

报名号 # 1088

1 问题重述

1.1 问题的背景

20 世纪以来，随着人口的持续增长，食品、水、资源等问题越来越多的暴露出来。英国科学家约翰·贝丁顿于 3 月 18 日在英国《卫报》发表文章以发出警告，称世界将面临“完美风暴”，出现大动荡。

1.2 问题的提出

文章中说：再过20多年，全球人口将达到大约83亿。这意味着全球资源需求将变得极大，资源短缺压力将显著增加，气候变化还可能使这些问题更加恶化。另外，目前全球的粮食储存量只有年消费量的14%，不足以抵抗干旱或者洪灾的发生。而且表示：我们的粮食储备是50 年来最低的，到2030 年，我们对粮食的需求会增加50%，同时，我们对能源的需求也会增加50%，淡水的需求会增加30%。题目要求建立评估中国在“完美风暴”中的地位 and 作用的合理模型，并结合通过收集相关数据论贝丁顿的观点的同时运用所建立“完美风暴”危机预测模型，判断出通过中国自身或几个国家的协同作用调整来有效地降低“完美风暴”发生的可能性。

2 模型假设

2.1 对全球人口预测模型的假设

- (1) 在预测时间内，不发生大的疫情，灾难或战争等引起人口重大变化的事件；
- (2) 在预测时间内，总和生育率、死亡率和出生性别比不会发生大的波动，可以用往年平均值代替预测值；
- (3) 人口预测的参数主要由当前的实际情况和今后的政策决定；
- (4) 生育情况在预测时间内保持不变，并且假设一胎只生一个；
- (5) 流动人口按农业人口和非农业人口占目前的人口总量的比例分配给他们；
- (6) 在达到某一范围后，人口增长将不受政府干预而是由各人所占的资源决定的；

2.2 其余假设

- (7) 由于数据搜寻的困难性，假设我们查到的数据真实可靠，对最终结果的预测不会产生太大误差；
- (8) 假设 2030 年前没发生重大变化，如全球性的自然灾害、传染病、大规模的移民等；
- (9) 在预测期间，没有大规模使用海水淡化技术来获取淡水资源，并且也没有人为的大规模破坏污染淡水资源；
- (10) 假设以和平和发展为主题的国际社会局势不会产生转折性的改变，各国际组织（如联合国粮农组织 FAO 等）对各种危机重视程度和解决措施不会产生转折性的提升。

报名号 # 1088

3 符号说明

为简化问题的分析和数据的处理，我们在以后的文字中将使用如下符号代表变量：

表 1 文中的部分变量符号说明

符号	描述
$x_m(i, t)$	第 t 年 i 岁 (满 i 周岁而不满 $i+1$ 周岁) 的男性人数 ($t=0, 1, 2, \dots$; $i=0, 1, \dots, m$)
$x_w(i, t)$	第 t 年 i 岁 (满 i 周岁而不满 $i+1$ 周岁) 的女性人数 ($t=0, 1, 2, \dots$; $i=0, 1, \dots, m$)
$d_m(i, t)$	第 t 年 i 岁的男性死亡率 ($t=0, 1, 2, \dots$; $i=0, 1, \dots, m$)
$d_w(i, t)$	第 t 年 i 岁的女性死亡率 ($t=0, 1, 2, \dots$; $i=0, 1, \dots, m$)
$b(i, t)$	第 t 年 i 岁女性的生育率 ($t=0, 1, 2, \dots$; $i=0, 1, \dots, m$)
$f(t)$	第 t 年的出生人口数 ($t=0, 1, 2, \dots$)
$f_w(t)$	第 t 年的出生人口数的女婴人数 ($t=0, 1, 2, \dots$)
$d_{w0}(t)$	第 t 年出生女婴的死亡率 ($t=0, 1, 2, \dots$)
$xf(i, t)$	非农业人口在第 t 年 i 岁的人口数量
$xn(i, t)$	农业人口在第 t 年 i 岁的人口数量
fg	到 2030 年粮食需求量的增长率
wg	到 2030 年淡水需求量的增长率
oc	到 2030 年石油需求量的增长比例

其它符号说明请见公式。

4 问题分析

本文的主要目的是解决以下三个问题：(1) 建立评估中国在“完美风暴”中的地位 and 作用的合理模型，(2) 搜寻有效数据并建立数学模型论证英国科学家约翰·贝丁顿的观点，即到 2030 全球对粮食和能源的需求会增加 50%，淡水的需求会增加 30% 等。要论证以上一些结论，需要分别建立全球人口预测模型以及全球粮食和能源需求等模型。根据这这些模型利用搜索得到的有效数据预测出 2030 年全球人口数量、全球粮食和能源等的需求量。因全球淡水需求量与全球人口数量成某一比例，由人口预测模型所得数据可以求得 2030 年时淡水需求量。由模型所得数据便可以判断贝丁顿结论的正确性的同时建立“完美风暴”危机预测模型。能引发“完美风暴”的影响因子有：人口数量、粮食需求量、淡水资源需求量、能源需求量等。其中人口数量因素是导致其他三个因素基数改变的主要诱因。用主成分分析方法便可确定以上几个因子间的相关关系。全球人口数量的增加、粮食需求量和能源需求量的增加都会导致全球淡水需求量的增加。(3) 判断出通过中国自身或几个国家的协同作用调整来有效地降低“完美风暴”发生的可能性。

报名号 # 1088

5 模型的分析、建立和求解

5.1 评价中国在“完美风暴”中地位和作用的模糊数学综合评判模型

5.1.1 评价中国在“完美风暴”中地位和作用模型的建立

气候变化和人口增长将导致食品、水和能源短缺，进而引发大规模移民、公共骚乱和国际冲突。如果未来数年内没有充分准备的话，到2030年，世界将面临“完美风暴”，出现大的动荡；中国也必然会遭受到“完美风暴”。那么，中国在世界“完美风暴”中的具有什么样的地位呢？为了更好的解决好这样问题，我们根据人口、粮食、淡水、能源等情况采用模糊数学综合评判来分析中国在“完美风暴”中的地位和作用。

5.1.1.1 模糊数学综合评判模型的建立

(1) 确定对象、因素集和评语集

首先中国为评判对象。根据所收集的数据，选取四项因素构成评判因素集 U ，即人口 u_1 、粮食 u_2 、淡水 u_3 、能源 u_4 ，评语集 V 为非常高、比较高、一般高、高、不高 5 级，即 $V = \{v_1, v_2, \dots, v_5\}$ 。

设评价问题的方案为 M ，指标标集为 $N = (N_1, N_2, N_3, N_4)$ ，并且方案 M 对指标 N_j 的属性值记为 X_{ij} ($i=1, j=1, 2, 3, 4$)，则矩阵 $X = (x_{ij})_{1 \times 4}$ 表示 4 个目标对评价对象的目标特征值矩阵：

$$X = \begin{bmatrix} x_{11} & x_{21} & x_{31} & x_{41} \end{bmatrix} \quad (5.1-1)$$

现将矩阵 (5.1-1) 中的特征指标值转化为相应的指标相对隶属度，这里就分为对越大且越优型指标的隶属度构造和对越小且越优型指标隶属度构造。对越大且越优型指标，其隶属度构造表示为：

$$r_{ij} = x_{ij} - \frac{x_j^{\min}}{(x_i^{\max} - x_i^{\min})} \quad (5.1-2)$$

同理，对越小且越优型指标，其隶属度构造表示为：

$$r_{ij} = x_i^{\min} - \frac{x_{ij}}{(x_i^{\max} - x_i^{\min})} \quad (5.1-2')$$

这样就可以将 (5.1-1) 式的目标值转化为指标隶属度矩阵：

$$R = \begin{bmatrix} r_{11} & r_{21} & r_{31} & r_{41} \end{bmatrix} \quad (5.1-3)$$

假如将矩阵 (5.1-3) 中每一行的最大值抽出来组成一个新的向量，记为 r_g ，并称

$$r_g = (r_{g1}, r_{g2}, r_{g3}, r_{g4}) = (\max r_{1i}, \max r_{2i}, \max r_{3i}, \max r_{4i}) \quad (5.1-4)$$

为理想的比较重要；

同理，假如将矩阵 (5.1-3) 中每一行最小值抽出组成一个新的向量，记为 r_b ，并称

$$r_b = (r_{b1}, r_{b2}, r_{b3}, r_{b4}) = (\min r_{1i}, \min r_{2i}, \min r_{3i}, \min r_{4i}) \quad (5.1-4')$$

为理想的不重要。

则评价对象都以一定的隶属度 u_{gi} ， u_{bi} 隶属于比较重要和不重要，这里称 u_{gi} ， u_{bi} 是

报名号 # 1088

方案的优属度和劣属度，这样就可以构造出最优模糊评判划分矩阵：

$$U = \begin{bmatrix} u_{g1} & u_{b1} \end{bmatrix} \quad (5.1-5)$$

其中，(5.1-5) 式中满足： $0 \leq u_{gj} \leq 1$, $0 \leq u_{bj} \leq 1$, $u_{gj} + u_{bj} = 1$, $j = 1$ 。

又设评价指标的加权向量为 $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)^T$, $\sum_{i=1}^4 \lambda_i = 1$ 。

