

Longitudinal data analysis: Assignment 3

Team B: Kendall Brown *r0773111*
Stefan Velev *r0924289*

Raïsa Carmen *s0204278*
Adhithya Unni Narayanan *r0776057*

1 Introduction

This report assesses the evolution of hearing thresholds over time for a sample of 546 healthy male volunteers. The data originates from the famous Baltimore Longitudinal Study of Aging (BLSA). Previous research showed a change in hearing threshold for all age groups but especially the older population (Brant and Fozard 1990). In this report, we will especially take care of missingness in the data.

First, the *TIME* variable is rounded to the nearest integer value. As such, we aim to balance the dataset with equally-spaced time instances when hearing thresholds are measured.

Hearing thresholds will be explored, both as a continuous variable and as a trichotomized (ordinal) variable with the following three levels:

- ≤ 6 dB: Excellent hearing
- over 6 and ≤ 25 dB: Normal hearing
- ≥ 26 dB: Hearing loss

1.1 Missingness exploration

After discretizing the *TIME* variable, we consider a subject to be missing at a certain time instance if there is no measurement for that subject at that time. It should be noted that, if the subject is not missing (18.16% of TIME-subject instances), we usually (16.88% of TIME-subject instances) have two measurements (one for each ear) at each time instance. In fact, the average number of measurements per subject at each time instance is 0.35, and maximum 4.

Figure 1 was created using the *visdat* package. It shows all subjects, ordered from youngest (in the top) to oldest (in the bottom) and whether or not their data is missing at a certain time instance (on the x-axis). The percentages on top shows the percentage of missingness at each time instance. It is clear from Figure 1 that the missingness is not monotone; subjects may be missing at one time instance and come back later. Since there are too many possible missingness patterns with 23 time instances (2^{23}), we do not give an overview of the number of subjects that follow each possible pattern. Instead, figure 2 shows, for each time instance, the number of subjects that:

- are *present*: when the subject's hearing is measured at time t and $t - 1$
- are *missing*: when the subject is missing at time t and $t - 1$
- *drop out*: when the subject's hearing is measured at time $t - 1$ but not at time t
- *return*: when the subject's hearing is measured at time t but not at time $t - 1$

Table 1: A mixed model to predict missingness.

Variable	Estimate
Intercept	2.22 ***
TIME	0.24 ***
sideright	-0.19 ***
age	0.10
R_{t-1}	-2.67 ***
sigma	2.31

Figure 2 clearly shows that subject rarely are measured two years in a row, most are not measured at $t = 1$, and the number of subjects that stay missing gradually increases as time passes.

Lastly we explore whether the missingness can be explained by the data by fitting a mixed model to a dataset where R_{it} is equal to one if the hearing threshold is missing and zero otherwise:

$$\begin{cases} \text{logit}(R_{it}) = \beta_0 + \beta_1 \text{TIME}_{it} + \beta_3 \text{side}_{it} + \beta_4 \text{age}_{it} + \beta_5 R_{it-1} + b_i \\ b_i \sim N(0, \sigma^2) \end{cases} \quad (1)$$

The variable *age* was standardized to get convergence in the model. Table 1 shows that TIME is significant; as time increases, subjects are more likely to be missing. We can therefore assume missingness at random (MAR). Left ear measurements are also more likely to be missing. A subject is also less likely to be missing at time t if he was missing at time $t - 1$. This can be seen especially in the first couple of years in figure 1: all but 3 subjects are measured at $t = 0$, almost no-one is measured at $t = 1$ and many are measured again at $t = 2$.

2 Methodology

First, a direct likelihood analysis is compared with multiple imputation in the continuous case. Next, weighted generalized estimating equations are compared with ‘multiple-imputation generalized estimating equations’. Lastly, a sensitivity analysis is performed.

For imputation, the *mice* library is used (Buuren and Groothuis-Oudshoorn 2011) and different imputation techniques were tested: Predictive mean matching, Bayesian linear regression, Unconditional mean imputation, and imputation by random forests.

All analysis was done in R. All scripts are freely available at this git repository.

