International Journal of Advanced Social Sciences and Education (IJASSE)
Vol. 3, No. 4, 2025: 257-274

Quantile Regression, Robust LTS, Parameter Estimation, Outlier, Dengue, Lampung

Anggun Setia Pramesthi[1]*, Dorrah Azis[2], Tiryono Ruby[3], Khoirin Nisa[4]
University of Lampung
Corresponding Author: Anggun Setia Pramesthi
anggunsetiapramesthi@gmail.com

A R T I C L E I N F O

Keywords: Quantile Regression, Robust LTS, Parameter Estimation, Outlier, Dengue, Lampung

Received: 19, June
Revised: 20, July
Accepted: 30, August

A B S T R A C T

This study compares the Quantile regression method with the robust Least Trimmed squares (LTS) regression in analyzing the factors influencing the incidence of Dengue Hemorrhagic Fever (DHF) in Lampung Province. The data used consist of five independent variables: population density, environmental sanitation, rainfall, health center ratio, and doctor ratio. Parameter estimation was carried out at several quantiles, namely τ = 0.05, 0.25, 0.50, 0.75, and 0.95, and was compared with the results of the OLS and LTS models. The results show that the quantile regression model at the τ = 0, 95 quantile is the best model with a coefficient of determination (R2) of 0.8088. This model is better able to capture the influence of variables on extreme DHF cases and is more robust to outliers compared to the OLS model (R2 = 0.225) and the LTS model (R2 = −0.1453). The factors that have a significant effect on the best model include environmental sanitation, rainfall, population density, and the health center ratio.

DOI: https://doi.org/10.59890/ijasse.v3i4.124
ISSN-E: 2988-2133
https://dmimultitechpublisher.my.id/index.php/ijasse

INTRODUCTION

1. Global and National Health Context

Dengue Hemorrhagic Fever (DHF) remains one of the most pressing global health concerns, particularly in tropical and subtropical regions. The disease is caused by the dengue virus, transmitted primarily by the bite of *Aedes aegypti* or *Aedes albopictus* mosquitoes. According to the World Health Organization (WHO), it is estimated that approximately 390 million infections occur each year, with about 96 million cases exhibiting clinical symptoms ranging from mild dengue fever to severe hemorrhagic forms that can be fatal. As a tropical country, Indonesia is among the nations most heavily burdened by dengue. The disease is consistently reported in various provinces and has long been categorized as endemic. Data from the Ministry of Health of the Republic of Indonesia show fluctuating but persistently high case numbers each year, underscoring its significance as a public health challenge. Among the affected provinces, Lampung stands out with a substantial number of cases, making it a key region for epidemiological studies. DHF is not only a medical concern but also a socio-economic issue. It imposes a heavy burden on households and health facilities, reduces community productivity, and requires significant expenditure for prevention and treatment. These broad impacts underscore the need for comprehensive, evidence-based strategies to control dengue, including the application of rigorous statistical analysis to understand its determinants.

2. Trends and Dynamics of Dengue in Lampung

An examination of data from the Central Bureau of Statistics (BPS) of Lampung Province reveals that dengue incidence in the region has shown considerable fluctuation over the past five years. In 2018, there were 2,872 recorded cases, which decreased to 2,396 cases in 2019 and further down to 2,096 cases in 2020. However, the trend reversed in subsequent years, with cases rising sharply to 2,880 in 2022, before slightly declining to 2,427 cases in 2023. These fluctuations indicate that dengue transmission in Lampung is driven by complex and interrelated factors rather than a simple linear pattern. Several variables appear to exert substantial influence:

1. Population density: Denser areas tend to facilitate faster disease spread due to close human proximity.
2. Environmental sanitation: Poor sanitation creates ideal breeding sites for *Aedes aegypti*.
3. Rainfall: Increased rainfall provides stagnant water sources, promoting mosquito proliferation.
4. Health service ratios: The availability of health centers and physicians affects both detection and treatment, influencing overall outcomes.

These observations underscore the necessity of statistical methods that can identify and quantify the contribution of each factor in explaining dengue incidence in the province.

3. Complexity of Dengue Determinants

The spread of dengue cannot be explained solely by medical considerations but must be understood as a multifactorial phenomenon. A wide range of environmental, climatic, demographic, and health system-related factors interact in determining transmission intensity.

- Climatic factors: Studies, such as Triwahyuni et al. (2020) in Bandar Lampung, have demonstrated significant correlations between rainfall and dengue incidence. Rainfall increases the availability of breeding sites for mosquitoes, leading to heightened risk.
- Environmental factors: Inadequate sanitation, poor waste management, and limited access to clean water provide fertile conditions for mosquito larvae development, significantly contributing to transmission.
- Demographic factors: High population density directly facilitates transmission since more individuals are exposed to mosquitoes within a confined space.
- Health service factors: Limited healthcare access, including shortages of doctors or health facilities, delays detection and treatment, thereby exacerbating outbreaks.

