مجله تخصصی اپیدمیولوژی ایران؛ ۱۳۹۳، دوره ۱۰، شماره ۳: صفحات ۹۱-۸۳.

مقاله مروری

مروری بر تکنیک‌های غیر مستقیم پاسخ‌گویی به سؤال‌های حساس دو حالتی

مرتضی بنای‌جدی[۱]، سودابه نواده[۲]، اکبر فتوحی[۳]، کامران یزدانی[۴]

[۱] دانشجوی کارشناسی ارشد اپیدمیولوژی، گروه اپیدمیولوژی و آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی تهران، تهران، ایران

[۲] دانشجوی دکتری اپیدمیولوژی، گروه اپیدمیولوژی و آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی تهران، تهران، ایران

[۳] استاد اپیدمیولوژی، گروه اپیدمیولوژی و آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی تهران، تهران، ایران

[۴] استادیار اپیدمیولوژی، گروه اپیدمیولوژی و آمار زیستی، دانشکده بهداشت، دانشگاه علوم پزشکی تهران، تهران، ایران

نویسنده رابط: کامران یزدانی، نشانی: تهران،دانشگاه علوم پزشکی تهران، دانشکده بهداشت، گروه اپیدمیولوژی و آمار زیستی، تلفن: ۸۸۹۵۱۴۰۶- ۰۲۱، دورنگار: ۸۸۹۸۹۱۲۷- ۰۲۱

پست الکترونیک: kyazdani@tums.ac.ir

تاریخ دریافت: ۹۲/۷/۱۴؛ پذیرش: ۹۳/۲/۶

پرسش‌های حساس در مورد برخی از رفتارهای افراد، اغلب در مطالعه‌های علوم انسانی، اجتماعی و پزشکی مطرح می‌شود. به دست آوردن پاسخ‌های معتبر از این گونه سؤال‌ها، از اصلی‌ترین چالش‌های مطرح در این گونه مطالعه‌ها می‌باشد. مهم‌ترین عوامل تهدید کننده اعتبار چنین مطالعه‌هایی تورش‌های پاسخ‌دهی نادرست و عدم پاسخ‌دهی است، که معمولاً به خاطر ترس از افشا شدن اطلاعات شخصی رخ می‌دهد. بدین منظور، پژوهشگران به دنبال راه‌کارهایی غیر از پرسش مستقیم هستند تا با کاهش این نوع تورش‌ها، موجب افزایش اعتبار نتایج حاصل از مطالعه‌ها شوند.

در این مقاله مروری، در خصوص مدل‌های وارنر، طرح سؤال نامرتبط، پاسخ‌دهی اجباری، مورس و منگات از تکنیک پاسخ‌دهی تصادفی شده و در مورد مدل‌های مثلثی (تری انگولار) و متقاطع (کراس وایز) از تکنیک پاسخ‌دهی غیر تصادفی شده، که همگی برای پاسخ‌گویی به سؤال‌های حساس دو حالتی به کار می‌روند، بحث می‌شود.

واژگان کلیدی: سؤال حساس، تکنیک پاسخ‌دهی تصادفی شده، تکنیک پاسخ‌دهی غیر تصادفی شده

مقدمه

موضوعات حساس اجتماعی یا پرسش‌های بسیار محرمانه در مورد برخی از رفتارهای افراد، اغلب در مطالعه‌های علوم انسانی، اجتماعی و پزشکی مطرح می‌شود. سقط غیر قانونی جنین، تولد ناخواسته نوزاد، ابتلا به بیماری ایدز و سوء مصرف دارو، نمونه‌هایی از این موضوعات هستند (۱،۲).

از مهم‌ترین چالش‌های موجود در نتایج این گونه مطالعه‌ها، نامعلوم بودن میزان پاسخ‌دهی‌های درست است. معمولاً وقتی افراد مستقیماً در مقابل سؤال‌های حساس قرار می‌گیرند، به دلایلی مانند ترس از خطر افشا شدن اطلاعات و احساس ورود به حریم خصوصی‌شان، خواسته یا ناخواسته، از ارایه اطلاعات شخصی خودداری می‌کنند که این امر می‌تواند زمینه‌ساز بروز تورش‌هایی به دلیل پاسخ‌دهی نادرست و عدم پاسخ‌دهی[۱] شده و اعتبار نتایج مطالعات را به مخاطره بیاندازد (۳). به منظور مقابله با چنین چالش‌هایی، پژوهشگران همواره به دنبال یافتن راه‌کارهایی هستند تا بتوانند با کاهش این نوع تورش‌ها، موجب افزایش اعتبار نتایج حاصل از مطالعه‌ها گردند. یکی از راه‌کارهای مطرح در این زمینه، استفاده از گزینه شانس و تصادف در پاسخ افراد شرکت کننده در مطالعه است. برای این‌که پژوهشگران بتوانند اطلاعات مفیدی از پاسخ‌های افراد به دست آورند، می‌توانند گزینه شانس را به پاسخ‌های آن‌ها اضافه نموده و سپس با انجام تجزیه و تحلیل‌های خاص، جواب سؤال‌های خود را بیابند.

