

新课标下初中统计领域新增内容教学探索

徐德同

(江苏省教育科学研究院,210013)

摘要:《义务教育数学课程标准(2022年版)》初中统计领域的课程,新增了“数据分组的原则”和“箱线图与百分位数”的内容。增加前者,立足于大数据时代的特征,体现了课程的时代性。相关的教学建议有:引导学生感悟数据分组的必要性;让学生适应信息化生态,形成大数据思维;激发学生探索分组准则的求知欲。增加后者,进一步完善了初中生对统计图表的认识,有利于形成整体的知识结构。相关的教学建议有:让学生通过实例制作,感悟箱线图的结构特征;让学生通过知识拓展,形成完整的认知结构。

关键词:初中数学;统计领域;数据分组;箱线图

与《义务教育数学课程标准(2011年版)》相比,《义务教育数学课程标准(2022年版)》(以下简称“新课标”)初中统计领域的课程,新增了“数据分组的原则”和“箱线图与百分位数”的内容,具体要求分别是“经历数据分类的活动,知道按照组内离差平方和最小的原则对数据进行分类的方法”^[1]和“会计算四分位数,了解四分位数与箱线图的关系,感悟百分位数的意义”^[2]。新增内容的学科内涵是什么?育人价值体现在哪里?教学实施有哪些注意点?厘清这些问题,是课程实施的

基本前提。

一、新增的“数据分组的原则”

(一) 内容解析

1. 数据分组的内涵及价值

数据分组是指根据统计研究的需要,将原始数据按照某种标准划分成不同的组别。在大数据分析中,对于通过调查得到的数据,虽然经过审核、排序等整理手段给予了处理,但是,由于数据庞杂,还不能直接进入对数据的分析阶段,有必要对数据进行分组处理,使大量无序、混沌的数据变为有序、层次分明、

显示总体特征的资料。在取得完整、正确的统计资料前提下,数据分组的优劣是决定整个统计研究成败的关键,直接关系到统计分析的质量。从一定意义上说,数据整理的中心任务就是分组和编制频数分布。

大数据时代的特征是数据渗透到每一个行业和职能领域,大数据给社会治理、工业发展、民生服务、日常生活等带来了越来越多的便利,正在改变人们的生活和理解世界的方式。就像望远镜能让我们感知宇宙,显微镜能让我们观察微生物一样,大数据开启了一次重大的时代转型,成为新发明和新服务的源泉。与传统的数据分析相比,大数据分析要在海量数据中寻找数据规律,发现数据异常,所以,数据预处理就显得分外重要。数据分组就是数据预处理的手段之一。因此,新课标增加“数据分组的原理”的内容,立足于大数据时代的特征,体现了课程的时代性。

2. 数据分组的原理

和数学中的分类一样,数据分组也必须做到不遗漏(穷尽性)和不重复(互斥性),即每一个数据都要且只能划归到某一组中去。在此基础上,数据分组的一个基本原则是,使组内的差距尽可能小,组间有明显的差异。

假设有 n 个数据 x_1, x_2, \dots, x_n , 不失一般性,假设这些数据都不相等。如果把这些数据分为两组,例如,前 m 个数据为第一组,后 $n-m$ 个数据为第二组。记 $\bar{x}_1 = \frac{x_1 + x_2 + \dots + x_m}{m}$, $\bar{x}_2 = \frac{x_{m+1} + x_{m+2} + \dots + x_n}{n-m}$, 则 $x_1 + x_2 + \dots + x_m = m\bar{x}_1$, $x_{m+1} + x_{m+2} + \dots + x_n = (n-m)\bar{x}_2$ 。于是:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n},$$

