Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v13n1p1

A statistical modelling of the visual acuity measurement and its multiple test procedure
By Mita et al.

Published: 02 May 2020

Electronic Journal of Applied Statistical Analysis
Vol. 13, Issue 01, May 2020, 1-15
DOI: 10.1285/i20705948v13n1p1

A statistical modelling of the visual acuity measurement and its multiple test procedure

Norihiro Mita[a], Hiroshi Sasaki[a], Kazutaka Kani[b], Akio Tabuchi[c], and Heihachiro Hara[*d]

[a] *Department of Ophthalmology, Kanazawa Medical University, Ishikawa-Uchinada, Japan*
[b] *Department of Orthoptics and Visual Science, Kyushu University of Health and Welfare, Nobeoka, Japan*
[c] *Department of Sensory Science, Kawasaki University of Medical Welfare, Kurashiki, Japan*
[d] *Department of Health Informatics, Kawasaki University of Medical Welfare, Kurashiki, Japan*

Published: 02 May 2020

To establish the computer assisted system of the visual acuity test, we propose a statistical modelling of the visual acuity measurement and its multiple test procedure. The psychometric functions for individual patients are produced by the logistic regression combined with the guessing rate. We adopt test statistics based on (i) psychometric functions (the Cochran-Mantel-Haenszel method) and (ii) psychophysical thresholds (the delta method). The multiple comparisons are performed by the step-down procedure with Ryan-Einot-Gabriel-Welsch (REGW) significance levels. To show the practical effectiveness of our system, we present a numerical example of four patient groups.

keywords: Cochran-Mantel-Haenszel test statistic, delta test statistic, psychometric functions, psychophysical thresholds, step-down procedure.

1 Introduction

The methodology of statistics is versatile and powerful to find a regularity or an irregularity from observed data, and many statistical modellings are applied to various problems.

*Corresponding author: hhrhara@gmail.com

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

To prove the effectiveness of new medical measurement and treatment, we need to obtain precise statistical characteristics. In the field of visual acuity measurement, we have some problems to be solved: obtaining precise statistical data for individual patients, and testing multiple groups of various type patients.

On the conventional visual acuity measurement, a trained inspector shows the Snellen chart to an individual patient and checks whether the patient can recognize letters or symbols (Landolt C, Tumbling E) correctly. In conventional methods, we often evaluate the overall trend of visual acuity values (pre-operative and post-operative values etc.) for several patients. We can also define "improvement" as two-line or three-line improvement of visual acuity. However, the former method does not enable observation of individual changes, and the latter method has no statistical basis (Tokutake et al., 2014) .

It is known that the cumulative probability distribution of the normal distribution can estimate the probability of visual acuity for an individual patient (Watson, 1979; Simpson, 1995; Tokutake et al., 2014). By developing the system on the personal computer, the automated measurements of the visual acuity are proposed (Bach, 1996; Beck et al., 2003; Schulze-Bonsel et al., 2006). The testing of significance difference between multiple thresholds is developed by Nagai et al. (2006). Some visual acuity testings for individual patients by using psychometric functions have been proposed (Mita et al., 2010, 2014; Tokutake et al., 2014). In the new method based on logistic regression, we can obtain a psychometric function to measure visual acuity more precisely than with conventional methods. The most important property of new method is that we can calculate both visual acuity and its variance in one measurement, which cannot be achieved by use of conventional methods.

For multiple comparison problems, several statistical approaches were proposed. Ryan, Einot-Gabriel and Welsch proposed some stepwise procedures for multiple comparison problems (Ryan, 1960; Einot and Gabriel, 1975; Welsch, 1977). Félix and Menezes (2018) reported comparisons of ten corrections methods for t-test in multiple comparisons via Monte Carlo study. By adopting the bootstrap logistic regression, Mita et al. (2014) developed an algorithm of multiple comparisons with a control, and Mita et al. (2017) proposed an algorithm of multiple test based on the step-down procedure.

In epidemiological study, the Mantel-Haenszel test statistic is developed by Mantel and Haenszel (1959), and it is applied to the problem of differential item functioning by Holland and Thayer (1988). The generalized Cochran-Mantel-Haenszel (CMH) statistics, which allows many strata, expands the applicable fields (Landis et al., 1978; Somes, 1986; Zhang and Boos, 1996; Penfield, 2001).

In the present article, we propose a statistical modelling of the visual acuity measurement and its multiple test procedure.

Step 1 Modelling of the visual acuity for individual patient: We present the logistic regression combined with a constant guessing rate, and define the psychometric function, the psychophysical threshold and its variance in section 2.

Step 2 Computing test statistics: We show the CMH test statistic based on psychometric functions in section 3, and the delta test statistic based on psychophysical thresholds and their variances in section 4.

Step 3 Multiple comparisons of groups: We show the multiple test by the step-down procedure with Ryan-Einot-Gabriel-Welsch (REGW) significance levels in section 5. Finally, in section 6 we present an application of our algorithm to the visual acuity measurement of four groups.

2 Modelling of the visual acuity for individual patient

We use the Landolt C as the optotype, and the size of the optotype is determined on the basis of a logarithmic scale. The patient is directed to choose one of four directions (left, right, bottom and top) of optotypes. Since the criterion for an "unreadable" measurement is subjective, the accuracy of the measurement is compromised if the patient answers "unreadable". Therefore "unreadable" answer is not allowed in our measurement. Thus we fix the guessing rate as $1/4 = 0.25$.