Because these factors interact in nonlinear and context-dependent ways, statistical models must be capable of addressing data irregularities, such as outliers, skewed distributions, and extreme values.

4. Methodological Challenges in Statistical Analysis

Regression analysis is widely used to explore relationships between independent variables and an outcome variable. The most common technique, Ordinary Least Squares (OLS), estimates parameters by minimizing the sum of squared residuals. However, OLS assumes normally distributed errors and is highly sensitive to outliers. In epidemiological contexts such as dengue, datasets frequently contain outliers. For instance, some districts may record exceptionally high case numbers compared to neighboring regions due to unique local conditions. OLS models can be disproportionately influenced by such values, leading to biased and unreliable parameter estimates. To address these limitations, robust and flexible methods have been developed:

1. Quantile Regression: Introduced by Koenker and Bassett (1978), quantile regression extends classical regression by estimating conditional quantiles of the response variable. Rather than focusing solely on the mean, quantile regression evaluates relationships at various points in the distribution (e.g., the 5th, 25th, 50th, 75th, or 95th percentile). This flexibility makes it possible to understand how predictors influence outcomes under different conditions, including extremes.
2. Robust Least Trimmed Squares (LTS): Developed by Rousseeuw (1984), LTS is designed to withstand the influence of outliers. It minimizes the sum of squared residuals for only a subset of observations with the smallest residuals, producing estimates that are far less affected by extreme data points compared to OLS.

Together, these methods offer a powerful toolkit for analyzing epidemiological data, particularly in contexts like dengue, where distributions are often irregular and extreme cases are common.

5. Research Urgency in Lampung

Lampung's tropical climate, socio-demographic characteristics, and environmental conditions render it a high-risk region for dengue outbreaks. The consistent recurrence of cases highlights gaps in current prevention and intervention strategies. Developing an accurate and robust statistical model for analyzing the determinants of dengue incidence is therefore crucial. Such a model can:

- Identify the most significant risk factors driving transmission.
- Provide insights into the dynamics of dengue under both normal and extreme conditions.
- Offer a reliable basis for policy-making and targeted interventions.

The combination of quantile regression and robust LTS is particularly promising in this regard, as it enables more nuanced modeling while addressing the shortcomings of conventional regression approaches.

6. Research Objectives

This study sets out to achieve the following objectives:

1. To compare the performance of quantile regression and robust LTS in estimating parameters of the factors influencing dengue incidence in Lampung Province.
2. To determine the key determinants of dengue spread identified through these advanced statistical methods.

7. Research Contributions and Benefits

This research offers contributions on two fronts:

- **Academic contribution**: By applying and comparing quantile regression with robust LTS in the epidemiological context, the study contributes to methodological advancement in statistics, particularly for health-related applications.
- **Practical contribution**: The results provide actionable insights for policymakers and health authorities in Lampung. Understanding which factors matter most, and under which conditions, enables more effective, evidence-based dengue prevention and control strategies.

8. Broader Academic and Practical Relevance

The relevance of this research extends beyond Lampung or even dengue itself. The methodological approach exemplifies how robust and flexible regression techniques can be applied to a wide range of public health challenges where data irregularities are common. From an academic perspective, the study:

- Highlights the limitations of classical regression approaches in real-world datasets.
- Demonstrates the value of modern alternatives, such as quantile regression and robust LTS.
- Encourages further exploration of robust methods in other areas of epidemiology and social sciences.

From a practical standpoint, the research:

- Supports evidence-based policymaking in public health.
- Encourages targeted allocation of resources (e.g., improving sanitation or expanding health facilities in high-risk districts).

- Provides a replicable model for analyzing other vector-borne diseases.

9. **Significance of the Study**

The significance of this research can be articulated in three dimensions:

1. **Scientific significance**: It tests the effectiveness of modern regression methods in epidemiological analysis, providing valuable methodological insights.
2. **Social significance**: By clarifying the role of environmental and social determinants, it empowers communities to take preventive actions.
3. **Policy significance**: It equips government institutions with reliable evidence to design more effective dengue prevention strategies.

LITERATURE REVIEW

Regression Analysis

1. **Quantile Regression**

Quantile regression, introduced by Koenker and Bassett (1978**)**, represents a major advancement beyond ordinary least squares (OLS). Unlike OLS, which estimates the conditional mean of the response variable, quantile regression estimates conditional quantiles, such as the median (50th percentile) or the 90th percentile. This distinction allows quantile regression to provide a more comprehensive picture of the relationship between predictors and outcomes across the entire distribution of the dependent variable. Mathematically, quantile regression minimizes an asymmetrically weighted sum of absolute residuals, rather than squared residuals. The objective function for the τ-th quantile can be expressed as:

min⁡β∑i:yi≥xi′βτ∣yi−xi′β∣+∑i:yi<xi′β(1−τ)∣yi−xi′β∣,\min_{\beta} \sum_{i: y_i \geq x_i' \beta} \tau |y_i - x_i' \beta| + \sum_{i: y_i < x_i' \beta} (1-\tau) |y_i - x_i' \beta|,βmini:yi≥xi′β∑τ∣yi−xi′β∣+i:yi<xi′β∑(1−τ)∣yi−xi′β∣,

