



Statistics

21世纪统计学系列教材

Statistics

统计学 (第二版)

金勇进 编著

中国人民大学出版社

014041956

C8
176-2



Statistics 21 世纪统计学系列教材



Statistics

统计学 (第二版)

金勇进 编著



北航

C1729521

中国人民大学出版社
· 北京 ·

C8
176-2

图书在版编目 (CIP) 数据

统计学/金勇进编著. —2 版. —北京: 中国人民大学出版社, 2014. 4
21 世纪统计学系列教材
ISBN 978-7-300-18660-3

I. ①统… II. ①金… III. ①统计学-高等学校-教材 IV. ①C8

中国版本图书馆 CIP 数据核字 (2014) 第 060698 号

21 世纪统计学系列教材
统计学 (第二版)
金勇进 编著
Tongjixue

出版发行	中国人民大学出版社	邮政编码	100080
社 址	北京中关村大街 31 号		
电 话	010-62511242 (总编室)	010-62511770 (质管部)	
	010-82501766 (邮购部)	010-62514148 (门市部)	
	010-62515195 (发行公司)	010-62515275 (盗版举报)	
网 址	http://www.crup.com.cn		
	http://www.ttrnet.com (人大教研网)		
经 销	新华书店	版 次	2010 年 1 月第 1 版
印 刷	北京密兴印刷有限公司		2014 年 4 月第 2 版
规 格	185 mm×260 mm 16 开本	印 次	2014 年 4 月第 1 次印刷
印 张	19 插页 1	定 价	32.00 元
字 数	402 000		

版权所有 侵权必究 印装差错 负责调换

第二版前言

统计学被称为一门“有关数据的科学”，它的作用、价值和社会公认度由其上升到一级学科而得到进一步彰显，这也是学科建设的正确定位和历史必然。当今，各种类型的数据量正在以几何级数形式增长，大数据时代的大门已经慢慢开启。在这样的背景下，学习和掌握一些统计学知识，具备运用统计方法描述和分析数据的基本能力，无疑是提升人生价值的一个重要环节。

本书是在第一版基础上修订而成的，除了保持第一版的内容体系和写作特点外，本书增添和改进的方面主要有：

1. 为了使内容的表现方式更为活泼，本版中增添了“小词典”、“知道吗”、“小常识”、“想一想”、“试一试”等栏目，将一些知识要点以启发方式提出，结合讲述的内容引导读者进一步思考和实践。

2. 为了便于使用计算机处理和分析数据，本版中采用了应用领域更为广泛的 Excel 软件，在各章最后，结合相关内容，给出使用 Excel 的具体操作步骤，配以图示，方便读者学习和使用。

3. 在本书中我们增添了一些著名统计学家的生平和学术贡献，以供读者了解。

4. 为配合第二版教材，我们还将推出与之配套的习题和解答（电子版），以方便学生和教师使用。

本书是在笔者的两名博士研究生胡丹丹、张喆的协助下完成的。他们做了大量工作，包括文字初稿的重新梳理、书中一些素材的收集，以及 Excel 操作步骤的介绍。对他们的辛勤劳动，我表示深深的谢意。

由于笔者水平有限，尽管十分努力，但错误和疏漏在所难免。对于书中不足，恳请各位专家和读者提出宝贵意见，以便进一步修订。

金勇进

第一版前言

翻开报纸，打开电视，登录网络，我们就会发现越来越多的数字，我们正生活在数字的海洋中。其实，每个人的生活都与这些数字息息相关，也许我们会想：

- 股市有回暖之势，股价指数不断上扬，想要投资现在是否应该进入？
- 官方公布全国职工平均工资比上年同期增长 18%，我的工资上涨了多少？
- 有关方面公布的央视春节联欢晚会调查中观众满意度达到 83%，这个数字是否可信？

● 类似的数据比比皆是，在好奇之余，我们不禁要询问，这些数字是怎么得出的，它说明了什么，怎样对这些数字进行恰当的分析？

统计学是一门关于数据的科学，是关于数据搜集和数据分析的科学。对于统计专业的学生，毫无疑问，统计学课程是他们最重要的专业基础课；对于非统计专业的学生，统计学课程也非常重要，它是分析和解读数据最重要的方法和工具。具备了统计学的知识，我们就可以在数字的海洋中遨游，并发现数字背后许多有趣的现象和规律。

