



Mc
Graw
Hill

(原书第4版)

大学经济管理
课程辅导教材

统计学原理 学习指南与习题集

Schaum's Outline of Theory and Problems of
Business Statistics
(4th Edition)

(美) 伦纳德 J. 卡茨米尔 (Leonard J. Kazmier) 著
亚利桑那州立大学
谢群 程程 译

全球销量
超过3000万册



机械工业出版社
China Machine Press



大学经济管理
课程辅导教材

(原书第4版)

统计学原理 学习指南与习题集

Schaum's Outline of Theory and Problems of
Business Statistics
(4th Edition)

(美) 伦纳德 J. 卡茨米尔 (Leonard J. Kazmier) 著
亚利桑那州立大学
谢群 程程 译



机械工业出版社
China Machine Press

本书以简洁易懂的语言介绍了商业统计中常用的描述、推断、预测和实际运用的方法。它的结构比较完整,首先包括了概率的基本知识和常用统计描述方法,然后进一步讲解了抽样分布、参数估计、假设检验、线性回归、方差分析以及非参数估计。最后,作为统计的一个直接应用,本书也引入了时间序列、决策分析和统计过程控制方面的知识。为了帮助读者更深入地理解概念和运用,书中还详解了大量例题。

本书的读者对象为经济管理类的本科生和MBA学生。

Leonard J. Kazmier. Schaum's Outline of Theory and Problems of Business Statistics.

ISBN 0-07-123679-1

Copyright © 2004, 1996, 1988, 1976 by The McGraw-Hill Companies, Inc.

Original language published by The McGraw-Hill Companies, Inc. No part of this publication may be reproduced or distributed in any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

Simplified Chinese translation edition jointly published by McGraw-Hill Education (Asia) Co. and China Machine Press.

All rights reserved.

本书中文简体字翻译版由机械工业出版社和美国麦格劳-希尔教育(亚洲)出版公司合作出版。未经出版者预先书面许可,不得以任何方式复制或抄袭本书的任何部分。

版权所有,侵权必究。

本书法律顾问 北京市展达律师事务所

本书封底贴有McGraw-Hill公司防伪标签,无标签者不得销售。

本书版权登记号:图字:01-2005-2441

图书在版编目(CIP)数据

统计学原理学习指南与习题集(原书第4版)/(美)卡茨米尔(Kazmier, L. J.)著;谢群,程程译.-北京:机械工业出版社,2005,8

(大学经济管理课程辅导教材)

书名原文:Schaum's Outline of Theory and Problems of Business Statistics.

ISBN 7-111-16752-X

I. 统… II. ①卡… ②谢… ③程… III. 统计学-高等学校-教学参考资料 IV. C8

中国版本图书馆CIP数据核字(2005)第067020号

机械工业出版社(北京市西城区百万庄大街22号 邮政编码 100037)

责任编辑:洪海山 版式设计:刘永青

北京昌平奔腾印刷厂印刷 · 新华书店北京发行所发行

2005年8月第1版第1次印刷

787mm×1020mm 1/16·19.5印张

定价:39.00元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换

本社购书热线:(010) 68326294

投稿热线:(010) 88379007

译者序

在美国念书的时候，发现他们的统计学教材内容和教学方法比国内的有趣多了，举的例子也有很多直接取自报刊杂志，所以非常贴近生活，供学生分析练习之用，意在提高学生的兴趣。尽管这样，美国的普通学生仍然是“谈统计色变”。与此同时，社会对定量分析的要求越来越高，不论是华尔街还是普通行业，不论是实际运用还是学术研究，统计都是其中最重要的分析工具之一。统计的思想本身帮助我们提高概率修养和对随机现象的认识。在国内早些时候，统计提供的更多的是对历史数据的总结，但是在国外的统计学习和应用中，人们用它来帮助了解现象的内在结构，对观察到的数据做出一定的解释，并且更进一步地，可以对未来数据进行预测。国外企业的很多商业决策都是建立在对数据的分析基础上的，并且常常有一支专门的数量分析的队伍。所以统计学习越来越重要。我们欣喜地看到，国内（尤其是商学院）在统计的教学和研究方面越来越接近国际趋势，尤其是对本科生、研究生和MBA学生的教育上。

国外的统计课本的一个特点就是非常厚，动辄800页之多。看完本书之后，我们深感它既简洁明了地处理了纷繁错综的统计概念，又不失对统计内容的完整覆盖（它甚至包括了决策分析和全面质量管理的章节）。尤其是作者花了相当的篇幅，对习题进行了比较详尽的解答，甚至对留给学生练习用的补充习题也给出了简要答案，它是一本学习、练习中不可多得的参考书。同时，我们认为也完全可以用它来作为教材，尤其对那些希望能在短期内掌握统计概要和方法的已经参加工作的人。

