

人口预测模型【gyj】

1. 适用范围

2. 常见类型

3.1 指数增长模型(Malthus模型)

3.1.1 模型假设与参数说明

3.1.2 推导过程

3.1.3 模型表达式

3.1.4 不同取值的图像

3.1.5 模型评价

3.2 阻滞增长模型(logostic模型)

3.2.1 模型假设与参数说明

3.2.2 推导过程

3.2.3 模型表达式

3.2.4 不同取值的图像

3.2.5 模型评价

3.2.6 代码

3.3 考虑年龄结构和生育模式的人口模型

3.3.1 人口发展方程

3.3.2 生育率和生育模式

3.3.3 人口指数

3.4 例题

4. 参考资料

人口预测模型【gyj】

1. 适用范围

用于预测未来某个时间里，某个地区的人口数量

2. 常见类型

- 指数增长模型(Malthus模型): 假设人口增长率为常数
- 阻滞增长模型(logostic模型): 假设人口增长率为与人口总数相关的先增后减的函数，考虑自然资源和环境对人口的阻滞作用
- 考虑年龄结构和生育模式的人口模型: 考虑人口的年龄结构，即各个年龄段的死亡率不相同

3.1 指数增长模型(Malthus模型)

3.1.1 模型假设与参数说明

- 人口数量的变化是封闭的，即人口数量的增加与减少只取决于人口中个体的生育和死亡，且每一个个体都具有相同的生育能力和死亡率。
- $x(t)$: t 时刻的人口数，可微且连续
- r : 人口增长率，常数
 - 增长率 = 出生率 - 死亡率

3.1.2 推导过程

由假设， t 时刻到 $t + \Delta t$ 时刻人口的增量为 $x(t + \Delta t) - x(t) = rx(t)\Delta t$ ，得：

$$\begin{cases} \frac{dx}{dt} = rx \\ x(0) = x_0 \end{cases}$$

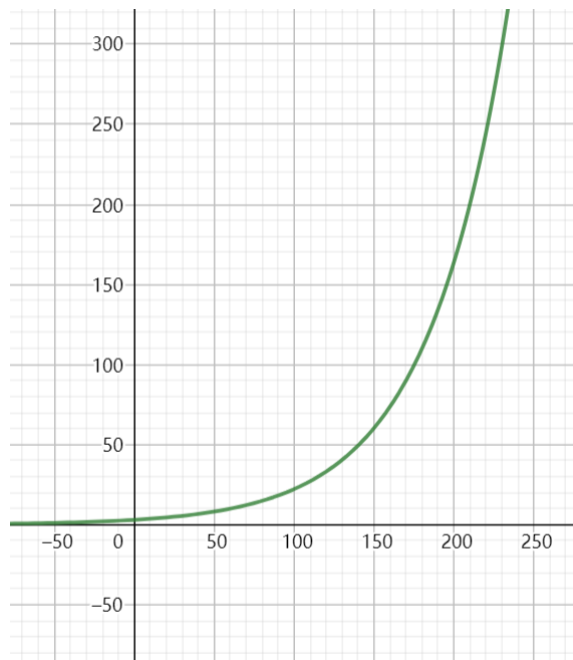
其解为：

$$x(t) = x_0 e^{rt}$$

3.1.3 模型表达式

$$x(t) = x_0 e^{rt} \quad (1)$$

3.1.4 不同取值的图像



3.1.5 模型评价

- 由图可见，随时间流逝，人口数量呈指数级增长。
- 考虑200多年来人口增长的实际情况，1961年世界人口总数为 $x_0 = 3.06 \times 10^9$ ，1961—1970年，每年平均的人口增长率为 $r = 2\%$ ，则(1)式就写成了 $x(t) = 3.06 \times 10^9 \cdot e^{0.02(t-1961)}$ 。利用这个式子就可以算出1961年之后任一年的世界人口总数。带入 $t = 2670$ ，算出 $x(t) = 4.4 \times 10^{15}$ ，即4400万亿，相当于地球上每平方米就要容纳至少20人，故这样的模型是不符合事实的。
- 用这一模型进行预测的结果远高于实际人口增长，误差的原因是对增长率 r 的估计过高。由此， r 不应该是一个常数。

