Available online at www.elixirpublishers.com (Elixir International Journal)

Discrete Mathematics

Elixir Dis. Math. 92 (2016) 39301-39304

ISSN: 2229-712X

The robustness of F-test in two-way interactive balanced design

Ehiwario, J.C[1], Osemeke, R.F[1] and Okafor Nnaemeka P[2]

[1]Department of Mathematics College of Education, Agbor, Delta State

[2]Institute of Ecumenical Education, in Affiliation with Enugu state University, Enugu Onitsha study centre, Anambra State.

ARTICLE INFO

Article history:
Received: 22 July 2012;
Received in revised form:
24 March 2016;
Accepted: 29 March 2016;

Keywords
Robust,
F-test,
Random.

ABSTRACT

The extent of departure of a set of data from the fundamental assumptions required by a test is often very important in the determination of the appropriate statistical test to be applied. Some test statistics are more robust to the departure from certain underlying assumptions than others. Hence, the study was aimed at examining the robustness or non-robustness of the F-test statistic in the two-way Interactive Balanced Design. The data used for the study were tested against the three basic assumptions of analysis of variance which include normality, independence and homogeneity of variance assumptions. The results show that even when the normality and homogeneity of variance assumptions were violated, the F-test still yields good results. Hence, it was concluded that the F-test is robust to the normality and homogeneity of variance assumptions. Based on the results of the study, it was recommended that large sample size should always be used in carrying out any experiment involving the applications of analysis of variance. It was also recommended that the rank transformation is a step in the right direction whenever our data fail to conform with the assumptions of ANOVA prior to the conventional ANOVA test since it yields result in test which are more robust to the non-normal and resistance to outliers and non-homogeneity of variance.

Introduction

Analysis of variance is widely used in industries and firms to help identify the source of potential problems in the production process and identify whether variations in measured output value is due to variability between various manufacturing process or within them by varying the factors in a predetermined pattern and analyzing the output, one can used statistical techniques to make an accurate assessment as to the causes of variation in a manufacturing process (Altman and Bland, 2006). An F-test is the test statistic, used for testing for differences in means of more than two populations for statistical significance

An F-test statistic is said to be robust if it has the capacity of providing adequate protection against deviation from the assumptions underlying the statistical procedure. According to Glass et al (1992), the ability of a test statistic to provide accurate estimation of probability of type I and type II errors, even when the underlying criteria are violated, is regarded as robustness.

In testing for statistical significance of difference between more than two group-means via F-test, certain criteria which guarantee the use of Analysis of Variance (ANOVA) are to be met. These criteria include normality, independence and homogeneity of variances. There are various statistical techniques that can be used in testing for these assumptions. According to Eze (2002), applying the ANOVA technique without testing for the conformity of the underlying assumptions is like treating an ailment without going through medical diagnosis. This underscores the need for testing the assumptions underlying any statistical procedure.

Assumptions of ANOVA and the Effects of their Violations

The three conceptual classes of models in which the experimental design is based include fixed, random and mixed effects models. Each of these models is based on the conditions of normality, homogeneity of variances and independence of the variables.

The F-test statistic is remarkably robust to the deviation from normality and homogeneity of variances. This is so since it can yield meaningful results even when its assumptions are violated by the true model from which the data is generated [(Lindman, 1974, Glass et al, 1992) Andy (1992) and Ehiwario (2008)] while Ferguson and Takae (2005) hold a contrary view of it.

The Skewness of the distribution usually does not have any sizeable effect on the F-statistic. The Central limit theorem allows us to assume that the criterion of normality is approximated even for the skewed distributions, if the sample sizes are large enough. However, moderate deviation from normality does not seriously affects the F-test [Winer(1971), Andy (1972) and Ezepue (2006)]. Similarly, with respect to homogeneity of variances, Andy (1992) said if the sample sizes are equal, this condition is readily met and there is no need to test whether the variances are homogeneous or not. He however, maintained that if the sample sizes are not equal, the homogeneity of variances can be tested by using appropriate test. Andy also noted that if larger samples also have larger variances, a conservative F-test is obtained for the null hypothesis, that is the probability of committing type I error is less than the stated α- level. On the other hand if the larger samples are associated with smaller variances, the probability of type I error increases much more than our stated α- level.

Hubert (1979) said the F-test itself does not test for the assumption of homogeneity of variance. According to him, instances where sample variances seem to differ considerably among themselves and independent test for equality of variances may have to be made. If the results of such test indicate that there are rather extreme departures from homogeneity of

Tele:
E-mail address: moluaogom@hotmail.com

variances, then Hubert opined that analysis of variance should not be used. He however, suggested that moderate departure from homogeneity of variance can be tolerated. Such departures can often be reduced considerably by the transformation of variables. Similarly, Andy (1992) suggested that if we suspect that the variances are unequal and there are different sample sizes, the raw data can be transformed by any method. The essence of the transformations is to reduce the differences between group variances, so that the assumption of homogeneity of variances is met. However, moderate deviation of such assumption does not seriously affects the F-test, hence, it is said to be robust to the assumption.

