# Multivariate statistics: Assignment 1

**Team B:** **Kendall Brown** *r0773111* **Raïsa Carmen** *s0204278*
**Stefan Velev** *r0924289* **Adhithya Unni Narayanan**

## Abstract

## 1 Introduction and data exploration

This report assesses hearing thresholds for a sample of 546 healthy male volunteers. The subjects were 52 years old on average at the start of the study and are followed for an average of 7.57 years. The hearing threshold is measured, on average, every 1.59 years. Table 1 describes the demographics in more detail. It can be seen that the data is highly unbalanced; there is a lot of variation in the time a volunteer is followed and in the number of times a volunteer's hearing threshold is measured. Normally, each ear is measured at each visit but this only happened in 93.22% of all visits. The left (right) ear was tested in 96.72% (96.5%) of all visits.

Previous research on the hearing data of males in the Baltimore Longitudinal Study of Aging (BLSA) showed a change in hearing threshold for all age groups but especially the older population (Brant and Fozard 1990). In contrast to our dataset, Pearson et al. (1995) and Morrell et al. (1997) consider females in their study. They similarly found a decrease in hearing sensitivity for all ages at 500Hz and included a quadratic function of age to predict the hearing threshold, as in Verbeke, Spiessens, and Lesaffre (2001). Additionally, a statistically significant learning effect from the first visit to subsequent visits was found (Morrell et al. 1997; Verbeke, Spiessens, and Lesaffre 2001). Morrell et al. (1997) observed that hearing levels are slightly poorer on average on the left compared to the right ear and that the variance in the hearing threshold is higher for people with a higher age.

Figures 1 and 2 show the trends in the hearing threshold for all volunteers over time. These figures shows that many volunteers' hearing threshold over time have an erratic pattern meaning there is likely high variability **within subjects**. Additionally, variability **between subjects** is also high, especially for older volunteers.

### 1.1 Mean, variance and correlation structure

Figure 3 shows the mean and 95% confidence interval for the hearing threshold (dB) for different age groups. The thresholds for the age groups are chosen such that each group is approximately the same size (between 141 and 149 measurements in each group). The graph shows that the variance increases with age and there doesn't seem to be a significant difference, on average, between left and right ears.

Figure 4 shows Loess smoothing curves by age category and side of the ear. The trend in the average is slightly increasing for most groups. For the oldest group, there is a more pronounced

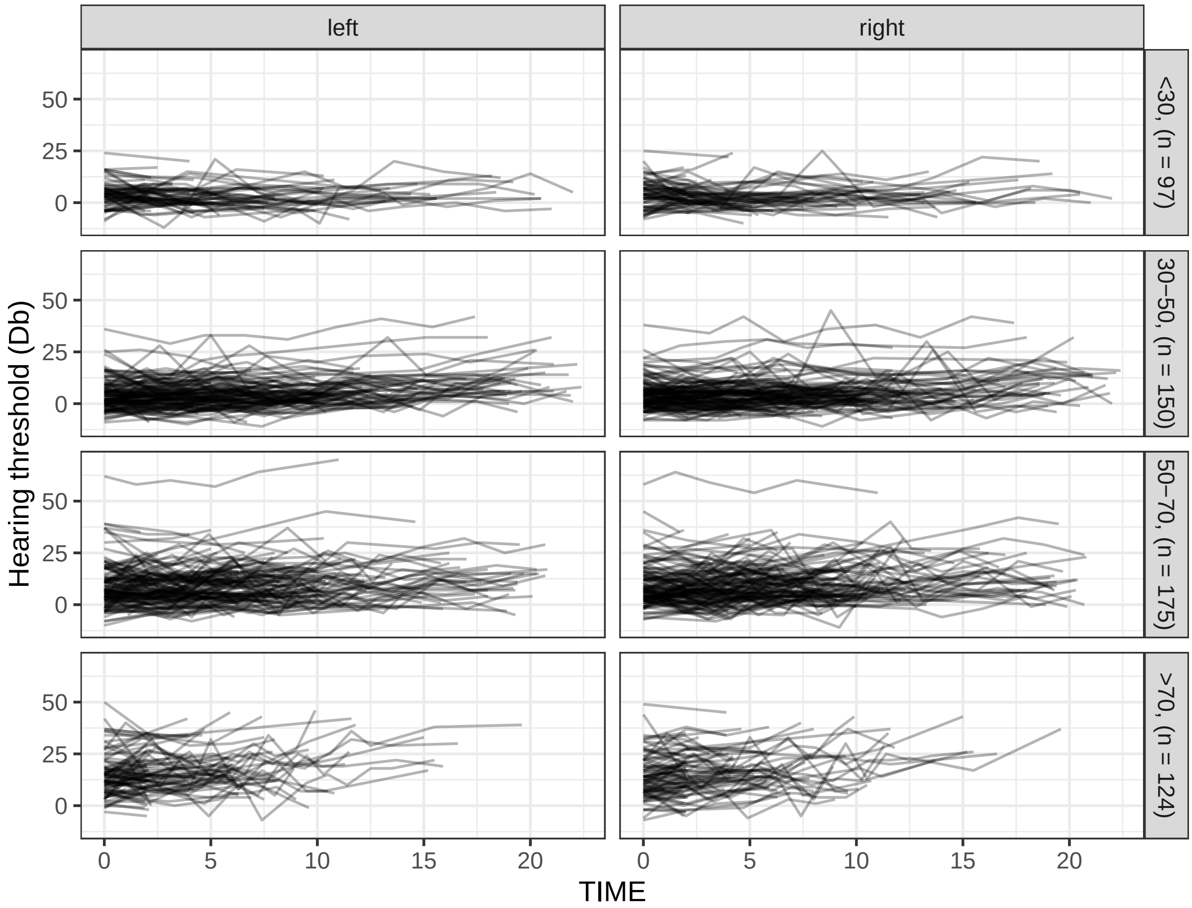



Figure 1: Hearing threshold over time, divided by left and right ear and by age group at the start of the study

