

Statistical Inference for a Novel Health Inequality Index

Cuizhen Niu¹, Shaoxin Hong², Xuejun Jiang^{3*}

¹School of Statistics, Beijing Normal University, Beijing, China

²Department of Mathematics and Statistics, University of North Carolina, Charlotte, USA

³Department of Mathematics, Southern University of Science and Technology of China, Shenzhen, China

Email: *jiangxj@sustc.edu.cn

How to cite this paper: Niu, C.Z., Hong, S.X. and Jiang, X.J. (2017) Statistical Inference for a Novel Health Inequality Index. *Theoretical Economics Letters*, 7, 251-262. <https://doi.org/10.4236/tel.2017.72021>

Received: January 16, 2017

Accepted: February 25, 2017

Published: February 28, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In this paper, we develop statistical inference for an important health inequality index proposed by Lv, Wang and Xu [1] for ordinal data. Asymptotic distributions of the indices are established. This allows us to make inference for the indices. Generalizations of the indices to multiple population setting are also studied. We demonstrate the effectiveness of our procedure using the health inequality data of several areas in Switzerland, and our results classify these areas into three classes based on their health inequalities.

Keywords

Health Inequality Index, Multiple Comparison, Self-Reported Health Status, Statistical Inference

1. Introduction

The qualitative nature of SRHS data prevents the straightforward use of conventionally developed indices for measuring income inequality. A reasonable index for SRHS data should be invariant to rescalings of variables which preserve the order of categories.

Assessment on health inequality for ordered data has received attention in the last ten years, [2] and [3] developed median-based concept of inequality. [4] proposed polarization measures, which are also median based. These methods are invariant to cardinal scaling on the categories. [5] proposed a method using income-health matrix to measure socioeconomic inequality in health. [6] introduced a family of sub-group decomposable indices and investigated the decomposability of the indices. [7] conducted an empirical study of the health inequality index for ordinal data from China. Reference [8] considered the tools and

choices to be made when measuring socioeconomic inequalities with rank-dependent inequality indices. [9] made an empirical comparison with several ordinal and cardinal measures of health inequality. [10] proposed a new measure for ordinal health data to monitor income-related health differences between regions in Great Britain. [11] defined a new ratio-scale health status variable and developed positional stochastic dominance conditions that could be implemented in a context of multi-dimensionality categorical variables. [12] examined the measurement of social polarization with categorical and ordinal data. [13] introduced two approaches to measure social polarization in the case where the distance between groups is based on an ordinal variable, such as self-assessed health status. More examples on ordinal inequality measurements can be seen in [14], [15] and so on. For statistical inference of these recent developed health inequality indices, some authors (e.g. [4], [16]) have derived standard errors for the inequality indices they have introduced. [17] presented a unified methodology for the estimation of inequality indices of the cumulative distribution function.

Recently, [1] proposed a class of measures of health inequality, which are easy to compute and have some desirable properties, such as additivity, invariance of parallel shifts, normalization and simple aversion to median-preserving spreads. However, it is designed only for one population and has not developed statistical inference for the index. This motivates us to work along this topic. In this paper, we establish asymptotic distributions of the indices introduced by [1] and extend the indices to multiple population settings. Our procedures allow dependence between the considered populations and different sample sizes. In particular, we answer several important questions, for example, whether the health inequality of one population is the same as others and is there a linear relationship among the health inequalities of different populations?

The remainder of the paper is organized as follows. In Section 2, we review the indices developed by [1] and derive asymptotic distribution of the indices. In Section 3, we develop the indices for multiple populations. Empirical results are reported in Section 4. Section 5 concludes the paper.

2. Inference for the Health Inequality Index

2.1. Review of the Indices

According to [1], denote $\mathbf{V} = (V_1, \dots, V_i, \dots, V_n)^T$ as the health statuses of n individuals. Let $\mathbf{h} = (h_1, \dots, h_m)^T$ be a finite given set of health categories with $m \geq 2$. Assume that health categories represent various health statuses and satisfy $h_1 < \dots < h_m$. The values of h_1, \dots, h_m are ordinally significant so that, if $h_j < h_k$, then h_j represents a lower health status than h_k .

