مجله علوم آماری، پاییز و زمستان ۱۳۹۳

جلد ۸، شماره ۲، ص ۲۴۵-۲۶۰

مدل پاسخ تصادفیده کمی اختیاری سه مرحله‌ای

زهرا یزاری، سید محمدرضا علوی

گروه آمار، دانشگاه شهید چمران اهواز

تاریخ دریافت: ۱۳۹۲/۱۰/۲۲ تاریخ آخرین بازنگری: ۱۳۹۳/۸/۳

چکیده : روش پاسخ تصادفیده روشی برای جمع‌آوری اطلاعات درباره ویژگی‌های حساس بدون افشای هویت پاسخ‌دهنده است. مدل‌های پاسخ تصادفیده اختیاری، براساس این فرض که یک سؤال حساس ممکن است برای یک پاسخ‌دهنده حساس باشد و برای دیگری حساس نباشد، بنا شده‌اند. در این مقاله یک مدل پاسخ تصادفیده اختیاری سه مرحله‌ای پیشنهاد و خواص آن در مطالعه‌ای شبیه‌سازی در محیط R بررسی شده است. با استفاده از این مدل، میانگین و سطح درآمد سرپرست خانوارهای دانشجویان دانشگاه شهید چمران اهواز برآورد شده است.

واژه‌های کلیدی : متغیر حساس کمی، روش پاسخ تصادفیده، روش پاسخ تصادفیده اختیاری، سطح حساسیت.

آدرس الکترونیک مسئول مقاله: سید محمدرضا علوی، alavi_m@scu.ac.ir

کد موضوع‌بندی ریاضی (۲۰۱۰): ۶۲D۰۵

۱ مقدمه

یکی از مسائلی که در تحقیق‌های مربوط به موضوع‌های حساس هنگام استفاده از پاسخ مستقیم وجود دارد، اریبی جامعه‌پسند است. روش‌های پاسخ تصادفیده[۱] که برای اولین بار توسط وارنر (۱۹۶۵) معرفی شد، از جمله روش‌هایی است که برای اصلاح این اریبی استفاده می‌شود. در این روش‌ها پاسخ واقعی در یک پاسخ تصادفی ادغام می‌شود و بر اساس این پاسخ‌های ادغام شده که به آن تصادفیده می‌گویند، پارامترهای جامعه برآورد می‌شوند. مدل‌های پاسخ تصادفیده کمی را می‌توان به سه دسته کامل، جزئی و اختیاری تقسیم نمود. در مدل کامل، تمام افراد نمونه و در مدل جزئی بخشی از افراد پاسخ تصادفیده ارائه می‌دهند. وارنر (۱۹۷۱) روش تصادفیده کامل جمعی را معرفی کرد. گوپتا و تورنتون (۲۰۰۲) مدل جزئی جمعی را بررسی نمودند. گوپتا و همکاران (۲۰۰۶) مدل اختیاری یک مرحله‌ای و گوپتا و همکاران (۲۰۱۰) مدل اختیاری دومرحله‌ای را معرفی کردند. مهتا و همکاران (۲۰۱۲) تحت یک محدودیت گونه‌ای از مدل اختیاری سه‌مرحله‌ای جمعی را معرفی کردند. کرومپال (۲۰۱۲) برای برآورد میزان رواج بیگانه هراسی و مخالفت با یهودیان، مقایسه‌ای میان روش پاسخ تصادفیده و روش پاسخ مستقیم انجام داد. همچنین کراس و همکاران (۲۰۱۰) به‌منظور به‌دست آوردن اطلاعات حساس در مورد میزان شیوع بیماری‌های حیوانی از روش پاسخ تصادفیده استفاده کردند.

در این مقاله از طریق افزودن یک مرحله تصادفی دیگر به مدل گوپتا و همکاران (۲۰۱۰) مدلی جدید معرفی می‌شود. انتظار می‌رود این روش، حفاظت از محرمانگی را افزایش و نرخ بی‌پاسخی را کاهش دهد. نشان داده می‌شود که مدل‌های ارائه شده قبلی، حالت‌های خاصی از مدل پیشنهادی هستند. در بخش ۲ برای برآورد میانگین متغیر حساس کمی، روش‌های پاسخ تصادفیده کامل و جزیی جمعی مرور می‌شوند. بخش ۳ به معرفی روش تصادفیده اختیاری دومرحله‌ای جمعی اختصاص دارد. در بخش ۴ روش پیشنهادی معرفی شده، سپس برآورد میانگین و سطح حساسیت متغیر حساس کمی، به‌همراه واریانس آن‌ها ارائه می‌گردند. در بخش ۵ در مطالعه‌ای

---

[۱] Randomized response techniques

شبیه‌سازی، مدل جدید پیشنهادی و مدل گوپتا و همکاران (۲۰۱۰) مقایسه می‌شوند. در پایان در بخش ۶، میانگین درآمد ماهانه سرپرست خانوار دانشجویان دانشگاه شهید چمران اهواز با روش پیشنهادی برآورد می‌شود.

