HW4

9.13

模拟 $\phi = 0.7$，$\theta = -0.5$，$\mu = 100$ 的 ARMA(1, 1) 模型。模拟 50 个值，但将最后的 10 个值搁置起来，以对预测值与真实值进行比较。

In [75]:
```
set.seed(114514)
series = arima.sim(n=50, model=list(ar=0.7, ma=0.5)) + 100

real = series[41:50]
series = series[1:40]
```

(a)

使用序列前 40 个值，求 ϕ、θ 和 μ 的极大似然估计值。

Solution.

In [76]:
```
model = arima(series, order=c(1,0,1), method='ML')
model
```

```
Call:
arima(x = series, order = c(1, 0, 1), method = "ML")

Coefficients:
         ar1     ma1  intercept
      0.6524  0.9015   100.1296
s.e.  0.1202  0.1466     0.7776

sigma^2 estimated as 0.9029:  log likelihood = -56.29,  aic = 120.58
```

由上可知，$\hat{\phi} = \boxed{0.6524}$, $\hat{\theta} = \boxed{0.9015}$, $\hat{\mu} = \boxed{100.1296}$。

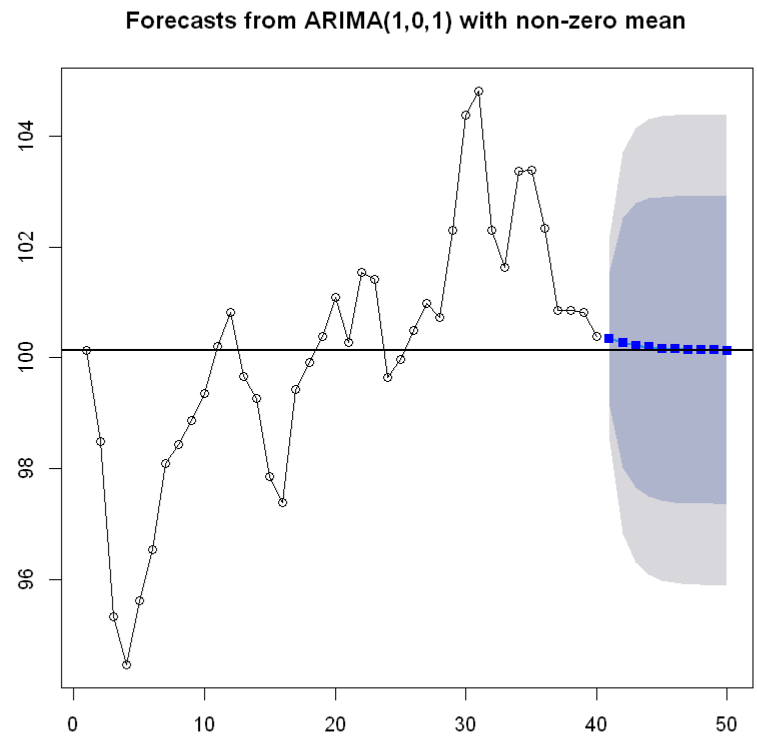
除了 $\hat{\theta}$ 之外，其他估计值与真实值相差不大。

(b)

使用所估计的模型，预测序列接下来的 10 个值，并画出带 10 个预测值的序列。在估计的序列均值上画一条水平线。

Solution.

In [77]:
```
library(forecast)
forecast_values = forecast(model, h=10)
plot(forecast_values, type='o')
points(forecast_values$mean, col='blue',pch=15)
abline(h=100.1296, col='black', lwd=2)
```



(c)

将 10 个预测值与所留出的真实值进行比较。

Solution.

In [78]:
```
pred = forecast_values$mean

cbind(real, pred)
```

A Time Series: 10 × 2

	real	pred
41	99.59074	100.3477
42	98.08429	100.2719
43	98.12495	100.2224
44	97.55493	100.1902
45	97.97085	100.1691
46	97.77657	100.1554
47	99.95185	100.1464
48	102.23360	100.1406
49	100.36248	100.1368
50	100.45562	100.1343

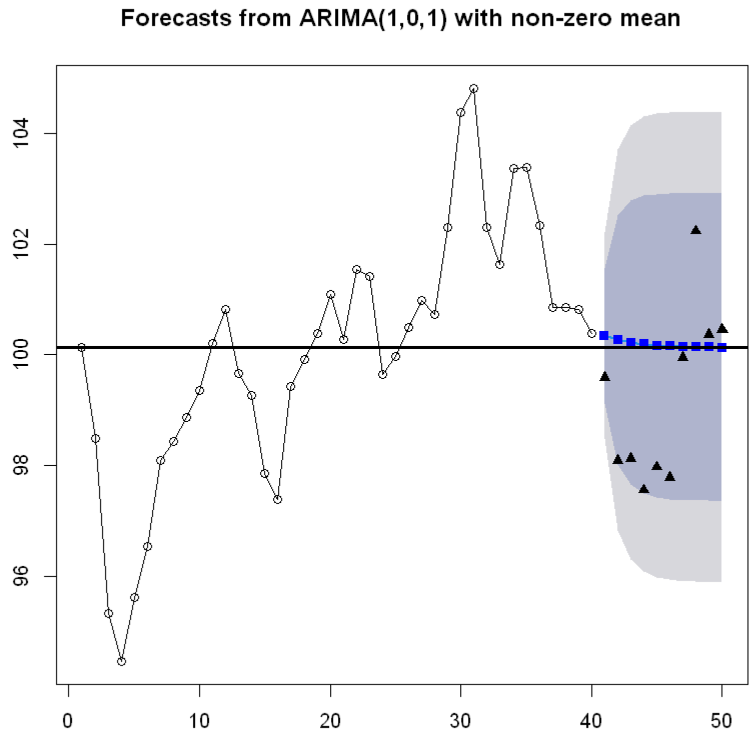
差别都还可以，不算太大。

(d)

画出预测及其 95% 预测极限。真实值是否落入预测极限的区间?

Solution.

In [79]:
```
plot(forecast_values, type='o')
points(forecast_values$mean, col='blue',pch=15)
points(x=(41:50), y=real, col= 'black',pch=17)
abline(h=100.1296, col='black', lwd=3)
```



真实值基本都在预测区间当中。

(e)

用同样的参数值和相同的样本容量，模拟一个新的序列，并重复 (a) 到 (d)。

Solution.

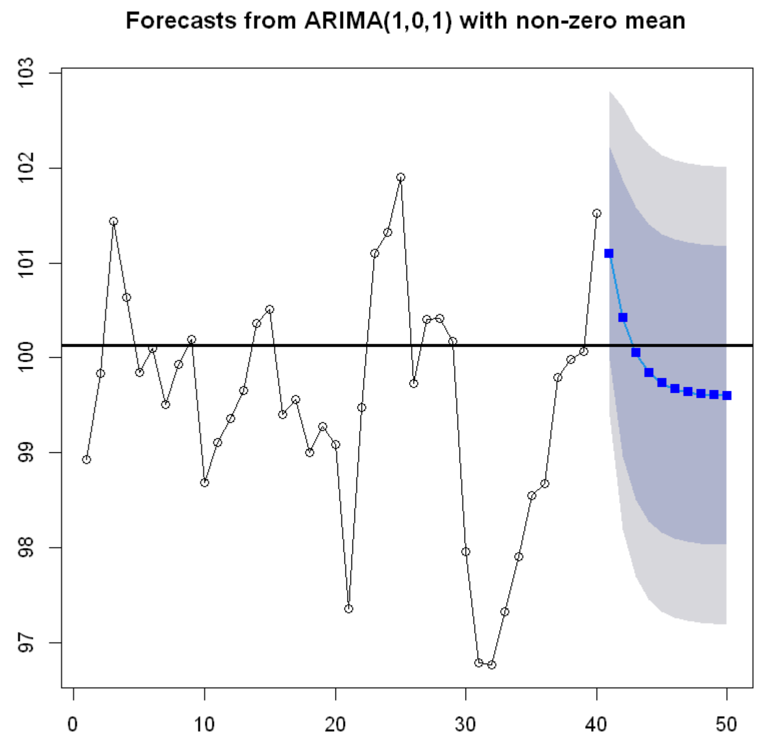
In [80]:
```
set.seed(10086)
series = arima.sim(n=50, model=list(ar=0.7, ma=0.5)) + 100
real = series[41:50]
series = series[1:40]
model = arima(series, order=c(1,0,1), method='ML')
print(model)

forecast_values = forecast(model, h=10)

plot(forecast_values, type='o')
points(forecast_values$mean, col='blue',pch=15)
abline(h=100.1296, col='black', lwd=3)
pred = forecast_values$mean
```

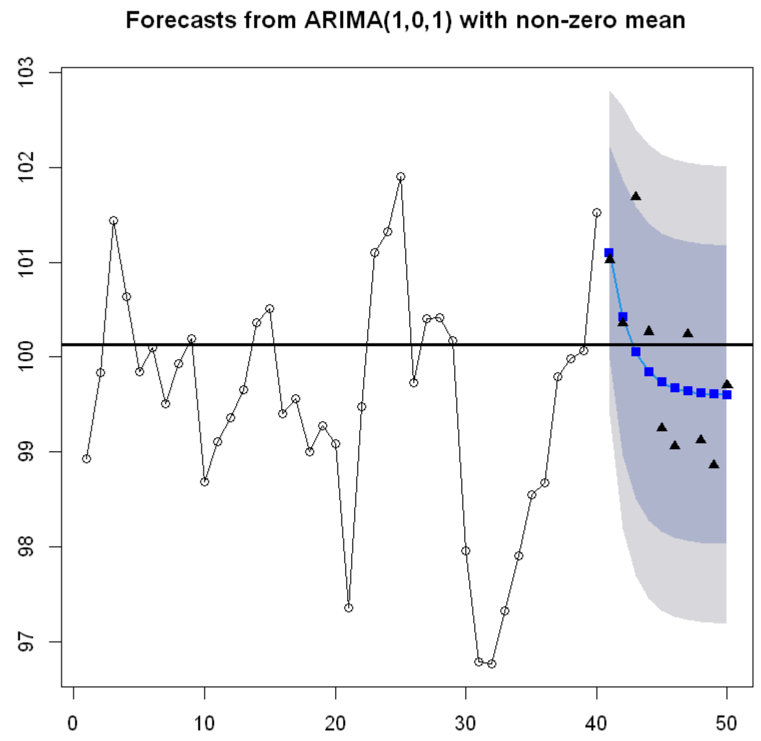
```r
cbind(real, pred)

plot(forecast_values, type='o')
points(forecast_values$mean, col='blue',pch=15)
points(x=(41:50), y=real, col= 'black',pch=17)
abline(h=100.1296, col='black', lwd=3)
```

```
Call:
arima(x = series, order = c(1, 0, 1), method = "ML")

Coefficients:
         ar1     ma1  intercept
      0.5480  0.2777    99.5971
s.e.  0.1961  0.2350     0.3782

sigma^2 estimated as 0.7639:  log likelihood = -51.73,  aic = 111.46
```

