

**ĐẠI HỌC QUỐC GIA TP HỒ CHÍ MINH**

**ĐẠI HỌC BÁCH KHOA**



## **BÀI TẬP LỚN XÁC SUẤT THỐNG KÊ**

**LỚP L10 --- NHÓM 02 --- HK 231**

---

**XÁC ĐỊNH MỨC ĐỘ ẢNH HƯỞNG CỦA CÁC THÔNG SỐ ĐIỀU CHỈNH TRONG  
MÁY IN 3D ĐẾN CHẤT LƯỢNG IN, ĐỘ CHÍNH XÁC VÀ ĐỘ BỀN. TRƯỜNG  
HỢP CÓ CHÍN THÔNG SỐ CÀI ĐẶT VÀ BA THÔNG SỐ ĐẦU RA ĐƯỢC ĐO.**

---

**Giảng viên hướng dẫn: Phan Thị Hường**

<b>STT</b>	<b>Họ Và Tên</b>	<b>MSSV</b>	<b>Khoa</b>
1	Trần Bá Gia Khiêm	2013487	BDCN
2	Nguyễn Văn Kiên	2113822	BDCN
3	Nguyễn Tường Minh Nhật	2212401	Cơ Khí
4	Lê Nhựt Quy	2114580	BDCN

Thành phố Hồ Chí Minh, Ngày 26 Tháng 11 năm 2023

## KẾT QUẢ LÀM VIỆC NHÓM

STT	Họ Và Tên	MSSV	Nhiệm Vụ	Đóng góp	Điểm Chia
1	Trần Bá Gia Khiêm	2013487	Phần 3 + 6	25%	-0.5
2	Nguyễn Văn Kiên	2113822	Phần 2 + 6 + word	25%	
3	Nguyễn Tường Minh Nhật	2212401	Phần 4 + 6	25%	-0.5
4	Lê Nhựt Quy	2114580	Phần 1 + 3 + 4 + 5	25%	+1.0

## Đánh Giá của Giảng Viên

STT	Họ Và Tên	Điểm	Nhận xét
1	Trần Bá Gia Khiêm		
2	Nguyễn Văn Kiên		
3	Nguyễn Tường Minh Nhật		
4	Lê Nhựt Quy		

*Họ Và Tên Nhóm Trưởng: Lê Nhựt Quy, Email: [quy.le3797@hcmut.edu.vn](mailto:quy.le3797@hcmut.edu.vn)*

GIẢNG VIÊN  
(Ký và ghi rõ họ, tên)

TS. Phan Thị Hường

NHÓM TRƯỞNG  
(Ký và ghi rõ họ, tên)

Lê Nhựt Quy

## MỤC LỤC

PHẦN 1. TỔNG QUAN VỀ ĐỀ TÀI .....	2
PHẦN 2. KIẾN THỨC NỀN.....	3
2.1. Máy in 3D.....	3
2.2. Hồi quy tuyến tính bội .....	3
2.2.1. Mô hình hồi quy tuyến tính .....	3
2.2.2. Phương trình hồi quy tổng thể .....	4
2.2.3. Phương trình hồi quy mẫu.....	4
2.2.4. Sử dụng phương pháp bình phương cực tiểu để xác định hệ số hồi quy.....	5
2.2.5. Kiểm định mức độ phù hợp của mô hình.....	5
a). Hệ số xác định bội.....	5
b). Hệ số xác định điều chỉnh .....	6
c). Đánh giá ý nghĩa toàn diện của mô hình .....	7
d). Ước lượng khoảng tin cậy cho hệ số hồi quy .....	8
2.5.6 Kiểm định hệ số hồi quy.....	9
PHẦN 3. TIỀN XỬ LÝ SỐ LIỆU.....	11
3.1 Đọc dữ liệu .....	11
3.2 Khai báo hai biến (infill_pattern) và (material) là dạng phân loại.....	11
3.3 Xử lý dữ liệu khuyết.....	12
PHẦN 4. THỐNG KÊ MÔ TẢ .....	13
4.1. Tính các giá trị thống kê mô tả: trung bình (mean), trung vị (median), độ lệch chuẩn (sd), giá trị nhỏ nhất (min), giá trị lớn nhất (max), Q1, Q3.....	13
4.2. Vẽ biểu đồ histogram với các biến liên tục và phụ thuộc đầu ra. ....	15
4.3. Vẽ đồ thị boxplot của các biến đầu ra theo từng biến đầu vào .....	16
4.4. Biểu đồ phân tán của các biến độc lập và phụ thuộc .....	20
4.5 Ma trận tương quan .....	24
PHẦN 5: THỐNG KÊ SUY DIỄN .....	24
5.1 Dựng mô hình hồi quy bội cho độ nhám bề mặt (Roughness) .....	24
5.2 Kiểm định các giả định của mô hình hồi quy tuyến tính bội.....	28
KẾT LUẬN.....	32
PHẦN 6: THẢO LUẬN VÀ MỞ RỘNG .....	33
PHẦN 7: NGUỒN CODE VÀ DỮ LIỆU.....	33
PHẦN 8: TÀI LIỆU THAM KHẢO .....	33

## PHẦN 1. TỔNG QUAN VỀ ĐỀ TÀI

Đề tài nghiên cứu về mức độ ảnh hưởng độ ảnh hưởng của các thông số điều chỉnh trong máy in 3d đến chất lượng in, độ chính xác và độ bền. Trong đó có chín thông số cài đặt và ba thông số đầu ra được đo.

**Bảng Các Thông Số Của Máy In 3D**

<b>Thông Số Đầu Vào</b>	<b>Tên Gọi</b>	<b>Đơn Vị</b>	<b>Kiểu Biến</b>	<b>Loại Biến</b>
Layer Height	Chiều cao lớp	(mm)	Liên tục	Độc lập
Wall Thickness	Độ dày của tường	(mm)	Liên tục	Độc lập
Infill Density	Mật độ lấp đầy	(%)	Liên tục	Độc lập
Infill Pattern	Mẫu điền	Không	Phân loại	Độc lập
Nozzle Temperature	Nhiệt độ vòi phun	(C°)	Liên tục	Độc lập
Bed Temperature	Nhiệt độ giường	(C°)	Liên tục	Độc lập
Print Speed	Tốc độ in	(mm/s)	Liên tục	Độc lập
Material	Vật liệu	Không	Phân loại	Độc lập
Fan Speed	Tốc độ quạt	(%)	Liên tục	Độc lập
<b>Thông Số Đầu Ra</b>	<b>Tên Gọi</b>	<b>Đơn Vị</b>	<b>Kiểu Biến</b>	<b>Loại Biến</b>
Roughness	Độ nhám bề mặt	( $\mu$ m)	Liên tục	Phụ Thuộc
Tension Strength	Độ bền kéo của vật liệu	(MPa)	Liên tục	Phụ Thuộc
Elongation	Độ giãn dài	(%)	Liên tục	Phụ Thuộc

Để cài đặt được các thông số đầu vào chúng ta dựa trên các cài đặt và dây tóc của máy in 3-D Ultimaker S5. Các thử nghiệm vật liệu và độ bền được thực hiện trên máy thử Sincotec GMBH có khả năng kéo 20 kN.

Các thông số đầu vào là độc lập với nhau vì nó là cài đặt in của máy in nên khi thay đổi bất kỳ thông số nào trong số đó thì các thông số khác sẽ không thay đổi. Các thông số đầu ra như độ nhám, độ bền kéo, độ giãn dài sẽ phụ thuộc vào chính thông số đầu vào, thay đổi tùy theo chín thông số cài đặt ở trên.

## **PHẦN 2. KIẾN THỨC NỀN**

### **2.1. Máy in 3D**

Có kiến thức cơ bản về máy in 3D, các khái niệm về các thông số đầu vào và đầu ra của máy in.

### **2.2 Hồi quy tuyến tính bội**

#### **2.2.1. Mô hình hồi quy tuyến tính**

Trong nghiên cứu, chúng ta thường phải kiểm định các giả thuyết về mối quan hệ giữa hai hay nhiều biến, trong đó có một biến phụ thuộc và một hay nhiều biến độc lập. Nếu chỉ có một biến độc lập, mô hình được gọi là mô hình hồi quy đơn biến SLR (Simple Linear Regression). Trường hợp có từ hai biến độc lập trở lên, mô hình được gọi là hồi quy bội MLR (Multiple Linear Regression).

Mô hình hồi quy tuyến tính bội có dạng tổng quát như sau:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon$$

**Trong đó:**

- Y: biến phụ thuộc, là biến chịu tác động của biến khác.
- X, X<sub>1</sub>, X<sub>2</sub>, X<sub>n</sub>: biến độc lập, là biến tác động lên biến khác.
- $\beta_0$ : hằng số hồi quy, hay còn được gọi là hệ số chặn. Đây là chỉ số nói lên giá trị của Y sẽ là bao nhiêu nếu tất cả X cùng bằng 0. Nói cách khác, chỉ số này cho

chúng ta biết giá trị của  $Y$  là bao nhiêu nếu không có các  $X$ . Khi biểu diễn trên đồ thị Oxy,  $\beta_0$  là điểm trên trục Oy mà đường hồi quy cắt qua.

- $\beta_1, \beta_2, \beta_n$ : hệ số hồi quy, hay còn được gọi là hệ số góc. Chỉ số này cho chúng ta biết về mức thay đổi của  $Y$  gây ra bởi  $X$  tương ứng. Nói cách khác, chỉ số này nói lên có bao nhiêu đơn vị  $Y$  sẽ thay đổi nếu  $X$  tăng hoặc giảm một đơn vị.
- $e$ : sai số. Chỉ số này càng lớn càng khiến cho khả năng dự đoán của hồi quy trở nên kém chính xác hơn hoặc sai lệch nhiều hơn so với thực tế. Sai số trong hồi quy tổng thể hay phần dư trong hồi quy mẫu đại diện cho hai giá trị, một là các biến độc lập ngoài mô hình, hai là các sai số ngẫu nhiên.

### 2.2.2. Phương trình hồi quy tổng thể

Với  $Y$  là biến phụ thuộc vào  $x_1, x_2, \dots, x_k$  là các biến độc lập,  $Y$  là ngẫu nhiên có một phân phối xác suất nào đó. Tồn tại  $E(Y|x_1, x_2, \dots, x_k) = f_Y(x_1, x_2, \dots, x_k)$  là hàm hồi quy tổng thể của biến phụ thuộc  $Y$  theo biến giải thích  $x_1, x_2, \dots, x_k$ .

