

## کاربردی از نمونه‌گیری برشی

مریم پورمحمود\* و حمیدرضا نواب‌پور

دانشگاه علامه طباطبائی

**چکیده.** برای به دست آوردن خصوصیات و ویژگی‌های کلی یک جامعه باید اطلاعاتی در مورد اعضای آن جامعه در دست باشد، یک روش جمع‌آوری این اطلاعات آمارگیری نمونه‌ای است. بر اساس نحوه‌ی انتخاب نمونه دو رهیافت در آمارگیری نمونه‌ای وجود دارد، رهیافت احتمالاتی و رهیافت نااحتمالاتی. در رهیافت نمونه‌گیری نااحتمالاتی برآوردها اریب هستند، لذا نتیجه‌ها باید در کلاس برآوردهای اریب بررسی شوند، ولی به سبب سرعت و هزینه‌ی کم برخی از این روش‌ها در بسیاری از آمارگیری‌ها مورد توجه قرار گرفته‌اند. یکی از روش‌های نمونه‌گیری نااحتمالاتی نمونه‌گیری برشی است. این شیوه‌ی گزینش نمونه به دلیل ویژگی‌های مناسبی که دارد در بسیاری از آمارگیری‌های اقتصادی اجتماعی مثل آمارگیری از شاخص قیمت‌ها و آمارگیری از کارگاه‌های صنعتی به کار گرفته شده است. در این مقاله ابتدا انواع نمونه‌گیری برشی معرفی می‌شود و سپس از نمونه‌گیری برشی نوع سوم برای برآورد میانگین ارزش افزوده‌ی کارگاه‌های صنعتی دارای ۱۰ کارکن و بیش‌تر در سال ۱۳۸۵ استفاده می‌شود. در این مطالعه عملکرد این نوع نمونه‌گیری با رهیافت مدل‌یار در برآورد میانگین آماره‌ی مذکور مقایسه می‌شود. نتیجه‌ی مطالعه نشان می‌دهد که رهیافت مدل‌یار تحت شرایطی برآوردهای دقیق‌تری ارائه می‌دهد. در نهایت نتایج به دست آمده بر اساس رهیافت مدل‌یار با نتایج برآورد نسبتی روش نمونه‌گیری تصادفی ساده مقایسه می‌شود.

واژگان کلیدی: نمونه‌گیری احتمالاتی؛ نمونه‌گیری نااحتمالاتی؛ نمونه‌گیری برشی؛ آستانه‌ی برش؛

رهیافت حذف؛ رهیافت مدل‌یار؛ اندازه‌ی نمونه.

دریافت: ۱۳۸۸/۲/۲۲، پذیرش: ۱۳۸۸/۸/۲۰

\* نویسنده‌ی عهده‌دار مکاتبات

## ۱- مقدمه

برای تولید اطلاعات در مورد عناصر جامعه از روش‌های مختلف تولید آمار استفاده می‌شود. تولید آمار به سه روش ممکن است انجام شود، تولید آمار به کمک ثبت داده‌ها، سرشماری و آمارگیری نمونه‌ای.

تولید آمار از طریق آمارگیری نمونه‌ای در صورتی معتبر و قابل قبول است که تمام مراحل بررسی به نحوی معقول و منطقی و بر اساس اصول و ملاک‌های علمی و به دور از هر گونه جهت‌گیری و یا سلیقه‌ی شخصی انجام گرفته باشد. بر اساس گفته‌ی سارندال [۶] نمونه‌گیری به دو شیوه‌ی نمونه‌گیری احتمالاتی (همه‌ی عناصر جامعه‌ی هدف شانس مثبت برای گزینش شدن را دارند) و نمونه‌گیری نااحتمالاتی (برخی عناصر جامعه‌ی هدف شانس گزینش شدن را در نمونه ندارند) صورت می‌گیرد. در واقع در عمل اجرای نمونه‌گیری‌های نااحتمالاتی در مقایسه با نمونه‌گیری‌های احتمالاتی بسیار کم‌هزینه‌تر و سریع‌تر هستند، ولی در نهایت نتایج آریب به دست می‌دهند. یکی از روش‌های نمونه‌گیری نااحتمالاتی نمونه‌گیری برشی است. در این روش بخشی از عناصر جامعه با توجه به درجه‌ی اهمیت آن‌ها (اطلاعات کمی که دارند) مورد توجه قرار نمی‌گیرند. به عبارت دیگر از چارچوب نمونه‌گیری حذف می‌شوند، لذا عناصر این بخش از جامعه شانس برای گزینش شدن در نمونه ندارند.

در ادامه در بخش ۲، نمونه‌گیری برشی و انواع آن معرفی می‌شود و نحوه‌ی تعیین اندازه‌ی نمونه و آستانه‌ی برش مورد بحث قرار می‌گیرد. در بخش ۳، از طرح نمونه‌گیری برشی نوع سوم برای برآورد میانگین ارزش افزوده‌ی کارگاه‌های صنعتی در سال ۱۳۸۵ استفاده می‌شود.

