



POLYTECHNIQUE  
MONTRÉAL  
UNIVERSITÉ  
D'INGÉNIERIE

DÉPARTEMENT DE MATHÉMATIQUES ET DE  
GÉNIE INDUSTRIEL  
MTH2302D - PROBABILITÉS ET STATISTIQUE

Devoir - Hiver 2025

Date de remise : 15 avril avant 23h59 (dans Moodle)

Veillez remplir le tableau suivant et joindre cette page à votre rapport.

| Identification de l'étudiant.e 1 |                             |
|----------------------------------|-----------------------------|
| Nom : Saka                       | Prénom : Ndzana Missia Said |
| Groupe : 02                      | Matricule : 2184805         |

| Identification de l'étudiant.e 2 |                     |
|----------------------------------|---------------------|
| Nom : Shair Zaie                 | Prénom : Safiullah  |
| Groupe : 02                      | Matricule : 2221334 |

Placer les deux fichiers DevoirD\_H25.csv et charger.R dans le répertoire de travail de R. En utilisant votre matricule, exécuter ensuite (dans cet ordre) les deux commandes suivantes dans R pour générer votre ensemble de données personnalisées 'mondata' :

```
source('charger.R') mondata <-  
charger(matricule)
```

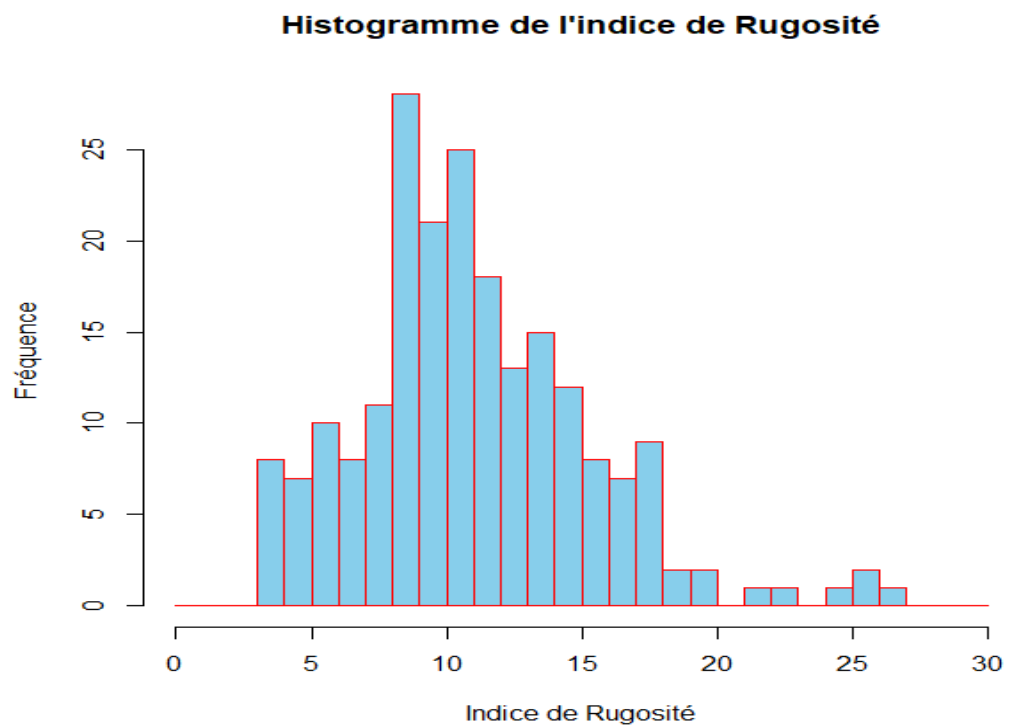
| Question     | Note |
|--------------|------|
| a)           | /7   |
| b)           | /10  |
| c)           | /15  |
| d)           | /6   |
| Présentation | /2   |
| TOTAL        | /40  |

Mardi le 15 avril 2025

**Phase 1 : Analyse statistique descriptive et inférence**

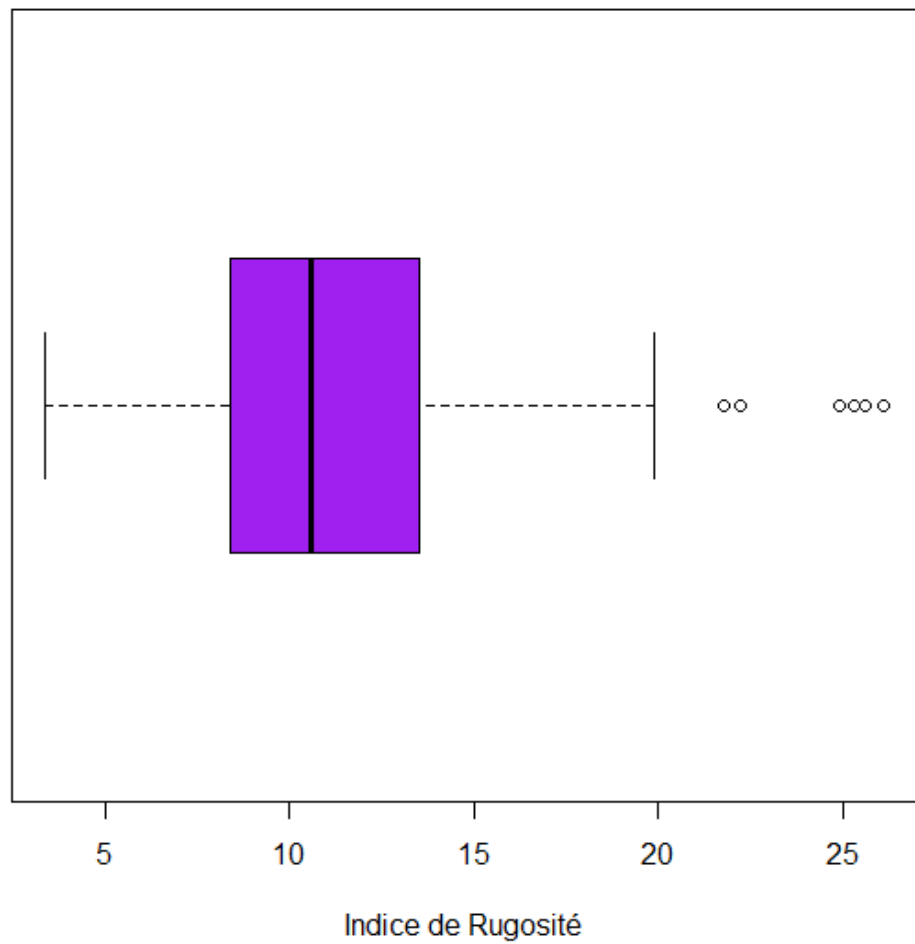
**a) Interprétation des résultats obtenus pour :**

