-Tách dữ liệu của 2 nhóm cần khảo sát:

+ Nhóm 1: Giá cả đề xuất của các Lithography có kích thước > 32 mm

+ Nhóm 2: Giá cả đề xuất của các Lithography có kích thước <= 32 mm

group1 <- CPU\_data$Recommended\_Customer\_Price[CPU\_data$Lithography > 32]

group2 <- CPU\_data$Recommended\_Customer\_Price[CPU\_data$Lithography <= 32]

-Hình dạng biểu đồ của 2 nhóm sau khi tách:

hist(group1, main = "Histogram of Group 1 (Lithography > 32)",

xlab = "Recommended Customer Price", col = "orange")

print(group2)

hist(group2, main = "Histogram of Group 2 (Lithography <= 32)",

xlab = "Recommended Customer Price", col = "orange")

print(group2)

-Tính toán trung bình độ lệch chuẩn và số lượng cho group1

mean\_group1 <- mean(group1, na.rm = TRUE)

sd\_group1 <- sd(group1, na.rm = TRUE)

n1 <- length(group1)

-Tính toán trung bình độ lệch chuẩn và số lượng cho group2

mean\_group2 <- mean(group2, na.rm = TRUE)

sd\_group2 <- sd(group2, na.rm = TRUE)

n2 <- length(group2)

- In kết quả

cat("Group 1: Mean =", mean\_group1, ", SD =", sd\_group1, ", n =", n1, "\n")

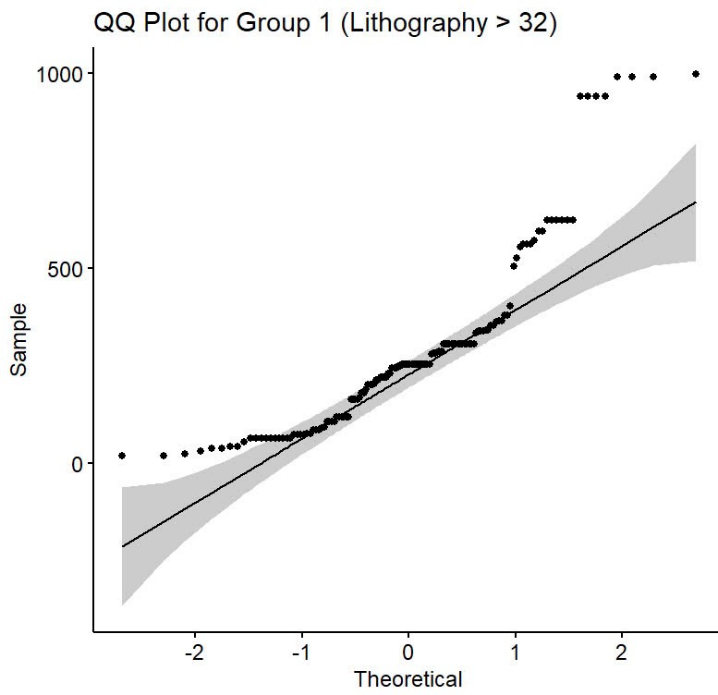
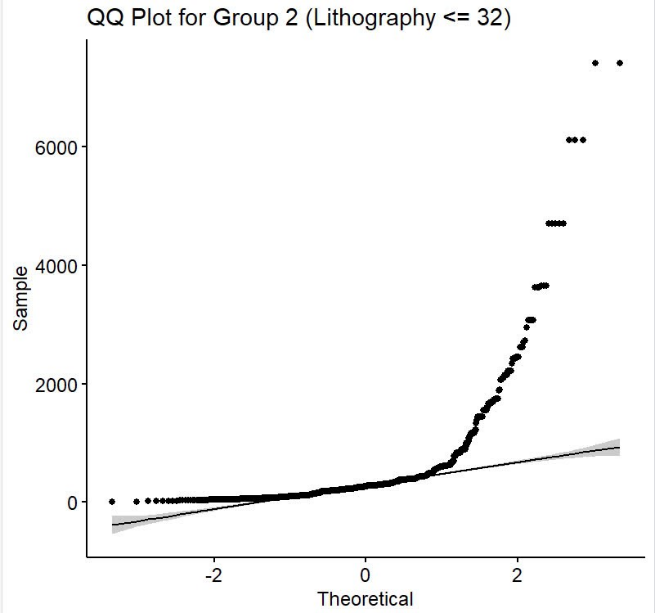
cat("Group 2: Mean =", mean\_group2, ", SD =", sd\_group2, ", n =", n2, "\n")

-Để biết được mỗi mẫu phân bố theo phân bố j ta sẽ khảo sát bằng cách vẽ sơ đồ QQ-plot:

+ Vẽ QQ-plot cho group1

qqplot1 <- ggqqplot(group1, main = "QQ Plot for Group 1 (Lithography > 32)")

print(qqplot1)

* VÌ số lượng điểm khảo sát nằm lệch với đường chéo của phân phối chuẩn nhiều nên mẫu này phân phối tùy ý.
* Tương tự mẫu trên cũng phân phối tùy ý.

