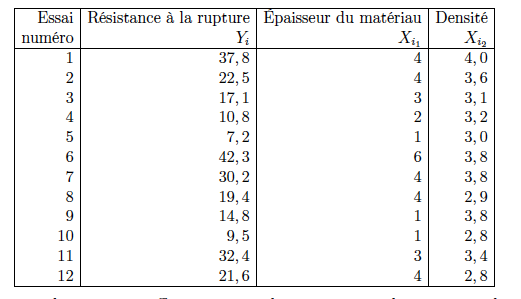
**Bài 2**

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, 6.3, 3.1, 3.2, 3.0, 3.8, 3.8, 2.9, 3.8, 2.8, 3., 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 2 là: **Y = 3.523 + 6.036\*X1**

*Nhận xét*: p-value của β1 trong 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 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  -13.181 -8.973 -2.131 8.428 19.331  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 10.678 12.987 0.822 0.430  X2 3.235 3.546 0.912 0.383  Residual standard error: 11.45 on 10 degrees of freedom  Multiple R-squared: 0.0768, Adjusted R-squared: -0.01552  F-statistic: 0.8319 on 1 and 10 DF, p-value: 0.3832 |

Như vậy, đường thẳng hồi quy mô hình 2 là: **Y = ~~10.678 + 3.235\*X~~~~2~~**

*Nhận xét*: p-value của cả hai hệ số hồi quy β0 và β1 trong mô hình này đều rất lớn là 0.430 và 0.383 - ở mức 43%. Mô hình này không có ý nghĩa thống kê (chính xác thì mô hình này chỉ có ở ý nghĩa ở mức 43% trở lên). Không có mối liên hệ tuyến tính giữa mức độ bền dẻo của nhựa và mật độ vật liệu.

* 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  -8.092 -5.003 -1.159 3.246 10.881  Coefficients:  Estimate Std. Error t value Pr(>|t|)  (Intercept) 2.8991 8.1368 0.356 0.72983  X1 5.9945 1.4188 4.225 0.00222 \*\*  X2 0.2121 2.2793 0.093 0.92789  ---  Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1  Residual standard error: 6.989 on 9 degrees of freedom  Multiple R-squared: 0.6906, Adjusted R-squared: 0.6218  F-statistic: 10.04 on 2 and 9 DF, p-value: 0.0051 |

Như vậy, mặt phẳng hồi quy mô hình 3 là: **Y = 2.8991 + 5.9945\*X1 + 0.2121\*X2**

*Nhận xét*: p-value khi kiểm định F cho sự tồn tại của mô hình rất nhỏ, nhở hơn 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.

Tuy nhiên, kiểm định T cho mỗi hệ số hồi quy, ta thấy chỉ có β1 là dễ dàng bác bỏ giả thuyết khác 0 – tồn tại mối liên hệ tuyến tính, còn đối với β0 và β2 với mức ý nghĩa tới 10% ta vẫn không có cở sở bác bỏ giả thuyết khác 0 của chúng, thậm chí lên đến mức 70% vẫn vậy. Ngoài ra, β1 cũng là hệ số lớn nhất, gấp 28 lần β2. Điều này cho thấy độ dày vật liệu là có ảnh hưởng lớn nhất đến mức độ bền dẻo của nhựa và nó cũng rất có ý nghĩa thống kê. Khi vật liệu khá mỏng việc đánh giá độ bền dẻo của nhựa là rất khó khăn.

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)$r.squared  summary(model3)$r.squared |

Kết quả:

|  |
| --- |
| > summary(model1)$r.squared  [1] 0.6902638  > summary(model1)$r.squared  [1] 0.6905616 |

Cả hai mô hình biến thiên của các biến độc lập giải thích 69% biến thiên của biến phụ thuộc. Ta thấy mô hình số 3 sử dụng thêm biến về mật độ vật liệu nhưng hệ số xác định tăng không đáng kể.

Tuy nhiên 2 mô hình khác nhau số biến độc, hệ số xác định hiệu chỉnh nên được sử dụng để so sánh:

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

Kết quả:

|  |
| --- |
| > summary(model1)$adj.r.squared  [1] 0.6592901  > summary(model1)$adj.r.squared  [1] 0.6217975 |

Mô hình 1 tuy chỉ có một biến nhưng dường như giải thích độ biến thiên của mức độ bền dẻo hơn mô hình 3.

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 20.0763 0.001531 \*\*  X2 1 0.42 0.42 0.0087 0.927887  Residuals 9 439.61 48.85  ---  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 | 981.02 | 490.51 | 10.04 |
| **E** | 9 | 439.61 | 48.85 |  |
| **T** | 11 | 1420.63 |  |  |

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 = 10.04 (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 = 10.04 > F(1-0.05, 2, 9) =4.26

Ta bác bỏ H0 với mức ý nghĩa 5% hay khả năng giả thuyết được chấp nhận là nhỏ hơn 5% khả năng giả thuyết bị bác bỏ có thể chấp nhận ở độ tin cậy 95%.

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 về 1 phía so vớigiá 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.

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) -15.507664 21.305793**  **X1 2.784947 9.203995**  **X2 -4.944025 5.368297** |

Chỉ duy nhất khoảng tin cậy 95% của hệ số hồi quy β1 (2.78;9.20)là hoàn toàn nằm bên trái giá trị 0. X1 – độ dày vật liệu có mối quan hệ tuyến tính đồng biến với Y – mức độ bền dẻo của nhựa. Khoảng tin cậy 95% của hệ số hồi quy β2 (-4.94;5.37) chứa giá trị 0. Ta không thể bác bỏ giả thuyết β2 = 0 với mức ý nghĩa 5%. Thậm chí giá trị 0 còn nằm khá sát gần trung điểm của khoảng tin cậy. Hay mật độ vật liệu X2 không có ý nghĩa thống kê trong mô hình tuyến tính với biến phụ thuộc mức độ bền dẻo của nhựa.