VOLUME - 11, ISSUE - 03, MARCH - 2022 • PRINT ISSN No. 2277 - 8160 • DOI : 10.36106/gjra

Original Research Paper

Statistics

A NOTE ON COW MILK DATA USING RANDOMIZED COMPLETE BLOCK DESIGN

R. Sasikumar

Department of Statistics, Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu, India.

ABSTRACT In a completely randomized design, the experimental material should be homogeneous. Usually, the experimental materials are not so homogeneous in nature, particularly in agricultural field experiments. In such a situation, the principle of local control is adopted and the experimental material is grouped into homogeneous sub-groups. In this paper, cow milk data are used and applied four different types of treatments. The yield of a cow is analyzed and interpreted using Randomized Complete Block Design (RCBD). Results are explored, tabulated and interpreted.

KEYWORDS : RCBD, critical difference, two-way analysis of variance, cow milk data

1. INTRODUCTION

Milk is approximately 3.3% protein and contains all of the essential amino acids. The protein content of some milk varieties is shown in the nutrient content tables. Proteins are the fundamental building blocks of muscles, skin, hair, and cellular components. Proteins are needed to help muscles contract and relax, and help repair damaged tissues.

They play a critical role in many body functions as enzymes, hormones, and antibodies. Proteins may also be used as an energy source by the body. Nine amino acids must be obtained from the diet and are referred to as the "essential" amino acids: leucine, isoleucine, valine, phenylalanine, tryptophan, histamines, threonine, methionine, and lysine. Proteins that contain all 9 essential amino acids are often called "complete" proteins. Proteins of animal origin and soy are complete proteins, whereas proteins from grains and legumes are missing 1 or more of the essential amino acids, which means that consumers must eat complementary foods in order to get all of the essential amino acids. Milk protein consists of approximately 82% casein and 18% whey (serum) proteins. Both casein and whey proteins are present in milk, yogurt, and ice cream. The content of specific nutrients in milk, important background information on the chemistry of milk energy, fat, water, protein and enzymes.

Also, milk prevented to heart health disease, diabetes, weight loss, inflammatory issues, growth and development cells, clear function of immune body systems and giving the strength for bones and teeth. In this paper, we analyzed cow milk for improving the yield through the four different types of treatments.

The Mann-Whitney test for two samples or the Kruskal-Wallis test for k samples could be used for testing differences in treatment effects under a completely randomized design (Kruskal and Wallis, 1953). Analysis of variance is the fundamental tool for analyzing data from designed experiments (Cochran and Cox, 1957; Searle, 1971). The randomization protocol reduces any bias in favor of particular treatments, while the blocking enables extraneous variation to be absorbed into block effects. Consequently, one obtains better estimates of treatment effects and more powerful tests for treatment differences.

Blocking refers to the division of experimental runs into smaller sub-groups, or blocks. Each treatment is applied randomly to a number of subjects within each block. This design, known as a Randomized Complete Block Design (RCBD), is commonly employed in biological experiments, where, for example, experimental runs on a given day may be treated as a block (Sokal and Rohlf, 1981). Mafra-Neto and Cardé (1998) utilized an RCBD to test the effect of treatments.

Linn et al. (1996) explained that they do not have a complete randomization protocol; however, they analyzed their experimental data using an ANOVA. Hollander and Wolfe (1999) proposed several well-known nonparametric tests exist for testing differences in treatment effects depending on the type of design used.

2. MATERIALS AND METHODS

Four different treatments have been applied to the four different cows and the milk yield were collected. Based on this data sets, RCBD with m=29 observation per cell have been applied and the results are discussed.

2.1. RCBD

In this design, the whole experimental materials is divided into homogeneous groups, each of which constitutes a single replications. Each of these groups is further divided into a number of experimental units which are equal in all respects. The treatments are applied to these units by any random process. In case of field experiments, if it is observed that the fertility gradient of the field is in one direction, the whole field may be divided into a number of equal blocks and then each block be divided into a number of equal plots. The number of plots in each block is equal to the number of treatments, so that each block is a replicate of each treatment.

Let there be k treatments. Each of the treatments is replicated the same number of times in this design. Let r denote the number of replications of each treatment. The total number of experimental units is therefore, kr. These units are arranged into r groups, each of size k. The error control measure in this design consists of making the units in each of these groups homogeneous. These groups are commonly known as blocks and experimental units in the blocks are known as plots.

The following points are important for this design.

(a) The number of blocks must be equal to the number of replications fixed for each treatment.
(b) The number of plots in each block should be equal to the number of treatments.
(c) An important and essential point, on which the attention is kept, is that the experimental errors within each block are to be kept as small as practically possible and the variation from block to block as great as possible. In this way all the treatments which are assigned to one block, experience the same type of environmental effects, and are, therefore, comparable.
(d) Randomization of treatments in each block should be afresh.

