STA 305
Project Report

1007249951
Jiawei Gong

1. Introduction

Sleep is crucial for both physical and mental health[1]. To measure the quality of sleep, sleep duration is one of the most significant indicator. Study shows there are different aspects affecting the sleep duration, such as lifestyle, mental stress, and sleeping environment[2]. For this factorial design, we focus on the lifestyle. I expect the lifestyle factors that can affect the sleep duration are physical activity, drinking alcohol and meal before sleep. Therefore, the objective of this factorial design is to analyze the effect of these three factors on the sleep duration measured in hours, so that we can gain insights of methods of increasing sleeping duration to keep health.

2. Materials and Methods

(1) Factors and levels

The factorial experiment is capable to study the effect of these three factors and their interactions on the sleep duration as the response. To conduct factorial design, the levels of factors need to be specified. The first factor 'physical activity' is 1 if this person have physical activity such as running during the daytime, and -1 if not. Second factor called 'Alcohol' is 1 if this person drink alcohol at dinner, and -1 if not. Third factor 'Meal' is 1 if this person have meal before sleep, and -1 if not. At last, the numerical response is the sleep duration measured in hours.

Table 1: Response and levels of factors

Factors & Interactions	Level 1	Level 2
Physical activity (X1)	-1 (No physical activity)	1 (Have physical activity)
Alcohol (X2)	-1 (No alcohol)	1 (Drink alcohol)
Meal (X3)	-1 (No meal)	1 (Have meal)
Response		
Sleep hours (Y)	Numerical	

(2) Data collection and experiment design

There are 2^3=8 combinations of different levels of three factors in total. In order to compute the variance of effect, the treatment of each of 8 combinations is performed twice, which means there are 16 observations in total. The specific data collection process is to firstly write repeated 8 combinations on the 16 lots. To ensure randomness, the sleep hours of 16 subjects are collected by asking 16 random person to draw lots. Each person is asked to follow the specific combination on their lot and record their sleep duration at that night. For example, two person may get the lot that requires them to have physical activity in the day, drink alcohol at dinner, and have meals before sleep. In order to minimize the effect of unknown variables on the final analysis, these 16 persons have similar age and same sex.

By factorial design, the individual effects of three lifestyle factors and their interactions can be analyzed simultaneously, which helps to increase the efficiency and to better understand the complex relationships that might not be apparent in simpler experiments. This aligns the objective of analyzing how three lifestyle factors affect the sleep duration.

(3) Statistical Analysis

Firstly, a linear regression model for this 2^3 factorial design is fitted in the form of:

$$Y_i = b_0 + b_1X_{i1} + b_2X_{i2} + b_3X_{i3} + b_4X_{i1}X_{i2} + b_5X_{i1}X_{i3} + b_6X_{i2}X_{i3} + b_7X_{i1}X_{i2}X_{i3} + \varepsilon_i$$

- Y_i represents the i^{th} response of observation of sleep hours
- ε_i represents the error of i^{th} observation
- X_{i1}, X_{i2}, X_{i3} are three binary factors whose value are -1 or 1, which represent the 'Physical activity', 'Alcohol', and 'Meal' for i^{th} observation

- $X_{i1}X_{i2}$, $X_{i1}X_{i3}$, $X_{i2}X_{i3}$, $X_{i1}X_{i2}X_{i3}$ represent all interaction terms between these three factors for i^{th} observation

Secondly, to make reliable inference on the significance of factors, normality and constant variance error assumptions are required to be verified using Normal Q-Q plot and residual plot.

Next, there are three methods to determine the significance of factors and interactions. The first method is to use hypothesis test, which is H_0: b=0 v.s. H_1: b≠0. If the corresponding p-value is less than 5%, then H_0 is rejected, and we conclude such factor or interaction term is statistically significant. Second method is to check if 0 is included in the 95% confidence interval constructed by $effect \pm t_{8,\,5\%/2} \times se(effect)$. If 0 is not included, then we can conclude such factor or interaction term is significant. The third method is to use half-normal plots find significant factors and interactions. Also, the cube plot can help to visualize the effect of changing level on the response. Every edge of this cube represent the change of level of one factor while other two factors stay unchanged.

