

Marcadores del Alzheimer, su relación usando Redes Bayesianas

Santiago Mora, Gabriel Reynoso, Guillermo Villegas

Abstract

En el siguiente trabajo se realiza un análisis de datos para una base de datos con diferentes métricas relacionadas al diagnóstico de la enfermedad de Alzheimer (AD) proveniente del Alzheimer's Disease Neuroimaging Initiative (ADNI). El estudio utiliza Directed Acyclic Graphs (DAGs) para establecer relaciones entre variables. Mismas que fueron seleccionadas y relacionadas en conjunto con tres especialistas de la salud. De acuerdo a nuestros modelos realizados encontramos una considerable probabilidad de presentar menores tamaños del cerebro o del hipocampo en presencia de niveles altos de las neurotoxinas consideradas (beta-amiloide, Tau y Tau fosforilada), así como una alta relación de encontrar niveles altos de estas neurotoxinas ante la no duplicidad del gen APOE4

Introducción

En 2020 se estimaba, que en el mundo más de 50 millones de personas vivían con demencia, aunque no es el único padecimiento el Alzheimer representa entre 60 y 70% de los casos (Gómez-Virgilio L 2022). La enfermedad de Alzheimer (EA) es una condición neurodegenerativa que afecta funciones cognitivas como lo son la memoria, el juicio, lenguaje entre otras (Kumar et al. 2024). Al ocurrir en un espectro de gravedad, un diagnóstico temprano es crucial para realizar cambios en el estilo de vida del paciente para mejorar sus condiciones. El problema radica en que no existe una sola prueba diagnóstica estandarizada (Gómez-Virgilio L 2022). Existen pruebas de imagen como la positron emission tomography (PET) , de biomarcadores como lo son el péptido beta-amiloide o las hormonas tau, o pruebas cognitivas como la Clinical Dementia Rating Scale (CDRS) (Rukmangadachar and Bollu 2024). En el siguiente artículo se busca explorar las relaciones entre estos marcadores con el uso de redes bayesianas. EL Alzheimer's Disease Neuroimaging Initiative (ADNI) es un estudio que busca apoyar la investigación, del desarrollo, la prevención y el tratamiento de la EA, se utilizó la información de su base de datos.

Métodos

Comenzamos el estudio con el entendimiento de la enfermedad, la base de datos y las métricas involucradas. Esto fue posible gracias a los especialistas de la salud con los que consultamos sobre conocimientos de medicina para la realización de los modelos: Diana Villegas Morales, Alexis Álvarez Cano, Dr. Victor Cárdenas García. Los tres coincidieron en que al tratar con el diagnóstico de Alzheimer, lo más importante es la imagenología y algunas pruebas conductuales que teníamos como variables en la base de datos. Los tres recalcaron la importancia de la duplicidad en el gen APOE4, que básicamente te hace más propenso a desarrollar Alzheimer. Esto se alinea con nuestras fuentes bibliográficas consultadas. A como fue explicado por los médicos no se sabe a ciencia cierta si la duplicidad en este gen propicie la codificación de las neurotoxinas que utilizamos para el estudio: Beta-Amiloide, Tau, y Tau fosforilada. Estas neurotoxinas son responsables de dañar las neuronas en el cerebro, que se expresa como un decremento en su tamaño y en el hipocampo. Aquí es donde los médicos divergen en qué neurotoxinas afectan qué partes del cerebro. Después de que se perjudique el cerebro, es cuando podemos observar un mayor nivel para el Clinical Dementia Rating, métrica usada para medir qué tanta demencia es percibida en el paciente por el médico. Después de haber tenido estas discusiones, elaboramos las Directed Acyclic Graph (DAGs) haciendo énfasis en las relaciones de dependencia entre variables y platicamos con los médicos sobre qué probabilidades les gustaría conocer en base a las variables seleccionadas. Después de esto, limpiamos la base de datos, buscamos conocer las distribuciones de las variables mediante métricas de cuartiles, tendencia central y gráficas. Después de conocer nuestros datos, creamos las DAGs propuestas por los médicos y ajustamos una red Gaussiana lineal para poder hacer los cálculos de probabilidad.

Aplicación

Variables

APOE4: Cantidad de copias del gen APOE4. Casi todos los pacientes con dos copias de APOE4 presentan Alzheimer o características físicas de este como en los biomarcadores del líquido cefalorraquídeo y PET (Fortea et al. 2024).

