A Project Report

on

Completely Randomized Block Design for Evaluating the Impact of Parental Satisfaction, Absenteeism, and Subject Choice on Student Performance with Gender as a Blocking Factor

College of Computing and Mathematic

جامعة الملك فهد للبترول والمعادن
King Fahd University of Petroleum & Minerals

Submitted By

Abdibasid

Abd-ur-Rahman

Submitted To

Dr. Jimoh Ajadi

College of Computing and Mathematic

Department of Mathematics, King Fahd University of Petrolium & Minerals, Dhahran

DECEMBER – 2023

INTRODUCTION

A Randomized Block Design is characterized by the division of subjects into subgroups, known as blocks, with the goal of minimizing variability within blocks compared to between blocks. In this design, one specific factor is of primary interest, while others are considered as nuisance factors. According to the Concise Encyclopedia of Statistics, treatments are randomly assigned to experimental units within each block, distinguishing between completely randomized block designs and incomplete randomized block designs based on whether all treatments appear in each block.

Gary (2010) describes the Randomized Complete Block Design (RCBD) as a design that involves the use of every treatment in every block, akin to a two-way ANOVA without interaction. On the other hand, incomplete block designs involve the omission of certain treatments in some blocks, aiming to mitigate systematic errors.

Randomized Block Designs, as likened to stratified sampling by Prof. William M.K. Trochim, involve the grouping of samples into homogeneous blocks or strata to reduce variance or nuisance effects. This design allows for more efficient estimates of treatment effects within blocks, contributing to an overall more efficient estimate when pooled across blocks.

In the context of educational research, the project focuses on a Completely Randomized Block Design to assess the impact of Parental Satisfaction, Absenteeism, and Subject Choice on Student Performance, with Gender serving as a blocking factor. The study particularly explores the correlation between attendance (categorized as 'Under-7' and 'Above-7' days) and academic outcomes, acknowledging the critical role of absenteeism in shaping student learning and performance.

The extensive dataset covers various subjects and includes records of parental satisfaction levels. By examining these variables alongside absenteeism data, the research aims to uncover patterns that could inform educational strategies. The RCBD, known for its efficiency in controlling for variability, is chosen for its suitability in this study.

In summary, the study adopts a Randomized Complete Block Design to investigate the multifaceted influences of Parental Satisfaction, Absenteeism, and Subject Choice on Student Performance, with gender as a key blocking factor. The chosen design allows for a comprehensive analysis of the complex interplay of variables in an educational setting.

Advantages of RCBD

Increased Accuracy: Grouping more homogeneous experimental material typically yields more accurate results compared to a completely randomized design.

Flexibility in Treatment and Replicates: RCBD allows for any number of treatments and replicates. Each treatment has the same number of replicates, ensuring uniformity across the experiment.

Straightforward Statistical Analysis: The design simplifies the statistical analysis process, making it more accessible for researchers.

Unbiased Error Estimation: In cases where variance differs across treatments, RCBD provides an unbiased error estimate for testing specific combinations of treatment means.

Disadvantages of RCBD

Handling Missing Values: The primary challenge arises with missing values, if an entire block or all data on a treatment are missing, it doesn't pose a significant issue. However, problems occur when individual units within a block are missing. The "missing-plot" technique is often used to maximize the use of available data.

Potential Loss of Precision: If there are no real differences among the blocks, blocking can reduce precision in estimating error variability. This is because blocking consumes degrees of freedom that would otherwise be used in estimating error variation.

The RCBD Model

The model for RCBD is defined as:

$Yij=\mu+\tau i+\beta j+\epsilon ij$

Where:

Yijrepresents the response for treatment i observed in block j.

μ is a constant, representing the overall mean.

τi is the additive effect of the *h*ith treatment (=1,2,…,i=1,2,…,k).

βj is the additive effect of the *h*jth block (=1,2,…,j=1,2,…,n).

ϵij is the random error for the *h*ith treatment in the *h*jth block.

Application in Research

In applying RCBD to research, it's crucial to:

- Identify Homogeneous Blocks: Determine the blocking factor that can best group the experimental units into homogeneous sets.
- Randomize Treatments Within Blocks: Ensure that each treatment is randomly assigned within each block to prevent systematic bias.
- Analyse Data Appropriately: Use statistical methods suited for RCBD, like ANOVA, to analyse the data, taking into account the block effects.

Aim and Objectives of the Study

Aim of the Research

The aim of this research is to examine the impact of parental satisfaction, absenteeism, and subject choice on student performance, utilizing a Completely Randomized Block Design with gender as a blocking factor. This study aims to understand how these variables individually and collectively influence academic outcomes in different gender groups.

