

Diagnosticke testy v medicíne

Keywords: kombinatorika, pravdepodobnosť a štatistika, podmienená pravdepodobnosť, počet výskytov daného javu

Každý človek počas svojho života podstúpi mnoho diagnostických testov, či už u lekára alebo doma. Nezáleží na tom, či ide o test na COVID-19, celiakiu, všeobecné mamografické skriningové vyšetrenie alebo domáci tehotenský test. V každom prípade chceme vedieť, či je výsledok testu spoľahlivý. Pre každý takýto test sú poskytnuté dve základné informácie:

- *Citlivosť testu*, t. j. pravdepodobnosť, že test zistí ochorenie, keď ho mám.
- *Špecifická testu*, t. j. pravdepodobnosť, že test bude negatívny, keď ochorenie nemám. Pre hodnotenie je tiež dôležitá takzvaná prevencia ochorenia, t. j. pomer počtu jedincov, ktorí majú ochorenie, k počtu všetkých jedincov v populácii.

Interpretácia výsledkov testu

V lekárňach a drogériách sa predáva množstvo domácich testov na alergie a intolerancie, vrátane testov na COVID-19 alebo tehotenských testov. Napríklad v príbalovom letáku testu na celiakiu (intoleranciu na lepok) sa uvádza, že citlivosť testu je 96,3 % a špecifická testu je 89,7 %. Vieme, že počet celiatikov v populácii je približne 1 %, takže prevencia ochorenia je 1/100.

Našou hlavnou otázkou bude: Ak mám pozitívny test, aká je pravdepodobnosť, že skutočne mám celiakiu? Skúste najprv odhadnúť odpoveď. V nasledujúcom cvičení vykonáme presný výpočet.

Úloha 1. Citlivosť testu je 96,3 %, špecifická testu je 89,7 % a prevencia ochorenia je 1/100. Aká je pravdepodobnosť, že pacient, ktorý má pozitívny test na celiakiu, skutočne má toto ochorenie?

Riešenie. Toto je príklad problému s podmienenou pravdepodobnosťou, ktorý sa dá jednoducho vyriešiť pomocou Bayesovej vety:

Nech A je náhodný jav a nech $B_1, \dots, B_n$ je úplny systém javov. Ak $P(A) > 0$, $P(B_1) > 0, \dots, P(B_n) > 0$, potom

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

Symbolom $P(B|A)$ označuje pravdepodobnosť, že nastane udalosť B za predpokladu, že nastala udalosť A .

Definujme nasledujúce pravdepodobnosti:

- $P(C)$ pravdepodobnosť, že som pozitívny (mám ochorenie),
- $P(N)$ pravdepodobnosť, že som negatívny (nemám ochorenie),
- $P(+|C)$ pravdepodobnosť, že test je pozitívny, ak som pozitívny,
- $P(-|C)$ pravdepodobnosť, že test je negatívny, ak som pozitívny,
- $P(+|N)$ pravdepodobnosť, že test je pozitívny, ak som negatívny,
- $P(-|N)$ pravdepodobnosť, že test je negatívny, ak som negatívny,

- $P(C|+)$ pravdepodobnosť, že som pozitívny, ak je test pozitívny.

Keďže prevalencia testu je $1/100$, potom

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

potom zo zadania

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

Podľa Bayesovej vety platí

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

Pravdepodobnosť, že pacient, ktorý má pozitívny test, skutočne má ochorenie je teda iba 0,086, t. j. 8,6 %.

Poznámka. Je veľmi pravdepodobné, že veľa študentov tento výsledok pred výpočtom správne neodhadne. To nie je prekvapujúce, pretože štúdie ukázali, že až 95% lekárov nedokáže správne odpovedať na takéto otázky.

Úloha 2. Aká je pravdepodobnosť, že pacient, ktorý má, negatívny test je skutočne negatívny?

Riešenie. Použijeme označenie z predchádzajúceho cvičenia a Bayesovu vetu. Dostaneme

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

Teda pravdepodobnosť, že negatívne testovaný pacient je v skutočnosti zdravý, je 0,999583 t. j. 99,96 %.

Interpretácia bez použitia pravdepodobnosti

Skúsme vyriešiť prvú úlohu znova, ale preformulujeme zadanie tak, aby sme sa vyhli pojmu podmienenej pravdepodobnosti.

Úloha 3. Zo 100 000 ľudí má približne 1 000 celiakiu. Z týchto 1 000 ľudí test zistí celiakiu v 960 prípadoch. Zo zostávajúcich 99 000 ľudí, ktorí nemajú celiakiu, bude mať 10 200 ľudí tiež pozitívny test. Predstavme si skupinu tých ľudí, ktorí majú pozitívny test. Aká časť z nich skutočne celiakiu má?

Riešenie. Celkovo máme

$$960 + 10\,200 = 11\,160$$

pozitívnych výsledkov, ale len 960 skutočne chorých:

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

Vidíme, že podiel ľudí s ochorením, medzi tými, ktorí mali pozitívny test je 8,6 %.

V tomto výpočte sme si zjednodušili prácu a pracovali sme s celými číslami a zaokrúhlením na desiatky. Tým sme obetovali trochu presnosti pre zjednodušenie. Okrem toho sme predpokladali, že všetko dopadne tak, ako predpovedá pravdepodobnosť, že presne 1 000 ľudí má celiakiu. To však nie je vždy pravda. Hoci je táto metóda jednoduchšia a často ľahšie pochopiteľná, v učebniciach sa zvyčajne nevyskytuje. Avšak pri tomto zadaní úlohy je úspešnosť lekárov pri odhadovaní výsledku takmer 100%.

Literatúra

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.