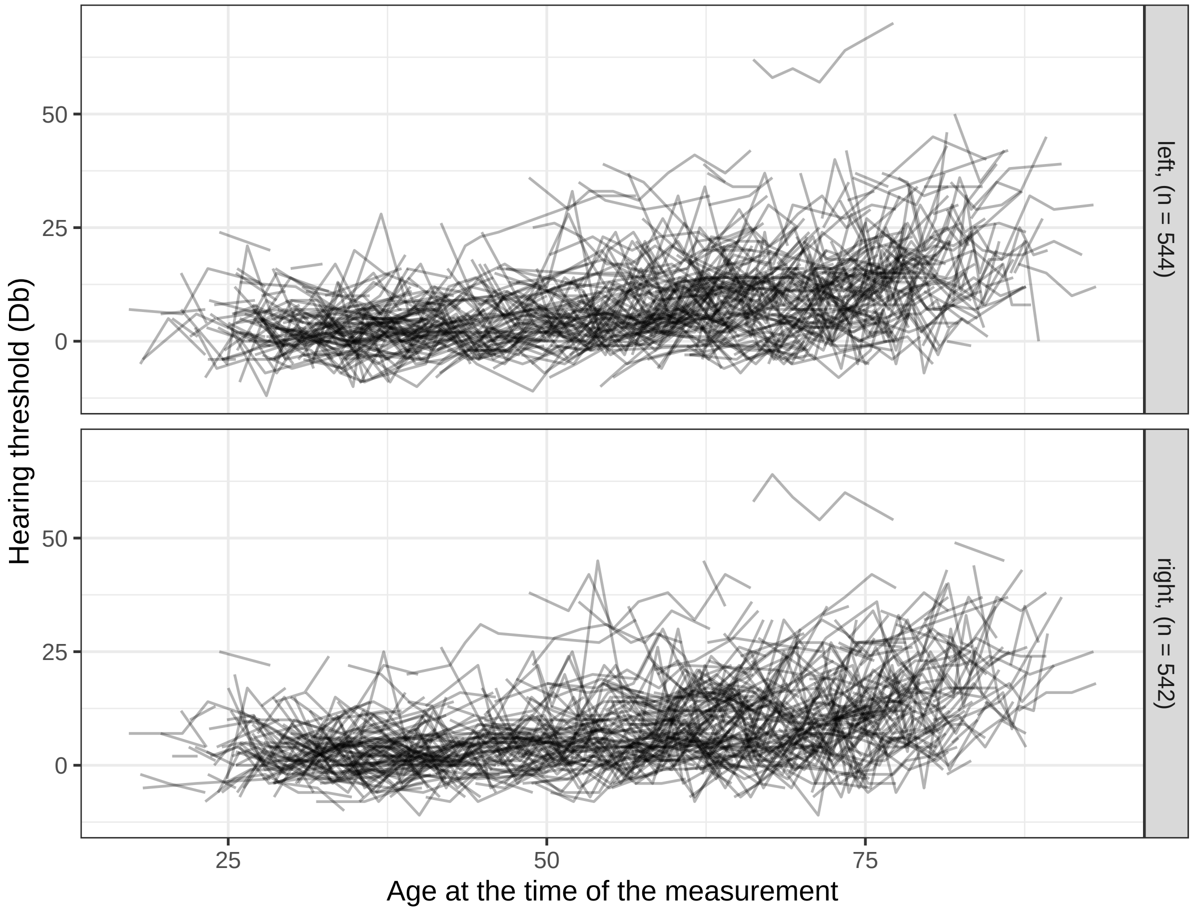



Figure 2: Hearing threshold over time, divided by left and right ear.