统计学与数学有联系，因为分析数据、寻找数字背后的规律、用已有数据对未知现象进行推论、对未来的情况进行预测，都需要利用统计模型。但统计学不是数学，数学研究的是抽象的数量关系，采用的是演绎思维；统计学研究的是具体的数量关系，具有不确定性，主要采用归纳思维。本书是统计学的基础性教材，笔者在写作过程中尽量避免复杂的数学推导，具有高中水平的读者就可以读懂。

全书共十二章，具体内容如下：

第一章引论，主要讨论了数据的类型，介绍了统计学中的最基本概念和几种最常用的统计软件，使读者对什么是统计学有了一个大致的了解。

第二章数据的搜集，主要介绍统计数据是怎么来的。其实，数据的来源主要有两个，通过调查和实验可以获得不同特征的数据，而通过误差分析，就可以从某个角度解读数据的质量。本章的核心是通过什么样的技术手段来获取高质量的统计数据。

第三章和第四章是姐妹章，名称分别为“数据的描述 1——数据的直观显示”和“数据的描述 2——重要的统计量”。第三章主要介绍了统计表和统计图，不同统计表、

统计图的特征和应用场合,以及如何制作规范的统计表、统计图;第四章介绍了一些重要的统计量以及每个统计量的作用和特点。数据的描述既是统计分析的一个组成部分,也为后面要讲到的统计推断做了铺垫。

第五章概率和概率分布,这是统计推断的理论基础。在高等学校中,概率论课程是一门重要的基础课程,学过概率论课程的读者可以略过本章。

第六章参数估计,本章讲授利用样本统计量对总体参数进行估计的方法,包括参数估计的基本原理,进行点估计、区间估计的方法,以及进行区间估计时确定合适样本量的方法。参数估计是统计推断的重要内容。

第七章假设检验,这是统计推断的另一个重要内容。该方法先对研究总体的参数提出假设,然后利用样本数据检验事先所做假设是否合理,实质上是通过判断样本信息与原假设是否有显著性差异来确定原假设的可信度。

第八章列联分析,这是统计推断方法在定性数据上的应用。列联分析包括拟合优度检验和独立性检验,本章对这些定性数据的检验内容进行了比较详细的讨论。

第九章方差分析,这种统计方法应用于分析分类型变量与数值型变量相互之间的关系,是假设检验方法在不同数据类型应用上的延伸。本章主要介绍了单因素方差分析和双因素方差分析,有助于加深我们对统计思想的理解。

第十章相关与回归,主要讨论变量之间的关系。本章通过统计建模的方法,将变量之间联系的密切程度数量化,并通过求解模型系数,实现统计预测的功能。本章内容有非常广阔的应用空间。

上面的第六章至第十章都是统计推断中的经典内容,而第十一章和第十二章主要是对动态统计数据描述和分析的方法。

第十一章时间序列分析,本章内容既包括对动态数据进行描述的方法,也包括对动态数据进行统计推断的方法,本章介绍的统计方法也可用于预测。

第十二章指数,这是另一类描述事物动态变化的方法和工具,在社会经济生活中,指数方法有着广泛的应用。本章主要介绍指数理论,包括指数的编制、指数的分解以及如何利用指数进行分析,本章还介绍了经济生活中几种重要的指数。

本书的写作具有如下几个特点:

(1) 尽量通俗易懂,避免复杂的数学推导。统计学中有许多定理,而不同的统计方法又有不同的数学性质。本书在写作时,尽量以通俗易懂的方式,采用逻辑叙述和分析的方法,解释公式背后的统计思想,避免枯燥的数学公式推导。一些重要且必要的公式推导,均放到各章最后,并以附录形式表现,既使略去这些推导,也不妨碍读者对全书的理解。

(2) 计算机的迅速发展,解决了统计计算中的诸多问题,使不同的统计方法有了更广阔的应用空间。面对这种情况,本书在相应的计算部分都应用了计算机,同时介绍了统计软件的操作步骤和输出结果,而不必进行繁杂的手工计算。本书的统计软件以使用最广泛的 SPSS 软件为模版,相应计算机操作的具体步骤均在附录中说明,以不影响全书正文表达的紧凑。



