Predicting Hearing Loss
Using Auditory Steady-State Responses

By

Yiwen Li

A Project

Submitted to the Faculty

of

Worcester Polytechnic Institute

In partial fulfillment of the requirement for the

Degree of Master of Science

in

Applied Statistics

_____

Dec 2008

APPROVED:

_____

Dr. Joseph D. Petruccelli, Project Advisor

_____

Dr. Bogdan Vernescu, Department Head

Abstract

Auditory Steady-State Response (ASSR) is a promising tool for detecting hearing loss. In this project, we analyzed hearing threshold data obtained from two ASSR methods and a gold standard, pure tone audiometry, applied to both normal and hearing-impaired subjects. We constructed a repeated measures linear model to identify factors that show significant differences in the mean response. The analysis shows that there are significant differences due to hearing status (normal or impaired) and ASSR method, and that there is a significant interaction between hearing status and test signal frequency. The second task of this project was to predict the PTA threshold (gold standard) from the ASSR-A and ASSR-B thresholds separately at each frequency, in order to measure how accurate the ASSR measurements are and to obtain a "correction function" to correct the bias in the ASSR measurements. We used two approaches. In the first, we modeled the relation of the PTA responses to the ASSR values for the two hearing status groups as a mixture model and tried two prediction methods. The mixture modeling was successful, but the predictions gave disappointing results. A second approach, using logistic regression to predict group membership based on ASSR value and then using those predictions to obtain a predictor of the PTA value, gave successful results.

Acknowledgement

I would like to express my deep gratitude to the following people for their academic and spiritual support in finishing this project and my master program study.

My advisor: Dr. Joseph D. Petruccelli .Without his guidance and help, there would not have been such a piece of work.

Professors in math department : Dr. Balgobin Nandram, Dr. Ryung S Kim, Dr. Jayson D. Wilbur, Dr. Marcel Blais, Dr. Hasanjan Sayit, Dr. Volodymyr Hrynkiv, Dr. Bogdan M. Vernescu, Dr. Suzanne L.Weekes.

Staff in math department: Ellen Mackin, Deborah M. K. Riel, Rhonda Podell, Mike Malone. They are so nice, patient and helpful.

My friends in America: Mei Jin, Dayang Liu, Yi Li, Zhipeng Lin, Dilli Raj Bhatta, Budhinath Padhy, Suwodi Dutta Bordoloi, Shiquan He, Ya Zhang, Ma. Criselda Toto, Jonathan Legare, Kristen Elaine Pedersen, Qin Huang, Su Chen, Li Zhu, Yinzhuo Fu, Shuai Shen, Yurong Mao, Xiao Zhong, Jonathan Adler, Hai Ling.

My parents in China and relatives in America: Jinxiang Li, Lihua Duan, Maozheng Dai, Yun Li, Hailei Dai, Bin Gui, Xiang Zhang, Li Jin.

Contents

Introduction

Auditory-evoked potentials (AEPs) are small electrical potential originating from the brain in response to an auditory stimulus such as different tones or speech sounds. AEPs are typically recorded using sensors placed on the scalp. In clinical practice, they are used to evaluate the hearing of human subjects.

There are two primary groups of patients who benefit from AEP testing: subjects with suspected neural problems and patients for whom accurate behavioral evaluation of hearing sensitivity is not possible. The second group is principally composed of infants; subjects who cannot be tested behaviorally for associated problems, and subjects who are suspected of exaggerating subjective audiometric thresholds (that is, the lower limit of the perception of the stimulus).

Among the different kinds of AEPs, auditory brainstem responses (ABRs) evoked by clicks are the most utilized in clinical practice, thanks to the high reproducibility and stability of the waveform. But since the early applications of click ABR, it was realized that the test couldn’t provide frequency-specific information since this signal has little frequency selectivity.

A newer alternative, auditory steady-state-evoked response (ASSR), has the potential to estimate the audiogram more efficiently than ABR. ASSRs are responses to single continuous tones modulated in amplitude (AMT) at rates between 75 and 110 Hz. This response, due to the synchronous discharge of auditory neurons in the brain stem, is periodic and phase locked to the modulation frequency of the carrier stimulus; it can be represented best in the frequency domain and not in the time domain like the other potentials.

Because they are a more recent development, the performance of ASSRs in detecting hearing thresholds has not been as well studied as that of ABRs. The research reported here seeks to evaluate the performance of two ASSR testing methods. Forty-eight subjects with normal hearing and twenty-two with impaired hearing are subjected to ASSR testing using the two different methods, which for confidentiality reasons we will call ASSR-A and ASSR-B. In addition, they are tested using Pure Tone Audiometry (PTA), which in this study is being used as the gold standard. Hearing thresholds are recorded for each subject at 1000, 2000 and 4000 Hz for each of ASSR-A, ASSR-B, and PTA. Our thanks go to Dr. Stavros Hatzopoulos and his associates at the University of Ferrara for providing the data.

The study has two main goals:

1. To explore the structure of the data and to identify factors that show significant differences in the mean response. In particular, we want to know whether there are significant differences between the two methods (ASSR-A and ASSR-B) and the gold standard (PTA), and between the two methods themselves, whether there are any differences at the different tested frequencies, and whether there are interactions between the two.

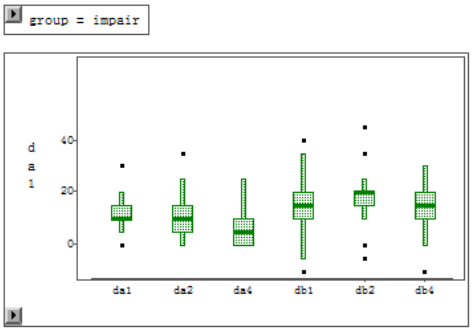
2. To predict the PTA threshold (gold standard) from the ASSR-A and ASSR-B thresholds separately at each frequency, in order to measure how accurate the ASSR measurements are and to obtain a “correction function” to correct the bias in the ASSR measurements.

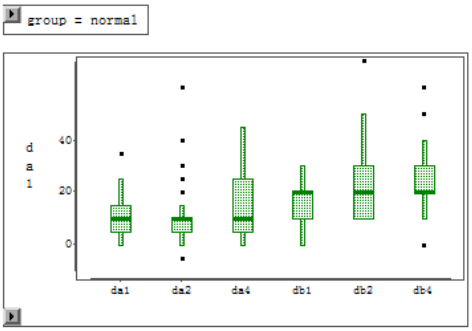
Exploratory Analysis

One problem we encountered early in the data exploration phase is that the PTA measurements for all subjects in the normal hearing group took the same value, 10 db HL. A check with the experimenters reveals that this is due to the clinical protocol: "The PTA value of 10 is considered an index of normality for frequency x, independently from the fact that a person might have different threshold sensitivity (ie 5, 0, -5 dB HL etc). So when you get a PTA of 10 you stop measuring."[1] This indicates that in statistical terms, these measurements are left-censored. However, since all PTA measurements in the normal group are censored, it would be very difficult, and perhaps impossible to model these data as censored.

As a result, we made two decisions. First, we decided to ignore the censoring, partly because we had little choice, and partly because these observations are regarded as "real" by the clinicians. Second, we decided to analyze the differences between the ASSR and PTA values. This is of little practical consequence for the analysis we conducted since we are interested in the differences between ASSR and PTA measurements, but is of great consequence for the model, since it allowed us to assume normality. In the following analyses, the differences between the ASSR-A and PTA measurements for each subject are labeled with the prefix "da", and those for the ASSR-B-PTA measurements are labeled "db". "da*i*" represents data measured under frequency1000*i*, as shown in the boxplots below:

[1] Stavros Hatzopoulos, personal communication.