为了求解评价对象相对比较重要的相对隶属度 u_{gj} 的最优值，我们引入如下的优化准则：方案 j 的欧式加权距最优距离平方与欧式加权距最劣距离平方之和为最小，即目标函数 Z 为：

$$\min \left\{ Z = u_{gj}^2 \sum [\lambda_i (r_{ij} - r_{gj})]^2 + (1 - u_{gj})^2 \sum [\lambda_i (r_{ij} - r_{bj})]^2 \right\} \quad (5.1-6)$$

其中，对象 j 的欧式加权距最优距离为：

$$S_{gj} = u_{gj} \sqrt{\sum_{i=1}^m [\lambda_i (r_{gi} - r_{ij})]^2} \quad (5.1-7)$$

同样，对象 j 的欧式加权距最劣距离为：

$$S_{bj} = u_{bj} \sqrt{\sum_{i=1}^m [\lambda_i (r_{ij} - r_{bi})]^2} \quad (5.1-7')$$

其中， $u_{bj} = 1 - u_{gj}$ 。

这里求目标函数的导数，且令导数为 0，可得到：

$$\begin{aligned} u_{gj} &= \frac{\sum_{i=1}^{34} [\lambda_i (r_{ij} - r_{bj})]^2}{\sum_{i=1}^{34} [\lambda_i (r_{ij} - r_{bj})]^2 + \sum_{i=1}^{34} [\lambda_i (r_{gj} - r_{ij})]^2} \\ &= \frac{1}{1 + \frac{\left[\sum_{i=1}^{34} \lambda_i (1 - r_{ij}) \right]^2}{\sum_{i=1}^{34} (\lambda_i r_{ij})^2}} \end{aligned} \quad (5.1-8)$$

(5.1-8)式就是目标模糊评判模型，其中的 u_{gj} 就是决策的优属度，从而可以确定中国在“完美风暴”中的地位和作用。

5.1.1.2 熵值法——权重确定模型

根据所给定的目标 $B_i (i = 1, 2, 3, 4)$ 的性质和特点，对所评价对象 $D_j (j = 1)$ 的各个目标值 $Z_{ij} (i = 1, 2, 3, 4; j = 1)$ 进行归一化的处理。

对于目标值越小越优的目标，其归一化公式为：

$$Z_{ij}^* = 1 - \frac{Z_{ij}}{Z_i' + Z_i''} \quad (i = 1, 2, 3, 4; j = 1) \quad (5.1-9)$$

其中， Z_i' 、 Z_i'' 分别为各个目标的最大值和最小值。

报名号 # 1088

对于任意给定的目标 B_i ，其目标值 Z_{ij}^* ($j=1$) 的差异越大，则表明不同的方案之间的目标值相对强度越强，目标 B_i 对方案的比较作用也就越大，即说明它包含和传输的决策信息越多。

熵是系统状态不确定性（混沌态）的度量，可以被用来度量上述信息量的大小。用熵表示所有方案对目标 B_i 的总贡献量为：

$$e(B_i) = -K \sum_{j=1}^1 \frac{Z_{ij}^*}{E_i} \ln \left(\frac{Z_{ij}^*}{E_i} \right) (i=1,2,3,4)$$

其中， $E_i = \sum_{j=1}^1 Z_{ij}^* (i=1,2,3,4)$ ， $K \neq 0$ 且为常数， \ln 为自然对数，并且 $e(B_i) \geq 0$ 。

如果 Z_{ij}^* 对于给定的目标 i 全都是相等的，那么 $\frac{Z_{ij}^*}{E_i} = \frac{1}{n}$ ，此时 $e(B_i)$ 取得最大值，即 $e_{\max} = K \times \ln(1)$ 。

我们假设 $K = \frac{1}{\ln(4)}$ ，于是就有 $0 \leq e(B_i) \leq 1$ 。

对于给定的目标 B_i ， $Z_{ij}^* (j=1)$ 的差异越小，则说明 $e(B_i)$ 越大；当它们全部都相等时， $e(B_i) = e_{\max} = 1$ ，即表明该评价对象对于该目标而言贡献基本上没有差异，此时该目标对对象的评价基本上不能够提供有用的信息，因此其权系数值也等于零，该目标在进一步决策中可以删除。而当各个方案对该目标而言贡献相差越大时，该目标对于方案的比较作用就越大，即目标传输的决策信息也就越大，它的权系数值也就越大。

因此，权系数 λ 与熵 $e(B)$ 成反比关系。

考虑到归一化，我们定义各个目标的权系数为：

$$\lambda_i = \frac{1 - e(B_i)}{4 - \sum_{i=1}^4 e(B_i)} (i=1,2,3,4) \quad (5.1-10)$$

5.1.2 模型的求解

首先，通过对熵值法——权重确定模型的运用，最终求得各个目标的权系数为： $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 0.25$ ，很显然满足 $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1$ 。

由于 EXCEL 自带有很强大的数据处理功能和函数运算功能，因此我们就用 EXCEL 求解模型。得到中国的优属度为 0.316571，由此可以分析出中国在“完美风暴”中具有比较重要的地位，并发挥着相当大的作用，毕竟全世界 200 多个国家中，中国的优属度就占了 31.6571%。

5.2 引发“完美风暴”的危机预测模型的分析、建立和求解

首先收集了人口、粮食、能源、淡水、环境相应指标的数据，对“完美风暴”是否来袭进行建立模型研究，为了方便就转化为对人口模型、能源需求模型、水资源需求模型和粮食需求模型等这些子模型的建立研究（这里在“完美风暴”主成分分析法模型中得

报名号 # 1088

到验证), 从而对英国科学家约翰·贝丁顿的预言进行检验, 然后由各子模型共同构建危机预测模型验证“完美风暴”来袭的可能性。

5.2.1 引发“完美风暴”的聚类 and 主成分分析

5.2.1.1 引发“完美风暴”指标集的建立和聚类相关步骤

引发“完美风暴”的过程通常可以视为一个时间序列。在聚类前, 需对引发“完美风暴”的过程进行特征提取, 即进行引发“完美风暴”指标的选择。选择的原则主要是要选择那些对引发“完美风暴”现象影响显著的特征因子。同时选择的方法主要有3种: (1) 理论分析; (2) 影响因子相关分析; (3) 模式识别。

首先, 进行“完美风暴”主成分分析, 在尽可能保留原有“完美风暴”指标所含有的信息的基础上, 将原来较多的且具有相关性的引发“完美风暴”的指标简化为少数几个新的综合性指标; 其次, 利用上面分析的结果, 根据几个新的综合指标上的相似性来进行“完美风暴”聚类分析。

5.2.1.2 主成分分析模型及求解

主成分分析是一种化繁为简, 尽可能压缩指标数的降维(即空间压缩)技术, 其结果是把多个指标简化为少数几个新的综合指标。新的综合指标称为原指标的主成分, 并且按其所含有的信息量的大小依次称为第1主成分, 第2主成分……新指标要求满足: (1) 尽可能保留原有指标所含有的信息; (2) 各指标之间不相关, 即各自含有的信息不重叠。文献[5]对主成分分析方法的原理和性质作了详细介绍, 本模型仅给出如下主成分分析的计算步骤:

a. 输入“完美风暴”特征的相关值。

设对“完美风暴”测得 p 个指标(变量)的数据, 这 $1 \times p$ 个数据构成“完美风暴”的特征相关矩阵 X :

$$X = \begin{bmatrix} x_{11} & \dots & x_{1p} \end{bmatrix}$$

其中, $x_{1j} (j=1, 2, \dots, p)$ 为“完美风暴”的第 j 个指标的相关值。

b. 计算各引发“完美风暴”指标的均值 \bar{x}_j 和标准差 s_j :

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2} \quad (j=1, 2, \dots, p)$$

c. 将特征相关矩阵 X 标准化, 得标准化数据阵 $Y = (y_{1j})_{1 \times j}$, 再据此计算其关联矩阵

$$R = (r_{jk})_{p \times p};$$

$$y_{1j} = \frac{x_{1j} - \bar{x}_j}{s_j} \quad (j=1, 2, \dots, p)$$

$$r_{jk} = \frac{1}{n-1} \sum_{i=1}^n y_{1j} y_{1k} = \frac{1}{n-1} \frac{\sum_{i=1}^n (x_{1j} - \bar{x}_j)(x_{1k} - \bar{x}_k)}{s_j s_k} \quad (j, k=1, 2, \dots, p)$$

d. 求 R 的特征值及特征向量。若能够通过正交换 Q , 使得 $Q^T R Q = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$, 则 $\lambda_1, \lambda_2, \dots, \lambda_p$ 为 R 的 p 个特征值。为了方便, 不妨设 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, 则 Q 的各列

报名号 # 1088

$l_j = (l_{1j}, l_{2j}, \dots, l_{pj})'$ ，即为 λ_j 所对应的正交化特征向量， $j=1, 2, \dots, p$ 。

- e. 确立主成分：按照累计方差贡献率 $con > 90\%$ （或者85%）的准则，确定k，从而取前k个主成分：

$$con = \sum_{j=1}^k \lambda_j / \sum_{j=1}^p \lambda_j \quad Z_j = l_j' Y = l_{1j} Y_1 + \dots + l_{pj} Y_p \quad (j=1, 2, \dots, k)$$