Let $\mathbf{d} = (d_1, \dots, d_m)^T$ and $\mathbf{f} = (f_1, \dots, f_m)^T$ be the empirical and population frequencies of the health categories, respectively, while $d_i = n^{-1} \sum_{l=1}^n I(V_l = h_i)$ represents the relative frequency of individuals with health statuses equal to h_i . Further let

$$\Omega_0(\mathbf{f}) = \text{diag}(\mathbf{f}) - \mathbf{f}\mathbf{f}^T, \quad (1)$$

and let G be a $m \times m$ matrix with the (i, j) th entry being $2g(|i - j|)$, where $g(\cdot)$ is a function of nonnegative integers, such that

$0 = g(0) < g(1) < g(2) < \dots < g(m-1)$. Lv, Wang and Xu [1] proposed the following classes of health inequalities:

$$I(\mathbf{f}) = \sum_{i=1}^m \sum_{j \neq i} g(|i - j|) f_i f_j. \quad (2)$$

Two typical choices for $g(\cdot)$ include the following:

$$g(i) = \frac{2i}{m-1}, \quad i = 1, \dots, m-1; \quad g(i) = 2\alpha^{m-1-i}, \quad i = 1, \dots, m-1, \quad 0 < \alpha < 1. \quad (3)$$

Intuitively, the index is estimated by $I(\mathbf{d})$, an empirical plug-in estimator in statistics.

2.2. Asymptotic Results

Base on the above indices, we establish the following asymptotic distribution.

Theorem 1. *Using the delta method, we can establish that*

$$\sqrt{n}(I(\mathbf{d}) - I(\mathbf{f})) \Rightarrow N(0, \sigma^2),$$

where $\sigma^2 = \mathbf{f}^T G^T \Omega_0(\mathbf{f}) G \mathbf{f}$.

In practice, σ^2 is unknown and must be estimated. Given that \mathbf{d} is a consistent estimator of \mathbf{f} , the asymptotic variance can be estimated by $\hat{\sigma}^2 = \mathbf{d}^T G^T \Omega_0(\mathbf{d}) G \mathbf{d}$. Based on the asymptotic result, the two-sided symmetric $100(1 - \alpha)\%$ asymptotic confidence interval for the health inequality index $I(\mathbf{f})$ can be constructed as

$$\left(I(\mathbf{d}) - z_{1-\alpha/2} \sqrt{\hat{\sigma}^2/n}, I(\mathbf{d}) + z_{1-\alpha/2} \sqrt{\hat{\sigma}^2/n} \right),$$

where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution.

3. Extension to Multiple Populations

3.1. Testing for Equivalence

We first consider two populations with $\mathbf{f}_i = (f_{i1}, \dots, f_{im})^T$, $\mathbf{d}_i = (d_{i1}, \dots, d_{im})^T$ and $\mathbf{V}_i = (V_{i1}, \dots, V_{im_i})^T$, $i = 1, 2$. Our analysis considers the cases of mutually dependent samples and independent samples, with the former being relevant in examining the evolution of health inequalities in a single group (e.g., changes in health inequality over time), while the latter being relevant in comparing health inequality between two groups (e.g., cross-national). The sampling is performed independently within each group.

Lemma 1. *Using the delta method, we have*

$$\begin{aligned} \sqrt{n}(I(\mathbf{d}) - I(\mathbf{f})) &= \sqrt{n} \mathbf{f}^T G^T (\mathbf{d} - \mathbf{f}) + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{l=1}^n \mathbf{f}^T G^T (I(\mathbf{V}_l = \mathbf{h}) - \mathbf{f}) + o_p(1) \end{aligned}$$

Theorem 2. *Let σ_{ij} be the (i, j) th entry of two populations' covariance*

matrix. Denote $\tilde{n} = \max\{n_1, n_2\}$, $\lambda_i = \lim_{\tilde{n} \rightarrow \infty} (n_i/\tilde{n})$, $0 < \lambda_i \leq 1$, $i = 1, 2$. The asymptotic distribution of $I(\mathbf{d}_1) - I(\mathbf{d}_2)$ is

$$\sqrt{\tilde{n}} \left[(I(\mathbf{d}_1) - I(\mathbf{d}_2)) - (I(\mathbf{f}_1) - I(\mathbf{f}_2)) \right] \Rightarrow N \left(0, \sigma_{11}^2/\lambda_1 + \sigma_{22}^2/\lambda_2 - 2\sigma_{12}^2/\sqrt{\lambda_1\lambda_2} \right).$$

Now we consider hypothesis testing problem,

$$H_{0,1} : I(\mathbf{f}_1) = I(\mathbf{f}_2) \text{ v.s. } H_{a,1} : I(\mathbf{f}_1) \neq I(\mathbf{f}_2).$$

We introduce the following Wald statistic:

$$T_n = \frac{\tilde{n} (I(\mathbf{d}_1) - I(\mathbf{d}_2))^2}{\sigma_{11}^2/\lambda_1 + \sigma_{22}^2/\lambda_2 - 2\sigma_{12}^2/\sqrt{\lambda_1\lambda_2}}. \tag{4}$$

Then under the null hypothesis H_0 , $T_n \Rightarrow \chi_1^2$ as $n \rightarrow \infty$. The corresponding p -value can be computed by the following formula:

$$p = Pr(T \geq t_{obs}) = 1 - F_{\chi_1^2}(t_{obs}), \tag{5}$$

where $F_{\chi_1^2}(\cdot)$ represents the cumulative distribution function of the chi-squared variable with one degree of freedom.

