 RAND STREET
RANCH  1550 3 MARKET STREET
```



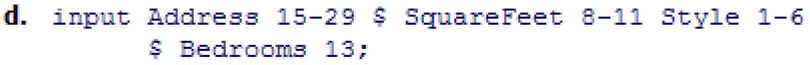
```
a. input Address 15-29 SquareFeet 8-11 Style 1-6
         Bedrooms 13;
```



```
b. input $ 15-29 Address 8-11 SquareFeet $ 1-6 Style
         13 Bedrooms;
```



```
c. input Address $ 15-29 SquareFeet 8-11 Style $ 1-6
         Bedrooms 13;
```



```
d. input Address 15-29 $ SquareFeet 8-11 Style 1-6
         $ Bedrooms 13;
```

Correct answer: c

Column input specifies the variable's name, followed by a dollar ($) sign if the values are character values, and the beginning and ending column locations of the raw data values.

2. Which is not an advantage of column input?

a. It can be used to read character variables that contain embedded blanks.

b. No placeholder is required for missing data.

c. Standard as well as nonstandard data values can be read.

d. Fields do not have to be separated by blanks or other delimiters.

Correct answer: c

Column input is useful for reading standard values only.

3. Which is an example of standard numeric data?

a. -34.245

b. $24,234.25

c. 1/2

d. 50%

Correct answer: a

A standard numeric value can contain numbers, scientific notation, decimal points, and plus and minus signs. Nonstandard numeric data includes values that contain fractions or special characters such as commas, dollar signs, and percent signs.

4. Formatted input can be used to read

a. standard free-format data

b. standard data in fixed fields

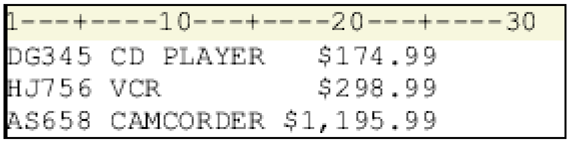
c. nonstandard data in fixed fields

d. both standard and nonstandard data in fixed fields

Correct answer: d

Formatted input can be used to read both standard and nonstandard data in fixed fields.

5. Which informat should you use to read the values in column 1-5?



```
1---+----10---+----20---+----30
DG345 CD PLAYER   $174.99
HJ756 VCR         $298.99
AS658 CAMCORDER $1,195.99
```

a. w.

b. $w.

c. w.d

d. COMMAw.d

Correct answer: b

The $w. informat enables you to read character data. The w represents the field width of the data value or the total number of columns that contain the raw data field.

6. The COMMAw.d informat can be used to read which of the following values?

a. 12,805

b. $177.95

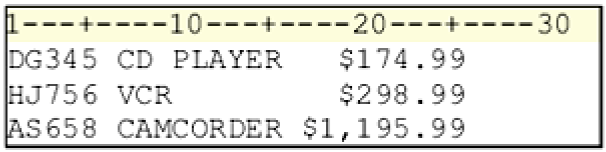
c. 18 %

d. all of the above

Correct answer: d

The COMMAw.d informat strips out special characters such as commas, dollar signs, and percent signs from numeric data, and stores only numeric values in a SAS data set.

7. Which INPUT statement correctly reads the values for `ModelNumber` (first field) after the values for `Item` (second field)? Both `Item` and `ModelNumber` are character variables.



```
1---+----10---+----20---+----30
DG345 CD PLAYER   $174.99
HJ756 VCR         $298.99
AS658 CAMCORDER $1,195.99
```

a. `input +7 Item $9. @1 ModelNumber $5.;`

b. `input +6 Item $9. @1 ModelNumber $5.;`

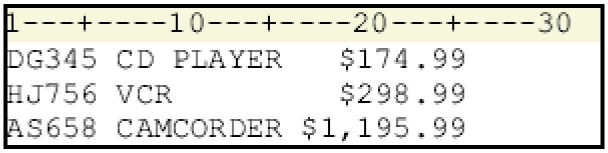
c. `input @7 Item $9. +1 ModelNumber $5.;`

d. `input @7 Item $9 @1 ModelNumber 5.;`

Correct answer: b

The +6 pointer control moves the input pointer to the beginning column of `Item`, and the values are read. Then the @1 pointer control returns to column 1, where the values for `ModelNumber` are located.

8. Which INPUT statement correctly reads the numeric values for `Cost` (third field)?



```
1---+----10---+----20---+----30
DG345 CD PLAYER   $174.99
HJ756 VCR         $298.99
AS658 CAMCORDER $1,195.99
```

a. `input @17 Cost 7.2;`

b. `input @17 Cost 9.2.;`

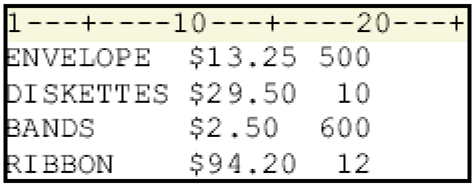
c. `input @17 Cost comma7.;`

d. `input @17 Cost comma9.;`

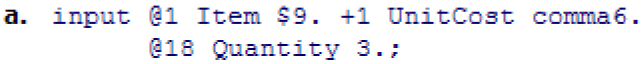
Correct answer: d

The values for `Cost` contain dollar signs and commas, so you must use the COMMAw.d informat. Counting the numbers, dollar sign, comma, and decimal point, the field width is 9 columns. Because the data value contains decimal places, a d value is not needed.

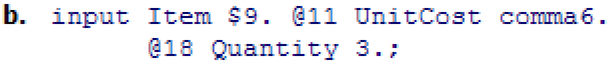
9. Which SAS statement correctly uses formatted input to read the values in this order: `Item` (first field), `UnitCost` (second field), `Quantity` (third field)?



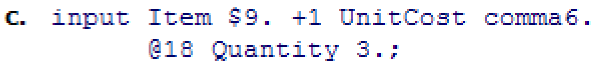
```
1---+----10---+----20---+
ENVELOPE  $13.25 500
DISKETTES $29.50  10
BANDS     $2.50  600
RIBBON    $94.20  12
```



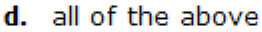
```
a. input @1 Item $9. +1 UnitCost comma6.
         @18 Quantity 3.;
```



```
b. input Item $9. @11 UnitCost comma6.
         @18 Quantity 3.;
```



```
c. input Item $9. +1 UnitCost comma6.
         @18 Quantity 3.;
```



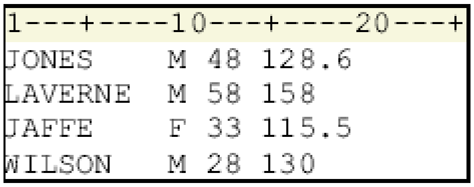
d. all of the above

Correct answer: d

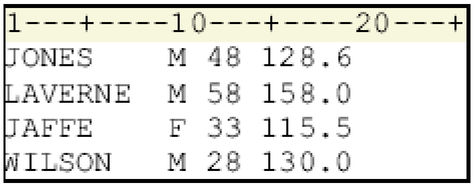
The default location of the column pointer control is column 1, so a column pointer control is optional for reading the first field. You can use the @n or +n pointer controls to specify the beginning column of the other fields. You can use the $w. informat to read the values for `Item`, the COMMAw.d informat for `UnitCost`, and the w.d informat for `Quantity`.

10. Which raw data file requires the PAD option in the INFILE statement in order to correctly read the data using either column input or formatted input?

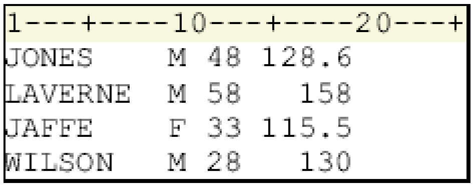
a.



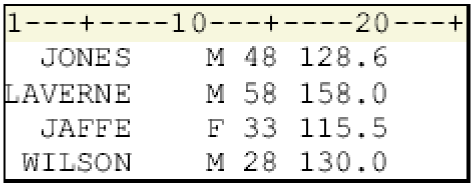
```
1---+----10---+----20---+
JONES     M 48 128.6
LAVERNE   M 58 158
JAFFE     F 33 115.5
WILSON    M 28 130
```

b. 

```
1---+----10---+----20---+
JONES   M 48 128.6
LAVERNE M 58 158.0
JAFFE   F 33 115.5
WILSON  M 28 130.0
```

c. 

```
1---+----10---+----20---+
JONES   M 48 128.6
LAVERNE M 58   158
JAFFE   F 33 115.5
WILSON  M 28   130
```

d. 

