ESTIMATE MISSING VALUE OF RANDOMIZED BLOCK DESIGN USING VARIABLE CONTROL CHARTS THROUGH RESPONSE SURFACE METHODOLOGY

Kokkilambigai S[1] and Pachamuthu M[2*]
•
[1]Research Scholar, Department of Statistics, Periyar University, Salem-11, India
[1]msamundiswariharini@gmail.com
[2*]Assistant Professor, Department of Statistics, Periyar University, Salem-11, India
[2*]pachamuthu@periyaruniversity.ac.in

Abstract

The complete block design of Randomised Block Design (RBD) and variable control charts play a vital role in the fields of agriculture and industry. The basic complete block designs are Completely Randomized Designs (CRD), RBD, and Latin Square Designs (LSD). An improvement in the CRD can be obtained by providing error control measures called an RBD. In statistics, missing values occur when a data value for a variable is not stored in an observation. There are several methods available to estimate the missing value of complete block designs, such as fuzzy techniques, traditional methodology, etc. In this paper, we are introducing the new methods to estimate the missing value using variable control charts such as 3-Sigma and 6-Sigma control charts through Response Surface Methodology (RSM) with the numerical example. RSM can be used to find factor systems that produce a desired maximum, minimum, or optimal response; to find factor systems that satisfy management specifications; and to model the relationship between quantitative factors and responses.

Keywords: *CRD, RBD, LSD*, 3-Sigma Control Charts, 6-Sigma Control Charts, *RSM.*

I. Introduction

Ronald A Fisher [1920] first introduced the Design of Experiments (*DOE*) in 1953. *DOE* is a systematic method used in various fields of design to tasks describe and explain data variation under hypothesized conditions. There are three basic principles of *DOE* such as randomization, replication and blocking or error control. Blocking (b) an experiment involves dividing or partitioning the observations into homogeneous groups called blocks so that the observations in each block are collected under relatively similar experimental conditions, treatments (v) are the different objects or procedures being compared in an experiment and replication (r) refers to the repetition of an experiment under identical conditions, but in the context of experimental designs, it refers to the number of distinct experimental units under the same treatment. The *DOE* may be broadly classified as two types are complete and incomplete block design. Designs in each block that receive all treatments are called complete block design otherwise its incomplete block designs. A *CRD* is an experimental design where treatments are assigned completely at random, ensuring that each experimental unit has an equal chance of receiving any one treatment. A *RBD* is a restricted randomized design in which experimental units are first organized into homogeneous blocks, and then the treatments are assigned at random to these units within these blocks. Due to some unforeseen reasons, observations may be missing in some observations. Observations of one or two units are too large or too small compared to other observations. The accuracy of such observations is often questionable. These observations greatly inflate the error variance and thereby distort the results. Therefore, the best ones are omitted, and the remaining data are then analyzed considering the missing data. The missing-plot technique is a standard formula for the estimation of a missing data observation. It is used within side the evaluation of the variance of data collected consistent with a diagnosed experimental design. If there is one missing observation, we can use to estimate the missing observation along with the sum of squares for testing the differential effect of the treatments. The main objective of *RSM* is to optimize reactions to achieve high yields and purity at low cost. Identify patterns, relationships, and optimal conditions for the response. The *RSM* is a technique for optimizing the response when two or more quantitative factors are involved. Dependent variables are referred to as responses, whereas independent variables or factors are primarily referred to as predictor variables in a response surface system. A 3*D* surface plot is a graphical representation used to visualize the relationship between the response variable and two independent variables. This plot helps to identify the optimal conditions for the response variable. Statistical Quality Control *(SQC)* is one of the methods used to determine whether data is normal or not. *SQC* is the use of statistical methods to monitor and control the quality of a manufacturing process.