The boxplots drawn for each group show us that both ASSR methods generally give greater values than PTA since the median of da (db) is above 0 for all frequencies, and, in fact the first quartile is above 0 for nearly all frequencies. Since the medians of da1, da2 and da4 are a little less than that of db1, db2 and db4 in both boxplots, we can tell that within each group, the difference between ASSR-A and PTA is a little less than the difference between ASSR-B and PTA. Finally, there is no certain pattern within each group, such as increasing trend or decreasing trend with respect to frequency.

Modeling

To assess the relation between the ASSR-PTA differences and method (ASSR-A and ASSR-B), group (normal and impaired hearing) and frequency (1000, 2000 and 4000 Hz), we built a general linear model. Because six measurements are obtained from each subject (two methods at each of three frequencies), we chose a repeated measures model to explore the data structure. The initial model specification is

$$y_{ijk} = u + g_i + f_j + gf_{ij} + m_k + gm_{ik} + fm_{jk} + gfm_{ijk} + \varepsilon_{ijkl}$$

where y_{ijk} is the ASSR-PTA difference for subject *l* (*l*=1,…,48 for the normal group and 1,…,22 for the impaired group) from group *i* (*i*=1,2), at frequency1000 *j* (*j*=1,2,3), using ASSR method *k* (*k*=1,2), u is the overall mean, $g_i$is the effect of group *i*, $f_j$is the effect of frequency *j*, $m_k$is the effect of method *k*, $\varepsilon_{ijkl}$is a random error term, and the other terms are interactions denoted in the usual way. The repeated measures are modeled by imposing a correlation structure on the ε_{ijkl}. The model can be written in general linear model form as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Y (dimension 420×1) represents the vector of observed responses, **β** (36×1) is an unknown vector of fixed-effects parameters with known design matrix **X** (420×36), and **ε** (420×1) is an unknown random error vector modeling the statistical noise around **X β**. A general within-subject covariance structures on **ε** models the possible dependence due to repeated observations on the same subject.

Since the subjects are considered independent, and each is assumed to have the same covariance structure, the overall covariance matrix for the epsilon vector is block diagonal with seventy identical 6 by 6 matrices on the diagonal. Assume, using the previous notation, the 6×1 vector of observations for subject *l* in group *i* has the form

$$\boldsymbol{y}_{il} = [\boldsymbol{y}'_{i1l}, \boldsymbol{y}'_{i2l}]' = [y_{i11l}, y_{i21l}, y_{i31l}, y_{i12l}, y_{i22l}, y_{i32l}]'$$

That is, the three observations for method 1 (at frequencies 1, 2, and 3) are followed by the three observations for method 2. This shows that the design is a doubly repeated measure with the repeated observations for the three frequencies repeated for the two methods.

There are a number of standard choices for the form of these matrices. Two that we considered are: compound symmetry (CS) and unstructured (UN).

In the compound symmetry structure, the covariance matrix of the vector of observations for each subject has the form

$$Compound\ Symmetry\ type = CS\text{:} \begin{pmatrix} \sigma_1^2+\sigma^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2+\sigma^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2+\sigma^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2+\sigma^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2+\sigma^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2+\sigma^2 \end{pmatrix}$$

The CS structure assumes equal variances for each observation and equal covariance between each pair of observations.

The unstructured covariance matrix assumes the most general structure possible:

$$Unstructured\ type = UN: \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} & \sigma_{16} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{24} & \sigma_{25} & \sigma_{26} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{34} & \sigma_{35} & \sigma_{36} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 & \sigma_{45} & \sigma_{46} \\ \sigma_{15} & \sigma_{25} & \sigma_{35} & \sigma_{45} & \sigma_5^2 & \sigma_{56} \\ \sigma_{16} & \sigma_{26} & \sigma_{36} & \sigma_{46} & \sigma_{56} & \sigma_6^2 \end{pmatrix}$$

[3, 4]

We also considered covariance matrices that incorporate the doubly repeated measures structure of the design. These use the direct or Kronecker product of two matrices. Specifically, we look at the product of a 2×2 unstructured matrix and a 3×3 compound symmetric, autoregressive of order 1, and unstructured matrix.k Their forms are as follows:

$$UN@CS = \begin{pmatrix} \delta_1^2 & \delta_{12}^2 \\ \delta_{12}^2 & \delta_2^2 \end{pmatrix} \otimes \begin{pmatrix} \sigma_1^2 + \sigma^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 + \sigma^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 + \sigma^2 \end{pmatrix}$$

$$= \begin{pmatrix} \delta_1^2(\sigma_1^2 + \sigma^2) & \delta_1^2\sigma_1^2 & \delta_1^2\sigma_1^2 & \delta_{12}^2(\sigma_1^2 + \sigma^2) & \delta_{12}^2\sigma_1^2 & \delta_{12}^2\sigma_1^2 \\ \delta_1^2\sigma_1^2 & \delta_1^2(\sigma_1^2 + \sigma^2) & \delta_1^2\sigma_1^2 & \delta_{12}^2\sigma_1^2 & \delta_{12}^2(\sigma_1^2 + \sigma^2) & \delta_{12}^2\sigma_1^2 \\ \delta_1^2\sigma_1^2 & \delta_1^2\sigma_1^2 & \delta_1^2(\sigma_1^2 + \sigma^2) & \delta_{12}^2\sigma_1^2 & \delta_{12}^2\sigma_1^2 & \delta_{12}^2(\sigma_1^2 + \sigma^2) \\ \delta_{12}^2(\sigma_1^2 + \sigma^2) & \delta_{12}^2\sigma_1^2 & \delta_{12}^2\sigma_1^2 & \delta_2^2(\sigma_1^2 + \sigma^2) & \delta_2^2\sigma_1^2 & \delta_2^2\sigma_1^2 \\ \delta_{12}^2\sigma_1^2 & \delta_{12}^2(\sigma_1^2 + \sigma^2) & \delta_{12}^2\sigma_1^2 & \delta_2^2\sigma_1^2 & \delta_2^2(\sigma_1^2 + \sigma^2) & \delta_2^2\sigma_1^2 \\ \delta_{12}^2\sigma_1^2 & \delta_{12}^2\sigma_1^2 & \delta_{12}^2(\sigma_1^2 + \sigma^2) & \delta_2^2\sigma_1^2 & \delta_2^2\sigma_1^2 & \delta_2^2(\sigma_1^2 + \sigma^2) \end{pmatrix}$$

$$UN@AR(1) = \begin{pmatrix} \delta_1^2 & \delta_{12}^2 \\ \delta_{12}^2 & \delta_2^2 \end{pmatrix} \otimes \begin{pmatrix} \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ \rho^2\sigma^2 & \rho\sigma^2 & \sigma^2 \end{pmatrix}$$

