University of Toronto
Department of Economics

Working Paper 804

Mean Location and Gini-like Inequality measures for Multivariate Ordinal Variates: Examining the Progress of Health Outcomes in Pre Covid United Kingdom

By Gordon Anderson

September 25, 2025

Mean Location and Gini-like Inequality measures for Multivariate Ordinal Variates: Examining the Progress of Health Outcomes in Pre Covid United Kingdom.

Gordon Anderson

University of Toronto Economics Department

October 2025

Summary

The ever-expanding use of ordinal data is usually facilitated by artificial attribution of cardinal scale to ordered categories. Such practices have been shown to lead to ambiguous and equivocal results. Here a probabilistic distance construct is employed to develop unambiguous level and inequality measures for ordinal situations analogous to the Mean and Gini coefficient used in cardinally measurable paradigms. The commonality of the probabilistic distance measure across dimensions means that the measures are readily extended to multidimensional situations. The measures are exemplified in an analysis of the progress of health outcomes in pre-covid 21st century United Kingdom.

Keywords: Ordered Categorical Data, Level and Inequality measurement, Health Outcomes.

JEL codes: I14 H11 C13 C43

Introduction.

The increasing use of opinion surveys[1], clinical trials with categorical outcomes (Koch et.al. 1998), political polling, and happiness and satisfaction surveys (Clark et.al. 2016, Lauderdale et.al. 2020), has heralded widespread growth in the use of Ordered Categorical Data (OCD). Difficulties arise with analyzing OCD because it is bereft of cardinal measure, a problem often circumvented by artificially attributing cardinal scale to the ordered categories (e.g. Cantril (1965), Likert (1932) or Rankin (1957) scales). However, the practice has recently met with criticism within the economics literature (Bond and Lang 2019, Schröder and Yitzhaki 2017) since such artificial attribution engenders ambiguity and equivocation in interpreting location and diversity measures (alternative equally valid scales can be shown to yield very different and sometimes contradictory conclusions). All of which highlights the need for unambiguous comparison instruments in OCD environments.

Here, using a probabilistic distance construct (Mendelson 1987), μ_{OC} and G_{OC}, location and diversity measures analogous to the Mean and Gini coefficient employed in cardinally measurable situations, are developed and exemplified in unambiguously measuring levels of, and the diversity in, ordered categorical outcome distributions. Intuitively the probabilistic distance between two outcome levels is the chance that an outcome occurs between the boundaries they define, the greater that chance, the greater the distance between them. The specificity of the application - an analysis of self-reported health outcomes in the United Kingdom - should not obscure the fact that these techniques can be used in a wide range of ordered categorical situations. Indeed, the commonality of the probabilistic metric that is employed across dimensions in multivariate OCD situations facilitates the extension of these techniques to multivariate environments in ways that have proved difficult in continuous multidimensional environments where different dimensions operate under different metrics.

[1] The number of active polling and survey agencies in North America more than doubled in the first 20 years of the 21st Century (Kennedy et.al. 2023).

In what follows, Section 1 develops Mean and Gini-like dispersion statics for univariate ordinal data and outlines additional analysis techniques used in that data environment. Section 2 exemplifies their use in analyzing the progress of health outcomes in pre-covid 21st Century United Kingdom. Conclusions are drawn in section 3 and bivariate extensions and appropriate inference procedures are outlined in the appendix.

1. Development of mean and Gini-like coefficient measures.

1.1 The Ordered Categorical Mean.

The Ordered Categorical Mean μ_{OC} is developed as an analogue of the classical mean employed in a cardinally measurable world. In that continuous world $f(y)$, the probability density function of the random variable Y, is defined on the real line such that $f(y) \geq 0$ with $\int_{-\infty}^{\infty} f(y)dy = 1$. The Cumulative Distribution Function $F(y) = \int_{-\infty}^{y} f(z)dz$, where $F(y) = P(Y \leq y)$ is the probabilistic distance of y from its lower bound, has a corresponding Survival Function $S(y) = 1 - F(y)$ where $S(y) = P(Y > y)$ is the probabilistic distance of y from its upper bound. The mean μ, where $\mu = \int_{-\infty}^{\infty} yf(y)dy$, also has a probabilistic interpretation. Integrating $\int_{-\infty}^{\infty} yf(y)dy$ by parts will show it to be equal to $\int_{-\infty}^{\infty} S(y)dy$ which implies that μ may be alternatively described as the cumulation of chances of higher outcomes than y over its range. In essence it is an index of the extent to which higher outcomes are probable throughout the range of the distribution.

