MethodsX 14 (2025) 103082

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/methodsx

# Modeling the hospitalization time of stroke patients at Abdul Wahab Sjahranie Hospital Samarinda using the Weibull Regression Model☆,★

Suyitno, Darnah*, Andrea Tri Rian Dani, Nurul Tri Oktavia

*Statistics Study Program, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Mulawarman University, Indonesia*

ARTICLE INFO

*Method name:*
*Weibull Regression Model and Applications*

*Keywords:*
Newton–Raphson iterative
Risk
Stroke patients
Survival data
Weibull Regression Model

ABSTRACT

The Weibull regression model is a regression model derived from the Weibull distribution, where the Weibull distribution is influenced by covariates. In this study, parameter estimation for the Weibull regression model was conducted using the Maximum Likelihood (ML) estimation. The aim of the study is to develop a Weibull regression model based on the hospitalization time of stroke patients at Abdul Wahab Sjahranie Hospital, Samarinda, during the period of 2021–2022, and to identify the factors affecting it. The event of interest in this study is patient recovery. The results indicate that the ML estimator of the Weibull regression model was obtained numerically using the Newton-Raphson iterative. The factors influencing the Weibull regression model include age, body mass index (BMI), and a history of diabetes mellitus. An increase in patient age and a history of diabetes mellitus are associated with an increase in the probability of the patient not recovering, a decrease in the likelihood of recovery, a lower recovery rate, and a longer recovery time. In contrast, an increase in BMI is associated with a decrease in the probability of the patient not recovering, an increase in the likelihood of recovery, a higher recovery rate, and a shorter recovery time. Some highlights in this article, the proposed method are:

- We present The Weibull distribution influenced by covariates is called the Weibull regression model
- The potential recovery of stroke disease and the factors that influence it can be analyzed through Weibull regression modeling.
- The chance of a patient not recovering is modeled through a Weibull survival regression model, the chance of a patient recovering is modeled through a Weibull cumulative distribution regression model, the patient's recovery rate is modeled through a Weibull hazard regression model, and the average patient hospitalization time is modeled through a Weibull mean regression model.

☆ **Related research article:** None.
★ ***For a published article:***
1. Suyitno, Purhadi, Sutikno, and Irhamah, "Multivariate Weibull regression model," *Far East Journal of Mathematical Sciences*, vol. 101, no. 9, pp. 1977–1992, May 2017, doi:10.17654/MS101091977.
2. Suyitno and N. W. W. Sari, "Parameter estimation of mixed geographically weighted weibull regression model," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Aug. 2019. doi:10.1088/1742-6596/1277/1/012046.

* Corresponding author.
*E-mail address:* darnah.98@gmail.com.

https://doi.org/10.1016/j.mex.2024.103082
Received 10 October 2024; Accepted 4 December 2024
Available online 12 December 2024

Specifications table

| | |
|---|---|
| Subject area: | Mathematics and Statistics |
| More specific subject area: | *Survival Analysis; Weibull Regression Model; Health* |
| Name of your method: | *Weibull Regression Model and Applications* |
| Name and reference of original method: | Suyitno, Purhadi, Sutikno, and Irhamah, "Multivariate Weibull regression model," *Far East Journal of Mathematical Sciences*, vol. 101, no. 9, pp. 1977–1992, May 2017, doi: 10.17654/MS101091977.<br>Suyitno and N. W. W. Sari, "Parameter estimation of mixed geographically weighted weibull regression model," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Aug. 2019. doi: 10.1088/1742-6596/1277/1/012046. |
| Resource availability: | *The research data is secondary data, namely medical record data for inpatients with stroke patients obtained from Abdul Wahab Sjahranie Samarinda Hospital in 2021–2022. The study sample size was 62 samples. The research data consisted of time data, namely data on the length of time patients with stroke were hospitalized ($Y$) and covariate data, namely age ($X_1$), gender ($X_2$), type of stroke ($X_3$), HDL levels ($X_4$), LDL levels ($X_5$), body mass index ($X_6$), systolic pressure ($X_7$), diastolic pressure ($X_8$), history of diabetes mellitus ($X_9$), and history of heart disease ($X_{10}$). The data analysis technique used was Weibull regression modeling, and computation using Octave software.* |

## Background

Survival analysis is a collection of statistical procedures for analyzing time data, which is the duration of time until the occurrence of a certain event [1,2]. The time variable in survival analysis is often referred to as survival time [3,4,5]. Time data consists of complete time data and incomplete time data. Complete time data is time data from individuals who experience events, while incomplete time data consists of censored data or truncated data. There are three types of censoring in survival analysis, that is right censoring, left censoring, and interval censoring [6]. The time data in this study is right censored time data. Time data in the field is often found to follow the Weibull distribution pattern [7].

The Weibull distribution initially contains three parameters, namely the scale parameter, shape parameter, and location parameter [8,9]. One special form of the Weibull distribution is the scale-shape version of the Weibull distribution, which is a Weibull distribution that contains two parameters, namely the scale and shape parameters [7]. Discussions about data distribution are generally limited to estimating distribution parameters and testing data distribution patterns. In fact, time data in the field is influenced by external factors (covariates), therefore, it is necessary to develop from the Weibull distribution to the Weibull distribution model influenced by covariates [1,10]. The Weibull distribution influenced by covariates is called the Weibull regression model [11,12].

The Weibull regression model in this study was applied to data on the hospitalization time of patients with stroke [13,14]. According to the Indonesian Ministry of Health, stroke is one of the second highest causes of death after coronary heart disease and the third cause of disability in the world [15,16]. According to the Indonesian Ministry of Health, based on the Basic Health Research (Riskesdas) in 2013 the national stroke prevalence was 12.1 per mile, while in the 2018 Riskesdas the stroke prevalence was 10.9 per mile, where the highest stroke prevalence was in East Kalimantan Province at 14.7 per mile and the lowest stroke prevalence was in Papua Province at 4.1 per mile.

The Indonesian government has made several efforts to deal with stroke disease and reduce stroke mortality by developing services and providing education about maintaining a healthy body [17,18]. Another effort that can be made to reduce stroke mortality is to provide information to the government and the public about the factors that affect stroke disease and determine the potential recovery of stroke patients through statistical modeling, namely Weibull regression modeling. The potential recovery of stroke disease and the factors that influence it can be analyzed through Weibull regression modeling. The chance of a patient not recovering is modeled through a Weibull survival regression model, the chance of a patient recovering is modeled through a Weibull cumulative distribution regression model, the patient's recovery rate is modeled through a Weibull hazard regression model, and the average patient hospitalization time is modeled through a Weibull mean regression model.

## Method details

### *Weibull distribution*

Suppose Y is a non-negative continuous random variable with a scaled version of the Weibull distribution, then the probability density function (FKP) of Y is

$$f(y) = \frac{\gamma}{\lambda}\left(\frac{y}{\lambda}\right)^{\gamma-1}\exp\left[-\left(\frac{y}{\lambda}\right)^{\gamma}\right]. \tag{1}$$

with $\gamma$ is the shape parameter, $\lambda$ is the scale parameter with $y \geq 0; 0 < \gamma, \lambda < \infty$ [19,20].

The cumulative distribution function of the scale-shape version of the Weibull distribution is given by

$$F(y) = P(Y \leq y) = 1 - \exp\left[-\left(\frac{y}{\lambda}\right)^{\gamma}\right], \tag{2}$$

and the survival function is defined by

$$S(y) = P(Y > y) = \exp\left[-\left(\frac{y}{\lambda}\right)^{\gamma}\right], \tag{3}$$

Based on Eq. (1) and Eq. (2), the hazard function is obtained, namely

$$h(y) = \frac{f(y)}{S(y)} = \frac{\gamma}{\lambda}\left(\frac{y}{\lambda}\right)^{\gamma-1} = \gamma\lambda^{-\gamma}y^{\gamma-1}. \tag{4}$$

The mean of a random variable $Y$ with a scale-shape version of the Weibull distribution is

$$\mu_Y = E(Y) = \lambda\Gamma\left(\frac{1}{\gamma}+1\right), \tag{5}$$

with $\Gamma$ is the gamma function [12].

