ggDoE: An R Package for Experimental Design with an application to Prostate Cancer Drugs

Casey Truong Orange Coast College

Advisors: Dr. Jessica Jaynes and Jose Toledo Luna California State University, Fullerton Department of Mathematics

Introduction

Experimental design is a critical part of the scientific process in which we learn how systems or processes work to help improve process yield, reduce variability, decrease development time and overall costs. One important type of experimental design is a two-level full factorial design, which explores all possible combinations of various factors at two levels.

The statistical software R is widely used to assist experimenters in selecting and constructing factorial designs, as well as implementing and analyzing them. A newly developed R package called ggDoE (Design of Experiments) is a visualization software for Experimental Design, which updates existing R functions for the illustration of factorial designs.

Objectives

This research focuses on how to implement the ggDoE package with an application pertaining to Prostate Cancer and five anti-cancer drugs by:

- Implementing functions within the newly developed R package: ggDoE
- Creating an R Vignette with RMarkdown for the ggDoE package based on the prostate cancer dataset

Methods: ggDoE Functions

An R Vignette is a demonstration manual for users on how to utilize functions from R Packages. The ggDoE package has several functions available for the implementation and visualization of factorial designs. The R Vignette being developed focuses on the following functions:

- *main_effect*: Obtains the main effect plot for a single factor from a factorial design.
- *interaction_effect*: Obtains the interaction effect plot between two factors from a factorial design.
- *diagnostic_plots*: Obtains linear regression diagnostic plots with ggplot2.
- *halfnormal*: Obtains a Half-normal plot.
- *pareto_plot*: Obtains a Pareto plot of the factorial effects.

Further information about ggDoE can be found here: ggDoE Documentation

Full Factorial Design Application: Prostate Cancer Dataset

In the study performed by Jia et al. (2017), a two-level full factorial design is considered to investigate the combinations of five anti-cancer drugs to minimize the survival of prostate cancer cells. The five anti-cancer drugs and their corresponding low (0) and high (1) levels are presented in the table below. For simplicity purposes we call the five drugs: $D1, D2, D3, D4, D5$.

The table below presents the full factorial deigns for the five anti-cancer drugs and their combinations $(2^5 = 32)$ and response survival (the percentage of prostate cancer cells):

	D1	D2	D3	D4	D5	Survival
1	1	1	1	1	1	0.42
2	1	1	1	1	0	0.41
3	1	1	1	0	1	0.41
4	1	1	1	0	0	0.40
5	1	1	0	1	1	0.40
6	1	1	0	1	0	0.47
7	1	1	0	0	1	0.43
8	1	1	0	0	0	0.50
9	1	0	1	1	1	0.53
10	1	0	1	1	0	0.53
11	1	0	1	0	1	0.53
12	1	0	1	0	0	0.56
13	1	0	0	1	1	0.44
14	1	0	0	1	0	0.51
15	1	0	0	0	1	0.57
16	1	0	0	0	0	0.67

	D1	D2	D3	D4	D5	Survival
17	0	1	1	1	1	0.68
18	0	1	1	1	0	0.69
19	0	1	1	0	1	0.69
20	0	1	1	0	0	0.36
21	0	1	0	1	1	0.50
22	0	1	0	1	0	0.48
23	0	1	0	0	1	0.46
24	0	1	0	0	0	0.50
25	0	0	1	1	1	0.68
26	0	0	1	1	0	0.86
27	0	0	1	0	1	0.77
28	0	0	1	0	0	0.82
29	0	0	0	1	1	0.94
30	0	0	0	1	0	0.87
31	0	0	0	0	1	0.91
32	0	0	0	0	0	0.99

Results: Factorial Effects

To determine which factors are statistically significant on the percent of cancer survival cells factorial effects can be calculated.

