

Prediktory progresu diabetu

Úvod a teoretické ukotvení

Diabetes mellitus 2. typu patří mezi onemocnění, která výrazně ovlivňují kvalitu života pacientů. Kromě nutnosti celoživotní změny životního stylu se u diabetiků častěji vyskytují psychické obtíže, včetně deprese a úzkosti (Anderson et al., 2001). Proto je důležité porozumět faktorům, které ovlivňují rychlost progresu onemocnění. Toto porozumění má přínos nejen z medicínského hlediska, ale i z hlediska psychologické prevence.

Mezi hlavní ovlivnitelné rizikové faktory progresu diabetu patří body mass index (BMI) a průměrný krevní tlak. Oba ukazatele jsou přitom do určité míry ovlivněny psychosociálními faktory, jako je stres nebo adherence k léčbě (Stephoe & Kivimäki, 2012). Věk a pohlaví pak představují neovlivnitelné demografické proměnné, které je potřeba v analýze kontrolovat.

Vícenásobná lineární regrese umožňuje kvantifikovat vztah mezi několika prediktory a jednou spojitou závislou proměnnou. Metoda nejmenších čtverců (OLS) odhaduje regresní koeficienty β tak, aby minimalizovala součet čtverců reziduí (Chatterjee & Hadi, 2006). Oproti jednoduché regresi má výhodu v tom, že koeficient β pro daný prediktor vyjadřuje jeho efekt při současné kontrole ostatních proměnných v modelu.

Cílem této analýzy bylo zjistit, které ze čtyř vybraných prediktorů (**BMI, krevní tlak, věk, pohlaví**) statisticky významně předpovídají míru progresu diabetu rok po vstupním měření, a kvantifikovat velikost jejich efektu.

Metoda

Data. Byl použit dataset Efron et al. (2004), dostupný v balíčku *sklearn*. Soubor obsahuje 442 pacientů s diabetem. Výslednou proměnnou je kvantitativní míra progresu onemocnění jeden rok po vstupním měření ($M = 152,1$; $SD = 77,0$; rozsah 25–346). Jako prediktory byly zvoleny čtyři klinicky nejintuitivnější ukazatele: BMI ($M = 26,4$; $SD = 4,4$ kg/m²), průměrný krevní tlak ($M = 94,7$; $SD = 13,8$ mm Hg), věk ($M = 48,5$; $SD = 13,1$ let) a pohlaví (47 % žen).

Statistická metoda. Byl odhadnut model vícenásobné lineární regrese metodou nejmenších čtverců (OLS):

$$\hat{y} = \beta_0 + \beta_1 \cdot BMI + \beta_2 \cdot TK + \beta_3 \cdot Věk + \beta_4 \cdot Pohlaví + \varepsilon$$

Koeficient β_j vyjadřuje, o kolik jednotek se změní predikovaná progresu, pokud se daný prediktor zvýší o 1 jednotku (při kontrole ostatních proměnných). Celková shoda modelu je hodnocena koeficientem determinace R^2 a jeho upravenou verzí (Radj²). Statistická významnost modelu jako celku je ověřena F-testem, významnost jednotlivých koeficientů t-testem. Splnění předpokladů regrese (normalita a homoskedasticita reziduí) je ověřeno graficky pomocí diagnostických grafů.

Výsledky

Fit modelu

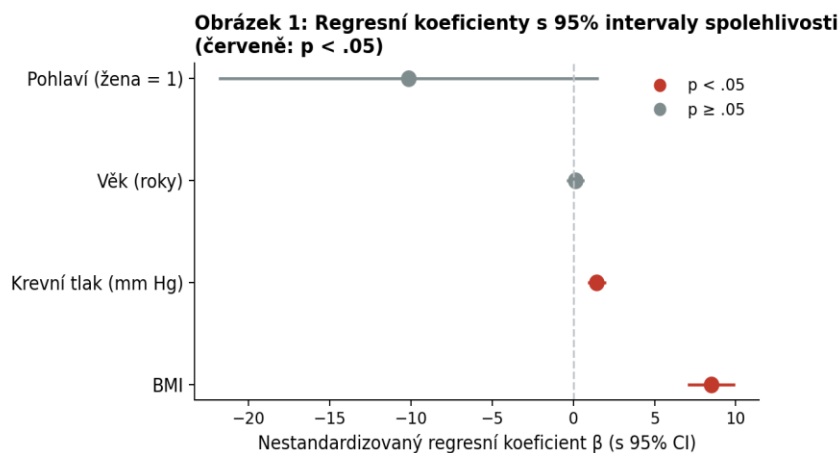
Model jako celek byl statisticky významný: $F(4, 437) = 72,91$, $p < ,001$. Koeficient determinace $R^2 = ,400$ znamená, že čtyři prediktory společně vysvětlují 40,0 % variance v progresu diabetu (Radj² = ,395). Zbytek variance připadá na faktory, které model nezahrnuje (např. genetik, adherence k léčbě, hladina cholesterolu).

