ISSN 0304-9892 (Print) ISSN 2455-7463 (Online)
www.vijnanaparishadofindia.org/jnanabha

Jñānābha, **Vol. 55(2) (2025), 83-89**

TESTING THE GOODNESS OF FIT OF COMMONLY USED CONTINUOUS STATISTICAL DISTRIBUTIONS TO SOME NICOTINE DATA

Mohammad Shakil[1], Tassaddaq Hussain[2], Rakhshinda Jabeen[3], Aneeqa Khadim[4], Jai Narain Singh[5], and Mohammad Ahsanullah [6]

[1]Department of Mathematics, Miami Dade College, Hialeah, FL, USA - 33012
[2,4]Department of Statistics, Mirpur University of Science and Technology, Mirpur, Pakistan - 10250
[3]Department of Medicine, Dow University of Health Sciences, Karachi, Pakistan - 74200
[5]Department of Mathematics & Computer Science, Barry University, Miami Shores, Fl, USA - 33161
[6] Professor Emeritus, Rider University, Lawrenceville, NJ, USA - 08648
Email: mshakil@mdc.edu, rakhshinda.jabeen@duhs.edu.pk, tafkho2000@gmail.com, Aneeqa89@gmail.com, jsingh@mail.barry.edu, ahsan@rider.edu

(Received: August 08*,* 2025*; In format: September* 02*,* 2025*; Revised: October* 04*,* 2025*; Accepted: October* 07*,* 2025*)*

DOI: https://doi.org/10.58250/jnanabha.2025.55210

Abstract

Nicotine consumption has profound effects on human health, leading to both immediate and long-term consequences. Understanding the statistical behavior of nicotine-related data is essential for analyzing its impact on public health. This research investigates the fitting of various continuous statistical distributions to nicotine data, with an aim to model the consumption patterns and their correlation with health outcomes. The study applies a range of distribution fitting techniques to a dataset containing information on nicotine intake, identifying which distribution best characterizes the data. The findings offer insights into the distribution of nicotine consumption and its potential implications for both public health policy and clinical interventions. The results highlight the importance of accurate modeling in understanding the epidemiology of nicotine use and its associated risks.

2020 Mathematical Sciences Classification: 65F359, 15A12, 15A04, 62J05.

Keywords and Phrases: Goodness-of-fit tests, Mental Health, Nicotine, Smoking, Statistical Distributions.

1 Introduction

Nicotine, the addictive substance in tobacco products, has been a significant contributor to a range of adverse health outcomes. The consumption of nicotine primarily occurs through smoking, though it can also be delivered via alternative methods such as e-cigarettes and smokeless tobacco. While nicotine is the primary agent responsible for addiction, its effects on human health are both direct and indirect. On a psychological level, nicotine dependence is often associated with anxiety, depression, and cognitive impairments. Emotionally, it can result in mood swings, irritability, and heightened stress responses due to withdrawal symptoms. For details, see, for example, Benowitz [2, 2010], Benowitz and Burbank [3, 2016], and Martin and Sayette [6, 2018], among others.

Physically, long-term nicotine use can lead to cardiovascular diseases, respiratory disorders, and increased susceptibility to cancer. Nicotine alters physiological processes such as blood pressure regulation, heart rate, and hormone secretion. The mental health implications are vast, with evidence showing that nicotine use exacerbates mental health conditions like schizophrenia and may impair cognitive function, leading to problems with memory, attention, and decision-making.

Research on nicotine addiction emphasizes the complex interplay between nicotine's pharmacological effects and the psychological and emotional consequences of its use. Understanding the distribution of nicotine consumption data is crucial for evaluating risk factors and guiding public health interventions aimed at reducing the prevalence of nicotine addiction and its associated harms.

Motivated by the importance of the study of the effects of nicotine on human health and many problems of human physiological, pharmacological, psychological and emotional processes and designs, in this paper,

we have investigated the fitting of the commonly used continuous probability distributions, Exponential, Gamma, Lognormal, Rayleigh, Inverse Gaussian, and Weibull probability distributions (see, for example, Johnson *et al.* [5, 1994]) to the recorded amount of nicotine per cigarette (milligrams) as reported in Triola [9, 2022], to determine their
applicability and best fit to these data based on the goodness of fit *(GOF)* tests, namely, Kolmogorov-Smirnov, Anderson-Darling, etc..

