Phân tích các yếu tố ảnh hưởng đến độ nhám bề mặt trong in 3D

## III. Tiền xử lý dữ liệu

data <- read.csv("~/BTL\_XSTK/data.csv", header = TRUE, sep = ",")  
head(data, 10) #Xuát 10 dòng đầu của dữ liệu

## layer\_height wall\_thickness infill\_density infill\_pattern nozzle\_temperature  
## 1 0.02 8 90 grid 220  
## 2 0.02 7 90 honeycomb 225  
## 3 0.02 1 80 grid 230  
## 4 0.02 4 70 honeycomb 240  
## 5 0.02 6 90 grid 250  
## 6 0.02 10 40 honeycomb 200  
## 7 0.02 5 10 grid 205  
## 8 0.02 10 10 honeycomb 210  
## 9 0.02 9 70 grid 215  
## 10 0.02 8 40 honeycomb 220  
## bed\_temperature print\_speed material fan\_speed roughness tension\_strenght  
## 1 60 40 abs 0 25 18  
## 2 65 40 abs 25 32 16  
## 3 70 40 abs 50 40 8  
## 4 75 40 abs 75 68 10  
## 5 80 40 abs 100 92 5  
## 6 60 40 pla 0 60 24  
## 7 65 40 pla 25 55 12  
## 8 70 40 pla 50 21 14  
## 9 75 40 pla 75 24 27  
## 10 80 40 pla 100 30 25  
## elongation  
## 1 1.2  
## 2 1.4  
## 3 0.8  
## 4 0.5  
## 5 0.7  
## 6 1.1  
## 7 1.3  
## 8 1.5  
## 9 1.4  
## 10 1.7

**1. Làm sạch dữ liệu**

Các biến trong dữ liệu gồm:

* layer\_height: Chiều cao layer
* wall\_thickness: Độ dày
* infill\_density: Mật độ lấp đầy
* infill\_pattern: Mật độ mô hình
* nozzle\_temperature: Nhiệt độ vòi phun
* bed\_temperature: Nhiệt độ khay
* print\_speed: Tốc độ in
* material: Vật liệu
* fan\_speed: Tốc độ quạt
* roughness: Độ nhám bề mặt

Ta tạo một dữ liệu mới tên **new\_data** gồm các biến ta quan tâm trong dữ liệu:

selected\_columns <- c("layer\_height", "wall\_thickness", "infill\_density", "infill\_pattern",  
 "nozzle\_temperature", "bed\_temperature", "print\_speed", "material",  
 "fan\_speed", "roughness")  
new\_data <- data[, selected\_columns]

**2. Kiểm tra dữ liệu khuyết:**

freq.na(new\_data)

## missing %  
## layer\_height 0 0  
## wall\_thickness 0 0  
## infill\_density 0 0  
## infill\_pattern 0 0  
## nozzle\_temperature 0 0  
## bed\_temperature 0 0  
## print\_speed 0 0  
## material 0 0  
## fan\_speed 0 0  
## roughness 0 0

**Nhận xét:** Không có dữ liệu nào bị khuyết.

## IV. Thống kê mô tả

**1. Tạo function và lập bảng tính thống kê mô tả cho các biến liên tục:**

conts\_var <- data[, c("layer\_height", "wall\_thickness", "infill\_density",   
 "nozzle\_temperature", "bed\_temperature", "print\_speed",   
 "fan\_speed", "roughness")]

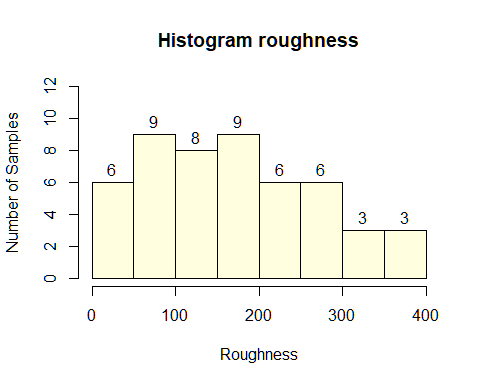
Trong bảng gồm các giá trị.

mean\_val <- apply(conts\_var, 2, mean)  
sd\_val <- apply(conts\_var, 2, sd)  
var\_val <- apply(conts\_var, 2, var)  
median\_val <- apply(conts\_var, 2, median)  
min\_val <- apply(conts\_var, 2, min)  
max\_val <- apply(conts\_var, 2, max)  
quantile1 <- apply(conts\_var, 2, function(x) quantile(x, probs = 0.25))  
quantile3 <- apply(conts\_var, 2, function(x) quantile(x, probs = 0.75))  
#Trung bình, độ lệch chuẩn, biến số, trung vị, giá trị nhỏ nhất, giá trị lớn nhất, phân vị đơn, phân vị 3  
  
summary\_stats <- data.frame(mean = mean\_val, sd = sd\_val, var = var\_val,   
 median = median\_val, min = min\_val, max = max\_val,   
 quantile1 = quantile1, quantile3 = quantile3)  
knitr::kable(t(summary\_stats))

|  | layer\_height | wall\_thickness | infill\_density | nozzle\_temperature | bed\_temperature | print\_speed | fan\_speed | roughness |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| mean | 0.1060000 | 5.220000 | 53.40000 | 221.50000 | 70.000000 | 64.0000 | 50.00000 | 170.58000 |
| sd | 0.0643967 | 2.922747 | 25.36348 | 14.82035 | 7.142857 | 29.6923 | 35.71429 | 99.03413 |
| var | 0.0041469 | 8.542449 | 643.30612 | 219.64286 | 51.020408 | 881.6327 | 1275.51020 | 9807.75878 |
| median | 0.1000000 | 5.000000 | 50.00000 | 220.00000 | 70.000000 | 60.0000 | 50.00000 | 165.50000 |
| min | 0.0200000 | 1.000000 | 10.00000 | 200.00000 | 60.000000 | 40.0000 | 0.00000 | 21.00000 |
| max | 0.2000000 | 10.000000 | 90.00000 | 250.00000 | 80.000000 | 120.0000 | 100.00000 | 368.00000 |
| quantile1 | 0.0600000 | 3.000000 | 40.00000 | 210.00000 | 65.000000 | 40.0000 | 25.00000 | 92.00000 |
| quantile3 | 0.1500000 | 7.000000 | 80.00000 | 230.00000 | 75.000000 | 60.0000 | 75.00000 | 239.25000 |

