IRAQI STATISTICIANS JOURNAL

https://isj.edu.iq/index.php/isj

**ISSN: 3007-1658 (Online)**

# Comparison of Simple Linear Regression and Binary Logistic Linear Regression for Digital Image Segmentation

Alaa Mohammed Ali Mustafa[1], Aseel Muslim Iesa[2], Zainab Falih Hamza[3] and Kareem Khalaf Aazer[4]

[1,2,4] College of Administration & Economics, University of Sumer
[3] College of Business Informatics, University of Information Technology& Communication

**ARTICLE INFO**

***Article history:***
Received 19/11/2024
Revised 20/11/2024
Accepted 13/01/2025
Available online 15/05/2025

***Keywords:***
Simple linear regression
Binary logistic regression
Image processing
Image segmentation
Threshold

**ABSTRACT**

Statistical methods play an important role in image processing. The most importan these methods are the simple linear regression function and the binary logistic regress function, which are used to study the relationship between the dependent variable and independent variable. They are also used in the process of predicting the value of dependent variable at a specific value for the independent variable. In this research, simple linear regression function and the binary logistic regression function were employ in image processing as a statistical technique whose purpose is not to study the relations between the dependent variable and the independent variable or in the prediction proce but rather a tool that works to segment images using the threshold technique, consider the sum of the simple linear regression vector and the binary logistic linear regress vector, which were estimated from the image data as the threshold limit for segment images. The two techniques were good in the segmentation process in terms of giving best segmented images containing important areas that have features that are useful for study, while removing the useless or unimportant areas. The two techniques were compa using the Jaccard scale, which is used to determine which technique was better in segmentation process. It was found that the simple linear regression technique gave clearer segmented image of the features, and thus it is better than the segmented ima using the binary logistic linear regression technique.

## 1. Introduction

Regression analysis is a model-building tool used to describe and understand the relationship between variables that can be used extensively in the prediction process.[4] In some studies, the analytical relationship may be between a quantitatively dependent variable, which depends on a single illustrative variable, called simple linear regression, or depends on a set of illustrative variables and is called multiple linear regression. But if the dependent variable is descriptive, that is, segmentation d into two or more categories and depends on explanatory variables that can be quantitative or descriptive, it is called binary logistic regression. Regression analysis can be used in other areas based on statistical theories such as image processing, which uses most statistical methods in the processes of Segmented, optimization, classification and others in the processing of digital images, so the regression analysis function can be used in the process of image Segmented using the threshold techni Which depends on most statistical tools such as the arithmetic mean and variance, giving satisfactory results to segmentation the image and find the areas that represent the objects or find the important areas of objects in the image by classifying each pixel according to the value of the colour image to group in areas representing these objects. Simple linear regression and binary logistic regression function from digital image data where local threshold, which is one of the threshold slicing methods based on the value of the estimated$\hat{y}$, was used in the image Segmented process [10], the local threshold is used effectively when the gradient effect is small relative to the size of the selected sub-image [5].

The research aims to employ the simple linear regression function and the binary logistic

* Corresponding author.Email: zainab.stat@uoitc.edu.iq
https://doi.org/10.62933/qn001441

regression function to segment selected images and compare the two methods.

In 2023, researcher Eesa [1] presented the use of the binary logistic regression function in the image segmentation process. As for researchers Eesa and Hamza [2] in (2024), they employed the simple linear regression function in the image segmentation process.

### 1.1 General linear regression model

The general linear regression model represents the relationship between the response variable which is connected to a variable or set of illustrative variables as follows [9]:

$$y_i = \beta_0 + \beta_1 x_{1k} + \cdots + \beta_k x_{ik} + u_i \ldots\ldots\ldots (1)$$

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_k x_{ij} + u_i \qquad \ldots\ldots\ \ldots\ldots\ldots (2)$$

$$\mathrm{y} = \mathrm{x}\beta + u \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.\ldots (3)$$

Where

$y_i$ :is variable response.

$\beta_j$: Model parameters that are estimated, j=1,...,k.

$x_{ij}$: Explanatory variables i=1,…,n, j=1,…,k.