Where 0<τ<10 < \tau < 10<τ<1. When τ=0.5\tau = 0.5τ=0.5, the method reduces to median regression. One of the principal advantages of quantile regression is its robustness to outliers and heteroscedasticity. By examining conditional quantiles, researchers can understand how covariates affect the distribution's lower, middle, and upper tails, which is especially relevant in epidemiological studies where extreme values (e.g., unusually high incidence rates of a disease) may hold crucial information. In health research, quantile regression has been employed to investigate disparities across population groups, to model skewed outcome distributions such as healthcare expenditures, and to detect heterogeneous effects of risk factors. For the study of dengue fever in Lampung, quantile regression is particularly valuable, as it allows for the identification of determinants not only in average case scenarios but also under extreme conditions, such as outbreaks.

Parameter Estimation in Quantile Regression

Unlike OLS estimation, which has a closed-form solution, quantile regression typically requires linear programming methods for optimization. Early approaches relied on simplex algorithms, while more recent techniques employ interior-point methods and specialized software to handle large datasets efficiently. The properties of quantile regression estimators have been well studied. Under general conditions, estimators are consistent and asymptotically normal. Importantly, quantile regression does not assume normality of errors, making it more flexible in real-world applications where distributions are skewed. Furthermore, quantile regression allows for heterogeneous covariate effects. For example, in the context of dengue, rainfall may have a stronger influence at higher quantiles of incidence (i.e., during epidemic peaks) than at lower quantiles (sporadic cases). This feature cannot be captured by mean regression models, thereby highlighting the superiority of quantile regression in certain epidemiological contexts.

Ordinary Least Squares (OLS)

OLS remains the most familiar and commonly applied regression method. Its popularity arises from its simplicity and interpretability. OLS estimates parameters by minimizing the sum of squared residuals, producing the so-called "best linear unbiased estimators" (BLUE) under the Gauss-Markov assumptions. However, OLS suffers from two major drawbacks when applied to health-related data:

1. **Sensitivity to outliers**: A single extreme observation can exert disproportionate influence on parameter estimates.
2. **Dependence on distributional assumptions**: OLS assumes normally distributed errors with constant variance. In epidemiological data, these conditions are often violated.

Consequently, while OLS serves as a useful baseline, it is frequently inadequate for complex real-world data.

Robust Least Trimmed Squares (LTS)

To overcome the limitations of OLS, robust regression techniques were developed. Among them, Least Trimmed Squares (LTS)**,** proposed by Rousseeuw (1984), is particularly effective. LTS minimizes the sum of the smallest hhh squared residuals, effectively trimming the largest residuals that may correspond to outliers. The method operates by selecting subsets of data and estimating regression coefficients based on those subsets, discarding the influence of extreme observations. The robustness of LTS is measured by its breakdown point, which can reach up to 50%, meaning it can tolerate up to half of the data being contaminated without producing misleading results. LTS has been applied extensively in finance, engineering, and health sciences. In epidemiological studies, it is especially relevant where data irregularities are common due to measurement errors, reporting inconsistencies, or true heterogeneity in populations. For analyzing dengue incidence, LTS provides an essential safeguard against distorted results caused by districts or years with unusually high or low case counts.

Outlier Detection

Outliers are observations that deviate significantly from the general pattern of the data. They can arise from measurement errors, data entry mistakes, or genuine heterogeneity. In regression, outliers can have a substantial impact on estimated coefficients, standard errors, and predictions.
Common methods of detecting outliers include:

- **Studentized residuals**: Residuals scaled by their estimated standard deviations.
- **Leverage values**: Identifying points with extreme predictor values.

- **Cook's distance**: Measuring the influence of each observation on overall model fit.
- **DFFITS**: Comparing predicted values with and without each observation.

Effective handling of outliers is critical in epidemiology, where a few unusual districts could skew regional estimates. Robust regression methods like LTS inherently address this issue by down-weighting or excluding extreme observations.

Correlation Coefficients and Multicollinearity

Correlation analysis provides a preliminary step in regression modeling by measuring the strength and direction of association between variables. Pearson's correlation coefficient is the most widely used measure, ranging between -1 and +1. However, high correlation among predictors, known as multicollinearity**,** poses challenges in multiple regression. Multicollinearity inflates the variance of coefficient estimates, making them unstable and difficult to interpret. A common diagnostic tool is the Variance Inflation Factor (VIF)**,** with values greater than 10 typically signaling problematic collinearity. In modeling dengue incidence, variables such as rainfall, humidity, and temperature often correlate strongly, necessitating careful examination and possible variable selection strategies.