فرض نمایید در نظر است شیوع سقط غیر قانونی جنین بین گروهی از زنان حاضر در یک مکان برآورد شود. بدین منظور می‌توان از دو روش کاملاً متفاوت استفاده نمود. در روش اول، از آنان خواسته می‌شود در صورت ارتکاب به عمل سقط غیرقانونی، دست خود را بلند نمایند. طبیعتاً در این روش در حالت بسیار خوش‌بینانه، تعداد کمی از زنان با وجود انجام این عمل، دست خود را بلند خواهند نمود و برآورد حاصل از این روش کم‌تر از مقدار واقعی خواهد بود. در روش دوم می‌توان به هر یک از زنان شرکت کننده یک تاس داد و از آنان خواست که در خلوت خود آن را پرتاب نموده و چنان‌چه روی تاس یکی از اعداد ۱ تا ۴ آمد یا مرتکب عمل سقط غیر قانونی شده‌اند، دست خود را بلند نمایند. در این حالت چون به غیر از شخص پرتاب کننده تاس،

[۱] Response bias and non-response bias

[Downloaded from irje.tums.ac.ir on 2024-07-03]

فرد دیگری نمی‌تواند از وضعیت واقعی وی نسبت به موضوع سقط مطلع گردد، بنابراین انتظار می‌رود در آن جمع تعداد بیش‌تری از افراد با اطمینان خاطر از عدم فاش شدن اطلاعات خصوصی‌شان دست خود را بلند نمایند، که در نهایت این امر منجر به برآورد شیوع سقط جنین، با خطای کم‌تری خواهد شد.

دو تکنیک پاسخ‌دهی تصادفی شده[۱] و پاسخ‌دهی غیر تصادفی شده[۲]، از تکنیک‌های غیر مستقیم عمده‌ای هستند که می‌توان از آن‌ها برای پاسخ‌گویی به سؤال‌های حساس دو حالتی استفاده نمود.

در این مقاله مروری، در خصوص مدل‌های وارنر[۳]، طرح سؤال نامرتبط[۴]، پاسخ‌دهی اجباری[۵]، مورس[۶] و منگات[۷]، از تکنیک پاسخ‌دهی تصادفی شده، و در مورد مدل‌های مثلثی[۸] و متقاطع[۹]، از تکنیک پاسخ‌دهی غیر تصادفی شده، بحث می‌گردد.

تکنیک پاسخ‌دهی تصادفی شده

مدل وارنر

اصول استفاده از مؤلفه شانس به منظور برآورد پاسخ سؤال‌های حساس، اولین بار توسط وارنر[۱۰] در سال ۱۹۶۵ میلادی معرفی شد (۴). در آن زمان، این مدل به عنوان یک شیوه جدید برای حفاظت از حریم خصوصی افراد در مقابل سؤال حساس به شمار می‌رفت.

در این مدل، پاسخ دهندگان در مقابل دو عبارت متضاد قرار گرفته، سپس از آنان درخواست می‌شود در خلوت خود با استفاده از وسیله تصادفی‌سازی که از قبل در اختیار داشتند، یکی از عبارت‌ها را به صورت تصادفی انتخاب و با گزینه‌های بلی یا خیر به آن پاسخ دهند. نکته قابل توجه این که، باید وسیله تصادفی‌سازی مورد استفاده (مانند یک تاس)، دارای توزیع احتمالی مشخص می‌بود.

به عنوان مثال فرض نمایید موضوع حساس مورد نظر، همان برآورد شیوع عمل سقط غیر قانونی جنین بین گروهی از زنان حاضر در یک مکان باشد. در این صورت عبارات اول و دوم به این صورت طرح می‌شود:

عبارت اول: من حداقل یک بار مرتکب سقط غیرقانونی جنین شده‌ام.

عبارت دوم: من هرگز مرتکب عمل سقط غیر قانونی جنین نشده‌ام.

الف: بلی ب: خیر

از پاسخ‌دهندگان درخواست می‌شود تاس را پرتاب کرده، اگر نتیجه آن اعداد ۱ تا ۴ بود، به عبارت اول و در غیر این‌صورت به عبارت دوم پاسخ دهند.

همان طوری که ملاحظه می‌شود، وقتی پاسخ‌دهندگان گزینه «الف» یا «ب» را انتخاب می‌کنند، برای هیچ‌کس مشخص نیست که گزینه انتخاب شده، پاسخ مربوط به عبارت اول است یا دوم، و به این ترتیب، حریم خصوصی افراد حفظ خواهد شد.

در مدل وارنر، هدف، برآورد نسبت پاسخ مثبت به عبارت اول به عنوان عبارت اصلی و حساس است. با استفاده از احتمالات مربوط، برآورد نااریب این نسبت و واریانس آن، به صورت زیر است:

$$\hat{\lambda} = p\pi + (1-p)(1-\pi)$$

$$\hat{\pi} = \frac{\hat{\lambda}+p-1}{2p-1} \quad , \quad p \neq 0.5$$

$$var(\hat{\pi}) = \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(2p-1)^2}$$

$\hat{\lambda}$= نسبت افرادی که گزینه الف را انتخاب نموده‌اند؛

p= احتمال پاسخ به عبارت اول؛

$\hat{\pi}$= نسبت برآورد شده برای پاسخ مثبت به عبارت اول؛ و

n= حجم نمونه.

واریانس‌های بالای برآوردهای حاصل از مدل وارنر، از اصلی‌ترین اشکالاتی بود که بر آن وارد می‌شد (۵). به همین خاطر پژوهشگران بسیاری با ارایه مدل‌های گوناگون در صدد حل مشکل مربوط و کسب نتایج معتبرتر برآمدند که برخی از مهم‌ترین آن‌ها به شرح زیر هستند.