$S(y)$, in its role in the First Order Stochastic Dominance condition formulation[2], features in the development of μ_{OC}'s properties in the ordinal world. Note that, for alternative distributions $f_a(y)$ and $f_b(y)$, $\mu_a - \mu_b = \int_{-\infty}^{\infty}\big(S_a(y) - S_b(y)\big)dy > 0$ does not imply that the chance of a higher than y outcome under $f_a(y)$ is at least as great as under $f_b(y)$ at every y. To establish that an unambiguity factor $UNAM = \int_{-\infty}^{\infty}\big(S_a(y) - S_b(y)\big)dy/$

---

[2] For any two distributions of y, $f_a(y)$ and $f_b(y)$, $S_a(y) \geq S_b(y)\ \forall\ y$ with strict inequality for some y guarantees the expected value of any monotonic increasing function of y to be at least as great under $f_a(y)$ as under $f_b(y)$. In essence the condition requires the chance of a higher outcome than y to always be at least as great under $f_a(y)$ than under $f_b(y)$ and strictly greater for some y.

$\int_{-\infty}^{\infty}|S_a(y) - S_b(y)|dy$, must equal 1, only then can it be said that chances of higher outcomes under $f_a(y)$ are unequivocably at least as good as under $f_b(y)$.

In the ordinal world an Ordered Categorical Mean can be analogously construed as the cumulation (i.e. sum) of the survival function over the ordered categories where the Survival Function at a particular category is the chance of being in any higher category. To fix ideas, let $\underline{p}$ be a K long vector of probabilities p_i of realizing the i'th of K ordered categories indexed $i = 1,..,K$ where $0 \leq p_i \leq 1 \; \forall \; i \; with \; \sum_{i=1}^{K} p_i = 1$. Let $\underline{P}$ be the K long vector of cumulative densities (with typical element P_i) where $P_i = \sum_{j=1}^{i} p_j$ and Let $\underline{S}$ be the K long vector of survival function probabilities (with typical element $S_i = 1 - P_i$). Then μ_{OC} the Ordered Categorical Mean can be defined as:

$$\mu_{OC} = \sum_{i=1}^{K} S_i \qquad [1]$$

Note that, μ_{OC} is a unique cardinally measurable number between 0 (when all the mass is at the lowest category) and K-1 (when all the mass is in the highest category), dividing μ_{OC} by $K - 1$ will render it comparable across variables with differing numbers of categories. Furthermore, the commonality of the distance measure across dimensions, means that [1] is readily extended to multi-dimensioned ordered categorical environments, the bivariate case is developed in the appendix.

Given a random sample of OCD observations, inference regarding [1] is straightforward and is outlined in the appendix. Note that, for any two distributions $\underline{p}_a$ and $\underline{p}_b$ over the same group of categories with respective ordered categorical means $\mu_{OC,a}$ and $\mu_{OC,b}$, $\mu_{OC,a} - \mu_{OC,b} = \sum_{i=1}^{K}(S_{a,i} - S_{b,i}) > 0$, even though such a difference may be statistically significant, does not imply that the chance of a higher outcome under $\underline{p}_a$ is greater than that under $\underline{p}_b$ at every category. For that $UNAM = \frac{\sum_{i=1}^{K}(S_{a,i}-S_{b,i})}{\sum_{i=1}^{K}|(S_{a,i}-S_{b,i})|} = 1$ needs to be established[3], only then can it be said that chances of higher outcomes under $\underline{p}_a$ are unequivocably at least as good as under $\underline{p}_a$.

1.2 The Ordered Categorical Gini coefficient.

When the random variable resides in the positive orthant, the classic Gini Coefficient formula is given by:

$$G = \frac{1}{\mu}\int_0^\infty f(y) \int_0^\infty f(x)\,|x-y|dxdy$$

Which, in words is the mean standardized expected distance between every pair of points in the range of the random variable. For a random sample of n observations $x_i\ , i = 1, .., n$ on the continuous random variable X, the Gini Coefficient can be estimated by:

$$G = \frac{1}{\mu n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|x_i - x_j\right|$$

where $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$ which in words is the mean standardized average distance between every pair of points in the sample.

An analogous Gini coefficient for Ordered Categorical Data can be developed using the probabilistic distance concept and the Ordered Categorical Mean developed above. The probabilistic distance between two outcomes is the chance that an outcome occurs between the boundaries they define, the greater that chance, the greater the distance. In the univariate case $\underline{P}$ can be used to define $d(i,j)$, the probabilistic distance between category i and category j. Suppose $i > j$ then $P_i - P_j = \sum_{k=j+1}^{i} p_k$ which includes the chance of being in category i but excludes the chance of being in category j, alternatively $P_{i-1} - P_{j-1} = \sum_{k=j}^{i-1} p_k$ includes the chance of being in category j and excludes the chance of being in category i, therefor define $d(i,j)$ the distance between categories i and j as:

$$d(i,j) = 0.5\left(\left(P_i - P_j\right) + \left(P_{i-1} - P_{j-1}\right)\right) = \left(P_i - P_j\right) + 0.5\left(p_j - p_i\right).$$

[3] When UNAM is close to 1 $\underline{p}_a$ can be said to be “Almost” Dominant (Leshno and Levy 2002).

