An Investigation of Twitter Users

Who Disclosed Their Personal Profile Items in Their Tweets Honestly

Yasuhiko Watanabe, Hiromu Nishimura, Yuuya Chikuki, Kunihiro Nakajima, and Yoshihiro Okada
Ryukoku University
Seta, Otsu, Shiga, Japan
Email: watanabe@rins.ryukoku.ac.jp, t160405@mail.ryukoku.ac.jp, t160389@mail.ryukoku.ac.jp,
nakajima.k216@gmail.com, okada@rins.ryukoku.ac.jp

***Abstract*—These days, many people use a Social Networking Service (SNS). Most SNS users are careful in protecting the privacy of personal information: name, age, gender, address, telephone number, birthday, etc. However, some SNS users disclose their personal information that can threaten their privacy and security even if they use non-real name accounts. In this study, we investigated tweets disclosing submitters' personal profile items which many of us think are not true. We collected 565 tweets where submitters used non-real name accounts and made promises to disclose their personal profile items, surveyed the details of their personal profile items disclosed by themselves, especially their ages, genders, and heights, and analyzed them statistically, to be specific, applied the Shapiro-Wilk test of normality and the Welch's test to them. The results of these tests showed that most of the submitters disclosed their ages, genders, and heights honestly.**

Keywords–personal information; Twitter; SNS; privacy risk; Shapiro-Wilk test of normality; Welch's test.

I. INTRODUCTION

These days, many people use a Social Networking Service (SNS) to communicate with each other and try to enlarge their circle of friends. SNS users are generally concerned about potential privacy risks [1]. To be specific, they are afraid that unwanted audiences will obtain information about them or their families, such as where they live, work, and play. As a result, SNS users are generally careful in disclosing their personal information. They disclose their personal information only when they think the benefits of doing it is greater than the potential privacy risks. However, some SNS users, especially young users, disclose personal information on their profiles, for example, real full name, gender, hometown and full date of birth, which can potentially be used to identify details of their real life, such as their social security numbers. In order to discuss this phenomenon, many researchers investigated how much and which type of information are disclosed in SNSs, especially, in Facebook [2] [3]. Researchers might think that personal information disclosed in Facebook is reliable, or it is possible to check whether personal information disclosed in Facebook is true. This is because

- Facebook users are required to register and disclose their real names when they first start using Facebook.
- Facebook users would be criticized by their friends if they disclose their information dishonestly.

On the other hand, a small number of researchers investigated how much and which type of information disclosed by non-real name account users, such as Twitter users. Researchers might think that personal information disclosed by non-real name account users is unreliable. This is because

- nobody criticizes non-real name account users when they disclose their personal information dishonestly.
- true personal information can threaten their privacy and security even if they use non-real name accounts.

As a result, many of us think that it is natural for non-real name account users not to disclose their personal information honestly. Figure 1 shows tweets submitted by non-real name account user, *Rina*. In these tweets, *Rina* disclosed her personal profile items: her age, gender, birthday, zodiac sign, and height. Many of us think that these personal profile items were not true. However, we do not check whether *Rina* disclosed her personal profile items honestly because it is difficult to do it. In this paper, we collect tweets where non-real name account users disclosed their personal profile items, analyze them statistically, and show that it is likely that most of the non-real name account users, especially young users, disclosed their

Figure 1. A non-real name account user, *Rina*, disclosed her personal profile items in her tweets.

 ISBN: 978-1-61208-800-6

Figure 2. A tweet promising to disclose the same number of submitters' personal profile items as likes to it.

personal information honestly.

The rest of this paper is organized as follows: In Section II, we survey the related works. In Section III, we show how to collect tweets disclosing submitters' personal profile items. In Section IV, we survey the details of submitters' personal profile items, analyze them statistically, and show that it is likely that most of the submitters disclosed their personal profile items honestly. Finally, in Section V, we present our conclusions.

