

¿Quiénes votaron la IVE?

Adriana Chazarreta - Germán Rosati - factor~data, IDAES/UNSAM

Introducción

Se habló mucho en estos días sobre la votación del proyecto de ley de “Interrupción Voluntaria del Embarazo” (IVE). Básicamente, se contaron votos y se dijeron cosas como las que aparecen en las [figuras 1 2, y 3](#):

Descompongamos el primer titular y tratemos de operacionalizar las variables que menciona:

- “los bonaerenses” -> habla de un efecto a nivel de “provincia” o “región”
- “oficialistas” -> hace referencia al bloque
- “jóvenes” -> bueno... éste es medio obvio

Ejemplos abundan y todos coinciden en general en señalar estas variables como determinantes. En algunos casos, se mencionó que los varones eran menos proclives a votar de forma positiva.

Ahora bien, cuando uno mira un poco la “metodología” atrás de estas variables se ve que en algunos casos están flojitas de papeles. Es decir, el proceso que lleva de los datos que se construyen a las conclusiones que se presentan parece estar un tanto sesgado hacia las prenaciones que se tienen respecto a la votación. El caso del titular de la “tribuna de doctrina” es bastante evidente.

Vamos a ver un ejemplo:

Fueron 50 los legisladores de la Provincia [de Buenos Aires] que se pronunciaron a favor, un número que permitió contrarrestar los votos opositores del norte del país, que fueron 49, dado que, en la región nortea, la mayoría rechazó la ley.

— La Nación 11/12/2020

Se observa cómo se utilizan conteos simples y distribuciones de frecuencias para llegar a las conclusiones. El problema acá es doble.

- Por un lado, no se tiene en cuenta la distribución relativa de la variable: es bastante probable que Buenos Aires sea la provincia que mayor cantidad de legisladores aporte... a cualquier votación! Simplemente porque es la que mayor cantidad de legisladores tiene.
- Por otro lado, estamos viendo de a una variable por vez... vemos primero la provincia, luego la edad, luego el bloque y así sucesivamente. El problema aquí se presenta si alguna de las variables está relacionada con otra. Por ejemplo, ¿qué pasaría si por algún motivo, las mujeres jóvenes tuvieran mayor probabilidad de votar a favor? ¿O si los legisladores del FDT de alguna provincia votaran de forma diferente al resto de sus compañeros de bloque? En términos estadísticos se llama a este hecho “interacción”, es decir, el efecto conjunto de dos variables sobre un mismo output.

ABORTO LEGAL CÁMARA DE DIPUTADOS DE LA NACIÓN

Aborto: el mayor aporte de votos a favor lo hicieron los bonaerenses, los oficialistas y los más jóvenes

Figura 1: Nota de La Nación del 11/12/2020

Cómo fueron los votos bloque por bloque

El bloque del Frente de Todos fue por lejos el que aportó la mayor cantidad de votos para conseguir la media sanción. De los 119 diputados que participaron de la sesión, 82 votaron de manera afirmativa. En contra votaron 32 diputados, mientras que tres se abstuvieron.

El PRO contabilizó 11 votos positivos, 42 negativos y una abstención, mientras que la Coalición Cívica sumó cuatro afirmativos, nueve negativos y una abstención. Los diputados Ignacio de Mendiguren (en uso de licencia por tener funciones en el BICE) y Eduardo Cáceres (PRO) estuvieron ausentes. Al comienzo de la sesión tampoco estaba Roxana Reyes, afectada por la muerte de un hijo, aunque finalmente la santacruceña inscribió su voto negativo.

Figura 2: Nota de Página12 del 11/12/2020

EL PAÍS

11 de diciembre de 2020

ABORTO LEGAL

CÁMARA DE DIPUTADOS DE LA NACIÓN

Río Negro fue el único distrito 100 por ciento verde; Formosa completamente celeste

Aborto Legal: cómo fue el voto provincia por provincia

Figura 3: Nota de Página12 del 11/12/2020

¿Cómo podríamos resolver ambos problemas? Una primera forma sería armar una bruta tabla que cruce todas las variables: género, edad, bloque, región y la variable dependiente (voto positivo o negativo). Para que nos demos una idea, esa tabla tendría aproximadamente unas 5.000 celdas. Así que, quizás no sea la forma más fácil de trabajar este problema.