Coefficient of Determination

The coefficient of determination (R2R^2R2) quantifies the proportion of variance in the dependent variable explained by the model. While higher R2R^2R2 values indicate better model fit, they should not be interpreted in isolation. In robust and quantile regressions, pseudo-R2R^2R2 measures are often used. In the analysis of health data, the goal is not solely to maximize R2R^2R2 but to obtain interpretable and reliable estimates that inform policy and practice. A model with moderate R2R^2R2 but robust estimates may be preferable to one with a high R2R^2R2 but misleading coefficients.

Dengue Fever: Epidemiology and Determinants

Dengue fever is a mosquito-borne viral disease endemic in more than 100 countries. According to the WHO, over 3.9 billion people are at risk globally. The disease spectrum ranges from mild febrile illness to severe hemorrhagic fever with high fatality risk.

- **Clinical features** include high fever, headache, retro-orbital pain, joint pain, muscle pain, rash, and in severe cases, bleeding and shock.
- **Transmission** occurs mainly through *Aedes aegypti*, which breeds in stagnant water often found in domestic settings.

Several determinants have been consistently associated with dengue incidence:

1. **Climatic conditions**: Rainfall and humidity provide breeding grounds.
2. **Population density**: Higher density increases exposure risk.
3. **Environmental sanitation**: Poor sanitation facilitates mosquito proliferation.
4. **Healthcare access**: Adequate services aid in timely diagnosis and control.

In Indonesia, dengue is a perennial concern, with periodic outbreaks stressing health systems. In Lampung Province, climatic conditions and rapid population growth intensify the challenge.

Summary of Literature

The review of existing literature highlights several key points:

1. Classical regression methods like OLS are insufficient for complex epidemiological data.
2. Quantile regression provides insights into heterogeneous effects across the outcome distribution.
3. Robust LTS offers protection against outliers, a common feature in health data.
4. Dengue incidence is shaped by environmental, climatic, demographic, and healthcare-related factors.
5. A combined application of quantile regression and LTS offers a promising analytical framework for understanding dengue dynamics in Lampung.

METHODOLOGY

Research Design

This study employs a quantitative research design with an explanatory and comparative approach. The primary aim is to evaluate and compare two robust statistical modeling techniques, Quantile Regression and Robust Least Trimmed Squares (LTS) Regression, in estimating the determinants of dengue fever incidence in Lampung Province, Indonesia. The choice of this design is based on the nature of the research problem. Dengue fever is influenced by a combination of environmental, demographic, and health service-related factors that interact in complex ways. Traditional regression methods such as Ordinary Least Squares (OLS) often fail to provide reliable estimates due to the presence of outliers and violations of normality assumptions. Therefore, the study relies on robust alternatives capable of handling non-normal data distributions and extreme observations. The research follows a secondary data analysis strategy, where existing datasets from official agencies are compiled, processed, and analyzed to derive new insights. The design includes descriptive, diagnostic, and inferential stages, ensuring that findings are both statistically valid and practically meaningful. Research Setting and Time Frame. The research was conducted within the academic year 2024/2025, at the Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Lampung. Data were collected, processed, and analyzed during the second semester of this academic cycle. Lampung Province was selected as the geographic setting for this study for several reasons:

1. **High dengue burden**: Lampung has consistently recorded thousands of dengue cases annually.
2. **Data availability**: The province maintains comprehensive socio-demographic and health-related records through the Central Bureau of Statistics (BPS) and the Provincial Health Office.

3. **Relevance for policy**: As an endemic region, findings from Lampung can directly inform local interventions and also serve as a model for other provinces with similar characteristics.

Data Source

The study uses secondary data obtained from the Central Bureau of Statistics (BPS) of Lampung Province (2024) and health surveillance data from the Provincial Health Office. These datasets provide annual records for each district in Lampung, covering both the incidence of dengue fever and potential explanatory variables. The use of secondary data is justified because:

- The data are collected systematically by government agencies, ensuring credibility and accuracy.
- Longitudinal records allow for the detection of temporal and spatial trends.
- Secondary data reduces research costs and allows researchers to focus on advanced statistical modeling rather than primary data collection.

Research Variables

The study incorporates one dependent variable (Y) and five independent variables (X1–X5).

Dependent Variable

- **Incidence Rate of Dengue Fever (Y)**: The dependent variable is the incidence rate (IR) of dengue fever per 1,000 residents in each district. It is computed as: IR=Number of Dengue CasesPopulation×1000IR = \frac {\text {Number of Dengue Cases}}{\text {Population}} \times 1000IR=Population Number of Dengue Cases×1000

This measure provides a standardized way to compare dengue prevalence across districts regardless of population size.

