REVSTAT – Statistical Journal
Volume 21, Number 2, April 2023, 267–281
https://doi.org/10.57805/revstat.v21i2.401

Median Distance Model for Likert-Type Items in Contingency Table Analysis

Authors: Serpil Aktas Altunay ✉
– Department of Statistics, Hacettepe University, Ankara, Turkey
spxl@hacettepe.edu.tr

Ayfer Ezgi Yilmaz
– Department of Statistics, Hacettepe University, Ankara, Turkey
ezgiyilmaz@hacettepe.edu.tr

Received: April 2021 Revised: November 2021 Accepted: November 2021

Abstract:

- Likert-type items (questions) are a widely used scale in questionnaire design. The "neutral" or "undecided" option may lead to misinterpretation and confusion about the results. This paper proposes two novel log-linear models to measure how much accumulation of the neutral option over the contingency tables at any question levels. These models also test the odds that a respondent's level how far from the median. These models will help the researchers how to incorporate the neutral option in conceptual frameworks.

Keywords:

- *Likert-type items; association models; ordinal variables; contingency table; log-linear models.*

AMS Subject Classification:

- 49A05, 78B26.

✉ Corresponding author.

1. INTRODUCTION

Likert-type scales or formally ordinal scales are psychometric scales used when there is an order in responses and distances between categories are not quantitative [4, 14]. Likert scale is widely used in medical, education, and many disciplines in social sciences.

There is a difference between the terms of Likert-type items and Likert scales [20]. Likert items are the single questions that use some aspect of the original Likert response alternatives and several of them built a Likert scale [10]. In this study, Likert-type items are considered as a part of a scale or not.

Likert-type items are usually formed in five responses: “1: strongly disagree”, “2: disagree”, “3: neutral”, “4: agree”, “5: strongly agree”. Similarly, a 7-point Likert scale includes seven responses such as; “1: strongly disagree”, “2: disagree”, “3: somewhat disagree”, “4: neither agree nor disagree”, “5: somewhat agree”, “6: Agree”, “7: strongly agree”.

The attitudes change from mildly positive to mildly negative. The neutral option that is sometimes referred to as “neither agree nor disagree” or “undecided” on a Likert scale means that respondents are not willing to answer a particular question or have no idea.

With regard to the neutral point on the scale, we should be aware that neutral does not imply the midpoint between the two extreme-scale scores.

Those respondents who check the neutral option might mislead the results and the main point might not be achieved. Hence, the question “neutral responses will be omitted or how to handle with neutral questions?” matter. In some surveys that there is often no neutral category included in the middle of the scale [7]. Sometimes it is placed at the end of the scale, and sometimes it is eliminated directly. The neutral means is the median or mid-point and the median is the 50% sample distribution and it means 50% of the participants have neutral to agree with opinions in a 5-point Likert scale. If the median is 4, it means 50% of participants have a positive opinion. The ordinal structure and the existence of a neutral category should be considered to model the Likert items. Despite the independence of the two Likert-type items is analyzed with the chi-square test, it does not accept the ordinal structure of the items. Linear-by-linear association model and its special form uniform association model are used to analyze the association between the variables of a contingency table with ordered categories [1, 8]. There are many extensions of association models (e.g. [5, 6, 18, 21, 22]). Even though all these models consider the ordinal structure of the variables, they ignore the ambiguous nature of the neutral category and treat it as if the neutral category has the same structure as other categories. Truebner [19] showed that changes in respondents’ characteristics do not affect median response with the exception of age. Even though the intervals between the categories should be regarded as subjectively equal, Oppenheim [15] states that “attitudes may be shaped more like concentric circles or overlapping ellipses or three-dimensional cloud formations, therefore, the model of the linear continuum or dimension is not always easy or appropriate”.

2. MATERIAL AND METHOD

A contingency table summarizes information of two or higher dimensions random variables. An example of the contingency table is given in Table 1 for the first question (Q_1) and second question (Q_2) in a questionnaire.

Table 1: Two-way classification table for a 5-point Likert scale questions.

Q_1	Q_2					**Total**
	1	2	3	4	5	
1	n_{11}	n_{12}	n_{13}	n_{14}	n_{15}	n_{1+}
2	n_{21}	n_{22}	n_{23}	n_{24}	n_{25}	n_{2+}
3	n_{31}	n_{32}	n_{33}	n_{34}	n_{35}	n_{3+}
4	n_{41}	n_{42}	n_{43}	n_{44}	n_{45}	n_{4+}
5	n_{51}	n_{52}	n_{53}	n_{54}	n_{55}	n_{5+}
Total	n_{+1}	n_{+2}	n_{+3}	n_{+4}	n_{+5}	n

Consider a two-way table in which both the row and column variables have R categories (levels). R denotes the R-point Likert scale. In an $R\times R$ table, n_{ij}'s denote the cell frequencies for the ith row and jth column where $i = 1, ..., R$. n_{i+} and n_{+j} are the row and column totals, respectively, satisfying

$$\sum_{i=1}^{R} n_{i+} = \sum_{j=1}^{R} n_{+j} = n.$$

The goal of the log-linear analysis is to determine which categorical variables represent the data. Log-linear models do not distinguish between response and explanatory variables. All variables in a log-linear model are treated as responses.