مدل طرح سؤال نامرتبط

طرح سؤال نامرتبط با توزیع احتمالی نامشخص[۱۱]

مهم‌ترین طرحی که به منظور بهبود کارایی مدل وارنر ارایه شد،

---

[۱] Randomized Response Technique, RRT
[۲] Non Randomized Response; NRR
[۳] The Warner Model
[۴] The Unrelated Question Model; UQM
[۵] The Forced Response Model; FRM
[۶] The Moors Model
[۷] The Mangat Model
[۸] The Triangular Model; TM
[۹] The Crosswise Model; CM
[۱۰] Warner

[۱] The Unrelated Question Model with an unknown distribution

[Downloaded from irje.tums.ac.ir on 2024-07-03]

مربوط به طرح هوروتز است که در سال ۱۹۶۷ میلادی با عنوان مدل طرح سؤال نامرتبط معرفی شد (۶). هوروتز[۱] پیشنهاد نمود سؤال یا عبارت دوم مدل وارنر که متضاد با سؤال اول است با یک عبارت غیر حساس و نامرتبط با آن، جایگزین شود.

در این مدل بر خلاف مدل وارنر به جای دو سؤال حساس از یک سؤال حساس استفاده می‌شود. بنابراین انتظار بر این بود که این مدل بتواند با کاهش تورش پاسخ‌دهی، موجبات همکاری بیش‌تر پاسخ‌دهندگان را فراهم نموده و در نهایت سبب افزایش کارایی مدل گردد.

به عنوان مثال

سؤال اول: آیا شما سابقه سقط غیرقانونی جنین داشته‌اید؟

سؤال دوم: آیا دیروز اخبار ساعت ۹ شب را دیده‌اید؟

الف: بلی ب: خیر

در این مدل، توزیع احتمال پاسخ مثبت به سؤال نامرتبط مشخص نبود. به همین سبب، این مدل با عنوان طرح سؤال نامرتبط با توزیع احتمالی نامشخص معروف شد (۷). در این مدل به منظور برآورد نسبت رفتار حساس ($\hat{\pi}_x$) و غیر حساس ($\hat{\pi}_y$) از دو نمونه مستقل استفاده می‌شد (x و y به ترتیب نشان دهنده صفت حساس و صفت غیر حساس می‌باشند). نکته قابل توجه این که، باید احتمال پاسخ به سؤال حساس در دو نمونه متفاوت از هم باشند ($p_1 \neq p_2$). بنابراین به منظور تأمین این شرط، از دو وسیله تصادفی سازی متفاوت مانند سکه و تاس استفاده می‌شود.

در این مدل، برآورد ناریب از نسبت پاسخ مثبت به سؤال حساس و واریانس آن، به صورت زیر محاسبه می‌گردد:

$$\hat{\pi}_x = \frac{\hat{\lambda}_1(1-p_2) - \hat{\lambda}_2(1-p_1)}{p_1 - p_2}$$

$$\mathrm{var}(\hat{\pi}_x) = \left[\frac{1}{(p_1-p_2)^2}\right]\left[\frac{\lambda_1(1-\lambda_1)(1-p_2)^2}{n_1} + \frac{\lambda_2(1-\lambda_2)(1-p_1)^2}{n_2}\right]$$

$\hat{\lambda}_1, \hat{\lambda}_2$= نسبت انتخاب گزینه الف در نمونه اول و دوم؛

p_1, p_2= احتمال پاسخ به سؤال حساس در نمونه اول و دوم؛

$\hat{\pi}_x$= نسبت برآورد شده برای پاسخ مثبت به سؤال حساس؛ و

n_1, n_2 = حجم نمونه اول و دوم.

طرح سؤال نامرتبط با توزیع احتمالی مشخص[۲]

مدل طرح سؤال نامرتبط توسط گرینبرگ در سال ۱۹۶۹ میلادی اصلاح شد (۸). طرح گرینبرگ[۳] برای اصلاح این مدل عبارت بود از استفاده از سؤال نامرتبطی که احتمال پاسخ مثبت به آن مشخص باشد. به عنوان مثال: «آیا شما متولد فصل بهار هستید؟»

لازم به ذکر است با فرض توزیع یکنواخت تولد در طول سال، احتمال تولد در فصل بهار برابر ۰/۲۵ می‌باشد.

روش اصلاحی گرینبرگ علاوه بر این‌که منجر به افزایش دقت برآورد نسبت رفتار حساس شد، تنها با یک نمونه قابل انجام بود و نیاز به دو نمونه را رفع نمود (۷).

در این مدل، برآورد ناریب از نسبت پاسخ مثبت به سؤال حساس و واریانس آن، به صورت زیر محاسبه می‌شود:

$$\hat{\pi}_x = \frac{\hat{\lambda} - \hat{\pi}_y(1-p)}{p} \qquad \mathrm{var}(\hat{\pi}_x) = \frac{\hat{\lambda}(1-\hat{\lambda})}{np^2}$$

$\hat{\lambda}$= نسبت افرادی که گزینه الف را انتخاب نموده‌اند؛

p= احتمال پاسخ به سؤال حساس؛

$\hat{\pi}_x$= نسبت برآورد شده برای پاسخ مثبت به سؤال حساس؛

$\hat{\pi}_y$= نسبت مشخص برای پاسخ مثبت به سؤال غیرحساس؛ و

n = حجم نمونه.

