学生学号		实验课成绩	

武漢理工大學

学 生 实 验 报 告 书

实验课程名称	违背基本假设的情况处理
开 课 学 院	理学院
指导教师姓名	李丹
学 生 姓 名	
学生专业班级	

20 -- 20 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：_违背基本假设的情况处理___

实验项目名称	违背基本假设的情况处理			实验成绩	
实 验 者		专业班级		组　　别	
同 组 者				实验日期	年　　月　　日

第一部分：数据：课本 P129 表 4-11，表 4-13

表 4-11

用户	X	Y	用户	X	Y
1	679	0.79	28	1748	4.88
2	292	0.44	29	1381	3.48
3	1012	0.56	30	1428	7.58
4	493	0.79	31	1255	2.63
5	582	2.70	32	1777	4.99
6	1156	3.64	33	370	0.59
7	997	4.73	34	2316	8.19
8	2189	9.50	35	1130	4.79
9	1097	5.34	36	463	0.51
10	2078	6.85	37	770	1.74
11	1818	5.84	38	724	4.10
12	1700	5.21	39	808	3.94
13	747	3.25	40	790	0.96
14	2030	4.43	41	783	3.29
15	1643	3.16	42	406	0.44
16	414	0.50	43	1242	3.24
17	354	0.17	44	658	2.14
18	1276	1.88	45	1746	5.71
19	745	0.77	46	468	0.64
20	435	1.39	47	1114	1.90
21	540	0.56	48	413	0.51
22	874	1.56	49	1787	8.33
23	1543	5.28	50	3560	14.94
24	1029	0.64	51	1495	5.11
25	710	4.00	52	2221	3.85
26	1434	0.31	53	1526	3.93
27	837	4.20			

表 4-13

周次	销售额 y	周演出场次 x_1,	周点击率 x_2,	周次	销售额 y	周演出场次 x_1,	周点击率 x_2,
1	893.93	5	292	27	668.3	4	173
2	1	5	252	28	915.03	5	360
3	1	5	267	29	565.92	4	340
4	1	5	379	30	1	5	380
5	997.24	5	318	31	930.24	6	285
6	1	6	393	32	379.38	4	232
7	1	5	331	33	500.74	5	294
8	747.24	4	204	34	83.65	5	220
9	866.43	5	266	35	982.94	6	391
10	603	5	253	36	722.28	4	279
11	343.52	5	315	37	1	5	322
12	472.1	6	271	38	1	4	231
13	171.79	4	166	39	1	6	368
14	135.79	4	204	40	1	5	357
15	925.95	5	335	41	767.64	5	260
16	1	5	352	42	1	5	298
17	1	5	274	43	1	5	350
18	971.27	4	333	44	1	6	320
19	1165.2	5	302	45	957.68	4	227
20	597.85	4	324	46	1	5	261
21	490.34	4	327	47	1	5	303
22	709.59	5	206	48	1	6	263
23	987.3	5	310	49	1	4	215
24	954.6	6	306	50	827.56	4	294
25	1	6	350	51	803.16	4	288
26	1	5	275	52	1	6	257

第二部分：实验要求

1、 用 R 软件完成下列的计算分析；
2、 用普通最小二乘法建立表 4-11 中的 x 与 y 之间的回归方程，画出残差散点图；
3、 诊断该数据是否存在异方差；如果存在异方差，用加权最小二乘回归消除异方差的影响。
4、 用普通最小二乘法建立表 4-13 中的 y 与 x1,x2 之间的回归方程，用残差图及 DW 检验诊断序列的自相关性；
5、 如果存在自相关性，用迭代法处理这种序列相关，并建立回归方程；比较 3 和 4 中建立的回归方程的优良性。