Nghĩa là hàm hồi quy của  $Y$  theo  $x_1, x_2, \dots, x_k$  chính là kỳ vọng có điều kiện của  $Y$  đối với  $x_j$ , có dạng:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \epsilon_i$$

**Trong đó:**

- $\beta_0$ : hệ số tung độ gốc (hệ số chặn)
- $\beta_1$ : hệ số góc của biến  $Y$  theo biến  $x_1$  giữ các biến  $x_2, \dots, x_k$  không đổi
- $\beta_2$ : hệ số góc của biến  $Y$  theo biến  $x_2$  giữ các biến  $x_1, x_3, \dots, x_k$  không đổi
- ...
- $\beta_k$ : hệ số góc của biến  $Y$  theo biến  $x_k$  giữ các biến  $x_1, x_2, \dots, x_{k-1}$  không đổi
- $\epsilon_i$ : thành phần ngẫu nhiên, có phân phối chuẩn với  $E(\epsilon_i) = 0, Var(\epsilon_i) = \sigma^2$

### 2.2.3. Phương trình hồi quy mẫu

Do không thể biết được tổng thể, nên chúng ta không biết được giá trị trung bình tổng thể của biến phụ thuộc là đúng ở mức độ nào. Vì thế chúng ta phải dựa vào dữ liệu mẫu để ước lượng.

Ta có hàm hồi quy mẫu tổng quát được viết dưới dạng như sau:

$$\hat{Y}_i = \hat{B}_0 + \hat{B}_1 x_{i1} + \hat{B}_2 x_{i2} + \dots + \hat{B}_k x_{ik}$$

**Trong đó:**

- $\hat{Y}_i$ : là giá trị ước lượng của biến phụ thuộc  $Y_i$  căn cứ vào mô hình hồi quy
- $\hat{B}_i$ : là giá trị ước lượng cho hệ số hồi quy  $\beta_i$  căn cứ vào dữ liệu mẫu, ( $i = 0, 1, \dots, k$ ).

#### 2.2.4. Sử dụng phương pháp bình phương cực tiểu để xác định hệ số hồi quy

Phương trình hồi quy tuyến tính có dạng:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \epsilon_i$$

Để xác định được các hệ số hồi quy  $\beta_j$  ( $j = 0, 1, 2, \dots, k$ ) ta sử dụng hàm phương pháp bình phương cực tiểu, và hàm bình phương cực tiểu được xác định:

$$L = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij})^2$$

Tìm giá trị nhỏ nhất của hàm L ta sẽ tìm được các hệ số hồi quy mẫu cho phương trình hồi quy.

Trong khi xây dựng mô hình hồi quy tuyến tính bội, ta cần kiểm tra các giả thuyết như: hàm hồi quy là hàm tuyến tính theo các tham số, sai số ngẫu nhiên độc lập với nhau tuân theo phân phối chuẩn với kỳ vọng bằng 0 và phương sai là  $\sigma^2$ .

#### 2.2.5. Kiểm định mức độ phù hợp của mô hình

##### a). Hệ số xác định bội

Để xác định được phần biến thiên trong biến phụ thuộc được giải thích bởi mối liên hệ giữa biến phụ thuộc và tất cả các biến độc lập trong mô hình, người ta đi xác định hệ số xác định bội  $R^2$  ( $0 \leq R^2 \leq 1$ ). Hệ số xác định bội sẽ giải thích trong 100% sự biến động của

Y so với trung bình của nó thì có bao nhiêu % là do biến các biến  $x_j$ , gây ra, công thức để tính hệ số xác định bội là:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

Trong đó:

- SST (Sum of Squares Total): là tổng bình phương tất cả các sai lệch giữa các giá trị của biến phụ thuộc và giá trị trung bình

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

- SSR (Sum of Squares in Regression): là tổng bình phương của tất cả các sai lệch giữa các giá trị của biến phụ thuộc Y nhận được từ hàm hồi quy mẫu và giá trị trung bình của chúng.

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

- SSE (Sum of Squares for Error): dùng để đo sự chênh lệch giữa từng giá trị quan sát với giá trị dự đoán. SSE được xem như sai số do những yếu tố ngoài X hoặc do lấy mẫu ngẫu nhiên.

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

- $R^2$  cao nghĩa là sự biến động của biến phụ thuộc Y so với trung bình của nó do biến X gây ra càng cao
- Nếu  $R^2 = 1$  nghĩa là đường hồi quy giải thích cho 100% sự thay đổi của Y
- Nếu  $R^2 = 0$  nghĩa là mô hình không đưa ra thông tin nào về sự thay đổi của biến phụ thuộc Y.

#### ***b). Hệ số xác định điều chỉnh***

Do việc đưa thêm biến độc lập vào mô hình sẽ luôn làm gia tăng hệ số xác định  $R^2$ , thậm chí ngay cả khi biến độc lập được đưa vào không có mối liên hệ hoặc có mối liên hệ không đáng kể với biến phụ thuộc. Vì thế khi số biến độc lập tăng lên chắc chắn  $R^2$  sẽ



luôn tăng lên, tuy nhiên mỗi biến độc lập được thêm vào sẽ làm mất đi một bậc tự do. Sự gia tăng trong  $R^2$  có thể không bù đắp được thiệt hại do mất thêm bậc tự do khi thêm biến, thế nhưng  $R^2_{adj}$  lại có xét đến và điều chỉnh giá trị của  $R^2_{adj}$  theo nó một cách phù hợp.  $R^2_{adj}$  sẽ luôn bé hơn  $R^2$ . Khi một biến độc lập được thêm vào không có đóng góp xứng đáng vào khả năng giải thích cho biến phụ thuộc thì  $R^2_{adj}$  sẽ luôn giảm mặc dù  $R^2$  thì tăng.

Hệ số xác định điều chỉnh là một đại lượng đo lường quan trọng khi số biến độc lập lớn một cách tương đối so với cỡ mẫu, nó tính đến mối liên hệ giữa cỡ mẫu và số biến, nếu số biến độc lập là khá lớn so với cỡ mẫu thì  $R^2$  sẽ thổi phồng khả năng giải thích cho biến phụ thuộc của mô hình một cách giả tạo, thế nên người ta cần xài đến hệ số xác định điều chỉnh để đánh giá được chính xác hơn sự thay đổi của biến phụ thuộc Y vào biến độc lập X. Hệ số xác định điều chỉnh được xác định bằng công thức:

$$R^2_{adj} = 1 - \frac{\frac{SSE}{n-k-1}}{\frac{SST}{n-1}} = 1 - (1 - R^2) \frac{n-1}{n-k-1}$$

### c). *Đánh giá ý nghĩa toàn diện của mô hình*

Kiểm định mức ý nghĩa của mô hình hồi quy là kiểm định để xác định xem có tồn tại mối quan hệ tuyến tính hay không giữa biến phụ thuộc  $Y_i$  và tập con của các biến độc lập  $x_1, x_2, \dots, x_k$ . Vì mô hình hồi quy mà chúng ta xây dựng là dựa trên dữ liệu của một mẫu lấy từ tổng thể vì vậy nó có thể bị ảnh hưởng của sai số lấy mẫu, chính vì thế mà chúng ta cần phải kiểm định ý nghĩa thống kê của mô hình. Như đã biết, hệ số xác định bội  $R^2$  là đại lượng cho biết có bao nhiêu phần trăm biến thiên trong biến phụ thuộc có thể được giải thích bởi mô hình hồi quy, là một số thống kê trên mẫu có thể sử dụng để suy diễn về việc mô hình toàn diện có ý nghĩa về mặt thống kê hay không trong việc giải thích cho biến thiên của biến phụ thuộc. Với ý tưởng này, chúng ta đặt ra các giả thuyết kiểm định như sau:

$$H_0: R^2 = 0$$

$$H_1: R^2 \neq 0$$

Bản chất của giả thuyết  $H_0$  này là mô hình hồi quy tuyến tính bội tổng thể mà chúng ta xây dựng thì các biến độc lập không giải thích được chút nào cho những biến thiên trong biến phụ thuộc. Tương tự, ta có thể đặt lại giả thuyết:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_k = 0$$

$$H_1: \text{Có ít nhất một hệ số } \beta_j \neq 0$$

Nếu ta bác bỏ giả thuyết  $H_0$  đồng nghĩa với việc có ít nhất một trong những biến hồi quy  $x_1, x_2, \dots, x_k$  có đóng góp đáng kể cho mô hình

Kiểm định thống kê cho giả thuyết  $H_0$  được xác định bằng công thức:

$$F_0 = \frac{\frac{SSR}{k}}{\frac{SSE}{n-k-1}} = \frac{MSR}{MSE}$$

Chúng ta sẽ bác bỏ giả thuyết  $H_0$  nếu như giá trị của kiểm định thống kê  $f_0$  là lớn hơn  $f_{\alpha, k, n-k-1}$ .

#### **d). Ước lượng khoảng tin cậy cho hệ số hồi quy**

Trong mô hình hồi quy tuyến tính bội, người ta thường xác định khoảng tin cậy của các hệ số hồi quy để xác định được phân bố xác suất của các hệ số hồi quy  $\beta_j$  ( $j=0,1,\dots,k$ ). Việc xây dựng khoảng tin cậy cho các hệ số hồi quy này đòi hỏi sai số  $\epsilon_i$  phải có phân phối chuẩn với kỳ vọng bằng 0 và phương sai là  $\sigma^2$

$$\epsilon_i \sim N(0, \sigma^2)$$

Do đó, từ giả định của sai số ngẫu nhiên, suy ra biến phụ thuộc  $Y_i$  sẽ có phân phối chuẩn với kỳ vọng là  $\beta_0 + \sum_{j=1}^k \beta_j x_{ij}$  và phương sai là  $\sigma^2$ . Vì ước lượng điểm cho các hệ số hồi quy  $\beta_j$  là một tổ hợp tuyến tính nên  $\beta_j$  sẽ có phân phối chuẩn với kỳ vọng là  $\beta_j$  phương sai là ma trận hiệp phương sai. Sau đó, mỗi thống kê:

$$T = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 c_{ij}}} \quad j = 0, 1, \dots, k$$

sẽ có phân phối T với bậc tự do là  $n - k - 1$

Trong đó:

- $\sigma^2$ : là ước lượng điểm cho phương sai lỗi, và được xác định bằng

- $\hat{\sigma}^2 = \frac{SSE}{n-k-1}$
- $\sqrt{\hat{\sigma}^2 C_{ij}}$ : là sai số chuẩn của hệ số hồi quy  $\hat{\beta}_j$ , kí hiệu  $se(\hat{\beta}_j)$

Điều này dẫn đến với độ tin cậy  $1 - \alpha$  thì khoảng ước lượng cho hệ số hồi quy  $\beta_j$ ,  $j=0,1,\dots,k$  cho mô hình hồi quy tuyến tính bội được đưa ra bởi:

$$\hat{\beta}_j - t_{\frac{\alpha}{2}, n-k-1} se(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + t_{\frac{\alpha}{2}, n-k-1} se(\hat{\beta}_j)$$

### 2.5.6 Kiểm định hệ số hồi quy

Thông thường, khi xây dựng mô hình hồi quy tuyến tính bội, người ta thường quan tâm đến việc kiểm tra các giả thuyết về các hệ số hồi quy riêng lẻ. Các thử nghiệm như vậy sẽ hữu ích trong việc xác định giá trị tiềm năng của từng biến hồi quy. Ví dụ, mô hình có thể có hiệu quả hơn nếu bao gồm các biến bổ sung hoặc có thể là việc xóa bỏ một biến hồi quy trong mô hình. Cũng giống như mô hình hồi quy tuyến tính đơn, kiểm định thống kê cũng yêu cầu sai số ngẫu nhiên  $\epsilon_i$  có phân phối chuẩn với kỳ vọng bằng 0 và phương sai là  $\sigma^2$ . Giả thuyết để kiểm tra hệ số hồi quy riêng lẻ, giả định rằng:

$$H_0: \beta_j = \beta_{j0}$$

$$H_1: \beta_j \neq \beta_{j0}$$

Và thống kê kiểm định cho giả thuyết này là:

$$T = \frac{\hat{\beta}_j - \beta_{j0}}{\sqrt{\hat{\sigma}^2 C_{ij}}} = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)}$$

Giả thuyết  $H_0$  sẽ bị bác bỏ nếu như  $|t_0| > t_{\frac{\alpha}{2}, n-k-1}$ .

