

高勇.有关概念测量的三个故事与一个教训[M/OL]//赵联飞,赵锋.社会研究方法评论:第2卷.重庆:重庆大学电子音像出版社.

有关概念测量的三个故事与一个教训

高勇^①

摘要:在定量研究中,“测量”是一个看似简单实则复杂的环节。研究过程中某些草率的决策,往往导致研究者对于经验事实产生错误的判断。本文简要地介绍了三个研究实例,分别涉及到“社会封闭”“地区收入不平等”“人类发展指数”等概念的测量。在这些实例中,测量结果的真实意蕴往往偏离了研究者原本的测量意图,出现了某种“非预期后果”。本文认为,测量本身就是一种具体的社会过程,不能只从抽象概念及概念维度的操作化来理解。只有把测量还原为具体的社会过程时,我们才能透彻地理解测量。

关键词:测量;社会过程;概念;操作化;社会闭合

Abstract: In quantitative research, “measurement” is a seemingly simple but complex process. Some sloppy decisions in the research process often lead to misleading conclusions about empirical facts. This paper briefly introduces three research examples, which involve the mea-

①作者简介:高勇,中国政法大学社会学院教授、博士生导师。中国社会学会社会分层与流动专业委员会理事,北京社会学会理事。联系方式:gaoyongvincent@163.com。

surement of concepts such as “social closure”, “regional income inequality” and “human development index”. In these examples, the true meaning of the measurement results deviates from the researcher’s original measurement intentions, with some “unintended consequence”. This article argues that measurement itself is a concrete social process that cannot be understood only from the perspective of abstract concepts and operationalization. It is only when measurements are considered as concrete social processes that we can understand measurements thoroughly.

Keyword: Measurement; Social Process; Concept; Operationalization; Social Closure

(社会科学的实际工作)不是“实验室”，而是“工作室”，是手工艺人或15世纪文艺复兴时期的画家们的那些工作室，也就是说它有着各种各样的错误，起步不当、摇摆不定、陷入困境、推倒重来，诸如此类，不一而足。研究者们尽管各自工作的进展程度不一，但都会把自己力图建构的对象摆出来，接受其他所有人的质疑与批评。这些人的行动方式，用传统行话来说，就像这行当里的老同事、老伙计，他们会各尽所能，把他们通过过去所有的尝试和失败教训而逐渐积累起来的集体性经验贡献出来。

——布尔迪厄，《反思社会学导引》

在几乎所有的社会研究方法教材中，测量都是非常重要的章节。它是“假设—演绎”研究范式中的重要一环，是连接理论层面与经验层面的关键节点。在定量研究中，“测量”是一个看似简单实则复杂的环节。研究过程中某些草率的决策，往往导致研究者

对于经验事实产生错误的判断。本文试图通过几个研究实例的讨论,说明测量环节可能存在的研究陷阱,指出认真考察测量过程的必要性。

一、“社会闭合”

(一)概念的提出

科尔曼(Coleman,1988)在《人力资本生成中的社会资本》一文中提出了他的“社会资本”概念:它就是某种社会结构,有着便利一些行动者采取某种行动的功能后果。^①缺了这种特定的社会结构,那些行动者的目标就没法达成。例如,某个行业是由内部有亲缘或地缘关系的小圈子构成的,这种特定的社会结构便利了他们做这一行业的生意,这就构成了一种社会资本;某个城市中有着邻里之间彼此照看的规范,这种特定的社会结构便利了他们可以让小孩自由在街上玩耍而不必费心监管,这也就构成了一种社会资本。

与以往的“非正式制度”“正式制度”“信息渠道”等概念相比,“社会资本”增加了哪些理论要素?第一,所有的“非正式制度”“正式制度”“信息渠道”都需要落实到某种实在的社会结构上,如有彼此互惠义务的群体、有共享规范和规则的组织、互通信息的网络圈子等等。第二,如同资本具有一定程度的“通用性”一样,原本用于某种目标而

