Journal of Statistical Sciences, Spring and Summer, 2024
Vol. 18, No. 1, pp 73-89
DOI: 10.61186/jss.18.1.05

# Estimating the Means of Two Sensitive Variables with a New Quantitative Randomized Response Method

Salemian, H.[1], Mahmoudi, E.[1], Alavi, S. M. R.[2]
[1]Department Statistics, Yazd University, Iran.
[2]Department Statistics, Shahid Chamran University, Ahvaz, Iran.

**Corresponding author:** .E. Mahmoudi, emahmoudi@yazd.ac.ir

**Received:** 13/8/2023 **Revised:** 30/1/2024 **Accepted and Published Online:** 3/2/2024.

## Introduction

In sample surveys, some questions are daunting for the respondent. These types of questions are known as sensitive questions. In most cases, people give unrealistic answers or refuse to answer questions that are asked directly. In such cases, indirect methods, such as the randomized response method, are used instead of direct ones. The randomized response method allows the possibility of preserving the respondent's confidentiality and obtaining a more reliable estimate.

## Material and Methods

A new quantitative randomized response method is introduced using a three-state random experiment. Suppose $Y$ is a sensitive variable with mean and standard deviation $\mu_y$ and $\sigma_y$ respectively, and S is an independent variable of $Y$ with known distribution. It is assumed that the range of both variables is the interval $(a, b)$. The randomized response method asks each person to do a three-state random experiment first. If $W = 1$ occurs, the response is $S$; if $W = 2$, the sensitive response is $Y$; otherwise, the sensitive response is $b - Y$ in the answer sheet. It is reported.

## Results and Discussion

In this article, a new quantitative randomized response method is introduced, and by conducting a series of simulation studies, we show that the proposed method is preferable to the collective method of Warner and the multiplicative method of Eichharn and Heyer. Then, we check the mean estimate of

two sensitive variables using the proposed randomized response method, and by using unbiased predictors, we estimate the covariance between two sensitive variables. In an experimental study using the proposed method, the mean number of cheating and the mean daily cigarette consumption of the Shahid Chamran University of Ahvaz students were estimated along with their variance. An estimate for the covariance between them is provided.

**Conclusion**
In this article, a new quantitative randomized response method was presented to estimate the mean of the offer's sensitive variable and its variance. Then, we compared the superiority of the proposed method with Warner's collective randomized response method and Eichhorn and Heyer's multiplicative randomized response method using simulation (with $R$ software) and showed that the proposed method is better than Warner's collective method and Eichhorn and Heyer's is preferable. Then, we used the randomized response method for two sensitive variables and provided an estimate for the covariance between two sensitive variables using unbiased predictors and, in an experimental study, an unbiased estimate for the mean number of frauds and an unbiased estimate for the mean amount of daily smoking students of the Shahid Chamran University of Ahvaz were obtained based on the new proposed randomized response method.

**Keywords:** Randomized esponse method, Sensitive variable, Unbiased predictor, Covariance estimate, The mean of the cheating number, The mean of the number of smoking cigarette.

**Mathematics Subject Classification (2010):** 62D05, 62F10.

©The Author(s). The Publisher is Iranian Statistical Society.
This is an open access article distributed under the terms and conditions of (CC BY-NC 4.0)

**مجله علوم آماری، بهار و تابستان ۱۴۰۳**
**جلد ۱۸، شماره ۱، ص ۷۳ -- ۸۹**
DOI: 10.61186/jss.18.1.5
**مقاله پژوهشی**

# برآورد میانگین دو متغیر حساس با یک روش پاسخ تصادفیده کمی جدید

حامد سالمیان[۱]، عیسی محمودی[۱]، سیدمحمدرضا علوی[۲]

[۱]گروه آمار،دانشکده علوم ریاضی،دانشگاه یزد
[۲]گروه آمار،دانشکده علوم ریاضی و کامپیوتر،دانشگاه شهید چمران اهواز

**نویسنده مسئول :** عیسی محمودی، emahmoudi@yazd.ac.ir

تاریخ دریافت: ۱۴۰۲/۵/۲۲ تاریخ بازنگری: ۱۴۰۲/۱۱/۱۰ تاریخ پذیرش و انتشار: ۱۴۰۲/۱۱/۱۴

**چکیده:** اغلب در بررسی‌های نمونه‌ای پاسخ دهندگان از پاسخ دادن به برخی سؤالات با ماهیت حساس، خودداری می‌کنند. روش‌های پاسخ تصادفیده برای فاش نشدن راز پاسخ دهنده طراحی شده‌اند. در این مقاله یک روش پاسخ تصادفیده کمی جدید معرفی شده و با انجام یک سری مطالعات شبیه‌سازی نشان می‌دهیم روش پیشنهادی نسبت به روش‌های جمعی و ضربی ارجحیت دارد. سپس برآوردگر میانگین دو متغیر حساس را با استفاده از روش پاسخ تصادفیده پیشنهادی معرفی کرده و با به‌کار بردن پیشگوهای نااریب، برآوردگری برای کوواریانس بین دو متغیرحساس ارائه می‌دهیم. در یک مطالعه تجربی با استفاده از روش پیشنهادی، میانگین تعداد تقلب و میانگین میزان مصرف روزانه سیگار دانشجویان دانشگاه شهید چمران اهواز همراه با واریانس آن‌ها برآورد شده و برآوردی برای کوواریانس بین آن‌ها ارائه می‌شود.

**واژه‌های کلیدی:** روش پاسخ تصادفیده، متغیر حساس، پیشگوی نااریب، برآورد کوواریانس، میانگین تعداد تقلب، میانگین میزان مصرف سیگار.
**کد موضوع‌بندی ریاضی (۲۰۱۰):** 62F10 ، 62D05.

## ۱ مقدمه

در بررسی‌های نمونه‌ای برخی از سؤالات برای پاسخ دهنده دلهره‌آور است. این نوع سؤالات به سؤالات حساس معروف‌اند. برخی از سؤالات حساس عبارتند از مصرف سیگار، عدم رعایت قوانین رانندگی، تقلب در امتحانات،

©نویسند(گان). ناشر انجمن آمار ایران است.
این مقاله با دسترسی آزاد تحت شرایط و ضوابط (CC BY-NC 4.0) توزیع شده است.

میزان درآمد افراد، سقط جنین، پرداخت مالیات، اختلاس، قمار بازی و غیره. در اغلب موارد افراد به سؤالاتی که به طور مستقیم از آن‌ها پرسیده می‌شود، پاسخ‌هایی غیر واقعی می‌دهند و یا از پاسخ دادن امتناع می‌ورزند. در این گونه موارد به جای استفاده از روش‌های مستقیم از روش‌های غیر مستقیم، مانند روش پاسخ تصادفیده استفاده می‌شود. روش پاسخ تصادفیده این امکان را فراهم می‌آورد که راز پاسخ دهنده حفظ شود و برآوردی با قابلیت اعتماد بیشتر حاصل شود. هدف اصلی از این روش‌ها دعوت به همکاری و تشویق پاسخ دهنده به دادن جواب صادقانه به سؤال‌هاست که این کار با توجه به محرمانگی جواب‌ها و حفاظت از حریم شخصی فرد صورت می‌گیرد.

