Relationship Between Wake-up time and the total screen time in the previous day using Federal and Meta Learning

Team Hot Pot; Xueting Tao, Isabel Gomez

2/14/2022

Abstract:

Introduction:

The use of cell-phones and their impact on our daily lives is an ongoing relationship that scientists try to understand. This study focused on the relationship between wake-up time and total screen time in the previous day.

Method:

The data was collected from two students of Biostatistics 620 class, using the summary report in the cellphone. Since we assume that the individual level data could not be shared, thus federal learning method was applied in the simple linear regression(SLR) and meta learning methods was applied to the multiple linear regression (MLR) adjusted for school days.

Results:

We found both insignificant relationship between wake-up time and total screen time in the previous day in both adjusted and unadjusted analysis. The School day is significantly related to earlier wake-up time. Federal learning method yields the same results and SLR using overall data while Meta learning gives slightly different results.

Conclusion:

We cannot varify the relationship between wake-up time and total screen time in the previous day in this study. Further study might be needed with adjusting more baseline confoundings and with bigger sample size.

Introduction:

The use of cell-phones and their impact on our daily lives is an ongoing relationship that scientists try to understand. Many studies have shown the impact of phones on various health outcomes such as sleep. A study conducted in 2015, showed that screen time is adversely associated with sleep time [1]. Other studies have shown that cellphone usage is associated with a later wake-up time [2]. Given these findings, we aim to study if a similar pattern arises among the screen time usage in master's level students at the University of Michigan. The specific data collected includes total screen time in minutes, first pick-up time and the number of pickups in a day. With this specific individual-level data collected, the goal is to better understand phone usage behavior in this group.

The hypothesis for this study is that screen-time usage is associated with later wake-up times. Here we use the first pick-up time to approximate the wake-up time. In order to test this, we fitted linear regression models using the Federal Learning and the Meta Learning process. In this study we found non-significant relationships between screen time usage and wake-up time.

Data Description:

The data was collected between January 3rd, 2022 – February 13, 2022 (20 days) from two students in the Biostatistics 620 class is used in this study. The data was collected both from xuetao's HUAWEI Mate 30 pro phone and isgomez's iphone. Overall following variables were collected for the study:

- ID: the umich unique name for the individual, recorded as string type.
- Date: Recorded everyday, recorded as date type
- Tot.Soc.Time: Total Social Screen Time for the day, recorded in MM format
- Pickups: Total numbers of pick-up, recorded as discrete numbers.
- Pickup.1st: Time at first pick-up, Recorded as time type

Table 1 above shows the drastic difference in screen usage and number of pickups between the two participants in this study. As a team, the average amount of total screen time between January 3rd - February 13, 2022 is 421 minutes or approximately 7 hours. The average social screen time is 105 minutes or 1 hour and 45 minutes and the average number of daily pickups was 96.

Figure 1 displays the pairwise scatter plots of total screen time, social screen time and number of pickups for the group. There is moderately weak significant negative correlation between total screen time and number of pickups (Pearson correlation = -0.292). Social screen time has a weak positive correlation with both total screen time (pearson correlation = 0.035) and number of pickups (pearson correlation = 0.181), both of these are non-significant.

Figure 2 shows the first pick-up time for each member, along with the average of all the participants together. From this figure we can see that the most popular pickup time is

around 7:30am. Whereas if we look at the individual pickup time, isgomez has a more variable pickup time with the earliest being at 8:30am and xuetao has a more consolidated pickup time around 7:30am.

Data Preprocessing:

The data for both individuals was manually combined using an excel spreadsheet. Previous work on Data validation was conducted at the individual level, therefore we assume that all the data reported in this spreadsheet is accurate. In order to ensure that all the units are the same, a data transformation was made in the first-pick up time to convert the hours into minutes with midnight being 0 minutes and 11:59pm being 1440 minutes. An additional variable was made by taking the date and converting it to weekdays (Sunday-Saturday).

Federated Learning:

Let D1 denote the data from team member 1 (Isabel) and D2 denote the data from the team member 2(Xueting). Let Y denote the time of first pick-up the following day and X denote the total screen time for the previous day. Following is the formula we used for the linear regression:

Overall Model:

$Y_i = \beta_0 + \beta_1 * X_i + \epsilon_i$

From the principle of Simple linear regression, we could get following formula for the regression coefficients:

- $\hat{\beta}_1 = \frac{SSXY(D1 \cup D2)}{SSX(D1 \cup D2)}$
- $\hat{\beta}_0 = \overline{Y}(D1 \cup D2) - \hat{\beta}_1 * \overline{X}(D1 \cup D2)$