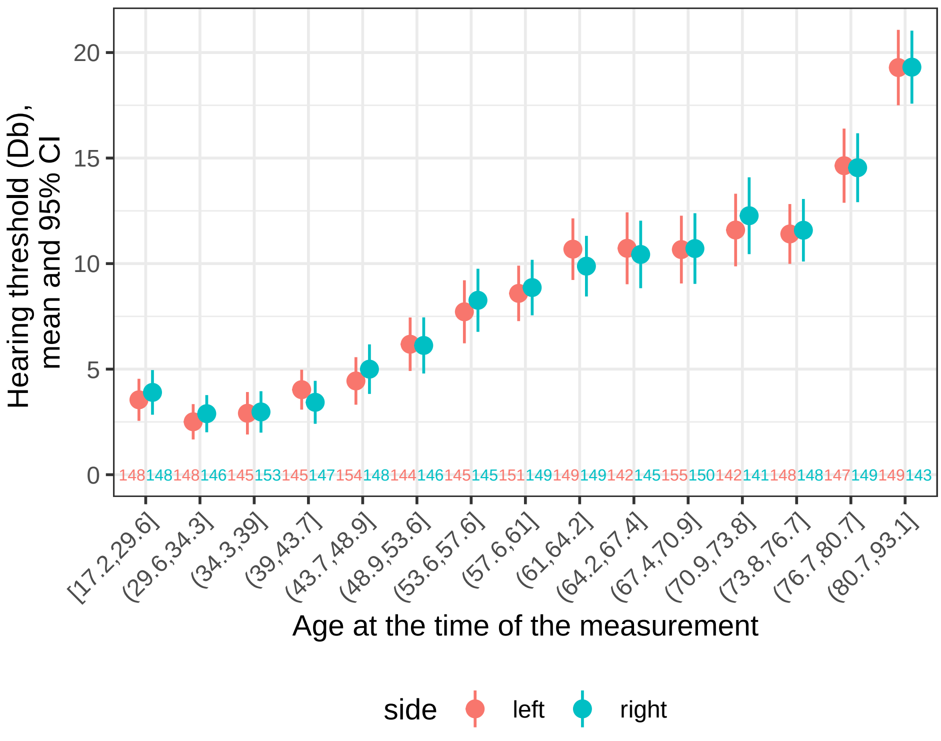



Figure 3: Hearing threshold over time, divided by left and right ear. The numbers in the bottom show the number of measurements that were taken.

increase in the hearing threshold. Additionally, the estimated standard errors around the mean are larger as time increases, likely due to dropout. Based on the estimated averages from 4, squared residuals can be calculated. Figure 5 shows the Loess smoothed variance function for each group. Overall, the variance seems quite constant over time and it is higher for older subjects. The downward trend at time $> 15$ for the oldest age group can possibly not be trusted due to limited data.

The data has a hierarchical (grouped) structure since each individual ($id$) is measured several times over the years. Some basic descriptive statistics were derived using the *statsBy* function from the psych package. The intraclass correlation (percentage of variance due to groups) is 0.69 for the hearing threshold ($y$) and, as expected, very high (0.93) for age at the time of measurement.

The total correlation between y and the age at the measurement, ignoring the hierarchical structure, is 0.45. Following Marzban et al. (2013) and Mongomery (2017), the correlation matrix for each subject that has at least 2 measurements for both ears is calculated, one can obtain within-group correlations and variances. Figure 6 shows the histograms over all individuals for the correlation between the age at the measurement and y, and the variances for both age and y (though the variance in age is not very informative). The mean or median of all within-group correlations is often used as a measure of within-group correlation. The mean (0.15) and median (0.23) of the within-group correlations are indicated in red and blue respectively on the graph. It can be seen that within-group correlation spans the entire range from -1 to 1, meaning that age has a strong positive relationship to the hearing threshold for some and a strong negative relationship for others. Age will be a good predictor for hearing threshold for some but not

Table 1: Demographics for all respondents

| | |
|---|---|
| **Age at the beginning of the study** | |
| min | 17.20 |
| max | 87.00 |
| median | 54.10 |
| mean (sd) | $51.99 \pm 18.70$ |
| **Years of follow-up** | |
| min | 0.00 |
| max | 22.40 |
| median | 6.30 |
| mean (sd) | $7.57 \pm 6.30$ |
| **Number of visits** | |
| min | 1 |
| max | 15 |
| median | 3.00 |
| mean (sd) | $4.19 \pm 2.88$ |

