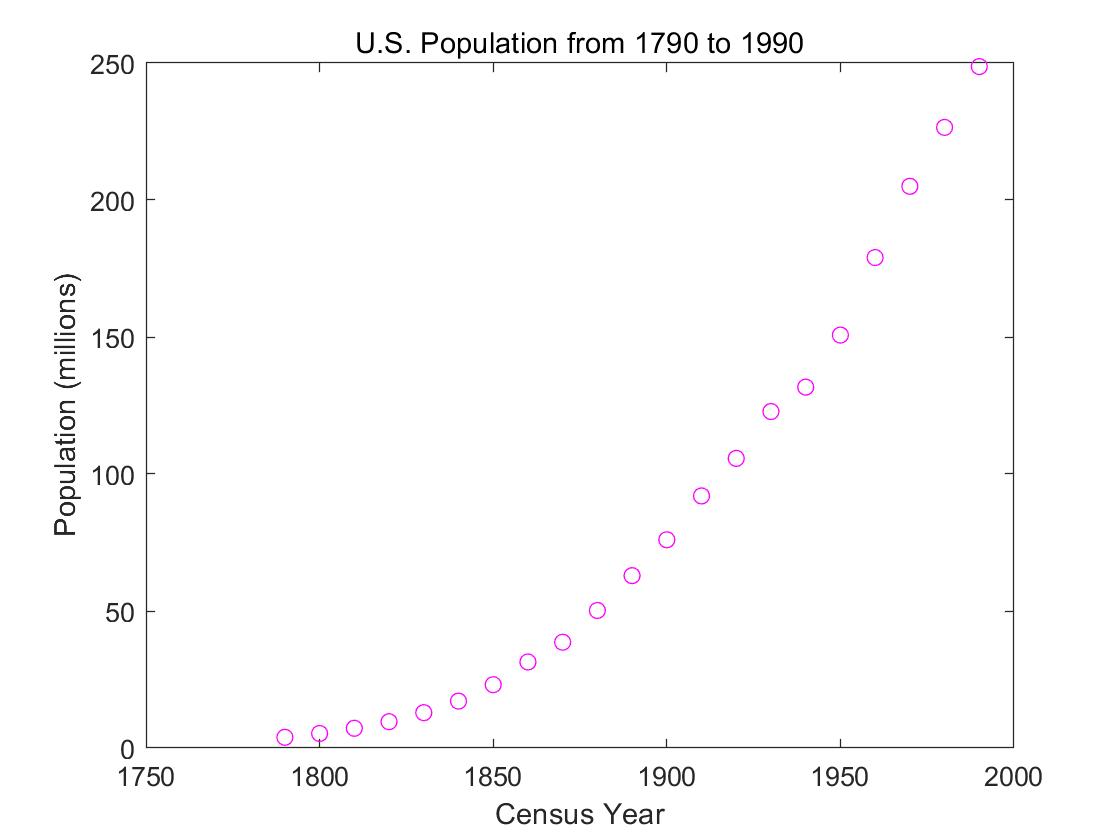
# 一．程序源码

1. %%针对matlab提供的census.mat数据文件，将其中的数据进行多项式拟合，
2. %%给出数学模型及详细的的拟合过程
3. clear;clc;
4. load census.mat  %导入数据
5. %计算样本的相关系数矩阵，并画散点图，证明二者有较强的相关性
6. figure(1);
7. plot(cdate,pop,'om')
8. title('U.S. Population from 1790 to 1990');
9. xlabel('Census Year');
10. ylabel('Population (millions)');
11. corrcoef(cdate,pop);
12. %样本数据标准化
13. sdate=(cdate-mean(cdate))./std(cdate);
14. %根据不同的拟合次数先画拟合图和残差图，以选择合适的模型
15. k=2;
16. **for** i=1:4
17. t=polyfit(sdate,pop,i);
18. fit\_y=polyval(t,sdate);
19. figure(k);k=k+1;
20. plot(cdate,fit\_y,'+-',cdate,pop,'om');
21. title('U.S. Population from 1790 to 1990')
22. legend('Polynomial Model','Data','Location','NorthWest');
23. xlabel('Census Year');
24. ylabel('Population (millions)');
25. hold off;
26. figure(k);k=k+1;
27. res = pop - fit\_y;
28. plot(cdate,res,'+m')
29. title(['Residuals for the Quadratic Polynomial Model(n=',num2str(i),')’]);
30. end
31. %计算3次拟合的残差和拟合优度
32. t=polyfit(sdate,pop,3);
33. fit\_y=polyval(t,sdate);
34. yresid = pop - fit\_y;
35. SSresid = sum(yresid.^2);
36. SStotal = (length(pop)-1) \* var(pop);
37. rsq = 1 - SSresid/SStotal

