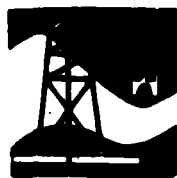


金刚石钻探成本预测的数学模型

沈洁清

(河北地质学院)



钻探技术

金刚石钻探的进一步发展不仅依赖于本身技术的进步,还取决于它的管理水平的提高。为提高其经济效益和管理水平,钻探成本的确定是十分必要的。

金刚石钻探是一个多环节的工艺过程,因此,对其成本的影响也是多因素的,这就为成本的精确制定增加了难度。到目前为止,有关金刚石钻探成本研究的文献所见甚少,本文采用了数理统计学中多元线性回归原理和电子计算机技术,探讨金刚石钻探成本的控制问题

影响成本预测因素的选取

应用多元线性回归预测成本,必须首先确定影响成本预测的诸因素。众所周知,影响成本预测的因素很多,如地层岩石情况,钻孔的孔径、孔深、倾角情况,钻探质量要求,钻探效率,等等。根据理论分析和实际资料,我们选取下列七个因素作为成本预测的影响因素:即钻探总进尺量(一个矿区的总工作量影响到基建投资,组成和撤销等不变成本,又影响到钻探熟练程度、钻进速度等可变成本);平均钻孔深度(一般讲随孔深增加,单位成本增加);钻孔倾角或顶角;岩石平均可钻性级别;岩石平均破碎程度;钻孔优质孔率(我们用优质孔率来替代岩心钻探质量指标,其优点是综合性强,用它对整个孔进行评价,比某一项钻探质量指标更能说明钻探质量。一般讲质量要求高,优质孔率大,则钻探成本也高);钻探钻月效率。钻孔口径因金刚石岩心钻探目前基本上是以60毫米一种规格为主,所以孔径的影响

不予考虑。岩石研磨性对钻头寿命、钻进速度影响很大,但是由于目前缺少这方面资料,故未列入。根据四个地矿局所属的十一个队的资料,设钻探单位成本为因变量 Y ,影响成本的因素分别为 X_1 (钻探总进尺米数), X_2 (平均孔深), X_3 (平均岩石可钻性级别), X_4 (岩石平均破碎程度), X_5 (钻孔顶角), X_6 (优质孔率), X_7 (钻月效率),列表如表1所示。

数学模型和统计量

多元线性回归方法是研究一个因变量与多个自变量之间成线性相关关系的数学理论和方法。因变量 Y 随自变量 $X_1, X_2, \dots, X_n$ 而变化,收集有 m 组数据后($Y_i, X_{1i}, X_{2i}, \dots, X_{ni}$) $i=1, 2, \dots, m$,应先根据这些资料检验 Y 与 $X_1, X_2, \dots, X_n$ 之间是否呈线性关系,只有它们之间大致呈线性关系时,才能利用线性回归方法。从散点图图1可以看出是呈大致呈线性关系的。这样,就可在 Y_i 与 $X_{1i}, X_{2i}, \dots, X_{ni}$ 之间确定一个线性回归方程来反映它们之间的相关关系。其方程为:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots + b_i X_i + \dots + b_n X_n \quad (1)$$

式中, Y —回归方程计算值,即单位成本计算值;

b_0 —回归方程常数项;

b_i — Y 对 X_i 的回归系数;

X_i —自变量,即影响单位成本因素。

我们采用HKC—8800微型电子计算机计算。其算法基本上分三步:第一步是根据输入的表1数据以最小二乘法为基础求出正规方程组。

表 1

收集数据组数	单位成本 (元/米)	总进尺数 (米)	平均孔深 (米)	平均岩 岩级别	岩石平均 破碎程度	钻孔顶角 (度)	优质孔率 (%)	钻月效率 (米/钻月)
m	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	70.02	15599	420	6.0	0.31	0	80	283
2	67.30	25206	390	6.0	0.28	0	85	244
3	69.28	19763	403	6.2	0.30	1.5	80	360
4	73.02	20457	370	6.5	0.40	6.0	90	382
5	75.40	25146	230	8.2	0.57	10.1	90	443
6	73.68	23566	357	7.2	0.48	15.0	85	451
7	68.35	17509	300	7.0	0.20	9.0	79	483
8	80.20	10322	305	8.5	0.68	10.0	80	496
9	93.75	8685	501	9.3	0.70	18.0	91	523
10	94.30	9024	520	8.7	0.96	20.6	95	531
11	86.50	11191	470	9.2	0.81	11.2	80	538