The Randomized Block Design (RBD) is often called the Randomized Complete Block Design (RCBD) because each block contains a complete set of treatments.

The design provides a two-way classified data according to the levels of two factors, viz., blocks and treatments. For its analysis the following model is taken:

$$y_{ij} = \mu + t_i + b_j + e_{ij} \qquad i=1,2,3,\ldots,k \text{ and } j=1,2,3,\ldots,r.$$

where y_{ij} is the yield obtained from the j^{th} block, receiving the ith treatment, μ is the general mean, ti is the ith treatment effect and b_j is the effect j^{th} block and eij is the error component. The error components are assumed to be independently and normally distributed with zero mean and constant variance σ^2

Table 1: Randomized Complete Block Design

Treatment	Blocks						Totals
	1	2	...	j	...	r	
1	y_{11}	y_{12}	...	y_{1j}	...	y_{1r}	$T_{1.}$
2	y_{21}	y_{22}	...	y_{2j}	...	y_{2r}	$T_{2.}$
3	y_{31}	y_{32}	...	y_{3j}	...	y_{3r}	$T_{3.}$
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
i	y_{i1}	y_{i2}	...	y_{ij}	...	y_{ir}	$T_{i.}$
:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:
k	y_{k1}	y_{k2}	...	y_{kj}	...	y_{kr}	$T_{k.}$
Total	$B_{.1}$	$B_{.2}$	...	$B_{.j}$	...	$B_{.r}$	$y_{..}=G$

Let $T_i = \sum_{j=1}^{r} y_{ij} (i = 1, 2, 3, 4, \ldots, k)$ be the total i^{th} – treatment and

$B_j = \sum_{i=1}^{k} y_{ij}$ $(j=1,2,3,\ldots,r)$ be the total of j^{th} block.

- Grand Total = $G = \sum_{i=1}^{k} T_i = \sum_{j=1}^{r} B_j$
- Correction Factor (C.F.) = $\frac{G^2}{(k*r)}$
- Raw Sum of Squares = $\sum_{i=1}^{k}\sum_{j=1}^{r} y_{ij}^2$
- Total Sum of Squares (TSS) = Raw Sum of Squares – C.F.
- Treatment Sum of Squares = TrSS = $\frac{1}{r}\sum_{i=1}^{k} T_i^2 - C.F.$
- Block Sum of Squares = BSS = $\frac{1}{k}\sum_{j=1}^{r} B_j^2 - C.F.$
- Error Sum of Squares ErSS = TSS – TrSS – BSS.

Table 2: Analysis of Variance of a Randomized Block Design

Source of Variation	Degrees of freedom	Sum of Squares	Mean Square	F
Blocks	r-1	$\frac{1}{k}\sum_{j=1}^{r} B_j^2 - C.F.$	S_B^2	$\frac{S_B^2}{S_E^2}$
Treatments	k-1	$\frac{1}{r}\sum_{i=1}^{k} T_i^2 - C.F.$	S_T^2	$\frac{S_T^2}{S_E^2}$
Error	(r-1)(k-1)	By subtraction	S_E^2	
Total	rk-1	$\sum_{i=1}^{k}\sum_{j=1}^{r} y_{ij}^2 - C.F.$		

The hypothesis that the treatments have equal effects is tested by F-test where F is the ratio of $\frac{S_T^2}{S_E^2}$ with (k-1) and (r-1)(k-1) degrees of freedom. If F is non-significant the data do not suggest that the treatment effects are different. When F is significant, we conclude that the treatment effects are different.

2.2. Critical Difference

If the treatments show significant effect, then we would be interested to find out which pair(s) of treatments differ significantly. For this, instead of calculating Student's t for different pairs of treatment means, we calculate the Least Significant Difference (L.S.D) at the given level of significance. This least significant difference is known as the Critical Difference (C.D).

The C.D. between any two treatment means, say $\bar{y}_{i.}$ and $\bar{y}_{j.}$ at level of significance 'α' is given by:

C.D. $(\bar{y}_{i.} - \bar{y}_{j.})$ = [The critical value at level of significance α and error d.f.] x $[S.E > (\bar{X}_i - \bar{X}_j)]$

But stands for $V(\bar{y}_{i.} - \bar{y}_{j.}) = \frac{\sigma_e^2}{n_i} + \frac{\sigma_e^2}{n_j}$ $\Rightarrow$ $S.E.(\bar{y}_{i.} - \bar{y}_{j.}) = \sigma_e\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$