۲ مدل‌های پاسخ تصادفیده کامل و جزئی در حالت جمعی

در مدل کامل جمعی که توسط وارنر (۱۹۷۱) معرفی شد، از تمام پاسخ‌دهندگان انتخاب شده در نمونه، در خواست می‌شود که با استفاده از یک توزیع معلوم مناسب مانند G، یک عدد تصادفی S با میانگین و انحراف معیار به‌ترتیب μ_S و σ_S انتخاب کنند و سپس پاسخ واقعی خود به سؤال حساس را به‌وسیله جمع زدن با آن عدد تصادفی مخفی نمایند. فرض کنید X متغیر حساس، مستقل از S با میانگین و انحراف معیار به‌ترتیب μ_X و σ_X باشد. در آن‌صورت متغیر پاسخ تصادفیده Z براساس مدل کامل جمعی به‌صورت $Z = X + S$ می‌باشد. یک برآورد نااریب برای میانگین متغیر حساس کمی، توسط وارنر (۱۹۷۱) به‌صورت

$$\hat{\mu}_X = \bar{Z} - \mu_S$$

معرفی شد، واریانس این برآوردگر برابر است با:

$$V(\hat{\mu}_X) = \frac{\sigma_Z^2}{n} = \frac{\sigma_X^2}{n} + \frac{\sigma_S^2}{n} \qquad (1)$$

جمله دوم در رابطه (۱) تاوان استفاده از مدل پاسخ تصادفیده است. گوپتا و تورنتون (۲۰۰۲) براساس مدل کامل جمعی، مدل جزئی جمعی را به‌صورت زیر پیشنهاد دادند:

از فرد iام در نمونه تقاضا می‌شود یک آزمایش برنولی با احتمال پیروزی T را انجام دهد، اگر پیروزی رخ دهد پاسخ واقعی خود، یعنی X_i را بیان کند در غیر این‌صورت با استفاده از روش جمعی پاسخ تصادفیده $X_i + S_i$ را بیان کند. فرض کنید U، متغیر تصادفی برنولی با احتمال پیروزی T باشد، در آن‌صورت پاسخ تصادفیده Z_i براساس مدل جزئی جمعی برابر است با:

$$Z_i = U_i X_i + (1 - U_i)(X_i + S_i)$$

آن‌ها یک برآوردگر نااریب را برای μ_X براساس یک نمونه تصادفی به حجم n به‌صورت

$$\hat{\mu}_X = \bar{Z} - (1 - T)\mu_S$$

معرفی کردند، واریانس این برآوردگر عبارت است از

$$V(\hat{\mu}_X) = \frac{\sigma_X^2}{n} + \frac{(1 - T)(\sigma_S^2 + T\mu_S^2)}{n}$$

۳ مدل‌های پاسخ تصادفیده اختیاری

به‌طور کلی مدل‌های اختیاری با این دیدگاه ارائه شدند که ممکن است سؤال مطرح شده از نظر برخی پاسخ‌دهندگان حساس نباشد و آن‌ها نیازی به مخفی کردن پاسخ‌های خود نداشته باشند. بنابراین انتخاب بین اعلام پاسخ تصادفیده یا پاسخ واقعی به اختیار پاسخ‌دهندگان واگذار می‌شود. در این روش با استفاده از دو نمونه مستقل علاوه بر میانگین حساس می‌توان سطح حساسیت را نیز برآورد نمود (گوپتا و همکاران، ۲۰۰۲). گوپتا و همکاران (۲۰۰۶) مدل تصادفیده اختیاری یک مرحله‌ای را معرفی کردند. طبق این مدل، اگر W سطح حساسیت سؤال حساس X باشد، پاسخ تصادفیده اختیاری یک مرحله‌ای جمعی فرد jام $(j = 1, \ldots, n_i)$ در نمونه‌ی iام $(i = 1, 2)$ به‌صورت

$$Z_{ij} = \begin{cases} X_{ij} & \text{با احتمال} (1 - W) \\ X_{ij} + S_{ij} & \text{با احتمال} W \end{cases}$$

است. گوپتا و همکاران (۲۰۱۰) مدل اختیاری دومرحله‌ای جمعی را پیشنهاد کردند، که در آن از دو نمونه مستقل با حجم‌های n_1 و n_2 استفاده می‌شود. از هر فرد در خواست می‌شود یک آزمایش برنولی با احتمال پیروزی T انجام دهد. اگر پیروزی رخ دهد، پاسخ واقعی و اگر شکست رخ دهد، پاسخ تصادفیده‌ی یک مرحله‌ای را بیان نماید. پاسخ تصادفیده‌ی فرد jام $(j = 1, \ldots, n_i)$ در نمونه‌ی iام $(i = 1, 2)$

به‌صورت

$$Z_{ij} = \begin{cases} X_{ij} & \text{با احتمال } T + (1-T)(1-W) \\ X_{ij} + S_{ij} & \text{با احتمال } (1-T)W \end{cases}$$

است، که در آن S_{ij} عدد تصادفی فرد jام است که از توزیع معلوم G_i با میانگین و واریانس معلوم به‌ترتیب μ_{S_i} و $\sigma^2_{S_i}$ تولید شده است. با فرض دو به‌دو مستقل بودن X، S_1 و S_2 آن‌ها برآوردهای μ_X و W را به‌صورت

$$\widehat{\mu}_X = \frac{\bar{Z}_1\mu_{S_2} - \bar{Z}_2\mu_{S_1}}{\mu_{S_2} - \mu_{S_1}}, \quad \mu_{S_1} \neq \mu_{S_2}$$
$$\widehat{W} = \frac{\bar{Z}_2 - \bar{Z}_1}{(\mu_{S_2} - \mu_{S_1})(1-T)}, \quad \mu_{S_1} \neq \mu_{S_2}$$

ارائه دادند. واریانس‌های این دو برآوردگر به‌صورت

$$V(\widehat{\mu}_X) = \frac{1}{(\mu_{S_2} - \mu_{S_1})^2}(\mu^2_{S_2}\frac{\sigma^2_{Z_1}}{n_1} + \mu^2_{S_1}\frac{\sigma^2_{Z_2}}{n_2})$$
$$V(\widehat{W}) = \frac{1}{(\mu_{S_2} - \mu_{S_1})^2(1-T)^2}(\frac{\sigma^2_{Z_1}}{n_1} + \frac{\sigma^2_{Z_2}}{n_2})$$

هستند، که در آن

$$\sigma^2_{Z_i} = \sigma^2_X + W(1-T)(\sigma^2_{S_i} + [1 - W(1-T)]\mu^2_{S_i}), \quad i = 1, 2$$