Let $P(x_i; a_j, b_j)$ be the probability that the patient j $(j = 1, \cdots, N)$ answers the visual acuity test i $(i = 1, \cdots, M)$ correctly:

$$P(x_i; a_j, b_j) = p(x_i; a_j, b_j) + c_0 \left(1 - p(x_i; a_j, b_j)\right) \quad (i = 1, \cdots, M; j = 1, \cdots, N),$$

where

$$p(x_i; a_j, b_j) = \left(1 + \exp(-(a_j + b_j x_i))\right)^{-1},$$

x_i is the visual target of test i, a_j and b_j are intercept and slope parameters, respectively, which depend on the patient j, and c_0 $(0 \leq c_0 < 1)$ is a prescribed constant which defines the guessing rate.

Let D_j $(j = 1, \cdots, N)$ be the set of patient data such that

$$D_j = \{(x_i, \mu_{ij}, \nu_{ij}) \ \ (i = 1, \cdots, M)\} \quad (j = 1, \cdots, N),$$

where μ_{ij} and ν_{ij} $\ (0 \leq \mu_{ij} \leq \nu_{ij};\ i = 1, \cdots, M; j = 1, \cdots, N)$ are responses of patient j to the test i such that μ_{ij} is the number of correct outcomes among ν_{ij} trials. The binomial likelihood $L(a_j, b_j)$ is given by

$$L(a_j, b_j) = \prod_{i=1}^{M} P_{ij}^{\mu_{ij}} (1 - P_{ij})^{\nu_{ij} - \mu_{ij}} \quad (j = 1, \cdots, N),$$

where $P_{ij} = P(x_i; a_j, b_j)$ $(i = 1, \cdots, M; j = 1, \cdots, N)$. Then we can obtain optimum values $\hat{a}_j$ and $\hat{b}_j$ for a_j and b_j, respectively, by adopting the Fisher score method. We define the psychometric function $\varphi_j(x)$ $(j = 1, \cdots, N)$ for individual patient j such that

$$\varphi_j(x) = P(x; \hat{a}_j, \hat{b}_j) \quad (-\infty < x < +\infty; j = 1, \cdots, N).$$

Then we define the psychophysical threshold ξ_j $(j = 1, \cdots, N)$ of patient j with guessing rate c_0:

$$\xi_j = \varphi_j^{-1}\left(\frac{1 + c_0}{2}\right) \quad (j = 1, \cdots, N).$$

The threshold ξ_j and its variance $\mathrm{var}(\xi_j)$ are obtained by

$$\xi_j = -\frac{\hat{a}_j}{\hat{b}_j}, \quad \mathrm{var}(\xi_j) = \xi_j^2 \left(\frac{\mathrm{var}(\hat{a}_j)}{\hat{a}_j^2} - 2\frac{\mathrm{cov}(\hat{a}_j, \hat{b}_j)}{\hat{a}_j \hat{b}_j} + \frac{\mathrm{var}(\hat{b}_j)}{\hat{b}_j^2} \right),$$

where $\mathrm{var}(\hat{a}_j)$ and $\mathrm{var}(\hat{b}_j)$ are variances of $\hat{a}_j$ and $\hat{b}_j$, respectively, and $\mathrm{cov}(\hat{a}_j, \hat{b}_j)$ is the covariance of $\hat{a}_j$ and $\hat{b}_j$.

3 Test statistic based on psychometric functions

Let r $(r = 1, \cdots, \rho)$ be the stratum of observations and let $u_{k\ell}^r$ $(r = 1, \cdots, \rho; k = 1, \cdots, \kappa; \ell = 1, \cdots, \lambda)$ be the number of observations at r-th stratum of the $\kappa \times \lambda$ contingency table. For applying the Cochran-Mantel-Haenszel (CMH) test statistic to the visual acuity results, we choose indices r, k, ℓ in the contingency table such that

(i) strata $r\,(r = 1, \cdots, \rho)$: visual acuity strata $(x^1, \cdots, x^\rho)$,
(ii) rows $k\,(k = 1, \cdots, \kappa)$: patient groups $(\Gamma_1, \cdots, \Gamma_\kappa)$,
(iii) columns $\ell\,(\ell = 1, \cdots, \lambda)$: responses of visual tests ($\ell = 1$ for "correct" and $\ell = 2$ for "incorrect").

Then we fix the number of columns at $\lambda = 2$ and choose the number of observations $u_{k\ell}^r$ $(r = 1, \cdots, \rho; k = 1, \cdots, \kappa; \ell = 1, 2)$ such that

$$u_{k1}^r = \sum_{j \in \Gamma_k} \varphi_j(x^r), \quad u_{k2}^r = n_k^r - u_{k1}^r \quad (r = 1, \cdots, \rho; k = 1, \cdots, \kappa),$$

where n_k^r is the number of patients in group Γ_k, and $\varphi_j(x^r)$ is the value of psychometric function of patient j at the visual acuity stratum x^r.

The null hypothesis is described such that there is no association between the row (patient group) and the column (response of visual test).

