Journal of Education Technology and Information System Vol (01) Ed (02) 53-62 (2025)
Received October 2025 / Revised February 2025 / Accepted June 2025

Journal of Education Technology and Information System

https://journal.unesa.ac.id/index.php/jetis/

I-Regs (Internet-Regression Analysis) as a Statistical Innovation in Nonparametric Regression Modeling

Andrea Tri Rian Dani[1*], I Nyoman Budiantara[2], Ulfa Siti Nuraini[3], Wiyli Yustanti[4], Sifriyani[5], Fachrian Bimantoro Putra[6]

[1,5,6] *Mulawarman University, Samarinda, 75119, Indonesia*
andreatririandani@fmipa.unmul.ac.id , sifriyani@fmipa.unmul.ac.id , putrafachrian9@gmail.com
[2] *Sepuluh Nopember Institute of Technology, Surabaya, 60111, Indonesia*
i_nyoman_b@statistika.its.ac.id
[3,4] *The State University of Surabaya, Surabaya, 60231, Indonesia*
ulfanuraini@unesa.ac.id , wiyliyustanti@unesa.ac.id

Abstract.

This research develops an information system based on the R-Shiny Dashboard, allowing users to perform nonparametric regression modeling. Internet-Regression Analysis (I-Regs) is the name of a dashboard that has been successfully developed. I-Regs provides a complete model library in regression analysis modeling, including parametric, nonparametric, and semiparametric regression. Based on the analysis, the I-Regs information system has been successfully developed with the main features: an intuitive user interface, a complete model library, and interactive data visualization. The development of the I-Regs information system has succeeded in carrying out regression analysis. I-Regs can be a comprehensive solution for researchers, practitioners and students in solving various data analysis problems.

Keywords: Dashboard; I-Regs; Nonparametric Regression; Regression Analysis; Statistics

1. Introduction

Some real-world problems, such as two or more inseparable variables, will be interesting if the nature of the relationship is investigated. The method that is often used is regression analysis. Regression analysis is one of the statistical methods used to determine the pattern of relationships between one or more variables (Suparti *et al.*, 2021). The variables used in regression analysis consist of response variables and predictor variables (Ratnasari, Budiantara and Dani, 2021). Initial identification is done to check for a pattern of relationships by using a scatter plot. (Budiantara *et al.*, 2019) in his research stated that other things need to be checked before conducting statistical modeling with regression analysis, namely investigating whether the variables are rationally correlated. If they are rationally correlated, then statistical modeling can be done using regression analysis. The primary purpose of regression analysis is to find the estimated form of the regression curve.

Using regression analysis, researchers created several statistical modeling techniques, including semiparametric, nonparametric, and parametric regression. (Ratnasari, Utama and Dani, 2024). Parametric regression is a regression where the form of the regression curve pattern is known. In contrast, nonparametric regression is a regression where the form of the curve pattern is unknown. Meanwhile, if some forms of the regression curve pattern are known and some are unknown, then the regression analysis used is semiparametric regression (Hardle, 1994; Sifriyani *et al.*, 2023). In practice, it is frequently uncertain how the response variable and the predictor variable relate to one another. The parametric regression model is less appropriate in certain situations, hence nonparametric regression is the suggested method. (Eubank, 1999; Lin and Ying, 2001). The nonparametric regression approach is highly flexible since it does not rely on the assumption of a specific regression curve shape. The data is anticipated to modify the predicted form of the regression curve itself without being impacted by the subjectivity of the researcher. (Biao and Pourahmadi, 2003; Du, Parmeter and Racine, 2012).In nonparametric regression modeling, researchers developed several estimators, including Kernel (Wong, 1983; Du, Parmeter and Racine, 2012; Hidayat *et al.*, 2019; Guidoum, 2020), Spline (Regier and Parker, 2015; Wayan Sudiarsa *et al.*, 2015; Sifriyani *et al.*, 2017; Widyastuti, Fernandes and Pramoedyo, 2021), and Fourier series (Bilodeau, 1992; Wayan Sudiarsa *et al.*, 2015; Suparti *et al.*, 2019; Mariati, Budiantara and Ratnasari, 2020).

