

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...