A Time Series: 10 × 2

	real	pred
41	101.02436	101.09532
42	100.35559	100.41818
43	101.67953	100.04708
44	100.26480	99.84371
45	99.24863	99.73225
46	99.05965	99.67117
47	100.23960	99.63769
48	99.12448	99.61935
49	98.85758	99.60929
50	99.70334	99.60378





模拟效果基本与第一次相同。效果都很好。

9.14

模拟 $\theta=0.8$，$\theta_0=0$ 的 IMA(1, 1) 模型。模拟 35 个值，但将最后的 5 个值搁置起来，以对预测值与真实值进行比较。

In [81]:
```
set.seed(1919810)
series = arima.sim(n=35, model=list(order=c(0,1,1), ma=-0.8))[-1]
# 去掉第一个值，因为第一个值是常数0
real= series[31:35]
series = series[1:30]
```

(a)

使用序列前 30 个值，求 θ 的极大似然估计值。

Solution.

In [82]:
```
model = arima(series, order=c(0,1,1), method='ML')
model
```

```
Call:
arima(x = series, order = c(0, 1, 1), method = "ML")

Coefficients:
          ma1
       -0.743
s.e.    0.097

sigma^2 estimated as 0.8141:  log likelihood = -38.57,  aic = 81.14
```

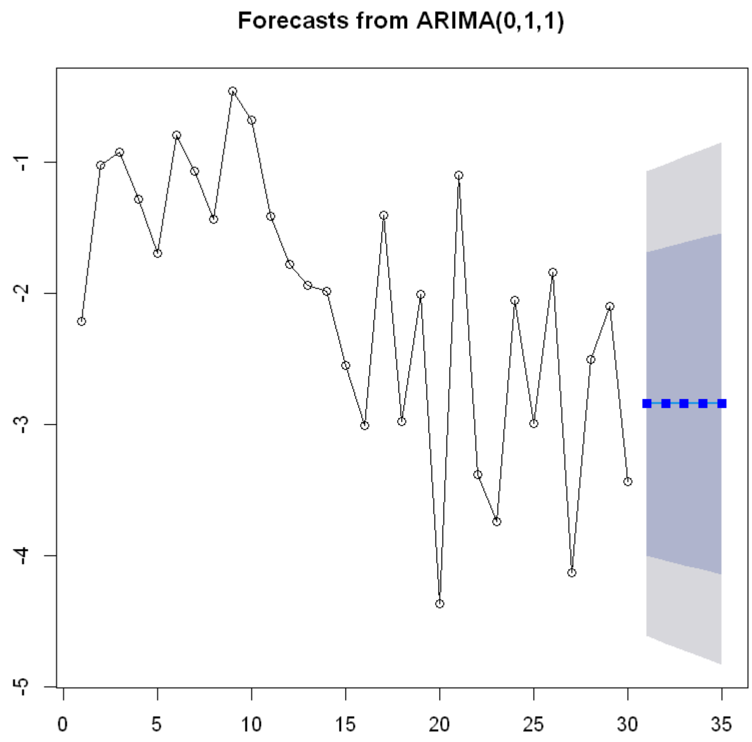
由上可知，$\hat{\theta} = \boxed{0.7430}$。

(b)

使用所估计的模型，预测序列接下来的 5 个值，并画出带 5 个预测值的序列。这些预测值有什么特殊之处吗?

Solution.

In [83]:
```
library(forecast)
forecast_values = forecast(model, h=5)
plot(forecast_values, type='o')
points(forecast_values$mean, col='blue',pch=15)
```



特殊之处在于，这五个预测值似乎并没有什么波动，都保持在水平线之上。

(c)

将 5 个预测值与所留出的真实值进行比较。

Solution.

In [84]:
```
pred= forecast_values$mean
cbind(real, pred)
```

A Time Series: 5 × 2

	real	pred
31	-0.8370999	-2.841928
32	-2.4948004	-2.841928
33	-2.9323672	-2.841928
34	-2.8074105	-2.841928
35	-2.5760650	-2.841928

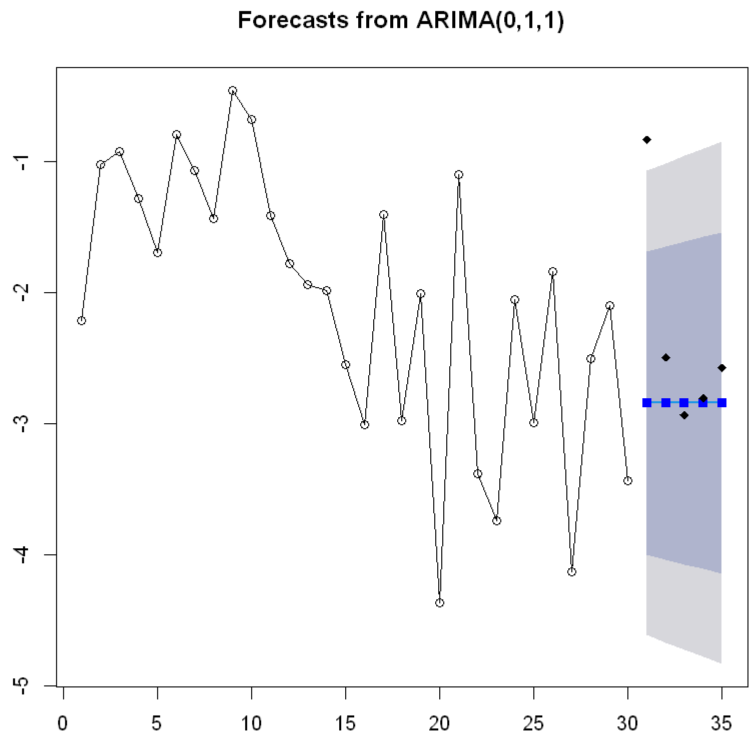
比较可知，模拟效果尚可，偏误不大。

(d)

画出预测及其 95% 预测极限。真实值是否落入预测极限的区间?

Solution.

In [85]:
```
forecast_values = forecast(model, h=5)
plot(forecast_values, type='o')
points(forecast_values$mean, col='blue',pch=15)
points(x=(31:35), y= real, col= 'black',pch=18)
```



第31个真实值明显偏离了预测区间，但其他真实值都在预测区间内。

甚至第34个真实值点基本和预测值重合了。

(e)

用同样的参数值和相同的样本容量，模拟一个新的序列，并重复 (a) 到 (d)。

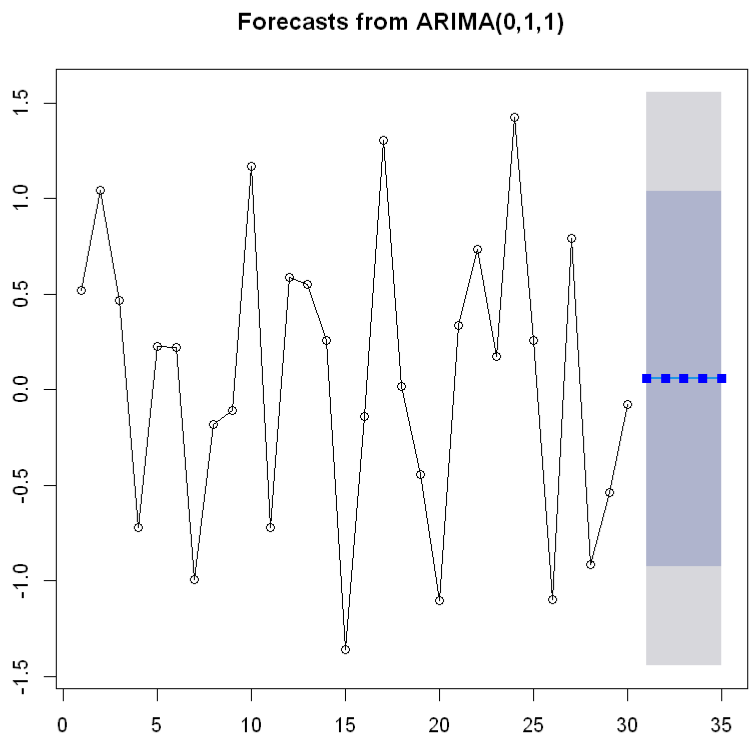
Solution.

