

22 省市汉族女青少年生长发育状况的分析

季成叶 叶恭绍 袁捷

(北京医科大学儿童青少年卫生研究所)

关键词 青少年;生长;发育;聚类分析

内 容 提 要

利用 SPSS 计算机软件的 PCA-Q 分析程序,从分析 7 项生长发育指标的相似与相异点入手,对我国 22 个省市区的 44 个汉族城乡女青少年样本进行聚类分析。分析结果表明,各地区女青少年在生长发育状况上存在着显著差异,这些差异与地理因素、城乡差别和社会经济发展水平有关。表现为北方女青少年的体格生长水平普遍高于南方,城市高于农村,经济发达地区高于相对欠发达地区。作者并就本结果在了解中国青少儿生长发育状况及趋势方面的作用,以及 PCA-Q 分析程序的实际应用体会进行了讨论。

在我们这个幅员辽阔、人口众多的多民族国家里,汉族约占人口总数的 93.7%。然而,生活在不同地区的汉族青少年由于受遗传、地理环境、社会经济、生活方式等诸因素的长期影响,在生长发育水平上存在着巨大差异。研究这类差异及其影响因素,不仅是人类学家关心的课题,在医学、卫生学、体质生理学,乃至有关国计民生的各行各业,均有现实意义。

1985 年由国家教委、民委、体委与卫生部联合组织的大规模“中国学生体质与健康调查”,曾对我国青少年生长发育的地区差异作过大量描述(中国学生体质与健康调研组,1987)。遗憾的是,这些描述主要是在对单项测量指标进行分析比较的基础上作的,不易全面反映体格状况,也很难表现出不同地区青少年在生长发育状态与趋势方面的特点和差异。

本研究是在该调查基础上进行的再分析,目的是利用聚类分析方法,对我国部分省市自治区女青少年的生长发育状况进行分类比较,并对产生差异的地理及社会经济因素作初步分析探讨。

一、材料与方法

分析对象

来自“中国学生体质与健康调查”中 22 省市区的¹⁾的女青少年样本。这些省、市和自治区是:北京、天津、河北、内蒙古、辽宁、吉林、黑龙江、山东、陕西、青海、宁夏、上海、浙江、

1) 除台湾省外,西藏因无汉族资料,山西、甘肃、新疆、湖北、江苏和四川因无月经初潮资料,故未列入本研究。

安徽、福建、江西、河南、湖南、广东、广西、贵州与云南。每一省市区又分城市、农村两片(各作一组样本),分别含7—18岁十二个连续年龄组,每组人数229—306人,合计140,448人,44组样本。

分析指标

计有:(1)身高;(2)坐高;(3)体重;(4)相对身高指数(身高/ $\sqrt[3]{\text{体重}}$);(5)相对胸围指数(胸围/身高 $\times 100$)(以上各项分别以18岁时的均值表示);(6)身高突增高峰(Peak height velocity)年龄;(7)月经初潮年龄等7项。

分析方法

使用 SPSS 计算机软件内的 PCA-Q 型分析程序(三宅一郎等,1982)进行聚类分析,步骤如下:(1)计算各样本内各指标的均值,建立指标间的相关系数矩阵;(2)根据相关关系抽取若干主成分,由此把各指标的均值转化成主成分得分;(3)利用下列公式计算每两两样本间的空间距离,以此作为判断样本间相似度的基准。

$$D_{ik} = \sqrt{\sum_{j=1}^m (F_{ij} - F_{kj})^2} / \sqrt{M}$$

式中, D_{ik} 为任意两样本 i 与 k 的空间距离; F_{ij} 与 F_{kj} 为该两样本内某指标的主成分得分; i 为指标的个数; M 则为主成分的个数。

根据 D_{ik} , 空间距离最短(相似度最高)的样本首先聚合成“类”;这些类逐步吸收与它们距离最近的样本,逐步形成更大的“类”;最后,“类”与“类”之间再按相似度的高低逐步结合,直到形成一个最终聚合体。其全部聚合过程可以系统图的方式表示。

