随机模拟方法与应用导论作业七

陈稼霖 45875852

2019-10-05

7.5 (Hubble's Law)

In 1929 Edwin Hubble investigated the relationship between distance and velocity of celestial objects. Knowledge of this relationship might give clues as to how the universe was formed and what may happen in the future. Hubble's Law is

$$\text{Recession Velocity} = H_0 \times \text{Distance}$$

where H_0 is Hubble's constant. This model is a straight line through the origin with slope H_0. Data that Hubble used to estimate the constant H_0 are given on the DASL web at http://lib.stat.cmu.edu/DASL/Datafiles/Hubble.html. Use the data to estimate Hubble's constant by simple linear regression.

首先读取储存有数据的文件`hubble.xls`，数据如下

```
hubble = read.table('hubble.txt',header = TRUE)
hubble
```

```
##     distance recession_velocity
## 1      0.032                170
## 2      0.034                290
## 3      0.214               -130
## 4      0.263                -70
## 5      0.275               -185
## 6      0.275               -220
## 7      0.450                200
## 8      0.500                290
## 9      0.500                270
## 10     0.630                200
## 11     0.800                300
## 12     0.900                -30
## 13     0.900                650
## 14     0.900                150
## 15     0.900                500
## 16     1.000                920
## 17     1.100                450
## 18     1.100                500
```

```
## 19    1.400              500
## 20    1.700              960
## 21    2.000              500
## 22    2.000              850
## 23    2.000              800
## 24    2.000             1090
```

然后调用函数lm，用过原点的一条直线拟合上面的数据，展示拟合直线的斜率和图像

```
M1 = lm(hubble$recession_velocity ~ 0 + hubble$distance)
M1
```

```
##
## Call:
## lm(formula = hubble$recession_velocity ~ 0 + hubble$distance)
##
## Coefficients:
## hubble$distance
##           423.9
```

```
plot(hubble,main = 'The Relationship Between Distance and Velocity of Celestial Objects')
abline(M1)
```

The Relationship Between Distance and Velocity of Celestial Object
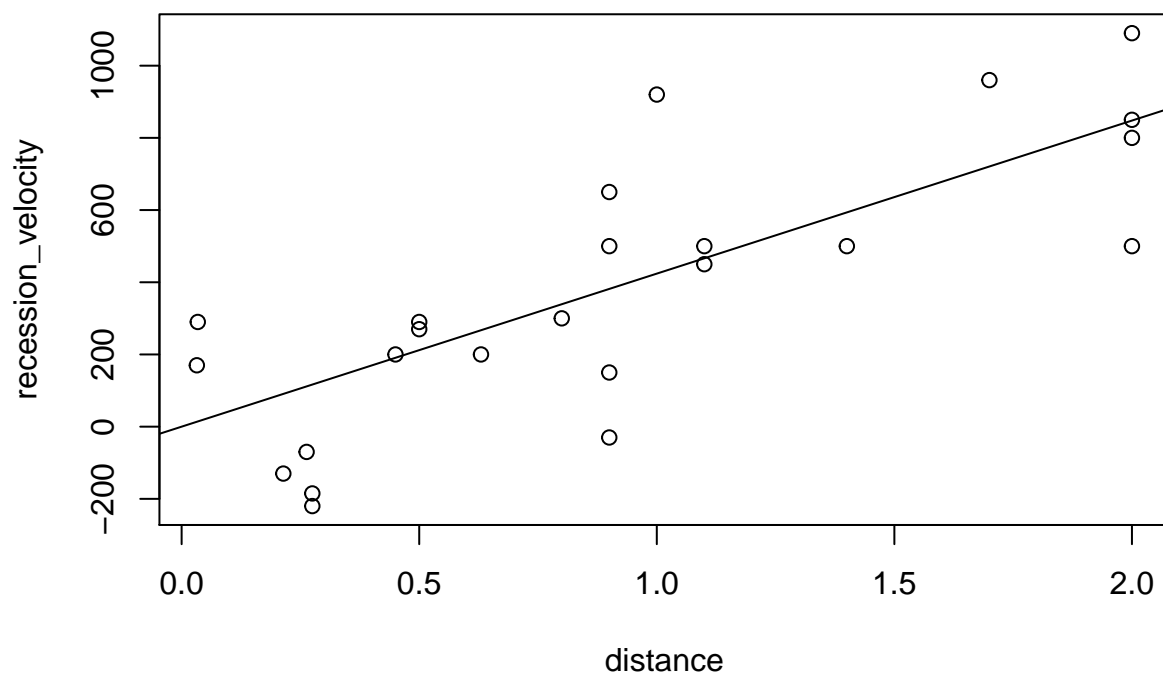
由此可见，哈勃常数（拟合直线的斜率）$H_0 = 432.9$

$$\text{Recession Velocity} = 432.9 \times \text{Distance}$$

这意味着天体距离每增加1Mpc，其推行速度增加432.9km/s。

7.10 (lunatics data)