ABETA: EL péptido beta amiloide que se deposita en las paredes extracelulares del cerebro, es un indicador de EA (Rukmangadachar and Bollu 2024)

TAU: Proteína Tau, un biomarcador de EA, se encuentre en el líquido cerebroespinal (Mecocci et al. 1998)

PTAU: Tau Fosforilada, marcador de la enfermedad encontrado en sangre (Ashton et al. 2023)

CDRSB: Clinical Dementia Rating. De una escala del 0 al 3, mide el grado de demencia percibido por la conducta del paciente. Siendo 0 no demencia y tres demencia grave. Este test se basa en darle al paciente preguntas de razonamiento (Khan 2016).

Hippocampus: Mediciones de hipocampo del paciente.

WholeBrain: Mediciones del cerebro del paciente.

Base de Datos

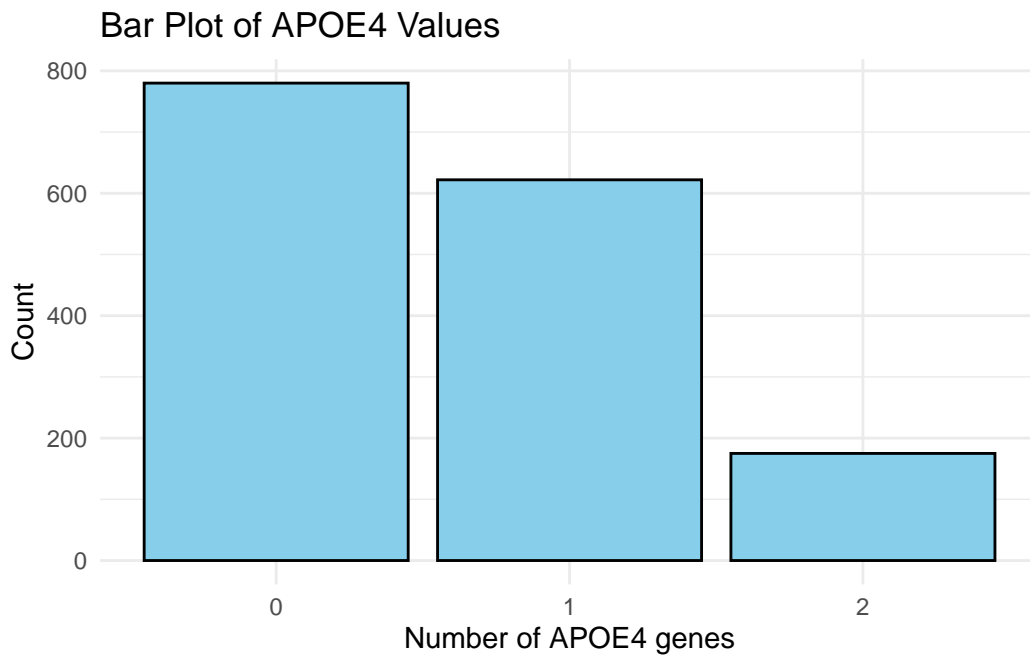
A tibble: 6 x 7

	APOE4	ABETA	TAU	PTAU	CDRSB	Hippocampus	WholeBrain
	<dbl>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>
1	0	<NA>	<NA>	<NA>	0	8336	1229740
2	1	741.5	239.7	22.83	4.5	5319	1129830
3	1	<NA>	<NA>	<NA>	6	5446	1100060
4	1	601.4	251.7	24.18	3.5	5157	1095640
5	1	<NA>	<NA>	<NA>	8	5139	1088560
6	0	1501	153.1	13.29	1	6869	1154980

