فصلنامه پژوهش های اقتصادی ایران/ شماره ۱۱ / تابستان ۸۱

پیشنهادی برای برآورد نسبت بیکاری در سطح استان‌های کشور

حمیدرضا نواب‌پور [*]

علیرضا خوشگویانفرد [**]

محمد طاهری منزه [***]

تاریخ ارسال: //۸۱   تاریخ پذیرش: //۸۱

چکیده

معمولاً دفاتر ملی آمار کشورها، به دلیل محدودیت‌هایی از قبیل بودجه و زمان سعی می‌کنند طرح‌های آمارگیری خود را در سطوح جغرافیایی بزرگ مثلاً ، در سطح ملی،طراحی و اندازه نمونه‌ها را بهینه کنند. این امر، سبب می‌شود که پژوهشگران اقتصادی، برنامه‌ریزان، و تصمیم گیرندگانی که نیاز به اطلاعات آماری در سطوح کوچکتر جغرافیایی و یا به تفکیک صفاتی خاص دارند از دستیابی به اطلاعات مورد نیاز محروم بمانند. کاربران این نوع اطلاعات با انتقال نیازهای آماری خود به واحدهای تولید کننده آمار موجب شده‌اند که این واحدها، طرح‌های آمارگیری خود را در سطوح کوچکتر جغرافیایی – مثلاً در سطح استان – طراحی و اندازه نمونه‌ها را بهینه کنند. این امر،افزون بر تحمیل هزینه به بودجه عمومی کشورها و طولانی‌تر شدن فرایند تهیه و توزیع آمار، موجب می‌شود که خطاهای غیر نمونه‌گیری سهم بیشتری در مراحل تهیه و توزیع آمارها داشته باشند. در این مقاله، سعی می‌شود با معرفی چند روش برآورد ناحیه کوچک[1] (SAE) روشی مناسب برای برآورد نسبت بیکاری در سطح استان‌های کشور پیشنهاد شود.

واژه‌های کلیدی: *برآوردگر بیز تجربی، برآوردگر غیرمستقیم، برآورد حافظ ساختار، برآوردگر ترکیبی، طبقات پسین،ناحیه کوچک.*

---

[*] استادیار دانشگاه علامه طباطبائی   Email:h.navvabpour@src.ac.ir

[**]محقق مرکز تحقیقات صدا و سیمای جمهوری اسلامی ایران

[***]کارشناس مرکز آمار ایران

1.Small Area Estimation

۱. مقدمه

بیشتر آمارگیری‌های نمونه‌ای که به طور مستمر اجرا می‌شوند، به منظور ارایه برآورد در مورد مناطق از پیش تعیین شده، مثلاً در سطح ملی، طراحی می‌شوند. این در حالی است که اکثر پژوهشگران اقتصادی، برنامه‌ریزان، وتصمیم گیرندگان نیازمند استفاده از آمارهای قابل اعتماد در سطوح کوچکتر جغرافیایی و یا به تفکیک صفتی خاص مثلا شهری و روستایی، هستند. به منظور صرفه‌جویی در زمان و هزینه‌ها، غالباً از این آمارگیری‌ها به طور مستقیم برای ارایه برآوردهای مربوط به زیر جامعه‌های (نواحی کوچک) مناطق مذکور بهره‌گیری می‌شود،اگرچه تعداد نمونه‌ها در ناحیهٔ کوچک به دلیل کوچک بودن نقش معرف بودن خود را ایفا نمی‌کنند. به همین دلیل از آمارشناسان و جمعیت شناسان انتظار می‌رفت که برآوردهای قابل اعتمادی را بدون در اختیار داشتن داده‌های لازم در مورد زیر ناحیه‌ها به کاربران ارایه کنند. از این رو،آنها برای تأمین این نیاز روش‌هایی را با عنوان روش‌های برآورد ناحیه کوچک ابداع کردند. وجه مشترک کلیه این روش ها که آنها را مطلوبتر از برآوردگر مستقیم ساخته، بهره‌گیری از داده‌های مکمل علاوه بر داده‌های نمونه است.

هدف اصلی از کاربرد روش‌های برآورد ناحیه کوچک، ارایه برآوردهای قابل اعتماد برای زیر مجموعه‌های یک جمعیت است، در صورتی که اندازهٔ نمونه برای کل جمعیت تعیین شده باشد. مثلاً، مرکز آمار ایران به علت نیاز فزاینده به برآورد نسبت بیکاری در سطح استان‌های کشور، چند سالی است که برای برآورد نسبت بیکاری در سطح استان ها، سالیانه اقدام به اجرای آمارگیری‌های نمونه‌ای در سطح استان ها می‌کند، در صورتی که اقتصاددانان به صورت ماهانه و یا حداقل به صورت فصلی به این آمارها نیاز دارند، و این نیاز به‌دلیل محدودیت‌های بودجه‌ای، زمانی، حرفه‌ای، اجرایی، پردازش و اطلاع‌رسانی قابل تأمین نیست مگر، با به کارگیری روش‌های برآورد ناحیه کوچک.

باید خاطر نشان ساخت که برآوردهای مستقیم این زیر مجموعه‌ها (نواحی کوچک)، نمی‌توانند به علت کوچکی اندازهٔ نمونه (خطای بزرگ نمونه‌گیری)،برآوردهای قابل اعتمادی به دست دهند. چندین برآوردگر غیرمستقیم در طی تقریباً چهل سال پیشنهاد شده است. برای تعریف مسئله و بررسی روش‌های برآورد ناحیه کوچک، خواننده می تواند به پرسل و کیش[۱] (۱۹۷۹،۱۹۸۰)، گوش و رائو[۲] (۱۹۹۴)، شیبل[۳] (۱۹۹۵)، مارکر[۴] (۱۹۹۹)، و پفرمان[۱] (۱۹۹۹) مراجعه کند.

