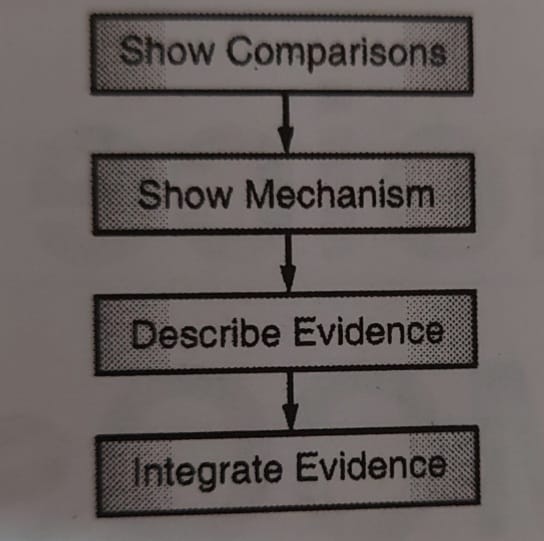
Unit 1

Q1. Which of the following represented in the below figure?



(a) Show Comparisons

(b) Integrate Evidence

(c) Describe Evidence

**(d) None of the mentioned**

Q2. Point out correct statement.

**(a) Raw data is original source of data**

(b) Preprocessed data is original source of data

(c) Raw data is the data obtained after processing steps

(d) None of the mentioned

Q3. Which of the following is performed by Data Scientist?

(a) Define the question

(b) Create reproducible code

(c) Challenge results

**(d) All of the mentioned**

Q 4. Point out the correct statement.

**(a) Least square is an estimation tool**

(b) Least square problems falls in to three categories

(c) Compound least square is one of the category of least square

(d) None of the mentioned

Q 5. Which of the following is the most important language for Data Science?

(a) Java

(b) Ruby

**(c) R**

(d) None of the mentioned

Q.6 Point out the wrong statement

(a) Merging concerns combining dataset on the same observations to produce a result with more variables

**(b) Data visualization is the organization of information according to preset specifications.**

(c)Subsetting can be used to select and exclude variables and observations

(d) All of the mentioned

Q 7. Which of the following approach should be used to ask Data Analysis question?

(a) Find only one solution for particular problem

**(b) Find out the question which is to be answered**

(c) Find out answer from dataset without asking question

(d) None of the mentioned

Q 8. Which of the following is one of the key data science skills?

(a) Statistics

(b) Machine Learning

(c) Data Visualization

**(d) All of the mentioned**

Q 9. Which of the following is a key characteristic of a hacker?

(a) Afraid to say they don't know the answer

**(b) Willing to find answers on their own**

(c) Not Willing to find answers on their own

(d) All of the mentioned

Q.10. Which of the following is characteristic of Processed Data?

(a) Data is not ready for analysis

**(b) All steps should be noted**

(c) Hard to use for data analysis

(d) None of the mentioned

Q.11 Raw data should be processed only one time.

(a) True

**(b) False**

Q. 12 Point out the correct statement

(a) Machine learning focuses on prediction, based on known properties learned from the training data

(b) Data Cleaning focuses on prediction, based on known properties learned from the training data

(c) Representing data in a form which both mere mortals can understand and get valuable insights is as much at science as much as it is ort

**(d) None of the mentioned**

Q. 13 Which of the following characteristic of big data is relatively more concerned to data science?

(a) Velocity

**(b) Variety**

(c) Volume

(d) None of the mentioned

Q. 14 Which of the following analytical capabilities are provided by information management company?

(a) Stream Computing

(b) Content Management

(c) Information Integration

**(d) All of the mentioned**

Q.15 Point out the wrong statement.

(a) The big volume indeed represents Big Data

(b) The data growth and social media explosion have changed how we look at the data

**(c) Big Data is just about lots of data**

(d) All of the mentioned

Q. 16 Which of the following step is performed by data scientist after acquiring the data?

**(a) Data Cleansing**

(b) Data Integration

(c) Data Replication

(d) All of the mentioned

Q. 17 3V's are not sufficient to describe big data.

**(a) True**

(b) False

Q.18 Which of the following focuses on the discovery of (previously) unknown properties on the data?

**(a) Data mining**

(b) Big Data

(c) Data wrangling

(d) Machine Learning

Q.19 Beyond Volume, variety and velocity are the issues of big data veracity.

**(a) True**

(b) False

Q.20 Point out the correct statement.

**(a) If equations are known but the parameters are not, they may be inferred with data** **analysis**

(b) If equations are not known but the parameters are. they may be inferred with data analysis

(c) If equations and parameter are not, they may be inferred with data analysis

(d) None of the mentioned

Q.21 Which of the following is the top most important thing in data science?

