

Example of Technical Report

Lúa Arconada Manteca y Alejandro Macías Pastor

20 November 2023

Introduction

The data set contains information about some individuals of a given species of animals. More specifically, it contains three variables: their weight (in pounds), their illness status, and their chronological age (in years). From the information contained in this data set we want to estimate the biological age of an individual of this species.

Method

We employ a multiple linear regression model to estimate the biological age from the weight and illness status, so that:

$$BA = \beta_0 + \beta_1 \cdot \text{weight} + \beta_2 \cdot \text{illness} + \beta_3 \cdot \text{weight} \cdot \text{illness}$$

Following the method explained in [1], a second linear regression model is performed, this time between the predicted biological age and the real chronological ages:

$$BA = a + b \cdot CA$$

The predicted biological age is then corrected by adding the deviation from the mean of the recorded chronological age times a weighting factor related to the regression coefficient of the second model:

$$BA_c = BA + z$$

where

$$z = (CA - \overline{CA}) (1 - b)$$

The predictors in the first model will be a numerical value indicating the weight of the individuals (in pounds) and a binary variable indicating their illness status (1 if they are sick; 0 otherwise). This model also includes a term accounting for the relation between the two predictors. As for the second model, the only predictor is a numerical variable containing the chronological age.

We expect the model to predict higher biological than chronological age for sick individuals, as well as for individuals with higher weights.

Results

The first model's output produces the following regression coefficients: -4.01948, 0.22526, 18.98889 and -0.09069, for the intercept, weight, illness status and relation between weights and illness status terms, respectively. On the other hand, the second model's output is 7.6141 for the intercept term and 0.8276 for the age coefficient. The values of these coefficients indicate that higher weights lead to a higher biological age, and so does getting sick. The last coefficient indicates that in sick individuals a higher weight might be a sign of a lower biological age.

The p-values of the first model's output indicate that none of the variables are individually significant within the framework of the model, although they are not all simultaneously zero. This means, if we removed one specific variable, the resulting model would be better at predicting the age.

Below, the predicted biological ages are shown:

```
##           1           2           3           4           5           6           7           8
## [1,] 31.91373 27.55251 53.75206 57.45055 57.15364 58.18584 41.57676 23.21821
##           9          10
## [1,] 51.28622 39.45251
```

These predicted biological ages do not show any specific pattern concerning the weight and illness status of the individuals.

Bibliography

[1] Bafei, S.E.C., Shen, C. Biomarkers selection and mathematical modeling in biological age estimation. npj Aging 9, 13 (2023). <https://doi.org/10.1038/s41514-023-00110-8>

Appendix

```
df = read.table('bioage.csv', header=T, sep=';', fill=T)
df=df[,-1]
df$sick=factor(df$sick)

model=lm(age~weight+sick+weight:sick, df)

summary(model)

##
## Call:
## lm(formula = age ~ weight + sick + weight:sick, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.611  -1.750   1.680   3.278   4.895
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -4.01948    20.63631  -0.195    0.852
## weight         0.22526     0.12413   1.815    0.119
## sick1        18.98889    26.94686   0.705    0.507
## weight:sick1  -0.09069     0.13919  -0.652    0.539
##
## Residual standard error: 6.3 on 6 degrees of freedom
## Multiple R-squared:  0.8276, Adjusted R-squared:  0.7413
```

```
## F-statistic: 9.598 on 3 and 6 DF, p-value: 0.01047
```

```
model2=lm(predict(model)~df$age)
```

```
summary(model2)
```

```
##
```

```
## Call:
```

```
## lm(formula = predict(model) ~ df$age)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

```
## -7.2541 -2.2385 -0.4396  3.2765  7.4615
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)   7.6141     6.1026   1.248 0.247426
```

```
## df$age         0.8276     0.1336   6.196 0.000261 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 4.963 on 8 degrees of freedom
```

```
## Multiple R-squared:  0.8276, Adjusted R-squared:  0.806
```

```
## F-statistic: 38.39 on 1 and 8 DF, p-value: 0.0002605
```

```
predict(model) + (df$age- mean(df$age))*(1-model2$coefficients[2])
```

```
##      1      2      3      4      5      6      7      8
```

```
## 31.91373 27.55251 53.75206 57.45055 57.15364 58.18584 41.57676 23.21821
```

```
##      9     10
```

```
## 51.28622 39.45251
```