按年龄分组的人口模型

20123211高雨晴 20123210王艺霏

2022 年 11 月 13 日

目录

1 问题重述 **2**
1.1 引言 2
1.2 问题 2

2 问题分析 **2**
2.1 概述 2
2.2 问题一的分析 3
2.3 问题二的分析 3

3 模型假设 **4**

4 定义与符号声明 **4**

5 模型的建立与求解 **4**
5.1 模型建立：Leslie矩阵人口模型 4
5.2 模型求解 6
5.2.1 对问题一的求解 6
5.2.2 对问题二的求解 8

6 模型评价 **9**
6.1 模型优点 9
6.2 模型缺点 9

参考文献 **9**

7 附录：MATLAB代码 **9**

1 问题重述

1.1 引言

生育问题一直是人口研究的重大问题。生育问题主要表现在生育水平和模式两个方面，不仅仅关系到对妇女生育历史的回顾，更关系到预测未来生育状况的趋势，基于此政府可以正确把握人口增长趋势，进行宏观调控，避免出现人口过多造成的粮食、土地稀缺问题和人口稀疏造成的资源浪费、劳动力不足等问题。

我国在不同时期出现过不同的人口问题和应对策略。在新中国成立初期，政府曾鼓励生育。但是1969 年中国总人口超过 8 亿。为控制人口过快增长，解决劳动力问题，我国于1982 年底，确立了计划生育政策，提倡晚婚晚育，有计划地控制人口数量。计划生育政策解决了当时人口数量过多的问题，得到了国内外的肯定，但也给中国人口带来了老龄化问题。

从上个世纪90 年代开始，我国人口增长越来越慢，而 64 岁以上老龄人口的增长速度却越来越快。随着我国从高生育率到低生育率的转变，出现人口老龄化、性别比失调等问题。

为解决人口增长率持续下降和老龄化问题，我国于 2015 年 10 月全面放开二孩政策。二孩政策放开之初，新生儿的数量确实在短时间内有所增加，但是时间很短暂。如今我国的出生率依然很低。为进一步解决人口增长速度急速下降和老龄化问题，中共中央政治局于 2021 年 5月宣布实施三孩政策，最大限度发挥人口对经济社会发展的能动作用。本文将建立人口模型，就三孩政策的必要性展开研究，提出一对夫妻生育几个孩子才能保持人口的平衡，从人口老龄化角度给出年轻人应该生或多生孩子的理由。

1.2 问题

问题一　根据历史人口统计数据，划分不同年龄组，并预测未来人口总量和年龄结构。

问题二　分析一胎、二胎、三胎政策对于未来人口结构的影响，给出合适的政策，使未来人口规模和结构更加合理。

2 问题分析

2.1 概述

人类因繁殖导致人口增加，因为自然死亡及其他因素而减少，不同年龄组的繁殖率、死亡率有较大的差别，因此在研究某一种群数量的变化时，需要考虑按年龄分组的种群增长。将人类按照年龄等间隔的分成若干组，时间也离散化为时段，根据各年龄组的繁殖率和死亡率，建立按照年龄分组的人口增长模型，以雌性个体数量为对象，根据每个雌性生育胎

儿的数量，预测未来各个年龄段的人口数量，并讨论时间充分长以后的变化趋势，以更好的制定相关政策，控制人口总量。

2.2 问题一的分析

通过分析我国历史人口相关数据，将年龄划分成不同组别，根据全国人口总数、男女性别比、不同年龄组的死亡率和生育率建立能够考虑年龄结构、性别比例、生育政策等因素的 Leslie 矩阵人口模型，预测未来达到的人口总量和不同年龄段之间所含的人口数量。 [1]

2.3 问题二的分析

分析育龄女性生育一胎、二胎、三胎对于人口总量产生的影响，结合人口老龄化、人口总量变化的趋势，找到最适合当前形势下的生育政策，以更好的控制未来人口总量，使人口总量更加合理，为政策的建立提建议。

