

管理统计及其应用研究

——以 SPSS15.0 软件应用为例

张 侨 蔡道成 著

湖南师范大学出版社

管理统计及其应用研究

——以 SPSS15.0 软件应用为例

张 侨 蔡道成 著

湖南师范大学出版社

图书在版编目 (CIP) 数据

管理统计及其应用研究——以 SPSS 15.0 软件应用为例/ 张
侨, 蔡道成著. —长沙: 湖南师范大学出版社, 2010. 6

ISBN 978 - 7 - 5648 - 0245 - 5

I. ①管… II. ①张… ②蔡… III. ①统计分析—软件包,
SPSS 15.0—高等学校: 技术学校—教材 IV. ①C819

中国版本图书馆 CIP 数据核字 (2010) 第 098019 号

管理统计及其应用研究——以 SPSS 15.0 软件应用为例

张 侨 蔡道成 著

◇责任编辑: 柳 丰

◇责任校对: 蒋旭东

◇出版发行: 湖南师范大学出版社

地址/长沙市岳麓山 邮编/410081

电话/0731. 88853867 88872751 传真/0731. 88872636

网址/http: //press. hunnu. edu. cn

◇经销: 湖南省新华书店

◇印刷: 长沙市华中印刷厂

◇开本: 850 × 1168 1/32

◇印张: 8

◇字数: 200 千字

◇版次: 2010 年 6 月第 1 版 2010 年 6 月第 1 次印刷

◇书号: ISBN 978 - 7 - 5648 - 0245 - 5

◇定价: 18.00 元

前 言

随着现代信息技术的发展和计算机的普及,以往很多依靠手工运算非常复杂或根本难以进行,使非数学专业和非计算机专业人士望而却步的统计理论和方法,在今天通过计算机软件运算,已经变得非常简单和迅速。并且借助计算机,人们能够处理的统计数据量也越来越大。统计理论和方法在社会科学领域内得到了进一步广泛和深入的研究与应用。统计方法也成为管理研究及应用的一种非常强大的工具。国内外很多学者应用统计方法对经济管理相关问题进行了比较广泛和深入的研究。但是,已有的研究要么偏重统计理论的推演,要么偏重于计算机软件的应用,系统地研究统计理论和方法在管理这一具体的社会科学领域内的应用性还是比较少见。本书以应用统计方法成熟,实用性、易用性和交互性良好的 SPSS 统计分析软件为例,以管理研究及应用为导向,比较系统地研究各种现代统计理论与方法及其在管理中的应用。其目的在于让非数学与非计算机专业的读者,掌握现代统计学的一般理论与方法,了解并学习如何使用 SPSS 统计分析软件针对管理研究及应用问题进行数据融合、数据分析和结果展示等工作。

全书共分为 9 章。各章结构具体如下:

第 1 章 概率统计基础知识,包括概率论基础知识和统计学基础知识两个部分。

第 2 章 SPSS 软件入门,包括 SPSS 简单介绍、SPSS 简单应用等两个部分。为后面各章节应用这个软件进行数据处理作准备。

第3章 统计数据的收集,包括数据及其测度类型、数据收集的调查方法、数据收集的实验方法、数据信度与效度检验、SPSS 在信度检验中的应用、案例分析等六个部分。

第4章 样本数据描述性分析,包括样本数据的基本特征、样本数据的集中特征、样本数据的离散特征、数据特征的综合表达、SPSS 在描述性分析中的应用、案例分析等六个部分。

第5章 参数估计,包括点估计、区间估计、用样本均值推断总体均值的样本容量问题、SPSS 在参数估计的应用、案例分析等五个部分。

第6章 参数假设检验,包括参数假设检验的基本思想、正态分布总体参数假设检验、0-1 分布总体参数假设检验、大样本两任意分布总体均值假设检验、SPSS 在参数假设检验中的应用、案例分析等六个部分。

