Bài 11.

Hồi quy tương quan

1. Mô hình hồi quy tuyến tính đơn

Mô hình hồi quy tuyến tính đơn cho các cặp dữ liệu (x_i, y_i) như sau

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Các hệ số β_0 và β_1 chưa biết và sẽ được ước lượng từ dữ liệu.

Người ta dùng phương pháp bình phương bé nhất để tìm các ước lượng này. Cụ thể,

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Ví dụ 1: Nhịp tim tối đa

Nhịp tim tối đa của một người được cho rằng (theo kinh nghiệm) có mối quan hệ với tuổi tác theo phương trình sau

Max = 220 – Age

Người ta khảo sát nhịp tim tối đa của15 người có độ tuổi khác nhau. Dữ liệu sau được ghi lại:

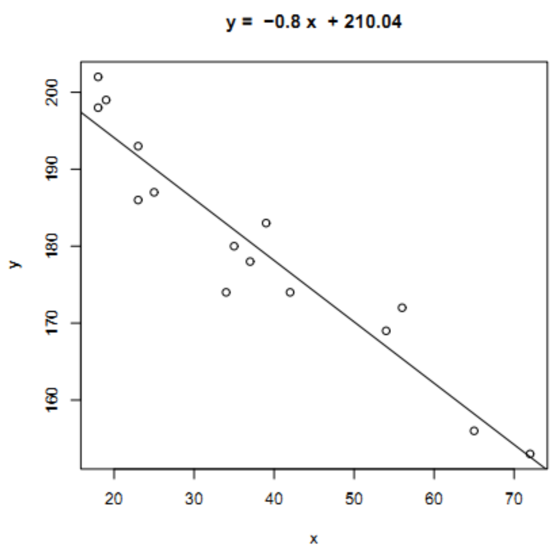
Age	18	23	25	35	65	54	34	56	72	19	23	42	18	39	37
Max Rate	202	186	187	180	156	169	174	172	153	199	193	174	198	183	178

Hãy biểu diễn dữ liệu trên và vẽ đường hồi quy đơn Max Rate theo Age trên cùng một đồ thị?

Giải:

```
x = c(18,23,25,35,65,54,34,56,72,19,23,42,18,39,37)    #nhập dữ liệu

y = c(202,186,187,180,156,169,174,172,153,199,193,174,198,183,178)

plot(x,y)              # vẽ đồ thị

abline(lm(y ~ x))      # vẽ đường hồi quy

> lm(y ~ x)            # các giá trị cơ bản của phân tích hồi quy

Call:
```

```
lm(formula = y ~ x)

Coefficients:

(Intercept)      x

210.0485   -0.7977
```



Hình: Hồi quy nhịp tim tối đa theo tuổi

hoặc

```
> result=lm(y ~ x)

> summary(result)

Call:

lm(formula = y ~ x)

...

Coefficients:

Estimate    Std. Error t value    Pr(>|t|)

(Intercept)210.04846  2.86694    73.27      < 2e-16 ***

x            -0.79773   0.06996    -11.40     3.85e-08 ***

...
```

Nếu muốn lấy thông tin về thặng dư (residuals) ta dùng `resid`, về hệ số (coefficients) ta dùng `coef`. Ví dụ,

```
> coef(result)              # hoặc sử dụng result[['coef']]
(Intercept)           x
210.0484584  -0.7977266
> res = resid(result)       # hoặc result[['resid']]
> summary(res)
      Min.    1st Qu.    Median       Mean    3rd Qu.      Max.
-8.926e+00 -2.538e+00 3.879e-01 -1.628e-16  3.187e+00  6.624e+00
```