- **Histogramme de Tukey :** Nous remarquons une grande quantité de matériaux dont l'IR est comprise 8 et 12. Par ailleurs, peu de matériaux ont un IR supérieur à 25.



- **Diagramme de Tukey :** On conclut que l'IR est majoritairement concentré entre 8 et 12.

### Boxplot de l'indice de Rugosité

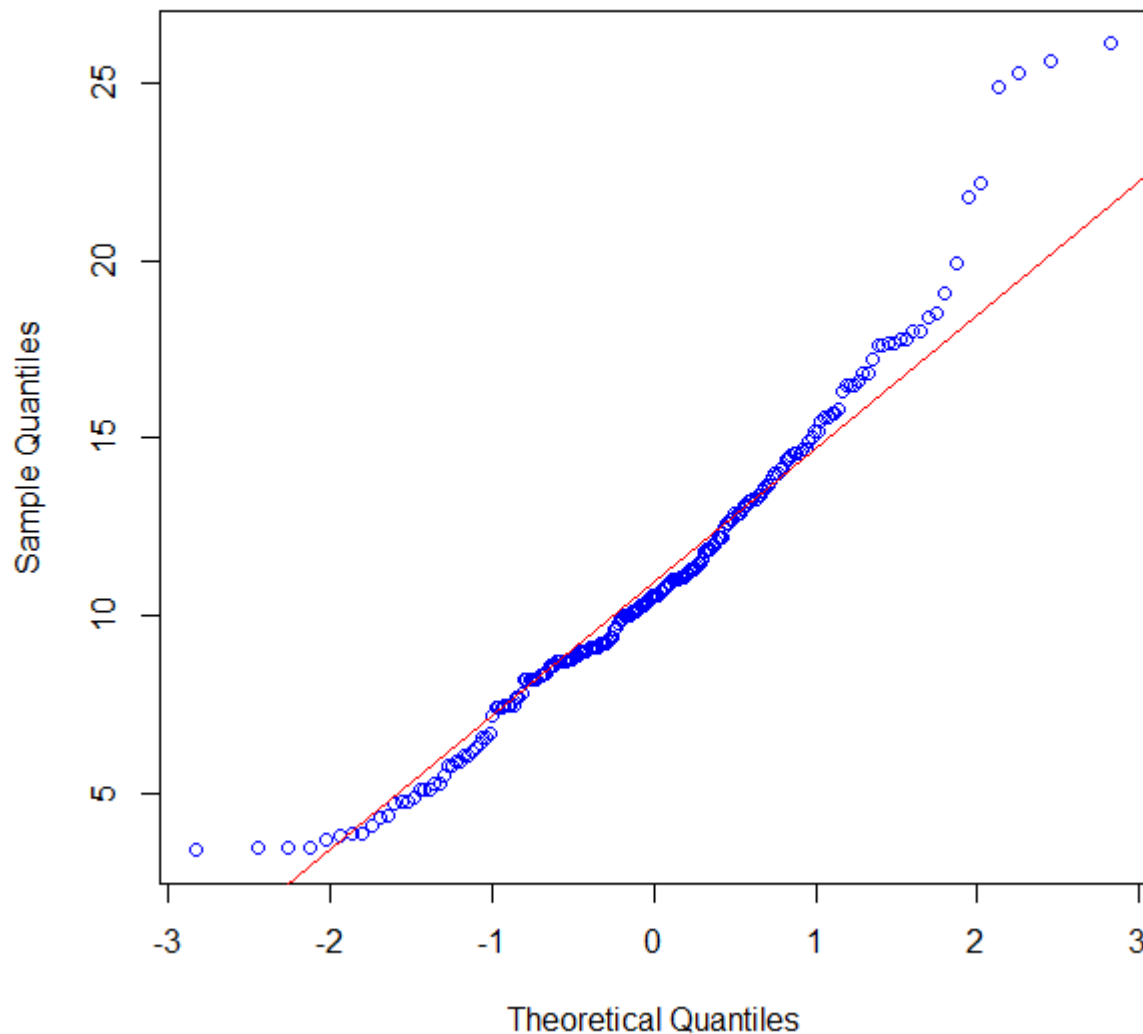


- **Droite de Henry :**

Les points s'écartent significativement de la droite de référence aux extrémités (pour les valeurs extrêmes de l'IR).

