

BÀI TẬP VỀ NHÀ  
MÔN HỌC: HỒI QUY TUYẾN TÍNH

**Bài 2:** Bảng số liệu

Số thứ tự	Độ bền dẻo ( $Y_i$ )	Độ dày vật liệu ( $X_{i1}$ )	Mật độ vật liệu ( $X_{i2}$ )
1	37,8	4	4,0
2	22,5	4	3,6
3	17,1	3	3,1
4	10,8	2	3,2
5	7,2	1	3,0
6	42,3	6	3,8
7	30,2	4	3,8
8	19,4	4	2,9
9	14,8	1	3,8
10	9,5	1	2,8
11	32,4	3	3,4
12	21,6	4	2,8

*(Học viên sử dụng phần mềm R để thực hiện bài tập)*

**Câu hỏi 1:** Tìm 2 phương trình đường thẳng hồi quy và 1 phương trình siêu phẳng (Nếu có)?

Trả lời:

Nhập liệu R:

```
Y<-c(37.8, 22.5, 17.1, 10.8, 7.2, 42.3, 30.2, 19.4, 14.8, 9.5, 32.4, 21.6)
```

```
X1<-c(4, 4, 3, 2, 1, 6, 4, 4, 1, 1, 3, 4)
```

```
X2<-c(4.0, 3.6, 3.1, 3.2, 3.0, 3.8, 3.8, 2.9, 3.8, 2.8, 3.4, 2.8)
```

Ta xây dựng mô hình hồi quy như sau:

a. **Mô hình 1:**  $Y = \beta_0 + \beta_1 X_1 + \epsilon$

Mô hình đường thẳng hồi quy:  $Y = \widehat{\beta}_0 + \widehat{\beta}_1 X_1$

Code:

```
Y1<-lm(Y~X1)
summary(Y1)

Call:
lm(formula = Y ~ X1)

Residuals:
    Min       1Q   Median       3Q      Max
-8.266 -4.887 -1.208  3.232 10.770

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.523      4.383   0.804 0.440237
X1             6.036      1.279   4.721 0.000816 ***
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.633 on 10 degrees of freedom
Multiple R-squared:  0.6903,    Adjusted R-squared:  0.6593
F-statistic: 22.29 on 1 and 10 DF,  p-value: 0.0008155
```

Kết quả:

$$\widehat{\beta}_0 = 3,523 \text{ và } \widehat{\beta}_1 = 6,036$$

$$\Rightarrow Y = \widehat{\beta}_0 + \widehat{\beta}_1 X_1 = 3,523 + 6,036 X_1$$

**b. Mô hình 2:  $Y = \beta_0 + \beta_2 X_2 + \epsilon$**

Mô hình đường thẳng hồi quy:  $Y = \widehat{\beta}_0 + \widehat{\beta}_2 X_2$

Code:

```
Y2<-lm(Y~X2)
summary(Y2)

Call:
lm(formula = Y ~ X2)

Residuals:
    Min       1Q   Median       3Q      Max
-15.1923  -5.1780  -0.2298   6.1123  12.3077

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -36.373     20.489  -1.775   0.1062
X2             17.464       6.069   2.878   0.0164 *
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.815 on 10 degrees of freedom
Multiple R-squared:  0.453,    Adjusted R-squared:  0.3983
F-statistic: 8.282 on 1 and 10 DF,  p-value: 0.01645
```

Kết quả:

$$\widehat{\beta}_0 = -36,373 \text{ và } \widehat{\beta}_2 = 17,464$$

$$\Rightarrow Y = \widehat{\beta}_0 + \widehat{\beta}_2 X_2 = -36,373 + 17,464 X_2$$

c. **Mô hình 3:**  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$

Mô hình siêu phẳng hồi quy:  $Y = \widehat{\beta}_0 + \widehat{\beta}_1 X_1 + \widehat{\beta}_2 X_2$