-Dựa vào khảo sát ta xác định được bài toán thuộc vào dạng 4d: 2 mẫu độc lập, phân phối tùy ý.

-Để kiểm tra sự khác biệt về **giá thành đề xuất (Recommended\_Customer\_Price)** giữa hai nhóm **CPU có Lithography > 32** và **<= 32** ta sẽ sử dụng kiểm định t cho 2 mẫu độc lập với cú pháp:

t.test(x, y, var.equal = FALSE)

+ **x và y**: Hai nhóm dữ liệu cần so sánh (trong trường hợp này là Recommended\_Customer\_Price cho hai nhóm CPU với **Lithography > 32** và **<= 32**).

+ **var.equal**: Nếu bạn giả định rằng phương sai của hai nhóm là **bằng nhau**, đặt var.equal = TRUE. Nếu không, đặt var.equal = FALSE.

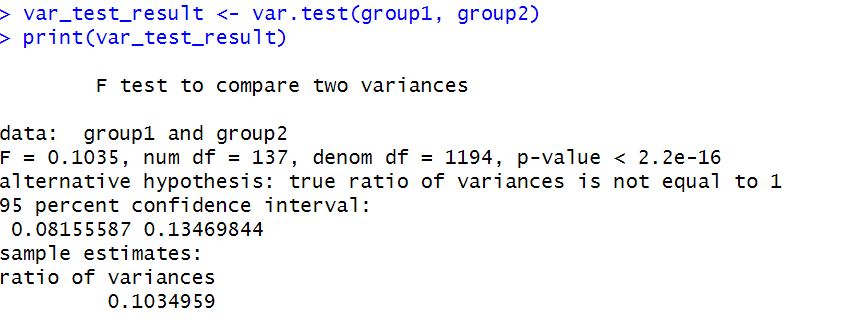
-Dựa vào cú pháp trên, ta cần khảo sát được 2 mẫu được xét có phương sai khác nhau hay giống nhau, ta sẽ sử dụng f-test để kiểm tra:

# Thực hiện kiểm định F-test

var\_test\_result <- var.test(group1, group2)

# In kết quả

print(var\_test\_result)

* Ta thấy giá trị của p-value < 2.2e-16, P-value cực kỳ nhỏ (nhỏ hơn 0.05), có nghĩa là bạn **bác bỏ giả thuyết không** (H₀) và kết luận rằng **phương sai của hai nhóm là khác nhau**.

+ **F = 0.1035**: Tỷ lệ giữa phương sai của hai nhóm.

+**95 percent confidence interval**: Khoảng tin cậy 95% cho tỷ lệ phương sai giữa hai nhóm là từ **0.08155587** đến **0.13469844**. Vì khoảng tin cậy này không bao gồm 1, bạn có thể kết luận rằng phương sai của hai nhóm là **khác nhau**.

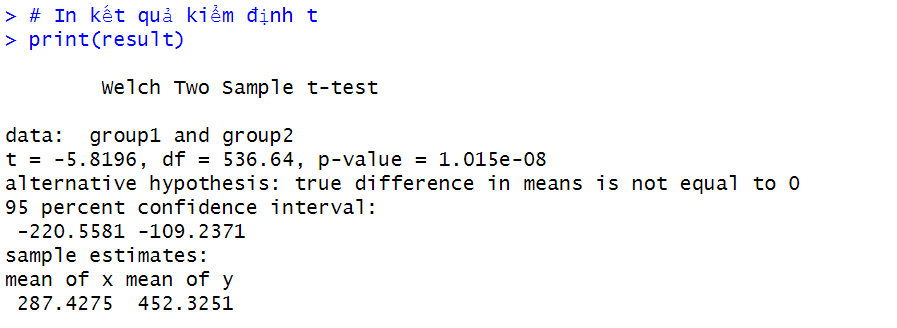
-Với giá trị 2 phương sai là khác nhau ta sẽ khảo sát bài toán với cú pháp:

# Kiểm định t cho hai mẫu độc lập với phương sai khác nhau

result <- t.test(group1, group2, var.equal = FALSE)

# In kết quả kiểm định t

print(result)

#-> Vi p < 0.05 nên Ho bị bác bỏ, dẫn đến giá thành trung bình của 2 nhóm kích thước là khác nhau.