^①不过可以想见,这种社会结构同时也一定阻碍了另一些行动者采取另一些行动。在这种意义上,特定社会结构对一些人采取一些社会行动是“资本”,对另一些人采取另一些社会行动来说就是“妨碍”。有利于父母管教孩子的一种社会结构,对父母是“社会资本”,但对那些顽皮孩子来说,无疑是一种“社会妨碍”。科尔曼本人也明确地意识到这一点,但并未展开讨论(Coleman,1988:S105)。

形成的社会结构,随后完全可以成为服务于另一目标的手段,它同样具有“通用性”。正是在这个意义上,那些社会结构才具有“资本”的特性。

人际关系网络就是这样一种结构。不过,它要有“通用性”,就必须具有“社会闭合”(social closure)的特点。你与工作同事有关系网络,你的孩子与同学有关系网络,但是如果你的孩子的同学正好是你工作同事的孩子,那么关系网络就“闭合”了。此时,原来因工作建立起来的关系网络就可能被摊开来监督和引导孩子,它便利了你的这一目标。这种人际关系网络有了通用性,因此它构成了父母的一种社会资本。在这样的社会结构中,人们之间的联结不是单一性的,而是多重联结(a multiplex relation);在某一背景下形成的联结可以非常方便地摊用到另一背景当中。现代社会中,这种多重联结的人际关系网络对于社会的运作发挥着至关重要的作用。

与“社会资本”概念相比,附属概念“社会闭合”有具体的指向,功能论的色彩也少一些。它强调了人际关系中多重联结的结构性作用,为社会结构研究增加了新的维度。某些研究者甚至认为,“社会闭合”这一概念更具有原创性(Carbonaro, 1998)。

(二)经验测量与证据

科尔曼的论文尽管有数据佐证,但是仍然重在提出思想,而不是经验测量。他以往一直在从事教育领域的研究,因此就顺带把这一思想放在教育领域中来阐释。由于手头的数据中并没有对于“社会闭合”的直接测量,科尔曼使用了“由于搬家而导致的转校次数”来作为对这一概念的近似指标。这一做法不能令人信服,有严重的“贴标签”嫌疑。另一个证据是宗教性私立学校的辍学率远低于公立学校。

宗教性私立学校中,学生父母的人际关系网络可能更具有多重联结的特点:既是教友,又是好友的父母。但这一证据也太于单薄,因为我们可以想象出很多备择假设来解释同一事实。

有可能受限于科尔曼当初的例证,有关“社会闭合”的后续研究基本都是在教育领域进行的,虽然这一概念本身有更深远的意义。在1988年美国教育纵贯研究(NELS:88)的第二轮追踪调查中^①,研究者特意对这一概念进行了操作化,设计了如下题项来测量“代际关系闭合”:

53. 你能说出你孩子的一个好朋友的名字(或外号)吗?

能1(继续回答54题)

不能2(跳答56题)

54. 请列出你孩子的好朋友的名字(或外号),然后说明:

(A)他(她)和你孩子是否在同一所学校上学?

(B)你是否认识他(她)的家长?

好朋友的名字	A 同一所学校上学			B 认识家长	
	是	否	不适用	是	否
1.	1.....	2.....	3	1.....	2
2.	1.....	2.....	3	1.....	2
3.	1.....	2.....	3	1.....	2
4.	1.....	2.....	3	1.....	2
5.	1.....	2.....	3	1.....	2

上述54题B中回答“是”的数目,就可以作为“社会闭合”的一种

^①NELS的基线调查是1988年实施的,第二轮追踪调查是在1992年实施的。读者现在仍然可以从以下网址下载到当年的问卷:<https://nces.ed.gov/surveys/nels88/questionnaires.asp>。

测量指标。它代表了家长认识多少位别的家长,这和科尔曼当初举的例子是吻合的。从概念操作化的角度来看,这似乎没有任何问题。基于 NELS 数据,研究者对于科尔曼的理论假说进行了经验检验(Carbonaro, 1998; Morgan、Sørensen, 1999)。卡博纳罗发现,在个体层面社会闭合对于学生的数学成绩的提升确实有正面影响。摩尔根和索伦森的研究使用多水平的随机效应模型,发现在学校层面上“社会闭合”对学习成绩基本上没有什么正向作用。

(三)我们究竟测量的是什么?

