IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

DESIGN AND IMPLEMENTATION OF ONE WAY ANALYSIS OF VARIANCE

Sanjib Das [*], Rashmi Ranjita Panda, Imkongyapang Jamir
[*] Department of IT & Mathematics, ICFAI University Nagaland, India
Department of IT & Mathematics, ICFAI University Nagaland, India
Department of IT & Mathematics, ICFAI University Nagaland, India

ABSTRACT

Analysis of variance (ANOVA) is the method used to compare continuous measurements to determine if the measurements are sampled from the same or different distributions. It is an analytical tool used to determine the significance of factors on measurements by looking at the relationship between a quantitative "response variable" and a proposed explanatory "factor." This method is similar to the process of comparing the statistical difference between two samples, in that it invokes the concept of hypothesis testing. Instead of comparing two samples, however, a variable is correlated with one or more explanatory factors, typically using the F-statistic. From this F-statistic, the P-value can be calculated to see if the difference is significant. For example, if the P-value is low (P-value<0.05 or P-value<0.01 - this depends on desired level of significance), then there is a low probability that the two groups are the same. The method is highly versatile in that it can be used to analyze complicated systems, with numerous variables and factors.

KEYWORDS: ANOVA, Single - Factor ANOVA (One-Way), Two - Factor ANOVA (Two-way), Degrees of Freedom, F – Distribution.

INTRODUCTION

ANOVA is a quantitative research method that tests hypotheses that are made about differences between two or more means. If independent estimates of variance can be obtained from the data, ANOVA compares the means of different groups by analyzing comparisons of variance estimates. There are two models for ANOVA, the fixed effects model, and the random effects model (in the latter, the treatments are not fixed).

ANOVA is a very useful technique for testing the equality of more than two means of population. The word analysis of variance is used because the technique involves first finding out the total variation among the observation in the collection data, then assigning causes of components of variation to various factors and finally drawing conclusion about the equality of means. It also used to test the significance of a regression equation as a whole i.e., whether all the equation are equal to zero.

Factor analysis is the process by which a complicated system of many variables is simplified by completely defining it with a smaller number of "factors." If these factors can be studied and determined, they can be used to predict the value of the variables in a system.

BACKGROUND OF THE STUDY

History

ANOVA was initially suggested by the British statistician Sir Ronald Aylmer Fisher in the 1920s. He coined the phrase "analysis of variance," defined as "the separation of variance ascribable to one group of causes from the variance ascribable to the other groups."1
Fisher was very interested in genetics. ANOVA uses Fisher's F-distribution as part of the test of statistical significance. Some of his famous papers include "On the mathematical foundations of theoretical statistics", published in the Philosophical Transactions of the Royal Society in 1922, and "Applications of Student's distribution" published in 1925.

Advantages and disadvantages of ANOVA

Advantages

- Robust design
- Increases statistical power

In addition a two way ANOVA

- Looks at interaction between factors
- Reduces random variability
- Can look at effect on second variable after controlling the first variable

Disadvantages

- If null hypothesis is rejected, we know at least one group differs from others, but with a one way ANOVA and multiple groups, it may be difficult to determine which group is different
- Assumptions need to be fulfilled

Procedure in hypothesis testing

There are five steps involved in testing a hypothesis:

1. Formulate a hypothesis: Here we have to set up two hypothesis instead of one, in such a way that if one hypothesis is true, the other is false. They are the null hypothesis and the alternative hypothesis.
2. Set up a suitable level of significance level: In this step we have to set up a level of significance with confidence for the purpose that a null hypothesis is rejected or accepted depends upon the significance level. A significance level at 5% means that in the long run, the risk of making mistake the wrong decision is about 5%.
3. Select text criterion: In this test, we have to select an appropriate statistical degree as a test criterion, when the sample size is more than 30 the Z-test is appropriate and when a sample size is less then equal to 30 T-test is appropriate.
4. Computation: In this step we have to apply all the related data collected before the application of the particular test.

It is the final step in hypothesis testing is to draw a statistical decision involving the acceptance or rejection of null hypothesis. This will depend on whether the computed value of the test criterion falls in the region of acceptance or in the region of rejection of given level of significance.

One way Analysis of Variance (ANOVA)

The analysis of variance is a partitioning of the total variance in a set of data into a number of component parts, so that the relative contributions of identifiable sources of variation to the total variation in measured responses can be determined. From this partition, suitable F-tests can be derived that allow differences between sets of means to be assessed.

Thus ANOVA is a bio-statistical method for determining whether a difference exists between the means of three or more independent populations. Expressed mathematically, it tests the null hypothesis- H_0: 41 = 42 = 43. The one-way ANOVA parametric test will result in either accepting or rejecting this null hypothesis. If we reject the null hypothesis, then we can conclude that the population means are not equal. We do not know however whether all the means are different from one another or only some of them are different. This additional specificity is determined by conducting multiple comparison procedures, i.e. additional statistical tests, it used to test general rather than specific differences among means.