```
1---+----10---+----20---+
  JONES   M 48 128.6
LAVERNE   M 58 158.0
  JAFFE   F 33 115.5
 WILSON   M 28 130.0
```

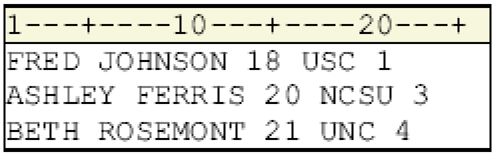
Correct answer: a

Use the PAD option in the INFILE statement to read variable-length records that contain fixed-field data. The PAD option pads each record with blanks so that all data lines have the same length.

Chapter 18: Reading Free-Format Data Answer Key

1. The raw data file referenced by the fileref Students contains data that is

Raw Data File Students



```
1---+----10---+----20---+
FRED JOHNSON 18 USC 1
ASHLEY FERRIS 20 NCSU 3
BETH ROSEMONT 21 UNC 4
```

a. arranged in fixed fields

b. free format

c. mixed format

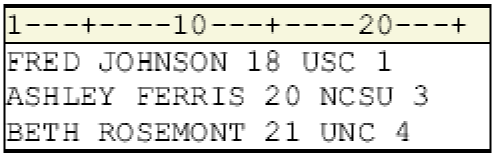
d. arranged in columns

Correct answer: b

The raw data file contains data that is free format, meaning that the data is not arranged in columns or fixed fields.

2. Which input style should be used to read the values in the raw data file that is referenced by the fileref Students?

Raw Data File Students



```
1---+----10---+----20---+
FRED JOHNSON 18 USC 1
ASHLEY FERRIS 20 NCSU 3
BETH ROSEMONT 21 UNC 4
```

a. column

b. formatted

c. list

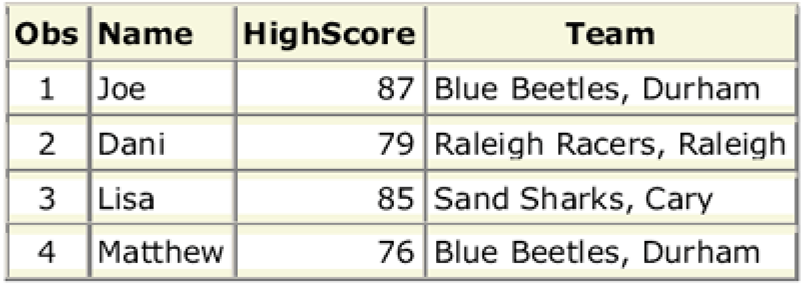
d. mixed

Correct answer: c

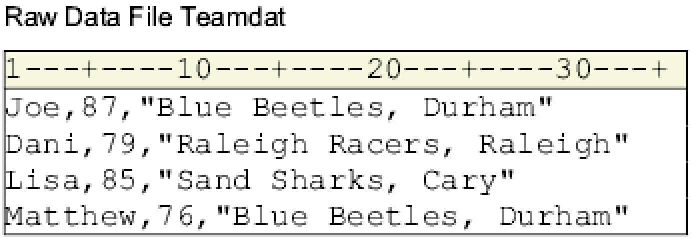
List input should be used to read data that is free format because you do not need to specify the column locations of the data.

3. Which SAS program was used to create the raw data file Teamdat from the SAS data set Work.Scores?

SAS Data Set Work.Scores

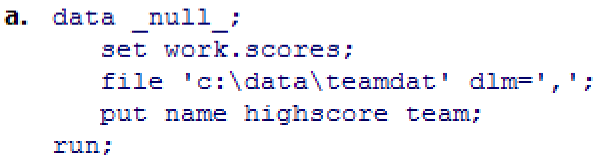


Obs	Name	HighScore	Team
1	Joe	87	Blue Beetles, Durham
2	Dani	79	Raleigh Racers, Raleigh
3	Lisa	85	Sand Sharks, Cary
4	Matthew	76	Blue Beetles, Durham

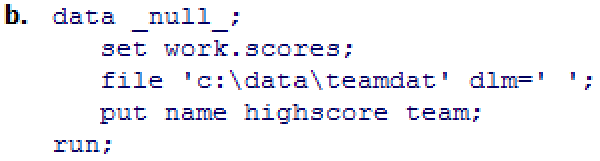


Raw Data File Teamdat

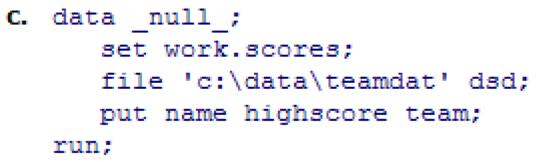
```
1---+----10---+----20---+----30---+
Joe,87,"Blue Beetles, Durham"
Dani,79,"Raleigh Racers, Raleigh"
Lisa,85,"Sand Sharks, Cary"
Matthew,76,"Blue Beetles, Durham"
```



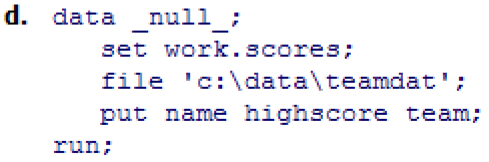
```
a. data _null_;
      set work.scores;
      file 'c:\data\teamdat' dlm=',';
      put name highscore team;
   run;
```



```
b. data _null_;
      set work.scores;
      file 'c:\data\teamdat' dlm=' ';
      put name highscore team;
   run;
```



```
c. data _null_;
      set work.scores;
      file 'c:\data\teamdat' dsd;
      put name highscore team;
   run;
```



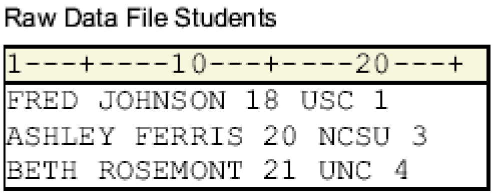
```
d. data _null_;
      set work.scores;
      file 'c:\data\teamdat';
      put name highscore team;
   run;
```

Correct answer: c

You can use the DSD option in the FILE statement to specify that data values containing commas should be enclosed in quotation marks. The DSD option uses a comma as the delimiter by default.

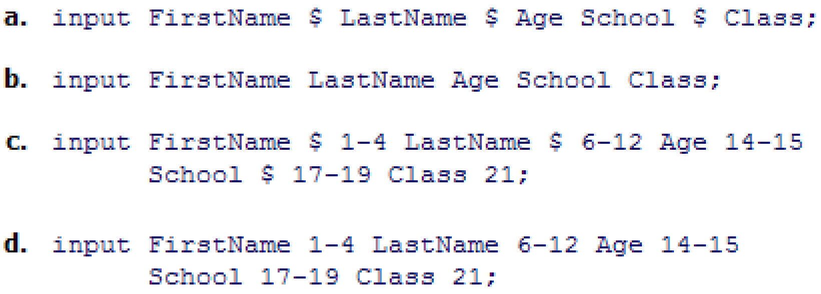
4. Which SAS statement reads the raw data values in order and assigns them to the variables shown below?

Variables: `FirstName` (character), `LastName` (character), `Age` (numeric), `School` (character), `Class` (numeric)



Raw Data File Students

```
1---+----10---+----20---+
FRED JOHNSON 18 USC 1
ASHLEY FERRIS 20 NCSU 3
BETH ROSEMONT 21 UNC 4
```

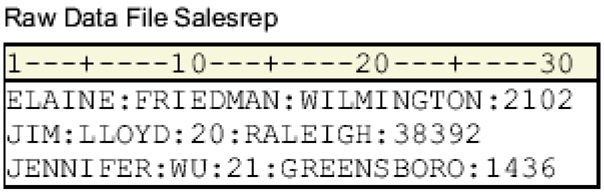


```
a. input FirstName $ LastName $ Age School $ Class;

b. input FirstName LastName Age School Class;

c. input FirstName $ 1-4 LastName $ 6-12 Age 14-15
         School $ 17-19 Class 21;

d. input FirstName 1-4 LastName 6-12 Age 14-15
         School 17-19 Class 21;
```

Correct answer: a

Because the data is free format, list input is used to read the values. With list input, you simply name each variable and identify its type.

5. Which SAS statement should be used to read the raw data file that is referenced by the fileref Salesrep?



Raw Data File Salesrep