These results are general, an assumption of independent populations is not required, this implies that our test work with the unbalanced designs case. If these two populations are treated as independent, then

$\text{Cov}\{I(V_{1l} = \mathbf{h}), I^r(V_{2l} = \mathbf{h})\} = 0$ and thus $\sigma_{12}^2 = 0$. For a particular circumstance, when the sample sizes of these two populations are equal, $n_1 = n_2 = n$, we can have $\lambda_1 = \lambda_2 = 1$ and then the asymptotic distribution in Theorem 2 reduces to

$$\sqrt{n} \left[(I(\mathbf{d}_1) - I(\mathbf{d}_2)) - (I(\mathbf{f}_1) - I(\mathbf{f}_2)) \right] \Rightarrow N \left(0, \sigma_{11}^2 + \sigma_{22}^2 - 2\sigma_{12}^2 \right). \tag{6}$$

We propose statistical inference procedures to test the equality between samples in terms of their health inequality indices. This equality issue often emerges when checking for the similarity of the health inequalities in the whole country or in a specified region. For example, China, a country consists of many administrative regions, such as Eastern China, North China, and Central Region, with each region having several provinces. Those provinces in the same region have similar economic and/or social behaviors. Therefore, those provinces in the same region are assumed to have the same health inequalities. We also examine whether the health inequality index of a province is the same as the average index of the entire region. The above two testing problems lead to another application. If the preceding analysis reveals that the provinces within each region have equal indices, then we can check whether the common means in two regions are also the same. Accordingly, we cluster the regions based on the test results. In other words, if several regions have the same health inequality, then we can view these regions as one cluster.

3.2. Global Test

Suppose there are $r (r \geq 3)$ populations with $\mathbf{f}_i = (f_{i1}, \dots, f_{im})^T$,

$\mathbf{d}_i = (d_{i1}, \dots, d_{im})^T$ and $\mathbf{V}_i = (V_{i1}, \dots, V_{im_i})^T$, $i = 1, \dots, r$. For the dependent samples, we can obtain the similar results as those presented in Section 2. However, the covariance structure becomes too complex to be practical when more samples are used. We only consider independent samples for simplification. A global test can be constructed as:

$$H_{0,2} : I(\mathbf{f}_1) = \dots = I(\mathbf{f}_r) \text{ v.s. } H_{a,2} : I(\mathbf{f}_i) \neq I(\mathbf{f}_j) \tag{7}$$

for some $i \neq j (i, j = 1, 2, \dots, r)$.

Define the matrix $R = \begin{bmatrix} I_{(r-1) \times (r-1)}, -L_{(r-1) \times 1} \end{bmatrix}$, where $I_{(r-1) \times (r-1)}$ is an identity matrix with $r-1$ dimension, and $L_{(r-1) \times 1}$ is a $r-1$ dimensional vector with all the elements being 1. Then, Hypothesis in (7) can be rewritten as follows:

$$H_{0,3} : RI(\mathbf{f}) = 0, \text{ v.s. } H_{a,3} : RI(\mathbf{f}) \neq 0,$$

where $I(\mathbf{f}) = (I(\mathbf{f}_1), \dots, I(\mathbf{f}_r))^T$.

Define $\tilde{n} = \max\{n_i, i = 1, \dots, r\}$, $\lambda_i = \lim_{\tilde{n} \rightarrow \infty} (n_i / \tilde{n})$, $0 < \lambda_i \leq 1$, $i = 1, \dots, r$. Given the independence of the r groups of samples, we can obtain

$$\sqrt{\tilde{n}}(I(\mathbf{d}) - I(\mathbf{f})) \Rightarrow N(0, \Sigma), \tag{8}$$

where $\Sigma = \text{diag}(\sigma_{11}^2 / \lambda_1, \dots, \sigma_{rr}^2 / \lambda_r)$. Therefore,

$$\sqrt{\tilde{n}}(RI(\mathbf{d}) - RI(\mathbf{f})) \Rightarrow N(0, R\Sigma R^T)$$

and

$$\tilde{n}(RI(\mathbf{d}) - RI(\mathbf{f}))^T (R\Sigma R^T)^{-1} (RI(\mathbf{d}) - RI(\mathbf{f})) \Rightarrow \chi_{r-1}^2. \tag{9}$$

