Communication, Society and Media
ISSN 2576-5388 (Print) ISSN 2576-5396 (Online)
Vol. 2, No. 1, 2019
www.scholink.org/ojs/index.php/csm

Original Paper

Gender Wise Distribution of Income Using L-Moments Method

Muhammad Alam[1], Saeed Ullah Jan[2*] & Alam Zeb[1]

[1] Faculty of Basic & Applied Sciences, Department of Statistics, International Islamic University, Islamabad, Pakistan

[2] Department of Computer Science & IT, University of Malakand, Chakdara, Pakistan

[*] Saeed Ullah Jan, Department of Computer Science & IT, University of Malakand, Chakdara, Pakistan

Received: January 31, 2018   Accepted: February 9, 2019   Online Published: February 16, 2019

doi:10.22158/csm.v2n1p29   URL: http://dx.doi.org/10.22158/csm.v2n1p29

Abstract

The main purpose of this work is to explore the income distribution of both male and female in Pakistan over the period of 2010-2011. For this purpose, the lognormal distribution with known parameters is used as a model and its unknown parameters are estimated by three methods that are likelihood, moments and L-moments. The results show that citizens of Pakistan are not equal in income and the probability plot suggested that the income of the male is greater than that of a female in Pakistan. Moreover, for small sample size, the best method of parameters estimation is the L-moments, while, for large sample size the best method is a maximum likelihood. Findings of the study suggest that suitable policy is required to maintain equality in income distribution in the country. It will consequently reduce the gap among rich and poor and will certainly improve social welfare.

Keywords

distribution, lognormal, likelihood, l-moments, ratio, population

1. Introduction

The importance of income throughout the world is undeniable. It is a backbone for a country, especially in Pakistan is important as breathing air, without air there is no hope for life. Without having a reasonable source of earning there cannot be even a single ray of hope for survival. The lognormal distribution (Note 1) is used in several fields of life for estimation of parameters, e.g., they are used in Statistics, Geology, Medical Science, Environmental Science, Technology, Ecology, Social science, and income. Incomes within countries generally adopt a skewed distribution with a long heavy tail. Lognormal distribution has been found best fit on the average for income distribution. This can also prove through the goodness of fit test such as Chi-square test, Anderson Darling test (Albreht & Klazinga, 2009).

The lognormal distribution is a continuous probability distribution. The probability density function of the three parametric lognormal distributions is given as:

$$f(x;\ \mu,\sigma,\gamma) = 1/((x-\gamma)\sigma\sqrt{2\pi})e^{\{\frac{-[\ln(x-\gamma)-\mu]^2}{2\sigma^2}\}} \quad (1)$$

Where $0 < x < \infty, 0 \leq \gamma < x, -\infty < \mu < \infty, \sigma > 0$

Whereas parameters,

$\mu =$ Scale,

$\sigma =$ Shape and

$\gamma =$ Location or Threshold

The L-moments (Note 2) can be used for characterization and explanation of theoretical probability distributions, narration, explanation of experimental data samples, assessment of different parameters, assessing of probability distributions and hypothesis testing. The major benefit of L-moments over traditional moment is that the L-moments illustrate a broad range of distributions, suffered less from the effect of sampling variability and are more powerful than when there are outliers in the data by (Bílková, 2008). It is used for extreme events such as rainfall, flood, earth quick, droughts, heat waves, snowfall, famine, and income distribution. The L-moments technique gives the correct result in the estimation of parameters as compared to the maximum likelihood, method of moments (Bílková, 2011).

Let suppose X is a vector of observations in a sample which can be described by the model $f(x,\theta)$, where θ is the parameter value. The given model $f(x,\theta)$, will give us the probabilities of the various values of X. When we define the maximum likelihood then we have to define the likelihood function first (Bílková, 2012). The likelihood function for a given value of X is $L(\theta)$ as a function of θ. The MLE is a reasonable method of estimation because it locates that value of θ for which the observed data are most probable. The MLE enjoys some good properties in large sample size such as normality, efficiency, consistency, and un-biasedness. Suppose a random sample $X = (x_1, x_2, \dots \dots x_n)$ are drawn from a distribution with probability density function $f(x,\theta)$, $\theta \epsilon \Theta$ where Θ is parameter space (Bílková, 2012). The likelihood function is:

$$L(\theta) = \prod_{i=1}^{n} f(xi,\theta) \quad (2)$$

Let suppose we have random samples $x_1, x_2, \dots . x_n$ from a distribution having probability density function is $f(x,\underline{\theta})$, and $\underline{\theta} = (\theta 1, \dots, \theta n)$, $\underline{\theta} \epsilon \Theta$ which are considered as unknown vector parameters. We then define the r_{th} population moment at the origin "0" as, for continuous distribution:

$$\acute{\mu} r = E(x^r) = \int_{-\infty}^{\infty} x^r f(x,\underline{\theta}) dx \quad (3)$$