Spline estimators, with their remarkable adaptability, can be extended to complex and intricate statistical modeling. Their unique advantage lies in their ability to adapt to the data pattern, regardless of its movement (Mariati, Budiantara and Ratnasari, 2021). In addition to Spline, the Kernel estimator has also garnered significant attention from researchers. This estimator, with its flexible form, excels in modeling data that lacks a specific pattern, such as data with irregular peaks and valleys(Afifah, Budiantara and Latra, 2017). Notably, the mathematical calculations involved are straightforward, adding to the appeal of Kernel estimators. Another nonparametric regression estimator that often takes the spotlight is the Fourier series. The Fourier series estimator is commonly employed when the data pattern forms a sine-cosine wave pattern (Budiantara *et al.*, 2019).

Estimators developed in nonparametric regression by researchers are currently limited by the availability of open-source and easy-to-use applications. However, there is a promising future in the potential development of an integrated, flexible, and easy-to-access system for nonparametric regression modeling. Until now, some researchers have constructed programs manually by translating manual processing algorithms into a programming language using open-source software to get results. This poses a challenge for laypeople who wish to study nonparametric and semiparametric regression, as they must learn theoretical concepts and program coding in software to get results. The data analysis process, especially nonparametric regression modeling, can be easily understood when applied in a system. An integrated system with the same goal of obtaining nonparametric regression modeling results can undoubtedly be developed based on the existing framework. The advantage of this system-based application is that it can provide information that can be used anywhere and anytime.

In this research, we developed a web-based R software with R-Shiny to make it easier to analyze data with a more interactive, straightforward, nonparametric regression approach that does not require users to master the R program. We designed this web-based application with R-shiny to solve regression analysis modeling problems with parametric, nonparametric, and semiparametric regression approaches. It is hoped that the innovations made can be helpful in the development of statistical science, especially in the fields of nonparametric and semiparametric regression and its applications.

2. Methods

In this section, we will discuss some theoretical foundations related to nonparametric regression.

2.1. Truncated Spline Nonparametric Regression

Suppose given paired data $(x_{1i}, x_{2i}, ..., x_{pi}, y_i)$ which can then be assumed if the relationship between the predictor with response variables follows a nonparametric regression model.

$$y_i = f(x_{1i}, x_{2i}, ..., x_{pi}) + \varepsilon_i \tag{1}$$

Regression curve $f(x_{1i}, x_{2i}, ..., x_{pi})$ assumed to be additive, so that we can write it in the form:

$$f(x_{1i}, x_{2i}, ..., x_{pi}) = f_1(x_{1i}) + f_2(x_{2i}) + ... + f_p(x_{pi}) \rightarrow \sum_{j=1}^{p} f_j(x_{ji}) \tag{2}$$

Regression curve $f_j(x_{ji})$ is approximated by a multivariable truncated spline function, so that:

$$f_j(x_{ji}) = \delta_{j0} + \sum_{l=1}^{m} \delta_{jl} x_{ji}^{l} + \sum_{k=1}^{r} \delta_{j(m+k)} (x_{ji} - K_{jk})_+^m \tag{3}$$

For $i = 1,2,...,n$; $j = 1,2,..., p$. The truncated function can be written:

$$(x_{ji} - K_{jk})_+^m = \begin{cases} (x_{ji} - K_{jk})^m & , x \geq K_{jk} \\ 0 & , x < K_{jk} \end{cases}$$

By substituting Equation (3) into Equation (1), we obtain a multivariable truncated spline nonparametric regression model in Equation (4).

$$y_i = \delta_0 + \sum_{j=1}^{p} \left(\sum_{l=1}^{m} \delta_{jl} x_{ji}^{l} + \sum_{k=1}^{r} \delta_{j(m+k)} (x_{ji} - K_{jk})_+^m \right) + \varepsilon_i \tag{4}$$

Equation (4) can be written in matrix form as:

$$\mathbf{y} = \mathbf{X}(K)\boldsymbol{\delta} + \boldsymbol{\varepsilon} \tag{5}$$

The parameter vector of $\boldsymbol{\delta}$ in Equation (5) can be estimated using the Maximum Likelihood Estimation (MLE) method. It is assumed that if *error* random ε_i are independent, identical, and normally distributed with mean 0 and variance σ^2 or can be written $\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$. Estimator for $\hat{\boldsymbol{\delta}}$ written in Equation (6).

$$\hat{\boldsymbol{\delta}} = \left(\mathbf{X}(K)^T \mathbf{X}(K)\right)^{-1} \mathbf{X}(K)^T \mathbf{y} \tag{6}$$

