Theory and Practice of Science and Technology
2025, VOL. 6, NO. 6, 59-62
DOI: 10.47297/taposatWSP2633-456912.20250606

Application of Investigation and Estimation Method for Sensitive Issues Based on Privacy Protection in Company Management

Ming Chen

School of General Education, Hubei College of Chinese Medicine, Jingzhou, Hubei, 434020, China

ABSTRACT
Due to the concerns of the respondents, it is difficult to obtain effective data for the investigation of sensitive issues. A method that can obtain valid data while protecting the privacy of the respondents is proposed. Using this method to establish a mathematical model, estimation formulas and accuracy related to sensitive issues are provided. The romantic relationships and procrastinating or shirking work among colleagues are selected as examples to estimate and analysis. The estimation accuracy can be adjusted by selecting appropriate parameters. This method and mathematical model can be extended and transplanted to investigate other socially sensitive issues.

KEYWORDS
Sensitive issues; Protecting privacy; Survey method; Company management

1 Introduction

Sensitive issues usually refer to those that may cause controversy, touch on sensitive social points, or involve personal privacy. These issues often attract widespread public attention and discussion due to their impact on the interests, values, or emotions of certain groups. Corresponding methods of investigation on sensitive issues are emerging [1-6]. There are many practical applications of survey methods on sensitive issues in education, healthcare, and social psychology [7-9]. The sensitive issues of managing employees in a company mainly include how to deal with their mood swings, sensitive personalities, and how to build an inclusive and supportive work environment. It includes salary issues, romantic relationships, personal emotions, work procrastination, and leadership questioning. These sensitive issues are a common phenomenon in companies, and a certain degree of quantitative analysis is needed to address this behavior in order to utilize company management. With the development of society, company employees pay more attention to the protection of their privacy rights at work, so research on sensitive issues related to employee management gradually increase and become a hot topic. However, as a sensitive issue, if these behaviors are investigated, respondents generally have concerns. Even if anonymous surveys are used, respondents are likely to avoid or intentionally give incorrect answers, which will greatly reduce the authenticity of the survey data [10]. Therefore, it is necessary to design a survey method that can both improve the response rate and reduce the false response rate. in this paper, the survey method proposed has two basic starting points. Firstly, the respondents randomly choose to answer one of several sensitive questions. Secondly, the investigator does not know which question the respondent is answering. This can protect the personal privacy of the respondents and enable them to answer the selected questions truthfully.

2 Survey model for sensitive issues

2.1 Survey method for sensitive issues

In the implementation aspect of sensitivity survey, sensitive issues cannot be separated from the collection of information in the real lives of respondents. The intrusive, dangerous and stigmatizing nature of sensitive issues can trigger a stressful reaction in respondents. Therefore, it is also important to choose the appropriate method of data collection. Generally speaking, the stress reaction of respondents is characterized by vulnerability, rational calculation and destructiveness. So choosing the right method can play a certain role in alleviating it. Generally, the following three principles are followed: questionnaire that avoids the vulnerability of the respondents, interview that downplays the rational calculation instinct of the respondents, and observation that reduces the destructive nature of the respondents' stress.

Firstly, several identical cards are prepared with1,2,3,...*m*printed on each card. The proportion of cards with the number*k*printed on them isP_k ($k = 1,2,3,\cdots,m$), and$P_k$is not exactly the same. Then the respondent randomly draws a card, and it is easy to know that the probability of drawing the number*k*isP_k. If the respondents have had this sensitive behavior before, they answer the difference between$m + 1$and the drawn number; on the contrary, if the respondents did not have this sensitive behavior before, they answer the drawn number.

