

Predikce celkového počtu najetých kilometrů

V oblasti dopravní psychologie představuje objem najetých kilometrů významný ukazatel expozice riziku. Vyšší kilometrická expozice zvyšuje pravděpodobnost kontaktu s dopravními situacemi, které mohou vést k přestupkům či nehodám (Elvik, 2006). Zároveň však celkový počet najetých kilometrů není pouze indikátorem rizika, ale také zkušenosti, mobility a životního stylu.

Z hlediska statistického modelování je důležité, že proměnné typu „celkový počet kilometrů“ obvykle vykazují silně pravostranné (pozitivně zešikmené) rozdělení. Většina osob uvádí relativně nízké hodnoty, zatímco menší část extrémně vysoké hodnoty. Takové rozdělení porušuje předpoklad normality reziduí lineární regrese (Osborne, 2010).

Jedním ze standardních řešení je logaritmická transformace závislé proměnné. Pokud je původní proměnná kladná a přibližně log-normálně rozdělená, pak logaritmus této proměnné může být modelován pomocí lineární regrese (Long & Freese, 2014). Tento přístup se označuje jako log-normální regrese.

Koeficienty lze následně interpretovat jako přibližné procentuální změny v původní škále (Wooldridge, 2016).

V kontextu řízení lze očekávat, že:

- věk bude pozitivně souviset s kumulativní kilometrickou expozicí,
- muži budou vykazovat vyšší mobilitu než ženy (Rosenbloom, 2006),
- vyšší počet přestupků bude spojen s vyšší expozicí (více příležitostí k porušení pravidel),
- vztah mezi chybami a expozicí může být komplexnější.

Cílem analýzy bylo ověřit, zda počet chyb a přestupků predikuje celkový počet najetých kilometrů při kontrole věku a pohlaví.

Analýza dat a její výsledky

Analyzovaný soubor tvořilo 2669 respondentů (po vyřazení nulových a chybějících hodnot).

Závislá proměnná

- km_celkem (logaritmicky transformováno)

Prediktory

- HS_chyby
- HS_prestupky
- věk
- pohlaví (0 = žena, 1 = muž)

Distribuce proměnné km_celkem byla nejprve analyzována z hlediska normality, přičemž byla zjištěna výrazná pravostranná šikmost, která odůvodňovala použití logaritmické transformace. Následně byl odhadnut lineární regresní model na logaritmicky transformované závislé proměnné ($\ln(\text{km_celkem})$) a regresní koeficienty byly interpretovány jako aproximace procentuálních změn v původní škále.

Pro analýzu byl použit log-normální regresní model. Závislá proměnná, celkový počet najetých kilometrů (km_celkem), vykazovala výraznou pravostrannou šikmost, a proto byla transformována pomocí přirozeného logaritmu. Následně byla provedena vícenásobná lineární regrese s logaritmicky transformovanou závislou proměnnou ($\ln(\text{km_celkem})$). Do modelu byly zahrnuty prediktory počet chyb (HS_chyby), počet dopravních přestupků (HS_prestupky), věk a pohlaví (0 = žena, 1 = muž). Tento přístup umožňuje modelovat logaritmus závislé proměnné

při zachování předpokladů lineární regrese a výsledné koeficienty lze interpretovat jako aproximace procentuálních změn v původní škále.

Distribuce `km_celkem` vykazovala výraznou pravostrannou šikmost. Po logaritmické transformaci se rozdělení významně aproximovalo normálnímu tvaru, což splňuje předpoklady lineární regrese.

Tabulka 1: *Log-normální regrese predikující $\ln(\text{km_celkem})$ ($N = 2669$)*

Proměnná	B	SE	β
Konstanta	5,583***	0,124	—
HS_chyby	-0,034***	0,004	-0,110
HS_prestupky	0,058***	0,003	0,253
Věk	0,141***	0,004	0,566
Pohlaví (muž = 1)	0,971***	0,064	0,183

$R^2 = 0,461$

Adj. $R^2 = 0,460$

$F(4, 2664) = 570,00***$

Pozn. * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

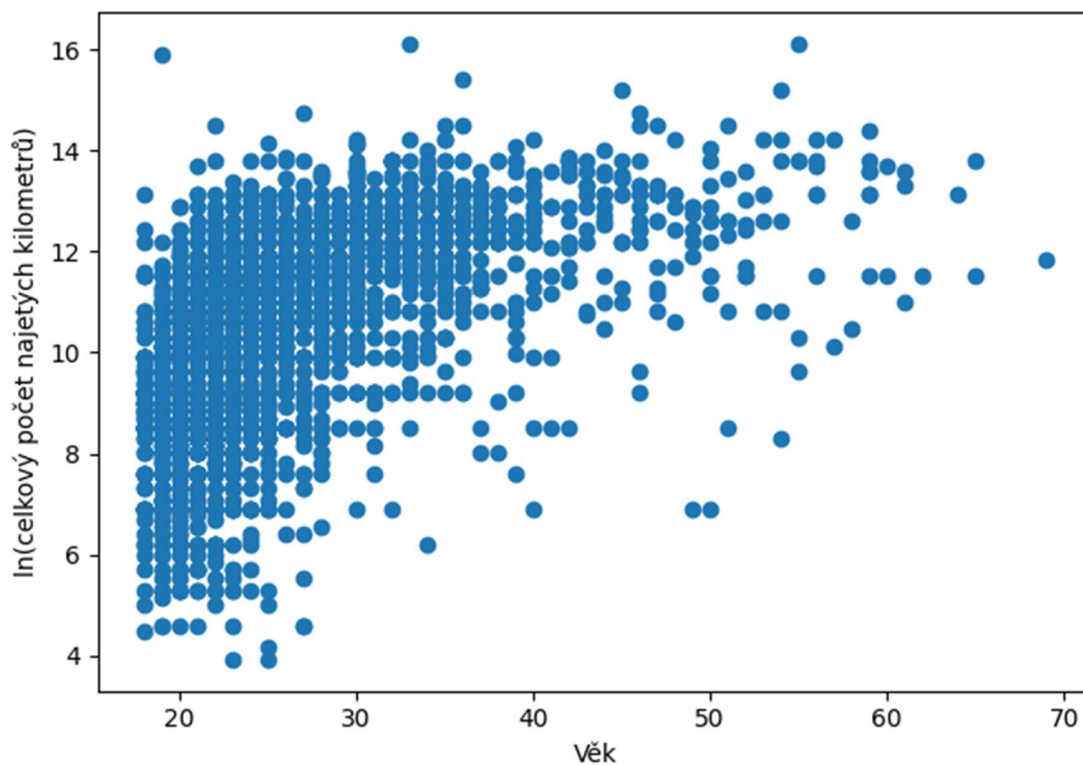
Model je statisticky významný. Model vysvětluje 46,1 % variability logaritmu celkového počtu najetých kilometrů, což představuje relativně vysokou míru vysvětlené variance v kontextu behaviorálních dat.

