工程统计学

茆诗松　周纪芗　编著

高等教育出版社

工程统计学

茆诗松　周纪芗　编著

高等教育出版社·北京

内容提要

本书是为高等学校工科各专业大学生编写的统计学教程。全书分八章。前两章讲统计学所需的概率知识，以一维概率分布为主。第三、四章讲两种统计推断形式，参数估计与假设检验。最后四章讲四种常用统计技术：统计过程控制、方差分析、回归分析和正交试验设计，可以全讲，也可选讲，视教学学时多少而定。

全书以讲授统计方法为主，不求理论完整，但求读者处理各类数据的能力得到不断提高。全书尽量选用启发性的例子编写，语言通俗，便于学生阅读。各章配以适量的习题，以供选用。

高等学校其他专业，如管理、经济、农林、医药等亦可选用，或作自学教材。

图书在版编目(CIP)数据

工程统计学/茆诗松，周纪芗编著．--北京：高等教育出版社，2018.12

ISBN 978-7-04-049645-1

Ⅰ. ①工… Ⅱ. ①茆… ②周… Ⅲ. ①工业统计学-高等学校-教材 Ⅳ. ①F402.4

中国版本图书馆CIP数据核字(2018)第084193号

策划编辑 田 玲　　责任编辑 田 玲　　封面设计 张申申　　版式设计 马敬茹
插图绘制 邓 超　　责任校对 吕红颖　　责任印制 尤 静

出版发行 高等教育出版社
社 址 北京市西城区德外大街4号
邮政编码 100120
印 刷 涿州市京南印刷厂
开 本 787mm×960mm 1/16
印 张 24
字 数 430千字
购书热线 010-58581118
咨询电话 400-810-0598
网 址 http://www.hep.edu.cn
http://www.hep.com.cn
网上订购 http://www.hepmall.com.cn
http://www.hepmall.com
http://www.hepmall.cn
版 次 2018年12月第1版
印 次 2018年12月第1次印刷
定 价 43.80元

物 料 号 49645-00

序　言

2000年，在我国工厂中，特别在一些大型企业中，兴起了一个质量工程师培训和六西格玛统计方法推广的小高潮，我们有幸被邀与一些高级工程师和质量管理人员经常在一起商议如何编写培训教材事项。大家认为统计部分应以统计方法为主，着力培养学员处理受随机干扰的数据的能力，为此应该确定哪些统计方法可进入培训教材，又需要讲清哪些概念和定理。通过讨论，定下大纲，分头编写。我们两人负责概率统计基础知识和常用统计方法两部分的编写任务，写完后，相互审阅修改。定稿时，大家认为要以概念清晰、方法实用、例子生动、语言朴实为准。出版后编写人员亲自授课，听取学员意见，在教学相长中我们不断改进教材和教法，使学员更易掌握。那时每年都要修改一次教材，这对更新教材，提高教学水平提供了极好的机会。

2014年质量工程师的考试停止了，从而培训也停止了，培训教材的不断更新也停止了，但我们的思路没有停止。我们想：既然对质量工程师的培训有用，那对一般工程师的培养也有用。不讲究数学体系完美，不追求数学证明完善，只求不断提高处理各类数据的能力，把统计学中很实用的部分集中起来形成教材，书中多处采用启发性的例子和宜于接受的方式编写，这不是对工科学生更实用吗？

十几年的实践让我们了解工程师的成长路径，他们学习统计，主要是培养处理各种数据的能力，从中发掘隐藏在数据后面的统计规律。据此想法，我们决定编写一本工科统计学教材。我们两人在原教材框架基础上着手改进，适当扩充一些内容，增加章节之间的联系，使其成为一本适应工科大学生阅读的教材。譬如，我们加强概率分布的描述，从一堆数归结成一个概率分布，在不减弱正态分布叙述的基础上又重点介绍对数正态分布、伽马分布和韦布尔分布。在概率论中分布只充当例子角色，可在统计中被看作模型来使用，一个一个仔细研究其形成背景和使用范围。一个分布就是一个统计模型，在慢吞细嚼中品出味道来。在回归分析中增加多元回归和回归诊断。在置信区间和假设检验中强化对分位数的理解和使用，并适当增加了样本量确定的讨论。在不讲究严格证明的前提下用图形来描述频率稳定于概率的大数定律。用实例来叙述中心极限定理的含义及应用。

