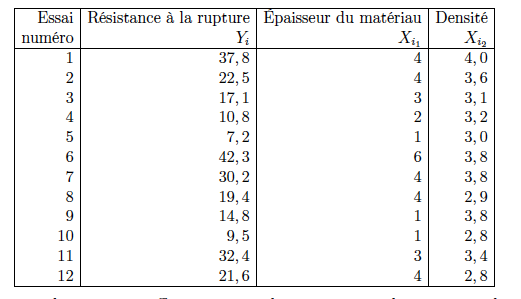
**Bài 2**

Nghiên cứu nhằm xác định mức độ bền dẻo của nhựa (Y ) có thể bị ảnh hưởng bởi độ dày của vật liệu (X1) và mật độ của vật liệu (X2). Có 12 thử nghiệm với kết quả như bảng:



Ta nhập dữ liệu của 12 mẫu thử nghiệm trên R như sau:

|  |
| --- |
| # Y: muc do ben deo cua nhua  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: do day cua vat lieu  X1 <- c(4, 4, 3, 2, 1, 6, 4, 4, 1, 1, 3, 4)  #X2: mat do cua vat lieu  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) |

1. **Các phương trình đường thẳng, mặt phẳng hồi quy**

* Mô hình 1: Biến phụ thuộc Y (mức độ bền dẻo của nhựa) bị ảnh hưởng bởi biến độc lập X1 (độ dày vật liệu).

Tính toán mô hình trên R:

|  |
| --- |
| model1 <- lm(Y~X1)  summary(model1) |

Kết quả trên R:

|  |
| --- |
| 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 |

Như vậy, đường thẳng hồi quy mô hình 1 là: **Y = 3.523 + 6.036\*X1**

*Nhận xét*: p-value của β1 trong mô hình rất nhỏ, nhỏ hơn 0.1%, rất có ý nghĩa thống kê. Có mối liên hệ tuyến tính giữa mức độ bền dẻo của nhựa và độ dày vật liệu.

Tuy nhiên, hệ số chặn β0 có p-value khá lớn 44%, nghĩa là sai số sẽ rất lớn khi dự đoán hệ số chặn β0 và khi đánh giá độ bền dẻo của nhựa Y, đặc biệt khi độ dày vật liệu nhỏ (vật liệu mỏng) thì tỷ trọng sai số càng lớn.

* Mô hình 2: Biến phụ thuộc Y (mức độ bền dẻo của nhựa) bị ảnh hưởng bởi biến độc lập X2 (mật độ vật liệu).

Tính toán mô hình trên R:

|  |
| --- |
| model1 <- lm(Y~X2)  summary(model1) |

Kết quả trên R:

|  |
| --- |
| 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 |

Như vậy, đường thẳng hồi quy mô hình 2 là: **Y =** -36.373 **+** 17.464**\*X2**

*Nhận xét*: p-value của β1 trong mô hình là 1.6%, rất có ý nghĩa thống kê. Có mối liên hệ tuyến tính giữa mức độ bền dẻo của nhựa và mật độ vật liệu.

Tương tự mô hình 1, trong mô hình 2 hệ số β0 cũng có p-value khá, hơn 10%, nghĩa là sai số sẽ rất lớn khi dự đoán hệ số chặn β0 và khi đánh giá độ bền dẻo của nhựa Y. Ngoài ra, β0 âm, giá trị gần gấp đôi β1 đặc biệt khi mật độ vật liệu nhỏ hơn 2 thì giá trị độ bền dẻo của nhựa là số âm, vô lý, ước lượng độ bền dẻo của nhựa có mật độ thấp sẽ rất khó khăn

* Mô hình 3: Biến phụ thuộc Y (mức độ bền dẻo của nhựa) bị ảnh hưởng bởi cả 2 biến độc lập X1 ((độ dày vật liệu)) và X2 (mật độ vật liệu).

Tính toán mô hình trên R:

|  |
| --- |
| model3 <- lm(Y~X1+X2)  summary(model3) |

Kết quả trên R:

|  |
| --- |
| 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 |

Như vậy, mặt phẳng hồi quy mô hình 3 là: **Y =** -30.081 **+** 4.905**\*X1 +** 11.072**\*X2**

*Nhận xét*: p-value khi kiểm định F cho sự tồn tại của mô hình rất nhỏ, nhở hơn 0.1%, rất có ý nghĩa thống kê. Có mối liên hệ tuyến tính giữa mức độ bền dẻo của nhựa với ít nhất một trong hai thông số là độ dày vật liệu và mật độ vật liệu.

Kiểm định T cho mỗi hệ số hồi quy, cũng cho kết quả tốt với mức ý nghĩa 5%. Điều này cho thấy mô hình giải thích rất tốt biến thiên mức độ bền dẻo của nhựa trên mối quan hệ tuyến tính với 2 biến độ dày vật liệu và mật độ vật liệu và nó cũng rất có ý nghĩa thống kê. Nhưng, β0 âm và giá trị khá lớn nên khi vật liệu khá mỏng và mật độ thấp thì việc đánh giá độ bền dẻo của nhựa là rất khó khăn, giá trị dự đoán âm là không chính xác.

1. **Phần trăm sự biến thiên của biến phụ thuộc của các mô hình:**

Ta chỉ xem xét Mô hình 1 và 3, mô hình 2 không có ý nghĩa thống kê.

Phần trăm sự biến thiên của biến của biến phụ thuộc mà mô hình giải thích được chính là hệ số xác định R2.