Một trường hợp đặc biệt quan trọng của giả thuyết  $H_0$ , là thông thường ta sẽ kiểm định giả thuyết  $H_0, H_1$  với  $\beta_{j0} = 0$

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

Nếu như giả thuyết  $H_0$  không bị bác bỏ, điều này đồng nghĩa biến hồi quy  $x_j$  có thể được xóa bỏ khỏi mô hình. Việc thêm biến vào mô hình hồi quy luôn làm cho tổng bình phương của tất cả các sai lệch giữa các giá trị của biến phụ thuộc  $Y$  nhận được từ hàm hồi quy mẫu và giá trị trung bình của chúng (SSR) tăng lên và tổng bình phương của lỗi (SSE) giảm xuống (điều này giải thích tại sao hệ số xác định bội  $R^2$  luôn tăng khi ta thêm biến hồi quy vào mô hình). Chúng ta phải xem xét rằng mức tăng trong SSR có đủ lớn khi ta thêm một biến mới vào hay không, hơn nữa việc thêm một biến mới không quan trọng vào mô hình có thể làm cho sai số ngẫu nhiên tăng lên. Điều này cho thấy rằng việc thêm một biến như vậy thực sự đã làm cho mô hình kém phù hợp hơn so với dữ liệu (Điều này lí giải tại sao hệ số xác định điều chỉnh  $R^2_{adj}$  là biện pháp tốt hơn trong việc kiểm tra độ phù hợp của mô hình). Ngoài ra, còn có các giả thuyết kiểm định về phía trái hay phía phải cho các hệ số hồi quy riêng lẻ.

Đối với phía phải ta có các giả thuyết

$$H_0: \beta_j \leq \beta_{j0}$$

$$H_1: \beta_j > \beta_{j0}$$

Thống kê kiểm định cũng giống như kiểm định về hai phía. Và giả thuyết  $H_0$  sẽ bị bác bỏ nếu như  $t_0 > t_{\frac{\alpha}{2}, n-k-1}$ . Việc bác bỏ giả thuyết  $H_0$  đồng nghĩa với việc ta chấp nhận giả thuyết  $H_1$ , điều đó có nghĩa là  $x_j$  có tác động thuận đối với mô hình.

Tương tự cho kiểm định về bên trái, ta cũng đặt các giả thuyết:

$$H_0: \beta_j \geq \beta_{j0}$$

$$H_1: \beta_j < \beta_{j0}$$

Giả thuyết  $H_0$  sẽ bị bác bỏ nếu như  $t_0 < t_{\frac{\alpha}{2}, n-k-1}$ . Điều này đồng nghĩa với  $x_j$  sẽ có tác động ngược đối với mô hình.

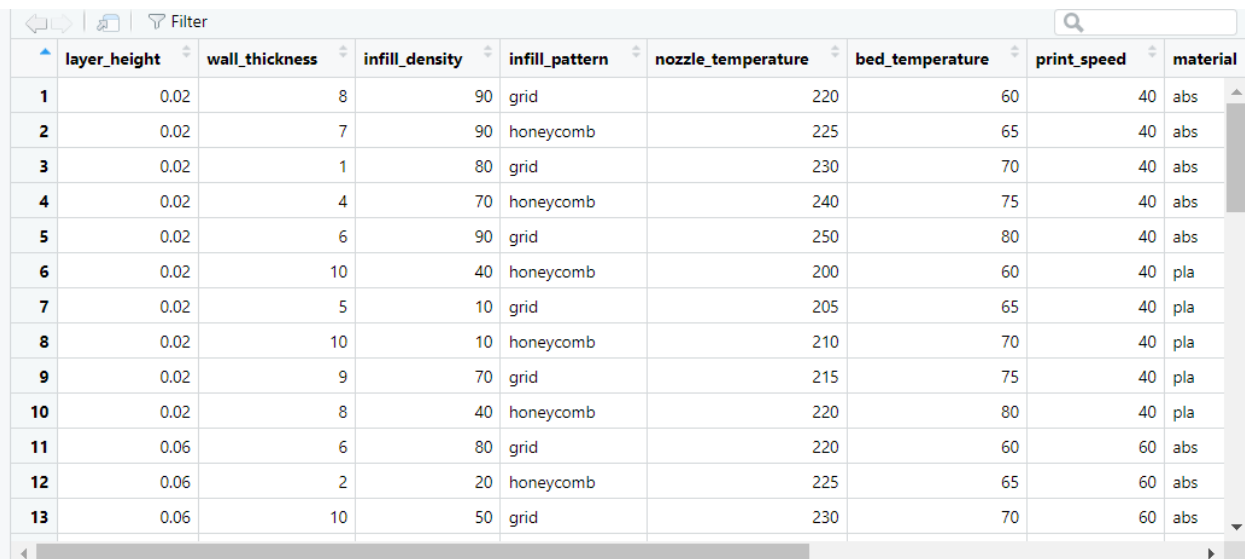
## PHẦN 3. TIỀN XỬ LÝ SỐ LIỆU

### 3.1 Đọc dữ liệu

Trong đề tài này chúng ta sẽ lấy số liệu nghiên cứu từ nghiên cứu của khoa Cơ khí Đại học TR/Selcuk cho hai loại vật liệu polymer ABS và PLA

Đọc dữ liệu từ file “ data.csv”

```
library(readr)
data <- read_csv("BTLxstk/data.csv")
View(data)
```



	layer_height	wall_thickness	infill_density	infill_pattern	nozzle_temperature	bed_temperature	print_speed	material
1	0.02	8	90	grid	220	60	40	abs
2	0.02	7	90	honeycomb	225	65	40	abs
3	0.02	1	80	grid	230	70	40	abs
4	0.02	4	70	honeycomb	240	75	40	abs
5	0.02	6	90	grid	250	80	40	abs
6	0.02	10	40	honeycomb	200	60	40	pla
7	0.02	5	10	grid	205	65	40	pla
8	0.02	10	10	honeycomb	210	70	40	pla
9	0.02	9	70	grid	215	75	40	pla
10	0.02	8	40	honeycomb	220	80	40	pla
11	0.06	6	80	grid	220	60	60	abs
12	0.06	2	20	honeycomb	225	65	60	abs
13	0.06	10	50	grid	230	70	60	abs

*Hình 1: Đọc dữ liệu từ file excel “ data.csv”*

### 3.2 Khai báo hai biến (infill\_pattern) và (material) là dạng phân loại.

Code:

```
#chuyển đổi 2 cột infill_pattern và cột material thành biến phân loại
data$infill_pattern <- factor(data$infill_pattern)
data$material <- factor(data$material)
```

❖ Kiểm tra lại kết quả đã đúng hay chưa bằng câu lệnh:

```
if (is.factor(data$infill_pattern)) {  
  print("Cột infill_pattern là biến phân loại.")  
} else {  
  print("Cột infill_pattern không phải là biến phân loại.")  
}  
if (is.factor(data$material)) {  
  print("Cột material là biến phân loại.")  
} else {  
  print("Cột material không phải là biến phân loại.")  
}
```

```
> if (is.factor(data$infill_pattern)) {  
+   print("Cột infill_pattern là biến phân loại.")  
+ } else {  
+   print("Cột infill_pattern không phải là biến phân loại.")  
+ }  
[1] "Cột infill_pattern là biến phân loại."  
> # Kiểm tra xem cột "material" có phải là biến phân loại hay không  
> if (is.factor(data$material)) {  
+   print("Cột material là biến phân loại.")  
+ } else {  
+   print("Cột material không phải là biến phân loại.")  
+ }  
[1] "Cột material là biến phân loại."  
> |
```

**Hình 2:** Kết quả kiểm tra biến phân loại

**Nhận xét:** Kết quả đã khai báo thành công 2 biến infill\_pattern và material là biến phân loại.

### 3.3 Xử lý dữ liệu khuyết.

Code:

```
missing_values <- sum(is.na(data))  
  
# In tổng số giá trị khuyết  
cat("Tổng số giá trị khuyết trong dữ liệu là:", missing_values, "\n")
```

```
> # In tổng số giá trị khuyết
> cat("Tổng số giá trị khuyết trong dữ liệu là:", missing_values, "\n")
Tổng số giá trị khuyết trong dữ liệu là: 0
```

**Hình 3:** Kết quả kiểm tra số dữ liệu khuyết

**Nhận xét:** Dựa vào kết quả kết luận dữ liệu này không có dữ liệu khuyết

## PHẦN 4. THỐNG KÊ MÔ TẢ

### 4.1. Tính các giá trị thống kê mô tả: trung bình (mean), trung vị (median), độ lệch chuẩn (sd), giá trị nhỏ nhất (min), giá trị lớn nhất (max), Q1, Q3

```
> summary(data)
 layer_height wall_thickness infill_density infill_pattern nozzle_temperature
Min. :0.020 Min. : 1.00 Min. :10.0 Length:50 Min. :200.0
1st Qu.:0.060 1st Qu.: 3.00 1st Qu.:40.0 Class :character 1st Qu.:210.0
Median :0.100 Median : 5.00 Median :50.0 Mode :character Median :220.0
Mean :0.106 Mean : 5.22 Mean :53.4 Mean :221.5
3rd Qu.:0.150 3rd Qu.: 7.00 3rd Qu.:80.0 3rd Qu.:230.0
Max. :0.200 Max. :10.00 Max. :90.0 Max. :250.0
bed_temperature print_speed material fan_speed roughness tension_strenght
Min. :60 Min. : 40 Length:50 Min. : 0 Min. : 21.0 Min. : 4.00
1st Qu.:65 1st Qu.: 40 Class :character 1st Qu.: 25 1st Qu.: 92.0 1st Qu.:12.00
Median :70 Median : 60 Mode :character Median : 50 Median :165.5 Median :19.00
Mean :70 Mean : 64 Mean : 50 Mean :170.6 Mean :20.08
3rd Qu.:75 3rd Qu.: 60 3rd Qu.: 75 3rd Qu.:239.2 3rd Qu.:27.00
Max. :80 Max. :120 Max. :100 Max. :368.0 Max. :37.00
elongation
Min. :0.400
1st Qu.:1.100
Median :1.550
Mean :1.672
3rd Qu.:2.175
Max. :3.300
```