2.2. Gaussian Kernel Nonparametric Regression

Suppose given paired data, namely x_i And y_i with observation observation $i = 1,2,...,n$ and follows a nonparametric regression model. The regression curve in Equation (7) is a curve whose shape is unknown and there is no past information regarding the relationship pattern. The regression curve $m(x_i)$ in nonparametric regression models can be approached by using the Kernel estimator.

$$\hat{m}_\lambda(x) = n^{-1} \sum_{i=1}^{n} \frac{K_\lambda(x - x_i)}{n^{-1} \sum_{i=1}^{n} K_\lambda(x - x_i)} y_i \rightarrow n^{-1} \sum_{i=1}^{n} W_{\lambda i}(x) y_i \tag{7}$$

with $K_\lambda(x - x_i)$ is a Kernel function : $K_\lambda(x - x_i) = \dfrac{1}{\lambda} K\left(\dfrac{x - x_i}{\lambda}\right)$

and $W_{\lambda i}(x)$ is a weighting function : $W_{\lambda i}(x) = \dfrac{K_\lambda(x - x_i)}{n^{-1} \sum_{i=1}^{n} K_\lambda(x - x_i)}$

Gaussian Kernel function is presented

$$K_\lambda(u)=\frac{1}{\sqrt{2\pi}}\exp\left(\frac{1}{2}\left(-u^2\right)\right)\quad ;I_{[-\infty,\infty]}(u)$$

Next, if we expand Equation (7), we can write:

$$\hat{m}_\lambda(x)=n^{-1}W_{\lambda 1}(x)y_1+n^{-1}W_{\lambda 2}(x)y_2+\ldots+n^{-1}W_{\lambda n}(x)y_n \tag{8}$$

Based on Equation (8), it applies to each $x=x_1,x=x_2,\ldots,x=x_n$, we can write it in matrix form as follows:

$$\begin{bmatrix}\hat{m}_\lambda(x_1)\\ \hat{m}_\lambda(x_2)\\ \vdots\\ \hat{m}_\lambda(x_n)\end{bmatrix}=\begin{bmatrix}n^{-1}\sum_{i=1}^{n}W_{\lambda i}(x_1)y_i\\ n^{-1}\sum_{i=1}^{n}W_{\lambda i}(x_2)y_i\\ \vdots\\ n^{-1}\sum_{i=1}^{n}W_{\alpha i}(x_n)y_i\end{bmatrix}$$

If denote into matrix form, it becomes:

$$\hat{\mathbf{m}}_\lambda(x)=\mathbf{X}(\lambda)\mathbf{y} \tag{9}$$

2.3. Fourier Series Nonparametric Regression

Suppose given a multivariable regression model as in Equation ()

This regression curve $f_j(x_{ji})$ will be approximated by the Fourier series function so that it is written in Equation (10)

$$f_j(x_{ji})=b_jx_{ji}+\frac{1}{2}\alpha_{0j}+\sum_{k=1}^{K}\alpha_{kj}\cos kx_{ji}\ ,\ j=1,2,\ldots,p \tag{10}$$

Furthermore, based on Equation (10), we can write it in matrix form as:

$$\mathbf{y}=\mathbf{W}(O)\boldsymbol{\alpha}+\boldsymbol{\varepsilon} \tag{11}$$

The parameter vector of $\boldsymbol{\alpha}$ in Equation (11) can be estimated using the Maximum Likelihood Estimation (MLE) method. The estimation of $\hat{\boldsymbol{\alpha}}$ shown in Equation (12)

$$\hat{\boldsymbol{\alpha}}=(\mathbf{W}(O)^T\mathbf{W}(O))^{-1}\mathbf{W}(O)^T\mathbf{y} \tag{12}$$

Estimator for regression curves $\hat{f}_j(x_{ji})$ can be written:

$$\hat{f}_j(x_{ji})=\hat{b}_jx_{ji}+\frac{1}{2}\hat{\alpha}_{0j}+\sum_{k=1}^{K}\hat{\alpha}_{kj}\cos kx_{ji} \tag{13}$$