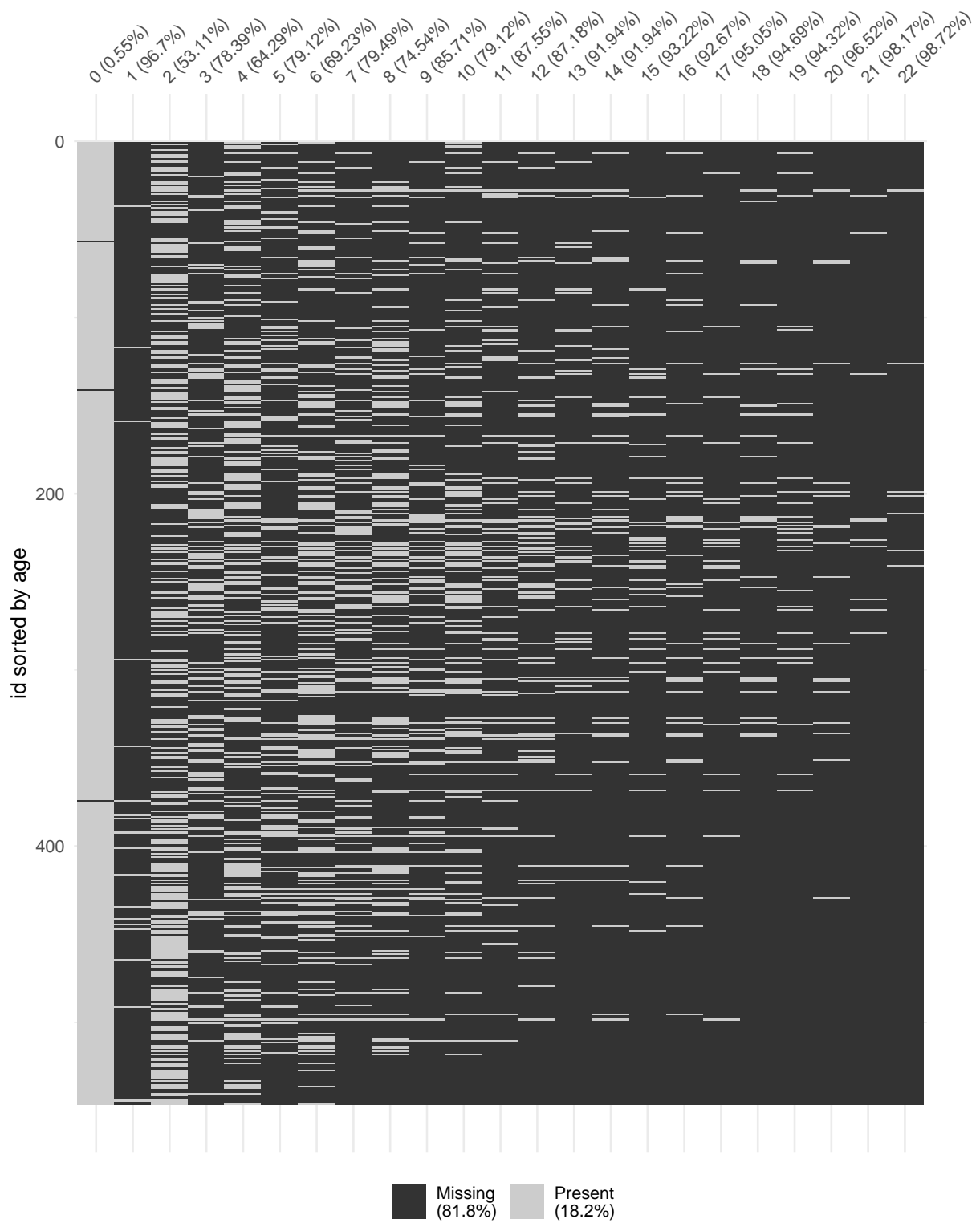


Figure 1: Visual inspection of missingness for different ages at different time instances.