**Hình 4:** Các giá trị thống kê mô tả của dữ liệu

❖ Hàm `summary()` không tính được độ lệch chuẩn do đó ta sử dụng đoạn code sau:

```
#tính độ lệch chuẩn
selected_columns <- c("layer_height", "wall_thickness", "infill_density", "nozzle_temperature",
  ["bed_temperature", "print_speed", "fan_speed", "roughness", "tension_strenght", "elongation"]
dt <- data[, selected_columns]
sd_values <- sapply(dt, sd)
result <- data.frame(Variable = names(sd_values), Standard_Deviation = sd_values)
print(result)
```

Variable	Standard_Deviation
layer_height	0.06439673
wall_thickness	2.92274682
infill_density	25.36348009
nozzle_temperature	14.82035280
bed_temperature	7.14285714
print_speed	29.69229956
fan_speed	35.71428571
roughness	99.03412935
tension_strength	8.92563380
elongation	0.78818831

**Hình 5:** Độ lệch chuẩn của các biến liên tục

❖ Đối với biến phân loại, cần thống kê các giá trị và tần số của từng giá trị.

```
#thống kê mô tả biến phân loại
categorical_variables <- c("material", "infill_pattern")
# Tạo và in bảng thống kê
for (variable in categorical_variables) {
  cat("Bảng thống kê cho biến", variable, ":\n")
  print(table(data[[variable]]))
  cat("\n")
}
```

Bảng thống kê cho biến material :

```
abs pla
25  25
```

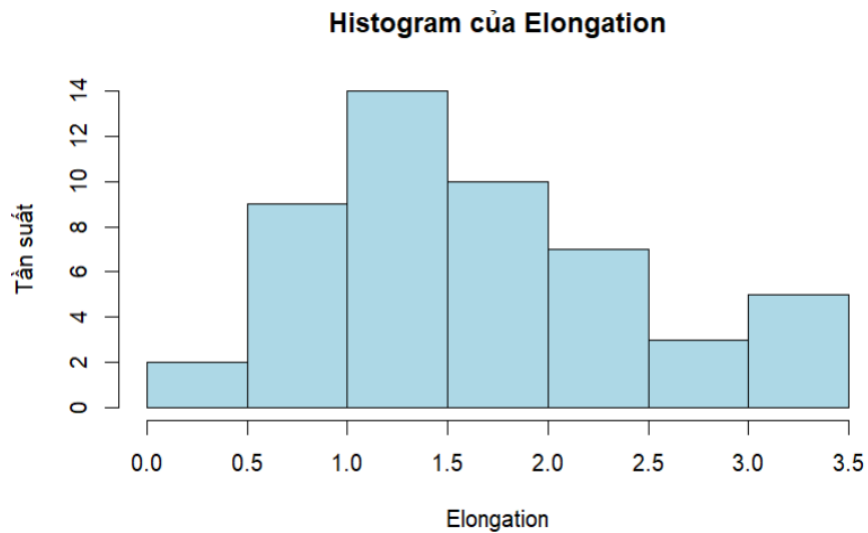
Bảng thống kê cho biến infill\_pattern :

```
grid honeycomb
25      25
```

**Hình 6:** Giá trị thống kê của biến phân loại

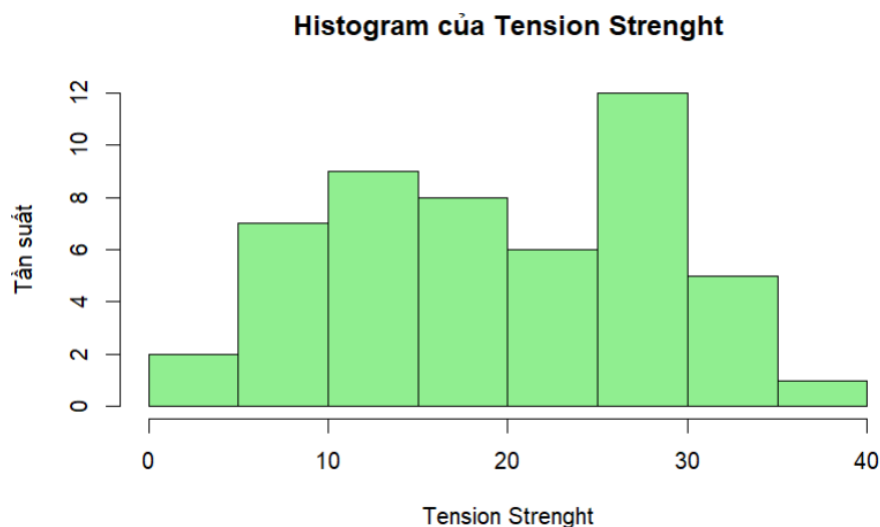


#### 4.2. Vẽ biểu đồ histogram với các biến liên tục và phụ thuộc đầu ra.



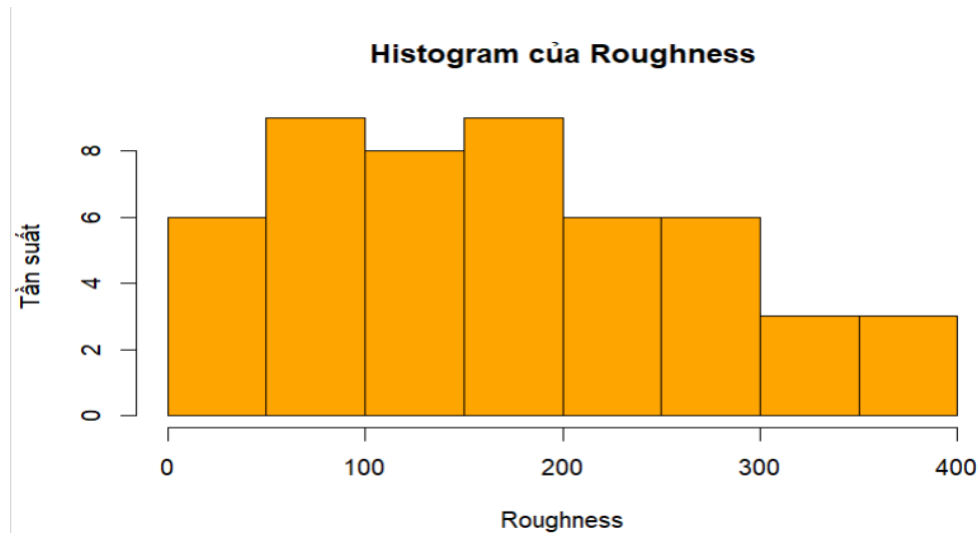
**Hình 7:** Biểu đồ histogram của biến elongation

**Nhận xét:** Biểu đồ histogram của biến Elongation chưa phải phân phối chuẩn (khá giống với hình cái chuông), và bị lệch về phía bên phải (  $\text{Mean} > \text{Median}$  ). Dữ liệu của thông số trải dài từ 0,0%- 3,5% . Dữ liệu từ 1,0% -1,5% chiếm tỉ lệ nhiều nhất với tần suất là 14, từ 0,0% - 0,5% chiếm tỉ lệ ít nhất với tần suất là 2.



**Hình 8:** Biểu đồ histogram của biến tension\_strength

**Nhận xét:** Biểu đồ histogram của biến tension\_strenght có dạng phân phối chuẩn (khá giống với hình cái chuông). Dữ liệu của thông số trải dài từ 0 đến 40MPa . Dữ liệu từ 25 - 30Mpa chiếm tỉ lệ nhiều nhất với tần suất là 12, từ 35 - 40MPa chiếm tỉ lệ ít nhất với tần suất là 1.

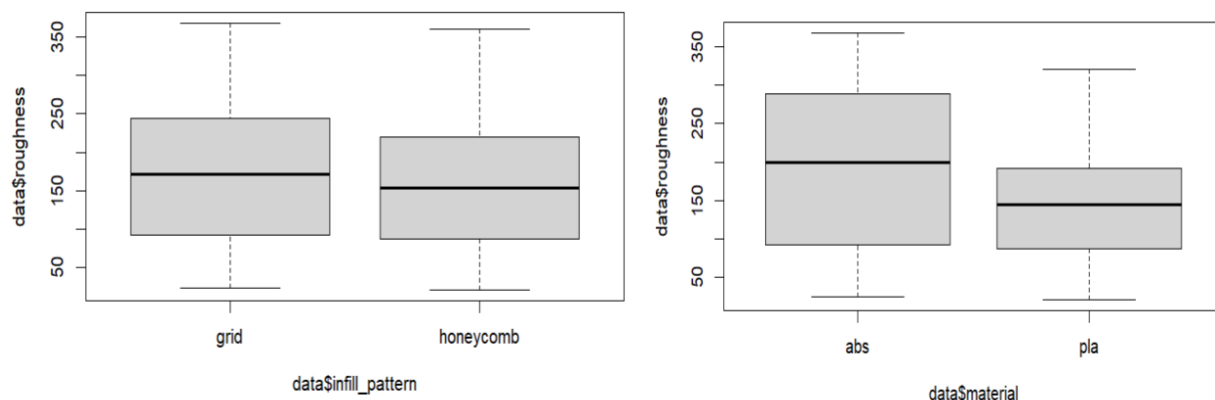


**Hình 9:** Biểu đồ histogram của biến roughness

**Nhận xét:** Biểu đồ histogram của biến roughness không có dạng phân phối chuẩn. Dữ liệu của thông số trải dài từ 0 đến 400 $\mu m$  . Dữ liệu từ 50 – 100 $\mu m$  và 150 - 200  $\mu m$  chiếm tỉ lệ nhiều nhất với tần suất là 9, từ 300 - 400  $\mu m$  chiếm tỉ lệ ít nhất với tần suất là

#### 4.3. Vẽ đồ thị boxplot của các biến đầu ra theo từng biến đầu vào

##### ❖ Biến Roughness với biến infill\_pattern và material



*Hình 10: Biểu đồ boxplot của biến Roughness với biến infill\_pattern và material*

**Nhận xét:**

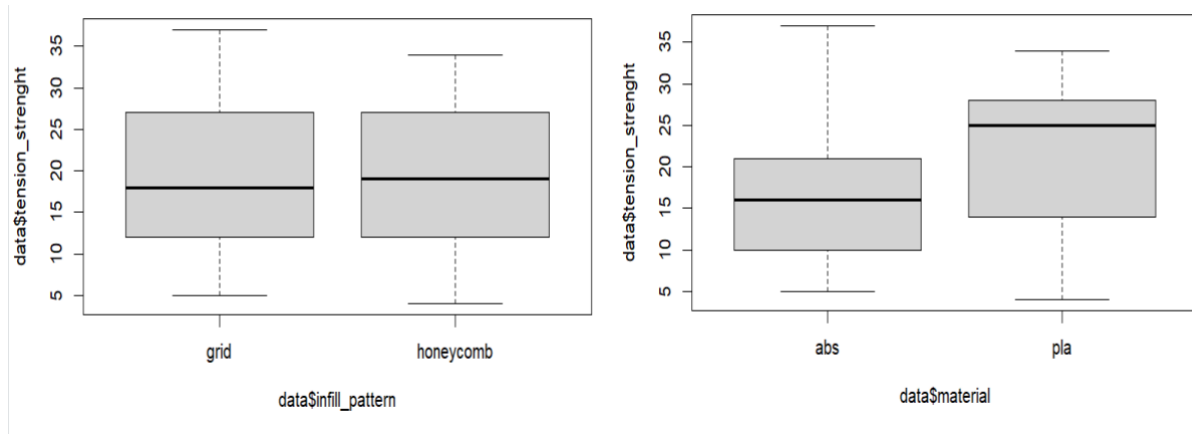
➤ **Đối với mẫu điền grid và honeycomb**

- Không có sự khác biệt giữa hai nhóm này.
- Giá trị lớn nhất của roughness: khoảng 370  $\mu\text{m}$  đối với loại grid và khoảng 360  $\mu\text{m}$  với loại honeycomb.
- Giá trị nhỏ nhất của roughness: khoảng 20  $\mu\text{m}$  đối với loại grid và khoảng 19  $\mu\text{m}$  với loại honeycomb.
- Mức phân vị Q1( có 25% giá trị đó trở xuống): khoảng 90  $\mu\text{m}$  đối với loại grid và khoảng 88  $\mu\text{m}$  với loại honeycomb.
- Mức phân vị Q2( có 50% giá trị đó trở xuống): Khoảng 170  $\mu\text{m}$  đối với loại grid và khoảng 160  $\mu\text{m}$  đối với loại honeycomb.
- Mức phân vị Q3( có 75% giá trị đó trở xuống): Khoảng đối 245  $\mu\text{m}$  với loại grid và khoảng 225  $\mu\text{m}$  honeycomb.

