Final Project

Stat 425

Group: Xiangchen Guo, Zijian Shen, Sean Pang

Date: 05/10/2025

Xiangchen Guo, Zijian Shen, Sean Pang

1 Research content

Explore how different personal behaviors and environmental factors affect reaction time, sleep duration, and cognitive performance. Through data cleaning and preprocessing, multiple categorical variables are transformed into factors suitable for analysis, and the data is further grouped and simplified. Descriptive statistics describe the basic situation of respondents, including the distribution of variables such as reaction time, age, and sleep duration, and analyze the differences in characteristics among different groups. Subsequently, exploratory data analysis was conducted to examine the impact of factors such as fatigue, stress, and exercise on reaction time, and to observe the potential effects of caffeine intake and video game duration on sleep duration.

The main issues raised in the study include:

(1) Do factors such as fatigue, stress, and exercise significantly affect reaction time, and do these factors exhibit a non-linear relationship?

(2) Is there a significant relationship between video game duration and sleep duration, and is it regulated by caffeine intake?

(3) Does cognitive performance differ under the influence of environmental variables such as device types and network stability?

2 data preprocessing

2.1 Data cleaning

Read the data file in CSV format, convert one of the text variables representing levels into an ordered factor, then merge multiple categories of this variable into three categories and create a new grouped variable. Next, create a binary variable based on a certain field to indicate whether specific conditions are met, converting it to a factor type. After that, remove a variable with extremely imbalanced information to simplify analysis. Then explicitly convert several other categorical variables to factor form, some of which are set with specific orders or levels. Finally, quickly preview the entire cleaned data structure and variable types.

```{r}
survey
```

A tibble: 145 x 29

First <dbl>	Second <dbl>	Third <dbl>	Class <chr>	Age <dbl>	AvgSleepTime <dbl>	LastNightSleep <dbl>	HoursAweak <dbl>	Fatigue <ord>	Stress <ord>
734	684	783	Senior	22	8.00	8.00	13.0	Moderately fatigued	Very Low
647	510	565	Sophomore	20	8.00	8.00	11.0	Moderately fatigued	Low
779	608	566	Senior	21	7.00	8.00	9.0	Moderately fatigued	Low
693	705	652	Junior	21	8.00	5.00	6.0	Slightly Fatigued	Moderate
310	292	298	Junior	21	7.00	6.00	3.0	Slightly Fatigued	Very Low
512	490	537	Senior	21	10.00	17.00	9.0	Not fatigued at all	Moderate
623	618	527	Junior	20	8.00	10.00	12.0	Slightly Fatigued	Low
386	360	362	Junior	21	6.00	8.00	6.0	Not fatigued at all	Very Low
473	402	391	Senior	22	5.00	5.00	2.0	Extremely fatigued	Moderate
588	549	514	Junior	22	6.50	10.00	4.0	Moderately fatigued	High

1-10 of 145 rows | 1-10 of 29 columns   Previous 1 2 3 4 5 6 ... 15 Next

2.2 Descriptive statistics

Description statistics for continuous variables							
variable name	**min**	**1 quartile**	**median**	**mean**	**3 quartile**	**max**	**meaning**
First	212	512	588	592.3	663	1050	Average time of first test hit (millisecond)
Second	292	491	561	559.8	625	868	The second test averaged the time of hits
Third	298	490	542	548.7	615	819	The third test measured the average time of hits
Age	18	20	21	20.88	21	26	age
AvgSleepTime	4	6.5	7	7.25	8	11	Average daily sleep time (hours)
LastNightSleep	0	6	7	7.16	8	17	Sleep duration last night (hours)
HoursAweak	0	4	6	7.45	10	24	Time spent awake (hours)
AvgHoursExercise	0	2	4	4.86	6	30	Average weekly exercise time (hours)

