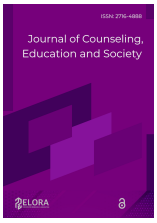
Vol. 7, No. 1, 2026, pp. 723-733
DOI: https://doi.org/10.29210/08jces698200

Contents lists available at Journal ELORA

Journal of Counseling, Education and Society
ISSN: 2716-4896 (Print) , ISSN 2716-4888 (Electronic)

Journal homepage: https://jces.eloracenter.org/jces



Higher education participation across 39 indonesian regions socioeconomic and digital correlates of distance

Mailani Hamdani*)[1], Yanuar Trisnowati[2], Sabna Ainazah Fatikhah[2]

[1] Management Study Program, Faculty of Economics and Business, Universitas Terbuka, Indonesia
[2] Entrepreneurship Study Program, Faculty of Economics and Business, Universitas Terbuka, Indonesia

Article Info

Article history:

Received Sept 11th, 2026
Revised Oct 15th, 2026
Accepted Nov 08th, 2026

Keyword:

Distance higher education,
Digital divide,
Participation,
Socioeconomic context,
Regression

ABSTRACT

Distance higher education is expected to reduce geographical barriers, yet regional participation may remain influenced by socioeconomic conditions and unequal digital access. This study examined the associations between regional participation in Universitas Terbuka (UT) and six socioeconomic, demographic, and digital-access indicators across 39 designated locations in Indonesia. Using secondary administrative and regional statistical data, the study calculated a crude participation ratio based on active UT students per 10,000 residents aged 19–24. Multiple linear regression was conducted using Ordinary Least Squares (OLS), complemented by HC3 robust inference, diagnostic tests, and sensitivity analyses. The baseline model explained 27.6% of regional variation (R^2 = .276; adjusted R^2 = .140), but was not statistically significant overall, $F(6, 32) = 2.03$, $p = .090$. None of the six predictors remained statistically significant under HC3 inference. Sensitivity analyses revealed substantial variation in model fit across alternative sample specifications, particularly when influential observations and provisional socioeconomic data were considered. The findings do not establish that socioeconomic and digital conditions are irrelevant; rather, they indicate that the available regional indicators provide limited evidence of independent associations with UT participation. Stronger geographic data validation, consistent socioeconomic measurements, and direct information on learners' circumstances and support services are needed to improve regional participation assessment and inform equitable distance higher education policies.

© 2026 The Authors.
This is an open access article under the CC BY-NC-SA license
(https://creativecommons.org/licenses/by-nc-sa/4.0)

Corresponding Author:

Mailani Hamdani,
Universitas Terbuka
Email: mailani@ecampus.ut.ac.id

Introduction

Open and distance education is widely recognized as a means of expanding higher education opportunities for people facing geographical, occupational, family, or financial constraints. However, removing physical attendance requirements does not necessarily eliminate inequalities in educational participation. Differences in economic resources, technological access, and local infrastructure may continue to shape learners' opportunities (UNESCO, 2020). OECD (2024) also highlights the importance of socioeconomic circumstances in determining access to and participation in education. For Universitas Terbuka (UT), which provides distance higher education across Indonesia, participation therefore needs to be examined not only through enrollment figures but also through the regional conditions in which learners live. Geographical location is an important dimension of higher education equity because it can influence students' educational choices and academic experiences (Brownie et al., 2025).

Indonesia presents a particularly relevant context because of its geographical dispersion, uneven economic development, and differences in digital infrastructure. The World Bank (2021) identifies affordability, infrastructure quality, and digital capabilities as important barriers to digital inclusion. Moreover, the digital divide extends beyond physical access to include digital skills and patterns of technology use (van Dijk, 2020). Van Deursen and van Dijk (2015) further distinguish different dimensions of Internet access, while Loh and Chib (2022) emphasize the importance of individuals' ability to appropriate digital resources. For adult learners, employment, financial circumstances, and family responsibilities may further constrain educational participation (OECD, 2024). These inequalities are also reflected in higher education, where geographical differences in broadband availability can intersect with students' socioeconomic conditions (Cullinan et al., 2021). Consequently, regional Internet access and device ownership should be understood as indicators of material digital access rather than comprehensive measures of individual digital readiness.

Research on online higher education also demonstrates that technology access alone cannot explain students' educational experiences. Bond et al. (2021), in their mapping of emergency remote teaching, identify substantial differences in students' experiences of online learning, although such findings cannot be directly generalized to established open universities. Research in Indonesia similarly identifies variations in how university students experience and use educational technologies (Hidayat et al., 2022). Therefore, understanding participation in distance higher education requires attention to the broader socioeconomic and infrastructural environments in which students study, rather than assuming that household connectivity automatically translates into educational readiness or success (UNESCO, 2023).

To examine these relationships, this study adopts a place-based analytical approach using 39 UT-labelled study locations across Indonesia as its units of analysis. Regional administrative student records are integrated with socioeconomic, demographic, and digital infrastructure indicators to examine differences in participation across locations. This approach is relevant because geographical conditions can shape access to educational opportunities, while regional broadband disparities may reinforce existing inequalities (Brownie et al., 2025). The analysis considers six explanatory variables: Gross Regional Domestic Product (GRDP) per capita, poverty, household Internet access, device ownership, population aged 19–24, and regional status. Household Internet access and device ownership represent the material-access dimensions of the digital divide, while individual digital skills and technology-use patterns are not directly measured. Furthermore, because the data are geographically aggregated, the findings are interpreted as regional associations rather than individual-level relationships, consistent with the methodological cautions concerning ecological effects discussed by Holt et al. (1996). The integration of administrative and regional datasets also presents measurement challenges that require careful reconciliation (Yunita et al., 2022).

