

Q.1. R is an \_\_\_\_\_ programming language?

- a) closed source      b) GPL      **c) Open source**      d) Definite source

Q.2. Packages are useful in collecting sets into a \_\_\_\_\_ unit ?

- a) single**      b) multiple      c) both      d) none of the above

Q.3. Who developed R?

- a) Dennis Ritchie      b) Bjarne Stroustrup      **c) John Chambers**      d) James Gosling

Q.4. \_\_\_\_\_ is used to make predictions about unknown future events?

- a) descriptive analysis      **b) predictive analysis**      c) both      d) none of the above

Q.5. R is an interpreted language so it can access through \_\_\_\_\_?

- a) disk operating system      b) user interface operating system  
c) operating system      **d) command line interpreter**

Q.6. How many steps does predictive analysis process contained?

- a) 5      b) 6      **c) 7**      d) 8

Q.7. Many quantitative analysts use R as their \_\_\_\_\_ tool?

- a) leading tool      b) programming tool      **c) both**      d) none of the above

Q.8. R was named partly after the first names of \_\_\_\_\_ R authors.?

- a) one      **b) two**      c) three      d) four

Q.10. Descriptive analysis tell about \_\_\_\_\_?

- a) past**      b) future      c) present      d) none of the above

Q.11. Vectors come in two parts: \_\_\_\_\_ and \_\_\_\_\_.

- a) atomic vectors and matrix      b) atomic vectors and array  
**c) atomic vectors and list**      d) none of them

Q.12. \_\_\_\_\_ initiates an infinite loop right from the start.

- a) never      **b) repeat**      c) break      d) set

Q.13. \_\_\_\_\_ programming language is a dialect of S.

- a) B      b) C      **c) D**      d) S

Q.14. Which of the following finds the maximum value in the vector x, exclude missing values

- a) rm(x)      b) all(x)      **c) max(x, na.rm=TRUE)**      d) x%in%y

Q.15. In 1991, R was created by Ross Ihaka and Robert Gentleman in the Department of Statistics at the University of \_\_\_\_\_.

- a) John Hopkins      b) California      c) Harvard      **d) Auckland**

- Q.16. Which of the following is a base package for R language ? a) util b) lang  
c) **tools** d) All of the mentioned
- Q.17. debug() flags a function for \_\_\_\_\_ mode in R mode.  
a) debug b) **run** c) compile d) All of the mentioned
- Q.18. The \_\_\_\_\_ function takes a vector or other objects and splits it into groups determined by a factor or list of factors.  
a. apply() b. **split()** c. lsplit() d) mapply()
- Q.19. lapply function takes \_\_\_\_\_ arguments in R language  
a. debug() b. trace() c. **4** d. 5
- Q.20. \_\_\_\_\_ is proprietary tool for predictive analytics.  
a. R b. **SAS** c. SSAS d. All of the mentioned
- Q.21. Which of the following is used for reading in saved workspaces ?  
a. **load** b. get c. unserialize d. None of the above
- Q.22. Which of the following argument denotes if the file has a header line ?  
a. sep b. file c. **header** d. None of the above
- Q.23. Which of the following is a wrong statement ?  
a. writeLines is used for for writing character data line-by-line to a file or connection  
b. dump is used for for dumping a textual representation of multiple R objects  
c. **All of the above**  
d. None of the above
- Q.24. Which of the following is used for outputting a textual representation of an R object ?  
a. dump b. dget c. **dput** d. None of the above
- Q.25. Which of the following is a correct statement ?  
a) read.data() function is one of the most commonly used functions for reading data  
b) unserialize is used for converting an R object into a binary format for outputting to a connection  
c) **save is used for saving an arbitrary number of R objects in binary format to a file**  
d) All of the above
- Q.26. Which of the following statement would read file foo.txt ?  
a data <- read.data("foo.txt") b **data <- read.table("foo.txt")**  
c read.data <- read.table("foo.txt") d All of the above
- Q.27. \_\_\_\_\_ is used to make predictions about unknown future events?  
a none of the above b both ( descriptive analysis & predictive analysis)  
c. **predictive analysis** d descriptive analysis

Q.28. Which of the following return a subset of the columns of a data frame?

