مجله علوم آماری، پاییز و زمستان ۱۳۹۸
جلد ۱۳، شماره ۲، ص ۴۹۷ – ۵۱۳
DOI: 10.29252/jss.13.2.497

مدل تعداد اقلام تصادفیده

سید محمدرضا علوی، سارا نیری مازی و محمدرضا آخوند
گروه آمار، دانشگاه شهید چمران اهواز

تاریخ دریافت: ۱۳۹۶/۱۲/۵ تاریخ آخرین بازنگری: ۱۳۹۷/۶/۱۶

چکیده: در بسیاری از آمارگیری‌های نمونه‌ای متغیّرهای مورد علاقه همچون تقلب دانشجو دارای ماهیت حسّاس هستند. در چنین موقعیت‌هایی افراد به سؤال مستقیم پاسخ‌های نادرست می‌دهند یا از پاسخ دادن امتناع می‌ورزند. روش‌های مختلف غیرمستقیم از جمله روش پاسخ تصادفیده و روش تعداد اقلام برای جمع‌آوری اطلاعات حساس معرفی شده است. در این مقاله ابتدا یک روش تعداد اقلام جدید معرفی شده و سپس گونه تصادفیده آن با نام مدل تعداد اقلام تصادفیده معرفی می‌شود. برآوردی نااریب برای نسبت حساس با استفاده از این مدل به‌دست آورده می‌شود. واریانس این برآوردکننده و برآوردی برای این واریانس معرفی می‌شود. یک معیار کمی برای مقایسه توأم کارایی و محرمانگی معرفی می‌شود. با استفاده از شبیه‌سازی مدل پیشنهادی ارزیابی و کارایی و محرمانگی آن با روش تصادفیده سیمونس مقایسه می‌شود. براساس این معیار برتری روش پیشنهادی بر سیمونس نشان داده می‌شود. نسبت تقلب دانشجویان در دانشگاه شهید چمران اهواز با استفاده از مدل پیشنهادی برآورد می‌شود.

واژه‌های کلیدی: پاسخ تصادفیده، تعداد اقلام تصادفیده، روش تعداد اقلام، متغیر حسّاس، نسبت تقلب دانشجویان.

۱ مقدمه

در همه جوامعی که با انسان سروکار دارند نیاز اساسی به روش‌هایی برای برآورد نسبت ویژگی‌های حسّاس، احساس می‌شود. برآورد نسبت ویژگی حسّاس می‌تواند برای ارائه برنامه‌های راهبردی و برنامه‌ریزی مورد

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

کد موضوع‌بندی ریاضی (۲۰۱۰): 62D05.

استفاده قرار گیرد. معمولاً در چنین مواردی برای اجتناب از بی‌پاسخی بیش از حد یا پاسخ‌های گمراه‌کننده از روش‌های غیرمستقیم برای جمع آوری‌اطلاعات استفاده می‌شود. روش‌های پاسخ‌تصادفیده و تعداد اقلام از جمله مهمترین روش‌های غیرمستقیم هستند. روش پاسخ تصادفیده اوّلین بار توسط وارنر (۱۹۶۵) ابداع شد. در این روش پاسخ دهنده با انجام یک آزمایش برنولی با احتمال پیروزی p سؤال حساس (متغیر برنولی S) و با احتمال شکست $۱-p$ ، مکمل آن $(۱-S)$ را پاسخ می‌دهد.

پاسخ تصادفیده به روش وارنر به‌صورت یک متغیر برنولی

$$Z = \begin{cases} S & \text{با احتمال } p \\ ۱-S & \text{با احتمال } ۱-p, \end{cases}$$

با احتمال موفقیت $\lambda = E(Z) = p\pi + (۱-p)(۱-\pi)$ نشان داده می‌شود ، که در آن π بیانگر نسبت ویژگی حساس است $E(S) = \pi$. بنابراین اگر $Z_۱, \dots, Z_n$ پاسخ‌های تصادفیده یک نمونه تصادفی از افراد براساس روش وارنر باشند، یک برآوردکننده ناریب برای π به صورت $\hat{\pi}_w = \frac{\bar{Z}-(۱-p)}{۲p-۱}$ است، که در آن $\bar{Z}$ میانگین نمونه پاسخ تصادفیده را نشان می‌دهد.

۱.۱ روش پاسخ تصادفیده سیمونس

گرینبرگ و همکاران (۱۹۶۹) روش سیمونس را با جایگذاری یک سؤال غیرحساس نامرتبط (متغیر برنولیT) به‌جای سؤال مکمل در روش وارنر معرفی کردند. آنها ادعا کردند که چون از جنبه تصادفی بیشتری استفاده شده، محرمانگی روش آنها بیشتر از روش وارنر است.

پاسخ تصادفیده به روش سیمونس می‌تواند به‌صورت یک متغیر برنولی

$$Z = \begin{cases} S & \text{با احتمال } p \\ T & \text{با احتمال } ۱-p, \end{cases}$$

با احتمال موفقیت $\lambda = E(Z) = p\pi + (۱-p)\theta$ بیان شود، که در آن θ بیانگر نسبت ویژگی غیرحساس است $E(T) = \theta$. بنابراین اگر $Z_۱, \dots, Z_n$ پاسخ‌های تصادفیده یک نمونه تصادفی از افراد براساس روش سیمونس و θ معلوم باشد، یک برآوردکننده نااریب برای π برابر $\hat{\pi}_S = \frac{\bar{Z}-(۱-p)\theta}{p}$ است. در روش‌های پاسخ تصادفیده از پاسخ‌دهنده خواسته می‌شود که پاسخ خود به سؤال حساس را

به‌گونه‌ای تصادفی بیان کند. این کار براساس ترفندهای تصادفی خاصی انجام می‌شود. برخی از ابزار تصادفی کردن شامل پرتاب تاس، پرتاب سکه، استفاده از دسته کارت و جدول اعداد تصادفی هستند. از آنجا که پاسخ به‌دست آمده توسط این ابزار تصادفی است، در نهایت پاسخ پاسخ‌دهنده مخفی می‌ماند. روش‌های پاسخ‌های تصادفیده را می‌توان به دو دسته اصلی کیفی[۱] و کمی[۲] طبقه‌بندی کرد. منگات و سینگ (۱۹۹۰)، منگات (۱۹۹۴)، کیم و وارد (۲۰۰۴)، جستوانگ و سینگ (۲۰۰۶)، چادری (۲۰۱۰)، علوی و تاج‌الدینی(۱۳۹۴ و ۲۰۱۶) از جمله نویسندگانی هستند که روش‌های کیفی را مطالعه کردند. ایچهارن و هایر (۱۹۸۳)، یزاری و علوی (۱۳۹۳)، مهتا و همکاران (۲۰۱۲)، گوپتا و همکاران (۲۰۱۳) از جمله کسانی هستند که روش‌های کمی را مطالعه کرده‌اند.

