

ELABORACIÓN DE UNA FÓRMULA PARA DETERMINAR EL RIESGO DE PACIENTES CON CORONAVIRUS

Álvaro de la Fuente, Alfonso Gordon,
Pablo Rodríguez y Daniel Anchuela

Abstract

En este documento explicaremos los fundamentos de los que partimos para construir la fórmula que utilizaremos en la determinación la puntuación del riesgo de los pacientes en nuestra aplicación Covidist, así como los pasos que utilizamos en su desarrollo haciendo uso de todos los datos estadísticos obtenidos y calculados por la aplicación.

Para obtener un modelado matemático lo más realista posible, utilizaremos los porcentajes de los datos utilizados, usaremos ponderaciones a la hora de asignar a un paciente a su caso, elevaremos la puntuación del dato a diferentes números según el caso para premiar más la puntuación que así lo requiera, y lo multiplicaremos por el valor de importancia que tenga en ese momento el tipo de dato que estamos analizando dependiendo de si las estadísticas provienen de los pacientes ingresados o de los pacientes que ya han muerto, que en este caso, tendrán más peso que los primeros.

Explicación

Los elementos fundamentales que necesita nuestra aplicación para devolver la puntuación de riesgo que tiene un paciente en un momento determinado son todos los datos estadísticos de los que dispone la base de datos en ese momento.

Dicho esto, empezaremos a modelar nuestra fórmula partiendo de los porcentajes de cada característica, previamente recibidos por parte de la clase que calcula todas las estadísticas.

Tomaremos el ejemplo más simple para empezar a modelar la formula. En este caso cogeremos el sexo, ya que solo tiene dos posibles casos: masculino y femenino. Una posible idea podría ser fijarnos en qué porcentajes son más altos para darles una mayor puntuación, pero esta idea no se correspondería a la realidad en algunos casos. Por ejemplo, en este caso la aplicación le daría más puntos a un 50% de hombres ingresados que a un 40% de pacientes que pudieran sufrir una misma enfermedad. Esto no tendría sentido, así que una posible idea sería fijarnos en la diferencia de porcentaje que hay entre los casos de un mismo dato. De esta manera obtendremos la diferencia entre el caso que mayor porcentaje tiene y el que menos para valorar esa diferencia a la hora de calcular una puntuación. De esta manera, obtenemos:

$$P_{max_i} - P_{min_i}$$

Donde P_{max} determina el porcentaje del caso más común y P_{min} determina el porcentaje del caso menos frecuente de entre todos los casos de la característica que estamos calculando (representada por el subíndice i).

Una vez tenemos la diferencia entre los porcentajes más y menos frecuentes, necesitamos asociar esa diferencia al tipo de paciente que estamos tratando. La manera que hemos razonado sería multiplicando esa diferencia que solo nos da información sobre los porcentajes de ese dato, por el porcentaje en tanto por uno del caso al que pertenece el paciente en cuestión, obteniendo:

$$(P_{max_i} - P_{min_i}) \cdot P_{p,i}$$

Donde $P_{p,i}$ es el porcentaje en tanto por uno al cual pertenece el paciente p en el dato en cuestión i .

En este momento nos encontramos con un problema: con esta fórmula le estamos dando una puntuación a un paciente, basándonos en los porcentajes que tiene ese dato. De esta manera, estamos obviando el hecho de que algunos datos tengan más importancia que otros dependiendo del número de casos que tiene. Volviendo al ejemplo

anterior con el dato *sexo*, puede ser importante valorar que por ejemplo haya un 70% de hombres ingresados con respecto a un 30% de mujeres, pero consideramos que el hecho de que sea posible que un 35% de todos los pacientes ingresados por coronavirus tenga cáncer, por ejemplo, con toda la variedad de enfermedades que podrían tener, es algo que, aunque la diferencia de porcentajes no sea tan grande, hay que valorar. Por eso, una buena manera de valorarlo sería tomando la puntuación ya explicada y elevándola al número de casos posibles que tiene ese dato. De todas formas, haciendo algunas pruebas nos dimos cuenta de que, haciéndolo de esta manera, la puntuación obtenida en datos que excedían los 4 casos era desproporcionada. De esta manera, para poder aplicar este requerimiento a nuestra fórmula manteniendo un rigor matemático, decidimos elevar la puntuación a “1.” seguido del número de casos del dato en cuestión. Por lo que quedaría así:

$$[(P_{max_i} - P_{min_i}) \cdot P_{p,i}]^{(1 + \frac{Nopt_i}{10})}$$

Donde $Nopt_i$ es el número de casos posibles que tiene el dato i .

Llegados a este punto, tenemos una fórmula que tiene en cuenta la diferencia de porcentajes, asocia esa diferencia al caso al que pertenece el paciente en cuestión y valora todo el abanico de posibilidades de cada dato para darle una mayor puntuación a aquellos datos que poseen más casos ya que el porcentaje está más repartido. Pero de momento no estamos teniendo en cuenta que nuestros datos pueden ser de dos tipos:

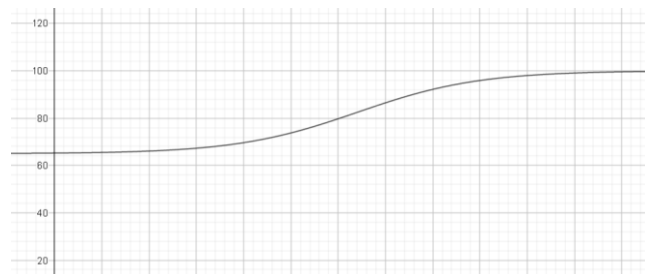
1. Datos provenientes de pacientes que ya han muerto.
2. Datos provenientes de pacientes que han sido contagiados de coronavirus.

Como lo que queremos es salvar vidas, vamos a darle un mayor peso a los datos provenientes de pacientes muertos. Sin embargo, este peso tiene que ser variable en el tiempo porque tiene que depender del número de pacientes muertos que tenemos en los últimos días. Si, por ejemplo, los datos dicen

que, en los últimos días, el número de pacientes de coronavirus que han muerto sube exponencialmente, tendremos que darle más peso del que tiene normalmente. Por eso mismo, para obtener la fórmula que determine la importancia que tienen los datos de los pacientes muertos, recibiremos los datos de estos pacientes por parte de la clase que calcula las estadísticas, al igual que lo hicimos anteriormente.

En este caso, recibiremos la media de la tasa de incremento o descenso de muertos de los últimos 7 días. Esto lo hacemos para prevenir la posibilidad de fuertes subidas y bajadas de muertes. No podemos fijarnos solo en la tasa de incremento del día anterior porque puede haber subido muy poco, y estaríamos ignorando el hecho de que el día anterior a ese podría haber subido mucho. De esta manera, partiendo de este último dato pasamos a modelar la fórmula de la importancia de los datos de los pacientes muertos.

En primer lugar, estableceremos un límite de importancia que en este caso estará en 65% y 35%. Lo que estamos haciendo en este momento es decir que, aunque haya un número de muertes muy bajo, queremos que como mínimo, la importancia de la información recibida por parte de los datos de los pacientes muertos sea del 65% frente al otro 35% que sería la importancia que recibirían los datos de los pacientes ingresados, siempre dependiendo de los anteriores. Además, vamos a determinar un límite del 200% de crecimiento de muertes, el cual, si se superase, la importancia de los datos de los pacientes muertos tendería a 100%. Dicho esto, la gráfica de la fórmula que obtenemos sería la siguiente:



De esta manera, la función que presenta esta gráfica y que depende de nuestra variable α_d que

representa la media de la tasa de incremento de los últimos 7 días sería:

$$f(a_d) = \frac{35}{1 + e^{-0.039 \cdot a_d + 5}} + 65$$

Aunque para nuestro caso lo dividiremos todo entre 100 porque es como lo hemos estado haciendo hasta el momento.

Por eso mismo, la fórmula que determinaría el valor de los datos de los pacientes ingresados sería la siguiente:

$$1 - \frac{\frac{35}{1 + e^{-0.039 \cdot a_d + 5}} + 65}{100}$$

Por último, como hemos explicado antes, lo juntaremos todo en la misma fórmula para dar el resultado final de la puntuación obtenida de un único dato, dependiendo de si este pertenece a los pacientes ingresados o a los que ya han muerto. Además, para obtener la puntuación total de un paciente teniendo en cuenta la puntuación de todos los datos obtenidos utilizaremos un sumatorio que irá desde el primer dato hasta n que representa el número total de datos de los que disponemos.

De esta manera, nuestra formula final quedaría así:

Score of a patient p in a day d

$$score(p, d) = \sum_{i=1}^n \left\{ [(Pmax_i - Pmin_i) \cdot Pp_{p,i}]^{\left(1 + \frac{Nopt_i}{10}\right)} \cdot \left[\underbrace{\frac{35}{1 + e^{-0.039 \cdot a_d + 5}} + 65}_{\text{if characteristic of dead people}} \cdot \underbrace{\left(1 - \frac{35}{1 + e^{-0.039 \cdot a_d + 5}} + 65\right)}_{\text{if characteristic of infected people}} \right] \right\}$$