Multivariate statistics: Assignment 1

Team B: **Kendall Brown** *stnumber* **Raïsa Carmen** *s0204278*
Stefan Velev *stnumber* **Adhithya Unni Narayanan** *stnumber*
Audrey-Justine Towo Kamga *stnumber*

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 visits. 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.

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 has an erratic pattern meaning there is likely high variability **within subjects**. Additionally, variability between subjects is also high, especially for older volunteers.

Figure 3 shows the mean and 95% confidence interval for the hearing threshold (dB) for different age group. The age is the age at the time the measurement was taken. While the top graph shows 15 age groups with approximately equal sample size (between 141 and 149 measurements in each group), the bottom graph shows age groups with the same interval range of about 5 years. Both graphs show that the variance increases with age and there doesn't seem to be a significant difference, on average, between left and right ears.

Describe the data, and use graphical techniques to explore the mean structure, the variance structure and the correlation structure. Summarize your conclusions. What are the implications with respect to statistical modeling ?

2 Methodology

All analysis was carried out with the statistical software R. All scripts are freely available at this git repository.

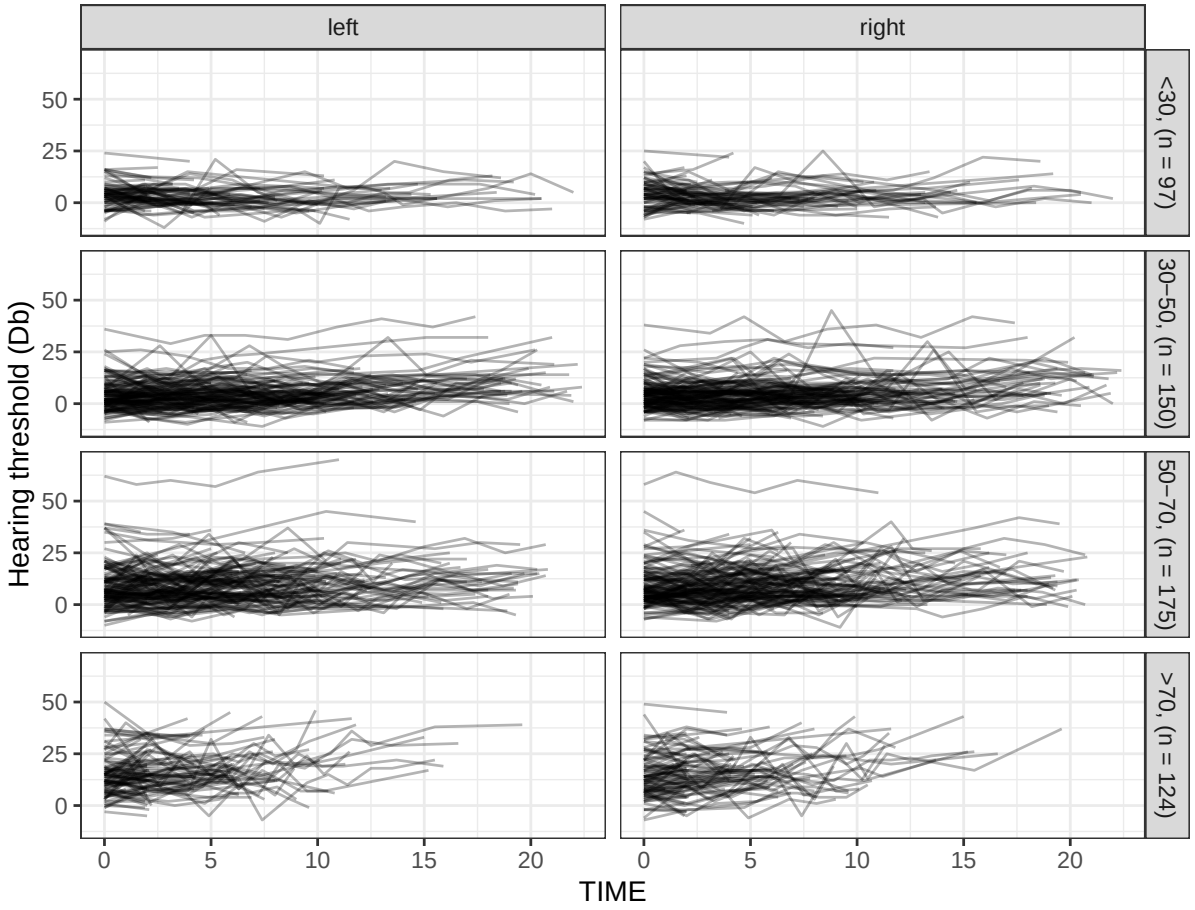

Figure 1: Hearing threshold over time, divided by left and right ear and by age group at the start of the study