The criterion for independence is more difficult to evaluate, although in most cases, there is no ground to doubt the validity of the independence. This is particularly so, if the subjects in one group are not the same as those in another and have been randomly selected (Andy, 1992). Eze (2002) and Ehiwario (2008) however said lack of independence could result in correlated error terms. Thus, the correlation between two terms should be zero.

Ranks on Analysis of Variance

Conover and Iman (1981) suggested that when a set of data do not satisfy the assumptions of analysis of variance, we can represent each of the original data value by its rank and then run a conventional ANOVA calculation on the transformed data. Helsel and Hirsch (2002) remarked that rank transformed data results in test which are more robust to non-normality and resistance to outliers and heterogeneity of variances, than in ANOVA without the transformation. Seaman (1994), however, noticed that the rank transformation of Conover and Imam (1981) was not appropriate for testing interactions among effects in a factorial design as it can cause an increase in type I error.

Method of Data Collection

The data for the study were collected from the daily production record of Camel Paint and Chemical Industries in Delta State as used in Ehiwario (2008). Hence, the data were purely secondary data. The data include the daily production of the various Machines per Crew for twenty days. Hence, in each of the cells there are twenty (20) observations. (See the appendix).

Method of Analysis

The data collected were analysed with both descriptive and inferential statistics. The raw data were transformed using the logarithmic transformation method. This was aimed at bringing the variances of data closer to equality and to reduce the difficulties involved in the computations of the variances of the various groups. The transformed data were then tested to ascertain the level of conformity to the basic assumptions of **ANOVA.**

Test for Normality Assumption

Here, our aim is to test the data for normality assumption using the goodness –of-fit. Table 1 below represents the distribution of the probability points and the expected frequencies

Hypothesis

Ho_1:

Data are normally distributed

H_{11}:

Data are not normally distributed

$$\chi^2 = \sum_{i=1}^{8} (O_i - E_j)/E_j,\ i = 1, 2, 8$$

$= 46.168$ (from table 1)

df = r = k – m – 1 where k = no of categories

$\chi^2_{(5,0.05)} = 11.070$

Comparing the two chi-squared values, we observed that the chi-squared calculated exceeded the critical value (tabulated). Hence, we reject Ho_1 and conclude that the data are not normally distributed.

Test for Homogeneity of Variances, using the Bartlett's Test on the Transformed Data

The test statistic is given by:

χ^2_{α} = 2.3026q/C, with a -1 degrees of freedom where

$q = (N – a) \log Sp^2 - \Sigma (n – 1) \log S_i^2$

$C = 1 + 1/3(a -1) [\Sigma (n –1)^{-1} – (N –a)^{-1}]$

$S_P^2 = (N –a)^{-1} \Sigma (n-1) S^2_1$

Let S^2_i denote the unbiased estimate of the population variance treatment i and SP^2 the unbiased estimate of the pooled variance.

The test hypotheses are:

$Ho_2: \delta_1^2 = \delta_2^2 = \delta_3^2$

$H_{11}: \delta_1^2 \neq \delta_2^2 \neq \delta_3^2$

We have the summary of the result as:

Sp^2 = 96.0443, q = 2.701 and C =1.00005 a = 3 and n_i =60

Substituting these values into the Bartlett's test statistic, we have:

$$\chi^2 = 2.3026 \times \frac{2.70100}{1.0005}$$

$\chi^2 = 6.23$

Testing at α = 0.05, we have $\chi^2_{3,0.05} = 5.99$.

Comparing the two chi-squared values, we noticed that the chi-squared calculated exceeds the critical value. Hence, we reject the null hypothesis and conclude that the homogeneity of variances assumption is moderately violated.

Independence

Test for Independence of the Data

Here, the test for the independence of the data via the goodness of-fit test is done on the transformed data. The result is shown in table 2 below.

The test hypothesis is

Ho_3:

The data are independent

The test statistic is given as:

$\chi^2 = \Sigma\ (O-E)^2/E$

$\chi^2 = 0.0000189$

Testing at 0.05 level of significance, we have our critical value as:

$\chi^2_{(4,0.05)} = 9.49$

Comparing the two chi-squared values, we observe that the chi-squared calculated is less than the critical value (tabulated). Hence, we accept H_{o3} and conclude that the data are independent.

The results of the above tests of the assumptions of analysis of variance via the appropriate test-statistic revealed that the data are not normally distributed, the variances are not homogeneous and the independent assumption is met.

Table 1: Distribution of the Probability Points and Expected Frequencies of the Transformed Data.