Interpreting this output:

1. A one-way analysis is used to compare the populations for one variable or factor. In this instance the one variable is seating and there are 3 populations, also called group or factor levels being compared: front, middle and back.
2. DF stands for degrees of freedom. - The DF for the variable (e.g. Seating) is found by taking the number of group levels (called k) minus 1 (i.e. $k – 1$). - The DF for Error is found by taking the total sample size, N, minus k (i.e. $N – k$). - The DF for Total is found by $N – 1$.
3. The SS stands for Sum of Squares. The first SS is a measure of the variation in the data between the groups and for the Source lists the variable name (e.g. seating) used in the analysis. This is sometimes referred to as SSB for "Sum of Squares Between groups". The next value is the sum of squares for the error often called SSE or SSW for "Sum of Squares Within". Lastly, the value for Total is called SST (or sometimes SSTO) for "Sum of Squares Total". These values are additive, meaning SST = SSB + SSW.

4. The test statistic used for ANOVA is the F-statistic and is calculated by taking the Mean Square (MS) for the variable divided by the MS of the error (called Mean Square of the Error or MSE). The F-statistic will always be at least 0, meaning the F-statistic is always nonnegative. This F-statistic is a ratio of the variability *between* groups compared to the variability *within* the groups. If this ratio is large then the *p*-value is small producing a statistically significant result.(i.e. rejection of the null hypothesis)
5. The *p*-value is the probability of being greater than the F-statistic or simply the area to the right of the F-statistic, with the corresponding degrees of freedom for the group (number of group levels minus 1, or here $3 - 1 = 2$) and error (total sample size minus the number of group levels, or here $30 - 3 = 27$). The F-distribution is skewed to the right (i.e. positively skewed) so there is no symmetrical relationship such as those found with the *Z* or *t* distributions. This *p*-value is used to test the null hypothesis that all the group population means are equal versus the alternative that at least one is not equal. The alternative is not “they are not all equal."

Characteristics of ANOVA

- ✓ It is used in the analysis of comparative experiments, those in which only the difference in outcomes is of interest.
- ✓ The statistical significance of the experiment is determined by a ratio of two variances.
- ✓ The ratio is independent of several possible alterations to the experimental observations.
- ✓ Adding a constant to all observations does not alter significance.
- ✓ Multiplying all observations by a constant does not alter significance.

So ANOVA statistical significance result is independent of constant bias and scaling errors as well as the units used in expressing observations. In the era of mechanical calculation it was common to subtract a constant from all observations (when equivalent to dropping leading digits) to simplify data entry

Hypothesis for the One way ANOVA

The **null hypothesis** (H_0) tested in the One-way ANOVA is that the population means from which the *K* samples are selected are equal or that each of the group means is equal.

$H_0: \mu_1 = \mu_2 = \ldots = \mu_k$ Where *K* is the number of levels of the independent variable.

The **alternative hypothesis** (H_1) is that at least one group mean significantly differs from the other group means. Or – that at least two of the group means are significantly different.

$H_1: \mu_1 \neq \mu_2$

Assumptions

When using one-way analysis of variance, the process of looking up the resulting value of *F* in an *F*-distribution table, is proven to be reliable under the following assumptions

- Each sample of size n is drawn randomly and each sample is independent of the other sample
- The populations are normally distributed.
- The populations from which samples are drawn have equal variance. This means that:

$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \ldots\ldots\ldots\ldots = \sigma_k^2$, for populations

Logic behind Analysis of Variance (ANOVA)

The estimate of population variance σ^2 is computed by two different estimates of σ^2, each one by different method. One approach is to compute an estimate of σ^2 in such a manner that even if population means are not equal it will have no effect on the value of this estimator. This means that the differences in the values of the population means does not alter the value of σ^2 as calculated by a given method. This estimator of σ^2 is the average of the variance found within each of the samples. For example, if we take 10 samples of size n, then each sample will have a mean and a variance. Then the mean of these 10 variance would be considered as an unbiased estimator of σ^2 , the population variance, and its value remains appropriate irrespective of whether the population means are equal or not. This is really done by pooling all the sample variance to estimate a common population variance, which is the average of all the sample variance. This common variance is known as variance within samples or σ^2_{within}.

The second approach to calculate the estimate of σ^2 is based upon the central Limit Theorem and is valid only under the null hypothesis assumption that all the population means are equal. This means that in fact, if there is no difference among the population means, then the compound value of σ^2 by the second approach should not differ significantly from the computed value of σ^2 by the first approach.