Note that under the null hypothesis, $RI(\mathbf{f}) = 0$ in (9). Consequently, a Wald type of test statistic can be defined as

$$T_r = \tilde{n}I^T(\mathbf{d})R^T(R\hat{\Sigma}R^T)^{-1}RI(\mathbf{d}), \tag{10}$$

where $\hat{\Sigma} = \text{diag}(\tilde{n}\hat{\sigma}_{11}^2/n_1, \dots, \tilde{n}\hat{\sigma}_{rr}^2/n_r)$ is an estimator of Σ . Given the central role of the test statistic T_r , we state the asymptotic behavior of T_r under the null hypothesis in the following theorem.

Theorem 3. Let $\tilde{n} = \max\{n_i, i = 1, \dots, r\}$, $\lambda_i = \lim_{\tilde{n} \rightarrow \infty} (n_i / \tilde{n})$, $0 < \lambda_i \leq 1$, $i = 1, \dots, r$, then under the null hypothesis $H_{0,1} : I(\mathbf{f}_1) = \dots = I(\mathbf{f}_r)$ in (7), we have $T_r \Rightarrow \chi_{r-1}^2$.

The corresponding p -value can be computed by:

$$p = Pr(T_r \geq t_{r,obs}) = 1 - F_{\chi_{r-1}^2}(t_{r,obs}), \tag{11}$$

where $F_{\chi_{r-1}^2}(\cdot)$ represents the cumulative distribution function of the chi-squared variable with $r-1$ degrees of freedom. The equality hypothesis (7) can be regarded as a generalization of the two-sample comparison case. The availability of this hypothesis can be seen clearly in our empirical application.

3.3. Hypothesis Testing within a Cluster

Another interesting problem in the multiple sample case is whether the health

inequality of a specified population is the same as the average health inequality of entire population. For instance, one may interest to investigate the health inequality level in Hebei province is higher or lower than the average level of all provinces in the North China region. Accordingly, we propose the following testing hypothesis:

$$H_{0,4} : I(\mathbf{f}_j) = I(\mathbf{f}_0) \text{ v.s. } H_{a,4} : I(\mathbf{f}_j) \neq I(\mathbf{f}_0) \tag{12}$$

for some $j (j = 1, 2, \dots, r)$. If the null hypothesis $H_{0,1}$ in (7) holds, then null hypothesis $H_{0,4}$ holds naturally. In other words, hypothesis $H_{0,4}$ only becomes meaningful when hypothesis $H_{0,1}$ is not true.

Define $\mathbf{a}_j = (-1/r, \dots, -1/r, 1-1/r, -1/r, \dots, -1/r)^T$, that is, \mathbf{a}_j is a $r \times 1$ vector with its j -th element being $1-1/r$ and other elements all being $-1/r$. Hypothesis (12) can be rewritten as follows

$$H_{0,4} : \mathbf{a}_j^T I(\mathbf{f}) = 0, \text{ v.s. } H_{a,4} : \mathbf{a}_j^T I(\mathbf{f}) \neq 0.$$

Recall that $\sqrt{\tilde{n}}(I(\mathbf{d}) - I(\mathbf{f})) \Rightarrow N(0, \Sigma)$ in (8) holds, we can obtain

$$\sqrt{\tilde{n}}(\mathbf{a}_j^T I(\mathbf{d}) - \mathbf{a}_j^T I(\mathbf{f})) \Rightarrow N(0, \mathbf{a}_j^T \Sigma \mathbf{a}_j).$$

Similar to the derivation of T_r , we can construct the following test statistic

$$T_{ra_j} = \tilde{n} I^T(\mathbf{d}) \mathbf{a}_j (\mathbf{a}_j^T \Sigma \mathbf{a}_j)^{-1} \mathbf{a}_j^T I(\mathbf{d}). \tag{13}$$

Under the null hypothesis in (12), $T_{ra_j} \Rightarrow \chi^2(1)$. Then the p -value can be determined similarly as that for T_r .

3.4. Hypothesis Testing between Clusters

Further, we discuss the hypothesis testing between clusters. Assume now that our preliminary analysis reveals that the provinces the corresponding region (cluster), such as Eastern China region, have the same health inequality indices. We may then examine whether the health inequalities between two regions are similar. To this end, we choose two representative provinces in each region and then compare their health inequality indices following the proposed approaches in Section 2. However, this method does not employ all information in these groups. To use all underlying information, we compare the common means of these two regions. We consider the following hypothesis:

$$H_{0,5} : I(\mathbf{f}_{01}) = I(\mathbf{f}_{02}) \text{ versus } H_{a,5} : I(\mathbf{f}_{01}) \neq I(\mathbf{f}_{02}) \tag{14}$$

where $I(\mathbf{f}_{01}) = \sum_{i=1}^{r_1} I(\mathbf{f}_i) / r_1$ and $I(\mathbf{f}_{02}) = \sum_{i=r_1+1}^{r_1+r_2} I(\mathbf{f}_i) / r_2$.