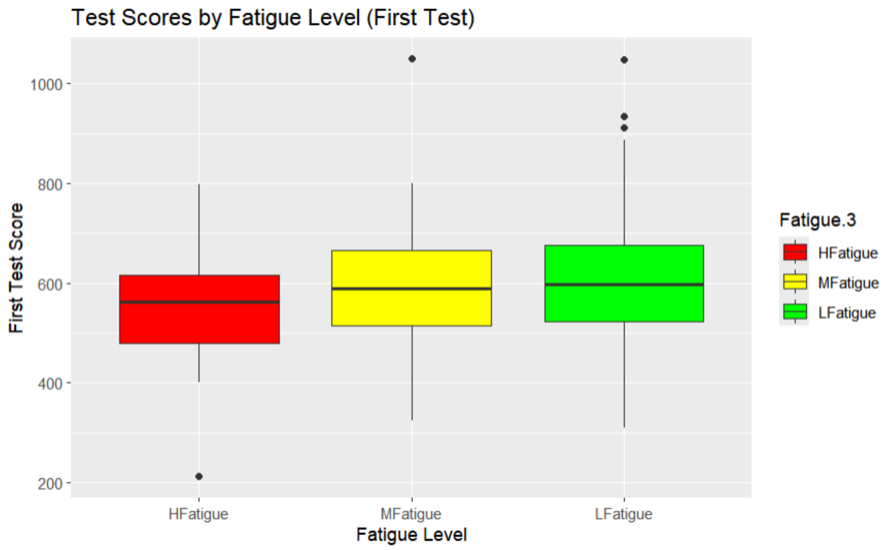
The average reaction time of the first test (592ms) was higher than that of the second and third tests, which may indicate an improvement effect of the exercise. The majority of respondents are concentrated between the ages of 20-21, with little fluctuation. The sleep data shows a healthy level, but there are a few extreme values (such as 0 hours of sleep, 17 hours). The distribution of exercise data is wide, and the maximum value of 30 hours may be an extreme value.

Category variable frequency statistics			
Variable Name	**Category Levels**	**Frequency / Proportion**	**Description**
Fatigue.3	HFatigue / MFatigue / LFatigue	19 / 49 / 77	Recoded fatigue level
Stress	Very Low / Low / Moderate / High / Very High	27 / 39 / 50 / 23 / 6	Perceived stress during gameplay
Distraction	Yes / No (not explicitly counted)	— (Total: 145 entries)	Whether distractions occurred
NoiseLevel	1–10 (most common: 3, 2, 1)	3(36), 2(31), 1(29), ...	Noise level in the environment
Temperature	Neutral / Cold / Warm / Very Warm	111 / 14 / 18 / 2	Perceived ambient temperature
Gamer	Yes / No	69 / 76	Whether the participant considers themselves a gamer
CaffeinIntake	Yes / No	36 / 109	Caffeine intake within 3 hours prior to gameplay
PrimarilyHand	Right / Left / Ambidextrous	133 / 7 / 5	Dominant hand used for tasks
RH	Y / N	133 / 12	Whether the participant is right-handed (simplified)
VisualAcuity	Excellent – Very Poor	— (Frequencies not shown)	Corrected visual clarity
InputDevice	Multiple types	—	Input device used for gameplay
RefreshRate	Standard / Mid / High	—	Screen refresh rate
DeviceOS	Multiple types	—	Operating system or device used
WiFi	Stable / Unstable	—	Stability of internet connection

Most respondents had moderate or low levels of fatigue (MFatigue and LFatigue accounted for 87%). The stress levels are concentrated in Moderate (50 people) and Low (39 people). Most people are at moderate noise (3) and neutral temperature. The proportion of gamers is about 47.6%, and caffeine intake accounts for 24.8%.