第二部分：实验过程记录

过程记录（包括操作的步骤或者代码，输出的结果或者图形）：

第一题

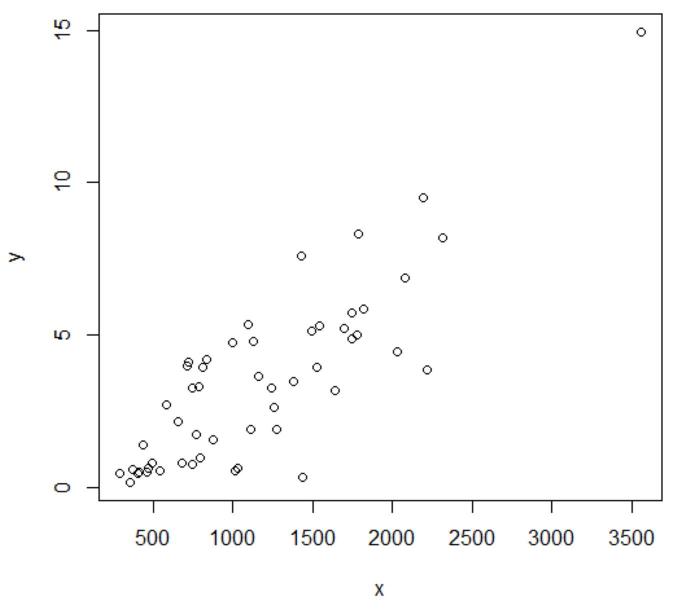
一、 用普通最小二乘法建立 x 与 y 之间的回归方程

1. 画出散点图，确立模型

1) 编程脚本

```
#1.读入数据
x<-scan(file="411x.txt");x
y<-scan(file="411y.txt");y
#2.画出散点图
plot(x,y)
```

2) 运行结果



3) 结果分析

根据散点图，可看出 x 和 y 大致呈线性关系，因此考虑一元线性回归模型。

2. 模型的参数估计

用最小二乘法对模型的参数进行估计，对每一个样本观测值 (x_i, y_i)，最小二乘法考虑观测值 y_i 与其回归值 $E(y_i)$ 的离差越小越好，综合考虑 n 个离差值，定义离差平方和为：

$$Q(\beta_0,\beta_1)=\sum_{i=1}^{n}(y_i-E(y_i))^2=\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2$$

所谓最小二乘法，就是寻找参数 β_0,β_1 的估计值 $\hat{\beta}_0,\hat{\beta}_1$ ，满足：

$$Q(\hat{\beta}_0,\hat{\beta}_1)=\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i))^2=\min\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2$$

求出的 $\hat{\beta}_0,\hat{\beta}_1$ 就称为回归参数的最小二乘估计。

1) 编程脚本

```
#1.读入数据
x<-scan(file="411x.txt");x
y<-scan(file="411y.txt");y
#2.得到回归方程 并查看回归结果
CG<-lm(y~x)
CG
summary(CG)
```

2) 运行结果

```
> x<-scan(file="411x.txt");x
Read 53 items
 [1]  679  292 1012  493  582 1156  997 2189 1097 2078 1818 1700  747 2030 1643
[16]  414  354 1276  745  435  540  874 1543 1029  710 1434  837 1748 1381 1428
[31] 1255 1777  370 2316 1130  463  770  724  808  790  783  406 1242  658 1746
[46]  468 1114  413 1787 3560 1495 2221 1526
> y<-scan(file="411y.txt");y
Read 53 items
 [1]  0.79  0.44  0.56  0.79  2.70  3.64  4.73  9.50  5.34  6.85  5.84  5.21  3.25
[14]  4.43  3.16  0.50  0.17  1.88  0.77  1.39  0.56  1.56  5.28  0.64  4.00  0.31
[27]  4.20  4.88  3.48  7.58  2.63  4.99  0.59  8.19  4.79  0.51  1.74  4.10  3.94
[40]  0.96  3.29  0.44  3.24  2.14  5.71  0.64  1.90  0.51  8.33 14.94  5.11  3.85
[53]  3.93
> #2.得到回归方程 并查看回归结果
> CG<-lm(y~x)
> CG

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
  -0.831304     0.003683

> summary(CG)

Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-4.1399 -0.8275 -0.1934  1.2376  3.1522

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.8313037  0.4416121  -1.882   0.0655 .
x            0.0036828  0.0003339  11.030 4.11e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.577 on 51 degrees of freedom
Multiple R-squared:  0.7046,    Adjusted R-squared:  0.6988
F-statistic: 121.7 on 1 and 51 DF,  p-value: 4.106e-15
```