1.Purcell & Kish
2. Ghash & Rao
3.Schaible
4.Marker

در ایران، مطالعه روش‌های برآورد ناحیه کوچک در سازمان‌های آماری و سایر سازمان‌های تولید کنندهٔ‌آمار نادیده گرفته شده است. در سال‌های اخیر این موضوع، زمینه‌های جالبی را برای پژوهش فراهم آورده است. در این مقاله، سه روش برآورد ناحیه کوچک ترکیبی، مرکب، و بیز تجربی به منظور برآورد نسبت بیکاری در سطح استان‌های کشور– در حالی که اندازهٔ نمونه در سطح کل کشور بهینه شده است– در نظر گرفته می شود.

ابتدا، از راه برآورد حافظ ساختار، برآوردگر ترکیبی مورد بررسی قرار می‌گیرد. استفاده از برآوردگر ترکیبی به علت سادگی محاسبات و سازگاری با طرح‌های رایج نمونه‌گیری بسیار متداول است. این برآوردگر، بر "ترکیب" اطلاعات پیشین ناحیه‌های کوچک و برآوردگرهای نمونه‌بخش‌های دوبدو ناسازگار و جامع از جامعه‌ای که نواحی کوچک را در برمی‌گیرد،متکی است. به همین دلیل است که این برآوردگر به عنوان برآوردگر ترکیبی شناخته شده است. مطالعات تجربی نشان داده‌است که این برآوردگر اریب ولی نسبت به برآوردگر مستقیم دارای واریانس کوچکتری است. استفاده از اصطلاح برآورد ترکیبی را به مرکز ملی آمارهای بهداشتی آمریکا منتسب می‌کنند که به نظر می‌رسد اولین استفاده مستند از این برآوردگر در آنجا صورت گرفته است. سپس، برآوردگر مرکب به صورت مجموع موزون برآوردهای مستقیم و ترکیبی بررسی و در نهایت، روش بیز تجربی به طور خلاصه ذکر می‌شود.

در تحلیل کاربردی، یک طرح نمونه‌گیری را که برای برآورد نسبت بیکاری در سطح کشور طراحی شده است مورد توجه قرار می‌دهیم. سپس، با بهره‌گیری از داده‌های به دست آمده از ۲۸ استان کشور برآوردهای خود را ارایه می‌کنیم و عملکرد مربوط به هر یک از آنها را مورد مقایسه قرار می‌دهیم. مبنای مقایسه این برآوردها، برآوردی است که با بهینه کردن اندازه نمونه در سطح استان به دست می‌آید.

۲. مروری بر برآوردگرها

۲-۱. برآورد حافظ ساختار[2]

1. Pfeffermann
2 ، Structure Preserving Estimation

این برآوردگر اساساً، مقادیر اولیه پارامترهای مربوط به نواحی کوچک را که از سرشماری قبل به دست آمده است، روز آمد می‌کند که این امر،از روی داده‌های حاصل از آمارگیری‌های نمونه‌ای بزرگ (یا سایر منابع روز آمد) انجام می شود. در اینجا روشی که منتسب به پرسل و کیش (۱۹۸۰) است و چارچوب آن در واقع برای داده‌های شمارشی طراحی شده است،در نظر گرفته می شود.

فرض کنید افزون بر متغیر مورد نظر که به عنوان برآوردنی[1] شناخته می شود، یک یا چند متغیر کمکی دیگر نیز که با این متغیر در حد بالایی مرتبط هستند، در اختیار باشد. برای به کار بردن روش مطلوب نیاز به دو مجموعه اطلاعات زیر است که رابطه‌های بین این برآوردنی و متغیرهای کمکی را سازمان دهد.

۱. ساختار پیوند

این ساختار، رابطه‌ها را در سطح ناحیه کوچک مورد بررسی، سازمان می‌دهد. نظر به اینکه داده‌های جاری برای نواحی کوچک موجود نیست، این ساختار، از روی داده‌های سرشماری ایجاد می‌شود و در قالب یک جدول پیشایندی سه بُعدی با فراوانی خانه‌های N_{ijh} که در آن اندیس‌ها به ترتیب به iامین ناحیه کوچک $(i=1,2,...,I)$، jامین گروه متغیر کمکی $(j=1,2,...,J)$ و hامین گروه برآوردنی $(h=1,2,...,H)$ مربوط می‌شود، به نمایش در می‌آیند.

۲.ساختار انتساب

این ساختار رابطه‌ها را در سطح جامعه (یا زیر جامعه‌ای که ناحیه‌های کوچک را شامل می‌شود) سازماندهی می‌کند. این ساختار، با یک جدول پیشایندی دوبعدی با فراوانی خانه‌های m_{jh} (که در آن اندیس‌ها به صورت قبل تعریف می شوند)، نشان داده می‌شود. توجه داشته باشید که می‌توان m_{jh}هارابه عنوان حاشیه‌های بهنگام شدهٔ جدول قبلی همراه با $N_{\bullet jh}$ها محسوب کرد($N_{\bullet jh}$ هاشمارش‌های جدول هستند که روی اندیسiجمع بسته شده‌اند)، با این دلیل که $N_{\bullet jh}$ها از روی داده‌های قدیمی و m_{jh}ها از روی داده‌های جاری به دست می‌آیند.