The relationship between two or more variables is examined in analyzing contingency tables. We will refer to the variables in two-way contingency tables as "question". In a two-way $R\times R$ contingency table, let $\{\mu_{ij}\}$ be the expected values corresponding to the observed values. The independence model for any pair of items is commonly defined for the two questions in Equation (2.1).

$$\text{Log}(\mu_{ij}) = \lambda + \lambda_i^{Q_1} + \lambda_j^{Q_2} + \lambda_{ij}^{Q_1 Q_2}, \qquad i, j = 1, ..., R, \tag{2.1}$$

where λ is the intercept term (overall mean of the natural log of the expected values), $\lambda_i^{Q_1}$ is the main effect for question Q_1, $\lambda_j^{Q_2}$ is the main effect for question Q_2, and $\lambda_{ij}^{Q_1 Q_2}$ is the interaction term. The parameters are set to satisfy the following restrictions:

$$\sum_{i=1}^{R} \lambda_i^{Q_1} = \sum_{j=1}^{R} \lambda_j^{Q_2} = \sum_{i=1}^{R}\sum_{j=1}^{R} \lambda_{ij}^{Q_1 Q_2} = 0.$$

Because concluding that respondents are neutral might be inaccurate, we suggest two models that measure the variability around the neutral option, namely in the third group for the 5-point Likert-type and the fourth group for the 7-point Likert-type as shown in Figures 1–3.

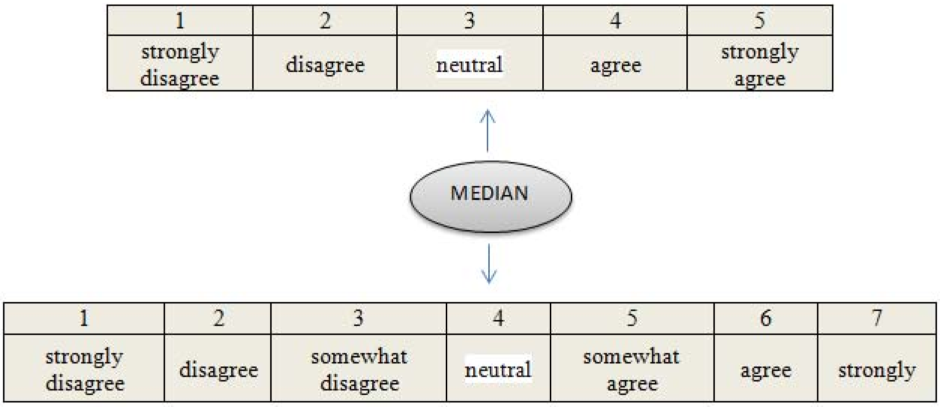

Figure 1: The position of the median in 5- and 7- point Likert scales.

Q_1 \ Q_2	1	2	3	4	5
1	n_{11}	n_{12}	n_{13}	n_{14}	n_{15}
2	n_{21}	n_{22}	n_{23}	n_{24}	n_{25}
3	n_{31}	n_{32}	n_{33}	n_{34}	n_{35}
4	n_{41}	n_{42}	n_{43}	n_{44}	n_{45}
5	n_{51}	n_{52}	n_{53}	n_{54}	n_{55}

Figure 2: Variability around the median in a 5×5 table.

Q_1 \ Q_2	1	2	3	4	5	6	7
1	n_{11}	n_{12}	n_{13}	n_{14}	n_{15}	n_{16}	n_{17}
2	n_{21}	n_{22}	n_{23}	n_{24}	n_{25}	n_{26}	n_{27}
3	n_{31}	n_{32}	n_{33}	n_{34}	n_{35}	n_{36}	n_{37}
4	n_{41}	n_{42}	n_{43}	n_{44}	n_{45}	n_{46}	n_{47}
5	n_{51}	n_{52}	n_{53}	n_{54}	n_{55}	n_{56}	n_{57}
6	n_{61}	n_{62}	n_{63}	n_{64}	n_{65}	n_{66}	n_{67}
7	n_{71}	n_{72}	n_{73}	n_{74}	n_{75}	n_{76}	n_{77}

Figure 3: Variability around the median in a 7×7 table.

The median of an R categories is calculated as

$$m = \frac{R+1}{2},$$

and median cell implies that the cell falls into the (m, m). The median cell falls into the (3,3) cell for a 5×5 table, fall into the (4,4) cell for a 7×7 table.

We built two novel log-linear models taking the main effects (Q_1, Q_2), association parameter, and distance parameter. The simple model is the Median Distance (MD) model as

$$\text{Log}(\mu_{ij}) = \lambda + \lambda_i^{Q_1} + \lambda_j^{Q_2} + \delta_{ij}, \qquad i, j = 1, ..., R. \tag{2.2}$$

The parameter δ is the median distance parameter which is defined in Equation (2.3) and the method to identify the log-linear parameters involves fixing the parameters to zero for one category of Q_1 and Q_2, respectively. For an $R\times R$ table, the MD model has m median distance parameters:

$$\delta_{ij} = \begin{cases} \delta_1, & i=j=m \text{ (median cell)},\\ \delta_2, & \text{one-step distance from the median cell},\\ \delta_3, & \text{two-step distance from the median cell},\\ \vdots & \quad\vdots \\ \delta_{m-1}, & (m-2)\text{-step distance from the median cell},\\ \delta_m, & (m-1)\text{-step distance from the median cell}. \end{cases} \tag{2.3}$$

For example, the light gray shaded area in Figure 2 represents one step from the midpoint, and the dark gray shaded area shows the two-step distance from the midpoint. The median distance parameters are set to satisfy the following restriction:

$$\sum_{i=1}^{m} \delta_i = 0.$$

This model has more $(m-1=(R-1)/2)$ parameters than the independence model, the residual degrees of freedom under the MD model is

$$\begin{aligned} df &= R\times R - \left[1-(R-1)+(R-1)+\left(\frac{R+1}{2}-1\right)\right] \\ &= \frac{2R^2-5R+3}{2}. \end{aligned}$$

