

# Analyse statistique de l'impact du système LAVIA sur l'occurrence des dépassements de vitesse



Stage réalisé par Mai Linh PHAM

Master 2 de statistique

Responsable du master : Mr Deheuvels

Responsable de stage : Mr Saint Pierre

Septembre 2010



# Remerciements

Ce rapport n'aurait pas été possible sans la contribution des personnes à qui je souhaite présenter ma reconnaissance.

J'adresse mes respectueux remerciements à Jacques Ehrlich, directeur du LIVIC, pour m'avoir accueilli au LIVIC.

Je tiens à adresser mes grands et sincères remerciements à mon responsable de stage, Guillaume Saint Pierre, pour sa confiance dans mon travail, sa permanente disponibilité, ses conseils et son aide tout au long de mon stage.

Je tiens aussi à remercier Cindie Andrieu pour tous ses conseils, son aide et sa disponibilité.

Je remercie toute l'équipe du LIVIC de m'avoir offert un environnement propice au travail.

Enfin je remercie mes camarades stagiaires Pierre-Emmanuel, Vincent, Soraya, Antoine et Alix pour avoir contribué à créer une ambiance de travail agréable, pour tous leurs conseils et pour leur aide.



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Présentation du LIVIC</b>	<b>3</b>
2.1	Historique . . . . .	3
2.2	Démarche scientifique du LIVIC . . . . .	3
2.2.1	Perception de l’environnement proche et lointain . . . . .	3
2.2.2	Risque, décision et action . . . . .	4
2.2.3	Simulation, intégration, test et évaluation . . . . .	4
2.3	Coordonnées utiles . . . . .	4
<b>3</b>	<b>Le projet LAVIA</b>	<b>5</b>
3.1	Les dispositifs d’évaluation . . . . .	5
3.2	Le plan expérimental . . . . .	6
3.3	Zone couverte . . . . .	6
3.4	Le dispositif LAVIA . . . . .	7
3.4.1	Limitation active . . . . .	8
3.4.2	Le kickdown . . . . .	8
3.5	Principales conclusions de l’étude LAVIA . . . . .	8
3.6	Avertissement . . . . .	10
3.6.1	Représentativité de la population testée lors de l’expérience LAVIA . . . . .	11
3.6.2	Plan d’expérience . . . . .	11
3.6.3	Autres considérations . . . . .	12
3.7	Données issues de l’expérience LAVIA . . . . .	13
3.7.1	Prétraitement . . . . .	13
3.7.2	Variables et individus . . . . .	14
3.7.3	Variables instrumentées . . . . .	14
<b>4</b>	<b>Détection d’événement de dépassement de vitesse</b>	<b>17</b>
4.1	Définitions et détection d’un dépassement de vitesse . . . . .	17
4.2	Analyse descriptive des résultats . . . . .	20
4.2.1	Le système LAVIA et la durée des dépassements de vitesse volontaires . . . . .	21
4.2.2	Le système LAVIA et la distance parcourue en dépassement de vitesse volontaire . . . . .	21
4.2.3	Le système LAVIA et le nombre de dépassements de vitesse volontaires . . . . .	22
4.2.4	Conclusions . . . . .	24

<b>5</b>	<b>Modèles linéaires généralisés (GLM)</b>	<b>25</b>
5.1	Généralités sur les modèles linéaires généralisés (GLM)	25
5.1.1	Hypothèses et notations	26
5.1.2	Estimation des paramètres $\beta_k$	28
5.2	Cas particulier: la régression logistique	28
<b>6</b>	<b>Différentes approches de l'analyse des données corrélées: équations d'estimation généralisées (GEE) et modèles linéaires généralisés mixtes (GLMM)</b>	<b>31</b>
6.1	Généralités sur les équations d'estimation généralisées (GEE)	31
6.1.1	Equations d'estimation généralisées	31
6.1.2	Spécification de la matrice de corrélation de travail	32
6.2	Généralités sur les modèles linéaires généralisés mixtes (GLMM)	34
6.2.1	Les modèles linéaires mixtes	34
6.2.2	Les modèles linéaires généralisés mixtes	35
<b>7</b>	<b>Applications aux données LAVIA: Modélisation de la probabilité d'occurrence des dépassements de vitesse volontaires</b>	<b>37</b>
7.1	Tirage aléatoire	37
7.2	Analyse descriptive des résultats du tirage aléatoire par tableau de contingence	38
7.2.1	Tableau de contingence de "DepassType2" et "ConsigneLimiteVitesse"	38
7.2.2	Tableau de contingence de "DepassType2" et "ModeLAVIA"	38
7.3	Modélisation de la probabilité d'occurrence des dépassements de vitesse volontaires dans un intervalle de temps de trente secondes	41
7.3.1	Modélisation par régression logistique	41
7.3.2	Modélisation par équations d'estimation généralisées (GEE)	43
7.3.3	Modélisation par modèle linéaire généralisé mixte (GLMM)	45
7.4	Comparaison des modèles	48
7.4.1	Comparaison des Odds Ratios	48
7.4.2	Comparaison de l'estimation des paramètres et des erreurs types	51
7.5	Perspectives	52
7.6	Conclusion	56
<b>8</b>	<b>Applications aux données LAVIA: Analyse de l'impact de tous les modes LAVIA sur le dépassement de vitesse volontaires</b>	<b>57</b>
8.1	Tableau de contingence de "DepassType2" et "ModeLAVIA"	58
8.2	Modélisation de la probabilité d'occurrence des dépassements de vitesse volontaires dans un intervalle de temps de trente secondes avec la variable "ModeLAVIA"	58
8.2.1	Modélisation par régression logistique	60
8.2.2	Modélisation par équations d'estimation généralisées (GEE)	61
8.2.3	Modélisation par modèle linéaire généralisé mixte (GLMM)	62
8.3	Comparaison des modèles	64
8.3.1	Comparaison des Odds Ratios	64
8.3.2	Comparaison de l'estimation des paramètres et des erreurs types	67
8.4	Conclusion	68
<b>9</b>	<b>Conclusion générale</b>	<b>69</b>

A	Estimation des paramètres du modèle logistique à trois variables	73
B	Estimation des paramètres du modèle GEE à trois variables	75
C	Estimation des paramètres des effets fixes du modèle linéaire généralisé mixte à trois variables	77





# Liste des figures

2.1	<i>Les bureaux du LIVIC</i> . . . . .	3
3.1	<i>Représentation schématique de la zone de recueil de données.</i> . . . . .	6
3.2	<i>Représentation schématique du fonctionnement du LAVIA.</i> . . . . .	7
3.3	<i>Exemple de trajet. La courbe bleue représente la vitesse instantanée, tandis que la ligne rouge représente la consigne de limitation de vitesse (50km/h ici) avec un intervalle en pointillés de <math>\pm 4</math>km/h. Les points verts représentent les rapports de boîte de vitesse en cours (échelle de droite). On peut se rendre compte des réactions du conducteur lorsque la limitation devient active. Le conducteur effectue un kickdown au moment de la seconde limitation active du trajet.</i> . . . . .	9
3.4	<i>Répartition des vitesses pratiquées lorsque la vitesse de consigne est de 70km/h. Dans ce cas précis, la moyenne du mode Debrayable est fortement influencée par la présence de trajets effectués à faible vitesse, vraisemblablement en période de fortes congestions. Le report des vitesses induit par le LAVIA n'induit pas forcément une réduction de la vitesse moyenne.</i> . . . . .	10
4.1	<i>Histogrammes du pourcentage de temps passé en dépassement de vitesse sur l'ensemble de tous les trajets considérés selon le mode LAVIA: Neutre et Contraint.</i> . . . . .	18
4.2	<i>Exemple de dépassement de vitesse involontaire. La courbe bleue représente la vitesse instantanée, tandis que les lignes oranges représentent les consignes de limitation de vitesse (130 km/h et 110 km/h ici). L'aire en rouge représente le dépassement de vitesse involontaire (passage de la limitation de vitesse 130 km/h à 110 km/h) qui dure 5 secondes.</i> . . . . .	18
4.3	<i>Histogramme de la durée des dépassements de vitesse sur l'ensemble de tous les trajets considérés.</i> . . . . .	19
4.4	<i>Histogrammes de la durée des dépassements de vitesse volontaires et involontaires pour le mode LAVIA <b>Contraint</b>: les dépassements de vitesse volontaires sont obligatoirement associés à un événement de kickdown.</i> . . . . .	20
4.5	<i>Boîtes à moustaches représentant la durée des dépassements de vitesse volontaires avec/sans le système LAVIA. L'ANOVA détecte une différence significative (<math>DDL1 = 1</math>, <math>DDL2 = 9994</math>, <math>F = 30.60</math>, <math>p\text{-value} &lt; 0.0001</math>) entre deux modes LAVIA Neutre (sans le système LAVIA) et Contraint (avec le système LAVIA).</i> . . . . .	21

4.6	<i>Boîtes à moustaches représentant la distance parcourue en dépassement de vitesse volontaires avec/sans le système LAVIA. L'ANOVA détecte une différence significative (DDL1 = 1, DDL2 = 9994, F = 43.01, p-value &lt; 0.0001) entre deux modes LAVIA Neutre (sans le système LAVIA) et Contraint (avec le système LAVIA).</i> . . . . .	22
4.7	<i>Boîtes à moustaches représentant le nombre de dépassements de vitesse volontaires par km avec/sans le système LAVIA. L'ANOVA détecte une différence significative (DDL1 = 1, DDL2 = 3851, F = 26.91, p-value &lt; 0.0001) entre deux modes LAVIA Neutre (sans le système LAVIA) et Contraint (avec le système LAVIA).</i> . . . . .	23
7.1	<i>Graphique des Odds Ratios bruts et des Odds Ratios de trois modèles (logistique, GEE et linéaire généralisé mixte) pour le dépassement de vitesse volontaire. La barre rouge correspond à la valeur neutre (i.e 1) d'Odds Ratio.</i> . . . . .	50
7.2	<i>Graphique des Odds Ratios de quatre modèles (logistique, GEE avec la structure de corrélation "échangeable", GEE avec la structure de corrélation "non structurée" et linéaire généralisé mixte) pour le dépassement de vitesse volontaire. La barre rouge correspond à la valeur neutre (i.e 1) d'Odds Ratio.</i> . . . . .	54
8.1	<i>Graphique des Odds Ratios bruts et des Odds Ratios de trois modèles (logistique, GEE et linéaire généralisé mixte) à une variable explicative "ModeLAVIA" pour le dépassement de vitesse volontaire. La barre rouge correspond à la valeur neutre (i.e 1) d'Odds Ratio.</i> . . . . .	66

# Liste des tableaux

3.1	<i>Répartition des conducteurs selon la tranche d'âge.</i>	11
3.2	<i>Données LAVIA : Les variables d'intérêt.</i>	15
3.3	<i>Données LAVIA : Les variables de repérage.</i>	16
3.4	<i>Données LAVIA : Les variables de contrôle.</i>	16
4.1	<i>Répartition du nombre de dépassements de vitesse volontaires par km pour l'ensemble des conducteurs avec/sans le système LAVIA.</i>	23
5.1	<i>Les principaux modèles linéaires généralisés.</i>	27
7.1	<i>Tableau de contingence de "DepassType2" et "ConsigneLimiteVitesse" (les données du tirage aléatoire).</i>	39
7.2	<i>Estimation des Odds Ratios (rapports de cotes) pour "ConsigneLimiteVitesse" à partir des données du tirage aléatoire.</i>	39
7.3	<i>Tableau de contingence de "DepassType2" et "ModeLAVIA" (les données du tirage aléatoire).</i>	40
7.4	<i>Estimation de l'Odds Ratio (rapport de cotes) à partir des données du tirage aléatoire: le mode LAVIA Contraint VS le mode LAVIA Neutre pour le dépassement de vitesse volontaire.</i>	40
7.5	<i>Estimation des paramètres du modèle logistique à deux variables explicatives.</i>	42
7.6	<i>Estimation des Odds Ratios (rapports de cotes) du modèle logistique à deux variables explicatives.</i>	42
7.7	<i>Test global des variables explicatives du modèle logistique à deux variables.</i>	42
7.8	<i>Estimation des paramètres du modèle GEE à deux variables explicatives, donnant l'effet du mode LAVIA et des limitations de vitesse sur la probabilité d'avoir un dépassement de vitesse volontaire.</i>	44
7.9	<i>Estimation des Odds Ratios (rapports de cotes) du modèle GEE à deux variables explicatives.</i>	44
7.10	<i>Test global des variables explicatives du modèle GEE à deux variables.</i>	44
7.11	<i>Test global des variables explicatives du modèle GEE à trois variables.</i>	45
7.12	<i>Estimation des paramètres des effets fixes du modèle linéaire généralisé mixte à deux variables explicatives.</i>	47
7.13	<i>Estimation des Odds Ratios (rapports de cotes) du modèle linéaire généralisé mixte à deux variables explicatives.</i>	47
7.14	<i>Test global des effets fixes du modèle linéaire généralisé mixte à deux variables explicatives.</i>	47

7.15	<i>Les Odds Ratios bruts (le tableau de contingence) et les Odds Ratios estimés du modèle logistique pour le dépassement de vitesse volontaire. . . . .</i>	49
7.16	<i>Les Odds Ratios estimés du modèle GEE et du modèle linéaire généralisé mixte pour le dépassement de vitesse volontaire. . . . .</i>	49
7.17	<i>Estimation des paramètres de trois modèles: le modèle logistique, le modèle GEE et le modèle linéaire généralisé mixte (GLMM). . . . .</i>	51
7.18	<i>Estimation des erreurs types de trois modèles: le modèle logistique, le modèle GEE et le modèle linéaire généralisé mixte (GLMM). . . . .</i>	51
7.19	<i>Les critères de sélection du modèle. . . . .</i>	53
7.20	<i>Estimation des paramètres de quatre modèles: le modèle logistique, le modèle GEE avec la structure de corrélation "échangeable", le modèle GEE avec la structure de corrélation "non structurée" et le modèle linéaire généralisé mixte (GLMM). Les valeurs en rouge ne sont pas significativement différentes de 0. . . . .</i>	55
7.21	<i>Estimation des erreurs types de quatre modèles: le modèle logistique, le modèle GEE avec la structure de corrélation "échangeable", le modèle GEE avec la structure de corrélation "non structurée" et le modèle linéaire généralisé mixte (GLMM). Les erreurs types en bleu sont les plus petites. . . . .</i>	55
8.1	<i>Tableau de contingence de "DepassType2" et "ModeLAVIA" (les données du tirage aléatoire contenant quatre modes LAVIA). . . . .</i>	59
8.2	<i>Estimation des Odds Ratios (rapports de cotes) à partir des données du tirage aléatoire contenant quatre modes LAVIA pour le dépassement de vitesse volontaire. . . . .</i>	59
8.3	<i>Estimation des paramètres du modèle logistique à une variable explicative ("ModeLAVIA"). . . . .</i>	60
8.4	<i>Estimation des Odds Ratios (rapports de cotes) du modèle logistique à une variable explicative ("ModeLAVIA"). . . . .</i>	60
8.5	<i>Test global de la variable explicative "ModeLAVIA" du modèle logistique. . . . .</i>	60
8.6	<i>Estimation des paramètres du modèle GEE à une variable explicative ("ModeLAVIA"). . . . .</i>	61
8.7	<i>Estimation des Odds Ratios (rapports de cotes) du modèle GEE à une variable explicative ("ModeLAVIA"). . . . .</i>	62
8.8	<i>Test global de la variable explicative "ModeLAVIA" du modèle GEE. . . . .</i>	62
8.9	<i>Estimation des paramètres du modèle linéaire généralisé mixte à une variable explicative ("ModeLAVIA"). . . . .</i>	63
8.10	<i>Estimation des Odds Ratios (rapports de cotes) du modèle linéaire généralisé mixte à une variable explicative ("ModeLAVIA"). . . . .</i>	63
8.11	<i>Test global de la variable explicative "ModeLAVIA" du modèle linéaire généralisé mixte. . . . .</i>	63
8.12	<i>Les Odds Ratios bruts (le tableau de contingence) et les Odds Ratios estimés du modèle logistique à une variable explicative "ModeLAVIA" pour le dépassement de vitesse volontaire. . . . .</i>	65
8.13	<i>Les Odds Ratios estimés du modèle GEE à une variable explicative "ModeLAVIA" et du modèle linéaire généralisé mixte (GLMM) à une variable explicative "ModeLAVIA". . . . .</i>	65
8.14	<i>Estimation des paramètres de trois modèles (logistique, GEE et linéaire généralisé mixte) à une variable explicative "ModeLAVIA". . . . .</i>	67

8.15	<i>Estimation des erreurs types de trois modèles (logistique, GEE et linéaire généralisé mixte) à une variable explicative "ModeLAVIA". . . . .</i>	67
A.1	<i>Estimation des paramètres du modèle logistique à trois variables explicatives. Le paramètre "ModeLAVIA_Constraint*ConsigneLimiteVitesse_90" n'est pas significatifs. . . . .</i>	73
B.1	<i>Estimation des paramètres du modèle linéaire généralisé (GLM) à trois variables explicatives. Les paramètres estimés ne sont pas tous significatifs. . .</i>	75
C.1	<i>Estimation des paramètres des effets fixes du modèle linéaire généralisé mixte (GLMM) à trois variables explicatives. Le paramètre de l'effet fixe "ModeLAVIA_Constraint*ConsigneLimiteVitesse_90" n'est pas significatifs. . . . .</i>	78



# Chapitre 1

## Introduction

Les activités du LIVIC (<http://www.inrets.fr/ur/livic/>) ont pour finalité l'amélioration des circulations automobiles par le développement de systèmes technologiques d'aide à la conduite. Avant tout déploiement à grande échelle de ces systèmes, il est indispensable d'en évaluer l'acceptabilité individuelle et sociale. Le recueil systématique de données sur les véhicules équipés de ces systèmes et leur analyse statistique permettent d'estimer leur impact sur le comportement du conducteur.

C'est pour atteindre ces objectifs que le LIVIC a entrepris une étude concernant le système LAVIA (Limiteur de vitesse s'Adaptant à la Vitesse Autorisée). Ce projet a concerné 90 conducteurs qui ont chacun testé le système pendant 8 semaines suivant les différents modes de fonctionnement : informatif, actif débrayable et actif non débrayable, avec au préalable deux semaines d'utilisation du véhicule dans un mode neutre, c'est-à-dire sans aucun fonctionnement du système LAVIA. Une importante quantité d'information a été recueillie grâce aux systèmes d'acquisition de données embarqués. Toutes les 480 ms ont été enregistrées des variables telles que vitesse instantanée, accélération, vitesse autorisée, actions du conducteur sur les différentes commandes du véhicule (frein, accélérateur, clignotants, etc.), trajectoires (GPS), etc. Toutes ces informations ont été centralisées dans une base de données qui a fait l'objet de nombreuses analyses statistiques dans les années précédentes, cherchant à étudier l'impact du système sur les vitesses pratiquées et la consommation de carburant (cf. LAVIA consortium [2006] et Saint Pierre [2010]).

Il s'agit dans le présent rapport de présenter une approche complémentaire de l'impact du système LAVIA sur le dépassement de vitesse, sans recommencer l'étude précise des vitesses pratiquées et des dépassements de consigne qui a fait l'objet du rapport de Sébastien Romon (Romon [2006]) dont les principaux résultats seront rappelés au chapitre 3.

Nous nous intéressons au dépassement de vitesse, en particulier à l'impact du système LAVIA sur ce genre d'événements, car il empêche théoriquement les excès de vitesse qui sont, rappelons le, la deuxième cause d'accidents en France, après l'alcool. Les dépassements de vitesse augmentent non seulement le risque d'accident, mais aussi le risque de blessure grave et de décès en cas d'accident.

Nous commencerons par décrire le projet LAVIA et détailler les différentes variables instrumentées et présentes dans les données recueillies. Puis, nous définirons les deux types

(volontaire et involontaire) de dépassement de vitesse. Avec ces définitions, une détection des dépassements de vitesses sera effectuée. Ensuite, une première analyse descriptive des résultats sera présentée afin d'étudier l'impact du système LAVIA sur cet événement. Par la suite, les approches approfondies seront présentées. Les deux chapitres suivants seront consacrés à l'application de ces approches aux données LAVIA afin d'analyser l'impact du système LAVIA de façon approfondie et plus précise. Enfin, nous ponctuerons notre analyse en récapitulant ses résultats.



# Chapitre 2

## Présentation du LIVIC

### 2.1 Historique

Le LIVIC (Laboratoire sur les Interactions Véhicules-Infrastructure-Conducteurs) est une unité mixte de recherche dont les tutelles sont :

- Le LCPC (Laboratoire Central des Ponts et Chaussées)
- L'INRETS (Institut National de Recherche sur les Transports et leur Sécurité)

Le LIVIC a été créé en 1999 et il est composé d'une trentaine de personnes (chercheurs, ingénieurs, techniciens, doctorants). Il dispose d'un réseau de 7km de pistes d'essais et de véhicules instrumentés. Ses activités ont pour finalité l'amélioration du fonctionnement des réseaux routiers par le développement de systèmes technologiques d'aides à la conduite.



FIGURE 2.1 – *Les bureaux du LIVIC*

### 2.2 Démarche scientifique du LIVIC

Les activités de recherche du LIVIC sont divisées en 4 orientations stratégiques :

- Rendre la conduite accessible à tous,
- Accroître la sécurité, la mobilité et réduire l'impact environnemental,
- Développer les « routes automatisées » pour le transport en commun (TC), le transport individuel (TI) et le fret,
- Evaluer les performances, l'usage et l'impact des systèmes.

Pour les mettre en oeuvre le LIVIC est organisé en 3 équipes :

#### 2.2.1 Perception de l'environnement proche et lointain

Sa mission est de développer les techniques de reconstruction de la scène routière dans l'environnement proche et lointain, de situer avec précision les objets mobiles dans cette scène et prévoir leur trajectoire y compris dans des situations environnementales dégradées (brouillard, nuit,...).

### 2.2.2 Risque, décision et action

Sa mission est la modélisation du risque, la définition de trajectoire optimale sous contraintes (interactions, infrastructure, environnement, conducteur), l'observation par capteurs proprioceptifs et la conception d'automate d'assistance à la conduite permettant leur mise en oeuvre dans différents modes de coopération avec le conducteur pouvant aller jusqu'à l'automatisation complète. Les différentes stratégies d'optimisation devront être évaluées aussi du point de vue de leur impact sur le trafic.

### 2.2.3 Simulation, intégration, test et évaluation

Sa mission est le développement des moyens d'essais et des méthodes d'évaluations de la conduite. Entrent dans son champ d'action :

- Le développement de moyens techniques : simulation, bancs de test. Ces systèmes sont principalement fondés sur les produits issus des deux autres équipes du LIVIC mais aussi sur les acquis de projets partenariaux notamment dans le domaine du handicap. Ceci inclus la construction et/ou l'intégration dans les véhicules et les équipements de bord de route de sous ensembles électroniques, le développement et/ou l'intégration logicielle selon les standards en vigueur (par exemple portage MATLAB vers RTMAPS).
- La conception de méthodes et d'outils pour l'évaluation : études méthodologiques, conception de plans d'expérience, réalisation d'expérimentations, recueil et analyse des données (statistiques).

L'équipe est également responsable de la gestion de l'Atelier Véhicule Instrumenté.