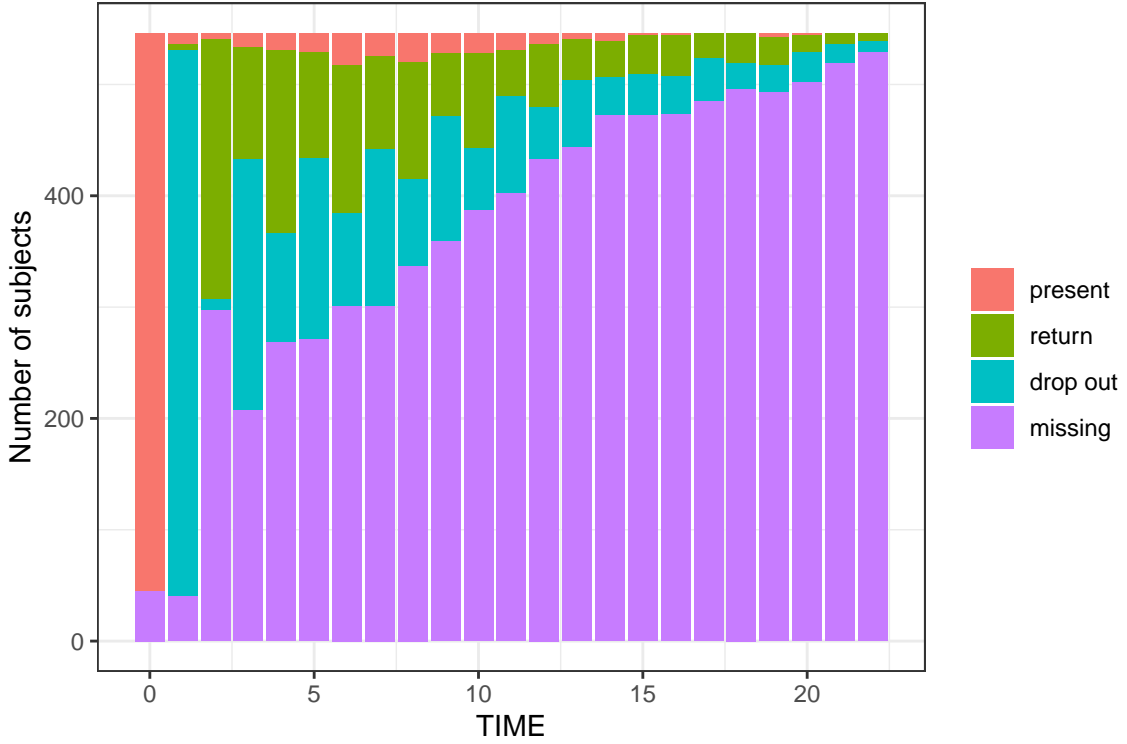


Figure 2: Number of subjects the are present, return, drop out or are missing at each time instance.

3 Results

3.1 Direct Likelihood Analysis

For the purposes of this section we will analyze the hearing data respective to the continuous response. Two popular ways to analyze data sets with missing data is to use either a direct likelihood method or an analysis through multiple imputation. In a direct likelihood analysis, missing values are assumed to be ignorable (Beunckens, Molenberghs, and Kenward 2005). Drafting a linear mixed effects model within R allows for the direct calculation for each patients' respective likelihood functions. The R-package *nlme* allows for this model to be fit in a way such that the parameter estimates are representative of their parameters' maximum log-likelihood. Additionally, we will compare models of differing covariance structures: "simple" and "compound symmetry." Summary statistics for these models can be seen in Table 2.

Parameter estimates from both covariance structures remain stable.

3.1.1 Multiple Imputation

We shall now draft and evaluate models generated by multiple imputation methods. Here, artificial data points are generated according to the algorithms: Predictive mean matching, Bayesian linear regression, Unconditional mean imputation, and imputation by random forests. As we are dealing with continuous data, we also consider two covariance matrix structures: "simple" and "compound symmetry". Being generated from an aggregation of 10 simulations, these sim-

Table 2: Direct likelihood with MI, simple cov. struct.

	Mean imputation	Random forest	Bayesian linear regression	Predictive mean matching
Variable	Estimate (std.error)	Estimate (std.error)	Estimate (std.error)	Estimate (std.error)
(Intercept)	4.294 (0.582)***	-1.449 (1.533)	3.657 (2.806)	4.378 (2.795)
age	0.070 (0.024)**	0.138 (0.064)*	-0.164 (0.119)	-0.190 (0.118)
TIME	0.268 (0.011)***	0.575 (0.025)***	0.319 (0.018)***	0.305 (0.017)***
learning	-0.516 (0.123)***	1.047 (0.243)***	0.884 (0.184)***	0.764 (0.181)***
I(age^2)	0.000 (0.000)	0.000 (0.001)	0.004 (0.001)***	0.004 (0.001)***
age:TIME	-0.005 (0.000)***	-0.009 (0.000)***	0.000 (0.000)	0.001 (0.000)*

Table 3: Direct likelihood with MI, compound symmetry cov. struct.