其中， Y_1, \dots, Y_p 为标准化后的指标向量， $1 \leq k \leq p$ 。

- f. 计算前k个主成分的样本值，即：

$$Z_{1j} = \sum_{i=1}^p y_{1i} l_{ij} \quad (j=1, 2, \dots, k)$$

从而得到新指标矩阵（即主成分矩阵）： $Z = (z_{1j})_{n \times k}$ ，以其代替原“完美风暴”特

征 相关矩阵 $X = (x_{1j})_{n \times p}$ 做聚类分析，就可以将问题简化了，与此同时，主成分的个数也得到了确定。

- g. 写出主成分并给出主成分的得分函数：

主成分的原始指标 x_1, x_2, \dots, x_p 表示为

$$f_m = a_{1m} x_1 + a_{2m} x_2 + \dots + a_{pm} x_p, \quad (m=1, 2, \dots, k) \quad (5.2-1)$$

将原始数据代入（5.2-1）式就可以得到主成分值矩阵：

$$F = \begin{bmatrix} f_{11} & \dots & f_{1n} \\ \dots & \dots & \dots \\ f_{n1} & \dots & f_{nk} \end{bmatrix} \quad (5.2-2)$$

- h. 用主成分进行综合评价

该样本（“完美风暴”）的表现由主成分反映，用主成分计算各个样本的综合评价价值，进而对该样本进行排序和比较。一般用加权算术平均来综合并且以各主成分的方差贡献率为权重，即：

$$f = \alpha_1 f_1 + \alpha_2 f_2 + \dots + \alpha_k f_k \quad (5.2-3)$$

将该样本的主成分值（5.2-2）式代入（5.2-3）式可以得到该样本的主成分综合评价价值，进而可以进行综合比较和排序。

在提取了四个主成分时，各个主成分的方差贡献用 b_1, b_2, b_3, b_4 表示；则累积方差贡献率 con 就可以求得，即：

$$con = b_1 + b_2 + b_3 + b_4 \quad (5.2-4)$$

若（5.2-4）的结果满足大于等于90%（或者85%），就可以说明分析效果显著，否则视为分析效果不明显。这样就可得到各主成分的得分函数及主成分载荷分析，再用主成分分析进行综合评价，即由（5.2-3）式计算出综合评价价值。

5.2.1.3 数据分析过程和求解

通过对数据的挖掘搜索和在SPSS软件中的分析处理，就可利用主成分分析法对世界人口、粮食需求量、淡水需求量、能源需求量（具体数值见附录1、2、3、4）进行主成分分析，得到相关系数矩阵如下表（表2）和人口、粮食、淡水资源、能源权值如下表（表3）：

报名号 # 1088

表 2 原始指标的相关系数矩阵

correlation	人口peop	粮食food	淡水资源wate	能源enme
人口peop	1.0	0.656	0.7169	0.589
粮食food	0.656	1.0	0.3741	0.1477
淡水资源wate	0.7169	0.3741	1.0	0.8204
能源enme	0.589	0.1477	0.8204	1.0

表 3 人口、粮食、淡水资源、能源权值

correlation	人口peo	粮食food	淡水资源wate	能源enm
权值	0.6439	0.0908	0.0349	0.2304

由上表可知，人口与淡水资源、粮食、能源的相关系数都较大，即人口对淡水资源、粮食、能源的影响较大，说明引起淡水资源、粮食、能源问题的主要原因在人口问题。

根据各个方面的数据和资料，以及对引发“完美风暴”主成分分析法模型的求解结果（表2和表3）表明：粮食问题、淡水资源问题、人口问题以及能源问题是引发“完美风暴”的四个重要因素。

5.2.2 全球人口预测模型的分析、建立和求解

5.2.2.1 模型一 考虑性别的人口发展方程模型的建立和求解

本模型将男性和女性的人口数当作两个独立的系统进行预测，约定总人口发展趋势由两者共同决定。

设 $x_m(i, t)$, $x_w(i, t)$ 分别表示为第 t 年 i 岁 (满 i 周岁而不满 $i+1$ 周岁) 的男性人数和女性人数; $d_m(i, t)$, $d_w(i, t)$ 分别表示为第 t 年 i 岁的男性死亡率和女性死亡率, 则有:

$$d_m(i, t) = \frac{x_m(i, t) - x_m(i+1, t+1)}{x_m(i, t)}; i = 0, 1, 2, \dots, m-1; t = 0, 1, 2, \dots$$

$$d_w(i, t) = \frac{x_w(i, t) - x_w(i+1, t+1)}{x_w(i, t)}; i = 0, 1, 2, \dots, m-1; t = 0, 1, 2, \dots$$

于是可得, 男性人口发展的差分方程为:

$$x_m(i+1, t+1) = [1 - d_m(i, t)]x_m(i, t); i = 0, 1, 2, \dots, m-1; t = 0, 1, 2, \dots$$

同理可得, 女性人口发展的差分方程为:

$$x_w(i+1, t+1) = [1 - d_w(i, t)]x_w(i, t); i = 0, 1, 2, \dots, m-1; t = 0, 1, 2, \dots$$

又设 $b(i, t)$ 表示为第 t 年 i 岁女性的生育率, 则 $b(i, t)$ 可以分解为:

$$b(i, t) = \beta(t)h(i, t) \quad (5.2-5)$$

其中 $h(i, t)$ 为第 t 年年龄为 i 周岁的女性生育形式, 表示育龄妇女在不同年龄阶段的生育率的高低, 育龄区间为 $[i_1, i_2]$, 而且有 $h(i, t)$ 满足:

$$\sum_{i=i_1}^{i_2} h(i, t) = 1$$

所以对 (5.2-5) 式进行求和就可以得到:

报名号 # 1088

$$\beta(t) = \sum_{i=i_1}^{i_2} b(i, t)$$

$\beta(t)$ 表示在第 t 年每个育龄妇女平均生育的婴儿数目，即生育率总和，或生育的胎次，是控制人口的重要参数。

则第 t 年的出生人口数 $f(t)$ 可表示为：

$$f(t) = \sum_{i=i_1}^{i_2} b(i, t)x_w(i, t) = \beta(t) \sum_{i=i_1}^{i_2} h(i, t)x_w(i, t)$$

其中，出生人口数中，女婴数 $f_w(t)$ 为：

$$f_w(t) = f(t) \frac{100}{\varphi(t) + 100}$$

其中， $\varphi(t)$ 为在第 t 年出生人数的性别比，即每出生 100 名女婴则出生 $\varphi(t)$ 名男婴。

设 $d_{w0}(t)$ 表示第 t 年出生女婴的死亡率，即在第 t 年出生但未活到人口统计时刻的女婴比例，则第 t 年存活下来的女婴数 $x_w(0, t)$ 为：

$$x_w(0, t) = [1 - d_{w0}(t)]f_w(t) = \frac{100}{\varphi(t) + 100} [1 - d_{w0}(t)]\beta(t) \sum_{i=i_1}^{i_2} h(i, t)x_w(i, t)$$

同理可得：设

$$b'_w(i, t) = \frac{100}{\varphi(t) + 100} [1 - d_{w0}(t)] [1 - d_w(0, t)] h(i, t)$$

为了方便计算，我们引入了年龄向量 $X(t)$ ，存活率矩阵 $A(t)$ ，出生率矩阵 $B(t)$ ，其中：

$$\begin{aligned} X(t) &= [x(1, t), x(2, t), \dots, x(m, t)]' \\ A(t) &= \begin{bmatrix} 0 & 0 & \dots & 0 \\ 1 - d(1, t) & 0 & \dots & \dots \\ \dots & 1 - d(2, t) & \dots & \dots \\ 0 & 0 & 1 - d(m-1, t) & 0 \end{bmatrix} \\ B(t) &= \begin{bmatrix} 0 & \dots & 0 & b'(i_1, t) & \dots & b'(i_2, t) & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & 0 \end{bmatrix}_{m \times m} \end{aligned} \quad (5.2-6)$$

其中， $A(t)$ 、 $B(t)$ 的取值由相关数据资料来确定，而后就可以进一步的求解得到 $\beta(t)$ 的取值。

综上所述，上面的关系可以表示为：

$$X(t+1) = A(t)x(t) + \beta(t)B(t)X(t)$$

同理，我们引用女性人口数量的向量 $X_w(t)$ ，女性出生率的矩阵 $B_w(t)$ ，以及女性存活率的矩阵 $A_w(t)$ 。很显然， $X_w(t)$ 、 $B_w(t)$ 、 $A_w(t)$ 的形式与 (5.2-6) 式是相同的，则可将上面的关系表示为：

报名号 # 1088

$$X_w(t+1) = A_w(t)x_w(t) + \beta(t)B_w(t)X_w(t)$$

同理可得：男性的人口数量向量 $X_m(t)$ 、出生率矩阵 $B_w(t)$ 、存活率矩阵 $A_m(t)$ 的关系可以表示为：

$$X_m(t+1) = A_m(t)x_m(t) + \beta(t)B_m(t)X_m(t)$$

其中：

$$b_m'(i,t) = \frac{100}{\varphi(t)+100} [1-d_{m0}(t)][1-d_m(0,t)]h(i,t)$$

所以，总人口的发展方程模型表示为：

$$X(t+1) = X_m(t+1) + X_w(t+1) \quad (5.2-7)$$

其中， $A_w(t)$ 、 $B_w(t)$ 、 $A_m(t)$ 、 $B_m(t)$ 的取值由相关数据资料来确定，而后就可以进一步的求解得到 $\beta(t)$ 、 $\varphi(t)$ 的取值，即可以预测人口的发展趋势。

