R软件作业1参考程序及有关输出结果

March 10, 2016

1.
(1)

```
x=rnorm(100,2,3)
x
```

```
  [1] -0.45721502 -0.79364375  1.95844260  0.22885352  5.03656929
  [6]  3.70930597 -4.37917791  7.22521619  2.25805016  1.57595302
 [11]  1.38004547  3.33533599 -0.39281426 -4.15405869 -0.25472426
 [16]  4.35766446  1.00696834  1.69762199  2.13247886 -0.37540446
 [21]  2.01375499  1.14447113  2.35857610  3.63040899  3.52408490
 [26]  6.02285211  4.90564814  9.06066514 -0.55123731  5.06829290
 [31] -0.89340833  2.02017498  4.81481491 -1.53360665  3.32955448
 [36]  0.30136744  6.47343797  5.89708028  4.53474240  2.95458993
 [41]  4.67217697 -1.98330458  0.50968374  5.76352168  2.50180038
 [46]  5.48461878 -2.06841406  6.13073489  7.89841261  2.56592142
 [51] -3.12971983  2.22551421 -2.86454431  3.08228440  2.34610521
 [56]  2.58892738  6.17303029  5.99610156  6.32187038  4.53273974
 [61]  1.07379924  0.57705484 -2.01235691  0.88151531 -0.14689318
 [66]  1.08425414 -1.43492115  2.02577081  8.02881868 -1.10230673
 [71]  0.09688577  0.16665187 -0.56875295  6.18638187  4.56697240
 [76]  5.37960704 -1.52455252  0.46694160 -2.04983304  4.27421191
 [81]  1.07692440  2.01640911  6.04399265 -2.54565918  0.93141882
 [86] -1.65161154  0.98749055  3.26220480 -0.38499177  1.86173321
 [91]  4.89069400 -2.37163228  2.45194845  2.29168277 -2.77718695
 [96] -1.61557022  3.36055797 -1.52135013  3.48274903 -0.36020287
```

(2)

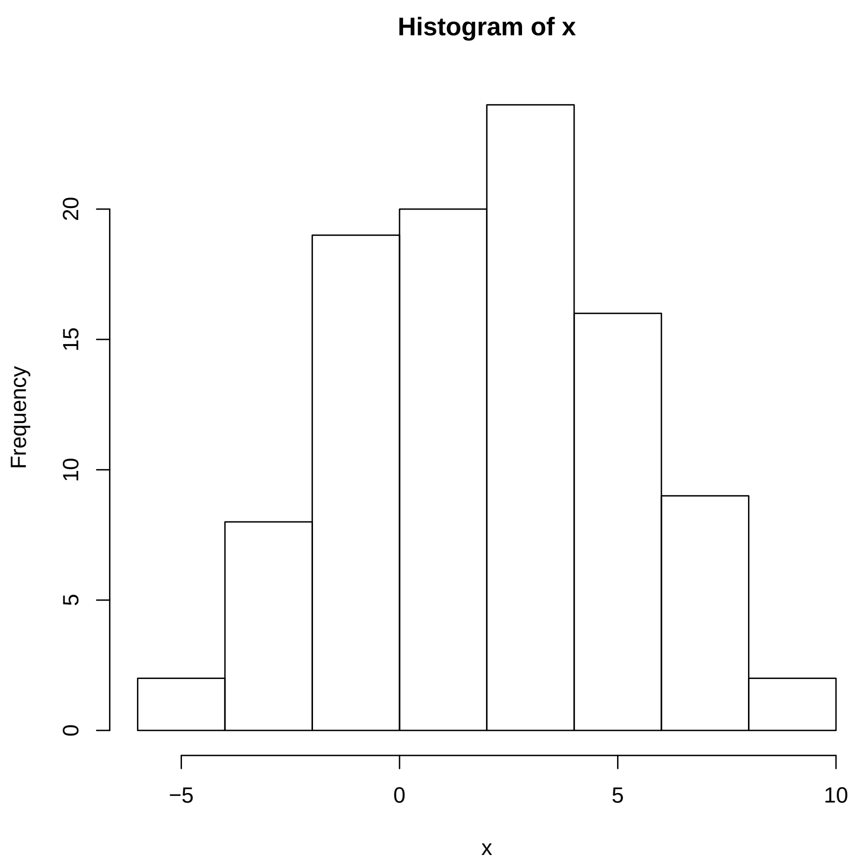
```
mean(x);var(x);sd(x);min(x);max(x);quantile(x,0.95);quantile(x,seq(0,1,0.1))
```

```
[1] 1.94318
[1] 8.871881
[1] 2.97857
[1] -4.379178
```