Finally, the effects of factors and interactions can be obtained by multiplying the estimated coefficients by 2 since we assign -1 and 1 to two levels of factors. Also, the standard error of effect is able to be calculated by multiplying the standard error of estimated coefficient by 2, because the experiment is replicated.

3. Results and Discussion

Firstly, a linear regression model is fitted, whose summary is outlined in the table 2. Then we generate Normal Q-Q plot and residual plot to check if $\varepsilon_i \sim N(0\ ,\ \sigma^2)$ for all observations.

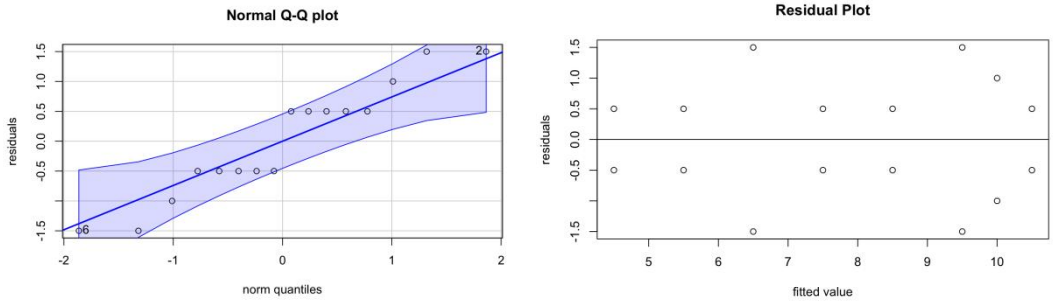

Figure 1: Normal Q-Q plot and residual plot for checking assumption

Although there are lots of points not on the line, these deviated points are in the range of confidence band, the normality is held. From the residual plots, the scatter has no fanning pattern, so the assumption of heterogeneity in the error variance is met.

Table 2: Summary of fitted model.

Factors & Interactions	Coefficients	Std.Error	t value	p-value	Effect
Interaction	b0=7.8125	0.3248	24.056	$9.5*10^{-9}$	
Physical activity(X_1)	b1=1.3125	0.3248	4.041	0.00373	2.625
Alcohol(X_2)	b2=1.3125	0.3248	4.041	0.00373	2.625
Meal(X_3)	b3=0.6875	0.3248	2.117	0.06714	1.375
Physical activity:Alcohol(X_1:X_2)	b4=-0.1875	0.3248	-0.577	0.57958	-0.375
Physical activity:Meal(X_1:X_3)	b5=0.1875	0.3248	0.577	0.57958	0.375
Alcohol:Meal(X_2:X_3)	b6=-0.3125	0.3248	-0.962	0.36410	-0.625

Physical activity:Alcohol:Meal(X_1:X_2:X_3)	b7=-0.3125	0.3248	-0.962	0.36410	-0.625
Std.Error(ε) = 1.299					

To determine the significance of factors and interactions, we can use the first method which is to conduct hypothesis test. Using a threshold of 5%, from table 2, we can observe that p-values of 'Physical activity' and 'Alcohol' are less than 5%, which indicates they are significant factors in affecting the sleep hours.

Table 3: 95% confidence interval

Factors & Interactions	2.5% Lower Bound	97.5% Upper Bound
Physical activity	1.1272	4.122
Alcohol	1.1272	4.122
Meal	-0.1228	2.8728
Physical activity:Alcohol	-1.8728	1.1228
Physical activity:Meal	-1.2228	1.8728
Alcohol:Meal	-2.1228	0.8728
Physical activity:Alcohol:Meal	-2.1228	0.8728

The 95% confidence interval can be used to check the significance. If 0 is included, then such term is not significant in affecting the sleep hours. From table 3, we can observe only the confidence intervals of 'Physical Activity' and 'Alcohol' do not contain 0, which aligns the previous results that only these two factors are significant.

