Neutrosophic 2²-Factorial Designs and Analysis

Pranesh Kumar

Department of Mathematics and Statistics, University of Northern British Columbia, Prince George, BC V2N 4Z9, Canada.

Email: Pranesh.kumar@unbc.ca Manuscript submitted April 8, 2023; revised May 8, 2023; accepted August 9, 2023; published May 23, 2024. doi: 10.17706/ijapm.2024.14.2.59-66

Abstract: In field or laboratory planned experiments, it is possible to observe vague, incomplete, or imprecise data due to known or unknown reasons. Thus, the analysis should take into consideration the imprecision in data vales. In recent past, researchers have proposed various approaches such as fuzzy, intuitionistic fuzzy and neutrosophic logic and analysis, which provide better understanding, analysis and interpretations of the imprecise data. Experimental design and analysis is a systematic, rigorous approach to problem solving that applies principles and techniques at the data collection stage so as to ensure the generation of valid, defensible, and supportable conclusions. Factorial designs are widely used in experiments that involve several factors and where it is necessary to study the joint effects of the factors on a response. Several special cases of the general factorial design are important because they are widely used in research work and also because they form the basis of other designs of considerable practical value. These designs are widely used in factor screening experiments as well. The most important of these special cases is that of *k* factors, each at only two levels. These levels may be quantitative or they may be qualitative. A complete replicate of such a design is called a 2^{k} -factorial design. In this paper, we consider the first design in the 2^k -series which is one with only two factors, say A and B, each run at two levels. The levels of the factors may be arbitrarily called low and high. This design is called a 2²-factorial design. For the imprecise response data, we will define a neutrosophic 2^2 -factorial design (N2²FD), neutrosophic model and neutrosophic analysis. As an illustration, we consider an investigation into the effect of the concentration of the reactant and the amount of the catalyst on the conversion (yield) in a chemical process. The objective of the experiment is to determine if adjustments to either of these two factors would increase the yield.

Keywords: Imprecise data, neutrosophic statistics, neutrosophic factorial design, neutrosophic analysis

1. Introduction

It is commonly noted in field or laboratory planned experiments that the collected data or information is vague, incomplete, or imprecise due to measurement errors or data collection errors. In the classical data analysis, methods assume that data values are precise and do not take into account the imprecise nature of the data. However, the analysis should take into consideration any likely imprecision in data set. To understand and analyze imprecision, Smarandache [1–3] has proposed neutrosophic logic and neutrosophic math which is an extension of fuzzy logic where a variable *x* is described by triplet values, i.e., x = (t, i, f), where *t* is the degree of truth, *f* is the degree of false and *i* is the level of indeterminacy. A neutrosophic data *x* can be expressed as x = d + i, where *d* is the determinate (sure) part of *x*, and *i* is the indeterminate (unsure) part of *x*. For example, suppose a measured value *x* is not precisely known but we

know that $x \in [8,8.6]$. This data value can be expressed as x = 8 + i, where $i \in [0,0.6]$. That means, for sure $x \ge 1$ 8 (meaning that the determinate part of x is 8), while the indeterminate part $i \in [0,0.6]$ means the possibility for number x to be greater than or equal to 8 but less than or equal to 8.6. In field or laboratory planned experiments, it is possible to observe vague, incomplete, or imprecise data due to known or unknown reasons. Thus, the analysis should take into consideration the imprecision in data vales. Kumari et al. [4] have introduced the neutrosophic completely randomized design that is a generalization of the completely randomized design. They studied the flexible way of handling imprecise elements in completely randomized design. AlAita et al. [5] have proposed a novel method for ANCOVA using neutrosophic statistics. They applied the Neutrosophic Analysis of Covariance (NANCOVA) in three different designs, neutrosophic completely randomized design, neutrosophic randomized complete block design, and neutrosophic split-plot design. Authors have implemented and explained proposed method using numerical examples. The proposed NANCOVA method is found to be flexible and effective in the presence of uncertainty when compared to the existing method. AlAita et al. [6] have proposed neutrosophic statistics analysis for split-plot and split-block designs. In their proposed method, neutrosophic hypothesis is formulated, a decision rule is suggested, and neutrosophic ANOVA table is given. A numerical example and a simulation study demonstrate the effectiveness of the proposed method.