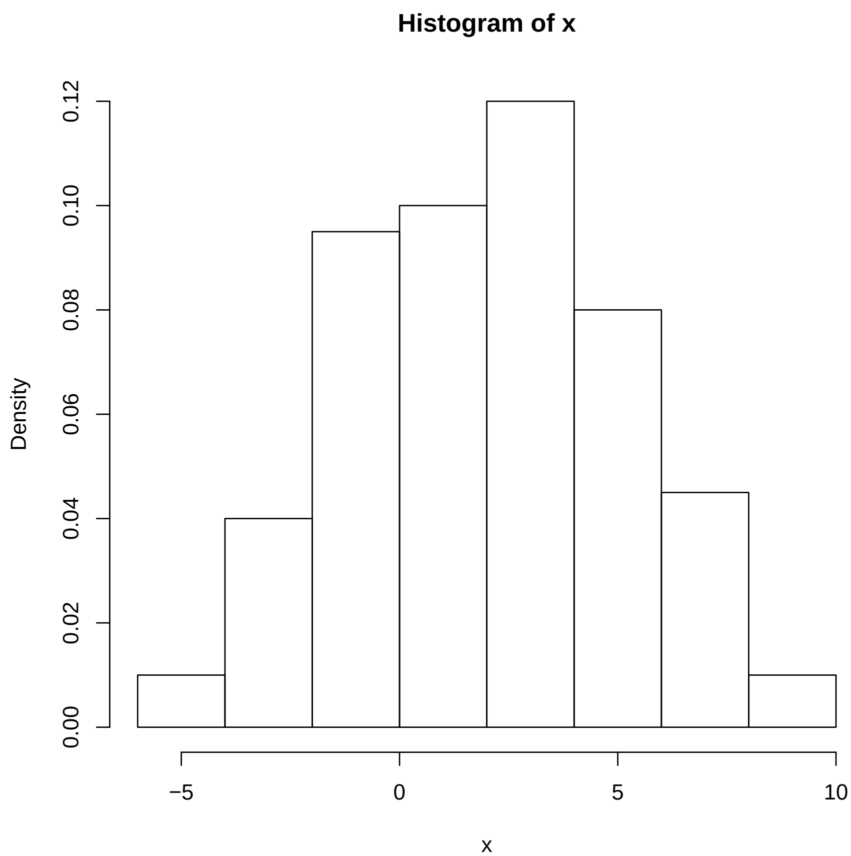
```
[1] 9.060665
      95%
6.329449
        0%        10%        20%        30%        40%        50%
-4.3791779 -1.9862098 -0.6137311  0.1457220  1.0470669  2.0150820
       60%        70%        80%        90%       100%
 2.4718892  3.4951498  4.8299907  6.0249662  9.0606651
```

