Using Maximum Likelihood Method to Estimate Parameters of the Linear Regression T Truncated Model

Elham Abduklreem Hussain*[1] *Ahmad N.S.Al-Shallawi*[2], *Haifa Abduljawaad Saied
e,a,hussain@ntu.edu.iq , a.n.s.al-shallawi@ntu.edu.iq [2] hayfaajwad.65@uomosul.edu.iq

[1,2] Administrative Technical College / Mosul - Northern Technical University,
[3] College of Computer Science and Mathematics - University of Mosul

Corresponding author: ***Elham Abduklreem Hussain***, e-mail: e,a,hussain@ntu.edu.iq
Co-authors: ANS, e-mail a.n.s.al-shallawi@ntu.edu.iq, HAS, email hayfaajwad.65@uomosul.edu.iq

Received: 29-07-2022, **Accepted:** 11-10-2022, **Published online:** 20-12-2022

Abstract. The estimation of the parameters is one of an important issues in the mathematical statistics. The development of estimation methods requires accurate estimation and finds the best estimator for parameters.

The aim of this research is to build a regression model. A dependent variable has a truncated t-distribution for this model.

It was depended to the method of maximum likelihood for finding the parameters of the model, the parameters β and σ^2 were estimated when they were unknown, the approximation of the cumulative function of the truncated t-distribution from two sides is used to represent the function of the dependent variable.

It was concluded that the value β is equal β_{ols} plus a certain amount, this amount is equal zero when the values of the truncated points a, b are equal in value and different in sign.

As a practical application, simulation was used in data generation by using the program (Matlab 2020a), the value of β was estimated, The comparison was between maximum likelihood method and according to the formula that we reached with its estimated value according to the least squares method based on the mean square errors, where its value according to the maximum likelihood method was less than its value according to the method of least squares, which indicates the advantage of the first method over the second.

Keywords: maximum likelihood, truncate, t-distribution, regression, cumulative function

Introduction

The truncated distributions appeared in several areas of application and largely in the industrial context. The final product is put under test before sending it to the customer. This test includes whether the product's performance falls within certain limits (tolerance). The product is judged by its quality and sent to the customer; otherwise, this product will be rejected and returned to production. In this case, the nature of the distribution to the customer is a truncated distribution [1].The truncated distributions are very important and have wide uses in statistics [2].

The truncation process affects the properties of the probability density function of the random variable Y, as the probability of the random variables space after truncation will be less than one, and this requires finding a new probability distribution, moments, measures, and estimating the parameters and characteristics of the new distribution without need to resort to the original distribution of the random variable [3].Intuitively, the general form of the truncated distribution is by dividing the probability density function by the cumulative function of the distribution at certain truncation points, so we will only address the method of estimating the truncated t-distribution, which is maximum likelihood estimation. [4] mentioned that the first results of the truncated distributions that developed later depended mainly on the multivariate for the standard normal distribution, and there were two main results presented in the years 1953 and 1961.Thus, the use of estimation for truncated regression models has increased, and these models are related to truncated distributions, whether they are truncated on one side or on two sides, and the estimation methods depended on estimation using the Bayesian method or maximum likelihood method [5].

Materials and Methods:

Mixed Model

The mixed distribution is called (compound distribution).It is doubling or aggregation of non-heterogeneous components of statistical distributions, which occurs when a sample is drawn from non-heterogeneous populations with different or similar probability functions with different parameters for each partial population and requires statistical tests to be conducted to find out the extent of the distribution Mixed belongs to the same family or not, rather than the singular distribution. Mixed distributions arise when the population is non-heterogeneous [6].

Since the t-distribution is the division of the normal distribution by the chi-square distribution with n degrees of freedom [7],but this method is complex to reach at the t-distribution function, so the t-distribution can be expressed as a mixed distribution of the normal distribution and the inverse gamma distribution, and this leads to a thicker tail than the tail of the normal distribution [8]. This formula is used in statistical modeling of the t-distribution in conventional statistics and Bayesian statistics [9].

Truncated t Regression Model

The classic linear regression model is as in the following equation:

$$y_i^* = \beta x_i^* + \varepsilon_i \qquad \dots(1)$$

whereas :

x_i^* :represents the independent variable that is not truncated.

β: the parameters of the model.