Then D, the total probabilistic distance in the distribution from lower outcomes to higher outcomes is given by:

$$D = \sum_{i=2}^{K} \sum_{j=1}^{i-1} d(i,j)$$

with an average probabilistic distance given by $2D/ND$ where ND is the number of distances computed which results G_{OC} an ordered categorical Gini coefficient where:

$$G_{OC} = D/ND\mu_{OC} \qquad [2]$$

With μ_{OC} is the Ordered Categorical Mean. Again, the commonality of the probabilistic distance metric across dimensions facilitates development of a multidimensional Ordered Categorical Gini which is also relegated to the appendix.

1.3 Additional Tools of analysis.

In the following exemplifying analysis of the progress of Health Outcomes in the UK some additional tools of analysis that have already appeared in the literature will be required, they are briefly outlined here.

Much of the examination of progress hinges upon whether health outcomes differ substantially over time and between groups. In order to establish whether or not outcome distributions are significantly different, Anderson, Linton and Whang (2012) developed an asymptotically normal test statistic for the commonality of two continuous distributions based upon OV, the extent to which they overlap[4]. OV can be shown to be equal to $1 - TR$ which is Ginis' Transvariation measure (Gini 1912) where $TR = 0.5 \int |f_1(x) - f_2(x)| dx$ is statistic measuring the extent of differences between $f_1(x)\ and\ f_2(x)$. Given two distributions $f_1(x)$ and $f_2(x)$, $OV = \int min(f_1(x), f_2(x))dx$. When the distributions are identical $OV = 1$ $(TV = 0)$, when they are completely segmented $OV = 0$ $(TV = 1)$. The ordered categorical variable analogues for two alternative distributions $\underline{p}_A$ and $\underline{p}_B$ are given by $OV = \sum_{k=1}^{K} min\left(p_{k,A}, p_{k,B}\right)$ and $TV = 1 - OV$, each is asymptotically normal with

---

[4] OV can be shown to be equal to $1 - TR$ which is Ginis' Transvariation measure (Gini 1912) where $TR = 0.5 \int |f_1(x) - f_2(x)| dx$ is statistic measuring the extent of differences between $f_1(x)\ and\ f_2(x)$

variance $OV(1-OV)/\left(n_A n_B/(n_A+n_B)\right)$. OV is a measure of commonality, TV is a measure of dissimilarity.

It is well known that health outcomes deteriorate with age (Deaton and Paxson, 1998; Kerkhofs and Lindeboom, 1997; Miller et al., 2019) however an egalitarian policy objective would have it that all should have similar chances of good health. Equality of opportunity has much to do with people with different circumstances beyond their control (such as age and gender), having the same chances of success. in essence circumstance conditioned outcome distributions should be similar in order to secure equal chances of success.

This has been examined by comparing the mean outcomes of circumstance groups (Ferreira and Peragine 2015). Suppose J circumstance groups indexed $j = 1, . . , J$ with respective distributions $\underline{p}_j$ and Ordered Categorical Means $\mu_{OC,j}$ then a means based index of inequality of opportunity is given by $MBG = 2\sum_{j=2}^{J}\sum_{l=1}^{j}\left|\mu_{OC,j} - \mu_{OC,l}\right| / (J^2 - J)\mu_{OCA}$ where μ_{OCA} is the average ordered categorical mean. However, it is readily shown that difference in means is not a sufficient statistic for the commonality of distribution required for Equality of Opportunity (think of a collection of distributions with identical means but different higher moments). However, this can be examined using a multilateral distributional variation measure of Inequality of Opportunity MDV (Anderson et. al. 2021). Letting $OV_{j,l} = \sum_{k=1}^{K} min\left(p_{k,j}, p_{k,l}\right)$ consider:

$$MDV = 2\sum_{j=2}^{J}\sum_{l=1}^{j}\left(1 - OV_{j,l}\right) / (J^2 - J)$$

2. An Exploration of Health Outcomes in 21st Century Pre-Covid U.K.

The Functionings and Capabilities approach to a nations wellbeing argues that the overall level of good health and the equality of access to its achievement are integral components of a nations’ wellbeing. To examine the pre–covid progress of health outcomes in the United Kingdom, data was sourced from the Understanding Society survey dataset (University of Essex, Institute for Social and Economic Research 2022), a large-scale longitudinal study conducted in the United Kingdom. The seven self-reported Health Status Categories were labelled CAT1 Completely dissatisfied, CAT2 Mostly dissatisfied, CAT3

Somewhat dissatisfied, CAT4 Neither dissatisfied nor satisfied, CAT5 Somewhat Satisfied, CAT6 Mostly Satisfied, CAT7 Completely Satisfied. Table 1 reports the Density Functions, Cumulative Distribution Functions and Survival Functions for the overall populations in 2010 and 2018 together with the corresponding Ordered Categorical Means, Gini Coefficients and Ambiguity Measures. With a value of 0.9602 and a standard deviation of 0.0014, the distributional similarity test OV clearly rejects the null hypothesis of identical distributions in 2010 and 2018 (z = 27.6551).

