



MACHINE LEARNING SUPERVISÉ

Estimer le coût
de la couverture
médicale d'un.e
américain.e

1



Sommaire


Analyse des données

- Variables qualitatives et quantitatives
- Statistique descriptive

Modélisation

- Régression Linéaire Multiple
- Forêts Aléatoires

Évaluation et comparaison des modèles



2

Analyse des données

- 1338 lignes / 7 colonnes
- Complet
- 4 variables quantitatives

- **charge**

- **bmi**

- **children**

- **age**

- 3 variables qualitatives

- **sex**

- **smoker**

- **region**

Variable expliquée

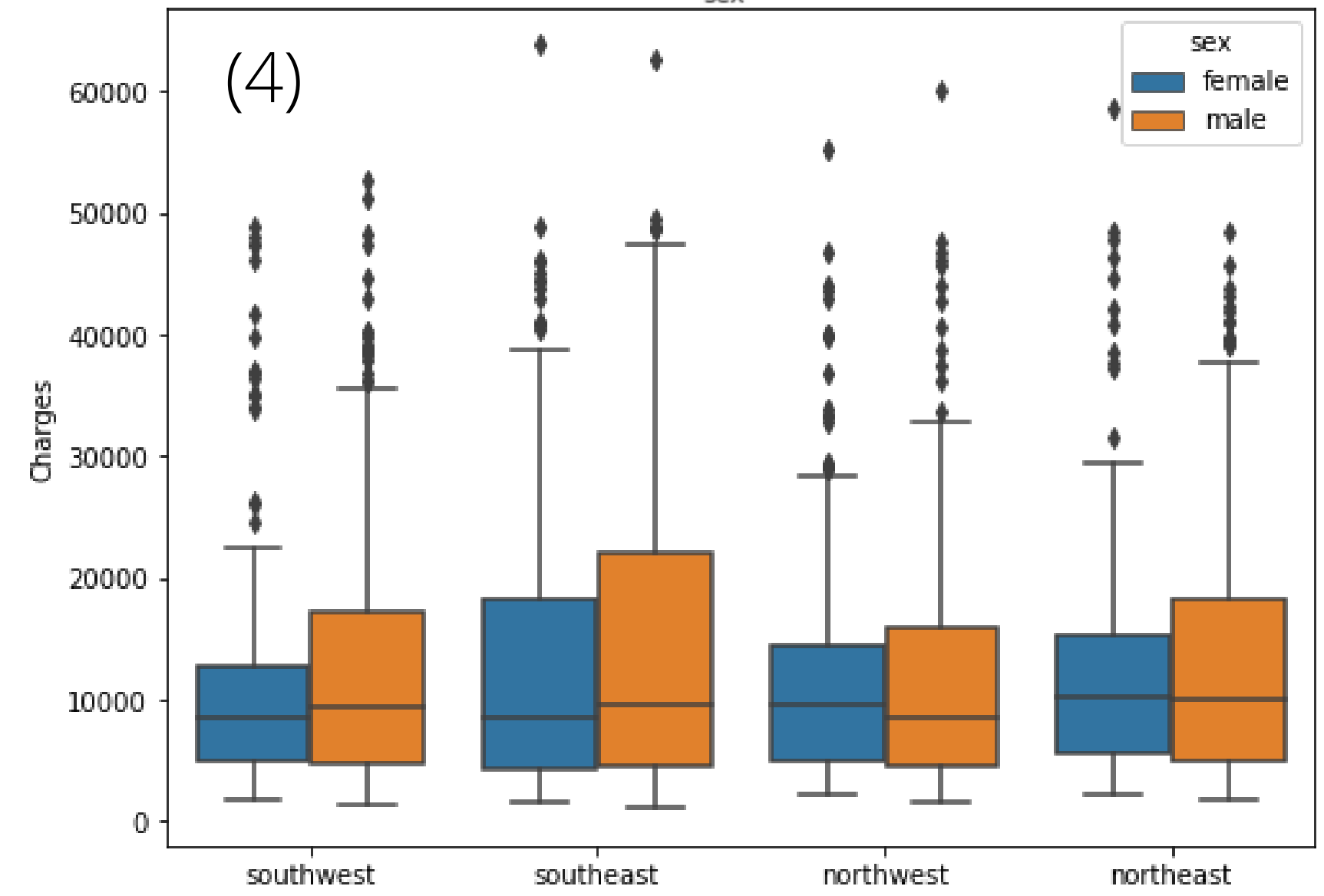
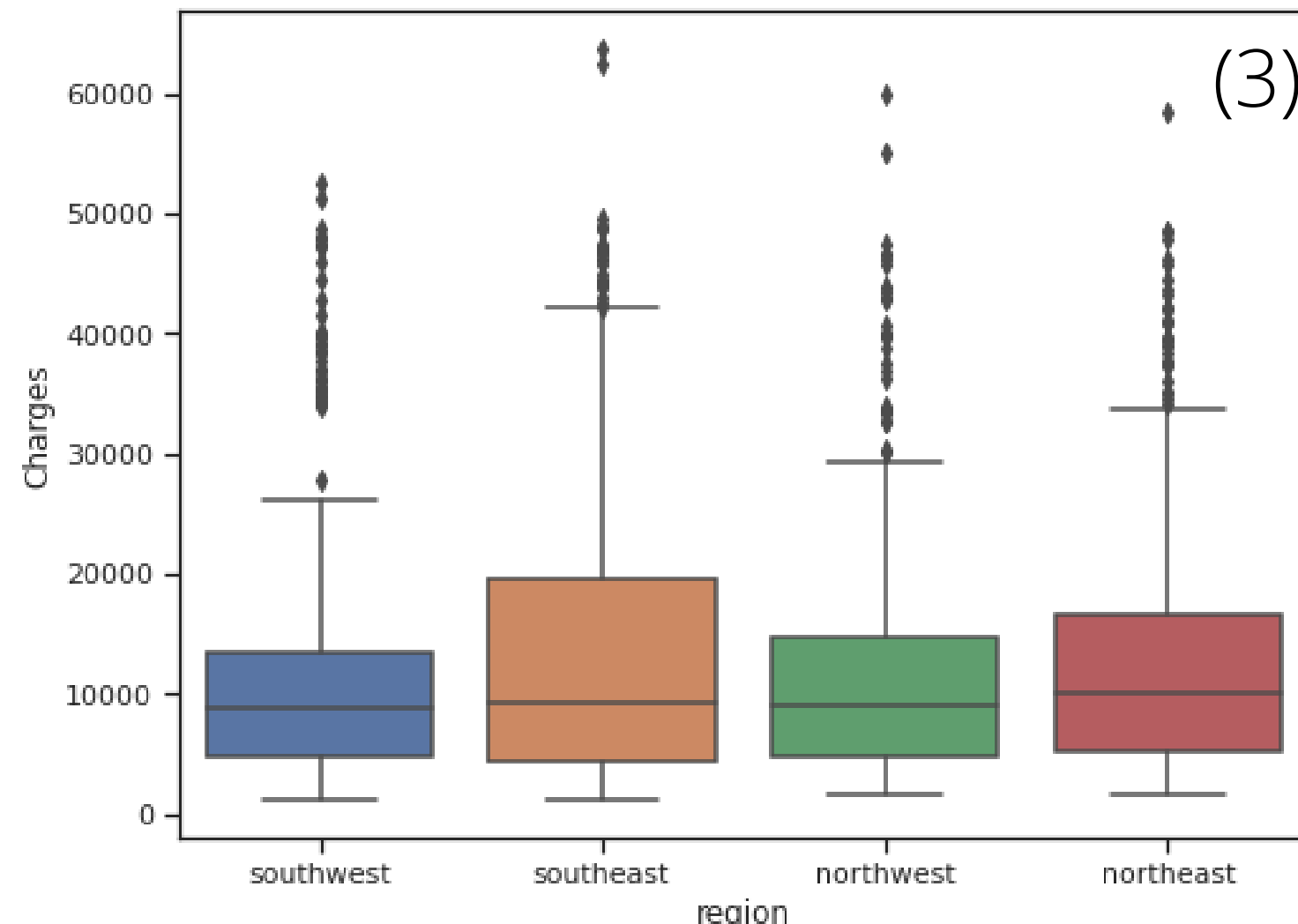
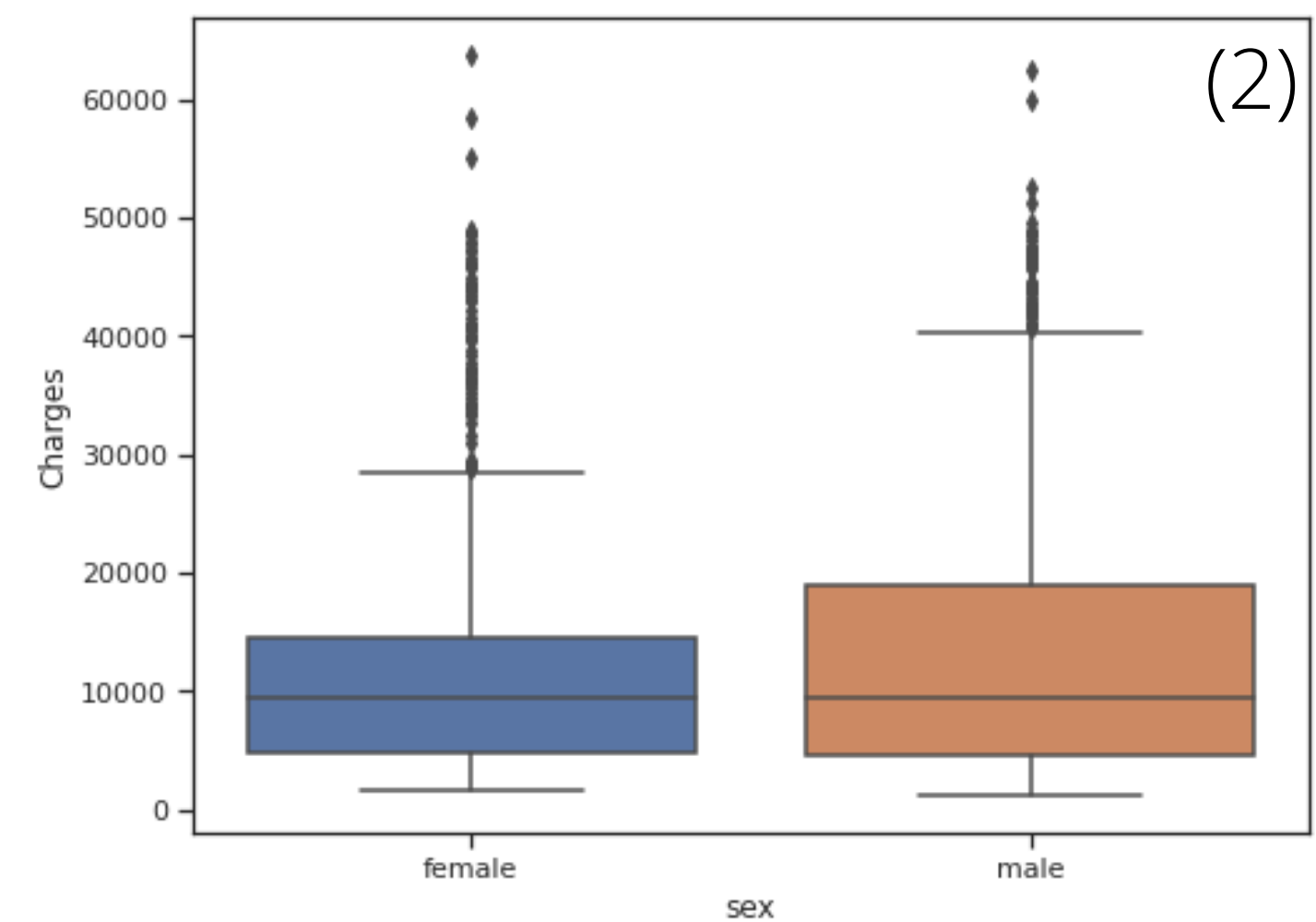
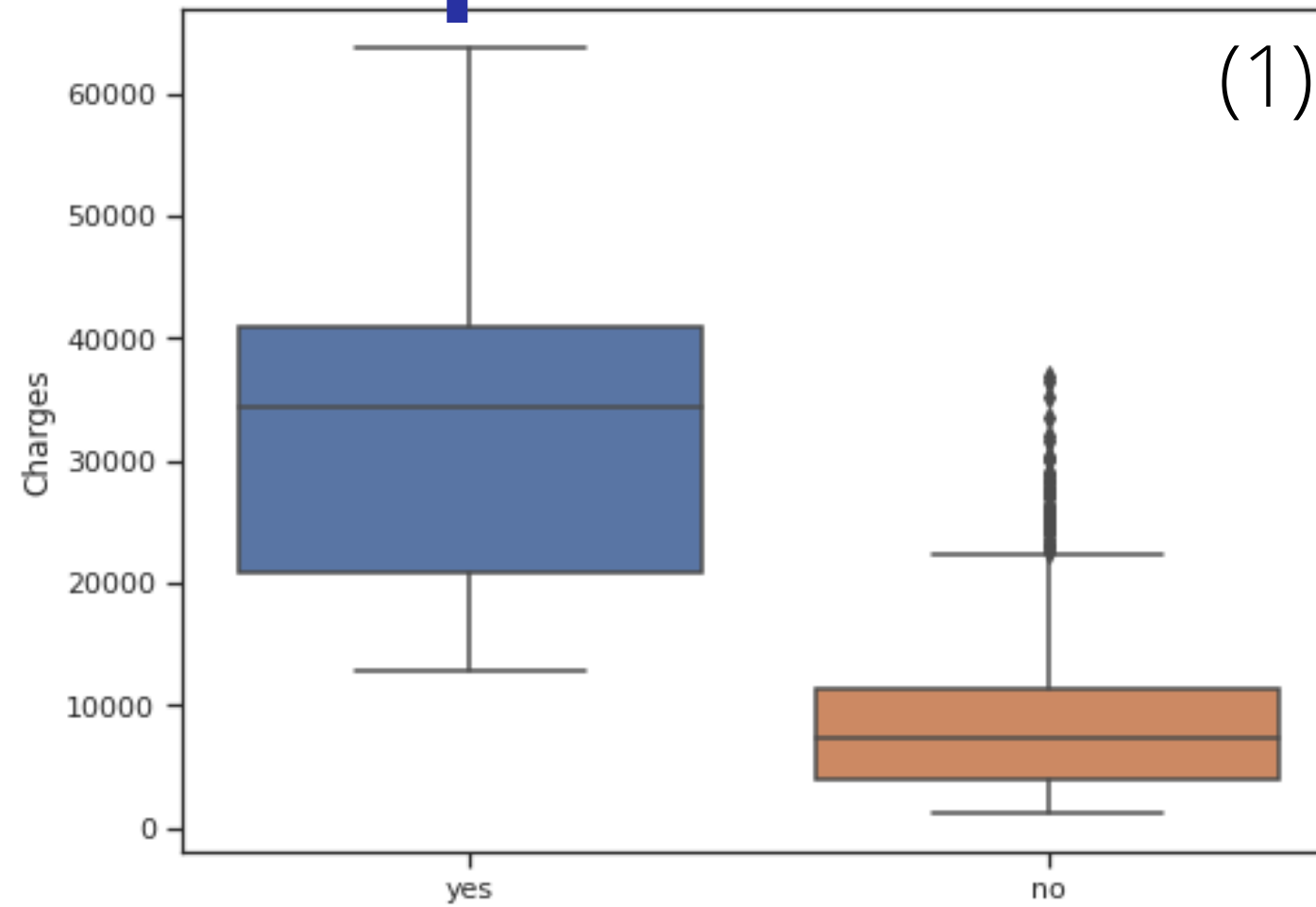
Variables explicatives
quantitatives

