Print

Quiz, Lesson 10: Combining SAS Data Sets

Your Score: 100%

Congratulations! Your score of 100% indicates that you've mastered the topics in this lesson. If you'd like, you can review the feedback for each question.

1. Which of the following statements is true about merging SAS data sets by using the DATA step?

   a. Merging combines observations from two or more data sets into a single observation in a new data set.
   b. SAS can merge data sets based on the position of observations in the original data set or by the values of one or more common variables.
   c. Match-merging is merging by values of one or more common variables.
   d. To match-merge data sets, all input data sets must be sorted or indexed on the BY variables.
   e. All of the above

   Your answer: e
   Correct answer: e

   All of these statements about merging SAS data sets by using the DATA step are true.

   Review: Merging Data Sets, The MERGE and BY Statements in the DATA Step

---

2. Which of the following programs concatenates the data sets, **sales** and **products**, in that order?

   a.
   ```
   data newsales;
      set products sales;
   run;
   ```
   b.
   ```
   data newsales;
      set sales products;
   run;
   ```
   c.
   ```
   data newsales;
      set sales;
      set products;
   run;
   ```

   Your answer: b
   Correct answer: b

   You list the data sets that you want to concatenate in the SET statement in the same order that you want to concatenate them. You list both data sets in one SET statement.

   Review: Specifying Multiple Data Sets in the SET Statement

---

3. If you run this DATA step, what observations does the data set **bonuses** contain?

```
data bonuses;
   merge managers staff;
   by EmpID;
run;
```

a. all of the observations from **managers**, and only those observations from **staff** with matching values for **EmpID**
b. all of the observations from **staff**, and only those observations from **managers** with matching values for **EmpID**
c. all observations from **staff** and all observations from **managers**, whether or not they have matching values
d. only those observations from **staff** and **manager** with matching values for **EmpID**

Your answer: c
Correct answer: c

By default, the output data set of a DATA step that includes a MERGE statement and a BY statement includes all of the observations from all of the input data sets that are listed in the MERGE statement.

Review: The MERGE and BY Statements in the DATA Step

✔ 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?

reps

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

close

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

a. missing
b. *$30,000*
c. *$40,000*
d. You cannot concatenate these data sets.

Your answer: a
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 **reps** does not contain a value for **Sale** because the variable does not exist in this data set, so a missing value appears for this variable.

Review: Concatenating Data Sets with Different Variables

✔ 5. What happens if you submit the following program to merge **donors1** and **donors2**, shown below?

```
data merged;
   merge donors1 donors2;
   by ID;
run;
```

donors1

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

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 eight observations.
c. The DATA step produces errors.

Your answer: c
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.

Review: The MERGE and BY Statements in the DATA Step

✔ 6. Suppose you want to concatenate these data sets. Which DATA step creates an output data set that combines the values of **Color** and **Hue** in the single variable **Color**?

widget1

Tag	Color	Model
77904	blue	AB42
56012	red	BA25
35499	orange	FC36

widget2

Tag	Hue	Model
89325	red	SP17
65888	yellow	BA12
00167	green	PG20

a.

```
data widgets_all;
   set widget1(rename=(Color=Hue))
       widget2;
run;
```

b.

```
data widgets_all;
   set widget1
       widget2(rename=(Hue=Color));
run;
```

c.

```
data widgets_all;
   set widget1
```

```
        widget2(Hue=color);
run;
```

Your answer: b
Correct answer: b

To combine two variables, you can use the RENAME= option to rename a variable for processing. You specify the RENAME= option next to the name of the data set that contains the variable that you want to rename.

Review: The RENAME= Data Set Option

---

7. What is the syntax error in this DATA step?

```
data returns_qtr1;
   set returns_jan(rename=(ID=CustID)
                   (Return=Item))
       returns_feb(rename=(Dt=Date))
       returns_mar;
run;
```

a. You cannot specify more than two data sets in the SET statement.
b. There are too many sets of parentheses in the RENAME= option.
c. You cannot specify multiple variables in the RENAME= option.
d. The BY statement is missing.

Your answer: b
Correct answer: b

The RENAME= option after the data set **returns_jan** has an extra set of parentheses. This is the correct code:

```
(rename=(ID=CustID Return=Item))
```

Review: The RENAME= Data Set Option

---

8. In the second iteration of this DATA step, after the data is merged, what are the values of **C** and **A**?

```
data client_amount;
   merge clients(in=C)
         amounts(in=A);
   by Name;
run;
```

clients

Name	EmpID
Ankerton	11123
Davis	22293
Masters	33351
Wolmer	44483

amounts

Name	Date	Amt
Ankerton	08OCT96	92
Ankerton	15OCT96	43
Davis	04OCT96	16
Masters	.	27
Thomas	21OCT96	15

a. C=*1*, A=*0*
b. C=*0*, A=*1*
c. C=*1*, A=*1*
d. missing
e. unknown

Your answer: c
Correct answer: c

Both of the data sets contribute observations on the second iteration of the DATA step, so the values of both IN= variables are *1*. The values of the IN= variables are always either *0* or *1*, depending on whether the data set contributes an observation to the output data set on that iteration.

Review: Using the IN= Data Set Option

---

9. If you run this DATA step, what observations does the data set **bonuses** contain?

```
data bonuses;
   merge managers (in=M)
         staff (in=S);
   by EmpID;
   if M=0 and S=1;
run;
```

a. only the observations from **staff** that have no match in **managers**
b. only the observations from **managers** that have no match in **staff**
c. all observations from both **managers** and **staff**, whether or not they match
d. no observations

Your answer: a
Correct answer: a

The subsetting IF statement selects all observations from **staff** that have no match in **managers**.

Review: Selecting Observations by Using the Subsetting IF Statement, Selecting Only Matches, Selecting Non-Matches

---

10. What is the relationship of the data set **first** to the data set **second** when merged by the variable **ID**?

first

Name	ID	Age
Togar	121150	39
Kylie	121152	34
Birin	121153	32
Gloria	121154	12
James	121155	36
Gene	121156	28
Tom	121157	27

second

ID	Date
121150	02/15/05
121152	05/22/07
121153	03/04/06
121154	11/22/05
121155	07/08/06
121156	12/15/06
121157	04/30/07

a. one-to-one
b. one-to-many
c. many-to-one
d. many-to-many
e. non-matching

Your answer: a
Correct answer: a

The SAS data sets **first** and **second** have a one-to-one matching relationship. Each observation in **first** matches one and only one observation in **second**.

Review: Knowing the Structure and Contents of Your Data

Close