برآوردگرهای کالیبره[1]، روشی مناسب برای برآورد شاخص‌های عمده نیروی کار

فاطمه هرندی[2]

چکیده

برآوردگرهای کالیبره، خانواده‌ای از برآوردگرها هستند که از اطلاعات کمکی برای تصحیح وزن‌ها به منظور بهبود برآورد استفاده می‌کنند. وزن‌های تصحیح‌شده که تحت عنوان وزن‌های کالیبره نامیده می‌شوند، براساس یک معیار فاصله‌ی مشخص، تا حد امکان به وزن‌های معمول در نمونه‌گیری ـ عکس احتمال انتخاب واحدها ـ نزدیک بوده و برآورد مجموع متغیر کمکی حاصل از کاربرد این وزن‌ها (مجموع نمونه‌ای مقادیر موزون متغیر کمکی)، برابر مقدار معلوم مجموع آن متغیر در جامعه است.

هر معیاری برای فاصله از طریق معادلات کالیبره به یک سیستم وزن‌دهی خاص و در نتیجه به یک برآوردگر کالیبره منتهی می‌شود. برآوردگرهای حاصل از معیارهای متفاوت، خانواده‌ی برآوردگرهای کالیبره را تشکیل می‌دهند. ویژگی عددی وزن‌ها و سادگی محاسبه‌ی آنها مهمترین نکاتی است که در انتخاب برآوردگر مناسب باید به آن توجه شود.

---

[1] *Calibration Estimators*

[2] عضو هیئت علمی پژوهشکده آمار

مقدمه

یکی از مباحث مهم در آمار، چگونگی تهیه برآوردهای مناسب براساس اطلاعات نمونه است. روش‌های برآورد مطرح در تئوری آمار، بر مبنای نمونه‌ای از یک جامعه نامحدود و غالباً با شکل تابع توزیع معلوم تعریف می‌شوند. در تئوری آمارگیری نمونه‌ای نیز یکی از مباحث اصلی، تعیین روش بهینه برای برآورد پارامترهای موردنظر در یک جامعه محدود براساس اطلاعات حاصل از نمونه است که به دلیل تنوع صفات مورد بررسی در یک آمارگیری، اصولاً بحثی در مورد شکل توزیع متغیرهای مورد بررسی مطرح نمی‌شود و این متغیرها آزاد از توزیع یا آزاد از مدل، فرض می‌شوند. این ویژگی در عمل به استفاده از روش‌های ساده برآورد در تئوری آمارگیری نمونه‌ای منجر می‌شود به نحوی که برای انواع توزیع فراوانی قابل کاربرد هستند.

روش‌های برآورد در آمارگیری نمونه‌ای را به دو دسته کلی می‌توان تقسیم نمود:

- روش‌هایی که صرفاً از داده‌های حاصل از آمارگیری در مورد متغیر موردنظر برای ساخت برآورد استفاده می‌کنند.
- روش‌هایی که از اطلاعات کمکی یک یا چندمتغیره برای بهبود برآورد پارامتر مورد مطالعه استفاده می‌کنند.

در صورت وجود اطلاعات کمکی همبسته با متغیر موردنظر، استفاده از روش‌های دسته دوم ارجحیت دارد زیرا به بهبود برآورد منجر می‌شود.

از جمله روش‌های دسته دوم، روش برآورد رگرسیونی است. برآوردگر رگرسیونی مجموع یا میانگین صفت موردنظر، براساس اطلاعات متغیرهای کمکی و از طریق ساخت معادله رگرسیون به دست می‌آید؛ اما شیوه‌ی دیگری برای به دست آوردن برآورد رگرسیونی وجود دارد که از طریق اصلاح وزن‌های برآورد بر مبنای اطلاعات متغیرهای کمکی می‌باشد. در این شیوه که بخصوص برای برآورد مجموع جامعه‌ی محدود کاربرد دارد، به هر داده حاصل از آمارگیری، وزنی مرتبط با اطلاعات کمکی موجود منتسب می‌شود.

به بیان دیگر، وزن‌های معمول در نمونه‌گیری که برای هر واحد، عکس احتمال انتخاب آن است، با استفاده از اطلاعات مربوط به متغیرهای کمکی به نحوی تصحیح می‌شوند که مجموع موزون مقادیر متغیر کمکی برابر مقدار معلوم آن متغیر در جامعه باشد. این وزنها، وزن‌های کالیبره شده و برآوردگرهای حاصل از کاربرد این وزنها، برآوردگرهای کالیبره نامیده می‌شوند.