Table 1.

Overall	CAT1	CAT2	CAT3	CAT4	CAT5	CAT6	CAT7
2010 PDF	0.04269	0.06846	0.13865	0.08337	0.14618	0.40395	0.11670
2018 PDF	0.03872	0.07657	0.13830	0.09845	0.16280	0.38811	0.09705
2010 CDF	0.04269	0.11115	0.24980	0.33316	0.47935	0.88330	1.00000
2018 CDF	0.03872	0.11529	0.25359	0.35204	0.51484	0.90295	1.00000
2010 SF	0.95731	0.88885	0.75020	0.66684	0.52065	0.11670	0.00000
2018 SF	0.96128	0.88471	0.74641	0.64796	0.48516	0.09705	0.00000

	μ_{OC}	G_{OC}
2010	3.90056	0.10244
2018	3.82258	0.10736
$UNAM$	0.90770	

Table 2

Females	CAT1	CAT2	CAT3	CAT4	CAT5	CAT6	CAT7
2010 PDF	0.04581	0.07201	0.14039	0.08247	0.14282	0.39852	0.11798
2018 PDF	0.04213	0.08358	0.14034	0.09708	0.16395	0.37454	0.09838
2010 CDF	0.04581	0.11782	0.25821	0.34068	0.48350	0.88202	1.00000
2018 CDF	0.04213	0.12571	0.26605	0.36313	0.52708	0.90162	1.00000
2010 SF	0.95419	0.88218	0.74179	0.65932	0.51650	0.11798	0.00000
2018 SF	0.95787	0.87429	0.73395	0.63687	0.47292	0.09838	0.00000

	μ_{OC}	G_{OC}
2010	3.87197	0.10282
2018	3.77427	0.10849
$UNAM$	0.92997	

Table 3.

Males	CAT1	CAT2	CAT3	CAT4	CAT5	CAT6	CAT7
2010 PDF	0.03872	0.06396	0.13644	0.08450	0.15045	0.41085	0.11507
2018 PDF	0.03452	0.06792	0.13578	0.10015	0.16139	0.40483	0.09541
2010 CDF	0.03872	0.10269	0.23913	0.32363	0.47408	0.88493	1.00000
2018 CDF	0.03452	0.10244	0.23822	0.33837	0.49976	0.90459	1.00000
2010 SF	0.96128	0.89731	0.76087	0.67637	0.52592	0.11507	0.00000
2018 SF	0.96548	0.89756	0.76178	0.66163	0.50024	0.09541	0.00000

	μ_{OC}	G_{OC}
2010	3.93682	0.10146
2018	3.88211	0.10601
$UNAM$	0.83626	

The decline in the overall level of health (though not unequivocal, $UNAM < 1$) and the rise in the inequality of health outcomes is noteworthy, a configuration that is replicated in the gender specific results reported in Table 2 for Females and Table 3 for Males. With respective OV (Standard Error) [z statistic] values of 0.9527 (0.0021) [22.5481] for Females and 0.9694 (0.0019) [16.0805] for males the intertemporal difference in distribution of outcomes cannot be rejected for both genders. Equally noteworthy is the evidence that, consistent with Case and Paxson (2005), Nusselder et. al. (2010), Oksuzyen et. al. (2009) and Van Oyen et. al. (2013) results, males have better average health (though with respective $UNAM$ stats of 0.9177 and 0.9478 not unequivocably so) and less inequality in outcomes than do females in both years. Turning to an age group analysis, Tables 4 and 5 report overall data for <35, 35-65 and > 65 age groupings for the years 2010 and 2018. The distributional similarity tests uniformly reject similarity between cohorts in both observation years[5] and the differences between the cohorts are unambiguous in all cases.

Table 4.

