On Two Tests for Main Effects in A Balanced Two-Way Interactive Model

F.C. Eze[1*] Adimonye F.0[2] Urama K. U[3]

1,2 Department of Statistics Nnamdi-Azikiwe University P.M.B 5025, Awka, Anambra State, Nigeria

3 Michael Okpara University of Agriculture Umudike-Umuahia Abia State Nigeria

*E-mail: ezefcc@yahoo.com

Abstract

Two methods of removing interaction in a two-way balanced design were considered. Removing interaction from the model/ data will result to a reduced model and data which consequently violates the assumptions of analysis of variance (ANOVA). To resolve this problem, a linear combination method approach was used which does not violates the assumptions of ANOVA and completely makes the presence of interaction to be zero.

Keywords: Additive effects, Homogeneity of variance, Normal distribution

1. Introduction

The interpretation of data based on analysis of variance (ANOVA) is valid only when the assumption of ANOVA is satisfied. These assumptions are additive effects, independence of errors, homogeneity of variance and normal distribution. The departures from these assumptions make the interpretation based on these statistical techniques invalid. Therefore, it is necessary to detect the deviations and apply the appropriate remedial measures.

The effects of two factors, say treatment and replication, are said to be additive if the effects of one factor remains constant over all the levels of the factors. A hypothetical example from set of data with 2 treatments and two replications with additive effects is shown in Table 1.

Table 1.Randomized complete block design with additive effects

Treatment	Replication		Replication effect
	1	11	1 - 11
A	85	70	15
B	60	45	15
Treatment effect (A-B)	25	25	

Here in Table 1, the treatment effect is equal to 25 for both replications and replication effect is 15 for both treatments.

If the factors are additive, automatically the interaction between them are zero. Overton (2001) argued that if the factors are additive, interactions between the factors are zero and their graph plot must be parallel as shown in figure 1 using the above hypothetical example.

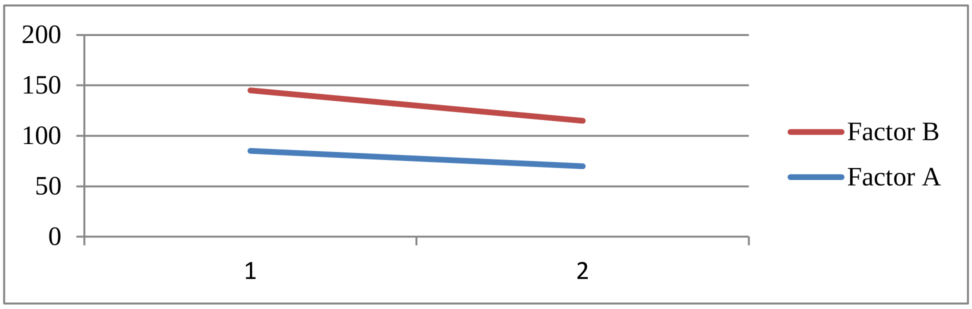

Figure 1 Graph showing zero interaction

The presence of interaction may obscure the result for the test of significance for the main effects according to Moore, et al (2004). One may be tempted to vehemently maintain that we should not even test for main effects once we know that interactions are present. Presence of interaction between the two factors means both effects are not independent.

Montgomery (2001) cited a similar example to Overton (2001). He demonstrated that interaction occurs between two factors when the difference in response between the levels of one factor is not the same at all levels of the other factors. According to Montgomery (2001), a significant interaction can mask the significance of the main effects. This means that, when interaction is present, the main effects of the factors involved in the interaction may not have much meaning.

Stockburger (1998) explained that change in the simple main effect of one variable over levels of the other is most easily seen in the graph of the interaction. If the lines describing the simple main effects are not parallel, then a possibility of an interaction exists he emphasized.

Ottenbacher (1991) in his paper entitled "Interpretation of interaction in factorial analysis of variance" argued that the validity of statistical conclusions in medical research depends on proper analysis and interpretation of collected data. According to Ottenbacher (1991), he argued that the potential area of invalidity is the inappropriate post hoc analysis of statistically significant interactions in the analysis of variance of factorial designs. He examined the statistical explanations included in 83 studies published in three leading medical journals where the findings indicated significant interaction effects. Only 24 per cent of the reported statistically significant interactions had an accompanying correct interpretation. The most common form of misinterpretation involved a comparison of individual cell means within a row or column of one factor used in the design. This interpretation according to him did not conform to the factorial ANOVA model with interaction.

