Specialization : Business Analytics

Course Code : 205 Course Name: Business Analytics using R Programming

MCQ

Sr No	Question	Answer
1	**Which of these measures are used to analyse the central tendency of data?** a. Mean and Normal Distribution b. Mean, Median and Mode c. Mode, Alpha & Range d. Standard Deviation, Range and Mean e. Median, Range and Normal Distribution	**B**
2	Five numbers are given: (5, 10, 15, 5, 15). Now, what would be the sum of deviations of individual data points from their mean? A) 10 B)25 C) 50 D) 0 E) None of the above	**D**
3	A test is administered annually. The test has a mean score of 150 and a standard deviation of 20. If Ravi's z-score is 1.50, what was his score on the test? A) 180 B) 130 C) 30 D) 150 E) None of the above	**A**
4	Business intelligence (BI) is a broad category of application programs which includes _____ a) Decision support b) Data mining c) OLAP d) All of the mentioned	**A**
5	Point out the correct statement. a) OLAP is an umbrella term that refers to an assortment of software applications for analyzing an organization's raw data for intelligent decision making b) Business intelligence equips enterprises to gain business advantage from data c) BI makes an organization agile thereby giving it a lower edge in today's evolving market condition d) None of the mentioned	**A**
6	BI can catalyze a business's success in terms of _____ a) Distinguish the products and services that drive revenues b) Rank customers and locations based on profitability c) Ranks customers and locations based on probability d) All of the mentioned	**d**
7	Which of the following areas are affected by BI? a) Revenue b) CRM c) Sales d) All of the mentioned	**D**
8	1. Business intelligence (BI) is a broad category of application programs which	**D**

	includes _____ a) Decision support b) Data mining c) OLAP d) All of the mentioned	
9	Which of the following measures of central tendency will always change if a single value in the data changes? A) Mean B) Median C) Mode D) All of these	**A**
10	Strong assessment items are made up of five elements: a) Standard b) Stimulus c) Stem d) Key e) Distractors	**A**
11	A good question is ---------- It focuses on recall of only the material covered in your lesson and aligns well with the overall learning objectives a) relevant. b) clear c) concise d) purpose	**B**
12	A good question is framed in a----------, easily understandable language, without any vagueness. Students should understand what is wanted from the question even when they don't know the answer to it. a) clear b) relevant c) concise d) purpose	**A**
13	A good question is usually crisp and----------. It omits any unnecessary information that requires students to spend time understanding it correctly. The idea is not to trick learners but assess their knowledge. a) concise b) clear c) relevant d) purpose	**A**
14	1. ____ programming language is a dialect of S. a) B b) C c) R d) K	**C**
15	Point out the WRONG statement? a) Early versions of the S language contain functions for statistical modeling b) The book Programming with Data by John Chambers documents S version of the language c) In 1993 Bell Labs gave StatSci (later Insightful Corp.) an exclusive license to develop and sell the S language d) The book Programming with Data by IBM documents S version of the language	**C**
16	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	**D**

17	Point out the wrong statement? a) R is a language for data analysis and graphics b) K is language for statistical modelling and graphics c) One key limitation of the S language was that it was only available in a commercial package, S-PLUS d) C is a language for data and graphics	**A**
18	Business analytics results in which of these? a. Evidence Based Decisions b. Data Driven Decisions c. Better Decisions **d.** All of these are correct	**D**
19	Which one of the following is not a type of Business Analytics? a. Descriptive Analytics b. Diagnostic Analytics c. Predictive Analytics **d.** Performance Analytics	**D**
20	What will be the output of the following R code snippet? > **paste**("a", "b", se = ":") a) “a+b” b) “a=b” c) “a b :” d) none of the mentioned	**D**
21	Point out the correct statement? a) In R, a function is an object which has the mode function b) R interpreter is able to pass control to the function, along with arguments that may be necessary for the function to accomplish the actions that are desired c) Functions are also often written when code must be shared with others or the public d) All of the mentione	**D**
22	The _____ function returns a list of all the formal arguments of a function. a) formals() b) funct() c) formal() d) fun()	**A**
23	What will be the output of the following R code snippet? > f <- **function**(num = 1) { + hello <- "Hello, world!**n**"	**A**

	+ **for**(i **in seq_len**(num)) { + **cat**(hello) + } + chars <- **nchar**(hello) * num + chars + } > f() a) Hello, world! [1] 14 b) Hello, world! [1] 15 c) Hello, world! [1] 16 d) Error	
24	Point out the wrong statement? a) A formal argument can be a symbol, a statement of the form 'symbol = expression', or the special formal argument b) The first component of the function declaration is the keyword function c) The value returned by the call to function is not a function d) Functions are also often written when code must be shared with others or the public	**A**
25	You can check to see whether an R object is NULL with the _____ function. a) is.null() b) is.nullobj() c) null() d) as.nullobj()	**A**
26	Which of the following code will print NULL? a) > args(paste) b) > arg(paste)	**A**

	c) > args(pastebin) d) > arg(bin)	
27	What will be the output of the following R code snippet? > **paste**("a", "b", sep = ":") a) “a+b” b) “a=b” c) “a:b” d) a*b	**A**
28	What will be the output of the following R code snippet? > f <- **function**(a, b) { + **print**(a) + **print**(b) + } > f(45) a) 32 b) 42 c) 52 d) 45	**A**
29	What will be the output of the following R code snippet? > f <- **function**(a, b) { + a^2 + } > f(2) a) 4 b) 3 c) 2 d) 5	**A**
30	Which of the following is a base package for R language? a) util b) lang c) tools d) All of the above	**C**

31	R comes with a _____ to help you optimize your code and improve its performance. a) Debugger b) Monitor c) Profiler d) None of the above	**A**
32	debug() flags a function for _____ mode in R mode. a) debug b) run c) compile d) None of the above	**B**
33	_____ suspends the execution of a function wherever it is called and puts the function in debug mode a) recover() b) browser() c) Both of the above	**C**
34	A matrix is ___dimensionsinal rectangular data set? a) 5 b) 4 c) 3 d) 2	**D**
35	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) isplit() d) mapply()	**B**
36	lapply function takes___ arguments in R language a) 1 b) 3 c) 4 d) 5	**C**
37	___is used to apply a function over subsets of a vector a) apply()	**d**

	b) lapply() c) mapply() d) tapply() **a)**	
38	_____applies a function over the margins of an array a) apply() b) lapply() c) tapply() d) mapply()	**A**
39	___function is same as lapply() in R b) apply() c) lapply() d) sapply() e) tapply()	**C**
40	_____ loop over a list and evaluate a function on each element a) apply() b) lapply() c) sapply() d) tapply()	**A**
41	_____ is proprietary tool for predictive analytics. a) R b) SAS c) SSAS d) SPSS	**B**
42	Data frames can be converted to a matrix by calling data._____ a) matr() b) mat() c) matrix() d) None of the above	**C**
43	Which of the following method make a vector of repeated values? a) rep() b) data()	**b**

