

Régression Linéaire en R

1 Introduction

La régression linéaire est une méthode statistique permettant de modéliser la relation entre une variable dépendante (Y) et une ou plusieurs variables explicatives (X). Elle repose sur plusieurs hypothèses fondamentales et permet d'effectuer des prédictions et d'analyser l'influence des variables explicatives sur la variable cible.

Il existe deux types principaux de régression linéaire :

- **Régression linéaire simple** : Une seule variable explicative (X).
- **Régression linéaire multiple** : Plusieurs variables explicatives (X_1, X_2, \dots, X_p).

Dans ce cours, nous allons utiliser **R** pour illustrer ces concepts avec un exemple appliqué en agriculture.

2 Régression Linéaire Simple

2.1 Importation des données

Nous utilisons un fichier de données contenant des observations sur le rendement agricole (Y) en fonction de la quantité d'engrais appliquée (X).

```
# Importation des données
RLS <- read.table(file.choose(), header=TRUE)
attach(RLS)
names(RLS) # Afficher les noms des variables
```

2.2 Visualisation des données

Avant d'ajuster le modèle, nous affichons un nuage de points pour observer la relation entre les variables :

```
plot(Rendement ~ Engrais, xlab="Engrais (Kg/ha)", ylab="Rendement (t/ha)")
```

2.3 Ajustement du modèle

Nous utilisons la fonction `lm()` pour ajuster une régression linéaire simple :

```
# Estimation des paramètres du modèle
modele1 <- lm(Rendement ~ Engrais)
summary(modele1) # Afficher les résultats du modèle
```

2.4 Visualisation de la droite de régression

Nous traçons la droite ajustée sur le nuage de points :

```
plot(Rendement ~ Engrais, xlab="Engrais (Kg/ha)", ylab="Rendement (t/ha)")
abline(modele1, col="blue")
```

2.5 Validation du Modèle

2.5.1 Test de normalité des résidus

```
par(mfrow=c(1,2))
hist(residuals(modele1), main="Histogramme des résidus")
qqnorm(resid(modele1))
qqline(resid(modele1), lty=2, col="blue")
shapiro.test(resid(modele1)) # Test de Shapiro-Wilk
```

2.5.2 Test de l'homogénéité des résidus

```
library(lmtest)
plot(residuals(modele1) ~ fitted(modele1), xlab="Valeurs prédites", ylab="Résidus")
bptest(modele1) # Test de Breusch-Pagan
```

2.5.3 Test d'autocorrélation des résidus

```
library(lmtest)
dwtest(modele1) # Test de Durbin-Watson
```

3 Régression Linéaire Multiple

3.1 Importation et Visualisation des données

Nous incluons ici plusieurs variables explicatives (*Engrais, Irrigation, Type_de_sol*).

```
RLM <- read.table(file.choose(), header=TRUE)
attach(RLM)
pairs(RLM) # Matrice de dispersion
```

3.2 Ajustement du Modèle

```
modele2 <- lm(Rendement ~ Engrais + Irrigation + Type_de_sol)
summary(modele2)
```

3.3 Validation du Modèle

3.3.1 Test de normalité des résidus

```
par(mfrow=c(1,2))
hist(residuals(modele2), main="Histogramme des r\'esidus")
qqnorm(resid(modele2))
qqline(resid(modele2), lty=2, col="blue")
shapiro.test(resid(modele2))
```

3.3.2 Test de l'homogénéité des résidus

```
plot(residuals(modele2) ~ fitted(modele2), xlab="Valeurs pr\'edites", ylab="R\'esidus")
bptest(modele2)
```

3.3.3 Test de multicollinéarité

```
library(car)
vif(modele2) # Facteur d'inflation de la variance
```

3.4 Conclusion

La régression linéaire est un outil puissant pour analyser les relations entre variables quantitatives. Les tests de validation sont essentiels pour s'assurer que les hypothèses du modèle sont respectées.

Extensions possibles :

- Régression polynomiale
- Régression logistique
- Régression ridge et lasso

Ces techniques permettent d'affiner les modèles et de mieux capturer les relations complexes entre les variables.