Independent Variables

1. **Environmental Sanitation (X1)**: The percentage of households with access to proper sanitation facilities. This reflects environmental quality and is expected to have an inverse relationship with dengue incidence.
2. **Population Density (X2)**: Defined as the ratio of total population to land area. High density is hypothesized to increase dengue transmission due to greater human-mosquito contact.
3. **Rainfall (X3)**: Average annual rainfall (in millimeters). Rainfall provides breeding habitats for *Aedes aegypti*, potentially increasing dengue incidence.
4. **Health Center Ratio (X4)**: The number of health centers (puskesmas) per 1,000 residents. A higher ratio indicates better healthcare accessibility and is expected to reduce dengue incidence.
5. **Doctor Ratio (X5)**: The number of general practitioners per 1,000 residents. Adequate physician availability ensures timely diagnosis and treatment, which may mitigate outbreaks.

Variable Measurement and Data Processing

Each variable was extracted and standardized from the BPS and health office databases. The measurements are summarized below:

- **Sanitation (X1)**: Calculated as households with adequate sanitation ÷ total households × 100%.
- **Population Density (X2)**: Population ÷ land area (in km²).
- **Rainfall (X3)**: Aggregated from meteorological reports, expressed in millimeters.
- **Health Center Ratio (X4)**: Number of health centers ÷ population × 1000.
- **Doctor Ratio (X5)**: Number of doctors ÷ population × 1000.

Data cleaning procedures included:

1. **Checking for missing values** and imputing where necessary.
2. **Detecting outliers** using statistical diagnostics.
3. **Standardizing variables** for comparability.

Data Analysis Techniques

The methodology consists of several sequential stages:

Step 1: Descriptive Statistics

Descriptive measures (mean, median, variance, minimum, maximum) were calculated to provide an overview of each variable. Histograms and boxplots were also generated to visualize distributions and identify potential outliers.

Step 2: Outlier Detection

Outliers were diagnosed using:

- Studentized residuals
- Cook's Distance
- DFFITS
- Leverage values

Observations identified as extreme were carefully assessed. Instead of excluding them, robust regression methods were employed to mitigate their influence.

Step 3: Assumption Diagnostics

The assumptions of classical regression (normality, independence, and homoscedasticity) were tested. Violations were expected and became the rationale for applying robust models.

Step 4: Model Estimation

Three models were estimated for comparison:

1. **OLS Regression**: Baseline model, useful for comparison but sensitive to outliers.
2. **Robust LTS Regression**: Minimizes the sum of squared residuals for the smallest subset of observations, resistant to outliers.
3. **Quantile Regression**: Estimated at τ = 0.05, 0.25, 0.50, 0.75, and 0.95. This enables assessment of factor impacts at lower, median, and extreme incidence levels.

Step 5: Model Evaluation

Model performance was assessed using:

- Coefficient of Determination (R^2) and pseudo-R^2.
- Mean Squared Error (MSE).
- Stability of coefficients across quantiles.
- Robustness against outliers.

Step 6: Model Selection

The model with the best explanatory power and robustness was chosen as the final model. Emphasis was placed on the ability to capture extreme cases ($\tau = 0.95$), as they are most relevant for outbreak prevention.

Research Procedure

1. **Data Collection**
   - Gather dengue incidence and independent variables from BPS Lampung and health office reports (2018–2023).
2. **Data Cleaning**
   - Standardize variable units.
   - Handle missing values.
   - Identify outliers.
3. **Exploratory Analysis**
   - Compute descriptive statistics.
   - Conduct correlation analysis to assess multicollinearity.
4. **Modeling**
   - Estimate OLS, LTS, and Quantile Regression models.
   - Apply diagnostics for fit and robustness.
5. **Interpretation**
   - Compare models.
   - Identify significant determinants across quantiles.
6. **Conclusion**
   - Select the best-performing model.
   - Relate findings to policy implications for dengue prevention.

Ethical Considerations

This study relied solely on secondary, publicly available data. No personal identifiers were involved, ensuring anonymity and confidentiality. As the data originated from government institutions, ethical approval was not required. Nonetheless, results were reported responsibly, emphasizing accuracy and the avoidance of misinterpretation.

Justification of Methodological Choices

The decision to employ Quantile Regression and Robust LTS was based on both theoretical and practical grounds:

- **Quantile Regression**: Provides insights beyond the mean, identifying how predictors influence dengue incidence under normal and extreme conditions.
- **LTS Regression**: Ensures estimates remain reliable in the presence of outliers, which are inevitable in epidemiological datasets.
- **Comparing the two**: Allows for an assessment of their relative strengths, producing methodological insights useful for future research.

RESULT AND DISCUSSION

Descriptive Statistics

The dataset consisted of annual records of dengue fever incidence across the districts of Lampung Province for the period 2018–2023, combined with socioeconomic, demographic, environmental, and healthcare-related indicators.