	c) view() d) None of the above	
44	R objects can have attributes, which are like _____ for the object a) metadata b) features c) expressions	**A**
45	Attributes of an object (if any) can be accessed using the _____ function. a) objects() b) attrib() c) attributes()	**C**
46	_____ involves predicting a response with meaningful magnitude, such as quantity sold, stock price, or return on investment. a) Regression b) Clustering c) Summarization	**A**
47	_____ provides needed string operators in R a) str b) forcast c) stringr	**C**
48	_____ splits a data frame and results in an array (hence the da). Hopefully, you're getting the idea here. a) apply b) daply c) stats	**B**
49	System.time function returns an object of class _____ which contains two useful bits of information. a) debug_time b) procedure_time c) proc_time	**C**
50	Which of the following will start the R program? a) $ R b) & R	**a**

	c) Rb	
	Unit 2	
1	The third step in decision making process is a linear predictions b dependent predictions c making predictions d independent predictions	**C**
2	The decision making step, which consists of organization goals, predicting alternatives and communicating goals is called a organization b alternation c planning d valuing	**C**
3	The fourth step in decision making process is a linear correlation b making decisions c implement decisions d evaluate performance	**B**
4	The costs that behaves as irrelevant costs in process of decision making are classified as a past costs b future costs c expected costs d sunk costs	**A**
5	Which of these is not a topic covered in a typical Business Analyst Aptitude Test? a. Analytical Thinking c. Data Interpretation b. Listening Skills d. Risk Management	**D**
6	If the test should be 30 minutes, Analytical Thinking is taken in how many minutes? a. 5 c. 10 b. 7 d. 15	**C**
7	Primary objective of a business analyst is to help businesses implement a. Business systems b. Business solutions c. Technology systems d. Technology solutions	**B**
8	Which business professional performs cost-benefit analyses of existing and potential customers a) Marketer b) Financial Analyst	**C**

	c) Business Analyst d) Sales Representative	
9	1. A Use Case is a set of steps, typically defining interactions between a role, True of False a. True b. False	**A**
10	Any fact that the solution can assume to be true when the use case begins is what? a. A win b. A Failure c. A success d. A Precondition	**C**
11	A State Diagram is used for what? a. Which Events cause a transition between states b. Which events cause a success between states c. Allowable behaviour d, All	**D**
12	A Solution Requirement is comprised of two types of requirements what are they? a, Functional b. Hard c. Existing d. Non-Functional	**A**
13	Which of the following is used for Statistical analysis in R language? a) Studio b) RStudio c) Heck	**B**
14	R functionality is divided into a number of _____ a) Packages b) Functions c) Domains	**A**
15	Which of the following is an example of vectorized operation as far as subtraction is concerned? > x <- 1:4 > y <- 6:9 a) x+y b) x-y	**b**

	c) x/y d) x*y	
16	What would be the output of the following code? > x <- 1:4 > y <- 6:9 > z <- x + y > z a) 7 9 11 13 b) 7 9 11 13 14 c) 9 11 13 d) Null	**A**
17	What would be the output of the following code? > x <- 1:4 > x > 2 a) FALSE FALSE TRUE TRUE b) 1 2 3 4 c) 1 2 3 4 5	**A**
18	What would be the value of the following expression? log(-1) a) Warning in log(-1): NaNs produced b) 1 c) Null d) 0	**A**
19	What will be the output of the following code? > g <- function(x) { + a <- 3 + x+a+y	**d**

	c + ## 'y' is a free variable + } > g(2) a) 8 b) 9 c) 42 d) Error	

20	What will be the output of the following code?
	
	function(p) {
	
	params[!fixed] <- p
	
	mu <- params[1]
	
	sigma <- params[2]
	
	## Calculate the Normal density
	
	a <- -0.5*length(data)*log(2*pi*sigma^2)
	
	b <- -0.5*sum((data-mu)^2) / (sigma^2)
	
	-(a + b)
	
	}
	
	> ls(environment(nLL))
	
	a) "data" "fixed" "param"
	b) "data" "variable" "params"
	c) "data" "fixed" "params"
	d) None of the above
		C

21	Which of the following is a principle of analytic graphics? a) Don't plot more than two variables at at time b) Make judicious use of color in your scatterplots c) Show box plots (univariate summaries)	**D**

	d) Show causality, mechanism, explanation	
22	R is an_____programming language? a) Closed source b) GPL c) Open source d) Definite sourc	**C**
23	Who developed R? a) Dennis Ritchie b) John Chambers c) Bjarne Stroustrup	**A**
24	R was named partly after the first names of___R authors? a) One b) Two c) Three d) Four	**B**
25	Packages are useful in collecting sets into a____unit ? a) Single b) Multiple	**C**
26	Many quantitative analysts use R as their___tool? a) Leading tool b) Programming tool c) Both the above	**D**
27	Predictive analysis is the branch of _____analysis? a) Advanced b) Core c) Both the above	**B**
28	_____ is used to make predictions about unknown future events? a) Descriptive analysis b) Predicitive analysis c) Both the above	**C**
29	How many steps does the predictive analysis process contained? a) 5	**d**

	b) 6 c) 7 d) 8	
30	Descriptive analysis tell about_____? a) Past b) Present c) Future	**A**
31	How many types of R objects are present in R data type? a) 4 b) 5 c) 6 d) 7	**C**
32	How many types of data types are present in R? a) 4 b) 5 c) 6 d) 7	**A**
33	Which of the following is a primary tool for debugging? a) debug() b) trace() c) browser() d) None of the above	**B**
34	Which function is used to create the vector with more than one element? a) Library() b) plot() c) c() d) par()	**C**
35	In R every operation has a _____call? a) System b) Function c) None of the above	**A**
36	The _____ in R is a vector. a) Basic data structure b) Basic datatypes	**b**

	c) Both	
37	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	**C**
38	Vectors come in two parts____ and ____. a) Atomic vectors and matrix b) Atomic vectors and array c) Atomic vectors and list	**A**
39	How many types of atomic vectors are present? a) 3 b) 4 c) 5 d) 6	**C**
40	How many types of vertices functions are peresent? a) 1 b) 2 c) 3 d) 4	**B**
41	_____and_____ are types of matrices functions? a) Apply and sapply b) Apply and lapply c) Both	**C**
42	How many control statements are present in R? a) 6 b) 7 c) 8 d) 9	**A**
43	Which of the following finds the maximum value in the vector x, exclude missing values a) rm(x) b) all(x)	**b**