$$= \begin{pmatrix} \delta_1^2\sigma^2 & \delta_1^2\rho\sigma^2 & \delta_1^2\rho^2\sigma^2 & \delta_{12}^2\sigma^2 & \delta_{12}^2\rho\sigma^2 & \delta_{12}^2\rho^2\sigma^2 \\ \delta_1^2\rho\sigma^2 & \delta_1^2\sigma^2 & \delta_1^2\rho\sigma^2 & \delta_{12}^2\rho\sigma^2 & \delta_{12}^2\sigma^2 & \delta_{12}^2\rho\sigma^2 \\ \delta_1^2\rho^2\sigma^2 & \delta_1^2\rho\sigma^2 & \delta_1^2\sigma^2 & \delta_{12}^2\rho^2\sigma^2 & \delta_{12}^2\rho\sigma^2 & \delta_{12}^2\sigma^2 \\ \delta_{12}^2\sigma^2 & \delta_{12}^2\rho\sigma^2 & \delta_{12}^2\rho^2\sigma^2 & \delta_2^2\sigma^2 & \delta_2^2\rho\sigma^2 & \delta_2^2\rho^2\sigma^2 \\ \delta_{12}^2\rho\sigma^2 & \delta_{12}^2\sigma^2 & \delta_{12}^2\rho\sigma^2 & \delta_2^2\rho\sigma^2 & \delta_2^2\sigma^2 & \delta_2^2\rho\sigma^2 \\ \delta_{12}^2\rho^2\sigma^2 & \delta_{12}^2\rho\sigma^2 & \delta_{12}^2\sigma^2 & \delta_2^2\rho^2\sigma^2 & \delta_2^2\rho\sigma^2 & \delta_2^2\sigma^2 \end{pmatrix}$$

$$UN@UN = \begin{pmatrix} \delta_1^2 & \delta_{12}^2 \\ \delta_{12}^2 & \delta_2^2 \end{pmatrix} \otimes \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{pmatrix} = \begin{pmatrix} \delta_1^2\sigma_1^2 & \delta_1^2\sigma_{12} & \delta_1^2\sigma_{13} & \delta_{12}^2\sigma_1^2 & \delta_{12}^2\sigma_{12} & \delta_{12}^2\sigma_{13} \\ \delta_1^2\sigma_{12} & \delta_1^2\sigma_2^2 & \delta_1^2\sigma_{23} & \delta_{12}^2\sigma_{12} & \delta_{12}^2\sigma_2^2 & \delta_{12}^2\sigma_{23} \\ \delta_1^2\sigma_{13} & \delta_1^2\sigma_{23} & \delta_1^2\sigma_3^2 & \delta_{12}^2\sigma_{13} & \delta_{12}^2\sigma_{23} & \delta_{12}^2\sigma_3^2 \\ \delta_{12}^2\sigma_1^2 & \delta_{12}^2\sigma_{12} & \delta_{12}^2\sigma_{13} & \delta_2^2\sigma_1^2 & \delta_2^2\sigma_{12} & \delta_2^2\sigma_{13} \\ \delta_{12}^2\sigma_{12} & \delta_{12}^2\sigma_2^2 & \delta_{12}^2\sigma_{23} & \delta_2^2\sigma_{12} & \delta_2^2\sigma_2^2 & \delta_2^2\sigma_{23} \\ \delta_{12}^2\sigma_{13} & \delta_{12}^2\sigma_{23} & \delta_{12}^2\sigma_3^2 & \delta_2^2\sigma_{13} & \delta_2^2\sigma_{23} & \delta_2^2\sigma_3^2 \end{pmatrix}$$

To determine the most appropriate covariance structure, we used the AIC criterion. The AIC value is twice the difference between the number of parameters in the fitted model and the log likelihood of that model. Models with smaller AIC values are preferred. For the present model, the direct product of UN with UN gives the minimum AIC among the covariance structures we tried.

The fitting method can also affect the value of AIC. For the model with the UN covariance matrix, if we use restrict maximum likelihood (REML) to estimate the unknown covariance parameters, we get AIC = 2922.2 If we use maximum likelihood (ML) estimation, we get AIC = 2948.6 Besides covariance structure of type “UN”, we also tried “CS”, “Direct Product UN”, “Direct Product CS” , “Direct Product AR(1)”. For each of the covariance structures we tried in this project, REML brings smaller AIC than ML. Thus we will do all the following analysis with REML.

“Group” and “Method” are certainly qualitative variables in our analysis, but we have the option of including “frequency” as either qualitative or quantitative. “Frequency” has three levels: 1, 2, 4 kHz. For the above model with unstructured covariance matrix and “frequency” modeled as quantitative, an REML fit gives AIC = 2922.2, while modeling “frequency” as qualitative gives AIC = 2900.3 For all covariance structures, treating “frequency” as qualitative resulted in smaller AIC. Thus we will consider “frequency” as a qualitative variable for the following analysis.

SAS statistical software, specifically proc mixed, was used to fit and analyze the repeated measures models. Here is the table for the REML solution of the fixed effects from the full model, considering “frequency” as qualitative variable with UN@UN as covariance structure.

Type 3 tests of fixed effects

Effect	DF	DF	F Value	Pr > F
group	1	68	5.36	0.0236 *
freq	2	136	0.83	0.4386
group*freq	2	136	13.81	<.0001 *
method	1	68	37.36	<.0001 *
group*method	1	68	2.93	0.0916
method*freq	2	135	1.27	0.2830
group*method*freq	2	135	0.09	0.9138

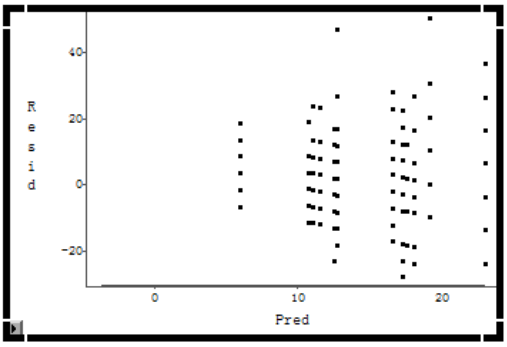
At the α=0.05 level of significance, “group”, “method”, and “group*freq” are significant. Since the interaction term “group*freq” is significant, we also included the main effect “freq” in our model. So finally we keep “group”, “freq”, “group*freq”, “method” in the model, while deleting “group*method”, “method*freq”, “group*method*freq”.

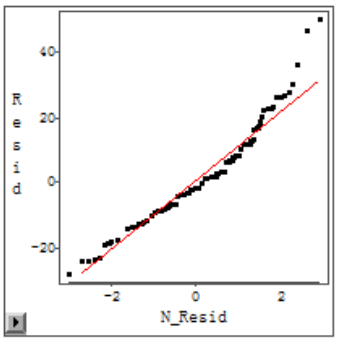
This is the model after variable selection:

$$y_{ijkl}= u+g_i+f_j+gf_{ij}+m_k+\varepsilon_{ijkl}$$

AIC becomes a little bit larger (2913.6) after our variable selection, but we still decided to use the reduced model since it brings great simplification.

Below is the plot of Studentized residuals versus predicted value and normal quantile plot of the Studentized residuals.

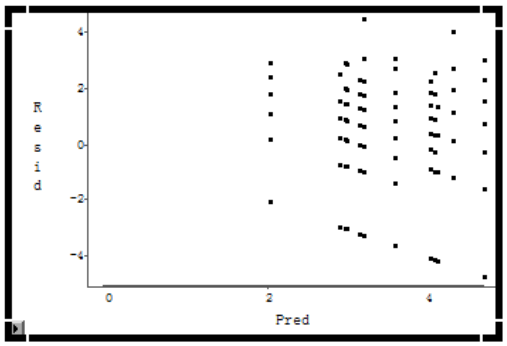


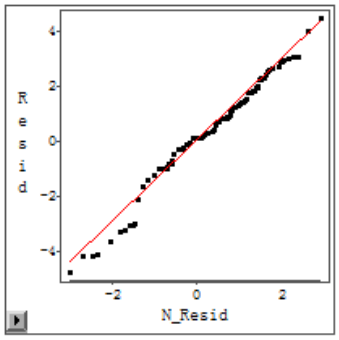


These plots show evidence of heteroscedasticity and nonnormality, so we considered transformations of the response variable. After exploring several transformations, we found a square root transformation to be most suitable.