3.2 阻滞增长模型(logistic模型)

3.2.1 模型假设与参数说明

- 考虑到地球资源对人口增长率的制约。随人口数量的增加，自然资源、环境条件等对人口再增长的限制作用将会越来越显著。这也就是环境等因素对人口的阻滞作用。
 - 在人口较少时，可以把增长率 r 看成常数；
 - 让人口达到一定数量时，就应当视 r 为一个随着人口的增加而减小的量，即增长率 r 为关于人口 x 的减函数
- 按照工程师的一般原则，我们会假设 $r(x)$ 为一线性减函数

$$r(x) = r - sx, \quad (r, s > 0)$$

- r ：固有增长率，即人口很少时看成常数的那个增长率
- $s = \frac{r}{x_m}$
 - x_m ：人口容量，即自然资源和环境所能容纳的最大人口量。当 $x = x_m$ ， $r = 0$ ，人口不再增长。

3.2.2 推导过程

写出 $r(x)$ 关于 x 的减函数

$$\frac{dx}{dt} = r(x)x \quad x(0) = x_0$$

将 $r(x) = r - \frac{rx}{x_m}$ 带入上式, 得

$$\frac{dx}{dt} = rx(1 - \frac{x}{x_m}) \quad x(t_0) = x_0$$

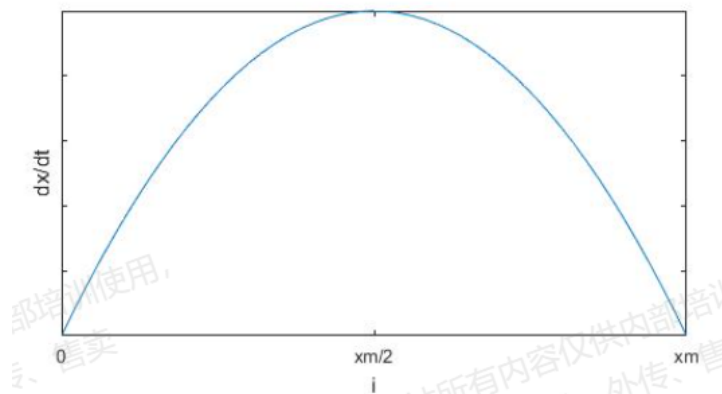
通过解常微分方程得到解析式。

3.2.3 模型表达式

$$x(t) = \frac{x_m}{1 + (\frac{x_m}{x_0} - 1)e^{-r(t-t_0)}} \quad (2)$$

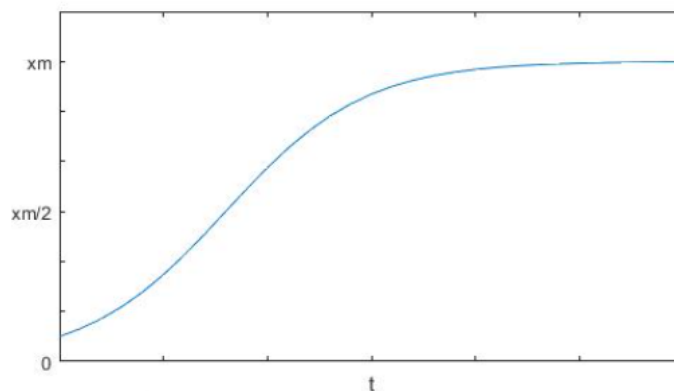
3.2.4 不同取值的图像

图一：阻滞增长模型 $\frac{dx}{dt} - x$, 体现了人口增长速度



增加速度先升高再递减, 在 $x = \frac{x_m}{2}$ 时达到最大增速

图二：阻滞增长模型 $x - t$ 曲线, 体现了人口增加的变化情况



总人口数先增加后稳定

3.2.5 模型评价

- 对公式(2)求二阶导, 可得

$$\frac{d^2x}{dt^2} = r^2(1 - \frac{x}{x_m})(1 - \frac{2x}{x_m})x$$

人口总数 $x(t)$ 有以下规律：

- $\lim_{t \rightarrow +\infty} x(t) = x_m$, 即无论人口初值 x_0 为何, 人口总数以 x_m 为极限
- 当 $0 < x_0 < x_m$ 时, $\frac{dx}{dt} = r(1 - \frac{x}{x_m})x > 0$, 这说明 $x(t)$ 是单调增加的, 又由上述二阶导公式可知, 当 $x < \frac{x_m}{2}$ 时, $\frac{d^2x}{dt^2} > 0$, $x = x(t)$ 为凹函数; 当 $x > \frac{x_m}{2}$ 时, $\frac{d^2x}{dt^2} < 0$, $x = x(t)$ 为凸函数