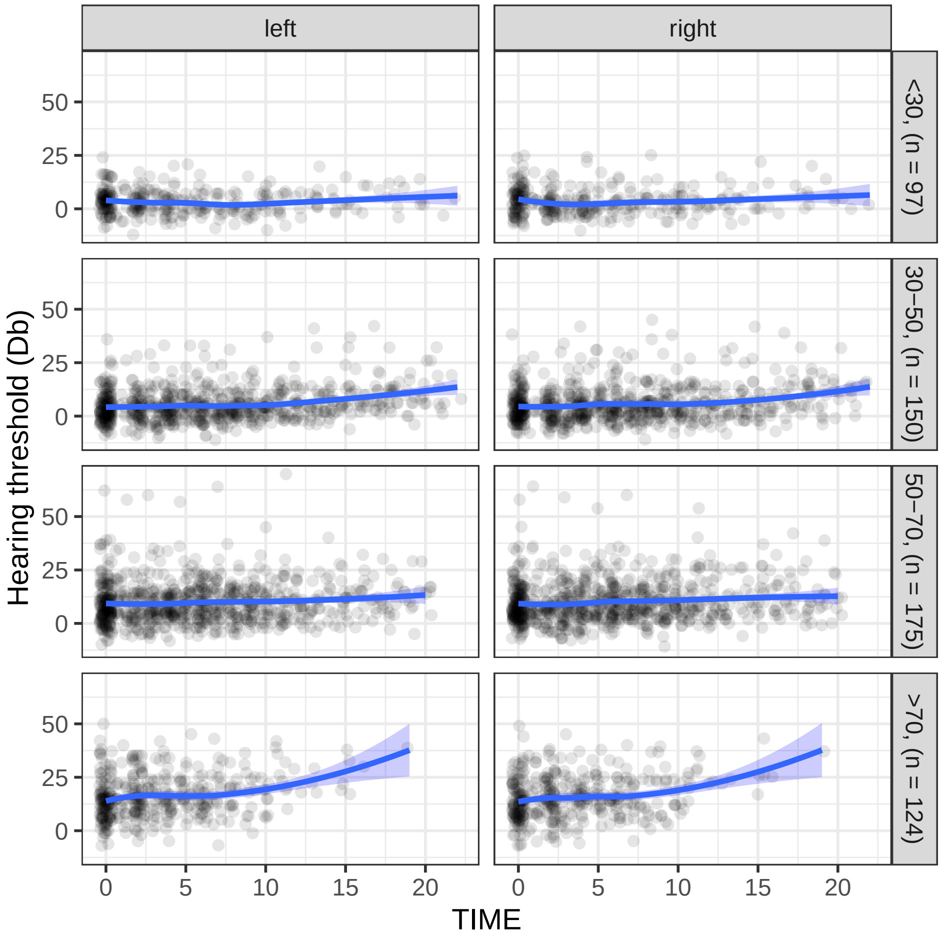



Figure 4: Loess smoothing on the hearing threshold since start of the study, divided by left and right ear and age group at the start of the study.

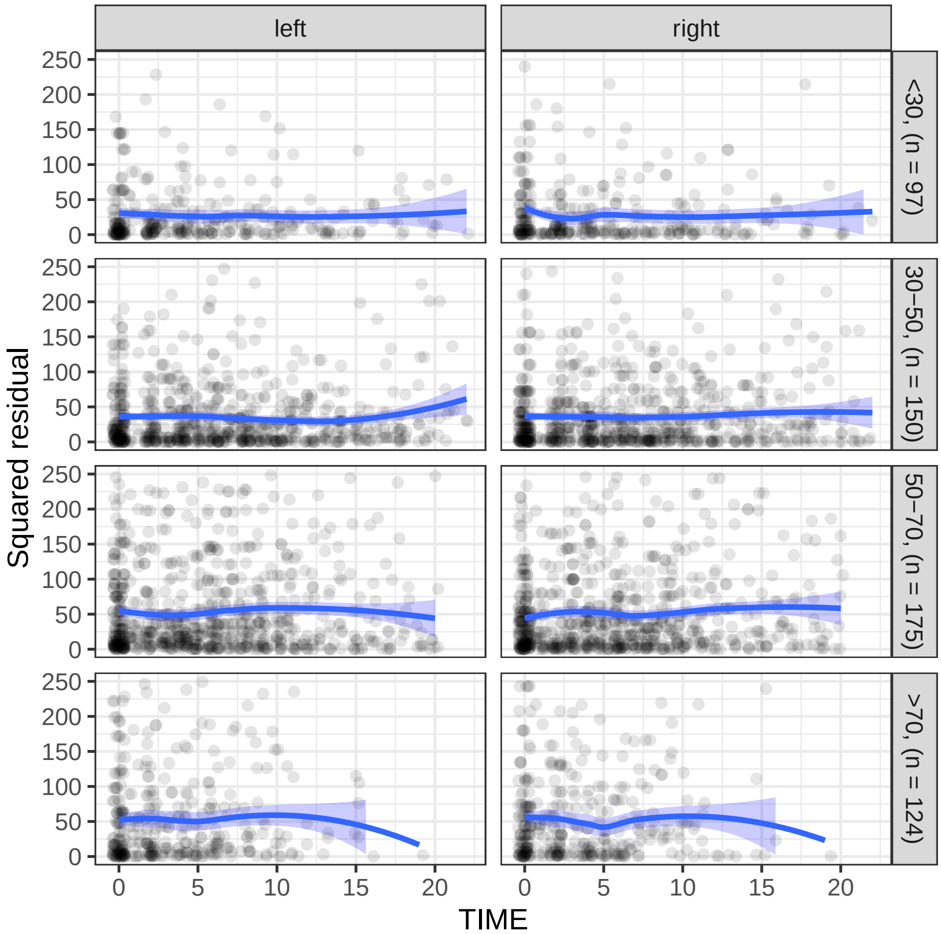



Figure 5: Loess smoothed variance in the hearing threshold since start of the study, divided by left and right ear and age group at the start of the study.

for others. Between-group correlation is obtained by averaging the age at measurement and the hearing threshold for each group (see scatterplot in Figure 6) and then computing the correlation across the groups, obtaining 0.54.

Panel B in Figure 6 shows that the correlation between the side of the ear that is measured and the hearing threshold ($y$) is high for some but small for most individuals. Panel C in Figure 6 shows that the variance in the hearing threshold is usually less then 50 but can be very large for some individuals.

