

A Technical Report under AMA^{-1} perspective *

Álvaro Novillo Correas 100483441@alumnos.uc3m.es
Paolo Salvatore Lodato 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} , Statistics

Introduction (A)

In this case, we will implement the AMA^{-1} methodology in R using the dataset “bioage.csv”¹. 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).

The dataset used includes data on a specific animal species, detailing their weight, illness status (1= sick), and chronological age. Our article aims to investigate if there exists a method of determining the biological age of these animals and how it compares to their chronological age.

Table 1 present the preliminary analysis of the dataset. The results indicate 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)

We will try to determine the biological age of these animals using a linear model with the variables at hand. In this case, the model used will be the following

$$age^* = \beta_0 + \beta_1 \cdot weight + \beta_2 \cdot sick$$

In this case, we will compare the predicted age (age^*), which we will consider as the biological age of the animals, with their chronological age. Using our 10 observations of our dataset, we get the following results:

Parameter	Coefficient	SE	95% CI	t(7)	p
(Intercept)	7.83	9.34	(-14.25, 29.92)	0.84	0.429
weight	0.15	0.05	(0.03, 0.28)	2.85	0.025
sick	2.09	6.99	(-14.44, 18.62)	0.30	0.774

*Replication files are available on the author’s Github account (<https://github.com/AlvaroNovillo>). **Current version:** noviembre 10, 2023; **Corresponding author:** alvanovi@ucm.es.

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

Based on these results, we can conclude that weight is a significant predictor of biological age in these animals, while illness status (sick) does not have a significant effect on biological age. The model can be used to predict the biological age of animals based on their weight (R^2 / R^2_{adj} 0.815 / 0.763)², and we can compare this predicted biological age to their chronological age to determine if they are older or younger than expected. Further analysis using innovative ideas such as machine learning algorithms or Bayesian inference could provide additional insights into the biological age of these animals.

Now we need to verify whether the assumptions of the linear model have been met. As seen in Fig. 1, the model fulfils the assumptions, albeit with some likely observations being flagged as problematic in the top right plot.

Results (A^{-1})

We can use the model equation to predict the biological age of each animal in the dataset and compare it to their chronological age. If the predicted biological age is higher than the chronological age, it indicates that the animal is aging faster than expected, while if it is lower, it indicates that the animal is aging slower than expected.

Fig. 2 represents the biological predicted age, with respect to the chronological age of the animals. Animals (points) located above the dotted diagonal line are aging more slowly than expected, while those situated below the line are aging at a faster rate than expected. The color of the data points denotes their illness status. As predicted, this variable has no significant effect on predicting the animals' biological age.

In our study, it was discovered that 70% of the animals in our sample are undergoing accelerated aging. Unfortunately, due to the lack of context provided, we are unable to provide clear justification for these findings. More context would also allow us to tackle the problem with a more sophisticated statistical approach

References

García-Portugués, E. 2023. "Notes for Predictive Modeling." <https://bookdown.org/egarpor/PM-UC3M/>.

²The R^2_{adj} is a measure of goodness of fit of the multiple linear regression models. See (García-Portugués (2023)) for further insight.

