

学生学号	0121906280724	实验课成绩	
------	---------------	-------	--

# 武汉理工大学

## 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	张逸敏
学生专业班级	统计 2001

2022 -- 2023 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	张逸敏	专业班级	统计 2001	组 别	
同 组 者	刘璇、马钟森、焦鼎云、李耀祖、危景熙			实验日期	2023 年 3 月 31 日

### 第一部分：数据

异方差数据：ca\_learning.txt

	n	Y	X
1	1	16	14
2	2	14	22
3	3	22	85
4	4	10	50
5	5	14	62
6	6	17	70
7	7	10	55
8	8	13	63
9	9	19	88
10	10	12	57
11	11	18	81
12	12	11	51

自相关数据：google\_stock

2/7/2005	196.03	3/30/2005	180.45	5/18/2005	239.16
2/8/2005	198.64	3/31/2005	180.51	5/19/2005	239.18
2/9/2005	191.58	4/1/2005	180.04	5/20/2005	241.61
2/10/2005	187.98	4/4/2005	185.29	5/23/2005	255.45
2/11/2005	187.40	4/5/2005	188.57	5/24/2005	256.00
2/14/2005	192.99	4/6/2005	189.22	5/25/2005	260.81
2/15/2005	195.23	4/7/2005	193.76	5/26/2005	259.20
2/16/2005	198.41	4/8/2005	192.05	5/27/2005	266.00
2/17/2005	197.90	4/11/2005	193.23	5/31/2005	277.27
2/18/2005	197.95	4/12/2005	193.96	6/1/2005	288.00
2/22/2005	191.37	4/13/2005	192.93	6/2/2005	287.90
2/23/2005	193.95	4/14/2005	191.45	6/3/2005	280.26
2/24/2005	188.89	4/15/2005	185.00	6/6/2005	290.94
2/25/2005	185.87	4/18/2005	186.97	6/7/2005	293.12
2/28/2005	187.99	4/19/2005	191.40	6/8/2005	279.56
3/1/2005	186.06	4/20/2005	198.10	6/9/2005	286.31
3/2/2005	185.18	4/21/2005	204.22	6/10/2005	282.50
3/3/2005	187.01	4/22/2005	215.81	6/13/2005	282.75
3/4/2005	185.90	4/25/2005	223.53	6/14/2005	278.35
3/7/2005	188.81	4/26/2005	218.75	6/15/2005	274.80
3/8/2005	185.20	4/27/2005	219.78	6/16/2005	277.44
3/9/2005	181.35	4/28/2005	219.45	6/17/2005	280.30
3/10/2005	179.98	4/29/2005	220.00	6/20/2005	286.70
3/11/2005	177.80	5/2/2005	222.29	6/21/2005	287.84
3/14/2005	174.99	5/3/2005	226.19	6/22/2005	289.30
3/15/2005	178.61	5/4/2005	228.50	6/23/2005	289.71
3/16/2005	175.60	5/5/2005	226.98	6/24/2005	297.25
3/17/2005	179.29	5/6/2005	228.02	6/27/2005	304.10

## 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模，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回归方程，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。