ویژگی مذکور این وزنها، برای آمارشناسان قابل توجه و حائز اهیمت است زیرا در صورت وجود همبستگی قوی بین متغیرهای کمکی و متغیر مورد مطالعه، می‌توان انتظار داشت، وزن‌هایی که برای متغیر کمکی مطلوب هستند، برای متغیر مورد مطالعه هم مطلوب باشند.

علت اطلاق عنوان «کالیبره» به این برآوردگرها به مفهوم کالیبره برمی‌گردد. کالیبره از ریشه «کالب» یا «قالب» به معنی قالب زدن یک ابزار اندازه‌گیری است[1]. برای درک مفهوم کالیبره کردن، آزمایش زیر را در نظر بگیرید:

مقادیر صحیح و معلوم $x_1,\dots,x_n$

مقادیر مشاهده شده دارای اختلال براساس مقیاس به کار رفته در یک ابزار اندازه‌گیری $Y_1,\dots,Y_n$

مسئله‌ی کالیبره کردن، استفاده از این داده‌ها برای برآورد مقدار صحیح و نامشخص x_0 براساس مشاهده‌ی Y_0 در آینده است.

یک کاربرد ویژه این موضوع، کالیبره کردن یک دماسنج صنعتی است به این ترتیب که در شرایط آزمایشگاهی به دقت کنترل شده، n مشاهده دقیق از حرارت محیط (x) تعیین و براساس آن، دماسنج مدرج می‌شود؛ سپس به کمک این مشاهدات، یک دماسنج صنعتی به نحوی کالیبره می‌شود که هنگام استفاده از آن در شرایط معمولی (خارج از محیط آزمایشگاه)، بتوان حرارت واقعی محیط (x_0) را از درجه‌ای که دماسنج نشان می‌دهد (y_0) به دست آورد. پس، انجام آزمایش در شرایط آزمایشگاهی، برآورد توزیع شرطی Y به شرط $X = x$ را براساس n مشاهده از x و Y امکان‌پذیر می‌سازد. اما مسئله‌ی کالیبره کردن برعکس است و در واقع جستجو برای برآورد مقدار واقعی x_0 به شرط داشتن $Y_0 = y_0$ است که عکس رگرسیون می‌باشد. هر چند نحوه‌ی کالیبره

---

[1] در جلد دوازدهم لغت نامه دهخدا، کالب به این مفهوم آمده است.

کردن کمی پیچیده است، اما با داشتن اطلاعات اضافی و مفروضاتی مشخص، پیچیدگی آن کاهش می‌یابد که چون خارج از موضوع بحث است در این‌جا مطرح نمی‌شود.

نحوه‌ی به دست آوردن برآوردگر رگرسیونی عمومی[1] به وسیله کالیبره کردن

فرض کنید

$$U = \{1, \ldots, k, \ldots, N\}$$

مجموعه عناصر جامعه موردنظر است. از این جامعه نمونه احتمالی s با طرح نمونه‌گیری $P(\cdot)$ انتخاب می‌شود. احتمال انتخاب نمونه s ، $P(s)$ است و

احتمال انتخاب واحد k ام در نمونه s : $\pi_k = P_r(k \in s) > 0$

مقدار متغیر موردنظر برای k امین واحد جامعه : y_k

بردار متغیرهای کمکی مربوط به y_k : $\boldsymbol{x}_k = (x_{k1}, \ldots, x_{kj}, \ldots, x_{kJ})'$

مقادیر y_k و $\boldsymbol{x}_k$ برای تمامی $k \in s$ مشاهده می‌شود. همچنین

مجموع معلوم x در جامعه : $\boldsymbol{t}_x = \sum_{k \in U} \boldsymbol{x}_k$

مقدار دقیق $\boldsymbol{t}_x$ از منابع دیگری غیر از آمارگیری (سرشماری، داده‌های ثبتی، ...) به دست آمده است.

هدف، برآورد مجموع y در جامعه یعنی:

$$t_y = \sum_{k \in U} y_k$$

است.

