

Name of the Teacher: Ms. P. S. Patil

**Class: BE
AY: 2020-21**

**Subject: Data Analytics
SEM: II**

UNIT-1

1) What is Big Data?

- a) Huge amount of data
- b) Small amount of data
- c) Huge File
- d) Big Storage

Ans:

a

Explanation: It is Huge amount of data

2) According to analysts, for what can traditional IT systems provide a foundation when they're integrated with big data technologies like Hadoop?

- a) Big data management and data mining
- b) Data warehousing and business intelligence
- c) Management of Hadoop clusters
- d) Collecting and storing unstructured data

Ans:

a

Explanation: Big data management and data mining

3) What are the main components of Big Data?

- a) MapReduce
- b) HDFS
- c) YARN
- d) All of these

Ans:

d

Explanation: All of these

4) The sources of Big Data are

- a) Stock Exchange
- b) Transport Data
- c) Banking Data
- d) All of the Above

Ans:

d

Explanation:

5) Big Data Characteristics are:

- a) Structured data
- b) Semi-structured data
- c) Quasi-structured data
- d) All of the above

Ans:

d

Explanation:

6) BI tends to provide reports, dashboards, and queries on business

	questions for the current period or in the past.
	a) True b) False
Ans:	a
Explanation:	
7)	Big data can come in multiple forms, including structured and non-structured data
	a) True b) False
Ans:	a
Explanation:	
8)	BI problems tend to require highly structured data organized
	a) Rows b) Columns c) Accurate Reporting d) All of the Above
Ans:	d
Explanation:	
9)	EDW achieves the objective of reporting and sometimes the creation of dashboards, perform analysis on unstructured data
	a) High-value data is hard to reach and leverage b) Data moves in batches from EDW to local analytical tools c) Data Science projects will remain isolated d) All of the Above
Ans:	d
Explanation:	
10)	Drivers of Big Data
	a) Medical information b) Photos and video footage uploaded to the World Wide Web c) data extracts d) Both a and b
Ans:	d
Explanation:	
11)	According to analysts, for what can traditional IT systems provide a foundation when they're integrated with big data technologies like Hadoop?
	a) Big data management and data mining b) Data warehousing and business intelligence c) Management of Hadoop clusters d) Collecting and storing unstructured data
Ans:	a
Explanation:	
12)	Select from option which is not the phase of data analytics

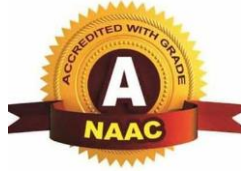
	<ul style="list-style-type: none"> a) model planning b) testing c) discovery d) operationalize
Ans:	b
Explanation:	
13)	Which phase of data analytics require more time to complete
	<ul style="list-style-type: none"> a) Data preparation b) model building c) communicate results d) Discovery
Ans:	a
Explanation:	
14)	What is analytic sandbox?
	<ul style="list-style-type: none"> a) Tool b) Separate repository c) data cleaning d) Data conditioning
Ans:	b
Explanation:	
15)	The person which provides analytic techniques and modeling is called as.
	<ul style="list-style-type: none"> a) Data Engineer b) Data scientist c) Business user d) Project manager
Ans:	b
Explanation:	
16)	What is task of Project manager?
	<ul style="list-style-type: none"> a) analytic modelling b) Provide requirement c) ensure meeting objectives d) creates DB environment
Ans:	c

Explanation:	
17)	Identifying Key Stakeholders this task is performed in which phase?
	a) Data preparation b) model building c) Discovery d) communicate results
Ans:	c
Explanation:	
18)	ETL process is performed in which phase
	a) Discovery b) communicate results c) model planning d) Data preparation
Ans:	d
Explanation:	
19)	How much data Data science teams prefer for analysis?
	a) too little b) average c) more d) more than average
Ans:	c
Explanation:	
20)	select from option tool which is not used in model planning phase
	a) Data wrangler b) R c) SQL Analysis service d) SAS/ACCESS
Ans:	c
Explanation:	

21)	if reports and dashboards will be impacted and need to change this task is performed by.
	<ul style="list-style-type: none"> a) Project sponsor b) BI Analyst c) Data Engineer d) Project manager
Ans:	b
Explanation:	
22)	What is need of data analytic lifecycle.
	<ul style="list-style-type: none"> a) Data cleaning b) To solve Big data problems c) Data conditioning d) Data Exploration
Ans:	b
Explanation:	
23)	How many phases are there in data analytic lifecycle?
	<ul style="list-style-type: none"> a) 4 b) 5 c) 6 d) 7
Ans:	c
24)	The person with technical skills is called as?
	<ul style="list-style-type: none"> a) Business user b) Data Engineer c) Data scientist d) Project sponsor
Ans:	b
25)	What is outcome of Model building phase?
	<ul style="list-style-type: none"> a) Analytic results b) Quality data c) Data d) Potential resources
Ans:	a



ZEAL EDUCATION SOCIETY'S
ZEAL COLLEGE OF ENGINEERING AND RESEARCH
NARHE | PUNE -41 | INDIA
DEPARTMENT OF COMPUTER ENGINEERING



