IA

- práctica 4 -

Por:

Diego Caballé Casanova (738712)

1. Estudio previo

a) Explica razonadamente cuáles de estas propiedades son ciertas:

1. I y H son independientes dado S
2. V y H son independientes dado S

1. Existen dos camino de tripletas que conectan I y H. El primer camino, compuesto de las tripletas, I→C→P y luego C → P→ H y el segundo camino I→C→S y luego C→ S → H.

Para poder decir que I y H son independientes, no puede haber ningún camino activo. Y a pesar de que el segundo camino tiene una tripleta inactiva (C→ S → H), el primer camino tiene todas sus tripletas activas, ya que todas las tripletas son cadenas causales y ningún nodo intermedio está bloqueandola. Luego existe un camino activo, luego I y H son dependientes.

2. En este caso sólo existe un camino V → G → S y G → S → H. Como sólo existe un camino y es inactivo, pues S bloquea la cadena causal en G S H, no existe ningún camino activo, luego V y H son independientes dado S.

b) Eliminación de variables

Probabilidad de Covid en un paciente que ha sido contacto estrecho de un infectado y presenta síntomas, y no está vacunado contra la gripe.

P(c | +e, +s , ~v)

El primer paso es ignorar variables. Las variables ignorables son las que no son ancestro ni de una variable Query ni de una evidencia. Tanto H como P son ignorables pues no son ancestros ni de c, ni de e, ni de s, ni de v. En cambio, i y g no se pueden ignorar, pues son ancestros de s y de c.

P(I) P(E) P(C|I, E) P(~V) P(G|~V) P(S| C, G)

Ahora eliminamos tanto I como G, todas las demás son querys o evidencias del problema.

Empezamos por I.

Primero hacemos JOIN I → P(c,i| e) = 0.001 , P(c,i|-e) = 0.0001 , P(c,i| e) = 0.18 , P(c,i|-e) = 0.018

Después, sumamos las filas para eliminar i → P(c|e)= 0.181, P(-c|e)=0,819, P(c|-e)= 0.0181 P(-c|-e) = 0,9819

Ahora, tenemos estas variables, de la que nos queda G para dejar el problema en querys y evid.:

P(E) P(C|+E) P(~V) P(G|~V) P(+S| C, G)

Para eliminar G y mantener ~V haremos los siguientes pasos:

JOIN G|~V → P(s,g| c, ~v) = 0,072, P(s, -g| c, ~v) = 0,552, P(s, g| -c, ~v) = 0,064 P(s, -g| -c, ~v) = 0,0184

Después, sumamos las filas para eliminar g → P(s | c, ~v)= 0,624, P(s| -c, ~v)=0,0824

Después de eliminar las variables ocultas, hemos terminado con P(S | C, ~V) P(C|E) P(E) luego para obtener nuestro objetivo sólo tenemos que:

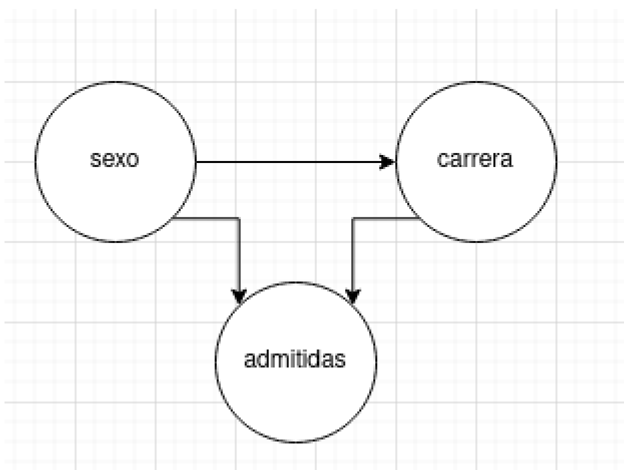
JOIN C → P(S,C|+E,~V)=0,112944, P(S, -C| E, ~V) = 0,0674856, P(S, C | -E, ~V)=0,01129,

P(S, -C | -E, ~V)=0,0809

Normalizamos S → P(S)=0,1804296 P(C|+S, +E, ~V) = 0,625972678

Segundo apartado

a..



b.