```
1---+----10---+----20---+----30
ELAINE:FRIEDMAN:WILMINGTON:2102
JIM:LLOYD:20:RALEIGH:38392
JENNIFER:WU:21:GREENSBORO:1436
```

a. `infile salesrep;`

b. `infile salesrep ':';`

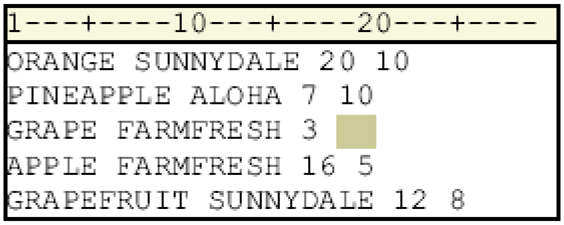
c. `infile salesrep dlm;`

d. `infile salesrep dlm=':';`

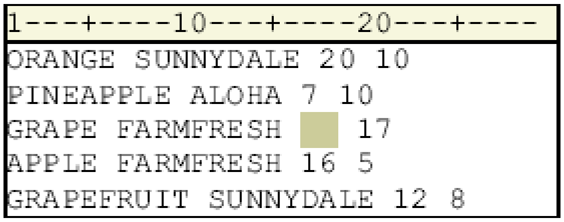
Correct answer: d

The INFILE statement identifies the location of the external data file. The DLM= option specifies the colon (:) as the delimiter that separates each field.

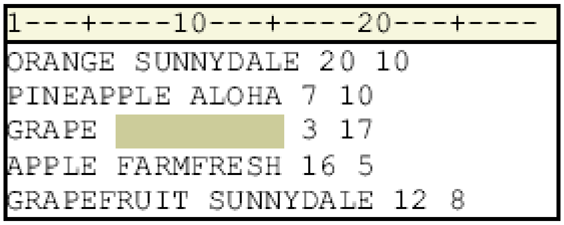
6. Which of the following raw data files can be read by using the MISSOVER option in the INFILE statement? Missing values are indicated with colored blocks.

a. 

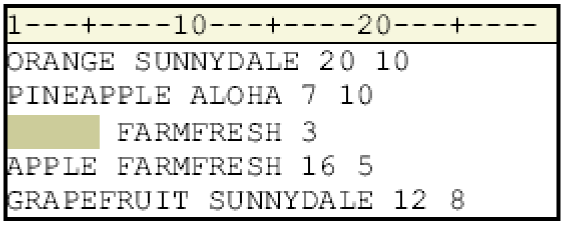
```
1---+----10---+----20---+----
ORANGE SUNNYDALE 20 10
PINEAPPLE ALOHA 7 10
GRAPE FARMFRESH 3
APPLE FARMFRESH 16 5
GRAPEFRUIT SUNNYDALE 12 8
```

b. 

```
1---+----10---+----20---+----
ORANGE SUNNYDALE 20 10
PINEAPPLE ALOHA 7 10
GRAPE FARMFRESH    17
APPLE FARMFRESH 16 5
GRAPEFRUIT SUNNYDALE 12 8
```

c. 

```
1---+----10---+----20---+----
ORANGE SUNNYDALE 20 10
PINEAPPLE ALOHA 7 10
GRAPE           3 17
APPLE FARMFRESH 16 5
GRAPEFRUIT SUNNYDALE 12 8
```

d. 

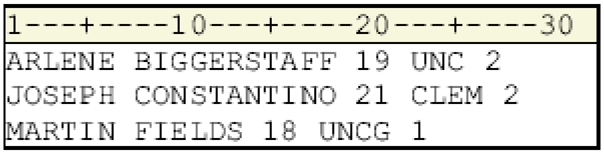
```
1---+----10---+----20---+----
ORANGE SUNNYDALE 20 10
PINEAPPLE ALOHA 7 10
      FARMFRESH 3
APPLE FARMFRESH 16 5
GRAPEFRUIT SUNNYDALE 12 8
```

Correct answer: a

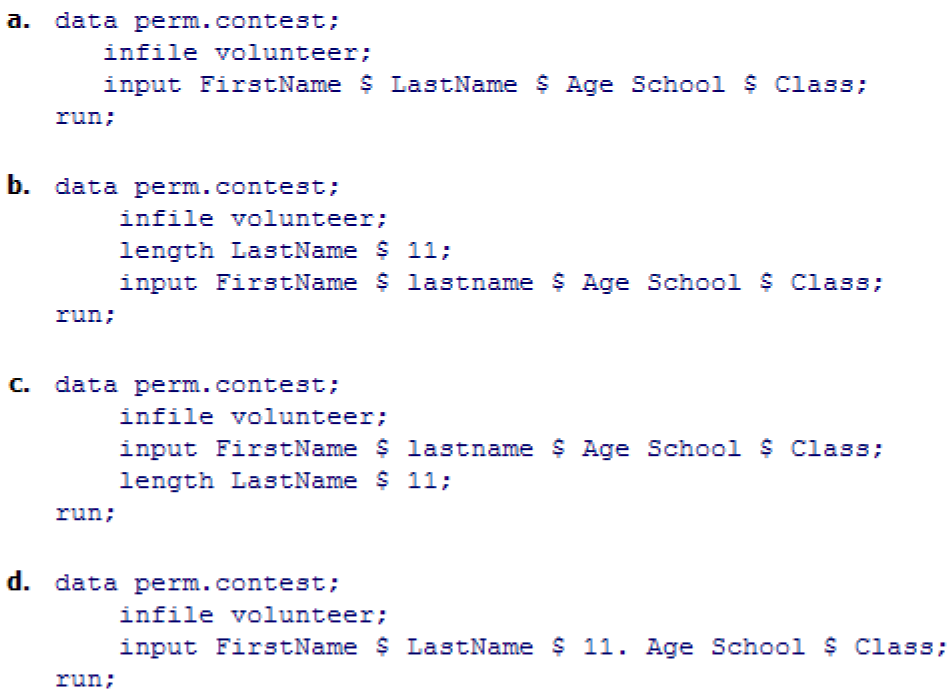
You can use the MISSOVER option in the INFILE statement to read the missing values at the end of a record. The MISSOVER option prevents SAS from moving to the next record if values are missing in the current record.

7. Which SAS program correctly reads the data in the raw data file that is referenced by the fileref Volunteer?

Raw Data File Volunteer



```
1---+----10---+----20---+----30
ARLENE BIGGERSTAFF 19 UNC 2
JOSEPH CONSTANTINO 21 CLEM 2
MARTIN FIELDS 18 UNCG 1
```



a.
```
data perm.contest;
   infile volunteer;
   input FirstName $ LastName $ Age School $ Class;
run;
```

b.
```
data perm.contest;
    infile volunteer;
    length LastName $ 11;
    input FirstName $ lastname $ Age School $ Class;
run;
```

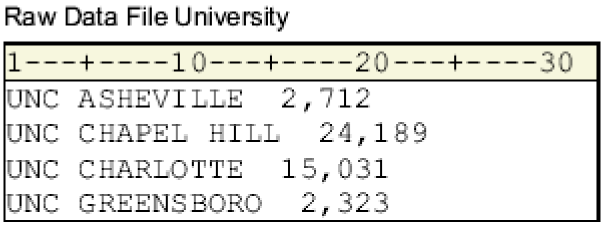
c.
```
data perm.contest;
    infile volunteer;
    input FirstName $ lastname $ Age School $ Class;
    length LastName $ 11;
run;
```

d.
```
data perm.contest;
    infile volunteer;
    input FirstName $ LastName $ 11. Age School $ Class;
run;
```

Correct answer: b

The LENGTH statement extends the length of the character variable `LastName` so that it is large enough to accommodate the data. Variable attributes such as length are defined the first time a variable is named in a DATA step. The LENGTH statement should precede the INPUT statement so that the correct length is defined.

8. Which type of input should be used to read the values in the raw data file that is referenced by the fileref University?



Raw Data File University

```
1---+----10---+----20---+----30
UNC ASHEVILLE  2,712
UNC CHAPEL HILL  24,189
UNC CHARLOTTE  15,031
UNC GREENSBORO  2,323
```

a. column

b. formatted

c. list

d. modified list

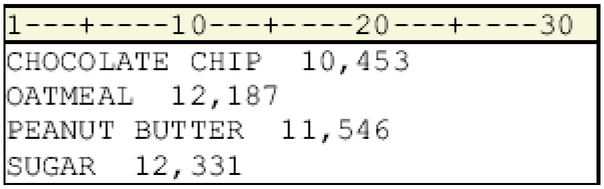
Correct d

answer:

Notice that the values for `School` contain embedded blanks, and the values for `Enrolled` are nonstandard numeric values. Modified list input can be used to read the values that contain embedded blanks and nonstandard values.

9. Which SAS statement correctly reads the values for `Flavor` and `Quantity`? Make sure the length of each variable can accommodate the values that are shown.

Raw Data File Cookies



```
1---+----10---+----20---+----30
CHOCOLATE CHIP  10,453
OATMEAL  12,187
PEANUT BUTTER  11,546
SUGAR  12,331
```

a. `input Flavor & $9. Quantity : comma.;`

b. `input Flavor & $14. Quantity : comma.;`

c. `input Flavor : $14. Quantity & comma.;`

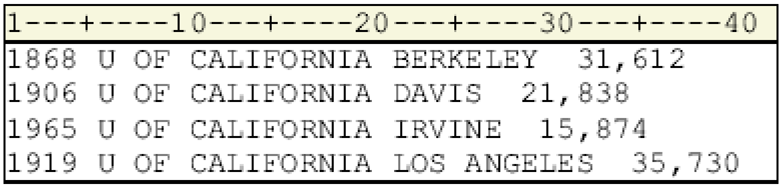
d. `input Flavor $14. Quantity : comma.;`

Correct answer: b

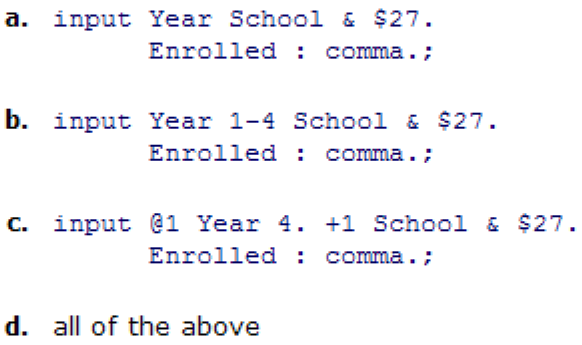
The INPUT statement uses list input with format modifiers and informats to read the values for each variable. The ampersand (&) modifier enables you to read character values that contain single embedded blanks. The colon (:) modifier enables you to read nonstandard data values and character values that are longer than eight characters, and that contain no embedded blanks.

10. Which SAS statement correctly reads the raw data values in order and assigns them to these corresponding variables: `Year` (numeric), `School` (character), `Enrolled` (numeric)?

Raw Data File Founding