Without loss of generality, we assume that the first r_1 populations are clustered in one group with a common health inequality $I(\mathbf{f}_{01})$, while the $r_1 + 1$ to $r_1 + r_2$ populations are clustered in another group with another common health inequality $I(\mathbf{f}_{02})$. $H_{0,5}$ only becomes meaningful when null hypothesis $H_{0,1}$ in (7) is not true. Define

$$\mathbf{b} = (1/r_1, \dots, 1/r_1, -1/r_2, \dots, -1/r_2, 0, \dots, 0)^T \dots$$

That is, \mathbf{b} is a $r \times 1$ vector with its first r_1 elements being $1/r_1$, the $r_1 + 1$ to $r_1 + r_2$ elements being $-1/r_2$ and the other elements being 0. Similar to the derivation of T_{ra} , we can construct the test statistic as follows:

$$T_{rb} = \tilde{n}I^T(\mathbf{d})\mathbf{b}(\mathbf{b}^T\Sigma\mathbf{b})^{-1}\mathbf{b}^T I(\mathbf{d}). \tag{15}$$

Under the null hypothesis in (14), $T_{rb} \Rightarrow \chi^2(1)$, thus p -value can be determined similarly as that for T_r .

4. Empirical Application

To illustrate our proposed procedures, we present a real application by using the data of the Swiss Health Survey [SHS] in 2002, conducted by Switzerland’s Federal Statistical Office. A total of 19,706 observations were collected from seven areas in Switzerland. The survey respondents were asked to rate their health statuses on a five-point scale ranging from very bad to very good. This dataset was also analyzed by [3] and [6]. We do not include the distributions of SHS in the seven regions in this paper, this information can be found in [6]. We use the health inequality indices proposed by [1] to analyze the survey data and yield new observations. Denote the index with $g(|i - j|) = 2|i - j|/(m - 1)$ by F1 and the index with $g(|i - j|) = 2\alpha^{m-1-|i-j|}$ by F2. For checking the robustness of the results obtained, we choose $\alpha = 0.9, 0.6$ and 0.3 , then these related indices are denoted as F2-1, F2-2 and F2-3, respectively.

Table 1 presents the health inequalities of seven areas in Switzerland based on F1 and F2 with different α . The standard errors are enclosed in parentheses, and the health inequalities are ranked based on the proposed measures. From this table, in all four different measures, we can find that Lemman is the region with the highest health inequality value, which implies that the health status exists the most significant difference between Lemman citizens. The other regions show ambiguous ranking. Specifically, for F1 and F2-2, Zurich is the region with least difference in health status, and Central has the second-to-the-lowest inequality. However, for F2-3, Central is identified as the least imbalanced region in health status, while Zurich has the second-to-the-lowest inequality. East and Ticino show the similar behavior.

Table 1. Health inequality in the seven statistical areas of Switzerland.

Area	F1	F2-1	F2-2	F2-3
Lemman	0.3934(0.0073)1	0.8985(0.0102)1	0.3226(0.0058)1	0.0750(0.0034)1
North-West	0.3589(0.0090)3	0.8277(0.0142)3	0.2937(0.0068)3	0.0651(0.0033)3
Central	0.3151(0.0084)6	0.7930(0.0159)4	0.2601(0.0065)6	0.0439(0.0018)7
Middle-Land	0.3665(0.0064)2	0.8572(0.0100)2	0.3013(0.0051)2	0.0662(0.0027)2
East	0.3211(0.0070)4	0.7899(0.0131)5	0.2637(0.0054)4	0.0467(0.0016)5
Ticino	0.3205(0.0170)5	0.7245(0.0303)7	0.2605(0.0134)5	0.0583(0.0056)4
Zurich	0.3138(0.0066)7	0.7735(0.0125)6	0.2579(0.0051)7	0.0456(0.0015)6

Due to the reason of random sampling of the data set, it is natural to ask questions, like, do East and Ticino have different health inequalities in fact? Do Central and Zurich have the same health inequality actually? We use statistical inferences to address these problems. To fully answer these questions, various interesting two-sample comparison tests are carried out, the results are reported in **Table 2**. We set the significance level to 5%. From **Table 2**, we can conclude that Leman is significantly more imbalanced than Middle-Land in health status. In contrast to the findings in **Table 1**, Middle-Land and North-West do not show statistically significant differences in their health inequalities. In other words, these two regions have the same health inequality level base on the data set we have. North-West is significantly more unbalanced in health status than East. Except for F2-3, all p-values for North-West and Ticino are all smaller than 5%. Therefore, the difference of health inequality between North-West and Ticino can be confirmed almost. Central and Zurich have the same inequality level, and the same finding has been observed for East and Ticino.