工程师在当今世界中发挥着重要作用。因为他们在为我们社会需要的最好产品中挑起了设计和制造的担子,而有一定统计修养和训练,对他们的今后工作会有较大影响。这次编写教材是让工科学生理解处理数据的方法,学会使用处理数据的基本技巧,提高他们解决工程问题的能力,而不要求他们知道统计学的数学原理。

概率统计的工科教材与理科教材是有差别的。现在讨论的问题是(1) 概率讲多少才够大多数工科学生的需要;(2) 统计方法选哪些入教材为宜。本书给出关于这两个问题的看法,是否妥当还望在今后的施行中不断进行修改。

本书共分八章,前五章由茆诗松编写,后三章由周纪芗编写,最后由茆诗松统稿。我们估计若有70多个学时,可基本上把全书讲完。若只有40个左右的学时,我们建议主要讲授前四章,再加第五章(统计过程控制)或再加第七章(回归分析)。我们的一切努力是否有效还要实践检验,还望诸位多提意见,以便不断改进。

茆诗松,周纪芗

2017年12月

目　录

第一章 基础概率

§1.1 事件与概率

1.1.1 随机现象

在一定条件下,并不总出现相同结果的现象称为随机现象. 从这个定义中可看出,随机现象有两个特点:

(1) 随机现象的结果至少有两个;

(2) 至于哪一个出现,事先并不知道.

抛硬币、掷骰子是两个最简单的随机现象的例子. 抛一枚硬币,可能出现正面,也可能出现反面,至于哪一面出现,事先并不知道. 又如掷一颗骰子,可能出现1点到6点中某一个,至于哪一点出现,事先也不知道. 这里"事先不知道"就是随机性的特征.

只有一个结果的现象称为确定性现象. 例如,太阳从东方升,水向低处流,同性电荷相斥,异性电荷相吸等都是确定性现象.

例 1.1.1 以下是另外一些随机现象的例子:

(1) 一天内进入某超市的顾客数;

(2) 一顾客在超市中购买的商品数;

(3) 一顾客在超市排队等候付款的时间;

(4) 一颗麦穗上长着的麦粒数;

(5) 新产品在未来市场的占有率;

(6) 一台电视机从开始使用到发生第一次故障的时间;

(7) 加工某机械轴直径的误差;

(8) 一罐午餐肉的净重.

随机现象在工农业生产和管理中到处可见. 各级研究和管理人员都要形成随机观念,并用它去看待和分析现实世界中产生的各种问题.

认识一个随机现象首要的是罗列出它的一切可能发生的基本结果. 这里的基本结果称为样本点,**随机现象的一切可能样本点的全体称为这个随机现象的样本空间**,常记为 Ω. 如

"抛一枚硬币"的样本空间 $\Omega=\{$正面,反面$\}$;

"掷一颗骰子"的样本空间 $\Omega=\{1,2,3,4,5,6\}$；

"一顾客在超市中购买商品的件数"的样本空间 $\Omega=\{0,1,2,\cdots\}$；

"一台电视机从开始使用到发生第一次故障的时间"的样本空间 $\Omega=\{t:t\geqslant 0\}$；

"测量某物理量的误差"的样本空间 $\Omega=\{x:-\infty<x<\infty\}$.

1.1.2 随机事件

随机现象的某些样本点组成的集合称为随机事件,简称事件,常用大写字母 A,B,C 等表示. 如在掷一颗骰子时,"出现奇数点"是一个事件,它由 1 点、3 点、5 点共三个样本点组成,若记这个事件为 A,则有 $A=\{1,3,5\}$.

从随机事件的定义可见,事件有如下几个特征:

(1) 任一事件 A 是相应样本空间 Ω 的一个子集. 在概率论中常用一个长方形示意样本空间 Ω,用其中一个圆(或其他几何图形)示意事件 A,如图 1.1.1 所示,这类图形称为维恩(Venn)图.

(2) 事件 A 发生当且仅当 A 中某一样本点发生,若记 ω_1,ω_2 是 Ω 中的两个样本点(见图 1.1.1):

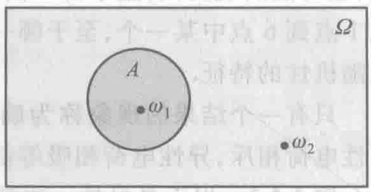

图 1.1.1 维恩图

当 ω_1 发生,且 $\omega_1\in A$(表示 ω_1 在 A 中),则事件 A 发生;

当 ω_2 发生,且 $\omega_2\notin A$(表示 ω_2 不在 A 中),则事件 A 不发生.

(3) 事件 A 的表示可用集合,也可用语言,但所用语言应是明白无误的.

