

设计探究问题链 构建统计推断方法

——以“独立性检验”的教学设计为例

唐 岚(浙江省杭州学军中学)

摘 要:由于样本的随机性,统计推断的结论只能评价好坏。新教材既要让学生认识到独立性检验的合理性,又要让学生体会到其结果的不确定性。设计探究问题链,让学生在独立性检验的学习中,重视统计思维的培养,进一步领悟其本质,促进数学抽象、逻辑推理、数学建模等学科核心素养的提升。

关键词:独立性检验;统计推断;探究问题链

文章编号:1002-2171(2022)9-0019-03

1 问题提出

2022年4月20日,杭州市教研室组织全市展开高二概率统计教研活动,笔者开设了一节研讨课,内容是人教A版《数学》(选择性必修三)第八章“成对数据分析”8.3节“列联表与独立性检验”的第二课时。这部分是新教材新要求的统计知识,其蕴含的统计方法并不容易理解。高中阶段的数学大多是确定性教学,即条件确定、结论确定、结果对错确定。但统计推断,由于样本的随机性,其结论只能评价好坏。新教材既要让学生认识到方法的合理性,又要其体会到结果的不确定性。

构建统计推断方法的问题串不是简单的问题组合,它应该是教师围绕统计相关思想,从学生的基础经验和能力出发,按照特定的教学逻辑设计将问题串融入探究活动中,能有效帮助学生培养统计思想,全方位地经历统计推断方法的构建。下面笔者就以“独立性检验”为例,设计核心素养导向下采用探究问题链的教学。

2 教学分析

2.1 内容分析

独立性检验是研究随机变量独立性的一种统计方法。为了解两分类变量是否相互独立,可从总体中抽取简单随机样本,整理成 2×2 列联表,用独立性检验的方法检验两分类变量是否相互独立。独立性检验本质上是一种假设检验,是一种“概率反证法”,通过样本事件是否属于小概率事件来判断总体假设的真伪。

独立性检验是成对样本数据统计分析的重要内容,是依据数据进行合理推理的典型方法,是体现统计思维与确定性思维的差异、归纳推断与演绎证明的

差异的主要素材,也是积累数据分析经验,培养数据分析、数学建模、逻辑推理等素养的媒介。

2.2 学情分析

学生经过古典概型、条件概率、正态分布等知识的学习,已经初步掌握了基本的统计思想方法。但本节内容是学生遇到为数不多的统计推断问题,需要以学生现有的知识为基础,借助类比、抽象、建模等思想方法,实现从逻辑推理能力向概率推理能力的提升。

2.3 教学目标

(1)能掌握假设检验的基本步骤。

(2)能说明 χ^2 统计量构造的合理性;能说出基于小概率原则的独立性检验的基本思想。

(3)能说出独立性检验的基本步骤,并能用独立性检验方法解决简单的实际问题。

2.4 教学重点与难点

教学重点:独立性检验的思想方法,利用独立性检验解决实际问题。

教学难点:正确理解小概率原则,理解样本随机性,培养学生统计推断的能力。

3 教学过程

3.1 情境设置,新课自然引入

探究1:女士品茶。

某聚会上,张三说:“我能尝出奶茶制作中是先加奶还是先加茶。”很多人觉得不可思议。张三就说:“你们不信,我们可进行实验。”于是人们准备了10杯奶茶,未告知张三先加奶还是茶。结果张三说对了9杯。你觉得张三有超能力吗?你是如何判断的?

设计意图:从张三品茶情境出发,为课堂增添趣味性。新教材特别重视情境的创设和问题的提出,注重从实际情境出发,抽象概括出数学问题。张三品茶

具有实际情境的铺垫,引导学生思考如何利用统计推断进行合理怀疑。

问题1:假设张三是随机猜测的,即张三说中的概率为 $\frac{1}{2}$,事件“10杯奶茶说对9杯”发生的概率有多少?

问题2:在假设前提下,计算得到事件“10杯奶茶说对9杯”发生的概率为0.0097,请你说说此时能不能推断假设是错的?

问题3:我们的判断是否一定正确,会不会是张三运气好猜中9杯,而非具有超能力?

