

Censura

En el sentido más conocido de la palabra, censura es la acción de examinar una obra destinada al público, suprimiendo o modificando la parte que no se ajusta a determinados planteamientos políticos, morales o religiosos, para determinar si se puede o no publicar o exhibir. Entonces, ¿qué queremos decir en estadística cuando hablamos de datos censurados?. Nada que ver con la política, moral ni la religión. Para explicar lo que es un **dato censurado** tendremos que hablar primero de las **variables de tiempo a suceso** y de los análisis de supervivencia.

De manera general, podemos decir que hay tres tipos de variables: cuantitativas, cualitativas y de tiempo a suceso. Las dos primeras se entienden bastante bien en general, pero las de tiempo a suceso son un poco más complicadas de entender.

Imaginemos que queremos estudiar la mortalidad de esa terrible enfermedad que es la fildulastrosis. Podríamos contar el número de fallecidos al final del periodo del estudio y dividirlos por la población total al inicio. Por ejemplo, si al inicio hay 50 enfermos y se nos mueren cuatro durante el seguimiento, podríamos calcular la mortalidad como $4/50 = 0,08$, o sea del 8%. Así, si hemos seguido a la población durante cinco años, podremos decir que la supervivencia de la enfermedad a los cinco años es del 92% ($100-8 = 92$).

Sencillo, ¿verdad? El problema es que esto solo es válido cuando todos los sujetos tienen el mismo periodo de seguimiento y no se producen pérdidas o abandonos a lo largo del estudio, situación que suele estar lejos de la realidad en la mayor parte de los casos.

En estos casos, lo correcto es medir no solo si se produce el fallecimiento (que sería una variable dicotómica), sino también cuándo se produce, teniendo en cuenta además el diferente periodo de seguimiento y las pérdidas. Así, utilizaríamos una variable de tiempo a suceso, que está compuesta por una variable dicotómica (el suceso que se mide) y una continua (el tiempo de seguimiento cuando se produce).

Siguiendo el ejemplo anterior, los participantes en el estudio podrían clasificarse en tres tipos: aquéllos que fallecen durante el seguimiento, los que permaneces vivos al final del estudio y los que se pierden durante el seguimiento.

De los que se mueren podemos calcular su supervivencia pero, ¿cuál es la supervivencia de los que están vivos al final del estudio? ¿Y cuál es la supervivencia de los que se pierden durante el seguimiento? Está claro que algunos de los perdidos pueden haber fallecido al final del estudio sin que

nosotros lo detectemos, por lo que nuestra medida de la mortalidad no será exacta.

Y aquí es donde nos encontramos con los datos censurados. Todos aquellos que no presentan el evento durante un estudio de supervivencia se denominan censurados (las pérdidas y los que acaban el estudio sin presentar el evento). La importancia de estos datos censurados es que hay que tenerlos en cuenta al hacer el estudio de supervivencia, tal como veremos a continuación.

La metodología a seguir es confeccionar una [tabla de supervivencia](#) que tenga en cuenta los sucesos (en este caso las muertes) y los datos censurados, tal como vemos en la tabla adjunta.

Las columnas de la tabla representan lo siguiente: x , el número de año del seguimiento; N_x , el número de participantes vivos al inicio de ese año; C_x , el número de pérdidas de ese año (censurados); M_x , el número de fallecidos durante ese periodo; PM , probabilidad de morir en ese periodo; PSP , la probabilidad de sobrevivir en ese periodo (la probabilidad de no presentar el evento); y PSG , la probabilidad de supervivencia hasta ese momento.

X	N_x	C_x	M_x	PM	PSP	PSG
0	50	0	1	0,02	0,98	0,98
1	49	0	0	0	1	0,98
2	49	2	1	0,0204	0,9796	0,96
3	46	5	2	0,0434	0,9566	0,9183
4	39	2	0	0,0512	1	0,9183

Como vemos, el primer año partimos de 50 participantes, de los cuales uno fallece. La probabilidad de fallecer en ese periodo es de $1/50 = 0,02$, con lo que la probabilidad de supervivencia en el periodo (que es igual a la global por ser el primer periodo) es de $1-0,02 = 0,98$.

En el segundo periodo partimos de 49 y no fallece ni se pierde nadie. La PM en el periodo es cero y la de supervivencia uno. Así, la probabilidad global será de $1 \times 0,98 = 0,98$.

En el tercer periodo seguimos con 49. Se pierden dos y fallece uno. La PM es de $1/49 = 0,0204$ y la PSP de $1-0,0204 = 0,9796$. Si multiplicamos la PSP por la global del periodo anterior, obtenemos la supervivencia global

de este periodo: $0,9796 \times 0,98 = 0,96$.

En el cuarto periodo partimos de 46 participantes, produciéndose cinco pérdidas y dos fallecimientos. La PM será de $2/46 = 0,0434$, la PSP de $1 - 0,0434 = 0,9566$ y la PSG de $0,9566 \times 0,96 = 0,9183$.

Por último, en el quinto periodo partimos de 39 participantes. Tenemos dos censurados y ningún evento (fallecimiento). PM es cero, PSP es igual a uno (no se muere nadie en este periodo) y PSG $1 \times 0,9183 = 0,9183$.

Finalmente, teniendo en cuenta los datos censurados, podemos decir que la supervivencia global de la fildulastrosis es del 91,83% a los cinco años.

Y con esto vamos a dejarlo por hoy. Hemos visto cómo se construye una tabla de supervivencia con datos censurados para tener en cuenta el seguimiento desigual de los participantes y las pérdidas durante el seguimiento.

Solo dos reflexiones antes de terminar. En primer lugar, aunque se hable de análisis de supervivencia, el evento no tiene porqué ser el fallecimiento de los participantes. Puede ser cualquier evento que se produzca a lo largo del seguimiento del estudio.

En segundo lugar, las variables de tiempo a suceso y los datos censurados son la base para realizar otras técnicas estadísticas que estiman la probabilidad de producirse el evento en estudio en un momento determinado, como los modelos de regresión de Cox. Pero esa es otra historia...