(3) 本书的写作以应用为背景,在统计方法介绍中突出应用的特点,同时强调读者的动手能力,各章均有一些思考题和计算题,并附之计算题的参考答案,供读者在练习中参考。

本书可以作为非统计专业统计学课程的教学用书,也可以作为其他人员学习统计方法的参考书。

本书是集体劳动的成果,由我和我的研究生们共同完成。其中,金勇进负责全书编写大纲的设计,并负责第一章和第八章的写作;其余各章由我的学生分别提供初稿,具体为谢佳斌(第一章与金勇进合作,第六章),张璞(第二章,第九章与汤琰合作),陶然、王俊合作(第三章,第四章,第十一章),艾小青(第五章),汤琰(第七章,第九章与张璞合作),黄媛(第十章),孙欣(第十二章),金勇进对全书进行了最后的修改和总纂,张璞为全书进行了技术处理。在此,向这些同学表示深深的谢意!同时,感谢中国人民大学出版社为出版本书所付出的辛勤劳动。

尽管我们十分努力,但书中一定会有不少错误和缺憾。对于书中的不足,恳请各位专家和读者提出宝贵意见。

金勇进

目 录

第1章 引 论	1
1.1 统计数据与统计学	1
1.2 一些基本概念	5
1.3 统计与统计软件	9
本章小结	10
习 题	10
第2章 数据的收集	12
2.1 数据来源	12
2.2 数据误差	20
2.3 数据文件	25
本章小结	26
习 题	27
第3章 数据的描述 1——数据的直观显示	29
3.1 用统计表描述数据	29
3.2 用统计图描述数据	36
本章小结	46
附 录	47
习 题	53
第4章 数据的描述 2——重要的指标	56
4.1 集中趋势的度量	56
4.2 离散趋势的度量	67
4.3 偏态与峰度的度量	73
4.4 数据的标准化处理	77
本章小结	79



附 录	79
习 题	80
第 5 章 概率和概率分布	83
5.1 概率的问题	83
5.2 离散变量的概率分布	92
5.3 连续变量的概率分布	97
本章小结	106
习 题	106
第 6 章 参数估计	108
6.1 抽样分布	108
6.2 点估计	110
6.3 区间估计	112
6.4 一个总体参数的区间估计	113
6.5 两个总体参数的区间估计	116
6.6 样本量	119
本章小结	122
附 录	123
习 题	124
第 7 章 假设检验	126
7.1 假设检验基本问题	126
7.2 一个总体参数的假设检验	131
7.3 两个总体参数的假设检验	136
本章小结	146
附 录	146
习 题	148
第 8 章 列联分析	150
8.1 定性数据与列联表	150
8.2 拟合优度检验	153
8.3 独立性检验	158
8.4 列联表中的相关测量	160
8.5 列联分析中需注意的问题	164
本章小结	167
附 录	167



习 题	170
第 9 章 方差分析	172
9.1 基本问题	172
9.2 单因素方差分析	176
9.3 双因素方差分析	181
本章小结	187
附 录	187
习 题	190
第 10 章 相关与回归	192
10.1 相关分析	192
10.2 回归分析	197
10.3 用回归进行预测	209
10.4 logistic 回归	212
本章小结	215
附 录	215
习 题	218
第 11 章 时间序列分析	221
11.1 时间序列的描述	221
11.2 时间序列的分解	230
11.3 时间序列的平滑法	243
11.4 时间序列的 ARMA 模型	249
本章小结	250
附 录	251
习 题	254
第 12 章 指 数	258
12.1 基本问题	258
12.2 总指数	262
12.3 指数体系	274
12.4 几种常用的指数	281
本章小结	288
习 题	289
参考文献	291

“初步核算，2013 年上半年国内生产总值 248 009 亿元，按可比价格计算，同比增长 7.6%。其中，一季度增长 7.7%，二季度增长 7.5%。分产业看，第一产业增加 18 622 亿元，增长 3.0%；第二产业增加 117 037 亿元，增长 7.6%；第三产业增加 112 350 亿元，增长 8.3%。从环比看，二季度国内生产总值增长 1.7%。”^①

我们每天在网络上、在报纸和杂志中都会看到类似的新闻，它们中间包含着大量的数据和图表。从中可以看出统计在日常生活中的作用越来越大，而且随着计算机的普及和发展，掌握统计理论和知识是人们必备的素质。通过对本章的学习，你将了解统计的基本含义、统计学中的基本概念和常用的统计软件。

