Mcs – 511
DCS-UOK
Supervisor: Ms. Shaista Rais

What is the Chi-square test of independence?

The Chi-square test of independence is a statistical hypothesis test used to determine whether two categorical or nominal variables are likely to be related or not.

Using the Chi-square test of independence

The Chi-square test of independence checks whether two variables are likely to be related or not. We have counts for two categorical or nominal variables. We also have an idea that the two variables are not related. The test gives us a way to decide if our idea is plausible or not.

Test of independence: (steps)

Steps:

To illustartae the concept, let's consider a random sample showing the gender of liberal arts college students and their favorite academic area.

1. Parameter of interest: (gender, favorite course area). Ho: preference of subject area is independent of the gender of the class.

Ha: preference of subject area is not independent of the gender of the class.

2) Hypothesis test area: a) check assumption

b) Identify the probability distribution and the test statistic to be used based on marginal totals

c) The chi-square distribution will be used with d.f = (2-1)*(3-1) = (1)*(2). Determine the level of significance alpha = 0.05

3) a) collect the sample information:

b) Calculate the value of the test statistic

4) The probability Distribution:

a) Determine the critical region and critical value(s)

b) Determine whether or not the calculated test statistic is in the critical region.

5) The Results:

a) State the decision about Ho

b) State the conclusion about Ha

Categorical variable: is a variable that classifies or categorizes each individual into exactly one of several cells or classes that are all inclusive and mutually exclusive.

Example:

Mcs – 511
DCS-UOK
Supervisor: Ms. Shaista Rais

The resulting face from a rolled die is a categorical variable. The list of outcomes {1,2,3,4,5,6} from a set of all-inclusive and mutually exclusive categories.

Contingency table: A contingency table is an arrangement of data into a two way classification. The data are sorted into cells, and the number of data in each cell is repeated. The contingency table involves two factors or variables and the usual question concerning such tables is whether the data indicate that two variables are independent or dependent.