Vamos a ir por una segunda opción. Vamos a tratar de construir un modelo simple que nos permita extraer algunas conclusiones sobre estas cinco variables y nos posibilite presentar la información de forma más compacta.

Modelando el voto a la IVE

Sin ponernos extremadamente técnicos, digamos que lo que intentamos hacer es estimar la probabilidad que cada diputada tiene de haber votado a favor del proyecto IVE. A su vez, esa probabilidad la “modelamos”¹ como una función del género, la región, el bloque y la edad de cada una.

La variable bloque la construimos de la siguiente forma:

- PRO, UCR y Coalición Cívica fueron clasificados como JxC y aliados
- El Frente de Todos quedó tal como está: FDT
- La izquierda, el socialismo y otros partidos provinciales fueron clasificados como resto y provinciales

A su vez, agrupamos las provincias en las siguientes regiones:

- NEA: Chaco, Corrientes, Formosa y Misiones
- NOA: Catamarca, Jujuy, La Rioja, Salta, Santiago del Estero y Tucumán
- Cuyo: Mendoza, San Juan y San Luis
- Centro: Córdoba, Entre Ríos y Santa Fe
- Patagonia: Chubut, La Pampa, Neuquén, Río Negro, Santa Cruz y Tierra del Fuego
- CABA
- Buenos Aires

Bueno, basta... ¡Mostrame la DATA!

Todo muy lindo pero vamos a ver entonces qué resultados arrojó el modelo. Un primer análisis que se puede intentar hacer es mirar qué variables son “significativas”. Es decir, cuáles son las X 's que tienen una “influencia” que podemos asumir diferente a cero. Eso lo podemos ver analizando los llamados **p-valores**².

En este caso particular, la aplicación de los p-valores en el sentido puramente estadístico del término puede ser confusa: dado que se trata de herramientas de la estadística inferencial y aquí estamos lidiando con una población (la totalidad de las diputadas) pareciera que aquí no corre la posibilidad de hacer inferencias.

No obstante, uno podría argumentar dos cosas:

¹ Que es una forma elegante de decir, “la analizamos como” o “la pensamos como”

Para ser más exactos (y hacernos les capes), modelamos la probabilidad con la siguiente fórmula:

$$P(y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4)}}$$

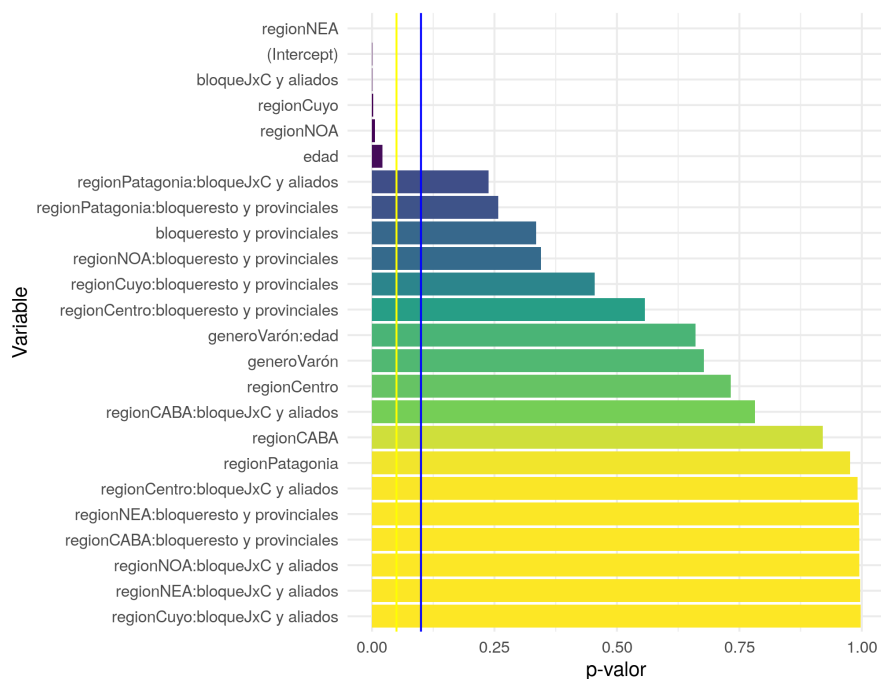
Donde $P(y = 1)$ es la probabilidad de voto afirmativo; X_1 es el género, X_2 es la edad, X_3 es el bloque, X_4 es la región.