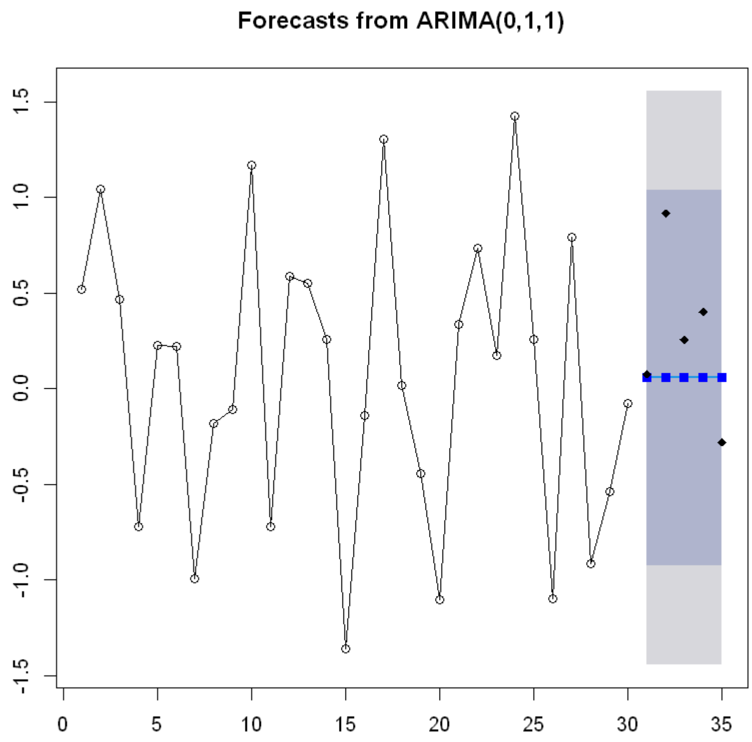
In [86]:
```
set.seed(7777777)
series = arima.sim(n=35, model=list(order=c(0,1,1), ma=-0.8))[-1]
real= series[31:35]
series = series[1:30]
model = arima(series, order=c(0,1,1), method='ML')
print(model)

forecast_values = forecast(model, h=5)
plot(forecast_values, type='o')
points(forecast_values$mean, col='blue',pch=15)
pred= forecast_values$mean
cbind(real, pred)

plot(forecast_values, type='o')
points(forecast_values$mean, col='blue',pch=15)
points(x=(31:35), y= real, col= 'black',pch=18)
```

```
Call:
arima(x = series, order = c(0, 1, 1), method = "ML")

Coefficients:
          ma1
      -1.0000
s.e.   0.1045

sigma^2 estimated as 0.5678:  log likelihood = -34.64,  aic = 73.28
```

A Time Series: 5 × 2

	real	pred
31	0.07555234	0.05628263
32	0.91543928	0.05628263
33	0.25717217	0.05628263
34	0.40096227	0.05628263
35	-0.27907410	0.05628263





这一次，模拟效果很好，没有再出现真实值点明显偏离了预测区间的情况。

P178

10.7 假设过程 $\{Y_t\}$ 满足 $Y_t = Y_{t-4} + e_t$，当 $t=1,2,3,4$ 时，$Y_t = e_t$。

(a) 求 $\{Y_t\}$ 的方差函数。

(b) 求 $\{Y_t\}$ 的自相关函数。

(c). 证明 $\{Y_t\}$ 的模型是季节 ARIMA 模型

Proof.

a). $Var(Y_t) = Var(Y_{t-4} + e_t) = Var(Y_{t-4}) + 2Cov(Y_{t-4}, e_t) + Var(e_t) \Rightarrow \gamma_t = \gamma_{t-4} + \sigma_e^2$

$$\Rightarrow \gamma_t = \gamma_{t-8} + 2\sigma_e^2 = \cdots = \begin{cases} \frac{t}{4}\sigma_e^2, & t-[\frac{t}{4}]=0 \\ \gamma_{t-4[\frac{t}{4}]} + [\frac{t}{4}]\sigma_e^2, & t-[\frac{t}{4}] \geq 1 \end{cases}$$

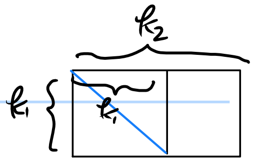
其中，$1 \leq t-4[\frac{t}{4}] \leq 3$，则 $\gamma_{t-4[\frac{t}{4}]} = \sigma_e^2$，即 $\gamma_t = \begin{cases} \frac{t}{4}\sigma_e^2, & t-[\frac{t}{4}]=0 \\ (1+[\frac{t}{4}])\sigma_e^2, & t-[\frac{t}{4}] \geq 1 \end{cases}$

也可写作：对于 $t=4k-m$ (其中 $k=1,2,3,\ldots$ $m=1,2,3$)，$\gamma_0 = k\sigma_e^2$

b). 关注 $\gamma_{t,s} = Cov(Y_t, Y_s)$，其中 $t = 4k_1 - m_1$，$s = 4k_2 - m_2$，$k_1, k_2 \geq 1$，$m_1 m_2 = 0,1,2,3$

由 (a)，知 $Y_{4k-m} = Y_{4k-m-4} + e_{4k-m} = \cdots = Y_{4-m} + e_{4k-m} + e_{4k-m-4} + \cdots + e_{8-m}$

即 $Y_{4k-m} = \sum_{i=1}^{k} e_{4i-m}$



$\therefore \gamma_{t,s} = Cov(Y_t, Y_s) = \sum_{i=1}^{k_1}\sum_{j=1}^{k_2} Cov(e_{4i-m_1}, e_{4j-m_2})$

① $m_1 = m_2$，$\gamma_{t,s} = \min\{k_1 k_2\}\sigma_e^2$

则 $\rho_{t,s} = \frac{Cov(Y_t Y_s)}{Var(Y_t)^{1/2} Var(Y_s)^{1/2}} = \frac{\min\{k_1, k_2\}\sigma_e^2}{\sqrt{k_1}\sigma_e\sqrt{k_2}\sigma_e} = \frac{\min\{k_1 k_2\}}{\sqrt{k_1}\sqrt{k_2}}$，$k_1 = [\frac{t}{4}]+1$，$k_2 = [\frac{s}{4}]+1$

② $m_1 \neq m_2$，$\gamma_{t,s} = 0 \Rightarrow \rho_{t,s} = 0$

综上，
$$\rho_{t,s} = \begin{cases} \left([\frac{\min\{t,s\}}{4}]+1\right)\left([\frac{t}{4}]+1\right)^{-1/2}\left([\frac{s}{4}]+1\right)^{-1/2}, & t-4[\frac{t}{4}] = s-4[\frac{s}{4}] \\ 0, & 否则 \end{cases}$$

c). $\nabla_4 Y_t = Y_t - Y_{t-4} = e_t = (1-B^4)Y_t$，即 $(1-B^4)\nabla^0\nabla_4^1 Y_t = e_t$

所以由定义知 $p=0$，$P=1$，$d=0$，$D=1$，$q=0$，$Q=0$，$S=4$

$\{Y_t\}$可被认为是 ARIMA$(0,0,0)\times(1,1,0)_4$ 过程

10.9，10.11 见R语言实现

P226

12.3 使用定义 $\eta_t = r_t^2 - \sigma_{t|t-1}^2$ [方程(12.2.4)] 并证明$\{\eta_t\}$是一个序列不相关过程，同样证明 η_t 与过去的收益率平方是不相关的，即证明对于 $k>0$，$Corr(\eta_t, r_{t-k}^2) = 0$.

Proof.

定义 $\underbrace{\sigma_{t|t-1}^2 = h_t = Var(r_t | F_{t-1})}_{\text{模型中，} h_t \text{ 是由 } r_{t-1} \dots r_{t-q} \text{ 定义的}}$，则 $\eta_t = r_t^2 - h_t = h_t(\varepsilon_t^2 - 1)$

$E(\eta_t \eta_{t-j}) = E[E(\eta_t\eta_{t-j} | r_{t-1} \dots r_1)] = E[\eta_{t-j} E(\eta_t | r_{t-1} \dots r_1)]$

其中 $E(\eta_t | r_{t-1}\dots r_1) = h_t E(\varepsilon_t^2 - 1 | r_{t-1} \dots r_1) = 0 \Rightarrow E(\eta_t \eta_{t-j}) = 0$

$\Rightarrow Cov(\eta_t \eta_{t-j}) = 0$，所以$\{\eta_t\}$序列不相关

$Cov(\eta_t, r_{t-k}^2) = Cov(\eta_t, \eta_{t-k} + h_{t-k}) = Cov(\eta_t \eta_{t-k}) + Cov(\eta_t, h_{t-k})$

$= 0 + Cov(\eta_t, h_{t-k}) = E(\eta_t, h_{t-k}) = E[E(\eta_t h_{t-k} | r_1 \dots r_{t-1})]$

$= E[h_{t-k} E(\eta_t | r_{t-1} \dots r_1)] = E[h_{t-k} h_t E(\varepsilon_t^2 - 1 | r_{t-1} \dots r_1)]$

$= E[h_{t-k} h_t \cdot 0] = 0$，即 $Cov(\eta_t, r_{t-k}^2) = 0 \Rightarrow Corr(\eta_t, r_{t-k}^2) = 0$

12.4 把 $\sigma^2_{t|t-1}=r_t^2-\eta_t$ 带入方程（12.2.2），推导得到方程（12.2.5）的代数运算过程。

（12.2.2）：$\sigma^2_{t|t-1}=\omega+\alpha r^2_{t-1}$　　（12.2.5）：$r_t^2=\omega+\alpha r^2_{t-1}+\eta_t$

Proof.

将 $\sigma^2_{t|t-1}=r_t^2-\eta_t$ 代入到（12.2.2）：$\sigma^2_{t|t-1}=\omega+\alpha r^2_{t-1}$ 中，

得（12.2.5）：$r_t^2=\omega+\alpha r^2_{t-1}+\eta_t$　得证.

12.5 证明方程（12.2.8）

（12.2.8）：　$\tau=\omega^2+2\omega\alpha\sigma^2+\alpha^2 3\tau$

Proof.

（12.2.8）来源于对（12.2.2）两边平方求期望，即：

$E(\sigma^4_{t|t-1})=E(\omega^2+2\omega\alpha r^2_{t-1}+\alpha^2 r^4_{t-1})$

$=\omega^2+2\omega\alpha E(\sigma^2_{t|t-1}\varepsilon_t^2)+\alpha^2E(\sigma^4_{t|t-1}\varepsilon_t^4)$，又由 ε_t 无关于 $\sigma^2_{t|t-1}$

原式 $=\omega^2+2\omega\alpha E(\sigma^2_{t|t-1})E(\varepsilon_t^2)+\alpha^2E(\sigma^4_{t|t-1})E(\varepsilon_t^4)$

$=\omega^2+2\omega\alpha\sigma^2[Var(\varepsilon_t)+E^2(\varepsilon_t)]+\alpha^2E(\sigma^4_{t|t-1})E(\varepsilon_t^4)$

由 $Var(\varepsilon_t)=1$，$E(\varepsilon_t)=0$，$E(\varepsilon_t^4)=3\cdot1^2=3$，继而代入 $E(\sigma^4_{t|t-1})=\tau$

故有 $\tau=\omega^2+2\omega\alpha\sigma^2+\alpha^2 3\tau$，得证.

10.9

最先由 Box 和 Jenkins(1976) 研究的月航线客运量时间序列被视为典型的时间序列。数据详见文件 airline.

