ANalysis Of VAriance

DR. AJAY KUMAR GAUTAM

TEACHING ASSOCIATE

DEPARTMENT OF AGRICULTURAL ECONOMICS AND STATISTICS

Outline

- Introduction
- Concept and Meaning
- Definitions
- Classifications
- Two- Way ANOVA Technique
- Assumptions
- Utility
- Its Advantages
- Limitations of ANOVA
- Conclusions

ANALYSIS OF VARIANCE (ANOVA)

- Analysis of variance (Abbreviated as ANOVA)
- Variance is defined as the expectation of the squared deviation of a random variable from its mean i.e. **S^2 or σ^2**
- Analysis of Variance (ANOVA) is a method for testing the hypothesis that there is no difference between two or more population means.
- The ANOVA technique enables us to perform the simultaneous test and as such is considered to be an important tool of analysis in the hands of a researcher.
- The significance of the difference of means of the two samples can be judged through either z-test or t-test.

- The technique of the analysis of variance was extremely useful in all types of researches.
- The variance of analysis studies the significance of the difference in means by analyzing variance.
- The variance would differ only when the means are significantly different.
- H0: Variability within group= Variability b/t group,
- Ha: Variability within groups $\neq$Variability b/t groups,

Concept and meaning

- The ANOVA technique was first developed by **Ronald A. Fisher**, a **British Scientist** in **1923**.
- For this technique **Fisher** is called the **Father of Modern Statistics.**
- It the most powerful techniques available in the field of statistical teaching.
- It is widely used in the experiments of behavioral and social sciences to test the significance of differences of means in different groups of a varied populations.
- Through this technique, it is possible to determine the significance of difference of different means in a single test rather than many.
- It minimizes the Type I error unlike in case of t- test.

Definitions of ANOVA

- "The analysis of variance is essentially a procedure for testing the difference between different groups of data for homogeneity".

  Yule and Kendal

- "Analysis of variance is the separation of the variance ascribe to one group of cause from the variance ascribe to other group."

  R.A. Fisher

Basic Principle of ANOVA

- The basic principle of ANOVA is to test for differences among the means of the populations by examining the amount of variation within each of these samples, relative to the amount of variation between samples.

Assumptions of Analysis of Variance

- Populations are normally distributed
- Population have same or equal variances
- Independent random samples are drawn