Against this background, this study investigates the extent to which regional UT participation ratios are associated with the six socioeconomic, demographic, and digital-access indicators. Using multiple regression across 39 study locations, the analysis evaluates these relationships while considering heteroskedasticity-robust inference and alternative sample specifications. Given the relatively limited number of observations, particular attention is required when interpreting coefficient precision and model complexity (Kelley & Maxwell, 2003; Riley et al., 2019). The study contributes a regional perspective to discussions of participation equity in open and distance higher education. It examines statistical associations rather than causal effects and does not construct a participation-equity index or decomposable inequality measures, which are reserved for a separate study.

Method

Design and observations. This cross-sectional, observational study used secondary data for 39 designated UT regional locations. The student dataset records 2026 active-student totals, while supporting administrative variables and regional indicators were compiled from the research dataset and official statistical sources. Each row was treated as the named municipality or district that matched the corresponding UT record, as provided by the research team. The exact one-to-one match between students' complete residence catchment and official district codes was not independently reverified; consequently, the unit of inference is the supplied matched regional observation, not every district in Indonesia.

Outcome definition. Let A_i denote active UT students assigned to study region i, and P_i the corresponding population aged 19–24. The dependent variable was $Y_i = 10{,}000 \times A_i / P_i$. This is a crude UT-specific representation ratio, not the national Gross Enrolment Ratio (GER) for all higher education. The official BPS tertiary GER uses ages 19–23 and includes tertiary enrollment beyond UT (BPS, 2025). UT serves learners older than traditional college age; hence, Y_i values should not be interpreted as probabilities of attendance among 19–

24-year-olds, and values above conventional percentage bounds would not imply erroneous individual enrollment probabilities.

Predictors and data preparation. Six regional indicators were specified a priori: GRDP per capita, poverty, household Internet access, household ownership of digital devices, the size of the 19–24 population, and the regional status code. GRDP was expressed in million rupiah per person; poverty and digital indicators were placed on percentage scales where source interpretation permitted; the 19–24 population was expressed in units of 100,000; and the regional status code retained its original 0/1 coding. Because the uploaded file did not provide a verifiable legend for this code, the coefficient is interpreted only as the difference between code 1 and code 0, not as a confirmed 3T versus non-3T comparison. Table 1 lists the operational definitions.

Quality control and missing information. Formatting and coverage inconsistencies were flagged before modeling. Values expressed as "million" were parsed as GRDP per capita where plausible; ambiguous "billion/trillion" entries were not silently converted to individual per-capita amounts. Poverty entries that appeared to be counts were conditionally converted to shares only when a corresponding population figure could be read; some entries remained ambiguous. Missing or unparseable GRDP and poverty data were replaced with complete-case medians in the full 39-region scenario. Eight GRDP entries, four poverty entries, and six age-denominator entries were flagged as imputed or estimated; 16 regions had at least one such flag. These substitutions produce an exploratory analytic scenario rather than officially harmonized BPS measurements. Original flags were retained for sensitivity analysis.

Analytic strategy. Ordinary Least Squares (OLS) was estimated with an intercept and all six covariates. Model fit was summarized using R^2, adjusted R^2, and an omnibus F-test. Coefficients were accompanied by conventional t-tests and HC3 heteroskedasticity-consistent standard errors, which serve as a sensitivity check for small samples and high-leverage observations (MacKinnon & White, 1985). Multicollinearity was examined using Variance Inflation Factors (VIF), heteroskedasticity using the Breusch–Pagan test, residual normality using the Shapiro–Wilk and Jarque–Bera tests, and influence using Cook's distance. Robustness was tested across 28 complete cases for the two ambiguous socioeconomic predictors, 18 higher-quality complete cases, 26 higher-quality regions with provisional imputations, all regions excluding Jakarta, and a log-transformed outcome model. These specifications test the stability of associations rather than serving as competing significance-seeking models. The conventional significance level was .05.

Table 1. Variables and measurement in the regional OLS analysis

Variable	Operational measurement	Analytical caveat
Participation (Y)	Active UT students / population aged 19–24 × 10,000	Crude UT-specific ratio; not BPS tertiary GER
GRDP per capita (X1)	Million Indonesian rupiah per person	Unit and reference year require confirmation
Poverty (X2)	Provisional poverty percentage	Some entries inferred from counts; review needed
Internet access (X3)	Households with Internet access (%)	Does not capture affordability or connection quality
Device ownership (X4)	Households owning relevant digital device (%)	Does not capture student-level device availability
Population (X5)	Residents aged 19–24, in 100,000 persons	The denominator of Y
Regional code (X6)	Indicator: 1 vs 0	Original category definition unverified

The estimated model was $Y_i = \beta_0 + \beta_1 GRDP_i + \beta_2 Poverty_i + \beta_3 Internet_i + \beta_4 Devices_i + \beta_5 Population_i + \beta_6 RegionCode_i + \varepsilon_i$. Coefficients correspond to the unit definitions in Table 1; cross-sectional associations should not be interpreted as the effect of an intervention on a student.