然而,这里面有一个最基本的问题需要澄清:我们测量出来的,究竟是什么?霍里南和库比切克(Hallinan & Kubitcheck, 1999)指出,这一变量很可能与“社会闭合”关系不大。

让我们回到事实层面,来仔细考察一下测量过程是怎么发生的:首先,被访者要列出他们孩子的朋友名字;然后,我们才继续提问有关这些朋友的进一步信息。第一步给后面的步骤定了上限:如果你只能列出孩子的3个朋友,那么就算你的社会关系网络都是闭合的,那你在后面的步骤中也最多只能说自己的3个朋友的父母;如果你能列出孩子的5个朋友,那么就算你的社会关系网络并不太闭合,你在后面的步骤中也有很大概率说出自己的3个朋友的父母。

变量的大部分变异,究竟是来自“知道的小孩朋友数”,还是来自“社会关系闭合度”?数据表明,前者更符合事实。列出小孩名字的个数,与列出父母名字的个数是强相关,相关系数接近0.75。家长平均能够说出孩子的4个好朋友的名字,平均能够说出孩子的3.3个好朋友的父母名字。基本上来说,知道自家小孩的朋友名字,一般就能

够认识他们的父母。

于是,我们了解到的信息,与其说是“社会关系闭合度”,不如说是“孩子的人缘”(Hallinan & Kubitcheck, 1999)。我们自以为看到的是,社会关系闭合度高的小孩成绩更好。但事实可能是,成绩好的孩子人缘就比较好,因此父母知道的朋友数也越多;这就同社会闭合度没有关系。或者,我们自以为看到的是,社会关系闭合度与成绩没有关系;事实上我们证明的只是,孩子的人缘和成绩没有关系。

还有另一种可能:我们了解到的信息,与其说是“社会关系闭合度”,不如说是拉鲁(2018)在《不平等的童年》中指出的“协作培养方式”(concerted cultivation)。父母越是与孩子多交流,他就越了解孩子的生活细节,能说出更多的孩子的朋友名字。当然,这种培养方式下,孩子的学业成绩也往往更高。无论如何,这和社会闭合度也没有什么关系。^①

在这里,测量出现了问题。然而,如果简单沿用教科书式的“假设—演绎”模型,我们并不太容易发现这里的问题。概念与测量之间的脱节,比我们以为的更容易发生(Martin, 2017:38–40)。

①在经历了这些学术争论之后,2002年和2004年美国的ELS(Education Longitudinal Study)调查重新设计了社会网络模块的提问方式。利用新的数据,摩尔根和陶德(Morgan & Todd, 2009)再次对“社会闭合”效应进行了检验。结果发现,在控制了家庭背景、学生交往网络等因素后,社会闭合在公立学校中对成绩并没有明显促进作用,但在天主教学校中对成绩有一定促进作用。

二、“收入不平等”与幸福感

(一)理论命题的提出

发展经济学家赫希曼(Hirschman & Rothschild, 1973)在《经济发展过程中对收入不平等容忍程度的变化》(*The Changing Tolerance for Income Inequality in the Course of Economic Development*)一文中提出了著名的“隧道效应”。人们对于不平等的容忍程度是随着发展的不同阶段而变化的。在发展的早期阶段,收入差距逐渐拉开,但是人们对于不平等具有相当的容忍程度。因为人们看到身边的高收入者时,会感觉到这提供了生活改善的希望与前景。如同处在交通堵塞的隧道中的司机,他在看到旁边车道开始移动时,即使自己还没有移动也会充满希望一样。收入差距的适度增大,增强了人们对于未来生活福祉提升的预期,从而提升了他/她的幸福感。但是在早期阶段过后,人们对不平等的理解方式就改变了。如同处在交通堵塞隧道中的司机,在看到旁边车道一直移动,自己却迟迟不能起步时一样会充满抱怨。