Refer to the "lunatics" data in Example 7.8. Repeat the analysis, after deleting the two counties that are offshore islands, `NANTUCKET` and `DUCKS` counties. Compare the estimates of slope and intercept with those obtained in Example 7.8. Construct the plots and analyze the residuals as in Example 7.8.

首选读取储存有数据的文件`lunatics.txt`，数据如下

```
lunatics = read.table('lunatics.txt',header = TRUE)
lunatics
```

```
##           COUNTY NBR DIST     POP PDEN PHOME
## 1     BERKSHIRE 119  97  26.656   56    77
## 2      FRANKLIN  84  62  22.260   45    81
## 3     HAMPSHIRE  94  54  23.312   72    75
## 4       HAMPDEN 105  52  18.900   94    69
## 5     WORCESTER 351  20  82.836   98    64
## 6     MIDDLESEX 357  14  66.759  231    47
## 7         ESSEX 377  10  95.004 3252    47
## 8       SUFFOLK 458   4 123.202 3042     6
## 9       NORFOLK 241  14  62.901  235    49
## 10      BRISTOL 158  14  29.704  151    60
## 11     PLYMOUTH 139  16  32.526   91    68
## 12    BARNSTABLE 78  44  16.692   93    76
## 13    NANTUCKET  12  77   1.740  179    25
## 14        DUKES  19  52   7.524   46    79
```

去除COUNTY为NANTUCKET和DUKES的行

```
lunatics = lunatics[which(lunatics$COUNTY != 'NANTUCKET' & lunatics$COUNTY != 'DUKES'),]
lunatics
```

```
##           COUNTY NBR DIST     POP PDEN PHOME
## 1     BERKSHIRE 119  97  26.656   56    77
## 2      FRANKLIN  84  62  22.260   45    81
## 3     HAMPSHIRE  94  54  23.312   72    75
## 4       HAMPDEN 105  52  18.900   94    69
## 5     WORCESTER 351  20  82.836   98    64
```

```
## 6    MIDDLESEX 357    14  66.759  231    47
## 7       ESSEX 377    10  95.004 3252    47
## 8     SUFFOLK 458     4 123.202 3042     6
## 9     NORFOLK 241    14  62.901  235    49
## 10    BRISTOL 158    14  29.704  151    60
## 11   PLYMOUTH 139    16  32.526   91    68
## 12 BARNSTABLE  78    44  16.692   93    76
```