The advantages of the moment's method are that it is an old and simple technique used for the estimation of parameters and a very simple method for finding the estimators of parameters. Its estimators are consistent and best method in case of point estimation parameters. It is an easy method to find an estimator and also in the case when the other methods fail to find. It is used in case of large sample size. Its first moment equal to mean, second to variance, third to skewness and forth one is equal to kurtosis (Arltová, 2013).

The drawbacks of the moment's method are that this method is not available for every distribution. Moments method estimators are not necessary to be sufficient.

The Hosking et al. (2005) is used for the data of Household Income and Expenditure Survey (HIES) and Pakistan Integrated Household Survey (PIHS) w.e.f 1963 to 2002 for both rural, urban citizen and was decided that the main goal of economic growth is to improve the living standard of ordinary men, but only economic growth is not sufficient for the standard of living of a common men, the improvement in the income distribution is also playing a vital role in the betterment of a layman life.

But unluckily the ratio of poverty becomes an increase in the 1990s and the economist focuses to find the ratio of the poor population instead of improving the measurement of the income distribution.

In Pakistan, the income inequalities have also been estimated for both methods, i.e., Pakistan Integrated Household Survey (PIHS) and Household Income and Expenditure Survey (HIES).

The wage data of the Czech Republic from 2004-2005 apply different methods of parameters for the aforesaid estimation such as quantile method, moment method, and maximum likelihood method. By applying lognormal distribution one can easily decide the results of the lognormal distribution (Kemal, 2009).

Rehman et al. (2008) analyzed that the distribution of income, growth and its development in Pakistan is because of income inequality and create financial hurdles to all them respectively. In the early periods, several attempts were made to make a relation between income inequality and economic growth, but there is no proper mechanism was developed. Rehman et al. (2008) concluded that there is a reverse correlation between income and economic growth.

The data on household income per capita in the Czech Republic from 1992-2008 decided that the method of L-moments for estimation of parameters gives us the correct result as compared to other methodologies like maximum likelihood, moment subject to the condition that the data should be separate, for grouped data, these methods will give similar results of Arltová (2013).

The Census data record from Czech Statistical Office, income data from statistical surveys and wage data taken from Czech's official website shows that the L-moments technique gives the correct result in the estimation of parameters as compared to the maximum likelihood, method of moments and quantile method for individual data subject to the condition that data should not be in group because the in grouped data all the four methods give a similar result like that of Bílková (2008).

Also, the income data of the Czech Republic discussed by Steinführer et al. (2010) suggested that method of L-moments for estimation of parameters provides precise consequence as compared to the other methods of estimation, such as the method of moment, a method of maximum likelihood in case of ungrouped data; for grouped data, these methods give the same results. The household per capita of the Czech Republic proposed by Kemal (2008) that for income and wage distribution the L-moments is the best method for estimation of parameters as compared with the method of moment, quantile method, and method of maximum likelihood.

Furthermore, the data of the Czech Republic's household income from 1992-2008 has also been used for the lognormal distribution. Also, apply four methods of parameters estimation, which includes L-moments, quantile, maximum likelihood, and moment method. They decided that the method of L-moments gives more accurate results than other methods by Albreht (2008).

There are two types of data; the first one is the Czech Republic per capita household income in years 1992, 1996 and 2002, and the second is the statistical survey micro-census of years 2005, 2006, 2007 and 2008. They used the three parameters of the lognormal distribution. And conclude that the accuracy of the method of L-moments is better than the other methods of estimation, that are maximum likelihood and the quantile method, a method of the moment for individual data. But for the grouped data the method of L-moments, a method of maximum likelihood, a method of moments, and the quantile method gives similar consequences by Langhamrová and Bílková (2008).

2. Materials & Methods

The Income data are taken from Pakistan Social and Living Standard Measurement Survey (PSLSMS) (Note 3) of 2010-2011 of the Pakistan Bureau of Statistics (PBS) Islamabad. The calculation of a national income is important. From the national income, we can find the performance of a country during that year. From the national income, we can also find out whether the financial system is increasing or decreasing. The national income depends upon the wages, profits, rents, interest, and business. Due to these factors, we can improve our national income and from national income, we improve the per capita income and also the standard of living. The income distribution also used to find the proportion of low, middle and high incomes workers in a country.

