QUESTION BANK- 2

1. Interviewing all members of a given population is called:

A. a sample
B. a Gallup poll
C. a census
D. a Nielsen audit.

Ans: Option C

Explanation: An audit tries to count every item, but it does not interview them, so it is not the same as the census. Nielsen made his name by counting stocks in stores across the USA and Gallup made his name in opinion research by taking samples (the term 'Gallup poll' has now entered dictionaries).

2. Sampling means following a sequence of stages. Which ONE of the following stages should come before the others?

A. proceed with the fieldwork
B. Find suitable source for the population members
C. Define the people of interest
D. Examine the objective of the study.
Ans: Option D

Explanation: All research should stem from the purpose of the study. Otherwise, we are following our instincts and interest, and our clients will not be satisfied.

3.For sampling, which ONE of the following should be up-to-date, complete and affordable?

A. Census.
B. CAPI machine
C. Sampling frame
D. Respondent.

Ans: Option C

4.A sampling frame is usually a list or directory. The statement cannot be applied to the other options.Which ONE of the following is the benefit of using simple random sampling?

A. We can calculate the accuracy of the results.
B. The results are always representative
C. Interviewers can choose respondents freely.
D. Informants can refuse to participate.

Ans:Option A.

Point to be noted: Remember that random does not necessarily mean representative, and random approaches mean that interviewers have strict instructions about who to choose. Refusals are common to all methods and are never a good thing.

5.Which ONE of the following is the main problem with using non-probability sampling techniques?

A. The expense.
B. The results are never representative.
C. Informants can refuse to participate
D. Human judgment error

Ans: Option D

Explanation: At any time when humans interact, there is scope for bias. Interviewers tend to select people who look likely to talk - but the busy person is as important to the results.

6.Which ONE of the following methods is generally used in qualitative sampling?

A. Random digit dialing
B. Quota
C. Stratified random
D. Simple random.

Ans: Option B

Explanation: quota sampling is also commonly used with quantitative research. Any qualitative study that uses the word 'random' should be inspected carefully, because there may be some misunderstanding.

7. Which ONE of these sampling methods is a probability method?

A. Quota

B. Judgment

C. Convenience

D. Simple random

Ans: Option D

8.The expected value or _____ of a random variable is the center of its distribution.
a) mode
b) median
c) mean
d) bayesian inference

Answer: c

Explanation: A probability model connects the data to the population using assumptions.

9.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

Answer: d

Explanation: Every cumulative distribution function F is non-decreasing and right-continuous.

10.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

Answer: a

Explanation: Densities with a higher variance are more spread out than densities with a lower variance.

11. The square root of the variance is called the _____ deviation.

a) empirical

b) mean

c) Continuous

d) standard

Answer: d

Explanation: Standard Deviation (SD) is the measure of spread of the numbers in a set of data from its mean value

12. 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) Python cannot approximate quantiles for you for common distributions

d) None of the Mentioned

Answer: c

Explanation: Python can approximate quantiles for you for common distributions

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

a) Chebyshev

b) Stautaory

c) Testory

d) All of the Mentioned

Answer: a

Explanation: Chebyshev’s inequality is also spelled as Tchebysheff’s inequality.

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

a) True

b) False

Answer: b

Explanation: For continuous random variables, the PDF is the derivative of the CDF.

15.Chebyshev’s inequality states that the probability of a “Six Sigma” event is less than :

a) 10%

b) 20%

c) 30%

d) 3%

Answer: d

Explanation: If a bell curve is assumed, the probability of a “six sigma” event is on the order of one ten millionth of a percent.

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

a) iid

b) id

c) pmd

d) all of the Mentioned

Answer: a
Explanation: Random variables are said to be iid if they are independent and identically distributed.

17.Cumulative distribution functions are used to specify the distribution of multivariate random variables.
a) True
b) False

Answer: a
Explanation: In the case of a continuous distribution, it gives the area under the probability density function from minus infinity to x.

18. 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

Answer: d
Explanation: Many random variables, properly normalized, limit to a normal distribution

19. Which of the following is incorrect with respect to use of Poisson distribution ?
a) Modeling event/time data
b) Modeling bounded count data
c) Modeling contingency tables
d) All of the Mentioned
Answer: b
Explanation: Poisson distribution is used for modeling unbounded count data.

20. _____ random variables are used to model rates.

a) Empirical

b) Binomial

c) Poisson

d) All of the Mentioned

Answer: c

Explanation: Poisson distribution is used to model counts.

21.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

Answer: a

Explanation: The normal distribution is symmetric and peaked about its mean.

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

a) Asymptotics

b) Symptotics

c) Asymmetry

d) All of the Mentioned

Answer: a

Explanation: Asymptotics is the term for the behavior of statistics as the sample size.

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

a) True

b) False

Answer: a

Explanation: The Bernoulli distribution arises as the result of a binary outcome.

24. 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

Answer: a

Explanation: The Central Limit Theorem (CLT) is one of the most important theorems in statistics.

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

a) True

b) False

Answer: a

Explanation: The binomial random variables are obtained as the sum of iid Bernoulli trials.

26. Which of the following can be considered as random variable ?

a) The outcome from the roll of a die

b) The outcome of flip of a coin

c) The outcome of exam

d) All of the Mentioned

Answer: d

Explanation: The probability distribution of a discrete random variable is a list of probabilities associated with each of its possible values.

27. Which of the following random variable that take on only a countable number of possibilities?

a) Discrete

b) Non Discrete

c) Continuous

d) All of the Mentioned

Answer: a

Explanation: Continuous random variable can take any value on the some subset of the real line.

