

# 序值在定量经验研究中的应用

夏传玲

**提要:** 本文通过序值分析, 试图解决定量经验研究中经常面临的两个重要问题: 一是定序变量的描述统计缺乏统计量及其统计检验; 二是量表分析仅关注测量结构, 不关注测量制式。这两个问题常导致参数解释和研究结论出现偏差。本文以阶层主观认同和国家认同研究为示例, 探讨了不同统计量和分析方法所带来的偏差以及采用序值分析来消除这些偏差的可行性。

**关键词:** 序值分析 主观阶层认同 国家认同

在社会学的经验研究中, 我们常常会遇到定序变量,<sup>①</sup>即只能区分高低无法区分大小的随机数组。例如, 在问卷中采用选项“很好、较好、一般、较差、很差”所得到的变量即为定序变量。尽管在编码时我们可以把这些变量方便地编码为1~5之间的整数, 但它们并不具有定距变量的测量属性(例如等距性和原点), 因而不能采用加减乘除等算术运算, 更不能采用反映定距变量集中度和变异度的统计量(例如均值和方差)来进行统计描述和推理(例如t检验)。

如何对定序变量进行描述和统计检验? 在统计学中, 这两个问题已经得到较好的解答。但在社会学的经验研究中, 因为缺乏相关的计算机程序, 统计学的解决办法并没有得到广泛的应用。首先, 本文将对这方面的统计学研究进行简述,<sup>②</sup>并编写一个Stata的程序riditei。<sup>③</sup>把相关统计学的理论和算法研究成果转变为经验研究中的实用工具。然后, 本文将利用这个新工具, 对一些已经获得发表的社会学经验研究进行再分析, 并以示例说明在分析定序变量时使用不恰当的统计量给经验研究结论所带来的潜在偏差。

---

① 这里需要区分排名(ranking)和定序(ordinal)。前者不会形成连续分布, 且个案取值随规模而变化; 后者可形成分布, 个案取值与规模无关(Borgatta, 1968)。

② 参见詹森(Jansen, 1984)的文章。

③ 相对于R平台上的ridit和ridittools, Stata平台下的wridit和fridit等软件包而言, riditei聚焦于序值均值的组间比较的统计检验, 弥补了wridit和fridit的不足; 同时, 输出两种界定的序值均值、其差值的置信区间、Z值、输赢平三个概率、输赢几率比等9个结果。在多组比较的情形下, 输出两两比较的统计检验, 方便确认组间差异的具体来源。读者如有需要, 请与本文作者联系(Email: xiacl@cass.org.cn)。

## 一、序值概念

如何表达定序变量的集中度? 布罗斯最早提出一个新统计量, 英文缩写为 *ridit*,<sup>①</sup> 表示一个定序分布累积概率分布的赋值 (Bross, 1958)。*ridit* 中文可译为“序值”, 它的取名模仿 *logit* (概率比对数制式) 和 *probit* (概率比正态制式), 表示相对于同一个分布而取得的概率度量值。在布罗斯的界定中, 一个变量的取值  $x$  的序值, 是将其投射到一个分布  $X$  上而取得的位次概率, 即:

$$\text{ridit}(x) = P(X < x) + \frac{1}{2}P(X = x)$$

这是常识中的“赢面”概念的数学表达, 布罗斯将它界定为胜出概率和平局概率的一半之和。此时, 序值的取值范围是  $[0, 1]$ , 它在一个均匀分布 (作为参照系) 的均值总是 0.5。此时, 参照组被看作一个无限总体, 其经验分布的统计量是总体参数 (Mielke et al., 2009)。

布罗斯的序值界定仅有一个假定, 即定序变量是一个连续潜变量的显现, 这个潜变量的分布可以是正态分布, 也可以是其他分布。但这个赢面的计算公式存在一个参照系任意性的问题。如果我们比较两个总体的序值时把它们当作一个总体, 就会带来统计推理上的误导性 (Uwawunkonye & Anaene, 2013), 因而不能处理不同组来自不同总体的情况 (Flora, 1974)。为此, 布罗克特和莱文提出了另外一种序值的界定 (Brockett & Levine, 1977), 即:

$$\text{ridit}'(x) = P(X < x) - P(X > x)$$

此时, 赢面是胜出概率和淘汰概率之差, 它的取值范围是  $[-1, 1]$ 。<sup>②</sup> 这个界定具有单一选项的无区分性、翻转顺序所具有的对称性、合并相邻类别不会影响其他未合并类别的赋值 (分叉属性)、两个类别的赋值差距不因概率增加而变小 (相对位次属性) (Brockett & Levine, 1977)。换言之, 不考虑平局的序值界定具

① 关于 *ridit* 的命名有两种解释。第一种认为 *rid* 是 *relative to an identified distribution* 的缩写, 表示相对于一个已知分布而言; *it* 是 *unit* 的缩写, 表示换算的类型; 两个缩写合并为 *ridit*。第二种解释则认为 *ridit* 是 *relative to an identified distribution integral transformation* 的缩写, 将一个已知的函数空间通过概率积分换算映射到另一个函数空间。前一种阐释强调转换的结果 (制式不同), 后一种强调转换的方法 (参见 Flora, 2006: 7280 - 7283)。

② 在 Stata 的 *wridit* 和 *fridit* 命令中, 布罗斯的界定称作未折叠 (*unfolded*), 布罗克特的界定称作折叠 (*folded*)。该命令的权重可取值为 0, 这样我们就可以依据一个目标总体来计算当前样本的序值。

有更符合定序测量的数学属性。

由于胜出、淘汰、平局三种选项的概率之和为 1 (选项的穷尽性), 这两种界定之间具有下列换算关系:

$$\text{ridit}(x) = \frac{1}{2}[\text{ridit}'(x) + 1]$$

尽管在数学上这两种界定存在上述换算关系, 但从应用的角度来看, 布罗克特和莱文的界定更清晰, 没有假定平局算“半赢”的任意性, 且其序值具有正负性, 能够反映对比双方的强弱方向。因此, 在本文中我们采用布罗克特和莱文的序值界定,<sup>①</sup>以下不再说明。

## 二、序值的统计检验

如何对序值均值之差进行统计推断? 不同的序值界定给出了不同的答案。如果以总体为参照组, 布罗斯的序值均值就会逼近服从正态分布, 其方差公式是:

$$\text{Var}(\text{ridit}) = \frac{1}{12n}$$

即序值均值的方差是其样本量 12 倍的倒数。在分组平均序值等于 0.5 的零假设下,  $\text{Mean}(\text{ridit})/\text{Var}(\text{ridit})$  服从自由度为 1 的卡方分布 (Uwawunkonye & Anaene, 2013)。但有学者指出, 布罗斯的序值分析只能用于描述统计, 因为上述方差公式计算的是一个  $[0, 1]$  区间内均匀分布变量的方差, 它让显著性检验变得保守 (Beder & Heim, 1990; Mantel, 1979)。

为此, 在康诺弗 (Conover, 1973) 的统计学理论研究的基础上, 福洛拉 (Flora, 1974) 提出了基于秩统计量的非参数检验程序。它不再假定定序变量是一个连续潜变量的显现, 更不用假定任何特定的连续统计分布 (例如标准正态分布)。换言之, 该非参数检验方法并不假定总体分布的数学属性, 但用于进行统计推断的统计量仍然是一种参数  $Z$ 。