## ۲- نمونه‌گیری برشی و انواع آن

در صورت اولیه‌ی این روش برخلاف روش‌های معمول نمونه‌گیری که در آن‌ها فرض می‌شود تمام اعضای جامعه شانس مثبتی برای حضور در نمونه دارند، بخشی از جامعه به صورت آگاهانه از مطالعه حذف می‌شوند (قسمت برش داده‌شده) و هیچ شانس حضوری

.....گزیده‌مطالب آماری، سال ۲۰، شماره‌ی ۱، بهار و تابستان ۱۳۸۸، صص ۹۵-۱۲۲.....

در نمونه ندارند. در بسیاری از موارد اطلاعات با ارزشی از قبیل نظر کارشناسان و افراد با تجربه در رابطه با صفت مورد نظر، چگونگی توزیع آن در جامعه و اطلاعات بررسی‌های پیشین و نیز اطلاعات سرشماری‌های قبلی در دست است به طوری که آمارگیر را ترغیب می‌کند تا بخشی از جامعه‌ی مورد نظر را که در مقدار کل یا میانگین صفت مورد بررسی در جامعه تأثیر کمی دارند از بررسی حذف کرده و توجه خود را به آن بخش از جامعه معطوف کند که معتقد است در مقدار پارامتر تأثیر قابل توجهی دارد.

بر اساس گفته‌ی ناب [۵] تجربه نشان داده است اگر توزیع صفت مورد بررسی  $Y$  در جامعه‌ای که دارای چولگی زیاد باشد، و برای واحدهای جامعه با مقادیر کوچک  $Y$  چارچوب قابل اعتمادی در دست نباشد یا ارائه‌ی چنین چارچوبی مستلزم صرف هزینه و دقت زیاد باشد. در این موارد استفاده از روش نمونه‌گیری برشی علاوه بر صرفه‌جویی در هزینه‌های مالی، زمانی و انسانی برآوردهای قابل اعتماد و نسبتاً دقیقی ارائه می‌دهد. به عبارت دیگر در بیش‌تر مواقع در صورت استفاده از طرح‌های نمونه‌گیری برشی آریبی نتیجه‌های به دست آمده در مقایسه با صرفه‌جویی‌های صورت‌گرفته قابل چشم‌پوشی می‌باشد. از دیگر عامل‌های استفاده از نمونه‌گیری برشی آن است که گاهی اوقات به دلیل شرایط اقتصادی، قابلیت دسترسی، امکانات و ... مجبور به انتخاب نمونه در مکان‌هایی هستیم که آمارگیر و سایر شرایط لازم برای آمارگیری فراهم است. در چنین شرایطی استفاده از نمونه‌گیری برشی اجتناب‌ناپذیر است (یکی از منابع مهم مقاله‌ی هان و همکاران [۳] می‌باشد). در نمونه‌گیری برشی بسته به نحوه‌ی برخورد با اعضای جامعه با مقدارهای کوچک  $Y$  یا واحدهایی از جامعه که دسترسی به آن‌ها امکان‌پذیر نیست می‌توان چهار نوع روش نمونه‌گیری برشی را تشخیص داد، که در ادامه به توضیح این روش‌ها پرداخته می‌شود. لازم به توضیح است که در هر یک از این روش‌ها از نتایج به دست آمده توسط حیدراقلو [۴] و بی و همکاران [۲] استفاده شده است.

### ۱-۲- نمونه‌گیری برشی نوع اول (سرشماری-حذف)

در این نوع روش نمونه‌گیری برشی که از متداول‌ترین روش‌های نمونه‌گیری برشی موجود است و در اکثر مطالعه‌های اقتصادی، کارگاهی و ... دارای کاربرد می‌باشد، جامعه‌ی  $N$

..... گزیده‌مطالب آماری، سال ۲۰، شماره ۱، بهار و تابستان ۱۳۸۸، صص ۹۵-۱۲۲ .....

عضوی  $U = \{1, 2, \dots, N\}$  به دو ناحیه  $U_C$  و  $U_E$  تقسیم می‌شود که در آن  $U_C$  بخشی از اعضای جامعه  $U$  را نشان می‌دهد که به طور کامل در آمارگیری شرکت داده می‌شوند. برای هر عضو  $j \in U_C$  داریم  $\pi_j = 1$ . همچنین  $U_E$  (ناحیه‌ی برش داده شده) بخشی از اعضای جامعه  $U$  را نشان می‌دهد که در آمارگیری شرکت داده نمی‌شوند و به طور کامل از بررسی حذف می‌شوند. در واقع برای  $j \in U_E$  داریم  $\pi_j = 0$  و  $U = U_E \cup U_C$ .

