

TP 2 : Correction

II-3-d) Les points $(Y_i, \hat{Y}_i)_{i=1, \dots, n}$ ne sont pas proches de la droite $y = x$, ce qui indique une mauvaise qualité de prédiction. Néanmoins, la corrélation entre les valeurs observées et prédites par validation croisée (égale à 0.62) est "significativement" positive (bien qu'il ne s'agisse pas d'un test!), ce qui semble indiquer un lien entre la consommation d'alcool et la mortalité.

II-4) Le graphe montre une très bonne qualité d'estimation, la courbe estimée et la vraie fonction de régression étant très proches.

II-5) Une bonne qualité d'estimation mais une mauvaise qualité de prédiction signifient que la consommation d'alcool est un paramètre insuffisant pour bien prévoir l'âge de décès. En d'autres termes, la variabilité individuelle non expliquée par la consommation d'alcool reste importante.

III-3) Les points $(Y_i, \hat{Y}_i)_{i=1, \dots, n}$ sont proches de la droite $y = x$, c'est à dire que les Y_i sont bien prédits par le modèle : le taux d'endocrine est bien prédit par le taux d'hormone administré.

III-4) Les points $(Y_i, \hat{Y}_i)_{i=1, \dots, n}$ forment un nuage non structuré, et la corrélation entre valeurs prédites et observées est négative (-0.2), ce qui indique que le modèle linéaire ne permet pas du tout de prédire le taux d'endocrine en fonction de la dose d'hormone.

L'erreur de validation croisée est de 0.01 avec l'estimateur par polynômes locaux, et est 12 fois plus élevée avec le modèle linéaire. Cela confirme les bien meilleures performances prédictives de l'estimateur par polynômes locaux.