赫希曼的这一理论命题有着非常重要的社会学意味。他表明,人们对收入不平等的容忍程度,并不是一种稳定的个体偏好,而是受到对外在社会情境的“定义”的影响。我们可以把它看成是社会学中“情境定义”命题在宏观经济发展中的应用。“情境定义”并不只是一个在微观人际互动中适用的命题,它完全可以推广到我们对宏观社会变迁的理解中。

(二)经验测量与证据

在中国高速发展的背景下,有很多研究者使用调查数据来讨论

收入不平等程度对于人们心态观念的影响。利用全国综合抽样调查数据(如CGSS)计算得出区县(甚至更小单位的)基尼系数(或泰尔指数等相似指标),然后引入回归模型进行分析,这是一种讨论收入不平等对其他因素影响的常见做法。研究者并没有得到一致的答案:有的发现地区基尼系数对幸福感有正向作用,有的发现有负向作用,还有的发现两者存在一种“倒U性曲线关系”。

我们用CGSS2013数据来看一下区县基尼系数与各区县幸福感均值的关系。初步的证据(图1)表明,两者之间似乎存在着一种正向关系:区县基尼系数越大,幸福感的评价值越高。这似乎也可以印证赫希曼的理论命题。为了直观起见,我们在此只用散点图来展示双变量关系,但是在更复杂的统计模型分析中,收入基尼系数对幸福感的效应也非常显著。

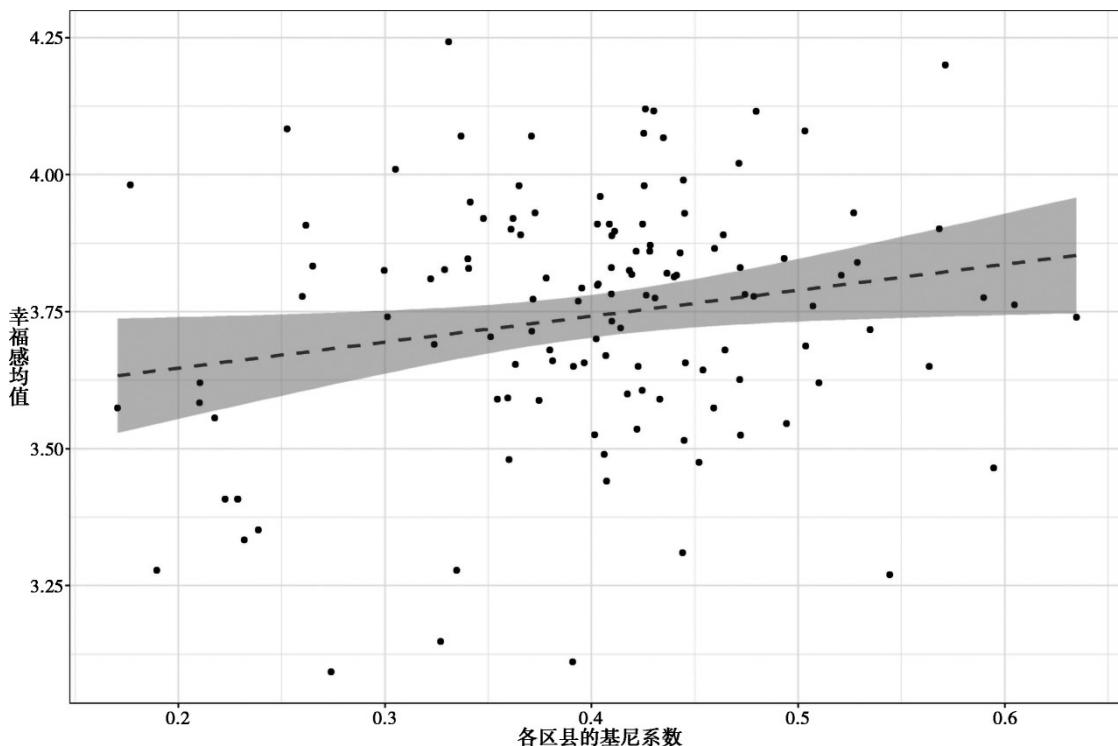


图1 区县基尼系数与幸福感均值

(三)我们究竟测量的是什么?