## 2.3 Coordonnées utiles

LIVIC

14, route de la Minière - Bâtiment 824 - Satory

78000 Versailles

Tél. : +33 (0)1 40 43 29 01 ou 02 - Fax : +33 (0)1 40 43 29 30

Site internet : <http://www.inrets.fr/ur/livic/>

**Directeur** : Jacques Ehrlich

Tél : +33 (0)1 40 43 29 03

e-mail : [jacques.ehrlich@lcpc.fr](mailto:jacques.ehrlich@lcpc.fr)

**Responsable de stage** : Guillaume Saint Pierre

Tél : +33 (0)1 40 43 29 33

e-mail : [guillaume.saintpierre@lcpc.fr](mailto:guillaume.saintpierre@lcpc.fr)

# Chapitre 3

## Le projet LAVIA

L'objectif du projet LAVIA était relativement ambitieux par rapport aux expérimentations de l'époque, en général cantonnées à des zones limitées le plus souvent urbaines, et à des solutions technologiques parfois rudimentaires. Cela a eu plusieurs conséquences sur le cahier des charges du projet :

- La mise en place d'une zone expérimentale suffisamment vaste, représentative d'un bassin de vie quotidienne ;
- La mobilisation d'un nombre suffisant de véhicules expérimentaux, sur une durée d'expérimentation assez longue, pour bénéficier d'une population statistiquement significative de conducteurs participants à l'expérimentation ;
- L'utilisation d'une technologie offrant les meilleures garanties d'ergonomie et de sécurité pour les conducteurs, même sur des routes à vitesse élevée (autoroutes), où les limiteurs actifs n'avaient pas encore été testés et où un (petit) risque pouvait exister ;
- La nécessité de pouvoir comparer, et donc d'intégrer à l'expérimentation, les deux grands modes de fonctionnement (informatif ou actif) des limiteurs de vitesse.

### 3.1 Les dispositifs d'évaluation

L'expérimentation devait permettre d'éclairer les conditions d'un futur déploiement des limiteurs intelligents en traitant tous les aspects importants. Des conducteurs utilisent dans leurs trajets quotidiens des véhicules équipés du LAVIA, leurs comportements sont observés et leurs attitudes et représentations sont recueillies pour être analysées. Les thèmes d'évaluation retenus portaient sur :

- Tester l'acceptabilité et l'usage du système par les utilisateurs ;
- Evaluer les changements de comportement individuel (réductions de vitesse induites ou écarts par rapport à la vitesse limite) ;
- Mesurer l'efficacité en terme de risque individuel et aussi détecter et apprécier les éventuels effets pervers ;
- Evaluer, mais par simulation, les impacts collectifs globaux sur la sécurité ;
- Estimer la pertinence de la solution technique.

## 3.2 Le plan expérimental

Pour évaluer l'utilité et l'acceptabilité du LAVIA, il a été adopté un plan d'expérience classique. Chaque participant à l'expérimentation conduit un véhicule (Renault ou Peugeot) pendant deux mois : quatre fois deux semaines d'affilée suivant une succession imposée des systèmes : mode neutre, informatif, actif débrayable et actif contraint. Il n'a pas été possible de tirer au hasard les ordres de passage des systèmes après le mode neutre. Cela peut entraîner un effet d'apprentissage mais aussi faciliter la prise en main du limiteur selon une progression naturelle des contraintes imposées par le système dans la conduite. Les participants ont été recrutés sur la base du volontariat et ensuite soumis à un processus de sélection suivant des critères liés à la mobilité automobile et des critères médicaux. 90 conducteurs ont participé à l'expérimentation au lieu des 100 visés initialement. Il a été estimé qu'une taille d'une centaine de conducteurs était un bon compromis en matière de coût/efficacité pour obtenir une variabilité suffisante mais non invalidante, c'est-à-dire contrôlée, en vue de tirer des conclusions fondées sur des analyses statistiques quantitatives. Cette estimation n'a pas été documentée par les initiateurs du projet LAVIA, et on peut penser qu'elle repose sur une connaissance d'expert associée à la taille des effets supposés, c'est-à-dire une estimation de l'impact du système sur les vitesses pratiquées. On peut qualifier l'évaluation de semi quantitative pour montrer qu'on se situe à la frontière entre une évaluation qualitative tirée d'un petit échantillon d'utilisateurs et une évaluation quantitative sur un très grand échantillon.

## 3.3 Zone couverte

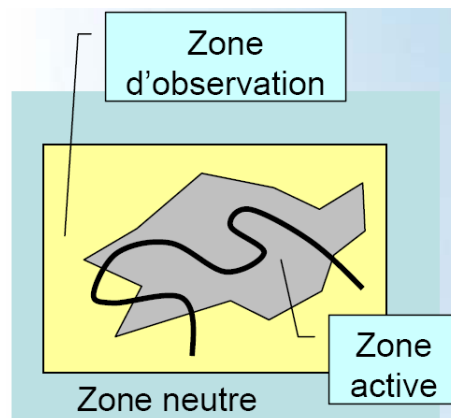


FIGURE 3.1 – Représentation schématique de la zone de recueil de données.

La zone d'expérimentation du système LAVIA est définie par un rectangle dont les coordonnées géographiques sont les suivantes :

- Coin Nord-Est : Latitude :  $48^{\circ} 55'$  ; Longitude :  $2^{\circ} 20'$
- Coin Sud-Ouest : Latitude :  $48^{\circ} 40'$  ; Longitude :  $1^{\circ} 50'$

Elle se décompose en deux sous-zones :

- La zone, dite "active", constituée d'un espace où le système LAVIA est actif et où les données sont recueillies de façon automatique ;

- La zone, dite d'observation, où le système LAVIA est inopérant mais où le système de recueil de données reste actif.

A l'extérieur de ces deux zones, ni le LAVIA, ni le système de recueil de données ne sont actifs. Si la zone d'expérimentation s'étend sur un large périmètre de la partie Ouest de l'Île-de-France, la zone active du système LAVIA est concentrée, quant à elle, dans le Sud-Ouest parisien. Entretenant une proximité immédiate avec la capitale par les autoroutes urbaines, la plupart de cette zone s'étend sur le département des Yvelines. Sur la zone active, un relevé des limitations de vitesse a été fait et l'information codée dans une base de données géographiques dite LUT par la société Navteq. Le kilométrage total de la zone active s'élève à 1289 km.

### 3.4 Le dispositif LAVIA

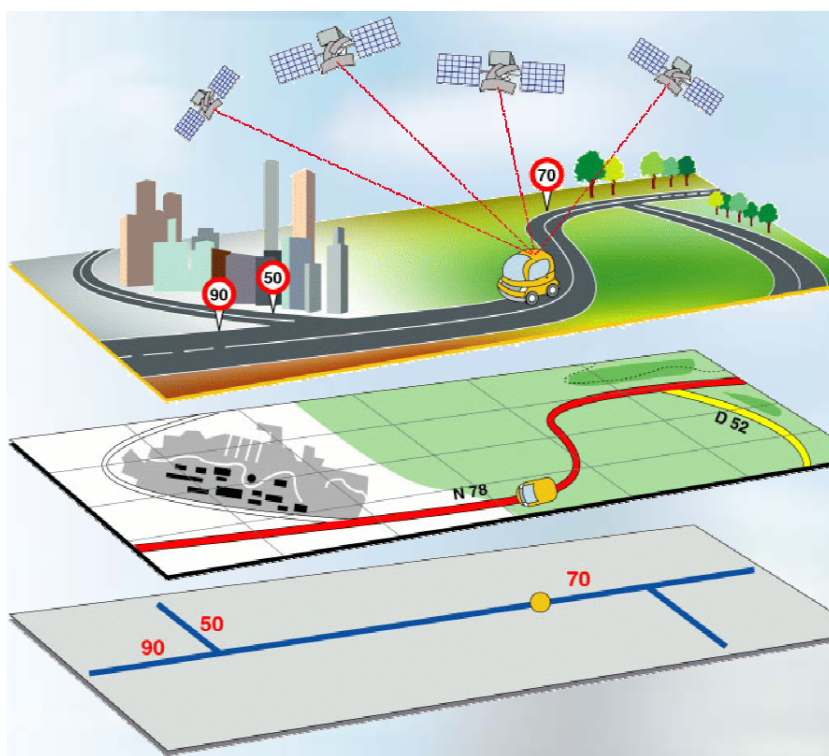


FIGURE 3.2 – Représentation schématique du fonctionnement du LAVIA.

Le principe du système LAVIA est récapitulé à la Figure 3.2. Le limiteur LAVIA, installé sur les véhicules expérimentaux Renault Laguna II et Peugeot 307, comporte :

- un système de localisation GPS ;
- une base de données embarquée couplée à un système de navigation, permettant d'identifier la voie empruntée et la vitesse maximale autorisée ;
- un affichage visuel de la vitesse limite ;
- un limiteur de vitesse (doté d'un découplage d'urgence "kick down").

Trois modes de fonctionnement du limiteur sont testés :

- le mode informatif (affichage de la vitesse limite, avertissement d'un dépassement) ;
- le mode actif débrayable (avec activation, si le conducteur le décide, du limiteur) ;
- le mode actif permanent ou contraint (avec activation permanente du limiteur, sauf usage du "kick down").

Les 90 conducteurs ont utilisé le LAVIA pendant 8 semaines chacun (45 hommes et 45 femmes mais répartition non équilibrée selon les marques de véhicule). Le projet disposait d'une flotte de 20 véhicules composée de 10 véhicules Renault Laguna II et de 10 véhicules Peugeot 307. Au total, cela représente plus de 1000 jours d'expérimentation par mode.

### 3.4.1 Limitation active

Le système se trouve dans l'état « Limitation active » lorsque le système LAVIA limite l'injection de carburant. Dans cet état, le système n'agit pas sur la pédale de frein, seule la pédale d'accélération demeure sans effet. On peut distinguer deux types de situations amenant le système à se trouver à l'état « Limitation active » :

- la vitesse du véhicule est supérieure à la vitesse de consigne suite à une accélération du véhicule. Cette accélération peut avoir deux origines : volontaire avec appui sur la pédale d'accélération, involontaire dans les descentes par exemple.
- le véhicule se trouve en infraction de vitesse suite à une transition de vitesse de consigne vers une limitation plus basse.

### 3.4.2 Le kickdown

On considère qu'un kickdown commence lorsque le conducteur enfonce la pédale d'accélération au delà du point dur. Ceci a pour conséquence la neutralisation du système LAVIA. Un kickdown se termine une fois que la vitesse du véhicule est inférieure à la vitesse de consigne. Pour plus de détails concernant les informations relatives aux kick-downs et au dépassements de vitesse volontaires ou non, on pourra se référer au document de Jacquet [2005] écrit par les ingénieurs de chez Renault. Un exemple de trajet obtenu avec le LAVIA en mode Contraint, avec un exemple de kick-down, est présenté à la figure 3.3.

## 3.5 Principales conclusions de l'étude LAVIA

Le document "Analyse statistique des données LAVIA" écrit par la ZELT (Romon [2006]) a fournit de précieuses informations sur l'impact du LAVIA. L'objectif de ce précédent rapport était centré sur l'analyse des vitesses pratiquées selon les différents modes d'utilisation du LAVIA. Les variables prises en compte étaient celles directement contenues dans les données, et les analyses proposées ne concernaient pas les trajets. Aucun indicateur statistique n'a donc été construit dans ce travail, et le point de vue choisit était celui d'une analyse globale. Nous présentons ici quelques unes des principales conclusions de cette étude.

- Pour une vitesse de consigne fixée, la répartition des vitesses instantanées est fortement influencée par le mode d'utilisation du LAVIA.
- Il y a un report des vitesses élevées vers la vitesse de consigne, mais aussi un report (moins important) des vitesses faibles vers la vitesse de consigne. C'est-à-dire que le conducteur a tendance à "coller" à la consigne puisque le système l'assure de ne pas être en infraction (figure 3.4).

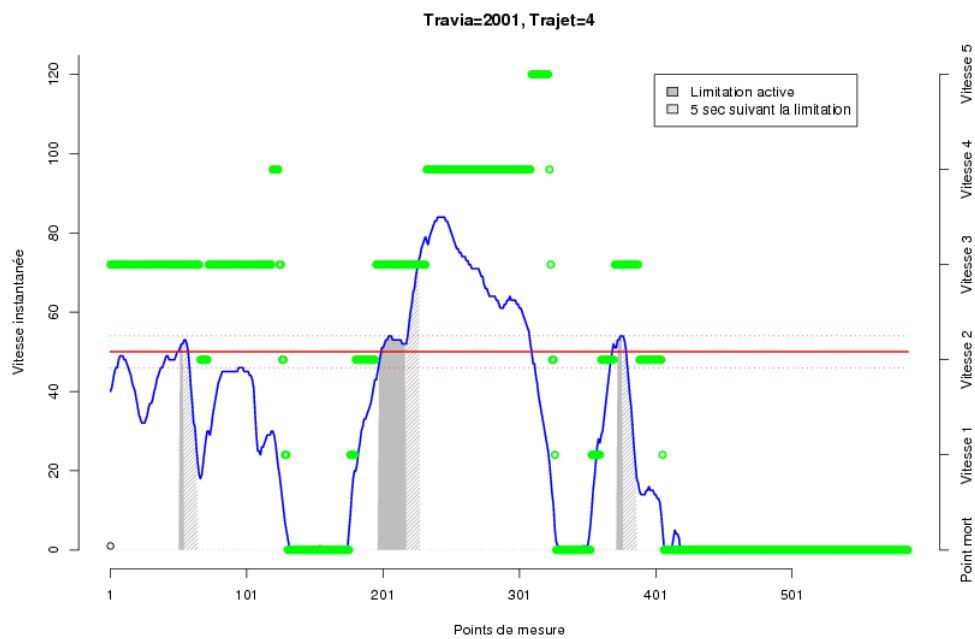


FIGURE 3.3 – Exemple de trajet. La courbe bleue représente la vitesse instantanée, tandis que la ligne rouge représente la consigne de limitation de vitesse (50km/h ici) avec un intervalle en pointillés de  $\pm 4$ km/h. Les points verts représentent les rapports de boîte de vitesse en cours (échelle de droite). On peut se rendre compte des réactions du conducteur lorsque la limitation devient active. Le conducteur effectue un kickdown au moment de la seconde limitation active du trajet.

- Lorsque le temps d'utilisation du système augmente, le conducteur a tendance à utiliser de plus en plus fréquemment le kickdown. On note une augmentation de l'utilisation du kickdown entre les modes debrayables et actifs pour 71% des conducteurs.
- Le mode informatif du LAVIA est le mieux accepté par les utilisateurs. En effet, l'affichage de la vitesse de consigne a un impact positif sur le comportement de l'utilisateur vis à vis de sa vitesse. Il y a un transfert de certaines vitesses élevées vers des vitesses plus faibles et plus proches de la vitesse de consigne, même si ce report est moins important que pour les modes actifs (cf. figure 3.4).
- Les modes actifs du LAVIA ont une influence positive sur les dépassements des vitesses de consignes.

Le lecteur pourra trouver de plus amples informations sur les conclusions du projet LAVIA dans le carnet de route du LAVIA (LAVIA consortium [2006]), téléchargeable sur le site <http://www.lavia.fr>. On pourra trouver par ailleurs une approche de type analyse coûts-bénéfices du système dans Driscoll et al. [2007].

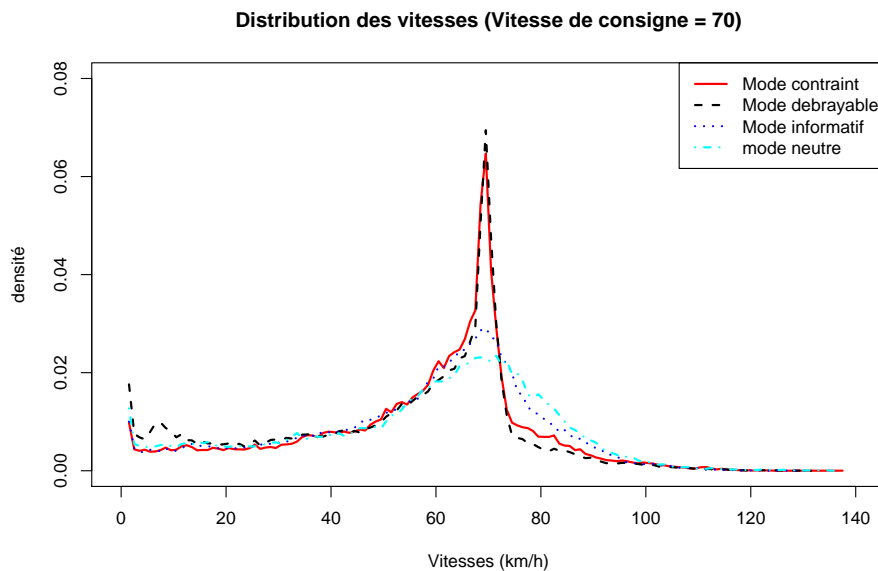


FIGURE 3.4 – Répartition des vitesses pratiquées lorsque la vitesse de consigne est de 70km/h. Dans ce cas précis, la moyenne du mode Debrayable est fortement influencée par la présence de trajets effectués à faible vitesse, vraisemblablement en période de fortes congestions. Le report des vitesses induit par le LAVIA n'induit pas forcément une réduction de la vitesse moyenne.

### 3.6 Avertissement

Le projet LAVIA préfigurait déjà les projets actuellement financés par l'Union Européenne que sont les "Field operational Test" (FOT), c'est-à-dire des évaluations des systèmes d'aides à la conduite à grande échelle et en conditions naturelles. Le meilleur exemple de ce type d'approche est le projet EuroFOT dans lequel est impliqué le LIVIC (UMR LCPC-



INRETS), dont on peut trouver une description détaillée sur le site web qui lui est dédié (<http://www.eurofot-ip.org>). La méthodologie de ces FOT est maintenant bien au point et a fait l'objet d'un projet préliminaire pour la mettre au point (cf. FESTA consortium [2008]). Cependant, lors du lancement du projet LAVIA, peu d'expériences de ce genre avaient été tentées en France. Certains aspects du projets, volontaires ou involontaires, conduisent donc à prendre quelques précautions concernant l'usage qui doit être fait des résultats obtenus.

### 3.6.1 Représentativité de la population testée lors de l'expérience LAVIA

Celle-ci peut être remise en question dans la mesure où la distribution des classes d'âges n'est pas suffisamment identique à celle de la population en âge de rouler (aucune personne âgée de plus de 59 ans par exemple, cf. table 3.1). De même, le choix des véhicules utilisés pour l'expérience s'est porté sur des berlines routières (Laguna II et Peugeot 307), certainement associées à des types de conduites particulières. Le dernier aspect concerne les types de trajets effectués lors de l'expérience, qui correspondent à des petits trajets à faible vitesse et fortement congestionnés, associés à des déplacements de type domicile-travail en banlieue parisienne.

A la lumière de ces limitations, il devient évident qu'inférer les résultats de l'étude LAVIA à la population française n'est pas une chose naturelle. Il faudra donc se garder de voir une quelconque conclusion quand à l'effet global des systèmes ISA (Intelligent Speed Adaptation), à laquelle appartient le LAVIA, sur la consommation de carburant. Néanmoins, les conclusions de l'étude restent valables pour la population étudiée (des conducteurs d'âge moyen, vivants en banlieue parisienne et conduisant une grosse berline) étant donnée l'ampleur de l'expérimentation.

Tranche d'âge	Nombre de conducteurs
20-29 ans	28
30-39 ans	23
40-49 ans	27
50-59 ans	12
Total	90

TABLE 3.1 – Répartition des conducteurs selon la tranche d'âge.

### 3.6.2 Plan d'expérience

Le plan d'expérience utilisé ne permet pas de garantir l'absence d'effets non contrôlés. En effet, la méthodologie d'expérimentation pour ce genre de projets a été précisée dans le projet FESTA, récemment achevé (FESTA consortium [2008]). Il existe pour ce genre d'études deux principales approches :

**Plan intra-sujets :** Cela signifie que les mêmes sujets sont exposés aux différentes conditions expérimentales (ici les différents modes du LAVIA). En outre, ce genre de plan d'expérience nécessite une condition de contrôle, c'est-à-dire une observation en conditions normales (sans le LAVIA par exemple). Il faut de plus contrebalancer les effets,

c'est-à-dire que la moitié des participants doivent commencer l'expérience par la situation normale et finir par l'exposition aux différentes conditions, tandis que l'autre moitié doit faire l'inverse.

**Plan inter-sujets :** Ce genre de plan implique que des groupes différents (et représentatifs) soient constitués par tester chaque condition expérimentale, ainsi que la condition de contrôle. Cela implique que chaque participant ne teste qu'une seule des conditions expérimentales.

Lorsqu'une manipulation est intra-sujets, on ne peut pas attribuer la différence entre deux conditions à des différences entre les sujets puisqu'ils ont chacun été confrontés aux mêmes manipulations expérimentales. A l'inverse, dans un plan inter-sujets, il est possible que l'effet observé (si il y en a un) soit dû à des caractéristiques particulières d'un des groupes testés. Dans le cas du LAVIA, c'est un plan intra-sujets qui a été adopté, c'est-à-dire que tous les participants ont testés tous les modes du LAVIA. Si l'on avait adopté un plan inter-sujets, la différence entre les différents modes aurait pu être attribuée au fait qu'un groupe de conducteurs avait une conduite plus économe, ou bien qu'ils faisaient en moyenne plus de trajets en congestion.

Le plan adopté pour l'expérience LAVIA est donc plutôt adapté excepté qu'une condition n'a pas été respectée : le contrebalancement des effets, c'est-à-dire une inversion de l'ordre d'expérience des différents modes pour la moitié des participants n'a pas pu être respecté. Les différents modes du LAVIA ont été testés dans le même ordre pour tous les participants, ce qui laisse la porte ouverte à un effet non contrôlé d'apprentissage. Les participants peuvent avoir appris à utiliser le système au cours de l'expérience, et il est possible d'attribuer un effet à un des modes du système, alors même qu'il s'agit simplement d'un effet d'apprentissage.

En outre, on sait que la région parisienne est sujette à des périodes de fortes congestions ou à des périodes de grèves de transports. Cela peut induire des conditions de circulations différentes selon les périodes de l'année, qu'il est nécessaire de contrôler dans l'expérience. Cela n'a pas été fait non plus, c'est-à-dire que lorsque plusieurs personnes testaient le mode actif du LAVIA par exemple, personne n'était observé en condition normale sur le même intervalle de temps. Il est donc possible qu'une période de congestion, de travaux ou de grèves, aie influencé les mesures de consommations pour certains sujets.

### 3.6.3 Autres considérations

Il n'a pas été prévu de continuer à enregistrer les données après la fin des tests du limiteur de vitesse, c'est-à-dire que l'impact sur le comportement à long terme n'a pas été étudié. Il devient donc difficile, au vu de cette expérience, d'étudier les éventuels effets pervers du système LAVIA après une longue période d'utilisation.

Au total, les 90 conducteurs de l'expérience LAVIA ont conduit 132679 km, ce qui fait une moyenne d'environ 1500km par conducteur, réparti entre les quatre modes du LAVIA. Il semble que cela soit relativement peu comparé à d'autres études du même genre, notamment celle de MONASH University (Regan et al. [2006]) qui a fait parcourir 16500km à chacun des 23 participants (soit 379500km au total), pour une durée total d'expérimentation de 10 mois avec seulement 15 véhicules. L'expérience LAVIA a donc privilégié la durée, tandis que l'étude de MONASH a privilégié la distance parcourue.

## 3.7 Données issues de l'expérience LAVIA

Les données recueillies pendant l'expérience LAVIA proviennent en grande partie du bus CAN des véhicules utilisés. Ces données concernent le véhicule en lui-même, c'est-à-dire les paramètres du moteur ou du véhicule (régime, vitesse instantanée, accélération, distance parcourue ...), ainsi que les actions du conducteur (freinage, essuie glaces, clignotants, phares). Un autre type d'information est recueillie grâce à l'instrumentation des véhicules avec un GPS. Il s'agit dans ce cas de données de position (Latitude, Longitude) associées à la notion d'arc topologique permettant de connaître le type d'infrastructure. Un certain nombre de paramètres concernant le conducteur lui-même, le mode d'utilisation du LAVIA ainsi que le motif du trajet ont aussi été recueillis.

### 3.7.1 Prétraitement

Cette étape est nécessaire à tout travail de traitement statistique de données. Cela consiste notamment à nettoyer et vérifier la validité des données recueillies. Cependant le pré-traitement peut être très variable et dépend essentiellement de la qualité des données fournies. Voici quelques unes des étapes classiques :

- Vérification et validation des données (recherches d'erreurs, de valeurs aberrantes, de données manquantes).
- Création/Transformation de variables.
- Recodages, transformations entre qualitatif et quantitatif.
- Mise en forme, stockage sous forme de fichiers nettoyés et accessibles facilement.