اخیراً با جایگذاری یک آزمایش سه‌حالتی به‌جای آزمایش دوحالتی در روش گوپتا و همکاران (۲۰۱۰)، یک مدل سه‌مرحله‌ای توسط مهتا و همکاران (۲۰۱۲) معرفی شده است. روش آن‌ها براساس دو نمونه مستقل است که در هرکدام از نسبت معلومی مانند T می‌خواهند پاسخ واقعی، از نسبت معلوم دیگری مانند p تقاضا می‌کنند پاسخ تصادفیده جمعی و از بقیه درخواست می‌کنند که پاسخ تصادفیده اختیاری را ارائه نمایند. پاسخ تصادفیده فرد jام $(j = 1, \ldots, n_i)$ از نمونه‌ی iام $(i = 1, 2)$ به‌صورت

$$Z_{ij} = \begin{cases} X_{ij} & \text{با احتمال } T + (1-T-p)(1-W) \\ X_{ij} + S_{ij} & \text{با احتمال } p + (1-T-p)W \end{cases}$$

است، که در آن S_{ij}ها همان اعداد با شرایط مطرح شده در روش گوپتا و همکاران (۲۰۱۰) هستند. در این روش لازم است که شرط $T + p < ۱$ برقرار باشد. این محدودیت قابلیت انعطاف این مدل را کاهش می‌دهد. لذا در بخش بعد یک مدل سه‌مرحله‌ای که دارای چنین محدودیتی نیست پیشنهاد شده است.

۴ مدل اختیاری کمی سه‌مرحله‌ای

با اضافه کردن یک مرحله تصادفی دیگر به مدل اختیاری دومرحله‌ای گوپتا و همکاران (۲۰۱۰) گونه دیگری از مدل سه‌مرحله‌ای معرفی می‌شود که مدل اختیاری کمی سه‌مرحله‌ای نامیده می‌شود و مزیت آن رفع محدودیت مدل سه‌مرحله‌ای مهتا و همکاران (۲۰۱۲) است. در این روش به‌دلیل اینکه سه‌بار از ابزار تصادفی کردن استفاده می‌شود، انتظار می‌رود محرمانگی از درجه بالاتری نسبت به مدل دومرحله‌ای برخوردار باشد. این روش براساس دو نمونه مستقل انجام می‌شود که در هر کدام از نسبت معلومی مانند T خواسته می‌شود پاسخ واقعی را ارائه نمایند و از بقیه درخواست می‌شود که اگر سؤال برای آن‌ها حساس نیست پاسخ واقعی را بیان کنند و در غیر این‌صورت یک آزمایش برنولی با احتمال پیروزی p انجام دهند اگر پیروزی رخ دهد پاسخ واقعی و گرنه پاسخ تصادفیده جمعی را گزارش کنند. بنابراین در این روش Z_{ij}، پاسخ تصادفیده فرد jام $(j = ۱, ..., n_i)$ در نمونه iام $(i = ۱, ۲)$، به‌صورت

$$Z_{ij} = \begin{cases} X_{ij} & \text{با احتمال } T + (۱ - W)(۱ - T) + Wp(۱ - T) \\ X_{ij} + S_{ij} & \text{با احتمال } W(۱ - p)(۱ - T) \end{cases}$$

خواهد بود، که در آن W سطح حساسیت است. در این روش نیز فرض می‌شود X، $S_۱$ و $S_۲$ دو به‌دو مستقل و میانگین $S_۱$ و $S_۲$ متفاوت هستند. میانگین پاسخ تصادفیده برای نمونه iام به‌صورت

$$\mu_{Z_i} = \mu_X + W(۱ - T)(۱ - p)\mu_{S_i} \qquad (۲)$$

به‌دست می‌آید. با جایگذاری کردن برآورد نااریب μ_{Z_i} یعنی $\bar{Z}_i$ در رابطه (۲) و حل آن‌ها بر حسب $\widehat{\mu}_X$ و $\widehat{W}$ داریم:

$$\widehat{\mu}_X = \frac{\bar{Z}_1\mu_{S_2} - \bar{Z}_2\mu_{S_1}}{\mu_{S_2} - \mu_{S_1}}, \quad \mu_{S_1} \neq \mu_{S_2} \tag{۳}$$

$$\widehat{W} = \frac{\bar{Z}_2 - \bar{Z}_1}{(1-T)(1-p)(\mu_{S_2} - \mu_{S_1})}, \quad \mu_{S_1} \neq \mu_{S_2} \tag{۴}$$

با توجه به اینکه دو نمونه مستقل هستند، میانگین‌های پاسخ تصادفیده آن‌ها نیز مستقلند. پس $Cov(\bar{Z}_1, \bar{Z}_2) = 0$. بنا به قضیه حد مرکزی (رایس، ۲۰۰۷)، توزیع $\widehat{\mu}_X$ به‌طور مجانبی نرمال با میانگین μ_X و واریانس

$$V(\widehat{\mu}_X) = \frac{1}{(\mu_{S_2} - \mu_{S_1})^2}(\mu_{S_2}^2 \frac{\sigma_{Z_1}^2}{n_1} + \mu_{S_1}^2 \frac{\sigma_{Z_2}^2}{n_2})$$

است، که در آن

$$\sigma_{Z_i}^2 = \sigma_X^2 + \alpha(\sigma_{S_i}^2 + [1-\alpha]\mu_{S_i}^2), \quad i = 1, 2 \tag{۵}$$

و $\alpha = W(1-T)(1-p)$ است. $\widehat{W}$ نیز به‌طور مجانبی نرمال با میانگین W و واریانس

$$V(\widehat{W}) = \frac{1}{(1-T)^2(1-p)^2(\mu_{S_2} - \mu_{S_1})^2}(\frac{\sigma_{Z_1}^2}{n_1} + \frac{\sigma_{Z_2}^2}{n_2})$$

است، که در آن $\sigma_{Z_i}^2$ در رابطه (۵) تعریف شد. برآورگرهای نااریب واریانس‌های $\widehat{\mu}_X$ و $\widehat{W}$ را می‌توان با استفاده از برآوردگر نااریب $\sigma_{Z_i}^2$ یعنی واریانس نمونه $S_{Z_i}^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}(Z_{ij} - \bar{Z}_i)^2$ برای $i = 1, 2$ به‌دست آورد.