Hayes and Matthes (2009) argued that researchers often hypothesize moderated effects, in which the effect of an independent variable on an outcome variable depends on the value of a moderator variable. Such an effect reveals itself statistically as an interaction between the independent and moderator variables in a model of the outcome variable. When an interaction is found, it is important to probe the interaction, for theories and hypotheses often predict not just interaction but a specific pattern of effects of the focal independent variable as a function of the moderator.

2. Methodology

2.1 Removal of interaction from the model/data

Given the model

$$X_{ijk} = \mu + \alpha_i + \beta_j + \lambda_{ij} + e_{ijk} \quad i = 1, 2, ..., 3; \; j = 1, 2, ..., q \text{ and } k = 1, 2, ..., k \qquad (1)$$

Where

X_{ijk} is the kth observation in ijth cell,

μ is a constant,

α_i is the average effect of factor A,

β_j is the average effect of factor B,

λ_{ij} is the interaction between factor A and factor B and

e_{ijk} is the error associated with X_{ijk} .

The least square estimates for the various parameters in Equation (1) can be shown to be

$$\mu = \bar{X}_{...}$$

$$\alpha_i = \bar{X}_{i..} - \bar{X}_{...}$$

$$\beta_j = \bar{X}_{.j.} - \bar{X}_{...}$$

$$\lambda_{ij} = \bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}_{...}$$

$$e_{ijk} = X_{ijk} - \bar{X}_{ij.}$$

Statement: To remove the interaction from Equation (1) does not change the least square estimates of the main effects.

Proof:

From Equation (1)

$$X^*_{ijk} = X_{ijk} - \bar{X}_{ij.} + \bar{X}_{i..} + \bar{X}_{.j.} - \bar{X}_{...}$$

Averaging over k we have $\bar{X}^*_{ij.} = \bar{X}_{i..} + \bar{X}_{.j.} - \bar{X}_{...}$

Next we take average over jk to have $\bar{X}^*_{i..} = \bar{X}_{i..}$

Finally, we take average ijk to have $\bar{X}^*_{...} = \bar{X}_{...}$

It then follows that $\bar{X}^*_{i..} - \bar{X}^*_{...} = \bar{X}_{i..} - \bar{X} = \hat{\alpha}_i$

and $\bar{X}^*_{.j.} - \bar{X}^*_{...} = \bar{X}_{.j.} - \bar{X} = \hat{\beta}_j$

When the interaction is removed from the model, Equation (1) reduces to

$$X_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk} \quad i = 1, 2, ..., 3; \;\; j = 1, 2, ..., q \text{ and } k = 1, 2, ..., k \qquad (2)$$

and consequently, the data will be reduced resulting to distorting the homogeneity condition of ANOVA. Performing ANOVA techniques on the reduced data will lead to a biased result.

2.2 Linear Combination

Eze et al (2013) developed a method of eliminating interaction from a two-way balanced interactive model/data by a method of linear combination. Any linear combination such that the differences between the corresponding yield row-wise as well as column-wise differences is a constant and yet the total sum of the yield remains unchanged eliminates the interaction.

Let $x_{11}, x_{12}, ..., x_{pq}$ be the yields or values in two-way crossed interactive data. The data format for a two-way balanced interactive data with one observation per cell is shown in Table 2 below.

Table 2. Data format for two-way balanced interactive data with one observation per cell.

	Factor B				
Factor A	1	2	3	… q	
1	x_{11}	x_{12}	x_{13}	x_{1q}	
2	x_{21}	x_{22}	x_{23}	x_{2q}	
3	x_{31}	x_{32}	x_{33}	x_{3q}	
.	.	.	.	.	
.	.	.	.	.	
.	.	.	.	.	
P	x_{p1}	x_{p2}	x_{p3}	x_{pq}	$x_{i.}$

From Table 2

$$x_{12} - x_{11} = k \Rightarrow x_{12} = k + x_{11}$$
$$x_{13} - x_{12} = k \Rightarrow x_{13} = k + x_{12} = 2k + x_{11}$$
$$x_{21} - x_{13} = k \Rightarrow x_{21} = k + x_{13} = 3k + x_{11}$$

In general

$$x_{11} + \{(pq-1)k + (pq-1)x_{11}\} = x_{i.} \quad (3)$$

where

x_i is the total observation per number of observation per cell.