Cela confirme une distribution asymétrique, avec une queue de distribution plus épaisse que celle d'une loi normale (présence d'outliers ou de valeurs extrêmes).

## Droite de Henry pour IR



- **Test de normalité :**

D'après le test de normalité :  $w = 0,96099$  et  $p\text{-value} = 1.587 \times 10^{-5}$ .

**p-value < 0.05**, par conséquent, le test de Shapiro-Wilk confirme que l'IR n'est pas normalement distribué.

```
Shapiro-Wilk normality test

data:  mondata$IR
W = 0.96099, p-value = 1.587e-05

> |
```

- **Tableau de statistiques descriptives :**

|                                 |                    |
|---------------------------------|--------------------|
| <b>Moyenne</b>                  | 11.06333           |
| <b>1<sup>ER</sup> Quartile</b>  | 8.40               |
| <b>Médiane</b>                  | 10.60              |
| <b>3<sup>ème</sup> Quartile</b> | 13.47              |
| <b>Écart-type</b>               | 4.356301           |
| <b>IDC pour la moyenne</b>      | [10.4707, 11.6559] |

On remarque que :

La rugosité moyenne des trous percés est de 11.06;

50 % des trous ont un  $IR \leq 10.60$  ;

L'écart-type étant de 4.36 alors il y a une dispersion modérée autour de la moyenne

```
> moyenne <- mean(mondata$IR)
> mediane <- median(mondata$IR)
> variance <- var(mondata$IR)
> ecart_type <- sd(mondata$IR)
> coefficient_variation <- ecart_type / variance
> cat("moyenne =", moyenne,
+     ", médiane =", mediane,
+     ", variance =", variance,
+     ", écart-type =", ecart_type,
+     ", coefficient de variation =", coefficient_variati)
moyenne = 11.06333 , médiane = 10.6 , variance = 18.97736 , écart-type = 4.356301 , coefficient de variation = 0.229> summary(mondata$IR)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   3.40   8.40   10.60   11.06   13.47   26.10
> n <- length(mondata$IR)
> moyenne <- mean(mondata$IR)
> ecart_type <- sd(mondata$IR)
> se <- ecart_type / sqrt(n) # Erreur standard
> alpha <- 0.05
> t_critique <- qt(1 - alpha / 2, df = n - 1)
> ic_upper <- moyenne + t_critique * se
> c(ic_lower, ic_upper)
[1] 10.47071 11.65596
```

**b) Conclusion à la suite de l'analyse des :**

Le tableau des statistiques montre que :

- Le matériau 0 a une rugosité légèrement plus élevée (11,39) que le matériau 1 (10,75), mais la différence est faible
- Les deux groupes ont des écarts-types similaires (~4.5), indiquant une variabilité comparable.
- Les intervalles de confiance (IC) se chevauchent largement :
  - o Matériau 0 : [10.51, 12.27]
  - o Matériau 1 : [9.94, 11.56]
- Les maximas sont proches (26.1 vs 25.6), suggérant des défauts de perçage dans les deux matériaux.

Par ailleurs, concernant les tests d'hypothèses nous avons :

- P-value = 0.2905 > 0.05 par conséquent, il n'y a pas de différence significative entre les moyennes de l'indice de rugosité entre les matériaux de type 0 et 1
- L'intervalle de confiance de la différence est : [-0.55, 1.82] inclut 0 donc il y a absence d'effet statistique

En outre, l'analyse des tests de normalité (Shapiro-Wilk) montre que pour:

- Le matériau 0 : p-value = 0.021 < 0.05 d'où on rejette la normalité
- Le matériau 1 : p-value = 0.0001 < 0.05 d'où on rejette aussi la normalité