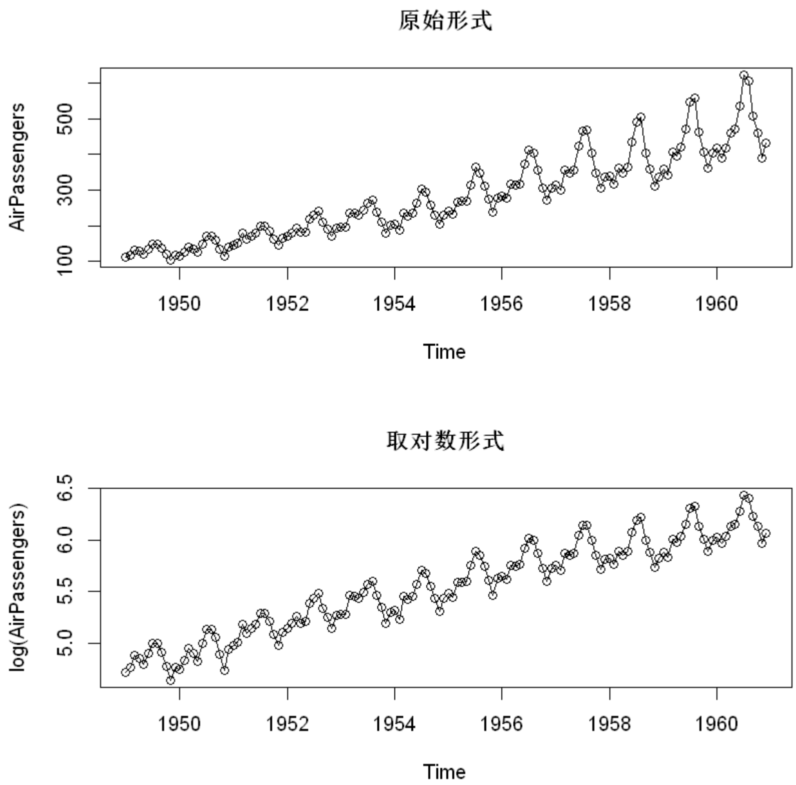
In [14]:
```r
library(TSA)
data(AirPassengers)
```

(a)

画出此序列的原始形式和取对数形式的时间序列图。说明对数变换在这里是恰当的。

Solution.

In [2]:
```r
par(mfrow = c(2,1))
plot(AirPassengers, type='o', main='原始形式')
plot(log(AirPassengers), type='o', main='取对数形式')
```



从图中可以看出，两个时间序列都表现出了明显的上升趋势。

但是，取对数后序列的上升趋势明显要更加的稳定。

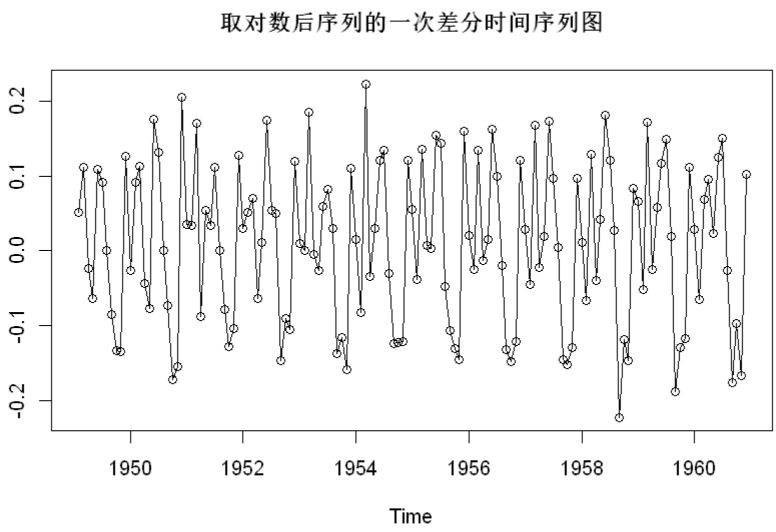
因此如果我们想对其进行分析，取对数后的序列是更合适的。

(b)

画出并解释取对数后序列的一次差分时间序列图。

Solution.

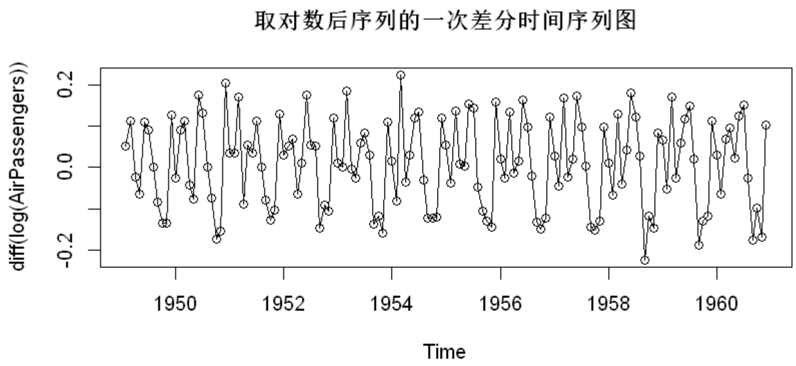
In [3]:
```
par(pin=c(6,3))

plot(diff(log(AirPassengers)),
     pch=1, type='o',
     main='取对数后序列的一次差分时间序列图')
```

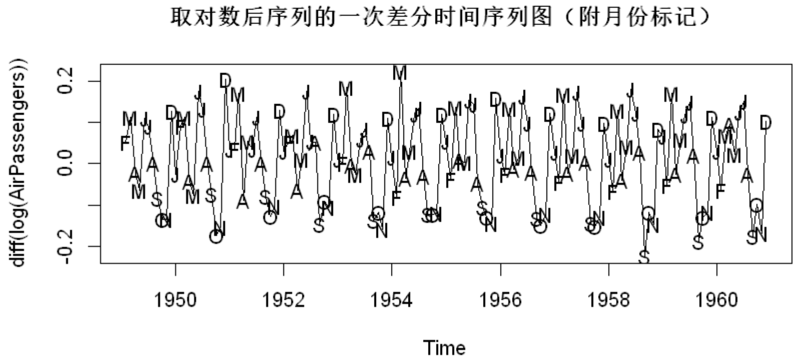


基本上识别不出明显的趋势。

但是，有理由怀疑，这个差分序列存在着某种季节性波动。

In [4]:
```
par(mfrow = c(2,1))

plot(diff(log(AirPassengers)),
     pch=1, type='o',
     main='取对数后序列的一次差分时间序列图')

plot(diff(log(AirPassengers)),
     type='l',
     main='取对数后序列的一次差分时间序列图（附月份标记）')
points(diff(log(AirPassengers)),
       x= time(diff(AirPassengers)),
       pch= as.vector(season(diff(log(AirPassengers)))))
```





季节性是较为明晰的。

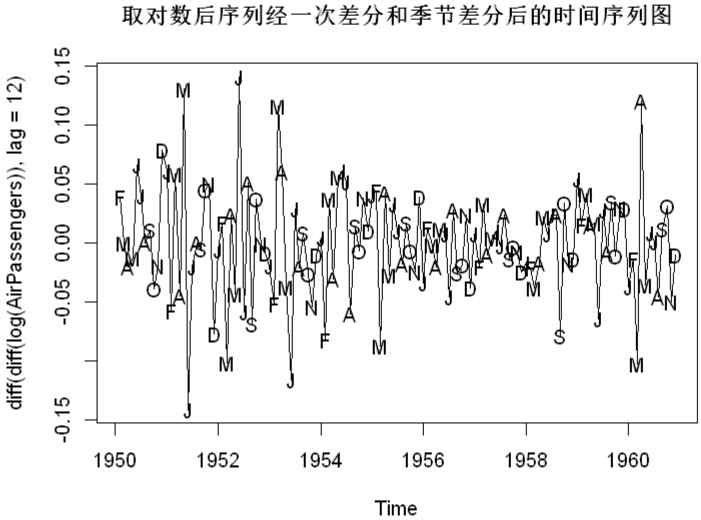
例如，标记为'M'、'J'、'D'的月份总比其他月份要高。标记为'O'、'S'的月份总是较低。

(c)

画出并解释取对数后序列经一次差分和季节差分后的时间序列图。

Solution.

In [5]:
```
par(pin=c(5,3))

plot(diff(diff(log(AirPassengers)), lag=12), type='l',
     main= '取对数后序列经一次差分和季节差分后的时间序列图')

points(diff(diff(log(AirPassengers)), lag=12),
       x=time(diff(diff(log(AirPassengers)), lag=12)),
       pch= as.vector(season(diff(log(AirPassengers)))))
```



季节周期性变得不那么明显了。

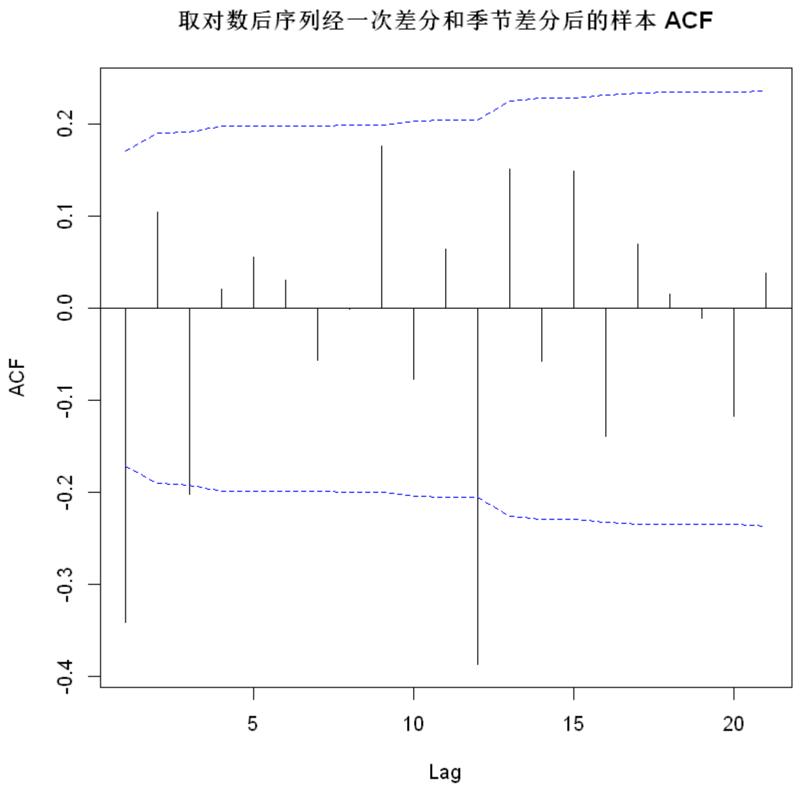
未经季节差分的序列中，标记为'M'、'J'、'D'的月份总比其他月份要高。标记为'O'、'S'的月份总是较低。但是，经过季节差分后，这些月份的波动变得不那么明显了。

(d)

计算并解释取对数后序列经一次差分和季节差分后的样本 ACF。

Solution.