Variables explicatives
qualitatives

Nécessite encodage

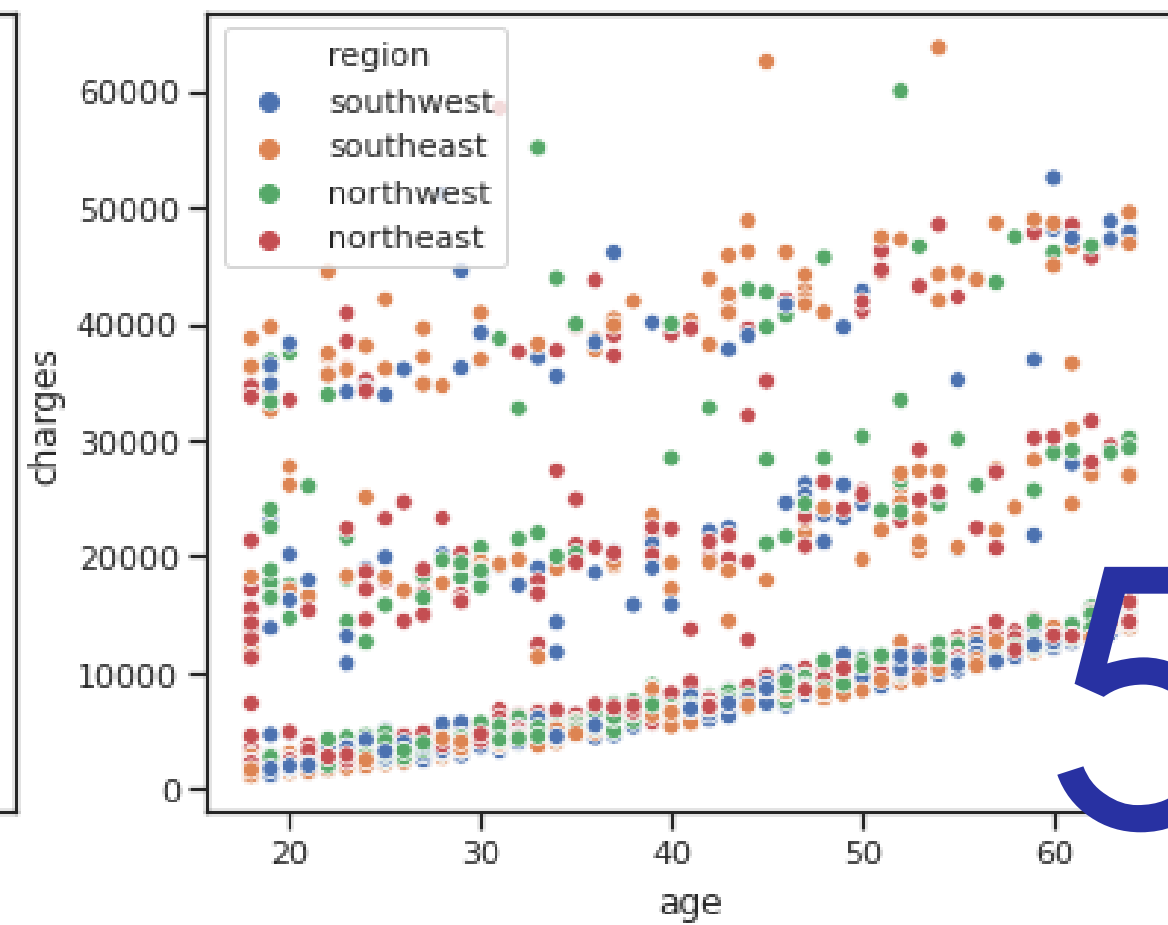
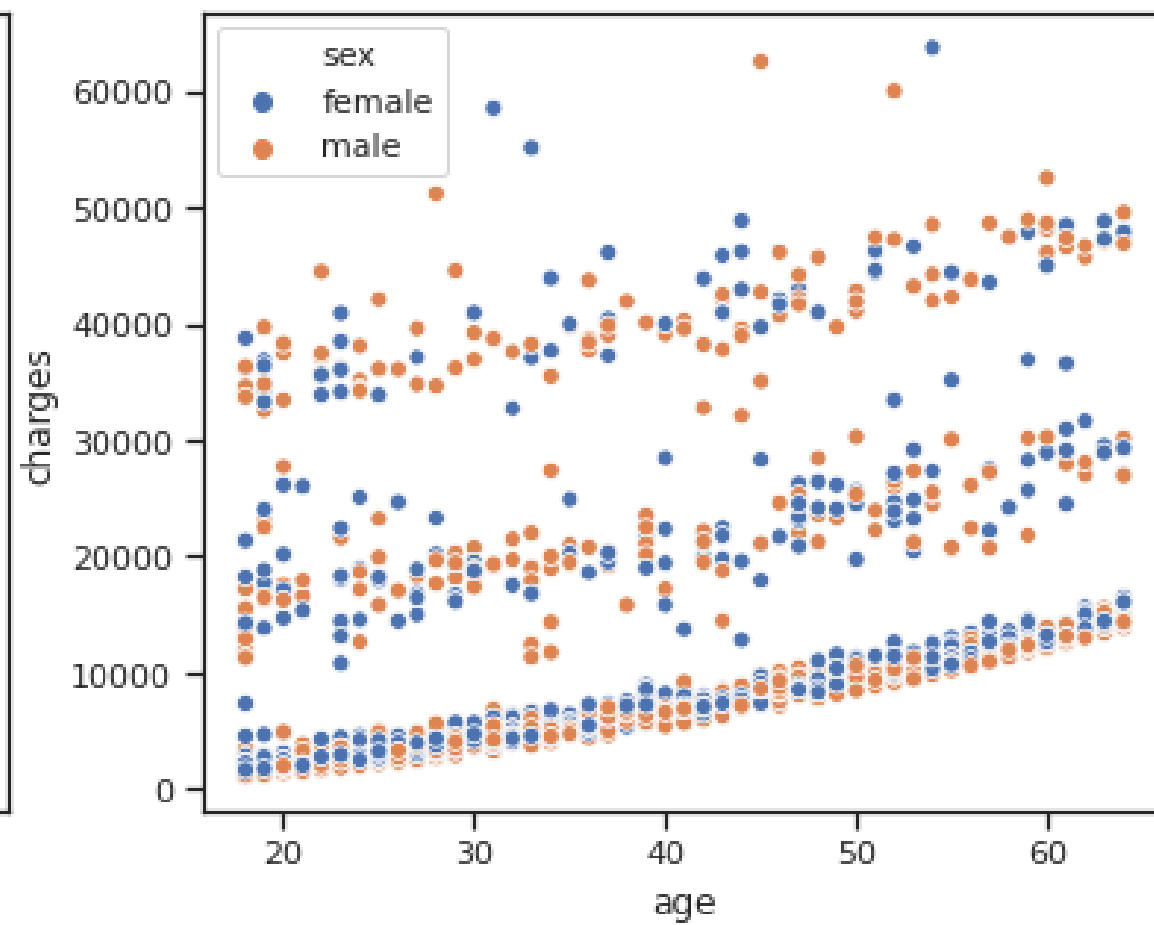
