Factorial Experiment Design on Key Factors Affecting BMI

Zichen Gong

1005682469

Materials and Methods

Experimental Design and Data

Factor	Column Name
A	Kidney Disease (Yes/No)
B	Smoking (Yes/No)
C	Alcohol Drinking (Yes/No)
D	Physical Activity (Yes/No)
E	Sleep Time (Numerical)

$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 1$

$A^{\alpha_1} B^{\alpha_2} C^{\alpha_3} D^{\alpha_4} E^{\alpha_5}$=ABCDE

$L = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 + \alpha_5 x_5$
$= x_1 + x_2 + x_3 + x_4 + x_5$

High Level: $x_i = 1$
Low Level: $x_i = 0$

Block 1: $L\ mod(2) = 0$
Block 2: $L\ mod(2) = 1$

2^5 factorial experiment design with highest order confounding and 2 blocks

	A	B	C	D	E	BMI	block
1	-1	-1	-1	-1	-1	37.23	1
2	-1	-1	-1	-1	1	32.92	2
3	-1	-1	-1	1	-1	36.58	2
4	-1	-1	-1	1	1	34.46	1
5	-1	-1	1	-1	-1	29.29	2
6	-1	-1	1	-1	1	28.19	1
7	-1	-1	1	1	-1	34.95	1
8	-1	-1	1	1	1	24.85	2
9	-1	1	-1	-1	-1	25.06	2
10	-1	1	-1	-1	1	23.89	1
11	-1	1	-1	1	-1	23.4	1
12	-1	1	-1	1	1	21.79	2
13	-1	1	1	-1	-1	35.43	1
14	-1	1	1	-1	1	35.26	2
15	-1	1	1	1	-1	28.13	2
16	-1	1	1	1	1	27.12	1
17	1	-1	-1	-1	-1	31.64	2
18	1	-1	-1	-1	1	33.91	1
19	1	-1	-1	1	-1	30.13	1
20	1	-1	-1	1	1	20.25	2
21	1	-1	1	-1	-1	48.42	1
22	1	-1	1	-1	1	38.01	2
23	1	-1	1	1	-1	26.16	2
24	1	-1	1	1	1	27.71	1
25	1	1	-1	-1	-1	31.09	1
26	1	1	-1	-1	1	31.31	2
27	1	1	-1	1	-1	23.69	2
28	1	1	-1	1	1	33.45	1
29	1	1	1	-1	-1	31.57	2
30	1	1	1	-1	1	19.67	1
31	1	1	1	1	-1	49.57	1
32	1	1	1	1	1	29.57	2

Table 1: Table 1: Effect Estimates for the Blocked 2^5 Design

	Regression Coefficient	Effect Estimate	Sum of Squares	Percent Contribution
(Intercept)	30.77	61.54	23.81	1.60
A1	0.86	1.72	62.44	4.21
B1	-1.40	-2.79	58.05	3.91
C1	1.35	2.69	52.74	3.55
D1	-1.28	-2.57	112.43	7.57
E1	-1.87	-3.75	32.16	2.17
A1:B1	1.00	2.01	23.32	1.57
A1:C1	0.85	1.71	55.60	3.74
B1:C1	1.32	2.64	2.59	0.17
A1:D1	-0.28	-0.57	71.88	4.84
B1:D1	1.50	3.00	0.95	0.06
C1:D1	0.17	0.35	8.82	0.59
A1:E1	-0.53	-1.05	2.11	0.14
B1:E1	0.26	0.51	66.99	4.51
C1:E1	-1.45	-2.89	1.46	0.10
D1:E1	-0.21	-0.43	149.82	10.09
A1:B1:C1	-2.16	-4.33	269.00	18.12
A1:B1:D1	2.90	5.80	21.16	1.42
A1:C1:D1	0.81	1.63	43.80	2.95
B1:C1:D1	1.17	2.34	11.42	0.77
A1:B1:E1	-0.60	-1.19	49.90	3.36
A1:C1:E1	-1.25	-2.50	36.68	2.47
B1:C1:E1	-1.07	-2.14	2.73	0.18
A1:D1:E1	0.29	0.58	1.60	0.11
B1:D1:E1	0.22	0.45	0.82	0.06
C1:D1:E1	-0.16	-0.32	126.64	8.53
A1:B1:C1:D1	1.99	3.98	69.03	4.65
A1:B1:C1:E1	-1.47	-2.94	0.48	0.03
A1:B1:D1:E1	-0.12	-0.24	10.19	0.69
A1:C1:D1:E1	0.56	1.13	29.95	2.02
B1:C1:D1:E1	-0.97	-1.93	86.26	5.81
A1:B1:C1:D1:E1	-1.64	-3.28	0.00	0.00

Statistical Analysis

$$\overline{BMI}_{block\ 1} - \overline{BMI}_{block\ 2}$$
$$= 32.41 - 29.13$$
$$= 3.28$$