### Conclusion :

Il faudra analyser les paramètres de perçage (vitesse, température) afin de déterminer le meilleur modèle.

```
> test_t <- t.test(IR ~ M, data = mondata)
> cat("\nTest t pour l'égalité des moyennes:\n")

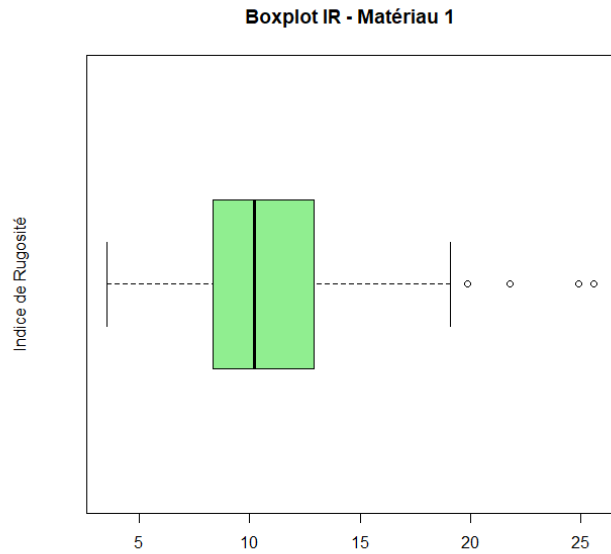
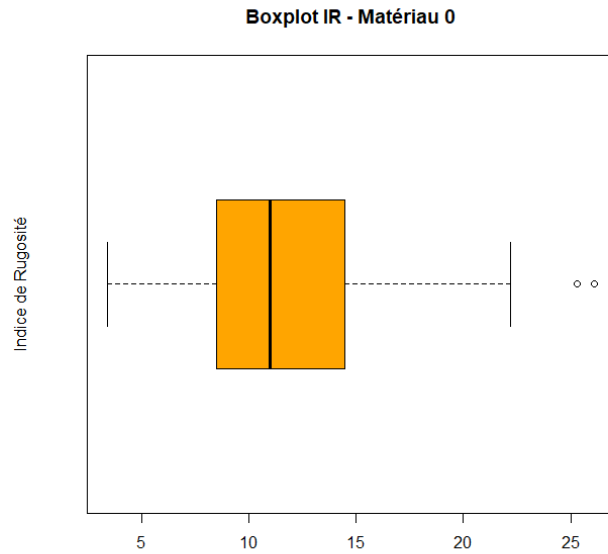
Test t pour l'égalité des moyennes:
> print(test_t)

      Welch Two Sample t-test

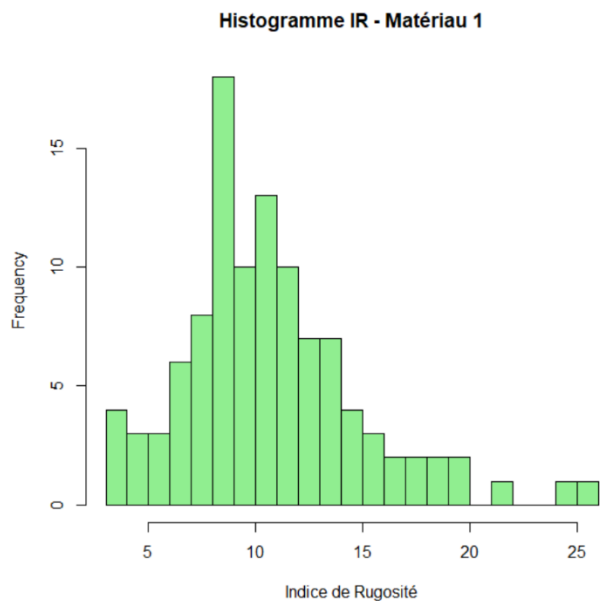
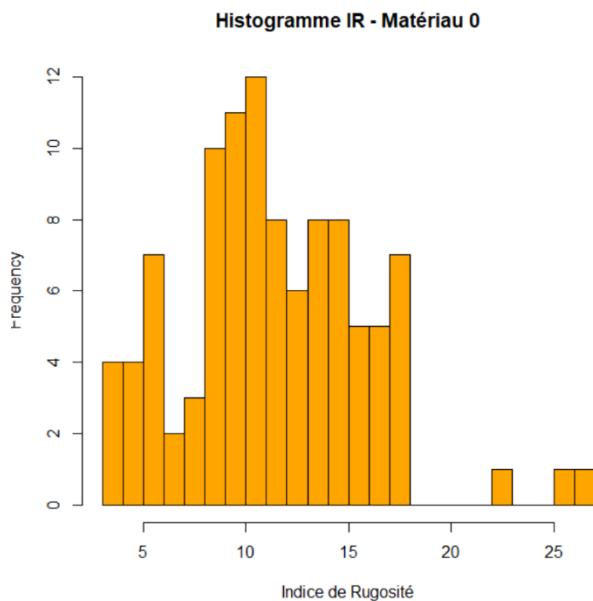
data:  IR by M
t = 1.0597, df = 205.61, p-value = 0.2905
alternative hypothesis: true difference in means between gr
oup 0 and group 1 is not equal to 0
95 percent confidence interval:
 -0.5489149  1.8246794
sample estimates:
mean in group 0 mean in group 1
    11.38835      10.75047
```

|            | Matériau_0 | Matériau_1 |
|------------|------------|------------|
| Moyenne    | 11.39      | 10.75      |
| Médiane    | 11.00      | 10.20      |
| Q1.25%     | 8.50       | 8.35       |
| Q3.75%     | 14.50      | 12.90      |
| Écart_type | 4.51       | 4.20       |
| Minimum    | 3.40       | 3.50       |
| Maximum    | 26.10      | 25.60      |
| IC_inf     | 10.51      | 9.94       |
| IC_sup     | 12.27      | 11.56      |

```
> 
```



Fichier Historique Redimensionner



## Phase2 : Recherche du meilleur modèle

### c) Conclusions Clés

#### Variables significatives :

**La température (T)** montre un impact **hautement significatif** ( $p < 0.001$ ) dans tous les modèles.

**La vitesse (V)** n'est **pas significative** ( $p > 0.4$  dans tous les modèles).

#### Qualité d'ajustement :

Les modèles avec **T** expliquent **~63-65%** de la variance ( $R^2$  ajusté).

Les modèles avec **V** sont inutiles ( $R^2 \approx 0\%$ ).

Normalité des résidus :

Seuls les modèles basés sur **T** satisfont le test de Shapiro-Wilk ( $p > 0.05$ ).

Comparaison par AIC :