$u_i$: The random error $u_i$ has a normal distribution with a mean of zero, and its variance is $\sigma^2$ $u_i$~$\mathrm{N}(0,\sigma^2)$ . It is also assumed that the errors are statistically independent, meaning that $cov(u_i,u_j) = 0$ .

The linear regression model shown in equation (1) assumes several assumptions:

1. The matrix of explanatory variables X is defined, and given Fixed, and its size is without errors.
2. It is assumed that there is no perfect or almost perfect linear relationship between the illustrative variables, i.e. $Rank(X) = (k+1) < n$.
3. There is no relationship between independent observations ($x_{ik}$) and random error $u_i$ i.e. the columns of the matrix x are linearly independent of the random error vector.
4. The random error $u_i$ has a normal distribution with a mean of zero, and its variance , $\sigma^2$ ($u_i$~$\mathrm{N}(0,\sigma^2)$, and the errors are assumed to be statistically independent, i.e. cov($u_i$, $u_j$) = 0.
5. The linear regression model is called a simple linear model if the response variable is related to one illustrative variable and is linear in its parameters and is in the form .

$$y_i = \beta_0 + \beta_1 x_i + u_i \ldots\ldots\ldots\ldots\ldots (4)$$

The features of the linear regression model are estimated by a set of methods and in this research the features were estimated by the usual least squares method as follows:

$$u'u = (y - x\beta)'(y - x\beta)$$

We derive the above equation for β and produce the following:

$$\frac{\partial u'u}{\partial \beta'} = -x'(y - x\hat{\beta}) = 0$$

$$x'x\hat{\beta} = x'y \qquad \ldots\ldots\ldots\ldots\ldots (5)$$

Multiply an equation (4)$(x'\mathrm{x})^{-1}$ Vintage

$$\hat{\beta} = (x'\mathrm{x})^{-1}x'y$$

Therefore, the estimated regression function will be as follows

$$\hat{y} = x\hat{\beta} \qquad \ldots\ldots\ldots\ldots\ldots (6)$$

The following algorithm illustrates the image segmentation process based on the regression function:

1- Enter the image and convert it to grayscale image.

2- Through the histogram, the dependent variable and the independent variable are calculated from the grayscale image images.

3- Estimating the parameters of the model through the following formula :

$$\hat{\beta} = (\mathrm{x}'\mathrm{x})^{-1}\mathrm{x}'\mathrm{y}$$

4- Find the estimated regression function $\hat{\mathrm{y}} = \mathrm{x}\hat{\beta}$

5- Use $\hat{y}$ as a local threshold for cropping the image

6- If $x(i,j) \geq \hat{\mathrm{y}}$ then y(i, j)=1 but if $x(i,j) < \hat{\mathrm{y}}$ then $\mathrm{y(i,j)} = 0$

7- Extract the segmented image.

### 1.2. The binary logistic regression model

The binary logistic regression function[7] is represented by the following formula:

$y_i \sim bernoulli(\pi_i)$ ……………………..…(7)

$\pi_i = pr(y_i = 1 / x_{i1}, x_{i2}, \dots, x_{ip})$

$$\pi_i = \frac{e^{x\beta}}{1 - e^{x\beta}} \quad \dots (8)$$

The probability $0 \le \pi_i \le 1$, while the ratio $\frac{\pi_i}{1-\pi_i}$ is a positive value confined between $(0.\infty)$ and when entering the natural logarithm, the estimated is $-\infty \le \ln\frac{\pi_i}{1-\pi_i} \le \infty$ Therefore, the model can be written as follows:

$$\ln\frac{\pi_i}{1 - \pi_i} = x\beta \quad \dots (9)$$

$$x\beta = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \quad \dots (10)$$

The $y_i$of the response variable and $\beta_j$ represent the parameters of the model that are estimated.
$x_{ij}$The explanatory variables are $i$=1,…,, $j$=1,…,$p$ .

Simple linear regression function used

$$x\beta = \beta_0 + \beta_1 x_i \quad \dots (11)$$

The features of the binary logistic regression model are estimated in the way of maximum likelihood, and this method chooses the values of estimating the features to be the most likely viewing results to be selected, and estimating the features requires the use of the iterative algorithm, and one of the most important features of estimating the method of the greatest possibility is that its distribution follows the asymptotic distribution in the case of the sample size is large. The difference between the bone possibility method and the method of estimation by the usual least squares method is that the square of the difference between the observation values and the prediction values of the dependent variable is less [6].