ε_i: model error, the t-distribution has a mean of zero and a specific variance $\varepsilon_i \sim t(0, \sigma^2, v)$ and is independent of x_i^*.

y_i^*: the dependent variable is untruncated,and the t-distribution is distributed with a specific mean, variance, and degree of freedom υ, as follows:

$$y_i^* | x_i^* \sim t(\beta x_i^*, \sigma^2, v)$$

As for the non-truncated regression model, it can be defined first as the model in which the sample observations of the dependent variable are missing at certain points, offset by the loss of the sample's observations of the independent variables at the same point[10].

If the truncation is from two sides at the values (a,b), the probability density function for the variable y is as follows:

$$f(y|x,\beta,\sigma^2) = \frac{f\left(\frac{y - x\beta}{\sigma}\right)}{F\left(\frac{a - x\beta}{\sigma}\right) - F\left(\frac{b - x\beta}{\sigma}\right)} \qquad \dots(2)$$

The numerator in equation (2) represents the probability density function of the variable y and its value is between two values (a,b), and the denominator represents the cumulative distribution at the truncating points a and b.

Since the dependent variable y represents a conditional mixed distribution with a second variable, the probability density function is as follows:

$$f(y|\tau, x, \beta, \sigma^2; (a, b)) = \frac{t(y|\tau, x, \beta, \sigma^2, v)}{\int_a^b t(y|\tau, x, \beta, \sigma^2, v)}$$

$$= \frac{\int_0^\infty \frac{1}{\sqrt{2\pi\sigma\tau}} e^{-\frac{\sum(y-x\beta)^2}{2\sigma^2\tau}} g(\tau)d\tau}{\int_a^b \int_0^\infty \frac{1}{\sqrt{2\pi\sigma\tau}} e^{-\frac{\sum(y-x\beta)^2}{2\sigma^2\tau}} g(\tau)d\tau dy} \quad a < y < b \quad \ldots (3)$$

Methods

Maximum likelihood estimations for the parameters of the truncated t regression model.

The estimating parameters is one of an important issues that have attracted the attention of researchers and those interested in mathematical statistics, because of the development of estimation methods, this requires accuracy in estimation and finding the best estimator for these parameters [11]. [12] mentioned that Cohen was the first to use a function The greatest possibility in estimating the parameters of the truncated distribution from one side and from two sides, where [13] presented in his research the problem of estimating the mean and variance of the truncated normal distribution with giving numerical examples as a practical application of his findings. Here, the maximum likelihood method will be used to find the parameters of the truncated t-regression model, and this method has certain advantages represented by consistency and efficiency [2].

The truncated t-regression model has a maximum likelihood function as follows[14]:

$$lnL = \ln\left(\frac{1}{\sqrt{2\pi\sigma\tau}} e^{-\frac{\sum(y-x\beta)^2}{2\sigma^2\tau}} g(\tau)\right) - \ln\left(F\left(\frac{b-x\beta}{\sigma}\right) - F\left(\frac{a-x\beta}{\sigma}\right)\right) \quad \ldots (4)$$

whereas[15] :

$$F\left(\frac{b-x\beta}{\sigma}\right) - F\left(\frac{a-x\beta}{\sigma}\right) = \frac{c_1}{\sigma} + \frac{c_2(a - x_i\beta)(b - x_i\beta)}{\sigma^2} \quad \ldots (5)$$

$g(\tau)$: Represents the inverse gama distribution.

From (4) it can be obtained the estimate of the parameters β and σ^2.

Estimation of β and σ^2 when they are unknown:

From the following equation:

$$L = \frac{1}{(2\pi)^{n/2}(\sigma^2\tau)^{n/2}} e^{-\frac{\sum(y_i - x_i\beta)^2}{2\sigma^2\tau}} \left(\frac{c_1}{(\sigma^2)^{1/2}} + \frac{c_2(a-x_i\beta)(b-x_i\beta)}{\sigma^2}\right)^{-n}$$

let Q be as in equation (6):

$$Q = \sum (y_i - x_i\beta)^2 \quad \ldots (6)$$

and :

$$h_i = (a - \beta x_i)(b - \beta x_i) \quad \ldots (7)$$

$$lnL = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{n}{2}\ln\tau - \frac{\sum Q_i}{2\sigma^2\tau} - n\ln\left(\frac{1}{\sigma^2}\left(c_1(\sigma^2)^{1/2} + c_2 h_i\right)\right)$$

