SCIREA Journal of Mathematics
ISSN: 2995-5823
http://www.scirea.org/journal/Mathematics
May 4, 2025
Volume 10, Issue 3, June 2025
https://doi.org/10.54647/mathematics110531

Utilizing the Chi-square Technique in Fitting a Logistic Distribution to the Heights of Students of Akwa Ibom State University

Itoro Tim Michael [1,*] , Iseh, Matthew Joshua [2]

[1] Department of Statistics, Akwa Ibom State University, Nigeria

[2] Department of Statistics, Akwa Ibom State University, Mkpat Enin, Nigeria, Email: matthewiseh@aksu.edu.ng ORCID ID: https://orcid.org/0000-0003-2696-7319

*Corresponding author E-mail: itoromichael@aksu.edu.ng

Abstract

Goodness of fit tests indicate whether or not it is reasonable to assume that a random sample comes from a specific probability distribution. Again, it is established in the literature that the logistic distribution is a well-known probability model that, with very few exceptions, behaves similarly to the normal distribution with comparable measures of dispersion. On the foregoing, this utilizes the chi-squared technique to fit a logistic distribution to the heights of Students of Akwa Ibom State University. The study comprises of the heights of 617 students collected from the Akwa Ibom State University Main Campus. A chi-squared test is used to ascertain whether or not the heights of Students are logistic distributed. The visual representation of the simulated and real data with the same parameter value are presented. It is observed from the results that the logistic distribution cannot fit the heights of students of the Akwa Ibom State University at the significance level $\alpha = 0.01\%$. The graphs of the heights

of students, the simulated heights of students and the logistic densities values also showed a great disparity.

Keywords: Chi-square test, logistic distribution, heights of students, maximum likelihood estimates

1.0Introduction

Statistical data are obtained from all areas of studies; Engineering, Medicine, Sciences, art etc and different probability models are available to aid in extracting useful information from the datasets. The data sets possess varieties of properties such as skewness, unimodality, bimodality and a couple of others. Some of the probability models include the normal, gamma, Weibull, Rayleigh, Logistic, and Two-parameter Burr Type X distributions.

In statistics, the logistic distribution is a well-known probability model that, with very few exceptions, behaves similarly to the normal distribution with comparable measures of dispersion. The median, mean, and mode are all equal and have characteristics in common with the normal distribution.[1], [2]. Following this, it is commonly recognized in the literature that, when plotted on a graph, human height often follows the normal distribution, forming a bell-shaped curve. Nonetheless, the study's goals are made possible by the small deviations in the logistic distribution's resilience and behavioral characteristics with regard to outliers.

Again, [3] and [4] fitted the normal and log-normal distributions to the weights of students of the Akwa Ibom State University using the Chi-squared approach by splitting the students' weights into different cells to obtain the observed values and using the raw data for the maximum likelihood estimation of normal and logistic models' parameters; the mean and standard deviation, thereafter, calculating the cells probability and the chi-squared value.

Goodness of fit tests indicate whether or not it is reasonable to assume that a random sample comes from a specific probability distribution. According to [5] and [6] measures of goodness of fit typically summarize the discrepancy between observed and expected values under the model considered and such measures can be used in statistical hypothesis testing to test for; normality of residuals, whether two samples are drawn from identical distributions or whether outcome frequencies follow a specified distribution and others.

In what follows, [7] introduced the Anderson-Darling test, a statistical test of whether a given sample data is drawn from a given probability distribution with no parameter to be estimated. [8] introduced the Shapiro-Wilk test to test the null hypothesis that the random samples constituting a random variable comes from a normally distributed population. [9] introduced the D'Agostino's K2 test, a goodness of fit measure of departure from normality; the test aims to establish whether or not the given sample comes from a normally distributed population. [10] investigated the properties of Pearson's chi-squared test. Pearson chi-squared test tests a null hypothesis that the frequency distribution of certain events observed in a sample is consistent with a particular theoretical distribution. [11] introduced the Lilliefors test, a normality test based on the Kolmogorov-Smirnov test. It is used to test the null hypothesis that data come from a normally distributed population, when the null hypothesis does not specify which normal distribution

This work fits the Logistic distribution to the heights of Akwa Ibom State University Students using the Chi-Squared test and graphical method. The heights of 617 students of the Akwa Ibom State University were collected from the Medical Centre, Main Campus, Ikot Akpaden.