但是,我们对于基尼系数的测量似乎有些不太对头。让我们把图1中最左端和最右端那些散点所代表的区县所在省份标注出来,由此得到图2。为什么那些基尼系数最低的区县都来自于北京、天津、广东?为什么那些基尼系数最高的区县却来自于云南、福建、安徽、山东?这既和我们的日常经验不相符合,也和赫希曼的理论推断不相符合。

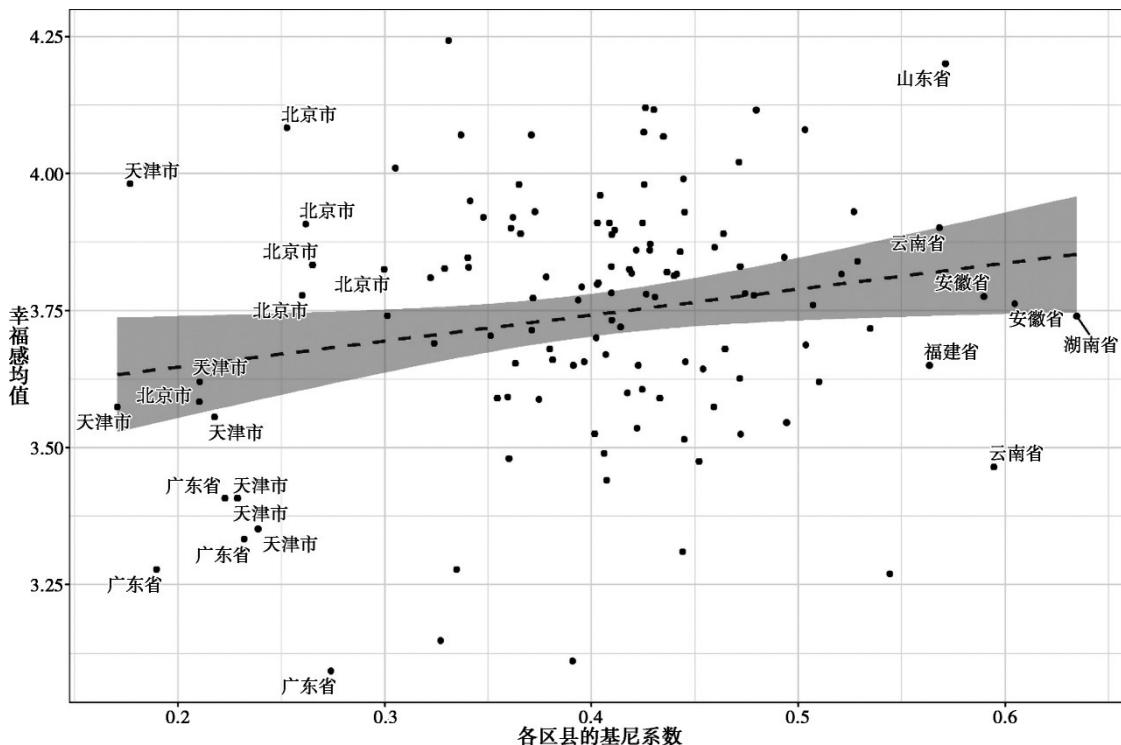


图2 奇怪的分布

原因出在了我们对基尼系数的测量方式上,而这又和CGSS调查的抽样方式有关。CGSS调查的抽样有必选层与抽选层的设计。^①北

^①更多细节请参阅《中国综合社会调查(CGSS)第二期(2010-2019)抽样方案》(中国综合社会调查网,<http://cgss.ruc.edu.cn/xmwd/cysj.htm>)。

京、上海、天津、广州、深圳五个大城市是必选层,必选层的初级抽样单元是街道:首先抽取 40 个街道,每个街道中抽取 2 个居委会,每个居委会抽取 25 个家庭户。因此,大城市的样本全部来自居委会,多数区县仅抽到 2 个居委会,这 2 个居委会一定在同一个街道。此外,由于城市房产价格的原因,大都市地区居住于同一居委会中的居民家庭收入往往较为接近。基于上述原因,CGSS 必选层区县内的样本具有很高的同质性,因此测算出来的收入基尼系数一定较小。