Results and Discussions

Regional Participation and Descriptive Characteristics

Across the 39 recorded locations, 877,760 active Universitas Terbuka (UT) students were observed. The mean regional participation ratio was 2,903.7 active students per 10,000 residents aged 19–24 (SD = 1,626.0), indicating substantial variation across locations. This figure represents an unweighted mean of regional ratios rather than the national UT enrollment rate. It should therefore not be interpreted as the national higher

education Gross Enrolment Ratio (GER), which uses a different age denominator and institutional coverage (BPS, 2025)

Table 2. Descriptive statistics of the 39 regional observations.

Indicator	Mean	Median	SD	Min	Max
Active UT students	22,507	16,691	20,013	3,170	107,032
UT participation ratio / 10,000	2,903.7	2,695.7	1,626.0	298.2	7,559.8
GRDP per capita (million IDR)	107.2	89.2	76.1	7.5	403.9
Poverty proxy (%)	13.9	8.7	12.6	1.0	42.3
Households with Internet (%)	81.1	80.0	10.1	53.7	100.0
Households with devices (%)	85.9	85.9	9.4	60.0	100.0
Population aged 19–24 (× 100,000)	1.01	0.60	1.32	0.17	7.97

The descriptive statistics indicate considerable geographical variation in regional UT participation, GRDP per capita, poverty, and digital infrastructure. The participation ratio ranged from 298.2 to 7,559.8 students per 10,000 residents aged 19–24. Similarly, GRDP per capita ranged from IDR 7.5 million to IDR 403.9 million, while the poverty proxy ranged from 1.0% to 42.3%. These differences suggest that the 39 locations represent diverse socioeconomic and demographic contexts. However, descriptive variation alone does not establish an explanatory relationship between regional conditions and higher education participation.

Household Internet access averaged 81.1%, while household device ownership averaged 85.9%. Although these figures suggest relatively widespread household access to digital technology, they do not necessarily reflect the actual learning conditions experienced by individual UT students. Access to devices and connectivity may differ from the availability of reliable Internet, digital skills, learning engagement, and the ability to use technology effectively. Michikyan et al. (2025) demonstrate that digital inequality in higher education involves interconnected dimensions of access, connectivity, digital skills, engagement, and motivation. Consequently, household-level Internet access and device ownership should be interpreted as indicators of material digital access rather than comprehensive measures of students' digital learning readiness.

Multiple Regression Results

The six-predictor Ordinary Least Squares (OLS) model explained 27.6% of the variation in regional UT participation ($R^2 = .276$), while the adjusted R^2 was .140. The overall regression model did not reach conventional statistical significance, $F(6, 32) = 2.032$, $p = .090$. This indicates that the six regional predictors, when considered jointly, did not provide sufficient statistical evidence of an association with the participation ratio at the .05 significance level.

Table 3. OLS coefficients for crude UT participation (N = 39).

Predictor	B	SE	p	HC3 SE	HC3 p
Intercept	446.22	2,793.74	0.874	3,117.22	0.887
GRDP / capita	1.34	4.41	0.763	3.92	0.735
Poverty (%)	-20.73	21.31	0.338	20.84	0.327
Internet (%)	37.79	32.46	0.253	37.02	0.315
Devices (%)	4.17	30.80	0.893	31.71	0.896
Population 19–24	-640.84	248.56	0.015	931.85	0.497
Region code (1 vs 0)	-970.21	686.42	0.167	826.23	0.249

The coefficient estimates showed different directions across the explanatory variables. GRDP per capita had a positive but statistically imprecise association with regional participation (B = 1.34; HC3 p = .735). Poverty had a negative coefficient (B = −20.73; HC3 p = .327), while household Internet access showed a positive coefficient (B = 37.79; HC3 p = .315). Device ownership had a coefficient close to zero (B = 4.17; HC3 p = .896). The regional-status contrast was also statistically uncertain (B = −970.21; HC3 p = .249). None of these coefficients reached the .05 significance threshold using either conventional or HC3 inference.

These findings do not establish that regional economic conditions, poverty, or digital infrastructure are unrelated to UT participation. Rather, they indicate that the available regional indicators do not provide sufficiently precise evidence of independent associations within the current sample and model specification. The relatively small number of observations, provisional socioeconomic measurements, and differences in regional characteristics may contribute to the uncertainty of the estimates.

The distinction between household access and effective digital participation is particularly important. Michikyan et al. (2025) show that reliable connectivity and digital learning capabilities represent related but distinct dimensions of digital inequality. Thus, the absence of statistically significant coefficients for Internet

access and device ownership should not be interpreted as evidence that digital infrastructure is unimportant. Instead, the results suggest that broad household-level indicators may be insufficient to explain regional differences in participation without additional information about connectivity quality, affordability, digital competence, and student engagement.

Population Denominator and Interpretation of Regression Estimates

The coefficient for population aged 19–24 requires particular caution. Under conventional OLS inference, the coefficient was negative and statistically significant (B = −640.84, p = .015). However, its significance disappeared when HC3 standard errors were applied (HC3 p = .497). The substantial difference between these two results indicates that the apparent association is sensitive to the covariance estimator and should not be interpreted as robust statistical evidence.

