TESTES POSITIVOS no COVID?

O que nos diz o Teorema de Bayes

Depois de ter chegado com a SDT (Signal Detection Theory) a conclusões surpreendentes (??) sobre o possível impacto dos falsos positivos na metodologia de testes, adoptada para o Covid 19, podendo criar uma enorme desconfiança nos valores de novos infectados com que somos bombardeados diariamente, amigos da estatística mencionaram-me a necessidade de recorrer nestes casos ao Teorema de Bayes.

Arregacei as mangas e fui estudar. E as conclusões parecem confirmar os resultados surpreendentes que obtivera com a SDT.

O teorema de Bayes

O teorema de Bayes permite-nos analisar o comportamento probabilístico de alguma coisa com base em informações que conhecemos de outras coisas que condicionam o evento a analisar

O Teorema de Bayes é realmente um corolário da lei da probabilidade total, expresso matematicamente na forma da seguinte equação:

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)},$$

A e B são eventos a que está associada um lei de probabilidades.

P(A|B) quer dizer “probabilidade de acontecer A sabendo que B aconteceu”, que será diferente de P(A), se os eventos A e B não forem independentes. Por exemplo a probabilidade de alguém ter gripe sabendo que tem febre é diferente da probabilidade de ter gripe.

Por outro lado é evidente que P(B) = P(B|A) x P(A) +P(B|~A) x P(~A) onde ~A represente não A,

Ou seja, a probabilidade de B acontecer é a soma da probabilidade de B acontecer tendo ocorrido A e não tendo ocorrido A.

O teorema de Bayes e os testes Covid 19

Vejamos um exemplo para vermos onde esta introdução teórica nos conduz:

Consideremos o exemplo de um teste de Covid 19. Tomemos como referências 99% de sensibilidade e 97% de especificidade. Isto é, o teste produzirá 99% de resultados verdadeiros positivos para pessoas infectadas e 97% de resultados verdadeiros negativos para não-infectadas. (ou seja, 1% de falsos negativos e 3% de falsos positivos).

Suponhamos agora que 2 % das pessoas na comunidade estão infectadas, valor que parece ser razoável para Portugal. Tal corresponderia a ter em Portugal cerca de 200.000 pessoas infectadas.

A pergunta é “Se um indivíduo selecionado aleatoriamente nesta comunidade testar positivo, qual a probabilidade de ele estar realmente infectado? Isto é, qual a probabilidade de não se cometer um falso positivo?”. O teorema responde a esta pergunta e o resultado é talvez surpreendente, pois não é 97%, é muito menos.

Façamos o cálculo que não é difícil:

$$
\begin{aligned}
P(infectado \mid +) &= \frac{P(+ \mid infectado)P(infectado)}{P(+)} \\
&= \frac{P(+ \mid infectado)P(infectado)}{P(+ \mid infectado)P(infectado) + P(+ \mid \text{não}\, infectado)P(\text{não}\, infectado)} \\
&= \frac{0{,}99 \times 0{,}02}{0{,}99 \times 0{,}02 + 0{,}03 \times 0{,}98} = \frac{0.0198}{0{,}0492} \\
&\approx 0.40 \dashrightarrow 40\%
\end{aligned}
$$

P(+) representa a probabilidade de o teste dar positivo.

Utilizemos agora um exemplo concreto para compreender melhor e aceitar a aparente surpresa a que chegámos.

Se nas condições prescritas 1000 indivíduos forem testados, espera–se que em média 980 indivíduos não estejam infectados (assumindo a incidência de 2% da infecção na população) e que 20 indivíduos estejam infectados.

Para os 980 não-infectados, são esperados 0.03 x 980 ≈ 29 falsos positivos. Para os 20 infectados, são esperados 0.99 x 20 ≈ 20 positivos verdadeiros. Isto é, dos 49 resultados positivos, apenas 20 (ou 40 %) são genuínos.

ESPANTO! Concluímos portanto que, naquelas condições de teste, a probabilidade de não estar infectado é MAIOR do que a de estar infectado, 40% versus 60%!

É fácil concluir que a probabilidade daquela pessoa que testou positivo não estar infectada aumenta rapidamente se a espeficidade do teste baixar, se em vez de 97% tivermos valores mais baixos, e portanto mais falsos positivos, mas também aumenta se a incidência da doença na população baixar.

Vejamos um outro exemplo com um teste Covid com a mesma Sensibilidade de 99% mas com Espeficidade = 95% e suponhamos uma incidência na população de 0.5%, ou seja, para a população portuguesa, haveria cerca de 50000 infectados.

Repitamos o cálculo:

$$
\begin{aligned}
P(infectado \mid +) &= \frac{P(+ \mid infectado)P(infectado)}{P(+)} \\
&= \frac{P(+ \mid infectado)P(infectado)}{P(+ \mid infectado)P(infectado) + P(+ \mid \text{não}\, infectado)P(\text{não}\, infectado)} \\
&= \frac{0{,}99 \times 0{,}005}{0{,}99 \times 0{,}005 + 0{,}05 \times 0{,}995} = \frac{0.00495}{0{,}0547} \\
&\approx 0.09 \dashrightarrow 9\%
\end{aligned}
$$

A probabilidade do nosso individuo estar realmente infectado baixaria drasticamente para 9%.

É fácil, repetindo o método anterior, confirmar a veracidade deste valor assustador.

Se, neste caso, os mesmos 1000 indivíduos da população portuguesa fossem testados, esperar–se-ia que em média 995 indivíduos não estivessem infectados (assumindo a incidência de 0.5 % da infecção na população) e que 5 indivíduos estivessem infectados.

Para os 995 não-infectados, são esperados 0.05 x 995 ≈ 50 falsos positivos. Para os 5 infectados, são esperados 0.99 x 5 ≈ 5 positivos verdadeiros. Isto é, dos 55 resultados positivos do teste, apenas 5 (ou seja 9 %) são genuínos.

Isto ilustra a importância da probabilidade condicional e como políticas de teste podem resultar em enormes equívocos, se as probabilidades condicionais forem negligenciadas.

Estes resultados permitem mais uma vez e por outro método evidenciar a importância dos falsos positivos, cujo impacto depende enormemente da espeficidade dos testes e da incidência da infecção na população.

É fácil demonstrar que, só testando 2ª ou mesmo 3ª vez ou mais, a probabilidade de erro baixa. Realmente, se fizermos um novo teste, teríamos de utilizar mais uma vez o teorema de Bayes calculando desta vez:

P(infectado| Testou positivo a 1ª vez)

P(infectado | Testou positivo 2 vezes)

…)

Deixo este exercício para os leitores desta nota