II. Related work

Personally identifiable information is defined as information which can be used to distinguish or trace an individual's identity such as social security number, biometric records, etc. alone, or when combined with other information that is linkable to a specific individual, such as date and place of birth, mother's maiden name, etc. [4] [5]. Internet users are generally concerned about unwanted audiences obtaining personal information. Fox et al. reported that 86% of Internet users are concerned that unwanted audiences will obtain information about them or their families [1]. Also, Acquisti and Gross reported that students expressed high levels of concern for general privacy issues on Facebook, such as a stranger finding out where they live and the location and schedule of their classes, and a stranger learning their sexual orientation, name of their current partner, and their political affiliations [2]. However, Internet users, especially young users, tend to disclose personal information on their profiles, for example, real full name, gender, hometown and full date of birth, which can potentially be used to identify details of their real life, such as their social security numbers. As a result, many researchers discussed the reasons why young users willingly disclose personal information on their SNS profiles. Dwyer concluded in her research that privacy is often not expected or undefined in SNSs [6]. Barnes argues that Internet users, especially teenagers, are not aware of the nature of the Internet and SNSs [3]. Hirai reported that many users had troubles in SNSs because they did not mind that strangers observed their communication with their friends [7]. Viseu et al. reported that many online users believe the benefits of disclosing personal information in order to use an Internet site is greater than the potential privacy risks [8]. On the other hand, Acquisti and Gross explain this phenomenon as a disconnection between the users' desire to protect their privacy and their actual behavior [2]. Also, Livingstone points out that teenagers' conception of privacy does not match the privacy settings of most SNSs [9]. Joinson et al. reported that trust and perceived privacy had a strong affect on individuals' willingness to disclose personal information to a website [10]. Also, Tufekci found that concern about unwanted audiences had an impact on whether or not students revealed their real names and religious affiliation on MySpace and Facebook [11]. The authors also think that most students are seriously concerned about their privacy and security. However, they often underestimate the risk of their online messages and submit them. For example, Watanabe et al. reported that many students submit tweets concerning school events and these tweets may give a chance to other people, including unwanted audiences, to distinguish which schools students go to [12].

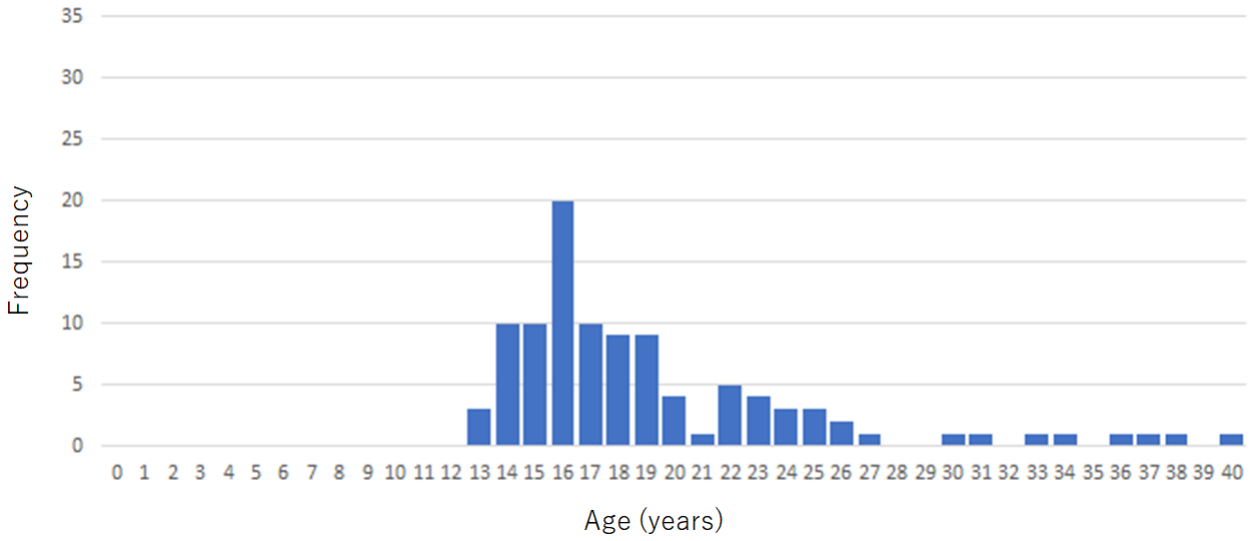
III. A collection of tweets disclosing submitters' personal profile items