Hence, *If these two values of σ^2 are approximately the same, then we can decide to accept the null hypothesis.*

ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785

The second approach results in the following computation.

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

or, the variance would be :

$$\sigma^2 = n\sigma_x^2$$

Thus by knowing the square of the standard error of the mean (σ_x^2), we could multiply it by n and obtain a precise estimate of σ^2. This approach of estimating σ^2, is known as $\sigma^2_{between}$. Now, if the null hypothesis is true, that is if all population means are equal then, value should be approximately the same as σ^2_{within} values. A significant difference between the two values would lead us to conclude that this difference is the result of difference between the population means. R.A.Fisher determined that the difference between the $\sigma^2_{between}$ and$\sigma^2_{between}$ σ^2_{within} values could be expressed as a ratio to be designated s the F-value, so that

$$F=\sigma^2_{between} / \sigma^2_{within}$$

In the above case, if the population means are exactly the same, then $\sigma^2_{between}$ will be equal to the σ^2_{within} and the value of F will be equal to 1. Accordingly, the larger the value of F, the more likely the decision to reject the null hypothesis. But how large the value of F be so as to reject the null hypothesis? The answer is that the computed value of F must be larger than the *critical* value of F, given in the table for a given level of significance and calculated number of degrees of freedom.

Degrees of Freedom

The F-distribution being a family of curves, each curve reflecting the degrees of freedom relative to both $\sigma^2_{between}$ and σ^2_{within}. This means that the degrees of freedom are associated both with the numerator as well as with the denominator of the F-ratio.

1. The numerator. Since the variance between samples $\sigma^2_{between}$ comes from many samples and if there are K number of samples, then the degrees of freedom, associated with the numerator would be (K-1)
2. The denominator is the *mean variance* of the various of K samples and since each variance in each sample is associated with the size of sample (n), then the degrees of freedom associated with each sample would be(n-1). Hence, the total degrees of freedom would be the sum of degrees of freedom of K samples or

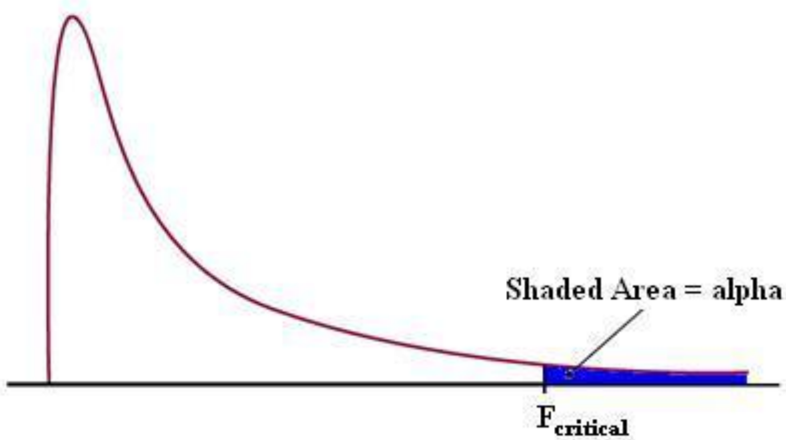
   Df=k(n - 1), when each sample is of size n.

The F-Distribution

The major characteristics of the F-distribution are as follows:

1. Unlike normal distribution, which is only one type of curve irrespective of the value of the mean and the standard deviation, the F distribution is the family of curves. A particular curve is determined by two parameters. These are the degrees of freedom in the numerator and the degrees of freedom in the denominator. The shape of the curve changes as the number of degrees of freedom change
2. It is continuous distribution and the value of F cannot be negative
3. The curve representing the F distribution is positively skewed.
4. The values of F theoretically range from zero to infinity.

A diagram of F distribution curve is shown below.



The rejection region is only in the right tail of the curve because unlike Z distribution and t distribution which had a negative values for areas below the mean, F distribution has only positive values by definition and only positive values of F that are larger than the critical values of F will lead to a decision to reject the null hypothesis.

Computation of F One way ANOVA

Since F ratio contains only two elements, which are the variance between the samples and the variance within the samples, the concepts which have been discussed before, let us recapitulate the calculation of these variance.

If all the means of samples were exactly equal and all samples were exactly representative of their respective populations so that all sample means were exactly equal to each other and to the population mean, then there will be no variance. However, this can never be the case. We always have variation both between samples and within samples, even if we take these samples randomly and from the same population.

This variation is known as the total variation. The total variation designated by $\sum(X - \bar{\bar{X}})^2$, where, X represents individual observation for all samples and $\bar{\bar{X}}$ is the grand mean of all sample means and equals (μ), the population mean, is also known as *the total sum of square* or SST.