Regresní koeficienty lze interpretovat jako aproximace procentuálních změn v původní škále. Koeficient věku ($B = 0,141$) naznačuje, že s každým dalším rokem věku se očekávaný počet najetých kilometrů zvyšuje přibližně o 14 %. Koeficient

dopravních přestupků ($B = 0,058$) odpovídá přibližně 5,8% nárůstu kilometrické expozice při zvýšení počtu přestupků o jednu jednotku. Negativní koeficient chyb ($B = -0,034$) naznačuje přibližně 3,4% pokles očekávaného počtu kilometrů při zvýšení počtu chyb o jednu jednotku. Koeficient pohlaví ($B = 0,971$) ukazuje, že muži mají v průměru přibližně 2,6krát vyšší celkový počet najetých kilometrů než ženy.

Grafická vizualizace vztahu mezi věkem a logaritmem celkového počtu najetých kilometrů je uvedena na Obrázku 1. Graf naznačuje pozitivní vztah mezi věkem a kilometrickou expozicí, který byl potvrzen také regresním modelem.

Obrázek 1: *Vztah mezi věkem a logaritmem celkového počtu najetých km*



Diskuse

Výsledky analýzy naznačují, že kumulativní kilometrická expozice je nejsilněji spojena s věkem respondentů. Starší řidiči mají logicky více příležitostí k řízení v průběhu života, což vede k vyššímu celkovému počtu najetých kilometrů. Tento výsledek je konzistentní s teorií dopravní expozice, podle které je pravděpodobnost účasti na dopravních událostech úzce spojena s množstvím absolvované jízdy (Elvik, 2006).

Významný vliv se ukázal také u pohlaví. Muži vykazovali výrazně vyšší kilometrickou expozici než ženy, což odpovídá dřívějším výzkumům dopravního chování, které dlouhodobě poukazují na vyšší mobilitu mužů (Rosenbloom, 2006). Tento rozdíl může souviset s rozdílnými pracovními rolami, životním stylem i preferencemi v oblasti mobility.

Pozitivní vztah mezi počtem dopravních přestupků a kilometrickou expozicí lze interpretovat v rámci teorie expozice riziku. Vyšší počet najetých kilometrů zvyšuje pravděpodobnost setkání s dopravními situacemi, ve kterých může dojít k porušení pravidel silničního provozu. Jinými slovy, řidiči s vyšší mobilitou mají jednoduše více příležitostí k dopravním přestupkům.

Naopak negativní vztah mezi počtem chyb a kilometrickou expozicí může odrážet rozdíly v řidičských zkušenostech. Řidiči s vyšším počtem najetých kilometrů mohou mít větší zkušenost a jistotu v řízení, což se může projevit nižším počtem subjektivně vnímaných chyb. Alternativně může tento vztah odrážet rozdíly ve vnímání vlastních řidičských schopností.

Celkově model vysvětlil přibližně 46 % variability logaritmu celkového počtu najetých kilometrů, což představuje relativně vysokou míru vysvětlené variance v kontextu behaviorálních dat. Výsledky zároveň potvrzují, že při analýze silně pravostranně zešikmených proměnných je logaritmická transformace vhodným postupem, který umožňuje splnit předpoklady lineární regrese. Výsledky je však třeba interpretovat s ohledem na korelační povahu dat, která neumožňuje vyvozovat kauzální vztahy mezi analyzovanými proměnnými.

Závěr

Log-normální regrese se ukázala jako metodologicky adekvátní postup pro modelování silně zešikmené závislé proměnné. Analýza prokázala významný vliv věku, pohlaví a dopravních přestupků na celkový počet najetých kilometrů.

Výsledky podtrhují význam správné volby statistického modelu při práci s ne-normálně distribuovanými daty a demonstrují praktickou aplikaci logaritmické transformace v behaviorálním výzkumu.

Zdroje

Elvik, R. (2006). Laws of accident causation. *Accident Analysis & Prevention*, 38(4), 742–747.

Long, J. S., & Freese, J. (2014). *Regression models for categorical dependent variables using Stata* (3rd ed.). Stata Press.

Osborne, J. W. (2010). Improving your data transformations. *Practical Assessment, Research & Evaluation*, 15(12).

Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach* (6th ed.). Cengage Learning.