

Analyse statistique des déterminants de la longueur atypique de la durée trajet domicile-travail en R

Source des données : Enquête Conditions de travail 2013

Charaf ZGUIOUAR - Etudiant Master I Trésorerie d'entreprise

Zgcharaf@gmail.com

Matière : **Économétrie en R**

Professeur : **Pr. Marc-Arthur Diaye**

Année scolaire : **2022-2023**

Introduction

Ce travail porte sur une analyse statistique de la longueur atypique du trajet travail-domicile de milliers de personnes avec l'objectif de déceler les déterminants de cette anomalie.

Les données de notre étude proviennent du volet "individu" de l'enquête sur les conditions de travail de l'année 2013. Elle fait partie d'une série d'enquêtes menées par la direction de l'animation, de la recherche, des études, et des statistiques (Dares), tous les 7 ans et puis tous les trois ans entre 2013 et 2019. Il s'agit d'une étude qui vise à obtenir une description concrète du travail, de son organisation et de ses conditions sous divers angles : horaire, rythme de travail, horaires, efforts physiques ou risques encourus, pénibilité, sécurité, conflit, coopération...

La Direction de l'Animation de la Recherche, des Études et des Statistiques (Dares) est une direction de l'administration publique centrale française, qui dépend du ministère du Travail. Elle produit des analyses, des études et des statistiques sur les thèmes du travail, de l'emploi, de la formation professionnelle et du dialogue social.

L'édition de 2013 a interrogé les individus sur :

1. l'activité professionnelle,
2. Les horaires et l'organisation du temps de travail,
3. L'organisation, les rythmes et les moyens de travail,
4. Les risques, les contraintes physiques et leur prévention,
5. Les contraintes psychosociales, les relations avec le public, la violence au travail,
6. Le parcours professionnel et familial,
7. La santé physique et mentale,
8. L'exposition aux risques psycho-sociaux.

Description des données

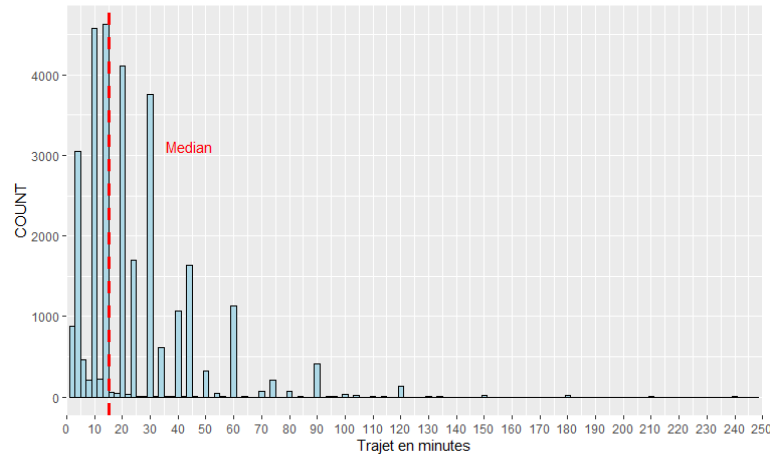
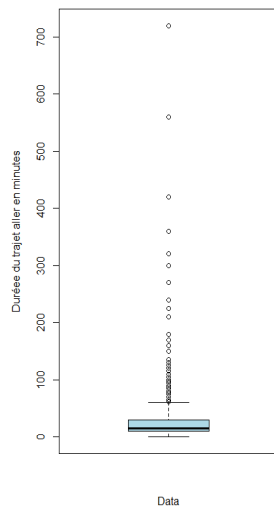
La table individus dispose de 33 673 observations et de 541 variables dont 76 qui sont quantitatives et 465 qui sont qualitatives.

L'objet de notre étude est la variable quantitative Trajet qui exprime la réponse en minutes concernant le voyage aller au lieu de travail le plus courant (hors détours éventuels) à la question "En combien de temps en moyenne faites-vous le trajet domicile-travail ?"

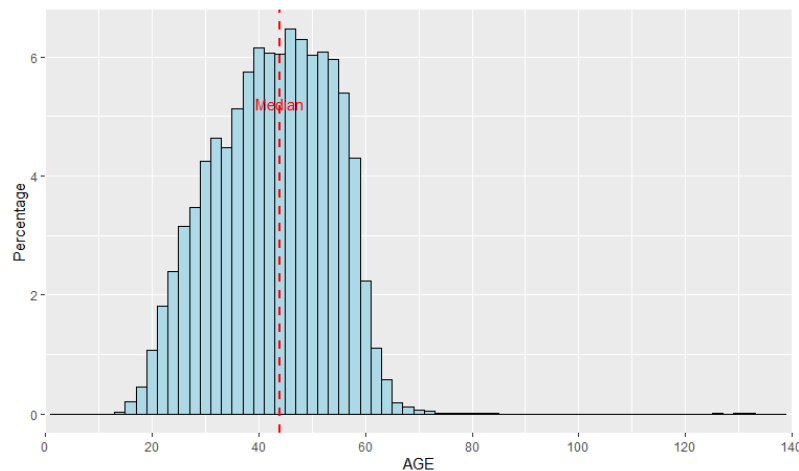
Afin de pouvoir étudier la variable trajet, nous avons enlevé les valeurs vides ainsi que les observations avec la valeur 997 qui signifie que le trajet n'est pas habituel. En faisant ainsi, nous avons désormais 32 208 observations.

