

Testy diagnostyczne w medycynie

Keywords: kombinatoryka, prawdopodobieństwo, statystyka, prawdopodobieństwo warunkowe, częstotliwość zjawisk

Każda osoba przechodzi wiele testów diagnostycznych w ciągu swojego życia, zarówno w gabinecie lekarskim, jak i w domu. Nie ma znaczenia, czy jest to test na COVID-19, celiakię, ogólne badanie mammograficzne czy domowy test ciążowy. W każdym przypadku chcemy wiedzieć, czy wynik testu jest wiarygodny. Dla każdego takiego testu dostarczane są dwie podstawowe informacje:

- *Czułość testu*: prawdopodobieństwo, że test wykryje chorobę, gdy na nią zachoruję.
- *Specyfika testu*: prawdopodobieństwo, że test da wynik negatywny, gdy nie mam choroby.

Tak zwana *częstotliwość występowania choroby*, tj. stosunek liczby osób cierpiących na chorobę do liczby wszystkich osób w populacji, jest również ważna dla oceny.

Interpretacja wyników testu

W aptekach i drogeriach sprzedawanych jest wiele domowych testów na alergię i nietolerancję, w tym testy na COVID-19 lub testy ciążowe. Na przykład, ulotka testu na celiakię (nietolerancję glutenu) stwierdza, że czułość testu wynosi 96,3 %, a specyfika 89,7 %. Wiemy, że liczba osób z celiakią w populacji wynosi około 1 %, więc częstotliwość występowania choroby wynosi 1/100.

Będziemy przede wszystkim zainteresowani pytaniem: Jeśli wynik testu jest pozytywny, jakie jest prawdopodobieństwo, że faktycznie mam celiakię? Spróbuj najpierw odgadnąć odpowiedź. W następnym ćwiczeniu wykonamy dokładne obliczenia.

Zadanie 1. Czułość testu wynosi 96,3 %, specyfika testu wynosi 89,7 %, a częstotliwość występowania choroby wynosi 1/100. Jakie jest prawdopodobieństwo że pacjent z pozytywnym wynikiem testu na celiakię rzeczywiście cierpi na tę chorobę?

Rozwiązanie. Jest to przykład problemu dotyczącego prawdopodobieństwa warunkowego, który można łatwo rozwiązać za pomocą twierdzenia Bayesa:

Niech A będzie zdarzeniem losowym i niech $B_1, \dots, B_n$ będzie kompletnym zbiorem zdarzeń. Jeśli $P(A) > 0$, $P(B_1) > 0, \dots, P(B_n) > 0$, to

$$P(B_k|A) = \frac{P(A \cap B_k)}{P(A)} = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^n P(A|B_i)P(B_i)}, \quad k = 1, \dots, n.$$

Symbol $P(B|A)$ oznacza prawdopodobieństwo wystąpienia zdarzenia B przy założeniu, że wystąpiło zdarzenie A .

Zdefiniujmy następujące prawdopodobieństwa:

- $P(C)$ prawdopodobieństwo, że jestem pozytywny (mam chorobę),
- $P(N)$ prawdopodobieństwo, że jestem negatywny (nie mam choroby),
- $P(+|C)$ prawdopodobieństwo, że test jest pozytywny, jeśli jestem pozytywny,

Results matter!

- $P(-|C)$ prawdopodobieństwo, że test jest negatywny, jeśli jestem pozytywny,
- $P(+|N)$ prawdopodobieństwo, że test jest pozytywny, jeśli jestem negatywny,
- $P(-|N)$ prawdopodobieństwo, że test jest negatywny, jeśli jestem negatywny,
- $P(C|+)$ prawdopodobieństwo, że jestem pozytywny, jeśli wynik testu jest pozytywny.