同时，根据该模型能够分别预测出未来的男性和女性人口数量的变化规律以及人口老龄化的发展规律。根据参数的不同取值，也就可以对人口增长进行中长期的预测。

5.2.2.2 模型二 在考虑非农业人口的基础上的人口预测的修改模型的建立和求解

模型一将男性和女性的人口量当作两个独立的方面进行模型的建立，根据它很容易计算出不同性别、不同年龄的人口分布规律。但是模型一是基于一个重要假设：仅考虑出生率死亡率的影响，不计农业人口转移为非农业人口等社会因素的影响，因而只适用于一个相对比较独立的人口发展预测系统。

模型二则是以模型一为基础，把非农业人口和农业人口作为两个预测系统，分别在应用模型一的同时考虑这两个预测系统之间的人口迁移情况，即农业人口转移为非农业人口，使非农业人口的比率符合当前社会人口发展的客观规律。

设非农业人口和农业人口在第 t 年 i 岁的人口数量分别为 $xf(i,t)$ 、 $xn(i,t)$ ，并且其余的变量、矩阵都做同模型一中相类似的定义。则可得到非农业人口的发展方程模型为：

$$\begin{aligned} XF(t+1) &= XF_m(t+1) + XF_w(t+1) \\ &= AF_m(t)XF_m(t) + \beta f(t)BF_m(t)XC_m(t) + AF_w(t)XF_w(t) + \beta f(t)BF_w(t)XC_w(t) \end{aligned} \quad (5.2-8)$$

同理可得：农业人口的发展方程模型为：

$$XN(t+1) = XN_m(t+1) + XN_w(t+1) \quad (5.2-9)$$

为了方便建立模型，我们这里定义非农业人口的比率为：

$$\text{非农业人口的比率} = \frac{\text{非农业人口}}{\text{全部人口}}$$

所以：

$$c(t) = \frac{\sum_i xf(i,t)}{\sum_i [xf(i,t) + xn(i,t)]}$$

其中， $c(t)$ 表示非农业人口的比率同时间变化而变化的函数，这是由当前的实际情况和

报名号 # 1088

今后的形势政策所决定的。

因此，由以上的式子可以推出第 t 到第 $t+1$ 年间由农业人口转为非农业人口的数量为 $move(t)$ ：

$$move(t) = \sum_i xn(i, t+1) - [1 - c(t+1)] \sum_i [xf(i, t) + xn(i, t)] \quad (5.2-10)$$

流动人口按照非农业人口和农业人口目前所占人口总量的比例分配给他们，则各个国家地区的人口总量可变更为：

$$\begin{cases} \sum_i xf'(i, t+1) = \sum_i xf(i, t+1) + move(t) \\ \sum_i xn'(i, t+1) = \sum_i xn(i, t+1) - move(t) \end{cases} \quad (5.2-11)$$

求解时，先确定函数 $c(t)$ ，按年迭代计算人口数量，再按式(5.2-10)和(5.2-11)变更，将变更后的结果迭代回式(5.2-8)和(5.2-9)即可。

根据以上的模型和相关的数据库（详见附录 1 全球人口的相关数据（单位：万人）），我们可以得到，到 2030 年全球人口的数量约为 814460.0132 万人。

5.2.2.3 全球人口预测模型验证

首先，为得到搜集到的 1987-2007 年世界人口数量（单位：万人）的一些关系，我们初步对数据进行分析 and 进一步处理，利用 MATLAB 的强大的绘图功能得到下图 1：

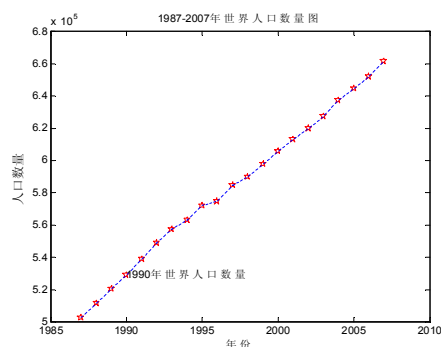


图 1 世界人口数量图

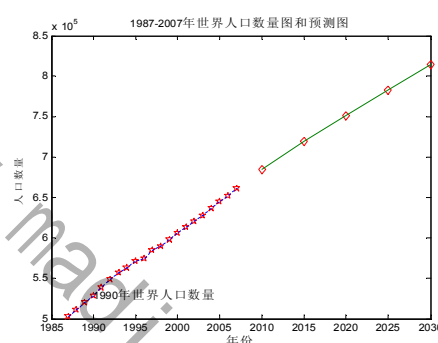


图 2 数据拟合后的世界人口数量图

为了能够更好的反应出该预测结果的相对准确性，我们对附录 1 1987-2007 年世界人口统计（单位：万人）利用数据拟合的 MATLAB 实现，对相关数据进行了进一步的分析，结果如下表 4 世界人口预测数(单位：万人 以 5 年为一个时间序列且取整)：

表 4 世界人口预测数(单位：万人 以 5 年为一个时间序列且取整)

年份	2010	2015	2020	2025	2030
预测人口数	684951	719044	751629	782648	836884

从而预测出 2030 年全球人口数量为 836884.4952 万人。再利用 MATLAB 的绘图功能得到数据拟合后的世界人口图（上图 2），其相关程序见附录 6。

总之，通过图1和图2则可以基本看出世界人口的增长趋势是符合以线性关系的，拟合效果较好。这样也进一步得到了世界人口 $peop$ 随年份 t 增长的函数表达式为：

$$peop(t) = 768.7t - 14768000$$

虽然以上的两种预测方法所得到的结果是有误差的，但是这样的误差在大数据处理中是可以被忽略不计的，因此，我们就得出了约翰·贝丁顿的观点（气候变化和人口增长将导致食品、水和能源短缺，进而引发大规模移民、公共骚乱和国际冲突。如果未来

报名号 # 1088

数年内没有充分准备的话，到2030年，世界将面临“完美风暴”，出现大的动荡，并且全球人口将增加到大约83亿）是相对正确的。

5.2.3 线性回归模型的建立和检验

线性回归预测模型是一种最基本的回归模型，其数学表达式如下：

$$y = a + bx + \varepsilon \quad (5.3-1)$$

其中： y 表示因变量； x 表示自变量； a, b 表示未知常数； ε 表示随机误差且 $\varepsilon \sim N(0, \delta^2)$ ，

δ^2 未知； $y = a + bx$ 称为回归直线， b 称为回归系数，此时 $y \sim N(a + bx, \delta^2)$

用于估计参数 a, b 的最常用方法是最小二乘法。首先需要因变量 y 和自变量 x 的样本资料（必须是实际数值）：见下表 5

表 5 因变量 y 和自变量 x 的样本资料

自变量 x	x_1	x_2	x_3	x_i	x_n
因变量 y	y_1	y_2	y_3	y_i	y_n

这样就可以将上表 1 中是数据通过 MATLAB 软件在坐标系中描绘出来了（ x 为横轴， y 为纵轴），当这 n 对数据点在坐标系中近似呈现为直线分布时，就可以用线性回归模型的数学表达式（5.3-1）。

当然，（5.3-1）式子中的 $y = a + bx$ 应该作为预测的实际取值，而对于 (x, y) 的样本 $(x_1, y_1), \dots, (x_n, y_n)$ 有：

$$\begin{cases} y_i = a + bx_i + \varepsilon_i, i = 1, 2, \dots, n, \\ \varepsilon_i \sim N(0, \delta^2) \varepsilon_1, \varepsilon_2, \dots, \varepsilon_n \text{ 相互独立,} \end{cases}$$

从而对所有样本资料中的每一个 x_i 值都会有一个 y_i 的估计值 y_i' 相对应，并且 y_i 的估计值都与其实际值存在一个偏差 $\varepsilon_i (i=1, 2, 3, \dots, n)$ ，这样就有：

$$\varepsilon_i = y_i - y_i' = y_i - a - bx_i$$

设参数 a, b 的函数 Q 为：

$$Q = Q(a, b) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - a - bx_i)^2$$

为了求得 Q 最小，则可以利用极值原理：

$$\begin{cases} \frac{\partial Q(a, b)}{\partial a} = \sum_{i=1}^n (y_i - a - bx_i)(-2) = 2 \sum_{i=1}^n (a + bx_i - y_i) = 0 \\ \frac{\partial Q(a, b)}{\partial b} = \sum_{i=1}^n (y_i - a - bx_i)(-2x_i) = 2 \sum_{i=1}^n x_i (a + bx_i - y_i) = 0 \end{cases}$$

经过整理可以得到：

$$\begin{cases} na + \left(\sum_{i=1}^n x_i \right) b = \sum_{i=1}^n y_i \\ \left(\sum_{i=1}^n x_i \right) a + \left(\sum_{i=1}^n x_i^2 \right) b = \sum_{i=1}^n x_i y_i \end{cases} \quad (5.3-2)$$

式（5.3-2）称为正规方程组。

报名号 # 1088

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

现令

于是，求解正规方程组式 (5.3-2) 可得：

$$\begin{cases} a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \\ b = \frac{\sum_{i=1}^n y_i}{n} - \frac{b \sum_{i=1}^n x_i}{n} \end{cases} \quad \text{或者} \quad \begin{cases} \tilde{a} = \bar{y} - b \bar{x} \\ \tilde{b} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \end{cases} \quad (5.3-3)$$