هدف، برآورد(یا اصلاح) N_{ijh} هایی است که با m_{jh}ها (حاشیه‌های بهنگام شده) سازگار باشند و روابط موجود در ساختار پیوند را حفظ کنند. چنانچه، فراوانی‌های خانه‌ای برآورد شده را با

1. Estimand

$X = \{X_{ijh}\}$ نشان دهیم، هدف نهایی روی حاشیه‌هایی از X است که مربوط به نواحی کوچک هستند یعنی $X_{i.h}$. حال،فرض کنید که $N = \{N_{ijh}\}$ و $M = \{m_{jh}\}$، مشخص باشند. برای برآورد، می‌توان از روش کمترین توان‌های دوم موزون برای مینیمم کردن مجموع توان‌های دوم موزون تفاوت‌های بین N_{ijh} و X_{ijh} نسبت به قیدهای حاشیه‌ای که با ساختار انتساب معین می‌شوند، استفاده کرد به این معنی که تابع لاگرانژ زیر را مینیمم کرد:

$$F\left(X_{ijh}, \lambda\right) = \sum_{i,j,h} W_{ijh}^{-1}\left(X_{ijh} - N_{ijh}\right)^2 - \lambda\left(\sum_i X_{ijh} - m_{jh}\right) \tag{۱}$$

که در آن W_{ijh} ها، وزن‌های از پیش تعیین شده‌اند و λ ضریب لاگرانژ است. چنانچه مشتقات $F\left(X_{ijh}, \lambda\right)$ نسبت به X_{ijh} و λ برابر صفر قرار داده شود، رابطه زیر به دست می‌آید:

$$X_{ijh} = W_{ijh}\, W_{.jh}^{-1}\left(m_{jh} - N_{.jh}\right) + N_{ijh} \tag{۲}$$

با انتخاب $W_{ijh} = N_{ijh}$ فراوانی خانه‌ای برآورد شده به صورت زیر است:

$$X_{ijh} = \frac{N_{ijh}}{N_{.jh}}\, m_{jh} \tag{۳}$$

با جمع بستن(۳) روی بعد متغیر کمکی، برآورد برای hامین گروه برآوردنی درiامین ناحیه کوچک از رابطهٔ زیر به دست می‌آید:

$$X_{i.h} = \sum_j \frac{N_{ijh}}{N_{.jh}}\, m_{jh} \tag{۴}$$

و در نهایت، برآورد حافظ ساختار برای نسبت مورد نظر به شرح زیر حاصل می‌شود:

$$\hat{P}_{i.h}^{s} = \frac{X_{i.h}}{N_{i..}} = \sum_{j} \frac{N_{ijh}}{N_{.jh}\ N_{i..}}\ m_{jh} \tag{۵}$$

۲-۲.برآوردگر ترکیبی[۱]

رابطه (۴) را به تعریف سنتی برآوردگر ترکیبی که به اختصار در مقدمه شرح داده شد، مرتبط می‌کنیم. برای برآورد، در درجهٔ اول، جامعه‌ای که بر نواحی کوچک احاطه دارد به چند طبقه تحت عنوان طبقه‌های پسین براساس متغیری که کاملاً با متغیر مورد نظر مرتبط است،افرازمی‌شود. چنانچه،از طبقه‌های پسین برآوردهای نااریب مستقیم دراختیار باشد، برآورد ترکیبی پارامتری که مورد بررسی است برابر با مجموع موزون برآوردهای طبقه پسین خواهد بود. در رابطه (۴)، $\frac{N_{ijh}}{N_{.jh}}$ ها و m_{jh}ها به ترتیب وزن ها و برآوردهای نااریب مستقیم طبقات پسین هستند که به وسیله متغیر کمکی تشکیل شده‌اند.

۲-۳. برآورد مرکب[۲]

برآوردگر مرکب، ترکیب خطی ازبرآوردگرهای ترکیبی و مستقیم ناحیه کوچک است. این برآوردگر دارای شکل کلی زیر است:

$$\hat{P}_{i}^{SH} = W_{i}\ \hat{P}_{i}^{D}\ + \left(1 - W_{i}\right)\ \hat{P}_{i}^{S} \tag{۶}$$

1، Synthetic Estimator
2، composite Estimator

که در آن، $\hat{P}_i^D$ برآوردگر مستقیم(نسبت) مبتنی بر طرح برای iامین (نسبت) ناحیه کوچک است و $0 \le W_i \le 1$. مهمترین مسئله، تعیین وزن مناسبی است که توازن بین اریبی برآوردگر ترکیبی و بی‌ثباتی برآوردگر مستقیم را برقرار کند. وزن‌های مختلفی در متون پیشنهاد شده اند در اینجا، وزن مناسب که با مینیمم کردن میانگین توان دوم خطا (MSE) نسبت به W_i به دست می آید، ذکر می‌شود. میانگین توان دوم خطا برای رابطه (۶) با فرض[۱] $Cov(\hat{P}_i^D \; , \; \hat{P}_i^S) \cong 0$ به صورت رابطه (۷) است:

$$\begin{aligned} MSE\left(\hat{P}_i^{SH}\right) &= E\left(\hat{P}_i^{SH} - P_i\right)^2 \\ &= E\left\{W_i\left(\hat{P}_i^D - P_i\right) + \left(1-W_i\right)\left(\hat{P}_i^S - P_i\right)\right\}^2 \\ &= W_i^2\, Var\left(\hat{P}_i^D\right) + \left(1-W_i\right)^2\, MSE\left(\hat{P}_i^S\right) \end{aligned} \quad (۷)$$

بنابراین، وزن بهینه به صورت زیر به دست می آید:

$$W_i^{opt} = \frac{MSE(\hat{P}_i^S)}{MSE(\hat{P}_i^S) + Var(\hat{P}_i^D)} \quad (۸)$$

