1. What is R?

a. A statistical programming language

b. A spreadsheet program

c. A web development language

d. An operating system

Answer: a. A statistical programming language

2. Which of the following is the correct syntax for assigning a value to a variable in R?

a. var = 10

b. 10 = var

c. var == 10

d. var := 10

Answer: a. var = 10

3. Which of the following is a valid variable name in R?

a. 2var

b. var2

c. var_2

d. var#2

Answer: b. var2

4. What is the output of the following code in R?

x <- 1:5 y <- x^2 plot(x, y)

a. A scatterplot of x versus y

b. A line plot of x versus y

c. A histogram of x

d. An error message

Answer: a. A scatterplot of x versus y

5. What is the output of the following code in R?

x <- c(1, 2, 3) y <- c(4, 5, 6) z <- x + y

a. An error message

b. The vector [5, 7, 9]

c. The vector [1, 2, 3, 4, 5, 6]

d. The vector [1, 4, 9]

Answer: b. The vector [5, 7, 9]

6. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- x[x > 2] print(y)

a. The vector [2, 3, 4, 5]

b. The vector [1, 2, 3]

c. The vector [3, 4, 5]

d. An error message

Answer: c. The vector [3, 4, 5]

7. What is the output of the following code in R?

`x <- c(1, 2, 3)

y <- c(4, 5, 6)

z <- cbind(x, y)

a. A matrix with two rows and three columns

b. A matrix with three rows and two columns

c. A list with two elements

d. An error message

Answer: a. A matrix with two rows and three columns

8. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) mean(x)

a. 3

b. 3.5

c. 4

d. 5

Answer: b. 3.5

9. What is the output of the following code in R?

x <- c(1, 2, 3) y <- c(2, 4, 6) cor(x, y)

a. -1

b. 0

c. 1

d. 2

Answer: c. 1

10. What is the output of the following code in R?

x <- c(1, 2, 3) y <- c(4, 5, 6) lm(y ~ x)

a. An error message

b. A linear regression model object

c. A scatterplot of x versus y

d. A summary of the regression model

Answer: b. A linear regression model object

11. Which of the following is a valid way to read in a CSV file in R?

a. read.csv("data.csv")

b. read.table("data.csv")

c. read.excel("data.csv")

d. load("data.csv")

Answer: a. read.csv("data.csv")

12. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- c(4, 5, 6, 7, 8) lm(y ~ x)$coefficients

a. A vector containing the intercept and slope of the linear regression model

b. A scatterplot of x versus y

c. A correlation matrix between x and y

d. An error message

Answer: a. A vector containing the intercept and slope of the linear regression model

13. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- c(4, 5, 6, 7, 8) summary(lm(y ~ x))$r.squared

a. The R-squared value of the linear regression model

b. A scatterplot of x versus y

c. A summary of the regression model

d. An error message

Answer: a. The R-squared value of the linear regression model

14. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- c(4, 5, 6, 7, 8) plot(x, y)

a. A scatterplot of x versus y

b. A line plot of x versus y

c. A bar plot of x versus y

d. An error message

Answer: a. A scatterplot of x versus y

15. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- c(4, 5, 6, 7, 8) cor(x, y)^2

a. The R-squared value of the linear regression model

b. A scatterplot of x versus y

c. The correlation coefficient between x and y

d. An error message

Answer: a. The R-squared value of the linear regression model

16. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- c(4, 5, 6, 7, 8) lm(y ~ x)

a. A scatterplot of x versus y

b. A summary of the regression model

c. The regression coefficients of the model

d. An error message

Answer: b. A summary of the regression model

17. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- c(4, 5, 6, 7, 8) summary(lm(y ~ x))$coefficients

a. A scatterplot of x versus y

b. A summary of the regression model

c. The regression coefficients of the model

d. An error message

Answer: c. The regression coefficients of the model

18. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- c(4, 5, 6, 7, 8) predict(lm(y ~ x))

a. A scatterplot of x versus y

b. The predicted values of y based on the regression model

c. A summary of the regression model

d. An error message

Answer: b. The predicted values of y based on the regression model

19. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- c(4, 5, 6, 7, 8) resid <- residuals(lm(y ~ x)) sum(resid)

a. The R-squared value of the linear regression model

b. A scatterplot of x versus y

c. The sum of squared residuals of the regression model

d. An error message

Answer: d. An error message

20. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- c(4, 5, 6, 7, 8) resid <- residuals(lm(y ~ x)) sum(resid^2)

a. The R-squared value of the linear regression model

b. A scatterplot of x versus y

c. The sum of squared residuals of the regression model

d. An error message

Answer: c. The sum of squared residuals of the regression model

21. Which of the following statements is true about missing values in R?

a. R automatically replaces missing values with the mean of the non-missing values

b. Missing values are denoted by the character “NA”

c. R automatically removes observations with missing values from analyses

d. All of the above

Answer: b. Missing values are denoted by the character “NA”

