

教育评估中的数据汇总合成

夏燕¹, 冯晖²

(1. 上海市教育评估院教育评估研究所, 2. 上海市教育评估院, 上海 200031)

摘要: 为使教育评估结果更具科学性、准确性, 文章探讨了教育评估过程中数据汇总合成的处理方法, 包含指标群体评判结果的汇总处理、评估指标属性值的规范化处理和多指标数据的合成处理等方面, 并结合实例具体介绍了合成的具体做法。

关键词: 教育评估; 群体评判; 规范化处理; 汇总合成

中图分类号: G40-058.1 **文献标识码:** A **文章编号:** 2095-3380(2020)02-0056-05

Data Integration and Synthesis in Education Evaluation

Xia Yan¹, Feng Hui²

(1. Education Evaluation Research Department, Shanghai Education Evaluation Institute;

2. Shanghai Education Evaluation Institute, Shanghai 200031)

Abstract: In order to make the education evaluation results more scientific and accurate, the paper introduces the methods of data integration and synthesis in the process of education evaluation, including the integration and processing of group evaluation results of indicators, the standard processing of attribute values of evaluation indicators and the synthesis processing of multi evaluation indicators.

Keywords: Education evaluation; Group evaluation; Standard processing; Integration and synthesis

目前,常用的教育评估工作中数据汇总合成的处理方法涉及以下问题。首先,当一些评估指标的内涵较为模糊抽象、缺乏精度度量标准时,往往采用主观评判方式,凭借专家的知识 and 经验,对评估对象进行评判,如何将专家的群体评判结果有效地汇总为评估对象在该指标上的价值。^[1]其次,对多个评估对象采取相对评估方式时,如何有效地将群体评判结果进行汇总。再次,评估指标涉及评估对象的各个方面,对不同的指标有不同的赋值方式,不同物理含义、不同量纲、不同类型的数据无法直接进行数据汇总合成,如何对各项指标的属性进行规范化处理,是进行有效数据汇总合成的先决条件。最后,如何根据多指标数据汇总合成

的要求,选取合适的评估模型,将所有指标数据汇总合成为评估对象的整体价值,从而得到科学的、准确的教育评估结果是当前教育评估亟待解决的重要问题。

针对以上问题,本文探讨了教育评估过程中数据汇总合成的主要处理方法,包含指标群体评判结果的汇总处理、评估指标属性值的规范化处理和多指标数据的合成处理等方面,并进行了实例介绍。希望能规范化教育评估工作的数据汇总合成,提高评估结果的公正性和科学性。

一、群体评判结果的汇总处理

在教育评估工作中,主观指标的属性值通常凭

借专家的知识 and 经验对评估对象进行价值评判,或者通过广泛的社会调查得到,然后将群体评判结果汇总为评估对象在该指标上的价值。^[2]在汇总群体评判结果时,受专家的知识与经验、对标准的理解和认知等的影响,每位专家的评判结果存在个体偏差和不确定性。为了集中不同领域专家的智慧,集成更多的知识优势,同时避免个人评判产生的随机性和偶然性,教育评估中通常采用群体评判。将各专家的评估结果采用数学计算方法进行汇总合成,使评估对象具有可比性。

在教育评估中,对多个评估对象采取绝对评估方式,常采用算术平均法、中间平均法和加权平均法,对多个评估对象采取相对评估方式时,常采用票选法和排序法,能有效地将群体评判结果进行汇总。考虑到群体评判中可能出现专家对评估对象给出的评判结果相差甚远的现象,需要分析群体评判结果的一致性,从而进一步保证和提高评估结果的科学性和准确性。^[3]

(一) 群体评判结果汇总合成方法

1. 群体评判结果汇总合成常用方法

在教育评估中,对于绝对评估方式,常用的群体评判结果汇总合成方法包含算术平均法、中间平均法和加权平均法,这三种方法比较简单。此外常用的群体评判结果汇总合成方法还有票选法和排序法。