2.0 Methodology

This paper uses two methods for testing or verifying if logistic distribution fits the heights of Akwa Ibom State University Students; the graphical method and the chi-squared methods.

2.1 The graphical method

The graphical methods involve the use of graphical tools to display box plots, histogram and density plot of the given data sets and comparing same with that of the theoretical distribution. In this research work, we display the logistic density plots for the raw and the simulated datasets and the heights of students of the Akwa Ibom State University

2.2 The chi-squared method

According to [12], [10] proposed the following test statistics, which is a function of the deviations of the observed counts from their expected values, weighted by the reciprocals of their expected values. Thus,

$$\chi^2_{k-1} = \sum_{i=1}^{k} \frac{[n_i - E(n_i)]^2}{E(n_i)} = \sum_{i=1}^{k} \frac{[n_i - np_i]^2}{np_i} \quad (1)$$

called the Pearson chi-squared test and denoted by χ^2_{k-1} with $k-1$ degrees of freedom.

Where:

n_i = an observed frequency (i.e. count) for n_i

$E(n_i) = np_i$ an expected frequency for n_i asserted by the null hypothesis.

n = the sample size

2.3 The Logistic Probability Distribution Model

According to [13], a random variable X is said to have a logistic distribution with parameter θ if and only if the density function of X is

$$f(x) = \begin{cases} \frac{\exp\{-(x-\theta)\}}{(1+\exp\{-(x-\theta)\})^2} & , -\infty < x < \infty, -\infty < \theta < \infty \\ 0, & otherwise \end{cases} \quad (2)$$

Consider a random sample $X_1, X_2, \ldots, X_n$ of size n from this distribution in Eq. 2, and if we denote the frequency of n_i, $i = 1,2,3,\ldots,k$ so that $n_1 + n_2 + \ldots + n_k = n$, then the random χ^2_{k-1} variable in Eq.1 cannot be computed once $n_1, n_2, \cdots, n_k$ have been observed, since each p_i, and hence χ^2_{k-1}, is a function of θ. The value of θ that minimize χ^2_{k-1} are difficult to compute therefore, their maximum likelihood estimates are used to evaluate p_i and χ^2_{k-1}. Using maximum likelihood estimates of the parameters in place of minimum chi-square estimates tend to lead to the rejection of the null hypothesis since the χ^2_{k-1} value is not minimized by maximum likelihood estimates, and as such the computed value is somewhat greater than it would be if minimum chi-square estimates are used.

2.4 Research hypothesis

The Null hypothesis (H_0): The height of students follows a logistic distribution

The Alternative Hypothesis (H_1): The height of students does not follow a logistic distribution.

2.5 Log-Likelihood function of the Logistic Model Parameter

The likelihood function of the logistic distribution is defined as

$$L = \prod_{i=1}^{n} f(x_i, \theta)$$

And the log-likelihood function is given as

$$\ell = \sum_{i=1}^{n} \log\left(f(x_i, \theta)\right)$$

by applying equations Eqs. 2 and 3 to Eq.1, we have

$$L = \frac{\exp\{-(x_i-\theta)\}}{\prod_{i=1}^{n}(1+\exp\{-(x_i-\theta)\})^2}$$

And

$$\ell(\theta) =-\sum_{i=1}^{n}((x_i-\theta)+2\log(1+\exp\{-(x_i-\theta)\})) \quad (3)$$

2.6 Estimation of the Logistic Model Parameter Using MaxLik package in R

[14] introduced the maxLik package in R for maximum likelihood estimation of a model's parameter. The package is very helpful in estimating the parameters of a given distribution and a data set. We make use of the paper.

```
X=HEIGHT of Students
library(maxLik)
logliklogis<-function(logis){
logis<-logis[1]
sum(-X+logis-2*log(1+exp(-(X-logis))))
}
mlelogis<-maxLik(logLik=logliklogis,start=c(logis=5))
summary(mlelogis) ##
logis=coef(mlelogis)
logis # θ̂ = 1.657743
```

3.0 Results and discussion

3.1 Graphical display

Figure 1 is a plot of the simulated data obtained from the logistic distribution for a sample 617 with $\hat{\theta} = 1.657743$. The graph of the simulated data points resembles a bell-shaped which shares similarity with that of the normal distribution.

