Table 2: Collected Data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **person** | **Physical activity** | **Alcohol** | **Meal** | **Sleep hours** |
| 1 | -1 | -1 | -1 | 4 |
| 2 | 1 | -1 | -1 | 8 |
| 3 | -1 | 1 | -1 | 8 |
| 4 | -1 | -1 | 1 | 6 |
| 5 | 1 | 1 | -1 | 11 |
| 6 | 1 | -1 | 1 | 8 |
| 7 | -1 | 1 | 1 | 8 |
| 8 | 1 | 1 | 1 | 11 |
| 9 | -1 | -1 | -1 | 5 |
| 10 | 1 | -1 | -1 | 5 |
| 11 | -1 | 1 | -1 | 7 |
| 12 | -1 | -1 | 1 | 5 |
| 13 | 1 | 1 | -1 | 9 |
| 14 | 1 | -1 | 1 | 11 |
| 15 | -1 | 1 | 1 | 9 |
| 16 | 1 | 1 | 1 | 10 |

The presentation is going to talk about a factorial design about the factors of sleep time.

It is divided into 4 parts

Firstly, let’s talk about some background, sleep is crucial for the both physical and mental health, and sleep duration is one of the most significant indicator of sleep quality. After doing some research, it shows lifestyle is one of the main aspects that affect sleeping time. We will focus on three lifestyle factors, which are physical activity, drinking alcohol and having meal before sleep. Hence, our objective is to analyze these three factors on the sleep duration, so that we can know how to increase sleeping duration to keep health.

The next part is the method

Firstly, let me introduce our experiment design. For our objective, we will conduct a factorial design with 3 factors and each factor has 2 levels. So there are 8 combination of these levels and factors. And this is a replicated design, By replications, we mean we will collect the data for each combination twice, so there are 16 observations in total. The specific data collection process is that at first random 16 persons are asked to draw lots to ensure randomness. Then each person is asked to follow the specific combination of their lot and record their sleep duration. Also, in order to minimize the effect of unknown variables on the final analysis, these 16 persons have similar age and same sex. So why we choose factorial design? Firstly, the effects of three lifestyle factors and their interactions can be analyzed at the same time. And this help us to increase the efficiency and to better understand the complex relationships that might not be apparent in simpler experiments.

Here is a table of the three factors and their levels, for example, physical activity =-1 means this person does not have physical activity, and 1 means this person have physical activity, and this factor is represented by X1. and our response is the sleep hour.

And there is a table of collected data

After collecting data, we will perform some statistical analysis to get the effect of our factors on sleep time. Firstly, we will fit a linear regression model in the form of this shown formula here. One notable thing is that εi represents the error, Then in order to make reliable inference on the effect, we need to check normality and constant error variance assumptions using Normal Q-Q plot and residual plot. Third, there are three methods to find significant factors and interaction terms, which are Hypothesis test, confidence interval and Half normal plot. Forth, we can generate a Cube Plot to visualize 8 combinations. Finally, we can eventually obtain the exact Effects of factors and interactions by multiplying the coefficients by 2, and we can calculate the variance of effect.

Next part is the results

Firstly, there is a summary table of the linear model we fit. you can see there are the estimated coefficients, standard error of each factor and interaction term. Also you can see there are p-value and effect. Which will be talked about in the next slides.

Firstly, we need to check the assumptions. As we can see In QQ plot, all points are in the range of confidence band, so the normality is held. And From the residual plots, the scatter has no fanning pattern, so the assumption of constant error variance is met. So we are good to make reliable inference.

Then let’s talk about the three methods to find significant factors and interaction. First method is the Hypothesis test at the significant level of 5%. we are looking for whose p-value is less than 5%, then the null hypothesis is rejected, and we conclude it’s significant. Back to our summary table, we can see the p-values of ‘Physical activity’ and ‘Alcohol’ are less than 5%, so they are significant factors in affecting the sleep hours.

Next method is to construct 95% confidence interval, and check if 0 is included in the

If 0 is not included, then we can conclude such factor is significant.

And it turns out the confidence intervals of ‘Physical Activity’ and ‘Alcohol’ do not contain 0, which aligns the previous result.

Last method is to use half normal plot, and it can help to identity significant factors, which are ‘Physical Activity’ and ‘Alcohol’ in our plot.

Also, we can generate a cube plot. This plot can 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.

Next, back to our summary table, we can get the effect of each factor and interaction by multiply their coefficient by 2, because the level are set to be -1 and +1. For our significant factors, ‘Physical activity’ and ‘Alcohol’. Their effect are both 2.625. This means, sleep hours will be increased by 2.625 if a person drink alcohol at dinner compared to not drinking alcohol, the interpretation is similar for ‘physical activity’.

Also, we can observe X1, X2, X3, X1:X3 have positive effects, while others have negative effects.

Lastly, we can calculate the variance of effect which is about 0.4, by two method. The first method use the pooled variance of observation, while second method is simply multiply the standard error of coefficient by 2 and square.

To sum up, a factorial experiment with 3 factors and 8 combinations, 16 observations is conducted to analyze the effect of three lifestyle factors and their interactions on the sleep duration. It’s found out ‘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 to keep healthy, but these approaches themselves may be unhealthy, which need further study