2010	CAT1	CAT2	CAT3	CAT4	CAT5	CAT6	CAT7
<35 PDF	0.02613	0.04998	0.11796	0.08203	0.16132	0.40817	0.15441
35-65 PDF	0.04880	0.07436	0.14654	0.07881	0.13910	0.41179	0.10059
>65 PDF	0.05304	0.08270	0.15089	0.09875	0.14100	0.37421	0.09940
<35 CDF	0.02613	0.07611	0.19406	0.27609	0.43741	0.84559	1.00000
35-65 CDF	0.04880	0.12316	0.26970	0.34852	0.48762	0.89941	1.00000
>65 CDF	0.05304	0.13574	0.28663	0.38539	0.52639	0.90060	1.00000
<35 SF	0.97387	0.92389	0.80594	0.72391	0.56259	0.15441	0.00000
35-65 SF	0.95120	0.87684	0.73030	0.65148	0.51238	0.10059	0.00000
>65 SF	0.94696	0.86426	0.71337	0.61461	0.47361	0.09940	0.00000

	μ_{OC}	G_{OC}
<35	4.14461	0.09494
35-65	3.82278	0.10495
>65	3.71221	0.10914

$UNAM$ <35 - 35-65 difference 1.00000, 35-65 - >65 difference 1.00000

[5] 2010 similarity test for <35 - 35-65 0.9207 (0.0033) [24.2145], 2010 similarity test for 35-65 - >65 0.9612 (0.0030) [12.7467], 2018 similarity test for <35 - 35-65 0.9104 (0.0044) [20.4146], 2018 similarity test for 35-65 - >65 0.9806 (0.00220 [8.6322]

Table 5.

2018	CAT1	CAT2	CAT3	CAT4	CAT5	CAT6	CAT7
<35 PDF	0.02915	0.05671	0.11600	0.09860	0.18055	0.37408	0.14490
35-65 PDF	0.04240	0.08605	0.14811	0.09670	0.15732	0.38902	0.08040
>65 PDF	0.04105	0.07717	0.14088	0.10229	0.15543	0.40176	0.08142
<35 CDF	0.02915	0.08586	0.20186	0.30047	0.48101	0.85510	1.00000
35-65 CDF	0.04240	0.12845	0.27657	0.37326	0.53058	0.91960	1.00000
>65 CDF	0.04105	0.11822	0.25910	0.36139	0.51682	0.91858	1.00000
<35 SF	0.97085	0.91414	0.79814	0.69953	0.51899	0.14490	0.00000
35-65 SF	0.95760	0.87155	0.72343	0.62674	0.46942	0.08040	0.00000
>65 SF	0.95895	0.88178	0.74090	0.63861	0.48318	0.08142	0.00000

	μ_{OC}	G_{OC}
<35	4.04655	0.09901
35-65	3.72915	0.11095
>65	3.78484	0.10928

$UNAM$ <35 - 35-65 difference 1.00000, 35-65 - >65 difference -1.00000

With one exception, seniors in 2018, health levels deteriorate with age group, consistent with Deaton and Paxson (1998) and Kerkofs and Lindeboom (1997). Similarly, inequality increases with age group again with the exception of the elderly in 2018. To consider the progress of equality of opportunity for good health treating age-group and gender as circumstances beyond control, Table 6 reports the gender specific age group results.

Table 6. Gender and Age results.

	2010		2018	
	G_{OC}	μ_{OC}	G_{OC}	μ_{OC}
Female <35	0.09602	4.11202	0.10079	4.00393
Female 35-65	0.10444	3.80583	0.11136	3.68498
Female >65	0.11043	3.67374	0.11075	3.73645
Male <35	0.09328	4.20533	0.09632	4.11366
Male 35-65	0.10504	3.85734	0.10998	3.79705
Male >65	0.10768	3.75710	0.10763	3.84041
Inequality of Opportunity Indices				
MDV	0.07057		0.06858	
MBG	0.06451		0.05213	

The same pattern of diminishing health levels and increasing inequality with age group except for the elderly in 2018 is observed in the age-group-gender-specific results, with decreasing health levels and increasing inequality (elderly males excepted) for all groups.

Note that while overall, male and female specific inequality went up over the period, age group and gender-based inequality of opportunity diminished whether measured by the difference in means measure or the difference in distributions measure. Notice also that the difference in distributions measures are always greater than the difference in distributional location measure since they capture more than just locational differences.

A multidimensional analysis.

An integral part of the Functionings and Capabilities Approach to a society's wellbeing (Nussbaum and Sen 1993, Sen 1999) is that the populace should live long and healthy lives. By Reversing the ordering of the age-groups the joint distribution of health and anticipated future life length can be estimated. A mean long and healthy life index can be obtained from the corresponding Survival Function, and a Gini coefficient can be obtained from the Cumulative Distribution Function. Tables 7 and 8 report the distributions for 2010 and 2018 respectively. With a distributional overlap statistic and standard error of 0.92457 and 0.00194 respectively the commonality of distribution in the two years is strongly rejected. The positive sum of the 2010-2018 Survival Function differences (0.40943) with an unambiguity factor of 0.96316 indicates that 2010 outcomes "almost" first order dominate 2018 outcomes (Leshno and Levy 2002) indicating a decline in the long and health lived lives of the nation which is reflected in the lower mean long and health life index in 2018 and a higher Gini coefficient in that year.

