

Zápočet do předmětu Vícerozměrné statistické metody

Úvod

V tomto textu se budu zabývat metodou vícenásobné lineární regrese, a to konkrétně ve spojitosti se srdečními onemocněními. Na srdeční choroby má vliv mnoho psychologických i nepsychologických faktorů, například deprese (Denollet & Van Heck, 2001). Dále také absence fyzické aktivity (Williams, 2001). Fyzická aktivita se může podílet také na našem well-beingu, může zvyšovat naše kompetence, sociální interakce a její přínos pro psychické zdraví se jeví jako významný, a fyzická aktivita může být přínosem jak pro zdravé osoby, tak pro osoby trpící depresí či úzkostí (Wankel, 1993).

Další faktor, který má vliv na vznik srdečních onemocnění je kouření (Gallucci, 2020). Kouření je také ovlivněno řadou psychologických faktorů. Rondina et al. (2007) uvádí, že kuřáci bývají více extravertní, impulzivní, úzkostní, a vykazují více rysů neuroticismu a psychotismu než bývalí kuřáci nebo nekuřáci. Dále uvádí souvislost mezi kouřením a duševními poruchami, jako je schizofrenie a deprese.

V následujícím textu budu ověřovat vztah mezi kouřením a fyzickou aktivitou, konkrétně jízdou na kole a srdečním onemocněním. Přestože se nejedná přímo o psychologické faktory, tak se na základě výše zmíněného domnívám, že s nimi velmi úzce souvisí.

Metoda analýzy dat

Vztah mezi kouřením, jízdou na kole a onemocněním srdce byl ověřován na následujících datech o 498 pozorování. Jedná se o 498 amerických měst, **nezávisle proměnnou** je procento obyvatel daného města, které má srdeční onemocnění. **Prediktory** jsou procento obyvatel, které kouří v konkrétním městě, a procento obyvatel, které jezdí na kole. Pro zpracování jsem použila program Jamovi 2.2.5.

Hypotézu, že mezi kouřením, jízdou na kole a srdečními onemocnění existuje vztah, lze ověřit pomocí vícenásobné lineární regrese. Předtím, než budeme hypotézu testovat, je nutné ověřit, zda data splňují předpoklady pro použití dané metody.

Data a další informace o této zprávě jsou dostupné na adrese: https://www.scribbr.com/wp-content/uploads//2020/02/heart.data_.zip

Předpoklady

Jedním z předpokladů pro použití metody je nepřítomnost *outliers*, tedy příliš odlehlých hodnot. Tento předpoklad ověříme, pomocí **Cook's distance** (Cookovy vzdálenosti), která by neměla být vyšší než 1. V tabulce 1 je patrné, že tento předpoklad je splněn, jelikož maximální Cookovy vzdálenosti je 0,0308.

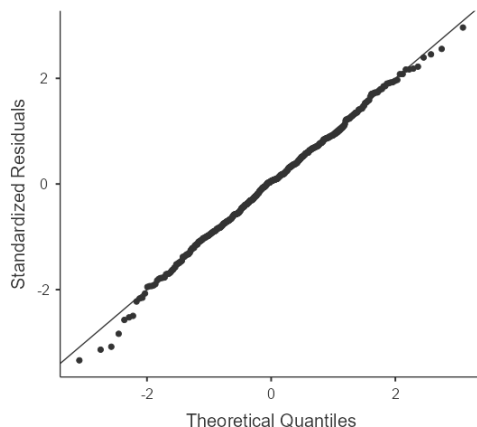
Tabulka 1

Cook's distance			Rozsah	
Mean	Median	SD	Min	Max
0.00200	7.90e-4	0.00335	7.32e-9	0.0308

Dalším předpokladem je normalita reziduí, kterou ověříme pomocí **Shapiro-Wilkova testu** a **Q-Q grafu**. V tomto případě test normality není statisticky významný ($p = 0,494$) a jednotlivá pozorování v Q-Q grafu téměř kopírují diagonálu, což naznačuje, že tento předpoklad je také splněn.