L'importance de l'étape de pré-traitement est fondamentale, dans la mesure où elle conditionne la qualité et la véracité des résultats obtenus lors de l'analyse proprement dite. Il est en effet nécessaire de s'assurer qu'aucune erreur de mesure ne s'est glissée au sein des 30 millions de lignes des données LAVIA par exemple. Avant de construire de nouvelles variables et indicateurs, il est absolument nécessaire de nettoyer les données brutes. Un premier travail a déjà été effectué par la ZELT, qui a appliqué les traitements suivants :

- Suppression des 200 premiers enregistrements.
- Suppression des 5 derniers enregistrements.
- Suppression des trajets dont la durée est inférieure à 2 minutes et 30 secondes.
- Suppression des trajets dont la distance est inférieure à 200 mètres.
- Correction des sauts d'odomètre supérieurs à 500 mètres.
- Suppression des enregistrements où certaines données sont figées.

Les véhicules peuvent sortir de la zone contrôlée au sein de laquelle la consigne de limitation de vitesse est connue, et le GPS peut aussi présenter des pertes de signaux qui se traduisent par une absence d'information au niveau de plusieurs paramètres. Une situation fréquemment rencontrée est la suivante : `VInstantanee=65535` ; `ConsigneLimiteVitesse=255` ; `NumArcTopo=65535`. Cette condition correspond à la valeur "Hors Zone" de la variable `ZoneLAVIA`. Ces enregistrements ont été supprimés des données préparées pour le traitement. De même, afin d'éviter de considérer des données contenant des erreurs de capteurs, nous avons supprimé les lignes dont le code de la variable `OKNOK` était inférieur à 88.

### 3.7.2 Variables et individus

En statistique, un individu peut être une personne (par exemple un consommateur d'un produit donné) ou un objet (par exemple un appartement à louer à Paris); et on appelle variable l'ensemble des valeurs observées sur les différents individus pour une caractéristique donnée. Une variable renseigne la qualité d'un individu (variable "couleur des yeux" prenant les valeurs bleus ou marrons par exemple) : elle a pour valeurs des modalités; ou bien une quantité mesurée sur cet individu (variable taille mesurée en cm par exemple) : ses valeurs sont des nombres représentant des quantités.

Dans le cas qui nous occupe ici, certaines variables sont constantes tout le long de l'enregistrement (le motif du trajet par exemple), et font donc référence à un individu qui est en fait le trajet complet. Cependant, une variable comme la vitesse instantanée ne fait pas référence à l'individu statistique "trajet", mais bien à l'intervalle de 480ms pendant lequel elle a été mesurée. Cette notion d'individu statistique est sous jacente à l'interprétation des variables. En effet, la totalité des données LAVIA représente approximativement 30 millions de lignes, chaque ligne étant associée à un intervalle de mesure de 480ms. Peut on considérer pour autant que ces lignes correspondent à autant de millions d'individus?

On peut donc se demander ce qu'est un individu statistique dans le cas présent. Celui ci doit être choisi parmi les possibilités suivantes : Un trajet, un conducteur ou un intervalle de temps. Si l'on souhaite étudier le comportement de conduite, on peut faire deux hypothèses différentes et complémentaires :

- **Première hypothèse** : Le comportement dépend essentiellement du conducteur. Un individu statistique est alors un conducteur. On construit donc des tableaux rassemblant les caractéristiques de chaque conducteur (un individu=une ligne de tableau)
- **Deuxième hypothèse** : Le comportement d'un conducteur peut varier selon les trajets. Un individu statistique est alors un Trajet (+ de 14000). On construit donc des tableaux rassemblant les caractéristiques de chaque trajet.
- D'autres hypothèses sont possibles ...

Dans le cadre de ce document, la seconde hypothèse a été privilégiée. C'est-à-dire que nous travaillerons essentiellement sur un tableau rassemblant de multiples indicateurs concernant les trajets.

### 3.7.3 Variables instrumentées

Cette section présente un récapitulatif des variables mesurées lors de l'expérience LAVIA, et présentes dans les données brutes que nous avons eu à disposition. Les différents types de variables ainsi que leur format de stockage sont présentés aux tables 3.2, 3.3, 3.4.

Variables d'intérêt		
Variable	Type	Description
MotifTrajet	Caractère	Variable renseignée par le conducteur au départ du trajet : "Domicile-Travail", "Professionnel" ou "Loisir"
ModeLAVIA	Caractère	Renseigne le mode d'utilisation du LAVIA : "Neutre", "Informatif", "Debrayable" ou "Contraint"
VInstantanee	Réel	Vitesse instantanée du véhicule (Bus CAN)
Acceleration	Réel	Accélération instantanée
RegimeMoteur	Entier	Régime du moteur en tours/min
Distance	Réel	Distance parcourue depuis le point de départ
PosAccelateur	Entier	Pourcentage d'enfoncement de la pédale d'accélérateur
Buteeaccel	Booléen	1 en état de "kickdown", 0 sinon
CptKickdown	Entier	Compteur de "kickdown"
ConsigneLimiteVitesse	Entier	Renseigne la limitation de vitesse en vigueur. 255 en cas d'absence de vitesse consigne.
Latitude	Entier	Projection WGS84 en minutes-secondes
Longitude	Entier	Projection WGS84 en minutes-secondes
NumArcTopo	Entier	Numéro de l'arc topologique considéré
Freinage	Caractère	Usage du frein : "oui" ou "non"
EssuieGlace	Caractère	Usage de l'essuie-glaces : "Inactif" ou "Actif"
Phares	Caractère	Usage de l'éclairage : "Eteint", "Phares" ou "Codes"
Clignotants	Caractère	Usage du clignotant : "Eteint", "gauche" ou "droite"

TABLE 3.2 – Données LAVIA : Les variables d'intérêt.

Variables de repérage		
Variable	Type	Description
NumEnr	Entier	Numéro de l'enregistrement du trajet
Date1/heure1	Date	Date de début du trajet
Date2/heure2	Date	Date et Heure du système
Travia	Entier	Code permettant d'identifier le conducteur
Trajet	Entier	Numéro du trajet
Conducteur	Entier	Code conducteur 1 : Conducteur principal 2 : Conducteur secondaire

TABLE 3.3 – Données LAVIA : Les variables de repérage.

Variables de contrôle		
Variable	Type	Description
VectEtat	Booléen	Variable logique définissant le bon fonctionnement du système. 1 : "Arrêt" ; 2 : "Marche"
Limitation	Caractère	"Limit" lorsque le limiteur de vitesse limite l'injection de carburant. "NoLim" sinon.
ActivationLAVIA	Caractère	"Marche" lorsque le LAVIA est en marche "Arrêt" lorsqu'il est désactivé
DesactivationDefinitive	Entier	Vaut 0 lorsque le LAVIA est désactivé volontairement s'incrémente de 1 à chaque désactivation (cad usage du bouton spécifique)
ZoneLAVIA	Caractère	"Active" lorsque le véhicule est à l'intérieur de la zone "Observation" dans un périmètre autour de la zone LAVIA "HorsZone" à l'extérieur des deux autres zones
OKNOK et MOK	Entier	Variables donnant un "score" de qualité à la ligne d'observation qui permet de repérer les dysfonctionnement des capteurs

TABLE 3.4 – Données LAVIA : Les variables de contrôle.

## Chapitre 4

# Détection d'événement de dépassement de vitesse

En France, la vitesse est la deuxième cause d'accidents, après l'alcool. Le dépassement de vitesse est l'un des principaux problèmes de sécurité routière. C'est également un facteur important de risque d'accident. Les dépassements de vitesse augmentent non seulement le risque d'accident, mais ils aggravent aussi le risque de blessure grave et de décès en cas d'accident. C'est pour cela que nous étudierons l'impact du système LAVIA sur le dépassement de vitesse qui est le sujet de ce chapitre. Nous considérons seulement le mode LAVIA **Neutre** (c'est-à-dire sans aucun fonctionnement du système LAVIA) et le mode LAVIA **Contraint** (c'est-à-dire avec activation permanente du limiteur, sauf usage du "kick down").

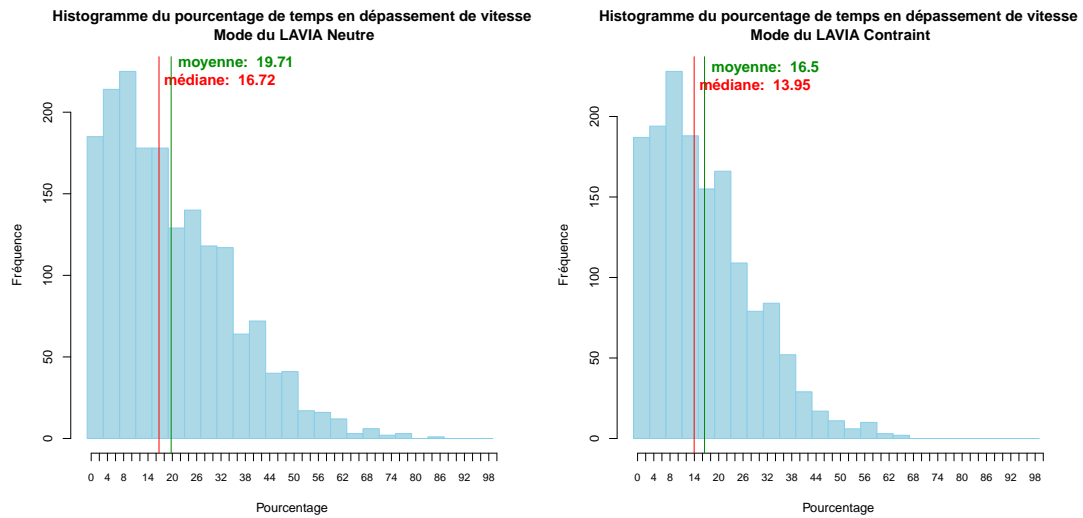
### 4.1 Définitions et détection d'un dépassement de vitesse

Nous appelons un **dépassement de vitesse** (ou un **excès de vitesse**) si la vitesse instantanée est nettement supérieure à la vitesse de consigne. Nous nous sommes fixés, en nous inspirant des travaux de Romon [2006], un seuil de 4 km/h, c'est-à-dire que nous ne considérons pas les dépassements de vitesse dont l'écart entre la vitesse instantanée et la limitation de vitesse reste inférieure à 4 km/h pendant toute la durée du dépassement de vitesse.

Après avoir calculé le pourcentage de temps en dépassement de vitesse (les résultats sont représentés à la figure 4.1), nous remarquons que ce pourcentage est élevé même avec le mode LAVIA Contraint (16.5% pour la moyenne et 13.95% pour la médiane).

Cela est surprenant pour un système qui empêche théoriquement ce genre d'événements. En effet, les dépassements de vitesse ont deux origines : volontaire et involontaire (par exemple, quand on passe à une limitation de vitesse plus faible comme l'illustre la figure 4.2). Notons que les dépassements de vitesse involontaires ont généralement une durée courte. Le seuil distinguant les dépassements de vitesse involontaires et les dépassements de vitesse volontaires est à choisir entre 5 et 10 secondes (cf. EuroFOT consortium [2009], annexe 4).

D'après des analyses descriptives sur la durée des dépassements de vitesse, nous en déduisons le seuil de 10 secondes pour distinguer les dépassements de vitesse involontaires et les dépassements de vitesse volontaires. En effet, les résultats sont présentés à la figure 4.3.



(a) Histogramme du pourcentage de temps en dépassement de vitesse, mode LAVIA Neutre. (b) Histogramme du pourcentage de temps en dépassement de vitesse, mode LAVIA Contraint.

FIGURE 4.1 – Histogrammes du pourcentage de temps passé en dépassement de vitesse sur l'ensemble de tous les trajets considérés selon le mode LAVIA : Neutre et Contraint.

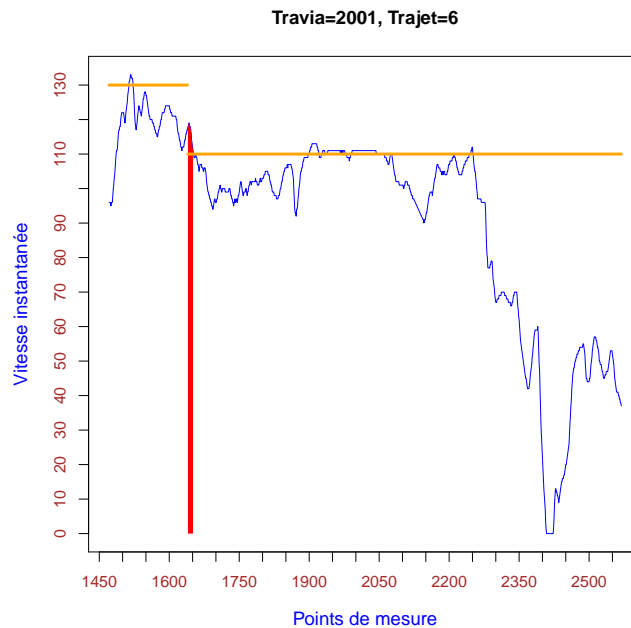


FIGURE 4.2 – Exemple de dépassement de vitesse involontaire. La courbe bleue représente la vitesse instantanée, tandis que les lignes oranges représentent les consignes de limitation de vitesse (130 km/h et 110 km/h ici). L'aire en rouge représente le dépassement de vitesse involontaire (passage de la limitation de vitesse 130 km/h à 110 km/h) qui dure 5 secondes.



En moyenne, un dépassement de vitesse dure 19.38 secondes. D'après la courbe de densité (la courbe en rouge), cette moyenne cache de nombreuses disparités puisqu'il semble nécessaire de distinguer les dépassements involontaires dont la durée n'excède pas 10 secondes (correspondant au pic observé sur la distribution des dépassements de vitesse) et les dépassements volontaires.

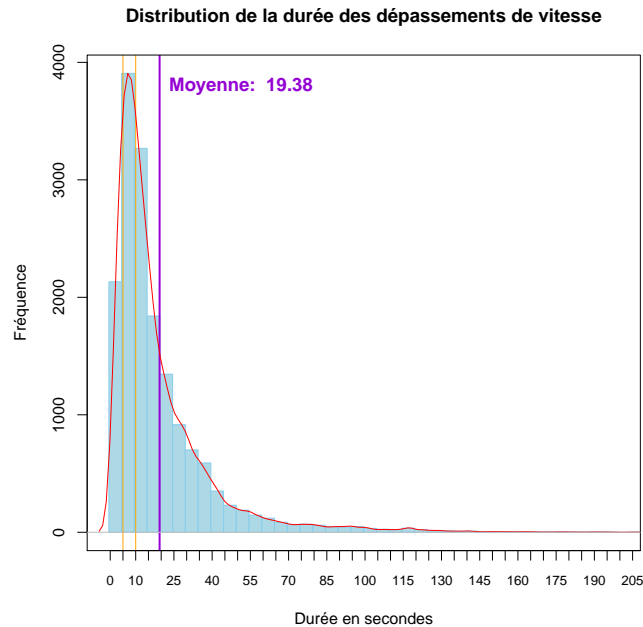
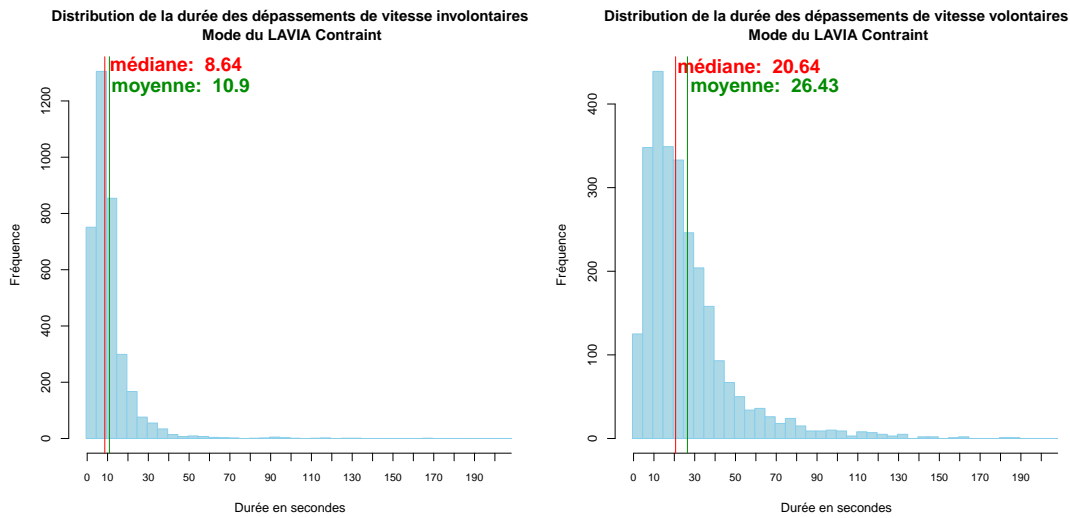


FIGURE 4.3 – Histogramme de la durée des dépassements de vitesse sur l'ensemble de tous les trajets considérés.

En outre, nous disposons, lors de l'utilisation du mode LAVIA Contraint, de la vérité de l'origine de dépassement de vitesse car les dépassements de vitesse volontaires pour ce mode LAVIA sont obligatoirement associés à un événement de kickdown (pression forte sur la pédale d'accélérateur pour débrayer le système (cf. chapitre précédent)). Nous avons donc, avec le mode du LAVIA **Contraint**, une autre définition basée sur l'événement de kickdown pour détecter des dépassements de vitesse involontaires et volontaires :

- Dépassement involontaire (type 1) : Dépassement de plus de 4 km/h de la vitesse autorisée, et non associé à un événement de kickdown.
- Dépassement volontaire (type 2) : Dépassement de plus de 4 km/h de la vitesse autorisée, et associé à un événement de kickdown.

Avec cette définition, pour le mode du LAVIA Contraint, nous détectons des dépassements de vitesse et déterminons leur type à l'aide des variables : "VInstantanee", "Consigne-LimiteVitesse" et "Buteeaccel". La distribution des dépassements de vitesse selon le type de dépassement est représenté à la figure 4.4. En moyenne, les dépassements de vitesse involontaires durent 10.9 secondes (médiane : 8.64) tandis que les dépassements de vitesse volontaires durent 26.43 secondes (médiane : 20.64). Ces résultats nous confirment ainsi le choix du seuil de 10 secondes.



(a) Histogramme des dépassements de vitesse involontaires (b) Histogramme des dépassements de vitesse volontaires.

FIGURE 4.4 – Histogrammes de la durée des dépassements de vitesse volontaires et involontaires pour le mode LAVIA **Contraint** : les dépassements de vitesse volontaires sont obligatoirement associés à un événement de *kickdown*.

Nous choisissons donc le **seuil de 10 secondes** afin de classer les événements correspondant à des dépassements de vitesse dans deux catégories pour les deux modes du LAVIA Neutre et Contraint comme suit :

- **Dépassement involontaire (type 1)** : Dépassement de plus de 4 km/h de la vitesse autorisée, et d'une durée inférieure à 10sec.
- **Dépassement volontaire (type 2)** : Dépassement de plus de 4 km/h de la vitesse autorisée, et d'une durée supérieure à 10sec.

Avec ces définitions, nous détectons l'événement de dépassement de vitesse à l'aide des variables : "VInstantanee", "ConsigneLimiteVitesse". Pour repérer le mode LAVIA, le conducteur, le trajet et pour calculer la durée ainsi que la distance parcourue des dépassements de vitesse, nous avons également besoin des variables : "ModeLAVIA", "Travia", "Trajet", "Date2/heure2" et "Distance".

D'abord, nous détectons des dépassements de vitesse sur l'ensemble des trajets considérés. Nous calculons ensuite la durée, la distance parcourue, l'horaire de début et de fin de chaque événement. A partir de la durée, nous repérons les dépassements de vitesse involontaires (codés 1 et appelés aussi dépassements de vitesse de type 1) et les dépassements de vitesse volontaires (codés 2 et appelés dépassements de vitesse de type 2).

## 4.2 Analyse descriptive des résultats

A partir de la définition précédente, nous pouvons identifier les dépassements de vitesse et étudier leur durée et leur longueur.

Nous ne nous intéressons ici qu'aux dépassements de vitesse volontaires.

### 4.2.1 Le système LAVIA et la durée des dépassements de vitesse volontaires

La figure 4.5 montre clairement l'intérêt du système LAVIA en terme de sécurité. En effet, la durée des dépassements de vitesse volontaires diminue en moyenne de 9.82% entre le mode LAVIA Contraint (avec le système LAVIA) et le mode LAVIA Neutre (sans le système LAVIA).

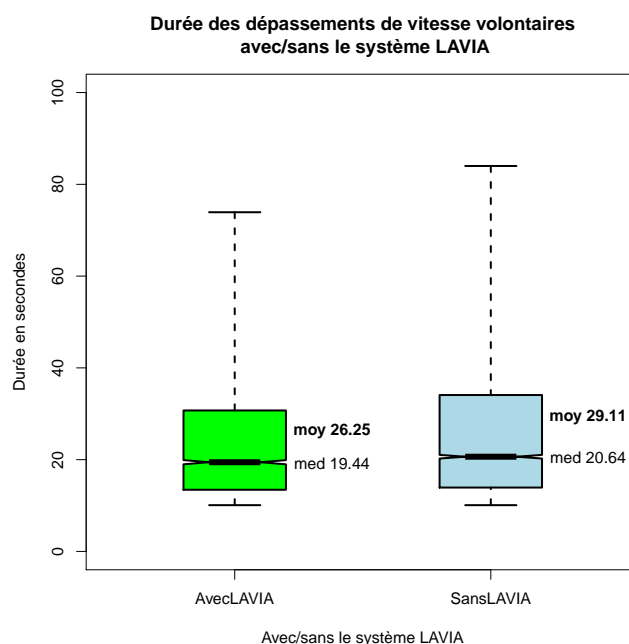


FIGURE 4.5 – Boîtes à moustaches représentant la durée des dépassements de vitesse volontaires avec/sans le système LAVIA. L'ANOVA détecte une différence significative ( $DDL1 = 1$ ,  $DDL2 = 9994$ ,  $F = 30.60$ ,  $p\text{-value} < 0.0001$ ) entre deux modes LAVIA Neutre (sans le système LAVIA) et Contraint (avec le système LAVIA).

### 4.2.2 Le système LAVIA et la distance parcourue en dépassement de vitesse volontaire

D'après la figure 4.6, la distance parcourue en dépassement de vitesse volontaire diminue en moyenne de 16.55% entre le mode LAVIA Contraint et le mode LAVIA Neutre. De plus, la médiane des distances parcourues en excès de vitesse pour les trajets où le mode LAVIA Contraint est utilisé est inférieure à la médiane des distances parcourues en excès de vitesse pour les trajets sans aucun fonctionnement du LAVIA (mode LAVIA Neutre). Ces résultats confirment les résultats du paragraphe précédent.

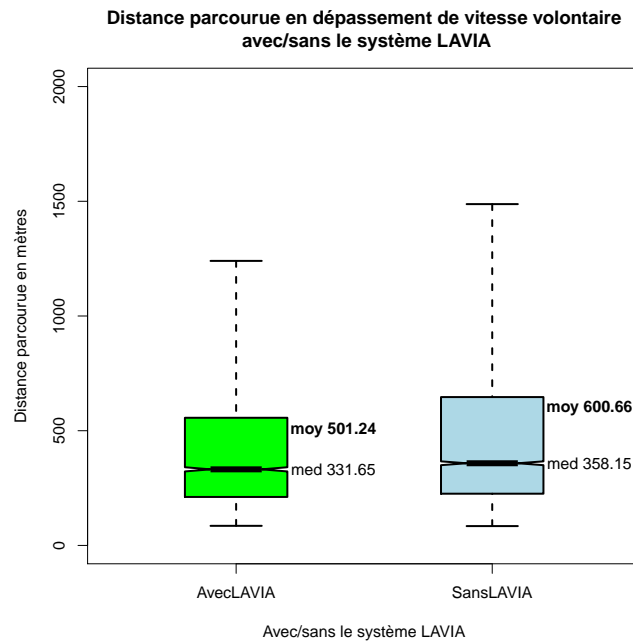


FIGURE 4.6 – Boîtes à moustaches représentant la distance parcourue en dépassement de vitesse volontaires avec/sans le système LAVIA. L'ANOVA détecte une différence significative ( $DDL1 = 1$ ,  $DDL2 = 9994$ ,  $F = 43.01$ ,  $p\text{-value} < 0.0001$ ) entre deux modes LAVIA Neutre (sans le système LAVIA) et Contraint (avec le système LAVIA).

