نردبان بازتولیدپذیری با برآورد دوبارهٔ طراحی‌آگاه برای پژوهش‌های سفر مبتنی بر پیمایش: مطالعهٔ موردی بیست‌ودو مطالعه از پیمایش ملی سفر خانوار آمریکا

سید صابر ناصرعلوی
گروه مهندسی عمران، دانشگاه شهید باهنر کرمان، کرمان ۷۶۱۶۹۱۴۱۱۱، ایران
saber_naseralavi@uk.ac.ir

چکیده

برنامه‌ریزان حمل‌ونقل از ضرایب رگرسیون و امتیازهای پیش‌بینی پیمایش‌های ملی سفر برای ارزیابی اثر سواری‌خواهی اینترنتی، حمل‌ونقل عمومی یا خودروهای برقی بر الگوی سفر بهره می‌گیرند. این عددها به انتخاب‌هایی بستگی دارند که مقاله‌ها کمتر کامل شرح می‌دهند: کدام رکوردها نگه داشته شد، آیا وزن‌های پیمایش به کار رفت، عدم‌قطعیت چگونه محاسبه شد و مدل پیش‌بینی چگونه آزموده شد. ما نردبانی چهارپله‌ای برای بازتولیدپذیری پیشنهاد می‌کنیم (بازسازی نمونه، بازسازی جدول‌های توصیفی، برازش مدل منتشرشده، و برازش آن با طرح نمونه‌گیری خود پیمایش) و آن را بر ۲۲ مطالعهٔ منتشرشده از پیمایش ملی سفر خانوار آمریکا (NHTS) به کار می‌بریم؛ فایل‌های عمومی اجازهٔ برازش دوبارهٔ مدل ۱۳ مطالعه را داد. نمونهٔ ۹ مطالعه خوب بازسازی شد و ۸۱ درصد از ۱٬۴۵۴ مقدار توصیفی چاپ‌شده بازتولید شد. مطالعه‌هایی که وزن‌های پیمایش را به کار برده بودند، تنها با وزن بازتولید شدند: در یک مطالعه، نسبت‌های بخت وزنی ۵۷ درصد از مقدارهای چاپی را بازتولید کرد و نسبت‌های بدون وزن ۸ درصد. وقتی وزن‌های تکرارِ همراه پیمایش به کار رفت، خطای معیار به‌طور میانه ۲٫۲ برابر خطای معیار بدون وزن بود؛ ۳۲ درصد از ضریب‌های معنادار در سطح پنج درصد معناداری خود را از دست دادند و ۸ درصد ضریب‌ها میان برازش بدون وزن و وزنی علامتشان عوض شد. هم‌خوشگی اشخاصِ یک خانوار، نه فرمول وزن‌های تکرار، بیشتر این تفاوت را توضیح داد. برای مدل‌های طبقه‌بندی، بیش‌نمونه‌گیری از رستهٔ کمیاب پیش از تقسیم داده، سطح زیر منحنی مشخصهٔ عملکرد گیرنده (ROC)، امتیازی که شانس را پنجاه درصد و کمال را صد درصد می‌داند، را برای یک جنگل تصادفی از ۶۴ درصد به ۹۹ درصد رساند، و امتیاز رگرسیونیِ چاپی (سهم واریانس توضیح‌داده‌شده) ۷۱ درصد روی خانوارهای کنار گذاشته‌شده تنها ۲۰ درصد بازتولید شد. این ابزار به‌صورت دو ماژول از بستهٔ متن‌باز پایتون pynhts منتشر شده است.

واژه‌های کلیدی: پیمایش ملی سفر خانوار؛ بازتولیدپذیری؛ وزن‌های پیمایش؛ واریانس با وزن‌های تکرار؛ نشت داده؛ رگرسیون دوجمله‌ای منفی

مقدمه

پیمایش ملی سفر خانوار آمریکا (NHTS) مهم‌ترین منبع عمومی دربارهٔ چگونگی سفر آمریکایی‌هاست. موج‌های ۲۰۰۱، ۲۰۰۹، ۲۰۱۷ و ۲۰۲۲ آن هر کدام میان حدود هفتاد هزار تا صد و پنجاه هزار خانوار را دربرمی‌گیرند و برای هر عضو، سفرهای یک روز تعیین‌شده، خودروها و فهرست بلندی از ویژگی‌های خانوار و فرد ثبت شده است. پژوهشگران از این فایل‌ها برای برآورد این‌که چه کسی سوار می‌شود، چه کسی هزینه می‌دهد و چه کسی آلایندگی ایجاد می‌کند استفاده می‌کنند و برنامه‌ریزان ضرایب رگرسیون، نسبت‌های بخت و امتیازهای پیش‌بینیِ حاصل را شاهدی دربارهٔ سواری‌خواهی اینترنتی، دوچرخهٔ اشتراکی، حمل‌ونقل عمومی، خودروهای برقی و خودروکیلومتر پیموده می‌خوانند (Alberini و همکاران, 2021, Conway و همکاران, 2018, Wu و MacKenzie, 2021). جمله‌ای مانند «هر بار استفادهٔ بیشتر از سواری‌خواهی در ماه با سفرهای روزانهٔ بیشتر همراه است» جمله‌ای کوتاه است که بر زنجیره‌ای بلند از تصمیم‌ها تکیه دارد: کدام پاسخ‌گویان وارد

تحلیل شدند، آیا وزن‌های پیمایش اعمال شد، خطای معیار چگونه ساخته شد و، برای مدل پیش‌بینی، چگونه روی داده‌ای که ندیده بود آزموده شد.

این تصمیم‌ها از آن رو اهمیت دارند که NHTS نمونهٔ تصادفی ساده نیست. خانوارها در لایه‌ها انتخاب می‌شوند، اشخاص درون خانوار خوشه‌ای‌اند، نرخ پاسخ در گروه‌ها فرق دارد، و به همین دلیل فایل‌ها وزن فرد و خانوار و، برای موج‌های ۲۰۰۱، ۲۰۰۹ و ۲۰۱۷، میان نود و هشت تا صد وزن تکرار برای هر رکورد دارند تا تحلیل‌گر بتواند برآوردی از واریانس که طرح نمونه‌گیری را رعایت کند به دست آورد (Rust و Rao, 1996, Wolter, 2007). این‌که رگرسیون وزن‌دار شود یا نه پرسشی دیرینه در آمار پیمایشی است: وزن‌ها وقتی انتخاب به پیامد وابسته باشد می‌توانند اریبی را بردارند، اما وقتی وابسته نباشد کارایی را کاهش می‌دهند (Bollen و همکاران, 2016, Gelman, 2007, Solon و همکاران, 2015, Winship و Radbill, 1994). این‌که هنگام محاسبهٔ خطای معیار از طرح نمونه‌گیری استفاده شود کمتر محل بحث است: نادیده‌گرفتن خوشه‌بندی و لایه‌بندی عدم‌قطعیت را کمتر از واقع نشان می‌دهد (Lumley, 2004). نمونه‌هایی از پیمایش‌های سلامت نشان می‌دهند ارتباطی که در تحلیل بدون وزن معنادار است ممکن است با رعایت وزن و طرح معنادار نباشد (Birrell و همکاران, 2019, Gard و همکاران, 2023). این‌که این اتفاق در پژوهش حمل‌ونقل، که رگرسیون روی داده‌های NHTS در آن امری عادی است، تا چه اندازه رخ می‌دهد هنوز اندازه‌گیری نشده است.

نگرانی دوم به بازتولیدپذیری در معنای وسیع‌تر برمی‌گردد. بازآفرینی نتیجه‌های محاسباتی از روی شرح منتشرشده اغلب دشوار است (Hardwicke و همکاران, 2021, Peng, 2011, Stodden و همکاران, 2018)، و وقتی تحلیل‌گران مستقل داده و پرسش یکسانی دریافت می‌کنند، انتخاب‌های موجه اما متفاوت می‌تواند به نتیجه‌های متفاوت و گاه در جهت‌های مخالف بینجامد (Schweinsberg و همکاران, 2021, Silberzahn و همکاران, 2018, Simonsohn و همکاران, 2020, Steegen و همکاران, 2016). مدل‌های پیش‌بینی لایه‌ای دیگر می‌افزایند. یادگیری ماشین با این ادعا که مدل‌های انعطاف‌پذیر از مدل‌های کلاسیک بهترند وارد پژوهش رفتار سفر شده است (Hagenauer و Martín-Baos, 2017, Helbich و همکاران, 2023, Wang و همکاران, 2024)، در حالی که ادبیات گسترده‌تر نشان می‌دهد نشت میان داده‌های آموزش و آزمون در بسیاری از رشته‌ها امتیازهای خوش‌بینانه می‌سازد (Kapoor و همکاران, 2024, Kapoor و Narayanan, 2023, Rosenblatt و همکاران, 2024). دو نشت اینجا ویژه اهمیت دارد. ساختن نمونه‌های مصنوعی از رستهٔ کمیاب پیش از تقسیم داده، نسخه‌هایی از ردیف‌های آزمون را وارد آموزش می‌کند و امتیازها را به‌شدت بالا می‌برد (Blagus و Lusa, 2013, Chawla و همکاران, 2002, Demircioğlu, 2024). تقسیم تصادفی ردیف‌ها وقتی ردیف‌ها خانوار، مکان یا فرد مشترک دارند، شبه‌نسخه‌ها را در دو سوی تقسیم قرار می‌دهد و طرح‌های اعتبارسنجی که این ساختار را رعایت می‌کنند امتیازهای پایین‌تر و صادقانه‌تر می‌دهند (Ploton و همکاران, 2020, Roberts و همکاران, 2017, Vabalas و همکاران, 2019). ردیف‌های NHTS دقیقاً از این نوع‌اند: اشخاص یک خانوار تقریباً همه چیز مهم را با هم دارند.

کارهای دربارهٔ خود NHTS بیشتر کیفیت داده را بررسی کرده‌اند، برای نمونه اریبی انتخاب میان پاسخ‌گویانی که پذیرفتند در پیمایش‌های پیگیری شرکت کنند (Wang و همکاران, 2022)، یا از آن برای پاسخ به پرسش‌های محتوایی بهره برده‌اند (Conway و همکاران, 2018). خط‌لوله‌های باز برای تقاضای سفر مصنوعی از پیمایش‌های ملی دیگر هم وجود دارد (Hörl و Balac, 2021). ما مطالعه‌ای نیافتیم که مجموعه‌ای از تحلیل‌های منتشرشدهٔ NHTS را بردارد، نمونه‌ها و جدول‌هایشان را بازسازی کند، مدل‌هایشان را دوباره برازش دهد و بپرسد آیا نتیجه‌هایشان از طرح نمونه‌گیری خود پیمایش و از اعتبارسنجی بدون نشت جان به‌در می‌برد. این شکافی دوگانه است: اندازهٔ اثر نادیده‌گرفتن طرح در رگرسیون‌های حمل‌ونقل ناشناخته است و میزان وابستگی امتیازهای پیش‌بینیِ منتشرشده به خط‌لولهٔ اعتبارسنجی نیز ناشناخته است.

این مقاله شکاف را با یک روش و یک مطالعهٔ موردی پر می‌کند. روش، نردبانی چهارپله‌ای برای بازتولیدپذیری است که هر پله‌اش با عددهای چاپی درجه‌بندی می‌شود، همراه با برآورد دوبارهٔ طراحی‌آگاه که سه شیوهٔ برازش یک مدل را مقایسه می‌کند و امتیازدهی دوبارهٔ طبقه‌بندها در چهار خط‌لوله. مطالعهٔ موردی آن را بر ۲۲ مطالعهٔ منتشرشده از NHTS، که به‌صورت مجموعه‌ای در دسترس گردآوری شده، به‌کار می‌برد و فایل‌های عمومی اجازهٔ برازش دوبارهٔ ۱۳ مطالعه را داد. چهار پرسش داریم: چه مقدار از هر مطالعه را می‌توان بازسازی کرد؛ آیا عددهای چاپی با برآوردگرِ اعلام‌شدهٔ مقاله بازتولید می‌شوند؛

نتیجه‌ها زیر طرح نمونه‌گیری چقدر تغییر می‌کنند؛ و امتیاز پیش‌بینی تا چه اندازه به خط لوله وابسته است. همین ابزار به‌صورت دو ماژول از بستهٔ متن‌باز pynhts منتشر شده تا بررسی به‌جای چند هفته در چند خط کد انجام شود.