<sup>①</sup> 两种界定的具体计算示例可参见福洛拉 (Flora, 1974) 一文的第三节。在田忌赛马的典故中, 比赛双方马匹的定序测量结果是上中下各 1 匹马, 此时, 田忌和齐王两组间的布罗斯序值均值是 0.50, 布罗克特和莱文的序值均值是 0.00, 表示在定序测量上两组旗鼓相当, 齐王组胜在定序内部的微弱差异上。孙臆的策略让定序测量临时变成 5 级, 从低到高排列, 齐王组  $[1, 1, 1, 0, 0]$ , 田忌组是  $[0, 0, 1, 1, 1]$ , 此时, 布罗斯序值均值是 0.94, 布罗克特和莱文的序值均值是 0.89, 田忌组在定序上胜出。

按照布罗克特和莱文的序值定义,我们先计算两个组别个体间两两比较的输赢结果,赋值赢为 1,输为 -1,平为 0,即:

$$D_{ij} = \begin{cases} 1 & \text{若 } X_{1i} > X_{2j}, \text{ 其中 } i = 1, \dots, n, \text{ 且 } j = 1, \dots, m \\ 0 & \text{若 } X_{1i} = X_{2j}, \text{ 其中 } i = 1, \dots, n, \text{ 且 } j = 1, \dots, m \\ -1 & \text{若 } X_{1i} < X_{2j}, \text{ 其中 } i = 1, \dots, n, \text{ 且 } j = 1, \dots, m \end{cases}$$

其中,  $X_{1i}$  表示第一组的个体  $i$  在定序变量  $X$  上的取值,  $X_{2j}$  表示第二组的个体  $j$  在定序变量  $X$  上的取值,  $D_{ij}$  表示第一组的个体  $i$  和第二组的个体  $j$  之间的对比结果。第一组的样本规模为  $n$ , 第二组的样本规模为  $m$ 。由此,我们得到一个统计量  $W$ :

$$W = \sum_{i=1}^n \sum_{j=1}^m D_{ij}$$

它是两组间赢和输之差的总和,由此可见,  $\frac{W}{mn}$  是两组间赢率和输率之差。对较大的  $m$  和  $n$  而言,  $W$  逼近一个正态分布,逼近程度取决于两组各自的样本量  $m$  和  $n$ ,以及每一个类别的个案数。对大多数社会调查数据而言,这种逼近性都能得到满足。若两个组别来自相同总体,则它们的  $W$  的期望值为 0,这是我们检验时所采用的假设。

为此,我们需要计算统计量  $W$  的方差  $Var(W)$ ,即:

$$Var(W) = \frac{mn(N+1)}{3} \left( 1 - \frac{\sum_{i=1}^k (X_i^3 - X_i)}{N^3 - N} \right)$$

其中,  $N$  是样本量,即  $m+n$  之和,是定序变量  $X$  在  $i$  类别上两组个案数之和。将  $W$  标准化,我们就得到统计量  $Z$ ,它的计算公式定义为:

$$Z = \frac{W}{\sqrt{Var(W)}}$$

依据标准正态分布表,取得它的显著性水平或计算它的置信区间(Flora, 1974)。统计模拟研究表明,和常用的 Mann-Whitney-Wilcoxon 统计检验相比,此统计检验具有较低的假阳性,对小样本具有较高的统计力(Marfak et al., 2020)。可见,它是较好的分析定序变量的统计工具。以不太恰当的比喻来说,从布罗斯的序值方差公式到福洛拉的  $Z$  分方差计算公式的变化过程,是从淘汰赛到循环

赛的赛制转变。

另一种思路是利用两个或多个随机变量的经验特征函数(Lukacs, 1970)来判定它们是否来自相同的分布,即判定它们是否存在统计上可甄别的分布差异。此时,我们忽略定序变量的有序性,只考察它们的分布属性(例如位置、尺度等)。其中一个 是埃普斯和辛格 尔顿提出的通用特征函数检验(Epps & Singleton, 1986),简称为CF检验,它对小样本、偏态分布、来自不同分布家族的样本、离散变量等情形均具有较高的统计力。而且,模拟分布表明,CF检验比Kolmogorov-Smirnov(KS检验)、Anderson-Darling(AD检验)和Cramer-von Mises(CM检验)具有更广的应用范围(可应用于非连续变量的情形),且具有较高的统计力。因此,我们也把它列为本文的统计推理工具,<sup>①</sup>和基于序值的统计检验形成参照关系。

### 三、描述统计中的序值分析

在社会学的经验研究中,我们经常会遇到定序变量。例如,当前的研究热点之一中国人的主观阶层认同就是一个比较典型的示例。我们在其他地方已经讨论过这类研究在测量上的不恰当性(夏传玲,2021),<sup>②</sup>在这里,本文再讨论一些其他研究在处理定序变量时分析上的缺憾。

第一种分析缺憾是割裂定序变量的分布整体性,不用一个统计量来表征定序变量的序值,而只讨论其中一个或几个层序的百分比。例如,有研究在讨论不同时期、不同国别的主观认同时,有时比较中层的比例,有时比较下层的比例,有时比较上层的比例,并作出“超过”“远远高于”等主观判断(张翼,2011),我们重复分析该研究表1的数据并加以分析对比。<sup>③</sup>

① 该检验在Stata中的命令是escftest。

② 测量主观阶层认同的题器至少需要两个,一个测量被访者的主观阶层数(例如“您认为中国社会存在几个阶层?”),另一个测量被访者在其中的位置(例如“您在其中的哪一个阶层上?”)。和国家认同不同,阶层既没有地理边界,也没有符号边界,在这个意义上,它是一些经验研究所自创的“想象的共同体”,特别是在事先固定主观阶层数量的情形下。

③ 原分析没有对调查数据进行加权处理,默认各国、各时期的样本构成和各自的目标总体构成一致。这个假定是值得商榷的,但因不在本文的讨论范围之内,故存而不论。原表1中还包含韩国,因为没有样本量信息,无法进行统计推断,此处略去不提。国外数据存在行百分比不等于100的现象,原表如此,因本文关注方法,故没有使用原始数据校对,原表照搬,换算时以个案数乘以百分比,取整数得到频次数,以保持分析结果的可比性。

由表1的序值分析可见,以原西德的主观阶层分布为参照,巴西和新加坡的主观阶层认同分布和原西德的分布没有统计上的显著差异,这三个国家具有相同的主观阶层认同分布。加拿大、澳大利亚和美国比原西德的主观阶层认同程度更高,序值均值分别高0.102、0.063和0.057。即和原西德的被访者相比,来自加拿大、澳大利亚和美国的被访者具有更高的主观阶层认同的概率分别是10.2%、6.3%和5.7%。意大利、法国、印度、日本和菲律宾比原西德的主观阶层认同更低,序值均值分别低0.116、0.076、0.055、0.048和0.047。即和原西德的被访者相比,来自意大利、法国、印度、日本和菲律宾的被访者具有更低的主观阶层认同的概率分别是11.6%、7.6%、5.5%、4.8%和4.7%。差距最大的国家是中国,来自中国的被访者具有更低的主观阶层认同的概率最低是17.8%,最高为38.1%。由此可见,主观阶层认同的国别高低次序如下:加拿大、澳大利亚、美国、巴西、原西德、新加坡、菲律宾、日本、印度、法国、意大利和中国,一些相邻国家的主观阶层认同不具有统计上的显著差异。