We define the total and partial summations of $u_{k\ell}^r$:

$$t^r = \sum_{k=1}^{\kappa} \sum_{\ell=1}^{2} u_{k\ell}^r \quad (r = 1, \cdots, \rho),$$

$$m_\ell^r = \sum_{k=1}^{\kappa} u_{k\ell}^r \ \ (r = 1, \cdots, \rho; \ell = 1, 2), \quad n_k^r = \sum_{\ell=1}^{2} u_{k\ell}^r \ \ (r = 1, \cdots, \rho; k = 1, \cdots, \kappa).$$

Let

$$\overline{\boldsymbol{u}}^r = \left(u_{11}^r, u_{21}^r, \cdots, u_{(\kappa-1)1}^r \right)' \quad (r = 1, \cdots, \rho)$$

be the vector of the number of observations, where the symbol $'$ means the transpose of vector or matrix. We define the vector of partial summation:

$$\boldsymbol{n}^r = (n_1^r, \cdots, n_{\kappa-1}^r)' \quad (r = 1, \cdots, \rho).$$

Then the expected value vector $\boldsymbol{e}^r$ and the variance-covariance matrix $\boldsymbol{V}^r$ are expressed as

$$\boldsymbol{e}^r = \frac{m_1^r}{t^r}\boldsymbol{n}^r, \quad \boldsymbol{V}^r = \frac{m_1^r m_2^r}{(t^r)^2(t^r-1)}\boldsymbol{N}^r \quad (r = 1, \cdots, \rho),$$

where

$$\boldsymbol{N}^r = t^r \operatorname{diag}(\boldsymbol{n}^r) - \boldsymbol{n}^r(\boldsymbol{n}^r)',$$

$\operatorname{diag}(\boldsymbol{n}^r)$ is the $(\kappa-1)\times(\kappa-1)$ diagonal matrix with diagonal elements of $\boldsymbol{n}^r$. The Cochran-Mantel-Haenszel (CMH) test statistic Ψ_K for the set of groups $K = \{\Gamma_1, \cdots, \Gamma_\kappa\}$ is given by

$$\Psi_K = \boldsymbol{u}'\,\boldsymbol{V}^{-1}\boldsymbol{u},$$

where

$$\boldsymbol{u} = \sum_{r=1}^{\rho}\overline{\boldsymbol{u}}^r - \sum_{r=1}^{\rho}\boldsymbol{e}^r, \quad \boldsymbol{V} = \sum_{r=1}^{\rho}\boldsymbol{V}^r.$$

This CMH test statistic Ψ_K has an asymptotic χ^2-distribution with $(\kappa-1)$ degrees of freedom under the null hypothesis (Agresti, 2002).

For the special case $\kappa = 2$, we adopt the Mantel-Haenszel test statistic with Yates correction Ψ_y (Mantel and Haenszel, 1959).

4 Test statistic based on psychophysical thresholds

For applying the delta test statistic to the problem of comparisons of visual acuity, we compute mean values of the psychophysical threshold $\overline{\xi}_k$ and its variance $\operatorname{var}(\overline{\xi}_k)$ $(k = 1, \cdots, \kappa)$ of group Γ_k such that

$$\overline{\xi}_k = \frac{1}{N_k}\sum_{j\in\Gamma_k}\xi_j, \quad \operatorname{var}(\overline{\xi}_k) = \frac{1}{N_k}\sum_{j\in\Gamma_k}\operatorname{var}(\xi_j) \quad (k = 1, \cdots, \kappa),$$

where N_k is the number of patients in the group Γ_k, and ξ_j and $\operatorname{var}(\xi_j)$ $(j = 1, \cdots, N_k)$ are psychophysical threshold and its variance of individual patient j in Γ_k, respectively. Let $\overline{\xi}$ be the total mean of $\overline{\xi}_k$ $(k = 1, \cdots, \kappa)$ defined by

$$\overline{\xi} = \frac{1}{\kappa}\sum_{k=1}^{\kappa}\overline{\xi}_k.$$

Let A_k $(k = 1, \cdots, \kappa)$ be the main effect defined by

$$A_k = \overline{\xi}_k - \overline{\xi} \quad (k = 1, \cdots, \kappa).$$

We shall test the null hypothesis for κ groups $(\kappa \geq 2)$: $\overline{\xi}_1 = \cdots = \overline{\xi}_\kappa$. If we define the following notations:

$$u_k = A_1 - A_{k+1} = \overline{\xi}_1 - \overline{\xi}_{k+1} \quad (k = 1, \cdots, \kappa-1),$$

then the null hyposesis can be rewritten such that $u_1 = \cdots = u_{\kappa-1} = 0$.

