

## VZTAH POHLAVÍ A RIZIKA VZNIKU SRDEČNÍ CHOROBY

Světová zdravotnická organizace odhaduje, že každý rok na celém světě zemře v důsledku srdečních chorob 12 milionů lidí.

Existuje mnoho studií, které ukazují, že muži mají vyšší riziko vzniku kardiovaskulárních chorob než ženy a že je mnoho faktorů, které to mohou způsobit.

Například studie <sup>1</sup> zveřejněná v "Journal of the American College of Cardiology" v roce 2019 prokázala, že muži mají vyšší pravděpodobnost vzniku aterosklerózy, což je stav, kdy se cholesterol a další látky hromadí na stěnách tepen.

Další studie <sup>2</sup> publikovaná v "European Heart Journal" v roce 2020 ukázala, že muži mají větší riziko kardiovaskulárních chorob kvůli genetickým faktorům.

Dle současných výzkumů lze tedy tvrdit, že muži mají mnohem vyšší pravděpodobnost vzniku srdečních chorob než ženy. My se nyní pomocí logistické regrese pokusíme toto tvrzení ověřit.

Zdrojem dat v tomto modelování je veřejně přístupný soubor <sup>3</sup> umístěný na webových stránkách Kaggle, který pochází z probíhající kardiovaskulární studie na obyvatelích města Framingham v Massachusetts.

Nyní budou data ze studie využita pro vysvětlení statistické metody **Logistická regrese**. Tato statistická metoda se využívá k predikci pravděpodobnosti výskytu určitého jevu (v našem případě vzniku kardiovaskulární choroby) na základě vstupních dat. To umožňuje odhalit faktory a posoudit nakolik ovlivňují, zda sledovaný jev nastane či nikoliv.

**Naším cílem je zkoumat vztah nezávislých – vysvětlujících proměnných X a závislé – vysvětlované proměnné Y.**

**Nezávislé proměnné X** jsou získány z demografické, behaviorální a zdravotní oblasti, přičemž každá proměnná z těchto oblastí je potenciálním rizikovým faktorem.

Demografické proměnné:

- Pohlaví: muž nebo žena (nominální)
- Věk: Věk pacienta

Behaviorální proměnné:

- Kuřák: zda pacient je nebo není současný kuřák (nominální)
- Cigaret/den: počet cigaret, které osoba vykouřila v průměru za jeden den. (Ize považovat za nepřetržité, protože člověk může mít libovolný počet cigaret, dokonce i polovinu cigarety.)

Zdravotní proměnné (historické údaje):

- Léky na krevní tlak: zda pacient užíval léky na krevní tlak (nominální)
- Prevalentní mrtvice: zda pacient v minulosti prodělal mrtvici (nominální)
- Prevalentní hyp: zda pacient byl nebo nebyl hypertenzní (nominální)
- Diabetes: zda pacient měl nebo neměl diabetes (nominální)

Zdravotní proměnné (aktuální údaje):

- CelkChol: hladina celkového cholesterolu
- SysTI: systolický krevní tlak
- DiaTI: diastolický krevní tlak
- BMI: Index tělesné hmotnosti
- Srdfre: srdeční frekvence
- Glukóza: hladina glukózy

**Závislá proměnná Y** je v našem případě:

- 10leté riziko budoucí ischemické choroby srdeční (ICHS), 1= „Ano“ X 0 = „Ne“

Následující výstup popisuje výsledky logistické regrese, která byla provedena s cílem odhadnout pravděpodobnost výskytu ischemické choroby srdeční (ICHS) po dobu 10 let. Zobrazeny jsou údaje o několika proměnných: věk, pohlaví, kouření, předchozí mrtvice, předchozí hypertenze a diabetes.

**Tab.1 – výstup Logistické regrese modelující pravděpodobnost vzniku ICHS v 10 letech**

Effect	10leté riziko ICHS – Odds Ratios (List1 in Dataset ICHS) Distribuce: BINOMIALNÍ Funkce: Logistická regrese modelující pravděpodobnost vzniku ICHS v 10 letech = 0			
	OR: Odds Ratio	Lower CL 95 %	Upper CL 95%	p
Intercept	<b>1</b>			
Věk	<b>0,933188</b>	0,922785	0,943708	0
<b>Pohlaví</b>	<b>1,666194</b>	<b>1,390501</b>	<b>1,996547</b>	<b>0</b>
Kuřák	<b>1,463802</b>	1,214384	1,764448	0,000064
Mrtvice	<b>2,671917</b>	1,137172	6,277978	0,02414
Hypertenze	<b>1,963927</b>	1,635845	2,357807	0
Diabetes	<b>2,372616</b>	1,561053	3,606094	0,000052

