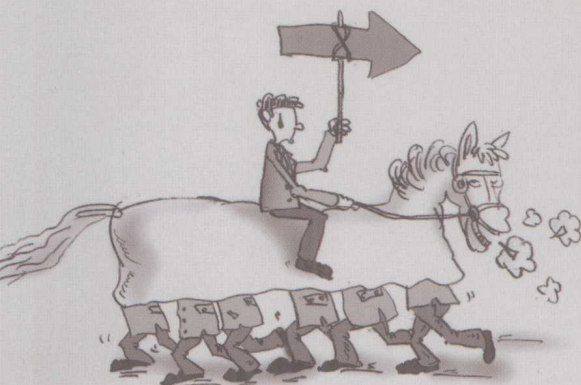




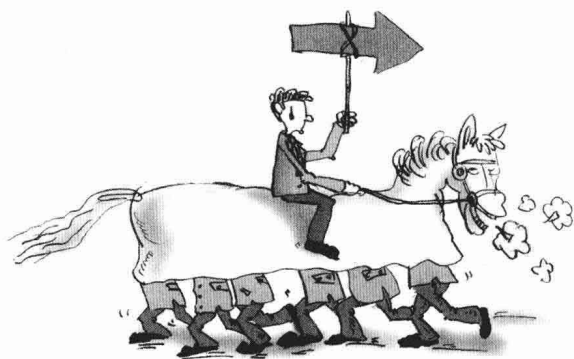
生活中的

STATISTICS AND LIFE

统计学

中国统计杂志社 编

 中国统计出版社
China Statistics Press



生活中的

STATISTICS AND LIFE



统计学

中国统计杂志社 编



中国统计出版社
China Statistics Press

(京)新登字 041 号

图书在版编目(CIP)数据

生活中的统计学 /《中国统计》杂志社编著. —北京:
中国统计出版社, 2010. 9
ISBN 978-7-5037-6056-3

I. ①生… II. ①中… III. ①统计学—通俗读物
IV. ①C8-49

中国版本图书馆 CIP 数据核字(2010)第 167452 号

生活中的统计学

作 者 / 中国统计杂志社 编
责任编辑 / 胡文华 张玉妹
装帧设计 / 杨 超
出版发行 / 中国统计出版社
通信地址 / 北京市西城区月坛南街 57 号 邮政编码 / 100826
办公地址 / 北京市丰台区西三环南路甲 6 号
网 址 / www.stats.gov.cn/tjshujia
电 话 / 邮购 (010)63376907 · 书店 (010)68783172
印 刷 / 河北天普润印刷厂
经 销 / 新华书店
开 本 / 710 × 1000mm 1 / 16
字 数 / 275 千字
印 张 / 21
版 别 / 2010 年 9 月第 1 版
版 次 / 2010 年 11 月第 2 次印刷
书 号 / ISBN 978-7-5037-6056-3/C · 2410
定 价 / 38.00 元

中国统计版图书, 版权所有, 侵权必究。

中国统计版图书, 如有印装错误, 本社发行部负责调换

在生活中正确理解和使用统计

天地玄黄，人间沧桑，万物有变，皆有因循。统计学在一定意义上就是探究因循，发现规律，管理社会、指导人生之学。

对每个人而言，从出生到死亡，是一个既短暂又漫长的过程。人生中的每一天，都面临着不同的问题。从一个幼小的生命有了感知能力开始，他就要学习、思考、判断、选择。成年后，在上大学、结婚、投资以及处理每天的工作中，都必须作出各种决策。这就要求每一个人都要有一种技能，要了解自然界和人类行为中的不确定性，在利用自己和他人的经验作出决策时能使风险最小化、利益最大化。统计知识对个人而言，就是这样的一笔宝贵财富。

统计学和统计数据对政府的政策决策也是非常重要的。政府的每一项政策都会给大量的人群带来不同的影响。为此，他们需要正确的信息，同时需要处理信息的正确方法。世界各国政府为了收集数据，都有一个庞大的官方统计机构。这些数据被用来指导日常行政工作以及制定经济发展和社会福利政策。政策制定者在做出决策时，也要了解和掌握统计专业技术知识。