- a. retrieve                      b. get                      **c. select**                      d. all of the mentioned

Q.29. Point out the correct statement :

- a. All of the mentioned**  
 b. There are packages on CRAN that implement data frames via things like relational databases that allow you to operate on very very large data frames  
 c. The data frame is a key data structure in statistics and in R  
 d. R has an internal implementation of data frames that is likely the one you will use most often

Q.30. \_\_\_\_\_ generate summary statistics of different variables in the data frame, possibly within strata

- a. subset                      b. set                      **c. summarize**                      d. rename

Q.31. \_\_\_\_\_ add new variables/columns or transform existing variables

- a. add                      **b. mutate**                      c. apped                      d. arrange

Q.32. The \_\_\_\_\_ operator is used to connect multiple verb actions together into a pipeline

- a. piper                      **b. pipe**                      c. start                      d. all of the mentioned

Q.33. The dplyr package can be installed from CRAN using :

- a. none of the mentioned                      **b. install.packages("dplyr")**  
 c. install.packages("dplyr")                      d. installed.packages("dplyr")

Q.34. The \_\_\_\_\_ function can be used to select columns of a data frame that you want to focus on.

- a. rename                      b. all of the mentioned                      c. get                      **d. select**

Q.35. \_\_\_\_\_ function is similar to the existing subset() function in R but is quite a bit faster.

- a. set                      b. rename                      c. subset                      **d. filter**

Q.36. Columns can be arranged in descending order too by using the special \_\_\_\_\_ operator.

- a. descending()                      **b. desc()**                      c. asc()                      d. subset

Q.37. Point out the wrong statement :

- a. Renaming a variable in a data frame in R is surprisingly hard to do  
**b. mute() function, which does the same thing as mutate() but then drops all non-transformed variables**  
 c. The mutate() function exists to compute transformations of variables in a data frame  
 d. None of the mentioned

Q.38. The \_\_\_\_\_ function is used to generate summary statistics from the data frame within strata defined by a variable.

- a. group()                      **b. group\_by()**                      c. groupby()                      d. arrange

Q.39. The \_\_\_\_\_ operator allows you to string operations in a left-to-right fashion.

- a. >%>%                      **b. %>%>**                      c. %>%                      d. All of the mentioned

Q.40. Which of the following extracts first element from the following vector ? > x <- c("a", "b", "c", "c", "d", "a")

- a. x[10].                      **b. x[1].**                      c. x[0].                      d. None of the mentioned

Q.41. Point out the correct statement :

- a. The [[ operator is used to extract elements of a list or data frame by string name  
**b. There are three operators that can be used to extract subsets of R objects**

c. The `[]` operator is used to extract elements of a list or data frame by literal name d. All of the mentioned

Q.42. Which of the following extracts first four element from the following vector ? `> x <- c("a", "b", "c", "c", "d", "a")`

- a. x[0:3].                      b. all of the mentioned                      **c. x[1:4].**                      d. x[0:4].

Q.43. What would be the output of the following code ? `> x <- c("a", "b", "c", "c", "d", "a") > x[c(1, 3, 4)]`

- a. "a" "c" "c"                      b. All of the mentioned                      c. "a" "c" "b"                      d. "a" "b" "c"

Q.44. Point out the wrong statement :

- a. All of the mentioned
- b. The `[]` operator always returns an object of the same class as the original
- c. `$` operator semantics are similar to that of `[]`
- d. **The `$` operator is used to extract elements of a list or a data frame**

Q.45. What would be the output of the following code ? `> x <- matrix(1:6, 2, 3) > x[1, 2]`

- a. 1                      **b. 3**                      c. 2                      d. 0

Q.46.What would be the output of the following code ?

```
> x <- matrix(1:6, 2, 3)
```

```
> x[1, ]
```

- a. 135                      b. 335                      c. file                      d. 235

Q.47. What will be the output of following code ?

```
> f <- function(a, b) {
```

```
+ print(a)
```

```
+ print(b)
```

 $+$   $\}$ 
$$> f(45)$$

- a. 42      b. 52      **c. 45**      d. 32

Q.48. What will be the output of following code snippet ?

```
> paste("a", "b", sep = ":")
```

- a. "a+b"      b. "a=b"      c. none of the mentioned      d. "a:b"

Q.49. Point out the correct statement :

- a. All of the mentioned
- b. The value NaN represents undefined value
- c. NaN can also be thought of as a missing value
- d. **Number Inf represents infinity in R**

Q.50. What would be the result of following code ?

```
> x <- vector("numeric", length = 10)
```

 $\geq x$ 

- a. 01                      b. 10                      c. None of the mentioned                      d. 0000000000

Q.51. The `c()` function can be used to create vectors of objects by concatenating things together.

