



**FACULTÉ
D'ÉCONOMIE**

Le degré d'impact des différentes caractéristiques techniques des voitures modernes sur leur consommation en carburant

Projet Individuel de Recherche

Master MBFA

Saunders Patrick (SIEF)

Chantrel Clement (Actuariat)

Sous la responsabilité de Benoit Mulkay

Table des matières

Résumé.....	
.....	3
Abstract.....	
.....	3
I.	
Introduction.....	
.....	4
II. La base de	
données.....	4
2.1 Data cleansing, Data	
tidying.....	4
2.2 Data	
mining.....	5
2.3 Biais	
d'échantillonnages.....	5
2.4 Descriptions des	
variables.....	5
III. Analyse	
descriptive.....	5
IV Modèle et	
régressions.....	5
Conclusion.....	
.....	5
Bibliographie.....	
.....	5

Résumé

L'objectif principal de ce projet est de déterminer les caractéristiques techniques ayant un impact sur la consommation des véhicules sortis en 2019 , non seulement en situation mixte, mais aussi en ville et sur autoroute. Nous sommes donc passé par une analyse d'une base de données des véhicules sortis en 2019 en fonction de leur caractéristiques techniques. À travers une analyse descriptive et de modèles économétriques nous avons pu déterminer les facteurs importants ayant un impact positif et négatif sur la consommation, mixte, en ville et sur l'autoroute.

I. Introduction

L'environnement étant un sujet d'actualité dans la politique et l'économie, et avec une conscience sociale de l'environnement menée par des études de plus en plus inquiétantes sur l'évolution des températures globales dans les années à suivre, les constructeurs de voitures cherchent à réduire la consommation par kilomètres de leurs voitures afin de répondre aux attentes des particuliers et leur désirs de voitures moins polluantes et plus efficiente.

Cette recherche d'efficience dans la consommation n'est désormais pas sans scandale, en 2015 l'affaire Volkswagen a fait surmonter une technique employée par VW de réduire la pollution engendrée par leur véhicules en lançant un programme activant leur contrôles d'émission seulement lors de tests en laboratoire, en conditions réelles leur pollution de NO_x étaient 40 fois plus élevées.

Les voitures étant au coeur du transport de gens dans le monde entier, pour le travail et le loisir, surtout en régions non-urbanisées, la réduction de pollution de ces voitures est un sujet important dans notre quotidien.

Cette étude porte sur l'ensemble de voitures sorties en 2019 aux états-unis, nous remarquons que ces voitures sont dans l'ensemble puissantes, ne représentant pas le marché européen de voitures. ^[1]

Après une étude et nettoyage de notre base de données, nous sommes passé par une analyse en appliquant les méthodes de statistiques descriptives, de data mining afin de retrouver des données pour des caractéristiques qui nous semblaient pertinentes, et des modèles économétriques (par regression par OLS et SUR) à travers le logiciel R.

II. La base de données

Nous avons utilisé la base de donnée de l'U.S Department of Energy

(<https://www.fueleconomy.gov/feg/download.shtml>).

Pour des raisons de simplicité, nous avons uniquement utilisé la base de données de 2019.

Cette base de données recense toutes les voitures sorties en 2019 sur le marché nord-américain, en fournissant leur caractéristique techniques et d'autres critères pertinent à l'étude de leur consommation de carburant.

La base de donnée contient initialement 1264 lignes (différents modèle de voitures), et 162 variables.

2.1 Data cleansing, Data tidying

Pour mener à bien notre étude, nous avons nettoyé la base de donnée initiale :

- Supprimer les variables avec trop de valeurs absentes, les variables qui nous sont inutiles, les variables redondantes et les variables dont on ne connaît pas la définition.
- Créer des variables pour condenser les informations et des variables indicatrices pour aider notre étude statistique .
- Convertir les mesures impériales en mesure métriques pour une analyse plus rapide de notre part.

La nouvelle base de donnée comprend maintenant 53 variables, dont 21 variables indicatrices.

2.2 Data mining

Lors de l'étude de la base de données sur les caractéristiques techniques de nouvelles voitures sorti en 2019, nous avons observé qu'il manquait des données sur le poids et la puissance de ces véhicules, en faisant l'hypothèse que ces différents facteurs seraient cruciaux à notre étude, nous avons pris la décision de tenter de récupérer ces informations afin de pouvoir les traiter. Nous nous sommes donc tournés vers le data mining.

La première étape a été de choisir un site web qui aurait ces informations disponibles, structuré de façon adéquate qui permettra la recherche successive de chaque voiture dans notre base de données et de pouvoir retirer les informations essentielles.

Nous avons donc créé un script en Python (**ANNEX X**) pour aller chercher ces informations sur nos 1264 modèles de voiture sur internet.

.....

Une fois retirés, nous les avons ajoutés à notre base de données.

L'étape finale a été de rajouter les valeurs manquantes.

2.3 Biais d'échantillonnages

Nous avons choisi cette base de données pour son accessibilité, sa source sûre, son nombre important d'informations et sa relative propreté mais après analyse nous avons été heurté à un problème : Les voitures du marché nord américain sont conçues de manière différente, elles sont bien plus spacieuses, lourdes et puissantes, équipées de moteurs de gros cylindrée, contrairement aux voitures du marché européen (et asiatique) .

Aussi, la technologie Diesel y est beaucoup moins populaire.

Sûrement pour des raisons économiques et politiques (le prix du pétrole y est généralement moins cher et les normes anti pollution sont bien moins sévères que chez nous) et culturelle, les constructeurs des USA privilégient en premier le confort et la puissance.

Cet échantillon représente donc mal la population qu'est le parc automobiles des pays développés.

De plus, cet échantillon contient une grande quantité de voitures de loisirs très performantes, donc qui consomment, alors que ces modèles sont moins utilisés que les modèles plus sobres.

III. Analyse descriptive

Nous avons analysés nos différentes variables en fonction de la consommation, pour se faire une idée de comment nous allons procéder pour les régressions.

La consommation est en litre par 100 km .

Différence de consommation entre les voitures essence et Diesel

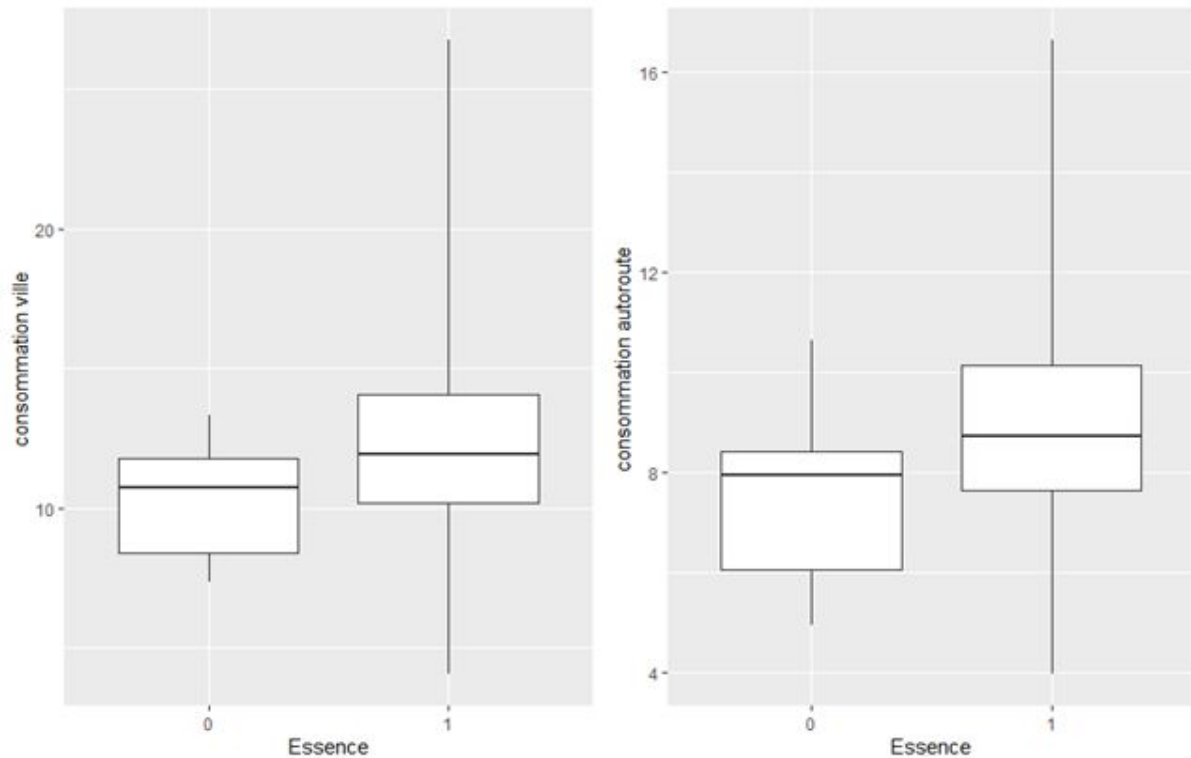


Illustration 1: consommation entre voitures essence et diesel en ville et autoroute