1.1 统计数据与统计学

1.1.1 统计数据

我们每天都可以在报纸上、网络上和电视中看到或听到各种数据。例如，“2013 年二季度企业景气指数为 120.6，比一季度低 5.0 点，但仍明显高于 100 的景气临界值，企业运行仍处在景气区间。其中，反映企业当前景气状态的即期企业景气指数为 117.0，比一季度低 2.0 点；反映企业未来景气预判的预期企业景气指数为 123.0，比一季度低 6.9 点。二季度，企业家信心指数为 117.0，比一季度低 5.4 点。”又如北美票房数据（见图 1—1）。

^① 数据来源于国家统计局。

北美 2013年第4周（1月25日-27日）							
本周	上周	周数	片名	出品公司	周末票房	累计票房	院线
1	W	1	《女巫猎人》Hansel and Gretel: Witch Hunters	派拉蒙	\$19,000,000	\$19,000,000	3,372
2	1	2	《妈妈》Mama	环球	\$12,860,000	\$48,648,000	2,160
3	3	11	《乌云背后的幸福线》Silver Linings Playbook	温斯坦	\$10,000,000	\$69,465,000	2,641
4	2	6	《猎杀本·拉登》Zero Dark Thirty	索尼	\$9,800,000	\$69,904,000	2,946
5	W	1	《帕克》Parker	FD	\$7,000,000	\$7,000,000	2,927
6	7	5	《被解放的姜戈》Django Unchained	温斯坦	\$5,005,000	\$146,295,000	3,102
7	W	1	《电影43》Movie 43	Rela	\$5,000,000	\$5,000,000	2,027
8	4	3	《匪帮传奇》Gangster Squad	华纳	\$4,200,000	\$39,647,000	3,103
9	5	2	《破碎之城》Broken City	福克斯	\$4,000,000	\$15,270,000	2,927
10	8	5	《悲惨世界》Les Miserables (2012)	环球	\$3,912,000	\$137,237,000	2,927

图 1—1 北美票房数据统计图

资料来源：<http://ent.163.com/special/2013pfb/>.

类似的还有股票行情、物价指数、房价波动、汇率变化，乃至交通事故数、死亡人数等各种数据。数据是统计的起点，数据绝大多数是由数字构成的，但是数据不能单纯地理解为数字。我们可以把数据理解为有内涵和逻辑结构的数字。例如，以 40 为例，数字 40 本身没有什么意义，当我们听说某人的体温为 40°C 时，这个人肯定是发烧了，因为人的正常体温平均在 $36^{\circ}\text{C}\sim 37^{\circ}\text{C}$ 。统计就是利用数据来进行分析和推断的，因此有必要了解数据的类型。

统计数据按照所采用计量尺度的不同可划分为三种类型（见图 1—2）。一是**数值型数据**（quantitative data），是指用数值尺度测量的观察值。例如，GDP 等宏观经济运行数据，每天进出海关的旅游人数，某地流动人口的数量等。数值型数据的表现就是具体的数值，统计处理中的大多数数据都是数值型数据。二是**分类型数据**（categorical variable），是指对数值进行分类的结果，例如人口按性别分为男、女两类，受教育程度也可以按不同类别来区分。这种数据通常用频数（率）表示分类的结果，例如某地区男性常住人口占 52%，女性占 48%。三是**顺序型数据**（ordinal data），是指数据不仅是分类的，而且类别是有序的，例如满意度调查中的选项有“非常满意”、“比较满意”、“比较不满意”、“非常不满意”等。在这三类数据中，数值型数据由于说明了事物的数量特征，因此可归为**定量数据**（quantitative data）；分类型数据和顺

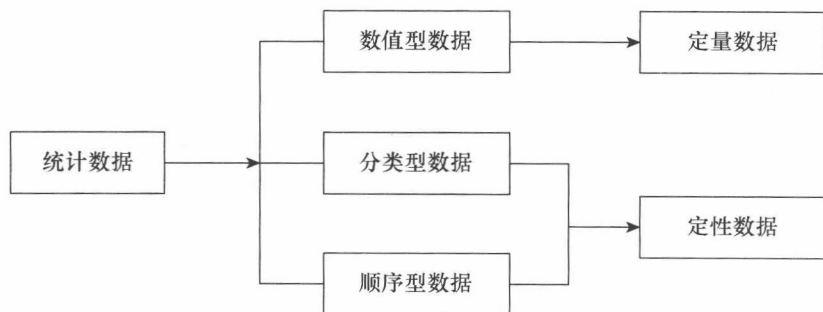


图 1—2 统计数据的分类

序型数据由于定义了事物所属的类别,说明了事物的品质特征,因此可统称为**定性数据**(qualitative data)。区分数据的类型非常重要,这是因为不同类型的数据在一些情况下,需要运用不同的统计方法进行处理。

试一试

在健康体检中会检查很多项目,试分析下面设计的数据属于哪种类型?