本书的读者对象为经济管理类的本科生和MBA学生，尤其对没有上过微积分的同学来说，其较简单的数学处理既不影响对统计方法的介绍，也比较容易为同学们所接受。作为一门实践科学，统计需要有软件来处理数据，这本书介绍了两个常见的统计软件：微软的Excel（中的数据分析师宏）和另外一个专业软件Minitab。前者能计算本书提到的绝大多数的统计程序，后者则对全面质量管理、非参数估计等有进一步的分析功能。（原书第16章有一些关于“Compustat”的软件在序列分析中的运用，考虑到国内没有这个软件，而且在国外它也并非序列分析主要软件，我们就没有进行翻译。）在这个意义上来说，本书是一本比较完备的学习指南。

在翻译本书的过程中，刘哲对Excel和Minitab的操作步骤进行了翻译和操作验证，特以致谢。

前 言

本书涵盖了商务统计学入门及中级课程中包含的描述统计学、推断统计学、决策分析以及过程管理的基本方法。

书中采用简洁明了的方式介绍了有关概念和方法，我们尽力缩短了大篇幅的解释，而将重点放在具体例子上。由于本书是专门针对那些希望运用统计技术的人，所以省略了数学推导。

本书用作教科书的补充时，大量的例题和习题将有助于更好地理解教科书中阐述的数学概念。同时它也可以作为一本出色的参考书，其简明的形式使您能非常方便地查找到需要的方法。最后，本书涵盖的内容非常完整，完全可以用作课程的教材。

在这一版中，我们做了大量的更新，加入了计算机软件的应用，包括Excel（Microsoft, Inc.版权所有）和Minitab（Minitab, Inc.版权所有）。

伦纳德 J. 卡茨米尔

目 录

译者序 前 言

第1章 分析商业数据

1.1 商务统计的定义	1
1.2 描述统计学与推断统计学	1
1.3 商务应用的种类	1
1.4 离散和连续变量	1
1.5 通过观测和调查获取数据	2
1.6 随机抽样的方法	2
1.7 其他抽样方法	3
1.8 使用Excel和Minitab产生随机数	3
习题与解答	3
补充习题	6

第2章 统计表达与图形表示

2.1 频数分布	8
2.2 组宽度	8
2.3 直方图和频数多边形	9
2.4 频数曲线	9
2.5 累积频数分布	10
2.6 相对频数分布	11
2.7 “大于等于 a , 小于 b ”类型的频数分布	11
2.8 茎叶图	11
2.9 点图	12
2.10 帕累托图	12
2.11 柱状图和线图	13
2.12 运行图	14
2.13 饼图	14

2.14 使用Excel和Minitab	14
习题与解答	15
补充习题	30

第3章 描述商业数据：位置的量度

3.1 数据集中位置的量度	33
3.2 算数均值	33
3.3 加权均值	33
3.4 中位数	34
3.5 众数	34
3.6 均值和中位数的关系	34
3.7 均值和中位数满足的数学准则	35
3.8 均值、中位数和众数的应用	35
3.9 统计过程控制中均值的应用	36
3.10 四分位数、十分位数和百分位数	36
3.11 使用Excel和Minitab	37
习题与解答	37
补充习题	41

第4章 描述商业数据：离散度的量度

4.1 数据集中离散度的量度	43
4.2 极差	43
4.3 修正极差	43
4.4 盒形图	44
4.5 平均绝对离差	44
4.6 方差和标准差	45
4.7 方差和标准差的简便计算方法	45
4.8 方差和标准差满足的数学准则	46
4.9 使用标准差描述数据	46
4.10 极差和标准差在统计过程控制中的应用	47

4.11 方差系数	48
4.12 皮尔逊偏斜系数	48
4.13 使用Excel和Minitab	48
习题与解答	49
补充习题	54

第5章 概率

5.1 概率的基本定义	56
5.2 概率的表示	56
5.3 互斥事件与非互斥事件	57
5.4 加法法则	57
5.5 独立事件、相关事件和条件概率	58
5.6 乘法法则	59
5.7 贝叶斯定理	61
5.8 联合概率分布表	61
5.9 排列	62
5.10 组合	63
习题与解答	64
补充习题	72

第6章 离散型随机变量的概率分布：二项分布、超几何分布和泊松分布

6.1 什么是随机变量	76
6.2 描述离散型随机变量	76
6.3 二项分布	77
6.4 用比例表示的二项分布变量	78
6.5 超几何分布	79
6.6 泊松分布	80
6.7 二项分布的泊松近似	81
6.8 使用Excel和Minitab	81
习题与解答	82
补充习题	91

第7章 连续型随机变量的概率分布：正态分布和指数分布

7.1 连续型随机变量	95
7.2 正态概率分布	95
7.3 正态分布随机变量的百分位点	97

7.4 二项分布的正态近似	98
7.5 泊松分布的正态近似	99
7.6 指数概率分布	100
7.7 使用Excel和Minitab	100
习题与解答	100
补充习题	108