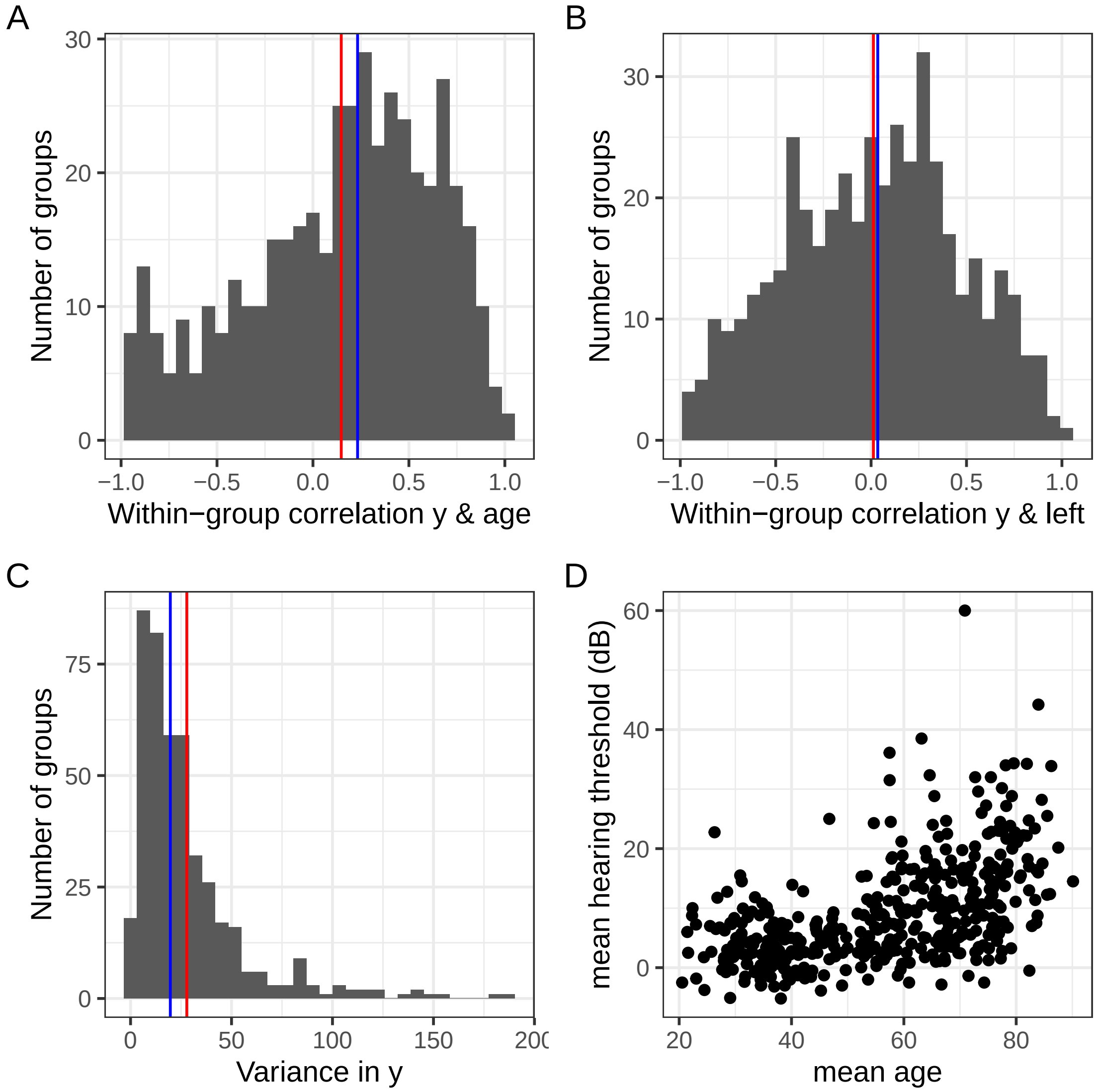
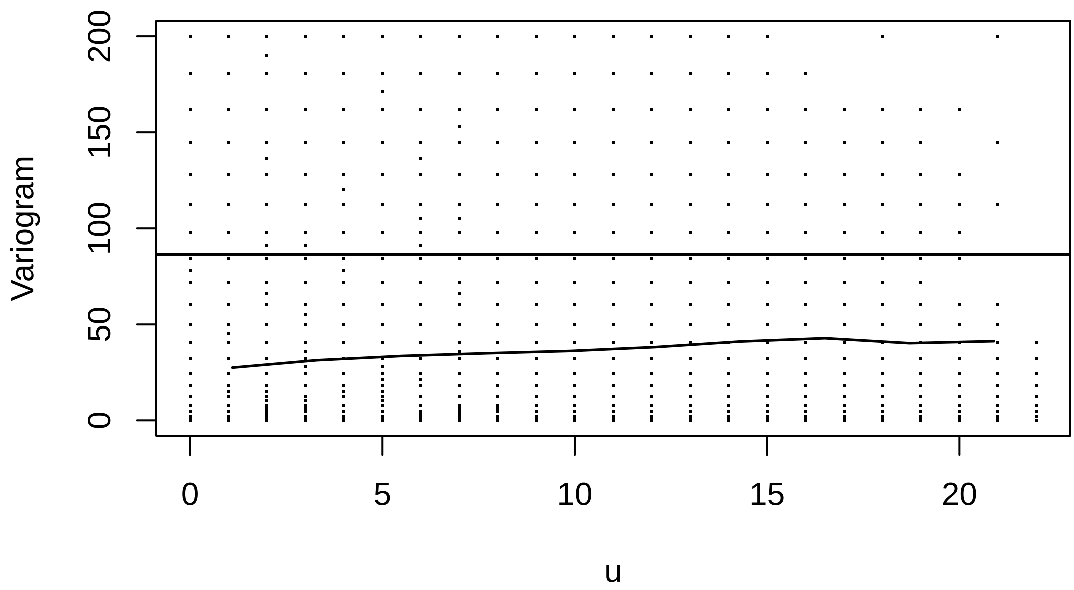



Figure 6: The mean is indicated in red, the median in blue




Lastly, a semi-variogram is constructed using the *joineR* package. The total variance is estimated to be 86.42. Measurement error is approximately 22.29, the serial correlation component is approximately 19.18 and the between-subject variability is thus 44.95.