Phân tích thặng dư

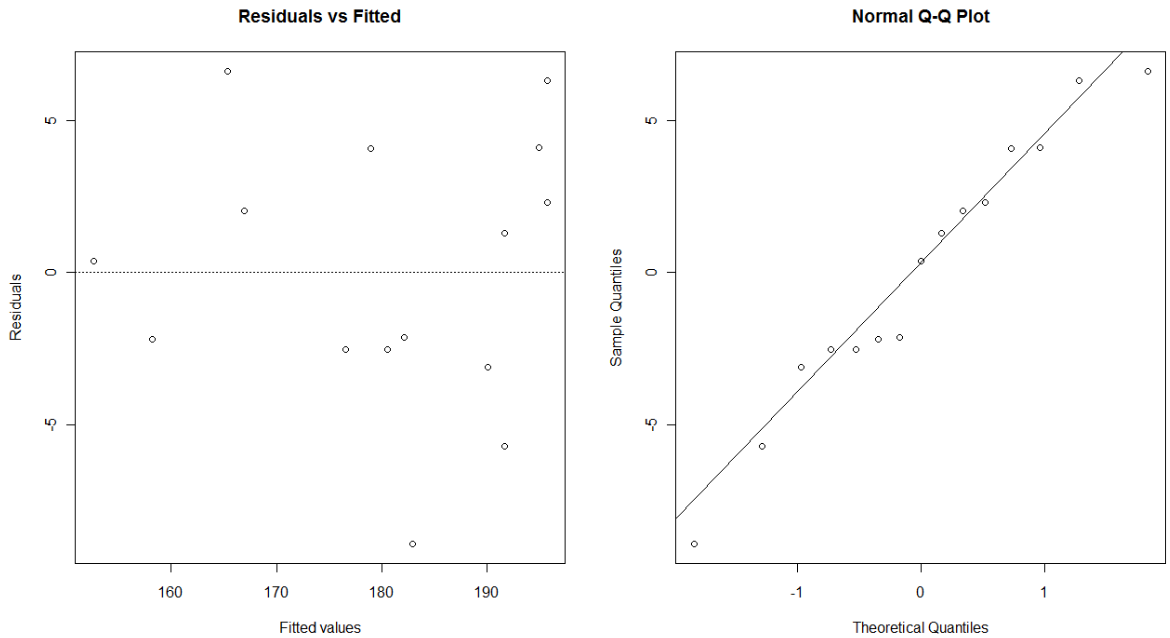
Ta dựa vào đồ thị biểu diễn thặng dư theo giá trị hồi quy ($\hat{y}_i$) và đồ thị Normal Q-Q.

Ví dụ 2: Nhịp tim tối đa (tt)

Hãy phân tích giá trị thặng dư của mô hình hồi quy.

Giải:

```
x = c(18,23,25,35,65,54,34,56,72,19,23,42,18,39,37)    #nhập dữ liệu
y = c(202,186,187,180,156,169,174,172,153,199,193,174,198,183,178)
result = lm(y ~ x)
par(mfrow=c(1,2))     #chuẩn bị vẽ hai đồ thị trên 1 cửa sổ
plot(result$fitted.values,resid(result),xlab ='Fitted values',
ylab ='Residuals', main = 'Residuals vs Fitted')
                #đồ thị thặng dư theo giá trị hồi quy
abline(h=0,lty=3)     #đường thẳng y = 0 với nét chấm
qqnorm(res)           #đồ thị Normal Q-Q
qqline(res)           #đường thẳng lí thuyết trên đồ thị Normal Q-Q
```



Hình: Phân tích thặng dư

Nhìn vào đồ thị **Residuals vs Fitted** ta thấy thặng dư phân tán quanh trục Ox một cách ngẫu nhiên đồng đều. Do đó thặng dư có kì vọng 0 và phương sai không đổi.

Nhìn vào đồ thị **Normal Q-Q Plot** ta thấy thặng dư gần xấp xỉ đường thẳng. Do đó thặng dư tuân theo phân phối chuẩn.