An additional concern arises from the mathematical construction of the dependent variable. Population aged 19–24 is used as the denominator of the participation ratio and simultaneously included as an explanatory variable. This creates mathematical overlap between the outcome and predictor, potentially generating a mechanical association that does not necessarily reflect a substantive relationship between population size and educational participation.

Accordingly, the negative population coefficient should not be interpreted as evidence that regions with larger youth populations experience lower access to UT. The result primarily highlights the limitations of using a ratio outcome while including its denominator as a predictor. Future analyses should consider modeling active student counts directly, with an appropriate population exposure term only when the exposure denominator is conceptually aligned with the population at risk. Cameron and Trivedi (2013) explain that count regression provides an alternative framework for analyzing nonnegative event counts through Poisson and negative binomial specifications. The choice between these models should also consider the variance structure of the observed counts. Ver Hoef and Boveng (2007) demonstrate that overdispersion can influence the relative performance and estimated effects of alternative count regression models.

Model Fit, Diagnostics, and Influential Regional Observations

The observed-versus-fitted comparison provides further evidence of the limitations of the baseline regression model. The residual standard error was approximately 1,507.8 active students per 10,000 residents aged 19–24, indicating substantial unexplained variation relative to the mean participation ratio.

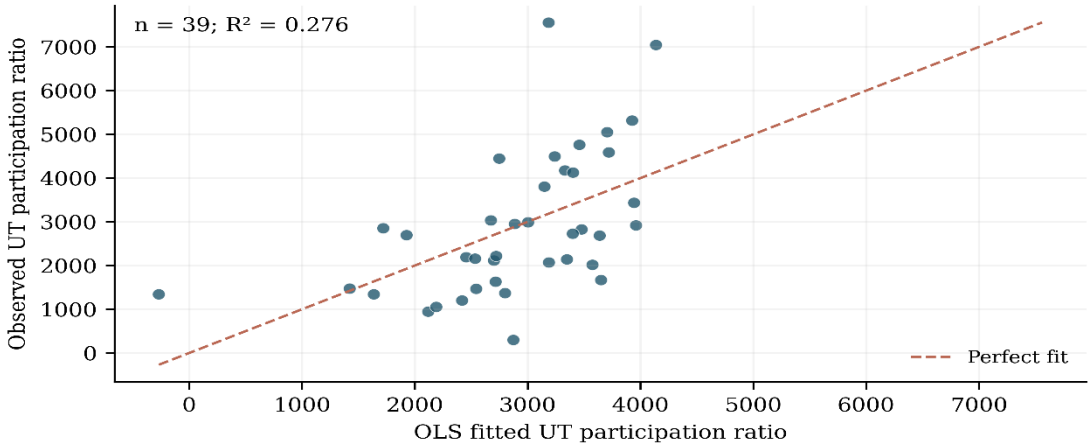


Figure 1. Observed versus fitted crude UT participation ratios in the baseline six-predictor OLS model. The dashed line indicates equality between observed and fitted values.

Figure 1 illustrates considerable dispersion around the equality line. Several regional observations deviate substantially from their fitted values, indicating that the six explanatory variables do not adequately capture the full variation in regional UT participation. Consequently, the baseline model should not be used to rank regions or establish policy priorities based solely on predicted participation ratios.

The diagnostic results indicated no severe multicollinearity among the explanatory variables, with VIF values ranging from 1.19 to 1.88. The Breusch–Pagan test did not detect statistically significant heteroskedasticity (p = .890). However, residual normality tests produced mixed results: the Shapiro–Wilk test was borderline (p = .053), whereas the Jarque–Bera test rejected normality (p = .035). These findings suggest that residual distribution assumptions should be considered cautiously, particularly given the limited sample size.

Cook's distance identified Jakarta as a highly influential observation, with a value of approximately 3.08, considerably exceeding the conventional screening reference of 4/n = 0.103. This indicates that Jakarta has substantial influence on the estimated regression results. Cook (1979) explains that influential observations can materially change regression coefficients, residual variances, and statistical tests when included or excluded from a model. Nevertheless, an influential observation should not automatically be removed from the analysis.