Table 7. Health and Unexpired Life Distributions 2010

PDF	CAT1	CAT2	CAT3	CAT4	CAT5	CAT6	CAT7
>65	0.01004	0.01576	0.02914	0.01891	0.02750	0.07314	0.01898
35-65	0.02550	0.03835	0.07611	0.04063	0.07162	0.21442	0.05291
<35	0.00714	0.01435	0.03340	0.02382	0.04706	0.11640	0.04480
CDF							
>65	0.01004	0.02580	0.05494	0.07385	0.10136	0.17449	0.19348
35-65	0.03554	0.08965	0.19490	0.25445	0.35357	0.64113	0.71302
<35	0.04269	0.11115	0.24980	0.33316	0.47935	0.88330	1.00000
Survival function							
>65	0.98996	0.97420	0.94506	0.92615	0.89864	0.82551	0.80652
35-65	0.96446	0.91035	0.80510	0.74555	0.64643	0.35887	0.28698
<35	0.95731	0.88885	0.75020	0.66684	0.52065	0.11670	0.00000

N=43411, MOCM = 14.98431, Gini = 0.01657.

Table 8. Health and Unexpired Life Distributions 2018

PDF	CAT1	CAT2	CAT3	CAT4	CAT5	CAT6	CAT7
>65	0.00995	0.01886	0.03392	0.02454	0.03775	0.09833	0.01973
35-65	0.02182	0.04408	0.07669	0.05007	0.08156	0.19978	0.04209
<35	0.00695	0.01362	0.02768	0.02385	0.04349	0.09000	0.03523
CDF							
>65	0.00995	0.02881	0.06273	0.08726	0.12502	0.22334	0.24308
35-65	0.03177	0.09471	0.20532	0.27993	0.39924	0.69734	0.75917
<35	0.03872	0.11529	0.25359	0.35204	0.51484	0.90295	1.00000
Survival function							
>65	0.99005	0.97119	0.93727	0.91274	0.87498	0.77666	0.75692
35-65	0.96823	0.90529	0.79468	0.72007	0.60076	0.30266	0.24083
<35	0.96128	0.88471	0.74641	0.64796	0.48516	0.09705	0.00000

N=32076, MOCM = 14.57492, Gini = 0.01768.

Conclusions.

The ambiguity problems associated with applying artificial scaling to ordinal data have been circumvented by applying the construct of probabilistic distance, whereby the distance between two categories is quantified in terms of the probability that an observation could be realized between those categories. It has been possible to develop unambiguous tools for measuring the locational level of, and inequalities within ordered categorical distributions analogous to the mean and Gini coefficient employed in cardinal world distributions. Furthermore, these tools are readily extended to multi-dimensional environments. A Simple exemplifying application to the study of self-reported health outcomes in the United Kingdom over the pre-covid period 2010-2018 highlighted the fact that health outcome levels were deteriorating with inequalities in those outcomes increasing. However, it was also determined that equality of opportunity in achieving good health levels improved over the period.

Appendix

Multidimensionality.

μ_{OC} is readily extended to the multidimensional case, for example when the situation is two dimensioned with ordered outcomes indexed $i = 1,..,K$ in one dimension and $j = 1,..,H$ in the other, $\underline{P}$, the cumulative density matrix, becomes a K x H matrix based upon the similarly dimensioned matrix $\underline{p}$ with typical element $p_{k,h}$ where $P_{k,h} = \sum_{l=1}^{k} \sum_{m=1}^{h} p_{l,m}$. The corresponding Survival function matrix $\underline{S}$ will have typical elements $S_{k,h} = 1 - P_{k,h}$ and μ_{MOC} the multidimensional ordered categorical mean will be of the form:

$$\mu_{MOC} = \sum_{j=1}^{H} \sum_{i=1}^{K} S_{j,i}$$

μ_{MOC} will have a minimum value of 0 when $p_{1,1} = 1$ and a maximum value of $LK - 1$ when $p_{L,K} = 1$.

A multidimensional Ordered Categorical Gini coefficient can be developed in a similar fashion. In the bivariate case the Cumulative Density Matrix $\underline{P}$ can be used to define $d\left(i(k_i, h_i), j(k_j, h_j)\right)$, the probabilistic distance between category combination i and category combination j. Suppose $k_i \geq k_j$ and $h_i \geq h_j$ with strict inequality somewhere, then $P_{k_i,h_i} - P_{k_j,h_j}$ = $\sum_{h=h_j+1}^{h_i} \sum_{k=k_j+1}^{k_i} p_k$ includes the chance of being in category combination $i(k_i, h_i)$ and excludes the chance of being in category combination $j(k_j, h_j)$, whilst $P_{k_i-1,h_i-1} - P_{k_j-1,h_j-1}$ = $\sum_{h=h_j+1}^{h_i} \sum_{k=k_j+1}^{k_i} p_k$ excludes the chance of being in category combination $i(k_i, h_i)$ and includes the chance of being in category combination $j(k_j, h_j)$, therefor letting $d\left(i(k_i, h_i), j(k_j, h_j)\right) =$

$$0.5\left(\left(P_{k_i,h_i} - P_{k_j,h_j}\right) + \left(P_{k_i-1,h_i-1} - P_{k_j-1,h_j-1}\right)\right) = \left(P_{k_i,h_i} - P_{k_j,h_j}\right) + 0.5\left(p_{k_j,h_j} - p_{k_i,h_i}\right).$$

