Date 30th June 2020

Dr. Rajesh Verma, Assistant professor and Head, U.G. Department of Zoology, D.K. College Dumraon, (Buxar)
Notes for B.Sc part 1st, paper (2A)

Question:- Write Notes on STANDARD-t test?

Answer:- **Student's t-test:-**

The t-test is any statistical hypothesis test in which the test statistic follows a Student's t-distribution under the null hypothesis.

A t-test is most commonly applied when the test statistic would follow a normal distribution if the value of a scaling term in the test statistic were known. When the scaling term is unknown and is replaced by an estimate based on the data, the test statistics (under certain conditions) follow a student's distribution. The t-test can be used, for example, to determine if means of two sets of data are significantly different from each other.

History:-

Date
20 June, 2020

The Term "t-statistic" is abbreviated from "hypothesis test statistic". In statistics, the t-distribution was first derived as a posterior distribution in 1876 by Helmert and Lüroth. The t-distribution also appeared in a more general form as Pearson type IV distribution in Karl Pearson's 1895 paper. However, the t-Distribution also known as student's T Distribution gets its name from William Sealy Gosset who first published in it English literature in his 1908 paper titled Biometrika using his pseudonym "student" because his employer preferred staff to use pen names when publishing scientific papers instead of their real name, so he used the name "Student" to hide his identity. Gosset worked at the Guinness Brewery in Dublin, Ireland, and was interested in the problems of small samples - for example, the chemical properties of barley with small samples - for example, the chemical properties of barley with small sample sizes. Hence a second version of the etymology of the term student is that Guinness did not want their competitors to know that they

were using the t-test to determine the quality of raw material.

Slope of a regression line :-

Suppose one is fitting the model

$$Y = \alpha + \beta x + \epsilon$$

Where x is known, α and β are unknown, ϵ is a normally distributed random variable with mean 0 and unknown variance σ^2, and Y is the outcome of interest. We want to test the null hypothesis that the slope β is equal to some specified value β_0 (often taken to be 0, in which case the null hypothesis is that x and y are uncorrelated).

Let,

$\hat{\alpha}, \hat{\beta}$ = least-squares estimators,

$SE_{\hat{\alpha}}, SE_{\hat{\beta}}$ = the standard errors of least-squares

Then

$$t \text{ score} = (\hat{\beta} - \beta_0) / SE_{\hat{\beta}} \sim T_{n-2}$$

Date 20-06-2020

has a t-distribution with $n-2$ degrees of freedom if the null hypothesis is true. The standard error of the slope coefficient:

$$SE_{\hat{\beta}} = \frac{\sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

can be written in terms of the residuals. let

$$\hat{\varepsilon}_i = y_i - \hat{y}_i = y_i - (\hat{\alpha} + \hat{\beta} x_i) = \text{Residuals} =$$

$$SSR = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \text{sum of squares of residuals.}$$

~~then term is given by :~~

can be written in terms of the residuals let

$$\hat{\varepsilon}_i = y_i - \hat{y}_i = y_i - (\hat{\alpha} + \hat{\beta} x_i) = \text{Residuals} =$$

$$SSR = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \text{sum of squares of residuals.}$$