$$lnL = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{n}{2}\ln\tau - \frac{\sum Q_i}{2\sigma^2\tau} - \sum_{i=1}^{n}\ln\left(c_1(\sigma^2)^{1/2} + c_2 h_i\right) + nln\sigma^2$$

$$lnL = -\frac{n}{2}\ln(2\pi) + \frac{n}{2}\ln(\sigma^2) - \frac{n}{2}\ln\tau - \frac{\sum Q_i}{2\sigma^2\tau} - \sum_{i=1}^{n}\ln\left(c_1(\sigma^2)^{1/2} + c_2 h_i\right) \quad \dots(8)$$

Let :

$$w(\beta, \sigma^2) = \ln\left(c_1(\sigma^2)^{1/2} + c_2 h_i\right) \quad \dots(9)$$

By using Tyler's series about the point $\sigma^2 = 1$, $\beta = 0$

$$w(\beta, \sigma^2) = w(0,1) + (\beta \quad \sigma^2)' \frac{\partial w(\theta)}{\partial \theta} \quad \dots(10)$$

Whereas:

$$\theta = \beta, \sigma^2$$

$$\therefore \; w(0,1) = \ln(c_1(1) + c_2(a-0)(b-0)) = \ln(c_1 + c_2 ab) \quad \dots(11)$$

$$\frac{\partial w(\theta)}{\partial \theta} = \begin{pmatrix} \frac{\partial w(\beta, \sigma^2}{\partial \beta} \\ \\ \frac{\partial w(\beta, \sigma^2}{\partial \sigma^2} \end{pmatrix} \quad \dots(12)$$

$$\frac{\partial w(\beta, \sigma^2}{\partial \beta} = \frac{-c_2\,(a+b)x_i' + 2\,c_2\,x_i' x_i \beta}{c_1(\sigma^2)^{1/2} + c_2 h_i} \quad \dots(13)$$

Whereas:

$$h_i = (a - \beta x_i)'(b - \beta x_i) = ab - a\beta x_i - b\beta x_i' + x_i'\beta'\beta x_i$$

$$\frac{\partial w(\beta, \sigma^2)}{\partial \sigma^2} = \frac{c_1}{2(\sigma^2)^{1/2}\left(c_1(\sigma^2)^{1/2} + c_2 h_i\right)} \quad \dots(14)$$

$$\frac{\partial w(\theta)}{\partial \theta} = \begin{pmatrix} \frac{-c_2\,(a+b)x_i' + 2\,c_2\,x_i' x_i \beta}{c_1(\sigma^2)^{1/2} + c_2 h_i} \\ \\ \frac{c_1}{2(\sigma^2)^{1/2}\left(c_1(\sigma^2)^{1/2} + c_2 h_i\right)} \end{pmatrix}_{\substack{\beta=0 \\ \sigma^2=1}}$$

$$\frac{\partial w(\theta)}{\partial \theta} = \begin{pmatrix} \frac{-c_2\,(a+b)x_i'}{c_1 + abc_2} \\ \\ \frac{c_1}{2\,(c_1 + abc_2\,)} \end{pmatrix}$$

Tyler's series becomes to $w(\,\theta\,)$ as follows :

$$w(\theta\,) = \ln(c_1 + abc_2\,) + (\,\beta' \quad \sigma^2\,) \begin{pmatrix} \frac{-c_2\,(a+b)x_i'}{c_1 + abc_2} \\ \\ \frac{c_1}{2\,(c_1 + abc_2\,)} \end{pmatrix}_{\substack{\beta=0 \\ \sigma^2}}$$

$$= ln(c_1 + abc_2\,) - \beta' \frac{c_2(a+b)x_i'}{f} + \frac{c_1}{f}\,\sigma^2 \qquad \ldots(15)$$

Whereas:

$$f = c_1 + abc_2 \qquad \ldots(16)$$

$$\therefore w(\,\theta\,) = \ln f - \frac{c_2(a+b)}{f}\,\beta' x_i' + \frac{c_1}{2f}\,\sigma^2 \qquad \ldots(17)$$