After data transformation, we refit the full model, using the same candidate covariance matrices and both REML and ML fitting methods. As before, we found that the REML solution of the fixed effects from the full model, considering "frequency" as qualitative variable with UN@UN as covariance structure gave the smallest AIC=1328.8

Below are a plot of Studentized residuals versus predicted value and a normal quantile plot of Studentized residuals after data transformation. These confirm the improvement in model assumptions obtained from the transformation.





Here is the table for the fixed effects from that model.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
group	1	68	4.39	0.0399
freq	2	136	2.21	0.1133
group*freq	2	136	16.43	<.0001
method	1	67	40.68	<.0001
group*method	1	67	0.80	0.3733
method*freq	2	130	2.47	0.0883
group*method*freq	2	130	1.26	0.2859

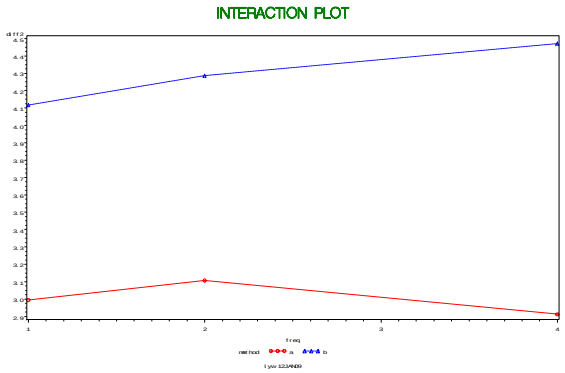
So after data transformation, we considered “freq” as a qualitative variable, kept “group”, “freq”, “group*freq”, “method” in the model and we chose type “UN@UN” as covariance structure. And we still chose the reduced model to do further analysis since AIC (1332.6) just becomes a little bigger while it brought great simplification for the model.

After data transformation, we estimated contrasts for main effects and interactions, to find out how they influence the responses in detail. Here is the table for the “Estimates”.

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
impair-normal	-0.6286	0.2676	68	-2.35	0.0217
method a - method b	-1.1215	0.1598	68	-7.02	<.0001
freq4-freq1	-0.1381	0.1440	136	-0.96	0.3394
freq2-freq1	0.1326	0.1435	136	0.92	0.3571
group*freq 12	0.1462	0.2871	136	0.51	0.6114
group*freq 24	1.3322	0.2921	136	4.56	<.0001

The table shows that mean difference (ASSR-PTA) for the impaired group is significantly less than that for the normal group; the mean difference (ASSR-PTA) for method A is significantly less than that for method B. And there is no significant difference (ASSR-PTA) due to frequencies, but there is a significant group-frequency interaction.

In order to explore the nature of the interaction, we drew two interaction plots.



Red line: Mean ASSR-PTA difference, a method

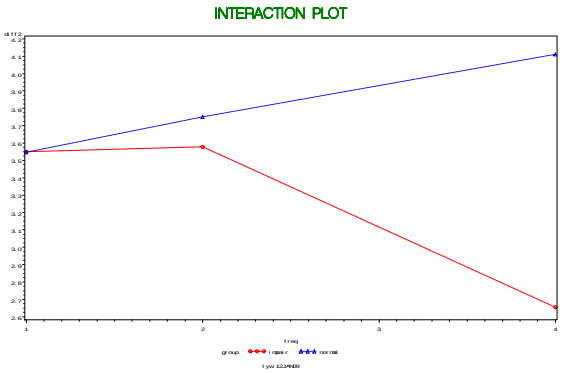
Blue line: Mean ASSR-PTA difference, b method

X Axis: frequency

Y Axis: Difference from PTA after data transformation

As frequency increases, the mean differences of both methods increase slowly and steadily. Overall, mean ASSR-B PTA difference is much larger than that for

ASSR-A PTA. The near-parallel lines confirm the nonsignificance of the method*freq interaction.



Red line: Mean ASSR-PTA difference, impaired group

Blue line: Mean ASSR-PTA difference, normal group

X Axis: frequency

Y Axis: Difference from PTA after data transformation

As frequency increases, the mean difference increases linearly for the normal group, while that of impaired group increases from 1 kHz to 2 kHz, then drops from 2 kHz to 4 kHz. This plot shows the nature of the significant group*freq interaction.

Finally we look at the table for the solution for fixed effects, and we can say since the estimate for intercept is 4.7199, we consider ASSR is generally bigger than PTA after data transformation.

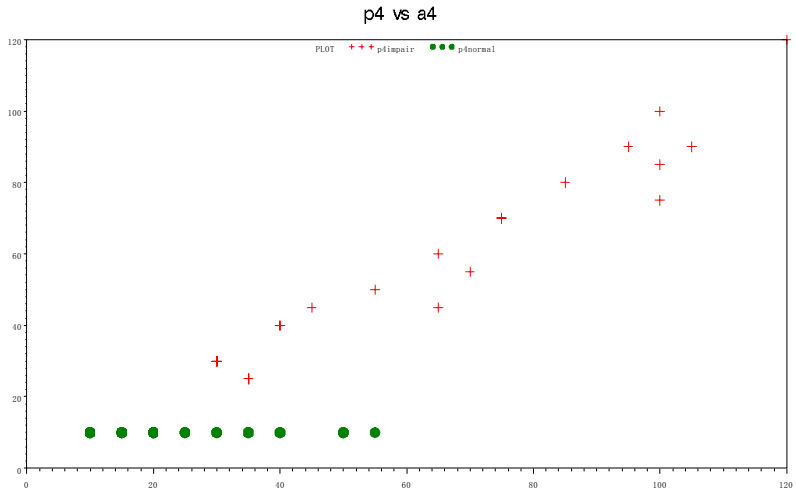
Effect	groupn	method	freq	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
Intercept				4.7199	0.2056	68	22.96	<.0001	0.05
groupn	1			-1.5654	0.3590	68	-4.36	<.0001	0.05
freq			1	-0.6011	0.1584	136	-3.80	0.0002	0.05
freq			2	-0.3954	0.1607	136	-2.46	0.0151	0.05
groupn*freq	1		1	1.4784	0.2881	136	5.13	<.0001	0.05
groupn*freq	1		2	1.3322	0.2921	136	4.56	<.0001	0.05
method		a		-1.1215	0.1598	68	-7.02	<.0001	0.05

Prediction

We want to predict the PTA threshold (gold standard) from the ASSR-A and ASSR-B thresholds separately at each frequency, in order to measure how accurate the ASSR measurements are and to obtain a “correction function” to correct the bias in the ASSR measurements.

Below is a plot of observed response from PTA (p_4) versus observed response from ASSR (a_4) at frequency of 4 kHz, with the two groups, impaired (red plus) and normal (green square), having different plotting symbols. Data from the two groups show different trends, which suggests that a mixture model might be used to do prediction for our project. [2, 6]

Red Plus: PTA for frequency 4, impaired group
Green Dot: PTA for frequency 4, normal group
Y Axis: PTA range from 0 to 120 by 20
X Axis: ASSR-A range from 0 to 120 by 20



Mixture distributions arise in practical problems when the measurements of a random variable are taken under two different conditions. For example, the distribution of heights in a population of adults reflects the mixture of males and females in the population. Mixture models can be used in problems, where the population of sampling units consists of a number of subpopulations within each of which a relatively simple model applies.

