ANOVA
One way & Two way classified data

Dr. Mukta Datta Mazumder

Associate Professor
Department of Statistics

ANOVA

The total variation present in a set of observable quantities may, **under certain circumstances**, be partitioned into a number of components associated with the nature of classification of the data

The systematic procedure of achieving this is called analysis of variance (ANOVA)

ANOVA

- ANOVA was developed by statistician and evolutionary biologist Ronald Fisher.

The purpose of ANOVA is to test for significant difference between means.

If we comparing two means ANOVA will produce the same results as t- test for independent (dependent) samples

ANOVA

The name is derived from the fact that in order to test for statistical significance between means , **we are actually comparing (analyzing) variances**.

Basic ANOVA Concepts:

A response variable related to **one or more explanatory variables** , usually categorical .

One way & Two way Classified data

- One way or two way refers to the no. of independent variables

✓ One way- one independent variable(2 level)

Ex. Brand of cereal

✓

✓ Two way- two independent variables(can have multiple levels)

Ex. Brand of cereal and calories

One way- ANOVA

- **Which science departments gives out lowest average grade?**-

  Explanatory variable- Department

  Response variable- student’s GPA for individual course

  -Which kind of promotional campaign leads to greatest store income at Christmas time

  Explanatory variable-Promotion type

  Response variable- daily store income

TWO way- ANOVA

- How do the **type of career and martial**
- **status** of a person relate to the total cost of
- annual claims she/he likely to make on
- her/his **health insurance**

- - **Explanatory variables-** Career and
- martial status

- - **Response variable-** health insurance payouts

Examples

- Students from different colleges take the same examination. – one can see if one college outperform other.

- A group of psychiatric patients are trying three different therapies- counseling, medication and biofeedback.- one can check if one therapy is better than the others.

ANOVA

- **Summary**
- Analysis of variance is a statistical method used to test two or more means
- Provide statistical test whether population means of several group are equal
- Generalizes the t-test to more than one group
- Inferences about means are drawn by analyzing the variance

Assumptions

- The experimental error are normally distributed
- Equal variances between treatments(Homoscedasticity)
- Independence of Samples

Assumptions

- -Independent Observations
- -Normality
- -Homogeneity :variance within all subpopulation must be equal
- -Independence of errors –errors are independently distributed

Assumptions

- -
- Absence of outliers- Outlying score have been removed from data

- If the assumptions hold then under null hypothesis F follows F distribution with DF between & DF within SS

Logic of ANOVA

- ANOVA focuses on variability, it involve calculation of several measures of variability

-Partitioning total variation into two components

-components due to difference between means & component due to within SS (true random error)

ANOVA Test

- To find out survey or experiment results are significant(reject null hypothesis or accept alternative hypothesis)

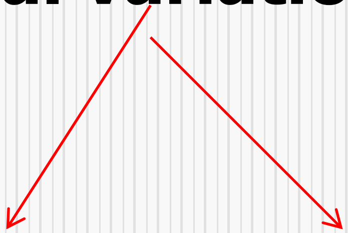
  Testing groups to check if there is a difference between them

Hypotheses

- Null hypothesis
- H0: $\mu_1 = \mu_2 = \mu_3 = \mu$
- Alternative hypothesis
- H1: at least one population mean is different from one another

Partitioning the total variation

- **Total variation**



- Between Gr variation Within Gr variation
- - variances are compared in F ratio to determine mean differences (MS between) are significantly bigger than chance (MS within)

- F= (MS bet gr)/(MS within gr)
- MS bet gr= SS bet gr/DF bet gr
- MS within gr= SS within gr/DF within gr
- Total SS= bet SS + within SS
- Total DF= bet DF+ within DF
- F ratio is always positive as F ratio computed from two variance

Numerical Example

- Suppose the National Transportation Safety Board wants to examine the safety of cars Type A, cars Type B and cars Type C . It collects a sample of three for each of the treatments (cars types).
- Using the hypothetical data provided below, test whether mean pressure applied to the driver's head during crash test is equal for each types of car.

Table

Cars Type A	Cars Type B	Cars Type C
643	469	484
655	427	456
702	525	402

STEPS

- 1. **State Null & Alternative hypotheses**

- In ANOVA null hypothesis – population means are equal
- H0: $\mu_1 = \mu_2 = \mu_3$ (Mean head pressure is statistically equal across the 3 types of cars)
- Since in null hypothesis assume all means are equal ,we could reject the null hypothesis if one mean is different thus

- -Alternative hypothesis –
- H1: at least one mean pressure is not statistically equal
- To test – we calculate appropriate test statistics

- under H0
- F= (MS bet gr)/(MS within gr) follows F distribution
- Total SS – Total variation in data.
- - It is the sum of between and within variation
- SST= $\Sigma\ \Sigma(Yij- \overline{Y})^2$ $\overline{Y}$ = 529.22
- = 96303.55

- Between SS (or Treatment SS) – Variation in the data between different samples (or treatments)
- SSTr = 86049.55
- Within variation (or error SS)- Variation in the data from each individual treatment)
- Error SS (SSE)= 10254

Mean squares

- -
- Next step in ANOVA - to compute Mean squares :
- Total mean square MST= SST/ N-1 , N= Total no of observations
- MST= 96303.55/(9-1) = 2037.94

- Mean square Treatment (MSTr) = SSTr/(k -1) ,
- k= No of treatments (in Ex no of columns)
- MSTr= 86049.55/(3 -1) = 43024.78
- Mean square error (MSE)= SSE/ (N-k)
- MSE= 10254/ (9 -3) = 1709
- NOTE :SST= SSTr + SSE but
- MST ≠ MSTr + MSE

TEST Statistic

- Next step – Calculate TEST Statistic
- F0= MSTr/MSE =25.17
- Obtain the critical value: To find critical value from F distribution it is required to know DF of numerator(DF1) & DF denominator (DF2) along with significance level

- - F (critical)has DF1=(k-1) &
- DF2= (N-k) ,
- ɑ = 5% or 1%
- In our example , DF1=3-1=2,
- DF2= (9-3) =6
- ɑ = 5%

F (critical)

- Hence we need to find F (critical) with
- 2 and 6 DF at 5% level of significance

- Using F table , F (critical) with 2 and 6 DF at 5% level of significance = 5.14

Decision Rule

- _ We reject H0 at α level of significance -
- F (observed) > F(critical)

- In our example 25.17 > 5.14
- We may reject the null hypothesis α level of significance

Interpretation

- We are 95% confident that the mean head pressure is statistically not equal for cars Type A, Cars Type B and cars Type C.
- However only one mean must be different to reject the null, we do not know which mean(s) is/are different.

- In ANOVA test provide us at least one mean is different ,additional test must be conducted to determine which mean(s) is/are different
- **Most common test – Least significant difference(LSD) test**

ANOVA Table

SV	SS	DF	MS	VR (F)	F (Cr)
Between Gr	86049.55	2	43024.78	25.17	5.14
Within Gr	10254	6	1709		
Total	96303.55	8			