Este modelo se llama *regresión logística*. Una explicación muy clara de esta técnica pueden encontrarla [acá](#) o una un poco más técnica [acá](#)

² Los p-valores (aplicados a este problema puntual) de cada variable vendrían a ser algo así como la probabilidad de haber obtenido los datos observados si la influencia de cada variable fuera cero. De esta forma, a menor p-valor, mayor significatividad de la variable en cuestión. Convencionalmente, p-valores menores a 0.05 se consideran significativos.

- Que se puede usar el p-valor de forma “conceptual” y no “estadística” (gracias a [Diego Koslowski](#) y [Octavio Bramajo](#) por sus comentarios al respecto). Es decir, lo usamos como una forma de ver qué variables tienen mayor influencia sobre la votación. También podríamos haber visto los tamaños de los coeficientes y ya.
- Que (y aquí la cosa se torna un poco más filosófica) la votación que observamos es una realización de un proceso aleatorio. Es decir que si realizáramos la votación muchas veces, obtendríamos un resultado diferente. Y en ese sentido la votación que observamos se trataría de una “muestra”, no de una población, sino de muchas otras votaciones posibles. Gary King en [este libro](#) lo explica de forma bastante clara.

Además de todo esto, hay una gran discusión en la comunidad estadística acerca de los p-valores y los problemas que genera su uso. Les recomendamos como primera aproximación [esta charla](#) que dio [Laura Ación](#) al respecto. [Acá está el video de la charla completa](#) (junto con las exposiciones de Walter Sosa Escudero y Germán Feierherd).



Veamos, entonces, la figura 4. Ahí marcamos un primer nivel de significatividad en color amarillo ($p\text{-value} < 0.05$) y otro en azul ($p\text{-value} < 0.10$). Si vemos las variables que están por debajo de la línea amarilla, veremos que la región (particularmente, NEA, CUYO y NOA), el bloque (puntualmente el de JxC) y la edad son las variables más relevantes.

A su vez, y contra algunas afirmaciones un tanto apresuradas que estuvieron circulando, no parece tan cierto que el género sea una variable relevante en

Figura 4: Variables relevantes para explicar el voto de les diputades

esta votación. Y aquí vemos una de las ventajas de este tipo de análisis (que desde ya, no es el único posible, ni el necesariamente mejor, ni tampoco uno definitivo).

Si vemos el conteo crudo de votos, efectivamente los hombres votaron en menor medida que las mujeres a favor de la IVE: 48 % contra 55 %. Pero pareciera que existen otras variables afectando.

Cuadro 1: Distribucion de votos por género

votacion_final1	Mujer	Varón
negativo	44.86 % (48)	51.68 % (77)
positivo	55.14 % (59)	48.32 % (72)

Efectivamente, al introducir las otras tres variables, la significatividad del género desaparece. Es por ello que este tipo de análisis multivariados son relevantes.

Ahora bien, esto solo nos habla de la importancia de cada variable. Pero no nos dice nada acerca de la FORMA en que cada variable independiente afecta a la dependiente. Para eso, vamos a tener que hacer algunas cuentas. Como puede verse en la nota 1 nuestro modelo da una ecuación que sirve para vincular las variables independientes con la dependiente. Dado que nuestro modelo tiene muchos coeficientes, vamos a pensar en un ejemplo más simple.

Supongamos, entonces, que la probabilidad de voto positivo dependiera solamente de la edad. En este ejemplo hipotético, corremos nuestra regresión y obtenemos el siguiente modelo:

$$P(y = 1|X) = \frac{1}{1 + e^{-(1-0,004 \times edad)}}$$

De esta forma, si imagináramos en una diputada que tiene 34 años podríamos calcular la probabilidad de voto afirmativo reemplazado $edad = 34$:

$$P(y = 1|X) = \frac{1}{1 + e^{-(1-0,004 \times 34)}} = 0,703$$

Es decir que este diputado tendría una probabilidad de 70 % de votar a favor de la IVE. Bancamos mucho a este diputado virtual.