第8章 抽样分布和均值的置信区间

8.1 总体或过程参数的点估计	111
8.2 抽样分布的概念	111
8.3 均值的抽样分布	112
8.4 中心极限定理	113
8.5 计算样本均值的概率值	113
8.6 用正态分布计算均值的置信区间	114
8.7 确定估计均值需要的样本容量	115
8.8 t 分布与均值的置信区间	115
8.9 总体均值区间估计的总结表	116
8.10 使用Excel和Minitab	117
习题与解答	117
补充习题	123

第9章 其他置信区间

9.1 使用正态分布时两均值之差的置信区间	125
9.2 t 分布与两均值之差的置信区间	126
9.3 总体比例的置信区间	126
9.4 计算估计比例时需要的样本容量	127
9.5 两总体比例之差的置信区间	128
9.6 χ^2 分布与方差和标准差的置信区间	128
9.7 使用Excel和Minitab	129
习题与解答	129
补充习题	134

第10章 总体均值的假设检验

10.1 引言	137
10.2 用临界值方法进行假设检验的基本步骤	137
10.3 用正态分布检验有关均值的假设	138

10.4 假设检验中的第一类错误和第二类错误	140
10.5 计算检验均值时需要的样本容量	142
10.6 使用 t 分布对均值进行假设检验	142
10.7 使用 P 值方法对总体均值进行假设检验	143
10.8 使用置信区间方法对均值进行假设检验	144
10.9 统计过程控制中均值的检验	144
10.10 均值假设检验的总结表	144
10.11 使用Excel和Minitab	144
习题与解答	145
补充习题	153

第11章 其他假设检验

11.1 用正态分布检验两个均值之差	156
11.2 用 t 分布检验均值之差	157
11.3 使用成对观测值时均值之差的检验	158
11.4 对总体比例的假设检验	159
11.5 检验比例时所需的样本容量	160
11.6 统计过程控制中过程比例的检验	160
11.7 对两个总体比例之差的检验	161
11.8 用 χ^2 分布对方差进行假设检验	161
11.9 检验统计过程控制中的过程波动	162
11.10 F 分布以及两总体方差相等的假设检验	162
11.11 检验原假设的其他方法	163
11.12 使用Excel和Minitab	164
习题与解答	164
补充习题	171

第12章 用 χ^2 检验分析定性数据

12.1 χ^2 检验的一般目的	174
12.2 检验拟合优度	174
12.3 两个分类变量独立性的检验(列联表检验)	175
12.4 关于比例的假设检验	176
12.5 使用计算机软件	178

习题与解答	178
补充习题	185

第13章 方差分析

13.1 检验几个总体均值之间差别的基本原理	188
13.2 单因素完全随机化设计(单因素ANOVA)	189
13.3 双因素方差分析(双因素ANOVA)	189
13.4 随机化区组设计(双因素ANOVA, 每个单元格中有一个观测值)	190
13.5 双因素完全随机化设计(双因素ANOVA, 每个单元格中有 n 个观测值)	190
13.6 其他的一些考虑	191
13.7 使用Excel和Minitab	191
习题与解答	192
补充习题	204

第14章 线性回归和相关分析

14.1 回归分析的目的和假设	206
14.2 散点图	206
14.3 拟合回归线的最小二乘方法	207
14.4 残差和残差图	208
14.5 估计量的标准误差	208
14.6 对于斜率的推断	208
14.7 条件均值的置信区间	209
14.8 因变量的个别值的预测区间	209
14.9 相关分析的目的和假设	209
14.10 决定系数	210
14.11 相关系数	210
14.12 从协方差角度理解相关系数	211
14.13 相关系数的显著性检验	212
14.14 回归分析和相关分析的缺陷和局限性	212
14.15 使用Excel和Minitab	212
习题与解答	212
补充习题	220

第15章 多元回归与相关

15.1 多元回归分析的目的和假设	222
15.2 多元回归分析中的其他概念	222
15.3 使用指示(虚拟)变量	223
15.4 残差和残差图	223
15.5 线性回归分析中的方差分析	223
15.6 多元相关分析的目的和假设	224
15.7 多元相关分析中的其他一些概念	225
15.8 多元回归分析和多元相关分析的缺陷和局限性	225
15.9 使用Excel和Minitab	225
习题与解答	226
补充习题	229

第16章 时间序列分析和商业预测

16.1 古典时间序列模型	232
16.2 趋势分析	232
16.3 循环变量分析	234
16.4 对季节变差的度量	234
16.5 运用季节调整	234
16.6 根据趋势和季节因子进行预测	235
16.7 周期预测与商业指数	235
16.8 基于移动平均的预测	236
16.9 指数平滑预测法	236
16.10 包含平滑的其他预测方法	237
16.11 使用计算机软件	237
习题与解答	237
补充习题	242

第17章 非参数统计

17.1 测量尺度	244
17.2 参数与非参数统计方法	244
17.3 随机性的队检验	245
17.4 一个样本: 符号检验	245
17.5 一个样本: 威尔科克森检验	245
17.6 两个独立样本: 曼-惠特尼检验	246
17.7 成对观测值: 符号检验	247
17.8 成对观测值: 威尔科克森检验	247
17.9 多个独立样本: 克鲁斯卡尔-沃利斯检验	247