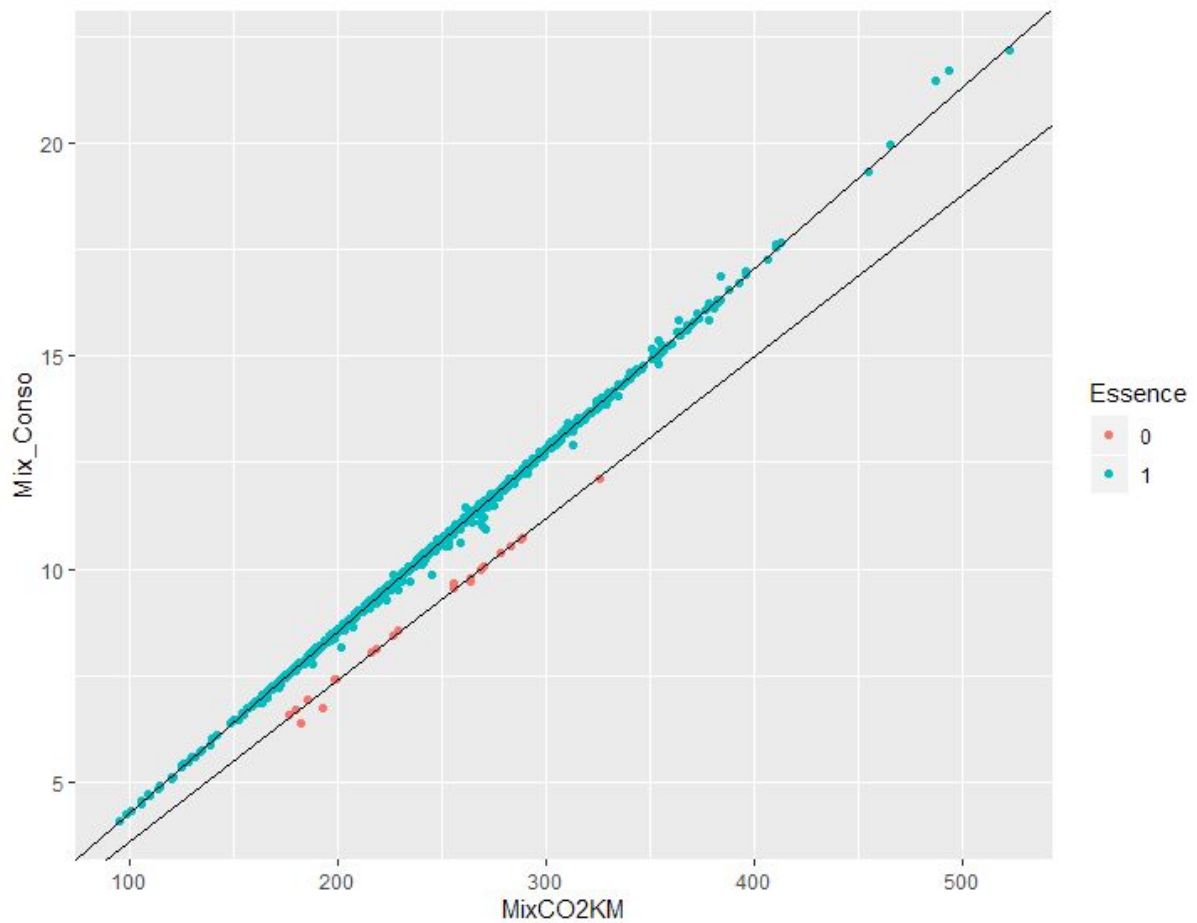
Nous pouvons voir que les voitures Diesel consomment moins que ce soit en ville ou sur autoroute (1 si essence, 0 sinon) .

Le taux de compression d'un moteur Diesel étant plus élevé, le rendement est meilleur. Leur consommation est donc moindre à celle des moteurs à essence classique. **[1]**

Malheureusement, notre échantillon contient très peu de voiture Diesel.

Nous pouvons aussi voir que que les voitures consomment plus en ville. La vitesse (et aussi le régime moteur) en ville étant beaucoup moins constante que sur autoroute.

Consommation et rejet de Co2

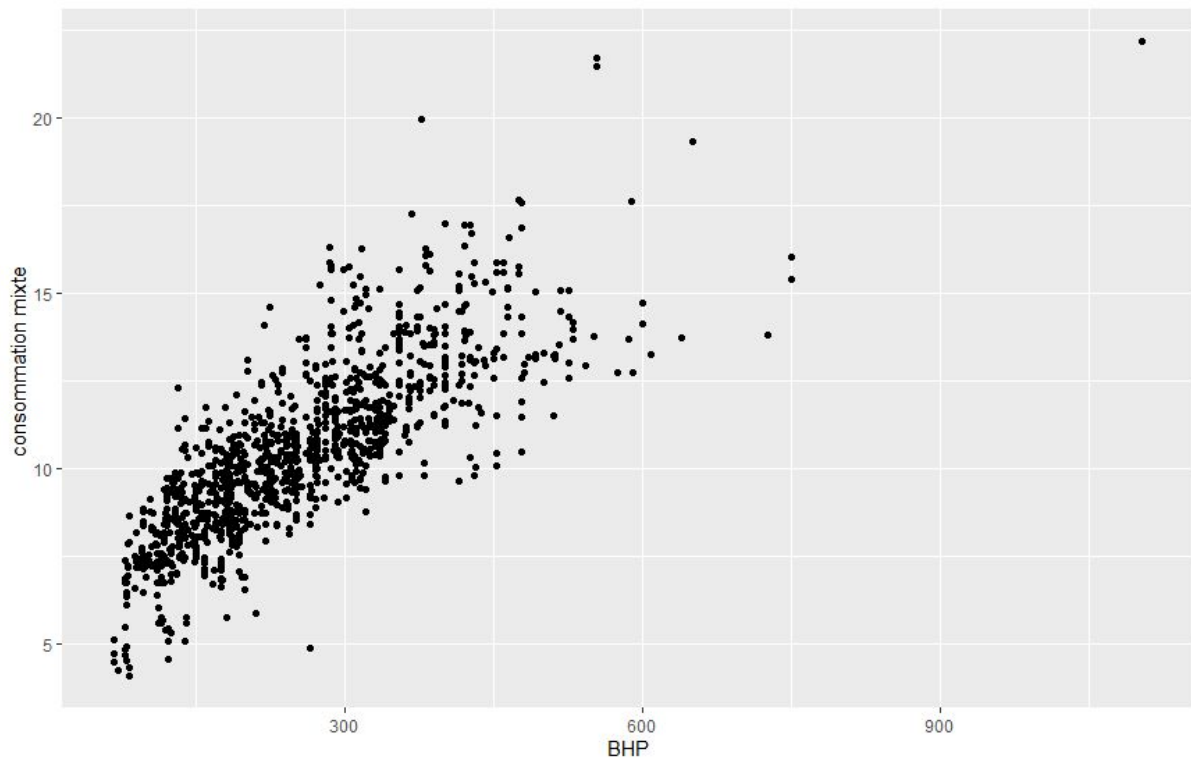


Sur ce graphique, nous avons la consommation mixte (l/100km) en ordonnée et l'émission de Co2 (g/km) en abscisse.

Nous pouvons voir la relation directe entre ces deux variables : le rejet de Co2 augmente avec la consommation.

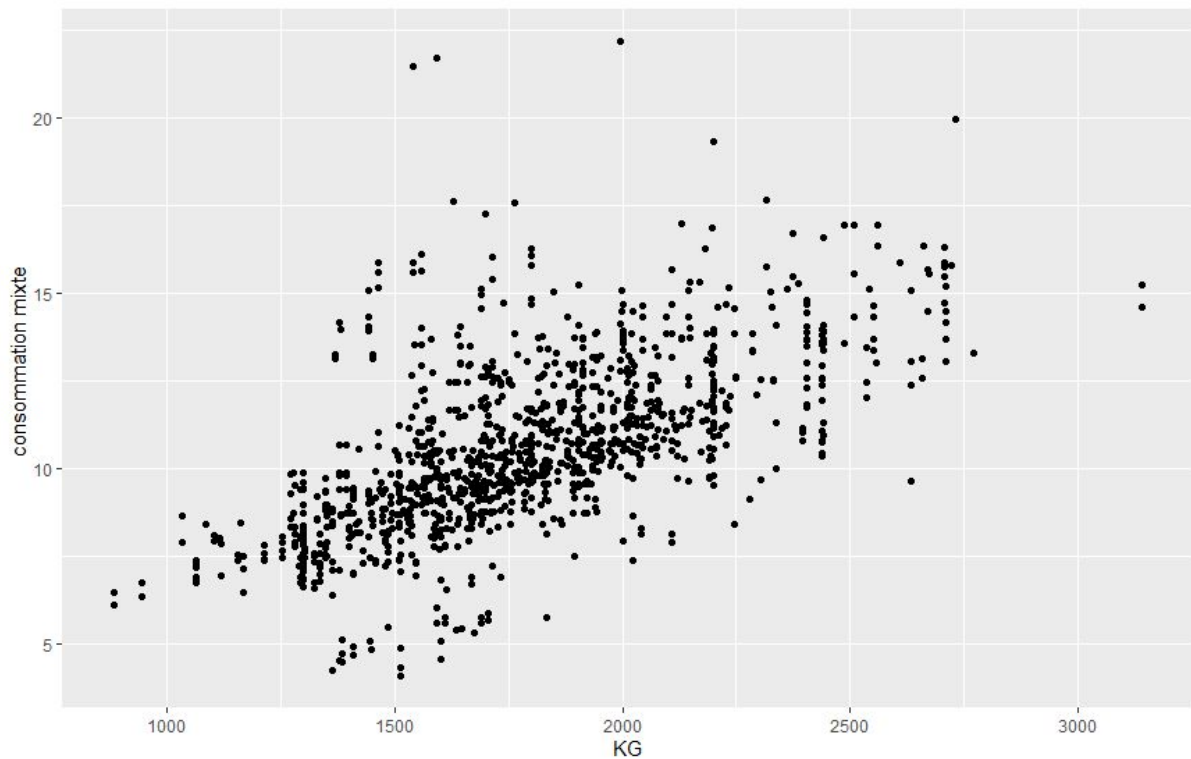
Comme vu précédemment, Les voitures diesels consomment moins, elles rejettent donc moins de Co2

Consommation selon la puissance de la voiture



Comme nous le montre l'illustration, la consommation augmente avec la puissance

Consommation selon la masse de la voiture



Il semble logique, pour des raisons mécaniques, que la consommation augmente avec la masse.

