矿业经济分析数学方法

汪光顺　舒　航　编著

国标准出版社

矿业经济分析数学方法

汪光顺　舒　航　编著

中 国 标 准 出 版 社

图书在版编目(CIP)数据

矿业经济分析数学方法/汪光顺,舒航编著.-北京:
中国标准出版社,1995.12
ISBN 7-5066-1183-X

Ⅰ.矿… Ⅱ.①汪… ②舒… Ⅲ.经济数学-应用-矿业
经济 Ⅳ.F407.1

中国版本图书馆 CIP 数据核字(95)第 19171 号

中国标准出版社出版
北京复兴门外三里河北街 16 号
邮政编码:100045
电　　话:8522112
中国标准出版社秦皇岛印刷厂印刷
新华书店北京发行所发行　各地新华书店经售

*

开本 850×1168 1/32　印张 10¼　字数 293 千字
1995 年 12 月第一版　1995 年 12 月第一次印刷

*

印数 1—500　定价 16.00 元

前　言

数学是经济分析的基础。矿业经济学作为一门新兴的交叉性边缘学科，更离不开现代数学理论与方法的支撑。近年来，我国在矿业经济，尤其是在微观技术经济领域，取得了令人瞩目的成绩；并正朝着理论体系化、方法多元化、手段现代化、成果实用化方向发展。为推动我国矿业技术经济研究迈上新的台阶，促进矿业经济学科的不断完善，我们编写了《矿业经济分析数学方法》一书。

本书分上、下两篇。上篇重点介绍了微观矿业经济研究中常用的几种现代数学方法。为加深读者对各种数学方法的应用有较深层次的理解，下篇刻意选编了大量具有代表性的实用案例，供读者参考。应当说明的是，限于篇幅等诸多原因，还有许多新发展的现代数学——决策方法及应用案例未作介绍，请读者明鉴。

本书的编写，得到北京科技大学矿产经济研究室陈希廉教授的鼓励与指导，采矿系刘华生教授、李国乔教授亦给予热忱支持并引用了他们的资料，地质系袁怀雨教授、黄凤吟副教授、曹乐农高级工程师也给予热情帮助，在此一并致谢！此外，作者还要向本书引用资料的所有作者、单位表示诚挚感谢！

由于作者学术水平不高，专业知识局限，书中难免有谬误出现，诚请读者批评指正。

作　者

1994 年 11 月

目　　录

上篇　基础篇

下篇　应用篇

上篇　基础篇

第一章 回 归 分 析

回归分析是数理统计中应用最广泛的一个重要分支，它是研究变量间相关关系的一种统计方法。所谓相关关系，是指变量间有一定的关系，但这种关系无法用一个精确的数学式子来表达。例如，人的身高与体重之间有一定关系，知道一个人的身高大致可以估计出体重，但不能精确地计算出体重，其原因是人有较大的个体差异。象这种是既密切，又不是完全确定的变量间的关系，称为相关关系。它与确定性关系是不同的（当然两者之间也并不存在一条严格的界限）。

回归分析方法是建立在对客观事物进行大量试验和观察的基础上，寻找隐藏在那些看上去是不确定的现象中的统计规律性，从而找出它们之间的相关关系。这种方法在实践中有着广泛的应用，例如求经验公式，寻找产品的产量或质量指标与生产条件之间的关系，天气预报，病虫害预报，建立自动控制中的数学模型，某些新标准的制订等等，都经常要用到回归分析这一工具。

在回归分析中，主要研究以下三个问题：

（1）从一组数据出发，确定变量间是否存在着相关关系，如果存在相关关系的话，需要确定它们之间的定量关系表达式，并对它的可信程度作统计检验。

（2）从共同影响一个变量的许多变量中，判断哪些变量的影响是显著的，哪些变量的影响是不显著的。

（3）利用所找到的定量关系表达式对变量进行预测或控制。

回归分析的内容是十分丰富的，限于篇幅，只介绍一元线性回归、可化为一元线性回归的非线性回归及多元线性回归。

第一节　一元线性回归

一、一元线性回归的数学模型

一元线性回归是处理两个变量之间的关系，下面通过一个例子来说明如何建立一元线性回归模型。

例 1　已知某合金的抗拉强度 y (kg/mm^2)与合金中含碳量 x (%)之间有一定的相关关系。为了了解其相关关系的表达形式，第一步是通过试验或从生产记录中去收集 n 组 y 与相应的 x 的值，表 1-1 就是收集到的 12 个 y 与相应的 x 的值。