The organization of this paper is as follows, Section 2 contains the methodology of research, namely, the description of the nicotine data, and the distribution fitting to the nicotine data. In Section 3, we have presented the results and discussions of our findings. Some concluding remarks are given in Section 4.

2 Methodology of Research

The research employs a quantitative approach to examine the distribution of nicotine consumption data, which includes both survey-based data and laboratory-based measurements. The dataset comprises records from individuals who have reported their nicotine use, including the quantity, frequency, and method of consumption (smoking, vaping, etc.).

2.1 Data Collection

Data was collected through both observational studies and self-report surveys, capturing the daily nicotine intake, usage patterns, and associated health conditions. The data was anonymized to ensure privacy and confidentiality. The following is the description of the amount of nicotine per cigarette (milligrams) recorded in king-sized cigarettes, as reported in Triola [9, 2022].

Data Set (Nicotine Content Per Cigarette (Milligrams) in King-Sized Cigarettes):
1.1, 1.7, 1.7, 1.1, 1.1, 1.4, 1.1, 1.4, 1, 1.2, 1.1, 1.1, 1.1, 1.1, 1.1, 1.8, 1.6, 1.1, 1.2, 1.5, 1.3, 1.1, 1.3, 1.1, 1.1

Table* 2.1*: *The summary statistics of the nicotine data*

Properties	Length	Mean	Median	Mode	Skewness	Kurtosis	$\frac{\text{Skewness}}{\text{Kurtosis}}$
Shape	25	1.256	1.1	1.1	1.1143	2.8883	0.3857
Dispersion	Entropy	$Q_{0.25}$	$Q_{0.75}$	QD	MD_{Mean}	Q. Skewness	SD
values	1.6633	1.1	1.4	0.15	0.1916	1	0.2328

From Table 2.1 it is evident that the distribution of nicotine levels in cigarettes is right skewed, with an average (mean) of 1.256 mg higher than the median of 1.1 mg , indicating some high-nicotine outliers. Most cigarettes contain around 1.1 mg of nicotine, with consistent levels across samples, evidenced by a tight interquartile range (1.1 mg to 1.4 mg) and low variability. The distribution has light tails, meaning few extreme outliers despite the skewness. Overall, a typical cigarette contains about 1.1 mg nicotine, but a small number with higher levels can influence the average and potentially increase addiction risk. Moreover, a violin plot of nicotine content with summary statistics (Table 2.1) has been provided in the following Figure 2.1:

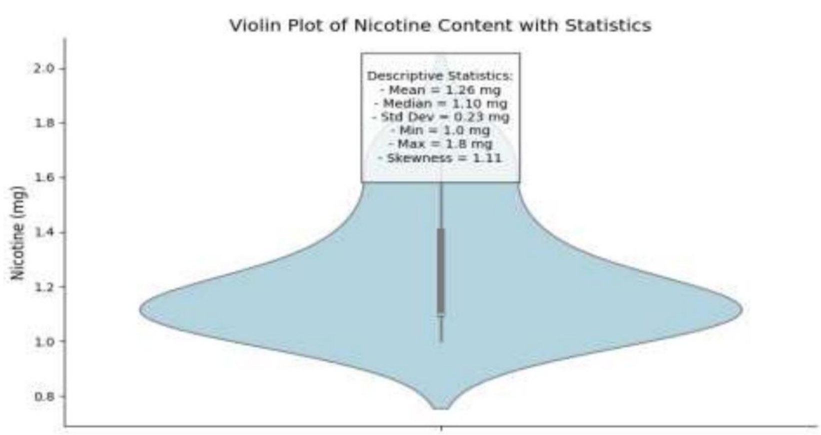

Figure* 2.1*: *Violin Plot of Nicotine Content with Summary Statistics*

It is obvious from Figure 2.1, as portrayed above, that the nicotine distribution is rightskewed, with a mean of 1.26 mg and a median of 1.10 mg , indicating most cigarettes contain around 1.1 mg , but some have higher levels up to 1.8 mg. The skewness of 1.11 confirms a moderate right tail, with fewer high-nicotine outliers. Variability is low to moderate ($SD = 0.23$ mg), and the range spans from 1.0 mg to 1.8 mg , with most data concentrated around 1.1 mg. The violin plot shows a dense cluster at 1.1 mg , tapering off toward higher values, and minimal density below 1.0 mg . While most cigarettes deliver typical nicotine levels, the presence of high nicotine outliers raises concerns about increased addiction risk and highlights the importance of monitoring and controlling these outliers through manufacturing oversight.

2.2 Applications of Probability Models and their Suitability

To analyze the nicotine data, various continuous statistical distributions will be fitted to the data. Moreover, goodness-of-fit tests, such as the Kolmogorov-Smirnov test, Anderson-Darling test, and Akaike Information Criterion *(AIC)*, will be used to evaluate the best-fitting model. For details, see, for example, Johnson *et al.* [5, 1994].

Continuous Statistical Distributions Considered for Nicotine Data Analysis

Sl. No.	Name of the Distributions	$f(x)$	Parameters
1	Exponential	$f(x;\lambda)=\lambda e^{-\lambda x}$ $x \geq 0, \lambda > 0$	$\lambda(>0)$: rate, or inverse scale parameter
2	Gamma	$f(x;\alpha,\beta)=\dfrac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$ $x \geq 0, \alpha > 0, \beta > 0$	$\alpha(>0)$: shape parameter $\beta(>0)$: rate parameter
3	Rayleigh	$f(x)=\left(\dfrac{x}{\sigma^2}\right)e^{-\frac{x^2}{2\sigma^2}}$ $x \geq 0, \sigma > 0$	$\sigma(>0)$: scale parameter
4	Inverse Gaussian	$f(x)=\sqrt{\frac{\lambda}{2\pi x^3}}\exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right)$	$\lambda(>0)$: shape parameter $\mu(>0)$: location parameter $E[X]=\mu$ is the mean, where $0 < x < +\infty$
5	Lognormal	$f(x)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2\right)$	$\sigma(>0)$: scale parameter μ (real): location parameter, and $0 < x < +\infty$
6	Weibull	$f(x)=\frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}\exp\left(-\left(\frac{x}{\beta}\right)^\alpha\right)$	$\alpha(>0)$: shape parameter $\beta(>0)$: scale parameter, and $0 \leq x < +\infty$

3 Results and Discussions

In what follows, we presented the results and discussions of our findings.

3.1 Distribution Fitting

Fitting of the above-said distributions to the nicotine data are carried as follows:

As a first step, we have tested the normality of the nicotine data by Ryan-Joiner Test, followed by plotting the histogram and probability plot of the data. These are provided in the following Figures 3.1 (a, b, c):

Ryan-Joiner Test (Similar to Shapiro-Wilk Test)
Test Statistic, Rp: 0.88406
Critical Value for 0.05 Significance Level: 0.95800
Critical Value for 0.01 Significance Level: 0.94000
Reject normality with a 0.05 significance level.
Reject normality with a 0.01 significance level.
Possible Outliers
Number of Data Values Below Q1 by More Than 1.5 IQR: 0
Number of Data Values Above Q3 by More Than 1.5 IQR: 0