表1 主观阶层认同的序值分析

组别	上层	中上层	中层	中下层	下层	个案数	序值差
原西德	1.8	11.2	62.5	20.0	3.6	1127	—
美国	1.9	15.7	60.7	17.4	3.6	987	0.057 **
法国	0.4	10.9	57.7	25.2	5.3	993	-0.076 ***
意大利	0.7	7.0	56.9	22.2	8.0	1000	-0.116 ***
澳大利亚	1.1	8.6	72.8	10.4	2.7	1104	0.063 ***
加拿大	1.2	14.2	68.8	11.8	2.2	1012	0.102 ***
巴西	4.4	13.1	57.4	17.2	5.5	1000	0.040
日本	1.1	12.5	56.0	24.4	5.0	1042	-0.048 **
新加坡	1.0	3.9	74.2	16.2	3.0	996	-0.024
印度	1.2	12.0	57.5	21.7	7.5	1020	-0.055 **
菲律宾	1.3	7.0	67.1	18.5	5.9	1574	-0.047 **
中国2002省会	1.6	10.4	46.9	26.5	14.6	10738	-0.178 ***
中国2006全国	0.5	6.2	41.0	29.3	23.1	6789	-0.333 ***
中国2006城镇	0.3	5.4	37.5	30.7	26.1	2822	-0.385 ***
中国2006农村	0.7	6.7	41.2	27.8	23.5	4142	-0.320 ***
中国2008全国	0.8	6.8	40.3	30.4	21.7	7014	-0.320 ***
中国2008城市	0.9	7.4	36.3	35.8	19.6	236	-0.338 ***
中国2008区县	0.8	2.1	41.6	36.7	18.9	186	-0.380 ***
中国2008城镇	0.6	6.1	35.5	33.9	24.0	2942	-0.381 ***
中国2008农村	0.9	7.5	44.4	27.5	19.7	3831	-0.269 ***

注:(1)各阶层栏中的单位是百分比。(2)\*\*  $P < 0.05$ , \*\*\*  $P < 0.01$ 。(3)所有CF检验均显著,说明该统计检验更“敏感”。

忽略统计检验是定序变量分析常见的第二种瑕疵。例如,上述示例依据认同中层的比例相近,得出了印度和法国“差不多”的结论(张翼,2011)。实际上,我们可以依据两组的序值均值差为 $-0.019$ 、 $Z$ 值为 $-0.807$ 判定两国在主观阶层认同上具有相同的分布,没有统计上显著的差异,让分析结论具有更扎实的统计推理依据。同样,在处理年份间的比较时,上述分析通过比较2006年度调查数据和2008年度调查数据,以上层、中上层、中层、下层的比例变化,得出2008年中国的主观阶层认同有“上移”趋势(张翼,2011)。但从表1可知,以原西德为参照组,2006年中国主观阶层认同的序值均值是 $-0.333$ ,2008年是 $-0.320$ ,两者有 $0.013$ 左右的差距。这个差距的 $Z$ 值是 $-1.444$ ,它的绝对值低于 $0.05$ 置信水平上 $1.96$ 的临界值,说明两者在主观阶层认同上的分布并没有统计上显著的差异。与2006年相比,尽管2008年经历了金融危机,但中国人的主观阶层认同依然未变。这个结论和原分析的“上移”不同,但更具有统计推理的依据,减少了主观判定的任意性。

但前文(张翼,2011)结论并没有在另外一个使用相同调查数据集的研究(陈光金,2013)中得到重复。在描述中国主观阶层认同的十年变迁时,该研究利用“中国社会状况综合调查”四次截面调查数据(参见表2),以各层比例的合并百分比为主观判断依据,得出“在最近的十年中,被调查者的主观经济社会地位认同结构<sup>①</sup>出现了前期明显下沉,然后不稳定上扬的趋势”的结论(陈光金,2013)。特别是在比较2006年和2008年间的变化时,该研究认为,“相对于2006年的调查结果,认同中等及以上地位的被调查者比重有所上升,而认同下等地位者所占比重则有所下降”,两个研究的结论相同,但数据显示的结果却是两种分析之间的不一致。两者报告的有效个案数不同,各主观阶层的百分比也有细微差异。以表2的数据来看,2008年和2006年的主观阶层认同的序值均值差是 $0.057$ ,对应的序值 $Z$ 值是 $6.212$ ,此值大于 $1.96$ ,表明这两个年份的主观阶层认同来自不同的分布总体,两者的差距具有统计上的显著性,存在“上升”的特征。表2的序值表明,以2001年的主观阶层认同分布为参照,2006年下降 $17.6\%$ ,在2008年、2011年则保持“回升”的态势,它们的序值均值的下降幅度分别是 $12.2\%$ 和 $8.5\%$ 。可见,依据统计检验所得到的变动模式应表述为:在2001—2011的十年间,被访者的经济社会地位主观认同在前5年间出现明显下沉,幅度达 $17.6\%$ ;后5年间呈现稳定上扬的态势,但上扬的幅度在衰减,总水平仍未“恢复”到比较基点的水平,仍比基点低 $8.6\%$ 。

<sup>①</sup> 实际上是模式,而不是结构。

表 2 主观阶层认同的序值分析

年份	上层	中上层	中层	中下层	下层	个案数	序值差
2001	1.9	11.0	45.9	23.5	17.7	6193	—
2006	0.4	5.5	38.5	29.4	26.1	6994	-0.176***
2008	0.7	7.2	39.9	30.8	21.5	7045	-0.122***
2011	0.5	7.7	43.0	30.2	18.7	6924	-0.086***

注:(1)各阶层栏中的单位是百分比。(2)\*\* $P < 0.05$ ,\*\*\* $P < 0.01$ 。(3)所有 CF 检验均显著。

#### 四、量表建构中的加权序值

量表是一种利用有形、可观察的外在特征(即指标)来表征无形、不可观察的理论概念(因子或潜变量)的测量工具。它是社会学由表及里、由外而内、理论和经验相互参验的重要认知工具。目前,社会学经验研究中对定序指标变量的常规处理方法有两种:升序和映射。升序是将指标变量的测量层次从定序上升为定距层次,映射是将指标变量从定序层次映射到一个定距潜变量上。

升序方法的开端是瑟斯通量表,它是由一组只有赞同和不赞同两个选项的题器构成,将赞同项编码为 1、不赞同编码为 0,量表分是各题器赞同项得分的总和(Thurstone & Chave, 1929)。在此基础上,李克特将题器的选项从两个扩展到多个(常见的是 5 个)并按照顺序进行编码(例如,1~5),以这些编码在各题器上的总和作为量表分(Likert, 1932)。李克特量表的选项可以是单极,也可以是双极,除了顺序编码之外,也可以采用其他编码和计分方法。

在上面这两种量表中,每个题器对量表的分值贡献是相等的。但古特曼(Guttman, 1944)认为,一种态度或其他属性可能有很多种行为或其他指标来表达,反映它们的题器总和构成一个总体。但其中的每一个题器和待测态度之间的关系可能有强弱之分。因此,他将量表中的每一个题器按照一个维度进行排序,确定相对重要性,然后考察被访者的应答模式,以确定其量表得分(Guttman, 1944)。在量表建构时,古特曼已经关注到了指标的难易度对建构量表的意义。