Maintenant nous allons pouvoir découvrir davantage notre variable. D'abord la durée moyenne de trajet aller de 22.1 minutes et un écart type de 21.63 minutes, la médiane est moins élevée à 15 minutes car notre distribution est asymétrique à droite. La valeur maximale est de 720 minutes et la valeur minimale est de 0 minutes. 25% des individus de notre échantillon disposent d'un temps de trajet d'aller inférieur à 10 minutes, et 50% des individus d'un temps inférieur à 15 minutes, et 75% inférieur à 30 minutes.

Boîte à moustache de la variable Trajet



Nous pouvons également s'intéresser à la répartition de l'Age notre échantillon statistique : la médiane se situe autour de 44 ans et avec 75% de notre échantillon se situant en dessous de 52 ans.



Afin d'étudier l'atypicité de la longueur du trajet nous avons créé une variable qualitative qui s'appelle **Trjt (à ne pas confondre avec Trajet)** qui prend la valeur 1 si la durée de trajet est supérieure ou égale au troisième quartile de la durée de trajet de l'échantillon (30 mins) et 0 sinon. Vu que la médiane de l'échantillon est à 15 minutes, en créant cette variable nous allons avoir plus de valeurs 0 que de 1 au sein de ce vecteur.

En faisant table (NDF\$Trjt), on obtient :

30% des individus de notre échantillon disposent d'une durée de trajet supérieure ou égale 30 minutes.

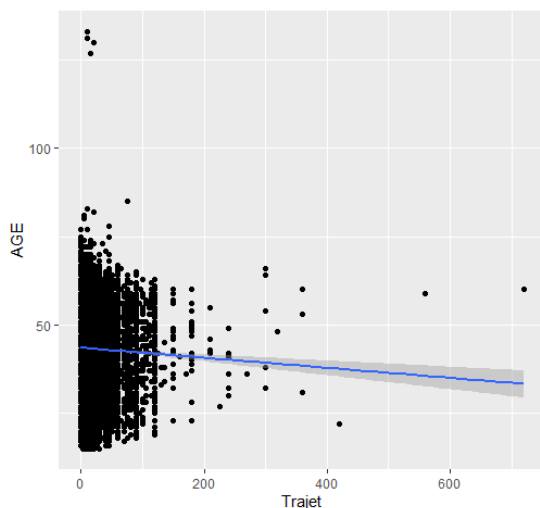
```
> table(NDF$Trjt)
 0    1
22519 9689
> 9689/22519
[1] 0.4302589
> 9689/(22519+9689)
[1] 0.3008259
>
```

Le choix des variables et hypothèses :

Afin d'expliquer les raisons sous-jacentes à la longueur atypique du trajet de certains individus, il va falloir d'abord explorer les relations des variables de notre table avec la variable **Trjt** et par ricochet **Trajet**. De prima bord, si nous étions dans une discussion autour d'amis et l'on nous a demandé d'expliquer les raisons qui expliqueront une durée de trajet atypiquement longue, l'on pensera d'emblée à l'endroit d'habitation, l'emplacement du lieu de travail, le revenu, le moyen de mobilité, le type d'emploi, les horaires de travail et ainsi de suite.

Explorons donc les corrélations entre certaines de nos variables quantitatives et la variable Trajet.

L'âge semblerait dans la pensée quotidienne être plutôt un élément favorisant la longueur de durée de trajet, durant les années de jeunesse il plus probable que l'individu ait une durée de trajet plus longue que la moyenne et au fur et à mesure cette durée se rétrécit en fonction de l'évolution des ressources financières.

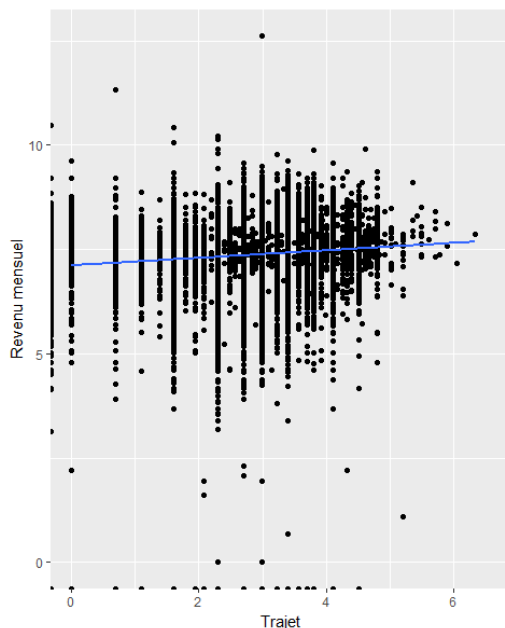
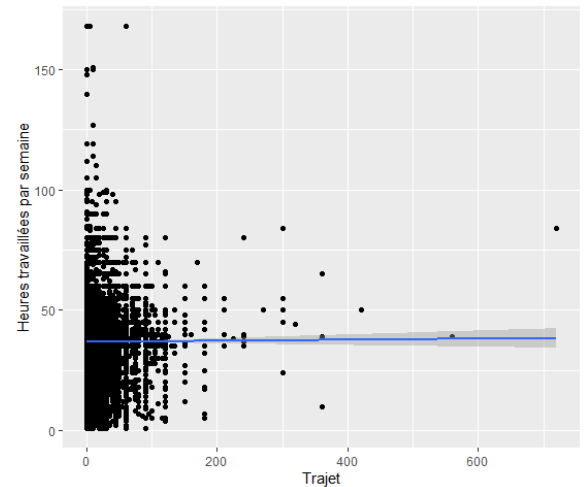