Ta lấy giá trị hệ số xác định của mô hình từ R như sau:

|  |
| --- |
| summary(model1)$r.squared  summary(model2)$r.squared  summary(model3)$r.squared |

Kết quả:

|  |
| --- |
| > summary(model1)$r.squared  [1] 0.6902638  > summary(model2)$r.squared  [1] 0.4530022  > summary(model3)$r.squared  [1] 0.8480931 |

Cả hai mô hình biến thiên của các biến độc lập giải thích 69% đối với mô hình 1 và 45% đối với mô hình 2 sự biến thiên của biến phụ thuộc. Ta thấy mô hình số 3 sử dụng cả 2 biến về hệ số xác định tăng rất đáng kể.

1. **Bảng ANOVA cho mô hình 2 biến:**

R cung cấp bảng ANOVA chi tiết cho mô hình tuyến tính đã xây dựng bằng lệnh như sau:

|  |
| --- |
| anova(model3) |

Bảng phân tích phương sai:

|  |
| --- |
| 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 |

Bảng phân tích phương sai đơn giản:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Df** | **SS** | **MS** | **F-value** |
| **R** | 2 | 1204.85 | 602.425 | 25.12314 |
| **E** | 9 | 215.81 | 23.97889 |  |
| **T** | 11 | 1420.66 |  |  |

1. **Kiểm định giả thiết thống kê:**

Giả thuyết:

H0: β1 = β2 = 0

H1: ít nhất 1 hệ số β1 hoặc β2 bằng 0

Thống kê cho kiểm định này:

Sử dụng kiểm định Fisher:

Fobs = 25.12314 (lấy từ bảng ANOVA đã lập ở câu 3)

Hoặc có thể tính toán thủ công bằng code R như sau:

|  |
| --- |
| ####Kiem dinh gia thuyet H0: Beta1 = Beta2 = 0  SSE <- anova(model3)[3,2]  SSR <- anova(model3)[1,2] + anova(model3)[2,2]  #p = 3 so luong he so hoi quy  #n = 12 tong so quan sat trong mau  F\_obs <- (SSR/2)/(SSE/9) #Gia tri thong ke  F\_obs |

Xác định miền bác bỏ với mức ý nghĩa 5%:

Vì Fobs = 25.12314 > F(0.05, 2, 9) =0.05158674 (Fobs lớn hơn rất nhiều giá trị tới hạn F)

Ta bác bỏ H0 với mức ý nghĩa 5% hay khả năng giả thuyết được chấp nhận là nhỏ hơn 5%. Có ít nhất 1 biến trong X1, X2 giải thích cho Y, hay là mức độ bền dẻo của nhựa có chịu ảnh hưởng tuyến tính của ít nhất 1 trong các yếu tố độ dày vật liệu và mật độ vật liệu, mô hình hồi quy tuyến tính có ý nghĩa.

1. **Khoảng tin cậy với mức ý nghĩa 5% cho β1 trong trường hợp mô hình chỉ có biến độc lập là độ dày vật liệu:**

Sử dụng hàm trong R để tính độ tin cậy như sau:

|  |
| --- |
| **confint(model1, level = 0.95)[2,]** |

Kết quá:

|  |
| --- |
| **> confint(model1, level = 0.95)[2,]**  **2.5 % 97.5 %**  **3.187036 8.884790** |

Khoảng tin cậy của hệ số hồi quy β1 là từ 3.19 đến 8.89 với mức ý nghĩa 5%. Hay với 100 lần ước lượng β1 thì có 95 lần rơi vào khoảng 3.19 đến 8.89

1. **Sử dụng khoảng tin cậy để khẳng định mô hình có ý nghĩa giữa mức độ bền dẻo của nhựa và độ dày của vật liệu và mật độ của vật liệu:**

Khoảng tin cậy 95% cho β1 cho mô hình hồi quy đơn mô tả sự phụ thuộc tuyến tính giữa mức độ bền dẻo của nhựa độ dày vật liệu từ 3.19 đến 8.89, khoảng tin cậy nằm khá xa và lệch hoàn toàn về phía phải so với giá trị 0. Điều này giúp ta dễ dàng bác bỏ giả thuyết β1 = 0 với mức ý nghĩa 5% tức là mô hình hồi quy tuyến tính này là có ý nghĩa. Có thể sử dụng để giải thích sự biến thiên của mức độ bền dẻo của nhựa dựa trên độ dày vật liệu

Ta xem xét khoảng tin cậy 95% của mô hình 2 biến:

|  |
| --- |
| **confint(model3, level = 0.95)** |

Kết quá:

|  |
| --- |
| **> confint(model3, level = 0.95)**  **2.5 % 97.5 %**  **(Intercept) -55.994798 -4.167046**  **X1 2.611442 7.197983**  **X2 2.881287 19.262828** |

Khoảng tin cậy 95% của hệ số hồi quy β1 là (2.61; 7.20) hoàn toàn nằm bên phải giá trị 0. X1 – độ dày vật liệu có mối quan hệ tuyến tính đồng biến với Y – mức độ bền dẻo của nhựa, độ dày tăng độ bền dẻo tăng và ngược lại.

Khoảng tin cậy 95% của hệ số hồi quy β2 là (2.88; 19.26) cũng là hoàn toàn nằm bên trái giá trị 0 nhưng rộng hơn khoảng 3 lần của β1 . Mật độ vật liệu có mối quan hệ tuyến tính đồng biến với mức độ bền dẻo của nhựa nhưng khả năng gây ra sai số ước lượng cao hơn so với độ dày vật liệu