3) 结果分析

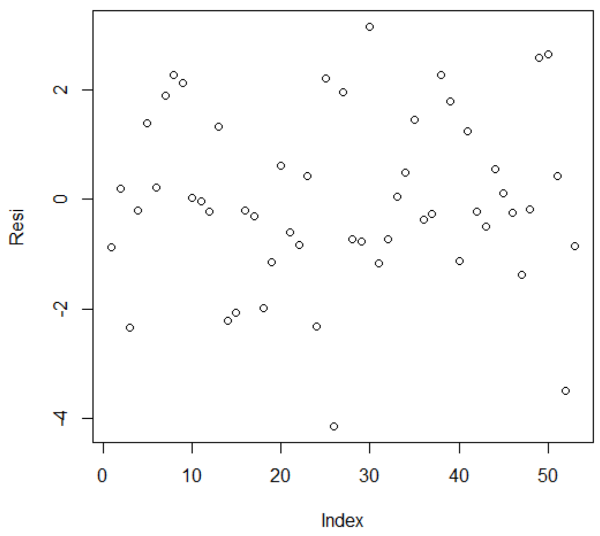
由输出结果可知，所得回归方程为 y=0.003683x-0.831304。

二、 画出残差散点图

1) 编程脚本

```
#1.读入数据
x<-scan(file="411x.txt");x
y<-scan(file="411y.txt");y
#2.得到回归方程 并查看回归结果
CG<-lm(y~x)
CG
summary(CG)
#3.计算残差 并作残差图
Resi<-residuals(CG)
Resi
plot(Resi)
```

2) 运行结果



三、 诊断该数据是否存在异方差

a) 残差图分析法

由上面得到的残差图，残差分布看起来没有一定的趋势，分布较为随机，初步判定不存在异方差。残差图虽然直观但是不够精确，检验的结果不够可靠，以下考虑用等级相关系数法。

b) 等级相关系数法

进行等级相关系数检验通常有三个步骤：

Step1：作 y 关于 x 的普通最小二乘回归，求出 ε_i 的估计值，即 e_i 的值

Step2：取 e_i 的绝对值，即 $|e_i|$，把 x_i 和 $|e_i|$ 按递增或递减的次序排列后分成等级，计算出等级相关系数 r_s

$$r_s = 1 - \frac{6}{n(n^2-1)}\sum_{i=1}^{n} d_i^2$$

Step3：做等级相关系数的显著性检验，在 n>8 的情况下，检验统计量为

$$t = \frac{\sqrt{n-2r_s}}{\sqrt{1-r_s^2}}$$

如果$|t| \le t_{\alpha/2}(n-2)$，可以认为异方差性问题不存在；如果$|t| > t_{\alpha/2}(n-2)$，异方差性问题存在。

1) 编程脚本

```
#1.读入数据
x<-scan(file="411x.txt");
y<-scan(file="411y.txt");
#2.计算残差
Resi<-residuals(CG)
Resi<-abs(Resi)
Resi
#3.计算等级相关系数
a=rank(Resi);a
b=rank(x);b
c=(a-b)^2;c
n=length(x);n
rs=1-6/(n*(n^2-1))*sum(c);rs
#4.计算t检验统计量及临界值
t=sqrt(n-2)*rs/(sqrt(1-rs^2));t
t_alpha1=qt(0.025,51,lower.tail=F);t_alpha1
```

2) 程序运行

```
> #1.读入数据
> x<-scan(file="411x.txt");
Read 53 items
> y<-scan(file="411y.txt");
Read 53 items
> #2.计算残差
> Resi<-residuals(CG)
> Resi<-abs(Resi)
> #3.计算等级相关系数
> a=rank(Resi);
> b=rank(x);
> c=(a-b)^2;
> n=length(x);n
[1] 53
> rs=1-6/(n*(n^2-1))*sum(c);rs
[1] 0.3175294
> #4.计算t检验统计量及临界值
> t=sqrt(n-2)*rs/(sqrt(1-rs^2));t
[1] 2.391371
> t_alpha1=qt(0.025,51,lower.tail=F);t_alpha1
[1] 2.007584
```

3) 结果分析

由输出结果可知，$r_s = 0.3175294$，得到的检验统计量值为$t = 2.391371$，比较该值与置信水平为 95%的t临界值，有$t = 2.391371 \ge t_{\alpha/2}(n-2) = 2.007584$，说明$x_i$与$|e_i|$之间存在系统关系，异方差性问题存在。

