

On Analysis of Two-way ANOVA using Data Transformation Techniques



Statistics

KEYWORDS : Two Way, ANOVA , Log Transformation, Square-Root Transformation , Arc Sine Transformation

Maidapwad S.L

Research Scholar , Dept. of Statistics , Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (Maharashtra)

Dr. Sananse S. L

Professor , Dept. of Statistics , Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (Maharashtra)

ABSTRACT

Generally most of the researchers use ANOVA technique considering normality in data. However, it is observed that many times data are not normal and violates assumptions of normality. In such cases data are analyzed using ANOVA and the significance of treatment may differ & recommendations of the experiments may change. This paper has thrown a light on this problem and presented here the consequences of the problem. In such cases data transformation may be one of the remedy for such problem. Here in this paper an attempt has been made to compare appropriate transformed data with normal ANOVA techniques and how significance of the treatment affects the recommendations of the experiment have been critically examined and consequences have been reported. The results are supported by numerical data. Some recommendations based on the present study are discussed. Further, the results obtained by these methods are compared and recommended.

1.0 Introduction :

In experimental situations we may have large number of treatments. Researcher interest will be to test whether all the treatments have the same effect or not. In other words, researcher intention will be to test the null hypothesis, $H_0 : \mu_1 = \mu_2 = \mu_3 = \dots = \mu_t$, against the alternative hypothesis, $\mu_1 \neq \mu_2 \neq \mu_3 \neq \dots \neq \mu_t$. For this purpose we may use Student's t-test for all possible combinations. But it is a tedious procedure and theoretically not valid. The appropriate method for such tests is the Analysis of Variance (ANOVA). The analysis of variance is the systematic algebraic procedure of decomposing the overall variation in the response observed in an experiment into different components. Each component is attributed to an identifiable cause or source of variation. The structure of these component parts is determined by the design of experiments. Basically two types of ANOVA : One-way ANOVA (Completely Randomized Design) & Two-way ANOVA (Randomized Block Design), these test have belongs to Parametric test. Maidawad s.l. & et.al [9] & Gupta S.C [7].

Parametric Test : Two-Way Analysis of Variance (ANOVA) ANOVA (Two-Way Classification)

Source of Variation	D.F.	S.S.	M.S.S.	F ratio
Treatment	t - 1	TSS	TMS	TMS / EMS
Replication	r - 1	RSS	RMS	RMS / EMS
Error	(t - 1) (r - 1)	ESS	EMS	
Total	rt - 1	Total S.S.		

If by F - test we find significant difference between treatment, then we can use critical difference (CD) for comparing pairs of treatments. The CD is given by,

$$CD = t \cdot SE(d)$$

Where, t = table value of t for a specified level of significance and error degrees of freedom.

$$SE.(d) = \sqrt{\frac{2EMS}{r}}$$

If the absolute difference $|T_i - T_j|$ for treatments i and j is greater than the critical difference, the treatments are considered to be significantly different otherwise non-significant.

The necessary assumptions in the Analysis of Variance are:

- 1) The samples are independently drawn.

- 2) The population are normally distributed, with a common variance.
- 3) The effects of various components are additive. R. Rangaswamy [13].

2.0 Data Transformation Techniques :

2.1 Logarithmic Transformation

The logarithmic transformation is most appropriate for data where the standard deviation is proportional to the mean or where the effects are multiplicative. To transform a data set into the logarithmic scale, simply take the logarithm of each and every component of the data set.

If the data set involves small values (e.g., < 10), $\log(X+1)$ should be used instead of $\log X$, where X is the original data, Kwanachai A. Gomez et.al [8].

2.2 Square -Root Transformation

Square-root transformation is appropriate for data consisting of small whole numbers. The square-root transformation is also appropriate for percentage data where the range is between 0 and 30 % or between 70 and 100 %. If most of the values in the data set are small (e.g. < 10), especially with zeros present, $(X + 0.5)^{1/2}$ should be used instead of $X^{1/2}$, where X is the original data, Kwanachai A. Gomez et.al [8].