Substituting equation (17) into equation (8) we get:

$$lnL\Big|\tau = -\frac{n}{2}\ln(2\pi\,) + \frac{n}{2}\ln(\,\sigma^2\Big) - \frac{n}{2}\ln\tau - \frac{\sum Q_i}{2\sigma^2\tau} - \sum_{i=1}^{n}\left(\ln f - \frac{c_2(a+b)}{f}\,\beta' x_i' + \frac{c_1}{2f}\,\sigma^2\right) \qquad \ldots(18)$$

$$Q_i = (y_i - x_i\beta\,)'(\,y_i - x_i\beta\,) = [\,y_i'y_i - \beta'\,x_i'y_i - \beta' x_i' y_i + \beta' x_i' x_i \beta\,] = [\,y_i'y_i - 2\beta'\,x_i'y_i + \beta' x_i' x_i\beta\,]$$

$$\frac{\partial lnL|\tau}{\partial \beta} = -\frac{\sum_{i=1}^{n}(-2\,x_i'y_i + 2\,x_i'x_i\beta)}{2\,\sigma^2\tau} + \frac{c_2(a+b)}{f}\sum_{i=1}^{n} x_i'$$

$$= -\frac{-2\,\sum_{i=1}^{n}(x_i'y_i - x_i'x_i\,\beta\,)}{2\,\sigma^2\tau} + \frac{c_2(a+b)}{f}\,n\bar{x}' \qquad \ldots(19)$$

Whereas:

$$\sum_{i=1}^{n} x_i' = n\bar{x}' \qquad \ldots(20)$$

$$\frac{\partial lnL|\tau}{\partial \beta} = \frac{\sum_{i=1}^{n}(\,x_i'y_i - x_i'x_i\beta)}{\sigma^2\tau} + \frac{n\,c_2(a+b)}{f}\,\overline{x_i}' \qquad \ldots(21)$$

$$\frac{\partial lnL|\tau}{\partial\sigma^2}=\frac{n}{2\,\sigma^2}+\frac{\sum Q_i}{2\tau(\sigma^2)^2}-\frac{nc_1}{2f} \quad \text{...(22)}$$

$$\therefore\ \frac{\partial lnL}{\partial\beta}=\int_0^\infty\frac{\partial lnL|\tau}{\partial\beta}\,f(\tau)d\tau$$

$$=\int_0^\infty\left(\frac{(\boldsymbol{x'y}-\boldsymbol{x'x\beta})}{\tau\sigma^2}+\frac{n\,c_2(a+b)}{f}\bar{x}'\right)f(\tau)d\tau$$

$$=d_1\frac{(\boldsymbol{x'y}-\boldsymbol{x'x\beta})}{\sigma^2}+d_2\bar{x}_i' \quad \text{...(23)}$$

Whereas:

$$d_1=\int_0^\infty\frac{1}{\tau}\,f(\tau)d\tau \quad \text{...(24)}$$

$$d_2=\int_0^\infty\frac{n\,c_2(a+b)}{f}f(\tau)d\tau \quad \text{...(25)}$$

$$\frac{\partial lnL}{\partial\beta}=0$$

$$\rightarrow\frac{d_1}{\sigma^2}(\boldsymbol{x'y}-\boldsymbol{x'x\beta})+d_2\bar{x}'=0 \quad \text{...(26)}$$

$$-\frac{d_1}{\sigma^2}(\boldsymbol{x'y}-\boldsymbol{x'x\beta})=d_2\bar{\boldsymbol{x}}'$$

Multiply both sides by $\frac{\sigma^2}{d_1}$,we get :

$$-(\boldsymbol{x'y}-\boldsymbol{x'x\beta})=\frac{d_2}{d_1}\sigma^2\bar{x}'$$

$$\boldsymbol{x'x\beta}-\boldsymbol{x'y}=\frac{d_2}{d_1}\sigma^2\bar{x}'$$

$$\boldsymbol{\beta}=(\boldsymbol{x'x})^{-1}x'y+\sigma^2\frac{d_2}{d_1}(\boldsymbol{x'x})^{-1}\ \bar{x}' \quad \text{...(27)}$$

$$\frac{\partial lnL}{\partial\sigma^2}=\int_0^\infty\frac{\partial lnL|\tau}{\partial\sigma^2}\,f(\tau)d\tau$$