مدل اختیاری یک‌مرحله‌ای گوپتا و همکاران (۲۰۰۶) و مدل اختیاری دومرحله‌ای گوپتا و همکاران (۲۰۱۰) حالت‌های خاصی از مدل پیشنهادی هستند که به‌ترتیب با درنظرگرفتن $T = p = 0$ و $p = 0$ حاصل می‌شوند. برای مقایسه کارایی مدل پیشنهادی و دومرحله‌ای گوپتا و همکاران (۲۰۱۰) مقادیر p، T و W در دو مدل یکسان در نظرگرفته شده‌اند. فرض کنید $\widehat{W}_N$ و $\widehat{W}_G$ به‌ترتیب بیانگر برآورد سطح حساسیت در مدل پیشنهادی و مدل دومرحله‌ای باشند و λ_1 به‌صورت

$$\lambda_1 = V(\widehat{W}_N) - V(\widehat{W}_G) = \frac{1}{(1-T)^2(1-p)^2(\mu_{S_2} - \mu_{S_1})^2}(\frac{1}{n_1}A_1 + \frac{1}{n_2}A_2)$$

تعریف شود، که در آن

$$A_i = p[(۲-p)\sigma_X^۲ + W(۱-p)(۱-T)(\sigma_{S_i}^۲ + \mu_{S_I}^۲)], \quad i = ۱, ۲$$

با توجه به‌اینکه $\lambda_۱ \geq ۰$، واریانس $\widehat{W}$ به‌دست آمده از مدل پیشنهادی همواره بزرگتر از واریانس $\widehat{W}$ از مدل دومرحله‌ای است. اختلاف واریانس برآوردگر میانگین متغیر حساس مدل دومرحله‌ای و مدل پیشنهادی به‌صورت

$$\lambda_۲ = V(\widehat{\mu}_G) - V(\widehat{\mu}_N) = \frac{۱}{(\mu_{S_۲} - \mu_{S_۱})^۲}(B_۱\frac{\mu_{S_۲}^۲}{n_۱} + B_۲\frac{\mu_{S_۱}^۲}{n_۲}) \qquad (۶)$$

است، که در آن

$$B_i = \alpha\frac{p}{۱-p}\sigma_{S_i}^۲ + \frac{\alpha p}{(۱-p)^۲}[(۱-۲\alpha) - (۱-\alpha)p]\mu_{S_i}^۲, \quad i = ۱, ۲$$

و α در رابطه (۵) تعریف شد. اگر

$$W > \max\{\frac{\sigma_{S_۱}^۲ + \mu_{S_۱}^۲}{(۲-p)(۱-T)\mu_{S_۱}^۲}, \frac{\sigma_{S_۱}^۲ + \mu_{S_۱}^۲}{(۲-p)(۱-T)\mu_{S_۱}^۲}\} \qquad (۷)$$

آن‌گاه واریانس برآوردگر میانگین متغیر حساس جامعه به‌دست آمده از مدل پیشنهادی، کوچکتر از واریانس نظیرخویش در مدل دومرحله‌ای گوپتا و همکاران (۲۰۱۰) خواهد بود. در بخش بعد نتایج مطالعه شبیه‌سازی برای مقدار $\lambda_۲$ در سطوح مختلف p، T و W ارائه شده است.

۵ مطالعه شبیه‌سازی

برای انتخاب مقادیر مناسب p و T به‌منظور حصول صحت و دقت مدل پیشنهادی، با استفاده از نرم‌افزار R، شبیه‌سازی با ۱۰۰۰۰ بار تکرار که در هر بار برای دو نمونه با حجم برابر n انجام شده است. برای هماهنگی با شبیه‌سازی انجام شده توسط گوپتا و همکاران (۲۰۱۰)، مقدار n، ۵۰۰ و ۱۰۰۰ و توزیع هر سه متغیر X، $S_۱$ و $S_۲$، پواسون با میانگین‌های به‌ترتیب ۴، ۲ و ۵ در نظرگرفته شد. در هر تکرار رابطه‌های (۳) و (۴) محاسبه گردید سپس میانگین و واریانس آن‌ها در ۱۰۰۰۰ بار

تکرار به‌عنوان امید ریاضی و واریانس برآوردها درنظرگرفته شد. باتوجه به جدول ۱، $E(\widehat{W})$ برای $p = ۰/۵$، به مقادیر واقعی W بسیار نزدیک هستند، اما اختلاف $E(\widehat{W})$ ها از مقدار واقعی با دور شدن از $p = ۰/۵$ بیشتر می‌شود. $E(\widehat{\mu}_X)$ در تمام حالت‌ها به مقدار واقعی نزدیک است. تغییر T تأثیر چندانی روی $E(\widehat{\mu}_X)$ و $E(\widehat{W})$ ندارد. در جدول ۲ واریانس شبیه‌سازی‌شده $\widehat{\mu}_X$ و $\widehat{W}$ آمده است. با افزایش p واریانس‌ها افزایش می‌یابند و با افزایش T عموماً کاهش واریانس $\widehat{\mu}_X$ و افزایش واریانس $\widehat{W}$ ملاحظه می‌شود. از طرفی در عمل، حفظ محرمانگی بیشتر در مقادیر نزدیک به ۰/۵ برای p (چون در آزمایش برنولی حاصل بیشترین محرمانگی بین پیروزی و شکست نهفته است) و نزدیک به صفر برای T (چون افراد مایل نیستند سؤال حساس را مستقیماً پاسخ دهند لذا مشارکت آن‌ها در بررسی با یک پرسشنامه با T کوچک، افزایش می‌یابد) قابل حصول می‌باشد. بنابراین توصیه می‌شود در عمل از چنین مقادیری استفاده شود. لذا در بخش کاربرد برای طراحی پرسشنامه از مقادیر $T = ۰/۲$ و $p = ۰/۵$ استفاده شده است.