## 第二部分：实验过程记录

### 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

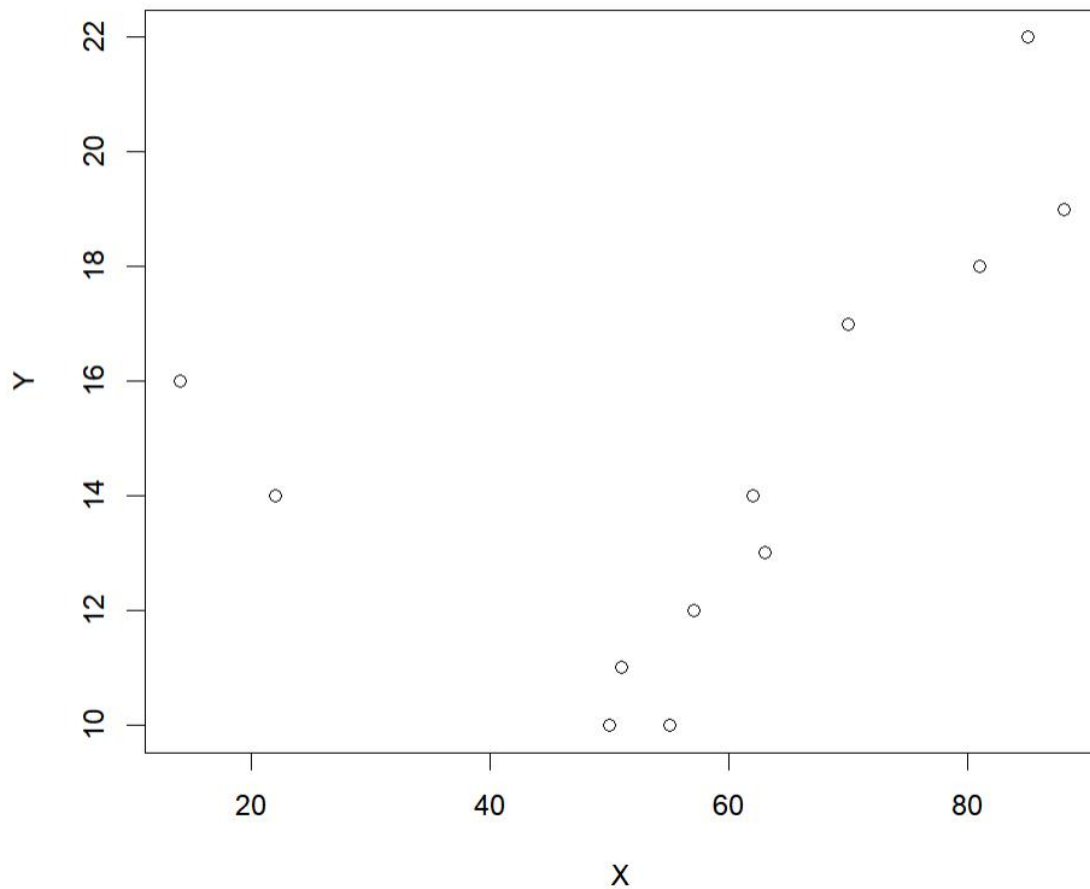
##### 1) 编程脚本

```

1 getwd()
2 setwd("D:/R/R2")
3 #读入数据
4 ca_learning <- read.delim("D:/R/R2/ca_learning.txt", header=FALSE)
5 names(ca_learning)<-c("n","Y","X")
6 #画出散点图，确立模型
7 attach(ca_learning)
8 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  $\beta_0, \beta_1$  的估计值  $\hat{\beta}_0, \hat{\beta}_1$ ，满足：

$$Q(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = \min \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$$

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

```
ca_learning <- read.delim("D:/R/R2/ca_learning.txt", header=FALSE)
names(ca_learning)<-c("n", "Y", "X")
#画出散点图，确立模型
attach(ca_learning)
plot(X,Y)
#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
#得到回归方程，查看回归结果
lm<-lm(Y~X)
lm
summary(lm)
```

### 2) 运行结果

```

> plot(X,Y)
> #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
> #得到回归方程, 查看回归结果
> lm<-lm(Y~X)
> lm

Call:
lm(formula = Y ~ X)

Coefficients:
(Intercept)          X
      9.9025         0.0819

> summary(lm)

Call:
lm(formula = Y ~ X)

Residuals:
    Min       1Q   Median       3Q      Max
-4.4073 -2.6983  0.1917  1.9913  5.1356

