

# PREVALENCIA A PREDIKTORY DIABETES MELLITUS

## LOGISTICKÁ REGRESIA

V tejto správe sa budeme zaoberať diagnostickými údajmi, ktoré môžu mať prediktívny potenciál určiť prítomnosť cukrovky. Dataset, s ktorým sa bude pracovať, obsahuje reálne údaje pacientov, ktoré boli zozbierané v National Institute of Diabetes and Digestive and Kidney Diseases. Pôvodne však nešlo o úplne správne nakódované dáta, nakoľko mnoho údajov nadobúdalo hodnoty 0, čo nebolo v realite možné. Preto sme v tejto práci nahradili reálne nemožné hodnoty 0 ako chýbajúce hodnoty. Budeme vytvárať logistickú regresiu. Ide o štatistickú metódu, ktorá odhaduje pravdepodobnosť javu na dichotomických premenných.

Diabetes mellitus, známy aj ako cukrovka, je závažné chronické ochorenie, ktoré postihuje milióny ľudí na celom svete. Patrí medzi najrozšírenejšie endokrinné ochorenia a jeho výskyt v posledných rokoch rýchlo narastá. Hlavnou príčinou cukrovky je nedostatočná produkcia inzulínu, hormónu, ktorý umožňuje bunkám v tele využívať glukózu z krvného obehu na výrobu energie. Pri diabetes sa buď inzulín vôbec neprodukuje (typ 1) alebo ho telo produkuje nedostatočné množstvo alebo ho nedokáže využiť efektívne (typ 2). Okrem toho existuje aj tzv. gestačný diabetes, ktorý sa vyskytuje u tehotných žien.

Cukrovka má vplyv na celkovú kvalitu života pacientov a zvyšuje ich riziko vzniku ďalších zdravotných problémov. Napriek tomu však existuje množstvo spôsobov, ako ju kontrolovať a predchádzať jej. V súčasnosti sa vyvíja množstvo liekov a terapeutických postupov, ktoré majú pomôcť pacientom s cukrovkou žiť plnohodnotný a zdravý život.

Nakoľko prevalencia cukrovky rapídne narastá, v oblasti predikcie a diagnostiky sa neustále uskutočňuje množstvo nezávislých výskumov, zameriavajúce sa na široké spektrum potencionálnych prediktorov. Existuje množstvo faktorov, ktoré môžu byť prediktormi vzniku cukrovky. Jedným z najvýznamnejších je nedostatočná aktivita, sedavý spôsob života, nadváha a obezita. Okrem týchto faktorov sú dôležité aj genetické faktory. Štúdie ukazujú, že ak jeden z rodičov má cukrovku, dieťa má tiež až 50% šancu na jej prepuknutie. Okrem toho sa ukázalo, že aj etnická príslušnosť môže byť prediktorom cukrovky, pričom niektoré etnické skupiny majú vyššie riziko vzniku ochorenia ako iné. V neposlednom rade hrá dôležitú úlohu aj životný štýl. Konkrétne fajčenie a nadmerná konzumácia alkoholu.

Na základe týchto údajov sme sa v tejto správe rozhodli vytvoriť logistickú regresiu, zameranú na potencionálne prediktory diagnostiky cukrovky a z toho vyplývajúci štatistický odhad, že človek bude pozitívne diagnostikovaný.

Na základe dostupných dát sme pracovali s ôsmimi nezávislými premennými, menovite:

- Vek
- BMI
- Tvorba inzulínu
- Dedičný predpoklad
- Hrúbka kože
- Glukóza v krvi
- Krvný tlak
- Tehotenstvo

Naša závislá premenná bola diagnostika ochorenia, ktorá bola v pozitívnom prípade kódovaná ako 1, v negatívnom ako 0. Dataset obsahoval reálne údaje od 768 pacientov, s minimálnym vekom 21 rokov. Dáta boli uložené vo formáte cvs. a spracované v štatistickom programe Jasp. 0.16.3.

Hodnoty glukózy, krvného tlaku, BMI, hrúbky kože a inzulínu však dosahovali hodnôt 0, čo je nemožné, preto boli tieto hodnoty kódované ako chýbajúce.

V nasledujúcej tabuľke č.1 je možné prezrieť výsledky logistickej regresie, ktorá zahŕňa ako údaje o odhade, tak aj p- hodnotu a údaje o pomere šancí na pozitívnu diagnostiku, ktoré sú pre nás pravdepodobne najľahšie interpretovateľné.

### Tabuľka č.1

Logistická regresia

	Odhad	SE	Pomer šancí	z	Wald. test		
					Wald. Štat.	df	p
Konštanta	-9.35	0.971	8.56e-5	-9.64	93.03	1	< .001
Vek	0.029	0.014	1.029	2.095	4.388	1	0.036
Inzulín	-0.001	0.001	0.999	-0.80	0.643	1	0.423
Hrúbka kože	0.006	0.015	1.006	0.409	0.167	1	0.683
Krvný tlak	-0.08	0.010	0.992	-0.84	0.718	1	0.397
BMI	0.091	0.023	1.095	3.943	15.545	1	< .001
Dedičnosť	1.297	0.357	3.658	3.634	13.209	1	< .001
Glukóza	0.033	0.004	1.033	7.250	52.567	1	< .001
Tehotenstvo	0.120	0.043	1.128	2.798	7.828	1	0.005

Výsledky naznačujú, že nezávislé premenné vek, BMI, dedičnosť, glukóza v krvi a tehotenstvo sú potencionálnym prediktorom prítomnosti alebo budúceho rozbehnutia cukrovky u jedinca. Z ďaleka najvýraznejšie sa to však ukazuje v kategórii dedičnosť, čo potvrdilo aj mnoho predchádzajúcich výskumov a štúdií.