与此同时,在斯皮尔曼的因子测量模型(指标方差是由因子和误差构成)的基础上,瑟斯通(Thorndike, 2005)开发了因子析取和因子转置技术,形成我们现在常用的探索性因子分析。指标和因子不是简单的直接对应,而是通过因子负

荷在两者之间建立回归关系。这样,每一个指标和因子的关系强度就会不同,但不必像古特曼量表那样在指标选择时就加以判定。因子分是各指标变量的回归方程的预测值,它将测量模型中的误差项排除在外,比瑟斯通量表和李克特量表的计分方法更加科学。同时,误差项的大小还是测量模型稳定性的重要指标,这是信度系数的来源。

但升序显然违反了定序测量的本质特征,只分大小、不问距离。相应地,在数学运算上,定序变量的赋值之间只能进行不等式运算,不能进行加减乘除等算术运算。有鉴于此,我们可以采用另一种考虑定序属性的分析方法——映射。映射是将一个定序变量的取值依据一定规则和另一个定距潜变量的取值之间建立一一对应关系,即假定定序指标变量  $X$  是一个定距潜变量  $U$  的粗略度量, $U$  的值域在正负无穷大之间,定序指标变量  $X$  的  $k$  个取值则是  $X$  在  $U$  的  $k-1$  个阈限所界定的区间上的投射。因为  $U$  是一个潜变量,我们无法了解其分布形态及其测量制式,统计学家们推荐的一种常用的解决方案是假定  $U$  服从一个标准正态分布  $N(0,1)$ 。这样,每一个定序指标变量都对应到一个标准正态分布上,即以 0 为均值,1 个标准差为测量单位的连续变量  $U$  (Muthén, 1984)。这种方法是常用的定序验证性因子分析,在计算上,它的基础是指标变量间的多相 (polychoric) 相关系数矩阵;而探索性因子分析的基础是指标变量间的皮尔逊线性相关系数矩阵。在心理学中,这种映射关系被称作“应答函数”,所采用的数学规则是几率比对数单位 (logit) 或几率比正态单位 (probit),所形成的测量模型称作“偏信用模型” (Masters, 1982)。在回溯的视野中,古特曼量表实际上是偏信用模型的特例。以应答函数为映射规则的因子分析被称作“答题论”,<sup>①</sup> 它将受访者的答题行为看作是两个参数的函数:一个参数是受访者的特质 (theta),这是经验研究所关注的主要成分;另一个是题器的难度系数 (beta),它是特质既定时间答题器本身难易程度的度量,在定序的情形下,这是由易到难的应答过程。答题论将古典测量模型中的真值 (tau) 分解为两个成分,因而是测量方法论上的重大进步。

定序验证性因子分析和偏信用模型避免了升序方法在数学运算上的错误,这样,它们对定序指标变量的处理在数学运算上就具有正当性。但它们也具有自己的弱点,主要有以下四个。一是正态转换的任意性,为什么是一个服从标准正态分布的潜变量  $U$ ? 特别是对一些明显偏态分布的定序变量来说,这种转换

<sup>①</sup> 对应英文表述为 item response theory,常被误译为“项目反应理论”。

的数学和经验正当性都值得商榷。二是各指标变量的均值相等假定,在一个自变量引发多个因变量的验证性因子分析模型中,各定序指标变量之间均以0为均值,这样,它们的原点位置就被人为设定在一个起点上,因而在总体上被假定为相等。第三是赋值的不可比性,每一个定序指标变量所取得的相同赋值的阈限却不相同。<sup>①</sup>换言之,相同的取值意味着不同的阈限值,这意味着不同的定序和定距之间的对应关系。也就是说,不同定序指标变量的取值尽管数值相同,但却不具有可比性或等价性。第四是因子值的方差在很多场合被限定在一个固定值上,即为了测量模型具有唯一的数学解(判定性问题),我们需要假定潜变量的方差为1。但在经验研究中,一个潜变量所表达的社会属性或状态的变化幅度很可能是一个非常重要的研究问题,它要求我们在理论和经验上加以探讨,却在分析方法上被一个计算假定所替代。

由前面的分析可见,从瑟斯通量表,经因子分析到答题论,我们已经很好地解决了量表的构型问题(因子数、因子和指标之间的关系、各指标之间的相对权重),也较好地解决了量表的守恒检验问题(量表对各群体、社区、组织、国家和文化的无偏性)问题,遗留的一个未解决问题是量表的制式问题,即如何确立原点和单位的问题。

在潜变量的测量模型中,原点一般确定为0,它是标准正态分布的均值。在心理学量表中,第一个样本的均值和方差被当作是常模,其他研究以它为参照点,按照标准分的计算公式获得自己的量表分值。但这种计算是以累加量表为原型的,没有吸收因子分析和答题论的研究成果。而且,以标准差为单位的标准分让量表的相同分数差具有不同的概率含义。更重要的是,标准分将一个经验分布投射到了一种理论分布上。

为解决量表的制式问题,我们建议使用序值分析。和标准分相比,序值是一种经验分布,0值(原点)具有明确的理论意义(即平局),分值之差在全值域中具有相同的概率含义。具体计算分为下列四个步骤。

第一,对构成量表的 $j$ 个定序题器进行广义偏信用模型分析,确定量表的单维度性,估计每一个题器的因子负荷(难度系数),作为总分的权重 $w_i$ 。

第二,将量表的所有指标的分布累加成一个总分布,作为序值的标杆分布 $D$ 。一旦发布,它就相当于确定了这个量表的原点和单位,类似于地球仪上的经纬线或地层学中的金钉子。

---

<sup>①</sup> 在很多统计软件中,如Stata和R中,这些阈值被报告为多个切割点(cut)。

第三,按照布罗克特和莱文的序值界定,计算每一个题器的序值  $R_i$ 。

第四,计算量表的加权序值  $R$ ,即:

$$R = \frac{\sum_{i=1}^j w_i R_i}{\sum_{i=1}^j w_i}$$

此加权序值的取值范围为  $[-1, +1]$ ,分数差具有概率解释。此后的其他研究可以依据标杆分布  $D$  和题器权重  $w_i$  计算自己的量表序值分数。即使量表的题器没有完全得到重复,例如少了一两个题器,我们仍可以计算出具有可比性的量表分数。这是因为量表的各题器已经由平行关系转变成串联关系,个案在每一个题器的分段竞赛过程中得到自己的总成绩。整个量表的建构过程就是在所有可能的题器中(题器总体)选择能够反映概念(及其测量上的潜变量)的一组指标,估计它们和潜变量的回归系数,以此作为分段赛胜负的权重系数,然后让有所个案在这组指标上进行竞赛,以确定其总成绩,即加权序值。<sup>①</sup>

目前,把序值分析和量表建构关联起来的研究还不是很多。有学者对二分定序指标变量的序值进行主成分分析,以其回归系数作为权重系数并合成一个新变量,以区分不同个案在一个潜变量上的数值高低(Brockett et al., 2002)。这实际上是基于序值的探索性因子分析,将各定序指标变量因子负荷作为其权重,计算一个总和分数,以反映一个潜变量的水平。有学者将这种方法应用到医疗服务领域,用于判定各医院所提供医疗服务质量的高低,并指出,和其他量表建构的统计分析相比(例如多指标多因模型、潜变量模型、贝叶斯模型等),序值分析的一个明显优势是不需要正确设定测量模型,也就是设定指标和因子之间的因果或相关关系,因此得以避开这个特别容易犯又无法检验的谬误(Lieberthal, 2008)。

