
¿Qué se aprende en el Magíster en Bioestadística?

JOSÉ ANTONIO RUIZ-TAGLE MATURANA^{1,2}

Desde el inicio del proceso de tesis, pasando por la graduación hasta hoy, me he dado el tiempo de reflexionar sobre qué aprendí en el programa. En este ensayo les contaré mi experiencia, la cual transitó por un camino muy pedregoso y con más bajos que altos. Quizás pueda serle útil a alguna persona que esté pasando por lo mismo.

Ingresé al Magíster inmediatamente después de terminar la carrera de Ciencia Política. Estoy consciente de que es una combinación extraña, sin embargo, durante el pregrado me acerqué a los “métodos cuantitativos” más que a cualquier otra asignatura. Las clases consistían en sentarnos frente a un computador con SPSS a analizar datos. Aprendí muchísimas cosas, pero me di cuenta de que no entendía lo que estaba haciendo, era muy hábil maniobrando el computador para tener el resultado, pero llegué a un momento en donde tenía tantas dudas, y a la vez tan pocas herramientas para contestarlas, que me vi en la necesidad de entrar al Magíster en Bioestadística. Además, mis intereses chocaban con los fenómenos que comúnmente se investigaban en la Ciencia Política, soy un renegado al igual que muchas personas. El primer año fue muy duro, y temía a causa de mi poco entendimiento matemático. Logré pasar el curso “Bases matemáticas para la Estadística” con un 4, teniendo que invertir en clases particulares para poder aprobarlo. En los cursos aplicados me iba relativamente bien, pero los matemáticos tuvieron una complejidad imposible de abordar desde mi posición. Me ausenté durante muchas clases porque estaba deprimido, sentía que asistir a clases era perder el tiempo. Sin embargo, nunca me quedaba de brazos cruzados, estudiaba de fuentes que me resultarían más inteligibles, trabajaba con bases de datos, simulaba comprobaciones de la teoría, etc. A pesar de que las matemáticas son una herramienta muy necesaria, personalmente no me conformaba con saber la demostración de un teorema, necesitaba entender qué significaba y por qué era tan importante. De alguna manera, percibí que los profesores estaban conscientes de eso y me tendieron la mano cuando lo necesité. Mis compañeros también me ayudaban mucho, y estaré eternamente agradecido por eso. Afortunadamente, no me ha tocado encontrar ningún estimador máximo verosímil o alguna función generadora de momentos.

Durante el programa fallé en todo lo que pude fallar: reprobé los hitos de integración 1 y 2, cambié mi

proyecto de tesis y tuve que alargar un semestre para poder terminar mi investigación. Las ganas de abandonar nunca me faltaron, sin embargo, el incentivo más fuerte que tenía para no rendirme era que de mi generación empezamos nueve y solo quedábamos cuatro. En el último semestre que me quedaba para terminar mi tesis, empezaron las clases de Doctorado en Políticas Públicas, las cuales eran en modalidad telemática. A partir de ese momento, comencé a reflexionar sobre todo lo que había aprendido en el Magíster en Bioestadística.

Resulta interesante pensar que, a pesar de que hay muchas maneras de aproximarse a un fenómeno, no puede quedar fuera de la ecuación el “¿qué me gustaría aprender?”. Particularmente, el programa de Magíster en Bioestadística no se trata sólo de abordar un fenómeno utilizando ciertos métodos, es estar al tanto de que existe una gama de posibilidades que podrían ser viables, por lo que debemos justificar cuidadosamente nuestras elecciones. A raíz de esto, se vuelve necesario darse un tiempo para pensar en cómo haríamos el estudio si tuviésemos recursos ilimitados. Por ejemplo, mi tesis buscaba estudiar la relación entre el consumo de sustancias con las conductas sexuales de riesgo. Utilicé como indicador el número de parejas sexuales en el último año. Para hacerlo utilicé datos del Instituto Nacional de la Juventud (INJUV), y apliqué un modelo de regresión binomial negativo generalizado (NB-P). El artículo resultante de este proceso aún se encuentra en proceso de edición. El diseño de muestreo de la encuesta no era óptimo para cumplir con los objetivos de mi investigación, porque correspondía a una muestra especialmente diseñada para estimar a través de intervalos de confianza. El problema es que ese tipo de diseños muestrales no considera dos elementos esenciales para obtener los anhelados p valores: potencia estadística y tamaño del efecto. Por otro lado, las encuestas rara vez preguntan lo que uno exactamente necesita, y nos tendremos que conformar con acercarnos lo más posible al fenómeno con la información disponible. Evidentemente, tener recursos hace posible recolectar nuestra propia muestra, donde las estimaciones tendrán una precisión deseada y establecida *a priori*, las mediciones serán más apegadas al fenómeno de interés y podríamos evitar algunos sesgos. Al estar consciente de esto, no pude evitar preguntarme por qué había tantos estudios que hacían lo mismo que yo. Esto me resultaba preocupante porque parecía que los

¹Magíster en Bioestadística (Universidad de Chile).