Based on the above analysis, we classify North-West and Middle-Land, East and Ticino, and Central and Zurich into three groups. However, can we combine two groups, such as the East and Ticino group with the Central and Zurich group? The question is equivalent to ask whether the average health inequality of the East and Ticino group is the same as that of the other group. The p-values of tests by using the above four measures are 0.5505, 0.1778, 0.7105 and 0.0140, respectively, which are all larger than 5% except for F2-3. Therefore, East, Ticino, Central, and Zurich may be clustered into one group. We also check whether these four regions have the same health inequality levels. The p-values for this global equality hypothesis testing are 0.8805, 0.1824, 0.8946 and 0.0942, respectively, which suggest that these regions have the same inequality levels. We then examine whether this four-member group can be enlarged by including the North-West and Middle-Land group? We propose two hypotheses to investigate this question. First, are the average inequalities of North-West and Middle-Land similar to those of the other groups? Second, do these six regions have the same health inequality levels? For these two hypotheses, all the p-values resulting from tests with the four measures are significantly smaller than 5%, which indicate that the average health inequality of the North-West and Middle-Land group is different from that of the four-member group. We then examine whether the

Table 2. *p*-values for two-sample comparison problems.

Area	F1	F2-1	F2-2	F2-3
Leman vs Middle-Land	0.0054	0.004	0.0056	0.0378
North-West vs Middle-land	0.4765	0.0902	0.3723	0.7969
North-West vs East	0.0006	0.0494	0.0005	4.04E-07
North-West vs Ticino	0.0436	0.002	0.0273	0.3002
Central vs Zurich	0.9032	0.3349	0.7904	0.4494
East vs Ticino	0.9739	0.0476	0.8275	0.0473

health inequality level of Leman is the same as the average level of the North-West and Middle-Land group. The p -values of all four measures are strongly smaller than 5%, which indicate that the health inequality level of Leman is different from the average level of the North-West and Middle-Land group. In sum, we classify these seven regions into three groups, that is, Leman, North-West and Middle-Land, and the other four regions.

5. Conclusion

In this paper, we propose several statistical inference procedures for the novel health inequality indices introduced in [1]. We consider one-, two-, and multiple-sample cases. Given that health surveys generally cover multiple regions, the health inequalities of multiple sample cases must be tested. The health inequality in various regions of Switzerland validates the availability of our proposed tools. Seven regions covered by SHS can be categorized into three groups after the numerical study; Leman has the highest health inequality followed by the North-West and Middle-Land group. The other four regions (*i.e.*, Central, East, Ticino, and Zurich) have the same health inequality. Our proposed procedures can also be applied to other recently proposed health inequality indices. The subjective well-being is influenced by many factors such as health inequality, education, environment and so on. The statistical inference on multi-dimensionality well-being inequality can be investigated ongoing.

Acknowledgements

This research was funded by the Fundamental Research Funds for the Central Universities, China Postdoctoral Science Foundation (2016M600951), National Natural Science Foundation of China (11101432) and Natural Science Foundation of Guangdong Province, China (2016A030313856). The authors thank the editor, the associate editor and the anonymous referee for their constructive comments and suggestions which led to a substantial improvement of an early manuscript. All correspondence should be addressed to Xuejun, Jiang, Department of Mathematics, Southern 6 University of Science and Technology, Shenzhen, China, E-mail: jiangxj@sustc.edu.cn.

References

- [1] Lv, G. M., Wang, Y. and Xu, Y. S. (2015) On a New Class of Measures for Health Inequality Based on Ordinal Data. *The Journal of Economic Inequality*, **13**, 465-477. <https://doi.org/10.1007/s10888-014-9289-4>
- [2] Allison, R.A. and Foster, J.E. (2004) Measuring Health Inequalities Using Qualitative Data. *Journal of Health Economics*, **23**, 505-524. <https://doi.org/10.1016/j.jhealeco.2003.10.006>
- [3] Abul Naga, R.H. and Yalcin, T. (2008) Inequality Measurement for Ordered Response Health Data. *Journal of Health Economics*, **27**, 1614-1625. <https://doi.org/10.1016/j.jhealeco.2008.07.015>
- [4] Apouey, B. (2007) Measuring Health Polarization with Self-Assessed Health Data. *Health Economics*, **16**, 875-894. <https://doi.org/10.1002/hec.1284>