17.10 使用Minitab	248
习题与解答	248
补充习题	254

第18章 决策分析: 损益表和决策树

18.1 损益表的结构	256
18.2 仅在概率基础上进行决策	257
18.3 仅在经济后果基础上进行决策	257
18.4 在概率和经济后果两者基础上进行决策: 期望损益准则	258
18.5 决策树分析	259
18.6 使用期望效用作为决策准则	260
习题与解答	261
补充习题	267

第19章 统计过程控制

19.1 全面质量管理	270
19.2 统计质量控制	270
19.3 过程中的变异类型	271
19.4 控制图	272
19.5 过程均值的控制图: $\bar{X}$ 图	272
19.6 用于解释 $\bar{X}$ 图的标准检验	273
19.7 过程标准差的控制图: s 图	274
19.8 过程极差的控制图: R 图	275
19.9 过程比例的控制图: p 图	275
19.10 使用Minitab	276
习题与解答	276
补充习题	282

附录A 随机数表	285
附录B 二项分布概率	286
附录C $e^{-\lambda}$ 值	289
附录D 泊松分布概率	290
附录E 标准正态分布的面积表	295
附录F t 分布的面积表	296
附录G χ^2 分布的面积表	297
附录H 概率为5%和1%的 F 分布值	298
附录I 控制图的要素	301
附录J 威尔科克森检验里的 T 临界值	302

第 1 章

分析商业数据

1.1 商务统计的定义

统计学 (statistics) 是一门用于收集、整理、分析和解释数据的技术。这些数据可能是用数字描述的定量数据,也可能是用品质特征表示的定性数据,比如消费者偏好。在商务应用中,通过更好地理解数据波动的来源,发现商务数据的模式及相互关系,统计学能帮助做出更好的决策。

1.2 描述统计学与推断统计学

描述统计学 (descriptive statistics) 对定量数据进行总结和描述使数据更易于理解。使用的方法可以是图形分析和计算分析 (参见第2、3、4章)。

例1 我们可以通过绘制一张条形图或者线图来表示一件产品在过去一年中的月销售量以使这些数据变得直观有意义 (参见2.11节)。通过计算月销售量指数,我们可以清楚地看出各个月份的相对销售量,只要比较该指数和100的差异,就可以看出该月与年平均月销售量的百分比差异。

推断统计学 (inferential statistics) 通过从一个样本观测到的数据得出对总体的一些结论。由于这些结论是在不确定的条件下得出的,我们必须引入概率这一概念。有关样本的测量值称为样本统计量 (sample statistics), 而统计总体的测量值称为总体参数 (population parameters)。如果一个总体的所有成员都被测量,这个过程就称为普查 (census)。用于过程控制时,样本抽取主要用于发现和控制输出数据的波动来源。第5~7章涵盖了概率方面的一些概念,之后的大部分章节都是这些概念在统计推断方面的应用。

例2 为了估计能使一种电子设备失效所需的电压,我们取一组这样的设备构成一个样本,逐渐升高电压直到每个设备都失效。基于这组样本的测量结果,来自同一总体的其他各个设备在各个电压水平上失效的概率就可以估计出来了。

1.3 商务应用的种类

经典统计学 (classical statistics) 主要通过分析样本数据推断总体性质。它明确排除对于数据的个人判断,并且暗含这样一个假定:样本来自于静态 (稳定的) 总体。决策分析 (decision analysis) 则将管理判断融合在统计分析中 (参见18章)。统计过程控制 (statistical process control) 则假定过程的输出可能是不稳定的,相反,由于各种可识别的原因,输出的质量随时间而变动,过程是动态的。

例3 使用经典的统计推断方法,在科学取样和市场调查后,我们可以来估计一种新产品的销量。在决策分析中,管理人员的判断将被量化并且融合到分析中。而统计过程控制则主要集中于识别在市场测试阶段销售量随时间的变动模式。

1.4 离散和连续变量

离散变量 (discrete variable) 只在一定范围数值中的一些离散的点上有观测值。在商务统计中,这些数据多在计数 (counting) 中出现,所以,这些数值一般用整数表示。连续变量可以在特定区间的任意分数点取值,连续型数据是在测量过程中产生的。

例4 离散数据的例子有：每户的人数、库存数量以及装配部件中次品个数。连续数据的例子有：一批货物的重量，设备直到第一次故障时使用的时间以及一个大的社区中平均每户人数。注意，虽然每户人数是一个离散变量，平均人数由于可以是分数而属于连续变量。

1.5 通过观测和调查获取数据

获取数据的一种方法是直接观测。在利用输出样本进行系统性估计的统计过程控制中，直接观测是采取干预的依靠。另外一种直接观测的形式是统计实验 (statistical experiment)，这里，我们控制部分或者全部可能影响研究结果的因素，以便识别造成特定结果的可能原因。