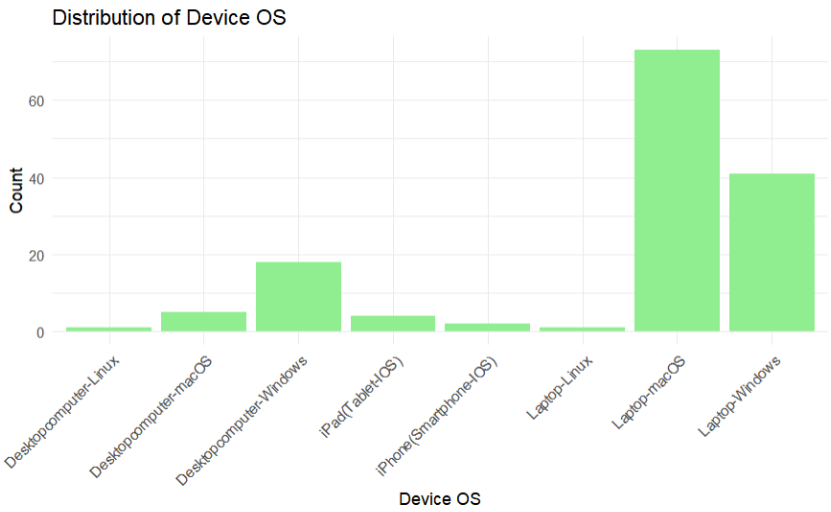
3. Exploratory data analysis

3.1 Comparison of fatigue levels and response time (first round)



The box plot shows the response time distribution of different fatigue levels during the first test. The overall pattern indicates that as the level of fatigue increases, the distribution of reaction time also expands, with higher variability observed in the extreme fatigue group. This indicates that participants with extreme fatigue often have more unstable and slower reaction times, which may be due to the negative effects of fatigue. In contrast, people with moderate fatigue exhibit more concentrated and consistent performance, indicating that their reaction time changes less. The mild fatigue group appears to have the most stable and predictable reaction time, with fewer outliers, indicating that mild fatigue does not significantly impair performance. This pattern reflects how different levels of fatigue affect reaction time, with extreme fatigue leading to more inconsistent results, while moderate to mild fatigue appears to have a smaller impact.

3.2 Distribution of Operating System Types



The distribution of devices using different operating systems is relatively concentrated in certain types. Most participants use laptops running Windows and macOS, each occupying a relatively large proportion. Next, there is also a certain distribution in the usage frequency of desktop computers running Windows and macOS, but they are fewer in number compared to the former. Other types of devices, such as desktops and laptops with Linux operating systems, iPads (iOS tablets), and iPhone (iOS smartphones), show characteristics of lower usage rates. Overall, the data reflects a device distribution dominated by Windows and macOS, with other devices having lower usage frequencies.

3.3 Comparison of weekly exercise duration and first round reaction time

From the data, it can be seen that there is a certain relationship between different exercise times and reaction times. The distribution of reaction time is relatively wide and covers multiple levels of exercise time from low to high, indicating that there may be certain patterns or trends between exercise time and reaction time. Some participants may have higher reaction times at low exercise times, while at higher exercise times, reaction times may be lower or more stable.

4. Model construction

4.1 Multiple Linear Regression

Predictor	Estimate	Std. Error	t value	P	Significance
(Intercept)	582.3575	154.0334	3.781	0.000257	***
Fatigue.3LFatigue	82.1195	38.4976	2.133	0.035181	*
Stress^4	61.319	20.327	3.017	0.003188	**
AvgHoursExercise	-4.7851	2.4451	-1.957	0.052931	.
RHN	100.9457	37.5348	2.689	0.008294	**
InputDeviceMouse	-167.1352	84.1193	-1.987	0.049468	*
InputDeviceTouch screen	-196.5561	101.8504	-1.93	0.056249	.
RefreshRateStandard	60.3629	29.7419	2.03	0.04486	*
WiFiUnstable	169.2559	87.0499	1.944	0.054454	.

The data in the above table are significant fields. The goal of the multiple linear regression model is to explore the impact of various personal behaviors and environmental factors on the first reaction time (First). The overall fit of the model is good, with a significant F-test result, indicating that the explanatory variables included in the model generally have statistical significance for reaction time. From the adjusted R^2 value, the model can explain about half of the variation in reaction time, suggesting that although multiple variables have been incorporated into the model, some variations still originate from factors outside the model, such as individual differences or other unmeasured variables.

At the variable level, changes in fatigue levels are associated with reaction times, with lower-fatigue groups responding faster compared to higher-fatigue groups; significant terms also appeared in the polynomial of stress, suggesting a possible nonlinear relationship between stress and reaction time; furthermore, right-handed (RH) users have faster reaction times, and there are differences in reaction times for certain types of input devices (such as mice and touchscreens), indicating that device type may affect task execution efficiency. Screen refresh rates and network stability also influence reaction performance to some extent, possibly reflecting the indirect effects of device performance or usage environment on job performance.

The model revealed the potential relationship between sleep, exercise, fatigue, device characteristics and other factors and cognitive response ability. The explanatory power of a single variable is limited, but through multi-factor comprehensive analysis, we can more comprehensively understand which behavioral and environmental variables may have an important impact on cognitive performance.