Classification

One Way ANOVA

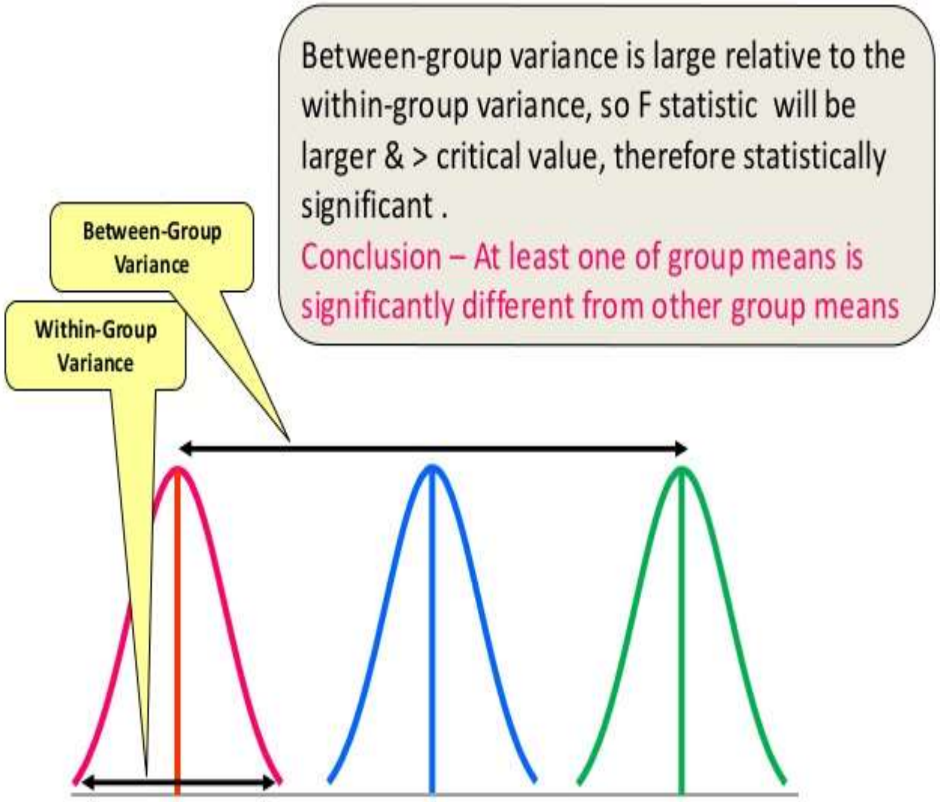
Two Way ANOVA

Two-Way Classification

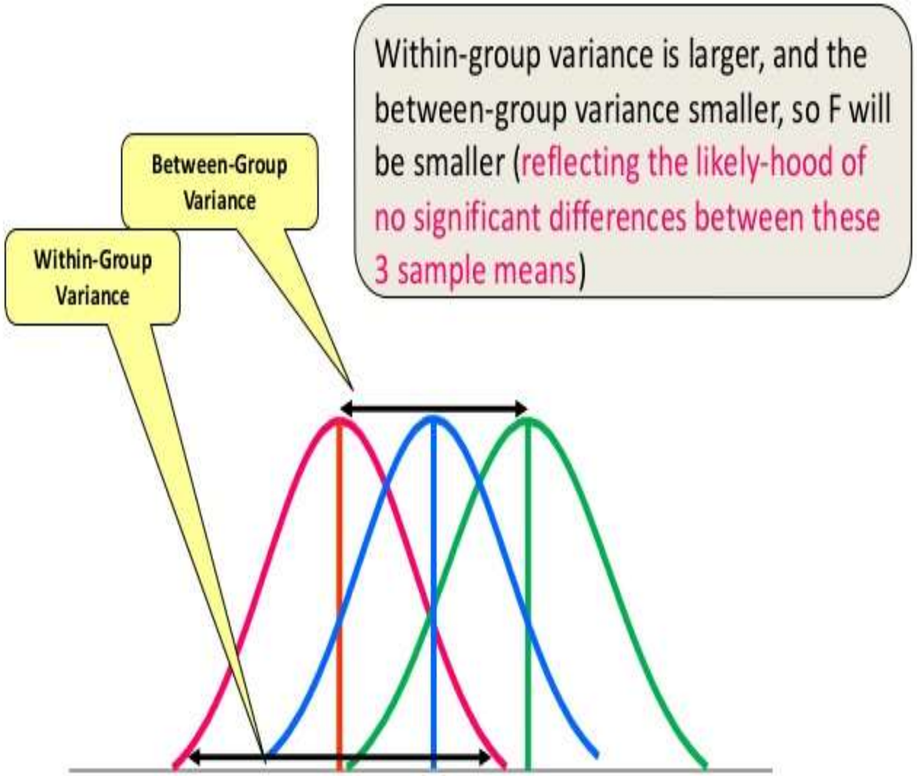
- In a one- way classification we take into account the effect of only one variable.
- If there is a two- way classification the effect of two variables or two factors can be studied.
- The procedure of analysis in a two- way technique is total both the columns and rows.
- The effect of one factor is studied through the column wise figures and totals and of the other through the row wise figures and totals.
- The variances are compared with the residual variance or error.

ANOVA Technique

- Obtain the mean of each sample
- Work out the mean of the sample means
- Calculate sum of squares for variance between the samples **(or SS between)**
- Obtain variance or mean squares **(MS) between samples**



- Calculate sum of squares for variance within the samples **(or SS within).**
- Obtain the variance or mean square **(MS) within samples**
- Find sum of squares of deviations for total variance
- Finally, find F-ratio



ANOVA Procedure

This is the ten step procedure for analysis of variance:

1. Description of data
2. Assumption: Along with the assumptions, we represent the model for each design we discuss.
3. Hypothesis
4. Test statistics
5. Distribution of test statistics
6. Description rule
7. Calculation of test statistics: the results of the arithmetic calculations will be summarized in a table called the ANOVA table. The entries in the table make it easy to evaluate the results of the analysis.
8. Statistical decision
9. Conclusion
10. Determination of p value

Utility

- ANOVA is used to test hypotheses about differences between two means
- The t-test can only be used to test difference between two means.
- When there are more than two means, it is possible to compare each mean with each other mean using t-tests.
- However, concluding multiple t-tests can be lead to severe inflation of the Type I error rate.
- ANOVA can be used for significance without increasing the Type I error rate.

Advantages of ANOVA

- It is improved technique over t-test & z-test.
- Suitable for multi- dimensional variables.
- Analysis various factors at a time.
- Can be used in three and more than three groups.
- Economical and good method of Parametric testing.
- It involve more than independent variables in studying the main impact & interaction effect.
- The experimental design (simple random design & level treatment design) are based on one way ANOVA technique.

Limitations of ANOVA

- It is difficult to analyze ANOVA under strict assumptions regarding the nature of data.
- It is not so helpful in comparison with t-test that there is no special interpretation of the significance level.
- It has a fixed and difficult set for designing experiments for the researcher.
- Requirement of post-ANOVA t-test for further testing.
- Sometimes, time consuming & also time requires knowledge & skills for solving numerical problems.
- It provides no additional information as compared to t-test.