One method of estimating Weibull distribution parameters is the maximum likelihood estimation (MLE) method. The first step in the MLE method is to define the likelihood function. Suppose the sample data is $y_1, y_2, \ldots, y_n$ independent and identically distributed namely $y_i \sim W(\lambda,\gamma),\ i = 1,2,3,\ldots,n$. Based on the FKP in Eq. (1), the likelihood function is defined by

$$L(\theta_1,\mathbf{y}) = \prod_{i=1}^{n} f(\theta_1,\mathbf{y}) = \prod_{i=1}^{n}\left(\frac{\gamma}{\lambda}\left(\frac{y_i}{\lambda}\right)^{\gamma-1} exp\left[-\frac{y_i}{\lambda}\right]^{\gamma}\right), \tag{6}$$

with $\theta_1 = [\lambda\ \gamma]^T$. It is known that $\hat{\boldsymbol{\theta}}_1 = [\hat{\lambda}\ \hat{\gamma}]^T$ which maximizes the likelihood function also maximizes the log-likelihood function and is $\hat{\boldsymbol{\theta}}_1$ is easily obtained through the log-likelihood function. Applying the natural logarithm to both segments of the likelihood function in Eq. (6) obtained the log-likelihood function, namely

$$\begin{aligned}\ell(\boldsymbol{\theta}_1,\mathbf{y}) &= \ln\left[L(\boldsymbol{\theta}_1,\mathbf{y})\right] = \sum_{i=1}^{n}\ln\left[f(\boldsymbol{\theta}_1,\mathbf{y})\right]\\ &= \sum_{i=1}^{n}\left(\left(\ln\gamma - \ln\lambda + (\gamma-1)\left[\ln y_i - \ln\lambda\right]\right) - \left(\frac{y_i}{\lambda}\right)^{\gamma}\right),\end{aligned} \tag{7}$$

ML estimator vector $\hat{\boldsymbol{\theta}}_1 = [\hat{\lambda}\ \hat{\gamma}]^T$ obtained from the solution of the first derivative of the log-likelihood function for all component parameters $\boldsymbol{\theta}_1$ then equated to zero ($\mathbf{0}$), namely

$$\frac{\partial\ell(\boldsymbol{\theta}_1,\mathbf{y})}{\partial\boldsymbol{\theta}_1} = 0 \tag{8}$$

The left segment of the Eq. (8) is called the gradient vector ($\mathbf{g}$), namely

$$\mathbf{g}(\boldsymbol{\theta}_1) = \frac{\partial\ell(\boldsymbol{\theta}_1,\mathbf{y})}{\partial\boldsymbol{\theta}_1} = \left[\frac{\partial\ell(\theta_1,\mathbf{y})}{\partial\lambda} \quad \frac{\partial\ell(\theta_1,\mathbf{y})}{\partial\gamma}\right]^T \tag{9}$$

It is known that Eq. (8) is a likelihood equation consisting of nonlinear equations so that the exact solution to get the ML estimator ($\hat{\boldsymbol{\theta}}_1$) cannot be done analytically. An alternative method to solve the likelihood equation in Eq. (1) to obtain the ML estimator is to use the Newton-Raphson iteration method. The Newton-Raphson iteration algorithm is

$$\hat{\boldsymbol{\theta}}_1^{(q+1)} = \hat{\boldsymbol{\theta}}_1^{(q)} - \left[\boldsymbol{H}\left(\hat{\boldsymbol{\theta}}_1^{(q)}\right)\right]^{-1}\mathbf{g}\left(\hat{\boldsymbol{\theta}}_1^{(q)}\right),\ q = 0,\ 1,\ 2,\ \ldots, \tag{10}$$

with $\mathbf{g}(\boldsymbol{\theta}_1)$ is the gradient vector given by Eq. (9) and $\boldsymbol{H}(\boldsymbol{\theta}_1)$ is the Hessian matrix, which is the matrix of second-order partial derivatives of the log-likelihood function on the combination of the components of the parameter vector $\theta_1$. The general form of the Hessian matrix is

$$\boldsymbol{H}(\boldsymbol{\theta}_1) = \begin{bmatrix}\frac{\partial^2\ell(\theta_1,\mathbf{y})}{\partial\lambda^2} & \frac{\partial^2\ell(\theta_1,\mathbf{y})}{\partial\lambda\partial\gamma}\\ \frac{\partial^2\ell(\theta_1,\mathbf{y})}{\partial\gamma\partial\lambda} & \frac{\partial^2\ell y}{\partial\gamma^2}\end{bmatrix}. \tag{11}$$

*Testing distribution of hospitalization time data*

Data distribution testing can be done using the Kolmogorov-Smirnov test. Suppose you want to test whether sample data with cumulative distribution function $\hat{F}(y)$ comes from a population with unknown $F(y)$ distribution. The distribution testing hypothesis formulation is as follows:

$H_0 : F(y) = \hat{F}(y)$

(Sample data is drawn from a population with distribution function $\hat{F}(y)$)

$H_1 : F(y)\hat{F}(y)$

(Sample data is drawn from a population with distribution function not $\hat{F}(y)$)

Test statistic is

$$GD = \sup\left|\hat{F}(y) - G(y)\right|, \tag{12}$$

with,

$$G(y_i) = \frac{\text{the number of } Y \text{ observations data } \leq y_i}{n}. \tag{13}$$

*Multicollinearity detection*

Multicollinearity cases can be detected using the Variance Inflation Factor (VIF) value [21]. A regression model is said to be free of multicollinearity if it has a VIF value of no more than 10. VIF for covariates $X_k$ is calculated using the formula

$$VIF_k = \frac{1}{1 - R_k^2}, \tag{14}$$

where $R_k^2$ is the coefficient of determination of the auxiliary regression model, which is the regression model of covariate $X_k$ regressed on other covariates. The formula to get the value $R_k^2$ is

$$R_k^2 = 1 - \frac{\sum_{i=1}^{n} \left(x_{ki} - \hat{X}_{ki}\right)^2}{\sum_{i=1}^{n} \left(x_{ki} - \bar{X}_k\right)^2}, \tag{15}$$

with,
$R_k^2$: Coefficient of determination of the covariate $X_k$ regressed on the other covariates
$x_{ki}$: Value of covariate $X_k$ at observation $i$
$\hat{X}_{ki}$: Predicted value of covariate $X_k$ at observation $i$
$\bar{X}_k$: Average of observation data $X_k$
$n$: Sample size

*Weibull Regression Model*

The Weibull regression model is mathematically formulated from the Weibull distribution with the scale parameter ($\lambda$) expressed in the regression model [22]. It is known that the scale parameter in Eq. (1) is positive real value so that it can be expressed as a function of regression parameters, namely

$$\lambda(\boldsymbol{x}) = \exp\left[\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p\right] = \exp\left[\boldsymbol{\beta}^T \boldsymbol{x}\right], \tag{16}$$

with $\boldsymbol{\beta}^T = [\beta_0\ \beta_1\ \ldots\ \beta_p]$ is a vector of $p+1$ dimensional regression parameters and $\boldsymbol{x} = [1\ X_1\ \ldots\ X_p]^T$. The Weibull survival regression model is obtained from substituting Eq. (16) into Eq. (3), namely

$$S(y, \boldsymbol{x}) = \exp\left[-\left(\frac{y}{\lambda(\boldsymbol{\beta}, \boldsymbol{x})}\right)^{\gamma}\right] = \exp\left(-y^{\gamma} \exp\left[-\gamma \boldsymbol{\beta}^T \boldsymbol{x}\right]\right). \tag{17}$$

The Weibull cumulative distribution regression model is obtained from substituting Eq. (16) into Eq. (2), namely

$$F(y, \boldsymbol{x}) = 1 - S(y, \boldsymbol{x}) = 1 - \exp\left(-y^{\gamma} \exp\left[-\gamma \boldsymbol{\beta}^T \boldsymbol{x}\right]\right) \tag{18}$$

The Weibull hazard regression model is obtained from substituting Eq. (16) into Eq. (4), namely

$$h(y, \boldsymbol{x}) = \gamma y^{\gamma - 1} \exp\left[-\gamma \boldsymbol{\beta}^T \boldsymbol{x}\right] \tag{19}$$