There are many reachers review the concepts we are given some of the them mixture-process variable experiments, including control and noise variables within a split plot structure Tae-Yeon Cho, et.al [14] have discussed the mixture-process variables experiments often involve larger ran due to increased coefficient parameters and difficult to controlled interactions. These experiments could have analyzed used meant and slope models within a split-plot structure. Low prediction variances were desirable. Methods for mixture-process variable designs with noise variables demonstrated, and some robust designs are evaluated used fraction of design space plots. A *G*-optimal design created using a genetic algorithm. Loss of efficiency in randomized block designs with two missing values Subramani. J and Balamurali. S [13] has investigated the efficiency of *RBD* with two missing values, obtaining explicit expressions for elementary treatment contrasts, average variance, and loss of efficiency. It also tabulated the loss of efficiency for different values. Optimum estimation of missing values in *RCBD* by genetic algorithm. Azadeh. A, et.al. [1] have discussed the efficiency of *RBD* with missing values, providing explicit expressions for treatment contrasts, average variance, and loss of efficiency. A fractional factorial screening experiment to determine factors affecting discrete part process control. Naugraiya. M and Drury, C.G [6], have utilized a 2^6 -1 fractional factorial design to investigate human factors in manufacturing processes. It found that task difficulty reduced performance, while individual experience and training improved it. Analysis of ranked data in randomized blocks when there are missing values Besta. D.J and Raynera. J.C.W [2] have introduced a novel method for analyzing *RBD* using rank data and utilized an *ANOVA* on ranks, which surpassed the Skillings-Mack test in terms of indicative sizes and powers. Missing plot techniques using regression analysis and Its comparison with *RBD* missing plot techniques derived by Richa Seth, et al [12] investigated the estimated values of one and two missing observations using the regression method analysis. Also estimate the value of the same missing observations by expressing the regression data. Analysed the *RBD* data format. Also analysed the data to obtain an *ANOVA* table using both methods, where I found the same significant result. Construction of control chart using fuzzy probabilistic approach for cotton sweater product Pachamuthu. M. and Shanmugasundram. V [7], had discussed, creation of fuzzy statistical controlled chart used the fuzzy membership function proposed. It illustrates the cotton sweater product of quality evaluation such as color, designs, quality and price of a product with consumer's demographic characters on a hedonic ruled. A statistical analysis of Latin square designs through fuzzy ranking approach using median value have discussed by Pachamuthu. M et al., [3]. Pachamuthu. M and Mariappan. A [9] have constructed the statistical quality control chart using fuzzy probabilistic approach and proposed different solutions to prevented air pollution. Also, developed a new methodology for construction of statistical quality-controlled chart used fuzzy probability approached. Application of this method had established through the air pollution controlled caused illustration with consumer's demographic characters on a hedonic ruled. Construction of 3σ control chart using fuzzy probabilistic approach for quality evaluation by Pachamuthu. M. et al., [8]. The 3σ fuzzy statistical quality-controlled charts played an important role for smart controlled of home appliances. Also, constructed the 3σ control charts used fuzzy probabilistic approached for quality evaluation. An optimization method of construction for warping copper plates and engines using complete block designs with some special types of graphs by Pachamuthu M et al., [5]. The optimization method for constriction of randomized blocked design and Latin square design used bipartite and regular graphs with applications. Use of control charts for conducting *ANOVA* Study Ghosh. D.K et al., [4] have explored was used to control charts for an empirical study of analysis of variance. The modified control charts were renamed as *FES* charts. The control limits were labelled as lower selection limit and upper selection limit. Also showed that the results of the control chart could be compared if the 3σ control limits were kept intact. *ANOVA* was conducted at the 0.27% significance level. On the construction of balanced incomplete block designs through factorization and coloring graphs using mutually orthogonal Latin square designs Pachamuthu. M and Ramya. A [10]. A statistical analysis of variance for one-way classification with addition operations using triangular fuzzy numbers discussed by Pachamuthu. M and Mariappan. A [11]. A statistical hypotheses technique of one-way analysis of variance using triangular fuzzy numbers: a comparative study Pachamuthu. M et al., [15] have discussed, a comparative studied on the statistical tested of *ANOVA* for one way with ∝-cut interval method, Exponential Triangular FuzzyNumbered *(ETFNs),* and Logarithmic Triangular Fuzzy Numbers *(LTFNs)* used the Triangular Fuzzy Numbers *(TFNs)* based on the decision ruled and illustrated with numerical examples. In this paper, a statistical analysis and estimate the one missing value of *RBD* using 3σ and 6σ control chart through *RSM* with numerical example.

II. Methods

I. Definition of *RBD*

Let us suppose that experimental material is divided into *r* blocks and *t* treatments. Each block is then divided into *t* units, and the treatments are allocated within a block randomly the resulting design is called an *RBD*.

II. Definition of Missing Plot Technique

In field experimentation, some of the experimentation from certain experimental units may not be available due to some practical reasons. In such the case of observations which are not available are called missing observation. The technique of estimating the missing value and carryout the statistical analysis in experimental designs is called as missing plot techniques.

III. Control Chart

Control charts are graphical representations of process data over time. They show whether a manufacturing process is stable and operating within expected parameters using statistical limits.

Whether a manufacturing process is in control or out of control can be determined through a technique called a control chart. The 3σ control charts limits are; Upper Control Limit (UCL) = $\mu + 3\sigma$, Center Limit (CL) = μ and Lower Control Limit (LCL) = $\mu - 3\sigma$. Then, 6σ control charts limits are Upper Control Limit (UCL) = $\mu+6\sigma$, Center Limit (CL) = μ and Lower Control Limit (LCL) = $\mu - 6\sigma$.

IV. Interaction Plot

An interactive plot shows that the relationship between one categorical factor and a continuous response depends on the value of the second categorical factor. This plot shows the means for levels of one factor on the axis and a separate line for each level of another factor.

V. Residual Plot and Contour Plot

A residual plot shows the difference between the predicted values and the actual values of the response variable. An ideal residual plot, called a null residual plot, shows a random scatter of points that form a band of approximately constant width around the identity line. Contour plots or contour maps are providing one of the most obvious ways to illustrate and interpret response surface structure. Contour plots are two-dimensional or three-dimensional graphs in which the axis structure is a specific pair of design variables, while the other design variables are held constant. Definitions are not created by finite equations. Each point on an edge has a standard error.

VI. 3*D* Surface

A 3*D* surface plot is a way of visualizing data with three dimensions such as length, width and height. In a 3*D* surface plot, these points are plotted in three-dimensional space, creating a surface that shows how the data varies at different levels. It's like looking at a three-dimensional map of data, where the height of the surface represents the value of the data at each point.

III.Methods

I. A Statistical Analysis and Estimate the one missing value of *RBD*

Let us consider y_{ij} is the general observations and the layout table is given below;

Table 1: *Layout of RBD*

	Blocks(B_j)					
Treatments(T_i)	1	…	j	…	r	Total
1	y_{11}	…	y_{1j}	…	y_{1r}	$y_{1.}$
⋮	⋮		⋮		⋮	
i	y_{i1}	…	y_{ij}	…	y_{ir}	$y_{i.}$
⋮	⋮		⋮		⋮	
T	y_{t1}	…	y_{tj}	…	Y_{tr}	$y_{t.}$
Total	$y_{.1}$	…	$y_{.i}$	…	$y_{.r}$	$y_{..} = G$

Linear model of *RBD*

$$y_{ij} = \mu + a_i + b_j + e_{ij} \qquad i = 1,2,\dots,t; \qquad j = 1,2,\dots,r$$

where, y_{ij} is the observation corresponding to i^{th} of treatment and j^{th} blocks, μ refers the general mean effect which is fixed, a_i is fixed effect due to the i^{th} treatment b_j is fixed effect due to the j^{th} blocks and e_{ij} random error effect $N(0,\sigma^2)$. Here, the parameters μ, a_i and b_j can be estimated by the principle of least squares. $\frac{\partial E}{\partial \mu} = 0$, $\frac{\partial E}{\partial a_i} = 0$ and $\frac{\partial E}{\partial b_j} = 0$. Then, the normal equations are

$$e_{ij} = y_{ij} - \mu - a_i - b_j$$

Summing and squaring on both sides;

$$E = \sum_{i=1}^{t} \sum_{j=1}^{r} e_{ij}{}^2 = \sum_{i=1}^{t} \sum_{j=1}^{r} \left(y_{ij} - a_i - b_j - \mu\right)^2$$