Figure* 3.1*(a): *Ryan-Joiner Test of Normality Assessment*

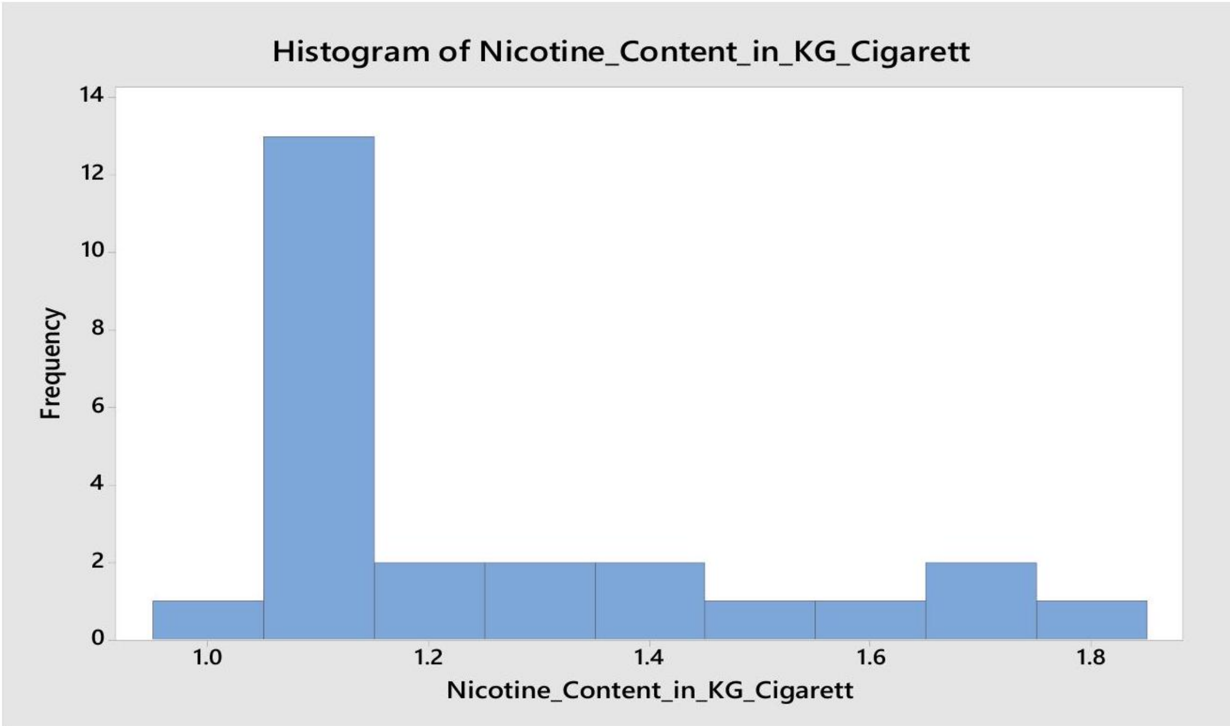

***Figure* 3.1*(b)*:** *Histogram*

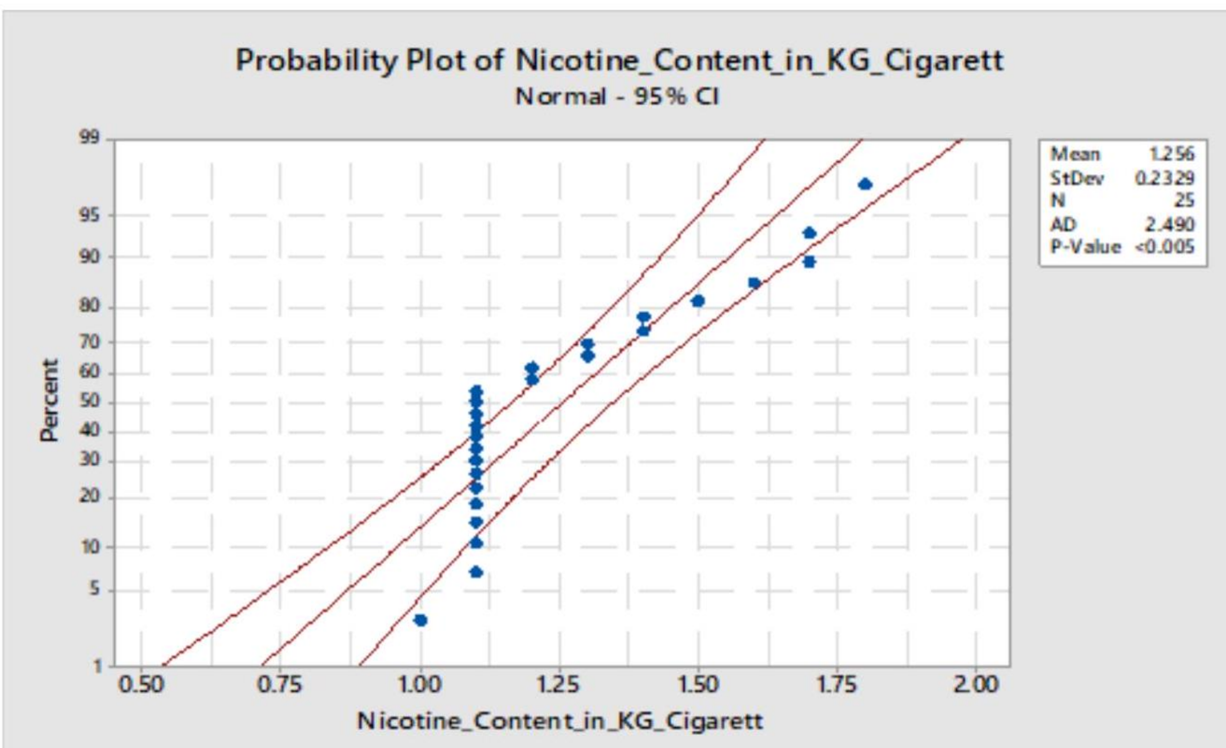

***Figure* 3.1*(c)*:** *Probability Plot*

From Figures 3.1 (a, b, c), it is obvious that the shape of the nicotine data is skewed to the right. This is also confirmed from the skewness of the summary statistics and the violin plot of the nicotine data in Table 2.1 and Figure 2.1, respectively.

3.2 Goodness of Fit

In this subsection, various discrimination criteria based on the log-likelihood function are employed, where: q represents the number of parameters, $\widehat{\Theta}$ denotes the *MLEs* of Θ, n is the sample size, k is the number of classes, $z_j = F_X(x_j), j = 1, 2, 3, \ldots$ are the ordered observations. The Akaike Information Criterion *(AIC)* by Akaike [1, 1973] optimizes model selection by balancing fit and complexity, while the Bayesian Information Criterion *(BIC)* by Schwarz [7, 1978] imposes a stricter complexity penalty, favoring simpler models in larger samples. These are defined as follows:

$$AIC = 2q - 2\ell(\widehat{\Theta}), \text{ and } BIC = q\ln(n) - 2\ell(\widehat{\Theta}).$$

The best model is accepted with the least values of the Kolmogorov-Smirnov ($K - S$) statistics:

$$KS = \max\left\{\frac{j}{k} - z_j, z_j - \frac{j}{k}\right\}.$$

Moreover, in order to compare various models, we will also use $-2\ln L$ critrion, the corrected Akaike information criterion *(CAIC)* (Sugiura [8, 1978]), the Bayesian information criterion *(BIC)* (Schwarz [7, 1978]), and Hannan-Quinn information criterion *(HQIC)* (Hannan and Quinn [4, 1979]), which are defined as follows:

$$CAIC = AIC + \frac{2q(q+1)}{n-q-1},$$

$$BIC = q\ln(n) - 2\ln L,$$

and

$$HQIC = 2q\ln[\ln(n)] - 2\ln L,$$

where q is the number of parameters in the statistical model, n is the sample size and $\ln L$ is the maximized value of the log- likelihood function under the considered model. The "best" distribution corresponds to the smallest values of $-2\ln L, AIC, BIC, CAIC, HQIC$ criteria. Table 3.1 shows the *MLEs* of the model parameters for the above-said nicotine data set.

Table 3.1: *Parameter Estimation and Goodness of Fit of Different Models*

Models	$\hat{\alpha}$	$\hat{\theta}$	A_0^*	W_0^*	KS	$p-$ value
ED	0.7962	-	3.5414	0.7315	0.5489	0.0088
WD	5.3847	1.3552	1.7890	0.2396	0.2666	0.5442
GD	33.7282	0.0372	2.2943	0.3092	0.3132	0.3402
RD	0.9026	-	2.2943	0.3092	0.3132	0.3402
IGD	1.2560	43.774	2.3362	0.3224	0.3221	0.3078
LND	0.2130	0.1683	0.8763	0.0993	0.2099	0.8224