مدل پاسخ‌دهی اجباری

مدل پاسخ‌دهی اجباری توسط بوروچ[۴] در سال ۱۹۷۱ میلادی معرفی شد (۹). در این مدل از پاسخ‌دهندگان درخواست می‌شد، با استفاده از دو وسیله تصادفی‌سازی مانند پرتاب هم‌زمان دو سکه، به سؤال‌های حساس پاسخ دهند. دستورالعمل پاسخ‌دهی در این مدل بدین صورت است که پاسخ‌دهندگان باید بدون توجه به موضوع سؤال، اگر نتایج پرتاب سکه ها شیر باشد، گزینه‌ی «بلی» و اگر خط باشد، گزینه «خیر» را انتخاب نمایند و صرفاً در صورت مخالف بودن نتایج، صادقانه به سؤال پاسخ دهند.

در مدل پاسخ‌دهی اجباری، چون نتایج پرتاب سکه‌ها کاملاً محرمانه بوده و فقط فرد پاسخ دهنده از آن اطلاع داشت، انتخاب گزینه «بلی» مشخص کننده وضعیت واقعی فرد در قبال موضوع حساس نبوده، و بدین ترتیب حریم خصوصی فرد حفظ می‌شد.

در این مدل، برآورد ناریب از نسبت پاسخ مثبت به سؤال حساس و واریانس آن، به صورت زیر محاسبه می‌شود:

---

[۱] Horvitz

[۲] The Unrelated Question Model with a known distribution

[۳] Greenberg

[۴] Boruch

[Downloaded from irje.tums.ac.ir on 2024-07-03]

$$\hat{\pi} = \frac{\hat{\lambda} - \theta}{p_{true}} \qquad var(\hat{\pi}) = \frac{\hat{\lambda}(1-\hat{\lambda})}{np^2}$$

$\hat{\lambda}$= نسبت افرادی که گزینه الف را انتخاب نموده‌اند؛

p_{true} = احتمال پاسخ به سؤال حساس؛

θ= احتمال انتخاب گزینه «بلی» بدون توجه به موضوع سؤال؛

$\hat{\pi}$= نسبت برآورد شده برای پاسخ مثبت به سؤال حساس؛ و

n= حجم نمونه.

مدل مورس

این مدل توسط مورس[1] در سال ۱۹۷۱ میلادی به منظور افزایش کارایی مدل "طرح سؤال نامرتبط با توزیع احتمالی نامشخص" پیشنهاد شد (۱۰). مورس پیشنهاد نمود احتمال پاسخ به سؤال حساس در نمونه دوم برابر صفر در نظر گرفته شود ($P_2=0$). با این تغییر، عملاً پاسخ دهندگان در نمونه دوم، به صورت مستقیم با یک سؤال غیر حساس مواجه می‌شدند و نتیجه به دست آمده نیز برابر با نسبت برآورد شده برای صفت غیر حساس می‌شد.

در مدل مورس به سبب نیاز به حجم نمونه کم در نمونه دوم نسبت به مدل طرح سؤال نامرتبط با توزیع احتمالی نامشخص، احتمال کم‌تری برای عدم رعایت شرایط لازم برای تکنیک پاسخ‌دهی تصادفی شده (RRT) از طرف پاسخ‌دهندگان وجود داشت که این امر باعث افزایش کارایی مدل می‌شد.

در این مدل، برآورد ناریب از نسبت پاسخ مثبت به سؤال حساس و واریانس آن، به صورت زیر محاسبه می‌گردد:

$$\hat{\pi}_x = \frac{\hat{\lambda}_1 - \hat{\lambda}_2(1-p_1)}{p_1}$$

$$var(\hat{\pi}_x) = \left[\frac{(1-p_1)\sqrt{\pi_y(1-\pi_y)} + \sqrt{\hat{\lambda}_1(1-\hat{\lambda}_1)}}{p_1\sqrt{n}}\right]^2$$

$\hat{\lambda}_1$= نسبت پاسخ مثبت مشاهده شده در نمونه اول؛

p_1= احتمال پاسخ به سؤال حساس در نمونه اول؛

$\hat{\pi}_x$= نسبت برآورد شده برای پاسخ مثبت به سؤال حساس؛

$\hat{\lambda}_2 = \pi_y$ = نسبت پاسخ مثبت مشاهده شده در نمونه دوم و نیز نسبت برآورد شده برای صفت غیر حساس؛ و

$\bar{n}$= متوسط حجم نمونه اول و دوم.

[1] Moors

مدل منگات

این مدل در سال ۱۹۹۴ میلادی توسط منگات[2] ارایه شد (۱۱). در مدل پیشنهادی ایشان، ابتدا از پاسخ‌دهندگان درخواست می‌شد در صورت ارتکاب رفتار حساس، صادقانه به سؤال حساس پاسخ مثبت دهند. در مرحله بعد، از آن دسته افرادی که هنوز پاسخ نداده بودند مجدداً درخواست می‌گردید تا این بار طبق مدل وارنر به دو سؤال موجود پاسخ دهند. تمامی این فرایند به صورت کاملاً پنهان و دور از چشم مصاحبه کننده انجام می‌گرفت.

