

# 高中数学核心素养的统计分析

张淑梅, 何雅涵, 保继光

(北京师范大学, 北京 100875)

摘要: 学生对六个核心素养的掌握情况有一定差异, 但素养之间均具有显著的相关性, 其中逻辑推理和数学运算的相关性最大, 并且数学运算对逻辑推理的影响比逻辑推理对数学运算的影响更大。数学建模与数据分析的素养对其他四种数学素养的依赖程度明显大于其他四种素养对它们的依赖程度, 说明数学抽象、逻辑推理、直观想象、数学运算是基本数学素养。可以将数学素养分为三类: 数学抽象、逻辑推理、直观想象、数学运算作为基本数学素养为一类, 数学建模与数据分析各成一类。

关键词: 数学核心素养; 聚类分析; 广义相关测量; 因子分析

中图分类号: G633.6 文献标志码: A 文章编号: 1000-0186(2017)10-0050-06

数学核心素养是适应个人终身发展和社会发展需要的具有数学特征的思维品质与关键能力。高中阶段数学核心素养包括: 数学抽象、逻辑推理、数学建模、直观想象、数学运算和数据分析。这些数学核心素养既相对独立, 又相互交融, 是一个有机的整体。本文将使用统计的方法, 基于测试结果数据研究这六个数学核心素养之间的相关关系, 在此基础上给出若干建议。

## 一、数据来源

2016 年, 教育部普通高中数学课程标准测评组有关专家编制了七套不同的试卷。试卷除了涉及六个数学核心素养之外, 也包括不同情境(现实情境、数学情境、科学情境)和不同难度(三个水平)。测评组将试卷下发给 2 807 名高中生, 并获得有效作答 2 697 份, 其中高二学生 2 184 人, 高三学生 513 人。测评组对试卷中每一道大题中的每一小题进行分割, 确定其考察的核心素

养。为了能对学生的素养情况判断更加准确, 又将每个素养的表现分为四个层次: 0 分为完全没有掌握该素养, 1~2 分对应水平一, 3~4 分对应水平二, 5~6 分对应水平三。最后对学生的作答结果进行评判, 获得每一个学生的素养得分。

由于七套试卷各个核心素养的满分不同, 为了对被试进行比较, 本文将每个素养得分除以该试卷的这个素养的满分作为这个被试素养的最终得分。这样各个素养的得分均限制在  $[0, 1]$  内。

## 二、试卷的质量的基本分析

### (一) 信度检验

测验的信度是指测验结果的稳定性和可靠性, 以及与考生真实水平的一致性。测验的信度水平越高, 测验结果的误差越小。本文使用原始数据采用同质性检验的方法分别对七套试卷进行信度检验, 计算出每套试卷的 *Cronbach's  $\alpha$*  系数, 结果如表 1。

收稿日期: 2017-04-29

作者简介: 张淑梅, 1963 年生, 吉林长春人, 北京师范大学教授, 博士, 从事应用统计和教育统计的研究; 何雅涵, 1993 年生, 辽宁人, 北京师范大学统计学院研究生, 研究方向为应用统计; 保继光, 1963 年生, 北京人, 北京师范大学数学科学学院教授, 研究方向为偏微分方程和数学教育。

表 1 信度检验汇总表

试卷编号	项数	Cronbach's $\alpha$	基于标准化项 Cronbach's $\alpha$
1	14	0.861	0.867
2	15	0.726	0.743
3	14	0.814	0.831
4	14	0.880	0.889
5	15	0.866	0.850
6	14	0.818	0.833
7	13	0.832	0.831

结果显示,除了试卷 2 以外,其他试卷的 Cronbach's  $\alpha$  系数均在 0.8 以上,试卷 2 的 Cronbach's  $\alpha$  系数也在 0.7 以上,表明该系列试卷可以通过信度检验、测试结果的可靠性比较高。

### (二) 效度检验

效度是指通过试卷能够测验到计划测验的内容的程度,及测验是否有效,有效的程度。效度主要分为内容效度、结构效度和效标关联效度三种类型,这里我们主要关注问卷的内容效度。