It is difficult to collect tweets disclosing submitters' personal profile items, such as tweets in Figure 1, directly. To solve this problem, we focused on tweets where submitters promised their followers to disclose the same number of their own personal profile items as likes to their tweets. Figure 2 shows a tweet submitted by *Rina* on September 3, 2019. In this tweet, *Rina* promised her followers to disclose the same number of her personal profile items as likes to her tweet. Actually, *Rina* submitted 35 replies disclosing her personal profile items to her tweet shown in Figure 2 from September 3 to 9, 2019. The five tweets shown in Figure 1 were the first five replies submitted by *Rina* to her tweets shown in Figure 2. As of November 20, 2019, we confirmed that 37 likes were given to her tweet shown in Figure 2. Furthermore, we found many tweets promising to disclose the same number of their own personal profile items as likes to their tweets. As a result, it is easy to collect tweets disclosing submitters' personal profile items when we collect tweets promising to disclose submitters' personal profile items. The reasons why many Twitter users submitted tweets promising to disclose submitters' personal profile items might be

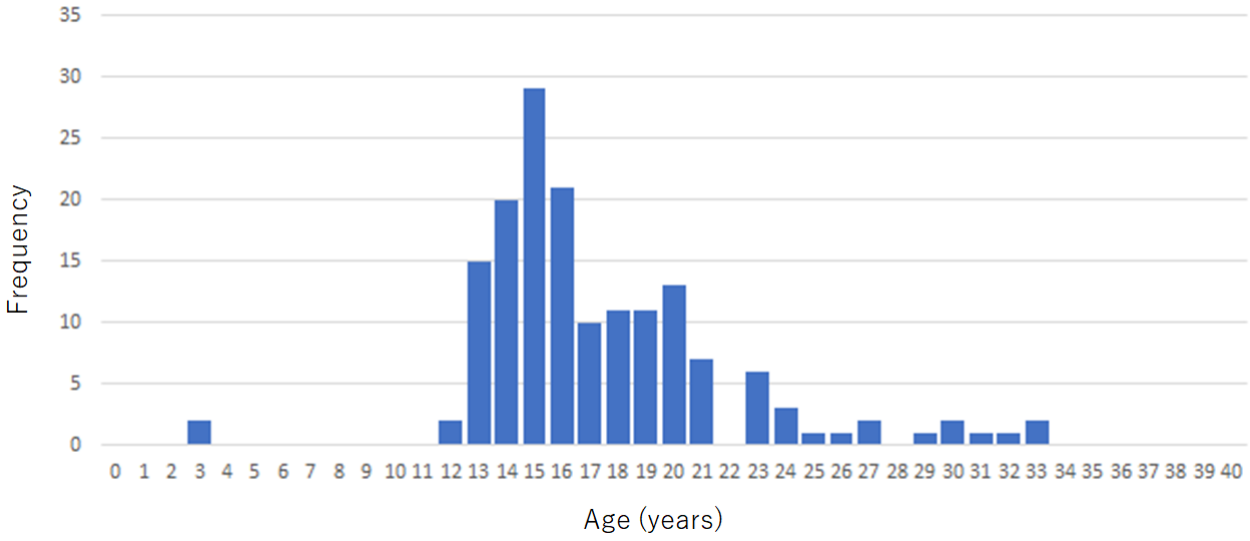
- they thought they looked fun,
- they wanted to draw attention, and
- they wanted to know how much attention was paid to their tweets.