در این مدل، برآورد ناریب از نسبت پاسخ مثبت به سؤال حساس و واریانس آن، به صورت زیر محاسبه می‌گردد:

$$\hat{\lambda} = \pi + (1-p)(1-\pi)$$

$$\hat{\pi} = \frac{\hat{\lambda} - 1 + p}{p}$$

$$var(\hat{\pi}) = \frac{\pi(1-\pi)}{n} + \frac{(1-\pi)(1-p)}{np}$$

$\hat{\lambda}$ = نسبت افرادی که گزینه «بلی» را انتخاب نموده‌اند؛

p= احتمال پاسخ به عبارت اول؛

$\hat{\pi}$= نسبت برآورد شده برای پاسخ مثبت به عبارت اول؛ و

n= حجم نمونه.

محدودیت‌های تکنیک پاسخ‌دهی تصادفی شده

RRT از لحاظ تئوری، اصول ساده‌ای دارد، اما از لحاظ اجرایی با محدودیت‌هایی مواجه است که کاربرد آن را پیچیده می‌سازد. عمده‌ترین محدودیت‌های RRT عبارت‌اند از:

۱. بالا بودن احتمال عدم رعایت دستورالعمل‌های موجود برای انواع مدل‌های آن از جانب پاسخ‌دهندگانی که دچار سوء فهم شده‌اند یا به آن اعتماد ننموده‌اند (۱۲،۱۳)؛

۲. افزایش هزینه‌های مطالعه به سبب نیاز به ارایه آموزش‌های گسترده برای ارتقای مشارکت واقعی پاسخ دهندگان (۱)؛

۳. وارد آوردن بار اضافی به پاسخ‌دهندگان به سبب استفاده اجباری آنان از یک وسیله تصادفی سازی؛ تا جایی که این امر می‌تواند باعث عدم انجام تصادفی‌سازی توسط پاسخ‌دهنگان به ویژه در موارد خود ایفا شود (۱۴)

[2] Mangat

[Downloaded from irje.tums.ac.ir on 2024-07-03]

۴. احتمال رخداد تورش خود حفاظتی خیر[1] توسط هر دو گروه از افراد (افرادی که مرتکب رفتار حساس شده‌اند و افرادی که آن رفتار را انجام نداده‌اند) وجود دارد. به عنوان مثال در مدل پاسخ‌دهی اجباری، وقتی بر اساس دستورالعمل مدل، فردی که مرتکب رفتار حساس نشده مجبور شود صرفاً بر اساس نتایج پرتاب سکه‌ها گزینه «بلی» را انتخاب کند، ممکن است به دلایل گوناگون از انتخاب آن امتناع نموده و گزینه «خیر» را علامت بزند (۱۴).

تکنیک پاسخ‌دهی غیر تصادفی شده

تکنیک پاسخ‌دهی غیر تصادفی شده (NRR) جایگزینی برای غلبه بر مشکلات یاد شده در RRT می‌باشد. مدل‌های مثلثی (TM or Triangular Method) و متقاطع (CM or Crosswise Model) دو مدل از این نوع تکنیک هستند که در سال ۲۰۰۸ میلادی توسط Yu، Tian و Tang معرفی شدند (۱).

عمده‌ترین مزایای استفاده از این مدل‌ها عبارتند از:

۱. سهولت اجرا در مطالعات، نسبت به مدل‌های RRT به دلیل عدم نیاز به وسیله تصادفی‌سازی؛
۲. کاربرد آسان در مصاحبه‌های رودررو و خود ایفا؛ و
۳. کاهش هزینه‌ها، ارتقای مشارکت و اطمینان از تکرارپذیری نتایج نسبت به مدل‌های RRT (۲،۱).

اساس طراحی این مدل‌ها از یک ایده ساده منشأ گرفته است، بدین صورت که پاسخ‌دهندگان در مقابل دو سؤال (یکی حساس و دیگری غیر حساس) قرار می‌گیرند. سپس از آنان درخواست می‌گردد با استفاده از یک دستورالعمل خاص به صورت هم‌زمان به آن دو سؤال پاسخ دهند.

فرض نمایید X متغیر حساس مورد نظر می‌باشد، X=۱ نشان دهنده گروهی از افراد است که دارای آن رفتار حساس هستند (مثلاً سقط غیرقانونی جنین) و X=۰ افرادی هستند که این رفتار را ندارند. علاوه بر آن فرض نمایید Y یک متغیر تصادفی دو حالتی غیر حساس (مثلاً تولد در فصل بهار) با دو ویژگی مستقل از متغیر حساس و توزیع احتمالی مشخص، می‌باشد. Y=۱ یعنی افراد در فصل بهار متولد شده‌اند و Y=۰ یعنی افراد در فصلی به غیر از فصل بهار متولد شده‌اند.

در این دو مدل همیشه فرض بر این است که احتمال پاسخ مثبت به سؤال غیر حساس، P=Pr (Y=1) از قبل مشخص است.