28. Point out the wrong statement:

a) A random variable is a numerical outcome of an experiment

b) There are three types of random variable

c) Continuous random variable can take any value on the real line

d) None of the Mentioned

Answer: b

Explanation: There are two types of random variable-continuous and discrete.

29. Which of the following condition should be satisfied by function for pmf ?

a) The sum of all of the possible values is 1

b) The sum of all of the possible values is 0

c) The sum of all of the possible values is infinite

d) All of the Mentioned

Answer: a

Explanation: A probability mass function evaluated at a value corresponds to the probability that a random variable takes that value.

30. Which of the following function is associated with a continuous random variable ?

a) pdf

b) pmv

c) pmf

d) all of the Mentioned

Answer: a

Explanation: pdf stands for probability density function.

31. Statistical inference is the process of drawing formal conclusions from data.

a) True

b) False

Answer: a

Explanation: Statistical inference requires navigating the set of assumptions and tools

32. 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

Answer: a

Explanation: If you can’t fix the variable,stratify it.

33. If X predicts Y,it does mean X causes Y.

a) True

b) False

Answer: b

Explanation: If X predicts Y,it does not mean X causes Y.

34. Point out the wrong statement:

a) In Sample Error is also called generalization error

b) Out of Sample Error is the error rate you get on the new dataset

c) In Sample Error is also called resubstitution error

d) All of the Mentioned

Answer: a

Explanation: Out of Sample Error is also called generalization error.

35. Which of the following is correct order of working ?

a) questions->input data ->algorithms

b) questions->evaluation ->algorithms

c) evaluation->input data ->algorithms

d) all of the Mentioned

Answer: a

Explanation: Evaluation is done in the last.

36. Which of the following shows correct relative order of importance ?

a) question->features->data->algorithms

b) question->data->features->algorithms

c) algorithms->data->features->question

d) none of the Mentioned

Answer: b

Explanation: Garbage in should be equal to garbage out.

37. Which of these measures are used to analyze 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

Solution: (B)

The mean, median, mode are the three statistical measures which help us to analyze the central tendency of data. We use these measures to find the central value of the data to summarize the entire data set.

38. Five numbers are given: (5, 10, 15, 5, 15). Now, what would be the sum ofdeviations of individual data points from their mean?

A) 10

B)25

C) 50

D) 0

Solution: (D)

The sum of deviations of the individual will always be 0.

39. 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) 150
D) None of the above

Solution: (A)

$X = \mu + Z\sigma$ where μ is the mean, σ is the standard deviation and X is the score we're calculating. Therefore $X = 150+20*1.5 = 180$

40. 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

Solution: (A)

The mean of the dataset would always change if we change any value of the data set. Since we are summing up all the values together to get it, every value of the data set contributes to its value. Median and mode may or may not change with altering a single value in the dataset.

41.Which of the following statements are true about Bessels Correction while calculating a sample standard deviation?

1. Bessels correction is always done when we perform any operation on a sample data.
2. Bessels correction is used when we are trying to estimate population standard deviation from the sample.
3. Bessels corrected standard deviation is less biased.

A) Only 2

B) Only 3

C) Both 2 and 3

D) Both 1 and 3

Solution: (C)

Contrary to the popular belief Bessel's correction should not be always done. It's basically done when we're trying to estimate the population standard deviation using the sample standard deviation. The bias is definitely reduced as the standard deviation will now(after correction) be depicting the dispersion of the population more than that of the sample.

42. [True or False] Standard deviation can be negative.

A) TRUE

B) FALSE

Solution: (B)

43. [True or False] The standard normal curve is symmetric about 0 and the total area under it is 1.

A)TRUE

B) FALSE

Solution: (A)

By the definition of the normal curve, the area under it is 1 and is symmetric about zero. The mean, median and mode are all equal and 0. The area to the left of mean is equal to the area on the right of mean. Hence it is symmetric.

44. If n = 10 and p= 0.8, then the standard deviation of the binomial distribution is _____.
i) 0.80
ii) 1.26
iii) 1.60
iv) 8.00
Solution: iii

45. If the outcomes of a discrete random variable follow a Poisson distribution, then their
i) mean equals the variance
ii) mean equals the standard deviation
iii) medain equals the variance
iv) median equals the standard deviation
Solution: i

46. Which of the following is not true about binomial probability distribution?
i) Each outcome is independent of each other
ii) Each outcome can be classified as either success or failure.
Iii) The probability of success must be constant from trial to trial.

iv) The random variable of interest is continuous.

Solution: iv

47. The mean of certain discrete random variable is 12, find out the mean if all of the observations are multiplied by 2 and added with 2.

i) 24

ii) 4

iii) 26

iv) 12

Solution: iii

48. The standard deviation of certain continous random variable is 4, find out the standard deviation if all of the observations are multiplied by 2 and added with 2.

i) 8

ii) 6

iii) 10

iv) 4

Solution: i

49.Suppose a discrete random variable has following distribution

X	4	3	12	2
P(X == x)	0.3	0.2	0.2	0.3

find out the mean and standard deviation of the distribution

i) 3.6 1.2

ii) 4.8 3.68

iii) 4.8 1.2

iv) 5 3.68

Solution: ii

50.PDF of a certain continuos random variable is defined by function where x can be anything in the interval [0,1]. Compute C(approx) using normalization condition on PDFs.

i) 12

ii) 6

iii) 2

iv) 24

Solution: i