例5 为了比较两种组装方法，我们选出两组工人分别使用这两种方法，第一组的每个成员都与第二组的相对应成员在年龄和经验等方面尽量严格地匹配。

在一些情况下，直接收集数据是不可能的，只能通过被调查个体的回应获取信息。统计调查 (statistical survey) 是通过对个体的询问获取并收集数据的过程。收集数据的可能方法包括个人采访调查、电话调查以及问卷调查。

例6 州经济保障部门的分析员可能需要确定该州各商业公司雇员数量的增减计划，获取这类数据的常规方法就是对商业公司进行调查。

1.6 随机抽样的方法

随机抽样 (random sampling) 是一种目标总体中每个个体都有已知的——而且常常是相同的——机会被选中作为样本的抽样方法。这样的抽样方法可以确保样本中的个体选择是无偏的，而且为和推断有关的置信度提供了统计基础 (参见第8、9章)。随机样本也称为概率样本或者科学样本。随机抽样的四个主要方法是简单抽样、系统抽样、分层抽样和整群抽样法。

简单抽样是指目标总体中的每个个体被选中的可能都是随机的。这种随机选择有些类似抽奖中中奖数字的随机抽取。然而，在统计抽样中，常常用随机数表或者计算机提供的随机数发生器来选择随机样本，这里需要先将总体中的个体编号。

例7 附录A是一张简化的随机数表。假设我们想要从一个含有90个编号为01~90的数字的总体中得到一个10个数字的简单随机样本，我们可以闭上眼睛，在随机数表中任意指定一个起始位置，然后沿任意方向每次读取两个数字，得到我们的样本数字。假设我们从第6行，第1列开始成对读取数字，那么样本中的10个数字就是66, 06, 59, 94, 78, 70, 08, 67, 12, 65。然而，由于总体中只有90个数字，94不能被包含在样本中，所以，下一个数字11就被选入样本。任何被重复选中的数字都只在样本中出现一次。

系统抽样的随机样本是通过从总体中按顺序每隔固定的区间选取一个个体产生的，比如抽取每逢十的账号作为样本。样本中的第一个数必须是随机选取的 (可以通过随机数表选择)，对于系统抽样，需要特别注意的是总体中周期性因素的存在会导致样本结果的系统偏差。

例8 在一个为了配置足够明亮的路灯系统而进行的居民调查中，如果每第12号房屋都位于拐弯处，那么每隔12家住宅就抽取一家的系统抽样样本就会产生系统误差。在这种情况下，要不所有的被调查住户都位于拐弯处，要不就都不位于拐弯处。

在分层抽样中，总体中的个体首先被研究人员根据一种或几种重要性质分成不同的子类或者层，然后在每层中运用简单抽样或者系统抽样选取样本。这种方法可以保证样本中每个子类的个体个数都是成比例的。并且，用这种方法达到同一精确度要求所需的样本数量比简单随机抽样要小，所以它可以减小抽样成本。

例9 在一个关于学生对校园内住房看法的研究中，我们有理由相信本科生和研究生、男生和女生之间存在很大的差异。所以，应该采用分层抽样的方法，在以下四个层次中进行简单随机抽样：男性研究生、女

性研究生、男性本科生和女性本科生。

在整群抽样中，总体自然地聚成几个子类。所有子类或者说聚类被随机选取作为样本。

例10 经济安全局的分析员想要研究一个大城区的平均小时工资，但是要得到所有赚取工资者的名单是非常困难的，而得到一份该地区所有公司的名单却容易得多。所以，分析人员可以对各个公司进行随机简单抽样，而这些公司的本身代表了雇员的聚类。这样就可以获得公司付给雇员的工资数据了。

1.7 其他抽样方法

虽然非随机样本也可以作为总体的代表，但是很难事先假定这种样本是无偏的，而且从这种样本中做出的推断精度也很难从统计意义上进行描述。

判断抽样 (judgment sample) 是依赖于个人进行样本选择的抽样。样本在多大程度上代表了总体，取决于样本选择者的判断，因此无法在统计意义上进行评价。

例11 会计师根据他的判断选取他认为能够代表普遍情况的记录作为审计样本而不是随机抽取待审计记录。因此，我们无法从统计意义上评价这样的样本是否无偏，或者说在多大程度上代表了总体的情况。

正如“方便”这个词的含义，方便抽样 (convenience sample) 选取那些最容易获得的测量值或者观测值作为样本。

例12 为了研究公众对于市中心购物中心的态度，社区发展办公室在购物中心的一个出口进行了一项民意调查。调查的结果显然不可能反映那些不在购物中心的人、在购物中心却没有参加这项民意调查或者在购物中心中没被取样调查部分的人的意见。

在统计过程控制中，严格意义上的随机样本常常是不可得的。因此，为了考察输出质量的变化，每隔固定的时间就取一个小样本，这种抽样方法叫做理性分组 (rational subgroups)。我们可以认为这种抽样是在每个时点上进行的随机抽样，但是应该对这种抽样可能造成偏差的因素有充分的了解和警觉。

