

Correct answers are colored red

1. In R, which of the following most closely aligns with our notion of a dataset in Stata? (Multiple Choice)

Answer 1: A matrix

Answer 2: A data frame

Answer 3: Arrays

Answer 4: None of the above

2. Which of the following objects are considered "heterogeneous" in that they allow for columns of data to have different core data types (for example, a column of integers and another column of character strings)? (Single Choice)

Answer 1: Matrices

Answer 2: Data frames

Answer 3: Neither

Answer 4: Both

3. Which core data type only allows for binary TRUE/FALSE values? (Single Choice)

Answer 1: Numeric

Answer 2: Logical

Answer 3: Character

4. Which of the following is NOT a value that R automatically recognizes as missing or not present in the data? (Single Choice)

Answer 1: .

Answer 2: NA

Answer 3: NULL

Answer 4: All of the above are automatically recognized by R as missing/not present

5. What type of information is a factor designed to handle? (Single Choice)

Answer 1: Continuous

Answer 2: Categorical

Answer 3: Either

6. Which is the correct syntax for subsetting the second row of matrixA? (Single Choice)

Answer 1: matrixA(2,0)

Answer 2: matrixA(2,)

Answer 3: matrixA[2,0]

Answer 4: matrixA[2,]

7. The process of compiling an R Markdown file into a document is called: (Single Choice)

Answer 1: Outputting

Answer 2: Knitting

Answer 3: Marking down

8. Which of the following must be included in an R Markdown file in order to knit a document? (Single Choice)

Answer 1: YAML header

Answer 2: Section headings

Answer 3: Code chunks

Answer 4: None of the above

9. Which Tidyverse function is used to sort rows by values within columns? (Single Choice)

Answer 1: order()

Answer 2: arrange()

Answer 3: sort()

Answer 4: none of the above

10. What symbol is used for the pipe operator that is heavily used with Tidyverse functions? (Single Choice)

Answer 1: <-

Answer 2: %>%

Answer 3: >>>