۲.۱ روش تعداد اقلام

روش تعداد اقلام[۳] (ICT) اولین بار توسط اسمیت و همکاران (۱۹۷۵) تحت عنوان روش پاسخ مجموع بلوکی[۴] معرفی شد و بعد از آن میلر (۱۹۸۴) آن را توسعه داد. در این روش مصاحبه شوندگان به‌طور تصادفی به دو زیرنمونه تقسیم می‌شوند و به هر کدام از اعضای نمونه اول فهرستی از g سؤال غیرحساس داده می‌شود و به هر کدام از اعضای زیرنمونه دوم فهرستی از $g + 1$ سؤال داده می‌شود که مشتمل بر همان g سؤال غیرحساس به همراه یک سؤال حساس است. تمام سوالات حساس و غیرحساس دارای پاسخ‌های دو حالتی (بله و خیر) هستند. از پاسخ دهندگان در هر زیر نمونه خواسته می‌شود که فقط تعداد پاسخ "بله" را به محقق گزارش دهند بدون اینکه پاسخ هر قلم از سوالات را بیان کند. برآوردگر نااریب نسبت حساس در جامعه را می‌توان با پیدا کردن تفاوت میانگین تعداد پاسخ "بله" در دو زیرنمونه تصادفی پیدا کرد. کاتس و جان (۲۰۱۱) این روش را با روش‌های پاسخ‌های تصادفیده مقایسه کرده‌اند. چادری و کریستوفیدز (۲۰۰۷)، تسچیا و همکاران (۲۰۰۷)، ایما (۲۰۱۱)، حسین و همکاران (۲۰۱۲) و تیان و همکاران (۲۰۱۵) از دیگر نویسندگانی هستند که روش تعداد اقلام را مطالعه کرده‌اند. انگیزه اصلی در اکثر تحقیقات در مطالعات ویژگی حساس، بیان روشی است که درجه محرمانگی را افزایش دهد. یک معیار شهودی که معمولاً استفاده می‌شود افزایش جنبه تصادفی کردن پاسخ به سؤال حساس است، به‌گونه‌ای که پاسخگو احساس کند پاسخ او به سؤال حساس محرمانه می‌ماند. هرچه جنبه تصادفی کردن در

[1]Qualitative Response Models
[2]Quantitative Response Models
[3]Item Count Technique
[4]Block Total Response Technique

روش غیرمستقیم بالا رود، حفاظت از محرمانگی بیشتر می‌شود. از طرفی افزایش جنبه تصادفی از لحاظ آماری سبب افزایش واریانس برآوردکننده می‌شود. بنابراین با مرور کلی روش‌های غیرمستقیم (روش‌های پاسخ تصادفیده و روش تعداد اقلام) نویسندگانی زیادی اقدام به معرفی روش‌های جدیدی کرده و ادعا کرده‌اند که روش آنها سبب افزایش محرمانگی می‌شود. دلیل اکثر آنها محاسبه یک معیار کمی صریح نبوده استفاده بیشتر از جنبه تصادفی ملاک آن‌ها بوده است، به‌عنوان مثال گرینبرگ و همکاران (۱۹۶۹) ادعا کردند که روش آنها (روش سیمونس) نسبت به روش وارنر سبب افزایش محرمانگی می‌شود. چون در مقایسه با روش وارنر یک جنبه تصادفی بیشتر دارد و آن استفاده از سؤال بدون ارتباط غیرحساس بود. با اضافه کردن مراحلی که جنبه تصادفی داشت روش‌های تصادفی دو مرحله‌ای زیادی از جمله منگات و سینگ (۱۹۹۰)، منگات (۱۹۹۴) و کیم و وارد (۲۰۰۴) و روش‌های چند مرحله‌ای چون جستوانگ و سینگ (۲۰۰۶)، چادری (۲۰۱۰)، علوی و تاج الدینی (۱۳۹۴ و ۲۰۱۶)، مهتا و همکاران (۲۰۱۲)، گوپتا و همکاران (۲۰۱۳) و یزاری و علوی (۱۳۹۳) برای افزایش محرمانگی توسط نویسندگانی زیادی ابداع شدند. با مطالعه این روش‌ها در می‌یابیم انگیزه ابداع آنها صرفاً افزایش محرمانگی بر اساس افزایش جنبه تصادفی بوده و لزوماً این روش‌ها کارایی بیشتری نسبت به روش‌های قبلی از جمله روش وارنر و سیمونس نداشته‌اند. البته برای روش‌های تصادفیده‌ای که پاسخ تصادفیده آن‌ها به‌صورت بله و خیر بود، دو معیار کمی توسط لانک (۱۹۷۶) و لیسیور (۱۹۷۹) که باز جنبه شهودی داشتند بر مبنای محاسبه احتمال شرطی پاسخ تصادفیده بله به شرط وقوع خصوصیت حساس معرفی شدند. آنها تحت فرض یکسان بودن این معیارها واریانس یک روش را نسبت به دیگری مقایسه می‌کردند. طبق این دو معیار گوئریرو و ساردینی (۲۰۰۷) بیان داشتند که به‌ندرت روشی پیدا می‌شود که به‌طور یکنواخت کمترین واریانس را داشته باشد و بیان کردند، که در مقایسه دو مدل تصادفیده با پاسخ بله و خیر، تحت بعضی شرایط مدل اول برتری دارد و تحت شرایط دیگر مدل دوم برتر است. به‌هرحال به دلیل پیچیدگی محاسبه این دو معیار در بسیاری از روش‌های پاسخ تصادفیده، و عدم محاسبه آن‌ها در روش‌هایی که پاسخ تصادفیده آنها به‌صورت یک عدد حقیقی است هم‌چون روش جستوانگ و سینگ (۲۰۰۶) و روش تعداد اقلام (به دلیل عدم محاسبه احتمال شرطی) مورد اقبال محققان زیادی قرار نگرفته است و هنوز مبنای شهودی افزایش جنبه تصادفی ملاک افزایش محرمانگی معرفی روش‌های جدید است.

در مطالعات ویژگی حساس، حفاظت از محرمانگی بیش از کارایی اهمیت دارد، زیرا با افزایش نمونه می‌توان واریانس را کاهش داد اما حتی با سرشماری نمی‌توان حفاظت از محرمانگی را افزایش داد. به عبارت دیگر انتظار می‌رود با افزایش حجم نمونه روش تصادفیده برآورد معتبری برای نسبت حساس تعیین

کند، در حالی‌که حتی با سرشماری، به روش مستقیم نمی‌توان نسبت معتبری برای نسبت حساس به‌دست آورد.

در بخش ۲ ابتدا یک گونه جدید تعداد اقلام براساس یک نمونه معرفی و سپس یک مدل جدید به‌نام تعداد اقلام تصادفیده برای برآورد نسبت حساس پیشنهاد و واریانس آن به‌همراه یک برآورد برای آن معرفی می‌شود. در بخش ۳ یک معیارجدید برای مقایسه توأم کارایی و محرمانگی معرفی می‌شود. با استفاده از شبیه‌سازی مدل پیشنهادی ارزیابی و کارایی و محرمانگی آن با روش تصادفیده سیمونس مقایسه می‌شود و بر اساس این معیار جدید نشان داده می‌شود که مدل پیشنهادی از مدل سیمونس بهتر است. در بخش ۴ نسبت تقلب در دانشگاه شهید چمران اهواز با استفاده از مدل پیشنهادی برآورد می‌شود. در بخش ۵ بحث و نتیجه‌گیری بیان می‌شود.

