

BE2- plans d'expériences et régression logistique
MOD - Statistique appliquée aux sciences de l'ingénieur.

Exercice 1 : Plan d'expériences - Criblage

On s'intéresse à l'apparition d'un défaut de fabrication récurrent lors de la fabrication de plaques de silicium (support des composants microélectroniques). Il s'agit d'un défaut de planéité de la plaque qui entraîne une détérioration des propriétés mécaniques des plaques fabriquées. Une partie des composants électroniques fabriqués sont alors rejetés, ce qui diminue le rendement de fabrication. L'objectif de l'étude est de détecter les facteurs influents sur la courbure des plaques de silicium. Les étapes les moins maîtrisées dans le processus de fabrication sont le laminage à froid de la plaque silicium et la cuisson de cette plaque en atmosphère réductrice. Les paramètres intervenant lors de ces deux étapes sont :

- le temps de laminage qui varie entre 10 et 25 secondes : L_{time}
- la température de laminage qui varie entre 55 et 75 degrés C : L_{temp}
- la pression de laminage qui varie entre 5 et 10 bars : L_{press}
- la température de cuisson qui se fait entre 1580 et 1620 degrés C : C_{temp}
- le temps de cuisson qui se fait entre 17,5 et 29 secondes : C_{time}
- l'atmosphère de cuisson dont la température de rosée est soit égale à 20 degrés C ou 26 degrés C : C_{atmos} .

Afin de sélectionner les variables influentes, un plan d'expériences a été réalisé. Chaque expérience est répliquée 4 fois. Les données sont contenues dans le fichier `silicium.txt`.

1. Importer le fichier `silicium.txt` sous R. Donner le nom du plan. Vérifier sous R les propriétés de ce plan. Ce plan permet-il d'estimer les effets principaux sans confusion ? les effets d'interactions ? Expliquer comment il a été construit.
2. Ajuster un modèle linéaire de la variable `Camber` en fonction des 6 facteurs. Analyser la sortie `summary` du modèle.
3. Retrouver par le calcul le chiffre de la colonne `Std Error`.
4. Estimer un modèle plus simple ne comprenant que les facteurs influents. Comparer l'estimation des coefficients avec le modèle précédent.
5. — quelles sont les conditions expérimentales qui permettent de minimiser la courbure `Camber` ?
— donner un intervalle de confiance pour la courbure moyenne en ce point de fonctionnement optimal
— quel est l'impact sur la courbure d'une augmentation de 5 degrés C de la température de laminage ?
6. Les hypothèses du modèle sont-elles vérifiées ? expliquer.

Exercice 2 : Régression logistique

Nous cherchons à prédire la variable **Pain** en fonction de 4 autres variables. Nous disposons pour ceci d'un fichier de 60 observations nommé `neuralgia.txt`.

1. Importer le fichier `neuralgia.txt` sous R. Décrire les variables.
2. Partager le fichier en un fichier d'apprentissage (80%) et un fichier de test (20%). On utilisera la commande `u = sample(1 : n, p)` pour choisir p éléments dans $1 : n$. L'ensemble `data[-u,]` est alors le complémentaire de l'ensemble `data[u,]`.
3. Réaliser sur le fichier d'apprentissage une régression logistique pour prédire la variable **Pain**. Quel est l'événement modélisé? Rappeler la forme du modèle logit. Dans cette question, on considère un modèle sans interaction.
4. Analyser le résultat des commandes `Anova(, type = "III", test.statistic = "LR")` et `Anova(, type = "III", test.statistic = "Wald")`. Interpréter la sortie de `summary`. Comparer à que donne la commande `anova(, test = "chisq")`. Quelles sont les variables influentes? Simplifier le modèle.
RQ : il faut charger et installer le package "car" pour faire fonctionner la routine `Anova`.
5. Réaliser maintenant une procédure forward pour le critère AIC. Expliquer le principe et donner le modèle obtenu. On adaptera le code vu en cours. Le modèle initial est réduit à la constante, le modèle le plus complet est le modèle contenant toutes les interactions.
6. A l'aide du fichier test, comparer les matrices de confusions pour les deux modèles (on utilisera la commande `predict` pour réaliser les prévisions sur le fichier test puis la commande `table` pour obtenir la matrice de confusion).
7. On se fixe un modèle. Etudier la sensibilité des qualités prédictives à l'échantillon (prendre 50 ensembles d'apprentissage constitué de 80% des individus) puis étudier la sensibilité des résultats à la proportion apprentissage/test (10%, 20%, 30%, 40%).