روش پاسخ تصادفیده نخستین بار توسط وانر (۱۹۶۵) به منظور حفظ محرمانگی و کاهش اریبی در پاسخ‌ها معرفی شد. مدل‌های پاسخ تصادفیده را می‌توان به دو دسته‌ی اصلی ۱–مدل‌های پاسخ تصادفیده کیفی[1] و ۲–مدل‌های پاسخ تصادفیده کمی[2] طبقه‌بندی کرد. پژوهش های زیادی در مورد مدل‌های پاسخ تصادفیده کمی و کیفی انجام شده است. از جمله گرین‌برگ و همکاران (۱۹۶۹) که به جای سؤال مکمل در روش وارنر از سؤال نامرتبط غیر حساس استفاده کردند. گاهی به این روش، روش سیمونس گفته می‌شود. گرین‌برگ و همکاران (۱۹۷۱) کاربرد روش پاسخ تصادفیده را در به‌دست آوردن میانگین داده‌های کمی حساس [3] مطالعه کردند. وارنر (۱۹۷۱) روش پاسخ تصادفیده جمعی را برای متغیر حساس کمی معرفی کرد. مدل پاسخ تصادفیده ضربی برای متغیرهای حساس کمی توسط ایچهارن و هایر (۱۹۸۳) معرفی شد. کیم و وارده (۲۰۰۴) یک مدل پاسخ تصادفیده با طبقه‌بندی به‌کار بردند. جست‌وانگ و سینگ (۲۰۰۷) یک مدل پاسخ تصادفیده سه مرحله‌ای پیشنهاد کردند. علوی (۱۳۸۶) روش پاسخ تصادفیده چندگزینه‌ای را پیشنهاد کرد. مهتا و همکاران (۲۰۱۲) تحت یک محدودیت گونه‌ای از مدل پاسخ تصادفیده اختیاری سه مرحله‌ای جمعی را معرفی کردند. علوی و یزاری (۱۳۹۳) یک مدل پاسخ تصادفیده کمی اختیاری سه مرحله‌ای را ارائه کردند. علوی و تاج‌الدینی (۱۳۹۴) و علوی و تاج‌الدینی (۲۰۱۶) روش پاسخ تصادفیده مکرر را معرفی و با استفاده از آن نسبت تقلب دانشجویان را برآورد کردند. علوی و همکاران (۱۳۹۸) مدل تعداد اقلام تصادفیده را معرفی کردند و علوی و همکاران (۱۳۹۸) یک مدل پاسخ تصادفیده مرکب جدید را پیشنهاد دادند. در این مقاله یک روش پاسخ تصادفیده کمی جدید برای برآورد میانگین متغیر حساس پیشنهاد و برآورد واریانس آن ارائه می‌گردد. با استفاده از معیار یان و همکاران (۲۰۰۸) سطح حفاظت از محرمانگی روش محاسبه می‌گردد و سپس با استفاده از معیار برتری $\delta$ که از نسبت واریانس به سطح محرمانگی به‌دست می‌آید، به مقایسه برتری روش پیشنهادی با روش پاسخ تصادفیده جمعی وارنر (۱۹۷۱) و روش پاسخ تصادفیده ضربی ایچهارن و هایر (۱۹۸۳) با استفاده از شبیه‌سازی (با نرم‌افزار $R$) می‌پردازیم و نشان می‌دهیم که روش پیشنهادی نسبت به روش جمعی وارنر و روش ضربی ایچهارن و هایر ارجحیت دارد. سپس روش پاسخ تصادفیده فوق را برای دو متغیر حساس به کار برده و برآوردی برای کوواریانس بین دو متغیر حساس با استفاده از پیشگوهای نااریب ارائه می‌کنیم. این موارد از جمله دستاوردهای جدید نسبت به تحقیقات پیشین است.

[1] Qualitative Randomized Response Models
[2] Quantitative Randomized Response Models
[3] Quantitive sensitive data

## ۲ روش پاسخ تصادفیده پیشنهادی

در این بخش برای حفاظت از محرمانگی سؤالات حساس کمی، با استفاده از یک آزمایش تصادفی سه حالتی یک روش پاسخ تصادفیده کمی جدید معرفی می‌شود. فرض کنید $Y$ یک متغیر حساس با میانگین و انحراف معیار به ترتیب $\mu_y$ و $\sigma_y$ بوده و $S$ یک متغیر مستقل از $Y$ با توزیع معلوم باشد. فرض می‌شود دامنه هر دو متغیر بازه $(a, b)$ باشد. روش پاسخ تصادفیده به اینصورت است که از هر فرد تقاضا می‌شود که در ابتدا یک آزمایش تصادفی سه حالتی انجام دهد اگر $W = ۱$ رخ دهد، پاسخ $S$ و اگر $W = ۲$ رخ دهد، پاسخ حساس $Y$ و در غیر اینصورت پاسخ حساس $b - Y$ در پاسخنامه گزارش می‌شود. بنابراین متغیر پاسخ تصادفیده

$$Z = \begin{cases} S & \text{با احتمال } p_۱ \\ Y & \text{با احتمال } p_۲ \\ b - Y & \text{با احتمال } p_۳ \end{cases}$$

تعریف می‌شود و امید ریاضی آن تحت ترفند تصادفیده برابر $E_R(Z) = p_۱\mu_s + p_۳b + (p_۲ - p_۳)Y$, است. بنابراین امید ریاضی متغیر پاسخ تصادفیده $Z$ بنا به قاعده امید ریاضی مضاعف برابر

$$E(Z) = E_P[E_R(Z)] = E_P\big(p_۱\mu_s + p_۳b + (p_۲ - p_۳)Y\big) = p_۱\mu_s + p_۳b + (p_۲ - p_۳)\mu_y,$$

است. اندیس $P$ بیانگر امید ریاضی تحت طرح نمونه‌گیری است. برای محاسبه واریانس $Z$ با استفاده از قاعده واریانس مضاعف داریم $V(Z) = V_P(E_R(Z)) + E_P(V_R(Z))$، که در آن

$$V_P(E_R(Z)) = V_P\big(p_۱\mu_s + p_۳b + (p_۲ - p_۳)Y\big) = (p_۲ - p_۳)^۲\sigma_y^۲.$$

برای محاسبه $E_P(V_R(Z))$ داریم

$$\begin{aligned} V_R(Z) &= p_۱E(S^۲) + p_۲Y^۲ + p_۳(b - Y)^۲ - \big(p_۱\mu_s + p_۳b + (p_۲ - p_۳)Y\big)^۲ \\ &= p_۱(\sigma_s^۲ + \mu_s^۲) + p_۲Y^۲ + p_۳(b^۲ + Y^۲ - ۲bY) - \big(p_۱\mu_s + p_۳b\big)^۲ \\ &- (p_۲ - p_۳)^۲Y^۲ - ۲\big(p_۱\mu_s + p_۳b\big)(p_۲ - p_۳)Y \\ &= [p_۲ + p_۳ - (p_۲ - p_۳)^۲]Y^۲ - ۲[p_۳b + \big(p_۱\mu_s + p_۳b\big)(p_۲ - p_۳)]Y \\ &+ p_۱\big(\sigma_s^۲ + \mu_s^۲\big) + p_۳b^۲ - \big(p_۱\mu_s + p_۳b\big)^۲. \end{aligned}$$

لذا واریانس متغیر پاسخ تصادفیده $Z$ به صورت

$$\begin{aligned} V(Z) &= V_P(E_R(Z)) + E_P(V_R(Z)) \\ &= (p_۲ - p_۳)^۲\sigma_y^۲ + \big(p_۲ + p_۳ - (p_۲ - p_۳)^۲\big)\big(\sigma_y^۲ + \mu_y^۲\big) \\ &- ۲[p_۳ b + \big(p_۱\mu_s + p_۳ b\big)(p_۲ - p_۳)]\mu_y + p_۱\big(\sigma_s^۲ + \mu_s^۲\big) \\ &+ p_۳ b^۲ - \big(p_۱\mu_s + p_۳ b\big)^۲. \end{aligned} \tag{۱}$$

است. یک برآوردگر برای میانگین متغیر حساس کمی $Y$، براساس یک نمونه تصادفی $Z_۱, \cdots, Z_n$ به حجم $n$ از پاسخ‌های تصادفیده تحت روش پیشنهادی

$$\hat{\mu}_y = \frac{\bar{Z} - p_۱\mu_s - p_۳ b}{p_۲ - p_۳}, \tag{۲}$$

معرفی می‌شود که در آن $p_۲ \neq p_۳$ و $\bar{Z}$ میانگین نمونه پاسخ‌های تصادفیده است. واریانس برآوردگر میانگین متغیر حساس برابر است با

$$V(\hat{\mu}_y) = \frac{V(\bar{Z})}{(p_۲ - p_۳)^۲} = \frac{V(Z)}{n(p_۲ - p_۳)^۲}. \tag{۳}$$

با جایگذاری (۱) در (۳) واریانس برآوردگر میانگین متغیر حساس به‌دست می‌آید. برآوردگر نااریب برای واریانس $\hat{\mu}_y$

$$\hat{V}(\hat{\mu}_y) = \frac{S_z^۲}{n(p_۲ - p_۳)^۲}, \tag{۴}$$

تعیین می‌شود، که در آن $S_z^۲ = \frac{\sum_{i=۱}^n (z_i - \bar{z})^۲}{n - ۱}$ واریانس نمونه‌ای پاسخ‌های تصادفیده است.