表 1-1

编号	x	y	编号	x	y
1	0.10	42.0	7	0.16	49.0
2	0.11	43.5	8	0.17	53.0
3	0.12	45.0	9	0.18	50.0
4	0.13	45.5	10	0.20	55.0
5	0.14	45.0	11	0.21	55.0
6	0.15	47.5	12	0.23	60.0

第二步是画一张散点图。这是在研究两个变量间相关关系时常用的一种直观办法，以 x 为横坐标，以 y 为纵坐标，每一对数据(x_i, y_i)作为一个点在坐标纸上以“×”表示出来($i=1,2,\cdots,n$)。本例的散点图见图 1-1。

第三步是观察这张散点图。从图 1-1 我们发现，这 n 个点基本上在一条直线 l 附近，从而我们可以认为 y 与 x 的关系基本上是线性的，而这些点与直线 l 的偏离是由其他一切随机因素的影响而造成的。故我们可以假定表 1-1 中的数据有如下结构：

$$y = \beta_0 + \beta x + \varepsilon \tag{1-1}$$

其中 $\beta_0+\beta x$ 表示 y 随 x 的变化而线性变化的部分。ε 是一切随机因素影响的总和，称为随机变量。并假定其数学期望 $E(\varepsilon)=0$，方差 $D(\varepsilon)=\sigma^2$。在涉及到分布时，可进一步假定 ε 服从正态分布 $N(0,\sigma^2)$。x 可以是随机变量也可以是一般变量，以下讨论中我们总认为 x 是一般变量，即它是可以精确测量或严格控制的。由以上假定可知 y 也是个随机变量，但其值是可以观测的，其数学期望是 x 的线性函数：

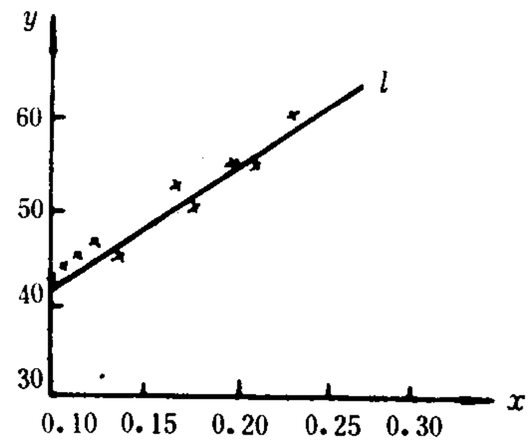

图 1-1

$$E(y)=\beta_0+\beta x \tag{1-2}$$

这便是 y 与 x 的相关关系的形式。

对表 1-1 的 n 组观测，由式(1-1)可认为有

$$\begin{cases} y_i=\beta_0+\beta x_i+\varepsilon_i \\ \text{各 } \varepsilon_i \text{ 相互独立，} E(\varepsilon_i)=0,\ D(\varepsilon_i)=\sigma^2 \\ \qquad i=1,2,\cdots,n。\end{cases} \tag{1-3}$$

式(1-3)是一元线性回归的数学模型。

我们的首要任务就是要根据表 1-1 去求出式(1-2)中未知参数 β_0 与 β 的估计 $\hat{\beta}_0$、$\hat{\beta}$，由此可得 $E(y)$ 的估计：

$$\hat{y}=\hat{\beta}_0+\hat{\beta}x \tag{1-4}$$

式(1-4)称为 y 关于 x 的一元线性回归方程。这便是我们要求的 y 与 x 之间的定量关系表达式，其图象便是图 1-1 中的直线 l。称此直线为回归直线，$\hat{\beta}$ 也称为回归系数，它是回归直线的斜率，$\hat{\beta}_0$ 为回归常数，它是回归直线的截距。

二、最小二乘估计

要求出回归方程(1-4)就是要求出 β_0 与 β 的估计，求此估计的一个直观想法便是希望对一切 x_i，观测值 y_i 与回归值 $\hat{y}_i=\hat{\beta}_0+\hat{\beta}x_i$