In order to collect tweets promising to disclose submitters' personal profile items, we focused on images attached to these tweets. This is because many submitters attached the same image to their tweets and many personal profile items were listed in the image. As shown in Figure 2, *Rina* attached an image to her tweet and showed the list of personal profile items that she promised her followers to disclose in the image. Many twitter users attached the same image to their tweets promising to disclose their personal profile items. As a result, we used these shared images as key to collect tweets promising to disclose submitters' personal profile items. To be specific, we

 ISBN: 978-1-61208-800-6



(a) The number of submitters who disclosed that they were men by age.



(b) The number of submitters who disclosed that they were women by age.

Figure 3. The number of submitters who disclosed their genders clearly by age.

collected these tweets by using Twigaten [13]. Twigaten helps us to collect tweets to which the same image is attached. By using Twigaten, we collected 565 Japanese tweets promising to disclose submitters' personal profile items on November 20, 2019. The obtained tweets were submitted from October 3, 2018 to November 20, 2019.

IV. An Analysis of Tweets Disclosing Submitters' Personal Profile Items

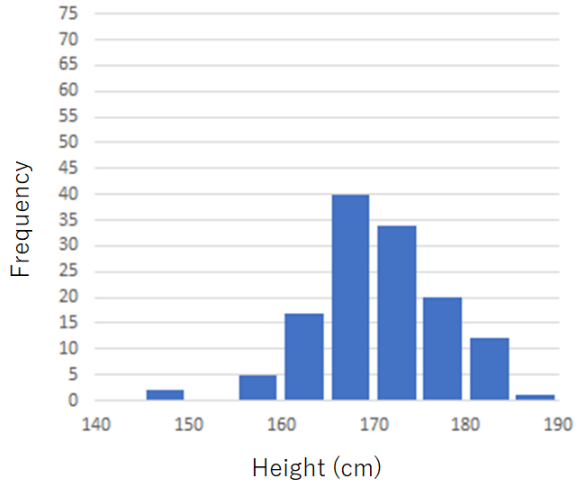
It is difficult to determine whether an individual submitter disclosed his/her personal profile items honestly. For example, it is difficult to determine whether *Rina*, who submitted tweets in Figure 1 and Figure 2, was a woman. In this study, we discuss whether submitters disclosed their personal profile items honestly when they made promises to disclose them. In order to discuss this problem, we analyze submitters' genders, ages, and heights statistically.

A. Submitters' genders

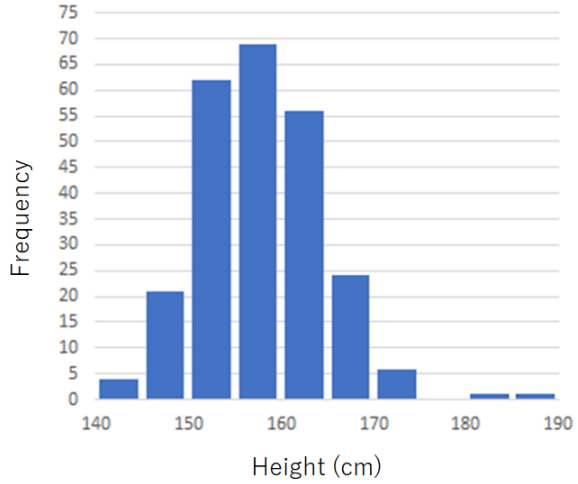
As mentioned in Section III, we obtained the 565 tweets promising to disclose their personal profile items. We surveyed these 565 tweets and their replies and, according to submitters' genders disclosed in the replies, classified them into

- 282 tweets (women)
- 156 tweets (men)
- 27 tweets (unclear)
- 100 tweets (no replies)

 ISBN: 978-1-61208-800-6



(a) the histogram of submitters' heights (disclosed genders: men).



(b) the histogram of submitters' heights (disclosed genders: women).

