

تعدیل‌هایی برای داده‌های گم‌شده در آمارگیری سرعت خودرو در سوئد

آنیکا ایساکسون[†]

مرکز آمار سوئد

مترجم: مهدی باقرپور سهزایی

مرکز آمار ایران

چکیده. در آمارگیری سرعت خودرو در سوئد، داده‌ها از یک نمونه چند مرحله‌ای از مکان‌های جاده‌ای با استفاده از یک دستگاه اندازه‌گیری که در جاده نصب‌شده است گردآوری می‌شوند. معمولاً برخی از خودروهای عبوری از یک مکان منتخب، مشاهده نمی‌شوند. بنا بر این، پیشنهاد می‌کنیم شدآمد (ترافیک) به رده‌های وزن‌دهی تقسیم شود. مشکل اصلی چگونگی تعدیل تعداد خودروهای مشاهده‌شده با توجه به داده‌های گم‌شده است. پیشنهاد می‌شود خودروهایی که به وسیله این دستگاه جانپی می‌شوند، در درون رده با هم جمع شوند؛ پیشنهاد دیگر، استفاده از وزن‌دهی احتمالاتی آمارهای ثبتی است. مدل‌هایی برای خطاهای موجود در تعداد خودروهای جانپی‌شده و خطاهای موجود در احتمال‌های ثبتی برآوردشده، امکان ارزیابی‌های نظری این پیشنهادها را فراهم می‌سازند. این مدل‌ها به کمک برخی داده‌های تجربی ارزیابی می‌شوند.

۱- مقدمه

برقراری امنیت شدآمد در سوئد با این تصور که در آینده مطلوب، هیچ‌کس در شدآمد

[†] Isaksson, A. (2005). Adjustments for Missing Data in a Swedish Vehicle Speed Survey. *Journal of Official Statistics* 21 (4), 605-615.

جاده‌ای کشته یا به‌طور جدی مجروح نخواهد شد، هدایت می‌شود. برای تبدیل این تصور به واقعیت، منابع زیادی صرف تغییر نگرش‌ها و رفتارهای کاربران جاده‌ها می‌شود. بر مسئولیت کاربران جاده‌ها از جمله رعایت سرعت مجاز به‌شدت تأکید می‌شود. اقدامات معمول برای کاهش سرعت عبارتند از تغییرات فیزیکی محدوده شدآمد (مثلاً، تبدیل تقاطع‌ها به دوربرگردان‌ها) و پیکارهایی در جهت هدایت عموم مردم به‌منظور بررسی نتایج این اقدامات. اداره ملی راه سوئد (Swedish National Road Administration: SNRA) از سال ۱۹۹۶ اقدام به اجرای آمارگیری سالانه از سرعت خودروها کرده است.

در آمارگیری سرعت، داده‌های مربوط به یک نمونه چند مرحله‌ای طبقه‌بندی‌شده از مکان‌های جاده‌ای به‌وسیله یک دستگاه اندازه‌گیری که در جاده نصب‌شده گردآوری می‌شوند. عبور چرخ‌های خودرو در این‌جا به‌صورت ضربه‌هایی ثبت، و این ضربه‌ها بر حسب خودروها با هم جمع می‌شوند. متغیرهای اصلی مطالعه عبارتند از جریان شدآمد، y ، و زمان تردد، z . جریان شدآمد برای یک مکان برابر تعداد خودروهای عبوری است، و زمان تردد، کل زمانی است که همه خودروها برای عبور از این مکان صرف می‌کنند. هدف اصلی این آمارگیری برآورد سرعت متوسط، $R = t_y / t_z$ ، است که t_y و t_z به‌ترتیب مجموع‌های جامعه‌ای y و z هستند.

به‌طور معمول، برخی از خودروهای عبوری از یک مکان منتخب، مشاهده نمی‌شوند. از یک طرف به‌وسیله جانپهنی‌های خودکار دستگاه، از طرف دیگر به‌وسیله کارایی اندازه‌گیری (Measurement Efficiency: ME) — یعنی نسبت ضربه‌های ثبت‌شده که با موفقیت بر حسب خودروها جمع شده‌اند — نشان داده می‌شود که نقص در مشاهده برخی خودروها کم است. مسلماً کم‌شماری خودروها، برآوردگرهای کل را اریب می‌کند، در صورتی که این تأثیر روی برآوردگر R نامشخص است.

فعلاً، از مسئله داده‌های گم‌شده به‌سادگی چشم‌پوشی می‌شود — رهیافتی که از این پس راهبرد ϕ نامیده می‌شود. در این مقاله، که به میزان زیادی مبتنی بر [۴] است، برای تعدیل مربوط به داده‌های گم‌شده در مرحله برآورد دو راهبرد معرفی می‌شوند. هر دو روش برای اجرای آسان طراحی شده‌اند که به شبیه‌سازی‌ها یا جمع‌آوری داده‌های کمکی جدید

نیاز ندارند، بلکه فقط نیاز به اصلاحات جزئی در برنامه‌های رایانه‌ای دارند که اکنون برای برآورد استفاده می‌شوند. یک راهبرد، جانپی‌ها را برای تعدیل‌ها به کار می‌برد؛ راهبرد دیگر، ME را برای همان منظور به کار می‌برد. ولی این دو راهبرد برای ساز و کار آمارهای ثبیتی ثبت خودرو به مدلی مشترک متکی‌اند.

۲- عملیات اصلی آمارگیری

در این قسمت، توضیحات مختصری درباره برخی عملیات مهم آمارگیری از سرعت-انتخاب نمونه، گردآوری داده‌ها و برآورد-ارائه می‌شود.

۲-۱- انتخاب نمونه

به منظور مشاهده، مکان‌های جاده‌ای به وسیله یک طرح سه‌مرحله‌ای با طبقه‌بندی در هر مرحله انتخاب می‌شوند. به‌ویژه، در مرحله آخر، مکان‌های جاده‌ای بر اساس سرعت مجاز و اولویت (جاده اصلی یا فرعی) طبقه‌بندی می‌شوند. در گزارش نهایی آمارگیری، برآوردهای حوزه‌ای برای طبقه‌های مختلف ارائه می‌شوند. برای آسانی کار، به هر حال ترجیح می‌دهیم که تمام طبقه‌بندی‌ها را در این مقاله نادیده بگیریم. این واقعیت که سه واحد بزرگ‌تر در مرحله اول، یک طبقه سرشماری تعریف می‌کنند را نیز نادیده می‌گیریم.