Gráficas

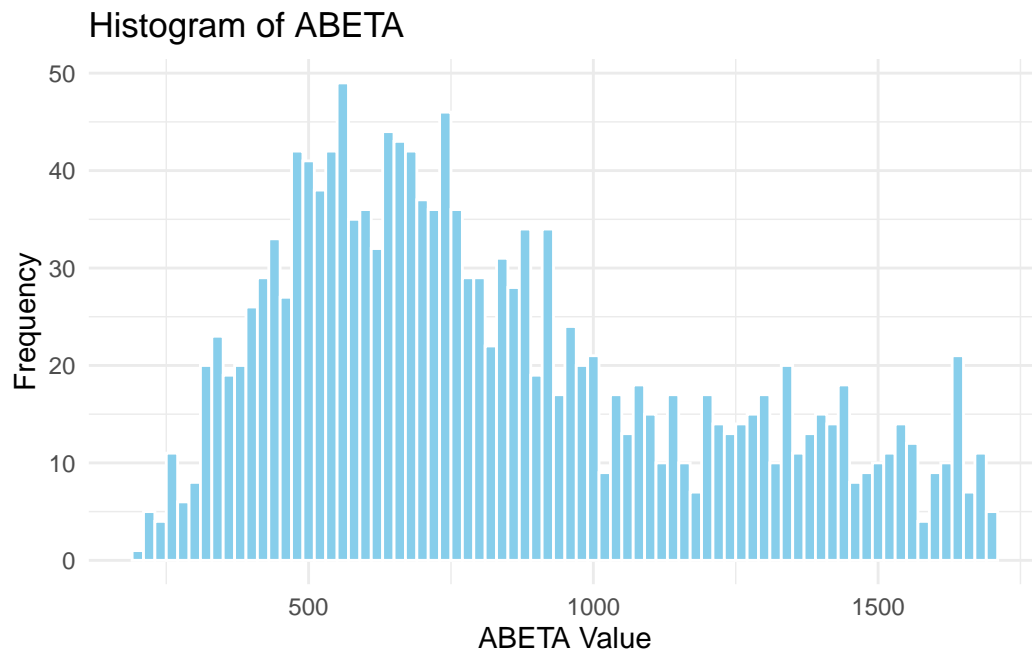
APOE4

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0000	0.0000	1.0000	0.6164	1.0000	2.0000



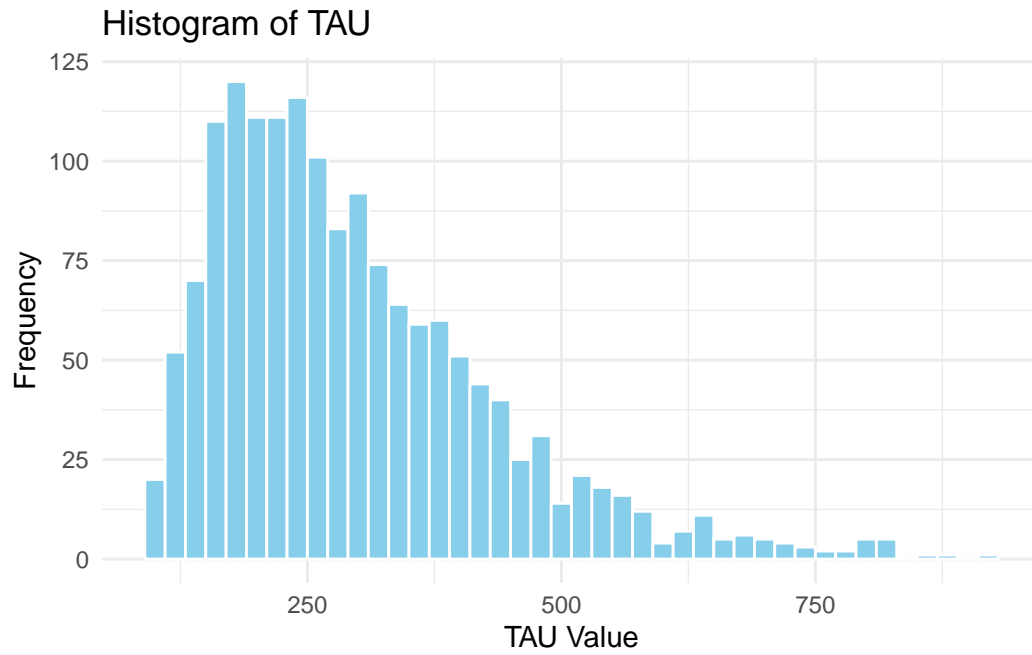
A-Beta

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
203.0	549.7	744.5	834.6	1079.0	1697.0



