

Prediktory onemocnění diabetes mellitus

Onemocnění diabetem je jedna z nejvíce rozšířených chorob, a právem se řadí mezi civilizační choroby. Existují dva typy, z nichž ten více rozšířený typ, diabetes mellitus 2. typu, je cukrovka vznikající až v dospělosti. Jedná se o metabolickou poruchu charakterizovanou zvýšenou hladinou glukózy v krvi při současné rezistenci na inzulin a relativním nedostatkem inzulinu (Kumar et al., 2004). Léčba je dlouhodobého charakteru a vzhledem k narůstajícímu trendu počtu pacientů s diabetem je stále větší podíl veřejných financí zdravotních pojišťoven vynakládán na pokrytí nákladů spojených s léčbou diabetu¹. Lama et al. (2021) zjistili, že indikátory jako je body mass index (BMI), index poměru boků a pasu, věk, systolický a diastolický krevní tlak a rodinná historie jsou nejvíce signifikantní prediktory výskytu diabetu.

Binomiální logistická regrese

Pomocí dat, která byla volně ke stažení², jsem se pokusil ověřit tento předpoklad pomocí binomiální logistické regrese. Binomiální logistická regrese je statistický model používaný ke zkoumání vztahu mezi jedním nebo více proměnnými a binárními odpověďmi. Model odhaduje pravděpodobnost úspěchu nebo selhání binární odpovědní proměnné na základě hodnot nezávislých proměnných (Hosmer et al., 2013). V našem případě máme tyto proměnné:

Závisle proměnná = Výstup - (ne)přítomnost diagnózy diabetes mellitus u pacienta

Nezávislé proměnné:

- **Glukóza** – hladina glukózy v krvi (mg/dl.)
- **Krevní tlak** - Diastolický krevní tlak (mm/Hg)
- **Tloušťka kůže** (mm)
- **Inzulín** – hladina inzulinu v krvi (U/mL)
- **BMI** – Body mass index (kg/m²)
- **Index vrozenosti** – index určující pravděpodobnost výskytu diabetu v rodinné historii pacienta
- **Věk**

¹ V roce 2020 bylo na péči o pacienty v České republice k dispozici 320 miliard Kč, z toho pouze na léčbu diabetu se spotřebovalo až 48 miliard Kč.

² <https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database?resource=download>

Datová matice obsahovala údaje celkem 768 pacientek indického Národního institutu pro léčbu diabetu, nemocí trávicího a vylučovacího traktu. Jelikož ne u všech pacientek byla data kompletní, musel jsem nejprve data vyčistit. Celkový počet pacientek po vyčištění dat činil 385. K výpočtu logistické regrese byl použitý program Statistica. Byla provedena binomiální logistická regrese týkající se proměnných vypsanych výše. Výsledky modelu a jednotlivé hodnoty – odhadu, šance a Waldovy statistiky ke každému regresoru spolu s hladinou významnosti p shrnuje Tabulka 1.

Tab. 1 pro závislou proměnnou „Výstup“ - parametrický odhad pravděpodobnosti, že Výstup = 1 (pozit. diag. diabetu)

	Šance	Odhad (b)	Waldova statistika	p-hodnota
konstanta		-10,20	70,32	0,00
Glukóza	1,04	0,04	43,69	0,00
Krevní tlak	1,00	0,00	0,02	0,88
Tloušťka kůže	1,01	0,01	0,63	0,43
Inzulín	1,00	0,00	0,70	0,40
BMI	1,07	0,07	6,21	0,01
Index vrozenosti	2,55	0,94	4,75	0,03
Věk	1,05	0,05	13,56	0,00

*Červeně jsou zvýrazněny signifikantní výsledky na hladině významnosti $p < 0,05$

Signifikantní výsledek byl prokázán u 4 proměnných. Jedná se o regresory: glukóza, body mass index, index vrozenosti a věk. Tyto proměnné mají velmi malou p hodnotu, a zároveň Waldova statistika dosahuje vysokých hodnot. Na základě těchto výsledků můžeme o těchto proměnných uvažovat jako o významných v souvislosti s detekcí onemocnění diabetu. Nejvyšší šanci má proměnná index vrozenosti, která nám udává, že za každý získaný bod nezávislé proměnné indexu vrozenosti se zvětší šance, že bude u této osoby diagnostikovaný diabetes 2,55krát. Je tedy pravděpodobný vysoký podíl heritability na rozvoji tohoto onemocnění.

Kontrola kvality modelu

Pro zjištění kvality modelu jsou využity ukazatele Nagelkerke R^2 a Cox-Snell R^2 . Hodnota Cox-Snell R^2 je rovna 0,32, tedy vysvětluje přibližně 32% a hodnota Nagelkerke R^2 je rovna hodnotě 0,45. Ta nám udává, že pomocí modelu sestaveného z našich regresorů (proměnných) jsme schopni vysvětlit 45% veškerého rozptylu vysvětlované proměnné R^2 . Celkově tedy tyto ukazatele naznačují, že binomiální logistický regresní model má střední až dobrou přizpůsobivost k datům a může být použit k vytvoření poměrně přesných predikcí.

Diskuze

Naše analýza biominální logistické regrese potvrzuje zjištění Lama et al. (2021), že indikátory jako je body mass index (BMI), věk, a rodinná historie má signifikantní vliv na výskyt diabetu s tím, že největší šanci výskytu nemoci zvyšuje právě tento hereditární vliv a to více než 2,5krát. Nicméně od zmíněných autorů se liší výsledky v proměnné diastolického krevního tlaku, která v naší analýze nenabyla hladiny významnosti.

Zdroje:

<https://www.foodnet.cz/cs/aktuality/2730-zhruba-milion-cechu-trpi-cukrovkou-a-nemocnych-neustale-pribyva>

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression. John Wiley & Sons.

Kumar, V., Abbas, A. K., & Fausto, N. (2004). *Robbins and cotran pathologic basis of disease*. (7th ed.). W B Saunders.

Lama L, Wilhelmsson O, Norlander E, Gustafsson L., Lager A., Tynelius P., Wärvik L., & Östenson C. G. (2021) Machine learning for prediction of diabetes risk in middle-aged Swedish people Heliyon 7:e07419. <https://doi.org/10.1016/j.heliyon.2021.e07419>