Differentiating with respect to the parameters are;

$$\frac{\partial E}{\partial \mu} = -2\sum_{i=1}^{t} \sum_{j=1}^{r} \left(y_{ij} - a_i - b_j - \mu\right) = 0$$
$$\frac{\partial E}{\partial a_i} = -2\sum_{j=1}^{r} \left(y_{ij} - \mu - a_i - b_j\right) = 0 \qquad \text{and}$$
$$\frac{\partial E}{\partial b_j} = -2\sum_{i=1}^{t} \left(y_{ij} - \mu - a_i - b_j\right) = 0$$

The estimate of μ is given by,

$$\hat{\mu} = \frac{G}{tr}$$

The estimate of a_i is given by,

$$\hat{a}_i = \frac{A_i}{r} - \frac{G}{tr}$$

Then b_j is given by,

$$\hat{b}_j = \frac{B_j}{t} - \frac{G}{tr}$$

These three equations (2), (3) and (4) substituting in (1), we get sum of square are;

$$Q = \text{TSS} = \sum_{i=1}^{t} \sum_{j=1}^{r} \left(y_{ij}{}^2 - \frac{G^2}{tr}\right) \quad \text{which has} \quad (tr-1)df$$

$$Q_t = \sum_{i=1}^{t} \left(\frac{T_i^2}{t} - \frac{G^2}{tr}\right) \quad \text{which has} \quad (r-1)df$$

$$Q_b = \sum_{i=1}^{b} \left(\frac{B_j^2}{r} - \frac{G^2}{tr}\right) \quad \text{which has} \quad (t-1)df$$

$$Q_E = Q - Q_t - Q_b \text{ which has } (t-1)\,(r-1)df$$

Sometimes, points earned in tests are lost due to unavoidable circumstances. Therefore, we can estimate the missing values.

Table 2: *Layout of one missing* in *RBD*

	Blocks(B_j)					
Treatments(T_i)	1	…	j	…	R	Total
1	x_{11}	…	x_{1j}	…	x_{1r}	
⋮	⋮		⋮		⋮	
i	x_{i1}	…	x_{ij}	…	x_{ir}	$T=T'+x$
⋮	⋮		⋮		⋮	
t	x_{t1}	…	x_{tj}	…	x_{tr}	
Total			$B=B'+x$			

Let $X = x_{ij}$ be the one missing observation of *RBD* corresponding to the i^{th} treatment and j^{th} block; Let G be the grant total of all tr number of observations; G denote the grant total of *(tr-1)* number of observations excluding the missing observation x, Then $G = G'+x$. Let T be the i^{th} treatment total of r observations. Let T' denote the i^{th} treatment total of $(t-1)$ observations excluding the missing observation x then,

$$T = T' + x$$

Let B be the j^{th} block total of r observations; B' denote the j^{th} block total of $(r-1)$ observations x then,

$$B = B' + x$$

The correction factor ($c\,f$) is define by,

$$cf = \frac{G^2}{N} = \frac{(G'+x)^2}{tr}, \qquad N = tr$$
$$TSS = Q = \text{ constant } + x^2 - cf$$
$$TrSS = Q_t = \text{ constant } + \frac{(T'+x)^2}{t} - cf$$
$$BSS = Q_b = \text{ constant } + \frac{(B'+x)^2}{r} - cf$$

Then the Error Sum of Square *(ESS)* can be obtained by subtraction

$$Q_E = Q - Q_t - Q_b$$

$$ESS = x^2 - \frac{(T'+x)^2}{t} - \frac{(B'+x)^2}{r} + \frac{(G'+x)^2}{tr} + \text{ constant}$$

To estimate the missing value of x such that error is minimum of this value and we must have,

$$\frac{\partial E}{\partial x} = 0 = 2x - \frac{2(T'+x)}{t} - \frac{2(B'+x)}{r} - \frac{2(G'+x)}{tr} = 0$$

Then

$$\hat{x} = \frac{rT'+tB'-G'}{(t-1)(r-1)}$$

The estimate of the missing observation is fitted in the given data and the statistical analysis of *RBD* is carried out. All the values are referring in the *ANOVA* table and inference is drawn.

Table 3: *ANOVA* table for *RBD* with one missing value

SV	*Df*	*SS*	*TSS*	*F - Ratio*
Treatments	$(t-1)$	$Q_t = \sum_{i=1}^{t}\left(\frac{T_i^2}{r} - \frac{G^2}{tr}\right)$	$M_t = \frac{Q_t}{(r-1)}$	$F_t = \frac{M_t}{M_e}$
Blocks	$(r-1)$	$Q_b = \sum_{j=1}^{r}\left(\frac{B_j^2}{t} - \frac{G^2}{tr}\right)$	$M_b = \frac{Q_b}{(t-1)}$	$F_b = \frac{M_b}{M_e}$
Error	$(t-1)(r-1) - 1$	$Q_E = Q$ -Q_t - Q_b	$M_e = \frac{Q_E}{(t-1)(r-1)-1}$	-
Total	$(tr-2)$	$Q = \sum_{i=1}^{t}\sum_{j=1}^{r}\left(y_{ij}^2 - \frac{G^2}{tr}\right)$	-	-

Inference: If the observed value of F_0 is greater than the expected value of F_e ($F_0 > F_e$) at α% level of significance then the null hypothesis of equal treatment is rejected. Otherwise, it is accepted. Similarly, the block effect may also be tested.

IV. Application

In our study, the data was taken from 3–5 months yield of pulses data on agricultural land in India. There are various variety of pulses for green gram, red gram and horse gram (in kg and 000's) and years for 2016, 2017, 2018, 2019 and 2020 respectively the data are given below and test whether there is any significant difference between yield of pulses in different year and also estimate the yield of pulse in year 2027.