The main goal of economic growth is to improve the standard of life, but only economic growth is not sufficient for the standard of living of peoples; the improvement in the income distribution is also playing an important role in the betterment of the human life. But unluckily the ratio of poverty becomes an increase in the 1990s and the economist focuses to find the ratio of the poor population instead of improving the measurement of the income distribution. Income inequality is a main financial problem to all the word. In the early periods, several attempts have been made to make out a relation between the income inequality and the economic growth, but there is no conclusion find between the income inequality and the economic growth but conclude that there is a reverse correlation between income and economic growth.

The conventional moments are used to estimate the parameters of a distribution but the conventional moments for small sample size are not always convenient, moreover, we can also specify a distribution by its L-moments even if its conventional moments do not exist. The sample convenient moments are used to summarize the mean, variance, skewness and kurtosis, similarly, the sample L-moments are used to summarize the location, scale, skewness, and kurtosis for a specified distribution.

Let Y be a random variable with cumulative distribution function $F(Y)$and quantile function $Y(F)$ and let $Y_{1:n} \le Y_{2:n} \le \cdots \le Y_{n:n}$ be the order statistics of a random sample of size "n" taken from the distribution of Y. The L-moments of Y can be defined

$$\lambda_1 = E(Y_{1:1})$$
$$\lambda_2 = \frac{1}{2}E(Y_{2:2} - Y_{1:2})$$
$$\lambda_3 = \frac{1}{3}E(Y_{3:3} - 2Y_{2:3} + Y_{1:3})$$
$$\lambda_4 = \frac{1}{4}E(Y_{4:4} - 3Y_{3:4} + 3Y_{2:4} - Y_{1:4})$$

The general term can be written as

$$\lambda_r = r^{-1}\sum_{i=0}^{r-1}(-1)^i\binom{r-1}{i}E(Y_{r-i:r}), \qquad r = 1,2,\ldots$$

Now the expectation of an order statistic may be defined as

$$E(Y_{r:n}) = \frac{n!}{(r-1)!(n-r)!}\int_0^1 y(v)v^{r-1}(1-v)^{n-r}dv \qquad (2)$$

The unique specification of L-moments is that it specifying the distribution if some of its conventional moments do not exist.

The L-moments ratios (Note 4) can be defined as:

$$\tau_r = \frac{\lambda_r}{\lambda_2},\ r = 3,4\ldots..$$

The L-CV is

$$\tau = \frac{\lambda_2}{\lambda_1},\quad 0 < \tau < 1$$

The probability weighted moments is defined as

$$M_{p,r,s} = E[Y^p\{F(Y)\}^r\{1 - F(Y)\}^s] \qquad (3)$$

The L-moments of a probability distribution are given as

$$\lambda_1 = E(Y_{1:1})$$
$$\lambda_2 = \frac{1}{2}E(Y_{2:2} - Y_{1:2})$$
$$\lambda_3 = \frac{1}{3}E(Y_{3:3} - 2Y_{2:3} + Y_{1:3})$$
$$\lambda_4 = \frac{1}{4}E(Y_{4:4} - 3Y_{3:4} + 3Y_{2:4} - Y_{1:4})$$

And for general

$$\lambda_r = r^{-1} \sum_{i=0}^{r-1} (-1)^i \binom{r-1}{i} E(Y_{r-i:r}) \qquad (4)$$

The sample L-moments may also be defined as

$$l_1 = b_0$$

$$l_2 = 2b_1 - b_0$$

$$l_3 = 6b_2 - 6b_1 + b_0$$

$$l_4 = 20b_3 - 30b_2 + 12b_1 - b_0$$

And for the general term

$$l_{r+1} = \sum_{k=0}^{r} P^{*}_{r,k}\, b_k \quad \text{Where}$$

$$r = 0,1,\ldots\ldots,n-1 \qquad (5)$$

3. Results and Discussion

In Table 1, l_1 is the mean of the data, the value of t is 0.4047 that indicates that there is inequality in the male income data. t_3 is the L-skewness and its value is 0.4399 which is positive so the male income data distribution will be positively skewed, it indicates that most of the peoples income will be on the left side of the mean and the extreme income means the high income will be on the right side of the mean, Q_1, Q_2, Q_3 are the three quartiles of the male income data as shown in the table given below.