### 4.2.3 Le système LAVIA et le nombre de dépassements de vitesse volontaires

A partir de la définition précédente, nous calculons aussi le nombre de dépassements de vitesse volontaires par km et par trajet pour l'ensemble des trajets considérés.

La figure 4.7 représentant le nombre de dépassements de vitesse volontaires par km et par trajet selon le mode LAVIA montre que le système LAVIA permet un gain en terme de sécurité. En effet, le nombre de dépassements de vitesse volontaires a diminué de 42.86% en moyenne entre le mode LAVIA Contraint et le mode LAVIA Neutre (0.24 de dépassements en moyenne pour le mode LAVIA Contraint et 0.42 pour le mode LAVIA Neutre). Nous observons aussi, pour le mode Contraint, que la médiane du nombre de dépassements de vitesse volontaires par km est égale à 0.15 tandis que pour le mode Neutre, elle est égale à 0.34. Nous constatons également d'après le tableau 4.1 que le nombre maximal de dépassements de vitesse volontaire par km de l'ensemble de conducteurs en mode LAVIA Neutre est de 2.60 alors qu'il n'est que de 2.02 en mode LAVIA Contraint.

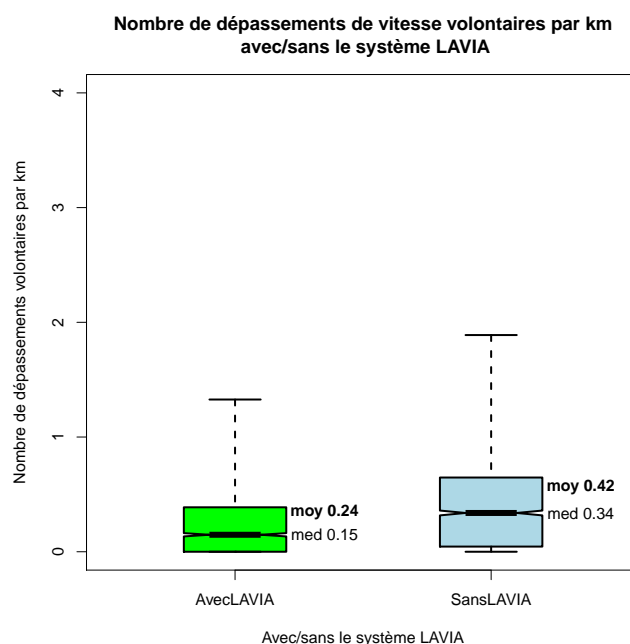


FIGURE 4.7 – Boîtes à moustaches représentant le nombre de dépassements de vitesse volontaires par km avec/sans le système LAVIA. L'ANOVA détecte une différence significative ( $DDL1 = 1$ ,  $DDL2 = 3851$ ,  $F = 26.91$ ,  $p\text{-value} < 0.0001$ ) entre deux modes LAVIA Neutre (sans le système LAVIA) et Contraint (avec le système LAVIA).

	Moyenne	Écart-type	Minimim	Maximum	Médiane
Nombre de dépassement volontaires par km sans le système LAVIA	0.42	0.39	0	2.60	0.34
Nombre de dépassement volontaires par km avec le système LAVIA	0.24	0.30	0	2.02	0.15

TABLE 4.1 – Répartition du nombre de dépassements de vitesse volontaires par km pour l'ensemble des conducteurs avec/sans le système LAVIA.

#### 4.2.4 Conclusions

Les impacts positifs du système LAVIA sur le dépassement de vitesse volontaire apparaissent donc clairement. Il augmente la sécurité routière en diminuant non seulement la durée, la distance parcourue mais aussi la fréquence des dépassements de vitesse volontaires. De plus, ces diminutions sont significatives puisque les ANOVAs détectent une différence significative entre les deux modes LAVIA Neutre (sans le système LAVIA) et Contraint (avec le système LAVIA).

## Chapitre 5

# Modèles linéaires généralisés (GLM)

Nous voulons étudier l'impact du système LAVIA sur le dépassement de vitesse volontaire. Pour cela, nous allons modéliser la probabilité qu'il y ait eu au moins un dépassement de vitesse volontaire dans un intervalle de trente secondes tiré aléatoirement parmi les trajets des modes LAVIA Neutre et Contraint. Nous disposons de plusieurs méthodes d'analyse statistique : l'analyse de tableau de contingence, la régression logistique (cas particulier des modèles linéaires généralisés), les équations d'estimation généralisées (GEE) et le modèle mixte.

L'analyse de tableau de contingence est la méthode la plus simple. En effet, le rapport de chance (Odds Ratio) et la variance correspondante peuvent être calculés facilement. C'est pour cela que nous l'utiliserons pour des analyses descriptives. En revanche, cette méthode ne considère qu'un seul facteur à la fois et ne peut pas s'adresser à l'interaction entre des effets. En outre, l'analyse de tableau de contingence suppose que les observations de chaque sujet sont indépendantes. Cette hypothèse d'indépendance doit aussi être vérifiée pour la régression logistique. Cependant, les données LAVIA ne sont pas indépendantes. En effet, **les observations (trajets) issus d'un même conducteur sont corrélées**. Ce type de données est appelé données longitudinales (mesures répétées dans le temps ou au sein d'un même cluster). Il est donc raisonnable de prendre en considération la corrélation des mesures pour un conducteur donné dans les analyses statistiques. Par conséquent, il n'est pas tout à fait rigoureux d'utiliser l'analyse de tableau de contingence et la régression logistique pour notre modélisation. Les deux principales méthodes pour analyser ce type de données corrélées sont les **équations d'estimation généralisées (GEE)** et les **modèles mixtes**.

Nous présentons d'abord, dans ce chapitre, les modèles linéaires généralisés et la régression logistique, un cas particulier que nous utiliserons. Nous consacrerons ensuite le chapitre suivant aux équations d'estimation généralisées (GEE) et aux modèles mixtes.

### 5.1 Généralités sur les modèles linéaires généralisés (GLM)

Lorsqu'il s'agit de modéliser des phénomènes naturels, les modèles linéaires généralement utilisés n'autorisent que des données indépendantes, de distribution normale et de variance constante. Malgré la prédominance bien justifiée de cette fameuse loi, il est néanmoins possible d'imaginer une foule de situations où supposer une distribution normale ne sera

pas approprié. C'est le cas, par exemple, de relevés sur des durées de vie de matériels, de l'observation du nombre d'individus dans une population ayant telle ou telle caractéristique, ou encore du décompte d'événements rares.

Ainsi, les modèles linéaires classiques ont été étendus à la classe plus large des **modèles linéaires généralisés (GLM)**, permettant de modéliser la distribution d'une variable **réponse**  $Y$  en fonction de variables explicatives ou **prédicteurs**  $X$ , tant que cette distribution fait partie de la famille exponentielle (binomiale, Poisson, normale, gamma, etc.).

Cette famille de modèles englobe le modèle linéaire général (régression multiple, analyse de la variance et analyse de la covariance), le modèle log-linéaire et des techniques de modélisation telles que la régression logistique ou la régression de Poisson. Les modèles linéaires généralisés sont formés de trois composantes :

- La variable de réponse  $Y$ , composante aléatoire à laquelle est associée une loi de probabilité,
- Les variables explicatives  $X_1, \dots, X_K$  utilisées comme prédicteurs dans le modèle, qui définissent, sous forme d'une combinaison linéaire, la composante déterministe,
- Le lien qui décrit la relation fonctionnelle entre la combinaison linéaire des variables  $X_1, \dots, X_K$  et l'espérance mathématique de la variable de réponse  $Y$ .

### 5.1.1 Hypothèses et notations

Notons  $(Y_1, \dots, Y_n)$  un échantillon aléatoire de taille  $n$  de la variable de réponse  $Y$ , les variables aléatoires  $Y_1, \dots, Y_n$  étant supposées indépendantes.

Supposons que pour chaque individu (ou groupe)  $i$  ( $i = 1, \dots, n$ ), nous avons  $n_i$  mesures répétées où  $n_i$  peut varier pour chaque sujet.

Ainsi, pour chaque  $i$  ( $i = 1, \dots, n$ ), nous avons un vecteur de  $n_i$  observations de la

variable réponse  $Y_i = \begin{bmatrix} Y_{i1} \\ \vdots \\ Y_{in_i} \end{bmatrix}$ , ainsi que la matrice de dimension  $n_i \times K$  :

$$X_i = \begin{bmatrix} X'_{i1} \\ \vdots \\ X'_{in_i} \end{bmatrix} = \begin{bmatrix} X_{i11} & \dots & X_{i1K} \\ \vdots & \ddots & \vdots \\ X_{in_i1} & \dots & X_{in_iK} \end{bmatrix}, \text{ où } X_{ij} = \begin{bmatrix} X_{ij1} \\ \vdots \\ X_{ijK} \end{bmatrix} \text{ représente la valeur des}$$

variables explicatives pour l'observation  $j$  de l'individu (ou groupe)  $i$ .

On suppose que la loi des  $Y_{ij}$  appartient à la famille des lois exponentielles. Plus précisément, on suppose que la fonction de densité de  $Y_{ij}$  s'écrit :

$$f(y_{ij}) = \exp \left\{ \frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{a(\phi)} + c(y_{ij}, \phi) \right\}, \quad i = 1, \dots, n; \quad j = 1, \dots, n_i;$$

où le paramètre  $\theta_{ij}$  est un **paramètre canonique** (appelé **paramètre naturel**) inconnu et  $\phi$  est un **paramètre de dispersion**. Les fonctions  $b(\cdot)$  et  $c(\cdot)$  sont spécifiques à chaque distribution, et généralement  $a(\phi) = \frac{\phi}{\omega_{ij}}$  où la valeur de  $\phi$  reste constante pour toutes les observations tandis que  $\omega_{ij}$  est une valeur connue qui peut varier d'observation en observation.

Dans certaines applications, chaque variable  $Y_i$  est binaire; on supposera alors que la composante aléatoire est distribuée selon une loi binomiale. Dans d'autres applications,



chaque réponse est un effectif distribué selon une loi de Poisson. Si chaque observation provient d'une variable continue, on peut supposer une distribution normale de la composante aléatoire.

La composante déterministe, exprimée sous forme d'une combinaison linéaire  $\beta_0 + \beta_1 X_1 + \dots + \beta_K X_K$  (appelée aussi **prédicteur linéaire**), précise quels sont les prédicteurs servant à décrire la moyenne de  $Y$  que l'on note  $\mu$ . Certaines des variables  $X_j$  peuvent se déduire de variables initiales utilisées dans le modèle. Par exemple, on pourra utiliser  $X_3 = X_1 \times X_2$  de façon à étudier l'interaction entre  $X_1$  et  $X_2$ .

Le lien entre la composante aléatoire et la composante déterministe spécifie comment l'espérance mathématique de  $Y$ , notée  $\mu$ , est liée au prédicteur linéaire construit à partir des variables explicatives. On peut modéliser une fonction monotone  $g(\mu)$  de l'espérance. On a alors :

$$g(E(Y)) = g(\mu) = \beta_0 + \beta_1 X_1 + \dots + \beta_K X_K \quad (5.1)$$

$$\text{autrement dit : } g(E(Y_{ij})) = g(\mu_{ij}) = X'_{ij} \beta \quad (5.2)$$

La fonction  $g$  est appelée **fonction de lien**.

### Quelques fonctions de lien

La fonction de lien  $g(\mu) = \log(\mu)$  permet par exemple de modéliser le logarithme de l'espérance et donne lieu aux modèles log-linéaires.

La fonction de lien  $g(\mu) = \log \frac{\mu}{1-\mu}$  modélise le logarithme du rapport des chances. Elle est appelée logit et est adaptée au cas où  $\mu$  est compris entre 0 et 1. C'est ce que l'on appelle la régression logistique.

### Choix du modèle

Composante aléatoire	Lien $g(\mu)$	Nature des variables de la composante déterministe	Modèle
Normale	Identité	Quantitatives	Régression
Normale	Identité	Qualitatives	Analyse de la variance
Normale	Identité	Mixtes	Analyse de la covariance
Binomiale	Logit	Mixtes	Régression logistique
Poisson	Log	Mixtes	Modèles log-linéaires
Multinomiale	Logit généralisé	Mixtes	Modèles à réponses multinomiales

TABLE 5.1 – Les principaux modèles linéaires généralisés.

Le choix du modèle linéaire généralisé dépend de la nature des données que l'on souhaite étudier. Le tableau 5.1 résume ces différents cas.

L'avantage des modèles linéaires généralisés est de fournir un cadre théorique adapté aussi bien à la modélisation de variables quantitatives que qualitatives. On dispose notamment de plusieurs tests statistiques permettant de tester l'adéquation du modèle aux données.

### 5.1.2 Estimation des paramètres $\beta_k$

Les estimateurs  $\hat{\beta}_k$  ( $k = 0, 1, \dots, K$ ) de  $\beta_k$  sont estimés par la méthode du maximum de vraisemblance. En effet, dans le cas d'un modèle linéaire généralisé, le logarithme de la vraisemblance s'écrit :

$$\ell = \sum_{i=1}^n \left[ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right]$$

avec :

$$E(Y_i) = \mu_i = b'(\theta_i) = \frac{\partial b(\theta_i)}{\partial \theta_i},$$

$$Var(Y_i) = a(\phi) b''(\theta_i) = a(\phi) \frac{\partial^2 b(\theta_i)}{\partial \theta_i^2},$$

$$g(\mu_i) = X_i' \beta = \eta_i.$$

L'estimateur  $\hat{\beta}$  est donc solution des équations de vraisemblance (équations du score) obtenues en dérivant la fonction de log-vraisemblance  $\ell$  au-dessus par rapport à  $\beta$  :

$$U_k = \frac{\partial \ell}{\partial \beta_k} = \sum_{i=1}^n \frac{(y_i - \mu_i) X_{ij}}{Var(Y_i)} \times \frac{\partial \mu_i}{\partial \eta_i} = 0 \quad (5.3)$$

Pour la plupart des modèles linéaires généralisés, ces équations sont non linéaires. On résout ainsi ces équations par un algorithme populaire appelé "Fisher scoring". Pour la régression logistique binomiale, et pour les modèles log-linéaires de Poisson, cet algorithme se simplifie, et n'est alors qu'une version du très connu algorithme de Newton-Raphson. L'estimateur  $\hat{\beta}$  est donc consistant (i.e.  $\hat{\beta} \xrightarrow{P} \beta$ ).

## 5.2 Cas particulier : la régression logistique

Ici, notre variable réponse  $Y$  ne prend que deux valeurs ( $Y = 1$  s'il y a eu au moins un dépassement de vitesse volontaire dans un intervalle de trente secondes et  $Y = 0$  sinon).  $Y$  suit alors une loi de Bernoulli caractérisée par les probabilités suivantes :  $p = P(Y = 1)$  et  $1 - p = P(Y = 0)$ . D'après le tableau 5.1, nous utiliserons alors la régression logistique dont la fonction de lien est la fonction *logit* tel que :

$$\text{logit}(x) = \log \frac{x}{1-x}.$$

Le modèle est donc :

$$g(E(Y)) = g(\mu) = \log \frac{\mu}{1-\mu} = \beta_0 + \beta_1 X_1 + \dots + \beta_K X_K.$$

Si on note  $p$  la probabilité que  $Y$  prenne la valeur 1 (i.e.  $p = P(Y = 1)$ ), on a alors  $\mu = p$  puisque  $Y$  est binaire. Le modèle se réécrit ainsi :

$$\text{logit}(p) = \log \frac{p}{1-p} = \beta_0 + \beta_1 X_1 + \dots + \beta_K X_K. \quad (5.4)$$



## Chapitre 6

# Différentes approches de l'analyse des données corrélées : équations d'estimation généralisées (GEE) et modèles linéaires généralisés mixtes (GLMM)

Nous avons remarqué que les données LAVIA ne sont pas indépendantes (des mesures pour un conducteur donné sont corrélées). Il n'est donc pas tout à fait rigoureux d'utiliser le modèle logistique qui suppose l'indépendance entre les observations. C'est pour cela que nous présenterons, dans ce chapitre, les deux principales méthodes pour analyser ce type de données corrélées (données longitudinales) : les équations d'estimation généralisées (GEE) et les modèles linéaires généralisés mixtes (GLMM).

### 6.1 Généralités sur les équations d'estimation généralisées (GEE)

#### 6.1.1 Equations d'estimation généralisées

Les **équations d'estimation généralisées (GEE)**, introduit par Liang et Zeger (1986), sont une méthode d'analyse des données corrélées qui, autrement, pourraient être modélisées à l'aide d'un modèle linéaire généralisé (GLM). Elles se sont imposées comme étant une stratégie essentielle dans l'analyse des données corrélées.

Nous reprenons les hypothèses et les notations du chapitre précédent. Nous supposons en plus une **matrice de corrélation de travail**  $R_i(\alpha)$  pour  $Y_i$  contenant certains paramètres inconnus représentés par le vecteur  $\alpha$ .

L'estimateur de  $\beta$  (vecteur des paramètres) est solution des équations de quasi-vraisemblance appelées **équations d'estimation généralisées** ou **generalized estimating equations (GEE)** :

$$\sum_{i=1}^n \frac{\partial \mu_i'}{\partial \beta} V_i^{-1} (Y_i - \mu_i(\beta)) = 0 \quad (6.1)$$

où  $\mu_i = (\mu_{i1}, \dots, \mu_{in_i})'$ , et  $V_i$  est un estimateur de la matrice de covariance de  $Y_i$  :

$$V_i = \phi A_i^{1/2} R_i(\alpha) A_i^{1/2}$$

où  $A_i$  est la matrice diagonale constituée des  $Var(\mu_{ij})$ .

L'estimation de  $\beta$  se fait à l'aide d'un processus itératif : algorithme de Fisher modifié. Le terme "modifié" signifie que l'on utilise  $V_i$  à la place de la vraie matrice de variances-covariances  $S_i$ .

L'intérêt de la méthode GEE repose sur la robustesse des estimations obtenues : les inférences sur  $\beta$  sont correctes même si la matrice  $V_i$  n'est pas correctement spécifiée. On utilise pour cela un estimateur "sandwich". Ainsi, sous des conditions de régularité faibles, l'estimateur  $\hat{\beta}$  est asymptotiquement distribué selon une loi multinormale :

$$\hat{\beta} \xrightarrow{d} N(\beta, I_0^{-1} I_1 I_0^{-1}) \quad (6.2)$$

où  $I_0 = \sum_i D_i' V_i^{-1} D_i$  et  $I_1 = \sum_i D_i' V_i^{-1} S_i V_i^{-1} D_i$  (avec  $D_i = \frac{\partial \mu_i}{\partial \beta'}$ ) évalué à  $\beta = \hat{\beta}$  et  $\alpha = \hat{\alpha}$ ,

et en estimant  $S_i$  par  $(Y_i - \mu_i(\hat{\beta}))(Y_i - \mu_i(\hat{\beta}))'$ .

Si la matrice de variances-covariances est correctement spécifiée (i.e.  $V_i = S_i$ ), alors  $I_0 = I_1$  et on retrouve le résultat classique :

$$\hat{\beta} \xrightarrow{d} N(\beta, I_0^{-1}) \quad (6.3)$$

$I_0$  étant la matrice d'information de Fisher.

### 6.1.2 Spécification de la matrice de corrélation de travail

#### Indépendance :

Dans ce cas, les mesures répétées pour un individu (ou groupe) sont indépendantes :

$$Corr(Y_{ij}, Y_{ij'}) = \begin{cases} 0 & \text{si } j \neq j' \\ 1 & \text{si } j = j' \end{cases}$$

et la matrice de corrélation de *travail* est la matrice identité :

$$R(\alpha) = I_{n_i \times n_i} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

#### Corrélation échangeable

Ici, on fait l'hypothèse que tout couple de mesures d'un sujet a pour corrélation  $\alpha$  :

$$\text{Corr}(Y_{ij}, Y_{ij'}) = \begin{cases} \alpha & \text{si } j \neq j' \\ 1 & \text{si } j = j' \end{cases}$$

Ce type de corrélation est la corrélation échangeable, et la matrice de corrélation de *travail* est :

$$R(\alpha) = \begin{bmatrix} 1 & \alpha & \dots & \alpha \\ \alpha & 1 & \dots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \dots & 1 \end{bmatrix}$$

Cette hypothèse est généralement appropriée pour des ensembles de données dans lesquels les mesures répétées n'ont aucune dépendance de temps. S'il y a peu d'individus et beaucoup d'observations par individu, une matrice de corrélation de *travail* échangeable est un bon choix. La corrélation échangeable assume un seul facteur de corrélation entre deux mesures répétées quelconques et la même variance pour chaque mesure répétée.

### Corrélation auto-régressive d'ordre 1 (AR(1))

Dans ce type de structure, on suppose que pour  $j \neq j'$  :

$$\text{Corr}(Y_{ij}, Y_{ij'}) = \alpha^{|j' - j|}$$

et la matrice de corrélation de *travail* est représentée sous la forme :

$$R(\alpha) = \begin{bmatrix} 1 & \alpha & \alpha^2 & \dots & \alpha^{n_i-1} \\ \alpha & 1 & \alpha & \dots & \alpha^{n_i-2} \\ \alpha^2 & \alpha & 1 & \dots & \alpha^{n_i-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{n_i-1} & \alpha^{n_i-2} & \alpha^{n_i-3} & \dots & 1 \end{bmatrix}$$

Comme cette structure implique que la corrélation diminue à mesure que l'écart entre  $j$  et  $j'$  augmente, ce type de corrélation est indiqué pour des ensembles de données dans lesquels les mesures répétées ont une dépendance temporelle, l'indice  $j$  dénotant l'ordre (chronologique) dans lequel les mesures ont été prises.

### Corrélation non structurée

Ce type de structure suppose que pour chaque paire d'observation d'un sujet a sa propre corrélation :

$$\text{Corr}(Y_{ij}, Y_{ij'}) = \alpha_{jj'}$$

La matrice de corrélation de travail non structurée est donc la plus générale des structures de corrélation discutées. Elle a la structure suivante :

$$R(\alpha) = \begin{bmatrix} 1 & \alpha_{1,2} & \dots & \alpha_{1,n_i} \\ \alpha_{1,2} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \alpha_{n_i-1,n_i} \\ \alpha_{1,n_i} & \dots & \alpha_{n_i-1,n_i} & 1 \end{bmatrix}$$

S'il y a peu d'observations par individu et plusieurs individus, une matrice de corrélation de *travail* non structurée est sans doute un bon choix.

Malheureusement, la matrice de corrélation ainsi estimée n'est pas nécessairement inversible et des problèmes numériques peuvent survenir, particulièrement pour des ensembles de données non équilibrés, c'est-à-dire le cas où l'on n'a pas le même nombre d'observations par individu (ou groupe) ou des données manquantes. Dans de tels cas, il est recommandé d'utiliser une structure de corrélation de travail plus simple, comme celles présentées aux sous-sections précédentes.

## 6.2 Généralités sur les modèles linéaires généralisés mixtes (GLMM)

### 6.2.1 Les modèles linéaires mixtes

Les modèles linéaires mixtes sont des extensions des modèles linéaires classiques qui ne contiennent que des effets fixes. Les modèles linéaires mixtes font intervenir des effets aléatoires spécifiques à chaque sujet. Il peut être plus vraisemblable qu'un sujet particulier soit associé à un risque plus élevé (si l'effet aléatoire est positif) ou à un risque plus faible (si l'effet aléatoire est négatif). Dans le cadre des modèles linéaires mixtes, c'est l'espérance de la variable réponse conditionnellement à ces effets aléatoires qui est modélisée sous la forme d'une combinaison linéaire des variables explicatives, incluant à la fois des facteurs fixes et des facteurs aléatoires. Les modèles mixtes modélisent non seulement la moyenne mais aussi la variance de la variable réponse. Les coefficients du modèle mesurent l'influence directe des variables explicatives sur la variable réponse pour des groupes de sujets hétérogènes et non pour la population.