En visualisant un nuage de points de nos observations contrastant l'âge et la durée de trajet, cela fait dégager une très légère corrélation négative, qui est plus ou moins conforme à notre hypothèse ci-dessus.

L'on pourrait s'intéresser également entre le nombre d'heures de travail effectuées chaque semaine et la longueur de trajet.

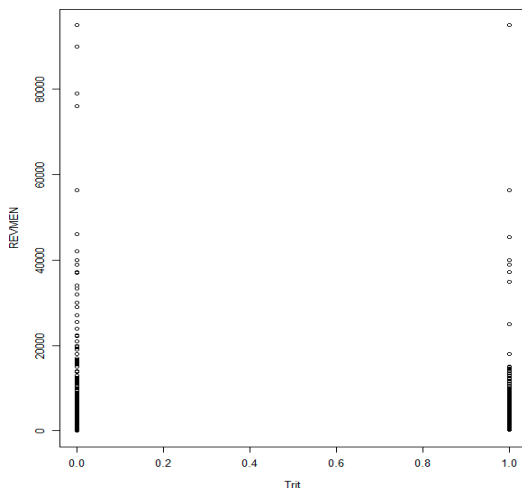
Il s'avère qu'il n'y a pas de corrélation entre les heures travaillées et la durée du trajet dans notre échantillon.

L'on pourrait également interroger la corrélation entre le revenu mensuel de l'individu et la durée du trajet pour vérifier si l'hypothèse de la réduction de la durée de trajet en raison de l'accroissement des revenus permettant la possibilité de se procurer de moyens de mobilités moins chronophages.

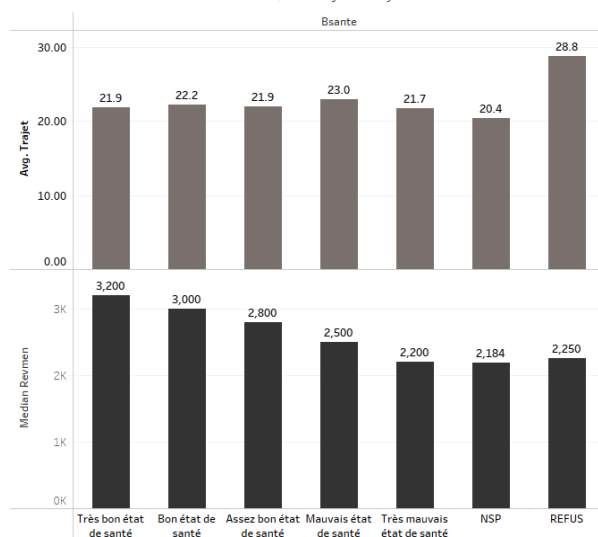


Comme l'on peut constater sur le nuage de points à gauche, la corrélation entre le log de la variable Trajet et de la variable REVMEN qui exprime le revenu mensuel moyen déclarée n'est pas très prometteuse avec un coefficient de corrélation de 0.06 ce qui n'est pas conséquent.

Cependant, en visualisant les revenus mensuels moyens de ceux qui disposent d'une durée de trajet au-dessus ou égale au troisième quartile et ceux dont la durée de trajet en dessous. Le médiane et la moyenne de la première catégorie sont respectivement à 3 400 EUR et 3 762 EUR (contre 3000 EUR et 3 300 EUR pour ceux qui disposent d'une durée de trajet en dessous du troisième quartile). En conséquence, il semblerait que la durée de trajet ne pourrait pas associer uniquement avec des revenus faibles car la réalité est plus nuancée que cela.



Relation entre le revenu median, le trajet moyen et l'état de santé.