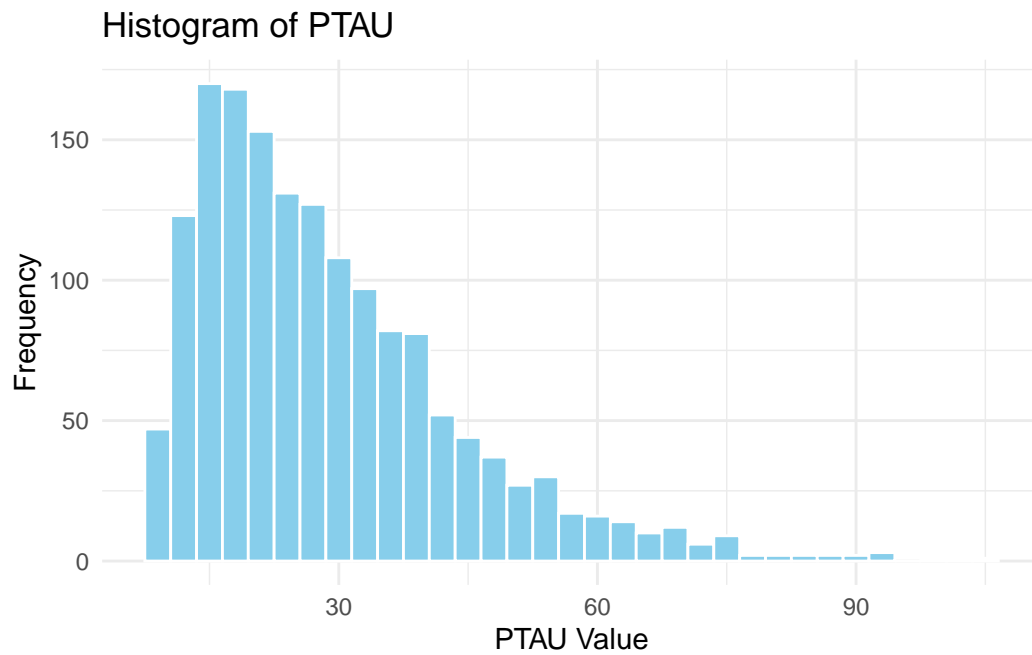
Tau

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
97.89	194.60	266.00	296.88	369.80	915.80



P-Tau

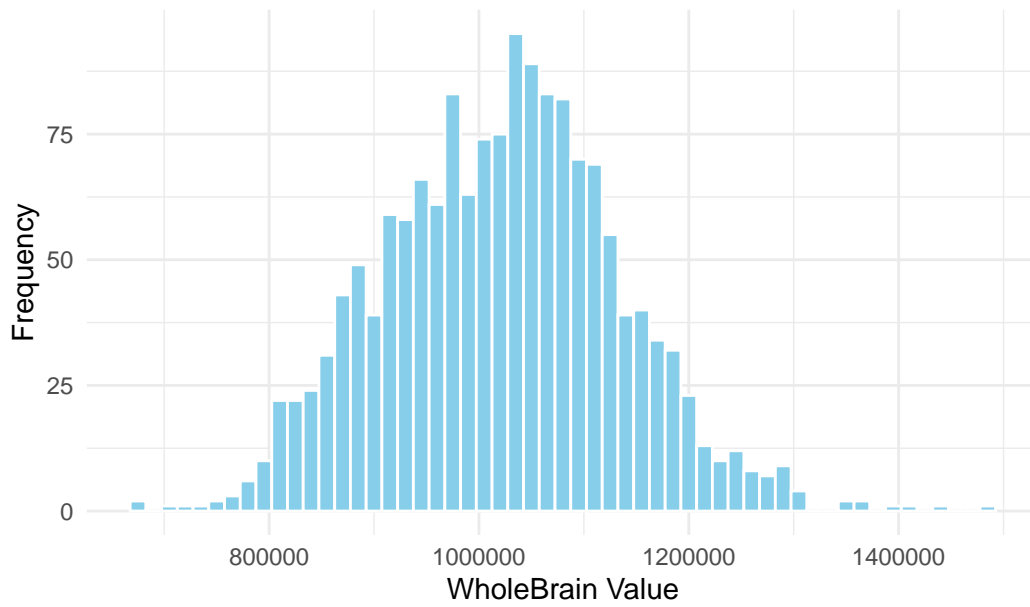
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
8.21	17.36	25.40	28.93	36.63	103.70