In [6]:
```
acf(as.vector(diff( diff(log(AirPassengers)), lag=12 ) ),
    ci.type='ma',
    main='取对数后序列经一次差分和季节差分后的样本 ACF')
```



由上图可知，在1月和12月，都存在显著自相关。

另外在3月似乎也由显著的自相关（相对不那么显眼）。

(e)

用“航线模型”($\mathrm{ARIMA}(0,1,1)\times(0,1,1)_{12}$ ）拟合对数化的序列。

Solution.

In [7]:
```
model = arima(log(AirPassengers),
              order=c(0,1,1),
              seasonal=list(order=c(0,1,1),
              period=12))

model
```

```
Call:
arima(x = log(AirPassengers), order = c(0, 1, 1), seasonal = list(order = c(0, 
    1, 1), period = 12))

Coefficients:
          ma1     sma1
      -0.4018  -0.5569
s.e.   0.0896   0.0731

sigma^2 estimated as 0.001348:  log likelihood = 244.7,  aic = -485.4
```

系数拟合结果都显著。

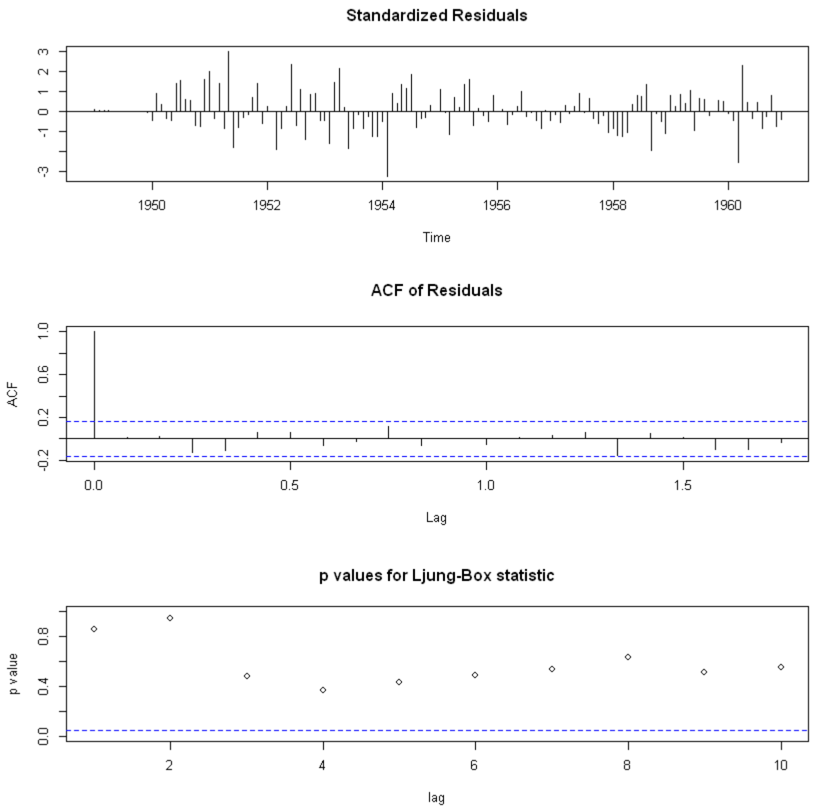
$\hat{\theta}=0.4018$, $\hat{\Theta}=0.5569$

(f)

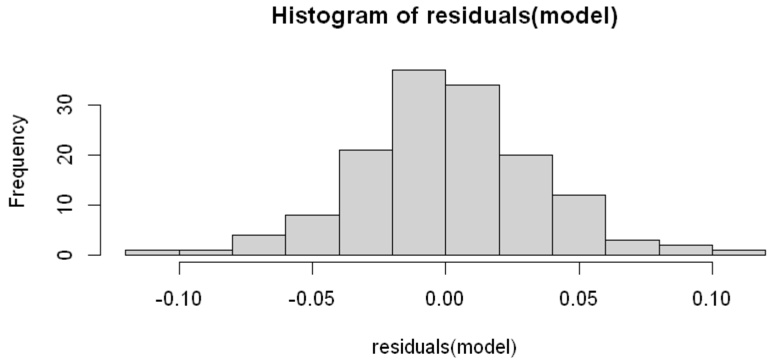
对模型及其自相关性和残差的正态性进行诊断。

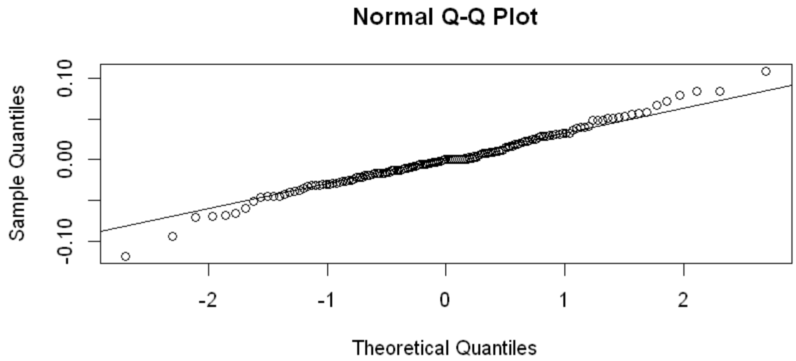
Solution.

In [8]:
```
tsdiag(model)
```



In [9]:
```
par(mfrow = c(2,1))
hist(residuals(model))

qqnorm(model$residuals)
qqline(model$residuals)
```





In [10]:
```
shapiro.test(residuals(model))
```

```
	Shapiro-Wilk normality test

data:  residuals(model)
W = 0.98637, p-value = 0.1674
```

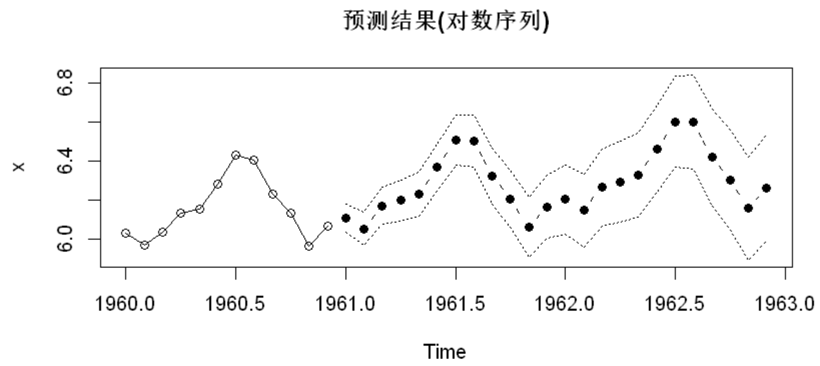
自相关性和残差的正态性表现都相对良好。

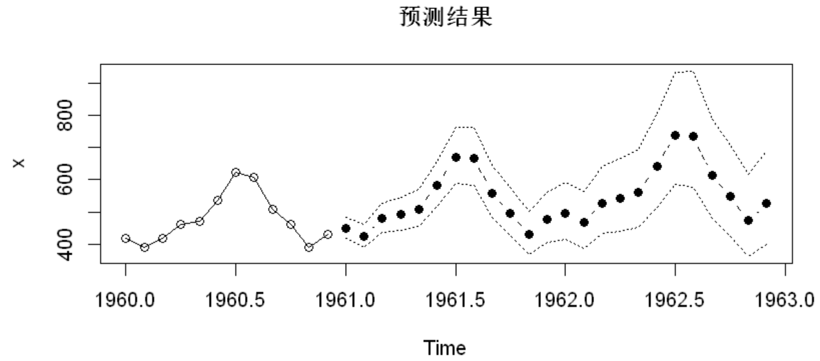
(g)

假设前置时间为两年，对此序列进行预测，并要求给出预测极限。

Solution.

In [11]:
```
par(mfrow=c(2,1))
plot(model,n1=c(1960,1), n.ahead=24, pch=19, main='预测结果(对数序列)')
plot(model,n1=c(1960,1), n.ahead=24, pch=19, main='预测结果', transform=exp)
```





黑色点为预测值，虚线为预测极限。

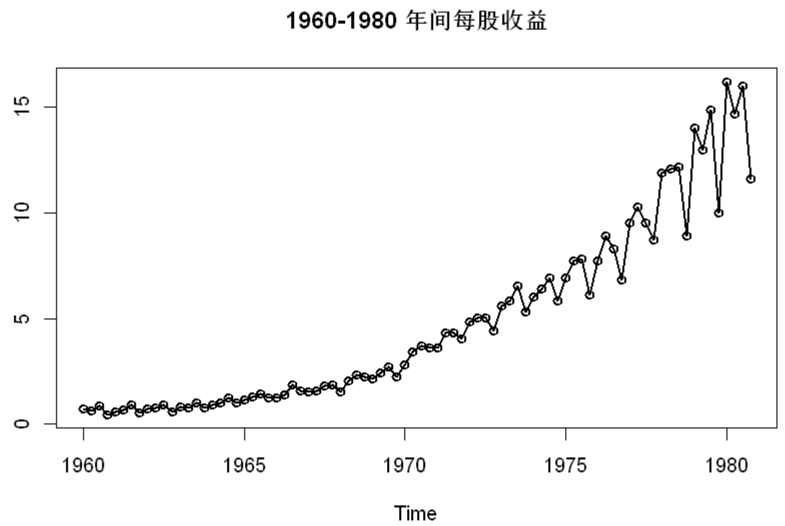
10.11

美国 Johnson & Johnson 公司于 1960～1980 年间每股收益的季度数据见于文件 JJ 中.