## ۳ معیار ارزیابی و مقایسه روش پاسخ تصادفیده پیشنهادی با سایر روش‌ها

روش‌های پاسخ تصادفیده به منظور حفظ محرمانگی طراحی شده‌اند. برای اینکه نشان داده شود که هر مدل چقدر محرمانگی را پوشش می‌دهد نیاز به معرفی یک معیار برای محرمانگی است. یان و همکاران (۲۰۰۸) برای مدل‌های پاسخ تصادفیده کمی یک معیار محرمانگی به صورت $P_\ell = E(Z - Y)^۲$ تعریف کردند که با استفاده از آن سطح محرمانگی مدل پاسخ تصادفیده پیشنهادی محاسبه

می‌شود. برای محاسبه معیار سطح محرمانگی در روش پیشنهادی داریم:

$$\begin{aligned} E_R[Z-Y]^۲ &= E_R(Z^۲) + Y^۲ - ۲YE_R(Z) \\ &= p_۱E(S^۲) + p_۲Y^۲ + p_۳(b-Y)^۲ + Y^۲ - ۲Y[p_۱\mu_s + p_۳b + (p_۲-p_۳)Y] \\ &= \big(-p_۲ + ۳p_۳ + ۱\big)Y^۲ - ۲p_۱\mu_sY - ۴Yp_۳b + p_۳b^۲ + p_۱\big(\sigma_s^۲ + \mu_s^۲\big). \end{aligned}$$

بنابراین

$$\begin{aligned} P_\ell &= E_PE_R(Z-Y)^۲ \\ &= \big(-p_۲ + ۳p_۳ + ۱\big)\big(\sigma_y^۲ + \mu_y^۲\big) - ۲p_۱\mu_s\mu_y - ۴\mu_yp_۳b + p_۳b^۲ + p_۱\big(\sigma_s^۲ + \mu_s^۲\big). \end{aligned} \tag{۵}$$

هر چه معیار سطح محرمانگی بیشتر باشد محرمانگی مدل نیز بیشتر است. برای مقایسه مدل پاسخ تصادفیده پیشنهادی از معیار برتری

$$\delta = \frac{V(\hat{\mu}_y)}{P_\ell}. \tag{۶}$$

ارائه شده توسط گوپتا و همکاران (۲۰۱۷) استفاده می‌شود. هرچه واریانس مدل کم (کارایی بالا) و محرمانگی بیشتر باشد مدل بهتر است. لذا برای مقایسه برتری دو مدل پاسخ تصادفیده برای برآورد میانگین متغیر حساس آن مدلی مناسب‌تر است که $\delta$ کمتری داشته باشد. برای محاسبه معیار برتری مدل پاسخ تصادفیده پیشنهادی با جایگذاری (۳) در (۶) داریم

$$\delta_p = \frac{V(Z)}{n(p_۲-p_۳)^۲P_\ell} \tag{۷}$$

با جایگذاری (۱) و (۵) در (۷)، $\delta_p$ محاسبه می‌شود. معیار برتری مدل پاسخ تصادفیده وارنر (۱۹۷۱) به صورت

$$\delta_w = \frac{V(\hat{\mu}_w)}{P_{lw}} = \frac{(\sigma_y^۲ + \sigma_s^۲)/n}{E(Z-Y)^۲} = \frac{\sigma_y^۲ + \sigma_s^۲}{n(\sigma_s^۲ + \mu_s^۲)},$$

و مدل پاسخ تصادفیده ایچهارن و هایر (۱۹۸۳) به صورت

$$\delta_E = \frac{V(\hat{\mu}_E)}{P_{lE}} = \frac{\sigma_s^2(\sigma_y^2+\mu_y^2)+\mu_y^2\mu_s^2/n}{E(Z-Y)^2}$$
$$= \frac{\sigma_s^2(\sigma_y^2+\mu_y^2)+\mu_y^2\mu_s^2/n}{(\sigma_y^2)^2\sigma_s^2\mu_y^2+2\mu_y^2\mu_s^2\sigma_y^2\sigma_s^2+\sigma_y^2\mu_y^2(\mu_s^2)^2+(\mu_s^2)^2(\mu_y^2)^2+(\mu_y^2)^2\mu_s^2+(\sigma_y^2+\mu_y^2)-2\mu_s(\sigma_y^2+\mu_y^2)}.$$

بدست می‌آیند. همانگونه که مشاهده می‌شود به دلیل پیچیدگی معیار برتری $\delta$ در مدل پاسخ تصادفیده پیشنهادی، برای مقایسه برتری مدل پاسخ تصادفیده پیشنهادی و مدل پاسخ تصادفیده جمعی وارنر و مدل پاسخ تصادفیده ضربی ایچهارن و هایر از شبیه‌سازی استفاده می‌کنیم.

## ۴ مطالعه شبیه‌سازی

در این بخش مطالعه شبیه‌سازی برای ارزیابی و مقایسه روش پاسخ تصادفیده پیشنهادی و روش های پاسخ تصادفیده جمعی و ضربی با استفاده از معیار برتری ($\delta$) ارائه می شود. به دلیل پیچیدگی واریانس برآوردکننده میانگین متغیر حساس یک مطالعه شبیه‌سازی برای ارزیابی واریانس و معیار حفاظت از سطح محرمانگی نیز با استفاده از نرم‌افزار $R$ انجام گرفته شده است. نتایج شبیه‌سازی روش پیشنهادی برای مقادیر مختلف $n$، $\mu_y$ و $\sigma_y$ در جدول ۱ قابل مشاهده است. با توجه به جداول ۱ تا ۴ نتایج زیر از مطالعه از شبیه‌سازی به‌دست می‌آیند.

۱. برآورد میانگین در تمام سطوح برای مدل پاسخ تصادفیده پیشنهادی و مدل جمعی و مدل ضربی به مقدار واقعی نزدیک است و مشاهده می‌شود که برآوردها نااریب هستند.

۲. برآورد واریانس به ازای هر مقدار از $n$ با افزایش $\mu$ (وقتی که $\sigma$ ثابت) و افزایش $\sigma$ (وقتی که $\mu$ ثابت) افزایش می‌یابد اما با توجه به جداول در کل با افزایش حجم نمونه و ثابت نگه داشتن $\sigma$ و $\mu$ شاهد کاهش واریانس و در نتیجه افزایش کارایی خواهیم بود.

۳. در جداول ۱، ۲ و ۳ با افزایش $n$ مقدار واریانس کاهش می‌یابد، با افزایش $\sigma$ و $\mu$ سطح محرمانگی افزایش می‌یابد و در نتیجه $\delta$ کاهش می‌یابد.

۴. با مراجعه به جدول ۴ می‌توان مشاهده کرد که به ازای تمام مقادیر $\mu$، $\sigma$ و $n$ مدل پاسخ تصادفیده پیشنهادی بر مدل پاسخ تصادفیده جمعی وارنر و مدل پاسخ تصادفیده ضربی ایچهارن و هایر برتری دارد.

## ۵ تعمیم روش پاسخ تصادفیده پیشنهادی برای دو متغیر حساس

فرض کنید $Y$، بیانگر پاسخ سؤال حساس اول (به طور مثال تعداد تقلب در امتحانات دانشگاه) و $X$، پاسخ سؤال حساس دوم (به طور مثال میزان مصرف روزانه سیگار) باشند و فرض کنید $S_1$ و $S_2$ دو متغیر تصادفی مستقل