```
1---+----10---+----20---+----30---+----40
1868 U OF CALIFORNIA BERKELEY  31,612
1906 U OF CALIFORNIA DAVIS  21,838
1965 U OF CALIFORNIA IRVINE  15,874
1919 U OF CALIFORNIA LOS ANGELES  35,730
```



a.
```
input Year School & $27.
      Enrolled : comma.;
```

b.
```
input Year 1-4 School & $27.
      Enrolled : comma.;
```

c.
```
input @1 Year 4. +1 School & $27.
      Enrolled : comma.;
```

d. all of the above

Correct answer: d

The values for `Year` can be read with column, formatted, or list input. However, the values for `School` and `Enrolled` are free-format data that contain embedded blanks or nonstandard values. Therefore, these last two variables must be read with modified list input.

Chapter 19: Reading Date and Time Values Answer Key

1. SAS date values are the number of days since which date?

a. January 1, 1900

b. January 1, 1950

c. January 1, 1960

d. January 1, 1970

Correct answer: c

A SAS date value is the number of days from January 1, 1960, to the given date.

2. A great advantage of storing dates and times as SAS numeric date and time values is that

a. they can easily be edited.

b. they can easily be read and understood.

c. they can be used in text strings like other character values.

d. they can be used in calculations like other numeric values.

Correct answer: d

In addition to tracking time intervals, SAS date and time values can be used in calculations like other numeric values. This lets you calculate values that involve dates much more easily than in other programming languages.

3. SAS does not automatically make adjustments for daylight saving time, but it does make adjustments for:

a. leap seconds

b. leap years

c. Julian dates

d. time zones

Correct answer: b

SAS automatically makes adjustments for leap years.

4. An input data file has date expressions in the form 10222001. Which SAS informat should you use to read these dates?

a. DATE6.

b. DATE8.

c. MMDDYY6.

d. MMDDYY8.

Correct answer: d

The SAS informat MMDDYYw. reads dates such as 10222001, 10/22/01, or 10-22-01. In this case, the field width is eight.

5. The minimum width of the TIMEw. informat is

a. 4

b. 5

c. 6

d. 7

Correct answer: b

The minimum acceptable field width for the TIMEw. informat is five. If you specify a w value less than five, you will receive an error message in the SAS log.

6. Shown below are date and time expressions and corresponding SAS datetime informats. Which date and time expresssion cannot be read by the informat that is shown beside it?

a. 30May2000:10:03:17.2 `DATETIME20.`

b. 30May00 10:03:17.2 `DATETIME18.`

c. 30May2000/10:03 `DATETIME15.`

d. 30May2000/1003 `DATETIME14.`

Correct answer: d

In the time value of a date and time expression, you must use delimiters to separate the values for hour, minutes, and seconds.

7. What is the default value of the YEARCUTOFF= system option?

a. 1920

b. 1910

c. 1900

d. 1930

Correct answer: a

The default value of YEARCUTOFF= is 1920. This enables you to read two-digit years from 00-19 as the years 2000 through 2019.

8. Suppose your input data file contains the date expression 13APR2009. The YEARCUTOFF= system option is set to 1910. SAS will read the date as

a. 13APR1909

b. 13APR1920

c. 13APR2009

d. 13APR2020

Correct answer: c

The value of the YEARCUTOFF= system option does not affect four-digit year values. Four-digit values are always read correctly.

9. Suppose the YEARCUTOFF= system option is set to 1920. An input file contains the date expression 12/08/1925, which is being read with the MMDDYY8. informat. Which date will appear in your data?

a. 08DEC1920

b. 08DEC1925

c. 08DEC2019

d. 08DEC2025

Correct answer: c

The w value of the informat MMDDYY8. is too small to read the entire value, so the last two digits of the year are truncated. The last two digits thus become 19 instead of 25. Because the YEARCUTOFF= system option is set to 1920, SAS interprets this year as 2019. To avoid such errors, be sure to specify an informat that is wide enough for your date expressions.

10. Suppose your program creates two variables from an input file. Both variables are stored as SAS date values: FirstDay records the start of a billing cycle, and LastDay records the end of that cycle. The code for calculating the total number of days in the cycle would be

a. `TotDays=lastday-firstday;`

b. `TotDays=lastday-firstday+1;`

c. `TotDays=lastday/firstday;`

d. You cannot use date values in calculations.

Correct answer: b

To find the number of days spanned by two dates, subtract the first day from the last day and add one. Because SAS date values are numeric values, they can easily be used in calculations.

Chapter 20: Creating a Single Observation from Multiple Records Answer Key

1. You can position the input pointer on a specific record by using

a. column pointer controls.

b. column specifications.

c. line pointer controls.

d. line hold specifiers.

Correct answer: c

Information for one observation can be spread out over several records. You can write one INPUT statement that contains line pointer controls to specify the records from which values are read.

2. Which pointer control is used to read multiple records sequentially?

a. @n

b. +n

c. /

d. all of the above

Correct answer: c

The forward slash (/) line pointer control is used to read multiple records sequentially. Each time a / pointer is encountered, the input pointer advances to the next line. @n and +n are column pointer controls.

3. Which pointer control can be used to read records non-sequentially?

a. @n

b. #n

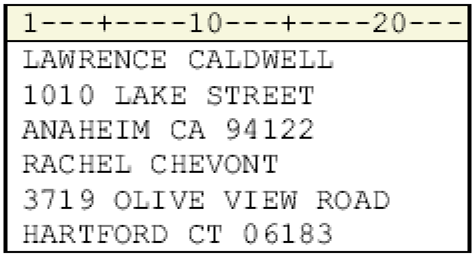
c. +n

d. /

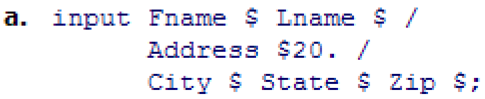
Correct answer: b

The #n line pointer control is used to read records non-sequentially. The #n specifies the absolute number of the line to which you want to move the pointer.

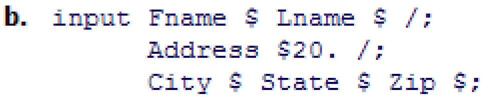
4. Which SAS statement correctly reads the values for `Fname`, `Lname`, `Address`, `City`, `State`, and `Zip` in order?



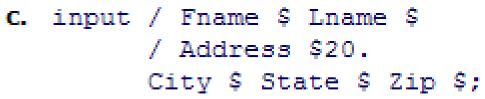
```
1---+----10---+----20---
LAWRENCE CALDWELL
1010 LAKE STREET
ANAHEIM CA 94122
RACHEL CHEVONT
3719 OLIVE VIEW ROAD
HARTFORD CT 06183
```



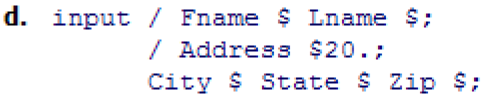
```
a. input Fname $ Lname $ /
         Address $20. /
         City $ State $ Zip $;
```



```
b. input Fname $ Lname $ /;
         Address $20. /;
         City $ State $ Zip $;
```



```
c. input / Fname $ Lname $
         / Address $20.
         City $ State $ Zip $;
```

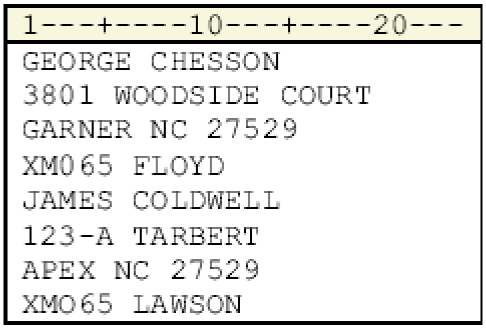


```
d. input / Fname $ Lname $;
         / Address $20.;
         City $ State $ Zip $;
```

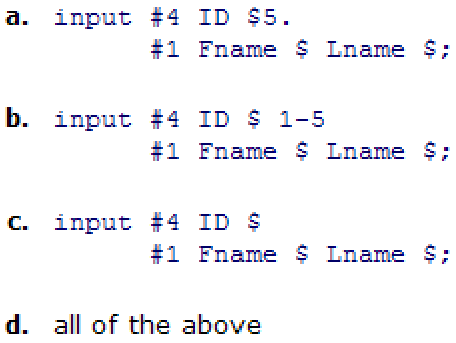
Correct answer: a

The INPUT statement uses the / line pointer control to move the input pointer forward from the first record to the second record, and from the second record to the third record. The / line pointer control only moves the input pointer forward and must be specified after the instructions for reading the values in the current record. You should place a semicolon only at the end of a complete INPUT statement.

5. Which INPUT statement correctly reads the values for `ID` in the fourth record, then returns to the first record to read the values for `Fname` and `Lname`?