در ادامه به‌دلیل وجود مقدار مجهول W، در شرط کارایی بیان شده در رابطه (۷)، با استفاده از شبیه‌سازی واریانس برآوردگرهای میانگین به‌دست آمده از مدل سه‌مرحله‌ای پیشنهادی و مدل دومرحله‌ای گوپتا و همکاران (۲۰۱۰) مقایسه می‌شوند. شبیه‌سازی همانند قبل با درنظرگرفتن توزیع پواسون برای متغیرهای $S_۱$ و $S_۲$ با میانگین‌های به‌ترتیب ۲ و ۵ و حجم‌های نمونه $n_۱ = n_۲ = ۵۰$ انجام شده است. در جدول ۳ مقدار $\lambda_۲$ بیان شده در رابطه (۶) به‌ازای مقادیر کم، متوسط و زیاد p و W در سطوح مختلف T آورده شده است. همان‌طور که مشاهده می‌شود در تمام این حالت‌ها کارایی مدل پیشنهادی از مدل گوپتا و همکارن (۲۰۱۰) بهتر است.

۶ کاربرد

به‌منظور برآورد میانگین درآمد ماهانه سرپرست خانوارهای دانشجویان دانشگاه شهید چمران اهواز، پرسشنامه‌ای براساس روش پیشنهادی تنظیم گردید و در میان

جدول ۱: امید ریاضی شبیه‌سازی‌شده برآورد میانگین و سطح حساسیت

T								
۰/۷		۰/۵		۰/۳				
$E(\widehat{\mu_X})$	$E(\widehat{W})$	$E(\widehat{\mu_X})$	$E(\widehat{W})$	$E(\widehat{\mu_X})$	$E(\widehat{W})$	W	p	n
۴/۰۰۱	۰/۰۴۱	۴/۰۰۰	۰/۰۴۳	۳/۹۹۹	۰/۰۴۳	۰/۱		
۴/۰۰۱	۰/۱۲۸	۴/۰۰۰	۰/۱۲۹	۴/۰۰۱	۰/۱۲۸	۰/۳		
۳/۹۹۸	۰/۲۱۶	۳/۹۹۸	۰/۲۱۶	۳/۹۹۸	۰/۲۱۵	۰/۵	۰/۳	
۳/۹۹۶	۰/۳۰۷	۳/۹۹۸	۰/۳۰۱	۴/۰۰۰	۰/۲۹۹	۰/۷		
۴/۰۰۰	۰/۳۸۶	۳/۹۹۸	۰/۳۸۷	۴/۰۰۱	۰/۳۸۶	۰/۹		
۴/۰۰۱	۰/۰۹۸	۴/۰۰۱	۰/۰۰۱	۳/۹۹۹	۰/۰۰۱	۰/۱		
۴/۰۰۱	۰/۲۹۹	۴/۰۰۱	۰/۲۹۸	۴/۰۰۰	۰/۳۰۰	۰/۳		
۳/۹۹۸	۰/۵۰۳	۴/۰۰۰	۰/۵۰۱	۳/۹۹۹	۰/۴۹۹	۰/۵	۰/۵	۵۰۰
۳/۹۹۷	۰/۷۰۷	۴/۰۰۱	۰/۷۰۱	۴/۰۰۱	۰/۶۹۸	۰/۷		
۳/۹۹۹	۰/۹۰۲	۳/۹۹۹	۰/۹۰۱	۴/۰۰۱	۰/۸۹۹	۰/۹		
۴/۰۰۲	۰/۲۲۵	۳/۹۹۶	۰/۲۳۸	۳/۹۹۸	۰/۲۳۸	۰/۱		
۳/۹۹۸	۰/۷۰۴	۴/۰۰۶	۰/۶۸۸	۳/۹۹۸	۰/۷۰۲	۰/۳		
۳/۹۹۸	۱/۱۷۳	۳/۹۹۷	۱/۱۶۹	۴/۰۰۳	۱/۱۶۵	۰/۵	۰/۷	
۴/۰۰۳	۱/۶۲۸	۴/۰۰۲	۱/۶۲۸	۳/۹۹۶	۱/۶۴۰	۰/۷		
۴/۰۰۰	۲/۱۰۰	۳/۹۹۷	۲/۱۰۶	۳/۹۹۸	۲/۱۰۰	۰/۹		
۳/۹۹۹	۰/۰۴۴	۳/۹۹۷	۰/۰۴۵	۴/۰۰۱	۰/۰۴۲	۰/۱		
۳/۹۹۹	۰/۱۳۱	۴/۰۰۰	۰/۱۲۸	۳/۹۹۹	۰/۱۳۰	۰/۳		
۴/۰۰۰	۰/۲۱۳	۳/۹۹۷	۰/۲۱۷	۳/۹۹۹	۰/۲۱۴	۰/۵	۰/۳	
۴/۰۰۰	۰/۳۰۰	۴/۰۰۰	۰/۳۰۰	۳/۹۹۹	۰/۳۰۰	۰/۷		
۳/۹۹۸	۰/۳۸۸	۳/۹۹۹	۰/۳۸۷	۴/۰۰۳	۰/۳۸۴	۰/۹		
۴/۰۰۲	۰/۰۹۸	۳/۹۹۷	۰/۱۰۴	۳/۹۹۹	۰/۱۰۱	۰/۱		
۴/۰۰۰	۰/۳۰۲	۴/۰۰۰	۰/۳۰۱	۴/۰۰۰	۰/۲۹۹	۰/۳		
۴/۰۰۱	۰/۵۰۰	۴/۰۰۰	۰/۵۰۲	۴/۰۰۲	۰/۴۹۸	۰/۵	۰/۵	۱۰۰۰
۴/۰۰۱	۰/۶۹۶	۴/۰۰۱	۰/۶۹۶	۴/۰۰۱	۰/۶۹۸	۰/۷		
۳/۹۹۸	۰/۹۰۶	۳/۹۹۹	۰/۹۰۲	۳/۹۹۸	۰/۹۰۲	۰/۹		
۴/۰۰۱	۰/۲۳۳	۳/۹۹۷	۰/۲۳۷	۳/۹۹۹	۰/۲۳۶۵	۰/۱		
۴/۰۰۱	۰/۷۰۰	۳/۹۹۹	۰/۷۰۲	۴/۰۰۰	۰/۶۹۹	۰/۳		
۴/۰۰۱	۱/۱۶۴	۴/۰۰۱	۱/۱۶۶	۳/۹۹۹	۱/۱۶۹	۰/۵	۰/۷	
۴/۰۰۱	۱/۶۲۹	۴/۰۰۱	۱/۶۳۱	۳/۹۹۸	۱/۶۳۴	۰/۷		
۳/۹۹۹	۲/۱۰۴	۴/۰۰۰	۲/۱۰۰	۳/۹۹۹	۲/۱۰۰	۰/۹		