**جدول ۱.** میانگین، قدر مطلق اریبی، واریانس، سطح محرمانگی و معیار برتری روش پیشنهادی

| $\delta_p$ | $P_\ell$ | $V(\hat{\mu}_y)$ | $\lvert Bias\rvert$ | $E(\hat{\mu}_y)$ | $\sigma_y$ | $\mu_y$ | $n$ |
|---|---|---|---|---|---|---|---|
| ۰٫۰۵۴۳ | ۳۱۶٫۱۲۷ | ۱۷٫۱۷۳۱ | ۰٫۰۸۱۱ | ۵٫۰۸۱۱ | ۵ | | |
| ۰٫۰۶۱۵ | ۹۹۱٫۴۸۰۱ | ۶۱٫۰۶۱۷۶ | ۰٫۰۲۵۴ | ۴٫۹۷۴۶ | ۱۰ | ۵ | |
| ۰٫۰۷۵۱ | ۱۷۸۲٫۰۶۱ | ۱۳۳٫۸۷۶۶ | ۰٫۰۳۲۷ | ۵٫۰۳۲۷ | ۱۵ | | |
| ۰٫۰۲۶۷ | ۱۱۷۰٫۲۸۴ | ۳۱٫۲۸۷۵۱ | ۰٫۰۲۸۷ | ۹٫۹۷۱۳ | ۵ | | |
| ۰٫۰۳۲۴۳ | ۲۳۸۶٫۰۰۱ | ۷۷٫۳۹۰۰ | ۰٫۰۳۹۵ | ۱۰٫۰۳۹۵ | ۱۰ | ۱۰ | $n = ۲۰$ |
| ۰٫۰۶۶۷ | ۳۸۹۵٫۰۴۵ | ۲۵۹٫۹۴۷۶ | ۰٫۰۳۷۱ | ۹٫۹۶۲۹ | ۱۵ | | |
| ۰٫۰۴۲۱ | ۳۳۴۹٫۲۳۲ | ۱۴۱٫۱۳۳۶ | ۰٫۱۳۵۲ | ۱۹٫۸۶۴۸ | ۵ | | |
| ۰٫۰۴۱۵ | ۴۴۶۹٫۵۱۴ | ۱۸۵٫۷۳۲۸ | ۰٫۱۹۳۳ | ۱۹٫۸۰۶۷ | ۱۰ | ۲۰ | |
| ۰٫۰۲۸۸ | ۹۳۱۷٫۱۸۷ | ۲۶۹٫۰۸۳ | ۰٫۰۷۰۱ | ۱۹٫۹۲۹۹ | ۱۵ | | |
| ۰٫۰۰۶۹ | ۴۸۲٫۸۵۸۲ | ۳٫۳۶۵۰ | ۰٫۰۶۶۷ | ۵٫۰۶۶۷ | ۵ | | |
| ۰٫۰۰۷۶ | ۱۶۳۸٫۴۲۶ | ۱۲٫۴۸۲۹ | ۰٫۰۱۰۲ | ۵٫۰۱۰۲ | ۱۰ | ۵ | |
| ۰٫۰۰۷۴ | ۳۶۹۹٫۸۵ | ۲۷٫۶۴۴۷ | ۰٫۰۰۲۱ | ۵٫۰۰۲۱ | ۱۵ | | |
| ۰٫۰۰۶۹ | ۹۰۸٫۲۰۰۵ | ۶٫۳۲۲۶ | ۰٫۰۱۸۸ | ۱۰٫۰۱۸۸ | ۵ | | |
| ۰٫۰۰۷۸ | ۱۹۶۵٫۸۹۷ | ۱۵٫۴۲۹۶ | ۰٫۰۱۸۹ | ۱۰٫۰۱۸۹ | ۱۰ | ۱۰ | $n = ۱۰۰$ |
| ۰٫۰۰۵۹ | ۵۱۰۷٫۶۱۸ | ۳۰٫۴۰۳۸ | ۰٫۰۲۸۰ | ۱۰٫۰۲۸۰ | ۱۵ | | |
| ۰٫۰۰۸۵ | ۳۳۱۶٫۷۹۲ | ۲۸٫۳۳۴۰ | ۰٫۰۱۵۹ | ۲۰٫۰۱۵۹ | ۵ | | |
| ۰٫۰۰۷۳ | ۵۱۳۲٫۴۸۸ | ۳۷٫۸۸۸۱ | ۰٫۰۳۵۳ | ۲۰٫۰۳۵۳ | ۱۰ | ۲۰ | |
| ۰٫۰۰۷۱ | ۷۵۷۰٫۶۲۸ | ۵۴٫۳۱۵۶ | ۰٫۰۳۰۱ | ۲۰٫۰۳۰۱ | ۱۵ | | |
| ۰٫۰۰۱۰ | ۶۴۶٫۳۱۰۱ | ۰٫۶۸۰۳ | ۰٫۰۰۳۲ | ۵٫۰۰۳۲ | ۵ | | |
| ۰٫۰۰۱۴ | ۱۷۳۰٫۸۵۷ | ۲٫۴۶۸۳ | ۰٫۰۰۸۰ | ۵٫۰۰۸۰ | ۱۰ | ۵ | |
| ۰٫۰۰۱۱ | ۴۹۱۸٫۶۲۲ | ۵٫۴۶۰۹ | ۰٫۰۰۰۷ | ۴٫۹۹۹۳ | ۱۵ | | |
| ۰٫۰۰۱۱ | ۱۱۳۳٫۶۵۵ | ۱٫۲۶۹۹ | ۰٫۰۰۹۲ | ۹٫۹۹۰۸ | ۵ | | |
| ۰٫۰۰۰۹ | ۳۳۶۶٫۰۸۳ | ۳٫۰۵۷۵ | ۰٫۰۰۶۸ | ۱۰٫۰۰۶۸ | ۱۰ | ۱۰ | $n = ۵۰۰$ |
| ۰٫۰۰۱۹ | ۳۲۳۴٫۶۸۸ | ۶٫۲۹۶۶ | ۰٫۰۰۱۴ | ۹٫۹۹۸۶ | ۱۵ | | |
| ۰٫۰۰۱۲ | ۴۳۷۲٫۲۸ | ۵٫۶۷۵۶ | ۰٫۰۰۵ | ۱۹٫۹۹۵۰ | ۵ | | |
| ۰٫۰۰۱۲ | ۵۹۲۷٫۹۳۹ | ۷٫۶۶۵۴ | ۰٫۰۱۱۹ | ۱۹٫۹۸۸۱ | ۱۰ | ۲۰ | |
| ۰٫۰۰۱۸ | ۵۴۱۹٫۷۲۶ | ۱۰٫۲۷۴۶ | ۰٫۰۱۴۳ | ۱۹٫۹۸۵۷ | ۱۵ | | |

با توزیع معلوم باشند که هر دو از $Y$ و $X$ مستقل هستند. اگر پاسخ‌های تصادفیده برای $Y$ و $X$ را به ترتیب با نمادهای $Z$ و $T$ نمایش دهیم داریم:

$$Z = \begin{cases} S_۱ & \text{با احتمال } p_۱ \\ Y & \text{با احتمال } p_۲ \\ b_۱ - Y & \text{با احتمال } p_۳ \end{cases} \qquad T = \begin{cases} S_۲ & \text{با احتمال } q_۱ \\ X & \text{با احتمال } q_۲ \\ b_۲ - X & \text{با احتمال } q_۳ \end{cases}$$

که در آن $p_۲ \neq p_۳$ و $q_۲ \neq q_۳$. اگر $Y_i$ و $X_i$ و $Z_i$ و $T_i$ به ترتیب پاسخ‌های واقعی متغیر حساس اول و دوم و پاسخ‌های تصادفیده آن‌ها طبق روش پیشنهادی باشند با توجه به رابطه (۲) پیشگوهای نااریب $X_i$ و $Y_i$ را تحت