的偏离达到最小，为此我们通常采用最小二乘法来求 β_0 与 β 的估计。令

$$Q(\beta_0,\beta)=\sum_{i=1}^{n}(y_i-\beta_0-\beta x_i)^2 \tag{1-5}$$

所谓 β_0 与 β 的最小二乘估计是指使下式成立的 $\hat{\beta}_0$ 与 $\hat{\beta}$：

$$Q(\hat{\beta}_0,\hat{\beta})=\min_{\beta_0,\beta}Q(\beta_0,\beta)$$

由于 $Q(\beta_0,\beta)$ 是 β_0,β 的非负二次函数，其最小值必存在，同时它是 β_0,β 的可微函数，故根据微积分学中的极值原理，$\hat{\beta}_0,\hat{\beta}$ 应是下列方程组的解：

$$\begin{cases}\left.\dfrac{\partial Q}{\partial \beta_0}\right|_{\beta_0=\hat{\beta}_0,\beta=\hat{\beta}}=-2\displaystyle\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}x_i)=0\\ \left.\dfrac{\partial Q}{\partial \beta}\right|_{\beta_0=\hat{\beta}_0,\beta=\hat{\beta}}=-2\displaystyle\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}x_i)x_i=0\end{cases} \tag{1-6}$$

式(1-6)可化简成

$$\begin{cases}n\hat{\beta}_0+(\sum_i x_i)\hat{\beta}=\sum_i y_i\\ (\sum_i x_i)\hat{\beta}_0+(\sum_i x_i^2)\hat{\beta}=\sum_i x_iy_i\end{cases} \tag{1-7}$$

式(1-7)称为正规方程组，其中“$\sum_i$”表示“$\sum_{i=1}^{n}$”，下面不再重复说明。

为求正规方程组(1-7)的解，我们可从第一式中求得

$$\hat{\beta}_0=\frac{1}{n}\sum_i y_i-\hat{\beta}\cdot\frac{1}{n}\sum_i x_i=\bar{y}-\hat{\beta}\bar{x}$$

将它代入第二式，得

$$(\sum_i x_i)(\bar{y}-\hat{\beta}\bar{x})+(\sum_i x_i^2)\hat{\beta}=\sum_i x_iy_i$$

经过整理，得

$$[\sum_i x_i^2-\frac{1}{n}(\sum_i x_i)^2]\hat{\beta}=\sum_i x_iy_i-\frac{1}{n}(\sum_i x_i)(\sum_i y_i)$$

故求得其解为：

$$\begin{cases} \hat{\beta} = \dfrac{\sum\limits_i x_i y_i - \dfrac{1}{n}(\sum\limits_i x_i)(\sum\limits_i y_i)}{\sum\limits_i x_i^2 - \dfrac{1}{n}(\sum\limits_i x_i)^2} \\ \hat{\beta}_0 = \overline{y} - \hat{\beta}\overline{x} \end{cases} \tag{1-8}$$

为简化记号，令

$$l_{xy} = \sum_i (x_i - \overline{x})(y_i - \overline{y}) = \sum_i x_i y_i - \frac{1}{n}(\sum_i x_i)(\sum_i y_i)$$

$$l_{xx} = \sum_i (x_i - \overline{x})^2 = \sum_i x_i^2 - \frac{1}{n}(\sum_i x_i)^2$$

其中

$$\overline{x} = \frac{1}{n}\sum_i x_i, \quad \overline{y} = \frac{1}{n}\sum_i y_i$$

则最小二乘估计为

$$\begin{cases} \hat{\beta}_0 = \overline{y} - \hat{\beta}\overline{x} \\ \hat{\beta} = \dfrac{l_{xy}}{l_{xx}} \end{cases} \tag{1-9}$$

下面我们来求例 1 的回归方程，计算过程通常用表格形式给出（表 1-2 与表 1-3）。为了下面进一步分析的需要，在此将 $l_{yy} = \sum\limits_i (y_i - \overline{y})^2 = \sum\limits_i y_i^2 - \dfrac{1}{n}(\sum\limits_i y_i)^2$ 一起求出了，它在求回归方程时是不需要的。

若将 $\hat{\beta}_0$ 的表达式代入一元线性回归方程

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}x$$

表 1-2

编号	x	y	x^2	xy	y^2
1	0.10	42.0	0.010 0	4.200	1 764.00
2	0.11	43.5	0.012 1	4.785	1 892.25
3	0.12	45.0	0.014 4	5.400	2 025.00
4	0.13	45.5	0.016 9	5.915	2 070.25
5	0.14	45.0	0.019 6	6.300	2 025.00