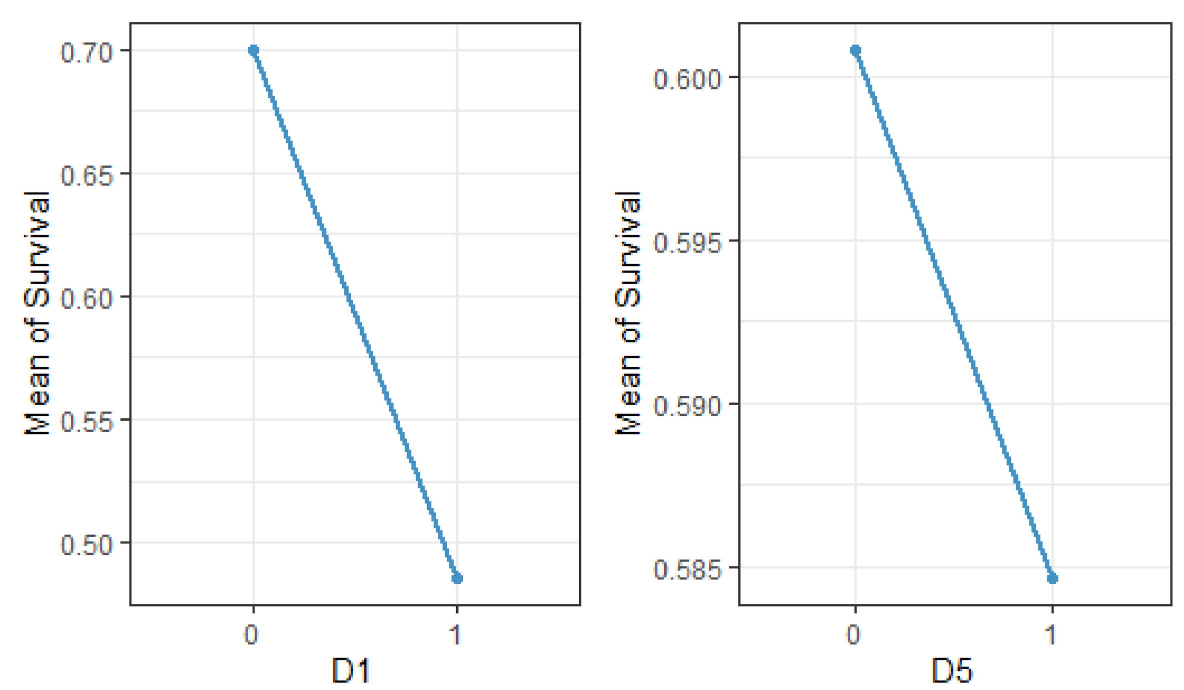


Figure 1. A main effect plot graphs the averages of all the observations at each level of the factor and connects them by a line.