Code:

```
Y3<-lm(Y~X1+X2)
summary(Y3)

Call:
lm(formula = Y ~ X1 + X2)

Residuals:
    Min       1Q   Median       3Q      Max
-6.897 -2.135 -1.126  1.714 10.122

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -30.081     11.455   -2.626  0.027542 *
X1              4.905       1.014    4.838  0.000923 ***
X2             11.072       3.621    3.058  0.013617 *
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.897 on 9 degrees of freedom
Multiple R-squared:  0.8481,    Adjusted R-squared:  0.8143
F-statistic: 25.12 on 2 and 9 DF,  p-value: 0.0002075
```

Kết quả:

$$\widehat{\beta}_0 = -30,081 \text{ và } \widehat{\beta}_1 = 4,905 \text{ và } \widehat{\beta}_2 = 11,072$$

$$\Rightarrow Y = \widehat{\beta}_0 + \widehat{\beta}_1 X_1 + \widehat{\beta}_2 X_2 = -30,081 + 4,905 X_1 + 11,072 X_2$$

**Câu hỏi 2:** Xác định tỉ lệ phần trăm sự biến thiên của biến phụ thuộc cho từng mô hình có thể có trên.

Trả lời:

a. **Mô hình 1:**  $Y = \beta_0 + \beta_1 X_1 + \epsilon$

Quan sát lại:

Call:

```
lm(formula = Y ~ X1)
```

Residuals:

Min	1Q	Median	3Q	Max
-8.266	-4.887	-1.208	3.232	10.770

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.523	4.383	0.804	0.440237
X1	6.036	1.279	4.721	0.000816 ***

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.633 on 10 degrees of freedom

Multiple R-squared: 0.6903, Adjusted R-squared: 0.6593

F-statistic: 22.29 on 1 and 10 DF, p-value: 0.0008155

- **Kết quả:** Dựa vào *hệ số xác định bội (Multiple R-squared)* là 0,6903. Vậy biến giải thích  $X_1$  giải thích 69,03% sự biến thiên của độ dẻo vật liệu (Y) phụ thuộc vào độ dày vật liệu ( $X_1$ ).

**b. Mô hình 2:  $Y = \beta_0 + \beta_2 X_2 + \epsilon$**

Quan sát lại:

Call:

```
lm(formula = Y ~ X2)
```

Residuals:

Min	1Q	Median	3Q	Max
-15.1923	-5.1780	-0.2298	6.1123	12.3077

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-36.373	20.489	-1.775	0.1062
X2	17.464	6.069	2.878	0.0164 *

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.815 on 10 degrees of freedom

Multiple R-squared: 0.453, Adjusted R-squared: 0.3983

F-statistic: 8.282 on 1 and 10 DF, p-value: 0.01645

- Kết quả: Dựa vào *hệ số xác định bội (Multiple R-squared)* là 0,453. Vậy biến giải thích  $X_2$  giải thích 45,3% sự biến thiên của độ dẻo vật liệu (Y) phụ thuộc vào mật độ vật liệu ( $X_2$ ).

**c. Mô hình 3:  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$**

Quay lại mô hình:

Call:

```
lm(formula = Y ~ X1 + X2)
```

Residuals:

Min	1Q	Median	3Q	Max
-6.897	-2.135	-1.126	1.714	10.122

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-30.081	11.455	-2.626	0.027542 *
X1	4.905	1.014	4.838	0.000923 ***
X2	11.072	3.621	3.058	0.013617 *

---

Signif. codes:

0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.897 on 9 degrees of freedom

Multiple R-squared: 0.8481, Adjusted R-squared: 0.8143

F-statistic: 25.12 on 2 and 9 DF, p-value: 0.0002075

- Kết quả: Dựa vào *hệ số xác định bội (Multiple R-squared)* là 0,8481. Vậy biến giải thích  $X_2$  giải thích 84,81% sự biến thiên của độ dẻo (Y) phụ thuộc vào độ dày ( $X_1$ ) và mật độ vật liệu ( $X_2$ ).