---

[1] ساخت معادله، مقید به استقلال متغیرهای مستقل نیست.

از طرفی با توجه به فرمول برآوردگر هورویتز ـ تامپسون، برآورد مجموع y برابر است با:

$$\hat{t}_{y\pi}=\sum_{k\in s}y_k/\pi_k=\sum_{k\in s}d_k y_k$$

در رابطه‌ی فوق، π_k ، احتمال انتخاب واحد k ام و $d_k=\dfrac{1}{\pi_k}$ ، وزن واحد k ام نمونه است. به عبارت دیگر، d_k نشان می‌دهد واحد k ام معرف چند واحد جامعه است.

آنچه که می‌خواهیم انجام دهیم این است که وزن‌های d_k را با استفاده از اطلاعات کمکی به نحوی تصحیح کنیم که وزن‌های حاصل از نمونه s (w_{ks})، براساس یک مقیاس فاصله‌ی مشخص تا حد امکان به d_k نزدیک بوده و در معادله کالیبره‌ی زیر صدق کند.

$$\sum_{k\in S}w_{ks}x_k=t_x \qquad (۱)$$

اگر به شکل معادلـه‌ی فوق توجه کنید و توضیحاتی را که در مورد مفهوم کالیبره کردن ذکر شد به یاد آورید، متوجه علت اطلاق عنوان برآوردگر کالیبره به برآوردهای حاصل از وزن‌های کالیبره شده (w_{ks})، می‌شوید. در واقع، در معادله‌ی فوق t_x مشخص است و براساس آن می‌خواهیم w_{ks} ها را به دست آوریم.

همان‌گونه که اشاره شد، وزن‌های w_{ks} باید براساس یک معیار فاصله‌ی مشخص تا حد امکان به وزن‌های d_k نزدیک باشند. علت ملحوظ نمودن این قید این است که چون وزن‌های d_k خواص مطلوب تولید برآوردهای ناریب را دارند، آمارشناسان تمایل دارند وزن‌های تعدیل شده هم تا حد امکان به این وزنها نزدیک باشند تا خواص مطلوب آنها را حفظ کنند.

ساده‌ترین معیار فاصله‌ی متوسط را که به برآوردگرهای کالیبره‌ی ساده منجر می‌شود می‌توان به صورت زیر تعریف نمود.

$$E_p\left\{\sum_{k\in s}\left(w_{ks}-d_k\right)^{۲}\Big/d_k q_k\right\}$$

در رابطه‌ی فوق، $E_p(\cdot)$، امید ریاضی مربوط به طرح نمونه‌گیری و q_k ها مقادیر مثبت، معلوم و مستقل از d_k هستند که برای عمومیت بیشتر معادله به آن اضافه شده است و در بسیاری از موارد برای کلیه‌ی مقادیر k ، برابر ۱ فرض می‌شوند. با می‌نیمم کردن فاصله‌ی متوسط فوق به شرط معادله (۱) و برای هر نمونه s، مقدار وزن‌های کالیبره به صورت زیر به دست می‌آید.

$$w_{ks} = d_k\left(1 + q_k x_k' \boldsymbol{\lambda}\right) \qquad (۲)$$

در معادله‌ی فوق، $\boldsymbol{\lambda}$ بردار ضرایب لاگرانژ می‌باشد و مقدار آن برابر است با

$$\boldsymbol{\lambda} = \frac{\boldsymbol{t}_x - \hat{\boldsymbol{t}}_{x\pi}}{\sum_{k \in s} d_k q_k \boldsymbol{x}_k \boldsymbol{x}_k'}$$

پس از محاسبه‌ی $\boldsymbol{\lambda}$ و در نتیجه‌ی مشخص شدن مقادیر w_{ks}، برآوردگر مجموع y به صورت زیر به دست می‌آید.

$$\hat{t}_{yreg} = \sum_{k \in s} w_{ks} y_k = \hat{t}_{y\pi} + \left(\boldsymbol{t}_x - \hat{\boldsymbol{t}}_{x\pi}\right)\hat{\boldsymbol{B}}_s \qquad (۳)$$