- a. cp()                      b. concat()                      c. none of the mentioned                      **d. c()**

Q.52. Which of the following statement is invalid ?

- a. `x <- c(TRUE, FALSE)`      **b. None of the mentioned**      c. `x <- c(T, F)`      d. `x <- c(1+0i, 2+4i)`

Q.53. What will the following code print ?

```
> x <- 0:6
```

```
as.logical(x)
```

a. 0 1 2 3 4 5 6

**b. FALSE TRUE TRUE TRUE TRUE TRUE TRUE**

c. All of the mentioned

d. "0" "1" "2" "3" "4" "5" "6"

Q.54. \_\_\_\_\_ function returns a vector of the same size as x with the elements arranged in increasing order

a. orderasc()

b. orderby()

**c. sort()**

d. none of the mentioned

Q.55. Which of the following is used for generating sequences ?

a. order()

b. none of the mentioned

c. sequence()

**d. seq()**

Q.56. Which of the following statement would print "0" "1" "2" "3" "4" "5" "6" for the following code ?

```
> x <- 0:6
```

**a. as.character(x)**

b. none of the mentioned

c. as.numeric(x)

d. as.logical(x)

Q.57. What would the following code print ?

```
> x <- c("a", "b", "c")
```

```
> as.logical(x)
```

**a. NA NA NA**

b. a b c

c. All of the mentioned

d. 0 1 2

Q.60. What would be the output of the following code ?

```
> m <- matrix(nrow = 2, ncol = 3)
```

```
> dim(m)
```

**a. 2 3**

b. None of the mentioned

c. 3 2

d. 2 2

Q.61. What would be the result of following code ?

```
> x <- 1:3
```

```
> y <- 10:12
```

```
rbind(x, y)
```

a. All of the mentioned

**b. [,1] [,2] [,3] x 1 2 3 y 10 11 12**

c. [,1] [,2] [,3] x 1 2 3 y 4 5 6

d. [,1] [,2] [,3] x 1 2 3 y 10 11

Q.62. What would be the result of following code ?

```
> x <- list(1, "a", TRUE, 1 + 4i)
```

```
> x
```

a. advertisement [[1]] [1] 3 [[2]] [1] "a" [[3]] [1] TRUE [[4]] [1] 1+4i

b. [[1]] [1] 2 [[2]] [1] "b" [[3]] [1] TRUE [[4]] [1] 1+4i

**c. [[1]] [1] 1 [[2]] [1] "a" [[3]] [1] TRUE [[4]] [1] 1+4i**

d. All of the mentioned

Q.63. What would the following code print ?

```
> x <- c(1, 2, NaN, NA, 4)
```

```
> is.na(x)
```

a. TRUE FALSE TRUE TRUE FALSE

b. None of the mentioned

c. FALSE TRUE TRUE TRUE FALSE

**d. FALSE FALSE TRUE TRUE FALSE**

Q.64. Which of the following would print the following output ?

foo bar 1 1 TRUE

2 2 TRUE

3 3 FALSE 4

4 FALSE

a. `> x <- data.frame(foo = 1:6, bar = c(F, T, F, F)) > x`

c. `> x <- data.frame(foo = 1:4, bar = c(T, T, F, F)) > x`

b. `> x <- data.frame(foo = 1:4, bar = c(F, T, F, F)) > x`

d. None of the mentioned

Q.65. What would the following code print ?

`> x <- data.frame(foo = 1:4, bar = c(T, T, F, F))`

`> ncol(x)`

a. 7

b. 4

c. All of the mentioned

d. 2

Q.66. What would be the output of the following code ?

`x <- 1:3`

`> names(x)`

a. NUL

b. 1

c. None of the mentioned

d. 2

Q.67. Which of the following argument denotes if the file has a header line ?

a. sep

b. all of the mentioned

c. header

d. file

Q.68. Which of the following code would read 100 rows ?

Select one:

a. all of the mentioned

b. `initial <- read.table("datatable.txt", nrow = 99)`

c. `tabAll <- read.table("datatable.txt", colClasses = classes)`

d. **`initial <- read.table("datatable.txt", nrow = 100)`**

Q.69. Which of the following return a subset of the columns of a data frame?

a. get

b. retrieve

c. all of the mentioned

d. select

Q.70. The dplyr package can be installed from CRAN using :

a. none of the mentioned

b. **`install.packages("dplyr")`**

c. `installed.packages("dplyr")`

d. `install.packages("dplyr")`

Q.71. Which of the following is example of vectorized operation as far as subtraction is concerned ?