In [1]:
```r
library(TSA)
data(JJ)
```

```
Warning message:
"package 'TSA' was built under R version 4.3.3"

Attaching package: 'TSA'


The following objects are masked from 'package:stats':

    acf, arima


The following object is masked from 'package:utils':

    tar
```

In [2]:
```r
par(pin=c(6,3))
plot(JJ, type='o', pch=1, lwd=2, main='1960-1980 年间每股收益')
```

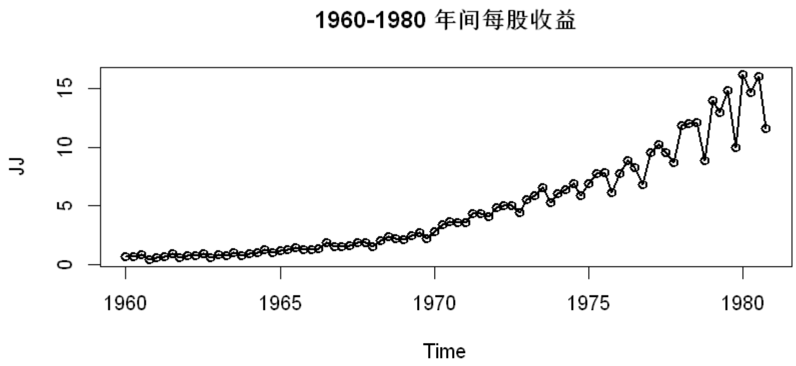


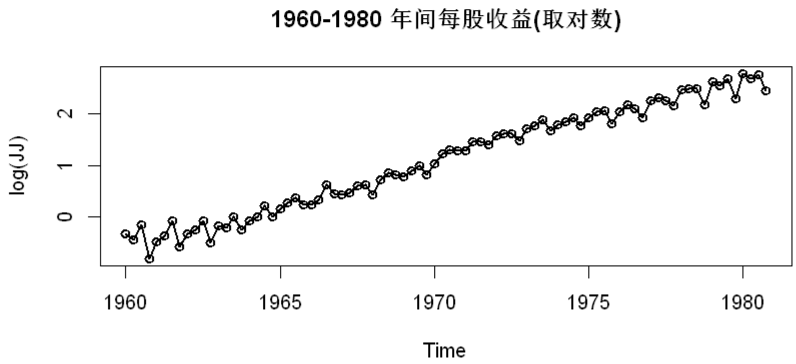
(a)

画出该序列及其取对数后的时间序列图。论证对序列进行对数变换的必要性。

Solution.

In [3]:
```
par(mfrow=c(2,1))
plot(JJ, type='o', pch=1, lwd=2, main='1960-1980 年间每股收益')
plot(log(JJ), type='o', pch=1, lwd=2, main='1960-1980 年间每股收益(取对数)')
```





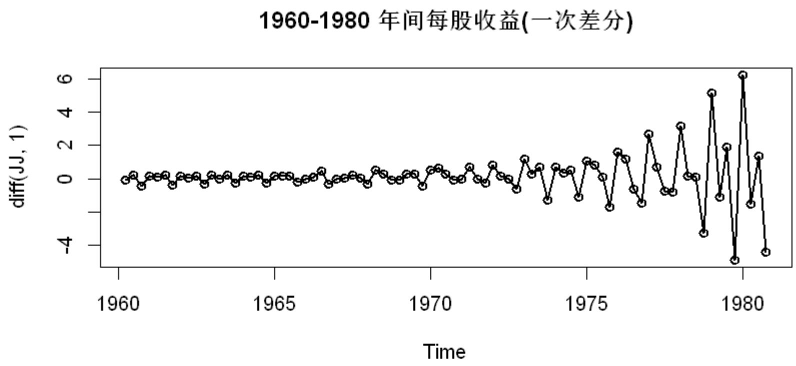
很明显，取对数之后，序列的增量趋势更加“稳定”。

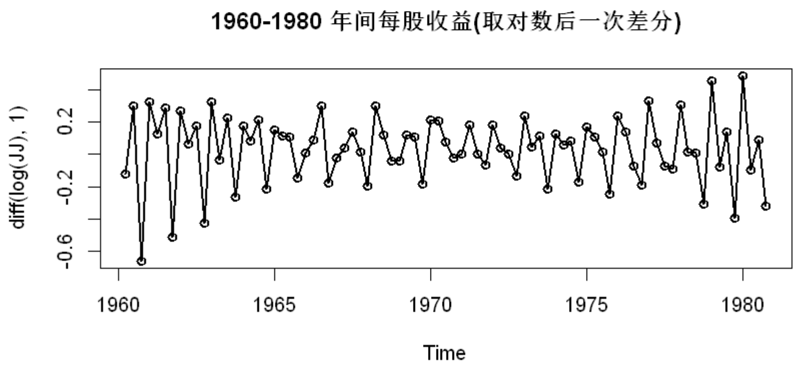
(b)

序列明显是非平稳的。对其进行一次差分变换并画出序列图。现在序列平稳性有无合理性？

Solution.

In [4]:
```
par(mfrow = c(2,1))
plot(diff(JJ,1), type='o', pch=1, lwd=2, main='1960-1980 年间每股收益(一次差分)'
plot(diff(log(JJ),1), type='o', pch=1, lwd=2, main='1960-1980 年间每股收益(取对数
```





未作对数处理时，差分序列的非平稳性是明显的。

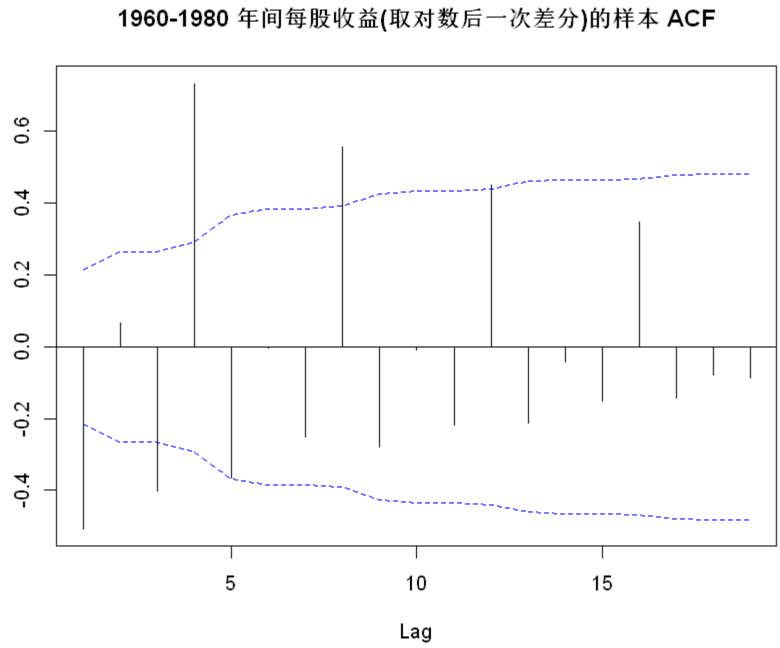
而作对数处理后，差分序列似乎相对平稳了一点，但是目测仍有些可疑。

(c)

计算并画出经一次差分后序列的样本 ACF，并解释结果。

Solution.

In [5]:
```
par(pin=c(6,4))
acf(diff(as.vector(log(JJ)),1), ci.type='ma',
    main='1960-1980 年间每股收益(取对数后一次差分)的样本 ACF')
```



在 4, 8, 12, 16 这几个时间点上，我们捕捉到了明显的自相关性。

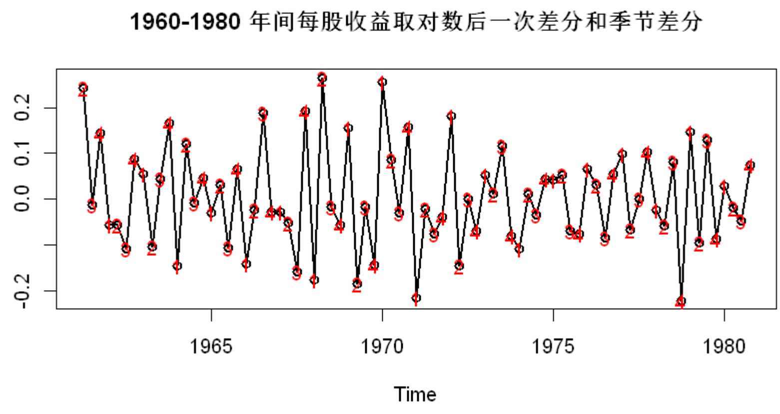
据此有理由怀疑，这个序列可能是一个季节性 ARIMA 模型，其季节周期长度可能为4。

(d)

画出并解释经过一次差分和季节差分后的序列图。牢记季度数据一季的长度为4。

Solution.

In [6]:
```
par(pin=c(6,2))

plot(diff(diff(log(JJ),4),1), type='o', pch=1, lwd=2,
     main='1960-1980 年间每股收益取对数后一次差分和季节差分')
points(y=diff(diff(log(JJ),4),1),
       x=time(diff(diff(log(JJ),4),1)),
       pch=as.vector(season(diff(diff(log(JJ),4),1))),
       col='red')
```



现在看来，经过季节差分和一次差分后的序列（相对于各期时间点）的分布已经较为随机，没有明显的模式。

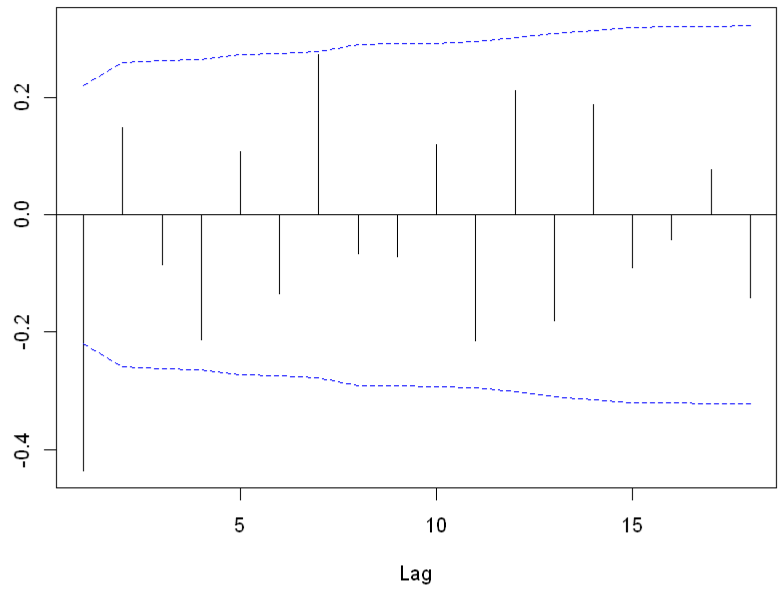
(e)

画出并说明经过一次差分和季节差分后的序列的样本 ACF。

Solution.