$$=\int_0^\infty\left(\frac{n}{2\,\sigma^2}+\frac{\sum Q_i}{2\tau(\sigma^2)^2}-\frac{nc_1}{2f}\right)f(\tau)d\tau \quad \text{...(28)}$$

$$\frac{\partial lnL}{\partial\sigma^2}=0$$

$$\rightarrow\frac{n}{2\,\sigma^2}+\frac{\sum Q_i}{2\tau(\sigma^2)^2}f(\tau)d\tau-\frac{nc_1}{2f}=0$$

$$\frac{n}{2\,\sigma^2}+\frac{\sum Q_i}{2(\sigma^2)^2}d_1-\frac{nc_1}{2f}=0 \quad \text{...(29)}$$

Multiply both sides by 2 $(\sigma^2)^2$:

$$n\,\sigma^2 + d_1\,(y'y - 2\,\beta'x'y + \beta'x'x\beta) - \frac{nc_1}{2f}(\sigma^2)^2 = 0 \qquad \dots(30)$$

$$n\,\sigma^2 + d_1\,y'y - 2d_1\,\beta'x'y + d_1\beta'x'x\beta - \frac{nc_1}{2f}(\sigma^2)^2 = 0$$

By substituting for β, we get:

$$n\sigma^2 + d_1\,\boldsymbol{y'y} - 2\,d_1\left[(\boldsymbol{x'x})^{-1}\,\boldsymbol{x'y} + \sigma^2\frac{d_2}{d_1}(\boldsymbol{x'x})^{-1}\,\bar{x}'\right]\boldsymbol{x'y} + d_1\left[(\boldsymbol{x'x})^{-1}\,\boldsymbol{x'y} + \sigma^2\frac{d_2}{d_1}(\boldsymbol{x'x})^{-1}\,\bar{x}'\right]'(\boldsymbol{x'x})\left[(\boldsymbol{x'x})^{-1}\,\boldsymbol{x'y} + \sigma^2\frac{d_2}{d_1}(\boldsymbol{x'x})^{-1}\,\bar{x}'\right] - \frac{nc_1}{2f}(\sigma^2)^2 = 0$$

$$n\sigma^2 + d_1\,\boldsymbol{y'y} - \left[2\,d_1(\boldsymbol{x'x})^{-1}\,\boldsymbol{x'y} + 2\,d_1\,\sigma^2\frac{d_2}{d_1}(\boldsymbol{x'x})^{-1}\,\bar{\boldsymbol{x}}'\right]\boldsymbol{x'y} + d_1\left[\boldsymbol{y'x}(\boldsymbol{xx'})^{-\mathbf{1}}(\boldsymbol{x'x}) + \sigma^2\frac{d_2}{d_1}\bar{\boldsymbol{x}}(\boldsymbol{xx'})^{-\mathbf{1}}(\boldsymbol{x'x})\right]\left[(\boldsymbol{x'x})^{-1}\,\boldsymbol{x'y} + \sigma^2\frac{d_2}{d_1}(\boldsymbol{x'x})^{-1}\,\bar{\boldsymbol{x}}'\right] - \frac{nc_1}{2f}(\sigma^2)^2 = 0$$

$$n\sigma^2 + d_1\,\boldsymbol{y'y} - 2d_1\boldsymbol{y'x}(\boldsymbol{x'x})^{-\mathbf{1}}\boldsymbol{x'y} - 2\sigma^2 d_2\bar{\boldsymbol{x}}(\boldsymbol{x'x})^{-1}\,\boldsymbol{x'y} + \left[d_1\boldsymbol{y'x} + \frac{d_1\sigma^2 d_2}{d_1}\bar{\boldsymbol{x}}\right]\left[(\boldsymbol{x'x})^{-1}\,\boldsymbol{x'y} + \sigma^2\frac{d_2}{d_1}(\boldsymbol{x'x})^{-1}\,\bar{x}'\right] - \frac{nc_1}{2f}(\sigma^2)^2 = 0$$

