

# Affecting Factors Analysis of Adolescents' Acknowledgement of Personal Data in China in the Big Data Era

Tianjian Sun

Xi'an Tieyi High School, Xi'an City, China

**Abstract.** The swift development of information technology brings people to the big data era and put the issue of personal data privacy on the agenda. This research focus on the affective factors (gender, age, living environment, school type) of adolescents' acknowledgement of personal data privacy in China. The samples were selected from five different schools in Shaanxi province, China. Descriptive statistics are used to analyze the result from the questionnaires. T-test and multiple linear regression analysis are applied to test the hypotheses, and it is examined that hypotheses involving age, region and school type are supported by both tests.

**Keywords:** Adolescents, personal data privacy, affective factors.

## 1. Introduction

One of the key features the 21st century possesses is the swift development of information technology. This rapid growth, however, brings individuals into a different stage than ever before-the big data era.

Big data is a term that describes the large volume of data – both structured and unstructured – that inundates a business on a day-to-day basis. In 2016, it's a market which worth around \$40 billion and is projected to reach \$66.8 billion by 2021. Although this transformation appears to be beneficial, it also brings challenges that cannot be ignored. One of the biggest problem individuals face nowadays is the loss of personal data, which might lead to many unnecessary problems like garbage messages, ID card stolen, crack call and so on.

Personal data, in today's usage, is any information that relates to an identified or identifiable living individual. Well there have already been dozens of research focusing on large companies' role in dealing with the issue of data privacy, this study aims to discover the affective factors of adolescents' acknowledgement of personal data privacy in China.

This research bases its study in China primarily for two reasons: 1. The writer's identity as a high school student in China allows him to get more in touch with adolescents at that age. 2. China, unlike the E.U under the regulation of GDPR, does not yet have a authorized published law that regulates personal data. Adolescents are chosen to be the subject of the research due to their uniqueness throughout individual's life time. Classified as people aged from 12-18, adolescents are more easily tricked by the outside world and getting their personal data hacked or used illegally compared to the adults. By discovering the various factors that affect adolescents' acknowledgement of the personal data privacy, this research could contribute to developing a better protection system for data privacy in China.

## 2. Research Hypotheses

Male and female, due to the difference in genetics, show a different degree of understanding the same thing. It is generally assumed that female, especially female adolescents, has a better understanding of privacy compared to male at the same age. It is probably the damaging effect of the long-held opinion to be inferior to men that makes them terrified and focus more on themselves, in general. The acknowledgment of personal data privacy, in this case, is also believed to be different under different gender. Based on this assumption, we form the hypotheses as follows.

## **2.1 Female Adolescents, Compared to Male Adolescents, Show a Better Acknowledgement of Personal Data Privacy.**

It is also commonly known that individual's knowledge and understanding of the world increases with age. Adolescent (aged 12-18) is a particular time period where individuals refresh themselves with new understanding and knowledge on a daily basis. Personal data privacy should be on the list among the newly introduced knowledge. The hypothesis is formed under this assumption.

## **2.2 Older Adolescents (Aged 13-15) Show a Better Acknowledgement of Personal Data Privacy Compared to Younger Adolescents (Aged 16-18).**

Since education resources are allocated unevenly in China, students from urban region, compared to students living in urban region, normally are expected to have a deeper understanding of the same topic. This phenomenon could be explained by the different levels of the textbook, the natural-born intelligence, the teaching ability of their teacher and so on. The hypothesis is formed under this belief.

## **2.3 Adolescents from the Urban Region Shows a Better Acknowledgement of Personal Data Privacy Compared to Adolescents from Rural Region.**

Currently, Chinese high school have two separate departments. One department is normal high school department where students take the Gaokao exam to get into university in China. The other is the international department where students experience foreign high school environment (mainly U.S high school) and apply for foreign university. Since students from the international department would spend a large part of their life in the U.S, it is reasonable to suspect that they simulate the way U.S teenagers live. and personal data privacy is valued significantly important in the U.S. In this case, the hypotheses is formed as follow.

## **2.4 Adolescents from International Department Show a Different Degree of Acknowledgement in Terms of Personal Data Privacy Compared to Adolescents from the Normal High School Department.**

### **2.4.1 Sample Source**

The data used for this research was collected from five different schools in Shaanxi province. Two of them are from rural regions (Hanzhong city and Weinan city, respectively). The other three of them are from Urban regions in Xi'an (Xi'an GaoxinNo.1 high school, Xi'an Tiewi high school and Middle school attached to Northwestern Industry University). Students taking this research ranges from 9th grade to 11th grade. Male and female have an average ratio of 4:5 in each of these schools. Data of students from international high school is taken from Xi'an Tiewi high school's international department.