Figure 4. The histogram of heights of submitters who disclosed their genders, men or women, clearly. (bin width = 5cm)

B. *Submitters' ages*

We also surveyed the 565 tweets and their replies and, according to whether submitters' ages were disclosed in their replies clearly, classified them into

- 276 tweets (clearly)
- 60 tweets (unclearly)
- 229 tweets (no replies)

When submitter's age was disclosed such as "early 20s" and "over thirty", we determined that submitter's age was disclosed unclearly. Among the 276 tweets where submitters' ages were disclosed clearly, we found 102 and 161 tweets where submitters' genders were also disclosed clearly, men and women, respectively. Figure 3 shows the number of submitters, who disclosed their genders clearly, men and women, by age. As shown in Figure 3, the most popular age of men and women were 16 and 15 years old, respectively.

TABLE I. THE RESULTS OF THE SHAPIRO-WILK TEST OF NORMALITY

gender	age	sample size	W value	p-value
men	15	10	0.885	0.147
men	16	18	0.929	0.190
men	17	9	0.977	0.946
women	14	17	0.933	0.244
women	15	24	0.971	0.697
women	16	19	0.961	0.587

C. *Submitters' heights*

We also surveyed the 565 tweets and their replies and, according to whether submitters' heights were disclosed in their replies clearly, classified them into

- 401 tweets (clearly),
- 8 tweets (unclearly), and
- 156 tweets (no replies).

Among the 401 tweets where submitters' heights were disclosed clearly, we found 131 and 244 tweets where submitters' genders were disclosed clearly, men and women, respectively. Figure 4 shows the histogram of heights of submitters who disclosed their genders, men or women, clearly.

It is difficult to determine whether an individual submitter disclosed his/her personal profile items honestly. In this study, we statistically examine whether submitters disclosed their personal profile items honestly when they made promises to disclose their personal profile items and disclosed them in the same way as *Rina* did.

It is well known that our heights follow a normal (Gaussian) distribution [14]. As a result, if most of submitters disclose their ages, genders, and heights honestly, their heights would follow a normal distribution. Also, the average of their heights would be equal to the national average height in Japan. To solve this problem, in this paper, we conduct the statistical analysis on

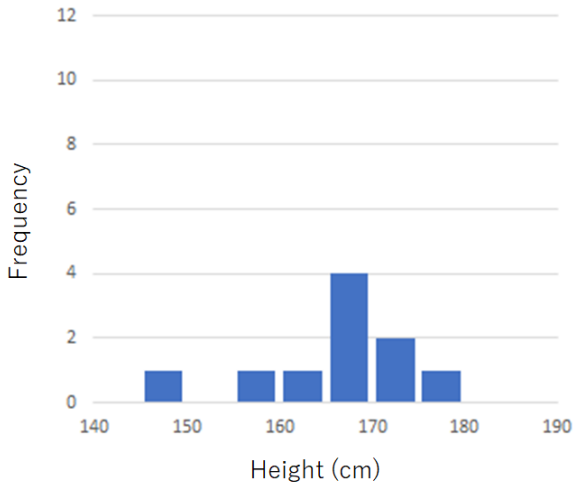
- 37 submitters who disclosed their genders (men), ages (15-17 years old), and heights clearly, and
- 60 submitters who disclosed their genders (women), ages (14-16 years old), and heights clearly.

As shown in Figure 3, men aged 15-17 and women aged 14-17 were the most popular segments in the submitters' ages.