The mean Weibull regression model is obtained from substituting Eq. (16) into Eq. (5), namely

$$\mu_y(y, \boldsymbol{x}) = \Gamma\left(\frac{1}{\gamma} + 1\right) \exp\left[\boldsymbol{\beta}^T \boldsymbol{x}\right] \tag{20}$$

Based on Eq. (6) and Eq. (1), the FKP that is directly affected by covariates is obtained, namely

$$f(y, \boldsymbol{x}) = \gamma y^{\gamma - 1} \exp\left[-\gamma \boldsymbol{\beta}^T \boldsymbol{x}\right] \exp\left(-y^{\gamma}\ exp\left[-\gamma \boldsymbol{\beta}^T \boldsymbol{x}\right]\right) \tag{21}$$

The Weibull regression model estimation method uses MLE method. The first step in the MLE method is to define the likelihood function. Suppose given $n$ data $(y_i, \boldsymbol{x}_i, \delta_i)$, $i = 1, 2, \ldots, n$, with time data for the $i$ individual is $y_i$ where the probability of the $i$ individual experiencing an event with status $\delta_i = 1$ is $P(Y = y_i) = f(y_i)$. The probability of the $i$ individual surviving with status $\delta_i = 0$ is $P(Y > y_i) = S(y_i)$. The likelihood function is defined by

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i)^{\delta_i} \left(S(y_i)\right)^{1-\delta_i} = \prod_{i=1}^{n} h(y_i)^{\delta_i}\ S(y_i), \tag{22}$$

with $\boldsymbol{\theta} = [\gamma\ \beta_0\ \beta_1\ \ldots\ \beta_p]^T$ . [4] Based on Eq. (17) and Eq. (19), the likelihood function in Eq. (22) is written into

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left[\gamma y_i^{\gamma-1} \exp\left(-\gamma \boldsymbol{\beta}^T \boldsymbol{x}_i\right)\right]^{\delta_i}\ \exp\left[-y_i^{\gamma} \exp\left(-\gamma \boldsymbol{\beta}^T \boldsymbol{x}_i\right)\right], \tag{23}$$

with $\boldsymbol{x_i} = [1\ x_{1i}\ x_{2i}\ \ldots\ x_{pi}]^T$. The ML estimator of the Weibull regression model is the vector value $\hat{\boldsymbol{\theta}}$ that maximizes the likelihood function in Eq. (23) and maximizes the natural logarithm function (*log-likelihood*). The log-likelihood function based on the likelihood function is

$$\ell(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta})$$

$$= \sum_{i=1}^{n} \left( \delta_i \left[ \ln \gamma + (\gamma - 1) \ln y_i - \gamma \boldsymbol{\beta}^T \boldsymbol{x}_i \right] - y_i^{\gamma} \exp\left( -\gamma \boldsymbol{\beta}^T \boldsymbol{x}_i \right) \right). \quad (24)$$

The ML estimator ($\hat{\boldsymbol{\theta}}$) is determined by solving the following likelihood equation

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0, \quad (25)$$

where $\mathbf{0}$ is a zero vector of dimension $p+2$ and the left segment in Eq. (25) is a gradient vector of dimension $p+2$, namely

$$\mathbf{g}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \left[ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \gamma} \ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_0} \ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_1} \cdots \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_p} \right]^T. \quad (26)$$

Based on Eq. (24), the likelihood in Eq. (25) consists of nonlinear equations, so the exact solution of the likelihood Eq. (25) to obtain the exact ML estimator cannot be found analytically. One method that can be used to obtain the ML estimator numerically is the Newton Raphson iteration method. To determine the ML estimator with the Newton-Raphson iteration method requires the calculation of the gradient vector and Hessian matrix. Gradient vector $\mathbf{g}(\boldsymbol{\theta})$ is given by Eq. (26) and Hessian matrix ($\mathbf{H}(\boldsymbol{\theta})$) is a matrix of second-order derivatives of the log-likelihood function for all combinations of vectors $\boldsymbol{\theta}$. The general form of the Hessian matrix to obtain the ML estimator of the Weibull regression model is as follows

$$\mathbf{H}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma^2} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma \partial \beta_0} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma \partial \beta_1} & \cdots & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \gamma \partial \beta_p} \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_0 \partial \gamma} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_0^2} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_0 \partial \beta_1} & \cdots & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_0 \partial \beta_p} \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_1 \partial \gamma} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_1 \partial \beta_0} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_1^2} & \cdots & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_1 \partial \beta_p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_p \partial \gamma} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_p \partial \beta_0} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_p \partial \beta_1} & \cdots & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_p^2} \end{bmatrix} \quad (27)$$

The Newton-Raphson iteration algorithm to obtain $\hat{\boldsymbol{\theta}}$, namely

$$\hat{\boldsymbol{\theta}}^{(q+1)} = \hat{\boldsymbol{\theta}}^{(q)} - \left[ \boldsymbol{H}\left( \hat{\boldsymbol{\theta}}^{(q)} \right) \right]^{-1} \mathbf{g}\left( \hat{\boldsymbol{\theta}}^{(q)} \right), \ q = 0,\ 1,\ 2,\ \ldots,\ p \quad (28)$$

The initial stage of the Newton-Raphson iteration is to determine the initial vector $\hat{\boldsymbol{\theta}}^{(0)} = [\hat{\gamma}_0^{(0)}\ \hat{\beta}_0^{(0)}\ \ldots\ \hat{\beta}_p^{(0)}\ ]^T$. Iteration is stopped at iteration to $q+1$, if $\|\hat{\boldsymbol{\theta}}^{(q+1)} - \hat{\boldsymbol{\theta}}^{(q)}\| < \varepsilon$, with $\varepsilon$ is a small positive real number, e.g. $\varepsilon = 10^{-12}$. Based on the Hessian matrix in Eq. (27), the Fisher information matrix is obtained, namely

$$\left[ \boldsymbol{I}_f(\hat{\boldsymbol{\theta}}) \right] = -\boldsymbol{E}\left[ \mathbf{H}(\hat{\boldsymbol{\theta}}) \right] = -\left[ \mathbf{H}(\hat{\boldsymbol{\theta}}) \right] \quad (29)$$

The inverse of the Fisher information matrix is the variance-covariance matrix of the estimator vector $\boldsymbol{\theta}$, i.e. [23]

$$cov(\hat{\boldsymbol{\theta}}) = \left[ \boldsymbol{I}_f(\hat{\boldsymbol{\theta}}) \right]^{-1} \quad (30)$$

where $\hat{\boldsymbol{\theta}} \sim \boldsymbol{N}(\boldsymbol{\theta},\ [\boldsymbol{I}_f(\hat{\boldsymbol{\theta}})]^{-1})$.

*Hypothesis testing of Weibull regression parameters*

Testing regression parameters simultaneously aims to determine whether the estimated parameters provide a fit regression model or not [23]. The hypothesis for testing regression parameters simultaneously is

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0$ (Weibull regression model is not feasible (not fit))

$H_1$ : at least one $\beta_k 0; k = 1,\ 2, \ldots,\ p$ (Weibull regression model is feasible (fit))

The test statistic is determined using the likelihood ratio test method, i.e.

$$G = 2\left[ \ell(\hat{\theta}) - \ell(\hat{\omega}) \right], \quad (31)$$

with $\hat{\theta} = [\widehat{\gamma},\ \hat{\beta}_0,\ \hat{\beta}_1,\ \ldots, \hat{\beta}_p]^T$ is the set of parameters under the population model that maximizes the log-likelihood function in Eq. (24). The maximum value of the log-likelihood function under the population model is

$$\ell(\hat{\theta}) = \sum_{i=1}^{n} \left( \hat{\delta}_i \left[ \ln \hat{\gamma} + (\hat{\gamma} - 1) \ln y_i - \hat{\gamma} \hat{\beta}^T x_i \right] - y_i^{\gamma} \exp\left( -\hat{\gamma} \hat{\beta}^T x_i \right) \right). \quad (32)$$