مدل مثلثی (Triangular)

دستورالعمل پاسخ‌دهی در مدل مثلثی به این صورت است که از پاسخ‌دهنده درخواست می‌شود چنان‌چه پاسخ وی به هر دو سؤال «خیر» باشد، گزینه «الف» و چنان‌چه پاسخ حداقل یکی از سؤال‌ها «بلی» باشد، گزینه «ب» را انتخاب نماید (جدول شماره ۱). بدین ترتیب گزینه «الف» به گروهی از افراد تعلق می‌گیرد که نه مرتکب سقط غیرقانونی جنین شده‌اند و نه در فصل بهار به دنیا آمده اند و گزینه «ب» نیز متعلق به افرادی می‌شود که یا مرتکب سقط غیرقانونی جنین شده‌اند، یا در فصل بهار به دنیا آمده‌اند، یا هر دو ویژگی را داشته‌اند. چون گزینه «ب» شامل دو گروه از افراد است، بنابراین نمی‌توان گفت افرادی که آن را انتخاب کرده‌اند، الزاما مرتکب رفتار غیرقانونی شده‌اند و به این ترتیب حریم خصوصی پاسخ‌دهندگان حفظ می‌گردد.

در این مدل، برآورد ناریب از نسبت پاسخ مثبت به سؤال حساس و واریانس آن به صورت زیر محاسبه می‌شود:

$$\hat{\theta} = (1-\pi)(1-p)$$

$$\hat{\pi} = 1 - \frac{\hat{\theta}}{1-p}$$

$$\mathrm{var}(\hat{\pi}) = \frac{\theta(1-\theta)}{n(1-p)^2} = \frac{\pi(1-\pi)}{n} + \frac{p(1-\pi)}{n(1-p)}$$

هم‌چنین برای برآورد ناریب واریانس نسبت رفتار حساس $(\mathrm{var}(\hat{\pi}))$، فرمول زیر مورد استفاده قرار می‌گیرد:

$$\widehat{\mathrm{var}}(\hat{\pi}) = \frac{\hat{\theta}(1-\hat{\theta})}{(n-1)(1-p)^2}$$

$\hat{\theta}$= نسبت افرادی که گزینه «الف» را انتخاب کرده‌اند؛

$\hat{\pi}$= نسبت افرادی که رفتار حساس را داشته اند؛

p= نسبت افرادی که به سؤال غیر حساس پاسخ مثبت داده‌اند؛ و

n= حجم نمونه

در مدل مثلثی یک راه‌کار خود حفاظتی برای پاسخ به سؤال‌ها وجود دارد. انتخاب گزینه «الف» بدون شک به این معنی است که افراد مرتکب رفتار حساس نشده‌اند و این می‌تواند برای افرادی که

[1] Self-protective response bias

[Downloaded from irje.tums.ac.ir on 2024-07-03]

بخواهند در هر شرایطی پاسخ «خیر» بدهند، گزینه مناسبی باشد.

مدل متقاطع (Crosswise)

چگونگی پاسخ‌دهی در مدل متقاطع، به این صورت است که از پاسخ‌دهنده درخواست می‌شود چنان‌چه پاسخ به هر دو سؤال یکسان باشد (هر دو بلی یا هر دو خیر)، گزینه «الف» و چنان‌چه پاسخ به هر دو سؤال متفاوت باشد (یکی بلی و دیگری خیر)، گزینه «ب» را انتخاب نماید (جدول شماره ۲).

چون در این حالت گزینه‌های «الف» و «ب» هر دو گروه از افراد را شامل می‌شوند، دیگر یک راه‌کار خود حفاظتی مشخص وجود نخواهد داشت. بنابراین در این مدل انتظار داریم برآوردهای به دست آمده معتبرتر از برآوردهای مدل مثلثی باشد.

باید توجه داشت که در مدل متقاطع، احتمال پاسخ مثبت به سؤال غیر حساس نباید مساوی با ۰/۵ باشد. اگر در این تکنیک P=۰/۵ در نظر گرفته شود، واریانس به سمت بی‌نهایت و اگر P=۰ یا P=۱ در نظر گرفته شود، واریانس به سمت حداقل میل خواهد نمود. به عبارتی زمانی که P ≤ ۱ و ۰/۵ < P یا ۰/۵ > P ≥ ۰ باشد، برای هر مقدار ثابت π و حجم نمونه ثابت n، Var (π) تابعی افزایشی یا کاهشی از p می‌باشد. به صورتی که هر چقدر مقدار P از ۰/۵ دورتر یا به آن نزدیک‌تر شود؛ متناسب با آن Var (π) نیز کم‌تر یا بیش‌تر خواهد شد. بنابراین این موضوع باید در طراحی سؤال غیر حساس مورد توجه قرار گیرد و متغیری وارد سؤال شود، که احتمال پاسخ مثبت به آن از ۰/۵ دورتر باشد.

به عنوان مثال، با فرض یکنواخت بودن توزیع تولد در طول سال، می‌توان به جای سؤال از فصل تولد اشخاص از ماه تولدشان پرسید، چون در این حالت مقدار p از ۰/۲۵ به حدود ۰/۰۸ کاهش پیدا می‌کند.