这种用最小二乘法求出的估计 \tilde{a} 、 \tilde{b} 分别称为 a, b 的最小二乘估计，此时，拟合得到的直线为：

$$\tilde{y} = \tilde{a} + \tilde{b}x = \bar{y} + \tilde{b}(x - \bar{x}) \quad (5.3-4)$$

下面用矩法求 δ^2 的估计，根据数学期望和方差的性质可得： $\delta^2 = D\varepsilon = E\varepsilon^2$ ，由矩估计法， $E\varepsilon^2$ 可用 $\frac{1}{n} \sum_{i=1}^n \varepsilon_i^2$ 进行估计，而 $\varepsilon_i = y_i - a - bx_i$ ， a, b 分别由 \tilde{a} 、 \tilde{b} 代入，故 δ^2 可用下式作估计：

$$\tilde{\delta}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{a} - \tilde{b}x_i)^2$$

对于估计量 \tilde{a} 、 \tilde{b} 、 $\tilde{\delta}^2$ 的分布，存在着如下四个的定理关系：

$$\tilde{a} \sim N \left(a, \frac{\delta^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} \right); \quad \tilde{b} \sim N \left(b, \frac{\delta^2}{n \sum_{i=1}^n (x_i - \bar{x})^2} \right); \quad \frac{n\tilde{\delta}^2}{\delta^2} \sim \chi^2(n-2); \quad \tilde{\delta}^2 \text{ 分别与 } \tilde{a}、\tilde{b} \text{ 独立。}$$

下面就进行检验：在回归模型建立以后，它与实际资料拟合度和模型的线性关系显著性是否显著及有效性是否高，就必须进行数理统计和经济意义上的检验。常规的统计检验方法如下：

(1) 回归模型预测精度的检验——标准差 s 的检验

标准差 s 的表达式为：

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \tilde{y}_i)^2}$$

$$\frac{s}{\tilde{y}} < 0.1$$

显然 s 的值应该是越小越好，一般的要求为：

(2) 回归模型的线性相关性的显著程度检验——相关系数检验

取检验统计量 R ，

报名号 # 1088

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5.3-5)$$

称R为样本相关系数。

在数学上证得： $|R| \leq 1$ 且有①当 $|R|=1$ 时， y_i 完全落在了回归直线上，此时y与x是完全线性相关的；②当 $0 < R < 1$ 时，y与x一定有正线性相关性，此时越接近1越好；③当 $-1 < R < 0$ 时，y与x一定有负线性相关性，此时越接近-1越好。

现按照式子(5.3-5)计算出相关系数R的值；再拟定显著性水平 α （ α 可取0.1, 0.05, 这里我们取 $\alpha=0.05$ ，即置信度为95%），在取自由度为n-1查表得相关系数的临界值 R_α ；若当 $|R| \geq R_\alpha$ 时，y与x在显著性水平 α 下显著相关，此时通过检验，否则y与x线性关系不明显，未通过检验。

(3)x与y之间是否存在显著的线性统计关系——F检验

取检验统计量F：

$$F = \frac{(n-2) \sum_{i=1}^n (\tilde{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \tilde{y}_i)^2} = \frac{(n-2)R^2}{(1-R^2)}$$

先按上式计算出F值；再拟定显著性水平 α （ α 可取0.1, 0.05, 这里我们取 $\alpha=0.05$ ，即置信度为95%），取自由度为n-2，查表得出F的临界值 F_α ；若 $F \geq F_\alpha$ 则x与y间在 α 显著水平下存在线性的统计关系，通过检验，也就可以说明所建立的模型是有效的，反之则未通过检验，模型无效。

5.2.4 全球粮食预测模型的分析、建立和求解

粮食在整个世界占有非常重要的地位，它作为一种生存必备物，既要保障粮食产量能够维持生存，又得尽可能的多的去生产粮食，这样就需要在符合现实情况下保障粮食供应充足时进行一些储存。

5.2.4.1 全球粮食需求量预测模型的建立和求解

为了能预测出到2030年世界粮食需求量的情况，我们根据搜集的数据（详见附录2），MATLAB的强大的绘图功能得到世界粮食的需求量年图。用根据数据所呈现出来的表面特征建立了线性回归模型来预测2030年淡水需求量情况。

首先,通过搜集到的 1987-2006 年世界粮食需求量（单位：百万吨）进行初步数据分析和处理,利用 MATLAB 的强大的绘图功能得到下图 3 世界粮食需求量图

报名号 # 1088

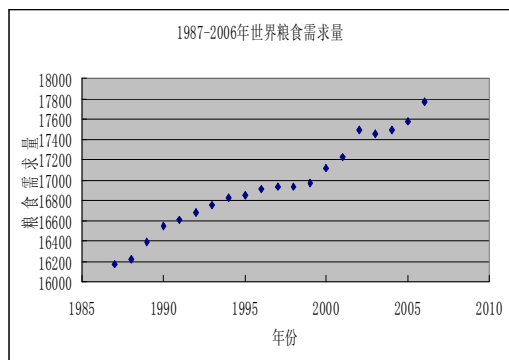


图3 世界粮食需求量图

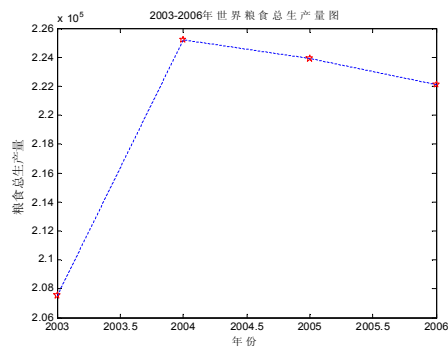


图4 世界粮食总产量图

由上图可见世界粮食需求量几乎呈现线性关系，但是不是很明显。于是我们就粗略的采用前面所建立的线性回归模型来预测 2030 年的粮食需求量。

再通过 MATLAB 拟合得出了全球粮食需求量 $food_1$ 与年份 t 之间的函数关系：

$$food_1(t) = 348338.3t - 676070263.4444$$

于是当 $t=2030$ 时，计算可得： $food_1(2030) = 310565.5322$ (百万吨)，即到 2030 年时全球的粮食总需求量为 310565.5322 百万吨。

所以到 2030 年粮食需求量的增长率 fg 为：

$$fg = \frac{food_1(t) - food_1(2003)}{food_1(2003)} \% = \frac{310565.5322 - 169461.1822}{169461.1822} \% = 45.43\%$$

5.2.4.2 全球粮食总生产量预测模型的建立和求解

经过数据的搜索，我们得到近几年来全球粮食总生产量（由于主要考虑粮食对于生存的贡献，指食用粮食总产量，下表 6），绘图得见上图 4：

表 6 世界粮食总生产量(单位：百万吨)

年份	2003	2004	2005	2006
粮食总生产量	207530.9	225201.7	223940	222112

我们通过一次 MATLAB 拟合得出了全球粮食总产量 $food_2$ 与年份 t 之间的函数关系：

$$food_2(t) = 4200t - 8.2957 \times 10^6$$

于是当 $t=2030$ 时，计算可得： $food_2(2030) = 203000$ (百万吨)，即到 2030 年时全球的粮食总产量为 203000 百万吨。

这样通过比较可得：到 2030 年时，粮食的需求量为 310565.5322 百万吨，而粮食总产量为 203000 百万吨，相比之下生产量与需求量之差为：107565.5322 百万吨，这样到 2030 年粮食的缺口比例为：34.63%。

5.2.5 淡水资源需求增长比例预测模型的分析、建立和求解

为了方便以后的模型建立，我们首先对所搜集到的数据中选取 1994-2006 年的数据（详见附录 3）进行了图示化处理，即用 MATLAB 软件进行了绘图，得到了下图 5。

报名号 # 1088

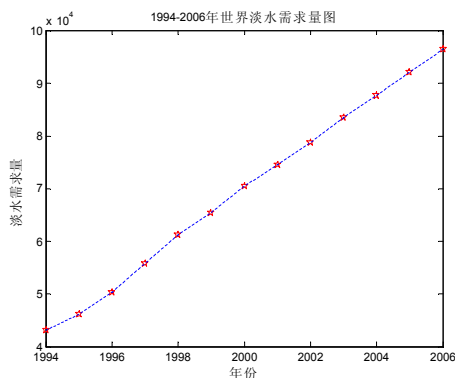


图 5 世界淡水需求量

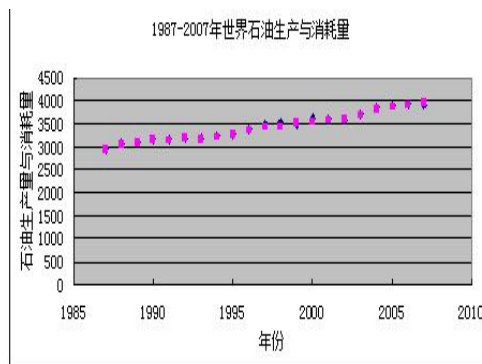


图 6 世界石油生产和需求（消耗）量

从中发现具有很大的线性相关性，于是我们用前面所建的线性回归模型进行求解。

首先算出， $\bar{x} = 2000, \bar{y} = 69585.846, b = 4547.368, a = -9125150.418$,

于是就得到了：线性回归模型的数学表达式为：

$$y(x) = a + bx = -9125150.418 + 4547.368x$$

$$\frac{s}{\bar{y}} = 0.0087 < 0.1$$

下面对其进行检验：（1）标准差 s 的检验：通过计算可得 $\frac{s}{\bar{y}}$ ，显然通过检验；（2）相关系数检验：计算得 $R=0.691$ ，即回归模型的线性相关性的显著程度较显著；（3）F 检验：计算得 $F=22.846 > F_{\alpha}=19.41$ ，即 x 与 y 间在 α 显著水平下存在线性的统计关系。也就是说通过检验，所建立的模型是有效的。

