Statistics worksheet solution

- Ans 1 true
- Ans2 a) central limit theorem
- Ans3 b) modelling bounded count data
- Ans4 d) all of the above
- Ans5) c) poisson
- Ans6) b) false
- Ans7) b) Hypothesis testing
- Ans8) a) at 0
- Ans9) c) outliers cannot confirm to the regression relationship

SUBJECTIVE QUESTION

Ans10) Normal distribution :

Normal distribution is also called as a Gaussian distribution or referred as bell curve. It is continuous in nature like age, marks ,salary etc. and cannot be used in discrete or categorical data. In this type of distribution every event is independent from other. The centre of distribution is always inclined towards mean.so it is a probability distribution which is symmetric about the mean which implies that density of data are more distributed around the mean as compared to far from mean.

Ans 11) there are several techniques to handle missing data like interpolation, regression analysis,

By mean and medium etc.

Correlation and regression method are best to solve missing data,

Ans12) A/B also called as split testing is basically use to compare between two version of anything

And then compare which version is better. In this method one component has to be changed

To observe the effects on its popularity and then accept the better one.

Few steps has to be followed to perform the test

a) Identify a problem
b) Analyse the data provided by user
c) Perform hypothesis testing
d) Again analyse the changes
e) Select the best one

Ans13) the process of replacing null values with the overall mean of the observation is consider as Data mean imputation

Mean imputation is generally not considered as best practice to replace null values as provided vague results and mainly biased one for example if BMI (BODY MASS INDEX) of a sample is provided in which age group above 60years is dominated in comparison to 18 years old then replacing null value will not provide satisfactory result.

Ans 14) What is linear regression in statistics?

Linear regression is basically used to predict the value of unknown variables by using the known one. in this method coefficient of independent variable in equation is find through various technique and through the analysis of this dependent variables are predicted.

Ans15) There are basically two branches of statistics.

a) Descriptive statistics = in this process detail of each and every sample is tested and finally summary has been prepared based on this.
b) Inferential statistics = in this process a random data of the sample out of whole lot has been collected and inference has been made on that basis. It is very useful method when its not possible to analyse whole set of data. example **e**lection poll survey.