Ahora bien, con esta idea podemos construir “perfiles hipotéticos” de diputados que nos ayuden a entender mejor cómo interactúan las diferentes variables. A esta operación se la suele llamar [post-estimación](#).

Veamos: las figuras 5 a 11 realizan esta operación para las diferentes regiones. Es decir, para cada región construyen “diputados frankenstein” de diferentes edades, género y bloque y sus respectivas probabilidades de votar a favor de la IVE. A título de ejemplo, veamos la figura de Buenos Aires

Cuando consideramos la pertenencia partidaria interactuando con la edad, el sexo y la región, podemos observar varias cuestiones.

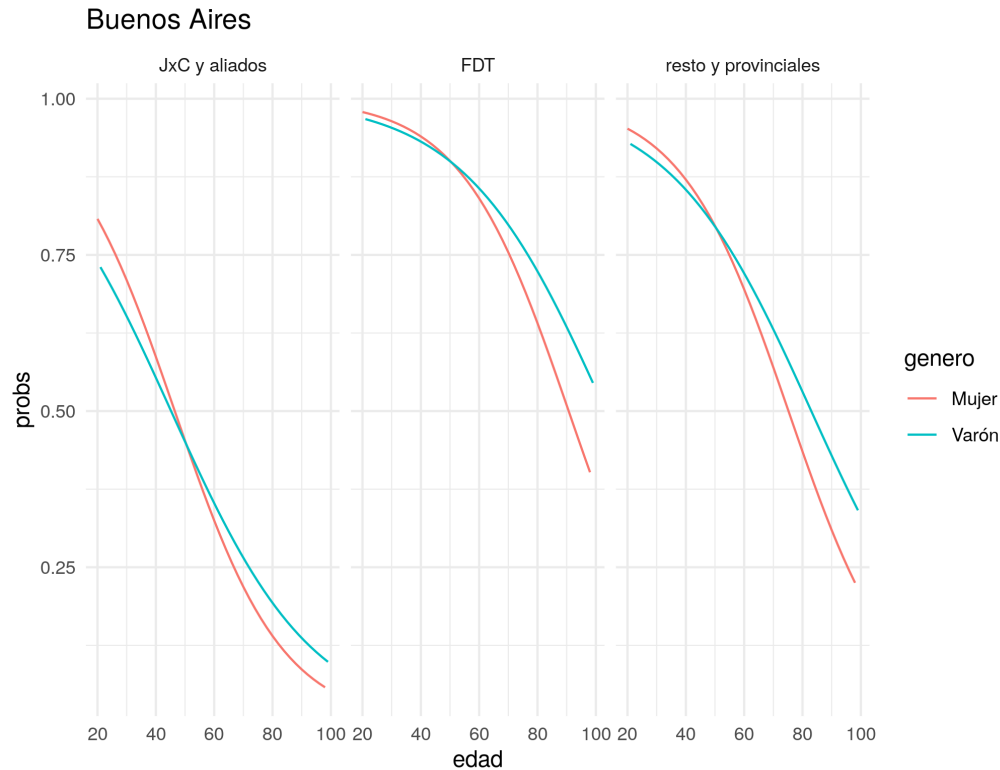


Figura 5: Interacciones entre variables para explicar el voto de les diputades de Buenos Aires

1. En términos generales, no pareciera registrarse diferencias significativas entre varones y mujeres.
2. A su vez, a medida que aumenta la edad disminuye la probabilidad de votar afirmativamente.
3. El FDT tiene mayor probabilidad de voto afirmativo a la IVE.
4. Sin embargo, en Centro, Cuyo, NEA y NOA, el efecto más fuerte es el del bloque de diputades de JxC y aliados: no se observan diferencias en la probabilidad de voto positivo por edad. En estas regiones, el voto de JxC es siempre negativo.

Por lo tanto, pareciera ser más relevante para determinar el voto, la región (a partir de la provincia representada) y la edad que la influencia de otras variables.

Para ir terminando...