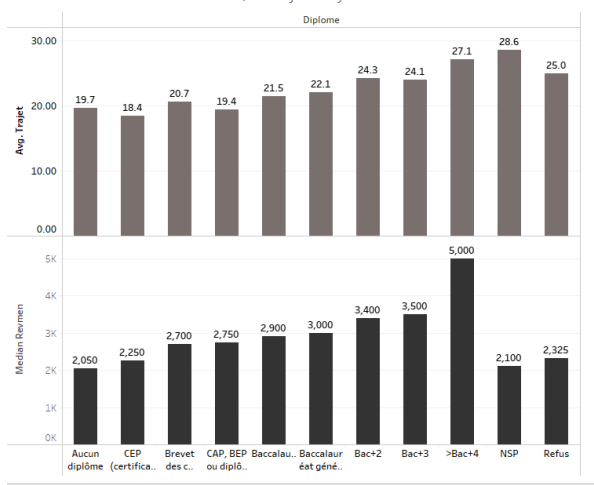


Au-delà de nos choix des variables, si on s'intéresse à la relation de la durée de trajet, le revenu mensuel avec l'état de santé de l'individu. Nous constatons qu'être dans un très bon ou dans un bon état de santé est associé avec des revenus mensuels élevés et inversement être dans un état de santé moins bien, ou très mauvais est associé à des revenus moins élevés, alors que la durée moyenne de trajet semble être plus ou moins constante entre toutes les catégories de l'état de santé hors la catégorie de refus.

Graphiquement, l'état de santé de l'individu ne semble pas avoir de corrélation avec la durée de trajet, l'on verra si statistiquement ce lien est

décelable.

Relation entre le revenu median, le trajet moyen et le niveau d'études.



Le niveau d'études accompli est positivement corrélé avec la durée de trajet, avec une différence de 10 minutes en moyenne de plus entre ceux qui disposent d'un diplôme supérieur à un diplôme bac +4 et ceux que n'en disposent d'aucun.

Modèle utilisé de régression logistique

Compte-tenu de la nature catégorique de notre variable catégorique, nous allons utiliser un modèle de régression logistique qui permet de donner comme output une pr

Méthodologie des choix des variables explicatives dans le modèle de régression logistique :

Dans notre démarche de choix des variables explicatives nous ne sommes appuyés d'abord sur la visualisation, les hypothèses que nous allons développer ensuite dans ce chapitre.

Hypothèses sous-jacentes à chaque fois de variables :

1. AGE : Toute chose égale par ailleurs le plus une personne est âgée, plus conséquent que son trajet devient.
2. CVFVP : Il s'agit d'une variable catégorique permettant de savoir si l'individu a un deuxième emploi en dehors de son activité principale ; notre hypothèse consiste à présumer que le fait d'avoir une deuxième activité professionnelle augmente la probabilité d'avoir un trajet au-dessus ou égale au troisième quartile.
3. B5D : il s'agit d'une variable catégorique qui évalue le niveau du sentiment d'être dépassé par les changements trop rapides ; Nous présumons que le sentiment d'être dépassé par la rapidité des événements de la vie quotidienne pourra être lié à des trajets domicile-travail très longs.
4. WHO : Il s'agit du score de bien-être de l'OMS qui varie entre 1 et 25, 0 étant le pire score et 25 étant le meilleur. Nous présumons que le fait d'être dans un mauvais état de bien être pourrait être lié à un trajet trop long, ou potentiellement qu'un état de mauvais bien-être ne permet pas d'accéder à des métiers facilement desservis par les transports.
5. SEXE : nous supposons que le sexe de la personne ne doit pas avoir de relation avec la durée de trajet.
6. MISSION : Il s'agit d'une variable catégorique qui jauge la fréquence de devoir dormir en dehors de chez soi en raison du travail ; nous présumons que des métiers impliquant devoir dormir en dehors de chez soi doivent impliquer également des trajets plus longs.
7. MAISON : Il s'agit d'une variable catégorique qui mesure la fréquence d'apporter du travail chez soi. Nous supposons que cela doit être lié à des longs trajets qui rendent les individus moins productifs au travail et doivent continuer pour finir leur travail chez eux.
8. COUPLE : Il s'agit d'une variable catégorique qui spécifie si la personne est en couple ou non. Nous supposons que le fait d'être en couple est lié à des trajets moins longs par rapport au 3ème quartile.
9. DETHOR : Il s'agit d'une variable catégorique qui précise les modalités de fixation des horaires au travail. Nous présumons que les métiers ayant les plus niveaux de contrôles horaires les plus strict doivent également être liés à des durées de trajet plus longues, potentiellement en raison de l'affluence sur les trajets durant les heures de pointes.