`> x <- 1:4 > y <- 6:9`

a. `x/y`

b. `x*y`

c. `x+y`

d. **`x-y`**

Q72. The nominal scale of measurement has the properties of the \_\_\_\_\_.

a. ordinal scale

b. only interval scale

c. ratio scale

d. **None of these alternatives is correct.**

Q73. Some hotels ask their guests to rate the hotel's services as excellent, very good, good, and poor. This is an example of the \_\_\_\_\_.

a. **ordinal scale**

b. ratio scale

c. nominal scale

d. interval scale

Q74. Categorical data \_\_\_\_\_.

a. indicate either how much or how many

b. cannot be numeric

c. **are labels used to identify attributes of elements**

d. must be nonnumeric

Q75. In a sample of 400 students in a university, 80, or 20%, are Business majors. Based on the above information, the school's paper reported that "20% of all the students at the university are Business majors." This report is an example of \_\_\_\_\_.

- a. a sample                      b. a population                      c. statistical inference                      d. **descriptive statistics**

Q76. A frequency distribution is \_\_\_\_\_.

- a. a tabular summary of a set of data showing the relative frequency  
b. a graphical form of representing data  
c. **a tabular summary of a set of data showing the frequency of items in each of several nonoverlapping classes**  
d. a graphical device for presenting categorical data

Q77. The relative frequency of a class is computed by \_\_\_\_\_.

- a. dividing the midpoint of the class by the sample size  
b. dividing the frequency of the class by the midpoint  
c. dividing the sample size by the frequency of the class  
d. **dividing the frequency of the class by the sample size**

Q78. A researcher is gathering data from four geographical areas designated: South = 1; North = 2; East = 3; West = 4. The designated geographical regions represent \_\_\_\_\_.

- a. **categorical data**                      b. quantitative data                      c. label data                      d. either quantitative or categorical data

Q79. A cumulative relative frequency distribution shows \_\_\_\_\_.

- a. **the proportion of data items with values less than or equal to the upper limit of each class**  
b. the proportion of data items with values less than or equal to the lower limit of each class  
c. the percentage of data items with values less than or equal to the upper limit of each class  
d. the percentage of data items with values less than or equal to the lower limit of each class

### Exhibit 1

A survey of 800 college seniors resulted in the following crosstabulation regarding their undergraduate major and whether or not they plan to go to graduate school.

Undergraduate Major				
Graduate School	Business	Engineering	Others	Total
Yes	70	84	126	280
No	182	208	130	520
<b>Total</b>	252	292	256	800

Q80. Refer to Exhibit 1 Those students who are majoring in business, what percentage plans to go to graduate school?

- a. 27.78                      b. 8.75                      c. 70                      d. 72.22

Q81. Refer to Exhibit 1. Among the students who plan to go to graduate school, what percentage indicated "Other" majors?

- a. 15.75                      b. 45                      c. 54                      d. 35

Q82. The collection of all possible sample points in an experiment is \_\_\_\_\_.

- a. **the sample space**                      b. a sample point                      c. an experiment                      d. the population

Q83. The intersection of two mutually exclusive events \_\_\_\_\_.

- a. can be any value between 0 to 1  
 c. **must always be equal to 0**  
 b. must always be equal to 1  
 d. can be any positive value

Q84. If  $P(A) = 0.4$ ,  $P(B | A) = 0.35$ ,  $P(A \cap B) = 0.69$ , then  $P(B) =$

- a. 0.14  
 b. **0.43**  
 c. 0.75  
 d. 0.59

### Exhibit 2

A survey of a sample of business students resulted in the following information regarding the genders of the individuals and their selected major.

Gender	Management	Marketing	Others	Total
Male	40	10	30	80
Female	30	20	70	120
<b>Total</b>	70	30	100	200

Q85. In the exhibit 2, Given that a person is male, what is the probability that he is majoring in Management?

- a. 0.20  
 b. 0.25  
 c. 0.50  
 d. 0.40

Q86. In the exhibit 2, what is the probability of selecting a male individual?