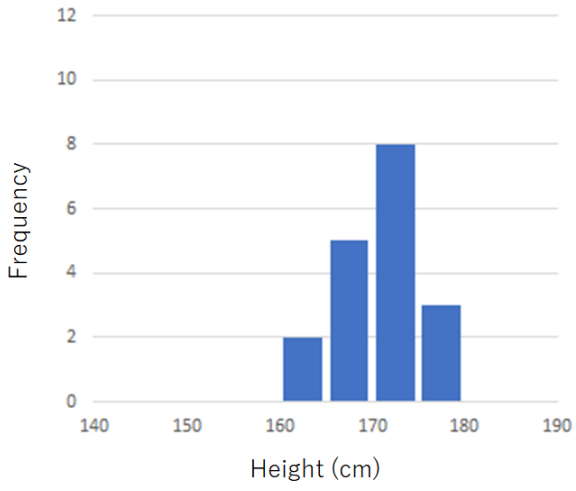
First, we discuss whether submitters' heights followed a normal distribution. Figure 5 shows the histograms of their heights. In order to discuss whether submitters' heights followed a normal distribution, we conducted the Shapiro-Wilk test of normality. The null hypothesis in this study was that submitters' heights followed a normal distribution. Table I shows the results of the Shapiro-Wilk test of normality. As shown in Table I, the p-value in each case was greater than 0.05. As a result, the null hypothesis in each case was not rejected. In other words, submitters' heights, in each case of men aged 15-17 and women aged 14-16, followed a normal distribution.

Next, we discuss whether the average of submitters' heights was equal to the national average height in Japan. Table II shows the average of submitters' heights. Table III shows the national average height in Japan [15]. In order to discuss

