

STATISTICS WORKSHEET-1

1. **Bernoulli random variables take (only) the values 1 and 0.**
Ans: (a) true.
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?**
Ans: (a) central limit theorem.
3. **Which of the following is incorrect with respect to use of Poisson distribution?**
Ans: (b) Modelling bounded count data.
4. **Point out the correct statement.**
Ans: (d) All of the mentioned.
5. **_____ random variables are used to model rates.**
Ans: (d) All of the mentioned.
6. **Usually replacing the standard error by its estimated value does change the CLT**
Ans: (b) False.
7. **Which of the following testing is concerned with making decisions using data?**
Ans: (b) Hypothesis.
8. **Normalized data are cantered at _____ and have units equal to standard deviations of the original data.**
Ans: (a) 0.
9. **Which of the following statement is incorrect with respect to outliers?**
Ans: (c) Outliers cannot conform to the regression relationship.
10. **What do you understand by the term Normal Distribution?**
Ans: Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve.
The point given are:
 1. A normal distribution is the proper term for a probability bell curve.
 2. In a normal distribution the mean is zero and the standard deviation is 1. It has zero skew and a kurtosis of 3.
 3. Normal distributions are symmetrical, but not all symmetrical distributions are normal.
 4. In reality, most pricing distributions are not perfectly normal.
11. **How do you handle missing data? What imputation techniques do you recommend?**
Ans: There are proven techniques to deal with missing data. When dealing with missing data, data scientists can use two primary methods to solve the error: imputation or the removal of data. The imputation method develops reasonable guesses for missing data. It's most useful when the percentage of missing data is low.

The data is handled by the Imputation vs. Removing, Imputation, Deletion, Multiple Imputation, K Nearest Neighbour.

The imputation techniques are:

1. Mean, Median and Mode.
2. Time-Series Specific Methods
3. Last Observation Carried Forward (LOCF) & Next Observation Carried Backward (NOCB)
4. Linear Interpolation
5. Seasonal Adjustment with Linear Interpolation.

12. **What is A/B testing?**

Ans: A/B testing (also known as split testing or bucket testing) is a method of comparing two versions of a webpage or app against each other to determine which one performs better. A/B testing is essentially an experiment where two or more variants of a page are shown to users at random, and statistical analysis is used to determine which variation performs better for a given conversion goal.

Running an A/B test that directly compares a variation against a current experience lets you ask focused questions about changes to your website or app and then collect data about the impact of that change.

Testing takes the guesswork out of website optimization and enables data-informed decisions that shift business conversations from "we think" to "we know." By measuring the impact that changes have on your metrics, you can ensure that every change produces positive results.

13. **Is mean imputation of missing data acceptable practice?**

Ans: 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. 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?** 15. **What are the various branches of statistics**

Ans:

Linear Regression: - Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable. For example, a modeler might want to relate the weights of individuals to their heights using a linear regression mode.

The simplest form of the regression equation with one dependent and one independent variable is defined by the formula $y = c + b \cdot x$, where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable.

The branches of statistics: -

1. **Econometric:** Econometric is one of the branches of statistics where it takes parts to resolve economic models and problem.
2. **Actuarial:** Actuarial is another applied statistical branch that focuses on studying and analysing risk in finance and insurance.
3. **Psychometrics:** Psychometrics is another interesting branch of statistics. This one focus on studying measurement technique and analysing in the education world and psychology. This thing included attitude, personality, emotion, and many others.

4. **Physics Statistics:** Physics statistics is one of the statistical branches that focuses on solving physics science. Usually, statistics take part in measurement and calculation with particle.
5. **Population Statistics:** The population of statistics is one of the most useful branches that study about many things related to society. It has many connections with another aspect of our life, such as health, education, migration, and so on calculation with particle.
6. **Official Statistics:** Not just people, the government also need statistics as a tool to make the right policy. In official statistics, we learn how to measure some strategic indicator that has a large impact on society. There are many indicators that the government's need to fulfil the policy based on evidence or data.
7. **Biostatistics:** Biology statistics or more popular with names Biostatistics is another branch of statistics that used to resolve many statistical problems in biology.
8. **Industrial Statistics:** In Industrial Statistics, we learn and study how to resolve an industrial problem such as quality control, queuing theory, optimizing variable, etc.
9. **Computing statistics:** Computing statistics is a branch of statistics that focuses on the use of information technology in producing more powerful statistics.