和布罗克特的方案相比,量表的加权序值将测量模型建立在答题论而不是古典测量模型上,更具有继承性。另外,将累加分布作为参照组也尽可能避免了单个题器分布对随机误差的敏感度不一造成的影响,因而更加稳健。

① 这里需要和锚定法做区分。锚定法试图解决的方法问题是评估结果在不同个体之间不具有可比性,其解决途径是让不同个体评估“相同”物体或情境,以锚定他们的主观“偏见”,然后假定这种偏见在所有评估中保持恒定,让其评估结果具有可比性。但评估结果可以是定类或定序(常见),也可以是定量(评分)。加权序值分析是在假定定序变量在个体间、群体间(组间)具有可比性的前提下的统计分析方法,解决定序变量分布的集中度量及其统计检验问题。两者没有“内在”关联。

## 五、以加权序值考察国家认同

我们选择国家认同作为加权序值在量表建构中的研究示例,虽然有点任意性,但它的政治重要性和理论争论都值得我们花一点笔墨加以进一步的探讨。当然,对一个量表的方法讨论都离不开对其背后概念的理论探讨,对国家认同的量表考察自然也不会例外。

本文认为,国家认同的背后是国家边界的概念,它是国家作为组织形式的区分在空间、人口、语言和文化的投射所形成的分界线,是国家作为系统的存在条件。在个体层次上,国家边界具有各种表现形式。例如,国家边界的制度形式是国籍,法定的符号表达是身份证或护照,<sup>①</sup>道德形式是公民的权利和义务等规范,经济形式是随国籍而享有的各种社会福利制度、就业机会、投资机会等经济特权,政治形式是因国籍而衍生的各种政治特权,例如集会、结社、选举、出版等(Noiriel, 1996),文化形式是人口普查、地图和图书馆等(Anderson, 1991)。国家认同是国家边界的心理形式,它是个体对国家的组织认同,是对自己处在一个国家边界内外的主观隶属判定,也是对国籍的心理确认。

和社会中的其他社会认同一样,国家边界的制度形式(国籍)和心理形式(国家认同)可能存在不对应的情形。有国籍无认同的个体,我们称作“叛国者”或“卖国贼”;有认同无国籍的个体,我们称作“爱国侨民”;无国籍无认同的个体,我们称作“外国人”;有国籍有认同的个体,我们称作“中国人”。在英语世界里有“香蕉人”之谓,说的是黄种人的外表、白种人的内心,表达的种族边界在体质和心理上的不一致,也是和国家认同相类似的情形。

有研究表明,国家认同一般是在青少年期就已经完成的心理归因过程(Erikson, 1993; Piaget & Smith, 1995)。但在国际移民等特殊人群中,这个心理过程也可以发生在成年阶段,变成一个国家边界的心理重构过程。有时候,国家边界本身发生变化,例如,分裂或统一所涉及的个体也会被动发生国家认同的重新建构。无论发生在人生的哪一个阶段或哪一种情形,在个体的日常生活中,国家认同都是一个自认(自问哪里是我的祖国)和他别(别人区分你是哪国人)两个渠道、长期潜移默化的心理归因过程。在出生就确定国籍的情形下,个体在社会化的生理发展过程中获得默认的国家认同,这是一种先赋的社会属性,个体的

---

<sup>①</sup> 在组织层次上,国家的符号表达是国旗、国歌等。

意愿、情感和态度等并不参与其成员资格的获得过程。在基于审批(例如入党)或交换(例如应聘)或学徒(例如行会)等进入组织边界的情形下,个体的意志、期望和态度等心理因素对社会认同具有非常重要的作用,它们是后致的社会属性,进入各种组织的过程会影响其后期的社会认同。在国际移民的情形下,国籍是一种后致的审批过程,有时还需要以永久居留权作为过渡阶段,这就影响国际移民的国家认同的心理过程。如果再伴以入籍仪式和入籍宣誓,国家认同的强化作用将进一步得到加强。

当然,更常见的、潜移默化的影响机制是正式教育和意识形态的宣传,特别是大众传媒对国别角色的塑造,有学者称之为“同化”或“熔炉”过程(Noiriel, 1996)。在殖民地时代,这种同化过程还伴随有强制拆散原有的文化、宗教、种族、宗族、社区、家庭纽带等压迫行为,以期在儿童的成长期塑造其默认的国家认同。在当代,国家间的冲突、竞赛、合作等跨国界事件会唤醒或强化个体的国家认同。在日常生活中,我们较频繁地需要承担的边界任务是管理互动边界(该不该出席、能不能加入别人的谈话)、群体边界(例如谁是好友、谁是家人)和组织边界(上下班、转岗和下岗、辞职和应聘等)。管理国家边界的心理任务只有在需要出示身份证、升国旗、唱国歌、国庆、办理签证、国际交往、出境旅游等场合才偶尔会浮现出来。因此,在和平时期的日常生活中,个体的国家认同常常处于一种休眠、默认的状态。

这些理论认识对我们的测量模型及其验证具有十分重要的方法论含义。首先,国家认同在概念上是一个三分变量,在边界内即认同,边界外即不认同,模糊时不确定。因此,最直接的测量题器是“你是一个中国人吗?”而选项只有“是”“否”和“不确定”三个。其次,我们需要区分和国家认同相关的一些概念,例如,国家忠诚、国家承诺或奉献、爱国主义、仇外心理、国家沙文主义、国家自豪感等是在国家认同既定的前提下所形成的国别心理现象,不是国家认同的构成成分或结构维度。国籍是一种社会类别,个体经国家间的比较而形成自己的取舍(Tajfel et al., 1971)。再次,国家认同具有很强的政治意涵,在意识形态和社会信任的双重压力下,它的公开自由表达会受到心理审查机制的压制和过滤。因此,“说不清”“不好说”“不确定”以及拒答等情形具有理论和测量上的含义,不仅仅只是信息的缺失。最后,因为国家认同处在一种默认、休眠的状态,社会调查中所获得的应答是一种诱导和压力下的被动表白。在社会调查中,被访者所表达的情绪、态度、行为、属性和状态等是一种社会公共产品,它们是被访者对陌生人在广义社会信任水平的制约下所披露的各种信息。在这个过程中,如果社

会调查具有明显的机构特征,组织信任还会加入进来。如果存在政治审查或政治正确等意识形态因素,制度信任也会加入其中。在有些场合下,访问员的访谈行为也会带来被访者的不信任反应,人际信任也会加入其中。不同于公开态度,私密态度是人们自省时所表现出的情绪或认知倾向。只有在隐私、信任的互动环境中,私密态度才有可能向亲朋好友吐露。简言之,在社会调查的情境下,国家认同作为一种私密态度的被动表白,会遭到一定程度的扭曲,总方向是向正面倾斜。考虑到测量过程,概念上的三分变量才转变成测量上的连续变量。