Factorial designs are widely used in experiments that involve several factors and where besides the effects of main factors, it is necessary to study the joint effects of the factors on an observed response. Several special cases of the general factorial design are important to the experimenter as they are widely used in research work and also because they form the basis of other designs of considerable practical value. These designs are widely used in factor screening experiments as well. The most important of these special cases is that of k factors, where each factor is at two levels. These levels may be quantitative or they may be qualitative. A complete replicate of such a design is called a 2^k -factorial design.

In this paper, we consider the first design in the 2^k -series which is one with only two factors, say *A* and *B*, each run at two levels. The levels of the factors are arbitrarily called low (–) and high (+). This design is called a 2^2 -factorial design. For the imprecise response data values, we will define a neutrosophic 2^2 -factorial design (N2²FD), neutrosophic model and neutrosophic analysis. As an illustration, we consider an investigation into the effect of the concentration of the reactant and the amount of the catalyst on the conversion (yield) in a chemical process. The objective of the experiment is to determine if adjustments to either of these two factors would increase the yield.

2. Neutrosophic 2²-Factorial Design

The Neutrosophic 2²-Factorial Design (N2²FD) is a factorial design with two factors *A* and *B*, each run at two levels, and have the imprecise response which is a neutrosphic variable. As an example [7], let a researcher wish to investigate the effect of the concentration of the reactant (*A*) and the amount of the catalyst (*B*) on the conversion yield (*y*) in a chemical process. The objective of the experiment is to determine if adjustments to either of these two factors would increase the yield. Further, let factor *A* be the reactant concentration with two levels, 15% (low level) and 25% (high level). The catalyst is factor *B* with two levels, the high level denoting the use of 2 pounds of the catalyst and the low level denoting the use of 1 pound. The experiment is replicated three times (i.e., n = 3), so there are 12 runs (observations). The order in which the runs are made is random, so this is a completely randomized experiment. The data obtained from this experiment is in Table 1 and also shown in Fig. 1. Note that some data values of the response variable conversion (yield) are neutrosophic values.

Factor Treatment Replicate Total A В Combination I Π III A low, B low 28 25 [26,28] [79,81] A high, B low [35,37] 32 32 [99,101] + A low, B high [16,20] [18,20] [20,26] [54,66] + A high, B high [89,91] [30,32] 30 29 +





Fig. 1. Conversion yield in the chemical process for treatment combinations: (a) Lower value; (b) Upper value.

By convention, the effect of a factor is denoted by a capital Latin letter. Thus, *A* refers to the effect of factor *A*, *B* refers to the effect of factor *B*, and *AB* refers to the *AB* interaction effect. In the 2^2 -design, the low and high levels of *A* and *B* are denoted by "+" and "-", respectively. The letter *a* represents the treatment combination of *A* at the high level and *B* at the low level, *b* represents *A* at the low level and *B* at the high level, and *ab* represents both factors at the high level. By convention, (1) is used to denote both factors at the low level. For the given data values, it is therefore noted that

$$(1) = [79,81], a = [99,101], b = [54,66], ab = [89,91].$$

3. Factor Effects

In a two-level factorial design, we define the main (average) effect of a factor as the change in response produced by a change in the level of that factor averaged over the levels of the other factor. Thus, the main effect of *A*:

$$A = \frac{ab+a-b-(1)}{2n} \tag{1}$$

The main effect of factor *B*:

$$B = \frac{ab+b-a-(1)}{2n} \tag{2}$$

We define the interaction effect *AB* as the average difference between the effect of *A* at the high level of *B* and the effect of *A* at the low level of *B*. Thus, the AB interaction effect:

$$AB = \frac{ab+(1)-a-b}{2n} \tag{3}$$

Using the treatment combination values: (1) = [79,81], a = [99,101], b = [54,66], ab = [89,91], we estimate the factor effects as