واحدهای نمونه‌گیری اولیه (Primary Sampling Units: PSUs) عبارتند از N_I مرکز جمعیت در سوئد. i امین PSU با برچسب i خود نشان داده می‌شود. بنابراین، مجموعه PSUها را به صورت $U_I = \{1, \dots, i, \dots, N_I\}$ نشان می‌دهیم. مرکز جمعیت $i \in U_I$ به N_{III} ناحیه کوچک، با برچسب $q = 1, \dots, N_{III}$ ، افزاز می‌شود که واحدهای نمونه‌گیری ثانوی (Secondary Sampling Units: SSUs) را نشان می‌دهند. مجموعه SSUهای تشکیل شده از تقسیم مجدد i به صورت $U_{III} = \{1, \dots, q, \dots, N_{III}\}$ نشان داده می‌شود. سرانجام، جاده‌های ناحیه کوچک q در مرکز جامعه i به صورت افزاز شده به N_{IIq} مکان جاده‌ای یک‌متری در نظر گرفته می‌شوند. نمونه s از مکان‌های جاده‌ای از جامعه U از مکان‌های جاده‌ای شهری به روش زیر

انتخاب می‌شود:

مرحله اول- یک نمونه با احتمال متناسب با اندازه PSUها با جایگذاری و با احتمال متناسب با تعداد ساکنان استخراج می‌شود. فرض کنید i_v نشان‌دهنده PSU انتخابی در استخراج v ام، $v = 1, \dots, m_I$ ، و p_{i_v} احتمال انتخاب i_v باشد. بردار PSUهای انتخابی، (i_1, \dots, i_{m_I}) ، نمونه مرتب OS_I را تشکیل می‌دهد.

مرحله دوم- به ازای هر i_v که مولفه‌ای از OS_I است، یک نمونه تصادفی ساده s_{III_v} از SSUهای با حجم n_{III_v} انتخاب می‌شود.

مرحله سوم- به ازای هر ناحیه کوچک $q \in s_{III_v}$ ، یک نمونه تصادفی ساده $s_{i_v,q}$ از مکان‌های جاده‌ای به اندازه $n_{i_v,q}$ استخراج می‌شود.

۲-۲- جمع‌آوری داده‌ها

دستگاه اندازه‌گیری از دو لوله بادی کشیده شده در عرض جاده و متصل به یک تحلیلگر شدآمد (یک رایانه ساده) تشکیل شده است. وقتی چرخ یک خودرو از روی لوله می‌گذرد، فشار هوای آن تغییر می‌کند. دفعات چنین پیشامدها، یا ضربه‌ها، توسط تحلیلگر شدآمد ثبت می‌شوند. تحلیلگر دستگاه اندازه‌گیری از روی جریان ضربه، خودروها را ایجاد و سرعت‌هایی را به آن‌ها منتسب می‌کند.

داده‌های گم‌شده زمانی رخ می‌دهند که ضربه‌های رسیده را نمی‌توان به‌طور واضح به خودروها تبدیل کرد. بر اساس ضربه‌های اضافی، خودروها به وسیله یک شیوه گام به گام اساساً ناتصادفی جانمایی می‌شوند. این شیوه شامل جمع و تفریق ضربه‌هاست، و ممکن است تعدادی خودروی جانمایی شده تولید کند که از تعداد واقعی خودروهای ثبت نشده کم‌تر یا بیش‌تر باشد. بر اساس سرعت‌های خودروهای ثبت شده قبلی، به خودروهای جانمایی شده نیز سرعت‌هایی منتسب می‌شود. ولی در حال حاضر، جانمایی‌ها در برآورد کنار گذاشته می‌شوند. دلیل اصلی این امر آن است که سرعت‌های آن‌ها غیر قابل اعتماد تلقی می‌شوند.

ما از نمادگذاری زیر استفاده می‌کنیم. مجموعه خودروهای عبوری از مکان

جاده‌ای k (در یک دوره زمانی معین) متشکل از y_k خودرو با برچسب $U = 1, \dots, y_k$ است. برای سادگی، فرض می‌کنیم خودروی U ام با برچسب خودش U نشان داده شود. بنابراین، برای این مکان و دوره زمانی مورد نظر، جامعه خودروهای عبوری به صورت $U_k = \{1, \dots, U, \dots, y_k\}$ نشان داده می‌شود. زمان تردد z_k برای مکان k به وسیله $z_k = \sum_{U \in U_k} x_U$ ارائه می‌شود که در آن x_U زمانی است که خودروی U برای عبور از این مکان صرف می‌کند. زیرمجموعه باموفقیت مشاهده شده از U_k به صورت r_k با اندازه n_{r_k} نشان داده می‌شود. تحت راهبرد o ، برآوردگرهای y_k و z_k به ترتیب عبارتند از $\hat{y}_k^{(o)} = n_{r_k}$ و $\hat{z}_k^{(o)} = \sum_{r_k} x_U$.

۳-۲- برآورد کردن با داده‌های کامل

فرض کنید $t_a = \sum_U a_k$ ، که a_k مقدار واقعی متغیر مورد مطالعه a (که می‌تواند y یا z باشد) برای مکان $k \in U$ است. به طور مطلوب، a_k به‌زای تمام $k \in S$ معلوم است. پس، از روی [۱۰]، یک برآوردگر نارایب مبتنی بر طرح از t_a به وسیله

$$(۱) \quad \hat{t}_a = \frac{1}{m_I} \sum_{i=1}^{m_I} \frac{\hat{t}_{mi_v}}{p_{i_v}}$$

ارائه می‌شود که در آن $\hat{t}_{mi_v} = (N_{III_v} / n_{III_v}) \sum_{s_{III_v}} \hat{t}_{mi_v,q}$ و اگر $i \in U_I$ در قرعه U ام انتخاب شود، $\hat{t}_{mi_v,q} = (N_{i_v,q} / n_{i_v,q}) \sum_{s_{i_v,q}} a_k$ آن‌گاه $N_{III_v} = N_{III}$ و $N_{i_v,q} = N_{i_v}$ از روی [۸]، یک برآوردگر تقریباً نارایب مبتنی بر طرح از R به وسیله

$$(۲) \quad \hat{R} = \frac{\hat{t}_y}{\hat{t}_z}$$

ارائه می‌شود. اگر برخی داده‌ها گمشده باشند، a_k ‌های واقعی مجهول‌اند، و برابری‌های (۱) - (۲) دیگر کاربرد نخواهند داشت. به این وضعیت در بخش ۴ می‌پردازیم.

۳- پیشنهادهایی برای تعدیل‌های داده‌های گم‌شده

در این بخش، بحث را با فرمول‌بندی یک مدل برای توزیعی آغاز می‌کنیم که مجموعه خودروهای ثبت‌شده r_k را برای مکان جاده‌ای مشاهده‌شده k تولید می‌کند. سپس این مدل نقطه شروع ساخت برآوردگرهایی می‌شود که خودروهای ثبت‌نشده را به حساب می‌آورد.

۳-۱- یک مدل ثبتي

راهبردهای تعدیل ما هر دو متکی به مدل زیر هستند که از همه جهت‌ها با مدل همگنی پاسخ گروهی در سارندال و سوونسون [۹] یا [۱۰] همسان است.