四、 用加权最小二乘回归消除异方差的影响

用一元加权最小二乘估计法消除异方差性，一元线性回归的加权最小二乘的离差平方和为：

$$Q_\omega(\beta_0,\beta_1)=\sum_{i=1}^{n}\omega_i(y_i-E(y_i))^2=\sum_{i=1}^{n}\omega_i(y_i-\beta_0-\beta_1x_i)^2$$

式中，ω_i 为给定的第 i 个观测值的权数。加权最小二乘估计就是寻找参数 β_0,β_1 的估计值 $\hat{\beta}_{0w},\hat{\beta}_{1w}$ 使上式的离差平方和最小。

取 X 构造权重矩阵，此时权函数为：$\omega_i=\dfrac{1}{x_i^{\ m}}$，为了找到消除异方差效果较好的权重，这里将 m 从 2 以步长-0.5 变化到-2，根据得到的加权最小二乘回归方程的 R^2 和调整后的 R^2 大小来选择这些模型中效果较好的一个。

1. 选择较优权重

1) 编程脚本

```
#1.读入数据
x<-scan(file="411x.txt");x
y<-scan(file="411y.txt");y
#2.得到加权回归方程 并查看回归结果
CG<-lm(y~x,weights=1/x)#权重为x值的倒数
CG
summary(CG)
```

加的权重与自变量的幂函数成比例，幂的大小从 2 到-2，以-0.5 的步长变化

2) 运行过程

对幂为-1 的权重，输出结果如下：

```
> #1.读入数据
> x<-scan(file="411x.txt");
Read 53 items
> y<-scan(file="411y.txt");
Read 53 items
> #2.得到加权回归方程 并查看回归结果
> CG<-lm(y~x,weights=1/x)#权重为x值的倒数
> CG

Call:
lm(formula = y ~ x, weights = 1/x)

Coefficients:
(Intercept)            x
  -0.664243     0.003538

> summary(CG)

Call:
lm(formula = y ~ x, weights = 1/x)

Weighted Residuals:
     Min       1Q   Median       3Q      Max
-0.10825 -0.02936 -0.01153  0.04231  0.08447
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.6642427  0.3369066  -1.972   0.0541 .
x            0.0035379  0.0003447  10.264 5.26e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04514 on 51 degrees of freedom
Multiple R-squared:  0.6738,    Adjusted R-squared:  0.6674
F-statistic: 105.4 on 1 and 51 DF,  p-value: 5.264e-14
```

其他幂函数下的输出结果不再一一呈现。

3) 结果分析

由输出结果可知，当幂为-1 时，加权最小二乘回归方程的 $R^2=0.6738$，调整 $R^2=0.6674$，对幂的大小从 2 到-2 依次进行回归，得到各自的 R^2 和调整后的 R^2 如下表：

m	2	1.5	1	0.5	0	-0.5	-1	-1.5	-2
R^2	0.6436	0.6591	0.6738	0.6883	0.7046	0.7257	0.7531	0.785	0.8175
调整 R^2	0.6366	0.6524	0.6674	0.6822	0.6988	0.7204	0.7483	0.7808	0.8139

由此可知，选 $m=-2$ 时模型的拟合程度最优，因此确定权函数为 $\omega_i = x_i^{\ 2}$

2. 验证是否消除异方差

1) 编程脚本

```
#1.读入数据
x<-scan(file="411x.txt");
y<-scan(file="411y.txt");
#2.得到加权回归方程  并查看回归结果
CG<-lm(y~x,weights=x^2)#权重为x值的倒数
#3.计算残差
Resi<-residuals(CG)
Resi<-abs(Resi)
#4.计算等级相关系数
a=rank(Resi);
b=rank(x);
c=(a-b)^2;
n=length(x);n
rs=1-6/(n*(n^2-1))*sum(c);rs
#4.计算t检验统计量及临界值
t=sqrt(n-2)*rs/(sqrt(1-rs^2));t
t_alpha1=qt(0.025,51,lower.tail=F);t_alpha1
```

2) 运行结果

```
> n=length(x);n
[1] 53
> rs=1-6/(n*(n^2-1))*sum(c);rs
[1] -0.1011127
> #4.计算t检验统计量及临界值
> t=sqrt(n-2)*rs/(sqrt(1-rs^2));t
[1] -0.7258091
> t_alpha1=qt(0.025,51,lower.tail=F);t_alpha1
[1] 2.007584
```