۲ مدل تعداد اقلام تصادفیده پیشنهادی

در این بخش، چون اجرای نمونه‌گیری با دو نمونه هزینه بیشتری دارد، برای کاهش هزینه و راحتی اجرای نمونه‌گیری، ابتدا یک روش جدید تعداد اقلام براساس یک نمونه معرفی می‌شود. در این روش به هر مصاحبه‌شونده پرسشنامه‌ای شامل $g + ۱$ سؤال ارائه می‌شود که j امین سؤال شامل پرسشی درباره یک متغیر برنولی غیرحسّاس T_j با احتمال پیروزی θ_j ($j = ۱, \ldots, g$) و آخرین سؤال درباره متغیر برنولی حسّاس S است. سپس از مصاحبه‌شونده درخواست می‌شود از مجموع پاسخ $g + ۱$ متغیر برنولی یک واحد را کم کند($X + S - ۱$) و نتیجه را در پاسخ‌نامه درج کند. اگر پاسخ تعداد اقلام را با U نمایش دهیم، $U = X + S - ۱$ که $X = \sum_{i=۱}^{g} T_i$ مجموع پاسخ g متغیر غیرحساس است. پاسخ تعداد اقلام U مقادیر از صفر تا g را اختیار می‌کند. فقط اگر پاسخ g باشد، محرمانگی حفظ نمی‌شود. میانگین U برابر

$$E(U) = \sum_{i=۱}^{g} \theta_i + \pi - ۱,$$

است. با فرض استقلال S و T_j ها واریانس U برابر

$$V(U) = \sum_{i=۱}^{g} \theta_i(۱ - \theta_i) + \pi(۱ - \pi),$$

است. بنابراین اگر $U_1, \dots, U_n$ پاسخ‌های تعداد اقلام یک نمونه تصادفی از افراد و θ_i ها معلوم باشند، یک برآورد کننده نااریب برای π برابر

$$\hat{\pi} = \bar{U} - \sum_{i=1}^{g} \theta_i + 1$$

و واریانس این برآوردکننده برابر

$$V(\hat{\pi}) = \frac{\sum_{i=1}^{g} \theta_i(1-\theta_i) + \pi(1-\pi)}{n}$$

است. متأسفانه در این روش اگر پاسخ تعداد اقلام g باشد، محرمانگی حفظ نمی‌شود. بنابراین برای حفظ هر چه بیشتر محرمانگی با تصادفیده کردن این روش تعداد اقلام، مدل جدیدی را با نام تعداد اقلام تصادفیده پیشنهاد می‌شود. در این مدل پیشنهادی از مصاحبه‌شونده درخواست می‌شود یک آزمایش برنولی با احتمال پیروزی p انجام دهد. اگر پیروزی رخ دهد، X و در غیر این‌صورت U را در پاسخ‌نامه ثبت کند. بنابراین اقلام تصادفیده برابر

$$Z = \begin{cases} X & \text{با احتمال } p \\ U & \text{با احتمال } 1-p \end{cases}$$

است. اگر W، T_j و S متغیرهای برنولی به‌ترتیب با احتمالات پیروزی p ، θ_j و π باشند، پاسخ تعداد اقلام تصادفیده به‌صورت

$$Z = WX + (1-W)(X+S-1) = X + (1-W)(S-1)$$

نیز قابل بازنویسی است. واضح است که Z یک متغیر تصادفی گسسته است که مقادیر صفر تا g را با تابع احتمال آمیخته

$$P(Z=z) = f_Z(z) = pf_X(z) + (1-p)f_U(z), \qquad (1)$$

اختیار می‌کند، که در آن $f_X(.)$ و $f_U(.)$ به‌ترتیب تابع احتمال متغیرهای تصادفی X و U هستند. بنا به قاعده میانگین و واریانس مضاعف میانگین و واریانس Z به‌ترتیب به صورت

$$E(Z) = \sum_{i=1}^{g} \theta_i - (1-p) + (1-p)\pi, \tag{2}$$

$$V(Z) = \sum_{i=1}^{g} \theta_i(1-\theta_i) + (1-p)(1-\pi)[1-(1-p)(1-\pi)], \tag{3}$$

هستند. در حالت خاص $\theta_j = \theta$ ، رابطه (۱) به‌صورت

$$\begin{aligned} f_Z(z) &= p\binom{g}{z}\theta^z(1-\theta)^{g-z} \\ &+ (1-P)\sum_{j=0}^{g}\binom{g}{j}\theta^j(1-\theta)^{g-j}\pi^{z-1-j}(1-\pi)^{2-z+j} \end{aligned} \tag{4}$$

است. فرض کنید $Z_1, \ldots, Z_n$ پاسخهای تصادفیده افراد یک نمونه تصادفی به حجم n باشند، بنا به (۲) به روش گشتاوری یک برآورد کننده گشتاوری به‌صورت

$$\hat{\pi}_P = \frac{\bar{Z} - \sum_{i=1}^{g}\theta_i + (1-p)}{1-p} \tag{5}$$

با واریانس

$$V(\hat{\pi}_P) = \frac{V(Z)}{n(1-p)^2} = \frac{\sum_{i=1}^{g}\theta_i(1-\theta_i) + (1-p)(1-\pi)[1-(1-p)(1-\pi)]}{n(1-p)^2} \tag{6}$$

معرفی می‌گردد. یک برآورد نااریب برای این واریانس برابر

$$\hat{V}(\hat{\pi}_P) = \frac{s_Z^2}{n(1-p)^2} \tag{7}$$

است، که در آن $s_Z^۲$ واریانس نمونه پاسخ تصادفیده است. در حالت خاص برای $\theta_j = \theta$ ، داریم:

$$\hat{\pi}_P = \frac{\bar{Z} - g\theta + (۱ - p)}{۱ - p} \quad (۸)$$

$$V(\hat{\pi}_P) = \frac{g\theta(۱ - \theta) + (۱ - p)(۱ - \pi)[۱ - (۱ - p)(۱ - \pi)]}{n(۱ - p)^۲} \quad (۹)$$

۳ مطالعه شبیه‌سازی

برای ارزیابی برآورد روش پیشنهادی در ۱۰۰۰۰ تکرار شبیه‌سازی نمونه‌هایی به حجم ۲۰، ۵۰ و ۱۰۰ از توزیع برنولی با احتمال پیروزی معلوم π انتخاب کرده سپس داده‌های تعداد اقلام تصادفیده را در حالت $\theta_j = \theta$ با مقادیر مختلف θ و g به ازای $p = ۰/۵$ بنا به (۱) تولید کرده سپس برآورد π را با استفاده از رابطه (۸) برای هر تکرار محاسبه کردیم. میانگین و واریانس این ۱۰۰۰۰ تکرار به‌عنوان امید ریاضی و واریانس برآورد پیشنهادی درجداول ۱ تا ۳ آمده است. جهت مقایسه کارایی روش پیشنهادی با روش سیمونس در ۱۰۰۰۰ تکرار شبیه‌سازی نمونه‌هایی به حجم ۲۰، ۵۰ و ۱۰۰ از توزیع برنولی با احتمال پیروزی معلوم π انتخاب کرده سپس داده‌های پاسخ تصادفیده سیمونس را برای سؤال نامرتبط با احتمال ویژگی غیر حساس θ به ازای $p = ۰/۵$ و با مقادیر مختلف θ تولید کرده سپس برآورد π را با استفاده از برآورد سیمونس رابطه (؟؟) محاسبه کردیم. میانگین و واریانس این ۱۰۰۰۰ تکرار به‌عنوان امید ریاضی و واریانس برآورد سیمونس درجداول ۱ تا ۳ آمده است.

نتایج شبیه‌سازی نااریبی برآوردکننده‌های پیشنهادی و سیمونس را تأیید می‌کنند و با افزایش n واریانس هر دو روش کاهش پیدا می‌کند. هر چه θ به ۰/۵ نزدیک می‌شود واریانس ها افزایش می‌یابند. به‌ازای هر مقدار از n با افزایش g واریانس روش پیشنهادی افزایش پیدا می‌کند. با توجه به جداول ۱ تا ۳ با افزایش حجم نمونه و مقادیر مختلف θ ، واریانس برآوردکننده‌ها کاهش می‌یابد و در همه موارد واریانس روش پیشنهادی بیشتر از روش سیمونس است. البته این بزرگی واریانس دور از انتظار نبود چون، روش سیمونس در مقایسه با روش پیشنهادی فقط یک جنبه تصادفی دارد، در حالیکه در روش پیشنهادی از $g + ۲$ جنبه تصادفی استفاده شده است که محافظت از محرمانگی را به‌شدت افزایش می‌دهد. با افزایش حجم نمونه انتظار می‌رود که برای مقادیر کوچک g کارایی مدل پیشنهادی تقریباً نصف در مقابل محرمانگی ۵ برابر مدل سیمونس شود.

