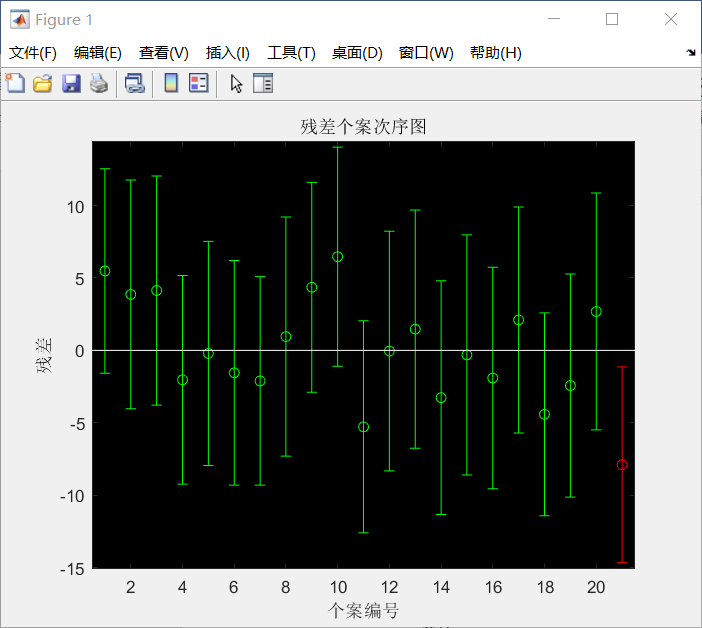
# 练习2 多元线性回归问题(用regress、polyfit或nlinfit求解)

某人记录了21天每天使用空调器的时间和使用烘干器的次数，并监视电表以计算出每天的耗电量，数据见下表，试研究耗电量（KWH，记作y）与空调器使用的小时数(AC，记作)和烘干器使用次数(DRYER，记作)之间的关系：

1. 建立y与x1、x2之间的线性回归模型，并分析模型效果的显著性；
2. x1=[1.5 4.5 5.0 2.0 8.5 6.0 13.5 8.0 12.5 7.5 6.5 8.0 7.5 8.0 7.5 12.0 6.0 2.5 5.0 7.5 6.0]'
3. x2=[1 2 2 0 3 3 1 1 1 2 3 1 2 2 1 1 0 3 0 1 0]'
4. y=[35 63 66 17 94 79 93 66 94 82 78 65 77 75 62 85 43 57 33 65 33]'
5. X=[ones(size(x1)),x1,x2];
6. [b,bint,r,rint,stats]=regress(y,X)
7. rcoplot(r,rint)



根据残差图，去掉异常点后

b =

9.7966

5.4160

12.5843

Y=9.7966+5.4160x1+12.5843x2

①r2=0.975878 （样本97.09%的信息被回归模型学习到）

②F=343.876

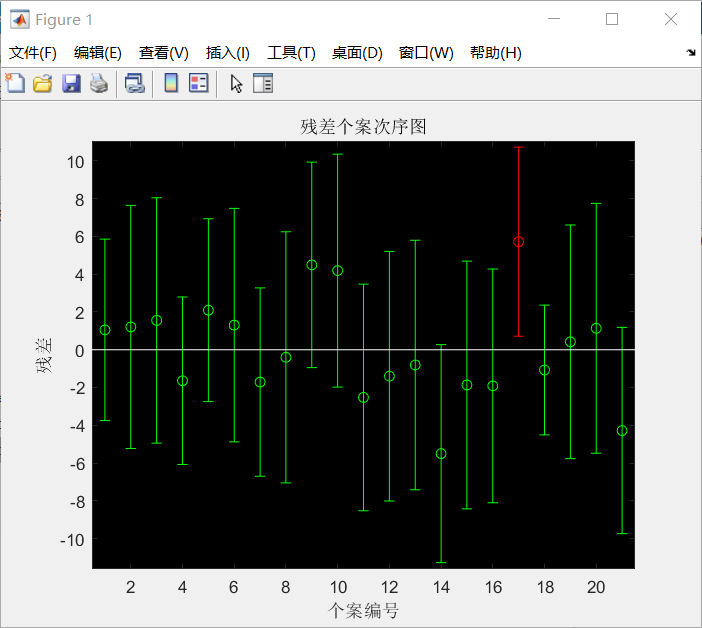
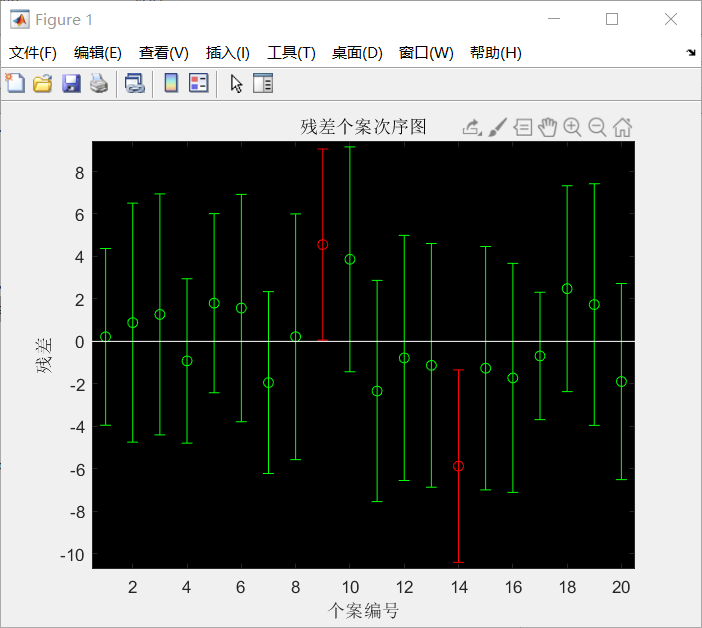
1. n1=1;n2=19;a=0.95
2. x\_a=finv(a,n1,n2)