Objective:

To assess the significance of parental satisfaction levels on the academic performance of students, this involves exploring how different levels of parental satisfaction correlate with student achievements in school.

Question:-

- ✓ What is the individual impact of parental satisfaction, absenteeism, and subject choice on student academic performance?
- ✓ How do these factors interact with each other, and what is the combined influence on student achievement?
- ✓ To what extent does gender act as a moderating factor in the relationship between the identified variables and academic performance?

Data Set

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	gender	NationalIT	PlaceofBir	StageID	GradeID	SectionID	Topic	Semester	Relation	raisedhan	VisITedRe	Announce	Discussior	ParentAns	Parentsch	StudentAb	Class
2	M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	15	16	2	20	Yes	Good	Under-7	M
3	M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	20	20	3	25	Yes	Good	Under-7	M
4	M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	10	7	0	30	No	Bad	Above-7	L
5	M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	30	25	5	35	No	Bad	Above-7	L
6	M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	40	50	12	50	No	Bad	Above-7	M
7	F	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	42	30	13	70	Yes	Bad	Above-7	M
8	M	KW	KuwaIT	MiddleSch	G-07	A	Math	F	Father	35	12	0	17	No	Bad	Above-7	L
9	M	KW	KuwaIT	MiddleSch	G-07	A	Math	F	Father	50	10	15	22	Yes	Good	Under-7	M
10	F	KW	KuwaIT	MiddleSch	G-07	A	Math	F	Father	12	21	16	50	Yes	Good	Under-7	M
11	F	KW	KuwaIT	MiddleSch	G-07	B	IT	F	Father	70	80	25	70	Yes	Good	Under-7	M
12	M	KW	KuwaIT	MiddleSch	G-07	A	Math	F	Father	50	88	30	80	Yes	Good	Under-7	H
13	M	KW	KuwaIT	MiddleSch	G-07	B	Math	F	Father	19	6	19	12	Yes	Good	Under-7	M
14	M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	5	1	0	11	No	Bad	Above-7	L
15	M	lebanon	lebanon	MiddleSch	G-08	A	Math	F	Father	20	14	12	19	No	Bad	Above-7	L
16	F	KW	KuwaIT	MiddleSch	G-08	A	Math	F	Mum	62	70	44	60	No	Bad	Above-7	H
17	F	KW	KuwaIT	MiddleSch	G-06	A	IT	F	Father	30	40	22	66	Yes	Good	Under-7	M