本次测试,课标组除了自己编制数学素养的测试题之外,还分别在北京、天津、河北、苏州等地成立了数学核心素养试题的研制小组收集和改编试卷,测试全面地考察了六个核心素养。为了保证试卷的效度,数学课标测试组还在北京、天津、河北和苏州进行了预测试,并根据预测试的结果完善问卷。测试组还邀请富有经验的高考命题专家(包括国家考试中心的专家)对试卷进行审核和调整,试题得到了专家的充分肯定。经过几轮修订之后,测试组又围绕着课标中数学核心素养的水平和指标体系对试题进行了分析和统计,以保证较好地考察考生的数学素养水平。

### (三) 测验难度

测验难度指测验试题的难易程度,是衡量测验试题对学生知识水平适宜程度的指标。一般来说难度处于 0.5 左右最为适宜。本文利用学生最终素养得分(得分经过预处理,限制在 [0, 1] 区间)来检验各个素养试题的难度,得到各个素养难度水平如下:

表 2 各素养难度水平汇总

	数学抽象	逻辑推理	数学建模
难度水平	0.263 656	0.263 309	0.115 209
	直观想象	数学运算	数据分析
难度水平	0.331 349	0.262 268	0.388 918

结果显示,各数学素养的难度水平均在 0.5

以下,对于学生来说试题的难度水平偏高,说明学生比较不适应基于核心素养的问题。

### (四) 测验区分度

试题的区分度是测验鉴别考生实际能力水平高低的量度,它是评价试题质量、筛选试题的主要指标与依据。本文采用极端值法对六个数学核心素养的最终得分进行了区分度检验,分别将六个核心素养升序排列,找出分数的上 27% 分位数和下 27% 分位数,将学生分为高分组和低分组,并对这两组学生进行独立样本  $t$  检验,结果显示六个核心素养均通过检验,表明各个素养试题总体上来说区分度比较好。

## 三、核心素养的描述性统计分析

### (一) 核心素养的分布情况

为了更加直观地展示学生对六个数学核心素养的掌握情况,我们通过对每个数学素养绘制直方图的方式来观察各个素养的分布情况,结果显示,六个数学核心素养得分分布情况大体一致,总体上来看,水平低的学生比例较大,水平高的学生比例较小。这种结果部分是学生在作答时不认真导致的,反馈上来的试卷也表明确实有一部分学生不认真作答甚至交白卷,使得各素养水平很低的学生比例很大。除此之外,也可能跟测评组设置的题目太难或者比较新颖导致学生作答情况不理想所致。如果对学生各素养最低水平要求过高,就会导致很大一部分学生达不到最低水平。

对比上述六个素养,学生掌握情况比较好的是直观想象和数据分析,水平低的学生比例不是特别大;掌握的最不好的一个素养是数学建模,绝大部分学生都处于最低水平。在今后的教学中建议:“学生能有意识地用数学语言表达现实世界,感悟数学与现实之间的关联;学会用数学模型解决实际问题,积累数学实践的经验;认识数学模型在科学、社会、工程技术诸多领域的作用,提升应用能力,增强创新意识和科学精神”。<sup>[1]</sup>

### (二) 核心素养的成对比较

两组比较方法通常被用来检验两个随机变量的均值之间是否有显著性差异,主要分为两种:独立组比较和成对组比较。独立组比较一般用于两个独立样本的观测数据,例如男女体脂含量的

随机样本；成对组比较一般应用于样本中每个个体包含成对测量，也就是说对每个个体进行两种测量。

为了研究学生对于六个核心素养掌握情况的差异性，我们对六个核心素养进行了两两分组比较。由于每组两个素养都是个体的成对测量，本文采用成对组比较的方法。

首先得到的是六个核心素养得分的均值与方差，如表 3。

表 3 六个核心素养的简单统计量

素养	观测数	求和	平均	方差
数学抽象	2 697	711.079 2	0.263 656	0.069 232
逻辑推理	2 697	710.143 5	0.263 309	0.067 149
数学建模	2 311	367.233 3	0.158 907	0.065 368
直观想象	2 697	893.648 6	0.331 349	0.057929
数学运算	2 697	707.335 5	0.262 268	0.059 838
数据分析	1 954	759.946 4	0.388 918	0.112 002