$\hat{\omega} = [\hat{\gamma}_0,\ \hat{\beta}_{00}]^T$ is the set of parameters under $H_0$ and based on Eq. (24) the maximum value of log-likelihood function under $H_0$ is

$$\ell(\hat{\omega}) = \sum_{i=1}^{n} \left( \hat{\delta}_i \left[ \ln \hat{\gamma} + (\hat{\gamma} - 1) \ln y_i - \hat{\gamma} \hat{\beta}_{00} \right] - y_i^{\gamma} \exp\left( -\hat{\gamma} \hat{\beta}_{00} \right) \right) \quad (33)$$

The test statistic G given by Eq. (31) is $\chi_p^2$ distributed. The test statistic G in Eq. (31) can be estimated by

$$G = \hat{\beta}^T \left[I_{22}(\hat{\theta})\right]^{-1} \hat{\beta}, \tag{34}$$

with $\hat{\beta} = [\hat{\beta}_1\ \hat{\beta}_2\ \ldots\ \hat{\beta}_p]^T$ and $[I_{22}(\hat{\theta})]$ is obtained from the inverse matrix of Fisher information given by Eq. (30) by removing the first row and the first column. Critical areas in this test is to reject $H_0$ at the significance level $\alpha$ if the value $G \geq \chi^2_{(\alpha,p)}$ or $p_{value} < \alpha$, where

$$p_{value} = P(G_v > G) \tag{35}$$

with $G_v$ is a distributed variable $\chi^2_{(p)}$ and $G$ is the value of test statistic in Eq. (31).

Partial regression parameter testing aims to determine whether covariates individually affect the regression model [23,24]. The hypothesis for testing the regression parameter $\beta_k$ with $k = 0,\ 1,\ 2,\ \ldots,\ p$ is

$H_0 : \beta_k = 0$ (Covariates $X_k$ has no effect on the Weibull regression model)

$H_1 : \beta_k 0$ (Covariates $X_k$ affects the Weibull regression model)

Test statistics on partial testing using Wald statistic, namely

$$W_0 = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \sim N(0,1), \tag{36}$$

where $SE(\hat{\beta}_k) = \sqrt{var(\hat{\beta}_k)}$ with $var(\hat{\beta}_k)$ is the main diagonal element to $k+1$ of the matrix $cov(\hat{\theta})$ in Eq. (30). Critical areas of this test is $H_0$ rejected at the significance level $\alpha$ if the value $|W_0| > Z_{1-\frac{\alpha}{2}}$ or $p_{value} < \alpha$, where

$$p_{value} = 1 - 2p\left(Z > |W_0|\right), \tag{37}$$

**$Z$ is a standard normal distributed random variable and $W_0$ is the value of the test statistic in Eq. (36).**

*Weibull regression model interpretation*

The interpretation of the Weibull regression model uses ratio values, namely the Weibull survival regression ratio value, the Weibull hazard regression ratio value and the Weibull mean regression ratio value [23,25]. The ratio values of Weibull survival regression based on continuous covariates $X_k$ is

$$RS_{X_k} = \frac{\exp\left[-y_i^{\hat{\gamma}} \exp\left[-\hat{\gamma}\left(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \ldots + \hat{\beta}_k\left(X_k + 1\right) + \ldots + \hat{\beta}_p X_p\right)\right]\right]}{\exp\left[-y_i^{\hat{\gamma}} \exp\left[-\hat{\gamma}\left(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \ldots + \hat{\beta}_k X_k + \ldots + \hat{\beta}_p X_p\right)\right]\right]} \tag{38}$$

Weibull cumulative distribution regression ratio values based on continuous covariates $X_k$ is

$$RF_{X_k} = \frac{F\left(y|X_k + 1\right)}{F(y)} = \frac{1 - S\left(y|X_k + 1\right)}{1 - S(y)} \tag{39}$$

Weibull hazard regression ratio values based on continuous covariates $X_k$ is

$$Rh_{X_k} = \frac{h\left(y|X_k + 1\right)}{h(y)} = \exp\left[-\hat{\gamma}\hat{\beta}_k\right] \tag{40}$$

and the ratio values of mean Weibull regression based on continuous covariates $X_k$ is

$$R\mu_{X_k} = \frac{\exp\left[\hat{\beta}_k\left(X_k + 1\right)\right]}{\exp\left[\hat{\beta}_k X_k\right]} = \exp\left[\hat{\beta}_k\right] \tag{41}$$

Weibull survival regression ratio values based on nominal covariates $X_k$ is

$$RS_{X_k} = \frac{\exp\left[-y_i^{\hat{\gamma}} \exp\left[-\hat{\gamma}\left(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \ldots + \hat{\beta}_k(1) + \ldots + \hat{\beta}_p X_p\right)\right]\right]}{\exp\left[-y_i^{\hat{\gamma}} \exp\left[-\hat{\gamma}\left(\hat{\beta}_0 + \hat{\beta}_1 X_1 + \ldots + \hat{\beta}_k(0) + \ldots + \hat{\beta}_p X_p\right)\right]\right]} \tag{42}$$

Weibull cumulative distribution regression ratio values based on nominal covariates $X_k$ is

$$RF_{X_k} = \frac{F\left(y|X_k = 1\right)}{F(y|X_k = 0)} = \frac{1 - S\left(y|X_k = 1\right)}{1 - S(y|X_k = 0)} \tag{43}$$

Weibull hazard regression ratio values based on nominal covariates $X_k$ is

$$Rh_{X_k} = \frac{h\left(y|X_k = 1\right)}{h\left(y|X_k = 0\right)} = \exp\left[-\hat{\gamma}\hat{\beta}_k\right] \tag{44}$$

and the ratio values of mean Weibull regression based on nominal covariates $X_k$ is

$$R\mu_{X_k} = \frac{\exp\left[\hat{\beta}_k(1)\right]}{\exp\left[\hat{\beta}_k(0)\right]} = \exp\left[\hat{\beta}_k\right] \tag{45}$$

**Table 1**
Descriptive statistics.

| Covariates | Minimum | Maximum | Average |
|---|---|---|---|
| Hospitalization Time ($\boldsymbol{Y}$) | 2 | 19 | 9 |
| Age ($\boldsymbol{X}_1$) | 25 | 81 | 56 |
| HDL levels ($\boldsymbol{X}_4$) | 21 | 94 | 43.06 |
| LDL levels ($\boldsymbol{X}_5$) | 50 | 291 | 144.34 |
| Body Mass Index ($\boldsymbol{X}_6$) | 17.58 | 37.78 | 25.99 |
| Systolic Pressure ($\boldsymbol{X}_7$) | 80 | 221 | 161.23 |
| Diastolic Pressure ($\boldsymbol{X}_8$) | 61 | 144 | 91.85 |

**Table 2**
Percentage of patients for gender covariates.

| Gender | Status | | | | Total | |
|---|---|---|---|---|---|---|
| | Event (Recovery) | | Censored | | | |
| | Frequency | % | Frequency | % | Frequency | % |
| Male | 34 | 83 | 7 | 17 | 41 | 66 |
| Female | 16 | 76 | 5 | 24 | 21 | 34 |

**Table 3**
Percentage of patients for stroke type covariates.

| Covariates | Ischemic Stroke | | Hemorrhagic Stroke | | Total | |
|---|---|---|---|---|---|---|
| | Frequency | % | Frequency | % | Frequency | % |
| Type of Stroke | 42 | 68 | 20 | 32 | 62 | 100 |

**Table 4**
Percentage of patients for history of diabetes mellitus covariates.

| Covariates | Has a History of Diabetes Mellitus | | No History of Diabetes Mellitus | | Total | |
|---|---|---|---|---|---|---|
| | Frequency | % | Frequency | % | Frequency | % |
| History of Diabetes Melitus | 25 | 40 | 37 | 60 | 62 | 100 |

**Table 5**
Percentage of patients for history of heart disease covariates.

| Covariates | Has a History of Heart Disease | | No History of Heart Disease | | Total | |
|---|---|---|---|---|---|---|
| | Frequency | % | Frequency | % | Frequency | % |
| History of Heart Disease | 10 | 16 | 52 | 84 | 62 | 100 |

## Method validation

*Data Description*

The research data is secondary data, namely medical record data for inpatients with stroke patients obtained from Abdul Wahab Sjahranie Samarinda Hospital in 2021–2022. The study sample size was 62 samples. The research data consisted of time data, namely data on the length of time patients with stroke were hospitalized ($\boldsymbol{Y}$) and covariate data, namely age ($\boldsymbol{X}_1$), gender ($\boldsymbol{X}_2$), type of stroke ($\boldsymbol{X}_3$), HDL levels ($\boldsymbol{X}_4$), LDL levels ($\boldsymbol{X}_5$), body mass index ($\boldsymbol{X}_6$), systolic pressure ($\boldsymbol{X}_7$), diastolic pressure ($\boldsymbol{X}_8$), history of diabetes mellitus ($\boldsymbol{X}_9$), and history of heart disease ($\boldsymbol{X}_{10}$). The data analysis technique used was Weibull regression modeling, and computation using Octave software.