10. CWDEBOU : il s'agit d'une variable catégorique mesurant si un individu reste debout ou non au travail. Nous supposons que des métiers impliquant rester debout longtemps doivent également impliquer des durées de trajet longues.
11. NBSLENTC : Il s'agit d'une variable catégorique du nombre des salariés de l'entreprise. Nous supposons que le fait de travailler dans un grand groupe a un impact sur la longueur du trajet en comparaison avec travailler dans une petite structure.
12. Ancienneté : Il s'agit d'une variable quantitative qui exprime en nombre d'année l'ancienneté au travail d'un individu, nous supposons une relation négative entre cette dernière et le fait d'avoir une probabilité au-dessus ou égale au troisième quartile.
13. Statut : est une variable catégorique qui exprime le rôle occupé au travail ainsi que le type de l'organisme employeur (Etat, hôpital public, établissement privé...). Nous supposons que les salariés de secteur privé ont plus de chances d'avoir un trajet au-dessus ou égale au troisième quartile.
14. IMMI : la variable IMMI exprime si l'individu est immigré ou non. Nous supposons que le fait d'être immigré augmente la probabilité d'avoir une durée de trajet au-dessus ou égale au troisième quartile.
15. BSANTE : il s'agit d'une variable catégorique exprimant l'état de santé perçu par l'individu. Nous supposons que le fait d'avoir une durée de trajet au-dessus ou égale au troisième quartile, augmente la probabilité d'avoir la perception d'être dans un état de santé moins bon.
16. TENSION3 : est une variable catégorique qui indique si l'individu a de la tension dans ses rapports avec ses collègues ou non.
17. DIPLOME : il s'agit d'une variable catégorique exprimant le niveau d'accomplissement académique. Nous supposons qu'être porteur de diplômes hautement qualifiés, augmente la probabilité d'avoir une durée de trajet plus longue. Potentiellement en raison des types de postes accessibles à travers ces diplômes.
18. REGION : est une variable qualitative qualifiant la région d'habitation de l'individu. Nous supposons que dans les régions très urbanisées et peuplées, la probabilité d'avoir un trajet au-dessus ou égale au 3ème est plus élevée par rapport aux régions moins peuplées et moins urbanisées. Nous supposons une neutralité d'impact sur la durée de trajet.
19. CONTROLE : Il s'agit d'une variable catégorique exprimant le type de contrôle horaire auquel l'individu est soumis au travail. Nous supposons qu'il y aurait plus de probabilité d'avoir une durée de trajet longue dans les cadres restrictifs d'horaires et inversement une durée de trajet moins élevée dans les cadres flexibles
20. OBJECTIF : Il s'agit d'une variable catégorique exprimant la contrainte de devoir faire face à des objectifs chiffrés ou non. Nous supposons une neutralité d'impact sur la longueur de la durée de trajet.

RESULTATS & DISCUSSION

Nous fixons notre seuil de p-value à 2.5%, c'est-à-dire qu'il y aura moins de 2.5% de chances que le résultat obtenu soit dû au hasard. Nous avons fait le choix de ce seuil afin de réduire le risque d'erreur et de faire relever que les variables explicatives les plus significative.

Notre plan de résultats se découpe en trois parties :

- Résultats du modèle de la régression logistique
- Résultats des odds ratios
- Résultats des résultats marginaux.

Résultats du modèle de la régression logistique :

- Nous constatons d'abord que la **région** d'habitation a un impact sur notre variable qualitative, par exemple habiter en Ile-de-France augmente la probabilité d'avoir une durée de trajet au-dessus du 3^{ème} quartile de 26% par rapport la référence de base (la région d'alsace), de 22% si l'on habite en Martinique ou de 21% si l'on habite en Guadeloupe et inversement habiter dans le Limousin réduit la probabilité de 21%, ou dans la Basse-Normandie par 10.5%.
- La variable **Diplôme** a également un impact sur notre variable dépendante, le fait d'être porteur d'un diplôme bac+2 augmente la probabilité d'avoir une durée de trajet au-dessus ou égale troisième quartile de 7% par rapport à ne pas avoir de diplôme, de 6% quand l'on est porteur d'un diplôme bac+3/4 et de 5% quand l'on est porteur d'un diplôme supérieur à un bac+4 par rapport à être porteur d'aucun diplôme.
- Nous constatons que la modalité avoir l'avis d'être en un mauvais état de santé augmente la probabilité d'avoir une durée de trajet au-dessus ou égale au troisième quartile de 5.2% par rapport à ceux qui ont répondu être dans un très bon état de santé. L'on pourrait s'interroger sur la causalité, certainement des longues durées de trajet impactent négativement l'état de bien être à long terme, mais également le fait d'être dans un mauvais état de santé rend l'expérience d'un long trajet moins supportable.
- La variable **CWDEBOU** (si le travail impose devoir rester debout longtemps) a un impact sur la probabilité sur (TRJT = 1) : ne pas devoir rester debout durant le travail augmente la probabilité d'avoir une durée de trajet au-dessus ou égale au troisième quartile de 7.2% par rapport à devoir rester debout longtemps durant le travail. Les métiers sédentaires sont souvent associés.
- **L'âge** augmente la probabilité d'avoir une durée trajet au-dessus ou égale au troisième quartile de 0.3%.
- Être soumis à un contrôle de ses horaires de travail par d'autres personnes (des collègues par exemple) augmente la probabilité d'avoir une durée de trajet au-dessus ou égale au 30 minutes de 3.8% par rapport à n'avoir aucun contrôle horaire.
- Le fait d'être une femme réduit la probabilité d'avoir une durée de trajet au-dessus ou égale au troisième quartile de 3.2% par rapport à être un homme.
- Travailler dans une structure avec un nombre de salariés entre 50 et 499 augmente la probabilité d'avoir une durée de trajet égale ou supérieure à 30 mins de 5% par rapport à être dans une structure avec un nombre d'employés entre 1 et 49 et la probabilité passe à 9.3% lorsque l'on est employé d'une structure avec plus de 500 employés par rapport à la modalité de base.
- L'ancienneté a un odds ration très proche de 1, ce qui implique une absence d'effet sur la probabilité d'avoir une durée de trajet supérieure à 30 minutes.
- La variable tenir chercher à savoir si l'individu peut exercer son métier jusqu'à la retraite, le fait de refuser à répondre à cette question semble augmenter la probabilité d'avoir une durée de

trajet supérieure à 30 de 66% par rapport à être en mesure d'exercer son métier jusqu'à la retraite.