3) 结果分析

由输出结果可知，$r_s=-0.1011127$，得到的检验统计量值为 $t=-0.7258091$，比较该值与

置信水平为 95%的 t 临界值，有 $|t| = 0.7258091 \le t_{\alpha/2}(n-2) = 2.007584$，说明异方差问题不 存在，加权最小二乘估计大大减弱了异方差的影响。

第二题

一、 用普通最小二乘法建立表 4-13 中的 y 与 x1,x2 之间的回归方程

被解释变量 y 与解释变量 x1,x2 均有关，因此建立的是多元回归方程，首先考虑最简单的多元线性回归方程。

1) 编程脚本

未命名 - R编辑器

```
y=scan(file="413y.txt")
x1=scan(file="413x1.txt")
x2=scan(file="413x2.txt")
lm.test<-lm(y~x1+x2)
summary(lm.test)
```

2) 运行过程

```
> y=scan(file="413y.txt")
Read 52 items
> x1=scan(file="413x1.txt")
Read 52 items
> x2=scan(file="413x2.txt")
Read 52 items
> lm.test<-lm(y~x1+x2)
> summary(lm.test)

Call:
lm(formula = y ~ x1 + x2)

Residuals:
    Min      1Q  Median      3Q     Max 
-747.71 -229.80   -2.15  267.23  547.68 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)  
(Intercept) -574.0624   349.2707  -1.644   0.1067  
x1           191.0985    73.3092   2.607   0.0121 *
x2             2.0451     0.9107   2.246   0.0293 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 329.7 on 49 degrees of freedom
Multiple R-squared:  0.2928,    Adjusted R-squared:  0.264 
F-statistic: 10.15 on 2 and 49 DF,  p-value: 0.0002057
```

3) 结果分析

由输出结果可知，所得回归方程为 y=191.0985x1+2.0451x2-574.0624

二、 诊断序列的自相关性

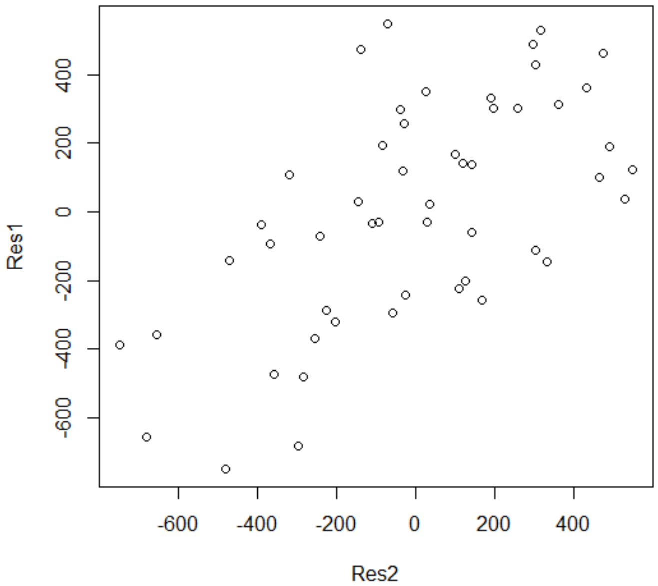
a) 图示检验法

1、 绘制 e_t, e_{t-1} 的散点图

1) 编程脚本

```
y=scan(file="413y.txt")
x1=scan(file="413x1.txt")
x2=scan(file="413x2.txt")
lm.test<-lm(y~x1+x2)
Resi<-residuals(lm.test)
Res1=Resi[-1]
Res2=Resi[-length(Resi)]
plot(Res2,Res1)
```