$$S^2 = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n$$

$$- \bar{x})^2 = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_m - \bar{x})^2 + (x_{m+1} - \bar{x})^2 + (x_{m+2} - \bar{x})^2 + \dots + (x_n - \bar{x})^2 = (x_1 - \bar{x}_1 + \bar{x}_1 - \bar{x})^2 + (x_2 - \bar{x}_1 + \bar{x}_1 - \bar{x})^2 + \dots + (x_m - \bar{x}_1 + \bar{x}_1 - \bar{x})^2 + (x_{m+1} - \bar{x}_2 + \bar{x}_2 - \bar{x})^2 + (x_{m+2} - \bar{x}_2 + \bar{x}_2 - \bar{x})^2 + \dots + (x_n - \bar{x}_2 + \bar{x}_2 - \bar{x})^2 = (x_1 - \bar{x}_1)^2 + (x_2 - \bar{x}_1)^2 + \dots + (x_m - \bar{x}_1)^2 + (x_{m+1} - \bar{x}_2)^2 + (x_{m+2} - \bar{x}_2)^2 + \dots + (x_n - \bar{x}_2)^2 + m(\bar{x}_1 - \bar{x})^2 + (n-m)(\bar{x}_2 - \bar{x})^2 + 2[(x_1 - \bar{x}_1)(\bar{x}_1 - \bar{x}) + (x_2 - \bar{x}_1)(\bar{x}_1 - \bar{x}) + \dots + (x_m - \bar{x}_1)(\bar{x}_1 - \bar{x})] + 2[(x_{m+1} - \bar{x}_2)(\bar{x}_2 - \bar{x}) + (x_{m+2} - \bar{x}_2)(\bar{x}_2 - \bar{x}) + \dots + (x_n - \bar{x}_2)(\bar{x}_2 - \bar{x})] = (x_1 - \bar{x}_1)^2 + (x_2 - \bar{x}_1)^2 + \dots + (x_m - \bar{x}_1)^2 + (x_{m+1} - \bar{x}_2)^2 + (x_{m+2} - \bar{x}_2)^2 + \dots + (x_n - \bar{x}_2)^2 + m(\bar{x}_1 - \bar{x})^2 + (n-m)(\bar{x}_2 - \bar{x})^2 = S_1^2 + S_2^2。$$

其中, $S_1^2 = (x_1 - \bar{x}_1)^2 + (x_2 - \bar{x}_1)^2 + \dots + (x_m - \bar{x}_1)^2 + (x_{m+1} - \bar{x}_2)^2 + (x_{m+2} - \bar{x}_2)^2 + \dots + (x_n - \bar{x}_2)^2$, 通常称为组内离差平方和; $S_2^2 = m(\bar{x}_1 - \bar{x})^2 + (n-m)(\bar{x}_2 - \bar{x})^2$, 通常称为组间离差平方和。^[3]

依据方差的统计学意义,组内离差平方和 S_1^2 越小,组内数据的差距就越小;组间离差平方和 S_2^2 越大,组间数据的差异就越明显。由于一组数据的离差平方和 S^2 (除以数据个数 n 即得方差)是一个定值,当组内离差平方和 S_1^2 最小时,组间离差平方和 S_2^2 最大。所以,一个合理的分组原理是使 S_1^2 最小,称为数据分组的“组内离差平方和最小原理”。

(二) 教学建议

1. 引导学生感悟数据分组的必要性

数据分组是确定数据类型的基础,是更好地认识数据特征、发现数据规律的前提。

教学中,要创设真实性情境让学生讨论,从而直观感悟数据分组的必要性。例如:统计得到全省八年级学生数学学业质量监测数据,对这一海量的数据如何分析呢?抛出问题后,让学生讨论,形成一些共识。比如:按地域标志,可以把数据划分为城区、城镇和郊区、农村等组类;按学校性质标志,可以把数据划分为公办学校、民办学校等组类。这样的分组可以更好地说明不同类型数据的特点,有利于更好地通过样本数据发现总体的规律。利用具体实例引导学生感悟:通过数据分组,可以根据影响因素和结果因素的对应,更好地揭示现象之间的相互制约、相互依存关系;也可以反映总体内部各部分之间的差别,表明总体的内部结构;还可以计算各组所占总体的比重,从总体的构成上认识总体各部分的作用,从而对总体作出正确的预测。

2. 让学生适应信息化生态,形成大数据思维

大数据与小数据的根本区别在于大数据采用全样思维方式,小数据强调抽样。抽样是数据采集、存储、分析、呈现技术达不到实际要求或成本远超过承受范围的情况下采取的统计方法。从理论上讲,抽样得到的结论是不稳定的。比如,比较两位篮球运动员的投篮成绩,一位是2投1中,另一位是10投5中,尽管投中的频率都是50%,但是由于样本大小的明显差异,说两者有相同的投篮命中率是不严谨的。随着科学技术的发展,过去不可能获取全样数据、不可能存储和分析全样数据的情况都不复存在,大数据时代是全样数据的时代,是精准分析的时代。“数据分组的原则”的教学要让学生了解抽样作为统计的一种方式有一定的局限性,全样本统计的时代已经到来;要让学生养成获取大数据