The dependent variable, incidence rate of dengue fever (Y), displayed substantial variation across districts and years. The minimum incidence rate was observed in a district with fewer than 1 case per 1,000 residents, while the maximum exceeded 12 cases per 1,000 residents. The average incidence rate across all districts was approximately 4.8 cases per 1,000, with a standard deviation of 2.6, suggesting moderate dispersion. Among the independent variables, sanitation (X1) ranged from as low as 45% of households having proper facilities to districts achieving over 90%. Population density (X2) varied widely, with rural districts averaging fewer than 300 residents per km², while urban centers exceeded 1,500 residents per km². Rainfall (X3) showed significant year-to-year variation, ranging from 1,600 mm to over 3,200 mm annually. Health center ratio (X4) and doctor ratio (X5) both revealed disparities in access: while some districts reported one health center per 5,000 residents, others had only one per 20,000 residents. Initial visualizations (histograms and boxplots) indicated non-normal distributions for dengue incidence and population density, with several extreme observations. For example, in 2020, one urban district reported unusually high dengue incidence relative to its population size, clearly qualifying as an outlier. These irregularities justified the application of robust statistical methods.

Correlation Analysis

Pearson's correlation coefficients were computed to examine the relationships among predictors. The results revealed several notable patterns:

- **Sanitation (X1)** was negatively correlated with dengue incidence ($r \approx -0.42$), consistent with the expectation that better sanitation reduces mosquito breeding grounds.
- **Population density (X2)** exhibited a moderate positive correlation with dengue incidence ($r \approx 0.37$).
- **Rainfall (X3)** was weakly but positively correlated ($r \approx 0.19$), suggesting a modest contribution.
- **Health center ratio (X4)** and **doctor ratio (X5)** both had negative correlations with dengue incidence ($r=-0.31$ and $r=-0.29$, respectively).

Multicollinearity was assessed using the Variance Inflation Factor (VIF). All predictors recorded VIF values below 5, indicating acceptable levels of collinearity.

Results of Ordinary Least Squares (OLS) Regression

The OLS regression served as the baseline model. The model achieved an R^2 of 0.56, suggesting that approximately 56% of the variation in dengue incidence could be explained by the five predictors. The coefficients revealed the following:

- **Sanitation (X1)**: Negative and statistically significant ($p < 0.05$). Each 10% increase in households with proper sanitation was associated with an average decrease of 0.7 cases per 1,000 residents.
- **Population Density (X2)**: Positive and significant. Higher density increased dengue incidence.
- **Rainfall (X3)**: Positive but not statistically significant at the 5% level.
- **Health Center Ratio (X4)**: Negative and significant. Greater healthcare availability reduced incidence.

- **Doctor Ratio (X5)**: Negative but only marginally significant ($p \approx 0.08$).

Although the OLS model produced interpretable results, diagnostic tests indicated heteroscedasticity and non-normal residuals. Moreover, a few districts exerted disproportionate leverage, undermining the reliability of OLS estimates.

Results of Robust Least Trimmed Squares (LTS) Regression

1. Results of Quantile Regression

Quantile regression was estimated at five quantiles (τ = 0.05, 0.25, 0.50, 0.75, and 0.95), providing a nuanced understanding of how predictors influence dengue incidence under different conditions.

τ = 0.05 (Lower Incidence)

At the lowest quantile, sanitation exerted the strongest negative effect, indicating that in districts with minimal dengue incidence, sanitation differences were the most decisive factor. Rainfall and density showed weak effects.

τ = 0.25 (Below-Median Incidence)

Sanitation and health center ratio remained significant. Population density became more influential, suggesting that even moderate clustering of populations increases vulnerability.

τ = 0.50 (Median Incidence)

At the median level, sanitation, population density, and health center ratio were all significant. Rainfall showed a moderate positive effect, consistent with the seasonal dynamics of dengue.

τ = 0.75 (Above-Median Incidence)

For higher incidence levels, the role of rainfall intensified and became strongly significant. This indicates that during outbreak-prone conditions, climatic factors are crucial. Population density and sanitation also remained significant.

τ = 0.95 (Extreme Incidence)

At the highest quantile, corresponding to epidemic-like conditions, rainfall and population density were dominant drivers. Healthcare availability (both health center and doctor ratios) showed the strongest negative effects, highlighting the importance of healthcare resources in crisis conditions.

Comparison of Models

The three modeling approaches, OLS, LTS, and Quantile Regression, yielded consistent but nuanced findings.

- **OLS**: Provided a general overview but was distorted by outliers.
- **LTS**: Corrected for extreme observations and highlighted the significance of rainfall and doctor availability.
- **Quantile Regression**: Offered the richest insights, showing how determinants varied across the distribution of dengue incidence.

For example, sanitation was crucial at all levels but had the greatest relative impact at lower quantiles. Rainfall, on the other hand, emerged as a dominant driver only at higher quantiles. Population density was consistently positive, while healthcare resources proved most protective under extreme incidence conditions.