Le modèle linéaire mixte est de la forme :

$$Y = X\beta + Zb + \varepsilon \tag{6.4}$$

avec :

- $Y$  est la variable de réponse ;
- $X$  est la matrice plan (en anglais : design matrix) connue des variables explicatives ;
- $\beta$  est un vecteur de paramètres fixes inconnus ;
- $Z$  est la matrice plan connue des effets aléatoires ;
- $b$  est un vecteur de paramètres aléatoires inconnus ;
- $\varepsilon$  est un vecteur des erreurs ;
- Les vecteurs  $b$  et  $\varepsilon$  sont supposés normalement distribués, indépendants, avec :

$$E \begin{bmatrix} b \\ \varepsilon \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$



$$\text{Var} \begin{bmatrix} b \\ \varepsilon \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}$$

### Remarque

- La moyenne conditionnelle à  $b$ , la moyenne, la variance conditionnelle à  $b$  et la variance de la variable réponse  $Y$  sont respectivement :

$$E(Y|b) = X\beta + Zb \quad (6.5)$$

$$E(Y) = X\beta \quad (6.6)$$

$$\text{Var}(Y|b) = R = \text{Var}(\varepsilon) \quad (6.7)$$

$$\text{Var}(Y) = ZGZ' + R \quad (6.8)$$

où  $Z'$  est la matrice transposée de  $Z$ .

Nous pouvons modéliser la variance de  $Y$  par un choix approprié de la matrice plan des effets aléatoires  $Z$  et par la spécification de la structure de covariance de  $G$  et  $R$ .

- Le modèle linéaire est un cas particulier du modèle mixte avec  $R = \sigma^2 I$  et  $Z = 0$ .

### 6.2.2 Les modèles linéaires généralisés mixtes

Comme son nom l'indique, les modèles linéaires généralisés mixtes sont une généralisation à la fois des modèles linéaires généralisés et des modèles linéaires mixtes. Dans un tel modèle, on suppose que :

$$Y = \mu + \varepsilon \quad (6.9)$$

et

$$g(\mu) = g(E(Y|b)) = X\beta + Zb \quad (6.10)$$

où, comme dans un modèle linéaire généralisé (cf. chapitre précédent),  $g$  est une fonction de lien monotone et, comme dans un modèle linéaire mixte,  $b$  est un vecteur de paramètres aléatoires inconnus,  $b \sim N(0, G)$ . On a encore  $\text{Var}(\varepsilon) = R$ .

Notons que l'équation 6.10 est équivalente à l'équation suivante :

$$g(E(Y_{ij}|b_i)) = g(\mu_{ij}) = X_{ij}\beta + Z_{ij}b_i \quad (6.11)$$

où l'indice  $ij$  indique la  $j$ -ième observation du sujet (ou groupe)  $i$ .



## Chapitre 7

# Applications aux données LAVIA : Modélisation de la probabilité d'occurrence des dépassements de vitesse volontaires

Nous modélisons ici la probabilité d'occurrence des dépassements de vitesse volontaires dans un intervalle de temps de trente secondes quelconque d'un trajet. Notre objectif est de savoir si le système LAVIA a de l'influence sur cette probabilité et si oui, sur quelle consigne de limitation de vitesse a-t-il le plus d'influence? Pour cela, nous effectuons d'abord des tirages aléatoires des intervalles de trente secondes afin de construire une base de données que nous utiliserons pour modéliser. Nous modélisons ensuite la probabilité d'occurrence des excès de vitesse volontaires. Nous considérons toujours le mode LAVIA **Neutre** et le mode LAVIA **Contraint**.

### 7.1 Tirage aléatoire

A partir de l'ensemble des trajets considérés (après avoir détecté et étiqueté les dépassements de vitesse, cf. chapitre 4), pour chaque mode LAVIA (neutre et contraint), nous tirons aléatoirement dix milles intervalles tel que chacun de ces intervalles appartienne à un seul trajet et à une seule consigne de limitation de vitesse. Nous fixons la taille de ces intervalles à trente secondes car un intervalle court permet de garantir l'homogénéité des conditions de circulation (type de route, intensité du trafic, etc.). Pour cela, pour chaque mode LAVIA (neutre et contraint), nous appliquons dix milles fois l'algorithme suivant :

1. Tirer au hasard un intervalle de trente secondes, compter ensuite le nombre de limitations de vitesse ainsi que le nombre de trajets auxquels appartient l'intervalle tiré ; en retirer un autre s'il y a plusieurs limitations de vitesse ou plusieurs trajets.  
Réitérer cette expérience jusqu'à trouver un intervalle appartenant à un seul trajet et à une seule consigne de limitation de vitesse.
2. Calculer la valeur de la nouvelle variable "*DepassType2*" (i.e. "occurrence des dépassements de vitesse volontaires (type 2)") pour l'intervalle obtenu à l'étape précédente tel que :

$$DepassType2 = \begin{cases} 1 & \text{s'il y a eu au moins un dépassement de vitesse volontaire} \\ 0 & \text{sinon} \end{cases}$$

## 7.2 Analyse descriptive des résultats du tirage aléatoire par tableau de contingence

Nos données, après avoir fait des tirages aléatoires décrits dans la section précédente, contiennent cinq variables ("Travia", "Trajet", "ModeLAVIA", "ConsigneLimiteVitesse", "DepassType2") et 20000 observations dont 10000 observations du mode LAVIA Neutre et 10000 observations du mode LAVIA Contraint. Nous utiliserons ces données pour modéliser la probabilité d'avoir un dépassement de vitesse volontaire dans un intervalle de temps de trente secondes. Avant de modéliser, nous faisons l'analyse descriptive des résultats obtenus après le tirage aléatoire.

### 7.2.1 Tableau de contingence de "DepassType2" et "ConsigneLimiteVitesse"

Le tableau de contingence de "*DepassType2*" et "*ConsigneLimiteVitesse*" ainsi que l'estimation de l'Odds Ratio (rapport de cotes) de toutes les limitations de vitesses par la limitation de vitesse 110 km/h sont présentés respectivement dans les tableaux 7.1 et 7.2.

Nous voyons que tous les Odds Ratios sont supérieurs à 1 (tableau 7.2). Cela se traduit par le fait que, par rapport de la limitation de vitesse 110 km/h, les limitations de vitesse inférieures (30, 50, 70, 90 km/h) sont associées à un risque de dépassement de vitesse volontaire plus élevé. Nous reamarquons aussi que le risque le plus élevé est associée à la limitation de vitesse 70 km/h (l'Odds Ratio le plus élevé).

### 7.2.2 Tableau de contingence de "DepassType2" et "ModeLAVIA"

Les tableaux 7.3 et 7.4 représentent respectivement le tableau de contingence de "*DepassType2*" et "*ModeLAVIA*" et l'estimation de l'Odds Ratio (rapport de cotes) du mode LAVIA Contraint par le mode LAVIA Neutre. L'Odds Ratio est inférieur à 1 indiquant la diminution du risque d'excès de vitesse volontaire en mode LAVIA Contraint, c'est-à-dire avec le système LAVIA. Ces résultats confirment les analyses que nous avons faites à la section 4.2 (page 20).

DepassType2	ConsigneLimiteVitesse					
Fréquence						
Pourcentage						
Pctage en ligne						
Pctage en col.	30	50	70	90	110	Total
<b>1</b>	202	2335	274	262	406	3479
	1.01	11.68	1.37	1.31	2.03	17.40
	5.81	67.12	7.88	7.53	11.67	
	18.62	17.44	28.05	20.94	12.32	
<b>0</b>	883	11056	703	989	2890	16521
	4.42	55.28	3.52	4.95	14.45	82.61
	5.34	66.92	4.26	5.99	17.49	
	81.38	82.56	71.95	79.06	87.68	
<b>Total</b>	1085	13391	977	1251	3296	20000
	5.43	66.96	4.89	6.26	16.48	100.00

TABLE 7.1 – Tableau de contingence de "DepassType2" et "ConsigneLimiteVitesse" (les données du tirage aléatoire).

ConsigneLimiteVitesse	Odds Ratio	IC 95%	
30 vs 110	1.6284	1.3536	1.9590
50 vs 110	1.5034	1.3426	1.6833
70 vs 110	2.7744	2.3313	3.3017
90 vs 110	1.8857	1.5889	2.2380

TABLE 7.2 – Estimation des Odds Ratios (rapports de cotes) pour "ConsigneLimiteVitesse" à partir des données du tirage aléatoire.

DepassType2	ModeLAVIA		
Fréquence			
Pourcentage			
Pctage en ligne			
Pctage en col.	Contraint	Neutre	Total
<b>1</b>	1277	2202	3479
	6.39	11.01	17.40
	36.71	63.29	
	12.77	22.02	
<b>0</b>	8723	7798	16521
	43.62	38.99	82.61
	52.80	47.20	
	87.23	77.98	
<b>Total</b>	10000	10000	20000
	50.00	50.00	100.00

TABLE 7.3 – Tableau de contingence de "DepassType2" et "ModeLAVIA" (les données du tirage aléatoire).

ModeLAVIA	Odds Ratio	IC 95%	
Contraint VS Neutre	0.5148	0.4808	0.5590

TABLE 7.4 – Estimation de l'Odds Ratio (rapport de cotes) à partir des données du tirage aléatoire : le mode LAVIA Contraint VS le mode LAVIA Neutre pour le dépassement de vitesse volontaire.

## 7.3 Modélisation de la probabilité d'occurrence des dépassements de vitesse volontaires dans un intervalle de temps de trente secondes

Nous cherchons à savoir si le système LAVIA a de l'influence sur la probabilité d'occurrence des dépassements de vitesse volontaires dans un intervalle de trente secondes et sur quelle consigne de limitation de vitesse a-t-il le plus d'influence. Pour cela, nous modélisons cette probabilité. Ici, notre variable réponse  $Y$  est la variable "*DepassType2*" (occurrence des dépassements de vitesse volontaires dans un intervalle de trente secondes) qui est binaire (elle prend 0 et 1 comme valeurs). Les méthodes que nous utiliserons pour modéliser cette probabilité ont été présentées au chapitre 5 et au chapitre 6.

### 7.3.1 Modélisation par régression logistique

Nous avons vu au chapitre 5 que la régression logistique est un cas particulier du modèle linéaire généralisé (GLM) avec la fonction de lien "*logit*" (car la variable réponse  $Y$  est binaire) et le modèle est :

$$g(E(Y_{ij})) = g(p_{ij}) = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = X'_{ij}\beta = \beta_0 + \beta_1 X_{ij1} + \dots + \beta_K X_{ijK}$$

ce qui équivaut en terme de probabilité à :

$$p_{ij} = \frac{\exp(X'_{ij}\beta)}{1 + \exp(X'_{ij}\beta)}$$

avec :

- L'indice  $ij$  indique le trajet  $j$  du conducteur  $i$  auquel l'intervalle tiré aléatoirement appartient ;
- $p_{ij} = P(Y_{ij} = 1) = E(Y_{ij})$  (car  $Y$  est binaire) ;
- $\beta = [\beta_0, \beta_1, \dots, \beta_K]$  est le vecteur des paramètres qui seront estimés ( $K$  est le nombre de variables explicatives).

Nous rappelons qu'ici, les observations sont considérées comme indépendantes.

### Modélisation

Nous commençons d'abord par modéliser avec deux variables explicatives : "*ModeLAVIA*", "*ConsigneLimiteVitesse*". Comme les variables explicatives sont qualitatives, il y aura un paramètre pour chaque modalité de chacune de ces variables. Il y a donc en total 8 paramètres à estimer dont 1 pour l'intercept, 5 pour cinq modalités de la variable "*ConsigneLimiteVitesse*" (30, 50, 70, 90 et 110 km/h) et 2 pour deux modalités de la variable "*ModeLAVIA*" (Neutre et Contraint). Les résultats obtenus à l'aide de la procédure GENMOD de SAS (sans option REPEATED) sont présentés dans les tableaux 7.5, 7.6, 7.7.

Toutes les variables sont significatives (les "p-values" sont inférieures à 5%, cf. le tableau 7.7). Si on observe le tableau 7.5, les lignes correspondantes à une consigne de 110 km/h et au mode LAVIA Neutre sont les références, à partir desquelles est calculée la probabilité d'avoir un dépassement de vitesse volontaire. D'après ce tableau, le mode LAVIA Contraint est

Paramètre	Estimation	Erreur type	IC 95%		p-value
<b>Intercept</b>	-1.6859	0.0550	-1.7951	-1.5793	<.0001
<b>ConsigneLimiteVitesse_30</b>	0.5329	0.0951	0.3453	0.7181	<.0001
<b>ConsigneLimiteVitesse_50</b>	0.4302	0.0581	0.3175	0.5451	<.0001
<b>ConsigneLimiteVitesse_70</b>	1.0463	0.0896	0.8702	1.2216	<.0001
<b>ConsigneLimiteVitesse_90</b>	0.6771	0.0881	0.5037	0.8493	<.0001
<b>ConsigneLimiteVitesse_110</b>	0.0000				
<b>ModeLavia_Constraint</b>	-0.6691	0.0386	-0.7450	-0.5935	<.0001
<b>ModeLavia_Neutre</b>	0.0000				

TABLE 7.5 – Estimation des paramètres du modèle logistique à deux variables explicatives.

	Odds Ratio	IC 95%		p-value
<b>ModeLAVIA</b>				
<b>Constraint vs Neutre</b>	0.5122	0.4748	0.5525	<.0001
<b>ConsigneLimiteVitesse</b>				
<b>30 vs 110</b>	1.7039	1.4143	2.0529	<.0001
<b>50 vs 110</b>	1.5376	1.3723	1.7229	<.0001
<b>70 vs 110</b>	2.8472	2.3886	3.3940	<.0001
<b>90 vs 110</b>	1.9681	1.6559	2.3392	<.0001

TABLE 7.6 – Estimation des Odds Ratios (rapports de cotes) du modèle logistique à deux variables explicatives.

Effet	DDL	Khi-2	p-value
<b>ConsigneLimiteVitesse</b>	4	152.25	<.0001
<b>ModeLAVIA</b>	1	309.25	<.0001

TABLE 7.7 – Test global des variables explicatives du modèle logistique à deux variables.



associé à une valeur négative (coefficient estimé à -0.6691) ce qui signifie que la probabilité d'avoir un dépassement de vitesse volontaire avec le système LAVIA en mode Contraint diminue par rapport au mode Neutre. En outre, on peut dire que cet effet est significatif car la "p-value" associée est inférieure à 5%, ce qui signifie que le paramètre estimé est significativement différent de 0. C'est d'ailleurs le cas de tous les autres paramètres estimés. Concernant l'effet des limitations de vitesse, elles augmentent toutes la probabilité d'avoir un dépassement de vitesse volontaire (paramètres positifs et significativement différents de 0) par rapport aux zones limitées à 110 km/h. Les excès de vitesse volontaires sont plus probables dans les zones limitées à 70 km/h (paramètre plus élevé). Nous observons aussi que ces résultats sont confirmés par les résultats de l'estimation des Odds Ratios (rapport de cotes) représentée au tableau 7.6. En effet, l'Odds Ratio estimé du mode LAVIA Contraint par rapport au mode LAVIA Neutre est inférieur à 1 (0.5122). Cela signifie que le système LAVIA en mode Contraint diminue la probabilité d'occurrence des dépassements de vitesse volontaires. Concernant les limitations de vitesse, tous les Odds Ratios estimés sont supérieurs à 1 ce qui signifie que par rapport à la limitation de vitesse 110 km/h, toutes les autres limitations augmentent cette probabilité. La limitation de vitesse 70 km/h a le taux d'augmentation le plus élevé (l'Odds Ratio le plus élevé).

Nous essayons aussi de modéliser un autre modèle logistique à trois variables explicatives : "ModeLAVIA", "ConsigneLimiteVitesse" et "ModeLAVIA\*ConsigneLimiteVitesse". En observant les résultats obtenus, nous remarquons que le paramètre associé à l'interaction entre le mode LAVIA Contraint et la consigne de limitation de vitesse 90 km/h de ce modèle n'est pas significatif ( $p\text{-value} = 0.8855 > 5\%$ , cf. le tableau A.1 de l'annexe A), ce qui signifie que l'influence du système LAVIA est indépendante de la limitation de vitesse. Le modèle à deux variables explicatives est donc préférable puisque tous ses paramètres sont significatifs. Nous choisissons finalement le modèle à deux variables explicatives : "ModeLAVIA" et "ConsigneLimiteVitesse".

### 7.3.2 Modélisation par équations d'estimation généralisées (GEE)

La régression logistique que nous venons d'utiliser dans le paragraphe précédent suppose que les observations sont toutes indépendantes entre elles. Or ici, nous sommes dans un cas de données corrélées : un même conducteur a plusieurs observations (plusieurs trajets), donc on crée un cluster pour chaque conducteur de taille variable. Ainsi, la régression logistique n'est plus valable. Nous utilisons ici la méthode basée sur les GEE (cf. le chapitre 6) qui tient compte des corrélations intra-conducteurs, c'est-à-dire les observations sont considérées comme indépendantes entre les conducteurs mais corrélées pour un même conducteur. Notre structure de corrélation est "échangeable" car nous n'avons pas le même nombre d'observations par individu (conducteur), nous ne pouvons donc pas utiliser la structure de corrélation "non structurée". Le modèle ne change pas. Il s'écrit donc de la même manière que précédemment (nous prenons les notations du paragraphe précédent et toujours pour les mêmes raisons que précédemment, la fonction de lien est encore la fonction "logit") :

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = X'_{ij}\beta = \beta_0 + \beta_1 X_{ij1} + \dots + \beta_K X_{ijK}.$$

## Modélisation

Nous commençons aussi par modéliser avec deux variables explicatives : "ModeLAVIA", "ConsigneLimiteVitesse". Les paramètres restent donc inchangés. Les résultats obtenus avec la structure de corrélation **échangeable** (i.e. la corrélation de tout couple de trajets d'un conducteur est la même) sont donnés aux tableaux 7.8, 7.9, 7.10. Ces résultats sont obtenus à l'aide de la procédure GENMOD de SAS avec l'option **REPEATED**.

Paramètre	Estimation	Erreur type	IC 95%		p-value
<b>Intercept</b>	-1.6714	0.1944	-2.0525	-1.2903	<.0001
<b>ConsigneLimiteVitesse_30</b>	0.7296	0.2472	0.2451	1.2140	0.0032
<b>ConsigneLimiteVitesse_50</b>	0.4710	0.1888	0.1010	0.8411	0.0126
<b>ConsigneLimiteVitesse_70</b>	1.0027	0.2162	0.5789	1.4265	<.0001
<b>ConsigneLimiteVitesse_90</b>	0.6285	0.1900	0.2561	1.0009	0.0009
<b>ConsigneLimiteVitesse_110</b>	0.0000				
<b>ModeLavia_Contraint</b>	-0.5632	0.0854	-0.7306	-0.3959	<.0001
<b>ModeLavia_Neutre</b>	0.0000				

TABLE 7.8 – Estimation des paramètres du modèle GEE à deux variables explicatives, donnant l'effet du mode LAVIA et des limitations de vitesse sur la probabilité d'avoir un dépassement de vitesse volontaire.

	Odds Ratio	IC 95%		p-value
<b>ModeLAVIA</b>				
<b>Contraint vs Neutre</b>	0.5694	0.4816	0.6731	<.0001
<b>ConsigneLimiteVitesse</b>				
<b>30 vs 110</b>	2.0742	1.2778	3.3670	0.0032
<b>50 vs 110</b>	1.6016	1.1063	2.3188	0.0126
<b>70 vs 110</b>	2.7255	1.7840	4.1640	<.0001
<b>90 vs 110</b>	1.8747	1.2918	2.7207	0.0009

TABLE 7.9 – Estimation des Odds Ratios (rapports de cotes) du modèle GEE à deux variables explicatives.

Effet	DDL	Khi-2	p-value
<b>ConsigneLimiteVitesse</b>	4	17.26	0.0017
<b>ModeLAVIA</b>	1	17.98	<.0001

TABLE 7.10 – Test global des variables explicatives du modèle GEE à deux variables.

La corrélation "échangeable" estimée est petite (0.0607) indiquant une corrélation marginale faible entre les observations d'un même conducteur.

Toutes les variables sont significatives (les "p-values" sont inférieures à 5%, cf. le tableau 7.10). Si on observe le tableau 7.8, comme le tableau 7.5, les lignes correspondant à une consigne de 110 km/h et au mode LAVIA Neutre sont les références, à partir desquelles est calculée la probabilité d'avoir un dépassement de vitesse volontaire. Ainsi, le mode LAVIA Contraint est associé à une valeur négative (coefficient estimé à -0.5632) ce qui signifie que la probabilité d'avoir un dépassement de vitesse volontaire avec le système LAVIA en mode Contraint diminue par rapport à l'usage du mode Neutre. En outre, cet effet est significatif car la "p-value" associée est inférieure à 5%. Ce paramètre estimé est donc significativement différent de 0. Tous les autres paramètres estimés le sont aussi. De la part de l'effet des limitations de vitesse, elles augmentent toutes la probabilité d'avoir un dépassement de vitesse volontaire (paramètres positifs et significativement différents de 0) par rapport à la limitation 110 km/h. Nous observons aussi que les zones limitées à 70 km/h possèdent une probabilité d'occurrence des excès de vitesse volontaires la plus élevée (paramètre le plus élevé). Les Odds Ratios estimés confirment ces résultats.

Nous essayons également de modéliser un autre modèle GEE à trois variables explicatives en ajoutant l'interaction "ModeLAVIA\*ConsigneLimiteVitesse". Le tableau 7.11 contient les résultats du test global des effets des variables explicatives du modèle avec la structures de corrélation échangeable. D'après ce tableau, **l'effet "ModeLAVIA\*ConsigneLimiteVitesse" n'est pas significatif**. En effet,  $\chi^2_{\alpha=0.05}(DDL) = \chi^2_{\alpha=0.05}(4) = 9.49 > \chi^2\text{-value} = 6.87$  (où  $\alpha$  est le risque de première espèce) et  $p\text{-value} = 0.1429$ . Cela confirme l'indépendance des effets de la limitation de vitesse et du système LAVIA.

Effet	DDL	Khi-2	p-value
ConsigneLimiteVitesse	4	18.11	0.0012
ModeLAVIA	1	15.35	<.0001
ModeLAVIA*ConsigneLimiteVitesse	4	6.87	0.1429

TABLE 7.11 – Test global des variables explicatives du modèle GEE à trois variables.

Par conséquent, le modèle à deux variables explicatives ("ModeLAVIA" et "Consigne-LimiteVitesse") est notre modèle GEE final.

### 7.3.3 Modélisation par modèle linéaire généralisé mixte (GLMM)

Nous avons vu au chapitre 6 que les modèles linéaires généralisés mixtes et les GEE sont deux principales méthodes pour analyser les données longitudinales (mesures répétées dans le temps). Nous venons de présenter les résultats obtenus par la méthode basée sur les GEE dans le paragraphe précédent. Dans le cadre de ce paragraphe, nous utilisons la dernière méthode, les modèles linéaires généralisés mixtes.

La fonction de lien ne change toujours pas (la fonction "logit"). En revanche, concernant les effets, nous introduisons un effet aléatoire qui est associé à l'intercept  $\beta_0$  et qui traduit

le fait que la probabilité d'occurrence des dépassements de vitesse volontaires varie entre les conducteurs. Le modèle est donc :

$$\text{logit}(E(Y_{ij}|b_{0i})) = \text{logit}(\mu_{ij}) = (\beta_0 + b_{0i}) + \beta_1 X_{ij1} + \dots + \beta_K X_{ijK}$$

avec  $\beta = [\beta_0, \beta_1, \dots, \beta_K]$  le vecteur des paramètres des effets fixes ( $K$  est le nombre de variables explicatives) et  $b = b_{0i}$  le paramètre d'effet aléatoire associé à l'intercept  $\beta_0$ .

Rappelons que  $\mu_{ij} = E(Y_{ij}|b_{0i})$  et  $p_{ij} = P(Y_{ij} = 1|b_{0i})$  sont égaux puisque notre variable réponse est binaire. Le modèle se réécrit ainsi :

$$\text{logit}(p_{ij}) = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = (\beta_0 + b_{0i}) + \beta_1 X_{ij1} + \dots + \beta_K X_{ijK}$$

où  $b_{0i} \sim N(0, \sigma^2)$ .