Table 4: *Different Variety of Pulses and Years*

Pulses Variety	Yield of Pulse(in Tonnes)				
	2016	2017	2018	2019	2020
Green Gram	-	56.0	78.2	76.8	76.4
Red Gram	59.9	37.0	53.7	51.6	53.2
Horse Gram	44.6	12.3	48.5	53.5	61.3

Null hypothesis(H_0): There is no significant difference between yield of pulses in different year.

Solution: The above data is one missing value problem. This problem may be using different methods. In this paper we use *RBD* one missing value method. First estimate the missing value,

$$\hat{x} = \frac{rT' + tB' - G'}{(t-1)(r-1)}$$

$$\hat{x} = \frac{287.4(3) + 104.5(5) - 763}{(3-1)(5-1)}$$

$$\hat{x} = 77.5$$

Table 5: *The Estimated Observation for Pulses and Years*

Pulses Variety	Yield of Pulse(in Tonnes)				
	2016	2017	2018	2019	2020
Green Gram	77.5	56.0	78.2	76.8	76.4
Red Gram	59.9	37.0	53.7	51.6	53.2
Horse Gram	44.6	12.3	48.5	53.5	61.3

Then substitute the missing value in above table and carried out the statistical analysis. The missing observation of the green ram will be 77.5 in 2016. It was found by using one missing observation of *RBD*. The above data was InControl or out of control. Using the 3σ control chart for different variety and different years. Upper Control Limit (UCL) = 111.7173, Center Limit *(CL)* = 56.033 and Lower Control Limit *(LCL)* = 0.3530 and also using the 6σ control chart for different variety and different years. Upper Control Limit (*UCL*) = 130.173, Center Limit (*CL*) = 56.033 and Lower Control Limit (*LCL*) = -18.106.

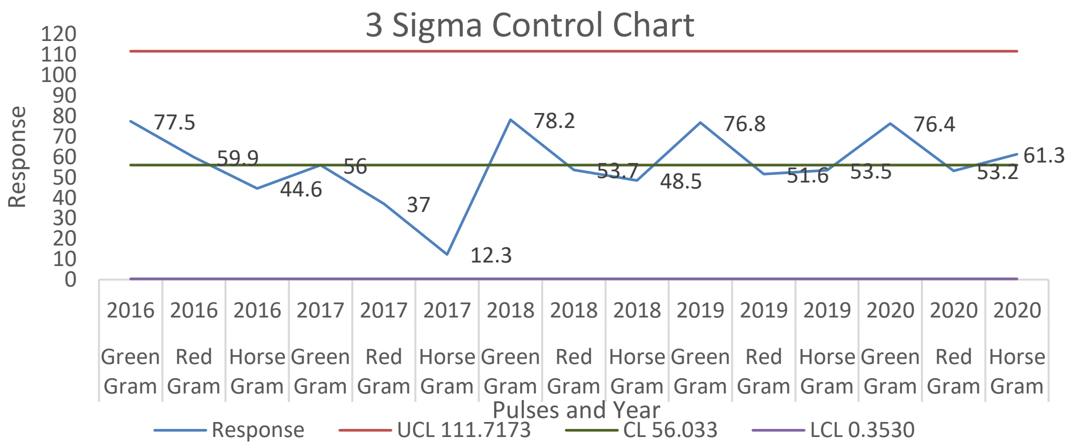

Figure 1: *3 σ Control Chart*

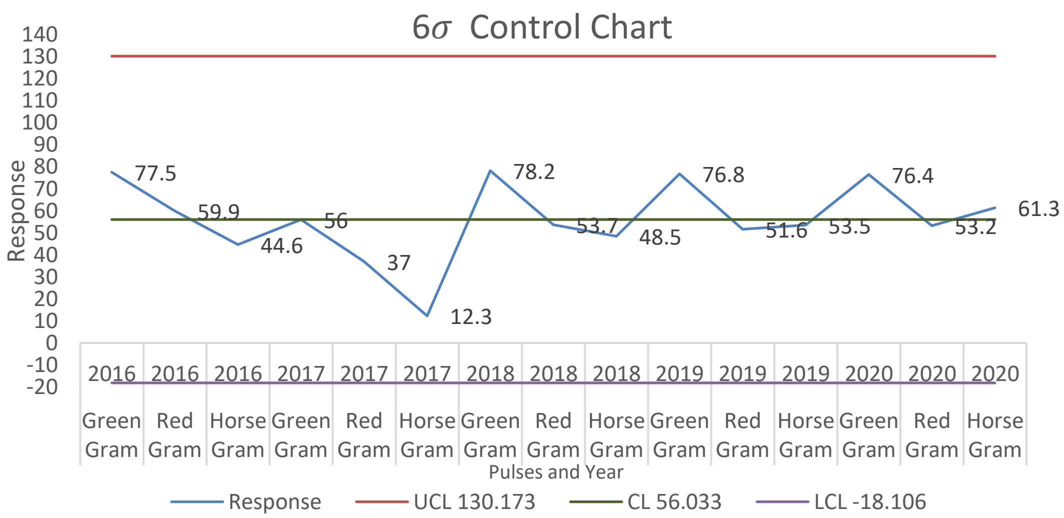

Figure 2: *6 σ Control Chart*

The above figures 1 and 2 represents the pulses of green gram, red gram and horse gram in 2016,2017,2018,2019 and 2020, which are under the control limit of $3\sigma$and 6σ. The data is in control, so we continue with further analysis.