### ۱-۱-۲- برآوردگر برشی نوع اول با رهیافت حذف

در این رهیافت پس از تقسیم‌بندی جامعه به دو ناحیه  $U_C$  و  $U_E$  برآوردگر ارائه شده برای پارامتر مورد نظر فقط با استفاده از مقدارهای صفت  $Y$  در ناحیه‌ی سرشماری  $U_C$  محاسبه می‌شوند. در این حالت برآوردگر مجموع برای  $t$  عبارت است از:

$$(۱) \quad \hat{t}_{\text{cut}} = \sum_{i \in U_C} Y_i = N_C \bar{Y}_C$$

که در آن  $N_C$  اندازه‌ی ناحیه‌ی سرشماری و  $\bar{Y}_C$  میانگین ناحیه‌ی سرشماری است. در این حال برآوردگر ارائه شده اریب می‌باشد و میزان اریبی آن عبارت است از:

$$(۲) \quad \begin{aligned} |\text{Bias}(\hat{t}_{\text{cut}})| &= |N_C \bar{Y}_C - t| = |N_C \bar{Y}_C - N \bar{Y}_N| \\ &= |N_C \bar{Y}_C - N_C \bar{Y}_C - N_E \bar{Y}_E| = |-t_E| = |t_E| \end{aligned}$$

که در آن  $N$  اندازه‌ی کل و  $\bar{Y}_N$  میانگین کل جامعه است [۱].

### ۲-۱-۲- برآوردگر برشی نوع اول با رهیافت مدل‌یار

در این روش از مقادیری که در ناحیه‌ی حذف قرار دارند به وسیله‌ی برقراری رابطه با مقادیر ناحیه‌ی سرشماری به صورت زیر استفاده می‌شود. واضح است که  $t_E = \delta t_C$  که در آن  $t_C$  مجموع در ناحیه‌ی سرشماری و  $\delta$  در فرمول (۴) در زیر تعریف شده است.

بنابراین می‌توان  $t_E$  را به وسیله  $\hat{t}_E = \tilde{\delta} t_C$  برآورد کرد. در نتیجه برآوردگر برشی نوع اول  $t$  با استفاده از رهیافت مدل‌یار به صورت زیر به دست می‌آید:

$$\begin{aligned} \hat{t}_{\text{cut}} &= t_C + \hat{t}_E = (1 + \tilde{\delta}) t_C \\ &= (1 + \tilde{\delta}) \sum_{i \in U} Y_i \end{aligned} \quad (۳)$$

در این حالت میزان اریبی  $\hat{t}_{\text{cut}}$  عبارت است از:

$$\begin{aligned} \text{Bias}(\hat{t}_{\text{cut}}) &= E(\hat{t}_{\text{cut}}) - t \\ &= t_C + \tilde{\delta} t_C - t_C - t_E \\ &= (\tilde{\delta} - \delta) t_C \end{aligned}$$

که در آن

$$\delta = \frac{\sum_{i \in U_E} Y_i}{\sum_{i \in U_C} Y_i} = \frac{t_E}{t_C} \quad (۴)$$

و همچنین برآورد آن به صورت زیر به دست می‌آید:

$$\tilde{\delta} = \frac{\sum_{i \in U_E} X_i}{\sum_{i \in U_C} X_i} = \frac{t_{X_E}}{t_{X_C}} \quad (۵)$$

که در آن  $X_E$  متغیر کمکی مورد استفاده در ناحیه‌ی حذف و  $X_C$  مقدار متغیر کمکی در ناحیه‌ی سرشماری است.

## ۲-۲- نمونه‌گیری برشی نوع اول تبدیل‌یافته

در بررسی صفت  $Y$  پس از تقسیم جامعه به دو ناحیه (بر حسب بزرگی مقدارهای  $Y$ )، به جای سرشماری از ناحیه‌ی  $U_C$  نمونه‌گیری به عمل می‌آید، یعنی ناحیه‌ای از جامعه

را نشان می‌دهد که در آن واحدهایی با مقادیرهای بزرگ  $Y$  قرار دارند و نمونه‌ای به اندازه‌ی  $n$  از آن برای بررسی انتخاب می‌شود. همچنین مشابه قبل  $U_E$  ناحیه‌ای از جامعه را نشان می‌دهد که در آن واحدهایی با مقادیرهای کوچک  $Y$  قرار دارند.

### ۱-۲-۲- برآوردگر برشی نوع اول تبدیل یافته با رهیافت حذف

در این روش برآوردگر برشی به صورت زیر معرفی می‌شود:

$$(۶) \quad \begin{aligned} t &= t_E + t_S \\ \bar{Y}_N &= W_E \bar{Y}_E + W_S \bar{Y}_S \end{aligned}$$

که  $W_S = \frac{N_S}{N}$ ،  $W_E = \frac{N_E}{N}$  و  $W_E + W_S = 1$  و در نهایت با نادیده گرفتن مقادیری که در ناحیه‌ی حذف قرار گرفته برآورد  $t$  برابر  $\hat{t}_{cut} = N \hat{Y}_N$  می‌شود. و مقدار اریبی آن برابر است با:

$$\begin{aligned} \text{Bias}(\hat{t}_{cut}) &= E(\hat{t}_{cut}) - t \\ &= t_S + \tilde{\delta} t_S - t_S - t_E \\ &= (\tilde{\delta} - \delta) t_S. \end{aligned}$$