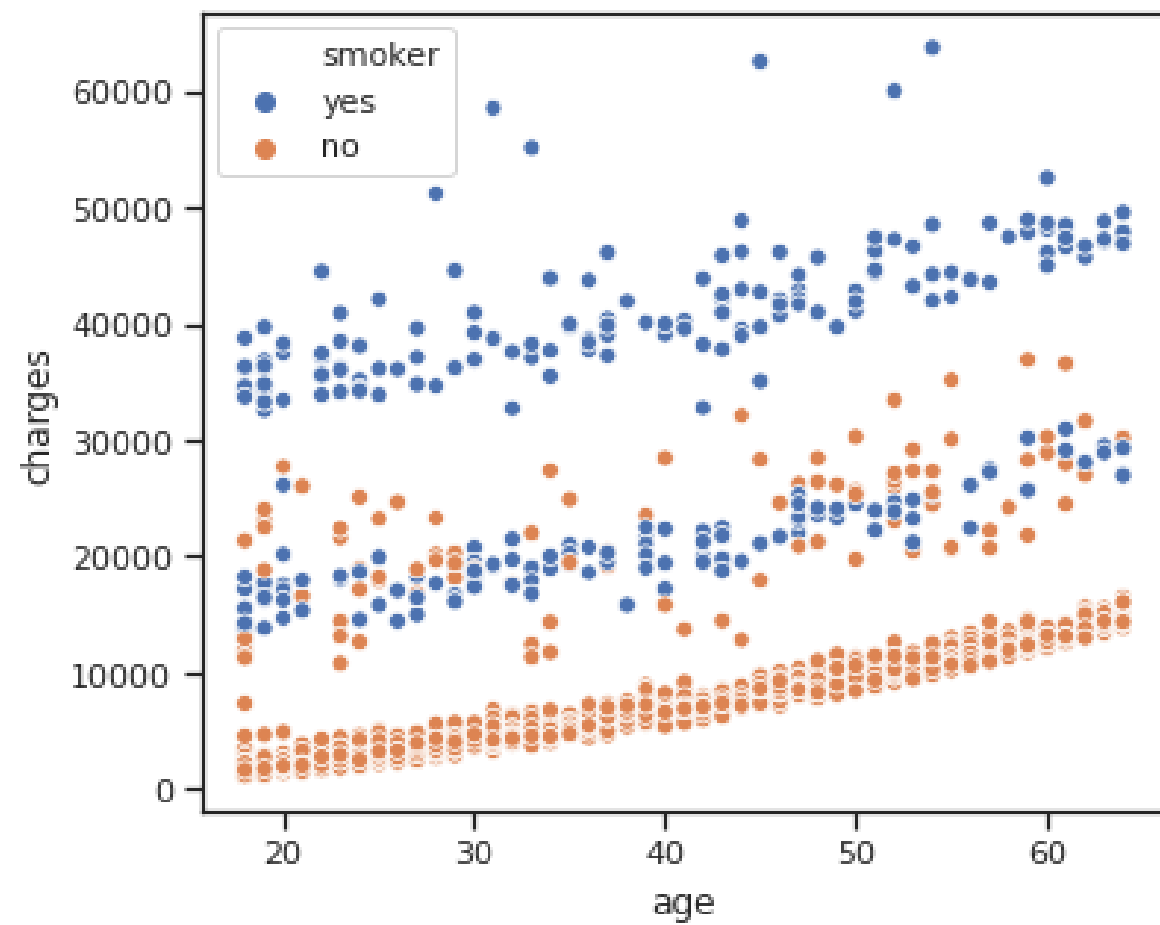