*岩石完整为0，岩石破碎为1。

本表资料由温国勇同学收集。

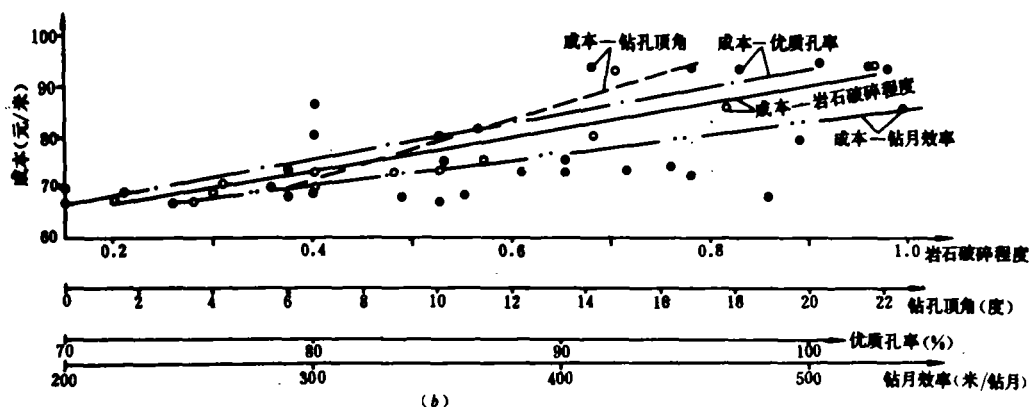
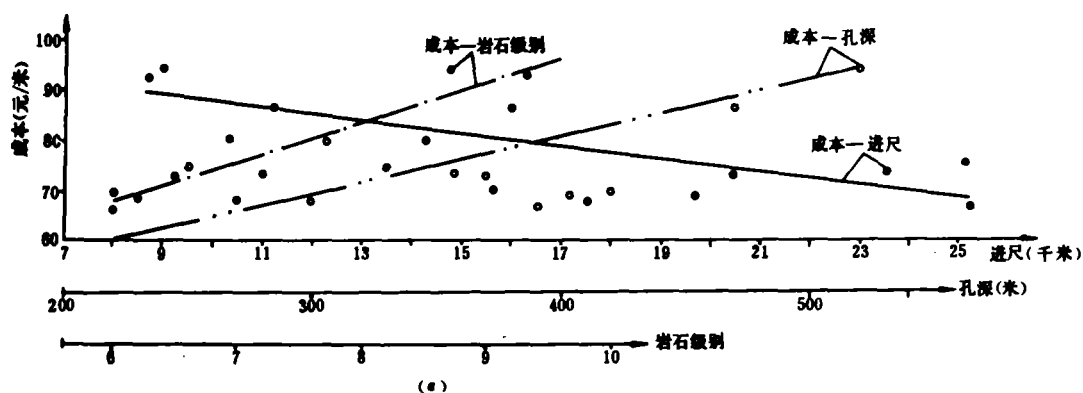


图 1 散点图

第二步是利用高斯消元法求解方程组，得回归方程。第三步是求出方差分析和 F 检验的统计量值

U 、 Q 、 S 、 R 和 F 值，并打印结果。简单流程图图为图 2 所示。

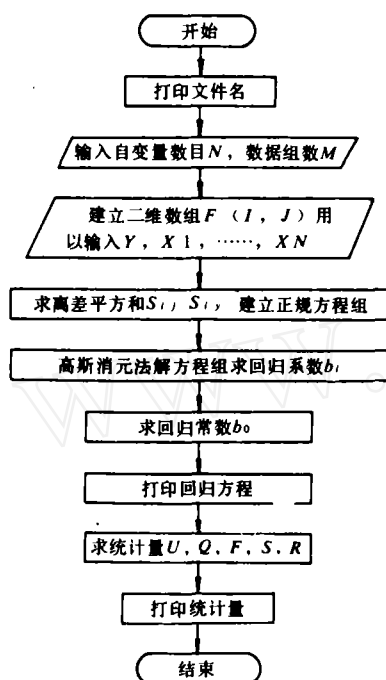


图2 简单流程框图

输入表1数据, 利用多元线性回归的程序, 得如下结果的数学模型和统计量:

$$Y = 13.79 - 3.62 \times 10^{-4} X_1 + 0.024 X_2 + 3.72 X_3 + 5.08 X_4 + 0.12 X_5 + 0.37 X_6 - 7.06 \times 10^{-3} X_7 \quad (2)$$

$$U = 983.2068$$

$$Q = 1.6953$$

$$F = 248.5546$$

$$S = 0.7517$$

$$R = 0.9991$$

对数学模型和统计量的讨论与分析

1. 各自变量对Y的关系

总进尺数 X_1 和钻月效率 X_2 与钻探单位成本 Y 成负相关。即总进尺数愈大, 钻探单位成本愈低, 这是与实际情况相符的。钻月效率愈高, 单位成本愈低。其他几个因素则成正相关。平均孔深愈深, 岩石可钻性等级愈高, 岩石破碎程度愈大, 钻孔顶角越大, 钻孔质量要求愈高, 则相应的要求钻探单位成本愈高。

2. 回归平方和 U 值和剩余平方和 Q 值

回归平方和 U 表示方程中考虑的 n 个自变量对 Y 的变差的总影响:

$$U = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \quad (3)$$

剩余平方和 Q 表示方程中考虑的 n 个自变量以外的因素对 Y 的影响:

$$Q = \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (4)$$

式中, $\bar{Y}$ — m 组数据 Y 的平均值。

U 愈大, Q 愈小, 表示 Y 与这些自变量的线性关系愈密切, 回归的规律性愈强, 效果也就愈好。由计算结果得知 $U \gg Q$, Q 值只有1.6953, 而 U 值高达983.2。这说明对钻探单位成本 Y 的影响因素, 从成本预测的角度来看几乎都包括在内了。总的讲, 选取上述七个因素作为单位成本的影响因素是可取的, 合理的。但是为了使得收集资料 and 计算工作更简单些, 在对单位成本计算值影响不大的情况下可以对七个因素作进一步的筛选, 以求抓住对成本影响的主要因素。这一问题将在“最优”成本数学模型一节中讨论。

3. 剩余标准差 S 值

剩余标准差 S 表示在排除了 X_i 对 Y 的影响后, 衡量 Y 随机波动大小的一个估计量。

$$S = \sqrt{\frac{Q}{m-n-1}} \quad (5)$$

若收集资料的组数 m 愈多, 统计的规律性愈大, S 愈小, 则 Y 值的波动愈小, 精度愈高。计算得 $S = 0.7517$, 是很小的, 表明公式(2)计算的单位成本预测值是相当精确的。

4. 复相关系数 R 值

复相关系数 R 表示回归平方和 U 在总平方和 S_{yy} 中所占的比值。 $S_{yy} = U + Q$ 。

$$R = \sqrt{U/S_{yy}} = \sqrt{1 - Q/S_{yy}} \quad (6)$$

R 愈大(愈接近于1), 表示 Y 与 $X_1, X_2, \dots, X_n$ 的线性关系愈密切。计算得 $R = 0.9991$, 很接近于1, 表示公式(2)中 Y 与七个自变量的线性关系很密切。

5. F 检验值

对多元回归方程进行显著性检验, 即 Y 与 n 个自变量 $X_1, X_2, \dots, X_n$ 之间线性关系是否密切的检验采用 F 检验法:

$$F = \frac{U/n}{Q/(m-n-1)} \quad (7)$$

设显著性水平为 α ，查 F 分布表 $F_{\alpha}(n, m-n-1)$ 的临界值 F_{α} 。把由上式计算的 F 值与 F_{α} 值进行比较，若 $F > F_{\alpha}$ 表示回归显著性较好，