$$A = \frac{ab + a - b - (1)}{2n} = \frac{[89,91] + [99,101] - [54,66] - [79,81]}{2(3)}$$
$$= \frac{[188,192] - [133,147]}{6} = \frac{[41,59]}{6} = [6.83,9.83]$$
$$B = \frac{ab + b - a - (1)}{2n} = \frac{[89,91] + [54,66] - [99,101] - [79,81]}{2(3)}$$
$$= \frac{[143,157] - [178,182]}{6} = \frac{[-39,-21]}{6} = [-6.6,-3.5]$$
$$AB = \frac{ab + (1) - a - b}{2n} = \frac{[89,91] + [79,81] - [99,101] - [54,66]}{2(3)}$$
$$= \frac{[168,172] - [153,167]}{6} = \frac{[1,19]}{6} = [0.17,3.17]$$

The effect of *A* (reactant concentration) is positive and [6.83,9.83]. This suggests that increasing *A* from the low level (15%) to the high level (25%) will increase the yield by [6.83,9.83]. The effect of *B* (catalyst) is [-6.6, -3.5]; this suggests that increasing the amount of catalyst added to the process from 1 pound to 2 pound will decrease the yield. The interaction effect [0.17,3.17] appears to be small relative to the two main effects. And, hence can be ignored in the design regression model.

4. Analysis of Variance (ANOVA) for the Neutrosophic 2²-Factorial Experiment

In factorial design experiments, it is useful to examine the magnitude and direction of the factor effects to determine which variables are likely to be important. The analysis of variance is used to confirm this interpretation. Effect magnitude and direction are considered along with the ANOVA, because the ANOVA alone does not convey this information.

We use the factor contrast to compute the sum of squares. Sum of squares due to factor A:

 $SS_A = \frac{[ab+a-b-(1)]^2}{4n} = \frac{([89,91]+[99,101]-[54,66]-[79,81])^2}{4(3)} = \frac{([41,59])^2}{12} = \frac{[1681,3481]}{12} = [140.08,290.08].$

Sum of squares due to factor *B*:

$$SS_B = \frac{[ab+b-a-(1)]^2}{4n} = \frac{([89,91]+[54,66]-[99,101]-[79,81])^2}{4(3)} = \frac{([-39,-21])^2}{12} = \frac{[441,1521]}{12} = [36.75,126.75]$$

Sum of squares due to interaction of factors, i.e., AB:

$$SS_{AB} = \frac{[ab+(1)-a-b]^2}{4n} = \frac{([89,91]+[79,81]-[99,101]-[54,66])^2}{4(3)} = \frac{([1,19])^2}{12} = \frac{[1,361]}{12} = [0.08,30.08]$$

Total Sum of Squares:

$$SS_{T} = \sum_{i,j,k} y_{ijk}^{2} - \frac{y_{..}^{2}}{4n} = [8979,9851] - \frac{[103041,114921]}{4(3)} = [274.25,392.25]$$
$$SS_{E} = SS_{T} - SS_{A} - SS_{B} - SS_{AB} = [274.25,392.25] - [140.08,290.08]$$
$$-[36.75,126.75] - [0.08,30.08] = [54.66,97.34]$$

The complete ANOVA for the neutrosophic 2²-factorial design is summarized in Table 2. On the basis of the *P*-values, we conclude that the main effects of factor A (reactant concentration) and factor B (amount of catalyst) are statistically significant. There is no significant interaction between factor A and factor B. This confirms initial interpretation of the data based on the magnitudes of the factor effects.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F _{1,8}	P-value
А	1	[140.08,290.08]	[140.08,290.08]	[11.51,42.47]	[0.0002,0.009]
В	1	[36.75,126.75]	[36.75,126.75]	[4.59,15.84]	[0.004,0.064]
AB	1	[0.08,30.08]	[0.08,30.08]	[0.01,3.76]	[0.088 ,0.923]
Error	8	[54.66,97.34]	[6.83,12.17]		