	c) max(x, na.rm=TRUE) d) x%in%y	
44	Which of the following sort dataframe by the order of the elements in B a) a.x[rev(order(x$B)),] b) b.x[ordersort(x$B), c) c.x[order(x$B),]	**A**
45	_____ initiates an infinite loop right from the start. a) Never b) Repeat c) Break d) Set	**B**
46	_____ is used to skip an iteration of a loop. a) Next b) Skip c) Group	**A**
47	____ programming language is a dialect of S. a) B b) C c) D d) S	**A**
48	In 1991, R was created by Ross Ihaka and Robert Gentleman in the Department of Statistics at the University of _____. a) Auckland b) Harvard c) California d) John Hopkins	**A**
49	Finally, in _____ R version 1.0.0 was released to the public. a) 2000 b) 2005 c) 2010 **d)** 2012	**D**
50	R is technically much closer to the Scheme language than it is to the original ____	**c**

	language. a) B b) S c) C d) C++	
	Unit-3	
1	They primary R system is available from the _____ a) CRAN b) CRWO c) GNU d) CRDO	**C**
2	Point out the wrong statement? a) Key feature of R was that its syntax is very similar to S b) R runs only on Windows computing platform and operating system c) R has been reported to be running on modern tablets, phones, PDAs, and game consoles d) R functionality is divided into a number of Packages	**D**
3	R functionality is divided into a number of _____ a) Packages b) Functions c) Domains d) Classes	**A**
4	Which Package contains most fundamental functions to run R? a) root b) child c) base d) parent	**A**
5	Which language is best for the statistical environment? a) C b) R c) Java d) Python	**B**
6	In order to use the R-related functionality in Dundas BI, you must have access to an existing _____ a) Console b) Terminal c) Packages d) R serve	**D**
7	The open source _____ software is available for Unix, Linux, and Windows platforms.	**A**

	a) Rserve b) BServe c) CServe d) Dserve	
8	Modification in Dundas BI is done _____ a) Directly b) Indirectly c) Need access to Server d) Not known	**A**
9	Is It possible to inspect the source code of R? a) Yes b) No c) Can't say d) Some times	**A**
10	_____ function is used to watch for all available packages in library. a) lib() b) fun.lib() c) libr() d) library()	**D**
11	The longer programs are called _____ a) Files b) Structures c) Scripts d) Data	**D**
12	Scripts will run on _____ a) Script Editors b) Console c) Terminal d) GCC Compiler	**A**
13	What will be the output of the following R function? ab <- list(1, 2, 3, "X", "Y", "Z") dim(ab) <- c(3,2) print(ab) a. 123 Xyz b. Error c. Xyz123 d. 123xyz	**A**
14	What is the meaning of the following R function? x <- c(4, 5, 1, 2, 3, 3, 4, 4, 5, 6) x <- as.factor(x) a) x becomes a factor b) x is a factor c) x does not exist	**A**

	d) x is not a vector	
15	What is the meaning of the following R function? print(sqrt(2)) a) 1.414314 b) 1.414214 c) Error d) 14.1414	**B**
16	What will be the output of the following R function? d <- date() a) Prints todays date b) Prints some date c) Prints exact present time and date d) Error	**C**
17	Which of the following commands will correctly read the above csv file with 5 rows in a dataframe? A) csv('Dataframe.csv') B) csv('Dataframe.csv',header=TRUE) C) dataframe('Dataframe.csv') D) csv2('Dataframe.csv',header=FALSE,sep=',')	**B**
18	R functionality is divided into a number of _____ a) Packages b) Functions c) Domains	**A**

19	**Consider the following function.**
	f <- function(x) {
	g <- function(y) {
	y + z
	}
	z <- 4
	x + g(x)
	}
	If we execute following commands (written below), what will be the output?
	z <- 10
	f(4)
		A

	A) 12 B) 7 C) 4 D) 16	
20	**The iris dataset has different species of flowers such as Setosa, Versicolor and Virginica with their sepal length. Now, we want to understand the distribution of sepal length across all the species of flowers. One way to do this is to visualise this relation through the graph shown below.** **Which function can be used to produce the graph shown above?** A) xyplot() B) stripplot() C) barchart() D) bwplot()	**B**
21	The plot above is of type strip whereas the options a, c and d will produce a scatter, bar and box whisker plot respectively. Therefore, option B is the correct solution. Alpha 125.5 0 Beta 235.6 1 Beta 212.03 0 Beta 211.30 0 Alpha 265.46 1 File Name – Dataframe.csv **Which of the following commands will correctly read the above csv file with 5 rows in a dataframe?** A) csv('Dataframe.csv') B) csv('Dataframe.csv',header=TRUE) C) dataframe('Dataframe.csv') D) csv2('Dataframe.csv',header=FALSE,sep=',')	**D**
22	**Excel file format is one of the most common formats used to store datasets. It is important to know how to import an excel file into R. Below is an excel file in which data has been entered in the third sheet.** **Alpha 125.5 0**	**D**

	Beta 235.6 1 Beta 212.03 0 Beta 211.30 0 Alpha 265.46 1 **File Name – Dataframe.xlsx** **Which of the following codes will read the above data in the third sheet into a dataframe in R?** A) Openxlsx::read.xlsx("Dataframe.xlsx",sheet=3,colNames=FALSE) B) Xlsx::read.xlsx("Dataframe.xlsx",sheetIndex=3,header=FALSE) C)XLConnect::readWorksheetFromFile("Dataframe.xlsx",sheet=3,header=FALSE) D)All of the above	
23	**A 10 Sam** B 20 Peter C 30 Harry D ! ? E 50 Mark **File Name – Dataframe.csv** **Missing values in this csv file has been represented by an exclamation mark ("!") and a question mark ("?"). Which of the codes below will read the above csv file correctly into R?** A) csv('Dataframe.csv') B) csv('Dataframe.csv',header=FALSE, sep=',',na.strings=c('?')) C) csv2('Dataframe.csv',header=FALSE,sep=',',na.strings=c('?','!')) D) dataframe('Dataframe.csv')	**C**
24	**Column 1 Column 2 Column 3** Row 1 15.5 14.12 69.5 Row 2 18.6 56.23 52.4 Row 3 21.4 47.02 63.21	**d**