(4) 任一样本空间 Ω 都有一个最大子集,这个最大子集就是 Ω,它对应的事件称为必然事件,仍用 Ω 表示. 如掷一颗骰子,"出现点数不超过 6"就是一个必然事件,因为它含有 $\Omega=\{1,2,3,4,5,6\}$ 中所有样本点.

(5) 任一样本空间 Ω 都有一个最小子集,这个最小子集就是空集,它对应的事件称为不可能事件,记为 $\varnothing$. 如掷一颗骰子,"出现 7 点"就是一个不可能事件,因为它不含有 $\Omega=\{1,2,3,4,5,6\}$ 中任一个样本点.

例 1.1.2 设产品只区分合格与不合格,且我们把注意力放在不合格品上. 若记合格品为 0,不合格品为 1,则检查两件产品的样本空间 Ω 由下列四个样本点组成:

$$\Omega=\{(0,0),(0,1),(1,0),(1,1)\}$$

其中样本点 $(0,1)$ 表示第一件产品为合格品,第二件产品为不合格品,其他样本点可类似解释. 下面几个事件可用集合表示,也可用语言表示.

A="至少有一件合格品"$=\{(0,0),(0,1),(1,0)\}$；

B="至少有一件不合格品"$=\{(0,1),(1,0),(1,1)\}$；

C="恰好有一件合格品"$=\{(0,1),(1,0)\}$；

Ω="至多有两件合格品"$=\{(0,0),(0,1),(1,0),(1,1)\}$；

$\varnothing$="有三件不合格品".

现在我们转入考察"检查三件产品"这个随机现象，它的样本空间 Ω 含有 $2^3=8$ 个样本点，

$$\Omega=\{(0,0,0),(0,0,1),(0,1,0),(1,0,0),(0,1,1),(1,0,1),(1,1,0),(1,1,1)\}$$

下面几个事件可用集合表示，也可用语言表示.

A="至少有一件合格品"$=\{\Omega$ 中剔去$(1,1,1)$的其余 7 个样本点$\}$；

B="至少有一件不合格品"$=\{\Omega$ 中剔去$(0,0,0)$的其余 7 个样本点$\}$；

C_1="恰有一件不合格品"$=\{(0,0,1),(0,1,0),(1,0,0)\}$；

C_2="恰有两件不合格品"$=\{(0,1,1),(1,0,1),(1,1,0)\}$；

C_3="全是不合格品"$=\{(1,1,1)\}$；

C_0="没有不合格品"$=\{(0,0,0)\}$.

1.1.3 随机变量

用来表示随机现象结果的变量称为随机变量，常用大写字母 X,Y,Z 等表示. 如把掷一颗骰子出现的点数记为 X，则 X 就是一个随机变量，它是仅可能取 1,2,3,4,5,6 这 6 个值的随机变量.

$\{X=2\}$ 表示"出现 2 点"的事件；

$\{X=4\}$ 表示"出现 4 点"的事件；

$\{X>4\}$ 表示"出现 5 点或 6 点"的事件；

$\{X\geqslant 4\}$ 表示"出现 4,5,6 点中某一个"的事件；

$\{2<X\leqslant 5\}$ 表示"出现 3,4,5 点中某一个"的事件.

由此可见，当随机变量 X 的含义设置清楚时，用等号或不等号把 X 与数联系起来就可表示许多不同的事件，甚至必然事件和不可能事件也可用随机变量的取值表示，如

$\{X\geqslant 1\}$ 就是必然事件 $\Omega=\{1,2,3,4,5,6\}$；

$\{X=3.5\}$ 就是不可能事件 $\varnothing$.

随机事件不仅可用集合和语言表达，还可用随机变量的取值表示. 今后在实际中遇到的事件大多都是用随机变量的取值表示的事件，因为此种表示简洁、明白、灵活，不会引起误解.

例 1.1.3 检查三个产品，若产品只分合格品（用 0 表示）和不合格品（用 1

表示)，这时其样本空间 Ω 共含 8 个样本点(见例 1.1.2)，其中每个样本点都可能是检查的一个结果. 若忽略合格品与不合格品出现的先后次序，只考察样本点中不合格品数 X，则 X 是一个随机变量，且样本点 ω 与 X 的取值 x 间有如下对应关系：一个 ω 对应一个 x，不同的 ω 可对应同一个 x.

样本点 ω	(0,0,0)	(0,0,1)	(0,1,0)	(1,0,0)	(0,1,1)	(1,0,1)	(1,1,0)	(1,1,1)
X 的取值 x	0	1	1	1	2	2	2	3

这种对应关系实质是函数关系，只是其自变量是样本点 ω，它可以是数也可以不是数，但因变量 X 一定是数.