### **2.4.2 Sample Selected**

Responses classified as useless(answer every multiple questions with the same choices with A or B) are deleted to ensure the accuracy of the data collected. In this case, different hypotheses have a different number of results collected.

## **3. Variables Analyzed**

### **3.1 Independent Variable**

Gender(male or female), age(13-15 or 16-18), residential places(urban or rural) and schools(normal high school or international department) are the four independent variables in this research. The four independent variables represent the four hypotheses, and they correspond to each dependent variables to test these hypotheses.

### **3.2 Dependent Variable**

This research set out in the form of questionnaire. The questionnaire consists of twenty multiple questions. Responses of questions 6,7,11,12,15,16,17,18,19 are collected. All the responses of these questionnaires are analyzed in a way that can represent the test-taker's acknowledgement of personal data privacy. The secret is to attach different points on each responses of the particular questions and then we calculate the total points this student gets on this questionnaire. There are two separate ways of calculations.

The first type of calculations is used on question 6,11,12,16,17,19, which shows a degree of the acknowledgement of personal data privacy by giving different choices depicting different degrees.

To be specific, question 6 asks for the familiarity of personal data privacy in four separate degrees, students score 1 point by answering never heard of it and score 4 points choosing by having a deep understanding. Question 11 asks for the level of trust of public Wi-Fi and students get 2 points by choosing having high degree of trust and 1 point choosing give no trust at all. Questions 12 asks whether it is legal for government agency to use individual's personal information for personal use. Students who choose illegal score 1 point while students choose legal score 2 point. Questions 16 asks the students whether has he heard of laws and regulations related to adolescents' data privacy, and students score 3 points for choosing very familiar and 1 point for never heard of it. Questions 17 asks whether it is necessary to publish specific laws to regulate the collection of personal data. Students score 3 points by saying it is desperately needed and 1 point by answering it is not of my business. Questions 19 asks whether this student satisfies with the current policy enacted on personal data privacy. Students score 4 points being satisfied and 1 point being does not bother.

The second type of calculations, applied to question 7, 15 and 18, is based on the number of choices the students choose. For question 7, it asks students to circle the choices that they think should be included in personal data privacy, students pick 1 choices score 1 point and score 7 points for picking 7 choices. Question 15 asks which of the following behavior should be listed as an act that protects personal data privacy. Students score 4 points for choosing 4 of them and 1 point by picking 1 of these. Questions 18 asks about which of the following laws does the student think relate to personal data privacy, students score 5 points by choosing all of these and 1 point by choosing 1 of these.

## **4. Combining the Results of These Two Types of Calculations, the Total Points Each Students can Get on a Personal Data Privacy Test Come out and are Used as the Dependent Variable.**

### **4.1 Method**

This research makes use of descriptive statistics to analyze the result. STATA 13.0 is used to analyze the data in an effective manner. The final results make use of T-test and multiple linear regression as means to examine the hypothesis.

### **4.2 Descriptive Statistics**

The results from the questionnaires are collected altogether in the following chart, the data here act as a fundamental step for the next test.

Table 1. Descriptive statistics chart

Dependent variables	Average value	Minimum value	Maximum value	Standard deviation	Sample size
Points scored on the questionnaire	19.69	9	28	3.48	965
Independent variables	Representation and proportion				Sample size
Gender	1=male,54.42%;0=female,45.58%				1064
Age	1=aged 16-18,55.94%;0=aged 13-15,44.06%				1060
Region	1=urban,45.99%;0=rural,54.01%				1035
School type	1=international department,20.44%;0=normal high school,79.56%				1037

### 4.3 T-test

T-test is applied to test the hypothesis. Before all the calculations, one null hypothesis and two alternate hypotheses are developed as below:

$$H_0: \text{diff} = 0$$

$$H_1: \text{diff} \neq 0$$

T-test is applied to examine whether statistical differences exist for the two item. The whole process starts by comparing the calculated t value with the corresponding t value in the t distribution chart. If the calculated t value is greater than the t value in the t distribution chart, the null hypothesis ( $H_0: \text{diff} = 0$ ) is calculated to be less than 0.05 (in terms of the confidence interval). Then p value would be calculated to be less than 0.1. This represent that the null hypothesis is statistically rejected and statistical differences exist between the variables tested. And if p value is calculated to be more than 0.1, the hypothesis failed from a statistical standpoint according to the t test.

(All the calculations involving t value and the relevant comparison are not presented in this chart, only the results were presented).