Descriptive statistics of ratio research variables can be seen in Table 1, descriptive statistics of nominal research variables based on gender covariates can be seen in Table 2, descriptive statistics of nominal research variables based on stroke type covariates can be seen in Table 3, descriptive statistics of nominal research variables based on history of diabetes mellitus covariates can be seen in Table 4, and descriptive statistics of nominal research variables based on history of heart disease covariates can be seen in Table 5.

Based on Table 1, it can be seen that the average hospitalization time of stroke patients is 9 days with the fastest hospitalization time is 2 days and the longest is 19 days. The average age of stroke patients is 56 years old with the youngest patient is 25 years old and the oldest patient is 81 years old. The average HDL levels of stroke patients was 43,06 mg/dL with the lowest HDL levels was 21 mg/dL and the highest was 94 mg/dL. The average LDL levels of stroke patients was 144,34 mg/dL with the lowest LDL levels was 50

**Table 6**
ML estimators of Weibull distribution parameters.

| Parameters | Estimated |
|---|---|
| Scale ($\lambda$) | 10.2037 |
| Shape ($\gamma$) | 2.8383 |

**Table 7**
Weibull Distribution Test Results.

| $D_{\text{count}}$ | $D_{(50;0,10)}$ | Decision |
|---|---|---|
| 0.1258 | 0.170 | Failed to reject $H_0$ |

mg/dL and the highest was 291 mg/dL. The average body mass index of stroke patients was 25,99 with the lowest body mass index was 17,58 and the highest was 37,78. The average systolic pressure of stroke patients was 161,23 mmHg with the lowest systolic pressure was 80 mmHg and the highest was 221 mmHg The average diastolic pressure of stroke patients was 91,85 mmHg with the lowest diastolic pressure was 61 mmHg and the highest was 144 mmHg

Based on Table 2, that there are 66 % or 41 patients out of 62 patients are male patients, while 34 % or 21 patients out of 62 patients are female patients. Male patients who recovered (experienced events) were 34 people or 83 % and those who died (censored) were 7 people or 17 %, while female patients who recovered were 16 people or 76 % and those who died were 5 people or 24 %.

Based on Table 3, that there are 42 patients or 68 % of 62 patients are patients who have ischemic stroke, while 20 patients or 32 % of 62 patients are patients who have hemorrhagic stroke.

Based on Table 4, that there are 25 patients or 40 % of 62 patients are patients who have a history of diabetes mellitus, while 37 patients or 60 % of 62 patients are patients who do not have a history of diabetes mellitus.

Based on Table 5, that there are 10 patients or 16 % of 62 patients are patients who have a history of heart disease, while 52 patients or 84 % of 62 patients are patients who do not have a history of heart disease.

*Weibull distribution parameter estimation*

The assumption of the Weibull regression model is that the time data must be Weibull distributed, therefore the time data in this study is assumed to be Weibull distributed. ML estimator of Weibull distribution parameters, based on the results of calculations using Octave software can be seen in Table 6.

Based on the results of the parameter estimation in Table 6, the estimated survival function is

$$\hat{S}(y) = \exp\left[-\left(\frac{y}{10.2037}\right)^{2.8383}\right], \tag{46}$$

and the estimated cumulative distribution function is

$$\hat{F}(y) = 1 - \exp\left[-\left(\frac{y}{10.2037}\right)^{2.8383}\right]. \tag{47}$$

*Testing the distribution of hospitalization time data*

The assumption that must be met in Weibull regression modeling is that time data must be Weibull distributed. To test that the time data meets the Weibull distribution assumption, distribution testing is carried out on the stroke patient hospitalization time data. The data distribution testing method in this study uses the Kolmogorov-Smirnov method. The distribution testing hypothesis is

$H_0 : F(y) = \hat{F}(y)$

(Sample data is drawn from a Weibull-distributed population with a distribution function given by Eq. 47)

$H_1 : F(y)\hat{F}(y)$

(Sample data is drawn from a population not Weibull distributed)

The calculation results using Octave software can be seen in Table 7.

Based on Table 7, obtained $D_{\text{count}} = 0.1258 < D_{(50;0,10)} = 0.170$, then it is decided to accept $H_0$ at the 10 % significance level. The conclusion of the test is that the stroke patient hospitalization time data is drawn from a Weibull-distributed population.

*Multicollinearity detection*

Multicollinearity detection in this study uses VIF criteria. The results of the calculation of the VIF value of each covariate using Octave software can be seen in Table 8.

**Table 8**
VIF Values of Covariates.

| Covariates | VIF value |
|---|---|
| Age ($X_1$) | 1.2862 |
| Gender ($X_2$) | 1.2370 |
| Type of Stroke ($X_3$) | 1.2460 |
| HDL Levels ($X_4$) | 1.3202 |
| LDL Levels ($X_5$) | 1.1828 |
| Body Mass Index ($X_6$) | 1.0837 |
| Systolic Pressure ($X_7$) | 4.4327 |
| Diastolic Pressure ($X_8$) | 4.1027 |
| History of Diabetes Mellitus ($X_9$) | 1.1877 |
| History of Heart Disease ($X_{10}$) | 1.1788 |

**Table 9**
Parameter Estimator of Weibull Regression Model.

| Covariates | Parameter | Estimated |
|---|---|---|
| – | $\gamma$ | 3.0633 |
| – | $\beta_0$ | 2.5201 |
| Age ($X_1$) | $\beta_1$ | 0.0065 |
| Body Mass Index ($X_6$) | $\beta_6$ | -0.0216 |
| History of Diabetes Mellitus ($X_9$) | $\beta_9$ | 0.1947 |

Based on Table 8, the VIF value for each covariate is less than 10 so it can be concluded that there is no multicollinearity between covariates. Based on the multicollinearity detection results, Weibull regression modeling can involve 10 covariates, namely age ($X_1$), gender ($X_2$), type of stroke ($X_3$), HDL levels ($X_4$), LDL levels ($X_5$), body mass index ($X_6$), systolic pressure ($X_7$), diastolic pressure ($X_8$), history of diabetes mellitus ($X_9$), and history of heart disease ($X_{10}$).

*Parameter estimation of the best Weibull regression model*

The best model is obtained through parameter estimation of all models from all combinations of 10 covariates in Table 8 which gives the minimum AIC model and all covariates in the model have the most effect. Based on the 10 covariates in Table 8, it can produce as many as 1023 Weibull regression models. Based on the selection results, the best model is the Weibull regression model with 3 covariates, namely age ($X_1$), body mass index ($X_6$), and history of diabetes mellitus ($X_9$). This model produces a minimum AIC value of -442.712 and the three covariates are all influential.

Parameter estimation of the Weibull regression model using the MLE method. The ML estimator of the Weibull regression model based on the calculation results using Octave software can be seen in Table 9.