(1) 票选法

票选法是指按照一定标准和要求,对 n 个评估对象做出“赞成”“弃权”“反对”的三选一式评判,评判人选择“赞成”的票数不得超过给定的限额 m ($1 \leq m \leq n$)。当 $m=n$ 时为等额票选;当 $m < n$ 时为差额票选。^[3]首先,进行有效性核定。若不符合要求,则该评判人的票选结果无效,例如赞成票数超过了给定的限额 m 。接着,对票选统计结果的汇总合成方法有得票数优先法和得票比例法。

得票数优先法根据得票数多少确定评估对象的胜出,分为赞成票多者优先和净胜票(赞成票与反对票之差)多者优先。得票比例法根据赞成票与反对票的占比来确定评估对象的胜出。例如,在某次教育评估实践中,对于赞成票达到实到人数的三分之二及以上,并且反对票未达三分之一的评估对象予以直接通过,对于其他评估对象进行现场答辩,进一步进行评估。

(2) 排序法

排序法是指根据一定的标准和要求,对 n 个评估对象按照从强到弱的顺序进行标记,若评判人认为若干评估对象强弱程度相当,可给予并列名次。例如,在某次学校内涵建设绩效评估中,邀请 10 位专家作为评判人对某项主观指标采用排序法进行评估,规范化处理后,评估次序如表 1 所示,进行数据汇总合成后,得到 R_j ,则评估对象的排列次序如排名列所示。

表 1 群体排序结果示例

评估对象	专家名单										R_j	排名
	专家1	专家2	专家3	专家4	专家5	专家6	专家7	专家8	专家9	专家10		
学校1	9	10	12	6	2	8	9	8	6	8	78	9
学校2	1.5	3	1	3	6	1	1	1	7	1	25.5	2
学校3	6	11	9	11	5	10	6.5	10	9	11	88.5	10
学校4	7	9	13	4	10	7	6.5	7	2	4	69.5	6
学校5	4	8	7	7	11	6	4	5.5	3	5	60.5	4
学校6	11	7	4	8.5	1	11	10	11	1	7	71.5	7
学校7	8	12	10	5	8	4	8	4	12	6	77	8
学校8	1.5	1	2	1	4	2	5	2	4	2	24.5	1
学校9	5	4	5	8.5	7	9	3	9	8	10	68.5	5
学校10	13	5	8	10	12	12	12	12	10	12	106	12
学校11	12	13	6	13	13	13	13	13	13	13	122	13
学校12	10	6	11	12	9	5	11	5.5	11	9	89.5	11
学校13	3	2	3	2	3	3	2	3	5	3	29	3

2. 群体评判结果汇总合成的数据过滤

应用以上方法进行群体评判结果汇总合成时,为有效获得评估对象的价值,需要对评估数据进行过滤处理,如过滤同一指标的最大值和最小值。但是,这样的处理还不够完善,过滤数据的标准可进一步考虑误差大小,而非数据自身大小。根据评判人给出的评判结果 x_{ij} 相应的标准化值 Z_{ij} 的误差进行数据过滤。 Z_{ij} 如式(1)所示进行计算。 Z_{ij} 是评判人给出的评判结果 x_{ij} 的标准化值,反映了各评判人给出的评判结果与群体评判均值的相对偏离程度。

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j} \quad (1)$$

$$S_j = \sqrt{\frac{1}{m-1} \sum_{i=1}^m (x_{ij} - \bar{x}_j)^2} \quad (2)$$

设定参数 t ,如果 $|Z_{ij}| > t$,则认为相应的 x_{ij} 误差过大而作为无效数据予以剔除;否则,认为 x_{ij} 的误差在允许的范围之内得以保留。在教育评估实践中,可根据实际情况适当简化误差大小的计算。 S_j 称为标准差,意义与 Z_{ij} 类似。