- Ne jamais apporter du travail chez soi ou l'apporter parfois semble réduire la probabilité d'avoir une durée de trajet au-dessus ou égale à 30 minutes respectivement de 6% et de 6.7% par rapport à apporter toujours du travail chez soi.
- Être en mesure de modifier ses horaires sous un système à la carte augmente la probabilité d'avoir une durée de trajet supérieure à 30 minutes de 4.3% par rapport à un système où les horaires sont déterminés par l'entreprise ou l'administration sans possibilité de modification.
- Ne pas avoir de collègues au travail (la modalité 3 de Tension3 qui cherche à savoir si la personne a des conflits dans ses rapports avec ses collègues) réduit la probabilité d'avoir une durée de trajet au-dessus ou égale à 30 minutes de 7.1% par rapport à avoir des collègues et des rapports nourris de tensions avec ces derniers.
- Ne pas savoir si l'on est en contact avec du public dans notre travail semble augmenter la probabilité d'avoir une durée de trajet égale ou supérieure à 30 minutes de 67%. Nous étions étonnés de ce résultat et en regardant les données sous-jacentes.

```
> long = df$TRAJET[df$PUBLIC ==9]
> long
[1] 5 45 0 15 0 12
> mean(long)
[1] 12.83333
> median(long)
[1] 8.5
```

- Il semble qu'il y a un effet d'une valeur aberrante de 45 qui a potentiellement conduit à ce résultat compte-tenu de la taille du nombre des individus avec cette modalité (6).
- La modalité NSP de la variable contrôle semble avoir un effet réducteur de la probabilité d'avoir une durée de trajet égale ou supérieure à 30 minutes de l'ordre de -25%. En regardant les valeurs sous-jacentes de cette modalité :

```
> long = df$TRAJET[df$CONTROLE ==9]
> long
[1] 180 0 NA 30 10 30 10 20 0 15 5 20 5 80 0 15 0 5 30 10 45 20 NA 0 NA 15
[27] 10 25 25
```

- Nous nous apercevons qu'il y a quelques valeurs aberrantes, des valeurs vides et des valeurs autour la médiane. La moyenne de cette série est de 23 minutes et la médiane est à 15 et il s'agit de 29 observations avec la modalité NSP. L'on pourrait également s'interroger sur un effet d'insuffisance d'échantillon avec de pouvoir généraliser.

Summary GLM

Modalité	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.75896324	0.17837509	4.25487231	0.00%
AGE	0.00698644	0.00136125	5.1323769	0.00%
CVFVPLes horaires du travail s'accordent bien avec ses engagements sociaux	0.11291953	0.02608023	4.32969816	0.00%
CVFVPLes horaires du travail ne s'accordent pas très bien avec ses engagements sociaux	0.25558575	0.03653997	6.99468966	0.00%

CVFVPLes horaires du travail ne s'accordent pas du tout avec ses engagements sociaux	0.25557043	0.05413531	4.72095653	0.00%
SEXE_FEMME	-0.08566345	0.02393896	-	0.03%
MISSION Jamais Dormir en dehors de chez soi en raison du travail	-0.14930387	0.06386048	-	1.94%
MAISON Sans objet (travail à domicile)	-0.26988371	0.07496676	-	0.03%
DETHOR Vos horaires sont modifiables par vous-même d'un jour à l'autre (système d'horaire à la carte)	0.15798517	0.03935697	4.0141604	0.01%
CWDEBOU Ne pas rester debout longtemps	0.20858582	0.02612997	7.98262817	0.00%
NBSALENTC 50 à 499	0.14660584	0.03244504	4.51859048	0.00%
NBSALENTC 500 et plus	0.26872646	0.03098064	8.67401304	0.00%
NBSALENTC NSP	0.19177234	0.04820469	3.97829198	0.01%
TENSION3 Sans objet (sans collègue)	-0.16058877	0.0590901	-	0.66%
anciennete	-0.00075575	0.00011529	-6.5550916	0.00%
DIPLOME_Bac + 2	0.1371743	0.05093804	2.69296376	0.71%
DIPLOME_Bac +3/4	0.13036975	0.05545177	2.35104744	1.87%
IMMI_Immigré	0.13283907	0.04457218	2.98031371	0.29%
region Auvergne	-0.34233943	0.10177923	-	0.08%
region Basse-Normandie	-0.27906808	0.10169508	-	0.61%
region Bretagne	-0.27472557	0.08100373	-	0.07%
region Champagne-Ardenne	-0.25140443	0.10059558	-	1.24%
region_Franche-Comté	-0.29223607	0.10673083	-	0.62%
region_Guadeloupe	0.50052522	0.08399227	5.95918182	0.00%
region_Île-de-France	0.65619152	0.06608805	9.929049	0.00%
region_La Réunion	0.21550377	0.08484719	2.53990462	1.11%
region_Limousin	-0.7337542	0.15790399	-	0.00%
region_Martinique	0.56013796	0.08305497	6.74418335	0.00%
region_Pays de la Loire	-0.21895541	0.07473445	-	0.34%
region_Picardie	0.24804593	0.08270735	2.99907981	0.27%