In [7]:
```
par(pin=c(6,4))
acf(as.vector( diff(diff(log(JJ),4),1) ), ci.type='ma',
    main='1960-1980 年间每股收益取对数后一次差分和季节差分后的样本 ACF')
```

1960-1980 年间每股收益取对数后一次差分和季节差分后的样本 ACF



在第1期出现显著自相关。其余时间点，仅有第7期较为接近置信边界，其余均不显著。

忽略这个，我们可以认为之前的差分操作是合理的。据此将模型识别为 $\text{ARIMA}(0,1,1)\times(0,1,1)_4$

(f)

拟合 $\text{ARIMA}(0,1,1)\times(0,1,1)_4$ 模型，并评估系数估计值的显著性。

Solution.

In [8]:
```
model = arima(log(JJ),
              order=c(0,1,1),
              seasonal=list(order=c(0,1,1), period=4)
              )

model
```

```
Call:
arima(x = log(JJ), order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), peri
od = 4))

Coefficients:
          ma1     sma1
      -0.6809  -0.3146
s.e.   0.0982   0.1070

sigma^2 estimated as 0.007931:  log likelihood = 78.38,  aic = -152.75
```

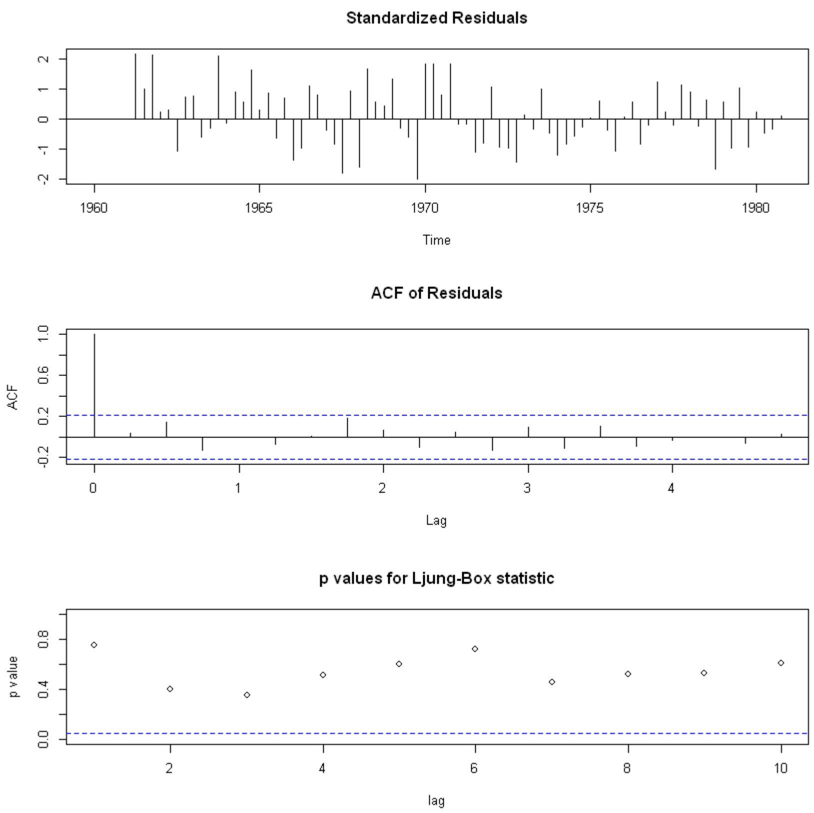
据此，模型拟合为 $\text{ARIMA}(0,1,1)\times(0,1,1)_4$。系数估计均显著。

(g)

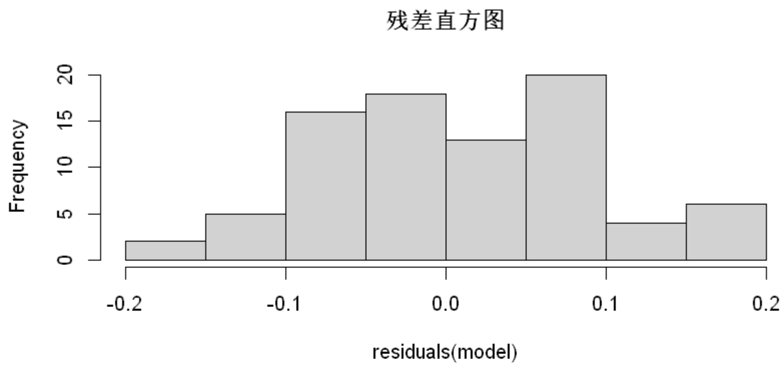
对残差进行所有的诊断性检验。

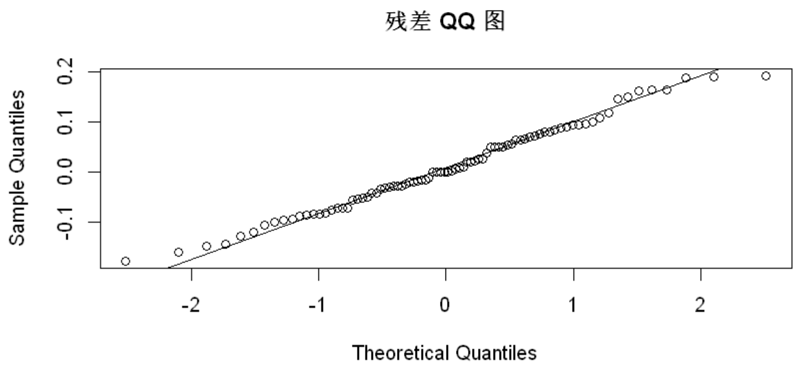
Solution.

In [9]:
```
tsdiag(model)
```



In [10]:
```
par(mfrow=c(2,1))

hist(residuals(model), main='残差直方图')
qqnorm(residuals(model), main='残差 QQ 图')
qqline(residuals(model))
```





In [11]:
```
shapiro.test(residuals(model))
```

```
	Shapiro-Wilk normality test

data:  residuals(model)
W = 0.98583, p-value = 0.489
```

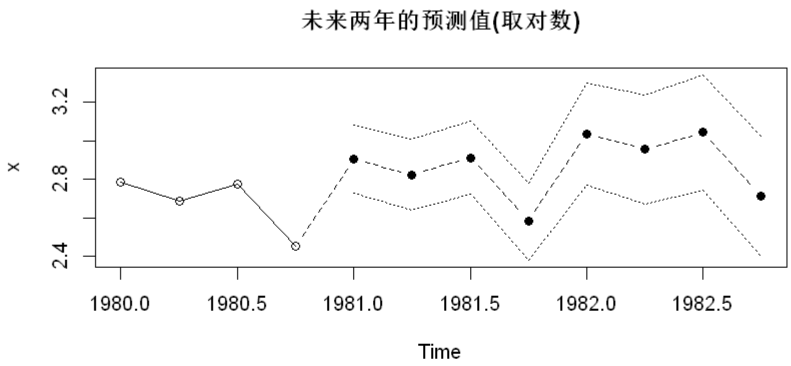
各种检验都表明残差正常。

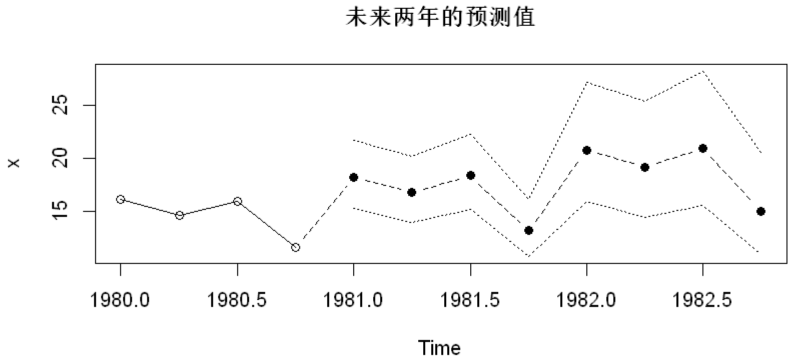
(h)

计算并画出序列未来两年的预测值，要求给出预测极限。

Solution.

In [12]:
```
par(mfrow = c(2,1))

plot(model,
     n1=c(1980,1),
     n.ahead=2*4,
     pch=19,
     main='未来两年的预测值(取对数)')

plot(model,
     n1=c(1980,1),
     n.ahead=2*4,
     pch=19,
     transform=exp,
     main='未来两年的预测值')
```





预测情况展示如上。

预测点为黑色实心点，预测极限为黑色虚线。

对GARCH(1,1)模型

$$r_t=\sqrt{h_t}\nu_t,\quad h_t=\alpha_0+\beta_1 h_{t-1}+\alpha_1 r_{t-1}^2$$

其中$\nu_t\sim t(d)/\sqrt{d/(d-2)}$（$d>2$），$t(d)$为自由度为$d$的$t$-分布，除以$\sqrt{d/(d-2)}$是为了单位方差。令$\boldsymbol{\theta}=(\alpha_0,\alpha_1,\beta_1)^\top$

- 计算对数似然函数$\ell(\boldsymbol{\theta})$
- 设定$\alpha_0=0.02,\alpha_1=0.25,\beta_1=0.35,h_1=0.05,d=4$，生成样本量为1000的服从上述GARCH(1,1)过程的序列，并画图
- 基于上面这个样本，写一个R程序找到$\boldsymbol{\theta}$的MLE和相应的标准差。

注：自由度为d的t分布的pdf为

$$f(t)=\frac{\Gamma(\frac{d+1}{2})}{\sqrt{d\pi}\Gamma(\frac{d}{2})}\left(1+\frac{t^2}{d}\right)^{-\frac{d+1}{2}},\ \Gamma(t)=\int_0^\infty x^{t-1}e^{-x}dx$$

Sol.