Pravin S.Patil

Subject Teacher

Name of the Teacher: Ms. P. S. Patil

Class: BE
AY: 2020-21

Subject: Data Analytics
SEM: II

UNIT-1I

1)	1. A statement made about a population for testing purpose is called?
	a) Statistic b) Hypothesis c) Level of Significance d) Test-Statistic
Ans:	b
Explanation:	
2)	If the assumed hypothesis is tested for rejection considering it to be true is called?
	a) Null Hypothesis b) Statistical Hypothesis c) Simple Hypothesis d) Composite Hypothesis
Ans:	a
Explanation:	
3)	A statement whose validity is tested on the basis of a sample is called?
	a) Null Hypothesis b) Statistical Hypothesis c) Simple Hypothesis d) Composite Hypothesis
Ans:	b
Explanation:	
4)	A hypothesis which defines the population distribution is called?
	a) Null Hypothesis b) Statistical Hypothesis c) Simple Hypothesis d) Composite Hypothesis
Ans:	c
Explanation:	
5)	If the null hypothesis is false then which of the following is accepted?
	a) Null Hypothesis b) Positive Hypothesis c) Negative Hypothesis d) Alternative Hypothesis.
Ans:	d
Explanation:	

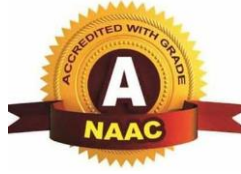
6)	The rejection probability of Null Hypothesis when it is true is called as?
	a) Level of Confidence b) Level of Significance c) Level of Margin d) Level of Rejection
Ans:	b
Explanation:	
7)	The point where the Null Hypothesis gets rejected is called as?
	a) Significant Value b) Rejection Value c) Acceptance Value d) Critical Value
Ans:	d
Explanation:	
8)	If the Critical region is evenly distributed then the test is referred as?
	a) Two tailed b) One tailed c) Three tailed d) Zero tailed
Ans:	a
Explanation:	
9)	The type of test is defined by which of the following?
	a) Null Hypothesis b) Simple Hypothesis c) Alternative Hypothesis d) Composite Hypothesis
Ans:	c
Explanation:	
10)	Which of the following is defined as the rule or formula to test a Null Hypothesis?
	a) Test statistic b) Population statistic c) Variance statistic d) Null statistic
Ans:	a
Explanation:	
11)	Type 1 error occurs when?
	a) We reject H_0 if it is True b) We reject H_0 if it is False c) We accept H_0 if it is True d) We accept H_0 if it is False
Ans:	a
Explanation:	
12)	The probability of Type 1 error is referred as?
	a) $1-\alpha$

	b) β c) α d) $1-\beta$
Ans:	c
Explanation:	
13)	Alternative Hypothesis is also called as?
	a) Composite hypothesis b) Research Hypothesis c) Simple Hypothesis d) Null Hypothesis
Ans:	b
Explanation:	
14)	Which of the following is required by K-means clustering?
	a) defined distance metric b) number of clusters c) initial guess as to cluster centroids d) all of the mentioned
Ans:	d
Explanation:	
15)	Point out the wrong statement.
	a) k-means clustering is a method of vector quantization b) k-means clustering aims to partition n observations into k clusters c) k-nearest neighbor is same as k-means d) none of the mentioned
Ans:	c
Explanation:	
16)	Hierarchical clustering should be primarily used for exploration.
	a) True b) False
Ans:	a
Explanation:	
17)	Which of the following function is used for k-means clustering?
	a) k-means b) k-mean c) heatmap d) none of the mentioned
Ans:	a
Explanation:	
18)	Which of the following clustering requires merging approach?
	a) Partitional b) Hierarchical c) Naive Bayes d) None of the mentioned
Ans:	b
Explanation:	

19)	K-means is not deterministic and it also consists of number of iterations.
	a) True b) False
Ans:	a
20)	Depending on acceptance and rejection of null hypothesis there are 2 types of error produced
	a) Type 1 b) Type 2 c) None of these d) All of these
Ans:	d
21)	The power of a test can be defined as a possibility of ...
	a) Rejecting null hypothesis b) Accepting null hypothesis c) Increasing null hypothesis d) Decreasing null hypothesis
Ans:	a
22)	For a fixed significance level, a greater sample size is mandatory to discover a
	a) Minor difference in mean b) Major difference in mean c) Average difference in mean d) None of the above
Ans:	a
23)	ANNOVA tests if any of the population means vary from other population means
	a) True b) False
Ans:	a
24)	Clustering is defined as group of same kind of objects which are gathered by use of
	a) Unsupervised method b) Supervised method c) Semi supervised method d) None of these
Ans:	a
25)	Following are the applications of Kmeans
	a) Image Processing b) Medical c) Customer Segmentation d) All of the above



ZEAL EDUCATION SOCIETY'S
ZEAL COLLEGE OF ENGINEERING AND RESEARCH
NARHE | PUNE -41 | INDIA
DEPARTMENT OF COMPUTER ENGINEERING



Ans:

d

Pravin S.Patil

Subject Teacher