重复例7.8的分析过程，首先绘制在家照料的疯人百分比PHOME到最近的精神卫生中心的距离DIST的散点图并计算两者的相关系数
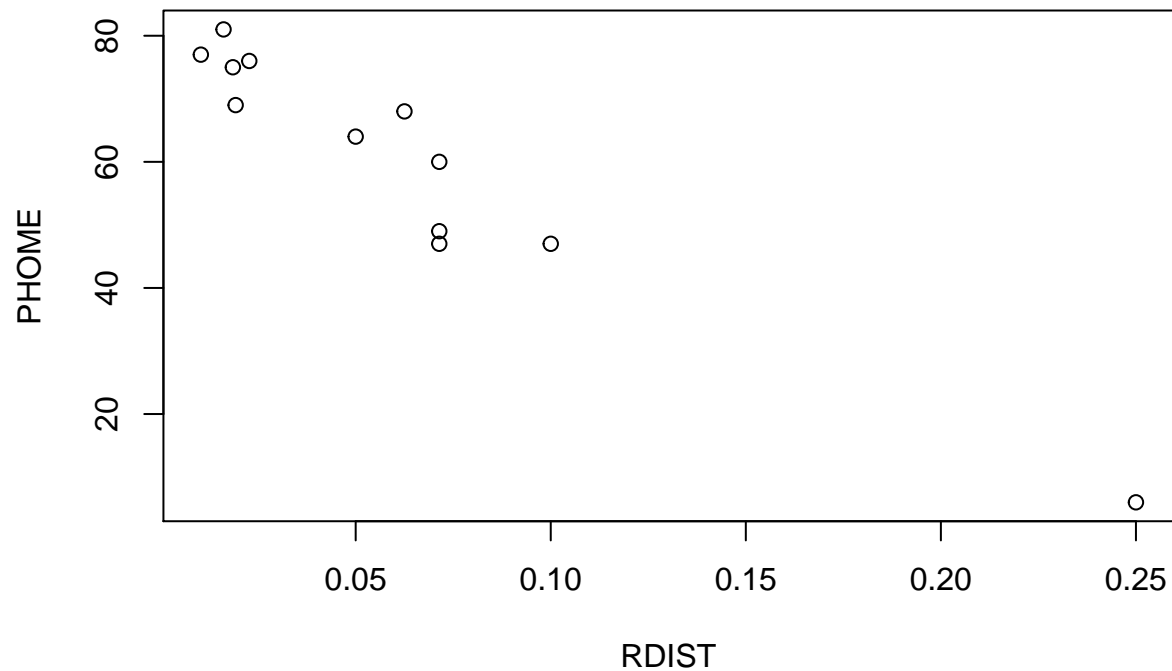
```
attach(lunatics)
plot(DIST,PHOME)
```
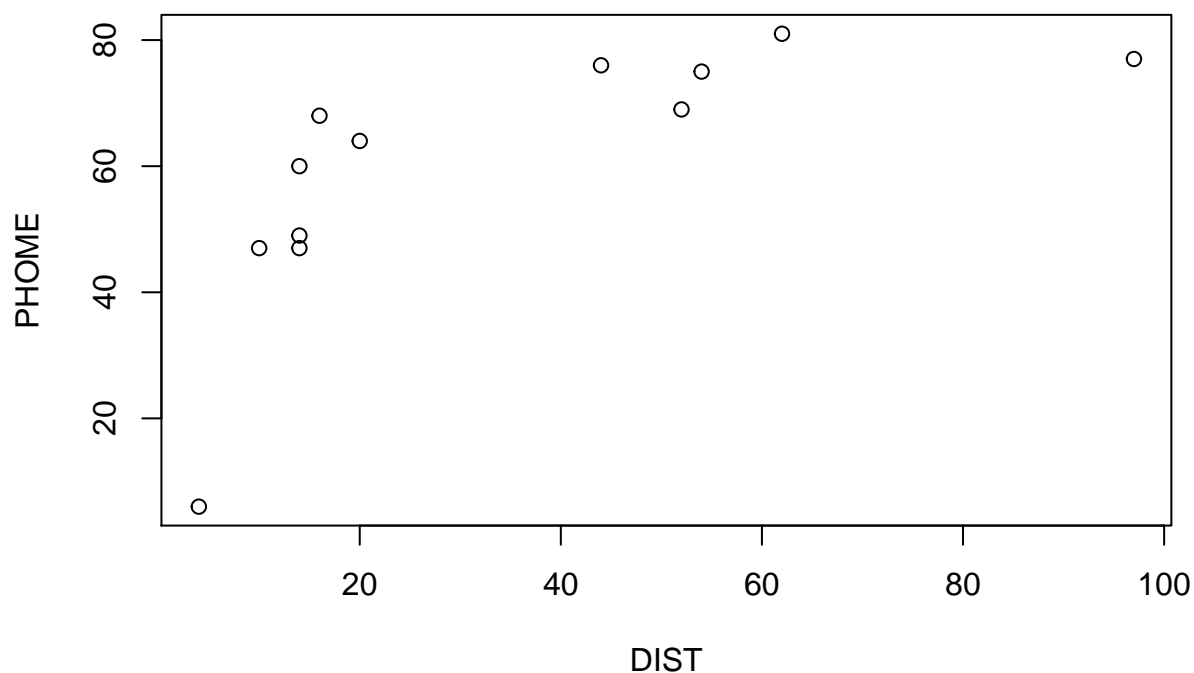


```
cor(DIST,PHOME)
```

```
## [1] 0.6963078
```

由此可见，PHOME和DIST之间的关系似乎并非线性，而更可能是双曲线。因此，取DIST的倒数RDIST，绘制PHOME关于RDIST的散点图，并计算两者之间的相关系数。

```
RDIST = 1 / DIST
plot(RDIST,PHOME)
```



```
cor(RDIST,PHOME)
```

```
## [1] -0.963456
```

