
统计数字的不确定因素及其控制

王 德 发

一般来说“统计数据”可以概括为两方面的含义，一是指统计调查的数字；二是经过统计加工分析的数字，如每年政府公布的经济发展的统计公报的数字。从理论上说要求这些数字正确、可靠，才能准确地反映现象总体多方面的特征。但是由于多种不确定因素的影响，使得统计数字很难满足上述的质量要求。因此，统计数字的质量控制主要是要最大程度地消除各种不确定因素的影响，使数字准确、可靠。为此，首先要对影响统计数字正确性的不确定因素作一番探讨，挖掘它们的产生根源，才能对症下药采取针对性的措施予以消除。

一、影响统计数字的不确定因素

(一) 指标定义的不确定。统计指标作为一种概念，有其内涵与外延两个方面。概念的内涵是指概念所反映的对象特有属性的总和；概念的外延是指概念所反映的对象的总和，也就是概念所固有的范围。因而指标定义的不确定性也有两个方面。

1. 定义内涵的不确定性。概念内涵如前所述指的是现象特有属性的总和，反映了现象“质”的规定性，导致了事物本身“非此即彼”的特征。但是，有些现象并不能以“非此即彼”的对立形式出现，相反以“亦此亦彼”的中介过渡形式来表现，导致了指标定义的模糊性，致使搜集统计数据时，很难确定数据是否属于某个指标。

2. 定义外延的不确定。定义外延是指概念的固有范围，如果由于条件不充分而使外延（范围）不明确，使某一现象在出现与否上表现出不确定性。从逻辑上看，这是因果律的破损，即冲破了“因为有A，必然有B”的约束，处于一种“纵然有A，也不能确定有B”的状况，因而导致难以确定某一指标的计算范围。

(二) 人为因素的不确定。所谓人为因素造成的不确定是指在搜集资料过程中，由于人的行为不当所造成的不确定性。而人的行为有主观意识支配与非主观意识支配之分。前者主要是指为了达到某些目的，或为了符合某种需要而故意扩大、缩小或多报、漏报，使数字不真的不确定；后者是人在调查过程中，由于行动不慎所造成的误差，即平时所说的登记性误差。

以上二种不确定性主要存在于调查数字中。对于分析预测的统计数字还应包括下面一种不确定因素。

(三) 模型构造的不确定。统计分析，无论是简单的或是复杂的，总要首先将一定的数据构造成相应的模型。因而，用以表现现象总体的变化及其发展趋势的函数形式及其对变量的解释是否科学就显得十分关键。构造模型时，一般都要假定某些变量恒定不变。但实际上由于社会结构的变化、科技的发展，特别是由于人们的价值观念的更新，要使上述假定在实践上完全符合理论是很困难的，在此基础上得到的统计分析数字就不可避免地产生了许多不确定性。

上述种种不确定性因素有些可根据其特点采取一定的处理方法。例如定义的不确定，可在统计设计阶段，运用模糊数学和概率论对指标概念的内涵与外延予以限定，避免不确定性的萌生；人为因素的不确定性，可在统计工作的全阶段，采用全面质量管理的方法，层层把关，予以消除。而对模型构造的不确定因素却很难消除，其原因是所有的模型的有关理论和数据都隐含了大量难以检验的判断和假定，越是复杂的模型，其分析的精度就越难控制。

然而，尽管采取了多种处理措施，消除了某些不确定因素，人们接触到的统计数据实质上还是一个既包含了事物的多种信息，又渗入了多种不确定因素的集合，很难把两者分离开来。一旦不确定因素的影响过份大时，统计数据便表现出一种异常，这就为如何发现统计数据中不确定性带来了可能性。

二、统计数据的质量控制——枝叶图

在质量管理中，人们常常用直方图来对生产工序进行过程控制，现在有一种称作“枝叶图”的方法，能比直方图提供更多的信息，它是由 J·W Tuky 提出来的，目的旨在对数据结构进行加工整理，以获得其统计规律性。这就为我们剔除不确定性因素影响下的异常数据，从而达到对统计数字的质量控制提供了方法。

(一) 枝叶图的基本原理。所谓枝叶图是将数据斩头去尾取其两位有效数，按高低位列成“枝”和“叶”，然后将一批数据分别填入枝和叶，形成图形。其中“枝”是高位数字的分组，如将十位数分为“10~14”与“15~19”两枝、“20~24”与“25~29”两枝，“30~34”与“35~39”两枝，余类推。填“叶”是将低位数按其大小作为“叶”分别填入各所属十位数的所在枝。例如有50个数据：

1.93	1.48	1.65	1.65	1.49	1.72	1.74	1.85	1.75	1.76
1.68	1.78	1.80	1.82	1.75	1.60	1.42	1.45	1.98	1.50
1.75	1.74	1.72	1.78	1.65	1.68	1.80	1.82	1.85	1.74
1.70	1.68	1.65	1.55	1.58	1.55	1.55	1.48	1.70	1.72
1.72	1.73	1.70	1.80	1.65	1.73	1.74	1.76	1.58	1.55

显见每个数字都有相同的个位数 1，可以暂不考虑，只取十分位数为“枝”，百分位数为“叶”。每个十分位数各分为两枝（见图 1）。百分位数分别按“0~4”，填入第一枝，“5~9”填入第二枝（见图 2）。然后将百分位数字按从小到大重新排列，并由小到大和由大到小计算累计次数至中位数“枝”（见图 3）。

