Mikailalsys Journal of
Mathematics and Statistics

e-ISSN : 3030-816X
p-ISSN : 3030-8399
Index : Harvard, Boston, Sydney University, Dimensions, Lens, Scilit, Semantic, Google, etc

https://doi.org/10.58578/MJMS.v2i3.3449

F-Test and Analysis of Variance (ANOVA) in Economics

Nand kishor Kumar
Trichandra Campus, Tribhuvan University, Nepal
nandkishorkumar2025@gmail.com

Article Info:

Submitted:	Revised:	Accepted:	Published:
Aug 1, 2024	Aug 17, 2024	Aug 20, 2024	Aug 23, 2024

Abstract

ANOVA remains a cornerstone of empirical economic research, providing economists with a robust framework to analyze differences between groups, evaluate policy interventions, and draw meaningful conclusions from data. Its versatility and applicability across diverse economic contexts underscore its significance in advancing economic theory and informing evidence-based policymaking. As data availability and computational capabilities continue to expand, ANOVA's role in economic analysis is expected to evolve, supporting increasingly sophisticated studies of economic phenomena and policy impacts. In economics, particularly in empirical research and data analysis, the F-test and Analysis of Variance (ANOVA) are fundamental statistical tools used to test hypotheses regarding the equality of means across two or more groups.

Key-words: Fest-test, Analysis of Variance, Time Series Analysis, Hypothesis

Introduction

In the field of economics, the F-test and Analysis of Variance (ANOVA) are powerful statistical tools used to analyze differences between groups and to make inferences about population parameters based on sample data. It is a hypothesis test that is parametric. It is a methodical estimating process used to verify the equality of two or more population means (Aigner,1971).

The rule of total variance, which divides the observed variance in a given variable into components attributed to various causes of variation, is the basis of the ANOVA. By offering an analytical examination of the variations in group means, ANOVA expands the scope of the t-test beyond two means. F-tests are used in ANOVA to statistically test for mean equality (Bridge,1971).

Understanding Variance

In many scientific fields, particularly statistical research, variance is a crucial tool. Variance is the expectation of a random variable's squared divergence from its mean in statistics and probability theory. In actuality, it is assessed to determine how widely distributed the series' data are around the average value. In statistics, variance is frequently employed in a variety of contexts, including statistical inference, hypothesis testing, and descriptive statistics (Cramer, 1969).

Objective

The objective of this article is to give a review of the use of analysis of variance (ANOVA) in the study of economics.

Economic Significance of ANOVA

ANOVA is a fundamental statistical technique that is used to compare means between different groups and test the equality hypothesis. Within the field of economics, this turns into a potent technique for evaluating policy efficacy, examining market segmentation, and investigating the economic effects of diverse elements across multiple populations or historical periods.

ANOVA offers a systematic method for examining economic issues, whether comparing the GDP growth rates of nations using various economic models or examining consumer buying trends across demographic groups (Christ, 1966).

Pertinence to the Analysis of Economic Data

Because of its richness and diversity, economic data frequently calls for a sophisticated analytical strategy. Economists may test theories and models against actual data using ANOVA, which makes it easier to draw conclusions that are both statistically significant and economically useful. This technique serves as a basis for evidence-based economic policymaking and strategic business choices. It also helps with hypothesis testing, model construction, and model refining.

ANOVA in Economics

In the field of economics, analysis of variance (ANOVA) is a vital technique that provides a statistical framework for testing hypotheses on differences in group means. It has numerous and diverse uses in economics, ranging from consumer behavior study to policy efficacy. This section highlights a number of significant economics use cases and applications for ANOVA, highlighting its adaptability and significance for economic study and analysis (Desai, 1976).

Key applications of ANOVA in Economics:

1. **Comparing Means Across Multiple Groups**: ANOVA allows economists to compare the means of more than two groups simultaneously. For example, it can be used to analyze whether there are differences in income levels across different sectors of the economy or across different regions.
2. **Experimental Studies**: In experimental economics, ANOVA is used to analyze the results of experiments involving multiple treatments or conditions. It helps researchers determine whether there are statistically significant differences in outcomes between experimental groups.
3. **Time Series Analysis**: ANOVA techniques can also be adapted for time series data to analyze variations in economic indicators over time and across different time periods.