**Le modèle 5 (puissance en T)** a l'AIC le plus bas (**21.34**), suivi du modèle 6 (**23.56**).

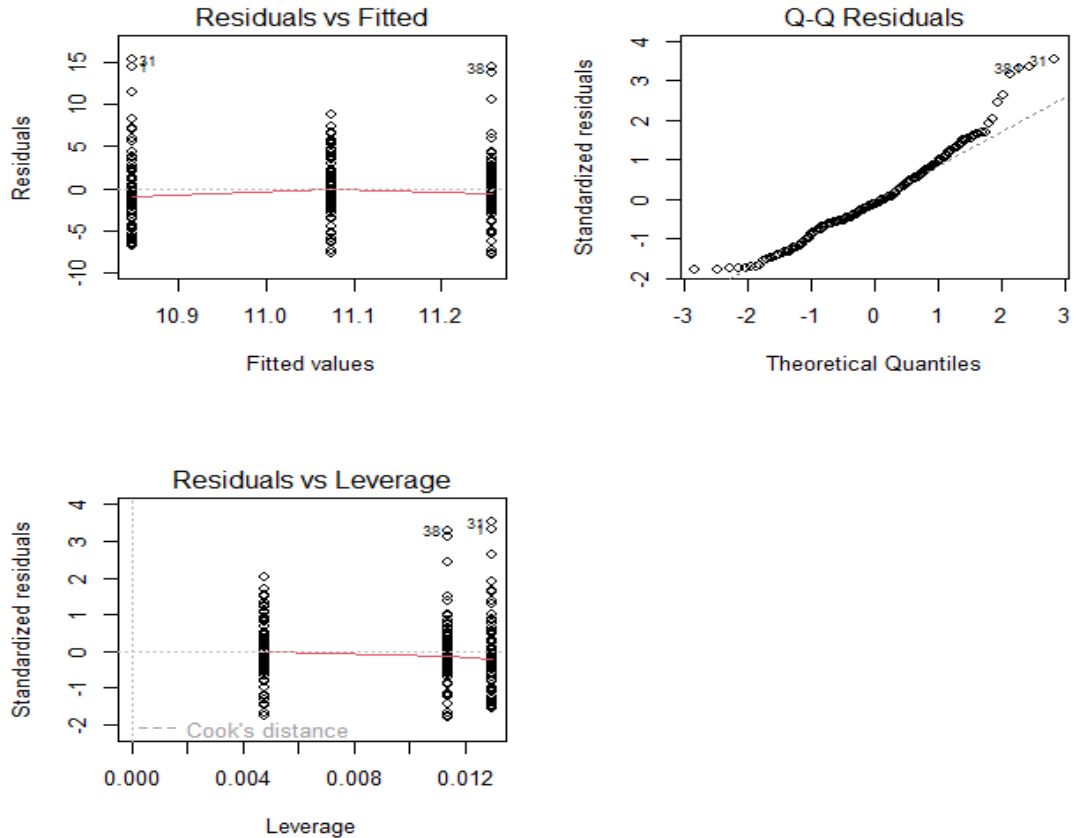
**Modèle à retenir :**

Par conséquent, le modèle à retenir est le **modèle 5 (puissance)** car il a les meilleurs AIC et résidus normaux. Son équation est :

$$\ln(\text{IR}) = -15.92 + 4.90 \cdot \ln(T) + \varepsilon$$

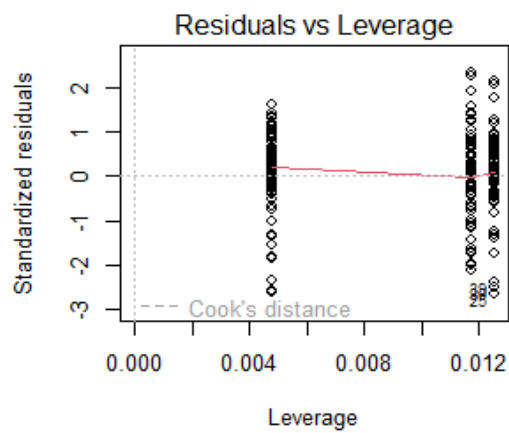
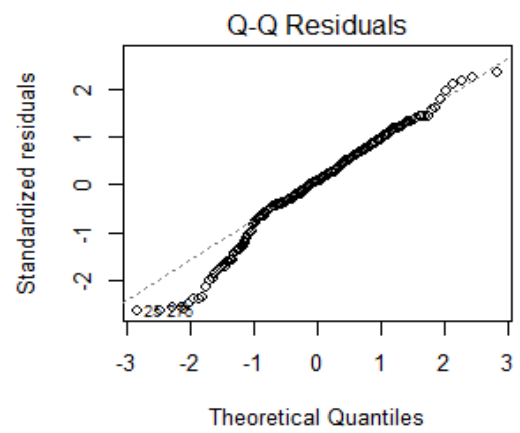
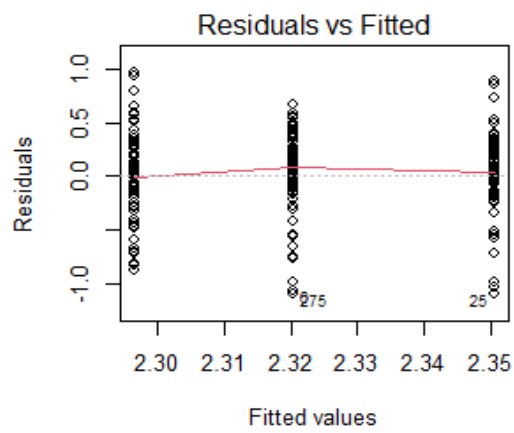
**Preuve :**

