

Demographic Correlates of Cybercrime Awareness among Respondents: A Multiple Linear Regression Analysis

Mr. Anil M. Panchal., Dr. Arti J. Rajyaguru

Department of Statistics, Veer Narmad South Gujarat University (VNSGU), Surat, Gujarat, India

DOI: <https://doi.org/10.51583/IJLTEMAS.2026.150900083>

Received: 27 September 2026; Accepted: 02 October 2026; Published: 15 October 2026

ABSTRACT

This study examines the association between selected demographic characteristics and cybercrime awareness among 445 mobile Internet users from five selected talukas of Surat District, Gujarat, India. Cybercrime awareness was measured using a seven-item composite scale and transformed to a 0–100 index. Area, education, gender, and age group were represented using dummy variables, with Umarpada, Graduate education, Male, and Older age group as the respective reference categories. Multiple linear regression was estimated using the Enter method. The overall model was statistically significant, $F(10, 434) = 6.860$, $p < .001$, with $R^2 = .1365$ and adjusted $R^2 = .1166$. The Breusch–Pagan diagnostic indicated heteroskedasticity; therefore, HC3 heteroskedasticity-consistent standard errors were used for final coefficient-level inference. Bardoli, Mangrol, and Female gender showed statistically significant negative associations with awareness relative to their reference categories, while Postgraduate education showed a statistically significant positive association relative to Graduate education. Up-to-School education, Mahuva, Mandvi, Doctoral/Professional education, Young, and Middle were not statistically significant at the 5% level under HC3 inference. The findings indicate adjusted observational associations rather than causal effects.

Keywords: cybercrime awareness; demographic characteristics; multiple linear regression; heteroskedasticity; HC3

INTRODUCTION

Cybercrime has become an important social and technological concern as individuals increasingly interact with digital platforms, online communication systems, electronic services, and internet-based transactions. Awareness of cybercrime and basic cybersecurity practices can help individuals recognize online risks and respond more safely to potentially harmful digital situations.

Cybercrime awareness may differ across population groups. Geographic context, educational attainment, gender, and age may be associated with differences in access to digital information, digital experience, exposure to online risks, and familiarity with cybersecurity practices. Examining these characteristics jointly is useful because the observed association of one characteristic with awareness may depend on the distribution of other characteristics.

Indian studies have examined cybercrime and cybersecurity awareness among students using demographic, educational, gender, and locality-related characteristics. Senthilkumar and Easwaramoorthy (2017) examined college students in Tamil Nadu; Thakur and Kang (2018) studied gender and locale differences among adolescents; and Verma and Kushwaha (2021) examined awareness among secondary school students. More recent studies have considered youth and higher-education populations in Delhi/NCR, Coimbatore, Navsari, and Western Odisha (Chawla et al., 2023; Kaleeshwari & Jegadeeshwaran, 2024; Prajapati & Sharma, 2024; Sahu et al., 2026). The present study extends this line of inquiry by examining the simultaneous association of area, education, gender, and age group with a continuous cybercrime awareness score using multiple linear regression. The analysis is observational and therefore does not establish causal effects.

Objectives of the Study

- To examine the association between selected demographic characteristics and cybercrime awareness.
- To assess whether awareness differs across area, education, gender, and age-group categories.
- To estimate a multiple linear regression model using dummy-variable coding.
- To assess important regression assumptions and diagnostic characteristics.
- To use heteroskedasticity-robust inference where required by the diagnostic results.

RESEARCH METHODOLOGY

Study Design and Sample

The study was conducted among mobile Internet users in five selected talukas of Surat District, Gujarat, India: Mandvi, Mangrol, Umarpada, Mahuva, and Bardoli. The target population comprised people who use the Internet through mobile devices. Primary data were collected through a structured questionnaire administered using Google Forms. A total of 445 valid responses were included in the analysis. The study used a random sampling approach as specified in the study design. The dependent variable was a continuous cybercrime awareness score expressed on a 0–100 scale. The regression model was estimated using IBM SPSS Statistics, and the same OLS specification was reproduced in R for HC3 robust inference.