F-Test

The designation "F-test" comes from Sir Ronald Fisher's name. One may express the F-statistic as a ratio of two variances. The square of the standard deviation equals variance. Because standard deviations are expressed in the same units as the data rather than squared units, they are simpler for the average person to grasp than variances. The ratio of mean squares is the foundation of F-statistics. Although the word "mean squares" may sound technical, it just refers to a population variance estimate that takes the degrees of freedom (DF) into consideration.

F – test or variance ratio test is usually applied to test the significant difference between the variances of independent normal populations. It can also be applied to test whether the two samples are taken from a. normal population having same variance or not (Halder, 2024).

Let $x_1, x_2 .... x_{n_1}$ and $y_1, y_2 ... y_{n_2}$ be two samples drawn from two independent normal populations Then, under null hypothesis;

$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2,$ the F – statistics is given by

$$F = \frac{S_1^2}{S_2^2} \sim F(n_1 - 1)(n_2 - 1)$$

$$S_1^2 = \frac{1}{n_1 - 1} \Sigma (x - \bar{x})^2 = \frac{1}{n_1 - 1} \left[\Sigma x^2 - \frac{(\Sigma x)^2}{n_1}\right] = \frac{(x_1 - \overline{x_1})^2}{n_1 - 1}$$

$$S_2^2 = \frac{1}{n_2 - 1} \Sigma (y - \bar{y})^2 = \frac{1}{n_2 - 1} \left[\Sigma y^2 - \frac{(\Sigma y)^2}{n_2}\right] = \frac{(x_2 - \overline{x_2})^2}{n_2 - 1}$$

Now, $S_1^2 > S_2^2$

and $F(n_1 - 1)\ (n_2 - 1)$ implies F - distribution with $(n_1 - 1$ and $(n_2 - 1)$ degrees of freedom.

$$F = \frac{\text{Larger estimat of variance}}{\text{Smaller estimate of variance}}$$

Example 1: Two random samples of size 8 and 7 drawn from normal distributions had the following values of variables.

Sample X 8 10 11 15 13 10 17 12

Sample Y 12 10 14 16 11 15 13

Do the two population have the same variance at the 5 % level of significance.

Solution:

$H_0 : {\sigma_x}^2 = {\sigma_y}^2$ i.e., The two population have same variance.

$H_1 : {\sigma_x}^2 \neq {\sigma_y}^2$ i.e., The two population do not have same variance. The given level of significance 5 %

Test statistics : To compare the difference in variance, then F-test is used.

Calculation: $F = \frac{\text{Larger estimat of variance}}{\text{Smaller estimate of variance}}$

X	(x- $\bar{x}$)	$(x-\bar{x})^2$	y	(y- $\bar{y}$)	$(y-\bar{y})^2$
8	-4	16	12	-1	1
10	-2	4	10	-3	9
11	-1	1	14	1	1
15	3	9	16	3	9
13	1	1	11	-2	4
10	-2	4	15	2	4
17	5	25	13	0	0
12	0	0			
92		60	92		28

$\bar{x} = \frac{\sum x}{n_1} = \frac{92}{8} = 12$ and $\bar{y} = \frac{\sum y}{n_1} = \frac{91}{7} = 13$

$${S_x}^2 = \frac{1}{n_1-1}\sum(x-\bar{x})^2 = \frac{60}{8-1} = 8.57$$

$${S_y}^2 = \frac{1}{n_2-1}\sum(y-\bar{y})^2 = \frac{28}{7-1} = 4.67$$

So, $\frac{{S_x}^2}{{S_y}^2} \sim F(n_1 - 1, n_2 - 2) = \frac{8.571}{4.667} = 1.836$

The tabulated value of $F_{0.05}\ v_1 = 7$, and $v_2 = 6 = 4.21$

As calculated F < tabulated F, null hypothesis is accepted. This means that both of the populations have same variance.