۳-۱-۱- مدل ثبتي

فرض کنید که خودروهای عبوری از مکان جاده‌ای $k \in S_{i,q}$ در یک روز منتخب به H_k گروه U_{kh} ($h = 1, \dots, H_k$) طوری افراز شوند که با فرض معلوم بودن $S_{i,q}$

- تمام y_{kh} خودرو در گروه U_{kh} دارای احتمال (مجهول) ثبت شدن یکسان $\theta_{kh} > 0$ باشند؛
- ثبت یک خودرو مستقل از بقیه باشد.

شاید تنها راه عملی گروه‌بندی جریان شدآمد به‌وسیله بازه‌های زمانی باشد. در آزمایش ما (ن.ک. بخش ۵)، ساعت کشیک به‌عنوان مبنای تقسیم استفاده شده است.

۳-۲- راهبرد ۱

پیشنهاد نخست ما برای تعدیل داده‌های گم‌شده، استفاده از خودروهای جانهدی شده است. به سرعت‌هایشان اعتماد نمی‌کنیم، بلکه اجازه می‌دهیم برآوردگرهای راهبرد ۱ ما تنها از تعداد آن‌ها استفاده کنند. برای برآوردگر y_k ،

$$(۳) \quad \hat{y}_k^{(1)} = \sum_{h=1}^{H_k} (n_{r_{kh}} + n_{l_{kh}}) = \sum_{h=1}^{H_k} \hat{y}_{kh}^{(1)}$$

را پیشنهاد می‌کنیم که در آن $n_{r_{kh}}$ و $n_{l_{kh}}$ به ترتیب تعداد خودروهای ثبت‌شده و جانمایی‌شده در گروه همگنی U_{kh} هستند. برای برآوردگر z_k

$$(۴) \quad \hat{z}_k^{(1)} = \sum_{h=1}^{H_k} \frac{\sum_{v=1}^{n_{r_{kh}}} x_{v, kh}}{\hat{\theta}_{kh}^{(1)}} = \sum_{h=1}^{H_k} \frac{\sum_{v=1}^{n_{r_{kh}}} x_{v, kh}}{n_{r_{kh}} / \hat{y}_{kh}^{(1)}} = \sum_{h=1}^{H_k} (n_{r_{kh}} + n_{l_{kh}}) \bar{x}_{r_{kh}}$$

را پیشنهاد می‌کنیم که در آن $\bar{x}_{r_{kh}} = \sum_{v=1}^{n_{r_{kh}}} x_{v, kh} / n_{r_{kh}}$.

برآوردگر $\hat{y}_k^{(1)}$ تابعی است از $n_{r_{kh}}$ ها، که ویژگی‌های تصادفی آن‌ها توسط مدل ثبتی تنظیم می‌شود، و $n_{l_{kh}}$ ها، که در اصل موجودی‌های ثابت‌اند (زیرا شیوه جانمایی قطعی است). به هر حال برای سهولت، با $n_{l_{kh}}$ ها نیز مثل متغیرهای تصادفی رفتار می‌کنیم. اگر حق انتخاب داشتیم، θ_{kh} را به جای $\hat{\theta}_{kh}^{(1)}$ ، با نرخ ثبت واقعی $n_{r_{kh}} / y_{kh}$ برآورد می‌کردیم. آن‌گاه، برآوردگر $\hat{z}_k^{(1)}$ از z_k روایت سرشماری برآوردگر وزن‌دهی مستقیم می‌شد (حالت خاصی که مشاهده تمام اعضای جامعه هدف عالی است، و بنا بر این داده‌های گم‌شده تنها منبع تصادفی بودن هستند) که در [۹] یا [۱۰] ارائه شده است. به شرط $s_{i, \alpha}$ ، و مشروط بر این که احتمال یک گروه همگن تهی قابل چشم‌پوشی باشد، آن‌گاه تحت مدل ثبتی، $\hat{z}_k^{(1)}$ برای z_k ناریب خواهد بود.

۳-۳- راهبرد ۲

اگر از خودروهای جانمایی‌شده استفاده نکنیم، به‌منظور تعدیل جریان مشاهده‌شده گزینه‌های معدودی برای داده‌های گم‌شده باقی گذاشته‌ایم. به هر حال یکی از امکانات باقی‌مانده این است که به‌روش مناسبی تعداد خودروهای ثبت‌شده را وزن‌دهی کنیم. برآورد $\hat{\theta}_{kh}^{(1)}$ دیگر یک گزینه نیست، و برآورد دیگری از θ_{kh} لازم است. امکان برآورد کردن احتمال‌ها (ی پاسخ) از روی داده‌های کمکی فقط به‌طور خیلی پراکنده در نوشتگان بحث شده است. برخی منابع نخستین عبارتند از [۱] و [۲]. [۴]

احتمال‌های پاسخ را به‌وسیله رگرسیون لوژستیک مدل‌بندی می‌کنند و آن‌ها را از روی مدل برازش‌شده برآورد می‌کنند. مثلاً روش‌های برآورد ناپارامتری توسط [۵] بحث می‌شوند.

اگر برآوردگرهای تعدیل‌یافته داده‌های گم‌شده y_k و z_k شامل پارامترهای مدل باشند، باید به‌گونه‌ای این پارامترها از روی داده‌های نمونه برآورد شوند. به‌وسیله یافتن یک متغیر کمکی که با θ_{kh} رابطه یک به یک داشته باشد گریز ساده‌ای از این مسئله را امتحان می‌کنیم. ME متغیری است که امیدواریم به بهترین صورت مناسب این شرح باشد. بنا بر این، دومین پیشنهاد ما برای برآوردگر y_k به‌صورت زیر است:

$$(۵) \quad \hat{y}_k^{(r)} = \sum_{h=1}^{H_k} \frac{n_{r_{kh}}}{\hat{\theta}_{kh}^{(r)}} = \sum_{h=1}^{H_k} \frac{n_{r_{kh}}}{(ME)_{kh}} \sum_{h=1}^{H_k} \hat{y}_{kh}^{(r)}$$

که در آن $(ME)_{kh}$ کارایی اندازه‌گیری گروه همگنی U_{kh} است. پیشنهاد متناظر ما برای برآوردگر z_k به‌صورت زیر است:

$$(۶) \quad \hat{z}_k^{(r)} = \sum_{h=1}^{H_k} \frac{\sum_{v=1}^{r_{kh}} x_{vD}}{\hat{\theta}_{kh}^{(r)}} = \sum_{h=1}^{H_k} \frac{\sum_{v=1}^{r_{kh}} x_{vD}}{(ME)_{kh}}$$

برآوردگر $\hat{z}_k^{(r)}$ طبق همان اصول مربوط به $\hat{z}_k^{(1)}$ ساخته می‌شود، فقط با این تفاوت که θ_{kh} به‌جای $\hat{\theta}_{kh}^{(1)}$ ، با $\hat{\theta}_{kh}^{(r)}$ برآورد می‌شود.