2) 运行结果



3) 结果分析

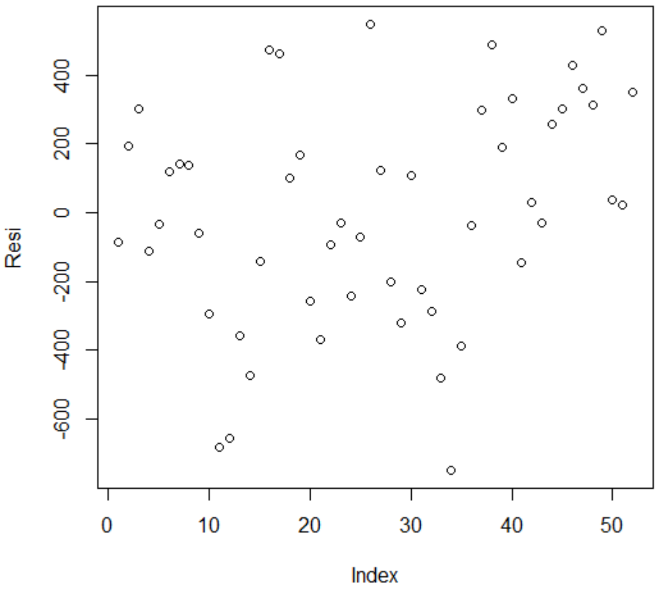
如图所示，大部分点落在第一、三象限，表明随机扰动项存在正的序列相关。

2、 按照时间顺序绘制回归残差项的图形

1) 编程脚本

```
y=scan(file="413y.txt")
x1=scan(file="413x1.txt")
x2=scan(file="413x2.txt")
lm.test<-lm(y~x1+x2)
Resi<-residuals(lm.test)
plot(Resi)
```

2) 运行结果



3)　结果分析

e_t 随着 t 的变化逐次有规律地呈现循环形状的变化，表明随机扰动项存在序列相关。

b） DW 检验

1)　编程脚本

D:\Documents\R\43.R - R编辑器

```
y=scan(file="413y.txt")
x1=scan(file="413x1.txt")
x2=scan(file="413x2.txt")
lm.test<-lm(y~x1+x2)
library(lmtest)
dwtest(lm.test)
```

2)　运行结果

```
> dwtest(lm.test)

        Durbin-Watson test

data:  lm.test
DW = 0.74526, p-value = 1.758e-07
alternative hypothesis: true autocorrelation is greater than 0
```

3)　结果分析

由以上输出可得到，$DW = 0.74526 \in (0,2)$，说明误差项呈正自相关。

一、　用迭代法处理这种序列相关，并建立回归方程

根据回归方程式，有 $y_{t-1} = \beta_0 + \beta_1 x_{t-1} + \varepsilon_{t-1}$，将该式两端乘以 ρ，用 $y_t - \rho y_{t-1}$ 得：

$$(y_t - \rho y_{t-1}) = (\beta_0 - \rho\beta_0) + \beta_1(x_t - \rho x_{t-1}) + (\varepsilon_t - \rho\varepsilon_{t-1})$$

根据 $\hat{\rho} \approx 1 - \frac{1}{2} DW$ 计算出 ρ 的估计值。

DW 与 $\hat{\rho}$ 的对应关系如下表

$\hat{\rho}$	DW	误差项的自相关性

-1	4	完全负相关
（-1,0）	(2，4)	负自相关
0	2	无自相关
（0,1）	(0，2)	正自相关
1	0	完全正自相关

1) 编程脚本

第一次迭代：

```
y=scan(file="413y.txt")
x1=scan(file="413x1.txt")
x2=scan(file="413x2.txt")
rou<-1-0.74526/2;rou
newy<-y[2:length(y)]-rou*y[1:length(y)-1]
newx1<-x1[2:length(y)]-rou*x1[1:length(y)-1]
newx2<-x2[2:length(y)]-rou*x2[1:length(y)-1]
lm.test<-lm(newy~newx1+newx2)
summary(lm.test)
library(lmtest)
dwtest(lm.test)
```

第 $k(k \geq 2)$ 次迭代：

```
y=newy
x1=newx1
x2=newx2
```

Rou 根据第 $k-1$次迭代后求出的 DW 值计算,其他代码与第一次迭代相同。

2) 运行过程

第一次迭代输出的结果如下：

```
> y=scan(file="413y.txt")
Read 52 items
> x1=scan(file="413x1.txt")
Read 52 items
> x2=scan(file="413x2.txt")
Read 52 items
> rou<-1-0.74526/2;rou
[1] 0.62737
> newy<-y[2:length(y)]-rou*y[1:length(y)-1]
> newx1<-x1[2:length(y)]-rou*x1[1:length(y)-1]
> newx2<-x2[2:length(y)]-rou*x2[1:length(y)-1]
> lm.test<-lm(newy~newx1+newx2)
> summary(lm.test)

Call:
lm(formula = newy ~ newx1 + newx2)

Residuals:
    Min       1Q  Median       3Q      Max
-496.06 -228.25   40.02   180.34   574.87

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -178.8440    90.3694  -1.979   0.0536 .
newx1        211.1096    47.7502   4.421  5.6e-05 ***
newx2          1.4365     0.6287   2.285   0.0268 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 257.9 on 48 degrees of freedom
Multiple R-squared:  0.4731,    Adjusted R-squared:  0.4511
F-statistic: 21.55 on 2 and 48 DF,  p-value: 2.099e-07

> library(lmtest)
> dwtest(lm.test)

        Durbin-Watson test

data:  lm.test
DW = 1.7162, p-value = 0.1806
alternative hypothesis: true autocorrelation is greater than 0
```

