

医药院校教学质量评价系统指标数据变换方法研究^{*}

王 轶 孔瑞珍 孙 焱

(山西医科大学图书馆 太原 030001)

〔摘要〕 简要介绍教学质量评价指标数据变换的目的与主要内容,对数据变换中的统计分布问题进行探讨,在对各种数据变换方法进行比较的基础上,认为标准差法能够将分布各异的原始数据变换成统一的标准正态分布的标准化数据,是一种比较科学合理的数据变换方法。

〔关键词〕 教学质量;评价指标;数据变换;统计分布

Research on Index Data Transformation Methods of Teaching Quality Evaluation System in Medical Colleges and Universities

WANG Yi, KONG Rui-zhen, SUN Yan, Library of Shanxi Medical University, Taiyuan 030001, China

〔Abstract〕 The paper briefly introduces the aims and main contents of index data transformation in teaching quality evaluation, discusses statistic distribution problems in data transformation. Based on the comparison among various data transformation methods, it generates a conclusion that standard deviation method would make different original data transform into standardized data which has unified standard normal distribution, it is a scientific and rational data transformation method.

〔Keywords〕 Teaching quality; Evaluation index; Data transformation; Statistic distribution

在对高校教学质量进行评价的相关研究中,问题探讨主要集中在评价方法的改进和指标权重的确定两个方面,而关于评价中指标数据的变换方法却关注较少^[1]。任何一种综合测评的方法都离不开数据的支撑,如果在评价中忽略了指数据的有效性,势必会对评价系统的整体科学性造成一定的影响,使评价缺乏客观公正的结果。医药院校教学有自己的特殊性,教学活动包括基础、临床、理论、实践等多个环节和人财物等因素,教学质量

评价的指标体系较为复杂,针对不同类型的评价对象有不同的评价内容。指标的原始数据往往也来源于不同的总体,因此科学合理的数据变换是实现评价结果客观准确的基本保证^[2-3]。

1 数据变换基本问题

1.1 目的

数据变换的目的主要在于通过无量纲化对不同计量单位的指标能够进行综合。所谓量纲就是计量单位,所谓无量纲化就是去掉统计指标计量单位的过程,无量纲化的结果就是将有计量单位的指标形式转换成无计量单位的指标形式,从而使各项指标数值可比、可加,最终达到能够综合的目的。

〔收稿日期〕 2010-08-24

〔作者简介〕 王轶,馆员,发表论文7篇。

〔基金项目〕 山西省教育厅高等学校科技研究开发项目“高等医药院校教学质量评价系统的开发与研究”(项目编号:200811014)。

1.2 两类问题

在对医药院校教学质量的综合评价中,涉及到数据变换的主要是两类问题:第 1 类是对各项指标的原始数据进行无量纲处理,使量纲不同的各项指标转化为可以直接计算的数值,一般多用于不同对象的横向比较;第 2 类是对教学质量的发展水平进行测算,以便反映指标数值的增长或下降的变化情况,一般多用于不同时间的纵向比较。在数据变换的两类问题中,第 1 类问题是基础和前提,第 2 类问题是在第 1 类问题的基础上进行测算。可以认为数据变换的第 1 类问题更为重要,本文只就第 1 类问题进行讨论。

1.3 教学质量评价指标数据变换的主要内容

根据本项目所设计的医药院校教学质量评价系统的指标体系,数据变换主要包括两个方面的内容:一是在学生评教中,由于学生总体各异,教师所教课程(教学环节)不同,取得的原始数据不具备可加性和可比性,在进行综合时必须进行数据变换。二是由于各种评价指标的量纲不同,不能直接相加,在进行综合时也必须进行数据变换。

2 数据变换中的统计分布问题

2.1 统计分布的重要性

统计分布是统计学中一个非常重要的概念,其实质是把总体的全部单位按标志所分的组进行分配所形成的变量数列,又称为分配数列、分布数列或次数分布^[4]。统计分布可分为概率分布和频率分布两种,概率分布是基于理论模型的理论分布,频率分布是基于实际统计数据的经验分布,可以根据经验分布为总体选配理论模型。很显然教学质量评价各项指标数据形成的分布是频率分布。统计分布是统计数据整理结果的一种重要表现形式和分析方法,它可以表明统计数据的数字特征、结构状况、平均水平、变异程度和变动规律等情况。在教学质量评价中,各评价对象作为总体单位在各项指标数值上形成的变量数列,反映了不同的统计分布。统

计理论认为,以相同统计分布为背景和参照系的变量数列中的数值才有可比性和可加性,综合才更加科学和合理^[5]。例如,在学生对教师教学质量的评价中,每位教师的评价指标原始数据是来源于不同的学生群体(总体),显然不经过数据变换而把原始数据直接综合作为评教排序的评价方式缺乏科学性,得出的评价结果也不具说服力。因此,数据变换的实质不仅仅是实现度量单位无量纲化,更为重要的是实现统计分布的一致性,达到各项指标的变换数值相加综合更加科学合理的目的。