- 人口变化率 $\frac{dx}{dt}$ 在 $x = \frac{x_m}{2}$ 时取得最大值, 即人口总是达到极限值的一般以前是加速增长时期, 经过这一点后, 增长率会逐渐变小, 最终达到零。

3.2.6 代码

Matlab

```
%人口预测模型
%下面有某地区30年的人口数据, 给出该地区人口增长的logistic模型
clear;clc
y=[33815 33981 34004 34165 34212 34327 ...
   34344 34458 34498 34476 34483 34488 ...
   34513 34497 34511 34520 34507 34509 ...
   34521 34513 34515 34517 34519 34519 ...
   34521 34521 34523 34525 34525 34527];
% T=年份-起始年份
T=1:30;
%对数据作线性处理
for i=1:30,
    x(i)=exp(-i);
    Y(i)=1/y(i);
end

%计算回归系数b
c=zeros(30,1)+1;
X=[c,X'];
b=inv(X'*X)*X'*Y'

for i=1:30,
    %计算拟合值
    z(i)=b(1,1)+b(2,1)*x(i);
    %计算离差
    s(i)=Y(i)-sum(Y)/30;
    %计算误差
    w(i)=z(i)-Y(i);
end

%计算离差平方和
S=s*s';
%计算回归误差平方和
Q=w*w';
%计算回归平方和
U=S-Q;
%计算并输出F检验值
F=28*U/Q
%计算非线性回归模型的拟合值
for j=1:30,
    p(j)=1/(b(1,1)+b(2,1)*exp(-j));
end
%输出非线性回归模型的拟合曲线
plot(T,y,'*')
```

```
hold on
plot(T,p,'r-');
legend('实际情况','预测情况')
```

3.3 考虑年龄结构和生育模式的人口模型

以上两个模型都是针对人口总数和总的增长率，不涉及年龄结构。但事实上，在人口预测中人口按年龄的分布状况是十分重要的，因为不同年龄的人的出生率和死亡率有很大区别。在考虑年龄结构的人口模型中，变量有时间和年龄。

3.3.1 人口发展方程

- 假设与参数说明
 - 为研究任意时刻不同年龄的人口数量，引入人口分布函数和密度函数。
 - 人口分布函数($F(r, t)$)：时刻 t 年龄小于 r 的人口。 t, r 均是连续变量，设 F 是连续的、可微的。
 - $N(t)$ ：时刻 t 的人口总数
 - r_m ：最高年龄
 - $p(r, t)$ ：人口密度； $p(r, t) = \frac{\partial F}{\partial r}$
 - $\mu(r, t)$ ： t 时刻年龄为 r 的人的死亡率