第一次迭代后的 $DW = 1.7162$， 仍然在 $(0,2)$ 之间，序列误差项仍成正相关，还需要继续进行迭代。

第二次迭代输出的结果：

```
> y=newy
> x1=newx1
> x2=newx2
> rou<-1-1.7162/2;rou
[1] 0.1419
> newy<-y[2:length(y)]-rou*y[1:length(y)-1]
> newx1<-x1[2:length(y)]-rou*x1[1:length(y)-1]
> newx2<-x2[2:length(y)]-rou*x2[1:length(y)-1]
> lm.test<-lm(newy~newx1+newx2)
> summary(lm.test)

Call:
lm(formula = newy ~ newx1 + newx2)

Residuals:
    Min      1Q  Median      3Q     Max 
-442.63 -191.12   47.06  183.15  558.82 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -158.342     75.789  -2.089   0.0421 *  
newx1        205.934     44.610   4.616 3.04e-05 ***
newx2          1.538      0.596   2.581   0.0130 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 256.5 on 47 degrees of freedom
Multiple R-squared:  0.5149,    Adjusted R-squared:  0.4942 
F-statistic: 24.94 on 2 and 47 DF,  p-value: 4.148e-08

> library(lmtest)
> dwtest(lm.test)

        Durbin-Watson test

data:  lm.test
DW = 2.0002, p-value = 0.5524
alternative hypothesis: true autocorrelation is greater than 0
```

第二次迭代后的 $DW = 2.0002 \approx 2$，序列误差项无自相关，不需要继续进行迭代，迭代法处理序列相关完成。

教师签字_______

第三部分　结果与讨论（可加页）

第一题

1、 用普通最小二乘法建立的回归方程分析

summary(CG)后的输出结果为

```
> summary(CG)

Call:
lm(formula = y ~ 1 + x)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.1399 -0.8275 -0.1934  1.2376  3.1522 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.8313037  0.4416121  -1.882   0.0655 .  
x            0.0036828  0.0003339  11.030 4.11e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.577 on 51 degrees of freedom
Multiple R-squared:  0.7046,    Adjusted R-squared:  0.6988 
F-statistic: 121.7 on 1 and 51 DF,  p-value: 4.106e-15
```

由输出结果可知，所得回归方程为 y=0.003683x-0.831304。

Residuals 部分列出了残差的最小值点、四分之一分位点、中位数点、四分之三分位点和最大值点，分别为-4.1399，-0.8275，-0.1934，1.2376，3.1522；

Coefficients 部分中 Estimate 是回归方程参数的估计值，Std. Error 表示回归参数的标准差，t value 即为 t 值，Pr(>|t|) 即为 p 值，由一次项系数的 p 值=4.11e-15<0.05 可知，该系数通过了显著性检验，而常数项的 p 值=0.0655>0.05,未通过显著性检验。

Multiple R-squared 为 0.7064，Adjusted R-squared 为 0.6988，F 检验统计量为 121.7，对应的 p 值 4.106e-15<0.05,落在拒绝域内，说明回归方程显著。

综上可得，第一问中建立的一元线性回归方程通过了回归系数和回归方程的显著性检验，是适用的。

2、 用加权最小二乘法建立的回归方程分析

根据得到的加权最小二乘回归方程的 R^2 和调整后的 R^2 大小来选择确权函数为 $\omega_i = x_i^2$

提取该加权最小二乘回归方程的信息如下：

```
> summary(CG)

Call:
lm(formula = y ~ x, weights = x^2)

Weighted Residuals:
    Min      1Q  Median      3Q     Max 
-8794.6  -628.3   395.4  1411.3  4877.7 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -2.4003802  0.6218314   -3.86  0.00032 ***
x            0.0045971  0.0003041   15.12  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2377 on 51 degrees of freedom
Multiple R-squared:  0.8175,    Adjusted R-squared:  0.8139 
F-statistic: 228.5 on 1 and 51 DF,  p-value: < 2.2e-16
```