$\therefore$ $C.D.(\bar{y}_{i.} - \bar{y}_{j.}) = t_{n-k}(\alpha/2).\sqrt{M.S.S.E\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} = t_{n-k}(\alpha/2).s_E\sqrt{\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$

where S_e^2 provides an unbiased estimate of the error variance σ_e^2

In particular, if $n_i = n\ \forall\ i = 1,2,\ldots,k$, *i.e*, if each treatment is replicated n times, then

$C.D.(\bar{y}_{i.} - \bar{y}_{j.}) = t_{n-k}(\alpha/2).S_E\sqrt{2/n}$

Here $t_v(\alpha)$ is the right-tailed critical value of t for v d.f. at level of significance , so that $P[t > t_v(\alpha)] = \alpha$

If the difference $|\bar{y}_{i.} - \bar{y}_{j.}|$ between any two treatments means is greater than CD (or LSD). It is said to be significant, otherwise it is not significant.

2.3. Two-way analysis of variance

As with the t-test and the one-way ANOVA, we assume that the variances are homogeneous. In a manner analogous to the one-way ANOVA's within-group sums of squares, we can calculate a within-cells sums of squares and corresponding degrees of freedom which under the assumption of constant variance, can be used to obtain a pooled variance common to all cells:

$$within-cells\ SS = \sum_{i=1}^{a}\sum_{j=1}^{b}[\sum_{1=1}^{n}(X_{ij1} - \bar{X}_{ij})^2]$$

within-cells DF = ab(n – 1),
where: a = number of levels in factor A
b = number of levels in factor B
n = number of replicates.

The pooled variance (s_p^2), which is the best estimate of σ2, is found by

$$\frac{within-cells\ SS}{within-cells\ DF} = \frac{error\ SS}{error\ DF} = \text{MSE}$$

We also need an estimate of the variability among the cells. This is analogous to "groups" in the one-way ANOVA:

cells SS = $\sum_{i=1}^{a}\sum_{j=1}^{b} n(\bar{X}_{ij} - \bar{X})^2$

cells DF = ab – 1.

Finally, we need to know the variability among all the data, N, which is also analogous to that of the one-way ANOVA:

total SS = SS = $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{1=1}^{n}(\bar{X}_{ij} - \bar{X})^2$

DF = N – 1.

In a two-way ANOVA, we typically want to know about the differences between the two factors when considered independently. Thus, we want to examine the specific components of the cells

SS and cells DF. We can examine these differences by only considering one factor at a time in the ANOVA. So, for factor A:

$SS(A) = an\sum_{j=1}^{b}(\bar{X}_{.j} - \bar{X})^2$

DF(B) = b – 1.

However, the sums of squares and degrees of freedom for

factor A and factor B will not exactly sum to the cells SS because of interaction between the factors. The relationship is additive and it can be expressed as:

$SS(AxB) = SS(Cells) - SS(A) - SS(B)$,
$DF(A \times B) = DF(Cells) - DF(A) - DF(B) = (a-1)(b-1)$.

An interaction between the two factors implies that they are not independent of each other. We will test interaction to determine if it is actually significant in an upcoming example.

As with the one-way ANOVA, you can divide the different SS by their corresponding DF to obtain a variance, called a *mean square:*

$$MS(A)=\frac{SS(A)}{DF(A)}$$
$$MS(B)=\frac{SS(B)}{DF(B)}$$
$$MS(AxB)=\frac{SS(AXB)}{DF(AXB)}$$
$$MS(Error)=\frac{SS(Error)}{DF(Error)}$$

Error MS is usually called the mean square error (MSE). As with the one-way ANOVA, these ratios are F-distributed. So, we can calculate corresponding F-statistics and compare to a one tailed critical value from the F-distribution for our hypothesis tests:

Table 3: Two-way Analysis

Source of Variation	Sum of Squares (SS)	Degrees of Freedom (DF)	Mean Square (MS)	F-statistic
Cells	$\sum_{i=1}^{a}\sum_{j=1}^{b} n(\bar{X}_{ij}-\bar{X})^2$	ab-1		$F=\frac{MS(A)}{MSE}$
Factor A	$bn\sum_{i=1}^{a}(\bar{X}_{i}-\bar{X})^2$	a-1	$\frac{SS(A)}{DF(A)}$	$F=\frac{MS(B)}{MSE}$
Factor B	$an\sum_{j=1}^{b}(\bar{X}_{j}-\bar{X})^2$	b-1	$\frac{SS(B)}{DF(B)}$	$F=\frac{MS(AXB)}{MSE}$
A x B	SS(Cells)–SS(A) – SS(B)	(a-1)(b-1)	$\frac{SS(AXB)}{DF(AXB)}$	
Error	$\sum_{i=1}^{a}\sum_{j=1}^{b}[\sum_{i=1}^{n}(X_{ijk}-\bar{X}_{ij})^2]$	ab(n-1)	$\frac{SS(Error)}{DF(Error)}$	
Total	$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{i=1}^{n}(X_{ij}-\bar{X})^2$	N-1		

$F=\frac{MS(A)}{MSE}$ (MS(A))/MSE , for Ho: no effect of factor A on response variable

$F=\frac{MS(B)}{MSE}$(MS(B))/MSE , for Ho: no effect of factor B on response variable

$F=\frac{MS(AXB)}{MSE}$ for Ho: no interaction between factor A and factor B.

