

STATISTICS WORKSHEET-1

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

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

a) True
b) False

Ans: a

2. 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

Ans: a

3. 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

Ans: b

4. 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

Ans: d

5. _____ random variables are used to model rates.

a) Empirical
b) Binomial
c) Poisson
d) All of the mentioned

Ans: c

6. 10. Usually replacing the standard error by its estimated value does change the CLT.

a) True
b) False

Ans: b

7. 1. Which of the following testing is concerned with making decisions using data?

a) Probability
b) Hypothesis
c) Causal
d) None of the mentioned

Ans: b

8. 4. Normalized data are centered at _____ and have units equal to standard deviations of the original data.

- a) 0
- b) 5
- c) 1
- d) 10

Ans: a

9. Which of the following statement is incorrect with respect to outliers?
- a) Outliers can have varying degrees of influence
 - b) Outliers can be the result of spurious or real processes
 - c) Outliers cannot conform to the regression relationship
 - d) None of the mentioned

Ans: b

Q10 and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What do you understand by the term Normal Distribution?

Ans: Normal Distribution

Common Continuous Probability Distribution is stated for Normal Distribution in statistics for advanced data evaluation and interpretation.

The normal distribution is a probability distribution that (roughly) describes many common datasets in the real world. It is the most common type of distribution, and it arises naturally in statistics through random sampling techniques.

Data utility - Random variable $\in \mathbb{R}$

classification -

Normal Distribution -

1.) Probability Distribution Function [P.D.F.]

2.) Cumulative Distribution Function [C.D.F.]

11. How do you handle missing data? What imputation techniques do you recommend?

Ans:

1. Deleting Rows with missing values.
2. Impute missing values for continuous variable.
3. Impute missing values for categorical variable.
4. Other Imputation Methods.
5. Using Algorithms that support missing values.
6. Prediction of missing values.

Common techniques that we are usually recommend as bellowed:

1. Mean or Median Imputation. When data is missing at random, we can use list-wise or pair-wise deletion of the missing observations. ...
2. Multivariate Imputation by Chained Equations (MICE) MICE assumes that the missing data are Missing at Random (MAR). ...
3. Random Forest.

12. What is A/B testing?

Ans:

also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variables are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drive business metrics.

Essentially, A/B testing eliminates all the guesswork out of website optimization and enables experience optimizers to make data-backed decisions. In A/B testing, A refers to 'control' or the original testing variable. Whereas B refers to 'variation' or a new version of the original testing variable.

13. Is mean imputation of missing data acceptable practice?

Ans: True, imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the mean remains unbiased. That's a good thing. ... Since most research studies are interested in the relationship among variables, mean imputation is not a good solution.

14. What is linear regression in statistics?

Ans: In statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables). The case of one explanatory variable is called simple linear regression.

Linear regression was the first type of regression analysis to be studied rigorously, and to be used extensively in practical applications. This is because models which depend linearly on their unknown parameters are

easier to fit than models which are non-linearly related to their parameters and because the statistical properties of the resulting estimators are easier to determine.

15. What are the various branches of statistics?

Ans: Usually Statistics plays a main role in the field of research. It helps us in the collection of data, analysis of data and presentation of data.

Various branches of statistics:

descriptive statistics

inferential statistics.

Collection of data

These are the main branches of statistics.



FLIP ROBO