Graf 1

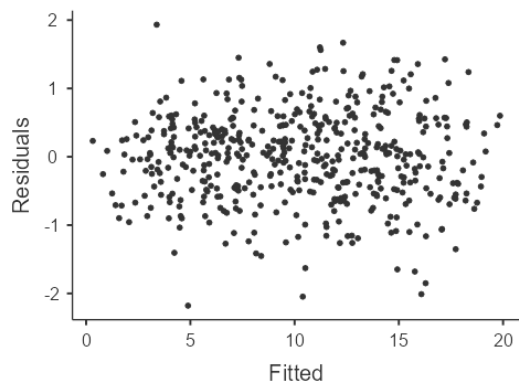
Q-Q Graf



Dále je nutné ověřit **homoskedasticitu**, kterou zjistíme z grafu reziduí. Předpoklad homogeneity rozptylu je splněn, což lze poznat podle toho, že naše naměřená data vytvářejí náhodný rozptyl kolem přímky procházející nulou na ose y. V Grafu 2 je graf reziduí pro celkový model.

Data a další informace o této zprávě jsou dostupné na adrese: https://www.scribbr.com/wp-content/uploads//2020/02/heart.data_.zip

Graf 2



Dalším předpokladem je **multikolinearita**, což znamená, že regresory nemají dokonalý lineární vztah. Tento předpoklad zjistíme pomocí *Tolerance value*, která by měla být vyšší než 0,2.

Tabulka 2

Collinearity Statistics

	VIF	Tolerance
smoking	1.00	1.000
biking	1.00	1.000

Výsledky

Všechny nutné předpoklady pro provedení vícenásobné lineární regrese data splňují. Výsledky analýzy jsou v tabulce 3 a tabulce 4.

Tabulka 3

Model Fit Measures

Model	R	R ²	Adjusted R ²	Overall Model Test			
				F	df1	df2	p
1	0.990	0.980	0.980	11895	2	495	< .001

Data a další informace o této zprávě jsou dostupné na adrese: https://www.scribbr.com/wp-content/uploads//2020/02/heart.data_.zip

Tabulka 4

Model Coefficients - heart.disease

Predictor	Estimate	SE	t	p
Intercept	14.985	0.08014	187.0	< .001
smoking	0.178	0.00354	50.4	< .001
biking	-0.200	0.00137	-146.5	< .001

Pomocí vícenásobné lineární regrese jsme testovali, zda procento obyvatel v daném městě, kteří jezdí na kole a procento obyvatel, kteří kouří, ovlivňuje procento lidí, kteří mají srdeční onemocnění. Prediktory mají s procentem obyvatel se srdečními chorobami signifikantní vztah, $p = < 0,001$, adjustované R^2 nabývá hodnoty 0,98, tato hodnota nám říká, kolik variability je vysvětleno modelem s dvěma prediktory.

Procento kuřáků (smoking) signifikantně předpovídá procento lidí se srdečními onemocněními, $\beta = 0,178$, $t(495) = -50,4$, $SE = 0,00354$, $p < .001$. Procento kolarů (biking) také signifikantně předpovídá procento lidí se srdečními onemocněními, $\beta = -0,2$, $t(495) = -146,5$, $SE = 0.00137$, $p < .001$.

Literatura

Denollet, J., & Van Heck, G. L. (2001). Psychological risk factors in heart disease: what Type D personality is (not) about. *Journal of psychosomatic research*, 51(3), 465-468.

Williams, P. T. (2001). Physical fitness and activity as separate heart disease risk factors: a meta-analysis. *Medicine and science in sports and exercise*, 33(5), 754.

Wankel, L. M. (1993). The importance of enjoyment to adherence and psychological benefits from physical activity. *International Journal of Sport Psychology*, 24(2), 151-169.

Rondina, R. D. C., Gorayeb, R., & Botelho, C. (2007). Psychological characteristics associated with tobacco smoking behavior. *Jornal Brasileiro de Pneumologia*, 33, 592-601.

Gallucci, G., Tartarone, A., Lerose, R., Lalinga, A. V., & Capobianco, A. M. (2020). Cardiovascular risk of smoking and benefits of smoking cessation. *Journal of thoracic disease*, 12(7), 3866.

Data a další informace o této zprávě jsou dostupné na adrese: https://www.scribbr.com/wp-content/uploads//2020/02/heart.data_.zip