### ۲-۲-۲- برآوردگر برشی نوع اول تبدیل یافته با رهیافت مدل یار

در این حالت  $t$  را با توجه به روابط ارائه شده‌ی  $t_E = \delta t_S$  و در نتیجه  $t = (1 + \delta) t_S$  که در آن  $\delta$  به صورت زیر برآورد می‌شود:

$$(۷) \quad \tilde{\delta} = \frac{\sum_{i \in U_E} X_i}{\sum_{i \in U_S} X_i} = \frac{t_{X_E}}{t_{X_S}}$$

حال پس از برآورد  $\hat{t}_S$  و برآورد  $\tilde{\delta}$  توسط  $\tilde{\delta}$  برآوردگر برشی به صورت  $\hat{t}_{cut} = (1 + \tilde{\delta}) \hat{t}_S$  به دست می‌آید. در نهایت مقدار اریبی به صورت زیر محاسبه می‌شود:

..... گزیده مطالب آماری، سال ۲۰، شماره‌ی ۱، بهار و تابستان ۱۳۸۸، صص ۹۵-۱۲۲

$$\begin{aligned} \text{Bias}(\hat{t}_{\text{cut}}) &= (1 + \tilde{\delta}) E(\hat{t}_S) - t = (1 + \tilde{\delta}) t_S - t \\ &= (\tilde{\delta} - \delta) t_S. \end{aligned}$$

### ۳-۲- نمونه‌گیری برشی نوع دوم (سرشماری- نمونه‌گیری)

در صورت به کارگیری روش‌های نمونه‌گیری برشی نوع اول و نوع تبدیل‌یافته‌ی آن برآوردهای به دست آمده اریب می‌باشند، در ضمن نمی‌توان با اطمینان مناسب نتیجه‌های به دست آمده در سطح نمونه را به جامعه تعمیم داد. در نهایت استفاده‌ی موفقیت‌آمیز این روش‌ها وابسته به صادق بودن فرض‌هایی است که گاهی وقت‌ها غیر قابل اثبات هستند.

در نتیجه یک روش جایگزین برای این روش‌ها، روش نمونه‌گیری برشی نوع دوم (سرشماری- نمونه‌گیری) است که پس از تعیین آستانه‌ی برش مناسب، جامعه‌ی مورد بررسی به دو ناحیه‌ی  $U_S$  و  $U_C$  برش داده می‌شود که در آن  $U_C$  شامل واحدهایی از جامعه است که متناظر با مقدارهای بزرگ و غیر قابل چشم‌پوشی  $Y$  هستند و به طور کامل و با احتمال یک در بررسی شرکت داده می‌شوند (ناحیه‌ی سرشماری). در حالی که ناحیه‌ی  $U_S$  شامل واحدهای باقی‌مانده‌ی جامعه است که از آن نمونه‌ای تصادفی اختیار شده و برای شرکت در آمارگیری استفاده می‌شود.

این نوع روش نمونه‌گیری برشی از نوع روش‌های نمونه‌گیری احتمالاتی است و برآوردهای به دست آمده در این روش نارایب هستند.

پس از استخراج نمونه‌ای به اندازه‌ی  $n_S$  از ناحیه‌ی نمونه‌گیری  $U_S$  که دارای  $N_S = N - N_C$  عضو است و مشاهده‌ی مقدارهای  $y_1, \dots, y_{n_S}$  یک برآوردگر برای  $\bar{Y}_N$  عبارت است از:

$$(A) \quad \hat{Y}_{\text{cut}} = W_C \bar{Y}_C + W_S \bar{y}_S, \quad W_S = \frac{N_S}{N}, \quad W_C = \frac{N_C}{N}.$$

همچنین باید توجه کرد که در ناحیه‌ی  $U_S$  (که می‌توان آن را یک جامعه‌ی  $N_S$  عضوی نیز در نظر گرفت)  $\bar{y}_S$  یک برآورد نارایب برای  $\bar{Y}_S$  است، بنا بر این :

$$\begin{aligned} E(\hat{Y}_{\text{cut}}) &= W_C \bar{Y}_C + W_S E(\bar{y}_S) \\ &= W_C \bar{Y}_C + W_S \bar{Y}_S \\ &= \bar{Y}_N \end{aligned}$$