P(hombre) = 0,6589 P(mujer) = 0,3410

P(hombre| carrera A) =0,3878 P(mujer| carrera A)=0,09

P(hombre| carrera B) =0,2632 P(mujer| carrera B)=0,022

P(hombre| carrera C) =0,1527 P(mujer| carrera C)=0,53

P(hombre| carrera D) =0,1960 P(mujer| carrera D)=0,340

P(hombre| carrera A, admitido)	=0,62	P(mujer| carrera A, admitida)=0,82
P(hombre| carrera B, admitido)	=0,63	P(mujer| carrera B, admitida)=0,68
P(hombre| carrera C, admitido)	=0,369	P(mujer| carrera C, admitida)=0,34
P(hombre| carrera D, admitido)	=0,33	P(mujer| carrera D, admitida)=0,349
P(hombre| carrera A, noAdmit)	=0,38	P(mujer| carrera A, noAdmit)=0,18
P(hombre| carrera B, noAdmit)	=0,37	P(mujer| carrera B, noAdmit)=0,32
P(hombre| carrera C, noAdmit)	=0,631	P(mujer| carrera C, noAdmit)=0,66
P(hombre| carrera D, noAdmit)	=0,67	P(mujer| carrera D, noAdmit)=0,651

c.

Si se entiende la discriminación de género como el no acceso a una carrera por el hecho de ser mujer, no se ve reflejado en los datos, ya que el porcentaje de mujeres admitidas es mayor o parejo al de los hombres, salvo en la carrera C que es ligeramente inferior. Si esta discriminación de género se entiende por el hecho de ser hombre o mujer y no con un transfondo feminista, en la carrera A se podría decir que entran más fácil las mujeres que los hombres, aunque, como es la carrera más demandada entre los hombre y la segunda menos demandada entre las mujeres, se podría llegar a argumentar que más hombres intentando entrar implica que más hombres fallaran. Pero se puede contraargumentar con que la carrera C es la más demandada por las mujeres y la diferencia de admisión no es tan grande.

En conclusión, podría existir una discriminación de género en la carrera A, en donde las mujeres lo son más admitidas que los hombres.

d.

queremos saber P(admision | mujer), luego tenemos que eliminar la carrera, para eso, mezclaremos las 3 tablas en una:

P(mujer|carrera, admitida) x p(mujer|carrera) x p(mujer) =

P(mujer, carrera A, admitida) = 0,02516	P(mujer, carrera A, noAdmit)=0,005524
P(mujer, carrera B, admitida) =0,005100136	P(mujer, carrera B, noAdmit)=0,00240
P(mujer, carrera C, admitida) =0,06144	P(mujer, carrera C, noAdmit)=0,11928
P(mujer, carrera D, admitida) =0,04046	P(mujer, carrera D, noAdmit)=0,07536

Ya tenemos la variable carrera eliminada, y tenemos la tabla P(mujer, admision)