Regresní koeficienty

Výsledky regresní analýzy shrnuje Tabulka 1. Zvýrazněné řádky označují statisticky významné prediktory ($p < ,05$).

Prediktor	β	SE	t	p	95% CI
(Intercept)	-209,23	22,63	-9,24	< ,001	[-253,59; -164,87]
BMI	8,48	0,71	12,03	< ,001	[7,10; 9,87]
Krevní tlak (mm Hg)	1,44	0,24	6,00	< ,001	[0,97; 1,90]
Věk (roky)	0,14	0,23	0,58	,562	[-0,32; 0,59]
Pohlaví (žena = 1)	-10,16	5,92	-1,72	,087	[-21,77; 1,45]

Tabulka 1. Výsledky vícenásobné lineární regrese. β = nestandardizovaný koeficient; SE = standardní chyba; zvýrazněné řádky: $p < ,05$.



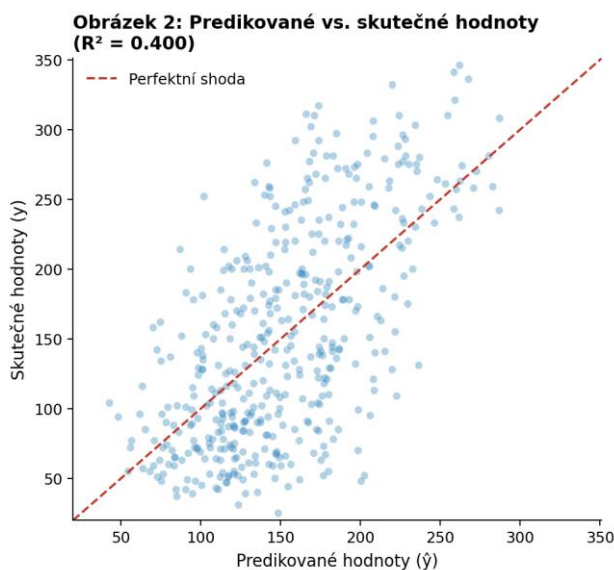
Zdroj: Efron et al. (2004), N = 442

Obrázek 1. Regresní koeficienty β s 95% intervaly spolehlivosti. Červeně: statisticky významné prediktory ($p < ,05$). Přerušovaná čára: $\beta = 0$.

BMI byl nejsilnějším prediktorem: každý 1 kg/m² nárůst BMI odpovídá průměrnému nárůstu progresního skóre o 8,48 bodu (95% CI [7,10; 9,87], $p < ,001$). Krevní tlak měl rovněž signifikantní efekt, kdy zvýšení o 1 mm Hg odpovídá nárůstu o 1,44 bodu (95% CI [0,97; 1,90], $p < ,001$). Věk ($\beta = 0,14$, $p = ,562$) ani pohlaví ($\beta = -10,16$, $p = ,087$) nebyly statisticky významné.

Shoda predikce se skutečností

Obrázek 2 zobrazuje vztah mezi predikovanými a skutečnými hodnotami progresu. Přerušovaná čára označuje perfektní shodu ($\hat{y} = y$). Body leží přiměřeně blízko přímce, ale rozptyl se zvětšuje pro vyšší hodnoty, což může ukazovat na mírnou heteroskedasticitu.



