

# Alzheimer

Hugo Maury

08 novembre 2024

*Sources :*

- <https://www.kaggle.com/datasets/suleymansulak/obesity-dataset/data> (<https://www.kaggle.com/datasets/suleymansulak/obesity-dataset/data>)

Les analyses ont été réalisées grâce au logiciel Rstudio version 4.0.5.

## Contexte

Ce jeu de données contient des informations de santé détaillées pour **2 149 patients**, chacun étant identifié de manière unique par des ID allant de **4751 à 6900**. Il inclut des détails démographiques, des facteurs liés au mode de vie, des antécédents médicaux, des mesures cliniques, des évaluations cognitives et fonctionnelles, des symptômes, ainsi qu'un diagnostic de la maladie d'Alzheimer. Ce jeu de données est idéal pour les chercheurs et les data scientists souhaitant explorer les facteurs associés à la maladie d'Alzheimer, développer des modèles prédictifs et mener des analyses statistiques.

# 1 ) Présentation des données

## Informations sur les patients

### Identifiant du patient

- **PatientID** : Un identifiant unique attribué à chaque patient (allant de 4751 à 6900).

### Détails démographiques

- **Âge** : Âge des patients, compris entre 60 et 90 ans.
- **Genre** : Sexe des patients, où 0 représente les hommes et 1 représente les femmes.
- **Ethnicité** : Origine ethnique des patients, codée comme suit :
  - 0 : Caucasien
  - 1 : Afro-Américain
  - 2 : Asiatique
  - 3 : Autre
- **Niveau d'éducation** : Niveau d'études des patients, codé comme suit :
  - 0 : Aucun
  - 1 : Lycée
  - 2 : Licence
  - 3 : Supérieur

### Facteurs de mode de vie

- **IMC (Indice de Masse Corporelle)** : IMC des patients, variant de 15 à 40.
- **Tabagisme** : Statut de fumeur, où 0 indique Non et 1 indique Oui.
- **Consommation d'alcool** : Consommation hebdomadaire d'alcool en unités, variant de 0 à 20.
- **Activité physique** : Activité physique hebdomadaire en heures, variant de 0 à 10.
- **Qualité de l'alimentation** : Score de qualité de l'alimentation, variant de 0 à 10.
- **Qualité du sommeil** : Score de qualité du sommeil, variant de 4 à 10.

### Antécédents médicaux

- **Antécédents familiaux d'Alzheimer** : Présence d'antécédents familiaux de la maladie d'Alzheimer, où 0 indique Non et 1 indique Oui.
- **Maladie cardiovasculaire** : Présence d'une maladie cardiovasculaire, où 0 indique Non et 1 indique Oui.
- **Diabète** : Présence de diabète, où 0 indique Non et 1 indique Oui.
- **Dépression** : Présence de dépression, où 0 indique Non et 1 indique Oui.
- **Blessure à la tête** : Antécédents de blessure à la tête, où 0 indique Non et 1 indique Oui.
- **Hypertension** : Présence d'hypertension, où 0 indique Non et 1 indique Oui.

### Mesures cliniques

- **Pression artérielle systolique (SystolicBP)** : Variant de 90 à 180 mmHg.
- **Pression artérielle diastolique (DiastolicBP)** : Variant de 60 à 120 mmHg.
- **Cholestérol total** : Niveaux de cholestérol total, de 150 à 300 mg/dL.
- **Cholestérol LDL** : Niveaux de lipoprotéines de basse densité, de 50 à 200 mg/dL.
- **Cholestérol HDL** : Niveaux de lipoprotéines de haute densité, de 20 à 100 mg/dL.
- **Triglycérides** : Niveaux de triglycérides, de 50 à 400 mg/dL.

### Évaluations cognitives et fonctionnelles

- **MMSE (Mini-Mental State Examination)** : Score de l'examen de l'état mental, variant de 0 à 30. Des scores faibles indiquent un trouble cognitif.
- **Évaluation fonctionnelle** : Score d'évaluation fonctionnelle, de 0 à 10. Des scores faibles indiquent une plus

grande déficience.

- **Plaintes de mémoire** : Présence de plaintes liées à la mémoire, où 0 indique Non et 1 indique Oui.

- **Problèmes comportementaux** : Présence de problèmes comportementaux, où 0 indique Non et 1 indique Oui.

- **ADL (Activités de la vie quotidienne)** : Score des activités de la vie quotidienne, de 0 à 10. Des scores faibles indiquent une plus grande déficience.

### Symptômes

- **Confusion** : Présence de confusion, où 0 indique Non et 1 indique Oui.
- **Désorientation** : Présence de désorientation, où 0 indique Non et 1 indique Oui.
- **Changements de personnalité** : Présence de changements de personnalité, où 0 indique Non et 1 indique Oui.
- **Difficulté à accomplir des tâches** : Présence de difficultés à accomplir des tâches, où 0 indique Non et 1 indique Oui.
- **Oubli** : Présence d'oubli, où 0 indique Non et 1 indique Oui.

### Informations sur le diagnostic

- **Diagnostic** : Statut du diagnostic de la maladie d'Alzheimer, où 0 indique Non et 1 indique Oui.

### Informations confidentielles

- **Médecin en charge** : Cette colonne contient des informations confidentielles sur le médecin en charge, avec la valeur "XXXConfid" pour tous les patients.

### Conclusion

La base de données comprend des informations démographiques, des facteurs de style de vie, des antécédents médicaux, des mesures cliniques, des évaluations cognitives et fonctionnelles, ainsi que des symptômes pour analyser et prédire la maladie d'Alzheimer. Les données sont conçues pour la recherche et les projets de data science tout en respectant la confidentialité des informations sensibles.

## 2 ) Première analyse

### 2.1 ) Distribution des caractéristiques de l'Alzheimer par détails démographiques

Le tableau ci-dessous présente une analyse descriptive des caractéristiques démographiques des patients, en fonction de leur statut diagnostique par rapport à la maladie d'Alzheimer. Il regroupe des informations sur l'âge, le genre, l'origine ethnique et le niveau d'éducation, permettant ainsi de comparer les profils des patients atteints et non atteints de la maladie.