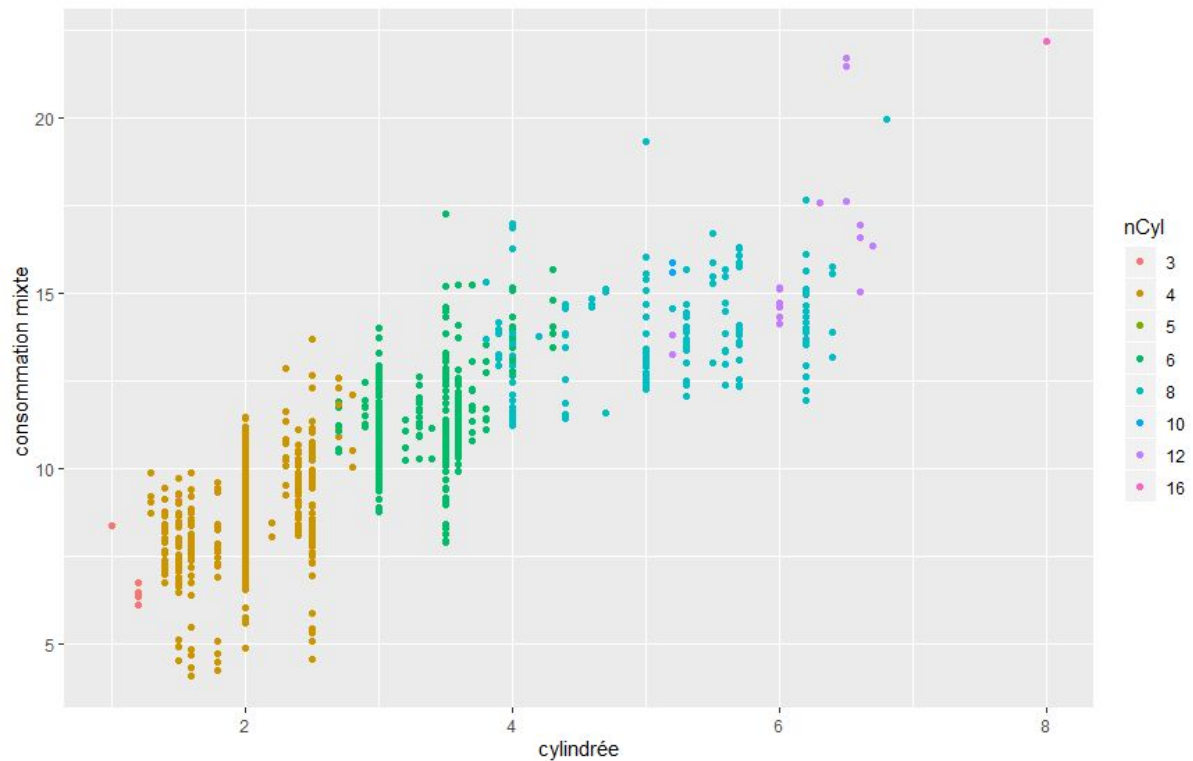
L'illustration nous montre l'allure de cette évolution.

Consommation, cylindrée et nombre de cylindres

Techniquement, le nombre de cylindres augmente un peu la consommation, car moins de couple à bas régime, plus de pièces en mouvement, plus de frottements ...

Un plus gros moteur consomme plus, il faut plus de carburant pour remplir les cylindres.

Pour les voitures, plus gros est le moteur et plus il y a de cylindres, pour des raisons de confort et de facilité de conduite.



Ici, l'illustration est un graphe qui nous montre effectivement une augmentation de la consommation lorsque la cylindrée (en litre) est plus importante, et l'allure de cette progression.

Le nombre de cylindres étant lié à la cylindrée, nous allons plutôt nous intéresser à la cylindrée voir au volume par cylindre.

Consommation selon un sous échantillon différencié par la taille du moteur

Nous allons diviser notre échantillon en deux, selon la taille du moteur en s'inspirant d'un article d'Eurostat « passengers cars in the UE » [2] , pour avoir une meilleure vision.

Nous allons donc avoir deux groupes :

- Petite cylindrée $\leq 2000\text{cc}$ (35% de notre base de données)
- Grosse cylindrée $> 2000\text{cc}$ (65 % de notre base de donnée)

Groupe	mean	sd
$\leq 2000\text{cc}$	8,58	1,31

> 2000cc	11,83	2,17
----------	-------	------

Tableau 2: Moyenne et variance de la consommation mixte en fonction des groupes

D'après ce que nous avons vu jusqu'ici, il n'est pas étonnant de voir sur ce tableau que le groupe avec la plus faible consommation est le groupe des petites cylindrées ($\leq 2000\text{cc}$), ce groupe est censé représenter la majorité du parc automobile (environ 89% du parc automobile européen en 2017 [2]) .

Mais en faisant de la sorte, la taille des échantillons est réduite. Le groupe petite cylindrée ne représente que 35% de notre échantillon, soit 448 voitures.

Différence de consommation entre les différents types de transmission:

Nous allons comparer 5 types de transmission différents :

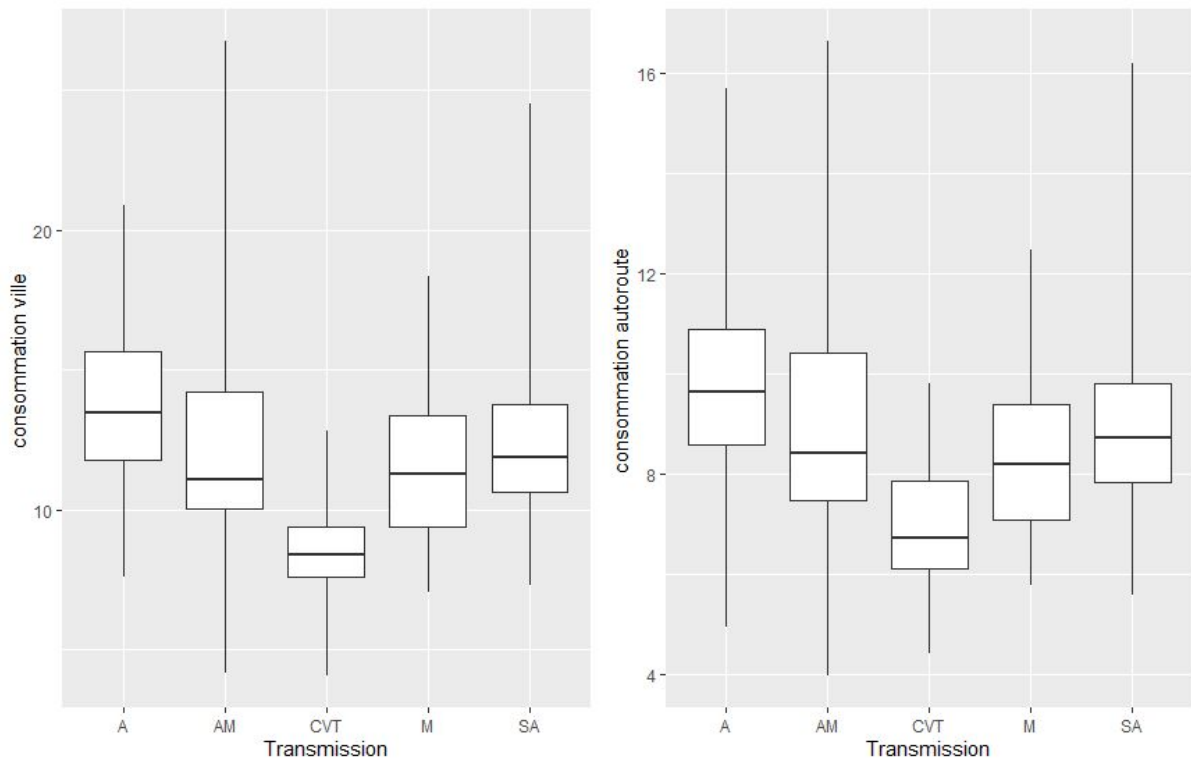
Manuelle (M)

Automatique (A)

Automatique Manuelle (AM)

Semi-automatique (SA)

Variable continue (CVT)



Transmission	mean	sd
A	11,88	2,09
SA	10,79	2,14
AM	10,68	3,11
M	10,11	2,15
CVT	7,1	1,59

Les boîtes de vitesses automatiques classiques ne permettent pas d'avoir un bon contrôle sur la consommation de sa voiture, car elle ne peuvent pas s'adapter à toutes les situations (légers sous et sur régimes), même si cette sur-consommation s'est réduite au fil des années. [3]

Les boîtes de vitesse semi-automatiques réduisent la consommation par rapport à une boîte automatique classique mais pas autant que la boîte manuelle, où le conducteur a entièrement le contrôle sur le régime moteur de sa voiture.