Boundary points	Prob. of points (p)	Expected freq. E	Observed freq. O	$(O-E)^2/E$
2.23035-2.24685	0	0	1	0
2.24685-2.26335	0.0010	0.180	1	3.736
2.26335-2.27985	0.0258	4.644	11	8.699
2.27985-2.29635	0.1968	35.624	15	11.776
2.29635-2.31285	0.4355	78.390	100	5.957
2.31285-2.32935	0.2827	50.886	36	4.355
2.32935-2.34585	0.0551	9.918	13	0.958
2.34585-2.36265	0.0031	0.558	3	10.687
			180	46.168

Table 2: Distribution of the Observed and Expected Frequencies of the Data

Observed freq. O_i	**Expected freq. (E_j)**	**$(O_i-E_j)^2/E_j$**
46.3076	46.2892	0.00000731
46.2089	46.2124	0.000000265
46.1392	46.1541	0.00000481
46.1937	46.2009	0.00000112
46.1262	46.1242	0.0000000867
46.0713	46.0661	0.000000587
46.1140	46.1251	0.00000267
46.0500	46.0485	0.0000000489
46.0002	45.9905	0.00000205
total		0.0000189

Table 3: ANOVA Table

Source of variation	df	Sum of squares (SS)	Mean squares (MS)	F-cal	F-tab
Machine, α	2	469.811	234.906	3.368	3.00
Crew, β	2	307.811	153.906	2.207	0.113
Machine * Crew, λ	4	11.689	2.922	0.042	2.37
Error	171	11926.00	69.737		
Total	179	12714.311			

Appendix

Transformed Data, using Logarithmic Transformation Method

1	2.3222	2.2788	2.2900	2.2788	2.3010	2.3010
	2.3324	2.3222	2.3222	2.3010	2.2944	2.3424
	22900	2.2900	2.3010	2.3118	2.3222	2.22742
	2.3118	2.3010	2.3222	2.3118	2.3010	2.322
	2.3010	2.3222	2.3222	2.3010	2.3010	2.2672
	2.3010	2.3010	2.3434	2.2788	2.3222	2.3424
	2.3617	2.3118	2.2833	2.3010	2.2672	2.3324
	2.3424	2.3617	2.3010	2.3324	2.3020	2.3010
	2.3324	2.2900	2.3010	2.3424	2.3010	2.3010
	2.3222	2.3118	2.3424	2.3222	2.3222	2.3222
	(46.3076)				(46.1392)	
2	2.3010	2.3617	2.2900	2.3010	2.3010	2.3010
	2.3222	2.3010	2.3010	2.3010	2.3222	2.3010
	2.3118	2.3324	2.3010	2.3010	2.3010	2.3222
	2.3010	2.3010	2.3118	2.3010	2.2900	2.2900
	2.3010	2.2672	2.3222	2.3010	2.3010	2.3010
	2.3222	2.3222	2.3010	2.3222	2.3118	2.3222
	2.3345	2.2553	2.3010	2.3010	2.3010	2.3010
	2.3010	2.3010	2.3222	2.3118	2.2900	2.3118
	2.3010	2.3222	2.3010	2.3118	2.2789	2.3010
	23010	2.3118	2.3222	2.3010	2.3010	2.3222
	(46.1937)		(46.1262)		(46.0713)	(46.0713)
3	2.3010	2.3222	2.3010	2.2798	2.2900	2.3222
	2.3010	2.3010	2.3010	2.2789	2.3010	2.3010
	2.3010	2.3222	2.3118	2.3118	2.3010	2.3010
	2.3222	2.3222	2.3010	2.3010	2.3222	2.3010
	2.2900	2.3010	2.3222	2.3010	2.3010	2.2304
	2.3010	2.3010	2.3010	2.3010	2.3224	2.3010
	2.3010	2.2788	2.2900	2.3010	2.3010	2.3010
	2.3010	2.3222	2.3010	2.3222	2.2900	2.3010
	2.3222	2.3010	2.3010	2.3010	2.2900	2.3010
	2.3010	2.3010	2.3222	2.3010	2.3010	2.3010
	(46.1140)		(46.0500)		(46.0002)	
$T_{\cdot j}$	138.6153		138.3851		138.2107	
$X_{\cdot j}$	2.3133		2.3064		2.3035	
S_2	96.33180		960154		957765	

At this point, we now proceed to carry out conventional ANOVA test on the subjects effects.

Analysis of Variance on the Subject Effects

Hypotheses

i) Ho_4: α: = 0 for all i = 1, 2 and 3

H_{14}: α: = 0 for at least one i

H_{04} means that the mean production levels of the machines are equal.

ii) H_{05}: $\beta_i = 0$, for all j = 1, 2 and 3

H_{15}: $\beta = 0$, for at least one j

H_{05}: implies that the mean production capacities of the crews are equal.

iii) Ho_6: $\lambda_{ij} = 0$, for at least one i, and $_j$ combination.

