Questions Module 2.6

In the text box beside each of the three variables, enter the letter of the appropriate modeling type.

Temp in C	a. continuous
Quality rating (1 through 5)	b. nominal
Blade material	c. ordinal

Incorrect.
The correct answers from top to bottom are **a**, **c**, and **b**.

In the text box beside each of the data quality issues below, enter the letter of the appropriate scenario.

incorrectly formatted data	a. A variable in your data set has typos.
missing data	b. You don't have data for all of the important variables.
dirty data	c. You are missing values for some of your variables.
incomplete data	d. A date variable is coded as nominal.

Incorrect.
The correct answers from top to bottom are **d**, **c**, **a**, and **b**.

You have a table with the following summary statistics.

Summary Statistics

5 Columns | Clear Select | Distribution

Columns	N	N Missing	N Categories	Min	Max	Mean	Std Dev	Median	Lower Quartile	Upper Quartile	Interquartile Range
Variable 1	5000	0	2	.	.	.	.	.	.	.	.
Variable 5	4760	240	2	.	.	.	.	.	.	.	.
Variable 6	4739	261	6	.	.	.	.	.	.	.	.
Variable 8	5000	0	.	0	5	0.1784	0.577096	0	0	0	0
Variable 13	3884	1116	.	0.52	203.31	33.482	8.327482	34.477	28.806	38.632	9.826

Which of the following statements is true?

- *a.* **Variable 13** is coded as nominal.
- *b.* **Variable 5** is not missing any values.
- *c.* **Variable 1** is coded as continuous.
- *d.* **Variable 8** contains mostly zeros.

Check My Answer

Correct.
Variable 8 is coded as continuous, but the minimum value, the median, the quartiles, and the interquartile range are all zero. From this, you can conclude that **Variable 8** includes mostly zeros.

You have the following data values for the categorical variable **Particle**

Size: *Small*, *small*, *Medium*, *Med.*, *medium*, *Medium*, *Small*. This is an example of which quality issue?

- *a.* incorrectly formatted data
- *b.* missing data
- *c.* messy or dirty data
- *d.* incomplete data

Incorrect.
The correct answer is **c**. This is messy, or dirty data. There is inconsistent use of capitalization and punctuation.

You have the following data values for the categorical variable **Particle**

Size: *Small*, *small*, *Medium*, *Med.*, *medium*, *Medium*, *Small*. You create a bar chart of **Particle Size** (with the existing data values). How many bars will you have in your bar chart?

- *a.* 2
- *b.* 3
- *c.* 5
- *d.* 7

Incorrect.
The correct answer is **c**. There are only two unique levels of **Particle Size**, but because there is inconsistent use of punctuation and capitalization, the bar chart will display five bars. In addition, you will have five levels of **Particle Size** in every analysis you run.

In the Missing Data Pattern table below, how many rows are not missing any data values?

	Count	Number of columns missing	Patterns	acceptable?	lot number	unit number	location	humidity	voltage	density
1	9	0	0000000	0	0	0	0	0	0	0
2	76	1	0000010	0	0	0	0	0	1	0
3	1	1	0000100	0	0	0	0	1	0	0
4	2	2	0000110	0	0	0	0	1	1	0
5	1	1	0001000	0	0	0	1	0	0	0
6	3	2	0001010	0	0	0	1	0	1	0
7	7	2	1000010	1	0	0	0	0	1	0
8	1	2	1000100	1	0	0	0	1	0	0

- *a.* 76
- *b.* 1
- *c.* 0
- *d.* 9

Check My Answer

Correct.
Nine rows in the original data table are not missing values for any of the variables.

You have data that are stored in separate columns in a data table. You want the labels to be stored in one column and the values to be stored in a separate column. What should you do?

- *a.* split the data
- *b.* stack the data
- *c.* transpose the data
- *d.* sort the data

Incorrect.
The correct answer is **b**. You should stack the data.

You have two data tables with information about two production lines. The data tables have the same structure and the same variable names. You would like to combine the data from these two tables into one file. What should you do?

- *a.* join the data tables
- *b.* transpose the data tables
- *c.* concatenate the data tables
- *d.* stack the data tables

Incorrect.
The correct answer is **c**. You should concatenate the data tables.