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.90253    2.85585   3.467  0.00605 **
X              0.08190    0.04598   1.781  0.10519
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.471 on 10 degrees of freedom
Multiple R-squared:  0.2409,    Adjusted R-squared:  0.165
F-statistic: 3.173 on 1 and 10 DF,  p-value: 0.1052

```

### 3) 结果分析

由输出结果可知, 所得回归方程为  $Y=0.0819X+9.9025$

## 二、画出残差散点图

### 1) 编程脚本

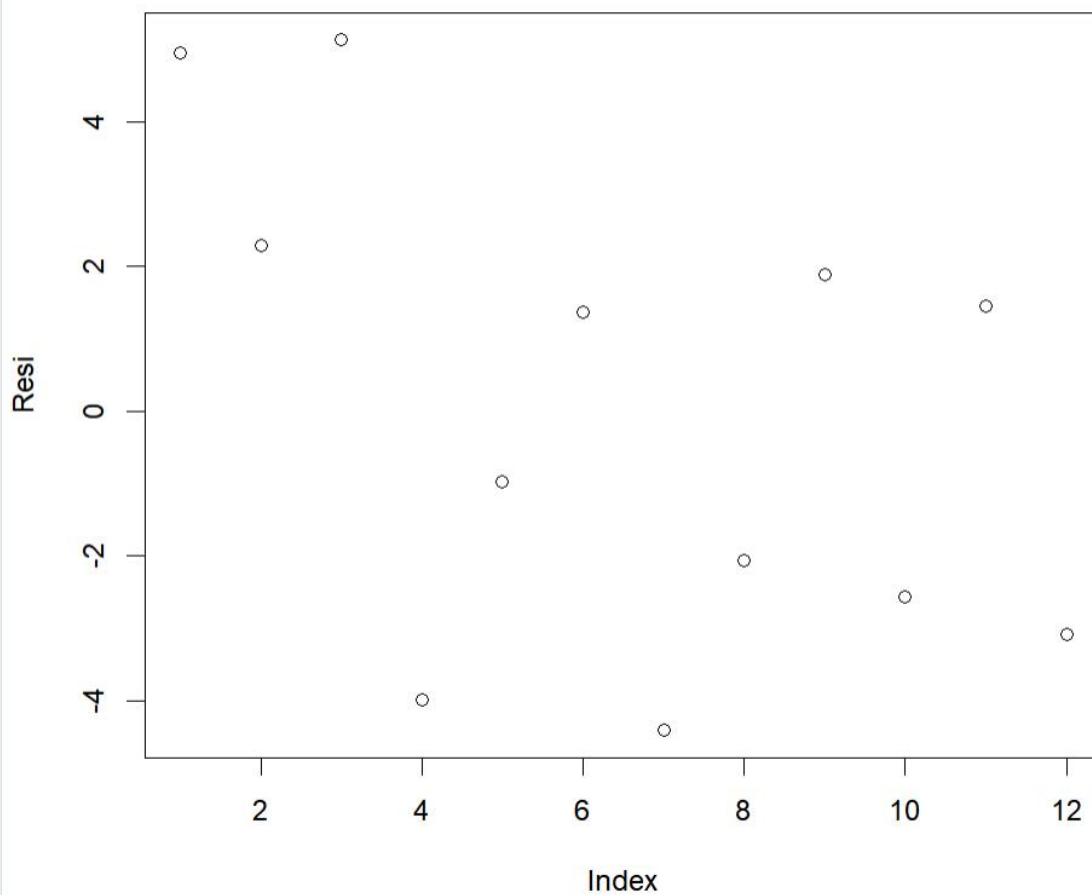
```
#得到回归方程，查看回归结果
lm<-lm(Y~X)
lm
summary(lm)
#计算残差，并作残差图
Resi<-residuals(lm)
Resi
plot(Resi)
```

## 2) 运行结果

```
> #计算残差，并作残差图
> Resi<-residuals(lm)
> Resi
```

1	2	3	4	5	6	7	8
4.9507986	2.2955596	5.1355526	-3.9977769	-0.9806353	1.3641257	-4.4073012	-2.0625402
9	10	11	12				
1.8898379	-2.5711110	1.4631721	-3.0796817				

```
> plot(Resi)
```





### 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  $\varepsilon_i$  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：

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

如果  $|t| \leq t_{\alpha/2}(n-2)$ ，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

```
#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
#得到回归方程，查看回归结果
lm<-lm(Y~X)
lm
summary(lm)
#计算残差，并作残差图
Resi<-residuals(lm)
Resi
plot(Resi)

#计算等级相关系数
a=rank(Resi);a
b=rank(X);b
c=(a-b)^2;c
m=length(X);m
rs=1-6/(m*(m*m-1))*sum(c);rs

