学生学号	0121618990118	实验课成绩	

学 生 实 验 报 告 书

武漢理工大學

实验课程名称	实用回归分析
开 课 学 院	理学院
指导教师姓名	李丹
学 生 姓 名	薛峰
学生专业班级	统计 1602

2018 -- 2019 学年 第 二 学期

实验教学管理基本规范

实验是培养学生动手能力、分析解决问题能力的重要环节；实验报告是反映实验教学水平与质量的重要依据。为加强实验过程管理，改革实验成绩考核方法，改善实验教学效果，提高学生质量，特制定实验教学管理基本规范。

1、 本规范适用于理工科类专业实验课程，文、经、管、计算机类实验课程可根据具体情况参照执行或暂不执行。

2、 每门实验课程一般会包括许多实验项目，除非常简单的验证演示性实验项目可以不写实验报告外，其他实验项目均应按本格式完成实验报告。

3、 实验报告应由实验预习、实验过程、结果分析三大部分组成。每部分均在实验成绩中占一定比例。各部分成绩的观测点、考核目标、所占比例可参考附表执行。各专业也可以根据具体情况，调整考核内容和评分标准。

4、 学生必须在完成实验预习内容的前提下进行实验。教师要在实验过程中抽查学生预习情况，在学生离开实验室前，检查学生实验操作和记录情况，并在实验报告第二部分教师签字栏签名，以确保实验记录的真实性。

5、 教师应及时评阅学生的实验报告并给出各实验项目成绩，完整保存实验报告。在完成所有实验项目后，教师应按学生姓名将批改好的各实验项目实验报告装订成册，构成该实验课程总报告，按班级交课程承担单位（实验中心或实验室）保管存档。

6、 实验课程成绩按其类型采取百分制或优、良、中、及格和不及格五级评定。

附表：实验考核参考内容及标准

	观测点	考核目标	成绩组成
实验预习	1. 预习报告 2. 提问 3. 对于设计型实验，着重考查设计方案的科学性、可行性和创新性	对实验目的和基本原理的认识程度，对实验方案的设计能力	20%
实验过程	1. 是否按时参加实验 2. 对实验过程的熟悉程度 3. 对基本操作的规范程度 4. 对突发事件的应急处理能力 5. 实验原始记录的完整程度 6. 同学之间的团结协作精神	着重考查学生的实验态度、基本操作技能；严谨的治学态度、团结协作精神	30%
结果分析	1. 所分析结果是否用原始记录数据 2. 计算结果是否正确 3. 实验结果分析是否合理 4. 对于综合实验，各项内容之间是否有分析、比较与判断等	考查学生对实验数据处理和现象分析的能力；对专业知识的综合应用能力；事实求实的精神	50%

实验课程名称：_实用回归分析____

实验项目名称	非线性回归模型的实现			实验成绩	
实 验 者	薛峰	专业班级	统计 1602	组　　别	
同 组 者	欧阳宵			实验日期	年　月　日

第一部分：实验数据及要求

数据：P236 9.6 中的数据

1、 列举一些非线性模型的数学形式，以及线性化的方法。

2、 构建乘性误差项模型：先对模型进行线性化，然后拟合模型；

3、 构建加性误差项模型：用非线性最小二乘法做拟合；

4、 比较乘性误差项模型与加性误差项模型的区别。

第二部分：实验过程记录（可加页）（包括实验原始数据记录，实验现象记录，实验过程发现的问题等）

1. 列举一些非线性模型的数学形式，以及线性化的方法。

(1)幂函数模型

如：$y_t = ax_t^b$

在上式等号两侧同取自然对数，得$\ln y_t = lna + blnx_t$

(2)指数函数模型

如：$y_t = ae^{bx_t}$

在上式等号两侧同取自然对数，得$\ln y_t = lna + bx_t$

(3)多项式方程模型

如：$y_t = b_0 + b_1x_t + b_2x_t^2 + b_3x_t^3$

分别令$x_{t1} = x_t$，$x_{t2} = x_t^2$，$x_{t3} = x_t^3$，

可得到三元线性函数：$y_t = b_0 + b_1x_{t1} + b_2x_{t2} + b_3x_{t3}$

(4)双曲线函数模型

如：$1/y_t = a + b/x_t$

分别令$y_t^* = 1/y_t$，$x_t^* = 1/x_t$

可得到：$y_t^* = a + bx_t^*$

(5)生长曲线模型

如：$y_t = \frac{k}{1+be^{-at}}$

两侧同时除y_t，同乘$1 + be^{-at}$，得到：$\frac{k}{y_t} = 1 + be^{-at}$，将 1 移动到方程左边，再在等式两侧取自然对数得到：$\ln\left(\frac{k}{y_t} - 1\right) = \mathrm{lnb} - \mathrm{at}$, 令$y_t^* = \ln\left(\frac{k}{y_t} - 1\right)$，$b^* = \mathrm{lnb}$

