Chi Square Tests

The chi-squared distribution allows for statistical tests of categorical data. Among these tests are those for goodness of fit independence and homogeneity.

Chi-squared goodness of fit tests

A goodness of fit test checks to see if the data has come from some specified population. This test allows us to test if categorical data corresponds to a model where the data is chosen from the categories according to some specified set of probabilities.

If these assumptions are satisfied, then the χ^2 statistic is approximately χ^2 distributed with n– 1 degrees of freedom. The null hypothesis is that the probabilities are as specified, against the alternative that some are not.

Question (1) of the assignment

f = c(615,383,198,804)

p = c(0.4,0.35,0.2,0.05)

chisq.test(f,p)

Chi-squared tests of independence

The same statistic can also be used to study if two rows in a contingency table are "independent". That is, the null hypothesis is that the rows are independent and the alternative hypothesis is that they are not independent.

Eg:- Following data has been tabulated on the severity of a crash for the cases where the passenger had a seat belt, or did not:

		Injury Level		
		None	**normal**	**critical**
Seat Belt	**Yes**	13,417	647	82
	No	55,973	8,000	352

Ybelt = c(13,417 ,647,82)

Nbelt = c(55,973,8,000,352)

chisq.test(data.frame(Ybelt,Nbelt))

Chi-squared tests for homogeneity

A test for homogeneity, tests to see if the rows come from the same distribution or appear to come from deferent distributions. Intuitively, the proportions in each category should be about the same if the rows are from the same distribution. The way of applying chi-square test of homogeneity is exactly same as for test of independence and only differs from the way of writing the hypothesis.