Based on Table 9 and Eq. (17), the estimated Weibull survival regression model which states the probability model of stroke patients not surviving after being treated for $y$ days is

$$S(y, \boldsymbol{x}) = \exp\left(-y^{3.0633} \exp\left(-7.7198 - 0.0199X_1 + 0.0662X_6 - 0.5964X_9\right)\right) \tag{48}$$

Based on Table 9 and Eq. (18), the Weibull cumulative distribution regression model which states the probability of stroke patients recovering after being treated for $y$ days is

$$F(y.\boldsymbol{x}) = 1 - \exp\left(-y^{3.0633} \exp\left(-7.7198 - 0.0199X_1 + 0.0662X_6 - 0.5964X_9\right)\right) \tag{49}$$

Based on Table 9 and Eq. (19), the Weibull hazard regression model which states the recovery rate model for stroke patients when treated for $y$ days is

$$h(y.\boldsymbol{x}) = 3.0633y^{2.0633} \exp\left(-7.7198 - 0.0199X_1 + 0.0662X_6 - 0.5964X_9\right) \tag{50}$$

and based on Table 9 and Eq. (20), the mean Weibull regression model which states the model of the average length of hospitalization time for stroke patients is

$$\mu_y(y.\boldsymbol{x}) = 0.8938 \exp\left(2.5201 + 0.0065X_1 - 0.0216X_6 + 0.1947X_9\right) \tag{51}$$

*Significance testing of the best Weibull Regression Parameters*

The hypothesis for testing the significance of parameters simultaneously is

$H_0 : \beta_1 = \beta_6 = \beta_9 = 0$

(Weibull regression model is not feasible (not fit))

$H_1$ : at least one $\beta_k 0; k = 1, 6, 9$

**Table 10**
Simultaneous Significance Test Results.

| Test Statistic $G$ | $\chi^2_{(0,10;3)}$ | $p_{value}$ | Keputusan |
|---|---|---|---|
| 9.5673 | 6.2514 | 0.022627 | Menolak $H_0$ |

**Table 11**
Partial Significance Test Results.

| Covariates | W count | $p_{value}$ | Decision |
|---|---|---|---|
| Constant ($\beta_0$) | 7.2169 | $5.3180 \times 10^{-13}$ | Reject $H_0$ |
| Age ($\beta_1$) | 1.8154 | 0.0695 | Reject $H_0$ |
| Body Mass Index ($\beta_6$) | 1.8656 | 0.0621 | Reject $H_0$ |
| History of Diabetes Mellitus ($\beta_9$) | 1.9994 | 0.0456 | Reject $H_0$ |

**Table 12**
Ratio Values of the Weibull Regression Model Based on Influential Covariates.

| Covariates | $R_s$ | $R_F$ | $R_h$ | $R_\mu$ |
|---|---|---|---|---|
| Age ($X_1$) | 1.0236 | 0.9896 | 0.9803 | 1.0065 |
| Body Mass Index ($X_6$) | 0.9222 | 1.0343 | 1.0684 | 0.9786 |
| History of Diabetes Mellitus ($X_9$) | 2.6256 | 0.7854 | 0.5508 | 1.2149 |

(Weibull regression model is feasible (fit))

The results of testing the significance of Weibull regression parameters simultaneously can be seen in Table 10.

Based on the calculation of simultaneous testing, it was decided to reject $H_0$ at the significance level $\alpha = 0.10$ because $G = 9.5637 > \chi^2_{(0,10;3)} = 6.2514$ and $p_{value} = 0.022627 < \alpha = 0.10$. The conclusion of the test is that the Weibull regression model with covariates of age, body mass index, and history of diabetes mellitus provides a feasible model (fit).

The partial test hypothesis for a given $\beta_k$, $k = 0,1,6,9$ is

$H_0 : \beta_k = 0; k = 0,1,6,9$

(Covariates $X_k$ has no effect on the Weibull regression model)

$H_1 : \beta_k 0; k = 0,1,6,9$

(Covariates $X_k$ affects the Weibull regression model)

The results of testing the significance of Weibull regression parameters partially can be seen in Table 11.

Based on the results of testing the Weibull regression parameters partially, the constant has a calculated W value of more than the critical value of $Z_{(0,95)} = 1.64$ or a $p_{value}$ of less than 0.10 so that the decision for the constant is to reject $H_0$ at the significance level $\alpha = 0.10$ and the conclusion is that the constant is significantly different from zero. Covariates of age ($X_1$), body mass index ($X_6$), and history of diabetes mellitus ($X_9$) have a calculated W value greater than the critical value $Z_{(0,95)} = 1.64$ or a $p_{value}$ of less than 0,10 so that the decision for the covariates age ($X_1$), body mass index ($X_6$), and history of diabates mellitus ($X_9$) s to reject $H_0$ at the significance level $\alpha = 0, 10$ and it is concluded that the covariates age ($X_1$), body mass index ($X_6$), and history of diabetes mellitus ($X_9$) affect the chances of patients not recovering and recovering, the rate of recovery of patients, and the average hospitalization time of stroke patients.

*Weibull Regression Model Interpretation*

The interpretation of the Weibull regression model uses the ratio value of Weibull regression based on influential covariates and is calculated based on one of the Eq. (38) to Eq. (45). The ratio values of survival regression, cumulative distribution regression, hazard regression and mean Weibull regression based on influential covariates can be seen in Table 12.

Based on the calculation results in Table 12, the Weibull survival regression ratio of patients based on age covariates is 1.0236. The interpretation is that an increase in patient age of 1 year will increase the chance of the patient not recovering to 1.0236 times or an increase of 2.36 %. Based on the calculation results, the Weibull survival regression ratio of patients based on body mass index covariates is 0.9222. The interpretation is that an increase in the patient body mass index by 1 % will reduce the patient chance of not recovering to 0.9222 times or decrease by 7.78 %. Based on the calculation results, the Weibull survival regression ratio of patients based on the covariate history of diabetes mellitus is 2.6256. The interpretation is that the chance of not recovering for patients who have a history of diabetes mellitus (assuming other covariate values are fixed) will increase to 2.6256 times the chance of not recovering for patients who do not have a history of diabetes mellitus. The survival regression graph of patients without a history of diabetes mellitus and with a history of diabetes mellitus can be seen in Fig. 1. Based on Fig. 1, the green graph is the Weibull survival regression graph of patients without a history of diabetes mellitus and the red graph is the Weibull survival regression graph of patients with a history of diabetes mellitus. The red graph is above the green graph, indicating that the chance of not recovering

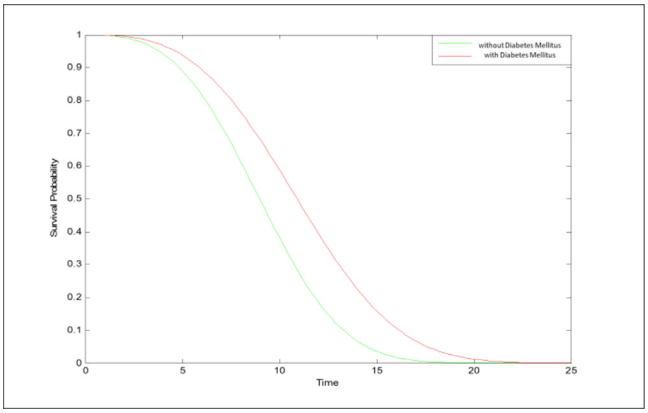


**Fig. 1.** Weibull survival regression graph without and with a history of diabetes mellitus ($\boldsymbol{X}_9$).

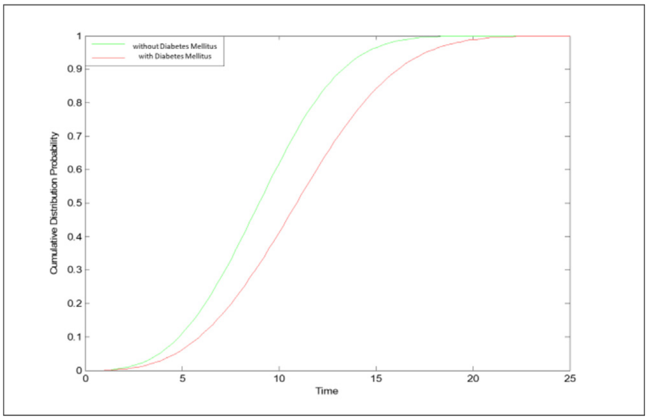


**Fig. 2.** Weibull cumulative distribution regression graph for patients without and with a history of diabetes mellitus ($X_9$).

for patients with a history of diabetes mellitus is greater than that of patients without a history of diabetes mellitus or patients with a history of diabetes mellitus recover more slowly.