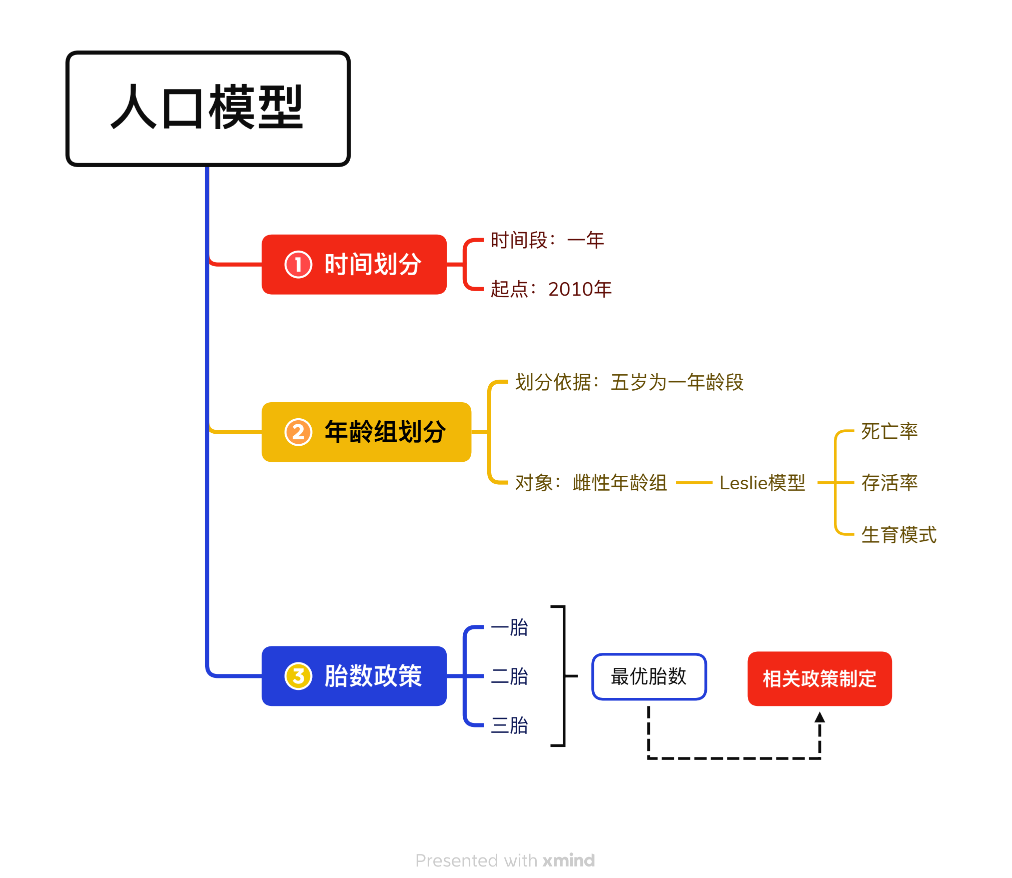

图 1: 建模流程

3 模型假设

假设一 种群按照年龄大小等分为n个年龄组，记为$i=0,1,2,...,n-1$.

假设二 把时间划分为时段，每年为一个时段，长度与年龄组区间相等，记$t=0,1,2,...$

假设三 模型所计算的时间范围内，女性群体的生育模式不发生变化。

4 定义与符号声明

符号	说明
$x_i(t)$	第t年i岁女性人口数量
$b_i(t)$	第t年i岁女性生育率
h_i	i岁女性生育模式
$\beta(t)$	总和生育率
$y_i(t)$	第t年i岁女性死亡数量
$d_i(t)$	第t年i岁女性死亡率
g_i	i岁女性死亡模式
$\alpha(t)$	总和死亡率

5 模型的建立与求解

5.1 模型建立：Leslie矩阵人口模型

我们用x_i表示第t年i岁的总女性数，其中$t=0,1,2,...$，$i=0,1,2,...,n-1$. 为计算方便，将所有年龄大于等于n的人数归于$i=n-1$. 人口的增长主要是由女性生育实现的。用$b_i(t)$表示第t年i岁女性的生育率，即每名女性平均生育婴儿的数目。则第$t+1$年出生的婴儿数目为

$$x_0(t+1)=\sum_{i=0}^{n}b_i(t)x_i(t) \tag{1}$$

其中当i不在育龄范围内（15-49岁）时，$b_i=0$. 于是定义生育模式h_i

$$h_i=\frac{b_i(t)}{\beta(t)},\quad \text{其中}h_i\text{满足}\sum_{i=0}^{n}h_i=1 \tag{2}$$