## Modélisation

Nous commençons encore par un modèle à deux variables explicatives : "ModeLAVIA" et "ConsigneLimiteVitesse". Les résultats obtenus à l'aide de la procédure GLIMMIX de SAS sont représentés dans les tableaux 7.12, 7.13, 7.14.

La variance  $\sigma^2$  de l'effet aléatoire associé à l'intercept est estimée à 0.4761.

Les lignes correspondant à une vitesse de consigne de 110 km/h et au mode du LAVIA Neutre du tableau 7.12 sont toujours les références. Tous les paramètres des effets fixes estimés sont significatifs (les "p-values" sont inférieures à 5% signifiant les paramètres estimés sont significativement différents de 0). Ces paramètres sont légèrement plus grands que ceux du modèle GEE que nous avons vu précédemment (à cause de l'ajout de l'effet aléatoire). Cependant, les résultats de ces deux modèles sont similaires. La probabilité d'avoir un dépassement de vitesse volontaire avec le système LAVIA en mode Contraint diminue par rapport à l'usage du mode Neutre et cet effet est significatif puisque le mode du LAVIA Contraint est associé à une valeur négative (coefficient estimé à -0.6228) et à un Odds Ratio inférieur à 1 (0.5364) et que les "p-values" associées sont inférieures à 5%. Les limitations de vitesse augmentent toutes significativement la probabilité d'avoir un dépassement de vitesse volontaire (paramètres positifs, Odds Ratios supérieurs à 1 et les "p-values" associées sont toutes inférieures à 5%) par rapport aux zones limitées à 110 km/h. Le risque le plus élevé est associé aux zones limitées à 70 km/h (paramètre et Odds Ratio les plus élevés).

Nous essayons ensuite de modéliser un autre modèle à trois variables explicatives en ajoutant la variable "ConsigneLimiteVitesse\*ModeLAVIA". En observant les résultats obtenus, nous voyons que le paramètre de l'effet fixe "ModeLAVIA\_Contraint\*ConsigneLimiteVitesse\_90" de ce modèle est non significatif car la p-value qui vaut 0.4836 est supérieure à 5% (cf. le tableau C.1 de l'annexe C) tandis que les paramètres des effets fixes estimés du modèle à deux variables explicatives sont tous significatifs. Le modèle à deux variables explicatives est donc meilleur que celui à trois variables explicatives. Nous choisissons ainsi le modèle à deux variables explicatives ("ModeLAVIA" et "ConsigneLimiteVitesse").

Paramètre (effet fixe)	Estimation	Erreur type	IC 95%		p-value
<b>Intercept</b>	-1.8625	0.1217	-2.1080	-1.6171	<.0001
<b>ConsigneLimiteVitesse_30</b>	0.8283	0.1044	0.6237	1.0329	<.0001
<b>ConsigneLimiteVitesse_50</b>	0.5310	0.06707	0.3995	0.6624	<.0001
<b>ConsigneLimiteVitesse_70</b>	1.1003	0.09731	0.9095	1.2910	<.0001
<b>ConsigneLimiteVitesse_90</b>	0.6996	0.09607	0.5113	0.8879	<.0001
<b>ConsigneLimiteVitesse_110</b>	0				
<b>ModeLavia_Constraint</b>	-0.6228	0.04158	-0.7043	-0.5413	<.0001
<b>ModeLavia_Neutre</b>	0				

TABLE 7.12 – Estimation des paramètres des effets fixes du modèle linéaire généralisé mixte à deux variables explicatives.

	Odds Ratio	IC 95%		p-value
<b>ModeLAVIA</b>				
<b>Constraint vs Neutre</b>	0.5364	0.4945	0.5820	<.0001
<b>ConsigneLimiteVitesse</b>				
<b>30 vs 110</b>	2.2894	1.8659	2.8091	<.0001
<b>50 vs 110</b>	1.7006	1.4911	1.9395	<.0001
<b>70 vs 110</b>	3.0050	2.4832	3.6364	<.0001
<b>90 vs 110</b>	2.0129	1.6674	2.4299	<.0001

TABLE 7.13 – Estimation des Odds Ratios (rapports de cotes) du modèle linéaire généralisé mixte à deux variables explicatives.

Effet	DDL Num.	DDL Res.	Valeur F	p-value
<b>Intercept</b>	1	43	200.37	<.0001
<b>ConsigneLimiteVitesse</b>	4	19951	37.22	<.0001
<b>ModeLAVIA</b>	1	19951	224.32	<.0001

TABLE 7.14 – Test global des effets fixes du modèle linéaire généralisé mixte à deux variables explicatives.

## 7.4 Comparaison des modèles

### 7.4.1 Comparaison des Odds Ratios

Les estimations d'Odds Ratio obtenues à partir de quatre méthodes utilisées (le tableau de contingence, le modèle logistique, le modèle GEE et le modèle linéaire généralisé mixte (GLMM)) sont présentées dans les tableaux 7.15, 7.16 et aussi illustrées dans la figure 7.1.

Nous rappelons que l'Odds Ratio est significativement différent de 1 si et seulement si son intervalle de confiance à 95% n'inclut pas la valeur neutre 1. Ainsi, les Odds Ratios des quatre méthodes sont tous significativement différent de 1 puisque les intervalles de confiance à 95% n'incluent pas la valeur neutre 1 (figure 7.1). Cela signifie que les modes du LAVIA et les limitations de vitesse ont des effets significativement différents sur la probabilité d'occurrence des excès de vitesse volontaires.

En effet, le système LAVIA (mode LAVIA Contraint) diminue de moitié la probabilité d'occurrence des excès de vitesse volontaires (l'Odds Ratio est d'environ 0.5 pour toutes les quatre méthodes), et par rapport à la limitation de vitesse 110 km/h, toutes les autres limitations augmentent cette probabilité (l'Odds Ratio est supérieur à 1 pour toutes les quatre méthodes). La limitation de vitesse 70km/h a l'impact le plus grand sur l'excès de vitesse volontaire car l'Odds Ratio associé est le plus élevé pour chacune de quatre méthodes utilisées (2.7744 pour l'Odds Ratio brut, 2.8472 pour le modèle logistique, 2.7255 pour le modèle GEE, 3.0050 pour le GLMM) et celle ayant l'impact le plus petit est 50 km/h (l'Odds Ratio associé est le plus petit pour chaque méthode). Du côté de l'impact des limitations de vitesse 30 km/h et 90 km/h, il y a une différence entre les quatre méthodes. On observe deux groupes qui ont des résultats opposés. Les excès de vitesse volontaires sont plus probables dans les zones limitées à 90 km/h que dans les zones limitées à 30 km/h pour le tableau de contingence et pour le modèle logistique (groupe de l'hypothèse d'indépendance entre les observations) et c'est le contraire pour le modèle GEE et le GLMM (groupe de l'hypothèse de corrélation intra-conducteur). De plus, on peut dire que tous ces impacts sont significatifs car les  $p$ -values associées sont inférieures à 5%.

L'Odds Ratio obtenu par le modèle GEE pour le mode LAVIA est le plus élevé. Cependant, pour les limitations de vitesse, ce sont les Odds Ratios obtenus par le modèle GLMM qui sont les plus élevés tandis que ceux du modèle GEE sont tantôt les plus petits (pour les limitations de vitesse 70 km/h et 90 km/h), tantôt plus grands que les Odds Ratios brut et ceux du modèle logistique (pour les limitations de vitesse 30 km/h et 50 km/h). Les valeurs des Odds Ratios bruts et celles des Odds Ratio obtenus par le modèle logistique sont proches. Ces résultats peuvent se traduire par le fait que les GEE et le modèle linéaire généralisé mixte (GLMM) supposent la corrélation intra-conducteur (i.e. les observations sont considérées comme corrélées pour un même conducteur et indépendantes entre les conducteurs) ce qui n'est pas le cas pour le tableau de contingence ainsi que pour la régression logistique (elles ne tiennent pas compte de la dépendance des données au sein d'un même cluster). De plus, l'effet aléatoire associé à l'intercept qui traduit le fait que la probabilité d'occurrence des dépassements de vitesse volontaires varie entre les conducteurs a été introduit dans le GLMM.

Les intervalles de confiance à 95% des Odds Ratios bruts et des Odds Ratios du modèle logistique sont plus étroits que ceux du modèle GEE et du GLMM. Le modèle GEE

donnent les intervalles de confiance à 95% plus grands. Cela signifie que la précision de l'estimation obtenue par le modèle GEE est plus faible. Ainsi, quand nous prenons en compte la corrélation intra-conducteur, le GLMM semble préférable au modèle GEE car les deux modèles tiennent compte de la corrélation intra-conducteur et donnent les mêmes résultats de l'impact des modes LAVIA et des limitations de vitesse sur le dépassement de vitesse volontaire mais la précision de l'estimation obtenue par le modèle GEE est plus faible. Le GLMM est préférable au modèle logistique ainsi qu'au tableau de contingence au niveau de l'hypothèse. Ainsi, le modèle linéaire généralisé mixte GLMM est le meilleur.

Facteur	Tableau de contingence Odds Ratio brut			Modèle logistique			
	Odds Ratio	IC 95%		Odds Ratio	IC 95%		p-value
<b>ModeLAVIA</b>							
<b>Contraint vs Neutre</b>	<b>0.5148</b>	0.4808	0.5590	<b>0.5122</b>	0.4748	0.5525	<.0001
<b>ConsigneLimiteVitesse</b>							
<b>30 vs 110</b>	<b>1.6284</b>	1.3536	1.9590	<b>1.7039</b>	1.4143	2.0529	<.0001
<b>50 vs 110</b>	<b>1.5034</b>	1.3426	1.6833	<b>1.5376</b>	1.3723	1.7229	<.0001
<b>70 vs 110</b>	<b>2.7744</b>	2.3313	3.3017	<b>2.8472</b>	2.3886	3.3940	<.0001
<b>90 vs 110</b>	<b>1.8857</b>	1.5889	2.2380	<b>1.9681</b>	1.6559	2.3392	<.0001

TABLE 7.15 – Les Odds Ratios bruts (le tableau de contingence) et les Odds Ratios estimés du modèle logistique pour le dépassement de vitesse volontaire.

Facteur	GEE				GLMM			
	Odds Ratio	IC 95%		p-value	Odds Ratio	IC 95%		p-value
<b>ModeLAVIA</b>								
<b>Contraint vs Neutre</b>	<b>0.5694</b>	0.4816	0.6731	<.0001	<b>0.5364</b>	0.4945	0.5820	<.0001
<b>ConsigneLimiteVitesse</b>								
<b>30 vs 110</b>	<b>2.0742</b>	1.2778	3.3670	0.0032	<b>2.2894</b>	1.8659	2.8091	<.0001
<b>50 vs 110</b>	<b>1.6016</b>	1.1063	2.3188	0.0126	<b>1.7006</b>	1.4911	1.9395	<.0001
<b>70 vs 110</b>	<b>2.7255</b>	1.7840	4.1640	<.0001	<b>3.0050</b>	2.4832	3.6364	<.0001
<b>90 vs 110</b>	<b>1.8747</b>	1.2918	2.7207	0.0009	<b>2.0129</b>	1.6674	2.4299	<.0001

TABLE 7.16 – Les Odds Ratios estimés du modèle GEE et du modèle linéaire généralisé mixte pour le dépassement de vitesse volontaire.

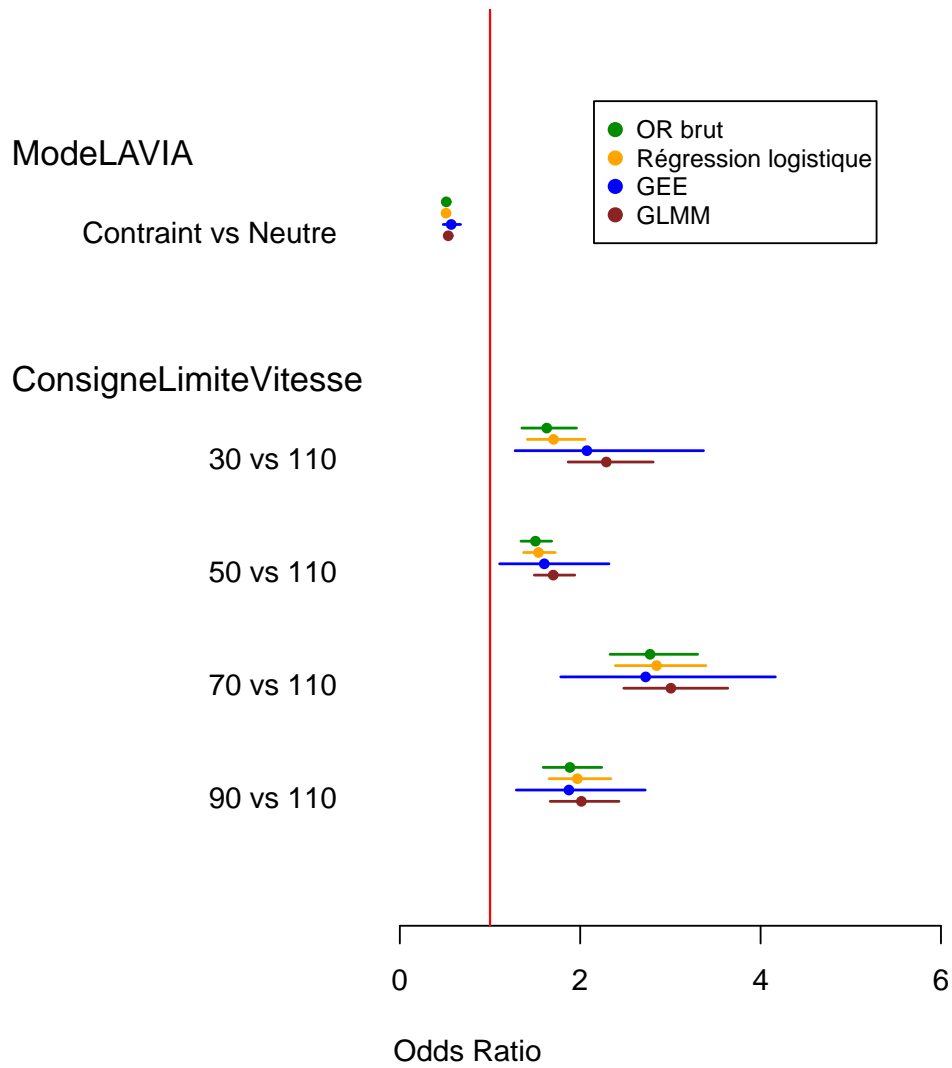


FIGURE 7.1 – Graphique des Odds Ratios bruts et des Odds Ratios de trois modèles (logistique, GEE et linéaire généralisé mixte) pour le dépassement de vitesse volontaire. La barre rouge correspond à la valeur neutre (i.e 1) d'Odds Ratio.



### 7.4.2 Comparaison de l'estimation des paramètres et des erreurs types

Les deux tableaux 7.17 et 7.18 représentent l'estimation des paramètres ainsi que les erreurs types de trois modèles : le modèle logistique, le modèle GEE et le modèle linéaire généralisé mixte (GLMM).

Paramètre	Estimation		
	Modèle logistique	GEE	GLMM
Intercept	-1.6859***	-1.6714***	-1.8625***
ConsigneLimiteVitesse_30	0.5329***	0.7296*	0.8283***
ConsigneLimiteVitesse_50	0.4302***	0.4710**	0.5310***
ConsigneLimiteVitesse_70	1.0463***	1.0027***	1.1003***
ConsigneLimiteVitesse_90	0.6771***	0.6285**	0.6996***
ConsigneLimiteVitesse_110	0.0000	0.0000	0.0000
ModeLavia_Contraint	-0.6691***	-0.5632***	-0.6228***
ModeLavia_Neutre	0.0000	0.0000	0.0000

\*  $p$ -value < 0.05; \*\*  $p$ -value < 0.01; \*\*\*  $p$ -value < 0.001

TABLE 7.17 – Estimation des paramètres de trois modèles : le modèle logistique, le modèle GEE et le modèle linéaire généralisé mixte (GLMM).

Paramètre	Erreur type		
	Modèle logistique	GEE	GLMM
Intercept	0.0550	0.1944	0.1217
ConsigneLimiteVitesse_30	0.0951	0.2472	0.1044
ConsigneLimiteVitesse_50	0.0581	0.1888	0.06707
ConsigneLimiteVitesse_70	0.0896	0.2162	0.09731
ConsigneLimiteVitesse_90	0.0881	0.1900	0.09607
ConsigneLimiteVitesse_110			
ModeLavia_Contraint	0.0386	0.0854	0.04158
ModeLavia_Neutre			

TABLE 7.18 – Estimation des erreurs types de trois modèles : le modèle logistique, le modèle GEE et le modèle linéaire généralisé mixte (GLMM).

Par rapport au modèle GEE, les paramètres du GLMM sont tous plus grands (à cause de l'ajout de l'effet aléatoire). Par rapport au modèle logistique, sauf le paramètre associé au mode LAVIA Contraint, les autres paramètres de ce modèle sont également plus grands. L'intercept et les paramètres associés aux limitations de vitesse 70 km/h et 90 km/h du

modèle logistique sont plus grands que ceux du modèle GEE, ses autres limitations de vitesse ont les paramètres les plus petits tandis que son paramètre associé au mode LAVIA Contraint est le paramètre le plus grand par rapport aux deux autres modèles (GEE et GLMM). Malgré toutes ces différences de valeur entre les paramètres des trois modèles, on obtient toujours le même résultat sur l'impact du système LAVIA : le mode LAVIA Contraint diminue significativement la probabilité d'occurrence des excès de vitesse volontaires car son coefficient est significativement négatif ( $p$ -value inférieure à 0.001) pour les trois modèles. Les résultats des trois modèles sont aussi les mêmes sur l'effet des autres limitations de vitesse par rapport à la limitation de vitesse 110 km/h. En effet, par rapport à la limitation de vitesse 110 km/h, le risque a significativement augmenté pour toutes les limitations de vitesse inférieures (coefficients positifs et  $p$ -values inférieures à 5%) et le risque le plus grand est associé aux zones limitées à 70 km/h (coefficient le plus élevé).

Tous les paramètres pour les trois modèles sont significatifs ( $p$ -values inférieures à 0.05). Cependant, ceux du modèle logistique et du GLMM sont plus significatifs car les  $p$ -values sont toutes inférieures à 0.001. Mais le GLMM est meilleur que le modèle logistique car il prend en compte la corrélation intra-conducteur qui n'est pas le cas pour le modèle logistique.

Concernant les erreurs types, celles du modèle logistique sont les plus petites et celles du modèle GEE sont les plus grandes. Ces résultats montrent que le modèle logistique ne considère pas la dépendance entre les observations d'un conducteur, ce qui provoque une sous-estimation des erreurs types. Avec une hypothèse tenant compte de la corrélation entre les observations sur les données, le GLMM est préférable au modèle logistique, et avec les erreurs types plus petites, le GLMM est aussi préférable au modèle GEE. Encore une fois, nous retrouvons que le GLMM est le meilleur.

## 7.5 Perspectives

Nous avons vu, aux deux paragraphes précédents, que le modèle linéaire généralisé mixte (GLMM) semble meilleur que le modèle GEE. Nous nous demandons si cela est dû à la structure de corrélation choisie (corrélation "échangeable") du modèle GEE. Nous cherchons alors à savoir si la structure de corrélation "non structurée" serait peut-être mieux adaptée. Rappelons que cette structure de corrélation est applicable s'il y a peu d'observations par individu (conducteur) et si chaque conducteur a le même nombre d'observations. Nous avons donc besoin d'une base de données où chaque conducteur a peu d'observations, et a le même nombre d'observations afin d'appliquer la structure de corrélation "non structurée". Pour cela, à partir des données du tirage aléatoire (cf. la section 7.1), nous avons pris d'abord 20 observations par conducteur et modélisons le modèle GEE ayant la structure de corrélation "non structurée" avec ces données. Malheureusement, avec ces données, le modèle n'a pas convergé. On a alors essayé de diminuer le nombre d'observations jusqu'à la convergence de ce modèle. Finalement, avec 14 observations par conducteur, le modèle converge. Nous utilisons ces données pour modéliser quatre modèles à deux variables explicatives ("ModeLavia" et "ConsigneLimiteVitesse") : le modèle logistique, le modèle GEE avec la structure de corrélation "échangeable", le modèle GEE avec la structure de corrélation "non structurée" et le modèle linéaire généralisé mixte contenant un effet aléatoire associé à l'intercept.

Il faut garder à l'esprit qu'ici, les résultats de l'impact du système LAVIA et des limi-

tations de vitesse sur le dépassement de vitesse volontaire ne sont pas corrects car on ne conserve qu'une petite partie des données.

Les estimations d'Odds Ratio obtenues des quatre modèles sont illustrées dans la figure 7.2. Tous les Odds Ratios estimés du modèle GEE avec la structure de corrélation "échangeable" ne sont pas significatifs ( $p$ -values supérieures à 5%). Le modèle logistique a 2 Odds Ratios significatifs ( $p$ -values inférieures à 5%). Il y en a 1 pour les deux modèles restant : le modèle GEE avec la structure de corrélation "non structurée" et le GLMM.

En général, les intervalles de confiance à 95% des Odds Ratios du modèle GEE avec la structure de corrélation "non structurée" sont plus étroits, et ceux des Odds Ratios du modèle GEE avec la structure de corrélation "échangeable" sont plus grands. Ainsi, le modèle GEE avec la structure de corrélation "non structurée" semble le meilleur au niveau de la précision de l'estimation.

Nous disposons aussi du critère **QIC** pour choisir la meilleure structure de corrélation pour le modèle GEE et du critère **QICu** pour sélectionner le meilleur modèle parmi des modèles GEE : **plus la valeur de QIC est petite, meilleur est le modèle**. Cependant, nous ne pouvons pas utiliser ce critère pour choisir le meilleur modèle parmi les deux modèles GEE puisque la valeur de QIC et celle de QICu pour les deux modèles GEE sont presque identiques (tableau 7.19).

	Critère d'ajustement GEE	
Type de corrélation	QIC	QICu
Échangeable	568.9372	565.7451
Non structurée	565.7048	566.4237

TABLE 7.19 – *Les critères de sélection du modèle.*

Les deux tableaux 7.20 et 7.21 représentent l'estimation des paramètres ainsi que les erreurs types des quatre modèles : le modèle logistique, le modèle GEE avec la structure de corrélation "échangeable", le modèle GEE avec la structure de corrélation "non structurée" et le modèle linéaire généralisé mixte (GLMM). Rappelons que les lignes correspondantes à une consigne de 110 km/h et au mode LAVIA Neutre sont les références.

Le modèle logistique a plus de paramètres significatifs (i.e.  $p$ -values inférieures à 5%) mais il ne tient pas compte de la corrélation entre les observations. Les trois autres modèles tiennent compte de cette corrélation. Parmi ces trois modèles, le modèle GEE avec la structure de corrélation "échangeable" a moins de paramètre significatif (un seul), les deux modèles restant ont deux paramètres significatifs. Ainsi, par rapport à la structure de corrélation "échangeable", la corrélation "non structurée" semble mieux adaptée.