Variance between samples

The variance between samples may be due to the effect of different *treatments*, meanings that the population means may be affected by the *factor* under consideration, thus making the population means actually different, and some variance may be due to the inter-sample variability. This variance is also known as the *sum of squares* between samples.

Then SSB is calculated by following steps:

1. Take k sample of size n each and calculate the mean of each simple i.e., $\bar{X}_{1,}\ \bar{X}_{2,}\ \bar{X}_{3,\ldots\ldots}\ \bar{X}_{k.}$
2. Calculate the grand mean $\bar{\bar{X}}$ of the distributed of the distribution of these sample means, so that:
$$\bar{\bar{X}} = \sum_{i=1}^{n} \bar{X}_1$$
3. Take the difference between the means of the various sample and the grand mean, i.e.,($\bar{X}_1 - \bar{\bar{X}}$), ($\bar{X}_2 - \bar{\bar{X}}$), ($\bar{X}_3 - \bar{\bar{X}}$),……, ($\bar{X}_k - \bar{\bar{X}}$)
4. Square these deviation or differences individually, multiply each of these squared deviations by its respective sample size and sum up all these products, so that we get: $\sum_{i=1}^{k} n_i(\bar{X}_I - \bar{\bar{X}})^2$, where n_i = size of i^{th} sample
This will be the value of SSB.
However, if the individual observation of all samples are not available, and only the various means of these samples are available, where the sample are either of the same size n or different sizes, n_1, n_2, n_3,…..n_k, then the value of SSB can be calculated as:
$$SSB = n_1(\bar{X}_I - \bar{\bar{X}})^2 + n_2(\bar{X}_2 - \bar{\bar{X}})^2 + n_3(\bar{X}_3 - \bar{\bar{X}})^2 + \ldots\ldots n_n(\bar{X}_n - \bar{\bar{X}})^2$$

Where:
n_1=number of items in sample1
n_2 = number of items in sample 2
n_k=number of items in sample k
$\bar{X}_1$=mean of sample 1
$\bar{X}_2$=mean of sample 2
$\bar{X}_k$=mean of sample k
$\bar{\bar{X}}$ =Grand mean or average of all items in all samples.

5. Divide SSB by the degrees of freedom, which are(k - 1), where k is the number of samples and this would give the value of $\sigma^2_{between}$, so that:

$$\sigma^2_{between} = \frac{SSB}{(k-1)}$$

(This is also known as mean square between samples or MSB)

Variation within samples

Even though each observation in a given sample comes from the same population and it's subjected to same treatment, some chance variation can still occur. This variation may be due to sampling errors or other natural causes. This variance or sum of squares is calculated through the following steps:

1. Calculate the mean value of each sample, i.e., $\bar{X}_1$, $\bar{X}_2$, $\bar{X}_3$,….. $\bar{X}_k$.
2. Take one sample at a time and take the deviation of each item in the sample from its mean.
3. Square these differences and take a total sum of all these squared differences (or deviations). This sum is also known as SSW or sum of squares within samples.
4. Divide the SSW by the corresponding degrees of freedom. df=(N-k), where N is the total number of items or observation, and k is the number of samples.
5. This figure SSW/df is also known as σ^2_{within}, or MSW (mean of sum of square within samples). Now the value of F can be computed as:

$$F = \frac{SSB/(k-1)}{SSW/(N-k)} = \frac{MSB}{MSW}$$

This value F is then compared with the critical value of F from the table and a decision is made about the validity of null hypothesis.

ANOVA Table:

After various calculations for SSB, SSW and the degrees of freedom have been made; these figures can be presented in a simple table called *Analysis of Variance* table or simply ANOVA table, as follows:

Source of variation	Sum of Squares	Degrees of freedom	Mean squares	F
Treatment	SSB	(k - 1)	MSB=SSB/(k - 1)	$\frac{MSB}{MSW}$
Within	SSW	(N - 1)	MSW=SSW/(N - k)	
Total	SST			

A short cut – Method

This formula developed above for the computation of the value of F-statistic is rather complex and time consuming when we have to calculate the variance between samples and the variance within samples. However, a short-cut simpler method for these *sum of squares* is available, which considerably reduces the computational work. This technique is used through the following steps.

ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785

1. Take the sum of all the observation of all the samples, either by adding all the individual values, or multiplying the mean of each sample ny its size and then adding up all these products as follows:
   The total sum (TS)=$n_1 \bar{X}_1 + n_2 \bar{X}_2 + \ldots\ldots n_k \bar{X}_k$, for k samples
2. Calculate the value of a *correction factor*. The correction factor (CF) value is obtained by squaring the total sum obtained above and dividing it by the total number of observation N, so that:

$$CF = \frac{(TS)^2}{N}$$

3. The total sum of squares is obtained by squaring all individual observations of all samples, summing up these values and subtracting from this sum, the correction factor(CF)

$$\text{Total sum of squares(SST)} = \sum X_1^2 + \sum X_2^2 + \ldots + \sum X_k^2 - CF$$

Where,

$\sum X^2$ = summation of squares for all X's in each samples.