# 2 Methodology

In this section, we explore a couple of different methods to analyze the data. All analysis was carried out with the statistical software R. All scripts are freely available at this git repository.

## 2.1 Summary statistics

One possibility to deal with the hierarchical structure of the data is to summarize the data and reduce the number of measurements per subject to one.

A simple paired-t test can check whether there is a significant difference between the left and right ear. For this test, the difference $\Delta_i^1 t$ is calculated between the hearing threshold of the left and right ear for each subject $i$ and at each time instance $t$ that both ears were measured (equation (1)).

$$\Delta_{it}^1 = Y_{ilt} - Y_{irt} \tag{1}$$

If the focus is on the change in hearing threshold over time, we can, for instance, calculate the average change in the hearing threshold for each subject $i$, $\Delta_i^2$ as in equation (2) where $Y_{ijk}$ is the $k^{th}$ measurement for ear $j \in \{l, r\}$ of subject $i$ and $n_{ij}$ is the total number of measurements of ear $j$ of subject $i$.

$$\Delta_i^2 = \frac{(Y_{iln_{il}} - Y_{il1}) + (Y_{irn_{ir}} - Y_{ir1})}{2} \tag{2}$$

Alternatively, if the difference between the left and right ear is an important effect to analyse, equation (3) shows an alternative with one summary measure for each ear $j$ of subject $i$ with at least two measurements over time. $n_{ij}$ is the total number of measurements of ear $j$ of subject $i$.

$$\Delta_{ij}^3 = Y_{ijn_{ij}} - Y_{ij1} \tag{3}$$

Equation (2) and (3) will allow us to see how the *evolution* of the hearing threshold differs for different age and side of the ear but will it not tell us anything about the *level* of the hearing threshold. To test whether the hearing threshold depends on age and side of the ear, equation (4) introduces $\Delta_{ij}^4$ as the average measured hearing threshold for ear $j$ of subject $i$. $n_{ij}$ is the total number of measurements that were recorded for ear $j$ of subject $i$.

$$\Delta_{ij}^4 = \frac{\sum_{t=1}^{t=n_{ij}} Y_{ijt}}{n_{ij}} \tag{4}$$

A first, obvious problem with this method is that it does not allow us to test all hypotheses of interest in one model. Additionally, a lot of information is lost:

- Only subjects with multiple measures on each ear are included in the summary measure of equation (2)
- Only ears that have been measured more than once are included in the summary measures of equation (3)
- Only subjects that were measured on both ears at a certain time instance are included in the summary measures of equation (1)
- While the summary measures of equation (2) and (3) can tell us something about the evolution in the long run because it focuses on the first and last measurement, all information on the trajectory between the first and last measurement is lost.
- Since the within-subject variability is rather large, relying on the first and last measurement as in equation (2) and (3) is risky.
- Previous literature found the hearing threshold was significantly higher for the first measurement (learning effect). Relying on the first measurement as in (2) and (3) may be worrisome.

To examine if the periodic shift in hearing capacity is affected by the patient's initial age and their time within the study, we examine their individual observed median absolute deviations (MAD) and their incremental shifts in hearing capability per year (SPY).

$$MAD = MEDIAN(|Y_{ijk} - \overline{Y_{ij}}|) \tag{5}$$

$$SPY = \frac{Y_{ijk_{max}} - Y_{ij1}}{TrialDuration} \tag{6}$$

The goal here is to see if patients of an older age or patients with a longer study period have larger degrees of variability in their hearing capacity. Examination of this relation could pose to provide insight into the nature of how hearing capability evolves over time. A positive relation between these metrics and a patients' starting age, in addition to their time spent in the study, would imply that hearing capabilities of older people have a more pronounced evolution compared to those of younger people. These variability measurements were taken in respect to both the left and right ears and then analyzed via an ANCOVA model.

### 2.2 Multivariate model

A multivariate model is constructed and we find the most parsimonious mean structure for it. Given the unbalanced data set, the *gee* and *geepack* packages are used to conduct Generalized Estimating Equations. We will conduct these under the strict assumptions of normality. We will compare models using Quasi-Information Criterium (QIC), the RMSE of the fitted residuals, and `ANOVA\MANOVA` tests. After some model searching in order to select the most parsimonious mean structure we will take care of selecting the appropriate covariance structure.

### 2.3 Two-stage analysis

In a two-stage analysis, subject-specific intercepts and time effects are first estimated. In the second step, these subject-specific parameters are analysed and related to additional covariates such as the age or the side of the ear.