We introduce the following partial derivatives (Nagai et al., 2006):

$$U_{kt} = \frac{\partial u_k}{\partial \overline{\xi}_t} = \delta_{t1} - \delta_{t,k+1} \quad (k = 1, \cdots, \kappa - 1;\ t = 1, \cdots, k),$$

where $\delta_{pp} = 1$ and $\delta_{pq} = 0$ $(p \neq q)$. By using u_k, U_{kt}, $\mathrm{var}(\overline{\xi}_t)$, we denote the vector $\boldsymbol{u}$ and matrices $\boldsymbol{U}, \overline{\boldsymbol{V}}$ such that

$$\boldsymbol{u} = (u_k)', \quad \boldsymbol{U} = (U_{kt}), \quad \overline{\boldsymbol{V}} = \mathrm{diag}(\mathrm{var}(\overline{\xi}_t)) \quad (k = 1, \cdots, \kappa - 1;\ t = 1, \cdots, \kappa),$$

where $\boldsymbol{u}$ is the column vector of the order of $(\kappa - 1)$, $\boldsymbol{U}$ is the $(\kappa - 1) \times \kappa$ matrix, and $\overline{\boldsymbol{V}}$ is the $\kappa \times \kappa$ diagonal matrix with diagonal elements of $\mathrm{var}(\overline{\xi}_t)$. Then we have the asymptotic variance-covariance matrix $\boldsymbol{V}$:

$$\boldsymbol{V} = \boldsymbol{U}\,\overline{\boldsymbol{V}}\boldsymbol{U}'.$$

It is known that the likelihood estimated parameters of psychometric function have the normal distribution in multiple variables (Pas and Koenderink, 2004; Kingdom and Prins, 2010; Tokutake et al., 2014). Since $\boldsymbol{u}$ has the normal distribution in multiple variables in our visual acuity testing, the delta test statistic Δ_K for the set of groups $K = \{\Gamma_1, \cdots, \Gamma_\kappa\}$ defined by

$$\Delta_K = \boldsymbol{u}'\,\boldsymbol{V}^{-1}\boldsymbol{u}$$

has an asymptotic χ^2-distribution with $(\kappa - 1)$ degrees of freedom under the null hypothesis.

5 Step-down procedure for multiple comparisons

Let G be the set of patient groups, and g be the number of groups in G. Let K be the subset of G, and κ be the number of groups in K. We shall test the following null hyposesis on the visual acuity problem of multiple comparisons for patient groups.

null hyposesis H_K^0: responses of visual tests are independent of groups in K.

Let $\mathcal{F}$ be the hierarchical family of null hypotheses for G. Let $\mathcal{F}_K$ be the subset of $\mathcal{F}$ such that

$$\mathcal{F}_K = \{H_{K'}^0 \mid H_{K'}^0 \subseteq H_K^0 \ \text{ for } H_{K'}^0 \in \mathcal{F}\} \quad (K, K' : \text{subsets of } G),$$

where $H_{K'}^0 \subseteq H_K^0$ means that $H_{K'}^0$ implies H_K^0.

Let W_K be the test statistic of K, where we choose W_K such that $W_K = \Psi_K$ for the CMH test and $W_K = \Delta_K$ for the delta test.

We adopt Ryan-Einot-Gabriel-Welsch (REGW) significance levels α_κ $(2 \leq \kappa \leq g)$ for the test of null hypothesis H_K^0 such that

$$\alpha_\kappa = \begin{cases} 1 - (1 - \alpha)^{\kappa/g} & (2 \leq \kappa \leq g - 2), \\ \\ \alpha & (g - 1 \leq \kappa \leq g), \end{cases}$$

where α is the type I familywise error rate (Ryan, 1960; Einot and Gabriel, 1975; Welsch, 1977).

Let $\chi^2_{\kappa-1}$ be the critical value for α_κ and let p_K be the p-value for K such that

$$P(X^2 > \chi^2_{\kappa-1}) = \alpha_\kappa \ \ (2 \le \kappa \le g), \quad p_K = P(X^2 > W_K),$$

where $P(X^2 > \chi^2_{\kappa-1})$ etc. are the upper probabilities of the chi-squared distribution X^2 with $(\kappa - 1)$ degrees of freedom. Then the null hypothesis H^0_K can be rejected with α if all hypotheses $H^0_{K'}$ in $\mathcal{F}_K$ are rejected with $\alpha_{\kappa'}$ $(2 \le \kappa' \le g)$ such that

$$W_{K'} > \chi^2_{\kappa'-1} \quad \text{for all } \ H^0_{K'} \in \mathcal{F}_K \quad (K, K' : \text{subsets of } G),$$

where κ' is the number of groups in K'.

The statistical power q_K and the error of the second kind β_K on K are the probabilities given by

$$q_K = 1 - \beta_K = P(\tilde{X}^2(\lambda_K) > \chi^2_{\kappa-1}) \quad (2 \le \kappa \le g),$$

where $P(\tilde{X}^2(\lambda_K) > \chi^2_{\kappa-1})$ is the upper probability of the non-central chi-squared distribution $\tilde{X}^2$ with $(\kappa - 1)$ degrees of freedom, and the non-centrality parameter λ_K is chosen as $\lambda_K = W_K$. Then we can obtain the cumulative statistical power $\mathrm{cum}(q_K)$ on the hypothesis H^0_K such that

$$\mathrm{cum}(q_K) = \prod_{K' \subseteq K} q_{K'} \quad (K, K' : \text{subsets of } G),$$

where $\prod_{K' \subseteq K}$ means the product for all K' satisfying $H^0_{K'} \in \mathcal{F}_K$.

6 Application to the visual acuity testing

For checking the reliability of our algorithm, we took the data from one individual with no visual abnormalities in complete refractive correction and also took the data from the same individual in $+0.5D$ incomplete refractive correction. Then by modifyng these data, we made replicas to have some samples in four groups. Althogh the data chosen here are not fully real, they provide evidence enough to know the effectiveness of our algorithm in multiple comparisons.