Length: 480

Area: E-learning, Education, Predictive models, Educational Data Mining

Attribute Characteristics: Integer/Categorical

Number of Attributes: 16

Date: 2016-11-8

Associated Tasks: Classification

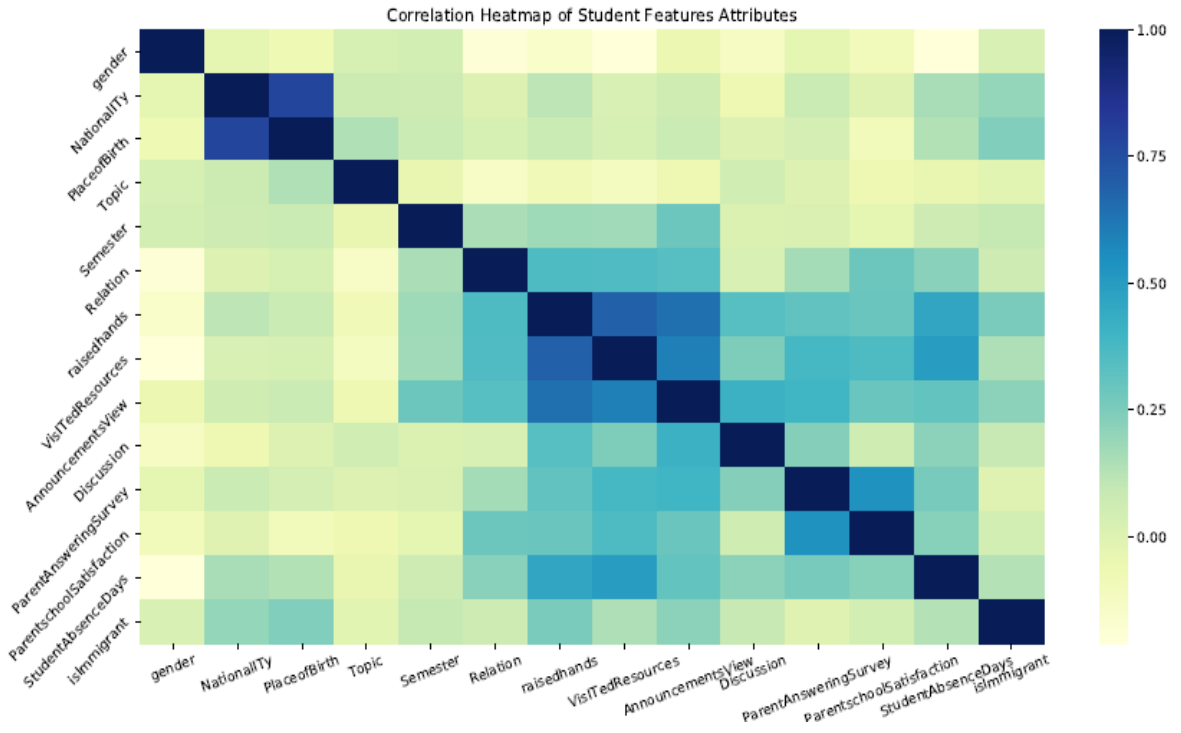
Missing Values? No

File formats: xAPI-Edu-Data.csv

Source:

Elaf Abu Amrieh, Thair Hamtini, and Ibrahim Aljarah, The University of Jordan, Amman, Jordan, http://www.Ibrahimaljarah.com
www.ju.edu.jo