- a. 0.15  
 b. 0.25  
 c. 0.45  
 d. 0.40

Q87. A description of the distribution of the values of a random variable and their associated probabilities is called a \_\_\_\_\_.

- a. **probability distribution**  
 b. random variance  
 c. random variable  
 d. expected value

Q88. An experiment consists of determining the speed of automobiles on a highway by the use of radar equipment.

The random variable in this experiment is a \_\_\_\_\_.

- a. discrete random variable  
 b. **continuous random variable**  
 c. complex random variable  
 d. simplex random variable

Q89. In the textile industry, a manufacturer is interested in the number of blemishes or flaws occurring in each 100 feet of material. The probability distribution that has the greatest chance of applying to this situation is the \_\_\_\_\_.

- a. normal distribution  
 b. binomial distribution  
 c. **Poisson distribution**  
 d. uniform distribution

### Exhibit 3

The following represents the probability distribution for the daily demand of computers at a local store.

Demand	Probability
0	0.1
1	0.2
2	0.3
3	0.2
4	0.2

Q90. Refer to Exhibit 3. The probability of having a demand for at least two computers is \_\_\_\_.



- a. 0.7                      b. 0.3                      c. 0.4                      d. 1.0

Q91. The closer the sample mean is to the population mean, \_\_\_\_\_.

- a. the larger the sampling error                      b. **the smaller the sampling error**  
c. the sampling error equals 1                      d. None of these alternatives is correct.

Q92. As the sample size becomes larger, the sampling distribution of the sample mean approaches a \_\_\_\_\_.

- a. binomial distribution                      b. Poisson distribution  
c. **normal distribution**                      d. chi-square distribution

Q93. A sample of 225 elements from a population with a standard deviation of 75 is selected. The sample mean is 180. The 95% confidence interval for  $\mu$  is \_\_\_\_\_.

- a. 105.0 to 225.0                      b. 175.0 to 185.0                      c. 100.0 to 200.0                      d. **170.2 to 189.8**

Q94. Whenever the population standard deviation is unknown and the population has a normal or near-normal distribution, which distribution is used in developing interval estimation?

- a. standard distribution                      b. z distribution                      c. alpha distribution                      d. **t distribution**

Q95. A normal distribution with a mean of 0 and a standard deviation of 1 is called \_\_\_\_\_.

- a. a probability density function                      b. an ordinary normal curve  
c. **a standard normal distribution**                      d. None of these alternatives is correct.

Q96. In a standard normal distribution, the probability that Z is greater than zero is \_\_\_\_\_.

- a. **0.5**                      b. equal to 1                      c. at least 0.5                      d. 1.96

### Exhibit 4

The weight of football players is normally distributed with a mean of 200 pounds and a standard deviation of 25 pounds.

Q97. Refer to Exhibit 4. What percent of players weigh between 180 and 220 pounds?

- a. 28.81%                      b. **0.5762%**                      c. 0.281%                      d. 57.62%

Q98. Refer to Exhibit 4. What is the minimum weight of the middle 95% of the players?

- a. 196                      b. **151**                      c. 249                      d. 190

Q99. The p-value \_\_\_\_\_.

- a. is the same as the Z statistic  
b. measures the number of standard deviations from the mean  
c. is a distance  
d. **is a probability**

Q100. A two-tailed test is performed at 95% confidence. The p-value is determined to be 0.09. The null hypothesis \_\_\_\_\_.

- a. must be rejected  
b. **should not be rejected**  
c. could be rejected, depending on the sample size  
d. has been designed incorrectly

Q101. A machine is designed to fill toothpaste tubes with 5.8 ounces of toothpaste. The manufacturer does not want any underfilling or overfilling. The correct hypotheses to be tested are \_\_\_\_\_.

**Ans H0:  $\mu = 5.8$  Ha:  $\mu \neq 5.8$**

Q102. The probability of committing a Type I error when the null hypothesis is true is

- a. the confidence level                      b. greater than 1                      c. **the Level of Significance**                      d.  $\beta$

Q103. Independent simple random samples are taken to test the difference between the means of two populations whose variances are not known, but are assumed to be equal. The sample sizes are  $n_1 = 32$  and  $n_2 = 40$ . The correct distribution to use is the \_\_\_\_\_.