1).

$$l(\theta)=\log f(r_1\dots r_n)=\log[f(r_n|r_{n-1}\dots r_1)\cdot f(r_{n-1}|\dots)\cdot\dots\cdot f(r_2|r_1)f(r_1)]$$

$$=\sum_{t=1}^n\log[f(r_t|r_{t-1}\dots r_1)]$$

由于 $f(r_t|r_{t-1}\dots r_1)=f(\sqrt{d/(d-2)}\,\nu_t=\frac{r_t}{\sqrt{h_t}}\cdot\sqrt{d/(d-2)}\,|\,r_{t-1}\dots r_1)$

由于 $\sqrt{d/(d-2)}\,\nu_t\sim t(d)$，所以 $f(r_t|\dots)=\frac{\Gamma(\frac{d+1}{2})}{\sqrt{d\pi}\,\Gamma(\frac{d}{2})}(1+\frac{1}{d-2}\frac{r_t^2}{h_t})^{-\frac{d+1}{2}}$

$\therefore \log f(r_t|r_{t-1}\dots r_1)=\log\frac{\Gamma(\frac{d+1}{2})}{\sqrt{d\pi}\,\Gamma(\frac{d}{2})}-\frac{d+1}{2}\log(1+\frac{r_t^2}{(d-2)h_t})$

$\therefore \log(f(r_1\dots r_n))=\sum_{t=1}^n[\log\frac{\Gamma(\frac{d+1}{2})}{\sqrt{d\pi}\,\Gamma(\frac{d}{2})}-\frac{d+1}{2}\log(1+\frac{r_t^2}{(d-2)h_t})]$

$\Rightarrow l(\theta)=n\log\frac{\Gamma(\frac{d+1}{2})}{\sqrt{d\pi}\,\Gamma(\frac{d}{2})}-\frac{d+1}{2}\sum_{t=1}^n\log(1+\frac{r_t^2}{(d-2)h_t})$ 为对数似然函数

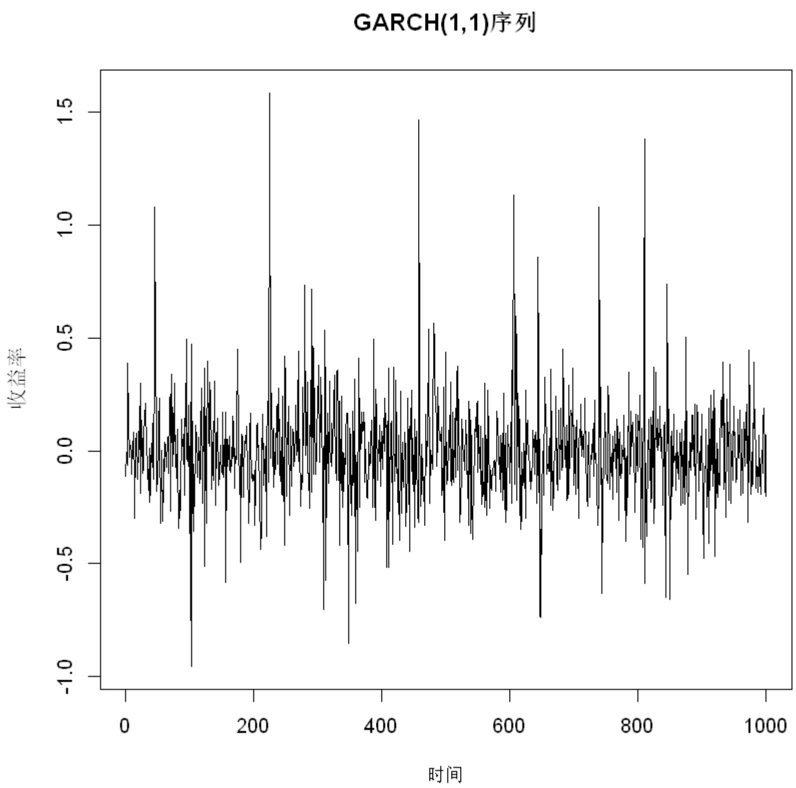
2)

设定 $\alpha_0 = 0.002,\ \alpha_1 = 0.25,\ \beta_1 = 0.35,\ h_1 = 0.05,\ d = 4$，生成样本量为 1000 的服从上述 GARCH(1,~1)过程的序列，并画图。

Solution.

In [16]:

```
# 设置参数
set.seed(123)  # 随机种子确保可重复性
n <- 1000      # 样本量
alpha0 <- 0.02
alpha1 <- 0.25
beta1 <- 0.35
h1 <- 0.05
d <- 4         # t分布自由度

# 初始化序列
h <- numeric(n)
r <- numeric(n)
h[1] <- h1
r[1] <- sqrt(h1) * rt(1, df = d) / sqrt(d / (d - 2))

set.seed(114514)
# 生成GARCH(1,1)序列
for (t in 2:n) {
  h[t] <- alpha0 + beta1 * h[t - 1] + alpha1 * r[t - 1]^2
  r[t] <- sqrt(h[t]) * rt(1, df = d) / sqrt(d / (d - 2))
}

# 绘制时间序列图
plot.ts(r, main = "GARCH(1,1)序列", xlab = "时间", ylab = "收益率", col = "black
```



3)

基于上面这个样本，写一个 `R` 程序找到 θ 的 MLE 和相应的标准差。

Solution.

本题中$\theta=(\alpha_0,\alpha_1,\ \beta_1)$，可对序列作如下处理：

In [17]:
```
library(tseries)

g1 = garch(r, order = c(1, 1))
summary(g1)
```

```
 ***** ESTIMATION WITH ANALYTICAL GRADIENT ***** 


     I     INITIAL X(I)        D(I)

     1     4.752933e-02     1.000e+00
     2     5.000000e-02     1.000e+00
     3     5.000000e-02     1.000e+00

    IT   NF      F         RELDF    PRELDF    RELDX   STPPAR   D*STEP   NPRELDF
     0    1 -9.824e+02
     1    5 -9.829e+02  5.13e-04  8.59e-04  1.9e-02  1.9e+05  2.1e-03  8.20e+01
     2    8 -9.876e+02  4.69e-03  5.53e-03  2.5e-01  5.3e+00  3.4e-02  2.14e+01
     3   12 -9.878e+02  2.18e-04  1.57e-03  2.8e-02  2.3e+00  4.9e-03  5.29e+00
     4   13 -9.888e+02  1.08e-03  1.18e-03  2.0e-02  2.0e+00  4.9e-03  3.00e+00
     5   17 -1.000e+03  1.15e-02  1.41e-02  3.4e-01  2.0e+00  1.3e-01  2.92e+00
     6   18 -1.005e+03  4.55e-03  8.15e-03  3.1e-01  2.0e+00  1.3e-01  2.08e+00
     7   19 -1.008e+03  3.09e-03  8.70e-03  1.9e-01  2.0e+00  1.3e-01  1.84e+00
     8   21 -1.009e+03  5.78e-04  5.99e-03  6.1e-02  2.0e+00  5.2e-02  9.63e-02
     9   22 -1.011e+03  2.16e-03  2.04e-03  3.0e-02  2.0e+00  2.6e-02  1.59e-02
    10   24 -1.011e+03  3.84e-05  1.16e-04  1.2e-02  2.0e+00  1.0e-02  4.48e-03
    11   26 -1.011e+03  1.77e-04  2.58e-04  3.5e-02  1.7e+00  2.9e-02  1.77e-03
    12   27 -1.011e+03  3.51e-04  8.31e-04  3.3e-02  1.7e+00  2.9e-02  7.14e-03
    13   28 -1.012e+03  4.88e-04  5.43e-04  3.2e-02  1.1e+00  2.9e-02  8.18e-04
    14   29 -1.012e+03  1.13e-04  1.12e-04  3.0e-02  0.0e+00  2.4e-02  1.12e-04
    15   30 -1.012e+03  1.98e-06  1.45e-06  2.9e-03  0.0e+00  2.6e-03  1.45e-06
    16   31 -1.012e+03  1.00e-07  1.05e-07  7.8e-04  0.0e+00  5.6e-04  1.05e-07
    17   32 -1.012e+03  2.44e-09  2.83e-11  1.6e-05  0.0e+00  1.2e-05  2.83e-11
    18   33 -1.012e+03 -4.37e-11  8.84e-15  2.6e-07  0.0e+00  1.9e-07  8.84e-15

 ***** RELATIVE FUNCTION CONVERGENCE *****

 FUNCTION    -1.012016e+03   RELDX        2.631e-07
 FUNC. EVALS      33         GRAD. EVALS      18
 PRELDF       8.841e-15      NPRELDF      8.841e-15

     I      FINAL X(I)        D(I)          G(I)

     1    1.971048e-02     1.000e+00     1.763e-03
     2    3.196469e-01     1.000e+00    -5.410e-06
     3    3.554318e-01     1.000e+00     2.046e-05
```

```
Call:
garch(x = r, order = c(1, 1))

Model:
GARCH(1,1)

Residuals:
     Min       1Q   Median       3Q      Max 
-4.55328 -0.59210 -0.02266  0.48529  7.62490 

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)    
a0  0.019710    0.003095    6.368 1.91e-10 ***
a1  0.319647    0.035416    9.025  < 2e-16 ***
b1  0.355432    0.072466    4.905 9.35e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Diagnostic Tests:
        Jarque Bera Test

data:  Residuals
X-squared = 1621.6, df = 2, p-value < 2.2e-16


        Box-Ljung test

data:  Squared.Residuals
X-squared = 0.64252, df = 1, p-value = 0.4228
```

则:

- $\hat{\theta}_{MLE} = (\hat{\alpha_0},\ \hat{\alpha_1},\ \hat{\beta_1}) = (0.0197,\ 0.3196,\ 0.3554)$。三个系数的拟合结果都是显著的。
- 相应的标准差分别为：$\hat{\sigma}_{\alpha_0} = 0.0031,\ \hat{\sigma}_{\alpha_1} = 0.0354,\ \hat{\sigma}_{\beta_1} = 0.0725$。