- 推导过程

- 在进行理论推导时，先假设 $r_m \rightarrow \infty$ ，于是对非负非降函数，有

$$F(0, t) = 0 \quad F(r_m, t) = N(t)$$

- 考察在时刻 t ，年龄在 $[r, r + dr)$ 内的人到时刻 $t + dt$ 的情况。他们中活着的那一部分人年龄变为 $[r + dr_1, r + dr + dr_1)$ ，这里 $dr_1 = dt$ 。这段时间死亡的人数为 $\mu(r, t)p(r, t)drdt$ ，于是

$$p(r, t)dr - p(r + dr_1, t + dt)dr = \mu(r, t)p(r, t)drdt$$

$$\Rightarrow p(r + dr_1, t + dt) - p(r, t + dt)dr + [p(r, t + dt) - p(r, t)]dr = -\mu(r, t)p(r, t)drdt$$

- 注意到 $dr_1 = dt$ ，得到人口发展方程：

$$\frac{\partial p}{\partial r} + \frac{\partial p}{\partial t} = -\mu(r, t)p(r, t)$$

- 设初始密度函数为： $p(r, 0) = p_0(r)$ ；婴儿出生率，即单位时间出生的婴儿数为 $p(0, t) = f(t)$ 。
- 各个年龄的人口数为

$$F(r, t) = \int_0^r p(s, t)ds$$

在社会安定局面下和不太长的时间内，死亡率与时间无关，于是可以近似的假设 $\mu(r, t) = \mu(r)$ ，这时方程的解为

$$p(r, t) = p_0(r - t)e^{-\int_{r-t}^r \mu(s)ds} \quad (0 \leq t \leq r)$$

$$p(r, t) = f(r - t)e^{-\int_0^r \mu(s)ds} \quad (t > r)$$

3.3.2 生育率和生育模式

- 假设与参数说明

- 控制人口发展状况的其中一个手段就是控制婴儿出生率 $f(t)$ ，下面将对其进行进一步分解。
- $k(r, t)$ ：女性性别比例函数
- $p(r, t)$ ：人口密度； $p(r, t) = \frac{\partial F}{\partial r}$
- $b(r, t)$ ：女性在单位时间内平均每人的生育数量
 - 定义 $b(r, t)$ 为： $b(r, t) = \beta(t)h(r, t)$ ；其中 $h(r, t)$ 满足 $\int_{r_1}^{r_2} h(r, t)dr = 1$ ，是年龄为 r 女性的生育加权因子，成为**生育模式**； $\beta(t)$ 的直接含义是时刻 t 单位时间内平均每个育龄女性的生育数。如果所有育龄女性在她的育龄期所及的时刻都保持这个生育数，那么 $\beta(t)$ 也表示平均每个女性一生的总和生育数。
- $[r_1, r_2]$ ：育龄区间

- 推导过程

- $$f(t) = \int_{r_1}^{r_2} b(r, t)k(r, t)p(r, t)dr$$

- 得到单位时间出生的婴儿数 $f(t)$

$$\beta(t) = \int_{r_1}^{r_2} b(r, t)dr$$
$$f(t) = \beta(t) \int_{r_1}^{r_2} h(r, t)k(r, t)p(r, t)dr$$

- 模型评价

- 在稳定的环境下，可近似地认为女性的生育模式与时间无关，即 $h(r, t) = h(r)$ 。 $h(r)$ 表示了哪些年龄生育率高，哪些年龄生育率低。
- 人口发展方程和单位时间出生的婴儿数 $f(t)$ 的表达式构成了我们的连续型人口模型。模型中的死亡率、性别比例函数、初始密度函数可以由人口统计资料直接查到。生育率和生育模式则是可以用于控制人口发展的两种手段。