#计算T检验统计量及临界值
t=sqrt(m-2)*rs/sqrt((1-rs^2));t
t_alpha1=qt(0.025,10,lower.tail=F);t_alpha1
```

## 2) 程序运行

```
> summary(lm)

Call:
lm(formula = Y ~ X)

Residuals:
    Min       1Q   Median       3Q      Max
-4.4073 -2.6983  0.1917  1.9913  5.1356

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.90253     2.85585   3.467  0.00605 **
X              0.08190     0.04598   1.781  0.10519
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.471 on 10 degrees of freedom
Multiple R-squared:  0.2409,    Adjusted R-squared:  0.165
F-statistic: 3.173 on 1 and 10 DF,  p-value: 0.1052

> #计算残差, 并作残差图
> Resi<-residuals(lm)
> Resi
      1      2      3      4      5      6      7      8
4.9507986 2.2955596 5.1355526 -3.9977769 -0.9806353 1.3641257 -4.4073012 -2.0625402
      9     10     11     12
1.8898379 -2.5711110 1.4631721 -3.0796817

> plot(Resi)
> #计算等级相关系数
> a=rank(Resi);a
 1  2  3  4  5  6  7  8  9 10 11 12
11 10 12 2  6  7  1  5  9  4  8  3
> b=rank(X);b
 [1]  1  2 11  3  7  9  5  8 12  6 10  4
> c=(a-b)^2;c
 1  2  3  4  5  6  7  8  9 10 11 12
100 64  1  1  1  4 16  9  9  4  4  1
> m=length(X);m
[1] 12
> rs=1-6/(m*(m-1))*sum(c);rs
[1] 0.2517483
> #计算T检验统计量及临界值
> t=sqrt(m-2)*rs/sqrt((1-rs^2));t
[1] 0.8225913
> t_alpha1=qt(0.025,10,lower.tail=F);t_alpha1
[1] 2.228139
```

## 3) 结果分析

由输出结果可知,  $r_s = 0.2517483$ , 得到的检验统计量的值  $t = 0.8225913$ , 比较该值与置信水平为 95% 的  $t$  临界值, 有  $t = 0.8225913 \leq t_{\alpha/2}(n-2) = 2.228139$ , 说明  $X_i$  与  $|e_i|$  之间不存在系统关系, 异方差性问题不存在。

## 第二题

### 一、用普通最小二乘法建立 google\_stock 中的 v1, v2 之间的回归方程

1、读入数据, 并将自变量设为 1 至样本个数

```

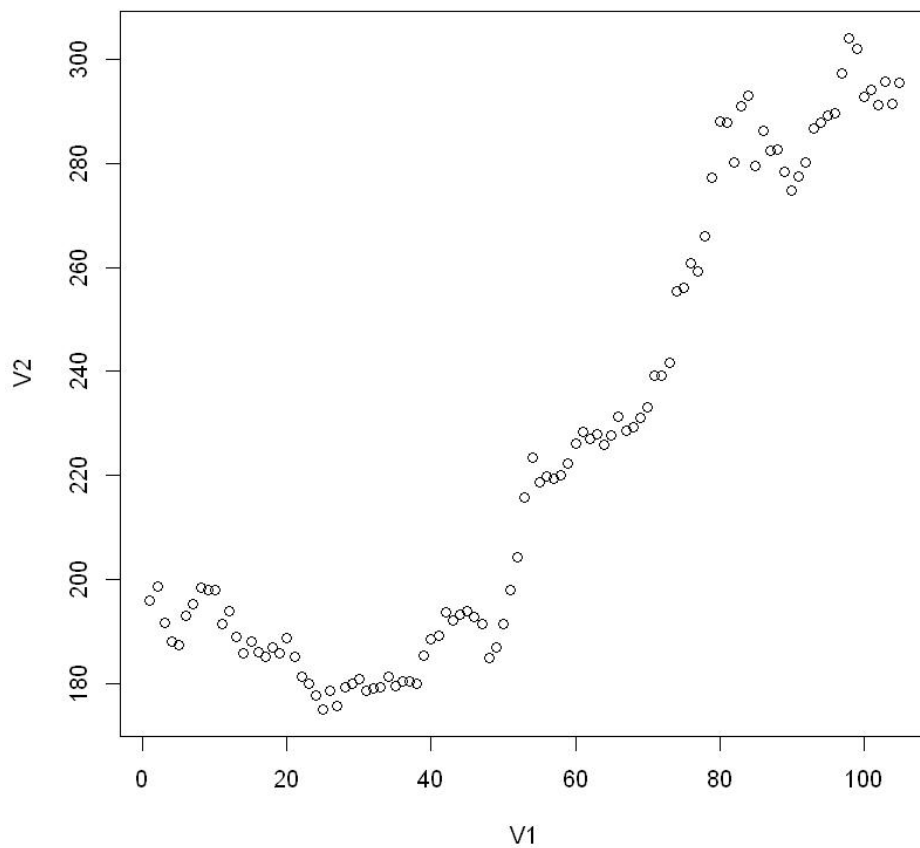
raw_data = read.table("./google_stock.txt")
google12 = raw_data[c(1,2)]
google34 = raw_data[c(3,4)]
google56 = raw_data[c(5,6)]
colnames(google34) = colnames(google12)
colnames(google56) = colnames(google12)
google = rbind(google12, google34, google56)
google$V1 = c(1: nrow(google))
google