4. The sum of squares between the samples (SSB) is obtained by the following formulas

$$SSB = \frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \ldots\ldots - CF$$

Where,
$(\sum X_1)^2$ = Squares of the total of all values in sample1.
$(\sum X_1)^2$ = Squares of the total of all values in sample2.

5. Then sum of squares within samples SSW can be calculated as:
   SSW=Total sum of squares – the sum of squares between samples=SST-SSB
6. The rest of the procedure will be same as previous method.

Solving problems using one way ANOVA

In this topic we will explain how to solve the problem step by step by giving necessary guide lines and explain how to calculate the f-test and Table value for F-distribution.

Now we will try to solve a problem by using the ANOVA logic to decide whether we will accept or reject the null hypothesis.

#Example:

To test whether all the professors teach the same material in different sections of the introductory statistics class or not, 4 sections of the same course were selected and a common test was administered to 5 students selected at random from each section. The scores for each student from each section where noted and are given below. We want to test for any difference in learning, as reflected in the average scores for each section.

Student#	Section 1(x_1)	Section 2(x_2)	Section 3(x_3)	Sectio4 (x_4)
1	8	12	10	12
2	10	12	13	15
3	12	10	11	13
4	10	8	12	10
5	5	13	14	10

The above given table of observations are solve using the *Short- cut Method* for finding the F-ratio.

*State the null hypothesis. We are assuming that there is no significance difference among the average scores of student from 4 sections and hence all professors are teaching the same material with the same effectiveness, i.e.

H_0: $\mu_1=\mu_2=\mu_3=\mu_4$

H_1: All means are not equal or at least two means differ from each other

*Establish a level of significance. Let $\alpha = 0.05$.

a. Total sum (TS)=$\sum X$ = 220

b. Correction factor=$(TS)^2/N = (220)^2/20 = 2420$

c. Total sum of squares (SST) = $\sum(X)^2$ - CF = 2522- 2420 = 102

d. Sum of squares between the samples SSB is obtained by:

$$SSB = \sum_{i=1}^{k} \frac{(X_i)^2}{n_i} - CF$$

ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785

$$= \frac{(X_1)^2}{n_1} + \frac{(X_2)^2}{n_2} + \frac{(X_3)^2}{n_3} + \frac{(X_4)^2}{n_4} - CF$$

$$= \frac{(45)^2}{5} + \frac{(55)^2}{5} + \frac{(60)^2}{5} + \frac{(60)^2}{5} - (2420)$$

$$= 405 + 605 + 720 + 720 - 2420$$

$$= 30$$

e. SSW can be calculated by:

SST – SSB= 102- 30 = 72

Now the F value can be calculated as:

$$F = \frac{SSB/df}{SSW/df}$$

$$= \frac{30/(k-1)}{27/(N-k)}$$

$$= (30/3)/(72/16)$$

$$= (10/4.5)$$

$$= 2.22$$

Now, we compare the value of F from the table for α=0.05 and df(numerator = 3), and df(denominator = 16), we get the critical value of F as 3.24.