Figure 2 is a plot of the logistic distribution density values and the heights of students obtained from the Medical Centre of the Akwa Ibom State University. The graph shows a great disparity from that of the simulated indicating that the data does not follows the logistic probability distribution.

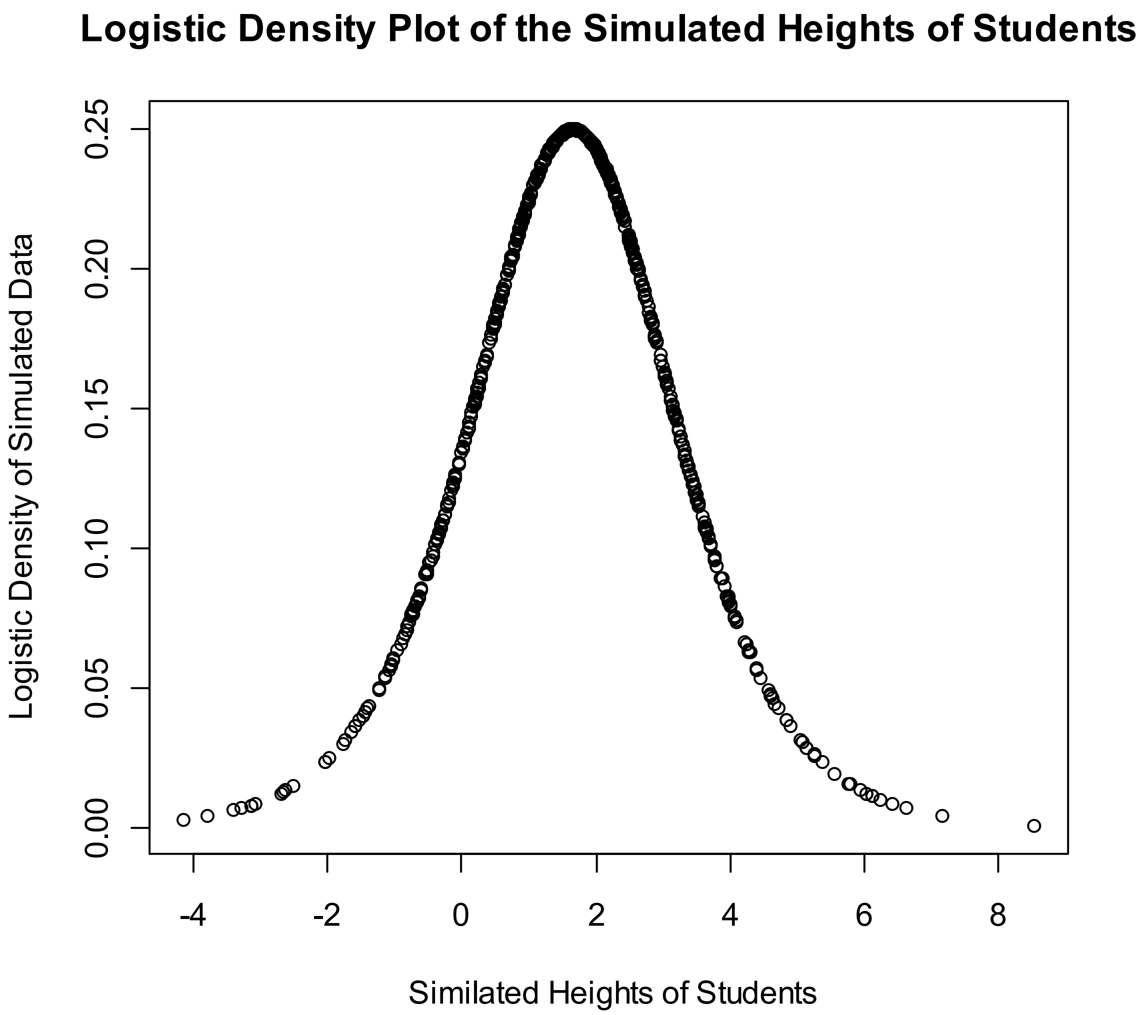
