

La liebre y la tortuga

Todos conocemos la fábula de la liebre y la tortuga. Resulta que, por alguna razón que no alcanzo a comprender, a alguien se le ocurre hacer una carrera con dos participantes: una tortuga y una liebre. Como es lógico, la gran favorita de la carrera es la liebre, infinitamente más rápida que la tortuga. Pero resulta que la liebre se confía en exceso y se echa a dormir, de forma que cuando quiere darse cuenta ya no puede recuperar la ventaja de la tortuga y pierde la carrera contra todo pronóstico. Moraleja: nunca menosprecies a los demás y no te duerma en los laureles, no sea que te pase como a la liebre.

En ocasiones, podemos pensar en el ensayo clínico como una carrera entre los participantes. Esto es así cuando la variable principal de resultado es una variable de tiempo a suceso. Estas variables miden cuántos participantes presentan el suceso en cuestión y, lo que es más importante, el tiempo que tardan en presentarlo. Las variables de tiempo a suceso se llaman también variables de supervivencia, aunque no tienen por qué estar relacionadas con la mortalidad.

Veamos un ejemplo. Supongamos que queremos saber la eficacia de un fármaco sobre el control de la presión arterial. Le damos el fármaco a los del grupo de intervención y un placebo a los del grupo control para ver cuántos se controlan mejor y en cuánto tiempo lo hacen.

Una posibilidad sería utilizar riesgos relativos. Dividimos la proporción de pacientes que se controlan en el grupo de intervención por la proporción que se controla en el de control y obtenemos nuestro riesgo relativo. El problema es que obtendremos información de cuántos más mejoran en un grupo que en otro, pero no nos dirá nada del aspecto temporal. No sabemos si lo hacen pronto o no.

Otra posibilidad es tomar como variable de resultado dicotómica el controlar la presión arterial (sí o no) y calcular un modelo de regresión logística. Con este modelo obtendremos una odds ratio, que nos dará una información similar a la del riesgo relativo, pero que tampoco tendrá en cuenta el aspecto temporal de la aparición del suceso.

El método adecuado de analizar este problema sería establecer la medida dicotómica de control arterial, pero calculando un modelo de regresión de riesgo proporcionales o regresión de Cox. Este modelo de regresión sí tiene en cuenta el tiempo que tarda en producirse el suceso.

El modelo de regresión de Cox calcula el riesgo de presentar el suceso en expuestos a la intervención frente a no expuestos en cada momento dado. Para esto calcula cuánto más probable es que se produzca el suceso, en el

intervalo temporal siguiente, entre los sujetos que todavía no lo han presentado. Llevando esta medida al límite, si vamos acortando el intervalo de tiempo hasta que vale cero llegamos al riesgo instantáneo, que va oscilando con el tiempo, pero del que el modelo nos calcula una extrapolación media. Este índice es el llamado cociente de riesgos instantáneos, más conocido por su nombre en inglés, [hazard ratio](#) (HR).

La HR puede tener valores entre cero e infinito. El valor neutro es uno, que indica el mismo riesgo en los dos grupos. Un valor menor que uno indica menor riesgo en el grupo expuesto. Por último, un valor mayor de uno indica mayor riesgo en expuestos, tanto mayor cuanto mayor sea el valor de la HR.

La HR no es una medida de probabilidad, sino una odds, por lo que su interpretación es similar a la de la odds ratio, solo que, además, tiene en cuenta el aspecto temporal. Un error frecuente es pensar que la HR informa de la duración temporal hasta el suceso. Por ejemplo, una $HR = 2$ no significa que los expuestos desarrollan el suceso dos veces más rápido, sino que los que todavía no lo han presentado tienen el doble de probabilidad de presentarlo que los no expuesto.

Si queremos información sobre la velocidad a la que se produce el suceso podemos recurrir a otro de los índices que nos proporciona la regresión de Cox, que es la [mediana de tiempo](#) en que presentan el suceso el 50% de los participantes.

Volviendo a nuestra fábula de la carrera, la HR nos diría quién tiene más probabilidades de ganar la carrera, mientras que la mediana nos diría cuánto ventaja sacaría el ganador al perdedor.

Y aquí dejamos el asunto de la liebre, la tortuga y la regresión de riesgos proporcionales. No hemos hablado nada de la forma de representar los resultados del modelo de regresión de Cox. Para ellos se utilizan un tipo especial de gráficos denominados [curvas de supervivencia](#) o de [Kaplan-Meier](#). Pero esa es otra historia...

[Ni tanto ni tan calvos](#)

¿Os habéis preguntado alguna vez por qué la gente se queda calva, especialmente los varones a determinada edad?. Creo que tiene algo que ver con las hormonas. El caso es que es algo que suele gustar poco al afectado, y eso que hay una creencia popular que dice que los calvos son más

inteligentes. A mí me parece que no tiene nada de malo ser calvo (es mucho peor ser gilipollas), claro que yo tengo todo mi pelo en la cabeza.