#### ۴-۲- نمونه‌گیری برشی نوع سوم (سرشماری- نمونه‌گیری- حذف)

در روش نمونه‌گیری برشی نوع دوم ممکن است در عمل واحدهایی در ناحیه‌ی  $U_S$  موجود باشند که نمونه‌گیری از آن‌ها منجر به صرف وقت و هزینه‌ی زیاد شود، بنا بر این گرچه برآوردهای ارائه‌شده برای پارامترهای مورد نظر در روش نمونه‌گیری برشی نوع دوم نارایب هستند، اما از نقطه نظر عملی از کارایی کم‌تری برخوردار هستند. یک روش جایگزین برای این روش که به ویژه در بررسی جامعه‌هایی با چولگی زیاد کاربرد دارد استفاده از روش نمونه‌گیری برشی نوع سوم است که در واقع تلفیقی از روش‌های نمونه‌گیری برشی نوع اول و دوم است. در این روش پس از تعیین آستانه‌های برش مناسب، جامعه‌ی مورد بررسی به سه ناحیه برش داده می‌شود که در آن یک ناحیه شامل واحدهایی از جامعه با مقدارهای بزرگ و غیر قابل چشم‌پوشی  $Y$  است که سرشماری می‌شوند ( $U_C$ ). ناحیه‌ی  $U_S$  شامل بخشی از جامعه با مقدارهای متوسط  $Y$  است که از آن نمونه‌گیری به عمل می‌آید و نهایتاً ناحیه‌ی  $U_E$  شامل بخشی از جامعه با مقدارهای کوچک و قابل چشم‌پوشی  $Y$  هستند. لازم به ذکر است در روش نمونه‌گیری برشی هنگام ارائه‌ی برآوردی از پارامتر مورد نظر عموماً دو نوع رهیافت در رابطه با مقدار صفت  $Y$  در قسمت حذف‌شده مورد استفاده قرار می‌گیرد که عبارت‌اند از رهیافت حذف (واحدهایی از جامعه که در ناحیه‌ی برش داده‌شده قرار می‌گیرند کلاً از بررسی حذف شده) و رهیافت مدل‌یار (پس از تعیین آستانه‌ی برش و تشکیل ناحیه‌ی  $U_E$  از طریق اطلاعات کمکی مقدارهای صفت  $Y$  برای قسمت حذف‌شده برآورد می‌شوند). در ادامه برآوردهای مربوطه به این رهیافت‌ها ارائه می‌شوند.



## ۱-۴-۲- برآوردگر برشی نوع سوم با رهیافت حذف

پس از تعیین آستانه‌های برش مناسب و تقسیم جامعه به ناحیه‌های  $U_C$ ،  $U_S$  و  $U_E$  برآوردگر  $\bar{Y}_N$  فقط به کمک مقادیرهای صفت  $Y$  در ناحیه‌ی  $U_C$  و نمونه‌ی به دست آمده از ناحیه‌ی  $U_S$  به صورت زیر محاسبه می‌شود. یعنی:

$$(۹) \quad \hat{Y}_{cut} = W_C \bar{Y}_C + W_S \bar{Y}_S$$

که در آن  $\bar{Y}_C$  میانگین ناحیه‌ی سرشماری،  $\bar{Y}_S$  میانگین ناحیه‌ی نمونه‌گیری،

$$W_C = \frac{N_C}{N} \text{ و } W_S = \frac{N_S}{N} \text{ هستند. بنا بر این}$$

$$(۱۰) \quad \text{var}(\hat{Y}_{cut}) = W_C^2 \text{var}(\bar{Y}_C) + W_S^2 \text{var}(\bar{Y}_S) = W_S^2 (1 - f_S) \frac{S_S^2}{n_S}$$

که در آن  $f_C$  عامل تصحیح جامعه‌ی محدود،  $S_S^2$  واریانس ناحیه‌ی نمونه‌گیری و  $n_S$  اندازه‌ی نمونه است و میزان اریبی آن به صورت زیر محاسبه می‌شود:

$$\text{Bias}(\hat{Y}_{cut}) = W_C \bar{Y}_C + W_S \bar{Y}_S - \bar{Y}_N = -W_E \bar{Y}_E$$

که در آن  $W_E = \frac{N_E}{N}$  است.

با توجه به نتیجه‌های به دست آمده میانگین توان‌های دوم خطا (MSE) برای  $\hat{Y}_{cut}$  عبارت است از:

$$(۱۱) \quad \text{MSE}(\hat{Y}_{cut}) = W_S^2 (1 - f_S) \frac{S_S^2}{n_S} + W_E^2 \bar{Y}_E^2$$

## ۲-۴-۲- برآوردگر برشی نوع سوم با رهیافت مدل‌یار

با استفاده از رهیافت مدل‌یار به جای حذف مقادیرهای متناظر  $Y$  با واحدهای جامعه در ناحیه‌ی  $U_E$  با استفاده از اطلاعات کمکی در دست و اعمال برخی فرضیه‌ها، مجموع مقادیرهای  $Y$  در ناحیه‌ی  $U_E$  برآورد می‌شود. برای این منظور رابطه‌ی بین واحدهای جامعه در ناحیه‌ی حذف  $U_E$  و ناحیه‌های  $U_C$  و  $U_S$  برآورد می‌شود تا به کمک آن

برآورد دقیق تری برای میانگین جامعه به دست آید. این رابطه را می توان به صورت زیر تعریف کرد [۱]:

$$(۱۲) \quad \delta = \frac{N_E \bar{Y}_E}{N_C \bar{Y}_C + N_S \bar{Y}_S}$$

که در آن میانگین ناحیه‌ی سرشماری،  $\bar{Y}_C$  میانگین ناحیه‌ی نمونه‌گیری،  $\bar{Y}_E$  میانگین ناحیه‌ی حذف،  $N_C$  اندازه‌ی ناحیه‌ی سرشماری،  $N_S$  اندازه‌ی ناحیه‌ی نمونه‌گیری و  $N_E$  اندازه‌ی ناحیه‌ی حذف هستند.