对某些领域如医学、经济学、科学和技术的专家来说，统计数据的解释和分析是他们研究工作中不可或缺的部分。

大家都生活在现实社会里，统计数据与人们的生活息息相关，分析任何问题都会涉及到统计数据。目前，就整个世界范围而言，统计工作所面临的社会环境越来越复杂，难度越来越大。社会各个层面对统计信息的需求激增，对统计的兴趣日益增强，当然也不乏质疑之声。因此，全社会正确地理解和使用统计数据非常重要。当下，除了不懈努力改进政府统计外，用通俗读物来宣传和普及统计知识尤其重要。

前些年,《中国统计》杂志开辟一些栏目,作为普及统计知识的窗口。美国、德国和台湾地区的一些统计大家热烈响应,于是乎就有了“海外掬粹”和“生活中的统计学”等栏目。其间,台湾统计研究所的赵民德教授、台湾辅仁大学谢邦昌教授、美国宾州大学林共进教授、台湾高雄大学的黄文璋教授等一批学者纷纷撰稿,就统计与人生、统计的出世与入世、概率等话题进行了深入浅出的探悉。这些深入浅出的文章给统计界带来一股清新之风。由此陆续引发出《统计使人豁达》、《解读平均数》、《样本的力量》、《闲话“偶然”与“必然”》等文章,用奇妙的构思,生动活泼的语言,将人们在投资、消费、婚恋、经商、求职等日常生活中能接触到的方方面面,用富于哲理的故事把其中蕴含的统计学原理和统计思想表达出来,把人类行为所包含的深刻哲理生活化。

2010年6月,第64届联合国大会第90次会议通过决议,确定今年10月20日为“世界统计日”并举行纪念活动。联合国庆祝活动的主题是:庆祝官方统计的众多成就,以及服务、诚信和专业精神等核心价值。中国政府统计响应这一号召,将庆祝活动的主题确定为“统计和您在一起”、“走向公开透明的统计”。此时,把散见于《中国统计》的优秀文章集成此书,作为向世界统计日的贺礼之一,并冀使您领略统计学的魅力,从中获得统计学的思维训练,从而掌握用统计学的观点看待生活的钥匙。

统计学不是空灵而抽象的学科,而是一门实用的艺术,了解并掌握它,就能拥有一双洞悉社会和生活的慧眼,运用它,我们会在生活中发现更多乐趣。愿《生活中的统计学》带给您生活的智慧。

统计载道。是为序。

陈皓光

2010年9月

编者说明

1.《生活中的统计学》一书是为配合“中国统计开放日”以及联合国首届“世界统计日”宣传活动而编辑出版的旨在普及统计知识的通俗读物。

2. 文章遴选自《中国统计》杂志自1999年至2010年开辟的“海外掬粹”、“生活中的统计学”等栏目的优秀文章。

3. 由于成文年代跨越10年有余，故文中有些提法已与当下有所差别，望大家必念及于此。

4. 10余年的积淀，文章很多，遗憾的是，我们无法全部收入，但我们保证用编辑的职业道德，将最好的文章呈现给大家。

5. 本书有来自台湾和美国的作者，个别文辞略有差异，我们大部分予以保留，以示尊重。

6. 由于本书文章皆在《中国统计》杂志上发表过，为规范，略去作者单位与刊期。

中国统计杂志社

2010年9月

目录

CONTENTS

万物有常 世事多变	(台湾)赵民德	1
必然性与随机性	(台湾)黄文璋	9
善用资讯	(台湾)黄文璋	17
挑战机遇	(台湾)黄文璋	31
统计学里无罪推定的精神	(台湾)黄文璋	36
统计与棒球	(台湾)黄文璋	41
几个有趣的统计话题	(台湾)谢邦昌	47
统计的A~Z	(台湾)谢邦昌	53
统计的出世及入世观	(台湾)谢邦昌	60
统计有用论	(台湾)谢邦昌	69
统计与人生	(台湾)谢邦昌	77
从铁达尼号沉船资料谈起	(美国)林共进	86
与时济进的统计未来	(美国)林共进	96
文字与数字的学问	(美国)林共进	100
0.618的统计漫话	陈仁恩	103
善用统计明白生活	戴稳胜 付 荣 谢邦昌	108
统计学“悖论”的深度观察	方国斌	113