The Goodness-of-Fit analysis (Table 3.1) reveals that the Lognormal *(LND)* and Weibull *(WD)* distributions provide the best fits for modeling nicotine content in cigarettes, as evidenced by their high p-values (0.8224 and 0.5442, respectively) and low test statistics (e.g., LND's $A_0{}^* = 0.8763$. In contrast, the Exponential Distribution *(ED)* is unsuitable (p = 0.0088), while the Gamma (GD) and Rayleigh (RD) distributions show inconsistencies, with identical metrics suggesting potential errors. The Inverse Gaussian *(IGD)* offers a moderate fit but is outperformed by LND and WD. These findings recommend using $\boldsymbol{LND}$ for robust modeling of nicotine data due to its alignment with the right-skewed distribution, while highlighting the need to verify *GD/RD* results. For practical applications, this implies prioritizing lognormal-based analyses for accurate predictions and regulatory assessments. Moreover, from Figure 3.1(d) it is evident that the nicotine content in KG cigarettes is best modeled by a right-skewed distribution, such as a log-normal or Gamma distribution. The data shows most cigarettes contain nicotine levels around $1.4 - 1.6$mg, with fewer instances of higher nicotine content up to 2.0 mg. The distribution's positive skewness indicates that while typical nicotine levels are clustered centrally, there are occasional higher outliers. These characteristics suggest that a right-skewed distribution accurately captures the data's pattern, aiding in better statistical modeling and understanding of nicotine variability across cigarette types.

3.3 Information Criteria Analysis

It is provided in the following Table 3.2:

Table 3.2: *Goodness of Fit of Different Continuous Distributions*

Models	$-\ell$	AIC	AICC	BIC	HQIC	CAIC
ED	30.6983	63.3966	63.5705	64.6155	63.7347	65.6155
WD	0.4795	4.9589	5.5044	7.3967	5.6351	9.3967
GD	-3.05631	-2.1126	-1.5671	0.3251	-1.4365	2.3251
RD	14.5538	31.1076	31.2816	32.3265	31.4457	33.3265
IGD	-3.7757	-3.5514	-3.0060	-1.1137	-2.8753	0.8863
LND	-3.7528	-3.5056	-2.9602	-1.0678	-2.8295	0.9321

From Table 3.2 it is clear that the information criteria analysis identifies the Lognormal *(LND)* model as the best fit for nicotine content data, demonstrating superior performance with the lowest *AIC* (-3.5056), *BIC* (-1.0678), and *HQIC* (-2.8295) values, which indicate an optimal balance between model fit and complexity. The Inverse Gaussian *(IGD)* is a close second, offering marginally higher but still competitive metrics, making it a viable alternative for robustness checks. Both models align well with the right-skewed nature of nicotine data, as previously confirmed by goodness-of-fit tests. In contrast, the Weibull *(WD)*, Gamma *(GD)*, Rayleigh *(RD)*, and Exponential *(ED)* distributions are less suitable due to higher information criteria values and poorer fits. For practical applications, the ***LND*** model is recommended for its consistency, interpretability, and statistical robustness in modeling cigarette nicotine content.

3.4 Visual Fit

The best fitting of the above-said considered probability models to the nicotine data also aligns with the visual evidence from the density plot as provided in the following Figure 3.1 (d):