(3)

```r
hist(x)
```

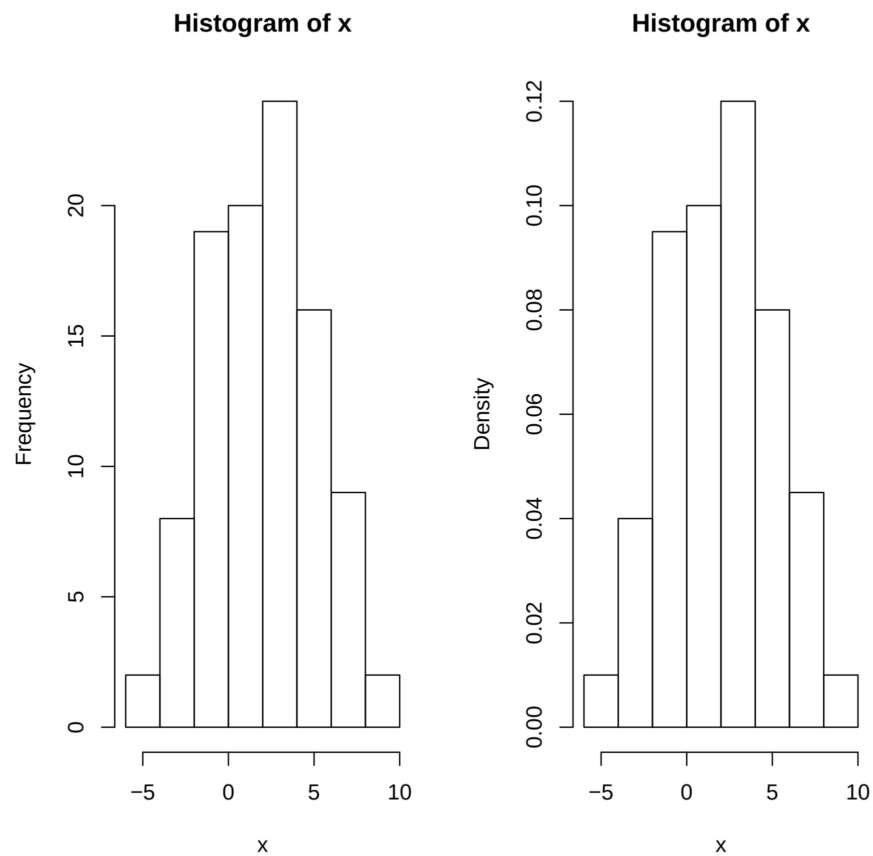



```r
hist(x,freq=F)
```



一页多图(排成一行)

```
par(mfrow=c(1,2))
hist(x);hist(x,freq=F)
```

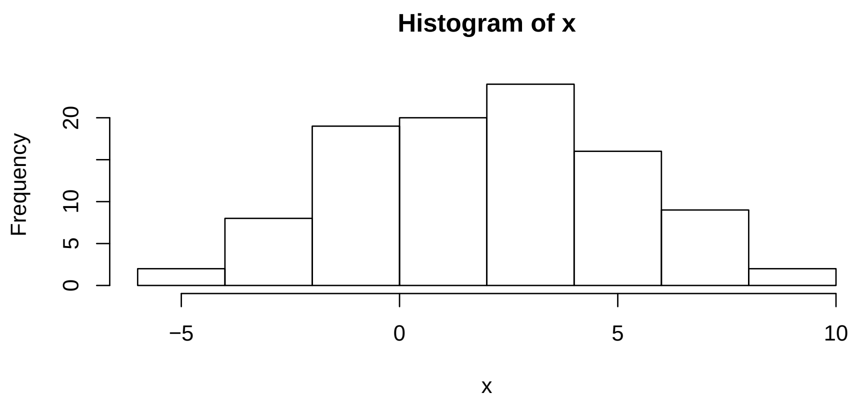


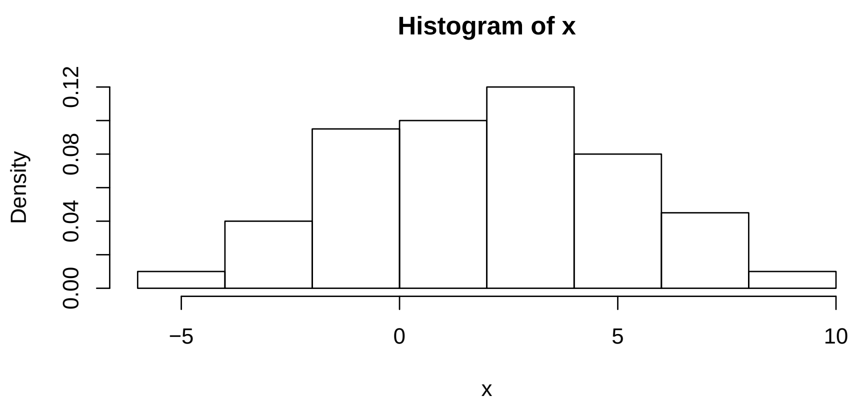
关掉该作图窗口

```
dev.off()

null device
          1
```