Table 2. Analysis of Variance Table for the Neutrosophic 2²-Factorial Design

5. Neutrosophic 2²-Factorial Design Regression Model

For the chemical process experiment considered above, the regression model without interaction term is

$$y = \beta_0 + \beta_0 x_1 + \beta_2 x_2 + \varepsilon, \tag{4}$$

where x_1 is a coded variable that represents the reactant concentration, x_2 is a coded variable that represents the amount of catalyst, and the β 's are regression coefficients. The relationship between the natural variables, the reactant concentration and the amount of catalyst, and the coded variables is:

$$x_1 = \frac{Conc - (Conc_{low} - Conc_{high})/2}{(Conc_{high} - Conc_{low})/2} = \frac{Conc - (15 + 20)/2}{(20 - 15)/2} = \frac{Conc - 20}{5}$$
(5)

Thus, if the concentration is at the high level (Conc = 25%), then $x_1 = +1$; if the concentration is at the low level (Conc = 15%), then $x_1 = -1$. Further,

$$x_2 = \frac{Catalyst - (Catalyst_{low} - Catalyst_{high})/2}{(Catalyst_{high} - Catalyst_{low})/2}$$
(6)

Thus, if the catalyst is at the high level (Catalyst = 2 pounds), then $x_2 = +1$; if the catalyst is at the low level (Catalyst = 1 pound), then $x_2 = -1$.

The fitted regression model with coded variables is

$$\hat{y} = [26.75, 28.25] + \left(\frac{[6.83, 9.83]}{2}\right) x_1 + \left(\frac{[-6.6, -3.5]}{2}\right) x_2 \tag{7}$$

$$\hat{y} = [26.75, 28.25] + [3.415, 4.915] x_1 + [-3.3, -1.75] x_2$$
 (8)

where the intercept is the average of all 12 response values, and the regression coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ are one-half the corresponding factor effect estimates. The regression coefficient is one-half the effect estimate because a regression coefficient measures the effect of a one-unit change in x on the mean of y, and the effect estimate is based on a two-unit change (from -1 to +1). This simple method of estimating the regression coefficients results in least squares parameter estimates.

The fitted regression model with natural factor levels is

$$\hat{y} = [26.75, 28.25] + \left(\frac{[6.83, 9.83]}{2}\right) \left(\frac{Conc-20}{5}\right) + \left(\frac{[-6.6, -3.5]}{2}\right) \left(\frac{Catalyst-1.5}{0.5}\right)$$
(9)

$$\hat{y} = \left[26.75, 28.25\right] + \left[3.415, 4.915\right] \left(\frac{Conc-20}{5}\right) + \left[-3.3, -1.75\right] \left(\frac{Catalyst-1.5}{0.5}\right)$$
(10)

$$\hat{y} = [12.34, 24.49] + [0.683, 0.983] Conc + [5.25, 9.9] Catalyst$$
 (11)

The regression model is used to predict values or calculate fitted value of y at the four points in the design. The residuals are the differences between the observed and fitted values of y. When the reactant concentration is at the low level ($x_1 = -1$) and the catalyst is at the low level ($x_2 = -1$), the predicted yield:

$$\hat{y} = [26.75, 28.25] + [3.415, 4.915] (-1) + [-3.3, -1.75] (-1)[23.585, 28.135]$$
(12)

There are three observations at this treatment combination, and thus, the residuals are

$$e_1 = [28, 28] - [23.585, 28.135] = [-0.135, 4.415]$$

 $e_2 = [25, 25] - [23.585, 28.135] = [-3.135, 1.415]$
 $e_3 = [26, 28] - [23.585, 28.135] = [-2.135, 4.415]$

When the reactant concentration is at the high level $(x_1 = +1)$ and the catalyst is at the low level $(x_2 = -1)$, the predicted yield is

$$\hat{y} = [26.75, 28.25] + [3.415, 4.915] (+1) = [31.915, 36.465])$$
 (13)

and the residuals are:

$$e_4 = [35, 37] - [31.915, 36.465] = [-1.465, 5.085]$$

 $e_5 = [32, 32] - [31.915, 36.465] = [-4.465, 0.085]$
 $e_6 = [32, 32] - [31.915, 36.465] = [-4.465, 0.085]$

When the reactant concentration is at the low level $(x_1 = -1)$ and the catalyst is at the high level $(x_2 = +1)$, the predicted yield is