ANOVA TABLE

TABLE 5.13: TWO-WAY CLASSIFIED DATA

Treatments (Rations)	Varieties of Cows 1	2	...	j	...	h	Row Totals $= \left(\sum_j y_{ij}\right)$	Row Means $= \left(\sum_j y_{ij}\right)/h$
1	y_{11}	y_{12}	...	y_{1j}	...	y_{1h}	$T_{1\cdot}$	$\bar{y}_{1\cdot}$
2	y_{21}	y_{22}	...	y_{2j}	...	y_{2h}	$T_{2\cdot}$	$\bar{y}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	y_{i1}	y_{i2}	...	y_{ij}	...	y_{ih}	$T_{i\cdot}$	$\bar{y}_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	y_{k1}	y_{k2}	...	y_{kj}	...	y_{kh}	$T_{k\cdot}$	$\bar{y}_{k\cdot}$
Column Totals	$T_{\cdot 1}$	$T_{\cdot 2}$	...	$T_{\cdot j}$	...	$T_{\cdot h}$	$G = \sum\sum y_{ij}$	
Column Means $= \left(\sum_i y_{ij}\right)/k$	$\bar{y}_{\cdot 1}$	$\bar{y}_{\cdot 2}$	...	$\bar{y}_{\cdot j}$	...	$\bar{y}_{\cdot h}$		

ANOVA for Fixed Effect Model

If we assume that, for both the factors the levels used are the only ones of interest, then the fixed effect or parametric model, is used.

Factor A: Treatments (Ration)

Factor B: Variety (Bread & Stock) of Cow

In the above illustration, the fixed effect model is

$y_{ij} = \mu_{ij} + \varepsilon_{ij} \quad \Rightarrow \varepsilon(.y_{ij}) = \mu_{ij}$; $(i=1,2,\ldots,k,\ j=1,2,\ldots,h)$

Where y_{ij} are independent N $(\mu_{ij}, \sigma e^2)$ and ε_{ij} are i.i.d.

1. The general mean effect μ given by $\mu = \sum_i \sum_j \mu ij/N$

The effect α_i, (i=1,2,,…,k) due to the i[th] ration given by :$\alpha_i=\mu_i-\mu$,

Where, $\mu_i = \frac{1}{n}\sum_{j=1}^{n} \mu ij$ μij: (i=1,2…,k)

2. The effect α_i, (i=1,2,,…,k) due to the i[th] ration given by :$\alpha_i=\mu_i-\mu$,

Where, $\mu_i = \frac{1}{n}\sum_{j=1}^{n} \mu_{ij}$: (i=1,2…,k)

3. The effect β_j, (j=1,2,…,h) due to the j[th] variety (breed of Cow) given by:

$\beta_j = \mu.j-\mu$, Where $\mu.j = \frac{1}{k}\sum_{i=1}^{k} \mu_{ij}$, (j=1,2,…h)

4. The interaction effect γ_{ij} when the i^{th} level of first factor (ration) & j^{th} level of second factor (breed of Cow) occur simultaneously & is given by:

$\gamma_{ij} = \mu_{ij} - \mu.i - \mu.j + \mu$

Where $\sum_{j} \gamma_{ij} = 0 \forall_{i=} = 1,2,\ldots,k$ & $\sum_{i} \gamma_{ij} = 0 \forall j = 1,2,\ldots,h$

Thus, we have

$\mu_{ij} = \mu + (\mu_i - \mu) + (\mu_{.j} - \mu) + (\mu_{ij} - \mu_{.i} - \mu_{.j} + \mu)$

and consequently the model becomes

$y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ij}$

where ε_{ij} is the error effect due to chance

Thus, we have

$\mu_{ij} = \mu + (\mu_i - \mu) + (\mu_{.j} - \mu) + (\mu_{ij} - \mu_{.i} - \mu_{.j} + \mu)$

and consequently the model becomes

$y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ij}$

where ε_{ij} is the error effect due to chance

Statistical Analysis of the Fixed Effect Model

- $\overline{y}_{i.}$= Mean yield of the i^th treatment (ration) =

$$\frac{1}{h}\sum_{j=1}^{n} yij = \frac{Ti}{h};(i=1,2,\ldots,k)$$

$\overline{y}_{ij}$ = Mean yield of the j^th variety=

$$\frac{1}{k}\sum_{i=1}^{k} yij = \frac{T.j}{k};(j=1,2,\ldots,h)$$

- $\overline{y}_{..}$= The overall mean=

$$\frac{1}{hk}\sum_{i}^{k}\sum_{j} yij = \frac{G}{N} = \frac{1}{k}\sum_{i} yi. = \frac{1}{h}\sum_{j} y.j)$$