**2. Vẽ đồ thị histogram thể hiện phân phối cho biến roughness:**

hist(new\_data$roughness, xlab = "Roughness", ylab = "Number of Samples", main = "Histogram roughness",  
 col = "lightyellow", labels = TRUE, ylim = c(0, 12))



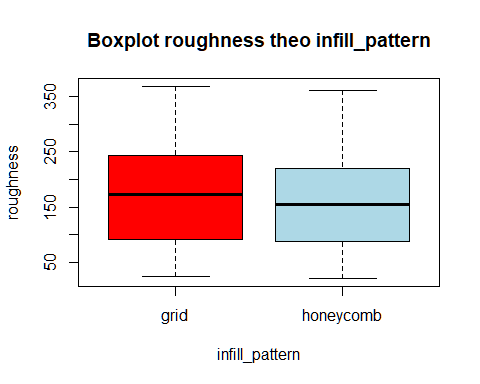
**Nhận xét:**

* Đồ thị phân phối hơi lệch về phải, không có phân phối chuẩn.
* Phân bố tần số cao nhất trong khoảng (50-200), và thấp nhât trong khoảng (300-400).

**3. Vẽ các đồ thị boxplot thể hiện phân phối của biến roughness theo các biến phân loại.**

**3.1. Biểu đồ hộp so sánh độ nhám giữa các loại biến mật độ mô hình:**

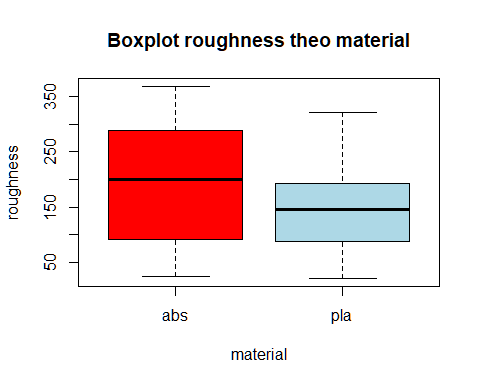
boxplot(roughness ~ infill\_pattern, data = new\_data, col = c("red", "lightblue"),  
 main = "Boxplot roughness theo infill\_pattern")



**Nhận xét:** Không có sự khác biệt nhiều về phân phối của độ nhám ở 2 nhóm infill\_pattern, ta dự đoán yếu tố infill\_pattern không ảnh hưởng đến độ.

**2.1. Biểu đồ hộp so sánh độ nhám theo vật liêu:**

boxplot(roughness ~ material, data = new\_data, col = c("red", "lightblue"),  
 main = "Boxplot roughness theo material")

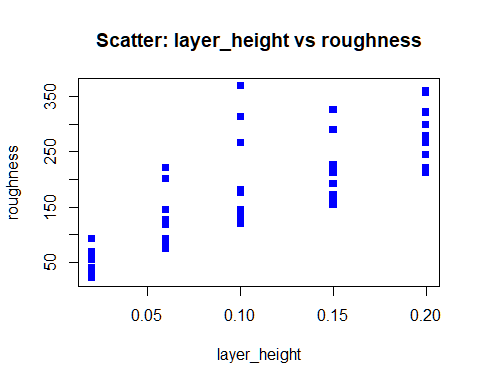


**Nhận xét:**

* Nhìn chung sự phân phối của 2 nhóm material cũng tương đồng nhưng ở loại vật liệu abs thì có độ nhám tối đa cao hơn loại vật liệu pla.
* Về khoảng phân bố thì loại vật liệu pla thấp hơn so với loại vật liệu abs, ta dự đoán yếu tố material có ảnh hưởng đến biến phụ thuộc roughness.

**3.3. Biểu đồ phân tán chiều cao và độ nhám:**

plot(new\_data$layer\_height, new\_data$roughness, col = "blue", pch = 15,  
 xlab = "layer\_height", ylab = "roughness", main = "Scatter: layer\_height vs roughness")



**Nhận xét:**

* Các điểm phân bố có xu hướng tăng lên khi biến layer\_height tăng dần.
* Các điểm không nằm quá sát nhau trên cùng 1 đường thẳng  
  ⇒ Có thể có quan hệ tuyến tính với nhau nhưng sẽ chỉ ở mức độ trung bình  
  ⇒ Khi biến layer\_height thay đổi thì biến roughness cũng thay đổi theo hay nói cách khác, biến layer\_height có ảnh hưởng đến biến roughness.

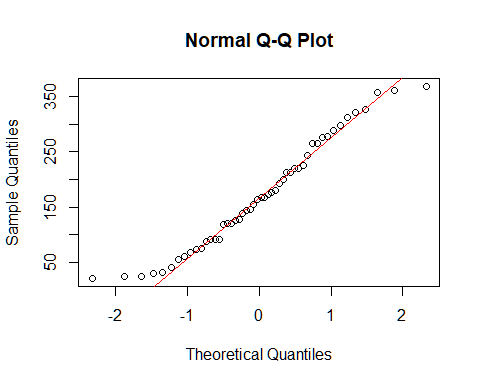
## V. Thống kê suy diễn.

**1. Bài toán 1 mẫu**

**Bài toán:** Với mức ý nghĩa 5%, cho kết luận độ nhám trung bình của bản in là 170 µm hay không?

*Kiểm tra giả định về phân phối chuẩn cho biến độ nhám:*

qqnorm(data$roughness)  
qqline(data$roughness, col = "red")



**Nhận xét:**

Dựa vào đồ thị ta nhận thấy đa số các quan trắc nằm xung quanh đường thẳng kì vọng phân phối chuẩn, ta có thể kết luận rằng độ nhám có phân phối chuẩn.

*Ngoài ra, ta có thể dùng kiểm định shapiro.test để kiểm tra:*

* Giả thiết H\_0: Độ nhám tuân theo phân phối chuẩn
* Giả thiết H\_1: Độ nhám không tuân theo phân phối chuẩn

shapiro.test(new\_data$roughness)

##   
## Shapiro-Wilk normality test  
##   
## data: new\_data$roughness  
## W = 0.95919, p-value = 0.08221

**Nhận xét:**

Vì p-value = 0.08221 > mức ý nghĩa 5% nên ta chưa bác bỏ được H\_0.