- a. t distribution with 73 degrees of freedom                      b. t distribution with 72 degrees of freedom  
c. t distribution with 71 degrees of freedom                      d. **t distribution with 70 degrees of freedom**

Q104. We are interested in testing to see if the variance of a population is less than 7. The correct null hypothesis is

- a)  $< 7$                       b.  $^2 7$                       c.  $S < 49$                       d.  $D. S > 49$

Q105. A regression analysis between sales (in \$1000) and price (in dollars) resulted in the following equation =  $60 - 8X$ . The equation implies that an \_\_\_\_\_.

- a. increase of \$1 in price is associated with a decrease of \$8 in sales  
b. increase of \$8 in price is associated with an decrease of \$52,000 in sales  
c. increase of \$1 in price is associated with a decrease of \$52 in sales  
d. **increase of \$1 in price is associated with a decrease of \$8000 in sales**

Q106. In regression analysis, an outlier is an observation whose \_\_\_\_\_.

- a. mean is larger than the standard deviation  
b. residual is zero  
c. mean is zero  
d. **residual is much larger than the rest of the residual values**

Q107. In a situation where the dependent variable can assume only one of the two possible discrete values, \_\_\_\_\_.

- a. we must use multiple regression  
b. there can only be two independent variables  
c. **logistic regression should be applied**  
d. all the independent variables must have values of either zero or one

Q108. A statistical test conducted to determine whether to reject or not reject a hypothesized probability distribution for a population is known as a \_\_\_\_\_.

- a. contingency test                      b. probability test  
b. c. **goodness of fit test**                      d. None of these alternatives is correct.

Q110. A collection of statistical methods that generally requires very few, if any, assumptions about the population distribution is known as \_\_\_\_\_.

- a. parametric methods                      b. **nonparametric methods**  
c. distribution-fixed methods                      d. normal

Q111. From a population of size 400, a random sample of 40 items is selected. The median of the sample \_\_\_\_\_.

- a. must be 200, since 400 divided by 2 is 200  
b. must be 10, since 400 divided by 40 is 10  
c. must be equal to the median of population, if the sample is truly random  
d. **None of these alternatives is correct.**

Q112. Fifteen people were given two types of cereal, Brand X and Brand Y. Two people preferred Brand X and thirteen people preferred Brand Y. We want to determine whether or not customers prefer one brand over the other. In the above said condition, to test the null hypothesis, the appropriate probability distribution to use is \_\_\_\_\_.

- a. normal                      b. chi-square                      c. Poisson                      d. **binomial**

113. The correlation coefficient is used to determine:

- a. A specific value of the y-variable given a specific value of the x-variable  
b. A specific value of the x-variable given a specific value of the y-variable  
c. **The strength of the relationship between the x and y variables**  
d. None of these

114. If there is a very strong correlation between two variables then the correlation coefficient must be

- a. any value larger than 1  
b. **much smaller than 0, if the correlation is negative**  
c. much larger than 0, regardless of whether the correlation is negative or positive  
d. None of these alternatives is correct.

115. In regression, the equation that describes how the response variable (y) is related to the explanatory variable (x) is:

- a. the correlation model  
b. **the regression model**  
c. used to compute the correlation coefficient  
d. None of these alternatives is correct.

116. The relationship between number of beers consumed (x) and blood alcohol content (y) was studied in 16 male college students by using least squares regression. The following regression equation was obtained from this study:

$$\hat{y} = -0.0127 + 0.0180x$$

The above equation implies that:

- a. each beer consumed increases blood alcohol by 1.27%  
b. on average it takes 1.8 beers to increase blood alcohol content by 1%  
c. **each beer consumed increases blood alcohol by an average of amount of 1.8%**  
d. each beer consumed increases blood alcohol by exactly 0.018

117. SSE can never be

- a. **larger than SST**                      b. smaller than SST                      c. equal to 1                      d. equal to zero

118. Regression modeling is a statistical framework for developing a mathematical equation that describes how

- a. one explanatory and one or more response variables are related  
b. several explanatory and several response variables response are related  
c. **one response and one or more explanatory variables are related**  
d. All of these are correct.