باید توجه کرد که وقتی برآوردگر مستقیم قابل اعتماد باشد، وزن بیشتری خواهد داشت، در غیر این صورت وزن بیشتری به برآوردگر ترکیبی داده خواهد شد. در عمل، وزن بهینه دررابطه(۸) نیز،اغلب پس از برآورد $MSE\left(\hat{P}_i^S\right)$ و $Var\left(\hat{P}_i^D\right)$ برآورد می‌شود. بر طبق طرح نمونه‌گیری چنانچه،داده‌های نمونه از روی iامین ناحیه کوچک موجود باشد، یک برآوردگر مبتنی بر مدل نااریب $Var\left(\hat{P}_i^D\right)$ نیز موجود خواهد بود. بنابراین، کافی است برآوردگری را برای $MSE\left(\hat{P}_i^S\right)$ به دست آورد:

---

۳. این فرض می تواند بر قرار باشد، زیرا، تجربه‌های عملی نشان داده‌است که تغییرات $\hat{P}_i^s$ نسبت به $\hat{P}_i^D$ بسیار کمتر بوده به طوری که گویا $\hat{P}_i^s$ در مقایسه با $\hat{P}_i^D$ ثابت است.

$$
\begin{aligned}
MSE(\hat{P}_i^S) &= E(\hat{P}_i^S - P_i)^2 \\
&= E(\hat{P}_i^s - \hat{P}_i^D + \hat{P}_i^D - P_i)^2 \\
&= E(\hat{P}_i^S - \hat{P}_i^D)^2 + E(\hat{P}_i^D - P_i)^2 + 2E\left[(\hat{P}_i^S - \hat{P}_i^D)(\hat{P}_i^D - P_i)\right]
\end{aligned}
\tag{۹}
$$

چنانچه $Cov(\hat{P}_i^D \ , \ \hat{P}_i^S) \cong 0$ خواهیم داشت:

$$
\begin{aligned}
E\left[(\hat{P}_i^S - \hat{P}_i^D)(\hat{P}_i^D - P_i)\right] &= E\left[(\hat{P}_i^S - P_i + P_i - \hat{P}_i^D)(\hat{P}_i^D - P_i)\right] \\
&= -Var(\hat{P}_i^D)
\end{aligned}
\tag{۱۰}
$$

در نتیجه، رابطه (۱۱) به دست می‌آید:

$$MSE(\hat{P}_i^S) = E(\hat{P}_i^S - \hat{P}_i^D)^2 - Var(\hat{P}_i^D) \tag{۱۱}$$

بدین ترتیب، می‌توان رابطه(۱۱) را براساس رابطه زیر که ممکن است منفی باشد به طور ناااریب برآورد کرد:

$$M\hat{S}E(\hat{P}_i^S) = (\hat{P}_i^S - \hat{P}_i^D)^2 - V\hat{a}r(\hat{P}_i^D) \tag{۱۲}$$

با قرار دادن رابطه (۱۲) و $V\hat{a}r(\hat{P}_i^D)$ به ترتیب به جای $MSE(\hat{P}_i^S)$ و $Var\left(\hat{P}_i^D\right)$ در رابطه (۸)، برآوردی از وزن بهینه که وابسته به داده‌ها است، به دست می‌آید. با این وجود، این وزن می‌تواند از قبل تعیین شده و یا وابسته به اندازه نمونه باشد. خواننده برای مشاهدهٔ سایر انواع وزن‌ها به کوپاس[1] (۱۹۷۲)، گوش و رائو[2] (۱۹۹۴)،وتامپسون وهولموی (۱۹۹۸) ارجاع داده می شود.

پرسل و کیش (۱۹۷۹) نیز، وزن مشترکی را برای کلیه ناحیه‌های کوچک پیشنهاد کرده‌اند. چنانچه نواحی کوچک خیلی متفاوت باشند،این وزن مشترک ممکن است به طور مطلوب عمل نکند

1.Copas
2.Thomssen & Holmoy

شیبل (۱۹۷۸) شیوه برآورد وزن در رابطه(۶) را از طریق مینیمم کردن MSE بررسی کرده است. وی، شرایطی را که تحت آن برآوردگر مرکب منجر به کاهش قابل ملاحظه‌ای درMSE می‌شود،ارایه داده است. تابع زیر را درنظر بگیرید:

$$f(W_i)=\frac{MSE(\hat{P}_i^{SH})}{MSE(\hat{P}_i^{S})}=(R_i+1)W_i^2-2W_i+1 \qquad (۱۳)$$

که در آن $R_i=\frac{Var(\hat{P}_i^{D})}{MSE(\hat{P}_i^{S})}$،این تابع دارای ویژگی مطلوب زیر است:

$$MSE(\hat{P}_i^{SH})\langle MSE(\hat{P}_i^{S})\Leftrightarrow f(W_i)\langle 1\Leftrightarrow(R_i+1)W_i-2\langle 0\Leftrightarrow W_i\langle\frac{2}{R_i+1} \qquad (۱۴)$$

این ویژگی به بازه‌ای به صورت $\left(0,\frac{2}{R_i+1}\right)$ برای W_i می‌رسد که در آن برآوردگر مرکب لزوماً بهتر از هر دو مؤلفه خود عمل می‌کند. برای مثال، چنانچه $R_i = 1$ باشد به ازای تمام $W_i\in(0,1)$، $MSE(\hat{P}_i^{SH})$ کوچکتر از $MSE(\hat{P}_i^{S})$ و $Var(\hat{P}_i^{D})$ است.

دو روشی که تا کنون در نظر گرفته شده‌اند تحت رهیافت فراوانی‌گرا[1] قرار دارند. اکنون روش بیز تجربی (EB) معرفی می‌شود.