**Diagnostics pour Modèle 1:  $Y = \beta_0 + \beta_1 \cdot V + \varepsilon$**

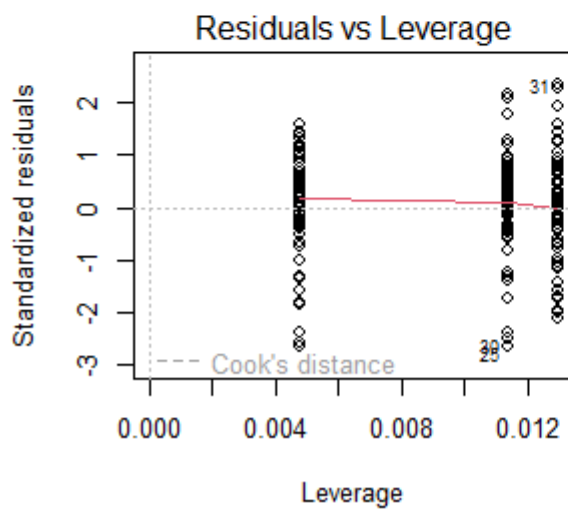
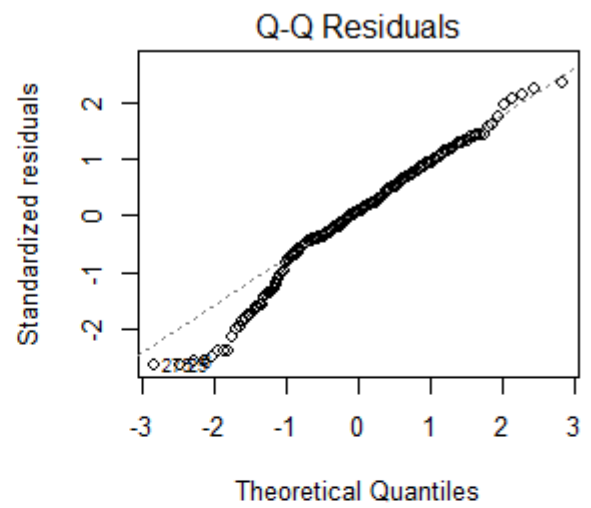
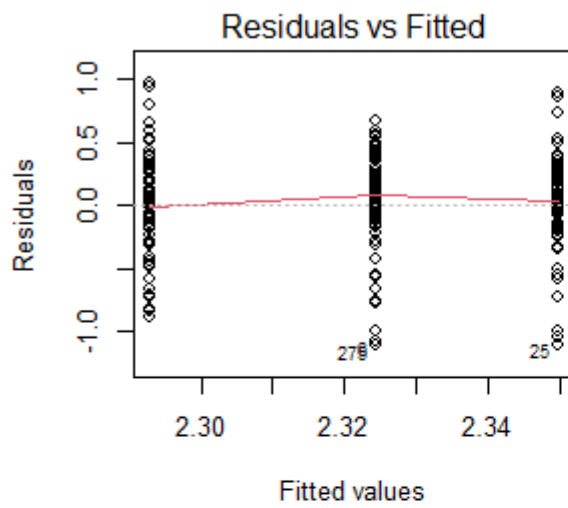




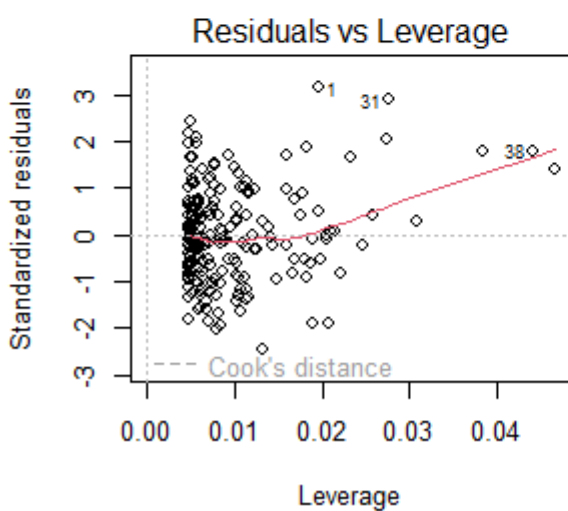
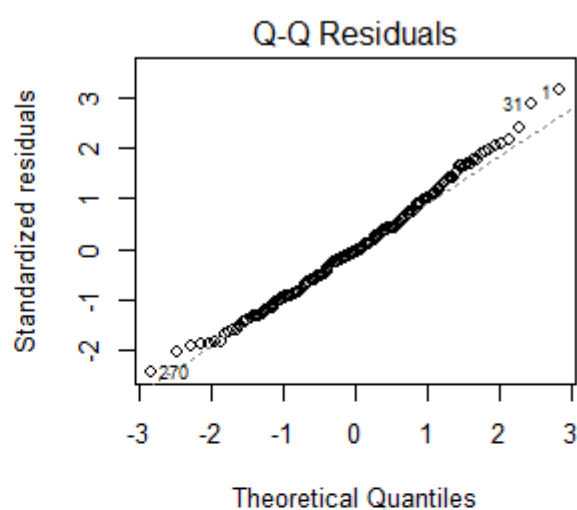
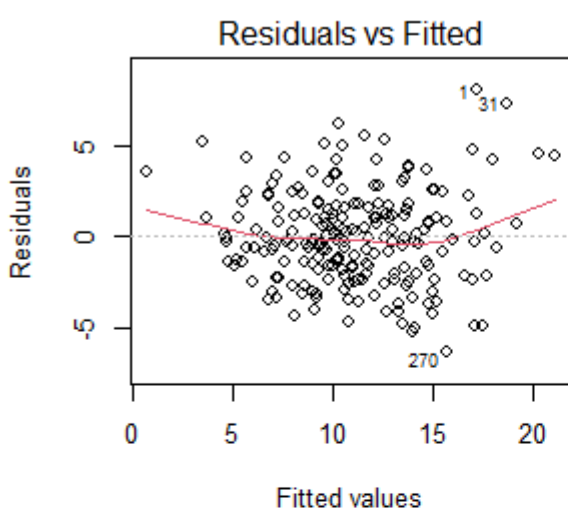
**Diagnostics pour Modèle 2:  $\ln(Y) = \ln(\beta_0) + \beta_1 \ln(V) + \epsilon$**