```
1---+----10---+----20---
GEORGE CHESSON
3801 WOODSIDE COURT
GARNER NC 27529
XM065 FLOYD
JAMES COLDWELL
123-A TARBERT
APEX NC 27529
XMO65 LAWSON
```



a.
```
input #4 ID $5.
      #1 Fname $ Lname $;
```

b.
```
input #4 ID $ 1-5
      #1 Fname $ Lname $;
```

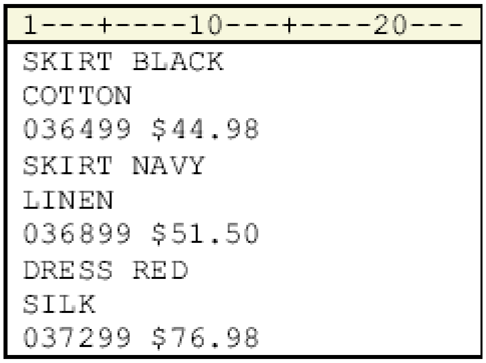
c.
```
input #4 ID $
      #1 Fname $ Lname $;
```

d. all of the above

Correct answer: d

The first #n line pointer control enables you to read the values for `ID` from the fourth record. The second #n line pointer control moves back to the first record and reads the values for `Fname` and `Lname`. You can use formatted input, column input, or list input to read the values for `ID`.

6. How many records will be read for each iteration of the DATA step?



```
1---+----10---+----20---
SKIRT BLACK
COTTON
036499 $44.98
SKIRT NAVY
LINEN
036899 $51.50
DRESS RED
SILK
037299 $76.98
```

```
data spring.sportswr;
   infile newitems;
   input #1 Item $ Color $
          #3 @8 Price comma6.
          #2 Fabric $
          #3 SKU $ 1-6;
run;
```

a. one

b. two

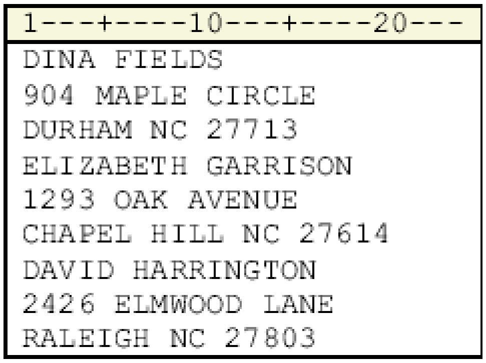
c. three

d. four

Correct answer: c

The first time the DATA step executes, the first three records are read, and an observation is written to the data set. During the second iteration, the next three records are read, and the second observation is written to the data set. During the third iteration, the last three records are read, and the final observation is written to the data set.

7. Which INPUT statement correctly reads the values for `City`, `State`, and `Zip`?



```
1---+----10---+----20---
DINA FIELDS
904 MAPLE CIRCLE
DURHAM NC 27713
ELIZABETH GARRISON
1293 OAK AVENUE
CHAPEL HILL NC 27614
DAVID HARRINGTON
2426 ELMWOOD LANE
RALEIGH NC 27803
```

a. `input #3 City $ State $ Zip $;`

b. `input #3 City & $11. State $ Zip $;`

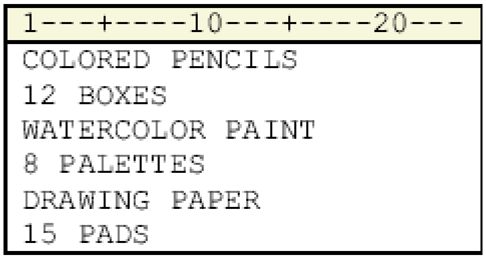
c. `input #3 City $11. +2 State $2. + 2 Zip $5.;`

d. all of the above

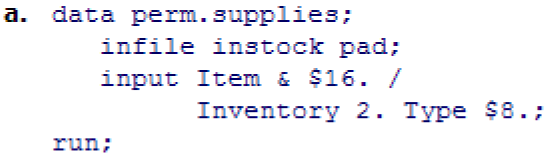
Correct answer: b

A combination of modified and simple list input can be used to read the values for `City`, `State`, and `Zip`. You need to use modified list input to read the values for `City`, because one of the values is longer than eight characters and contains an embedded blank. You cannot use formatted input, because the values do not begin and end in the same column in each record.

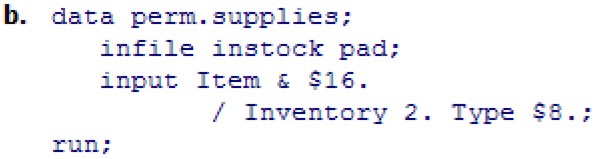
8. Which program does not read the values in the first record as a variable named `Item` and the values in the second record as two variables named `Inventory` and `Type`?



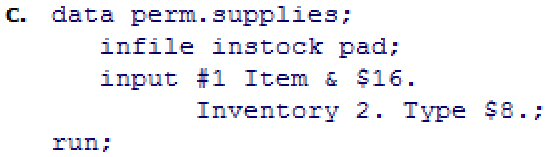
```
1---+----10---+----20---
COLORED PENCILS
12 BOXES
WATERCOLOR PAINT
8 PALETTES
DRAWING PAPER
15 PADS
```



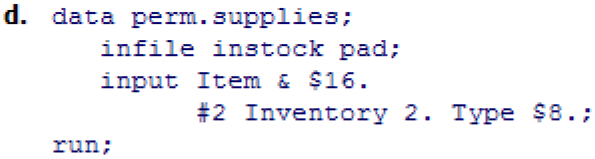
a.
```
data perm.supplies;
   infile instock pad;
   input Item & $16. /
         Inventory 2. Type $8.;
run;
```



b.
```
data perm.supplies;
   infile instock pad;
   input Item & $16.
          / Inventory 2. Type $8.;
run;
```



c.
```
data perm.supplies;
   infile instock pad;
   input #1 Item & $16.
         Inventory 2. Type $8.;
run;
```

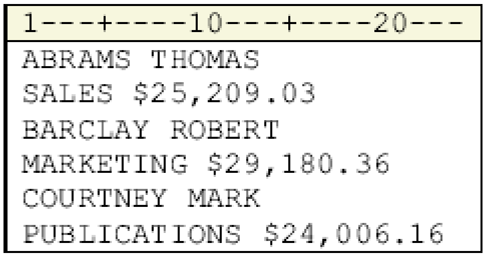


d.
```
data perm.supplies;
   infile instock pad;
   input Item & $16.
         #2 Inventory 2. Type $8.;
run;
```

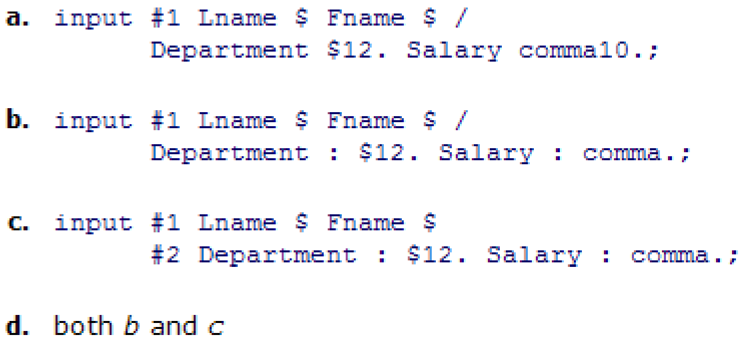
Correct answer: c

The values for `Item` in the first record are read, then the following / or #n line pointer control advances the input pointer to the second record to read the values for `Inventory` and `Type`.

9. Which INPUT statement reads the values for `Lname`, `Fname`, `Department`, and `Salary` (in that order)?