所以，当 $x=2030$ 时， $y(2030)=-9125150.418+4547.368 \times 2030=106007$ （亿立方米）或者 $wate(2030)=106007$ （亿立方米），即到 2030 年时淡水的需求量为 106007 亿立方米。

于是有 1994 年到 2006 年的淡水需求量平均值 $\bar{y}=6958.846$ （亿立方米），

所以到 2030 年淡水需求量的增长率 wg 为：

$$wg = \frac{y(x) - \bar{y}}{\bar{y}} \% = \frac{106007 - 69585.846}{69585.846} \% = 34.36\%$$

5.2.6 能源需求预测模型的分析、建立和求解

当今世界，各国家都在一定程度上存在着能源缺乏的问题，部分国家尤其严重。能源也此存在着危机——能源危机，它随时都有可能发生。现在我们要研究的问题是到未来 2030 年对能源的需求量问题，为了建立模型方便，我们选择当前世界具有能源代表性的石油来作为研究对象进行研究。同时为了更好的预测出未来能源的需求状况，我们从全世界每年对石油的需求量情况和每年对石油的生产量情况两个方面考虑入手进行建模。

首先通过搜集的 1987-2007 年世界石油生产与消耗量（单位：百万吨）数据（详见附录 4）进行了图示化处理，即用 MATLAB 软件绘图得到上图 6

从上图 6 可以看出，近 20 年来石油的生产量和消耗量基本保持着平稳的增长，由此就可以利用线性拟合来预测未来若干年全世界对石油及对能源的需求和生产情况。

通过上面的模型准备，则可设石油需求量函数表达式 $od(t)=a+bt$ ，其中 a, b 为常量。这样利用前面所建立的线性回归模型的求解方法类似的进行求解，得到 $a=-93460, b=48.5159$ 。对该模型进行检验时可以发现线性相关系数的检验 $R=0.9567$ 很接近于 1，即线性相关性的显著程度明显。这样对未来若干年的石油需求量就可通过下面的方程进行预测：

报名号 # 1088

$$od(t) = -93460 + 48.5159t.$$

于是当 $t=2030$ 时, $od(2030) = -93460 + 48.5159 \times 2030 = 5027.46$ (百万吨), 即 2030 年时全球的石油需求量为 5027.46 百万吨。

同理, 假设石油生产量函数表达式 $op(t) = c + dt$, 其中 c, d 为常量。同样求得 $c = -92304.3, d = 47.9457$ 。对该模型进行检验时可以发现线性相关系数的检验 $R = 0.9589$ 很接近于 1, 即线性相关性的显著程度明显。这样对未来若干年的石油生产量就可通过下面的方程进行预测: $op(t) = c + dt = -92304.3 + 47.9457t$;

当 $t=2030$ 时, $op(2030) = -92304.3 + 47.9457 \times 2030 = 5025.523$ (百万吨), 即 2030 年时全球的石油生产量为 5025.523 百万吨。

再求得近 20 年的石油需求量平均值 $\overline{od} = 3426.433$ 百万吨。而 2030 年的石油需求量或者能源需求量为 5027.46 百万吨。于是到 2030 年石油需求量的增长比例 oc 为:

$$oc = \frac{od(2030) - \overline{od}}{od(2030)} \% = \frac{5027.46 - 3426.433}{5027.46} \% = 31.85\%$$

5.2.7 结果论证

通过以上的各个模型我们得到了到 2030 年时世界人口、能源、粮食、淡水需求量的预测值, 并且我们将以搜集到所有年份的数据平均值作为人类生活环境的现状, 与 2030 年时的所有预测结果进行比较 (见下表 7)

表 7 人类生活现状与预测值比较

年份	2030
世界人口 (万人)	583212
粮食需求量 (百万吨)	836884
淡水需求量 (亿立方米)	310565.5322
能源需求量 (石油, 百万吨)	106007
	5025.523

这样就可得到①世界人口达到 83.6884 亿, 已超过 83 亿; ②粮食需求增长率为 45.43%, 接近 50%; ③淡水需求量增长率为 34.36%, 超过 30%; ④能源需求量增长率为 31.85%, 似乎小于 50%, 但是由于数据本身不够精确和模型的一些粗糙性完全能造成这样的差异。

同时, 我们采用预测出 2030 年的粮食需求量、淡水资源需求量、能源需求量与人口数量的比值之和来进一步预测出到 2030 年时发生“完美风暴”的可能性 gl ,

$$gl = \left(\frac{food(2030)}{3peop(2030)} + \frac{wate(2030)}{3peop(2030)} + \frac{od(2030)}{3peop(2030)} \right) \% = 42.3485\%$$

综上所述: 在建立“完美风暴”危机预测模型的同时初步论证了英国科学家约翰·贝丁顿的“完美风暴”理论观点的正确性, 且到 2030 年时发生“完美风暴”的可能性为 42.3485%

5.3 有效协调降低“完美风暴”发生可能性模型

5.3.1 模型的准备

根据评价中国在“完美风暴”中地位和作用的模糊数学综合评判模型结果和引发“完美风暴”危机预测模型中建立的主成分分析方法, 对网站中得到的近年世界各国国家作

报名号 # 1088

为评价对象，通过主成分分析进行四个指标（即人口（万人）、能源量（百万吨）、淡水资源量（亿立方米）、粮食量（万吨））在加上一个国土面积（万平方公里）指标分析后得到近年世界 20 个国家的评价指标的情况表 8：

表 8 近年世界 20 个国家的评价指标情况

国家	中国	印度	伊朗	日本	韩国	泰国	土耳其
面积	960	328.726	164.82	37.791	9.926	51.312	78.356
人口	131179.8	110981.1	6915.254	12756.49	4841.808	6472.442	7293.532
能源	36320000	8608800	22237.9	63117.6	22153.6	11544.1	10415.2
淡水	28282.365	12785.023	898.9830	4292.559	650.739	2115.841	2297.463
粮食	44535.516	23913	2280.992	1174.148	665.309	3314.629	3459.828
国家	埃及	南非	加拿大	墨西哥	美国	阿根廷	巴西
面积	100.145	121.909	998.467	195.82	963.203	278.04	851.488
人口	7539.691	4739.09	3255.646	10422.14	29898.81	3912.046	18869.44
能源	8252.7	17093.9	38952.5	20427.4	293130.6	8561.5	21191.1
淡水	376.985	452.583	28727.169	4134.463	28242.42	2786.55	54845.914
粮食	2273.643	945.325	5089.45	3195.916	34656.195	3355.633	5901.747
国家	法国	德国	意大利	波兰	俄罗斯	西班牙	
面积	55.15	35.703	30.134	31.269	1709.824	50.537	
人口	6103.751	8241.144	5857.1	3811.241	14236.76	4354.63	
能源	24889.5	42108	25701.6	12708.7	86974.3	16849.5	
淡水	1789.619	1068.876	1823.901	535.098	42902.476	1115.656	
粮食	6181.269	4347.479	2014.543	2177.593	7686.605	1935.315	

再经过所建立的主成分分析模型的步骤，一一求得指标的相关系数矩阵、相关矩阵 R 的特征值与特征向量，确定主成份、方差贡献率以确定主成份个数、主成份进行综合评价、主成分的得分计算等后得到上面 20 个国家的得分和排序情况，表 9，：

表 9 近年世界 20 个国家的评价体系得分及排名

国家	人口得分	能源得分	淡水得分	粮食得分	综合得分	得分排名
中国	1.3201	-0.0582	-0.2962	0.0744	1.04	1
印度	0.6086	0.0307	-0.0587	0.012	0.5926	2
伊朗	0.0579	0.0046	0.019	0.0053	0.0868	8
日本	0.0666	-0.1304	0.0172	-0.0696	-0.1162	20
韩国	0.0277	-0.0289	0.0064	-0.0211	-0.0159	14
泰国	0.0625	0.0175	0.0189	0.001	0.0998	6
土耳其	0.0581	-0.0029	0.0141	-0.0006	0.0687	9
埃及	0.0967	0.0662	0.0436	-0.0025	0.2041	4
南非	0.0273	-0.0151	0.0066	0.0021	0.0209	13
加拿大	0.0495	-0.2433	0.0157	0.2046	0.0282	12
墨西哥	0.0708	-0.0238	0.0186	0.0023	0.0679	10
美国	0.4008	-0.4233	0.102	-0.1206	-0.0412	15
阿根廷	0.0596	-0.0148	0.0226	0.0228	0.0901	7
巴西	0.1281	-0.0988	0.0378	0.096	0.163	5
法国	0.0523	-0.112	0.0011	-0.0453	-0.1039	18
德国	0.0567	-0.113	0.0068	-0.0573	-0.1068	19
意大利	0.0341	-0.0808	0.0038	-0.0375	-0.0804	17
波兰	0.0357	-0.0059	0.009	-0.0075	0.0313	11
俄罗斯	0.1697	-0.1644	0.0705	0.1534	0.2293	3
西班牙	0.0288	-0.0627	0.0033	-0.027	-0.0576	16