(a) answer

**(b) question**

(c) data

(d) none of the mentioned

Q.22 Which of the following approach should be used if you can't fix the variable?

**(a) randomize it**

(b) non stratify it

(c) generalize it

(d) none of the mentioned

Q. 23 Point out the wrong statement.

**(a) Randomized studies are not used to identify causation**

(b) Complication approached exist for inferring causation

(c) Causal relationships may not apply to every individual

(d) All of the mentioned

Q.24 Which of the following is a good way of performing experiments in data science?

(a) Measure variability

(b) Generalize to the problem

(c) Have Replication

**(d) All of the mentioned**

Q.25 Which of the following is commonly referred to as 'data fishing'?

(a) Data bagging

(b) Data booting

(c) Data merging

**(d) None of the mentioned**

Q 26. Which of the following data mining technique is used to uncover patterns in data?

(a) Data bagging

(b) Data booting

(c) Data merging

**(d) Data Dredging**

Q 27 If X predicts Y, it does mean X causes Y.

(a) True

**(b) False**

Q 28 Point out the correct statement.

**(a) Time deltas are differences in times, expressed in difference units**

(b) You can construct a Time delta scalar through various argument

(c) Date Offsets cannot be used in construction

(d) All of the mentioned

Q. 29 How many principles of analytical graphs exist?

(a) 3

(b) 4

**(c) 6**

(d) None of the mentioned

Q. 30 Which of the following is not a step in data analysis?

(a) Obtain the data

(b) Clean the data

(c) EDA

**(d) None of the mentioned**

Q.31 Which of the following technique comes under practical machine learning?

(a) Bagging

**(b) Boosting**

(c) Forecasting

(d) None of the mentioned

Q. 32 The process of converting a continuous analog signal to a discrete digital signal is called?

(a) Discretisation

**(b) Sampling**

(c) Preemphasis

(d) Reduction

Q. 33 What is the error that occurs when the number of bits is not sufficient enough to represent the analog voltages?

(a) Data error

(b) Sampling error

(c) Acquisition error

**(d) Quantization error**

Q. 34 Pick the odd one out?

(a) R-2R converters

(b) String converters

(c) Weighted current source converters

**(d) Flash conversion**

Q. 35 Point out the correct statement.

**(a) If equations are known but the parameters are not, they may be inferred with data** **analysis**

(b) If equations are not known but the parameters are, they may be inferred with data analysis

(c) If equations and parameter are not, they may be inferred with data analysis

(d) None of the mentioned

Unit 2

Q. 2.1 The expected value or \_\_\_\_\_\_ of a random variable is

the center of its distribution.

(a) mode

(b) median

**(c) mean**

(d) bayesian inference

Q. 2.2 Which of the following of a random variable is a measure of spread?

**(a) variance**

(b) standard deviation

(c) empirical mean

(d) all of the mentioned

Q. 2.3 Point out the correct statement.

(a) Some cumulative distribution function F is non decreasing and right-continuous

(b) Every cumulative distribution function F is decreasing and right-continuous

(c) Every cumulative distribution function F is increasing and left-continuous

**(d) None of the mentioned**

Q. 2.4 Which of the following of a random variable is a measure of spread?

**(a) variance**

(b) standard deviation

(c) empirical mean

(d) all of the mentioned

Q. 2.5 The square root of the variance is called the \_\_\_\_\_\_\_\_\_ deviation.

(a) empirical

(b) mean

(c) continuous

**(d) standard**

Q.2.6 Point out the wrong statement.

(a) A percentile is simply a quantile with expressed as a percent

(b) There are two types of random variable

**(c) R cannot approximate quantiles for you for common distributions**

(d) None of the mentioned

Q.2.7 Which of the following inequality is useful for interpreting variances?

**(a) Chebyshev**

(b) Stautaory

(c) Testory

(d) All of the mentioned

Q. 2.8 For continuous random variables, the CDF is the derivative of the PDF.

(a) True

**(b) False**

Q. 2.9 Chebyshev's inequality states that the probability of a "Six Sigma" event is less than\_\_\_\_\_\_\_\_\_\_.

(a) 10%

(b) 20%

(c) 30%

**(d) 3%**

Q. 2.10 Which of the following random variables are the default model for random samples?

**(a) iid**

(b) id

(c) pmd

(d) all of the mentioned

Q. 2.11 Cumulative distribution functions are used to specify the distribution of multivariate random variables.

**(a) True**

(b) False

Q. 2.12 Point out the correct statement.

(a) The exponent of a normally distributed random variables follows what is called the log-normal distribution