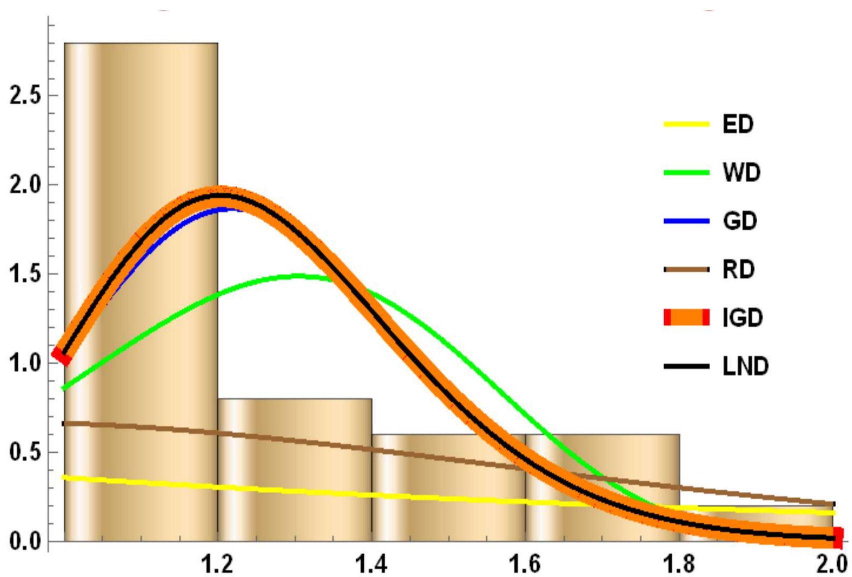

Figure 3.1(d): *Histogram of fitted Models*

4 Conclusion

This study explored the suitability of various continuous statistical distributions in modeling nicotine consumption data, with the objective of identifying the distribution that best characterizes the observed patterns. By applying robust distribution fitting techniques, the analysis provided a statistical foundation for understanding how nicotine intake is distributed
across a population. The results demonstrated that certain distributions more accurately represent the data, offering valuable insights into the underlying behavior of nicotine use.

These findings are not only statistically significant but also carry important implications for public health policy and clinical practice. Accurate modeling of nicotine consumption can support better prediction of usage trends, facilitate risk assessments, and inform targeted interventions aimed at reducing nicotine-related harm. As such, this research underscores the critical role of statistical analysis in public health and contributes to the growing body of knowledge on the epidemiology of nicotine use.

Future studies may build upon this work by incorporating additional variables such as demographic or behavioral factors to enhance model precision and relevance. Ultimately, the application of well-fitted statistical models is a vital tool in shaping effective strategies to address nicotine consumption and its associated health risks.

Acknowledgement
The authors are grateful to the Executive Editor of *Jñānābha*, Professor Dr. R. C. Singh Chandel, for his kind encouragement, support, and valuable and helpful comments and suggestions for modifying and updating the manuscript, which greatly and considerably helped in improving the quality and presentation of the manuscript. Finally, the authors are thankful to the anonymous referees for their helpful comments.

Declarations: Authors' Contributions: All the authors have equally made their contributions, and have read and approved the final manuscript.

Funding: The authors state that they have no funding source for this paper. Availability of Data and Materials: Not applicable.

Declarations Conflict of Interest: The authors declare that they have no competing interests. The authors state that no funding source or sponsor has participated in the realization of this work.

References

[1] H. Akaike, Information Theory as an Extension of the Maximum Likelihood Principle, In: B. N. Petrov, and F. Csaki, Eds., Second International Symposium on Information Theory, Akademiai Kiado, Budapest, (1973), 267-281.

[2] N. L. Benowitz, Nicotine addiction, The New England Journal of Medicine, **362**(24) (2010), 2295-2303. https://doi.org/10.1056/NEJMra0809890

[3] N. L. Benowitz, and A. D. Burbank, Cardiovascular toxicity of nicotine: Implications for electronic cigarette use, Trends Cardiovasc Med., **26**(6) (2016), 515-23. doi: 10.1016/j.tcm.2016.03.001.

[4] E. J. Hannan, and B. G. Quinn, The Determination of the order of an autoregression, Journal of the Royal Statistical Society, Series B, **41**(2) (1979), 190 - 195.

[5] N. L. Johnson, S. Kotz, and N. Balakrishnan, Continuous Univariate Distributions, Volume 1 (6th edition), John Wiley & Sons, New York, 1994.

[6] L. M. Martin, and M. A. Sayette, A review of the effects of nicotine on social functioning, Exp Clin Psychopharmacol, **26**(5) (2018), 425-439. doi: 10.1037/pha0000208.

[7] G. Schwarz, Estimating the dimension of a model, The annals of statistics, **6**(2) (1978), 461-464.

[8] N. Sugiura. Further analysis of the data by Akaike's information criterion and the finite corrections, Communications in Statistics - Theory and Methods, **7**(1) (1978), 13 - 26.

[9] M. F. Triola, Elementary Statistics Using Excel, 7th edition, Pearson Education, Inc., NJ, USA, 2022.