D, the total probabilistic distance in the distribution will be given by $D = \sum_{m=2}^{H} \sum_{h_j=1}^{m-1} \sum_{i=2}^{K} \sum_{k_j=1}^{i-1} d\left(i(k_i, h_i), j(k_j, h_j)\right)$ whilst the average probabilistic distance is given by $D/(ND)$ where ND is the number of distributional distances computed making G_{MOC} a Gini-like inequality measure where:

$$G_{MOC} = D/(ND)\mu_{MOC})$$

And where μ_{MOC} is the multidimensional Ordered Categorical Mean.

Inference.

Assume an independently observed sample from population a with n_a observations and let the true k'th outcome level probability $p_{a,k}$ for $k = 1, .., \mathrm{K}$ be stacked in the K x 1 vector $\underline{p}_a$ and let $\hat{\underline{p}}_a$ be the corresponding relative frequency estimates of those probabilities. Then, following Rao (2009), $\sqrt{n_t}\left(\hat{\underline{p}}_a - \underline{p}_a\right) \sim_{asymp} N\left(0, V\left(\underline{p}_a\right)\right)$ where:

$$V\left(\underline{p_t}\right)=\begin{pmatrix} p_{\mathrm{a},1} & 0 & . & 0 \\ 0 & p_{\mathrm{a},2} & . & 0 \\ . & . & . & . \\ 0 & 0 & . & p_{\mathrm{a,K}} \end{pmatrix}-\begin{pmatrix} p_{\mathrm{a},1}^2 & p_{\mathrm{a},1}p_{\mathrm{a},2} & . & p_{\mathrm{a},1}p_{\mathrm{a,K}} \\ p_{\mathrm{a},2}p_{\mathrm{a},1} & p_{\mathrm{a},2}^2 & . & p_{\mathrm{a},2}p_{\mathrm{a,K}} \\ . & . & . & . \\ p_{\mathrm{a,K}}p_{\mathrm{a},1} & p_{\mathrm{a,M}}p_{\mathrm{a},2} & . & p_{\mathrm{a,K}}^2 \end{pmatrix}$$

Given the $K \times K$ dimensioned cumulating matrix D, where:

$$D=\begin{pmatrix} 1 & 0 & . & 0 \\ 1 & 1 & . & 0 \\ . & . & . & . \\ 1 & 1 & . & 1 \end{pmatrix}$$

$\underline{F_a}$, the vector of CDF values are such that, given $\underline{I}$ is an K dimensioned vector of ones:

$$\underline{F_t}=D\underline{p_a} \text{ and } \underline{S_a}=\underline{I}-D\underline{p_a}$$

Each will have variance $DV\left(\underline{p_a}\right)D'$, so that $\sqrt{n_a}\left(\underline{\hat{F}_a}-\underline{F_a}\right)\sim_{asym}N\left(0,DV\left(\underline{p_a}\right)D'\right)$ and $\sqrt{n_a}\left(\underline{\hat{S}_a}-\underline{S_a}\right)\sim_{asym}N\left(0,DV\left(\underline{p_a}\right)D'\right)$. Since $\mu_{OC,\mathrm{a}}=\underline{I}'\underline{S_a}$, it follows that:

$$\sqrt{n_t}\left(\hat{\mu}_{OC,\mathrm{a}}-\mu_{OC,\mathrm{a}}\right)\sim_{asym}N\left(0,\underline{I}'DV\left(\underline{p_a}\right)D'\underline{I}\right).$$

When populations are independently sampled, under the null hypothesis of common means:

$$\sqrt{\frac{n_a n_b}{(n_a+n_b)}}\left(\hat{\mu}_{OC,\mathrm{a}}-\hat{\mu}_{OC,\mathrm{b}}\right)\sim_{asym}N\left(0,\underline{I}'D\left(V\left(\underline{p_a}\right)+V\left(\underline{p_b}\right)\right)D'\underline{I}\right)$$

References.

Anderson, Gordon, Oliver Linton, and Yoon-Jae Whang (2012) “Nonparametric Estimation and Inference about the Overlap of Two Distributions” Oliver Linton and Yoon-Jae Whang. Journal of Econometrics 171, 1-23

Anderson, Gordon, Oliver Linton, Maria Grazia Pittau, Yoon-Jae Whang, and Roberto Zelli, (2021), “On Unit Free Assessment of Multilateral Distributional Variation,” The Econometrics Journal 24, 502–518.