3.3.3 人口指数

以下是在人口统计学中，常用的一些人口指数来表示一个国家和地区的人口特征。

- 人口总数

$$N(t) = \int_0^{r_m} p(r, t)dr$$

- 平均年龄

$$R(t) = \frac{1}{N(t)} \int_0^{r_m} rp(r, t)dr$$

- 平均寿命

$$S(t) = \int_t^\infty e^{-\int_0^{\tau-t} \mu(r, t)dr} d\tau$$

- 老龄化指数

$$\omega(t) = \frac{R(t)}{S(t)}$$

3.4 例题

题目：利用下表给出的近两个世纪的美国人口统计数据(单位：百万)，建立人口预测模型，最后用它预报2010年美国人口。

年	1790	1800	1810	1820	1830	1840	1850	1860
人口	3.9	5.3	7.2	9.6	12.9	17.1	23.2	31.4
年	1870	1880	1890	1900	1910	1920	1930	1940
人口	38.6	50.2	62.9	76.0	92.0	106.5	123.2	131.7
年	1950	1960	1970	1980	1990	2000		
人口	150.7	179.3	204.0	226.5	251.4	281.4		

1. 建模与求解：

记 $x(t)$ 为第 t 年的人口数量，设人口年增长率 $r(x)$ 为 x 的线性函数， $r(x) = r - sx$ 。自然资源能容纳的最大人口数为 x_m ，即当 $x = x_m$ 时，增长率 $r(x_m) = 0$ ，可得 $r(x) = r(1 - \frac{x}{x_m})$ 。建立Logistic人口模型

$$\begin{cases} \frac{dx}{dt} = r(1 - \frac{x}{x_m})x \\ x(t_0) = x_0 \end{cases}$$

其解为

$$x(t) = \frac{x_m}{1 + (\frac{x_m}{x_0} - 1)e^{-r(t-t_0)}}$$

2. 参数估计：

利用matlab软件求得 $r = 0.0247$, $x_m = 373.5135$ ，2010年人口预测值为264.9119百万
matlab代码：

```
clc,clear,a=readmatrix('data.txt');
x=a([2:2:6],:); x=x(~isnan(x));
fn=@(r,xm,t)xm./(1+(xm/3.9-1)*exp(-r*(t-1790))); %定义匿名函数
ft=fittype(fn,'independent','t'); t=[1800:10:2000]';
[f,st]=fit(t,x(2:end),ft,'StartPoint',rand(1,2),...'Lower',[0,280],'Upper',[0.1,1000]) %由先验知识主管确定参数界限
xs=coeffvalues(f) %显示拟合的参数值
xh=f(2010) %求2010年的预测值

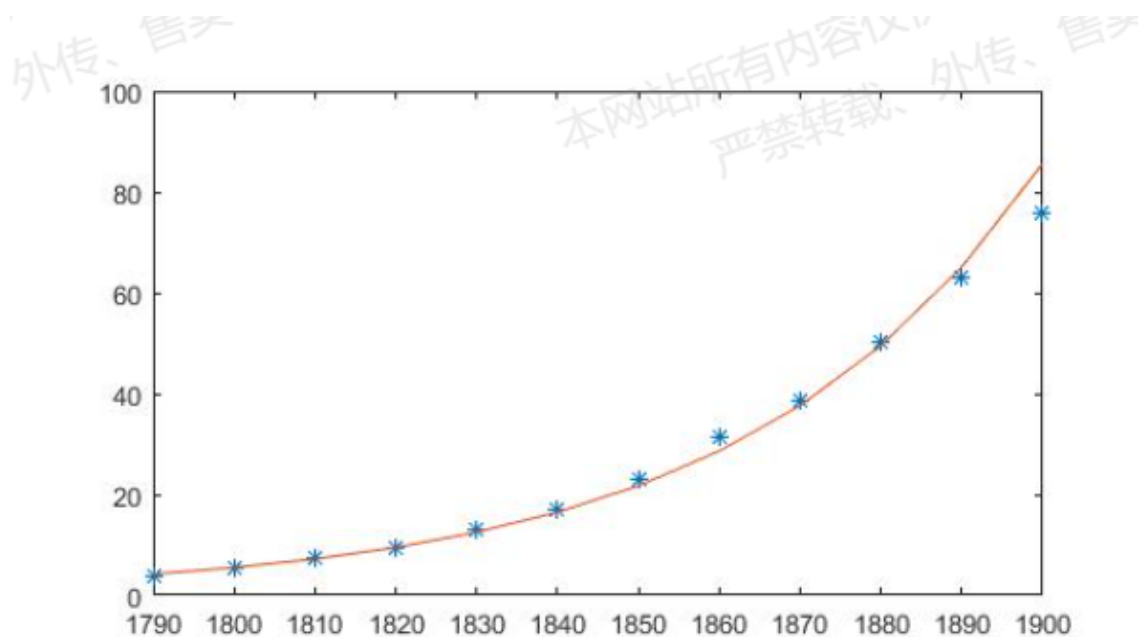
a=[ones(21,1),-x(1:end-1)]; %向前差分
b=diff(x),/x(1:end-1)/10;
cs=a\b; r=cs(1), xm=r/cs(2)
xh=fn(r,xm,2010) %求2010年的预测值

a=[ones(21,1),-x(2:end)]; %向后差分
b=diff(x),/x(2:end)/10;
cs=a\b; r=cs(1), xm=r/cs(2)
xh=fn(r,xm,2010) %求2010年的预测值
```