可得到：$y_t^* = lnb - at$

2. 构建乘性误差项模型：先对模型进行线性化，然后拟合模型

本文在导入数据后，首先构建了乘性误差项模型如下：

$$\mathrm{y} = \mathrm{A}e^{ut}K^aL^be^{\varepsilon}$$

然后对模型进行了对数线性化处理得到方程为：

$$\mathrm{lny} = \mathrm{lnA} + \mathrm{ut} + \mathrm{alnK} + \mathrm{blnL}$$

接着对上述模型进行了线性拟合，再接着对于拟合后的模型重新对其基本假设进行了检验，最后对于模型的部分不足进行了改进，具体代码及分析如下：

2.1 程序代码

```
#导入数据
data=read.csv("C:/Users/lenovo/Desktop/ 实 验 四 / 实 验 四 数
```

```
据.csv",head=T);data;
y=data[,4]
t=data[,2]
K=data[,5]
L=data[,6]

#乘性误差
fit=lm(log(y)~log(K)+log(L)+t)
fit
summary(fit)

#正态性及异方差性检验
par(mfrow=c(2,2))
plot(fit1)

#检查 fit 多重共线性
XX=cor(cbind(t,K,L));XX;
kappa(XX,exact=TRUE);
eigen(XX);

#检查 fit 自相关性
library(zoo)
library(lmtest)
dwtest(fit)
```

2.2 结果及分析

```
> data=read.csv("C:/Users/lenovo/Desktop/实验四/实验四数据.csv",head=T);data;
   年份  t     CPI      GDP      K     L
1  1978  1  100.00  3624.1 1377.9 40125
2  1979  2  101.90  3962.9 1446.7 41024
3  1980  3  109.54  4124.2 1451.5 42361
4  1981  4  112.28  4330.6 1408.1 43725
5  1982  5  114.53  4623.1 1536.9 45295
6  1983  6  116.82  5080.2 1716.4 46436
7  1984  7  119.97  5977.3 2057.7 48197
8  1985  8  131.13  6836.3 2582.2 49873
9  1986  9  139.65  7305.4 2754.0 51282
10 1987 10  149.85  7983.2 2884.3 52783
11 1988 11  178.02  8385.9 3086.8 54334
12 1989 12  210.06  8049.7 2901.5 55329
13 1990 13  216.57  8564.3 2975.4 64749
14 1991 14  223.94  9653.5 3356.8 65491
15 1992 15  238.27 11179.9 4044.2 66152
16 1993 16  273.29 12673.0 5487.9 66808
17 1994 17  339.16 13786.9 5679.0 67455
18 1995 18  397.15 14724.3 6012.0 68065
19 1996 19  430.12 15782.8 6246.5 68950
20 1997 20  442.16 16840.6 6436.0 69820
21 1998 21  438.62 17861.6 6736.1 70637
22 1999 22  432.48 18975.9 7098.9 71394
23 2000 23  434.21 20604.7 7510.5 72085
24 2001 24  437.25 22256.0 8567.3 73025
25 2002 25  433.75 24247.0 9764.9 73740
```

```
> fit=lm(log(y)~log(K)+log(L)+t)
> fit

Call:
lm(formula = log(y) ~ log(K) + log(L) + t)

Coefficients:
(Intercept)       log(K)       log(L)            t
    5.15483      0.46006     -0.02737      0.04184
```

```
> summary(fit)

Call:
lm(formula = log(y) ~ log(K) + log(L) + t)

Residuals:
      Min        1Q    Median        3Q       Max
-0.036638 -0.010917  0.001029  0.016708  0.044192

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.154834   1.014673   5.080 4.96e-05 ***
log(K)       0.460063   0.045620  10.085 1.67e-09 ***
log(L)      -0.027374   0.091963  -0.298    0.769
t            0.041839   0.004622   9.052 1.08e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02235 on 21 degrees of freedom
Multiple R-squared:  0.9988,    Adjusted R-squared:  0.9986
F-statistic:  5664 on 3 and 21 DF,  p-value: < 2.2e-16
```