直观上来看，数据分析平均得分最高，直观想象平均得分其次，数学建模平均得分最低，而数学抽象、逻辑推理和数学运算的平均得分比较接近。假设六个素养得分的均值相等，对其两两进行成对组比较，得到检验结果如表 4 所示。

表 4 两组比较的  $p$  值

	数学抽象	逻辑推理	数学建模
数学抽象	—	0.946 3	<0.000 1
逻辑推理	0.946 3	—	<0.000 1
数学建模	<0.000 1	<0.000 1	—
直观想象	<0.000 1	<0.000 1	<0.000 1
数学运算	0.747 2	0.732 8	<0.000 1
数据分析	<0.000 1	<0.000 1	<0.000 1
	直观想象	数学运算	数据分析
数学抽象	<0.000 1	0.747 2	<0.000 1
逻辑推理	<0.000 1	0.732 8	<0.000 1
数学建模	<0.000 1	<0.000 1	<0.000 1
直观想象	—	<0.000 1	<0.000 1
数学运算	<0.000 1	—	<0.000 1
数据分析	<0.000 1	<0.000 1	—

观察表 4，可以得到数学抽象、逻辑推理、数学运算这三个素养两两之间的  $p$  值较大，不能拒绝原假设，即可以认为这三个素养得分的均值大致相等，从数学素养直方图中也可以看出学生对于这三个素养的掌握情况差不多。而这三个

素养与其他素养以及其他素养之间的均值均可以认为是显著不相等的，其中数学建模的得分最低，因此现今学生的数学建模素养有待提高。

### (三) 核心素养的 Pearson 相关系数分析

六个核心素养的掌握情况存在着一定的差异，下面采用 Pearson 相关系数的方法来检验这些核心素养之间是否存在着一一定的相关性。

基于处理后的数据，我们计算六个素养的 Pearson 相关矩阵，结果见表 5。

表 5 六个核心素养的 Pearson 相关矩阵

	数学抽象	逻辑推理	数学建模
数学抽象	1.000 00 2 697	0.475 71 <0.001 2 697	0.498 29 <0.001 2 697
逻辑推理	0.475 71 <0.001 2 697	1.000 00 2 697	0.425 14 <0.001 2 311
数学建模	0.498 29 <0.001 2 311	0.425 14 <0.001 2 311	1.000 00 2 311
直观想象	0.564 37 <0.001 2 697	0.740 65 <0.001 2 697	0.434 50 <0.001 2 311
数学运算	0.614 18 <0.001 2 679	0.803 83 <0.001 2 679	0.532 44 <0.001 2 311
数据分析	0.259 25 <0.001 1 954	0.388 28 <0.001 1 954	0.410 14 <0.001 1 568
	直观想象	数学运算	数据分析
数学抽象	0.564 37 <0.001 2 697	0.614 18 <0.001 2 697	0.259 25 <0.001 1 954
逻辑推理	0.740 65 <0.001 2 697	0.803 83 <0.001 2 697	0.388 28 <0.001 1 954
数学建模	0.434 50 <0.001 2 311	0.532 44 <0.001 2 311	0.410 14 <0.001 1 568
直观想象	1.000 00 2 679	0.655 46 <0.001 2 697	0.345 67 <0.001 1 954
数学运算	0.655 46 <0.001 2 679	1.000 00 2 697	0.389 82 <0.001 1 954
数据分析	0.345 67 <0.001 1 954	0.389 82 <0.001 1 954	1.000 00 1 954

所有六个素养之间均具有显著的正相关性，取值落在 (0.25, 0.81) 中。逻辑推理与数学运算的相关性最大；数据分析与数学抽象的相关系数最小。数据分析与其他五个素养的相关性偏