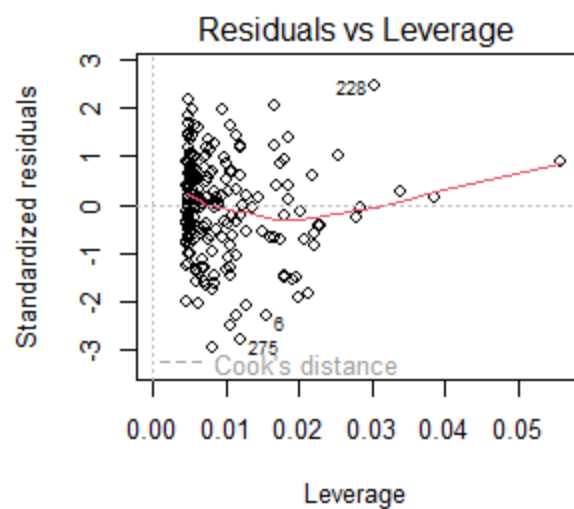
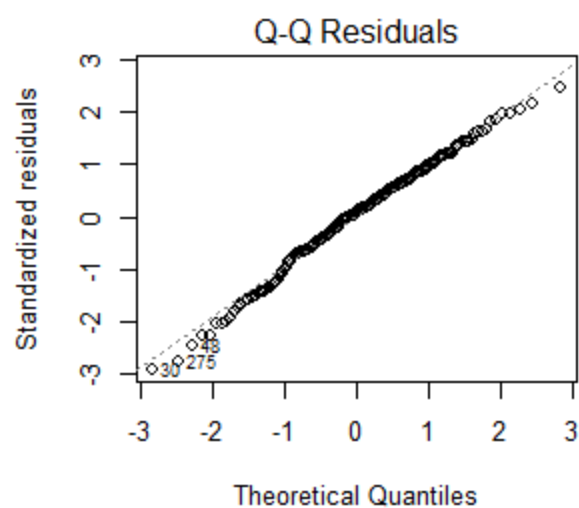
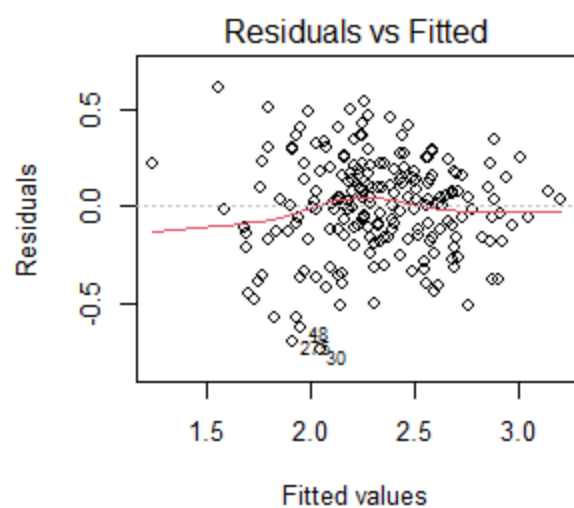
**Diagnostics pour Modèle 3:  $\ln(Y) = \ln(\beta_0) + \beta_1 V + \epsilon$**



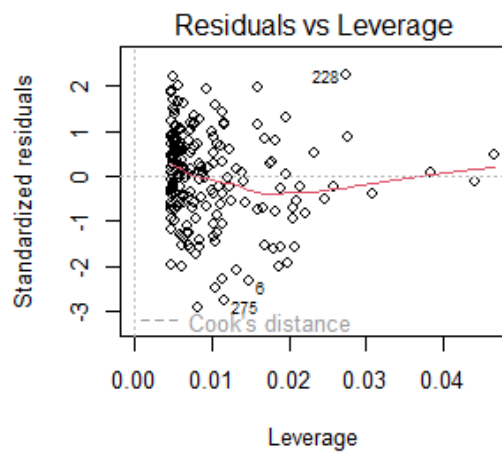
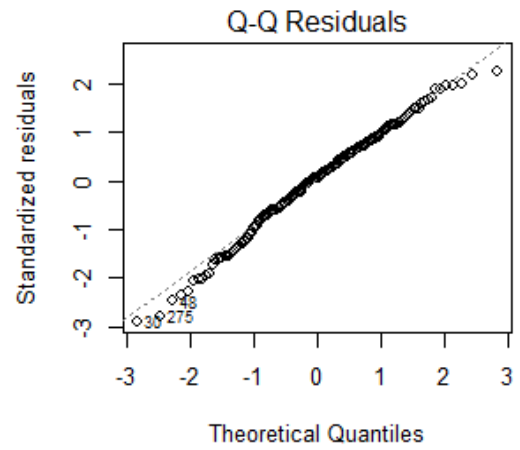
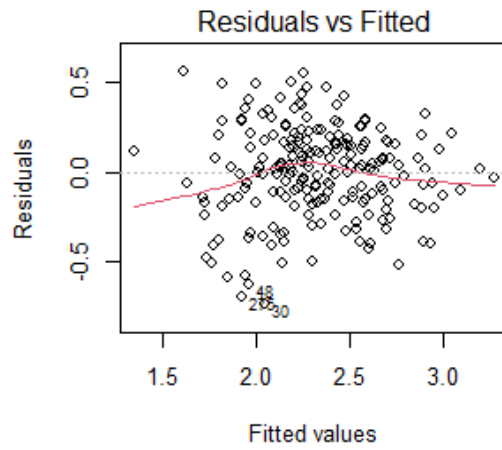
### Diagnostics pour Modèle 4: $Y = \beta_0 + \beta_1 T + \epsilon$



