

Modèle linéaire

Des procédés de hautes températures sont souvent utilisés pour traiter des pièces métalliques telles que les engrenages, en modifiant leurs caractéristiques de surface. La profondeur de trempe est une propriété importante de la pièce, et contribue à sa fiabilité. Des essais destructifs sont effectués en découpant la pièce pour mesurer la profondeur de trempe au niveau des dents de l'engrenage. Une ingénieure méthodes et procédés cherche à expliquer la profondeur de trempe y en fonction de quatre variables `soaktime`, `soakpct`, `difftime`, `diffpct`, à partir d'un jeu de données de 32 observations qui forment un modèle régulier ¹.

1. Analyser les graphes de la dernière page.
2. Modéliser par une régression linéaire multiple. Préciser la matrice du plan d'expérience, la dimension du modèle. En supposant les données gaussiennes, quelle est la loi de l'estimateur des moindres carrés?
3. Les résultats de l'estimation par moindres carrés sont consignés dans le tableau suivant.

| Parameter | Estimate | Std_Error | t_value | Pr(> t) |
|-----------|------------|-----------|---------|----------|
| Intercept | 0.0205609 | 0.0089236 | 2.304 | 0.0291 |
| soaktime | 0.0023847 | 0.0001533 | 15.560 | 5.28e-15 |
| soakpct | -0.0038039 | 0.0057357 | -0.663 | 0.5128 |
| difftime | 0.0083840 | 0.0012055 | 6.955 | 1.79e-07 |
| diffpct | -0.0057911 | 0.0069046 | -0.839 | 0.4090 |

Déterminer l'équation de la fonction de régression.

- (a) Rappeler ce que représente chacune des colonnes.
La variable `diffpct` est-elle significative?
Le coefficient θ_{DT} est-il positif?
- (b) Déterminer un intervalle de confiance de β_{DT} , puis un intervalle de confiance de β_{DP} .
- (c) Estimer par intervalle de confiance de niveau 95% l'espérance moyenne de la profondeur de trempe quand `soaktime=0.6`, `soakpct=1`, `difftime=0.6`, `diffpct=0.9`, puis prédire par intervalle de confiance de niveau 95% une valeur individuelle de profondeur de trempe sous les mêmes conditions. Pouvez-vous faire l'application numérique?
Un stagiaire calcule les intervalles observés suivants: [0.016; 0.02] pour l'IC de la moyenne et [0.013; 0.023] pour l'intervalle de prédiction. Cela vous semble-t-il cohérent?

¹Generalized Linear Models. R. Myers, D. Montgomery, G. Vining. Wiley 2002

- (d) Rappeler la construction du test de significativité globale du modèle iid (constant) contre le modèle d'étude (modèle complet).

Vérifier le calcul de la statistique de test à partir du tableau d'analyse de variance.

| | Res.Df | SCR | Df | SCM | F |
|-----------------|--------|---------|----|---------|-----|
| Modèle constant | 31 | 0.00433 | | | |
| Modèle complet | 27 | 0.00014 | 4 | 0.00419 | 202 |

Déterminer une estimation ponctuelle de la variance du bruit dans le modèle complet.

- (e) Le coefficient de détermination vaut $R^2 = 0.96767$. Les graphes des résidus sont représentés en annexe. Commenter la validation de ce modèle.

Retrouver la statistique de Fisher de significativité globale en fonction du R^2 .

- (f) On considère le modèle (M2) qui ne contient que les deux variables `soaktime` et `difftime`. Compléter le tableau d'analyse de la variance. Quel modèle retenir-vous?

| | Res.Df | SCR | Df | SCM | F |
|---------|--------|----------|-----|-----|-----|
| M2 | ??a | 0.000147 | | | |
| complet | ??b | ??c | ??d | ??e | ??f |

Calculer le coefficient de détermination de (M2).

- (g) On considère maintenant les variables `x1=soaktime*soakpct` et `x2=difftime*diffpct`. Un modèle utilisant ces variables est-il toujours linéaire?

L'ajustement du modèle (M3) n'utilisant que ces deux variables (et l'intercept) donne les résultats suivants:

Coefficients:

| Parameter | Estimate | Std_Error | t_value |
|--------------------|----------|-----------|---------|
| Intercept | 0.01089 | 0.00112 | 9.72 |
| soaktime * soakpct | 0.00269 | 0.00015 | 17.93 |
| difftime * diffpct | 0.00932 | 0.00146 | 6.38 |

Residual standard error: 0.002581 on ?? degrees of freedom

R2: 0.9553

Rappeler la définition de l'erreur standard résiduelle.

Est-il possible de tester ce modèle contre les autres?

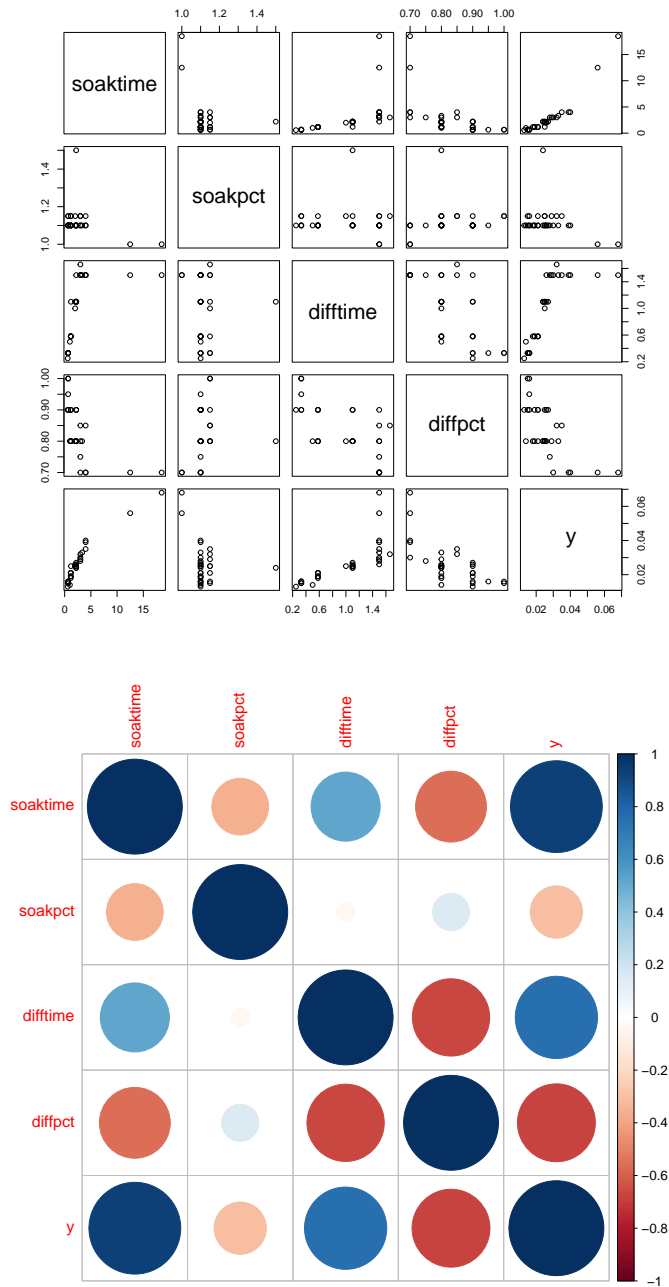


Figure 1: En haut: scatter plot des données - En bas: représentation des corrélations