

Domácnosti bez přístupu k automobilu - významné prediktory.

Přístup k automobilu je pro mnoho domácností klíčový v mnoha aspektech každodenního života. Zdá se, že vlastnictví automobilu může být přímo vázané se socioekonomickým statusem (Lucas et al., 2016). Domácnosti, které automobil nemají, patří zpravidla k těm nejchudším (Lansley, 2016). Bergstad et al. (2011) argumentují, že vlastnictví auta je spojeno se symbolickým projevem sociálního statusu a ekonomické nezávislosti.

Naproti tomu výsledky jiné studie naznačují, že drtivá většina amerických domácností bez přístupu k automobilu žije v bezprostřední blízkosti veřejné dopravy (Tomer, 2011). Governing (2017) dokonce uvádí seznam velkých měst s nízkým podílem domácností s automobily, kterým vévodí New York City s více než polovičním zastoupením takových domácností.

Tyto poznatky a případné další významné prediktory jsme ověřovali pomocí logistické regrese na základě amerických dat "GSS Data Survey". Tyto data umožňují komplexní vhled do různých částí sociální problematiky. K analýze byl použit edukační program univerzity Berkeley (2023) pro GSS data, který umožňuje přímou analýzu data setu. Odkaz na program zde: <https://sda.berkeley.edu/sdaweb/analysis/?dataset=gss21rel3>

Jedná se o modelování logistické regrese. Představme si uvažované regresory a kovariáty.

Závislá proměnná

oth5 - Přístup k automobilu (0: ne, 1: ano)

Regresory

xnorsiz - Velikost místa bydliště (1: město nebo předměstí velkého města; 2: středně velká a malá města; 3: vesnice a malé celky)

income - Příjem domácnosti (0: méně než 25.000 \$ / rok, 1: více než 25.000 \$ / rok)

Kovariáty

pray - Náboženství (0: modlí se méně než jednou týdně; 1: modlí se alespoň jednou týdně)

sex - Pohlaví (0: žena; 1: muž)

race - Rasa (1: Jiná než afroamerická; 2 :afroamerická)

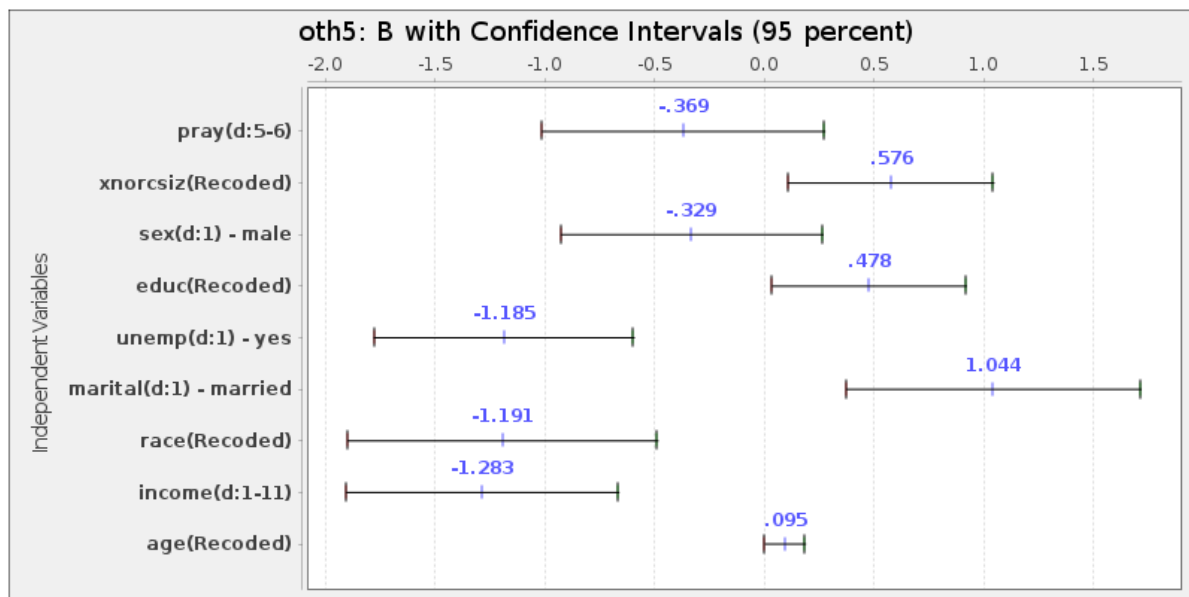
age - Věk (1 : 18-30 let; 5 = 31-50 let; 10 = 51-90 let)

marital - Manželství (0: ne; 1: ano)

educ - vzdělání (1: základní; 2: středoškolské; 3: vysokoškolské)

unemp - nezaměstnanost v poslední době (0: ne; 1: ano)

V tomto výchozím modelu jsme jako doplňující kovariáty použili základní demografické údaje o respondentovi, které mohou mít vliv na závislou proměnnou. Po pečlivém překódování těchto kovariát jsme se dostali k výsledkům logistické regrese. Použité kódování je vysvětleno výše a bylo otestováno na více způsobech grupování.