	Row 4 36.1 56.63 36.12 **File Name – Dataframe.csv** **6) The above csv file has row names as well as column names. Which of the following code will read the above csv file properly into R?** A) delim('Train.csv',header=T,sep=',',row.names=TRUE) B) csv2('Train.csv',header=TRUE, row.names=TRUE) C) dataframe('Train.csv',header=TRUE,sep=',') D) csv('Train.csv',,header=TRUE,sep=',')	
25	**Column 1 Column 2 Colum** Row 1 15.5 14.12 69.5 Row 2 18.6 56.23 52.4 Row 3 21.4 47.02 63.21 Row 4 36.1 56.63 36.12 **File Name – Dataframe.csv** **Which of the following codes will read only the first two rows of the csv file?** A) csv('Dataframe.csv',header=TRUE,row.names=1,sep=',',nrows=2 B) csv2('Dataframe.csv',row.names=1,nrows=2) C) delim2('Dataframe.csv',header=T,row.names=1,sep=',',nrows=2) D) dataframe('Dataframe.csv',header=TRUE,row.names=1,sep=',',skip.last=2)	**A**
26	Dataframe1 Dataframe2 **Feature1 Feature2 Feature3 Feature4 Feature1 Feature2 Feature3** A 1000 25.5 10 E 5000 65.5 B 2000 35.5 34 F 6000 75.5 C 3000 45.5 78 G 7000 85.5 D 4000 55.5 3 H 8000 95.5 **There are two dataframes stored Dataframe1 and Dataframe2 shown above. Which of the following codes will produce the output shown below?**	**D**

	Feature1 **Feature2** **Feature3** A 1000 25.5 B 2000 35.5 C 3000 45.5 D 4000 55.5 E 5000 65.5 F 6000 75.5 G 7000 85.5 H 8000 95.5 A) merge(dataframe[,1:3],dataframe2) B) merge(dataframe1,dataframe2)[,1:3] C) merge(dataframe1,dataframe2,all=TRUE) D) Both 1 and 2 E) All of the above	
27	**V1** **V2** 1 121.5 461 2 516 1351 3 451 6918 4 613 112 5 112.36 230 6 25.23 1456 7 12 457	**e**

	dataframe **A data set has been read in R and stored in a variable "dataframe". Which of the below codes will produce a summary (mean, mode, median) of the entire dataset in a single line of code?** A) summary(dataframe) B) stats(dataframe) C) summarize(dataframe) D) summarise(dataframe) E) None of the above	
28	**A dataset has been read in R and stored in a variable "dataframe". Missing values have been read as NA.** A 10 Sam B NA Peter C 30 Harry D 40 NA E 50 Mark dataframe**Which of the following codes will not give the number of missing values in each column?** A) colSums(is.na(dataframe)) B) apply(is.na(dataframe),2,sum) C) sapply(dataframe,function(x) sum(is.na(x)) D) table(is.na(dataframe))	**D**
29	**One of the important phase in a Data Analytics pipeline is univariate analysis of the features which includes checking for the missing values and the distribution, etc. Below is a dataset and we wish to plot histogram for "Value" variable.** **Parameter State Value Deper** Alpha Active 50 2 Beta Active 45 5 Beta Passive 25 0 Alpha Passive 21 0	**D**

	Alpha	Passive	26	1	
	Beta	Active	30	2	
	Beta	Passive	18	0	

dataframed

Which of the following commands will help us perform that task ?

A) hist(dataframed$Value)

B) ggplot2::qplot(dataframed$Value,geom="Histogram")

C)ggplot2::ggplot(data=dataframed,aes(dataframe$Value))+geom_histogram()

D) All of the above

30

Parameter	**State**	**Value**	**Usage**
Alpha	Active	50	0
Beta	Active	45	1
Beta	Passive	25	0
Alpha	Passive	21	0
Alpha	Passive	26	1
Beta	Active	30	1
Beta	Passive	18	0

D

Certain Algorithms like XGBOOST work only with numerical data. In that case, categorical variables present in dataset are first converted to DUMMY variables which represent the presence or absence of a level of a categorical variable in the dataset. For example After creating the Dummy Variable for the feature "Parameter", the dataset looks like below.

Parameter_Alpha	**Parameter_Beta**	**State**	**Value**	**U**
1	0	Active	50	0
0	1	Active	45	1

	0	1	Passive	25	0	
	1	0	Passive	21	0	
	1	0	Passive	26	1	
	0	1	Active	30	1d	
	0	1	Passive	18	0d	

Which of the following commands will help us to achieve this?
A) dummies:: dummy.data.frame(dataframe,names=c('Parameter'))
B) dataframe$Parameter_Alpha=0
dataframe$Gende_Beta=0
dataframe$Parameter_Alpha[which(dataframe$Parameter=='Alpha')]=1
dataframe$Parameter_Beta[which(dataframe$Parameter=='Alpha')]=0
dataframe$Parameter_Alpha[which(dataframe$Parameter=='Beta']=0
dataframe$Parameter_Beta[which(dataframe$Parameter=='Beta']=1
C) contrasts(dataframe$Parameter)
D) Both 1 and 2

		Column1	**Column2**	**Column3**	**Column4**	**Column5**	**d**
	Name1	Alpha	12	24	54	0	
	Name2	Beta	16	32	51	1	
	Name3	Alpha	52	104	32	0	
	Name4	Beta	36	72	84	1	
	Name5	Beta	45	90	32	0	
	Name6	Alpha	12	24	12	0	
	Name7	Beta	32	64	64	1	
	Name8	Alpha	42	84	54	0	
	Name9	Alpha	56	112	31	1	
31							

	Dataframe **We wish to calculate the correlation between "Column2" and "Column3" of a "dataframe". Which of the below codes will achieve the purpose?** A) corr(dataframe$column2,dataframe$column3) B) (cov(dataframe$column2,dataframe$column3))/(var(dataframe$column2)*sd(dataframe$column3)) C) (sum(dataframe$Column2*dataframe$Column3)- (sum(dataframe$Column2)*sum(dataframe$Column3)/nrow(dataframe)))/(sqrt((sum(dataframe$Column2*dataframe$Column2)- (sum(dataframe$Column2)^3)/nrow(dataframe))* (sum(dataframe$Column3*dataframe$Column3)-(sum(dataframe$Column3)^2)/nrow(dataframe)))) D) None of the Above	
32	**Parameter** **State** **Value** **Deper** Alpha Active 50 2 Beta Active 45 5 Beta Passive 25 0 Alpha Passive 21 0 Alpha Passive 26 1 Beta Active 30 2 Beta Passive 18 0 Dataframe **The above dataset has been loaded for you in R in a variable named "dataframe" with first row representing the column name. Which of the following code will select only the rows for which parameter is Alpha?** A) subset(dataframe, Parameter='Alpha') B) subset(dataframe, Parameter=='Alpha') C) filter(dataframe,Parameter=='Alpha') D) Both 2 and 3 E) All of the above	**D**
33	**15) Which of the following function is used to view the dataset in spreadsheet like format?** A) disp() B) View() C) seq() D) All of the Above	**B**