 $\hat{y} = [26.75, 28.25] + [3.415, 4.915] (-1) + [-3.3, -1.75] (+1) = [18.535, 23.085]$ (14) and the residuals are:

$$e_7 = [16,20] - [18.535,23.085] = [-7.085,1.465]$$

 $e_8 = [18,20] - [18.535,23.085] = [-5.085,1.465]$
 $e_9 = [20,26] - [18.535,23.085] = [-3.085,7.465]$

When the reactant concentration is at the high level $(x_1 = +1)$ and the catalyst is at the high level $(x_2 = +1)$, the predicted yield is

$$\hat{y} = [26.75, 28.25] + [3.415, 4.915] (+1) + [-3.3, -1.75] (+1) = [26.865, 31.415]$$
 (15)

and the residuals are:

$$e_{10} = [30,32] - [26.865,31.415] = [-1.415,5.135]$$

 $e_{11} = [30,30] - [26.865,31.415] = [-1.415,3.135]$
 $e_{12} = [29,29] - [26.865,31.415] = [-2.415,2.135]$

The predicted values and residuals are plotted in Fig. 2.

6. Conclusion

Neutrosophic logic, neutrosophic math, and neutrosophic statistics are important tools to understand and analyze the incomplete, vague and imprecise information/data. There is a huge scope for fuzzy data derived from experimental design set up. The work can easily be extended to other areas of designing too. We have shown the application of neutrosophic analysis of the neutrosophic 2²-factorial design by considering the experimental data that was collected to investigate the effect of the concentration of the reactant and the amount of the catalyst on the conversion yield in a chemical process. The objective of the experiment is to determine if adjustments to either of these two factors would increase the yield. It was concluded that the effect of reactant concentration is positive and [6.83,9.83]. This suggests that increasing reactant concentration from the low level (15%) to the high level (25%) will increase the yield by [6.83,9.83]. The effect of catalyst is [-6.6,-3.5]; this suggests that increasing the amount of catalyst added to the process from 1 pound to 2 pounds will decrease the yield. The interaction effect [0.17,3.17] appears to be small relative to the two main effects. And, hence can be ignored. As a future research, author plans to extend the neutrosophic analysis to the 2^k -factorial designs and response surface models.



Fig. 2. Predicted and observed conversion yield: (a) Lower observations; (b) Upper observations.

Conflict of Interest

The author declares no conflict of interest.

References

- [1] Smarandache, F. (2007). A Unifying Field in Logics: Neutrosophic Logic. Neutrosophy, Neutrosophic Set, Neutrosophic Probability and Statistics (6th ed.). USA: InfoLearnQuest. Retrieved from https://digitalrepository.unm.edu/cgi/viewcontent.cgi?article=1109&context=math_fsp
- [2] Smarandache, F. (2014). *Introduction to Neutrosophic Statistics*. Sitech & Education Publishing. Retrieved from https://digitalrepository.unm.edu/math_fsp/33
- [3] Smarandache, F. (2019). *Nidus Idearum. Scilogs, VII: Superluminal Physics* (3rd ed.). Bruxelles, Belgium: Andrew Thuysbaert E-publishing. Retrieved from http://fs.unm.edu/NidusIdearum7-ed3.pdf
- [4] Kumari, S., Azarudheen, S., & Jincy, J. (2022). A study on neutrosophic completely randomised design.

Mathematical Statistician and Engineering Applications, 71(4), 3738–3747.

- [5] AlAita, A., & Aslam, M. (2023). Analysis of covariance under neutrosophic statistics. *Journal of Statistical Computation and Simulation*, 93(3), 397–415. https://doi.org/10.1080/00949655.2022.2108423
- [6] AlAita, A., Talebi, H., Aslam, M., Al-Sultan, K. (2023). Neutrosophic statistical analysis of split-plot designs. *Soft Computing*, *27*, 7801–7811. https://doi.org/10.1007/s00500-023-08025-y
- [7] Montgomery, D. C. (2012). *Design and Analysis of Experiments* (8th ed.). John Wiley & Sons Inc.

Copyright © 2024 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (<u>CC BY 4.0</u>).