有研究把国家认同分为认知和情感两个成分(Wright et al., 2012)。但仔细考察其题器内涵,可以发现,所谓国家认同的认知成分实际上是国民形象,一种在特质上对国籍进行描述的刻板印象,例如,美国人的国民形象是信基督、土生土长、遵政守法、平等待人(Wright et al., 2012)。<sup>①</sup>情感成分则是国家自豪感的别称,一种对自己国家的正向价值判断。国民形象和国家自豪感是国家认同形成过程中“他别”途径的重要依据,它们影响个体对他人国家认同的判定,但不是国家认同本身。在个体层次上,国籍确定了政权运行的人口边界,国家认同确立了政权运行的心理边界,国家忠诚关系到政权运行的抗逆性,国家承诺和政权运行的动力性关联。忠诚和承诺以认同为前提,认同和国籍的人口比例是政权通胀的度量。我们在测量上可以借用反映这些概念的题器作为国家认同的替代变量,但不能因为这种借用关系而混淆概念本身的区分。

在美国1996年的综合社会调查(General Social Survey, GSS)中,<sup>②</sup>有一组题器和国家认同有关。有研究利用其中的三个题器建构了一个国家认同量表,并将它与国家自豪感和爱国主义关联在一起。这三个题器分别是:(1)“是一个美国人对您有多重要?”(对应数据变量名为 amimp)0~10分编码;(2)“你感觉美国离您有多近?”(变量名为 clseusa)和(3)“是美国人的感觉对您多重要?”(变量名为 amfeel),四级定序测量加“无法选择”选项(Huddy & Khatib, 2007)。在这个测量模型中,两个题器指向重要性,但重要性和认同的关联来自禀赋效应(Kahneman et al., 1991)或辛劳效应(Norton et al., 2012),前者是指因拥有而高估的心理机制,后者是因自己在获得国籍的过程中付出辛苦而高估的心理机制。

<sup>①</sup> 对美国GSS2016调查数据的广义偏信用模型分析显示,被访者判断真假美国人的标准依重要性排序依次是:美式生活方式(3.05)、生在美国(2.19)、有美国国籍(1.60)、说英语(1.38)、有美国人的感觉(0.83)、信基督(0.85)、尊重美国政体和法律(0.53)。括号中的数值是该题器的区分度。

<sup>②</sup> 调查数据和问卷来源请参见美国芝加哥大学网站(gss.norc.org)美国NORC公开提供该数据,在此致谢。

在该数据集中,有美国国籍的被访者的国家认同水平比无国籍者高 0.91 个标准分,这是存在禀赋效应的经验证据。16 岁之前生活在国外、现在获得美国国籍的被访者的国家认同水平比 16 岁之前生活在国外、调查之前尚无美国国籍的被访者高出 1.06 个标准分,显示禀赋效应和辛劳效应的双重作用。

在另一个调查中,该作者使用另外四个题器来测量国家认同概念,除了 amimp 相同之外,其他三个题器分别是“多大程度上你把自己看作是一个典型的美国人?”<sup>①</sup>“‘美国人’这个词多大程度上能用到你身上?”“在说美国人时,你是否经常说我们而不是他(她)们?”题器选项该文未说明(Huddy & Khatib, 2007)。从我们对国家认同的理解来看,这三个题器具有更高的概念效度,因为它们是国家认同的语言表达。在美国的学生样本中,该量表的标准分和升国旗、唱国歌的感受高度相关(0.74),可作为一个经验证据(Huddy & Khatib, 2007)。

国内有研究采用五个题器来测量国家认同,分别是“当别人批评中国人的时候,我觉得像在批评自己”,“我经常因国家现存的一些问题而感到丢脸”,“我经常为国家取得的成就而感到自豪”,“如果有下辈子,我还是愿意做中国人”和“不管中国发生什么事情,即使有机会离开,我也会留在中国”。选项是“很不符合”“不太符合”“比较符合”和“很符合”,另有一个“不好说”选项(李春玲、刘森林,2018)。前三个题器实际上是国家自豪感的指标变量,后两个变量是国家忠诚的指标变量,但对应题器的表述存在较大问题。国家认同和国家自豪感的联系途径是群体移情,是被访者站在全体国人的立场上来体验他们的情绪反应,是一种爱屋及乌的情感机制。国家忠诚是国家认同的心理内化,从群体相认上升到自我和国家在心理上的高度一体。这个概念分析具有测量上的证据,即前三题题器的因子负荷相近,后两个题器的因子负荷相近。<sup>②</sup> 具体来说,在验证性因子分析模型中,顺序编码的 1~5(把“不好说”处理为中间状态,编码为 3)的五个指标变量的因子负荷分别是 1.00、0.66、1.05、1.66 和 1.77;在定序验证性因子分析模型中,因子负荷分别是 1.00、0.58、1.50、2.95 和 2.15;在广义偏信用模型中,难度系数分别是 0.61、0.27、1.31、2.81 和 2.26。三个测量分析模型的参数指向同一个分析结论,在概念效度上,这五个题器不是一个概念的指标。

然而,上述结论并没有考虑对“不好说”的处理。如前所述,在国家认同的测量过程中,拒答和“不好说”不是中性的表达,而是政治敏感度或题器理解困难的表达。在广义偏信用模型中,这表现在题器响应函数上。当把“不好说”编

① 去掉“典型的”更佳。

② 数据来源于中国社会科学院社会学研究所“中国社会状况综合调查”公开数据,在此致谢。

码为中间项时,第一个题器的响应参数分别是-3.55、0.46、-3.94和1.31,没有显现出递减或递增的顺序性;当把“不好说”编码为缺失值时,第一个题器的响应参数分别是-2.66、-2.10和0.92,显示出递增性。其余四个题器呈现相同的变动模式,这说明,我们对“不好说”的理解更具有测量模型分析上的证据。

在讨论了国家认同量表的建构效度之后,经典测量模型和答题论的结论是我们无法采用这个数据来分析国家认同问题,因为量表无效。但序值分析把测量过程看作是一群人在一组题器上的竞赛过程,每一个题器都是一个竞赛项目,只要这些项目具有内在关联性,它们的加权总成绩就会反映这种内在关联。前面的分析表明,这五个题器分别是国家自豪感和国家忠诚的测量指标,它们都与国家认同相关联。因此,我们可以用序值来分析它们,讨论国家认同问题,只不过结论不是肯定性的,而是提示性的,还是因为量表构成不精确。<sup>①</sup>

结合前面的分析,我们将这个国家认同量表的测量层次定位在四级定序测量上。按照前面的计算步骤,我们通过广义偏信用模型,获得这五个题器的权重 $w_i$ ,分别是0.961、0.519、1.809、3.744和2.608。然后,我们计算这五个题器的总分布,从“很不符合”到“很符合”的四级定序分布依次是957、4915、23882和18316,它们构成国家认同的标杆分布。高于标杆分布得正分,低于标杆分布得负分。需要注意的是,和标准正态分布相比,这是一个右偏、非对称的离散分布,表示在和平时代、社会调查的语境下中国人对国家认同五个题器的应答模式。在题器权重和标杆分布的基础上,我们就可以计算每一个题器的序值以及量表的加权序值,对2013年参与调查的8684个被访者来说,他们的序值均值是0.08,表示有8.1%的个体对国家认同高于标杆分布,95%的置信区间是[0.073,0.090],这是一个左长尾的偏态分布(偏度为-0.163),比标准正态分布更高尖(峭度为1.817)。下面,我们以受教育程度为例来比较简单加总、标准因子分和加权序值的分析结果,如表3所示。