The explanatory variable x in our measurement is the logarithmic visual acuity (LogVA). We shall test differences of visual acuities for patients in 4 groups Γ_k $(k = 1, \cdots, 4)$. The number of Landolt-C targets M_k and the number of patients N_k in groups Γ_k are chosen such that

$$M_k = 7 \ (k = 1, \cdots, 4); \quad N_1 = 22, \ \ N_2 = 14, \ \ N_3 = 11, \ \ N_4 = 13.$$

For obtaining precise data by the constant stimuli method, we choose the number of trials ν_{ij} for patient j to the test i satisfying such that

$$\nu_{ij} = 0 \ \text{ or } \ 20 \ \ (i = 1, \cdots, M_k; j = 1, \cdots, N_k), \quad \sum_{i=1}^{M_k} \nu_{ij} = 120 \qquad (j = 1 \cdots, N_k).$$

For each patient j ($j = 1, \cdots, N_k$) in group Γ_k ($k = 1, \cdots, 4$), we obtain optimum values of parameters $\hat{a}_j$, $\hat{b}_j$, and psychophysical threshold ξ_j by adopting the Fisher score method. The psychometric functions $\varphi_j(x)$ ($j = 1, \cdots, N_k$) for indivisual patients in groups Γ_k ($k = 1, \cdots, 4$) are shown in Figs. 1 (Γ_1), 2 (Γ_2), 3 (Γ_3), 4 (Γ_4).

The hierarchical family of null hypotheses for 4 groups is

$$\mathcal{F} = \left\{H_{12}^0, H_{13}^0, H_{14}^0, H_{23}^0, H_{24}^0, H_{34}^0, H_{123}^0, H_{124}^0, H_{134}^0, H_{234}^0, H_{1234}^0\right\}.$$

The subsets $\mathcal{F}_{i_1 i_2}$ ($1 \le i_1 < i_2 \le 4$) of $\mathcal{F}$ are

$$\mathcal{F}_{12} = \left\{H_{12}^0,\ H_{123}^0,\ H_{124}^0,\ H_{1234}^0\right\}, \quad \mathcal{F}_{13} = \left\{H_{13}^0,\ H_{123}^0,\ H_{134}^0,\ H_{1234}^0\right\},$$

$$\mathcal{F}_{14} = \left\{H_{14}^0,\ H_{124}^0,\ H_{134}^0,\ H_{1234}^0\right\}, \quad \mathcal{F}_{23} = \left\{H_{23}^0,\ H_{123}^0,\ H_{234}^0,\ H_{1234}^0\right\},$$

$$\mathcal{F}_{24} = \left\{H_{24}^0,\ H_{124}^0,\ H_{234}^0,\ H_{1234}^0\right\}, \quad \mathcal{F}_{34} = \left\{H_{34}^0,\ H_{134}^0,\ H_{234}^0,\ H_{1234}^0\right\}.$$

The contingency table for computation of Cochran-Mantel-Haenszel (CMH) test is constructed by choosing strata $x^r = -0.2 + 0.05(r-1); (r = 1, \cdots, 9)$. The contingency table is shown in Table 1, where data are presented only for strata $r = 1, 5, 9$ in the case of the null hypothesis H_{1234}^0. We adopt Ryan-Einot-Gabriel-Welsch (REGW) significance levels for the multiple test of null hypotheses H_K^0. We test hypotheses with the type I familywise error rate $\alpha = 0.05$. Then REGW significance levels are $\alpha_2 = 0.0253$ and $\alpha_3 = \alpha_4 = 0.05$. The results of multiple tests are shown in Table 2, where "s" and "n" indicate significant (H_K^0 is rejected) and not significant (H_K^0 is not rejected), respectively, and p_K and cum(q_K) are p-values and cumulative powers, respectively, for the null hypotheses H_K^0. We have that null hypotheses H_{12}^0 and H_{34}^0 are not significant, and the cumulative powers are very small (0.065, 0.112) for H_{12}^0 and H_{34}^0. On the other hand, H_{23}^0 is significant, but the cumulative power is not enough large (0.493).

We also compute the delta test statistic based on psychophysical thresholds and their variances of groups Γ_k ($k = 1, \cdots, 4$). The mean values of logistic regression results for each group Γ_k ($k = 1, \cdots, 4$) are shown Tables 3 (intercepts $\overline{a}_k$ and slopes $\overline{b}_k$) and 4 (psychophysical thresholds $\overline{\xi}_k$). In Table 3, $p(\overline{b}_k)$ and $q(\overline{b}_k)$ are p-values and powers for type I error rate $\alpha = 0.05$, respectively, where the null hypotheses are $\overline{b}_k = 0$. In Table 4, min($\overline{\xi}_k$) and max($\overline{\xi}_k$) are lower and upper bounds of 95% confidence intervals of $\overline{\xi}_k$, respectively. The representative psychometric functions $\overline{\varphi}_k(x)$ ($k = 1, \cdots, 4$) for groups Γ_k ($k = 1, \cdots, 4$) are shown in Fig. 5, where paremeters $\overline{a}_k$ and $\overline{b}_k$ are adopted in the values of Table 3. The results of multiple test by delta test statistic are shown in Table 5. We can say that results of CMH and delta test statistics are nearly equivalent.

