DATE 26-6-2020

Dr. Rajesh Verma.
Assistant professor and Head UG
Department of Zoology D.K College
Dumraon (Buxar)

Notes for B.Sc part 1st paper 2(A).

Que. (1.) Write Notes on Goodness of fit?

The goodness of fit of a Statistical model describes how well it fits a set of Observations. Measures of goodness of fit typically Summarize the discrepancy between observed values and the values expected under the model in question. Such Measures can be used in Statistical hypothesis testing. e.g. to test for Normality of residuals, to test whether two Samples are drawn from identical distributions (See kolmogorov-smirnov test), or wether outcome frequencies follow a Specified distribution (See pearson's Chi-Squared test). In the analysis of Variance, one of the Components into which the Variance is partitioned may be a lack-of-fit Sum of Squares.

* Regression Analysis —

In regression analysis, the following topics relate to goodness of fit:

* Coefficient of determination (the R-Squared measure of goodness of fit)

* Lack-of-fit Sum of Squares:

* Reduced Chi-Squared

* Regression Validation

* Mallows's Cp Criterion

Categorical data —

The following are examples that arise in the Context of Categorical data.

Pearson's Chi-Squared test —

Pearson's Chi-Squared test uses a Measure of goodness of fit which is the Sum of differences between observed and expected outcome frequencies (that is, Counts of observations), each Squared and divided by the expectation:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

* Fit of distributions —

In assessing whether a given distribution is suited to a data set, the following tests and their underlying measures of fit can be used.

* Bayesian information Criterion
* Kolmogorov - Smirnov test
* Cramer – von Mises Criterion
* Anderson — Darling test
* Shapiro — Wilk test
* Chi – Squared test
* Akaike information Criterion.
* Hosmer — Lemeshow test
* Kuiper's test
* Kernelized Stein discrepancy
* Zhang's Z_K, Z_C and Z_A tests
* Moran test.

Where:
O_i = an observed count for bin i

E_i = an expected count for bin i, asserted by the null hypothesis.
The expected frequency is calculated by:

$$E_i = (F(Y_u) - F(Y_l))N$$

Where:
F = the cumulative distribution function for the probability distribution being tested.
Y_u = the upper limit for class i,

Y_l = the lower limit for class i, and

N = the sample size.

Example: equal frequencies of men and women —

For example, to test the hypothesis that a random sample of 100 people has been drawn from a population in which men and women are equal in frequency, the observed number of men and women would be compared to the theoretical frequencies of a ~~men~~

$$\chi^2 = \frac{(44-50)^2}{50} + \frac{(56-50)^2}{50} = 1.4$$