3. 基于评判准确性的数据汇总合成方法

为进一步提高数据汇总合成的准确性,借鉴数据过滤的方法,将所有评判结果的准确性逐一量化。如图1所示,数据过滤处理过程中随着参数 t 的缩小,逐步增大保留下来的数据的权重。评判准确性高的评判结果赋予较大权重,将评判准确性低的评判结果赋予较小权重。

权重 w_{ij} 的呈阶梯型模型。根据设定的权重 w_{ij} 和原始评判结果 x_{ij} 代入加权平均模型,即可得第 j 个评估对象群体评判结果的汇总合成。评判结果标准化值 Z_{ij} 的绝对值与权重 w_{ij} 成反比,两者基本符合标准正态分布。采用标准正态分布模型代替阶梯模型,则权重 w_{ij} 如式(3)~(4)所示。

$$w_{ij} = \frac{f(Z_{ij})}{f(0)} = e^{-\frac{1}{2}Z_{ij}^2} \quad (3)$$

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad (4)$$

权重 w_{ij} 的取值范围为 $[0, 1]$ 。当评判结果 x_{ij} 与群体评判结果的平均值一致时,标准化值 Z_{ij} 为0,则权重 w_{ij} 为1;当评判结果 x_{ij} 与群体评判结果的平均

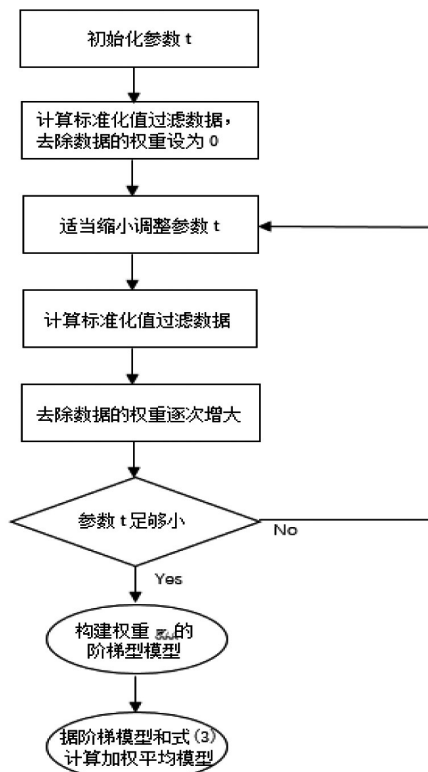


图1 基于评判准确性的数据汇总合成流程图
值相差越大时,权重 w_{ij} 越接近于0。因此,权重 w_{ij} 反映了评判结果 x_{ij} 的准确性程度。

(二) 群体评判结果一致性分析

对于各评估对象,多位评判人给出的评判结果是一组数据。当评判人数足够多时,评判结果趋于正态分布。根据正态分布的特征,理想的群体评判结果呈现一定的分布规律和集中趋势,以此检验群体评判结果的一致性程度。例如, m 位评判人对某一评估对象 j 给出的评判结果为 $x_{1j}, x_{2j}, \dots, x_{mj}$,以群体评判结果的均值为参考基准,可以用平均差 M_D 、标准差 S 和差异系数 V 等参数分析群体评判结果的一致性程度,如式(5)~(8)所示。

$$\bar{x}_j = \frac{1}{m} \sum_{i=1}^m x_{ij} \quad (5)$$

$$M_{Dj} = \frac{1}{m} \sum_{i=1}^m |x_{ij} - \bar{x}_j| \quad (6)$$

$$S_j = \sqrt{\frac{1}{m-1} \sum_{i=1}^m (x_{ij} - \bar{x}_j)^2} \quad (7)$$

$$V_j = \frac{S_j}{\bar{x}_j} \quad (8)$$

平均差 M_d 、标准差 S 和差异系数 V 均反映了群体评判结果的离散程度。平均差、标准差和差异系数越小,表示各位评判人给出的评判结果越趋近于群体评判结果的均值,该评估对象的群体评判结果一致性程度越高;反之,说明此群体评判的一致性越差,存在意见分歧,需进一步进行评估处理。^[4]