جدول ۲: واریانس‌های شبیه‌سازی‌شده برآورد میانگین و سطح حساسیت

T								
۰/۷		۰/۵		۰/۳				
$V(\widehat{\mu_X})$	$V(\widehat{W})$	$V(\widehat{\mu_X})$	$V(\widehat{W})$	$V(\widehat{\mu_X})$	$V(\widehat{W})$	W	p	n
۰/۰۵۲	۰/۰۸۲	۰/۰۵۳	۰/۰۳۱	۰/۰۵۵	۰/۰۱۶	۰/۱		
۰/۰۶۵	۰/۰۹۲	۰/۰۵۸	۰/۰۳۵	۰/۰۶۰	۰/۰۱۹	۰/۳		
۰/۰۵۷	۰/۰۹۶	۰/۰۶۰	۰/۰۳۸	۰/۰۶۳	۰/۰۲۱	۰/۵	۰/۳	
۰/۰۶۰	۰/۱۰۳	۰/۰۶۳	۰/۰۴۲	۰/۰۶۸	۰/۰۲۴	۰/۷		
۰/۰۶۰	۰/۱۰۶	۰/۰۶۶	۰/۰۴۴	۰/۰۷۳	۰/۰۲۶	۰/۹		
۰/۰۵۳	۰/۱۶۵	۰/۰۵۴	۰/۰۶۳	۰/۰۵۶	۰/۰۳۴	۰/۱		
۰/۰۵۸	۰/۱۹۴	۰/۰۶۱	۰/۰۷۵	۰/۰۶۴	۰/۰۴۲	۰/۳		
۰/۰۶۰	۰/۲۰۸	۰/۰۹۵	۰/۰۸۵	۰/۰۷۰	۰/۰۴۸	۰/۵	۰/۵	۵۰۰
۰/۰۶۴	۰/۲۲۹	۰/۰۶۹	۰/۰۹۴	۰/۰۷۶	۰/۰۵۵	۰/۷		
۰/۰۶۶	۰/۲۴۲	۰/۰۷۴	۰/۱۰۴	۰/۰۸۱	۰/۰۶۰	۰/۹		
۰/۰۵۴	۰/۴۷۳	۰/۰۵۵	۰/۱۸۰	۰/۰۵۷	۰/۰۹۸	۰/۱		
۰/۰۵۹	۰/۵۵۴	۰/۰۶۳	۰/۲۲۶	۰/۰۶۷	۰/۱۲۷	۰/۳		
۰/۰۶۲	۰/۶۳۰	۰/۰۷۰	۰/۲۶۴	۰/۰۷۸	۰/۱۵۵	۰/۵	۰/۷	
۰/۰۶۷	۰/۶۹۴	۰/۰۷۷	۰/۲۹۵	۰/۰۸۳	۰/۱۷۱	۰/۷		
۰/۰۷۱	۰/۷۵۴	۰/۰۸۲	۰/۳۲۹	۰/۰۸۶	۰/۱۸۰	۰/۹		
۰/۰۲۶	۰/۰۴۲	۰/۰۲۶	۰/۰۱۵	۰/۰۲۷	۰/۰۰۸	۰/۱		
۰/۰۲۷	۰/۰۴۵	۰/۰۲۹	۰/۰۱۸	۰/۰۲۹	۰/۰۰۹	۰/۳		
۰/۰۲۸	۰/۰۴۸	۰/۰۳۰	۰/۰۱۹	۰/۰۳۱	۰/۰۱۰	۰/۵	۰/۳	
۰/۰۳۰	۰/۰۵۱	۰/۰۳۲	۰/۰۲۱	۰/۰۳۴	۰/۰۱۲	۰/۷		
۰/۰۳۱۰	۰/۰۵۳	۰/۰۳۳	۰/۰۲۳	۰/۰۳۶	۰/۰۱۳	۰/۹		
۰/۰۲۶	۰/۰۸۲	۰/۰۲۸	۰/۰۳۲	۰/۰۲۸	۰/۰۱۷	۰/۱		
۰/۰۲۸	۰/۰۹۳	۰/۰۲۹	۰/۰۳۷	۰/۰۳۱	۰/۰۲۱	۰/۳		
۰/۰۲۸	۰/۰۹۳	۰/۰۲۹	۰/۰۳۷	۰/۰۳۱	۰/۰۲۵	۰/۵	۰/۵	۱۰۰۰
۰/۰۳۱	۰/۱۱۲	۰/۰۳۴	۰/۰۴۶	۰/۰۳۸	۰/۰۲۷	۰/۷		
۰/۰۳۲	۰/۱۲۱	۰/۰۳۸	۰/۰۵۲	۰/۰۳۹	۰/۰۲۹	۰/۹		
۰/۰۲۸	۰/۰۹۲	۰/۰۲۹	۰/۰۵۰	۰/۰۳۰	۰/۰۳۱	۰/۱		
۰/۰۳۲	۰/۱۱۴	۰/۰۳۳	۰/۰۶۳	۰/۰۳۴	۰/۰۴۰	۰/۳		
۰/۰۳۷	۰/۱۳۱	۰/۰۳۸	۰/۰۷۴	۰/۰۴۱	۱/۰۵۰	۰/۵	۰/۷	
۰/۰۳۷	۰/۱۴۷	۰/۰۴۲	۰/۰۸۷	۰/۰۴۳	۰/۰۵۶	۰/۷		
۰/۰۳۹	۰/۱۵۹	۰/۰۴۴	۰/۰۹۱	۰/۰۴۵	۰/۰۵۸	۰/۹		