其余全国所有家庭户构成抽选层,抽选层的初级抽样单元是区县:首先抽取 100 个区县,每个区县中抽取 4 个居委会或村委会,每个居委会或村委会抽取 25 个家庭户。在这样区县样本中,既有居委会也有村委会(根据城市化水平进行配比),所有区县都被抽中了 4 个村委会或居委会,这 4 个村委会或居委会未必在同一街道或乡镇。此外,在农村地区,隶属于同一村委会的村民因收入提升而搬离本村庄的可能性固然存在,但相对城市而言较小。基于上述原因,CGSS 抽选层区县内样本异质性程度更高,因此测算出来的收入基尼系数一定更大。

这样一种抽样方式导致了一个奇特的“非预期后果”:测算出来的区县收入基尼系数与区县城市化水平有很强的相关性。城市化水平越高,测算出来的区县收入基尼系数越低。在这种情况下,我们测量的与其说是区县的“收入不平等程度”,不如说是区县的“城市化水平”。我们以为看到的是,收入不平等程度越高,主观幸福感越高;而我们真正看到的有可能只是,大城市中的居民的主观幸福感较低。事实上,在区分了必选层和抽选层之后,回归模型中收入基尼系数对

幸福感的效应就大大衰减,失去了显著性。^①

以“基尼系数”的公式测算出来的数字,并不一定测量的就是“收入不平等程度”。数据的抽样方式以隐蔽的方式扭曲了我们的测量。

三、“人类发展指数”

(一)新的发展理念

早期对于发展的看法,将它等同于经济增长,同于财富增长。阿玛蒂亚·森(2002)在《以自由看待发展》中提出了不同的理念:“发展可以视为一种扩张人们享有的真实自由的过程。它的焦点是人类的自由,并与发展的狭隘观点形成强烈对比。这些狭隘的观点,像是把国民生产总值的成长、个人所得的增加、工业化、技术进步或社会现代化,看成是发展。国民生产总值或个人所得的增加,对于扩张社会成员的自由而言,当然可以是很重要的工具。但是自由也取决于其他因素,譬如社会和经济的安排以及政治和公民的权利。”这种发展观也被称为发展的“可行能力”视角。

联合国开发计划署(UNDP)编写的《1990年人类发展报告》中将这种理念明确化,正式提出了“人类发展”的概念,以与以前的发展观相区分:“人是国家真正的财富。发展的基本目标是创造有利环境,使人类长寿、健康和富有创造力……人类发展是一个不断扩大人们选择权的过程。最关键的选择权包括拥有健康长寿的生命、受教育和享受高品质生活的权利。其他选择权包括政治自由、有保障的人

^①有兴趣的读者可以在如下网址 <http://gao-yong.cn/> 查看更详尽的细节。

权与自尊……人类发展将商品生产与商品分配、拓展人类能力与利用人类能力结合起来。人类发展也关注选择权——关注人们为了确保生计应该有什么,是什么,做什么。此外,人类发展不仅与满足人类基本需要有关,同时也是一个动态参与过程。无论是在欠发达国家还是在发达国家,都是如此。”

(二)新的测量方式

在提倡新理念的同时,联合国开发计划署(UNDP)也提出了测量发展的新指标:人类发展指数。人类发展指数是三个分项指数的综合:健康指数(人均预期寿命)、教育指数(文盲率与平均受教育年限)、人均收入指数(以平价购买力计算的人均收入)。它对应于发展的三种基础性能力:长寿而健康的生命、获得知识的渠道、有尊严的生活水准。三个分项指数进行标准化之后的几何平均数,就是最终的人类发展指数。其中的技术细节,可以参看UNDP的说明。人类发展指数的推出,对于改变国际社会中的发展理念,更准确地测度某一国家的发展水平,起到了积极的作用。