Measurement of Cybercrime Awareness

Cybercrime awareness was measured using seven items covering awareness of the term cybercrime, the Information Technology Act, 2000, the National Cyber Security Policy, 2013, M-Kavach, Cyber Swachhta Kendra, cyber forensic training centres/laboratories, and the national cybercrime reporting helpline 1930. Each item used a four-point ordinal response scale: “I don't know,” “I don't know, but I have heard of it,” “I know a little,” and “I know very well,” with higher scores indicating greater awareness. A preliminary version of the questionnaire was pilot tested and yielded Cronbach's $\alpha = 0.762$. In the final sample of 445 respondents, the seven-item scale demonstrated good internal consistency (Cronbach's $\alpha = 0.854$; standardized $\alpha = 0.855$), with corrected item-total correlations ranging from 0.455 to 0.746. All seven items were retained. The seven responses were summed to obtain Awareness_Total. Because seven items were scored from 1 to 4, the theoretical minimum and maximum total scores were 7 and 28, respectively. The score was transformed to a 0–100 awareness index as follows:

Awareness Index = $((\text{Awareness_Total} - 7) / 21) \times 100$. Higher index values indicate higher measured cybercrime awareness.

Dummy Variable Coding

Table 1. Dummy Variable Coding and Reference Categories

Variable	Categories in model	Reference
Area	Bardoli, Mahuva, Mandvi, Mangrol	Umarpada
Education	Up-to-School, Postgraduate, Doctoral/Professional	Graduate
Gender	Female	Male
Age Group	Young, Middle	Older

One category from each categorical predictor was omitted as the reference category. Each coefficient therefore represents the adjusted difference in awareness score between that category and its reference category, controlling for the other predictors. Age was grouped as Young (under 18, 18–24, and 25–34 years), Middle (35–44 and 45–54 years), and Older (55 years and above). Older was used as the reference category. For education, Graduate was used as the reference category in the final re-estimated model.

Model Specification

$$Y = \beta_0 + \beta_1(\text{Area}) + \beta_2(\text{Education}) + \beta_3(\text{Gender}) + \beta_4(\text{Age Group}) + \varepsilon$$

For estimation, ten dummy variables were used: X_1 – X_4 for Area, X_5 – X_7 for Education (Up-to-School, Postgraduate, and Doctoral/Professional, with Graduate as the reference), X_8 for Female, and X_9 – X_{10} for Young and Middle. The use of one fewer indicator than the number of categories follows the standard dummy-variable specification (DeMaris, 2004).

Estimation and Diagnostics

Multicollinearity was assessed using tolerance and variance inflation factor (VIF) statistics. In the final regression model, tolerance values ranged from 0.349 to 0.878 and VIF values ranged from 1.139 to 2.868. These values indicated no evidence of problematic multicollinearity among the predictors.

All ten dummy predictors were entered simultaneously using the Enter method. Model fit was evaluated using R , R^2 , adjusted R^2 , the standard error of the estimate, and the overall F-test. Coefficients were reported with standard errors, t-values, p-values, and 95% confidence intervals. Diagnostics included VIF, tolerance, condition index, Durbin–Watson, Cook’s distance, Mahalanobis distance, standardized residuals, normality diagnostics, residual plots, and the Breusch–Pagan test.

Robust Inference and Interpretation. Because the Breusch–Pagan test indicated heteroskedasticity, HC3 heteroskedasticity-consistent standard errors were used for final coefficient-level inference. Regression coefficients were interpreted as adjusted associations or estimated differences relative to the specified reference categories, controlling for the other variables included in the model. As the study is observational, the regression results were not interpreted as causal effects.

Results

The multicollinearity diagnostics further supported the adequacy of the final model. Tolerance values ranged from 0.349 to 0.878, while VIF values ranged from 1.139 to 2.868. Thus, none of the predictors showed evidence of problematic multicollinearity.

