



HAL
open science

Détection des erreurs de mesure dans une enquête sur la mobilité

Cédric Garcia, Jimmy Armoogum

► **To cite this version:**

Cédric Garcia, Jimmy Armoogum. Détection des erreurs de mesure dans une enquête sur la mobilité. 10ème colloque francophone sur les sondages, Oct 2018, Lyon, France. hal-04318570

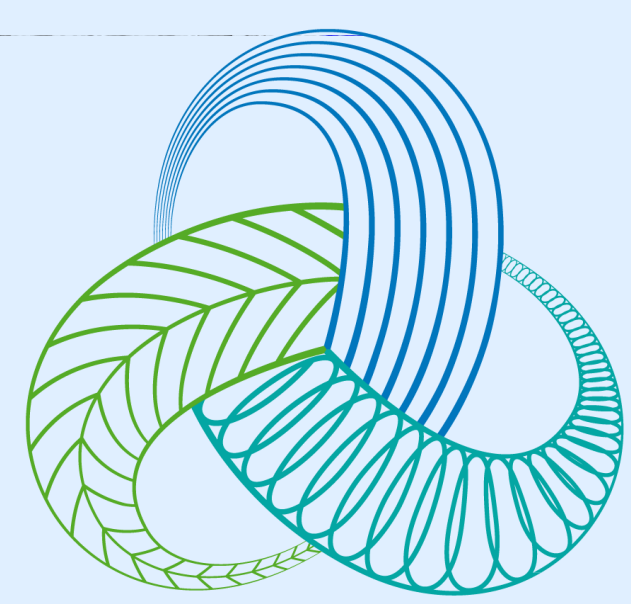
HAL Id: hal-04318570

<https://univ-eiffel.hal.science/hal-04318570v1>

Submitted on 1 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Détection des erreurs de mesure dans une enquête sur la mobilité

Cédric Garcia¹, Jimmy Armoogum¹

¹IFSTTAR-AME-DEST, 14-20 Boulevard Newton, Cité Descartes, Champs-sur-Marne, F-77447

Introduction

Contexte : Dans une enquête sur la mobilité, des personnes peuvent répondre être immobiles sur un jour, voire plusieurs. Ces immobiles peuvent être classés en 2 catégories :

Vrais immobiles

Faux immobiles

Oubli involontaire Omission volontaire

Les erreurs de mesure correspondent aux faux immobiles.

Objectif : Déterminer les vrais et les faux immobiles.