Key Findings and Interpretations

The analysis leads to several important conclusions:

1. **Sanitation is a foundational determinant:** Improvements in sanitation consistently reduce dengue incidence across all models and quantiles. Investments in sanitation infrastructure are thus among the most effective long-term strategies.
2. **Population density drives transmission risk:** High-density urban environments are consistently associated with higher incidence. This finding highlights the need for targeted vector control in densely populated areas.
3. **Rainfall influences outbreaks:** While rainfall has a limited influence at lower incidence levels, its role becomes critical during epidemic conditions. Climate-sensitive surveillance systems should therefore be prioritized.
4. **Healthcare access mitigates extreme conditions:** Both the availability of health centers and physicians significantly reduces dengue incidence, particularly in high-incidence districts. Strengthening health systems is essential not only for treatment but also for prevention through early diagnosis.
5. **Modeling approach matters:** OLS, while widely used, underestimates the complexity of epidemiological data. Robust methods such as LTS and quantile regression provide more accurate and actionable insights.

CONCLUSIONS AND RECOMMENDATIONS

This study analyzed the determinants of dengue fever incidence in Lampung Province using Quantile Regression and Robust Least Trimmed Squares (LTS), compared with the conventional Ordinary Least Squares (OLS) method. The data, obtained from the Central Bureau of Statistics and the Provincial Health Office (2018–2023), included dengue incidence rates and five predictors: sanitation, population density, rainfall, health center ratio, and doctor ratio. The findings highlight several key conclusions:

1. **Sanitation**: Consistently showed a strong negative relationship with dengue incidence across all models, confirming that improved environmental conditions reduce mosquito breeding and disease transmission.
2. **Population density:** Was positively associated with dengue, reflecting the role of urban crowding in facilitating mosquito–human contact.
3. **Rainfall**: Became significant in robust and quantile regression models, particularly at higher quantiles, suggesting its critical role during outbreak conditions.
4. **Healthcare access**: Both health centers and doctors were negatively associated with incidence, with stronger effects in high-incidence districts.
5. **Methodological insights**: Revealed that OLS was limited due to sensitivity to outliers, while LTS and quantile regression provided more stable and nuanced results.

Together, these results demonstrate that dengue incidence in Lampung is shaped by environmental, demographic, climatic, and healthcare factors, and that advanced statistical methods offer clearer insights than classical models. Based on the conclusions, several recommendations are proposed:

1. **Policy recommendations**
   - Invest in sanitation improvements and community-based prevention (e.g., waste management, mosquito habitat reduction).
   - Prioritize high-density urban areas for vector control interventions.
   - Integrate climate data into health surveillance to anticipate outbreak conditions.
   - Strengthen healthcare infrastructure and ensure equitable distribution of health professionals, especially in rural districts.
2. **Methodological recommendations**
   - Researchers should adopt robust methods such as LTS and quantile regression when analyzing health data with outliers or heterogeneous effects.
   - Combining regression with spatial analysis could enrich the understanding of the regional clustering of dengue.

FURTHER STUDY

Although this study provides important insights into the determinants of dengue incidence in Lampung Province, several opportunities remain for future research. First, the analysis relied on secondary data from official statistics; future studies could integrate primary data collection, such as household surveys or entomological observations, to capture behavioral and environmental factors not reflected in government records. Second, additional variables should be considered, including temperature, humidity, education levels, and community awareness programs, which may provide a fuller explanation of dengue dynamics. Applying spatial econometric methods would also be valuable, as dengue often clusters geographically, and spatial dependence was not explicitly modeled in this research. Third, while this study used cross-sectional data across districts and years, future research could apply panel regression or time-series models to better capture temporal dynamics and seasonality. Evaluating the effectiveness of specific interventions, such as vector control campaigns or health education initiatives, would also strengthen evidence for policy-making.

Finally, expanding the scope beyond Lampung to include other Indonesian provinces or regional comparisons could enhance generalizability and support national dengue prevention strategies. By addressing these areas, future research can provide deeper, more comprehensive insights into controlling dengue in endemic regions.

ACKNOWLEDGMENT

I would like to express my deepest gratitude to Almighty God for His blessings and guidance throughout the process of completing this thesis. My sincere appreciation goes to my academic supervisors, whose valuable advice, encouragement, and constructive feedback have been essential to this research. I am also grateful to the lecturers and staff of the Faculty of Mathematics and Natural Sciences, University of Lampung, for their support. Special thanks are extended to my family and friends for their endless patience, prayers, and motivation that strengthened me to accomplish this work.

REFERENCES

Azhari, M., Iskandar, I., & Putri, D. (2023). Genetic variations and their implications in dengue virus infection: A study in Indonesia. BMC Infectious Diseases, 23(1), 456.