34	**The below dataframe is stored in a variable named data.** **A** **B** 1 Right 2 Wrong 3 Wrong 4 Right 5 Right 6 Wrong 7 Wrong 8 Right Data **Suppose B is a categorical variable and we wish to draw a boxplot for every level of the categorical level. Which of the below commands will help us achieve that?** A) boxplot(A,B,data=data) B) boxplot(A~B,data=data) C) boxplot(A|B,data=data) D) None of the above	**B**
35	**Which of the following commands will split the plotting window into 4 X 3 windows and where the plots enter the window column wise.** A) par(split=c(4,3)) B) par(mfcol=c(4,3)) C) par(mfrow=c(4,3)) D) par(col=c(4,3))	**B**
36	**A Dataframe "df" has the following data:** **Dates** 2017-02-28 2017-02-27 2017-02-26 2017-02-25 2017-02-24 2017-02-23	**D**

	2017-02-22 2017-02-21 After reading above data, we want the following output: **Dates** 28 Tuesday Feb 17 27 Monday Feb 17 26 Sunday Feb 17 25 Saturday Feb 17 24 Friday Feb 17 23 Thursday Feb 17 22 Wednesday Feb 17 21 Tuesday Feb 17 **Which of the following commands will produce the desired output?** A) format(df,"%d %A %b %y") B) format(df,"%D %A %b %y") C) format(df,"%D %a %B %Y") D) None of above	
37	**Which of the following command will help us to rename the second column in a dataframe named "table" from alpha to beta?** A) colnames(table)[2]='beta' B) colnames(table)[which(colnames=='alpha')]='beta' C) setnames(table,'alpha','beta') D) All of the above	**D**
38	**A majority of work in R uses systems internal memory and with large datasets, situations may arise when the R workspace cannot hold all the R objects in memory. So removing the unused objects is one of the solution. Which of the following command will remove an R object / variable named "santa" from the workspace?** A) remove(santa) B) rm(santa) C) Both D) None	**C**
39	**"dplyr" is one of the most popular package used in R for manipulating data and it contains 5 core functions to handle data. Which of the following is not one of the core functions of dplyr package?** A) select() B) filter() C) arrange() D) summary()	**D**
40	**During Feature Selection using the following dataframe (named table), "Column1" and "Column2" proved to be non-significant. Hence, we would not like to take these two features into our predictive model.** **Column1 Column2 Column3 Column4 Column5** **Name1** Alpha 12 24 54 0	**D**

	Name2	Beta	16	32	51	1	Beta
	Name3	Alpha	52	104	32	0	Gamma
	Name4	Beta	36	72	84	1	Delta
	Name5	Beta	45	90	32	0	Phi
	Name6	Alpha	12	24	12	0	Zeta
	Name7	Beta	32	64	64	1	Sigma
	Name8	Alpha	42	84	54	0	Mu
	Name9	Alpha	56	112	31	1	Eta
	Table **Which of the following commands will select all the rows from column 3 to column 6 for the below dataframe named table?** A) dplyr::select(table,Column3:Column6) B) table[,3:6] C) subset(table,select=c('Column3','Column4','Column5','Column6')) D) All of the above						
41		**Column1**	**Column2**	**Column3**	**Column4**	**Column5**	**C**
	Name1	Alpha	12	24	54	0	
	Name2	Beta	16	32	51	1	
	Name3	Alpha	52	104	32	0	
	Name4	Beta	36	72	84	1	
	Name5	Beta	45	90	32	0	
	Name6	Alpha	12	24	12	0	
	Name7	Beta	32	64	64	1	

	Name8 Alpha 42 84 54 0 **Name9** Alpha 56 112 31 1 table **Which of the following commands will select the rows having "Alpha" values in "Column1" and value less than 50 in "Column4"? The dataframe is stored in a variable named table.** A) dplyr::filter(table,Column1==’Alpha’, Column4<50) B) dplyr::filter(table,Column1==’Alpha’ & Column4<50) C) Both of the above D) None of the above	Mu Eta
42	**Column1 Column2 Column3 Column4 Column5** **Name1** Alpha 12 24 54 0 **Name2** Beta 16 32 51 1 **Name3** Alpha 52 104 32 0 **Name4** Beta 36 72 84 1 **Name5** Beta 45 90 32 0 **Name6** Alpha 12 24 12 0 **Name7** Beta 32 64 64 1 **Name8** Alpha 42 84 54 0 **Name9** Alpha 56 112 31 1 Table **Which of the following code will sort the dataframe based on "Column2" in ascending order and "Column3" in descending order?** A) dplyr::arrange(table,desc(Column3),Column2) B) table[order(-Column3,Column2),] C) Both of the above D) None of the above	**C**
43	**What will be the output of the following command** grepl("neeraj",c("dheeraj","Neeraj","neeraj","is","NEERAJ")) A) [FALSE TRUE TRUE FALSE TRUE]	**B**

	B) [FALSE TRUE TRUE FALSE FALSE] C) [FALSE FALSE TRUE FALSE FALSE] D) None of the above	
44	**Sometimes as a Data Scientist working on textual data we come across instances where we find multiple occurrences of a word which is unwanted. Below is one such string.** A<-c("I can use because thrice in a sentence because because is a special word.") A) gsub("because","since",A) B) sub("because","since",A C) regexec("because","since",A) D) None of the above	**C**
45	**Imagine a dataframe created through the following code.** **Which of the following command will help us remove the duplicate rows based on both the columns?** A) df[!duplicated(df),] B) unique(df) C) dplyr::distinct(df) D) All of the above	**A**
46	**Grouping is an important activity in Data Analytics and it helps us discover some interesting trends which may not be visible easily in the raw data. Suppose you have a dataset created by the following lines of code.** table<-data.table(foo=c("A","B","A","A","B","A"),bar=1:6) **Which of the following command will help us to calculate the mean bar value grouped by foo variable?** A) aggregate(bar~foo,table,mean) B) table::df[,mean(bar),by=foo] C) dplyr::table%>%group_by(foo)%>%summarize(mean=mean(bar)) D) All of the above	**D**
47	**Dealing with strings is an important part of text analytics and splitting a string is often one of the common task performed while creating tokens, etc. What will be the output of following commands?** A<-paste("alpha","beta","gamma",sep=" ") B←paste("phi","theta","zeta",sep="") parts←strsplit(c(A,B),split=" ") A) alpha B) beta C) gamma D) phi E) theta F) zeta	**D**
48	**If I have two vectors x<- c(1,3, 5) and y<-c(3, 2), what is produced by the expression cbind(x, y)?** A) a matrix with 2 columns and 3 rows B) a matrix with 3 columns and 2 rows C) a data frame with 2 columns and 3 rows D) a data frame with 3 columns and 2 rows	**D**
49	**Which of the following commands will convert the following dataframe named maverick into the one shown at the bottom?**	**A**