Con este ejercicio quisimos mostrar con un caso concreto cómo partiendo de hipótesis intuitivas (y no tan intuitivas) se puede construir un problema de investigación (muy acotado) que permite resolverse a partir de técnicas estadísticas. Los resultados de este ejercicio posibilitan analizar cómo interactúan

diferentes variables y ser más preciso a la hora de interpretar las dimensiones que influyen al momento de votar un proyecto “polémico” como es el de la IVE.

Obviamente, la idea no es sacar conclusiones definitivas sino aportar una de las formas posibles de realizar una lectura de la votación en Diputados. Estamos seguros que éste es sólo un primer ejercicio y que quedan pendientes algunos aspectos: por un lado, la posibilidad de incluir otras variables relevantes que también pueden influir en la decisión de la votación como por ejemplo la creencia o la práctica religiosa y por el otro, la aplicación de otras técnicas y modelos que complementarían y enriquecerían los resultados alcanzados aquí.

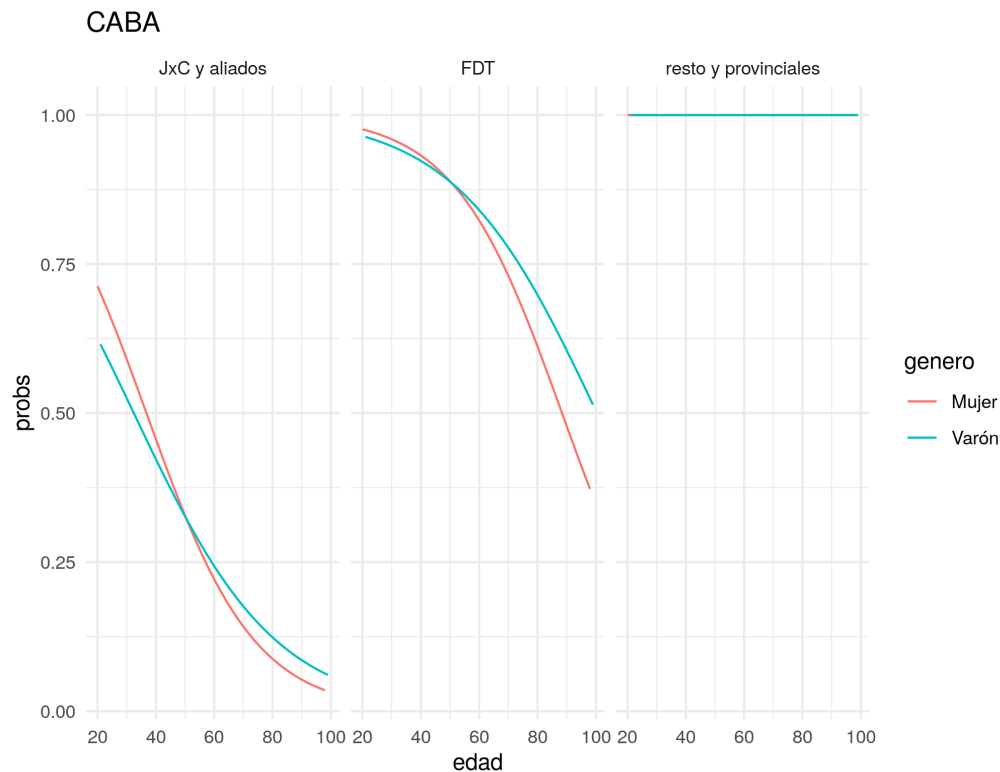


Figura 6: Interacciones entre variables para explicar el voto de les diputades de CABA