The odds ratios matrix under the MD model for a 5×5 table is shown below

$$\begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} & \theta_{14} \\ \theta_{21} & \theta_{22} & \theta_{23} & \theta_{24} \\ \theta_{31} & \theta_{32} & \theta_{33} & \theta_{34} \\ \theta_{41} & \theta_{42} & \theta_{43} & \theta_{44} \end{bmatrix} = \exp \begin{bmatrix} \delta_2-\delta_3 & 1 & 1 & \delta_3-\delta_2 \\ 1 & \delta_1-\delta_2 & \delta_2-\delta_1 & 1 \\ 1 & \delta_2-\delta_1 & \delta_1-\delta_2 & 1 \\ \delta_3-\delta_2 & 1 & 1 & \delta_2-\delta_3 \end{bmatrix}$$

and for a 7×7 table is given as

$$\begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} & \theta_{14} & \theta_{15} & \theta_{16} \\ \theta_{21} & \theta_{22} & \theta_{23} & \theta_{24} & \theta_{25} & \theta_{26} \\ \theta_{31} & \theta_{32} & \theta_{33} & \theta_{34} & \theta_{35} & \theta_{36} \\ \theta_{41} & \theta_{42} & \theta_{43} & \theta_{44} & \theta_{45} & \theta_{46} \\ \theta_{51} & \theta_{52} & \theta_{53} & \theta_{54} & \theta_{55} & \theta_{56} \\ \theta_{61} & \theta_{62} & \theta_{63} & \theta_{64} & \theta_{65} & \theta_{66} \end{bmatrix} = \exp \begin{bmatrix} \delta_3-\delta_4 & 1 & 1 & 1 & 1 & \delta_4-\delta_3 \\ 1 & \delta_1-\delta_3 & 1 & 1 & \delta_3-\delta_2 & 1 \\ 1 & 1 & \delta_1-\delta_2 & \delta_2-\delta_1 & 1 & 1 \\ 1 & 1 & \delta_2-\delta_1 & \delta_1-\delta_2 & 1 & 1 \\ 1 & \delta_3-\delta_2 & 1 & 1 & \delta_1-\delta_3 & 1 \\ \delta_4-\delta_3 & 1 & 1 & 1 & 1 & \delta_3-\delta_4 \end{bmatrix}$$

When both the column and row variables of a two-dimensional table are ordinal, a simple log-linear model that utilizes the orderings of the rows and the columns is the linear-by-linear association model [1]. This ordinarily of the data needs an extra parameter that gives the association of two ordinal variables. Hence, adding an association model to the MD model, the median distance + association (MDA) model is defined in a log-linear form as in Equation (2.4):

$$\text{(2.4)} \qquad \operatorname{Log}(\mu_{ij}) = \lambda + \lambda_i^{Q_1} + \lambda_j^{Q_2} + \beta u_{1i} u_{2j} + \delta_{ij}, \qquad i,j = 1,...,R,$$

where β is the linear-by-linear association parameter and δ is the median distance parameter which is defined in Equation (2.3). The necessity of reflecting the ordinarily of the variables, assigning scores to the ordinal categories are fulfilled by the row and column scores, by u_{1i} and u_{2j} scores. The integer scores, meanly $u_{1i}, u_{2j} = 1,...,R$ are the frequently used scores. This model has more one more parameter than the MA model, the residual degrees of freedom under the MDA model is

$$\begin{aligned} df &= R \times R - \left[1 - (R-1) + (R-1) + \left(\frac{R+1}{2} - 1\right) + 1\right] \\ &= \frac{2R^2 - 5R + 1}{2}. \end{aligned}$$

The matrix of odds ratios under the MDA model for a 5×5 table is

$$\begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} & \theta_{14} \\ \theta_{21} & \theta_{22} & \theta_{23} & \theta_{24} \\ \theta_{31} & \theta_{32} & \theta_{33} & \theta_{34} \\ \theta_{41} & \theta_{42} & \theta_{43} & \theta_{44} \end{bmatrix} = \exp \begin{bmatrix} \beta + \delta_2 - \delta_3 & \beta & \beta & \beta + \delta_3 - \delta_2 \\ \beta & \beta + \delta_1 - \delta_2 & \beta + \delta_2 - \delta_1 & \beta \\ \beta & \beta + \delta_2 - \delta_1 & \beta + \delta_1 - \delta_2 & \beta \\ \beta + \delta_3 - \delta_2 & \beta & \beta & \beta + \delta_2 - \delta_3 \end{bmatrix}.$$

The matrix of odds ratios under the MDA model for a 7×7 table is

$$\begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} & \theta_{14} & \theta_{15} & \theta_{16} \\ \theta_{21} & \theta_{22} & \theta_{23} & \theta_{24} & \theta_{25} & \theta_{26} \\ \theta_{31} & \theta_{32} & \theta_{33} & \theta_{34} & \theta_{35} & \theta_{36} \\ \theta_{41} & \theta_{42} & \theta_{43} & \theta_{44} & \theta_{45} & \theta_{46} \\ \theta_{51} & \theta_{52} & \theta_{53} & \theta_{54} & \theta_{55} & \theta_{56} \\ \theta_{61} & \theta_{62} & \theta_{63} & \theta_{64} & \theta_{65} & \theta_{66} \end{bmatrix} = \exp \begin{bmatrix} \beta + \delta_3 - \delta_4 & \beta & \beta & \beta & \beta & \beta + \delta_4 - \delta_3 \\ \beta & \beta + \delta_1 - \delta_3 & \beta & \beta & \beta + \delta_3 - \delta_2 & \beta \\ \beta & \beta & \beta + \delta_1 - \delta_2 & \beta + \delta_2 - \delta_1 & \beta + \beta & \beta \\ \beta & \beta & \beta + \delta_2 - \delta_1 & \beta + \delta_1 - \delta_2 & \beta & \beta \\ \beta & \beta + \delta_3 - \delta_2 & \beta & \beta & \beta + \delta_1 - \delta_3 & \beta \\ \beta + \delta_4 - \delta_3 & \beta & \beta & \beta & \beta & \beta + \delta_3 - \delta_4 \end{bmatrix}.$$

The goodness of fit hypothesis is tested by the likelihood ratio test statistic as

$$G^2 = 2 \sum_{i=1}^{R} \sum_{j=1}^{R} n_{ij} \log\left(\frac{n_{ij}}{\hat{\mu}_{ij}}\right).$$

Under the null hypothesis is true, likelihood ratio statistic has an asymptotic chi-square distribution with associated degrees of freedom.