119. In regression analysis, the variable that is being predicted is the

- a. **response, or dependent, variable**
- b. independent variable
- c. intervening variable
- d. is usually x

120. Regression analysis was applied to return rates of sparrowhawk colonies. Regression analysis was used to study the relationship between return rate (x: % of birds that return to the colony in a given year) and immigration rate (y: % of new adults that join the colony per year). The following regression equation was obtained.

$$\hat{y} = 31.9 - 0.34x$$

121. Based on the above estimated regression equation, if the return rate were to decrease by 10% the rate of immigration to the colony would:

- a. increase by 34%
- b. **increase by 3.4%**
- c. decrease by 0.34%
- d. decrease by 3.4%

122. In least squares regression, which of the following is not a required assumption about the error term  $\epsilon$ ?

- a. **The expected value of the error term is one.**
- b. The variance of the error term is the same for all values of x.
- c. The values of the error term are independent.
- d. The error term is normally distributed.

123. Larger values of  $r^2$  ( $R^2$ ) imply that the observations are more closely grouped about the

- a. average value of the independent variables
- b. average value of the dependent variable
- c. **least squares line**
- d. origin

124. In a regression analysis if  $r^2 = 1$ , then

- a. SSE must also be equal to one
- b. **SSE must be equal to zero**
- c. SSE can be any positive value
- d. SSE must be negative

125. The coefficient of correlation

- a. is the square of the coefficient of determination
- b. **is the square root of the coefficient of determination**
- c. is the same as r-square
- d. can never be negative

126. In regression analysis, the variable that is used to explain the change in the outcome of an experiment, or some natural process, is called

- a. the x-variable
- b. the independent variable
- c. the predictor variable
- d. the explanatory variable
- e. **all of the above (a-d) are correct**
- f. none are correct

127. In the case of an algebraic model for a straight line, if a value for the x variable is specified, then

- a. **the exact value of the response variable can be computed**

- b. the computed response to the independent value will always give a minimal residual
- c. the computed value of  $y$  will always be the best estimate of the mean response
- d. none of these alternatives is correct.

128. A regression analysis between sales (in \$1000) and price (in dollars) resulted in the following equation:

$$\hat{y} = 50,000 - 8X$$

The above equation implies that an

- a. increase of \$1 in price is associated with a decrease of \$8 in sales
- b. increase of \$8 in price is associated with an increase of \$8,000 in sales
- c. increase of \$1 in price is associated with a decrease of \$42,000 in sales
- d. **increase of \$1 in price is associated with a decrease of \$8000 in sales**

129. In a regression and correlation analysis if  $r^2 = 1$ , then

- a.  $SSE = SST$
- b.  $SSE = 1$
- c.  $SSR = SSE$
- d.  **$SSR = SST$**

130. If the coefficient of determination is a positive value, then the regression equation

- a. must have a positive slope
- b. must have a negative slope
- c. **could have either a positive or a negative slope**
- d. must have a positive  $y$  intercept

131. If two variables,  $x$  and  $y$ , have a very strong linear relationship, then

- a. there is evidence that  $x$  causes a change in  $y$
- b. there is evidence that  $y$  causes a change in  $x$
- c. **there might not be any causal relationship between  $x$  and  $y$**
- d. None of these alternatives is correct.

132. If the coefficient of determination is equal to 1, then the correlation coefficient

- a. must also be equal to 1
- b. **can be either -1 or +1**
- c. can be any value between -1 to +1
- d. must be -1

133. In regression analysis, if the independent variable is measured in kilograms, the dependent variable

- a. must also be in kilograms
- b. must be in some unit of weight
- c. cannot be in kilograms
- d. **can be any units**

134. The data are the same as for question 4 above. The relationship between number of beers consumed ( $x$ ) and blood alcohol content ( $y$ ) was studied in 16 male college students by using least squares regression. The following regression equation was obtained from this study:

$$\hat{y} = -0.0127 + 0.0180x$$

Suppose that the legal limit to drive is a blood alcohol content of 0.08. If Ricky consumed 5 beers the model would predict that he would be:

- a. 0.09 above the legal limit
- b. **0.0027 below the legal limit**

c. 0.0027 above the legal limit

d. 0.0733 above the legal limit

135. In a regression analysis if  $SSE = 200$  and  $SSR = 300$ , then the coefficient of determination is

a. 0.6667

b. **0.6000**

c. 0.4000

d. 1.5000

136. If the correlation coefficient is 0.8, the percentage of variation in the response variable explained by the variation in the explanatory variable is

a. 0.80%

b. 80%

c. 0.64%

d. **64%**

137. If the correlation coefficient is a positive value, then the slope of the regression line

a. **must also be positive**

b. can be either negative or positive

c. can be zero

d. can not be zero

138. If the coefficient of determination is 0.81, the correlation coefficient

a. is 0.6561

b. **could be either + 0.9 or - 0.9**

c. must be positive

D. must be negative

139. A fitted least squares regression line

a. **may be used to predict a value of y if the corresponding x value is given**

b. is evidence for a cause-effect relationship between x and y

c. can only be computed if a strong linear relationship exists between x and y

d. None of these alternatives is correct.