22. Which of the following statements is true about factors in R?

a. Factors are used to represent continuous variables

b. Factors are used to represent categorical variables

c. Factors are always stored as integers

d. Factors can be used in mathematical calculations

Answer: b. Factors are used to represent categorical variables

23. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- c(4, 5, 6, 7, 8) z <- c(“A”, “B”, “C”, “D”, “E”) df <- data.frame(x, y, z) head(df)

a. A scatterplot of x versus y

b. A summary of the data frame

c. The first six rows of the data frame

d. An error message

Answer: c. The first six rows of the data frame

24. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- c(4, 5, 6, 7, 8) z <- c(“A”, “B”, “C”, “D”, “E”) df <- data.frame(x, y, z) subset(df, y > 6)

a. A scatterplot of x versus y with only the points where y is greater than 6

b. A summary of the data frame with only the rows where y is greater than 6

c. The first six rows of the data frame

d. An error message

Answer: b. A summary of the data frame with only the rows where y is greater than 6

25. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- c(4, 5, 6, 7, 8) z <- c(“A”, “B”, “C”, “D”, “E”) df <- data.frame(x, y, z) levels(df$z)

a. The number of unique values in the z variable

b. The names of the levels in the z factor

c. The values of x where z is equal to “A”

d. An error message

Answer: b. The names of the levels in the z factor

26. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- c(4, 5, 6, 7, 8) z <- c(“A”, “B”, “C”, “D”, “E”) df <- data.frame(x, y, z) cor(dfx, dfy)

a. The correlation between x and y

b. A scatterplot of x versus y

c. The coefficients of a linear regression model predicting y from x

d. An error message

Answer: a. The correlation between x and y

27. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- c(4, 5, 6, 7, 8) z <- c(“A”, “B”, “C”, “D”, “E”) df <- data.frame(x, y, z) summary(lm(y ~ x, data = df))

a. A scatterplot of x versus y with a linear regression line

b. A summary of the data frame with the coefficients of a linear regression model predicting y from x

c. The p-value of a t-test for the slope of the linear regression model predicting y from x

d. An error message

Answer: b. A summary of the data frame with the coefficients of a linear regression model predicting y from x

28. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- c(4, 5, 6, 7, 8) z <- c("A", "B", "C", "D", "E") df <- data.frame(x, y, z) t.test(dfx, dfy)

a. A scatterplot of x versus y with a t-test for the difference in means

b. A summary of the data frame with the p-value of a t-test for the difference in means of x and y

c. The coefficients of a linear regression model predicting y from x with a t-test for the slope

d. An error message

Answer: b. A summary of the data frame with the p-value of a t-test for the difference in means of x and y

29. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- c(4, 5, 6, 7, 8) z <- c("A", "B", "C", "D", "E") df <- data.frame(x, y, z) library(ggplot2) ggplot(df, aes(x = x, y = y)) + geom_point()

a. A scatterplot of x versus y

b. A summary of the data frame

c. The first six rows of the data frame

d. An error message

Answer: a. A scatterplot of x versus y

30. Which of the following functions in R can be used to create a histogram?

a. scatterplot()

b. boxplot()

c. density()

d. hist()

Answer: d. hist()

31. Which of the following statements about factors in R is true?

a. Factors are used to represent numeric data in R.

b. Factors are used to represent categorical data in R.

c. Factors are used to represent missing values in R.

d. Factors are used to represent character data in R.

Answer: b. Factors are used to represent categorical data in R.

32. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- c(4, 5, 6, 7, 8) z <- c("A", "B", "C", "D", "E") df <- data.frame(x, y, z)

subset(df, x > 2)

a. A subset of the data frame where x is greater than 2

b. A subset of the data frame where y is greater than 2

c. A subset of the data frame where z is greater than 2

d. An error message

Answer: a. A subset of the data frame where x is greater than 2

33. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- c(4, 5, 6, 7, 8) z <- c("A", "B", "C", "D", "E") df <- data.frame(x, y, z)

aggregate(df$y, by = list(df$x), mean)

a. A summary of the data frame with the mean value of y for each value of x

b. A summary of the data frame with the median value of y for each value of x

c. A summary of the data frame with the maximum value of y for each value of x

d. An error message

Answer: a. A summary of the data frame with the mean value of y for each value of x

34. Which of the following functions in R can be used to generate a sequence of numbers?

a. rep()

b. seq()

c. sort()

d. length()

Answer: b. seq()

35. Which of the following functions in R can be used to calculate the standard deviation of a vector of numbers?

a. mean()

b. median()

c. var()

d. sd()