	Input Dataframe – "maverick" **Grade** **Male** **Female** A 10 15 B 20 15 A 30 35 **Output dataframe** **Grade** **Sex** **Count** A Male 10 A Female 15 B Male 30 B Female 15 A Male 30 A Female 35 A) tidyr::Gather(maverick, Sex,Count,-Grade) B) tidyr::spread(maverick, Sex,Count,-Grade C) tidyr::collect(maverick, Sex,Count,-Grade) D) None of the above	
50	**Which of the following command will help us to replace every instance of Delhi with Delhi_NCR in the following character vector?** C<-c("Delhi is","a great city.","Delhi is also","the capital of India.") A) gsub("Delhi","Delhi_NCR",C) B) sub("Delhi","Delhi_NCR",C) C) Both of the above D) None of the above	**C**
	Unit -4	
1	1. R has how many atomic classes of objects? a) 1 b) 2 c) 3 d) 5	**C**
2	Point out the correct statement?	**D**

	a) Empty vectors can be created with the vector() function b) A sequence is represented as a vector but can contain objects of different classes c) "raw" objects are commonly used directly in data analysis d) The value NaN represents undefined value	
3	Numbers in R are generally treated as _____ precision real numbers. a) single b) double c) real d) imaginary	
4	If you explicitly want an integer, you need to specify the ____ suffix. a) D b) R c) L d) K	
5	R is an_____programming language? a) Closed source b) GPL c) Open source d) Definite source	**C**
6	.Solve varx<-23, 34->vary print(varx+vary) a. 57 b. 2334 c. 3423 d. 66	**A**
7	find the output varx<-23, 34->vary print(varx == vary) a. True b. False c. None of the above d. Error	**B**
8	Below, we have represented six data points on a scale where vertical lines on scale represent unit. Which of the following line represents the mean of the given data points, where the scale is divided into same units? A) A B) B C) C D) D	**C**
9	If a positively skewed distribution has a median of 50, which of the following statement is true? A) Mean is greater than 50 B) Mean is less than 50	**E**

	C) Mode is less than 50 D) Mode is greater than 50 E) Both A and C F) Both B and D	
10	Which of the following is a possible value for the median of the below distribution? A) 32 B) 26 C) 17 D) 40	**B**
11	Which of the following statements are true about Bessels Correction while calculating a sample standard deviation? Bessels correction is always done when we perform any operation on a sample data. Bessels correction is used when we are trying to estimate population standard deviation from the sample. Bessels corrected standard deviation is less biased. A) Only 2 B) Only 3 C) Both 2 and 3 D) Both 1 and 3	**C**
12	If the variance of a dataset is correctly computed with the formula using (n – 1) in the denominator, which of the following option is true? A) Dataset is a sample B) Dataset is a population C) Dataset could be either a sample or a population D) Dataset is from a census E) None of the above	**A**
13	What would be the critical values of Z for 98% confidence interval for a two-tailed test ? A) +/- 2.33 B) +/- 1.96 C) +/- 1.64 D) +/- 2.55	**A**
14	Studies show that listening to music while studying can improve your memory. To demonstrate this, a researcher obtains a sample of 36 college students and gives them a standard memory test while they listen to some background music. Under normal circumstances (without music), the mean score obtained was 25 and standard deviation is 6. The mean score for the sample after the experiment (i.e With music) is 28. What is the null hypothesis in this case? A) Listening to music while studying will not impact memory. B) Listening to music while studying may worsen memory. C) Listening to music while studying may improve memory. D) Listening to music while studying will not improve memory but can make it worse.	**D**
15	Studies show that listening to music while studying can improve your memory. To demonstrate this, a researcher obtains a sample of 36 college students and gives them a standard memory test while they listen to some background music. Under normal circumstances (without music), the mean score obtained was 25 and standard deviation is 6. The mean score for the sample after the experiment (i.e With music) is	**B**

	28. What would be the Type I error? A) Concluding that listening to music while studying improves memory, and it's right. B) Concluding that listening to music while studying improves memory when it actually doesn't. C) Concluding that listening to music while studying does not improve memory but it does.	
16	Studies show that listening to music while studying can improve your memory. To demonstrate this, a researcher obtains a sample of 36 college students and gives them a standard memory test while they listen to some background music. Under normal circumstances (without music), the mean score obtained was 25 and standard deviation is 6. The mean score for the sample after the experiment (i.e With music) is After performing the Z-test, what can we conclude ___ ? A) Listening to music does not improve memory. B)Listening to music significantly improves memory at p C) The information is insufficient for any conclusion. D) None of the above	**B**
17	A researcher concludes from his analysis that a placebo cures AIDS. What type of error is he making? A) Type 1 error B) Type 2 error C) None of these. The researcher is not making an error. D) Cannot be determined	**D**
18	What happens to the confidence interval when we introduce some outliers to the data? A) Confidence interval is robust to outliers B) Confidence interval will increase with the introduction of outliers. C) Confidence interval will decrease with the introduction of outliers. D) We cannot determine the confidence interval in this case	**B**
19	A medical doctor wants to reduce blood sugar level of all his patients by altering their diet. He finds that the mean sugar level of all patients is 180 with a standard deviation of 18. Nine of his patients start dieting and the mean of the sample is observed to 175. Now, he is considering to recommend all his patients to go on a diet. Note: He calculates 99% confidence interval. What is the standard error of the mean? A) 9 B) 6 C) 7.5 D) 18	**B**
20	----------is function in R to get number of observation in a data frame a) n( ) b) ncol( ) c) nobs( ) d) nrow( )	**D**