4.2 ANOVA (Effects of different caffeine intake and game duration on sleep duration)

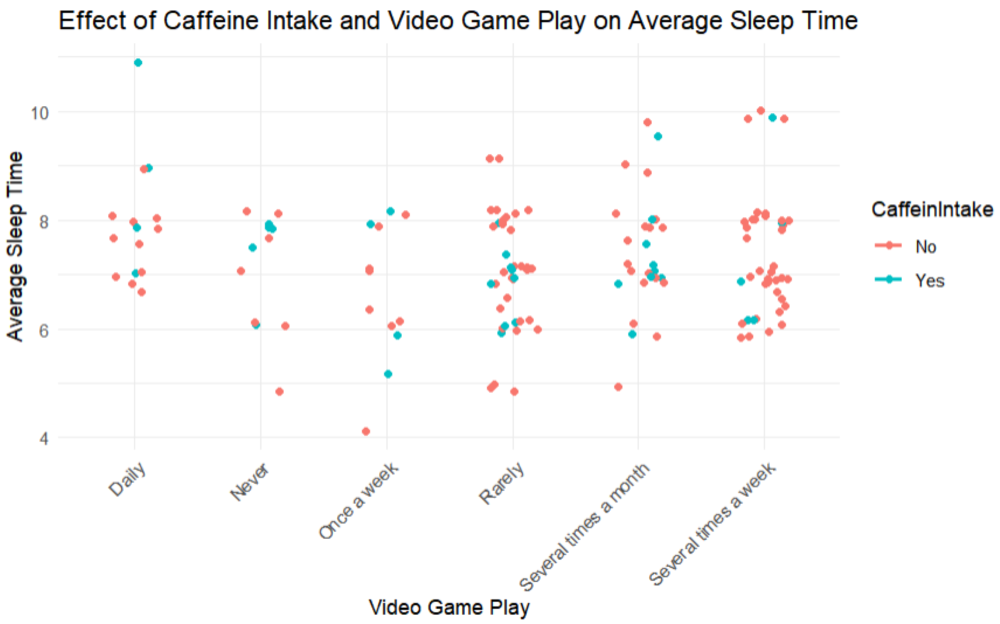
ANOVA (Analysis of Variance) is a statistical method used to compare whether the mean differences between multiple groups are significant, typically for testing the influence of one or more independent variables on the dependent variable. Its fundamental principle involves dividing total variance into between-group variation and within-group variation. The F-value (the ratio of between-group variation to within-group variation) is calculated to determine if there are significant differences between different groups. If the F-value is large and the corresponding p-value is less than the significance level (usually 0.05), the null hypothesis is rejected, indicating that there are significant differences in the means between the groups.

Source	**Df**	**Sum Sq**	**Mean Sq**	**F value**	**Pr(>F)**
CaffeinIntake	1	0.35	0.3493	0.281	0.5968
VideoGamePlay	5	15.61	3.1219	2.513	0.0329 *
CaffeinIntake:VideoGamePlay	5	5.23	1.0467	0.843	0.5217
Residuals	133	165.2	1.2421		

According to the results of ANOVA, caffeine intake had no significant effect on average sleep duration (AvgSleepTime) (p=0.5968), indicating that caffeine intake did not significantly alter the sleep duration of the subjects in this study. In contrast, the duration of video games has a significant impact on sleep duration (p=0.0329), indicating a certain relationship between the duration of video games and sleep duration. Although there is an interaction term (the interaction effect between caffeine intake and video game duration), the interaction effect did not reach significance (p=0.5217), indicating that the combined effect of caffeine intake and video game duration does not have a significant impact on sleep duration.

Comparison	Difference	Lower Bound	Upper Bound	p-Value
Caffeine Intake (Yes vs No)	0.1136	-0.3101	0.5374	0.5968
Video Game Play				
Never - Daily	-0.817	-2.0652	0.4311	0.4111
Once a week - Daily	-1.2826	-2.5307	-0.0344	0.0402 *
Rarely - Daily	-0.9087	-1.8879	0.0704	0.0853
Several times a month - Daily	-0.4816	-1.5195	0.5562	0.761
Several times a week - Daily	-0.5152	-1.4909	0.4606	0.6477
Caffeine Intake and Video Game Play Interaction				
Yes:Daily - No:Daily	1.1591	-1.0059	3.3241	0.8252
No:Never - No:Daily	-0.7766	-2.5694	1.0162	0.9529
Yes:Never - No:Daily	-0.0909	-2.0909	1.9091	1