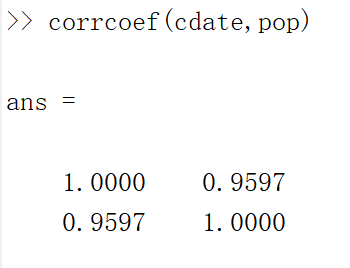
# 二．代码分析

## 2.1相关性分析

首先画出散点图，直观上看出原始样本数据具有较强的相关性，可以进行多项式拟合。

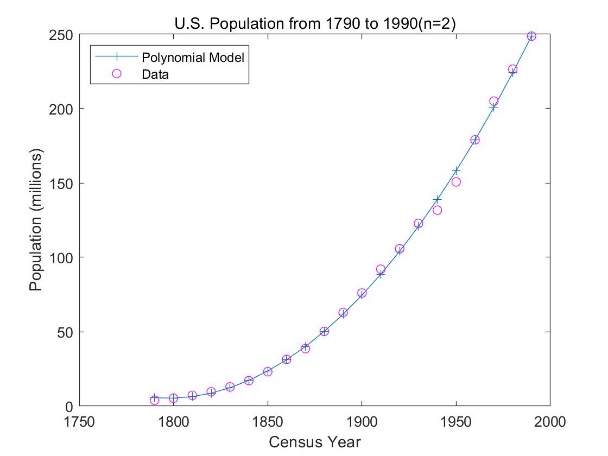
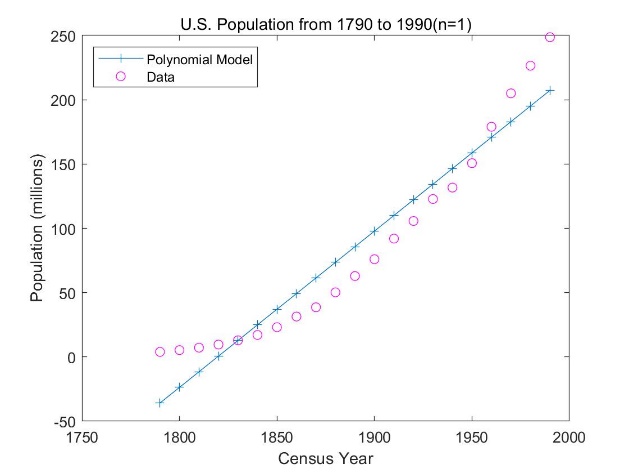


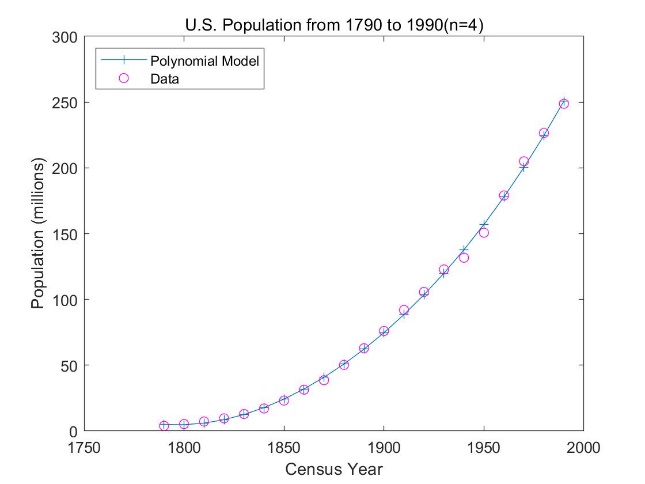
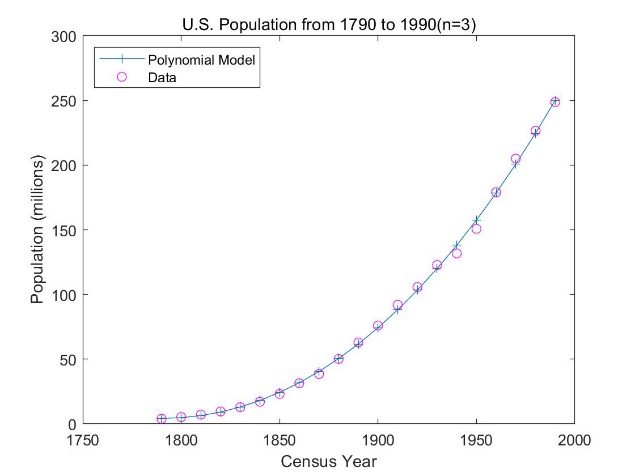
随后，计算原始样本数据的协方差矩阵，进一步证明数据的强相关性