```
1---+----10---+----20---
ABRAMS THOMAS
SALES $25,209.03
BARCLAY ROBERT
MARKETING $29,180.36
COURTNEY MARK
PUBLICATIONS $24,006.16
```



a.
```
input #1 Lname $ Fname $ /
      Department $12. Salary comma10.;
```

b.
```
input #1 Lname $ Fname $ /
      Department : $12. Salary : comma.;
```

c.
```
input #1 Lname $ Fname $
      #2 Department : $12. Salary : comma.;
```

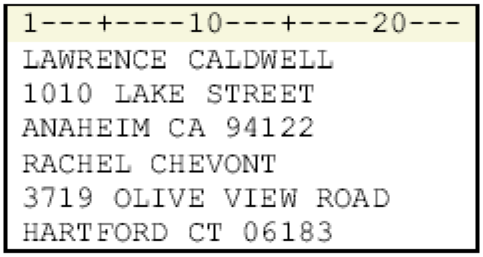
d. both *b* and *c*

Correct answer: d

You can use either the / or #n line pointer control to advance the input pointer to the second line, in order to read the values for `Department` and `Salary`. The colon (:) modifier is used to read the character values that are longer than eight characters `(Department)` and the nonstandard data values (`Salary`).

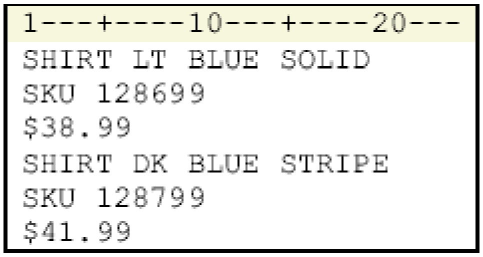
10. Which raw data file poses potential problems when you are reading multiple records for each observation?

a.

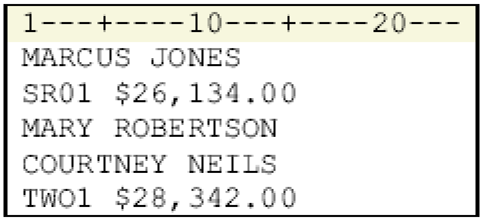


```
1---+----10---+----20---
LAWRENCE CALDWELL
1010 LAKE STREET
ANAHEIM CA 94122
RACHEL CHEVONT
3719 OLIVE VIEW ROAD
HARTFORD CT 06183
```

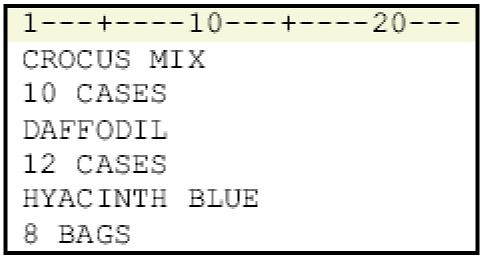
b.



```
1---+----10---+----20---
SHIRT LT BLUE SOLID
SKU 128699
$38.99
SHIRT DK BLUE STRIPE
SKU 128799
$41.99
```

c. 

```
1---+----10---+----20---
MARCUS JONES
SR01 $26,134.00
MARY ROBERTSON
COURTNEY NEILS
TWO1 $28,342.00
```

d. 

```
1---+----10---+----20---
CROCUS MIX
10 CASES
DAFFODIL
12 CASES
HYACINTH BLUE
8 BAGS
```

Correct answer: c

The third raw data file does not contain the same number of records for each observation, so the output from this data set will show invalid data for the ID and salary information in the fourth line.

Chapter 21: Creating Multiple Observations from a Single Record Answer Key

1. Which is true for the double trailing at sign (@@)?

a. It enables the next INPUT statement to read from the current record across multiple iterations of the DATA step.

b. It must be the last item that is specified in the INPUT statement.

c. It is released when the input pointer moves past the end of the record.

d. All of the above.

Correct answer: d

The double trailing at sign (@@) enables the next INPUT statement to read from the current record across multiple iterations of the DATA step. It must be the last item that is specified in the INPUT statement. A record that is being held by the double trailing at sign (@@) is not released until the input pointer moves past the end of the record, or until an INPUT statement that has no line-hold specifier executes.

2. A record that is being held by a single trailing at sign (@) is automatically released when

a. the input pointer moves past the end of the record.

b. the next iteration of the DATA step begins.

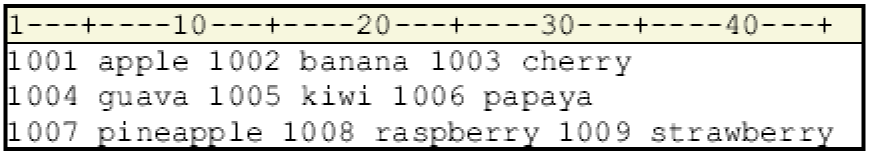
c. another INPUT statement that has a single trailing at sign (@) executes.

d. another value is read from the observation.

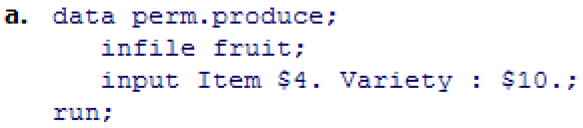
Correct answer: b

Unlike the double trailing at sign (@@), the single trailing at sign (@) is automatically released when control returns to the top of the DATA step for the next iteration. The trailing @ does not toggle on and off. If another INPUT statement that has a trailing @ executes, the holding effect is still on.

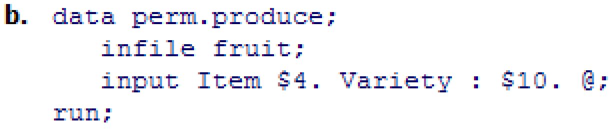
3. Which SAS program correctly creates a separate observation for each block of data?



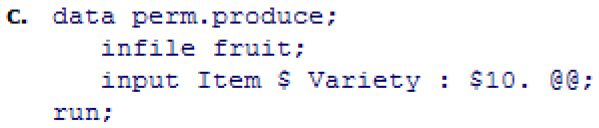
```
1---+----10---+----20---+----30---+----40---+
1001 apple 1002 banana 1003 cherry
1004 guava 1005 kiwi 1006 papaya
1007 pineapple 1008 raspberry 1009 strawberry
```



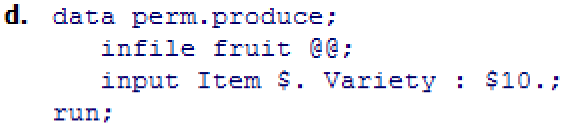
```
a. data perm.produce;
      infile fruit;
      input Item $4. Variety : $10.;
   run;
```



```
b. data perm.produce;
      infile fruit;
      input Item $4. Variety : $10. @;
   run;
```



```
c. data perm.produce;
      infile fruit;
      input Item $ Variety : $10. @@;
   run;
```

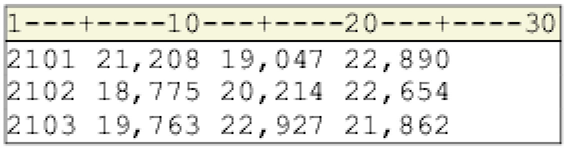


```
d. data perm.produce;
      infile fruit @@;
      input Item $. Variety : $10.;
   run;
```

Correct answer: c

Each record in this file contains three repeating blocks of data values for `Item` and `Variety`. The INPUT statement reads a block of values for `Item` and `Variety`, and then holds the current record by using the double trailing at sign (@@). The values in the program data vector are written to the data set as the first observation. In the next iteration, the INPUT statement reads the next block of values for `Item` and `Variety` from the same record.

4. Which SAS program segment reads the values for ID and holds the record for each value of `Quantity`, so that three observations are created for each record?



```
1---+----10---+----20---+----30
2101 21,208 19,047 22,890
2102 18,775 20,214 22,654
2103 19,763 22,927 21,862
```

```
a. data work.sales;
      infile unitsold;
      input ID $;
      do week=1 to 3;
         input Quantity : comma.;
         output;
      end;
   run;

b. data work.sales;
      infile unitsold;
      input ID $ @@;
      do week=1 to 3;
         input Quantity : comma.;
         output;
      end;
   run;

   c. data work.sales;
      infile unitsold;
      input ID $ @;
      do week=1 to 3;
         input Quantity : comma.;
         output;
      end;
   run;