Analysis of Variance

Analysis of Variance is of great use in following cases:

(i) If there are more than two samples

(ii) If there are more than one factor affecting the response variable of interest

In ANOVA, the total variance is split up into two causes viz (i) assignable cause and (ii) chance cause. Thus, Total variance due to assignable cause variance due to chance cause. The variance due to chance cause cannot be controlled & is called error variance. On the other hand, the variance due to assignable cause can be controlled and this variance is further split up into various number of variances based upon the total number of factors influencing the response variable of interest (Mehata, 2005).

On this basis, it can be concluded that ANOVA is classified into one way, two way, three way and so-on, depending upon the number of factors affecting to the response variable of interest.

Assumptions of ANOVA

(i) The observations, say x, are independent.

(ii) The parent population from which observations are taken is normal.

(iii) Various effects are additive in nature.

One way- ANOVA

In one way ANOVA, only one factor is considered to be affecting the response variable of interest. For example, suppose the agricultural production depends upon the seed variety only. This means other factors are not involved in the process of inferences. Now we can only take the various types of the seed i.e., usually more than two types and test the difference between the average production in different cases. Thus, one way ANOVA can be computed by following two methods.

1. Direct method

2. Short-cut method

Since direct method is some lengthy and tedious so short-cut method is demonstrated in this paper.

Short-cut Method

Compute correction factor C.F. = $\frac{T^2}{N}$

Where, T = $\sum_{i=1}^{k}(\sum x_i)$ = Grand Total and N = $n_1 + n_2 + \cdots\cdots\cdots + n_k = \sum_{i=1}^{k} n_i$

Computing SSC = $\frac{(\sum x_1)^2}{n_1} + \frac{(\sum x_2)^2}{n_3} + \cdots\cdots\cdots\cdots + \frac{(\sum x_k)^2}{n_k} - \frac{T^2}{N}$

Computing TSS = $\sum {x_1}^2 + \sum {x_2}^2 + \cdots\cdots + \sum {x_k}^2 - \frac{T^2}{N}$. Now computing

SSE = SST – SSC

Example 2. Find the grades given by three professors in an advanced course are follows:

A: 63 45 73 77 72

B: 67 45 76 80 70 70

C: 97 97 87 87 84 74 64

Does the difference in sample means appear to be due to chance variation or is there sufficient evidence to indicate that not all population means are the same? Use 5 % level of significance.

Solution: Now setting hypothesis

$H_0 : \mu_A = \mu_B = \mu_C$ i.e. The difference in sample mean is due to chance variation.

$H_1 : \mu_A \neq \mu_B \neq \mu_C$ i.e. The population means are not same. So, the difference in sample mean is not due to chance variation. The level of significance is given 5% and test statistics is F- test. Now, computation:

For the simplicity, the data is coded by 70. Let A = x_1, B = x_2 and C = x_3.

x_1 x_3^2	x_1^2	x_2	x_2^2	x_3	
-7	49	-3	9	27	729
-25	625	-25	625	27	729
3	9	6	36	17	289
7	49	10	100	17	289
2	4	0	0	14	196
		0	0	4	16
				-6	36
-20	736	-16	720	100	2284

Computing Correction Factor

C.F. = $\frac{T^2}{N}$, where T = -20-16 +100 = 64, so C.F. = $\frac{(64)^2}{18}$ = 227.556

Now, computing SSC = $\frac{(\Sigma x_1)^2}{n_1} + \frac{(\Sigma x_2)^2}{n_3} + \cdots\cdots\cdots\cdots + \frac{(\Sigma x_k)^2}{n_k} - \frac{T^2}{N}$

$= \frac{(-20)^2}{5} + \frac{(-16)^2}{6} + \frac{(100)^2}{7}$ - 227.556

= 1323.6854, Now computing TSS

TSS = $= \sum x_1{}^2 + \sum x_2{}^2 + \cdots \cdots + \sum x_k{}^2$ - $\frac{T^2}{N}$ = 736+ 720+ 2284- 227.556 = 2712.4444