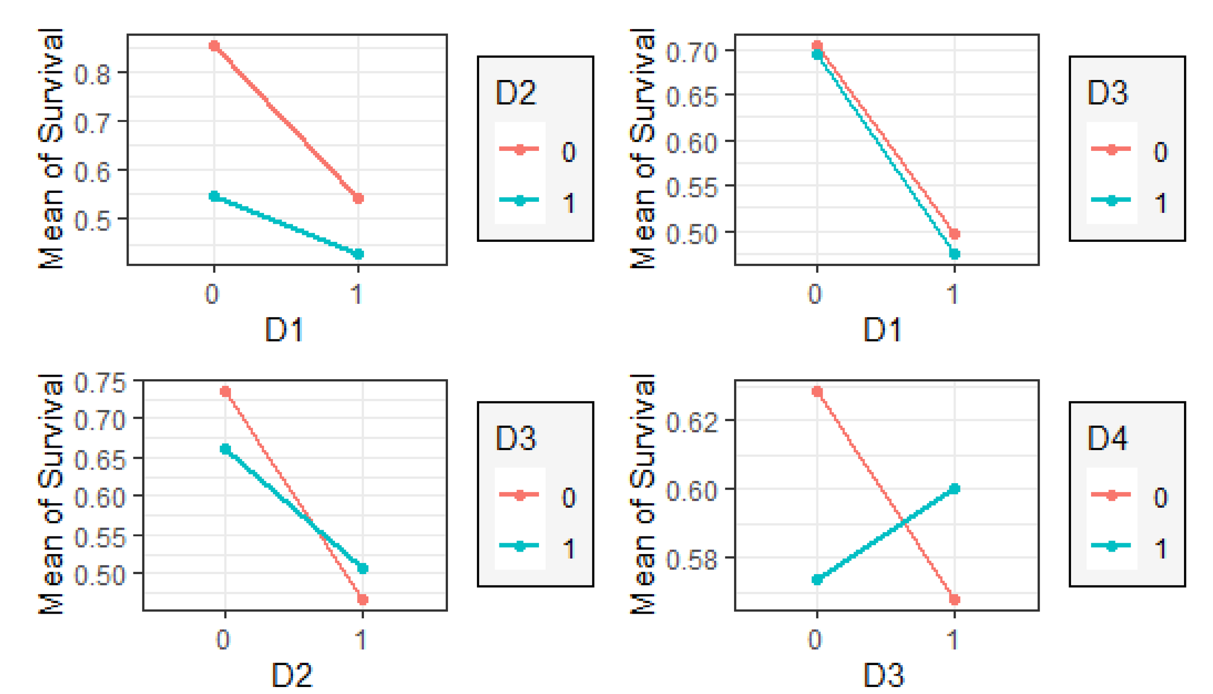
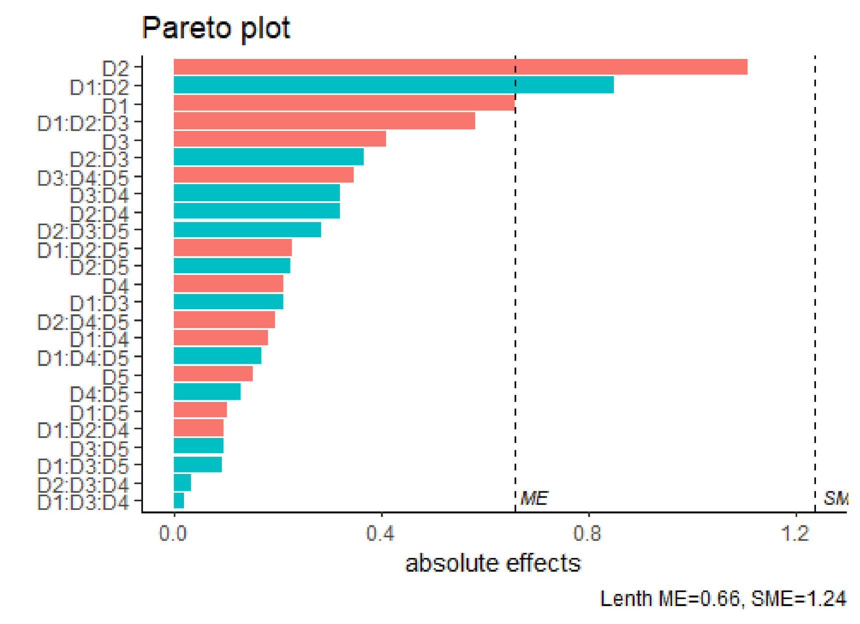


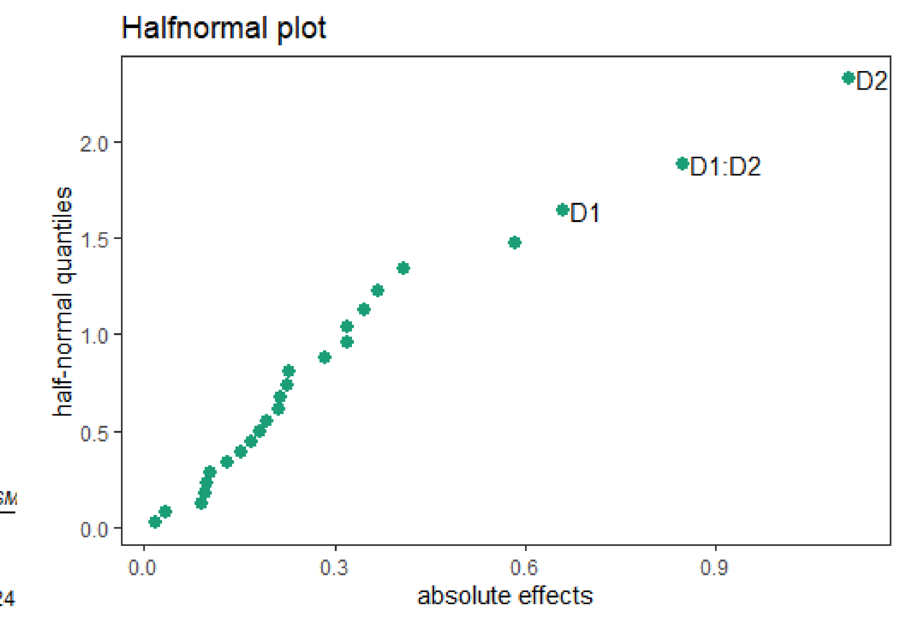
Figure 2. A interaction effect plot examines the relationship between two factors and the response.