由(2)可知，生育模式h_i代表了i岁的女性生育婴儿的数目在第t年所有育龄女性生育婴儿的数目$\beta(t)$（即：总和生育率）中的占比。由(1)和(2)有

$$\begin{aligned} x_0(t+1) &= \beta(t) \sum_{i=0}^{n} h_i x_i(t) \\ \beta(t) &= \sum_{i=0}^{n} b_i(t) \end{aligned} \tag{3}$$

除去人口增长，人口会因为死亡而减少。记第t年年龄为i的女性死亡率为$d_i(t)$.则第t年的总死亡人口$y(t+1)$为

$$y(t+1) = \sum_{i=0}^{n} d_i(t) x_i(t) \tag{4}$$

类似(2)，我们定义死亡模式g_i，它代表了i岁女性人口死亡率在所有年龄段女性人口死亡率之和$\alpha(t)$（即：总和死亡率）中的占比，即

$$\begin{aligned} g_i &= \frac{d_i(t)}{\alpha(t)}, \quad \text{其中} g_i \text{满足} \sum_{i=0}^{n} g_i = 1 \\ \alpha(t) &= \sum_{i=0}^{n} d_i(t) \end{aligned} \tag{5}$$

由(5)自然地定义存活率$s_i(t) = 1 - \alpha(t) g_i$.则第$t+1$年，$i+1$岁的女性人口数为第$t$年$i$岁存活的女性人口数，为

$$x_{i+1}(t+1) = s_i(t) x_i(t) \tag{6}$$

考虑到$i+1=n$时表示年龄大于$n-1$的人，她们被归为年龄为$n-1$的类别中，故年龄为

$$x_{n-1}(t+1) = s_{n-1}(t) x_{n-1}(t) + s_n(t) x_n(t) \tag{7}$$

至此我们得到了所有年龄段人口数量随t变化的递推关系。将人口数量按年龄分组写成列向量

$$x(t) = [x_0(t), x_1(t), x_2(t), ..., x_{n-1}(t)]^\top \tag{8}$$

引入存活率矩阵$\mathbf{S}$，定义为

$$\mathbf{S} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1-\alpha(t)g_0 & 0 & \cdots & 0 & 0 \\ 0 & 1-\alpha(t)g_1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ \vdots & \vdots & & \vdots & 0 \\ 0 & 0 & \cdots & 1-\alpha(t)g_{n-2} & 1-\alpha(t)g_{n-1} \end{bmatrix} \tag{9}$$

以及出生矩阵$\mathbf{B}$

$$\mathbf{B}=\begin{bmatrix} 0 & \cdots & h_0 & \cdots & h_{n-1} & \cdots & 0 \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \tag{10}$$

则相邻两年，女性人口按年龄分组的向量的递推关系可改写为一个双线性方程

$$x(t+1)=\mathbf{S}x(t)+\beta(t)\mathbf{B}x(t) \tag{11}$$

5.2 模型求解

我们从国家统计局官方网站获得了2011年至2015年，15至49岁育龄女性的生育率 [2]。根据假设，我国妇女的生育模式短时间内不会发生过大变化，故将这五年数据取平均值作为生育模式的参考。由(2)和(10)，获得生育矩阵$\mathbf{B}$.

在第六次人口普查数据的官网上，我们获得了各个年龄女性人口的死亡率统计数据 [3]。由(5)，将各年龄女性人口死亡率向量归一化便得到死亡模式向量g_i. 为计算存活矩阵$\mathbf{S}$，我们还需要获知总和死亡率$\alpha(t)$. 显然，$\alpha(t)$和人口总数有关。由常识，在环境资源有限的情况下，当种群有过度繁殖倾向时，内部的竞争关系将变得激烈，从而导致种群死亡率上升。当t=0，即2010年时，按5岁为一个年龄组计算得出的总和死亡率为0.614. 我们假设同一年龄组内每个年龄的死亡率是相同的，则$\alpha(0)=0.614\times 5=3.07$.