问题4:请你总结假设检验的一般步骤。

设计意图:通过问题1,让学生了解假设检验是进行科学怀疑的有效手段。通过问题2,让学生理解假设检验的基本思想是小概率事件。通过问题3,突破本节课的难点,假设检验结果并不是一定正确的,体现统计思维与确定性思维的差异。进而用问题4结合张三品茶的实例,得到假设检验的一般流程。

问题5:假设检验与之前学习的什么方法比较相似?

问题6:假设检验的思想与我们之前常用的反证法有异曲同工之妙,能不能总结一下反证法和假设检验的相同之处。

3.2 类比探究,统计量自然构建

探究2:提出零假设。

上节课我们已经知道两分类变量无关联的条件概率表示为 $P(Y=1|X=0)=P(Y=1|X=1)$ 。请阅读教材,提出两分类变量无关联的零假设。

设计意图:类比假设检验得到独立性检验的步骤,培养学生的类比迁移能力。用概率语言,经过严格的推理,将零假设中分类变量的关联性转化为分类变量的独立性,为后续构造卡方统计量打下基础。

探究3:构建统计量。

在零假设 H_0 :分类变量 X 与 Y 独立前提下,随机抽样所产生的列联表是小概率事件,则拒绝 H_0 。如何构造一个刻画列联表的统计量?

问题7:

(1)算一算:根据列联表(表1),完成概率表2、概率表3中统计的真实值。

表1 列联表

X	Y		合计
	Y=0	Y=1	
X=0	a	b	a+b
X=1	c	d	c+d
合计	a+c	b+d	a+b+c+d

(2)说一说:在零假设 H_0 设前提下,即 $P(X=0, Y=0)=P(X=0)P(Y=0)$, $P(X=0, Y=1)=P(X=0)P(Y=1)$, $P(X=1, Y=0)=P(X=1)P(Y=0)$, $P(X=1, Y=1)=P(X=1)P(Y=1)$ 。完成概率表3中的假设期望值。

表2 概率表

概率	值
$P(X=0)$	
$P(X=1)$	
$P(Y=0)$	
$P(Y=1)$	

表3 概率表

概率	真实值	假设期望值
$P(X=0, Y=0)$		
$P(X=0, Y=1)$		
$P(X=1, Y=0)$		
$P(X=1, Y=1)$		

(3)填一填:利用概率表3中的概率假设期望值,将列联表(表4)补充完整,将此类数据称为期望值。

表4 列联表

X	Y		合计
	Y=0	Y=1	
X=0			a+b
X=1			c+d
合计	a+c	b+d	a+b+c+d

(4)想一想:两列联表中观测值和期望值是否相等?之前学习过程中有没有出现过比较观测值和期望值去解释模型的知识?

(5)试一试:你们能否利用观测值和期望值偏离程度构造一个刻画样本列联表的统计量?

(6)推一推:与残差平方和类似,独立性检验有专门的统计量,构造方法稍显不同,我们称为卡方统计量,即卡方分布可以用来检验模型的适合性、变量的独立性等。假设前提下得到统计量,假设数据服从某一分布、某一模型,然后计算出数据应该有的样子,叫期望数据。对象是类别变量,卡方的检验公式是: $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$, O_i 是观察数据, E_i 是期望数据。请推导 2×2 列联表的卡方统计量。

设计意图:学生目前只遇到过概率统计量,如何根据列联表构造新的统计量是本节课的一大难点。探究3将构建统计量先分割成算一算、说一说、填一填、想一想、试一试、推一推共六个小问题,立足卡方统计量的本质进行研究,引导学生发现卡方统计量其实就是将“现实与理论不相符就拒绝理论”这一朴素的思想数学化。引入观测值和期望值的概念,与线性回归联系,便于学生联想到用残差去构造统计量,从

而逐步将统计量公式表示出来。小组活动贯穿整个教学过程,让学生充分体验到数学的严谨性,也让他们由浅入深地经历知识的形成过程。卡方分布其实就是将现实观察和理论不相符就拒绝理论这一朴素的思想数学化了。 2×2 列联表的卡方统计量可化简

$$\chi^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}.$$

3.3 借助技术,难点自然突破

探究4:小概率事件的定义。

利用GGB语言的卡方分布概率密度函数,结合图像解释如何根据卡方统计量的值确定是否为小概率事件。

问题8:回顾正态分布,概率密度函数有什么意义?如何计算 $P(x > \mu + 3\sigma)$ 。

问题9:(1) $\forall \alpha \in (0, 1)$,是否可以找到唯一的 x_α ,使得 $P(X > x_\alpha) = \alpha$?例如, $\alpha = 0.1$ 时。

(2)以小概率值 $\alpha = 0.1$ 为标准,若计算得 $\chi^2 = 2.5$,能否说在零假设前提下发生的这件事是小概率事件?

(3)以小概率值 $\alpha = 0.1$ 为标准,若计算得 $\chi^2 = 3.2$,能否说在零假设前提下发生的这件事是小概率事件?

(4)以小概率值 $\alpha = 0.05$ 为标准,若计算得 $\chi^2 = 3.2$,能否说在零假设前提下发生的这件事是小概率事件?

设计意图:探究4,教师根据GGB绘制的卡方分布概率密度函数,利用问题串逐次递进地研究,恰当地运用信息技术使学生直观地感知卡方分布小概率原则、确定临界值等;渗透统计借助信息技术、以数释形的研究策略,有利于提升学生数理统计分析能力,促进他们统计思维的发展,提高学习统计和应用统计的兴趣。

问题10:(1)你能总结基于小概率值 α 的检验规则吗?

(2)请你总结独立性检验的一般步骤。

设计意图:让学生用概率语言描述独立性检验判断标准,体验有理有据地表达见解的过程,认识到独立性检验也是有可能犯错的,但能将犯错的概率控制在小概率范围内。类比已有知识结构对新知识进行网状构建,逐步使新知识融入已有知识体系中。

3.4 学以致用,体现育人价值

探究5:独立性检验的应用。

例1 依据小概率值 $\alpha = 0.1$ 的 χ^2 独立性检验,分析以下抽样数据(表5),能否据此推断两校学生的数学成绩优秀率有差异。

表5 抽样数据

学校	数学成绩		合计
	不优秀 $Y=0$	优秀 $Y=1$	
甲校 $X=0$	33	10	43
乙校 $X=1$	38	7	45
合计	71	17	88

问题11:利用条件概率和等高堆积条形图,你会得到什么结论?

变式:基于例1的总体,再次随机抽样,抽取样本容量为880的样本,如表6。在相同的检验标准下,用独立性检验推断学校与数学成绩之间的关联性,结论还一样吗?请你解释原因。

表6 样本容量

学校	数学成绩		合计
	不优秀 $Y=0$	优秀 $Y=1$	
甲校 $X=0$	330	100	430
乙校 $X=1$	380	70	450
合计	710	170	880

问题12:(1)为什么独立性检验结果与例1不一致?

(2)有学生已经发现了表格中的数据不止样本总体扩大了十倍,其他数据也扩大了十倍。这里教师没有进行随机抽样,而是将数据简单扩大了十倍,相当于“学术造假”,你还愿意相信这次的结果吗?

4 教学思考

4.1 直观感知,发挥探究问题链的情境意义

设计探究问题链,创造情境,能更好地衔接学生已有的经验,更好地激发学生主动探索的兴趣。本节课,探究1“张三品茶”为学生理解和认识假设检验提供了平台,体验假设检验的构建过程;探究问题链5体验独立性检验的应用,理解统计基于数据真实性的特点。直观感知的探究问题链给予学生从特殊到一般的情境支持,极大程度地激发了学生的研究兴趣,同时彰显了统计研究的价值。

4.2 层层递进,凸显探究问题链的引导作用

探究问题链的教学展示了化整为零的思想,即通过小步骤完成阶段性目标,由浅入深、逐层递进的处理方式能有效降低学习的难度。本节课中,探究3“构建统计量”让学生经历了第一次统计量的构建,体会观测值与期望值的差异;探究4借助信息技术引导学生发现卡方分布小概率的所在位置,真正获得对独立性检验的深度理解。探究问题链并非是提出问题让学生浅显的总结,而是通过问题链让学生站在更高层次逐步击破统计推断知识体系中的难点。

4.3 激发潜能,发挥探究问题链的育人价值

探究5让学生真正意识到统计要基于数据真实的基础,如果数据是伪造和错误的,不管结果如何都不具有可信度,这也体现了数学的育人价值,促进学生数学素养的提升和道德品质的升华。