

Diagnosticke testy v medicíne

Keywords: kombinatorika, pravděpodobnost a statistika, podmíněná pravděpodobnost, četnost jevů

Každý člověk podstupuje za život mnoho diagnostických testů a to až už u lékaře nebo doma. A je celkem jedno, zda-li se jedná o test na COVID-19, celiakii, plošný screening na mamografu nebo domácí těhotenský test. V každém případě chceme po takovém testu vědět, jestli je jeho výsledek spolehlivý. U každého takového testu se uvádí dva základní údaje

- *senzitivita testu*, tj. pravděpodobnost, že test odhalí nemoc, když ji mám;
- *specifita testu*, tj. pravděpodobnost, že test je negativní, když nemoc nemám.

Pro vyhodnocení je však důležitá i tzv. *prevalence nemoci*, tj. podíl počtu jedinců, kteří nemoc mají, a počtu všech jedinců v populaci.

Interpretace výsledku testu

V lékárnách a drogeriích se prodává řada domácích testů na alergie a intolerance, ale také na COVID-19 nebo těhotenské testy. Na příbalovém letáku testu na celiakii (intoleranci na lepek) se například můžeme dočíst, že senzitivita testu je 96,3 % a specifita testu je 89,7 %. Víme přitom, že počet celiaků v populaci je zhruba 1 %, tedy prevalence nemoci je 1/100.

Zajímat nás bude hlavně otázka: Jestliže je u mě test pozitivní, jaká je pravděpodobnost, že trpím celiakií? Odpověď zkuste nejprve odhadnout. V následující úloze provedeme přesný výpočet.

Úloha 1. Senzitivita testu je 96,3 %, specifita testu je 89,7 % a prevalence nemoci je 1/100. Jaká je pravděpodobnost, že pozitivně testovaný pacient je skutečně nemocný?

Řešení. Jedná se o příklad na podmíněnou pravděpodobnost, který se dá snadno vyřešit pomocí Bayesovy věty:

Nechť A je náhodný jev a nechť $B_1, \dots, B_n$ je úplný systém jevů. Jestliže $P(A) > 0$, $P(B_1) > 0, \dots, P(B_n) > 0$, pak

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

Symbolem $P(B|A)$ označujeme pravděpodobnost, že nastal jev B , když nastal jev A .

Označme

- $P(C)$ pravděpodobnost, že jsem pozitivní,
- $P(N)$ pravděpodobnost, že jsem negativní,
- $P(+|C)$ pravděpodobnost, že test je pozitivní, když jsem pozitivní,
- $P(-|C)$ pravděpodobnost, že test je negativní, když jsem pozitivní,
- $P(+|N)$ pravděpodobnost, že test je pozitivní, když jsem negativní,
- $P(-|N)$ pravděpodobnost, že test je negativní, když jsem negativní,

- $P(C|+)$ pravděpodobnost, že jsem pozitivní, když test je pozitivní.

Jelikož prevalence je 1/100, pak

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

a ze zadání

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

Podle Bayesovy věty platí

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

Pravděpodobnost, že pozitivně testovaný pacient je skutečně nemocný, je tedy pouze 0,086, tj. 8,6 % .

Poznámka. Je velmi pravděpodobné, že tento výsledek před výpočtem mnoho žáků neodhadne. Není divu, jelikož podle provedených studií na takto položené otázky nedokáže správně odpovědět až 95 % lékařů.

Úloha 2. Jaká je pravděpodobnost, že negativně testovaný pacient je opravdu negativní?

Řešení. Využijeme značení z předchozí úlohy a Bayesovu větu. Dostaneme pak

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

Pravděpodobnost, že negativně testovaný pacient je skutečně zdravý, je tedy 0,999583, tj. 99,96 % .

Interpretace bez použití pravděpodobnosti

Zkusíme nyní první úlohu vyřešit ještě jednou, ale pozměníme zadání tak, abychom se vyhnuli pojmu podmíněné pravděpodobnosti.

Úloha 3. Ze 100 000 lidí jich asi 1 000 má celiakii. Z této tisícovky lidí test odhalí celiakii v 960 případech. Ze zbylých 99 000 lidí, kteří celiakii nemají, bude mít 10 200 lidí též pozitivní výsledek testu. Představme si vzorek lidí, kteří mají pozitivní výsledek. Jaká část z nich skutečně trpí celiakií?

Řešení. Celkem máme tedy

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

pozitivních výsledků, ale jen 960 opravdu nemocných, tj.

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

Results matter!

Při tomto výpočtu jsme se dopustili dvou zjednodušení. Místo s desetinnými čísly pracujeme s čísly celými a zaokrouhlenými na desítky. Tedy trochu přesnosti jsme obětovali jednoduchosti. Navíc jsme předpokládali, že všechno dopadne tak, jak pravděpodobnost předpovídá, tj. například, že právě 1 000 lidí má celiakii. To nemusí být vždy pravda. V učebnicích tedy tento postup obvykle nenajdeme, ale lidé mu zpravidla lépe rozumí. Při tomto zadání úlohy je úspěšnost lékařů při odhadu výsledku téměř stoprocentní.

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.