در رابطه‌ی فوق، $\hat{\boldsymbol{B}}_s$ برآوردگر وزنی ضریب رگرسیون چندگانه و $\hat{\boldsymbol{t}}_{x\pi}$ برآوردگر هورویتز ـ تامپسون بردار $\boldsymbol{x}$ است که برابر است با :

$$\hat{\boldsymbol{t}}_{x\pi} = \sum_{k \in s} d_k \boldsymbol{x}_k$$

همان‌طور که ملاحظه می‌کنید $\hat{t}_{yreg}$ مشابه یک برآوردگر رگرسیونی است که پس از محاسبه‌ی وزن‌های w_{ks}، به سادگی قابل محاسبه می‌باشد و در واقع از مجموع وزنی مشاهدات y_k از نمونه به دست می‌آید. مورد اخیر محاسبه‌ی برآورد واریانس $\hat{t}_{yreg}$ را ساده می‌نماید.

روش عمومی به دست آوردن برآوردگرهای کالیبره

در قسمت قبل ملاحظه نمودید که برای رسیدن به $\hat{t}_{yreg}$ از $(w_{ks} \quad d_k)^2 / d_k h_k^t$ به طور اختیاری برای فاصله بین وزن اصلی d_k و وزن جدید w_{ks} استفاده شد. بنابراین طبیعی است که بتوان معیارهای دیگری را نیز برای فاصله در نظر گرفت، البته مشروط

بر اینکه معیارهای انتخابی بتوانند از طریق معادلات کالیبره، به یک سیستم وزن‌دهی مناسب و در نتیجه به یک برآوردگر مفید منتهی شوند.

فرض کنید:

تابع فاصله برای واحد k ام نمونه: $G_k(w,d)$

این تابع برای $d>0$ ، غیر منفی، نسبت به w مشتق‌پذیر و اکیداً محدب[1] است.

این تابع روی فاصله $D_k(d)$ (شامل d) تعریف می‌شود و $G_k(d,d)=0$

همچنین :

$$g_k(w,d)=\frac{\partial G_k(w,d)}{\partial w}$$

پیوسته است و تابعی اکیداً افزایشی از w است و $g_k(d,d)=0$.

با مفروضات فوق، فاصله‌ی متوسط به صورت زیر تعریف می‌شود.

$$E_p\left\{\sum_{k\in s} G_k(w_{ks},d_k)\right\}$$

برای می‌نیمم کردن این مقیاس به شرط معادله (۱) و برای هر نمونه s، کافی است w_{ks} هایی را بیابیم که $\sum_{k\in s} G_k(w_{ks},d_k)$ را به شرط معادله (۱) و برای هر نمونه s ، می‌نیمم کند. با مشتق‌گیری از G نسبت به w داریم :

$$g_k(w_{ks},d_k)-\mathbf{x}'_k\boldsymbol{\lambda}=0$$

با حل این معادله مقدار $\boldsymbol{\lambda}$ به دست می‌آید. اگر این معادله راه‌حل داشته باشد، به دلیل مفروضات ذکر شده، این راه‌حل یکتا است. مقدار وزن‌های کالیبره شده نیز به صورت زیر می‌باشد.

$$w_{ks}=d_k F_k(\mathbf{x}'_k\boldsymbol{\lambda})$$

در رابطه‌ی بالا $d_k F_k(\cdot)$ نگاشت معکوس $g_k(\,.\,,d_k)$ و $\boldsymbol{\lambda}$ بردار ضرایب لاگرانژ است و با توجه به مفروضات فوق:

$$F_k(0)=1 \qquad , \qquad F'_k(0)>0$$

---

[1] مشتق دوم تابع مثبت است و امید ریاضی آن بزرگتر یا مساوی تابع امید ریاضی است یعنی $E(G_k)\ge G_k(E)$

$F_k'(\circ)$ در واقع نقش q_k در مقیاس فاصله قبلی را دارد. پس $F_k'(\circ)=q_k$:
معادلات کالیبره‌ی‌موردنیاز برای‌تعیین بردار ضرایب لاگرانژ، $\boldsymbol{\lambda}=(\lambda_1,\ldots,\lambda_j,\ldots,\lambda_J)$، با توجه به معادله (۱) عبارتست از :

$$t_x=\sum_{k\in s}w_{ks}x_k=\sum_{k\in s}d_kF_k(x_k'\lambda)x_k \qquad (۴)$$

در نتیجه :

$$\sum_{k\in s}d_k\{F_k(x_k'\lambda)-1\}x_k=t_x-\hat{t}_{x\pi} \qquad (۵)$$

قسمت راست معادلـــه (۵) برای هر نمونه s مشخص است و با حل معادله، $\boldsymbol{\lambda}$ به دست می‌آید. براساس مقدار $\boldsymbol{\lambda}$، وزن‌های کالیبره (w_{ks}) محاسبه می‌شود و در نتیجه می‌توان براساس مقادیر w_{ks}، برآوردگر کالیبره t_y را به صورت زیر به دست آورد.