Pour les boîtes CVT, il n'y a pas de vitesse fixe, mais une vitesse variable, gérée par un ordinateur afin d'avoir en permanence un couple et un régime optimaux. Mais il est aussi

plus coûteux à l'achat et à l'entretien qu'une boîte mécanique classique et engendre des avis très disparates quant au confort de conduite. [3] [4]

Mais il faut prendre en compte que les transmissions variables continues ne peuvent être installées que sur des voitures peu puissantes (à cause du couple exercé sur la courroie) ou non sportives et sont souvent utilisées sur des voitures hybrides, ce qui explique une moyenne aussi basse dans notre échantillon : en effet, ici, 87 % des voitures équipées de cette transmission sont dans le groupe petite cylindrée, les voitures qui consomment le moins, donc.

Les boîtes de vitesses « manuelle automatique » ou boîtes de vitesses robotisées ont la même architecture qu'une boîte de vitesse manuelle, mais ce n'est pas le conducteur qui embraye et change la vitesse mais un « robot » aussi gérée par un ordinateur pour passer les vitesses au rapport optimal. Cette boîte de vitesse est censé donner une consommation aussi faible voir plus faible qu'une boîte manuelle en gardant le confort d'une boîte automatique. [3]

Dans notre échantillon, 87% des voitures équipées de boîte de vitesses manuelle automatique sont dans le groupe grosse cylindrées, mal grès une moyenne relativement basse dans le tableau.

Différence de consommation entre les différents types d'injection:

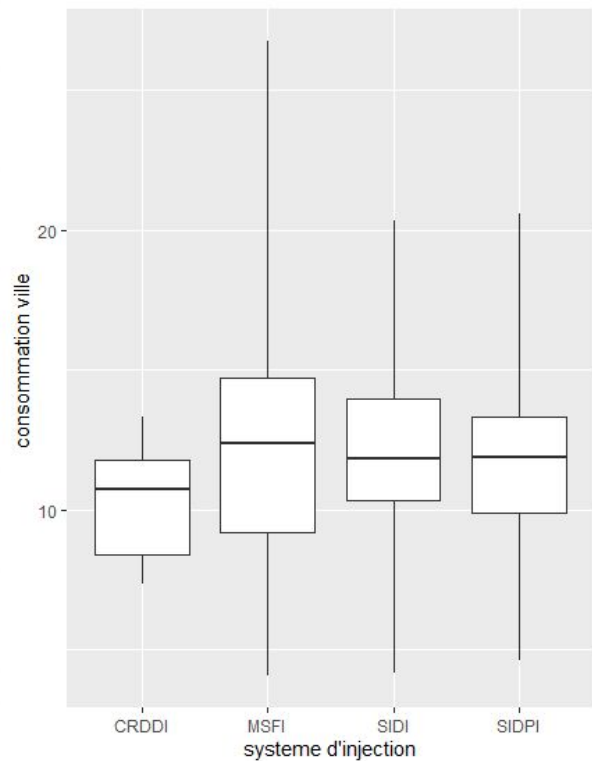
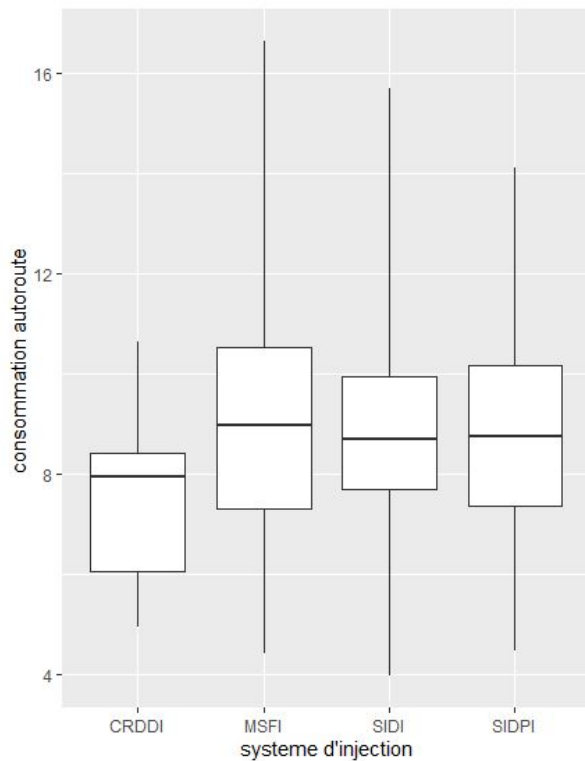
Nous allons comparer 4 types d'injection différents :

Injection directe (SIDI)

Injection directe & indirecte (SIDPI)

Injection groupée et séquentielle (MSFI)

Injection directe à rampe commune (CRDDI)



Système d'injection	mean	sd
CRDDI	8,91	1,62
MSFI	10,89	3,22
SIDI	10,71	2,2
SIDPI	10,43	2,7

Le système d'injection directe à rampe commune concerne uniquement les voitures Diesel, nous allons donc nous intéresser aux trois autres systèmes.

On remarque que la différence est faible. Chacun de ces trois systèmes est relativement récent et améliore la consommation à sa manière par rapport à des systèmes plus anciens (comme l'injection indirecte voire même l'alimentation par carburateur). **[8]**

La différence pourrait se faire à long terme ;

En effet :

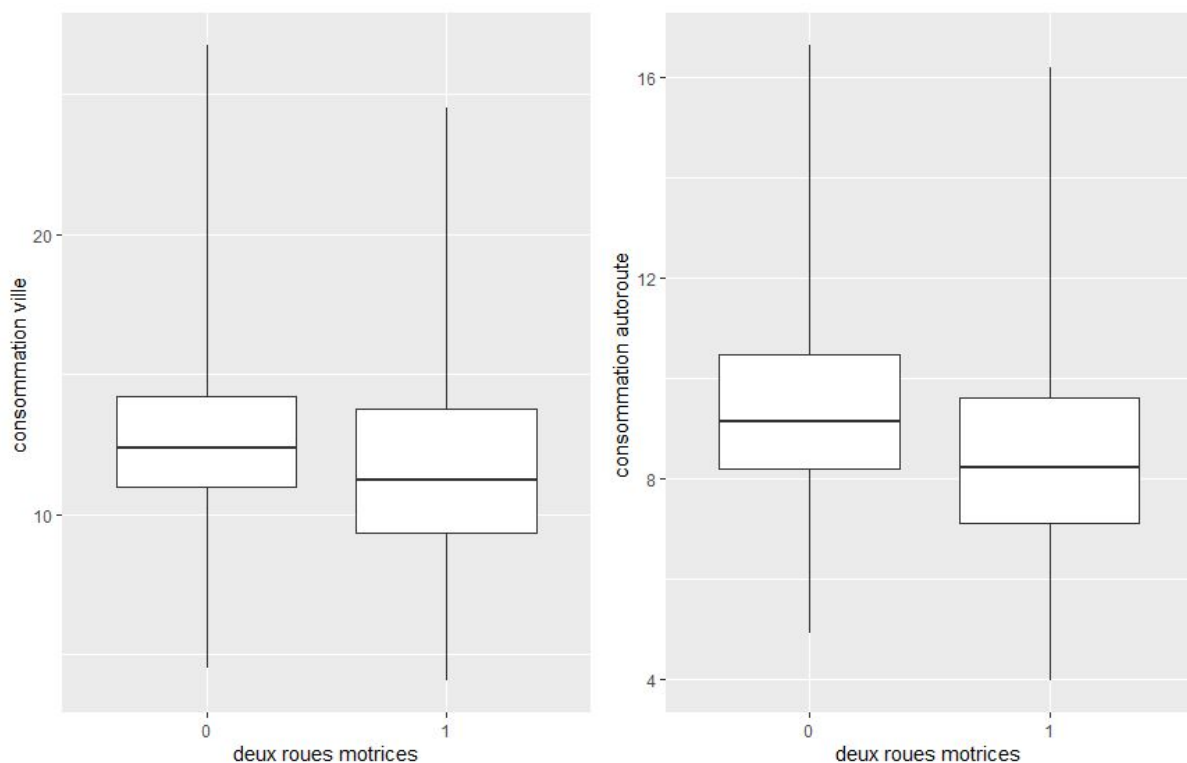
L'injection directe améliore le rendement par rapport à l'injection indirecte car elle aide le calculateur à mieux gérer le mélange, mais émet plus de particules fines nocives et abîme l'intérieur du cylindre avec des dépôts de carbone avec le temps (ce qui rendra le moteur moins efficace) car la combustion est moins bonne. [5] [6]

Combiner les deux systèmes permet d'en tirer les avantages et réduire les inconvénients, mais augmente le coût de fabrication car plus de contraintes. [7]

L'injection groupée et séquentielle (qui est un type d'injection indirect) réduit légèrement la consommation par rapport à l'injection indirecte simple mais pas autant que l'injection directe. [8] [9]

Consommation selon la motricité des roues.