Here we assume normal distribution for data in each group (although this is not really true for the normal group). Since the PTA values for the impaired group seem to be linearly statistically related to the ASSR values while the PTA values for the normal group are constant, we consider a two-component mixture model in which a regression model applies in one component and a constant mean in the other. We have

$$f(y_i) = pf_1(y_i) + (1-p)f_2(y_i)$$

Where

$$f_1(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right)$$

$$f_2(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i-u)^2}{2\sigma^2}\right) \qquad i = 1, \dots, n$$

The likelihood equations are

$$\hat{p} = \frac{\sum_{i=1}^{n} \hat{P}\,(1|y_i)}{n}$$

$$\hat{\beta} = (X'PX')^{-1}X'Py$$

$$\hat{\mu} = \frac{\sum_{i=1}^{n} y_i\, \hat{P}(2|y_i)}{\sum_{i=1}^{n} \hat{P}\,(2|y_i)}$$

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{2}\sum_{i=1}^{n} e_{ji}^2 \hat{P}(j|y_i)}{n}$$

Where

$$e_{1i} = y_i - \beta_0 - \beta_1 x_i$$

$$e_{2i} = y_i - \hat{u}$$

$$\hat{P}(1|y_i) = \frac{\hat{p}\hat{f}_1(y_i)}{\{\hat{p}\hat{f}_1(y_i) + (1-\hat{p})\hat{f}_2(y_i)\}} = 1 - \hat{P}(2|y_i)$$

X is the matrix of predictor variables x_{ri},with $x_{0i} = 1$, $y' = (y_1, \dots, y_n)$, and $P = diag(\hat{P}(1|y_i))$. $\hat{P}(1|y_i)$ is the ML estimate of the posterior probability that the i-th observation comes from the 1-st component. Thus $\hat{\beta}$ is a weighted least squares estimator using weights$\hat{P}(1|y_i)$.

The EM algorithm is used for finding maximum likelihood estimates of parameters in probabilistic models, where the model depends on unobserved latent variables. [5] EM alternates between performing an expectation (E) step, which computes an expectation of the likelihood by including the latent variables as if they were observed, and a maximization (M) step, which computes the maximum likelihood estimates of the parameters by maximizing the expected likelihood found on the E step. The parameters found on the M step are then used to begin another E step, and the process is repeated.

In our project, the EM algorithm begins with initial estimates of the parameters as the first M-step, and then calculates the $\hat{P}(1|y_i)$ based on these estimates (E-step). New weighted least squares estimates of the parameters are computed using the $\hat{P}(1|y_i)$ in a new M-step, and the sequence of alternate E- and M-steps continues until convergence occurs to the EM estimates.

First M-step:

Give certain values $(p_0, u_0, \beta_0, \beta_1, \sigma_0)$ as initial estimates of the parameters based on the data. p_0 is the initial probability that a subject belongs to the impaired group. u_0 is the initial mean of the data for the normal group and β_0, β_1 are the slope and intercept estimates and σ_0 the square root of the MSE of the simple linear regression of PTA on ASSR for the frequency of interest.

$$p_0 = \frac{22}{70} \qquad u_0 = 10 \qquad resid\ (model\ y = x) \qquad e_2 = y - u_0 \qquad \sigma_0$$

$$\Downarrow$$

$$f_1(y_i) = \frac{1}{\sqrt{2\pi{\sigma_0}^2}} \exp\left(-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2{\sigma_0}^2}\right) = \frac{1}{\sqrt{2\pi{\sigma_0}^2}} \exp\left(-\frac{{e_{1i}}^2}{2{\sigma_0}^2}\right)$$

$$f_2(y_i) = \frac{1}{\sqrt{2\pi{\sigma_0}^2}} \exp\left(-\frac{(y_i - u_0)^2}{2{\sigma_0}^2}\right) = \frac{1}{\sqrt{2\pi{\sigma_0}^2}} \exp\left(-\frac{{e_{2i}}^2}{2{\sigma_0}^2}\right)$$

$$\Downarrow$$

$$\hat{P}(1|y_i) = \frac{p_0 f_1(y_i)}{\{p_0 f_1(y_i) + (1 - p_0) f_2(y_i)\}} = 1 - \hat{P}(2|y_i)$$

First E-step:

Compute the new weighted least squares estimates of the parameters,$\hat{p}, \hat{u}, \hat{\sigma}, \hat{\beta}_0, \hat{\beta}_1$, using the $\hat{P}(1|y_i)$

$$\hat{p} = \frac{\sum_{i=1}^{n} \hat{P}\,(1|y_i)}{n} \quad \hat{\mu} = \frac{\sum_{i=1}^{n} y_i\, \hat{P}(2|y_i)}{\sum_{i=1}^{n} \hat{P}\,(2|y_i)} \quad resid\ \left(model\ y = x, weight\ \hat{P}(1|y)\right)\ e_{2i} = y_i - \hat{u}$$

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{2} \sum_{i=1}^{n} e_{ji}^2 \hat{P}(j|y_i)}{n} = \frac{\sum_{i=1}^{n} \{{e_{1i}}^2 P(1|y_i) + e_{2i}^2 P(2|y_i)\}}{n}$$

Succeeding steps:

Calculate the new $\hat{P}(1|y_i)$ based on these estimates. The sequence of alternate E- and M-steps continues until convergence occurs to the EM estimates.
When the maximum difference between the estimates from successive iterations becomes less than 0.01, we consider the algorithm converges. The EM algorithm was run on the data until it converged within 10 iterations in each case, giving the estimates:

For a_1 and p_1:

$\hat{\beta}_0 = -5.1666 \quad \hat{\beta}_1 = 0.93328 \quad \hat{p} = 0.6223 \quad \hat{\mu} = 10.5614 \quad \hat{\sigma} = 4.19856$

For a_2 and g_2:

$\hat{\beta}_0 = -5.98716 \quad \hat{\beta}_1 = 0.92097 \quad \hat{p} = 0.39294 \quad \hat{\mu} = 10.3609 \quad \hat{\sigma} = 4.04938$

For a_4 and g_4:

$\hat{\beta}_0 = -1.75653 \quad \hat{\beta}_1 = 0.92218 \quad \hat{p} = 0.42920 \quad \hat{\mu} = 10.0020 \quad \hat{\sigma} = 3.95950$

For b_1 and g_1:

$\hat{\beta}_0 = -17.8464 \quad \hat{\beta}_1 = 1.07223 \quad \hat{p} = 0.38748 \quad \hat{\mu} = 10.9721 \quad \hat{\sigma} = 5.99132$

For b_2 and g_2:

$\hat{\beta}_0 = -12.3397 \quad \hat{\beta}_1 = 0.91797 \quad \hat{p} = 0.41353 \quad \hat{\mu} = 10.4465 \quad \hat{\sigma} = 6.31467$

For b_4 and g_4:

$\hat{\beta}_0 = -15.4444 \quad \hat{\beta}_1 = 1.02264 \quad \hat{p} = 0.45810 \quad \hat{\mu} = 10.2865 \quad \hat{\sigma} = 6.66587$