The design matrix of the MDA model for a 5×5 table is constructed as below. If we subtracted the last column from the design matrix the MDA model would turn into the

MA model. This implies that the MD model has one less parameter than the MDA model:

$$
\log \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{14} \\ \mu_{15} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \\ \mu_{24} \\ \mu_{25} \\ \mu_{31} \\ \mu_{32} \\ \mu_{33} \\ \mu_{34} \\ \mu_{35} \\ \mu_{41} \\ \mu_{42} \\ \mu_{43} \\ \mu_{44} \\ \mu_{45} \\ \mu_{51} \\ \mu_{52} \\ \mu_{53} \\ \mu_{54} \\ \mu_{55} \end{bmatrix} = \begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & -1 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & -1 & 2 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & -1 & 3 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 4 \\
1 & 1 & 0 & 0 & 0 & -1 & -1 & -1 & -1 & -1 & -1 & 5 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & -1 & 2 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 4 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 6 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 8 \\
1 & 0 & 1 & 0 & 0 & -1 & -1 & -1 & -1 & -1 & -1 & 10 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & -1 & -1 & 3 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 6 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 9 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 12 \\
1 & 0 & 0 & 1 & 0 & -1 & -1 & -1 & -1 & -1 & -1 & 15 \\
1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & -1 & -1 & 4 \\
1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 8 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 12 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 16 \\
1 & 0 & 0 & 0 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & 20 \\
1 & -1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & -1 & -1 & 5 \\
1 & -1 & -1 & -1 & -1 & 0 & -1 & 0 & 0 & -1 & -1 & 10 \\
1 & -1 & -1 & -1 & -1 & 0 & 0 & -1 & 0 & -1 & -1 & 15 \\
1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & -1 & -1 & -1 & 20 \\
1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 25
\end{bmatrix} \begin{bmatrix} \lambda \\ \lambda_1^{Q_1} \\ \lambda_2^{Q_1} \\ \lambda_3^{Q_1} \\ \lambda_4^{Q_1} \\ \lambda_1^{Q_2} \\ \lambda_2^{Q_2} \\ \lambda_3^{Q_2} \\ \lambda_4^{Q_2} \\ \delta_1 \\ \delta_2 \\ \beta \end{bmatrix}.
$$

3. NUMERICAL EXAMPLES

In this section, we provide three data sets to illustrate the methods presented in this paper. Two of these data sets are artificial and one is real-life data. The observed frequencies in the artificial tables were generated so that the data set fits the model adequately, by adjusted according to the expected frequencies calculated under the models hold true. Models are applied to these numerical examples and the results are highlighted for the researchers to be able to understand and interpret the information more strategically and usefully. The models were analyzed using "General Log-linear models" in IBM SPSS 23 by entering the design matrix properly. In the design matrix, the δ and β parameters are defined as the covariates [12].

Example 1

An artificial 5×5 contingency table is given in Table 2 which displays for any two questions from a questionnaire, say Q_1 and Q_2.

The Independence, symmetry, quasi-symmetry, MD, and MDA models are applied to the data in Table 2 and the log-linear model results are summarized in Table 3 (see [1] and [3] for the details of symmetry and quasi-symmetry models). The quasi-symmetry, MD, and MDA models fit data (Table 3, $p>0.05$). The quasi-symmetry model implies that there is an agreement between Q_1 and Q_2.

Table 2: The frequencies (expected values) of a 5×5 table.

Q_1	Q_2					**Total**
	1	2	3	4	5	
1	8 (5.89)	6 (8.82)	9 (11.49)	11 (11.32)	13 (9.48)	47
2	15 (15.56)	51 (50.84)	65 (66.25)	67 (65.28)	25 (25.06)	223
3	18 (15.79)	51 (51.57)	150 (150.00)	67 (66.22)	23 (25.06)	309
4	13 (14.24)	46 (46.51)	59 (60.60)	61 (59.72)	25 (22.93)	204
5	5 (7.52)	15 (11.26)	20 (14.67)	11 (14.45)	9 (12.11)	60
Total	59	169	303	217	95	843

Table 3: Model results for the 5×5 table.

Model	G^2	***df***	***p*-value**	**AIC**	**BIC**
Independence	54.065	16	<0.001	—	—
Symmetry	25.347	10	0.005	—	—
Quasi-symmetry	7.118	6	0.310	−4.882	−33.304
MD	10.062	14	0.758	−17.938	−84.256
MDA	9.691	13	0.719	−16.309	−77.890

Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) [2, 17] are calculated for quasi-symmetry, MD, and MDA models to find the best fitting model to data. The MD model has the lowest AIC and BIC, it is considered as the best-fitted model. The parameter estimates under the MD model are summarized in Table 4.

Table 4: The parameter estimates under the MD model for the 5×5 table.

Parameter	**Estimate**	**Std. Error**	**Z**	***p*-value**	**95% CI**
Constant	3.282	0.320	10.253	<0.001	[2.654; 3.909]
$[Q_1 = 1]$	−0.244	0.195	−1.254	0.210	[−0.626; 0.138]
$[Q_1 = 2]$	0.728	0.218	3.336	0.001	[0.300; 1.155]
$[Q_1 = 3]$	0.742	0.220	3.374	0.001	[0.311; 1.173]
$[Q_1 = 4]$	0.639	0.219	2.915	0.004	[0.209; 1.068]
$[Q_1 = 5]$	$0^{(a)}$				
$[Q_2 = 1]$	−0.476	0.166	−2.874	0.004	[−0.801; −0.151]
$[Q_2 = 2]$	−0.073	0.226	−0.322	0.747	[−0.516; 0.370]
$[Q_2 = 3]$	0.192	0.225	0.853	0.394	[−0.249; 0.632]
$[Q_2 = 4]$	0.177	0.223	0.793	0.428	[−0.260; 0.615]
$[Q_2 = 5]$	$0^{(a)}$				
δ_1	0.795	0.119	6.683	<0.001	[0.562; 1.029]
δ_2	−0.008	0.096	−0.079	0.937	[−0.196; 0.181]

(a): This parameter is set to zero because it is redundant.