Descriptive Summary

Table 2. Descriptive Summary of Regression Variables

Variable	Mean	SD	N
Cybercrime Awareness Score	34.6388	21.1038	445
Bardoli dummy	0.2899	0.4542	445
Mahuva dummy	0.1551	0.3624	445
Mandvi dummy	0.2202	0.4149	445
Mangrol dummy	0.2382	0.4265	445

Up-to-School dummy	0.2404	0.4278	445
Doctoral/Professional dummy	0.0427	0.2022	445
Postgraduate dummy	0.2809	0.4499	445
Female dummy	0.4202	0.4942	445
Young dummy	0.4562	0.4986	445
Middle dummy	0.3775	0.4853	445

Model Fit

Table 3. Overall Model Fit

Statistic	Value
N	445
R	0.369
R ²	0.1365
Adjusted R ²	0.1166
Standard Error of Estimate	19.84
F(10, 434)	6.860
p-value	<0.001
Durbin–Watson	1.860

The overall model was statistically significant, $F(10, 434) = 6.860$, $p < 0.001$. The model explained 13.65% of the observed variation in awareness scores, with adjusted $R^2 = 0.1166$. This represents statistically significant but modest explanatory power.

Predictor	B	HC3 SE	t	p	95% CI
Constant	40.018	4.736	8.449	<0.001	[30.709, 49.327]
Bardoli	-14.251	4.235	-3.365	0.0008	[-22.576, -5.927]
Mahuva	-7.615	4.592	-1.658	0.098	[-16.640, 1.410]
Mandvi	-6.522	4.456	-1.464	0.144	[-15.279, 2.236]
Mangrol	-13.600	4.253	-3.198	0.0015	[-21.960, -5.241]
Up-to-School	-4.973	2.588	-1.921	0.0554	[-10.060, 0.115]
Postgraduate	8.986	2.482	3.620	0.0003	[4.108, 13.865]

Doctoral/Professional	9.068	7.020	1.292	0.1971	[-4.729, 22.865]
Female	-4.551	2.080	-2.188	0.0292	[-8.638, -0.463]
Young	5.681	3.022	1.880	0.0608	[-0.259, 11.621]
Middle	5.864	3.056	1.919	0.0556	[-0.142, 11.870]

ANOVA

Table 4. ANOVA for the Regression Model

Source	SS	df	MS	F	p
Regression	26989.948	10	2698.995	6.860	<0.001
Residual	170755.014	434	393.445		
Total	197744.962	444			

Heteroskedasticity and HC3 Inference

The Breusch–Pagan test indicated heteroskedasticity. Therefore, conventional OLS standard errors and p-values were not treated as the final basis for coefficient-level inference. HC3 heteroskedasticity-consistent standard errors were calculated. The OLS coefficient estimates remain unchanged, while the standard errors, t-statistics, p-values, and confidence intervals are adjusted for heteroskedasticity.

HC3-Robust Regression Coefficients

Table 5. Multiple Linear Regression Results with HC3 Robust Standard Errors

Note. Reference categories are Umapada, Doctoral, Male, and Old. HC3 = heteroskedasticity-consistent standard errors. Coefficients are OLS estimates; HC3 is used for inference.

Interpretation of Coefficients

Bardoli was negatively associated with awareness ($B = -14.251$, $p = 0.0008$). Holding the other variables constant, respondents from Bardoli had an estimated awareness score 14.25 points lower than respondents from Umapada.

Mangrol was negatively associated with awareness ($B = -13.600$, $p = 0.0015$), corresponding to an estimated 13.60-point lower score than Umapada, conditional on the other predictors.

Up-to-School education was not statistically significant at the 5% level under HC3 inference ($B = -4.973$, $p = 0.0554$) relative to the Graduate reference category.

Female gender was negatively associated with awareness ($B = -4.551$, $p = 0.0292$), corresponding to an estimated 4.55-point lower score than Male respondents, conditional on the other predictors.

Postgraduate education was positively associated with awareness relative to Graduate education ($B = 8.986$, $p = 0.0003$). Mahuva, Mandvi, Up-to-School education, Doctoral/Professional education, Young, and Middle were not statistically significant at the 5% level after HC3 adjustment.

