

Exercice - Session 1

Le pourcentage de viande maigre de carcasses de porc (LMP pour Lean Meat Percentage) est le principal critère sur lequel la valeur commerciale de la viande est calculée. Dans certains abattoirs, ce critère est prédit à partir des données issus d'un scanner à imagerie tomographique qui fournit un profil du pourcentage de viande maigre le long de la carcasse. Pour estimer un modèle de prédiction du pourcentage de viande maigre à partir de ce profil, on dispose de données expérimentales consistant en 117 observations conjointes du profil fourni par le scanner et du pourcentage de viande maigre. Ces données sont disponibles dans le fichier `scanner.txt`

La problématique est donc la suivante : comment prédire le pourcentage de viande maigre à partir des données du scanner ?

1. Dans cette problématique, quelle est la variable réponse ? Quelles sont les variables explicatives ? Vous donnerez aussi la nature (quantitative ou catégorielle) de ces variables.
2. Quelle est la population concernée par cette problématique ?
3. Proposez un modèle permettant de prédire le pourcentage de viande maigre à partir des variables explicatives.
4. Ajustez le modèle complet par la méthode des moindres carrés en utilisant la fonction `glm`. En examinant les coefficients estimés de ce modèle, expliquez la procédure réellement exécutée par la fonction `glm`.
5. Donnez une valeur numérique ou un graphique permettant d'évaluer la qualité d'ajustement de ce modèle.
6. Selon vous, la méthode consistant à estimer le modèle complet comme ci-dessus permet-elle d'obtenir un bon outil pour la prédiction. Argumentez votre réponse.
7. Si vous deviez ne choisir qu'une seule variable explicative parmi toutes celles obtenues par le scanner, laquelle choisiriez-vous ?
8. La méthode consistant à estimer le modèle construit après sélection de la meilleure variable explicative comme ci-dessus permet-elle d'obtenir un bon outil pour la prédiction. Argumentez votre réponse.