در این مدل، برآورد نااریب از نسبت پاسخ مثبت به سؤال حساس و واریانس آن، و هم‌چنین براورد واریانس این نسبت، خطای معیار و حدود اطمینان ۹۵٪ آن، به صورت زیر محاسبه می‌گردد:

$$\hat{\pi} = \frac{\hat{\lambda} + p - 1}{2p - 1} \quad , p \neq 0.5$$

$$\text{var}(\hat{\pi}) = \frac{\lambda(1-\lambda)}{n(2p-1)^2} = \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(2p-1)^2}$$

$$\widehat{\text{var}}(\hat{\pi}) = \frac{\hat{\lambda}(1-\hat{\lambda})}{(n-1)(2p-1)^2} = \frac{\hat{\pi}(1-\hat{\pi})}{n-1} + \frac{p(1-p)}{(n-1)(2p-1)^2}$$

$$SE(\hat{\pi}) = \sqrt{\frac{\hat{\lambda}(1-\hat{\lambda})}{(n-1)(2p-1)^2}}$$

$$95\%\ CI(\pi) = \hat{\pi} \pm z_{\alpha/2}\, SE(\hat{\pi})$$

$\hat{\lambda}$= نسبت افرادی که گزینه «الف» را انتخاب کرده‌اند؛

$\hat{\pi}$= نسبت افرادی که رفتار حساس را داشته‌اند؛

p= نسبت افرادی که به سؤال غیر حساس پاسخ مثبت داده‌اند؛ و

n= حجم نمونه.

به عنوان مثال، فرض نمایید موضوع سقط غیر قانونی جنین در قالب یک پرسش‌نامه مدل متقاطع، از ۱۰۰ نفر از زنان سنین باروری حاضر در یک مکان پرسیده شده است. سؤال‌های طرح شده به این صورت می‌باشند:

آیا پدر شما متولد ماه فروردین است؟

آیا شما تاکنون سابقه سقط غیرقانونی جنین داشته‌اید؟

گزینه الف: هر دو بلی یا هر دو خیر

گزینه ب: یکی بلی و دیگری خیر

پس از جمع‌آوری پرسشنامه‌ها مشخص می‌شود ۷۸ نفر گزینه «الف» و ۲۲ نفر گزینه «ب» را انتخاب نموده‌اند.

با توجه به احتمال پاسخ مثبت به سؤال غیر حساس طرح شده در پرسشنامه (با فرض توزیع یکنواخت احتمال تولد در طول سال)، نتایج تجزیه و تحلیل داده‌ها به قرار زیر می‌باشد:

$$p = \frac{1}{12}$$

$$\hat{\pi} = \frac{\frac{78}{100} + \frac{1}{12} - 1}{2 * \frac{1}{12} - 1} = 0.164 = 16.4\%$$

$$\widehat{\text{var}}(\hat{\pi}) = \frac{\frac{78}{100} * \left(1 - \frac{78}{100}\right)}{(100-1) * \left(2 * \frac{1}{12} - 1\right)^2} = 0.00249$$

$$SE(\hat{\pi}) = \sqrt{\frac{\frac{78}{100} * \left(1 - \frac{78}{100}\right)}{(100-1) * \left(2 * \frac{1}{12} - 1\right)^2}} = 0.0499$$

$$95\%\ CI(\pi) = [\,0.066\ ,0.262\,] = [\ 6.6\%, 26.2\%\ ]$$

حجم نمونه مورد نیاز در مدل متقاطع با استفاده از فرمول

$$n = \left(\frac{z_{\alpha/2}}{L_0(2p-1)}\right)^2$$

[Downloaded from irje.tums.ac.ir on 2024-07-03]

تعیین می‌گردد که در آن L_0 حداکثر دامنه اطمینان حول شیوع صفت حساس می‌باشد که بسته به نظر پژوهشگر می‌تواند تغییر نماید.

برای آنالیز روابط بین داده‌های حاصل از پرسشنامه مستقیم، زمانی که متغیر پاسخ دو حالتی باشد، می‌توان از مدل‌های رگرسیون لجستیک معمولی استفاده نمود. اما برای آنالیز این روابط در داده‌های حاصل از مدل متقاطعT به سبب این‌که متغیر پاسخ از ترکیب یک متغیر حساس با یک متغیر غیر حساس تشکیل شده است، نمی‌توان از آن مدل‌ها استفاده نمود. اما با توجه به این‌که توزیع احتمالی سؤال غیر حساس در مدل متقاطع مشخص بوده و دو متغیر مذکور نیز مستقل هستند، می‌توان از یک مدل رگرسیونی سازگار با این نوع داده‌ها با عنوان Crosswise Logistic Regression (۱۶-۱۴) استفاده نمود. این آنالیز را می‌توان با استفاده از دستورات زیر در نرم‌افزار Stata انجام داد.

"cwlogit y x_1 x_2 ..., pyes (#) or"
"xi:cwlogit y i.x_1 x_2 ..., pyes (#) or"

y= متغیر پاسخ؛
x_1 x_2 ,... = متغیرهای مستقل موجود در مطالعه؛ و
pyes (#) = احتمال پاسخ مثبت به سؤال غیر حساس.

نتیجه‌گیری

ماهیت سؤال‌های حساس و نگرانی‌های اساسی موجود در درستی پاسخ‌هایی که به این نوع سؤال‌ها داده می‌شود، باعث شده همواره پژوهشگران به دنبال یافتن تکنیک‌هایی باشند که آن‌ها را به سمت واقعیت سوق دهد. بنابراین وقتی نخستین تکنیک برخورد با چنین سؤال‌هایی با وجود تمام محدودیت‌ها توسط وارنر مطرح شد، مسیری پیش روی پژوهشگران باز شد تا در طول زمان بتوانند با اصلاح محدودیت‌ها و اضافه نمودن بخش‌هایی به آن باعث افزایش جذابیت و قابلیت انجام، هم بین پاسخ‌دهندگان و هم بین پژوهشگران این حیطه شود و البته این مسیر هم‌چنان تداوم دارد.