A data.frame: 105 × 2      v1v2    <int><dbl>      1196.03    2198.64
3191.58    4187.98    5187.40    6192.99    7195.23    8198.41    9197.90
  10197.95  11191.37  12193.95  13188.89  14185.87  15187.99
  16186.06  17185.18  18187.01  19185.90  20188.81  21185.20
  22181.35  23179.98  24177.80  25174.99  26178.61  27175.60
  28179.29  29180.04  30180.88  ::: 76260.81 77259.20 78266.00
79277.27    80288.00  81287.90  82280.26  83290.94  84293.12
85279.56    86286.31  87282.50  88282.75  89278.35  90274.80
91277.44    92280.30  93286.70  94287.84  95289.30  96289.71
97297.25    98304.10  99302.00 100292.72 101294.15 102291.25
  103295.71 104291.52 105295.54

```

2、画出散点图

```
plot(google)
```



### 3、拟合一元线性回归模型

代码:

```
mdl0 = lm(V2~V1, data=google)
summary(mdl0)
```

运行结果:

Call:

```
lm(formula = v2 ~ v1, data = google)
```

Residuals:

Min	1Q	Median	3Q	Max
-33.297	-13.967	1.335	12.056	39.395

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	156.67805	3.52517	44.45	<2e-16 ***
V1	1.28372	0.05774	22.23	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17.93 on 103 degrees of freedom