Table 6 : *Conversion of table from variety to Pulses variety*

Year	Variety	Response	Pulses Variety
2016	Green gram	77.5	1
2017	Green gram	56.0	1
2018	Green gram	78.2	1
2019	Green gram	76.8	1
2020	Green gram	76.4	1
2016	Red gram	59.9	2
2017	Red gram	37.0	2
2018	Red gram	53.7	2
2019	Red gram	51.6	2
2020	Red gram	53.2	2
2016	Horse gram	44.6	3
2017	Horse gram	12.3	3
2018	Horse gram	48.5	3
2019	Horse gram	53.5	3
2020	Horse gram	61.3	3

In the table above, we convert the Pulses varieties of green gram, red gram and horse gram into number 1, 2 and 3 for our convenience.

Table 7 : *Model Summary*

S	*sq*	*R-sq(adj)*	*R-sq(pred)*
7.06029	90.82%	89.93%	67.72%

The above table represent the model summary of R^2 value is 90.82% and R^2 value is 89.93%.

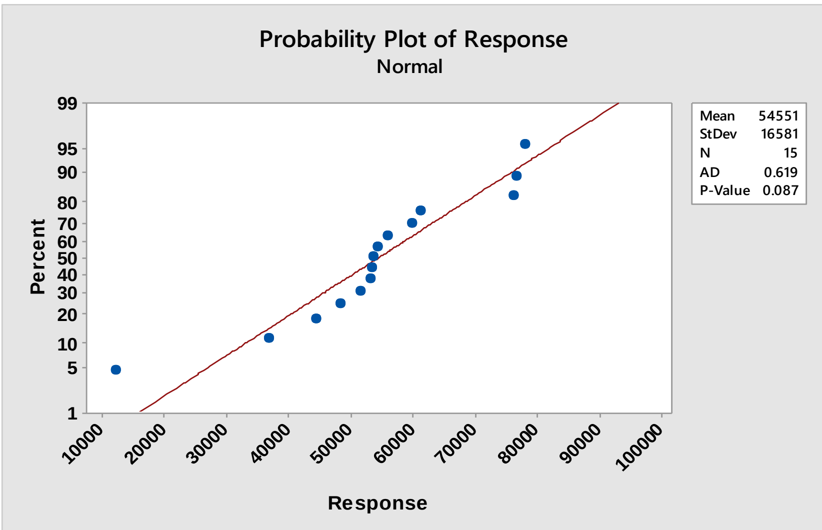

Figure 3: *Normal Plot*

This figure 3 is a probability plot of a response variable, showing that the data fit a normal distribution. It has a red line representing the expected normal distribution and blue dots representing the actual data points. The data points follow the reference line, indicating a reasonable fit to a normal distribution. The p-value is 0.139. Therefore, a p-value greater than 0.05 indicates that we have failed to reject the null hypothesis, and then the data conform to a normal distribution.

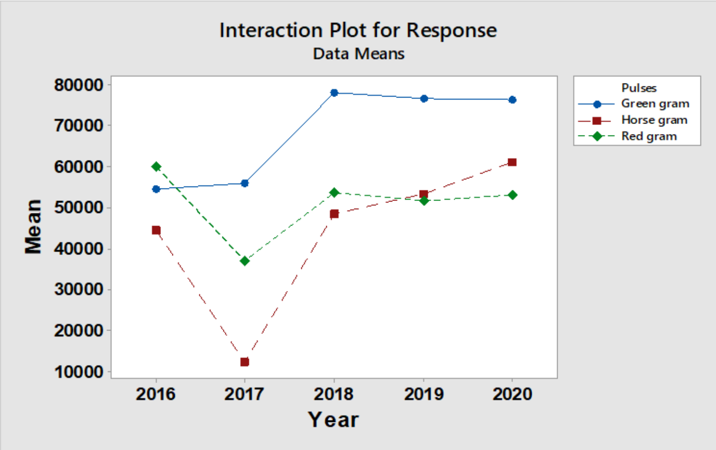

Figure 4: *Interaction Plot*

This figure 4 is a correlation graph for the response showing data for different types of pulses and different years. The graph showing the average values for each type of pulse has changed over the years, with significant interaction and trends across the different types.

Table 8 : *Coefficients*

Term	*Effect*	*Coef*	*SE Coef*	*T-Value*	*p-Value*	*VIF*
Constant		46.42	6.74	6.89	0.000	
Year	12.59	6.29	4.53	1.39	0.198	1.00
Pulsesvariety	-28.94	-14.47	3.92	-3.69	0.005	1.00
Year*Year	18.63	9.31	7.65	1.22	0.254	1.00
Pulsesvariety*Pulsesvariety	14.86	7.43	6.79	1.09	0.302	1.00
Year*Pulsesvariety	11.20	5.60	5.54	1.01	0.339	1.00

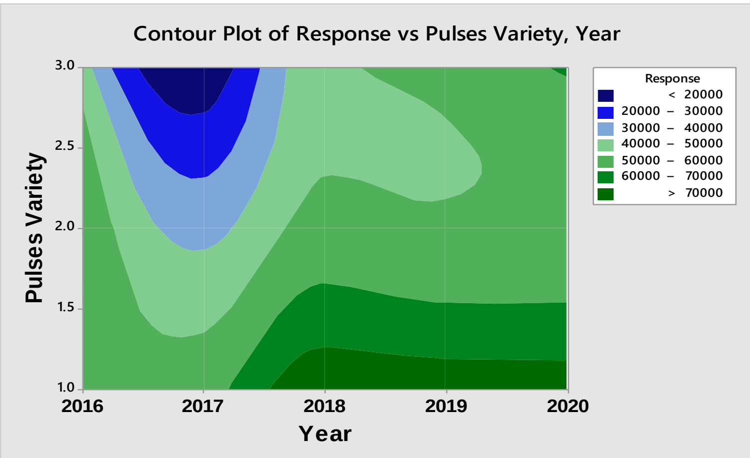

Figure 5: *Contour Plot*

The above figure 5 shows the relationship between the variables pulses variety and year, and the response variable is represented by different color gradients. These darker colors represent higher values in this graph.