So the estimate for the regression curve $f(x_{1i},x_{2i},\ldots,x_{pi})$ is given by:

$$\begin{aligned}f(x_{1i},x_{2i},\ldots,x_{pi})&=\sum_{j=1}^{p}\hat{f}_j(x_{ji})\\&=\sum_{j=1}^{p}\left(\hat{b}_jx_{ji}+\frac{1}{2}\hat{\alpha}_{0j}+\sum_{k=1}^{K}\hat{\alpha}_{kj}\cos kx_{ji}\right)\end{aligned} \tag{14}$$

2.4. R-Shiny

R Shiny is a framework built on the R programming language. R Shiny is intended to create interactive web applications quickly and easily. The RShiny program components are divided into two large groups, namely:

1. User Interface. This section is useful for:

a) Control Panel, which is a panel to control input in the form of data, variables, and models based on module complexity. The control display can be in the form of a slider, radio button, check-box, and others.
b) Input value request entry (data with various types of variables required, model selection, type, and statistical test criteria).
c) Presentation of output related to analysis results.

2. Server. This part is the brain of the program that is charge of simulating various data analysis according to user choices and then sending the results to the output section. This section is supported by various procedures and data analysis that are generally available in various R packages.

R Shiny is generally used as an interactive data visualization and data analysis tool. Allows users to explore data in real time.

2.5. I-Regs Development

I-Regs (Internet-Regression Analysis) is built based on R-Shiny which consists of a user interface system and a server. The server contains operations to process data and perform nonparametric regression modeling and its applications. Data inputted into I-Regs will be processed or processed via a web server. The basic framework that forms the architecture of I-Regs development is shown in Figure 1.

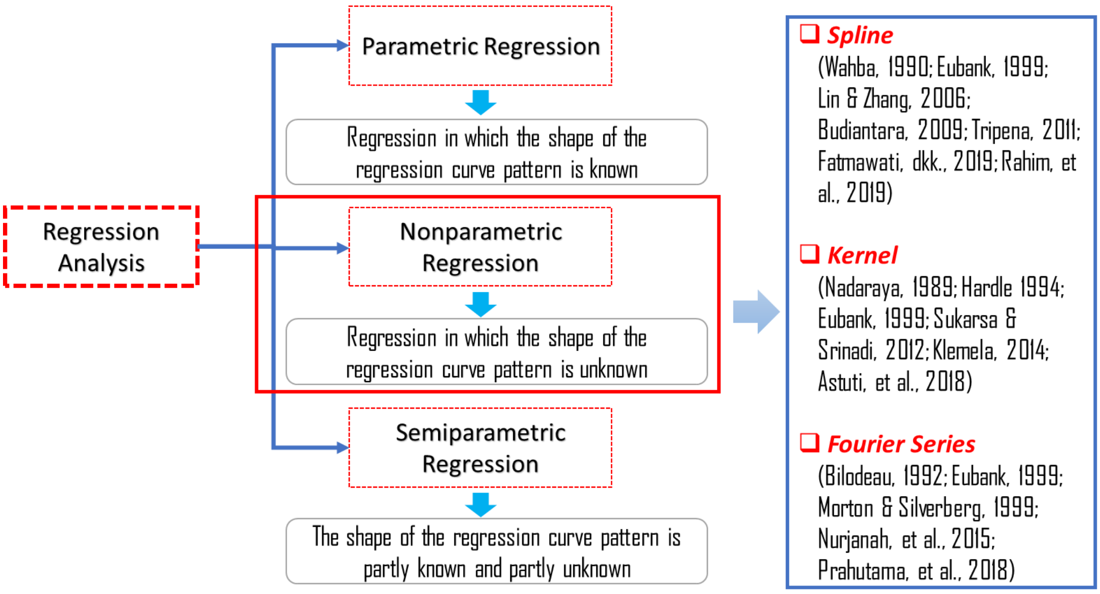


Figure 1 Basic framework that forms the architecture of I-Regs

The I-Regs interface is designed to provide input forms and display output from data processing results performed on the web server. In the data analysis process, the user interface plays a role in displaying estimation results, simultaneously and partially significance testing, and displaying comparison graphs of actual data and predictions from the best nonparametric regression model. In addition, the user interface can also display data exploration using descriptive statistics and scatter plots as an initial identification of regression analysis modeling.

3. Results and Discussion

3.1. Software Systems and Design

The I-Regs display is designed from input design, process design, and output design. In I-Regs, input and process design will be made into several parts. The input and process parts include entering data, exploring data, and determining what approach to use. The output design is designed according to the results of the analysis carried out.