另一个决定$\alpha(t)$的量是总和生育率$\beta(t)$. 在本题中，为求解保持人口数量稳定发展的最小总和生育率，我们令β的取值从1（即模拟严格执行一胎政策的情况）以0.1的步长取到3（即模拟严格执行三胎政策的情况）。由于模型只考虑女性，故在实际计算中，以$\frac{\beta}{2}$计。总和死亡率可由

$$\alpha(t)=\alpha(0)+e^{\beta(t)} \tag{12}$$

给出。

5.2.1 对问题一的求解

使用**MATLAB**对问题进行求解，代码见附录。结合实际情况，在本问中，我们将女性年龄划为四个区间：

0-14岁：儿童期 0-14岁的女性既无生育能力，也无生产能力，需要成年女性的抚养。

15-49岁：生育期 国际上通常认为15-49岁的女性处于生育年龄。这一部分的女性是人口增长的动力来源。

50-64岁：中年期　处于这一年龄段的女性往往虽然没有了生育能力，但仍然有生产能力，尚不需要子女赡养，也不会对社会产生过大压力。

65-99岁：老年期　高于65岁的女性认为进入老年期，这一年龄段的女性通常已经失去了生产能力，需要子女的赡养，并且可能由于身体状况欠佳而为社会带来一定的压力。

我们模拟了严格执行一胎政策、严格执行二胎政策和严格执行三胎政策的情况，统计2060年时的女性人口年龄结构，与2010年初始时比较。计算结果如图5.2.1所示。

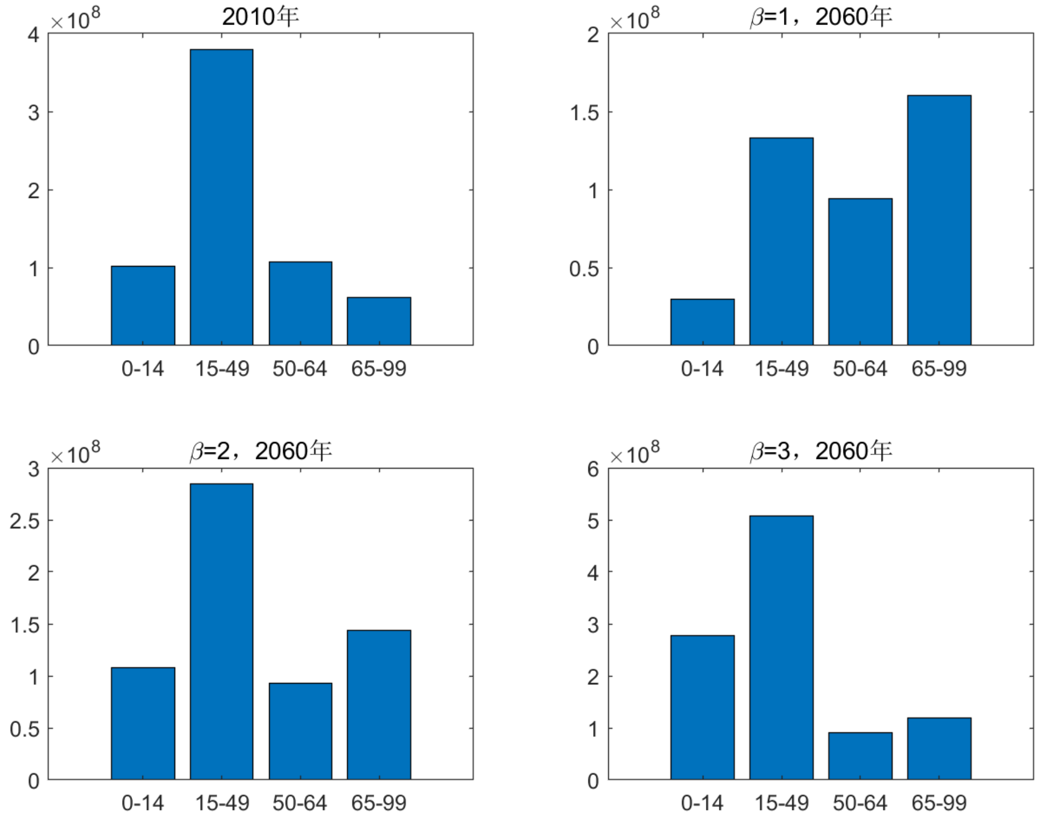

图 2: 三种政策影响下未来的年龄结构

由图可见，如果平均每位女性严格一生只生育一个孩子，那么在模型所假设的环境下，到2060年，中、老年期的女性将显著多余另外两个年龄组，儿童期的女性尤其少，这意味着社会将全面进入老龄化；如果平均每位女性严格一生生育两个孩子，则到2060年，老年期的女性占比略多于2010年，但总体上年龄结构和2010年的情况是相当的——即生育期的女性在数量上占优；如果平均每位女性严格一生生育三个孩子，则在50年后不仅所有年龄组的女性人口数都显著增加，且年龄结构也很健康，儿童期和生育期的女性占大多数。

