突破概率统计解题障碍:审题、计算与模型识别

张隆亿
(福建省永春第一中学)

概率统计解答题考查特点鲜明且“统计”的味道浓,考查内容包括随机变量的分布列及期望、正态分布、线性回归、独立性检验等.题目类型或是“说理说事”,或是“决策问题”,或是与函数等结合,形式新颖多样,既充分体现出概率统计是为生产、生活相关决策提供帮助的重要工具,又着重考查学生的数学建模能力、计算求解能力、创新能力等.这类试题的难度可控,受到命题者的青睐,然而却有相当一部分学生的成绩并不理想,因此本文梳理了破解概率统计中解题障碍点的方法.

1 提升数学建模能力

数学建模能力是指发现问题、提出问题、分析问题、建立模型、求解模型、测试结果、改进模型的一系列能力,通过把数学中的实际问题抽象出来,用数学语言表达,并建立数学模型来解决问题.数据分析是数学建模的重要内容,包括数据的收集、分析和处理,通过提取信息,构建模型和推理得到结论.

实测分析表明:数据分析能力低,是导致学生得分率低的根本原因.在面对篇幅长、信息量大的试题时,学生难以在短时间内把握解题的关键点.因此,提高数学建模能力是解决这一问题的关键.为提升数学建模能力,一般建议学生从问题入手,先确定求解目标,再分析解决问题所需的基本要素,然后在审题时快速阅读试题背景等信息,着重寻找解题的要素,以便能够快速入题,突破概率统计问题的阅读障碍.

对于回归分析题型,一般要明确具有线性关系的变量,如 x,y,再结合回归方程 $\hat{y}=\hat{b}x+\hat{a}$ 的斜率公式

$$\hat{b}=\frac{\sum\limits_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}=\frac{\sum\limits_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sum\limits_{i=1}^{n}x_i^2-n\bar{x}^2}$$

找出或计算出

$$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}),\ \sum_{i=1}^{n}(x_i-\bar{x})^2,$$

或

$$\sum_{i=1}^{n}x_iy_i,\sum_{i=1}^{n}x_i^2,\bar{x},\bar{y},$$

进而正确求解 $\hat{a}$ 和 $\hat{b}$.对于非线性回归题型,通常转化为线性回归,然后用线性回归方法处理.

对于独立性检验题型,一般要抓住两个分类变量,根据每一类的数量,列 2×2 列联表,并结合公式 $K^2=\dfrac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$ 求解 K^2 的观测值,进而完整表达统计结论.

例 1 某疫苗进行安全性临床试验.该疫苗安全性的一个重要指标是:注射疫苗后人体血液中的高铁血红蛋白(MetHb)的含量(简称“M 含量”).若 M 含量不超过 1%,则为阴性,认为受试者没有出现高铁血红蛋白血症(简称“血症”);若 M 含量超过 1%,则为阳性,认为受试者出现血症.若同时满足一批受试者的 M 含量平均数不超过 0.65%,出现血症的被测试者的比例不超过 5%这两个条件,则认为该疫苗在 M 含量指标上是“安全的”;否则认为“不安全”.现有接受了该疫苗注射的男、女志愿者各 400 名,经数据整理,制得频率分布直方图,如图 1 所示.(注:同一组的数据用区间中点值表示)

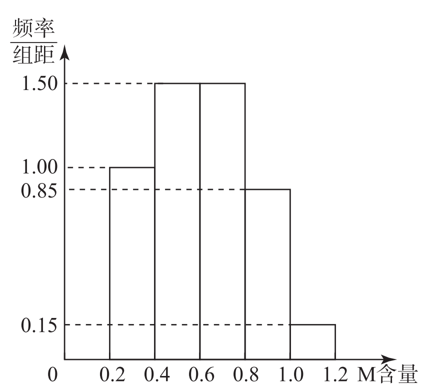

图 1

(1)估算这 800 名志愿者 M 含量平均数;

(2)按照性别分层随机抽样随机抽取 100 名志愿者,其中恰有 2 名男性志愿者检测阳性.用样本估计总体的思想,完成这 800 名志愿者的 2×2 列联表,并判

断是否有超过95%的把握认为血症与性别有关?

分析 解题的关键是通过问题:"是否有超过95%的把握认为血症与性别有关",抓住两个分类变量:血症和性别.易知这800名志愿者中阳性的频率为0.03,则阳性的人数共有$800\times0.03=24$人,所以男性志愿者中阳性的人数为$400\times\frac{2}{50}=16$,阴性的人数$400-16=384$;女性志愿者中阳性的人数为$24-16=8$,阴性的人数$400-8=392$,并制作2×2列联表.结合临界值表,根据$K^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$求解得观测值$\frac{800\times(16\times392-8\times384)^2}{400\times400\times24\times776}\approx2.749<3.841$,据此得出结论:由参考数据可得,没有超过95%的把握认为注射疫苗后血症与性别有关.

2 提升创新能力

创新能力是指结合日常生活、其他学科和学习实践的素材,发现并提出问题,以及能够灵活运用所学的数学知识和思维方法,独立思考、探索和研究,分析和解决问题的能力.

实测分析表明:识别概率模型能力低,是导致学生得分率低的主要原因.

超几何分布与二项分布是两个非常重要且应用广泛的概率模型.超几何分布模型对两类对象(物、人或事)进行不放回抽样,由于每取出一个,总体中就少一个,所以每次抽取的概率是不同的,题干中总体容量已知.二项分布模型对两类对象(物、人或事)进行放回抽样,由于每次抽取时总体不变,所以每次抽取的概率是p,且保持不变,题干中总体容量一般未知.

例2 某宿舍有7名学生,其中学习用功的有3名,不用功的有4名.从7名学生中随机抽取3名学生,若不用功的学生人数为X,求随机变量X的数学期望.

分析 总体容量即该宿舍的学生总数已知.本例题是对两类不同的对象,即学习用功与否的学生进行放回抽样的问题,所以$X\sim H(7,4,3)$,则

$$P(X=k)=\frac{C_4^k\cdot C_3^{3-k}}{C_7^3}(k=0,1,2,3),$$

$$E(X)=\frac{3\times4}{7}=\frac{12}{7}.$$

例3 若某班级每名学生因学习不用功考差的概率为$\frac{2}{3}$,任一学生之间学习不用功与否相互独立.若7名学生中因学习不用功考差的人数为Y,求随机变量Y的数学期望.

分析 总体容量未知,即该班级学生的总数.本例题是对两类不同的对象,即因学习不用功考差与否的放回抽样问题,则$Y\sim B(7,\frac{2}{3})$,所以

$$E(Y)=7\times\frac{2}{3}=\frac{14}{3}.$$

例4 若某地区学生因学习不用功考差人数占比为$\frac{2}{3}$,现从该地区中随机抽取7名学生,这7名学生因学习不用功考差的人数为Z,求随机变量Z的数学期望.

分析 总体容量即该地区学生总数.本题是对两类不同的对象未知,即因学习不用功考差与否的学生进行不放回抽样问题.然而,当总体容量很大时,由于每次抽取时总体几乎不变,所以抽取概率也几乎不变,近似看成独立重复试验,则$Z\sim B(7,\frac{2}{3})$,所以

$$E(Z)=7\times\frac{2}{3}=\frac{14}{3}.$$

例5 某宿舍有7名学生,每名学生因学习不用功考差的概率为$\frac{2}{3}$,任一学生学习不用功与否相互独立.7名学生中因学习不用功考差的人数为Y,ξ表示该宿舍的学习激励金为ξ元,且$\xi=200-30Y$,求随机变量ξ的数学期望.

分析 7名学生的学习激励金ξ不是对两类对象的抽样问题,所以随机变量ξ不服从二项分布或超几何分布.但Y与ξ满足线性关系且$Y\sim B(7,\frac{2}{3})$,因此,利用$E(ax+b)=aE(x)+b$,得出$E(\xi)=E(200-30Y)=200-30E(Y)=200-30\times\frac{14}{3}=60$.

3 提升运算求解能力

运算求解能力是指会根据法则、公式进行正确运算、变形和数据处理;以及根据问题的条件寻找合理、简捷的运算途径,并且根据要求对数据进行估计和近似计算.

概率与统计试题实测分析表明,计算求解能力低

是导致学生得分率低的重要原因.运算能力主要体现在运算的准确性和速度上.为确保准确性,我们需要知道运算之间的相互联系;为确保速度,我们还要掌握一些巧算的方法.例如,计算平均数可以采用预设平均数的方法:$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=x_0+\frac{1}{n}\sum_{i=1}^{n}(x_i-x_0)$.如例1第(1)问,志愿者$M$含量的平均数为

$$\bar{x}=0.3\times0.2+0.5\times0.3+0.7\times0.3+0.9\times0.17+1.1\times0.03=0.606.$$

采用取区间中点值0.7作为预设平均值后

$$\bar{x}=0.7+(-0.4)\times0.2+(-0.2)\times0.3+0+0.2\times0.17+0.4\times0.03=0.606,$$

简化了计算,提高了正确率.有时题干中的参考数据有较强的提示性,据此对运算式进行适当估算,既能避免繁杂的计算,保证准确性,又能提高计算速度.

例6 表1是对生产线零部件尺寸的16次抽样.

表1

抽取次序	1	2	3	4	5	6	7	8
零件尺寸/cm	9.95	10.12	9.96	9.96	10.01	9.92	9.98	10.04
抽取次序	9	10	11	12	13	14	15	16
零件尺寸/cm	10.26	9.91	10.13	10.02	9.22	10.04	10.05	9.95

经计算得

$$\bar{x}=\frac{1}{16}\sum_{i=1}^{16}=9.97,$$

$$s=\sqrt{\frac{1}{16}\sum_{i=1}^{16}(x_i-\bar{x})^2}=\sqrt{\frac{1}{16}(\sum_{i=1}^{16}x_i^2-16\bar{x}^2)}\approx0.212,$$

$$\sqrt{0.008}\approx0.09.$$

剔除$(\bar{x}-3s,\bar{x}+3s)$之外的所有数据,求剩余零件尺寸的平均数与标准差.

分析 因为第13个数据$9.22\notin(\bar{x}-3s,\bar{x}+3s)$,所以剔除后剩余15个数据.在没有数据参考的情况下,采用直接代入公式求样本的平均数和方差的方法,很难在短时间内求出准确结果.然而,抓住剔除前后的数据差异以及求平均数和方差所需的数据:$\sum_{i=1}^{12}x_i+\sum_{i=14}^{16}x_i,\sum_{i=1}^{12}x_i^2+\sum_{i=14}^{16}x_i^2$,利用参考数据中,剔除前的16个数据的平均数和方差,就可快速得到答案.这15个数据的平均数为$\frac{1}{15}(16\times9.97-9.22)=10.02$,方差为

$$\frac{1}{15}(16\times0.212^2+16\times9.97^2-9.22^2-15\times10.02^2)\approx0.008.$$

根据参考数据$\sqrt{0.008}\approx0.09$,可得标准差为0.09.

4 小结

在概率与统计的学习中,审题是一个非常重要的环节.由于题目中往往包含大量的背景信息和数据,因此学生需要具备敏锐的审题能力,准确地理解题目的要求和所给的数据.然而,在实际的学习中,很多学生常常因为审题不当而出现错误,所以提升学生的审题能力是至关重要的.计算是概率与统计学习中的另一个重要环节,由于概率与统计涉及大量的数值计算和统计分析,因此学生需要具备扎实的计算能力.然而,在实际的学习中,很多学生常常因为计算错误或计算方法不当而丢分,所以提高学生的计算能力也是非常必要的.模型识别也是概率与统计学习中的重要环节.由于概率与统计中涉及大量的统计模型和概率模型,因此学生需要具备敏锐的模型识别能力.然而,在实际的学习中,很多学生常常因为模型识别错误而无法得出正确的结论,所以提高学生的模型识别能力也是非常关键的.针对以上问题,我们提出以下建议.

4.1 提高学生的审题能力

学生应该仔细阅读题目,并尝试将题目中的信息和数据转化为数学语言.同时,学生还应该注意题目中的关键词和条件,确保自己完全理解了题目的要求.

4.2 提高学生的计算能力

学生应该掌握基本的数值计算和统计分析方法,并尝试通过解决实际问题来提高自己的计算速度和准确性.同时,学生还应该注意计算方法的选择,确保自己能够根据问题的具体情况选择合适的计算方法.此外,学生还应该注意计算结果的合理性和准确性,确保自己能够得到正确的结论.

4.3 提高学生的模型识别能力

学生应该熟悉各种统计模型和概率模型的特性,并尝试通过对比和分类来识别不同的模型.同时,学生还应该注意模型的应用条件和范围,确保自己能够正确地应用模型进行分析和预测.

实践表明,以上建议可以有效地帮助学生在概率与统计问题的求解中提高分析和解决问题的能力.通过有针对性的训练和实践经验的积累,学生可以获得更为敏锐的审题能力、准确的计算能力、高效的模型识别能力以及更强的数学建模能力、计算求解能力和创新能力,提升数学核心素养.

(完)