Table 1. L-Moments Ratios and Quartiles for Male Income in Pakistan

l_1	t	t_3	t_4	Q_1	Q_2	Q_3
131900	0.4047008	0.4399626	0.3190783	62400	96000	151200
79352.53	0.5651828	0.5235219	0.3122305	24000	36000	96000

Table 2. Parameters Estimation by Different Methods of Three Parameters Lognormal Distribution for Male Income in Pakistan

Methods of estimation	μ	σ	γ
L-moments	11.14	0.94	24366.17
Moments	11	1	31449.8
Maximum likelihood	11.5	0.7	-2001

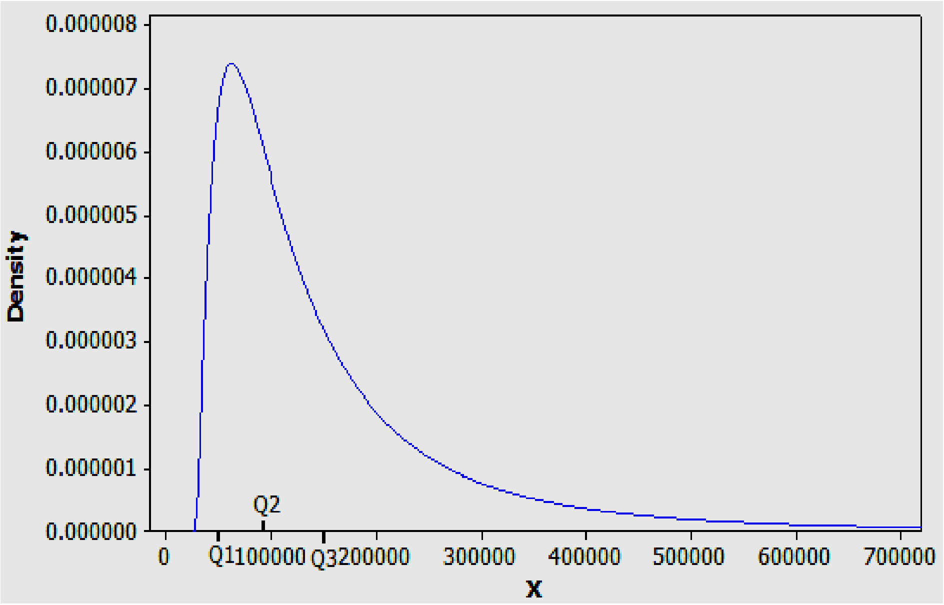

Figure 1. Probability Distribution Plot of Male Income in Pakistan

From the above figure, we conclude that 25% of male income is less than 62000, 50% of peoples have income below than 96000. And 75% of the total population income is less 150000, and the last 25% of the population that have income above than 150000.

Table 3. L-Moments Ratios and Quartiles for Female Income in Pakistan

l_1	t	t_3	t_4	Q_1	Q_2	Q_3
79352.53	0.5651828	0.5235219	0.3122305	24000	36000	96000
131900	0.4047008	0.4399626	0.3190783	62400	96000	151200

In Table 2, the average income of female is less than the male income; the value of t is 0.5652 that indicates that there is inequality in the female income data. As we compare the variation of male income with female income then there is a large variation in female income as compared with male income, the L-CV of male income is less than the female, and also the skewness of female income is greater than male income, t_3 is the L-skewness and its value is 0.5235 which is positive so the female income data distribution will be positively skewed. It indicates that most of the female income will be on the left side of the mean and the extreme income means the high income will be on the right side of the mean, Q_1, Q_2, Q_3 are the three quartiles of the female income data.

Table 4. Parameters Estimation by Different Methods of Three Parameters Lognormal Distribution for Female Income in Pakistan

Methods of estimation	μ	σ	γ
L-moments	10.58	1.15	2663.74
Moments	11.4	0.8	-38872.5
Maximum likelihood	10.7	1.1	605

From the below female income plot shown in Figure 2, we see that the first 25% of the population has income below than 24000 and 50% of the population income less than 36000. And 75% of the total female population income is less than 96000 and the last 25% of the female population that have income greater than 96000. As we see that the male income is greater than female income in Pakistan.