$$\hat{t}_{yw}=\sum_{k\in s}w_{ks}y_k=\sum_{k\in s}d_kF_k(x_k'\lambda)y_k \qquad (۶)$$

ثابت می‌شود که اگر همبستگی قوی بین x و y وجود داشته باشد، این برآوردگر بسیار نزدیک به مقدار واقعی مجموع y در جامعه یعنی $t_y=\sum_{k\in U}y_k$ خواهد بود.

برای اثبات این مدعا فرض کنید $y_k=x_k'\alpha$ برای تمامی مقادیر k برقرار باشد. α بردار ضرایب ثابت است. این رابطه به این معنی است که y کاملاً توسط x بیان می‌شود. بنابر این با توجـه به معادله‌ی (۴)، برای هر نمونه s داریم:

$$\hat{t}_{yw}=t_y$$

و در نتیجه واریانس برآوردهای حاصل از نمونه‌های مختلف، صفر است.

یک ویژگی دیگر برآوردگرهای کالیبره این است که خواص مطلوب $\hat{t}_{y\pi}$ را حداقل به صورت مجانبی حفظ می‌کنند.

در یک آمارگیری نمونه‌ای، روش کار برای دستیابی به این برآوردها به شرح زیر است :

تابع $F_k(\cdot)$ مشخص می‌شود. با انتخاب این تابع و در دست بودن نتایج نمونه s، معادله (۵) حل می‌شود تا $\boldsymbol{\lambda}$ به دست آید. ممکن است برای حل معادله مذکور، تکرار لازم باشد. پس از مشخص شدن $\boldsymbol{\lambda}$، برآوردگر کالیبره با فرمول (۶) محاسبه می‌شود.

در انتخاب معیار مناسب برای فاصله چند نکته باید مدنظر قرار گیرد. اولاً باید معیاری انتخاب شود که برای آن، معادله (۵) جواب داشته باشد. ثانیاً w_{ks}هایی که براساس معیار انتخابی تولید می‌شوند، نباید مقادیر غیرقابل قبول داشته باشند و آخرین نکته، سادگی محاسبه برآوردگر است.

آخرین نکته در واقع به این موضوع توجه می‌دهد که چنانچه معیارهایی برای فاصله وجود داشته باشند که ویژگی‌های اول و دوم را دارا باشند، معیاری باید انتخاب شود که محاسبه برآوردگر را ساده‌تر می‌سازد، زیرا ثابت می‌شود تمامی برآوردگرهای $\hat{t}_{yw}$ که با معیارهای متفاوت برای فاصله تولید می‌شوند به طور مجانبی معادل برآوردگر رگرسیونی $\hat{t}_{y_{reg}}$ هستند و انتخاب‌های متفاوت تنها تأثیر ملایمی روی واریانس برآوردگر می‌گذارد.

کاربرد

در این قسمت یکی از پرکاربردترین برآوردگرهای کالیبره معرفی می‌شود.

فرض کنید:

$\boldsymbol{x}_k = x_k$: یک عدد مثبت

$$q_k = \frac{1}{x_k}$$

به بیان دیگر برای هر مشاهده y_k فقط از مقدار یک متغیر کمکی (x_k) استفاده می‌شود. در نتیجه

$$\boldsymbol{x}_k'\lambda = x_k\lambda$$

ثابت می‌شود فارغ از نوع مقیاس فاصله‌ی مورد استفاده :

$$w_k = d_k \frac{t_x}{\hat{t}_{x\pi}}$$

$$\hat{t}_{yw} = \hat{t}_{y\pi} \frac{t_x}{\hat{t}_{x\pi}}$$

اثبات : با توجه به مفروضات فوق

$F_k(x'_k\lambda) = F(\lambda) =$ یک عدد ثابت

و با توجه به رابطه (۴)