وقتی مقدار  $\delta$  از قبل معلوم نباشد به کمک مقادیر متغیر کمکی  $X$  که دارای رابطه‌ی مستقیم با  $Y$  است، می توان  $\delta$  را به کمک  $\tilde{\delta}$  و به صورت زیر برآورد کرد:

$$(۱۳) \quad \tilde{\delta} = \frac{\sum_{i \in U_E} X_i}{\sum_{i \in U_C} X_i + \sum_{i \in U_S} X_i} = \frac{N_E \bar{X}_{X_E}}{N_C \bar{X}_{X_C} + N_S \bar{X}_{X_S}}$$

که در آن میانگین مقدار متغیر کمکی در ناحیه‌ی سرشماری،  $\bar{X}_{X_S}$  میانگین مقدار متغیر کمکی در ناحیه‌ی نمونه‌گیری و  $\bar{X}_{X_E}$  میانگین مقدار متغیر کمکی در ناحیه‌ی حذف است.

که در نهایت میانگین جامعه برابر است با

$$\bar{Y}_N = W_C \bar{Y}_C + W_S \bar{Y}_S + W_E \bar{Y}_E.$$

حال با فرض آن که  $\delta \cong \tilde{\delta}$ ، با برآورد  $\bar{Y}_S$  توسط  $\bar{y}_S$ ، برآوردگر برشی نوع سوم مدل یار برای  $\bar{Y}_N$  به صورت زیر به دست می آید:

$$(۱۴) \quad \hat{Y}_{\text{cut}} = (1 + \tilde{\delta})(W_C \bar{Y}_C + W_S \bar{y}_S)$$

میزان ارزیابی این برآوردگر به صورت زیر است:

$$\begin{aligned} \text{Bias}(\hat{Y}_{\text{cut}}) &= E(\hat{Y}_{\text{cut}}) - \bar{Y}_N = W_C \bar{Y}_C + W_S \bar{Y}_S + \tilde{\delta}(W_C \bar{Y}_C + W_S \bar{Y}_S) - \bar{Y}_N \\ &= (\tilde{\delta} - \delta)(W_C \bar{Y}_C + W_S \bar{Y}_S) \end{aligned}$$

همچنین

$$(۱۵) \quad \text{var}(\hat{Y}_{\text{cut}}) = (1 + \tilde{\delta})^2 W_S^2 (1 - f_S) \frac{S_S^2}{n_S}$$

میانگین توان دوم خطا برای برآوردگر برشی نوع سوم مدل یار ارائه‌شده‌ی بالا عبارت است از:

$$(۱۶) \quad \text{MSE}(\hat{Y}_{\text{cut}}) = (1 + \tilde{\delta})^2 W_S^2 (1 - f_S) \frac{S_S^2}{n_S} + (\tilde{\delta} - \delta)^2 (W_C \bar{Y}_C + W_S \bar{Y}_S)^2$$

اگر  $0 \leq \tilde{\delta} / \delta \leq 2$  و یا  $\tilde{\delta} \cong 0$  در این صورت اندازه‌ی میانگین توان دوم خطای برآوردگر مورد نظر در رهیافت مدل یار کم‌تر از رهیافت حذف بوده و با اریبی کم‌تر روش مناسب‌تری برای استفاده از نمونه‌گیری برشی نوع سوم می‌باشد.

### ۳-۴-۲- تعیین اندازه‌ی نمونه در رهیافت حذف

با استفاده از  $\text{MSE}(\hat{Y}_{\text{cut}}) \leq v_0$ ، کم‌ترین مقدار اندازه‌ی نمونه برای دستیابی به دقت مورد نظر  $v_0$  در برآورد  $\bar{Y}_N$  به صورت زیر به دست می‌آید:

$$(۱۷) \quad n_S \cong \frac{W_S^2 S_S^2}{\frac{W_S^2}{N_S} S_S^2 + (v_0 - W_E^2 \bar{Y}_E^2)}$$

اگر اطلاعات کمکی قابل اطمینانی در ارتباط با صفت مورد نظر  $Y$  در دسترس باشد اندازه‌ی نمونه به طور تقریبی به صورت زیر برآورد می‌شود:

$$(۱۸) \quad n_S \cong \frac{W_S^2 S_{X_S}^2}{\frac{W_S^2}{N_S} S_{X_S}^2 + (v_0 - W_E^2 \bar{X}_E^2)}$$