例13 在一次为符合最少包装量而进行的过程控制中，每隔固定的时间就取一组4包土豆并测量它们的重量。这些理性分组提供了统计基础，用来判断过程是否稳定并处于控制中，还是样本结果显示了不寻常的偏移——这时需要找到原因并做出改正。

1.8 使用Excel和Minitab产生随机数

计算机软件可以产生任意指定范围内的随机数字。习题1.10和1.11分别给出了用Excel和Minitab选择随机样本的例子。

习题与解答

描述统计学与推断统计学

1.1 指出下列哪些项目或操作与样本或抽样 (S) 有关，哪些与总体 (P) 有关：

- (a) 称为参数的成组测量值；
- (b) 使用推断统计；
- (c) 进行普查；
- (d) 通过检查其中的几箱判断购入的大批水果的质量。

解：(a) P； (b) S； (c) P； (d) S。

商务应用的种类

1.2 指出下列各种信息更适合用于经典统计推断 (CI)、决策分析 (DA) 还是统计质量管理 (PC)：

4 统计学原理学习指南与习题集

- (a) 对于一种新产品可能销量的管理性判断;
- (b) 选取组装线上每第15台汽车用于质量评估;
- (c) 对购买某种汽车模型的人简单随机抽样的调查结果;
- (d) 通过对银行账户的系统抽样检查账户是否平衡。

解: (a) DA; (b) PC; (c) CI; (d) CI。

离散和连续变量

1.3 指出下列变量属于离散变量 (D) 还是连续变量 (C):

- (a) 一包谷物的重量;
- (b) 一个轴承的直径;
- (c) 次品数量;
- (d) 一个地区接受失业补助的人数;
- (e) 在过去一个月中平均每个销售代表接触的期望客户数;
- (f) 销售量的美元数额。

解: (a) C; (b) C; (c) D; (d) D; (e) C; (f) D。

(注意: 虽然钱数是离散的, 但是当这种数量以分为单位计量时, 常常被作为连续数据处理。)

通过观测和调查获取数据

1.4 指出下列数据收集过程分别属于试验 (E) 还是调查 (S):

- (a) 一个在即将到来的选举中选民意向的民意测验;
- (b) 采访一个购物中心的顾客, 询问他们为什么在那里购物;
- (c) 通过在两个可比较地区分别采用不同方法, 比较两种不同的养老金促销方法。

解: (a) S; (b) S; (c) E。

1.5 在统计测量领域, 比如调查问卷, 可靠性 (reliability) 指测量手段的一致性, 而有效性 (validity) 指该手段的精确性。因此, 如果两组相同的对象对一份调查问卷的回答基本相同, 就可以认为这份调查问卷是可靠的。那么, 测量手段的可靠性可以保证它的有效性吗?

解: 一种测量手段的可靠性并不能保证它对于某一特定目的的有效性。一种可靠的测量手段在重复测量中产生一致性的结果, 但是可能每次测量都包含同样的错误或者偏差的成分。(参见下一题。)

1.6 参见上一题, 一种不可靠的测量手段在用于某一特定目的时可能是有效的吗?

解: 一种不可靠的测量手段用于任何目的时都不可能是有效的。没有了可靠性, 我们得到的测量结果是不一致的。我们用对气枪打靶的类比来解释这个概念。在打靶时, 如果弹孔集中在目标附近, 就说它具有可靠性 (一致性)。在这种情况下, 通过视觉调整使弹孔集中于靶心就可以改进有效性 (精确性)。但是, 如果弹孔十分分散, 说明缺乏可靠性, 那么, 无论怎么进行视觉调整都无法得到好成绩。

随机抽样的方法

1.7 为了进行统计推断, 我们要设计一个具有代表性的样本。那么, 为什么统计推断中只允许使用随机样本呢?

解: 没有一种抽样方法可以保证样本的代表性。我们能做到的只能是通过随机 (概率) 抽样来避免一致性的或者系统性的误差。尽管随机样本很少能完全代表它来自的那个目标总体, 采用这种抽样方法可以保证样本和总体之间的差异只由偶然因素产生。

1.8 一家石油公司在—个地区进行试验, 希望找出影响消费者选择加油站的因素, 为此, 需要取得该地区所有拥有汽车居民的姓名、地址和其他可获得的个人信息的—份清单。描述如何通过本章介绍的四种方法获得这份清单的—个样本。

解: 使用简单抽样, 可以对名单中的姓名顺序编号, 然后用随机数表选择随机数以抽取样本。使用系统抽样,

可以从名单中每 n 个人(比如每5个)抽一个组成样本,第一个人需要从前五个中随机抽取,等等。使用分层抽样,我们可以按照汽车类型、价值以及车主的性别、年龄对名单上的人进行分类,然后在每层中进行简单随机抽样或者系统抽样。使用整群抽样,我们可以在测试地区随机抽取几个街区,对这些街区的车主进行调查,因为分群是基于地理位置的,所以这种整群抽样也称为地区抽样。