在某次学科评估中,各专家对6门学科进行评估,各门学科群体评判结果的平均差、标准差和差异系数,如表2所示。各门学科群体评判结果的平均差在2-4之间,说明各评判人给出的评判结果平均偏离均值2-4分,其中学科5的平均差相对较大一些,各学科群体评判结果的差异系数均较小,说明其群体评判结果较一致,其中学科5的差异系数较大一些。

表2 群体评判结果的一致性分析

统计量	评估对象					
	学科1	学科2	学科3	学科4	学科5	学科6
均值	83.21	86.57	92.68	76.89	72.39	66.37
平均差	3.216	2.612	2.529	2.816	3.541	2.796
标准差	3.842	3.729	3.283	4.226	4.512	3.503
差异系数	0.041	0.043	0.036	0.052	0.059	0.051

二、指标属性的规范化处理

评估指标涉及评估对象的各个方面,对不同的指标有不同的数据采集方式、赋值方式,不同物理含义、不同量纲、不同类型、不同赋值标准的数据无法直接进行数据汇总合成。^[3]因此,将指标体系的各项指标数据合成为评估对象的总价值前,通常需要对各项指标的属性值进行规范化处理。如何对各项指标的属性进行有效的规范化处理,是进行有效数据汇总合成的先决条件。

考虑将指标的原始属性值 x 通过函数变换 $y=f(x)$,转换为某一个区间内的相对数据 y ,相对数据无量纲、各数据的特性和类型一致。转换函数 $f(x)$ 称为效用函数,取值范围为 $[0,1]$, $y=f(x)$ 越接近于1表示该项指标表现得越强,对评估对象总价值的贡献越大;反之,越接近于0则表现越弱,对评估对象总价值的贡献越小。按照效用函数的性质和特

征,可分为线性效用函数和非线性效用函数。本文主要讨论线性效用函数。常用的线性效用函数有直线型、折线型和阶梯型。

1. 指标属性的直线型规范化处理

直线型效用函数一般形式如式(9)所示。其中 b 和 a 为相应指标属性值的上、下阈值,分别表示指标属性值的满意界限和达标界限。当指标属性值小于达标界限 a 时,效用函数为0;当指标属性值大于满意界限 b 时,效用函数为1。

在式(9)中,当 $x=a$ 时 $f(x)=0$ 。有时为了给最低者一个基础分即 $f(a)=h$,则可以有下列计算式(10)

$$f(x)=\begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases} \quad (9)$$

$$y=f(x)=\frac{1-h}{b-a}x+\frac{bh-a}{b-a} \quad (10)$$

2. 指标属性的折线型规范化处理

折线型效用函数的一般形式如(11)所示。折线型效用函数是直线型效用函数的变形,在直线型效用函数的 $[a,b]$ 区间范围内设置某个分界点 c ,将 $[a,b]$ 区间划分为 $[a,c]$ 和 $[c,b]$ 两个子区间,在 $[a,c]$ 子区间和 $[c,b]$ 子区间内各自是一个直线型效用函数。

$$f(x)=\begin{cases} 0 & x < a \\ \frac{(x-a)d}{c-a} & a \leq x < c \\ \frac{(x-c)(1-d)}{b-c}+d & c \leq x < b \\ 1 & x \geq b \end{cases} \quad (11)$$

在教育评估实践中,由于难以精确地确定分界点 c 及其函数值 d ,一般情况下不太使用折线型效用函数。

3. 指标属性的阶梯型规范化处理

阶梯型效用函数的一般形式如(12)所示,将指标属性值分为若干个子区间,当指标属性值处于某子区间时,效用函数为某一常数,其中 $0 < k_1 < k_2 \cdots < k_{i+1}=1$

$$f(x) = \begin{cases} 0 & x < a_1 \\ k_1 & a_1 \leq x < a_2 \\ k_2 & a_2 \leq x < a_3 \\ \vdots & \vdots \\ 1 & x \geq a_j \end{cases} \quad (12)$$