在这里$|\text{cor(RDIST,PHOME)}| > |\text{cor(DIST,PHOME)}|$，说明PHOME和RDIST之间相对于PHOME和RDIST之间具有更强的线性相关性。因此，我们用简单的线性回归模型

$$\text{PHOME}_i = \beta_0 + \beta_1 \text{RDIST}_i + \varepsilon_i$$

来拟合两者之间的关系，调用函数lm计算相关的参数并绘制相应的拟合线

```
M2 = lm(PHOME ~ RDIST)
M2
```
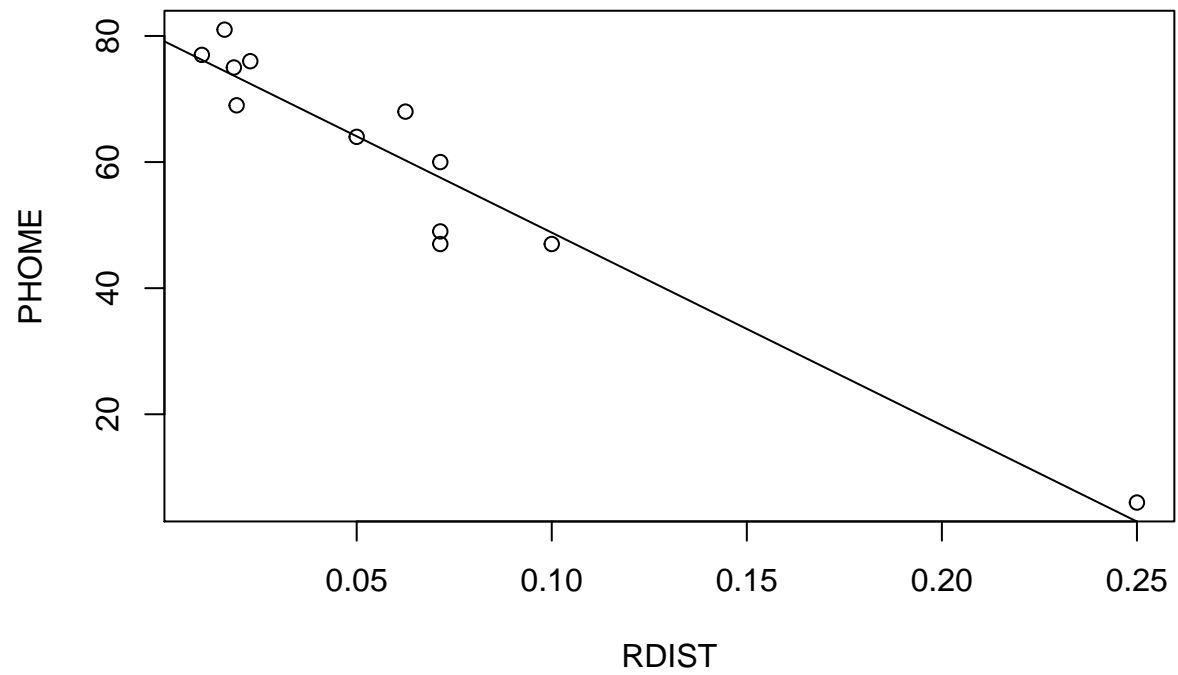
```
##
## Call:
## lm(formula = PHOME ~ RDIST)
##
## Coefficients:
## (Intercept)        RDIST
```
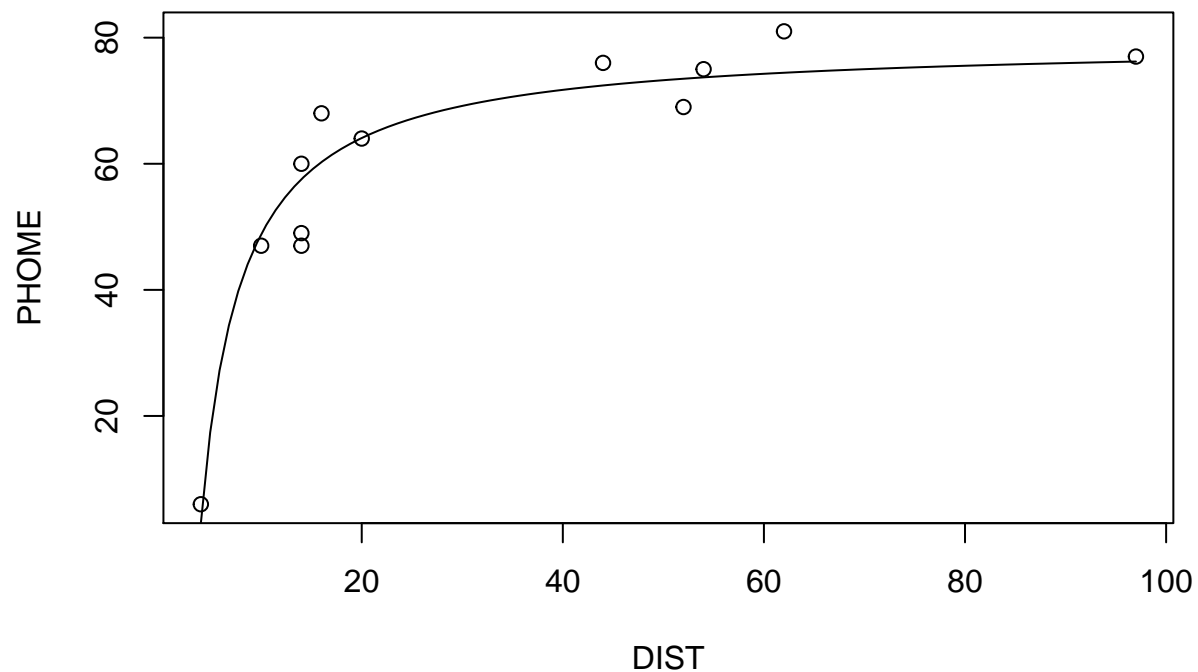
```