Rapport des cotes

Modalité	OR	2.50%	97.50%	p
(Intercept)	0.46815154	0.33028158	0.66310856	0.00%
AGE	1.0070109	1.0043431	1.00968322	0.00%
CVFVPLes horaires du travail s'accordent bien avec ses engagements sociaux	1.11954184	1.06383687	1.17822086	0.00%
CVFVPLes horaires du travail ne s'accordent pas très bien avec ses engagements sociaux	1.29121773	1.20188062	1.38713685	0.00%
CVFVPLes horaires du travail ne s'accordent pas du tout avec ses engagements sociaux	1.29119795	1.1610197	1.43557364	0.00%
SEXE_FEMME	0.9179031	0.87576034	0.96205475	0.03%
MISSION Jamais Dormir en dehors de chez soi en raison du travail	0.86130735	0.76090933	0.97532553	1.94%
MAISON Sans objet (travail à domicile)	0.76346827	0.65955779	0.88394408	0.03%
DETHOR Vos horaires sont modifiables par vous même d'un jour à l'autre(système d'horaire à la carte)	1.17114883	1.08417966	1.26494328	0.01%
CWDEBOU Ne pas rester debout longtemps	1.23193465	1.17032073	1.29680675	0.00%
NBSALENTC 50 à 499	1.15789748	1.08653271	1.23394022	0.00%
NBSALENTC 500 et plus	1.30829722	1.2312151	1.39028248	0.00%
NBSALENTC NSP	1.21139469	1.10221977	1.33091852	0.01%
TENSION3 Sans objet(sans collègue)	0.85164222	0.75833341	0.95563989	0.66%
anciennete	0.99924453	0.99901925	0.9994697	0.00%
DIPLOME_Bac + 2	1.14702806	1.03811466	1.26764579	0.71%
DIPLOME_Bac +3/4	1.13924954	1.02210296	1.27000519	1.87%
IMMI_Immigré	1.14206619	1.04646495	1.24618071	0.29%
region_Auvergne	0.71010713	0.58075848	0.86639675	0.08%
region_Basse-Normandie	0.7564884	0.61876162	0.92297698	0.61%
region_Bretagne	0.7597806	0.64794248	0.89076224	0.07%
region_Champagne-Ardenne	0.77770778	0.63776776	0.94651478	1.24%
region_Franche-Comté	0.74659227	0.60519445	0.91875872	0.62%
region_Guadeloupe	1.64958745	1.39955277	1.9451077	0.00%
region_Île-de-France	1.92743773	1.69372659	2.19525276	0.00%
region_La Réunion	1.24048666	1.05049041	1.46505182	1.11%
region_Limousin	0.4801032	0.34902346	0.6497976	0.00%
region_Martinique	1.75091404	1.48875602	2.06024649	0.00%
region_Pays de la Loire	0.80335754	0.69380192	0.9304933	0.34%
region_Picardie	1.28151879	1.08914762	1.50829019	0.27%

Le rapport des effets marginaux

	dF.dx	Std..Err.	z	P..z.
PUBLIC NSP	66.7%	0.01435024	46.4903958	0.00%
TENIR_REFUS	66.7%	0.01628644	40.9624602	0.00%
region_Île-de-France	26.0%	0.02523138	10.3096794	0.00%
region_Martinique	21.9%	0.03248478	6.75345512	0.00%
region_Guadeloupe	19.4%	0.0327688	5.92217821	0.00%
region_Picardie	9.5%	0.03163153	2.99950964	0.27%
NBSALENTC 500 et plus	9.3%	0.01143162	8.10255605	0.00%
region_La Réunion	8.8%	0.03274198	2.6787125	0.74%
NBSALENTC NSP	7.5%	0.01805426	4.12746138	0.00%
CWDEBOU Ne pas rester debout longtemps	7.2%	0.00931234	7.75805815	0.00%
DIPLOME_Bac + 2	7.0%	0.01877771	3.73708774	0.02%
DIPLOME_Bac +3/4	6.0%	0.02049766	2.92580324	0.34%
BSANTE_Mauvais	5.2%	0.02174829	2.36901331	1.78%
DIPLOME_>bac+4	5.0%	0.02198684	2.28775453	2.22%
NBSALENTC 50 à 499	5.0%	0.01190038	4.1651743	0.00%
DETHOR Vos horaires sont modifiables par vous-même d'un jour à l'autre (système d'horaire à la carte)	4.3%	0.01461492	2.94864378	0.32%
CONTROLE Être soumis à un contrôle par d'autres personnes, par exemple vos collègues	3.8%	0.01394997	2.69573227	0.70%
AGE	0.3%	0.00049924	5.39770004	0.00%
CONGE	0.1%	0.00027277	2.44000666	1.47%
anciennete	-0.03%	4.14E-05	-7.0420658	0.00%
OBJECTIF Ne pas devoir atteindre des objectifs précis	-2.2%	0.00890198	-2.41586439	1.57%
SEXE_FEMME	-3.2%	0.00855139	-3.7339584	0.02%
MAISON Apporter jamais le travail chez soi	-6.0%	0.02581645	-2.30726823	2.10%
MISSION Jamais Dormir en dehors de chez soi en raison du travail	-6.5%	0.02371036	-2.73171132	0.63%
MAISON Apporter parfois le travail chez soi	-6.7%	0.02470917	-2.70225462	0.69%
TENSION3 Sans objet(sans collègue)	-7.1%	0.01908646	-3.71620386	0.02%
region_Pays de la Loire	-8.3%	0.02396062	-3.46242035	0.05%
region_Champagne-Ardenne	-8.8%	0.03113111	-2.81780753	0.48%
region_Bretagne	-9.3%	0.02523853	-3.67739229	0.02%