3.2. Implementation and Unit Testing

In this section, the implementation of the designed software system design, I-Regs, will be carried out.

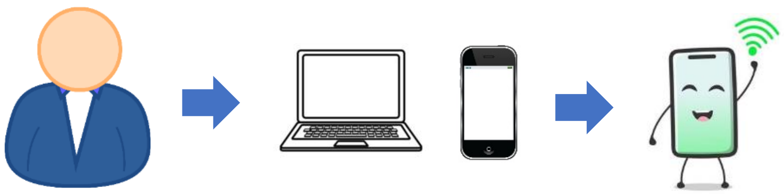


Figure 2 I-Regs software system design

To be able to access the I-Regs Program, either via PC or mobile phone can be done with the condition of being connected to the internet network. The IRegs program is published via shinnyapps.io with the following link:https://s.id/I-RegS.

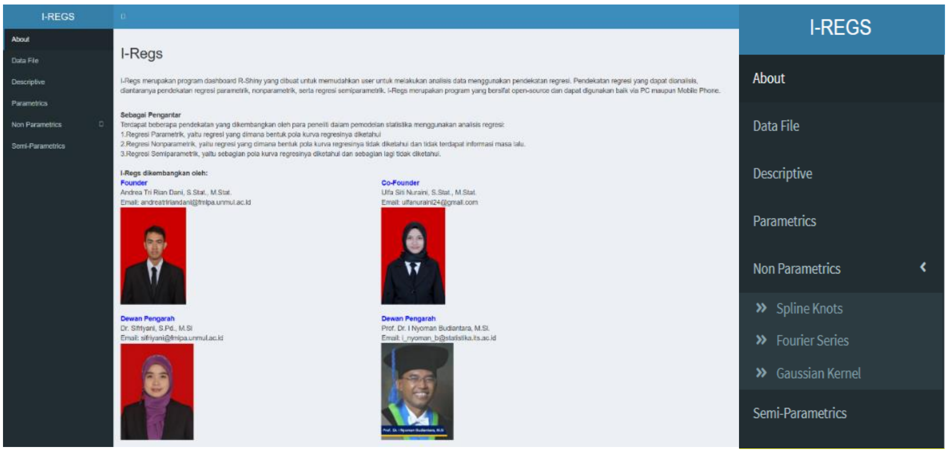


Figure 3 I-Regs on web

a) The interface section when successfully logging into the I-Regs system displays a brief description and profile of the I-Regs founder. I-Regs can be used for regression analysis with three approaches, parametric, nonparametric, and semiparametric regression.
b) *Tab Files* of the the IRegs program consists of About, Data File, Descriptive, Parametrics, Nonparametrics, and Semiparametrics.

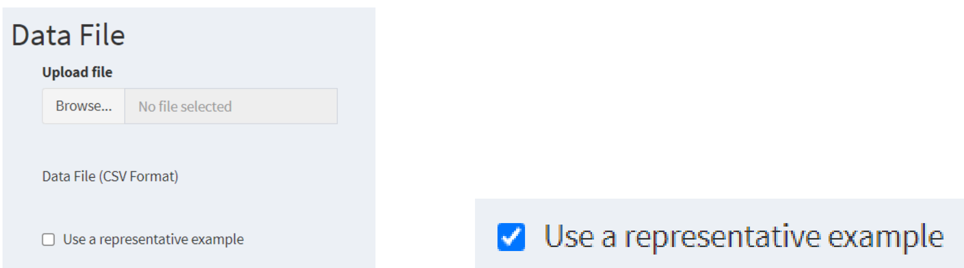
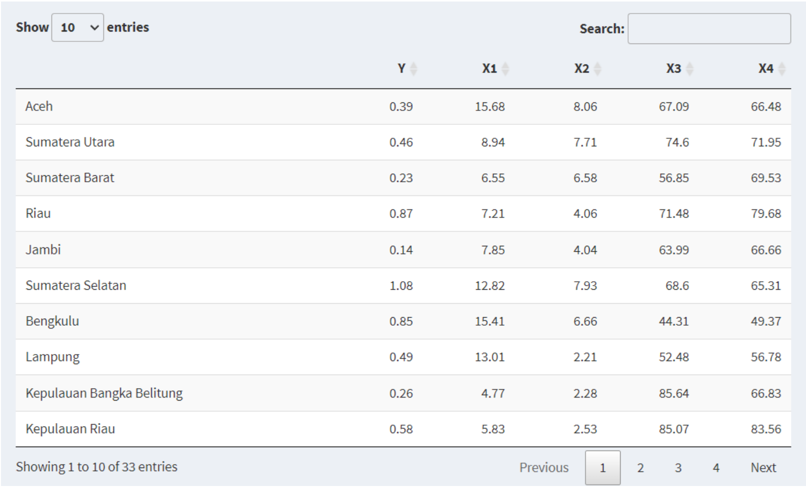