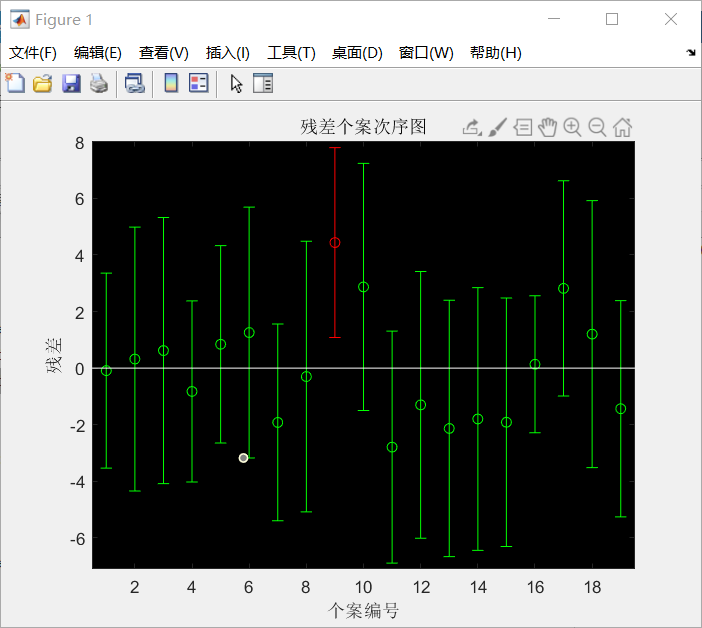
根据上述代码可解得F1-α（n-2）=4.3807，F≥F1-α（n-2），拒绝H0，模型正确

③p=1.78034e-14(p值为接受回归模型的风险，此犯错概率极低

由上可知，回归显著

1. 如有必要，考虑引入非线性项（平方项x\_1^2，x\_2^2以及交叉项x1\*x2），建立新的回归模型；
2. x1=[1.5 4.5 5.0 2.0 8.5 6.0 13.5 8.0 12.5 7.5 6.5 8.0 7.5 8.0 7.5 12.0 6.0 2.5 5.0 7.5 6.0]'
3. x2=[1 2 2 0 3 3 1 1 1 2 3 1 2 2 1 1 0 3 0 1 0]'
4. y=[35 63 66 17 94 79 93 66 94 82 78 65 77 75 62 85 43 57 33 65 33]'
5. x3=x1.\*x1
6. x4=x2.\*x2
7. x5=x1.\*x2
8. X=[ones(size(x1)),x1,x2,x3,x4,x5]
9. [b,bint,r,rint,stats]=regress(y,X)
10. rcoplot(r,rint)
11. y(17)=[]
12. X(17,:)=[]
13. y(14)=[]
14. X(14,:)=[]
15. stepwise(X(:,2:end),y)





根据残差去掉异常点后（第一次去掉17，第二次去掉14,第三次去掉9）

b =

9.7563

4.0644

22.0671

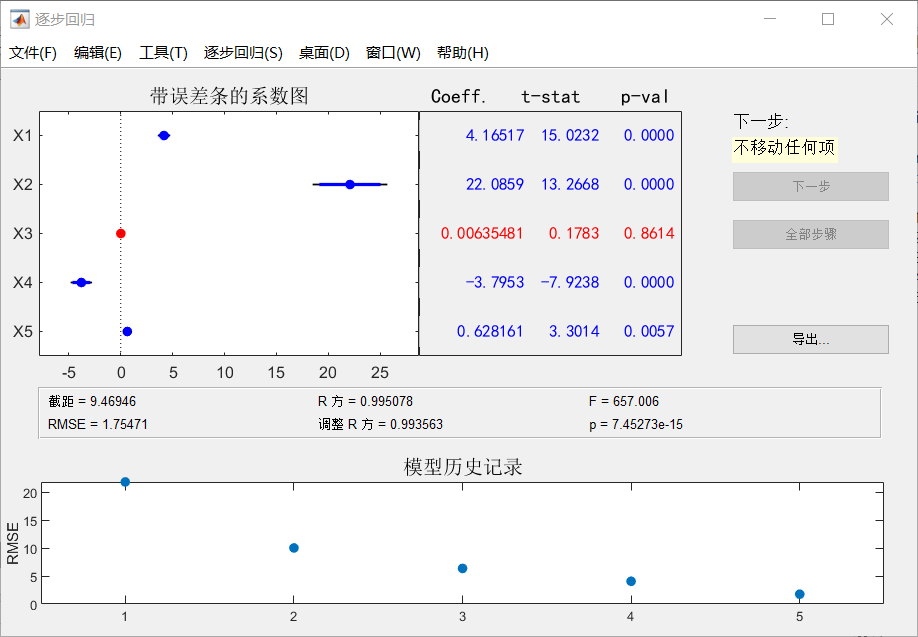
0.0064

-3.7979

0.6364

回归模型：y=9.7563+4.0644x1+22.0671x2+0.0064x3-3.7979x4+0.6364x5

（3）分析模型中新引入的非线性项是否都是必要的，若不是，请去掉多余项，建立新的模型，并分析新模型的效果。



X3不必要，X4（x\_2^2）和x5（x1\*x2）对回归模型显著，引入x4和x5

新的回归模型：y=9.7563+4.0644x1+22.0671x2-3.7979x4+0.6364x5新回归模型的r2=0.995078，更多的样本信息被回归模型学习到

F=657.006比F1-α（n-2）大，可证明模型正确

P=7.45273e-15犯错概率也更低。