$$t_x = \sum_{k \in s} d_k x_k F(\lambda) = \hat{t}_{x\pi} F(\lambda)$$

در نتیجه

$$F(\lambda) = \frac{t_x}{\hat{t}_{x\pi}}$$

بنابراین

$$w_k = d_k \frac{t_x}{\hat{t}_{x\pi}}$$

و

$$\begin{aligned} \hat{t}_{yw} &= \sum_{k \in s} d_k F(\lambda) y_k \\ &= \frac{t_x}{\hat{t}_{x\pi}} \sum_{k \in s} d_k y_k \\ &= \hat{t}_{y\pi} \frac{t_x}{\hat{t}_{x\pi}} \end{aligned}$$

بنابراین با فروض تعیین شده در این قسمت، انتخاب تابع فاصله تأثیری روی برآوردگر حاصل ندارد[1] و فارغ از تابع فاصله‌ی انتخاب شده، برآوردگر کالیبره و برآوردگر نسبتی یکی می‌شوند. چون همان‌گونه که توجه کرده‌اید، برآوردگـر فوق، یک برآوردگر نسبتی

[1] *البته این یک اصل کلی نیست و انتخاب توابع فاصله متفاوت به برآوردگرهای متفاوتی منتهی می‌شود،اما می‌توان انتظار داشت که در صورت وجود نمونه‌ای با حجم مناسب (متوسط تا بزرگ) ، تفاوت برآوردگرها زیاد نباشد.*

است. به این برآوردگر کالیبره، برآوردگر کالیبره ساده گفته می‌شود.
همان‌گونه که قبلاً ذکر شد بین وزن‌های کالیبره و مجموع x در جامعه رابطه زیر برقرار است :

$$t_x = \sum_{k\in s} w_k x_k$$

بنابراین برآوردگر کالیبره‌ی ساده، برآوردگر خوبی است اگر:

- x و y همبستگی خطی داشته باشند و
- t_x حاصل از منابع خارج از طرح، معتبر و قابل اعتماد باشد.

از این برآوردگر درطرح پیشنهادی برای آمارگیری نیروی کار (هرندی و دیگران، ۱۳۸۱) استفاده شد، به این ترتیب که برای k امین فرد نمونه در زیرگروه جمعیتی i ام، وزن‌های کالیبره به صورت زیر تعریف گردید.

$$w_{ki} = d_k \frac{t_i}{\hat{t}_i}$$

d_k ، وزن مشاهده k ام براساس طرح نمونه‌گیری است که برای بی‌پاسخی کامل نیز تصحیح شده است.

t_i ، نشان‌دهنده جمعیت خانوارهای معمولی ساکن در زیرگروه جمعیتی i ام براساس پیش‌بینی‌های جمعیتی است. در واقع برای دقت بیشتر، به جای استفاده از رقم کل جمعیت درتصحیح وزنها، از ارقام زیرگروه‌های جمعیتی برحسب جنس، سن و شهری/ روستایی استفاده می‌شود. به این ترتیب که ۲ زیرگروه بر حسب جنس، ۳ زیرگروه برحسب سن و ۲ زیرگروه بر حسب شهری/ روستایی و در مجموع ۱۲ زیرگروه جمعیتی تشکیل می‌شود. جمعیت این زیرگروهها با روش‌های جمعیت‌شناختی و با دقت بالا محاسبه می‌شود. بنابراین $\boldsymbol{t}$، بردار جمعیت زیرگروه‌های جمعیتی براساس‌پیش‌بینی‌های جمعیتی است که عبارتست از:

$$\boldsymbol{t} = \left(t_{۱۱۱}, t_{۱۱۲}, t_{۱۱۳}, \ldots, t_{۲۲۱}, t_{۲۲۲}, t_{۲۲۳}\right)'$$

هر عضو این بردار، جمعیت یک زیرگروه از زیرگروه‌های ۱۲ گانه‌ی مذکور است و برای هر عضو بردار، اندیس اول، شهری/ روستایی، اندیس دوم، جنس و اندیس سوم،

گروه سنی را نشان می‌دهد. برای مثال، t_{111}عبارتست از جمعیت پیش‌بینی شده مردان گروه سنی ۱ در نقاط شهری.