²Candidato a Doctor en Políticas Públicas (Universidad Mayor).

artículos científicos no respetaban al pie de la letra lo que dictaminaba la teoría estadística. A raíz de esto, pude concluir dos cosas: 1) es mejor estar consciente del error y no ignorarlo, y 2) mientras más cosas sean controladas en el diseño, más simples son los métodos para el análisis.

Elegí el modelo de regresión binomial negativa porque la primera vez que vi una variable que tenía una desviación estándar mayor a la media me llamó profundamente la atención. En ese momento no sabía que uno de los supuestos de la distribución de Poisson era la igualdad entre la media y la varianza. De alguna manera, los datos me estaban diciendo que había un comportamiento heterogéneo en torno a esa variable. Me pareció interesante, yo pensaba que era un problema, y que simplemente se debía a personas mal intencionadas que buscaban sesgar los resultados. Después entendí que no es sencillo esclarecer las causas de esta heterogeneidad, y que frente a este escenario, es mejor considerarlo una característica. Para poder aplicar el método, tuve que dedicar muchas horas de estudio a los modelos de tipo conteo, teniendo que responder preguntas como ¿por qué quiero usar una binomial negativa y no una Poisson?, ¿cómo puedo distinguirlas? Yo sabía por tacto y experiencia que la variable dependiente seguía esa distribución, pero debía demostrarlo con evidencia. Cuando llegué a la binomial negativa, debía buscar la parametrización correcta, y esto no fue tarea fácil. Particularmente, la regresión binomial negativa tiene más de 20 parametrizaciones diferentes, y cada una sirve para algo específico. Si no hubiera investigado esto, podría haber corrido el comando `nbreg` en Stata y se acababa el problema. No obstante, ese comando modela por defecto una parametrización NB2, es decir, la varianza es una función cuadrática de la media. Al hacer eso, me di cuenta de que la sobredispersión no se solucionaba, por ende, tuve que estimar el exponente utilizando una NB-P. Todo esto mientras me preguntaba si las investigaciones que utilizaban este tipo de métodos hicieron lo mismo que yo, ¿estarán seguros de que usaron la parametrización correcta?

Aprender Estadística se asemeja a introducirse en el arte de la magia. Tiene la virtud de impresionar y dar un gran espectáculo, pero la gran desventaja es que

mientras más trucos se aprenden, más se pierde la capacidad de asombro. Después de pasar por el programa, veía errores en todos lados, ¿por qué hace una regresión lineal con esos datos?, ¿por qué está hablando de que la muestra es representativa si quiere hacer comparaciones?, ¿para qué ponen la media y la desviación estándar de una variable binaria en la tabla de descriptivos? En contraste, también me he preguntado ¿por qué le añaden tanta complejidad a algo que puede ser tan simple? Ahora que soy candidato a Doctor en Políticas Públicas y Magíster en Bioestadística, sigo sin entender fórmulas con integrales dobles y sigo sin saber cuándo es el mejor momento para añadir una constante a ambos lados de la ecuación. Sin embargo, entiendo que todas las investigaciones tienen fortalezas y debilidades, y soy capaz de reconocer la diferencia entre error y limitación.

La solución a todos nuestros problemas está en hacer un buen diseño, idealmente desde la extracción de la muestra. Si esto no es posible, debemos estar alerta de todos los errores no forzados que tuvimos que cometer en el trayecto. Porque finalmente los métodos estadísticos deberán ser confecciones de sastrería para llevar a cabo nuestro diseño de investigación. Estoy consciente de que puede ser difícil considerando el arduo proceso que conlleva conseguir una publicación. Desde hace un tiempo que trabajo haciendo asesorías estadísticas. Mi labor consiste en ayudar a que otras personas se gradúen o publiquen. Perdí la cuenta de las veces que me han pedido analizar datos para ver si encuentro algo interesante, ojalá un p valor que muestre significancia. Ser Magíster en Bioestadística me permite tomarme la molestia de investigar métodos que sean adecuados para dar cuenta de los posibles errores que estoy cometiendo, haciendo de esto un proceso reflexivo y no automático.

Información de la tesis:

Título: “Relación entre el consumo de alcohol y otras drogas con el número de parejas sexuales en el último año en jóvenes residentes en Chile de 15 a 29 años”

Año: 2020

Autor: José Antonio Ruiz-Tagle Maturana

Profesor guía: Dr. Dante Cáceres Lillo, Programa de Salud Ambiental, Escuela de Salud Pública