Once we obtained the mixture model fit, we tried two ways to do prediction at a given value of the predictor $x=x_{new}$. First, we used the conditional expectation $E(y| x_{new})$ as the predictor. It can be calculated for the mixture model as

$$E(y|x_{new}) = \int yf(y) = \hat{p}\int yf_1(y) + (1-\hat{p})\int yf_2(y) = \hat{p}E_1(y|x_{new}) + (1-\hat{p})E_2(y|x_{new}) = \hat{p}(\hat{\beta}_0 + \hat{\beta}_1 x_{new}) + (1-\hat{p})\hat{u}$$

Another way to predict y is to use minus twice the log of the likelihood function, with the addition of the new "observation" $x=x_{new}$ having a missing response value. We then predict the missing response by finding the value that minimizes -2logL

$$-2logL = nlog\hat{\sigma}^2 - 2\sum_{i=1}^{n} \log\left\{\hat{p}\exp\left(-\frac{e_{1i}^2}{2\hat{\sigma}^2}\right) + (1-\hat{p})\exp\left(-\frac{e_{2i}^2}{2\hat{\sigma}^2}\right)\right\}$$

Consider the results of using a_1 to predict p_1. We set x_{new} as 10, 25, 50, 70 and we get y_{pred1} as 7.39, 12.46, 20.91, 27.67, while we get y_{pred2} as 10.00, 12.09, 10.56 and 10.56 For x_{new} =10, the ML method is better. For x_{new} = 25, the values predicted from the two methods above are pretty similar and consistent with the data. For x_{new} = 50 and above, both methods give severely underestimate, with the ML method being worse. As x_{new} increases, the results from ML become worse and worse. For example, the mean value of observations at x_{new} = 70 is close to 50 (compare with predictions of 27.68 and 10.56). We did prediction with these two methods for other a_i and p_i pairs (i=2, 4) , b_i and p_i pairs (i=1,2,4) and we got results similar to those for a_1 and p_1.

The reason for the unsatisfactory prediction results is that even though $\hat{p}$ clearly depends on the x values, we didn't take x_{new} into consideration when calculating $\hat{p}$. In order to put the information of x_{new} into consideration when we calculate $\hat{p}$, we use a logistic regression model.

Logistic regression is a method used for prediction of the probability of occurrence of an event by fitting data to a logistic curve. [4] It makes use of one or more predictor variables that may be either numerical or categorical. For example, the probability that a person has a heart attack within a specified time period might be predicted from knowledge of the person's age, sex and body mass index.

For our model, assume that group membership of patient *i* is represented by the binary random variable Y_i, where $Y_i = 1$ if the patient has impaired hearing and $Y_i = 0$ if the patient has normal hearing. The probability distribution of Y_i , called Bernoulli, is

$$f(y_i|p_i) = {p_i}^{y_i}(1 - p_i)^{1-y_i}$$

In this model, p_i is the probability patient *i* has impaired hearing. We use the logistic regression model to relate the probability distribution of Y_i to the predictor (for us, an ASSR measurement) X_i. We can state the simple logistic regression model in the following fashion: Y_i are independent Bernoulli random variables with expected values $E\{Y_i\} = p_i$, where

$$E\{Y_i\} = p_i = \frac{\exp(\beta_0+\beta_1 X_i)}{1+\exp(\beta_0+\beta_1 X_i)} \quad (1)$$

The likelihood function becomes:

$$L(\beta_0, \beta_1 | y_1 \dots y_n) = \prod_{i=1}^{n} \left(\frac{\exp(\beta_0+\beta_1 X_i)}{1+\exp(\beta_0+\beta_1 X_i)}\right)^{y_i} \left(\frac{1}{1+\exp(\beta_0+\beta_1 X_i)}\right)^{1-y_i}$$

We can maximize this likelihood function to get the maximum likelihood estimates of β_0 and β_1.

Once the maximum likelihood estimates, b_0 and b_1, of β_0 and β_1 are gotten, we substitute these values into the response function (1) to obtain the fitted response function.

$$\hat{p} = \frac{\exp(b_0 + b_1 X)}{1 + \exp(b_0 + b_1 X)}$$

We shall use $\hat{p}_i$ to denote the fitted value for the *i*th case:

$$\hat{p}_i = \frac{\exp(b_0 + b_1 X_i)}{1 + \exp(b_0 + b_1 X_i)}$$

For a_1 and p_1, our fitted model is:

$$\hat{p} = \frac{\exp(-4.5469 + 0.1363X)}{1 + \exp(-4.5469 + 0.1363X)}$$

The logistic regression can also be used in improving the prediction of PTA values. In this, we use the prediction function:

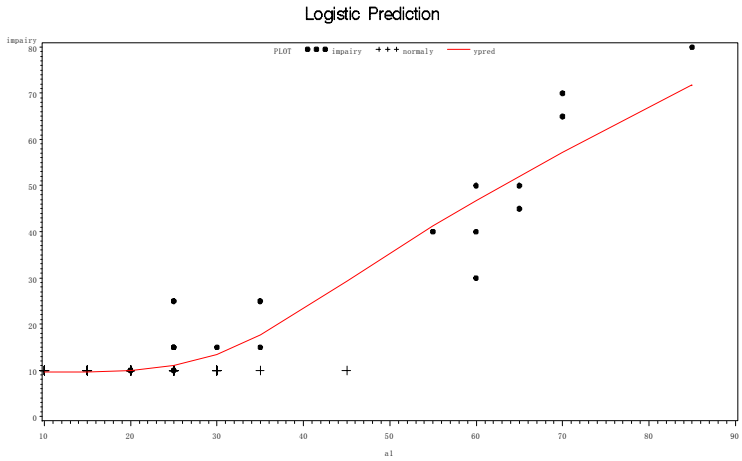
$$E(y|x_{new}) = E(y|x_{new}, normal) \times \hat{p}(normal|x_{new}) + E(y|x_{new}, impaired) \times \hat{p}(impaird|x_{new})$$
$$= 10 \times (1 - \hat{p}(impaird|x_{new}) + (b_0 + b_1 x_{new}) \times \hat{p}(impaird|x_{new})$$

where $\hat{p}(impaird|x_{new})$ is gotten from the simple logistic regression model and b_0, b_1 are gotten from simple linear regression fit to the impaired data only.

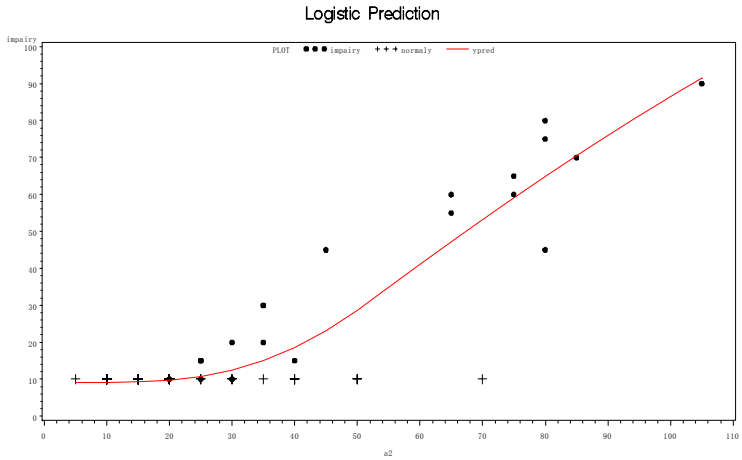
As an example, we consider p_1 as predictor variable in the logistic model and We fit a logistic regression to obtain $\hat{p}$ as a function of x. We set x_{new} as 10, 25, and 50 and we get the predicted PTA values marked as $y_{pred3,}$ which are 9.58741, 10.9850, 35.4944 and 57.0776, which are much better results compared with the results gotten from Mixture Model. Similar improvements in prediction were obtained for ASSR-A for the other frequencies and for ASSR-B for all frequencies. Calculated predictions for other pairs of a_i and p_i (i=1,2,4), the results we got are all better than predictions from the two mixture methods.