region_Franche-Comté	-9.5%	0.03248802	-	2.92851957	0.34%
region_Auvergne	-10.4%	0.03053937	-	3.40381957	0.07%
MAISON Sans objet (travail à domicile)	-10.4%	0.0245749	-	4.24913829	0.00%
region_Basse-Normandie	-10.5%	0.03067232	-	3.426193	0.06%
region_Limousin	-20.6%	0.03277913	-	6.27382544	0.00%
CONTROLE NSP	-25.3%	0.070426	-	3.59140068	0.03%
TENSION3 Refus	-33.3%	0.03215998	-	10.3450776	0.00%

En vue de ces résultats, force est de constater que la région d'habitation impacte fortement la durée de trajet, ceci était décelable en amont car nous sommes familiers avec les contraintes de mobilité en Ile-de-France et dans les métropoles. Aussi, le coût de location ou d'acquisition d'immobilier est plus conséquent dans les métropoles et donc les entreprises font le choix de s'installer à l'extérieure de la ville, augmentant ainsi la durée de trajet moyenne de leurs salariés, ceci est le cas par exemple dans le bassin du Rhône où un certain nombre de sociétés industrielles de chimie et de produits pharmaceutiques se sont implantés en région lyonnaise bien loin du centre de la ville. Bien évidemment ce n'est pas le cas pour toutes les entreprises mais pour les grandes multinationales il est plus fréquent de trouver ce cas de figure. Cependant dans certaines régions comme la Martinique et la Guadeloupe, nous avons relevé une augmentation de la probabilité d'avoir une durée de trajet plus élevée ou égale à 30 minutes, ceci pourra être liée à l'infrastructure et l'efficacité des transports ou en commun et la congestion. Par exemple en Guadeloupe, la voiture est le mode préféré pour se rendre au travail pour 91% des individus actifs et le premier poste des dépenses des ménages en Guadeloupe est les dépenses liées à leurs voitures soit 3400 euros par an (+34% à +56% de plus qu'en Hexagone).

L'éducation est également un facteur important, le plus élevé l'accomplissement académique, le plus l'on a tendance à travailler au sein de zones lointaines et le plus l'on a tendance à accepter un temps de trajet plus long pour une contrepartie financière plus ou moins compensatrice. Néanmoins, la généralisation est à considérer en tenant toutes les autres variables constantes, alors que la réalité est plus dynamique que cela. Nous avons également vu que les femmes avaient moins de chances d'avoir une durée de trajet supérieure ou égale à 30 minutes. La raison derrière cela pourrait être complexe à expliquer.

Par ailleurs, nous avons également calculé le pseudo R, qui est de 8.31%, un pourcentage qui n'est pas très élevé.

Interprétation du pseudo-R²

```

> #pseudo R
> ll.null = full_model$null.deviance/-2
> ll.proposed = full_model$deviance/-2
> pseudoR = (ll.null-ll.proposed)/ll.null
>
> pseudoR
[1] 0.08315668
>

```

Interprétation du AIC et BIC

Lors du test des combinaisons de variables afin de trouver le AIC le plus faible, la plus grande réduction du AIC est arrivée après l'ajout de la variable NBSALENTC qui exprime Nombre de salariés dans l'entreprise (y compris entreprises monoétablissement) en tranches. Le AIC est passé de 36 102 à 18 719.

Le AIC effectif de notre modèle final est 17390.97 vs un AIC d'un modèle vide de 39 396.

```

> AIC(glm(Trjt~1,data=NDF, family = binomial(link = 'probit')))
[1] 39396.41

```

Le BIC effectif de notre modèle final est 18074.06 vs un BIC d'un modèle vide de 39 404.97.

```

> BIC(glm(Trjt~1,data=NDF, family = binomial(link = 'probit')))
[1] 39404.79

```


Conclusion :

Nous avons pu démontrer parmi avec une p-value de 2.5% les facteurs déterminants d'une durée de trajet supérieure ou égale à 30 minutes. Certainement nous n'avons pas pu se servir de toutes les variables et nous n'avons pu choisir qu'un simple échantillon de variables.

Bien entendu, nous pourrions utiliser d'autres modèles pour des fins explicatives et prédictives comme les random forests, PCA, SVM, ou de clustering.