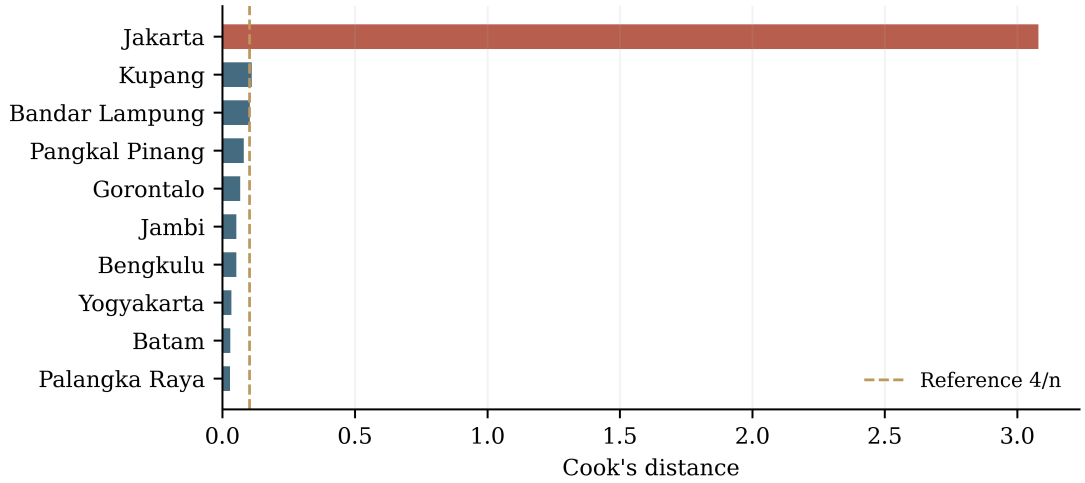


Figure 2. Ten largest Cook's distance values in the baseline regression. The horizontal reference line represents the conventional screening threshold of 4/n.

The influence of Jakarta is particularly important because its economic and infrastructural characteristics differ considerably from those of many other regional locations. Bollen and Jackman (1985) emphasize that regression diagnostics should not be applied mechanically when identifying influential observations. Instead, the substantive characteristics of each observation and the potential consequences of its exclusion should be considered. Therefore, Jakarta was retained in the baseline model, while its influence was examined through an alternative specification excluding the location.

The use of HC3 standard errors also provides an important robustness consideration. Long and Ervin (2000) show that HC3-type heteroskedasticity-consistent covariance estimators can improve statistical inference in relatively small samples compared with conventional asymptotic estimators. In the present study, the difference between conventional and HC3 results reinforces the need to interpret the regression coefficients cautiously rather than relying exclusively on conventional significance tests.

Sensitivity Analysis and Stability of Model Estimates

To examine the robustness of the findings, the regression model was evaluated under alternative sample restrictions and outcome specifications. These analyses were intended to assess whether the estimated relationships remained stable under different data-quality and geographical assumptions.

Table 4. Sensitivity of overall six-predictor regression fit across alternative analytical specifications.

Analytic sample/outcome	N	Adjusted R^2	Omnibus p
39 regions; baseline provisional imputations	39	.140	.090
28 regions; no GRDP/poverty imputation	28	−.065	.634
18 high-quality, complete regions	18	.226	.183
38 regions; Jakarta excluded	38	.268	.013
26 high-quality regions; provisional imputations	26	.506	.002
39 regions; log-transformed participation	39	.146	.083

The baseline model covering all 39 regions produced an adjusted R^2 of .140 and an omnibus p-value of .090. Excluding regions with missing or imputed GRDP and poverty observations reduced the sample to 28 regions, resulting in a negative adjusted R^2 (−.065) and a nonsignificant overall model (p = .634). The higher-quality complete-case subset of 18 regions produced an adjusted R^2 of .226, although the overall model remained statistically nonsignificant (p = .183).

In contrast, excluding Jakarta increased the adjusted R^2 to .268 and produced a statistically significant overall model (p = .013). This change demonstrates that the baseline results are sensitive to the inclusion of an influential geographical observation. However, excluding Jakarta is not automatically preferable because the region is

substantively important for understanding urban access to higher education and may represent genuine contextual variation.

The 26-region higher-quality subset with provisional imputations produced a substantially higher adjusted R^2 of .506 and an omnibus p-value of .002. This result should nevertheless be interpreted cautiously because the subset differs from the 18-region complete-case specification in its sample composition and data treatment. A higher coefficient of determination does not necessarily indicate better measurement validity or greater generalizability. The contrasting results also demonstrate that the classification of data quality and the handling of provisional socioeconomic indicators can substantially influence model estimates.

The log-transformed participation model produced an adjusted R^2 of .146 and an omnibus p-value of .083, indicating that the overall model remained statistically nonsignificant at the .05 level. Taken together, these alternative specifications suggest that the estimated associations are sensitive to geographical composition, data preprocessing, and influential observations. Bollen and Jackman (1985) similarly emphasize that influential-case diagnostics should be considered alongside substantive and measurement-related explanations rather than treated as automatic grounds for case deletion.

The sensitivity analysis therefore does not identify a single regression specification that provides consistently robust evidence of regional socioeconomic or digital predictors of UT participation. Instead, it highlights the importance of improving regional data quality and validating the underlying administrative measurements before drawing stronger conclusions about the determinants of participation.

Discussion: Digital Inequality and Regional Participation in Distance Higher Education

The findings provide a qualified perspective on the relationship between regional digital infrastructure and participation in open and distance higher education. Although household Internet access and device ownership were relatively high in several locations, neither indicator demonstrated a statistically significant independent association with regional Universitas Terbuka (UT) participation in the baseline model (Marin et.al., 2026). This result is consistent with the broader understanding that digital inclusion involves more than the physical availability of devices or Internet connections. Recent research increasingly conceptualizes digital inequality as a multidimensional phenomenon involving access, material resources, digital skills, patterns of technology use, and the outcomes obtained from digital engagement (Mirazchiyski, 2025).

