UNIVERSIDAD DE GRANADA

Universidad de Jaén

VI Jornadas de Estadística como Herramienta Científica
4-5 Febrero 2026 / Jaén (España)

Normal or skewed normal? Models of mixed distributions of suicide age in Galicia

Andrea Ruiz-Vega[1], Elena Cruz-Guardia[2] Miguel Ángel Montero-Alonso[3], Juan de Dios Luna del Castillo[4]

[1]arvega@ugr.es, [2]elenacruz@correo.ugr.es, [3]mmontero@ugr.es, [4]jdluna@ugr.es
Department of Statistics and Operational Research, University of Granada

biostatnet

Introduction

The analysis of suicide mortality requires models capable of reflecting heterogeneity in the age of occurrence. This study compares two mixture approaches: two-component normal distributions and two-component skewed normal distributions, applied to data from Galicia (1998–2022). Suicide counts by age are interpreted under a multinomial model and expanded into individual observations, allowing for direct adjustment of continuous mixtures. The results evaluated using AIC, BIC, and log-likelihood, show that the mixture of two normals provides a better fit than models with asymmetry.

Selection of the model

Δ BIC	Rejection of the highest BIC
0 a 2	Simple mention
2 a 6	Positive
6 a 10	Strong
> 10	Very strong

Bayesian Information Criteria

BIC	Men		Women	
	Norm	Skew	Norm	Skew
98-02	9413.781	**9401.566**	**3278.993**	3291.445
03-07	**9956.969**	9971.858	**3662.490**	3675.247
08-12	**9984.366**	10000.743	**3324.264**	3336.943
13-17	**10552.316**	10570.937	**3920.098**	3933.686
18-22	**9386.328**	9398.461	**3822.891**	3831.857