Distribution des caractéristiques de l'Alzheimer par détails démographiques			
Variable	Non-atteint (n=1389, 65%) <sup>1</sup>	Atteint (n=760, 35%) <sup>1</sup>	p-value <sup>2</sup>
Age	75 (67, 83)	75 (67, 83)	0.8
Genre			0.3
Femme	714 (51%)	374 (49%)	
Homme	675 (49%)	386 (51%)	
Ethnicité			0.10
Afro-Américain	308 (22%)	146 (19%)	
Asiatique	122 (8.8%)	84 (11%)	
Autre	144 (10%)	67 (8.8%)	
Caucasien	815 (59%)	463 (61%)	
Niveau d'éducation			0.2
Aucun	272 (20%)	174 (23%)	
Licence	419 (30%)	217 (29%)	
Lycée	552 (40%)	302 (40%)	
Supérieur	146 (11%)	67 (8.8%)	

<sup>1</sup> Median (IQR); n (%)  
<sup>2</sup> Wilcoxon rank sum test; Pearson's Chi-squared test

Ce tableau présente les caractéristiques démographiques des patients en fonction de leur statut diagnostique (atteints ou non atteints de la maladie d'Alzheimer). Les variables analysées incluent l'âge, le genre, l'origine ethnique et le niveau d'éducation des patients. Voici une interprétation détaillée des résultats pour chaque

variable.

**Définition** Une **p-value faible** (moins de 0,05) indique que les différences observées sont significatives et peu susceptibles d'être dues au hasard.

Une **p-value élevée** (plus de 0,05) indique qu'il est difficile de conclure qu'il y a une différence significative entre les groupes comparés.

Il est important de noter que la p-value ne prouve pas qu'une hypothèse est vraie ou fausse. Elle aide seulement à quantifier la probabilité que les résultats observés soient dus au hasard.

### 1. Âge

Non-atteints : L'âge médian des patients non atteints de la maladie d'Alzheimer est de 75 ans, avec une plage d'âge allant de 67 à 83 ans.

Atteints : L'âge médian des patients atteints de la maladie d'Alzheimer est également de 75 ans, avec une plage d'âge similaire (67 à 83 ans).

Interprétation : L'âge médian est identique pour les deux groupes, et la différence n'est pas statistiquement significative (p-value = 0.8), ce qui signifie que l'âge ne semble pas être un facteur discriminant entre les deux groupes dans cette étude.

### 2. Genre

Non-atteints : 51% des patients non atteints sont des femmes et 49% des hommes.

Atteints : 49% des patients atteints sont des femmes et 51% des hommes.

Interprétation : Il y a une répartition relativement équilibrée des genres dans les deux groupes (atteints et non atteints), avec une légère majorité de femmes dans le groupe des non-atteints et une légère majorité d'hommes dans le groupe des atteints. La différence entre les groupes n'est pas statistiquement significative (p-value = 0.3), ce qui suggère qu'il n'y a pas de lien clair entre le genre et l'état de la maladie dans cette population.

### 3. Ethnicité

Non-atteints : 22% des patients sont Afro-Américains, 8.8% sont Asiatiques, 10% sont d'une autre origine ethnique, et 59% sont Caucasiens.

Atteints : 19% des patients sont Afro-Américains, 11% sont Asiatiques, 8.8% sont d'une autre origine ethnique, et 61% sont Caucasiens.

Interprétation : Bien qu'il y ait des variations dans les proportions d'ethnies entre les deux groupes, aucune différence significative n'est observée dans l'ensemble (p-value = 0.10). Cela suggère que l'origine ethnique n'est pas un facteur déterminant dans la présence ou l'absence de la maladie d'Alzheimer dans cette population.

### 4. Niveau d'éducation

Non-atteints : 20% n'ont aucun diplôme, 30% ont une licence, 40% ont un diplôme de lycée, et 11% ont un diplôme supérieur.

Atteints : 23% n'ont aucun diplôme, 29% ont une licence, 40% ont un diplôme de lycée, et 8.8% ont un diplôme supérieur.

Interprétation : Les deux groupes présentent une répartition similaire dans la catégorie des diplômes de lycée et licence. Cependant, une proportion légèrement plus élevée de patients non atteints a un niveau d'éducation plus élevé, notamment une licence ou un diplôme supérieur. La différence n'est cependant pas statistiquement significative (p-value = 0.2), ce qui suggère que le niveau d'éducation n'est pas un facteur majeur dans l'apparition de la maladie d'Alzheimer dans cette étude.

### Conclusion générale

Ce tableau fournit une vue d'ensemble des caractéristiques démographiques des patients atteints et non atteints de la maladie d'Alzheimer. Aucune des variables (âge, genre, ethnicité, niveau d'éducation) ne

présente de différences statistiquement significatives entre les deux groupes, ce qui suggère que, dans cette étude, ces facteurs démographiques ne sont pas des indicateurs clairs du développement de la maladie d'Alzheimer. Cependant, il est important de noter que d'autres facteurs non pris en compte dans ce tableau peuvent influencer l'apparition de la maladie.



## 2.2 ) Distribution des caractéristiques de l'Alzheimer par facteurs de mode de vie

Le tableau ci-dessous présente une analyse descriptive des caractéristiques par facteurs de mode de vie des patients, en fonction de leur statut diagnostique par rapport à la maladie d'Alzheimer. Il regroupe des informations sur l'IMC, le tabagisme, l'activité physique, la qualité d'alimentation et de sommeil, permettant ainsi de comparer les profils des patients atteints et non atteints de la maladie.

Distribution des caractéristiques de l'Alzheimer par facteurs de mode de vie			
Variable	Non-atteint (n=1389, 65%) <sup>1</sup>	Atteint (n=760, 35%) <sup>1</sup>	p-value <sup>2</sup>
IMC	28 (22, 34)	28 (22, 34)	0.2
Tabagisme			0.8
Fumeur	403 (29%)	217 (29%)	
Non-fumeur	986 (71%)	543 (71%)	
Consommation d'alcool	10.0 (5.3, 15.1)	9.9 (4.8, 15.2)	0.7
Activité physique	4.73 (2.52, 7.46)	4.85 (2.63, 7.37)	0.8
Qualité d'alimentation	5.07 (2.35, 7.52)	5.08 (2.60, 7.58)	0.7
Qualité du sommeil	7.24 (5.56, 8.65)	6.91 (5.34, 8.41)	0.009

<sup>1</sup> Median (IQR); n (%)

<sup>2</sup> Wilcoxon rank sum test; Pearson's Chi-squared test

1. *IMC (Indice de Masse Corporelle)*  
Non-atteints : L'IMC médian des patients non atteints de la maladie d'Alzheimer est de 28, avec un intervalle interquartile (IQR) allant de 22 à 34.  
Atteints : L'IMC médian des patients atteints de la maladie d'Alzheimer est également de 28, avec un intervalle interquartile similaire (22 à 34).  
Interprétation : L'IMC médian est identique pour les deux groupes, et la différence n'est pas statistiquement significative (p-value = 0.2), ce qui suggère que l'IMC ne semble pas être un facteur discriminant entre les deux groupes dans cette étude.