身高(厘米)、体重(公斤)、血压(毫米汞柱)、辨色(是色盲/不是色盲)、有无眼疾(有/无)、肝功能(阴性/阳性)。

此外,还可以从其他角度对统计数据进行划分。

按照收集方法的不同,可将统计数据分为**观测数据**(observation data)和**试验数据**(experimental data)两类。观测数据在没有对事物进行人为控制的条件下得到,是通过调查或观测而收集到的数据,主要集中在社会经济领域。例如,通过对商品零售价格变动水平的测量可以得到商品零售价格指数,通过对股票价格变动水平的测量可以得到股票价格指数。试验数据指的是通过在试验中控制试验对象而收集到的数据,主要集中在自然科学领域。比如某种新型电池的使用寿命、一种新型降压药疗效的试验数据等。

按照是否与时间相联系,还可以将统计数据分为**截面数据**(cross-sectional data)和**时间序列数据**(time series data)。表1—1中的数据为截面数据,描述了现象在某一时刻的变化情况。表1—2中的数据则为时间序列数据,描述了现象随时间变化的情况。

表 1—1 2013 年 5 月汽车销量排行榜

厂商	品牌	车型	本月销量	本年累计
上海大众	大众	大众朗逸	30 534	185 957
上海通用	别克	别克凯越	24 756	124 832
长安福特	福特	福克斯	30 856	187 605
东风日产	日产	日产轩逸	19 428	93 986
一汽大众	奥迪	奥迪 A4L	9 856	58 960

资料来源:汽车销量网, <http://www.qichexl.com>。

表 1—2 2013 年 1—5 月我国全社会客货运输量数据

月份	1	2	3	4	5
客运量(亿人)	33.23	34.00	32.39	32.36	32.72
货运量(亿吨)	33.68	30.15	34.91	35.64	35.27

资料来源:国家统计局, http://www.stats.gov.cn/tjsj/jdsj/t20130626_402907403.htm。

1.1.2 统计学

统计学是一门收集、整理和分析数据的科学,只要有数据的地方,就有统计学的



应用。收集数据研究如何得到数据，与之对应的是统计学中的抽样调查和试验设计等理论；整理数据指的是将数据用图或表的形式展现出来，与之对应的是描述统计的方法；分析数据指的是选择适当的统计方法研究数据，并从数据中提取有用的信息进而得出结论，更多地对应于推断统计的理论与方法，包含参数估计、假设检验、相关分析、回归分析、时间序列分析等诸多内容。

统计学的应用领域非常广泛，它是一门几乎适用于所有学科领域的关于通用数据分析方法的科学。无论是学术研究、政府管理，还是公司或企业的生产经营管理，都离不开统计方法的应用。表 1—3 列出了统计学的一些应用领域，可以看到，统计无处不在。

表 1—3

统计的应用领域

Actuarial work (精算)	Hydrology (水文学)
Agriculture (农业)	Industry (工业)
Animal science (动物学)	Linguistics (语言学)
Anthropology (人类学)	Literature (文学)
Archaeology (考古学)	Manpower planning (劳动力计划)
Auditing (审计学)	Management science (管理科学)
Crystallography (晶体学)	Marketing (市场营销学)
Demography (人口统计学)	Medical diagnosis (医学诊断)
Dentistry (牙医学)	Meteorology (气象学)
Ecology (生态学)	Military science (军事科学)
Econometrics (经济计量学)	Nuclear material safeguards (核材料安全管理)
Education (教育学)	Ophthalmology (眼科学)
Election forecasting and projection (选举预测和策划)	Pharmaceutics (制药学)
Engineering (工程学)	Physics (物理学)
Epidemiology (流行病学)	Political science (政治学)
Finance (金融)	Psychology (心理学)
Fisheries research (水产渔业研究)	Psychophysics (心理物理学)
Gambling (博彩业)	Quality control (质量控制)
Genetics (遗传学)	Religious studies (宗教研究)
Geography (地理学)	Sociology (社会学)
Geology (地质学)	Survey sampling (抽样调查)
Historical research (历史研究)	Taxonomy (分类学)
Human genetics (人类遗传学)	Weather modification (气象改善)