**جدول ۲.** میانگین، قدر مطلق اریبی، واریانس، سطح محرمانگی و معیار برتری روش جمعی

| $\delta_w$ | $P_{lw}$ | $V(\hat{\mu}_{yw})$ | $\|Bias\|$ | $E(\hat{\mu}_{yw})$ | $\sigma_y$ | $\mu_y$ | $n$ |
|---|---|---|---|---|---|---|---|
| ۱٫۲۷۱۴ | ۰٫۹۹۳۶ | ۱٫۲۶۳۳ | ۰٫۰۰۴۶ | ۵٫۰۰۴۶ | ۵ | | |
| ۵٫۱۴۰۳ | ۰٫۹۹۵۹ | ۵٫۱۱۹۳ | ۰٫۰۱۷۱ | ۵٫۰۱۷۱ | ۱۰ | ۵ | |
| ۱۱٫۱۸۶۵ | ۱٫۰۰۳۹ | ۱۱٫۲۳۰۲ | ۰٫۰۰۴۶ | ۴٫۹۹۵۳ | ۱۵ | | |
| ۱٫۳۳۷۷ | ۰٫۹۷۱۲ | ۱٫۲۹۹۲ | ۰٫۰۰۸۷ | ۱۰٫۰۰۸۷ | ۵ | | |
| ۵٫۳۰۲۰ | ۰٫۹۸۷۱ | ۵٫۲۳۳۷ | ۰٫۰۲۹۹ | ۱۰٫۰۲۹۹ | ۱۰ | ۱۰ | $n = ۲۰$ |
| ۱۱٫۳۸۲۶ | ۰٫۹۹۹۹ | ۱۱٫۳۸۱۵ | ۰٫۰۱۱۳ | ۱۰٫۰۱۱۳ | ۱۵ | | |
| ۱٫۳۴۴۶ | ۰٫۹۷۱۱ | ۱٫۳۰۵۸ | ۰٫۰۲۸۵ | ۱۹٫۹۷۱۵ | ۵ | | |
| ۵٫۲۱۲۰ | ۰٫۹۶۲۸ | ۵٫۰۱۸۲ | ۰٫۰۱۶۲ | ۲۰٫۰۱۶۲ | ۱۰ | ۲۰ | |
| ۱۱٫۳۸۴۲ | ۰٫۹۹۲۸ | ۱۱٫۳۰۲۳ | ۰٫۰۱۶۱ | ۲۰٫۰۱۶۱ | ۱۵ | | |
| ۰٫۲۶۶۶ | ۰٫۹۸۰۳ | ۰٫۲۶۱۴ | ۰٫۰۰۲۴ | ۴٫۹۹۷۶ | ۵ | | |
| ۱٫۰۴۱۱ | ۰٫۹۷۰۸ | ۱٫۰۱۰۷ | ۰٫۰۰۷۳ | ۵٫۰۰۷۳ | ۱۰ | ۵ | |
| ۲٫۱۹۹۰ | ۱٫۰۰۲۷ | ۲٫۲۰۵۰ | ۰٫۰۰۱۰ | ۵٫۰۰۱۰ | ۱۵ | | |
| ۰٫۲۶۱۰ | ۰٫۹۸۴۲ | ۰٫۲۵۶۹ | ۰٫۰۰۴۷ | ۹٫۹۹۵۳ | ۵ | | |
| ۱٫۰۴۶۷ | ۰٫۹۸۰۸ | ۱٫۰۲۶۷ | ۰٫۰۱۶۳ | ۱۰٫۰۱۶۳ | ۱۰ | ۱۰ | $n = ۱۰۰$ |
| ۲٫۱۹۹۰ | ۱٫۰۱۵۴ | ۲٫۲۳۲۹ | ۰٫۰۰۹۵ | ۱۰٫۰۰۹۵ | ۱۵ | | |
| ۰٫۲۵۶۹ | ۰٫۹۹۹۷ | ۰٫۲۵۶۹ | ۰٫۰۰۵۳ | ۲۰٫۰۰۵۳ | ۵ | | |
| ۱٫۰۳۱۰ | ۰٫۹۷۵۲ | ۱٫۰۰۵۵ | ۰٫۰۰۸۳ | ۲۰٫۰۰۸۳ | ۱۰ | ۲۰ | |
| ۲٫۲۵۷۴ | ۱٫۰۱۹۳ | ۲٫۳۰۱۰ | ۰٫۰۰۴۸ | ۲۰٫۰۰۴۸ | ۱۵ | | |
| ۰٫۰۵۲۲ | ۰٫۹۸۸۹ | ۰٫۰۵۱۷ | ۰٫۰۰۱ | ۴٫۹۹۹۰ | ۵ | | |
| ۰٫۲۰۷۵ | ۰٫۹۷۲۶ | ۰٫۲۰۱۹ | ۰٫۰۰۱۳ | ۵٫۰۰۱۳ | ۱۰ | ۵ | |
| ۰٫۴۵۱۹ | ۱٫۰۰۸۹ | ۰٫۴۵۶۰ | ۰٫۰۰۰۴ | ۴٫۹۹۹۵ | ۱۵ | | |
| ۰٫۰۵۱۲ | ۰٫۹۸۹۵ | ۰٫۰۵۰۷ | ۰٫۰۰۰۴ | ۱۰٫۰۰۰۴ | ۵ | | |
| ۰٫۲۰۰۹ | ۰٫۹۹۴۳ | ۰٫۱۹۹۸ | ۰٫۰۰۰۸ | ۹٫۹۹۹۲ | ۱۰ | ۱۰ | $n = ۵۰۰$ |
| ۰٫۴۶۰۵ | ۰٫۹۹۴۰ | ۰٫۴۵۷۸ | ۰٫۰۰۱۵ | ۹٫۹۹۸۴ | ۱۵ | | |
| ۰٫۰۵۲۸ | ۰٫۹۷۰۸ | ۰٫۰۵۱۳ | ۰٫۰۰۱۲ | ۱۹٫۹۹۸۸ | ۵ | | |
| ۰٫۲۰۹۸ | ۰٫۹۷۴۴ | ۰٫۲۰۴۵ | ۰٫۰۰۰۷ | ۱۹٫۹۹۹۳ | ۱۰ | ۲۰ | |
| ۰٫۴۴۶۸ | ۱٫۰۱۷۲ | ۰٫۴۵۴۵ | ۰٫۰۰۳۲ | ۲۰٫۰۰۳۲ | ۱۵ | | |

ترفند تصادفیده از روابط زیر به‌دست می‌آوریم:

$$\hat{X}_i = \frac{T_i - q_۱\mu_{s_۲} - q_۳b_۲}{(q_۲ - q_۳)}, \quad \hat{Y}_i = \frac{Z_i - p_۱\mu_{s_۱} - p_۳b_۱}{(p_۲ - p_۳)}.$$

**قضیه ۱.** پیشگوهای $\hat{X}_i$ و $\hat{Y}_i$ به ترتیب برای $X_i$ و $Y_i$نااریب هستند.

**جدول ۳.** میانگین، قدر مطلق اریبی، واریانس، سطح محرمانگی و معیار برتری روش ضربی

| $\delta_E$ | $P_{lE}$ | $V(\hat{\mu}_{yE})$ | $\|Bias\|$ | $E(\hat{\mu}_{yE})$ | $\sigma_y$ | $\mu_y$ | $n$ |
|---|---|---|---|---|---|---|---|
| ۰/۰۸۰۰ | ۴۶/۵۲۱۵ | ۳/۷۲۳۶ | ۰/۰۱۵۰ | ۵/۰۱۵۰ | ۵ | | |
| ۰/۰۹۴۳ | ۱۱۸/۱۰۲ | ۱۱/۱۳۹۵ | ۰/۰۰۳۳ | ۴/۹۹۶۶ | ۱۰ | ۵ | |
| ۰/۰۹۹۸ | ۲۳۵/۵۴۱۷ | ۲۳/۵۲۸۳ | ۰/۰۰۱۳ | ۴/۹۹۸۶ | ۱۵ | | |
| ۰/۰۶۲۸ | ۱۱۸/۹۴۶۳ | ۷/۴۷۲۷ | ۰/۰۰۱۸ | ۱۰/۰۰۱۸ | ۵ | | |
| ۰/۰۷۸۵ | ۱۹۱/۴۰۴۲ | ۵/۲۳۳۷ | ۰/۰۰۱۴ | ۱۰/۰۰۱۴ | ۱۰ | ۱۰ | $n = ۲۰$ |
| ۰/۰۷۹۶ | ۳۵۰/۳۷۵۴ | ۲۷/۹۰۶۷ | ۰/۰۰۱۲ | ۱۰/۰۰۱۲ | ۱۵ | | |
| ۰/۰۵۱۵ | ۴۳۱/۹۹۸ | ۲۲/۲۶۱ | ۰/۰۰۳۰ | ۲۰/۰۰۳۰ | ۵ | | |
| ۰/۰۶۳۷ | ۴۷۹/۶۴۱۱ | ۳۰/۵۸۷۴ | ۰/۰۰۲۹ | ۲۰/۰۰۲۹ | ۱۰ | ۲۰ | |
| ۰/۰۶۹۴ | ۶۲۰/۰۲۷۳ | ۴۳/۰۷۰۷ | ۰/۰۰۰۶ | ۱۹/۹۹۹۴ | ۱۵ | | |
| ۰/۰۱۵۲ | ۴۸/۹۶۲۲ | ۰/۷۴۴۹ | ۰/۰۰۵۷ | ۵/۰۰۵۷ | ۵ | | |
| ۰/۰۱۸۷ | ۱۲۰/۷۶۱۳ | ۲/۲۶۸۵ | ۰/۰۰۵۴ | ۴/۹۹۴۶ | ۱۰ | ۵ | |
| ۰/۰۲۰۰ | ۲۴۰/۳۴۹۷ | ۴/۸۱۲۷ | ۰/۰۰۰۲ | ۴/۹۹۹۸ | ۱۵ | | |
| ۰/۰۱۲۴ | ۱۱۹/۲۴۲۱ | ۱/۴۸۶۸ | ۰/۰۰۸۶ | ۹/۹۹۱۳ | ۵ | | |
| ۰/۰۱۵۴ | ۱۹۱/۱۶۸۹ | ۲/۹۵۷۴ | ۰/۰۰۴۰ | ۱۰/۰۰۴۰ | ۱۰ | ۱۰ | $n = ۱۰۰$ |
| ۰/۰۱۷۳ | ۳۱۷/۵۷۰۱ | ۵/۵۱۰۸ | ۰/۰۰۳۸ | ۱۰/۰۰۳۸ | ۱۵ | | |
| ۰/۰۱۰۶ | ۴۱۸/۷۸۱ | ۴/۴۶۶۵ | ۰/۰۰۹۱ | ۱۹/۹۹۰۸ | ۵ | | |
| ۰/۰۱۲۲ | ۴۹۹/۵۸۲۳ | ۶/۱۰۵۰ | ۰/۰۱۱۹ | ۲۰/۰۱۱۹ | ۱۰ | ۲۰ | |
| ۰/۰۱۴۱ | ۶۲۴/۴۴۲۲ | ۸/۸۴۰۶ | ۰/۰۰۵۳ | ۲۰/۰۰۵۳ | ۱۵ | | |
| ۰/۰۰۳۰ | ۴۸/۷۴۶۷ | ۰/۱۴۷۷ | ۰/۰۰۱۳ | ۵/۰۰۱۳ | ۵ | | |
| ۰/۰۰۳۶ | ۱۲۲/۱۲۱ | ۰/۴۴۸۴ | ۰/۰۰۱۵ | ۵/۰۰۱۵ | ۱۰ | ۵ | |
| ۰/۰۰۳۹ | ۲۴۳/۳۹۳ | ۰/۹۶۸۹ | ۰/۰۰۶۶ | ۵/۰۰۶۶ | ۱۵ | | |
| ۰/۰۰۲۴ | ۱۲۳/۱۷۷ | ۰/۳۰۶۱ | ۰/۰۰۰۰۲ | ۱۰/۰۰۰۰۲ | ۵ | | |
| ۰/۰۰۳۱ | ۱۹۰/۳۴۵۶ | ۰/۶۰۷۲ | ۰/۰۰۳۳ | ۱۰/۰۰۳۳ | ۱۰ | ۱۰ | $n = ۵۰۰$ |
| ۰/۰۰۳۴ | ۳۱۸/۷۶۱۶ | ۱/۱۰۳۵ | ۰/۰۰۴۷ | ۱۰/۰۰۴۷ | ۱۵ | | |
| ۰/۰۰۲۱ | ۴۱۵/۴۱۴۵ | ۰/۹۱۲۸ | ۰/۰۰۲۵ | ۱۹/۹۹۷۴ | ۵ | | |
| ۰/۰۰۲۴ | ۴۹۰/۷۷۶۲ | ۱/۱۹۲۰ | ۰/۰۰۴۸ | ۱۹/۹۹۵۱ | ۱۰ | ۲۰ | |
| ۰/۰۰۲۸ | ۶۰۶/۱۱۵۶ | ۱/۷۱۰۸ | ۰/۰۰۵۳ | ۲۰/۰۰۵۳ | ۱۵ | | |