### 2.4 Random-effects model

Lastly, a random-effects model is fit. This model is similar to the two-stage model but both stages are now combined in one model. We use the *glmmTMB* / *lme4* ????which one???? packages in R to fit the models.

## 3 Results

### 3.1 Summary statistics

First, a two-sided paired t-test is done, comparing the hearing threshold between the left and right ear. With a p-value of 0.44, there doesn't seem to be a significant difference between the hearing threshold for the left and right ear. Simple linear regression models that relate the age (on a continuous or discrete scale) to $\Delta^1_{it}$ (equation (1)) also did not find any significant effects.

Next, regression models are made for $\Delta^2$ and $\Delta^3$. The variables of interest are age at the start of the study (a model with a continuous and discrete version is tested), the side of the ear (can only be tested for $\Delta^3$), and the time difference between the first and last measurement (see

Table 2: Overview of the regression results for summary statistics

| Variable | $\Delta^3_{continuous}$ | $\Delta^3_{discrete}$ | $\Delta^2_{continuous}$ | $\Delta^2_{discrete}$ |
|---|---|---|---|---|
| (Intercept) | -3.422 *** | -3.074 *** | -3.65 *** | -3.344 *** |
| age | 0.089 *** | | 0.089 *** | |
| sideright | -0.197 | -0.195 | | |
| timediff | 0.32 *** | 0.341 *** | 0.332 *** | 0.352 *** |
| age30-50 | | 1.126 | | 1.173 |
| age50-70 | | 3.248 *** | | 3.275 *** |
| age>70 | | 5.338 *** | | 5.375 *** |
| Rsquared | 0.085 | 0.098 | 0.114 | 0.132 |

Table 3: Overview of the regression results for summary statistics

| Variable | $\Delta^4_{continuous}$ | $\Delta^4_{discrete}$ |
|---|---|---|
| (Intercept) | -5.858 *** | 3.417 *** |
| age | 0.254 *** | |
| sideright | 0.071 | 0.049 |
| age30-50 | | 0.111 |
| age50-70 | | 5.841 *** |
| age>70 | | 10.558 *** |
| Rsquared | 0.242 | 0.226 |

Table 2). The intercept is always negative meaning the first measurement is higher than the last measurement in young subjects. This is unexpected and may be due to the learning effect that has been described in previous literature. Both age and the time between the first and last measurement have significantly positive coefficients meaning the difference between the first and last measurement is larger and may even become positive for older patients with long time between the first and last measurement. There is no significant difference between the left and right ear.

Table 3 shows the results for two regression models based on the summary statistic of equation (4). The age in this model refers to the age of the subject in the middle of his follow-up time (i.e. $age + \frac{max(TIME)}{2}$). The intercept now represents the expected hearing threshold for the left ear of a subject that is 0 years on (in the continuous age model) or a subject that is in the youngest age category ($<$ 30 years old). The older the subject, the higher the expected hearing threshold. There is no significant difference between the left and right ear.

Table 4 details the results of the ANCOVA where the variance of the hearing capabilities. In agreement with the previously derived summary statistics, the starting age of the patient as well as their time spent within the study both have an effect on the MAD and SPY statistics. In further

Table 4: Overview of the regression results for SPY and MAD summary statistics

| | MAD | | SPY | |
|---|---|---|---|---|
| Variable | Ancova Sum Sq.(Df) | lm coef | Ancova Sum Sq.(Df) | lm coef |
| (Intercept) | 200.73***(1) | 1.764*** | 41.299***(1) | -0.8*** |
| StartingAge | 174.406***(1) | 0.024*** | 68.276***(1) | 0.015*** |
| duration | 258.225***(1) | 0.093*** | 19.857**(1) | 0.026** |
| side | 17.464 (1) | -0.278 | 0.276 (1) | -0.035 |
| Residuals | 6003.637(903) | | 2193.765(903) | |

agreement with past results the side of the ear remains unimportant when inferring the expected periodic change of subjects' hearing capabilities. From these results it can be stated that both age and the time spent within the study have a positive effect on the observed variance of the measured response. As was in the case with previous analyses, these results imply a potentially large shift in hearing capabilities over time. As was the case with previous analyses, their does not appear to be a significant difference between the two ears.

## 3.2 Multivariate model

We find that the model $Y_{ijk} = age_{ij} + age_{ij}^2 + learning_{ijk} + TIME_{ijk}$ is the most parsimonious model. We further find that the effect of the *side* variable does not significantly influence the hearing threshold. These conclusions hold for a much wider selection of models (we present here only sample of what we tested) or even when we filter the data under different conditions (Removing subjects with less than $n = \{2, 4, 5, 6\}$ observations,looking at only the right/left ear, treating each ear-subject pair as single subject, balancing the data, grouping the subjects in cohorts of $\{1, 5, 10, 15, 20\}$ years and creating dummy variables for them).

Now we will take a look at alternative ways to optimize our models. The covariance structure of the model denotes the observed relation between the model variables. Certain structures may lead to a stronger model fit, whereas others are inappropriate given the nature of our data. We will compare the same model mean structure i.e. $Y_{ijk} = age_i + TIME_{jk}$ under different covariance structures and evaluate them according to Quasi Information Criterion (`QIC`) value. This is a performance metric which details the relative performance between a set of models.

