1. Conclusion

In this assignment we have analyzed the hearing data using a Generalized Estimating Equations approach with marginal log odds and a Cumulative Link Model with mixed effects. Then we have conducted a confirmatory analysis using an empirical Bayes methodology. Finally, we discussed the viability of transition models.

First, we had to trichonometrize the hearing data in order for it to be treated as an ordinal multinomial variable. We have selected our cut off points based expert evaluation and the existing medical literature used to diagnose hearing loss and have decided on the following values. If the hearing threshold is less than or equal to 6dB then the subject is classified as having Excellent hearing, if the threshold is between 6dB and 25dB then they are classified as having Normal hearing and if the threshold is more than 25 dB they are classified as having hearing loss. The majority of the subjects in dataset are healthy and only around 6% suffer from hearing loss.

In the next step of the analysis we evaluated the data using a Generalized Estimating Equations approach. There we see that the main drivers of hearing loss are the actual persons age, with a small additional learning effect from the first time they take the evaluation. We observe that as people age they are more likely to transition from having excellent hearing to having normal hearing, while the probability of hearing loss becomes tangible only after the age of 50. The learning effect constitutes a slight increase in probability of the subjects to be classified into a lower category if it is their first time being measured.

On the flipside we also conduct a Cumulative Link Model with mixed effects. This type of model has the advantage that is can be considered in a sense as continuous model where the different ordinal categories are cut of points of a latent continuous variable. Model convergence sometimes becomes an issue with these models. That is a logit-link model with random effects will create a mixed likelihood that combines the Normal distribution of the random effect with the Logistic distribution assumed for the latent responses. We end up with conditional probabilities for each subject, and come to the same conclusion as before, that is younger people are more likely to have Excellent hearing, with the probability of having Normal increasing as the get older, and the probability of hearing loss becoming tangible after 50 years of age.

It is clear that the two models show difference in the scale of their coefficients, i.e. the marginal probability coming from the GEE model and the conditional probability coming from the CLMM model. These differences arise from the fact that the mean of a nonlinear function of a random variable does not equal the nonlinear function of the mean. When we compare the subject specific curves to the population averaged curves it is clear that the slopes of the former are steeper than those of the latter. This characterizes further that the CLMM model results describe changes in the odds for an individual from the population, while the GEE model describes the population odds.

In the next step we analyse the data using an empirical Bayes approach and the coefficients from the CLMM model

Finally, we consider transition models. Transition models represent a natural way model categorical data, while both marginal and mixed effects model do so in a circumspect way. In transition models we consider the longitudinal data as a stochastic process with the appropriate Markov assumption, and from there we can use standard likelihood methodology to characterize the transition probabilities. Variates can be accommodated using (e.g.) a (multinomial-)logistic regression