Anderson, Gordon, Thierry Post, and Yoon-Jae Whang. (2020), “Somewhere Between Utopia and Dystopia: Choosing from Multiple Incomparable Prospects,” Journal of Business and Economic Statistics 38, 502-515.

Bond, Timothy N., and Kevin Lang, (2019), “The Sad Truth about Happiness Scales,” Journal of Political Economy 127, no. 4, 1629-1640.

Cantril, Hadley, (1965), “The Pattern of Human Concerns,” New Brunswick, NJ: Rutgers University Press.

Case, Anne, and Christina Paxson, (2005), “Sex Differences in Morbidity and Mortality,” Demography 42, 189–214

Andrew E. Clark, Sarah Flèche, and Claudia Senik (2016) Economic Growth Evens Out Happiness: Evidence from Six Surveys Rev Income Wealth. 2016 September ; 62(3): 405–419.

Deaton, Angus S., and Christina H. Paxson, (1998), “Aging and Inequality in Income and Health,” The American Economic Review 88(2), 248–253.

Ferreira F.H.G and Peragine, V. (2015) “Individual responsibility and Equality of Opportunity” Chapter 25 in the Oxford Handbook of Wellbeing and Public Policy edited by M.D. Adler and M. Fluerbaey.

Kennedy C, Popky D and S. Keeter (2023) “How Public Polling Has Changed in the 21st Century” Pew Research Centre Report.

Kerkhofs, Marcel, and Maarten Lindeboom, (1997), "Age Related Health Dynamics and Changes in Labour Market Status," Health economics 6, no. 4: 407-423.

Gary G. Koch, Catherine M. Tangen, Jin-Whan Jung, Ingrid A. Amara (1998) “Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them” Statistics in Medicine: 17, Issue 15-16 1661-1892

Benjamin E. Lauderdale, Delia Bailey , Jack Blumenau, Douglas Rivers (2020) “Model-based pre-election polling for national and sub-national outcomes in the US and UK” International Journal of Forecasting 36, 2, 399-413

Leshno, M. and Levy, H. (2002). Preferred by “all” and preferred by “most” decision makers: Almost stochastic dominance. Management Science, 48(8):1074–1085.

Likert, Rensis, (1932),"A Technique for the Measurement of Attitudes," Archives of Psychology 140(22), 5-55.

Mendelson, Haim, (1987), "Quantile-preserving Spread," Journal of Economic Theory 42, no. 2: 334-351.

Miller, Ray, Neha Bairoliya, and David Canning, (2019),"Health Disparities and the Socioeconomic Gradient in Elderly Life-cycle Consumption." The Journal of the Economics of Ageing 14: 100176

Nussbaum, Martha, and Amartya Sen (Eds.), (1993), “The Quality of Life,” Clarendon Press, Oxford.

Nusselder, W. J., C. W. N. Looman, Herman Van Oyen, Jean-Marie Robine, and Carol Jagger. (2010), "Gender Differences in Health of EU10 and EU15 Populations: the Double Burden of EU10 Men." European Journal of Ageing 7: 219-227.

Oksuzyan, Anna, Inge Petersen, Henrik Stovring, Paul Bingley, James W. Vaupel, and Kaare Christensen. (2009), "The Male–female Health–survival Paradox: A Survey and Register Study of the Impact of Sex-specific Selection and Information Bias," Annals of Epidemiology 19, no. 7: 504-511.

Van Oyen, Herman, Wilma Nusselder, Carol Jagger, Petra Kolip, Emmanuelle Cambois, Jean-Marie Robine, (2013), “Gender Differences in Healthy Life Years within the EU: an Exploration of the ‘Health–survival’ Paradox,” International Journal of Public Health 58, 143–155.

Rankin, John. (1957), "Cerebral vascular accidents in patients over the age of 60: II. Prognosis." Scottish medical journal 2, no. 5: 200-215.

Rao, Calyampudi Radhakrishna, (2009), “Linear Statistical Inference and its Applications,” 2 edn. New York: Wiley, 263-270.

Schröder, Carsten, and Shlomo Yitzhaki, (2017), "Revisiting the Evidence for Cardinal Treatment of Ordinal Variables," European Economic Review 92: 337-358.

Sen, Amartya, (1999), “Development as Freedom,” Knopf Press, New York.

University of Essex, Institute for Social and Economic Research. (2022). Understanding Society: Waves 1-12, 2009-2021 and Harmonised BHPS: Waves 1-18, 1991-2009. [data collection]. 17th Edition. UK Data Service. SN: 6614, http://doi.org/10.5255/UKDA-SN-6614-18.