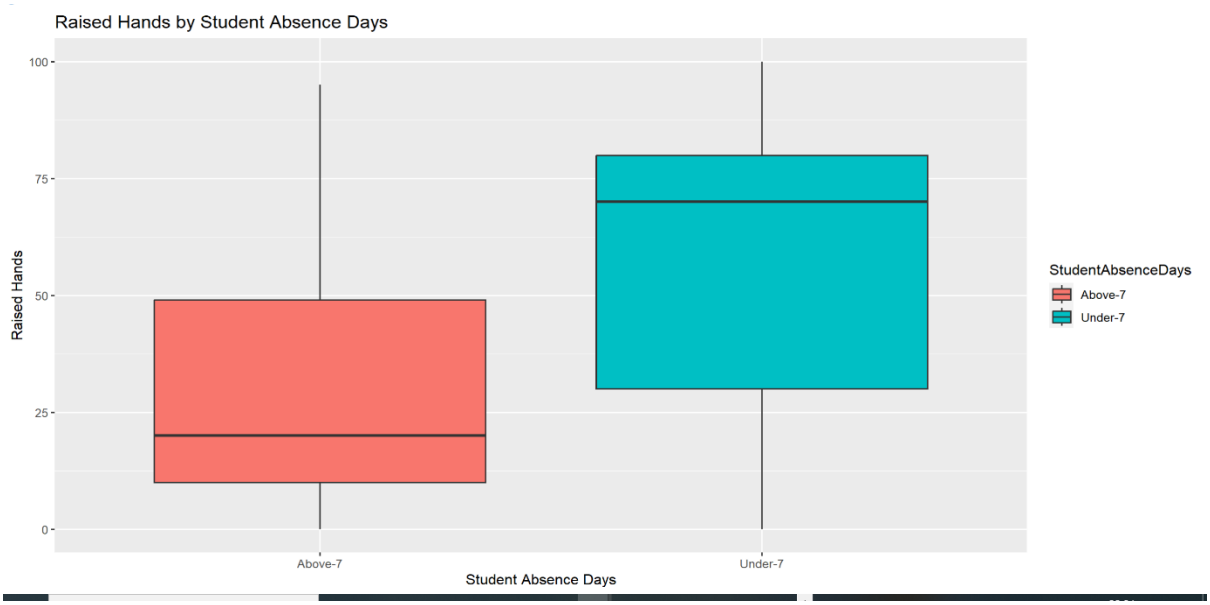
Correlation b/w Predictors



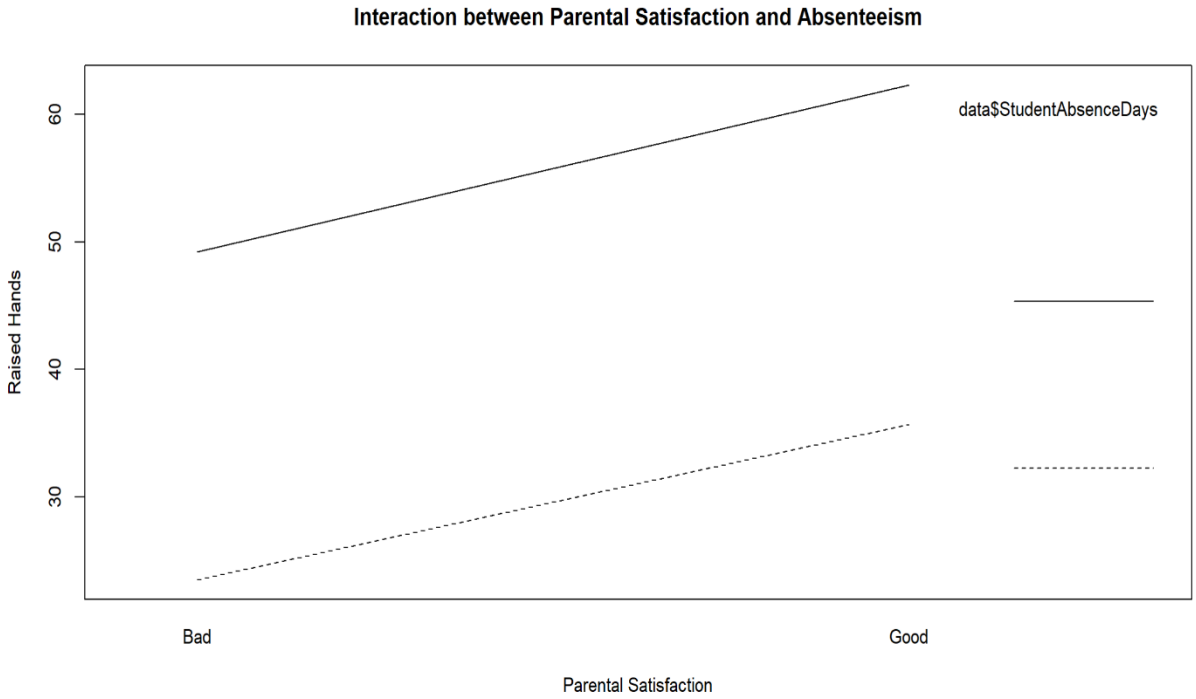
The heat map visually shows the degree of correlation between student attributes. It is not difficult to find that the most popular of these attributes include 'Relation, raised hands, Visited Resources, Announcements View, Discussion, Parent Answer Survey, Parents school Satisfaction and Student Absence Days.