Whole Brain

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
669364	941642	1025490	1021647	1094210	1486040

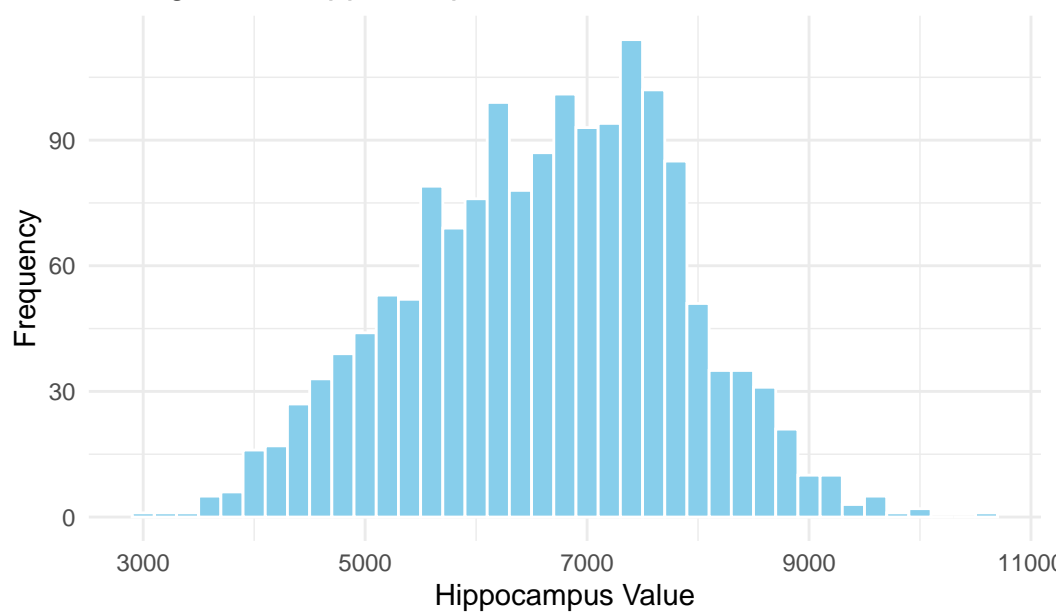
Histogram of WholeBrain



Hippocampus

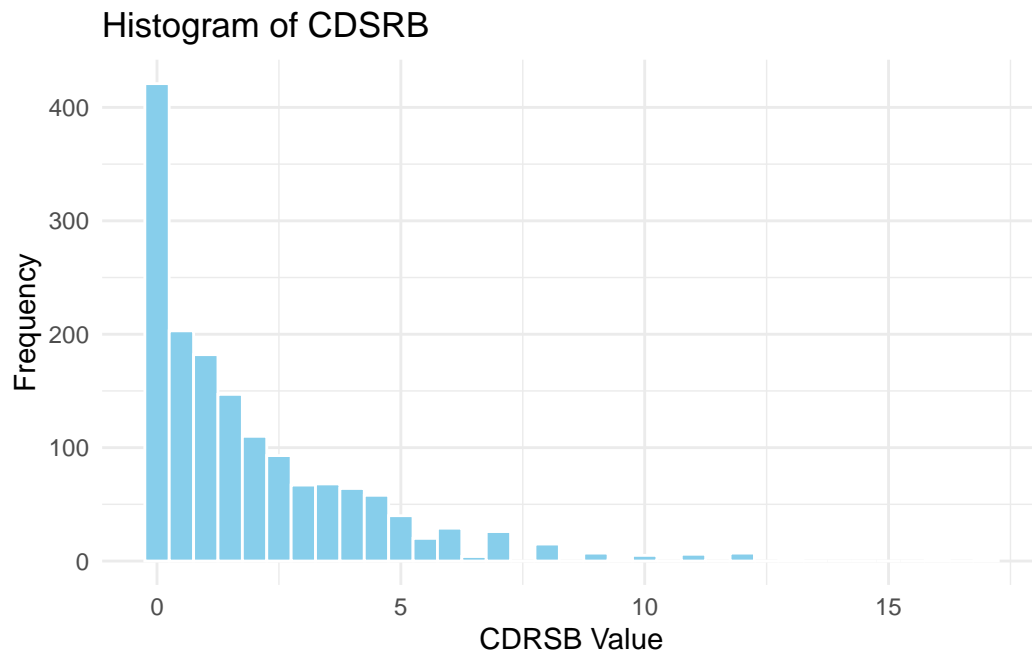
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
3054	5766	6707	6621	7498	10602