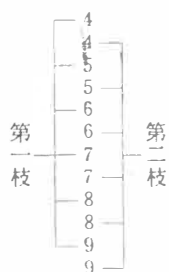


图1

4	2
4	8 9 5 8
5	0
5	5 8 5 5 8 5
6	0
6	5 5 8 5 8 8 5 5
7	2 4 4 2 4 0 0 2 2 3 0 3 4
7	5 6 8 5 5 8 6
8	0 2 0 2 0
8	5 5
9	3
9	8

图2

向下累计

1	.4	2
5	.4	5 8 8 9
6	.5	0
12	.5	5 5 5 5 8 8
13	.6	0
21	.6	5 5 5 5 5 8 8 8
29	.7	0 0 0 2 2 2 2 3 3 4 4 4 4
16	.7	5 5 5 6 6 8 8
9	.8	0 0 0 2 2
4	.8	5 5
2	.9	3
1	.9	8

向上累计 图3

(二) 异常数据的确定。通过枝叶图可以反映出数据结构的分布状况。为了进一步表现它的分布特征，可以计算出中位数、四分位数和四分之三位数。以 X_i ($i=1, 2, 3$) 为第 i 个顺序统计量，据枝叶图 3 可以查得：

$$\text{中位数 } X_2 = 1/2 [X_{(25)} + X_{(26)}] = 1.72$$

$$\text{四分位数 } X_1 = X_{(13)} = 1.6$$

$$\text{四分之三位数 } X_3 = X_{(38)} = 1.76$$

假定中位数的概率为 50%，两个四分位数的概率分别为 25%，则一个三点分布可以近似看作正态分布。那么在一定的概率保证下，可以确定正常数据的控制范围，而具有不确定影响因素的异常数据则在控制范围之外。由于考虑的是数据间的差异，为了使控制中心不受极端数值的影响，三点分布的中心轴取定中位数而不取平均数，与之相对应，采用 6 倍的四分位差来确定正常数据的控制界限，其概率保证程度为 99.73%。本例的四分位差为：

$$Q \cdot D = 1/2 (X_3 - X_1) = 1/2 (1.76 - 1.6) = 0.08$$

$$3Q \cdot D = 3 \times 0.08 = 0.24$$

于是控制界限 $X_2 \mp 3Q \cdot D$ 为:

1.72 ∓ 0.24 即 $1.48 \sim 1.96$

由此可见,数据中的1.42, 1.45, 1.98落在控制界限之外,因而可以认为它们是受不确定因素影响的异常数据。当然发生第 I 类错误的可能性也是有的(即把正常数据误认作异常数据),但弃真概率仅为 $\alpha = 1 - P = 1 - 99.73\% = 0.27\%$,相对而言是较小的。

以上分析仅是对数据结构的内在差异性的相互关系的一种技术性的处理,而统计数据作为一种统计指标的数值或标志值,有其经济内涵,因此确定了哪些是异常数据后还要进一步分析是由于哪一种不确定因素的影响,以便采取措施进行调整。

高尚全、蒋一苇被聘为我校兼职教授并应邀作学术报告

国家体改委副主任、中国城市经济学会副会长、中国工业经济协会副会长、综合开发研究院副理事长高尚全教授和全国人大法律委员会委员、综合开发研究院副理事长、重庆市社科院院长、中国工业经济协会副会长蒋一苇高级研究员,最近被上海财经大学聘为兼职教授。4月16日举行颁证仪式,上海财经大学党委书记兼校长金炳华向他们颁发了聘书。然后,两位兼职教授应邀作了学术报告。高尚全教授着重谈了我国经济体制改革的目标、基本方向和任务。蒋一苇高级研究员谈了《中共中央关于制定国民经济和社会发展十年规划和“八五”计划的建议》的重要意义以及如何深化企业改革。他们的报告获得了与会者的热烈欢迎和一致好评。

颁证仪式和报告会由金炳华校长主持。出席颁证仪式和报告会的还有上海财经大学副校长汤云为和谈敏、前党委书记叶麟根、前校长张君一、前校务委员会副主任叶孝理,以及各系主任、教师、研究生共100多人。

(一丁)

德国联邦统计局专家来我校讲学

3月15日德意志联邦共和国联邦统计局卡斯頓·史达沫博士在国家统计局和上海市统计局领导同志的陪同下访问了我校,并向统计专业和数量经济专业的教师和研究生作了题为《德意志联邦共和国1980年能源流量投入产出表》的学术报告。

史达沫博士是研究国民核算体系方面的专家,他为这次来我国讲学,事先作了充分的准备,讲授内容丰实,气氛生动活泼,取得了良好的效果,为促进中德人民的友谊和进行国际学术交流作了有益的贡献。

(余澄扬)

日本中央大学和我校开展会计学术交流

日前,日本国中央大学商务部主任根本光明教授等一行三人,应邀来我校进行会计学术交流,并与石人瑾教授、林宝怀教授研讨了中日比较会计研究成果。根本光明教授还在我校和上海市会计学会先后作了《会计电算化研究》的学术报告,受到与会者的重视和欢迎。

(朱)