三、多指标数据的合成处理

多指标数据的属性值经过规范化处理转换为没有量纲的规范化值以后,可以通过一定的模型或合成函数将多指标数据汇总成为评估对象的总价值。采用不同的模型将影响评估对象的评估结果,因此,可采用多种模型对比分析后,进行综合评估。

教育评估中常用的多指标数据合成处理方法包含主成分分析法、投入产出比法、数据包络法、聚类分析法、模糊综合评估法、人工神经网络等。本文通过某次学校内涵建设绩效评估的实践,对主成分分析法、投入产出比法进行详细说明。

1. 主成分分析法

主成分分析法(Principal Component Analysis)简称PCA,是一种通过数学变换将一组可能存在相关性的变量转换为一组线性不相关的变量来代替原来的变量的方法,在数学上常用来处理降维。^[5]在教育评估中,应用主成分分析法能将原先存在相关性的指标,重新组合成互相无关的综合指标。

例如,某次学校内涵建设绩效评估的实践涉及29所学校,初步设计的指标体系包含81个指标,23个投入指标,58个产出指标。使用SPSS软件对指标体系进行主成分分析法,对指标进行精简。首先,对指标属性值进行群体评判结果汇总合成、规范化处理;其次,构建产出指标的相关系数矩阵,如表3所示;再次,计算产出指标的相关系数矩阵的特征根与特征向量,如表4所示;随后,确定产出指标可综合为21个主成分指标,累计贡献率达92.981%;最后,得到各学校产出主成分指标的得分,如表5所示。投入指标亦可用此法进行分析精简。

在评估指标较多的情况下,主成分分析法可以在保证精度的前提下降低指标空间的维数,并尽量消除指标间的关联性。

2. 投入产出比法

表3 产出指标相关系数矩阵

产出指标					
指标1	指标2	指标3	...	指标57	指标58
0.6083	0.5108	0.8923	...	0.3138	0.1203
0.5407	0.4036	1.0000	...	0.4007	0.3303
0.3518	0.5036	0.8485	...	0.0683	0.1958
0.8116	0.6738	0.8531	...	0.1403	0.0538
...	...	...	...	...	...
...	...	...	...	...	...
0.8163	0.8361	0.6943	...	0.0030	0.0863
1.0000	0.3927	1.0000	...	1.0000	0.0086
0.5888	0.6360	1.0000	...	1.0000	0.0100

表4 相关系数矩阵的特征根与特征向量

主成分	特征根	贡献率	累计贡献率
1	4.125	37.059%	35.065%
2	2.392	18.097%	54.111%
...	...	...	...
20	0.726	4.538%	89.412%
21	0.548	3.901%	92.981%

表5 产出指标主成分指标的得分

评估对象	主成分指标				
	主成分1	主成分2	...	主成分20	主成分21
学校1	0.247	0.597	...	0.520	0.364
学校2	-0.245	-0.307	...	0.404	0.771
学校3	-0.246	0.323	...	0.570	-0.158
...	...	...	...	...	...
学校28	0.484	-0.095	...	-0.782	0.071
学校29	0.411	-0.463	...	0.626	-0.442

投入产出比法是衡量评价活动效益的基本方法,是教育领域中进行教育评价的经典方法之一。^[6]学校绩效同样注重办学效益、资源利用率等。因此,某次学校内涵建设绩效评估的实践采用投入产出比模型,即以学校的产出指标价值与投入指标价值的比率作为度量其绩效的结果,如式(13)所示。

$$\text{内涵建设绩效} = \frac{\text{学校产出成果}}{\text{学校投入资源}} \quad (13)$$

采用投入产出比模型,有利于既注重产出成果的数量和质量等绝对成绩,更注重办学效益,提高资源利用率。

(下转第80页)