2.2 数据变换的基本性质

教学质量评价系统中的许多指标都表现为强度相对指标、结构相对指标、平均指标。如:学生人均教学科研仪器设备值(元/人)、具有高级职称教师占专任教师的比例(%)、学生人均图书(册/人)。显而易见,由于这些指标计量单位不同,是不能直接比较和相加的。即便是学生直接打分的指标,也由于学生总体差异、教师所上课程(环节)不同,也是不能直接比较和相加的。因此,首先必须对各种指标的原始数据进行无量纲处理,变换成对比基础相同的无量纲单位的无名数。而无论哪一种数据变换方法,其基本性质都是线性变换,即变换后不改变指标原始数据大小的先后顺序和分布形状,其线性变换的一般数学表达式为 $x_i = A + Bx_i$ 。

2.3 数据变换中的统计分布类型

根据统计分布原理和数据变换的基本性质,用各种变换方法对指标的原始数据进行变换,就形成了新的不同的变量数列。对于归一化法变换形成的变量数列,其分布的平均数为 $\bar{x} = \frac{1}{n} \sum (x_i - x_{\min}/x_{\max} - x_{\min})$, 方差为 $\sigma^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$ 。可以看出,对于不同指标形成的新的变量数列,其平均数和方差是不一样的,说明不是同一种统计分布。对于用对数法、均值化法、功效系数法等方法变换形成的新的变量数列,也可以证明都不是同一种统计分布。而只有用标准差法进行数据变换,对于不同指标形成的变量数列服从同一种统计分布,即标

准正态分布 $N(0, 1^2)$ ，其平均数为 0，方差为 1^2 ，其数据变换的计算公式为 $x_i - \bar{x}/\sigma$ ，其中 x_i 为指标的原始数据值， $\bar{x}$ 为某项指标的平均值： $\bar{x} = \frac{1}{n} \sum x_i$ ， σ 为其标准差 $\sigma = [\frac{1}{n} \sum (x_i - \bar{x})^2]^{\frac{1}{2}}$ 。

3 基于统计分布的数据变换方法比较研究

3.1 数据变换方法实例分析

以山西医大教学质量评价系统中学生对教师课堂教学质量评价的两项指标（X₁ 教学方法，X₂ 学习态度）原始数据为例，进行数据变换，见表 1。

表 1 几种数据变换方法统计分布计算

样本	原始数据		归一化法		对数法		均值化法		标准差法	
			$\frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \times 100$		$\frac{\log x_i - \log x_{\min}}{\log x_{\max} - \log x_{\min}} \times 100$		$x_i/\bar{x} \times 100$		$(x_i - \bar{x})/\sigma$	
	X ₁	X ₂	X ₁	X ₂	X ₁	X ₂	X ₁	X ₂	X ₁	X ₂
教师 1	88	82	62.50	23.08	63.57	24.47	100.74	96.31	0.27	-0.75
教师 2	91	84	100.00	38.46	100.00	40.28	104.17	98.66	1.54	-0.27
教师 3	89	82	75.00	23.08	75.85	24.47	101.88	96.31	0.69	-0.75
教师 4	89	89	75.00	76.92	75.85	78.24	101.88	104.53	0.69	0.93
教师 5	86	80	37.50	7.69	38.59	8.26	98.45	93.96	-0.57	-1.23
教师 6	84	83	12.50	30.77	13.01	32.42	96.16	97.48	-1.41	-0.51
教师 7	86	79	37.50	0.00	38.59	0.00	98.45	92.79	-0.57	-1.47
教师 8	89	92	75.00	100.00	75.85	100.00	101.88	108.05	0.69	1.65
教师 9	88	81	62.50	15.38	63.57	16.41	100.74	95.13	0.27	-0.99
教师 10	89	87	75.00	61.54	75.85	63.32	101.88	102.18	0.69	0.45
教师 11	83	86	0.00	53.85	0.00	55.73	95.01	101.01	-1.84	0.21
教师 12	86	91	37.50	92.31	38.59	92.83	98.45	106.88	-0.57	1.41
教师 13	85	87	25.00	61.54	25.88	63.32	97.30	102.18	-0.99	0.45
教师 14	90	89	87.50	76.92	87.99	78.24	103.03	104.53	1.11	0.93
平均($\bar{x}$)	87.36	85.14	54.46	47.25	55.23	48.43	100.00	100.00	0.00	0.00
方差(σ^2)	5.63	17.36	879.98	1 027.37	880.20	1 027.68	7.38	23.95	1.00	1.00