二、结 果

1. 各生长发育指标间的相关关系与主成分分析

表1是由七项指标所建的相关系数矩阵。除体重外,各指标间均有较显著的相关,体现了它们彼此之间的紧密关系。例如,她们18岁时的身高、坐高与相对身高指数、相对胸围指数分别有较高的正(或负)相关,提示那些身材较高的女青少年群体体型趋细长,那些身料较矮的群体体型则趋粗壮。身高突增高峰年龄、月经初潮年龄与身高、坐高和相对胸围指数呈负相关,与相对身高指数呈正相关,提示在我国女青少年中那些身材占优势的群

表1 七项生长发育指标、指数的相关系数矩阵

指标(或指数)	V_1	V_2	V_3	V_4	V_5	V_6	V_7
V_1 (18岁身高)	1.000						
V_2 (18岁坐高)	0.935▲▲▲	1.000					
V_3 (18岁体重)	0.351▲▲	0.440▲▲▲	1.000				
V_4 (身高突增高峰年龄)	-0.589▲▲▲	-0.511▲▲▲	-0.246	1.000			
V_5 (相对身高指数)	0.724▲▲▲	0.594▲▲▲	-0.388▲▲	-0.759▲▲▲	1.000		
V_6 (相对胸围指数)	-0.641▲▲▲	-0.615▲▲▲	0.246	0.692▲▲▲	-0.811▲▲▲	1.000	
V_7 (月经初潮年龄)	-0.720▲▲▲	-0.675▲▲▲	0.095	0.680▲▲▲	-0.777▲▲▲	0.620▲▲▲	1.000

▲ $P < 0.05$, ▲▲ $P < 0.01$, ▲▲▲ $P < 0.001$ 。

体一般青春期发育也比较早。

表 2 显示主成分分析的结果。在指定条件下(所抽出的主成分的特征值均须大于 1.0), 共抽取 F_1 、 F_2 两个主成分, 以代表七项生长发育指标。它们的特征值之和为 6.0814。可用以说明上述七指标的变异的 86.9%。

表 2 七项生长发育指标、指数的主成分解

指标、指数	抽出主成分		共性方差
	F_1	F_2	
V_1		0.6362	0.9556
V_2		0.7152	0.9622
V_3		0.9355	0.9701
V_4	0.8597		0.7411
V_5	-0.9633		0.9331
V_6	0.8822		0.7785
V_7	0.8426		0.7410
特征值	4.4485	1.6329	6.0814
累积百分率(%)	63.5%	23.5%	86.9%

2. 聚类分析结果——系统图及其解释

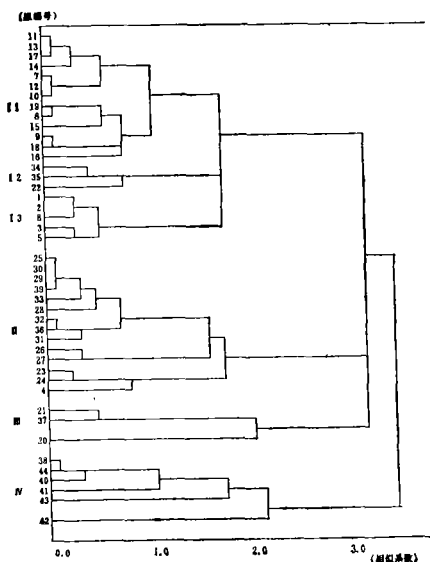


图 1 22 省市区女青少年生长发育聚类图

1.北京城; 2.天津城; 3.河北城; 4.内蒙城; 5.辽宁城; 6.吉林城; 7.黑龙江城; 8.山东城; 9.陕西城; 10.青海城; 11.宁夏城; 12.上海城; 13.浙江城; 14.安徽城; 15.福建城; 16.江西城; 17.河南城; 18.湖南城; 19.广东城; 20.广西城; 21.贵州城; 22.云南城; 23.北京农; 24.天津农; 25.河北农; 26.内蒙农; 27.辽宁农; 28.吉林农; 29.黑龙江农; 30.山东农; 31.陕西农; 32.青海农; 33.宁夏农; 34.上海农; 35.浙江农; 36.安徽农; 37.福建农; 38.江西农; 39.河南农; 40.湖南农; 41.广东农; 42.广西农; 43.贵州农; 44.云南农

图 1 为 PCA-Q 聚类分析结果。在较高的相似度(相似系数 0.2)下,形成四类,分别聚合了 21、14、3 和 16 个样本。这四类的最终聚合是在相似系数为 0.3—0.4 时。