其他抽样方法

1.9 指出下列样本分别属于判断抽样(J)、方便抽样(C)还是理性分组抽样(R):

- 在生产过程中每隔20分钟取5只灯泡来判断它们对于高电压的抵抗能力;
- 一家饮料公司通过在公司所在城市的酒馆内进行品尝试验,来估计消费者对一种计划推出的无酒精啤酒的反应;
- 一项选民意向测试从不同地理位置中挑选出选民进行调查,认为从不同位置选取得的人代表了整个地区选民的意向。

解: (a) R; (b) C; (c) J。

使用Excel和Minitab产生随机数

1.10 州经济学家想要从位于该州某特定部分的435家公司中随机抽取30家。方便起见,这些公司用ID号1~435标识。使用Excel产生研究中需要的30家样本公司的ID号。

图1-1显示了Excel的输出,30家样本公司的ID号分别列于第2列的第1~30行。随机抽样的性质决定了你选择的样本公司是不同的。按照下面步骤使用Excel选择这个例子中所需的 $n=30$ 的随机样本:

- 打开Excel,在A列产生数字1~435,你可以先在单元格A1中输入1,激活单元格A1(比如,可以点击其他地方再回到A1点击),点击编辑→填充→序列,打开序列对话框。
- 选择序列产生在列,步长值填1,终止值填435。点击确定,数字1~435就出现在A列中了。
- 为了产生30个样本数据,点击工具→数据分析→抽样。在输入区域中填入\$A\$1:\$A\$435,选择抽样方法为随机,样本数为30,输出区域为\$B\$1。点击确定,随机选取的公司的ID号码就会出现在B列的1到30行。

1.11 某州的经济学家想要从位于该州某特定部分的435家公司中随机抽取30家。方便起见,这些公司用ID号1~435标识。使用Minitab产生研究中需要的30家样本公司的ID号。

图1-2显示了Minitab的输出,30家样本公司的ID号分别列于C2列的第1~30行。随机抽样的性质决定了你选择的样本公司是不同的。按照下面步骤使用Minitab选择这个例子中所需的 $n=30$ 的随机样本:

- 打开Minitab。点击Calc(计算)→Make Patterned Data(产生成组数据)→Simple Set of Numbers(简单数集)。然后选择Store patterned data in(保存成组数据于): C1, From first value(起始值): 1, To last value(最末值): 435, In steps of(步长为): 1。点击OK(确定),数字1~435就会出现在列C1中。
- 为了产生30个样本数据,点击Calc(计算)→Random Data(随机数)→Samples from Columns(样本从列产生)。然后选择Sample(样本量):

1	186
2	70
3	371
4	307
5	343
6	1
7	427
8	292
9	261
10	126
11	392
12	316
13	12
14	184
15	44
16	107
17	435
18	338
19	178
20	14
21	382
22	349
23	269
24	79
25	330
26	76
27	109
28	317
29	115
30	247
31	
32	
.	
.	
.	
433	
434	
435	

图1-1 一个简单随机样本的Excel输出

6 统计学原理学习指南与习题集

30, from column[s] (产生列): C1; Store samples in (保存样本于): C2。点击OK (确定), 随机选择的ID号码就会出现在C2列的1~30行。

补充习题

描述统计学与推断统计学

1.12 指出下列哪些项目或操作与样本或抽样 (S) 有关, 哪些与总体 (P) 有关:

- (a) 宇宙;
- (b) 称为统计量的成组测量值;
- (c) 概率概念的应用;
- (d) 对装配线上每个产品的检查;
- (e) 对装配线上每10个产品的检查。

解: (a) P; (b) S; (c) S; (d) P; (e) S。

商务应用的种类

1.13 指出下列各种信息更适合用于经典统计推断 (CI), 决策分析 (DA) 还是统计过程控制 (PC):

- (a) 对一个学术组织现有成员进行问卷调查的结果;
- (b) 按月收集的由消费者对汽车服务部门做出的排名;
- (c) 投资分析家对于新兴企业的排名;
- (d) 一个大城区雇员工资数据的样本。

解: (a) CI; (b) PC; (c) DA; (d) CI。

离散和连续变量

1.14 指出下列变量属于离散变量 (D) 还是连续变量 (C):

- (a) 库存产品数量;
- (b) 流动资产对流动负债的比率;
- (c) 货物总重量;
- (d) 按个计量的货物数量;
- (e) 一条收费公路上的交通量;
- (f) 公司年会的出席人数。

解: (a) D; (b) C; (c) C; (d) D; (e) D; (f) D。

通过观测和调查获取数据

1.15 指出下列数据收集过程分别属于试验 (E) 还是调查 (S):

- (a) 培训航空机票代理人的新方法与传统方法的比较结果;
- (b) 通过让两组可比较的孩子分别使用两种不同的组装说明组装玩具来比较这两种组装说明;
- (c) 一份产品评价杂志给它的订阅者邮寄调查问卷, 请他们为近期购买的产品排名。