Siguiendo el hilo de la calvicie, supongamos que queremos saber si el color de pelo tiene algo que ver con quedarse calvo antes o después. Montamos un ensayo absurdo en el que reunimos 50 rubios y 50 morenos para estudiar cuántos se quedan calvos y en qué momento lo hacen.

Este ejemplo nos sirve para ilustrar los diferentes tipos de variables que podemos encontrarnos en un ensayo clínico y los diferentes métodos que debemos utilizar para comparar cada una de ellas.

Algunas variables son de tipo cuantitativo continuo. Por ejemplo, el peso de los participantes, su talla, su sueldo, el número de pelos por centímetro cuadrado, etc. Otras son de tipo cualitativo, como el color de pelo. En nuestro caso lo simplificaríamos a una variable binaria: rubio o moreno. Por último, encontramos variables llamadas de tiempo a evento, que nos muestran el tiempo que tardan los participantes en sufrir el evento en estudio, en nuestro caso, la calvicie.

Pues bien, a la hora de comparar si existen diferencias entre estas variables entre los dos grupos el método que elijamos vendrá determinado por el tipo de variable que estemos considerando.

Si queremos comparar una variable continua como la edad o el peso entre calvos y peludos, o entre rubios y morenos, tendremos que utilizar la prueba de la [t de Student](#), siempre que nuestros datos se ajusten a una distribución normal. En el caso de que no sea así, la prueba no paramétrica que tendríamos que utilizar es la de [Mann-Whitney](#).

¿Y qué pasa si queremos comparar varias variables continuas a la vez?. Pues que podremos utilizar la regresión lineal múltiple para hacer las comparaciones entre variables.

En el caso de las variables cualitativas el enfoque es diferente. Para saber si existe dependencia estadísticamente significativa entre dos variables tendremos que construir la tabla de contingencia y recurrir a la [prueba de la chi-cuadrado](#) o a la [prueba exacta de Fisher](#), según la naturaleza de los datos. Ante la duda podemos hacer siempre la prueba de Fisher. Aunque implica un cálculo más complejo, esto no es problema para cualquiera de los paquetes estadísticos disponibles hoy en día.

Otra posibilidad es calcular una medida de asociación como el [riesgo relativo](#) o la [odds ratio](#) con sus correspondientes intervalos de confianza. Si los intervalos no cruzan la línea de efecto nulo (el uno), consideraremos que la asociación es estadísticamente significativa.

Pero puede ocurrir que lo que queramos comparar sean varias variables

cualitativas. En estos casos podremos utilizar un modelo de [regresión logística](#).

Por último, vamos a hablar de las variables de tiempo a evento, algo más complicadas de comparar. Si utilizamos una variable como puede ser el tiempo que tardan en quedarse calvos nuestros sujetos podemos construir una [curva de supervivencia](#) o de [Kaplan-Meier](#), que nos muestra de forma gráfica que porcentaje de sujetos queda en cada momento sin presentar el evento (o que porcentaje ya lo ha presentado, según como la leamos). Ahora bien, podemos comparar las curvas de supervivencia de rubios y morenos y ver si existen diferencias en la velocidad a la que se quedan calvos los dos grupos. Para esto utilizamos la [prueba de los rangos logarítmicos](#), más conocida por su nombre en inglés: [log rank test](#).

Este método se basa en la comparación entre las dos curvas en base a las diferencias entre los valores observados y los esperados si la supervivencia (la producción del evento en estudio, que no tiene porqué ser muerte) fuese igual en los dos grupos. Con este método podemos obtener un valor de p que nos indica si la diferencia entre las dos curvas de supervivencia es o no estadísticamente significativa, aunque no nos dice nada de la magnitud de la diferencia.

El caso de cálculo más complejo sería el supuesto de que queramos comparar más de dos variables. Para el análisis multivariado hay que servirse de un modelo de [regresión de riesgos proporcionales de Cox](#). Este modelo es más complejo que los anteriores pero, una vez más, los programas informáticos lo llevan a cabo sin la menor dificultad si les introducimos los datos adecuados.

Y vamos a dejar a los calvos tranquilos de una vez. Podríamos hablar más acerca de las variables de tiempo a evento. Las curvas de Kaplan-Meier nos dan una idea de quién va presentando el evento a lo largo del tiempo, pero no nos dicen nada del riesgo de presentarlo en cada momento. Para eso necesitamos otro indicador, que es el [cociente de riesgos instantáneos](#) o [hazard ratio](#). Pero esa es otra historia...