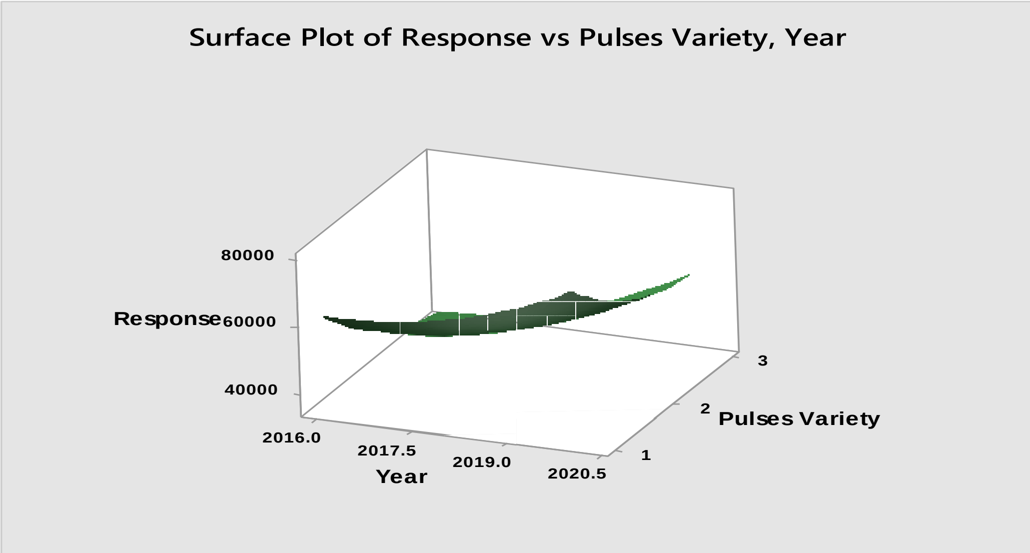

Figure 6: *3D Surface Plot*

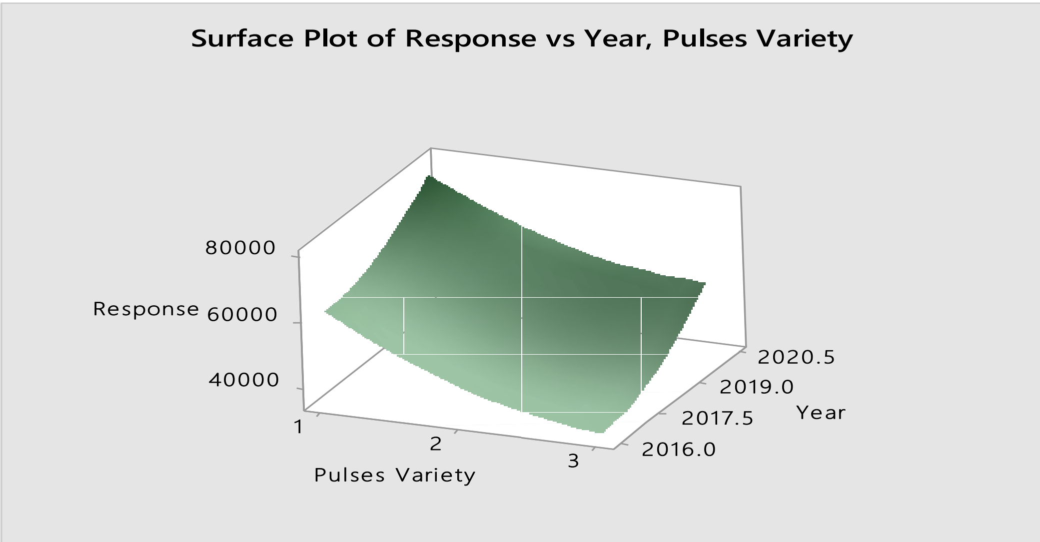

Figure 7: *3D Surface Plot*

The above figure 6 and 7 are represent the *3D* surface plot of response variable year and Pulses variety.

Table 9 : *ANOVA table for RBD layout*

SV	Df	*SS*	*MSS*	*F- Ratio*	*p-Value*
Variety	2	2277.8	1138.91	22.85	0.000
Year	4	1666.2	416.55	8.36	0.006
Error	8	687319021	85914878	-	-
Total	14	3849123322	-	-	-

Inference: Since the *p*-values of variety and years are less than 0.05, we reject the null hypothesis. Therefore, there is a significant difference in the yield of pulses across different years. However, the data is under control. Next, we forecasting for green gram, red gram and horse gram for next seven year (2027) are given below.