$$n\sigma^2 + d_1\,\boldsymbol{y'y} - 2d_1\boldsymbol{y'x}(\boldsymbol{x'x})^{-\mathbf{1}}\boldsymbol{x'y} - 2\sigma^2 d_2\bar{\boldsymbol{x}}(xx')^{-1}\,\boldsymbol{x'y} + d_1\boldsymbol{y'x}(\boldsymbol{x'x})^{-1}\,\boldsymbol{x'y} + d_1\boldsymbol{y'x}\,\sigma^2\frac{d_2}{d_1}(\boldsymbol{x'x})^{-1}\,\bar{\boldsymbol{x}}' + \sigma^2 d_2\bar{\boldsymbol{x}}(\boldsymbol{x'x})^{-1}\,\boldsymbol{x'y} + \sigma^2 d_2\bar{\boldsymbol{x}}\,\sigma^2\frac{d_2}{d_1}(\boldsymbol{x'x})^{-1}\,\bar{x}' - \frac{nc_1}{2f}(\sigma^2)^2 = 0$$

$$n\sigma^2 + d_1\,\boldsymbol{y'y} - d_1\boldsymbol{y'x}(\boldsymbol{x'x})^{-\mathbf{1}}\boldsymbol{x'y} - \sigma^2 d_2\bar{\boldsymbol{x}}(\boldsymbol{x'x})^{-1}\,\boldsymbol{x'y} + \sigma^2 d_2\boldsymbol{y'x}(\boldsymbol{x'x})^{-1}\bar{\boldsymbol{x}}' + \frac{(\sigma^2)^2(d_2)^2}{d_1}\bar{\boldsymbol{x}}(\boldsymbol{x'x})^{-1}\bar{\boldsymbol{x}}' - \frac{nc_1}{2f}(\sigma^2)^2 = 0$$

$$n\sigma^2 + d_1\,\boldsymbol{y'y} - d_1\boldsymbol{y'x}(\boldsymbol{x'x})^{-\mathbf{1}}\boldsymbol{x'y} - \sigma^2 d_2\bar{\boldsymbol{x}}(\boldsymbol{x'x})^{-1}\,\boldsymbol{x'y} + \frac{(\sigma^2)^2(d_2)^2}{d_1}\bar{\boldsymbol{x}}(\boldsymbol{x'x})^{-1}\bar{\boldsymbol{x}}' + (\sigma^2 d_2\bar{\boldsymbol{x}}(\boldsymbol{x'x})^{-1}\boldsymbol{x'}y)' - \frac{nc_1}{2f}(\sigma^2)^2 = 0$$

$$n\sigma^2 + d_1\boldsymbol{y'}\,[I - \boldsymbol{x}(\boldsymbol{x'x})^{-1}\boldsymbol{x'}]\boldsymbol{y} + (\sigma^2)^2\,\frac{(d_2)^2}{d_1}\,\bar{\boldsymbol{x}}(\boldsymbol{x'x})^{-1}\bar{\boldsymbol{x}}' - \frac{nc_1}{2f}(\sigma^2)^2 = 0$$

$$\left(\frac{(d_2)^2}{d_1}\,\bar{\boldsymbol{x}}(\boldsymbol{x'x})^{-1}\bar{\boldsymbol{x}}' - \frac{nc_1}{2f}\right)(\sigma^2)^2 + n\sigma^2 + d_1\,\boldsymbol{y'}[I - \boldsymbol{x}(\boldsymbol{x'x})^{-1}\boldsymbol{x'}]\,\boldsymbol{y} = 0$$

$$\rightarrow\ a_1(\sigma^2)^2 + n\sigma^2 + d_1 els'els = 0 \qquad \dots(31)$$

Where:

$$els = y - x\hat{\beta}_{ols} = y - \boldsymbol{x}(\boldsymbol{x'x})^{-1}\boldsymbol{x'y} \qquad \dots(32)$$

$$\rightarrow\ a_1(\sigma^2)^2 + n\sigma^2 + a_2 = 0 \qquad \dots(33)$$

Whereas:

$$a_2 = d_1 els'els$$

$$\rightarrow \ -a_1(\sigma^2)^2 - n\sigma^2 - a_2 = 0 \qquad \dots(34)$$

$$-a_1(\sigma^2)^2 - n\sigma^2 = a_2$$

By extracting ($-a_1$) as a common factor from the left side, we get:

$$-a_1\left((\sigma^2)^2 + \frac{n}{a_1}\sigma^2\right) = a_2$$

As for :

$$-a_1 = a_2 \qquad \dots(35)$$

This value is neglected because it is not needed.