	Mean imputation	Bayesian linear regression	Predictive mean matching
Variable	Estimate (std.error)	Estimate (std.error)	Estimate (std.error)
(Intercept)	4.294 (0.582)***	3.657 (2.806)	4.378 (2.795)
age	0.070 (0.024)**	-0.164 (0.119)	-0.190 (0.118)
TIME	0.268 (0.011)***	0.319 (0.018)***	0.305 (0.017)***
learning	-0.516 (0.123)***	0.884 (0.184)***	0.764 (0.181)***
I(age^2)	0.000 (0.000)	0.004 (0.001)***	0.004 (0.001)***
age:TIME	-0.005 (0.000)***	0.000 (0.000)	0.001 (0.000)*

ulated data points are then used in conjunction with the real data to draft appropriate models. Their summary statistics are displayed in Table 2 and 3.

Respective to the different covariance structures, parameter estimates are largely the same between model pairs of opposing covariance patterns. For this reason, the following analysis will only consider the models fit according to a simple covariance structure. Comparing model parameter estimates shows that the models do differ in regards to how the fixed effects regression coefficients are computed; with no two models agreeing perfectly on every parameter estimate in regards to sign or magnitude. As these are mixed models, interpreting the fixed effects without consideration of the random effects is inappropriate. Additionally, as data sets are non-standard amongst each of these models, many traditional metrics such as AIC and BIC would also be inappropriate. Instead we shall focus on predictive metrics as these should inform us of each of the models potential practical implications.

Table 4: A comparison of direct likelihood and MI methods

Method	Min SR	Max SR	SR Int. Length
Predictive mean matching	-4.711	7.049	11.761
Bayesian linear regression	-5.041	13.342	18.384
Unconditional mean imputation	-4.518	6.766	11.284
Random forests	-6.766	5.563	12.329
Direct likelihood	-4.039	6.709	10.748

3.1.2 Direct Likelihood vs. Multiple Imputation

To cross evaluate these models and the model drafted from a direct likelihood approach, we shall consider an analysis of the empirically observed within-group standardized residuals. For the models drafted through multiple imputation, to minimize potential bias inflation from the imputation processes, we shall consider a worst case scenario; where these residuals are assumed to rest within the largest possible observed interval between the 10 models (Table 4).

From this metric, we observe that the direct likelihood (DL) method results in the smallest interval of empirically observed within-group standardized residuals, with unconditional mean imputation (Mean) potentially resulting in largest degree of error. As will be discussed in further sections, the impact of predictive bias will be further explored to generate a much more thorough understanding of how well these models may preform.

3.2 Weighted generalized estimating equations versus ‘multiple-imputation generalized estimating equations’

Using frequentist methods, Q5

3.3 Sensitivity analysis

It is likely that different assumption on the missingness mechanism will lead to different conclusions, especially if the dropout is higher than 10% (Li et al. 2017).

Bibliography

- Beunckens, Caroline, Geert Molenberghs, and Michael G Kenward. 2005. “Direct Likelihood Analysis Versus Simple Forms of Imputation for Missing Data in Randomized Clinical Trials.” *Clinical Trials (London, England)* 2 (5): 379–86.
- Brant, Larry J., and James L. Fozard. 1990. “Age Changes in Pure-tone Hearing Thresholds in a Longitudinal Study of Normal Human Aging.” *The Journal of the Acoustical Society of America* 88 (2): 813–20. <https://doi.org/10.1121/1.399731>.
- Buuren, Stef van, and Karin Groothuis-Oudshoorn. 2011. “Mice: Multivariate Imputation by Chained Equations in r.” *Journal of Statistical Software* 45 (3): 1–67. <https://doi.org/10.>

18637/jss.v045.i03.

Li, Meijuan, Nan Chen, Yang Cui, and Hongyun Liu. 2017. "Comparison of Different LGM-Based Methods with MAR and MNAR Dropout Data." *Frontiers in Psychology* 8. <https://doi.org/10.3389/fpsyg.2017.00722>.