小，均在 (0.25, 0.41) 之间，与数学建模的相关系数最大，说明数据分析与数学建模有一定的相关性。

为了更加直观地观察六个核心素养的相关性大小，我们绘制了散点矩阵图（图略），可以看出，逻辑推理和数学运算之间具有很强的相关性，直观想象与逻辑推理和数学运算也有较强的相关性，而数据分析和数学建模与其他素养之间均不具有明显的相关关系。

#### (四) 核心素养的聚类分析

我们利用对变量聚类的统计方法将相关性较大的核心素养分成一类，得到谱系聚类图，见图 1。六个核心素养可以分为三类：数学抽象、逻辑推理、数学运算、直观想象为一类，它们是数学的基本素养；数学建模单独为一类，偏重数学知识在实际问题中的应用；数据分析自成一类，主要关注数据的分析处理能力。

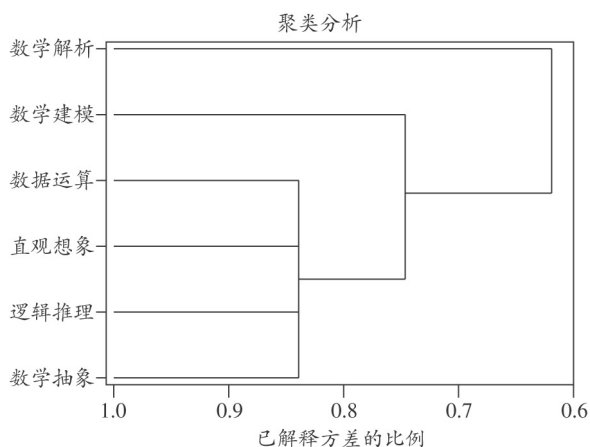


图 1 类平均法谱系聚类

#### 四、核心素养的非对称相关性

以上对于六个核心素养的相关性分析只能描述两两之间的对称关系，并不能得出哪一些核心素养更重要的结论。现在利用郑术蓉、史宁中等人提出的广义相关测量方法（GMC）来计算核心素养两两之间的非对称相关系数。<sup>[2]</sup>对于两个随机变量  $X$  和  $Y$ ，可以采用  $X$  方差中能用  $Y$  解释的部分与变量  $X$  方差的比值来表示  $Y$  解释  $X$  的程度，即  $Y$  对  $X$  的影响，计算公式为

$$GMC(X | Y) = \frac{\text{var}(E(X | Y))}{\text{var}(X)}$$

其中， $E(X | Y)$  表示给定变量  $Y$  时变量  $X$

的条件期望， $\text{var}(X)$  表示变量  $X$  的方差。类似地，可以计算  $X$  对  $Y$  的影响，得到非对称相关系数公式为：

$$\{GMC(X | Y), GMC(Y | X)\} = \left\{ \frac{\text{var}(E(X | Y))}{\text{var}(X)}, \frac{\text{var}(E(Y | X))}{\text{var}(Y)} \right\}$$

$GMC(Y | X)$  不一定等于  $GMC(X | Y)$ ，也就是说广义相关测量可以描述两个变量之间非对称的相关关系，同时广义相关测量并不仅仅反应变量之间的线性关系，也可以刻画出非线性关系。

一般来说，当  $GMC(Y | X) = GMC(X | Y)$  时，我们认为  $X$ 、 $Y$  这两个变量之间的影响是一样的；当  $GMC(Y | X) > GMC(X | Y)$  时， $X$  对  $Y$  的解释比  $Y$  对  $X$  的解释要大，说明  $X$  对  $Y$  的影响比  $Y$  对  $X$  的影响要大。

六个数学素养的非对称相关系数的计算结果如表 6 所示，其中第一行第二列的数字 0.455 表示数学抽象对逻辑推理的影响，第二行第一列的数字 0.408 表示逻辑推理对数学抽象的影响，以此类推。

表 6 数学六个核心素养的非对称相关系数

	数学抽象	逻辑推理	数学建模
数学抽象	1	0.455	0.482
逻辑推理	0.408	1	0.408
数学建模	0.332	0.231	1
直观想象	0.529	0.666	0.461
数学运算	0.520	0.826	0.536
数据分析	0.161	0.237	0.206
	直观想象	数学运算	数据分析
数学抽象	0.476	0.448	0.449
逻辑推理	0.623	0.790	0.450
数学建模	0.257	0.337	0.336
直观想象	1	0.541	0.569
数学运算	0.580	1	0.494
数据分析	0.242	0.244	1