#### ۴-۴-۲- تعیین اندازه‌ی نمونه در رهیافت مدل‌یار

کم‌ترین اندازه‌ی نمونه برای برآورد  $\bar{Y}_N$  به کمک روش نمونه‌گیری برشی نوع سوم با رهیافت مدل‌یار و دقت مورد نظر  $v_0$  عبارت است از:

$$(19) \quad n_S = \frac{W_S^r S_S^r}{\frac{W_S^r}{N_S} S_S^r + \psi_{v_0}}$$

که در آن

$$\psi_{v_0} = \frac{v_0 - (\tilde{\delta} - \delta)^r (W_C \bar{Y}_C + W_S \bar{Y}_S)^r}{(1 + \tilde{\delta})^r}$$

#### ۴-۴-۵- تعیین آستانه‌ی برش

میزان آریبی برآوردهای به دست آمده و نیز دقت برآوردهای ارائه‌شده برای  $t$  و  $\bar{Y}_N$  در هر دو رهیافت حذف یا مدل‌یار به آستانه‌ی برش و تعداد عناصر درون ناحیه‌های حذف، سرشماری و نمونه‌گیری بستگی دارد. برای این منظور ابتدا آستانه‌ی برش برای ناحیه‌ی حذف تعیین شده، سپس برای تعیین ناحیه‌های سرشماری و نمونه‌گیری اقدام می‌شود.

پس از مرتب کردن مقادیرهای  $Y$  در جامعه به صورت  $Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(N)}$  و تشکیل  $S_y(l) = \sum_{j=1}^l Y_{(j)}$  ها، که در آن ناحیه‌ی  $U_{CS}$  شامل  $100(\varepsilon)\%$  از اعضای جامعه با

مقادیرهای بزرگ و متوسط  $Y$  است و تمام واحدهای جامعه که برای آن‌ها  $S_y(l) \leq (1 - \varepsilon) S_y(N)$  (که در آن  $S_y(N)$  میانگین مقادیر واحدهای جامعه و  $S_y(l)$  میانگین مقادیر تا واحد  $l$ ام جامعه است)، در ناحیه‌ی  $U_E$  قرار می‌گیرند. در عمل مقدار  $\varepsilon$  از تجربه‌های گذشته و با در نظر گرفتن نظر کارشناسان و متخصصین تعیین می‌شود. هدف تعیین آستانه‌ی برش  $Y_S^*$  به گونه‌ای است که  $U_{CS}$  به ناحیه‌های  $U_C$  و  $U_S$  تقسیم شود.

**الف) تعیین  $Y_S^*$  برای دقت مفروض با رهیافت حذف**

در این روش دقت برآورد از قبل معلوم فرض شده و هدف ارائه‌ی آستانه‌ی برش بهینه می‌باشد که می‌توان آستانه‌ی برش را به صورت زیر به دست آورد: برای سادگی فرض کنید اندازه‌ی ناحیه‌ی سرشماری  $N_C = m$  باشد.

$$(20) \quad Y_S^* = \bar{Y}_S(m) + \sqrt{\frac{N-l-m-1}{(N-l-m)^2} N^2 v_o + S_S^2(m)}$$

که در آن  $N$  اندازه‌ی جامعه،  $l$  اندازه‌ی ناحیه‌ی حذف،  $\bar{Y}_S(m)$  میانگین ناحیه‌ی نمونه‌گیری و  $S_S^2(m)$  واریانس ناحیه‌ی نمونه‌گیری است. لازم به ذکر است یکی از ایرادهای آستانه‌ی برش ارائه‌شده آن است که محاسبه آن و ارائه‌ی مقدار  $m$  بهینه به مقدارهای  $\bar{Y}_S(m)$  و  $S_S^2(m)$  بستگی دارد، بنا بر این در عمل نمی‌توان از آن استفاده کرد. برای همین منظور می‌توان کران بالایی برای  $Y^*$  به صورت زیر ارائه کرد که به  $\bar{Y}_S(m)$  و  $S_S^2(m)$  وابسته نباشد:

$$(21) \quad Y_S^* < \bar{Y}_{CS} + \sqrt{\frac{N^2}{N-l} v_o + S_{CS}^2}$$

که در آن  $S_{CS}^2$  واریانس ناحیه‌ی نمونه‌گیری و سرشماری است.