导入数据后，对模型进行对数线性化处理，最终得到的回归方程如下：

$$\ln y = 5.155 + 0.46\ln K - 0.027\ln L + 0.0418t$$

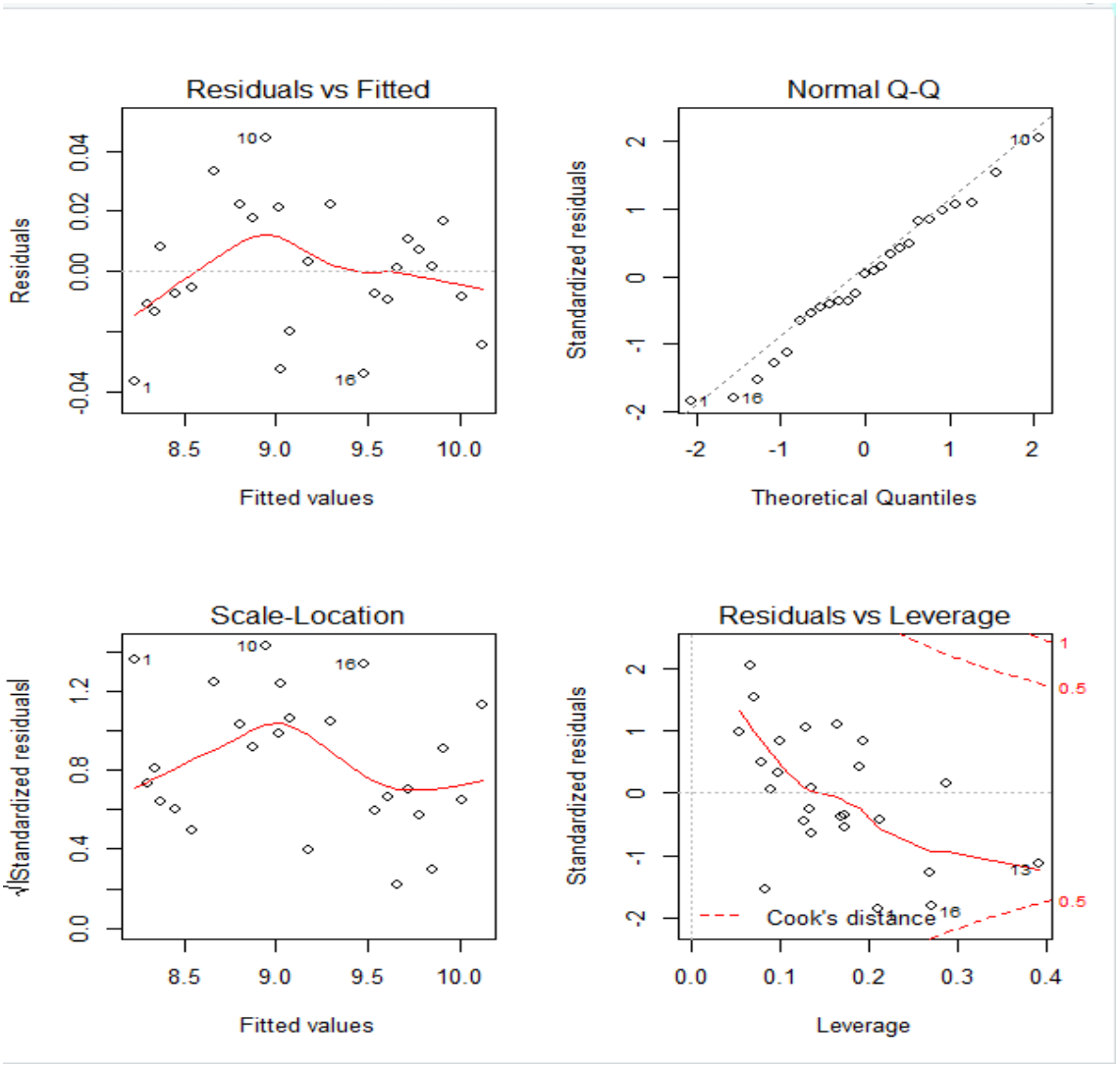
此时，模型的决定系数为 0.9988，调整后的决定系数为 0.9986，标准估计量的