注:数据来源于 2009 年度山西医科大学医学信息管理系青年教师课堂教学质量综合测评。

从计算结果的数字特征看，只有用标准差法变换形成的变量数列均服从 $N(0, 1^2)$ 的标准正态分布，这正是保证数据可比可加的重要基础。而用其它方法变换形成的变量数列则服从数字特征不同的统计分布。从变换后指标的取值范围看，归一化法和对数法的最小值为 0，最大值为 100，取值范围的界限非常清楚，也符合人们用百分制评价的习惯；标准差法变换的取值范围有 99.9999% 的数值在 $[-5, 5]$ 的区间上^[6]，经过线性变换也可在 0 ~ 100 之间取值；而其它变换方法的取值范围界限就非常不清楚，不能给出一个明确的印象。从指标数值的综合（X₁ + X₂）情况看，原始数据直接综合的结果和经标准差法变换后的综合结果是不一样的，见表 2。可以看出，用标准差法数据变换方法排序，有 7 名教师的顺序发生了变化。以上综合排

序没有加权，只是简单算术平均，而且只有两项指标，如果是加权平均则对综合排序的结果影响更大。

表 2 数据变换后综合指标计算比较

样本	原始数据		标准差法	
	(X ₁ + X ₂) / 2	排序	(X ₁ + X ₂) / 2	排序
教师 8	90.50	1	1.17	1
教师 14	89.50	2	1.02	2
教师 4	89.00	3	0.81	3
教师 12	88.50	4	0.42	6
教师 10	88.00	5	0.57	5
教师 2	87.50	6	0.63	4
教师 13	86.00	7	-0.27	9
教师 3	85.50	8	-0.03	7
教师 1	85.00	9	-0.24	8
教师 9	84.50	10	-0.36	10
教师 11	84.50	11	-0.82	11
教师 6	83.50	12	-0.96	13
教师 5	83.00	13	-0.90	12
教师 7	82.50	14	-1.02	14

3.2 标准差法数据变换的特点

如前所述,各种数据变换都是线性变换,其计算公式都是由一个分式形式构造的数学模型。而标准差法变换所构造的数学模型,之所以能够使新的变量数列统计分布一致,其主要原因是这种变换的数学模型符合由一般正态分布变换成标准正态分布的原理,即如果 $x \sim N(\mu, \sigma^2)$, 则 $x - \bar{x}/\sigma \sim N(0, 1^2)^{[7]}$ 。服从标准正态分布就是同一种参照系,其数据之间是等距的,因此可以进行加减运算,这一良好性质是数据之间可比可加的根本保证,也是不同于其它数据变换方法的主要特点。

3.3 标准差法数据变换的问题及解决方案

在标准差法数据变换中,还有两个主要问题需要解决:其一,数据变换的前提是教学质量评价指标的原始数据分布应呈正态分布或近似正态分布,如果有的指标分布不是正态分布该如何处理;其二,用计算公式 $(x_i - \bar{x})/\sigma$ 进行数据变换后得到的数据会出现负数,且取值范围在 $[-5, 5]$ 区间上,这与人们对事物评价的百分制习惯方式不太一样。

对于第 1 个问题,一般情况下可以先对原始数据进行正态化处理,把非正态分布的原始数据转换成百分等级数据,然后查对应的正态分布表就可以转换为标准正态分布的数据;对于第 2 个问题,只要对取得的数据再进行一次线性变换就可以解决,常用的数学模型为 $T = 50 + 10(x - \bar{x})/\sigma$ 。从正态分布的性质可知, $-5 \leq (x - \bar{x})/\sigma \leq 5$ 占正态分

布面积的 99.9999%, 因此可以使 T 在 0 ~ 100 范围之间取值。

4 结语

教学质量评价系统中评价指标数据变换不仅要考虑变换方法的科学性和合理性,还要考虑其它因素对数据变换结果的影响,比如指标的性质,是总量指标、平均指标,还是相对指标。指标的性质不同,各项指标原始数据分布的离散情况会对数据变换结果产生何种影响,这些都需要联系实际认真思考,从而促进数据变换方法研究的进一步深入。

参考文献

- 1 候三军,陈雅,郑建民. 基于 RITE 测度的数据变换研究[J]. 情报学报, 2007, (4): 315-320.
- 2 王秀平,卢莉,孔瑞珍,等. 医学信息管理实践教学体系的建立与实施[J]. 医学信息学杂志, 2007, (2): 198-200.
- 3 唐品,杨双琪. 医学信息检索课课堂教学质量评价研究[J]. 医学信息学杂志, 2009, (3): 85-87.
- 4 黄良文,陈仁恩. 统计学原理[M]. 北京:中央广播电视大学出版社, 2000: 88-91.
- 5 王汉澜. 教育测量学[M]. 郑州:河南大学出版社, 1988: 374-384.
- 6 王耀文,刘永胜. 专题文献著者群评价——一种“标准化”方法的探讨[J]. 图书情报工作, 2007, (4): 97-99.
- 7 白富志,张贵恩. 概率论与数理统计[M]. 北京:机械工业出版社, 1999: 40-44.

敬告作者

《医学信息学杂志》网站现已开通,投稿作者请登录期刊网站: <http://www.yxxxx.ac.cn>,在线注册并投稿。

《医学信息学杂志》编辑部