Overall SSXY and SSX could be obtained by the following way:

- $SSXY(D1 \cup D2) = (n_1 + n_2) * \left\{ \overline{XY}(D1 \cup D2) - \overline{X}(D1 \cup D2) * \overline{Y}(D1 \cup D2) \right\}$
- $SSX(D1 \cup D2) = (n_1 + n_2) * \left\{ \overline{X^2}(D1 \cup D2) - \overline{X}^2(D1 \cup D2) \right\}$
- $\overline{XY}(D1 \cup D2) = \frac{n_1}{n_1+n_2} * \overline{XY}(D1) + \frac{n_2}{n_1+n_2} * \overline{XY}(D2)$
- $\overline{X}(D1 \cup D2) = \frac{n_1}{n_1+n_2} * \overline{X}(D1) + \frac{n_2}{n_1+n_2} * \overline{X}(D2)$
- $\overline{Y}(D1 \cup D2) = \frac{n_1}{n_1+n_2} * \overline{Y}(D1) + \frac{n_2}{n_1+n_2} * \overline{Y}(D2)$
- $\overline{X^2}(D1 \cup D2) = \frac{n_1}{n_1+n_2} * \overline{X^2}(D1) + \frac{n_2}{n_1+n_2} * \overline{X^2}(D2)$

Also, we learned that the Standard Diviation for β_1 is defined as the following:

- $SD(\hat{\beta}_1) = \sqrt{\hat{Var}(\hat{\beta}_1)} = \sqrt{\frac{MSE(D1 \cup D2)}{SSX(D1 \cup D2)}}$
- $SD(\hat{\beta}_0) = \sqrt{\hat{Var}(\hat{\beta}_0)} = \sqrt{MSE * (\frac{1}{n} + \frac{\overline{X}^2(D1 \cup D2)}{SSX(D1 \cup D2)})}$

$MSE(D1 \cup D2)$ could be obtained by the following formula:

- $MSE(D1 \cup D2) = \frac{SSE(D1)+SSE(D2)}{n_1+n_2-2}$
- $SSE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$

To test Whether the slope (β_1) is significant from 0, we use the following t-test:

- $\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{MSE(D1 \cup D2)}{SSX(D1 \cup D2)}}} \sim t_{n_1+n_2-2}$

where $\beta_1 = 0$

Following Table 2 is the result utiliziing the method above:

From the result above, we found that the estimated $\hat{\beta}_1$ is -0.0350177, with SD=0.0793686. The calculated P value is 0.6607064 which is more than 0.05, which is non-significant. This results shows that there might be positive relationship between the first pick-up and the total screen time in the previous day but the relationship is not significant.

4

Table 1: Federal learning result

	Coefficient	SD	T-score	P-value
Intercept	484.7357	35.7674	13.5524	<0.001
Total Screen Time	-0.0350	0.0794	-0.4412	0.661

Table 2: Meta learning result

	Coefficient	SD	T-score	P-value
Intercept	442.0019	15.7599	28.0460	<0.001
Total Screen Time	0.0286	0.0330	0.8678	0.389
School Day	-41.3847	11.5500	-3.5831	0.001

Meta Learning:

Overall Model:

$Y_i = \beta_0 + \beta_1 * X_{i1} + \beta_2 * X_{i2} + \epsilon_i$

where :

- Y_i refers to First Pickup time.
- X_{i1} refers to Total Screen time in the previous day
- X_{i2} refers to School day(Which Monday to Thursday as school day and Friday to Sunday as non-school day)

First we calculate the 2 individual $\hat{\beta}$ and $\hat{Var}(\hat{\beta})$ to get the $\hat{\beta^{(1)}}$, $\hat{\beta^{(2)}}$ and $\hat{Var}(\hat{\beta^{(1)}})$, $\hat{Var}(\hat{\beta^{(2)}})$.

To get the best unbiased estimator, we need to minimum variance as the following:

$minVar[c_1\hat{\theta}_1 + c_2\hat{\theta}_2]$ and set $c_1 + c_2 = 1$. As a result, we could calculate the estimate and variance as the following:

Then the $\hat{\beta}^{meta}$ could be calculated using the following formula:

$\hat{\beta}^{meta} = \frac{\beta^{\hat{(1)}}/Var(\hat{\beta^{(1)}})+\beta^{\hat{(2)}}/Var(\hat{\beta^{(2)}})}{1/Var(\hat{\beta^{(1)}})+1/Var(\hat{\beta^{(2)}})}$

$Var(\hat{\beta^{meta}}) = \frac{1}{1/Var(\hat{\beta^{(1)}})+1/Var(\hat{\beta^{(2)}})}$

Below Table 4 is the meta learning result.