2.2 Estimation and analysis of sensitive issues

For the purpose of mathematical analysis, the variable $x_i\ (i = 1,2,3,\cdots,n)$ is introduced and let the

$$x_i = \begin{cases} m+1 \text{The } i\text{th respondent has engaged in sensitive behavior,} \\ 0 \text{The } i\text{th respondent did not have engaged in sensitive behavior.} \end{cases}$$

Let $\mu = P\left(x_i = m+1\right)$ be the probability that the respondents have had sensitive behavior. Assuming the variable y_i is the number drawn by the ith respondent on the card, it is easy to know that $y_i\ (i = 1,2,3,\cdots,n)$ is independent, identically distributed, and independent of x_i. According to the survey method, the number answered by the ith respondent is $r_i = |y_i - x_i|$, and it is easy to know that $1 \le r_i \le m$. Let n_i be the number of respondents who answer the number k. It is easy to obtain $n_1 + n_2 + \cdots + n_m = n$. Obviously when $r_i = 1$, i. e., the number answered is 1, then the following formula is obtained from the full probability formula.

$$P(r_i = 1) = \left(1-\mu\right)P_1 + \mu P_{m+1-1} \tag{1}$$

By extension, when $r_i = k$, we can obtain

$$P(r_i = k) = \left(1-\mu\right)P_k + \mu P_{m+1-k}. \tag{2}$$

Then the mathematical expectation of r_i is

$$E(r_i) = \sum_{k=1}^{m} kP(r_i = k) = \sum_{k=1}^{m} kP_k + \mu\left(m+1-2\sum_{k=1}^{m} kP_k\right). \tag{3}$$

Noting that $\sum_{k=1}^{m} kP_k = E(y_i)$, equation (3) becomes

$$E(r_i) = \mu(m+1-2E(y_i)) + E(y_i). \tag{4}$$

Further the variance of r_i is

$$D(r_i) = D(y_i) + \mu(1-\mu)(m+1-2E(y_i))^2. \tag{5}$$

Let the mean value of the numbers $r_1, r_2, \cdots, r_n$ answered by n respondents be $\bar{r} = \frac{1}{n}\sum_{i=1}^{m} r_i = \sum_{k=1}^{m} \frac{n_k}{n} k$, here we use $\bar{r}$ to estimate $E(r)$. By substituting it into equation (3), we can obtain

$$\hat{\mu} = \frac{\bar{r} - E(y_i)}{m+1-2E(y_i)}, m+1-2E(y_i) \neq 0. \tag{6}$$

It is not difficult to calculate $E(\hat{\mu}) = \mu$, i.e., $\hat{\mu}$ is an unbiased estimate of μ. The variance of $\hat{\mu}$ is further calculated as

$$D(\hat{\mu}) = \frac{D(\bar{r})}{(m+1-2E(y_i))^2} = \frac{\mu(1-\mu)}{n} + \frac{D(y_i)}{n(m+1-2E(y_i))^2}, \tag{7}$$

where $D(\bar{r}) = \frac{1}{n}\sum_{k=1}^{m}(k-\bar{r})^2\frac{n_k}{n}$, and let $\mathrm{A} = \frac{\mu(1-\mu)}{n}$, $\mathrm{B} = \frac{D(y_i)}{n(m+1-2E(y_i))^2}$.

Here, A is the variance generated when sensitive issues are directly surveyed, while B is the variance generated when respondents randomly select issues.

3 Application examples

3.1 Survey for office romance

Office romance is always a controversial topic in the modern workplace. Many companies explicitly prohibit romantic relationships between employees, especially when such relationships involve superiors, subordinates, or within the same department. Most companies prohibit office romance as they may bring negative impacts such as decreased work efficiency, damaged team atmosphere, conflicts of interest, damaged company image, and legal risks. Although most companies prohibit office romances, romantic relationships can still occur in the workplace. By establishing clear rules and regulations, strengthening employee education, establishing complaint mechanisms, providing psychological support, and encouraging employee self-discipline, companies can create a healthier and more positive work environment for promoting employee growth and development. In order to effectively manage the internal affairs of the company and optimize relationships among employees, it is necessary to conduct a preliminary investigation into the romantic relationships of employees within the company. Therefore, the previously mentioned variable x_i can be changed to

$$x_i = \begin{cases} m+1 \text{The } i\text{th respondent has had a romantic relationship with colleagues} \\ \text{in the company,} \\ 0 \text{The } i\text{th respondent did not have a romantic relationship with colleagues} \\ \text{in the company.} \end{cases}$$

Of course, the romantic status includes both past and present.