or:

$$(\sigma^2)^2 + \frac{n}{a_1}\sigma^2 = a_2 \qquad \dots(36)$$

$$\rightarrow (\sigma^2)^2 + \frac{n}{a_1}\sigma^2 - a_2 = 0$$

Using the constitution method and after ignoring the negative sign of the root, we get σ^2:

$$\sigma^2 = \frac{-\frac{n}{a_1} + \sqrt{\left(\frac{n}{a_1}\right)^2 + 4a_2}}{2\,(1)} \qquad \dots(37)$$

whereas:

$$A = 1\,; B = \frac{n}{a_1}\,; C = -a_2$$

Simulation :

Simulation is defined as imitation of a complex real reality by arranging a set of programming and mathematical steps to build a model that represents the imposed reality, researchers use it when they face some problems in obtaining real data. There are simulation methods, including computer simulation, where a program is written for the thing to be examined that matches its specifications in reality, and then this program is placed in software conditions that resemble reality.

The Monte Carlo simulation depends on random numbers, and a random number is a number whose probability of occurrence is equal to the probability of any other random number from a set of random numbers distributed as a standard Uniform distribution [0,1], this is because the random numbers generated by electronic machines fall within the period [0,1].

The sample is generated in several ways, such as the inverse function method, where this function depends on generating data by finding the cumulative function and then finding its inverse, in short, to build a linear model, 50 observations were generated, then we first assumed the initial values for the parameters of the regression model where [1.5,4].We have assumed that the independent variable is normally distributed with a mean equal to 3 and a standard deviation equal to 0.5.The random error is distributed as a truncated t-distribution, since the truncation values that have been relied upon are [-2, 1], and the degrees of freedom are 4.Note that all the values mentioned are randomly imposed values.

Table 1: Parameter and error values for the methods used

Method	β	MSE
MLE	2.5112	0.0443
OLS	6.667	0.1423

Conclusion:

In this paper, we have used a maximum likelihood to estimate the parameters of the t-truncated linear regression by depending on an approximation to the cumulative function. The method which is used leads to estimate the parameters in a simple way without using the statistic tables.

References

[1] Venables, William N and Smith, David M and R. Development Core Team and others. (2009). An introduction to R.

[2] Al-Athari, F. M. and Al-Sarraf, Z.J. (2007). Maximum Likelihood Estimation of Truncated Normal Regression Model.

[3] Tokmachev, M. (2018). Modeling of truncated probability distributions. In IOP Conference Series: Materials Science and Engineering 441 012056. IOP Publishing.

[4] Arismendi, J. C. (2013). Multivariate truncated moments. Journal of Multivariate Analysis 117 41–75.

[5] Marsh, T. L. and Mittelhammer, R. C. (2000). Truncated Regression in Empirical Estimation Technical Report.

[6] Karlis,D. and Xekalaki,E. (2005). Mixed poisson distributions. International Statistical Review/Revue International de Statistique 35–58.

[7] Shaw, W.(2006). New Methods for Managing “Student’s” T Distribution. Preprint King’s College. Cite seer.

[8] Mahbobi, M. and Tiemann, T. K. (2015). Making Estimates. Introductory Business Statistics with Interactive Spreadsheets-1st Canadian Edition.

[9] Arellano-Valle, R. B., Castro, L. M., Gonz'alea-Fari'as, G. and Munoz-Gajardo, K. A. (2012). Student-t censored regression model: properties and inference. Statistical Methods & Applications 21 453–473.

[10] Breen, R. et al. (1996). Regression models: Censored, sample selected, or truncated data 111. Sage.

[11] Khan, K. (2010). The Evaluation of Well-known Effort Estimation Models based on Predictive Accuracy Indicators.

[12] Del Castilo, J. (1994). The singly truncated normal distribution: a non-steep exponential family. Annals of the Institute of Statistical Mathematics 46 57–66.

[13] Cohen JR, A. C. (1950). Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples. The Annals of mathematical statistics 557–569.

[14] Kim, H.J. (2008). Moments of truncated Student-t distribution. Journal of the Korean Statistical Society 37 81–87.

[15] Shallawi, A.N (2022). Approximation of cumulative function for t-distribution.Unpuplished research. Northern Technical University