Figure 4 Tab files of the IRegs program

To import data, the dataset that needs to be provided must be in CSV format, or in this case the user can also use the sample data that has been provided by the Founder by checking Use a Representative Example.



Show 10 entries Search:

	Y	X1	X2	X3	X4
Aceh	0.39	15.68	8.06	67.09	66.48
Sumatera Utara	0.46	8.94	7.71	74.6	71.95
Sumatera Barat	0.23	6.55	6.58	56.85	69.53
Riau	0.87	7.21	4.06	71.48	79.68
Jambi	0.14	7.85	4.04	63.99	66.66
Sumatera Selatan	1.08	12.82	7.93	68.6	65.31
Bengkulu	0.85	15.41	6.66	44.31	49.37
Lampung	0.49	13.01	2.21	52.48	56.78
Kepulauan Bangka Belitung	0.26	4.77	2.28	85.64	66.83
Kepulauan Riau	0.58	5.83	2.53	85.07	83.56

Showing 1 to 10 of 33 entries Previous 1 2 3 4 Next

Figure 5 Dataset import on IRegs

The following is a data tabulation using the sample data provided. If using the data import feature, then with this data tabulation, users can see and ensure whether the dataset imported into the I-Regs program is correct or not.

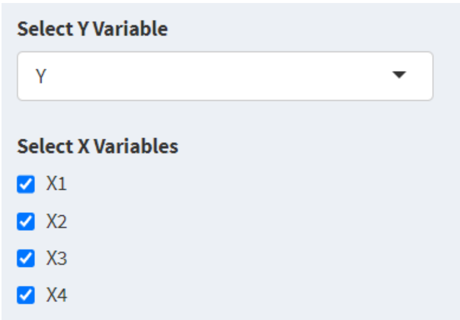
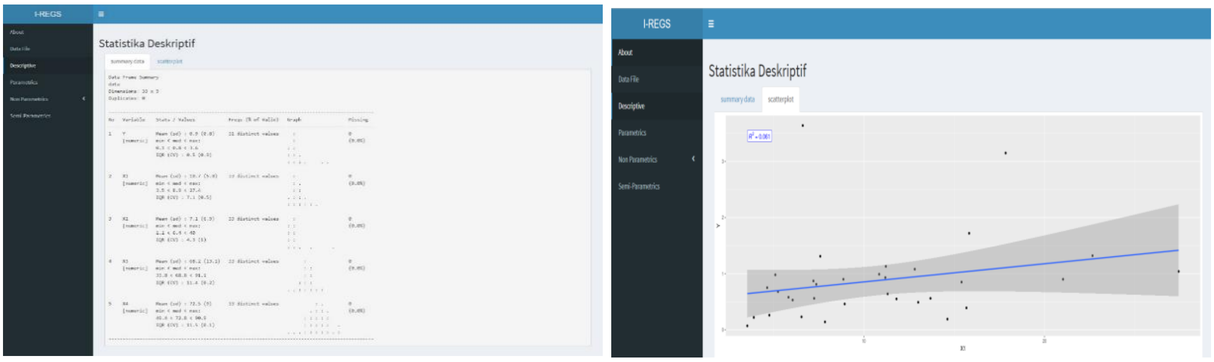


Figure 6 Variabel selection

In this section, the user must determine or select which variables are the response variables and also the predictor variables.

3.1.1. Descriptive Tab Menu Display



(a) **(b)**

Figure 7 (a) Summary data; (b) Scatter plot

In this section, Figure 7 the program will display descriptive statistics as initial information to the user about the data used. The information displayed includes data type, central tendency measure, histogram visualization and missing values for each variable.