- Taylor & D M Cowley, (Eds). Readings in Curriculum Evaluation [M]. Chicago: University of Chicago, 1972:84-86.
- [3] Stufflebeam D L. Toward a technology for evaluating evaluation (ERIC Document Reproduction Service No. ED090319), 1974.
- [4] Stufflebeam D L. Metaevaluation: concept, standards, and uses [C]// R A Berk (Eds). Educational Evaluation Methodology: The State of the Art [M]. Baltimore, MD: Johns Hopkins University, 1981:146-163.
- [5] 严芳. 教育元评估的理论与实践[M]. 上海: 华东师范大学出版社, 2013:63;108.
- Yan Fang. Theory and practice of education meta-evaluation [M]. Shanghai: East China Normal University Press, 2013:63;108.
- [6] Yarbrough D B, Shulha L M, Hopson R, et al. The program evaluation standards: A guide for evaluators and evaluation users (3rd ed.) [EB/OL]. <https://jcsee.org/program/>.
- [7] Keun-bok Kang, Chan-goo Yi. A design of the meta evaluation model. Poster session presented at the annual meeting of the Canadian Evaluation Society, Montréal, Canada, 2000.
- [8] 王致和. 高等学校教育评估[M]. 北京: 北京师范大学出版社, 1995:62.
- Wang Zhihe. Higher education evaluation [M]. Beijing: Beijing Normal University Press, 1995:62.
- [9] 吴钢. 现代教育评价教程[M]. 北京: 北京大学出版社, 2008:99.
- Wu Gang. Modern education evaluation course [M]. Beijing: Peking University Press, 2008:99.
- [10] 严芳. 从美国教育评估标准与原则看我国教育评估专业化发展[J]. 黑龙江高教研究, 2011(10):24-27.
- Yan Fang. Viewing the professional development of education evaluation in China from the standards and principles of education evaluation in the United States [J]. Heilongjiang Researches on Higher Education, 2011 (10):24-27.

(上接第60页)

参考文献

- [1] 张伟江. 上海教育评估工作规程(2008年版)[M]. 北京: 高等教育出版社, 2009.
- Zhang Weijiang. Regulations of Shanghai Education Evaluation (2008 version) [M]. Beijing: Higher Education Press, 2009.
- [2] 冯晖, 王奇. 专家评判准确性分析及专家星级评定方法[J]. 复旦教育论坛, 2012, 10(6):59-63.
- Feng Hui, Wang Qi. Analysis of the Accuracy of Expert Judgement and Method to Identify the Star-level of Expert [J]. Fudan Education Forum, 2012, 10(6): 59-63.
- [3] 冯晖. 教育评估计算学[M]. 北京: 高等教育出版社出版, 2012.
- Feng Hui. Education Evaluation Computing [M]. Beijing: Higher Education Press, 2012.
- [4] 冯晖. 教育评价中的敏感性问题研究[J]. 华东师范大学学报(教育科学版), 2007, 25(2):37-41.
- Feng Hui. Research on Sensitivity Problems in Education Evaluation [J]. Journal of East China Normal University (Educational Sciences), 2007, 25(2): 37-41.
- [5] 林海明, 杜子芳. 主成分分析综合评价应该注意的问题[J]. 统计研究, 2013, 30(8):25-31.
- Lin Haiming, Du Zifang. Some Problems in Comprehensive Evaluation in the Principal Component Analysis [J]. Statistical Research, 2013, 30(8): 25-31.
- [6] 杨雪, 冯晖. 基于主成分分析的高校投入产出绩效评价[J]. 上海管理科学, 2012, 34(5):107-109.
- Yang Xue, Feng Hui. An Evaluation on the Input-Output Performance of Universities Based on Principal Component [J]. Shanghai Management Science, 2012, 34(5): 107-109.