误差为0.02235，且此时除lnL项外，其余回归系数均十分显著，说明模型拟合效果很好。但由于存在系数不显著且产出与劳动力投入成反比，与实际经济意义不符，故认为模型仍存在不合理之处，于是本文又对拟合后的模型进行了诊断，主要包括自相关性诊断与多重共线性诊断，具体代码及分析如下：

```
> #检查fit多重共线性
> XX=cor(cbind(t,K,L));XX;
          t         K         L
t 1.0000000 0.9647929 0.9774590
K 0.9647929 1.0000000 0.9136937
L 0.9774590 0.9136937 1.0000000
> kappa(XX,exact=TRUE);
[1] 330.1796
> eigen(XX);
eigen() decomposition
$values
[1] 2.904226649 0.086977450 0.008795901

$vectors
           [,1]         [,2]        [,3]
[1,] 0.5849646  0.07460502  0.8076203
```

```
> library(lmtest)
> dwtest(fit)

        Durbin-Watson test

data:  fit
DW = 1.2847, p-value = 0.006203
alternative hypothesis: true autocorrelation is greater than 0
```

得到的结果如下所示：



首先，左上角是残差-拟合值点图，左下角是标准化残差-拟合值点图。通过结果图可以发现，残差点在图中随机分布，不具有任何模式性，故认为对原始数据的的多元回归方程不存在异方差性。右上角是残差 Q-Q 图，主要用来判断样本是否近似于正态分布。可以看到图中数据点按对角直线排列，趋于一条直线，并被对角穿过，直观上可判定该回归模型中随机项通过正态分布的假设。右下角是标准化残差和杠杆值，出现了等高线，说明数据中存在特别影响结果的异常点。

其次，通过相关系数矩阵可以发现，自变量之间的相关性都比较大，相关系数全在 0.9 以上。同时通过特征根判定法可以看到，模型的最大条件数高达 330.1796，大于 100，且存在特征根近似为零的情况，故可认为模型存在严重的多重共线性。

最后，本文对模型进行自相关性诊断结果后，发现其 DW 值为 1.2847，而显著性水平为 0.05，自由度为 25，解释变量个数为 4 时，DW 检验的上下界为：$d_L = 1.12$，$d_U = 1.66$，通过 DW 检验无法判定其是否存在自相关性。但是，此时模型的 P 值很小，为 0.006203，拒绝原假设，认为模型存在自相关性。考虑到经济数据往往随时间存在共同的变化趋势，容易存在自相关性，故本文认为题目所给数据有自相关性存在。