In addition, the visualization of the scatter diagram as initial information to the user about the pattern of the relationship between the predictor variables and the response variables. The information displayed from this menu will be useful as an initial identification to determine what approach will be used, whether the parametric regression approach, nonparametric regression approach, or semiparametric regression approach.

3.1.2. Nonparametrics Tab File

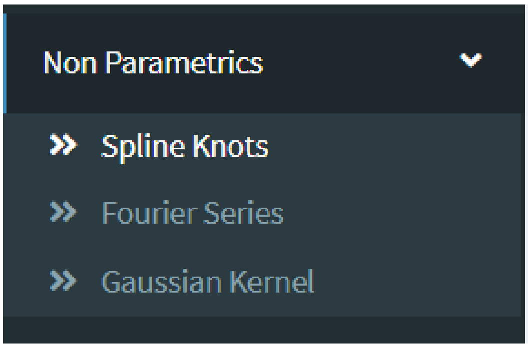


Figure 8 Nonparametrics tab

In this section, there are 3 choices that refer to the estimator in the nonparametric regression approach. The IRegs program that is currently being built has provided 3 estimators that can be selected, including the Spline Truncated estimator, Fourier series estimator, and Gaussian Kernel estimator.

1. Functional Testing

Table 1 I-Regs Summary

No	Feature	Scenario	Expected results	Work
1	Import Dataset	Users enter data in .csv format into the R-Shiny I-Regs dashboard.	The dataset can be imported properly and can be displayed in the Data Tabulation. Selection of response variables (Y) and predictor variables (X)	Yes
2	Displays descriptive statistics and scatterplots	User selects the data summary or scatter split menu	Descriptive statistics for each variable and scatter plot between the response variable (Y) and all predictor variables (X)	Yes
3	Modeling with Parametric Regression	User selects modeling with linear regression analysis	Parameter estimation, simultaneous and partial hypothesis testing, and regression model assumptions.	Yes
4	Modeling with Nonparametric Regression	The user selects the estimator used in the nonparametric regression analysis.	Parameter estimation, simultaneous and partial hypotheses, and graphical	Yes

No	Feature	Scenario	Expected results	Work
			visualization of actual vs. predicted comparisons.	
5	Modeling with Semiparametric Regression	The user selects the estimator used in the semiparametric regression analysis.	Parameter estimation, simultaneous and partial hypotheses, and graphical visualization of actual vs. predicted comparisons.	Yes

4. Conclusions

This research has succeeded in developing an application/system that can be used for regression analysis modeling, especially nonparametric regression and its applications. Internet-Regression Analysis (I-Regs) is an R-Shiny dashboard with features that make it easy for users to get data analysis results. The I-Regs program can be used by users, including students or related stakeholders whose analysis approach uses nonparametric regression. Currently, modeling using a nonparametric regression approach is very much needed, considering that in the era of the Industrial Revolution 4.0, the Big Data era, there are many patterns of relationships between the two variables (ex, predictor variables with response variables) that are not patterned (random). With the existence of IRegs, it can be an alternative that previously had no program that could accommodate, become available, and become a generally applicable, interactive, and informative program, making it easier for users in the data analysis process with a nonparametric regression approach.

References

Afifah, N., Budiantara, I.N. and Latra, I.N. (2017) 'Mixed Estimator of Kernel and Fourier Series in Semiparametric Regression', in *Journal of Physics: Conference Series*. Institute of Physics Publishing. Available at: https://doi.org/10.1088/1742-6596/855/1/012002.

Biao, W.W. and Pourahmadi, M. (2003) *Nonparametric estimation of large covariance matrices of longitudinal data*, *Biometrika*. Available at: http://biomet.oxfordjournals.org/.

Bilodeau, M. (1992) 'Fourier smoother and additive models', *Canadian Journal of Statistics*, 20(3), pp. 257–269. Available at: https://doi.org/10.2307/3315313.

Budiantara, I.N. *et al.* (2019) 'Modeling Percentage of Poor People In Indonesia Using Kernel and Fourier Series Mixed Estimator In Nonparametric Regression', *REVISTA INVESTIGACION OPERACIONAL*, 40(4), pp. 538–550.

Du, P., Parmeter, C.F. and Racine, J.S. (2012) *Nonparametric Kernel Regression with Multiple Predictors and Multiple Shape Constraints*. Department of Economics; McMaster University; Canada.