۲-۳. برآوردگر بیز تجربی [2]

1. Frequentist
2. Empirical Bayes Estimator

ابتدا، جامعه را به ناحیه‌های کوچک مختلف ۱،۲،...،L که L معلوم است تقسیم می‌کنیم. فرض می‌شود هدف، برآوردنسبت واقعی جامعه برای ناحیه کوچـــک i ام(P_i) است(i=۱،۲،...،L). برآوردگـر مستقیم (P_i)ازنمونه‌ی موجود در نمونه‌گیری کشوری در دسترس است، همچنیـن، بردار اطلاعات کمـکی $x_i = (x_{i1},..., x_{is})'$ از نمونه‌گیری کشوری یا داده‌های ثبتی و یا سرشماری‌های گذشته در اختیار است. باید توجه داشت که این اطلاعات کمکی با P_i مرتبط هستند.

تبدیل g که تابعی ازیک متغیرساده و دارای مشتق مرتبه اول پیوسته و غیر صفر است، در نظر گرفته می‌شود:

$$g_i = g(P_i) \qquad\qquad i = 1,2,...,L$$

فرض می شود $\hat{g}_i$ برآورد g_i است. اکنون،مدل ناحیه کوچک زیر در نظر گرفته می شود:

الف) با شرط گذاری روی $g = (g_1, g_2,..., g_L)'$ ، $\hat{g}_i$ ها دوبدو مستقـــل هســـــــتند، و برای اعضایی که به ناحیه کوچــــک i ام تعلق دارند، $E(\hat{g}_i \mid g_i) = g_i$ و $V(\hat{g}_i \mid g_i) = \delta_i^2 (> 0)$ که در آن δ_i^2 ها مقادیرمعلوم هستند.

ب) g_i ها مستقل و هم توزیع هستند با توزیع $N(x_i' \beta, \tau^2)$.

این مدل را به صورت زیر نیز می‌توان نوشت:

$$\hat{g} = X\beta + t + e \qquad (۱۵)$$

که در آن $\hat{g}$، t و e بردارهای (L×۱)، t بـردار اثرهای تصـــادفی ناحیه ای، e بردار خطاهای تصادفی نمونه‌گیری، و $\hat{g}$ دارای یک توزیع نرمال چند متغیره است. ماتریس X ماتریس طرح(L× s)و β بردار(s×۱)از پارامترهای مجهول است. فرض می‌کنیم بردارهای t و e مستقل و $\sum$ و ∇ ماتریس‌های قطری بوده که مؤلفه‌های روی قطر آنها به ترتیب τ^2 و δ_i^2 هستند. همچنین

$$E(e \mid \hat{g}) = 0 \qquad , \qquad Var(e \mid \hat{g}) = \nabla$$

و

$$t \sim N(0, \Sigma)$$

مؤلفه‌های واریانس δ_i^2 به وسیله فرمول‌های مناسب واریانس نمونه‌گیری کشوری به دست می آیند.
در این بخش،تبدیل لوژستیک رابه صورت زیر تعریف می‌کنیم:

$$\hat{g}_i = Ln \frac{\hat{P}_i}{1-\hat{P}_i} \qquad i = 1,2,\ldots,L$$

برآورد بیز $\hat{g}_i$ به شکل زیر است:

$$\hat{g}_i^B = (1-B_i)\hat{g}_i + B_i x_i' \beta \; ; \qquad i = 1,2,\ldots,L \tag{۱۶}$$

که در آن x_i' ،Iامین سطر ماتریس X بوده و:

$$B_i = \frac{\delta_i^2}{\delta_i^2 + \tau^2} \tag{۱۷}$$

ضریب $(1-B_i)$ میزان عدم قطعیت در الگوسازی θ_iها(یعنی نسبت واریانس (δ_i^2) به واریانس کل ($\delta_i^2 + \tau^2$)) را اندازه می گیرد. بنابراین،برآوردگر بیز تجربی نسبت تغییرات بین ناحیه‌های کوچک را به دقت برآوردگر مستقیم به درستی مورد استفاده قرار می دهد.درضمن، این برآوردگر برای عموم طرح‌های نمونه‌گیری معتبر است، زیرا، به جای تک تک اعضای جامعه، تنها P_iها الگوسازی شده‌اند.

در تحلیل بیز تجربی τ^2 و β نامعلوم هستند و بایستی آنها را به‌وسیله داده ها برآورد کرد. این برآوردها با حل تکراری معادلات زیر به دست می‌آیند:

$$\begin{cases}\tau^2 = (g - X\beta)' V^{-1}(g - X\beta) / L - s \\ \beta = (X'V^{-1}X)^{-1} X'V^{-1} g\end{cases} \qquad (۱۸)$$

که در آن، V ماتریس قطری (s × s) است که درایه (i,i) آن برابر $\delta_i^2 + \tau^2$ است.

پس از اینکه τ^2 با حل تکراری معادلات بالا به دســـت آمد، $\hat{\beta}$ بهترین برآوردگر نااریب خطی (BLUE)، از رابطه زیر به دست می آید:

$$\hat{\beta} = (X'U^{-1}X)X'U^{-1}\hat{g}$$

که درآن، U ماتریس قطری (s× s) است که از ماتـــریسV با قرار دادن $\hat{\tau}^2$ به جای τ^2 نتیجه می‌شود.

اکنون، با توجه به رابطه (۱۶)، برآوردگر بیز تجربی $\hat{g}_i$ به صورت زیر است:

$$\hat{g}_i^{EB} = (1 - \hat{B}_i)\hat{g}_i + \hat{B}_i x_i' \hat{\beta} \qquad i = 1,2,\ldots,L \qquad (۱۹)$$

که در آن، $\hat{B}_i = \delta_i^2 / (\delta_i^2 + \hat{\tau}^2)$.

درنهایت، برآوردگر بیز تجربی $\hat{P}_i^{EB}$، برای مدل لوژستیک از رابطه زیر به دست می آید:

$$\hat{P}_i^{EB} = \frac{e^{\hat{g}_i^{EB}}}{1 + e^{\hat{g}_i^{EB}}} \qquad i = 1,2,\ldots,L \qquad (۲۰)$$

دراینجابه این بحث خاتمه می دهیم و خواننده برای آگاهی از جزییات بیشتر به چاند و آلکساندر[1] (۱۹۹۵) ارجاع داده می شود.