Ponieważ częstotliwość występowania wynosi $1/100$, to

$$P(C) = 0,01 \quad P(N) = 0,99$$

a z podanych danych

$$P(+|C) = 0,963 \quad P(-|C) = 0,037 \quad P(+|N) = 0,103 \quad P(-|N) = 0,897.$$

Zgodnie z twierdzeniem Bayesa jest tak, że

$$\begin{aligned} P(C|+) &= \frac{P(+|C)P(C)}{P(+|C)P(C) + P(+|N)P(N)} = \\ &= \frac{0,963 \cdot 0,01}{0,963 \cdot 0,01 + 0,103 \cdot 0,99} = 0,086. \end{aligned}$$

Prawdopodobieństwo, że pacjent z pozytywnym wynikiem testu rzeczywiście cierpi na chorobę wynosi zatem tylko 0,086, tj. 8,6 %.

Uwaga: Jest bardzo prawdopodobne, że wielu studentów nie odgadnie tego wyniku poprawnie przed wykonaniem obliczeń. Nie jest to zaskakujące, ponieważ badania wykazały, że nawet 95% lekarzy nie potrafi poprawnie odpowiedzieć na tego typu pytania.

Zadanie 2. Jakie jest prawdopodobieństwo, że pacjent z ujemnym wynikiem testu jest naprawdę ujemny?

Rozwiązanie. Użyjemy notacji z poprzedniego ćwiczenia i twierdzenia Bayesa. Otrzymujemy

$$\begin{aligned} P(N|-) &= \frac{P(-|N)P(N)}{P(-|N)P(N) + P(-|C)P(C)} = \\ &= \frac{0,897 \cdot 0,99}{0,897 \cdot 0,99 + 0,037 \cdot 0,01} = 0,999583. \end{aligned}$$

Zatem prawdopodobieństwo, że pacjent, który uzyskał wynik negatywny, jest w rzeczywistości zdrowy, wynosi 0,999583, i.e., 99,96 %.

Interpretacja bez użycia prawdopodobieństwa

Spróbujmy ponownie rozwiązać pierwsze ćwiczenie, ale przeformułujemy zadanie, aby uniknąć koncepcji prawdopodobieństwa warunkowego.

Zadanie 3. Na 100 000 osób około 1 000 ma celiakię. Spośród tych 1000 osób test wykrywa celiakię w 960 przypadkach. Z pozostałych 99 000 osób, które nie mają celiakii, 10200 osób również uzyska pozytywny wynik testu. Wyobraźmy sobie grupę osób, które uzyskały pozytywny wynik testu. Jaki odsetek z nich rzeczywiście ma celiakię?

Results matter!

Rozwiązanie. W sumie mamy $960 + 10\,200 = 11\,160$ pozytywnych wyników testów, ale tylko 960 to prawdziwe pozytywne przypadki, więc:

$$\frac{960}{11\,160} = 0,086 \implies 8,6\%.$$

Ponownie widzimy, że odsetek osób cierpiących na tę chorobę wśród tych, którzy uzyskali pozytywny wynik testu, wynosi 8,6 %.

W tym obliczeniu uprościliśmy, pracując z liczbami całkowitymi i zaokrąglając do najbliższej dziesiątki. W ten sposób poświęciliśmy nieco precyzji na rzecz prostoty. Ponadto założyliśmy, że wszystko potoczy się tak, jak przewiduje prawdopodobieństwo, tj. że dokładnie 1000 osób zachoruje na celiakię. Nie zawsze jest to prawdą. Choć ta metoda jest prostsza i często łatwiejsza do zrozumienia, nie jest zwykle spotykana w podręcznikach. Jednak dzięki temu ćwiczeniu wskaźnik sukcesu lekarzy w oszacowaniu wyniku wynosi prawie 100%.

Literatura

1. Singh P., Arora A., Strand T.A., Leffler D.A., Catassi C., Green P.H., Kelly C.P., Ahuja V., Makharia G.K., *Global Prevalence of Celiac Disease: Systematic Review and Meta-analysis*. Clin Gastroenterol Hepatol. 2018 Jun; 16(6):823-836. doi: 10.1016/j.cgh.2017.06.037.
2. Gigerenzer, G., *Calculated risks*, Simon and Schuster, 2003.