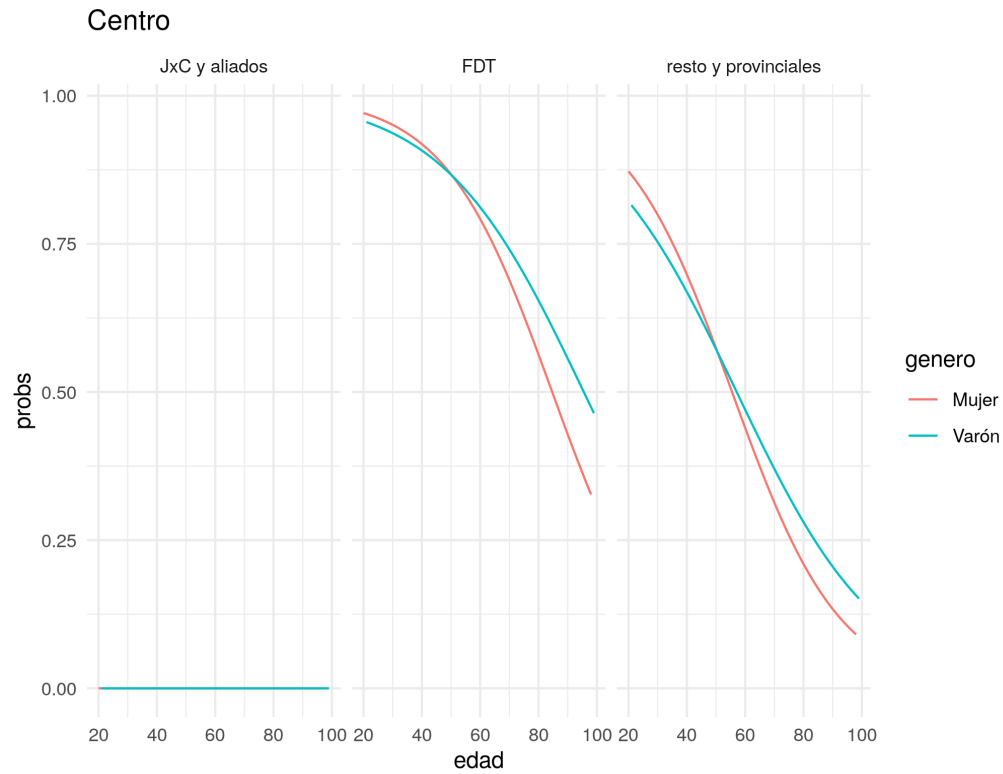


Figura 7: Interacciones entre variables para explicar el voto de los diputados de la región Centro

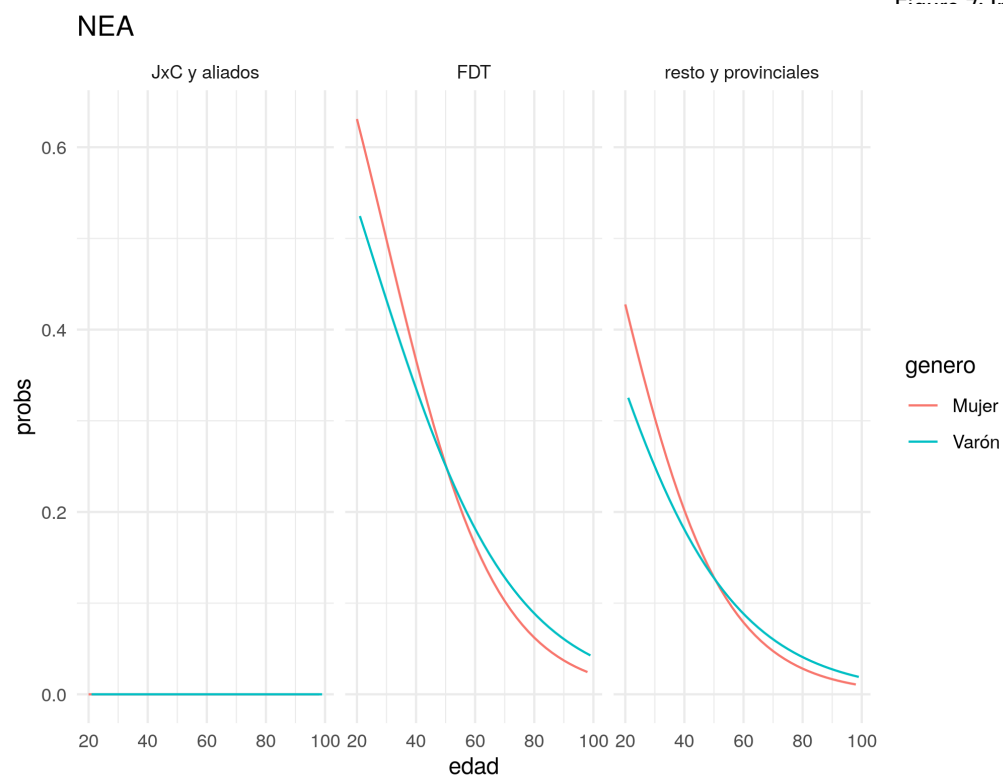


Figura 8: Interacciones entre variables para explicar el voto de los diputados de la región NEA