Histogram of Hippocampus



CDRSB

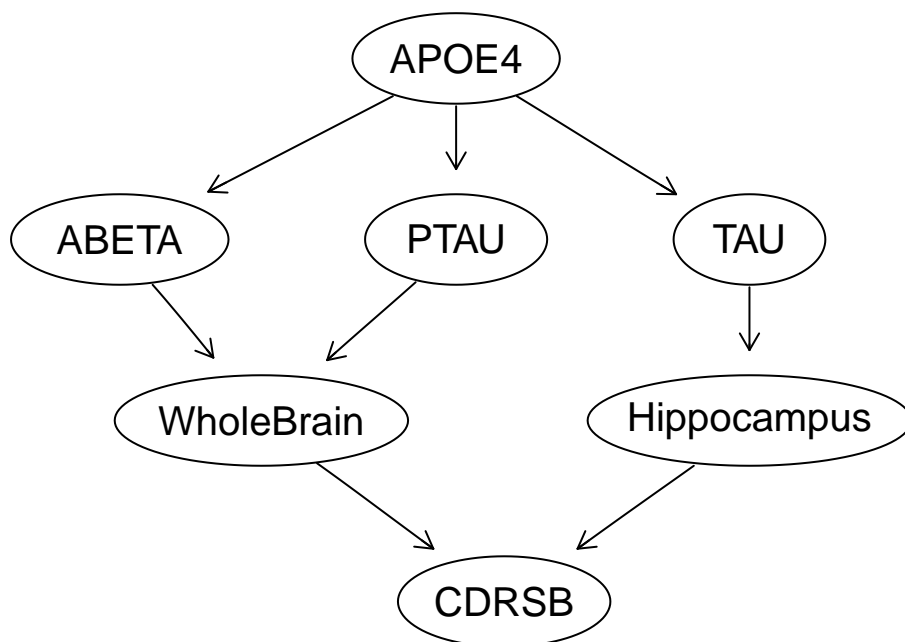
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	0.00	1.00	1.94	3.00	17.00



DAGs

DAG propuesta por Diana Villegas Morales, egresada de Medicina General

Loading required namespace: Rgraphviz



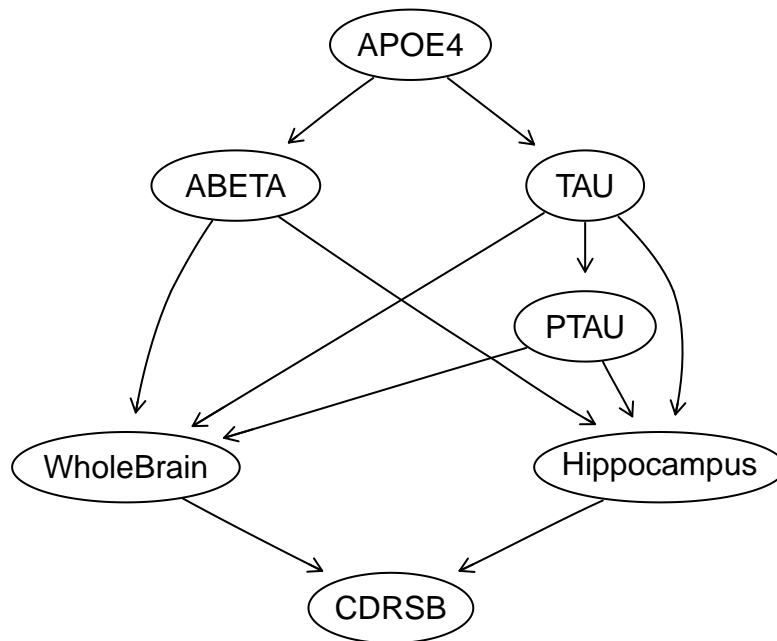
Score BIC para la DAG en cuestión.

[1] -66673.18

Score AIC para la DAG en cuestión.