جدول ۱: اریبی و واریانس مدل‌های پیشنهادی و سیمونس برای $\theta = ۰/۱$ (با شرط $\theta_j = \theta$) و $p = ۰/۵$ برای مقادیر g و π با نمونه‌های به حجم ۲۰، ۵۰ و ۱۰۰ در ۱۰۰۰۰ تکرار شبیه‌سازی.

n	g	$\sum_{i=1}^{g}\theta_i$	π									
			۰/۱		۰/۳		۰/۵		۰/۷		۰/۹	
			$E(\hat{\pi}_p)$	$V(\hat{\pi}_p)$	$E(\hat{\pi}_p)$	$V(\hat{\pi}_p)$	$E(\hat{\pi}_p)$	$V(\hat{\pi}_p)$	$E(\hat{\pi}_p)$	$V(\hat{\pi}_p)$	$E(\hat{\pi}_p)$	$V(\hat{\pi}_p)$
۲۰	۳	۰/۳	۰/۱۰۱	۰/۱۰۷	۰/۲۹۸	۰/۰۹۹	۰/۵۰۰	۰/۰۸۹	۰/۶۹۶	۰/۰۸۰	۰/۸۹۰	۰/۰۹۴
	۴	۰/۴	۰/۰۹۹	۰/۱۲۱	۰/۲۹۳	۰/۱۱۳	۰/۵۰۳	۰/۱۰۹	۰/۷۰۱	۰/۰۹۷	۰/۸۹۴	۰/۰۸۲
	۵	۰/۵	۰/۰۹۶	۰/۱۴۰	۰/۳۰۲	۰/۱۳۷	۰/۴۹۷	۰/۱۲۳	۰/۶۹۹	۰/۱۸۸	۰/۹۰۵	۰/۱۰۰
	۶	۰/۶	۰/۱۰۰	۰/۱۵۶	۰/۳۰۳	۰/۱۵۴	۰/۵۰۷	۰/۱۴۵	۰/۷۰۵	۰/۱۳۴	۰/۸۹۸	۰/۱۱۶
		سیمونس	۰/۱۰۳	۰/۰۱۸	۰/۳۰۲	۰/۰۳۱	۰/۵۰۵	۰/۰۴۳	۰/۷۰۰	۰/۰۴۸	۰/۸۹۷	۰/۰۴۹
۵۰	۳	۰/۳	۰/۰۹۸	۰/۰۴۲	۰/۲۹۹	۰/۰۳۴	۰/۵۰۲	۰/۰۳۶	۰/۷۰۰	۰/۰۳۲	۰/۸۹۹	۰/۰۲۶
	۴	۰/۴	۰/۱۰۰	۰/۰۴۷	۰/۲۹۵	۰/۰۴۳	۰/۴۹۷	۰/۰۴۳	۰/۷۰۰	۰/۰۳۹	۰/۹۰۰	۰/۰۳۳
	۵	۰/۵	۰/۰۹۹	۰/۰۵۵	۰/۳۰۲	۰/۰۵۰	۰/۵۰۰	۰/۰۵۰	۰/۷۰۰	۰/۰۴۶	۰/۸۹۹	۰/۰۳۸
	۶	۰/۶	۰/۰۹۸	۰/۰۶۴	۰/۳۰۰	۰/۰۶۰	۰/۴۹۹	۰/۰۵۸	۰/۷۰۰	۰/۰۵۴	۰/۹۰۲	۰/۰۴۷
		سیمونس	۰/۰۹۹	۰/۰۰۷	۰/۳۰۲	۰/۰۱۳	۰/۴۹۸	۰/۰۱۶	۰/۶۹۹	۰/۰۱۹	۰/۹۰۱	۰/۰۲۰
۱۰۰	۳	۰/۳	۰/۱۰۲	۰/۰۲۱	۰/۳۰۰	۰/۰۲۰	۰/۵۰۲	۰/۰۱۸	۰/۶۹۹	۰/۰۱۶	۰/۹۰۱	۰/۰۱۳
	۴	۰/۴	۰/۱۰۰	۰/۰۲۴	۰/۲۹۵	۰/۰۲۴	۰/۴۹۸	۰/۰۲۲	۰/۶۹۷	۰/۰۱۸	۰/۹۰۱	۰/۰۱۶
	۵	۰/۵	۰/۱۰۲	۰/۰۲۹	۰/۳۰۱	۰/۰۲۷	۰/۴۹۷	۰/۰۲۵	۰/۷۰۰	۰/۰۲۲	۰/۸۹۹	۰/۰۱۹
	۶	۰/۶	۰/۱۰۳	۰/۰۳۱	۰/۲۹۹	۰/۰۳۱	۰/۴۹۹	۰/۰۲۹	۰/۷۰۱	۰/۰۲۶	۰/۸۹۹	۰/۰۲۳
		سیمونس	۰/۱۰۰	۰/۰۰۴	۰/۳۰۰	۰/۰۰۶	۰/۵۰۰	۰/۰۰۸	۰/۷۰۰	۰/۰۰۹	۰/۹۰۱	۰/۰۱۰

جدول ۲: اریبی و واریانس مدل‌های پیشنهادی و سیمونس برای $\theta = ۰/۵$ (با شرط $\theta_j = \theta$) و $p = ۰/۵$ برای مقادیر g و π با نمونه‌های به حجم ۲۰، ۵۰ و ۱۰۰ در ۱۰۰۰۰ تکرار شبیه‌سازی.