Table 2. T-test analysis on the acknowledgement of personal data privacy

Affecting factors		Sample size	Average value	T value	P value
Gender	Male	521	19.45	-2.03	0.043
	Female	444	19.90		
Age	aged 16-18	394	21.48	-14.75	0.000
	aged 13-15	571	18.45		
Region	rural	412	21.43	-15.33	0.000
	urban	541	18.31		
School type	International department	158	20.54	-3.20	0.002
	Normal high school	798	19.51		

In the chart above, all the affecting factors listed have a p value less than 0.1, which makes all of them correct under the t-test examination. In this case, students show significant difference in the acknowledgement of personal data under different gender ( $0.043 < 0.1$ ), age ( $0.000 < 0.1$ ), region ( $0.000 < 0.1$ ), school type ( $0.000 < 0.1$ ).

### 4.4 Multiple Linear Regression

Since there are more than one independent variables tested in this research, using multiple linear regression analysis could act more effectively than the simple linear regression analysis in terms of testing these hypotheses. The principle that test the hypothesis is the same as t-test. Here are the results analyzed.

Table 3. Multiple linear regression analysis on the acknowledgement of personal data privacy (N=948)

	$\beta$	$S_{\bar{x}}$	T value	P value	90%CI
Gender(reference group: female)	0.311	0.198	1.57	0.118	-0.016~0.637
Age(reference group: aged 13-15)	1.520	0.303	5.02	0.000	1.021~2.018
Region(reference group: rural)	2.502	0.308	8.13	0.000	1.995~3.009
School type(reference group: normal high school)	-1.483	0.306	-4.48	0.000	-1.988~-0.978
Intercept	18.033	0.168	107.35	0.000	17.757~18.310

As it is shown in the chart, students does not show a significant differences in the acknowledgement of personal data privacy in terms of different genders( $0.118 > 0.1$ ), but students' acknowledgement of personal data privacy show a significant differences in terms of age, school type and region.( $0.000 < 0.1$ ).

## 5. Conclusion

Personal data privacy, as the product of the rapid growth of information technology, relates to adolescents' life to a large extent. By enhancing adolescent's acknowledgement of personal data privacy, they can have a larger probability to avoid from unnecessary problems. It is suggested that age, gender, region, school type are some of the leading factors that correlates when it comes to human's acknowledgement of new knowledge. These four factors are chosen as the independent variables for this research.

T-test and multiple linear regression are applied to test the hypotheses. The results of t-test portray that all four independent variables result in significant difference for the dependent variables. But multiple linear regression line reviews that the hypotheses involving age, region, school type proved to be correct and gender does not meet the standard in terms of creating a significant difference (statistically).

Personal data privacy, derives as the arisen of the big data era, present inevitable problems that the society must face. Well it is true that China does not yet have an authorized personal data protection law, relevant research like this could provide valuable insight to speed up the process. The acknowledgement of personal data privacy is still on the agenda that requires significant attention paid, especially for adolescents.

## Acknowledgements

Thank you for the guidance of Professor Quan Xiaojuan of Northwest Polytechnic University, and for the support of project team teachers and students of Northwest Polytechnic University and Xi'an Tieyi High School.

## References

- [1]. Trunomi, n.d. The EU General Data Protection Regulation (GDPR) is the most important change in data privacy regulation in 20 years. [Online] Available at: <https://eugdpr.org/>.
- [2]. Anon., 2016. what is personal data. [Online] Available at: [https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data\\_en](https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en).
- [3]. Campbell, M., n.d. What Are the Consequences of Data Loss?.
- [4]. Available at: <https://www.unitrends.com/blog/what-are-the-consequences-of-data-loss>.

- [5]. Nagrath, C., 2011. What Makes a School International? Available at: [https:// tieonline. com/ view\\_ article.cfm?articleid=87](https://tieonline.com/view_article.cfm?articleid=87).
- [6]. SAS, 2017. Big Data. Available at: [https://www.sas.com/en\\_us/insights/big-data/what-is-big-data.html#dmworld](https://www.sas.com/en_us/insights/big-data/what-is-big-data.html#dmworld).
- [7]. Techopedia, n.d. Information privacy. Available at: [https:// www. techopedia. Com /definition / 10380/information-privacy](https://www.techopedia.com/definition/10380/information-privacy).
- [8]. Zheng, H., 2016. National people's congress proposes the publication of personal data privacy. Available at: [http://www.xinhuanet.com/politics/2015-03/02/c\\_127534032.htm](http://www.xinhuanet.com/politics/2015-03/02/c_127534032.htm).