Results

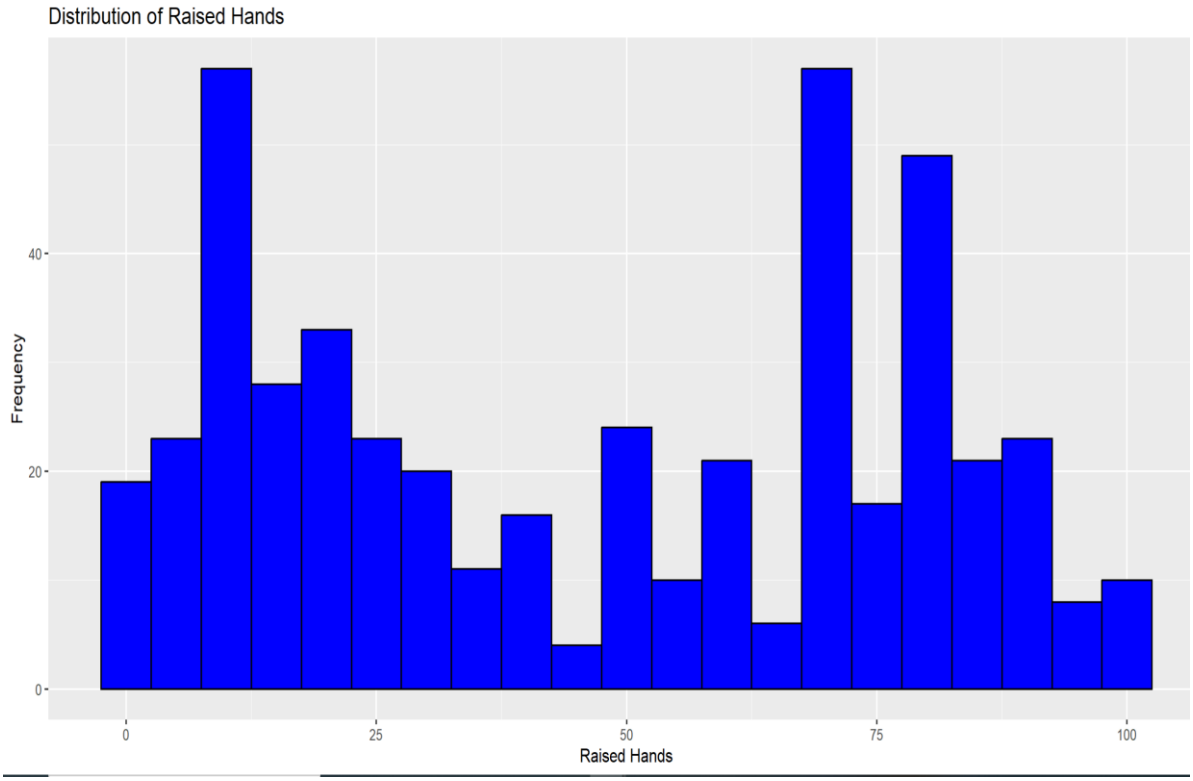
```
> summary(result)
                         Df Sum Sq Mean Sq F value   Pr(>F)
ParentschoolSatisfaction  1  40032   40032  64.879 6.75e-15 ***
StudentAbsenceDays        1  75095   75095 121.706  < 2e-16 ***
Topic                    11  51359    4669   7.567 4.90e-12 ***
gender                    1    386     386   0.625     0.43
Residuals               465 286914     617
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```



This is the graph between the raised hand and the absence of the student in the class there are most of student are present in the class during the whole semester and the small portion of the student are not present. we use the two level of absence one is grater than 7 and other one is less than. We see from the graph that there is almost 50 student which are absent more than 7 days.in the semester this graph tells the presence of the student in the whole semester



This is the interaction graph between parent satisfaction and the absent during the semester we that there is no relation between these two values so we can say that they are independent from each other they are no intersect to each other.



This is the distribution of raised hand by the student in the class

Result Interpretation

1. The variable "ParentschoolSatisfaction" shows a significant effect on student performance (F value = 64.879, p-value = 6.75e-15). This implies that there is a strong association between parental satisfaction with the school and the academic performance of students.

2. The variable "StudentAbsenceDays" is highly significant (F value = 121.706, p-value < 2e-16). This suggests a strong relationship between student absenteeism and academic performance. The result implies that the number of days a student is absent from school has a considerable impact on their overall performance.

3. The variable "Topic" shows a significant effect on student performance (F value = 7.567, p-value = 4.90e-12). This suggests that the academic performance of students varies across different topics. Some topics may positively or negatively influence student outcomes, indicating the importance of the curriculum in understanding and predicting performance.

4. The variable "gender" does not exhibit a significant effect on student performance (F value = 0.625, p-value = 0.43). This suggests that, in this analysis, there is insufficient evidence to conclude that gender has a notable impact on academic performance. Other factors may play a more crucial role in determining student outcomes.

Conclusion

In summary, parental satisfaction, student absenteeism, and the topics covered in the curriculum significantly influence student performance. On the other hand, gender does not appear to be a significant factor in this analysis. The unexplained variance in the residuals indicates the presence of other factors affecting student performance.

The model we propose facilitates the prediction of student academic performance, offering valuable technical assistance for teaching management tasks, such as tailoring instruction based on individual aptitude early in the course and issuing academic early warnings. This contributes a theoretical foundation and technical backing for the effective management of students in higher education institutions.

Future Aspects

Extend the analysis over multiple time points to observe trends and changes in student performance, attendance, and class participation over time.

Develop predictive models to forecast future student performance based on historical data. This could involve machine learning algorithms to identify patterns and make predictions

References

Namoun, A.; Alshanqiti, A. Predicting student performance using data mining and learning analytics techniques: A systematic literature review. Appl. Sci. **2020**, 11, 237. [CrossRef]

Ang, K.L.M.; Ge, F.L.; Seng, K.P. Big educational data & analytics: Survey, architecture and challenges. IEEE Access **2020**,8, 116392–116414.

Zohair, A. and Mahmoud, L., 2019. Prediction of Student's performance by modelling small dataset size. International Journalof Educational Technology in Higher Education, 16(1), pp.18

Shah, M.B., Kaistha, M. and Gupta, Y., 2019, November. Student performance assessment and prediction system using machine learning. In 2019 4th International Conference on Information Systems and Computer Networks (ISCON) (pp. 386-390). IEEE.

H. M.R. Hasan, A. S.A. Rabby, M.T. Islamand S. A. Hossain, "Machine Learning Algorithm for Student's Performance Prediction,"2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 2019, pp. 1-7,doi:10.1109/ICCCNT45670.2019.8944629