

# Projet - Modèles linéaires

Axel Struys - Alexis Buckens

December 28, 2016

## 1 Introduction

Pour réaliser ce projet, nous avons choisi une base de données en provenance de l'UCI machine learning repository. Ce dataset décrit plusieurs modèles de voitures en fonction de différentes caractéristiques. Plusieurs variables dépendantes pourraient être choisies pour ce dataset. Afin de restreindre notre analyse à une seule variable dépendante, nous avons choisi une question de recherche: "Quelles caractéristiques prédisent la consommation de carburant sur autoroute d'une voiture?". Le dataset contenait à l'origine 26 variables: 16 continues et 10 nominales. Pour l'analyse, nous avons choisi 13 variables continues et 2 variables nominales. Nous avons exclu 3 variables continues car elles étaient des variables dépendantes. Nous avons exclus 8 variables nominales afin de simplifier le modèle. Il y avait 205 observations à l'origine, mais pour 46 d'entre-elles, il y avait des valeurs manquantes. Nous avons supprimé ces observations. Le dataset ainsi nettoyé contient 159 observations.

Dans la table 1 nous présentons les statistiques descriptives pour les 13 variables continues, ainsi que les unités correspondantes. La table 2 présente les fréquences observées des deux variables discrètes : Aspiration et Engine type. Ces deux variables ont deux niveaux. On peut remarquer que les fréquences observées ne correspondent pas aux fréquences attendues: 40 observations par cellule. Pour l'analyse de la variance, nous sommes dans une situation où le design n'est pas balancé.

	mean	std.dev	median	min	max
wheelbase	98.24	5.16	96.90	86.60	115.60
length	172.32	11.55	172.20	141.10	202.60
width	65.60	1.95	65.40	60.30	71.70
height	53.88	2.28	54.10	49.40	59.80
curbweight	2459.45	480.90	2338.50	1488.00	4066.00
enginesize	119.09	30.41	110.00	61.00	258.00
bore	3.30	0.27	3.27	2.54	3.94
stroke	3.24	0.29	3.27	2.07	4.17
compressionratio	10.15	3.88	9.00	7.00	23.00
horsepower	95.88	30.63	88.00	48.00	200.00
peakrpm	5116.25	465.29	5200.00	4150.00	6600.00
highwaympg	32.07	6.44	32.00	18.00	54.00
price	11427.68	5863.79	9164.00	5118.00	35056.00

Table 1: Statistiques descriptives des 13 variables continues

Aspiration	Engine type	
	Diesel	Gas
Standard	5	127
Turbo	10	17

Table 2: Fréquences observées pour les variables Aspiration & Engine type