由输出结果可知，所得回归方程为 y=0.0045971x-2.4003802。

对于回归系数的显著性检验，常数项的 p 值=0.00032<0.05,x 的一次项系数的 p 值=2e-16<0.05,两个回归系数均通过了显著性检验；

对于回归方程的显著性检验，Multiple R-squared 为 0.8175，Adjusted R-squared 为 0.8139，F 检验统计量为 225.8，对应的 p 值 2.2e-16<0.05,落在拒绝域内，说明回归方程显著。

3、 比较两种回归方程

对普通最小二乘和加权最小二乘得到的回归方程进行比较，包括是否通过相应的显著性检验以及是否存在异方差，比较结果列于下表

	R^2	调整 R^2	方程	β_0	β_1	异方差
普通	0.7064	0.6988	是	否	是	存在
加权	0.8175	0.8139	是	是	是	不存在

由此可知，加权最小二乘不仅解决了异方差问题，在方程的拟合程度和系数的显著性检验方面也有了提高和改善，使回归的整体效果更好了。

第二题

1、 用普通最小二乘法建立的回归方程分析

该回归方程的信息如下：

```
Residuals:
    Min      1Q  Median      3Q     Max
-747.71 -229.80   -2.15  267.23  547.68

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -574.0624   349.2707  -1.644   0.1067
x1           191.0985    73.3092   2.607   0.0121 *
x2             2.0451     0.9107   2.246   0.0293 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 329.7 on 49 degrees of freedom
Multiple R-squared:  0.2928,    Adjusted R-squared:  0.264
F-statistic: 10.15 on 2 and 49 DF,  p-value: 0.0002057
```

由输出结果可知，所得回归方程为 y=191.0985x1+2.0451x2-574.0624。

对于回归系数的显著性检验，x1 的一次项系数的 p 值=0.0121<0.05,x2 的一次项系数的 p 值=0.0293<0.05,两个回归系数均通过了显著性检验，而常数项系数的 p 值=0.1607>0.05,未通过显著性检验。

对于回归方程的显著性检验，Multiple R-squared 为 0.2928，Adjusted R-squared 为 0.264，F 检验统计量为 10.15，对应的 p 值 0.0002057<0.05,落在拒绝域内，说明回归方程显著。

2、 迭代法处理自相关后建立的回归方程分析

该回归方程的信息如下：

```
Call:
lm(formula = newy ~ newx1 + newx2)

Residuals:
    Min      1Q  Median      3Q     Max
-442.63 -191.12   47.06  183.15  558.82

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -158.342     75.789  -2.089   0.0421 *
newx1        205.934     44.610   4.616 3.04e-05 ***
newx2          1.538      0.596   2.581   0.0130 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 256.5 on 47 degrees of freedom
Multiple R-squared:  0.5149,    Adjusted R-squared:  0.4942
F-statistic: 24.94 on 2 and 47 DF,  p-value: 4.148e-08
```

由输出结果可知，所得回归方程为 y=205.934x1+1.538x2-158.342。

对于回归系数的显著性检验，x1 的一次项系数的 p 值=3.04e-05<0.05,x2 的一次项系数的 p 值=0.0130<0.05,常数项系数的 p 值=0.0421<0.05, 三个回归系数均通过了显著性检验。

对于回归方程的显著性检验，Multiple R-squared 为 0.5149，Adjusted R-squared 为 0.4942，F 检验统计量为 24.92，对应的 p 值 4.148e-08<0.05,落在拒绝域内，说明回归方程显著。

3、 比较两种回归方程

对普通最小二乘和迭代法得到的回归方程进行比较，包括是否通过相应的显著性检验以及是否存在自相关性，比较结果列于下表

	R^2	调整 R^2	方程	β_0	β_1	自相关性
普通	0.2928	0.264	是	否	是	存在
加权	0.5149	0.4942	是	是	是	不存在

迭代法处理后建立的回归方程 R-squared 和 Adjusted R-squared 均有了明显的提高，这说明该回归方程不仅解决了自相关性问题，在方程的拟合程度和系数的显著性检验方面也有了提高和改进。