➤ **Đối với loại vật liệu là abs và pla.**

- Ta thấy phân phối của loại pla nhỏ hơn so với loại abs.
- Giá trị lớn nhất của roughness: khoảng 370  $\mu\text{m}$  đối với loại abs và khoảng 320  $\mu\text{m}$  với loại pla.
- Giá trị nhỏ nhất của roughness: khoảng 20  $\mu\text{m}$  đối với loại abs và khoảng 19  $\mu\text{m}$  với loại pla.
- Mức phân vị Q1( có 25% giá trị đó trở xuống): khoảng 90  $\mu\text{m}$  đối với loại abs và khoảng 88  $\mu\text{m}$  với loại pla.
- Mức phân vị Q2( có 50% giá trị đó trở xuống): Khoảng 200  $\mu\text{m}$  đối với loại abs và khoảng 145  $\mu\text{m}$  đối với loại pla
- Mức phân vị Q3( có 75% giá trị đó trở xuống): Khoảng 290  $\mu\text{m}$  đối với loại abs và khoảng 190  $\mu\text{m}$  với loại pla.

❖ **Biến `tension_strength` với biến `infill_pattern` và `material`**



*Hình 11: Biểu đồ boxplot của biến `tension_strength` với biến `infill_pattern` và `material`*

**Nhận xét:**

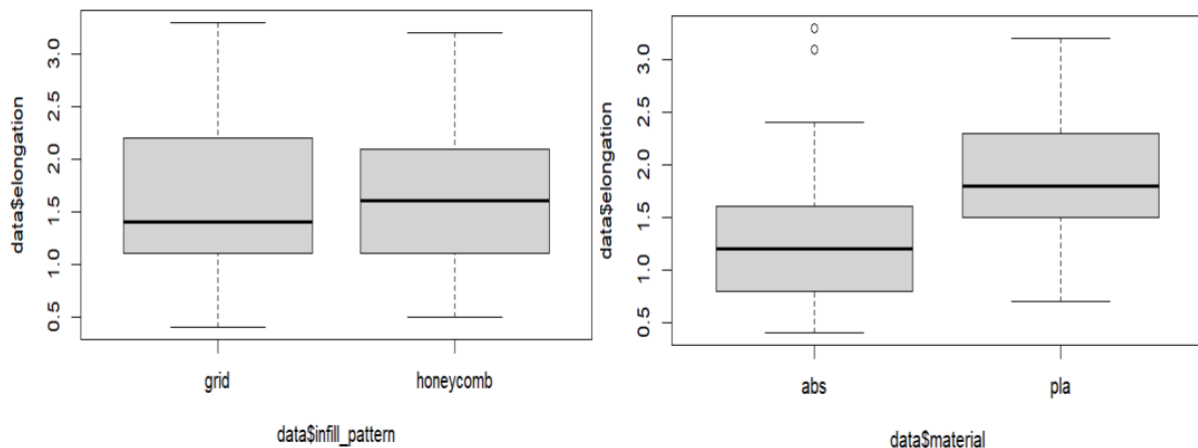
➤ **Đối với mẫu điền `grid` và `honeycomb`**

- Không có sự khác biệt giữa hai nhóm này.
- Giá trị lớn nhất của roughness: khoảng 36 MPa đối với loại `grid` và khoảng 34 MPa với loại `honeycomb`.
- Giá trị nhỏ nhất của roughness: khoảng 5 MPa đối với loại `grid` và khoảng 4 MPa với loại `honeycomb`.
- Mức phân vị Q1( có 25% giá trị đó trở xuống): khoảng 12 MPa đối với loại `grid` và loại `honeycomb`.
- Mức phân vị Q2( có 50% giá trị đó trở xuống): Khoảng 16 Mpa đối với loại `grid` và khoảng 17 MPa đối với loại `honeycomb`.
- Mức phân vị Q3( có 75% giá trị đó trở xuống): Khoảng 27 Mpa đối với loại `grid` và `honeycomb`.

➤ **Đối với loại vật liệu là abs và pla.**

- Phân phối của abs nhỏ hơn pla.
- Giá trị lớn nhất của roughness: khoảng 37 MPa đối với loại abs và khoảng 35 MPa với loại pla.
- Giá trị nhỏ nhất của roughness: khoảng 5 MPa đối với loại abs và khoảng 4 MPa với loại pla.
- Mức phân vị Q1( có 25% giá trị đó trở xuống): khoảng 10 MPa đối với loại abs và khoảng 13 với loại pla.
- Mức phân vị Q2( có 50% giá trị đó trở xuống): Khoảng 16 Mpa đối với loại abs và khoảng 25 MPa đối với loại pla.
- Mức phân vị Q3( có 75% giá trị đó trở xuống): Khoảng 21 Mpa đối với loại abs và khoảng 26 Mpa với loại pla.

❖ **Biến elongation với biến infill\_pattern và material**



**Hình 12:** Biểu đồ boxplot của biến elongation với biến infill\_pattern và material

**Nhận xét:**

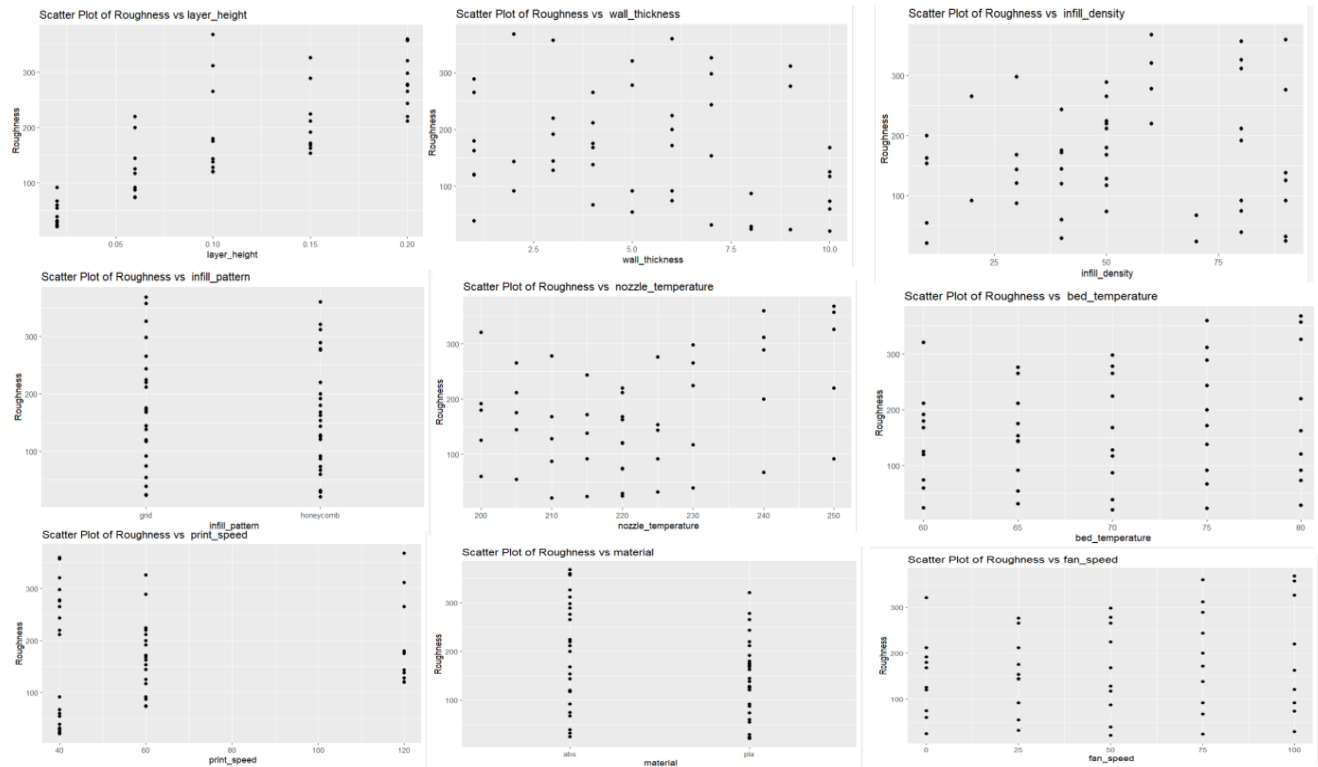
➤ **Đối với mẫu điền grid và honeycomb**

- Không có sự khác biệt giữa hai nhóm này.

- Giá trị lớn nhất của roughness: khoảng 3.3% đối với loại grid và khoảng 3.2% với loại honeycomb.
  - Giá trị nhỏ nhất của roughness: khoảng 0.25% đối với loại grid và khoảng 0.5% với loại honeycomb.
  - Mức phân vị Q1( có 25% giá trị đó trở xuống): khoảng 1.1% đối với loại grid và khoảng 1.1% với loại honeycomb.
  - Mức phân vị Q2( có 50% giá trị đó trở xuống): Khoảng 1.4% đối với loại grid và khoảng 1.6% đối với loại honeycomb.
  - Mức phân vị Q3( có 75% giá trị đó trở xuống): Khoảng đối 2.2% với loại grid và khoảng 2.1% honeycomb.
- **Đối với loại vật liệu là abs và pla.**
- Phân phối của abs nhỏ hơn pla.
  - Giá trị lớn nhất của roughness: khoảng 2.4% và có 2 điểm ngoại lai là 3.1% và 3.4% với loại abs và khoảng 3.25% với loại pla.
  - Giá trị nhỏ nhất của roughness: khoảng 0.25% đối với loại abs và khoảng 0.75% với loại pla.
  - Mức phân vị Q1( có 25% giá trị đó trở xuống): khoảng 0.8% đối với loại abs và khoảng 1.5% với loại pla.
  - Mức phân vị Q2( có 50% giá trị đó trở xuống): Khoảng 1.3% đối với loại abs và khoảng 1.75% đối với loại pla.
  - Mức phân vị Q3( có 75% giá trị đó trở xuống): Khoảng 1.6% đối với loại abs và khoảng 2.3% với loại pla.