[1] -66614.18

DAG propuesta por Alexis Álvarez Cano, estudiante de medicina



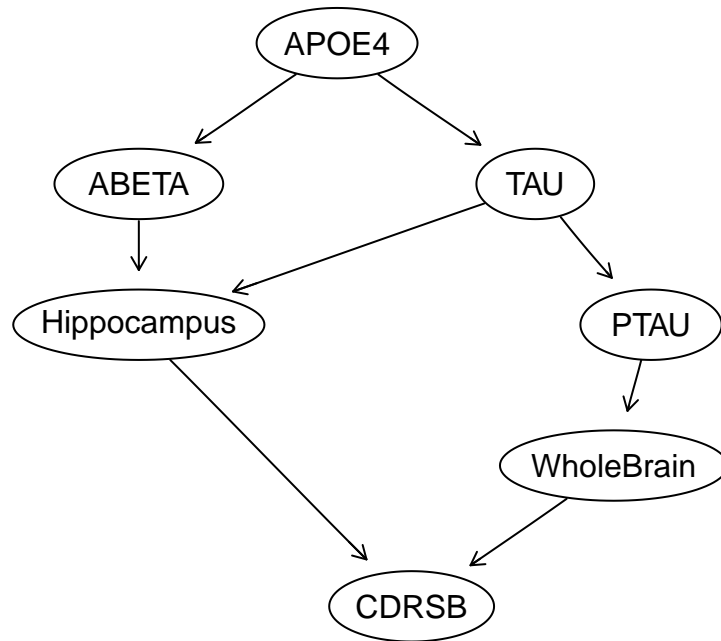
Score BIC para la DAG en cuestión.

[1] -63887.5

Score AIC para la DAG en cuestión.

[1] -63820.46

DAG propuesta por el Dr. Victor Cárdenas García, especialista en tratamiento gerátrico



Score BIC para la DAG en cuestión.

[1] -63919.75

Score AIC para la DAG en cuestión.

[1] -63860.75

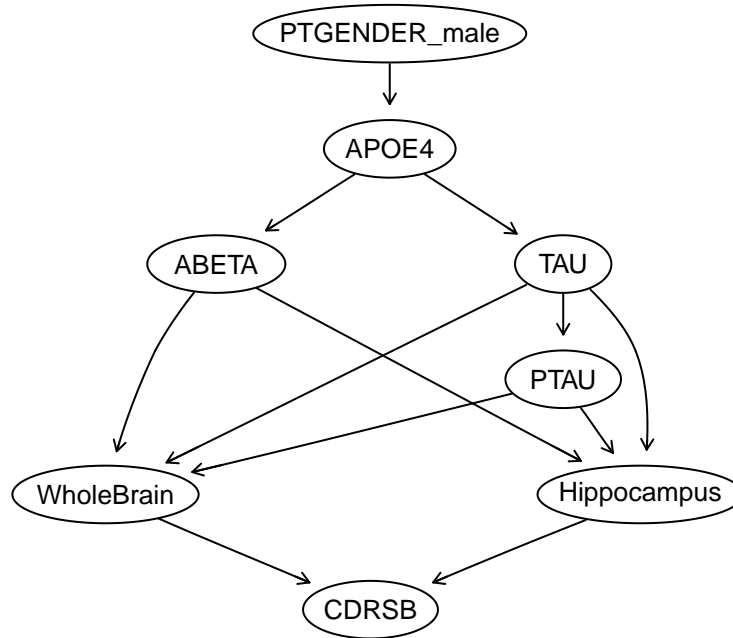
Inclusión de Variables categóricas en las DAGs

Si se quisieran agregar variables categóricas de interés a la Red Bayesiana Gaussiana, debería realizarse alguna de las siguientes alternativas:

La primera es discretizar las variables categóricas, es decir transformar la variable categórica en una numérica discreta. Esto se puede lograr mediante el uso de algún Encoding conocido, como lo es el One-Hot Encoding, que asigna 1 si se cumple que la variable sea cierta categoría o 0 en caso contrario. El problema con esto es que si hay C categorías, se tendrían que tener $C - 1$ variables diferentes para representarlas a todas, lo cual es práctico para variables como Sexo, pero puede no serlo para otras.

La otra alternativa es cambiar completamente modelo y, en vez de usar una Red Bayesiana Gaussiana, utilizar una Red Bayesiana Condicional (Híbrida). En este tipo de red, las variables numéricas continuas son modeladas como Gaussianas dadas las variables categóricas, mediante el condicionamiento de las distribuciones Gaussianas a cada una de las categorías. Es un modelo más sofisticado, el cual sería interesante utilizar a futuro para comparar resultados.