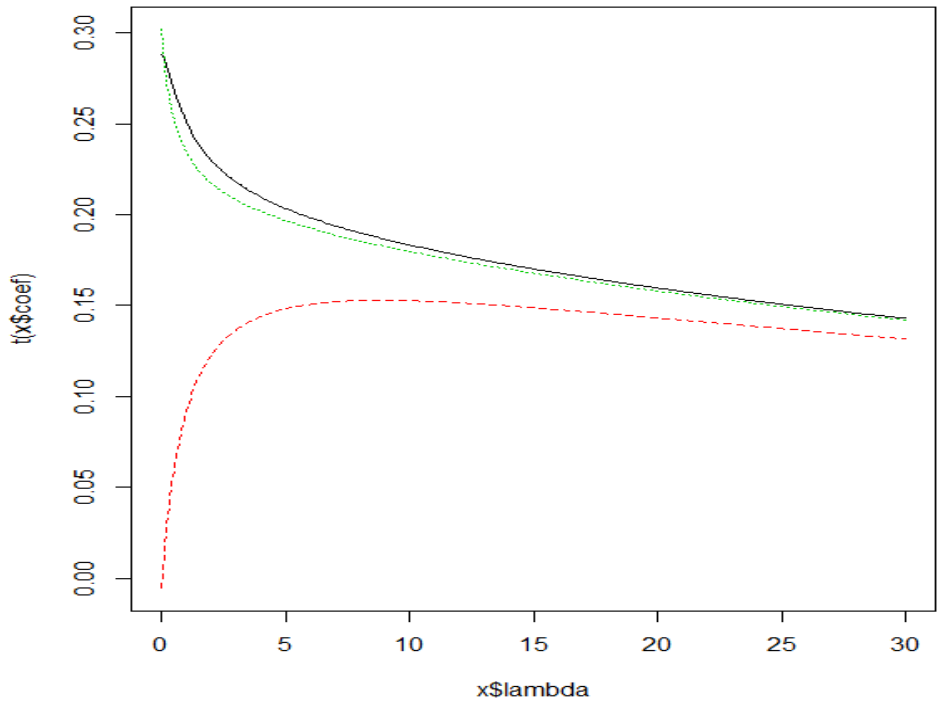
以下，本文针对多重共线性，采用岭回归法给出了具体的改进方案。

2.3.1 改进：利用岭回归法消除多重共线性

代码如下

```
#解决 fit 多重共线性
#岭迹法
library(MASS)
ridge<-lm.ridge(log(y)~log(K)+log(L)+t,lambda=seq(0,30,0.1))
beta=coef(ridge)
beta

plot(ridge)#绘制岭迹图

select(ridge)#自动挑选出 K

library(ridge)
mymod <- linearRidge(log(y) ~ log(K)+log(L)+t,lambda=0.01)#K 选定的是 0.01
summary(mymod)#使用岭迹法得出了回归方程，各个系数显著
fit_y=predict(mymod,interval="prediction")
R_2=(cor(y,fit_y))^2
R_2
```

结果如下：



```
Call:
linearRidge(formula = log(y) ~ log(K) + log(L) + t, lambda = 0.01)


Coefficients:
            Estimate Scaled estimate Std. Error (scaled) t value (scaled) Pr(>|t|)
(Intercept)  3.17030              NA                 NA               NA       NA
log(K)       0.44661         1.39877            0.09024            15.50   <2e-16 ***
log(L)       0.16919         0.17320            0.07907             2.19   0.0285 *
t            0.03720         1.34126            0.09668            13.87   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Ridge parameter: 0.01

Degrees of freedom: model 2.331 , variance 1.924 , residual 2.738

> fit_y=predict(mymod,interval="prediction")
> R_2=(cor(y,fit_y))^2
> R_2
[1] 0.9399606
```

本文通过岭回归法对模型进行了改进。从岭迹图可以看出，劳动力投入、资本和时间的岭回归系数均较为稳定，通过岭迹法来确定岭迹参数 k 的选择，最终确定 $k=0.1$，此时，岭迹大体上已达到稳定。得到改进的回归方程为：

$$\ln y = 3.1703 + 0.4466\ln K + 0.1692\ln L + 0.0372t$$

此时回归方程的各项系数在显著性水平为 0.05 的条件下均显著。且此时产出与资本、劳动力投入和随时间的科技进步均成正相关，符合实际经济意义。同时计算出的拟合优度达到了 0.9399，表明改进的回归方程是极为成功的。

2.3.2 改进：利用差分法消除自相关性

由于在 2.1 中得到的结果无法判别是否存在自相关性，且一般认为时间序列数据存在自相关性，所以本文做了消除自相关性的操作。

代码如下：

```
#消除自相关
#差分法
difflogy <- diff(log(y))
difflogK <- diff(log(K))
difflogL <- diff(log(L))
difft <- diff(t)
diff.reg <- lm(difflogy~difflogK+difflogL+difft-1)
summary(diff.reg)
durbinWatsonTest(diff.reg)
```

所得结果：

```
 lag Autocorrelation D-W Statistic
   1       -0.1215184       2.186018
```

从结果可知，DW 的值为 2.186。n=25，k=3，显著性水平 $\alpha=0.05$，查看 DW 表得，$d_U=1.55, d_L=1.21$，由于 $d_U<2.186<4-d_U$，所以误差项之间无自相关。通过 1 阶差分法成功地消除了自相关性。

3. 构建加性误差项模型：用非线性最小二乘法做拟合

本文在导入数据后，首先构建了加性误差项模型如下：

$$y = Ae^{ut}K^aL^b + \varepsilon$$

对于加性误差项模型，不能通过变量变换转化成线性模型，只能用非线性最小二乘求解未知参数。以上面乘性误差项的参数为初始值做非线性最小二乘，计算代码如下以及结果分析如下：