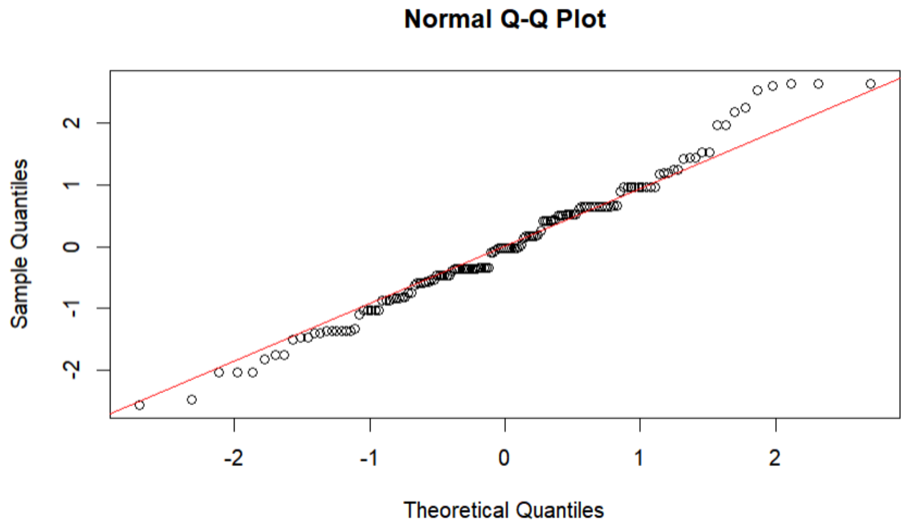
The results of TukeyHSD showed that there was no significant difference in the comparison of caffeine intake between "yes" and "no" (p=0.5968), indicating that there was no significant correlation between caffeine intake and average sleep duration. In terms of game duration (VideoGamePlay), there was a significant difference only between "once a week" and "every day" (p=0.0402), indicating that the sleep duration of the group who played games once a week was significantly shorter than that of the group who played games every day. However, comparisons between other game duration groups (such as "never" versus "every day", "occasionally" versus "every day", etc.) did not show significant differences (p>0.05). In addition, the interaction between caffeine intake and game duration did not show a significant effect (p-values for all interactions were greater than 0.05), indicating that the combined effect of these two factors had no significant impact on sleep duration.



From the scatter plot, it can be seen that video game duration and caffeine intake have some impact on average sleep time, but this effect is not entirely consistent. People with shorter video game durations (such as "Never," "Rarely," "Once a week") have a more varied distribution of sleep times, while those with longer game durations (such as "Several times a week," "Daily") sometimes have lower sleep times, especially among individuals who do not consume caffeine. Additionally, individuals who consume caffeine tend to have more concentrated sleep times, and in some cases, those who consume caffeine have longer sleep times. However, overall, the combination of video game duration and caffeine intake affects sleep time through various factors, leading to significant individual differences.

4.3 Residual diagnostic diagram

Residual analysis is part of regression model diagnostics, aiming to check whether the model's assumptions hold true. Residuals are the differences between actual observations and predicted values. Ideally, residuals should be randomly distributed with no discernible pattern. If residuals exhibit non-normal distribution, it may indicate that the model has misspecification or other anomalies have not been captured.



By comparing the quantiles of the residuals with those of the standard normal distribution, the points in the graph are arranged along a straight line, indicating that the residuals follow a normal distribution.

5 Conclusion

This study explores the complex relationships between multiple variables by exploring the effects of different individual behaviors and environmental factors on cognitive response ability and sleep duration. Firstly, factors such as fatigue, stress, and exercise have a significant impact on reaction time, with higher levels of fatigue leading to greater fluctuations in reaction time. Participants who experience extreme fatigue have slower and more unstable reaction times. Moderate exercise time can help reduce reaction time and improve reaction consistency. Through multiple regression analysis, we found that participants who use their right hand have faster reaction times compared to those who use their left or non dominant hand, and device type and network stability also have a certain impact on cognitive performance.

For the analysis of sleep duration, research has shown that there is a significant negative correlation between video game duration and sleep duration, especially among participants who play games daily and have significantly lower sleep duration than those who occasionally or do not play games. However, there was no significant direct effect between caffeine intake and sleep duration, indicating that caffeine's impact on sleep may be regulated by other factors. In addition, although the interaction between video game duration and caffeine intake did not reach statistical significance, there are individual differences in the combined effect of these two factors on sleep duration, suggesting that future research may need to further explore the interaction effects of multiple factors.

The results of this study indicate that individual behavioral habits (such as fatigue, exercise, and game duration) and environmental variables (such as devices and networks) have a profound impact on cognitive performance and health behavior, providing a theoretical basis for optimizing individual health management and improving cognitive performance.