To see the whole trend for each a_i and p_i pair (i=1,2,4), we draw a plot of $E(y|x_{new})$ vs y as below. For each plot, the red line shows predicted y vs x, the data for the impaired group are plotted with dots and those for the normal group with pluses.

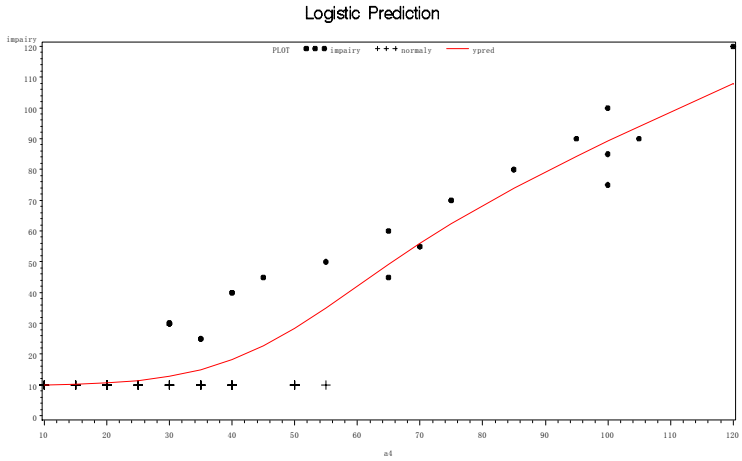
For a_1 and p_1



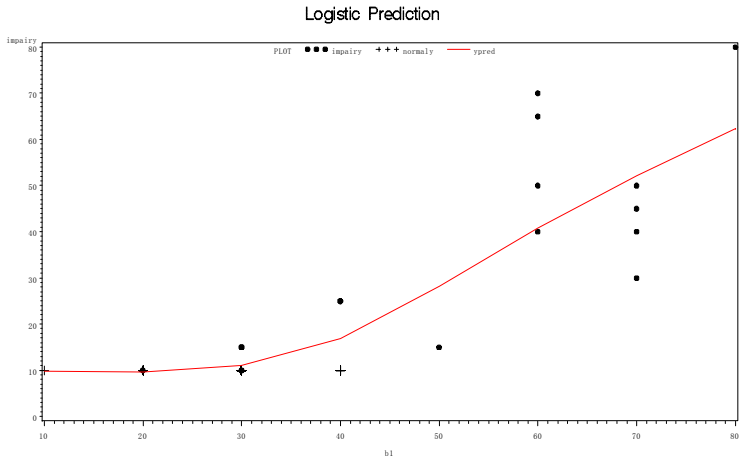
For a_2 and p_2



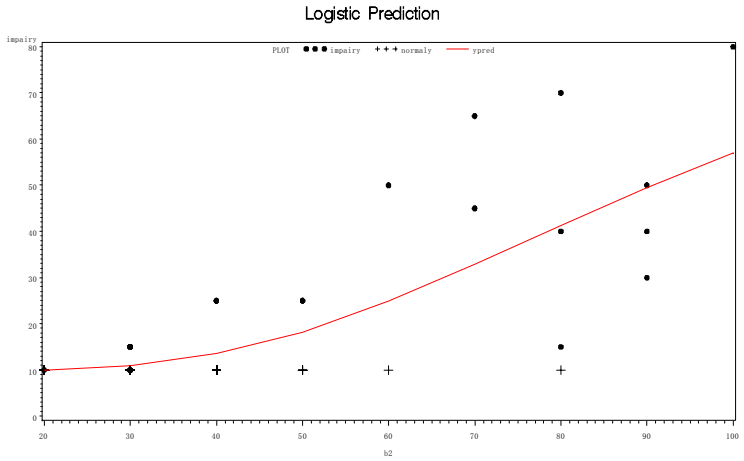
For a_4 and p_4



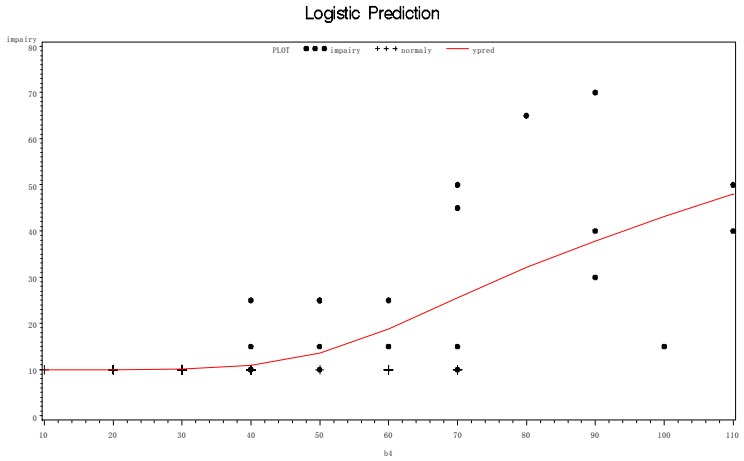
For b_1 and p_1



For b_2 and p_2



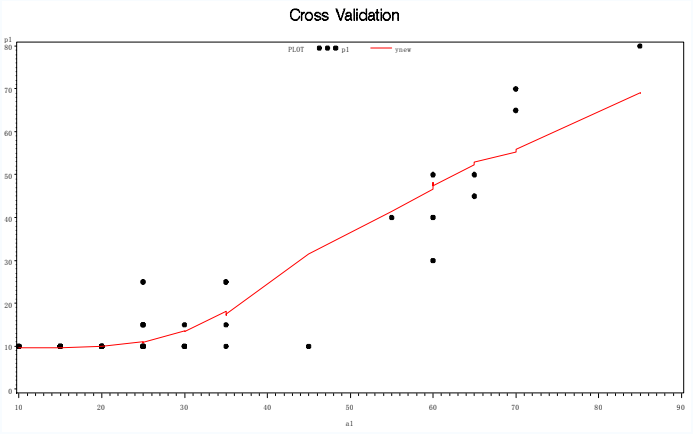
For b_4 and p_4



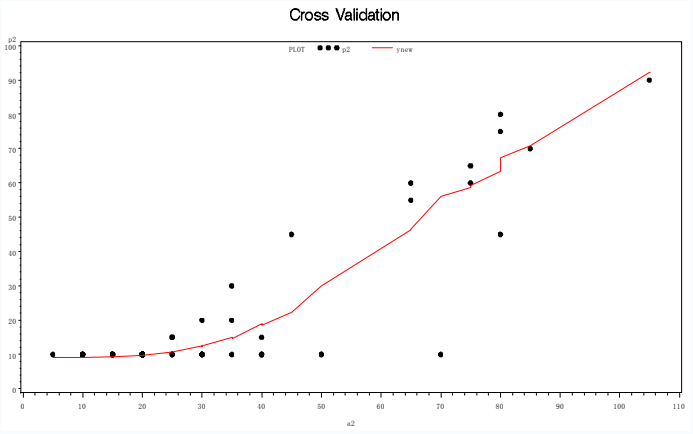
Finally, we did cross validation to give a more honest evaluation of the performance of the prediction method on future data. Cross-validation is the statistical practice of partitioning a sample of data into subsets such that the analysis is initially performed on a single subset, while the other subset(s) are retained for subsequent use in confirming and validating the initial analysis. The initial subset of data is called the training set; the other subset(s) are called validation or testing sets. In leave-one-out cross validation of a data set of size n, n pairs of subsets are selected, one of each pair consisting of one of the observations and the other of all the remaining data.