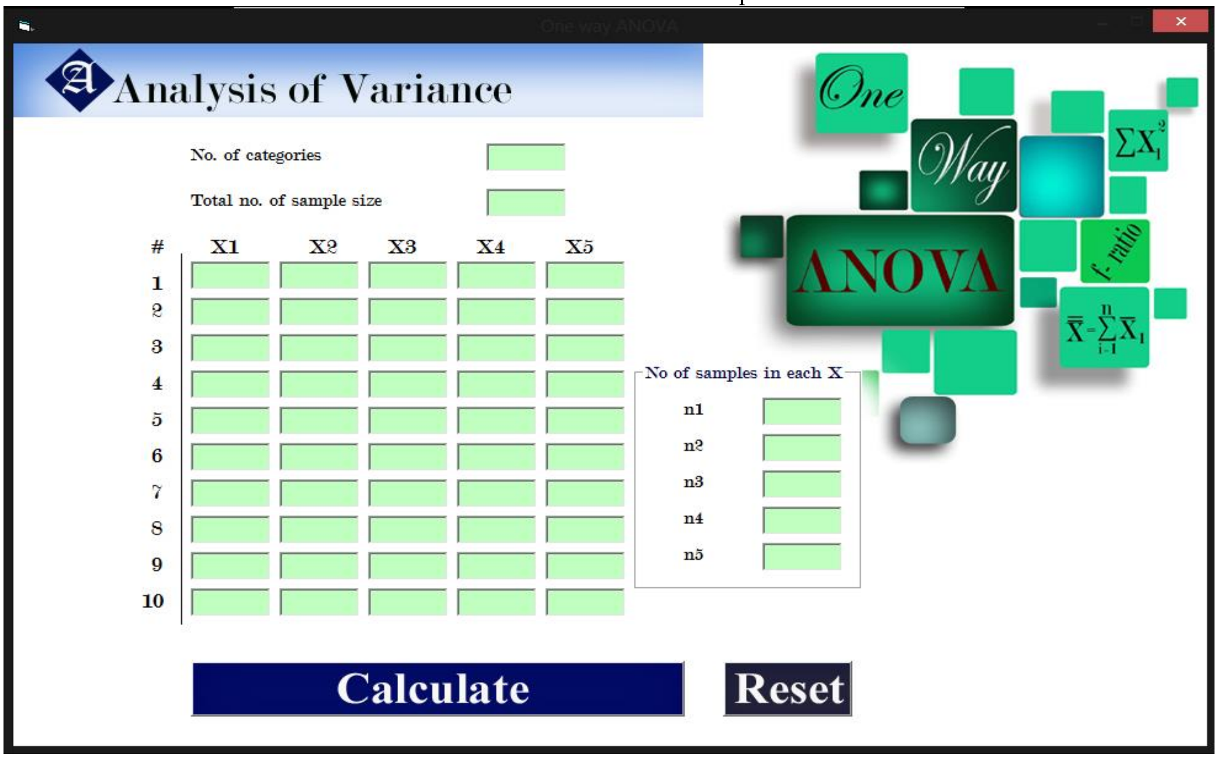
Since the calculated value is less than the table value we accept the null hypothesis and reject the alternative hypothesis due to less evidence.

ANOVA Table

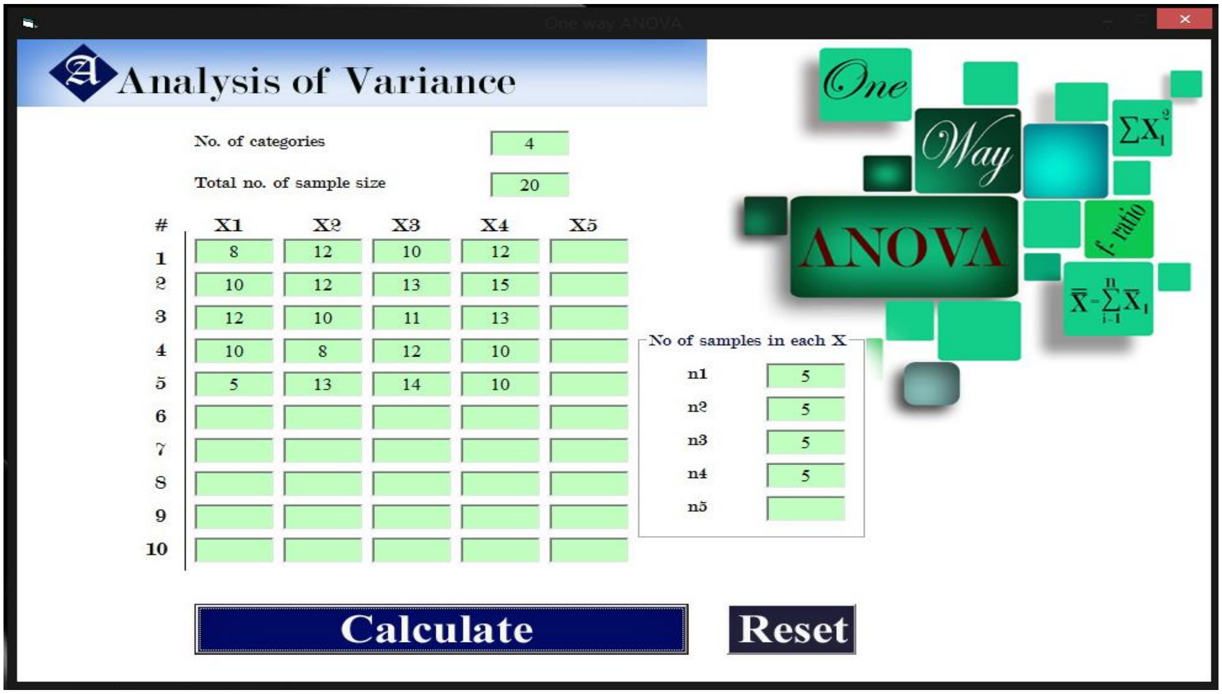
Source of variation	Sum of Squares	Degrees of freedom	Mean square	F
Treatment	SSB=30	(k-1)=3	MSB=10	2.22
Within (or error)	SSW= 72	(N-k)=16	MSW=4.45	
Total	SST=102			