Vậy ta có thể kết luận rằng độ nhám có phân phối chuẩn.

⇒ Đây là dạng bài kiểm định trung bình 1 mẫu, X có phân phối chuẩn, chưa biết 𝜎2

Gọi 𝜇 là là độ nhám trung bình của bản in thực tế.

* Giả thiết H\_0: 𝜇 =170
* Giả thiết H\_1: 𝜇 ≠170

*Thực hiện kiểm định bằng t-test:*

t.test(new\_data$roughness, mu = 170)

##   
## One Sample t-test  
##   
## data: new\_data$roughness  
## t = 0.041412, df = 49, p-value = 0.9671  
## alternative hypothesis: true mean is not equal to 170  
## 95 percent confidence interval:  
## 142.4348 198.7252  
## sample estimates:  
## mean of x   
## 170.58

**Nhận xét:** Vì p-value = 0.9671 > mức ý nghĩa 5% nên ta chưa bác bỏ được H\_0.

Vậy nên ta có thể kết luận độ nhám trung bình của bản in là 170 µm, xét mức ý nghĩa 5%

*Ngoài ra ta có thể đưa ra kết luận bằng cách:*

Tiêu chuẩn kiểm định t0 = 0,0414.

Miền bác bỏ: RR = (−∞;−𝑡𝛼/2;𝑛−1) U (𝑡𝛼/2;𝑛−1;+ ∞)

* Với 𝑡𝛼/2;𝑛−1 được tính bằng công thức:

qt(p = 0.05/2, df = nrow(new\_data)-1, lower.tail = FALSE)

## [1] 2.009575

RR = (−∞; - 2.0096) U (2.0096 ;+ ∞)

Vì t0 ∉ RR nên ta chưa bác bỏ H\_0.

Vậy *ta có thể kết luận độ nhám trung bình của bản in là 170 µm, xét mức ý nghĩa 5%*.

**2. Kiểm định 2 mẫu.**

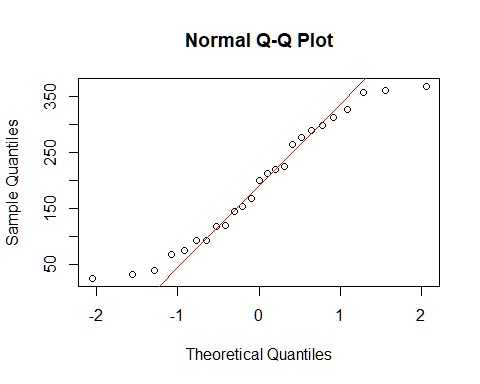
**Bài toán:** Với mức ý nghĩa 0.05%, hãy so sánh độ nhám trung bình của bản in khi sử dụng vật liệu abs và pla.

* Chia bộ dữ liệu theo 2 nhóm vật liệu:

abs\_data <- subset(data, material == "abs")  
pla\_data <- subset(data, material == "pla")

*Kiểm tra giả định về phân phối chuẩn cho biến độ nhám ở loại vật liệu abs:*

qqnorm(abs\_data$roughness); qqline(abs\_data$roughness, col = "red")



**Nhận xét:**

Dựa vào đồ thị ta nhận thấy đa số các quan trắc nằm xung quanh đường thẳng kì vọng phân phối chuẩn, ta có thể kết luận rằng độ nhám ở vật liệu abs có phân phối chuẩn.

*Ngoài ra, ta có thể dùng kiểm định shapiro.test để kiểm tra:*

* Giả thiết H\_0: Độ nhám ở vật liệu abs tuân theo phân phối chuẩn
* Giả thiết H\_1: Độ nhám ở vật liệu abs không tuân theo phân phối chuẩn

shapiro.test(abs\_data$roughness)

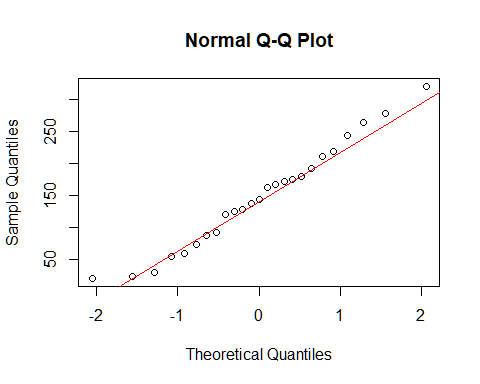
##   
## Shapiro-Wilk normality test  
##   
## data: abs\_data$roughness  
## W = 0.94235, p-value = 0.1677

**Nhận xét:** Vì p-value = 0.1677 > mức ý nghĩa 5% nên ta chưa bác bỏ được H\_0.

Vậy ta có thể kết luận rằng độ nhám ở vật liệu abs có phân phối chuẩn.

*Kiểm tra giả định về phân phối chuẩn cho biến độ nhám ở loại vật liệu pla:*

qqnorm(pla\_data$roughness); qqline(pla\_data$roughness, col = "red")



**Nhận xét:**

Dựa vào đồ thị ta nhận thấy đa số các quan trắc nằm xung quanh đường thẳng kì vọng phân phối chuẩn, ta có thể kết luận rằng độ nhám ở vật liệu pla có phân phối chuẩn.

*Ngoài ra, ta có thể dùng kiểm định shapiro.test để kiểm tra:*

* Giả thiết H\_0: Độ nhám ở vật liệu pla tuân theo phân phối chuẩn
* Giả thiết H\_1: Độ nhám ở vật liệu pla không tuân theo phân phối chuẩn

shapiro.test(pla\_data$roughness)

##   
## Shapiro-Wilk normality test  
##   
## data: pla\_data$roughness  
## W = 0.97437, p-value = 0.7561

**Nhận xét:** Vì p-value = 0.1677 > mức ý nghĩa 5% nên ta chưa bác bỏ được H\_0.

Vậy ta có thể kết luận rằng độ nhám ở vật liệu pla có phân phối chuẩn.