2. Khoảng tin cậy cho mô hình hồi quy:

- Cho β_1

$$\hat{\beta}_1 - t_{1-\alpha/2}^{n-2}\sqrt{\frac{MSE}{S_{xx}}} \le \beta_1 \le \hat{\beta}_1 + t_{1-\alpha/2}^{n-2}\sqrt{\frac{MSE}{S_{xx}}}$$

Trong đó,

$$MSE = \frac{SSE}{n-2}$$

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2$$

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

- Cho β_0

$$\hat{\beta}_0 - t_{1-\alpha/2}^{n-2}\sqrt{MSE\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)} \le \beta_0 \le \hat{\beta}_0 + t_{1-\alpha/2}^{n-2}\sqrt{MSE\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}$$

- Cho trung bình biến đáp ứng

$$\hat{\mu}_{Y|x_0} - t_{1-\alpha/2}^{n-2}\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}}\right)} \le \mu_{Y|x_0} \le \hat{\mu}_{Y|x_0} + t_{1-\alpha/2}^{n-2}\sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}}\right)}$$

Trong đó,

$$\hat{\mu}_{Y|x_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

$$\hat{\sigma}^2 = \frac{SSE}{n-2}$$

- Dự báo giá trị quan trắc mới Y_0 ứng với x_0

$$\hat{Y}_0 - t_{1-\alpha/2}^{n-2}\sqrt{\hat{\sigma}^2\left(1 + \frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}}\right)} \le Y_0 \le \hat{Y}_0 + t_{1-\alpha/2}^{n-2}\sqrt{\hat{\sigma}^2\left(1 + \frac{1}{n} + \frac{(\bar{x} - x_0)^2}{S_{xx}}\right)}$$

Ví dụ 3: Nhịp tim tối đa (tt)

Xác định khoảng tin cậy 95% cho β_1

Giải:

```
x = c(18,23,25,35,65,54,34,56,72,19,23,42,18,39,37)    #nhập dữ liệu
y = c(202,186,187,180,156,169,174,172,153,199,193,174,198,183,178)
n = length(x)
result = lm(y~x)
res   = resid(result)              # các thặng dư của result
b1    = (coef(result))[['x']]      # hệ số độ dốc của đường hồi quy
MSE   = sum( res^2 ) / (n-2)
Sxx   = sum((x-mean(x))^2)
eps   = qt(1-0.05/2,n-2)*sqrt(MSE/Sxx) #sai số ước lượng
ci    = c(can.duoi = b1 - eps, can.tren = b1 + eps)
ci                                 #khoảng tin cậy 95% cho hệ số độ dốc
```

3. Kiểm định giả thuyết cho mô hình hồi quy:

- Cho β_0

Các giả thuyết

$$\begin{cases} H_0: \beta_0 = b_0 \\ H_1: \beta_0 \neq b_0 \end{cases} \qquad \begin{cases} H_0: \beta_0 = b_0 \\ H_1: \beta_0 < b_0 \end{cases} \qquad \begin{cases} H_0: \beta_0 = b_0 \\ H_1: \beta_0 > b_0 \end{cases}$$

Thống kê kiểm định

$$t = \frac{\hat{\beta}_0 - b_0}{\sqrt{MSE\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}} \sim t(n-2)\ khi\ H_0\ \text{đúng}$$

Miền bác bỏ (mức ý nghĩa α)

Đối thuyết	Miền bác bỏ	p-giá trị
$H_0: \beta_0 \neq b_0$	$\lvert t \rvert > t_{1-\alpha/2}^{n-2}$	$2P(T_{n-2} \geq \lvert t \rvert)$
$H_0: \beta_0 < b_0$	$t < -t_{1-\alpha}^{n-2}$	$P(T_{n-2} \leq t)$
$H_0: \beta_0 > b_0$	$t > t_{1-\alpha}^{n-2}$	$P(T_{n-2} \geq t)$

- Cho β_1

Các giả thuyết

$$\begin{cases} H_0: \beta_1 = b_1 \\ H_1: \beta_1 \neq b_1 \end{cases} \qquad \begin{cases} H_0: \beta_1 = b_1 \\ H_1: \beta_1 < b_1 \end{cases} \qquad \begin{cases} H_0: \beta_1 = b_1 \\ H_1: \beta_1 > b_1 \end{cases}$$

Thống kê kiểm định

$$t = \frac{\hat{\beta}_1 - b_1}{\sqrt{\frac{MSE}{S_{xx}}}} \sim t(n-2)\ khi\ H_0\ \text{đúng}$$