Assume that the total number of respondents is$n = 250$ and$m = 5$, i.e., the numbers on the cards are$k = 1,2,...,5$on the card. The corresponding percentage of cards in the total number are$P_1 = 0.4$, $P_2 = 0.3$, $P_3 = 0.25$, $P_4 = 0.03$, $P_5 = 0.02$. The number of respondents answering the numbers1,2,...,5are$n_1 = 75, n_2 = 71, n_3 = 69$, $n_4 = 22, n_5 = 13$. The calculations are as follows.

$$E(y_i) = \sum_{k=1}^{5} kp_k = 1.970, D(y_i) = \sum_{k=1}^{5} (k - E(y_i))^2 p_k = 0.949100,$$

$$\bar{r} = \sum_{k=1}^{5} \frac{n_k}{n} k = 2.308, D(\bar{r}) = \frac{1}{n} \sum_{k=1}^{5} (k - \bar{r})^2 \frac{n_k}{n} = 0.005205.$$

By substituting them into equations (6) and (7), it can be seen that the estimated value of the proportion with office romance is

$$\hat{\mu} = \frac{2.308 - 1.970}{5 + 1 - 2 \times 1.970} = 0.1641.$$

The corresponding variance and standard deviation are respectively

$$D(\hat{\mu}) = \frac{0.005205}{(5 + 1 - 2 \times 1.970)^2} = 0.001226$$

and

$$\sqrt{D(\hat{\mu})} = \sqrt{0.001226} = 0.035021.$$

Further calculations show that A = 0.000332, B = 0.000895. that is to say, the standard deviation generated by the random selection of cards by the respondents is 2.99%, while the variance generated by the issue of office romance itself is 1.822%.

If the precision is set to 1x standard deviation, the percentage of employees who had an office romance is 16.41± 3.50%; if the precision is set to 2x standard deviation, the percentage of employees who had an office romance is 16.41± 7.00%. According to the above accuracy standards, this indicates that the range of the number of employees who have had office romances is 32-50 and 24-59, respectively.

3.2 Survey for procrastinating or shirking work

In the workplace, it is common to encounter employees who procrastinate or shirk responsibility for their work. This can be manifested in three situations: procrastinating execution time, shifting responsibility to others, and evading responsibility. These have consequences such as affecting team efficiency, reducing work quality, and negatively impacting the team's collaborative atmosphere. The behavior of procrastinating or shirking work is usually manifested by using subtle or hidden means. For these reason, in order to improve work efficiency, strengthen the responsibility of employee responsibilities and enhance team cohesion, it is necessary to conduct a survey on whether employees have engaged in any behavior of procrastinating or shirking work. Therefore, the previously mentioned variable$x_i$can be changed to

$$x_i = \begin{cases} m + 1 \text{The} i \text{threspondenthasprocrastinatedorshirkedworkresponsibilities}, \\ 0 \text{The} i \text{threspondentdidnothaveprocrastinatedorshirkedworkresponsibilities}. \end{cases}$$

Of course, the work state here includes both past and present.

Assume that the total number of respondents is$n = 300$and$m = 6$, i.e., the numbers on the cards are$k = 1,2,...,6$on the card. The corresponding percentage of cards in the total number are$P_1 = 0.5, P_2 = 0.3, P_3 = 0.1, P_4 = 0.03, P_5 = 0.04, P_6 = 0.03$. The number of respondents answering the numbers1,2,...,6 are$n_1 = 111$, $n_2 = 73$, $n_3 = 61$, $n_4 = 22$, $n_5 = 18$, $n_6 = 15$. The calculations are as follows.

$$E(y_i) = \sum_{k=1}^{6} kp_k = 1.900, D(y_i) = \sum_{k=1}^{6} (k - E(y_i))^2 p_k = 1.550,$$

$$\bar{r} = \sum_{k=1}^{6} \frac{n_k}{n} k = 2.360, D(\bar{r}) = \frac{1}{n} \sum_{k=1}^{6} (k - \bar{r})^2 \frac{n_k}{n} = 0.006924.$$