Estimated Regression Equation

$$\hat{Y} = 40.018 - 14.251(\text{Bardoli}) - 7.615(\text{Mahuva}) - 6.522(\text{Mandvi}) - 13.600(\text{Mangrol}) - 4.973(\text{Up-to-School}) + 8.986(\text{Postgraduate}) + 9.068(\text{Doctoral/Professional}) - 4.551(\text{Female}) + 5.681(\text{Young}) + 5.864(\text{Middle})$$

The intercept of 40.018 is the predicted awareness score for the combined reference profile: Umarpada, Graduate education, Male, and Older age group.

Diagnostic Assessment

Multicollinearity

Table 6. Multicollinearity Diagnostics

Diagnostic	Observed result	Assessment
VIF	Assessed in final model	Multicollinearity was evaluated using variance inflation factors.
Tolerance	Assessed in final model	Tolerance was examined alongside VIF.
Maximum Condition Index	Assessed in final model	Condition index was examined as an additional diagnostic.

Multicollinearity was assessed using VIF, tolerance, and condition index. The education reference category was re-levelled to Graduate for the final model; therefore, the numerical VIF values from the earlier Doctoral-reference specification were not carried forward. Final numerical VIF values should be regenerated from the final model if the journal requires them in the diagnostic table.

Residual and Influence Diagnostics

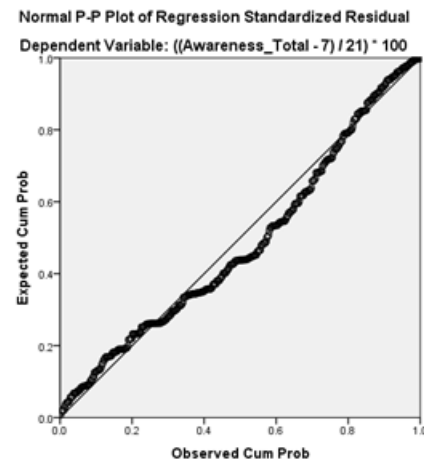
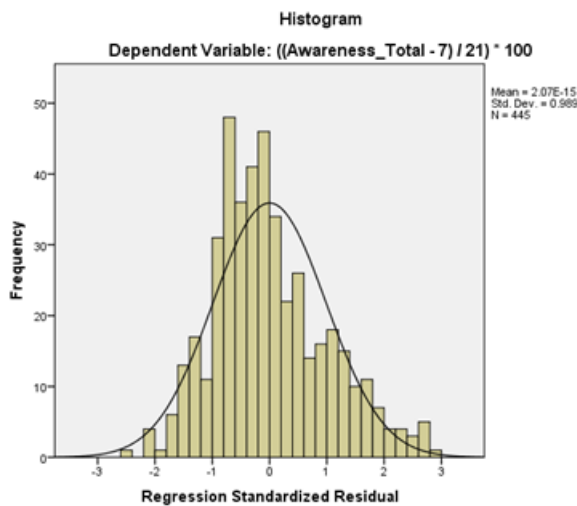
Table 7. Residual and Influence Diagnostics

Diagnostic	Result	Interpretation
Durbin–Watson	1.860	No strong indication of first-order residual autocorrelation; cautious interpretation
Maximum Cook's Distance	0.042	No large influential observation under conventional <1 screening
Maximum Mahalanobis Distance	34.822	Multivariate extremeness examined; no case judged influential by Cook's distance
Standardized residual range	–2.442 to 2.834	No standardized residual exceeded ± 3

The diagnostic results did not identify a major influence problem. Durbin–Watson does not indicate a strong first-order autocorrelation pattern, although it is not a complete independence test. Cook's distance values were small.