*Thực hiện so sánh phương sai độ nhám ở 2 nhóm vật liệu:*

* Giả thiết H\_0: σ\_1^2= σ\_2^2 hay (σ\_1^2≤ σ\_2^2)
* Giả thiết H\_1: σ\_1^2> σ\_2^2

var.test(abs\_data$roughness, pla\_data$roughness, alternative = "greater")

##   
## F test to compare two variances  
##   
## data: abs\_data$roughness and pla\_data$roughness  
## F = 1.8454, num df = 24, denom df = 24, p-value = 0.07024  
## alternative hypothesis: true ratio of variances is greater than 1  
## 95 percent confidence interval:  
## 0.9302656 Inf  
## sample estimates:  
## ratio of variances   
## 1.845423

**Nhận xét:** Vì p-value = 0.07024 > mức ý nghĩa 5% nên ta chưa bác bỏ được H\_0.

Vậy phương sai độ nhóm ở 2 nhóm vật liệu bằng nhau.

⇨ Đây là dạng bài kiểm định trung bình 2 mẫu độc lập , X1, X2 có phân phối chuẩn, chưa biết σ\_1^2, σ\_2^2 với σ\_1^2= σ\_2^2

Gọi 𝜇1,𝜇2 lần lượt là độ nhám trung bình của bản in khi sử dụng vật liệu abs và pla.

* Giả thiết H\_0: 𝜇\_1 = 𝜇\_2
* Giả thiết H\_1: 𝜇\_1 > 𝜇\_2

*Thực hiện kiểm định bằng t-test:*

t.test(roughness ~ material, data = data, alternative = "greater", var.equal = TRUE)

##   
## Two Sample t-test  
##   
## data: roughness by material  
## t = 1.6613, df = 48, p-value = 0.05159  
## alternative hypothesis: true difference in means between group abs and group pla is greater than 0  
## 95 percent confidence interval:  
## -0.4392901 Inf  
## sample estimates:  
## mean in group abs mean in group pla   
## 193.44 147.72

**Nhận xét:** Vì p-value = 0.05159 > mức ý nghĩa 5% nên ta chưa bác bỏ được H\_0.

Vậy nên độ nhám trung bình ở 2 nhóm vật liệu bằng nhau, xét với mức ý nghĩa 5%.

**3. Bài toán phân tích phương sai**

**3.1. Tính toán các giá trị**

**Bài toán:** So sánh độ nhám trung bình ở 5 nhóm layer\_height.

**Phân tích:** ANOVA là bài toán so sánh trung bình của các tổng thể nhưng với yêu cầu mỗi biến ảnh hưởng phải có 3 phân loại trở lên.

Trong tập dữ liệu ta có 2 biến phân loại là infill\_patten và material nhưng mỗi biến thì chỉ có 2 phân loại là biến infill\_pattern bao gồm grid và honeycomb, biến material bao gồm pla và abs.

⇒ Giải pháp: Dùng biến layer\_heght và chia các giá trị thành các nhóm.

data$layer\_height <- as.factor(data$layer\_height)  
print(data$layer\_height)

## [1] 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.06 0.06 0.06 0.06 0.06  
## [16] 0.06 0.06 0.06 0.06 0.06 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1   
## [31] 0.15 0.15 0.15 0.15 0.15 0.15 0.15 0.15 0.15 0.15 0.2 0.2 0.2 0.2 0.2   
## [46] 0.2 0.2 0.2 0.2 0.2   
## Levels: 0.02 0.06 0.1 0.15 0.2

Lúc này các giá trị trong biến layer\_height đã được chuyển sang dạng factor để phân loại.

Biến layer\_height đã có 5 nhóm lần lượt là: 0.02, 0.06, 0.1, 0.15, 0.2. Đặt giả thuyết

H\_0: 𝜇\_1=𝜇\_2=𝜇\_3=𝜇\_4=𝜇\_5

H\_1: Tồn tại 𝜇\_i≠𝜇\_j (i≠j)

*Code thực hiện hàm anova, ta sử dụng hàm aov() để tính toán các giá trị.*

anova\_model <- aov(roughness ~ layer\_height, data)  
summary(anova\_model)

## Df Sum Sq Mean Sq F value Pr(>F)   
## layer\_height 4 326957 81739 23.94 1.17e-10 \*\*\*  
## Residuals 45 153624 3414   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Ta thấy giá trị của Pr(>F) = 1.17e-10 nhỏ hơn mức ý nghĩa 5%

⇒ Bác bỏ H\_0, chấp nhận H\_1

⇒ Có sự khác biệt về độ nhám trung bình giữa 5 nhóm chiều cao trong biến layer\_height.

**3.2. Kiểm tra điều kiện mô hình ANOVA**

Điều kiện thứ nhất:

* Các quan sát từ các tổng thể dược lấy độc lập.
* Các giá trị độ nhám trong dự liệu phải được lấy độc lập.
* Điều kiện thứ nhất là hiển nhiên thỏa mãn do mỗi đơn vị dữ liệu được thu thập sau mỗi lần in.

Điều kiện thứ hai:

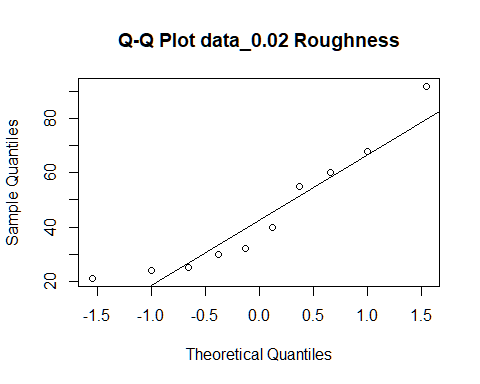
* Các tổng thể phải có phân phối chuẩn
* Độ nhám ở các nhóm có phân phối chuẩn.

data\_0.02 <- subset(data, layer\_height=="0.02")  
data\_0.06 <- subset(data, layer\_height=="0.06")  
data\_0.1 <- subset(data, layer\_height=="0.1")  
data\_0.15 <- subset(data, layer\_height=="0.15")  
data\_0.2 <- subset(data, layer\_height=="0.2")

Ta tách các nhóm dữ liệu ra từng tập dữ liệu riêng sau đó dùng hàm qqnorm, qqline và shapiro.test để kiểm tra.