一页多图(排成一列)

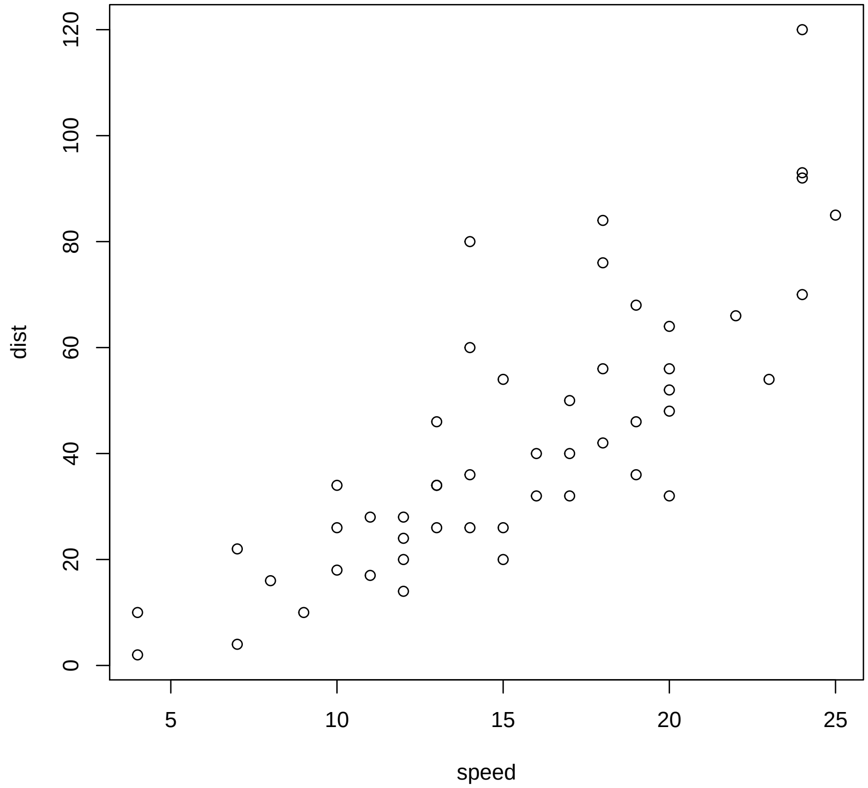
```
par(mfrow=c(2,1))
hist(x);hist(x,freq=F)
```