We applied leave-one-out cross validation to the logistic regression prediction algorithm. Specifically, we used all the data minus the one left out to compute the predictor of the omitted data value. Below are the combined plots of predicted PTA values from leave-one-out cross validation (red line) superimposed on the plot of PTA versus ASSR for each method and frequency.

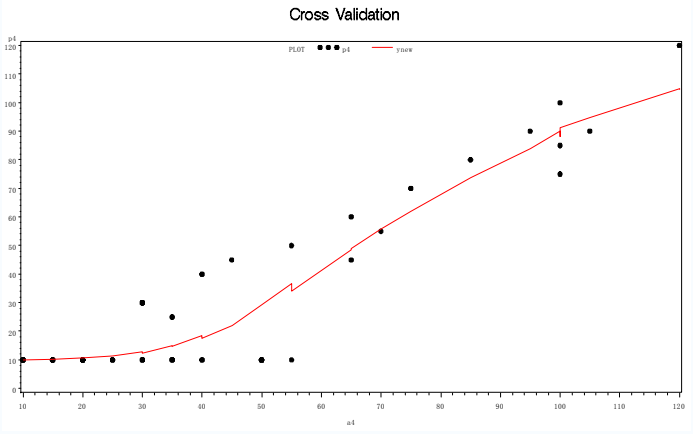
For a_1 and p_1



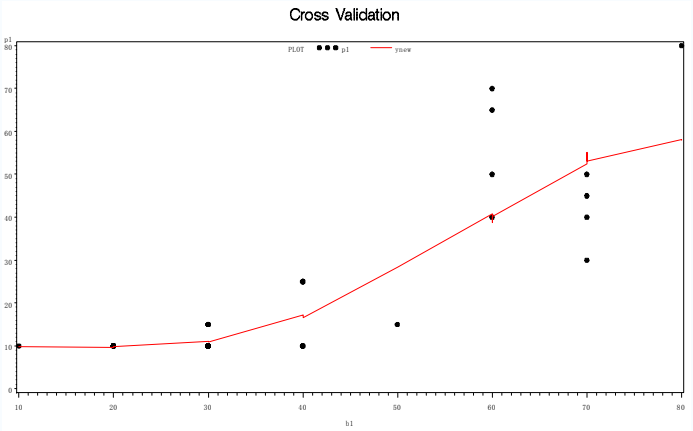
For a_2 and p_2



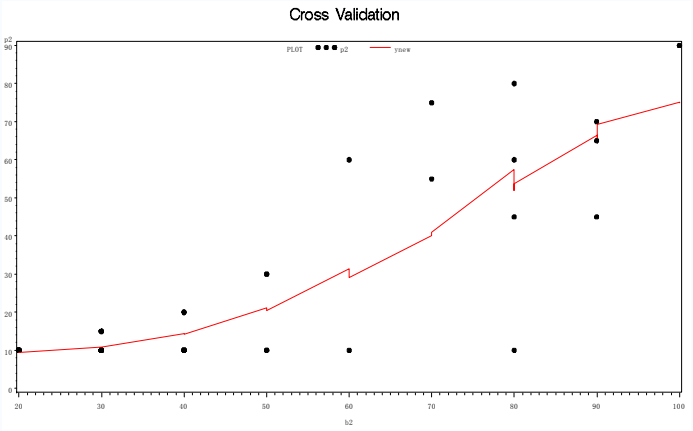
For a_4 and p_4



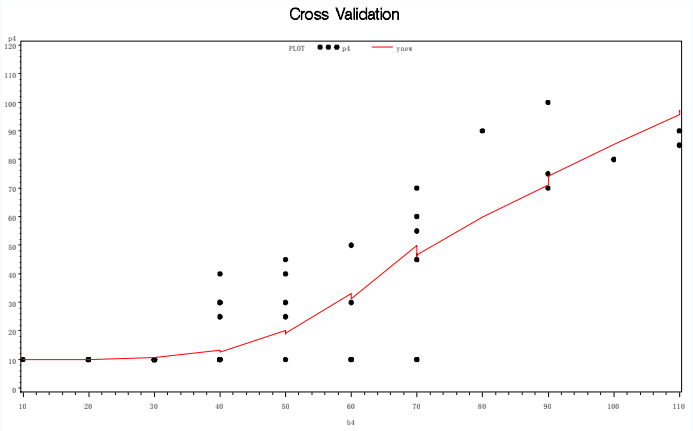
For b_1 and p_1



For b_2 and p_2



For b_4 and p_4



These plots further demonstrate the success of our third prediction method.

Conclusion

In this project, we had two goals:

1. To explore the structure of data relating two methods for obtaining auditory steady-state-evoked responses (ASSRs) to pure tone audiometry (PTA, the gold standard) in the evaluation of human hearing. In particular, we wanted to identify factors that show significant differences in the mean response. We also wanted to know whether there are significant differences between the two methods (ASSR-A and ASSR-B) and the gold standard (PTA), and between the two methods themselves, whether there are any differences at the different tested frequencies, and whether there are interactions between the two.
2. To predict the PTA threshold (gold standard) from the ASSR-A and ASSR-B thresholds separately at each frequency, in order to measure how accurate the ASSR measurements are and to obtain a "correction function" to correct the bias in the ASSR measurements.

We accomplished the first goal by exploratory data analysis and by constructing a repeated measures linear model. The model shows that the mean difference in ASSR and PTA is significantly related to patient status (normal or impaired hearing) and that this relation differs by frequency. The model also shows that the mean difference differs significantly for the two ASSR methods.

In order to address the second goal (predicting the PTA threshold from the ASSR-A and ASSR-B thresholds separately at each frequency), we initially modeled the relation of the PTA responses to the ASSR values for the two patient status groups as a mixture model. We then tried two prediction methods based on this model,

with disappointing results. A second approach, using a logistic regression model to predict group membership based on ASSR value and then using those predictions to obtain a predictor of the PTA value, gave successful results.

Bibliography

[1] Audrea Canale, Michelangelo lacilla, Andrea Luigi Cavalot and Roberto Albera (Published online 2006), *Auditory Steady-State Responses and Clinical Applications* Website: http://www.springerlink.com/content/8604417001050552/fulltext.html

[2] Aitkin, Murray and Granville Tunnicliffe Wilson (1980), Mixture Models, Outliers, and the EM Algorithm. *Technometrics*, 22:3, 325-331

[3] Verbeke, Geert and Molenberghs, Geert (2001), **Linear Mixed Models for Longitudinal Data.** New York: Springer-Verlag

[4] Mc Culloch, Charles E. and Shayle R. Searle (2003), **Generalized Linear and Mixed Models.** Beachwood: Institute of Mathematical Statistics

[5] Casella, George and Roger L. Berger (2002), **Statistical Inference.** Beijing: Thomson Learning

[6] Gelman, Andrew, John B. Carlin, Hal S. Stern and Donald B. Rubin (2004), **Bayesian Data Analysis.** New York: CRC Press