Nous allons comparer la différence de consommation entre les voitures à deux roues motrices et à quatre roues motrices.



Motricité	Mean	sd
Deux roues	10,19	2,54
Quatre roues	11,26	2,23

La transmission intégrale augmente la consommation, car : plus de poids, plus de pièces en mouvement, plus de frottements ...

Même si les nouvelles technologies permettent de réduire ce coût, utiliser deux roues motrices fait baisser la consommation.

Consommation selon la suralimentation.

Nous allons comparer les moteurs suivants :

- Atmosphérique (NA)
- Avec turbocompresseur (TC)
- Avec compresseur (SC)

Initialement, l'ajout d'un système de suralimentation se faisait pour augmenter la puissance du moteur : en comprimant l'air, on peut augmenter la quantité d'oxygène pour un même volume.

Mais dernièrement, de plus en plus de constructeurs utilisent le « downsizing », c'est à dire réduire la cylindrée des moteurs pour faire baisser la consommation et y ajouter un turbocompresseur pour ne pas perdre de puissance, le turbocompresseur ayant pour fonction d'améliorer le rendement du moteur sans altérer sa vitesse de rotation. **[5] [6] [10]**

Nous allons donc comparer cette différence de consommation entre les deux groupes cités plus haut.

mean	≤ 2000cc	> 2000cc
NA	7,39	11,69
TC	8,98	11,8
SC	0	13,18

Tableau 6: Moyenne de consommation mixte selon la suralimentation et le groupe

<i>sd</i>	$\leq 2000cc$	$> 2000cc$
<i>NA</i>	1,39	2,37
<i>TC</i>	0,52	1,85
<i>SC</i>	0	2,17

Tableau 7: Écart type de la consommation mixte selon la suralimentation et le groupe

Grâce au tableau des moyennes, nous pouvons voir pour notre échantillon que la moyenne de consommation est plus faible avec un moteur atmosphérique pour le groupes Petite cylindré alors que la moyenne de consommation avec un moteur turbo-compressé pour le groupe grosse cylindré ne baisse que très légèrement. et qu'un moteur équipé d'un compresseur semble plus consommer.

Consommation avec le système Stop & Start.

Ce système arrête le moteur lorsque la voiture est à l'arrêt et au point mort. Il a été conçu pour baisser la consommation. **[11]**

Nous allons voir les moyennes de consommations en ville et sur autoroute avec et sans ce système de notre échantillon.

mean	Ville	Autoroute	Mixte
sans	12,3	8,93	10,78
avec	12	8,86	10,58

sd	Ville	Autoroute	Mixte
sans	3,24	2,03	2,66
avec	2,77	1,78	2,26

Le tableau des moyennes nous montre que ce système semble être plus efficace en ville. Ce qui est normal, la voiture est plus souvent à l'arrêt en ville. Mais la différence entre les deux moyennes en ville est très faible (0,3 litres)

Autres technologies.

Il nous restent encore deux technologies :

- Le système de désactivation de cylindre, qui désactive un ou plusieurs cylindres lorsqu'on sollicite peu le moteur, pour économiser du carburant [12]
Mais nous n'avons que trop peu de voitures équipées de ce système pour pouvoir l'étudier
- Le système de distribution variable, qui contrôle le temps d'ouverture d'une soupape et son espace d'ouverture selon la sollicitation du moteur, qui permet aussi d'économiser du carburant
Mais nous n'avons que trop de voitures qui ne sont pas équipées de ce système pour pouvoir l'étudier.

IV Modèle et régressions

Grâce à l'analyse préventive de notre base de données nous avons pu trier et transformer les variables qui étaient, et nous semblaient, les plus pertinentes.

Variables étudiées

EngCyl: le volume par cylindre en litre, nous avons tester cette variable car elle nous semblait intéressante et nous avons trouvé qu'elle était significative, nous avons donc décider de la garder, mais elle remplacera le volume du moteur et le nombre de cylindre.

Truck: Qualification donnée à une voiture selon certaine dimension du chassie c'est variable booléenne : 1 si truck , 0 sinon

Essence: Variable qualitative qui indique si la voiture tourne avec de l'essence ou du Diesel : 1 si essence, 0 sinon

Manual: Variable qualitative qui indique si la voiture à une transmission manuelle ou non : 1 si manuelle, 0 sinon

AM.all: Variable indicatrice qui indique si la transmission de la voiture est manuelle automatique ou non, 1 si automatique manuelle, 0 sinon. Elle prend en compte les transmissions manuelle automatique avec et sans commandes

CVT.all: Variable indicatrice qui indique si la transmission de la voiture est variable continue ou non, 1 si variable continue, 0 sinon. Elle prend en compte les transmissions variables continues avec et sans commandes

SA: Variable indicatrice qui indique si la transmission de la voiture est semi-automatique ou non, 1 si semi-automatique, 0 sinon

A: Variable indicatrices qui indique si la transmission de la voiture est automatique ou non, 1 si automatique, 0 sinon

Forced.Induction: Variable booléenne qui indique si la voiture est équipée d'une suralimentation ou non : 1 si suralimentation, 0 sinon

TC: Variable indicatrice qui indique si la voiture est équipé d'un turbocompresseur ou non, 1 si turbocompresseur, 0 sinon

SC: Variable indicatrice qui indique si la voiture est équipé d'un compresseur ou non, 1 si compresseur, 0 sinon

Type d'injection:

MSFI: Variable indicatrice qui indique si la voiture est équipé d'un système d'injection MSFI ou non, 1 si MSFI, 0 sinon

SIDI: Variable indicatrice qui indique si la voiture est équipé d'un système d'injection SIDPI ou non, 1 si SIDPI, 0 sinon

CRDDI : Variable indicatrice qui indique si la voiture est équipé d'un système d'injection CRDDI ou non, 1 si CRDDI, 0 sinon

Stop.Start : Variable qualitative qui indique si la voiture est équipée d'un système Stop & Start ou non : 1 si Stop & Start, 0 sinon

Deux.Roues : Variable qualitative qui indique si la transmission de la voiture se fait sur deux roues ou quatre roues : 1 si deux roues, 0 sinon

Présentation du premier modèle:

Pour avoir une idée générale de notre modèle nous sommes passé par une régression préliminaire, en regroupant quelques variables indicatrices pour mesurer un effet plus regroupé de nos variables, et surtout nos variables catégoriques.

Nous passerons ensuite à une analyse plus précise en détachant certaines variables pour pouvoir capturer leur effets isolé sur notre variable dépendante.

Après une transformation logarithmique des variables dépendantes que nous allons étudier, (c'est-à-dire la consommation en ville, la consommation sur l'autoroute et la consommation mixte).

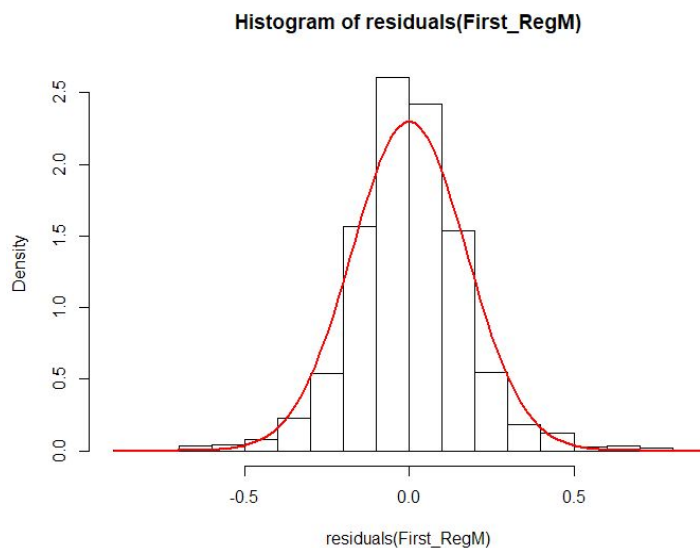
Afin de réduire l'hétéroscédasticité qui se présentera possiblement et pour garder une interprétation possible des différents coefficients des variables dépendantes, nous sommes passé premièrement par un modèle des moindres carrés ordinaires :

$$\log(MixConso) = \beta_0 + \beta_1 EngCyl + \beta_2 Truck + \beta_3 Essence + \beta_4 Manual + \beta_5 Forced.Induction + \beta_6 Stop.Start + \beta_7 Deux.Roues + \varepsilon$$

Pour vérifier que la méthode des Moindres Carrés Ordinaires soit approprié, notre modèle devra répondre aux différentes hypothèses du théorème de Gauss-Markov:

- H1: Linéarité du modèle
- H2: Absence de multicollinéarité significative entre les régresseurs.
- H3: Exogénéité des régresseurs.
- H4: Homoscédasticité.
- H5: Absence de autocorrelation.
- H6: Normalité des erreurs.

Car nous travaillons sur un échantillon assez élevé (1264 observations), nous pouvons présumer une distribution normale des erreurs et une absence d'autocorrélation grâce au central limit theorem. Nous pouvons vérifier cette supposition en modélisant nos résidus sur laquelle on superpose une courbe $N(0, \sigma_{résidus})$.



Nous remarquons une distribution assez normale des résidus et donc nous pouvons accepter l'hypothèse de normalité, et d'absence d'autocorrélation des résidus.