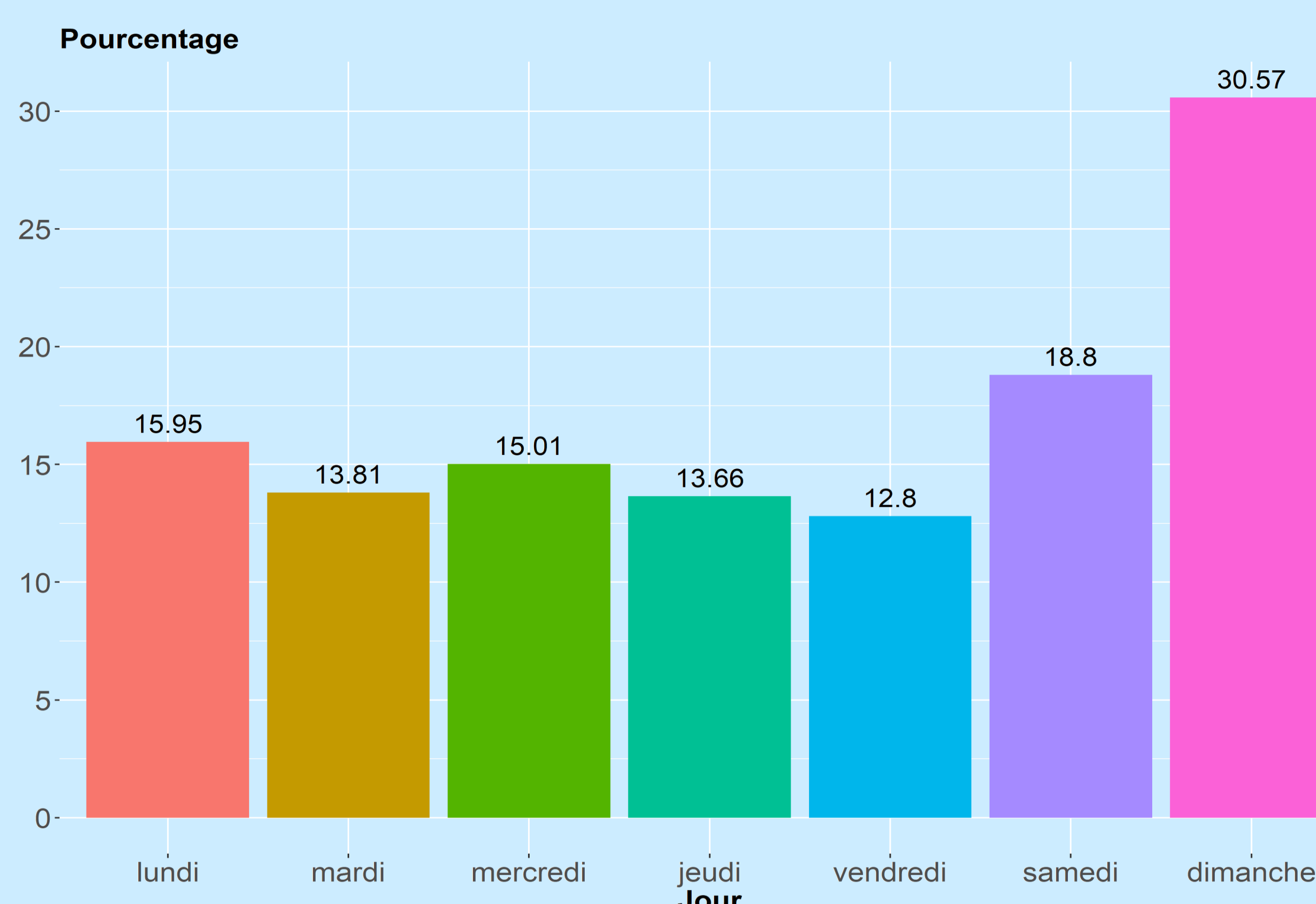
Matériels

Données : Enquête Nationale Transport Déplacements (ENTD) de 2007-2008. Enquête nationale donnant l'ensemble des déplacements réalisés pendant un jour ouvré et un jour de weekend.

- Champs d'action : France métropolitaine
- 18 632 ménages interrogés, individu Kish tiré au sort
- Question posée sur l'immobilité le jour J d'interrogation : « Depuis une semaine, y a-t-il des jours où vous n'êtes pas sorti(e) de chez vous (ou de votre jardin) ? »

- Hier (le jour J-1)
- Avant-hier (le jour J-2)
- Le jour J-3 dernier
- Le jour J-4 dernier
- Le jour J-5 dernier
- Le jour J-6 dernier
- Le jour J-7 dernier
- Je suis sorti(e) tous les jours

Immobilité observée par jour :



Méthodes

• **Etude des raisons d'immobilités données :**

En semaine (lundi à vendredi)

Incapacité provisoire
Incapacité permanente
Autre raison

En weekend

Pas besoin
Trop jeune pour me déplacer seul
Incapacité provisoire
Incapacité permanente
Nécessité de rester sur place
Pas de véhicule à disposition
Véhicule en panne
Aucune personne pour conduire
Mauvaises conditions atmosphérique
Autre

• **Tests d'homogénéités pour modéliser les erreurs de mesure :**

- Comparaison des probabilités d'être immobiles K jours précédant le jour J d'interrogation, sachant qu'on déclare être immobile les jours J-1 à J-(K-1).
- Test Q de Cochran pour déterminer si les proportions d'immobilités K jours avant le jour J d'interrogation sont les mêmes.
- Test du χ^2 d'homogénéité pour déterminer si le nombre de jours d'immobilité est influencé par le jour d'interrogation.

• **Modèle de régression logistique avec pénalisation lasso**

- Variable Y à expliquer : 0 (mobile au moins un jour sur les 7), 1 (immobile pendant les 7 jours). Ici $P(Y = 1) = 2,79\%$
- X vecteur des variables explicatives sélectionnées par la pénalisation lasso

Le modèle est :

$$P(Y = 1|X = x) = \frac{e^{\alpha + x^T \beta}}{1 + e^{\alpha + x^T \beta}}$$

Coefficients déterminés par :

$$\max_{\alpha, \beta} \sum_{i=1}^n \{y_i(\alpha + x_i^T \beta) - \log(1 + e^{\alpha + x_i^T \beta})\} \text{ sous la contrainte } \|\beta\|_1 \leq t$$