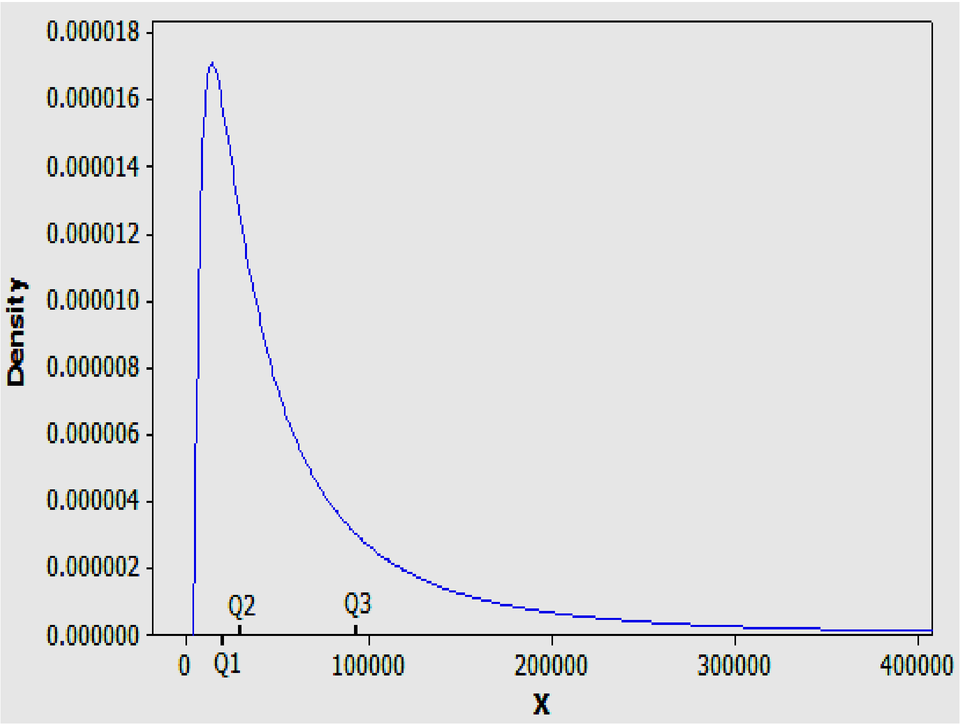

Figure 2. Probability Distribution Plot of Female Income in Pakistan

Table 5. Absolute Bias and RMSE for Different Methods of Estimation of Parameters for Sample Size n=20

Methods of estimation	μ	σ	γ	$\lvert Bias \rvert$	**RMSE**
L-moments	11.59	0.76	-8390	35892	162128
Moments	12.40	0.40	-127937	151951	231421
Maximum likelihood	11.64	0.72	-11172	69622	211036

When n=20, small sample size the best method of estimation is L-moments because its absolute Bias and RMSE are less than the Moments method and Maximum likelihood method.

In last we compare the parameters of lognormal distribution obtained by L-moments method, Moments method, and Maximum likelihood method, with respect to absolute Bias and RMSE. We take different sample sizes from the data and also take the random numbers of the same size that we selected from the real income data and then find the absolute Bias and RMSE, so the best method will be that whose absolute Bias and RMSE are small. So for a small sample size the best the L-moments method, and for large sample size the best method is the Maximum Likelihood.

Table 6. Absolute Bias and RMSE for Different Methods of Estimation of Parameters for Sample Size n=30

Methods of estimation	μ	σ	γ	$\lvert Bias \rvert$	**RMSE**
L-moments	11.25	1.10	15673	39066	233841
Moments	12.06	0.73	-68074	74522	293810
Maximum likelihood	11.22	1.18	17360	83800	479023

When n=30 then, in this case, the L-moments method for estimation of parameters is best because its absolute Bias and RMSE are less than the other methods.

Similarly, for n=50 the best method of parameters estimation is the maximum likelihood because its absolute Bias and RMSE are less than the L-moments method and moments method as shown in the below table.

Table 7. Absolute Bias and RMSE for Different Methods of Estimation of Parameters for Sample Size n=50

Methods of estimation	μ	σ	γ	$\|Bias\|$	**RMSE**
L-moments	11.59	0.71	-8310	44740	182374
Moments	12.34	0.39	-116996	129186	192291
Maximum likelihood	11.48	0.78	980	12203	152556

Table 8. Absolute Bias and RMSE for Different Methods of Estimation of Parameters for Sample Size n=100

Methods of estimation	μ	σ	γ	$\|Bias\|$	**RMSE**
L-moments	11.51	0.70	380	13826	145428
Moments	11.18	0.88	23264	27413	157953
Maximum likelihood	11.62	0.63	-8860	3176	119216

But when n=100 means when sample size increases the best method of estimation of parameters is maximum likelihood method because its absolute Bias and RMSE are less than other methods that are L-moments and moments method.