**Câu hỏi 3:** Nếu chúng ta chỉ quan tâm đến cả hai biến giải thích, hãy lập bảng ANOVA?

Trả lời:

Bảng ANOVA

```
> anova(lm(Y~X1+X2))
```

Analysis of Variance Table

Response: Y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
x1	1	980.63	980.63	40.8959	0.000126 ***
x2	1	224.22	224.22	9.3509	0.013617 *
Residuals	9	215.81	23.98		

---

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

**Câu hỏi 4:** Kiểm định giả thiết sau với mức ý nghĩa 5%,

$$H_0: \beta_1 = \beta_2 = 0$$

Vậy đối thuyết:

$$H_1: \beta_1 \neq 0 \text{ hoặc } \beta_2 \neq 0$$

Trả lời:

Nhắc lại kiến thức:

Kiểm định giả thiết biến độc lập có ảnh hưởng lên biến phụ thuộc không?	Giả thiết: $H_0: \beta = 0$ $H_1: \beta \neq 0$ PP giá trị tới hạn: B1: Tính $T_{qs} = \frac{\hat{\beta}}{se(\hat{\beta})}$ B2: Tra bảng t-student giá trị $t_{\frac{\alpha}{2}}^{n-2}$ B3: so sánh $ T_{qs} $ và $t_{\frac{\alpha}{2}}^{n-2}$ + $ T_{qs}  > t_{\frac{\alpha}{2}}^{n-2}$ : bác bỏ $H_0 \Rightarrow$ biến độc lập ảnh hưởng lên biến phụ thuộc Y + $ T_{qs}  < t_{\frac{\alpha}{2}}^{n-2}$ : chấp nhận $H_0$	Giả thiết: $H_0: \beta = 0$ $H_1: \beta \neq 0$ PP giá trị tới hạn: B1: Tính $T_{qs} = \frac{\hat{\beta}}{se(\hat{\beta})}$ B2: Tra bảng t-student giá trị $t_{\frac{\alpha}{2}}^{n-k}$ B3: so sánh $ T_{qs} $ và $t_{\frac{\alpha}{2}}^{n-k}$ + $ T_{qs}  > t_{\frac{\alpha}{2}}^{n-k}$ : bác bỏ $H_0 \Rightarrow$ biến độc lập ảnh hưởng lên biến phụ thuộc Y + $ T_{qs}  < t_{\frac{\alpha}{2}}^{n-k}$ : chấp nhận $H_0$
	PP P-value: Lấy giá trị p-value tương ứng với biến độc lập mình đang xét Tiến hành so sánh p-value và $\alpha$ : + p-value < $\alpha$ : bác bỏ $H_0 \rightarrow$ <b>biến độc lập (X) ảnh hưởng lên biến phụ thuộc (Y)</b> + p-value > $\alpha$ : chấp nhận $H_0$	PP P-value: Lấy giá trị p-value tương ứng với biến độc lập mình đang xét Tiến hành so sánh p-value và $\alpha$ : + p-value < $\alpha$ : bác bỏ $H_0 \rightarrow$ <b>biến độc lập (X) ảnh hưởng lên biến phụ thuộc (Y)</b> + p-value > $\alpha$ : chấp nhận $H_0$

**Sử dụng P-Value trong mô hình hồi quy (3):**

Call:

```
lm(formula = Y ~ X1 + X2)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-6.897 -2.135 -1.126  1.714 10.122
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -30.081      11.455   -2.626 0.027542 *
X1              4.905       1.014    4.838 0.000923 ***
X2              11.072       3.621    3.058 0.013617 *
```

