STATISTICS WORKSHEET-1 solutions

Ans 1 : - a - true. As Bernoulli distribution is binary I.e 0, 1..

Ans 2 : - a - Central Limit Theorem

Ans 3 : - c - Modeling bounded count data.

Ans 4 : - All of the mentioned.

Ans 5 : - Poisson random variables are used to model rates or counts

Ans 6 : - b - False. It doesn't change the CLT.

Ans 7 : - b - Hypothesis as null hypothesis is true and statistical data is needed to prove it wrong.

Ans 8 : - a - 0, in statistics it is difficult to do calculations every time we have a new set of values.

Ans 9 : - c- Outliers cannot conform to the regression relationship. It confirms with regression relationship.

Ans 10 : -
<u>Normal Distribution</u>: It is also called as Bell curve or Gaussian distribution. It has a probability value of $\pm 3\sigma$, where 0 being the mean, $\pm 1\sigma$ area under curve contains 68% of distribution data, $\pm 2\sigma$ area under curve contains 95% of distribution data and $\pm 3\sigma$ area under curve contains 99% of distribution data.

 It is most wide known and used distribution as it approximates most of the phenomenon very well.

Ans 11: -
There are 2 methods to handle missing data:
1) Imputation:- it is a technique used to replace missing data with some substitute value to retain most of the data in dataset.
.
2) Removal of data.(not preferred as it might reduce data making it useless or incomplete for correct analysis)

Imputation techniques recommended are as further:

A) Arbitrary value imputation.
B) Frequent category imputation.
C) Mean median imputation.
D) Complete case analysis.

Ans 12: -
 A/B testing is a most premium method and widely used stats tool. It is comparison method which helps for decision making in between two variables to find out who performs better under standard conditions.
We create a hypothesis(mean), then create 2 groups for the variables, then select samples randomly from the population of groups created. Further we calculate probability of the data collected of 2 groups and check the value with significance level (0.05).

Ans 13: - Is Mean imputation of missing data acceptable practice. It is a method which being simple, but does not preserve inter-relation between variables.
Also the standard errors are wrongly minimized due to deviation in population being very near to nearby values.

Ans 14: -
Linear regression is an algorithmic prediction between two variables' relations.
In linear regression examples 2 variables are examined viz.
Independent variable and Dependent variable.
Independent variables does not fluctuate or does not influence by other variables, whereas **dependent variables** fluctuate as independent variables changes.
Dependent variable is being studied or being predicted

Ans 15: -   Various Branches in stats
1) Mathematical or Theoretical statistics. (It helps forming experimental and statistical distribution)
2) Statistical methods or functions. (It helps in collection, interpretation and tabulation of data. It helps analyze data and give insight of data).
3) Descriptive statistics. (Helps summarize and organize any dataset characteristics, also represents data in classification and diagrammatic way.)
4) Inferential statistics. (Helps concluding the data after analyzing the sample drawn from it.)