پژوهشگاه علوم انسانی و مطالعات فرهنگی

۳-۴- مدل‌های خطا برای $n_{I_{kh}}$ و $\hat{\theta}_{kh}^{(r)}$

فرض کنید $\hat{a}_k^{(c)}$ ، $k \in s_{i,q}$ ، برآوردگر a_k تحت راهبرد c ($c = 0, 1, 2$) باشد. به‌عنوان ابزاری برای ارزیابی ویژگی‌های آماری برآوردگرهای مبتنی بر $\hat{a}_k^{(1)}$ ، مدل خطای زیر را برای $n_{I_{kh}}$ به‌عنوان برآوردگر $y_{kh} - n_{r_{kh}}$ فرمول‌بندی می‌کنیم.

۳-۴-۱- مدل خطای جانپی

فرض کنید بردار تمام $n_{r_{kh}}$ ‌ها در مکان $k \in s_{i,q}$ به‌صورت

برای $\mathbf{n}_{r_k} = (n_{r_k}, \dots, n_{r_{kh}}, \dots, n_{r_{khk}})$ نشان داده شود. با فرض معلوم بودن $s_{i,q}$ و \mathbf{n}_{r_k} ، برای گروه همگنی U_{kh} ($h = 1, \dots, H_k, k \in s_{i,q}$)،

- تعداد $n_{I_{kh}}$ خودروی جانهدی شده متشکل از تعداد خودروهای ثبت‌نشده ضرب در خطای تصادفی ε_{kh} است: $n_{I_{kh}} = (y_{kh} - n_{r_{kh}}) \varepsilon_{kh}$ ؛
- امید ریاضی و واریانس ε_{kh} مستقل از $s_{i,q}$ ، n_{r_k} و مکان جاده‌ای k هستند؛
- $n_{I_{kh}}$ ها مستقل‌اند.

از آن‌جا که هر چه خودروهایی که ثبت نمی‌شوند بیشتر باشند، کار جانهدی دشوارتر، و مخاطره وقوع خطاهای بزرگ بیشتر می‌شود وجود یک مدل خطای جانهدی ضربی به‌نظر معقول است.

برای این‌که بتوانیم ویژگی‌های آماری برآوردگرها را بر اساس $\hat{a}_k^{(r)}$ ارزیابی کنیم، مدل خطای زیر را برای $\hat{\theta}_{kh}^{(r)}$ به‌عنوان برآوردگر θ_{kh} فرمول‌بندی می‌کنیم.

۲-۴-۳- مدل خطا برای $\hat{\theta}_{kh}^{(r)}$

برای گروه همگنی U_{kh} ($h = 1, \dots, H_k, k \in s_{i,q}$)،

- برآوردگر $\hat{\theta}_{kh}^{(r)} = (ME)_{kh}$ تابعی از θ_{kh} و یک خطای تصادفی ε_{kh} است؛
- امید ریاضی و واریانس ε_{kh} مستقل از $s_{i,q}$ ، n_{r_k} و مکان جاده‌ای k هستند؛
- $\hat{\theta}_{kh}^{(r)}$ ها مستقل‌اند.

توجه خود را به دو رابطه تابعی ساده محدود می‌کنیم: مدل خطای

$$\text{جمعی، } \hat{\theta}_{kh}^{(r)} = \theta_{kh} + \varepsilon_{kh}, \text{ و مدل خطای ضربی، } \hat{\theta}_{kh}^{(r)} = \theta_{kh} \varepsilon_{kh}.$$

۴- برآورد کردن با داده‌های گم‌شده

فرض کنید برآوردگر t_a که به‌وسیله جایگذاری $\hat{a}^{(c)}$ به‌جای a در برابری (۱) به دست آمده است با $\hat{t}_{\hat{a}^{(c)}}$ نشان داده شود. ما به ویژگی‌های آماری برآوردگر داده‌های گم‌شده

$$(۷) \quad \hat{R}^{(c)} = \frac{\hat{t}_{\hat{y}^{(c)}}}{\hat{t}_{\hat{z}^{(c)}}}$$

از R علاقه‌مندیم. در این جا توزیع احتمال توأم (به شرط $s_{i,q}$) $\hat{a}_k^{(c)}$ ، مدل ξ نامیده می‌شود. (مدل ξ تحت راهبرد ۰، با مدل ثبتي، تحت راهبرد ۱ با مدل خطای ثبت و جانهي توأم، و تحت راهبرد ۲ با مدل ثبتي توأم و مدل خطا برای $\hat{\theta}_{kh}^{(y)}$ متناظر است.)

فرض کنید E_p و V_p نمایانگر امید ریاضی و واریانس در طرح نمونه‌گیری p توصیف‌شده در بخش ۱.۲ باشند. برای برآوردگرهای غیرخطی، مانند نسبت دو مجموع جامعه‌ای برآوردشده، استفاده از واریانس یک آماره خطی شده به‌عنوان تقریبی برای واریانس دقیق معمول است. فرض کنید AV_p نمایانگر چنین واریانس تقریبی در طرح p است. (برای جزئیات بیش‌تر درباره فن خطی‌سازی، [۱۰]. متناظراً، اگر امیدهای ریاضی و واریانس‌ها نسبت به مدل ξ اختیار شوند با زیروند ξ و اگر به‌طور توأم نسبت به طرح نمونه‌گیری p و مدل ξ اختیار شوند با زیروند $p\xi$ نشان داده می‌شوند.

تحت طرح نمونه‌گیری p و مدل ξ با هم، برآوردگر $\hat{R}^{(c)}$ دارای مقدار امید تقریبی

زیر است:

$$(۸) \quad E_{p\xi}(\hat{R}^{(c)}) \approx \frac{E_{p\xi}(\hat{t}_{\hat{y}^{(c)}})}{E_{p\xi}(\hat{t}_{\hat{z}^{(c)}})} = \frac{\sum_U E_{\xi}(\hat{y}_k^{(c)} | s_{iq})}{\sum_U E_{\xi}(\hat{z}_k^{(c)} | s_{iq})}$$

به‌طور کلی، علامت اربیی $\hat{R}^{(c)}$ به‌عنوان برآوردگر R (همچنین علامت تغییر واریانس ناشی از به کار بردن $\hat{R}^{(c)}$ به‌جای \hat{R}) نامعلوم است. با وجود این حالت مطلوبی که $\hat{y}_k^{(c)}$ و $\hat{z}_k^{(c)}$ به‌ترتیب برای y_k و z_k ناریب می‌شوند را در نظر بگیرید. آن‌گاه، $\hat{R}^{(c)}$ تقریباً برای R ناریب است. همچنین، همان‌طور که در [۶] نشان داده شده است، افزایش واریانس ناشی از به کار بردن $\hat{R}^{(c)}$ به‌جای \hat{R} به‌عنوان برآوردگر R به‌صورت زیر ارائه می‌شود:

$$(9) \quad AV_{p\xi}(\hat{R}^{(c)}) - AV_p(\hat{R}) = \frac{1}{t_z^y} \frac{1}{m_I} \sum_{i=1}^{N_I} \frac{1}{p_i} \frac{N_{III}}{n_{III}} \sum_{U_{III}} \frac{N_{Iq}}{n_{Iq}} \sum_{U_{Iq}} V_\xi(\hat{y}_k^{(c)} - R^{(c)} \hat{z}_k^{(c)} | S_{Iq}).$$