- Modèle final choisi selon la valeur t qui minimise le critère AIC

$$AIC = -2 \log(L) + 2K$$

L : vraisemblance non pénalisée du modèle

K : nombre de paramètres non nuls

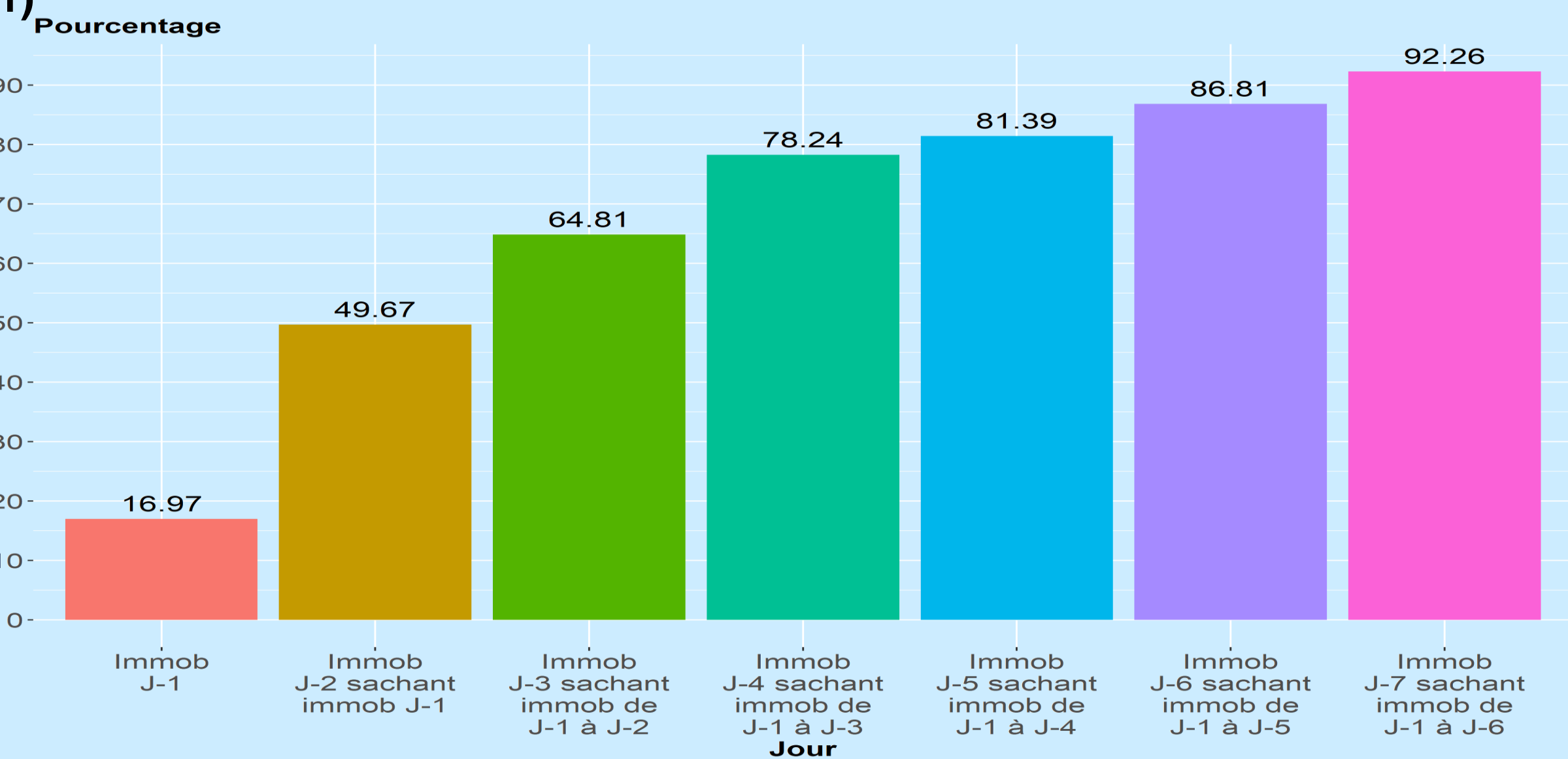
Résultats