n	g	$\sum_{i=1}^{g}\theta_i$	π									
			۰/۱		۰/۳		۰/۵		۰/۷		۰/۹	
			$E(\hat{\pi}_p)$	$V(\hat{\pi}_p)$	$E(\hat{\pi}_p)$	$V(\hat{\pi}_p)$	$E(\hat{\pi}_p)$	$V(\hat{\pi}_p)$	$E(\hat{\pi}_p)$	$V(\hat{\pi}_p)$	$E(\hat{\pi}_p)$	$V(\hat{\pi}_p)$
۲۰	۳	۱/۵	۰/۰۹۹	۰/۲۰۳	۰/۳۰۳	۰/۱۹۰	۰/۵۰۴	۰/۱۸۸	۰/۷۰۲	۰/۱۷۷	۰/۸۹۸	۰/۱۵۷
	۴	۲	۰/۱۰۳	۰/۲۴۹	۰/۲۸۶	۰/۲۴۴	۰/۴۴۹	۰/۲۳۶	۰/۷۰۲	۰/۲۲۴	۰/۸۹۷	۰/۲۰۸
	۵	۲/۵	۰/۱۰۳	۰/۲۹۷	۰/۲۹۷	۰/۳۰۷	۰/۵۰۱	۰/۲۲۸	۰/۷۰۰	۰/۲۷۶	۰/۸۹۸	۰/۲۵۰
	۶	۳	۰/۰۹۸	۰/۳۳۵	۰/۳۰۱	۰/۳۴۲	۰/۴۹۸	۰/۳۳۸	۰/۶۹۱	۰/۳۳۰	۰/۹۰۱	۰/۳۰۶
		سیمونس	۰/۱۰۰	۰/۰۴۲	۰/۲۹۸	۰/۰۴۹	۰/۵۰۰	۰/۰۵۰	۰/۷۰۲	۰/۰۴۸	۰/۹۰۰	۰/۰۴۳
۵۰	۳	۱/۵	۰/۰۹۸	۰/۰۸۰	۰/۳۰۴	۰/۰۷۷	۰/۴۹۵	۰/۰۷۶	۰/۶۹۷	۰/۰۶۸	۰/۹۰۲	۰/۰۶۲
	۴	۲	۰/۱۰۱	۰/۰۹۸	۰/۲۹۸	۰/۰۹۶	۰/۴۹۷	۰/۰۹۲	۰/۶۹۸	۰/۰۹۲	۰/۸۹۸	۰/۰۸۴
	۵	۲/۵	۰/۰۹۷	۰/۱۲۲	۰/۲۹۸	۰/۱۲۲	۰/۵۰۶	۰/۱۱۶	۰/۷۰۴	۰/۱۱۰	۰/۹۰۷	۰/۱۰۳
	۶	۳	۰/۰۹۶	۰/۱۳۸	۰/۳۰۳	۰/۱۳۵	۰/۴۹۳	۰/۱۳۷	۰/۷۰۲	۰/۱۳۰	۰/۹۰۳	۰/۱۲۴
		سیمونس	۰/۰۹۸	۰/۰۱۷	۰/۳۰۱	۰/۰۱۹	۰/۴۹۸	۰/۰۲۰	۰/۷۰۲	۰/۰۱۹	۰/۹۰۲	۰/۰۱۷
۱۰۰	۳	۱/۵	۰/۱۰۱	۰/۰۴۰	۰/۲۹۸	۰/۰۳۸	۰/۵۰۰	۰/۰۳۸	۰/۷۰۴	۰/۰۳۵	۰/۹۰۲	۰/۰۳۲
	۴	۲	۰/۰۹۹	۰/۰۵۰	۰/۳۰۱	۰/۰۵۱	۰/۴۹۶	۰/۰۴۷	۰/۶۹۹	۰/۰۴۵	۰/۹۰۱	۰/۰۴۱
	۵	۲/۵	۰/۱۰۲	۰/۰۶۰	۰/۳۰۰	۰/۰۶۰	۰/۴۹۸	۰/۰۵۸	۰/۷۰۲	۰/۰۵۴	۰/۸۹۷	۰/۰۵۲
	۶	۳	۰/۱۰۲	۰/۰۷۰	۰/۳۰۶	۰/۰۶۹	۰/۵۰۱	۰/۰۶۷	۰/۶۹۴	۰/۰۶۵	۰/۸۹۹	۰/۰۶۱
		سیمونس	۰/۱۰۱	۰/۰۰۸	۰/۳۰۰	۰/۰۱۰	۰/۵۰۱	۰/۰۱۰	۰/۷۰۰	۰/۰۱۰	۰/۹۰۱	۰/۰۰۸

جدول ۳: اریبی و واریانس مدل‌های پیشنهادی و سیمونس برای $\theta = ۰٫۹$ (با شرط $\theta_j = \theta$) و $p = ۰٫۵$ برای مقادیر g و π با نمونه‌های به حجم ۲۰، ۵۰ و ۱۰۰ در ۱۰۰۰۰ تکرار شبیه سازی.

			π									
			۰٫۱		۰٫۳		۰٫۵		۰٫۷		۰٫۹	
n	g	$\sum_{i=1}^{g} \theta_i$	$E(\hat{\pi}_p)$	$V(\hat{\pi}_p)$	$E(\hat{\pi}_p)$	$V(\hat{\pi}_p)$	$E(\hat{\pi}_p)$	$V(\hat{\pi}_p)$	$E(\hat{\pi}_p)$	$V(\hat{\pi}_p)$	$E(\hat{\pi}_p)$	$V(\hat{\pi}_p)$
۲۰	۳	۲٫۷	۰٫۰۹۸	۰٫۱۰۲	۰٫۳۰۰	۰٫۰۹۹	۰٫۵۰۲	۰٫۰۹۲	۰٫۷۰۰	۰٫۰۷۸	۰٫۹۰۰	۰٫۰۶۲
	۴	۳٫۶	۰٫۱۰۰	۰٫۱۲۱	۰٫۲۹۵	۰٫۱۱۸	۰٫۵۰۱	۰٫۱۰۸	۰٫۷۰۰	۰٫۰۹۶	۰٫۹۰۲	۰٫۰۸۰
	۵	۴٫۵	۰٫۰۹۷	۰٫۱۴۰	۰٫۲۹۳	۰٫۱۳۴	۰٫۵۰۰	۰٫۱۲۷	۰٫۶۹۷	۰٫۱۱۶	۰٫۸۹۸	۰٫۱۰۳
	۶	۵٫۴	۰٫۱۰۸	۰٫۱۵۹	۰٫۳۰۰	۰٫۱۵۴	۰٫۴۹۷	۰٫۱۵۴	۰٫۷۰۰	۰٫۱۳۵	۰٫۹۰۴	۰٫۱۲۱
	سیمونس		۰٫۱۰۲	۰٫۰۴۹	۰٫۳۰۰	۰٫۰۴۷	۰٫۵۰۲	۰٫۰۴۱	۰٫۷۰۳	۰٫۰۳۱	۰٫۸۹۷	۰٫۰۱۸
۵۰	۳	۲٫۷	۰٫۰۹۸	۰٫۰۴۰	۰٫۲۹۹	۰٫۰۳۹	۰٫۵۰۰	۰٫۰۳۶	۰٫۶۹۸	۰٫۰۳۲	۰٫۸۹۹	۰٫۰۲۶
	۴	۳٫۶	۰٫۱۰۰	۰٫۰۴۹	۰٫۳۰۱	۰٫۰۴۶	۰٫۴۹۸	۰٫۰۴۳	۰٫۷۰۳	۰٫۰۳۸	۰٫۹۰۰	۰٫۰۳۲
	۵	۴٫۵	۰٫۱۰۱	۰٫۰۵۷	۰٫۲۹۷	۰٫۰۵۳	۰٫۵۰۰	۰٫۰۵۱	۰٫۶۹۸	۰٫۰۴۶	۰٫۸۹۹	۰٫۰۳۹
	۶	۵٫۴	۰٫۱۰۵	۰٫۰۶۳	۰٫۲۹۹	۰٫۰۶۱	۰٫۵۰۰	۰٫۰۵۷	۰٫۷۰۱	۰٫۰۵۵	۰٫۹۰۲	۰٫۰۴۷
	سیمونس		۰٫۱۰۰	۰٫۰۲۰	۰٫۳۰۰	۰٫۰۱۹	۰٫۵۰۱	۰٫۰۱۷	۰٫۷۰۰	۰٫۰۱۳	۰٫۹۰۰	۰٫۰۰۷
۱۰۰	۳	۲٫۷	۰٫۱۰۱	۰٫۰۲۱	۰٫۳۰۰	۰٫۰۲۰	۰٫۵۰۱	۰٫۰۱۸	۰٫۶۹۸	۰٫۰۱۶	۰٫۸۹۸	۰٫۰۱۲
	۴	۳٫۶	۰٫۱۰۰	۰٫۰۲۴	۰٫۲۹۹	۰٫۰۲۳	۰٫۴۹۹	۰٫۰۲۲	۰٫۶۹۹	۰٫۰۱۹	۰٫۸۹۹	۰٫۰۱۷
	۵	۴٫۵	۰٫۰۹۸	۰٫۰۲۸	۰٫۳۰۱	۰٫۰۲۷	۰٫۴۹۸	۰٫۰۲۶	۰٫۷۰۴	۰٫۰۲۳	۰٫۸۹۸	۰٫۰۱۹
	۶	۵٫۴	۰٫۱۰۱	۰٫۰۳۲	۰٫۲۹۸	۰٫۰۳۱	۰٫۵۰۱	۰٫۰۲۹	۰٫۷۰۰	۰٫۰۲۷	۰٫۸۹۸	۰٫۰۲۳
	سیمونس		۰٫۰۹۹	۰٫۰۱۰	۰٫۳۰۰	۰٫۰۱۰	۰٫۴۹۹	۰٫۰۰۸	۰٫۷۰۰	۰٫۰۰۶	۰٫۸۹۹	۰٫۰۰۴