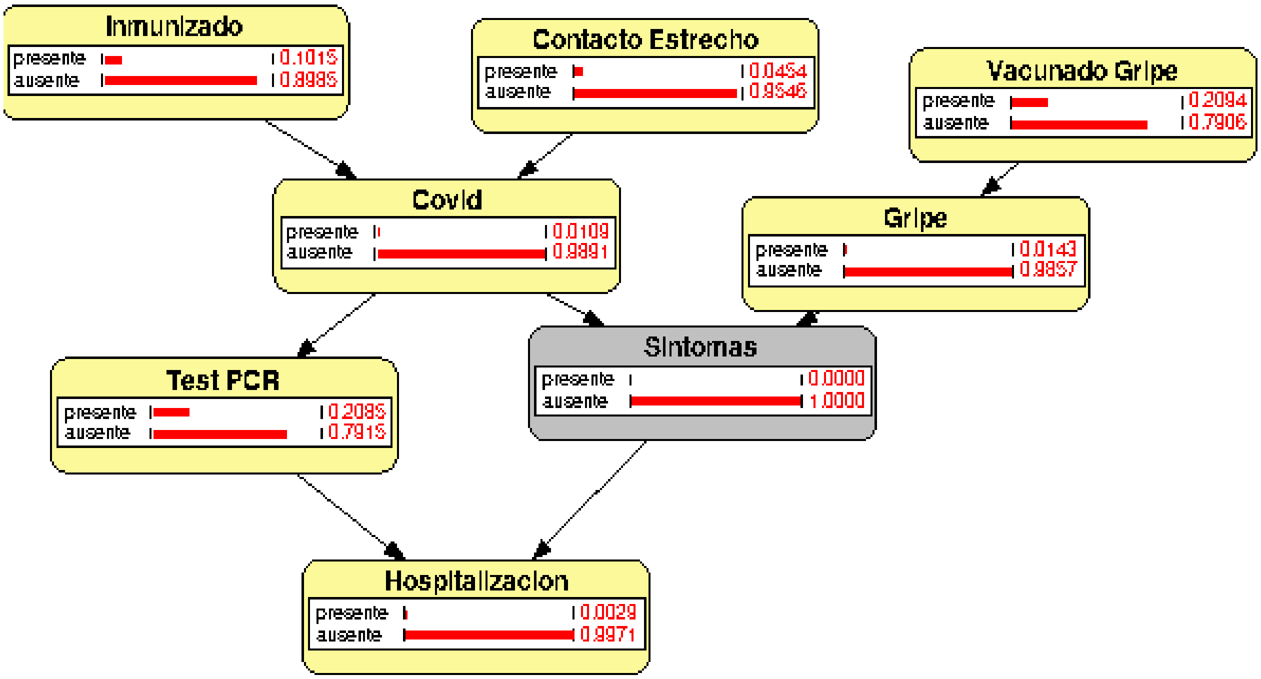
P(mujer, admitida)=0,132116	P(mujer,noAdmit)=0,202564