• **Raisons potentielles d'immobilité correspondant à des erreurs de mesure**

En semaine (lundi à vendredi)
Autre raison

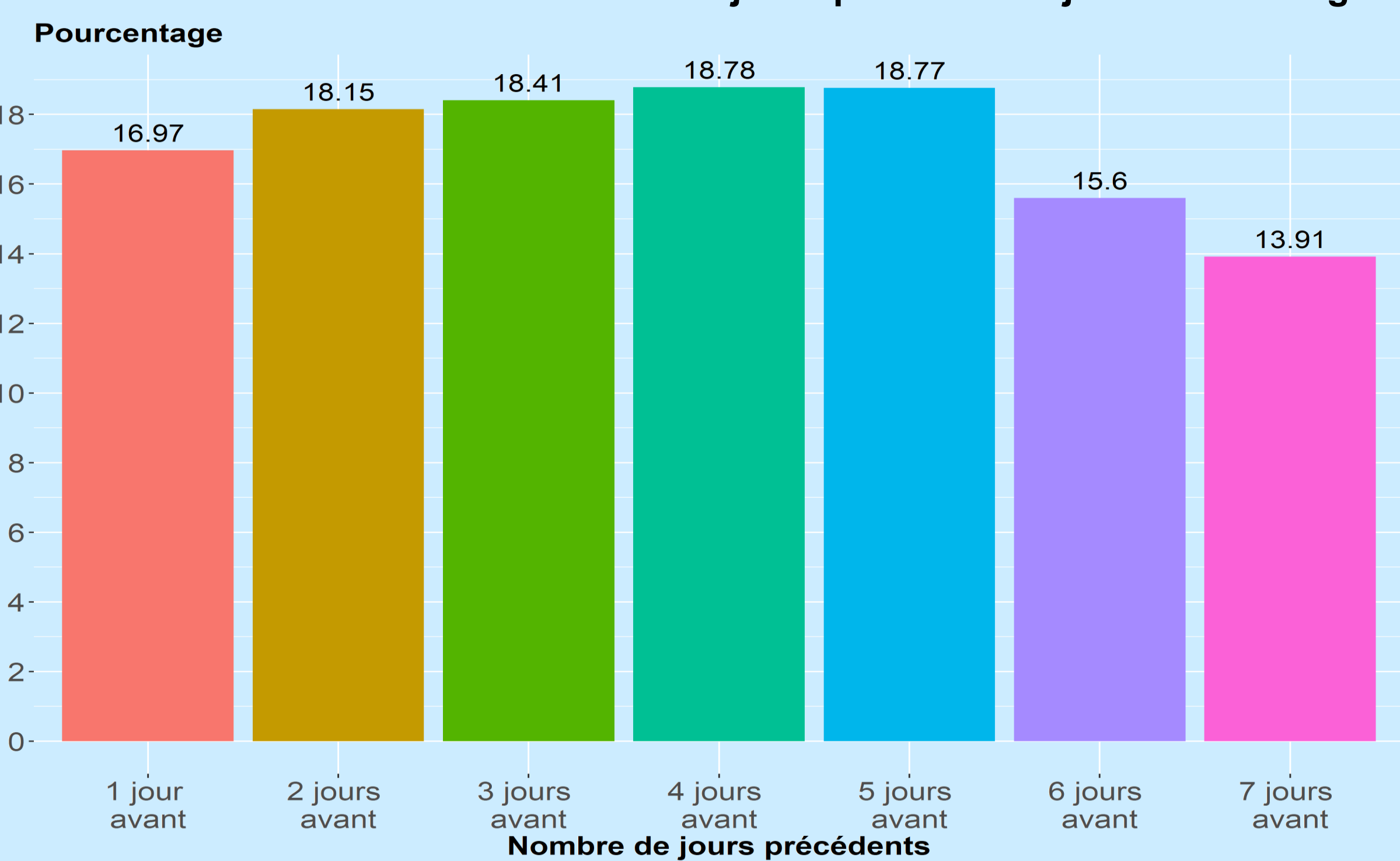
En weekend
Pas besoin
Autre

• **Probabilités conditionnelles d'être immobile jour J-K sachant immobile entre J-1 et J-(K-1)**