Results and Discussion

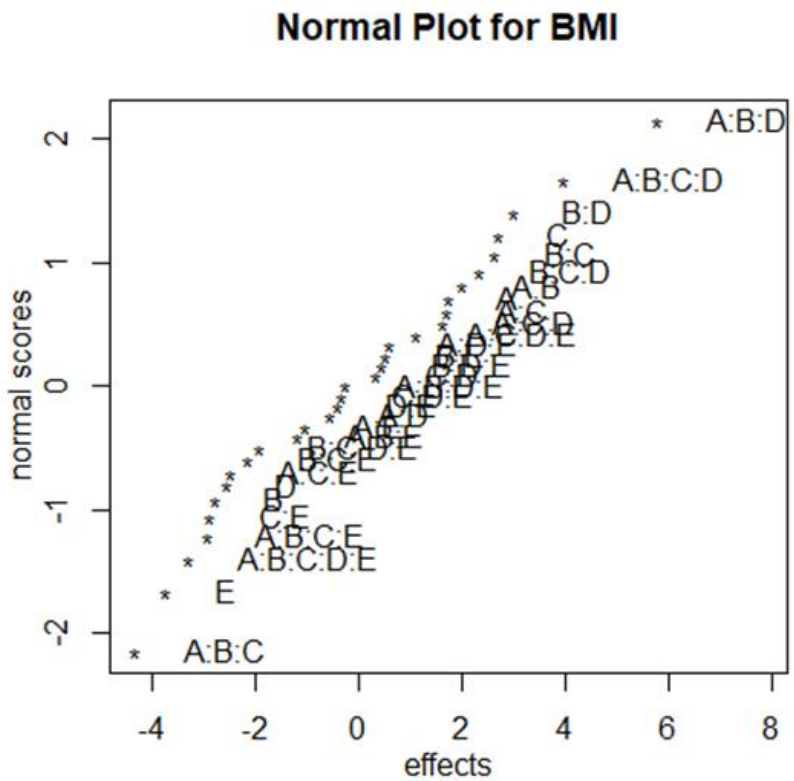
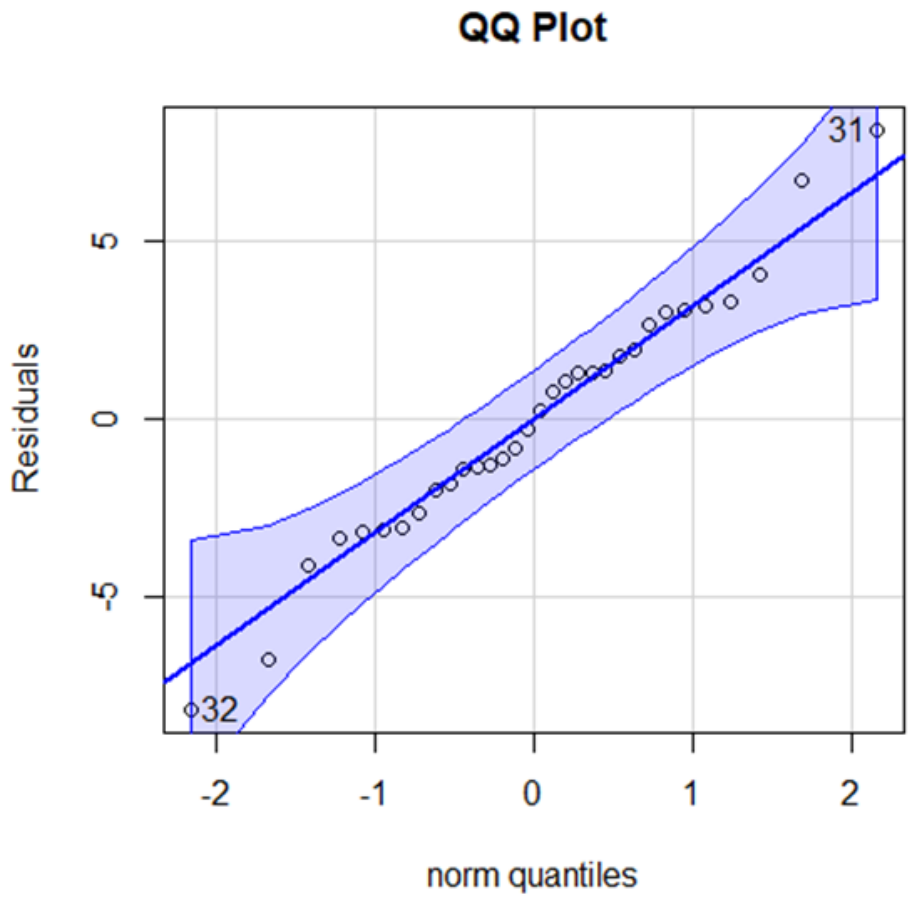
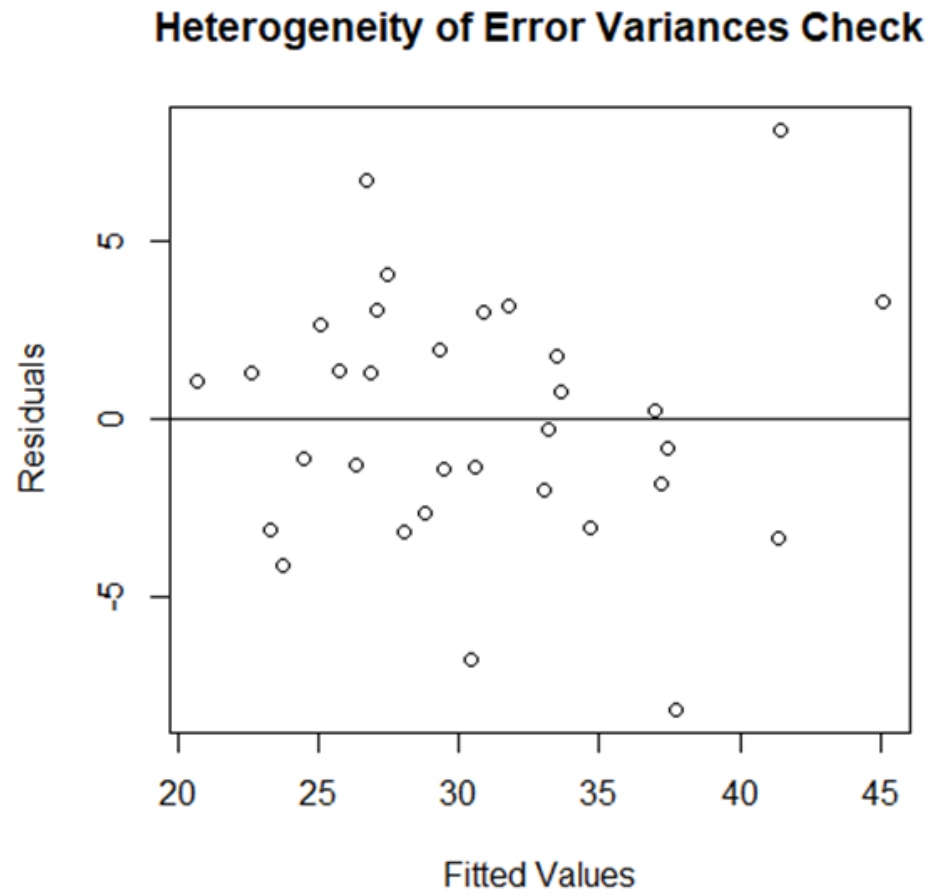