2.1 非线性最小二乘程序代码

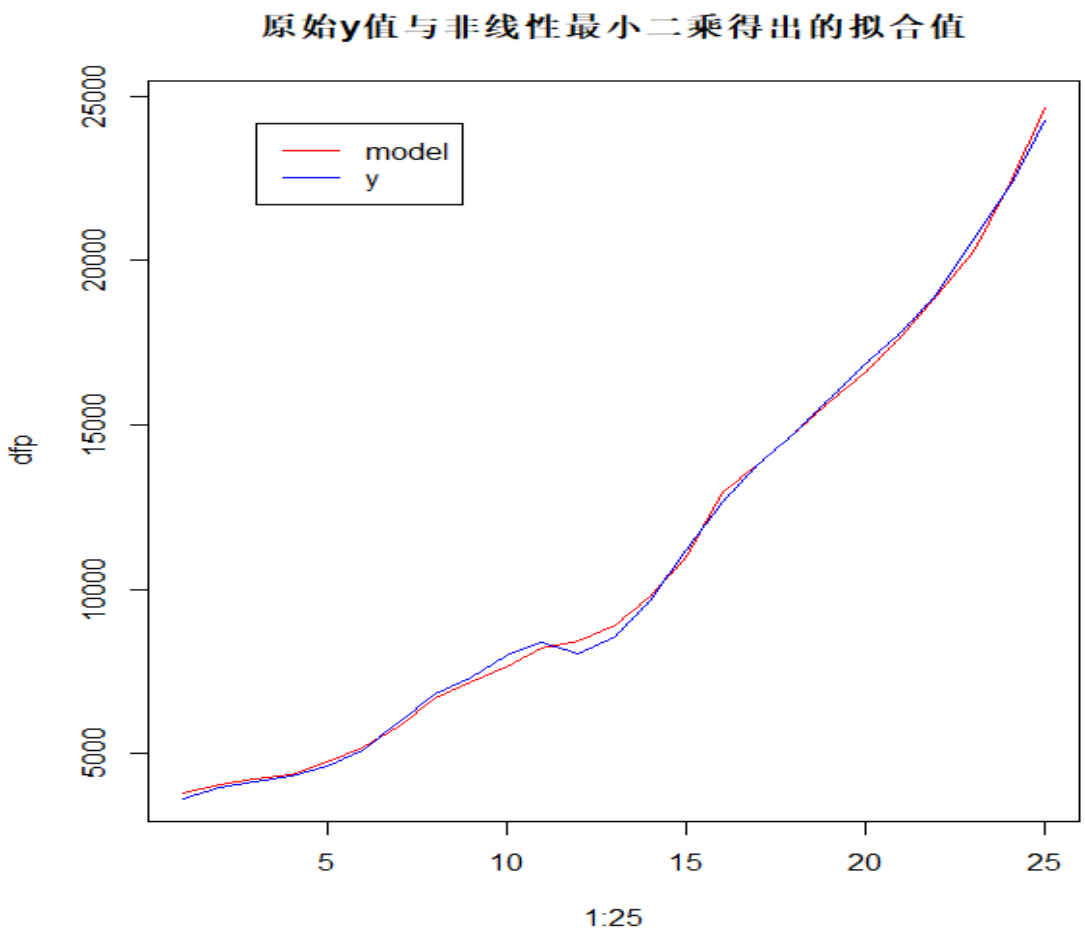
```
model=nls(y~A*exp(u*t)*(K^a)*(L^b),start=list(A=3.17,u=0.03,a=0.44,b=0.16))
model
summary(model)
par(mfrow=c(1,1))#设置画面布局
dfp=predict(model,interval="prediction")
dfp
plot(1:25,dfp,type = 'l',col='red',main = '原始 y 值与非线性最小二乘得出的拟合值')
lines(1:25,y,type='l',col='blue')
legend(locator(1),c("model","y"),lty=c(1,1),col=c("red","blue"))
```

2.2 结果及分析

```
> model=nls(y~A*exp(u*t)*(K^a)*(L^b),start=list(A=3.17,u=0.03,a=0.44,b=0.16))
> model
Nonlinear regression model
  model: y ~ A * exp(u * t) * (K^a) * (L^b)
   data: parent.frame()
        A         u         a         b
2.024e+02 4.577e-02 3.951e-01 2.459e-03
 residual sum-of-squares: 1003549

Number of iterations to convergence: 8
Achieved convergence tolerance: 1.475e-06
```

```
> summary(model)

Formula: y ~ A * exp(u * t) * (K^a) * (L^b)

Parameters:
   Estimate Std. Error t value Pr(>|t|)
A 2.024e+02  1.819e+02   1.112    0.279
u 4.577e-02  3.695e-03  12.388 4.04e-11 ***
a 3.951e-01  4.401e-02   8.978 1.24e-08 ***
b 2.459e-03  8.593e-02   0.029    0.977
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 218.6 on 21 degrees of freedom

Number of iterations to convergence: 8
Achieved convergence tolerance: 1.475e-06
```

```
> par(mfrow=c(1,1))#设置画面布局
> dfp=predict(model,interval="prediction")
> dfp
  [1]  3781.500  4035.756  4230.626  4376.300  4742.851  5186.759  5833.515
  [8]  6680.383  7173.979  7648.913  8225.264  8402.807  8887.605  9758.163
 [15] 10995.633 12986.265 13779.857 14753.953 15680.668 16610.505 17704.835
 [22] 18922.616 20255.320 22336.673 24623.831
> R_2=(cor(y,dfp))^2
> R_2
[1] 0.9989455
>
```