140. Regression analysis was applied between \$ sales (y) and \$ advertising (x) across all the branches of a major international corporation. The following regression function was obtained.

$$\hat{y} = 5000 + 7.25x$$

If the advertising budgets of two branches of the corporation differ by \$30,000, then what will be the predicted difference in their sales?

a. **\$217,500**

b. \$222,500

c. \$5000

d. \$7.25

141. Suppose the correlation coefficient between height (as measured in feet) versus weight (as measured in pounds) is 0.40. What is the correlation coefficient of height measured in inches versus weight measured in ounces? [12 inches = one foot; 16 ounces = one pound]

a. **0.40**

b. 0.30

c. 0.533

d. cannot be determined from information given

e. none of these

142. Assume the same variables as in question 28 above; height is measured in feet and weight is measured in pounds. Now, suppose that the units of both variables are converted to metric (meters and kilograms). The impact on the slope is:

a. the sign of the slope will change

b. **the magnitude of the slope will change**

c. both a and b are correct

d. neither a nor b are correct

143. Suppose that you have carried out a regression analysis where the total variance in the response is 133452 and the correlation coefficient was 0.85. The residual sums of squares is:

- a. **37032.92**
- b. 20017.8
- c. 113434.2
- d. 15%

144. This question is related to questions 4 and 21 above. The relationship between number of beers consumed (x) and blood alcohol content (y) was studied in 16 male college students by using least squares regression. The following regression equation was obtained from this study:

$$\hat{y} = -0.0127 + 0.0180x$$

Another guy, his name Dudley, has the regression equation written on a scrap of paper in his pocket. Dudley goes out drinking and has 4 beers. He calculates that he is under the legal limit (0.08) so he decides to drive to another bar. Unfortunately Dudley gets pulled over and confidently submits to a roadside blood alcohol test. He scores a blood alcohol of 0.085 and gets himself arrested. Obviously, Dudley skipped the lecture about residual variation. Dudley's residual is:

- a. +0.005
- b. -0.005
- c. **+0.0257**
- d. -0.0257

145. You have carried out a regression analysis; but, after thinking about the relationship between variables, you have decided you must swap the explanatory and the response variables. After refitting the regression model to the data you expect that:

- a. the value of the correlation coefficient will change
- b. **the value of SSE will change**
- c. the value of the coefficient of determination will change
- d. the sign of the slope will change
- e. nothing changes

146. Suppose you use regression to predict the height of a woman's current boyfriend by using her own height as the explanatory variable. Height was measured in feet from a sample of 100 women undergraduates, and their boyfriends, at Dalhousie University. Now, suppose that the height of both the women and the men are converted to centimeters. The impact of this conversion on the slope is:

- a. the sign of the slope will change
- b. the magnitude of the slope will change
- c. both a and b are correct
- d. **neither a nor b are correct**

147. A residual plot:

- a. displays residuals of the explanatory variable versus residuals of the response variable.
- b. displays residuals of the explanatory variable versus the response variable.
- c. **displays explanatory variable versus residuals of the response variable.**
- d. displays the explanatory variable versus the response variable.
- e. displays the explanatory variable on the x axis versus the response variable on the y axis.

148. When the error terms have a constant variance, a plot of the residuals versus the independent variable x has a pattern that

- a. fans out



- a. **0.8045**                      b. -0.8045                      c. 0                      d. 1
- b) The least squares estimate of  $b_1$  equals
- a. 0.923                      b. **1.991**                      c. -1.991                      d. -0.923
- c) The least squares estimate of  $b_0$  equals
- a. 0.923                      b. 1.991                      c. -1.991                      d. **-0.923**
- d) The sum of squares due to regression (SSR) is
- a. 1434                      b. 505.98                      c. 50.598                      d. **928.02**



e) The coefficient of determination equals

- a. **0.6471**      b. -0.6471      c. 0      d. 1

f) The point estimate of  $y$  when  $x = 0.55$  is

- a. **0.17205**      b. 2.018      c. 1.0905      d. -2.018      e. -0.17205

USM VITA