高等学校文科教学参考书

统计平均分析

周兆麟 著

国统计出版社

高等学校文科教学参考书

统计平均分析

周兆麟 著

中国统计出版社

统计平均分析

TONGJI PINGJUN FENXI

周兆麟 著

*

中国统计出版社出版
新华书店北京发行所发行
北京顺义振华胶印厂印刷

*

850×1168毫米 32开本 8印张 20万字
1989年5月第1版 1989年6月北京第1次印刷
印数：1—6,000
ISBN 7—5037—0128—5／C·51(课)
定价：2.05元

前　　言

统计平均分析是统计分析的重要内容，在统计理论研究和实际工作中起着重要的作用。在各个领域里，包括社会经济、生物医学、自然技术等等，均有极广泛的应用。

为适应祖国现代化建设的需要，统计分析工作亟待充实，统计工作质量亟待提高。本书着眼于现代统计平均分析，努力开拓与普及卓有成效的平均分析方法的知识及其运用，以期有益于统计工作，充实统计分析内容。

本书稿部分内容曾在中南财经大学统计专业85、86两届研究生"应用统计专题"课程讲授中利用过。美国Tulane大学数学系的M · Susanna、Tulane Medicine Centre周幼文先生等为本书写作提供了资料，本书也引用了某些作品的材料，谢鸿光、张璋、陈清华及有关同志为本书的完成和出版提供了宝贵的意见和帮助，谨表示衷心的感谢。

由于作者水平有限，书中缺点错误难免，敬请读者批评指正。

周兆麟
于广东商学院
1988年10月

目　　录

第一章　统计平均指标的一般问题 …………………… (1)
一、平均数的概念及意义 …………………… (1)
二、平均数的一般形式 …………………… (3)
三、平均数使用中应注意的问题 …………………… (7)
第二章　常用平均数的种类及计算 …………………… (9)
一、算术平均数 …………………… (9)
二、调和平均数与几何平均数 …………………… (19)
三、平方平均数与标志变异度 …………………… (21)
四、中位数与众数 …………………… (29)
第三章　抽样平均数的概率分布 …………………… (34)
一、随机变量的期望值 …………………… (34)
二、样本均值的概率分布 …………………… (39)
第四章　均值间差异的统计推断 …………………… (46)
一、U 检验法 …………………… (46)
二、t 检验法 …………………… (48)
三、大样本均值的正态检验法 …………………… (61)
四、基本极差的小样均值t_r检验法 …………………… (64)
第五章　非参数方法的应用 …………………… (72)
一、托给快速检验法 …………………… (72)
二、符号检验法 …………………… (75)
三、威尔柯克斯秩和检验法 …………………… (78)
四、曼—怀特列 U 检验法 …………………… (82)
五、配对符号秩检验法 …………………… (87)

六、柯尔莫哥洛夫—斯米诺夫检验法 …………………………… (89)
七、游程检验法 …………………………………………………… (92)
八、列联表属性独立性检验 ……………………………………… (94)
第六章　多总体均值间差异的统计推断 ……………………… (118)
一、方差分析法 …………………………………………………… (118)
二、方差分析中几个问题的处理 ………………………………… (136)
三、多总体均值的多重比较 ……………………………………… (145)
四、方差分析的快速检验 ………………………………………… (158)
第七章　偏离特征均值指标的统计推断 ……………………… (185)
一、矩的意义 ……………………………………………………… (185)
二、两个总体的方差齐性检验 …………………………………… (187)
三、多总体方差齐性检验 ………………………………………… (196)
四、偏度与峰度检验法 …………………………………………… (201)
第八章　统计均值的动态分析与因素分析 …………………… (212)
一、动态数列与序时平均数 ……………………………………… (212)
二、时间序列的因素分析 ………………………………………… (217)
三、平均指数分析 ………………………………………………… (227)
四、变量间关系的回归分析 ……………………………………… (232)

附　录

附表1　标准正态概率积分表 ………………………………… (239)
附表2　x^2分布表 ……………………………………………… (241)
附表3　F分布表 ……………………………………………… (242)
附表4　t分布表 ……………………………………………… (246)