Una posible DAG que integre la variable categórica Sexo podría ser la siguiente:



Querying ajustando un modelo lineal

Probabilidad de tener un *clinical dementia rating* mayor a 2 cuando existe duplicidad en el gen APOE4 $\mathbb{P}(\text{CDRSB} > 2 \mid \text{APOE4} = 2)$

[1] 0.5792

Probabilidad de presentar tamaños pequeños de cerebro o de hipocampo en presencia de altos niveles de neurotoxinas $\mathbb{P}(\text{Hippocampus} < 5,889 \cup \text{WholeBrain} < 947027 \mid \text{Tau} > 34.48 \cup \text{aBeta} > 1076 \cup \text{pTau} > 34.48)$

[1] 0.4498819

Probabilidad de que se presenten niveles altos de neurotoxinas dado que no se presenta duplicidad en el gen APOE4

$$\mathbb{P}(Tau > 34.48 \cup aBeta > 1076 \cup pTau > 34.48 | APOE4 \neq 2)$$

[1] 0.98032

Conclusiones

De las tres relaciones planteadas por los expertos la de Alexis Alvarez fue la mejor en base a los puntajes de BIC y AIC. En esta DAG se supone que todas las neurotoxinas observadas en el estudio repercuten en el tamaño del cerebro y del hipocampo, a diferencia de los otros modelos. Además este modelo permitió responder algunas preguntas que podrían contribuir a la estandarización de pruebas de diagnóstico. Por último la exploración de incorporar variables categóricas a las Redes Bayesianas Gaussianas, permite la expansión de posibles redes permitiendo modelos más complejos.

Referencias

- Ashton, NJ, WS Brum, G Di Molfetta, AL Benedet, B Arslan, E Jonatis, RE Langhough, et al. 2023. “Diagnostic Accuracy of the Plasma ALZpath pTau217 Immunoassay to Identify Alzheimer’s Disease Pathology.” *medRxiv [Preprint]* 2023 (07.11.23292493). <https://doi.org/10.1101/2023.07.11.23292493>.
- Fortea, Juan, Jordi Pegueroles, Daniel Alcolea, Olivia Belbin, Oriol Dols-Icardo, Lúdia Vaqué-Alcázar, Laura Videla, et al. 2024. “APOE4 Homozygosity Represents a Distinct Genetic Form of Alzheimer’s Disease.” *Nature Medicine* 30 (5): 1284–91. <https://doi.org/10.1038/s41591-024-02931-w>.
- Gómez-Virgilio L, Silva-Lucero MdC, Reyes-Gutiérrez GS. 2022. “Etiología, Factores de Riesgo, Tratamientos y Situación Actual de La Enfermedad de Alzheimer En México.” *Gac Med Mex*, no. 158(4): 235–41. <http://dx.doi.org/10.24875/GMM.22000023>.
- Khan, T. K. 2016. “Chapter 2 - Clinical Diagnosis of Alzheimer’s Disease.” In *Biomarkers in Alzheimer’s Disease*, edited by T. K. Khan, 27–48. Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-12-804832-0.00002-X>.
- Kumar, Amit, Jasmine Sidhu, Frankie Lui, and John W. Tsao. 2024. *Alzheimer Disease*. StatPearls [Internet]. Treasure Island, FL: StatPearls Publishing.
- Mecocci, Patrizia, Annalisa Cherubini, Maddalena Bregnocchi, Fabio Chionne, Riccardo Cecchetti, Douglas T. Lowenthal, and Ugo Senin. 1998. “Tau Protein in Cerebrospinal Fluid: A New Diagnostic and Prognostic Marker in Alzheimer Disease?” *Alzheimer Disease and Associated Disorders* 12 (3): 211–14. <https://doi.org/10.1097/00002093-199809000-00015>.

Rukmangadachar, Lakshmi Anand, and Prashanth C. Bollu. 2024. “Amyloid Beta Peptide.” In *StatPearls*, Updated 2023 Aug 28. Treasure Island (FL): StatPearls Publishing. <https://www.ncbi.nlm.nih.gov/books/NBK459119/>.