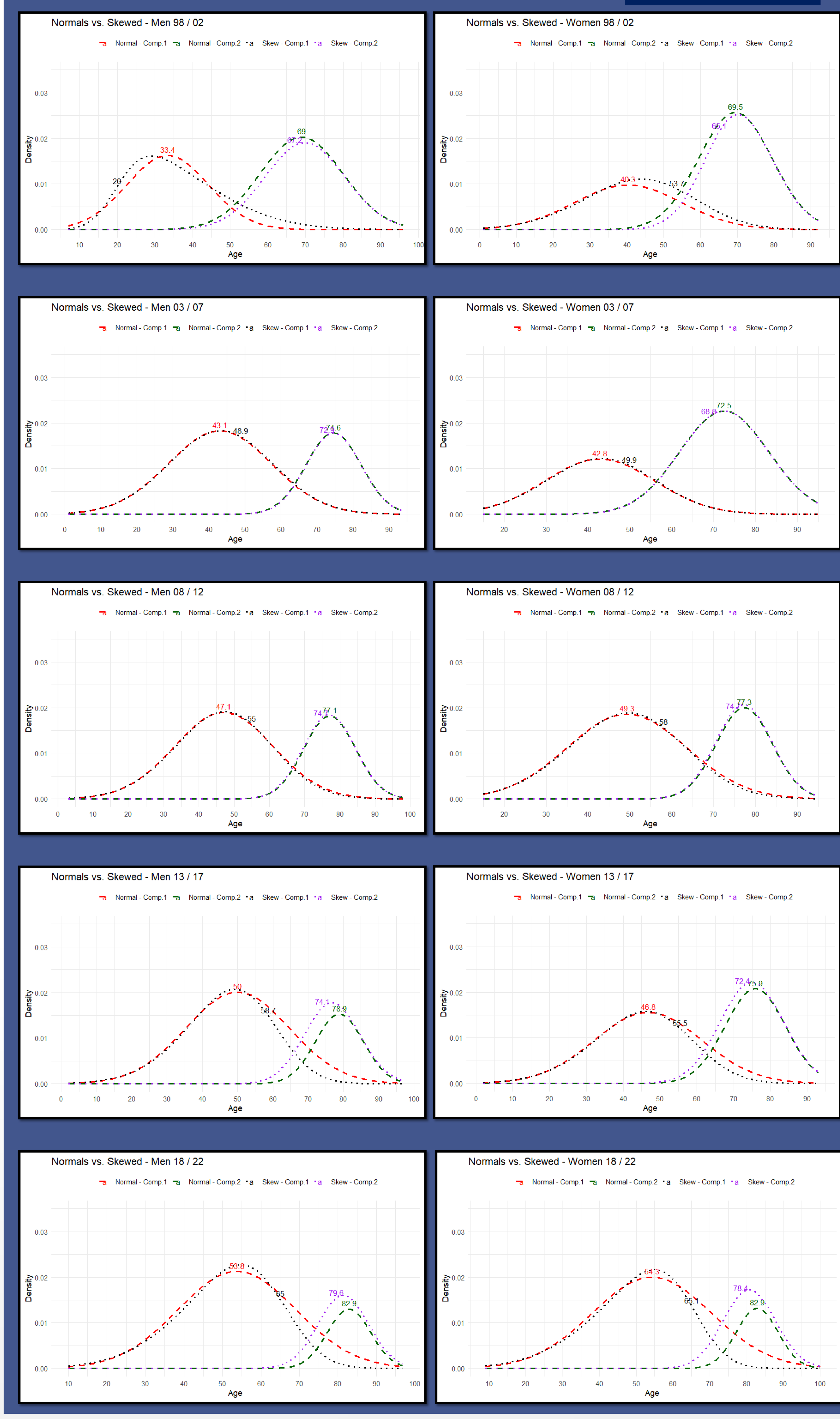
Methodology

- Disaggregate data by age and sex based on official counts, interpreted using a multinomial model.
- Estimate mixture models of normal and skewed normal distributions, each with two components.
- Calculate the Bayesian information criterion (BIC) for each model.
- Select the model with the lowest and/or most efficient BIC.
- Estimate the model parameters using maximum likelihood.
- Accept the final model for interpretation and analysis.

Age groups have different weights in the distribution to be adjusted, since they are made up of different populations. Rates allow us to compare between different populations or groups that have different population sizes, i.e., that are not homogeneous. There are several reasons why it is preferable to model rate data rather than direct count data:

- *Normalization of exposure size. Allows comparison of groups of different sizes.*
- *Tendency to reduce variance. Useful when counts are low.*
- *Likelihood of meeting assumptions of normality of errors. Errors are assumed to follow a normal distribution.*

Results



Conclusions

Analysis of the age distribution reveals a clear bimodality, more pronounced in men than in women, which justifies the use of two-component mixture models. The model selection was confirmed using the Bayesian Information Criterion (BIC). The current BIC values comparing Normal (Norm) and Skewed Normal (Skew) distributions show that Normal models are slightly inferior (and therefore preferable) in all periods and sexes, supporting the efficient choice of the two-component fit with Normal distributions. Using EM and maximum likelihood algorithms, the parameters $(\mu_i, \lambda_i, \sigma_i)$ of these subpopulations were obtained.

In men, the two components are consistently better separated, with the means further apart than in the graphs for women. This reinforces the conclusion that bimodality is more evident or pronounced in men.

The Skew-Normal model often shows a slight visual advantage. Skew curves tend to capture the shape of the distribution better than Normal curves, especially in Component 1 (young age), which tends to be more skewed to the left. This is consistent with the practice where young age distributions tend to be skewed.

In the graphs, the visual difference between the Normal and Skew fits is minimal in many periods. This visualizes and justifies why the BIC chose the Normal model, as the improvement in fit offered by the Skew model (with an extra parameter per component) does not compensate for the added complexity. The graphs confirm the robustness of the two-component structure in the age distribution. They visualize the greater separation in men and the small difference in fit between the Normal and Skew models, which explains why the BIC favored the simplicity of the Normal model.

References

- Azzalini, A. (1985). A Class of Distributions Which Includes the Normal Ones. Scandinavian Journal of Statistics, 12(2), 171–178. http://www.jstor.org/stable/4615982.
- Azzalini A. (2020). Further Results on a Class of Distributions which Includes the Normal Ones – Looking Back. Statistica (Bologna), 80(2), 161–175.
- Everitt, B. (1996). An introduction to finite mixture distributions. Statistical Methods in Medical Research, 5(2), 107-127. https://doi.org/10.1177/096228029600500202.
- INE - Instituto Nacional de Estadística. (s.f.). Defunciones por causas (lista reducida) por sexo y grupos de edad. INE. Recuperado 30 de noviembre de 2024, de https://www.ine.es/jaxiT3/Tabla.htm?t=10803.
- INE - Instituto Nacional de Estadística. (s.f.). Población por comunidades, edad (año a año), Españoles/Extranjeros, Sexo y Año. INE. Recuperado 30 de julio de 2025, https://www.ine.es/jaxi/Tabla.htm?path=/t20/e245/p08/l0/&file=02003.px&L=0.
- McLachlan, G., & Peel, D. (2000). Finite Mixture Models. Wiley-Interscience.
- Ruiz-Vega, A., Montero-Alonso, M.Á., & Luna, J.D. (2024). Mixture of distribution models: Application to the age distribution of people who committed suicide. Investigación Operacional, 45(3), 396-409.
- Ruiz-Vega, A., Montero-Alonso, M.Á., Luna, J. (2025). Mortalidad por suicidio en España. Distribución por edad, sexo y ámbito geográfico (1998-2022), DIGIBUG, https://hdl.handle.net/10481/104569.
- Schlattmann, P. (2009). Medical Applications of Finite Mixture Models. Springer Science & Business Media.