The δ_k parameters in Equation (2.2) have straightforward interpretations in terms of departures from the median category. The distance parameter estimates are $\hat{\delta}_1 = 0.795$, $\hat{\delta}_2 = -0.008$, and $\hat{\delta}_3 = 0 - [\hat{\delta}_1 + \hat{\delta}_2] = -0.787$. Odds ratios are calculated either taking the expected values in Table 2 or from the parameter estimates under the underlying model given in Table 4. For example for $\hat{\theta}_{11}$ is obtained as:

$$\hat{\theta}_{11} = \frac{5.89 \times 50.84}{8.82 \times 15.56} = \exp(\hat{\delta}_2 - \hat{\delta}_3) = 2.18.$$

This can be interpreted as: the respondent's response is 2.18 times more likely to fall into the neutral category than a category two-step away from the median category. The matrix of odds ratios:

$$\hat{\theta} = \begin{bmatrix} 2.18 & 1 & 1 & 0.46 \\ 1 & 2.23 & 0.45 & 1 \\ 1 & 0.45 & 2.23 & 1 \\ 0.46 & 1 & 1 & 2.18 \end{bmatrix}$$

$\exp(\hat{\delta}_1 - \hat{\delta}_2) = 2.23$ can be interpreted as: a respondent's response is 2.23 times more likely to fall into the neutral category than a category one-step away from the median category.

Example 2

Table 5 displays an artificial 7×7 contingency tables for any two questions from a questionnaire, say Q_1 and Q_2.

Table 5: The frequencies (expected values) of an hypothetical 7×7 table.

Q_1	Q_2							**Total**
	1	2	3	4	5	6	7	
1	5 (9.79)	15 (14.76)	18 (17.78)	24 (19.18)	22 (20.42)	13 (15.18)	9 (8.88)	106
2	16 (14.41)	27 (28.58)	32 (33.46)	29 (35.07)	45 (36.28)	28 (26.21)	8 (11.00)	125
3	17 (15.83)	21 (30.50)	75 (73.49)	80 (74.84)	82 (75.24)	21 (24.94)	9 (10.17)	305
4	20 (19.33)	45 (36.19)	87 (84.73)	95 (95.00)	70 (81.91)	27 (26.38)	10 (10.46)	354
5	21 (20.24)	40 (36.82)	82 (83.77)	80 (80.56)	75 (76.47)	25 (23.93)	8 (9.22)	331
6	19 (18.32)	32 (32.38)	35 (33.80)	31 (31.58)	25 (29.13)	21 (18.76)	8 (7.02)	171
7	10 (10.07)	12 (12.78)	11 (12.97)	9 (11.77)	11 (10.55)	7 (6.60)	8 (3.25)	68
Total	108	192	340	348	330	142	60	1520

The Independence, symmetry, quasi-symmetry, MD, and MDA models are applied to the data in Table 5 and the log-linear model results are summarized in Table 6. The symmetry, quasi-symmetry, MD, and MDA models fit data (Table 6, $p>0.05$). The symmetry and quasi-symmetry model implies that there is an agreement between Q_1 and Q_2.

Table 6: Model results for the 7 × 7 table.

Model	G^2	*df*	*p*-value	AIC	BIC
Independence	87.455	36	<0.001	—	—
Symmetry	14.355	21	0.854	−27.645	−139.501
Quasi-symmetry	8.840	15	0.886	−21.160	−101.057
MD	33.818	33	0.428	−32.182	−207.955
MDA	25.648	32	0.779	−38.352	−208.799

AIC and BIC are calculated for symmetry, quasi-symmetry, MD, and MDA models. The MDA model has the lowest AIC and BIC, thus it is considered as the best-fitted model. The parameter estimates under the MDA model are summarized in Table 7.

Table 7: The parameter estimates under the MDA model for the 7 × 7 table.

Parameter	Estimate	Std. Error	Z	*p*-value	95% CI
Constant	3.219	0.505	6.369	<0.001	[2.228; 4.209]
$[Q_1 = 1]$	−0.200	0.275	−0.730	0.466	[−0.739; 0.338]
$[Q_1 = 2]$	0.215	0.292	0.734	0.463	[−0.358; 0.787]
$[Q_1 = 3]$	0.337	0.263	1.280	0.201	[−0.179; 0.853]
$[Q_1 = 4]$	0.566	0.245	2.308	0.021	[0.085; 1.046]
$[Q_1 = 5]$	0.640	0.232	2.761	0.006	[0.186; 1.095]
$[Q_1 = 6]$	0.569	0.233	2.446	0.014	[0.113; 1.025]
$[Q_1 = 7]$	$0^{(a)}$				
$[Q_2 = 1]$	−0.074	0.282	−0.263	0.793	[−0.628; 0.479]
$[Q_2 = 2]$	0.365	0.297	1.228	0.220	[−0.217; 0.947]
$[Q_2 = 3]$	0.580	0.267	2.172	0.030	[0.057; 1.103]
$[Q_2 = 4]$	0.684	0.249	2.749	0.006	[0.196; 1.172]
$[Q_2 = 5]$	0.775	0.235	3.298	0.001	[0.315; 1.236]
$[Q_2 = 6]$	0.508	0.238	2.129	0.033	[0.040; 0.975]
$[Q_2 = 7]$	$0^{(a)}$				
δ_1	0.545	0.120	4.541	<0.001	[0.310; 0.780]
δ_2	0.420	0.083	5.054	<0.001	[0.257; 0.583]
δ_3	−0.331	0.093	−3.537	<0.001	[−0.514; −0.147]
β	−0.029	0.010	−2.842	0.004	[−0.048; −0.009]

(a): This parameter is set to zero because it is redundant.