Table 3: Validate of Federal learning

		Estimate	Std. Error	t value	Pr(>|t|)
Regression					
	Intercept	484.7357	35.7674	13.5524	<0.001
	Total Screen time	-0.035	0.0794	-0.4412	0.6607
Federal learning					
	Intercept	484.7357	35.7674	13.5524	<0.001
	Total Screen time	-0.0350	0.0794	-0.4412	0.661

Table 4: Validate of Meta learning

		Estimate	Std. Error	t value	Pr(>|t|)
Regression					
	Intercept	495.3828	35.2272	14.0625	<0.001
	Total Screen time	0.0057	0.0798	0.0712	0.9434
	School Day	-52.1179	25.6112	-2.035	0.0465
Federal learning					
	Intercept	442.0019	15.7599	28.0460	<0.001
	Total Screen time	0.0286	0.0330	0.8678	0.389
	School day	-41.3847	11.5500	-3.5831	0.001

Confirmation analysis:

Confirm the analysis:

Federal learning:

The coefficients and SD calculated in the federal learning is the same as the one calculated using overall data.

Meta learning:

Results using the overall data:

As a result, the federal learning estimate and variance is the same as the one we calculated using the overall data while the meta learning result is a little bit different from the one we calculated using the overall result.

Conclusion & Discussion:

To conclude, this study shows that there is a non-significant relationship between wake-up time and total screen time both in crude analysis and after adjusting for the School days. However, the School day has a significantly associated with earlier wake-up time, which might result from the need of early class.

This might be happening because the two individuals in the team have very different patterns of screen usage. As shown in figure3, xuetao has a more constrained wake-up time and isgomez has a more spread wake-up time. This might happen because the two individuals have different schedules for class. Thus, we further do a separated linear regression for both individuals.

As shown in the Appendix 1, For isgomez, in both crude and adjusted analysis, the significant level raised a little bit for total screen time but still not significant. The school day turns to be non-significant in the adjusted analysis. For xuetao, the significant level of total screen time decrease a little bit in the unadjusted analysis and raised a little bit in the adjusted analysis. The significant level for school day raised a little bit in the adjusted analysis. This indicates that the two individual indeed have different habit of screen usage and the combined result might be unreliable because of the difference between two individuals.

Also, the non-significant relationship might because of small sample size of the study. Since each of us only have 30 observations which leads to higher variance, thus more uncertainty. Moreover, there might be other baseline covraites like class hours, and demographic variables that influence the study. Furthermore, since we are using the first pick-up time to approximate the wake-up time, this approximation could also be biased some time. Further study might be needed to adjusted for more covariates , increase the sample size and have a better approximation of wake-up time.

What was your experience of data analysis, especially in the aspect of team collaboration?

From this data analysis experience, we noticed the importance of alignment of data collecting. For example, although we have the same data frozen time, we didn't actually have the same data enrollment date. This results in different lines of data for the two individuals. As a result, we decided to keep the same date that we both have data in order to make both of our data have similar weight. In the future study, it is not only important to control the study end time but also to control study enrollment time.

Also, from this analysis, we decided to separate the analysis into data description and federal/meta learning. Both of us contribute to our strength and after communicate after we finished our parts, we both learned a lot from each other.

Table 5: Appendix 1a: Stratified analysis for isgomez

		Coef	p-value
Ajusted			
	Total Screen time(in previous day)	0.208	0.254
	School day	-79.622	0.086
Unadjusted			
	Total Screen time	0.13	0.475

Table 6: Appendix 1b: Stratified analysis for xuetao

		Coef	p-value
Ajusted			
	Total Screen time(in previous day)	0.022	0.512
	School day	-38.643	0.003
Unadjusted			
	Total Screen time	-0.008	0.832

Summarize your main contributions and findings in this project.

Xueting worked on the Federal Learning and Meta Analysis,and edited the write-up report. Isabel worked on the data description, writing and editing the report. Both teammates contribute to the written-up of the team report. This study highlighted the difference in cell-phone usage among individuals, even when they have the same occupation (student).

Acknowledgement:

Appendix:

Reference:

- [1]. Hale, L., & Guan, S. (2015). Screen time and sleep among school-aged children and adolescents: a systematic literature review. Sleep medicine reviews, 21, 50–58. https://doi.org/10.1016/j.smrv.2014.07.007

- [2]. Hiltunen P, Leppänen MH, Ray C, Määttä S, Vepsäläinen H, Koivusilta L, Sajaniemi N, Erkkola M, Roos E. Relationship between screen time and sleep among Finnish preschool