تاکنون یک معیار کمی برای محاسبه میزان محرمانگی در روش‌هایی که پاسخ تصادفیده آنها یک عدد حقیقی است توسط محققین ارائه نشده است تا بتوان با استفاده از آن معیار محرمانگی روش پیشنهادی را تعیین کرد. در این مقاله بر اساس معیار شهودی تعداد جنبه تصادفی به‌کار رفته در روش‌های غیر مستقیم برای مقایسه برتری دو مدل تصادفیده، معیار جدید مجموع کارایی نسبی و محرمانگی نسبی [5] به‌صورت

$$SEP = \frac{V(۲)}{V(۱)} + \frac{PRIV(۱)}{PRIV(۲)}$$

ارائه می‌شود، که در آن $PRIV(i)$ و $V(i)$ به‌ترتیب محرمانگی و واریانس مدل i را نشان می‌دهند. دامنه این معیار مقادیر نامنفی است. اگر دو روش محرمانگی و کارایی یکسان داشته باشند مقدار این معیار ۲ است و مقدار بیشتر از ۲ بهتر بودن مدل اول را نسبت به دوم نشان می‌دهد. به صورت شهودی، حفظ محرمانگی روش پیشنهادی به‌دلیل استفاده از $g + ۲$ جنبه تصادفی $g + ۲$ برابر روش سیمونس است. هر چه g بزرگتر شود درجه محرمانگی افزایش و طبق رابطه (۱۶) کارایی کاهش می‌یابد. با کاهش g ، درجه محرمانگی کاهش و کارایی افزایش می‌یابد.

[5] Sum of efficiency and privacy

معیار SEP برای برتری مدل پیشنهادی نسبت به مدل سیمونس به‌صورت

$$SEP = \frac{V(S)}{V(P)} + \frac{PRIV(P)}{PRIV(S)}$$

است، که در آن $PRIV(P)$ و $PRIV(S)$ به‌ترتیب میزان محرمانگی روش پیشنهادی و روش سیمونس را نشان می‌دهند، و $V(P)$ و $V(S)$ به‌ترتیب واریانس برآورگر نسبت حساس روش پیشنهادی و روش سیمونس هستند. در جداول ۴ تا ۶ این معیار به ازای مقادیر مختلف حجم نمونه، g، θ و π محاسبه شده است، که نشان می‌دهد مدل پیشنهادی بهتر از مدل سیمونس می‌باشد. قسمت صحیح مقادیر این معیار نسبت محرمانگی و قسمت اعشار آن کارایی نسبی را نشان می‌دهد. برای اندازه نمونه بزرگ و g ثابت با افزایش π در بازه یکدهم تا هفت دهم، مقدار این معیار (قسمت اعشار) افزایش می‌یابد، یعنی برای موضوع حساس غیر نادر روش پیشنهادی بهتر از گونه حساس نادر، نسبت را برآورد می‌کند. چون محرمانگی به حجم نمونه بستگی ندارد، افزایش حجم نمونه فقط سبب افزایش کارایی می‌شود. بنابراین با افزایش حجم نمونه و g و π در بازه سه‌دهم و هفت دهم معیار افزایش می‌یابد.

جدول ۴: مقادیر معیار برتری مدل پیشنهادی نسبت به مدل سیمونس (SEP) برای $\theta = ۰/۱$ (با شرط $\theta_j = \theta$) و $p = ۰/۵$ برای مقادیر g و π با نمونه‌های به حجم ۲۰، ۵۰ و ۱۰۰ در ۱۰۰۰۰ تکرار شبیه‌سازی.

n	g	$\sum_{i=1}^{g}\theta_i$	π = ۰/۱	۰/۳	۰/۵	۰/۷	۰/۹
۲۰	۳	۰/۳	۵/۱۶۸	۵/۳۱۳	۵/۴۸۳	۵/۶۰۰	۵/۷۷۶
	۴	۰/۴	۶/۱۴۸	۶/۲۷۷	۶/۳۹۴	۶/۴۹۶	۶/۵۹۶
	۵	۰/۵	۷/۱۲۹	۷/۲۲۶	۷/۳۵۰	۷/۲۵۵	۷/۴۹۰
	۶	۰/۶	۸/۱۱۵	۸/۲۰۱	۸/۲۹۷	۸/۳۵۸	۸/۴۲۲
۵۰	۳	۰/۳	۵/۱۶۷	۵/۳۸۲	۵/۴۴۴	۵/۵۹۳	۵/۷۶۹
	۴	۰/۴	۶/۱۴۹	۶/۳۰۲	۶/۳۷۲	۶/۴۸۷	۶/۶۰۶
	۵	۰/۵	۷/۱۲۷	۷/۲۶۰	۷/۳۲۰	۷/۴۱۳	۷/۵۲۶
	۶	۰/۶	۸/۱۰۹	۸/۲۱۷	۸/۲۷۶	۸/۳۵۲	۸/۴۲۶
۱۰۰	۳	۰/۳	۵/۱۹۰	۵/۳۰۰	۵/۴۴۴	۵/۵۶۲	۵/۷۶۹
	۴	۰/۴	۶/۱۶۷	۶/۲۵۰	۶/۳۶۴	۶/۵۰۰	۶/۶۲۵
	۵	۰/۵	۷/۱۳۸	۷/۲۲۲	۷/۳۲۰	۷/۴۰۹	۷/۵۲۶
	۶	۰/۶	۸/۱۲۹	۸/۱۹۴	۸/۲۷۶	۸/۳۴۶	۸/۴۳۵

جدول ۵: مقادیر معیار برتری مدل پیشنهادی نسبت به مدل سیمونس (SEP) برای $\theta =$ ۰٫۵ (با شرط $\theta_j = \theta$) و $p =$ ۰٫۵ برای مقادیر g و π با نمونه‌های به حجم ۲۰، ۵۰ و ۱۰۰ در ۱۰۰۰۰ تکرار شبیه‌سازی.

		π					
۰٫۹	۰٫۷	۰٫۵	۰٫۳	۰٫۱	$\sum_{i=1}^{g}\theta_i$	g	n
۵٫۲۷۴	۵٫۲۷۱	۵٫۲۲۶	۵٫۲۲۱	۵٫۲۰۷	۱٫۵	۳	۲۰
۶٫۲۰۷	۶٫۲۱۴	۶٫۲۱۲	۶٫۱۷۲	۶٫۱۶۹	۲	۴	
۷٫۱۷۲	۷٫۱۷۴	۷٫۱۷۴	۷٫۱۳۷	۷٫۱۴۱	۲٫۵	۵	
۸٫۱۴۱	۸٫۱۴۵	۸٫۱۴۸	۸٫۱۲۳	۸٫۱۲۵	۳	۶	
۵٫۲۷۴	۵٫۲۷۹	۵٫۲۶۳	۵٫۲۴۷	۵٫۲۱۲	۱٫۵	۳	۵۰
۶٫۲۰۲	۶٫۲۰۷	۶٫۲۱۷	۶٫۱۹۸	۶٫۱۷۳	۲	۴	
۷٫۱۶۵	۷٫۱۷۳	۷٫۱۷۲	۷٫۱۵۶	۷٫۱۳۹	۲٫۵	۵	
۸٫۱۳۷	۸٫۱۴۶	۸٫۱۴۶	۸٫۱۴۱	۸٫۱۲۳	۳	۶	
۵٫۲۵۰	۵٫۲۸۸	۵٫۲۶۳	۵٫۲۶۳	۵٫۲۰۰	۱٫۵	۳	۱۰۰
۶٫۱۹۵	۶٫۲۲۲	۶٫۲۱۳	۶٫۱۹۶	۶٫۱۶۰	۲	۴	
۷٫۱۵۴	۷٫۱۸۵	۷٫۱۷۲	۷٫۱۶۷	۷٫۱۳۳	۲٫۵	۵	
۸٫۱۳۱	۸٫۱۵۴	۸٫۱۴۹	۸٫۱۴۸	۸٫۱۱۴	۳	۶	

جدول ۶: مقادیر معیار برتری مدل پیشنهادی نسبت به مدل سیمونس (SEP) برای $\theta =$ ۰٫۹ (با شرط $\theta_j = \theta$) و $p =$ ۰٫۵ برای مقادیر g و π با نمونه‌های به حجم ۲۰، ۵۰ و ۱۰۰ در ۱۰۰۰۰ تکرار شبیه‌سازی.