##      79.36       -305.52
```

```
plot(RDIST,PHOME)
abline(M2)
```
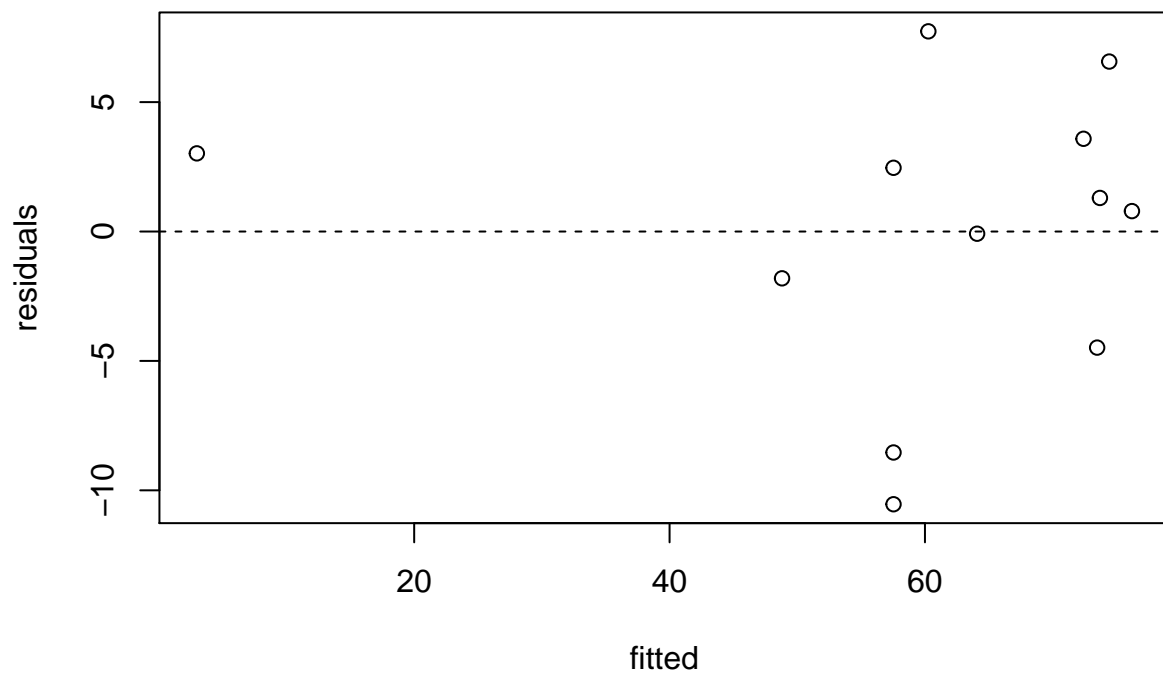


```
plot(DIST,PHOME)
curve(M2$coef[1] + M2$coef[2] / x,add = TRUE)
```

与例7.8中斜率73.93，截距−266.32相比，当去除NANTUCKET和DUCKS两个点后，获得的截距79更大，截距−305.52更小，由图可见，拟合得到的曲线更加贴合数据的走势。

仿照例7.8，绘制残差图

```
plot(M2$fitted.values,M2$residuals,xlab = 'fitted',ylab = 'residuals')
abline(h = 0,lty = 2)
```

相比例7.8，当去除NANTUCKET和DUCKS两个点后，残差范围更小，说明拟合效果更优。