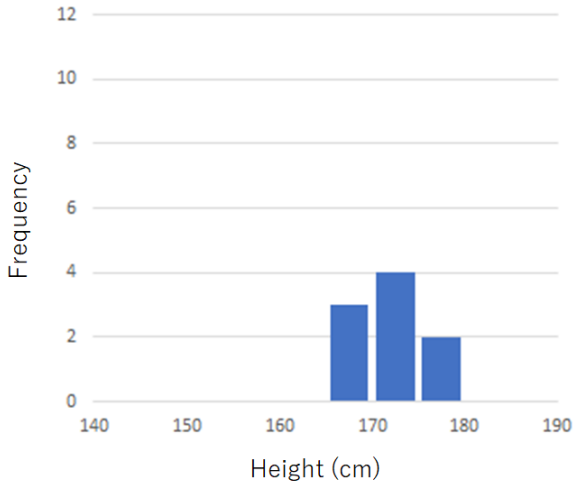
 ISBN: 978-1-61208-800-6



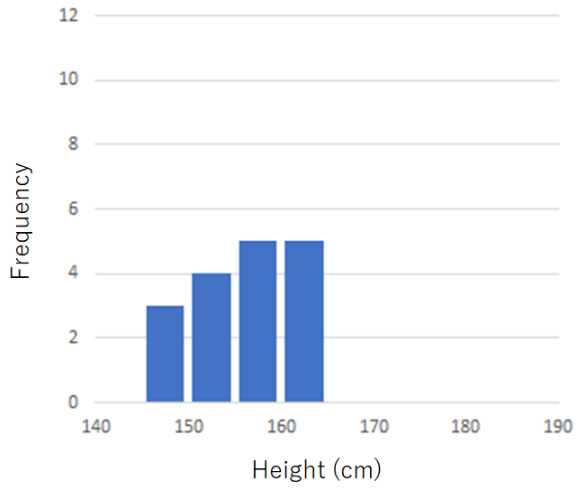
(a) men aged 15



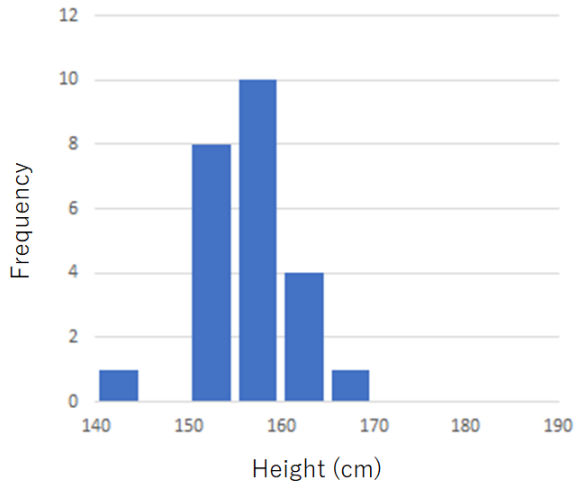
(b) men aged 16



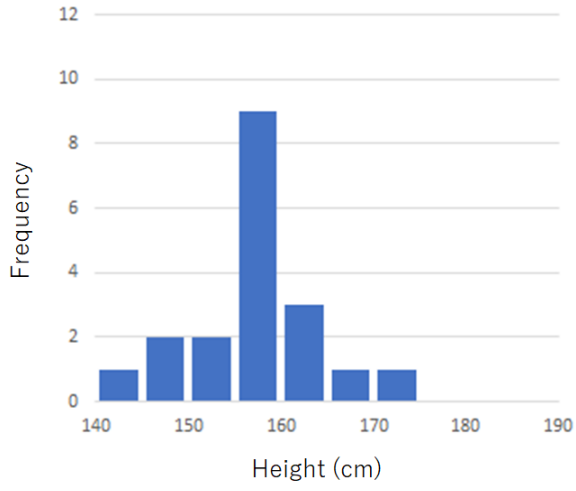
(c) men aged 17



(d) women aged 14



(e) women aged 15



(f) women aged 16

Figure 5. The histograms of heights of submitters who disclosed that they were men aged 15-17 and women aged 14-16 (bin width = 5cm).

 ISBN: 978-1-61208-800-6

TABLE II. THE AVERAGE AND STANDARD DEVIATION OF SUBMITTERS' HEIGHTS

gender	age	sample size	average	standard deviation
men	15	10	165.5	7.58
men	16	18	169.2	4.08
men	17	9	171.3	3.84
women	14	17	155.0	6.11
women	15	24	155.7	4.99
women	16	19	156.9	6.54

TABLE III. THE NATIONAL AVERAGE AND STANDARD DEVIATION OF HEIGHTS IN JAPAN

gender	age	sample size	average	standard deviation
men	15	1411	168.37	5.75
men	16	1428	169.59	5.70
men	17	1427	170.46	5.82
women	14	1386	156.36	5.24
women	15	1413	156.76	5.36
women	16	1419	157.16	5.17

TABLE IV. THE RESULTS OF WELCH'S TEST

gender	age	Degrees of freedom	test statistic T	p-value
men	15	9.07	1.195	0.262
men	16	17.84	0.380	0.708
men	17	8.23	-0.675	0.518
women	14	16.29	0.914	0.374
women	15	23.91	1.060	0.300
women	16	18.30	0.179	0.860

whether the average of their heights was equal to the national average height in Japan, we conducted the Welch's test. The null hypothesis in this study was that the average of submitters' heights was equal to the national average height in Japan. Table IV shows the results of the Welch's test. As shown in Table IV, the p-value in each case was greater than 0.05. As a result, the null hypothesis in each case was not rejected. In other words, in each case of men aged 15-17 and women aged 14-16, the average of submitters' heights was equal to the national average height in Japan.

The results of the Shapiro-Wilk test of normality and the Welch's test rarely happened when many submitters disclosed their ages, genders, and heights dishonestly. As a result, it is assumed that most of the submitters disclosed their ages, genders, and heights honestly. Furthermore, age, gender, and height were important personal information. It is likely that they disclosed not only their ages, genders, and heights but also other personal profile items honestly.

V. CONCLUSION

In this paper, we investigated tweets disclosing submitters' personal profile items and analyzed submitters' ages, genders, and heights statistically. The results of the statistical analysis showed that it is likely that most of the submitters disclosed their personal profile items honestly. These personal profile items can threaten their privacy and security even if they use non-real name accounts. We are investigating whether submitters were concerned about their privacy and security risks caused by submitting tweets disclosing their personal profile items honestly. Furthermore, we intend to conduct the same statistical analysis on tweets in languages other than Japanese.