*data\_0.02*

qqnorm(data\_0.02$roughness, main = "Q-Q Plot data\_0.02 Roughness")  
qqline(data\_0.02$roughness)

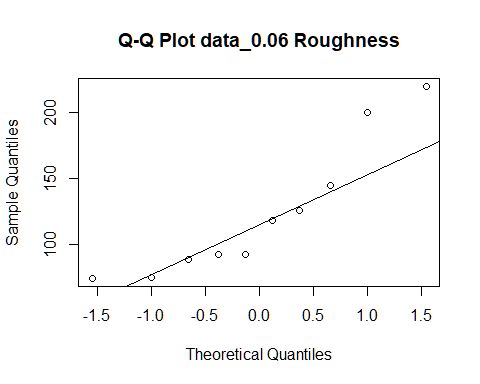


shapiro.test(data\_0.02$roughness)

##   
## Shapiro-Wilk normality test  
##   
## data: data\_0.02$roughness  
## W = 0.89205, p-value = 0.1788

*data\_0.06*

qqnorm(data\_0.06$roughness, main = "Q-Q Plot data\_0.06 Roughness")   
qqline(data\_0.06$roughness)

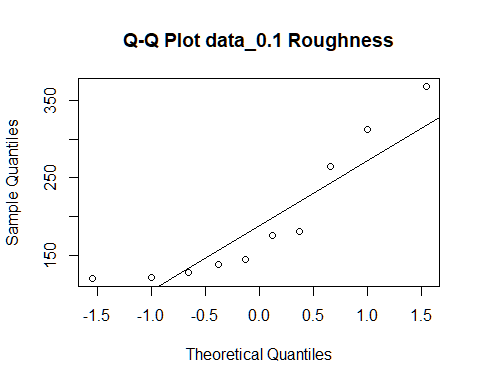


shapiro.test(data\_0.06$roughness)

##   
## Shapiro-Wilk normality test  
##   
## data: data\_0.06$roughness  
## W = 0.85434, p-value = 0.0654

*data\_0.1*

qqnorm(data\_0.1$roughness, main = "Q-Q Plot data\_0.1 Roughness")  
qqline(data\_0.1$roughness)

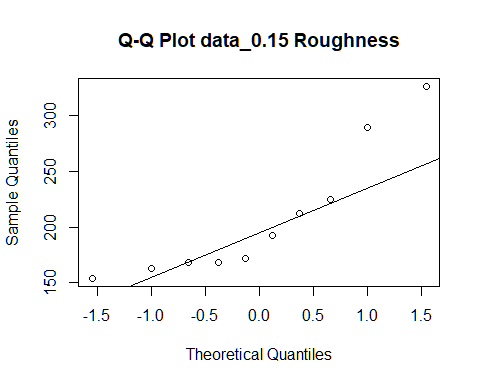


shapiro.test(data\_0.1$roughness)

##   
## Shapiro-Wilk normality test  
##   
## data: data\_0.1$roughness  
## W = 0.82279, p-value = 0.02739

*data\_0.15*

qqnorm(data\_0.15$roughness, main = "Q-Q Plot data\_0.15 Roughness")  
qqline(data\_0.15$roughness)

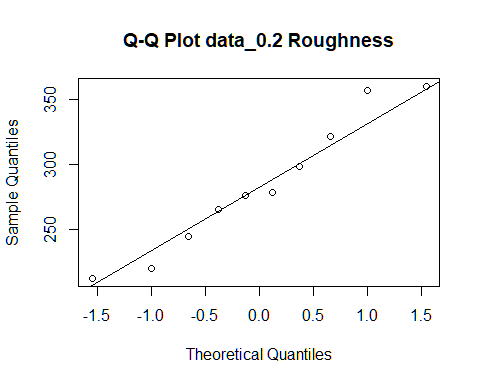


shapiro.test(data\_0.15$roughness)

##   
## Shapiro-Wilk normality test  
##   
## data: data\_0.15$roughness  
## W = 0.825, p-value = 0.02913

*data\_0.2*

qqnorm(data\_0.2$roughness, main = "Q-Q Plot data\_0.2 Roughness")  
qqline(data\_0.2$roughness)



shapiro.test(data\_0.2$roughness)

##   
## Shapiro-Wilk normality test  
##   
## data: data\_0.2$roughness  
## W = 0.9451, p-value = 0.611

Theo kết quả của code R, nhóm data có layer\_height là 0.1 và 0.15 không tuân theo phân phối chuẩn, còn các nhóm còn lại tuân theo phân phối chuẩn do có giá trị p-value lớn hơn mức ý nghĩa 5%.

Điều kiện thứ ba: Phương sai ở các tổng thể phải bằng nhau.

Ta có:

* H\_0: Phương sai độ nhám ở các nhóm bằng nhau.
* H\_1: Có ít nhất 2 nhóm có phương sai độ nhám khác nhau.

*Ta dùng hàm leveneTest để tính toán các giá trị về phương sai.*

library(car)  
leveneTest(roughness ~ layer\_height, data)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)  
## group 4 1.4956 0.2195  
## 45

Ta thấy giá trị của Pr(>F) = 0.2195 lớn hơn mức ý nghĩa 5%

⇒ Chấp nhận H\_0, phương sai độ nhám ở các nhóm bằng nhau.

**Kết luận:** Ta thấy dữ liệu thỏa điều kiện thứ nhất, điều kiện thứ ba nhưng ở điều kiện thứ hai còn một số nhóm data không thỏa mãn

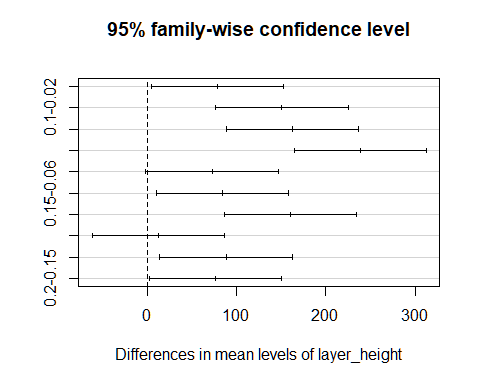
⇒ ANOVA trong dữ liệu này có thể mang tính chất tham khảo và có thể hoàn toàn không chính xác.