۳.یک کاربرد

در ایران، ۲۸ استان وجود دارد. در سال‌های اخیر به منظور برآورد نسبت‌های بیکاری در سطح استان ها، مرکز آمار ایران طرح‌های نمونه‌گیری مجزا برای هر استان به صورت سالانه طراحی کرده است.

1. Chand & Aletander

برای اجرای این طرح‌های نمونه‌گیری، در کل کشور بیش از ۱۰۰۰۰۰ نمونه انتخاب می‌شود. هزینه ارایه برآورد نسبت بیکاری با این روش به مرکز آمار ایران اجازه نمی‌دهد که این برآوردها را ماهانه یا حداقل به‌صورت فصلی ارایه‌کند. از آنجا که این اطلاعات به صورت ماهانه و یا فصلی مورد نیاز اقتصاددانان و پژوهشگران است و با توجه به محدودیت‌های اجرایی، بودجه‌ای، و... و طولانی شدن زمان تهیه و اطلاع رسانی آن، به نظر می رسد استفاده از روش‌های برآورد ناحیه کوچک امکان مناسبی را برای تأمین این نیاز مهیا کند، لذا، برای برآورد این نسبت در سال ۱۳۸۰ از روش‌های برآورد ناحیه کوچک، معرفی شده در بخش قبل، استفاده شد تا با مقایسه عملکرد آنها به پیشنهادی در به کارگیری یک و یا چند روش دست یابیم. این کاربرد در زیر توضیح داده می شود.

طرح نمونه‌گیری برای برآورد نسبت بیکاری در سطح ملی برای سال ۱۳۸۰ طراحی شد. بر این مبنا، نمونه‌ای در حدود ۱۳۰۰۰ در سطح کل کشور انتخاب شد. نظر به اینکه نمونه‌ها بین تمام ۲۸ استان براساس اندازهٔ آنها تخصیص داده شدند، لذا تعدادی نمونه از هریک از استان ها وجود داشت. این امر، ما را قادر ساخت تا برآورد مستقیمی از نسبت بیکاری در هر استان داشته باشیم. افزون براین، طرح‌های نمونه‌گیری مجزا، همانند سال‌های پیش نیز برای هر استان به منظور ارایه برآوردهای مبتنی بر طرح نسبت‌های بیکاری طراحی شد. بنابراین، دو برآورد یعنی برآورد مستقیم و برآورد مبتنی بر طرح را برای هر استان قبل از بهره‌گیری از روش‌های برآورد ناحیه کوچک، در اختیار داشتیم. برآوردهای مبتنی بر طرح در ستون دوم جدول (۱) نشان داده شده اند. این برآوردها که با بهینه کردن نمونه ها در سطح استان ها برای اردیبهشت ۱۳۸۰(تعداد نمونه ۱۰۴۵۶۶) محاسبه شده‌اند[۱]، مبنای مقایسه برآوردهای ناحیه کوچک هستند.

برآورد ترکیبی که در ستون سوم جدول(۱)به نمایش در آمده است اولین برآورد غیرمستقیمی است که با استفاده از روش‌های برآورد ناحیه کوچک به دست آمد. وسیله به کار رفته برای ارایه این برآوردها بر حسب چهار گروه سنی صورت گرفته است زیرا انتظار می رود که متغیر سن تا حدودی با وضعیت بیکاری مرتبط باشد. وزن $\frac{N_{ijh}}{N_{.jh}}$ در فرمول (۴) از روی داده‌های سرشماری سال ۱۳۷۵ محاسبه شد. همان طور که ملاحظه می شود خطای مطلق(قدرمطلق تفاوت دو برآورد) بین

۲. منبع: نشریه آمارگیری ازویژگی‌های اشتغال و بیکاری خانوار– اردیبهشت ۱۳۸۱ – مرکز آمار ایران

برآورد ترکیبی و برآورد مبتنی بر طرح از ۰/۰۰۰۹ برای استان مازندران تا ۰/۱۶۶۹۹ برای استان سیستان و بلوچستان تغییر می‌کند.شرایطی که تحت آن برآوردگرهای ترکیبی رضایت بخش خواهند بود به شرح زیر است:

۱. مناسب بودن متغیری که طبقه‌ها(طبقه‌های پسین) را تشکیل می‌دهد. چنانچه این متغیر درطبقات تغییرات کمی ایجاد کند، به نظرمی رسد که برآوردگر ترکیبی مناسب‌تر باشد.

۲. استفاده از وزن‌های بهنگام. چنانچه داده‌های ثبتی مرتبطی وجود داشته باشد، می‌توان برای برآوردگر ترکیبی وزن‌های بهنگام‌تری را به کار گرفت.

این دو مشکل، استفاده از برآوردگر ترکیبی را محدود می‌کند. بنابراین، به جای تمرکز روی این برآوردگر، آنرا برای ارایه برآورد مرکب که در ستون چهارم جدول (۱) نشان داده شده‌است، به دست آوردیم. در این حالت، خطای مطلق از ۰/۰۰۰۸۲ برای استان ایلام تا۰/۰۴۲۲ برای استان هرمزگان تغییر می‌کند.

برای محاسبه برآورد بیز تجربی (EB) نسبت بیکاری استان‌های کشور در ابتدا، ماتریس طرح X با $s=۲$ به صورت زیر تعریف می‌شود:

$$\begin{cases} x_{i1} = 1 \\ x_{i2} = N_i \end{cases}$$

که درآن، N_i جمعیت فعال اقتصادی مربوط به سال (۱۳۷۵) هر یک ازاستان‌های کشور است.