The negative value of β indicates that there is a negative relationship between Q_1 and Q_2 ($\hat{\beta} = -0.029$). The distance parameter estimates are $\hat{\delta}_1 = 0.545$, $\hat{\delta}_2 = 0.420$, $\hat{\delta}_3 = -0.331$, and $\hat{\delta}_4 = 0 - [\hat{\delta}_1 + \hat{\delta}_2 + \hat{\delta}_3] = -0.634$.

Odds ratios can be calculated over either the expected values in Table 5 or the parameter estimates in Table 7. For instance, $\hat{\theta}_{11}$ is calculated as

$$\hat{\theta}_{11} = \frac{9.79 \times 28.58}{14.76 \times 14.41} = \exp(\hat{\beta} + \hat{\delta}_3 - \hat{\delta}_4) = 1.32.$$

This can be interpreted as: the respondent's response is 1.32 times more likely to fall into the neutral category than a category three-step away from the median category, respectively. The matrix of odds ratios:

$$\hat{\theta} = \begin{bmatrix} 1.32 & 0.97 & 0.97 & 0.97 & 0.97 & 0.72 \\ 0.97 & 2.06 & 0.97 & 0.97 & 0.46 & 0.97 \\ 0.97 & 0.97 & 1.10 & 0.86 & 0.97 & 0.97 \\ 0.97 & 0.97 & 0.86 & 1.10 & 0.97 & 0.97 \\ 0.97 & 0.46 & 0.97 & 0.97 & 2.06 & 0.97 \\ 0.72 & 0.97 & 0.97 & 0.97 & 0.97 & 1.32 \end{bmatrix}$$

$\exp(\hat{\beta} + \hat{\delta}_1 - \hat{\delta}_2) = 1.10$ can be interpreted as: a respondent's response is 1.10 times more likely to fall into the neutral category than a category one-step away from the median category. The respondent's response is $\exp(\hat{\beta} + \hat{\delta}_1 - \hat{\delta}_3) = 2.06$ times more likely to fall into the neutral category than a category two-step away from the median category.

Real-Life Data

The study of hostel life data [16] is used to illustrate the proposed models. The project aims to measure the satisfaction level of the students towards facilities given in hostels. 5-point Likert items are used as: "1: very dissatisfied", "2: dissatisfied", "3: neutral", "4: satisfied", "5: very satisfied". Three items, "Overall Satisfaction about Hostel", "Management System of Mess", and "24 Hours Electricity" are selected. The answers of 184 students are given in Table 8.

Table 8: The study of hostel life data.

Overall Satisfaction	**Management System (24 Hours Electricity)**					**Total**
	1	2	3	4	5	
1	2 (0)	3 (2)	0 (5)	6 (5)	1 (0)	12
2	2 (3)	5 (3)	6 (8)	14 (14)	3 (2)	30
3	2 (1)	11 (13)	24 (10)	36 (35)	6 (20)	79
4	3 (2)	10 (3)	13 (12)	19 (20)	7 (15)	52
5	1 (0)	0 (0)	0 (2)	4 (3)	6 (6)	11
Total	10 (6)	29 (21)	43 (37)	79 (77)	23 (43)	184

The Independence, MD, and MDA models are applied to the overall satisfaction x management system of mess and overall satisfaction x 24 hour electricity tables. The log-linear model results are summarized in Table 9. For overall satisfaction x management system of mess table, both MD and MDA models fit the data well ($p>0.05$). For overall satisfaction x 24 hour electricity, only the MDA model fit the data well ($p>0.05$).

For overall satisfaction x management system of mess table, MD model has the lowest BIC and MDA model has the lowest AIC. We considered BIC. We follow the BIC results and decide that the MD model is the best-fitted model. The expected values under the best-fitted models are summarized in Table 10.

Table 9: Log-linear model results for hostel life data.

Table	Model	G^2	*df*	*p*-value	AIC	BIC
Overall satisfaction-Management system of mess	Independence	32.268	16	0.009	—	—
	MD	19.942	14	0.132	−8.058	−53.067
	MDA	17.750	13	0.167	−8.250	−50.044
Overall satisfaction-24 hours electricity	Independence	32.807	16	0.008	—	—
	MD	27.805	14	0.015	—	—
	MDA	16.426	13	0.227	−9.574	−51.368

Table 10: The expected values of hostel life data.

Overall Satisfaction	Management System (24 Hours Electricity)					Total
	1	2	3	4	5	
1	1.58 (1.07)	1.40 (2.58)	1.56 (4.00)	3.82 (3.37)	3.64 (0.98)	12
2	1.39 (1.58)	5.25 (4.65)	5.86 (9.39)	14.30 (10.27)	3.20 (4.12)	30
3	3.17 (2.60)	11.96 (9.93)	24.00 (10.00)	32.58 (37.11)	7.29 (19.37)	79
4	2.41 (0.69)	9.10 (3.41)	10.15 (11.66)	24.79 (21.58)	5.54 (14.65)	52
5	1.45 (0.06)	1.28 (0.44)	1.43 (1.94)	3.50 (4.67)	3.33 (3.89)	11
Total	10 (6)	29 (21)	43 (37)	79 (77)	23 (43)	184

The parameter estimates for overall satisfaction x management system of mess table under the MD model and overall satisfaction x 24 hours electricity table under the MDA model are summarized in Table 11 and Table 12, respectively.

Table 11: The parameter estimates under the MD model for overall satisfaction x management system of mess table.