**3.3. Phân tích sâu sau ANOVA (so sánh bội)**

TukeyHSD(anova\_model)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = roughness ~ layer\_height, data = data)  
##   
## $layer\_height  
## diff lwr upr p adj  
## 0.06-0.02 78.3 4.053227 152.54677 0.0341507  
## 0.1-0.02 150.5 76.253227 224.74677 0.0000069  
## 0.15-0.02 162.2 87.953227 236.44677 0.0000015  
## 0.2-0.02 238.4 164.153227 312.64677 0.0000000  
## 0.1-0.06 72.2 -2.046773 146.44677 0.0602239  
## 0.15-0.06 83.9 9.653227 158.14677 0.0196501  
## 0.2-0.06 160.1 85.853227 234.34677 0.0000020  
## 0.15-0.1 11.7 -62.546773 85.94677 0.9914005  
## 0.2-0.1 87.9 13.653227 162.14677 0.0130196  
## 0.2-0.15 76.2 1.953227 150.44677 0.0416948

plot(TukeyHSD(anova\_model))



Ta có:

* H\_0: 𝜇\_i = 𝜇\_j (i≠j)
* H\_1: 𝜇\_i ≠ 𝜇\_j (i≠j)

Ta xét giá trị của p-adj nếu giá trị bé hơn mức ý nghĩa 5% thì sẽ bác bỏ H\_0 và chấp nhận H\_1

Dựa vào giá trị sau khi tính toán hàm TukeyHSD() ta có: cặp giá trị giữa 0.1 – 0.06 và 0.15 – 0.1 là lớn hơn mức ý nghĩa 5%

⇒ Ta có: 𝜇\_0.1=𝜇\_0.6; 𝜇\_0.15=𝜇\_0.1

Bên cạnh đó, dựa trên các kết quả tính được, ta có thể sắp xếp theo thứ tự giảm dần của các trung bình tổng thể như sau: 𝜇\_0.2>𝜇\_0.15=𝜇\_0.1>𝜇\_0.06>𝜇\_0.02

**Kết luận:** Muốn độ nhám thấp nhất thì layer\_height cũng phải thấp nhất là 0.02, muốn độ nhám giảm đi thì chiều cao lớp in layer\_height cũng phải giảm đi.

**4. Bài toán hồi quy tuyến tính đơn**

**Mô hình:** Hồi quy tuyến tính đa biến.

**Bài toán:** Phân tích mức độ ảnh hưởng của các thông số điều chỉnh trong máy in 3D đến độ nhám (biến roughness) của bản in như thế nào? Và dự báo thông số độ nhám của bản in dựa trên các thông số điều chỉnh trong máy in 3D cho ngẫu nhiên.

**4.1. Tính toán các giá trị, xây dựng và đánh giá mô hình**

*Đây là code đi tính hồi quy tuyến tính của các biến theo độ nhám.*

model\_1 <- lm(roughness ~ layer\_height + wall\_thickness + infill\_density + infill\_pattern +   
 nozzle\_temperature + bed\_temperature + print\_speed + material + fan\_speed, new\_data)  
summary(model\_1)

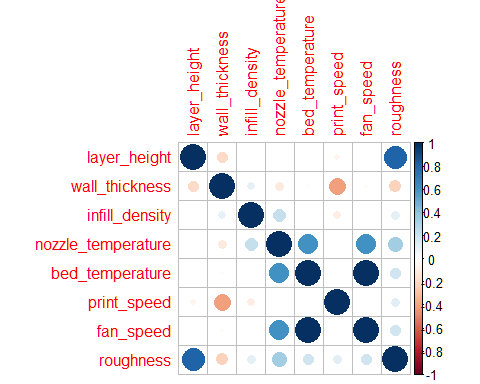
##   
## Call:  
## lm(formula = roughness ~ layer\_height + wall\_thickness + infill\_density +   
## infill\_pattern + nozzle\_temperature + bed\_temperature + print\_speed +   
## material + fan\_speed, data = new\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -72.746 -24.332 -1.641 20.304 96.552   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.371e+03 3.716e+02 -6.379 1.25e-07 \*\*\*  
## layer\_height 1.269e+03 8.765e+01 14.483 < 2e-16 \*\*\*  
## wall\_thickness 2.334e+00 2.189e+00 1.066 0.29259   
## infill\_density -4.231e-02 2.341e-01 -0.181 0.85742   
## infill\_patternhoneycomb -1.255e-01 1.128e+01 -0.011 0.99117   
## nozzle\_temperature 1.506e+01 2.529e+00 5.953 5.05e-07 \*\*\*  
## bed\_temperature -1.613e+01 3.251e+00 -4.962 1.27e-05 \*\*\*  
## print\_speed 6.496e-01 2.060e-01 3.153 0.00302 \*\*   
## materialpla 2.985e+02 5.836e+01 5.114 7.78e-06 \*\*\*  
## fan\_speed NA NA NA NA   
## ---  
## 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

*Phát hiện biến fan\_speed có hệ số NA → có đa cộng tuyến*

Kiểm tra đa cộng tuyến

*Ta sử dụng thư viện corrplot để mô hình hóa ma trận tương quan giữa các biến với nhau.*

data\_filter<-new\_data[, c("layer\_height", "wall\_thickness", "infill\_density", "nozzle\_temperature", "bed\_temperature", "print\_speed", "fan\_speed", "roughness")]  
  
cor\_data<-cor(data\_filter)  
corrplot(cor\_data)



Ta dễ dàng nhìn thấy giữa biến bed\_temperature và fan\_speed có 2 vòng tròn xanh đậm, chứng tỏ có mối quan hệ tuyến tính mạnh với nhau.

**Kết luận:** Loại bỏ biến fan\_speed.

Chạy lại mô hình hồi quy sau khi loại bỏ biến fan\_speed

model\_1 <- lm(roughness ~ layer\_height + wall\_thickness + infill\_density + infill\_pattern +   
 nozzle\_temperature + bed\_temperature + print\_speed + material , new\_data)  
summary(model\_1)

##   
## Call:  
## lm(formula = roughness ~ layer\_height + wall\_thickness + infill\_density +   
## infill\_pattern + nozzle\_temperature + bed\_temperature + print\_speed +   
## material, data = new\_data)  
##   
## 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 \*\*\*  
## layer\_height 1.269e+03 8.765e+01 14.483 < 2e-16 \*\*\*  
## wall\_thickness 2.334e+00 2.189e+00 1.066 0.29259   
## infill\_density -4.231e-02 2.341e-01 -0.181 0.85742   
## infill\_patternhoneycomb -1.255e-01 1.128e+01 -0.011 0.99117   
## nozzle\_temperature 1.506e+01 2.529e+00 5.953 5.05e-07 \*\*\*  
## bed\_temperature -1.613e+01 3.251e+00 -4.962 1.27e-05 \*\*\*  
## print\_speed 6.496e-01 2.060e-01 3.153 0.00302 \*\*   
## 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

Lúc này kết quả đã không còn hiện tượng đa cộng tuyến.