Table 2: Table 2: Analysis of Variance

	Degree of Freedom	Sum of Squares	Mean Square	F Value	P Value
E	1	112.425012	112.425012	4.4559795	0.0519755
A	1	23.805000	23.805000	0.9435142	0.3467795
B	1	62.440312	62.440312	2.4748296	0.1365338
C	1	58.050312	58.050312	2.3008314	0.1500949
D	1	52.736450	52.736450	2.0902158	0.1688145
A:B	1	32.160200	32.160200	1.2746736	0.2766250
A:C	1	23.324450	23.324450	0.9244675	0.3515490
B:C	1	55.598512	55.598512	2.2036540	0.1583934
A:D	1	2.587813	2.587813	0.1025683	0.7531870
B:D	1	71.880050	71.880050	2.8489748	0.1121087
C:D	1	0.952200	0.952200	0.0377406	0.8485715
A:B:C	1	149.818050	149.818050	5.9380572	0.0277506
A:B:D	1	269.004012	269.004012	10.6620078	0.0052174
A:C:D	1	21.157512	21.157512	0.8385807	0.3742907
B:C:D	1	43.804800	43.804800	1.7362088	0.2073914
A:B:C:D	1	126.643613	126.643613	5.0195355	0.0406275
Residuals	15	378.452187	25.230146	NA	NA


QQ Plot
Residuals
5
0
-5
31
32
-2
-1
0
1
2
norm quantiles


Heterogeneity of Error Variances Check
Residuals
5
0
-5
20
25
30
35
40
45
Fitted Values

References

- Are BMI charts different for men & women? (2019). Retrieved April 2, 2022, from https://gasparinutrition.com/blogs/fitness-facts/are-bmi-charts-different-for-men-women#:~:text=Recent%20studies%20have%20found%20that,21%20for%20women%20is%20healthy.
- Hadley Wickham, Romain Fran鏃is, Lionel Henry and Kirill Müller (2022). dplyr: A Grammar of Data Manipulation. R package version 1.0.8. https://CRAN.R-project.org/package=dplyr (Last Accessed: April 3, 2022)
- How Much Sleep Do I Need? (2017). Retrieved April 2, 2022, from https://www.cdc.gov/sleep/about_sleep/how_much_sleep.html
- John Fox and Sanford Weisberg (2019). An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/ (Last Accessed: April 3, 2022)
- Kaggle. (2022). *Personal Key Indicators of Heart Disease* [Data file]. Retrieved April 2, 2022, from https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/ (Last Accessed: April 3, 2022)
- Ulrike Gr”omping (2014). R Package FrF2 for Creating and Analyzing Fractional Factorial 2-Level Designs. Journal of Statistical Software, 56(1), 1-56. URL http://www.jstatsoft.org/v56/i01/ (Last Accessed: April 3, 2022)
- Yihui Xie (2021). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.37. (Last Accessed: April 3, 2022)