Concernant les erreurs types, celles du modèle GEE avec la structure de corrélation "échangeable" sont les plus grandes, sauf celle du paramètre "ConsigneLimiteVitesse\_70". Les erreurs types les plus petites sont partagées entre le modèle logistique et le modèle GEE avec la structure de corrélation "non structurée". Cependant, le modèle logistique suppose l'indépendance entre les observations. Nous comparons donc les erreurs types entre

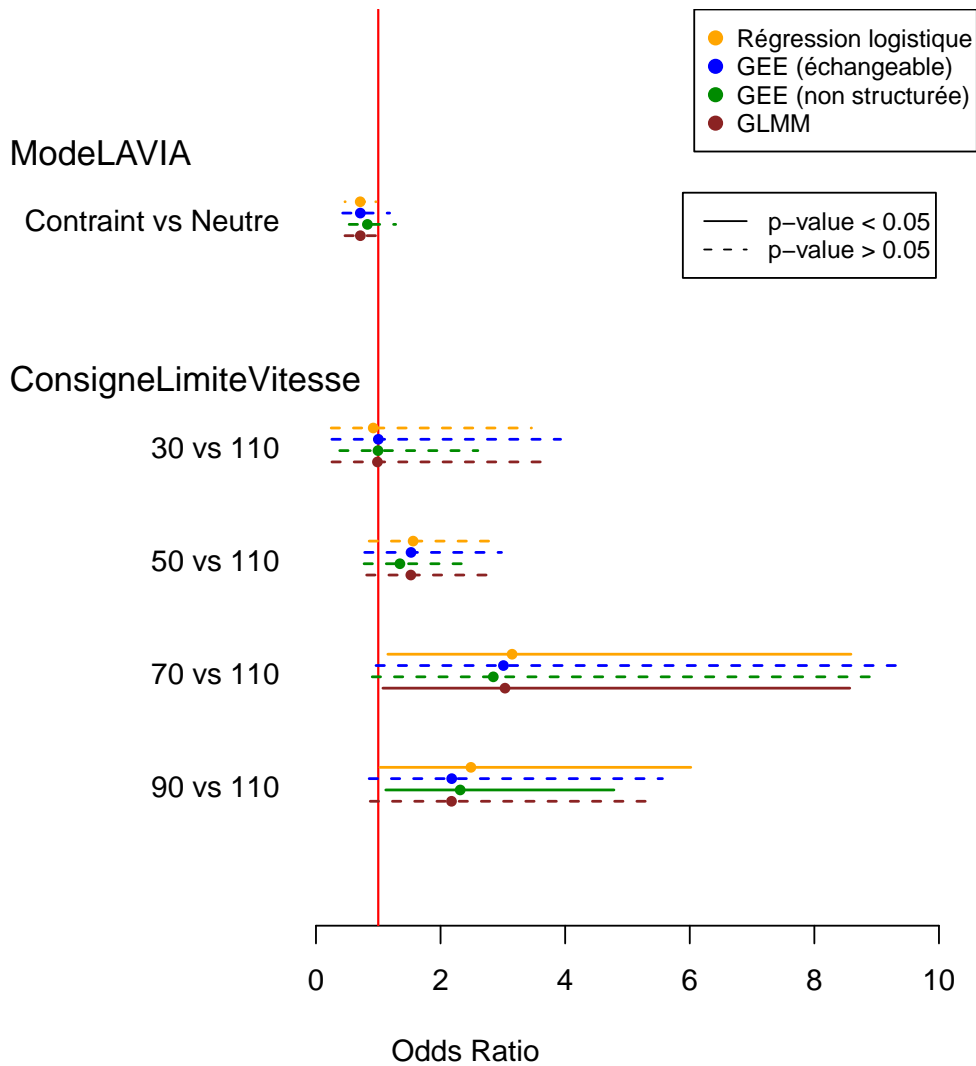


FIGURE 7.2 – Graphique des Odds Ratios de quatre modèles (logistique, GEE avec la structure de corrélation "échangeable", GEE avec la structure de corrélation "non structurée" et linéaire généralisé mixte) pour le dépassement de vitesse volontaire. La barre rouge correspond à la valeur neutre (i.e 1) d'Odds Ratio.

Paramètre	Estimation			
	Modèle logistique	GEE Échangeable	GEE Non structurée	GLMM
Intercept	-1.7417***	-1.7184***	-1.6909***	-1.7419***
ConsigneLimiteVitesse_30	-0.0840	0.0011	-0.0041	-0.01336
ConsigneLimiteVitesse_50	0.4444	0.4228	0.2980	0.4204
ConsigneLimiteVitesse_70	1.1468*	1.1015	1.0461	1.1100*
ConsigneLimiteVitesse_90	0.9112*	0.7788	0.8394*	0.7772
ConsigneLimiteVitesse_110	0.0000	0.0000	0.0000	0.0000
ModeLavia_Constraint	-0.3420	-0.3397	-0.1917	-0.3392
ModeLavia_Neutre	0.0000	0.0000	0.0000	0.0000

En rouge :  $p$ -value  $\geq 0.05$  ; \*  $p$ -value  $< 0.05$  ; \*\*  $p$ -value  $< 0.01$  ; \*\*\*  $p$ -value  $< 0.001$

TABLE 7.20 – Estimation des paramètres de quatre modèles : le modèle logistique, le modèle GEE avec la structure de corrélation "échangeable", le modèle GEE avec la structure de corrélation "non structurée" et le modèle linéaire généralisé mixte (GLMM). Les valeurs en rouge ne sont pas significativement différentes de 0.

Paramètre	Erreur type			
	Modèle logistique	GEE Échangeable	GEE Non structurée	GLMM
Intercept	<b>0.2903</b>	0.3527	0.3182	0.3082
ConsigneLimiteVitesse_30	0.6761	0.6973	<b>0.4893</b>	0.6861
ConsigneLimiteVitesse_50	0.3073	0.3411	<b>0.2838</b>	0.3190
ConsigneLimiteVitesse_70	<b>0.5120</b>	0.5805	0.5857	0.5284
ConsigneLimiteVitesse_90	0.4507	0.4781	<b>0.3702</b>	0.4658
ConsigneLimiteVitesse_110				
ModeLavia_Constraint	<b>0.2158</b>	0.2589	0.2234	0.2171
ModeLavia_Neutre				

TABLE 7.21 – Estimation des erreurs types de quatre modèles : le modèle logistique, le modèle GEE avec la structure de corrélation "échangeable", le modèle GEE avec la structure de corrélation "non structurée" et le modèle linéaire généralisé mixte (GLMM). Les erreurs types en bleu sont les plus petites.

les trois modèles tenant compte de la corrélation entre les observations sur les données (les deux modèles GEE et le GLMM). Parmi ces trois modèles, les erreurs types les plus petites sont partagées entre le modèle GEE avec la structure de corrélation "non structurée" et le GLMM. Le meilleur modèle est ainsi discuté entre ces deux modèles au niveau de l'erreur type.

Après ces comparaisons de l'estimation des Odds Ratios, des paramètres et des erreurs types sur les modèles, nous trouvons que pour le modèle GEE, la structure de corrélation "non structurée" est mieux adapté que la structure de corrélation "échangeable". Ainsi, nous pouvons dire qu'**avec nos données du tirage aléatoire (section 7.1), le GLMM est meilleur que le modèle GEE avec la corrélation "échangeable", cela peut être dû à la structure de corrélation choisie (corrélation "échangeable") du modèle GEE.** Si nous pouvons appliquer la structure de corrélation "non structurée", le modèle GEE pourrait être amélioré.

## 7.6 Conclusion

Nous avons donc apporté la preuve statistique que le système LAVIA en mode Contraint réduit significativement la probabilité d'occurrence des excès de vitesse volontaires (i.e. de plus de 10 secondes et de plus de 4 km/h par rapport à la vitesse limite). En effet, comme nous avons vu, quelque soit la méthode, l'analyse de tableau de contingence, la régression logistique, les équations d'estimation généralisées (GEE) ou le modèle linéaire généralisé mixte (GLMM), nous avons toujours obtenu ce résultat. Il confirme les analyses au chapitre 4.

Nous avons vu aussi que par rapport à la limitation de vitesse 110 km/h, les limitations de vitesse inférieures (30, 50, 70, 90 km/h) augment la probabilité d'occurrence des dépassements de vitesse volontaires. Ainsi, l'effet du système LAVIA sur l'occurrence des excès de vitesse volontaires est plus marqué dans les zones limitées à 110 km/h.

## Chapitre 8

# Applications aux données LAVIA : Analyse de l'impact de tous les modes LAVIA sur le dépassement de vitesse volontaires

Au chapitre 7, nous avons étudié l'impact du système LAVIA et des limitations de vitesse sur le dépassement de vitesse volontaire. Pour cette analyse, nous avons considéré seulement deux modes LAVIA : **Neutre** (aucun fonctionnement du système LAVIA) et **Contraint** (avec activation permanente du limiteur de vitesse, sauf usage du "kickdown"). Dans le cadre de ce chapitre, nous modélisons aussi la probabilité d'occurrence des dépassements de vitesse volontaires mais ici, notre objectif est d'analyser non seulement l'impact du mode LAVIA **Contraint** mais aussi l'impact des autres modes LAVIA : **Informatif** (affichage de la vitesse limite, avertissement d'un dépassement), **Débrayable** (avec activation, si le conducteur le décide, du limiteur de vitesse) sur cet événement sans l'effet des limitations de vitesse. Pour cela, nous utilisons les mêmes méthodes que précédemment : le tableau de contingence, la régression logistique, les équations d'estimation généralisées (GEE) et les modèles mixtes. La variable réponse reste inchangée (variable "*DepassType2*", cf. le chapitre 7). Cependant, il y a une seule variable explicative ("*ModeLAVIA*") et cette fois-ci, cette variable explicative a quatre modalités (Informatif, Débrayable, Contraint et Neutre) au lieu de deux comme avant. Il nous manque donc des données pour les deux modes LAVIA Informatif et Débrayable. Pour chacun de ces deux modes LAVIA, nous utilisons la définition de dépassement volontaire (type 2) définie à la page 20 pour détecter l'événement de dépassement de vitesse volontaire et nous réutilisons ensuite l'algorithme de tirage aléatoire décrit à la section 7.1 du chapitre 7 (page 37). Nos données, après avoir fait ces tirages aléatoires, contiennent cinq variables ("*Travia*", "*Trajet*", "*ModeLAVIA*", "*ConsigneLimite-Vitesse*", "*DepassType2*") et 40000 observations dont 10000 observations pour chaque mode LAVIA (Informatif, Débrayable, Contraint et Neutre). Nous les utilisons pour modéliser la probabilité d'avoir un dépassement de vitesse volontaire dans un intervalle de temps de trente secondes avec une seule variable explicative "*ModeLAVIA*" afin d'analyser l'impact de chacun des quatre modes LAVIA sur le dépassement de vitesse volontaire.

Nous faisons d'abord des analyses descriptives de ces données par tableau de contingence. Ensuite, nous utilisons la régression logistique, les GEE, le modèle linéaire généralisé

(GLMM) pour modéliser.

Rappelons que nos données sont corrélées puisqu'il y a plusieurs observations (trajets) pour un conducteur donné.

## 8.1 Tableau de contingence de "DepassType2" et "ModeLAVIA"

Le tableau de contingence de "*DepassType2*" et "*ModeLAVIA*" et l'estimation des Odds Ratios (rapports de cotes) des modes LAVIA Informatif, Debrayable et Contraint par le mode LAVIA Neutre sont présentés respectivement dans les tableaux 8.1 et 8.2.

Par rapport au mode LAVIA Neutre, nous avons vu à la section 7.2 du chapitre 7 (page 38) que le mode LAVIA Contraint diminue le risque de dépassement de vitesse volontaire. C'est aussi le résultat que nous obtenons ici. Les deux autres modes LAVIA (Informatif, Debrayable) diminuent aussi ce risque. Le mode Debrayable est associé au risque le plus faible et le mode Informatif est associé au risque le plus élevé. En effet, d'après le tableau 8.1, la fréquence du dépassement de vitesse volontaire est la plus élevée en mode Neutre (22.02% en mode Neutre et que 16.94%, 10.83%, 12.77% respectivement en mode Informatif, Debrayable et Contraint) et elle est la plus basse en mode Debrayable (pourcentage le plus petit). De plus, d'après le tableau 8.2, les Odds Ratios associés sont inférieurs à 1 (0.7222, 0.4301, 0.5184), l'Odds Ratio du mode Debrayable est le plus petit et celui du mode Informatif est le plus grand. Par rapport au mode Contraint, le mode Informatif augmente d'environ 39% (Odds Ratios = 1.3931) tandis que le mode Debrayable diminue d'environ 17% le risque d'excès de vitesse (Odds Ratio = 0.8296). On observe aussi que le mode Informatif augmente d'environ 68% ce risque par rapport au mode Debrayable.

Parmi les trois modes LAVIA Informatif, Debrayable et Contraint, le risque le plus élevé est associé au mode Informatif. Cela semble normal, parce que le limiteur de vitesse n'est jamais activé en mode Informatif. Le risque le plus faible est associé au mode Debrayable mais pas au mode Contraint. Cela peut être dû à l'effet d'apprentissage du mode Debrayable.

Cela se traduit par le fait qu'en mode Debrayable, les conducteurs apprennent à utiliser le système tandis qu'en mode Contraint, ils ont déjà expérimenté le système, et en connaissent donc l'usage.

## 8.2 Modélisation de la probabilité d'occurrence des dépassements de vitesse volontaires dans un intervalle de temps de trente secondes avec la variable "ModeLAVIA"

Ici, nous modélisons la probabilité d'occurrence des dépassements de vitesse volontaires dans un intervalle de temps de trente secondes par plusieurs méthodes afin d'analyser l'effet des quatre modes LAVIA (Informatif, Debrayable, Contraint et Neutre) sur le dépassement de vitesse. Les méthodes de modélisation que nous utiliserons sont toujours les mêmes : la régression logistique, les GEE et le GLMM (elles ont été représentées au chapitre 5 et au chapitre 6). Nous reprenons ici toutes les notations précédentes.



DepassType2	ModeLAVIA				
Fréquence Pourcentage Pctage en ligne Pctage en col.	Informatif	Debrayable	Contraint	Neutre	Total
<b>1</b>	1694 4.24 27.08 <b>16.94</b>	1083 2.71 17.31 <b>10.83</b>	1277 3.19 20.41 <b>12.77</b>	2202 5.51 35.20 <b>22.02</b>	6256 15.64
<b>0</b>	8306 20.77 24.61 83.06	8917 22.29 26.43 89.17	8723 21.81 25.85 87.23	7798 19.50 23.11 77.98	33744 84.36
<b>Total</b>	10000 25.00	10000 25.00	10000 25.00	10000 25.00	40000 100.00

TABLE 8.1 – *Tableau de contingence de "DepassType2" et "ModeLAVIA" (les données du tirage aléatoire contenant quatre modes LAVIA).*

ModeLAVIA	Odds Ratio	IC 95%	
<b>Informatif VS Neutre</b>	<b>0.7222</b>	0.6731	0.7750
<b>Debrayable VS Neutre</b>	<b>0.4301</b>	0.3975	0.4654
<b>Contraint VS Neutre</b>	<b>0.5184</b>	0.4808	0.5590
<b>Debrayable VS Contraint</b>	<b>0.8296</b>	0.7611	0.9043
<b>Informatif VS Contraint</b>	<b>1.3931</b>	1.2878	1.5071
<b>Informatif VS Debrayable</b>	<b>1.6792</b>	1.5472	1.8226

TABLE 8.2 – *Estimation des Odds Ratios (rapports de cotes) à partir des données du tirage aléatoire contenant quatre modes LAVIA pour le dépassement de vitesse volontaire.*

### 8.2.1 Modélisation par régression logistique

Nous rappelons qu'ici, on suppose que les observations sont indépendantes entre elles et que le modèle s'écrit comme suivant :

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = X'_{ij}\beta = \beta_0 + \beta_1 X_{ij1} + \dots + \beta_K X_{ijK}.$$

Comme la variable "ModeLAVIA" est qualitative, il y aura un paramètre pour chaque modalité de "ModeLAVIA".

Les tableaux 8.3, 8.4 et 8.5 représentent les résultats obtenus à l'aide de la procédure GENMOD de SAS (sans option REPEATED).

Paramètre	Estimation	Erreur type	IC 95%		p-value
<b>Intercept</b>	-1.2645	0.0241	-1.3120	-1.2174	<.0001
<b>ModeLavia _ Informatif</b>	-0.3254	0.0360	-0.3959	-0.2550	<.0001
<b>ModeLavia _ Debrayable</b>	-0.8437	0.0402	-0.9228	-0.7651	<.0001
<b>ModeLavia _ Contraint</b>	-0.6569	0.0385	-0.7325	-0.5817	<.0001
<b>ModeLavia _ Neutre</b>	0.0000				

TABLE 8.3 – Estimation des paramètres du modèle logistique à une variable explicative ("ModeLAVIA").

ModeLAVIA	Odds Ratio	IC 95%		p-value
<b>Informatif VS Neutre</b>	0.7222	0.6731	0.7750	<.0001
<b>Debrayable VS Neutre</b>	0.4301	0.3975	0.4654	<.0001
<b>Contraint VS Neutre</b>	0.5184	0.4808	0.5590	<.0001
<b>Debrayable VS Contraint</b>	0.8296	0.7611	0.9043	<.0001
<b>Informatif VS Contraint</b>	1.3931	1.2878	1.5071	<.0001
<b>Informatif VS Debrayable</b>	1.6792	1.5472	1.8226	<.0001

TABLE 8.4 – Estimation des Odds Ratios (rapports de cotes) du modèle logistique à une variable explicative ("ModeLAVIA").

Effet	DDL	Khi-2	p-value
<b>ModeLAVIA</b>	3	551.50	<.0001

TABLE 8.5 – Test global de la variable explicative "ModeLAVIA" du modèle logistique.

La ligne correspondante au mode LAVIA Neutre du tableau 8.3 est la référence, à partir de laquelle est calculée la probabilité d'occurrence des dépassements de vitesse volontaires. Ainsi, cette probabilité avec le système LAVIA en mode Informatif diminue par rapport à l'usage du mode Neutre puisque le paramètre associé est négatif (-0.3254). C'est d'ailleurs le cas pour les deux autres modes LAVIA Debrayable et Contraint. Les excès de vitesse sont le plus probable en mode Informatif et ils sont le moins probable en mode Debrayable. En plus, toutes les  $p$ -values correspondantes sont inférieures à 5% ce qui signifie que tous les paramètres estimés sont significativement différent de 0. Tous ces effets sont donc significatifs. L'estimation des Odds Ratios (tableau 8.4) confirment ces résultats. De plus, par rapport au mode Contraint, le mode Informatif augmente la probabilité d'occurrence des excès de vitesse volontaires et c'est le contraire pour le mode Debrayable. Les résultats obtenus ici sont similaires à ceux nous avons obtenus au paragraphe précédent.

D'après le tableau 8.5, la variable explicative "ModeLAVIA" est significative.

### 8.2.2 Modélisation par équations d'estimation généralisées (GEE)

Les résultats obtenus par le tableau de contingence et par la régression logistique dans les deux paragraphes précédents sont similaires. Cependant, ces deux méthodes ne tiennent pas compte de la corrélation entre les observations d'un même conducteur (corrélation intra-classe). Ainsi, les résultats obtenus peuvent être incorrects. C'est pour cela que les équations d'estimation généralisées GEE et le modèle linéaire généralisé mixte (GLMM) qui tiennent compte de cette corrélation sont utilisés. Nous présentons ici les résultats obtenus par les GEE et le paragraphe suivant sera consacré aux résultats obtenus par le GLMM.

Nous rappelons que le modèle est comme le modèle logistique :

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = X'_{ij}\beta = \beta_0 + \beta_1 X_{ij1} + \dots + \beta_K X_{ijK}$$

mais ici, la corrélation intra-classe est prise en compte.

Les résultats obtenus avec la structure de corrélation "échangeable" à l'aide de la procédure GENMOD de SAS avec l'option REPEATED sont représentés aux tableaux 8.6, 8.7, 8.8 et la corrélation échangeable est estimée à 0.0568 qui est petit.

La variable "ModeLAVIA" est significative (tableau 8.8).

Paramètre	Estimation	Erreur type	IC 95%		p-value
<b>Intercept</b>	-1.2535	0.1017	-1.4528	-1.0543	<.0001
<b>ModeLavia_ Informatif</b>	-0.2694	0.0909	-0.4475	-0.0913	0.0030
<b>ModeLavia_ Debrayable</b>	-0.7954	0.0855	-0.9631	-0.6278	<.0001
<b>ModeLavia_ Contraint</b>	-0.5883	0.0870	-0.7588	-0.4178	<.0001
<b>ModeLavia_ Neutre</b>	0.0000				

TABLE 8.6 – Estimation des paramètres du modèle GEE à une variable explicative ("ModeLAVIA").

ModeLAVIA	Odds Ratio	IC 95%		p-value
<b>Informatif VS Neutre</b>	0.7639	0.6392	0.9128	0.0030
<b>Debrayable VS Neutre</b>	0.4514	0.3817	0.5338	<.0001
<b>Contraint VS Neutre</b>	0.5553	0.4682	0.6585	<.0001
<b>Debrayable VS Contraint</b>	0.8129	0.6987	0.9458	0.0073
<b>Informatif VS Contraint</b>	1.3756	1.1930	1.5862	<.0001
<b>Informatif VS Debrayable</b>	1.6922	1.4394	1.9896	<.0001

TABLE 8.7 – Estimation des Odds Ratios (rapports de cotes) du modèle GEE à une variable explicative ("ModeLAVIA").

Effet	DDL	Khi-2	p-value
<b>ModeLAVIA</b>	3	28.50	<.0001

TABLE 8.8 – Test global de la variable explicative "ModeLAVIA" du modèle GEE.

La ligne correspondante au mode LAVIA Neutre du tableau 8.6 est toujours la référence. Les paramètres estimés sont tous négatifs et les  $p$ -values associées sont toutes inférieures à 5% comme dans le modèle logistique. Cela signifie que nous obtenons les mêmes résultats. Par rapport le mode Neutre, tous les autres modes LAVIA diminuent significativement la probabilité d'occurrence des excès de vitesse volontaires. Cet effet est plus remarquable en mode Debrayable et en mode Contraint, et il est le plus remarquable en mode Debrayable (coefficient négatif et le plus petit). L'estimation des Odds Ratios confirment ces résultats.

### 8.2.3 Modélisation par modèle linéaire généralisé mixte (GLMM)

Nous rappelons que les modèles linéaires généralisés mixtes (GLMM) sont une des méthodes pour analyser les données corrélées. Le modèle est :

$$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = (\beta_0 + b_{0i}) + \beta_1 X_{ij1} + \dots + \beta_K X_{ijK}$$

où  $b_{0i} \sim N(0, \sigma^2)$ .

Nous rappelons également que l'effet aléatoire  $b_{0i}$  associé à l'intercept  $\beta_0$  traduit le fait que la probabilité d'occurrence des dépassements de vitesse volontaires varie entre les conducteurs.

Les résultats obtenus à l'aide de la procédure GLIMMIX de SAS sont présentés dans les tableaux 8.9, 8.10, 8.11.

La variance  $\sigma^2$  de l'effet aléatoire  $b_{0i}$  associé à l'intercept est estimée à 0.5241.

La référence de l'estimation des paramètres est toujours le mode LAVIA Neutre. Tous les paramètres sont significatifs ( $p$ -values inférieures à 1). Les paramètres estimés sont encore tous négatifs; indiquant une réduction de la probabilité d'occurrence des excès de vitesse volontaires de tous autres les modes LAVIA par rapport au mode Neutre. Le paramètre

Paramètre	Estimation	Erreur type	IC 95%		p-value
<b>Intercept</b>	-1.3820	0.1124	-1.6086	-1.1554	<.0001
<b>ModeLavia_ Informatif</b>	-0.2885	0.03806	-0.3631	-0.2139	<.0001
<b>ModeLavia_ Debrayable</b>	-0.8553	0.04207	-0.9378	-0.7729	<.0001
<b>ModeLavia_ Contraint</b>	-0.6360	0.04067	-0.7157	-0.5563	<.0001
<b>ModeLavia_ Neutre</b>	0.0000				

TABLE 8.9 – Estimation des paramètres du modèle linéaire généralisé mixte à une variable explicative ("ModeLAVIA").

ModeLAVIA	Odds Ratio	IC 95%		p-value
<b>Informatif VS Neutre</b>	0.7494	0.6955	0.8074	<.0001
<b>Debrayable VS Neutre</b>	0.4251	0.3915	0.4617	<.0001
<b>Contraint VS Neutre</b>	0.5294	0.4889	0.5734	<.0001
<b>Debrayable VS Contraint</b>	0.8030	0.7339	0.8786	<.0001
<b>Informatif VS Contraint</b>	1.4155	1.3034	1.5372	<.0001
<b>Informatif VS Debrayable</b>	1.7627	1.6175	1.9209	<.0001

TABLE 8.10 – Estimation des Odds Ratios (rapports de cotes) du modèle linéaire généralisé mixte à une variable explicative ("ModeLAVIA").