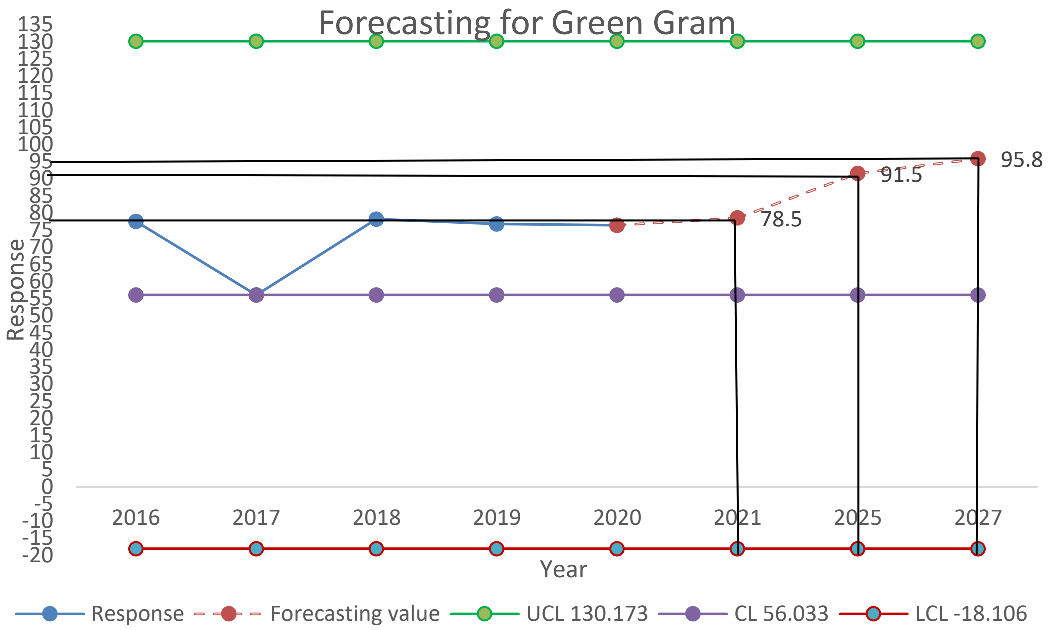

Figure 8: *Forecasting for Green Gram from 2021 to 2027*

The above figure 8 shows the green crop yield from 2016 to 2020. A predicted the future yields for the following years: 2021: 78.5, 2022: 86.1, 2023: 84.5, 2024: 88.0, 2025: 91.5, 2026: 94.1, and 2027: 95.8.

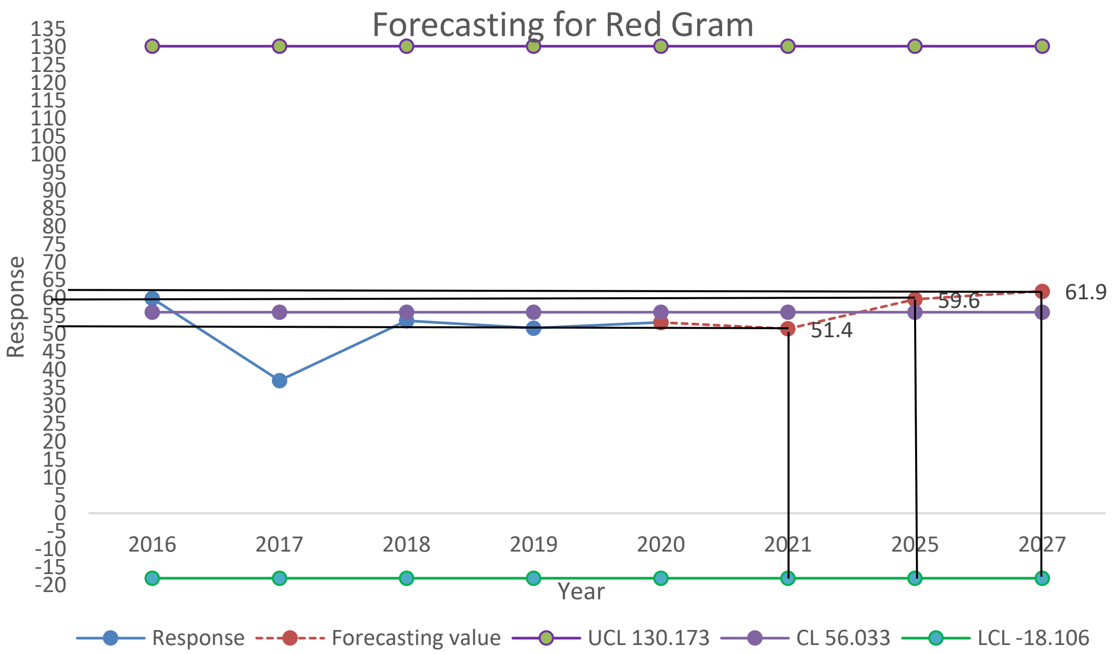

Figure**9:** *Forecasting for Red Gram from 2021 to 2027*

The above figure 9 shows the yield of red gram from 2016 to 2020. A predicted the subsequent yield for the following years: 2021: 51.4, 2022: 57.9, 2023: 56.0, 2024: 58.1, 2025: 59.6, 2026: 61.6 and 2027: 61.9.

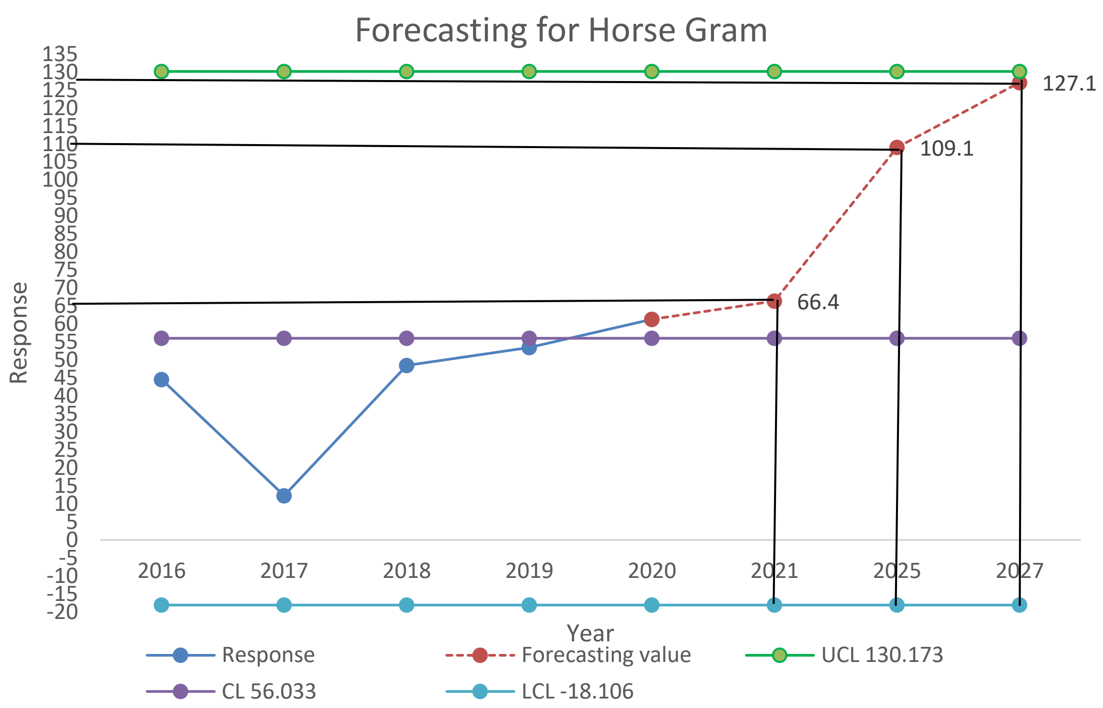

Figure 10: *Forecasting for Horse Gram from 2021 to 2027*

The above figure 10 shows the yield of the horse gram from 2016 to 2020. A predicted the yield for the following years: 2021: 66.4, 2022: 84.7, 2023: 88.4, 2024: 98.9, 2025: 109.1, 2026: 119.4, and 2027: 127.1.