Multiple R-squared: 0.8276, Adjusted R-squared: 0.8259

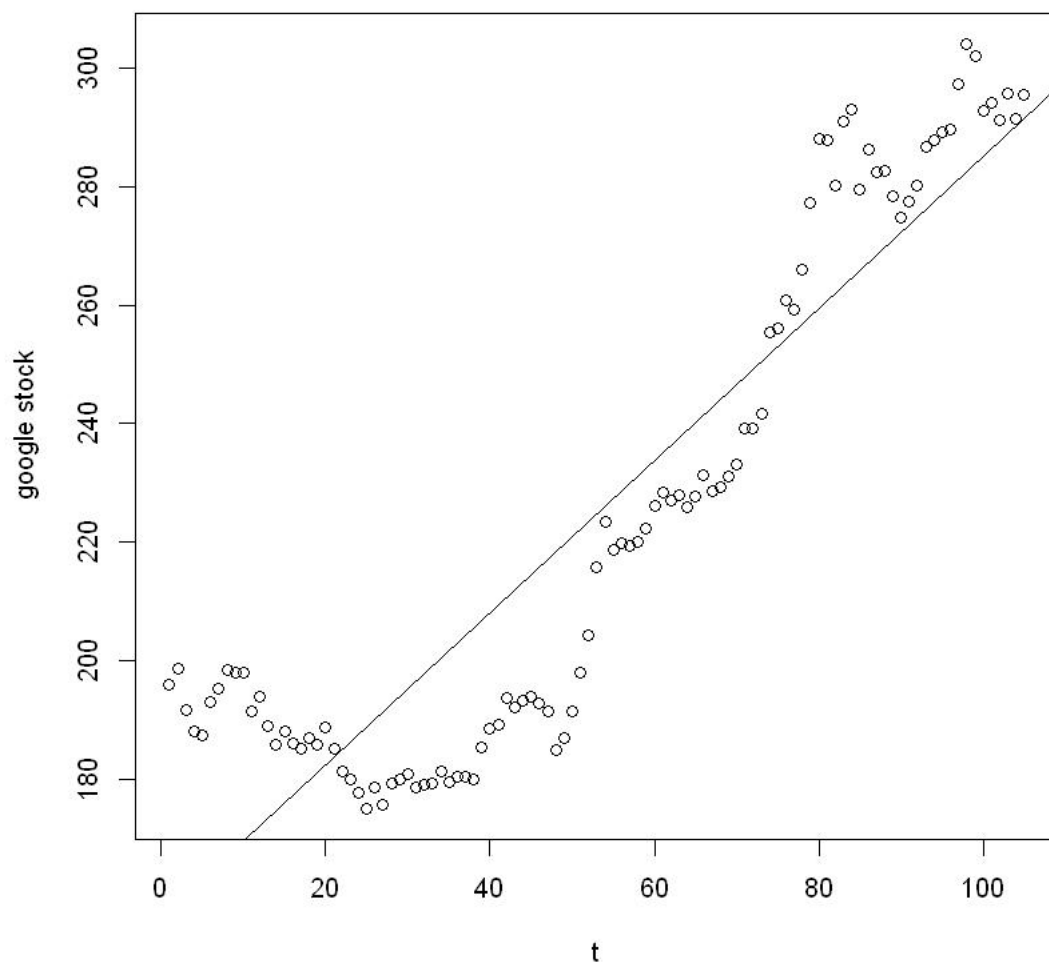
F-statistic: 494.3 on 1 and 103 DF, p-value: < 2.2e-16

4、画出拟合直线

代码:

```
plot(google$V1, google$V2,  
xlab = "t", ylab = "google stock")  
abline mdl0)
```

运行结果:



5、利用 Durbin-Watson 检验判断序列自相关性

代码:

```
library(lmtest)
dwtest(md10)
运行结果:
Durbin-Watson test
data: md10
DW = 0.061858, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is greater than 0
```

```
dw0 = as.numeric(dwtest(md10)$statistic) # 将 dw 统计量保存到变量 dw0 里,
计算迭代法时需要用来估计 rho
```

```
dw0
```

```
0.0618582452309325
```

## 6、DW 检验结果

$DW = 0.061858$

$p - value < 2.2e - 16$  存在正自相关

## 7、利用迭代法处理自相关

首先假设模型存在一阶自相关

需要做变换  $y_t' = y_t - \rho y_{t-1}$  和  $x_t' = x_t - \rho x_{t-1}$   
 $\rho$  值不知道, 用  $1-dw/2$  来估计

代码:

```
rho0 = 1-dw0/2 # 用 1-dw/2 来估计 rho
n0 = nrow(google) # 样本个数
y0 = google$V2
x0 = google$V1
y1 = y0[2:n0]-rho0*y0[1:n0-1] # 变换后的因变量, 共 n0-1 个
x1 = x0[2:n0]-rho0*x0[1:n0-1] # 变换后的自变量, 共 n0-1 个
mdl1 <- lm(y1~x1) # 用变换后的变量建立一元线性回归模型
summary(mdl1)
```

运行结果:

Call:

```
lm(formula = y1 ~ x1)
```

Residuals:

Min	1Q	Median	3Q	Max
-14.3720	-2.6945	-0.2969	2.4957	12.1304

Coefficients:

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.522      1.272   1.983  0.0501 .
x1              2.045      0.457   4.474  2e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Residual standard error: 4.327 on 102 degrees of freedom  
Multiple R-squared: 0.1641, Adjusted R-squared: 0.1559  
F-statistic: 20.02 on 1 and 102 DF, p-value: 1.998e-05

8、利用 DW 检验判断迭代一次后的模型的自相关性

```
dwtest(md11)
```

```
dw1 = dwtest(md11)$statistic
```

```
dw1
```

Durbin-Watson test

data: md11

DW = 1.8973, p-value = 0.2644

alternative hypothesis: true autocorrelation is greater than 0

DW: 1.89732294556596

$DW = 1.8973, p - value = 0.2644$

$p > 0.05$ , 没有自相关

迭代法结束

9、迭代一次后的模型估计值

$\beta_0' = 2.522, \beta_1' = 2.045, \hat{\rho} = 0.969$

根据变换前后模型系数的关系:  $\beta_0' = \beta_0(1 - \rho), \beta_1' = \beta_1$ , 可以计算原模型的系数估计值

$$b_0 = 2.522 / (1 - \rho_0) = 81.54$$

$$b_1 = 2.045$$

10、原一元线性回归方程为

$$y = 81.541 + 2.045x$$

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(lm)后的输出结果为

```
> summary(lm)

Call:
lm(formula = Y ~ X)

Residuals:
    Min       1Q   Median       3Q      Max
-4.4073 -2.6983  0.1917  1.9913  5.1356

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   9.90253     2.85585   3.467  0.00605 **
X              0.08190     0.04598   1.781  0.10519
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.471 on 10 degrees of freedom
Multiple R-squared:  0.2409,    Adjusted R-squared:  0.165
F-statistic: 3.173 on 1 and 10 DF,  p-value: 0.1052
```

由输出结果可知，所得回归方程为： $Y=0.0819X+9.90253$

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-4.4073，-2.6983，0.1917，1.9913，5.1356；

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，由常数项系数的 p 值=0.00605<0.05 可知，该系数通过了显著性检验。

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。

#### 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：



```

Call:
lm(formula = V2 ~ V1, data = google)

Residuals:
    Min       1Q   Median       3Q      Max
-33.297 -13.967   1.335  12.056  39.395

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 156.67805     3.52517   44.45  <2e-16 ***
V1           1.28372     0.05774   22.23  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17.93 on 103 degrees of freedom
Multiple R-squared:  0.8276, Adjusted R-squared:  0.8259
F-statistic: 494.3 on 1 and 103 DF, p-value: < 2.2e-16

```

由输出结果可知，所得回归方程为  $V2=1.28372V1+156.67805$ 。

对于回归系数的显著性检验，V1 的系数的 p 值  $<2e-16 < 0.05$ ，常数项系数的 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 0.8276，Adjusted R-squared 为 0.8259，F 检验统计量为 494.3，对应的 p 值  $<2.2e-16 < 0.05$ ，落在拒绝域内，说明回归方程显著。

## 2、利用迭代法处理自相关

该方程的回归信息如下：

```

Call:
lm(formula = y1 ~ x1)

Residuals:
    Min       1Q   Median       3Q      Max
-14.3720  -2.6945  -0.2969   2.4957  12.1304

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.522      1.272    1.983  0.0501 .
x1             2.045      0.457    4.474  2e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.327 on 102 degrees of freedom
Multiple R-squared:  0.1641, Adjusted R-squared:  0.1559
F-statistic: 20.02 on 1 and 102 DF, p-value: 1.998e-05

```

由输出结果可知，所得回归方程为  $y_1 = 2.045x_1 + 2.522$

对于回归系数的显著性检验， $x_1$  的系数的  $p$  值  $= 2e-05 < 0.05$ , 常数项系数的  $p$  值  $= 0.0501 > 0.05$ , 常数项系数未通过了显著性检验。

对于回归方程的显著性检验，Multiple R-squared 为 0.1641，Adjusted R-squared 为 0.1559，F 检验统计量为 20.02，对应的  $p$  值  $1.998e-05 < 0.05$ , 落在拒绝域内，说明回归方程显著。