(b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent

(c) The square of a standard normal random variable follows what is called chi-squared distribution **(d) All of the mentioned**

Q. 2.13 Which of the following goal is incorrectly represented in the below figure?

Statistical Modeling

nferences

Relationship

Distribution

Casual

Fig. Q. 2.13

(a) Relationship between variables

(b) Distribution of variables

(c) Inference about relationships

**(d) Causal**

Q. 2.14 Which of the following is incorrect with respect to use o Poisson distribution?

(a) Modeling event/time data

**(b) Modeling bounded count data**

(c) Modeling contingency tables

(d) All of the mentioned

Q. 2.15 \_\_\_\_\_\_\_\_\_\_\_ random variables are used to model rates.

(a) Empirical

(b) Binomial

**(c) Poisson**

(d) All of the mentioned

Q. 2.16 Point out the wrong statement.

**(a) The normal distribution is asymmetric and peaked about its mode**

(b) A constant times a normally distributed random variable is also normally distributed

(c) Sample means of normally distributed random variables are again normally distributed

(d) None of the mentioned

Q. 2.17 Which of the following form the basis for frequency interpretation of probabilities?

**(a) Asymptotics**

(b) Symptotics

(c) Asymmetry

(d) All of the mentioned

Q. 2.18 1Bernoulli random variables take (only) the values 1 and 0.

**(a) True**

(b) False

Q.2.19 The\_\_\_\_\_\_\_\_\_\_ basically states that the sample mean is consistent.

(a) LAN

**(b) LLN**

(c) LWN

(d) None of the mentioned

Q. 2.20 Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases?

**(a) Central Limit Theorem**

(b) Central Mean Theorem

(c) Centroid Limit Theorem

(d) All of the mentioned

Q. 2.21 The binomial random variables are obtained as the sum of iid Gaussian trials.

**(a) True**

(b) False

Q. 2.22 The \_\_\_\_\_\_\_ of the Chi-squared distribution is twice the degrees of freedom.

**(a) variance**

(b) standard deviation

(c) mode

(d) none of the mentioned

Q. 2.23 Point out the correct statement.

(a) Asymptotics are incredibly useful for simple statistical inference and approximations

(b) Asymptotics often lead to nice understanding of procedures

(c) An estimator is consistent if it converges to what you want to estimate

**(d) All of the mentioned**

Q. 2.24 Gosset's distribution is invented by which of the following scientist?

**(a) William Gosset.**

(b) William Gosling

(c) Gosling Gosset

(d) All of the mentioned

Q. 2.25 Point out the wrong statement.

**(a) Asymptotics generally give assurances about finite sample performance**

(b) The sample variance and the sample standard deviation are consistent as well

(c) The sample mean and the sample variance are unbiased as well

(d) None of the mentioned

Q. 2.26 Which of the following is a property of likelihood?

(a) Ratios of likelihood values measure the relative evidence of one value of the unknown parameter to another

(b) Given a statistical model and observed data, all of the relevant information contained in the data regarding the unknown parameter is contained in the likelihood

(c) The Resultant likelihood is multiplication of individual likelihood

**(d) All of the mentioned**

Q.2.27 CLT is mostly useful as an approximation.

**(a) True**

(b) False

Q. 2.28 The beta distribution is the default prior for parameters between

(a) 0 and 10

(b) 1 and 2

**(c) 0 and 1**

(d) None of the mentioned

Q. 2.29 Which of the following mean is a mixture of the MLE and the prior mean?

(a) interior

(b) exterior

**(c) posterior**

(d) all of the mentioned

Q. 2.30 Usually replacing the standard error by its estimated value does change the CLT.

(a) True

**(b) False**

Q. 2.31 Which of the following testing is concerned with making decisions using data?

(a) Probability

**(b) Hypothesis**

(c) Causal

(d) None of the mentioned

Q. 2.32 Point out the correct statement.

(a) Power of a one sided test is lower than the power of the associated two sided test

(b) Power of a two sided test is greater than the power of the associated one sided test

(c) Hypothesis testing is less commonly used

**(d) None of the mentioned**

Q. 2.33 Which of the following value is the most common measure of "statistical significance"?

**(a) P**

(b) A

(c) L

(d) All of the mentioned

Q. 2.34 Point out the wrong statement with respect to FDR.

**(a) FDR is difficult to calculate**

(b) FDR is relatively less conservative

(c) FDR allows for more false positives

(d) None of the mentioned

Q. 2.35 Which of the following is the oldest multiple testing correction?

**(a) Bonferroni correction**

(b) Bernoulli correction

(c) Likelihood correction

(d) All of the mentioned