第一章　统计平均指标的一般问题

一、平均数的概念及意义

统计平均分析是统计分析的重要内容，它是利用平均数进行定量分析的方法。

例如，全企业职工各人工资标准多种多样，要了解一个企业职工工资的一般水平的高低，就不能仅仅以个别职工的工资标准为准；即使提供该企业某月工资总额 57 600 元的总量指标资料，也不说明问题，因为工资总额与职工人数的多少有密切关系，工资总额大也不一定是工资一般水平就高；提供一份全体职工工资清单，虽然这是对职工工资情况最完善的描述，但也不会得到要领。如果以全企业 640 名职工去除工资总额，得到该企业职工月平均工资为 90 元的平均指标，这样，一般工资水平是高是低，了解就比较具体，就可以用它与过去时期的水平相比，或以它和其他企业工资水平相比，作各种分析来说明问题了。

一般说来，平均数就是将被研究的同种类现象的某种数量标志的各个体数量差异抽象化，用一个概括的指标综合说明现象的有代表性的典型水平，它是最常见也最重要的综合性指标。如平均年龄、单位产品平均成本、平均工资等，就是大家所熟知的平均指标。

在现象的发展过程中，某种共同的必然性的东西，总是和特殊的偶然性的因素交织结合在一起的。为了揭示一般典型的特征，需要消除偶然因素变动的影响，而在平均数中，偶然变动的影响相互抵消，而充分反映出某种必然性。因此，平均数比个别

数值更能充分反映出同种类现象的本质。

平均数在统计理论研究和其他许多领域中有极为广泛的应用。

许多社会经济现象和范畴，就是通过平均值来表现的。马克思认为："总的说来，在整个资本主义生产中，一般规律作为一种占统治地位的趋势，始终只是以一种极其错综复杂和近似的方式，作为不断波动中得到的但永远不能确定的平均情况来发生作用"①。例如，他谈到一般利润率时，指出它对于"由不同生产部门的不同利润率有决定作用，而这些不同的利润率的平均数形成一般利润率"②，谈到社会平均劳动时，他指明"物化为价值的劳动，是社会平均性质的劳动，也就是平均劳动力的表现。但是平均量始终只是同种的许多不同的个别量的平均数"③。诸如大量生产中产品成本水平、报告期生产者劳动生产率水平、职工工资水平、生产定额、设备利用定额、某时期某种商品价格水平等等，都是经济管理中重要的平均指标。

在自然技术科学领域里，也经常运用平均指标。研究物体运动的速度、气温、气压的形成与变动，各种计算数值的准确测定，元素原子量的研究等等，就要用到平均数。

就统计理论来说，平均数更是占着主要的地位。期望值就是总体平均数；测定数据变异程度要用到平均数；大数定律按其内容来说，无非是均值与期望值联系的定律；抽样法理论按其实质来说，也就是推算总体均值与抽样均值之间的差异等等。

概括起来说，统计平均分析可以解决这样几个方面的问题：(1) 说明现象发展的典型水平；(2) 就两个或多个水平间进行静态比较；(3) 说明现象水平在时间发展上的动态变动；(4) 进行

---

① 《资本论》，人民出版社1975年版，第3卷，第181页。

②同①。

③ 《资本论》，人民出版社1975年版，第1卷，第359页

与计划、设计、技术管理等有关的核算与估算问题等。

二、平均数的一般形式

平均数既然就是将被研究现象的某种数量标志的各个体数量差异抽象化，用一个概括的指标来综合说明现象的典型水平，它的各种表达式正是这样形成的。

例如，某车间有 n 名工人，其各人工资额分别为 x_1，x_2，…，x_n，则 $x_1+x_2+\cdots+x_n=\sum^{n} x_i$ 为全车间工人工资总额。如果我们能求得一个平均工资水平 $\bar{x}$，它是一个有代表性的典型水平，有它去取代各个体工资额，那么，下述关系式应该成立：

$$x_1+x_2+\cdots+x_n=\bar{x}+\bar{x}+\cdots+\bar{x}$$

即有

$$\sum_{i=1}^{n} x_i=n\bar{x}$$

$$\bar{x}=\frac{1}{n}\sum^{n} x_i \qquad (1—1)$$

这就是我们最熟悉的算术平均数。

又如，我国第一个五年计划期间每年的钢产量及环比发展速率资料如下：

年　份	1952	1953	1954	1955	1956	1957
钢产量(万吨)	134.9	177.4	222.5	285.3	446.5	535.0
环比发展速率		1.315	1.254	1.282	1.565	1.198