SSE = SST – SSC = 3512.44- 1323.6854 = 2188.759. Hence the ANOVA Table is

Source of variation	Sum of square	Degree of freedom	Mean square	F-Value
Between Sample SSC	1323.685	12	MSC= 1323.68=661.84	F= $\frac{MSC}{MSE}$= 453
Within Sample SSE	2188.76	15	MSE=2188.76=145.9	
Total	35412.444	17		

Tabulated value of $F_{0.05}$ at v_1 = 2 and v_2= 15 is 3.68. Since calculated F >Tabulated F, so null hypothesis is rejected. This means the difference of sample means is not due to chance variation.

Two way analysis of variance

In a situation, when the response variable of interest is affected by two factors, two-way ANOVA analysis is performed. For example, if the quantity demanded of any particular good is taken as affected by price of that good and the income of the consumer, there are now two factors and hence we perform the two-way ANOVA. Thus, the total sum of squares of variation, now, becomes the addition of three factors as follows.

TSS = SSC+ SSR + SSE

Where, TSS = Total sum of square

SSC = Sum of squares of variation in columns

SSR = Sum of squares of variation in Rows

SSE = Sum of squares of variation due residual

It should be noted that in two ways ANOVA, we obtain two F- values. As in ANOVA in one way classification, the ANOVA in two way classification can be carried out into two ways i.e.

(i) Direct method

(ii) Short-cut method

The procedure for computing ANOVA for two way classification by the short-cut method is as follows:

Computing correlation factor

C.F. = $\frac{T^2}{N}$ where T = Grand total = $\sum_{i=1}^{k}\sum_{j=1}^{C} x_{ij}$

It means the grand total taken from row as were or column wise are same.

N = Total number of observations.

T = $x_{c_1} + x_{c_2} + \cdots\cdots\cdots + x_{c_r}$

Now, computing SSC,

SSC = $\frac{\Sigma(\Sigma x_c)^2}{r}$, where r is number of rows. Computing SSR

SSR = $\frac{\Sigma(\Sigma x_r)^2}{c}$, where c is number of columns. Computing TSS

TSS= $\sum\sum x_{ij}{}^2$ - $\frac{T^2}{N}$. Now computing SSE = SST-(SSC+SSR)

Example 3: A tea company appoints salesman, A, B, C and D and observes their sales in three seasons-summer, winter and monsoon. The table are given below;

Seasons	Salesman				
	A	B	C	D	Total
Summer	36	36	21	35	128
Winter	28	29	31	32	120
Monsoon	26	28	29	29	112
Total	90	93	81	96	360

Find out if there is difference in the sales of different salesmen and seasons.

$F_{0.05}$ for 3 and 6 d. f. = 4.76

$F_{0.05}$ for 2 and 6.d.f = 5.14

Solution:

(a) Setting hypothesis :

H_o : (i) $\mu_A = \mu_B = \mu_C = \mu_D$ i.e., There is no significant difference in the mean sales made by 4 salesmen .

(ii) $\mu_S = \mu_W = \mu_M$ i.e., There is no difference in the sales made during three different seasons.

$H_1; (i) = \mu_A \neq \mu_B \neq \mu_C \neq \mu_D$

(ii) $\mu_S \neq \mu_W \neq \mu_M$ i.e ., there is significant difference in mean sale made by 4 salesmen as well as during there difference season.

(b) Level of significance : 5 percent

(c) Test statistics : F- test

(d) Computation: Since there are two – factors affecting the sales i.e., sales men and season, this is two – way ANOVA. For simplicity we code the data and apply the short cut method.

Now, coding the data by 30 and obtain the table as below.