表 3 对四类女孩各自的生长发育特点进行分析比较。第 I 类女孩身高、坐高最高,身高突增高峰和月经初潮的出现也最早。与此相比,第 II 类女孩的身高、坐高比较低,但体重重,相对身高指数低而相对胸围指数高,显得相对矮而粗壮。第 III 类女孩的身高、坐高和体重值更低,但从其他项指标的比较来看,其生长与青春发育趋势都最接近第 I 类女孩的水平。第 IV 类女孩生长发育水平最低。例如,她们的身高、坐高比第 I 类平均要低 4.37 厘米和 2.20 厘米,身高突增高峰年龄与月经初潮年龄也较晚,与第 II 类接近。

表 3 四类女青少年生长发育状况的比较*

类别	身高 (厘米)	坐高 (厘米)	体重 (公斤)	相对身 高指数	相对胸 围指数	身高突增 高峰年龄(岁)	月经初潮 年龄(岁)
I 类 (21 组)	158.37 (1.12)	86.11 (0.44)	49.76 (0.92)	43.06 (0.21)	49.25 (0.60)	10.5 (0.73)	13.1 (0.29)
II 类 (14 组)	157.05 (0.73)	85.77 (0.35)	51.40 (0.69)	42.36 (0.30)	51.23 (0.52)	11.8 (0.41)	13.7 (0.30)
III 类 (3 组)	155.39 (0.68)	84.79 (0.52)	47.64 (0.52)	42.85 (0.26)	48.47 (1.00)	10.7 (0.94)	13.5 (0.36)
IV 类 (6 组)	154.00 (0.83)	83.91 (0.50)	49.22 (0.82)	42.00 (0.29)	51.95 (0.73)	12.0 (0.50)	14.1 (0.24)

* 括号内为标准差。

从图 1 的第 I 类内,还可看到在更高相似度时形成的 I₁(含 13 组)、I₂(含 3 组)、I₃(含 5 组)三个亚类。从表 4 的比较来看,这些地区的女孩具有相同的生长发育特点,但在部分指标上仍有一定差异。例如,I₃与 I₂相比,身高、坐高和体重均值分别高出 2.70 厘米、1.08 厘米和 1.62 公斤,身高突增高峰年龄平均早一年。

表 4 I 类内三亚类女青少年生长发育状况的比较*

类别	身高 (厘米)	坐高 (厘米)	体重 (公斤)	相对身 高指数	相对胸 围指数	身高突增 高峰年龄(岁)	月经初潮 年龄(岁)
I ₁ 亚类 (13 组)	158.14 (0.79)	86.05 (0.34)	49.30 (0.56)	43.12 (0.19)	49.81 (0.97)	10.4 (0.74)	13.1 (0.27)
I ₂ 亚类 (3 组)	157.06 (0.13)	85.53 (0.17)	49.48 (0.75)	42.80 (0.18)	49.83 (0.85)	11.3 (0.47)	13.3 (0.32)
I ₃ 亚类 (5 组)	159.76 (0.72)	86.61 (0.21)	51.10 (0.16)	43.09 (0.16)	49.32 (0.36)	10.2 (0.40)	12.9 (0.16)

* 括号内为标准差。

3. 各地女孩生长发育差异的原因粗析

城乡差异:聚类分析结果与表 3、表 4 的比较表明,我国女孩的生长发育存在着显著城乡差异。身材高而瘦长、青春发育早的第 I 类女孩多数来自城市;而那些因具备相似生长发育趋势(身材相对矮而粗壮,青春发育较晚)而聚入第 II、IV 类的除一例(内蒙城女)外均来自农村。聚入第 III 类的三例中两例来自广西和贵州城女,她们除身材矮小外,其余

指标均更接近主要由城市女孩组成的第 I 类。

地理差异: 聚类分析还表明, 许多生长发育水平接近的样本来自毗邻地区, 北方女孩的身材普遍高于南方。例如, 生长发育水平最高的 I₃ 亚类来自北京、天津、河北、山东、辽宁等北方毗邻城市地区; 生长水平最低的第 IV 类来自江西、湖南、广东、广西、贵州、云南等南方毗邻农村地区。来自北方农村地区的样本全部集中在第 II 类。I₁ 亚类混杂着南北城市样本, 但仍可发现吉林与黑龙江、陕西、青海和宁夏, 上海、浙江和安徽, 以及江西、福建、广东和湖南等都是相互毗邻的。