Michikyan et al. (2025) demonstrate that digital access, connectivity, digital skills, and learning engagement represent distinct but interconnected dimensions of digital inequality in higher education. Similarly, the World Bank (2021) identifies affordability, infrastructure quality, and digital capabilities as important barriers to digital inclusion in Indonesia. These perspectives are reinforced by van Deursen and van Dijk (2019), who explain that inequalities persist even when basic Internet access is widely available because individuals differ in device quality, material resources, maintenance expenses, and opportunities to use technology effectively. Kartiasih et al. (2023) further demonstrate substantial spatial disparities in Indonesia's regional digital development, particularly between metropolitan areas and more peripheral regions. These findings suggest that regional household access indicators may not adequately capture the actual conditions under which UT students access learning materials, participate in tutorials, or complete academic activities.

The present study does not directly measure individual digital literacy, mobile data affordability, platform engagement, or the quality of student support services. Therefore, the absence of significant coefficients for household Internet access and device ownership should be interpreted as a limitation of the available regional indicators rather than evidence that digital infrastructure has no relevance to distance education participation. Mirazchiyski (2025) emphasizes that digital literacy and the ability to use information technologies meaningfully are essential dimensions of educational digital inclusion. UNESCO (2023) also cautions that technology does not automatically produce equitable educational outcomes. Effective participation depends on the broader educational and social conditions within which technology is used. Accordingly, the findings suggest that measuring digital inclusion in distance higher education requires indicators that extend beyond household connectivity and device ownership to include the quality, affordability, and educational utilization of digital resources.

The socioeconomic results require similar caution. GRDP per capita showed a positive but statistically imprecise coefficient, while the poverty proxy showed a negative association that did not reach statistical significance. These directions are broadly compatible with the expectation that economic resources may facilitate access to higher education. Wanti et al. (2022), in their systematic review of higher education access and equity, identify financial resources, institutional conditions, government policies, and student-level circumstances as important determinants of educational participation. In the Indonesian context, Ogawa and Iimura (2010) demonstrate that household income and the educational background of household heads are positively associated with decisions to pursue tertiary education. However, the current findings do not establish

that regional economic prosperity independently explains UT participation. The provisional nature of several GRDP and poverty observations may have reduced the precision of the estimates or obscured underlying relationships.

Furthermore, regional socioeconomic indicators cannot fully represent individual students' financial circumstances. Students living in the same district may differ substantially in employment, household responsibilities, educational backgrounds, and capacity to finance their studies. Wakefield (2008) explains that ecological studies based on aggregated geographical data can lose important within-area information, creating ecological bias when group-level associations are interpreted as individual-level relationships. Crawford and McKenzie (2023) similarly illustrate how geographical location, financial constraints, social belonging, and responsibilities outside university influence the higher education experiences of mature-aged students in regional and remote communities. These perspectives highlight the importance of considering individual and community circumstances alongside regional socioeconomic indicators. Consequently, the present findings should be understood as regional contextual associations rather than evidence about individual students' educational decisions or experiences.

The sensitivity of the model to Jakarta further demonstrates the importance of geographical heterogeneity. The improvement in model fit after excluding Jakarta suggests that the regional relationships are not uniform across all locations. However, Jakarta should not be considered an erroneous observation simply because it exerts substantial statistical influence. Its distinctive urban, economic, demographic, and infrastructural characteristics may represent a meaningful part of the geographical variation in distance higher education participation. Li et al. (2022), examining distance higher education across Chinese provinces, demonstrate that the expansion of distance education does not necessarily eliminate regional inequality, as disparities may persist within and between geographical areas. Gunter (2025) also emphasizes that geographical location and spatial accessibility influence participation and engagement in distance education, particularly among students living in underserved areas. Therefore, the appropriate interpretation is that the estimated regional associations depend partly on the inclusion of highly distinctive locations, not that those locations should necessarily be excluded.

These geographical considerations are also relevant to understanding the institutional and material dimensions of distance education. Gunter et al. (2020) conceptualize distance education as a socio-material assemblage in which students, technologies, institutional resources, communication networks, and geographical environments interact to make learning possible. This perspective suggests that the effectiveness of distance education depends not only on the availability of educational services but also on the networks and resources connecting students with their institutions. In the context of UT, regional differences in learning support, technological infrastructure, institutional accessibility, and students' surrounding environments may therefore contribute to participation patterns that cannot be fully explained by the socioeconomic and digital indicators included in the present model.

Overall, the findings indicate that regional socioeconomic and digital-access indicators, as currently measured, do not provide sufficiently stable evidence to explain differences in UT participation across the 39 locations. This does not negate the importance of economic conditions or digital infrastructure. Rather, it suggests that participation in open and distance higher education involves a wider combination of geographical, institutional, demographic, and learner-level conditions than can be captured by the available regional indicators. The findings contribute to the literature by highlighting the limitations of relying exclusively on aggregated socioeconomic and household digital-access measures when assessing regional participation in distance higher education. Future research should integrate more detailed regional data with individual-level information on students' socioeconomic backgrounds, digital literacy, learning engagement, affordability, and institutional support. Such an approach would provide a more comprehensive understanding of the conditions shaping equitable participation in distance higher education.

Policy Implications for Universitas Terbuka

Although the regression results do not establish causal relationships, three practical implications can be identified for UT management.