环比发展速率计算式是相邻各期产量（a_i）之比，即

$$x_i=\frac{a_i}{a_{i-1}} \qquad (i=1,\ 2,\ \cdots,\ 5)$$

五年的总速率是 $x_1\cdot x_2\cdot x_3\cdot x_4\cdot x_5$，即按照每年的速率

发展，钢产量由 1952 年的水平$a_0=134.9$，发展到 1957 年的$a_5=535.0$的水平：

$$a_0\cdot x_1\cdot x_2\cdot x_3\cdot x_4\cdot x_5=a_5$$

如果能求得一个典型水平$\bar{x}$，用它去取代各年的发展速率，下述关系式应当成立：

$$a_0\cdot x_1\cdot x_2\cdot x_3\cdot x_4\cdot x_5=a_0\cdot\bar{x}\cdot\bar{x}\cdot\bar{x}\cdot\bar{x}\cdot\bar{x}=a_5$$

即有 $$a_0\bar{x}^5=a_5$$

$$\bar{x}=\sqrt[5]{\frac{a_5}{a_0}}$$

一般地有 $$\bar{x}=\sqrt[n]{\frac{a_n}{a_0}} \qquad (1—2)$$

这就是几何平均数。其一般形式是：

$$\bar{x}=\sqrt[n]{x_1x_2\cdots x_n} \qquad (1—3)$$

再如，甲种苹果每斤 0.5 元，乙种苹果每斤 0.75 元，各买 1 元，则每元能买甲种苹果$\frac{1}{0.5}$斤，能买乙种苹果$\frac{1}{0.75}$斤，甲、乙种苹果各买 1 元，共买 $(\frac{1}{0.5}+\frac{1}{0.75})$ 斤。如果能求得一个典型水平$\bar{x}$，用它去代替各种苹果的单位价格，则有：

$$\frac{1}{0.5}+\frac{1}{0.75}=\frac{1}{\bar{x}}+\frac{1}{\bar{x}}$$

$$\frac{1}{0.5}+\frac{1}{0.75}=\frac{2}{\bar{x}}$$

$$\bar{x}=\frac{2}{\frac{1}{0.5}+\frac{1}{0.75}}\approx 0.6\text{ 元}$$

平均单价为 0.6 元，这就是调和平均数。其一般形式是：

$$\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}=\frac{1}{\bar{x}}+\frac{1}{\bar{x}}+\cdots+\frac{1}{\bar{x}}$$

$$\bar{x}=\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}} \tag{1-4}$$

在计算若干个正方形面积的平均边长、圆面积的平均半径，以及标识变异度研究、热过程分子动能研究等实际问题中，我们还会遇到所谓的平方平均数。以正方形平均边长计算为例，设有边长分别为 5、10、15、25 单位的 4 个正方形，其总面积是$5^2+10^2+15^2+25^2$，如果能求得典型的平均边长$\bar{x}$，用$\bar{x}$去取代这些正方形的边长，下述关系式应成立：

$$5^2+10^2+15^2+25^2=\bar{x}^2+\bar{x}^2+\bar{x}^2+\bar{x}^2$$

推得

$$\bar{x}=\sqrt{\frac{5^2+10^2+15^2+25^2}{4}}\approx 15.6\text{ 单位}$$

平方平均数的一般形式是：

$$\bar{x}=\sqrt{\frac{x_1^2+x_2^2+\cdots+x_n^2}{n}} \tag{1-5}$$

统计研究中有时也会遇到立方平均数及其它高次方平均数的情况，如统计分布曲线偏倚性研究中就用到立方平均数，统计分布曲线峰态研究中就用到四次方的平均数等。

上面列举的这些平均数形式，实际上可一般概括为：

$$\bar{x}=\left(\frac{\sum_{i=1}^{n}x_i^k}{n}\right)^{\frac{1}{k}} \tag{1-6}$$

当 k = 1 时，即为算术平均数；

当 k = 2 时，即为平方平均数；

当 k = 3 时，即为立方平均数；

当 k = 0 时，即为几何平均数。

当 k = 0 时，(1-6)式表现为未定式

$$\bar{x}=1^{\infty}$$

(1-6)式两边取对数

$$\ln\bar{x}=\ln\sqrt[k]{\frac{\Sigma x^k}{n}}=\frac{\ln\left(\frac{\Sigma x^k}{n}\right)}{k}$$