经济发展差异: 表 5 对城市地区部分社会经济指标的比较(中国国家统计局, 1985) 表明, 由历史原因造成的经济发展不平衡与我国女青少年生长发育的地区差异也有关。表内以人均社会生产总值为限, 将 22 个省市区分为 A、B、C 三类。其中 A 类主要分布于沿海, 人口密集, 工业发达, 人均社会生产总值、商品零售额和生活消费水平明显较高, 城乡消费水平的差距也相对较小。这些地区的女孩(除南方的上海、浙江城女外) 多属于生长发育水平最高的 I₃ 类。相反, C 类地区经济水平最低, 其中的内蒙、广西、贵州城女被分别聚入第 II、III 类, 生长发育水平在 22 个城市组中是最低的。B 类地区的经济发展界于两者之间, 由它们作为主体构成的 I₁ 类生长发育水平也属中等。农村样本间因缺乏可比性统计资料无法作类似比较, 但仍可发现不少相同表现。其中最明显的是来自农村经济最发达的上海和浙江乡女。她们被同时聚入第 I₂ 亚类, 其生长水平比内蒙、广西、贵州等城女更高。

表 5 22 省市区的城市地区部分社会经济指标的比较***

类别	人口密度 (人/平方公里)	城市人口 占总人口 比(%)	城市增长占 省市内总生 产比(%)	社会生产 总值 (元/人)	社会商品 零售额 (元/人)	生活消费 水平 (元/人)	城乡消费 水平比 (%)
A 类 (6 组)	726.2 (562.6)	54.2 (14.6)	83.7 (13.8)	6621.3 (1449.7)	1791.6 (335.6)	853.7 (133.2)	1.73 (0.28)
B 类 (11 组)	229.1 (115.0)	32.0 (12.3)	68.3 (6.2)	3975.9 (729.2)	1512.7 (464.9)	696.9 (81.3)	2.25 (0.36)
C 类 (5 组)	83.2 (70.1)	31.6 (5.8)	63.8 (3.5)	2709.75 (590.2)	1372.9 (171.9)	679.0 (77.1)	2.43 (0.38)

* 资料选自 1985 年中国统计年鉴, 括号内为标准差。

** A 类构成地区(人均社会生产总值 >5001 元): 北京、天津、辽宁、山东、上海、浙江。

B 类构成地区(人均社会生产总值 3500—5000 元): 吉林、黑龙江、河北、陕西、安徽、福建、江西、河南、湖南、云南、广东。

C 类构成地区(人均社会生产总值 <3500 元): 贵州、广西、内蒙古、宁夏、青海。

三、讨 论

多变量聚类分析, 是探讨生长发育群体差异的重要手段之一。早在 1949 年, Mahalanobis 等就曾利用多项人体测量数据, 作过不同地区、种族的形态学聚类分析。Panrose (1954) 则首次把多项测量指标转化成“体格”、“体型”两大主成分, 从分析样本间的相似点入手, 探讨不同人群的生长差异。Vasula 等 (1989) 利用同方法, 分析了印度 Yanadis

种族的生长地区差异,并对产生差异的地理、生态环境、文化、生活习惯等因素作了分析。本研究沿用以上理论,同时根据女孩特点,改选成七项指标。其中,身高、坐高和体重反映形态生长的三个侧面;相对身高指数和相对胸围指数虽不能全面反映体型,但它们分别通过比较人体的胖瘦程度和胸廓发育状况(进而可推测心肺功能),在反映女孩生长趋势方面仍起较重要作用。它们与另两项衡量青春发育的重要指标——“身高突增高峰年龄”和“月经初潮年龄”结合,可较全面反映群体女孩的发育水平。本研究所进行的聚类分析,是在主成分分析的基础上进行的,其优点是:(1)所抽出的两个主成分, F_1 代表两项指数和两项年龄指标(见表2),其特征值均较高,累积起来可说明全体变异的63.5%,是聚类分析的主要依据, F_2 代表身高、坐高和体重,所说明的变异(23.5%)较低,原因是样本间在这些指标上的差异较小。两主成分的累积百分率高达86.9%,说明其对全体指标的代表性较好。(2)通过主成分抽取,把由七项性质各不相同的指标组成的多维空间转化成低维空间,使样本间的距离计算在数学上更趋合理(Everitt, 1980);(3)在确定相似度时既考虑到各变量的大小差异又重视变量间的相关,避免了以往一些单用相关系数或单用距离测定来衡量相似度的方法中易偏重某一侧面而忽视另一侧面的倾向(Aldenderfer *et al.*, 1984)。