分析上述结果可以得到以下结论：

1. 同 Pearson 相关系数的结果一致，在六个核心素养中，逻辑推理对数学运算的影响是 0.790，数学运算对逻辑推理的影响是 0.826，这两个系数是所有系数中最大的，说明逻辑推理和数学运算这两组变量之间的相关性较大，在训练其中一个素养时，另外一个素养的能力也会相

应得到提升。并且数学运算对逻辑推理的影响比逻辑推理对数学运算的影响更大。

2. 数学抽象、逻辑推理、直观想象、数学运算这四个核心素养对数学建模的影响都明显大于数学建模对这四个核心素养的影响（差均大于0.150）。类似地，其他五个素养对数据分析的影响要明显大于数据分析对其他五个素养的影响。说明数学抽象、逻辑推理、直观想象、数学运算这四个核心素养是数学的基本素养。

3. 数学抽象、逻辑推理、直观想象、数学运算这四个核心素养两两之间的相互影响之差的绝对值都小于0.08，说明它们中任意两个的相

互影响相差不大。

### 五、核心素养的因子分析

根据上文对六个核心素养的统计分析可以发现，数学抽象、逻辑推理、数学运算、直观想象四个核心素养之间有较强的相关关系，它们对数学建模和数据分析的影响比较大，下面采用因子分析的方法进一步分析六个数学素养的结构特征。

因子分析主要研究相关阵或协差阵的内部依赖关系，将多个具有高度相关关系的变量归纳为少数几个因子的影响。<sup>[3]</sup>我们对六个核心素养进行因子分析得到因子累积贡献率如表7。

表7 因子累积贡献率

	特征值	差值	比例	累积贡献率
1	3.751 895 52	2.994 662 11	0.625 3	0.625 3
2	0.757 233 41	0.217 938 26	0.126 2	0.751 5
3	0.539 295 15	0.074 249 94	0.089 9	0.841 4
4	0.465 045 21	0.107 176 24	0.077 5	0.918 9
5	0.357 868 96	0.229 207 21	0.059 6	0.978 6
6	0.128 661 75		0.021 4	1.000 0

根据表7，保留三个因子能够解释超过六个数学素养的84.14%的信息时，所以选取三个公因子。此时，方差极大正交旋转后的因子载荷阵如表8所示。

表8 方差极大正交旋转后的因子载荷阵

	Factor1	Factor2	Factor3
数学抽象	0.801 78	0.158 26	0.098 43
逻辑推理	0.833 53	0.351 26	0.176 32
直观想象	0.796 67	0.135 46	0.282 66
数学运算	0.816 86	0.369 92	0.140 58
数学建模	0.320 51	0.913 57	0.197 05
数据分析	0.212 53	0.177 36	0.954 99

第一公因子的因子载荷主要是逻辑推理、数学抽象、直观想象和数学运算，它们是数学学习最基本的素养，是学好数学的基础，可以将其命名为基本素养；

第二公因子的因子载荷最大值为数学建模的载荷0.913 57，主要强调学生使用数学方法来解决实际问题的能力，其他数学素养对该因子也是有贡献的；

第三公因子的因子载荷最大值为数据分析的载荷0.954 99，主要强调学生对于数据分析处理

的能力。

获得最终的公因子方差估计如表9。

表9 公因子方差估计

逻辑推理	数学抽象	数学建模
0.849 244 34	0.677 589 80	0.976 165 23
直观想象	数学运算	数据分析
0.732 930 15	0.823 862 52	0.988 632 03

观察表9可以发现，使用这三个因子来划分数学核心素养比较好。因此可以用上述三个核心素养：基本素养、数据分析、数学建模来解释原来的六个核心素养。

### 六、结论

本文主要使用聚类分析、广义相关测量、因子分析等统计方法来对六个数学核心素养进行数据分析，得到以下结论。

1. 学生对六个核心素养的掌握情况有一定差异，但素养之间均具有显著的相关性，其中逻辑推理和数学运算的相关性最大，并且数学运算对逻辑推理的影响比逻辑推理对数学运算的影响更大。