谁与预测有约	宫春子	118
统计上的“伴侣”——话说绝对数与相对数	宫春子	122
经济生活中奇妙的整数与小数	郭翰超	126
统计是啥，啥是统计	何锦义	129
统计“陷阱”会把我们引入歧途	胡桂华	134
感悟家计调查	李 欧	140
生男，生女？	李 欧	146
数字与感觉	李 欧	151
学会用概率知识解析生活	李 欧	158
在对比中认识世界	李 欧	162
身边的统计学	李小明	169
运用统计分组法教育孩子	李玉玲	172
均数情结与统计推断	罗建国	174
被误解误用甚至盗用了的统计	毛有丰	179
统计与国事、家事天下事	潘 璠	183
我们应该做得更好	潘 璠	188
指数——怎样诠释我们的生活	潘 璠	191
女大三 抱金砖	庞智强	194

平均数的背后.....	齐 勤	198
统计使人豁达.....	邱 东	200
从汶川地震说统计绝对数.....	嵇 冉	203
激发统计趣味.....	孙 蕾	206
妙用贝叶斯统计三则	孙 蕾	212
从GDP说开去.....	王思彤	217
统计视野.....	魏瑾瑞 孙秋碧	220
如何用统计方法选择最佳伴侣.....	吴云勇	225
用统计数字说话点滴	向蓉美	228
世界杯中的数字——绿茵场上的统计学	邢 莉	230
统计——我们生活中的一把钥匙	严建辉	234
解读平均数	严建辉	240
漫谈统计相关——从开车系安全带说起.....	严建辉	247
“三讲”教育与统计	杨昌斌	251
人人都来学点统计知识.....	杨俊才	256
平均地看问题是一种生活态度.....	杨映霜	259
样市的力量	杨映霜	263
井水不犯河水 你与谁无关?	杨映霜	268



麻辣鸡丁、交通事故与概率论.....	杨映霜	273
你不说我也知道	俞肖云	278
统计的导向	俞肖云	281
统计之眼.....	俞肖云	284
总统选举、央视标王与彩电业亏损.....	俞肖云	287
身边的统计学.....	张汉斌	291
人类认知的死穴——兼论生活中的概率思维	张 浩	295
爱情统计学	郑 辉	302
解脱自杀心理 统计学帮你.....	郑 辉	308
公平的“运气”	郑京平	314
三思而行——浅议生活中的统计决策.....	钟守洋	317
闲话“偶然”与“必然”	钟守洋	321



万物有常 世事多变

(台湾) 赵民德

有一位歌手曾经唱过一支歌，大概是叫做“我的志愿”。歌词的意思说，小的时候，他的志愿是做总统，做发明家，后来日渐长大，才知道总统只有一个，真正的发明家也不太多。志愿在联考时填一百多个之后，也就不成其为志愿了。而当年纪更大了，才知道天下的事情都是正正常常地运作的，大多数人都是普通人，既不是总统，也不是发明家。虽不是混混沌沌地过日子，但是快乐是自己定义的；不快乐也是。因此我们可以悲伤，可以欢笑，和普通人一样。

这个世界本来就是这样的。不管你自己觉得你是如何的了不起或特立独行，但大家都是在同样的环境之下培养出来的，吃同样的米，喝同样的水，呼吸同样脏的空气。因此不论是在你的身上做什么测量——比方说腰围吧——那么测量出来的数字，如果放在一大堆类似的人的同样的测量里，则看不出有多少不同。因为你是一个群体里面典型的一员。典型的意思就是说：大家都是普通人，你当然也不会特别地特殊。

这当然是悲哀，也是无奈，但也不见得一定就是坏事。不论什么，我是说不论什么，只要不是强迫性的（例如二中选一的结果），只要是可量测出来（例如可以到小数点后二位）的事物，大概就是这样。你不论去看一个罗马时代的军团，或是去看一个成吉思汗的万人队，大概就是这样：大部分人都是普普通通的。

普通或特殊是在和类似的人相比之后的结果：多半的人一切都普普通通。因为那就是普通的定义，你若是在前面的 3%，那你当然不那么普通了，你若是在后面的 3%，那你也不那么普通了，但 94% 的人既不在前面三个百分点，也不在后面的三个百分点里。94% 的意思再明白不过，在街上随便找一百个人，平均有 94 个是普普通通的。