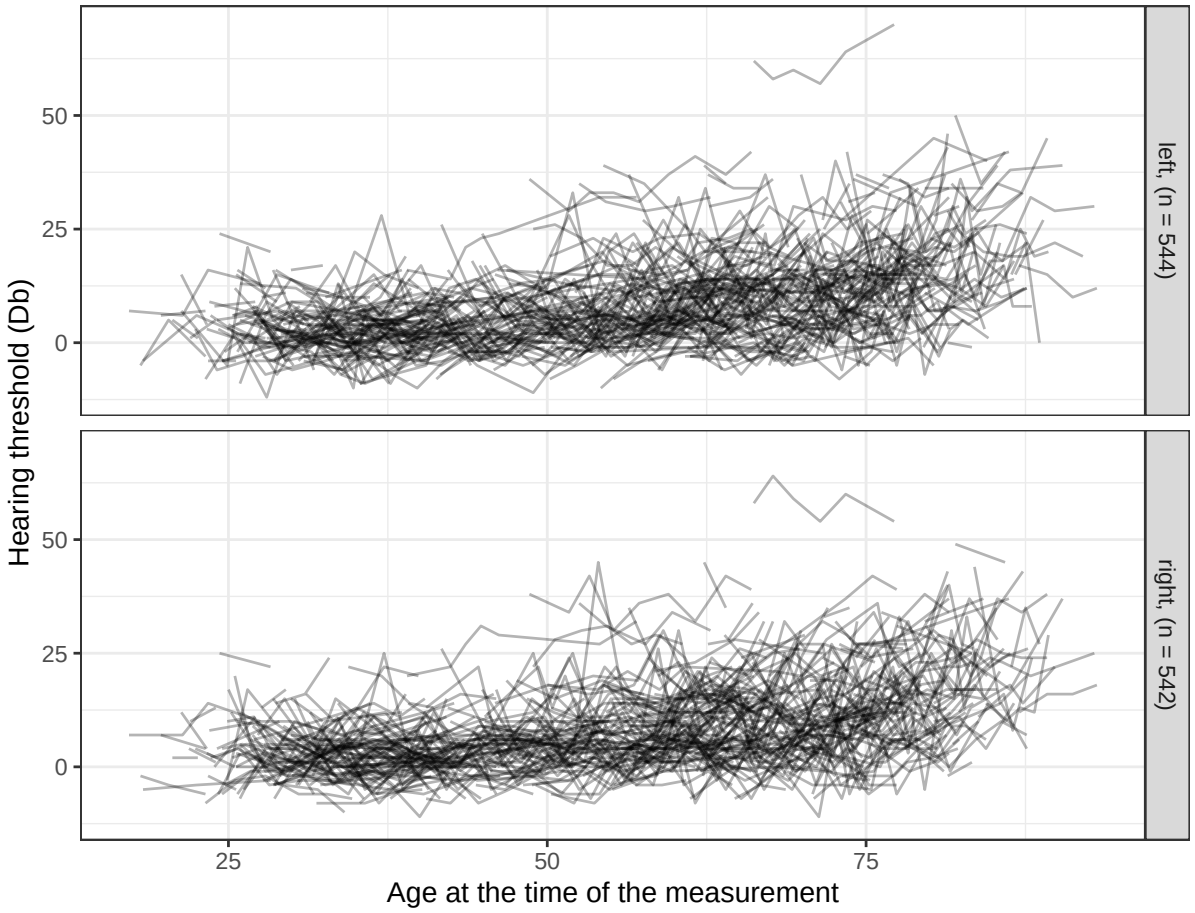

Figure 2: Hearing threshold over time, divided by left and right ear.