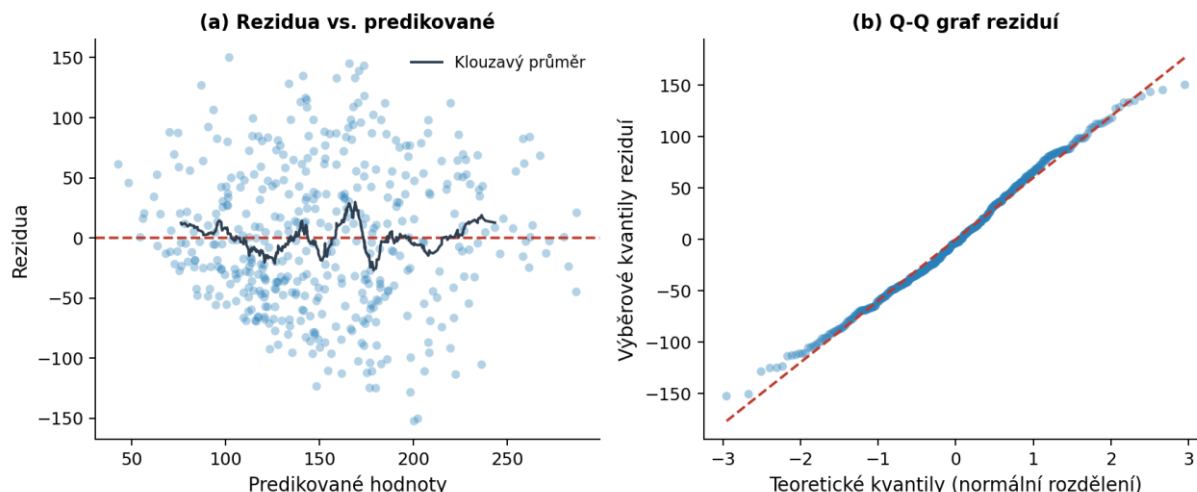
Zdroj: Efron et al. (2004), N = 442

Obrázek 2. Predikované hodnoty (\hat{y}) vs. skutečné hodnoty (y). Přerušovaná čára: ideální shoda ($\hat{y} = y$). $R^2 = ,400$.

Diagnostika reziduí

Obrázek 3 ukazuje dva diagnostické grafy. Panel (a) zobrazuje rezidua v závislosti na predikovaných hodnotách. Přerušovaná čára na nule by měla být bez systematického vzorce; klouzavý průměr ale naznačuje mírný nelineární trend, který by šlo řešit např. zahrnutím kvadratického členu. Panel (b) je Q-Q graf pro posouzení normality reziduí. Body přibližně sledují přímku s mírným odklonem v ocasech, což je u regresního modelu s tímto vzorkem akceptovatelné.

Obrázek 3: Diagnostika reziduí



Obrázek 3. Diagnostika reziduí: (a) rezidua vs. predikované hodnoty, (b) Q-Q graf normality reziduí.

Diskuse a limity

Výsledky ukazují, že BMI a krevní tlak jsou významnými prediktory progresu diabetu i při vzájemné statistické kontrole. Efekt BMI je podstatně silnější než efekt krevního tlaku, i když je třeba mít na paměti, že oba prediktory jsou měřeny v různých jednotkách a přímé srovnání nestandardizovaných koeficientů proto není zcela vypovídající. Nesignifikantní efekt věku je poněkud překvapivý, protože stárnutí je v klinické literatuře považováno za rizikový faktor (Chatterjee & Hadi, 2006). Jedním z možných vysvětlení je vzájemná korelace věku s ostatními prediktory v modelu.

Analýza má několik omezení. Model vysvětluje 40 % variance, což znamená, že velká část variability progresu závisí na faktorech, které jsme do modelu nezahrnuli (genetické dispozice, míra stresu, životní styl apod.). Diagnostické grafy navíc naznačují mírnou heteroskedasticitu, která může zkreslovat standardní chyby koeficientů. Data jsou také průřezová, takže z výsledků nelze vyvozovat kauzální závěry.

Závěr

Vícenásobná lineární regrese ukázala, že **BMI a krevní tlak** jsou statisticky významnými prediktory progresu diabetu ($R^2 = ,400$; $F(4, 437) = 72,91$, $p < ,001$). Věk a pohlaví se v modelu neukázaly jako významné. Analýza ilustruje, jak regresní model umožňuje posoudit vliv jednotlivých prediktorů při vzájemné kontrole a jak diagnostické grafy pomáhají ověřit předpoklady metody.

Zdroje

- Anderson, R. J., Freedland, K. E., Clouse, R. E., & Lustman, P. J. (2001). The prevalence of comorbid depression in adults with diabetes: A meta-analysis. *Diabetes Care*, 24(6), 1069–1078. <https://doi.org/10.2337/diacare.24.6.1069>
- Chatterjee, S., & Hadi, A. S. (2006). *Regression analysis by example* (4th ed.). Wiley. <https://doi.org/10.1002/0470055464>
- Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32(2), 407–499. <https://doi.org/10.1214/009053604000000067>
- Steptoe, A., & Kivimäki, M. (2012). Stress and cardiovascular disease. *Nature Reviews Cardiology*, 9(6), 360–370. <https://doi.org/10.1038/nrcardio.2012.45>