Miền bác bỏ (mức ý nghĩa α)

Đối thuyết	Miền bác bỏ	p-giá trị
$H_0: \beta_1 \neq b_1$	$\lvert t \rvert > t_{1-\alpha/2}^{n-2}$	$2P(T_{n-2} \geq \lvert t \rvert)$
$H_0: \beta_1 < b_1$	$t < -t_{1-\alpha}^{n-2}$	$P(T_{n-2} \leq t)$
$H_0: \beta_1 > b_1$	$t > t_{1-\alpha}^{n-2}$	$P(T_{n-2} \geq t)$

Ví dụ 4: Nhịp tim tối đa (tt)

Có người cho rằng độ dốc của đường hồi quy là -1. Với dữ liệu thu thập được, ta có thể chấp nhận ý kiến này không? $\alpha = 5\%$

Giải:

Ta sẽ kiểm định các giả thuyết $\begin{cases} H_0: \beta_1 = -1 \\ H_1: \beta_1 \neq -1 \end{cases}$ bằng thủ tục trong R như sau

```
x = c(18,23,25,35,65,54,34,56,72,19,23,42,18,39,37)    #nhập dữ liệu

y = c(202,186,187,180,156,169,174,172,153,199,193,174,198,183,178)
```

```
n = length(x)

result = lm(y~x)

res   = resid(result)             # các thặng dư của result

b1    = (coef(result))[['x']]     # hệ số độ dốc của đường hồi quy

MSE   = sum( res^2 ) / (n-2)

Sxx   = sum((x-mean(x))^2)

t     = (b1 - (-1) ) / sqrt(MSE/Sxx)

alpha = 0.05

ifelse(abs(t) > qt(1-alpha/2,n-2), 'Bác bỏ H0', 'Chấp nhận H0')

[1] "Bác bỏ H0"
```

4. Kiểm định giả thuyết cho hệ số tương quan:

Các giả thuyết

$$\begin{cases} H_0: \rho = 0 \\ H_1: \rho \neq 0 \end{cases} \qquad \begin{cases} H_0: \rho = 0 \\ H_1: \rho < 0 \end{cases} \qquad \begin{cases} H_0: \rho = 0 \\ H_1: \rho > 0 \end{cases}$$

Thống kê kiểm định

$$t = r\sqrt{\frac{n-2}{1-r^2}} \sim t(n-2)\ khi\ H_0\ \text{đúng}$$

Với

$$r^2 = \frac{SSR}{SST}$$

$$SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Miền bác bỏ (mức ý nghĩa α)

Đối thuyết	Miền bác bỏ	p-giá trị
$H_1: \rho \neq 0$	$\|t\| > t_{1-\alpha/2}^{n-2}$	$2P(T_{n-2} \geq \|t\|)$
$H_1: \rho < 0$	$t < -t_{1-\alpha}^{n-2}$	$P(T_{n-2} \leq t)$
$H_1: \rho > 0$	$t > t_{1-\alpha}^{n-2}$	$P(T_{n-2} \geq t)$

Ví dụ 5: Nhịp tim tối đa (tt)

Có người cho rằng nhịp tim tối đa và tuổi tác không có tương quan gì với nhau. Với dữ liệu thu thập được, ta có thể chấp nhận ý kiến này không? $\alpha = 5\%$

Giải:

Ta sẽ kiểm định các giả thuyết $\begin{cases} H_0: \rho = 0 \\ H_1: \rho \neq 0 \end{cases}$ bằng thủ tục trong R như sau

```
x = c(18,23,25,35,65,54,34,56,72,19,23,42,18,39,37)    #nhập dữ liệu

y = c(202,186,187,180,156,169,174,172,153,199,193,174,198,183,178)

n = length(x)

result = lm(y~x)

res   = resid(result)       # các thặng dư của result

SST   = sum((y - mean(y))^2)

SSR   = SST - sum(res^2)          # SSR = SST - SSE

r.sq  = SSR/SST                   # tính r^2

t     = sqrt(r.sq*(n-2)/1-r.sq)

alpha = 0.05

ifelse(abs(t) > qt(1-alpha/2,n-2), 'Bác bỏ H0', 'Chấp nhận H0')

[1] "Bác bỏ H0"
```

