Constructing a Structural Equation Model for Predicting Biological Age: A Technical Report under $AMA\hat{}-1$ perspective *

Álvaro Novillo Correas *100483441@alumnos.uc3m.es*
Paolo Salvatore Lodato Olano *100493512@alumnos.uc3m.es*

This article discusses the implementation of AMA^{-1}, a methodology that emphasizes the idea that statistics is used to solve real-world problems. The acronym AMA^{-1}, represents the three fundamental steps in this methodology: understanding the real-world problem and related data (A), translating the real-world problem into a statistical problem and solving it (M), and translating the statistical results back into terms of the original problem (A^{-1}). The article presents basic aspects of each step along with an example of its implementation

Keywords: AMA^{-1}, SEM

Introduction (A)

In this article, we will implement the AMA-1 methodology in R, dealing with the challenge of estimating biological age using observed health indicators. The dataset[1] comprises information on weight, sickness status, and chronological age. To do so, we will follow the three fundamental steps of the methodology: understanding the real-world problem and related data (A), translating the real-world problem into a statistical problem and solving it (M), and translating the statistical results back into terms of the original problem (A-1).

Making a preliminary analysis of the dataset, we can see that over half of the animals in our sample are affected by illness, with an illness status average of 0.6. Additionally, the age variable is left-skewed, suggesting a higher frequency of middle-aged animals.

Method (M)

Employing Structural Equation Modeling (SEM)[2], we construct a model to predict biological age as a latent variable. The relationship between the latent variable and each observed variable is quantified by the factor loadings. Standardization is applied to the observed variables to facilitate meaningful interpretation of coefficients and improve numerical stability in the estimation process.

Next, to make a comparison between biological age and measured chronological age, a regression analysis will be conducted using obtained biological age and chronological age values. If the regression coefficient is found to be significant and approaches one, it would demonstrate the correlation between the latent and observed variables, indicating that the animals are undergoing regular and healthy aging processes.

Table 1: # Loading

Link	Coefficient	SE	95% CI	z	p
bio_age =~ weight	1.00	0.00	(1.00, 1.00)		< .001
bio_age =~ sick	0.86	0.20	(0.48, 1.24)	4.39	< .001

*Replication files are available on the author's Github account (https://github.com/AlvaroNovillo). **Current version**: noviembre 14, 2023; **Corresponding author**: alvanovi@ucm.es.

[1]The dataset can be found at: https://aulaglobal.uc3m.es/mod/resource/view.php?id=4711061

[2]The lavaan package in R is employed for model estimation using the Maximum Likelihood (ML) method. The algorithm iteratively refines parameter estimates to maximize the likelihood of observing the given data under the specified model. See Rosseel (2012) for further insight.

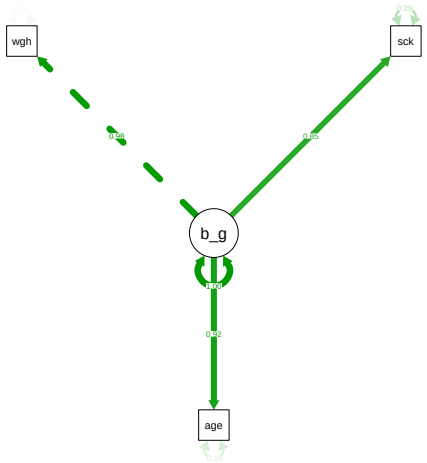
Table 2: # Regression

Link	Coefficient	SE	95% CI	z	p
age ~ bio_age	0.93	0.17	(0.61, 1.26)	5.66	< .001

Based on these results, given the factor loading (Table. 1) for weight and sick we can conclude that they are a significant predictor of biological age in these animals. Furthermore, the regression coefficient (Table. 2) of 0.934 indicates a positive correlation between biological age and chronological age, confirming a direct relationship between the latent and observed variables. Given the 95% confidence interval calculated for this coefficient, it is not possible to make any claims regarding whether the animals' chronological age corresponds to their biological age. To do so, it would be necessary to narrow the confidence interval, which could be achieved by increasing the sample size of the population.

Results (A^{-1})

The constructed Structural Equation Model (SEM) serves as a valuable framework for unraveling the intricate connections between observed health indicators and the latent construct of biological age. The model, supported by strong factor loadings and a significant regression coefficient, underscores the pivotal role of sickness status and weight in capturing the essence of biological aging.



Using the `semPlot` package we create path diagrams to visualize the structural equation model (SEM). Here, nodes represent variables and arrows indicate directional relationships between them. The numbers on arrows represent the path coefficients (regression weights) between variables.

While the model exhibits commendable fit indices, acknowledging its limitations is imperative for a comprehensive interpretation. The current sample size, while sufficient for initial exploration, beckons for expansion to enhance robustness and generalizability. A larger and more diverse dataset would fortify the model's capacity to discern nuanced patterns within the aging process.

Furthermore, recognizing the influence of contextual factors is crucial. Incorporating a richer contextual understanding of the database could empower us to address the intricacies of biological age estimation with a more sophisticated statistical approach.

References

Rosseel, Yves. 2012. "lavaan: An R Package for Structural Equation Modeling." *Journal of Statistical Software* 48 (2): 1–36. https://doi.org/10.18637/jss.v048.i02.