By substituting them into equations (6) and (7), it can be seen that the estimated value of the proportion with procrastinating or shirking work is

$$\hat{\mu} = \frac{2.360 - 1.900}{6 + 1 - 2 \times 1.900} = 0.1438.$$

The corresponding variance and standard deviation are respectively

$$D(\hat{\mu}) = \frac{0.006924}{(6 + 1 - 2 \times 1.900)^2} = 0.000676$$

and

$$\sqrt{D(\mu)} = \sqrt{0.000676} = 0.026003.$$

Further calculations show that A = 0.000172, B = 0.000505. that is to say, the standard deviation generated by the random selection of cards by the respondents is 2.246%, while the variance generated by the issue of procrastinated or shifted work itself is 1.310%.

If the precision is set to 1x standard deviation, the percentage of employees who had procrastinated or shirked work is 14.38 ± 2.60%. If the precision is set to 2x standard deviation, the percentage of employees who had procrastinated or shirked work is 14.38 ± 5.20%. According to the above accuracy standards, this indicates that the range of the number of employees who have had procrastinated or shirked work is 35-51 and 28-59, respectively.

4 Conclusion

Firstly, it was mentioned earlier that the estimation precision$D(\hat{\mu})$can be decomposed into the variance generated directly during the investigation of sensitive issues and the variance generated by the random selection of cards by respondents. Therefore, when the total number of respondents$n$is fixed, the precision can be improved by adjusting the parameters. For example, the size of$P_k$can be changed appropriately to improve the precision. Secondly, the mathematical model is not only limited to the survey of sensitive issues in company management, but also is applied to other sensitive issues of the same type. Finally, this paper mainly discusses whether there is sensitive behavior for respondent, and the next research will focus on improving the method to enable to survey and analysis for the frequency of sensitive behavior.

About the Author

* Corresponding Author : Ming Chen, 010431@hbzyy.edu.cn

References

[1] Tan, M. T., Tang, M.-L., Tian, G.-L. & Yuen, K. C. Bayesian non-randomized response models for surveys with sensitive questions. Statistics and its interface 2, 13–25 (2009).

[2] Coutts, E. & Jann, B. Sensitive questions in online surveys: Experimental results for the randomized response technique (RRT) and the unmatched count technique (UCT). Sociological Methods & Research 40, 169–193 (2011).

[3] Singh, H. P. & Tarray, T. A. A dexterous randomized response model for estimating a rare sensitive attribute using Poisson distribution. Statistics & Probability Letters 90, 42–45 (2014).

[4] Yan, T. Consequences of asking sensitive questions in surveys. Annual Review of Statistics and Its Application 8, 109–127 (2021).

[5] Kowalczyk, B., Niemiro, W. & Wieczorkowski, R. Item count technique with a continuous or count control variable for analyzing sensitive questions in surveys. Journal of Survey Statistics and Methodology 11, 919–941 (2023).

[6] Liu, Y., Ji, P. & Wu, G. Application of Joint Ratio Estimation Method in Investigation of Sensitive Issue. Statistics & Decision 40, 41–45 (2024).

[7] Zang, L., wang, D. & Zhang, J. Sensitive Issues in Social Science Research: Judging Features and Handling Methods. Study & Exploration 36-42+186 (2023).

[8] Wang, P. & Ji, C. List Experiment in Measuring Sensitive Issue—Example from Non-institutionalized Political Participation. China Soft Science 9, 135–143 (2016).

[9] Watimin, N. H., Zanuddin, H., Rahamad, M. S. & Yadegaridehkordi, E. Content framing role on public sentiment formation for pre-crisis detection on sensitive issue via sentiment analysis and content analysis. Plos one 18, e0287367 (2023).

[10] Wolter, F. & Diekmann, A. False positives and the "more-is-better" assumption in sensitive question research: new evidence on the Crosswise Model and the Item Count Technique. Public Opinion Quarterly 85, 836–863 (2021).