Answer: d. sd()

36. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- c(4, 5, 6, 7, 8) z <- c("A", "B", "C", "D", "E") df <- data.frame(x, y, z) colnames(df) <- c("X", "Y", "Z") head(df)

a. A data frame with three columns named X, Y, and Z

b. A data frame with three columns named x, y, and z

c. An error message

d. A vector with the names of the columns in the data frame

Answer: a. A data frame with three columns named X, Y, and Z

37. Which of the following functions in R can be used to generate a random sample from a vector?

a. mean()

b. median()

c. var()

d. sample()

Answer: d. sample()

38. Which of the following is a valid way to select the first row of a data frame named df in R?

a. df[0,]

b. df[1,]

c. df[,1]

d. df[,0]

Answer: b. df[1,]

39. Which of the following functions in R can be used to remove missing values from a vector?

a. na.rm()

b. na.omit()

c. na.fill()

d. na.exclude()

Answer: b. na.omit()

40. What is the output of the following code in R?

x <- 1:10 y <- 11:20 z <- cbind(x, y) z[,2]

a. A vector containing the values 1 through 10

b. A vector containing the values 11 through 20

c. A matrix containing the values 1 through 10 in the first column and the values 11 through 20 in the second column

d. An error message

Answer: b. A vector containing the values 11 through 20

41. Which of the following functions can be used to convert a factor variable in R to a character variable?

a. as.factor()

b. as.character()

c. as.numeric()

d. as.logical()

Answer: b. as.character()

42. Which of the following functions in R can be used to calculate the standard deviation of a vector?

a. var()

b. sd()

c. mean()

d. min()

Answer: b. sd()

43. Which of the following functions in R can be used to generate random numbers from a normal distribution?

a. rnorm()

b. runif()

c. rpois()

d. rbeta()

Answer: a. rnorm()

44. What is the output of the following code in R?

x <- list(1:3, 4:6, 7:9) lapply(x, sum)

a. A list containing the sums of the vectors in x

b. A list containing the mean of the vectors in x

c. A list containing the median of the vectors in x

d. An error message

Answer: a. A list containing the sums of the vectors in x

45. Which of the following functions in R can be used to calculate the correlation between two vectors?

a. cor()

b. cov()

c. sd()

d. var()

Answer: a. cor()

46. Which of the following functions in R can be used to create a scatterplot matrix?

a. plot()

b. pairs()

c. hist()

d. boxplot()

Answer: b. pairs()

47. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- c(4, 5, 6, 7, 8) cor(x, y)

a. The correlation coefficient between x and y

b. A scatterplot of x vs. y with a regression line

c. The R-squared value for the regression of y on x

d. An error message

Answer: a. The correlation coefficient between x and y

48. Which of the following functions in R can be used to perform principal component analysis (PCA)?

a. prcomp()

b. cor()

c. lm()

d. t.test()

Answer: a. prcomp()

49. Which of the following functions in R can be used to read in data from a CSV file?

a. read.csv()

b. read.table()

c. read.delim()

d. All of the above

Answer: d. All of the above

50. Which of the following functions in R can be used to create a boxplot?

a. boxplot()

b. hist()

c. plot()

d. lines()

Answer: a. boxplot ()

51. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- c(4, 5, 6, 7, 8) summary(lm(y ~ x))

a. A summary of the linear regression of y on x

b. A summary of the correlation between x and y

c. A summary of the principal components of x and y

d. An error message

Answer: a. A summary of the linear regression of y on x

52. Which of the following functions in R can be used to calculate the median of a vector?

a. mean()

b. median()

c. mode()

d. var()

Answer: b. median()

53. What is the output of the following code in R?

x <- c(1, 2, 3, 4, 5) y <- c(4, 5, 6, 7, 8) z <- x + y mean(z)

a. The arithmetic mean of the vector z

b. The arithmetic mean of the vector x

c. The arithmetic mean of the vector y

d. An error message

Answer: a. The arithmetic mean of the vector z

54. Which of the following functions in R can be used to create a cumulative distribution function (CDF) plot?

a. hist()

b. ecdf()

c. density()

d. qqnorm()

Answer: b. ecdf()

55. What is the output of the following code in R?

x <- c (1, 2, 3, 4, 5) y <- c (4, 5, 6, 7, 8) cor(x, y, method = “spearman”)

a. The Spearman correlation coefficient between x and y

b. The Pearson correlation coefficient between x and y

c. The Kendall correlation coefficient between x and y

d. An error message

Answer: a. The Spearman correlation coefficient between x and y

56. What is the use of the c () function in R?

a) It combines values into a vector or list.

b) It creates a new class.

c) It checks conditions.

d) It concatenates strings.