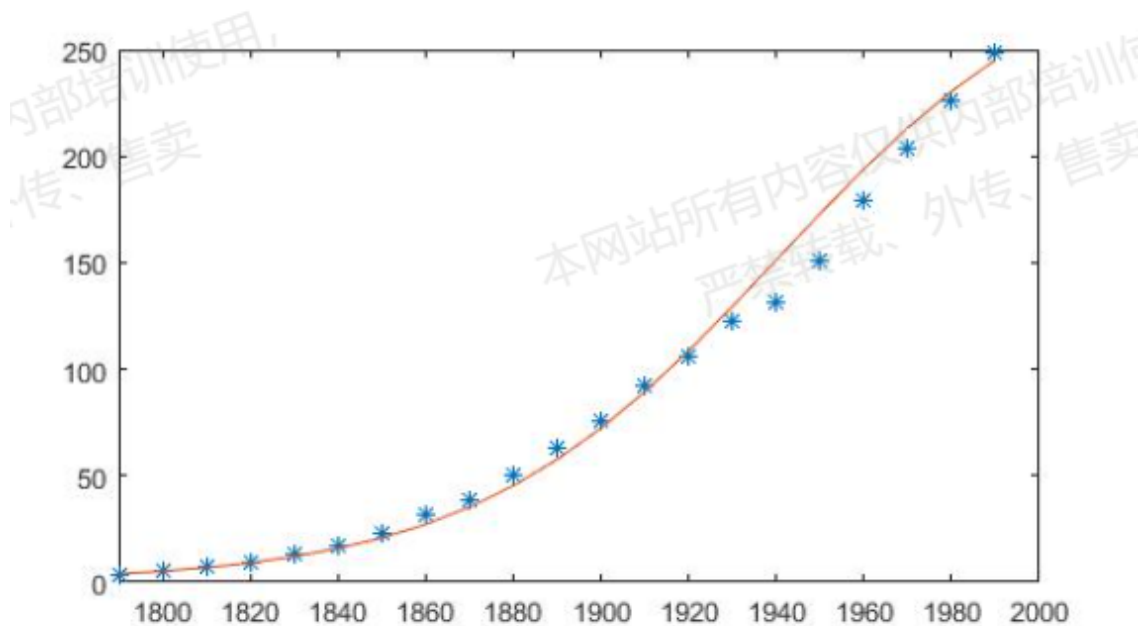
3. 参数的检验和预报：

分别对指数增长模型和阻滞增长模型拟合曲线，得到如下两张图

指数增长模型拟合曲线：



阻滞增长模型拟合曲线：



由两张图看见，指数增长模型从1900年起开始有较大的偏差，阻滞增长模型在1930-1960年一段拟合的不太好，猜测知道可能主要受第二次世界大战影响。这就是对数据预测结果的检验。

4. 参考资料

1. [数模官网 - 人口预测模型](#)
2. 《数学建模算法与应用》p.170~173.