资料来源：贾俊平、何晓群、金勇进：《统计学》，4 页，北京，中国人民大学出版社，2007。

当今世界，随着信息技术，特别是数据库技术的发展，社会各行业和领域收集、存储数据的能力有了很大的提升，同时也积累了大量的数据，这为统计学自身理论的进一步发展和在社会各领域中的更广泛应用提供了机遇。在知识经济和信息时代，统计学具有十分广阔的前景。

1.1.3 统计学的类型

统计学有两个分支，分别是描述统计 (descriptive statistics) 和推断统计 (infer-

ential statistics)。

小词典

描述统计是统计学的一个分支，主要指收集、分析和展示数据的方法。

日常生活中遇到的数据大多是杂乱无章的，描述统计的目的就是利用图形、表格和指标等形式展示数据，帮助读者理解数据要表达的含义。这些内容在后面的第2章、第3章、第4章会有详细介绍。

小词典

推断统计是统计学的一个分支，主要指根据收集到的样本对总体进行推断的方法。

描述统计方法能够实现一些基本目标，但是受到人力、物力和财力的影响，在大多数情况下，我们不能获得全部总体的资料，因此我们需要利用得到的部分样本信息来对总体进行推断，这就是推断统计。可见推断统计在实际生活中有重要的应用，这部分内容在本书第5章、第6章、第7章会为大家详细介绍。

你知道吗

陈希孺 (1934—2005)，国际著名数理统计学家、中国科学院院士、《中国科学院研究生院学报》主编、中国科学院研究生院教授，出生于长沙市北湘江之滨。研究领域主要为线性模型、 U 统计量、参数估计与非参数密度、回归估计和判别等数理统计学若干分支。他长期在中国科学技术大学工作，影响并培养了不少数理统计学家。

1.2 一些基本概念

1.2.1 随机性和规律性

许多社会现象是随机的，带有不确定性。例如，买一注“双色球”福利彩票，能否中奖是不确定的；在耕好的土地中播下1000颗种子，有多少会发芽也是不确定的。也有许多社会现象是有规律的，比如，人口密度高的地方犯罪率通常也比较高。

现实中，随机性与规律性并非完全对立，社会现象通常是随机性和规律性的有机结合，随机之中带有规律性。抛一枚质地均匀的硬币，可能出现正面朝上，也可能出现反面朝上，其结果是随机的。但是，随机之中又有规律，如果重复抛该硬币若干次，则硬币正面朝上出现的频率会随着重复次数的增加逐渐趋于0.5这个数值。图1—3记录了500次抛硬币试验中正面朝上出现频率的波动情况。在重复次数较少时，