برای به دست آوردن برآورد δ_i^2 ازروش دلتا استفاده شد. این روش،به صورت زیر است: فرض می‌کنیم y_1 ، y_2 ,…, y_n یک نمونه تصادفی و $\hat{\theta}_n$ برآورد پارامتر مورد نظر باشد به طوری که برای n بزرگ داشته باشیم:

$$\sqrt{n}\,(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma^2)$$

آن گاه، برای تابعی مانند $g(.)$ با مقدار حقیقی و مشتق مرتبه اول پیوسته در همسایگی θ داریم:

$$\sqrt{n}\,(g(\hat{\theta}_n) - g(\theta)) \xrightarrow{d} N(0, g'^2(\theta)\sigma^2)$$

که در آن، $\sigma^2 = \mathrm{var}(\hat{\theta}_n)$.

برآورد واریانس (σ^2) برای نسبت بیکاری با توجه به اینکه نمونه‌ها از طریق نمونه‌گیری تصادفی ساده به دست آمده‌اند، از رابطه زیر به دست می‌آید:

$$\hat{\sigma}^2 = \hat{\text{var}}(p_i^D) = \frac{P_i^D(1-P_i^D)}{n_i - 1}$$

که در آن n_i تعداد نمونه تخصیص یافته به استان iام و مقدار P_i^D برآورد مستقیم نسبت بیکاری دراستان iام است.

با توجه به برآورد $\hat{\sigma}^2$، مقدار δ_i^2 به صورت زیر برآورد می شود:

$$\hat{\delta}_i^2 = g'^2(\hat{P}_i)\hat{\sigma}^2$$

برای برآورد τ^2 از روش حل تکراری معادله (۱۸) و برنامه نوشته شده در محیط برنامه‌نویسی SAS/IML بهره‌گرفته‌شده است. برآورد τ^2 برای مسئله برآورد بیز تجربی نسبت بیکاری استان‌های کشور در سال ۱۳۸۰ برابر است با:

$$\hat{\tau}^2 = ۰/۵۵۹۶$$

و برآورد بردار پارامترها به صورت زیر است:

$$\hat{\beta} = \begin{pmatrix} -2/07 \\ -1/2E-7 \end{pmatrix}$$

ستون آخر جدول(۱) به برآوردهای بیز تجربی(EB) اختصاص داده شده است. این برآوردها پس از تقریباً ۵۰۰۰ بار حل‌تکراری معادله‌های (۱۸) با استفاده از برنامهٔ SAS/IML محاسبه شده‌اند. متغیر کمکی که در مدل (۱۵) به کار رفته است جمعیت فعال اقتصادی[1] است. همان گونه که ملاحظه می شود خطای مطلق از ۰/۰۰۰۶۸ برای استان گیلان تا ۰/۰۴۰۵۳ برای استان هرمزگان تغییر می‌کند.

1. Economically Active Population (EAP)

برای ارزیابی سه برآوردگر یادشده، میانگین قدر مطلق خطا[۱] که در سطر آخر جدول(۱) مشخص شده، محاسبه شده‌است. این معیار،نشان می‌دهد که برآوردهای مرکب و EB بهترازبرآوردهای ترکیبی عمل می‌کنند زیرا، در مقایسه با برآوردهای ترکیبی، آنها به برآوردهای مبتنی بر طرح(برآورد استانی) نزدیکتر هستند. برآوردهای مرکب و EB خیلی به هم نزدیک هستند. البته،این بررسی دقیقاً نشان نمی‌دهد که برآوردگر EB بهبود مناسبتری نسبت به برآوردگرمرکب به دست می‌دهد.

براساس یافته‌های این پژوهش،هم برآورد بیز تجربی و هم برآورد مرکب را می‌توانیم به طور یکسان برای برآورد نسبت بیکاری در سطح استان‌ها بدون اینکه یکی بر دیگری برتری داشته باشد، به کار ببریم.

2، $ARE = 100\left[\frac{1}{L}\sum_{i=1}^{L}\left|p_i - P_i^{Des}\right| / \left|P_i^{Des}\right|\right]$ که در P_i^{Des} برآورد مبتنی بر طرح است