Grafické znázornění výsledků prvního modelu.

Na první pohled je patrné, že v modelu jsou na statistické hladině významnosti 0.05 význačné jak proměnné spjaté se statusem a úrovní života (příjem, manželství, nezaměstnanost, vzdělání), tak i velikost města, ve kterém respondent žije. K zamyšlení je, že rasa respondenta je v tomto ohledu rovněž signifikantní. Naopak náboženství a překvapivě také věk a pohlaví se zdají být poněkud nevýznamné. To může být způsobeno tím, že v domácnostech zpravidla žijí generace a pohlaví dohromady. Než přistoupíme k finálním závěrům a interpretaci výsledků, je dobré se zamyslet nad korelací některých proměnných.

Correlation Matrix									
	pray(d:1-4)	xnorcsiz(d:1) - a large central city (over 250,000)	sex(d:1) - male	educ(Recoded)	unemp(d:1) - yes	marital(d:1) - married	race(Recoded)	income(d:12) - \$25,000 or more	age(Recoded)
pray(d:1-4)	1.00	.00	-.21	-.03	-.07	.07	.12	-.04	.14
xnorcsiz(d:1) - a large central city (over 250,000)	.00	1.00	-.01	.02	.05	-.11	.15	-.05	-.05
sex(d:1) - male	-.21	-.01	1.00	-.01	.04	.03	-.04	.08	.00
educ(Recoded)	-.03	.02	-.01	1.00	-.03	.03	-.05	.32	-.09
unemp(d:1) - yes	-.07	.05	.04	-.03	1.00	-.15	.07	-.15	-.26
marital(d:1) - married	.07	-.11	.03	.03	-.15	1.00	-.15	.26	.21
race(Recoded)	.12	.15	-.04	-.05	.07	-.15	1.00	-.11	-.03
income(d:12) - \$25,000 or more	-.04	-.05	.08	.32	-.15	.26	-.11	1.00	.07
age(Recoded)	.14	-.05	.00	-.09	-.26	.21	-.03	.07	1.00

Korelační matice nezávislých proměnných.

Z korelační matice lze vyzorovat přítomnost několika slabých korelací mezi proměnnými, jako například mezi věkem/příjmem a manželstvím nebo vzděláním a příjmem. Po zvážení jsme se rozhodli odstranit proměnné věk a pohlaví z důvodu jejich nízké signifikance. Bez výrazné ztráty vysvětlené variance v modelu jsme tak přešli k menšímu modelu bez těchto kovariát. Při pokusu o odstranění dalších kovariát již model ztrácel vysvětlující hodnotu výrazněji.

Logit Coefficients				Test That Each Coefficient = 0	
	B	SE(B)	Exp(B)	T-statistic	Probability
educ(Recoded)	.121	.042	1.129	2.885	.004
xnorcsiz(Recoded)	.613	.236	1.846	2.594	.010
pray(d:5-6)	-.469	.323	.626	-1.451	.147
unemp(d:1) - yes	-1.322	.298	.267	-4.431	.000
marital(d:1) - married	1.177	.333	3.243	3.539	.000
race(Recoded)	-1.119	.356	.327	-3.139	.002
income(d:12) - \$25,000 or more	1.145	.315	3.143	3.637	.000
Constant	1.402	.763	4.064	1.837	.067

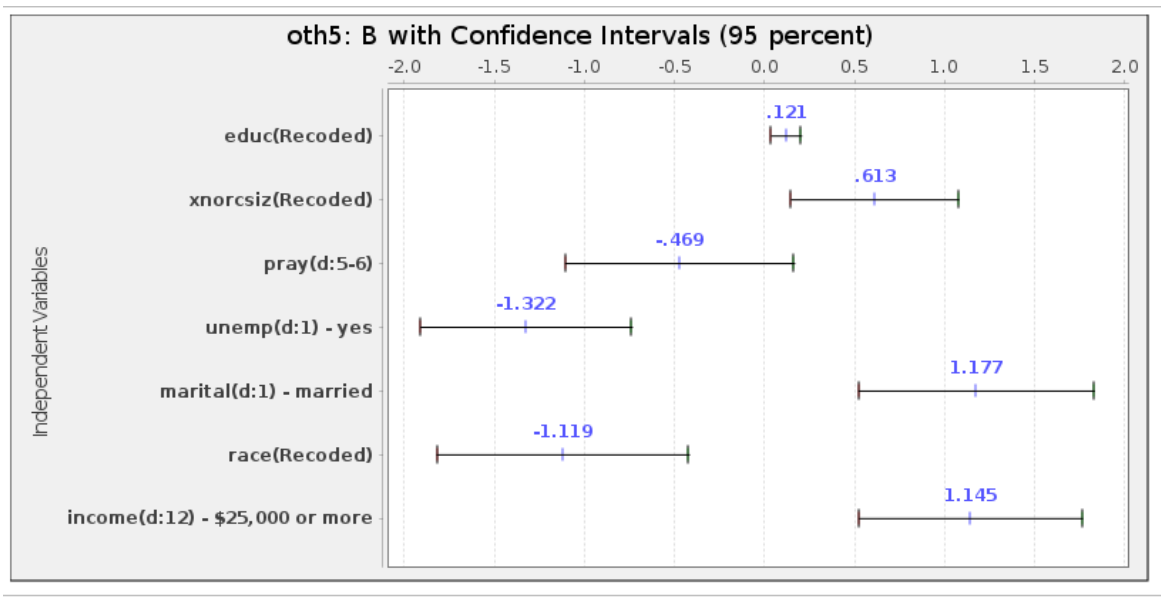
Color coding:	<-2.0	<-1.0	<0.0	>0.0	>1.0	>2.0	T
Effect of each variable:	Smaller than average		Larger than average				

Log Likelihood = -165.912 Pseudo R-sq = .242

Logit Coefficients			Diff. in prob. that Dep Var = 1 with Indep Vars (X's) at means	
	B	SE(B)	for 1 Unit of X	for 1 Std Dev of X
educ(Recoded)	.121	.042	.004	.012
xnorcsiz(Recoded)	.613	.236	.017	.013
pray(d:5-6)	-.469	.323	-.022	-.008
unemp(d:1) - yes	-1.322	.298	-.093	-.030
marital(d:1) - married	1.177	.333	.027	.017
race(Recoded)	-1.119	.356	-.071	-.015
income(d:12) - \$25,000 or more	1.145	.315	.026	.015
Constant	1.402	.763		

Log Likelihood = -165.912 Pseudo R-sq = .242

Výsledky finálního modelu.



Grafické zobrazení výsledků finálního modelu.

Výše uvedené výsledky naznačují, že jsme úspěšně replikovali poznatky výše uvedených studií na zvolených datech. Prokázali jsme, že výše příjmu je statisticky významným regresorem na hladině významnosti 0,05. Při přechodu ze skupiny s nižším příjmem než 25.000 \$ do skupiny s vyšším příjmem se šance na přístup domácnosti k vozidlu zvyšuje $\exp(1,145)$, tedy přibližně 3,14krát, za předpokladu, že všechny ostatní proměnné jsou konstantní. Podobně, nezaměstnaný jedinec má přibližně 4krát nižší šanci na přístup k automobilu. S narůstající velikostí města se pravděpodobnost přístupu k vozidlu sníží 1,83x při každém přechodu z vesnice směrem k větším sídlům.

Může se zdát, že jednou z nevýhod modelu může být poměrně nízká schopnost vysvětlení variance dat. Je ovšem nerealistické očekávat vysoké hodnoty vysvětlení variance v případě závislé proměnné, u které většina respondentů náleží do jedné ze dvou skupin, v tomto případě do skupiny domácností s přístupem k automobilu.

Literatura:

Bergstad, C. J., Gamble, A., Hagman, O., Polk, M., Gärling, T., & Olsson, L. E. (2011). Affective–symbolic and instrumental–independence psychological motives mediating effects of socio-demographic variables on daily car use. *Journal of Transport Geography*, 19(1), 33-38. doi: 10.1016/j.jtrangeo.2009.11.006

Berkeley (2023). Release 3b of the 1972-2021 General Social Survey (GSS) Cumulative Dataset has been added to the SDA Archive. <https://sda.berkeley.edu/sdaweb/analysis/?dataset=gss21rel3>

Governing (2017). Retrieved from <https://www.governing.com/archive/gov-car-free-households-cities.html>

Lansley G. (2016). Cars and socio-economics: understanding neighbourhood variations in car characteristics from administrative data. *Regional Studies, Regional Science*, 3(1), 264-285. doi: 10.1080/21681376.2016.1177466

Lucas, K., Bates, J., Moore, J., & Carrasco, J. A. (2016). Modelling the relationship between travel behaviours and social disadvantage. *Transportation Research Part A: Policy and Practice*, 85, 157-173. doi: 10.1016/j.tra.2016.01.008

Tomer, A. (2011). *Transit access and zero-vehicle households*. Metropolitan Policy Program at Brookings. Retrieved from https://www.brookings.edu/wp-content/uploads/2016/06/0818_transportation_tomer.pdf