- [5] Zheng, B. (2011) A New Approach to Measure Socioeconomic Inequality in Health. *The Journal of Economic Inequality*, **9**, 555-577. <https://doi.org/10.1007/s10888-010-9148-x>
- [6] Kobus, M. and Miloš, P. (2012) Inequality Decomposition by Population Subgroups for Ordinal Data. *Journal of Health Economics*, **31**, 15-21. <https://doi.org/10.1016/j.jhealeco.2011.11.005>
- [7] Wang, H.L. and Yu, Y.W. (2016) Increasing Health Inequality in China: An Empirical Study with Ordinal Data. *The Journal of Economic Inequality*, **14**, 41-61. <https://doi.org/10.1007/s10888-015-9315-1>
- [8] Erreygers, G. and Van Ourti, T. (2011) Measuring Socioeconomic Inequality in Health, Health Care and Health Financing by Means of Rank-Dependent Indices: A Recipe for Good Practice. *Journal of Health Economics*, **30**, 685-694. <https://doi.org/10.1016/j.jhealeco.2011.04.004>
- [9] Madden, D. (2010). Ordinal and cardinal measures of health inequality: an empirical comparison. *Health Economics*, **19**, 243-250. <https://doi.org/10.1002/hec.1472>
- [10] Allanson, P. (2017) Monitoring Income-Related Health Differences between Regions in Great Britain: A New Measure for Ordinal Health Data. *Social Science and Medicine*, **175**, 72-80. <https://doi.org/10.1016/j.socscimed.2016.12.033>
- [11] Makdissi, P. and Yazbeck, M. (2014) MEASURING Socioeconomic Health Inequalities in Presence of Multiple Categorical information. *Journal of Health Economics*, **34**, 84-95. <https://doi.org/10.1016/j.jhealeco.2013.11.008>
- [12] Permanyer, I. and D'ambrosio, C. (2015) Measuring Social Polarization with Ordinal and Categorical Data. *Journal of Public Economic Theory*, **17**, 311-327. <https://doi.org/10.1111/jpet.12093>
- [13] Fusco, A. and Silber, J. (2014) On Social Polarization and Ordinal Variables: The Case of Self-Assessed Health. *The European Journal of Health Economics*, **15**, 841-851. <https://doi.org/10.1007/s10198-013-0529-5>
- [14] Erreygers, G. (2009) Correcting the Concentration Index. *Journal of Health Economics*, **28**, 504-515. <https://doi.org/10.1016/j.jhealeco.2008.02.003>
- [15] Lazar, A. and Silber, J. (2013) On the Cardinal Measurement of Health Inequality When Only Ordinal Information Is Available on Individual Health Status. *Health Economics*, **22**, 106-113. <https://doi.org/10.1002/hec.1821>
- [16] Cowell, F. and Flachaire, E. (2014) Inequality with Ordinal data. Unpublished.
- [17] Abul Naga, R.H. and Stapenhurst, C. (2015) Estimation of Inequality Indices of the Cumulative Distribution Function. *Economics Letters*, **130**, 109-112. <https://doi.org/10.1016/j.econlet.2015.03.004>

Appendix. Proofs of Theorems

Proof of Theorem 1.

Given that d is a consistent estimator of f , for the empirical frequency of the health categories, we can easily obtain the following:

$$\sqrt{n}(d - f) \Rightarrow N(0, \Omega_0(f)), \tag{A.1}$$

where

$$\Omega_0(f) = \text{diag}(f) - ff^T.$$

Note that $g(|i - j|)f_i f_j = g(|j - i|)f_j f_i$. Therefore,

$$I(f) = 2 \sum_{i=1}^{m-1} \sum_{j=i+1}^m g(|i - j|)f_i f_j.$$

Define $J = \left(\frac{\partial I(f)}{\partial f_1}, \dots, \frac{\partial I(f)}{\partial f_m} \right)^T$ It can be easily shown that

$$\frac{\partial I(f)}{\partial f_i} = 2 \sum_{l \neq i} g(|l - i|)f_l. \tag{A.2}$$

Alternatively, we can have $J = Gf$. Keeping only the first two terms of the Taylor expansion, we can estimate $I(f)$ as

$$I(d) \approx I(f) + J^T \cdot (d - f).$$

Then the variance of $I(d) - I(f)$ is approximated by

$$\begin{aligned} \text{Var}(J^T \cdot (d - f)) &= J^T \cdot \text{Cov}(d - f, d - f) \cdot J \\ &= f^T G^T (\Omega_0/n) Gf. \end{aligned}$$

Also $E(I(d)) = E(I(f))$ since d is a consistent estimator of f , it follows that

$$\sqrt{n}(I(d) - I(f)) \Rightarrow N(0, \sigma^2),$$

where $\sigma^2 = f^T G^T \Omega_0(f) Gf$. □

Proof of Theorem 2.