Bài tập:

1. Giá một căn nhà (đv: 1000 USD) phụ thuộc vào số phòng ngủ trong căn nhà đó. Giả sử rằng dữ liệu sau được ghi lại cho các căn nhà ở một thành phố

Price	300	250	400	550	317	389	425	289	389	559
No. bedrooms	3	3	4	5	4	3	6	3	4	5

(a) Vẽ đồ thị phân tán và đường hồi quy trên cùng một hệ trục tọa độ.
(b) Kiểm định giả thuyết cho rằng khi thêm một phòng ngủ thì chi phí tăng thêm 60.000 USD với đối thuyết là chi phí cao hơn.

2. Ta biết rằng càng uống nhiều bia, mức cồn trong máu (blood alcohol level - BAL) càng tăng. Giả sử ta có dữ liệu sau về mức tiêu thụ bia và mức cồn trong máu tương ứng ở 10 sinh viên được khảo sát

Student	1	2	3	4	5	6	7	8	9	10

Beers	5	2	9	8	3	7	3	5	3	5
BAL	0.10	0.03	0.19	0.12	0.04	0.095	0.07	0.06	0.02	0.05

(a) Vẽ đồ thị phân tán và đường thẳng hồi quy trên cùng một hệ trục tọa độ.
(b) Kiểm định giả thuyết cho rằng uống thêm một chai bia làm tăng BAL lên 0.02 với đối thuyết cho rằng nhỏ hơn. Chọn mức ý nghĩa 5%
(c) Kiểm định giả thuyết cho rằng khi không uống bia thì mức cồn trong máu bằng 0, với đối thuyết hai phía. Chọn mức ý nghĩa 5%

3. Khi độ cao tăng lên nhiệt độ không khí giảm xuống. Theo kinh nghiệm của một số người, nhiệt độ sẽ giảm 9.8 °C/km. Để kiểm tra xem nhận định này có đúng hay không, một nhóm sinh viên mang theo nhiệt kế và thiết bị đo độ cao và ghi lại được dữ liệu sau trong một chuyến leo núi của họ.

Elevation(ft)	600	1000	1250	1600	1800	2100	2500	2900
Temperature(F)	56	54	56	50	47	49	47	45

(a) Vẽ đồ thị phân tán và đường hồi quy trên cùng hệ trục tọa độ.
(b) Kiểm định giả thuyết tốc độ giảm nhiệt là 9.8 °C/km với đối thuyết cho rằng nhỏ hơn.

4. Một động cơ tên lửa được chế tạo bằng cách kết hợp hai loại chất nổ đẩy là kíp nổ và nhiên liệu đốt cháy. Lực đẩy *y* của cách kết hợp này được cho rằng là một hàm tuyến tính theo tuổi (đv: tuần) của chất nổ đẩy *x* khi động cơ được thử nghiệm. Dữ liệu được ghi lại trong file "*rocket.motor.csv*".
   (a) Vẽ đồ thị phân tán cho dữ liệu thu được. Mô hình hồi quy đơn có phù hợp trong trường hợp này?
   (b) Tìm ước lượng bình phương bé nhất cho hệ số độ dốc (slope) và hệ số chặn (intercept) của đường hồi quy. Tìm ước lượng cho σ^2.
   (c) Ước lượng lực đẩy của động cơ được làm từ chất nổ đẩy với 20 tuần tuổi.
   (d) Vẽ đồ thị $\hat{y}_i$ (giá trị hồi quy) theo y_i (giá trị quan sát). Quan hệ tuyến tính giữa lực đẩy và tuổi có được xác định một cách hoàn hảo (không sai số) hay không? Từ đồ thị, hãy cho biết biến hồi quy tuổi (age) có phải là một lựa chọn hợp lí cho mô hình này?
   (e) Kiểm định giả thuyết $H_0: \beta_1 = -30$ và đối thuyết $H_1: \beta_1 \neq -30$ với $\alpha = 0.01$. Tìm p-value cho kiểm định này.
   (f) Kiểm định giả thuyết $H_0: \beta_0 = 0$ và đối thuyết $H_1: \beta_0 \neq 0$ với $\alpha = 0.01$. Tìm p-value cho kiểm định này.
   (g) Kiểm định giả thuyết $H_0: \beta_0 = 2500$ và đối thuyết $H_1: \beta_0 > 2500$ với $\alpha = 0.01$. Tìm p-value cho kiểm định này.
   (h) Tìm khoảng tin cậy 95% cho hệ số độ dốc β_1 và hệ số chặn β_0.
   (i) Tìm khoảng tin cậy 95% cho lực đẩy trung bình khi $x = 20$ tuần tuổi.
   (j) Tìm khoảng dự báo 95% cho lực đẩy khi $x = 20$ tuần tuổi.