[Doctor, ¿es grave?](#)

Me pregunto cuántas veces habré escuchado esta pregunta o alguna de sus muchas variantes. Porque resulta que siempre estamos pensando en ensayos clínicos y en preguntas sobre diagnóstico y tratamiento, pero pensad si

algún paciente os preguntó alguna vez si el tratamiento que le proponíais estaba refrendado por un ensayo clínico aleatorizado y controlado. A mí, al menos, no me ha pasado nunca. Pero sí que a diario me preguntan qué les va a ocurrir en el futuro.

Y de aquí deriva la importancia de los [estudios sobre pronóstico](#). Tened en cuenta que no siempre se puede curar y que, por desgracia, muchas veces lo único que podemos hacer es acompañar y aliviar lo que podamos ante el anuncio de graves secuelas o de la muerte. Pero para esto es fundamental disponer de información de buena calidad sobre el futuro de la enfermedad de nuestro paciente. Esta información nos servirá también para calibrar los esfuerzos terapéuticos en cada situación en función de los riesgos y los beneficios. Y, además, los estudios sobre pronóstico sirven para comparar resultados entre servicios u hospitales diferentes. A nadie se le ocurre decir que un hospital es peor que otro porque su mortalidad es mayor sin comprobar antes que el pronóstico de sus pacientes sea semejante.

Antes de meternos con la [lectura crítica](#) de los artículos sobre pronóstico aclaremos la diferencia entre [factor de riesgo](#) y [factor pronóstico](#). El factor de riesgo es una característica del ambiente o del sujeto que favorece el desarrollo de la enfermedad, mientras que el factor pronóstico es aquél que, una vez que se produce la enfermedad, influye sobre su evolución. Factor de riesgo y factor pronóstico son cosas diferentes, aunque a veces pueden coincidir. Lo que sí comparten los dos son el mismo diseño de tipo de estudio. Lo ideal sería utilizar [ensayos clínicos](#), pero la mayor parte de las veces no podemos o no es ético aleatorizar los factores pronóstico o de riesgo, por lo que lo habitual es que se usen [estudios de cohortes](#). En los casos que precisan seguimientos muy largos o en los que el efecto que queremos medir es muy raro se pueden usar [estudios de casos y controles](#), pero siempre serán menos potentes por tener más riesgo de sesgo.

Un estudio de pronóstico nos debe informar de tres aspectos: qué resultado queremos valorar, qué probabilidad hay de que suceda y en qué periodo de tiempo esperamos que pase. Y para valorarlo, como siempre, nos asentaremos sobre nuestros [tres pilares](#): validez, importancia y aplicabilidad.

Para valorar la **VALIDEZ** tendremos primero en cuenta si cumple una serie de **criterios primarios** o de eliminación. Si la respuesta es no, tirad el artículo y mirad a ver qué chorrada nueva han escrito vuestros amigos en Facebook.

¿Está bien definida la muestra de estudio y es representativa de pacientes en un momento similar de la enfermedad?. La muestra, que se suele denominar [cohorte incipiente](#) o [cohorte de inicio](#), debe estar formada por un grupo amplio de pacientes en el mismo momento de la enfermedad, idealmente

al inicio, y que se sigue de forma prospectiva. Debe estar bien especificado el tipo de pacientes incluidos, los criterios para diagnosticarlos y el método de selección. Además, debemos comprobar **que el seguimiento haya sido lo suficientemente largo y completo** como para observar el evento que estudiamos. Cada participante debe seguirse desde el inicio hasta que sale del estudio, ya sea porque se cure, porque presenta el evento o porque el estudio se acaba. Es muy importante tener en cuenta las pérdidas durante el estudio, muy habituales en diseños con seguimiento largo. El estudio debe proporcionar las características de los pacientes perdidos y los motivos para la pérdida. Si son similares a los que no se pierden, probablemente los resultados sean válidos. Si las pérdidas son de más de un 20% se suele hacer un análisis de sensibilidad utilizando el escenario de **"el peor de los casos"**: consideramos que todas las pérdidas han tenido mal pronóstico y recalculamos los resultados para ver si se modifican, en cuyo caso quedaría invalidado el estudio.

Una vez vistos estos dos aspectos, pasamos a los **criterios secundarios** de validez interna o rigor científico.

¿Se han medido los resultados de forma objetiva y no sesgada?. Debe especificarse con claridad qué se va a medir y cómo antes de iniciar el estudio. Además, lo ideal es que la medición de los resultados se haga de forma ciega para el experimentador, que debe desconocer si el sujeto en cuestión está sometido a alguno de los factores pronósticos para evitar el **sesgo de información**.