7 Conclusion

We proposed a statistical modelling of the visual acuity measurement and its multiple test procedure. The main properties of our algorithm are summarized:

Step 1: modelling of visual acuity for individual patient by logistic regression including the guessing rate,

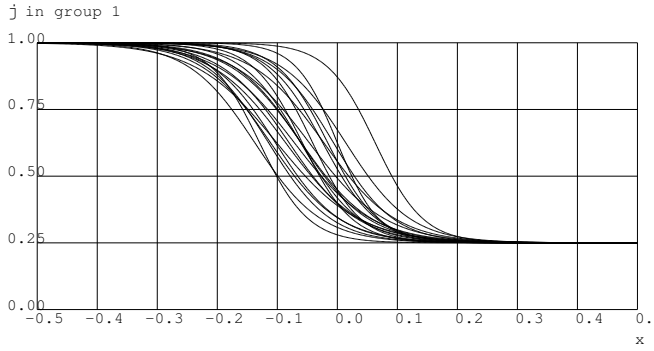

Figure 1: Psychometric functions φ_j $(j = 1, \cdots, 22)$ in group 1

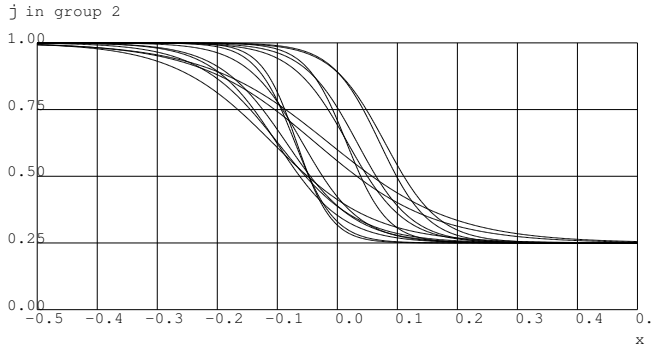

Figure 2: Psychometric functions φ_j $(j = 1, \cdots, 14)$ in group 2

Step 2a: computing the Cochran-Mantel-Haenszel (CMH) test statistic based on psychometric functions,
Step 2b: computing the delta test statistic based on psychophysical thresholds and their variances,
Step 3: multiple test based on the step-down procedure with Ryan-Einot-Gabriel-Welsch (REGW) significance levels.

We applied our algorithm to the visual acuity measurement of four group problem. We find that test results of Step 2a and Step 2b are nearly equivalent in this example. For real optometric application, more studies shall be implemented to know with certainty about the merit/demerit of proposed approaches: Step 2a (psychometric functions) and Step 2b (psychophysical thresholds).

Acknowledgement

The authors are grateful to the editor and anonymous reviewers for their valuable comments and suggestions.

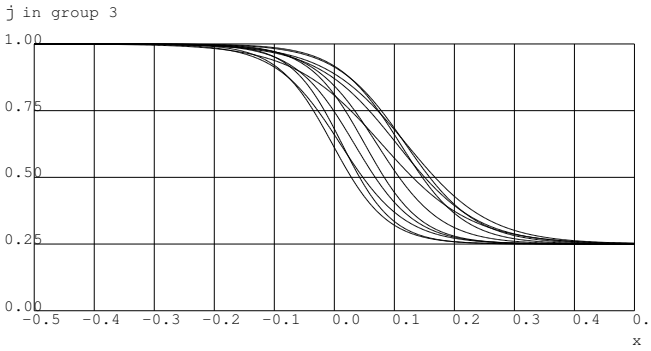

Figure 3: Psychometric functions φ_j $(j = 1, \cdots, 11)$ in group 3

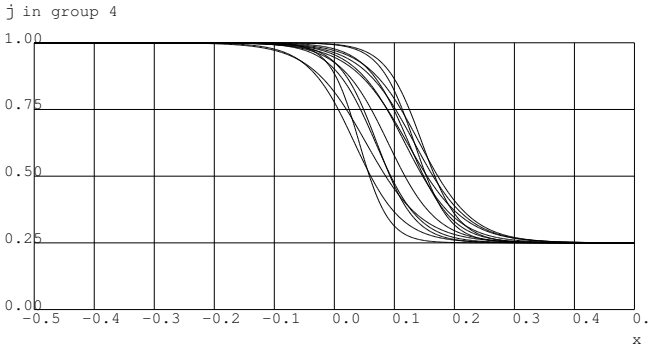

Figure 4: Psychometric functions φ_j $(j = 1, \cdots, 13)$ in group 4

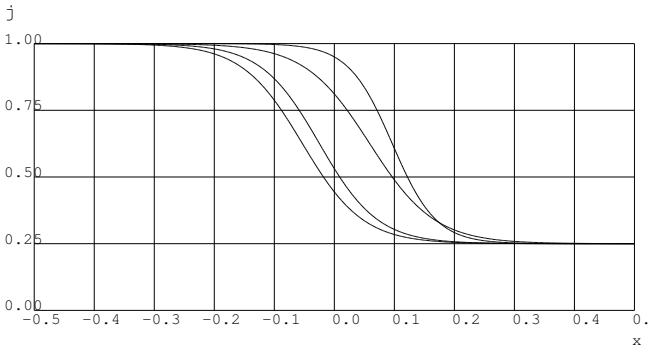

Figure 5: Representative psychometric functions $\overline{\varphi}_k$ (left to right: k=1, 2, 3, 4)