		π					
۰٫۹	۰٫۷	۰٫۵	۰٫۳	۰٫۱	$\sum_{i=1}^{g}\theta_i$	g	n
۵٫۲۹۰	۵٫۳۹۷	۵٫۵۱۶	۵٫۴۷۵	۵٫۴۸۰	۲٫۷	۳	۲۰
۶٫۲۲۵	۶٫۳۲۳	۶٫۲۸۷	۶٫۳۹۸	۶٫۴۰۵	۳٫۶	۴	
۷٫۱۷۵	۷٫۲۶۷	۷٫۲۴۴	۷٫۳۵۱	۷٫۳۵۰	۴٫۵	۵	
۸٫۱۴۹	۸٫۲۳۰	۸٫۳۰۱	۸٫۳۰۵	۸٫۳۰۸	۵٫۴	۶	
۵٫۲۶۹	۵٫۴۰۶	۵٫۴۷۲	۵٫۴۸۷	۵٫۵۰۰	۲٫۷	۳	۵۰
۶٫۲۱۹	۶٫۳۴۲	۶٫۳۹۵	۶٫۴۱۳	۶٫۴۰۸	۳٫۶	۴	
۷٫۱۷۹	۷٫۲۸۳	۷٫۳۳۳	۷٫۳۵۸	۷٫۳۵۱	۴٫۵	۵	
۸٫۱۴۹	۸٫۲۳۶	۸٫۲۹۸	۸٫۳۱۱	۸٫۳۱۷	۵٫۴	۶	
۵٫۳۳۳	۵٫۳۷۵	۵٫۴۴۴	۵٫۵۰۰	۵٫۴۷۶	۲٫۷	۳	۱۰۰
۶٫۲۳۵	۶٫۳۱۶	۶٫۳۶۴	۶٫۴۳۵	۶٫۴۱۷	۳٫۶	۴	
۷٫۲۱۱	۷٫۲۶۱	۷٫۳۰۸	۷٫۳۷۰	۷٫۳۷۵	۴٫۵	۵	
۸٫۱۷۴	۸٫۲۲۲	۸٫۲۷۶	۸٫۳۲۲	۸٫۳۱۲	۵٫۴	۶	

از طرفی حداکثر درجه محرمانگی برای انتخاب p در هر دو مدل در $p =$ ۰٫۵ رخ می‌دهد، چون در انتخاب بین دو گزینه $p =$ ۰٫۵ هیچکدام را بر دیگری ترجیح نمی‌دهد (گرینبرگ و همکاران، ۱۹۶۹). بنابراین در تمام جداول از $p =$ ۰٫۵ استفاده شده است. انتظار می‌رود که با استفاده از روش پیشنهادی افراد مشارکت بیشتری در بررسی داشته باشند و با ارائه داده‌های واقعی اعتبار استنباط را افزایش دهند. بنابراین در بخش کاربردی در طراحی پرسشنامه تعداد اقلام تصادفیده، از یک آزمایش تصادفی استفاده می‌شود که در آن احتمال پیروزی $p =$ ۰٫۵ باشد.

۴ برآورد نسبت تقلب در دانشگاه با روش پیشنهادی

در این بخش با استفاده از روش پیشنهادی نسبت تقلب دانشجویان در دانشگاه شهید چمران برآورد می‌شود. جامعه آماری در این پژوهش، دانشجویان دانشگاه شهید چمران اهواز در نظر گرفته شده‌است. داده‌ها از مهر ماه تا دی ماه سال ۱۳۹۵ جمع‌آوری شدند. ابزار پژوهش، پرسشنامه مبتنی بر روش پیشنهادی است که توسط نویسندگان طراحی شده و از سکه به عنوان سؤال غیرحساس و از تاس به عنوان ترفند تصادفی استفاده شده‌است. پرسشنامه به‌صورت زیر طراحی شده بود:

دانشجوی عزیز لطفاً یک سکه را چهار بار بیندازید و سؤالات زیر را در نظر بگیرید:

سؤال اول: آیا سکه در بار اول شیر آمده است؟

سؤال دوم: آیا سکه در بار دوم شیر آمده است؟

سؤال سوم: آیا سکه در بار سوم شیر آمده است؟

سؤال چهارم: آیا سکه در بار چهارم شیر آمده است؟

سؤال پنجم: آیا شما تاکنون در امتحانات تقلب کرده‌اید؟

اکنون تاس را بیندازید اگر زوج رخ دهد فقط تعداد بله چهار سؤال اول را در کادر پاسخنامه درج و در غیر اینصورت از تعداد بله پنج سؤال یک واحد کم و فقط آنرا در کادر پاسخنامه ثبت کنید.

پاسخنامه: $\boxed{}$

روش نمونه‌گیری به‌کار رفته در این پژوهش از نوع روش نمونه‌گیری طبقه‌بندی است که دانشکده‌ها به‌عنوان طبقه‌ها در نظر گرفته شده‌اند. از ۱۰ دانشکده نمونه‌ای به حجم ۶۸۱ دانشجو انتخاب و از آنها تقاضا شده پرسشنامه طراحی شده را پاسخ دهند. با توجه به پرسشنامه طراحی شده در این تحقیق $g =$ ۴ و در چهار سؤال اول $\theta_j = \theta =$ ۰٫۵ است. چون مناسبترین مقدار p برای حفظ محرمانگی مقدار ۰٫۵ است، برای تصادفی کردن تعداد اقلام از رخداد ''زوج بودن'' تاس استفاده شده است. با جایگذاری $g =$ ۴ ، $p =$ ۰٫۵

و $\theta = ۰/۵$ در رابطه‌های (۸) و (۷) برآورد نسبت تقلب و برآورد واریانس آن برای طبقه h (دانشکده) به‌ترتیب برابر

$$\hat{\pi}_{Ph} = \frac{\bar{Z}_h - ۱/۵}{۰/۵},$$
$$\hat{V}(\hat{\pi}_{Ph}) = \frac{\hat{V}_h(Z)}{۰/۲۵n_h} = \frac{(۱ - f_h)s_Z^۲}{۰/۲۵n_h},$$

هستند، که در آنها $(۱ - f_h)$ ضریب تصحیح واریانس برای نمونه‌گیری بدون جایگذاری در طبقه h است. در صورتی که تعداد طبقات L باشد، برآورد نسبت تقلب برای جامعه براساس روش پیشنهادی برابر

$$\hat{\pi}_P = \sum_{h=۱}^{L} W_h \hat{\pi}_{Ph}, \tag{۱۰}$$

است، که در آن $W_h = \frac{N_h}{N}$ وزن، N_h حجم طبقه h و $N = \sum_{h=۱}^{L} N_h$ حجم کل جامعه است. واریانس $\hat{\pi}_P$ برابر $\hat{V}(\hat{\pi}_P) = \sum_{h=۱}^{L} W_h^۲ \hat{V}(\hat{\pi}_{Ph})$ است. در این پژوهش تعداد طبقات (دانشکده‌ها) $L = ۱۰$ است.