Effet	DDL Num.	DDL Res.	Valeur F	p-value
<b>Intercept</b>	1	43	273.46	<.0001
<b>ModeLAVIA</b>	3	39953	166.10	<.0001

TABLE 8.11 – Test global de la variable explicative "ModeLAVIA" du modèle linéaire généralisé mixte.

le plus petit (risque le plus bas) et celui le plus grand (risque le plus élevé) sont toujours respectivement associés au mode Debrayable et au mode Informatif. Les Odds Ratios estimés nous donnent aussi les mêmes résultats.

Il n'y a donc pas de différence entre les résultats des trois modèles (le modèle logistique, le modèle GEE et le GLMM) sur l'impact des modes LAVIA sur le dépassement de vitesse volontaire.

## 8.3 Comparaison des modèles

### 8.3.1 Comparaison des Odds Ratios

Les estimations d'Odds Ratio obtenues à partir de quatre méthodes utilisées (le tableau de contingence, le modèle logistique, le modèle GEE et le modèle linéaire généralisé mixte (GLMM)) sont présentées dans les tableaux 8.12, 8.13 et aussi illustrées dans la figure 8.1. Les écarts entre les Odds Ratios de ces quatre méthodes sont petits et les résultats des quatre méthodes convergent. Cela se traduit par le fait que la corrélation entre les observations d'un conducteur est très petit.

Les Odds Ratios des quatre méthodes sont tous significativement différent de 1 puisque ses intervalles de confiance à 95% n'incluent pas la valeur neutre 1 (figure 8.1). Cela signifie que les modes du LAVIA et les limitations de vitesse ont des effets significativement différents sur la probabilité d'occurrence des excès de vitesse volontaires.

La probabilité d'occurrence des excès de vitesse volontaires diminue dès l'utilisation du mode Informatif. Cet effet est plus marqué pour les modes Debrayable et Contraint. En effet, les Odds Ratios du mode LAVIA Informatif, du mode LAVIA Debrayable et du mode LAVIA Contraint par le mode LAVIA Neutre sont tous inférieurs à 1 pour les quatre méthodes, ce qui signifie que par rapport au mode Neutre, ces trois modes LAVIA diminuent la probabilité d'occurrence des excès de vitesse volontaires. Parmi ces trois modes LAVIA, le mode Informatif est associé à un risque plus élevé (Odds Ratio associé le plus élevé pour chaque méthode), ce risque est plus petit lors de l'utilisation du mode Debrayable (Odds Ratio associé le plus petit pour chaque méthode, réduction de 57% pour l'Odds Ratio brut et le modèle logistique, et de 55% et 57% respectivement pour le modèle GEE et le GLMM). Par rapport au mode Contraint, d'après l'Odds Ratio brut et celui estimé par le modèle logistique, l'usage du mode Debrayable diminue la probabilité d'occurrence des excès de vitesse volontaires d'environ 17% (les Odds Ratios correspondant sont égaux à 0.8296) et il le diminue d'environ 19% et 20% respectivement pour le modèle GEE et le GLMM (Odds Ratio = 0.8129 et Odds Ratio = 0.8030 respectivement). De plus, ces effets sont significatifs puisque les  $p$ -values associées de quatre méthodes sont tous inférieures à 5%.

Les Odds Ratios bruts et ceux du modèle logistique sont identiques. Cela peut être dû à la même hypothèse d'indépendance entre les observations de ces deux méthodes. Le modèle GEE et le GLMM prennent en compte la corrélation de nos données. Le GLMM ajoute en plus un effet aléatoire traduisant le fait que la probabilité d'occurrence des dépassements de vitesse volontaires varie entre les conducteurs. Cela induit que les Odds Ratios de ces deux modèles sont plus ou moins différents. Par exemple, les Odds Ratios des trois modes Informatif, Debrayable et Contraint par le mode Neutre du modèle GEE sont plus grands que ceux du GLMM. La précision de l'estimation obtenue par le modèle GEE est plus faible.

Les intervalles de confiance à 95% des Odds Ratios du modèle GEE sont les plus grands signifiant la précision la plus faible de l'estimation obtenue par le modèle GEE. Ayant une meilleure hypothèse que le modèle logistique et ayant une meilleure précision que le modèle GEE, le GLMM semble préférable.

ModeLAVIA	Tableau de contingence Odds Ratio brut			Modèle logistique			p-value
	Odds Ratio	IC 95%		Odds Ratio	IC 95%		
<b>Informatif vs Neutre</b>	<b>0.7222</b>	0.6731	0.7750	<b>0.7222</b>	0.6731	0.7750	<.0001
<b>Debrayable vs Neutre</b>	<b>0.4301</b>	0.3975	0.4654	<b>0.4301</b>	0.3975	0.4654	<.0001
<b>Contraint vs Neutre</b>	<b>0.5184</b>	0.4808	0.5590	<b>0.5184</b>	0.4808	0.5590	<.0001
<b>Debrayable vs Contraint</b>	<b>0.8296</b>	0.7611	0.9043	<b>0.8296</b>	0.7611	0.9043	<.0001
<b>Informatif vs Contraint</b>	<b>1.3931</b>	1.2878	1.5071	<b>1.3931</b>	1.2878	1.5071	<.0001
<b>Informatif vs Debrayable</b>	<b>1.6792</b>	1.5472	1.8226	<b>1.6792</b>	1.5472	1.8226	<.0001

TABLE 8.12 – Les Odds Ratios bruts (le tableau de contingence) et les Odds Ratios estimés du modèle logistique à une variable explicative "ModeLAVIA" pour le dépassement de vitesse volontaire.

ModeLAVIA	GEE				GLMM			
	Odds Ratio	IC 95%		p-value	Odds Ratio	IC 95%		p-value
<b>Informatif vs Neutre</b>	<b>0.7639</b>	0.6392	0.9128	0.0030	<b>0.7494</b>	0.6955	0.8074	<.0001
<b>Debrayable vs Neutre</b>	<b>0.4514</b>	0.3817	0.5338	<.0001	<b>0.4251</b>	0.3915	0.4617	<.0001
<b>Contraint vs Neutre</b>	<b>0.5553</b>	0.4682	0.6585	<.0001	<b>0.5294</b>	0.4889	0.5734	<.0001
<b>Debrayable vs Contraint</b>	<b>0.8129</b>	0.6987	0.9458	0.0073	<b>0.8030</b>	0.7339	0.8786	<.0001
<b>Informatif vs Contraint</b>	<b>1.3756</b>	1.1930	1.5862	<.0001	<b>1.4155</b>	1.3034	1.5372	<.0001
<b>Informatif vs Debrayable</b>	<b>1.6922</b>	1.4394	1.9896	<.0001	<b>1.7627</b>	1.6175	1.9209	<.0001

TABLE 8.13 – Les Odds Ratios estimés du modèle GEE à une variable explicative "ModeLAVIA" et du modèle linéaire généralisé mixte (GLMM) à une variable explicative "ModeLAVIA".

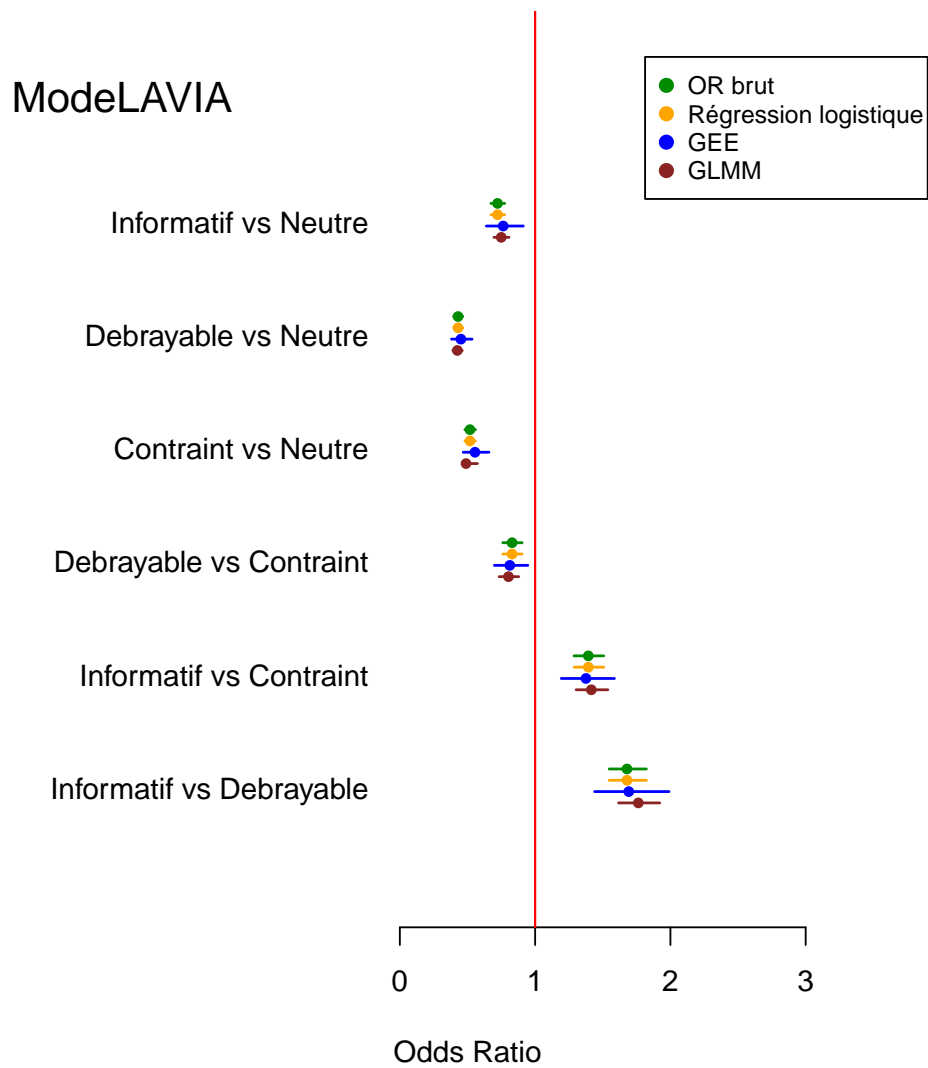


FIGURE 8.1 – Graphique des Odds Ratios bruts et des Odds Ratios de trois modèles (logistique, GEE et linéaire généralisé mixte) à une variable explicative "ModeLAVIA" pour le dépassement de vitesse volontaire. La barre rouge correspond à la valeur neutre (i.e 1) d'Odds Ratio.



### 8.3.2 Comparaison de l'estimation des paramètres et des erreurs types

L'estimation des paramètres ainsi que les erreurs types de trois modèles : le modèle logistique, le modèle GEE et le modèle linéaire généralisé mixte (GLMM) sont présentées dans les tableaux 8.14 et 8.15. Tous les paramètres sont négatifs pour les trois méthodes. Cela signifie que par rapport au mode Neutre, les dépassements de vitesse volontaires sont moins probables avec l'usage des autres modes LAVIA. Avec l'usage du mode Débrayable, le risque est le plus petit. De plus, ces impacts sont significatifs car les  $p$ -values sont toutes inférieures à 0.05 pour les trois méthodes.

Concernant les erreurs types, on observe encore que celles du modèle logistique sont les plus petites. Cette sous-estimation peut être dû à l'hypothèse d'indépendance entre les observations de ce modèle. Comme nos données sont corrélées (plusieurs observations pour un conducteur), le modèle GEE et le GLMM sont meilleurs que le modèle logistique car leurs hypothèses sont meilleures (tient compte de la corrélation intra-conducteur). De plus, par rapport au modèle GEE, les erreurs types du GLMM sont en général plus petites. Anisi, nous pouvons dire que le GLMM semble le meilleur.

Paramètre	Estimation		
	Modèle logistique	GEE	GLMM
Intercept	-1.2645***	-1.2535***	-1.3820***
ModeLavia_Informatif	-0.3254***	-0.2694**	-0.2885***
ModeLavia_Débrayable	-0.8437***	-0.7954***	-0.8553***
ModeLavia_Contraint	-0.6569***	-0.5883***	-0.6360***
ModeLavia_Neutre	0.0000	0.0000	0.0000

\*  $p$ -value < 0.05; \*\*  $p$ -value < 0.01; \*\*\*  $p$ -value < 0.001

TABLE 8.14 – Estimation des paramètres de trois modèles (logistique, GEE et linéaire généralisé mixte) à une variable explicative "ModeLAVIA".

Paramètre	Erreur type		
	Modèle logistique	GEE	GLMM
Intercept	0.0241	0.1017	0.1124
ModeLavia_Informatif	0.0360	0.0909	0.03806
ModeLavia_Débrayable	0.0402	0.0855	0.04207
ModeLavia_Contraint	0.0385	0.0870	0.04067
ModeLavia_Neutre			

TABLE 8.15 – Estimation des erreurs types de trois modèles (logistique, GEE et linéaire généralisé mixte) à une variable explicative "ModeLAVIA".

## 8.4 Conclusion

Le modèle linéaire généralisé mixte (GLMM) semble le meilleur. Il est préférable au modèle logistique au niveau de l'hypothèse (le GLMM tient compte de la corrélation entre les observations), et il est préférable au modèle GEE au niveau de la qualité d'estimation (erreurs types plus petites et intervalles de confiance à 95% des Odds Ratios plus petits).

L'usage du système LAVIA diminue la probabilité d'occurrence des dépassements de vitesse volontaires. En effet, tous les trois modes Informatif, Debrayable et Contraint, chacun est associé à un risque plus petit que le mode Neutre. Cet effet du système LAVIA sur l'occurrence des dépassements de vitesse volontaires est plus marqué en mode Debrayable.

## Chapitre 9

# Conclusion générale

Ce travail montre clairement que le système LAVIA permet un gain en terme de sécurité routière. Il augmente la sécurité en diminuant la fréquence des excès de vitesse volontaires ainsi que leur durée et la distance parcourue en excès de vitesse. Des analyses approfondies sur l'impact de ce système sur le dépassement de vitesse volontaire comme le modèle logistique, les équations d'estimation généralisées, les modèles mixtes ont été effectuées. Nous avons commencé par le modèle logistique. Cependant, comme les données utilisées présentent une corrélation, il n'est pas tout à fait rigoureux d'utiliser le modèle logistique qui suppose l'indépendance entre les observations. Cela nous a poussé à chercher d'autres méthodes pour des données corrélées. Nous avons orienté les recherches vers les équations d'estimation généralisées et les modèles mixtes. Après les comparaisons des modèles à partir des résultats obtenus, nous constatons que le modèle linéaire généralisé mixte (GLMM) semble le meilleur pour notre application. Nous avons vu aussi que cela peut être dû à la structure de corrélation choisie (corrélation "échangeable") du modèle GEE, il est donc probable que la structure de corrélation "non structurée" soit la mieux adaptée. Malheureusement, nous n'avons pas pu appliquer cette structure de corrélation à nos données car il faut avoir le même nombre d'observations par conducteur, ce qui n'est pas le cas pour nos données.

Les interprétations que l'on peut tirer des résultats des trois modèles obtenus (le modèle logistique, le modèle GEE avec la structure de corrélation "échangeable" et le modèle linéaire généralisé mixte) sont similaires. Cela s'explique probablement par le fait que la corrélation intra-conducteurs est petite. Cependant, si on regarde en détail ces résultats, on constate que le tableau de contingence et le modèle logistique sous estiment les effets des facteurs et leur variabilité (à cause de l'hypothèse d'indépendance entre les observations de ce modèle).

En terme de résultats, le système LAVIA a porté ses fruits puisque des impacts positifs ont été observés sur le dépassement de vitesse volontaire (i.e. de plus de 10 secondes et de plus de 4 km/h par rapport à la vitesse limite). En effet, le système LAVIA diminue la probabilité d'occurrence des dépassements de vitesse volontaires dès l'usage du mode Informatif. Cet effet est plus marqué en mode Debrayable et en mode Contraint. Il est intéressant de noter que le mode Debrayable a un impact plus marqué sur le dépassement de vitesse volontaire. Cela est vraisemblablement dû à l'effet d'apprentissage induisant une réduction de ce gain lors du passage au mode contraint.

Concernant les limitations de vitesse, l'effet du système LAVIA en mode Contraint sur l'occurrence des excès de vitesse volontaires peut être différent selon la consigne de limitation de vitesse. Cette hypothèse a été testée via l'introduction de l'interaction entre ces deux variables dans le modèle. L'absence de significativité a montré que les effets du système LAVIA sur l'occurrence des dépassements de vitesse sont constants sur tous types de routes (toutes consignes de limitation de vitesse).

# Bibliographie

- Demidenko E.** (2004). *Mixed models : Theory and applications*. John Wiley & Son, Inc.
- Driscoll R., Page Y., Lassarre S. et Ehrlich J.** (2007). *Lavia, an evaluation of the potential safety benefits of the french intelligent speed adaptation project*. 51st annual proceeding of the association for the advancement of automotive medicine. pages 485–505.
- Droesbeke J.-J., Lejeune M. et Saporta G.** (2005). *Modèles statistiques pour données qualitatives*. Editions TECHNIP.
- EuroFOT consortium** (2009), *Performance indicators, events and situational variables within eurofot*. Rapport technique.
- FESTA consortium** (2008), *Festa handbook*. Rapport technique.
- Guo F. et Hankey J.** (2009), *Modeling 100-car safety events : A case-based approach for analyzing naturalistic driving data*. Rapport technique.
- Hu F. B., Goldberg J., Hedeker D., Flay B. R. et Pentz M. A.** (1998). *Comparison of population-averaged and subject-specific approaches for analysing repeated binary outcomes*. American Journal of Epidemiology, vol. Vol.147, No.7. pages 694–700.
- Jacquet C.** (2005), *Lavia : Dépassements de la vitesse limite*. Rapport technique 64250/2005/050712.
- LAVIA consortium** (2006). *Carnet de route du lavia*. <http://www.lavia.fr>.
- Regan M. A., Triggs T. J., Young K. L., Tomasevic N., Mitsopoulos E., Stephan E. et Tingvall C.** (2006), *On-road evaluation of intelligent speed adaptation, following distance warning and seatbelt reminder systems : Final results of the tac safecar project*. Rapport technique.
- Romon S.** (2006), *Lavia : Traitement des données*. Rapport technique.
- Saint Pierre G.** (2010), *Le système lavia et la consommation de carburant : Méthodes exploratoires pour données de conduite en situation naturelle*. Rapport technique.
- SAS Institute** (2004). *Base sas 9.1 procedures guide*. SAS Publishing.



## Annexe A

# Estimation des paramètres du modèle logistique à trois variables

Paramètre	Estimation	IC 95%		p-value
Intercept	-1.5283	-1.6532	-1.4067	<.0001
ConsigneLimiteVitesse_30	0.3380	0.0954	0.5759	0.0058
ConsigneLimiteVitesse_50	0.2146	0.0794	0.3524	0.0021
ConsigneLimiteVitesse_70	0.9387	0.7172	1.1589	<.0001
ConsigneLimiteVitesse_90	0.7295	0.5155	0.9421	<.0001
ConsigneLimiteVitesse_110	0			
ModeLavia_Constraint	-1.1797	-1.4228	-0.9448	<.0001
ModeLavia_Neutre	0			
ModeLavia_Constraint*ConsigneLimiteVitesse_30	0.5963	0.2053	0.9880	0.0028
ModeLavia_Constraint*ConsigneLimiteVitesse_50	0.6540	0.4018	0.9132	<.0001
ModeLavia_Constraint*ConsigneLimiteVitesse_70	0.3883	0.0134	0.7638	0.0424
ModeLavia_Constraint*ConsigneLimiteVitesse_90	-0.0278	-0.4072	0.3501	0.8855
ModeLavia_Constraint*ConsigneLimiteVitesse_110	0			
ModeLavia_Neutre*ConsigneLimiteVitesse_30	0			
ModeLavia_Neutre*ConsigneLimiteVitesse_50	0			
ModeLavia_Neutre*ConsigneLimiteVitesse_70	0			
ModeLavia_Neutre*ConsigneLimiteVitesse_90	0			
ModeLavia_Neutre*ConsigneLimiteVitesse_110	0			

TABLE A.1 – Estimation des paramètres du modèle logistique à trois variables explicatives. Le paramètre "ModeLAVIA\_Constraint\*ConsigneLimiteVitesse\_90" n'est pas significatifs.





## Annexe B

# Estimation des paramètres du modèle GEE à trois variables

Paramètre	Estimation	IC 95%		p-value
Intercept	-1.5211	-1.9401	-1.1020	<.0001
ConsigneLimiteVitesse_30	0.5162	-0.0514	1.0839	0.0747
ConsigneLimiteVitesse_50	0.2425	-0.1634	0.6483	0.2416
ConsigneLimiteVitesse_70	0.8680	0.3078	1.4281	0.0024
ConsigneLimiteVitesse_90	0.6286	0.1478	1.1094	0.0104
ConsigneLimiteVitesse_110	0			
ModeLavia_Constraint	-1.1184	-1.7606	-0.4762	0.0006
ModeLavia_Neutre	0			
ModeLavia_Constraint*ConsigneLimiteVitesse_30	0.6329	-0.1821	1.4479	0.1280
ModeLavia_Constraint*ConsigneLimiteVitesse_50	0.6794	0.0466	1.3121	0.0354
ModeLavia_Constraint*ConsigneLimiteVitesse_70	0.4716	-0.3502	1.2935	0.2607
ModeLavia_Constraint*ConsigneLimiteVitesse_90	0.1157	-0.5594	0.7909	0.7369
ModeLavia_Constraint*ConsigneLimiteVitesse_110	0			
ModeLavia_Neutre*ConsigneLimiteVitesse_30	0			
ModeLavia_Neutre*ConsigneLimiteVitesse_50	0			
ModeLavia_Neutre*ConsigneLimiteVitesse_70	0			
ModeLavia_Neutre*ConsigneLimiteVitesse_90	0			
ModeLavia_Neutre*ConsigneLimiteVitesse_110	0			

TABLE B.1 – Estimation des paramètres du modèle linéaire généralisé (GLM) à trois variables explicatives. Les paramètres estimés ne sont pas tous significatifs.



## Annexe C

# Estimation des paramètres des effets fixes du modèle linéaire généralisé mixte à trois variables

Paramètre (effet fixe)	Estimation	IC 95%		p-value
Intercept	-1.6740	-1.9276	-1.4205	<.0001
ConsigneLimiteVitesse_30	0.6009	0.3407	0.8612	<.0001
ConsigneLimiteVitesse_50	0.2843	0.1303	0.4383	0.0003
ConsigneLimiteVitesse_70	0.9459	0.7076	1.1842	<.0001
ConsigneLimiteVitesse_90	0.6841	0.4557	0.9124	<.0001
ConsigneLimiteVitesse_110	0			
ModeLavia_Contraint	-1.2022	-1.4477	-0.9566	<.0001
ModeLavia_Neutre	0			
ModeLavia_Contraint*ConsigneLimiteVitesse_30	0.6602	0.2572	1.0632	0.0013
ModeLavia_Contraint*ConsigneLimiteVitesse_50	0.7195	0.4568	0.9821	<.0001
ModeLavia_Contraint*ConsigneLimiteVitesse_70	0.5049	0.1160	0.8939	0.0110
ModeLavia_Contraint*ConsigneLimiteVitesse_90	0.1395	-0.2508	0.5297	<b>0.4836</b>
ModeLavia_Contraint*ConsigneLimiteVitesse_110	0			
ModeLavia_Neutre*ConsigneLimiteVitesse_30	0			
ModeLavia_Neutre*ConsigneLimiteVitesse_50	0			
ModeLavia_Neutre*ConsigneLimiteVitesse_70	0			
ModeLavia_Neutre*ConsigneLimiteVitesse_90	0			
ModeLavia_Neutre*ConsigneLimiteVitesse_110	0			

TABLE C.1 – Estimation des paramètres des effets fixes du modèle linéaire généralisé mixte (GLMM) à trois variables explicatives. Le paramètre de l'effet fixe "ModeLAVIA\_Contraint\*ConsigneLimiteVitesse\_90" n'est pas significatifs.