续表 1-2

编号	x	y	x^2	xy	y^2
6	0.15	47.5	0.022 5	7.125	2 256.25
7	0.16	49.0	0.025 6	7.840	2 401.00
8	0.17	53.0	0.028 9	9.010	2 809.00
9	0.18	50.0	0.032 4	9.000	2 500.00
10	0.20	55.0	0.040 0	11.000	3 025.00
11	0.21	55.0	0.044 1	11.550	3 025.00
12	0.23	60.0	0.052 9	13.800	3 600.00
Σ	1.90	590.5	0.3194	95.925	29 392.75

表 1-3

$\Sigma x_i=1.90$	$\Sigma y_i=590.5$	$n=12$
$\bar{x}=0.158\ 3$	$\bar{y}=49.208\ 3$	
$\Sigma x_i^2=0.319\ 4$	$\Sigma x_i y_i=95.925$	$\Sigma y_i^2=293\ 92.75$
$\frac{1}{n}(\Sigma x_i)^2=0.300\ 8$	$\frac{1}{n}(\Sigma x_i)(\Sigma y_i)=93.495\ 8$	$\frac{1}{n}(\Sigma y_i)^2=290\ 57.520\ 8$
$l_{xx}=0.018\ 6$	$l_{xy}=2.429\ 2$	$l_{yy}=335.229\ 2$
	$\hat{\beta}=\frac{l_{xy}}{l_{xx}}=130.602\ 2$	
	$\hat{\beta}_0=\bar{y}-\hat{\beta}\bar{x}=28.534\ 0$	
	$\hat{y}=28.534\ 0+130.602\ 2x$	

则可得另一种表达形式：

$$\hat{y}=\bar{y}+\hat{\beta}(x-\bar{x}) \tag{1-10}$$

由此可知回归直线过$(\bar{x},\bar{y})$这一点，对作回归直线是很有帮助的。

当观测值较大(或过小)时，为简化计算，可以对数据作如下线性变换，令

$$x'=\frac{x-c_x}{d_x},\quad y'=\frac{y-c_y}{d_y} \tag{1-11}$$

其中 c_x,c_y,d_x,d_y 都是适当选取的常数，且 $d_x\neq0,d_y\neq0$。为建立 y 关于 x 的线性回归方程，可以先建立 y'关于 x'的线性回归方程，然后

再将变换式(1-11)代入,为此先对$(x_i, y_i)(i=1,2,\cdots,n)$,作变换(1-11),得到$(x_i', y_i')(i=1,2,\cdots,n)$,以这 n 组观测值,利用公式(1-9)求得一元线性回归方程:

$$\hat{y}' = \hat{\beta}_0' + \hat{\beta}' x' \tag{1-12}$$

再将变换式(1-11)代入式(1-12),整理后得:

$$\begin{aligned}
\hat{y} &= c_y + d_y\left(\hat{\beta}_0' + \hat{\beta}' \frac{x - c_x}{d_x}\right) \\
&= c_y + d_y\hat{\beta}_0' - \frac{d_y}{d_x}\hat{\beta}' c_x + \frac{d_y}{d_x}\hat{\beta}' x \\
&= \hat{\beta}_0 + \hat{\beta} x
\end{aligned}$$

其中

$$\hat{\beta}_0 = c_y + d_y\hat{\beta}_0' - \frac{d_y}{d_x}\hat{\beta}' c_x$$

$$\hat{\beta} = \frac{d_y}{d_x}\hat{\beta}'$$

例 2　为研究黄铜延性 $y(\%)$关于退火温度 $x(℃)$之间的关系,现收集了如下数据:

$x(℃)$	300	400	500	600	700	800
$y(\%)$	40	50	55	60	67	70

试求 y 关于 x 的一元线性回归方程。

解　由于 x 的取值为 100 的倍数,y 的取值大部分大于 50,故为了简化起见,可以令

$$c_i = 500,\ d_x = 100,\ c_y = 50,\ d_y = 1,$$

即令

$$x' = \frac{x - 500}{100},\quad y' = y - 50 \tag{1-13}$$

这样我们可以利用表 1-4 与表 1-5 求出 y'关于 x'的回归方程。

表 1-4

序号	x	y	x'	y'
1	300	40	-2	-10
2	400	50	-1	0
3	500	55	0	5
4	600	60	1	10
5	700	67	2	17
6	800	70	3	20
Σ			3	42