⇒ Phương trình hồi quy tuyến tính có dạng:

*Y = B\_1.layer\_height+B\_2.wall\_thickness+…*

**Đánh giá:** Phương trình tổng quát thì không tìm được chỉ có thể thông qua 1 bộ dữ liệu mẫu và ở đây là dữ liệu với 50 lần quan sát nên chỉ có thể có phương trình ước lượng mà thôi.

⇒ Phương trình hồi quy tuyến tính ước lượng có dạng:

*Y =B\_1.layer\_height+B\_2.wall\_thickness+…*

Muốn kiểm định xem các thông số của máy in 3D có ảnh hưởng như thế nào đến với độ nhám của lớp in hay không, ta đi kiểm định hệ số của từng thông số, nếu hệ số = 0 thì thông số không có ảnh hưởng đến với độ nhám của lớp in và ngược lại.

Ta có:

* H\_0: B\_i=0
* H\_1: B\_i≠0

**Ta sẽ đi so sánh:** p-value = Pr(>|t|) nếu nhỏ hơn mức ý nghĩa 5% thì sẽ bác bỏ H\_0 và chấp nhận H\_1. Khi đó thông số có giá trị p-value nhỏ hơn mức ý nghĩa 5% sẽ có ảnh hưởng đến với độ nhám của lớp in và ngược lại.

Dựa vào bảng giá trị ở trên, ta thấy có 3 thông số là: wall\_thickness, infill\_density và infill\_pattern là có p-value lớn hơn mức ý nghĩa 5% , chứng tỏ chấp nhận H0\_: B\_i=0

⇒ 3 thông số đó không có ảnh hưởng đến độ nhám của lớp in.

⇒ 3 thông số sẽ được loại bỏ khỏi mô hình.

**Ta xét đến yếu tố:**

* R^2 : Thể hiện phần trăm biến động của độ nhám lớp in được giải thích bởi biến độc lập có trong mô hình.
* R^2 có phạm vi từ 0 đến 1. Càng tiến về 1 thì chứng tỏ mô hình giải thích rất tốt đối với sự biến động của độ nhám lớp in do những biến độc lập (thông số máy in 3D) gây ra.

**Đánh giá:**  Đối với mô hình hồi quy tuyến tính đa bội thì khi đánh giá mô hình ta sẽ dựa vào thông số R^2 hiệu chỉnh (Adjusted R-squared) để đánh giá mô hình mà không sử dụng thông số R^2

**Giải thích:**  Do đối với thông số R2 thì cứ mỗi lần cung cấp 1 biến độc lập vào mô hình thì thông số R2 sẽ tăng mà không quan tâm biến độc lập đó có vi phạm điều kiện gì hay không.

⇒ Dẫn đến nếu cung cấp các biến độc lập nhưng không ảnh hưởng đến mô hình vào thì lúc này thông số R2 sẽ không còn đáng tin cậy nữa.

Vì vậy đối với mô hình hồi quy tuyến tính đa bội, khi đánh giá mô hình sẽ dựa vào thông số R^2 hiệu chỉnh để đánh giá vì thông số R^2 hiệu chỉnh sẽ cân bằng lại so với việc khi đưa nhiều biến độc lập không ảnh hưởng tới mô hình.

*Ta bắt đầu xây dựng mô hình sau khi đã loại bỏ 3 biến không ảnh hưởng đến mô hình.*

model\_2 <- lm(roughness ~ layer\_height + nozzle\_temperature + bed\_temperature + print\_speed + material, data = new\_data)  
summary(model\_2)

##   
## Call:  
## lm(formula = roughness ~ layer\_height + nozzle\_temperature +   
## bed\_temperature + print\_speed + material, data = new\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -74.084 -26.500 -1.662 22.585 92.356   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2310.7356 353.2009 -6.542 5.38e-08 \*\*\*  
## layer\_height 1246.5353 83.1780 14.986 < 2e-16 \*\*\*  
## nozzle\_temperature 14.7774 2.3979 6.163 1.95e-07 \*\*\*  
## bed\_temperature -15.8078 3.0895 -5.117 6.55e-06 \*\*\*  
## print\_speed 0.5538 0.1804 3.070 0.00366 \*\*   
## materialpla 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

Lúc này ta thấy giá trị của R2 hiệu chỉnh đã tăng hơn so với ban đầu và lớn hơn 0.8 chứng tỏ mô hình đang khá tốt.

*Ta sẽ dự đoán mô hình dựa trên những giá trị ngẫu nhiên của các biến độc lập (các thông số của máy in 3D)*

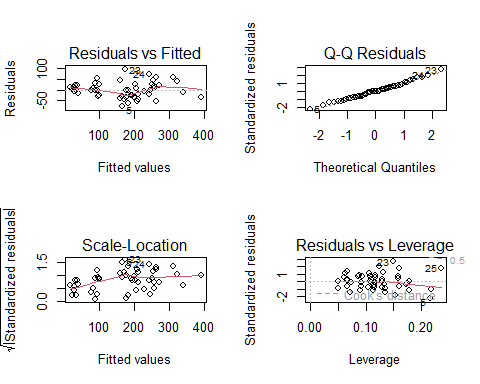
data\_test <- data.frame(layer\_height = 0.03, nozzle\_temperature = 250, bed\_temperature = 80, print\_speed = 80, material = "abs")  
data\_test$predicted\_value <- predict(model\_2, newdata = data\_test, interval = "confidence", level = 0.95)  
print(data\_test)

## layer\_height nozzle\_temperature bed\_temperature print\_speed material  
## 1 0.03 250 80 80 abs  
## predicted\_value.fit predicted\_value.lwr predicted\_value.upr  
## 1 200.7033 167.2929 234.1136

**4.2. Đánh giá các điều kiện của mô hình hồi quy**

*Ta sẽ dùng hàm plot() để đánh giá các điều kiện của mô hình hiện tại.*

par(mfrow = c(2, 2))  
plot(model\_2)



**Điều kiện thứ nhất: Sai số phải tuân theo phân phối chuẩn.**

Ta sẽ dựa vào đồ thị Q-Q Residuals để đánh giá.