**برهان:**

$$
\begin{aligned}
E_R(\hat{X}_i) &= \frac{E_R(T_i) - q_۱ E(S_۲) - q_۳ b_۲}{q_۲ - q_۳} \\
&= \frac{q_۱ E(S_۲) + q_۲ X_i + q_۳(b_۲ - X_i) - q_۱ E(S_۲) - q_۳ b_۲}{q_۲ - q_۳} = X_i, \\
E_R(\hat{Y}_i) &= \frac{E_R(Z_i) - p_۱ E(S_۱) - p_۳ b_۱}{p_۲ - p_۳} \\
&= \frac{p_۱ E(S_۱) + p_۲ Y_i + p_۳(b_۱ - Y_i) - p_۱ E(S_۱) - p_۳ b_۱}{p_۲ - p_۳} = Y_i.
\end{aligned}
$$

**جدول ۴.** مقایسه روش پاسخ تصادفیده پیشنهادی و روش های پاسخ تصادفیده جمعی و ضربی

| $\delta_E$ | $\delta_w$ | $\delta_p$ | $\sigma_y$ | $\mu_y$ | $n$ |
|---|---|---|---|---|---|
| ۰/۰۸۰۰ | ۱/۲۷۱۴ | ۰/۰۵۴۳ | ۵ | | |
| ۰/۰۹۴۳ | ۵/۱۴۰۳ | ۰/۰۶۱۵ | ۱۰ | ۵ | |
| ۰/۰۹۹۸ | ۱۱/۱۸۶۵ | ۰/۰۷۵۱ | ۱۵ | | |
| ۰/۰۶۲۸ | ۱/۳۳۷۷ | ۰/۰۲۶۷ | ۵ | | |
| ۰/۰۷۸۵ | ۵/۳۰۲۰ | ۰/۰۳۲۴۳ | ۱۰ | ۱۰ | $n = ۲۰$ |
| ۰/۰۷۹۶ | ۱۱/۳۸۲۶ | ۰/۰۶۶۷ | ۱۵ | | |
| ۰/۰۵۱۵ | ۱/۳۴۴۶ | ۰/۰۴۲۱ | ۵ | | |
| ۰/۰۶۳۷ | ۵/۲۱۲۰ | ۰/۰۴۱۵ | ۱۰ | ۲۰ | |
| ۰/۰۶۹۴ | ۱۱/۳۸۴۲ | ۰/۰۲۸۸ | ۱۵ | | |
| ۰/۰۱۵۲ | ۰/۲۶۶۶ | ۰/۰۰۶۹ | ۵ | | |
| ۰/۰۱۸۷ | ۱/۰۴۱۱ | ۰/۰۰۷۶ | ۱۰ | ۵ | |
| ۰/۰۲۰۰ | ۲/۱۹۹۰ | ۰/۰۰۷۴ | ۱۵ | | |
| ۰/۰۱۲۴ | ۰/۲۶۱۰ | ۰/۰۰۶۹ | ۵ | | |
| ۰/۰۱۵۴ | ۱/۰۴۶۷ | ۰/۰۰۷۸ | ۱۰ | ۱۰ | $n = ۱۰۰$ |
| ۰/۰۱۷۳ | ۲/۱۹۹۰ | ۰/۰۰۵۹ | ۱۵ | | |
| ۰/۰۱۰۶ | ۰/۲۵۶۹ | ۰/۰۰۸۵ | ۵ | | |
| ۰/۰۱۲۲ | ۱/۰۳۱۰ | ۰/۰۰۷۳ | ۱۰ | ۲۰ | |
| ۰/۰۱۴۱ | ۲/۲۵۷۴ | ۰/۰۰۷۱ | ۱۵ | | |
| ۰/۰۰۳۰ | ۰/۰۵۲۲ | ۰/۰۰۱۰ | ۵ | | |
| ۰/۰۰۳۶ | ۰/۲۰۷۵ | ۰/۰۰۱۴ | ۱۰ | ۵ | |
| ۰/۰۰۳۹ | ۰/۴۵۱۹ | ۰/۰۰۱۱ | ۱۵ | | |
| ۰/۰۰۲۴ | ۰/۰۵۱۲ | ۰/۰۰۱۱ | ۵ | | |
| ۰/۰۰۳۱ | ۰/۲۰۰۹ | ۰/۰۰۰۹ | ۱۰ | ۱۰ | $n = ۵۰۰$ |
| ۰/۰۰۳۴ | ۰/۴۶۰۵ | ۰/۰۰۱۹ | ۱۵ | | |
| ۰/۰۰۲۱ | ۰/۰۵۲۸ | ۰/۰۰۱۲ | ۵ | | |
| ۰/۰۰۲۴ | ۰/۲۰۹۸ | ۰/۰۰۱۲ | ۱۰ | ۲۰ | |
| ۰/۰۰۲۸ | ۰/۴۴۶۸ | ۰/۰۰۱۸ | ۱۵ | | |

بنا به قاعده امید ریاضی مکرر داریم:

$$E(\hat{X}_i) = E_P E_R(\hat{X}_i) = E_P(X_i) = \mu_x, \quad E(\hat{Y}_i) = E_P E_R(\hat{Y}_i) = E_P(Y_i) = \mu_y.$$

**قضیه ۲.** یک برآوردگر نااریب برای کوواریانس $\hat{X}$ و $\hat{Y}$ عبارت است از

$$\widehat{\mathrm{Cov}}(\hat{X}, \hat{Y}) = \frac{\sum_{i=۱}^{n} \hat{X}_i \hat{Y}_i - \frac{\sum_{i=۱}^{n} \hat{X}_i \sum_{i=۱}^{n} \hat{Y}_i}{n}}{n - ۱}.$$

**برهان:** چون برای نمونه تصادفی ساده با جایگذاری مشاهده $i$ام مستقل از مشاهده $j$ام است

$$
\begin{aligned}
E(\widehat{\mathrm{Cov}}(\hat{X},\hat{Y})) &= \frac{۱}{n-۱}\big[\sum_{i=۱}^{n} E(\hat{X}_i\hat{Y}_i) - nE(\bar{\hat{X}}\bar{\hat{Y}})\big] \\
&= \frac{۱}{n-۱}\big[\sum_{i=۱}^{n} E(\hat{X}_i\hat{Y}_i) - \frac{۱}{n}E\big[\sum_{i=۱}^{n}\hat{X}_i\hat{Y}_i + \sum_{i\neq j}\sum \hat{X}_i\hat{Y}_j\big]\big] \\
&= \frac{۱}{n-۱}\big[\sum_{i=۱}^{n}(۱-\frac{۱}{n})E(\hat{X}_i\hat{Y}_i) - \frac{۱}{n}E\big(\sum_{i\neq j}\sum \hat{X}_i\hat{Y}_j\big)\big] \\
&= \frac{۱}{n-۱}\big[\frac{n-۱}{n}nE(\hat{X}\hat{Y}) - \frac{۱}{n}n(n-۱)E(\hat{X})E(\hat{Y})\big] \\
&= \frac{۱}{n-۱}\big[(n-۱)(E(\hat{X}\hat{Y}) - E(\hat{X})E(\hat{Y}))\big] = \mathrm{Cov}(\hat{X},\hat{Y}).
\end{aligned}
$$