得到上表后，可看出上面的模型是较准确合理的，而且在以中国为代表的许多发展

报名号 # 1088

中国是很容易引起“完美风暴”的，从而也初步的反映了能通过中国自身或几个国家的协调来有效降低“完美风暴”发生的可能性。毕竟当今的发展中国家还存在着制度不够健全、社会保障不够完善，经济条件够好，且人口较多等状况，是较符合现状的。

有了这样的认识，仔细观察表9会发现其结果是比较令人满意的，这也为后面进行协调降低“完美风暴”发生可能性奠定了扎实的基础，尽可能的减小初始误差。

5.3.2 聚类分析模型的建立和求解

上面通过主成份分析法，初步筛选出了 20 个国家并对其进行了合乎常理的排名，但考虑到国家的数目仍然太多，并且一些国家在评判时相似度很接近，即对“完美风暴”的发生影响性很接近，我们吸取分层抽样的思想，再进一步通过聚类分析，找出几个比较典型国家代表这 20 个国家，进一步代表全球。聚类分析是研究分类问题的一种多元统计方法，类即指相似元素的集合。

聚类分析的基本思想：是根据一批样品的多个观测指标，具体地找出一些能够度量样品或指标之间相似程度的统计量，然后利用统计量将样品或指标进行归类。把相似的样品或指标归为一类，把不相似的归为其他类。直到把所有的样品（或指标）聚合完毕。它又有对样本和对变量的分类。

这里我们采用对样本分类（即 Q 型聚类分析）对样本进行聚类，它使具有相似性特征的样本聚集在一起，使差异性大的样本分离开来。

间隔尺度变量的相似性度量——距离

设有 n 个样本单元，每个样本测有 p 个指标（变量），原始资料阵为：

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (18)$$

并且每个样本都可以看成 p 维空间中的一点， n 个样本就是 p 维空间中的 n 个点；

第 i 个样本与第 j 个样本之间的距离记为 d_{ij} ；聚类过程中，相距较近的点归为一类，相距较远的点归为不同的类。

于是第 i 个和第 j 个样品之间的距离 d_{ij} 有如下四个条件：

(1) $d_{ij} \geq 0$ ， $d_{ij} = d_{ji}$ ， $d_{ij} \leq d_{ik} + d_{kj}$ 对一切的 i 和 j 都成立；

(2) $d_{ij} = 0$ 当且仅当 $i=j$ 时成立；

这里的距离可以用很多的方法来表示，为了方便运用，我们采用欧氏距离(Euclidean distance)，即当 $q=2$ 时：

$$d_{ij}(2) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (19)$$

则原始资料阵就可以用上表 9 中数据得到。

现用 $G1=\{\text{中国}\}$ 、 $G2=\{\text{印度}\}$ 、 $G3=\{\text{伊朗}\}$ 、 $G4=\{\text{日本}\}$ 、 $G5=\{\text{韩国}\}$ 、 $G6=\{\text{泰国}\}$ 、 $G7=\{\text{土耳其}\}$ 、 $G8=\{\text{埃及}\}$ 、 $G9=\{\text{南非}\}$ 、 $G10=\{\text{加拿大}\}$ 、 $G11=\{\text{墨西哥}\}$ 、 $G12=\{\text{美国}\}$ 、 $G13=\{\text{阿根廷}\}$ 、 $G14=\{\text{巴西}\}$ 、 $G15=\{\text{法国}\}$ 、 $G16=\{\text{德国}\}$ 、 $G17=\{\text{意大利}\}$ 、 $G18=\{\text{波兰}\}$ 、 $G19=\{\text{俄罗斯}\}$ 、 $G20=\{\text{西班牙}\}$ ；

采用欧式距离方法，得结果为：墨西哥、加拿大、阿根廷、法国为一类，巴西、西

报名号 # 1088

班牙、日本、波兰为一类，埃及、南非、土耳其为一类，德国、韩国、美国为一类，伊朗、意大利为一类，俄罗斯、中国、印度单独成一类；即共分为 8 类。如下图 7 所示：

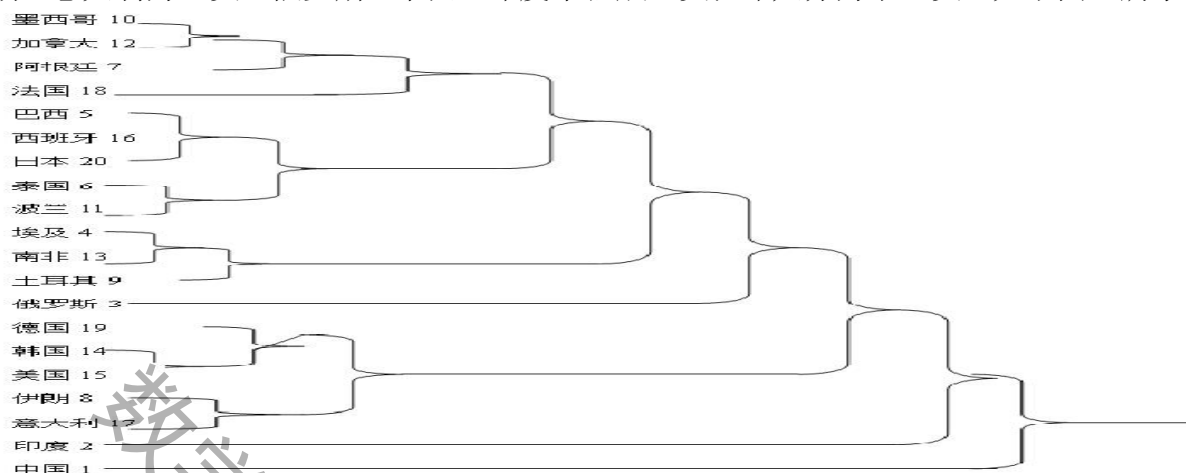


图 7 聚类分析结果图

最终根据以上结果再结合分层抽样的原则从 20 个具有代表性的国家里选出了以下 8 个最具有代表性的国家来作为通过协调能有效降低“完美风暴”发生的国家，并计算其可能性，分别为：中国（评价体系得分排名 1）、印度（评价体系得分排名 2）、俄罗斯（评价体系得分排名 3）、巴西（评价体系得分排名 5）、阿根廷（评价体系得分排名 7）、南非（评价体系得分排名 13）、美国（评价体系得分排名 15）、日本（评价体系得分排名 20）。

下面再运用在评价中国在“完美风暴”中地位和作用的模糊数学综合评判模型中的熵值法来求取 8 个主要国家在引起“完美风暴”发生的权系数如下表 10：

表 10 8 个主要国家在引起“完美风暴”发生的权系数

国家	中国	印度	俄罗斯	巴西	阿根廷	南非	美国	日本
熵值法权系数	0.0759	0.0102	0.1036	0.0936	0.1247	0.3611	0.1246	0.1060

很显然，上表中的各个权系数之和为：

$$0.0759+0.0102+0.1036+0.0936+0.1247+0.3611+0.1246+0.1060=0.998 \approx 1$$

是满足要求的，同时我们也可以看出的是在上表中，中国所占的权系数是相当小的，这也反面的验证了我们在选择协调能有效降低“完美风暴”发生的国家时是选择的多个国家而不只是中国一个国家，虽然只选择中国一个国家也有可能在一定程度上降低“完美风暴”发生的可能性，但是可以肯定的是没有多个国家联合协调的效果好。

由于在评价中国在“完美风暴”中地位和作用的模糊数学综合评判模型中已经算得了中国在“完美风暴”中的优属度，因此也就可以估计出其余 7 个国家的优属度，其中 6 个国家在引发“完美风暴”的权系数大于中国。也就可以估计出这 6 个国家在“完美风暴”中是优属度也大于中国。

这样我们就引进 8 个国家的权系数之和除以权系数大于中国的国家（含中国）的优属度（这里就取最小值 0.316571）之和，这样就可以得到 8 个国家进行协调后发生“完美风暴”的可能性 gl' ，其最大值为 39.3813%，即 $gl' \leq 39.3813\%$ 。

在与危机预测模型中得到的发生“完美风暴”可能性 $gl = 42.3485\%$ 相减，就得到了 2030 年前 8 个国家协调作用后能有效降低“完美风暴”发生可能性大小为：

$$gl - gl'_{\max} = 42.3485\% - 39.3813\% = 2.9672\%$$

即其可能性至少能够降低 2.9672%。

报名号 # 1088

6 模型改进与推广

模型比较全面的考虑了引发“完美风暴”的各个因素，并分别对其进行了分析和数据挖掘搜索，建立出多个子模型进行预测和分析危机，而后就综合评价考虑了其各个影响因子间的关系，进一步的扩展了模型的实用价值并判断出“完美风暴”发生的可能性和能够降低其发生的可能性。