Through the Odds Ratio (OR) $\frac{\pi_i}{1-\pi_i}$ for the success case ($y_i$=1), the ratio is estimated by the following formula:

$$OR = e^{x\hat{\beta}} = \frac{\hat{\pi}_i}{1 - \hat{\pi}_i} \quad \dots (12)$$

The individual number i must be classified to failure , $y$=1 if $OR > 1$.

The individual number i must be classified to success, $y = 0$ if $< 1$ .

The binary logistic regression function was used in the process of image Segmented[12,11] as a tool that works to segmentation the image into parts or objects that make it up, and it is based on the properties of discontinuity and similarity, and most of the similarity algorithms are based on thresholding[12,4]in addition to the algorithms for expanding (magnification) areas, splitting and merging areas. Slicing algorithms based on reproach are based on the basic assumption that objects and backgrounds in the image have distributions that are grayscale image-level. The selection of the threshold is one of the methods to separate objects from the background by identifying the areas that have grayscale image levels below the threshold in the background and identifying the areas that have grayscale image levels higher than the threshold value in objects, or vice versa for the size of the selected sub image [5].

The following algorithm illustrates the image segmentation process based on the binary logistic regression function:

1- Enter the image and convert it to grayscale image.
2- Through the histogram, the dependent variable and the independent variable are calculated from the grayscale image images.
3- Estimating the parameters of the model in a way that can be boned.
4- Finding the Estimated Binary Logistic Regression Function $\hat{\pi}_i = \frac{e^{x\hat{\beta}}}{1-e^{x\hat{\beta}}}$
5- Use π ̂as a local threshold to crop the image.
6- If $x_{(i,j)} > \hat{\pi}$ then $y_{(i,j)} = 1$ but if If $x_{(i,j)} < \hat{\pi}$ then $y_{(i,j)} = 0$
7- Extract the segmented image .

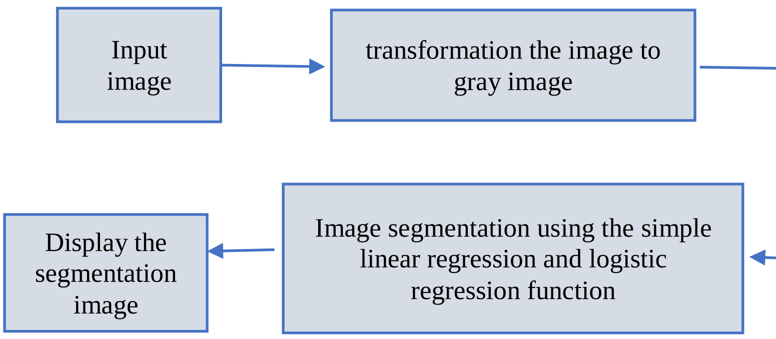


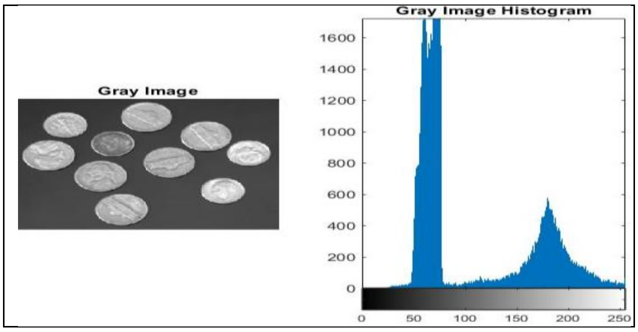
Figure (1) Flow chart of the process of segmentation the image

## 2. Methodology

To find out the performance of the threshold methods mentioned in the theoretical aspect, which works to segmentation the approved image into only two areas front and back, these methods were applied in the ATLAB2017a program, where the image was segmentation using the simple linear regression function and the binary logistic regression function and comparing the two methods with each other, and the techniques were applied to a set of images mentioned above:


Figure (2) shows the original images

In Figure (3), the two original images in Figure (2) were converted to grayscale image images as shown below:



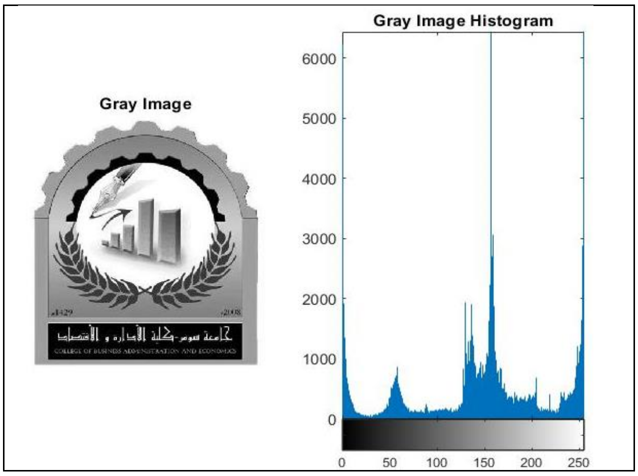
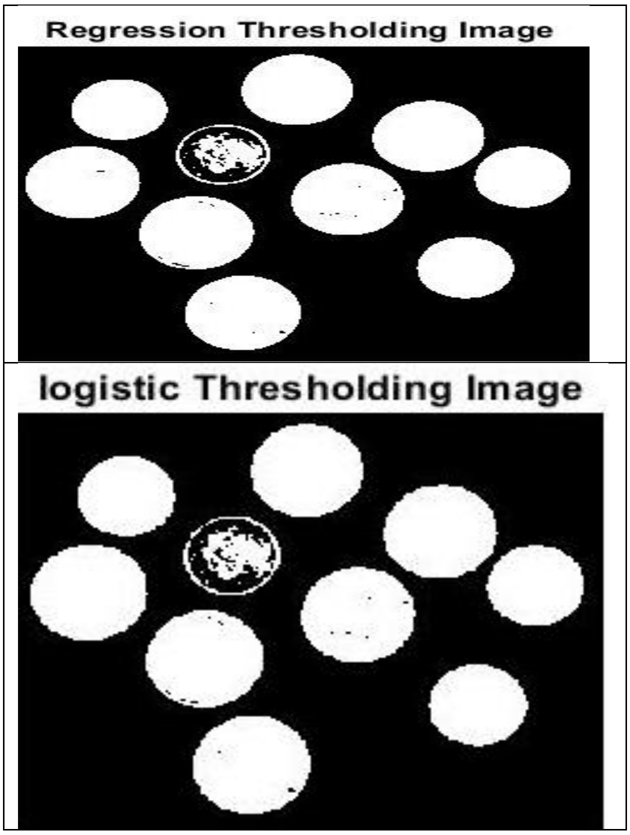
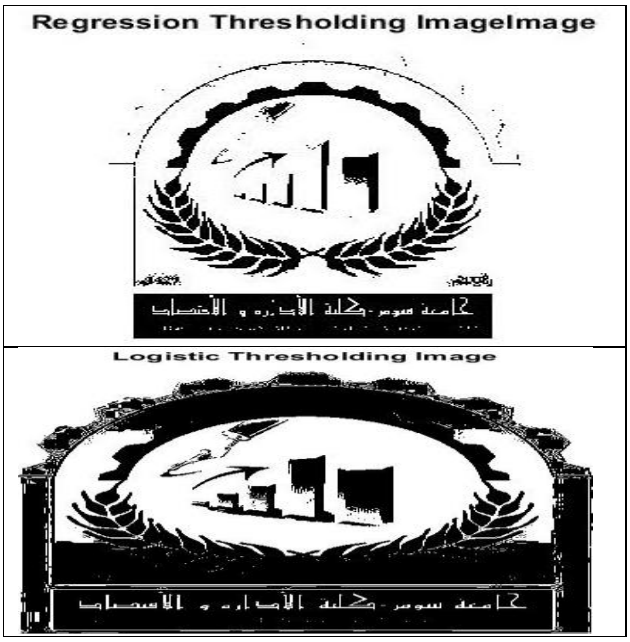




Figure (3) shows two grayscale image images on the left side and on the right side the frequency distribution of the two images.