Null and Alternative Hypothesis

Null Hypothesis-

The treatments as well as varieties are homogeneous,

$H_0t: \mu_1. = \mu_2. = \ldots = \mu_k. = \mu$; $H_0v: \mu._1 = \mu._2 = \ldots \mu._n = \mu$

or and, their equivalents;

$H_0t: \alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$; $H_0v: \beta_1 = \beta_2 = \ldots = \beta_h = 0$

Alternative Hypothesis:

H_{it}: At least two of the $\mu_i.$'s are different; H_1v: At least two of the $\mu._j$'s are different, or their equivalent:

H_{it}: At least one of the α_i's is not zero; H_1v:At least one of β_j's is not zero.

ANOVA Table for Two-Way Classified Data (one Observation per cell)

Source of Variation	Sum of Squares	Degree of Freedom	Mean Sum of Square	Ratio of F
Treatments (Ration)	$S_t^2 = h \sum_{j} (\bar{y}_{i.} - \bar{y}_{..})^2$	$k-1$	$s_t^2 = \frac{S_t^2}{(k-1)}$	$F_t = \frac{s_t^2}{s_E^2} \sim F[k-1, (h-1)(k-1)]$
Varieties (Breeds of Cow)	$S_V^2 = k \sum_{i} (\bar{y}_{.j} - \bar{y}_{..})^2$	$h-1$	$s_V^2 = \frac{S_V^2}{(h-1)}$	$F_V = \frac{s_V^2}{s_E^2} \sim F[h-1, (h-1)(k-1)]$
Residual	$S_E^2 = \sum_{i}\sum_{j} (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$	$(h-1) \times (k-1)$	$s_E^2 = \frac{S_E^2}{(h-1)(k-1)}$	
Total	$\sum_{i}\sum_{j} (y_{ij} - \bar{y}_{..})^2$	$hk-1$		

Example

To study the performance of three detergents and three different water temperature, the following whiteness readings were obtained with specially designed equipment.

Water Temp	**A**	**B**	**C**
Cold Water	47	45	50
Warm Water	39	42	52
Hot Water	44	36	48

Perform a two way ANOVA, using 5% level of Significance..

1) Calculation of Grand Total & Correction factor.

Data is coded by subtracting any guessed mid value (i.e. 40) for easy calculation

Water Temp	A	B	C	Total
Cold Water	47- 40= +7	45- 40= +5	50- 40= +10	**+22**
Warm Water	39- 40= -1	42- 40= +2	52- 40= +12	**+13**
Hot Water	44- 40= +4	36- 40= -4	48- 40= +8	**+8**
Total	**+10**	**+3**	**+30**	**43**

Correction factor $=T^2/N= (43)2/9= 1849/9=$ **205.44**

2) Calculation of SSC

$\mathbf{SSC= A^2/n_A+B^2/n_B+C^2.n_C- T^2/N}$

$SSC= (10)^2/3+ (3)^2/3+ (30)^2/3-205.44$

$SSC= 100/3+ 9/3+ 900/3-205.44= 33.33+3+300-205.44=$ **130.89**

Source of Variation	Sum of Squares	D.F.	Mean Square	Ratio of F
Between the columns	SSC=130.89	ν=(c-1) =3-1=2	MSC= SSC/(c-1) =130.89/2 =65.45	MSC/MSE =6545/12.28 =5.32
Between the Rows	SSR=33.55	ν=(r-1) =3-1=2	MSR=SSR/(r-1) =33.55/2 =16.78	MSR/MSC =16.78/12.28= 1.37
Residual or Error	SSE= 49.12	ν=(c-1) (r-1) =(3-1) (3-1)	MSE= SSE/(c-1) (r-1) =49.12/4 =12.28	
	SST= 213.5	ν=n-1		

3) **Calculation of SSR**

$SSR=C^2/n_C+ W^2/n_W+ H^2/N_H- T^2/N$

$SSR=(22)^2/3+ (13)^2/3+ (8)^2/3 -205.44$

SSR=484/3+ 169/3 +64 - 205.44

SSR=161.33+ 56.33+ 21.33- 205.44 =33.55

4) **Calculation of SST**

$SST=(7)^2+(-1)^2+(4)^2+(5)^2+(2)^2+(-4)^2+(10)^2+(12)^2+(8)^2-205.44$

SST=49+1+16+25+4+16+100+144+64-205.44

SST=419-205.44=213.56

5) **Calculation of SSE**

SSE= SST-(SSC+SSR)= 213-(130.89+33.55)=48.12

Tabulated F Value: $\nu_1=4$, $\nu_2=2$ F0.05=6.94


Thank you!