H_{06} means that there is no interaction effect between the machines and the crews.

Decision Rules

Reject the null hypotheses at 0.05 level of significance, if the calculated chi-squared value exceeds the critical (tabulated) value.

The summary of the result of the conventional ANOVA test on the data are shown in Table 3 below:

The computational details are suppressed as the results can easily be generated from statistical software.

From Table 3, we observe that $SS\alpha$, $SS\beta$ and $SS\lambda$ corresponding to machines, crews and machine-crew interactions respectively have the following respective values: 469.811, 307.811 and 11.689. Similarly, their respective degrees of freedoms are 2,2 and 4. While SSE = 11926 with corresponding degrees of freedom, 171. On the other hand, the mean squares for the row, column, interaction effects and error are 234.906, 153.906, 2.922 and 69.737 respectively. We noticed also, from the table that the F-calculated for the row, column and interactions are 3.368, 2.207 and 0.042 respectively. Conversely, the corresponding F-tabulated are 3.00, 3.00 and 2.37 respectively.

Based on these results and our decision rule, we reject Ho_4 and conclude that the mean production levels of the machines are not equal. On the other hand, we accept H_{05} and H_{06} and conclude that the average production capacity of the crews are equal and also that there is apparently no interaction between the machines and the crews that uses them respectively.

Discussion of Results

When we compare the last two columns of Table 3, we observed that the difference, (D) between the F-calculated and the F-tabulated are very small. That is for the machine, 3.368 – 3.00 = 0.368, for the crews, we have 2.207 – 0.113 = 2.094 and for the interactions, we have 2.37 – 0.042 = 2.328.

The difference between the F-calculated and F-tabulated is said to be significant if $D \geq 5$ (Ukpom, 2001). Since the difference (D) between the F-calculated and F-tabulated in the test of the three hypotheses, are all less than 5, we assert that the F-test statistic is robust on deviation from the normality and homogeneity of variances. This result is in line with that of Lindman (1974), Andy, 1992, Ehiwario, (2008) and Ezepue (2006) which stressed that provided that the sample sizes are large, the deviation from normality assumption has no effect on the F-test. This according to them is sequel to the Central limit theorem. Similarly, when the homogeneity of variance assumption is moderately violated, the F-test is not seriously affected and hence it is robust to this criterion Andy (1992) and Ehiwario (2008)

Conclusion and Recommendation

Based on the result of the study, we conclude that the F-test statistic is robust to the normality, homogeneity ofvariances. This is so because, moderate departures from both assumptions do not adversely affects the F-test.

Following the outcome of our findings, we recommend that in collecting data for any experiment, efforts should be made to always have large set of data to enable the Central limit theorem to come into play should the normality assumption be violated. We equally recommend that the rank transformation method should be adopted before carrying out ANOVA test whenever the ANOVA assumptions are violated. This is so because, according to Helsel and Hirsch (2002), rank transformation yields results in test which are more robust to the non-normality and resistant to outliers and heterogeneity of variance.

References

Attman, D.G and J.M Bland (2996). Statistics notes: Comparing Several Groups, using ANOVA. http://en.wikipedia.org/wiki/analysis of variance

Andy, I.J. (1992). Fundamental Statistics for Education and the Behavioural Science. Ibadan: kraft Books Ltd.

Conover, W.J and R.L. Iman (1981). Rank Transformation as bridge between Parametric and non-Parametric Statistics: Concept and Method Lexington: DC Health and Company.

EZe, F.C. (2002). Introduction to Analysis of Variance vol.1, Enugu: Lano Publishers.

Ferguson, G.A and Y. Takane (2005). Statistical Analysis in Psychology and Education, 6th Edition Montreal Quebec: McGraw-Hill Ryerson Limited.

Helsel, D.R. and R.M. Hirsch (2002). Statistical Methods in Water Resources: Techniques of Water resources Investigations. Trends Ecol. Eval. 9.

Hubert M.B (1979). Social Statistics 2nd Edition. Washington: McGraw-Hill, Inc.

Lindman, H.R. (1974). Analysis of Variance in Complex Experimental Design. New York: W.H. freeman and Co.

Seamen, J.W, S.C. Walls, S.E Wide and Jaeger (1994). Caveat Emptor: Rank Transformation Methods and Interactions. Trends Ecol. Evol.9.

Ehiwario, J.C (2008). Common F-test Denominator for two Way Interactive Balances Design. An M.Sc Thesis, Nnamdi Azikiwe University, Awka.

Ezepue, P.O. and L.A. Udo (2006). Conversation in Applied Statistical Modeling Part 1: Stochastic Models for Operations and profitability Assessment in Barbers Shops. Hawal International Conference on Statistics, Mathematics and Related Fields.