仍属于未定式$\frac{0}{0}$的情况。我们运用洛必达法则来求极限:

$$\begin{aligned}\lim_{k\to 0}\ \ln\bar{x}&=\lim_{k\to 0}\frac{\ln\left(\frac{\Sigma x^k}{n}\right)}{k}\\&=\lim_{k\to 0}\frac{\left(\ln\frac{\Sigma x^k}{n}\right)'}{k'}\\&=\lim_{k\to 0}\frac{\left(\frac{1}{n}\Sigma k\ln x\right)'}{k'}\\&=\Sigma\ln x^{\frac{1}{n}}\\&=\ln(\Pi x)^{\frac{1}{n}}\end{aligned}$$

即有 $$\bar{x}=\sqrt[n]{\Pi x_i}$$

当 k=-1 时，(1-6)式即为调和平均数。即:

$$\bar{x}=\frac{n}{\Sigma\frac{1}{x_i}}$$

从数学上看，这种乘方平均数公式中的幂指数越大，则同属一个体系的个别数值的平均数也越大，统计上常从形式上确立优函数定则关系:

$$\bar{x}_{倒数}<\bar{x}_{几何}<\bar{x}_{算术}$$

我们还可更进一步更为一般地来概括平均数在数量上的表达式。相应于标志的各个具体变量的某种函数关系Φ $(x_1, x_2, \cdots, x_n)$，如果$\bar{x}$是其综合的典型水平，则应

有：

$$\Phi(x_1, x_2, \cdots, x_n,) = \Phi(\bar{x}, \bar{x}, \cdots, \bar{x})$$

即 $\Phi(x_1, x_2, \cdots, x_n,) = \Psi(\bar{x})$

如果反函数存在，则有：

$$\bar{x} = \Psi^{-1}[\Phi(x_1, x_2, \cdots, x_n,)] \qquad (1-7)$$

这就是平均数在数量关系上的最一般形式。

三、平均数使用中应注意的问题

前面谈到，平均数就是将被研究的同种类现象的某种数量标志的各个体数量差异抽象化，用一个概括的指标综合说明现象的有代表性的典型水平。马克思说过："平均量始终只是同种的许多不同的个别量的平均数"①。我们利用平均数形式，消除个别事物的偶然变动，抽象地综合揭示现象的共同特性。要做到这一点，一定要注意，只有在同种类（或同质性）现象中才能利用平均数这个前提条件。否则，不但得不到正确结论，还会歪曲事实。例如，将资本家或资方代理人的高薪和收入低下的一般工人工资混同一起，计算所谓的职工平均工资，以"提高"工资水平，抹煞阶级差别，就会粉饰真相。

在同质总体中计算的平均指标，对于认识事物的共同特性是有重要意义的。但也正因为它抽象掉了局部的特征和属性，往往只能给人们以一个总的、笼统的一般概念。在利用平均数时，我们常常不能仅仅满足于这个总的平均数，而要借助于反映一定特征的分组局部平均数来作补充说明，这对更进一步认识事物是很有帮助的。例如，有生产同种产品的两个车间工人平均日产量资料如下表所示：

① 《马克思恩格斯全集》，人民出版社1972年第1版，第23卷，第359页。

	甲车间			乙车间		
	人数(人)	旬产量合计(件)	平均日产量(件)	人数(人)	旬产量合计(件)	平均日产量(件)
生产工人	120	9 600	8	150	10 500	7

表中总平均日产量数字给我们的印象是，甲车间劳动效率高。但如果进一步就熟练和不熟练两组工人分别计算组平均数便得如下资料：

	甲车间			乙车间		
	人数(人)	旬产量合计(件)	平均日产量(件)	人数(人)	旬产量合计(件)	平均日产量(件)
生产工人	120	9 600	8	150	10 500	7
其中：						
熟练工人	90	7 800	8.7	30	2 940	9.8
非熟练工人	30	1 800	6	120	7 560	6.3

上表给我们印象就很不一样：两个车间各类不同工人的劳动效率均是乙车间高于甲车间，之所以与总的劳动效率表现不一致，是由于乙车间的工人均较年轻，非熟练工人比重大所致。将总平均数和组平均数结合补充分析，所使分析工作更切合具体情况，更深入地揭示客观事物内在性质和发展变化的原因。

正因为平均数是将个别变动差异抽象化，所以运用平均数以综合描述现象时，还要注意利用分配数列和个别典型事例资料来结合说明问题，以发现萌芽的新生事物或潜在的异常情况等。