Based on the calculation results in Table 12, the regression ratio of the Weibull cumulative distribution of patients based on age covariates is 0.9896. The interpretation is that an increase in patient age of 1 year will reduce the patient chance of recovery to 0.9896 times or decrease by 1.04 %. Based on Table 12, the regression ratio of the cumulative Weibull distribution of patients based on the covariate of body mass index is 1.0343. The interpretation is that a 1 % increase in the patient body mass index will increase the patient chance of recovery to 1.0343 times or an increase of 3.43 %. Based on Table 12, the regression ratio of the cumulative Weibull distribution of patients based on the covariate history of diabetes mellitus is 0.7854. The interpretation is that the chance of recovery for patients who have a history of diabetes mellitus (assuming other covariate values are fixed) will decrease to 0.7854 times the chance of recovery for patients who do not have a history of diabetes mellitus. The regression graph of the cumulative distribution of patients with a history of diabetes mellitus and without a history of diabetes mellitus can be seen in Fig. 2.

Based on Fig. 2, the green graph is the Weibull cumulative distribution regression graph of patients without a history of diabetes mellitus and the red graph is the Weibull cumulative distribution regression graph of patients with a history of diabetes mellitus. The red graph is below the green graph, indicating that the chance of recovering patients who have a history of diabetes mellitus is smaller than patients who do not have a history of diabetes mellitus or patients who have a history of diabetes mellitus recover more slowly.

Based on the calculation results in Table 12, the Weibull hazard regression ratio for each patient based on age covariates is 0.9803. The interpretation is that an increase in patient age of 1 year will reduce the patient recovery rate to 0.9803 times or decrease by 1.97 %. Based on Table 12, the Weibull hazard regression ratio of each patient based on the body mass index covariate is 1.0684. The interpretation is that a 1 % increase in the patient body mass index will increase the recovery rate to 1.0684 times or an increase of 6.84 %. Based on Table 12, the Weibull hazard regression ratio for each patient based on the covariate history of diabetes mellitus is 0.5508. The recovery rate of stroke patients who have a history of diabetes mellitus will decrease by 0.5508 times the recovery rate for patients who do not have a history of diabetes mellitus. This means that the recovery rate of patients with a history of diabetes mellitus is lower than that of patients without a history of diabetes mellitus. The red graph is the Weibull hazard function graph of patients without a history of diabetes mellitus and the green graph is the Weibull hazard graph of patients with a history of diabetes

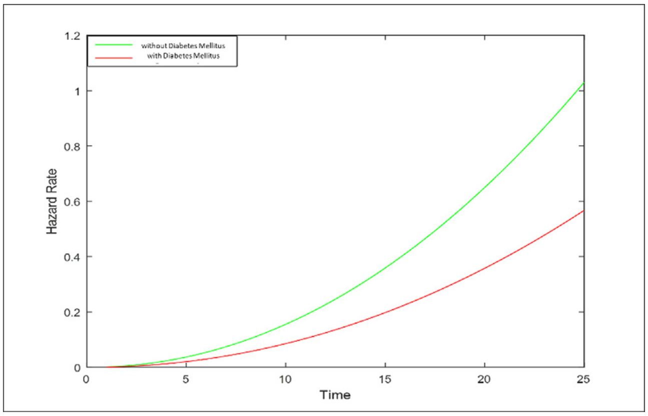


**Fig. 3.** Weibull hazard regression graph without and with history of diabetes mellitus ($\boldsymbol{X}_9$).

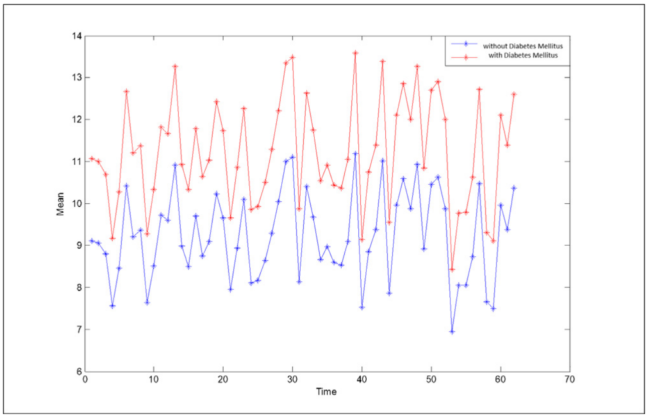


**Fig. 4.** Mean Weibull regression graph without and with history of diabetes mellitus ($X_9$).

mellitus. The red graph is above the green graph, meaning that patients without a history of diabetes mellitus recover faster than patients who have a history of diabetes mellitus (Fig. 3).

Based on the calculation results in Table 12, the mean Weibull regression ratio for each patient based on the age covariate is 1.0065. The interpretation is that an increase in patient age of 1 year will increase the average length of hospitalization of patients to 1.0065 times or an increase of 0.65 %. Based on Table 12, the mean Weibull regression ratio of each patient based on the body mass index covariate is 0.9786. The interpretation is that a 1 % increase in the patient body mass index will reduce the average length of hospitalization to 0.9786 times or decrease by 2.14 %. Based on Table 12, the mean Weibull regression ratio of each patient based on the covariate history of diabetes mellitus is 1.2149. The average length of hospitalization for stroke patients who have a history of diabetes mellitus will increase 1.2149 times the average length of hospitalization for patients who do not have a history of diabetes mellitus. This means that the average length of hospitalization of patients with a history of diabetes mellitus is greater than that of patients without a history of diabetes mellitus. The mean regression graph without and with a history of diabetes mellitus can be seen in Fig. 4. The mean Weibull regression graph represents the average length of stay of patients. Based on Fig. 4, the red graph is the average length of stay of patients with a history of diabetes mellitus and the blue graph is the average length of stay of patients without a history of diabetes mellitus. The red graph is above the blue graph, meaning that patients who have a history of diabetes mellitus the length of hospitalization of these patients will increase or get longer.

## Conclusion

The conclusions obtained from this study are as follow

1. The Weibull survival regression model that models the chance of stroke patients not recovering after being treated for $y$ days is:

$$\hat{S}(y) = \exp\left(-y^{3.0633} \exp\left(-7.7198 - 0.0199X_1 + 0.0662X_6 - 0.5964X_9\right)\right).$$

where $X_1$ is age, $X_6$ is body mass index, and $X_9$ is history of diabetes mellitus.

The Weibull cumulative distribution regression model that models the chance of a stroke patient recovering after being treated for $y$ days is

$$\hat{F}(y) = 1 - \exp\left(-y^{3.0633}\exp\left(-7.7198 - 0.0199X_1 + 0.0662X_6 - 0.5964X_9\right)\right).$$

The Weibull hazard regression model that models the recovery rate of stroke patients when treated for $y$ days is

$$\hat{h}(y) = 3.0633y^{2.0633}\exp\left(-7.7198 - 0.0199X_1 + 0.0662X_6 - 0.5964X_9\right).$$

The mean Weibull regression model which states the model of the average length of hospitalization of patients with stroke is

$$\hat{\mu}_y(y) = 0.8938\exp\left(2.5201 + 0.0065X_1 - 0.0216X_6 + 0.1947X_9\right).$$

2. Factors that affect the Weibull regression model in patients with stroke at Abdul Wahab Sjahranie Hospital are covariates of age ($X_1$), body mass index ($X_6$), and history of diabetes mellitus ($X_9$).
3. Interpretation of Weibull regression models for influential covariates, as follows:

   a. An increase in patient age of 1 year will increase the chance of the patient not recovering to 1.0236 times or an increase of 2.36 %. An increase in patient body mass index by 1 % will decrease the patient's chance of not recovering to 0.9222 times or decrease by 7.78 %. The chance of not recovering for patients who have a history of diabetes mellitus (assuming the value of other covariates is fixed) will increase to 2.6256 times the chance of not recovering for patients who do not have a history of diabetes mellitus.
   b. An increase in patient age of 1 year will decrease the patient chance of recovery to 0.9896 times or decrease by 1.04 %. An increase in patient body mass index by 1 % will increase the patient chance of recovery to 1.0343 times or an increase of 3.43 %. The chance of recovery for patients who have a history of diabetes mellitus (assuming the value of other covariates is fixed) will decrease to 0.7854 times the chance of recovery for patients who do not have a history of diabetes mellitus.
   c. An increase in patient age of 1 year will decrease the patient recovery rate to 0.9803 times or decrease by 1.97 %. An increase in patient body mass index by 1 % will increase the recovery rate to 1,0684 times or an increase of 6.84 %. The recovery rate of stroke patients who have a history of diabetes mellitus will decrease by 0.5508 times the recovery rate for patients who do not have a history of diabetes mellitus. This means that the recovery rate of patients who have a history of diabetes mellitus is lower than patients who do not have a history of diabetes mellitus.
   d. An increase in patient age of 1 year will increase the average length of hospitalization of patients to 1.0065 times or an increase of 0.65 %. An increase in patient body mass index by 1 % will reduce the average length of hospitalization of patients to 0.9786 times or decrease by 2.14 %. The average length of hospitalization for stroke patients who have a history of diabetes mellitus will increase 1.2149 times the average length of hospitalization for patients who do not have a history of diabetes mellitus. This means that the average length of hospitalization of patients who have a history of diabetes mellitus is greater than patients who do not have a history of diabetes mellitus.