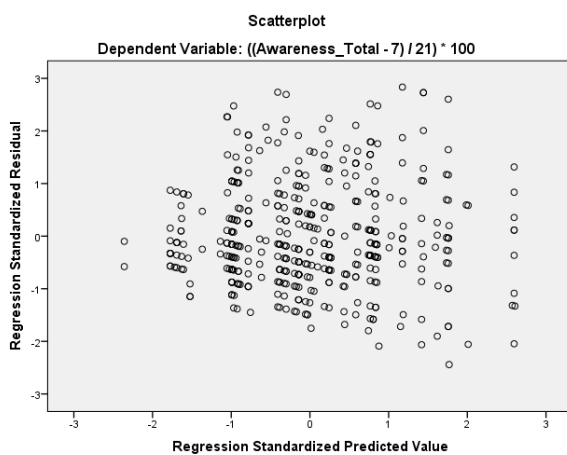
Normality

The Shapiro–Wilk test was statistically significant ($p < 0.001$), indicating departure from exact residual normality. With $N = 445$, formal tests can detect relatively small deviations. The residual histogram and normal Q–Q–P–P plot should therefore be considered together with the formal test rather than relying on the p-value alone.



Linearity and Heteroskedasticity

The residual-versus-predicted scatterplot should be used as a visual assessment of linearity and residual variance.



The formal Breusch–Pagan result indicated heteroskedasticity; therefore, HC3 robust inference was adopted for the final coefficient-level conclusions.

DISCUSSION

The selected demographic variables were jointly associated with cybercrime awareness, but the R^2 of 0.1365 indicates that most variation remains unexplained by the current specification. Bardoli and Mangrol showed lower adjusted awareness scores than Umarpada, while Postgraduate education showed a higher adjusted awareness score than Graduate education. Female respondents also showed lower adjusted awareness scores than Male respondents.

The positive Young and Middle coefficients did not remain statistically significant at the 5% level after HC3 adjustment. This illustrates the importance of robust inference when heteroskedasticity is present. The education block also showed higher VIF values, so education-related coefficients warrant cautious interpretation.

The modest explanatory power suggests that awareness may also be associated with factors not included in the present model, such as internet-use patterns, digital literacy, occupation, income, prior exposure to cybercrime, cybersecurity training, and other digital experiences. These factors were not tested here and are therefore directions for future research rather than findings of this study.

Practical Implications

The results may help identify groups and areas that could receive greater attention in cybercrime-awareness initiatives. However, because the study is observational, the findings should not be used to claim that a demographic characteristic causes a change in awareness. Programme design should also consider digital literacy, internet use, previous training, and local context.

Robustness and Further Analysis

The principal robustness analysis was the use of HC3 heteroskedasticity-consistent standard errors following the Breusch–Pagan result. The OLS coefficient estimates remained unchanged, while the inferential conclusions became more conservative. In particular, the Middle-age coefficient did not remain significant at the 5% level after HC3 adjustment. The final model also re-levelled education to use Graduate, the largest education group, as the reference category, improving the substantive interpretability of the education comparisons.

Interaction effects and hierarchical regression were considered as possible extensions but were not estimated in the primary model. They should be undertaken only with a clear theoretical rationale. Education could also be examined as an ordinal predictor in a sensitivity analysis, but the present model retains categorical dummy coding so that equal spacing between education categories is not imposed.

Limitations

- The model includes only variables available in the questionnaire and omits other potentially relevant determinants.
- The model explains a modest proportion of awareness variation ($R^2 = 0.1365$).
- Heteroskedasticity was detected, so HC3 robust inference is preferred over conventional OLS inference.
- Education predictors have VIF values above 5, indicating residual multicollinearity concentrated in the education block.
- The Shapiro–Wilk test indicates departure from exact residual normality, although formal tests can be sensitive at $N = 445$.
- The study is observational and coefficients represent adjusted associations rather than causal effects.
- Interaction and hierarchical specifications were not estimated in the primary analysis.

CONCLUSION

A dummy-variable multiple linear regression model was used to examine cybercrime awareness among 445 mobile Internet users. The model was statistically significant and explained approximately 13.65% of the variation in awareness. After heteroskedasticity was detected, HC3 robust standard errors were used for final inference. Significant negative associations were observed for Bardoli and Mangrol areas and Female gender, while Postgraduate education showed a significant positive association relative to Graduate education. Overall, the findings indicate that selected demographic characteristics are associated with cybercrime awareness, but they do not fully explain individual differences.