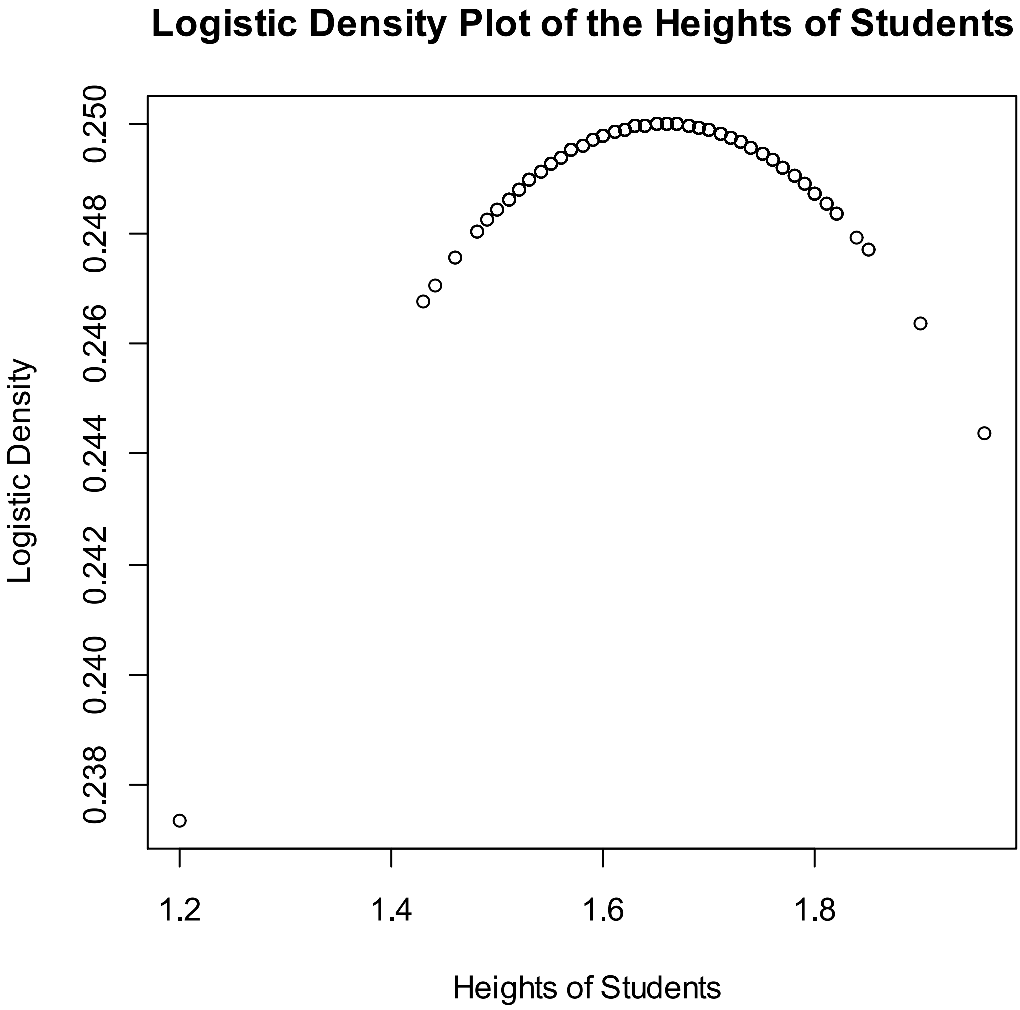


Figure 1: Logistic Density Plot of the Simulated Heights of Students



3.2 Computation of The Logistic Distribution Cells Probabilities

The random variable X, denoting the heights of students is partitioned into the following $k = 8$ mutually disjoint sets:

$M_1 = \{-\infty < X \leq 1.5\}, M_2 = \{1.5 < X \leq 1.55\}, M_3 = \{1.55 < X \leq 1.6\}, M_4 = \{1.6 < X \leq 1.65\}, M_5 = \{1.65 < X \leq 1.7\}, M_6 = \{1.7 < X \leq 1.75\}, M_7 = \{1.75 < X \leq 1.8\}, M_8 = \{1.8 < X \leq \infty\}$

Let $p(M_i) = p_i$, $i = 1,2, \ldots, k,$ where p_i is the probability that the outcome of the random experiment is an element of the set M_i from the logistic probability distribution. The probabilities are obtained as follows:

$$p_i = \int_a^{\mathrm{b}} \frac{\exp\{-(x-\theta)\}}{(1+\exp\{-(x-\theta)\})^2} dx \, , i = 1,2,3, \ldots, 8 \qquad (4)$$

where a and b are the lower and upper limits for each M_i, $i = 1,2,3, \ldots, 8$

The Table 1 shows the calculated probabilities, Observed Frequencies and Expected Frequencies of the Heights of Students and the Logistic Distribution obtained from (4)

Table 1. The calculated probabilities, Observed Frequencies and Expected Frequencies

Cells(i)	Sets (M_i)	Observed Frequencies (X_i)	Probabilities (p_i)	Expected Frequencies ($\mathrm{n}p_i$)	$X_i - \mathrm{n}p_i$	$(X_i - \mathrm{n}p_i)^2$	$\frac{(X_i - \mathrm{n}p_i)}{\mathrm{n}p_i}$
1	(-∞, 1.5]	13	0.4606	284.2185	-271.2185 +	73559.484	258.8131
2	(1.5,1.55]	57	0.0124	7.6782	49.3218	2432.637	316.8226
3	(1.55,1.6]	82	0.0125	7.6989	74.3011	5520.651	717.0687
4	(1.6,1.65]	146	0.0125	7.7100	138.2900	19124.115	2480.4197
5	(1.65,1.7]	150	0.0125	7.7115	142.2885	20246.010	2625.4226
6	(1.7,1.75]	101	0.0125	7.7034	93.2966	8704.258	1129.9265
7	(1.75,1.8]	53	0.0125	7.6857	45.3144	2053.390	267.1718
8	(1.8,∞)	15	0.4645	286.5937	-271.5937	73763.161	257.3788
Total							8053.024

The Test Statistic

$$\chi^2_{\mathrm{k}-2} = \sum_{i=1}^{\mathrm{n}} \frac{(\mathrm{n}_i - \mathrm{n}p_i)^2}{\mathrm{n}p_i} \tag{5}$$

The test statistic in Eq.5 where n_i and np_i denote the observed and expected frequencies respectively with $\mathrm{k} - 2 =$ the degree of freedom.

$$\chi^2_{\mathrm{k}-2} = \sum_{i=1}^{\mathrm{n}} \frac{(\mathrm{n}_i - \mathrm{n}p_i)^2}{\mathrm{n}p_i} = 8053.024 \tag{6}$$

3.6 Significant levels and critical values

The degree of freedom $(df) = k - s - 1 = 8 - 1 - 1 = 6$, where k is the number of cells and s, the number of parameters estimated. Some significance levels and their corresponding critical values are presented in Table 3.

Table 2: Significance Levels and Corresponding Critical Values for $df = 6$

Significance Level	Critical Values	Degree of Freedom
0.0001	27.8563	6
0.0011	22.2299	6
0.0021	20.6729	6
0.0031	19.7245	6
0.0041	19.0380	6
0.0051	18.4985	6
0.0061	18.0535	6
0.0071	17.6742	6
0.0081	17.3436	6
0.0091	17.0503	6

The Decision Rule

Reject H_o if $\chi^2_{k-2} > \chi^2_{crit}$, where χ^2_{k-2} is the computed value of the test statistic and χ^2_{crit} is the critical value as given in table 3.