Table 9. Absolute Bias and RMSE for Different Methods of Estimation of Parameters for Sample Size n=250

Methods of estimation	μ	σ	γ	$\|Bias\|$	**RMSE**
L-moments	11.11	0.91	20654	17299	233256
Moments	11.10	0.98	14758	3962	199615
Maximum likelihood	11.52	0.65	-6640	2095	158376

And when n=250 or above than 250 means by increases sample size the best method of estimation of parameters is maximum likelihood method than L-moments method and moments method because its absolute Bias and RMSE are less.

4. Conclusion

Incomes within countries generally adopt a skewed distribution with a long heavy tail; lognormal distribution has been found best fit on the average for income distribution in Pakistan. We estimate the parameters of lognormal distribution by different methods of estimation that are the L-moments method, moments method and maximum likelihood method. We find the parameters of these three methods for all male and female in the whole Pakistan. We also find the probability distribution plot for male and female income data. The citizens of Pakistan are not equal in income, a few groups of peoples are very rich and some are very poor. Finally, we compare the parameters of lognormal distribution obtained by L-moments method, moments method, and maximum likelihood method, with respect to absolute Bias and RMSE for different sample size. So the best method of parameters estimation will be that whose absolute Bias and RMSE are less. So in the case of small sample size, the best method of parameters estimation is the L-moments, and for large sample size the maximum likelihood method. The moments method is also used in case of large sample size but its estimators are not efficient and not necessary to be sufficient.

References

Albreht, T., & Klazinga, N. (2009). Privatisation of health care in Slovenia in the period 1992-2008. *Health policy*, *90*(2-3), 262-269. https://doi.org/10.1016/j.healthpol.2008.10.007

Arltová, M., Langhamrová, J., & Langhamrová, J. (2013). Development of life expectancy in the Czech Republic in years 1920-2010 with an outlook to 2050. *Prague Economic Papers*, *22*(1), 125-143. https://doi.org/10.18267/j.pep.444

Bílková, D. (2008). Application of lognormal curves in the modeling of wage distributions. *Journal of Applied Mathematics*, *1*(2), 341-352.

Bílková, D. (2011). L-moments and their use in modeling the distribution of income and wage. *ISI held on*, 21-26.

Bílková, D., & Malá, I. (2012). Application of the L-moment method when modeling the income distribution in the Czech Republic. *Austrian Journal of Statistics*, *41*(2), 125-132. https://doi.org/10.17713/ajs.v41i2.180

Hosking, J. R. M., & Wallis, J. R. (2005). *Regional frequency analysis: An approach based on L-moments*. Cambridge University Press.

Kemal, A. R. (2006). Income Inequalities in Pakistan and a strategy to reduce income inequalities. *Background Paper for PRSP-II, PRSP Secretariat, Government of Pakistan*.

Langhamrová, J., & Bílková, D. (2011). Analysis of the distribution of income in recent years in the Czech Republic by region. *International Days of Statistics and Economics at VŠE, Prague*, 2011-2023.

Rehman, H. U., Khan, S., & Ahmed, I. (2008). Income distribution, growth, and financial development: A cross countries analysis. *Pakistan Economic and Social Review*, 1-16.

Steinführer, A., Bierzynski, A., Großmann, K., Haase, A., Kabisch, S., & Klusácek, P. (2010). Population decline in Polish and Czech cities during post-socialism? Looking behind the official statistics. *Urban Studies*, *47*(11), 2325-2346. https://doi.org/10.1177/0042098009360224

Notes

Note 1. The Lognormal distribution is used to claim that the distribution is positively skewed or not. The distribution of income is also positively skewed. So we use the lognormal distribution for income data.

Note 2. L-moments are the outlook of statistics of the order of certain linear combinations and can be defined for every random variable whose mean exists. And can be used for characterization and explanation of theoretical probability distributions

Note 3. The data are taken from Pakistan Social and Living Standard Measurement Survey (PSLSMS) of 2010-2011.

Note 4. The L-moment ratio can be obtained by higher order L-moment divided by a measure of dispersion that as $t_r=l_r/l_2$. The t_3 used for skewness and t_4 used for kurtosis, l_1 denoted for the mean of the data, t is for L-CV which is similar to the coefficient of variation and the value of t will $0 \le t < 1$.