的意识,掌握大数据分组的方法,了解分析大数据的工具,形成大数据的思维。

3. 激发学生探索分组准则的求知欲

组内离差平方和最小原则是最基本的数据分组原则。如果学生对这类问题感兴趣,可以适度介绍模型选择的其他准则。比如,AIC准则——赤池信息准则的简称,用于衡量统计模型的拟合度是否优良,以对多个模型作出选择;BIC准则——贝叶斯信息准则的简称,与AIC准则相似,也用于模型选择;CV准则——交叉验证法的简称,也是一种分类的统计分析方法,基本思想是将原始的数据集分为两个部分——训练集与验证集,先对训练集进行训练,再用验证集对训练的模型进行测试,进一步进行分类评价。^[4]也可以引导学生查阅关于AIC、BIC、CV准则的资料,或者介绍其他关于大数据分类方法的图书。由此,激发学生进一步探索分组准则的求知欲。

二、新增的“箱线图与百分位数”

(一) 内容解析

1. 箱线图的内涵及作用

箱线图又称为盒式图、盒状图,是一种常用的描述数据分布情况的统计图。它有5个特征数:最小值、下四分位数、中位数、上四分位数与最大值。如下页图1所示,在箱线图中,最上方和最下方的线段分别表示这组数据的最大值和最小值;箱子上方和下方的线段分别表示这组数据的上四分位数和下四分位数,中间的线段表示这组数据的中位数;最上方和最下方的圆圈表示数据中的异常值。箱线图一般有两种类型,即单式箱线图和复式箱线图。前者用于分析只有一个变量的数据分布,后者用以分析具有两个或以上变量的数据分布。

箱线图制作方便,常见于品质管理等报

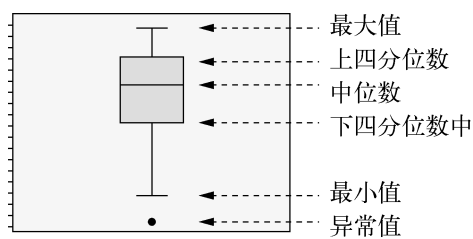


图 1

告中。它能够直观地呈现数据的离散分布情况(包括数据的对称性和数据中的异常值),直观地比较几组数据的分布特征,有利于数据的清洗。因此,新课标增加“箱线图与百分位数”的内容,进一步完善了初中生对统计图表的认识,有利于形成整体的知识结构。

2. 箱线图的制作步骤

第一步:把一组数据从小到大排列为 x_1, x_2, \dots, x_n ,找出这组数据的中位数,用 m_{50} 表示。 m_{50} 把这组数据分为前后(小大)两部分。

第二步:分别找出前后两部分数据的中位数,用 m_{25} 和 m_{75} 表示。 m_{25} 、 m_{50} 、 m_{75} 把这组数据分为个数相等的四个部分,因此被称为四分位数,其中 m_{75} 被称为上四分位数, m_{25} 被称为下四分位数。

第三步:绘制箱子的上下范围,上四分位数为上限,下四分位数为下限,在箱子内部中位数的位置绘制横线。

第四步:计算上四分位数和下四分位数的差值,即四分位数差,记为 δ 。将大于上四分位数 1.5δ 的值和小于下四分位数 1.5δ 的值划为异常值。超出 3δ 的异常值称为极端异常值,常用实心点表示;不到 3δ 的异常值称为温和异常值,常用空心点表示。

第五步:在异常值外,靠近最大值和最小值处划横线,作为箱线图的触须。

第六步:为箱线图添加名称、数轴。

3. 单个箱线图分析

中位数反映一组数据的集中趋势:中位

数高,表示平均水平较高;中位数低,表示平均水平较低。四分位数的差反映一组数据的离散情况:箱子短,表示数据集中;箱子长,表示数据分散。中位数在箱子的正中间时,数据一般呈正态分布(如图 2 所示);中位数在箱子的上半部分时,数据一般呈左偏分布(如图 3 所示);中位数在箱子的下半部分时,数据一般呈右偏分布(如图 4 所示)。