聚类分析结果把22省市区的女青少年群体划分成四种在生长发育趋势上有显著差异的类型。同时又体现出,产生这些地区差异的主要原因既有地理因素、城乡差别,也有经济发展不平衡所造成的长期影响。这些因素纵横交错,十分复杂,有时相辅相成,有时又互相排斥,谁也无法单独起作用。例如,一般来说北方女孩的生长发育优于南方女孩,但作为南方城女的安徽(No.14)、浙江(No.13)、上海(No.12)却和北方的青海(No.10)、黑龙江(No.7)、宁夏(No.11)等城女相聚合。又如,城乡女孩在生长发育水平上有明显差异,但上海乡女(No.34)、浙江乡女(No.35)却和多数城女样本同聚在水平最高的一类里;相反,内蒙城女(No.4)只能与乡女样本聚合,其生长水平比上海、浙江等乡女样本更低。

以上分析仅是在我国部分地区的汉族女青少年中进行,用于反映我们这个幅员辽阔、地理环境复杂,拥有11亿人口的多民族国家的青少年生长全貌是远远不够的。何况还有不少影响生长发育的因素,如日照、海拔、气温、气湿、生活习惯和方式等在本文中尚未得到充分讨论。营养状况对青少年生长的影响最为直接,但由于缺少统计资料,在诸如饮食结构,热量、蛋白质、钙和维生素等营养物质的摄入等方面无法作进一步的比较分析,只能通过对部分社会经济指标(见表5)的比较来粗略反映,成为一个明显缺憾。尽管如此,本研究的结果是有积极意义的。它有助于从宏观角度全面了解我国青少年的生长发育特点及地区差异,同时也为各级政府和科研部门从本地区基本条件出发,制订青少年身体素质评价和体育锻炼标准,采取必要的医疗保健、改善营养、卫生监督和各种行政干预手段提供了科学依据。

(1990年1月22日收稿)

参 考 文 献

三宅一郎等,1977。SPSS 解析编。217—254,东洋经济新报社。

- 中国学生体质与健康调研组, 1987. 全国学生体质与健康调查. 1202—1437, 人民教育出版社。
- 中国国家统计局, 1985. 1985 年中国统计年鉴. 54—129, 中国统计出版社。
- Aldenderfer, N. S. and R. K. Lashfield, 1984. *Cluster Analysis*. 1—17, Sage Publications, California.
- Everitt, B. S., 1980. *Cluster Analysis*. 2—15, Halsted Press, New York.
- Mahalanobis, P. C., D. N. Majumdar, and C. R., Rao, 1949. Anthropometric survey of the United provinces, 1941: A statistic study. *Sankhya.*, 9:89—324.
- Panrose, L. S., 1954. Distance, size and shape. *Ann. Eug.*, 13: 228—237.
- Vasula, T. S. and M., Pal, 1989. Size and shape components of anthropometric differences among the Yanadis. *Ann. Hum. Biol.*, 16: 449—462.

CLUSTER ANALYSES ON GROWTH AND DEVELOPMENT AMONG 22 PROVINCES' HAN GIRLS

Ji Chengye Ye Gongshao Yuan Jie

(*Institute of Child & Adolescent Health, Beijing Medical University*)

Key words Adolescent; Growth; Development; Cluster analysis

Abstract

This study was designed to study the morphological differences and similarities of Chinese Han girls' growth and development by using PCA-Q mode cluster analysis (SPSS language). The data consisted of 44 urban and rural cases came from 22 provinces. In each subject, seven anthropometric measures and indices were used for analysing. The results showed significant variations of Chinese Han girls' growth and development existed not only between north and south China, but also between urban and rural areas, and between developed and underdeveloped areas. These variations were shown clearly on the dendrogram, and the geographic and socio-economic factors causing these variations were investigated. The practical use of PCA-Q cluster analysis and its merits in studying Chinese girls' growth and development were also discussed by the authors.