HETEROSCEDASTICITE

Ayant vérifié la normalité des erreurs nous pouvons ensuite vérifier l'homoscédasticité de notre modèle, parce que nous avons transformé la variable dépendante en log et nous avons une dataset assez grand, nous allons passer test de Breusch-Pagan à 7 degrés de libertés.

Après avoir effectué ce test, nous avons un $\chi^2(7) = 96.856$. Nous avons donc une p-value largement inférieure à notre niveau critique de $\alpha=5\%$, donc nous rejetons H_0 , nous aurons besoin de corriger pour l'hétéroscédasticité. Parce que on ne peut pas être sûr que l'hétéroscédasticité soit une fonction linéaire des variables explicatives, pour vérifier la présence d'hétéroscédasticité nous avons effectué un test de White qui nous a conduit au même résultat que le test de Breusch-Pagan.

En ne connaissant pas l'origine de l'hétéroscédasticité, nous pouvons pas faire recours au moindres carrés pondérées, et donc nous allons passer par une correction de la matrice des variances-covariances à travers les écarts-types robustes de White.

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.0899530	0.0436505	24.9700	< 2.2e-16 ***
EngCyl	1.7198282	0.0527727	32.5893	< 2.2e-16 ***
Truck	0.0851088	0.0110677	7.6898	2.962e-14 ***
Essence1	0.2446109	0.0257443	9.5015	< 2.2e-16 ***
Auto	-0.0189375	0.0127048	-1.4906	0.13632
Forced_Induction	0.1832117	0.0163954	11.1746	< 2.2e-16 ***
Stop_Start	-0.0254813	0.0132580	-1.9220	0.05484 .
Deux_Roues	-0.0576416	0.0099803	-5.7756	9.661e-09 ***

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

D'après la statistique $F(7, 1255) = 167.8$ nous rejetons l'hypothèse H_0 :

$$\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$$

Donc notre modèle explique plus de variance dans la variable dépendante qu'un modèle avec juste la constante.

VIF

Pour vérifier l'hypothèse d'absence de multicollinéarité significative entre les régresseurs, nous allons déterminer les variance inflation factors, ils pourront nous indiquer s'il existe une multicollinéarité importante dans notre modèle.

	<i>EngCyl</i>	<i>Truck</i>	<i>Essence</i>	<i>Auto</i>	<i>Forced.Induction</i>	<i>Stop.Start</i>	<i>Deux.Roues</i>
<i>VIF</i>	1.572	1.364	1.058	1.202	2.526	1.856	1.171

Parce que nous avons aucun VIF qui dépasse 3 nous pouvons conclure de ce modèle ne souffre pas trop de problèmes de multicollinéarité.

Nous nous retrouvons avec des résultats cohérent et des d'autres surprenant. Par exemple nous avons la consommation mixte qui augmente de 171,98% si nous augmentons EngCyl de 1, d'après nos statistiques descriptives ceci est un résultat cohérent. Nous avons les voitures essence de cette base de données qui consomment 24,46% que leur contrepartie diesel.

Les résultats surprenant sont que Auto et Stop.Start ne soient pas significatif à 5%, Dans ce premier modèle nous avons plusieurs types de transmission intégré dans la variable Auto, et parce que nous allons par la suite développer ces modèles en ayant un système de régression avec consommation en ville et consommation sur autoroute comme variables dépendantes, nous avons une raison de les préserver afin de pouvoir possiblement observer des effets statistiquement significatif.

2e MODELE

Nous allons développer le modèle surtout en ouvrant les variables catégoriques et tout leur termes et supprimer la variable Truck, sans définition précise attribué à cette variable nous avons pris la décision de la supprimer, afin de préserver juste des variables importantes à l'aspect techniques du véhicule.

Avant de passer au système de régression, nous allons faire une régression préliminaire à partir de consommation mixte pour vérifier l'intégralité du modèle. Par la suite, l'intégralité des hypothèses de linéarité, exogénéité des régresseurs, absence d'autocorrélation et normalité des résidus sont vérifiés. Nous employons des écarts-types robustes par la suite afin de corriger pour l'hétéroscédasticité dans les modèles suivants.

$$\begin{aligned} \log(MixConso) = & \beta_0 + \beta_1 EngCyl + \beta_2 Essence + \beta_3 KG + \beta_4 BHP + \beta_5 AM.all + \beta_6 CVT.all + \beta_7 SA + \beta_8 A \\ & + \beta_9 TC + \beta_{10} SC + \beta_{11} MSFI + \beta_{12} SIDPI + \beta_{13} CRDDI + \beta_{14} Stop.Start \\ & + \beta_{15} Deux.Roues + \varepsilon \end{aligned}$$

D'après une analyse du VIF nous nous apercevons d'une multicollinéarité parfaite entre Essence et CRDDI, c'est-à-dire, que tous les véhicules diesels n'ont qu'un système d'injection unique et propre à eux, le CRDDI. Nous avons donc le choix entre:

- Supprimer les variables binaires associée à la catégorie 'type d'injection'.
- Supprimer les voitures diesels et la variable binaire CRDDI.

En faisant l'hypothèse que le type d'injection est un facteur important à la consommation d'une voiture, et en prenant en compte que l'effectif des voitures diesels n'est que 30, nous avons décidé de supprimer les voitures diesels du modèle ainsi que la variable binaire CRDDI afin de pouvoir approfondir l'étude sur les voitures à essence.

Nous posons donc le nouveau modèle:

$$\begin{aligned} \log(MixConso) = & \beta_0 + \beta_1 EngCyl + \beta_2 KG + \beta_3 BHP + \beta_4 AM.all + \beta_5 CVT.all + \beta_6 SA + \beta_7 A + \beta_8 TC \\ & + \beta_9 SC + \beta_{10} MSFI + \beta_{11} SIDPI + \beta_{12} Stop.Start + \beta_{14} Deux.Roues + \varepsilon \end{aligned}$$

Annex 2a: Résultats de l'estimation

Annex 2b: Significativité des paramètres prenant en compte les écarts-types robustes de White

Nous remarquons une augmentation du R^2 et du R^2 ajusté, notre modèle explique donc plus de variance que notre modèle principale.

Nous remarquons que toutes nos variables non-indicatrice sont statistiquement significatives, *AM.all* et *SIDPI* ne montre aucune significativité mais afin de ne pas changer l'interprétation des autres variables appartenant à leur mêmes variables catégoriques initiales, nous sommes obligé de les préserver.

Nous établissons ensuite nos deux équations:

$$\begin{cases} \log(\text{CityConso}) = \beta_{1,0} + \beta_{1,1}\text{EngCyl} + \beta_{1,2}\text{KG} + \beta_{1,3}\text{BHP} + \beta_{1,4}\text{AMall} + \beta_{1,5}\text{CVTall} + \beta_{1,6}\text{SA} + \beta_{1,7}\text{A} + \beta_{1,8}\text{TC} \\ \quad + \beta_{1,9}\text{SC} + \beta_{1,10}\text{MSFI} + \beta_{1,11}\text{SIDPI} + \beta_{1,12}\text{StopStart} + \beta_{1,13}\text{DeuxRoues} + \varepsilon_1 \\ \log(\text{HwyConso}) = \beta_{2,0} + \beta_{2,1}\text{EngCyl} + \beta_{2,2}\text{KG} + \beta_{2,3}\text{BHP} + \beta_{2,4}\text{AMall} + \beta_{2,5}\text{CVTall} + \beta_{2,6}\text{SA} + \beta_{2,7}\text{A} + \beta_{2,8}\text{TC} \\ \quad + \beta_{2,9}\text{SC} + \beta_{2,10}\text{MSFI} + \beta_{2,11}\text{SIDPI} + \beta_{2,12}\text{StopStart} + \beta_{2,13}\text{DeuxRoues} + \varepsilon_2 \end{cases}$$

Parce que nous souhaitons tester l'inégalité possible de l'effet d'une variable en ville et sur l'autoroute, nous sommes à priori posé la question de si c'était possible de passer par un test standard d'égalité des paramètres, c'est à dire, de tester:

$$\frac{\beta_{1,1} - \beta_{2,1}}{\sigma_{\beta_{1,1}} - \sigma_{\beta_{2,1}}}$$

La complication arrive au calcul de la matrice de variance-covariance car nous pouvons pas calculer $\text{Cov}(\beta_{11} - \beta_{21})$, du faite que les variables proviennent de deux régressions distinctes.

Nous passons donc par un modèle SUR, de la forme:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}$$

Les modèles SUR se modélisent par FGLS car ils nous donnent des estimateurs BLUE, c'est à dire des estimateurs plus efficient qu'à travers OLS dû au fait du théorème de Gauss-Markov, mais parce que nous avons un cas particulier où nos variables explicatives sont identiques dans les deux modèles, nous pouvons passer par une estimation par OLS.

Annex 3: Résultat des estimations.

Nous posons maintenant chaque hypothèse linéaire que nous allons traiter, c'est à dire nous testons les restrictions: $\forall n \in [1, 13], \beta_{1,n} - \beta_{2,n} = 0$, à travers une test de Wald à grand échantillon (asymptotiquement équivalent à une distribution χ^2).