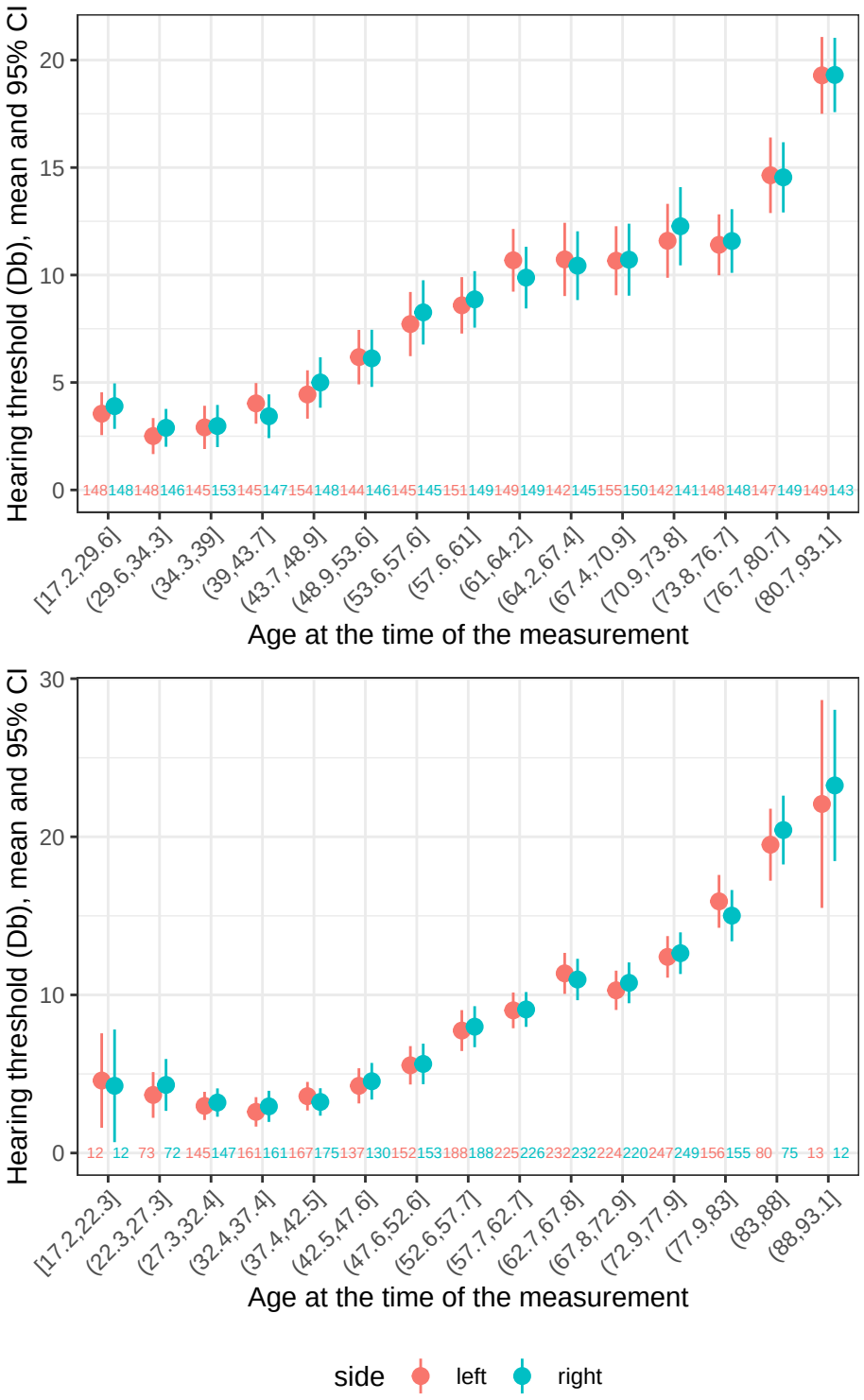

Figure 3: Hearing threshold over time, divided by left and right ear. numbers in the bottom show the number of measurements that were taken.

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