Parameter	Estimate	Std. Error	Z	*p*-value	95% CI
Constant	2.366	0.635	3.728	<0.001	[1.122; 3.610]
$[Q_1 = 1]$	0.087	0.417	0.208	0.835	[−0.731; 0.905]
$[Q_1 = 2]$	−0.041	0.468	−0.088	0.930	[−0.958; 0.876]
$[Q_1 = 3]$	0.782	0.448	1.746	0.081	[−0.096; 1.660]
$[Q_1 = 4]$	0.509	0.452	1.125	0.261	[−0.378; 1.396]
$[Q_1 = 5]$	$0^{(a)}$				
$[Q_2 = 1]$	−0.833	0.379	−2.199	0.028	[−1.575; −0.091]
$[Q_2 = 2]$	−0.953	0.468	−2.038	0.042	[−1.870; −0.036]
$[Q_2 = 3]$	−0.844	0.480	−1.761	0.078	[−1.784; 0.095]
$[Q_2 = 4]$	0.049	0.444	0.110	0.913	[−0.821; 0.919]
$[Q_2 = 5]$	$0^{(a)}$				
δ_1	0.875	0.272	3.214	0.001	[0.341; 1.408]
δ_2	0.287	0.208	1.381	0.167	[−0.120; 0.695]

(a): This parameter is set to zero because it is redundant.

The distance parameter estimates in Table 11 are $\hat{\delta}_1 = 0.875$, $\hat{\delta}_2 = 0.287$, and $\hat{\delta}_3 = 0 - [\hat{\delta}_1 + \hat{\delta}_2 + \hat{\delta}_3] = -1.162$. The odds ratios of overall satisfaction x management system of mess table can be calculated by the expected values in Table 10 or by the parameter estimates in Table 11. For example, $\hat{\theta}_{11}$ is calculated as:

$$\hat{\theta}_{11} = \frac{1.58 \times 5.25}{1.40 \times 1.39} = \exp(\hat{\delta}_2 - \hat{\delta}_3) = 4.26.$$

This can be interpreted as: the student's response is 4.26 times more likely to fall into the neutral category than a category two-step away from the median category. The matrix of odds ratios for overall satisfaction x management system of mess table:

$$\hat{\theta} = \begin{bmatrix} 4.26 & 1 & 1 & 0.23 \\ 1 & 1.80 & 0.56 & 1 \\ 1 & 0.56 & 1.80 & 1 \\ 0.23 & 1 & 1 & 4.26 \end{bmatrix}$$

$\exp(\hat{\delta}_1 - \hat{\delta}_2) = 1.80$ can be interpreted as: a student's response is 1.80 times more likely to fall into the neutral category than a category one-step away from the median category.

Table 12: The parameter estimates under the MDA model for overall satisfaction x 24 hours electricity.

Parameter	Estimate	Std. Error	Z	p-value	95% CI
Constant	−5.581	2.209	−2.527	0.012	[−9.910; −1.252]
$[Q_1 = 1]$	3.882	1.239	3.132	0.002	[1.452; 6.311]
$[Q_1 = 2]$	4.005	1.133	3.536	<0.001	[1.785; 6.225]
$[Q_1 = 3]$	4.238	0.883	4.797	<0.001	[2.506; 5.969]
$[Q_1 = 4]$	2.643	0.633	4.176	<0.001	[1.403; 3.884]
$[Q_1 = 5]$	$0^{(a)}$				
$[Q_2 = 1]$	1.148	1.046	1.098	0.272	[−0.902; 3.198]
$[Q_2 = 2]$	1.760	1.006	1.750	0.080	[−0.211; 3.732]
$[Q_2 = 3]$	1.937	0.832	2.328	0.020	[0.306; 3.569]
$[Q_2 = 4]$	1.500	0.630	2.381	0.017	[0.265; 2.735]
$[Q_2 = 5]$	$0^{(a)}$				
δ_1	−0.660	0.316	−2.090	0.037	[−1.278; −0.041]
δ_2	0.300	0.231	1.297	0.195	[−0.153; 0.752]
β	0.263	0.083	3.177	0.001	[0.101; 0.425]

(a): This parameter is set to zero because it is redundant.

The distance parameter estimates in Table 12 are $\hat{\delta}_1 = -0.660$, $\hat{\delta}_2 = 0.300$, and $\hat{\delta}_3 = 0 - [\hat{\delta}_1 + \hat{\delta}_2 + \hat{\delta}_3] = -0.634$. Similarly, the odds ratios of overall satisfaction x 24 hour electricity table can be calculated either from the expected values in Table 10 or from the parameter estimates in Table 12. For the odds ratio $\hat{\theta}_{11}$, is obtained as:

$$\hat{\theta}_{11} = \frac{1.07 \times 4.65}{2.58 \times 1.58} = \exp(\hat{\beta} + \hat{\delta}_2 - \hat{\delta}_3) = 1.22.$$

The odds ratio can be interpreted as: the students' response is 1.22 times more likely to fall into the neutral category than a category two-step away from the median category. The matrix of odds ratios for overall satisfaction x 24 hours electricity table:

$$\hat{\theta} = \begin{bmatrix} 1.22 & 1.30 & 1.30 & 1.38 \\ 1.30 & 0.50 & 3.39 & 1.30 \\ 1.30 & 3.39 & 0.50 & 1.30 \\ 1.38 & 1.30 & 1.30 & 1.22 \end{bmatrix}$$

$1/\exp(\hat{\beta}+\hat{\delta}_1-\hat{\delta}_2)=2$ can be interpreted as: a respondent's response is 2 times more likely to fall into a category one-step away from the median category than the neutral category. The positive value of β means that there is a positive effect of 24 hour electricity on overall satisfaction about Hostel ($\hat{\beta}=0.263$).

