TRANSCRIPT   MY CLASS   NOTES

Slide 1

In the final part of the statistic section, we reviewed multiple sample tests, ANOVA and Chi square. Multiple sample tests are used when we want to check for differences in sample means across more than two samples. We first looked at ANOVA which is used when a target variable or outcome is continuous and the factors or the independent variables that impact the target variable are discrete. Within ANOVA, we can conclude only that the means are different between the samples or not between the samples.

But we cannot be sure directly from an ANOVA which of the sample mean is different. In order to figure that out we need to run Post Hoc Test. ANOVA uses an F distribution which is the ratio of two Chi square distributions. It looks at the variances between the samples to come up with an estimate of whether or not the sample means are statistically, significantly different from one another.

Slide 2

Another kind of a multiple sample tests are Chi square tests which are used for testing differences across samples when we are deal with frequency or count data. There are many applications of Chi Square test including test of

association and goodness of fit tests. All the Chi square tests test of association, goodness of fit test essentially rely on the fact that the sum of observed - expected square / expected will follow a Chi square distribution.

So to implement a Chi square test typically what we do is

- We look at the observed frequencies in the sample that we have created
- We calculate the expected frequencies assuming that the null hypothesis is true
- We calculate a Chi square test static and come up with the appropriate P value

These Chi square tests test of association, goodness of fit test that rely on the observed – expected^2/expected summation are essentially what are called non-parametric tests. They are non-parametric because they don't require the data to follow any particular distribution. Irrespective of the distribution of the underlying data mathematically we can show that observed – expected^2/expected summation will follow a Chi square distribution. But there is also a parametric Chi square test which is a Chi square test of variance.

It is used to check whether or not it is different from the population variance. Remember we used a central limit theorem to approximate

TRANSCRIPT   MY CLASS   NOTES

sample mean distributions to normal. However the variation of sample means will not follow a normal distribution. The variance will actually follow a Chi square distribution. So if we are interested in checking a hypothesis test about variance, we will use a Chi square test of variance.