注：查看par的帮助，了解作图参数的含义.
2.
(1)

```
attach(cars)
plot(speed,dist)
```

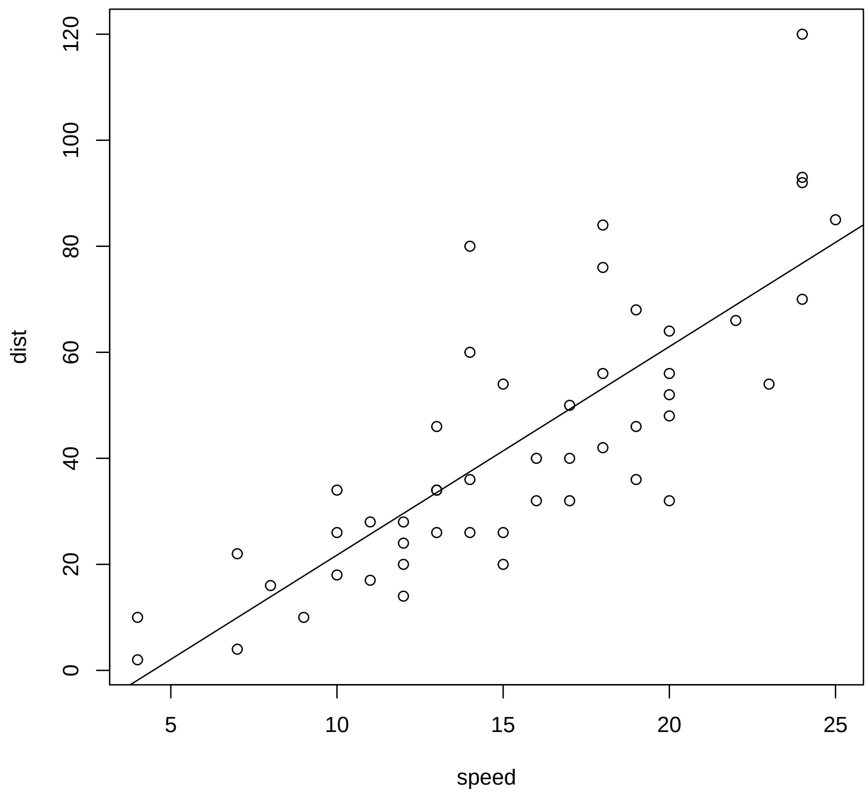


(2)

```
lm.sol=lm(dist~speed)
summary(lm.sol)


Call:
lm(formula = dist ~ speed)

Residuals:
    Min      1Q  Median      3Q     Max 
-29.069  -9.525  -2.272   9.215  43.201 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -17.5791     6.7584  -2.601   0.0123 *  
speed         3.9324     0.4155   9.464 1.49e-12 ***
```

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

Residual standard error: 15.38 on 48 degrees of freedom
Multiple R-squared:  0.6511,Adjusted R-squared:  0.6438
F-statistic: 89.57 on 1 and 48 DF,  p-value: 1.49e-12

summary(lm.sol)$coefficients[,1]

(Intercept)       speed
 -17.579095    3.932409

#or
lm.sol$coefficients

(Intercept)       speed
 -17.579095    3.932409

#or
coefficients(lm.sol)

(Intercept)       speed
 -17.579095    3.932409
```

```
#添加估计的回归直线
plot(speed,dist)
abline(lm.sol)
```



(3)

```
x=cbind(1,speed);y=dist
beta.hat=solve(t(x)%*%x,t(x)%*%y)
beta.hat

            [,1]
      -17.579095
speed   3.932409
```

(4)

```
n=length(y);p=dim(x)[2]
sigma2.hat=sum((y-x%*%beta.hat)^2)/(n-p)
sigma.hat=sqrt(sigma2.hat)
beta1.hat=beta.hat[2]
sxx=sum((x[,2]-mean(x[,2]))^2)
```