#### **4.4. Biểu đồ phân tán của các biến độc lập và phụ thuộc**

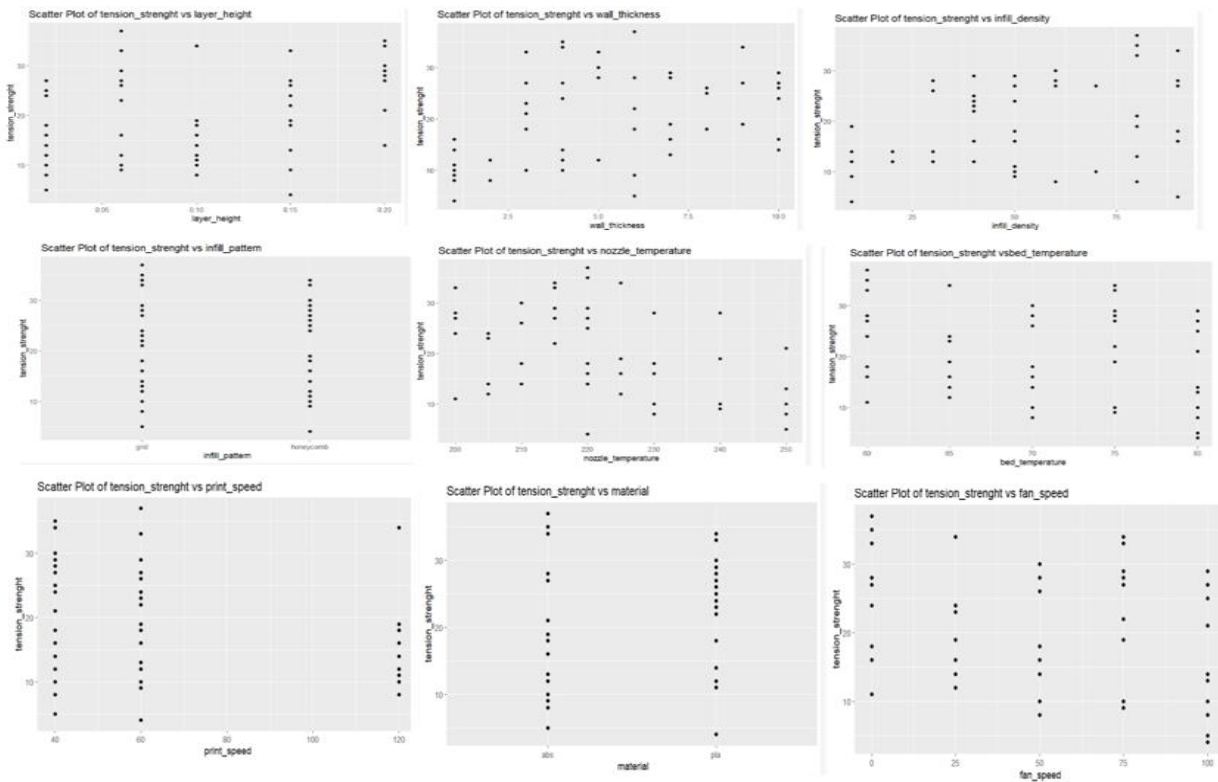
##### **❖ Biến roughness và các biến độc lập**



**Hình 13:** Đồ thị phân tán của biến roughness và các biến độc lập

**Nhận xét:** Dựa vào đồ thị trên ta thấy rằng mối quan hệ của biến roughness và các biến độc lập còn lại. Cụ thể biến roughness và biến layer-height có mối quan hệ tuyến tính mạnh còn đối với các biến khác không thể hiện rõ ràng cho lắm.

### ❖ Biến `tension_strength` và các biến độc lập

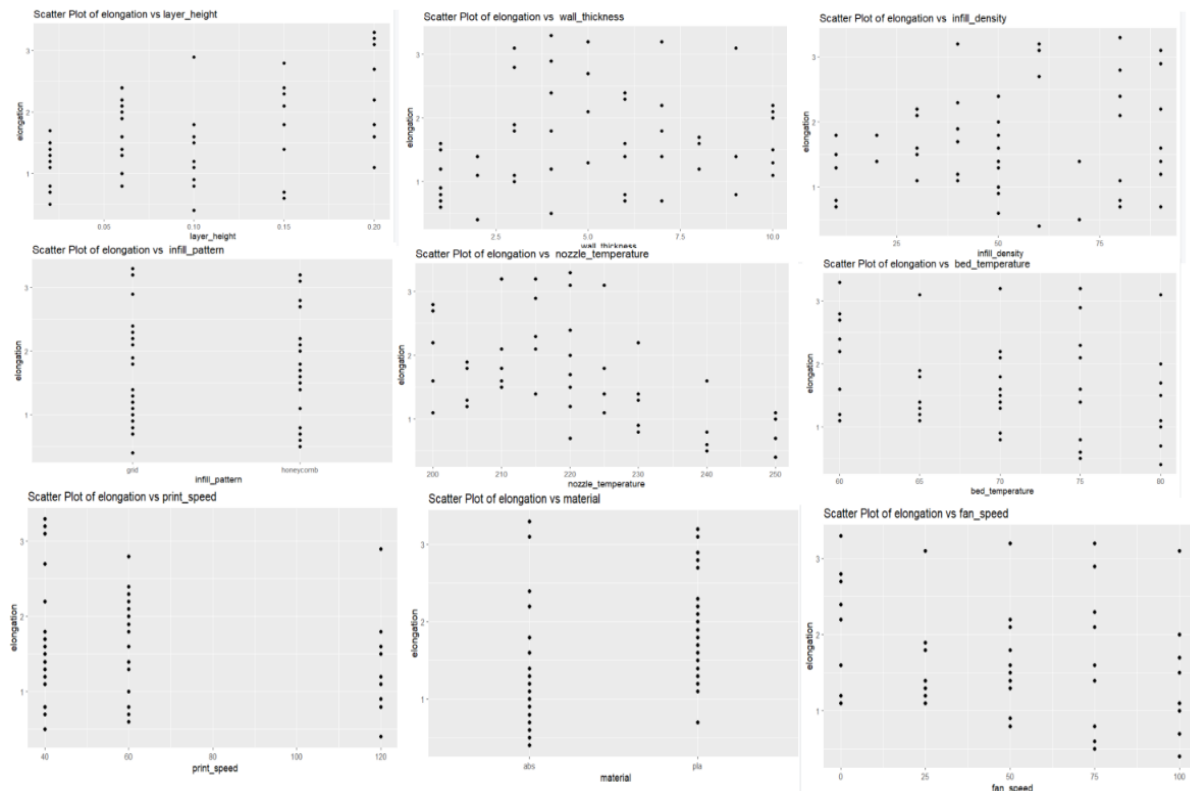


**Hình 14:** Đồ thị phân tán của biến `tension_strength` và các biến độc lập

**Nhận xét:** Dựa vào đồ thị phân bố của biến `tension_strength` và các biến độc lập ta thấy biến `tension_strength` không có mối quan hệ tuyến tính với các biến độc lập này.



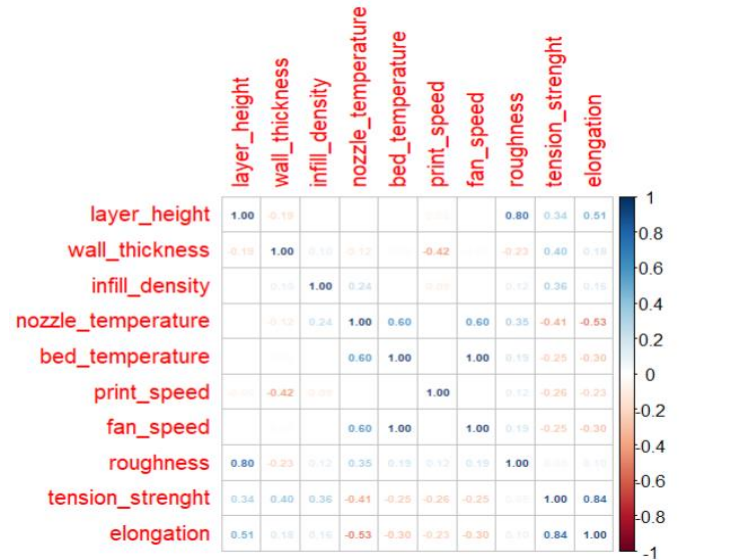
## ❖ Biến elongation và các biến độc lập



**Hình 15:** Đồ thị phân tán của biến elongation và các biến độc lập

**Nhận xét:** Dựa vào đồ thị phân bố của biến elongation và các biến độc lập ta thấy biến elongation không có mối quan hệ tuyến tính với các biến độc lập này.

## 4.5. Ma trận tương quan giữa các biến



**Hình 16:** Ma trận tương quan giữa các biến

### Nhận xét:

Biến roughness và biến layer\_height có mối quan hệ tuyến tính mạnh với nhau hệ số tương quan là 0,8 và không có mối quan hệ tuyến tính mạnh với các biến còn lại.

Biến tension\_strength có mối quan hệ tuyến tính mạnh với biến elongation với hệ số tương quan là 0.84, ngoài ra không có mối quan hệ tuyến tính mạnh với các biến khác.

Biến bed\_temperature và biến fan\_speed có hệ số tương quan bằng 1, có nghĩa 2 biến này có mối quan hệ phụ thuộc lẫn nhau, điều này rây ảnh hưởng đến mô hình, vì vậy cần loại bỏ 1 trong 2 biến này khi xây dựng mô hình.

## **PHẦN 5: THỐNG KÊ SUY DIỄN**

### **5.1 Dựng mô hình hồi quy bội cho độ nhám bề mặt (Roughness)**

Xây dựng mô hình hồi quy đa biến gồm.

- + Biến phụ thuộc: Roughness
- + Biến dự báo ( Biến độc lập): layer\_height; wall\_thickness; infill\_density; infill\_pattern; nozzle\_temperature; bed\_temperature; print\_speed; material.

```
dabien<-lm(data$roughness~data$layer_height+data$wall_thickness+data$infill_density
+data$infill_pattern+data$nozzle_temperature+data$bed_temperature+data$print_speed+data$material)
summary(dabien)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-72.746 -24.332  -1.641   20.304   96.552

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -2.371e+03  3.716e+02  -6.379 1.25e-07 ***
data$layer_height    1.269e+03  8.765e+01  14.483 < 2e-16 ***
data$wall_thickness    2.334e+00  2.189e+00   1.066  0.29259
data$infill_density   -4.231e-02  2.341e-01  -0.181  0.85742
data$infill_patternhoneycomb -1.255e-01  1.128e+01  -0.011  0.99117
data$nozzle_temperature  1.506e+01  2.529e+00   5.953 5.05e-07 ***
data$bed_temperature  -1.613e+01  3.251e+00  -4.962 1.27e-05 ***
data$print_speed     6.496e-01  2.060e-01   3.153  0.00302 **
data$materialpla     2.985e+02  5.836e+01   5.114 7.78e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 38.24 on 41 degrees of freedom
Multiple R-squared:  0.8752,    Adjusted R-squared:  0.8509
F-statistic: 35.95 on 8 and 41 DF,  p-value: 3.834e-16
```

**Hình 17:** Kết quả xây dựng mô hình hồi quy bội 1 cho biến Roughness.

**Kết quả:** Ta có được phương trình hồi quy như sau:

$$\begin{aligned} \text{Roughness} = & -2371 + 1269 \cdot \text{layer\_height} + 2.334 \cdot \text{wall\_thickness} - 0.0423 \cdot \text{infill\_density} \\ & - 0.1225 \cdot \text{infill\_pattern} + 15.06 \cdot \text{nozzle\_temperature} - 16.13 \cdot \text{bed\_temperature} - \\ & 0.6496 \cdot \text{print\_speed} + 298.5 \cdot \text{material}. \end{aligned}$$

**Nhận xét:**

Giá trị  $P\_value = 3.834e-16 \ll$  so với mức ý nghĩa 5%, điều này chỉ ra rằng có ít nhất 1 biến độc lập trong mô hình có ý nghĩa giải thích rất cao cho biến Roughness.