在实际统计工作中运用平均数时，还要根据研究任务和所掌握的资料具体情况，来对各种平均数的计算形式加以比较选择和利用。

第二章　常用平均数的种类及计算

统计平均数的种类很多，从表达形式及含义的类型上看，有算术平均数、调和平均数、几何平均数、中位数、众数、平方平均数等；从计算平均数包含的个体的范围大小上看，有总体平均数、抽样平均数、组平均数等；从时间因素上看，则有动态平均数与静态平均数之分等等。

一、算术平均数

这是平均数中最基本也最常用的形式，也是为大众所最熟知的。它是总体单位的某种标志数值的总和除以与之相适应的单位总数所求得的商。

$$\text{算术平均数}=\frac{\text{某种标志总量}}{\text{总体总量}} \tag{2-1}$$

平均数的计量单位和标志总量的计量单位是一致的。

例如：

$$\text{职工平均工资(元)}=\frac{\text{职工工资总额(元)}}{\text{职工总数(人)}}$$

$$\text{职工平均年龄(岁)}=\frac{\text{职工年龄总和(岁)}}{\text{职工人数(人)}}$$

算术平均数为什么是最常用的基本形式呢？因为实际中遇到的变量，各个体的某种标志值(个人工资、苹果单价等)相加求总和，就等于总体的标志总量(工资总额、苹果销售总额等)，算术平均数正好适用于这种情况。

(一)简单算术平均数与加权算术平均数

第一章的(1-1)式，就是**简单算术平均数**的计算公式:

$$\overline{X}=\frac{1}{n}\sum^{n}x_{i} \qquad (2-2)$$

【例 1】 9 个工人日产零件中的优等品件数为: 298，253，183，148，275，291，188，470，306，试求工人生产的平均优等品件数。

由(2-2)式有:

$$\overline{x}=\frac{298+253+183+148+275+291+188+470+306}{9}$$

$$=\frac{2\,412}{9}=268(\text{件})$$

【例 2】 假定需要平均的仍然是例 1 中的那 9 个标志数值，但生产各个优等品件数的人数不同，如下表:

优等品件数(件)	298	253	183	148	275	291	188	470	306
工人数(人)	4	8	12	4	2	2	4	2	2

求工人生产的平均优等品件数。

能不能仍将这 9 个标志值相加，并用 9 除得到 268 件，作为生产的平均优等品件数呢？这是不行的，因为各标志值的人数不一样。我们应当把这 40 个工人的所有标志值相加除以 40，按(2-2)式来计算$\overline{x}$，这实际上就是:

$$\overline{x}=\frac{9\,240}{40}$$

$$=231(\text{件})$$

同样是这 9 个标志值，为什么平均值 231 件比例 1 的 268 件少了许多呢？这主要是各标志值的人数不一样，生产优等品件数较少的工人人数的比重要大些的缘故。可见，这里的人数多少在

平均数计算中起着权衡轻重的作用，在平均数计算中，区别各个标志值相对重要性的计量数值，称之为**权数**，又称权重．如果用 f_i 表示第 i 个标志值的次数数值，即有：

$$\bar{x}=\frac{\sum x_i f_i}{\sum f_i} \tag{2-3}$$

$$或\quad \bar{x}=\sum x_i\cdot\frac{f_i}{\sum f_i}=\sum x_i W_i \tag{2-4}$$

式中 $W_i=\frac{f_i}{\sum f_i}$ 表示各标志值个体数在总体个数中所占的比重，即为权数，且有 $\sum W_i=1$。这种平均数形式称为**加权算术平均数**。

当我们只知道各标志值个体数在总体中所占的比重，而不知道这些具体的个体数时，同样可以按(2–4)式来计算加权平均数。

【例 3】 某生产组中工资标准为 105 元的有 40%的工人，工资标准为 90 元的有 50%的工人，工资标准为 80 元的有 10%的工人。求该生产组的平均工资水平。

$$\begin{aligned}\bar{x}&=105(40\%)+90(50\%)+80(10\%)\\&=42+45+8\\&=95(元)\end{aligned}$$

对于权数的理解，它只是各个标志值相对重要性的一种度量，而不要作别的理解。上面的例子是以各个标志值的次数数值的比重作为权数的，但次数数值并不一定是当然的权数，当次数数值不能反映出各个标志值的相对主要性时，就不能以它作为权数，而要另行考虑。

【例 4】 甲、乙、丙、丁四名分析员对同一批原料的有用成分含量进行测定，得资料如下：