

统计活动设计及活动分析

统计是一门特色鲜明的学科,是数学的一个分支。从日常生活到高新科技领域,都与统计有着密切的联系。

新课程标准提出了高中数学要重点培养学生的六大核心素养,其中之一就是数据分析素养。数据分析是指针对研究对象获取数据,运用数学方法对数据进行整理、分析和推断,形成关于研究对象的知识的素养。数据分析是研究随机现象的重要数学技术,是大数据时代数学应用的主要方法,也是“互联网+”相关领域的主要数学方法。数据分析已经深入科学、技术、工程和现代社会生活的各个方面^[1]。

系统、深度的统计学习活动中是培养学生数据分析素养的重要途径。令人遗憾的是,统计学科的深度没有有效地开展起来。因为教材中的涉及统计内容的例题、习题和高考中的相应考题比较简单,所以统计教学活动基本停留在概念学习和基本统计量运算的浅层学习上。诚如张留芳和张玉环在高中概率与统计教学现状的调查研究中发现:有45.3%的教师认为统计教学内容安排“受高考的压力,只能考什么就教什么,怎么考就怎么教”,实际授课时数低于课标的要求^[2]。而大多数其他章节教学实际课时是超出课标建议授课课时的。这样的教学方式不仅造成了学生统计知识的匮乏和理解片面,还严重偏离了课标要求:“统计学的教学活动应通过典型案例进行。教学中应通过对一些典型案例的处理,使学生经历较为系统的数据处理全过程,在此过程中学习数据分析的方法,理解数据分析的思路,运用所学知识和方法解决实际问题。”针对上述情况,我们设计了人教B版必修(第二册)第五章第一节统计的教学活动,让学生在典型案例的处理中学习知识,体验统计的基本方法和思想,培养学生数学学科的核心素养。



1.1 活动设计目标

对于人教 B 版必修(第二册)第五章第一节统计,课标要求有四点:获取数据的基本途径及相关概念、抽样、统计图表、用样本估计总体。在传统的统计教学中,教学重点主要集中在从形式上介绍简单随机抽样、分层抽样和利用信息技术进行抽样这三种不同的抽样方法,以及这三种方法在形式上的不同点,通过具体数学问题认知统计图表和掌握统计量的运算。传统的统计教学缺少了学生的实际运用和体验,忽视了引导学生对随机性的理解,只是使学生停留在对一些简单概念的理解和记忆层面上,无法深度学习统计思想的具体应用和实践价值。因此,这次统计活动设计的目标,是把教学重点集中在学生对统计方法中的随机思想的体验和理解上,让学生通过统计实践提高其对统计思想的掌握和运用。

1.2 活动设计方案

活动设计分三个步骤:第一个步骤是教师带领学生利用 Excel 进行抽样的模拟实验;第二个步骤是学生分小组进行数据收集、整理和分析工作;第三个步骤是各小组的汇报分享及交流反思,在这一过程中使学生能够达到对统计思想的正确理解。

首先利用 Excel 的随机功能让学生对随机抽样有一个清晰的认识。为此,我们把某一次大型考试的 9000 多名考生成绩利用 Excel 的随机功能做抽样分析,通过数据演示让学生对抽样中的样本容量与抽样结果的准确性有一个比较理性的认识。

在实验开始前,先让学生根据其已有的知识和对抽样方法的理解,讨论样本容量分别为 50,100,200,800 时,所获取的抽样数据与数据真值的关系。经过比较和分析之后,学生们一致认为,样本容量越大,样本数据越接近真值。随后,利用 Excel 的随机功能,分别对样本容量为 50,100,200,800 的样本进行抽样统计,通过数十次的实验,发现这四种样本的数据值并没有显著差异。这样一来,学生就会直观并真切地感受到:只要样本容量达到一定数量,并且统计方法正确,统计数据与真值差异就很小。

这一过程不仅让学生真正感受到了统计中随机思想的强大,也为学生调研自己感兴趣的社会现象提供了科学可行、恰当有效的统计方法。

在此步骤之后,让学生分组进行统计活动,从设计数据收集方案开始,再进行数据整理和计算,进而分析数据,到最后得出结论,都由小组同学合作完成。

1.3 学生统计活动的实施展示

学生们自由组合了8个小组,每个小组有5名左右的成员。下面是各小组的数据收集方法。

A组身高统计的方法是在上课间操时,从每个班级第一个人开始,5人为一个单位,取每个单位中间的人记录身高,即记录第3,8,13,⋯名同学的身高(系统抽样),并将男女生分开(分层抽样)。本组的出发点是为了确保调查的同学有各个身高段的人,从而进一步确保最后所得平均身高的数据与真实数据更加接近。为了保持数据的真实性,他们选择去掉一个班(体育特长班,半数篮球运动员)同学的数据,因为他们的平均身高远远高于其他班级的同学。

B组以每班人数的个位数字作为所抽学号的个位数字,每班抽取6人,调查他们的身高和他们父母的身高,数据精确到厘米。然后分开统计男女生的数据,使用Excel中的数据处理和制作统计图的功能,求出所需数据。

C组在每个班获取样本时是按照人数比例和男女生比例抽查,为了减小误差,根据目测选择有代表性的同学,而不是都选高的或都选矮的。他们认为这样做可能会更具有代表性。

D组抽样方法:全年级人数大约有500人,12个班,抽取其中约20%的数据,每个班大约抽取 $500 \times 20\% \div 12 \approx 8$ (人),运用随机数表法采取简单随机抽样。

E组按照学号来随机抽取,抽取每班学号为7,10,12,18,21和22的同学。样本为所抽取的72名同学的身高,共调查男生36人,女生36人。

F组以班为单位,1~8,11~12班人数差不多,9,10班人数明显和其他班不一样,所以进行分层抽样,每层的概率是 $1/8$ 。1~8,11~12班为一层,每班按40人算,随机剔除多余的人。在这一层中,以班为单位,再分成10个组,进行抽样。获取数据后,小组考虑到各班实际人数不同,又对各班的数据按各班实际人数进行加权处理,得到年级的平均身高为171.1cm。

G组第一步通过数字校园得到各班人数;第二步将所有的人编号;第三步将所有编号输入Excel,并用Excel的统计功能随机抽出72个编号;第四步将编号对应到具体的同学,然后进到各班记录编号对应的同学的身高。

H组采取每个小组成员负责两个班,各自对所负责的班级数据抽样调查,然后汇总各个成员的数据。在汇总中他们发现每位组员所选的样本数量都不相同,甚



至有人直接调查了所负责两个班的全体同学,考虑到样本数量不同,他们只是各自分析了所负责班级的身高情况。

最终,各小组的统计结果如下表所示。

| 小 组 | A 组 | B 组 | C 组 | D 组 | E 组 | F 组 | G 组 | H 组 |
|-------------|-------|-------|-------|-------|-------|-------|-------|-----|
| 男生平均身高统计/cm | 178.3 | 174.6 | 177.7 | 176.3 | 177.4 | | 177.7 | |
| 女生平均身高统计/cm | 163.6 | 166.9 | 168.8 | 165.5 | 165.5 | | 166.1 | |
| 年级平均身高统计/cm | 171.5 | 171.1 | 172.8 | 171.1 | 171.5 | 171.1 | 172.1 | |

在学生完成了初步的设计,得出本小组的结论之后,进行活动的第三步,即在课堂上,各个小组通过统计图表展示各组调查结果,并对本组的活动进行评价。各组都认真分析了统计活动误差产生的原因,如 A 组分析数据可能产生误差的原因:首先是因为没有加入 8,9 班同学身高而出现误差(8 班是学生自己所在的班级);其次是因为他们是在按身高排序的队伍中进行的分层抽样,所以取得数据的随机性较差。D 组认为本次抽样统计存在一定误差的原因:首先是由于人不全等原因,无法完全使用简单随机抽样;其次是体育特长班(篮球运动员多,身高要明显高于其他班级)的数据不全,导致平均身高偏低;最后是感觉应该按班级人数及男女比例分层算加权平均数。

接着各组同学和老师一起对抽样方法和结果进行了充分交流和讨论,很快就形成了以下三点共识。

(1) 获取数据要尽可能满足随机性,如对于 C 组的方案,很多学生提出,他们通过观察选择中等身材的学生进行调查,主观性太强,而且在面对同学的时候经常会有倾向性,如选择自己认识的同学做调查。A 组也承认,在排好高矮顺序之后再选择样本,随机性也会很差。

(2) 三种随机抽样并不是完全独立的,在本次活动中可以同时使用,也可以独立使用。

(3) 调查对象有明显差异时要进行分层抽样,很多组在活动中采用的是随机抽样,并没有区分男女生进行分层抽样,G 组在数据分析时发现他们抽出样本的男女比例有一定的差异,而年级男女生比例非常相近,但是男女生的身高是有着显著差异的,这造成了他们的年级平均身高偏大(样本数据中男生较多)。事实上,男生和女生身高的统计值的确是 G 组最接近年级男生和女生的身高实际值,全年级的平均身高统计值也确实与实际值有着非常明显的差异。

1.4 活动分析

显而易见,上面的教学活动,充分调动了学生学习统计的积极性。不仅完成了课标对统计教学的要求,还让学生充分明白了统计方法和思想在数学及实际应用中的重要价值。学生通过活动真切地感受到了抽样方法对数据的影响,为将来统计学习和应用打好基础。学生们在活动中充分利用了“统计图表”,真切体会到直方图、折线图和茎叶图等图表的作用。学生的活动也为后续统计的学习打下了良好基础。有了调查数据,学生非常关心孩子的身高与父母身高是否有相关性,几乎所有小组都自学了相关内容,主动分析调查数据的相关性,对数据相关性有了深切感受和更加直观的认识。其中一个小组结合相关性和自己调查的数据,利用 Excel 的统计功能,创造性地提出子女身高计算公式 $z = ax + (1-a)y$, 其中 z 为子女身高, x 为父亲身高, y 为母亲身高。通过他们的调查数据算出公式中男生对应的 $a = 1.50$, 女生 $a = 0.2$, 他们对这个发现非常兴奋。在本节学完后,还有一些同学利用相关性分析,对校园生活中遇到的问题进行了进一步探究,并取得了很好的成果。

总而言之,传统的统计教学重点是在计算数据的特征值(也就是计算平均数、中位数和方差等),用样本的数字特征来估计总体特征。学习列频率分布表、画频率分布直方图和频率折线图等方法也只是从中获取计算的数据,并没有真正应用这些图表的直观性分析数据特征。实际研究过程恰好与之相反,通过实验、查阅资料和设计调查问卷等方法收集数据,然后将数据分类整理,并根据研究问题的需要绘制图表。如果学生通过独立思考,积极灵活地运用所学的随机抽样方法进行调查,自主学习统计软件去做统计分析,那么他们所学到的知识就已经不仅仅是简单的数学上的逻辑推理和计算了。这也让我们真切地感受到,数学老师教给学生的不应该仅仅是数学概念、逻辑思维和解题方法,更应该注重学生数学思想的构建,特别是数学思想在生活中的具体应用,这样可以使学生在数学学习中融入自己的思想,灵活运用数学知识、数学思维解决各种问题,提高学生的综合素质。

1.5 结语

由此可见,想引导学生进行深度学习,要重点关注以下几个方面。

(1) 学生的学习要以教材为基准。教材是学生在过程中接受知识的源泉。对于数学知识的深度学习,学生要从课本出发,通过生动的情境和活动进行拓



展,了解本章将要学习的内容,做到心里有数。

(2) 通过对概念、定理等知识的理解,学习必要的数学知识,挖掘其中所包含的数学思想。从课本给出的例子中体会所学知识的运用,做到学以致用。通过课后解题检验学生对基本知识、基本技能和基本思想的掌握情况,能够进行自我检测。

(3) 力求使学生体验数学在解决实际问题中的作用、数学与日常生活及其他学科的联系,促进学生逐步形成和发展数学应用意识,提高学生数学学科素养。如果深度学习能够进入我们常态化的数学教学中,相信学生的学习能力和创造力一定能够被激发出来,我们培养的拔尖创新人才也一定能够达到更高的水平并具有更强的能力。

参考文献

- [1] 中华人民共和国教育部. 普通高中数学课程标准[M]. 北京: 人民教育出版社, 2020.
- [2] 张留芳, 张玉环. 高中概率与统计教学现状的调查研究[J]. 中国数学教育, 高中版, 2018(10): 6.



统计探究活动1——从午饭里省出时间

2.1 问题提出

中午总是需要花很长的时间打午饭,打完饭后又需要在大厅里找位置吃饭,排队占据了中午大量时间,有什么方法可以提高效率,从午饭里省出时间?

2.2 现状及分析

造成学校食堂排长龙的因素很多,有食堂窗口数与师生数之比、食堂师傅打饭的速度、特殊活动和特殊天气等影响因素。其中影响最大的因素还是人太多,师生都集中在上午下课后的某一段时间内进入食堂。因此想解决这个问题要聚焦到人的集中程度,用食堂中所含的学生数代表拥挤程度,然后这个问题就变成了关于食堂中人数与时间关系的讨论。首先对进入食堂的人数进行统计,下面是分时段、分年级出入食堂人数的统计结果。

| 时间 | 进入食堂的人数 | | | | | | 总计 | 离开食堂的人数 |
|-------|---------|----|----|----|----|----|----|---------|
| | 初一 | 初二 | 初三 | 高一 | 高二 | 高三 | | |
| 12:03 | 8 | 7 | 8 | 2 | 25 | 1 | 51 | 0 |
| 12:04 | 22 | 8 | 5 | 4 | 16 | 0 | 55 | 0 |
| 12:05 | 23 | 15 | 10 | 6 | 4 | 5 | 63 | 0 |
| 12:06 | 30 | 12 | 9 | 6 | 8 | 2 | 67 | 0 |
| 12:07 | 16 | 9 | 12 | 14 | 18 | 10 | 79 | -1 |
| 12:08 | 13 | 11 | 12 | 2 | 21 | 4 | 63 | 0 |

续表

| 时间 | 进入食堂的人数 | | | | | | 总计 | 离开食堂的人数 |
|-------|---------|----|----|----|----|----|----|---------|
| | 初一 | 初二 | 初三 | 高一 | 高二 | 高三 | | |
| 12:09 | 1 | 13 | 9 | 8 | 12 | 11 | 54 | 0 |
| 12:10 | 5 | 12 | 7 | 9 | 0 | 8 | 41 | 0 |
| 12:11 | 10 | 6 | 12 | 12 | 2 | 2 | 44 | 0 |
| 12:12 | 2 | 9 | 8 | 13 | 6 | 5 | 43 | 0 |
| 12:13 | 4 | 5 | 9 | 0 | 4 | 1 | 23 | -7 |
| 12:14 | 0 | 4 | 4 | 4 | 0 | 0 | 12 | -9 |
| 12:15 | 2 | 3 | 0 | 3 | 4 | 2 | 14 | -18 |
| 12:16 | 2 | 3 | 3 | 1 | 0 | 5 | 14 | -22 |
| 12:17 | 1 | 2 | 3 | 3 | 2 | 7 | 18 | -27 |
| 12:18 | 0 | 0 | 0 | 0 | 2 | 0 | 2 | -18 |
| 12:19 | 0 | 0 | 2 | 0 | 0 | 2 | 4 | -24 |
| 12:20 | 5 | 0 | 1 | 4 | 2 | 14 | 26 | -12 |
| 12:21 | 0 | 6 | 2 | 0 | 0 | 7 | 15 | -28 |
| 12:22 | 1 | 2 | 3 | 2 | 1 | 6 | 15 | -39 |
| 12:23 | 0 | 8 | 2 | 0 | 2 | 3 | 15 | -12 |
| 12:24 | 4 | 1 | 0 | 3 | 0 | 12 | 20 | -20 |
| 12:25 | 2 | 2 | 11 | 0 | 3 | 0 | 18 | -22 |
| 12:26 | 0 | 0 | 5 | 2 | 3 | 5 | 15 | -24 |
| 12:27 | 1 | 4 | 0 | 2 | 0 | 4 | 11 | -40 |
| 12:28 | 1 | 2 | 0 | 0 | 2 | 6 | 11 | -21 |
| 12:29 | 0 | 4 | 2 | 5 | 2 | 8 | 21 | -20 |

运用 Excel 对上述数据进行处理,可得到图 1。

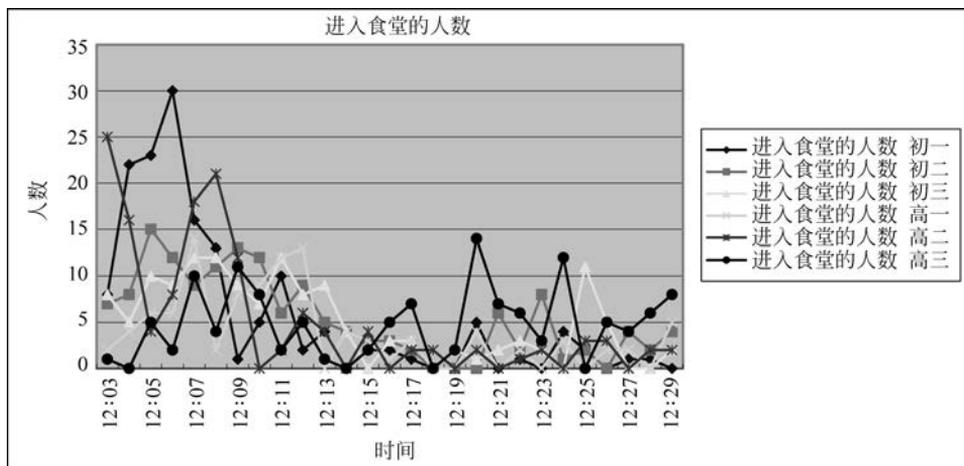


图 1

把学生总数合在一起得到图 2。



图 2

对上面两张图表进行分析,可以明显地看出,进入食堂的人数高峰时期为 12:05—12:08。在这 4 分钟内进入食堂共 272 人,占 12:29(含)之前进入食堂的总人数的 $272 \div 814 \times 100\% \approx 33.415\%$,而这 4 分钟仅为全部统计时间的 14.8%。再进一步分析,得到这段高峰期进入的学生数比例:

初一年级 $82 \div 153 \times 100\% \approx 53.595\%$;

初二年级 $47 \div 148 \times 100\% \approx 31.757\%$;

初三年级 $43 \div 139 \times 100\% \approx 30.935\%$;

高一年级 $28 \div 105 \times 100\% \approx 26.667\%$;

高二年级 $51 \div 139 \times 100\% \approx 36.691\%$;

高三年级 $21 \div 130 \times 100\% \approx 16.154\%$ 。

即图 3。

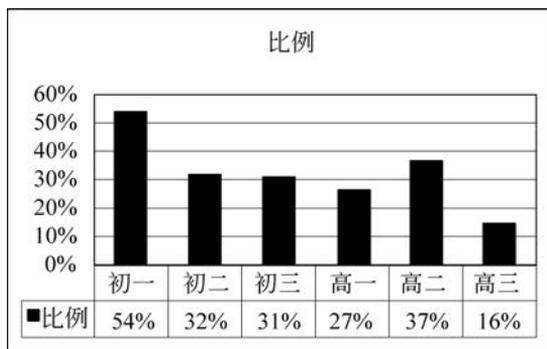


图 3

可以看出,进入食堂的学生数具有明显峰值特点,即短时间内有大量学生同时进入。这点对于初一年级尤为明显,超过半数的学生是在这4分钟内进入食堂的。而较高年级的学生这种情况就略微好一些。这很可能是因为初一年级的学生习惯于一下课就到食堂就餐,而高年级中的一小部分学生由于有了前几年的经验,具备避开高峰的意识,因而使本年级的峰值有所降低。但是,这对于食堂整体的拥挤情况并没有太大作用。

现在我们再来看看离开食堂的人数,见图4。

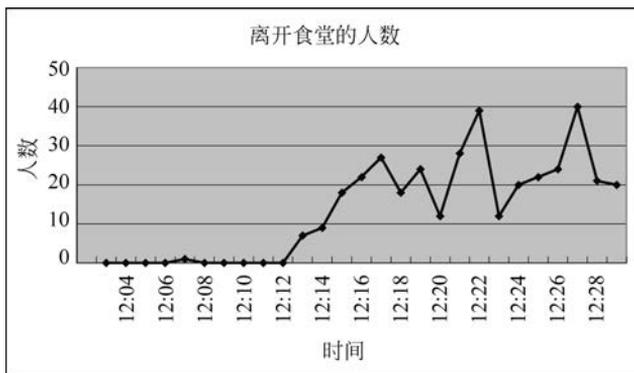


图 4

然后把进入食堂的人数合在一起,就得到了食堂内的学生人数,如图5所示。

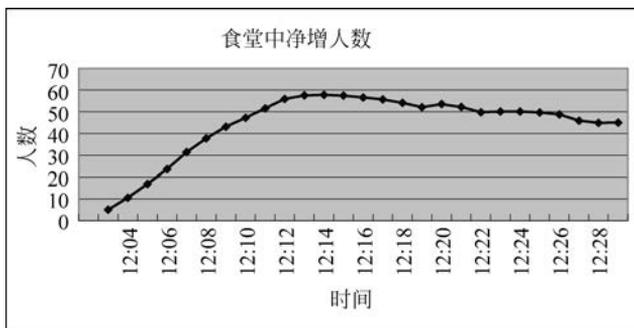


图 5

由此可以看出,单独避开进入食堂的人流高峰还是不够的,由于打饭速度和吃饭速度等因素的影响,食堂的拥挤度即使在12:09后还会持续相当一段时间。那么究竟什么时候食堂拥挤度会显著降低?