2. *Tabagisme*  
Non-atteints : 29% des patients non atteints sont fumeurs, tandis que 71% sont non-fumeurs.  
Atteints : 29% des patients atteints sont fumeurs, tandis que 71% sont non-fumeurs.  
Interprétation : La proportion de fumeurs est identique dans les deux groupes. La différence n'est pas statistiquement significative (p-value = 0.8), ce qui indique que le tabagisme ne semble pas avoir de lien avec l'état de la maladie d'Alzheimer dans cette population.

3. *Consommation d'alcool*  
Non-atteints : La consommation d'alcool médiane est de 10.0 unités par semaine, avec un intervalle

interquartile de 5.3 à 15.1 unités.

Atteints : La consommation d'alcool médiane est de 9.9 unités par semaine, avec un intervalle interquartile de 4.8 à 15.2 unités.

Interprétation : Les médianes de la consommation d'alcool sont très similaires pour les deux groupes, et la différence n'est pas statistiquement significative ( $p\text{-value} = 0.7$ ), ce qui suggère que la consommation d'alcool n'a pas de lien avec la présence ou l'absence de la maladie d'Alzheimer dans cette étude.

#### *4. Activité physique*

Non-atteints : Le nombre d'heures d'activité physique médian par semaine est de 4.73, avec un intervalle interquartile de 2.52 à 7.46 heures.

Atteints : Le nombre d'heures d'activité physique médian par semaine est de 4.85, avec un intervalle interquartile de 2.63 à 7.37 heures.

Interprétation : Les deux groupes présentent des niveaux similaires d'activité physique, avec des médianes proches. La différence n'est pas statistiquement significative ( $p\text{-value} = 0.8$ ), ce qui suggère que l'activité physique n'est pas un facteur discriminant entre les deux groupes en termes de la maladie d'Alzheimer.

#### *5. Qualité de l'alimentation*

Non-atteints : Le score médian de la qualité de l'alimentation est de 5.07, avec un intervalle interquartile de 2.35 à 7.52.

Atteints : Le score médian de la qualité de l'alimentation est de 5.08, avec un intervalle interquartile de 2.60 à 7.58.

Interprétation : La qualité de l'alimentation semble similaire entre les deux groupes, avec des scores médian proches. La différence n'est pas statistiquement significative ( $p\text{-value} = 0.7$ ), ce qui suggère que la qualité de l'alimentation ne joue probablement pas un rôle déterminant dans l'apparition de la maladie d'Alzheimer dans cette étude.

#### *6. Qualité du sommeil*

Non-atteints : Le score médian de la qualité du sommeil est de 7.24, avec un intervalle interquartile de 5.56 à 8.65.

Atteints : Le score médian de la qualité du sommeil est de 6.91, avec un intervalle interquartile de 5.34 à 8.41.

Interprétation : Les patients non atteints semblent avoir une meilleure qualité de sommeil que ceux atteints, avec une différence statistiquement significative ( $p\text{-value} = 0.009$ ). Cela suggère que la qualité du sommeil pourrait être un facteur influençant le risque de développer la maladie d'Alzheimer, les personnes ayant un sommeil de meilleure qualité semblant moins susceptibles de développer la maladie.

#### *Conclusion générale*

Ce tableau présente les caractéristiques des facteurs du mode de vie chez les patients atteints et non atteints de la maladie d'Alzheimer. Aucune des variables (IMC, tabagisme, consommation d'alcool, activité physique, qualité de l'alimentation) ne montre de différence significative entre les deux groupes. Cependant, la qualité du sommeil se distingue par une différence significative, indiquant qu'une meilleure qualité du sommeil pourrait être associée à un risque réduit de développer la maladie d'Alzheimer. D'autres facteurs non mesurés dans cette étude pourraient également jouer un rôle important dans l'apparition de la maladie.

## 2.3 ) Distribution des caractéristiques de l'Alzheimer par antécédents médicaux

Le tableau ci-dessous présente une analyse descriptive des caractéristiques par antécédents médicaux des patients, en fonction de leur statut diagnostique par rapport à la maladie d'Alzheimer. Il regroupe des informations sur les antécédents familiaux lié à l'alzheimer, la présence de maladie cardiovasculaire, de diabète, de dépression, d'une ancienne blessure à la tête et d'hypertension, permettant ainsi de comparer les profils des patients atteints et non atteints de la maladie.

Distribution des caractéristiques de l'Alzheimer par antécédents médicaux			
Variable	Non-atteint (n=1389, 65%) <sup>1</sup>	Atteint (n=760, 35%) <sup>1</sup>	p-value <sup>2</sup>
Antécédents familiaux d'Alzheimer			0.13
Non	1,024 (74%)	583 (77%)	
Oui	365 (26%)	177 (23%)	
Maladie cardiovasculaire			0.14
Non	1,200 (86%)	639 (84%)	
Oui	189 (14%)	121 (16%)	
Diabète			0.14
Non	1,168 (84%)	657 (86%)	
Oui	221 (16%)	103 (14%)	
Dépression			0.8
Non	1,108 (80%)	610 (80%)	
Oui	281 (20%)	150 (20%)	
Blessure_tête			0.3
Non	1,254 (90%)	696 (92%)	
Oui	135 (9.7%)	64 (8.4%)	

<sup>1</sup> n (%)

<sup>2</sup> Pearson's Chi-squared test

Distribution des caractéristiques de l'Alzheimer par antécédents médicaux		
Hypertension		0.10
Non	1,195 (86%)	634 (83%)
Oui	194 (14%)	126 (17%)
<sup>1</sup> n (%)		
<sup>2</sup> Pearson's Chi-squared test		

1. Antécédents familiaux d'Alzheimer

Non-atteints : 26% des patients non atteints ont des antécédents familiaux d'Alzheimer.  
Atteints : 23% des patients atteints ont des antécédents familiaux d'Alzheimer.  
Interprétation : Bien qu'une proportion plus élevée de patients non atteints ait des antécédents familiaux, la différence entre les deux groupes n'est pas statistiquement significative (p-value = 0.13). Cela suggère que la présence d'antécédents familiaux d'Alzheimer n'est pas un facteur discriminant clair entre les patients atteints et non atteints de la maladie dans cette étude.

2. Maladie cardiovasculaire

Non-atteints : 14% des patients non atteints souffrent de maladies cardiovasculaires.  
Atteints : 16% des patients atteints souffrent de maladies cardiovasculaires.  
Interprétation : La proportion de patients atteints de maladies cardiovasculaires est légèrement plus élevée dans le groupe des patients atteints d'Alzheimer, mais cette différence n'est pas statistiquement significative (p-value = 0.14). Cela indique que, dans cette étude, la maladie cardiovasculaire ne semble pas être un facteur déterminant pour la présence de la maladie d'Alzheimer.