Seasons	A	B	C	D	Seasons total
Summer	6	6	-9	+5	8
Winter	-2	-1	+1	2	0
Monson	-4	-2	-1	-1	-8
Salesman's total	0	3	-9	6	0

Correction factor $= \frac{T^2}{N} = \frac{0^2}{12} = 0$

SSC $= \frac{\Sigma(\Sigma_{x_c})^2}{r} - \frac{T^2}{N}$

$= \frac{(0)^2+(3)^2+(-9)^2+(6)^2}{3} - \frac{0^2}{12}$

∴ SSC = 42 with d. f. (c – 1) = (4- 1) = 3

Similarly calculating SSR

SSR $= \frac{\Sigma(\Sigma x_r)^2}{c} - \frac{T^2}{N} = \frac{(8)^2+(0)^2+(-8)^2}{4} - 0$

∴ SSR 32 with d. f. (r-1) = 2

Then, TSS = $\Sigma\Sigma_{x_{ij}} 2$ - $\frac{T^2}{N} = (\text{x}11^2 + \cdots + x^2 tc) - \frac{T^2}{N}$

$= [6^2 + (-2)^2 + (-4)^2 + (6)^2 + (-1^2 + (-2)^2 + (-9)^2 + (1)^2$

$+ (-1)^2 + (5)^2 + (2)^2 + (-1)^2] - \frac{0^2}{12}$ ∴ TSS = 210 with d. f = 12-1 = 11

Therefore, we can estimate SSE as ,

SSE = SST - (SSC+ SSR) = 210 – (42 + 32) = 136 with d. f. = (r- 1) (c-1) = 6

Then, the ANOVA table is as:

Sources of variation	Sum of square	d. f.	Mean square	F-value
Between columns Salesman	42	3	14	$\frac{22.67}{14}$= 1.62
Between rows Seasons	32	2	16	$\frac{22.67}{16}$= 1.42
Residual error	136	6	22.67	
Total	210	10		

Decision Making: since we have to calculated F values they are

(i) F = $\frac{MSE}{MSC}$ = 1.62 with d.f. $v_1 = 6$ & $v_2 = 3$

For this the tabulated value of Fo.05 at v_1 = 6 and $v_3 = 3$ degree of freedom is 8.94. As calculated F <tabulated F, null hypothesis is accepted. This means there is significant difference in the average sales of salesman A, B, C and D.

F = $\frac{MSE}{MSR}$ = 1.42 with d. f. v_1 = 6 & v_2 = 12

For this, the tabulated value of $F_{0.05}$ at v_1 = 6 and v_2 = 2 is 19.33. Hence, it also shows that null hypothesis is accepted. This means that there is no significant difference in the average sales in different seasons.

Practical Example in Economics

Imagine an economist conducting a study on the effectiveness of different monetary policies across several regions. They collect GDP growth rates from ten regions over five years. To determine if there are significant differences in growth rates between regions due to different policies, the economist can use ANOVA. The F-test derived from ANOVA will indicate whether there are statistically significant differences in growth rates among the regions.

Conclusion

In conclusion, the F-test and ANOVA are indispensable tools in economic analysis, providing rigorous methods to test hypotheses about group means and understand the relationships within economic data. By applying these statistical techniques, economists can draw reliable conclusions about the effectiveness of policies, the impact of interventions, and the differences between economic variables across different groups or periods.

References

Aigner, DJ. (1971). *Basic Econometrics*. N.J.: Prentice Hall, Inc., Englewood Cliffs.

Bridge, J.T. (1971). *Applied Econometrics*. Amsterdam: North-Holland Publishing Co.

Cramer, J.S. (1969). *Empirical Econometrics*. Amsterdam: North-Holland Publishing Co.

Christ, C.R. (1966). *Econometric Models and Methods*. New York: John Wiley & Sons, Inc.

Desai, M. (1976). *Applied Econometrics*. New York: McGraw-Hill Book Co.

Goldberger, A.S. (1966). *Econometric Theory*. New York: John Willey & Sons, Inc.

Gujarathi, D. (1978). *Basic Econometrics*. New York: McGraw-Hill Book Co.

Halder, N. (2024). *Unlocking Economic Insights: Mastering Analysis of Variance* (ANOVA). Analyst corner.

Mehata, B.C. & Kapoor, K. (2005). *Fundamentals of Economics*. Himalaya Publishing House, II,1- 496.