Hệ số hiệu chỉnh R – Squared = 0.8509 tức là mô hình này giải thích được 85,09% sự ảnh hưởng của các biến độc lập đến độ nhám roughness và 14,91% còn lại là do sai số hoặc sự ảnh hưởng của các yếu tố khác.

Roughness sẽ bị ảnh hưởng  $\beta_i$  lần khi ta thay đổi các biến tương ứng lên một đơn vị. Giá trị  $\beta_i$  dương tương ứng với sự ảnh hưởng theo chiều tăng và ngược lại.

### Kiểm định giả thuyết:

Giả thuyết  $H_0: \beta_i = 0$  ( hệ số hồi quy không có ý nghĩa thống kê)

Giả thuyết  $H_1: \beta_i \neq 0$  ( hệ số hồi quy có ý nghĩa)

Giá trị P\_value của wall\_thickness; infill\_density và infill\_pattern lớn hơn mức ý nghĩa 5% vì vậy chúng ta không đủ chứng cứ để bác bỏ giả thuyết  $H_0$  vậy giả thuyết này là hệ số hồi quy không có ý nghĩa thống kê là đúng, vì vậy chúng ta có thể loại bỏ các biến này ra khỏi mô hình.

**Xây dựng mô hình hồi quy bội thứ 2 cho biến roughness sau khi lược bỏ các biến wall\_thickness; infill\_density và infill\_pattern.**

```
dabien2<-lm(data$roughness~data$layer_height+data$nozzle_temperature+data$bed_temperature
+data$print_speed+data$material)
summary(dabien2)
```

Coefficients:					
	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-2310.7356	353.2009	-6.542	5.38e-08	***
data\$layer_height	12.4654	0.8318	14.986	< 2e-16	***
data\$nozzle_temperature	14.7774	2.3979	6.163	1.95e-07	***
data\$bed_temperature	-15.8078	3.0895	-5.117	6.55e-06	***
data\$print_speed	0.5538	0.1804	3.070	0.00366	**
data\$material1	294.1610	56.1586	5.238	4.38e-06	***

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

Residual standard error: 37.44 on 44 degrees of freedom  
 Multiple R-squared: 0.8717, Adjusted R-squared: 0.8571  
 F-statistic: 59.78 on 5 and 44 DF, p-value: < 2.2e-16

**Hình 18: Kết quả mô hình hồi quy bội thứ 2 cho biến roughness**

**Kết quả:**

$$\text{Roughness} = -2310.736 + 12.4654 \cdot \text{layer\_height} + 14.7774 \cdot \text{nozzle\_temperature} - 15.8075 \cdot \text{bed\_temperature} + 0.5538 \cdot \text{print\_speed} + 294.1610 \cdot \text{material}.$$

**Nhận xét:**

Giá trị  $P\_value = 2.2e-16 \ll$  so với mức ý nghĩa 5%, điều này chỉ ra rằng có ít nhất 1 biến độc lập trong mô hình có ý nghĩa giải thích rất cao cho biến Roughness.

Roughness sẽ bị ảnh hưởng  $\beta_i$  lần khi ta thay đổi các biến tương ứng lên một đơn vị. Giá trị  $\beta_i$  dương tương ứng với sự ảnh hưởng theo chiều tăng và ngược lại.

Giá trị của hệ số hiệu chỉnh R-Squared = 0.8571. Mô hình này giải thích được 85,71% sự ảnh hưởng của các yếu tố đầu vào đến độ nhám roughness và 14,29% còn lại là do sai số hoặc các yếu tố khác.

**So sánh hiệu quả của 2 mô hình**

Đặt giả thuyết:

$H_0$ : Hai mô hình có hiệu quả bằng nhau.

$H_1$ : Hai mô hình có hiệu quả khác nhau

```
#so sánh hiệu quả của 2 mô hình.  
anova(dabien, dabien2)
```

```

Analysis of Variance Table

Model 1: data$roughness ~ data$layer_height + data$wall_thickness + data$infill_density +
  data$infill_pattern + data$nozzle_temperature + data$bed_temperature +
  data$print_speed + data$material
Model 2: data$roughness ~ data$layer_height + data$nozzle_temperature +
  data$bed_temperature + data$print_speed + data$material
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      41 59968
2      44 61667 -3   -1699.4 0.3873 0.7627

```

**Hình 19:** Kết quả so sánh 2 mô hình

### **Nhận xét:**

Khi so sánh hai mô hình ta thấy giá trị P-value lớn hơn so với mức ý nghĩa 5% vì vậy ta chưa thể bác bỏ  $H_0$ . Nhưng ta nhận thấy rằng giá trị R-Squared = 0.8571 ở mô hình 2 lớn hơn so với mức 0.8509 ở mô hình một. Vậy ta có thể kết luận rằng mô hình 2 hiệu quả hơn mô hình một.

## **5.2 Kiểm định các giả định của mô hình hồi quy tuyến tính bội.**

Các giả định của mô hình hồi quy:  $Y_i = \beta_0 + \beta_1 * X_1 + \dots + \beta_i * X_i + \varepsilon_i, i=1,2,\dots,n$

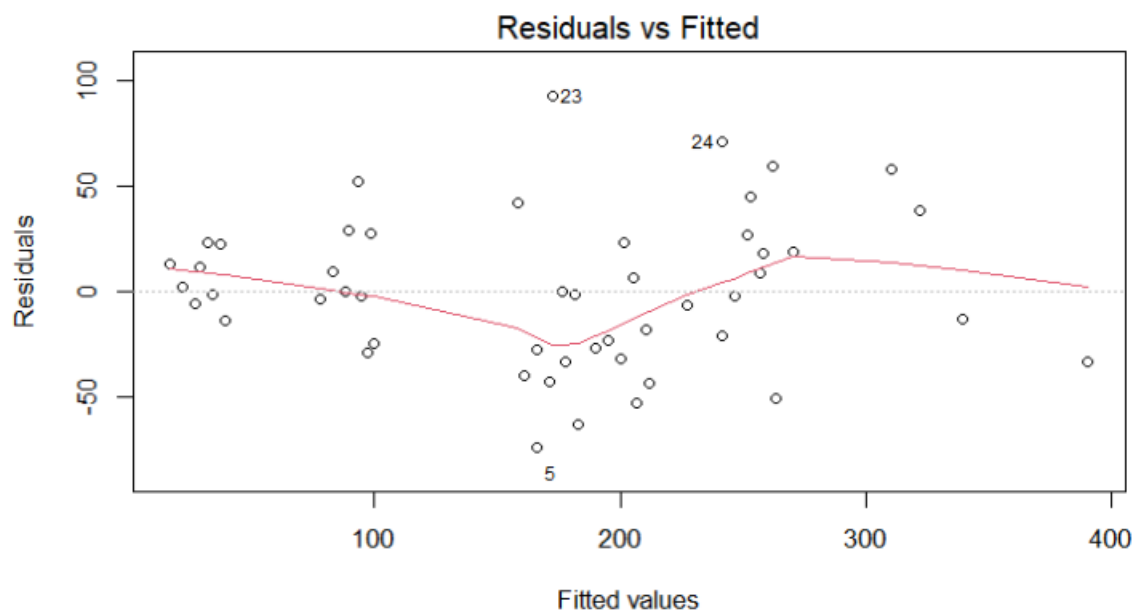
- Giả định 1: Tuyến tính của dữ liệu, mối quan hệ giữa biến dự báo X và biến phụ thuộc Y dự báo là tuyến tính ( đã được thực hiện ở trên )
- Các sai số có kỳ vọng bằng 0
- Sai số có phân phối chuẩn
- Phương sai của các sai số là hằng số

Thực hiện phân tích phần dư và vẽ các biểu đồ để kiểm tra các giả định trên.

```

fitted(dabien2)
resid(dabien2)
vephandu<-par(mfrow=c(2,2))
plot(dabien2)

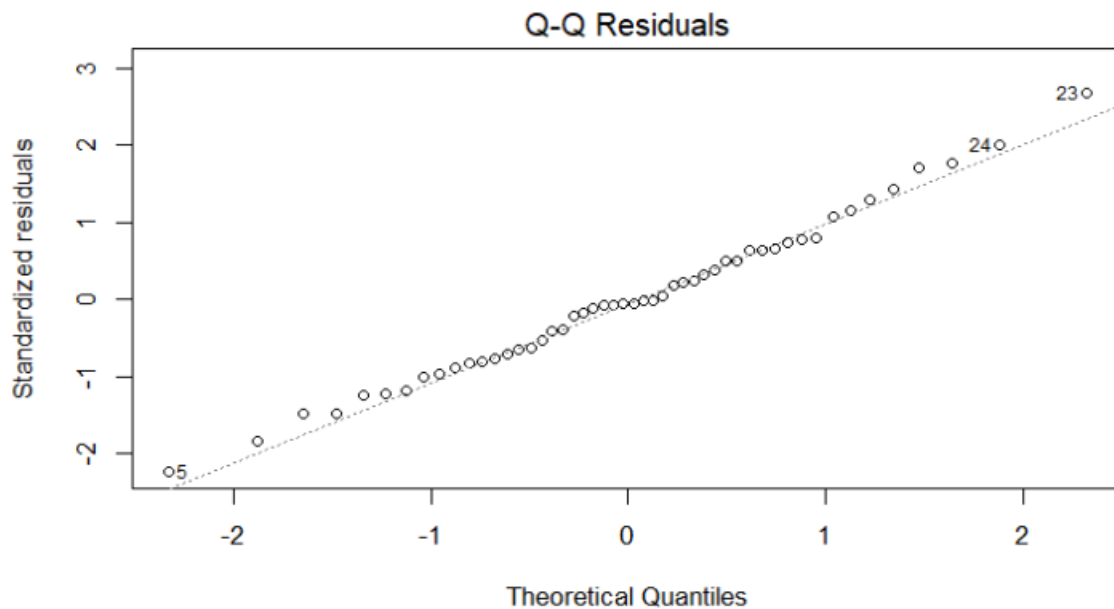
```



**Hình 20:** Các giá trị thặng dư tương ứng với các giá trị dự báo

### Nhận xét:

Đồ thị (Residuals vs Fitted) kiểm tra tính tuyến tính và các sai số  $\epsilon_i$  có giá trị trung bình bằng không. Trục tung biểu thị giá trị của phần dư, trục hoành biểu thị giá trị tiên lượng ( $\hat{y}_i$ ) của biến phụ thuộc. Nếu như giả thiết về tính tuyến tính của dữ liệu KHÔNG thỏa, ta sẽ quan sát thấy rằng đường màu đỏ trên đồ thị sẽ phân bố theo một hình mẫu (pattern) đặc trưng nào đó (ví dụ parabol). Nếu đường màu đỏ trên đồ thị phân tán là đường thẳng nằm ngang mà không phải là đường cong, thì giả thiết tính tuyến tính của dữ liệu được thỏa mãn. Giả thiết phần dư có trung bình bằng 0 thỏa mãn nếu đường màu đỏ gần với đường nằm ngang (ứng với phần dư = 0). Vậy giả định 1 và 2 là đúng.