由表3可见,这三种计算方式在统计检验上所得到的结论相同,95%的置信区间显示,受过高等教育和没有受过高等教育的人群在国家认同上具有显著差异,前者低于后者。但在具体数值的阐释上,不同计算方式给人的印象不同。简单加总的分值最大为10,最小为-10,似乎暗示国家认同有很大上升空间;因子分以0为均值,似乎暗示没有受过高等教育人群的国家认同接近总体,没有很大上升空间。加权序值以标杆分布为参照点,它的数值说明,没有受过高等教育人

<sup>①</sup> 如果竞赛项目(问卷题器)完全异质,加权序值也不能反映任何概念。换言之,分析方法的改进不能弥补测量模型概念效度的不足。

表 3 受教育程度和国家认同的关系

	简单加总	验证性因子分	加权序值
中专及以下	5.365 [5.301, 5.429]	0.013 [0.007, 0.018]	0.105 [0.096, 0.114]
本科及以上	4.627 [4.439, 4.815]	-0.108 [-0.125, -0.091]	-0.061 [-0.084, -0.037]
个案数	10168	10189	8669

注:[ ]中的数值是 95% 置信区间的上下限。

群的国家认同有 10.5% 的概率高于标杆,受过高等教育人群的国家认同有 6.1% 的概率低于标杆,而标杆是和平时期、处于高速发展阶段的国家认同分布,只有相对值,没有绝对值。更重要的是,加权序值分析可用于比较不同时期、不同国别、不同测量题器的国家认同。以 2013 年的中国调查和 1996 年的美国调查为例,我们得到表 4 的结果。

表 4 国家认同的国别比较

	序值均值	95% 置信区间	偏度	峭度	个案数
2013 年中国调查	0.081	[0.073, 0.090]	-0.163	1.817	8684
1996 年美国调查	0.087	[0.063, 0.112]	-0.492	2.333	1277

美国 1996 年调查序值采用相同的标杆分布  $D$ , 它的国家认同量表有三个题器,相近题器 (clseusa)、美国人感 (amfeel)、重要性题器 (amimp) 的权重系数分别是 0.755、0.986 和 0.672,因为重要性题器是 0~10 的计分方式,我们仅采用前两个定序题器计算该量表的加权序值。由表 4 的结果可见,在这两次调查中,中美被访者的国家认同没有统计上显著的差异,但相较而言,美国被访者的国家认同分布变异度更大、更左偏、更高尖,显示出更高的异质性。

因为拥有了可比的加权序值,美国调查数据中的其他变量可以弥补中国调查中的不足。例如,没有国籍的美国被访者 (citizen) 的国家认同的加权序值均值是 -0.166,即有 16.6% 的人国家认同低于标杆分布  $D$ 。这既显示出国籍的效应,也反映了美国意识形态运作的效率。在“愿意成为美国公民” (amcitizn) 这个变量上,被访者从“很不愿意”到“很愿意”的五级定序测量上国家认同的加权序值均值从 -0.612 逐步上升到 0.209,有 80% 以上的概率会提升。若合并“愿意”和“很愿意”两类,其余为一类,那么,拥有国籍的被访美国人中只有 2% 左右不愿意成为美国公民,没有国籍的被访者中却有 77% 左右的人愿意成为美国公民。该交互表从另一个侧面佐证了前面有美国籍和意识形态效应的结论。

在解释国籍效应时,我们还需要注意到,因为大多数经验研究所采用的调查

数据是在同一个国境内进行的,国籍效应严重衰减。如果我们在美国的调查问卷中添加对中国的国家认同量表,国籍对中国国家认同的效应才会较全面地显现出来。当然,最完整的显现是在全世界人口的代表性调查中询问被访者的中国国家认同。正是因为存在国家边界对国家认同的衰减效应,我们在使用回归方程分析国别调查数据时才会出现特别低的模型决定系数,有时候甚至低于0.05(李春玲、刘森林,2018)。这表示我们仅在关注细枝末节的因素,主因仍在分析的视野之外。

## 六、讨 论

序值分析是在布罗克特和莱文所阐明的四个约束(Brockett & Levine, 1977)的基础上,将定序尺度映射到概率尺度上,但概率转换的依据来自定序变量本身的经验分布;而logit和probit则是将定序尺度投射到对数尺度和正态分布的尺度上,即将一种定序测量从一种经验分布投射到一种统计上的理论分布。换言之,序值分析是一种基于经验分布的转换,logit和probit是一种基于统计学理论分布的转换。

当然,从数学上看,我们也可以把序值分析看作解决定序变量赋值难题的众多方案之一,序值比整数赋值更复杂,有时分析的结果也没有明显差异(Fielding, 1993)(例如本文表3中的结果)。在二分变量和两组比较的情形下,序值分析、非参数检验和均值比较还具有等价性(Vigderhous, 1979)。但从应用的角度看,序值分析让测量制式扎根于社会测量过程中,使得数据的大小具有行动者的意义锚定,即一组个案在一个或多个定序指标上的竞赛过程中胜出其他个案组的概率(赢面大小)。这样,我们就避免了犯错误设定(指标和因子之间关系设定错误)和错误假定(错误猜测经验分布背后的统计分布)两种谬误,让定量分析具有更可靠的行动者意义基础,让观察者和分析者的偏见更有可能代入到分析过程和结果中。

同时,在描述统计中,序值分析是变量分布的集中指标,辅以相应的方差计算,我们就可以将序值分析建立在严格的统计推理上,从而避免了定量分析中的主观随意性。

当然,我们也不能只要看见定序变量就盲目应用序值分析。如果定序变量的类别赋值具有理论和经验含义,就没有必要进行序值分析或加权序值分析

(Mantel, 1979)。另外,序值分析的前提是定序变量的取值在个体间具有可比性,对于不具有个体间可比性的定序变量,例如各种以定序变量形式出现的主观评估,在应用序值分析时应格外谨慎。当序值分析应用于测量模型时,测量模型的概念效度是加权序值分析的前提,虽然在本文中,我们在经验示例中展示了加权序值分析的可扩展性,但从概念到指标、从经验到数据的理论依据仍是定量分析科学性的基础。有研究先把定序因变量进行序值转换,然后再做多重线性回归,这种做法仍有许多数学问题需要厘清,例如回归方程两边的值域并不相等。因此,我们在本文中并没有做相应的介绍和推荐。最后,在使用各种平台上序值分析的相关软件时,我们必须对其序值界定和所采用的方差或标准差计算公式给予特别的关注,它们可能直接影响统计推断的结论。

在本文选择的两个关于主观认同研究的经验案例中,阶层主观认同的十层测量模型是科学研究的人造物,在社会中没有对应的符号边界,社会学的观察者有必要对自己的“造物”保持必要的批判力。国家认同是国家边界和心理折射物,在社会中具有相应的物理和符号边界,社会学的观察者也不能假“想象的共同体”(Anderson, 1991)之名忽略它们。在这个意义上,社会学家和现代巫师之间的角色界线并非泾渭分明。如果我们的测量工具仅是龟板上的随机纹路,我们的测量制式又来自统计学天堂中的一个理论分布时,我们在这两个角色之间的跃迁就是一个大概率事件。