## ۶ برآورد میانگین تعداد تقلب و برآورد میانگین میزان مصرف روزانه سیگار در بین دانشجویان با روش پیشنهادی

تقلب در امتحانات و مصرف سیگار در بین دانشجویان از موضوعات مهمی است که در امر آموزش وجود دارد. همین مسئله تا حدودی باعث افت تحصیلی دانشجویان شده است و علاوه بر آن نتایج فرهنگی و اجتماعی جبران‌ناپذیری را به دنبال خواهد داشت. میانگین تعداد تقلب و میانگین میزان مصرف روزانه سیگار در هر دانشگاه می‌توانند ملاک‌هایی جهت تایید برنامه‌های آموزشی آن دانشگاه باشند. در این بخش با استفاده از روش پاسخ تصادفیده کمی جدید ارائه شده در بخش ۲ میانگین تعداد تقلب و میانگین میزان مصرف روزانه سیگار را در دانشگاه شهید چمران اهواز در نیمسال اول تحصیلی سال ۹۸–۹۷ برآورد کرده‌ایم. دانشجویان مقطع کارشناسی دانشگاه را به عنوان جامعه آماری در نظر گرفته‌ایم و هر کدام از دانشکده‌ها به عنوان یک طبقه از جامعه در نظر گرفته شده‌اند. از هر طبقه نمونه‌ای به تصادف انتخاب شده و با استفاده از تکمیل پرسش‌نامه و مصاحبه حضوری داده ها جمع‌آوری شده است.

در پرسش‌نامه از دو ابزار تصادفی کردن پرتاب سکه و دسته کارت استفاده شده است. در مرحله اول ابتدا از پاسخ دهنده تقاضا می‌شود از دسته کارت که یکی از اعداد ۰تا ۹ بر روی این کارت‌ها نوشته شده است یک کارت را به تصادف انتخاب کند و بدون آنکه کسی متوجه عدد روی کارت شود آن را به خاطر بسپارد و سپس به دور از چشم دیگران یک سکه را دو مرتبه بیندازد. سکه مورد نظر همگن بوده و احتمال شیر و خط آمدن یکسان است. اگر با پرتاب سکه هر دو بار شیر باشد در پاسخنامه در قسمت پاسخ تصادفیده فقط عدد روی کارت را وارد نماید که در اینصورت برای حالت اول $p_۱ = ۰/۲۵$ است. اگر هردو بار خط بیاید در پاسخنامه قسمت پاسخ تصادفیده فقط عدد تعداد تقلب در ترم گذشته را وارد کند. برای این حالت $p_۲ = ۰/۲۵$ است. اگر یک شیر و یک خط بیاید در پاسخنامه در قسمت پاسخ تصادفیده از عدد ۹ تعداد تقلب‌های خود را کم کرده و فقط حاصل آن را وارد نماید. برای

این حالت $p_۳ = ۰/۵$ است. سپس در مرحله دوم ابتدا از پاسخ دهنده تقاضا می‌شود از دسته کارت که یکی از اعداد ۰ تا ۲۰ بر روی این کارت‌ها نوشته شده است یک کارت را به تصادف انتخاب کند و بدون آنکه کسی متوجه عدد روی کارت شود آن را به خاطر بسپارد و سپس به دور از چشم دیگران یک سکه را دو مرتبه بیندازد. سکه مورد نظر همگن بوده و احتمال شیر و خط آمدن یکسان است. اگر با پرتاب سکه هر دو بار شیر بیاید در پاسخنامه در قسمت پاسخ تصادفیده فقط عدد روی کارت را وارد نماید که در این صورت برای حالت اول $q_۱ = ۰/۲۵$ است. اگر هردو بار خط بیاید در پاسخنامه قسمت پاسخ تصادفیده فقط میزان مصرف سیگار در روز را وارد کند. برای این حالت $q_۲ = ۰/۲۵$ است. اگر یک شیر و یک خط بیاید در پاسخنامه در قسمت پاسخ تصادفیده از عدد ۲۰ میزان مصرف سیگار در روز خود را کم کرده و فقط حاصل آن را وارد نماید. برای این حالت $q_۳ = ۰/۵$ است. در آخر نیز نام، نام‌خانوادگی، جنسیت، رشته تحصیلی، معدل و تعداد واحدهای پاس شده از افراد نمونه، سؤال شده است.

با جایگذاری $b_۱ = ۹$، $p_۱ = ۰/۲۵$، $p_۲ = ۰/۲۵$، $p_۳ = ۰/۵$ و $\mu_{s_۱} = ۴/۵$ در رابطه‌های (۲) و (۴) برآورد میانگین و برآورد واریانس تعداد تقلب براساس پاسخ‌های تصادفیده به صورت

$$\hat{\mu}_y = \frac{\bar{Z} - ۵/۶۲۵}{-۰/۲۵} \quad \text{و} \quad \hat{V}(\hat{\mu}_y) = \frac{s_z^۲}{n(-۰/۲۵)^۲}.$$

محاسبه و با به‌کار بردن آنها برای طبقه (دانشکده) $h$ام داریم:

$$\hat{\mu}_{y_h} = \frac{\bar{Z}_h - ۵/۶۲۵}{-۰/۲۵} \quad \text{و} \quad \hat{V}(\hat{\mu}_{y_h}) = \frac{s_{z_h}^۲}{n_h(-۰/۲۵)^۲},$$

که در آن $n_h$ حجم نمونه انتخابی از طبقه $h$ام است. برآورد میانگین تعداد تقلب دانشجویان هر دانشکده به همراه برآورد واریانس آن در ستون پنجم و ششم جدول ۵ آمده است. با جایگذاری $b_۲ = ۲۰$، $q_۱ = ۰/۲۵$، $q_۲ = ۰/۲۵$، $q_۳ = ۰/۵$ و $\mu_{s_۲} = ۱۰$ در رابطه‌های (۲) و (۴) برآورد میانگین و واریانس میزان مصرف روزانه سیگار براساس پاسخ‌های تصادفیده به صورت

$$\hat{\mu}_x = \frac{\bar{T} - ۱۲/۵}{-۰/۲۵} \quad \text{و} \quad \hat{V}(\hat{\mu}_x) = \frac{s_t^۲}{n(-۰/۲۵)^۲}.$$

محاسبه و با به‌کار بردن آنها برای طبقه (دانشکده) $h$ام داریم:

$$\hat{\mu}_{x_h} = \frac{\bar{T}_h - ۱۲/۵}{-۰/۲۵} \quad \text{و} \quad \hat{V}(\hat{\mu}_{x_h}) = \frac{s_{t_h}^۲}{n_h(-۰/۲۵)^۲}.$$

برآورد میانگین میزان مصرف سیگار در روز دانشجویان هر دانشکده به همراه برآورد واریانس آن در ستون هفتم و هشتم جدول ۵ آمده است. اگر تعداد طبقات را برابر $L$ در نظر بگیریم، برآورد میانگین تعداد تقلب و برآورد میانگین

میزان مصرف روزانه سیگار به همراه برآورد واریانس آن‌ها برای جامعه (دانشگاه) نیز از رابطه‌های

$$\hat{\mu}_{stc} = \sum_{h=1}^{L} W_h \hat{\mu}_{y_h}, \qquad \hat{V}(\hat{\mu}_{stc}) = \sum_{h=1}^{L} W_h^2 \hat{V}(\hat{\mu}_{y_h}),$$
$$\hat{\mu}_{sts} = \sum_{h=1}^{L} W_h \hat{\mu}_{x_h}, \qquad \hat{V}(\hat{\mu}_{sts}) = \sum_{h=1}^{L} W_h^2 \hat{V}(\hat{\mu}_{x_h}),$$

حاصل می‌شوند، که در آن $\hat{\mu}_{stc}$ و $\hat{V}(\hat{\mu}_{stc})$ و $\hat{\mu}_{sts}$ و $\hat{V}(\hat{\mu}_{sts})$ به‌ترتیب برآورد میانگین و برآورد واریانس تعداد تقلب و برآورد میانگین و برآورد واریانس میزان مصرف روزانه سیگار به روش طبقه‌بندی هستند. $W_h = \frac{N_h}{N}$ برابر وزن طبقه $h$ام، که $N_h$ حجم طبقه $h$ و $N$ حجم کل افراد طبقات ($\sum_{h=1}^{\ell} N_h$) است، که $\hat{\mu}_{y_h}$ و $\hat{V}(\hat{\mu}_{y_h})$ و $\hat{\mu}_{x_h}$ و $\hat{V}(\hat{\mu}_{x_h})$ به ترتیب برآورد میانگین و برآورد واریانس تعداد تقلب و برآورد میانگین میزان مصرف روزانه سیگار به همراه برآورد واریانس آن براساس روش پیشنهادی در طبقه $h$ام هستند.