Luego P(admitida | mujer) = 0,3947

2. Desarrollo de la práctica

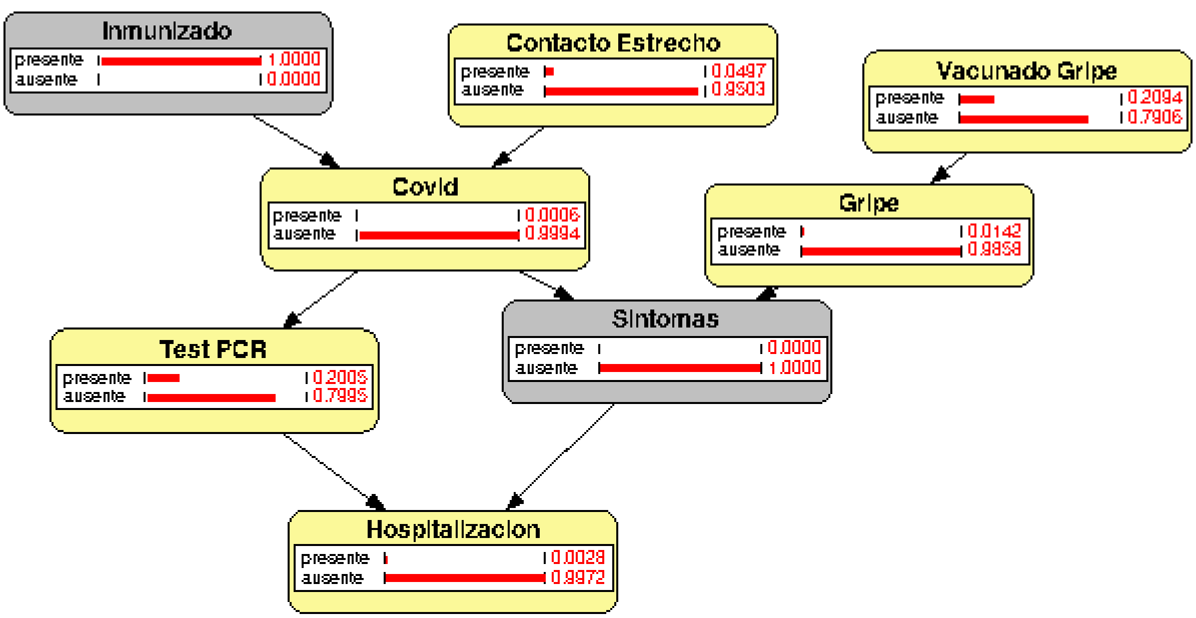
1 Covid

a) Partimos de esta situación

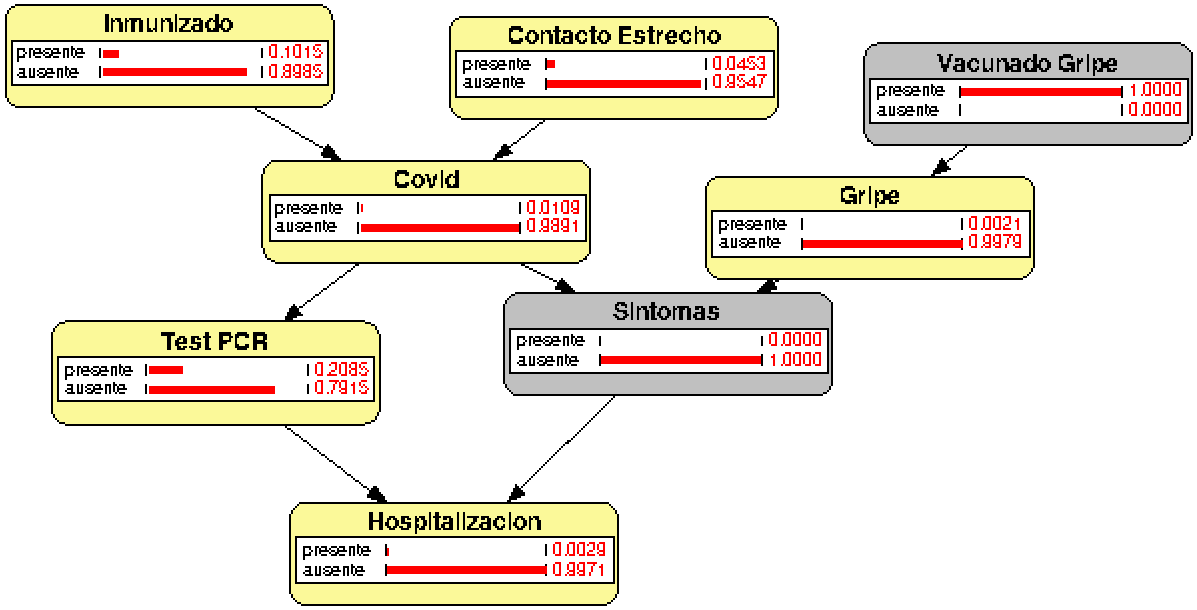


y queremos ver si I y H son independientes y V y H son independientes.

Al activar I, H cambia:

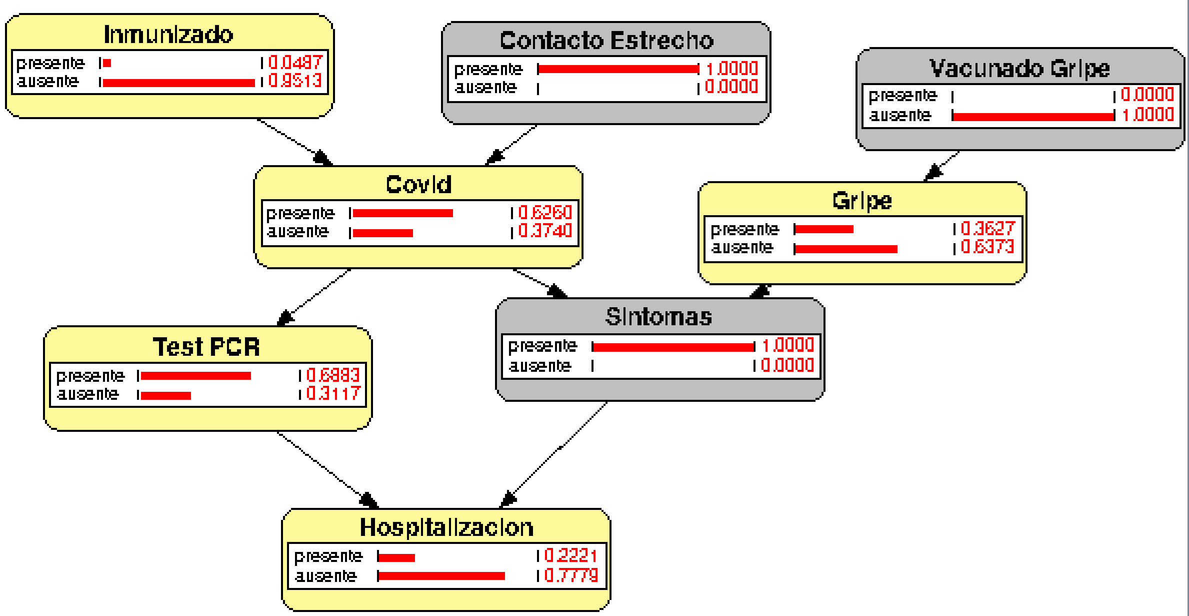


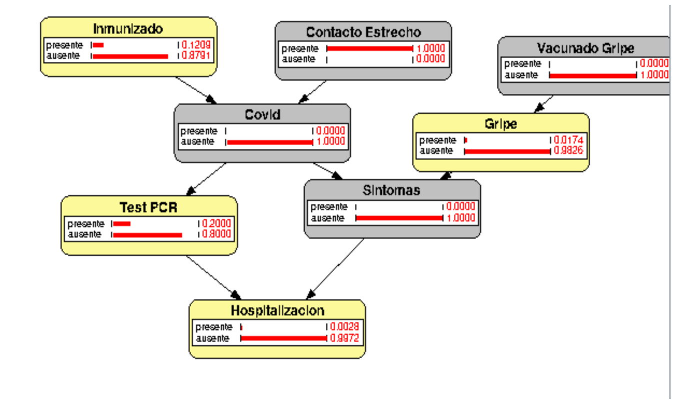
En cambio, con G activo, H no cambia:

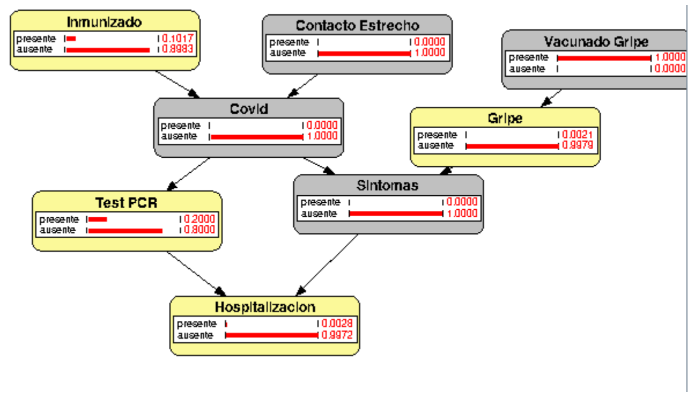


Si probamos con S presente, todos interactuan con todos, ya que todos intentan explicar los sintomas, estariamos en una causa común siendo un triplete activo, así que Covid y Gripe son dependientes para competir entre quien es la causa de los sintomas, haciendo que inmunizado y vacunado de gripe sean dependientes. Lo que tenemos que hacer es desactivar ese camino ausentando sintomas, entonces, sí que vemos como vacunado no influye en la hospitalización, pero inmunizado si que influye en la hospitalización.

b) La probabilidad de C es 0,6260 como lo previsto en el estudio previo







Si, como antes, desactivamos el camino del Covid, y el de sintomas, que influyen directamente sobre los resultados del PCR y sobre la hospitalización, contacto estrecho y no vacunado no alteran PCR y Hospitalización. Luego P y H son ignorables pues no cambian para todas las evidencias.