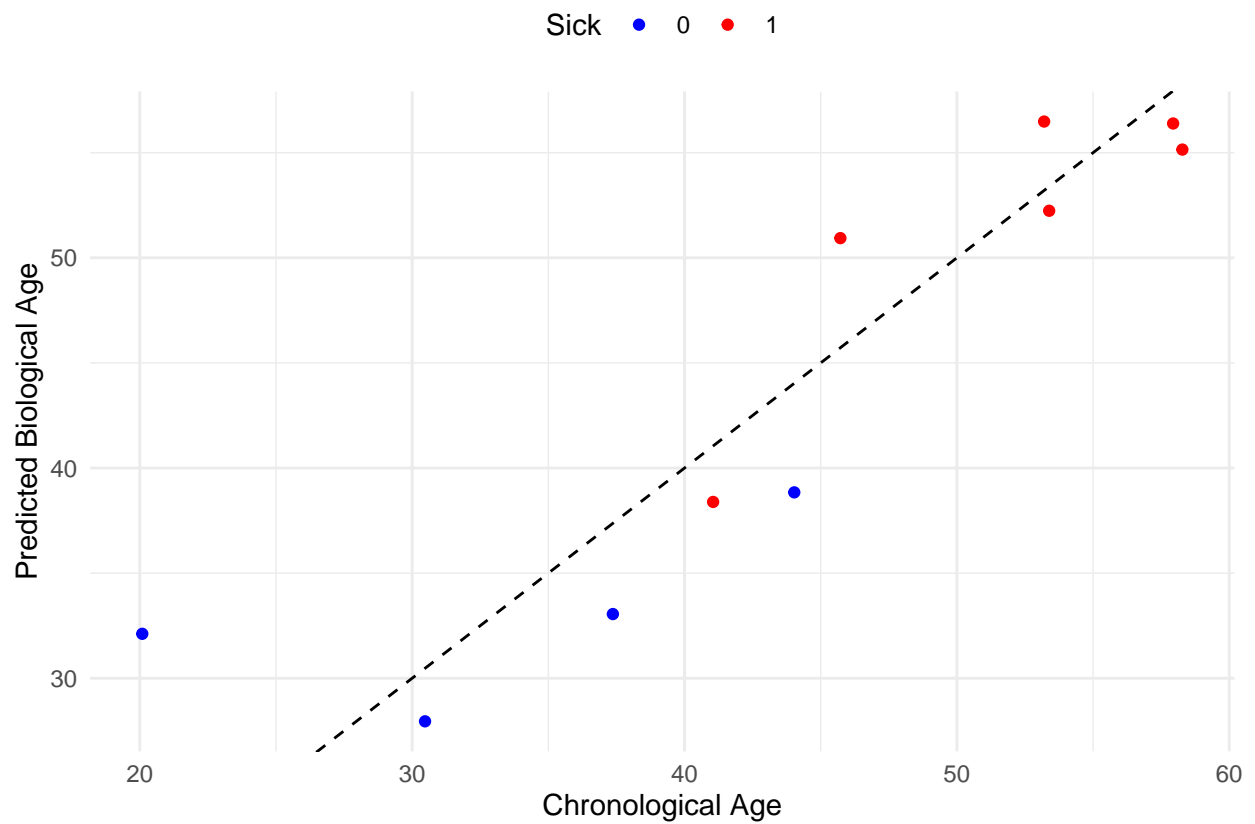


Figure 1: Scatterplot of the predicted biological age of the animals based on the linear regression model versus their chronological age. The illness status of each subject is represented as red if their are sick, and blue if they are healthy.

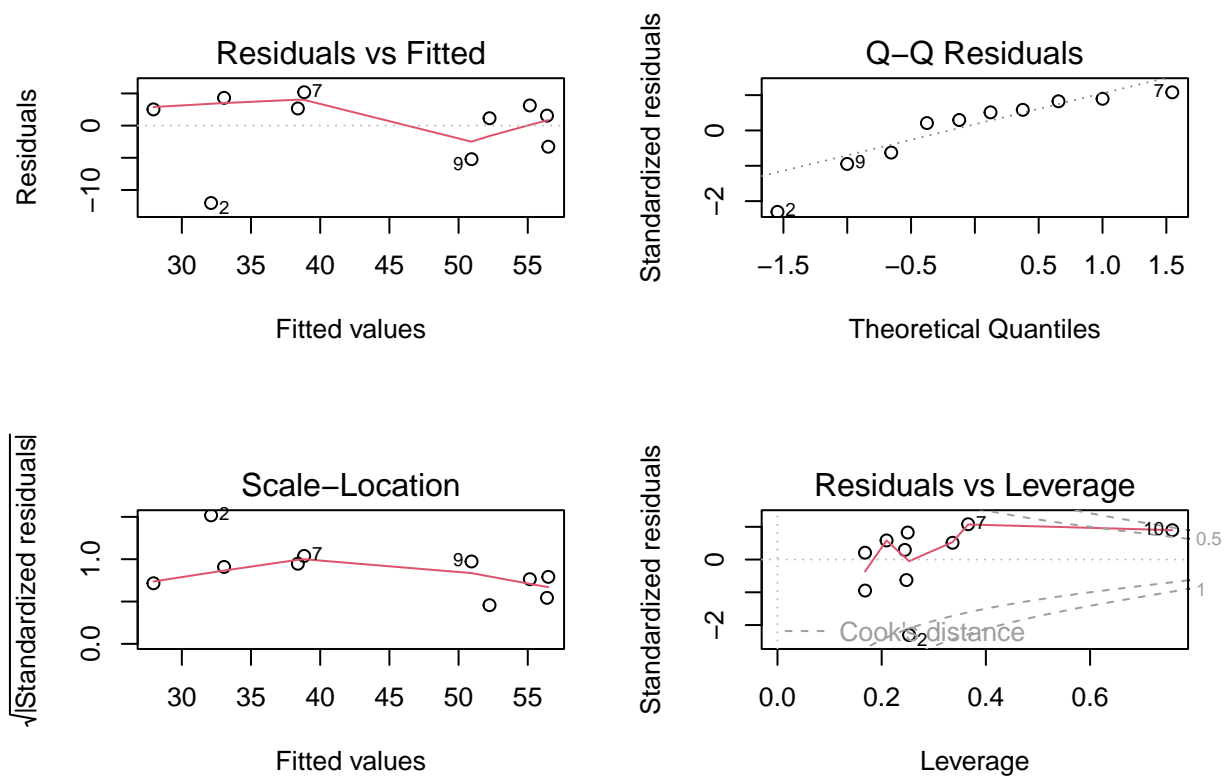


Figure 2: Model assumptions