جدول شماره ۱- مدل مثلثی و احتمالات مربوط به خانه‌های آن

Cat	Y=0	Y=1	Cat	Y=0	Y=1	Total
X=0	الف	ب	X=0	(1- π)(1-p)	(1- π)p	1- π
X=1	ب	ب	X=1	π (1- p)	π p	π
			Total	(1-p)	p	1

جدول شماره ۲- مدل متقاطع و احتمالات مربوط به خانه‌های آن

Cat	Y=0	Y=1	Cat	Y=0	Y=1	Total
X=0	الف	ب	X=0	(1- π)(1-p)	(1- π)p	1- π
X=1	ب	الف	X=1	π(1- p)	πp	π
			Total	(1-p)	p	1

منابع

1. Yu JW, Tian GL, Tang ML. Two new models for survey sampling with sensitive characteristic: design and analysis. Metrika 2008; 67: 251-63.
2. Tian GL, Yuen KC, Tang ML, Tan MT. Bayesian non-randomized response models for surveys with sensitive questions. Statistics and Its Interface 2009; 2: 13-25.
3. Tourangeau R, Yan T. Sensitive questions in surveys. Psychological bulletin 2007; 133: 859-83.
4. Warner SL. Randomized-response: A survey technique for eliminating evasive answer bias. Journal of the American Statistical Association 1965; 60: 63-9.
5. Houston J, Tran A. A Survey of Tax Evasion Using the Randomized Response. Advances in Taxation 2001;13: 69-94.
6. Horvitz DG, Shah BV, Simmons WR. The unrelated question randomized response model, In: proceedings of the Social Statistics Section. American Statistical Association 1967; 65–72.
7. Lensvelt-Mulders GJLM, Hox JJ, Van Der Heijden PGM. How to Improve the Efficiency of Randomised Response Designs. Quality and Quantity 2005; 39: 253-65.

[Downloaded from irje.tums.ac.ir on 2024-07-03]

8. Greenberg BG, Abul-Ela AA, Simmons WR, Horvitz DG. The unrelated question randomized response model: theoretical framework. Journal of the American Statistical Association 1969; 64: 520–39.
9. Boruch RF. Assuring confidentiality of responses in social research: A note on strategies. The American Sociologist 1971; 6: 308–11.
10. Moors JJA. Optimization of the unrelated question randomised response model. Journal of the American Statistical Association 1971; 66: 627–9.
11. Mangat NS. An improved randomised response strategy. Journal of the Royal Statistic society of Britain 1994; 56: 93–5.
12. Cruyff MJLF, Van Den Hout A, Van Der Heijden PGM, Böckenholt U. Log-linear randomized-response models taking self-protective response behavior into account. Sociological Methods and Research 2007; 36: 266-82.
13. Lensvelt-Mulders GJLM, Boeije HR. Evaluating compliance with a computer assisted randomized response technique: a qualitative study into the origins of lying and cheating. Computers in Human Behavior 2007; 23: 591-608.
14. Jann B, Jerke J, Krumpal I. Asking Sensitive Questions Using the Crosswise Model: An Experimental Survey Measuring Plagiarism. Public Opinion Quarterly 2012; 76: 32-49.
15. Banayejeddi M. Prevalence estimation of “Iron tablets intake” and “Nutrition Education” in the weekly Iron supplementation programs conducted by the Ministry of Health, using crosswise method and its comparison with direct method in Chalderan high school girl students in 90-91 [dissertation]. Tehran (Iran): Tehran University of Medical Sciences; 2013.
16. Shamsipour M, Yunesian M, Fotouhi A, Jann B, Rahimi-Movaghar A, Asghari F, Akhlaghi AA. Estimating the prevalence of illicit drug use among students using the crosswise model. Subst Use Misuse 2014; 49: 1303-10.

[Downloaded from irje.tums.ac.ir on 2024-07-03]

Iranian Journal of Epidemiology 2014; 10(3): 83-91.

Review Article

A Review on Indirect Answering Techniques to Sensitive Binomial Questions

Banayejeddi M[1], Navadeh S[2], Fotouhi A[3], Yazdani K[4]

1- MSc Student in Epidemiology, Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

2- PhD Student in Epidemiology, Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

3- Professor of Epidemiology, Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

4- Assistant Professor of Epidemiology, Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

Corresponding author: Yazdani K., kyazdani@tums.ac.ir

Sensitive questions about people's behaviors are often discussed in studies in the fields of human, social, and medical sciences. Obtaining valid answers to these kinds of questions is one of the main challenges in the mentioned studies. The most important risk factors for the validity of these studies are response bias and non-response bias, which occur because of concerns about divulging personal information. Therefore, researchers are to find solutions other than direct questioning to increase the validity of obtained results through reducing these biases.

In this review article, we discuss the Warner Model, the Unrelated Question Model, the Forced Response Model, the Moors Model, and the Mangat Model that are subsets of the Randomized Response Technique, and also the Triangular Model and Crosswise Model that are subsets of the Non-Randomized Response Technique. The models are all used in answering sensitive binomial questions.

Keywords: Sensitive Question, Randomized Response Technique, Non Randomized Response Technique

[Downloaded from irje.tums.ac.ir on 2024-07-03]