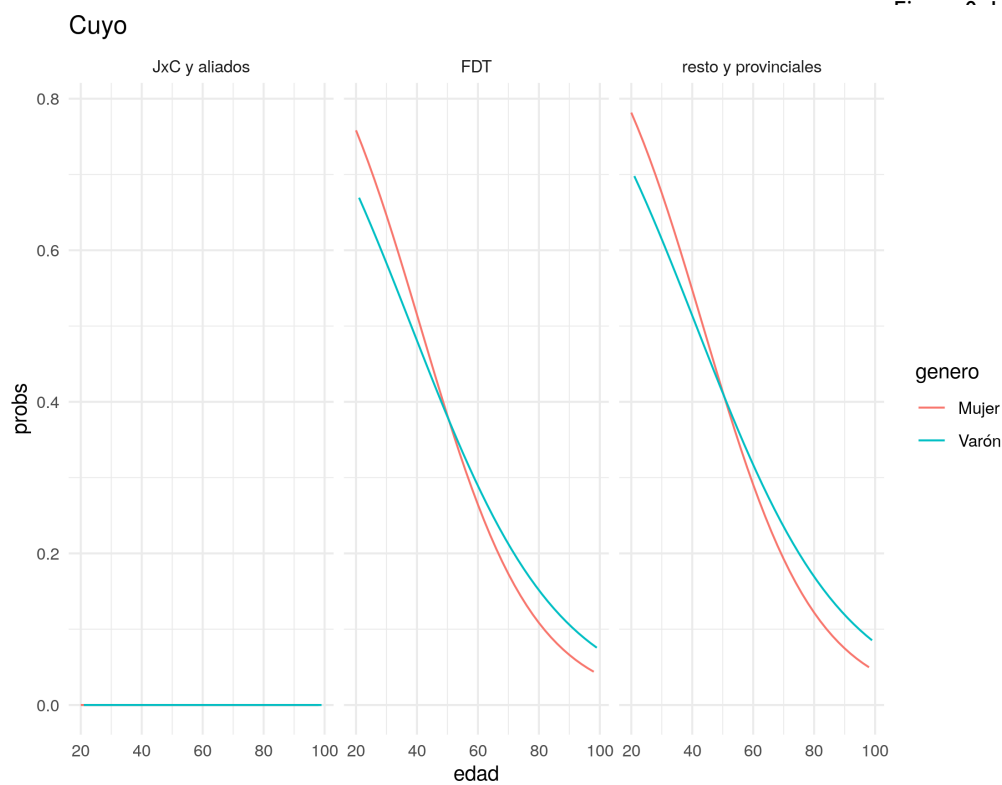
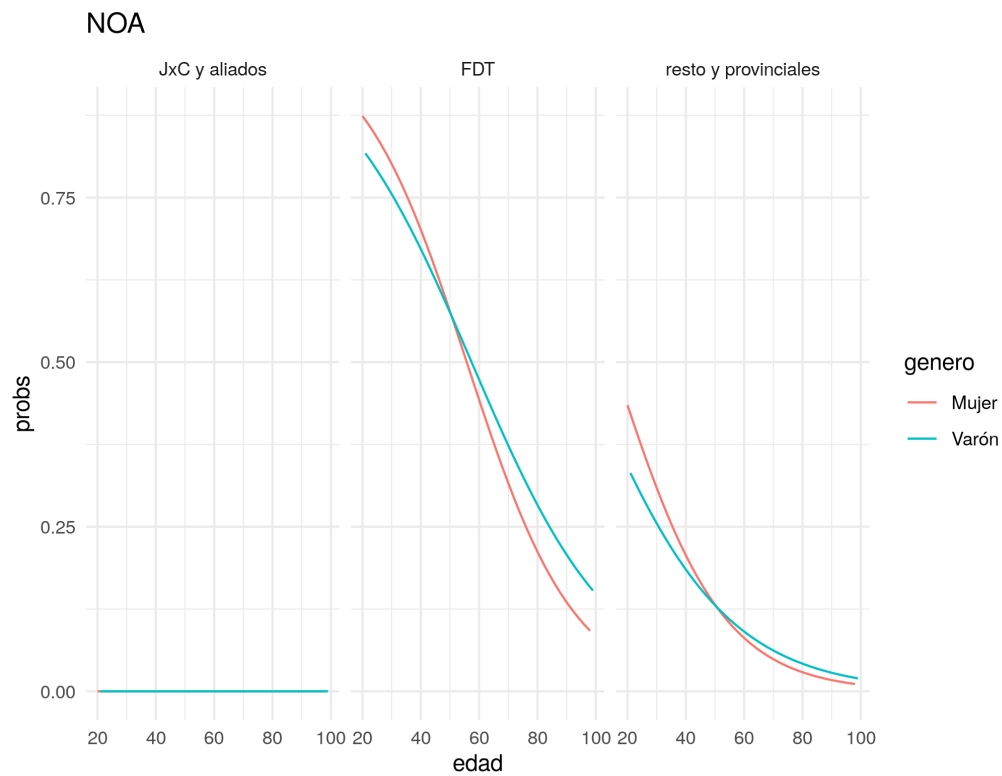