4.0 Conclusion

It is observed from Table 2 that $\chi^2_{k-2} = 8053.024 > \chi^2_{crit} = 27.8563$ when the significance level $\alpha \geq 0.0001$ or 0.01% . Hence, the heights of students of Akwa Ibom State University does not follow a logistic distribution using the chi-squared test. This result is also supported by the graphs of the simulated and real data obtained from the Medical Centre of the Akwa Ibom State University.

Acknowledgement

The first author appreciates the Director of Health Services, Dr Geoffrey C. Udoh and the Head of Medical Laboratory Unit, Mr. Aniekan B. Okure of the Akwa Ibom state University Medical Centre for their help during the data collection process from the Medical Centre for this study.

R codes for the graphs

```
plot(logissim,ysim,
main="Logistic Density Plot of the Simulated Heights of Students",
ylab="Logistic Density of Simulated Data",xla=" Similated Heights of Students")
plot(X,y,main="Logistic Density Plot of the Heights of Students",
ylab="Logistic Density",xlab=" Heights of Students")
```

References

[1] Iseh, M., Ekong, N., Usoro, A., and Ukpe, I. (2022). Juxtaposing Vertically Transmitted Infections (VTIs) and the Spread of HIV/AIDS in a Typically Infection Prevalent Region in Nigeria. *Journal of the Society of Physical Sciences*, DOI:10,46481/jnsps.2022.418. Vol. 4(2022) 99-104.

[2] Iseh, Matthew Joshua & Udoh, Nse Sunday (2022). Modelling The Risk Index Of Infant Mortality In Nigeria. Abacus (Mathematics Science Series) Vol. 49 (2):350-373.

[3] Michael, I. T., Ikpang, I. N. & Isaac, A. A (2017). Goodness of fit test: a chi-squared approach to fitting of a normal distribution to the weights of students of Akwa Ibom State University, Nigeria. *Asian Journal of Natural and Applied Sciences* 6(4) 107 – 113.

[4] Michael, I.T., Iseh, M.J., Ikpang, I. N, George, E. U. & Bassey, D. I. (2021). A Goodness of fit test: Using chi-squared approach to fitting of a log-normal probability model to the weights of students of Akwa Ibom State University, Nigeria. *International Journal of Applied Science a*nd Research 4(6) 146 – 152.

[5] Michael, I. T., Usoro,A. E., Ikpang, I. N., George, E. U. & Bassey, D. I.(2019). Fitting a Gamma Distribution using a chi-squared approach to the Heights of students of Akwa Ibom State University, Nigeria. *International Journal of Advanced Statistics and Probability* 7(2) 42 – 46.

[6] Ekong N., Moffat I., Usoro A., and Iseh M. (2021). A comparative study study of the impact of dummy variables on regression coefficients and canonical correlation

indices: an empirical perspective. *International Journal of Analysis and application*, Vol. 19(4): 576-586.

[7] Anderson, T. W. & Darling, D.A (1952). Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *Annals of Mathe-matical Statistics.* Vol.23: 193–212. https://doi.org/10.1214/aoms/1177729437.

[8] Shapiro, S. S. & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika.* 52 (3–4): 591–611. https://doi.org/10.1093/biomet/52.3-4.591.

[9] D'Agostino, R. B. (1970). Transformation to normality of the null distribution of g1. *Biometrika.* 57 (3): 679–681. JSTOR 2334794 https://doi.org/10.1093/biomet/57.3.679.

[10] Pearson, K. (1990). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5*. 50(302): 157-175. https://doi.org/10.1080/14786440009463897.

[11] Lilliefors, H. (1967). On the Kolmogorov–Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Asso-ciation*. 62; 399–402. https://doi.org/10.1080/01621459.1967.10482916.

[12] Wackerly, D. D., Mendenhall, W. & Scheaffer, L. R. (2008). Mathematical statistics with applications, 7th Ed., USA: Thomson Higher Education, Inc.

[13] Hogg, R. V., McKean, J. W. & Craig, A. T. (2013). *Introduction to Mathematical Statistics, 7th Ed.*, Boston: Pearson Education, Inc.

[14] Henningsen, A. & Toomet, O. (2009). MaxLik: Tools for Maximum Likelihood Estimation. R package version 0.5, Retrieved September 10, 2017 from *http://CRAN.R- project.org.*