در جدول ۷ خلاصه‌ای از اطلاعات نمونه‌گیری پاسخ تصادفیده روش پیشنهادی آورده شده است. برآورد نسبت تقلب دانشکده‌ها به‌همراه برآورد واریانس آنها در دو ستون آخر جدول ۷ نشان داده شده است.

جدول ۷: خلاصه اطلاعات نمونه‌گیری و برآورد نسبت تقلب در دانشکده‌ها

دانشکده h	N_h	n_h	$\bar{Z}_h$	s_h	$\hat{\pi}_P h$	$\hat{V}\hat{\pi}_P h$
ادبیات	۷۸۴	۶۶	۱/۹۲۴	۰/۷۷۱	۰/۳۹۳	۰/۰۳۳
اقتصاد	۱۵۰۶	۵۴	۱/۶۲۹	۷۸۴	۰/۲۵۹	۰/۰۴۴
الهیات	۶۰۰	۶۰	۱/۶۰۰	۰/۸۲۴	۰/۲۰۷	۰/۰۴۷
تربیت بدنی	۱۹۰	۴۳	۱/۷۴۴	۰/۸۵۹	۰/۴۸۸	۰/۰۵۳
علوم‌پایه	۷۶۳	۱۰۷	۱/۷۶۶	۰/۸۴۹	۰/۵۳۲	۰/۰۲۶
علوم آب	۱۸۳	۴۳	۲/۳۰۲	۰/۹۸۶	۱/۶۰۴	۰/۰۶۷
علوم تربیتی	۹۹۰	۹۳	۱/۸۴۹	۱/۲۶۳	۰/۶۹۸	۰/۰۶۲
علوم ریاضی	۸۰۰	۷۷	۱/۷۲۹	۰/۹۳۱	۰/۵۸۴	۰/۰۴۱
کشاورزی	۵۵۵	۵۰	۱/۸۶۰	۱/۰۴۹	۰/۷۲۰	۰/۰۸۰
مهندسی	۱۶۵۲	۸۸	۱/۷۳۸	۰/۸۵۰	۰/۴۷۷	۰/۰۳۱

بنابراین یک برآورد نااریب برای نسبت تقلب دانشجویان دانشگاه براساس روش پیشنهادی بنا بر رابطه (۱۰) برابر ۰/۴۰ است. با جایگذاری مقادیر در رابطه (؟؟) و جذرگیری از آن خطای معیار برآورد ۰/۱۷۵ می‌باشد.

۵ بحث و نتیجه‌گیری

در این مطالعه ابتدا برای کاهش هزینه نمونه‌گیری در بررسی ویژگی حساس، یک روش جدید تعداد اقلام براساس یک نمونه معرفی شد. برای افزایش محرمانگی روش، گونه تصادفیده آن با نام تعداد اقلام تصادفیده معرفی شد. برآورد نااریبی برای ویژگی حساس به همراه واریانس آن به‌دست آورده شد و یک برآورد نااریب برای واریانس نیز معرفی شد. یک معیار جدید برای مقایسه توأم کارایی و محرمانگی معرفی گردید. به‌کمک شبیه‌سازی، روش پیشنهادی با روش سیمونس مقایسه و نشان داده شد که با این معیار روش پیشنهادی از روش سیمونس بهتر است. با استفاده از روش پیشنهادی نسبت تقلب دانشجویان در دانشکده‌های دانشگاه شهید چمران اهواز برآورد براساس یک نمونه طبقه‌بندی به حجم ۶۸۱ دانشجو ۰/۴۰ با خطای معیار ۰/۱۷۵ محاسبه شد.

تقدیر و تشکر

نویسندگان از پیشنهادهای ارزنده داوران، هیئت‌تحریریه و ویراستار محترم مجله علوم آماری که باعث ارتقای مقاله شد، کمال تشکر و قدردانی را دارند.

مراجع

علوی، س. م. ر. و تاج الدینی، م. (۱۳۹۴)، یک روش پاسخ تصادفیده جدید و مقایسه آن با روش سیمونس، مجله علوم آماری ایران، جلد ۹، شماره ۲، ۲۲۷–۲۳۹.

یزاری، ز و علوی، س. م. ر . (۱۳۹۳)، مدل پاسخ تصادفیده ی کمّی اختیاری سه مرحله‌ای، مجله علوم آماری ایران، جلد ۸، شماره ۲، ۲۴۵–۲۶۰.

Alavi, S. M. R. and Tajodini (2016), Maximum likelihood Estimation of Sensitive Proportion Using Repeated Randomized Response Techniques, *Journal of Applied Statistics*, **43**, 563-571.

Chaudhuri, A .(2010), *Randomized Response and Indirect Questioning Techniques in Surveys*, New York, CRC Press.

Chaudhuri, A. and Christofides, T. (2007), Item Count Technique in Estimating the Proportion of People with a Sensitive Feature, *Journal of Statistical Planning and Inference*, **137**, 589-593.

Coutts, E. and Jann, B. (2011), Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT), *Sociological Methods & Research*, **40**,169-193.

Eichhorn, B. H. and Hayer, L. S. (1983), Scrambled Randomized Response Methods for Obtaining Sensitive Quantitative Data, *Journal of the American Statistical Association*, **66**, 243-250.

Gjestvang, C. R., and Singh, S. (2006), A New Randomized Response Model, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **68** , 523-530.

Greenberg, B. G., Abul-Ela, A. L. A., Simmons, W. R. and Horvitz, D. G. (1969), The Unrelated Question Randomized Response Model: Theoretical framework, *Journal of the American Statistical Association*, **64**, 520-539.

Gupta, S., Mehta, S., Shabbir, J. and Dass, B. K. (2013), Generalized Scrambling in Quantitative Optional Randomized Response Models, *Communications in Statistics-Theory and Methods*, **42**, 4034-4042.

Guerriero, M., Sandri M.F. (2007), A Note on the Comparison of Some Randomized Response Procedures, *Journal of Statistical Planning and Inference*, **137**, 2184–2190.

Hussain, Z., Ali Shah, E. and Shabir, J. (2012), An Alternative Item Count Technique in Sensitive Surveys, *Revista Colombiana de Estadistica*, **35**, 39-54.

Imai, K. (2011), Multivariate Regression Analysis for the Item Count Technique, *Journal of the American Statistical Association*, **106**, 407-416.

Kim, J. M. and Warde, W. D. (2004), A Stratified Warner's Randomized Response Model, *Journal of Statistical Planning and Inference*, **120**, 155-165.

Lanke, J. (1976), On the Degree of Protection in Randomized Interviews, *International Statistical Review*, **44**, 197–203.

Leysieffer, F. W. and Warner, S. L. (1976), Respondent Jeopardy and Optimal Designs in Randomized Response models, *Journal of the American Statistical Association*, **71**, 649–656.

Mangat, N. S. (1994), An Improved Randomized Response Strategy, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **56**, 93-95.

Mangat, N. S. and Singh, R. (1990), An Alternative Randomized Response Procedure, *Biometrika*,**77**, 439-442.

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

Miller, J. D. (1984), A New Survey Technique for Studying Deviant Behavior, *University Microfilms*, 1984 - 1989.

Smith, L. L., Federer, W. T., and Raghavarao, D. (1975). A Comparison of Three Techniques for Eliciting Truthful Answers to Sensitive Questions, *Proceedings of the Social Statistics Section, American Statistical Association*, 447-452.

Tian, G. L., Tang, M. L., Wu, Q. and Liu, Y. (2015), Poisson and Negative Binomial Item Count Techniques for Surveys with Sensitive Questions, *Statistical Methods in Medical Research*, in press.

Tsuchiya, T., Hirai, Y. and Ono, S. (2007), A Study of the Properties of the Item Count Technique, *Public Opinion Quarterly*, **71**, 253–272.

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