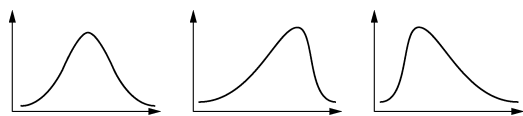


图 2

图 3

图 4

4. 多个箱线图对比分析

以某校九年级 8 个班级学生一次单元测试成绩的箱线图为例进行比较,结果如下页图 5 所示。

(二) 教学建议

1. 让学生通过实例制作,感悟箱线图的结构特征

弗赖登塔尔认为:“数学学习是一种活动,这种活动与游戏、骑自行车是一样的,不经过亲身体验,仅仅从看书本、听讲解、观察他人的演示,是学不会的。”“箱线图与百分位数”的教学要留出足够的时间,让学生在了解箱线图制作步骤的基础上制作箱线图,通过实例制作感悟箱线图的结构特征。从近几年江苏省八年级学生数学学业质量监测结果看,依据语言的描述画出图形是学生的薄弱环节。所以,要鼓励学生动手做,着力培养学生从语言(文字、数据)中提取信息、用图形表征信息的能力,逐步帮助学生形成“画图是分析问题的有效策略”的意识,体会到箱线图的价值。通过动手实践,箱线图的特征会越来越清晰、关系会越来越明白,学生的思考会越来越深入。当然,实际教学中,还要创设合理的信息化学习环境,探索利用信息技术(如

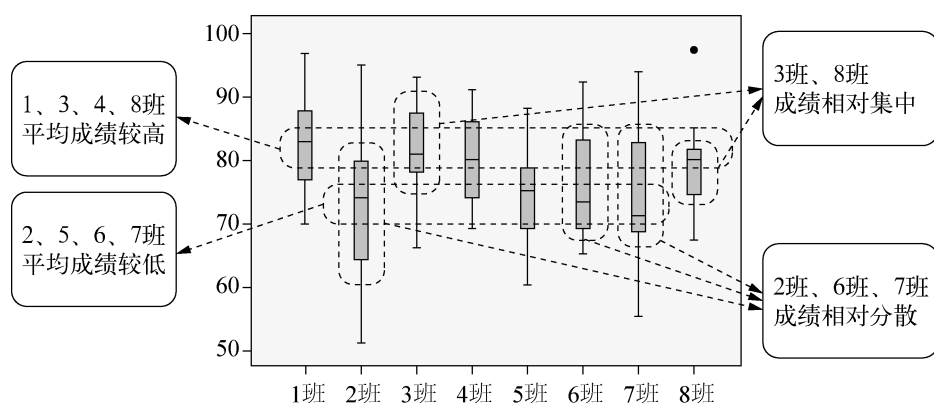


图 5

Excel、SPSS 等软件)制作箱线图,提升学生的探究热情,开阔学生的视野,激发学生的想象,提高学生的信息素养。

2. 让学生通过知识拓展,形成完整的认知结构

直方图、折线图和箱线图都可以用来描述单变量数据,当有两个或两个以上变量时,用什么样的图来描述数据呢?教学“箱线图与百分位数”内容后,引导学生思考这样的问题是很有意义的。一方面,这种思考问题的方式是“学过数学的人应该具备的基本思维特征”,这样的引导也是培养学生发现问题、提出问题能力的好办法;另一方面,由此介绍多变量统计图,有利于学生形成整体的统计图知识结构。实际上,学生对多变量统计图并不陌生:学校呈现学生学业成绩时就经常用雷达图(例如如图 6)来描述。无论从知识基础来看,还是从关键能力和思维品质来看,了解、认识甚至制作雷达图对初中生来说都是可行的,而且贴近他们的学习生活,能够激发他们的探究兴趣,是有意义、有价值的拓展。

当然,实际教学中,也要引导学生比较分析各种统计图的优势和缺点,在形成完整知识结构的同时,提升合理运用各种统计图表整理数据、分析数据的能力。

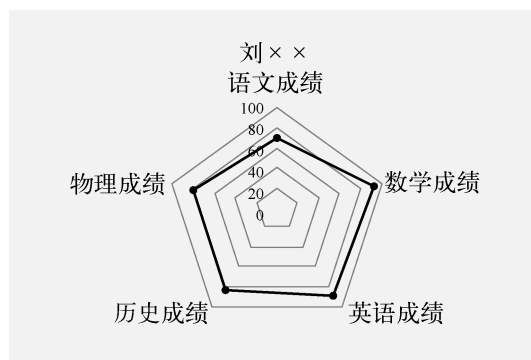


图 6

参考文献:

- [1][2][3] 中华人民共和国教育部. 义务教育数学课程标准(2022 年版)[S]. 北京:北京师范大学出版社,2022:74,74,162-163.
- [4] 王俊艳. 基于 AIC、BIC、CV 准则的模型选择[J]. 统计学与应用,2020(4):546-565.