$F < F_{\alpha}$ 表示回归不显著。

计算得 $F=248.55$ 。若设显著性水平 $\alpha=0.01$ ，则 $F_{\alpha}(n, m-n-1)=F_{0.01}(7, 3)$ ，查 F 检验表 $F_{0.01}(7, 3)=27.7$ 。 $F \gg F_{\alpha}$ ，说明公式(2)有高度的显著性。

由上述统计量分析可知，所求得的数学模型回归方程(2)选取的单位成本影响因素是全面的，它们与单位成本 Y 的线性关系很密切，公式(2)式的精确度是相当高的，有使用价值。

“最优”单位成本数学模型

所谓“最优”回归方程是指只包含所有对 Y 有显著影响的变量，而不包含对 Y 不显著的变量。

本文用标准回归系数 b_i 来分清多个因素(即自变量)对因变量 Y 影响的主次关系。因为按照回归系数 b_i 的意义，它表示在其他所有因素不变情况下 X_i 对 Y 的影响程度。但由于自变量所取单位有

所不同，为消除这个影响，用标准回归系数 b_i' 这个概念。它的绝对值愈大，相应的 X_i 因素对 Y 的影响就愈大。 b_i' 与 b_i 的关系如下：

$$b_i' = b_i \sqrt{\frac{S_{yy}}{S_{xx}}} \quad (8)$$

式中： S_{xx} — X_i 的离差平方和，

S_{yy} — Y 的离差平方和(总平方和)。

根据上述公式(8)分别求得： $b_1' = -0.1665$ ， $b_2' = 0.3763$ ， $b_3' = 0.4827$ ， $b_4' = 0.1270$ ， $b_5' = 0.0840$ ， $b_6' = 0.2116$ ， $b_7' = -0.0717$ 。从七个标准回归系数来看，其绝对值最小的为 b_7' (钻月效率)。其次是 b_5' (钻孔顶角)。再其次是 b_1' (岩石破碎程度)。

分别剔除 X_7 、 X_5 、 X_1 时的回归方程略，统计量见表2。

当同时剔除 X_7 、 X_5 时，得回归方程：

$$Y = 6.72 - 3.47 \times 10^{-4} X_1 + 0.026 X_2 + 3.70 X_3 + 5.17 X_4 + 0.42 X_6 \quad (9)$$

当同时剔除 X_7 、 X_5 、 X_1 时，得回归方程：

$$Y = -0.54 - 3.60 \times 10^{-4} X_2 + 0.028 X_3 + 4.42 X_4 + 0.47 X_6 \quad (10)$$

公式(9)(10)的统计量值见表2。

不同自变量时的统计量值

表 2

剔除自变量	回归平方和 [*] U	剩余平方和 Q	F 检验值 F	剩余标准差 S	复相关系数 R
未剔除	983.21	1.70	248.55	0.75	0.9991
剔除钻月效率 X_7	982.88/0.33	2.02	321.03	0.71	0.9990
剔除钻孔顶角 X_5	982.62/0.59	2.29	286.64	0.76	0.9988
剔除岩石平均破碎程度 X_1	981.67/1.54	3.83	170.63	0.98	0.9981
剔除 X_7 、 X_5	982.61/0.60	2.29	429.29	0.68	0.9988
剔除 X_7 、 X_5 、 X_1	980.36/2.85	4.55	323.44	0.87	0.9977

* 分子表示 U 值，分母表示未剔除自变量时 U 值—剔除某自变量时 U 值之差。

由表2可见，预测成本时，考虑总进尺数、平均孔深、岩石平均可钻性级别、岩石平均破碎程度、钻孔顶角、优质孔率、钻月效率七个因素是很全面的。由上述因素中剔除钻月效率 X_7 和钻孔顶角 X_5 两个因素后所建立的回归方程(9)的统计量 U 减少、 Q 增加是很小的。 S 值比七因素时更小，即成本预测值受随机波动影响的程度更小。 R 值变化也很小，因此，公式(9)可以看作是本

文所述的“最优”数学模型，用于预测成本。

成本区间估计

用公式(2)、公式(9)我们可以计算出钻探单位成本的估算值，但毕竟不是实际的单位成本值，这两者之间有多大误差？估算值的误差范围多大？这一误差范围的可靠性多大？这些问题对于加强成本管理无疑是必须回答的。求出单位成本的区