مواد و روش‌ها

مجموعهٔ مطالعه‌های موردی و داده

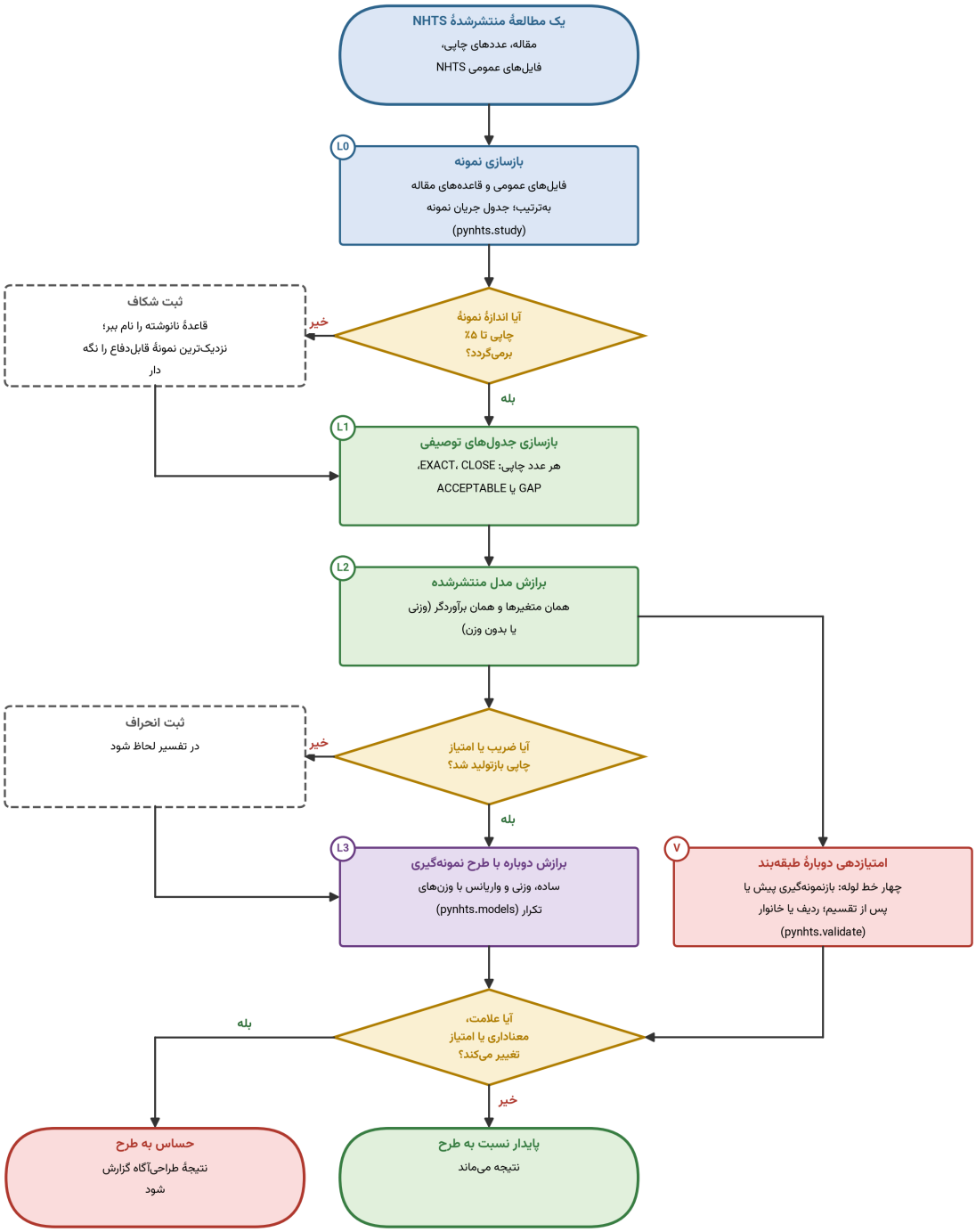
مجموعه از ۲۲ مطالعهٔ منتشرشده میان سال‌های ۲۰۱۱ تا ۲۰۲۵ تشکیل شده که NHTS را تحلیل می‌کنند و عدد کافی برای مقایسه چاپ کرده‌اند (جدول 1). این مطالعه‌ها (Alberini و همکاران, 2021؛ Barua و همکاران, 2021؛ Deka و Fei, 2019؛ Fu و همکاران, 2020؛ Jia, 2019؛ Jia و همکاران, 2020؛ Jiao و همکاران, 2020؛ Krizek و McGuckin, 2019؛ Lidbe و همکاران, 2021؛ Naseralavi و همکاران, 2025؛ Quinn و همکاران, 2016؛ Sabouri و همکاران, 2020؛ Sadeghvaziri و Shirgaokar, 2020؛ Tawfik و Su, 2021؛ Nobler و همکاران, 2021؛ Wu و MacKenzie, 2021؛ Xue و همکاران, 2021؛ Yang و همکاران, 2011؛ Zhang و Zhou; Zhang, 2018a, 2018b, 2018c؛ Wang, 2014) هنگام ساختن بستهٔ pynhts گردآوری شدند، بنابراین مجموعه‌ای در دسترس است و نمونهٔ سامانمند از ادبیات نیست؛ دربارهٔ سهم همهٔ مقاله‌های NHTS که چنین رفتاری دارند نتیجه‌ای نمی‌گیریم. یکی از مطالعه‌ها از نویسندهٔ نخست و توسعه‌دهندهٔ بسته است؛ در فهرست آمده اما دوباره برآورد نشد و مدل‌های آن برای مقاله‌ای جدا نگه داشته شده است. فایل‌های عمومی NHTS برای موج‌های ۲۰۰۱، ۲۰۰۹ و ۲۰۱۷ از پایگاه پیمایش و از راه pynhts دریافت شد که هر فایل را با کدنامه‌اش راستی‌آزمایی می‌کند. هر ۲۲ مطالعه از موج ۲۰۰۱، ۲۰۰۹ یا ۲۰۱۷ استفاده می‌کنند.

جدول 1: بیست‌ودو مطالعهٔ موردی و پله‌ای که هر یک به آن رسید. L۱ شمار عددهای توصیفی چاپی درجه‌بندی‌شده و در پرانتز سهم عددهای EXACT یا CLOSE را می‌دهد.

شناسه	مطالعه	موج	مدل	نمونه (L0)	عددهای چاپی منطبق (L1)	مدل (L2)
P01	Yang ۲۰۱۱	۲۰۰۱	logit	کامل	۱۵۴ (۹۵٪)	برازش شد
P02	Zhou ۲۰۱۴	۲۰۰۹	SEM	جزئی	۲۳ (۴۸٪)	برازش شد
P03	Quinn ۲۰۱۶	۲۰۰۹	logit	جزئی	۲۱ (۴۸٪)	برازش شد
P05	Deka ۲۰۱۹	۲۰۱۷	ZINB	جزئی	۸۶ (۷۰٪)	اجرا نشد
P06	Jia ۲۰۱۹	۲۰۱۷	ML classifiers	کامل	۴۲ (۸۶٪)	برازش شد
P07	Krizek ۲۰۱۹	۲۰۱۷	descriptive	جزئی	۱۸ (۲۲٪)	اجرا نشد
P09	Fu ۲۰۲۰	۲۰۱۷	simulation	ناکافی	۱۳ (۸٪)	اجرا نشد
P10	Lidbe ۲۰۲۱	۲۰۰۱/۲۰۱۷	descriptive	کامل	۳۳۶ (۹۰٪)	اجرا نشد
P11	Alberini ۲۰۲۱	۲۰۱۷	OLS	کامل	۵۰ (۹۴٪)	برازش شد
P12	Xue ۲۰۲۱	۲۰۱۷	SEM	ناکافی	۲۰ (۳۵٪)	اجرا نشد
P13	Zhang ۲۰۱۸a	۲۰۱۷	ZINB	جزئی	۸۰ (۶۹٪)	برازش شد
P14	Zhang ۲۰۱۸b	۲۰۱۷	ZINB	کامل	۹۷ (۹۷٪)	برازش شد
P15	Zhang ۲۰۱۸c	۲۰۱۷	ZINB	کامل	۸۵ (۹۶٪)	برازش شد
P16	Sabouri ۲۰۲۰	۲۰۱۷	multilevel Poisson	جزئی	۳۰ (۷۰٪)	اجرا نشد
P17	Jia ۲۰۲۰	۲۰۱۷	ML classifiers	کامل	۴۰ (۸۵٪)	برازش شد
P18	Jiao ۲۰۲۰	۲۰۱۷	NB	کامل	۵۵ (۸۷٪)	برازش شد
P19	Sadeghvaziri ۲۰۲۰	۲۰۰۹/۲۰۱۷	descriptive	جزئی	۱۱ (۸۲٪)	اجرا نشد
P20	Shirgaokar ۲۰۲۱	۲۰۱۷	NB	جزئی	۶۳ (۴۱٪)	برازش شد
P21	Barua ۲۰۲۱	۲۰۰۹/۲۰۱۷	GBM	جزئی	۴۰ (۴۸٪)	برازش شد
P22	Wu ۲۰۲۱	۲۰۱۷	PSM	کامل	۷۱ (۹۷٪)	برازش شد
P23	Su ۲۰۲۱	۲۰۱۷	motif clustering	جزئی	۶۵ (۶۵٪)	اجرا نشد
P24	Naseralavi ۲۰۲۵	۲۰۱۷	ML per state	جزئی	۵۴ (۹۱٪)	اجرا نشد

نردبان بازتولیدپذیری

نردبان چهار پله دارد (شکل 1). **L0**: نمونهٔ تحلیل را از فایل‌های عمومی با اعمال قاعده‌های مقاله به همان ترتیب بازسازی و اندازهٔ نمونه را با عدد چاپی مقایسه کن. **L1**: جدول‌های توصیفی را بازسازی کن و هر مقدار چاپی را EXACT (برابر در دقت چاپی)، CLOSE (تا دو درصد)، ACCEPTABLE (تا پنج درصد) یا GAP درجه‌بندی کن. **L2**: مدل منتشرشده را با همان متغیرها و برآوردگرِ اعلام‌شدهٔ مقاله (وزنی یا بدون وزن) برازش بده و هر ضریب یا امتیاز چاپی را به همان شیوه درجه‌بندی کن، با رواداری دست‌کم یک رقم آخر چاپی. **L3**: مدل را زیر طرح نمونه‌گیری دوباره برازش بده. هیچ قاعده‌ای برای رسیدن به عدد چاپی تغییر نکرد: تفاوت به‌صورت شکاف ثبت می‌شود و طبیعی‌ترین خوانش یک قاعدهٔ نانوشته همراه با شاهدی که آن را پشتیبانی می‌کرد مستند می‌شود. پله‌های L0 و L1 پیش از اجرای مدل‌های این مقاله کامل شده بودند و درجه‌بندی‌شان از پروندهٔ ممیزی مخزن بسته بازتولید می‌شود.



شکل 1: روش مقاله. پله‌های L0 تا L3 بر هر مطالعهٔ منتشرشده اعمال می‌شود؛ جعبهٔ V که طبقه‌بند را دوباره امتیاز می‌دهد برای مطالعه‌های یادگیری ماشین است. لوزی‌ها تصمیم‌اند؛ جعبه‌های خطچین آنچه را به مرحلهٔ بعد منتقل می‌شود ثبت می‌کنند.

سه شیوه برای برازش یک مدل

هر رگرسیون به سه شیوه برازش شد. برازش بدون وزن از بیشینهٔ درست‌نمایی با خطای معیار مبتنی بر مدل استفاده می‌کند، که همان است که بیشتر مطالعه‌ها گزارش می‌کنند. برازش وزنی از وزن فرد، خانوار یا فرد-روز مشاهده (مقیاس‌شده به اندازهٔ نمونه برای برآورد) همراه با واریانس ساندویچی بهره می‌برد که در آن مشاهده‌های یک خانوار یک خوشه‌اند. برازش طراحی‌آگاه همان برآورد وزنی را با طرح واریانس موج به کار می‌برد. برای موج‌هایی که وزن تکرار دارند این روش جک‌نایف است: مدل با هر یک از R وزن تکرار w_r دوباره برازش می‌شود و واریانس ضریب $\hat{\beta}$ برابر است با

$$\hat{V}(\hat{\beta}) = f \sum_{r=1}^{R} \left(\hat{\beta}_r - \hat{\beta}\right)^2,$$

با $R = ۹۸$ و $f = 6/7$ برای ۲۰۱۷، $R = ۱۰۰$ و $f = 99/100$ برای ۲۰۰۹ و $R = ۹۹$ و $f = 98/99$ برای ۲۰۰۱، مانند راهنمای کاربر. رگرسیون‌های لجستیک، خطی و پواسون از آغاز گرم دقیقاً دوباره برازش می‌شوند. رگرسیون‌های دوجمله‌ای منفی و دوجمله‌ای منفی صفرافزوده که برازش دوبارهٔ آن‌ها دقیقه‌ها طول می‌کشد، از گام‌های نیوتن از بهینهٔ نمونهٔ کامل بهره می‌گیرند، گامی که با برازش‌های دوبارهٔ دقیق سنجیدیم. درست‌نمایی مدل‌های شمارشی و گرادیان آن‌ها از خود ماست و با تفاضل متناهی و با کتابخانهٔ statsmodels سنجیده شد (Lambert, 1992, Zeileis و همکاران, 2008). برای سنجش مستقل، برازش‌های لجستیک، پواسون و خطی روی بزرگسالان فایل فرد ۲۰۱۷ با بستهٔ survey در R (طرح وزن تکرار با همان ضریب) مقایسه شد: ضریب‌ها تا یک در صد هزار و خطای معیارها تا ۰٫۲ درصد برابر بودند. رژیمی که عددهای چاپی یک مطالعه در آن بازتولید شد برآوردگر اعلام‌شدهٔ آن مطالعه گزارش می‌شود.

مدل‌های مطالعه‌های موردی

هشت مطالعهٔ رگرسیونی منتشرشده دوباره برازش شد: رگرسیون لجستیک پیاده‌روی و دوچرخه‌سواری (دو مطالعه)؛ کمترین مربعات معمولی با اثر ثابت برای سازندهٔ خودرو، ایالت و چگالی (یک)؛ رگرسیون دوجمله‌ای منفی سفرهای روزانه (دو)؛ و رگرسیون دوجمله‌ای منفی صفرافزودهٔ استفادهٔ ماهانه از تحرک اشتراکی (سه). یک مدل معادلات ساختاری با دو گرایش پنهان، همتاسازی امتیاز گرایش گروه‌های کاربر سواری‌خواهی، رگرسیون تقویت گرادیان خرید اینترنتی خانوار و پنج طبقه‌بند پذیرش خودروی سوخت جایگزین و برقی مجموعه را کامل می‌کنند. هر جا جدول ضرایب چاپی وجود دارد هر ضریب چاپی هدف L2 است؛ هر جا فقط یک نتیجهٔ کلیدی چاپ شده (اثر حمل‌ونقل عمومی در دو مطالعهٔ ZINB، کشش در مطالعهٔ کمترین مربعات)، همان نتیجه هدف است و هر جا از درصدی در چکیده به دست آمده در مواد تکمیلی گفته شده است.

مدل‌های یادگیری ماشین، طرح‌های اعتبارسنجی و تنظیم

پنج طبقه‌بند دوباره روی نمونهٔ ۳۰٬۵۱۱ مالک خودروی نو در ۲۷٬۴۱۳ خانوار اجرا شد، با پیامد خودروی سوخت جایگزین (۱٬۲۳۶ AFV، ۳۸۹ EV مالک) یا خودروی برقی EV مالک). قاعدهٔ نمونه (نوترین خودروی هر راننده اصلی، حداکثر دو سال مدل، متغیرهای کلیدی کامل) از اندازهٔ نمونهٔ چاپی بازیابی شد و اندکی کوچک‌تر از نمونهٔ چاپی است. مدل‌ها در جدول 2 توضیح داده شده‌اند. رگرسیون لجستیک جمع وزنی ورودی‌ها را به احتمال تبدیل می‌کند؛ بیز ساده فرض می‌کند هر ورودی در هر رده زنگوله‌ای و مستقل از دیگر ورودی‌هاست؛ درخت تصمیم رشته‌ای از پرسش‌های بله/خیر دربارهٔ ورودی‌ها می‌پرسد؛ جنگل تصادفی رأی درختان بسیاری را که روی بخش‌های تصادفی داده رشد کرده‌اند میانگین می‌گیرد؛ و ماشین بردار پشتیبان با هستهٔ شعاعی رده‌ها را با مرزی هموار که از نزدیک‌ترین نقطه‌های آموزشی ساخته می‌شود از هم جدا می‌کند. خانوادهٔ ششم، تقویت گرادیان، که درخت‌های کوچک را یکی پس از دیگری می‌افزاید تا هر یک خطای مجموع پیشین را اصلاح کند، برای رگرسیون خرید اینترنتی در سطح خانوار همراه با رگرسیون خطی، رگرسیون بردار پشتیبان و جنگل تصادفی به کار رفت. این چهار مدل رگرسیونی تنظیم نشدند: با تنظیمات ثابت و مستند اجرا شدند، همان‌ها که مطالعه چاپ کرده بود (هزینهٔ ده و پهنای هستهٔ یک‌صدم برای رگرسیون بردار پشتیبان، کمینهٔ بیست‌وپنج ردیف در برگ برای جنگل)، زیرا پرسش این بود که آیا با تنظیمات چاپی می‌توان به امتیازهای چاپی نزدیک شد.

جدول 2: پنج طبقه‌بند و فضای جست‌وجوی تنظیمات آن‌ها (برای همهٔ مدل‌ها همان جست‌وجوی تصادفی ده‌آزمایشی).

مدل	ایده به زبان ساده	ابرپارامترهای جست‌وجوشده (ده آزمایش تصادفی برای هر مدل)
LR	رگرسیون لجستیک: جمع وزنی ورودی‌ها که به احتمال تبدیل می‌شود	قدرت جریمه C (۰٫۰۰۱ تا ۱۰۰، مقیاس لگاریتمی)؛ نوع جریمه (L1 یا L2)
NB	بیز ساده با توزیع گاوسی: هر ورودی در هر رده زنگوله‌ای و مستقل از دیگران فرض می‌شود	کف واریانس (۱e-۱۱ تا ۱e-۳، مقیاس لگاریتمی)
DT	درخت تصمیم: رشته‌ای از پرسش‌های بله/خیر دربارهٔ ورودی‌ها که به برگی با یک رده می‌رسد	بیشینهٔ عمق (۳، ۵، ۸، ۱۲، ۲۰)؛ کمینهٔ اندازهٔ برگ (۱، ۵، ۲۰، ۵۰)؛ معیار شکست (جینی یا آنتروپی)
RF	جنگل تصادفی: درختان بسیار روی بخش‌های تصادفی داده که رأی‌شان میانگین‌گیری می‌شود	شمار درختان (۵۰، ۱۰۰، ۲۰۰)؛ بیشینهٔ عمق (۵، ۱۰، ۲۰، ۴۰)؛ کمینهٔ اندازهٔ برگ (۱، ۲، ۵، ۲۰)؛ سهم ورودی‌های آزموده در هر شکست (۰٫۳، ۰٫۵، ۱)
SVM	ماشین بردار پشتیبان با هستهٔ شعاعی: رده‌ها را با مرزی هموار که از نزدیک‌ترین نمونه‌های آموزشی ساخته می‌شود جدا می‌کند	هزینهٔ C (۰٫۱ تا ۱۰۰، مقیاس لگاریتمی)؛ پهنای هسته گاما (۰٫۰۰۱ تا ۱، مقیاس لگاریتمی)؛ آموزش روی حداکثر ۱۲٬۰۰۰ ردیف

هر مدل یکسان تنظیم شد. فضای جست‌وجوی هر مدل در جدول 2 آمده است. روش جست‌وجو، جست‌وجوی تصادفی با بودجهٔ یکسان (ده آزمایش) برای هر مدل است؛ هر آزمایش با اعتبارسنجی سه‌بخشی درون ردیف‌های آموزش امتیاز می‌گیرد، در خط لوله‌هایی که خانوارها را تقسیم می‌کنند گروه‌بندی‌شده بر پایهٔ خانوار و در خط لوله‌های تقسیم ردیف بر پایهٔ ردیف، و بیش‌نمونه‌گیری تنها درون بخش‌های آموزش درونی در خط لوله‌هایی که پس از تقسیم بیش‌نمونه‌گیری می‌کنند؛ ملاک سطح زیر منحنی ROC است. پیکربندی برگزیده روی همهٔ ردیف‌های آموزش دوباره برازش می‌شود و یک بار روی ردیف‌های کنار گذاشته‌شده امتیاز می‌گیرد (اعتبارسنجی تودرتو). آزمایش با دو بذر اصلی اجرا شد. چهار خط لوله فقط در جای بیش‌نمونه‌گیری و آنچه تقسیم می‌شود فرق دارند. در بیش‌نمونه‌گیری و سپس تقسیم ردیف‌ها رستهٔ کمیاب روی کل فایل با درون‌یابی میان همسایه‌های نزدیک بیش‌نمونه‌گیری می‌شود (Chawla و همکاران, 2002) و سپس ردیف‌ها تصادفی تقسیم می‌شوند؛ در تقسیم ردیف‌ها و سپس بیش‌نمونه‌گیری تقسیم اول می‌آید و تنها ردیف‌های آموزش بیش‌نمونه‌گیری می‌شوند؛ در تقسیم خانوارها و سپس بیش‌نمونه‌گیری هر خانوار کاملاً در یک سو قرار دارد؛ و در تقسیم خانوارها بدون بیش‌نمونه‌گیری ردیف‌های آموزش همان‌گونه که هستند به کار می‌روند و امتیاز آزمون با وزن پیمایش وزن‌دهی می‌شود. در هر تقسیم از طرح‌های خانوار حدود ۹٬۱۸۹ ردیف از حدود ۲۷٬۴۱۳ خانوار کنار گذاشته می‌شود (جدول 3) و هیستوگرام فراوانی رستهٔ کمیاب در همهٔ ردیف‌ها، ردیف‌های آموزش و ردیف‌های کنار گذاشته‌شده در شکل 8 مواد تکمیلی آمده است. چون خانوارهای کنار گذاشته‌شده زیرمجموعه‌ای تصادفی از همهٔ خانوارها هستند، این طرح‌ها تعمیم به خانوارهای تازه در همان نمونه را می‌آزمایند؛ هیچ گروه کنار گذاشته‌شده‌ای بیرون از دامنهٔ گروه‌های آموزش نیست، پس هیچ‌کدام به معنای پوش محدب برون‌یابی نیست. برای مطالعهٔ رگرسیونی طرح دوم کل ایالت‌ها را کنار می‌گذارد که آزمونی سخت‌تر برای انتقال است.

جدول 3: ردیف‌های کنار گذاشته‌شده و نگه‌داشته‌شده در تقسیم خانوار و سهم هر رستهٔ خودرو در آن‌ها (فراوانی رستهٔ کمیاب پیش و پس از کنار گذاشتن).

گروه	ردیف	سهم AFV	سهم EV
همهٔ مالکان خودروی نو	۳۰٬۵۱۱	۴٫۱٪	۱٫۳٪
خانوارهای آموزش (میانگین سه تقسیم)	۲۱٬۳۲۲	۴٫۱٪	۱٫۳٪
خانوارهای کنار گذاشته‌شده (میانگین سه تقسیم)	۹٬۱۸۹	۳٫۹٪	۱٫۳٪

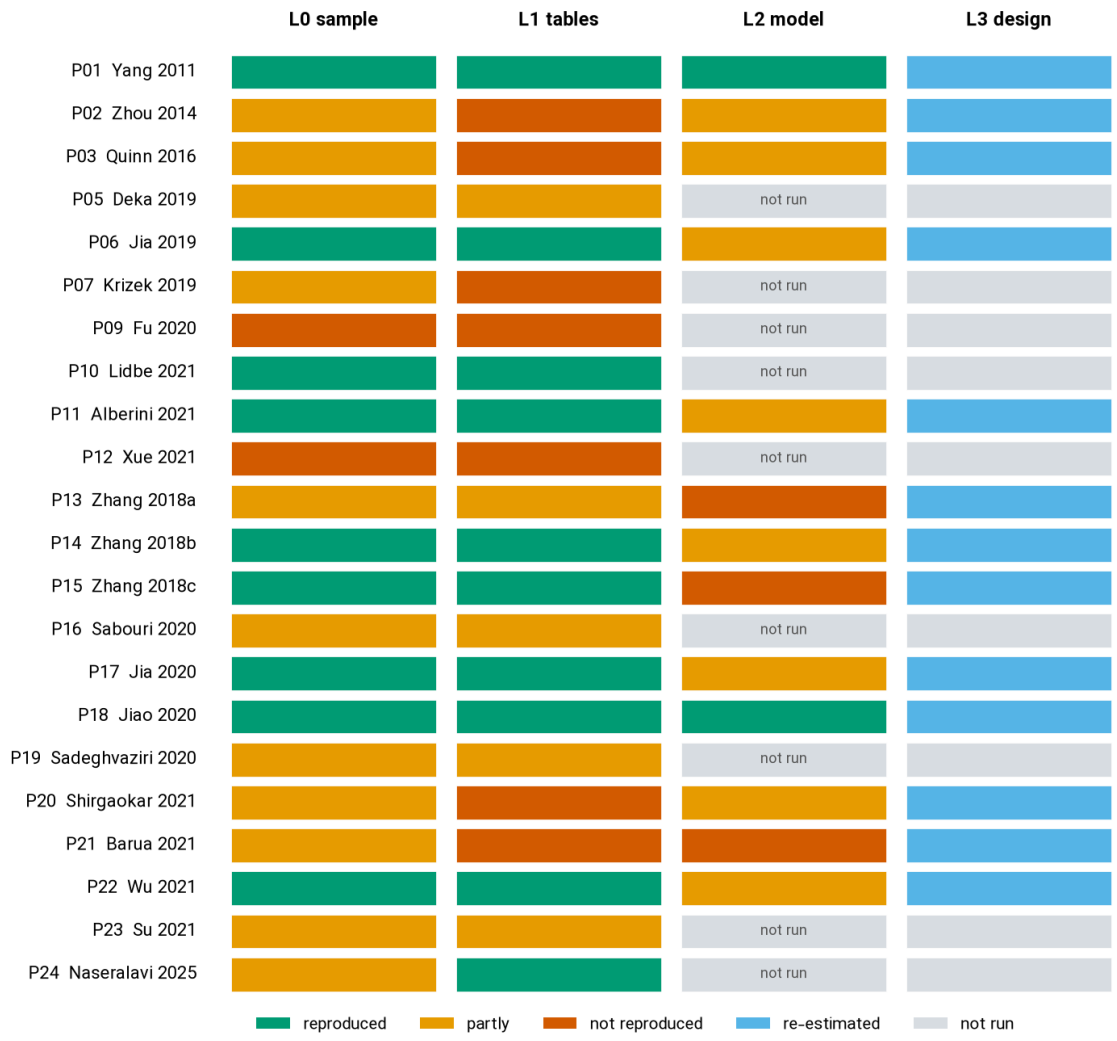
نرم‌افزار و بازتولیدپذیری

همهٔ عددها، جدول‌ها و شکل‌ها با کد از پرونده‌های نتیجه ساخته می‌شوند؛ متن هیچ نتیجهٔ دست‌تایپ‌شده‌ای ندارد. کد تحلیل، موتور رژیم‌های رگرسیون و خط لوله‌های اعتبارسنجی در مخزن است و با یک فرمان اجرا می‌شود، و ابزارهای رگرسیون و اعتبارسنجی بخشی از pynhts هستند (ماژول‌های `pynhts.models` و `pynhts.validate`) با آزمون واحد، آزمون دادهٔ واقعی در برابر R و مستندات. طرح تحلیل، با فرضیه‌ها و آستانه‌ها، پیش از اجرای مدل‌های نیمهٔ دوم مطالعه‌ها نوشته شد و تاریخ دارد.

نتایج

آنچه بازسازی شد

نردبان در شکل 2 دیده می‌شود. از ۲۲ مطالعه، ممیزی فایل‌های عمومی را برای بازسازی داده‌های ۹ مطالعه کافی، برای ۱۱ مطالعه تا حدی کافی و برای ۲ مطالعه ناکافی دانست که در آن‌ها جامعه یا قاعدهٔ نمونه‌گیری ننوشته شده است. در مجموع، ۸۷۹ مقدار از ۱٬۴۵۴ مقدار توصیفی چاپی دقیقاً در دقت چاپی بازتولید شد و ۸۱ درصد دقیقاً یا تا دو درصد. شکاف‌ها بیشتر از خود مقاله‌ها برخاستند: قاعده‌های نانوشتهٔ حذف رکورد، حساب‌هایی که در جدول جمع نمی‌شود و متغیرهایی که در فایل عمومی نیستند. ۱۳ مطالعه را می‌شد تا پلهٔ مدل برد؛ دلیل دیگر مطالعه‌ها در جدول 1 آمده است.

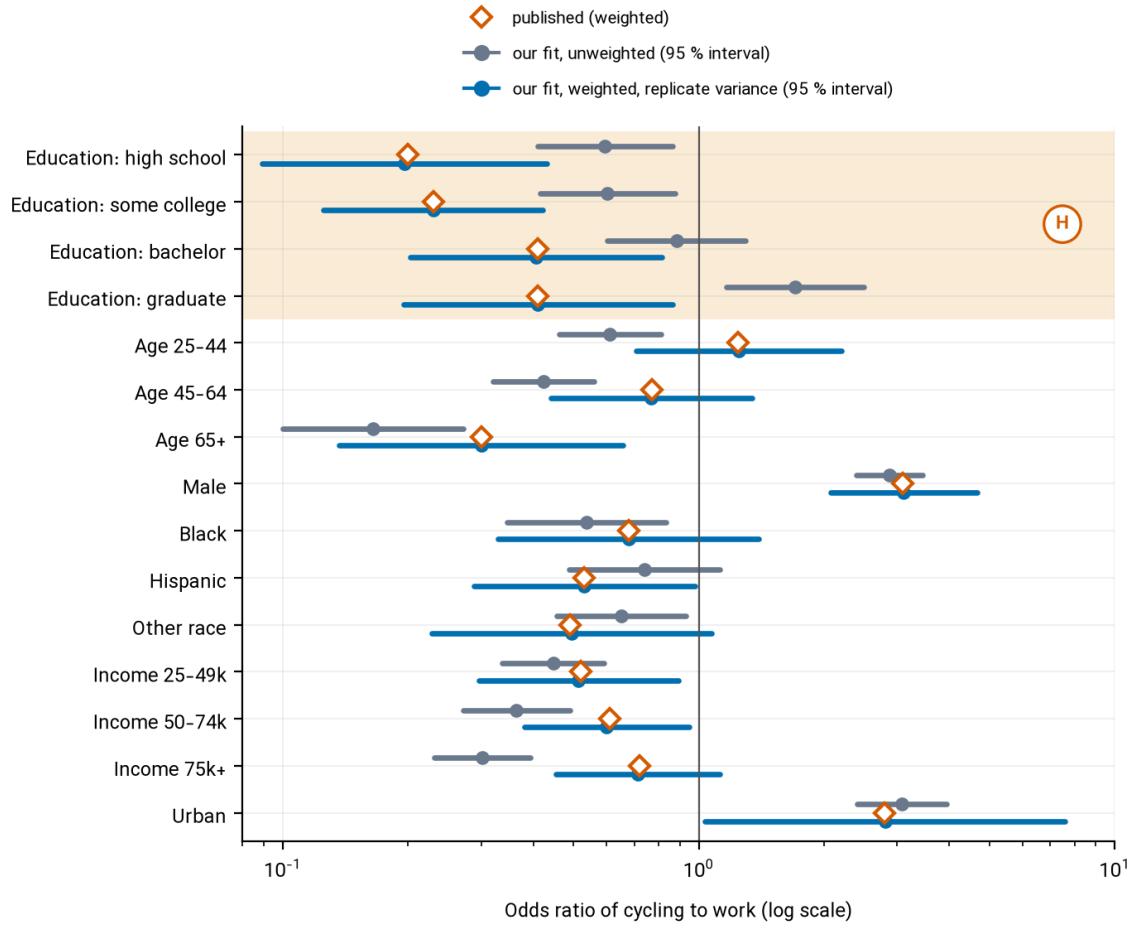


شکل 2: نردبان برای ۲۲ مطالعه. هر سطر یک مطالعه و هر ستون یک پله است؛ رنگ‌ها درجه را نشان می‌دهند. ستون L3 نشان می‌دهد کدام مدل‌ها دوباره برآورد شدند و درجه نیست.

برآوردگر اعلام‌شده عددهای چاپی را بازتولید می‌کند

مطالعه‌هایی که وزن را اعلام کرده بودند با وزن بازتولید شدند و آن‌هایی که نه، بدون وزن. در مطالعهٔ پیاده‌روی و دوچرخه‌سواری کوین و همکاران، که می‌گوید وزن‌های جمعیتی و جک‌نایف به کار رفته، نسبت‌های بخت وزنی ۳۴ مقدار از ۶۰ مقدار چاپی را بازتولید کرد و نسبت‌های بدون وزن ۵ مقدار؛ برای دو مدل رفت‌وآمد، ۲۹ مقدار از ۳۰ مقدار تا دقت چاپی یا دو درصد برابر شد. در مطالعهٔ فعالیت بدنی بر پایهٔ فصل، نسبت‌های بخت وزنی ۲۱ مقدار از ۲۱ مقدار چاپی را بازتولید کرد و نسبت‌های بدون وزن ۶ مقدار. در جهت دیگر، مطالعهٔ دوجمله‌ای منفی تحرک اشتراکی و سفرها، که برازش بدون وزن در R گزارش می‌کند، با برازش بدون وزن بازتولید شد: ۲۱ ضریب از ۳۲ ضریب در دقت چاپی برابر بود. مطالعهٔ دوجمله‌ای منفی صفرافزودهٔ مالکیت خودرو و سواری‌خواهی با برازش بدون وزن برای ۱۷ ضریب از ۳۲ ضریب تا دو درصد و برای ۲۶ ضریب تا پنج درصد بازتولید شد. کشش مایل سالانه نسبت به هزینهٔ سوخت در هر مایل در مطالعهٔ مایل پیموده −۰٫۵۱۵ بود در برابر −۰٫۵۱۰ چاپی. این نتایج فرضیهٔ نخست را پشتیبانی می‌کنند: عددهای چاپی وقتی برآوردگر اعلام‌شده به کار رود از فایل‌های عمومی بازیافتنی‌اند و در غیر این صورت نه.

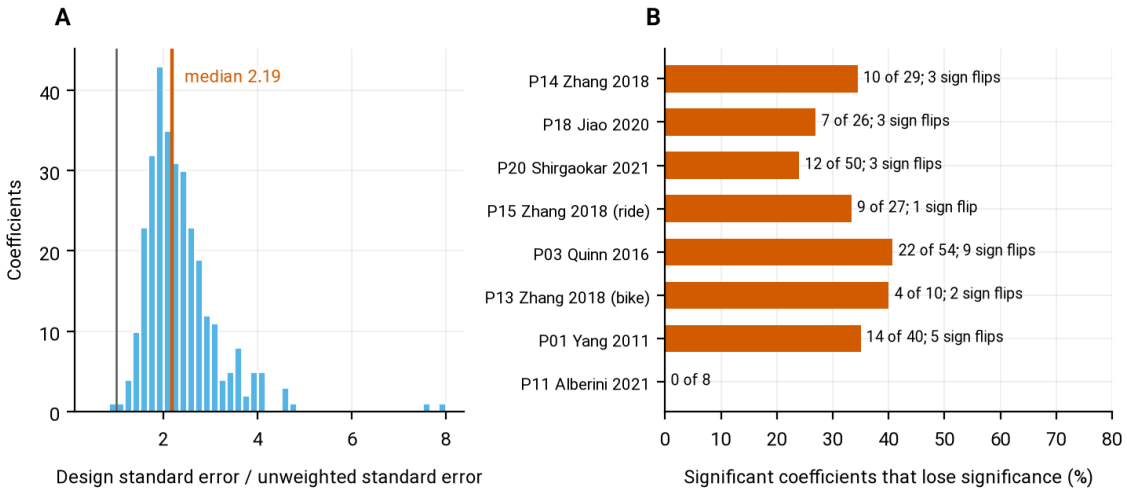
برآوردگر تنها در مطابقت نقش ندارد، در محتوا هم مؤثر است. برجسته‌ترین نمونه اثر تحصیلات بر دوچرخه‌سواری به محل کار است. نسبت بخت چاپی برای مدرک تحصیلات تکمیلی در برابر کمترین گروه تحصیلی ۰٫۴۱ است، برازش وزنی ۰٫۴۱ می‌دهد و برازش بدون وزن ۱٫۷۱: علامت ارتباط با تحصیلات برعکس می‌شود (شکل 3، حرف مرجع H). همین در اثر منطقهٔ میانه‌غربی بر سفر فعال در ۲۰۰۱ دیده می‌شود، با ۰٫۶۷ چاپی، ۰٫۶۷ وزنی و ۰٫۸۳ بدون وزن. وزن‌ها در این داده‌ها دربارهٔ پیامد آگاهی‌بخش‌اند، زیرا پاسخ و انتخاب با سن، تحصیلات و منطقه فرق می‌کند و برازش بدون وزن ارتباط را در نمونه برآورد می‌کند، نه در جمعیت.



شکل 3: نسبت‌های بخت دوچرخه‌سواری به محل کار در مطالعهٔ کوین و همکاران: مقدار چاپی، برازش بدون وزن ما و برازش وزنی ما با واریانس وزن تکرار. سطرهای سایه‌دار (حرف مرجع H) اثرهای تحصیلات‌اند که در متن بحث شده است.

خطای معیار و معناداری

در ۳۱۰ ضریبِ دوباره برآوردشدهٔ هشت مطالعهٔ رگرسیونی، خطای معیار طراحی‌آگاه به‌طور میانه ۲٫۱۹ برابر خطای معیار بدون وزن بود (دامنهٔ میان‌چارکی ۱٫۸۶ تا ۲٫۶۶) و برای ۹۵ درصد ضرایب این نسبت از یک‌ونیم فراتر رفت (شکل 4، پنل A). از ۲۴۴ ضریبی که در برازش بدون وزن در سطح پنج درصد معنادار بودند، ۷۸ ضریب (۳۲ درصد) در برازش طراحی‌آگاه معنادار نبودند، در حالی که ۱۱ ضریب معنادار شد؛ ۲۶ ضریب (۸٫۴ درصد) میان دو برازش علامتش عوض شد. سهم‌ها از مطالعه‌ای به مطالعهٔ دیگر فرق دارد (شکل 4، پنل B؛ جدول 4). این نتایج فرضیه‌های دوم و سوم را پشتیبانی می‌کنند: میانهٔ نسبت از یک‌ونیم بیشتر است و بیش از یک‌دهم ضرایب معنادار زیر طرح معنادار نبودند.



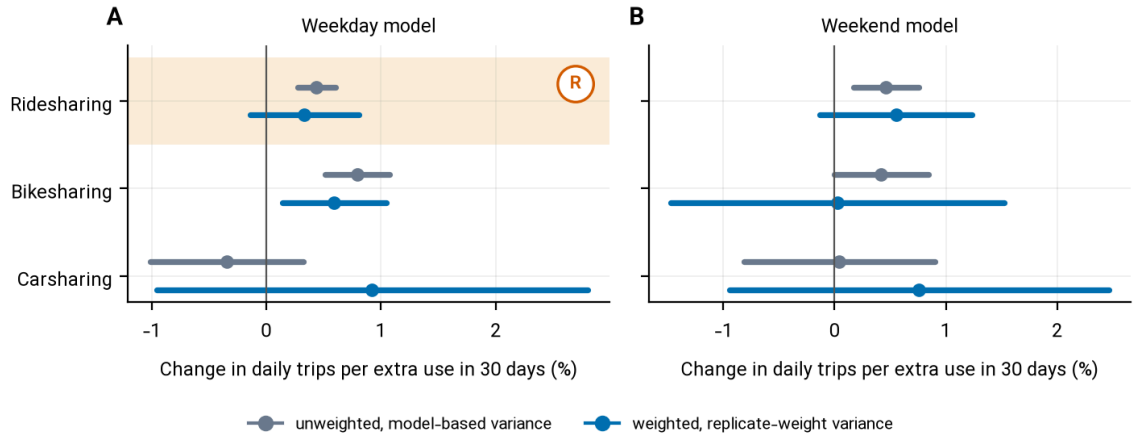
شکل 4: اثر طرح نمونه‌گیری بر خطای معیار و معناداری. پنل A: نسبت خطای معیار طراحی‌آگاه به خطای معیار بدون وزن در همهٔ ضرایب دوباره برآوردشده. پنل B: سهم ضریب‌های معنادار در برازش بدون وزن که در برازش طراحی‌آگاه معناداری خود را از دست می‌دهند، به تفکیک مطالعه.

جدول 4: برآورد دوبارهٔ هشت مطالعهٔ رگرسیونی. نسبت خطای معیار: خطای معیار طراحی‌آگاه بر خطای معیار بدون وزن. معناداری ازدست‌رفته: ضرایب معنادار در سطح ۵٪ در برازش بدون وزن که در برازش طراحی‌آگاه نیستند. تغییر علامت: علامت برآورد بدون وزن و طراحی‌آگاه فرق دارد.

مطالعه	ضرایب	مقدار بازتولیدشده: وزنی / وزن / چاپی	چاپی بدون	میانهٔ نسبت خطای معیار	معناداری ازدست‌رفته	تغییر علامت
P01 Yang ۲۰۱۱	۵۰	۲۱ / ۲۱ / ۶		۱٫۸۶	۴۰ / ۱۴	۵
P03 Quinn ۲۰۱۶	۷۲	۶۰ / ۳۴ / ۵		۲٫۳۳	۵۴ / ۲۲	۹
P11 Alberini ۲۰۲۱	۸	۴ / ۰ / ۱		۱٫۸۴	۸ / ۰	۰
P13 Zhang ۲۰۱۸ (bike)	۳۰	۲ / ۰ / ۰		۱٫۹۶	۱۰ / ۴	۲
P14 Zhang ۲۰۱۸	۳۲	۳۲ / ۶ / ۱۷		۲٫۵۴	۲۹ / ۱۰	۳
P15 Zhang ۲۰۱۸ (ride)	۳۰	۲ / ۰ / ۰		۲٫۴۲	۲۷ / ۹	۱
P18 Jiao ۲۰۲۰	۳۲	۳۲ / ۱۳ / ۳۱		۲٫۴۵	۲۶ / ۷	۳
P20 Shirgaokar ۲۰۲۱	۵۶	۱۴ / ۷ / ۹		۲٫۴۳	۵۰ / ۱۲	۳

نمونه‌ای ملموس، مطالعهٔ دوجمله‌ای منفی تحرک اشتراکی و سفرهاست. نتیجهٔ آن این است که استفاده از سواری‌خواهی پیش‌بین معنادار سفرهای روزانه است. ضریب بازتولید می‌شود (۰٫۳۳ درصد سفر بیشتر برای هر بار استفادهٔ اضافی در ماه گذشته در روزهای هفته) و در برازش بدون وزن به‌شدت معنادار است، با p کمتر از ۰٫۰۰۱. با واریانس طراحی‌آگاه خطای معیار ۰٫۰۰۲۴ است به‌جای ۰٫۰۰۰۸ و p برابر برابر ۰٫۱۷ می‌شود؛ در آخر هفته برابر ۰٫۱۱ است (شکل 5، حرف مرجع R). برآورد نقطه‌ای جابه‌جا نمی‌شود؛ شاهد برای آن جابه‌جا می‌شود. در مقابل، دوچرخهٔ اشتراکی در روزهای هفته در هر دو برازش معنادار است. نمونهٔ دوم تعداد خودروهای خانوار در مدل ZINB سواری‌خواهی است که پایدار است، در حالی که اثر

رانندهبودن بر تعداد سواریها، که در برازش بدون وزن معنادار است (p کمتر از ۰/۰۰۱)، در برازش طراحیآگاه نامطمئن میشود (p برابر ۰/۱۴).



شکل 5: جملههای تحرک اشتراکی در مدلهای سفر روزهای هفته (پنل A) و آخر هفته (پنل B) در مطالعهٔ جیائو و همکاران، با بازههای نودوپنجدرصدی از برازش بدون وزن و طراحیآگاه. حرف مرجع R سطر سواریخواهی را که در متن بحث شده نشان میدهد.

کدام لایهٔ طرح اهمیت دارد

برازش طراحیآگاه در سه لایه با برازش بدون وزن فرق دارد: وزنها برآورد را عوض میکنند؛ خوشهدانستن اشخاص یک خانوار خطای معیار را بزرگ میکند؛ و فرمول تکرار لایهبندی و ساختار باقیمانده را میگیرد. بیشتر تفاوت خطای معیار از لایهٔ دوم است. برازش وزنی با واریانس ساندویچی خوشهایِ خانوار خطای معیارهایی داد که میانهٔ نسبتشان به خطای معیارهای وزن تکرار ۰/۹۹ بود؛ از ۱۸۲ ضریب معنادار در آن برازش، فقط ۱۶ ضریب (۹ درصد) زیر واریانس تکرار معناداری خود را از دست داد. در عمل، برازش وزنی با خطای معیار خوشهایِ خانوار در این دادهها تقریب خوبی از واریانس تکرار است، که برای موج ۲۰۲۲ که وزن تکرار ندارد و طرح با لایه و خوشه بیان میشود سودمند است.

مدلهای بدون جدول ضرایب چاپی

مدل معادلات ساختاری خرید اینترنتی و سفرهای خرید، که ضرایب استانداردشده چاپ میکند، با همان متغیرها به جز دو متغیر که از منبع دیگرند برازش شد. دو اثر متقابل آن در برازش وزنی ما −۰/۰۸ و ۰/۱۷ بود در برابر ۰/۱۲ و ۰/۲۱ چاپی، با خطای معیار طراحی حدود ۰/۰۲؛ علامت اثرهای برونزا تا علامت دلخواه مقیاس پنهان با چاپی سازگار است، بنابراین آن را تنها از نظر کیفی بازتولیدشده درجهبندی میکنیم. برای مطالعهٔ همتاسازی امتیاز گرایش سواریخواهی و خودروکیلومتر، همتاسازی حریصانهٔ نزدیکترین همسایه در همان لایهها ۲٬۶۶۴ سهتایی همتاشده از اشخاص دارای دسترسی به خودروی خانوار داد، در برابر حدود پنج هزار چاپی. میانگین مایل روزانهٔ همتاشده با میانگینهای چاپی فرق داشت (برای کاربران پرتکرار ۲۹/۸ در برابر ۲۴/۸ مایل)، اما کاهش تعداد خودروهای خانوار هنگام گذر از غیرکاربر به کاربر گاهبهگاه به مقدار چاپی نزدیک بود (−۰/۰۶۷ در برابر −۰/۰۶۵). آزمون زوجی مقاله برای این تضاد آمارهٔ t برابر ۲/۴۵ میدهد؛ با واریانس وزن تکرار ۰/۹۵ است. دو تضاد خودروی دیگر زیر طرح آمارههای ۰/۰۰ و ۱/۶۲ دارند، پس هیچیک از سه تفاوت خودروی خانوار در نمونههای همتاشدهٔ ما زیر طرح از صفر قابلتمایز نیست.

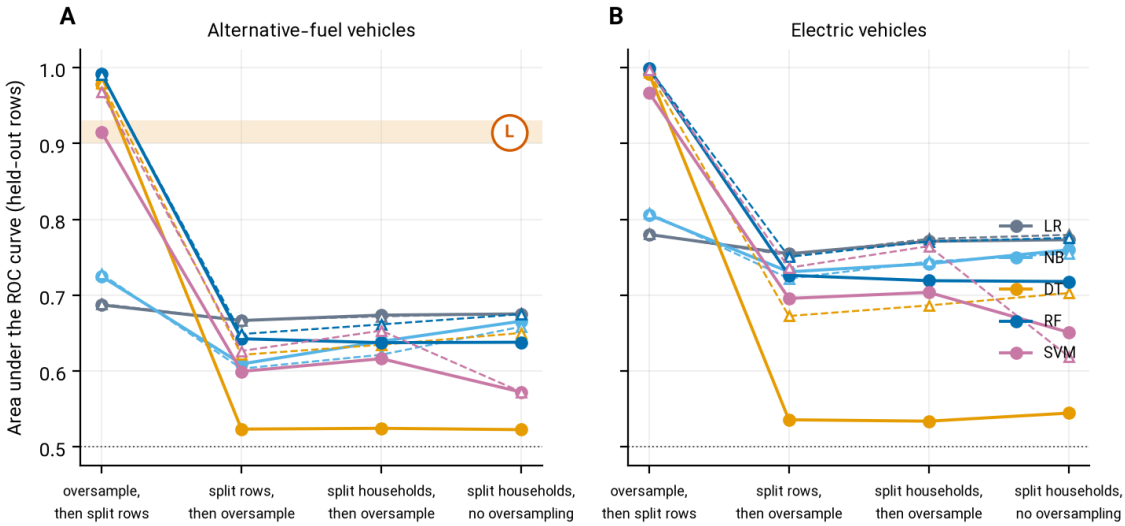
مدلهای یادگیری ماشین

طبقهبندهای پذیرش خودروی سوخت جایگزین و برقی بیشترین وابستگی را به خط لوله نشان دادند. وقتی رستهٔ کمیاب پیش از تقسیم ردیفها بیشنمونهگیری شد، سطح زیر منحنی جنگل تصادفی ۰/۹۹ (خودروی سوخت جایگزین) و ۱/۰۰ (خودروی

برقی) و درخت تصمیم ۰/۹۸ و ۰/۹۹ بود، نزدیک به ۰/۹۲ و ۰/۹۱ چاپی (شکل 6، حرف مرجع L). وقتی تقسیم اول آمد، همان مدل‌ها برای جنگل تصادفی ۰/۶۴ و ۰/۷۳ و برای درخت ۰/۵۲ و ۰/۵۴ گرفتند و تقسیم بر پایهٔ خانوار این را عوض نکرد (۰/۶۴ و ۰/۷۲؛ با وزن پیمایش ۰/۶۴ و ۰/۷۲). بنابراین بادکردگی برای جنگل تصادفی ۰/۳۵ (سوخت جایگزین) و ۰/۲۸ (برقی) و برای درخت تصمیم ۰/۴۵ و ۰/۴۶ بود. درخت تصمیم که در خط لولهٔ نشت‌دار ۰/۹۸ گرفت، هنگام ارزیابی صادقانه به‌ندرت بهتر از شانس است (۰/۵۲). این فرضیهٔ چهارم را پشتیبانی می‌کند: امتیازهای چاپی را تنها در خط لوله‌ای می‌توان نزدیک کرد که نسخه‌های ردیف‌های آزمون به داده‌های آموزش راه یابند. تنظیم یکسان همهٔ مدل‌ها شکاف را از میان نبرد. جنگل تصادفی تنظیم‌شده وقتی بیش‌نمونه‌گیری مقدم بود ۰/۹۹ و با تقسیم بر پایهٔ خانوار ۰/۶۶ گرفت (بادکردگی ۰/۳۳؛ برای خودروی برقی ۰/۲۳)، و درخت تصمیم تنظیم‌شده از ۰/۵۲ به ۰/۶۳ رسید اما بسیار پایین‌تر از امتیاز نشت‌دارش (۰/۹۸) ماند. بذر اصلی دوم همین الگو را داد (بزرگ‌ترین انحراف معیار میانگین روی بذرها ۰/۰۲۶).

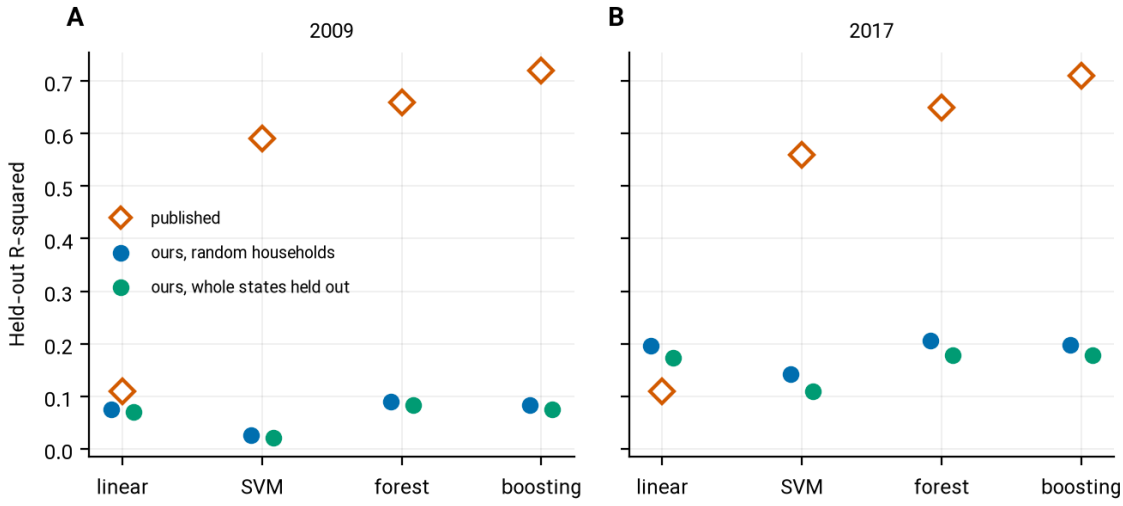
جدول 5: سطح زیر منحنی ROC روی ردیف‌های کنار گذاشته‌شده (میانگین سه بذر برای تنظیمات پیش‌فرض و دو بذر برای مدل‌های تنظیم‌شده). نشت‌دار: بازنمونه‌گیری و سپس تقسیم ردیف‌ها. ردیف‌ها: تقسیم ردیف‌ها و سپس بازنمونه‌گیری. خانوارها: تقسیم خانوارها و سپس بازنمونه‌گیری. وزنی: تقسیم خانوارها بدون بازنمونه‌گیری و امتیاز وزنی.

پیامد	مدل	نشت‌دار	ردیف‌ها	خانوارها	وزنی	تنظیم‌شده، نشت‌دار	تنظیم‌شده، خانوارها
سوخت جایگزین	LR	۰/۶۹	۰/۶۷	۰/۶۷	۰/۶۸	۰/۶۹	۰/۶۷
سوخت جایگزین	NB	۰/۷۲	۰/۶۱	۰/۶۴	۰/۶۷	۰/۷۳	۰/۶۲
سوخت جایگزین	DT	۰/۹۸	۰/۵۲	۰/۵۲	۰/۵۲	۰/۹۸	۰/۶۳
سوخت جایگزین	RF	۰/۹۹	۰/۶۴	۰/۶۴	۰/۶۴	۰/۹۹	۰/۶۶
سوخت جایگزین	SVM	۰/۹۱	۰/۶۰	۰/۶۲	۰/۵۷	۰/۹۷	۰/۶۵
برقی	LR	۰/۷۸	۰/۷۵	۰/۷۷	۰/۷۷	۰/۷۸	۰/۷۷
برقی	NB	۰/۸۱	۰/۷۳	۰/۷۴	۰/۷۶	۰/۸۱	۰/۷۴
برقی	DT	۰/۹۹	۰/۵۴	۰/۵۳	۰/۵۴	۰/۹۹	۰/۶۹
برقی	RF	۱/۰۰	۰/۷۳	۰/۷۲	۰/۷۲	۱/۰۰	۰/۷۷
برقی	SVM	۰/۹۷	۰/۷۰	۰/۷۰	۰/۶۵	۱/۰۰	۰/۷۶



شکل 6: سطح زیر منحنی ROC پنج طبقه‌بند در چهار خط لوله، برای خودروهای سوخت جایگزین (پنل A) و خودروهای برقی (پنل B). دایره‌های توپر: تنظیمات پیش‌فرض آزمایش نخست؛ مثلث‌های توخالی: همهٔ مدل‌ها یکسان تنظیم‌شده. حرف مرجع L امتیاز چاپی جنگل تصادفی مطالعه را نشان می‌دهد.

مطالعهٔ رگرسیونی خرید اینترنتی خانوار ضریب تعیین کنار گذاشته‌شدهٔ ۰/۷۱ را برای تقویت گرادیان در ۲۰۱۷ و ۰/۷۲ را در ۲۰۰۹ چاپ کرده بود. مدل تقویت گرادیان ما روی متغیرهای عمومی متناظر با فهرست چاپی به ۰/۲۰ و ۰/۰۸ روی خانوارهای تصادفی کنار گذاشته‌شده و ۰/۱۸ هنگام کنار گذاشتن کل ایالت‌ها رسید (شکل 7). شکاف حدود ۰/۵۱ با طرح اعتبارسنجی توضیح داده نمی‌شود؛ جدول متغیرهای چاپی درصدهای بالاتر از صد و متغیری دوبار فهرست‌شده دارد، بنابراین نمی‌توانیم بگوییم کدام متغیرها وارد مدل چاپی شدند و دربارهٔ درستی آن ادعایی نمی‌کنیم.



شکل 7: ضریب تعیین کنار گذاشته‌شدهٔ چهار مدل رگرسیونی برای خرید اینترنتی خانوار در ۲۰۰۹ (پنل A) و ۲۰۱۷ (پنل B): مقدارهای چاپی و مقدارهای ما برای کنار گذاشتن تصادفی و کنار گذاشتن کل ایالت‌ها.

بحث

نردبان چه می‌گوید. برای یک رگرسیون معمولی NHTS، عددهای چاپی از فایل‌های عمومی بازیافتنی‌اند، به شرط آن‌که خواننده بداند نویسندگان وزن داده‌اند یا نه، و این اطلاعاتی است که مقاله‌ها را گاه باید دوبار خواند تا بدهند. پلهٔ بعد جایی است که یافته‌ها جابه‌جا می‌شوند. خطای معیار میانه زیر طرح حدود ۲٫۲ برابر می‌شود، ۳۲ درصد ضرایب معنادار معناداری خود را از دست می‌دهند و ۸ درصد علامتشان عوض می‌شود. این‌ها نشانهٔ خطا در مقاله‌های اصلی نیستند، که بیشترشان آنچه کرده‌اند گفته‌اند؛ نشان می‌دهند برآوردگر انتخابی بوده با پیامدهایی که مقاله کم دربارهٔ آن بحث می‌کند. خواننده‌ای که ستارهٔ معناداری یک رگرسیون بدون وزن NHTS را می‌بیند به کمیتی نگاه می‌کند که بیان احتمالی‌اش دربارهٔ جمعیتی که پیمایش توصیف می‌کند صدق نمی‌کند.

تفسیر مهندسی. برای برنامه‌ریز پرسش این است که آیا ضریب را می‌توان برای پیش‌بینی تقاضا به کار برد. نمونهٔ سواری‌خواهی آموزنده است. ضریب حدود ۰٫۳ درصد سفر بیشتر برای هر بار استفادهٔ اضافی کوچک است و علامتش زیر هیچ رژیمی عوض نمی‌شود، اما معناداری‌اش، که سرخط مقاله بر آن استوار است، عوض می‌شود. نتیجهٔ بدون وزنی که تنها در نمونهٔ پاسخ‌گویان برقرار است توصیف پاسخ‌گویان است؛ نتیجهٔ طراحی‌آگاه عدم‌قطعیتی را دارد که پیش‌بینی لازم دارد. برای مطالعهٔ همتاسازی خودروکیلومتر، این نتیجه که مردم با گذر از ناکاربری به کاربری پرتکرار خودروی خانوار را عوض می‌کنند در نمونه‌های همتاشدهٔ ما زیر طرح پشتیبانی نمی‌شود؛ یعنی روایت کیفی دربارهٔ کنار گذاشتن خودرو به آزمونی در سطح نمونه تکیه دارد. هیچ‌یک از دو نتیجه رد نمی‌شود؛ هر دو پشتیبانی‌ای را که آزمون بدون وزن به آن‌ها می‌داد از دست می‌دهند.

چرا اثر طرح بیشتر از خوشه‌بندی برمی‌خیزد. اشخاص یک خانوار مکان، درآمد، خودرو و فرصت‌های سفر را مشترک دارند، بنابراین اندازهٔ نمونهٔ مؤثر برای یک متغیر در سطح خانوار از شمار ردیف‌ها کمتر است. وزن‌های تکرار لایه‌بندی و انتخاب نابرابر را می‌افزایند، اما در این پیمایش خوشه‌بندی بر پایهٔ خانوار تقریباً همهٔ اصلاح را توضیح می‌دهد. این برای کاربران موج ۲۰۲۲، که ناچارند با لایه و خوشه به‌جای وزن تکرار کار کنند، امیدوارکننده است و قاعده‌ای عملی می‌دهد: هر جا وزن تکرار در دسترس نیست، در سطح خانوار خوشه‌ای کن و وزن بده.

امتیازهای پیش‌بینی. امتیازهای چاپی طبقه‌بندهای پذیرش خودرو را تنها در خط لوله‌ای می‌توان نزدیک کرد که نسخه‌های مصنوعی ردیف آزمون به آموزش راه می‌یابند، نشتی که در رشته‌های دیگر ثبت شده است (Demircioğlu, 2024؛ Kapoor و Narayanan, 2023). از روی مقاله‌ها نمی‌توانیم بدانیم کدام ترتیب به کار رفته و نتیجهٔ ما ثابت نمی‌کند خط لولهٔ اصلی نشت داشته است؛ نشان می‌دهد عددهای چاپی در دامنهٔ نشت‌دارند. خواننده می‌تواند ادعا را با `pynhts.validate.audit_classifier` بسنجد که چند ثانیه می‌برد. وقتی داده بر پایهٔ خانوار تقسیم و فقط ردیف‌های آموزش بیش‌نمونه‌گیری شود، جنگل تصادفی پذیرش سوخت جایگزین، تقریباً هم‌دقت رگرسیون لجستیک (۰٫۶۴ و ۰٫۶۷)، سطح زیر منحنی‌ای ناچیز دارد، که پیش از به‌کارگیری مدل برای هدف‌گذاری سیاست پذیرش دانستنی است. شکاف امتیاز رگرسیونی مطالعهٔ خرید اینترنتی از نوع دیگری است: طرح اعتبارسنجی آن را توضیح نمی‌دهد و بدون تعریف متغیرهای چاپی بی‌توضیح می‌ماند.

پیامد برای بسته. این مطالعه به سه ابزار نیاز داشت که در بسته نبود: رگرسیونی که برآورد وزنی را با طرح واریانس موج برمی‌گرداند، بررسی نشت در خط لوله‌های پیش‌بینی، و راهی برای مقایسهٔ عدد چاپی با عدد بازسازی‌شده. دو ابزار نخست در pynhts با نام‌های `pynhts.models` و `pynhts.validate` منتشر شده، در برابر R و با آزمون اعتبارسنجی شده‌اند و قاعده‌های راهنمای عامل بسته اکنون می‌گوید از آن‌ها استفاده شود. سه توسعهٔ دیگر از شواهد برمی‌آید و برنامه‌ریزی شده است: رابط فرمول و خانواده‌های ترتیبی و چندحالته برای مقاله‌های انتخاب گسسته که مجموعهٔ کنونی ندارد؛ مدل‌های چندسطحی در سطح خانوار با ادغام جزئی برای مقایسه‌های منطقه‌ای؛ و کتابخانه‌ای از دستورهای بازتولید، یکی برای هر مطالعهٔ منتشرشده، تا شکاف‌های ثبت‌شده به آزمون‌های رگرسیونی تبدیل شوند که هر کاربر بتواند اجرا کند.

محدودیت‌ها. مجموعهٔ مطالعه‌ها مجموعه‌ای در دسترس از ۲۲ مطالعه است و ۱۳ مورد از آن‌ها دوباره برازش شد؛ سهم نتیجه‌های شکننده در همهٔ مقاله‌های NHTS بنابراین برآورد نشده است. برای دو مطالعه هدف چاپی نتیجه‌ای کلیدی است که از درصدی در چکیده به دست آمده و درجه‌بندی آن‌ها تقریبی است. برازش‌های بدون وزن ما بعضی مقدارهای چاپی را فقط تقریبی بازتولید می‌کنند، زیرا نمونه‌های چاپی دقیقاً بازسازی نمی‌شوند. مطالعهٔ همتاسازی با همتاسازی حریصانه تکرار شد، نه همتاسازی بهینهٔ اصلی. آزمایش‌های یادگیری ماشین از پنج طبقه‌بند با ده آزمایش تنظیم برای هر کدام استفاده می‌کنند، جست‌وجویی متوسط که مقایسه را منصفانه می‌کند اما جامع نه؛ مدل‌های چاپی ممکن است بیشتر تنظیم شده باشند. پیمایش در ۲۰۲۲ وزن تکرار ندارد، پس برازش طراحی‌آگاه آنجا از خطی‌سازی استفاده می‌کند که در بسته پیاده شده اما در این مقاله به کار نرفته است. سرانجام، نتیجه‌ها را حساس به طرح می‌نامیم؛ نمی‌گوییم نادرست‌اند.

توصیه‌ها. برای نویسندگان: بگویید وزن به کار رفته یا نه و چرا؛ خطای معیار را از طرح پیمایش گزارش کنید یا، هر جا فقط لایه و خوشه هست، با خطای معیار خوشه‌ای خانوار؛ و جدول جریان نمونه را با شمار رکوردهایی که هر قاعده حذف می‌کند بدهید. برای مدل‌های پیش‌بینی: پیش از هر کار بر پایهٔ خانوار تقسیم کنید، تنها ردیف‌های آموزش را بیش‌نمونه‌گیری کنید، امتیاز وزنی گزارش کنید و هر مدل را یکسان تنظیم کنید. برای داوران: ستارهٔ معناداری از رگرسیون بدون وزن NHTS و سطح زیر منحنی از تقسیم تصادفی ردیف‌های بیش‌نمونه‌گیری‌شده هر یک باید پرسشی برانگیزد.

نتیجه‌گیری

نردبان چهارپلهٔ بازتولیدپذیری با برآورد دوبارهٔ طراحی‌آگاه نگرانی کلی دربارهٔ پژوهش حمل‌ونقل مبتنی بر پیمایش را به عدد تبدیل می‌کند. بر ۲۲ مطالعهٔ منتشرشدهٔ NHTS، عددهای چاپی وقتی برآوردگر اعلام‌شده به کار می‌رفت بازیافتنی بودند؛ خطای معیار زیر طرح پیمایش حدود ۲٫۲ برابر بود؛ ۳۲ درصد ضرایبی که بدون طرح معنادار بودند معنادار نماندند؛ و امتیازهای پیش‌بینی چاپیِ پذیرش خودرو تنها در خط لوله‌ای بازتولید شد که ردیف‌های آزمون به آموزش نشت می‌کنند. واریانس ساندویچی خوشه‌ایِ خانوار بیشتر اصلاح وزن تکرار را بازیافت. این ابزار به‌صورت `pynhts.models` و `pynhts.validate` منتشر شده تا دیگر پژوهشگران همین بررسی را روی تحلیل‌های خود انجام دهند.

اعلان‌ها

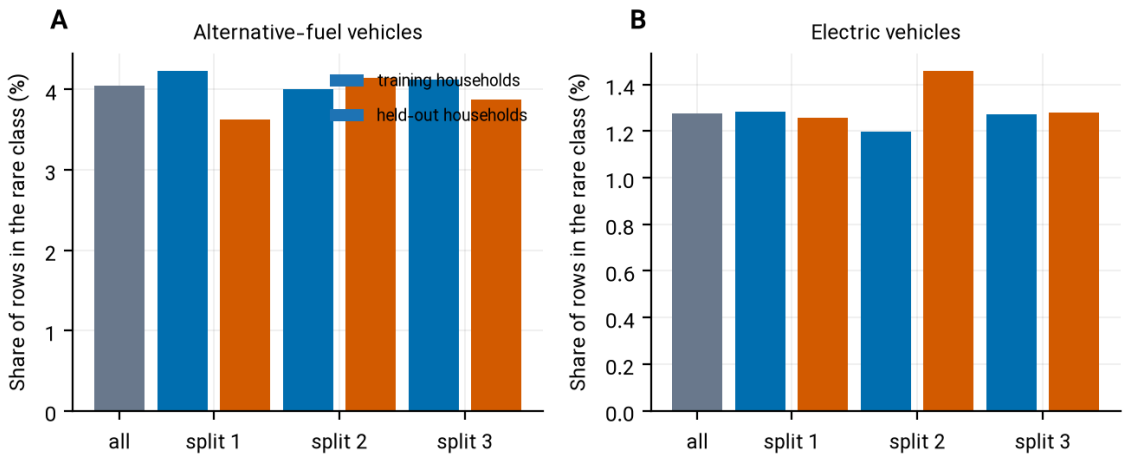
اعلان استفاده از هوش مصنوعی مولد و فناوری‌های یاری‌رسان در تهیهٔ دست‌نوشته. در تهیهٔ این اثر نویسنده از Claude (Anthropic) به‌عنوان دستیار برای برنامه‌ریزی مطالعه، نوشتن و اجرای کد تحلیل، غربالگری منابع، و پیش‌نویس و ویرایش متن استفاده کرد. پس از استفاده از این ابزار، نویسنده مواد تولیدشده (کد، نتایج، منابع و متن) را بازبینی، راستی‌آزمایی و ویرایش کرد و مسئولیت کامل محتوای اثر را می‌پذیرد. عددهای متن را کد از پرونده‌های نتیجه وارد می‌کند و ارجاع‌ها با پایگاه‌های کتابشناختی سنجیده شده‌اند.

تعارض منافع. نویسنده توسعه‌دهندهٔ بستهٔ pynhts است که ماژول‌هایش اینجا شرح شده و نویسندهٔ یکی از مقاله‌های موردی است؛ آن مقاله در فهرست آمد و دوباره برآورد نشد. [[AUTHOR-TODO: جملهٔ استاندارد تعارض منافع و هم‌نویسندگان تأیید شود.]]

منابع مالی. این پژوهش از هیچ نهاد تأمین مالی در بخش دولتی، تجاری یا غیرانتفاعی کمک مالی ویژه‌ای دریافت نکرده است. [[AUTHOR-TODO: تأیید شود.]]

دسترسی به داده و کد. فایل‌های عمومی NHTS از پایگاه پیمایش در دسترس‌اند. کد، جدول‌های نتیجه و بسته در مخزن باز pynhts (https://github.com/sabernaseralavi-60/pynhts) است و با شناسهٔ پایدار در Zenodo بایگانی شده است.

مواد تکمیلی



شکل 8: هیستوگرام فراوانی رستهٔ کمیاب (سهم خودروهای سوخت جایگزین، پنل A، و برقی، پنل B) در همهٔ ردیف‌ها، در خانوارهای آموزش و در خانوارهای کنار گذاشته‌شدهٔ هر یک از سه تقسیم.

جدول 6: ابرپارامترهای برگزیدهٔ مدل‌های تنظیم‌شده در خط لولهٔ تقسیم خانوار (بذر اصلی نخست). C: هزینه یا قدرت جریمه؛ vs: کف واریانس؛ crit ،leaf ،depth: عمق درخت، کمینهٔ اندازهٔ برگ و معیار شکست؛ n، mf: شمار درختان و سهم ورودی در هر شکست؛ gamma: پهنای هسته.

پیامد	مدل	مقدارهای برگزیده	AUC اعتبارسنجی درونی
سوخت جایگزین	LR	penalty_l1 = False C = ۰/۰۲۵۶’	۰/۶۷
سوخت جایگزین	NB	vs = ۱/۷۹e-۰۹	۰/۶۳
سوخت جایگزین	DT	crit = entropy leaf = ۵۰’ depth = ۳’	۰/۶۳
سوخت جایگزین	RF	mf = ۰/۳ leaf = ۱’ depth = ۱۰’ n = ۲۰۰’	۰/۶۵
سوخت جایگزین	SVM	gamma = ۰/۰۰۴۰۷ C = ۲/۵۹’	۰/۶۲
برقی	LR	penalty_l1 = True C = ۰/۰۱۷۵’	۰/۷۵
برقی	NB	vs = ۱/۷۹e-۰۹	۰/۷۲
برقی	DT	crit = entropy leaf = ۵۰’ depth = ۳’	۰/۶۸
برقی	RF	mf = ۰/۳ leaf = ۱’ depth = ۱۰’ n = ۲۰۰’	۰/۷۴
برقی	SVM	gamma = ۰/۰۰۴۰۷ C = ۲/۵۹’	۰/۷۲

کتاب‌نامه

Alberini, A., Teja Burra, L., Cirillo, C., Shen, C., 2021. Counting vehicle miles traveled: What can we learn from the NHTS?. Transportation Research Part D: Transport and Environment 98, 102984.. https://doi.org/10.1016/j.trd.2021.102984

Barua, L., Zou, B., Zhou, Y., Liu, Y., 2021. Modeling household online shopping demand in the U.S.: a machine learning approach and comparative investigation between 2009 and 2017. Transportation 50, 437–476.. https://doi.org/10.1007/s11116-021-10250-z

Birrell, C.L., Steel, D.G., Batterham, M.J., Arya, A., 2019. How to use replicate weights in health survey analysis using the National Nutrition and Physical Activity Survey as an example. Public Health Nutrition 22, 3315–3326.. https://doi.org/10.1017/s1368980019001927

Blagus, R., Lusa, L., 2013. SMOTE for high-dimensional class-imbalanced data. BMC Bioinformatics 14, 106.. https://doi.org/10.1186/1471-2105-14-106

Bollen, K.A., Biemer, P.P., Karr, A.F., Tueller, S., Berzofsky, M.E., 2016. Are Survey Weights Needed? A Review of Diagnostic Tests in Regression Analysis. Annual Review of Statistics and Its Application 3, 375–392.. https://doi.org/10.1146/annurev-statistics-011516-012958

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research 16, 321–357.. https://doi.org/10.1613/jair.953

Conway, M.W., Salon, D., King, D.A., 2018. Trends in Taxi Use and the Advent of Ridehailing, 1995–2017: Evidence from the US National Household Travel Survey. Urban Science 2, 79.. https://doi.org/10.3390/urbansci2030079

Deka, D., Fei, D., 2019. A comparison of the personal and neighborhood characteristics associated with ridesourcing, transit use, and driving with NHTS data. Journal of Transport Geography 76, 24–33.. https://doi.org/10.1016/j.jtrangeo.2019.03.001

Demircioğlu, A., 2024. Applying oversampling before cross-validation will lead to high bias in radiomics. Scientific Reports 14, 11563.. https://doi.org/10.1038/s41598-024-62585-z

Fu, Z., Liu, W., Zhu, J., Yuan, Y., 2020. Trip Chain Simulation of Electric Vehicles based on NHTS Data. 2020 12th IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC) 1–6.. https://doi.org/10.1109/appeec48164.2020.9220726

Gard, A.M., Hyde, L.W., Heeringa, S.G., West, B.T., Mitchell, C., 2023. Why weight? Analytic approaches for large-scale population neuroscience data. Developmental Cognitive Neuroscience 59, 101196.. https://doi.org/10.1016/j.dcn.2023.101196

Gelman, A., 2007. Struggles with Survey Weighting and Regression Modeling. Statistical Science 22, 153–164.. https://doi.org/10.1214/088342306000000691

Hagenauer, J., Helbich, M., 2017. A comparative study of machine learning classifiers for modeling travel mode choice. Expert Systems with Applications 78, 273–282.. https://doi.org/10.1016/j.eswa.2017.01.057

Hardwicke, T.E., Bohn, M., MacDonald, K., Hembacher, E., Nuijten, M.B., Peloquin, B.N., deMayo, B.E., Long, B., Yoon, E.J., Frank, M.C., 2021. Analytic reproducibility in articles receiving open data badges at the journal Psychological Science : an observational study. Royal Society Open Science 8, 201494.. https://doi.org/10.1098/rsos.201494

Hörl, S., Balac, M., 2021. Synthetic population and travel demand for Paris and Île-de-France based on open and publicly available data. Transportation Research Part C: Emerging Technologies 130, 103291.. https://doi.org/10.1016/j.trc.2021.103291

Jia, J., 2019. Analysis of Alternative Fuel Vehicle (AFV) Adoption Utilizing Different Machine Learning Methods: A Case Study of 2017 NHTS. IEEE Access 7, 112726–112735.. https://doi.org/10.1109/ACCESS.2019.2934780

Jia, J., Shi, B., Che, F., Zhang, H., 2020. Predicting the Regional Adoption of Electric Vehicle (EV) With Comprehensive Models. IEEE Access 8, 147275–147285.. https://doi.org/10.1109/ACCESS.2020.3014851

Jiao, J., Bischak, C., Hyden, S., 2020. The impact of shared mobility on trip generation behavior in the US: Findings from the 2017 National Household Travel Survey. Travel Behaviour and Society 19, 1–7.. https://doi.org/10.1016/j.tbs.2019.11.001

Kapoor, S., Cantrell, E.M., Peng, K., Pham, T.H., Bail, C.A., Gundersen, O.E., Hofman, J.M., Hullman, J., Lones, M.A., Malik, M.M., Nanayakkara, P., Poldrack, R.A., Raji, I.D., Roberts, M., Salganik, M.J., Serra-Garcia, M., Stewart, B.M., Vandewiele, G., Narayanan, A., 2024. REFORMS: Consensus-based Recommendations for Machine-learning-based Science. Science Advances 10, eadk3452.. https://doi.org/10.1126/sciadv.adk3452

Kapoor, S., Narayanan, A., 2023. Leakage and the reproducibility crisis in machine-learning-based science. Patterns 4, 100804.. https://doi.org/10.1016/j.patter.2023.100804

Krizek, K.J., McGuckin, N., 2019. Shedding NHTS Light on the Use of 'Little Vehicles' in Urban Areas. Transport Findings.. https://doi.org/10.32866/10777

Lambert, D., 1992. Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. Technometrics 34, 1.. https://doi.org/10.2307/1269547

Lidbe, A., Adanu, E.K., Penmetsa, P., Jones, S., 2021. Changes in the travel patterns of older Americans with medical conditions: A comparison of 2001 and 2017 NHTS data. Transportation Research Interdisciplinary Perspectives 11, 100463.. https://doi.org/10.1016/j.trip.2021.100463

Lumley, T., 2004. Analysis of Complex Survey Samples. Journal of Statistical Software 9.. https://doi.org/10.18637/jss.v009.i08

Martín-Baos, J.Á., López-Gómez, J.A., Rodriguez-Benitez, L., Hillel, T., García-Ródenas, R., 2023. A prediction and behavioural analysis of machine learning methods for modelling travel mode choice. Transportation Research Part C: Emerging Technologies 156, 104318.. https://doi.org/10.1016/j.trc.2023.104318

Naseralavi, S., Soltanirad, M., Ranjbar, E., Lucero, M., Gorzin, F., Hakiminejad, Y., Azimi, S., Baghersad, M., Mazaheri, A., 2025. Machine Learning Modeling of Household Trip Generation by State Using NHTS Data. Urban Science 9, 353.. https://doi.org/10.3390/urbansci9090353

Peng, R.D., 2011. Reproducible Research in Computational Science. Science 334, 1226–1227.. https://doi.org/10.1126/science.1213847

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., Pélissier, R., 2020. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. Nature Communications 11, 4540.. https://doi.org/10.1038/s41467-020-18321-y

Quinn, T.D., Jakicic, J.M., Fertman, C.I., Barone Gibbs, B., 2016. Demographic factors, workplace factors and active transportation use in the USA: a secondary analysis of 2009 NHTS data. Journal of Epidemiology and Community Health 71, 480–486.. https://doi.org/10.1136/jech-2016-207820

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F., Dormann, C.F., 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40, 913–929.. https://doi.org/10.1111/ecog.02881

Rosenblatt, M., Tejavibulya, L., Jiang, R., Noble, S., Scheinost, D., 2024. Data leakage inflates prediction performance in connectome-based machine learning models. Nature Communications 15, 1829.. https://doi.org/10.1038/s41467-024-46150-w

Rust, K., Rao, J., 1996. Variance estimation for complex surveys using replication techniques. Statistical Methods in Medical Research 5, 283–310.. https://doi.org/10.1177/096228029600500305

Sabouri, S., Brewer, S., Ewing, R., 2020. Exploring the relationship between ride-sourcing services and vehicle ownership, using both inferential and machine learning approaches. Landscape and Urban Planning 198, 103797.. https://doi.org/10.1016/j.landurbplan.2020.103797

Sadeghvaziri, E., Tawfik, A., 2020. Using the 2017 National Household Travel Survey Data to Explore the Elderly's Travel Patterns. International Conference on Transportation and Development 2020 86–94.. https://doi.org/10.1061/9780784483169.008

Schweinsberg, M., Feldman, M., Staub, N., Akker, O.R. van den, Aert, R.C. van, Assen, M.A. van, Liu, Y., Althoff, T., Heer, J., Kale, A., Mohamed, Z., Amireh, H., Venkatesh Prasad, V., Bernstein, A., Robinson, E., Snellman, K., Amy Sommer, S., Otner, S.M., Robinson, D., Madan, N., Silberzahn, R., Goldstein, P., Tierney, W., Murase, T., Mandl, B., Viganola, D., Strobl, C., Schaumans, C.B., Kelchtermans, S., Naseeb, C., Mason Garrison, S., Yarkoni, T., Richard Chan, C., Adie, P., Alaburda, P., Albers, C., Alspaugh, S., Alstott, J., Nelson, A.A., Rubia, E. Ariño de la, 2021. Same data, different conclusions: Radical dispersion in empirical results when independent analysts operationalize and test the same hypothesis. Organizational Behavior and Human Decision Processes 165, 228–249.. https://doi.org/10.1016/j.obhdp.2021.02.003

Shirgaokar, M., Nobler, E., 2021. Differences in daily trips between immigrants and US-born individuals: Implications for social integration. Transport Policy 105, 103–114.. https://doi.org/10.1016/j.tranpol.2021.03.008

Silberzahn, R., Uhlmann, E.L., Martin, D.P., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š., Bai, F., Bannard, C., Bonnier, E., Carlsson, R., Cheung, F., Christensen, G., Clay, R., Craig, M.A., Dalla Rosa, A., Dam, L., Evans, M.H., Flores Cervantes, I., Fong, N., Gamez-Djokic, M., Glenz, A., Gordon-McKeon, S., Heaton, T.J., Hederos, K., Heene, M., Hofelich Mohr, A.J., Högden, F., Hui, K., Johannesson, M., Kalodimos, J., Kaszubowski, E., Kennedy, D.M., Lei, R., Lindsay, T.A., Liverani, S., Madan, C.R., Molden, D., Molleman, E., Morey, R.D., 2018. Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results.

Advances in Methods and Practices in Psychological Science 1, 337–356.. https://doi.org/10.1177/2515245917747646

Simonsohn, U., Simmons, J.P., Nelson, L.D., 2020. Specification curve analysis. Nature Human Behaviour 4, 1208–1214.. https://doi.org/10.1038/s41562-020-0912-z

Solon, G., Haider, S.J., Wooldridge, J.M., 2015. What Are We Weighting For?. Journal of Human Resources 50, 301–316.. https://doi.org/10.3368/jhr.50.2.301

Steegen, S., Tuerlinckx, F., Gelman, A., Vanpaemel, W., 2016. Increasing Transparency Through a Multiverse Analysis. Perspectives on Psychological Science 11, 702–712.. https://doi.org/10.1177/1745691616658637

Stodden, V., Seiler, J., Ma, Z., 2018. An empirical analysis of journal policy effectiveness for computational reproducibility. Proceedings of the National Academy of Sciences 115, 2584–2589.. https://doi.org/10.1073/pnas.1708290115

Su, R., Xiao, J., McBride, E.C., Goulias, K.G., 2021. Understanding senior's daily mobility patterns in California using human mobility motifs. Journal of Transport Geography 94, 103117.. https://doi.org/10.1016/j.jtrangeo.2021.103117

Vabalas, A., Gowen, E., Poliakoff, E., Casson, A.J., 2019. Machine learning algorithm validation with a limited sample size. PLOS ONE 14, e224365.. https://doi.org/10.1371/journal.pone.0224365

Wang, S., Mo, B., Zheng, Y., Hess, S., Zhao, J., 2024. Comparing hundreds of machine learning and discrete choice models for travel demand modeling: An empirical benchmark. Transportation Research Part B: Methodological 190, 103061.. https://doi.org/10.1016/j.trb.2024.103061

Wang, X., Shaw, F.A., Mokhtarian, P.L., Watkins, K.E., 2022. Response willingness in consecutive travel surveys: an investigation based on the National Household Travel Survey using a sample selection model. Transportation 50, 2339–2373.. https://doi.org/10.1007/s11116-022-10312-w

Winship, C., Radbill, L., 1994. Sampling Weights and Regression Analysis. Sociological Methods & Research 23, 230–257.. https://doi.org/10.1177/0049124194023002004

Wolter, K.M., 2007. Introduction to Variance Estimation. Statistics for Social, Behavioral Sciences.. https://doi.org/10.1007/978-0-387-35099-8

Wu, X., MacKenzie, D., 2021. Assessing the VMT effect of ridesourcing services in the US. Transportation Research Part D: Transport and Environment 94, 102816.. https://doi.org/10.1016/j.trd.2021.102816

Xue, C., Wu, Q., Sun, M., Bai, P., Chen, Y., 2021. The Interaction between E-Shopping and Shopping Trips: An Analysis with 2017 NHTS. Complexity 2021, 8247158.. https://doi.org/10.1155/2021/8247158

Yang, Y., Diez Roux, A.V., Bingham, C.R., 2011. Variability and seasonality of active transportation in USA: evidence from the 2001 NHTS. International Journal of Behavioral Nutrition and Physical Activity 8, 96.. https://doi.org/10.1186/1479-5868-8-96

Zeileis, A., Kleiber, C., Jackman, S., 2008. Regression Models for Count Data in R. Journal of Statistical Software 27.. https://doi.org/10.18637/jss.v027.i08

Zhang, Y., Zhang, Y., 2018a. Associations between Public Transit Usage and Bikesharing Behaviors in The United States. Sustainability 10, 1868.. https://doi.org/10.3390/su10061868

Zhang, Y., Zhang, Y., 2018b. Examining the Relationship between Household Vehicle Ownership and Ridesharing Behaviors in the United States. Sustainability 10, 2720.. https://doi.org/10.3390/su10082720

Zhang, Y., Zhang, Y., 2018c. Exploring the Relationship between Ridesharing and Public Transit Use in the United States. International Journal of Environmental Research and Public Health 15, 1763.. https://doi.org/10.3390/ijerph15081763

Zhou, Y., Wang, X.(., 2014. Explore the relationship between online shopping and shopping trips: An analysis with the 2009 NHTS data. Transportation Research Part A: Policy and Practice 70, 1–9.. https://doi.org/10.1016/j.tra.2014.09.014