   d. data work.sales;
      infile unitsold;
      input ID $ @;
      do week=1 to 3;
         input Quantity : comma. @;
         output;
      end;
   run;
```

Correct answer: d

This raw data file contains an ID fieldthat is followed by repeating fields. The first INPUT statement reads the values for `ID` and uses the @ line-hold specifier to hold the current record for the next INPUT statement in the DATA step. The second INPUT statement reads the values for `Quantity`. When all of the repeating fields have been read, control returns to the top of the DATA step, and the record is released.

5. Which SAS statement repetitively executes several statements when the value of an index variable named `Count` ranges from 1 to 50, incremented by 5?

a. `do count=1 to 50 by 5;`

b. `do while count=1 to 50 by 5;`

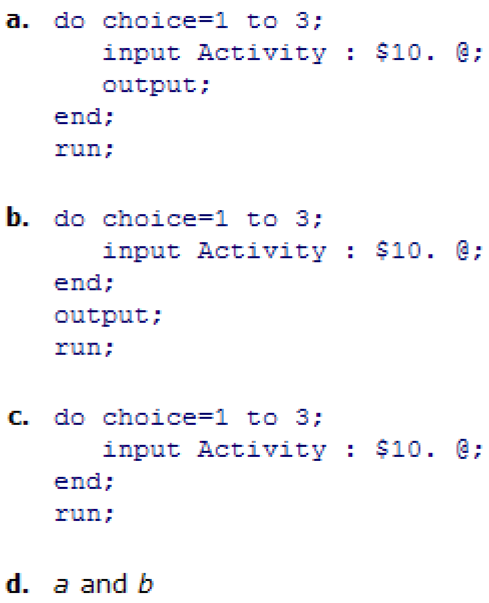
c. `do count=1 to 50 + 5;`

d. `do while (count=1 to 50 + 5);`

Correct answer: a

The iterative DO statement begins the execution of a loop based on the value of an index variable. Here, the loop executes when the value of `Count` ranges from 1 to 50, incremented by 5.

6. Which option below, when used in a DATA step, writes an observation to the data set after each value for `Activity` has been read?



a.
```
do choice=1 to 3;
   input Activity : $10. @;
   output;
end;
run;
```

b.
```
do choice=1 to 3;
   input Activity : $10. @;
end;
output;
run;
```

c.
```
do choice=1 to 3;
   input Activity : $10. @;
end;
run;
```

d. *a* and *b*

Correct answer: a

The OUTPUT statement must be included in the loop so that each time a value for `Activity` is read, an observation is immediately written to the data set.

7. Which SAS statement repetitively executes several statements while the value of `Cholesterol` is greater than 200?

a. `do cholesterol > 200;`

b. `do cholesterol gt 200;`

c. `do while (cholesterol > 200);`

d. `do while cholesterol > 200;`

Correct answer: c

The DO WHILE statement checks for the condition that `Cholesterol` is greater than 200. The expression must be enclosed in parentheses. The expression is evaluated at the top of the loop, before any statements are executed. If the condition is true, the DO WHILE loop executes. If the expression is false the first time it is evaluated, then the loop never executes.

8. Which choice below is an example of a Sum statement?

a. `totalpay=1;`

b. `totalpay+1;`

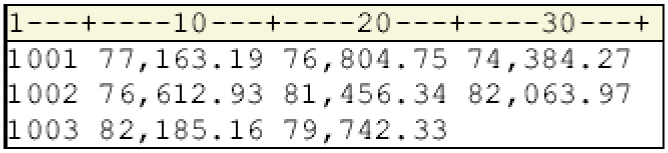
c. `totalpay*1;`

d. `totalpay by 1;`

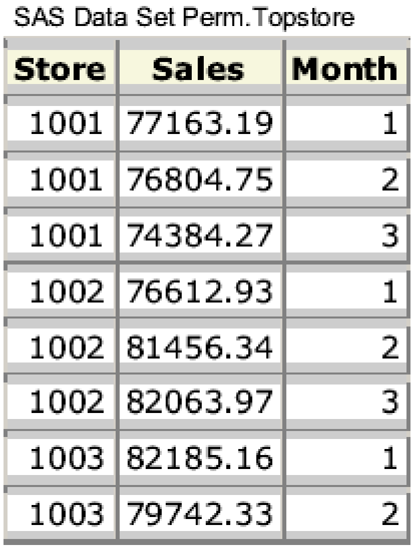
Correct answer: b

The Sum statement adds the result of an expression to a counter variable. So the + sign is an essential part of the Sum statement. Here, the value of `TotalPay` is incremented by 1.

9. Which program creates the SAS data set Perm.Topstore from the raw data file shown below?



```
1---+----10---+----20---+----30---+
1001 77,163.19 76,804.75 74,384.27
1002 76,612.93 81,456.34 82,063.97
1003 82,185.16 79,742.33
```



SAS Data Set Perm.Topstore

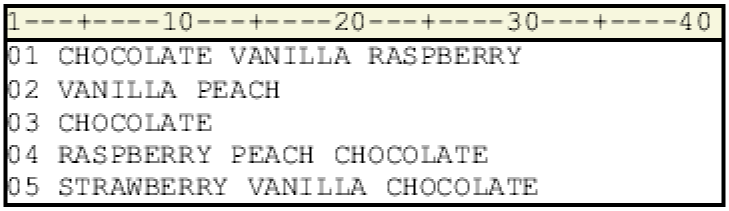
Store	Sales	Month
1001	77163.19	1
1001	76804.75	2
1001	74384.27	3
1002	76612.93	1
1002	81456.34	2
1002	82063.97	3
1003	82185.16	1
1003	79742.33	2

```
a. data perm.topstores;
   infile sales98 missover;
   input Store Sales : comma. @;
   do while (sales ne .);
      month + 1;
      output;
      input sales : comma. @;
   end;
run;

b. data perm.topstores;
   infile sales98 missover;
   input Store Sales : comma. @;
   do while (sales ne .);
      Month=0;
      month + 1;
      output;
      input sales : comma. @;
   end;
run;

c. data perm.topstores;
   infile sales98 missover;
   input Store Sales : comma. Month @;
   do while (sales ne .);
      month + 1;
      input sales : comma. @;
   end;
   output;
run;

d. data perm.topstores;
   infile sales98 missover;
   input Store Sales : comma. @;
   Month=0;
   do while (sales ne .);
      month + 1;
      output;
      input sales : comma. @;
   end;
run;
```

Correct answer: d

The assignment statement that precedes the DO WHILE loop creates the counter variable `Month` and assigns an initial value of zero to it. Each time the DO WHILE loop executes, the Sum statement increments the value of `Month` by 1.

10. How many observations are produced by the DATA step that reads this external file?



```
1---+----10---+----20---+----30---+----40
01 CHOCOLATE VANILLA RASPBERRY
02 VANILLA PEACH
03 CHOCOLATE
04 RASPBERRY PEACH CHOCOLATE
05 STRAWBERRY VANILLA CHOCOLATE
```

```
data perm.choices;
   infile icecream missover;
   input Day $ Flavor : $10. @;
   do while (flavor ne ' ');
      output;
      input flavor : $10. @;
    end;
run;
```

a. 3

b. 5

c. 12

d. 15

Correct answer: c

This DATA step produces one observation for each repeating field. The MISSOVER option in the INFILE statement prevents SAS from reading the next record when missing values occur at the end of a record. Every observation contains one value for `Flavor`, paired with the corresponding value for `ID`. Because there are 12 values for `Flavor`, there are 12 observations in the data set.

Chapter 22: Reading Hierarchical Files Answer Key

1. When you write a DATA step to create one observation per detail record you need to

   a. distinguish between header and detail records.

   b. keep the header record as part of each observation until the next header record is encountered.

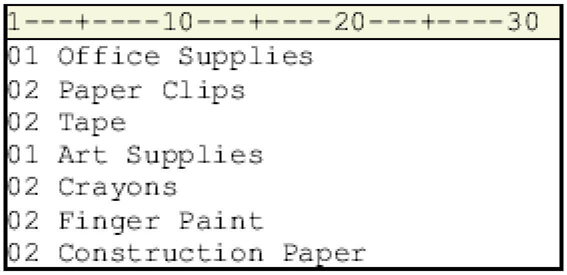
   c. hold the current value of each record type so that the other values in the record can be read.

d. all of the above

Correct answer: d

In order to create one observation per detail record, it is necessary to distinguish between header and detail records. Use a RETAIN statement to keep the header record as part of each observation until the next header record is encountered. You also need to use the @ line-hold specifier to hold the current value of each record type so that the other values in the record can be read.

2. Which SAS statement reads the value for `code` (in the first field), and then holds the value until an INPUT statement reads the remaining value in each observation in the same iteration of the DATA step?



```
1---+----10---+----20---+----30
01 Office Supplies
02 Paper Clips
02 Tape
01 Art Supplies
02 Crayons
02 Finger Paint
02 Construction Paper
```

a. `input code $2. @;`

b. `input code $2. @@;`

c. `retain code;`

d. none of the above

Correct answer: a

An INPUT statement is used to read the value for `code`. The single @ sign at the end of the INPUT statement holds the current record for a later INPUT statement in the same iteration of the DATA step.

3. Which SAS statement checks for the condition that `Record` equals C and executes a single statement to read the values for `Amount`?

a. `if record=c then input @3 Amount comma7.;`

b. `if record='C' then input @3 Amount comma7.;`

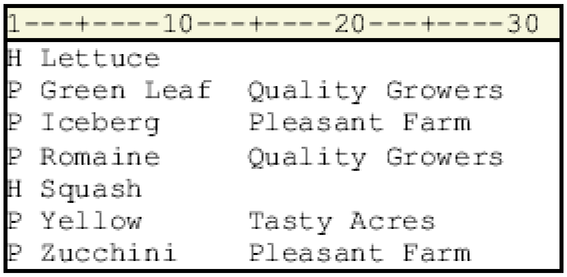
c. `if record='C' then do input @3 Amount comma7.;`

d. `if record=C then do input @3 Amount comma7.;`

Correct answer: b

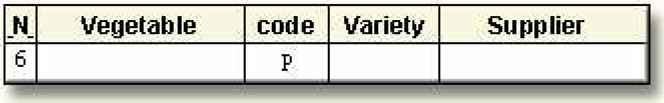
The IF-THEN statement defines the condition that `Record` equals C and executes an INPUT statement to read the values for `Amount` when the condition is true. C must be enclosed in quotation marks and must be specified exactly as shown because it is a character value.

4. After the value for `code` is read in the sixth iteration, which illustration of the program data vector is correct?



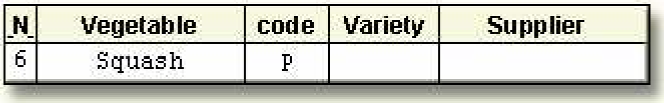
```
1---+----10---+----20---+----30
H Lettuce
P Green Leaf  Quality Growers
P Iceberg     Pleasant Farm
P Romaine     Quality Growers
H Squash
P Yellow      Tasty Acres
P Zucchini    Pleasant Farm
```