فرض کنید که مدل ثبتي برقرار است. اگر ε_{kh} دارای امید ریاضی یک باشد، آن‌گاه تحت راهبرد ۱، $\hat{y}_k^{(c)}$ و $\hat{z}_k^{(c)}$ به ترتیب برای y_k و z_k ناریب‌اند. برآوردگرهای z_k و y_k که نسبت‌های متغیرهای تصادفی بودند، تحت راهبرد ۲ به‌طور نظری پیچیده‌اند. اگر $\hat{\theta}_{kh}^{(r)}$ به‌وسیله تقریب درجه اول تیلور برای θ_{kh} ناریب باشد، برآوردگرهای $\hat{y}_k^{(r)}$ و $\hat{z}_k^{(r)}$ به هر حال به‌ترتیب برای y_k و z_k تقریباً ناریب‌اند.

تحت همه راهبردها، بیان سرراست $V_\xi(\hat{y}_k^{(c)} - R \hat{z}_k^{(c)} | S_{Iq})$ کاملاً پیچیده است. برای جزئیات بیشتر، به [۶] نگاه کنید.

۵- مطالعه تجربی

به‌منظور ارزیابی مدل‌های خطای ارائه‌شده در بخش ۴.۳، بررسی رفتارهای تجربی $\hat{a}_k^{(c)}$ ، آزمایشی در تابستان ۲۰۰۱ انجام دادیم. داده‌های مربوط به پنج مکان جاده‌ای در شهر لینگوینگ سوئد جمع‌آوری شدند. این مکان‌ها برای بازنمایی انواع گوناگون محدوده‌های شدآمد انتخاب شدند. ولی برای سادگی، مطالعه به خیابان‌های دوطرفه دوباندی با سرعت مجاز ۵۰ کیلومتر در ساعت، محدود شده بود.

۵-۱- جمع‌آوری داده‌ها

در هر مکان منتخب، داده‌ها طی ۲۴ ساعت متوالی با استفاده از دو جفت لوله بادی و سه تحلیلگر شدآمد جمع‌آوری شدند. یک جفت از لوله‌ها، متصل به تحلیلگر شدآمد M ، برای مشاهده همزمان خودروهای عبوری از هر دو باند استفاده شد. جفت دوم لوله‌ها به‌صورت موازی با جفت اول، تنها با اندک جابه‌جایی کنارین، نصب شده بود. این لوله‌ها در خط وسط خیابان با استفاده از شیرفلکه‌هایی بسته شده بودند: شیوه‌ای که می‌تواند اندازه‌گیری جداگانه هر باند را میسر سازد. انتهای لوله‌ها در هر طرف شیرفلکه‌ها به‌ترتیب به

تحلیلگرهای شدامد M_1 و M_2 وصل شدند.

روش بستن به‌عنوان ابزار بهبود کیفیت داده‌ها در SNRA توسعه داده شد. کار ثبت M_1 و M_2 آسان‌تر از کار M_0 (و بنا بر این کمتر در معرض خطاهای اندازه‌گیری) است: خودروها همزمان از روی لوله‌ها عبور نمی‌کنند، خودروهای کمتری عبور می‌کنند، و جهت‌شان از قبل معلوم است. با وجود این، در آمارگیری سرعت از این روش به‌ندرت استفاده می‌شود. دلیل اصلی این است که استفاده از آن نسبت به روش بسته نشدن زمان برتر است؛ باید شیرفلکه‌ها در لوله‌ها تعبیه شوند و خواباندن لوله‌ها مراقبت بیش‌تری را می‌طلبد. آسیب‌پذیری شیرفلکه‌ها عیب دیگر این روش است. مثلاً، اگر شیرفلکه‌ای با آب باران پر شود، یا توسط چرخ خودرو فشرده شود، ممکن است از کار بیفتد.

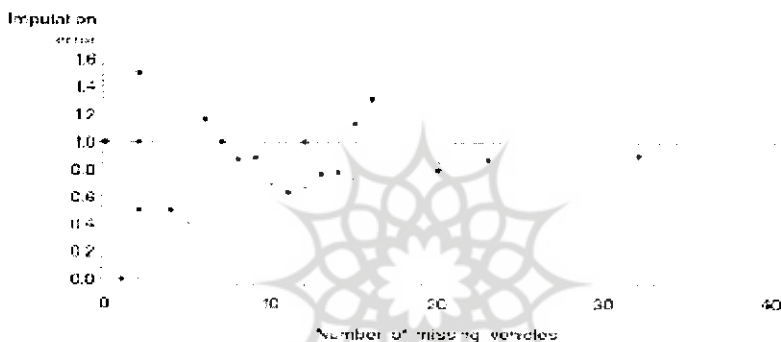
۲-۵- تحلیل

در تحلیل‌هایمان، به‌ازای هر مکان، مجموعه داده‌های تولیدشده توسط M_0 ، خروجی حاصل از یک اندازه‌گیری منظم را نشان می‌دهد. از سوی دیگر، مجموعه داده‌های با هم تولیدشده توسط M_1 و M_2 حقایق را نشان می‌دهد.

اگر مرحله جمع‌آوری داده‌ها خیلی خوب انجام شده بود، مجموعه داده‌های حاصل از M_1 و M_2 کامل بودند. برخی داده‌های گم‌شده، و در نتیجه برخی خودروهای جانپوشیده نیز به هر حال در اثر اندازه‌گیری‌های شیرفلکه به وجود آمد. در حالت‌های خاص، جانپوشی‌ها در داده‌های M_1 یا M_2 می‌توانستند با خودروهایی جور شوند که توسط M_0 به درستی ثبت شده بودند. بروز این موقعیت‌ها زمانی محتمل‌تر بود که چرخ‌های خودروهای عبوری در طرفین شیرفلکه‌ها قرار می‌گرفت. به‌ازای هر مکان، فایل‌های داده‌های حاصل از M_0 ، M_1 و M_2 را در جستجوی جانپوشی‌هایی در M_1 و M_2 مقایسه کردیم که با اطمینان قابل قبولی می‌توانستند با خودروهای ثبت‌شده در M_0 جور شوند. سپس خودروهای ثبت‌شده جایگزین این جانپوشی‌ها شدند. چون این جورسازی همواره ممکن نبود، ما تحلیل‌هایمان را با جانپوشی‌های مانده که نگه داشته و حذف می‌شوند انجام می‌دهیم (به عبارت دیگر، با دو مجموعه از مقادیر «واقعی» کار می‌کنیم).

۱-۲-۵- مدل خطای ضربی برای جانپی

وقتی ε_{kh} های مشاهده شده در مقابل تعداد خودروهای گم‌شده رسم می‌شوند، برای برخی از مکان‌ها با افزایش تعداد خودروهای گم‌شده واریانس خطا گرایش به کاهش دارد (که متناقض با مدل است) - در شکل ۱ مثالی ارائه شده است. در هر حال به دلیل کمبود مشاهدات برای تعدادهای زیاد خودروهای گم‌شده، هر گونه نتیجه‌گیری‌های قطعی سخت است. وقتی خطاها در مقابل تعداد خودروهای ثبت شده رسم می‌شوند، هیچ نوع ساختارهای غیر عادی پیدا نیست.



شکل ۱- خطاهای جانپی برآورد شده در مقابل تعداد « واقعی » خودروهای گم‌شده برای مکان ۴. یک نقطه‌ی داده‌ها متناظر با یک ساعت است. نقطه‌های گرد نشان می‌دهند که مقادیر « واقعی » همراه جانپی‌های حذف شده استفاده می‌شوند؛ علامت‌های مجموع نشان می‌دهند که مقادیر « واقعی » همراه جانپی‌های نگه داشته شده استفاده می‌شوند.

برای بررسی این که آیا واریانس خطاها مستقل از مکان است یا نه (همان‌طور که مدل بیان می‌کند)، یک مدل ANOVA فرمول‌بندی می‌کنیم:

$$(10) \quad \hat{\varepsilon}_{kh} = \alpha + \beta_k + e_{kh} \quad \begin{cases} k = 1, 2, \dots, b \\ h = 1, 2, \dots, c \end{cases}$$

که b تعداد مکان‌های آزمایش، و c تعداد ساعات مشاهده شده در مکان است. در عمل، $b = 5$ و $c = 24$. پارامتر α میانگین کلی است، β_k اثر تصادفی مکان k ام است، و e_{kh} خطای تصادفی است. فرض می‌کنیم که β_k ها به صورت نرمال و به‌طور مستقل با

میانگین صفر و واریانس σ_{β}^2 و e_{kh} ها به صورت نرمال و به طور مستقل با میانگین صفر و واریانس σ_e^2 توزیع می‌شوند، و این که β_k و e_{kh} مستقل اند. در واقع پیش فرض مدل اثرهای تصادفی (مثلاً، [۷]) این است که مکان‌های آزمایش ما به طور تصادفی از تمام مکان‌های ممکن (تمام سنجش‌گرهای جاده‌ای شهری در سوئد) انتخاب شده‌اند. پس، استنباط درباره تمام مکان‌ها امکان‌پذیر است. در مورد بررسی ما، از آن جا که مکان‌ها به طور هدفمند انتخاب شده‌اند باید نتایج را با احتیاط تفسیر کنیم.

وقتی فرض $H_0: \sigma_{\beta}^2 = 0$ در مقابل $H_1: \sigma_{\beta}^2 > 0$ را آزمون می‌کنیم، نتیجه‌گیری‌های ما به‌ازای تیمارهای مختلف از جانهی‌ها در اندازه‌گیری‌های ۶۷ فرق می‌کنند. اگر جانهی‌ها حذف شوند، فرض صفر در سطح معنی‌داری ۰/۰۵ رد نمی‌شود. از سوی دیگر اگر جانهی‌ها را داشته باشیم، فرض صفر رد می‌شود. بنا بر این، ما در مورد این که آیا بین مکان‌ها تغییرپذیری وجود دارد یا نه به گواه روشنی نمی‌رسیم. به‌علاوه، ما به میانگین $\mu_{\hat{\epsilon}} = \alpha$ از $\hat{\epsilon}_{kh}$ علاقه‌مندیم. وقتی $H_0: \mu_{\hat{\epsilon}} = 1$ در مقابل $H_1: \mu_{\hat{\epsilon}} \neq 1$ را آزمون می‌کنیم، اگر جانهی‌ها حذف شوند، فرض صفر در سطح معنی‌داری ۰/۰۵ رد نمی‌شود. از سوی دیگر اگر جانهی‌ها را داشته باشیم، این فرض رد می‌شود. بنا بر این، یک سؤال باز باقی‌می‌ماند که آیا تعداد خودروهای جانهی‌شده برای تعداد واقعی خودروهای گم‌شده به‌طور مشروط ناریب است یا نه.

۲-۲-۵- مدل خطا برای $\hat{\theta}_{kh}^{(r)}$

تحت دو مدل خطای جمعی و ضربی برای $\hat{\theta}_{kh}^{(r)}$ ، به‌نظر می‌رسد ϵ_{kh} ها مستقل از احتمال‌های « واقعی » و نیز تعداد خودروهای ثبت شده باشند.

هر دو مدل خطای جمعی و ضربی بیان می‌کنند که واریانس خطاها مستقل از مکان است. برای بررسی این موضوع، از همان مدل ANOVA در برابری ۱۰- فقط با تعویض $\hat{\epsilon}_{kh}$ با $\hat{\epsilon}_{kh} - \bar{\epsilon}_{kh}$ استفاده می‌کنیم. باز هم هدف، آزمون کردن فرض $H_0: \sigma_{\beta}^2 = 0$ در مقابل $H_1: \sigma_{\beta}^2 > 0$ است. جانهی‌ها در اندازه‌گیری‌های شیرفلکه به هر شکلی تیمار شوند، چه مدل جمعی و چه ضربی باشد، فرض صفر در سطح معنی‌داری ۰/۰۵ رد می‌شود. به

عبارت دیگر، بر خلاف آنچه مدل‌های ما می‌گویند، به نظر می‌رسد یک تغییرپذیری ناشی از مکان در خطای $\hat{\theta}_{kh}^{(r)}$ وجود دارد. این مخالفت با مدل مستلزم بررسی بیش‌تری است. ما همچنین آزمون کردیم که آیا $\hat{\theta}_{kh}^{(r)}$ برای θ_{kh} نارایب است. در سطح معنی‌داری ۰/۰۵ برای مدل خطای جمعی، فرض $\mu_{\varepsilon} = 0$ (در مقابل $\mu_{\varepsilon} \neq 0$) رد نشد. همچنین، برای مدل ضربی، فرض $\mu_{\varepsilon} = 1$ (در مقابل $\mu_{\varepsilon} \neq 1$) رد نشد. جانیهی‌ها در اندازه‌گیری‌های شیرفلکه به هر شکلی تیمار شوند، این نتایج برقرارند.

۳-۲-۵- رفتار تجربی برآوردهای u_k و z_k ، y_k

داده‌های محدود ما اجازه نمی‌دهند عملکردهای درازمدت برآوردهای جریان شدآمد و زمان تردد را مطالعه کنیم، اما می‌توانیم نشانه‌ای از آن‌ها را ارائه کنیم. در جدول ۱، برآوردهای z_k ، y_k و $u_k = y_k / z_k$ به وسیله رقم‌های « واقعی » استانداردسازی شده‌اند (متوسط‌ها روی پنج مکان مورد مطالعه به دست آمده‌اند). برآوردهای حاصل از استانداردسازی به وسیله مقادیر « واقعی » با حذف جانیهی‌ها در ستون « بدون جانیهی‌ها » نشان داده شده‌اند؛ برآوردهای حاصل از استانداردسازی به وسیله مقادیر « واقعی » با ماندن جانیهی‌ها در ستون « با جانیهی‌ها » نشان داده شده‌اند. برآوردهای z_k تنها به وسیله مقادیر « واقعی » با حذف جانیهی‌ها استانداردسازی شده‌اند، زیرا به زمان‌های تردد خودروهای جانیهی شده یقین نداریم.

همان‌طور که انتظار می‌رود، برآوردهای راهبرد ۰ از y_k و z_k همه زیر یک هستند، در صورتی که برآوردهای تعدیل‌شده داده‌های گم‌شده تحت راهبرد ۱ و راهبرد ۲ خیلی خوب به نظر می‌رسند. برای هر دو راهکار، با توجه به این‌که چه معیاری برای استانداردسازی استفاده می‌شود (آیا اندازه‌گیری‌های شیرفلکه شامل جانیهی‌ها هست یا نه)، برآوردهای متوسط کمی زیر یا بالای یک هستند. در کل، اصلاً معلوم نیست که کدام راهکار تعدیل (۱ یا ۲) بایستی توصیه شود.

نسبت u_k عنصر به عنصر با R نظیر است. همه برآوردهای استاندارد شده در جدول ۱ خیلی به یک نزدیک‌اند. به‌طور رسمی، نمی‌توانیم از این برآوردها برای ارزیابی عملکردهای برآوردهای موجود یا پیشنهادی R استفاده کنیم. هر چند به‌صورت اشاره‌ای کوچک، نتیجه می‌گیریم که هنگام برآورد R تعدیل داده‌های گم‌شده ضرورت ندارد.

جدول ۱- متوسط برآوردهای استاندارد شده y_k ، Z_k و u_k

پارامتر	بدون جانپها			با جانپها		
	راهبرد			راهبرد		
	۲	۱	۰	۲	۱	۰
y_k	۱/۰۱۱۹۶	۱/۰۱۵۳۰	۰/۹۴۶۰۱	۰/۹۹۲۹۵	۰/۹۳۰۱۲	۰/۹۹۶۳۱
Z_k	۱/۰۰۸۲۱	۱/۰۱۱۴۲	۰/۹۴۰۳۹	-	-	-
u_k	۱/۰۰۱۴۶	۱/۰۰۱۹۹	۱/۰۱۳۰۸	۰/۹۸۵۷۶	۰/۹۹۷۲۰	۰/۹۸۶۲۹

۶- مطالعه شبیه‌سازی

آزمایش تشریح‌شده در بخش ۵ به تعداد کمی از مکان‌های جاده‌ای محدود شده بود و بنا بر این اصلاً نتایج قاطعی ارائه نکرد. از آن‌جا که منابع لازم برای اجرای آزمایشی بزرگ‌تر را نداشتیم، کوشش کردیم بیش‌تر داده‌هایمان را از طریق استفاده آن‌ها در یک مطالعه شبیه‌سازی تولید کنیم. از روی داده‌های «واقعی» خودروها (با جانپهای حذف‌شده)، برای مکان‌های جاده‌ای شماره ۱، ۲، ۴، و ۵ (مکان ۳ به‌خاطر وجود تعداد زیاد داده‌های گم‌شده در اندازه‌گیری‌های شیرفلکه از مطالعه خارج شده بود)، از نمونه‌گیری برنولی طبقه‌بندی‌شده (STBE) جهت انتخاب خودروها استفاده شده بود. طبقه‌بندی برحسب ساعت کشیک بود، و شیوه نمونه‌گیری به‌طور مستقل و با جایگذاری $D=1000$ مرتبه تکرار شد. قرار بود مجموعه خودروهای به دست آمده در یک تکرار مفروض یک زیرمجموعه با موفقیت مشاهده‌شده از همه خودروهای عبوری از این مکان را نشان دهد. درون ساعت کشیک h و مکان جاده‌ای k ، سه احتمال انتخاب θ_{kh} به کار رفته بودند:

نسبت p_{kh} خودروهای ثبت‌شده در آزمایش اصلی؛ $0/9 p_{kh}$ و $0/8 p_{kh}$ به‌علاوه، در مکان ۲ (مکان دارای بیش‌ترین جریان شدآمد: نزدیک به ۱۵۰۰۰ خودرو در دوره ۲۴ ساعته مورد مشاهده)، بدینانه‌ترین سناریو را آزمودیم: احتمال‌های ثبت $0/8 p_{kh}$ در ساعات غیر اوج (که به‌صورت ساعات با جریان شدآمد « واقعی » کمتر از ۱۰۰۰ خودرو تعریف می‌شود) و $0/5 p_{kh}$ در غیر این ساعت‌ها.

گیریم $\hat{y}_{hkd}^{(c)}$ جریان شدآمد برآوردشده برای ساعت کشیک h ، مکان جاده‌ای k و تکرار d تحت راهبرد c است و گیریم $\hat{y}_{kd}^{(c)} = \sum_{h=1}^{T^*} \hat{y}_{hkd}^{(c)}$. برآوردهای زمان تردد $\hat{z}_{kd}^{(c)}$ و $\hat{z}_{kd}^{(c)}$ به‌طور متناظر تعریف می‌شوند. به‌ازای هر مکان و گزینه احتمال ثبت و برای راهبرد ۲ و $c = 0$ ، را u_k به‌وسیله

$$(11) \quad \hat{u}_k^{(c)} = \frac{1}{D} \sum_{d=1}^D \hat{u}_{kd}^{(c)}$$

برآورد می‌کنیم که $\hat{u}_{kd}^{(c)} = \hat{y}_{kd}^{(c)} / \hat{z}_{kd}^{(c)}$. مقدار تقریبی واریانس $\hat{u}_k^{(c)}$ به‌صورت زیر محاسبه می‌شود:

$$(12) \quad V(\hat{u}_k^{(c)}) = \frac{1}{D} \sum_{d=1}^D (\hat{u}_{kd}^{(c)} - \hat{u}_k^{(c)})^2$$

چون هیچ راهی نداریم که از تعداد خودروهایی که هر ساعت در هر مجموعه داده‌ها به‌جای خودروهای گم‌شده جانپی شده‌اند، اطلاع پیدا کنیم راهبرد ۱ در این مطالعه حذف شده است. در برآوردهای راهبرد ۲ از u_k ، مجبور به استفاده از یک مقدار ناهموار برای ME هستیم: ME در صورتی حاصل می‌شود که هر خودروی گم‌شده دقیقاً به تعداد چرخ‌های ضربه تولید کند.

این برآوردها در جدول ۲ نشان داده شده‌اند. می‌بینیم که همه برآوردهای u_k خیلی به مقدار واقعی نزدیک‌اند، و این که تغییرپذیری در برآوردها خیلی کم است (حتی در بدینانه‌ترین سناریوی ما).

۷- خلاصه

در آمارگیری سرعت دو راهبرد ممکن برای تعدیل‌های داده‌های گم‌شده مطرح کرده‌ایم. در بررسی ویژگی‌های نظری برآوردهای حاصل، از چندین مدل استفاده می‌کنیم. به‌نظر می‌رسد اکثر فرض‌های مدل به‌طور منطقی با داده‌های آزمایشی ما به‌خوبی سازگار باشند. همچنین، به‌نظر می‌رسد برآوردهای پیشنهادی منجر به برآوردهای t_x و t_z با اریبی کمتر نسبت به برآوردهای تعدیل‌نشده امروزی می‌شوند. اگرچه این موضوع لزوماً در مورد R درست نیست: ممکن است برآوردهای تعدیل‌نشده موجود از سرعت متوسط به‌طوری حیرت‌انگیز در برابر اریبی ناشی از داده‌های گم‌شده مقاوم باشد. بنا بر این نتیجه می‌گیریم که وقتی هدف ما برآورد کل مسافت پیموده‌شده خودرو یا کل زمان تردد است، باید برای تعدیل داده‌های گم‌شده از راهبرد ۱ یا راهبرد ۲ استفاده شود، در حالی که وقتی هدف ما برآورد سرعت متوسط است، تعدیل‌های داده‌های گم‌شده ضروری به‌نظر نمی‌رسند (راهبرد ۰ برای برآورد سرعت متوسط درست به‌خوبی راهبردهای ۱ یا ۲ است).

جدول ۲- برآوردهای u_k تحت راهبرد ۰ و راهبرد ۲، به‌تفکیک مکان، بر اساس ۱۰۰۰ نمونه STBE. * $8p_{kh}$ در ساعات غیر اوج؛ و $5p_{kh}$ در غیر این ساعات‌ها

شماره‌ی مکان	θ_{kh}	u_k	$\hat{u}_k^{(0)}$	$V(\hat{u}_k^{(0)})$	$\hat{u}_k^{(2)}$	$V(\hat{u}_k^{(2)})$
۱	p_{kh}	۵۲/۴۴	۵۲/۴۴	۰/۰۰۰۱۱	۵۲/۴۴	۰/۰۰۰۱۲
۱	$9p_{kh}$	۵۲/۴۴	۵۲/۴۴	۰/۰۰۱۶۲	۵۲/۴۴	۰/۰۰۱۶۳
۱	$8p_{kh}$	۵۲/۴۴	۵۲/۴۴	۰/۰۰۳۳۴	۵۲/۴۴	۰/۰۰۳۳۰
۲	p_{kh}	۵۲/۱۸	۵۲/۲۶	۰/۰۰۰۳۸	۵۲/۱۸	۰/۰۰۰۳۸
۲	$9p_{kh}$	۵۲/۱۸	۵۲/۲۶	۰/۰۰۱۱۵	۵۲/۱۸	۰/۰۰۱۱۳
۲	$8p_{kh}$	۵۲/۱۸	۵۲/۲۶	۰/۰۰۲۰۷	۵۲/۱۸	۰/۰۰۱۹۸
۲	*	۵۲/۱۸	۵۲/۶۸	۰/۰۰۳۳۳	۵۲/۱۸	۰/۰۰۴۷۸
۴	p_{kh}	۴۵/۳۷	۴۵/۳۷	۰/۰۰۰۰۸	۴۵/۳۷	۰/۰۰۰۰۸
۴	$9p_{kh}$	۴۵/۳۷	۴۵/۳۷	۰/۰۰۰۵۳	۴۵/۳۷	۰/۰۰۰۵۲
۴	$8p_{kh}$	۴۵/۳۷	۴۵/۳۷	۰/۰۰۱۱۵	۴۵/۳۷	۰/۰۰۱۱۲
۵	p_{kh}	۵۴/۳۱	۵۴/۳۲	۰/۰۰۰۱۳	۵۴/۳۱	۰/۰۰۰۱۳
۵	$9p_{kh}$	۵۴/۳۱	۵۴/۳۲	۰/۰۰۰۶۳	۵۴/۳۱	۰/۰۰۰۶۱
۵	$8p_{kh}$	۵۴/۳۱	۵۴/۳۲	۰/۰۰۱۳۰	۵۴/۳۱	۰/۰۰۱۲۶

توضیح

^۱ در عمل، x_{ij} ها به صورت معکوس سرعت‌های ثبت‌شده محاسبه می‌شوند.

مرجع‌ها

- [1] Chapman, D.W. (1976). A Survey of Nonresponse Imputation Procedures. *Proceedings of the American Statistical Association*, Social Statistics Section, 245–251.
- [2] Drew, J.H.; Fuller, W.A. (1980). Modeling Nonresponse in Surveys with Callbacks. *Proceedings of the American Statistical Association*, Section on Survey Research Methods, 639–642.
- [3] Drew, J.H.; Fuller, W.A. (1981). Nonresponse in Complex Multiphase Surveys. *Proceedings of the American Statistical Association*, Section on Survey Research Methods, 623–628.
- [4] Ekholm, A.; Laaksonen, S. (1991). Weighting via Response Modeling in the Finnish Household Budget Survey. *Journal of Official Statistics* 7, 325–337.
- [5] Giommi, A. (1987). Nonparametric Methods for Estimating Individual Response Probabilities. *Survey Methodology* 13, 127–134.
- [6] Isaksson A. (2003) *Survey Models for a Vehicle Speed Survey*, Doctoral Thesis. Linköping Studies in Statistics No.2, Linköping University, Sweden.
- [7] Montgomery, D.C. (1997). *Design and Analysis of Experiments. 4th Edition*: Wiley. New York.
- [8] Raj, D. (1968). *Sampling Theory*. McGraw-Hill. New York:
- [9] Särndal, C.-E.; Swensson, B. (1987). A General View of Estimation for Two Phases of Selection with Applications to Two-Phase Sampling and Nonresponse. *International Statistical Review* 55, 279–294.
- [10] Särndal, C.-E.; Swensson, B.; Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer. New York.



شېرشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی