

How Different Factors Influence Undergraduates on Playing Computer Games

May 1, 2019

Contents

1	Purpose and Sampling	2
2	Data Pre-Analysis	4
2.1	Data Prior Analysis	4
2.2	Data Pre-Processing	5
3	Data Analysis	5
3.1	Correlation Matrix	5
3.2	Wilcox Test of Playtime Between Different Grades	7
3.3	Quantitative Analysis	7
4	Superiority and Weakness	9
5	Summary	9

1 Purpose and Sampling

In general, playing computer games is supposed to highly correlated with the undergraduates' study. In order to investigate how different factors influence undergraduates on playing computer games, we conducted a survey and focused on how long a **SUSTech** student spent in playing computer games within one week under different preferences. Questionnaire was designed and 93 responses were received before data analysis.

Before implement our questionnaire, we first identified the target population, element, sampling unit, frame, sampling method of our sampling survey, which are listed below.

- Population: the whole undergraduates of **SUSTech**
- Element: an individual student of **SUSTech**
- Sampling unit: an individual student of **SUSTech**
- Frame: a list of all undergraduates of **SUSTech**
- Sampling method: Stratified random sampling
- Stratas: split the whole undergraduates into 4 grades (freshman,sophomore,junior,senior), each grade is defined as a strata.

The questionnaire consists of 18 questions intended to go to detail of undergraduates playing computer games. For instance, what kinds of computer games does the interviewees like most, or would they make a plan about their playtime allocation. The complete questionnaire is listed below.

- | | |
|---------------------------|------------------------------|
| 1. 您的性别 | 11. 您会熬夜打游戏吗(熬夜指晚上 12:00 以后) |
| 2. 您的年龄 | 12. 在哪些情况下, 您会熬夜打游戏 |
| 3. 您的年级 | 13. 您一般几点睡 |
| 4. 您所在院系或意向院系 | 14. 您熬夜是因为打游戏吗 |
| 5. 您每周大概玩多长时间游戏(整数, 以小时计) | 15. 您觉得自己打游戏影响到自己的学习生活了吗 |
| 6. 您平时玩游戏的类型 | 16. 您觉得自己打游戏有没有影响到他人的学习生活 |
| 7. 您玩游戏的原因是 | 17. 您觉得别人打游戏有没有影响到您的学习生活 |
| 8. 您是自己玩还是和朋友开黑 | 18. 有因为自己或他人打游戏和周围的人发生过矛盾吗 |
| 9. 玩游戏前有对玩游戏这件事进行计划吗 | |
| 10. 玩游戏有固定的时间吗 | |

Futhermore, the explanation of symbols can be seen in table 1.

Symbol	Explanation
X_1	Gender
X_2	Age
X_3	Grade
X_4	Apartment or desired apartment
X_5	Hours spend on playing games per week
$X_6^{(1)}$	The game type usually play is MOBA
$X_6^{(2)}$	The game type usually play is FPS
$X_6^{(3)}$	The game type usually play is RPG
$X_6^{(4)}$	The game type usually play is card games
$X_6^{(5)}$	The game type usually play is webgames
$X_6^{(6)}$	The game type usually play is console games
$X_6^{(7)}$	Usually play other games
$X_6^{(8)}$	Student who don't play games
$X_7^{(1)}$	The reason of playing games is to release pressure
$X_7^{(2)}$	The reason of playing games is to kill time
$X_7^{(3)}$	The reason of playing games is having passion for games
$X_7^{(4)}$	The reason of playing games is to follow e-sport trend
$X_7^{(5)}$	Driven by friends
$X_7^{(6)}$	Invited by friends
$X_7^{(7)}$	Other reasons
$X_7^{(8)}$	Student who don't play games
$X_8^{(1)}$	Tend to play alone
$X_8^{(2)}$	Tend to play with friends
$X_8^{(3)}$	Can play alone or play with friends
$X_8^{(4)}$	Student who don't play games
X_9	Whether have plan before play games
X_{10}	Whether have settled playing time
X_{11}	Whether stay-up to play
$X_{12}^{(1)}$	Will stay-up to play when there is little learning task
$X_{12}^{(2)}$	Will stay-up to play when want to achieve certain goals in games
$X_{12}^{(3)}$	Will stay-up to play when made a promise with friends
$X_{12}^{(4)}$	Others
$X_{12}^{(5)}$	Don't stay-up to play games
X_{13}	Time to sleep (Time interval between choices is 1 hour)
$X_{14}^{(1)}$	Due to playing games
$X_{14}^{(2)}$	Don't due to playing games
$X_{14}^{(3)}$	Sometimes due to playing games
$X_{14}^{(4)}$	Don't stay-up
X_{15}	The influence to the player brought by playing games
X_{16}	The influence to others brought by playing games
X_{17}	The influence brought by others
X_{18}	Whether had conflicts with others caused by games

Table 1: Symbols Explanation

2 Data Pre-Analysis

2.1 Data Prior Analysis

Our data pre-processing goal is to obtain a prior sample variance of each grade. After that we are able to obtain the sample size n such that the estimated average playtime is within an error bound B .

Play Time	Freshman	Sophomore	Junior	Senior
2.5	5	1	15	3
7.5	1	4	10	4
12.5	0	1	2	0
17.5	2	0	1	2
22.5	0	0	1	0
27.5	3	2	4	1
Average Play Time	12.5	12.5	8.71	10

Table 2: Data of playtime of different grades

Until the data pre-processing, 63 responses were received, which is listed as table 2. We used the 63 data to estimate the sample variance of playtime of each strata, noted by s_i^2 ($i = 1, 2, 3, 4$) respectively.

$$s_1^2 = 113.64, \quad s_2^2 = 81.25, \quad s_3^2 = 69.74, \quad s_4^2 = 61.25$$

From the official website of **SUSTech**, we obtain the population sizes of the whole undergraduates and of different grades, noted by N and N_i ($i = 1, 2, 3, 4$) respectively.

$$N_1 = 1005, \quad N_2 = 994, \quad N_3 = 937, \quad N_4 = 609, \quad N = \sum_{i=1}^4 N_i = 3545$$

It is reasonable to use N_i/N to estimate n_i/n , which is noted by ω_i .

$$\begin{aligned} \omega_1 &= \frac{N_1}{N} = \frac{1005}{3545} & \omega_2 &= \frac{N_2}{N} = \frac{994}{3545} \\ \omega_3 &= \frac{N_3}{N} = \frac{937}{3545} & \omega_4 &= \frac{N_4}{N} = \frac{609}{3545} \end{aligned}$$

Different error bounds B have been attempted, distinguish n has been obtained as table 3 by formula(1)

$$n \approx \frac{\sum_{i=1}^4 \frac{N_i^2 \sigma_i^2}{\omega_i}}{N^2 D + \sum_{i=1}^4 N_i \sigma_i^2}, \quad \text{where } D = \frac{B^2}{z_{\alpha/2}^2} \quad (1)$$

B	1	1.77	2	3
n	296	100	79	36
n_1	84	28	22	10
n_2	83	28	22	10
n_3	78	27	21	10
n_4	51	17	14	6

Table 3: Values of B and n_i

We want to minimize the error bound B with a relatively small sample size n , so we choose the error bound equals to 2 and the sample size equals to 79.

Estimating σ_i^2 by s_i^2 in formula (1), we compute that $n = 79$ under a 95% confidence level. we conclude that $n_1 = 22$, $n_2 = 22$, $n_3 = 21$, $n_4 = 14$ by $n_i = n \left(\frac{N_i}{N} \right)$.

Theoretically, it will suffice to draw 79 responses from the whole received 93 responses, with 22 responses from freshman, 22 responses from sophomore, 21 responses from junior and 14 responses from senior. The estimated average playtime will be within an error bound $B = 2$.

However, there exists data missing because only 18 responses from sophomore have been received. Thus we set $n_2 = 18$ and therefore $n = 75$.

In summary, $n_1 = 22$, $n_2 = 18$, $n_3 = 21$, $n_4 = 14$ and $n = 75$.

2.2 Data Pre-Processing

In the primitive table, the value of some variables (question 9, 10, 11, 15, 16, 17, 8 and 14) were determined by which option the responder chose, that is, the values of these variables were exactly the rank of option they chose. In this case, this variables will be non-meaningful for our data analysis. Thus, on the one hand, for question 9, 10, 11, 15, 16, 17, we utilize 0, 1, 2, 3, 4 to represent the increasing degree of the options. Figure 1a and figure 1b have shown this pre-processing.

15、您觉得自己打游戏影响到自己的学习	16、您觉得自己打游戏有没有影响到他人	17、您觉得别人打游戏有没有影响到你			
3	2	3	x15	x16	x17
3	2	2	3	2	3
2	2	2	3	2	2
5	5	5	0	0	5
5	5	5	0	0	5
2	2	1	2	2	1
1	2	2	1	2	2
5	5	5	0	0	5
3	2	2	3	2	2
5	5	2	0	0	2
2	2	1	2	2	1
2	2	2	2	2	2
4	2	1	4	2	1
1	2	5	1	2	5
2	2	1	2	2	1
1	2	1	1	2	1
	2	1	1	2	1

(a) Before data pre-processing

(b) After data pre-processing

Figure 1: Data with and without pre-processing

On the other hand, the options of some other questions (question 8 and 14) does not show apparent degree increasing. Therefore, we use dummy variables to represent different options. Figure 2a and figure 2b have shown this pre-processing.

18、您是自己玩还是和朋友开黑	x8_1	x8_2	x8_3	x8_4
3	0	0	3	0
3	0	0	3	0
4	0	0	0	4
4	0	0	0	4
3	0	0	3	0
1	1	0	0	0
4	0	0	0	4
3	0	0	3	0
4	0	0	0	4
3	0	0	3	0
3	0	0	3	0
3	0	0	3	0
3	0	0	3	0
2	0	2	0	0
1	1	0	0	0
2	0	2	0	0

(a) Before data pre-processing

(b) After data pre-processing

Figure 2: Data with and without pre-processing

3 Data Analysis

3.1 Correlation Matrix

Before data prior processing, correlation matrix of all variables is plotted as figure 3a. As the plot shown, this plot could not represent the correlation of different factors with playtime very well. Figure 3b shows

the correlation matrix after data prior processing. The correlation of various factors with playtime could be analysis evidently. Furthermore, we delete some factors that have few correlation with playtime, which is exhibited as figure 3c.

From figure 3b, we can find more factors that correlated to playtime compared to figure 3a (X_5 represents playtime). From figure 3a we find play MOBA($X_6^{(1)}$) has the strongest correlation with playtime, which means MOBA is the most addictive game type; while single-player game($X_6^{(5)}$) is the least addictive game type since it has the weakest correlation with playtime. Since $X_7^{(2)}$, $X_7^{(3)}$, $X_7^{(4)}$ have positive correlation with playtime, we can safely draw the conclusion that students who love game tend to spend more time on playing computer games.

Analysis 3.1: Negative correlation analysis

From correlation matrix figure 3c, we can find factors “Nonplayer”, “Nonplayer₁”, “Nonplayer₂”, “Whether stay-up to play”, “Condition₅”, “Reason for stay-up₂” have strong negative correlation with “PlayTime”.

There is an interpretation of this conclusion. It is reasonable that non-player will have a shorter play time. Setting the value of stay-up to play games be 1 and not stay-up to play games be 3, we find out that the larger the value of this factor is, the smaller the value of playtime is, which is coincident with the fact that students who don't stay-up to play tend to have a shorter playtime. “Condition₅” and “Reason for stay-up₂” also represent situations that students would not stay-up to play games.

Analysis 3.2: Positive correlation analysis

“Play time” is positively correlated with “plan” and “settle time”, that is, those who have no plans and who often stay up late tend to spend more time on games. “Play time” positively correlated with X_{13} and $X_{14}^{(1)}$ also prove this conjecture. Comparing “Play time” with X_{15} and X_{16} , we conclude that those who care about their lives or others also have less game time.