解: (a) E; (b) E; (c) S。

随机抽样的方法

1.16 指出下列抽样方法属于简单抽样 (R) 还是系统抽样 (S):

	C1	C2
1	1	425
2	2	256
3	3	403
4	4	322
5	5	349
6	6	124
7	7	395
8	8	315
9	9	334
10	10	381
11	11	72
12	12	100
13	13	236
14	14	303
15	15	253
16	16	213
17	17	34
18	18	365
19	19	58
20	20	300
21	21	10
22	22	83
23	23	211
24	24	279
25	25	416
26	26	311
27	27	325
28	28	110
29	29	275
30	30	179
31	31	
32	32	
.	.	.
.	.	.
.	.	.
.	.	.
433	433	
434	434	
435	435	

图1-2 一个简单随机样本的Minitab输出

- (a) 使用随机数表选择一个进入游乐园者的样本；
- (b) 对进入游乐园者每隔100人采访一人，随机选取第55个进入者作为起始值。

解：(a) R； (b) S。

1.17 对于下列基于分组的取样方法，指出是用了分层抽样 (St) 还是整群抽样 (C) 方法：

- (a) 估计不同生活区居民的投票偏好；
- (b) 在确信不同年龄和性别的消费者有很大不同的基础上研究消费者意向。

解：(a) C； (b) St。

其他抽样方法

1.18 指出下列样本分别属于判断抽样 (J)、方便抽样 (C) 还是理性分组 (R)：

- (a) 一个房地产评估员在一个地区选择了一个似乎可以代表该地区所有房屋的样本，以便估计该地区的房价水平；
- (b) 在一家电池制造厂，每隔半个小时测量一次电池寿命以确保产品符合要求；
- (c) 一家快餐店让公司雇员评价一种新鸡肉汉堡的味道和感知价值。

解：(a) J； (b) R； (c) C。

使用计算机软件产生随机数

1.19 一个审计员想要从一家大公司的5 250个应收款账户中随机抽取50个。这些账户从0001到5250顺序编号。使用合适的计算机软件生成所需的50个随机数字。

第 2 章

统计表达与图形表示

2.1 频数分布

频数分布 (frequency distribution) 是将变量的所有可能值分成若干组, 记录落入每组中的观测变量个数得到的表格。按频数分布组织的数据称为分组数据 (grouped data), 相反, 未分组数据 (ungrouped data) 将变量的每个观测值都列出来。

例1 表2-1列出的是一份周工资的频数分布。注意工资数额按照离它最近的整数计算。而当小数部分“恰好为0.5”(在这里即恰好为\$0.50)时, 习惯上将它近似为最近的偶数, 所以在数据分组过程中, 周工资\$259.50近似计为\$260。

表2-1 100名工人周工资的频数分布

周工资 (\$)	工人数 (f)
240~259	7
260~279	20
280~299	33
300~319	25
320~339	11
340~359	4
总计	100

2.2 组宽度

对于频率分布中的每一个组, 申明组区界限 (stated class limits) 的上下界表明了该组中涵盖的数据值。(参见表2-1的第一列。)相反, 精确分组界限 (exact class limits), 或称为分组边界 (class boundaries), 是可以将连续变量按其量度值分别归入相邻两组的确切数值。精确分组界限可以通过相邻组的标明的上限和下限值的中间点来确定。分组宽度 (class interval) 表示一组中包含的变量范围, 用该组的精确上限减去下限可以得到。当精确组边界无法确定时, 可以用相邻的下一组的申明下限值减去该组的申明下限值得到组宽度。最后, 为了某些特别的目的常用分组中值 (class midpoint) 表示一个组, 用组宽度的一半加上该组的精确下限值就可以得到分组中值。

例2 表2-2列出了表2-1中频数分布的精确组边界和分组中值。

例3 分别采用两种方法, 表2-2中第一组的组宽度为: $\$259.50 - \$239.50 = \$20$ (用该组的精确上限减去精确下限值), $\$260 - \$240 = \$20$ (相邻的下一组的申明下限值减去该组的申明下限值)。

表2-2 100名工人的周工资

周工资 (分组界限) (\$)	精确分组界限 ^① (\$)	分组中值 (\$)	工人数
240~259	239.50~259.50	249.50	7
260~279	259.50~279.50	269.50	20
280~299	279.50~299.50	289.50	33
300~319	299.50~319.50	309.50	25
320~339	319.50~339.50	329.50	11
340~359	339.50~359.50	349.50	4
总计			100

① 一般来说, 精确分组界限比申明的分组界限只精确一位, 但是由于对于货币来说比“每美元”更精确的度量单位就是“每美分”, 所以这里表示时精确到两位小数。

一般来说, 为了计算方便, 一个频数分布的各组宽度都取相同值。下面的公式可以用来确定近似的组宽度:

$$\text{近似组宽度} = [(\text{分组数据中的最大值}) - (\text{分组数据中的最小值})] / \text{期望分组数} \quad (2-1)$$