2.3 Arc Sine Transformation

An arc sine or angular transformation is appropriate for data on proportions, data obtained from a count, and data expressed as decimal fractions or percentages. Note that percentage data that are derived from count data. The mechanics of data transformation are greatly facilitated by using a table of the arc sine transformation. The value of 0 % should be substituted by $(1/4n)$ and the value of 100 % by $(100 - 1/4n)$, where n is the number of units upon which the percentage data was based (i.e., the denominator for used in computing the percentage).

Step 1 : For percentage data lying within the range of 30 to 70 %, no transformation is needed.

Step 2 : For percentage data lying within the range of either 0 to 30 % or 70 to 100 %, but not both, the square-root transformation should be used.

Step 3 : For percentage data that do not follow the ranges specified in either Step 1 or Step 2, the arc sine transformation should be used. Kwanachai A. Gomez et.al [8].

These transformations are generally followed by the researcher. In this paper an attempt is made to suggest these appropriate data transformations considering the mean and variance of the data. Further an attempt has been made to verify the results of earlier researchers for which type of transformations use to particular type of data.

3.0 Case Study :

In an experiment on Cowpea in randomized blocks design in plant heights were observed. The results (observations in round-up form) are presented in Table 1. Test whether the differences between the treatments are significant. (R. Rangaswamy [13]).

Table 1 : Height (cm) of Cowpea

Treatment	Block			
	I	II	III	IV
T ₁	35	32	67	23
T ₂	36	34	72	37
T ₃	43	38	69	42
T ₄	88	85	96	98
T ₅	92	96	98	97

3.1 Parametric Test :

Firstly apply Parametric Two-way ANOVA on Table 1 data and it's result are in following manner,

ANOVA (Two-Way Classification)

Source of Variation	D.F.	S.S.	M.S.S.	F ratio	P value
Treatment	4	12068.8	3017.20	35.90	0.000
Block	3	1830.6	610.20	7.26	0.005
Error	12	1008.4	84.03		
Total	19	14907.8			

On the basis p value is less than 0.05 i.o.s. value , so all the treatment effects are statistical significant. If by F- test we find significant difference between treatment, then we can use critical difference (CD) for comparing pairs of treatments.. The CD is given by,

$$CD = t. SE (d)$$

$$CD = 14.12$$

Treatment Mean : T1 = 39.25 , T2 = 44.75 , T3 = 48 , T4 = 91.75 , T5 = 95.75

Comment : Treatment pairs (T1 - T2), (T1 - T3), (T2 - T3), (T4 - T5) are Non-significant and other remaining treatment pairs are Significant .

By using Anderson-Darling normality test (<0.005), p values are less than 0.05, so this data follows non-normal distribution. By using following Data transformation techniques, we can shift non-normal data tends to normal. After data transformation, again check normality and if it follows normal then apply Parametric ANOVA on transformed data . If after transformation it follows non-normal then should be use Nonparametric methods such as Friedman test.

3.2 Data Transformation Techniques

Generally most of the researchers use ANOVA technique considering normality and other assumptions in data, however it is observed that many times data are not normals and violates assumptions of normality, in such cases the significant of treatment may differ. In such cases data transformation may be remedy for such problem. Mainly three types of data transformation

such as, 1) Log transformation 2) Square root transformation 3) Arc sine transformation. Hence transformation describe have been applied to the same data results are presented below.

3.2.1 Log Transformation

Table 2 : Criteria for using Log transformation

Treatment	Mean	Range	Variance	Standard Deviation
T ₁	39.25	44	368.25	19.19
T ₂	44.75	38	331.58	18.21
T ₃	48.00	31	200.67	14.17
T ₄	91.75	13	38.92	6.24
T ₅	95.75	6	6.92	2.63

Using result of Table 2 and as per guidelines of Kwanachai A. Gomez et.al [7] , it shows non-linear relationship between the range and mean i.e. the range increases but not proportionally with the mean. So for this data , Logarithmic transformation is not appropriate. But for research purpose , here apply log transformation and use Two- way ANOVA

Below data are log transformation of the original data presented in Table 1 .

Treatment	Block			
	I	II	III	IV
T ₁	1.5	1.5	1.8	1.4
T ₂	1.6	1.5	1.9	1.6
T ₃	1.6	1.6	1.8	1.6
T ₄	1.9	1.9	2.0	2.0
T ₅	2.0	2.0	2.0	2.0

By using Anderson-Darling normality test (<0.005), p values are less than 0.05, so this log transformed data do not follow normal distribution. After transformation it follows non-normal then use Nonparametric methods such as Friedman test. But we use here ANOVA instead of Non-Parametric for study purpose.