**جدول ۵.** برآوردهای میانگین تعداد تقلب و میانگین مصرف روزانه سیگار دانشجویان و برآورد واریانس آن‌ها

| دانشکده | $h$ | $N_h$ | $n_h$ | $\hat{\mu}_{y_h}$ | $\hat{V}(\hat{\mu}_{y_h})$ | $\hat{\mu}_{x_h}$ | $\hat{V}(\hat{\mu}_{x_h})$ | $\widehat{\mathrm{Cov}}(\hat{X},\hat{Y})$ |
|---|---|---|---|---|---|---|---|---|
| ادبیات و علوم انسانی | ۱ | ۷۶۸ | ۴۰۵ | ۵/۴۲۴ | ۰/۳۱۸۹ | ۱۳/۹۳۲ | ۱/۵۷۳۲ | −۸/۸۶۷ |
| اقتصاد و علوم اجتماعی | ۲ | ۱۳۴۶ | ۳۰۳ | ۵/۹۴۴ | ۰/۴۳۱۲ | ۱۴/۲۱۲ | ۲/۰۱۶۲ | −۲۲/۸۷۶ |
| الهیات و معارف اسلامی | ۳ | ۶۶۹ | ۵۰ | ۲/۵۸ | ۲/۶۰۵۴ | ۰ | ۱۳/۱۴۰۸ | ۰/۷۹۰ |
| تربیت بدنی و علوم ورزش | ۴ | ۲۱۳ | ۱۷۹ | ۰/۵۸ | ۰/۳۵۲۸ | ۰/۵۹۲ | ۱/۷۲۳۸ | ۵/۳۶۹ |
| علوم تربیتی و روانشناسی | ۵ | ۶۸۶ | ۱۰۰ | ۵/۱۴ | ۱/۲۴۸۳ | ۱۶/۸۴ | ۵/۳۱۸۰ | −۱۱/۷۷۵ |
| علوم | ۶ | ۸۷۴ | ۴۷۴ | ۵/۲۲۴ | ۰/۲۸۷۷ | ۱۳/۵۶ | ۱/۳۸۸۶ | ۴/۲۹۷ |
| علوم ریاضی و کامپیوتر | ۷ | ۴۹۰ | ۳۰۱ | ۳/۴۸۴ | ۰/۶۷۷۱ | ۶/۹۷۲ | ۱/۷۶۰۰ | ۲۲/۴۷۲ |
| علوم زمین | ۸ | ۴۰۳ | ۵۰ | ۵/۰۶ | ۲/۷۷۸۸ | ۳/۳۶ | ۱۳/۷۸۷۵ | ۳۳/۰۳۸ |
| فنی و مهندسی | ۹ | ۱۵۵۲ | ۳۰۰ | ۵/۳۱۲ | ۰/۳۸۲۸ | ۹/۰۲۸ | ۱/۸۵۹۲ | −۱۹/۸۵۴ |
| کشاورزی | ۱۰ | ۵۶۸ | ۲۰۰ | ۴/۹۲ | ۰/۶۴۷۹ | ۱۴/۷۴ | ۲/۸۸۰۸ | −۱۳/۲۳۲ |
| مهندسی علوم آب | ۱۱ | ۲۱۷ | ۵۰ | ۱/۵۴ | ۰/۵۹۵۲ | ۰ | ۱/۵۷۷۹ | ۹/۸۳۵ |

## بحث و نتیجه‌گیری

در این مقاله یک روش پاسخ تصادفیده کمی جدید برای برآورد میانگین متغیر حساس پیشنهاد و برآورد واریانس آن ارائه گردید. سپس به مقایسه برتری روش پیشنهادی با روش پاسخ تصادفیده جمعی وارنر (۱۹۷۱) و روش پاسخ تصادفیده ضربی ایچهارن و هایر (۱۹۸۳) با استفاده از شبیه‌سازی (با نرم افزار $R$) پرداختیم و نشان دادیم که روش پیشنهادی نسبت به روش جمعی وارنر و روش ضربی ایچهارن و هایر ارجحیت دارد. سپس روش پاسخ تصادفیده فوق را برای دو متغیر حساس به کار برده و برآوردگری برای کوواریانس بین دو متغیر حساس با استفاده از پیشگوهای نااریب ارائه کردیم و در یک مطالعه تجربی یک برآورد نااریب برای میانگین تعداد تقلب و یک برآورد نااریب

برای میانگین میزان مصرف سیگار در روز دانشجویان دانشگاه شهید چمران اهواز براساس روش پاسخ تصادفیده پیشنهادی جدید بدست آورده شد که برآورد میانگین تعداد تقلب و برآورد میانگین میزان مصرف روزانه سیگار به ترتیب، $\hat{\mu}_{stc} = ۴٫۷۷۹۲$ و $\hat{\mu}_{sts} = ۹٫۸۸۲۸$ می‌باشند. این اعداد بیانگر این هستند که دانشجویان دانشگاه شهید چمران اهواز در امتحانات به طور میانگین ۴٫۷۷۹۲ تقلب کردند و به طور میانگین میزان مصرف روزانه سیگار آن‌ها ۹٫۸۸۲۸ بوده است. برآورد واریانس آن‌ها به‌ترتیب $\hat{V}(\hat{\mu}_{stc}) = ۰٫۸۰۸۰$ و $\hat{V}(\hat{\mu}_{sts}) = ۳٫۷۵۱۶$ حاصل شد. همچنین یک برآورد برای کوواریانس بین تعداد تقلب و میزان مصرف روزانه سیگار دانشجویان دانشگاه شهید چمران اهواز با استفاده از پیشگوهای نااریب ۲٫۸۳۷ به‌دست آمده است که بیانگر وجود ارتباط بین تقلب کردن در امتحانات و مصرف سیگار دانشجویان دانشگاه شهید چمران اهواز است.

## تقدیر و تشکر

نویسندگان از سردبیر، داوران و ویراستار محترم مجله علوم آماری بابت نظرات ارزشمند و مفیدشان نهایت تشکر و قدردانی را دارند.

## مراجع

علوی، س. م. ر. (۱۳۸۶)، تصادفی کردن پاسخ سؤالات حساس چندگزینه‌ای و برآورد نسبت گزینه‌های نسبت تقلب دانشجویان در دانشگاه شهید چمران اهواز، مجله اندیشه آماری، **۱۲**، ۱۹-۱۳.

علوی، س. م. ر. و تاج الدینی، م. (۱۳۹۴)، معرفی یک روش پاسخ تصادفیده جدید و مقایسه آن با روش سیمونس، مجله علوم آماری، **۹**، ۲۳۹-۲۲۷.

یزاری، ز. و علوی، س. م. ر. (۱۳۹۳)، مدل پاسخ تصادفیده کمی اختیاری سه مرحله‌ای، مجله علوم آماری، **۸**، ۲۶۰-۲۴۵.

علوی، س. م. ر.، جوهرزاده. م. و چینی‌پرداز، ر. (۱۳۹۸)، یک مدل پاسخ تصادفیده مرکب جدید ، مجله علوم آماری، **۱۳**، ۱۸۴-۱۷۳

علوی، س. م. ر.، نیری‌مازی، س. و آخوند، م. ر. (۱۳۹۸)، مدل تعداد اقلام تصادفیده و مقایسه آن با مدل تصادفیده سایمون، مجله علوم آماری، **۱۳**، ۵۱۳-۴۹۷.

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

Eichhorn, B. H. and Hayre, L. S. (1983), Scrambled Randomized Response Methods for Obtaining Sensitive Quantitative Data, *Journal of Statistical Planning and Inference*, **7**, 307-316.

Gjestvang, C. R. and Singh, S. (2007), Forced Quantitative Randomized Response Model: a New Device, *Metrika*, **66**, 243-257.

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.

Greenberg, B. G., Kuebler Jr, R. R., Abernathy, J. R. and Horvitz, D. G. (1971), Application of the Randomized Response Technique in Obtaining Quantitative Data, *Journal of the American Statistical Association*, **66**, 243-250.

Gupta, A., Jha, R. K. and Jain, S. (2017), Attack Modeling and Intrusion Detection System for 5G Wireless Communication Network. *International Journal of Communication Systems*, **30**, 32-37.

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

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.

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.

Yan, Z., J. Wang, and J. Lai. (2008), An Efficiency and Protection Degree-Based Comparison Among the Quantitative Randomized Response Strategies, *Communications in Statistics- Theory and Methods*, **38**, 400-408.