这一指数随后也被用来测量某一国家内部不同地区的发展水平。1997年开始,UNDP发布的《中国人类发展报告》也开始用这种方式来测量中国各省、直辖市和自治区的“人类发展指数”。这一指数被认为超越了单纯用GDP来衡量发展的旧观念,能够促使各省更加重视发展的非经济维度。图3显示了2017年中国各省的人类发展指数情况(联合国开发计划署等,2019)。

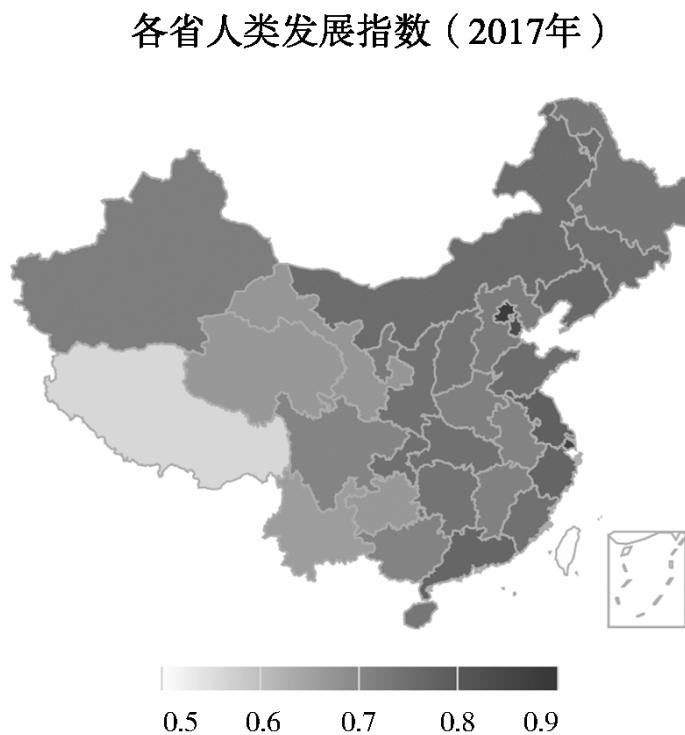


图 3 中国各省人类发展指数(2017年)

资料来源：联合国开发计划署、清华大学中国发展规划研究院、国家信息中心联合编制，2019，《中国人权发展报告特别版——历史转型中的中国人类发展40年：迈向可持续未来》，中国出版集团、中译出版社。

(三)我们究竟测量的是什么？

在现实当中，人们对于各个省人均发展指数的具体数值并不敏感，各省真正在意的是它们在人均发展指数上的排名。我们把各省在人类发展指数上的排名与他们在人均收入上的排名绘图，结果呈现在图4中。图4显示，尽管人类发展指数的构建中包含了除人均收入之后的其他两个指标（健康与教育），但是它的排名与收入排名是强相关。原因在于，各省在预期寿命和受教育年数上的差异并没有收入那样大；预期寿命和受教育年数本身与人均收入也是强相关。

尽管在理论上人类发展指数这一指标确实是诸多维度的综合加

总,但是在这一具体情境中,它对案例的区分与按人均收入进行的区分基本类似。如格尔曼等所言,在这一意义上我们测量的其实主要还是“人均收入”,虽然它经过了一些“神秘的转换”而获得了更加“炫目的名称”(Gelman, Hill, & Vehtari, 2021:22)。