实验分析：从输出结果来看，使用非线性最小二乘法估计时，用乘性误差项的参数作为初值，经过数次迭代以后得到如下方程：

$$y = 202.4e^{0.04557t}K^{0.3951}L^{0.002455}$$

其中L的指数的参数估计值为 0.002455，其 P 值为略大于 0.05，不是很显著，因此不能认为 b 值非 0。从模型的拟合效果图来看预测值与观测值的线性基本一致，偏差较小，同时拟合优度达到了 0.99，说明非线性最小二乘得出的结果是非常合理的。

4. 比较乘性误差项模型与加性误差项模型的区别

乘性误差项模型和加性误差项模型所得的结果有一定的差异，其中乘性误差项

模型认为y_t本身是异方差的，而$\ln y_t$是等方差的。加性误差项模型认为y_t是等方差的。从统计性质看两者的差异，前者淡化了y_t值大的项的作用，强化了y_t值小的项的作用，对早期数据拟合的效果较好，而后者则对近期数据拟合的效果较好。

影响模型拟合效果的统计因素主要包括异方差、自相关、共线性这三个方面。异方差可以通过选择乘性误差项模型和加性误差项模型解决，必要时还可以使用加权最小二乘。时间序列数据通常都会存在自相关性。

乘性误差项模型的形式为：$y = AK^a L^b e^{ut} e^{\varepsilon}$，该模型可以通过两边取对数转化为线性模型，此题中通过取对数将其变为了多元线性方程：$y = \beta_0 + \mu t + \mathrm{a}x_1 + bx_2 + \varepsilon$，对线性方程回归得到乘性误差项的C-D生产函数。加性误差项模型的形式为$y = AK^{\alpha} L^{\beta} e^{ut} + \varepsilon$，不能将其转化为线性形式，只能用非线性最小二乘法求解未知参数，加性误差项模型需要用到乘性误差项的参数作为初值做非线性最小二乘回归。

乘性误差项在数学处理上具有一定的优势，但是在实际应用中，与之相适应的经济现象很少；加性误差项处理上复杂一些，但实际应用广泛。

教师签字_____

第三部分　结果与讨论（可加页）

本文通过构建乘性误差项模型，得到的回归方程为：

$$\ln y = 5.155 + 0.46\ln K - 0.027\ln L + 0.0418t$$

通过分析，发现lnL的回归系数为负，而劳动力投入与产出应是正向关系，与实际经济意义不符。该模型虽然拟合程度高，决定系数和修正决定系数都达到了 0.99，但存在回归系数不显著的情况，因此本文对模型进行了诊断，然后在此基础上针对多重共线性和自相关性对模型进行了改进，最终得到回归方程为

$$\ln y = 3.1703 + 0.4466\ln K + 0.1692\ln L + 0.0372t$$

此时，回归方程中各项系数均显著，且回归系数的符号与实际经济意义相一致，且拟合优度达到了 0.93。

本文对所用进行了基本假设的检验，发现数据不存在异方差性，且符合正态性假设，但是存在自相关性，所以本文通过一阶差分法最终解决了自相关性这个问题。

对于误差项模型进行建模，本文得到的回归方程为：

$$y = 202.4e^{0.04557t}K^{0.3951}L^{0.002455}$$

其中L的指数的参数估计值为 0.002455，其 P 值为略大于 0.05，不是很显著，因此不能认为 b 值非 0；但是其余参数均通过了显著性检验，而且拟合优度达到了 0.99。从模型的拟合效果图来看预测值与观测值的线性基本一致，偏差较小，说明拟合效果良好。