RESULT & PERFORMANCE EVALUATION

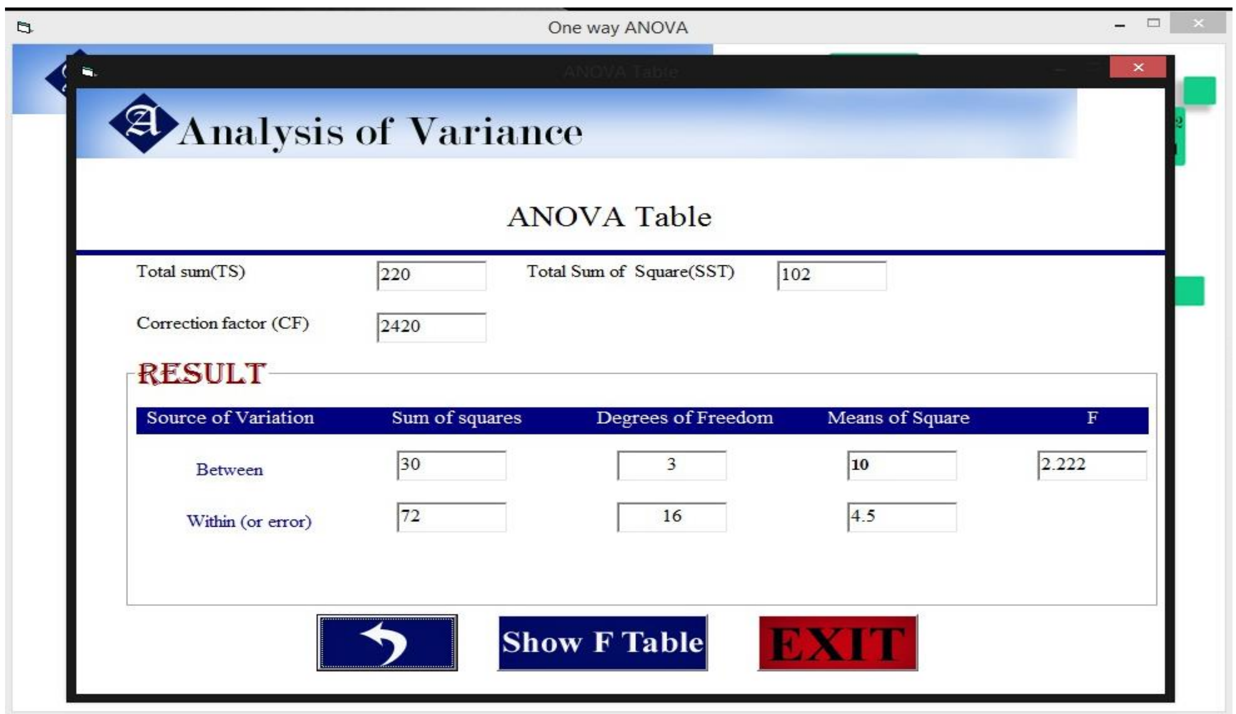
The screenshot shows below are the main screen where the user will input the data's