(k) Phân tích thặng dư của mô hình.

5. Một nghiên cứu về sự xuất hiện của Natri và clorua trên bề mặt các dòng sông ở trung tâm Rhode Island. Dữ liệu được ghi lại trong file *"chloride.csv"* là nồng độ clorua (mg/l) y và diện tích đường giao thông ở đầu nguồn x (theo tỉ lệ phần trăm).
   (a) Vẽ đồ thị phân tán cho dữ liệu. Mô hình hồi quy tuyến tính đơn có phù hợp ở đây hay không?
   (b) Vẽ đường hồi quy trên cùng đồ thị ở câu (a). Tìm ước lượng của σ^2.
   (c) Ước lượng nồng độ clorua trung bình cho một đầu nguồn có diện tích giao thông chiếm 1%.
   (d) Tìm giá trị hồi quy tương ứng với $x = 0.47$ và giá trị thặng dư của nó.
   (e) Kiểm định giả thuyết $H_0: \beta_1 = 0$ và đối thuyết $H_1: \beta_1 \neq 0$ với $\alpha = 0.01$. Tìm p-value cho kiểm định này.
   (f) Kiểm định giả thuyết $H_0: \beta_0 = 0$ và đối thuyết $H_1: \beta_0 \neq 0$ với $\alpha = 0.01$. Hãy rút ra kết luận. Mô hình có phù hợp với dữ liệu hơn nếu ta loại bỏ hệ số chặn (intercept) ?
   (g) Tìm khoảng tin cậy 99% cho β_1 và β_0.
   (h) Tìm khoảng tin cậy 99% cho trung bình nồng độ clorua khi diện tích đường giao thông $x = 1.0\%$.
   (i) Tìm khoảng dự báo 99% cho nồng độ clorua khi diện tích đường giao thông $x = 1.0\%$.
   (j) Phân tích thặng dư của mô hình

6. Dữ liệu của một nghiên cứu về mối quan hệ giữa phơi nhiễm tiếng ồn và tăng huyết áp được ghi lại trong file *"noise.csv"*.
   (a) Vẽ đồ thị phân tán của y (huyết áp tính bằng milimet thủy ngân) theo x (cường độ âm thanh tính bằng decibels). Mô hình hồi quy đơn có phù hợp trong trường hợp này?
   (b) Vẽ đường thẳng hồi quy trên cùng hệ trục tọa độ ở câu (a). Tìm ước lượng của σ^2.
   (c) Tìm mức huyết áp trung bình tương ứng với cường độ âm thanh 85 decibals.
   (d) Kiểm định giả thuyết $H_0: \beta_0 = 0$ và đối thuyết $H_1: \beta_0 \neq 0$ với $\alpha = 0.05$. Tìm p-value cho kiểm định này.
   (e) Tìm khoảng tin cậy 95% cho β_0 và β_1.
   (f) Tìm khoảng tin cậy 95% cho huyết áp trung bình khi cường độ âm thanh là 85 decibals.
   (g) Tìm khoảng dự báo 95% cho huyết áp khi cường độ âm thanh là 85 decibals.
   (h) Một người cho rằng phơi nhiễm tiếng ồn và tăng huyết áp không tương quan với nhau. Hãy kiểm định giả thuyết trên với mức ý nghĩa 5%.
   (i) Phân tích thặng dư của mô hình