第7章 方差分析,包括单因素方差分析、双因素方差分析、SPSS 在方差分析中的应用、案例分析等四个部分。

第8章 相关分析,包括线性相关分析、非线性相关分析、偏相关分析、相关系数显著性检验、SPSS 在相关分析中的应用、案例分析等六个部分。

第9章 回归分析,包括一元线性回归分析、多元线性回归分析、曲线回归分析、SPSS 在回归分析中的应用、案例分析等五个部分。

本书的重要特点在于依托 SPSS 软件,把现代统计理论与方法和管理研究及应用问题紧密结合起来,略去复杂演算证明过程,增强理论与方法的实用性。

张 侨 蔡道成

2010 年 4 月于琼州学院

目 录

第 1 章 概率统计基础知识	(1)
1.1 概率论基础知识	(1)
1.2 统计学基础知识	(9)
思考练习题	(12)
第 2 章 SPSS 软件入门	(14)
2.1 SPSS 简单介绍	(14)
2.2 SPSS 简单应用	(26)
思考练习题	(30)
第 3 章 统计数据的收集	(32)
3.1 数据及数据类型	(32)
3.2 收集数据的调查方法	(34)
3.3 收集数据的实验方法	(54)
3.4 数据信度与效度检验	(55)
3.5 SPSS 在信度检验中的应用	(59)
3.6 案例分析	(61)
思考练习题	(63)
第 4 章 样本数据描述性分析	(65)
4.1 样本数据的基本特征	(65)
4.2 样本数据的集中特征	(74)
4.3 样本数据的离散特征	(80)
4.4 数据特征的综合表达	(85)

4.5	SPSS 在描述性分析中的应用	(91)
4.6	案例分析	(105)
	思考练习题	(107)
第 5 章	参数估计	(109)
5.1	点估计的概念	(109)
5.2	区间估计	(114)
5.3	用样本均值推断总体均值的样本容量问题	(122)
5.4	SPSS 在参数估计中的应用	(123)
5.5	案例分析	(131)
	思考练习题	(132)
第 6 章	参数假设检验	(134)
6.1	假设检验的基本思想	(134)
6.2	正态分布总体的参数假设检验	(138)
6.3	0-1 分布总体的参数假设检验	(148)
6.4	大样本两任意总体的均值检验	(150)
6.5	SPSS 在参数假设检验中的应用	(152)
6.6	案例分析	(159)
	思考练习题	(161)
第 7 章	方差分析	(163)
7.1	单因素方差分析	(163)
7.2	双因素方差分析	(170)
7.3	SPSS 在因素分析中的应用	(181)
7.4	案例分析	(196)
	思考练习题	(199)
第 8 章	相关分析	(202)
8.1	线性相关分析	(202)
8.2	非线性相关分析	(204)
8.3	偏相关分析	(205)

8.4	相关系数显著性检验	(206)
8.5	SPSS 在相关分析中的应用	(207)
8.6	案例分析	(214)
	思考练习题	(216)
第9章	回归分析	(218)
9.1	一元线性回归分析	(218)
9.2	多元线性回归分析	(224)
9.3	曲线回归分析	(227)
9.4	SPSS 在回归分析中的应用	(228)
9.5	案例分析	(237)
	思考练习题	(241)
参考文献		(245)

第1章 概率统计基础知识

1.1 概率论基础知识

1.1.1 几个基本概念

1. 随机试验

随机试验是概率论的一个基本概念。在概率论中把符合下面三个特点的试验叫做随机试验(Random experiment):(1)每次试验的结果不止一个,并且能事先明确试验的所有可能结果;(2)试验之前无法确定哪一个结果会出现;(3)在同一条件下可重复进行试验。

2. 基本事件

一次随机试验中每一种可能出现的结果称为基本事件。基本事件必须具有下面两个特点:(1)任何两个基本事件是互斥的。(2)除不可能事件外,任何事件都可以表示为若干个基本事件的和。

3. 随机事件

一些基本事件组成的集合称为随机事件,简称为事件。

4. 样本空间

所有基本事件组成的集合称为样本空间或基本空间。

5. 相容事件

不可能同时发生的几个事件称为不相容事件。反之,可能同

时发生的事件则称为相容事件。

1.1.2 概率基础

概率是随机事件出现的可能性大小的一种度量。

1. 概率运算的主要性质

(1) 设 $\bar{A}$ 是 A 的对立事件, 则 $P(\bar{A}) = 1 - P(A)$ 。当 A 为样本空间时, $\bar{A}$ 为空集, 记为 $\emptyset$ 。由 $P(\bar{A}) = 1 - P(A) = 0$, 可以得出空集的概率为 0。