```
test.stat=beta1.hat*sqrt(sxx)/sigma.hat
test.stat

[1] 9.46399
```

```
#判断是否显著:检验统计量与临界值
alpha=0.05
critical.value=qt(1-alpha/2,n-p)
critical.value

[1] 2.010635

abs(test.stat)>=critical.value

[1] TRUE
```

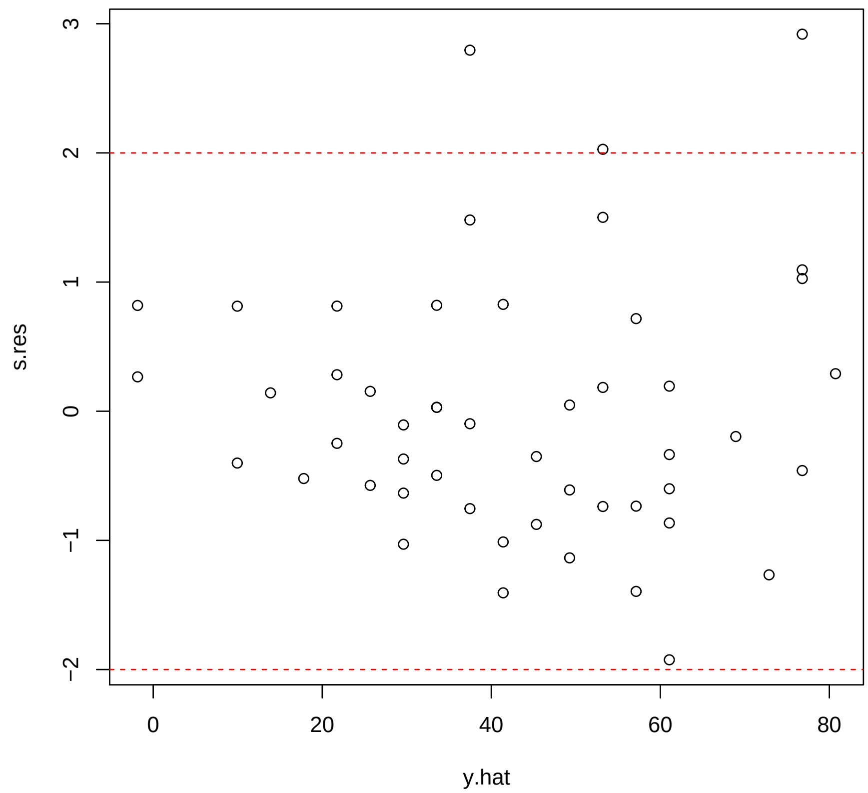
```
#判断是否显著:p-value与显著性水平
p_value=2*(1-pt(abs(test.stat),n-p))
p_value

[1] 1.489919e-12

p_value<alpha

[1] TRUE
```

(5)

```
y.hat=x%*%beta.hat;
res=y-y.hat
H=x%*%solve(t(x)%*%x)%*%t(x)
h=diag(H)
s.res=res/(sigma.hat*sqrt(1-h));s.res

             [,1]
 [1,]  0.26604155
 [2,]  0.81893273
 [3,] -0.40134618
 [4,]  0.81326629
 [5,]  0.14216236
 [6,] -0.52115255
 [7,] -0.24869180
 [8,]  0.28256008
 [9,]  0.81381197
[10,] -0.57409795
```

```
[11,]  0.15366341
[12,] -1.02971654
[13,] -0.63392061
[14,] -0.37005667
[15,] -0.10619272
[16,] -0.49644946
[17,]  0.03013240
[18,]  0.03013240
[19,]  0.82000518
[20,] -0.75422016
[21,] -0.09692637
[22,]  1.48057874
[23,]  2.79516632
[24,] -1.40612757
[25,] -1.01201579
[26,]  0.82717256
[27,] -0.87627072
[28,] -0.35074918
[29,] -1.13552237
[30,] -0.60956963
[31,]  0.04787130
[32,] -0.73777117
[33,]  0.18409193
[34,]  1.50103921
[35,]  2.02781813
[36,] -1.39503525
[37,] -0.73502818
[38,]  0.71698735
[39,] -1.92452335
[40,] -0.86524066
[41,] -0.60041999
[42,] -0.33559931
[43,]  0.19404203
[44,] -0.19590672
[45,] -1.26671228
[46,] -0.45938126
[47,]  1.02713306
[48,]  1.09470190
[49,]  2.91906038
[50,]  0.29053451
```

```
plot(y.hat,s.res)
abline(h=c(-2,2),col=2,lty=2)
```



可见绝大多数点都落在了两条虚线内，说明符合高斯马尔科夫假设.

3.

(1)

```
student=read.table("F:/第二章(2015yang)/exam0203.txt",head=T)
student

      Name Sex Age Height Weight
1     Alice   F  13   56.5   84.0
2     Becka   F  13   65.3   98.0
3      Gail   F  14   64.3   90.0
4     Karen   F  12   56.3   77.0
5     Kathy   F  12   59.8   84.5
6      Mary   F  15   66.5  112.0
7     Sandy   F  11   51.3   50.5
8    Sharon   F  15   62.5  112.5
9     Tammy   F  14   62.8  102.5
```