If there are more than one observations per cell, Equation (3) extends to

$$x_{111} + \{(pq-1)k + (pq-1)x_{111}\} = x_{1..} \quad (4)$$

$$x_{121} + \{(pq-1)k + (pq-1)x_{121}\} = x_{2..} \quad (5)$$

and so on.

The value of k is then added or subtracted from the first term of the observation.

3. Illustrative Example:

An engineer is designing a battery for use in a device that will be subjected to some extreme variations in temperature. Since the temperature can be controlled in the product development laboratory for the purposes of a test, the engineer decides to test all three plate materials at three temperature levels – 15^0F, 70^0F and 125^0F – as these temperature levels are consistent with the product end-use environment. Four batteries are tested at each combination of plate material and temperature and all 36 tests are run in random order. The data are shown in Table 3 below.

Table 3
Life (in hours) Data for Battery Design

	Temperature (^{0}F)						
Material type	15		70		125		
1	130	155	34	40	20	70	
	74	180	80	75	82	58	
2	150	188	136	122	25	70	
	159	126	106	115	58	45	
3	138	110	174	120	96	104	
	168	160	150	139	82	60	
							$x_{1..} = 903$ $x_{2..} = 979$ $x_{3..} = 959$ $x_{4..} = 958$ $x_{...} = 3799$

Source: D.C. Montgomery (2001). Design and Analysis of Experiments.

The above data was analyzed using SPSS version 15 and the result presented in ANOVA Table 4

Table 4. ANOVA Table

Dependent Variable:Observations

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	59154.000[a]	8	7394.250	11.103	.000
Intercept	398792.250	1	398792.250	598.829	.000
Material	10633.167	2	5316.583	7.983	.002
Temperature	39083.167	2	19541.583	29.344	.000
Material * Temperature	9437.667	4	2359.417	3.543	.019
Error	17980.750	27	665.954		
Total	475927.000	36			
Corrected Total	77134.750	35			

The interaction between material and temperature is significant.

Using the method in section 2.1, the interaction was removed from the data to have the reduced data presented in Table 5.

ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.4, No.10, 2014

Table 5. Reduced Life (in hours) Data for Battery Design

	Temperature (^{0}F)						
Material type	15		70		125		
1	170.72 61.72	142.72 167.72	61.97 107.97	67.97 102.97	35.69 97.69	85.69 73.69	
2	158.12 167.12	196.12 134.12	145.37 115.37	131.37 124.37	42.47 75.47	87.47 62.47	
3	61.8 91.8	33.8 83.8	58.8 34.8	4.8 23.8	-2.36 -16.36	5.64 -38.36	
							$x_{1..} = 679.58$ $x_{2..} = 755.58$ $x_{3..} = 735.58$ $x_{4..} = 734.58$ $x_{...} = 2905.32$

The analysis of the above reduced data is presented in the ANOVA Table 6.

Table 6. ANOVA Table

Dependent Variable:Observation

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	90826.910[a]	8	11353.364	16.814	.000
Intercept	234469.008	1	234469.008	347.252	.000
Material	53228.746	2	26614.373	39.416	.000
Temperature	34319.714	2	17159.857	25.414	.000
Material * Temperature	3278.450	4	819.613	1.214	.328
Error	18230.750	27	675.213		
Total	343526.669	36			
Corrected Total	109057.660	35			

From the ANOVA Table above, the interaction between the material and temperature is now non-significant but not zero.

The plot of the data entries are shown in figure 2.

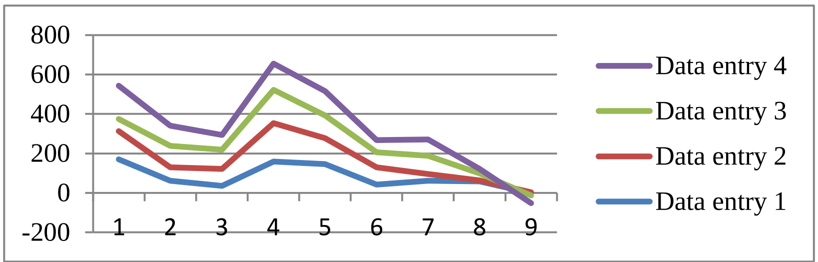

Figure 2: Graph of reduced data entry without interaction

From Figure 2, none of the data entries are parallel to each other, indicating lack of additive effects.

Using the method of linear combination described in section 2.2, the original data in Table 3 are transformed and presented in Table 7.