Declarations: The author declares that there is no conflict of interest. The study used questionnaire-based data collected from respondents, with confidentiality maintained and the data used for research purposes. No formal institutional ethical approval or formal informed-consent procedure was reported for the survey; this should be disclosed and handled in accordance with the journal's requirements before submission. Generative AI was used to assist with academic language refinement, manuscript organization, and formatting; the author

reviewed and edited the AI-assisted content and takes full responsibility for the accuracy, originality, statistical analysis, interpretation, and final content of the manuscript.

REFERENCES

1. Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47(5), 1287–1294. <https://doi.org/10.2307/1911963>
2. Chawla, V., Kapoor, Y., & Chawla, T. (2023). Cybersecurity awareness amongst youth—A survey in Delhi/NCR. *Indian Journal of Computer Science*, 8(2), 42–52. <https://doi.org/10.17010/ijcs/2023/v8/i2/172777>
3. Choudhary, D. M. (2020). Cyber crime awareness among higher education students from Haryana with respect to various demographical variables. *PalArch's Journal of Archaeology of Egypt/Egyptology*, 17(7), 14454–14461.
4. DeMaris, A. (2004). Multiple regression with categorical predictors: ANOVA and ANCOVA models. In W. A. Shewhart, S. S. Wilks, & A. DeMaris (Eds.), *Regression with social data*. Wiley. <https://doi.org/10.1002/0471677566.ch4>
5. Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics* (5th ed.). McGraw-Hill.
6. Kaleeshwari, S., & Jegadeeshwaran, M. (2024). Quantitative assessment of the cybersecurity and cybercrime awareness among postgraduate commerce students in the Coimbatore District, India. *Asian Journal of Probability and Statistics*, 26(9), 185–197. <https://doi.org/10.9734/ajpas/2024/v26i9653>
7. Panjani, H., & Mudgal, A. (2025). A study of cyber safety awareness among students and educational initiatives. *Indian Journal of Educational Technology*, 7(2), 246–256.
8. Prajapati, D. H., & Sharma, A. C. (2024). A comparative study of cyber-crime awareness in college students of Navsari District. *International Journal of Multidisciplinary Educational Research*, 13(11[2]), 84–89.
9. Sahu, R., Rao, K. S. K., & Mahalik, Y. (2026). Cybercrime awareness among undergraduate students of Western Odisha: A survey on gender and locality based differences. *Research Review International Journal of Multidisciplinary*, 11(5), 108–118. <https://doi.org/10.31305/rrijm.2026.v11.n05.011>
10. Senthilkumar, K., & Easwaramoorthy, S. (2017). A survey on cyber security awareness among college students in Tamil Nadu. *IOP Conference Series: Materials Science and Engineering*, 263, 042043. <https://doi.org/10.1088/1757-899X/263/4/042043>
11. Thakur, A., & Kang, T. K. (2018). Gender and locale differences in cyber crime awareness among adolescents. *Indian Journal of Health and Wellbeing*, 9(8–9), 906–916.
12. Verma, M. K., & Kushwaha, S. S. (2021). Awareness towards cybercrime among secondary school students: The role of gender and school management. *Safer Communities*, 20(3), 150–158. <https://doi.org/10.1108/SC-07-2020-0026>
13. Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th ed.). Cengage Learning.
14. Indian Cyber Crime Coordination Centre. (n.d.). National Cyber Crime Reporting Portal. Ministry of Home Affairs, Government of India. <https://www.cybercrime.gov.in/>

APPENDIX A. Variable Coding

Variable group	Dummy variables	Reference
Area	Bardoli, Mahuva, Mandvi, Mangrol	Umarpada
Education	Up-to-School, Postgraduate, Doctoral/Professional	Graduate
Gender	Female	Male
Age Group	Young, Middle	Older