First, strengthening administrative data governance should become a priority. The integration of student enrollment records with official regional statistics requires standardized geographical identifiers, consistent reference years, verified variable definitions, and transparent documentation of data transformations. Geographic matching should also be independently checked against official district codes wherever the underlying student information permits. These improvements would strengthen the reliability of regional participation monitoring and reduce uncertainty in future statistical analyses.

Second, student support should be informed by more detailed local evidence than household connectivity percentages alone. Information on broadband quality, mobile data costs, access to study facilities, tutorial

attendance, help-desk availability, and students' digital learning capabilities would provide a more comprehensive understanding of participation barriers. Such information could help UT identify locations where students require additional academic and technological support.

Third, regional learning services should consider the diversity of student circumstances. Low-bandwidth learning materials, flexible asynchronous tutorials, accessible digital support, and opportunities to use local learning facilities may help accommodate students who face employment, geographical, or technological constraints. These measures represent reasoned service-design priorities based on the broader literature and observed data limitations; they should not be interpreted as interventions whose effectiveness has been demonstrated by the present regression model.

Conclusions

This study examined the associations between regional socioeconomic, demographic, and digital-access conditions and UT participation across 39 designated study locations in Indonesia. The baseline six-predictor OLS model explained 27.6% of the regional variation ($R^2 = .276$; adjusted $R^2 = .140$), but the overall model was not statistically significant ($p = .090$), and none of the six predictors demonstrated a statistically significant association under HC3 robust inference. Sensitivity analyses involving alternative sample specifications ($N = 18–38$), data imputation, and the exclusion of Jakarta produced substantially different model estimates, indicating limited stability across analytical conditions. The sensitivity power analysis indicated that, with 39 observations and six predictors, the study could detect an overall regression effect of approximately Cohen's $f^2 = 0.423$ at 80% power and a 5% significance level. This suggests limited capacity to detect small associations. Jakarta was identified as an influential observation, with a Cook's distance of approximately 3.08, substantially exceeding the conventional screening threshold of $4/n = 0.103$. These findings highlight the importance of considering sample size, influential observations, and data quality when interpreting regional regression estimates. The results do not demonstrate that socioeconomic and digital conditions are irrelevant to participation in distance higher education. Rather, they indicate that the available regional secondary-data model cannot establish reliable independent associations. Future research should prioritize validated administrative–BPS geographic matching, consistent regional indicators, and direct measures of learners' socioeconomic circumstances, digital capabilities, and academic support needs. Alternative count-based models and more comprehensive learner-level data may provide a stronger basis for understanding participation patterns and informing equitable distance higher education policies.

Acknowledgments

This manuscript was developed from the 2026 applied research project at Universitas Terbuka on equity in participation in distance higher education. The authors should verify and include the applicable grant number, exact funding statement, acknowledgments, and corresponding-author details before submission.

References

Badan Pusat Statistik. (2025). Angka partisipasi kasar (APK) perguruan tinggi (PT) menurut jenis kelamin. https://www.bps.go.id/id/statistics-table/2/MTQ0NiMy/angka-partisipasi-kasar

Bollen, K. A., & Jackman, R. W. (1985). Regression diagnostics: An expository treatment of outliers and influential cases. Sociological Methods & Research, 13(4), 510–542. https://doi.org/10.1177/0049124185013004004

Bond, M., Bedenlier, S., Marín, V. I., & Händel, M. (2021). Emergency remote teaching in higher education: Mapping the first global online semester. International Journal of Educational Technology in Higher Education, 18, Article 50. https://doi.org/10.1186/s41239-021-00282-x

Brownie, S., Yan, A.-R., Broman, P., Comer, L., Blanchard, D., et al. (2025). Geographic location of students and course choice, completion, and achievement in higher education: A scoping review. Equity in Education & Society, 4(1), 92–112. https://doi.org/10.1177/27526461231200280

Cameron, A. C., & Trivedi, P. K. (2013). Regression analysis of count data (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139013567

Cook, R. D. (1979). Influential observations in linear regression. Journal of the American Statistical Association, 74(365), 169–174. https://doi.org/10.1080/01621459.1979.10481634

Crawford, N., & McKenzie, L. (2023). Localised learning: Mobilising belonging among mature-aged students in low socio-economic status regional and remote areas. Higher Education, 85, 1021–1040. https://doi.org/10.1007/s10734-022-00877-x

Cullinan, J., Flannery, D., Harold, J., Lyons, S., & Palcic, D. (2021). The disconnected: COVID-19 and disparities in access to quality broadband for higher education students. International Journal of Educational Technology in Higher Education, 18, Article 26. https://doi.org/10.1186/s41239-021-00262-1

Gunter, A. (2020). Distance education as socio-material assemblage: Place, distribution, and aggregation. Population, Space and Place, 26(3), Article e2320. https://doi.org/10.1002/psp.2320

Gunter, A. (2025). The geography of distance education: Spatial disparities, accessibility, and impact across place. South African Geographical Journal, 107(4), 498–516. https://doi.org/10.1080/03736245.2025.2472653

Hidayat, D. N., Lee, J. Y., Mason, J., & Khaerudin, T. (2022). Digital technology supporting English learning among Indonesian university students. Research and Practice in Technology Enhanced Learning, 17, Article 23. https://doi.org/10.1186/s41039-022-00198-8