从这一例子看出，"用来表示随机现象结果的随机变量"就是**定义在样本空间 Ω 上的一个实值函数**，即 $X=X(\omega)$. 由于样本点 ω 的出现是随机的，故随机变量 $X(\omega)$ 的取值也是随机的. 随机变量的进一步讨论将在第二章展开，这里只要会用随机变量表示事件.

例 1.1.4 检查 10 个产品时，记其中的不合格品数为 Y，则 Y 是一个随机变量. 可用 Y 表示不同事件.

"至少有 1 件不合格品" $=\{Y\geqslant 1\}$，这里等号两端的事件所含的样本点相同，故称这两个事件相等；

"至少有 2 件不合格品" $=\{Y\geqslant 2\}$；

"至多有 2 件不合格品" $=\{Y\leqslant 2\}$；

"没有 1 件是合格品" $=\{Y=10\}$.

特别，记检查 1 个产品时的不合格品数为 Z，则有

$A=\{Z=1\}=$"产品不合格"；

$B=\{Z=0\}=$"产品合格"；

$C=\{Z\leqslant 1\}=\Omega$，这是必然事件；

$D=\{Z<0\}=\varnothing$，这是不可能事件.

1.1.4 事件的概率

任一随机事件的发生是带有偶然性的. 但其发生的可能性还是有大小之别的，是可以度量的. 这个度量尺度就是概率.

一个随机事件 A 发生可能性的大小称为这个事件的概率，记为 $P(A)$. 概率是介于 0 到 1 之间的一个数. 概率愈大，事件发生的可能性就愈大，特别规定，必然事件 Ω 的概率为 1，不可能事件 $\varnothing$ 的概率为 0，即

$$P(\Omega)=1,\quad P(\varnothing)=0,\quad 0\leqslant P(A)\leqslant 1$$

例 1.1.5 确定概率的例子.

(1) 抛一枚硬币,出现正面与出现反面的机会(概率)是相等的,各为$\frac{1}{2}$. 足球裁判员就用抛硬币的方法让双方队长选择场地,以示公正.

(2) 掷一颗骰子,其样本空间 Ω 含有6个样本点,它们出现的机会是等可能的(说不出哪个样本点出现的可能性更大一些),因此每个样本点出现的概率各为$\frac{1}{6}$. 若记 X 为掷一颗骰子出现的点数,则有

$$P(X=1)=\frac{1}{6},\quad P(X\leqslant 2)=\frac{2}{6}=\frac{1}{3},\quad P(X>2)=\frac{4}{6}=\frac{2}{3}$$

(3) 设由 100 个产品组成的一个产品批次中含有 4 个不合格品,则从该批次中随机抽取一个产品为不合格品的概率为 0.04 或 4%. 这里"随机抽取"一词意味着每个产品都有相同的机会被取出. 若该批次中没有不合格品,则其概率为 0 或 0%. 若该批次产品全是不合格品,则其概率为 1 或 100%.

以上几个例子表明:在经验事实(等可能性)基础上,用逻辑分析获得一些(不是所有)事件的概率,这种确定概率的方法在历史上称为**古典方法**. 该方法的要点如下:**在有限样本空间场合,如诸样本点是等可能的,则其事件 A 的概率可定义为**

$$P(A)=\frac{\text{事件 }A\text{ 所含样本点的个数}}{\Omega\text{ 中所有样本点的个数}}$$

确定概率的另一个方法是**统计方法**,它的要点是:用多次重复试验所获得的事件 A 发生的频率 $\hat{P}(A)$ 去估计该事件的概率,即

$$\hat{P}(A)=\frac{\text{事件 }A\text{ 出现的次数}}{\text{重复试验的总次数}}$$

当重复试验次数增加时,频率 $\hat{P}(A)$ 愈稳定于概率 $P(A)$. 下面用例子说明统计方法.

(4) 在足球比赛中罚点球是激动人心的场合,因"点球命中"是一个随机事件,若记该事件为 A,则 $P(A)$ 等于多少,如何来获得这个概率? 有人对 1930—1988 年世界各地的 53274 场重大足球比赛进行了数据统计:在判罚的 15382 个点球中有 11172 个命中,由此得到点球命中的概率 $P(A)$ 的估计值为

$$\hat{P}(A)=\frac{11172}{15382}=0.726$$

这个频率是颇有说服力的,因为统计对象是世界级的足球运动员,有权威性.