Define $\tilde{n} = \max\{n_1, n_2\}$, $\lambda_i = \lim_{\tilde{n} \rightarrow \infty} (n_i/\tilde{n})$, $0 < \lambda_i \leq 1$, $i = 1, 2$. From Lemma 1, we have

$$\begin{aligned} \sqrt{\tilde{n}}(I(d) - I(f)) &= \sqrt{\tilde{n}}f^T G^T (d - f) + o_p(1) \\ &= \frac{1}{\sqrt{\tilde{n}}} \sum_{l=1}^n f^T G^T (I(V_l = h) - f) + o_p(1). \end{aligned}$$

Let $\sigma_h^2 = \text{Cov}(I(V_{1l} = h)I^T(V_{2l} = h))$. After that, similarly as the proof of Theorem 1, it can be derived that

$$\begin{aligned} \sqrt{\tilde{n}} \begin{pmatrix} I(d_1) - I(f_1) \\ I(d_2) - I(f_2) \end{pmatrix} &= \begin{pmatrix} \sqrt{\frac{\tilde{n}}{n_1}} \times \sqrt{n_1} (I(d_1) - I(f_1)) \\ \sqrt{\frac{\tilde{n}}{n_2}} \times \sqrt{n_2} (I(d_2) - I(f_2)) \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{\tilde{n}}{n_1}} \times \frac{1}{\sqrt{n_1}} \sum_{l=1}^{n_1} f_1^T G^T (I(V_{1l} = h) - f_1) \\ \sqrt{\frac{\tilde{n}}{n_2}} \times \frac{1}{\sqrt{n_2}} \sum_{k=1}^{n_2} f_2^T G^T (I(V_{2k} = h) - f_2) \end{pmatrix} + o_p(1) \\ &\Rightarrow N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11}/\lambda_1 & \sigma_{12}/\sqrt{\lambda_1 \lambda_2} \\ \sigma_{21}/\sqrt{\lambda_1 \lambda_2} & \sigma_{22}/\lambda_2 \end{pmatrix} \right). \end{aligned} \tag{A.3}$$

Here,

$$\begin{aligned}\sigma_{11}^2 &= \mathbf{f}_1^T \mathbf{G}^T \Omega_0(\mathbf{f}_1) \mathbf{G} \mathbf{f}_1, \\ \sigma_{12}^2 = \sigma_{21}^2 &= \mathbf{f}_1^T \mathbf{G}^T \text{Cov}(I(V_{1l} = \mathbf{h}) I^T(V_{2l} = \mathbf{h})) \mathbf{G} \mathbf{f}_2 = \mathbf{f}_1^T \mathbf{G}^T \sigma_h^2 \mathbf{G} \mathbf{f}_2, \\ \sigma_{22}^2 &= \mathbf{f}_2^T \mathbf{G}_0^T(\mathbf{f}_2) \mathbf{G} \mathbf{f}_2\end{aligned}$$

Plug in the consistent estimators of \mathbf{f}_i and σ_h^2 by \mathbf{d}_i and

$$\hat{\sigma}_h^2 = \left\{ \min(n_1, n_2) \right\}^{-1} \sum_{l=1}^{\min\{n_1, n_2\}} I(V_{1l} = \mathbf{h}) I^T(V_{2l} = \mathbf{h}) - \mathbf{d}_1 \mathbf{d}_2^T,$$

respectively, thus we can easily estimate $\sigma_{i,j}^2, i, j = 1, 2$ consistently.

Combining the above result in (A.3), we can obtain

$$\begin{aligned}& \sqrt{\tilde{n}} \left[(I(\mathbf{d}_1) - I(\mathbf{d}_2)) - (I(\mathbf{f}_1) - I(\mathbf{f}_2)) \right] \\ &= \sqrt{\tilde{n}} \left[(I(\mathbf{d}_1) - I(\mathbf{f}_1)) - (I(\mathbf{d}_2) - I(\mathbf{f}_2)) \right] \\ &\Rightarrow N\left(0, \sigma_{11}^2/\lambda_1 + \sigma_{22}^2/\lambda_2 - 2\sigma_{12}^2/\sqrt{\lambda_1 \lambda_2}\right).\end{aligned}$$

□



Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact tel@scirp.org