2. 数学建模与数据分析的素养对于其他四

个数学素养的依赖程度明显大于其他四个素养对于它们的依赖程度,说明数学抽象、逻辑推理、直观想象、数学运算是基本数学素养。

3. 通过因子分析和聚类分析的统计方法都说明可以将数学素养分为三类:数学抽象、逻辑推理、直观想象、数学运算作为基本数学素养为一类,数学建模与数据分析各成一类。

参考文献:

[1] 许世红, 胡中锋. 数学试卷分析方法 [M]. 上海:

华东师范大学出版社, 2009.

[2] Shurong Zheng, Ning-Zhong Shi, Zhengjun Zhang. Generalized measures of correlation for asymmetry, nonlinearity, and beyond [J]. Journal of the American Statistical Association, 2012 (107): 1239-1252.

[3] 高惠璇. 实用统计方法与 SAS 系统 [M]. 北京: 北京大学出版社, 2001.

(责任编辑: 李冰)

## Statistical Analysis of Mathematical Key Competency in Senior High School

Zhang Shumei, He Yahan, Bao Jiguang

(Beijing Normal University, Beijing 100875, China)

**Abstract:** There are significant differences between the six mathematical key competencies, but there are significant correlation between them. The correlation between logical reasoning and mathematical operation is largest; and the influence of mathematical operation to logical reasoning is bigger than the influence of logical reasoning to mathematical operation. The dependence of mathematical modeling and data analysis literacy on other four core competencies is significant deeper than dependence of the four key competencies on the mathematical modeling and data analysis. We can conclude that mathematical abstraction, logical reasoning, intuitive imagination and mathematical operation are the basic mathematical competencies, and we can divide the mathematical key competencies into three categories: mathematical abstraction, logical reasoning, intuitive imagination. Mathematical operation and mathematical modeling is one class separately, and data analysis is one class.

**Key words:** mathematical key competency; cluster analysis; generalized measures of correlation; factor analysis

### 本刊参考文献规范格式

#### 一、普通图书(包括教材)、学位论文

[序号] 主要责任者. 文献题名 [文献类型标志]. 出版地: 出版者, 出版年: 页码 (当整体引用时不注). 例:

[1] 杜威. 我们怎样思维·经验与教育 [M]. 姜文闵, 译. 北京: 人民教育出版社, 2005: 256-257.

#### 二、期刊文章

[序号] 主要责任者. 文献题名 [J]. 刊名, 年, 卷(期): 起止页码. 例:

[2] 王林. 小学渗透数学思想方法的实践与思考 [J]. 课程·教材·教法, 2010 (9): 53-58.

#### 三、报纸文章

[序号] 主要责任者. 文献题名 [N]. 报纸名, 出版日期 (版次). 例:

[3] 温儒敏. 守住并发扬重视教学的好传统 [N]. 中国教育报, 2008-11-20 (11).

#### 四、析出文献

[序号] 析出文献主要责任者. 析出文献题名 [文献类型标志] // 原文献主要责任者. 原文献题名. 出版地: 出版者, 出版年: 析出文献起止页码. 例:

[4] 叶圣陶. 关于探讨教材教法的几点想法 [C] // 课程教材研究所. 教材制度沿革篇 (上册). 北京: 人民教育出版社, 2004: 377-378.

#### 五、电子文献

[序号] 主要责任者. 电子文献题名 [电子文献及载体类型标志]. (发表或更新日期) [引用日期]. 电子文献的网址. 例:

[5] OECD (2013). PISA 2012 Assessment and Analytical Framework Mathematics Reading Science Problem Solving and Financial Literacy. [EB/OL]. (2013-02-11) [2015-04-18]. [http://www.oecd.org/pisa/pisaproducts/PISA%202012%20framework%20e-book\\_final.pdf](http://www.oecd.org/pisa/pisaproducts/PISA%202012%20framework%20e-book_final.pdf).