Results

Pereto plot using Lenth method: Lenth's method matches what we concluded from the half-normal plot above.

Half-normal Plot: A half-normal plot can be used to detect effect significance. A half-normal plot graphs the absolute value of the factorial effects against half-normal quantiles.





Model fit: We fit a linear model with 5 main effects, 10 two-factor interactions, and 10 three-factor interactions.

	Estimate	Std. Error	t value	Pr(>|t|)		Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.00	0.07	14.98	0.00	D3:D4	0.16	0.10	1.53	0.18
D1	-0.33	0.087	-3.79	0.01	D3:D5	0.05	0.10	0.46	0.66
D2	-0.56	0.088	-6.37	0.00	D4:D5	0.07	0.10	0.62	0.56
D3	-0.20	0.09	-2.35	0.06	D1:D2:D3	-0.29	0.10	-2.78	0.03
D4	-0.11	0.09	-1.23	0.27	D1:D2:D4	-0.05	0.10	-0.47	0.66
D5	-0.08	0.09	-0.87	0.42	D1:D2:D5	-0.12	0.10	-1.09	0.32
D1:D2	0.42	0.10	4.04	0.01	D1:D3:D4	0.01	0.10	0.09	0.93
D1:D3	0.11	0.10	1.01	0.35	D1:D3:D5	0.05	0.10	0.44	0.68
D1:D4	-0.09	0.10	-0.86	0.42	D1:D4:D5	0.08	0.10	0.81	0.45
D1:D5	-0.05	0.10	-0.49	0.64	D2:D3:D4	0.02	0.10	0.16	0.88
D2:D3	0.18	0.10	1.75	0.13	D2:D3:D5	0.14	0.10	1.36	0.22
D2:D4	0.16	0.10	1.52	0.18	D2:D4:D5	-0.1	0.10	-0.93	0.39
D2:D5	0.11	0.10	1.08	0.32	D3:D4:D5	-0.17	0.10	-1.65	0.15

Conclusion and Future Work

Based on the results from the Pareto plot and halfnormal plot, the main effect of factor $D2$ and the two-factor interaction $D1 : D2$ appear to be active effects in minimizing the prostate cancer cells survival. Whereas in the model fit, three main effects: $D1, D2, D3,$ and interactions between $D1 : D2$ and $D1 : D2 : D3$ are significant. The results between these two methods differ as they are based on different criteria for determining significance. Overall, the ggDoE package is an incredibly important and easy tool to use for factorial design related research. The application to the prostate cancer dataset illustrates the usefulness of the ggDoE package in generating output for factorial designs. The ggDoE package is actively being developed.

References

[1] Xiaolong Jia, Yiyang Li, Alok Sharma, Yulong Li, Guohai Xie, Guoyao Wang, Junhui Jiang, Yue Cheng, and Xianting Ding. Application of sequential factorial design and orthogonal array composite design (oacd) to study combination of 5 prostate cancer drugs. *Computational biology and chemistry*, 67:234–243, 2017.

[2] Douglas C Montgomery. *Design and analysis of experiments*. John wiley & sons, 2017.

[3] CF Jeff Wu and Michael S Hamada. *Experiments: planning, analysis, and optimization*, volume 552. John Wiley & Sons, 2011.