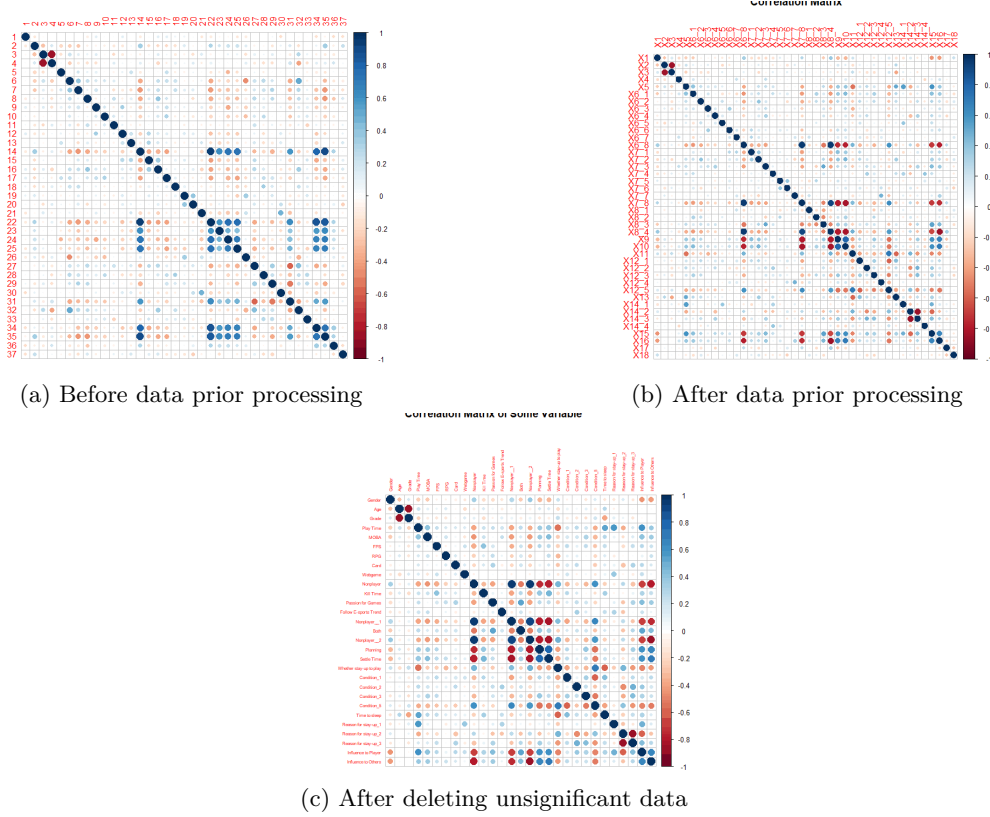


Figure 3: Data with and without prior processing

3.2 Wilcox Test of Playtime Between Different Grades

We utilize wilcox test to judge whether there exist difference between distinguish grades' average playtime.

From wilcox test, we conclude that there are not sufficient evidences to reject the null hypothesis, that is, for grade₁ and grade₂, grade₁ and grade₃, grade₂ and grade₃, grade₂ and grade₄, there are no difference between these grade pairs' average playtime.

However, there are sufficient evidences to reject null hypothesis for grade₃ and grade₄ as long as grade₁ and grade₄, which means that there are clearly differences between these grade pairs' average playtime.

3.3 Quantitative Analysis

Average game time for four grades(senior, junior, sophomore, freshman) are respectively 20.21, 7.14, 12.72 and 7.55. Estimated playtime for students in **SUSTech** is 11.07 hours per week. Boxplot of playtime vs. the grades are shown as figure 4.

$$\bar{y}_{st} = \frac{1}{N} \sum_{i=1}^L N_i \bar{y}_i \quad (2)$$

Applying the CLT, $\frac{\bar{y} - \mu}{\sqrt{\text{Var}(\bar{y}_{st})}} \sim N(0, 1)$. Thus a 95% confidence interval for the estimated game time is [8.08, 14.05]. The 95% confidence interval for the average game time for the four grades(senior, junior, sophomore, freshman) are respectively [11.03, 29.40], [3.76, 10.53], [5.89, 19.55] and [2.60, 12.48]