```
data perm.produce (drop=code);
   infile orders;
   retain Vegetable;
   input code $1. @;
   if code='H' then input @3 vegetable $6.;
   if code='P';
   input @3 Variety : $10. @15 Supplier : $15.;
run;
proc print data=perm.produce;
run;
```



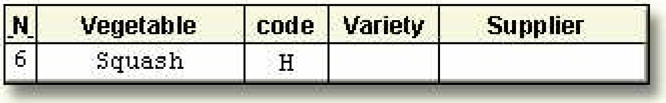
_N_	Vegetable	code	Variety	Supplier
6		P		

a.

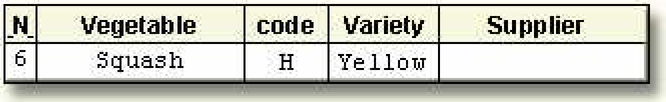


_N_	Vegetable	code	Variety	Supplier
6	Squash	P		

b.

c. 

_N_	Vegetable	code	Variety	Supplier
6	Squash	H		

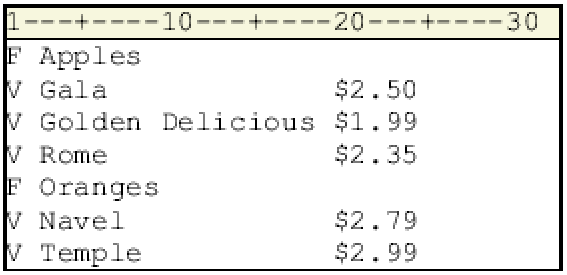
d. 

_N_	Vegetable	code	Variety	Supplier
6	Squash	H	Yellow	

Correct answer: b

The value of `Vegetable` is retained across iterations of the DATA step. As the sixth iteration begins, the INPUT statement reads the value for `code` and holds the record, so that the values for `Variety` and `Supplier` can be read with an additional INPUT statement.

5. What happens when the fourth iteration of the DATA step is complete?



```
1---+----10---+----20---+----30
F Apples
V Gala             $2.50
V Golden Delicious $1.99
V Rome             $2.35
F Oranges
V Navel            $2.79
V Temple           $2.99
```

```
data perm.orders (drop=type);
   infile produce;
   retain Fruit;
   input type $1. @;
   if type='F' then input @3 fruit $7.;
   if type='V';
   input @3 Variety : $16. @20 Price comma5.;
run;
```

a. All of the values in the program data vector are written to the data set as the third observation.

b. All of the values in the program data vector are written to the data set as the fourth observation.

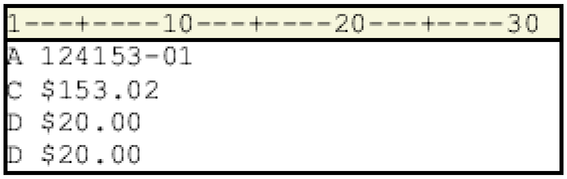
c. The values for `Fruit`, `Variety`, and `Price` are written to the data set as the third observation.

d. The values for `Fruit`, `Variety`, and `Price` are written to the data set as the fourth observation.

Correct answer: c

This program creates one observation for each detail record. The RETAIN statement retains the value for `Fruit` as part of each observation until the values for `Variety` and `Price` can be read. The DROP= option in the DATA statement prevents the values for `type` from being written to the data set.

6. Which SAS statement indicates that several other statements should be executed when `Record` has a value of A?



```
1---+----10---+----20---+----30
A 124153-01
C $153.02
D $20.00
D $20.00
```

a. `if record='A' then do;`

b. `if record=A then do;`

c. `if record='A' then;`

d. `if record=A then;`

Correct answer: a

The IF-THEN statement defines the condition that `Record` equals A and specifies a simple DO group. The keyword DO indicates that several executable statements follow until the DO group is closed by an END statement. The value A must be enclosed in quotation marks and specified exactly as shown because it is a character value.

7. Which is true for the following statements (X indicates a header record)?

```
if code='X' then do;
   if _n_ > 1 then output;
   Total=0;
   input Name $ 3-20;
end;
```

a. _N_ equals the number of times the DATA step has begun to execute.

b. When `code='X'` and `_n_ > 1` are true, an OUTPUT statement is executed.

c. Each header record causes an observation to be written to the data set.

d. a and b

Correct answer: d

`_N_` is an automatic variable whose value is the number of times the DATA step has begun to execute. The expression `_n_ > 1` defines a condition where the DATA step has executed more than once. When the conditions `code='X'` and `_n_ > 1` are true, an OUTPUT statement is executed, and `Total` is initialized to zero. Thus, each header record except for the first one causes an observation to be written to the data set.

8. What happens when the condition `type='P'` is false?

```
if type='P' then input @3 ID $5. @9 Address $20.;
else if type='V' then input @3 Charge 6.;
```

a. The values for `ID` and `Address` are read.

b. The values for `Charge` are read.

c. `Type` is assigned the value of V.

d. The ELSE statement is executed.

Correct answer: d

The condition is false, so the values for `ID` and `Address` are not read. Instead, the ELSE statement is executed and defines another condition which might or might not be true.

9. What happens when `last` has a value other than zero?

```
data perm.househld (drop=code);
   infile citydata end=last;
   retain Address;
   input type $1. @;
   if code='A' then do;
      if _n_ > 1 then output;
      Total=0;
      input address $ 3-17;
   end;
   else if code='N' then total+1;
   if last then output;
run;
```

a. `Last` has a value of 1.

b. The OUTPUT statement writes the last observation to the data set.

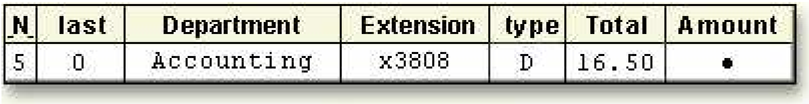
c. The current value of `last` is written to the DATA set.

d. a and b

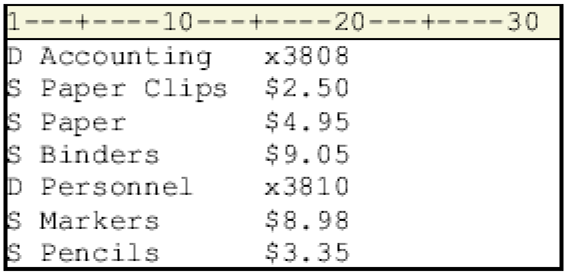
Correct answer: d

You can determine when the current record is the last record in an external file by specifying the END= option in the INFILE statement. `Last` is a temporary numeric variable whose value is zero until the last line is read. `Last` has a value of 1 after the last line is read. Like automatic variables, the END= variable is not written to the data set.

10. Based on the values in the program data vector, what happens next?



N	last	Department	Extension	type	Total	Amount
5	0	Accounting	x3808	D	16.50	•



```
1---+----10---+----20---+----30
D Accounting   x3808
S Paper Clips  $2.50
S Paper        $4.95
S Binders      $9.05
D Personnel    x3810
S Markers      $8.98
S Pencils      $3.35
```

```
data work.supplies (drop=type amount);
   infile orders end=last;
   retain Department Extension;
   input type $1. @;
   if type='D' then do;
      if _n_ > 1 then output;
      Total=0;
      input @3 department $10. @16 extension $5.;
   end;
   else if type='S' then do;
      input @16 Amount comma5.;
      total+amount;
      if last then output;
  end;
run;
```

a. All the values in the program data vector are written to the data set as the first observation.

b. The values for `Department`, `Total`, and `Extension` are written to the data set as the first observation.

c. The values for `Department`, `Total`, and `Extension` are written to the data set as the fourth observation.

d. The value of `last` changes to 1.

Correct answer: b

This program creates one observation for each header record and combines information from each detail record into the summary variable, `Total`. When the value of `type` is D and the value of `_N_` is greater than 1, the OUTPUT statement executes, and the values for `Department`, `Total`, and `Extension` are written to the data set as the first observation. The variables `_N_`, `last`, `type`, and `Amount` are not written to the data set.