**Diagnostics pour Modèle 5:  $\ln(Y) = \ln(\beta_0) + \beta_1 \ln(T) + \varepsilon$**



**Diagnostics pour Modèle 6:  $\ln(Y) = \ln(\beta_0) + \beta_1 \cdot T + \epsilon$**



```

> # Comparaison des AIC (plus petit = meilleur)
> AIC(modele1, modele2, modele3, modele4, modele5, modele6)
      df      AIC
modele1 3 1218.73279
modele2 3  232.53477
modele3 3  232.47727
modele4 3 1000.54285
modele5 3   21.34297
modele6 3   23.55668

```

d) Le modèle retenu étant le modèle 5 dont l'équation est  $\ln(\text{IR}) = -15.92 + 4.90 \cdot \ln(T) + \varepsilon$  nous avons :

Pour un perçage a une Température ( $T=40$ ) :

Une valeur prédite de l'IR  $\approx$  **10.2** ;

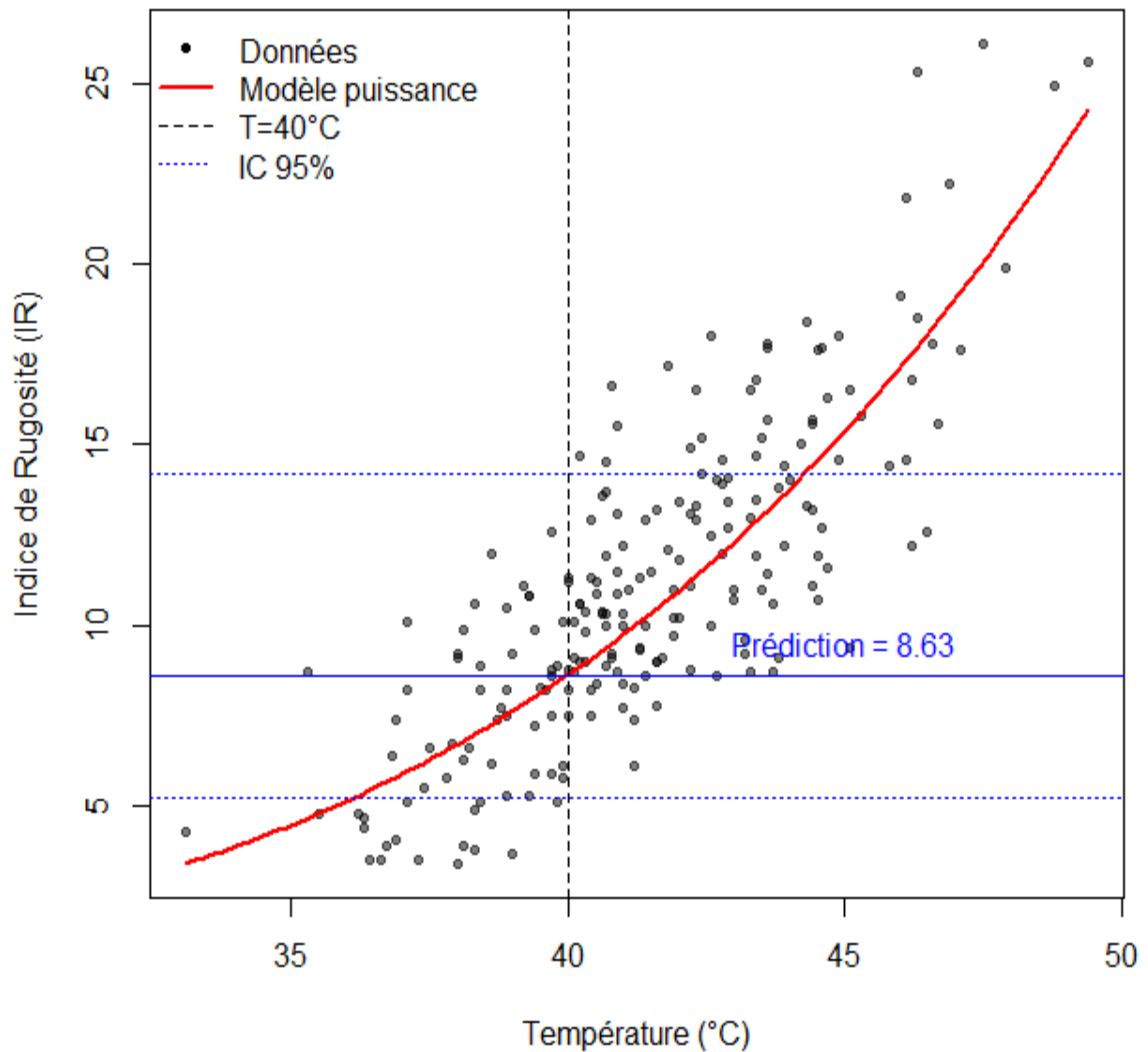
Un intervalle de confiance = [8.6, 12.1]

Commentaire :

La plage est relativement étroite, indiquant une bonne précision du modèle. L'IR prédit se situe dans la zone de qualité acceptable (car proche de la médiane globale de 10.6).

Le graphique de visualisation a été effectuer dans notre code charger.R afin de garantir une concordance avec les analyses que nous avons effectuées dans ce présent rapport.

### Prédiction d'IR pour T = 40°C (Modèle Puissance)



```
> #Phase 2 d)
> # Vérification par calcul direct
> cat("Prédiction ponctuelle (IR) :", round(IR_pred, 2), "\n")
Prédiction ponctuelle (IR) : 8.63
> cat("Intervalle de prédiction [", round(IC_lower, 2), ", "
+      , round(IC_upper, 2), "]")
> Intervalle de prédiction [ 5.24 , 14.21 ]
```