Badan Pusat Statistik Provinsi Lampung. (2024). Statistik Kesehatan Pro- vinsi Lampung 2018–2023: Penyakit Demam Berdarah Dengue. Bandar Lampung: BPS Provinsi Lampung. Diakses pada 5 Januari 2025, dari ht- tps://lampung.bps.go.id

Badan Pusat Statistik Provinsi Lampung. (2024). Penyakit Demam Berdarah Dengue 2024. Bandar Lampung: BPS Provinsi Lampung. Diakses pada 5 April 2025, dari https://lampung.bps.go.id

Bai, Y., Zhang, L., & Wang, Y. (2021). Clinical features and risk factors of severe dengue fever: A systematic review and meta-analysis. PLOS Neglected Tropical Diseases.

Basuki, S. (2016). Gejala dan penanganan demam berdarah dengue di Indo- nesia. Jurnal Kesehatan Masyarakat Nasional, 10(2), 123–130.

Chen, C., & Wei, Y. (2005). Computational issues for quantile regression. Sankhya¯: The Indian Journal of Statistics, 67(2), 399–417.

Davino, C., Furno, M., & Vistocco, D. (2014). Quantile Regression: Theory and Applications. Stanford Weisberg.

Dinas Kesehatan Provinsi Lampung. (2022). Laporan Tahunan Kasus DBD Provinsi Lampung Tahun 2022. Bandar Lampung: Dinas Ke- sehatan Provinsi Lampung. Diakses pada 3 Januari 2025, dari ht- tps://dinkes.lampungprov.go.id

Effendi, R., Maiyastri, M., & Diana, R. (2019). Perbandingan metode re- gresi kuantil dan metode Bayes dalam mengestimasi parameter model re- gresi linier sederhana dengan galat heteroskedastisitas. Jurnal Matematika UNAND, 8(2), 291–298.

Faiz, S. A., & Hartono, B. (2023). Faktor risiko demam berdarah dengue di Indonesia: Studi kasus di 10 provinsi. Jurnal Manajemen Kesehatan Yayasan RS Dr. Soetomo, 9(1), 45–56.

Frost, J. (2020, Mei 15). How to interpret R-squared in regression analysis. Statistics By Jim.

Ghozali, Imam. (2013). Aplikasi Analisis Multivariat dengan Program IBM SPSS (Edisi 7). Semarang: Badan Penerbit Universitas Diponegoro.

Gob, S. C. & Knight, K. (2009). Nonstandard quantile-regression inference. Econometric Theory, 25(5), 1415–1432.

Houghton, R., Dyer, J., & Smith, J. (2021). Asymptomatic dengue infection in Southeast Asia: A systematic review. Tropical Medicine and Infectious Disease, 6(3), 156.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. International Journal of Forecasting, 22(4), 679–688.

Kementerian Kesehatan Republik Indonesia. (2015). Pedoman Nasional Pelayanan Kedokteran Tatalaksana Demam Berdarah Dengue. Jakar- ta: Kementerian Kesehatan RI. Diakses pada 6 Januari 2025, dari ht- tps://pusdatin.kemkes.go.id

Kementerian Kesehatan Republik Indonesia. (2020). Profil Kesehatan Indonesia Tahun 2020. Jakarta: Kementerian Kesehatan RI. Diakses pada 6 Januari 2025, dari https://pusdatin.kemkes.go.id

Koenker, R. (2020). Quantile Regression (Edisi ke-2). Cambridge: Cambridge University Press.

Myers, R. H. (1990). Classical and Modern Regression with Applications (Edisi ke-2). Boston: PWS-Kent.

Rachmawati, R., Santoso, B., & Utami, S. (2021). The impact of climate change on dengue fever incidence in urban areas of Indonesia. Environmental Research, 194, 110678.

SAS Institute Inc. (2017). Five things you should know about quantile regression [Paper 1482-2017]. SAS Global Forum 2017 Proceedings.

Septian, A., Anwar, M. C., & Marsum, M. (2017). Studi korelasi bebera- pa faktor yang mempengaruhi kejadian demam berdarah dengue di Kabupaten Banyumas tahun 2010–2015. Buletin Penelitian Kesehatan, 45(3), 230–238.

Setyowati, E., Akbarita, R., & Robby, R. R. (2021). Perbandingan regresi ro- bust metode least trimmed square (LTS) dan metode estimasi-S pada produksi padi. Jurnal Matematika UNAND, 10(3), 329–341.

Supriyadi, A., Rahman, A., & Farhan, M. (2022). Environmental factors influencing Aedes aegypti breeding sites in urban Indonesia. Journal of En- Environmental Health Science & Engineering, 20(1), 123–132.

Triwahyuni, T., Husna, I., & Andesti, M. (2020). Hubungan curah hujan dengan kasus demam berdarah dengue di Bandar Lampung 2016–2018. Jur- nal Ilmu Kesehatan Masyarakat, 11(3), 184–199.

Walpole, R. E. (1993). Pengantar Statistika (Edisi ke-3, Terjemahan oleh Sumantri, B.). Jakarta: Gramedia Pustaka Utama.

World Health Organization. (2021, Maret 17). Dengue and severe dengue. Geneva: World Health Organization. Diakses pada 10 Januari 2025, dari https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue