

基于 R 的概率统计实验教学模式探讨

刘东海, 彭丹

(湖南科技大学 数学与计算科学学院, 湖南 湘潭 411201)

摘要:从概率统计课程的特点、学生学习的主要目的以及 R 软件的优势等方面强调概率统计课程引入 R 软件的必要性, 结合具体实例探讨了如何通过 R 软件优化概率统计课程教学, 探索提高教学效果和教学质量的新型概率统计实验教学模式。

关键词: 概率统计; R 软件; 实验教学

中图分类号: G642

文献标志码: A

文章编号: 1674-5884(2014)07-0043-03

概率统计的任务是研究有关收集、整理、分析数据, 并对所考察的问题做出一定的结论的方法与理论^[1]。在概率统计教学过程中, 应使学生掌握概率统计的基本知识和基本思想, 具有一定的统计应用能力, 并能借助计算机及软件完成统计计算, 分析统计结果, 做出统计推断, 为学生今后的工作打下初步的基础。

重理论、轻实践是目前概率统计课程教学的主要特点, 教师过度重视计算技巧的演练, 注重如何将教学内容讲透、讲细, 从而忽视了统计思想和数据处理能力的培养^[1-2], 这就致使学生很难将所学知识点与实际联系, 无法运用所学知识去分析和解决实际问题。另一方面, 使用传统的教学方法和教材, 无论是老师讲解例题, 还是学生完成习题都要花大量的时间进行手工计算, 如果使用软件可使数据分析更具有直观性、灵活性和可重复性, 起到举一反三的作用, 提高学生的学习兴趣。本文介绍基于 R 软件的概率统计实验教学, 主要原因是 R 软件有强大的数据图形展示和统计分析功能, 能免费使用和更新软件, 并可随时加载有针对性的软件包, 而 SAS、Matlab、SPSS、Eviews、S-Plus 却都是收费软件, 同时 R 软件高效的代码、简洁的输出和强大的帮助系统使 R 软件辅助概率统计教学成为可能。

1 基于 R 的概率统计实验教学模式研究

1.1 课程的任务驱动教学

概率统计是研究随机现象统计规律性的一门学科, 而要想获得随机现象的统计规律性, 就必须进行大量的重复试验, 这在有限的课堂时间内是难以实现的, 为此,

教师可以任务驱动教学法指导学生用 R 软件设计概率统计中的随机试验^[3], 根据教学主题把教学内容分解成多个具体的任务, 每个任务中都蕴含了学生必须掌握的相关知识和技能, 引导学生自主探索, 分析问题, 提出解决问题的方法, 并通过计算机用 R 实现图形显示动画模拟和数值计算等问题, 形成一个生动、直观的教学环境。

1.2 基于 R 的概率统计实验教学模式

将 R 实验引入到概率统计教学是一种全新的教学理念, 使概率统计教学从单纯的教师讲课、学生听课的模式发展到利用 R 软件实现师生共同参与的学习模式^[4-5]。同时利用 R 软件对随机试验的动态过程进行演示和模拟, 如投掷骰子实验、点估计相关性试验等, 再现了抽象理论的研究过程, 加深了学生对理论的理解及方法的运用, 这必将激发学生解决实际问题的兴趣, 培养学生应用概率统计知识解决实际问题的能力。

基于 R 的概率统计实验教学模式, 为概率统计和数学实验的应用提供了广阔的前景, 给概率统计课程教学注入了活力, 更能给学生一个“完整的概率统计”^[6]。该教学模式可以概括为四个环节: 创设情景—随机实验—归纳猜想—推理论证; 其主要内容是: 提出问题—分析问题—解决问题—拓展问题, 该模式在概率统计教学以及培养学生的统计建模能力方面显得尤为有效。

1.3 基于 R 的概率统计实验教学模式实践案例

众所周知, 概率论早期研究的是游戏或赌博随机现象中有关的概率问题, 我们可以借助 R 软件演示随机试验, 让学生直接观察并参与到试验中, 可编制如下实验:

例 1 (掷骰子) 掷一颗质地均匀的骰子 15 次, 令 X

表示出现的点数。

```
R 程序如下:sample(1:6,15,replace = T)
输出结果为:1 3 3 6 3 4 2 5 1 3 2 1 5 2 2(第一次实验)
              3 6 3 2 4 6 2 5 5 1 2 4 3 2 1(第二次实验)
```

该实验重复多次,可让学生了解各点数出现的随机性,又可以分析各点数出现的频率的稳定性和变化规律。

例 2 (随机游动)假设我们进行掷硬币实验,如果掷到人像就赢 2 元,掷到文字就输 2 元,这是一种简单的随机游动,我们可以用 R 设计如下随机实验。

设最初的金额是 $W(0) = 0$, $W(t)$ 表示在时间 t 累积的金额,则

$$W(t + 1) = W(t) + r(t) \quad t = 0, 1, 2, \dots, n$$

其中 $r(t) = 2$ 或 -2 ,机会是 0.5;

```
R 程序如下:
set.seed(13 579);
r = sample(c(-2,2), size = 100, replace = T, prob = c(0.5,0.5));
W = cumsum(r);
plot(W);
W 的值如图 1 所示。
```

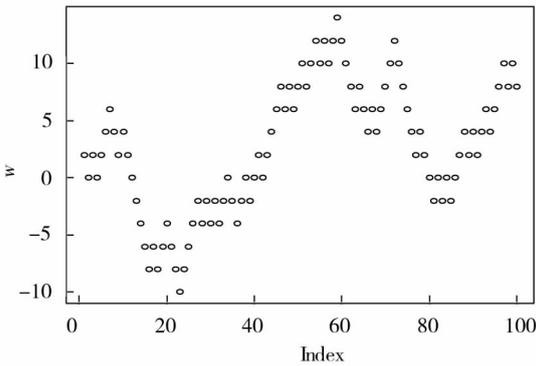


图 1 累积金额

在概率统计中对于一个具体的问题,通常归纳为对一个随机变量的取值及取值概率的研究,即对于事件 $P(X \leq x)$ 的研究,这就是随机变量的累积分布函数,我们可以借助 R 求随机变量的概率分布。

例 3^[7] (正态分布)设随机变量 $X \sim N(0,1)$,求 $P(X < 1.96)$ 的概率。

如果我们应用 R,可以直接设计

```
R 程序如下:p = pnorm(1.96)
              p = 0.975 002 1
```

这避免了对标准正态分布的密度函数求积分的复杂运算。

又如在讲授矩法估计时,我们知道矩估计可能不是唯一的,这是矩法估计的一个缺点,一般情况下用低阶矩估计给出未知参数的估计,为了使学生形象直观地了解为什么一般用低阶矩来估计未知参数,可借助 R 设置如下随机试验。

例 4 (矩估计)下面的观察值来自指数分布的一个样本:

```
0.593 275 4    0.128 549 35    0.469 002 28
0.298 359 80    0.243 414 62    0.065 666 37    0.400 855
36    2.996 871 23    0.052 789 12    0.098 985 944
```

我们来估计参数 λ 。

如果采用一阶矩进行估计,则 R 程序如下:

```
x = c(0.591 327 54, 0.128 549 35,0.469 002 28,
0.298 359 80 ,0.243 414 62,0.065 666 37 ,0.400 855 36,
2.996 871 23,0.052 789 12,0.098 985 944)
```

```
Lambda = 1/mean(x)
```

```
Lambda = 1.870 62
```

如果采用二阶矩进行估计,则 R 程序如下:

```
Lambad = 1/sd( lambda )
```

```
Lambad = 1.131 03
```

实际上上面的数据是模拟参数为 2 的指数分布,一阶矩估计为 1.870 62,二阶矩估计为 1.131 03,从上述实验结果中可直接观察到在矩法估计时采用低阶矩估计未知参数更精确。

在讲述相关性时,我们知道相关关系是指两个变量的数值变化存在不完全确定的依存关系,它们之间的数值不能用方程表示出来,但可用某种相关性度量来刻画,这时我们可以适当引申本内容,以实际问题为背景,让学生有机会脱离书本,利用自己学过的知识去认识问题,进一步激发学生学习的积极性。

例 5 某医生测定了 10 名孕妇的 15~17 周及分娩时脐带血 TSH 水平如下表 1 所示,试问变量 X 与 Y 是否相关?

表 1 脐带血和母血水平关系

母血 TSH(X)	1.21	1.30	1.39	1.42	1.47	1.56	1.68	1.72	1.98	2.10
脐带血(Y)	3.90	4.50	4.20	4.83	4.16	4.93	4.32	4.99	4.70	5.20

我们可以运用 R 画出 (X, Y) 的图:

```
R 程序如下:X = c(1, 21, 1.30, 1.39, 1.42, 1.47,
1.56, 1.68, 1.72, 1.98, 2.10)
              Y = c(3.90, 4.50, 4.20, 4.83, 4.16, 4.93, 4.32, 4.99,
4.70, 5.20)
```

```
Level = data.frame(X, Y)
```

```
plot(Level)
```

运行结果如图 2 所示,从图中我们只能推测 X 和 Y 之间有某种关系,但如何验证呢?这时可以进行第二步:

```
R 程序如下:attach(level)
```

```
cor.test(X,Y)
```

运行结果为

```
Pearson's product - moment correlation
```

```
data: x and y
```

```
t = 2.628 4, df = 8, p - value = 0.030 25
```

```
alternative hypothesis: true correlation is not equal to
```

```
0.95 percent confidence interval:
```

```
0.089 433 59 0.917 227 01
```

```
sample estimates:
```

```
cor
```

```
0.680 728 3
```

因为 p 值为 $0.0325 \leq 0.05$, 故拒绝原假设, 从而认为

X 和 Y 相关。

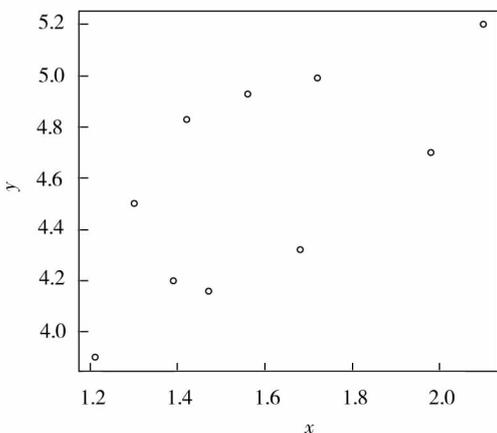


图2 X和Y的散点图

2 基于R的概率统计实验教学的意义

结合目前形势,实行概率统计的实验教学,是改变学

生对概率统计课程呆板枯燥的认识,提高学生学习概率统计兴趣的有效途径。在概率统计实验教学中,运用R对实际数据处理的过程演示,不仅能帮助学生充分理解概率统计问题的实质以及与实际问题之间的联系,而且能加深学生对概念、理论及方法的认识和理解;并进一步使学生熟练应用R的使用方法,且能使用R软件进行统计分析。

基于R的概率统计实验教学模式能将传统的教学方法与现代教学方法有机地联系起来,相辅相成,互为补充。让学生在学习中有所感悟、有所启发、有所体验,让教师通过教学做出更优化更利于学生发展的教学安排,以更好地指导学生开展研究性学习。

参考文献:

- [1] 刘琼荪. 将数学建模思想融入工科概率统计教学中[J]. 大学数学, 2006(2): 152 - 154.
- [2] 李大潜. 漫谈大学数学教学的目标与方法[J]. 中国大学数学, 2009(1): 7 - 10.
- [3] 顾光同, 张香云, 徐光辉. 统计实验寓于概率统计教学的探索与实践[J]. 统计与决策, 2007(21): 165 - 167.
- [4] 姜启源. 数学实验与数学建模[J]. 数学的实践与认识, 2015(15): 613 - 617.
- [5] 尹江丽. 数理统计课程教学方法探讨[J]. 数学教学研究, 2013(78): 707 - 715.
- [6] 代金辉. 随机模拟寓于概率统计教学中的探索与实践[J]. 高师理科学刊, 2009(29): 90 - 93.
- [7] 邓华玲. 概率统计方法与应用[M]. 北京: 中国农业出版社, 2006.

(责任校对 晏小敏)