REFERENCES

[1] S. Fox et al., Trust and Privacy Online: Why Americans Want to Rewrite the Rules, The Pew Internet & American Life Project, 2000. [Online]. Available: http://www.pewinternet.org/2000/08/20/trust-and-privacy-online/ [accessed: 2020-09-01]

[2] A. Acquisti and R. Gross, Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 36–58.

[3] S. B. Barnes, "A privacy paradox: Social networking in the United States." First Monday, vol. 11, no. 9, 2006. [Online]. Available: http://firstmonday.org/article/view/1394/1312 [accessed: 2020-09-01]

[4] C. Johnson III, Safeguarding against and responding to the breach of personally identifiable information, Office of Management and Budget Memorandum, 2007. [Online]. Available: http://www.whitehouse.gov/omb/memoranda/fy2007/m07-16.pdf [accessed: 2016-10-04]

[5] B. Krishnamurthy and C. E. Wills, "On the leakage of personally identifiable information via online social networks," Computer Communication Review, vol. 40, no. 1, 2010, pp. 112–117. [Online]. Available: https://doi.org/10.1145/1672308.1672328 [accessed: 2020-09-01]

[6] C. Dwyer, "Digital relationships in the "myspace" generation: Results from a qualitative study," in Proceedings of the 40th Annual Hawaii International Conference on System Sciences, ser. HICSS '07. Washington, DC, USA: IEEE Computer Society, 2007, p. 19.

[7] T. Hirai, "Why does "Enjyo" happen on the Web? : An Examination based on Japanese Web Culture," Journal of Information and Communication Research, vol. 29, no. 4, mar 2012, pp. 61–71. [Online]. Available: http://doi.org/10.11430/jsicr.29.4_61 [accessed: 2020-09-01]

[8] A. Viseu, A. Clement, and J. Aspinall, "Situating privacy online: Complex perception and everyday practices," Information, Communication & Society, 2004, pp. 92–114.

[9] S. Livingstone, "Taking risky opportunities in youthful content creation: teenagers' use of social networking sites for intimacy, privacy and self-expression." New Media & Society, vol. 10, no. 3, 2008, pp. 393–411.

[10] A. N. Joinson, U.-D. Reips, T. Buchanan, and C. B. P. Schofield, "Privacy, trust, and self-disclosure online." Human-Computer Interaction, vol. 25, no. 1, 2010, pp. 1–24. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/07370020903586662 [accessed: 2020-09-01]

[11] Z. Tufekci, "Can You See Me Now? Audience and Disclosure Regulation in Online Social Network Sites," Bulletin of Science, Technology & Society, vol. 28, no. 1, 2008, pp. 20–36.

[12] Y. Watanabe, H. Onishi, R. Nishimura, and Y. Okada, "Detection of school foundation day tweets that can be used to distinguish senders' schools," in Proceedings of the Eleventh International Conference on Evolving Internet (INTERNET 2019), Nov 2016, pp. 30–35. [Online]. Available: https://www.thinkmind.org/index.php?view=article&articleid=internet_2019_2_30_40026 [accessed: 2020-09-01]

[13] twigaten.204504byse.info. TwiGaTen. [Online]. Available: https://twigaten.204504byse.info/ [accessed: 2020-09-01]

[14] National Centre for Research Methods (NCRM). Using Statistical Regression Methods in Education Research. [Online]. Available: http://www.restore.ac.uk/srme/www/fac/soc/wie/research-new/srme/index.html [accessed: 2020-09-01]

[15] the Ministry of Education, Culture, Sports, Science and Technology (MEXT). the survey on physical strength and sporting ability (2018). [Online]. Available: https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00402102&tstat=000001088875&cycle=0&tclass1=000001133904&stat_infid=000031872003 [accessed: 2020-09-01]

 ISBN: 978-1-61208-800-6