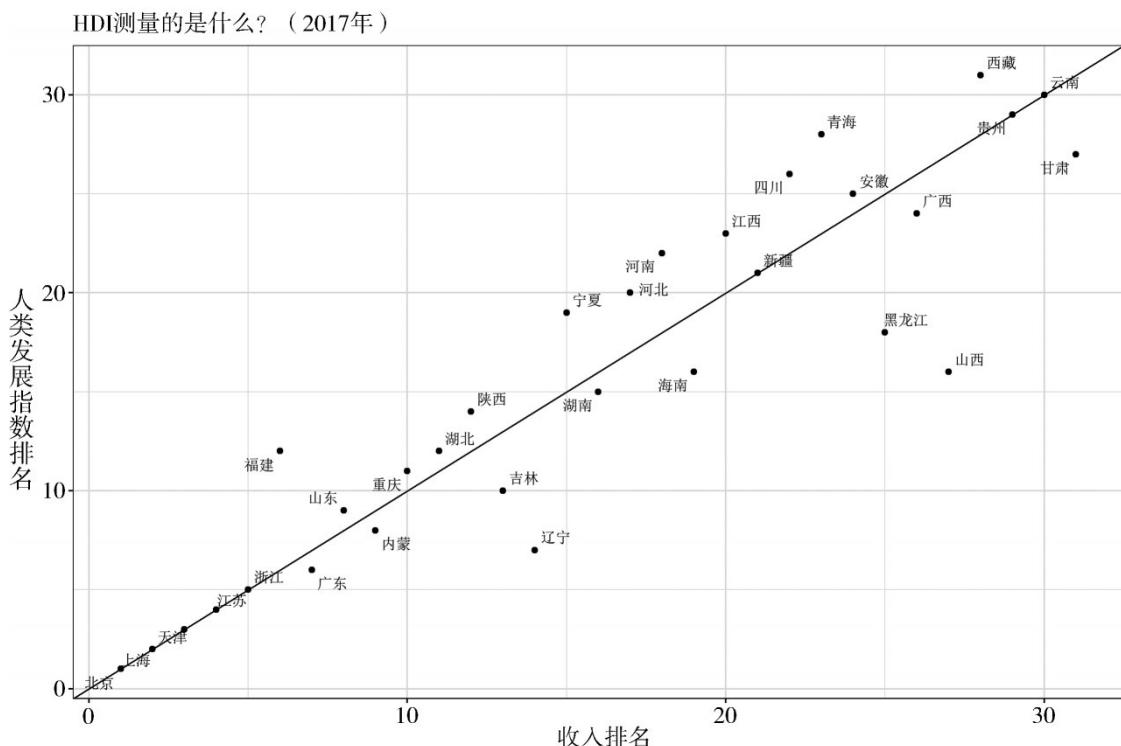


图 4 2017 年各省人类发展指数的排名与收入排名

四、一个教训

社会科学方法首先要克服的问题,就是各种形式的“自欺”(赵锋,2019)。在把概念与经验观察相关联的过程中,我们尤其要防范各种“自欺”。研究本身就是一个社会过程。我们要想了解作为研究对象的“大”的社会过程,前提是首先要充分理解自身研究这一“小”的社会过程。社会学要成为科学,必须要用社会学的眼光来看待自

己的社会研究过程。只有真正理解研究过程中究竟发生了什么,我们才能够免于各种“自欺”。

测量的通常流程是从概念出发,逐步进行概念化与操作化,最后得到反映概念要素的变量指标。“假设—演绎”的研究过程强调对于理论的检验,因此研究者容易急于去证明自己心仪的某些理论是正确的,而容易忽略测量背后潜藏的一些基本事实。由于多种原因,我们意图测量的与实质测量的内容未必是一致的。测量出来的信息究竟是什么,不仅是理论问题,也是经验问题。验证理论很重要,但前提是先要搞清楚基本事实。有必要回到事实层面,原原本本地去看数据,去琢磨现实生活中的社会过程可能是怎么回事,而不是只停留在抽象概念上。

为了防止测量过程中的种种失误,我们有必要在测量完成之后,去认真考察这一变量与其他各种变量的相关结构。如果这一变量与其他某个理论上完全独立的变量呈现出较强相关关系时,我们就需要警惕测量是否出现了某种偏差。此时,无论我们怎样为理论进行辩护,如果测量结果可以被解读为另外一种意义,其结论也就不再牢靠了。

参考文献

布尔迪厄、华康德,2015,《反思社会学导引》,李猛、李康译,北京:商务印书馆。

拉鲁,2018,《不平等的童年:阶层,种族和家庭生活(第2版)》,宋爽、张旭译,北京:北京大学出版社。

联合国开发计划署、清华大学中国发展规划研究院、国家信息中心联合编