5.2.2 对问题二的求解

合理调整总和生育率β，得到2060年不同总和生育率下的女性人口总数，与2010年女性人口总数比较，如5.2.2所示。

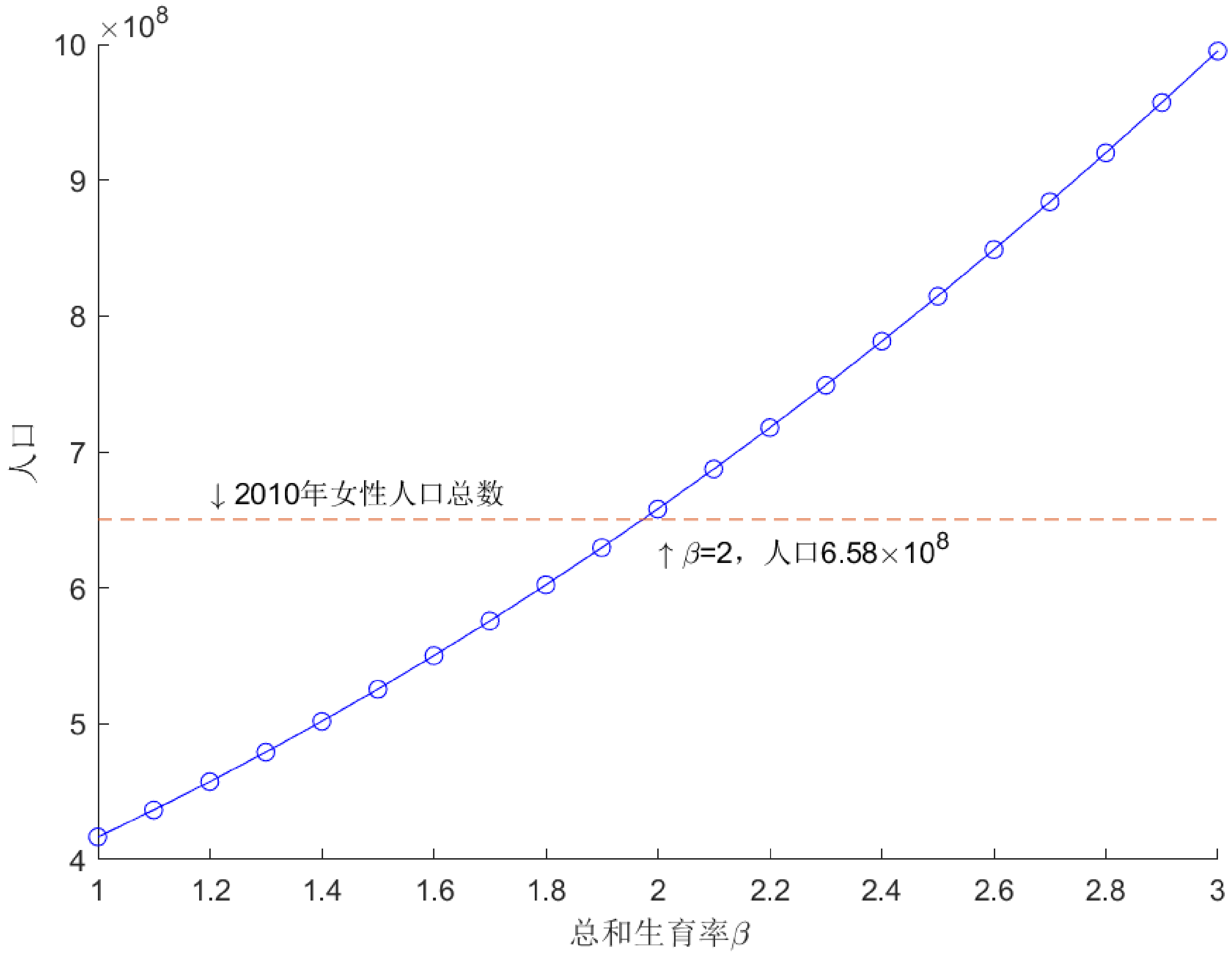

图 3: 不同总和生育率下50年后的人口变化

可见，当总和生育率$\beta = 2$时，人口刚好不发生减少，故最佳总和生育率应在2-2.2之间。

然而在政策的拟定中，政策提出者需要面对的问题是复杂的。例如，由于现在生育孩子的经济成本和时间成本巨大，再加上房贷、车贷以及人口老龄化所产生的双方老人的住院费用、抚养费用，这一系列的问题都让疲惫不堪的年轻人不敢要二胎。所以即便制定了二胎政策，也很难使总人口生育率达到2，不少家庭会因为各种原因没有生育二胎。但若是国家制定了三胎政策，在原有二胎政策的基础上给予人们更多奖励措施和福利待遇，许多有条件的家庭就会生育第三胎，弥补了一些因为经济压力、心理因素等没有生育二胎的家庭所造成的生育率不足，因此解释了三胎政策的合理性。