جدول ۳: مقدار $\lambda_۲$ برای سطوح مختلف p، T و W

		p		
		۰/۲	۰/۵	۰/۸
W	T	$\lambda_۲$	$\lambda_۲$	$\lambda_۲$
	۰/۱	۰/۰۱۶	۰/۰۴۳	۰/۰۷۳
	۰/۲	۰/۰۱۵	۰/۰۳۹	۰/۰۶۶
	۰/۳	۰/۰۱۴	۰/۰۳۵	۰/۰۵۹
	۰/۴	۰/۰۱۲	۰/۰۳۱	۰/۰۵۱
۰/۲	۰/۵	۰/۰۱۰	۰/۰۳۱	۰/۰۵۱
	۰/۶	۰/۰۰۹	۰/۰۲۲	۰/۰۳۶
	۰/۷	۰/۰۰۷	۰/۰۱۷	۰/۰۲۷
	۰/۸	۰/۰۰۵	۰/۰۱۱	۰/۰۱۹
	۰/۹	۰/۰۰۲	۰/۰۰۶	۰/۰۰۹
	۰/۱	۰/۰۲۲	۰/۰۶۸	۰/۱۳۰
	۰/۲	۰/۰۲۲	۰/۰۶۷	۰/۱۲۴
	۰/۳	۰/۰۲۲	۰/۰۶۴	۰/۱۱۶
	۰/۴	۰/۰۲۲	۰/۰۵۶۰	۰/۱۰۶
۰/۵	۰/۵	۰/۰۲۰	۰/۰۵۴	۰/۰۹۳
	۰/۶	۰/۰۱۸	۰/۰۴۷	۰/۰۷۹
	۰/۷	۰/۰۱۴	۰/۰۳۸	۰/۰۶۲
	۰/۸	۰/۰۱۰	۰/۰۲۷	۰/۰۴۴
	۰/۹	۰/۰۰۶	۰/۰۱۴	۰/۰۲۳
	۰/۱	۰/۰۰۳	۰/۰۴۳	۰/۱۲۴
	۰/۲	۰/۰۱۱	۰/۰۵۵	۰/۱۳۲
	۰/۳	۰/۰۱۷	۰/۰۶۳	۰/۱۳۵
	۰/۴	۰/۰۲۱	۰/۰۶۷	۰/۱۳۲
۰/۸	۰/۵	۰/۰۲۲	۰/۰۶۷	۰/۱۲۴
	۰/۶	۰/۰۲۲	۰/۰۶۲	۰/۱۱۰
	۰/۷	۰/۰۲۰	۰/۰۵۳	۰/۰۹۱
	۰/۸	۰/۰۱۵	۰/۰۳۹	۰/۰۶۶
	۰/۹	۰/۰۰۹	۰/۰۲۲	۰/۰۳۶