The screenshot shows below after the user gives the input

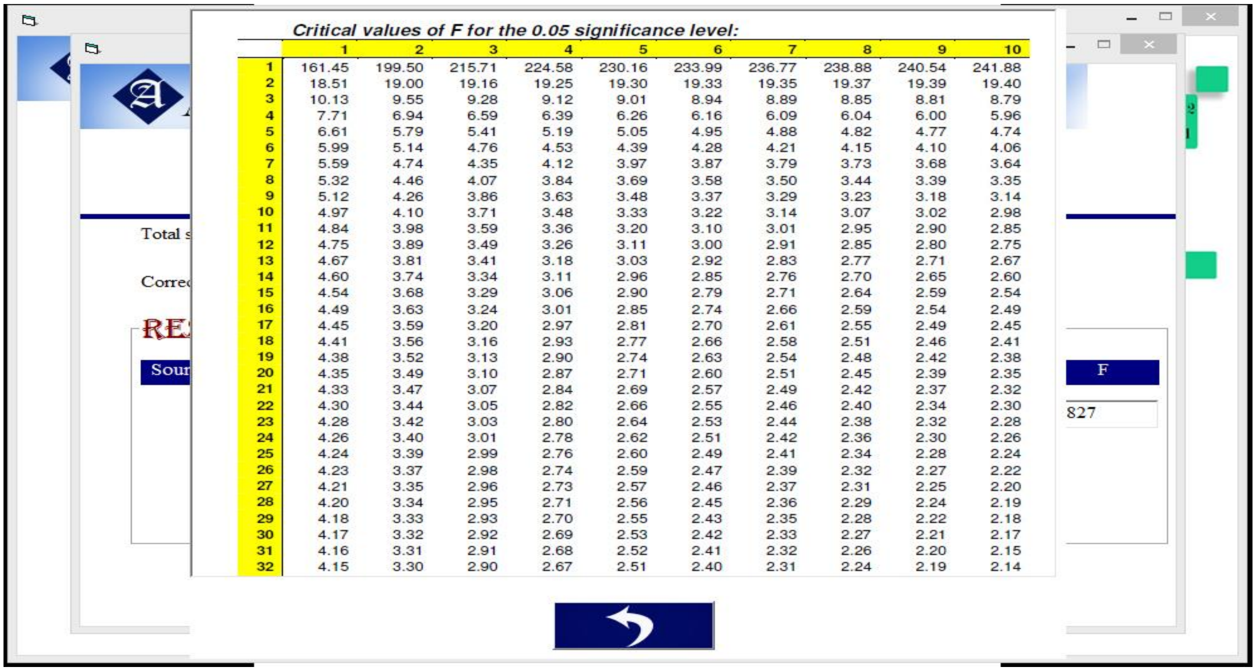


When the user click on the CALCULATE button it will given the output as shown below



The screenshot shows below the table value when the user click the back button

ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785



Critical values of F for the 0.05 significance level:

	1	2	3	4	5	6	7	8	9	10
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.39	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.97	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.10	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.97	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.56	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.33	3.47	3.07	2.84	2.69	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.38	2.32	2.28
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.26
25	4.24	3.39	2.99	2.76	2.60	2.49	2.41	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.17
31	4.16	3.31	2.91	2.68	2.52	2.41	2.32	2.26	2.20	2.15
32	4.15	3.30	2.90	2.67	2.51	2.40	2.31	2.24	2.19	2.14

CONCLUSION

The Design and Implementation of **One way Analysis of Variance (ANOVA)** is a software program developed by using Microsoft visual studio 2006, which is user friendly environment. It is a Graphical User Interface (GUI) which would make the user more comfortable to operate the application. This software can calculate some input data of a limited number of sample size and categories gives the final result of F and also the different elements output in a table. And also the user can view the F-table after the final result is computed.

With limited number of days for developing such application it takes lots of hard work, energy and logical concepts. Though the software has been completed, the inputs are limited in number of categories (5) and sample size (10), and the table value can be viewed only in the next form. The user can refer the table value in the other form which means it is not calculated automatically when the final result is executed.

REFERENCES

1. Statistics for Business and Economics by JIT S CHANDAN
2. Software Engineering by Rajib Mall
3. https://www.calvin.edu/~scofield/courses/m143/materials/handouts/anova1And2.pdf
4. https://www.ischool.utexas.edu/~wyllys/IRLISMaterials/anova.pdf
5. http://www.csun.edu/~amarenco/Fcs%20682/When%20to%20use%20what%20test.pdf
6. https://statistics.laerd.com/statistical-guides/one-way-anova-statistical-guide.php
7. http://ron.dotsch.org/degrees-of-freedom/
8. https://en.wikipedia.org/wiki/Degrees_of_freedom_(statistics)
9. http://www.differencebetween.net/miscellaneous/difference-between-t-test-and-anova/
10. https://controls.engin.umich.edu/wiki/index.php/Factor_analysis_and_ANOVA
11. http://www.vbforums.com/showthread.php?731995-Help-how-can-i-define-a-square-number-using-Visual-basic-language
12. https://social.msdn.microsoft.com/Forums/en-US/3da5c3bc-e49b-4353-82f0-e87a1855528b/how-to-calculate-average-of-values-in-vbnet?forum=vblanguage
13. http://forums.codeguru.com/showthread.php?379936-How-to-make-text-box-accept-numerical-value-only
14. http://stackoverflow.com/questions/18862674/how-to-create-an-error-message-box-for-an-empty-text-box-vb-2010-express
15. http://en.wikipedia.org/wiki/test_plan

AUTHOR BIBLIOGRAPHY

SANJIB DAS
Assistant Professor
Deptt. of IT & Mathematics
ICFAI University Nagaland
Sovima, 6th Mile, Dimapur – 797112, Nagaland
dassanjib1990@gmail.com

RASHMI RANJITA PANDA
Senior Lecturer
Deptt. of IT & Mathematics
ICFAI University Nagaland
Sovima, 6th Mile, Dimapur – 797112, Nagaland
rashmiranjita2006@gmail.com

IMKONGYAPANG JAMIR
Deptt. of IT & Mathematics
ICFAI University Nagaland
Sovima, 6th Mile, Dimapur – 797112, Nagaland