STATISTICS WORKSHEET - 1

ANS.1 True

ANS.2 Central limit theorem

ANS.3 Modeling bounded count data

ANS.4 All of the mentioned

ANS.5 Poisson

ANS.6 False

ANS.7 Hypothesis

ANS.8 0

ANS.9 Outliers cannot conform to the regression

relationship.

ANS.10 As with any probability distribution , the normal

distribution describes how the values of a

variable are distributed. It is the most important

probability distribution in statistics because it fits

many natural phenomena for example – height ,

blood pressure , measurement error and IQ

scores follow the normal distribution.

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.

Ans.11 A common technique is to use the mean or

median of the non-missing observations. This

can be useful in cases where the number of

missing observations is low. However , for large

number of missing values using mean or median

can result in loss of variation in data and it is

better to use imputation.

IMPUTATION TECHNIQUES

- COMPLETE CASE ANALYSIS(CCA)- this is a quite

straight forward method of handling the missing

data , which directly removes the rows that have

missing data. i.e we consider only those rows

where we have complete data i.e data is not

missing.

- Arbitray value imputation.

* Frequent category imputation.

Ans.12 A/B testing (also known as bucket testing or split-

run testing) is a user experience research meth-

odology. A/B tests consist of randomized

experiment with two variants, A and B.It includes

application of statistical hypothesis testing or two

sample hypothesis testing as used in the field of

of statistics. A/B testing is a way to compare two

versions of a single variable, typically by testing a

subject’s response to variant A against variant B

and determining which of the two variants is

more effective.

Ans.13 True , imputing the mean preserves the mean of

the observed data. So if the data are missing co-

mpletely at random the estimate of the mean

remains unbiased.

* Bad practice in general .

Ans.14 Linear regression attempts to model the relation-

ship between two variables by fitting a linear

equation to observed data. One variable is

considered to be a dependent variable for ex-

A modeler might want to relate the weights of

individuals to their heights using a linear regression

model.

Linear regression attempts to model the

relationship between two variable by fitting a

linear equation to observed data. A linear regres-

sion line has an equation of the form y=a+bx , where

X is the explanatory variable and Y is the dependent

variable.

Ans.15 There are three real branches of statistics : data

collection, descriptive statistics and inferential

statistics.

-Data collection – is all about how the actual data is

collected. For the most part ,this

need not concern us too much in

term of the mathematics but there

are significant issue to consider when

actually collecting data.

-Descriptive statistics – is the part of statistics that deal

we have. This can take two basic

forms.Presenting aspects of the

numerically [via averages and so on].

-Inferential statistics – is the aspects that deals with

making conclusions about the data

this is quite a wide area; essentially

you are asking ‘what is this data

telling us ,and what should we do?’