¿Se han ajustado los resultados según todos los valores pronósticos relevantes?. Hay que tener en cuenta todas las **variables confusoras** y los factores pronósticos que puedan influir en los resultados. En el caso de que se conozcan por estudios previos pueden tenerse en cuenta los factores conocidos. En caso contrario, los autores determinarán los efectos mediante **análisis estratificado** de los datos (el método más sencillo) o mediante el **análisis multivariante** (más potente y complejo), habitualmente mediante un modelo de riesgos proporcionales o de **regresión de Cox**. Aunque no vamos a entrar ahora en los modelos de regresión, sí que hay dos cosas sencillas que podemos tener en cuenta. La primera, estos modelos necesitan de un número determinado de eventos por cada variable incluida en el modelo, así que desconfiad cuando se analicen muchas variables, sobre todo con muestras pequeñas. La segunda, las variables las decide el autor y son diferentes de un trabajo a otro, por lo que tendremos que valorar si no se ha incluido alguna que pueda ser relevante para el resultado final.

¿Se han validado los resultados en otros grupos de pacientes?. Cuando hacemos grupos de variables y empezamos a comparar unos con otros corremos el riesgo de que el azar nos juegue una mala pasada y nos muestre asociaciones que realmente no existen. Por eso, cuando se describe un

factor de riesgo en un grupo ([grupo de entrenamiento o derivación](#)), conviene replicar los resultados en un grupo independiente ([grupo de validación](#)) para estar seguros de la relación.

A continuación, debemos fijarnos en **cuáles son los resultados** para determinar su **IMPORTANCIA**. Para esto comprobaremos si se proporciona la estimación de la probabilidad de que suceda el desenlace de estudio, la precisión de esta estimación y el riesgo asociado a los factores que modifican el pronóstico.

¿Se especifica la probabilidad del suceso en un periodo de tiempo determinado? Hay varias formas de presentar el número de sucesos que se producen durante el periodo de seguimiento. La más sencilla sería dar una [tasa de incidencia](#) (sucesos/persona/unidad de tiempo) o la [frecuencia acumulada](#) en un momento dado. Otra forma es dar la [mediana de supervivencia](#), que no es más que el momento del seguimiento en el cuál el suceso se ha producido en la mitad de la cohorte (recordad que aunque hablemos de supervivencia, el suceso no tiene que ser obligatoriamente la muerte).

Para determinar la probabilidad de que se produzca el suceso en cada periodo y el ritmo al cual se va presentando pueden utilizarse [curvas de supervivencia](#) de varios tipos. Las [tablas actuariales](#) o de vida se utilizan para muestras grandes, cuando no sabemos el momento exacto del evento y con periodos de tiempo fijos. Sin embargo, probablemente nos encontremos con más frecuencia con las [curvas de Kaplan-Meier](#), que miden mejor la probabilidad del suceso para cada momento concreto con muestras más pequeñas. Con este método se pueden proporcionar los [cocientes de riesgos instantáneos](#) en cada momento (los hazard ratios) y la mediana de supervivencia, además de otros parámetros según el modelo de regresión utilizado.

Para valorar la **precisión de los resultados** buscaremos, como siempre, los [intervalos de confianza](#). Cuanto mayor sea el intervalo, menos precisa será la estimación de la probabilidad del suceso en la población general, que es lo que realmente nos interesa saber. Hay que tener en cuenta que el número de pacientes suele ser menor según pasa el tiempo, por lo que es habitual que las curvas de supervivencia sean más precisas al comienzo que al final del seguimiento. Por último, valoraremos **cuáles son los factores que modifican el pronóstico**. Lo correcto es representar todas las variables que puedan influir sobre el pronóstico con sus correspondientes [riesgos relativos](#), que serán los que nos permitan evaluar la importancia clínica de la asociación.

Por último, tendremos que considerar la **APLICABILIDAD** de los resultados. **¿Son aplicables a mis pacientes?** Buscaremos las similitudes entre los pacientes del estudio y los nuestros y evaluaremos si las diferencias que

encontremos nos permiten extrapolar los resultados a nuestra práctica. Pero además, **¿son útiles los resultados?**. El que sean aplicables no quiere decir que tengamos que ponerlos en práctica obligatoriamente, sino que tendremos que valorar cuidadosamente si nos van a ayudar a decidir qué tratamiento aplicar o a cómo informar a nuestro paciente o a sus familiares.

Como siempre, os recomiendo que uséis alguna plantilla, como las que proporciona [CASPe](#), para realizar la lectura crítica de forma sistemática y no dejar ningún aspecto importante sin valorar.

Ya veis que los trabajos sobre pronóstico tienen mucha miga. Y eso que no hemos comentado prácticamente sobre modelos de regresión y curvas de supervivencia, que muchas veces son el núcleo del estudio estadístico de este tipo de trabajos. Pero esa es otra historia...