Le test de Wald s'effectue de la manière suivante:

$$W = (R\hat{\beta} - q)^T (R\widehat{\text{COV}}[\hat{\beta}] R^T)^{-1} (R\hat{\beta} - q).$$

Nous comparons ensuite la valeur calculé à la valeur critique du χ^2 à un degré de liberté de 1 (parce que nous testons une restriction) et un niveau critique de $\alpha = 5\%$. Nous allons donc comparer nos valeurs calculé à un $\chi^2 = 3,841$.

Les résultats des différents tests sont les suivants:

Hypothèse H0	$W_{calculé}$	$Prob(W_{calculé} > 3.841)$
$\beta_{1,1} - \beta_{2,1} = 0$	6.3465	0.01176
$\beta_{1,2} - \beta_{2,2} = 0$	2.1563	0.142
$\beta_{1,3} - \beta_{2,3} = 0$	31.068	2.492e-08
$\beta_{1,4} - \beta_{2,4} = 0$	5.8454	0.01562
$\beta_{1,5} - \beta_{2,5} = 0$	25.157	5.285e-07
$\beta_{1,6} - \beta_{2,6} = 0$	0.0563	0.8124
$\beta_{1,7} - \beta_{2,7} = 0$	0.3201	0.5716
$\beta_{1,8} - \beta_{2,8} = 0$	8.9235	0.002815
$\beta_{1,9} - \beta_{2,9} = 0$	1.3237	0.2499
$\beta_{1,10} - \beta_{2,10} = 0$	0.1853	0.6669
$\beta_{1,11} - \beta_{2,11} = 0$	6.998	0.00816
$\beta_{1,12} - \beta_{2,12} = 0$	10.489	0.001195
$\beta_{1,13} - \beta_{2,13} = 0$	1.8394	0.175

En étudiant les résultats des différents tests, nous pouvons en déduire que, dans le modèle SUR, *EngCyl*, *BHP*, *AM.all*, *CVT.all*, *TC*, *SIDPI* et *Stop.Start* ont des effets significativement différents entre la consommation en ville et la consommation sur l'autoroute.

Nous pouvons en conclure qu'à partir modèle SUR:

EngCyl à un effet d'augmenter la consommation en ville de 56,23% en augmentant le ratio de 1, sur l'autoroute la même augmentation augmente la consommation de 36,32%.

BHP à un effet d'augmenter la consommation en ville de 1,17% en augmentant les BHP de 10, sur l'autoroute la même augmentation augmente la consommation de 0,85%.

AM.all à un effet de réduire la consommation en ville de 4,13% par rapport à une transmission manuel, sur l'autoroute une transmission Automated Manual à un effet d'augmenter la consommation de 0,79% par rapport à une voiture à transmission manuel.

CVT.all à un effet de réduire la consommation en ville de 23% par rapport à une transmission manuel, sur l'autoroute une transmission variable continu à un effet de réduire la consommation de 12,98% par rapport à une voiture à transmission manuel.

Parce que 87% des voitures équipé d'une boîte *CVT* sont des petites cylindrées, ceci explique la baisse considérable dans la consommation, car d'après nos statistiques descriptives, les petites cylindrées ont une consommation réduite.

TC à un effet d'augmenter la consommation en ville de 5,21% par rapport à un moteur atmosphérique, sur l'autoroute l'ajout d'un turbocompresseur augmente la consommation de 1.05% par rapport à un moteur atmosphérique.

SIDPI à un effet de réduire la consommation en ville de 2,19% par rapport à une injection de type SIDI, sur l'autoroute une injection de type SIDPI à un effet d'augmenter la consommation de 0,17% par rapport à une voiture avec une injection de type SIDI.

Les voitures comprenant un système *Stop.Start* avaient une consommation en ville réduite de 8,70% par rapport à des voitures comprenant pas ce système, sur l'autoroute, les voitures comprenant un système Stop-Start avaient une consommation réduite de 5,14% par rapport au voitures sans ce système.

Nous observons un effet très significatif dans la réduction de la consommation sur l'autoroute à travers le système Stop-Start, nous pouvons abduire que le producteurs construisant des voitures comprenant des systèmes Stop-Start, et parce que notre échantillon de voiture est comporté des voitures très puissante et pas forcément construite pour réduire la consommation, que les voitures qui comprennent des systèmes Stop-Start sont conçu de base pour être plus efficiente, mais sans multicollinéarité importante, nous pouvons pas en déduire d'où est l'origine de cette baisse de consommation.

Pour l'interprétation des coefficients ayant un coefficient statistiquement non-inégale, nous passons aux coefficients déterminé du modèle mixte (annex 2a).

À partir du modèle OLS avec variable dépendante *MixConso* :

KG à un effet de réduire la consommation de 1,67% en augmentant le poids de 100kg

Les voiture à transmission de type *SA* aient une consommation réduite de 3,59% par rapport à une voiture à transmission manuel

Les voiture à transmission de type *A* aient une consommation réduite de 2,22% par rapport des voitures à transmission manuel.

Les voitures comprenant une équipé d'une suralimentation de type *SC* aient une consommation augmenté de 3,63% par rapport à des voitures à aspiration atmosphérique.

Les voitures comprenant un type d'injection de type *MSFI* aient une consommation réduite de 2,12% par rapport à des voitures à type d'injection *MSFI*.

Les voitures à *Deux.Roues* aient une consommation réduite de 4.52% par rapport à des voitures à 4 roues motrices à temps plein ou partiel.

Conclusion

À travers notre étude nous avons pu déterminer et comparer les caractéristiques techniques affectent la consommation des voitures sorti en 2019 sur le marché des états unis d'Amérique.

D'après notre étude, pour avoir une consommation optimale, une voiture devrait être équipé de :

- un petit moteur pas très puissant et atmosphérique ;
- un système d'injection MSFI
- d'une transmission variable continue sur deux roues
- d'un système Stop & Start

Et d'après nos recherches, elles peuvent aussi s'équiper des systèmes de distribution variable et de désactivation des cylindres.

Mais aussi , d'après nos recherches, le turbocompresseur et les types d'injection SIDPI et SIDI ont un effet positif sur la baisse de consommation.

Il ne faut pas oublier que nous n'avons que 1264 valeurs sur un marché en particulier et sur une année, nos résultats sont possiblement biaisés par ces facteurs.

Il existe aussi d'autre paramètres qui influencent la consommation, comme le comportement du conducteur, la charge de la voiture, la température, un moteur mal réglé ou un terrain difficile.

Bibliographie/Webographie

[1] Europe and the U.S. Share a Lot, Except When It Comes to Cars

<https://www.nytimes.com/2020/03/04/automobiles/european-us-cars-automakers.html>

[2] US Department of Energy, Office of Energy Efficiency and Renewable Energy

https://www.fueleconomy.gov/feg/di_diesels.shtml?fbclid=IwAR0rVdrpbFWyMTXESn0UH6p7iICONn7E2twZ9IUvYqXAAIjto0JtgpfCaM

[3] Eurostat, données sur les voitures Européenes

https://ec.europa.eu/eurostat/statistics-explained/index.php/Passenger_cars_in_the_EU?fbclid=IwAR3IziLJ0a7qWX5ve9vDI7-dNSuGSqjV7nbEdT57x98zU7ihNf0zrOW-BGk#Small_petrol_engines_more_common_than_medium-sized_and_large_engines

[4] Cost, Effectiveness, and Deployment of Fuel Economy Technologies for Light-Duty Vehicles

<https://www.nap.edu/read/21744/chapter/7>

[5] Transmission Systems: A comparative View, Gunter Jürgens

https://www.schaeffler.com/remotemedien/media/_shared_media/08_media_library/01_publications/schaeffler_2/symposia_1/downloads_11/6_Transmission_Systems_1.pdf

[6] Spark Ignition Gasoline Engines, Chapitre 6

<https://www.nap.edu/read/12924/chapter/6>

[7] Cost, Effectiveness, and Deployment of Fuel Economy Technologies for Light-Duty Vehicles (2015)

<https://www.nap.edu/read/21744/chapter/4>

[8] Caranddriver, Why some cars have both Port and Direct Injection

<https://www.caranddriver.com/news/a15342328/explained-why-some-engines-have-both-port-and-direct-injection/>

[9]Autoportal, What is GDI, FSI, CGI, SIDI, Direct Injection?

