

ANOVA:

Anova is performed to know if two or more different groups have the same means. So, we apply anova for one categorical variable with one numeric variable.

Eg: Does yoga has impacts logical ability. (Two groups- control group no yoga and other group yoga. Then test is administered and their scores are recorded. Are the scores different- We perform ANOVA.

We require

1. Total Sum of Square Errors
2. Between Sum of Squares
3. Within Sum of Squares
4. Degrees of Freedom for total, within and between

Step 1:

Compute the grand mean of the Data $\bar{X} = \sum x_i / N$ (total number of values in the table)

Compute the mean for each column/group $\bar{x} = \sum x_i / n$ where n is number of elements in each group

Step 2:

Compute Total Sum of Square $SST = \sum (x_i - \bar{X})^2$ for $i=1$ to N

Step 3:

Between Sum of Squares $SSB = \sum [n * (\bar{x} - \bar{X})^2]$ where l is number of elements in each group]

Step 4:

Compute within sum of squares $SSE = \sum (x_i - \bar{x})^2$ where i is number of elements in each group and x bar is the mean of that corresponding group

Step 5:

Total Mean Squares $MST = SST / (N-1)$

Mean Square Between $MSB = SSB / (\text{number of groups} - 1)$

Mean Squares Within $MSE = SSE / (N - \text{number of groups})$

Steps 6:

$F\text{-calculated} = MSB / MSE$

F critical value – $df1$ (groups-1) , $df2$ (Total number of elements-number of groups) and alpha is our desired significance level

Chi-square test

Chi Square test is done in two scenarios:

1. Test of independence: There are two categorical attributes and we you like to find if there is any influence of one categorical attribute over the other.

Eg: Does a particular community is inclined towards a particular political party- So here we have two categories. Religion and Political party. We observe the frequencies for each combination to check if the hypothesis is true.

2. Goodness of fit: I have a particular distribution for an event (a theoretical model). I observed the event happening. Does the new values fit in my actual distribution?

Eg: if a chocolate company make a particular chocolate in 6 different colours. Do all colours have equal proportions for it.

Step 1:

Compute Expected values

For each value in the table we compute joint probabilities.

Type/group	A	B	Total
C	x		Row Total1
D		y	Row Total2
	Column total 1	Column total 2	Grand Total

Exp value for cell x = $(\text{Row total1} * \text{column total1}) / \text{Grand Total}$

Expected value for cell y = $(\text{Row total2} * \text{column total2}) / \text{Grand total}$

Step 2:

Chi-square value = $\sum (O-E)^2/E$

To compute Chisquare critical value

Degrees of freedom $Df = (\text{rows}-1) * (\text{columns}-1)$

Significance value is our desired value

Similarly for goodness of fit. We have the expected values generally in probabilities. From which we compute the frequencies for each level and perform computations using formulae in step 2.