Holt, D., Steel, D. G., Tranmer, M., & Wrigley, N. (1996). Aggregation and ecological effects in geographically based data. Geographical Analysis, 28(3), 244–261. https://doi.org/10.1111/j.1538-4632.1996.tb00933.x

Kartiasih, F., Djalal Nachrowi, N., Wisana, I. D. G. K., & Handayani, D. (2023). Inequalities of Indonesia's regional digital development and its association with socioeconomic characteristics: A spatial and multivariate analysis. Information Technology for Development, 29(2–3), 299–328. https://doi.org/10.1080/02681102.2022.2110556

Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. Psychological Methods, 8(3), 305–321. https://doi.org/10.1037/1082-989X.8.3.305

Li, F., Wu, F., & Wang, Z. (2022). Distance higher education and regional equality in China. European Journal of Education, 57, 273–288. https://doi.org/10.1111/ejed.12502

Loh, Y. A. C., & Chib, A. (2022). Reconsidering the digital divide: An analytical framework from access to appropriation. Information Technology & People, 35(2), 647–676. https://doi.org/10.1108/ITP-09-2019-0505

Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. The American Statistician, 54(3), 217–224. https://doi.org/10.1080/00031305.2000.10474549

MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. Journal of Econometrics, 29(3), 305–325. https://doi.org/10.1016/0304-4076(85)90158-7

Martin, F., Ceviker, E., & Gezer, T. (2026). From digital divide to digital equity: Systematic review of two decades of research on educational digital divide factors, dimensions, and interventions. Journal of Research on Technology in Education, 58(2), 396–421. https://doi.org/10.1080/15391523.2024.2425442

McIntyre, N. A. (2023). Access to online learning: Machine learning analysis from a social justice perspective. Education and Information Technologies, 28, 3787–3832. https://doi.org/10.1007/s10639-022-11280-5

Megawanti, M., Wesselink, R., Biemans, H., & den Brok, P. (2022). Determining factors of access and equity in higher education: A systematic review. Equity in Education & Society, 1(2), 279–296. https://doi.org/10.1177/27526461221092429

Michikyan, M., Subrahmanyam, K., Regan, P., Cano, M. Á., Castillo, L. G., Harkness, A., & Schwartz, S. J. (2025). Differences in and relationship between digital access, use, connectivity, skill, engagement, and motivation: Implications for digital inequality in online learning among first-generation and continuing-generation undergraduate students. Discover Education, 4, Article 9. https://doi.org/10.1007/s44217-024-00371-9

Mirazchiyski, P. V. (2025). Contemporary gaps in research on digital divide in education: A literature review. Universal Access in the Information Society, 24, 991–1008. https://doi.org/10.1007/s10209-024-01166-3

OECD. (2024). Equity in education and on the labour market: Main findings from Education at a Glance 2024 (OECD Education Policy Perspectives No. 107). OECD Publishing. https://doi.org/10.1787/b502b9a6-en

Ogawa, K., & Iimura, K. (2010). Determinants of access and equity in tertiary education: The case of Indonesia. Excellence in Higher Education, 1(1–2), 3–22. https://doi.org/10.5195/ehe.2010.12

Riley, R. D., Snell, K. I. E., Ensor, J., Burke, D. L., Harrell, F. E., Jr., Moons, K. G. M., & Collins, G. S. (2019). Minimum sample size for developing a multivariable prediction model: Part I – Continuous outcomes. Statistics in Medicine, 38(7), 1262–1275. https://doi.org/10.1002/sim.7993

UNESCO. (2020). Global education monitoring report 2020: Inclusion and education—All means all. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000373718

UNESCO. (2023). Global education monitoring report 2023: Technology in education—A tool on whose terms? UNESCO. https://doi.org/10.54676/UZQV8501

van Deursen, A. J. A. M., & van Dijk, J. A. G. M. (2015). Toward a multifaceted model of Internet access for understanding digital divides: An empirical investigation. The Information Society, 31(5), 379–391. https://doi.org/10.1080/01972243.2015.1069770

van Deursen, A. J. A. M., & van Dijk, J. A. G. M. (2019). The first-level digital divide shifts from inequalities in physical access to inequalities in material access. New Media & Society, 21(2), 354–375. https://doi.org/10.1177/1461444818797082

van Dijk, J. A. G. M. (2020). The digital divide. Polity Press.

Ver Hoef, J. M., & Boveng, P. L. (2007). Quasi-Poisson vs. negative binomial regression: How should we model overdispersed count data? Ecology, 88(11), 2766–2772. https://doi.org/10.1890/07-0043.1

Wakefield, J. (2008). Ecologic studies revisited. Annual Review of Public Health, 29, 75–90. https://doi.org/10.1146/annurev.publhealth.29.020907.090821

World Bank. (2021). Beyond unicorns: Harnessing digital technologies for inclusion in Indonesia. World Bank. https://www.worldbank.org/en/country/indonesia/publication/beyond-unicorns-harnessing-digital-technologies-for-inclusion-in-indonesia

Yunita, A., Santoso, H. B., & Hasibuan, Z. A. (2022). 'Everything is data': Towards one big data ecosystem using multiple sources of data on higher education in Indonesia. Journal of Big Data, 9, Article 91. https://doi.org/10.1186/s40537-022-00639-7