It appears that a covariance structure that assumes independence across the variables produces the lowest `QIC` value, nevertheless it behooves us to take a closer look at these two models. First let's consider the model with a covariance structure that assumes equal covariances between variables with equivalent variances (compound symmetry).

In the above output we notice that we get two standard errors: Naive and Robust. Normally we would want to use the Robust estimates because the variances of coefficient estimates tend to be too small when responses within subjects are correlated (Bilder and Loughin, 2015), however in this case there is little difference between the Naive and Robust estimates. This further sug-

Table 5: Comparison of multivariate models

| var | gee1 | gee2 | gee3 | gee4 | gee5 |
|---|---|---|---|---|---|
| (Intercept) | -4.735*** | 4.192 | 4.559. | -5.754*** | -4.232*** |
| age | 0.237*** | -0.175 | -0.169 | 0.258*** | 0.251*** |
| TIME | -0.048 | -0.08 | -0.111 | 0.199*** | |
| learning | | 0.99* | | | |
| I(age^2) | | 0.004** | 0.004** | | |
| age:TIME | 0.005 | 0.008* | 0.007* | | |
| sideright | 0.259 | | | | |
| age:sideright | -0.005 | | | | |
| TIME:sideright | 0.012 | | | | |
| QIC | 336295.14 | 331467.63 | 331953.69 | 337007.96 | 341929.28 |
| RMSE | 6.54 | 6.54 | 6.46 | 6.56 | 6.59 |

gests that the independence assumption of the correlation structure seems realistic (Hothorn and Everitt, 2014). In the bottom we see a `Working correlation` output which shows the upper $4 \times 4$ of the Variance-Covariance matrix used for this model (the size depends on the number of observations per individual with the largest beeing $29 \times 29$ ). Below we will report the best model we could find under the conditions of this task. It has the same parameter estimates as a least squares model with the same mean structure specification, however the robust errors in this model are more accurate, as some of the conditions of the least squares model are not fulfilled (e.g. Normally distributed residuals)

### 3.3 Two-stage analysis

As previous analyses showed that individuals may have a different time effect, the two-stage analysis first fits individual-specific intercept and time effect (slope). As all previous analyses did not uncover a significant effect of the side of the ear, it is no longer considered in this analysis.

The first stage analysis fits the linear model in equation (7); the variable *learning* is 1 if it's the first visit for the subject ($TIME == 0$). Next, the first stage results are used in the second stage (equation (8)) to relate the estimated subject-specific parameters to age.

$$Y_{ij} = \beta_{1i} + \beta_{2i}TIME_{ij} + \beta_{2i} * learning_{ij} + \varepsilon_{ij} \quad j = 1, ...., n_i \tag{7}$$

$$\begin{cases} \beta_{1i} = \beta_0 + \beta_1 age_i + \beta_2 age_i^2 + b_{1i} \\ \beta_{2i} = \beta_3 + \beta_4 age_i + \beta_5 age_i^2 + b_{2i} \end{cases} \tag{8}$$

Table 6 shows the results for the second stage. The first row shows estimates for $\beta_0$, $\beta_1$ and $\beta_2$.

Table 6: Second-stage results

| model | intercept | age | age^2 |
|---|---|---|---|
| Intercept | 7.081 . | -0.253 | 0.004 ** |
| TIME | -1.573 * | 0.059 . | 0 |
| learning | -2.153 | 0.082 | -0.001 |

The estimated hearing threshold on the first visit for a person that is 0 years old is 7.081 -2.153 = 4.927dB and increases exponentially for each life year $age^2$ is highly significant. The expected slope increases with age (0.059 (p-value of 0.054 )).

### 3.4 Random-effects model

## 4 Discussion and conclusion

### 4.1 Further research

The fact that the data set is unbalanced, complicates the analysis significantly. Inviting the subjects to the clinic for a hearing test at fixed time instances and consistently measuring both ears in each visit can alleviate these problems.

For future research, a different covariance structure for older/younger people.

## Bibliography


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.

Marzban, Caren, Paul R Illian, David Morison, and Pierre D Mourad. 2013. "Within-Group and Between-Group Correlation: Illustration on Non-Invasive Estimation of Intracranial Pressure." *Viewed Nd, from Http://Faculty. Washington. Edu/Marzban/Within_Between_simple. Pdf.*

Mongomery, DC. 2017. "Design and Analysis of Experiments." *John Willy & Sons*.

Morrell, Christopher H, Jay D Pearson, Larry J Brant, and Sandra Gordon-Salant. 1997. "Construction of Hearing Percentiles in Women with Non-Constant Variance from the Linear Mixed-Effects Model." *Statistics in Medicine* 16 (21): 2475–88.

Pearson, Jay D., Christopher H. Morrell, Sandra Gordon-Salant, Larry J. Brant, E. Jeffrey Metter, Lisa L. Klein, and James L. Fozard. 1995. "Gender Differences in a Longitudinal Study of Age-associated Hearing Loss." *The Journal of the Acoustical Society of America* 97 (2): 1196–1205. https://doi.org/10.1121/1.412231.

Verbeke, Geert, Bart Spiessens, and Emmanuel Lesaffre. 2001. "Conditional Linear Mixed Models." *The American Statistician* 55 (1): 25–34.