<https://autoportal.com/articles/what-is-gdi-fsi-cgi-sidi-direct-injection-6965.html>

[10] CNet, What's So Great About Direct Injection Abc's of Car Tech

<https://www.cnet.com/roadshow/news/whats-so-great-about-direct-injection-abcs-of-car-tech/>

[11] US Departement of Energy, Office of Energy Efficiency and Renewable Energy

Turbocharge with engine downsizing

https://www.fueleconomy.gov/feg/tech_engine_more.shtml#turbo

[12] US Departement of Energy, Office of Energy Efficiency and Renewable Energy

Hybrid Technologies and Start-Stop Systems

<https://www.fueleconomy.gov/feg/tech-other.shtml#ss>

[13] US Departement of Energy, Office of Energy Efficiency and Renewable Energy

Cylinder Deactivation

https://www.fueleconomy.gov/feg/tech_engine_more.shtml

Systèmes SUR en R:

<https://cran.r-project.org/web/packages/systemfit/vignettes/systemfit.pdf>

Econometric Analysis, William H. Greene, Eight Edition

Chapitre 10

ANNEXE

Annex 1:

Data Mining Script

```
from bs4 import BeautifulSoup as bs
import requests
import re
import numpy as np
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
import time
from selenium.common.exceptions import NoSuchElementException

data =
pd.read_excel('C:\\Users\\Patrice\\Desktop\\Classes\\PIR\\Dataset_V1.2.xlsx')
data.head()
carlist=[]

for name in data.DivCar:
    dict_name = {'name':name, 'kg':float(0), 'bhp':float(0)}
    carlist.append(dict_name)

#définir le webdriver et la page web qui va nous intéresser

driver = webdriver.Chrome('C:\\Users\\Patrice\\Desktop\\chromedriver.exe')
driver.get('https://www.carfolio.com/search/');

#pour chaque voiture dans le dictionnaire, on va faire une recherche
```

```

for car in carlist:
    driver.get('https://www.carfolio.com/search/')
    in_y1      = driver.find_element_by_xpath("//div[3]/div/div/input")
    in_y2      = driver.find_element_by_xpath("//input[2]")
    in_name     = driver.find_element_by_xpath("//div/div/input")
    search_but  = driver.find_element_by_xpath("//p[29]/input")
    weight     = []
    bhp        = []

    #on cherche entre 2015 et 2019, et on cherche le nom de la voiture

    in_y1.send_keys('2015')
    in_y2.send_keys('2019')
    in_name.send_keys(car.get('name').lower())
    search_but.click()
    time.sleep(1)

    #cherche s'il y a au moins un résultat, sinon on passe à la prochaine recherche

    try:
        sort_year=driver.find_element_by_css_selector("[class^='longlist ui eight cards']")
        print(1)
    except NoSuchElementException:
        print(0)
        continue

    #si il y a seulement un résultat on peut pas trier par an, d'ailleurs on doit voir si le bouton existe, et puis l'exécuter

    try:
        sort_year=driver.find_element_by_xpath("//a/i").click()
    except NoSuchElementException:
        pass

    #chercher toute les informations différentes qui nous intéressent des résultats

    page = requests.get(driver.current_url)
    soup = bs(page.text, 'html.parser')

    for bhp_result in soup.findAll("abbr", {"title":"brake horsepower"}):
        bhp.append(bhp_result.next_sibling)

    for w_result in soup.findAll("span", {"class":"weight"}):

```

```

weight.append(re.sub("\D","",str(w_result)))

car.update({'bhp':np.mean(list(np.float_(bhp))),
'kg':np.mean(list(np.float_(weight)))})
print(car)

#time.sleep(5)
#search_box = driver.find_element_by_name('q')
#search_box.send_keys('ChromeDriver')
#search_box.submit()

import csv

donelist = carlist

keys = donelist[0].keys()
with open('newcarlist','w') as output_file:
    dict_writer = csv.DictWriter(output_file, keys)
    dict_writer.writeheader()
    dict_writer.writerows(donelist)

```

Annex 2a:

Résultats de l'estimation par OLS du deuxième modèle

```

Residuals:
    Min       1Q   Median       3Q      Max
-0.57417 -0.06057 -0.00107  0.06210  0.45082

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.587e+00  3.040e-02  52.186 < 2e-16 ***
EngCyl       4.835e-01  5.304e-02   9.115 < 2e-16 ***
BHP          1.051e-03  3.787e-05  27.757 < 2e-16 ***
KG           1.673e-04  1.388e-05  12.052 < 2e-16 ***
AM_ALL      -2.286e-02  1.366e-02  -1.673  0.094540 .
CVT_ALL     -1.897e-01  1.437e-02 -13.198 < 2e-16 ***
SA          -3.589e-02  1.095e-02  -3.277  0.001079 **
A           -2.222e-02  1.270e-02  -1.750  0.080401 .
TC           3.645e-02  9.359e-03   3.894  0.000104 ***
SC           3.634e-02  1.758e-02   2.067  0.038898 *
MSFI        2.592e-02  9.421e-03   2.752  0.006020 **
SIDPI       -1.497e-02  1.206e-02  -1.241  0.214726
Stop_Start  -7.292e-02  7.373e-03  -9.890 < 2e-16 ***
Deux_Roues  -4.527e-02  6.751e-03  -6.706  3.05e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1099 on 1219 degrees of freedom
Multiple R-squared:  0.7845,    Adjusted R-squared:  0.7822
F-statistic: 341.3 on 13 and 1219 DF,  p-value: < 2.2e-16

```

Annex 2b:

Significativité des paramètres avec les écarts-types robustes de White

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.5865e+00	3.3935e-02	46.7519	< 2.2e-16	***
EngCyl	4.8348e-01	6.4884e-02	7.4514	1.744e-13	***
BHP	1.0513e-03	5.5496e-05	18.9431	< 2.2e-16	***
KG	1.6728e-04	1.5067e-05	11.1024	< 2.2e-16	***
AM_ALL	-2.2864e-02	1.6724e-02	-1.3671	0.1718333	
CVT_ALL	-1.8970e-01	1.7039e-02	-11.1331	< 2.2e-16	***
SA	-3.5886e-02	9.5735e-03	-3.7485	0.0001862	***
A	-2.2216e-02	1.1321e-02	-1.9623	0.0499492	*
TC	3.6446e-02	1.2597e-02	2.8931	0.0038819	**
SC	3.6342e-02	1.9744e-02	1.8407	0.0659074	.
MSFI	2.5923e-02	9.9888e-03	2.5952	0.0095663	**
SIDPI	-1.4973e-02	1.4586e-02	-1.0265	0.3048413	
Stop_Start	-7.2917e-02	8.5705e-03	-8.5079	< 2.2e-16	***
Deux_Roues	-4.5270e-02	6.6533e-03	-6.8041	1.590e-11	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Annex 3:

Résultats du modèle SUR estimé par OLS

Model Formula: $\log(\text{City_Conso}) \sim \text{EngCyl} + \text{KG} + \text{BHP} + \text{AM_ALL} + \text{CVT_ALL} + \text{SA} + \text{A} + \text{TC} + \text{SC} + \text{MSFI} + \text{SIDPI} + \text{Stop_Start} + \text{Deux_Roues}$

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.66115e+00	3.31134e-02	50.16540	< 2.22e-16	***
EngCyl	5.62389e-01	5.77752e-02	9.73409	< 2.22e-16	***
KG	1.55738e-04	1.51187e-05	10.30099	< 2.22e-16	***
BHP	1.16932e-03	4.12540e-05	28.34431	< 2.22e-16	***
AM_ALL	-4.13372e-02	1.48840e-02	-2.77729	0.0055655	**
CVT_ALL	-2.32452e-01	1.56563e-02	-14.84719	< 2.22e-16	***
SA	-3.73148e-02	1.19285e-02	-3.12821	0.0018004	**
A	-2.56779e-02	1.38290e-02	-1.85681	0.0635791	.
TC	5.21438e-02	1.01945e-02	5.11489	3.6431e-07	***
SC	4.70388e-02	1.91461e-02	2.45683	0.0141554	*
MSFI	2.72064e-02	1.02620e-02	2.65118	0.0081251	**
SIDPI	-2.18643e-02	1.31380e-02	-1.66420	0.0963298	.
Stop_Start	-8.70225e-02	8.03077e-03	-10.83614	< 2.22e-16	***
Deux_Roues	-4.03575e-02	7.35305e-03	-5.48854	4.9259e-08	***

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.119677 on 1219 degrees of freedom

Number of observations: 1233 Degrees of Freedom: 1219

SSR: 17.459133 MSE: 0.014323 Root MSE: 0.119677

Multiple R-Squared: 0.785573 Adjusted R-Squared: 0.783287

Model Formula: $\log(\text{Hwy_Conso}) \sim \text{EngCyl} + \text{KG} + \text{BHP} + \text{AM_ALL} + \text{CVT_ALL} + \text{SA} + \text{A} + \text{TC} + \text{SC} + \text{MSFI} + \text{SIDPI} + \text{Stop_Start} + \text{Deux_Roues}$

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.48736e+00	3.09380e-02	48.07552	< 2.22e-16	***
EngCyl	3.63199e-01	5.39797e-02	6.72843	2.6319e-11	***
KG	1.86120e-04	1.41255e-05	13.17620	< 2.22e-16	***
BHP	8.54628e-04	3.85439e-05	22.17286	< 2.22e-16	***
AM_ALL	7.91061e-03	1.39062e-02	0.56885	0.5695605	
CVT_ALL	-1.24984e-01	1.46278e-02	-8.54428	< 2.22e-16	***
SA	-3.34399e-02	1.11449e-02	-3.00048	0.0027504	**
A	-1.49710e-02	1.29206e-02	-1.15870	0.2468071	
TC	1.04670e-02	9.52480e-03	1.09892	0.2720193	
SC	1.68920e-02	1.78883e-02	0.94430	0.3452023	
MSFI	2.11614e-02	9.58787e-03	2.20710	0.0274925	*
SIDPI	1.67034e-03	1.22750e-02	0.13608	0.8917828	
Stop_Start	-5.14128e-02	7.50320e-03	-6.85212	1.1521e-11	***
Deux_Roues	-5.40055e-02	6.87000e-03	-7.86107	8.4377e-15	***

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.111815 on 1219 degrees of freedom

Number of observations: 1233 Degrees of Freedom: 1219

SSR: 15.240572 MSE: 0.012503 Root MSE: 0.111815

Multiple R-Squared: 0.723743 Adjusted R-Squared: 0.720796