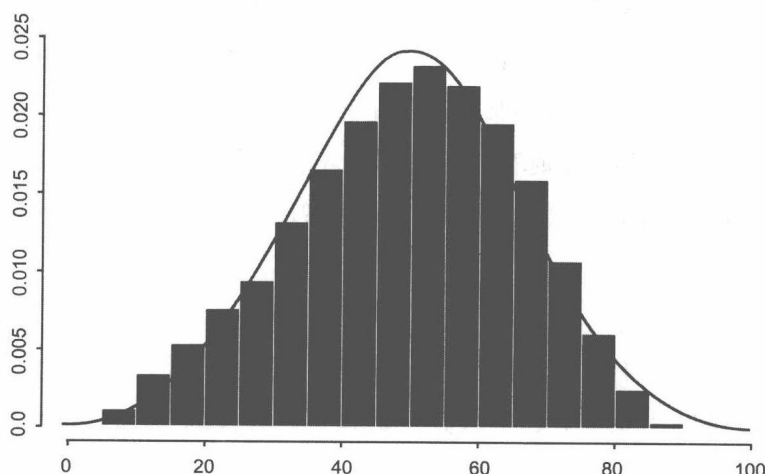
这里面有学问吗？这样的简单道理，粗粗一看，是没有什么学问的，但用一点点画图的技术，问题马上就不一样了。当然，也就有了前、后的 3% 以及当中的 94%。图一是 1988 年度台湾大专联考里两万多名考生的英文成绩，所画的是所谓最简单的长条图。

这个图像什么？这差不多是但又不全是常态分布的曲线。但从图上可清楚地看到，除了极差和极好的考生之外，都是普通考生。普通者，占大部分也。

计算一下这两万多名学生的平均值和标准差。图一的曲线是利用这两个值所算出来的常态分布曲线，只是相差不多罢了！

我们轻轻松松地，几乎是“芝麻开门”一般地打开了一个重要的宝库：常态分布。

* 你在中学有没有要证明 $\sqrt{2}$ 是无理数？



图一 28233 名考生的英文成绩和配适的分布曲线

* 你有没有在微积分里学过 $\frac{\sin(x)}{x} \rightarrow$ ，而这里的“正弦函数”

得要用“弧度制”，而弧度制，得靠圆周率 $\pi=3.14159\cdots$ 才能定义出来。

* 你有没有学过复利，而用到 $e=2.71828\cdots$

如果你觉得这三件事都有一点学问的话，那么常态分布的学问可能更大，它的标准型（意思是说最简单的）公式是：

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\}$$

我们既用到 $\sqrt{2}$ ，也用到 π ，还用到 e 。把任何一组数目，如果它们有一个共同的题目，如果它们由测量——不论是自然的方式（如量你的腰围），还是人为的硬性规定（如考试的分数在 0 与 100 之间），——如果将这一组数字画一个长条图，那么图一的形状，便很有可能出来。

有什么用？中国人做学问，基本上上焉者是为了救世济民，下焉者为了自己吃饭。为好奇而做的学问就少了。中国人有兴趣的是用处。但纯粹自用处的观点来找重要的结果，是很难的。重要的发明或发现，原动力来自“为什么”，而非“有什么用处”。一大堆纯粹因为好奇而导出的事物也许都是没有用处的。但偶然也有一些极有用的，常态分布是其中之一。

一颗刻磨成 58 面的钻石有用吗？它可以切割玻璃，它也可以讨好美女，但它的价值来自它的完善，硬度无出其右，折光率也无出其右。何况研磨的技术是人类能力的尖端所为，它的价值来自它的性质，它是自然生成的，在极难的条件生成。

对统计学者而言，常态分布的存在就如钻石一般：它不是我们发明的，它是自然产生的，而人类用极大的心力从自然现象里逐步提粹精炼，而最后得到 $\varphi(x)$ 这样的简洁形式：有如一座皇冠，镶上了 e 及 π 两粒闪亮的宝石，而尚嫌不足，又配了一个用 $\sqrt{2}$ 做的链子。

但这只是形式上的赞美，常态分布的精要，来自它所具有的性质：是这些性质让这个分布成为一个重要的分布，而不是任何的人为因素。例如它又叫做高斯 (Gauss) 分布，但并不是因为高斯导出它它才伟大 (事实上高斯的推导，还有一些绕圈子的逻辑未能厘清)，它的重要来自它本身的光辉。