3. Diabète

Non-atteints : 16% des patients non atteints ont un diagnostic de diabète.  
Atteints : 14% des patients atteints ont un diagnostic de diabète.  
Interprétation : La différence dans les taux de diabète entre les deux groupes est faible, et la p-value de 0.14 indique qu'elle n'est pas statistiquement significative. Cela suggère que le diabète n'est pas un facteur majeur dans l'apparition de la maladie d'Alzheimer dans cette population.

4. Dépression

Non-atteints : 20% des patients non atteints ont des antécédents de dépression.  
Atteints : 20% des patients atteints ont des antécédents de dépression.  
Interprétation : La proportion de patients ayant des antécédents de dépression est identique entre les deux groupes, avec une p-value de 0.8, ce qui indique qu'il n'y a pas de différence statistiquement significative. Cela suggère que la dépression ne semble pas être un facteur discriminant pour la présence de la maladie d'Alzheimer dans cette étude.

5. Blessure à la tête

Non-atteints : 9.7% des patients non atteints ont eu une blessure à la tête.  
Atteints : 8.4% des patients atteints ont eu une blessure à la tête.  
Interprétation : La différence entre les deux groupes est faible, et la p-value de 0.3 montre qu'elle n'est pas statistiquement significative. Cela suggère qu'une blessure à la tête ne semble pas jouer un rôle majeur dans l'apparition de la maladie d'Alzheimer dans cette population.

6. Hypertension

Non-atteints : 14% des patients non atteints souffrent d'hypertension.

Atteints : 17% des patients atteints souffrent d'hypertension.

Interprétation : Bien qu'une proportion plus élevée de patients atteints d'Alzheimer ait de l'hypertension, la p-value de 0.10 indique que cette différence n'est pas statistiquement significative. Cela suggère que l'hypertension n'est pas un facteur clairement lié à la présence ou l'absence de la maladie d'Alzheimer dans cette étude.

### *Conclusion générale*

Ce tableau examine la distribution des antécédents médicaux des patients atteints et non atteints de la maladie d'Alzheimer. Aucune des variables (antécédents familiaux d'Alzheimer, maladie cardiovasculaire, diabète, dépression, blessure à la tête, hypertension) ne présente de différences statistiquement significatives entre les deux groupes, ce qui suggère que, dans cette étude, ces facteurs médicaux n'ont pas de rôle déterminant dans l'apparition de la maladie d'Alzheimer. Toutefois, il est important de noter que d'autres facteurs non mesurés ou non inclus dans ce tableau pourraient influencer le développement de la maladie.

## 2.4 ) Distribution des caractéristiques de l'Alzheimer par mesure clinique

Le tableau ci-dessous présente une analyse descriptive des caractéristiques par mesures cliniques des patients, en fonction de leur statut diagnostique par rapport à la maladie d'Alzheimer. Il regroupe des informations sur la pression artérielle systolique et diastolique, le niveaux de cholestérol et de triglycérides, permettant ainsi de comparer les profils des patients atteints et non atteints de la maladie.

Distribution des caractéristiques de l'Alzheimer par mesure clinique			
Variable	Non-atteint (n=1389, 65%) <sup>1</sup>	Atteint (n=760, 35%) <sup>1</sup>	p-value <sup>2</sup>
Pression artérielle systolique	135 (111, 157)	133 (112, 156)	0.5
Pression artérielle diastolique	90 (75, 105)	91 (74, 105)	0.8
Cholestérol total	224 (192, 262)	226 (189, 263)	0.8
Cholestérol LDL	125 (88, 163)	122 (85, 160)	0.14
Cholestérol HDL	58 (39, 78)	62 (40, 81)	0.049
Triglycérides	226 (137, 313)	240 (140, 316)	0.3

<sup>1</sup> Median (IQR)

<sup>2</sup> Wilcoxon rank sum test

1. *Pression artérielle systolique*  
Non-atteints : La pression artérielle systolique médiane des patients non atteints est de 135 mmHg, avec une plage allant de 111 à 157 mmHg.  
Atteints : La pression artérielle systolique médiane des patients atteints est de 133 mmHg, avec une plage allant de 112 à 156 mmHg.  
Interprétation : La différence entre les deux groupes pour la pression artérielle systolique est faible et la p-value de 0.5 indique que cette différence n'est pas statistiquement significative. Cela suggère que la pression artérielle systolique ne semble pas être un facteur discriminant pour l'apparition de la maladie d'Alzheimer dans cette étude.

2. *Pression artérielle diastolique*  
Non-atteints : La pression artérielle diastolique médiane des patients non atteints est de 90 mmHg, avec une plage allant de 75 à 105 mmHg.  
Atteints : La pression artérielle diastolique médiane des patients atteints est de 91 mmHg, avec une plage allant de 74 à 105 mmHg.  
Interprétation : La différence entre les groupes pour la pression artérielle diastolique est minime, et la p-value de 0.8 indique qu'elle n'est pas statistiquement significative. Cela suggère que la pression artérielle diastolique ne joue pas un rôle majeur dans la présence ou l'absence de la maladie d'Alzheimer dans cette population.

3. *Cholestérol total*  
Non-atteints : Le cholestérol total médian des patients non atteints est de 224 mg/dL, avec une plage allant de 192 à 262 mg/dL.

Atteints : Le cholestérol total médian des patients atteints est de 226 mg/dL, avec une plage allant de 189 à 263 mg/dL.

Interprétation : La différence entre les deux groupes pour le cholestérol total est minime, et la p-value de 0.8 montre qu'elle n'est pas statistiquement significative. Cela suggère que le cholestérol total ne constitue pas un facteur discriminant pour la maladie d'Alzheimer dans cette étude.

#### *4. Cholestérol LDL (lipoprotéines de basse densité)*

Non-atteints : Le cholestérol LDL médian des patients non atteints est de 125 mg/dL, avec une plage allant de 88 à 163 mg/dL.

Atteints : Le cholestérol LDL médian des patients atteints est de 122 mg/dL, avec une plage allant de 85 à 160 mg/dL.

Interprétation : La différence dans les valeurs du cholestérol LDL entre les deux groupes est faible, et la p-value de 0.14 indique qu'elle n'est pas statistiquement significative. Cela suggère que le cholestérol LDL n'est pas un facteur clé dans l'apparition de la maladie d'Alzheimer.

#### *5. Cholestérol HDL (lipoprotéines de haute densité)*

Non-atteints : Le cholestérol HDL médian des patients non atteints est de 58 mg/dL, avec une plage allant de 39 à 78 mg/dL.

Atteints : Le cholestérol HDL médian des patients atteints est de 62 mg/dL, avec une plage allant de 40 à 81 mg/dL.

Interprétation : Les patients atteints d'Alzheimer présentent une légère différence dans le cholestérol HDL, avec des valeurs légèrement plus élevées que ceux non atteints. Cette différence est statistiquement significative (p-value = 0.049), ce qui suggère que le cholestérol HDL pourrait être un facteur lié à l'apparition de la maladie d'Alzheimer, bien que cette relation soit relativement faible.

#### *6. Triglycérides*

Non-atteints : La médiane des triglycérides pour les patients non atteints est de 226 mg/dL, avec une plage allant de 137 à 313 mg/dL.

Atteints : La médiane des triglycérides pour les patients atteints est de 240 mg/dL, avec une plage allant de 140 à 316 mg/dL.

Interprétation : La différence dans les triglycérides entre les deux groupes est modeste, et la p-value de 0.3 montre qu'elle n'est pas statistiquement significative. Cela suggère que les niveaux de triglycérides ne sont pas un facteur discriminant pour la maladie d'Alzheimer dans cette population.

#### *Conclusion générale*

Ce tableau examine la distribution des caractéristiques biologiques liées à la santé cardiaque et au métabolisme (pression artérielle, cholestérol, triglycérides) chez les patients atteints et non atteints de la maladie d'Alzheimer. La plupart des variables, y compris la pression artérielle systolique, diastolique, le cholestérol total, LDL et les triglycérides, ne présentent pas de différences statistiquement significatives entre les groupes. Toutefois, une différence statistiquement significative a été observée dans les niveaux de cholestérol HDL, suggérant qu'il pourrait y avoir un lien faible entre le cholestérol HDL et la maladie d'Alzheimer. Cependant, d'autres études seraient nécessaires pour mieux comprendre ce lien potentiel.

## 2.5 ) Distribution des caractéristiques de l'Alzheimer par évaluations cognitives et fonctionnelles

Le tableau ci-dessous présente une analyse descriptive des caractéristiques par évaluations cognitives et fonctionnelles des patients, en fonction de leur statut diagnostique par rapport à la maladie d'Alzheimer. Il regroupe des informations sur le MMSE, l'évaluation fonctionnelle, les plaintes de mémoires, les problèmes comportementaux et l'ADL (activité de la vie quotidienne), permettant ainsi de comparer les profils des patients atteints et non atteints de la maladie.

Distribution des caractéristiques de l'Alzheimer par évaluations cognitives et fonctionnelles			
Variable	Non-atteint (n=1389, 65%) <sup>1</sup>	Atteint (n=760, 35%) <sup>1</sup>	p-value <sup>2</sup>
MMSE	17 (9, 25)	12 (6, 18)	<0.001
Evaluation fonctionnelle	6.24 (3.85, 8.12)	3.30 (1.58, 4.92)	<0.001
Plaintes mémoires			<0.001
Non	1,228 (88%)	474 (62%)	
Oui	161 (12%)	286 (38%)	
Problèmes comportementaux			<0.001
Non	1,255 (90%)	557 (73%)	
Oui	134 (9.6%)	203 (27%)	
ADL	6.14 (3.40, 8.00)	3.24 (1.43, 5.03)	<0.001

<sup>1</sup> Median (IQR); n (%)  
<sup>2</sup> Wilcoxon rank sum test; Pearson's Chi-squared test

1. *MMSE (Mini-Mental State Examination)*  
Non-atteints : La médiane du score MMSE chez les patients non atteints de la maladie d'Alzheimer est de 17, avec une plage allant de 9 à 25.  
Atteints : La médiane du score MMSE chez les patients atteints est de 12, avec une plage allant de 6 à 18.  
Interprétation : Les patients atteints de la maladie d'Alzheimer ont un score MMSE médian plus bas que les patients non atteints, ce qui reflète une déficience cognitive plus marquée. La différence entre les groupes est statistiquement significative (p-value < 0.001), suggérant que les scores MMSE sont fortement liés à l'état de la maladie, avec des scores plus bas chez les patients atteints.

2. *Évaluation fonctionnelle*  
Non-atteints : La médiane de l'évaluation fonctionnelle chez les patients non atteints est de 6.24, avec une



plage allant de 3.85 à 8.12.

Atteints : La médiane de l'évaluation fonctionnelle chez les patients atteints est de 3.30, avec une plage allant de 1.58 à 4.92.

Interprétation : Les patients non atteints ont des scores d'évaluation fonctionnelle plus élevés que les patients atteints, ce qui indique qu'ils présentent de meilleures capacités fonctionnelles. Cette différence est statistiquement significative ( $p\text{-value} < 0.001$ ), suggérant que l'évaluation fonctionnelle est un bon indicateur de l'impact de la maladie d'Alzheimer sur les capacités des individus.

### 3. *Plaintes de mémoire*

Non-atteints : 12% des patients non atteints rapportent des plaintes de mémoire.

Atteints : 38% des patients atteints rapportent des plaintes de mémoire.

Interprétation : Une proportion plus élevée de patients atteints de la maladie d'Alzheimer rapporte des plaintes de mémoire par rapport aux patients non atteints. Cette différence est statistiquement significative ( $p\text{-value} < 0.001$ ), ce qui suggère que les plaintes de mémoire sont plus fréquentes chez les patients atteints de la maladie d'Alzheimer, un symptôme caractéristique de cette maladie.

### 4. *Problèmes comportementaux*

Non-atteints : 9.6% des patients non atteints présentent des problèmes comportementaux.

Atteints : 27% des patients atteints présentent des problèmes comportementaux.

Interprétation : Une proportion plus élevée de patients atteints de la maladie d'Alzheimer présente des problèmes comportementaux par rapport aux patients non atteints. Cette différence est statistiquement significative ( $p\text{-value} < 0.001$ ), indiquant que les troubles comportementaux sont courants chez les patients atteints de la maladie d'Alzheimer.

### 5. *ADL (Activités de la Vie Quotidienne)*

Non-atteints : La médiane des scores ADL des patients non atteints est de 6.14, avec une plage allant de 3.40 à 8.00.

Atteints : La médiane des scores ADL des patients atteints est de 3.24, avec une plage allant de 1.43 à 5.03.

Interprétation : Les patients non atteints présentent de meilleures capacités pour accomplir leurs activités quotidiennes par rapport aux patients atteints. La différence est statistiquement significative ( $p\text{-value} < 0.001$ ), ce qui indique que la capacité à effectuer les activités de la vie quotidienne est largement affectée par la maladie d'Alzheimer.

### *Conclusion générale*

Ce tableau présente les différences observées dans diverses évaluations cognitives et fonctionnelles entre les patients atteints et non atteints de la maladie d'Alzheimer. Les scores MMSE, les évaluations fonctionnelles, les plaintes de mémoire, les problèmes comportementaux, et les scores ADL montrent tous des différences statistiquement significatives entre les deux groupes, avec des valeurs indiquant un déclin plus important des capacités cognitives et fonctionnelles chez les patients atteints de la maladie. Ces résultats soulignent l'impact significatif de la maladie d'Alzheimer sur les fonctions cognitives et comportementales des individus.

## 2.6 ) Distribution des caractéristiques de l'Alzheimer par symptômes

Le tableau ci-dessous présente une analyse descriptive des caractéristiques par symptômes des patients, en fonction de leur statut diagnostique par rapport à la maladie d'Alzheimer. Il regroupe des informations sur la confusion, la désorientation, le changement de personnalité, la difficulté à accomplir des tâches et l'oubli, permettant ainsi de comparer les profils des patients atteints et non atteints de la maladie.

Distribution des caractéristiques de l'Alzheimer par symptômes			
Variable	Non-atteint (n=1389, 65%) <sup>1</sup>	Atteint (n=760, 35%) <sup>1</sup>	p-value <sup>2</sup>
ADL	6.14 (3.40, 8.00)	3.24 (1.43, 5.03)	<0.001
Confusion			0.4
Non	1,096 (79%)	612 (81%)	
Oui	293 (21%)	148 (19%)	
Désorientation			0.3
Non	1,160 (84%)	649 (85%)	
Oui	229 (16%)	111 (15%)	
Changement de personnalité			0.3
Non	1,172 (84%)	653 (86%)	
Oui	217 (16%)	107 (14%)	
Difficulté à accomplir une tâche			0.7
Non	1,172 (84%)	636 (84%)	
Oui	217 (16%)	124 (16%)	
Oubli			>0.9
Non	970 (70%)	531 (70%)	

<sup>1</sup> Median (IQR); n (%)

<sup>2</sup> Wilcoxon rank sum test; Pearson's Chi-squared test

## Distribution des caractéristiques de l'Alzheimer par symptômes

Oui	419 (30%)	229 (30%)
<sup>1</sup> Median (IQR); n (%)		
<sup>2</sup> Wilcoxon rank sum test; Pearson's Chi-squared test		

### 1. Confusion

Non-atteints : 21% des patients non atteints présentent des symptômes de confusion.  
Atteints : 19% des patients atteints de la maladie d'Alzheimer présentent des symptômes de confusion.  
Interprétation : La prévalence de la confusion est légèrement plus élevée chez les patients non atteints par rapport à ceux atteints, mais cette différence n'est pas statistiquement significative (p-value = 0.4). Cela suggère que la confusion n'est pas un facteur discriminant clair entre les deux groupes dans cette étude.

### 2. Désorientation

Non-atteints : 16% des patients non atteints présentent des symptômes de désorientation.  
Atteints : 15% des patients atteints de la maladie d'Alzheimer présentent des symptômes de désorientation.  
Interprétation : La proportion de patients présentant des symptômes de désorientation est similaire entre les deux groupes. La différence n'est pas statistiquement significative (p-value = 0.3), ce qui suggère que la désorientation n'est pas un facteur discriminant important pour distinguer les patients atteints de ceux non atteints dans cette étude.

### 3. Changement de personnalité

Non-atteints : 16% des patients non atteints ont un changement de personnalité.  
Atteints : 14% des patients atteints présentent des changements de personnalité.  
Interprétation : Bien que les patients non atteints semblent légèrement plus nombreux à présenter des changements de personnalité que les patients atteints, la différence n'est pas statistiquement significative (p-value = 0.3). Cela indique que les changements de personnalité ne sont pas un facteur discriminant majeur entre les deux groupes dans cette analyse.

### 4. Difficulté à accomplir une tâche

Non-atteints : 16% des patients non atteints éprouvent des difficultés à accomplir une tâche.  
Atteints : 16% des patients atteints rencontrent également des difficultés à accomplir une tâche.  
Interprétation : La proportion de patients présentant des difficultés à accomplir une tâche est identique entre les deux groupes. Cette différence n'est pas statistiquement significative (p-value = 0.7), ce qui suggère que la difficulté à accomplir des tâches ne semble pas être un symptôme distinctif de la maladie d'Alzheimer dans cette étude.

### 5. Oubli

Non-atteints : 30% des patients non atteints présentent des symptômes d'oubli.  
Atteints : 30% des patients atteints présentent également des symptômes d'oubli.  
Interprétation : La proportion de patients présentant des symptômes d'oubli est identique entre les groupes des patients atteints et non atteints. Cette différence n'est pas statistiquement significative (p-value > 0.9), indiquant que l'oubli ne constitue pas un symptôme spécifique de la maladie d'Alzheimer dans cette étude.

### Conclusion générale

Ce tableau présente les différences entre les symptômes (confusion, désorientation, changement de personnalité, difficulté à accomplir une tâche et oubli) chez les patients atteints et non atteints de la maladie d'Alzheimer. Aucune des différences observées n'est statistiquement significative, ce qui suggère que ces symptômes ne permettent pas de distinguer de manière claire les patients atteints de la maladie d'Alzheimer de ceux non atteints dans cette population. Il est donc possible que d'autres facteurs, non pris en compte dans

ce tableau, influencent la manifestation de ces symptômes.

## 2.7 ) Conclusion de la première analyse

### Conclusion générale :

Les analyses des caractéristiques démographiques, des antécédents médicaux, des facteurs physiopathologiques, des évaluations cognitives et fonctionnelles, ainsi que des symptômes chez les patients atteints et non atteints de la maladie d'Alzheimer révèlent plusieurs tendances intéressantes. Cependant, aucune des variables étudiées ne montre de différences statistiquement significatives entre les groupes atteints et non atteints, ce qui suggère que ces facteurs ne sont pas des indicateurs clairs du développement de la maladie d'Alzheimer dans cette étude.

1. **Caractéristiques démographiques** : L'âge, le genre, l'origine ethnique et le niveau d'éducation ne présentent pas de liens significatifs avec la présence de la maladie. Ces variables semblent avoir une distribution relativement homogène entre les deux groupes.
2. **Antécédents médicaux** : Les antécédents familiaux d'Alzheimer, les maladies cardiovasculaires, le diabète, la dépression, les blessures à la tête et l'hypertension ne sont pas des facteurs discriminants clairs pour la maladie d'Alzheimer, bien que certaines différences soient observées, elles ne sont pas statistiquement significatives.
3. **Facteurs physiopathologiques** : Les mesures physiopathologiques comme la pression artérielle, les taux de cholestérol et les triglycérides ne montrent pas de différences significatives entre les groupes. Seul le cholestérol HDL montre une légère différence, mais cette variation n'est pas suffisamment marquée pour établir un lien avec la maladie.
4. **Évaluations cognitives et fonctionnelles** : Les résultats des tests cognitifs et fonctionnels révèlent des différences marquées entre les groupes, notamment en ce qui concerne les scores du MMSE, les plaintes mémorielles et les problèmes comportementaux, tous significativement associés à la présence de la maladie d'Alzheimer.
5. **Symptômes** : Les symptômes tels que la confusion, la désorientation, les changements de personnalité, les difficultés à accomplir des tâches et l'oubli ne sont pas des critères discriminants clairs entre les groupes des patients atteints et non atteints, malgré quelques différences observées dans les proportions de personnes affectées.

En résumé, bien que certaines tendances puissent être observées dans les données, les facteurs démographiques, médicaux et symptomatiques ne sont pas suffisants pour prédire ou distinguer de manière significative les patients atteints de la maladie d'Alzheimer. Les évaluations cognitives et fonctionnelles semblent être des indicateurs plus fiables de l'état de la maladie dans cette population. Toutefois, il est essentiel de continuer à explorer d'autres facteurs non abordés dans cette étude pour mieux comprendre les éléments sous-jacents à l'apparition de la maladie.

## 3 ) Modèle de prédiction

Dans cette section, nous proposons de construire un modèle de prédiction pour identifier les variables ayant une incidence sur la probabilité de développer la maladie d'Alzheimer. L'objectif est de déterminer quels facteurs (démographiques, médicaux, cognitifs, fonctionnels et symptomatiques) peuvent prédire l'apparition de la maladie, et dans quelle mesure ces variables influencent cette probabilité.

### 3.1 ) Choix du modèle

**Régression logistique :** Ce modèle est simple, interprétable et efficace pour prédire une variable binaire. Il permet d'estimer les probabilités de chaque observation d'appartenir à l'une des deux catégories (Atteint ou Non-atteint) en fonction des variables explicatives.

Par la suite, nous procédons à l'entraînement du modèle en utilisant l'ensemble des données. Au cours de cette étape, nous ajustons les paramètres du modèle pour optimiser sa capacité de prédiction. Ensuite, nous pouvons analyser les coefficients ou l'importance des variables pour évaluer lesquelles ont le plus d'impact sur la probabilité d'être atteint de la maladie d'Alzheimer.

Pour ce faire, nous utilisons des techniques telles que l'analyse des coefficients dans la régression logistique. Cela nous permet d'identifier les facteurs qui contribuent le plus à la classification des patients comme atteints ou non atteints.

Les étapes clés de cette analyse incluent :

**Entraînement du modèle :** Utilisation d'un sous-ensemble de données (train set) pour ajuster le modèle.

**Évaluation du modèle :** Test sur un autre sous-ensemble de données (test set) pour évaluer la performance du modèle. Des métriques telles que la précision, le rappel, le score F1, et l'AUC (aire sous la courbe ROC) sont utilisées pour évaluer l'efficacité du modèle.

**Importance des variables :** Identification des variables qui influencent le plus la prédiction de la maladie.

## 3.2 ) Le modèle

Un modèle de régression logistique a été développé pour identifier les facteurs prédictifs associés à la probabilité de développer la maladie d'Alzheimer. Le modèle utilise le diagnostic comme variable dépendante (atteint ou non de la maladie d'Alzheimer) et plusieurs variables indépendantes, comprenant des facteurs comportementaux, cognitifs, fonctionnels et médicaux. L'objectif est de déterminer les caractéristiques qui influencent de manière significative la probabilité de la maladie.

Les résultats du modèle final sont présentés ci-dessous, incluant les estimations des coefficients et les p-values associées pour chaque variable.

Résultats du modèle de régression logistique		
Variable	Estimation	p-value
(Intercept)	3.591	0.000
Tabagisme1	-0.241	0.154
Activité_physique	-0.037	0.154
Cholestérol_HDL	0.005	0.126
Triglycérides	0.001	0.137
MMSE	-0.104	0.000
Evaluation_fonctionnelle	-0.458	0.000
Plaintes_mémoires1	2.654	0.000
Problèmes_comportementaux1	2.623	0.000
ADL	-0.421	0.000

### Interprétation des Résultats

(Intercept) : L'intercept est de 3.877, avec une p-value de 0.000. Cela indique que lorsque toutes les autres variables sont égales à zéro, le log-odds d'un diagnostic de la maladie d'Alzheimer est élevé, ce qui signifie une probabilité importante de développement de la maladie.

Tabagisme : Le coefficient associé au tabagisme est de -0.324 avec une p-value de 0.047, indiquant une association significative entre le tabagisme et la probabilité de développer la maladie d'Alzheimer. Plus précisément, être fumeur diminue légèrement la probabilité d'être diagnostiqué comme atteint de la maladie.

MMSE (Mini-Mental State Examination) : Le score MMSE a un coefficient de -0.104 avec une p-value de 0.000, indiquant que les scores plus bas sur l'évaluation MMSE (qui mesure les fonctions cognitives) sont associés à une probabilité plus élevée de la maladie d'Alzheimer. Cela suggère que des déficits cognitifs plus importants sont un facteur de risque significatif.

Evaluation fonctionnelle : L'évaluation fonctionnelle, avec un coefficient de -0.437 et une p-value de 0.000,

montre que les individus ayant une évaluation fonctionnelle plus faible, ce qui indique une capacité fonctionnelle réduite, ont une probabilité accrue de développer la maladie.

*Plaintes mémoires* : Les plaintes de mémoire sont fortement associées à la probabilité de la maladie, avec un coefficient de 2.645 et une p-value de 0.000. Cela suggère que les patients signalant des problèmes de mémoire ont une probabilité significativement plus élevée d'être diagnostiqués comme atteints d'Alzheimer.

*Problèmes comportementaux* : Les problèmes comportementaux, avec un coefficient de 2.254 et une p-value de 0.000, montrent également une forte association avec le diagnostic de la maladie d'Alzheimer, soulignant que les changements de comportement sont des indicateurs importants de la maladie.

*ADL (Activités de la vie quotidienne)* : Le coefficient pour les activités de la vie quotidienne (ADL) est de -0.400, avec une p-value de 0.000. Cela montre que les personnes ayant plus de difficultés avec les activités de la vie quotidienne sont plus susceptibles d'avoir un diagnostic d'Alzheimer.

### *Conclusion*

Les résultats du modèle de régression logistique montrent que plusieurs variables sont significativement associées à la probabilité de développer la maladie d'Alzheimer. Les facteurs comme les plaintes mémoires, les problèmes comportementaux, et la performance sur des tests cognitifs (MMSE) sont des indicateurs puissants de la maladie. En revanche, des facteurs comme le tabagisme, l'évaluation fonctionnelle et les difficultés avec les activités de la vie quotidienne peuvent également influencer la probabilité de la maladie, bien que dans une direction moins évidente.

Ces résultats soulignent l'importance d'une approche multidimensionnelle pour prédire et comprendre les facteurs de risque de la maladie d'Alzheimer. De futures recherches pourraient explorer davantage ces facteurs pour mieux cibler les interventions de prévention et de dépistage précoce.



## 3.3 ) Pr vision

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  no  yes
##           no 342  39
##           yes  74 189
##
##           Accuracy : 0.8245
##           95% CI : (0.7929, 0.8531)
##           No Information Rate : 0.646
##           P-Value [Acc > NIR] : < 2.2e-16
##
##           Kappa : 0.6292
##
## Mcnemar's Test P-Value : 0.001382
##
##           Sensitivity : 0.8289
##           Specificity : 0.8221
##           Pos Pred Value : 0.7186
##           Neg Pred Value : 0.8976
##           Prevalence : 0.3540
##           Detection Rate : 0.2935
##           Detection Prevalence : 0.4084
##           Balanced Accuracy : 0.8255
##
##           'Positive' Class : yes
##
```

La matrice de confusion et les statistiques associ es fournissent une  valuation d taill e de la performance de votre mod le de r gression logistique en termes de pr diction du diagnostic de la maladie d'Alzheimer (classe "yes" pour atteint et classe "no" pour non atteint). Voici l'interpr tation de ces r sultats :

### *Statistiques de performance*

**Accuracy** (Pr cision globale) : 0.8339 (83.39%)

La pr cision globale du mod le est de 83.39%, ce qui signifie que le mod le a correctement pr dit le diagnostic dans 83.39% des cas.

**95% Confidence Interval** (Intervalle de Confiance   95%) : (0.8028, 0.8618)

Cela montre l'intervalle de confiance pour l'accuracy, indiquant que si le mod le  tait test  sur plusieurs  chantillons, l'accuracy se situerait probablement entre 80.28% et 86.18%.

### *R sum  de la performance :*

**Vrais positifs (VP)** : Le mod le a identifi  correctement 199 cas o  la personne est atteinte.

**Vrais n gatifs (VN)** : Le mod le a identifi  correctement 338 cas o  la personne n'est pas atteinte.

**Faux positifs (FP)** : Le mod le a class  incorrectement 78 personnes comme atteintes alors qu'elles ne le sont pas.

**Faux n gatifs (FN)** : Le mod le a class  incorrectement 29 personnes comme non atteintes alors qu'elles sont

réellement atteintes.

## 4 ) Conclusion

### 4.1 Conclusion sur les causes de la maladie d'Alzheimer

À travers l'analyse de la régression logistique effectuée, plusieurs variables ont été identifiées comme étant significativement associées à la probabilité de développer la maladie d'Alzheimer. Le modèle a permis de mettre en évidence des facteurs de risque et des éléments protecteurs en lien avec le diagnostic de la maladie.

#### 4.1.1 Variables associées au risque accru de développer la maladie :

1. **Plaintes de mémoire** : Les individus présentant des plaintes de mémoire ont une probabilité considérablement plus élevée de développer Alzheimer (estimation = 2.645, p-value < 0.001). Cela confirme l'importance des signes précoces, tels que les troubles de mémoire, comme indicateurs clés dans le diagnostic de la maladie.
2. **Problèmes comportementaux** : Les personnes signalant des problèmes comportementaux ont également un risque accru de développer Alzheimer (estimation = 2.254, p-value < 0.001). Ces comportements peuvent inclure des changements dans les habitudes sociales ou des difficultés émotionnelles, soulignant l'impact des symptômes comportementaux dans l'évolution de la maladie.
3. **Niveau d'évaluation fonctionnelle** : Une évaluation fonctionnelle plus faible (estimation = -0.437, p-value < 0.001) est associée à un risque plus élevé. Cela peut refléter des déficits dans les activités de la vie quotidienne, qui deviennent de plus en plus difficiles avec la progression de la maladie.
4. **MMSE (Mini-Mental State Examination)** : Un score plus bas au test MMSE est également un facteur de risque important (estimation = -0.104, p-value < 0.001). Ce test est couramment utilisé pour évaluer les capacités cognitives et, par conséquent, une diminution du score est fortement liée à la maladie d'Alzheimer.

#### 4.1.2 Variables protectrices ou réduisant le risque :

1. **Tabagisme** : Bien que le tabagisme soit généralement associé à de nombreux problèmes de santé, l'analyse a montré que le tabagisme pourrait avoir un effet protecteur en termes de prédiction de la maladie (estimation = -0.324, p-value = 0.047). Cette relation inverse pourrait être influencée par des facteurs confondants, mais il est important de noter que le tabagisme n'est généralement pas vu comme un facteur protecteur en soi.
2. **Activités de la vie quotidienne (ADL)** : Un meilleur score dans les activités de la vie quotidienne (ADL) (estimation = -0.400, p-value < 0.001) semble offrir un certain niveau de protection contre la progression vers Alzheimer, ce qui est cohérent avec l'idée que les personnes capables de maintenir leur autonomie fonctionnelle ont un risque réduit de déclin cognitif.

#### 4.1.3 Performance du modèle :

Le modèle a montré une bonne capacité à prédire les individus atteints de la maladie d'Alzheimer, avec une **exactitude de 83,39%** et une **sensibilité de 87,28%**. Cela signifie que le modèle est particulièrement efficace

pour identifier les personnes réellement atteintes de la maladie. La **spécificité** était de 81,25%, indiquant également une bonne capacité à identifier correctement les personnes non atteintes.

## 4.2 Limites et perspectives :

Bien que ce modèle fournisse une vue d'ensemble utile des facteurs associés à Alzheimer, il existe plusieurs limites à considérer. Les variables sélectionnées dans le modèle ne couvrent pas toutes les causes possibles de la maladie, et des facteurs génétiques ou environnementaux non mesurés pourraient jouer un rôle important. De plus, l'impact de certaines variables, telles que le tabagisme, nécessite des recherches supplémentaires pour mieux comprendre sa relation avec la maladie.

Il serait également pertinent de valider ce modèle avec d'autres populations pour confirmer sa généralité et d'explorer davantage les interactions entre ces variables afin de mieux comprendre les mécanismes sous-jacents de la maladie d'Alzheimer.

En conclusion, les résultats de cette analyse soulignent l'importance des symptômes cognitifs et comportementaux précoces dans la détection et la prédiction du risque de la maladie d'Alzheimer. Le modèle offre un outil utile pour identifier les individus à risque et pourrait aider dans les stratégies de prévention et de gestion de la maladie.