V tabuľke č.2 vidíme presnosť odhadu modelu, ktorý nám program Jasp vypočítal. Ukazuje sa že v 78% prípadov program predikoval diagnózu správne.

## Tabuľka č.2

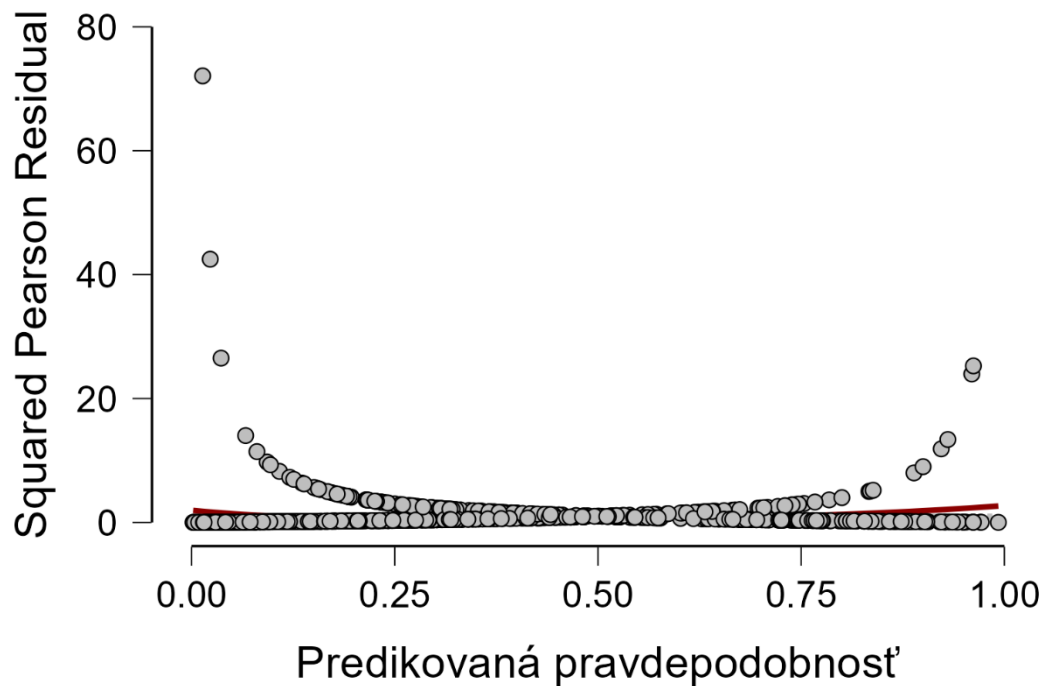
Presnosť odhadu modelu

pozorované	predikcia		% správne
	0	1	
0	320	39	89.136
1	77	102	56.983
celková presnosť			78.439

poznámka: cut-off skóre je 0,5

## Graf č.1

Graf pearsonových reziduí



V závere teda konštatujeme, že logistická regresia poukázala na 5 potencionálne významných diagnostických premenných. Presnosť odhadu modelu je vcelku vysoký a tak získané údaje vieme považovať za potencionálny odrazový mostík pre ďalšie bádanie. Graf nám naznačuje, že v našej vzorke nie je veľký rozptyl. Limity tejto práce však spočívajú v starších dátach a veľkosti vzorky, ktorá bola zameraná len na jednu etnicitu.

## Zdroje

Kennelly, M. A., & McAuliffe, F. M. (2016). Prediction and prevention of Gestational Diabetes: an update of recent literature. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 202, 92-98. <https://www.sciencedirect.com/science/article/pii/S0301211516301221>

Khan, F. A., Zeb, K., Al-Rakhami, M., Derhab, A., & Bukhari, S. A. C. (2021). Detection and prediction of diabetes using data mining: a comprehensive review. *IEEE Access*, 9, 43711-43735. <https://ieeexplore.ieee.org/abstract/document/9354154>

Thom, G., Messow, C. M., Leslie, W. S., Barnes, A. C., Brosnahan, N., McCombie, L., ... & Lean, M. E. J. (2021). Predictors of type 2 diabetes remission in the Diabetes Remission Clinical Trial (DiRECT). *Diabetic Medicine*, 38(8), e14395. <https://onlinelibrary.wiley.com/doi/full/10.1111/dme.14395>

Dáta získané z: <https://www.kaggle.com/code/mohammadabdulwahed/linear-discriminant-analysis-on-diabetes-dataset/notebook#Linear-Discriminant-Analysis-on-Diabetes-Data>