Eubank, R.L. (1999) 'Nonparametric Regression and Spline Smoothing'. New York: Marcel Dekker.

Guidoum, A.C. (2020) *Kernel Estimator and Bandwidth Selection for Density and its Derivatives: The kedd Package*. Algeria.

Hardle, W. (1994) 'Applied Nonparametric Regression.', in *Humboldt-Universitat zu Berlin, Institut fur Statistik und Okonometrie*. Berlin: Humboldt-Universitat zu Belrin. Available at: https://doi.org/10.2307/2348990.

Hidayat, R. *et al.* (2019) 'Kernel-Spline Estimation of Additive Nonparametric Regression Model', in *IOP Conference Series: Materials Science and Engineering*. Institute of Physics Publishing. Available at: https://doi.org/10.1088/1757-899X/546/5/052028.

Lin, D.Y. and Ying, Z. (2001) ‘Semiparametric and nonparametric regression analysis of longitudinal data’, *Journal of the American Statistical Association*, 96(453), pp. 103–113. Available at: https://doi.org/10.1198/016214501750333018.

Mariati, N.P.A.M., Budiantara, I.N. and Ratnasari, V. (2020) ‘Combination Estimation of Smoothing Spline and Fourier Series in Nonparametric Regression’, *Journal of Mathematics*, 2020. Available at: https://doi.org/10.1155/2020/4712531.

Mariati, N.P.A.M., Budiantara, I.N. and Ratnasari, V. (2021) ‘The application of mixed smoothing spline and fourier series model in nonparametric regression’, *Symmetry*, 13(11). Available at: https://doi.org/10.3390/sym13112094.

Ratnasari, V., Budiantara, I.N. and Dani, A.T.R. (2021) ‘Nonparametric Regression Mixed Estimators of Truncated Spline and Gaussian Kernel based on Cross-Validation (CV), Generalized Cross-Validation (GCV), and Unbiased Risk (UBR) Methods’, *International Journal on Advanced Science, Engineering and Information Technology*, 11(6).

Ratnasari, V., Utama, S.H. and Dani, A.T.R. (2024) ‘Toward Sustainable Development Goals (SDGs) with Statistical Modeling: Recursive Bivariate Binary Probit’, *IAENG International Journal of Applied Mathematics*, 54(8), pp. 1515–1521.

Regier, M.D. and Parker, R.D. (2015) ‘Smoothing using fractional polynomials: An alternative to polynomials and splines in applied research’, *Wiley Interdisciplinary Reviews: Computational Statistics*, 7(4), pp. 275–283. Available at: https://doi.org/10.1002/wics.1355.

Sifriyani *et al.* (2017) ‘Geographically weighted regression with spline approach’, *Far East Journal of Mathematical Sciences*, 101(6), pp. 1183–1196. Available at: https://doi.org/10.17654/MS101061183.

Sifriyani, S. *et al.* (2023) ‘Statistical Modeling: A New Regression Curve Approximation using Mixed Estimators Truncated Spline and Epanechnikov Kernel’, *Engineering Letters*, 31(4), pp. 1–7.

Suparti *et al.* (2019) ‘Modeling longitudinal data based on Fourier regression’, in *Journal of Physics: Conference Series*. Institute of Physics Publishing. Available at: https://doi.org/10.1088/1742-6596/1217/1/012105.

Suparti *et al.* (2021) ‘Biresponses kernel nonparametric regression: inflation and economic growth’, *International Journal of Criminology and Sociology*, 10, pp. 465–471. Available at: https://doi.org/10.6000/1929-4409.2021.10.54.

Wayan Sudiarsa, I. *et al.* (2015) ‘Combined estimator fourier series and spline truncated in multivariable nonparametric regression’, *Applied Mathematical Sciences*, 9(97–100), pp. 4997–5010. Available at: https://doi.org/10.12988/ams.2015.55394.

Widyastuti, D.A., Fernandes, A.A.R. and Pramoedyo, H. (2021) ‘Spline estimation method in nonparametric regression using truncated spline approach’, in *Journal of Physics: Conference Series*. IOP Publishing Ltd. Available at: https://doi.org/10.1088/1742-6596/1872/1/012027.

Wong, W.H. (1983) ‘On the Consistency of Cross-Validation in Kernel Nonparametric Regression’, *The Annals of Statistics*, 11(4), pp. 1136–1141.