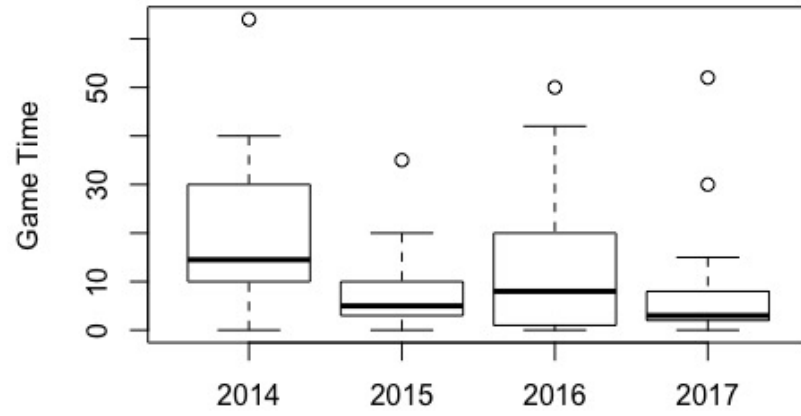


Figure 4: Boxplot of gametime vs. the grades

As it is shown in the correlation plot, playtime and MOBA player are positively correlated, box-plot of MOBA players' game time and non-MOBA players' game time are shown as figure 5:

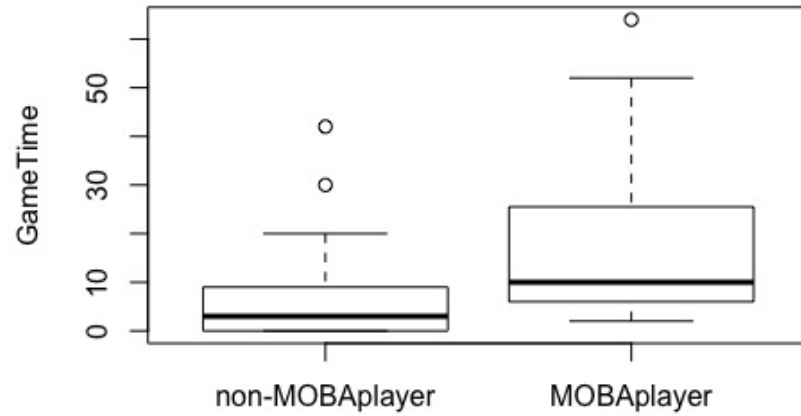


Figure 5: Gametime vs. MOBA

A Wilcoxon test is taken and the result is shown as figure 6. P-value is less than 0.01, it can be concluded that MOBA players' game time is greater than non-MOBA players' playtime with confidence level 0.99. This result implies that MOBA players' playtime tends to be longer.


```

Wilcoxon rank sum test with continuity correction

data: A$GameTime[A$MOBA == 1] and A$GameTime[A$MOBA == 0]
W = 1088.5, p-value = 8.318e-06
alternative hypothesis: true location shift is greater than 0

```

Figure 6: Correlation between parameters

Another pair of highly negatively correlated parameters are game time and gender, it suggests the understandable conclusion that boys' game time is much longer than girls' game time.

4 Superiority and Weakness

Analysis 4.1: Superiority

From above analysis, we know that the sample mean for playtime spent on games is an unbiased estimator. Furthermore, we obtain a prior sample variance of each grade to identify the sample size n .

Analysis 4.2: Weakness

There are also some problems that will effect the accuracy of result. Firstly, there exists missing data in sample, we didn't find a proper way to process these data. Secondly, we think that the sampling method will lose a little precision for the reason that some questions are too subjective.

5 Summary

Our analysis on data can mainly divided into three parts.

Analysis 5.1: Prior Analysis

Firstly, we performed the prior analysis on data, in this section we determined the sample size n_i for each grades such that the estimated average playtime is within an error bound B .

Analysis 5.2: Qualitative Analysis

Secondly, we performed the qualitative analysis on data, in this section we first reset the value of some variables to make it meaningful for the aim to get the correlation matrix. From the pictures above we can clearly point out that whether playing MOBA games is the biggest factor affecting the playtime. Furthermore, we can also conclude that boy's playtime is much longer than girl's playtime.

Analysis 5.3: Quantitative Analysis

Last, we perform the quantitative analysis. In this section, we mainly analysis the value of playtime for each grades. We first calculate the mean value and the confidence interval, then we perform a wilcoxon rank test for all four mean values of playtime for four grades, from which we can conclude that there are clearly difference between the average playtime for the two pairs: grade₃ and grade₄, grade₁ and grade₄.