大家都知道的一条性质是：“常态分布加常态分布，仍然为常态分布。”这经常是标准统计或机率教本里一定会讲到的性质。我们现在当然不用工夫讨论中央极限定理，但是这一条性质以及它的逆定理“如果有两种分布的和恰好是常态分布，则这两种分布都非同时是常态分布不可。”这两项重要的性质，左右夹击，使得中央极限定理不止是非成立不可，并且它极限分布也非得是常态不可。

我们先自可看到的一组一组的数据，来看是不是有这样的分布，然后再经由一些整套的数学推导 (以及近 200 年一大批学者的努力)，其结论是：我们所能收集的一大堆数据里来看，如果我们铁口直断地说它们是来自某一个常态分布的，那么，我们不会太错。

这个道理的重要就和生物学家要懂得细胞结构，物理学家要知道原子结构是一样的。常态分布是我们所最最了解的数据的基本结构。而我们用它来描绘人类最最不能避免的问题：误差及变异。变异所代表的意思是同样的物件或事情，当它一再重复发生的时候，不论我们如何用心测量，总是不能得到同样的结果，例如生产线上的规格，绝不可能 100% 地合格。再用心挑选的仪仗队成员，其身高体重都不会一样，只是有的变异大些，有些变异小些，而变异在常态分布的时候，可以容易地用一个尺度 σ (sigma) 来表示：

$$\varphi(x; \sigma), \sigma = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{1}{2\pi\sigma^2} X^2\right\}$$

和前面的 $\varphi(x)$ 相比， $\varphi(x)=\varphi(x; \sigma), \sigma=1$ 。我们知道 σ 愈小，则变异愈小， σ 愈大，则变异愈大。图二所示的是 $\sigma=0.5, 1$ 及 2 三个情形。

我们当然也可以将误差想像成变异，但我们还可以更多想一想。自从牛顿发明了第一到第三定律，我们就比较愿意去想象宇宙万物，不论是自然的，还是人文的，大概都有一个方程式可以表现出来。

这当然是一件了不起的想法，我们只要想一下凭几个力学的公式就能计算出多少道理，以由此而产生的近代科技文明，只要有一点想法的学者甚至政客，都会不禁地想：假如其他的事情也可以用公式来表示，那会有多厉害！

但世事岂能如此顺心！我们才学懂高中物理的时候，多半以为世上的事情都可以用公式计算，等到年纪愈大，见得愈多，才知道人类真正能够计算的东西，远比他不会计算的东西为少。但这并不能挡住我们想去“抓住”一些现象并寄望由此能做预测的想法。通常，我们能抓住，或者至少希望能抓住的那一部分，叫做模型，而抓不住的部分，叫做误差，而见到的现象 = 模型 + 误差。这里，误差若是用常态分布来表现——其实也不会非常离谱——我们就差不多概括了目前人类知道的大多数的计量模型。我们当然是将问题极度简化，才有以上的叙述。但精神上这差不多就是目前大多数计量方法的精义。这些得过诺贝尔奖的，还没有得但非常想得这个奖的学者，过去一两百年甚至未来一二百年，多多少少地用这样的基本想法来利用资料与模型，既归纳又演绎地寻找人间万事的规则。这些工作，若没有常态分布，免谈。

人们之所以能够这样做，一个基本的道理是我们已经非常了解常态分布，以至于我们可以充分地透过它的良好性质，一方面安心地设立模型，另一方面由这样的模型里做出推论：我们对大部分天下事物的了解，并不是由于我们真的懂得其中的运行规则，倒反而是因为我们比较了解问题中的变异及误差！

早年的农业实验，希望能找出最佳的肥料比例以使得稻米的产量更高。这一系列的活动，背后都是基于我们对常态分布的了解而达成。至于二次大战以后就一直不曾停止的工业品质与工程的工作，也



一直在常态分布的支撑之下有成。农业的改良让我们吃饱,工业让我们的衣、住和行,甚至娱乐得到满足。那么,新药的发展——为什么它们能够既有效又无害?为什么我们对它的效应及害处都能够用计量的方法加以诠注:这也得力于我们对常态分布的了解。至于人文及经济上的活动更不必说了。几乎可以说,没有常态分布,而这些学问都只能停留在原始的形象描述的阶段而已,哪里谈得上任何解析的结果。