间范围并确定包含这个范围的可靠程度,是我们要进行区间估计的目的和意义。

根据概率论正态分布的情况,设置信度(可靠程度)为 $1 - \alpha$, 则 Y 在这一置信度的区间为:

$$\hat{Y} \pm t_{\alpha/2} (m-1) \times \frac{S}{\sqrt{m}}$$

根据收集数据组数 m 及 α 值,查 t 分布表即可求出 $t_{\alpha/2} (m-1)$ 的值,从而计算出置信区间范围。

若设 $1 - \alpha = 0.95$, 即要求单位成本的可靠程度达95%时, $\alpha/2$ 即为0.025, 则 $t_{0.025} (11-1) = 2.23$ 。利用公式(2)的统计量 S 求得:

$$t_{\alpha/2} (m-1) \frac{S}{\sqrt{m}} = t_{0.025} (11-1) \frac{0.75}{\sqrt{11}} =$$

$$2.23 \times 0.75 / \sqrt{11} = 0.5 \text{ 元/米。}$$

利用“最优”成本模型公式(9)时的统计量 S , 求得:

$$t_{0.025} (11-1) \times 0.68 / \sqrt{11} = 0.45 \text{ 元/米。}$$

即分别利用公式(2)、(9)求出单位成本预测值的可靠程度为95%时,其区间估计范围分别为: $[Y \pm 0.5 \text{ 元}]$ 和 $[Y \pm 0.45 \text{ 元}]$ 。

要求可靠程度为90%时,区间估计范围分别为: $[Y \pm 0.41 \text{ 元}]$ 和 $[Y \pm 0.37 \text{ 元}]$ 。

要求可靠程度为99%时,区间估计范围分别为: $[Y \pm 0.75 \text{ 元}]$ 和 $[Y \pm 0.65 \text{ 元}]$ 。

即要求单位成本区间估计范围越小,其可靠程度越低。

若再考虑到回归方程本身的稳定性和各自变量取值距离其统计数平均值的远近不同,单位成本的区间估计将会比上述区间更大。两值愈靠近,区间估计范围愈小,两者愈远,区间估计愈大。

应用实例

1. 广西地矿局第九地质钻探工程队1983年金刚石岩心钻探钻进尺9778米,平均孔深465.62

米,岩石平均可钻性级别6.5级,优质孔率70%,实际单位成本68.19元/米。岩石破碎程度没有资料,选用公式(10)计算,求得成本预测值为70.60元/米,比实际成本高2.41元/米,误差为实际成本的3.53%。

2. 北京地矿局101队1984年总进尺11191.34米,平均孔深315米,岩石平均可钻性为8.02级,岩石平均破碎程度0.62,优质孔率92%,实际钻进成本81.42元/米。用“最优”模型公式(9)计算得成本预测值为82.88元/米,比实际成本高1.46元/米,误差为实际成本的1.79%。

例1误差较大,由表2的统计量可知,这是由该情况下, U 减小较多, S 较大, R 较小所致。但是由于实际生产中常常对岩石破碎程度没有定量数值,因此,公式(10)还有一定使用价值。至于应用“最优”方程的例2,其误差是较小的,有使用价值。

应用多元线性回归预测钻探成本是可行的,有实际使用价值。由于它以实际统计资料为依据所求得,因此实际结果总会有一半在回归线以上,一半在回归线以下,所以该法所得结果基本上是反映平均水平的,是合适的。剔除那些明显的非正常因素,以保证统计数据本身的真实性是所求成本精确与否的基础。统计的数组愈多。计算成本愈精确。

由以上两例可见,利用“最优”模型公式(9)能取得接近实际的数值;但在实际生产中,往往会缺少某项统计指标(例1就缺少岩石破碎程度指标),这时,也可使用其他公式。如例1就是使用公式(10)计算的,虽然其误差比例2稍大,但仍可基本反映出实际成本。说明,本文提出的数学计算模型公式,对成本预测有其实用价值。

另外,用该法也同样可以对钻探成本进行控制,为实现成本目标管理,提供出科学的方法。