7. Khối lượng hơi nước được sử dụng hàng tháng của một nhà máy hóa chất được cho rằng có liên quan đến nhiệt độ môi trường trung bình (°F) của tháng đó. Dữ liệu về lượng hơi nước sử dụng và nhiệt độ được ghi lại trong file *"temperature.csv"*.
   (a) Giả sử rằng mô hình hồi quy tuyến tính đơn là phù hợp, hãy vẽ đồ thị phân tán và đường hồi quy lượng hơi nước sử dụng (y) theo nhiệt độ trung bình (x) trên cùng một đồ thị. Tìm ước lượng của σ^2.
   (b) Lượng hơi nước sử dụng khi nhiệt độ trung bình là 55°F là bao nhiêu?
   (c) Khi nhiệt độ trung bình hàng tháng tăng 1°F thì lượng hơi nước sử dụng trung bình thay đổi bao nhiêu?
   (d) Giả sử nhiệt độ trung bình hàng tháng là 47°F. Tính giá trị hồi quy của y và giá trị thặng dư tương ứng.
   (e) Kiểm định giả thuyết $H_0: \beta_1 = 10$ và đối thuyết $H_1: \beta_1 \neq 10$ với $\alpha = 0.01$. Tìm p-value cho kiểm định này.
   (f) Kiểm định định giả thuyết $H_0: \beta_0 = 0$ và đối thuyết $H_1: \beta_0 \neq 0$ với $\alpha = 0.01$. Tìm p-value cho kiểm định này và rút ra kết luận.
   (g) Tìm khoảng tin cậy 99% cho β_1 và β_0.
   (h) Tìm khoảng tin cậy 95% cho mức hơi nước sử dụng trung bình khi nhiệt độ trung bình là 55°F.
   (i) Tìm khoảng dự báo cho mức hơi nước sử dụng khi nhiệt độ là 55°F. Giải thích tại sao khoảng dự báo này rộng hơn khoảng tin cậy trong câu (h).
   (j) Phân tích thặng dư của mô hình.

8. Dữ liệu về giá nhà và thuế hàng năm cho 24 căn nhà được ghi lại trong file *"house.price.csv"*.
   (a) Giả sử rằng mô hình hồi quy tuyến tính là phù hợp, hãy vẽ đồ thị phân tán và đường hồi quy giá nhà theo thuế trên cùng một đồ thị. Tìm ước lượng của σ^2.
   (b) Tìm giá bán nhà trung bình nếu biết thuế phải trả hàng năm là $x = 7.50$
   (c) Tính giá trị hồi quy y tương ứng với $x = 5.8980$. Tìm thặng dư tương ứng.
   (d) Tính giá trị $\hat{y}_i$ cho mỗi giá trị x_i. Vẽ đồ thị $\hat{y}_i$ theo y_i và nhận xét về hình dạng của đồ thị nếu quan hệ giữa y và x là một đường thẳng (không có sai số ngẫu nhiên). Từ đồ thị ta có thể cho rằng thuế hàng năm là một biến hồi quy hiệu quả để dự đoán giá nhà hay không?
   (e) Kiểm định giả thuyết $H_0: \beta_1 = 0$, với $\alpha = 0.05$
   (f) Kiểm định giả thuyết $H_0: \beta_0 = 0$, với $\alpha = 0.05$
   (g) Tìm khoảng tin cậy 95% cho β_1 và β_0.
   (h) Tính khoảng dự báo cho giá nhà khi tiền thuế phải trả hàng năm là $x = 7.50$
   (i) Từ dữ liệu, có thể cho rằng giá nhà và thuế hàng năm không tương quan nhau hay không? Mức ý nghĩa 5%
   (j) Phân tích thặng dư của mô hình.