Table 1: The 4×2 contingency table based on psychometric functions

strata	groups	$\ell = 1$ (correct) u^r_{k1}	$\ell = 2$ (incorrect) u^r_{k2}	total
$x^1 = -0.2$	$k = 1\,(\Gamma_1)$	20.569	1.431	$n^1_1 = 22.000$
	$k = 2\,(\Gamma_2)$	13.182	0.818	$n^1_2 = 14.000$
	$k = 3\,(\Gamma_3)$	10.918	0.082	$n^1_3 = 11.000$
	$k = 4\,(\Gamma_4)$	12.991	0.009	$n^1_4 = 13.000$
	total	$m^1_1 = 57.660$	$m^1_2 = 2.340$	$t^1 = 60.000$
		.....		
$x^5 = 0.0$	$k = 1\,(\Gamma_1)$	10.283	11.717	$n^5_1 = 22.000$
	$k = 2\,(\Gamma_2)$	7.711	6.289	$n^5_2 = 14.000$
	$k = 3\,(\Gamma_3)$	8.748	2.252	$n^5_3 = 11.000$
	$k = 4\,(\Gamma_4)$	12.030	0.970	$n^5_4 = 13.000$
	total	$m^5_1 = 38.772$	$m^5_2 = 21.228$	$t^5 = 60.000$
		.....		
$x^9 = 0.2$	$k = 1\,(\Gamma_1)$	5.670	16.330	$n^9_1 = 22.000$
	$k = 2\,(\Gamma_2)$	3.800	10.200	$n^9_2 = 14.000$
	$k = 3\,(\Gamma_3)$	3.616	7.384	$n^9_3 = 11.000$
	$k = 4\,(\Gamma_4)$	4.144	8.856	$n^9_4 = 13.000$
	total	$m^9_1 = 17.230$	$m^9_2 = 42.770$	$t^9 = 60.000$

Table 2: Multiple test based on psychometric functions

	groups	REGW	critical values	CMH	p-values		powers
H_K^0	κ	α_κ	$\chi^2_{\kappa-1}$	Ψ_K	p_K	test	cum(q_K)
H_{1234}^0	4	0.05	7.814	32.673	3.778×10^{-7}	s	0.999
H_{234}^0	3	0.05	5.991	16.943	2.094×10^{-4}	s	0.965
H_{134}^0	3	0.05	5.991	30.529	2.348×10^{-7}	s	0.999
H_{124}^0	3	0.05	5.991	26.243	2.002×10^{-6}	s	0.997
H_{123}^0	3	0.05	5.991	12.924	1.562×10^{-3}	s	0.905
H_{34}^0	2	0.0253	5.002	1.082	2.982×10^{-1}	n	0.112
H_{24}^0	2	0.0253	5.002	14.080	1.752×10^{-4}	s	0.900
H_{23}^0	2	0.0253	5.002	5.743	1.655×10^{-2}	s	0.493
H_{14}^0	2	0.0253	5.002	23.918	1.006×10^{-6}	s	0.992
H_{13}^0	2	0.0253	5.002	11.709	6.220×10^{-4}	s	0.798
H_{12}^0	2	0.0253	5.002	0.592	4.416×10^{-1}	n	0.065

Table 3: Mean values of logistic regression results (intercepts and slopes)

	patients	intercepts	slopes	variances	p-values	powers
groups	N_k	$\overline{a}_k$	$\overline{b}_k$	var$(\overline{b}_k)$	$p(\overline{b}_k)$	$q(\overline{b}_k)$
Γ_1	22	-1.052	-19.823	38.951	1.454×10^{-3}	0.888
Γ_2	14	-0.510	-20.568	41.657	1.401×10^{-3}	0.890
Γ_3	11	1.089	-18.478	23.451	1.212×10^{-4}	0.968
Γ_4	13	2.641	-27.370	60.263	3.964×10^{-4}	0.941

$q(\overline{b}_k)$: powers of $\overline{b}_k$ for $\alpha = 0.05$, where the null hypotheses are $\overline{b}_k = 0$.

Table 4: Mean values of logistic regression results (psychophysical thresholds)

	patients	thresholds	variances	SD		
groups	N_k	$\overline{\xi}_k$	var$(\overline{\xi}_k)$	$\overline{\sigma}_{\xi_k}$	min$(\overline{\xi}_k)$	max$(\overline{\xi}_k)$
Γ_1	22	-0.0570	3.342×10^{-4}	0.0183	-0.0928	-0.0212
Γ_2	14	-0.0301	3.817×10^{-4}	0.0195	-0.0684	0.0082
Γ_3	11	0.0642	3.915×10^{-4}	0.0198	0.0254	0.1030
Γ_4	13	0.0981	2.474×10^{-4}	0.0157	0.0672	0.1289

min$(\overline{\xi}_k)$ and max$(\overline{\xi}_k)$ are lower and upper bounds of 95% confidence intervals