Figura 10: Interacciones entre variables para explicar el voto de los diputados de la región NOA

Figura 10: Interacciones entre variables para explicar el voto de los diputados de la región Patagonia

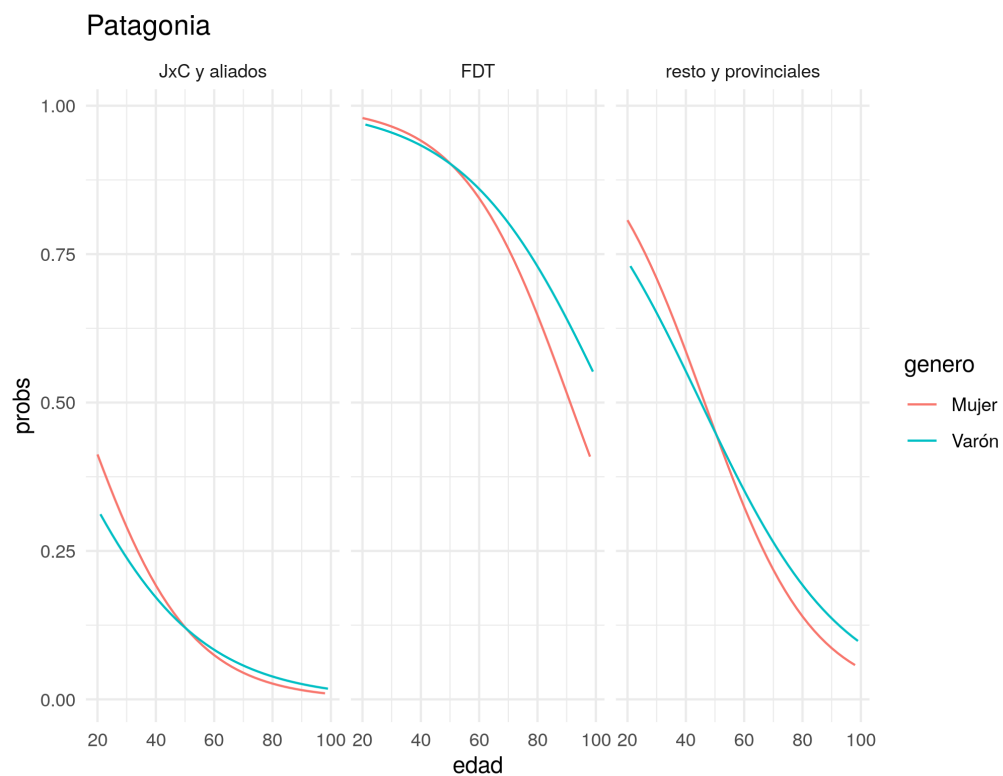


Figura 11: Interacciones entre variables para explicar el voto de les diputades de la región Patagonia