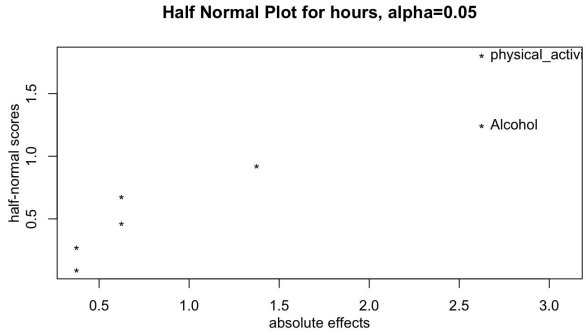

Figure 2: Half normal plot for hours at α=0.05

From half normal plot, we can observe the absolute effects of 'physical_activity' and 'Alcohol' are relatively large, so they are marked as significant factors at α=0.05, which also aligns the previous result.

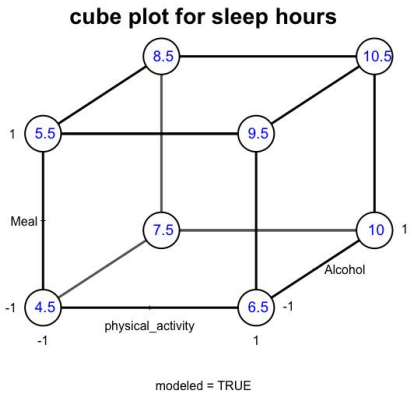

Figure 3: Cube plot

Cube plot can visualize 2^3 possible factors combinations, which helps us understand how changes in one factor affect the sleep hours at various levels of the other factors. For example, the numbers on the four right corners are larger than numbers on the four left corners, which indicates having 'Physical activity' increases the sleep time while other two factors stay unchanged.

From table 2, the effect of factors and interactions can be obtained by multiplying the estimated coefficients in the linear model by 2. We have found out the only two significant factors are 'Physical activity' and 'Alcohol'. Their effect are both 2.625. This means the sleep hours will be increased by 2.625 if a person drink alcohol at dinner compared to not drinking alcohol, and sleep hours will be increased by 2.625 if a person have physical activity during the day compared to not having physical activity. Also, we can observe X_1, X_2, X_3, X_1:X_3 have positive effects, while others have negative effects.

As last, as we have replicated runs, we can calculate the variance of effect = 0.4219

1) $s^2_i = \frac{(y_{i1}-y_{i2})^2}{2}$, y_{i1} and y_{i2} are the replicated two responses under same combinations of factors

2) $S^2_{pool} = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2+ \dots + (n_8-1)s_8^2}{(n_1-1)+(n_2-1)+ \dots + (n_8-1)} = 1.6875$

3) $\text{var(effect)} = \frac{S^2_{pool}}{4} = 0.4219$

Secondly, the standard error of effect can also be obtained by multiplying the standard error of estimated coefficient by 2, which is 0.3248×2=0.6496. Then the var(effect) is obtained by squaring, $0.6496^2 = 0.4219$

4. Conclusion

A 2^3 factorial experiment is conducted to analyze the effect of three lifestyle factors and their interactions on the sleep duration. The experiment is replicated in order to obtain the variance of effect. 16 observations are collected by randomly asking16 person to draw lots. Then they follow a specific combination and record the sleep time. A linear regression model is fitted to find significant factors through methods of hypothesis test, 95% confidence interval and half normal plot. We observe that 'physical activity' and 'alcohol' are significant factors. They both have the effect to increase the sleep hours by 2.625. This slightly contradicts my expectation, because 'Meal' is proved to be insignificant. Although having 'physical activity' and drinking 'alcohol' can prolong sleep time, but these approaches themselves may not be healthy, which needs further study. These discovery indicates objective is successful achieved.

5. References

[1] St-Onge, M.-P., Grandner, M. A., Brown, D., Conroy, M. B., Jean-Louis, G., Coons, M., & Bhatt, D. L. (2016). Sleep Duration and Quality: Impact on Lifestyle Behaviors and Cardiometabolic Health: A Scientific Statement From the American Heart Association. Circulation (New York, N.Y.), 134(18), e367–e386. https://doi.org/10.1161/CIR.0000000000000444

[2] Al-Hazzaa, H. M., Musaiger, A. O., Abahussain, N. A., Al-Sobayel, H. I., & Qahwaji, D. M. (2014). Lifestyle correlates of self-reported sleep duration among Saudi adolescents: a multicentre school-based cross-sectional study. Child : Care, Health & Development, 40(4), 533–542. https://doi.org/10.1111/cch.12051