(2) 对任意两个事件 A 和 B , 有:

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

(3) 如果事件 $A \subset B$, 则 $P(A) \leq P(B)$ 。

2. 等概率随机试验

若一个随机试验的基本的事件个数有限, 其每个基本事件出现的概率相等, 则该随机试验称为等概率随机试验。在等概率随机试验中, A 事件发生的概率计算公式为:

$$P(A) = \frac{A \text{ 事件所包含的基本事件的个数}}{\text{试验样本空间基本事件的总个数}}$$

3. 条件概率

事件 A 发生的条件下, 事件 B 发生的概率, 称为事件 B 的条件概率, 记为 $P(B|A)$ 。

例 1-1 有一个盒子里有 6 个球, 其中白球 4 个, 红球 2 个, 事件 A 为“第一次抽取并抽到红球”, 事件 B 为“第二次抽取并抽到红球”。

则: 事件 A 发生的条件下, 再从余下 5 个球中抽取一个球, 事件 B 发生的概率为 $\frac{1}{5}$ 。对条件概率有:

$$P(B|A) = \frac{P(AB)}{P(A)}$$

由条件概率可知: $P(AB) = P(B|A)P(A) = P(A|B)P(B)$

4. 贝叶斯公式

由条件概率公式可以得到以下贝叶斯公式:

$$\begin{aligned} P(A|B) &= \frac{P(AB)}{P(B)} = \frac{P(B|A)P(A)}{P(BA) + P(B\bar{A})} \\ &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})} \end{aligned}$$

5. 相互独立事件

对任意两个事件 A 和 B , 且 $P(B) > 0$, 若 $P(A|B) = P(A)$, 则称为事件 A 与 B 是相互独立的。

对于任意两个相互独立的事件 A 和 B , 有:

$$P(AB) = P(A|B)P(B) = P(A)P(B)$$

1.1.3 随机变量及其分布

1. 随机变量

表示一个随机试验结果的变量, 其数值是由一次随机试验的结果所决定的。根据取值不同随机变量可以分为离散型随机变量和连续型随机变量。

2. 离散型随机变量

(1) 离散型随机变量的定义。

由有限个数或可列个数字或符号组成的变量称为离散型随机变量。

(2) 离散型随机变量的概率。

离散型随机变量取某个具体数值(对应于一个实验结果 x) 所对应的概率, 即 $P(X=x)$, 称为离散型随机变量的概率。如: 抛一次硬币, 出现任一面的概率均为 0.5, $P(X=\text{正面})=0.5$ 和 $P(X=\text{反面})=0.5$ 。

(3) 离散型随机变量的概率分布。

离散型随机变量的概率分布是指离散型随机变量, 取遍每一个试验结果的概率的分布情况。

例 1-2 有一个盒子中有 20 个除颜色外,其他特征完全一样的小球,其中红色的有 5 个,蓝色的有 8 个,黄色的有 7 个,随机从盒子里摸到红色小球的概率为 0.25,摸到蓝色小球的概率 0.40,摸到黄色小球的概率为 0.35。这就是这一问题的随机变量的概率分布。

用表格和图形表达这一随机变量的概率分布,可以得到表 1-1 和图 1-1。

表 1-1 离散型随机变量的概率分布

X 的取值	1 = 红色	2 = 蓝色	3 = 黄色
X 的概率 $P(X=x)$	0.25	0.40	0.35

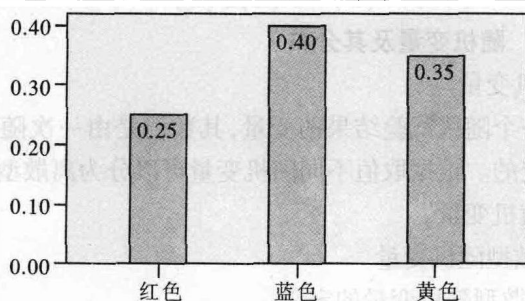


图 1-1 离散型随机变量的概率分布

(4) 离散型随机变量的累积概率。

随机变量 X 小于等于 x 的概率 $P(X \leq x)$ 称为 X 的累积概率。

例 1-2 中,随机变量 $X \leq 2$ 的累积概率 $P(X \leq 2) = 0.65$ 。

(5) 离散型随机变量的累积概率分布。

离散型随机变量的累积概率分布是指离散型随机变量小于等于于每一个试验结果 x 的概率 $P(X \leq x)$ 的分布情况。

例 1-2 中有:

$$P(X \leq 1) = 0.25$$

$$P(X \leq 2) = 0.65$$

$$P(X \leq 3) = 1.00$$

这个结果用表格表达如表 1-2 所示。

表 1-2 离散型随机变量的累积概率分布

X 的取值	1 = 红色	2 = 蓝色	3 = 黄色
X 的累积概率 $P(X \leq x)$	0.25	0.65	1.00

这个结果用条形图和折线图表示如图 1-2 和图 1-3 所示。

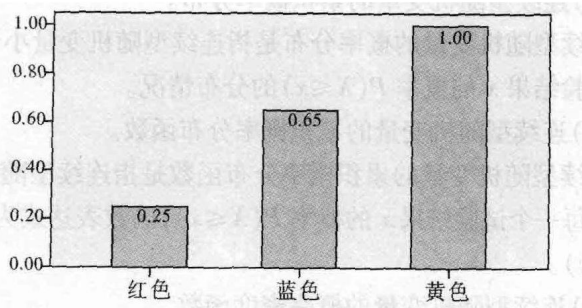


图 1-2 离散型随机变量累积概率分布条形图

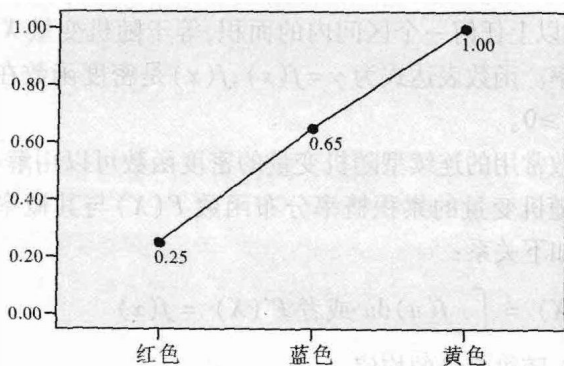


图 1-3 离散型随机变量累积概率分布折线图

3. 连续型随机变量

(1) 连续型随机变量的定义。

取值域为一个连续区间的随机变量称为连续型随机变量。

(2) 连续型随机变量的概率。

连续型随机变量只有在在一个区间上取值,其概率才可能为正值,即才可能有下式:

$$0 < P(x_1 \leq X \leq x_2) \leq 1, \text{ 式中 } x_1 < x_2。$$

(3) 连续型随机变量的累积概率。

随机变量 X 小于等于 x 的概率 $P(X \leq x)$ 称为 X 的累积概率。

(4) 连续型随机变量的累积概率分布。

连续型随机变量的概率分布是指连续型随机变量小于等于每一个试验结果 x 的概率 $P(X \leq x)$ 的分布情况。

(5) 连续型随机变量的累积概率分布函数。

连续型随机变量的累积概率分布函数是指连续型随机变量小于等于每一个试验结果 x 的概率 $P(X \leq x)$, 函数表达式为 $F(X) = P(X \leq x)$ 。

(6) 连续型随机变量的概率密度函数。

连续型随机变量的概率密度函数是这样一个函数:在曲线以下,横轴以上任何一个区间内的面积,等于随机变量 X 在该区间上的概率。函数表达式为 $y = f(x)$, $f(x)$ 是密度函数在 x 处的取值, $f(x) \geq 0$ 。

多数常用的连续型随机变量的密度函数可以用解析式表达。连续型随机变量的累积概率分布函数 $F(X)$ 与其概率密度函数 $f(x)$ 有如下关系:

$$F(X) = \int_{-\infty}^x f(u) du \text{ 或者 } F'(X) = f(x)$$

(7) 随机变量的均值。

随机变量 X 的均值 μ 由下式给出。

对于离散型随机变量:

$$\mu = E(X) = \sum_{i=1}^{\infty} x_i p_i$$

式中 x_i 是随机变量 X 所有可能取值中第 i 个值, $p_i = P(X = x_i)$, $i = 1, 2, \dots$ 。

对于连续型随机变量:

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x) dx$$

式中 $f(x)$ 为随机变量 X 的概率密度函数。

(8) 随机变量的方差。

随机变量 X 的方差 σ^2 由下式给出。

对于离散型随机变量:

$$\sigma^2 = \text{Var}(x) = \sum_{i=1}^{\infty} (x_i - \mu)^2 p_i$$

对于连续型随机变量:

$$\sigma^2 = \text{Var}(x) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

1.1.4 几个典型概率分布

1.0-1 分布

如果 X 服从 0-1 分布, 记为 $X \sim B(1, p)$, 即

$$P(X=1) = p$$

$$P(X=0) = 1 - p$$

显然, 0-1 分布的均值为 p , 即 $E(X) = p$, 方差为 $p(1-p)$, 即 $D(X) = p(1-p)$ 。

2. 二项分布

例如抛不均匀的硬币 n 次, 正面出现 k 次 ($0 \leq k \leq n$) 的概率为:

$$P(X=k) = P_n(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

其中系数 $\binom{n}{k} = n(n-1)\cdots(n-k+1)/k!$ 。

则称 X 服从二项分布, 可记为 $X \sim B(n, p)$, 二项分布的均值

为 np , 方差为 $np(1-p)$ 。

3. 泊松分布

随机变量 X 取值为 $1, 2, \dots$, 如果 $X=k$ 的概率为

$$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

则称 X 服从泊松分布。随机变量 X 的均值为 λ , 即 $E(X) = \lambda$, 方差也是 λ , 即 $\text{Var}(X) = \lambda$ 。

4. 均匀分布

如果 X 的分布函数如下:

$$F(X) = P(X \leq x) = \begin{cases} 0 & \text{当 } x < a \text{ 时} \\ \frac{x-a}{b-a} & \text{当 } a \leq x < b \text{ 时} \\ 1 & \text{当 } b \leq x \text{ 时} \end{cases}$$

则称随机变量 X 服从均匀分布。

可以求得概率密度函数:

$$f(x) = F'(x) = \frac{1}{b-a}$$

均匀分布的均值为:

$$E(X) = \frac{a+b}{2}$$

均匀分布的方差为:

$$\text{Var}(X) = \frac{(b-a)^2}{12}$$

5. 正态分布

正态分布的概率密度函数为:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

式中 x 取值为 $(-\infty, +\infty)$ 。

1.2 统计学基础知识

1.2.1 统计与统计学

1. 统计的含义

根据理解不同,“统计”一词有三种含义,即统计活动、统计数据和统计学。统计活动是对各种统计数据进行搜集、整理并做出相应的推断、分析的活动,通常被划分为统计调查、统计整理和统计分析三个阶段;统计数据是通过统计活动所获得的、用以表现研究现象特征的各种形式的数据;统计学则是指导统计活动的理论和方法,是关于如何搜集、整理和分析统计数据的一门科学。显然,统计的三种含义以统计数据为核心而紧密联系在一起。统计数据与统计活动是统计成果与过程的关系,统计活动与统计学则是统计实践与理论的关系。由于任何统计活动都有一定的针对性,所以不难发现,统计就是要围绕研究目的和任务,运用科学的统计方法,去获取真实客观的有关统计数据,做出必要的统计分析,以了解和认识事物的真相。所以,统计的本质就是关于为何统计、统计什么和如何统计的思想。

2. 统计学的产生和发展

统计学是随着统计的产生而产生的,而统计则起源很早,是随社会生产的发展和国家管理的需要而逐步产生和发展起来的。在原始社会,人类最初的一般计数活动蕴藏着统计萌芽。在奴隶社会,统治阶级为了征兵征税,开始了人口、土地和财产的统计。到了封建社会,统计有了一定的发展,封建君主和精明的政治家日益意识到统计对于治国强邦的重要性,统计范围有所扩大。但是由于封建经济的封闭割据和保守性,统计活动的范围受到了限制,统计方法也很不完善。到了资本主义社会,随着社会生产力的迅速