ANOVA (Two-Way Classification)

Source of Variation	D.F.	S.S.	M.S.S.	F ratio	P value
Treatment	4	0.648	0.162000	22.09	0.000
Block	3	0.132	0.044000	6.00	0.010
Error	12	0.088	0.007333		
Total	19	0.868			

On the basis p value is less than 0.05 i.o.s. value , so all the treatment effects are statistical significant. If by F- test we find significant difference between treatment, then we can use critical difference (CD) for comparing pairs of treatments. The CD is given by,

$$CD = t. SE (d)$$

$$CD = 0.131$$

Treatment Mean : T1 = 1.55 , T2 = 1.65, T3 = 1.65 , T4 = 1.95 , T5 = 2

Comment : Treatment pairs (T1 - T2), (T1 - T3), (T2 - T3), (T4 - T5) are Non-significant and other remaining treatment pairs are Significant .

3.2.2 Square-Root Transformation

Table 1. data does follows small whole number and percentage

data not lying within the range of either 0 to 30 % or 70 to 100 % , but not both so the square –root transformation should not be used.

Below data are square-root transformation of the original data presented in Table 1 .

Treatment	Block			
	I	II	III	IV
T ₁	5.9	5.7	8.2	4.8
T ₂	6.0	5.8	8.5	6.1
T ₃	6.6	6.2	8.3	6.5
T ₄	9.4	9.2	9.8	9.9
T ₅	9.6	9.8	9.9	9.8

By using Anderson-Darling normality test (<0.005), p values are less than 0.05, so this square-root transformed datado not followsnormal distribution. After transformation it follows non-normal then use Nonparametric methods such as Friedman test. But we use here ANOVA instead of Non-Parametric for study purpose.

ANOVA (Two-Way Classification)

Source of Variation	D.F.	S.S.	M.S.S.	F ratio	P value
Treatment	4	48.095	12.0237	26.25	0.000
Block	3	8.728	2.9093	6.35	0.008
Error	12	5.497	0.4581		
Total	19	62.320			

On the basis p value is less than0.05 l.o.s. value , so all the treatment effects are statistical significant. If by F- test we find significant difference between treatment, then we can use critical difference (CD) for comparing pairs of treatments. The CD is given by,

$$CD = t. SE (d)$$

$$CD = 1.04$$

Treatment Mean : T₁ = 6.15 , T₂ = 6.6, T₃ =6.9, T₄ =9.57, T₅ =9.8

Comment : Treatment pairs (T₁ – T₂), (T₁ – T₃), (T₂ – T₃), (T₄ – T₅) are Non-significant and other remaining treatment pairs are Significant .

3.2.3 Angular Transformation

Table 1. data does not follow rules or ranges of log and square-root transformation , so for this type of data, Arc sine transformation should be proper. Below data are square-root transformation of the original data presented in Table 1 .

Treatment	Block			
	I	II	III	IV
T ₁	36.27	34.40	54.94	60.00
T ₂	36.87	35.67	58.10	28.66
T ₃	40.98	38.06	56.17	37.47
T ₄	69.73	67.21	78.46	40.40
T ₅	93.57	78.46	81.87	15.34

By using Anderson-Darling normality test (<0.005), p values are less than 0.05, so this Angular transformed datado not follows-normal distribution. After transformation it follows non-normal

then use Nonparametric methods such as Friedman test. But we use here ANOVA instead of Non-Parametric for study purpose.

ANOVA (Two-Way Classification)

Source of Variation	D.F.	S.S.	M.S.S.	F ratio	P value
Treatment	4	2538.57	634.644	2.18	0.134
Block	3	2256.06	752.018	2.58	0.102
Error	12	3501.44	291.787		
Total	19	8296.07			

On the basis p value is greater than0.05 l.o.s. value , so all the treatment effects are non-statistical significant.

4.0 Specific Comments on use of Data transformations :

- Generally researcher useTwo-way ANOVA without considering normality assumption. In this paper, we apply Two-way ANOVA on sample percentage data and it result shows all treatment effects are statistical significant. But this data do not follow normal distribution. So we apply following data transformation methods on non-normal data for shifting to normal and then apply regular Two-way ANOVA on transformed data.