4. CONCLUDING REMARKS

Attitudinal questions are a fundamental part of surveys in the social sciences. The items in a Likert scale are designed to measure respondent's attitudes to a particular question. Likert-type data is ordinal data, and a score is higher or lower than another. In any survey, if people feel that they really have no idea upon a question or feel that they are urged to make a choice, they choose the random or intentionally choose the neutral option. Neutral states that the respondent has neither a positive response nor a negative response. The researchers prefer to use a neutral category or midpoint so as to one side of which lay the favorable categories and to the other side the unfavorable categories. If the researcher does not set to a midpoint and respondents actually have a neutral opinion, they either tend to give a response that does not represent their actual attitude or avoid answering the question because the respondents sometimes tend to avoid using extreme categories. Essentially age and education are believed the two most relevant demographic factors which have been associated with a neutral option [13]. For instance, unlike the results that Harzing [11] showed that a higher neutral response for women than men, Grimm and Church [9] had found no gender effect.

The neutral point is the most difficult to locate and even more difficult to interpret. Moreover, the Likert scales tend to perform well with regard to a particular attitude of respondents that is in rough order. Assuming that we employ a 5-point or 7-point Likert scale and our questionnaire comprises a neutral option, with this regard we would mainly wish to know if there is any agglomeration in the neutral option. In fact, being in the neutral option would also imply that those users might be moved towards the satisfied group in some senses. This would cause a misinterpretation and deviates from the real context. Statistical modeling is a very essential part of data analysis. With this point of view, this paper proposes two log-linear models that take the ordinal information into account, besides the distance from the median category in Likert scale data. These models test whether the frequencies accumulate over the median group by subtracting the association. The distance parameters indicate that whether a subject is in favor to decide neutral, or measures how far a subject from the median. If the models hold true, the researcher will be able to draw conclusions from the evidence presented in the findings which are the results of the parameter estimates. It is noteworthy that the δ parameters and their associated odds ratios in the MDA model give evidence that how the frequencies in a two-way contingency table are distributed around the median category, moreover, how far the frequencies are from the median or midpoint. Interpretation of the log-odds coefficient gives the odds that a respondents' response falls in the median group than being an m-step distant from the midpoint category.

The models have a limitation that addresses the cognitive bias. As a consequence of cognitive bias, individuals make decisions according to their own perspectives, and therefore, cognitive biases may sometimes lead to inaccurate inferences or illogical interpretations. The impact of cognitive bias might be reduced by helping the participants to understand the consequences of the inference at the beginning.

REFERENCES

[1] Agresti, A. (2010). *Analysis of Ordinal Categorical Data*, John Wiley and Sons, New Jersey.

[2] Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, **19**(6), 716–723.

[3] Aktaş Altunay, S.; Yilmaz, A.E.; Bahcecitapar, M. and Bakacak Karabenli, L. (2021). *SPSS ve R Uygulamalı Kategorik Veri Çözümlemesi*, Seçkin Yayıncılık, Ankara.

[4] Allen, I.E. and Seaman, C.A. (2007). Likert scales and data analyses, *Quality Progress*, **40**(7), 64–65.

[5] Altun, G. and Aktas, S. (2016). Asymmetry and skew-symmetry models for square contingency tables, *Turkiye Klinikleri Journal of Biostatistics*, **8**(2), 152–161.

[6] Bagheban, A.A. and Zayeri, F. (2010). A generalization of the uniform association model for assessing rater agreement in ordinal scales, *Journal of Applied Statistics*, **37**(8), 1265–1273.

[7] Barry, D. (2017). Do not use averages with likert scale data.
`https://bookdown.org/Rmadillo/likert/`

[8] Goodman, L.A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories, *Journal of the American Statistical Association*, **74**(367), 537–552.

[9] Grimm, S.D. and Church, A.T. (1999). Cross-cultural study of response biases in personality measures, *Journal of Research in Personality*, **33**(4), 415–441.

[10] Guerra, A.L.; Gidel, T. and Vezzetti, E. (2016). *Toward a common procedure using likert and likert-type scales in small groups comparative design observations.* In "Proceedings of International Design Conference", **33**(4), 23–32.

[11] Harzing, A.W. (2006). Response styles in cross-national survey research: a 26-country study, *International Journal of Cross Cultural Management*, **6**(2), 243–266.

[12] IBM Corp. Released (2015). *IBM SPSS Statistics for Windows*, Version 23.0, IBM Corp, New York.

[13] Knauper, B. (1999). The impact of age and education on response order effects in attitude measurement, *Public Opinion Quarterly*, **63**(3), 347–370.

[14] Likert, L. (1932). A technique for the measurement of attitudes, *Archives of Psychology*, **22**(140), 1–55.

[15] Oppenheim, A.N. (1992). *Questionnaire Design, Interviewing and Attitude Measurement*, Pinter Pub. Ltd., London.

[16] Sable, V. (2018). The study of hostel life (version 2).
`https://www.kaggle.com/vinayaksable/the-study-of-hostel-life`

[17] Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics*, **6**(2), 461–464.

[18] Tomizawa, S. (1991). A model of uniform association plus two-diagonals parameter and its application to occupational mobility table data, *Statistical Paper*, **32**(1), 243–252.

[19] Truebner, M. (2019). Dynamics of "Neither Agree nor Disagree" answers in attitudinal questions, *Journal of Survey Statistics and Methodology*, **9**(1), 51–72.

[20] Tutz, G. (2020). Hierarchical models for the analysis of likert scales in regression and item response analysis, *International Statistical Review*, **89**(1), 18–35.

[21] Valet, F.; Guinot, C. and Mary, J.Y. (2007). Log-linear non-uniform association models for agreement between two ratings on an ordinal scale, *Statistics in Medicine*, **26**(3), 647–662.

[22] Yilmaz, A.E. and Saracbasi, T. (2015). Symmetric disagreement + exponential score association model in square tables, *Turkiye Klinikleri Journal of Biostatistics*, **7**(2), 96–102.