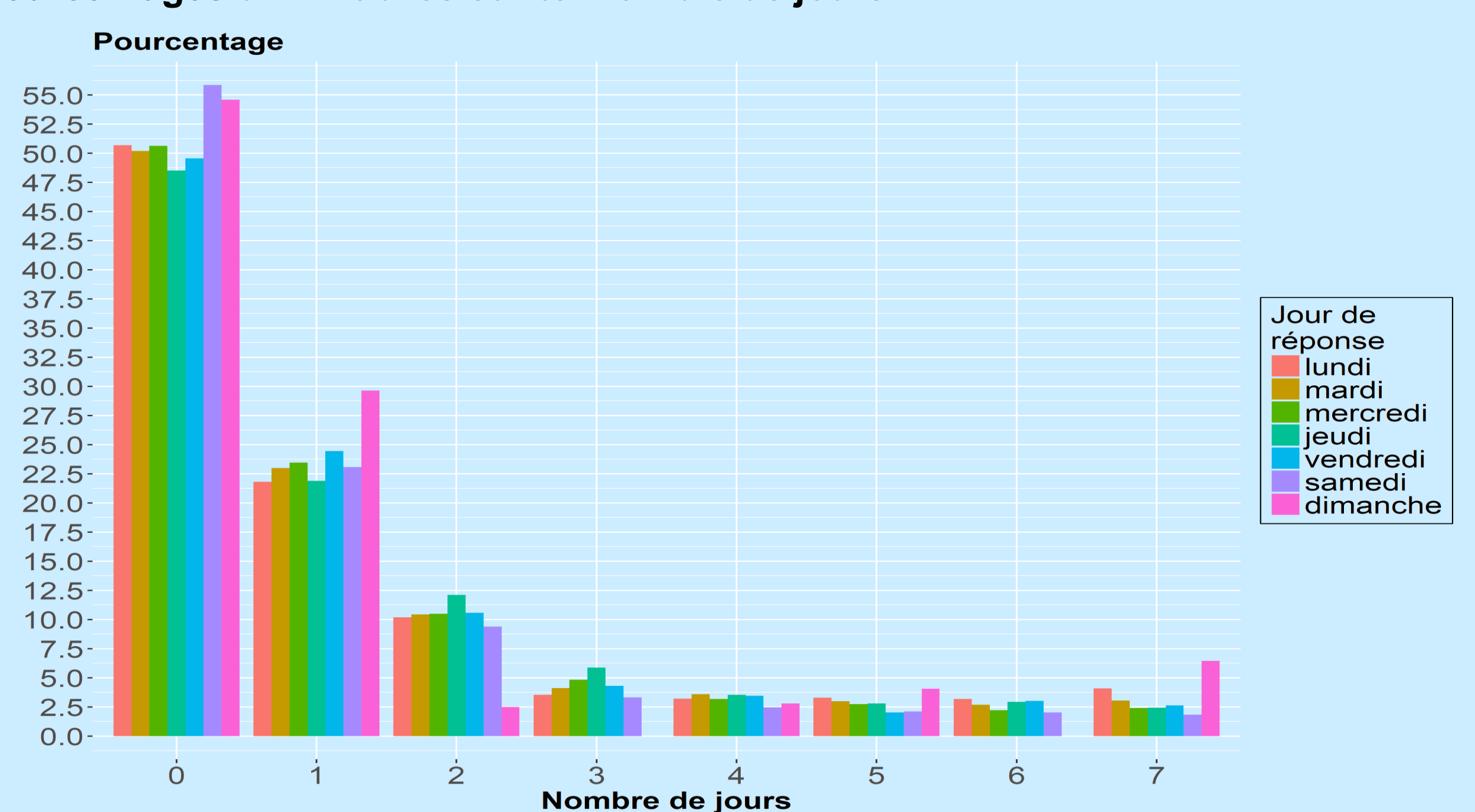
La dépendance des probabilités conditionnelles peut traduire qu'il y a des personnes qui ont des difficultés pour sortir de chez eux.

• **Probabilités d'être immobile tels nombre de jours précédant le jour J d'interrogation**



Proportions égales de 2 jours avant à 5 jours avant par Test Q de Cochran. L'oubli de déplacements ne semble pas apparaître ici.

• **Pourcentages d'immobilités sur tel nombre de jours**



Homogénéité pour ceux qui sont interrogés le mardi, mercredi, vendredi et dimanche. Ceux qui sont interrogés le samedi en particulier sont plus enclin à être mobile.

• **Performance du modèle logistique au regard de la sensibilité et la spécificité**

	Valeur observée	Mobile au moins un jour	Immobilier les 7 jours	Total
Prédiction				
Mobile au moins un jour		99,7 %	79,9 %	99,2 %
Immobilier les 7 jours		0,3 %	20,1 %	0,8 %

Facteurs qui augmentent le plus la probabilité d'être immobile les 7 jours :

- Principal apporteur de ressource hors du ménage (réf. « dans le ménage »)
- Non réponse dans problème de santé pour actions habituelles (réf. « non »)
- Véhicule personnel non remplacé pour cause de déménagement (réf. « non »)
- Commerçant et assimilés (réf. « ancien employé ou ouvrier »)
- Etat de santé général très mauvais (réf. « très bon »)

Conclusion

Cette étude met en évidence l'existence de faux immobiles, notamment ceux qui omettent de décrire leurs déplacements en déclarant être immobile durant les 7 jours. On propose un modèle prédictif à base d'apprentissage automatique afin de détecter les faux immobiles.

Les faux négatifs données par le modèle logistique sont ceux qui sont le plus susceptibles d'être ces faux immobiles. Le modèle estime un taux de faux immobiles sur 7 jours de l'ordre de 2,23%. Cette information donnée par le modèle permettrait de mieux cibler les personnes sur lesquels il faut effectuer des relances.