Figure(4)shows the two images segmentation by the simple linear regression function and the binary logistic regression function

Figure (3) shows the segmentation images by the threshold process using the simple linear regression function and the binary logistic regression function after the segmentation process, we note that the resulting images are black and white binary images, where the two methods were good in giving segmentation images containing the most important features and properties while removing unimportant features. To find out which methods gave better results, the MSE scale and the Jaccard scale were used, and the values are shown in Table (1).

**Table (1)** shows the values of the estimators of the two images

| Method | Images | | Estimate | Jaccard | MSE |
|---|---|---|---|---|---|
| Regression linear method | Image1 | $\beta_0$ | 0.5375 | 0.6987 | 0.2618 |
| | | $\beta_1$ | -0.0001 | | |
| | Image2 | $\beta_0$ | 124.7001 | 0.2079 | 0.5006 |
| | | $\beta_1$ | 0.0041 | | |
| binary logistic regression method | Image1 | $\beta_0$ | -6.186 | 0.7382 | 0.3013 |
| | | $\beta_1$ | 1.0289 | | |
| | Image2 | $\beta_0$ | -5.805 | 0.4811 | 0.7642 |
| | | $\beta_1$ | -1.1537 | | |

The table above shows the values of the estimators and the test measures extracted from the segmented images.

## 3. Results and discussion

It is known that images contain pixels, which will be the sample items. Since the image has a size of $n \times m$, the more pixels in the image, the higher the resolution of the image and the clearer and more features, so a high-resolution image will be produced. In the practical aspect, the original image was converted to a grayscale image, and the number of pixels was determined between 0 and 255. We note that both methods gave good results in segmenting the image, but the simple linear regression technique was better than the binary logistic regression technique, as proven by Table (1), where the Jaccard and MSE measures were lower.

## 4. Conclusion

In this research, the estimated regression function and the binary logistic regression function were used to segment the image based on the threshold, where the two functions were considered a local threshold. The two techniques proved their efficiency in providing a segmented image containing the most important features while removing unimportant features and extracting all image features. When using comparison metrics, it was found that the image segmentation technique based on the regression function estimated to compare two images was the best because it was able to give the most important features.

## References

[1] A. M. Eesa ,(2023),"Using the Binary Logistic Regression Analysis Function for Image Processing", Journal of Positive Sciences, Issue:2, Volume(3), Page(13-16), ISSN:2582-9351.

[2] A.M.Eesa. and Z.F. Hamza, (2023) ,Using A Simple Linear Regression Function to Segmentation Image of the Hamrin Marshes for the Years (2020 /2021 -2021 / 2022) , Issue:54, Page (68-74) , ISSN:2790-2293.

[3] A. Khashaa Mahmoud, 1987, "The Introduction to Regression Analysis", University of Mosul, Ministry of Higher Education and Scientific Research.

[4] C.Ashutosh Kumar , " Comparison of The Local and Global Thresholding Methods in Image Segmentation", World Journal of Research and Review (WJRR) ISSN:2455-3956, Volume-2, Issue-1, January 2016 Pages 01-04 .

[5] D. Barten " Contrast Sensitivity OF Human Eye and its Effect on Image Quality " , SPIE – The

International Society For Optical Engineering Belling Ham , 1999.

[6] D. J. Hand, (2012)," Statistics a very short introduction".

[7] H. R. Talib, (2019), " Use of multiple response logistic regression to identify factors affecting eye disease ",

[8] J. Yuan, D. Wang, and Rongxing Li,(2011),"Image Segmentation Based on Local Spectral Histograms and Linear Regression", Proceedings of International Joint Conference on Neural Networks, San Jose, California, USA.

[9] Joel Best, Damned Lies and Statistics: Untangling Numbers from the Media, Politicians, and Activists (Berkeley: University of California Press, 2001).

[10] Joko Hariaji,(2021)," Simple Linear Regression (SLR) Model and Multiple Linear Regression (MLR) Model" .

[11] K. Bhargavi and S. Jyothi,( 2014)," A Survey on Threshold Based Segmentation Technique in Image Processing", International Journal of Innovative Research & Development, November, (Special Issue) Vol 3 Issue 12.

[12] R.C. Gonzalez, R.E. Woods (2008), Digital Image Processing, Prentice Hall, Englewood Cliffs, Nj