---

Signif. codes:

```
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 4.897 on 9 degrees of freedom

Multiple R-squared: 0.8481, Adjusted R-squared: 0.8143

F-statistic: 25.12 on 2 and 9 DF, p-value: 0.0002075



- P-value của  $X_1 = 0,000923 < 0.05$ , hoặc
- P-value của  $X_2 = 0,013617 < 0.05$
- Nên ta bác bỏ giả thuyết

$$H_0: \beta_1 = \beta_2 = 0$$

- Cháo nhận đối thuyết:

$$H_1: \beta_1 \neq 0 \text{ hoặc } \beta_2 \neq 0$$

**Câu hỏi 5:** Xác định khoảng tin cậy với mức ý nghĩa 5% cho  $\beta_1$  trong trường hợp mô hình chỉ có biến độc lập là độ dày của vật liệu.

Trả lời:

Nhắc lại kiến thức:

Ước lượng khoảng	Dùng công thức cho đa biến với (j=1,2)	Với độ tin cậy (1 - $\alpha$ ), khoảng tin cậy đối xứng, tối đa, tối thiểu của $\beta_j$ là: $\hat{\beta}_j - Se(\hat{\beta}_j)t_{\alpha/2}^{(n-k)} < \beta_j < \hat{\beta}_j + Se(\hat{\beta}_j)t_{\alpha/2}^{(n-k)}$ $\beta_j < \hat{\beta}_j + Se(\hat{\beta}_j)t_{\alpha}^{(n-k)}$ $\hat{\beta}_j - Se(\hat{\beta}_j)t_{\alpha}^{(n-k)} < \beta_j$
------------------	----------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

- Quan sát lại kết quả từ mô hình hồi quy:  $y = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 = 3,523 + 6,036x_1$

$$\widehat{\beta}_1 = 6,036$$

- Khoảng tin cậy với mức ý nghĩa 5% của  $\beta_1$  là:

$$[\widehat{\beta}_1 - SE(\widehat{\beta}_1)t_{2,5}^{11} < \beta_1 < \widehat{\beta}_1 + SE(\widehat{\beta}_1)t_{2,5}^{11}]$$

$$[3,187036; 8,88479]$$

Sử dụng code R, ta có kết quả:

```
> confint(Y1)
```

	2.5 %	97.5 %
(Intercept)	-6.242858	13.28806
X1	3.187036	8.88479

**Câu hỏi 6:** Với khoảng tin cậy vừa tìm được ở câu 5, chúng ta có thể khẳng định rằng hồi quy tuyến tính là có ý nghĩa giữa mức độ bền dẻo của nhựa và độ dày của vật liệu và mật độ của vật liệu không? Chứng minh điều khẳng định của bạn.

Trả lời:

- Hồi quy tuyến tính là có ý nghĩa giữa mức độ bền dẻo của nhựa và độ dày của vật liệu và mật độ của vật liệu, Vì:

- Quan sát bảng kết quả cho mô hình hồi quy hai biến:  
  
 Call:  
`lm(formula = Y ~ X1 + X2)`  
  
 Residuals:  

Min	1Q	Median	3Q	Max
-6.897	-2.135	-1.126	1.714	10.122

  
  
 Coefficients:  

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-30.081	11.455	-2.626	0.027542	*
X1	4.905	1.014	4.838	0.000923	***
X2	11.072	3.621	3.058	0.013617	*

 ---  
 Signif. codes:  
 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
 Residual standard error: 4.897 on 9 degrees of freedom  
 Multiple R-squared: 0.8481, Adjusted R-squared: 0.8143  
 F-statistic: 25.12 on 2 and 9 DF, p-value: 0.0002075

- Ta thấy  $\widehat{\beta}_1 = 4,905 \in [3,187036; 8,88479] \Rightarrow \widehat{\beta}_1$  vẫn thuộc khoảng tin cậy dành cho ước lượng hệ số hồi quy của biến  $X_1$  trong mô hình hồi quy đơn.
- Ta tính khoảng tin cậy 95% cho  $\widehat{\beta}_2$  trong mô hình hồi quy đơn:  $y = \widehat{\beta}_0 + \widehat{\beta}_2 x_2 = -36,373 + 17,464x_2$

$$\widehat{\beta}_2 = 17,464$$

- Khoảng tin cậy với mức ý nghĩa 5% của  $\beta_2$  là:

$$[\widehat{\beta}_2 - SE(\widehat{\beta}_2)t_{2,5}^{11} < \beta_2 < \widehat{\beta}_2 + SE(\widehat{\beta}_2)t_{2,5}^{11}]$$

$$[3,942487; 30,98642]$$

- Sử dụng code R, ta có kết quả:

```
> confint(Y2)
```

	2.5 %	97.5 %
(Intercept)	-82.024671	9.27949
X2	3.942487	30.98642

- Ta thấy  $\widehat{\beta}_2 = 11,072 \in [3,942487; 30,98642] \Rightarrow \widehat{\beta}_2$  vẫn thuộc khoảng tin cậy dành cho ước lượng hệ số hồi quy của biến  $X_2$  trong mô hình hồi quy đơn.
- Do đó ta kết luận như trên.