We reject any Ho if $F \geq F\_critical$; otherwise, we do not reject H0

At this point, we can now construct our theoretical ANOVA table:

3. RESULT AND DISCUSSIONS

Analysis of Variance is a hypothesis testing procedure that tests whether two or more means are significantly different from each other. A statistic, F, is calculated that measures the size of the effects by comparing a ratio of the differences between the means of the groups to the variability within groups. The larger the value of F, the more likely that there are real effects. The obtained F-ratio is compared to a model of F-ratios that would be found given that there were no effects. If the obtained F-ratio is unlikely given the model of no effects, the hypothesis of no effects is rejected and the hypothesis of real effects is accepted. If the model of no effects could explain the results, then the null hypothesis of no effects must be retained.

Table 4: Result of RCBD

Number of Observations Read	32
Number of Observations Used	32

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	92794.56500	9279.45650	41.26	<.0001
Error	21	4723.20375	224.91446		
Corrected Total	31	97517.76875			

Hypothesis:

Null Hypothesis:

There is no significance difference between blocks yield

Alternative Hypothesis:

There is significance difference between blocks yield

From the above table, the null hypothesis states that the mean blocks milk yield values of 4 different blocks are not equal. Because the p-value is 0.001, which is less than the significance level of 0.05, you can reject the null hypothesis and conclude that some significance difference between blocks cow milk yields.

Table 5: Result of Critical Difference

R-Square	Coeff Var	Root MSE	yield Mean
0.951566	10.49715	14.99715	142.8688

R^2 is the percentage of variation in the response that is explained by the model. R^2 is always between 0% and 100%. The higher the R^2 (0.95) value, the better the model fits for milk yield data. Average mean yield is 142.86 liters.

Table 6: Result of Two-way analysis of Variance

Source	DF	Anova SS	Mean Square	F Value	Pr > F
block	7	56537.87375	8076.83911	35.91	<.0001
tr	3	36256.69125	12085.56375	53.73	<.0001

The above table shows that the "Sig." value (.001) is less than .05 and the null hypothesis must be rejected. If the alpha level had been set at .01, or even .001, the results of the hypothesis test would be statistically significant.

4. CONCLUSIONS

There is a significant difference between the month to month yields also the treatment effects were found to be significant. However, the exact significance level and let the reader set his or her own significance level. It shows that food B is giving better yields compare with others.

REFERENCES

1. Cochrn, W. G and Cox, G.M (1957) Experimental Design John Wiley & Sons, New York.
2. Das, M.N. And Giri, N.S (1986) Design and Analysis Of Experiments (2nd Edition).Wiley Eastern Ltd., New Delhi.
3. Dey, A (2010) Incomplete Block Design. World Scientific Publishing Company.
4. Fisher, R.A (1953) Design and Analysis of Experiments. Oliver and Boyd, London.
5. Hollander, M. and D.A. Wolfe (1999). Nonparametric Statistical Methods. (2ndedn), John Wiley and Sons, New York, USA.
6. Kruskal, W.H and W.A. Wallis. (1953). Use of Ranks in One-Criterion Variance Analysis, J Am Stat Assoc, 47, 583-621.
7. Linn, C. E., Campbell, M. G., Poole, K. R., Wu, W.-Q. and W.L. Roelofs. (1996). Effect of photoperiod on the circadian timing of pheromone response in male Trichoplusia ni: relationship to the modulatory action of octopamine, J. Insect. Physiol, 42, 881-891.
8. Mafra-Neto, A. and R. T. Cardé, (1998). Rate of realized interception of pheromone pulses in different wind speeds modulates almond moth orientation, J. Comp. Physiol. A, 182, 563-572.
9. Mann, H.B and D.R. Whitney (1947). On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other, Annals of Mathematical statistics, 18, 50-60.
10. Montgomery, D.C (2012) Design and Analysis of Experiments John Wiley & Sons, New York.
11. Searle, S. R. (1971). Linear Models. Wiley, New York.
12. Sokal, R. R. and F.J. Rohlf. (1981). Biometry: The Principles and Practice of Statistics in Biological Research, Second edition. New York: W. H. Freeman and Company.