21	A key property of vectors in R language is that a. A vector cannot have attributes like dimensions b. Elements of a vector can be of different classes c. Elements of a vector can only be a character or numeric d. Elements of a vector all must be of the same class	**D**
22	The definition of free software consists of four freedoms (freedoms 0 through 3). Which of the following is NOT one of the freedoms that are part of the definition? a. The freedom to study how the program works, and adapt it to your needs. b. The freedom to improve the program, and release your improvements to the public, so that the whole community benefits. c. The freedom to run the program, for any purpose. d. The freedom to sell the software for any price.	
23	Point out the correct statement : a) Blocks are evaluated until a new line is entered after the closing brace b) Single statements are evaluated when a new line is typed at the start of the syntactically complete statement c) The if/else statement conditionally evaluates two statements d) All of the mentioned	**C**
24	Which will be the output of following code ? x - 3 switch(6, 2+2, mean(1:10), rnorm(5)) a) 10 b) 1 c) NULL d) All of the mentioned	**C**
25	_____ is used to continue an iteration of a loop. A. next B. skip C. group D. All of the mentioned	**A**
26	Point out the correct statement : a) R has a number of ways to indicate to you that something's not right b) Executing any function in R may result in the condition c) "condition" is a generic concept for indicating that something unexpected	**D**

	has occurred d) All of the mentioned	
27	. Which of the following is primary tool for debugging ? a) debug() b) trace() c) browser() d) All of the mentioned	
28	Point out the correct statement : a) Vectorizing the function can be accomplished easily with the Vectorize() function b) There are different levels of indication that can be used, ranging from mere notification to fatal error c) Vectorizing the function can be accomplished easily with the vector() function d) None of the mentioned	**A**
29	Functions are defined using the _____ directive and are stored as R objects a) function() b) funct() c) functions() d) All of the mentioned	**A**
30	The _____ function returns a list of all the formal arguments of a function a) formals() b) funct() c) formal() d) All of the mentioned	**A**
31	Which of the following is multivariate version of lapply ? a) apply() b) lapply() c) sapply() d) mapply()	**D**
32	Point out the correct statement : a) split() takes elements of the list and passes them as the first argument of the function you are applying b) You can use tsplit() to evaluate a function single time each with a same argument	**C**

	c) Sequence of operations is sometimes referred to as “map-reduce” d) None of the mentioned	
33	A function, together with an environment, makes up what is called a _____ closure. a) formal b) function c) reflective d) All of the mentioned	**B**
34	The _____ function is used to plot negative likelihood. a) plot() b) graph() c) graph.plot() d) None of the mentioned	**A**
	Unit-5	
1	____ is a subset of ____ a) Information design, visual modality b) Information design, data visualization c) None of the answers are correct. d) Data visualization, information design	**A**
2	Which of the answers is an example of the kinesthetic modality? a) A speech b) A movie c) A picture d) The rain on our face	**B**
3	What area represents information in a graphical or pictorial form? a) Data design b) None of the answers are correct. c) Information design d) Data visualization	**C**
4	Which of the following is an example of a temporal data visualization? a) A Gnatt chart that is use in project management b) A histogram that represents proportions c) A matrix representing interconnecting data among various entities	**d**

	d) A 3D molecular rendering of a protein a)	
5	By definition, Tableau displays measures over time as a _____ a) Bar b) Line c) Histogram d) Scatter Plots	**D**
6	How do you identify a continuous field in Tableau? a) It is identified by a blue pill in the visualization b) It is identified by a green pill in a visualization c) It is preceded by a # symbol in the data window d) When added to the visualization, it produces distinct values	**A**
7	For creating variable size bins we use _____ a) Sets b) Groups c) Calculated fields d) Table Calculations	**B**
8	Which of the following is not a Trend Line model a) Linear Trend Line b) Exponential Trend Line c) Binomial Trend Line d) Logarithmic Trend Line	**C**
9	Data cleaning consists primarily in implementing ……..strategies before they occur a) error prevention b) error detection c) indicating error d) none of the above	**A**
10	data errors will be detected incidentally during activities a) When collecting or entering data b) When transforming/extracting/transferring data c) When exploring or analysing data d) When submitting the draft report for peer review	**A**
11	Data cleaning involves repeated cycles of a) screening, b) diagnosing, c) treatment and d) documentation of this process. e) All the above	**E**
12	After measurement, ……..are the object of a sequence of typical activities: a) Data b) Information , c) Record d) None of the above	**C**

13	1. Under the lattice graphics system, what do the primary plotting functions like xyplot() and bwplot() return? a) nothing; only a plot is made b) an object of class "lattice" c) an object of class "trellis" d) an object of class "plot"	**A**

14	What is produced by the following code?
	`library(nlme)`
	`library(lattice)`
	`xyplot(weight ~ Time | Diet, BodyWeight)`
	a) A set of 16 panels showing the relationship between weight and time for each rat.
	b) A set of 3 panels showing the relationship between weight and time for each diet.
	c) A set of 11 panels showing the relationship between weight and diet for each time.
	d) A set of 3 panels showing the relationship between weight and time for each rat…	

15	Which of the following functions can be used to annotate the panels in a multi-panel lattice plot? a) axis() b) text() c) panel.abline() d) points() e) lines()	**B**
16	In the lattice system, which of the following functions can be used to finely control the appearance of all lattice plots? a) par() b) splom() c) print.trellis() d) trellis.par.set()	**B**
17	What is ggplot2 an implementation of? a) a 3D visualization system b) the Grammar of Graphics developed by Leland Wilkinson c) the base plotting system in R d) the S language originally developed by Bell Labs	**C**
18	What is a geom in the ggplot2 system? • a method for mapping data to attributes like color and size • a method for making conditioning plots • a statistical transformation • a plotting object like point, line, or other shape	**A**
19	The following code creates a scatterplot of 'votes' and 'rating' from the movies dataset in the ggplot2 package. After loading the ggplot2 package with the library() function, I can run	**A**

	qplot(votes, rating, data = movies) How can I modify the the code above to add a smoother to the scatterplot? a) qplot(votes, rating, data = movies) + stats_smooth("loess") b) qplot(votes, rating, data = movies, panel = panel.loess) c) qplot(votes, rating, data = movies, smooth = "loess") d) qplot(votes, rating, data = movies) + geom_smooth()	
20	When I run the following code I get an error: library(ggplot2) library(ggplot2movies) g <- ggplot(movies, aes(votes, rating)) print(g) I was expecting a scatterplot of 'votes' and 'rating' to appear. What's the problem? a) The dataset is too large and hence cannot be plotted to the screen. b) There is a syntax error in the call to ggplot. c) ggplot does not yet know what type of layer to add to the plot. d) The object 'g' does not have a print method	**B**
21	**Sometimes creating a feature which represents whether another variable has missing values or not can prove to be very useful for a predictive model. Below is a dataframe which has missing values in one of its columns.** **Feature1** **Feature2** B NA C 30 D 40 E 50 **Which of the following commands will create a column named "missing" with value 1 where variable "Feature2" has missing values?** **Feature1** **Feature2** **Missing** B NA 1 C 30 0 D 40 0	**C**