## Limitations

'Not applicable'

## Ethics statements

The data used in this research are secondary data, namely medical record data for inpatients with stroke patients obtained from Abdul Wahab Sjahranie Samarinda Hospital in 2021–2022. The study sample size was 62 samples. The research data consisted of time data, namely data on the length of time patients with stroke were hospitalized ($Y$) and covariate data, namely age ($X_1$), gender ($X_2$), type of stroke ($X_3$), HDL levels ($X_4$), LDL levels ($X_5$), body mass index ($X_6$), systolic pressure ($X_7$), diastolic pressure ($X_8$), history of diabetes mellitus ($X_9$), and history of heart disease ($X_{10}$).

## CRediT author statement

**Suyitno**: Conceptualization, Methodology, Validity tests, Writing-Preparation of the first draft, and Supervision. **Darnah**: Conceptualization, Validity Test, Data curation, Analysis, Visualization. **Andrea Tri Rian Dani:** Analysis, Visualization, Editing Draft, and writing original draft. **Nurul Tri Oktavia**: Data curation, Analysis, Visualization.

## Supplementary material *and/or* additional information [OPTIONAL]

None.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## References

[1] R.L. Smith, Weibull Regression Models for Reliability Data, Reliab. Eng. Syst. Saf. 34 (1991) 55–77.
[2] T. Cavalcante, R. Ospina, V. Leiva, X. Cabezas, C. Martin-Barreiro, Weibull Regression and Machine Learning Survival Models: Methodology, Comparison, and Application to Biomedical Data Related to Cardiac Surgery, Biology. (Basel) 12 (3) (Mar. 2023), doi:10.3390/biology12030442.
[3] A.R. Baghestani, S.S. Moghaddam, H.A. Majd, M.E. Akbari, N. Nafissi, K. Gohari, Survival analysis of patients with breast cancer using weibull parametric model, Asian Pacific J. Cancer Prev. 16 (18) (2016) 8567–8571, doi:10.7314/APJCP.2015.16.18.8567.
[4] A. Kottas, Nonparametric Bayesian survival analysis using mixtures of Weibull distributions, J. Stat. Plan. Inference 136 (3) (Mar. 2006) 578–596, doi:10.1016/j.jspi.2004.08.009.
[5] S.A. Khan, Exponentiated Weibull regression for time-to-event data, Lifetime Data Anal. 24 (2) (Apr. 2018) 328–354, doi:10.1007/s10985-017-9394-3.
[6] B. George, S. Seals, I. Aban, Survival Analysis and Regression Models, Springer, New York LLC, 2014, doi:10.1007/s12350-014-9908-2.
[7] A.J. Hallinan, A Review of the Weibull Distribution, J. Qual. Technol. 25 (2) (Apr. 1993) 85–93, doi:10.1080/00224065.1993.11979431.
[8] M.S. Shama, et al., Modified generalized Weibull distribution: theory and applications, Sci. Rep. 13 (1) (Dec. 2023), doi:10.1038/s41598-023-38942-9.
[9] D.L. Wilson, The analysis of survival (mortality) data: Fitting Gompertz, Weibull, and logistic functions, Mech. Ageing Dev. 74 (1994) 15–33.
[10] J.P. Klein, H.C. van Houwelingen, J.G. Ibrahim, Handbook of Survival Analysis, T.H. Scheike (Ed.), Chapman and Hall/CRC, 2013, doi:10.1201/b16248.
[11] S.Y. Sohn, I.S. Chang, T.H. Moon, Random effects Weibull regression model for occupational lifetime, Eur. J. Oper. Res. 179 (1) (May 2007) 124–131, doi:10.1016/j.ejor.2006.03.008.
[12] Suyitno, N.W.W. Sari, Parameter estimation of mixed geographically weighted weibull regression model, Journal of Physics: Conference Series, Institute of Physics Publishing, Aug. 2019, doi:10.1088/1742-6596/1277/1/012046.
[13] B.W. Negasa, T.W. Wotale, M.E. Lelisho, L.K. Debusho, K. Sisay, W. Gezimu, Modeling Survival Time to Death among Stroke Patients at Jimma University Medical Center, Southwest Ethiopia: A Retrospective Cohort Study, Stroke Res. Treat. 2023 (2023), doi:10.1155/2023/1557133.
[14] D. Mukherjee, C.G. Patil, Epidemiology and the Global Burden of Stroke, Elsevier Inc, 2011, doi:10.1016/j.wneu.2011.07.023.
[15] A.S. Kim, S.C. Johnston, Global variation in the relative burden of stroke and ischemic heart disease, Circulation 124 (3) (Jul. 2011) 314–323, doi:10.1161/CIRCULATIONAHA.111.018820.
[16] J. xiao Li, Q. qiong Zhong, S. xiang Yuan, F. Zhu, Trends in deaths and disability-adjusted life-years of stroke attributable to low physical activity worldwide, 1990–2019, BMC. Public Health 23 (1) (Dec. 2023), doi:10.1186/s12889-023-17162-w.
[17] F.Sughra Zaidi, U. Fatima, A. Usmani, A.Raza Jafri, Comprehending Nodes Essentiality through Centrality Measures in Biological Networks, IJCSNS Int. J. Comput. Sci. Netw Secur. 19 (9) (2019).
[18] U. Fatima, S. Hina, M. Wasif, A novel global clustering coefficient-dependent degree centrality (GCCDC) metric for large network analysis using real-world datasets, J. Comput. Sci. 70 (2023) 102008, doi:10.1016/j.jocs.2023.102008.
[19] L. Attardi, M. Guida, G. Pulcini, A mixed-Weibull regression model for the analysis of automotive warranty data, Reliab. Eng. Syst. Saf. 87 (2) (Feb. 2005) 265–273, doi:10.1016/j.ress.2004.05.003.
[20] M. Heo, M.S. Faith, and D.B. Allison, “Power and sample size for survival analysis under the Weibull distribution when the whole lifespan is of interest,” 1998.
[21] M. Arashi, M. Norouzirad, M. Roozbeh, N.Mamode Khan, A high-dimensional counterpart for the ridge estimator in multicollinear situations, Mathematics 9 (23) (Dec. 2021), doi:10.3390/math9233057.
[22] Z. Zhang, Parametric regression model for survival data: Weibull regression model as an example, Ann. Transl. Med. 4 (24) (Dec. 2016), doi:10.21037/atm.2016.08.45.
[23] Purhadi Suyitno, Sutikno, Irhamah, Multivariate Weibull regression model, Far East J. Math. Sci. 101 (9) (May 2017) 1977–1992, doi:10.17654/MS101091977.
[24] Purhadi Suyitno, Sutikno, Irhamah, Parameter estimation of geographically weighted trivariate Weibull regression model, Appl. Math. Sci. 10 (17–20) (2016) 861–878, doi:10.12988/ams.2016.6129.
[25] D.D. Hanagal, A Bivariate Weibull Regression Model, Econ. Q. Control 20 (1) (2005) 143–150.