Ta thấy các điểm trên hình tập trung gần với đường thẳng phân phối chuẩn

⇒ Sai số tuân theo phân phối chuẩn.

⇒ Thỏa mãn điều kiện thứ nhất.

**Điều kiện thứ hai: Sai số có kì vọng bằng 0.**

Ta sẽ dựa vào đồ thị Residuals vs Fitted để đánh giá.

Nếu đường màu đỏ nằm gần đường bằng 0 thì có thể nói sai số có kì vọng bằng 0.

Ở đây ta thấy đường màu đỏ không nằm sát đường bằng 0

⇒ Mô hình có sai số có kì vọng khác 0.

⇒ Vi phạm điều kiện thứ hai.

**Điều kiện thứ ba: Phương sai của các sai số là hằng số (sai số đồng nhất)**

Nếu những điểm này nó phân bố ngẫu nhiên dọc theo đường màu đỏ

⇒ Ta nói phương sai của các sai số là hằng số

Nhưng khi nhìn vào đồ thị của hình thì ta thấy nó chỉ phân bố tập trung nhiều ở đoạn đầu đường màu đỏ và khu vực giữa đường đỏ

⇒ Vi phạm điều kiện thứ ba.

**Điều kiện thứ tư: Không có hiện tượng đa cộng tuyến xảy ra.**

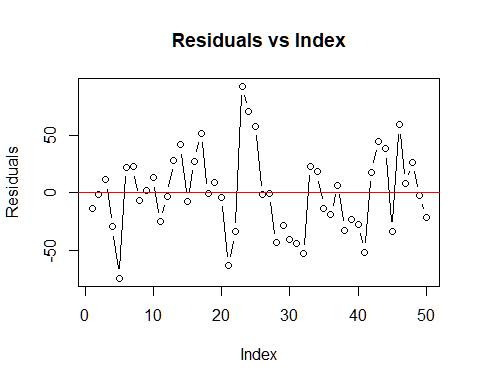
Ở đây model\_2 đã được loại bỏ các biến gây nên hiện tượng đa cộng tuyến

⇒ Thỏa mãn điều kiện thứ tư.

**Điều kiện thứ năm: Các sai số độc lập với nhau.**

Ta sử dụng biểu đồ Residuals vs Index.

plot(residuals(model\_2), type = "b", main = "Residuals vs Index",  
 xlab = "Index", ylab = "Residuals")  
abline(h = 0, col = "red")



Dựa trên mô hình kết quả, ta thấy các điểm phân bố theo ngẫu nhiên mà không có một quy tắc rõ ràng.

⇒ Các sai số độc lập với nhau.

⇒ Thỏa mãn điều kiện thứ năm.

## VI. Thảo luận và mở rộng.

Từ việc phân tích có thể thấy các yếu tố ảnh hưởng đến độ nhám là: layer\_height, nozzle\_temperature, bed\_temperature, material, cuối cùng là print\_speed.

R^2 hiệu chỉnh = 0.8571 cho thấy mô hình ta xây dựng tương đối phù hợp trong việc thực hiện dự báo.

Ưu điểm của phương pháp hồi quy chính là một phương pháp thống kê để thiết lập mối quan hệ giữa một biến phụ thuộc và một nhóm tập hợp các biến độc lập. Mô hình với một biến phụ thuộc với hai hoặc nhiều biến độc lập được gọi là hồi quy bội (hay còn gọi là hồi quy đa biến).

Hồi quy tuyến tính bội dễ hiểu và dễ triển khai, đặc biệt khi các biến có mối quan hệ tuyến tính và hiệu quả tính toán của phương pháp này có thể được tính toán nhanh chóng.

Dự đoán chính xác các biến độc lập có mối quan hệ tuyến tính mạnh với biến phụ thuộc, phương pháp hồi quy tuyến tính bội có thể cung cấp các dự đoán chính xác, các hệ số hồi quy tuyến tính bội cung cấp thông tin về mối quan hệ giữa từng biến độc lập và biến phụ thuộc, giúp hiểu rõ hơn về ảnh hưởng của từng yếu tố. và tiện lợi cho kiểm định giả thuyết, dễ dàng thực hiện các kiểm định giả thuyết về các hệ số hồi quy để kiểm tra sự ảnh hưởng của các biến độc lập ngay cả với dữ liệu lớn, nhờ các thuật toán tối ưu hóa hiệu quả.

Tuy nhiên mô hình cũng có một số nhược điểm như: Phương pháp này giả định rằng mối quan hệ giữa các biến độc lập và biến phụ thuộc là tuyến tính, điều này không phải lúc nào cũng đúng trong thực tế, hồi quy tuyến tính bội rất nhạy cảm với các điểm dữ liệu ngoại lệ, có thể làm sai lệch mô hình.

Khi các biến độc lập có tương quan cao với nhau, nó có thể gây ra đa cộng tuyến, làm cho các ước lượng hệ số hồi quy không ổn định và khó diễn giải. Phương pháp này giả định rằng các sai số có phân phối chuẩn và có phương sai không đổi. Nếu các giả định này không được thỏa mãn, kết quả hồi quy có thể không tin cậy.

Không phù hợp cho các mối quan hệ phi tuyến tính: Đối với các mối quan hệ phi tuyến tính giữa các biến, hồi quy tuyến tính bội không phải là lựa chọn tốt nhất. Các phương pháp khác như hồi quy phi tuyến hoặc mô hình cây quyết định có thể phù hợp hơn.

Ngoài ra, đề tài ta có thể phân tích xây dựng bổ sung hai mô hình hồi quy của hai biến roughness và elongation dựa trên cách thức giống như xây dựng mô hình ứng với biến tension\_strenght.

Từ những phương pháp này mà đề tài đã đánh giá được hiệu quả những yếu tố của biến ảnh hướng đến các sức căng bề in 3D theo như bài toán đã được đặt ra.