表 1-5

$\Sigma x_i'=3$	$\Sigma y_i'=42$	
$\bar{x}'=0.5$	$\bar{y}'=7$	
$\Sigma x_i'^2=19$	$\Sigma x_i'y_i'=124$	$\Sigma y_i'^2=914$
$\frac{1}{n}(\Sigma x_i')^2=1.5$	$\frac{1}{n}(\Sigma x_i')(\Sigma y_i')=21$	$\frac{1}{n}(\Sigma y_i')^2=294$
$l_{x'x'}=17.5$	$l_{x'y'}=103$	$l_{y'y'}=620$
	$\hat{\beta}'=5.885\ 7$	
	$\hat{\beta}_0'=4.057\ 1$	
	$\hat{y}'=4.057\ 1+5.885\ 7x'$	

再将式(1-13)代入 $\hat{y}'$ 并化简，得

$$\hat{y}-50=4.057\ 1+5.885\ 7\cdot\frac{x-500}{100}$$

故
$$\hat{y}=24.628\ 6+0.058\ 857x$$

这与直接用 x 与 y 求得的回归方程是一致的，但这里计算较简单。当然，对此问题，数据变换方式很多，不一定非用式(1-13)。例如也可令 $y'=\frac{y-40}{10}$，$x'=\frac{x-300}{100}$，使 x'，y' 中不出现负数。总之，变换是以方便为原则的。

三、回归方程的显著性检验

从上面介绍的求 y 关于 x 的一元线性回归方程的过程可知，在

计算过程中并不需要假定 y 与 x 一定有相关关系。即使是对于 n 对杂乱无章的数据(x_i,y_i)，同样可以按式(1-8)求得回归系数的最小二乘估计，从而获得 y 关于 x 的一元线性回归方程。但是这种方程毫无实际意义。前面曾经指出，只有在散点图上看到 n 个点落在一直线附近时才能认为 y 与 x 之间可配一元线性回归方程。为此我们必须定量地给出在什么情况下可以认为"n 个点落在一直线附近"，从统计上讲也就是 $E(y)$必须随 x 的变化而线性变化，即式(1-2)中的 β 不能等于 0。所以问题就变成了去检验假设 $\beta=0$ 是否为真。若 $\beta=0$ 为真，说明不管 x 如何变化，$E(y)$并不是随 x 而线性变化的；反之若 $\beta\neq0$，则当 x 变化时，$E(y)$是随 x 而线性变化的只有这时回归方程才是有意义的。

为检验假设 $\beta=0$ 是否为真，我们可以从分析引起各 $y_i(i=1,2,\cdots,n)$的不同原因着手。n 个 y_i 的值之所以不同，可能有两个方面的原因：其一，若 $E(y)$确是随 x 线性变化的话，那么 x 的取值不同就是一个原因；其二是其他一切因素的影响，若前一方面的影响是主要的，那么 $\beta\neq0$，方程是有意义的，否则方程就没有意义，为此必须把这两个原因所引起的 y_i 的波动大小从 y_i 的总的波动中分解出来。

记

$$S_{\mathrm{T}}=l_{yy}=\sum_i(y_i-\bar{y})^2$$

称它为总的偏差平方和，它反映了各 y_i 的波动大小。由于

$$\begin{aligned}S_{\mathrm{T}}&=\sum_i(y_i-\bar{y})^2\\&=\sum_i(y_i-\hat{y}_i+\hat{y}_i-\bar{y})^2\\&=\sum_i(y_i-\hat{y}_i)^2+\sum_i(\hat{y}_i-\bar{y})^2+2\sum_i(y_i-\hat{y}_i)(\hat{y}_i-\bar{y})\end{aligned}$$

由式(1-6)可知

$$\begin{aligned}&\sum_i(y_i-\hat{y}_i)(\hat{y}_i-\bar{y})\\&=\sum_i(y_i-\hat{\beta}_0-\hat{\beta}x_i)(\hat{\beta}_0+\hat{\beta}x_i-\bar{y})=0\end{aligned}$$

故