该频率波动剧烈；随着重复次数的增加，该频率波动的幅度在逐渐减小，并逐渐向 0.5 靠拢。

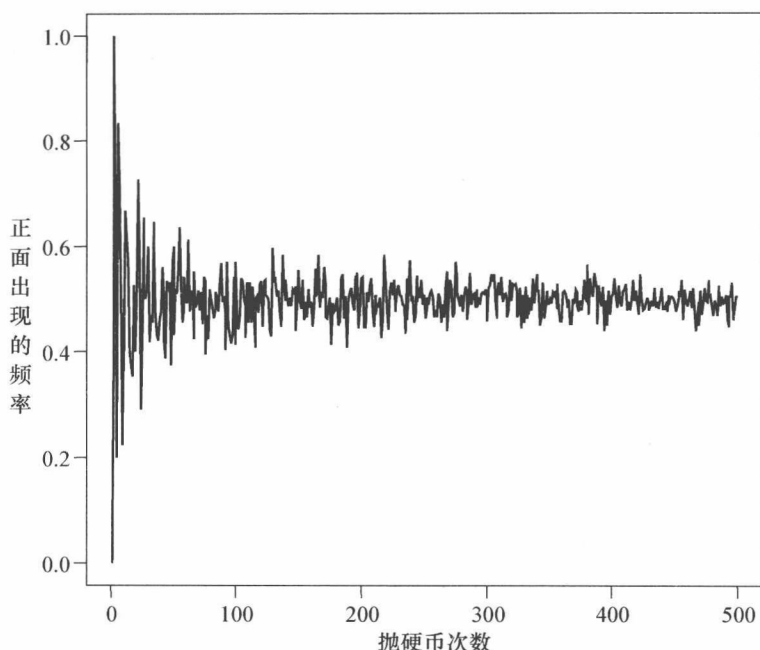


图 1—3 抛一枚质地均匀的硬币，正面朝上出现频率的变化趋势

类似的例子还有很多。你很难预测股价在某一天是上涨还是下跌，但是，作为国民经济的晴雨表，在经济向好的那些年份，股票价格指数总的趋势是上涨的。人的身高很难预先确定，受生活习惯、基因、体育锻炼等因素的影响，存在较大的个体差异，带有一定的随机性。但是，从总体来说，我国公民的平均身高却是非常稳定的，并随着生活水平的提高在逐渐增大。这都说明了随机之中又有规律，这种规律称为**统计规律**（statistical regularity）。

对统计数据进行分析，就是利用数据产生的随机性和统计规律进行推断和决策。例如，要比较甲、乙两所高中的英语教学水平，分别从两所学校随机抽取 30 名学生进行试卷测试。一般情况下，两个学校学生的成绩会有差异。这种差异，有可能来源于抽取到的学生水平不同，也有可能来源于两所高中英语教学水平的不同。在很多情况下，我们会面临不同背景的观测数据，需要研究这两组不同背景的数据是否来自于同一随机现象，即需要研究两组数据之间的差异是否大到超过随机性本身所能解释的程度，这种研究的理论是概率论。

1.2.2 概率和机会

概率是对机会的描述，度量了某件事情发生的可能性，其取值在 0 和 1 之间（也

可能是0和1)。概率为0对应那些绝对不可能发生的事情,比如在标准大气压下,水加热到 60°C 就沸腾,这是不可能的。概率为1对应那些一定会发生的事情,比如在真空状态下,自由落体在经过 t 秒钟后,落下的距离 s 必定是 $gt^2/2$ 。现实中,概率为0或1的事件都比较少,大部分事件发生的概率都介于0和1之间,为随机事件。这样的例子有:

- (1) 随意掷一枚骰子,出现的点数为6;
- (2) A, B两队进行一场足球比赛,比赛结果为平局;
- (3) 在某交易日,上证指数以红盘报收;
- (4) 某对新婚夫妇生下的是一名男孩;
- (5) 某天出现雷雨天气。

对这些随机事件以概率的形式进行表述,可以为进一步的统计推断提供基础。

1.2.3 参数和统计量

统计学经常涉及参数和统计量的概念。为研究某一问题,需要对研究对象进行界定。在统计学中,将所要研究的全部个体(数据)的集合称为**总体**(population),其特征的一些概括性数字度量称为**参数**(parameter)。例如,某一地区的平均受教育年限、一批袋装牛奶的合格品率等。

作为总体特征概括性的数字度量,参数的种类可以有很多,但研究者通常关心的主要有以下几种:总体平均数(记为 μ)、总体方差(记为 σ^2)、总体比例(记为 π)等。

现实中,研究对象的范围虽然容易(有时候也不容易)界定,但是要收集与这些研究对象相关特征相联系的全部数据却很困难,这涉及时间、人力、物力、财力等诸多因素。例如,想要知道全国的总人口数,若逐个进行统计,实施普查,则花费的时间和费用是惊人的,现阶段只能每十年进行一次人口普查。有时,对应的研究对象还可能是无限总体,即该总体所包括的元素是无限、不可数的,此时收集总体的全部数据根本不可行。例如,在科学试验中,每一个试验数据可以看做总体的一个元素,而试验可以无限地进行下去。为此,一个很自然的想法是我们不必去收集总体的全部数据,而是从总体中随机抽取一小部分元素形成**样本**(sample),根据样本提供的信息来推断总体的特征。比如,在我国每年进行一次1%人口调查;从出厂的某批次灯泡中随机抽出几个检测其寿命等。

与参数相对应,用来描述样本特征的概括性数字度量称为**统计量**(statistic)。统计量根据样本数据计算得出,是样本的函数。常用的统计量和参数类似,主要有样本平均数(记为 $\bar{x}$)、样本方差(记为 s^2)、样本比例(记为 p)等。

统计学中的绝大多数问题都是研究如何根据统计量去推断参数的。例如,如何用样本平均数($\bar{x}$)去估计总体平均数(μ),如何用样本方差(s^2)去估计总体方差(σ^2),如何用样本比例(p)去估计总体比例(π)等。图1—4形象地展示了这一