V. Conclusion

In this paper, new methods were introduced to estimate a missing value in an *RBD* and to determine whether the data is in control or out of control using control charts, specifically 3σ and 6σ control charts, along with *RSM* in a numerical example. First, one missing value was estimated as $\hat{x} = 77.5$ using the *RBD* one-missing-value formula, and the original dataset was then fitted. Statistical analysis was carried out using statistical software, confirming a significant difference in the yield of pulses across different years. Also, the study provides future yield predictions for pulses up to the year 2027, offering valuable insights for agricultural planning and decision-making. The predicted yields are as follows: green gram: 2021: 78.5, 2022: 86.1, 2023: 84.5, 2024: 88.0, 2025: 91.5, 2026: 94.1, 2027: 95.8. red gram: 2021: 51.4, 2022: 57.9, 2023: 56.0, 2024: 58.1, 2025: 59.6, 2026: 61.6, 2027: 61.9. horse gram: 2021: 66.4, 2022: 84.7, 2023: 88.4, 2024: 98.9, 2025: 109.1, 2026: 119.4, 2027: 127.1. The integration of control charts and *RSM* proves to be an effective approach for handling missing values and optimizing responses in agricultural datasets. For future research, this methodology can be extended to other crop varieties and environmental conditions to enhance predictive accuracy. Furthermore, incorporating advanced machine learning techniques and hybrid statistical models may further refine missing value estimations and improve yield forecasting. Exploring the impact of climate change, soil conditions, and other agronomic factors alongside *RSM*-based estimations could provide a more comprehensive framework for agricultural yield analysis.

References

[1] Azadeh, A. et al (2013). Optimum estimation of missing values in randomized complete block design by genetic algorithm. *ScienceDirect Elsevier.* 37, 37-47, https://doi.org/10.1016/j.knosys.2012.06.014.

[2] Best, D.J and Rayner, J.C.W (2016). Analysis of ranked data in randomized blocks when there are missing values. *Journal of Applied Statistics*. 44(1), 16-23, https://doi.org/10.1080/02664763.2016.1158245.

[3] Pachamuthu, M. et al (2018). A statistical analysis of latin square designs through fuzzy ranking approach using median value. *International journal of agricultural & statistical sciences*. Vol. 17(1), pp. 163-170, https://connectjournals.com/03899.2021.17.163.

[4] Ghosh, D.K. et.al (2019). Use of control charts for conducting anova study. *International Journal of Agricultural and Statistical Science*. 15(1), 375-384, ISSN: 0973-1903.

[5] Pachamuthu, M. et al (2018). An Optimisation Method of Construction for Warping Copper Plates and Engines Using Complete Block Designs with Some Special Types of Graphs. *Advances in Materials Science and Engineering*. Volume 2023, Article ID 9002743, 12 pages https://doi.org/10.1155/2023/9002743.

[6] Naugraiya, M. and Drury, C.G. (2014). A fractional factorial screening experiment to determine factors affecting discrete part process control. *Theoretical Issues in Ergonomics Science*. 10:1, 1-17, http://dx.doi.org/10.1080/14639220802075832.

[7] Pachamuthu, M. and Shanmugasundram, V. (2017). Construction of control chart using fuzzy probabilistic approach for cotton sweater product. *Research Journal of Mathematical and Statistical Sciences*. Vol. 5(4), 1-11.

[8] Pachamuthu, M. et al (2018). Construction of 3-Sigma Control Chart Using Fuzzy Probabilistic Approach for Quality Evaluation. *Science and Technology*. Volume 4, Issue 5, ISSN: 2395-6011, Online ISSN: 2395-602X.

[9] Pachamuthu, M. and Mariappan, A. (2018). On the Construction of Statistical Quality Control Chart Using Fuzzy Probabilistic Approach. *Science and Technology*. Volume 4, Issue 5, ISSN: 2395-6011, Online ISSN: 2395-602X.

[10] Pachamuthu, M. and Ramya, A. (2019). On the construction of balanced incomplete block designs through factorization and coloring graphs using mutually orthogonal Latin square designs. *International journal of agricultural and statistical sciences*. Vol. 15(2).

[11] Pachamuthu, M. and Mariappan, A. (2021). A Statistical Analysis of Variance for One Way Classification with Addition Operations using Triangular Fuzzy Numbers. *International Journal of Statistics and Reliability Engineering*. Volume 8, Issue 1, 45-56.

[12] Richa Seth, et.al (2017). Missing plot techniques using regression analysis and its comparison with randomized block design missing plot techniques. *International Journal of Applied Mathematics and Statistical Sciences.* 6(4), 55-66. ISSN(P): 2319-3972; ISSN(E): 2319-3980.

[13] Subramani, J and Balamurali, B. (2012). Loss of efficiency in randomized block designs with two missing values. *Elixir International*. 49, 10042-10044.

[14] Tae-Yeon ChoGhosh, et.al (2011). Mixture-process variable experiments including control and noise variables within a split-plot structure. *International Journal of Quality Engineering and Technology.* 2(1), 1-28, https://doi.org/10.1504/IJQET.2011.038719.

[15] Pachamuthu, M. et al (2022). A statistical hypotheses technique of one-way analysis of variance using triangular fuzzy numbers. *Advances and Applications in Mathematical Sciences.* Volume 21, Issue 6, Pages 2989-3009.