Table 7. Life (in hours) Data for Battery Design

	Temperature (^{0}F)						
Material type	15		70		125		
1	130	155	122.58	143.44	115.16	131.88	
	74	180	82	161.61	90.28	143.22	
2	107.74	120.32	100.32	108.76	92.90	97.20	
	98.42	124.83	106.56	106.44	114.70	88.05	
3	85.48	85.64	78.06	74.08	70.64	62.52	
	122.84	69.66	130.98	51.27	139.12	32.88	
							$x_{1..} = 902.88$
							$x_{2..} = 978.84$
							$x_{3..} = 958.90$
							$x_{4..} = 957.96$
							$x_{...} = 3798.58$

The analysis of variance of the above data is shown in ANOVA Table 8.

Table 8. ANOVA Table

Table 8 ANOVA Table

Dependent Variable: Observations

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	12815.894[a]	8	1601.987	1.870	.107
Intercept	400840.934	1	400840.934	467.952	.000
Material	11534.304	2	5767.152	6.733	.004
Temperature	1281.589	2	640.795	.748	.483
Material * Temperature	.000	4	.000	.000	1.000
Error	23127.799	27	856.585		
Total	436784.627	36			
Corrected Total	35943.692	35			

From the ANOVA Table 8, the interaction between material and temperature is non- significant and zero.

The plot of the data entries are shown in figure 2.

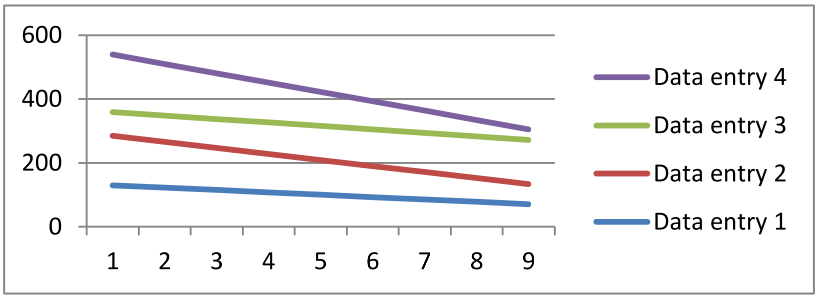

Figure 3: Graph of data without interaction

4. Discussions and Conclusion

We have seen that using the method in section 2.1 the presence of interaction is non-significant but not zero. The basic assumptions of ANOVA especially the homogeneity of variance and additive effects are violated.

For instance, in Table 5, 170.72 – 61.97 ≠ 61.97 – 35.69 and 170.72 – 158.12 ≠ 61.97 – 145.37 and so on. The total data entries for each cell are reduced as well as sum total for the entire cell totals. Also the graph of the reduced data in figure 2 is not parallel.

In contrast, in Table 7, 130 – 122.58 = 122.58 – 115.16 and 130 – 107.74 = 122.58 – 100.32 and so on. This obeys the assumption of additive effect. The total data entries for each cell as well as sum total for the entire cell totals remained unchanged. In figure 3, data entry 1 is parallel to data entry 3 and data entry 2 and data entry 4 shows sign of parallelism.

In conclusion the method in section 2.2 is highly recommended.

References

Eze, F.C., Adimonye, F.O, Nnanwa, C.P., and Ezeani, M.I. (2013). An appropriate F-Test For Two-way Balanced Interactive Model. International Journal Of Computational Engineering Research, Vol. 3 Issue 2, pp 121 – 127.

Hayes, A.F., and Mathhes, J.C (2009). Computational Procedures for Probing Interactions in OLS and Logistic Regression: SPSS and SAS Implementations. Behavior Research Meyhos 41(3): pp 924-936.

Moore, D.S., McCable, G., Duckworth, W., and Sclove, S. (2004). Practice of Business Statistics, Part iv pp15-17. Palgrave Macmillan.

Montgomery, D.C. (2001). Design and Analysis of Experiments. John Wiley & Sons New York p.176.

Ottenbacher, K.J (1991). Interpretation of Interaction in Factorial Analysis of Variance Design. Statistics in Medicine. Vol. 10, no. 10 pp 1565-1571.

Overton, R.C. (2001). Moderated Multiple Regression for Interaction Involving Categorical Variables – A Statistical ontrol for Heterogeneous Variable across Groups. Psychol. Methods, 6(3). pp218-233.

Stockburger, D.W (1998). Multivariate Statistics: Concepts, Models, and Applications. Missouri State University (www version 1.0)