دو نمونه تصادفی ساده با جایگذاری مستقل به حجم‌های ۴۹ و ۵۰ دانشجو توزیع گردید. برای نمونه اول یک دسته کارت تهیه گردید که بر روی ۲۰ درصد آن‌ها جمله «لطفاً درآمد ماهانه واقعی سرپرست را در پرسشنامه درج کنید» نوشته شده بود و بقیه کارت‌ها دارای اعدادی با توزیع فراوانی معلوم با میانگین $\mu_{S_1} = ۶۲۵۰۰۰$ بودند. برای نمونه دوم نیز یک دسته کارت دیگر تهیه گردید که بر روی ۲۰ درصد آن‌ها جمله «لطفاً درآمد ماهانه واقعی سرپرست را در پرسشنامه درج کنید» نوشته شده بود و بقیه کارت‌ها دارای اعدادی با توزیع فراوانی معلوم با میانگین $\mu_{S_2} = ۵۶۲۵۰۰$ بودند. توزیع فراوانی اعداد دو دسته کارت متفاوت درنظر گرفته شد. برای انجام آزمایش تصادفی دوم در هردو نمونه از زوج یا فرد بودن یکان شناسنامه دانشجو استفاده گردید. از هر دانشجو تقاضا شده بود که به‌تصادف از دسته کارت یک کارت انتخاب کند اگر کارت با جمله «لطفاً درآمد ماهانه واقعی سرپرست را در پرسشنامه درج کنید» انتخاب می‌شد، مقدار درآمد در پاسخ‌نامه توسط دانشجو یادداشت می‌شد، در غیر این‌صورت اگر پرسیدن درآمد سرپرست خانواده برای او حساس نبود مقدار درآمد را در پاسخ‌نامه ثبت می‌کرد اما اگر پرسیدن درآمد برای او حساس می‌بود از او تقاضا می‌شد اگر یکان شناسنامه‌اش زوج است درآمد و در غیر این‌صورت مجموع درآمد و عدد کارت را در پاسخ‌نامه درج کند. با توجه به دستور العمل پرسشنامه، برای برآورد میانگین درآمد و سطح حساسیت در روابط (۳) و (۴) مقادیر $p = ۰/۵$ و $T = ۰/۲$ باید جایگذاری شوند. میانگین‌های پاسخ برای نمونه اول و دوم به‌ترتیب $\bar{Z}_1 = ۱۷۶۶۰۰۰$ و $\bar{Z}_2 = ۱۷۹۴۳۶۰$ به‌دست آمدند. برآورد میانگین و سطح حساسیت به ترتیب ۲۰۴۹۶۰۰ تومان و ۱/۱۳− به‌دست آمد. خطای معیار برای میانگین ۲۵۴۳۸۳۳ برآورد گردید. متأسفانه به‌دلیل کوچک بودن حجم نمونه‌ها ملاحظه می‌شود که برآورد W خارج از فضای پارامتر است و برآورد قابل قبولی نیست. البته همان‌طور که در نتایج شبیه‌سازی آشکار شد و داده‌های واقعی نیز مؤید همین مطلب هستند، چنین امری در استفاده از روش‌های پاسخ تصادفیده دور از انتظار نیست (گوپتا و همکاران، ۲۰۱۰). هر چه جنبه تصادفی در مدل افزایش یابد به‌دلیل افزایش واریانس ممکن است برآوردها در فضای پارامتر قرار نگیرند ولی انتظار می‌رود با

افزایش حجم نمونه‌ها این مشکل برطرف شود (گوپتا و شبیر، ۲۰۰۴). البته به‌دلیل منفی بودن سطح حساسیت ممکن است کسی آن را صفر درنظر بگیرد و توجیه کند که سؤال درباره درآمد حساس نیست.

بحث و نتیجه‌گیری

در این مقاله، مدل تصادفیده کمی سه‌مرحله‌ای برای برآورد میانگین و سطح حساسیت متغیرهای کمی حساس معرفی و با استفاده از شبیه‌سازی اریبی و کارایی آن بررسی شد. نشان داده شد که کارایی مدل پیشنهادی از مدل گوپتا و همکاران (۲۰۱۰) بهتر است. با استفاده از این روش براساس دو نمونه تصادفی به حجم‌های به‌ترتیب ۴۹ و ۵۰ میانگین درآمد ماهانه سرپرست خانوار دانشجویان دانشگاه شهید چمران ۲۰۴۹۶۰۰ تومان با خطای معیار ۲۵۴۳۸۳۳ برآورد گردید. اما برآورد سطح حساسیت در دامنه صفر تا یک قرار نگرفت. البته گاهی به‌دلیل کم بودن حجم نمونه ممکن است در استفاده از روش‌های پاسخ تصادفیده با برآوردهایی که خارج از فضای پارامتر قرار می‌گیرند، مواجه شویم. در پژوهش‌های آینده می‌توان از روش‌های تصادفیده اجباری برای تلفیق با این مدل‌ها استفاده کرد و نیز می‌توان این روش‌ها را برای متغیرهای کیفی و یا کمی با مقیاس ترتیبی به‌کار برد.

تقدیر و تشکر

نویسنگان از راهنمایی‌های سازنده داوران محترم کمال تشکر و قدردانی را می‌نمایند. آن‌ها از دانشگاه شهید چمران اهواز نیز برای حمایت‌های مالی کمال تشکر را دارند.

مراجع

Cross, P., Edwards-Jones, G., Omed, P. and Williams, A. P. (2010), Use of a Randomized Response Technique to Obtain Sensitive Information on Animal Disease Prevalence, *Preventive Veterinary Medicine*, **96**, 252-562.

Gupta, S. N. and Thornton, B. (2002), Circumventing Social Desirability Response Bias in Personal Interview Surveys, *American Journal of Mathematical and Management Sciences*, **70**, 275-278.

Gupta, S. N., Gupta, B. C. and Singh, S. (2002), Estimation of Sensitivity Level of Personal Interview Survey Questions, *Journal of Statistical Planning and Inference*, **100**, 239-247.

Gupta, S. and Shabbir, J. (2004), Sensitivity Estimation for Personal Interview Survey Questions, *Statistica* , **64**, 643-653.

Gupta, S. N., Thornton, B., Shabbir, J. and Singhal, S. (2006), A Comparison of Multiplicative and Additive Optional RRT Models, *Journal of Statistical Theory and Applications*, **5**, 226-239.

Gupta, S., Shabbir, J. and Sehra, S. (2010), Mean and Sensitivity Estimation in Optional Randomized Response Models, *Journal of Statistical Planning and Inference*, **140**, 2870-2874.

Krumpal, I. (2012), Estimating the Prevalence of Xenophobia and Anti-Semitism in Germany: A Comparison of Randomized Response and Direct Questioning,*Social Science Research*, **41**, 1387-1403.

Mehta, S., Dass, B. K., Shabbir, J. and Gupta, S. (2012), A Three-Stage Optional Randomized Response Model, *Journal of Statistical Theory and Practice*, **6**, 417-427.

Rice, J. (2007), *Mathematical Statistics and Data Analysis*, 3rd Ed., Duxbury Press, Belmont.

Warner, S. L. (1965), Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias, *Journal of the American Statistical Association*, **60**, 63-69.

Warner, S. L. (1971), The Linear Randomized Response Model, *Journal of the American Statistical Association*, **66**, 884-888.