2 Berkley

a)

Hombre	0.657768
Mujer	0.342232

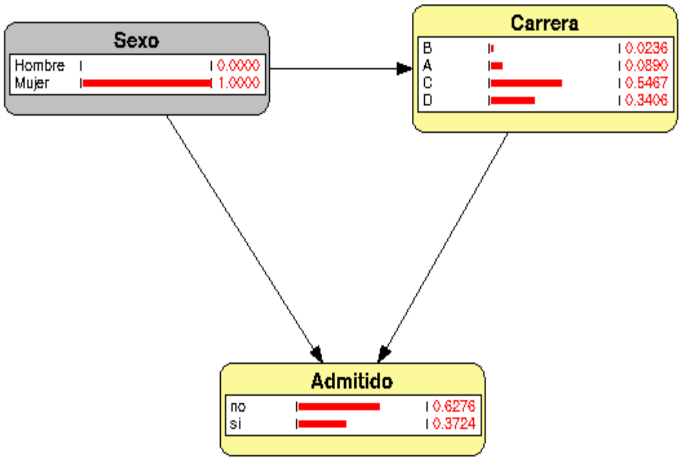
Sexo	Mujer	Hombre
B	0.023643	0.254178
A	0.089025	0.387572
C	0.546702	0.164236
D	0.34063	0.194014

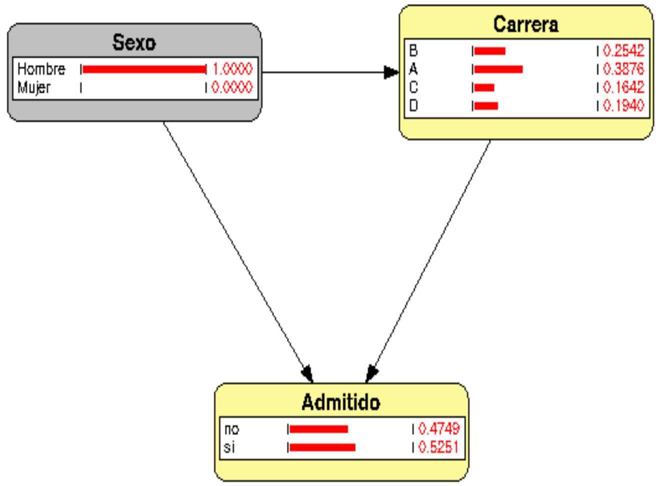
Sexo	Mujer	Mujer	Mujer	Mujer	Hombre	Hombre	Hombre	Hombre
Carrera	D	C	A	B	D	C	A	B
no	0.678938	0.680363	0.186275	0.329268	0.678404	0.620148	0.375784	0.376941
si	0.321062	0.319637	0.813725	0.670732	0.321596	0.379852	0.624216	0.623059

Los datos son bastante parejos, las ligeras diferencias (margen de error entre lo predicho y lo aprendido es del 0,01) que se pueden dar es porque la base de datos tiene más datos que nuestra tabla, todas las solicitudes son 3228, mientras que la BD tiene 5000 instancias con sus resultados.

b)

La probabilidad de admisión de una mujer es de 0,3724, la de un hombre es de 0,5251





c) las probabilidades son distintas y a favor de los hombres, pero eso nos sale al calcular las probabilidades globales, en cambio, en las probabilidades particulares de las carreras, nos surgía que las mujeres era mucho más probable que entrarán en la carrera A y las demás estuvieran bastante parejas. Es posible que estemos ante una paradoja de simpson, en donde en los datos agregados nos ha aparecido la tendencia contraria a la tendencia del grupo con la carrera A. Esto es porque a la tendencia de los datos le hemos dado una interpretación causal (la discriminación de género).