6 模型评价

6.1 模型优点

模型对各年全国女性人口总数进行预测时，结合了实际情况，用数年的育龄女性生育年龄数据拟合了生育模式，使模型的计算结果更加准确。并且，考虑了女性人口的死亡模式，并引入了总和死亡率α，以及其与总和生育率的动态关系。

6.2 模型缺点

实际上，用$\beta/2$来模拟女性只生育女儿的情况是不科学的，应当根据男女出生比来考虑。并且，总和死亡率α随总和生育率β的变化模型比较粗糙，仍需更多数据来归纳普遍规律。

参考文献

[1] 姜启源, 数学模型（第二版）． 高等教育出版社, 2011, vol. 208-213.

[2] 国家统计局“按年龄分生育率”. [Online]. Available: https://data.stats.gov.cn/easyquery.htm?cn=C01&zb=A03060H&sj=2021

[3] 中国2010年人口普查资料. [Online]. Available: http://www.stats.gov.cn/tjsj/pcsj/rkpc/6rp/indexch.htm

7 附录：MATLAB代码

Listing 1: **main.m**

```
clc;
load("deathrate.mat");
deathrate=deathrate./sum(deathrate); % deathmode
% sum(deathrate)
count=1;
deathrate_perage=zeros(1,100);
for i=1:100
    deathrate_perage(i)=deathrate(count)./5;
    if mod(i,5)~=0
        continue
```

```
    else
        count=count+1;
    end
end
% sum(deathrate_perage)
D=diag(deathrate_perage(1:99),-1);
D(100,100)=D(100,99);
C=diag(ones([1,99]),-1);
C(100,100)=1;
% alpha=?; % total death rate
% S=C-alpha.*D; % survival matrix

% total fertility rate
beta=ones(1,21)./2;
for i=1:20
    beta(i+1)=beta(i)+0.05;
end
load("childbearingage.mat");
y=mean(childbearingage(:,2:6),2);
h=y/sum(y); % bearing mode
B=zeros(100,100); % bearing matrix
for i=15:49
    B(1,i)=h(i-14);
end
result=zeros(40,21);
load("population.mat");

figure(1)
subplot(2,2,1)
popstructure(population);
title('2010 year')

x1=sum(population);
for i=1:21
    population2=population;
    alpha=0.614*5+exp(beta(i)); % total death rate
```

```
    for j=1:50 % 2010-2060 year
        population2=(C-alpha.*D)*population2+beta(i)*B*
           population2;
        result(j,i)=sum(population2);
    end

    switch(i)
        case(1)
            figure(1)
            subplot(2,2,2)
            popstructure(population2);
            title('\beta=1, 2060 year')
        case(10)
            figure(1)
            subplot(2,2,3)
            popstructure(population2);
            title('\beta=2, 2060 year')
        case(21)
            figure(1)
            subplot(2,2,4)
            popstructure(population2);
            title('\beta=3, 2060 year')
    end
end
start_pop=x1.*ones(1,21);

figure(2)
hold on
xlabel('total fertility rate\beta')
ylabel('population')
plot(1:0.1:3,result(50,:),'b-o')
plot(1:0.1:3,start_pop,'--')
text(1.2,6.7e8,'\downarrow women population in 2020')
text(2,6.3e8,'\uparrow \beta=2, population=6.58\times10^{8}')
hold off
```

Listing 2: **popstructure.m**

```
function [structure] = popstructure(population)
% input: population with age of 0-99
% output: population in 0-14, 15-49, 50-64, 65-99
structure=zeros(1,4);
structure(1)=sum(population(1:15));
structure(2)=sum(population(16:50));
structure(3)=sum(population(51:65));
structure(4)=sum(population(66:100));
bar(structure);
set(gca,'xticklabel',{'0-14','15-49','50-64','65-99'});
end
```