```
10  Alfred   M  14   69.0  112.5
11    Duke   M  14   63.5  102.5
12   Guido   M  15   67.0  133.0
13   James   M  12   57.3   83.0
14 Jeffrey   M  13   62.5   84.0
15    John   M  12   59.0   99.5
16  Philip   M  16   72.0  150.0
17  Robert   M  12   64.8  128.0
18  Thomas   M  11   57.5   85.0
19 William   M  15   66.5  112.0
```

(2)

```
mean(student$Height);sum(student$Height)

[1] 62.33684
[1] 1184.4

#or
mean(student[,4]);sum(student[,4])

[1] 62.33684
[1] 1184.4
```

(3)

```
tapply(student$Weight,student$Sex,mean)

        F         M
 90.11111 108.95000
```

(4)

```
tapply(student$Weight,student[,2:3],mean)

   Age
Sex   11     12 13     14     15  16
  F 50.5  80.75 91  96.25 112.25  NA
  M 85.0 103.50 84 107.50 122.50 150
```

4.
(1)

```
x=-10:10
x

 [1] -10  -9  -8  -7  -6  -5  -4  -3  -2  -1   0   1   2   3   4   5   6
[18]   7   8   9  10
```

(2)

```
n=length(x)
y=numeric(n)
y[x>=0]=x[x>=0]^2+1
y[x<0]=sin(x[x<0])
y

 [1]   0.5440211  -0.4121185  -0.9893582  -0.6569866   0.2794155
 [6]   0.9589243   0.7568025  -0.1411200  -0.9092974  -0.8414710
[11]   1.0000000   2.0000000   5.0000000  10.0000000  17.0000000
[16]  26.0000000  37.0000000  50.0000000  65.0000000  82.0000000
[21] 101.0000000
```

5.
(1)

```
P=matrix(c(0.1,0.2,0.3,0.4,0.4,0.1,0.2,0.3,0.3,0.4,0.1,0.2,0.2,0.3,0.4,0.1),4,4,byrow=T)
apply(P,1,sum)

[1] 1 1 1 1
```

(2)

```
A=P
n=2;i=1  #取不同的n,可得P^n
while(i<=n-1){
  A=A%*%P
  i=i+1
}
A

     [,1] [,2] [,3] [,4]
[1,] 0.26 0.28 0.26 0.20
[2,] 0.20 0.26 0.28 0.26
[3,] 0.26 0.20 0.26 0.28
[4,] 0.28 0.26 0.20 0.26
```

编写函数

```
matrix.power=function(P,n){
  A=P
  i=1
  while(i<=n-1){
    A=A%*%P
    i=i+1
```

```r
  }
  A
}
matrix.power(P,n=13)
```

```
     [,1] [,2] [,3] [,4]
[1,] 0.25 0.25 0.25 0.25
[2,] 0.25 0.25 0.25 0.25
[3,] 0.25 0.25 0.25 0.25
[4,] 0.25 0.25 0.25 0.25
```

尝试不同的n,当n >= 13 时，矩阵中的所有元素均为0.25。

方法二：利用矩阵对角化 $\mathrm{P}=\mathrm{DBD}^{-1}$，其中 B 为由特征值构成的对角阵, D 为对应的特征向量构成的正交阵,则 $\mathrm{P}^{\mathrm{n}}=\mathrm{DB}^{n}D^{-1}$。

```r
ev=eigen(P)
B=ev$values
D=ev$vectors
Pn=D%*%(diag(B))^3%*%solve(D)  #例：取n=3
Re(Pn) #此处矩阵P的特征值和特征向量为复数，故取其实部
```

```
      [,1]  [,2]  [,3]  [,4]
[1,] 0.256 0.244 0.240 0.260
[2,] 0.260 0.256 0.244 0.240
[3,] 0.240 0.260 0.256 0.244
[4,] 0.244 0.240 0.260 0.256
```