在全球人口预测模型中，我们直接运用第 t 年 i 岁（满 i 周岁而不满 $i+1$ 周岁）的人数减去下一年的人数之差除以第 t 年 i 岁（满 i 周岁而不满 $i+1$ 周岁）的人数来表示死亡率，这显然是模糊粗糙的，这就可以运用 Lee-Carter 预测模型对人口的死亡率进行预测。模型如下：

$$\ln[m_i(t)] = a_i + b_i k(t) + \varepsilon_i(t)$$

式中， $m_i(t)$ 是 t 年 i 岁人的死亡率； $k(t)$ 是时间因子，作为预测死亡率整体水平上的指标； a_i 是年龄别的死亡率取自然对数以后的平均值，描述了死亡率年龄模式的时间趋势； b_i 是年龄因子，用来描述当时间因子变化时，各个年龄组死亡率相对变化的系数； $\varepsilon_i(t)$ 表示 t 年 i 岁的人的死亡率的残差。

其次，根据世界的实际情况，人口率很大一部分是按照各国政府的措施变化的。我们改换思路从各国政府的角度考虑未来人口的走势。作为政府对待人口增长的态度主要有三种：1. 人口净增长率过高，为了限制自然增长率而实行类似计划生育的政策，如中国。2. 人口净增长率过低，为了增加自然增长率而实行鼓励生育的政策，如一些发达国家。3. 人口净增长率约为0，政府对人口增长处于一种监察状态，并未采取措施控制，如部分拉美国家。

显然，三种状态政府的投入由高到低，而人民满意度却是由低到高。我们把第三种状态定为政府期望状态，即政府希望通过调整政策并采取一定的措施将人口逐步控制在某一范围内。以使人口净增长率能够在没有政府干预的前提下基本稳定在0左右。以减少政府的出面干预。该模型的建立如下：

从人口生态学的角度出发，随着人口数量 x 的不断增长，人口数量日益接近环境条件所允许的饱和水平，因此，抑制人口继续增长的作用也在加强，人口增长率也逐渐变小，并最终趋于零，当人口具有零增长率的时候，人口数量就达到了该国家地区所能承受的最大

人口数量 (k)，设人口出生率 $W(x)$ ，死亡率 $U(x)$ 和人均增长率 $\frac{1}{x} \frac{dx}{dt}$ 分别为：

$$\begin{cases} W(x) = a_1 - b_1 x \\ U(x) = a_2 - b_2 x \\ \frac{1}{x} \frac{dx}{dt} = W(x) - U(x) = a_1 - b_1 x - a_2 + b_2 x \end{cases}$$

则人口演化过程可以用逻辑斯蒂 Logistic 方程来表达，即：

$$\begin{cases} \frac{dx}{dt} = rx \left(l - \frac{x}{k} \right) \\ x|_{t=0} = x_0 \end{cases}$$

报名号 # 1088

其中， $r = a_1 - a_2, k = \frac{a_1 - a_2}{b_1 + b_2}$ 。方程中的非线性项 $-\frac{rx^2}{k}$ 称为抑制项，对人口增长起到抑制的作用。解得：

$$x = \frac{k}{1 + \left(\frac{k}{x_0} - 1\right)e^n}$$

由于时间和数据不足的原因，我们就没有对模型进行求解。但是，可以看到的是模型是从政府的角度建立的，因此对于中长期总体人口预测有很大的借鉴意义。

7 模型评价

本文的优点是（1）模糊综合评判—综合评判就是对受到多个因素制约的事物或对象作出一个总的评价，由于从多方面对事物进行评价难免带有模糊性和主观性，采用模糊数学的方法进行综合评判将使结果尽量客观从而取得更好的实际效果；（2）分别建立了引发“完美风暴”的各个因子问题的模型，这样可以尽可能的符合全球各个地区的实际情况，利于我们从整体上来把握客观规律。同时还运用主成分分析的方法综合评价考虑了人口、粮食、淡水、能源等因素对“完美风暴”的影响性高低，也就进一步的扩展了模型的实用价值。

本文的不足之处有：（1）由于部分搜索历史数据来源有限，本文所建立的模型在进行数据的运用的过程中的精度十分不足；（2）建立各个因子之间的预测模型时可以运用的方法很多，各种方法所得出的结果也不相同，因此模型有许多可以改进的地方；（3）在运用主成分分析方法的综合考虑各个因子之间的关系时，我们对全球的实际情况和国际政策了解得不足，从而就考虑问题是上显得略有不足，因此也使得模型具有许多的局限性。（4）由于在前面所列模型及其改进中存在着由于累积误差而使各因素的预测可能出现较大偏离的可能。

8 参考文献

- [1]、姜启源，谢金星等，数学模型，北京：高等教育出版社，2003
- [2]、韩中庚，数学建模方法及其应用，北京：高等教育出版社，2005
- [3]、赫孝良，戴永红，周义仓. 数学建模竞赛 [M]. 西安：西安交通大学出版社，2003
- [4]、宋来宗，王志明，数学建模与实验，北京：科学出版社，2005
- [5]、刘卫国，MATLAB 程序设计与应用（第二版），北京：高等教育出版社，2006年7月
- [6]、吴传生，经济数学—概率论与数理统计，北京：高等教育出版社，2008年12月
- [7]、朱道元，多元统计分析与软件 SAS[M]，南京：东南大学出版社，1999
- [8]、DUNHAM M H. 数据挖掘教程[M]，郭崇慧，田凤占，靳晓明等译，北京：清华大学出版社，2005
- [9]、孙蕾、赵兴、李云辉，丁克现象与中国人口增长，工程数学学报，第24卷，2007年12月
- [10]、郑福寿、沈长松、林益才，主成分回归分析法在混凝土坝初次蓄水期应力统计模型中的应用，大坝与安全，第一期：38-40，2001

报名号 # 1088

- [11]、陈守煜, 系统模糊决策理论与应用[M], 大连: 大连理工大学出版社, 1994
- [12]、李洪兴、汪培庄, 模糊数学[M], 北京: 国防工业出版社, 1994
- [13]、李子奈, 计量经济学, 北京: 高等教育出版社, 2003 年 4 月
- [14]、刘增洁, 未来 20 年世界能源供需预测, 中国能源-研究与探讨, 2002 年第 5 期: 33-35, 2002
- [15]、国际统计年鉴 2009, 中国统计出版社, 2009
- [16]、世界各国年中人口统计表, <http://www.bojianbook.com/news.asp?id=101>, 2009 年 5 月 29 日
- [17]、世界淡水概况, <http://old.blog.edu.cn/user2/tiankongyizu/index.shtml>, 2009 年 5 月 29 日
- [18]、联合国粮农组织 FAO, 世界粮食产量统计数据, <http://www.faoatas.org>, 2009 年 5 月 29 日

报名号 # 1088

附录

附录 1 1987-2007 年世界人口统计 (单位: 万人)

附表 1 全球人口的相关数据 (单位: 万人)

年份	1987	1988	1989	1990	1991	1992	1993
全球总人口	502400	511200	520100	528500	538500	548000	557200
年份	1994	1995	1996	1997	1998	1999	2000
全球总人口	562980	571606	574560	584590	589848	597773	607580.372
年份	2001	2002	2003	2004	2005	2006	2007
全球总人口	614806	622498	627700	634513	646158.3892	653816.7861	661204.0394

附录 2 1987~2006 年世界粮食需求量 (单位: 万吨)

附表 2 全球粮食需求量 (单位: 万吨)

年份	1987	1988	1989	1990	1991	1992	1993
粮食需求量	16171	16222	16388	16542	16608	16683	16754
年份	1994	1995	1996	1997	1998	1999	2000
粮食需求量	16824	16850	16910	16928	16930	16973	17118
年份	2001	2002	2003	2004	2005	2006	2007
粮食需求量	17230	17492	17460	17495	17573	17770	

附录 3 1975~2006 年世界淡水需求量 (单位: 亿立方米)

附表 3 全球淡水需求量 (单位: 亿立方米)

年份	1994	1995	1996	1997	1998	1999	2000
需求量	43021	46012	50234	55743	61081	65279	70523
年份	2001	2002	2003	2004	2005	2006	
需求量	74533	78674	83423	87654	91987	96452	

附录 4 1987-2007 年全球石油生产与需求量 (单位: 百万吨)

附表 4 全球石油生产与需求量 (单位: 百万吨)

年份	1987	1988	1989	1990	1991	1992	1993
生产量	2949.3	3068.9	3102.7	3170.4	3160.5	3190.1	3188.9
需求量	2949.3	3041.8	3094.3	3154.9	3149.1	3186.3	3162.9
年份	1994	1995	1996	1997	1998	1999	2000
生产量	3237.4	3281.3	3377.1	3479.9	3547.3	3481.1	3614.1
需求量	3218.6	3264.2	3346.6	3433.3	3449.3	3518.1	3558.7
年份	2001	2002	2003	2004	2005	2006	2007
生产量	3600.3	3575.3	3701.1	3866.7	3897	3914.3	3905.9
需求量	3576.2	3611.3	3681.8	3823.7	3871	3910.9	3952.8

附录 6 世界人口数量和数据拟合后的世界人口数量图的程序

```

x1=1987:1:2007;
y1=[502400,511200,520100,528500,538500,548500,557200,562980,571606,574560,584590,589848,597773,605412,613010,619973,627252,636969,644598,651776,661204];
x2=2010:5:2030;
y2=[684951,719044,751629,782648,836884];
plot(x1,y1,'rp',x1,y1,': ',x2,y2,'rd',x2,y2)
title('1987-2007年世界人口数量图和预测图');
xlabel('年份');
ylabel('人口数量');
text(1990,528500,'1990年世界人口数量');

```