Table 5: Multiple test based on psychophysical thresholds

	groups	REGW	critical values	delta	p-values		powers
H_K^0	κ	α_κ	$\chi^2_{\kappa-1}$	Δ_K	p_K	test	cum(q_K)
H_{1234}^0	4	0.05	7.814	53.358	1.086×10^{-3}	s	1.000
H_{234}^0	3	0.05	5.991	26.620	1.658×10^{-6}	s	0.998
H_{134}^0	3	0.05	5.991	43.270	4.017×10^{-10}	s	1.000
H_{124}^0	3	0.05	5.991	48.729	2.622×10^{-11}	s	1.000
H_{123}^0	3	0.05	5.991	21.737	1.905×10^{-5}	s	0.991
H_{34}^0	2	0.0253	5.002	1.793	1.806×10^{-1}	n	0.184
H_{24}^0	2	0.0253	5.002	26.114	3.221×10^{-7}	s	0.996
H_{23}^0	2	0.0253	5.002	11.509	6.928×10^{-4}	s	0.866
H_{14}^0	2	0.0253	5.002	41.339	4.695×10^{-7}	s	1.000
H_{13}^0	2	0.0253	5.002	20.246	6.809×10^{-6}	s	0.979
H_{12}^0	2	0.0253	5.002	1.009	3.151×10^{-1}	n	0.109

References

Agresti, A. (2002). *Categorical data analysis, Second Edition.* Hoboken (NJ): Wiley.

Bach, M. (1996). The Freiburg visual acuity test – automatic measurement of visual acuity. *Optometry and Vision Science*, 73, 49–53.

Beck, R.W., Mokexi, P.S., Turpin, A.H., Ferris 3rd, F.L., SanGiovanni, J.P., Johnson, C.A., Birch, E.E., Chandler, D.L., Cox, T.A., Blair, R.C. and Kraker, R.T. (2003). A computerized method of visual acuity testing: Adaptation of the early treatment of diabetic retinopathy study testing protocol. *American Journal of Ophthalmology*, 135, 194–205.

Einot, I. and Gabriel, K.R. (1975). A study of the powers of several methods of multiple comparisons. *Journal of the American Statistical Association*, 70, 574–583.

Félix, V.B. and Menezes, A.F.B. (2018). Comparisons of ten corrections methods for t-test in multiple comparisons via Monte Carlo study. *Electronic Journal of Applied Statistical Analysis*, 11, 74–91.

Holland, P.W. and Thayer, D.T. (1988). Differential item performance and the Mantel-Haenszel procedure. In Wainer, H. and Braun, H.I., eds. *Test Validity*, pages 129–145. Hillsdale (NJ): Lawrence Erlbaum Associates.

Kingdom, F.A.A. and Prins, N. (2010). *Psychophysics – A Practical Introduction.* London: Academic Press.

Landis, J.R., Heyman, E.R. and Koch, G.G. (1978). Average partial association in three-way contingency tables: a review and discussion of alternative tests. *International Statistical Review*, 46, 237-254.

Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719-748.

Mita, N., Hara, H., Kani, K. and Tabuchi, A. (2010). Use of statistical analysis for visual acuity measurement (in Japanese). *Japanese Journal of Visual Science*, 31, 19–25.

Mita, N., Sasaki, H., Kani, K., Tabuchi, A. and Hara, H. (2014). Multiple comparisons for a psychophysical test in bootstrap logistic regression. *Journal of Algorithms and Computational Technology*, 8, 339-355.

Mita, N., Sasaki, H., Kani, K., Tabuchi, A. and Hara, H. (2017). Multiple comparisons in two-factor testing of psychophysical thresholds by the delta test method. *Communications in Statistcs – Simulation and Computation*, 46, 669-678.

Nagai, T., Hoshino, T. and Uchikawa, K. (2006). Statistical significance testing of thresholds estimated by constant stimuli method (in Japanese). *Vision*, 18, 113–123.

Pas, S.F. and Koenderink, J.J. (2004). Visual descrimination of spectral distributions. *Perception*, 33, 1483–1497.

Penfield, R.D. (2001). Assesing differential item functioning among multiple groups: A comparison of three Mantel-Haenszel procedures. *Applied Measurement in Education*, 14, 235–259.

Ryan, T.A. (1960). Significance tests for multiple comparison of proportions, variances

and other statistics. *Psychological Bulletin*, 57, 318–328.

Schulze-Bonsel, K., Feltgen, N., Burau, H., Hansen, L. and Bach, M. (2006). Visual acuities "hand motion" and "counting fingers" can be quatified with the Freiburg visual acuity test. *Investigative Ophthalmology and Visual Science*, 47, 1236–1240.

Simpson, T.L. (1995). Vision thresholds from psychometric analyses: alternatives to probit analysis. *Optometry and Visual Science*, 72, 371–377.

Somes, G.W. (1986). The generalized Mantel-Haenszel Statistic. *Amercan Statistician*, 40, 106–108.

Tokutake, T., Yamauchi, T., Mita, N., Kawamoto, K., Tabuchi, H., Kani, K., Hara, H., Tabuchi, A. and Sasaki, H. (2014). Repeatability of visual acuity testing using a psychometric function. *Japanese Journal of Ophthalmology*, 58, 381-387.

Watson, A.B. (1979). Probability summation over time. *Vision Research*, 19, 515–522.

Welsch, R.E. (1977). Stepwise multiple comparison procedures. *Journal of the American Statistical Association*, 72, 566–575.

Zhang, J. and Boos, D.D. (1996). Generalized Cochran-Mantel-Haenszel test statistics for correlated categorical data. *Institute of Statistics Mimeo Series*, No. 2284.