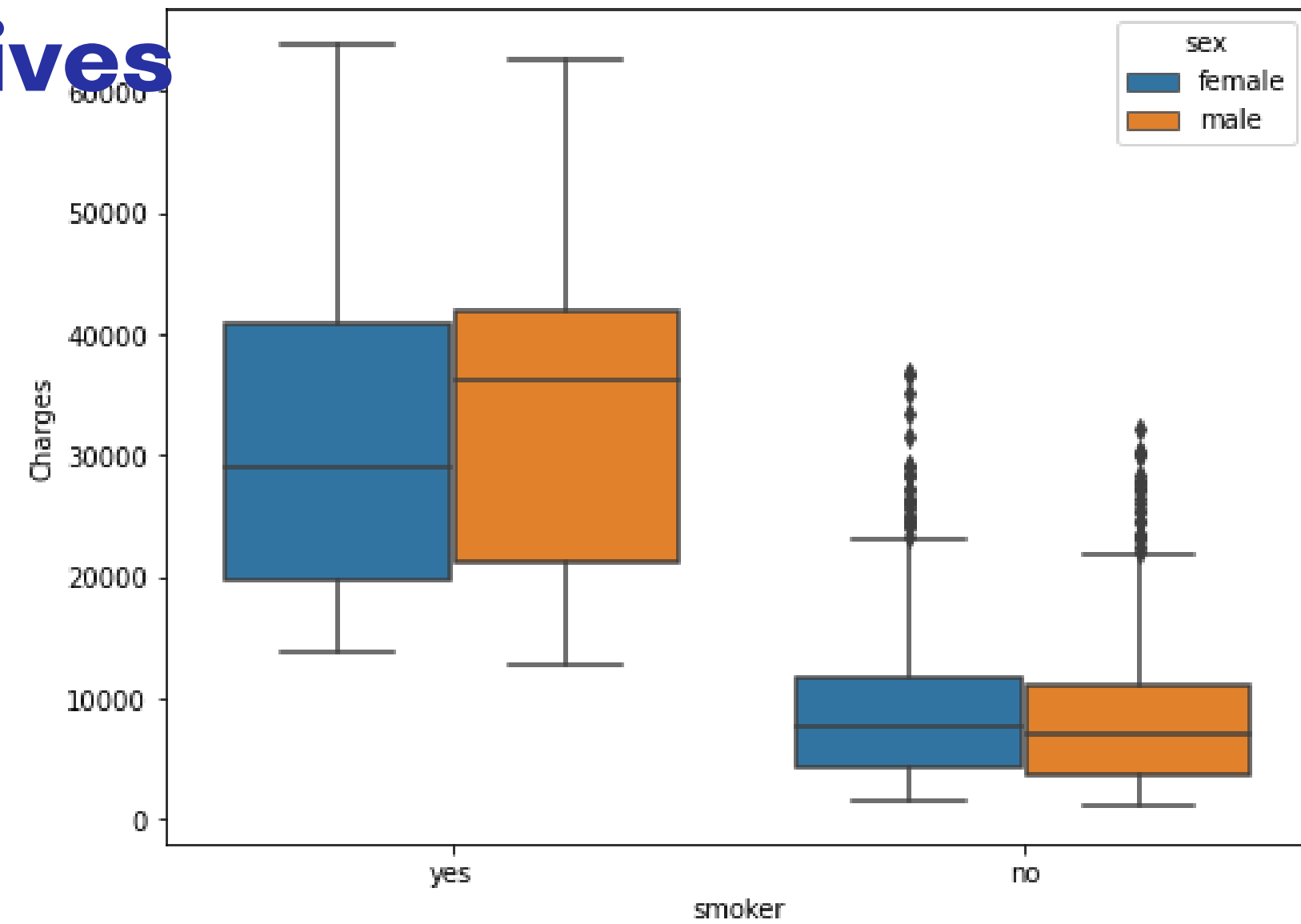
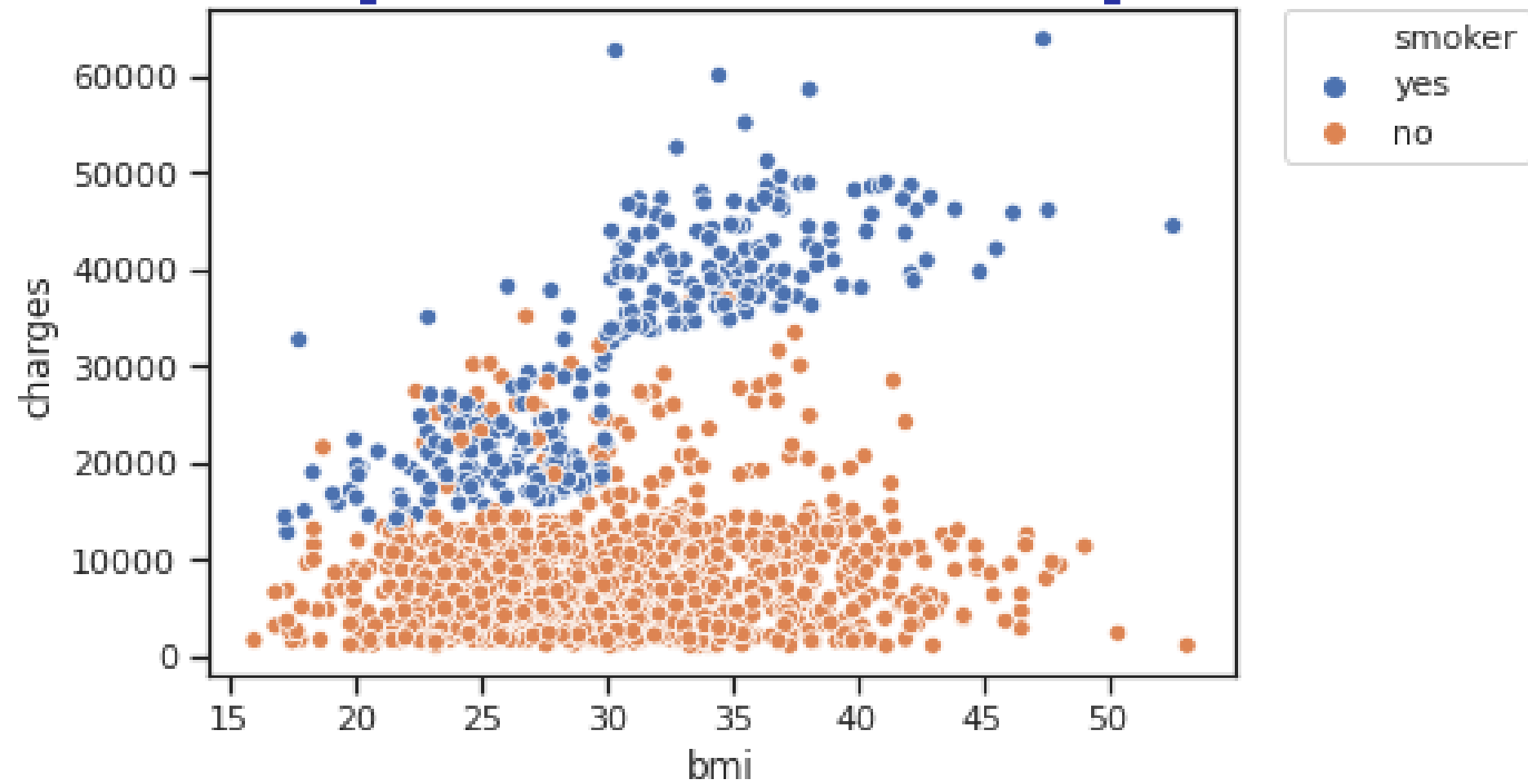
3