提高量表的效度需要提高我们的理论水平,扩展我们的经验视野,优化我们的知识结构,提高我们的理论和方法觉悟,这需要较长时间的学术积累和更多的探索研究。改变我们的量表制式则相对较容易,本文以经验分布作为制式的基础,探索以序值作为定序指标量表制式的可能性,以便让量表分值具有理论和经验基础,让各时点、各国别的社会调查数据具有制式上的可比性。当然,目前的解决方案也有自己的缺点,例如,以序值作为量表的制式要求一个量表的不同指标变量具有相同数量的定序类别。更重要的是,序值制式是以经验现象为参照的。例如,在测量温度时,摄氏度以水在一个大气压结冰的温度为0度,以水沸腾为100度,这也是一种经验参照;但摄氏度没有认识到温度的理论意义,即温度是分子的动量水平的测度,因而不可能有绝对零度的概念。每一个量表的加权序值背后的理论概念是什么?或者说,每一种标杆分布D代表社会处于什么状态?这些问题还需要我们进一步拓展研究思路,寻找办法来克服和解释它们,但第一步的学术工作是正视量表的制式问题。

## 参考文献:

- 陈光金,2013,《不仅有“相对剥夺”,还有“生存焦虑”——中国主观认同阶层分布十年变迁的实证分析(2001—2011)》,《黑龙江社会科学》第5期。
- 李春玲、刘森林,2018,《国家认同的影响因素及其代际特征差异——基于2013年中国社会状况调查数据》,《中国社会科学》第4期。
- 夏传玲,2021,《建构社会分层结构新潜类模型》,《中国社会科学报》12月8日。
- 张翼,2011,《中国社会阶层结构变动趋势研究——基于全国性CGSS调查数据的分析》,《中国特色社会主义研究》第3期。
- Anderson, Benedict 1991, *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso.
- Beder, Jay H. & Robert C. Heim 1990, “On the Use of Redit Analysis.” *Psychometrika* 55(4).
- Borgatta, Edgar F. 1968, “My Student, the Purist: A Lament.” *The Sociological Quarterly* 9(1).
- Brockett, Patrick L. & Arnold Levine 1977, “On a Characterization of Ridits.” *The Annals of Statistics* 5(6).
- Brockett, Patrick L., Richard A. Derrig, Linda L. Golden, Arnold Levine & Mark Alpert 2002, “Fraud Classification Using Principal Component Analysis of Ridits.” *The Journal of Risk and Insurance* 69(3).
- Bross, Irwin D. J. 1958, “How to Use Redit Analysis.” *Biometrics* 14(1).
- Conover, W. J. 1973, “Rank Tests for One Sample, Two Samples, and K Samples without the Assumption of a Continuous Distribution Function.” *The Annals of Statistics* 1(6).
- Epps, T. W. & Kenneth J. Singleton 1986, “An Omnibus Test for the Two-Sample Problem Using the Empirical Characteristic Function.” *Journal of Statistical Computation and Simulation* 26(3).
- Erikson, Erik H. 1993, *Childhood and Society*. New York: Norton.
- Fielding, A. 1993, “Scoring Functions for Ordered Classifications in Statistical Analysis.” *Quality and Quantity* 27(1).
- Flora, Jairus D. 1974, “A Note on Redit Analysis.” (<https://deepblue.lib.umich.edu/bitstream/handle/2027.42/1530/30927.0001.001.pdf>)
- 2006, “Redit Analysis.” In Samuel Kotz (ed.), *Encyclopedia of Statistical Sciences*. New Jersey: John Wiley & Sons.
- Guttman, Louis 1944, “A Basis for Scaling Qualitative Data.” *American Sociological Review* 9(2).
- Huddy, Leonie & Nadia Khatib 2007, “American Patriotism, National Identity, and Political Involvement.” *American Journal of Political Science* 51(1).
- Jansen, Marlies E. 1984, “Redit Analysis, a Review.” *Statistica Neerlandica* 38(3).
- Kahneman, Daniel, Jack L. Knetsch & Richard H. Thaler 1991, “Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias.” *The Journal of Economic Perspectives* 5(1).
- Lieberthal, Robert D. 2008, “Hospital Quality: A Predit Approach.” *Health Services Research* 43(3).
- Likert, R. 1932, “A Technique for the Measurement of Attitudes.” *Archives of Psychology* 22(140).
- Lukacs, Eugene 1970, *Characteristic Functions* (second edition). London: Griffin.
- Mantel, Nathan 1979, “Redit Analysis and Related Ranking Procedures—Use at Your Own Risk.” *American Journal of Epidemiology* 109(1).

- Marfak, Abdelghafour, Ibtissam Marfak, El Achhab Youness & Elmadani Saad 2020, "Improved Ridit Statistic Approach Provides More Intuitive and Informative Interpretation of EQ-5D Data." *Health and Quality of Life Outcomes* 18(1).
- Masters, Geoff N. 1982, "A Rasch Model for Partial Credit Scoring." *Psychometrika* 47(2).
- Mielke, Jr., P. W., Michael A. Long, Kenneth J. Berry & J. E. Johnston 2009, "G-Treatment Ridit Analyses; Resampling Permutation Methods." *Statistical Methodology* 6(3).
- Muthén, Bengt 1984, "A General Structural Equation Model with Dichotomous, Ordered Categorical, and Continuous Latent Variable Indicators." *Psychometrika* 49.
- Noiriel, Gérard 1996, *The French Melting Pot: Immigration, Citizenship, and National Identity*. Minneapolis: University of Minnesota Press.
- Norton, Michael I., Daniel Mochon & Dan Ariely 2012, "The Ikea Effect: When Labor Leads to Love." *Journal of Consumer Psychology* 22(3).
- Piaget, Jean & Leslie Smith 1995, *Sociological Studies*. London and New York: Routledge.
- Tajfel, Henri, Michael Billig, Robert P. Bundy & Claude Flament 1971, "Social Categorization and Intergroup Behaviour." *European Journal of Social Psychology* 1(2).
- Thorndike, Robert M. 2005, "History of Factor Analysis: A Psychological Perspective." In Brian S. Everitt & David C. Howell (eds.), *Encyclopedia of Statistics in Behavioral Science*. Chichester: John Wiley & Sons.
- Thurstone, L. L. & Ernest J. Chave 1929, *The Measurement of Attitude: A Psychophysical Method and Some Experiments with a Scale for Measuring Attitude toward the Church*. Chicago, Ill.,: The University of Chicago Press.
- Uwawunkonye, Ebu & Oyeka I. C. Anaene 2013, "A Comparative Study between Ridit and Modified Ridit Analysis." *American Journal of Theoretical and Applied Statistics* 2(6).
- Vigderhous, G. 1979, "Equivalence between Ordinal Measures of Association and Tests of Significant Differences between Samples." *Quality and Quantity* 13(3).
- Wright, Matthew, Jack Citrin & Jonathan Wand 2012, "Alternative Measures of American National Identity: Implications for the Civic-Ethnic Distinction." *Political Psychology* 33(4).

作者单位:中国社会科学院大学社会与民族学学院、  
中国社会科学院社会学研究所、  
中国社会科学院社会学研究所社会调查与数据处理研究中心  
责任编辑:刘保中