Každá proměnná má svůj vlastní řádek v tabulce, kde jsou uvedeny následující informace:

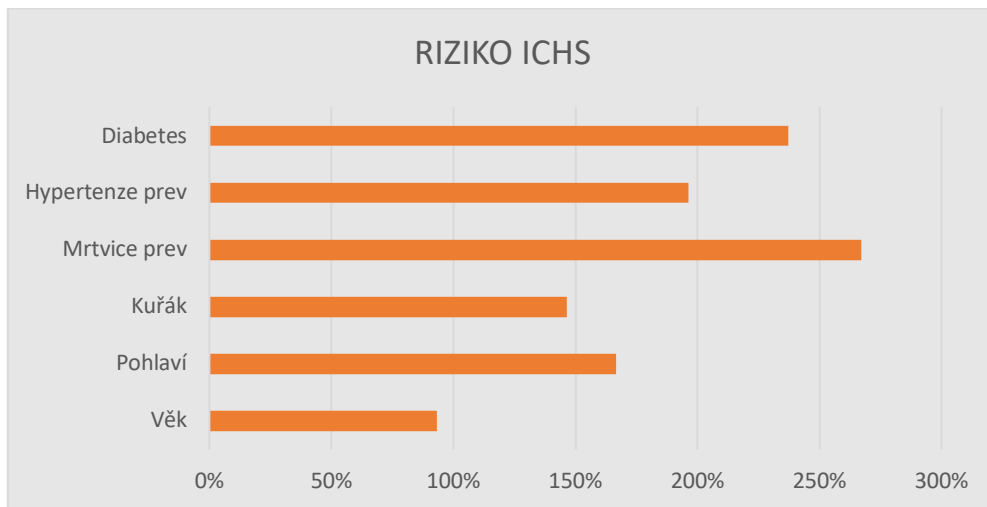
- Název proměnné (např. Věk, Pohlaví atd.)
- Hodnota OR (Odds ratio), která ukazuje, jaký vliv má každý faktor na pravděpodobnost výskytu ICHS v porovnání s referenční hodnotou (interceptem= 1).
- Dolní a horní hranice 95 % intervalu spolehlivosti (CL) pro odhad hodnoty Odds ratio.
- P hodnota, která ukazuje, zda je vztah mezi faktorem a ICHS statisticky významný (p <0,05 je považováno za statisticky významné).

Dle našich výsledků je každý faktor statisticky významně spojen s rizikem výskytu ICHS po dobu 10 let, protože p hodnota pro každý faktor je méně než 0,05.

Námi sledovaný **vztah mezi pohlavím a rizikem vzniku ICHS se potvrdil**, neboť dle údajů mají ženy nižší riziko ICHS než muži (**odhad Odds ratio = 1,67**). Podobně, nekuřáci mají nižší riziko ICHS než kuřáci (odhad odds ratio = 1,46), stejně jako lidé bez předchozí mrtvice nebo hypertenze (odhady odds ratio = 2,67 a 1,96).

Výsledky shrnuje následující graf:

**Graf.1 – výstup Logistické regrese modelující pravděpodobnost vzniku ICHS v 10 letech**



Celkově lze tedy konstatovat, že každý z těchto faktorů může ovlivnit pravděpodobnost výskytu ICHS po dobu 10 let, a proto by měly být brány v úvahu při hodnocení rizika tohoto onemocnění.

#### Zdroje:

1. Singh A., Collins B.L., Gupta A., et al. (2019). Sex Differences in Coronary Atherosclerosis and Risk Factors in Young Patients With Myocardial Infarction. *J Am Coll Cardiol.* 74 (13), 1613-1622. doi: 10.1016/j.jacc.2019.07.079
2. Folkersen L., Van't Hooft F., Chernogubova E., et al. (2020). Genetic predisposition to coronary artery disease in Swedish and Polish populations. *Eur Heart J.* 41 (19), 1828-1838. doi:10.1093/eurheartj/ehz745
3. Logistic reegtresion to predict heart desease. Získáno 23.2.2023 z: <https://www.kaggle.com/datasets/dileep070/heart-disease-prediction-using-logistic-regression?resource=download>
4. MUNI – Math and stats support centre. (n.d). Jednoduchá lineární regrese. Získáno 23.3.2023 z: <https://mathstat.econ.muni.cz/media/19031/linearni-regrese.pdf>

---

Data a další informace o této zprávě jsou dostupné na adrese <https://dostal.vyzkum-psychologie.cz/stat4?i=66>