	E 50 0 A) dataframe$missing<-0 dataframe$Missing[is.na(dataframe$Feature2)]<-1 B) dataframe$missing<-0 dataframe$Missing[which(is.na(dataframe$Feature2))]<-1 C) Both of the above D) None of the above	
22	**Suppose there are 2 dataframes "A" and "B". A has 34 rows and B has 46 rows. What will be the number of rows in the resultant dataframe after running the following command?** merge(A,B,all.x=TRUE) A) 46 B) 12 C) 34 D) 80	**C**
23	**The very first thing that a Data Scientist generally does after loading dataset is find out the number of rows and columns the dataset has. In technical terms, it is called knowing the dimensions of the dataset. This is done to get an idea about the scale of data that he is dealing with and subsequently choosing the right techniques and tools.** **Which of the following command will not help us to view the dimensions of our dataset?** A) dim() B) str() C) View() D) None of the above	**C**
24	**Sometimes, we face a situation where we have two columns of a dataset and we wish to know which elements of the column are not present in another column. This is easily achieved in R using the setdiff command.** **Column1** Alpha 12 24 54 0 **Name1** **Name2** Beta 16 32 51 1 **Name3** Alpha 52 104 32 0 **Name4** Beta 36 72 84 1	**C**

	Name5	Beta	45	90	32	0	Phi
	Name6	Alpha	12	24	12	0	Zeta
	Name7	Beta	32	64	64	1	Sigma
	Name8	Alpha	42	84	54	0	Mu
	Name9	Alpha	56	112	31	1	Eta

Dataframe
What will be the output of the following command?
setdiff(dataframe$Column1,dataframe$Column6)==setdiff(dataframe$Column6,dataframe$Column1)
A) TRUE
B)FALSE
C) Can't Say

B

25

The below dataset is stored in a variable called **"frame"**.

A	**B**
alpha	100
beta	120
gamma	80
delta	110

Which of the following commands will create a bar plot for the above dataset. Use the values from Column B to represent the height of the bar plot.
A) ggplot(frame,aes(A,B))+geom_bar(stat="identity")
B) ggplot(frame,aes(A,B))+geom_bar(stat="bin")
C) ggplot(frame,aes(A,B))+geom_bar()
D) None of the above

A

26

A	**mpg**	**cyl**	**disp**	**hp**	**drat**	**wt**	**qsec**	**vs**	**am**	**gear**
Mazda	21.	6	160	110	3.9	2.62	16.4	0	1	4

	RX4	0				0	0	6				
	Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
	Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
	Hornet Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
	Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
	Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1
	We wish to create a stacked bar chart for cyl variable with stacking criteria Being vs Variable. Which of the following commands will help us perform this action? A)qplot(factor(cyl),data=mtcars,geom="bar",fill=factor(vs) B) ggplot(mtcars,aes(factor(cyl),fill=factor(vs)))+geom_bar() C) All of the above D) None of the above											
27	**What is the output of the command – paste(1:3,c("x","y","z"),sep="") ?** A) [1 2 3x y z] B) [1:3x y z] C) [1x 2y 3z] D) None of the above											**C**
28	R has a rich library reserve for drawing some of the very high end graphs and plots and many a times you want to save the graphs for presenting your findings to someone else. Saving your plots to a PDF file is one such option. If you want to save a plot to a PDF file, which of the following is a correct way of doing that? A) Construct the plot on the screen device and then copy it to a PDF file with dev.copy2pdf(). B) Construct the plot on the PNG device with png(), then copy it to a PDF with dev.copy2pdf(). C) Open the PostScript device with postscript(), construct the plot, then close the device with dev.off(). D) Open the screen device with quartz(), construct the plot, and then close the device with dev.off().											**C**

29	Given $X_1=12,X_2=19,X_3=10,X_4=7$, then $\sum_{i=1}^4 X_i$ equals? a) 36 b) 48 c) 29 d) 41	**B**
30	The number of accidents in a city during 2010 is a) Discrete variable b) Continuous variable c) Qualitative variable d) Constant	**A**
31	The mean of a distribution is 23, the median is 24, and the mode is 25.5. It is most likely that this distribution is: a) Positively Skewed b) Symmetrical c) Asymptotic **d)** Negatively Skewed	**A**
32	Data collected by NADRA to issue computerized identity cards (CICs) are a) Unofficial data b) Qualitative data c) Secondary data d) Primary data **e)** None of these	**C**
33	Sum of dots when two dice are rolled is a) A discrete variable b) A continuous variable c) A constant **d)** A qualitative variable	**A**
34	A chance variation in an observational process is a) Dispersion/ Variability b) Measurement error c) Random error **d)** Instrument error	**C**
35	**If a distribution is abnormally tall and peaked, then is can be said that the distribution is:** **a) Leptokurtic** **b) Pyrokurtic** **c) Platykurtic** **d) Mesokurtic**	**A**
36	The mean of a distribution is 14 and the standard deviation is 5. What is the value of the coefficient of variation? a) 60.4% b) 48.3% c) 35.7% **d)** 27.8%	**C**
37	The first hand and unorganized form of data is called a) Secondary data b) Organized data	**C**

	c) Primary data **d)** None of these	
38	Questionnaire survey method is used to collect a) Secondary data b) Qualitative variable c) Primary data **d)** None of these	
39	The data which have already been collected by someone are called a) Raw data b) Array data c) Secondary data **d)** Fictitious data	**C**
40	The grouped data is also called a) Raw data b) Primary data c) Secondary data d) Qualitative data	**C**
41	A constant variable can take values a) Zero b) Fixed c) Not fixed d) Nothing	**B**
42	A parameter is a measure which is computed from a) Population data b) Sample data c) Test statistics **d)** None of these	**A**
43	According to the empirical rule, approximately what percent of the data should lie within $\mu \pm \sigma$? a) 75% b) 68% c) 99.7% d) 90% **e)** 95%	**E**
44	Primary data and _____ data are same a) Grouped b) Secondary data c) Ungrouped d) None of these	**C**
45	Which one of the following measurement does not divide a set of observations into equal parts? a) Quartiles b) Standard Deviations c) Percentiles d) Deciles **e)** Median	**B**
46	In descriptive statistics, we study a) The description of the decision-making process b) The methods for organizing, displaying and describing data	**A**

	c) How to describe the probability distribution **d)** None of the above	
47	Which of the following is not based on all the observations? a) Arithmetic Mean b) Geometric Mean c) Harmonic Mean d) Weighted Mean **e)** Mode	**E**
48	Which one is the not measure of dispersion. a) The Range b) 50th Percentile c) Inter-Quartile Range d) Variance	**B**
49	When data are collected in a statistical study for only a portion or subset of all elements of interest we are using: a) A sample b) A Parameter c) A Population **d)** Both b and c	**A**
50	In inferential statistics, we study a) The methods to make decisions about the population based on sample results b) How to make decisions about mean, median, or mode c) How a sample is obtained from a population **d)** None of the above	**A**