

DA

Q1. a) A null hypothesis is a type of hypothesis used in statistics that proposes that there is no difference between certain characteristics of a population (or data-generating process).

Alternative hypothesis :- It states that there is statistical significance between two variables.

b) Level of significance :- The significance level, also denoted as alpha or α , is the probability of rejecting the null hypothesis when it is true. eg:- a signif. level of 0.05 indicates a 5% risk of concluding that a difference exists when there is no actual diff.

c) Type I and II error :- Type I errors often assimilated with false positives - happen in hypothesis testing when the null hypothesis is true but rejected.

Type II errors happens when the null hypothesis is false and you subsequently fail to reject it.

Ans 2. Logistic regression is an extension of simple linear regression

Where the dependent variable is dichotomous or binary in nature, we cannot use simple linear regression. It is the statistical technique used to predict the relationship between predictors (our independent variables) and a predictable variable (the dependent variable) when the dependent variable is binary (eg:- response yes or no, score high or low).

There must be two or more independent variables or predictors for a logistic regression. The IVs, or predictors, can be continuous or categorical.

All predictor variables are tested in one block to assess their predictive ability while controlling for the effects of other predictors in the model.

Assumption for a Logistic regression

1. Adequate sample size (too few participants for too many predictors)

is bad !)

2. Absence of multicollinearity i.e., high intercorrelation among the predictors
3. No outliers

The statistic $-2 \log L$ is a badness of fit indicator, that is, large numbers mean poor fit of the model to the data.

When taken from large samples, the dif. b/w two values of $-2 \log L$ is chi-square.

Ans 3. Discriminate Analysis :- Acc. to this method distinct classes' data are based on distinct Gaussian distribution. For the purpose of creating a classifier the parameters of Gaussian distribution are estimated by the fitting function for every classes.

In order to predict new data classes the classes having the low cost of misclassification is found by the trained classifier.

Types :-

1. Linear Discriminant Analysis :- It is used for performing dimensionally ~~test~~ reduction where as preserving as much as possible the information of class discrimination.
2. Multiple Discriminant Analysis :- It is a technique used to compress a multivariate signal for producing a low dimensional signal that is open to classfcⁿ.
3. Quadratic Disc. Analysis :- The assumption of groups with matrices having equal covariances is not present in Quadratic Discrim. Analysis.
4. Canonical :- It is a method of dimension reduction liked with canonical correlation and Principal Component Analysis.
5. Gaussian Discrim Anal. :- It is used when data can be approximated with normal distribution.