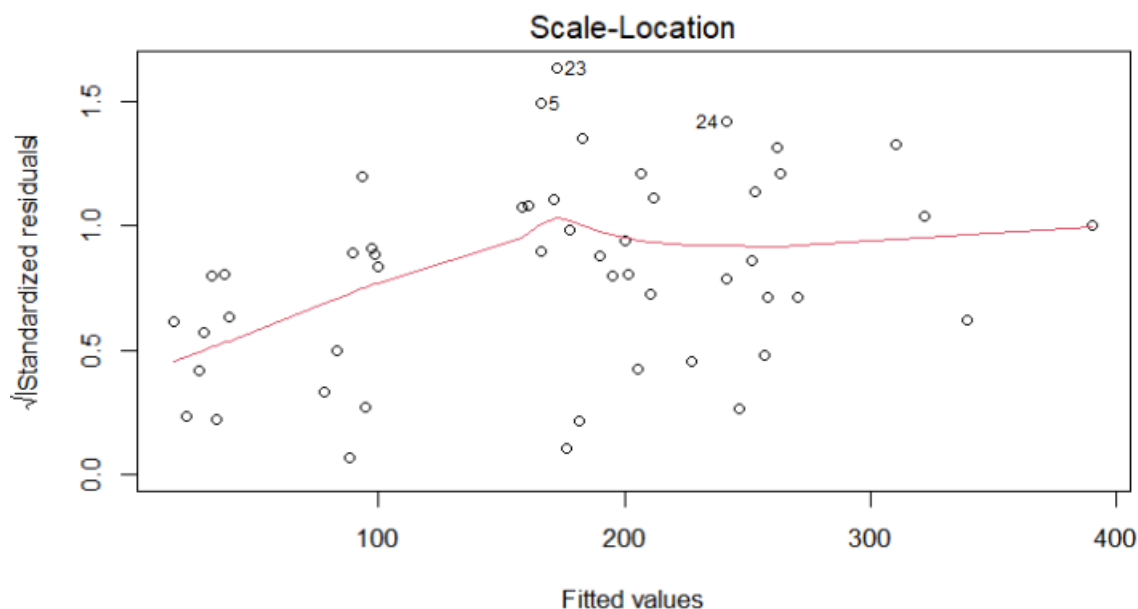


**Hình 21:** Thể hiện các sai số đã được chuẩn hóa

**Nhận xét:**

Đồ thị Normal Q-Q dùng để kiểm tra giả thiết phần dư có phân phối chuẩn. Nếu các điểm thặng dư nằm trên cùng 1 đường thẳng thì điều kiện về phân phối chuẩn được thỏa. Nhìn vào đồ thị trên xem xét được giả định 3 là tạm chấp nhận được.

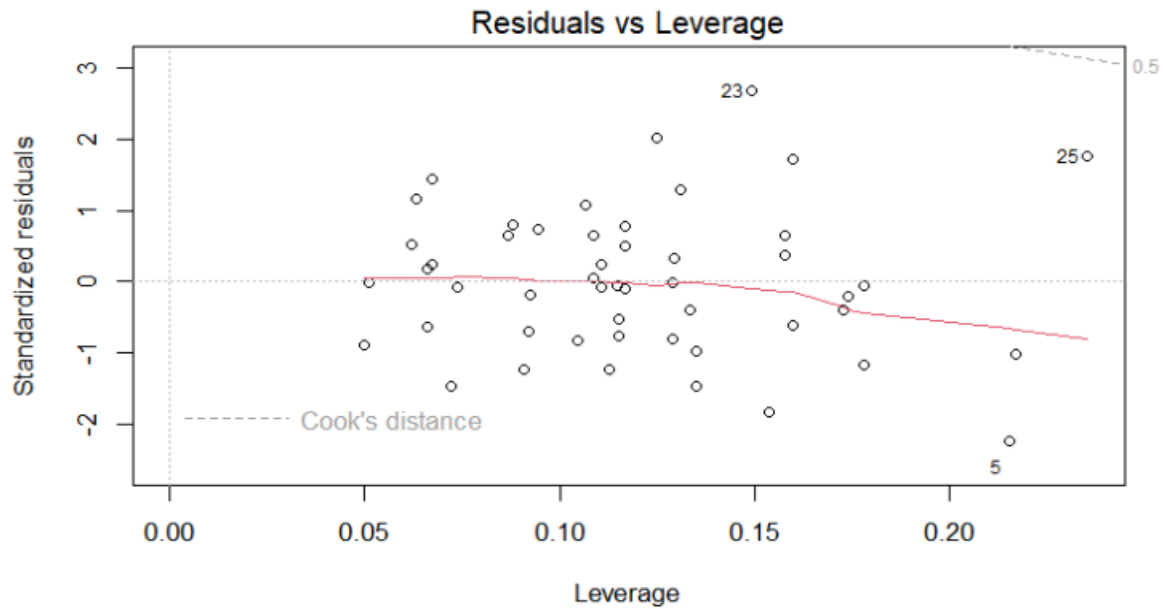




**Hình 22 :** Thể hiện phương sai của phần dư ( sai số )

**Nhận xét:**

Đồ thị (Scale – Location) dùng để kiểm định giả thiết phương sai của phần dư là không đổi. Trục tung là căn bậc hai của giá trị của phần dư (đã được chuẩn hóa), trục hoành là giá trị tiên lượng ( $\hat{y}_i$ ) của biến phụ thuộc từ mô hình. Nếu như đường màu đỏ trên đồ thị là đường thẳng nằm ngang và các điểm thặng dư phân tán đều xung quanh đường thẳng này thì giả thiết thứ 4 được thỏa. Nếu như đường màu đỏ có độ dốc (hoặc cong) hoặc các điểm thặng dư phân tán không đều xung quanh đường thẳng này, thì giả thiết này sai. Vậy theo các đồ thị trên thì giả thuyết phương sai của các sai số là hằng số là đúng.



**Hình 23:** Thể hiện những điểm có ảnh hưởng cao trong dữ liệu

### **Nhận xét :**

Đồ thị Residuals vs Leverage vẽ những điểm ảnh hưởng cao trong bộ dữ liệu.

### ***KẾT LUẬN***

Các mô hình hồi quy tuyến tính ta xây dựng được cho các biến phụ thuộc gồm độ nhám bề mặt ( Roughness ) là thỏa mãn các giả định của một mô hình hồi quy tuyến tính bội.

Độ nhám bề mặt (Roughness) sẽ chịu sự ảnh hưởng của các yếu tố như chiều cao lớp (layer\_height); nhiệt độ vòi phun (nozzle\_temperature); nhiệt độ giường (bed temperature) ; tốc độ in (print\_speed); và loại vật liệu (material). Sự ảnh hưởng của các yếu tố này đến độ nhám được thể hiện qua mô hình hồi quy bội bên dưới và mô hình này giải thích được 85,71% sự ảnh hưởng của các yếu tố đầu vào đến độ nhám roughness và 14,29% còn lại là do sai số hoặc các yếu tố khác.

$$\text{Roughness} = -2310.736 + 12.4654 \cdot \text{layer\_height} + 14.7774 \cdot \text{nozzle\_temperature} - 15.8075 \cdot \text{bed\_temperature} + 0.5538 \cdot \text{print\_speed} + 294.1610 \cdot \text{material}.$$

Bằng cách xây dựng mô hình hồi quy tuyến tính bội tương tự như Roughness cho hai biến phụ thuộc còn lại là tension\_strength và elongation ta sẽ tìm được sự ảnh hưởng của các biến đầu vào đến chúng.

## PHẦN 6: THẢO LUẬN VÀ MỞ RỘNG

Từ việc xây dựng mô hình hồi quy tuyến tính bội để xác định được sự ảnh hưởng của các yếu tố đầu vào đến chất lượng của sản phẩm, cụ thể là sự ảnh hưởng của chiều cao lớp (layer\_height); nhiệt độ vòi phun (nozzle\_temperature); nhiệt độ giường (bed temperature) ; tốc độ in (print\_speed); và loại vật liệu (material) đến độ nhám của sản phẩm (roughness) sẽ đem lại những lợi ích sau:

- Tối ưu hóa được chất lượng sản phẩm.
- Tiết kiệm được nguyên vật liệu, thời gian và chi phí.
- Tăng khả năng điều chỉnh các thông số hợp lý.
- Nâng cao hiệu suất của máy in 3D.
- Xác định được sự ảnh hưởng của các yếu tố đầu vào đến chất lượng sản phẩm.

Tuy nhiên cần phải xác định chính xác các giá trị của các thông số có vai trò quan trọng trong việc xác định được sự ảnh hưởng của các yếu tố đầu vào đến đầu ra.

## PHẦN 7: NGUỒN CODE VÀ DỮ LIỆU

Link code và dữ liệu:

Link dữ liệu: <https://www.kaggle.com/datasets/afumetto/3dprinter>

Link code: [https://cran.r-project.org/doc/contrib/Intro\\_to\\_R\\_Vietnamese.pdf](https://cran.r-project.org/doc/contrib/Intro_to_R_Vietnamese.pdf)

## PHẦN 8: TÀI LIỆU THAM KHẢO

1. AWS. *Mô hình hồi quy*. Truy cập từ: <https://aws.amazon.com/vi/what-is/linear-regression/>
2. In3D. *Máy in 3-D Ultimaker S5*. Truy cập từ: <https://in3ds.com/may-in-3d/ultimaker/c-hat-luong-cao-ultimaker-s5/>
3. Kaggle. *3D Printer Dataset for Mechanical Engineers*: Truy cập từ: <https://www.kaggle.com/datasets/afumetto/3dprinter>

4. Nghiên cứu giáo dục. *Mô hình hồi quy tuyến tính bội*. Truy cập từ:  
<https://nghiencuugiaoduc.com.vn/bai-3-phan-tich-hoi-quy-tuyen-tinh-boi-multiple-linear-regression/>
5. Nguyễn Thanh Nga – Học viện Ngân Hàng. (22/09/2021). *Mô hình hồi quy bội*. Truy cập từ: <https://rpubs.com/TKUD/810985>