Variables qualitatives



4

Variables qualitatives et quantitatives

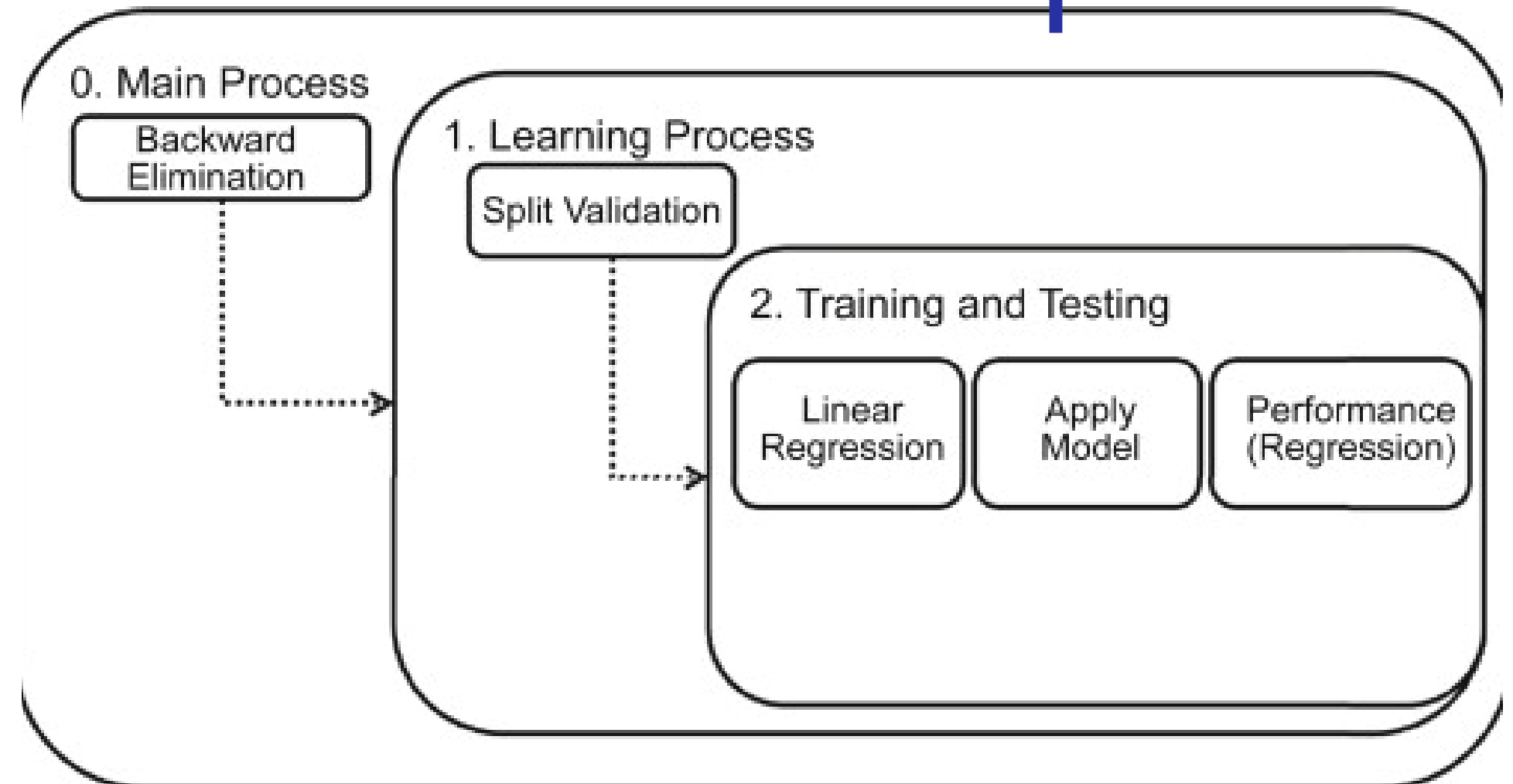


Régression Linéaire Multiple

$$\hat{y} = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n + \epsilon$$

Diagram illustrating the components of the multiple linear regression equation:

- \hat{y} : target (indicated by a red arrow)
- $\beta_0, \beta_1, \dots, \beta_n$: coefficients (indicated by a grey arrow)
- X_1, \dots, X_n : inputs (indicated by a blue arrow)
- ϵ : random error (indicated by a green arrow)



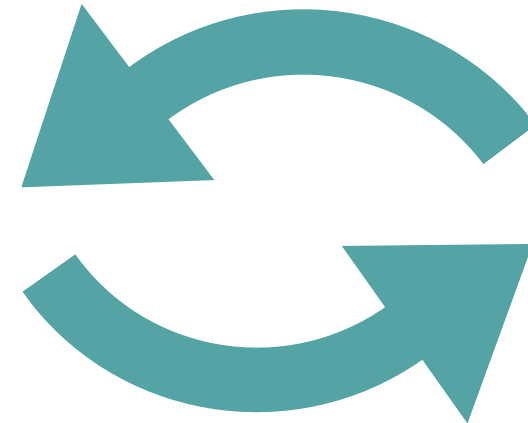
age/bmi/children/sex/smoker/region

(R^2 -ajusté) égale à 0.762 et une précision de 53 % vs (R^2 -ajusté) égale à 0.763 et une précision de 54 %. age/bmi/children/smoker/region

Charges = -12466.156 + 23347.627***smoker_yes** + 880.117***region_northeast** + 244.7259***age** + 340.0924***bmi** + 619.166***children**.

Random Forests

Création du modèle



Trouver ses paramètres optimaux

Paramètre par défaut

SCORE R^2 :

0.831

Paramètres optimisés

SCORE R^2 :

0.854

'max_depth': 4
"n_estimators": 38'

Traitement des features

SCORE R^2 :

0.852

"smoker" 70%
"bmi" 16%
"age" 11%

7

Comparaison

Régression linéaire

Forêts aléatoires

SCORE R^2 :

0.76

VS

SCORE R^2 :

0.85

Le modèle "Forêts aléatoires" l'emporte !