1) **Log transformation& Square root transformation:**After use of log transformation & square root transformation for sample data, then transformed data again shows non-normal so we should use here Non-Parametric method such as Friedman Test. But without using this non-parametric methods we use Parametric Two-way ANOVA on transformed data for verify Parametric method result instead use of Non-parametric method. On the basis of Two-way ANOVA result, it's p value is less than 0.05 l.o.s. value , so all the treatment effects are statistical significant. So we can use critical difference (CD) for comparing significant pairs of treatments . Before apply transformation and After apply log & square root transformation, same treatment pairs are significant and non-significant such as treatment pairs (T₁ – T₂), (T₁ – T₃), (T₂ – T₃), (T₄ – T₅) are Non-significant and other remaining treatment pairs are Significant .

2) **Angular or Arc sine transformation :** After angular transformation for sample data, then transformed data shows normal. We use Parametric Two-way ANOVA on transformed data. On the basis of Two-way ANOVA result, it's p value is greater than 0.05 l.o.s. value , so all the treatment effects are non- statistical significant.

5.0 Suggestions based on research study :

- Most of time, when researcher directly use Two-way ANOVA on data without test of normality then it's results are misleading because in this paper we use Two-way ANOVA for sample data and it result shows all treatment effects are statistical significant i.e. all treatment effects are different. But after check testof normality this data do not follow normal. After usingangular transformation of sample data , it shifting to normal. Then apply Two-way ANOVA on transformed data and it shows treatment effects are non-statistical significant i.e. all treatment effects are statistically same. In such type of situation, researcher mislead his results, this paper focus on result comparison of without using data transformation and after transformation.
- In this paper , sample data is in percentage data form and it most of values lies in 30 to 100 percentage. For such types data, we should properly use angular transformation instead of log and square-root transformation methods.

REFERENCE

- 1) Bartlett, M.S. (1947) : The use of transformation. *Biometrics*, 3, pp 39-52. | 2) Box, G.E.P. and Cox, D.R. (1964) : An analysis of transformation. *J. Roy. Statist. Soc.B*, 26 , pp 211 – 252. | 3) Conover, W.J., and Iman, R.L. (1976) : In some Alternative Procedure using ranks for the analysis of experimental designs. *Commun. Statist. Theor. Math.* A5(14), pp 1349-1368. | 4) Dolby, J.L. (1963) : A quick method for choosing a transformation. *Technometrics*, 5, pp 317 – 326. | 5) Draper, N.R. and Hunter, W.G. (1969) : Transformations : Some examples revisited. *Technometrics*, 11, pp 23 – 40. | 6) Gore, A.P and Deshpande, J.V : (Book) *Statistical Analysis of Non-normal data* pp 87 – 100. | 7) S.C. Gupta and V.K. : (Book) *Fundamentals of Applied Statistics*, S. Chand Publication. | 8) Kwanachai A. Gomez, Arturo A. Gomez (2010) : (Book) *Statistical Procedures for Agricultural Research* , Second Edition, pp 209 – 308. | 9) Maidapwad, S.L.and Sananse, S.L. Jan-Dec (2010) On Parametric and Non-Parametric Tests for Randomized Block Design, *Global Journal of Applied Mathematics and Mathematical Sciences*, Vol. 3 No. 1 , pp 1 -13. | 10) Maidapwad, S.L.and Sananse, S.L. Dec. (2009) On Estimation of Growth | Rates using Linear Models, *Inter.J. Agril.Statist.Sci* Vol.5 ,No.2,pp 463-469. | 11) Montgomery, D.C. (1997) : (Book) *Design and Analysis of Experiment*, John wiley& Sons. | 12) Pearson, E.S. (1929) : The analysis of variance in case of non-normal variation. *Biometrika*. 23, 259-286. | 13) R. Rangaswamy (2010) : (Book) *A Textbook of Agricultural Statistics*, New Age Publication, Second Edition. pp. 136-140. | 14) RajendraPrasad : *Non-Parametric Methods in Analysis of Data*.I.A.S.R.I, New Delhi. |