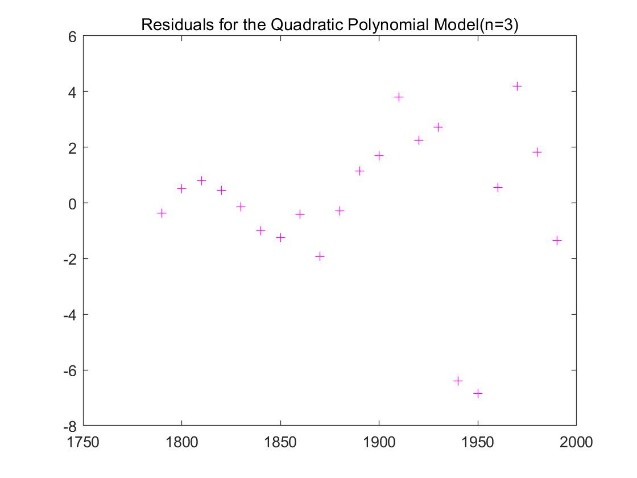
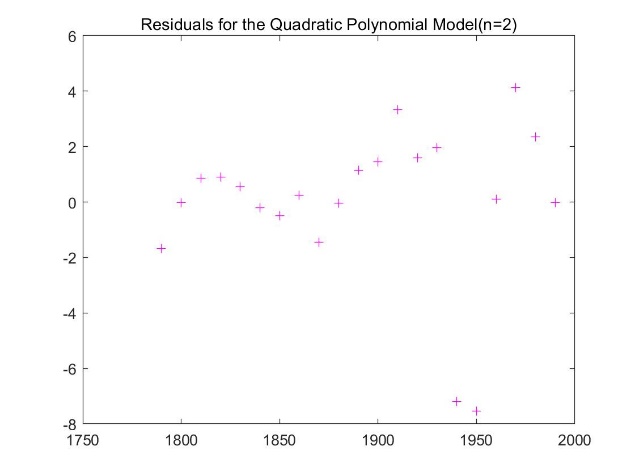
## 2.2多项式拟合

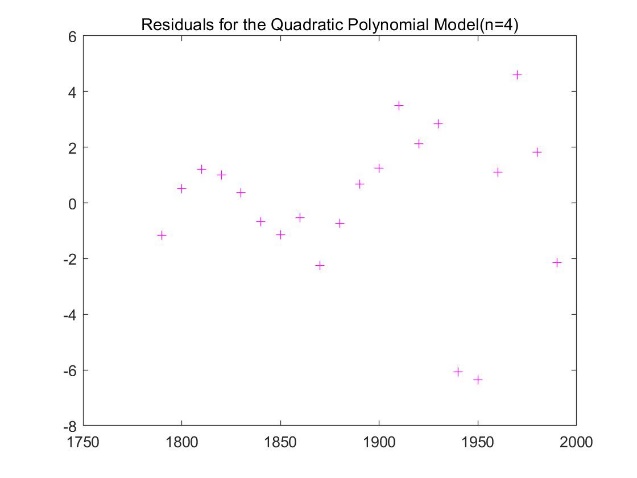
根据不同的拟合次数分别画图：





可以看出，n=2,3,4时，拟合效果差别不大，下面进行残差分析，根据4次拟合效果，画出残差图。





残差是因变量的观测值与模型的预测值之间的差。当拟合的模型适合数据时，残差接近独立随机误差。即，残差分布不应该呈现出可辨识的模式。如果残差图具有一定的模式（即残差数据点未呈现随机分布，该随机性表明该模型并未适当地拟合数据。n=和n=4d 残差图较相似,n=3比n=2时的更呈现一种随机模式，于是选取n=3时作为多项式拟合。

## 2.3拟合优度分析

拟合优度的一个度量是决定系数或 。该统计量表明您通过拟合模型得到的值与模型 可预测的因变量的匹配程度。统计人员通常利用拟合模型的残差方差定义 ：  ，SSresid 是与回归的残差的平方和。SStotal 是与因变量均值的差的平方和。计算得=0.9988，拟合效果较好。

# 三．最小二乘拟合

t=polyfit(sdate,pop,3)

pop1=polyval(t,sdate)

plot(cdate,pop,'o',cdate,pop1,'-');

f=@(a,sdate)a(1)\*sdate.^3+a(2)\*sdate.^2+a(3)\*sdate+a(4) %建立一个匿名函数模型

[a,e]=lsqcurvefit(f,[1,0,0,0],sdate,pop1) %最小二乘法

由于样本点大部分落在三阶拟合曲线上，所以，最小二乘法拟合也同样表明三阶拟合是较好的模型。

# 四．数学模型

由以上分析，拟合次数为3时多项式系数分别为，0.921，25.183，73.859，61.744可得到最终的三阶多项式拟合的美国人口增长数学模型

## 