$\hat{t}$ هم بردار جمعیت برآورد شده زیرگروه‌های جمعیتی براساس نتایج آمارگیری است و برابر است با :

$$\hat{t} = \left(\hat{t}_{111}, \hat{t}_{112}, \hat{t}_{113}, \ldots, \hat{t}_{221}, \hat{t}_{222}, \hat{t}_{223}\right)'$$

در اینجا هم برای مثال $\hat{t}_{221}$ ، جمعیت برآورد شده زنان گروه سنی ۱ در نقاط روستایی براساس نتایج آمارگیری را نشان می‌دهد.

پس از تعیین مقادیر t_i و محاسبه $\hat{t}_i$ برای هر زیرگروه جمعیتی، می‌توان وزن‌های کالیبره را محاسبه نمود.

$$w_{ki} = d_k \frac{t_i}{\hat{t}_i}$$

w_{ki} در واقع نشان می‌دهد که k امین فرد نمونه‌ی معرف چند نفر در زیرگروه جمعیتی i ام جامعه است زیرا برای هر فرد بسته به اینکه در چه زیرگروهی قرار می‌گیرد از وزن مربوط به همان زیرگروه استفاده می‌شود.

براساس مقادیر w_{ki}، برآوردهای کالیبره تعداد و نسبت را می‌توان به صورت زیر محاسبه نمود.

برآورد تعداد $\hat{T} = \sum_{k \in C} w_{ki}$

برآورد نسبت $\hat{P} = \sum_{k \in C} w_{ki} \Big/ \sum_{k \in D} w_{ki}$

که در آن i ، اندیس زیرگروه جمعیتی ، C مجموعه افراد دارای صفت موردنظر و D مجموعه افراد واجد شرایط را نشان می‌دهد. برای مثال، برای برآورد تعداد افراد شاغل، کافی است وزن‌های افراد شاغل نمونه با هم جمع شوند. همچنین برای برآورد نسبت بیکاری، کافی است حاصل تقسیم مجموع وزن‌های افراد بیکار درنمونه به مجموع وزن‌های افراد فعال در نمونه محاسبه شود.

نتیجه‌گیری

به طور کلی توانایی برآوردگر کالیبره این است که سازگاری برآوردهای حاصل از آمارگیری را با اطلاعات مربوط به سایر منابع که فرض می‌شود معتبرتر از آمارگیری هستند، تضمین می‌کند. البته در صورتی که اطلاعات سایر منابع، موثق نباشد یا نسبت به نتایج طرح، از اعتبار کمتری برخوردار باشد، کالیبره کردن اثر منفی خواهد داشت و نباید استفاده شود. در حال حاضر، کالیبره کردن برآوردها که یکی از شاخص‌های مطلوبیت یک آمارگیری تلقی می‌شود، در بسیاری از کشورها مرسوم است و معمول‌ترین آن، برآوردگر کالیبره‌ی ساده است. شایان ذکر است که فقط تعداد معدودی از کشورهای پیشرفته آماری هستند که از روش‌های پیچیده‌تر کالیبره کردن برآورد استفاده می‌کنند. مثلاً کانادا از برآوردگر کالیبره‌ی مبتنی بر اطلاعات کمکی چندمتغیره دارای ارتباط غیرخطی با متغیر مورد بررسی استفاده می‌کند. در فرانسه نیز معادلات کالیبره را مستقیماً در طرح نمونه‌گیری وارد می‌کنند که تحت عنوان نمونه‌گیری متعادل شده[1] نامیده می‌شود و روشی کاملاً جهت‌گیری شده است که ظاهراً در حال حاضر فقط فرانسه از آن استفاده می‌کند.

[1] *Balanced Sampling*

مراجع:

[۱] هرندی، فاطمه و دیگران، *طرح پیشنهادی برای آمارگیری نیروی کار*، تهران، پژوهشکده آمار، ۱۳۸۱.

[۲] Deville, J. C. and Sarndal, C. E. (1992). "Calibration Estimators in Survey Sampling", *Journal of the American Statistical Association*, **87**, 376-382.

[۳] Kitsos, C. P. 1999, *Encyclopedia of Statistical Sciences*, Update Volume 3, John Wiley & Sons.

[۴] Verma, V. 2000, *Sampling Methods*, United Nation, SIAP.