جدول –۱. برآوردهای نسبتهای بیکاری درسال۱۳۸۰دراستان‌های[1] کشور

استان	طرح	ترکیبی	مرکب	بیزتجربی
اردبیل	۰/۰۹۹۹۶	۰/۱۴۷۸۹	۰/۱۱۷۴۶	**۰/۱۱۲۶۳**
بوشهر	۰/۱۷۰۷۵	۰/۱۰۴۹۱	**۰/۱۷۲۲۰**	۰/۱۷۸۸۰
چهارمحال و بختیاری	۰/۰۹۱۱۱	**۰/۱۰۸۶۴**	۰/۰۶۳۵۱	۰/۰۶۶۵۹
آذربایجان شرقی	۰/۰۸۷۵۲	**۰/۰۸۷۷۱**	۰/۰۸۳۱۷	۰/۰۷۸۵۷
اصفهان	۰/۱۳۸۷۰	۰/۱۱۵۲۳	**۰/۱۴۵۹۸**	۰/۱۵۲۶۷
فارس	۰/۱۵۸۰۰	۰/۱۴۴۲۱	۰/۱۴۰۰۳	**۰/۱۵۲۶۷**
قزوین	۰/۱۳۹۸۷	۰/۰۹۱۷۸	۰/۱۱۱۱۹	**۰/۱۲۳۱۲**
قم	۰/۱۰۵۳۲	۰/۰۷۴۱۶	**۰/۱۰۰۶۸**	۰/۰۷۳۰۸
گیلان	۰/۱۳۴۹۳	۰/۲۰۵۷۵	۰/۱۳۶۰۹	**۰/۱۳۴۲۵**
گلستان	۰/۱۴۸۱۸	۰/۱۷۰۱۲	**۰/۱۳۷۸۷**	۰/۱۳۱۴۳
همدان	۰/۱۰۷۷۴	۰/۱۲۱۶۵	**۰/۱۰۰۰۹**	۰/۰۹۶۰۰
هرمزگان	۰/۲۲۲۹۶	۰/۱۱۱۱۱۳	۰/۲۶۵۱۶	**۰/۲۶۳۴۹**
ایلام	۰/۱۵۰۲۶	۰/۲۴۴۹۵	**۰/۱۵۱۰۸**	۰/۱۴۲۳۲
کرمان	۰/۱۵۲۶۵	۰/۱۲۰۹۸	۰/۱۳۰۸۲	**۰/۱۴۱۳۸**
کرمانشاه	۰/۱۷۵۶۳	۰/۲۶۴۹۸	**۰/۱۵۳۶۷**	۰/۱۵۱۰۸
خراسان	۰/۱۳۶۳۷	۰/۰۹۰۸۴	۰/۱۲۹۷۶	**۰/۱۳۱۰۸**
خوزستان	۰/۱۹۸۰۵	۰/۲۵۷۳۷	۰/۲۲۷۲۸	**۰/۲۲۲۳۱**
کهگیلویه و بویراحمد	۰/۲۰۰۳۸	۰/۲۱۶۸۱	۰/۲۲۸۰۷	**۰/۱۸۲۵۷**
کردستان	۰/۱۲۱۶۹	**۰/۱۲۵۳۸**	۰/۱۳۰۶۵	۰/۱۱۳۳۲

۱. نزدیکترین برآورد به برآورد مبتنی برطرح برجسته شده است.

لرستان	۰/۳۰۴۹۰	۰/۲۶۶۲۹	۰/۲۶۶۷۳	**۰/۲۸۶۱۱**
مرکزی	۰/۱۰۵۲۹	**۰/۱۰۹۲۴**	۰/۰۹۷۳۹	۰/۱۲۲۰۵

ادامه جدول شماره ۱.

مازندران	۰/۱۳۱۱۴	**۰/۱۳۰۲۴**	۰/۱۲۰۲۹	۰/۱۳۷۶۴
سمنان	۰/۱۵۷۰۶	۰/۰۶۸۱۳	۰/۱۳۳۱۵	**۰/۱۴۴۱۱**
سیستان و بلوچستان	۰/۲۸۶۸۳	۰/۱۱۹۸۴	**۰/۲۷۵۲۰**	۰/۲۷۳۷۲
تهران	۰/۱۲۱۸۷	۰/۰۸۴۴۴	۰/۱۱۹۵۲	**۰/۱۲۰۳۷**
آذربایجان غربی	۰/۰۸۷۹۶	۰/۱۲۹۷۶	**۰/۰۸۵۳۵**	۰/۰۸۴۷۹
یزد	۰/۱۲۱۲۳	۰/۰۶۷۷۴	۰/۱۰۹۵۶	**۰/۱۱۸۳۴**
زنجان	۰/۱۰۱۹۸	**۰/۰۹۱۱۵**	۰/۱۲۰۰۹	۰/۰۸۷۸۲
حداقل خطای مطلق[۱]	–	۰/۰۰۰۱۹۰۰	۰/۰۰۰۸۱۵۴	۰/۰۰۰۶۸۲۳
حداکثر خطای مطلق[۲]	–	۰/۱۶۶۹۹۰۰	۰/۰۴۲۱۹۸۰	۰/۰۴۰۵۳۲۴
میانگین قدر مطلق خطا[۲]	–	۰/۰۴۴۶۵	۰/۰۱۴۸۲	۰/۰۱۳۴۷

۲. این مقادیر قبل از گرد کردن برآوردها محاسبه شده‌اند.

منابع

Chand, N., and Alexander, C.H. (1995).Indirect Estimation of Rates and Proportions for Small Areas with Continuous Measurement. *In Proceeding of the Section on Survey Research Methods, American Statistical Association,* 549-554.

Copas, J. B. (1972). Empirical Bayes methods and the repeated use of a standard, *Biometrika*, 59, 349-360.

Ghosh, M., and Rao, J.N.K. (1994). Small Area Estimation: an appraisal (with discussion). *Statistical Science*, 9, 65 - 93.

Marker, D. A. (1999).Organization of Small Area Estimators Using a Generalized Linear Regression Framework. *Journal of Official Statistics*, 15, 1 - 24.

Pfeffermann, D. (1999). Small Area Estimation-Big Developments. *Paper presented at the international association of survey statisticians satellite conference*, Riga, Latvia, 20-21 August 1999.

Purcell, N. J., and Kish, L. (1979).Estimation for Small Domains. *Biometrics*, 35, 365 - 384.

Purcell, N. J., and Kish, L. (1980).Postcensal Estimates for Local Areas (or Domains). *International Statistical Review*, 48, 3 - 18.

SAS Institute Inc.(1990). *SAS/IML Software: Reference, Version 6*, First Edition, Cary, Nc: SAS Institute Inc.

Schaible, W. L. (1978).Choosing Weight for Composite Estimators for Small Area Statistics. *In Proceedings of the Section on Survey Research Methods, American Statistical Association*, 741-746.

Schaible, W.L., ed. (1995), *Lecture Notes in Statistics: Indirect Estimators in U.S. Federal Programs,* New York: Springer.

Thomsen, I., and Holmoy A.M.K. (1998).Combining data from surveys and administrative record system: The Norwegian experience. *International Statistical Review*, 66, 201-221.

U.S. National Center for Health Statistics. (1968). *Synthetic State Estimations of Disability.* Washington, D.C., U.S. Public Health Service.