

Estudio de la duración de la estancia hospitalaria de pacientes con cáncer de mama en España

Sara Barbeito Herrero

Curso 2023-2024

Por motivos de confidencialidad no es posible la publicación de la memoria completa del Trabajo Fin de Máster titulado “Estudio de la duración de la estancia hospitalaria de pacientes con cáncer de mama en España”, por lo que en el presente documento se incluye un resumen del mismo.


Resumen

El cáncer de mama es una de las enfermedades oncológicas más comunes y representa una preocupación significativa para la salud pública a nivel mundial. En España concretamente, según la Sociedad Española de Oncología Médica (SEOM) es el tipo de cáncer más diagnosticado en mujeres por delante del cáncer colorrectal, de pulmón, cuerpo uterino, tiroides y páncreas, con una incidencia estimada de 132 casos por cada 100000 habitantes en España. Esta alta incidencia subraya, entre otras cosas, la necesidad de una buena la gestión hospitalaria de los/as pacientes que sufren esta

enfermedad.

La duración de la estancia hospitalaria es un indicador clave tanto de la eficiencia del sistema de salud como de la calidad de la atención proporcionada. Además, puede tener implicaciones importantes en la planificación y asignación de recursos hospitalarios. Este Trabajo de Fin de Máster se centra en analizar la duración de la estancia hospitalaria de pacientes diagnosticados/as de cáncer de mama en hospitales públicos de España entre los años 2016 y 2021. Concretamente, se llevará a cabo un estudio de regresión, ajustando diferentes modelos, con el fin de analizar qué características de los/as pacientes determinan más, y de qué manera, el tiempo de ingreso, así como de realizar predicciones para anticiparlo.

Inicialmente, se nos ha proporcionado una base de datos que contenía información de pacientes con todo tipo de diagnósticos, por lo que hemos filtrado estos datos, quedándonos únicamente con los/as pacientes de cáncer de mama. Estos son casi doscientos mil, exactamente, 196123, que se miden en las siguientes variables: sexo, edad, provincia de hospitalización, provincia de residencia, fecha de alta hospitalaria, días de estancia, motivo del alta (pudiendo ser curación, traslado, muerte u otro), y diagnóstico de entrada (ordinario o urgente).

En el Capítulo 1 se presenta primero la base de datos, indicando brevemente cómo se ha procesado en  para su posterior análisis. Después de esto, se realiza un análisis descriptivo de los datos, con el que inicialmente se busca estudiar cada variable separadamente mediante resúmenes, diagramas de barras, e histogramas entre otros métodos según sea apropiado. Se emplean herramientas como mapas cloropléticos de España para estudiar la incidencia estimada por provincias de los casos de cáncer a partir de nuestros

datos.

También se intentan buscar relaciones entre algunos pares de variables, por ejemplo entre edad y días de estancia, relación para la cual a través de una nube de puntos se identifica ya de manera flagrante cómo los días de estancia aumentan al aumentar la edad de los pacientes. Finalmente se trata de responder a algunas preguntas importantes para la gestión hospitalaria a través de análisis estadísticos sencillos como diagramas de caja y bigotes, gráficos de violín, o tablas de contingencia, sin profundizar en conceptos complejos. Por ejemplo, se investiga si el hecho de hospitalizarse fuera de la provincia de residencia influye en los días de estancia, o si la entrada por diagnóstico urgente aumenta la probabilidad de alta por fallecimiento.

En el Capítulo 2 se expone el fundamento teórico de los modelos de regresión que van a ser aplicados a los datos para llevar a cabo el objetivo principal del presente trabajo: estudiar y predecir los días de estancia hospitalaria a partir de algunas de las otras variables disponibles. En concreto, es necesario en primer lugar un modelo de regresión múltiple, donde la variable respuesta es un conteo y, por tanto, debe ser modelada a través de una familia de Poisson. No obstante, se recalca la necesidad de un modelo Binomial Negativo ante la muy posible presencia de sobre-dispersión en el caso de la práctica.


Además, para dar más flexibilidad al modelo que ajustaremos, se introducen los modelos aditivos. Por otro lado, los datos están agrupados en provincias, lo que sugiere una posible relación de dependencia entre aquellos de una misma provincia. Esto plantea el problema de que el número de días de estancia hospitalaria podría ser diferente para individuos con las mismas características pero de provincias distintas, lo cual sería injusto y destacaría

la desigualdad en el sistema de salud entre ciertas zonas, y subrayaría la necesidad de mejoras y ayudas a la gestión hospitalaria en las más desfavorecidas. La presencia de datos agrupados en provincias sugiere entonces la necesidad de un modelo de regresión multinivel.

A continuación, en el Capítulo 3 se estudian los días de estancia hospitalaria a través de varios modelos de regresión mixtos aplicados a la base de datos disponible, siguiendo la línea teórica expuesta en el Capítulo 2. Inicialmente se contrasta el efecto grupo, llegando a la conclusión de que efectivamente existe el efecto provincia sobre los días de estancia, lo que justifica la necesidad de un modelo mixto. Luego, se ajusta un modelo de Poisson mixto sobre los datos, pero mediante un test se diagnostica sobre-dispersión en ellos, concluyendo que este modelo no es adecuado. Se procede entonces con el ajuste de un modelo Binomial Negativo Mixto, que presenta mejores resultados en cuanto a la diagnosis, pero todavía no es el más adecuado. Finalmente, se ajusta entonces un modelo Binomial Negativo Mixto Aditivo, con suavización *spline* sobre las variables edad del/la paciente y año de hospitalización.

Con este último modelo se observa que estas variables presentan en efecto una relación no lineal con la variable respuesta, que se representa e interpreta en cada caso. Dado que este último resulta ser el modelo más adecuado para los datos y el que más información aporta, es el que se emplea para predecir los días de estancia hospitalaria dada una submuestra aleatoria de pacientes de diferentes provincias de nuestra base de datos. También se representa el efecto estimado de la variable edad sobre los días de estancia distinguiendo entre provincias, es decir, dibujando una curva para algunas provincias diferentes, que permite comprobar el efecto del grupo sobre la relación entre estas dos variables. Finalmente, se resumen los resultados generales obtenidos, y

se exponen las conclusiones principales de este trabajo en el Capítulo 4.

Finalmente, nótese que todo el código de  que ha sido usado para llevar a cabo los diferentes análisis estadísticos desarrollados se ha incluido en un Anexo del trabajo.