其实我所知也是有限。但是以常态分布为基础的统计模型,的确在证券分析市场里大卖特卖。其中最有名的一个模型,叫做“美国 *BARRA* 多因子净值模型”。它自 1970 开始有了基本理论基础以来,目前已使其公司达到年营业额 4000 万美元以上。至少 600 家主要的投资公司(它的总资产额超过 1 兆美元)长期订购这个模型。其目前所卖的 *E2* 模型,却全然是传统的统计建模 (*modelingbuilding*) 工作。

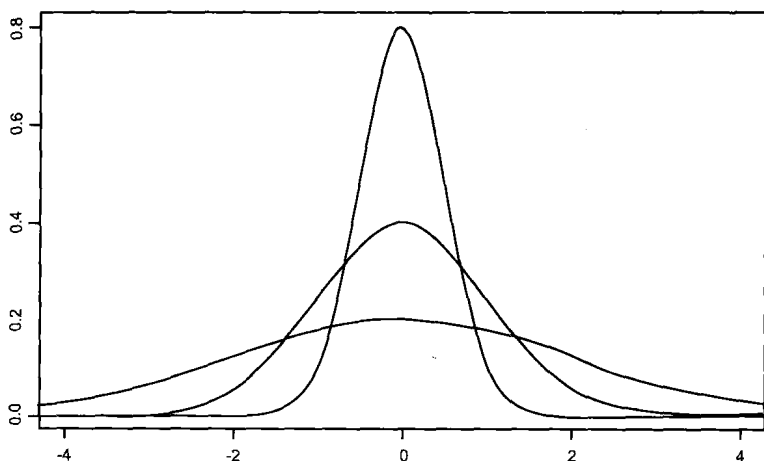
这样的工作有多难呢?在技术上,并没有想象的那样难。如果有源源不断而来的定期数据,如果我手上有一个像样的主计处里面一个技术科的人力,努力两三年就应该有可以出售的商品了。事实上, *BARRA* 模型里用了不下 60 个因子,算是一个相当复杂的多因子模型。如果只有一个因子,即所谓的单因子模型,那么各证券的投资报酬率针对那个唯一因子(多半指某一种市场指标,例如台湾的发行量加权指数)的敏感度,就是目下较高级一点的证券分析师口中所说的“贝它”(*beta*) 值。贝它有什么用?它可用来测量某一证券针对某一证券组合的影响。还太难了是不是?且不必管这些技术名词,更且不必管这些名词的真正意义。在投资工业里,贝它值是常年地被各家各派的高手计算以及估计着的量。“贝它服务”是可以像订报纸一样地按时送达的。

整个投资分析业是大量用到常态分布的行业。只是传统的统计学者,都太带着学者气息,不愿意真正去碰这个行业,因此,表面上

它的统计意味不那样重。但在这里，无数个聪明精巧的心灵，极尽所能地从如海的资料中，希望比他人更早一步找出重要的规则与信息。在另一方面，一个有效率的市场，其定义却正是“应该无法自公众皆可得到的资料中找出必然赚钱的方法”的市场。这是讯号和杂讯交战的杀场，也是一个近乎零和游戏中永不休止的震荡与徘徊。这是一个随机的世界，而随机乃是常态。

人类因为有数字的概念而比其他动物高明。我们曾经以为这方面的知识是绝对的：“给我够长的杠杆，我可以举起地球！”但现代的人类要谦卑多了，他们知道多半的时候，力有未逮。而在他们承认了误差及变异的存在之后，反而海阔天空地走入另一片开阔的领域。万物有常，这是我们想抓住的部分：世事多变，这是我们用适当地常态分布来表现的部分。但是，将常及变放在一起，加以计算及演绎，并由此做科学的验证，却纯然是近五六十年统计学及相关计量科学和广用科学的发展。

我们都是普通人，这是“常”的部分。普通人也有高有矮，有胖有瘦，这是变的部分。这些工作，只是尽量抓住“常”的部分，而让“变”的部分受到控制。而这一切都希望用真实的数据和确切的计算来支持我们的论点。我们当然不只是谈身高体重，我们谈国民所得，失业率，国家竞争力，品质工程，由报税单里看出异常而加以防止逃



图二 不同尺度 σ 之下的常态分布