**ب) تعیین  $Y_S^*$  برای دقت مفروض با رهیافت مدل‌یار**

در حالتی که از اطلاعات موجود در ناحیه‌ی حذف استفاده شود با استفاده از فرمول اندازه‌ی نمونه آستانه‌ی برش به صورت زیر حاصل می‌شود [۱]:

$$(22) \quad Y_S^* = \bar{Y}_S(m) + \sqrt{\frac{N-l-m-1}{(N-l-m)^2} N^2 \Psi_{v_o} + S_S^2(m)}$$

که در آن

$$\Psi_{v_o} = \frac{v_o - (\tilde{\delta} - \delta)^2 (W_C \bar{Y}_C + W_S \bar{Y}_S)^2}{(1 + \tilde{\delta})^2}$$

همچنین یک کران بالا برای آستانه‌ی برش بالا عبارت است از:

$$(23) \quad Y_S^* \leq \bar{Y}_{CS} + \sqrt{\frac{N^2}{N-1} \left( \frac{v_0 - (\tilde{\delta} - \delta)^2 (\bar{Y}_N - W_E \bar{Y}_E)^2}{(1 + \tilde{\delta})^2} \right)} + S_{CS}^2.$$

### ۳- برآورد میانگین ارزش افزوده‌ی کارگاه‌های صنعتی ۱۰ کارکن و بیش‌تر در سال ۱۳۸۵

نظر به نقشی که نتایج حاصل از اجرای طرح آمارگیری از کارگاه‌های صنعتی کشور در روند توسعه‌ی صنعتی کشور ایفا می‌کند، حصول نتایج دقیق از طرح‌های موجود در این زمینه برای دستیابی به اهداف رشد و توسعه در این بخش ضروری است و این موضوع از دیدگاه صرفه جویی در هزینه همراه با حفظ کیفیت نتایج، اهمیت ویژه‌ای پیدا می‌کند. یکی از طرح‌های مهم مرکز آمار ایران آمارگیری از قیمت تولیدکننده‌ی بخش صنعت می‌باشد. این طرح با هدف کلی تعیین شاخص قیمت تولیدکننده‌ی بخش صنعت اجرا می‌شود، که میانگین قیمت نسبی قلم و یا مجموعه‌ای از قلم‌های مورد نظر جامعه‌ی هدف به منظور استفاده در محاسبه‌ی شاخص قیمت تولیدکننده‌ی بخش صنعت در سطح موضوعی<sup>۱</sup> و جغرافیایی<sup>۲</sup> را برآورد می‌کند.

در این طرح جامعه‌ی هدف تمام محصولات و خدمات صنعتی تولید شده در کارگاه‌های صنعتی کشور است و هر یک از محصولات و خدمات صنعتی تولید شده در کارگاه‌های مورد نظر واحد آماری مزبور می‌باشد. زمان آمارگیری، هر یک از ۱۲ ماه سال، از بیستم شروع و تا سی‌ام همان ماه ادامه دارد و در نهایت اطلاعات مورد نظر از طریق مراجعه به منابع اطلاعات منتخب (کارگاه نمونه) و مصاحبه‌ی مستقیم با پاسخگوی مطلع جمع‌آوری می‌شود. روش آمارگیری مرکز آمار ایران در این طرح به این صورت است که در ابتدا از کارگاه‌های صنعتی ۱۰ کارکن و کم‌تر با توجه به چارچوب به دست آمده از سال ۱۳۸۱، نمونه‌گیری و از مابقی جامعه یعنی کارگاه‌های صنعتی ۱۰ کارکن و بیش‌تر سرشماری به عمل می‌آورد. لازم به ذکر است در هر رشته فعالیت صنعتی نمونه‌گیری

.....گزیده‌مطالب آماری، سال ۲۰، شماره‌ی ۱، بهار و تابستان ۱۳۸۸، صص ۹۵-۱۲۲.....

به صورت مستقل انجام می‌پذیرد و انتخاب اقلام مورد نظر با رعایت ضوابط خاصی انجام می‌گیرد.

با توجه به توضیحات داده شده یکی از روش‌هایی که می‌توان برای نمونه‌گیری از کارگاه‌های صنعتی استفاده کرد، نمونه‌گیری برشی است. در این بخش از طرح نمونه‌گیری برشی نوع سوم برای برآورد میانگین ارزش افزوده کارگاه‌های صنعتی سال ۱۳۸۵ استفاده می‌شود. چارچوب مورد استفاده، با توجه به محدودیت‌های موجود (عدم وجود اطلاعات سرشماری چارچوب کارگاه‌های صنعتی کم‌تر از ۱۰ کارکن برای سال ۱۳۸۵ و همچنین عدم دسترسی به اطلاعات ارزش افزوده‌ی چارچوب به دست آمده برای کارگاه‌های صنعتی کم‌تر از ۱۰ کارکن سال ۱۳۸۱) اطلاعات حاصل از سرشماری برای کارگاه‌های صنعتی ۱۰ کارکن و بیش‌تر مربوط به سال ۱۳۸۵ می‌باشد.

به دنبال آستانه‌ی برش و اندازه‌ی نمونه‌ی لازم برای اجرای نمونه‌گیری برشی نوع سوم در سال ۱۳۸۶ با استفاده از داده‌های کارگاه‌های صنعتی ۱۰ نفر کارکن و بیش‌تر سال ۱۳۸۵ تعیین می‌شود. از آن‌جا که دسترسی به داده‌های این آمارگیری در سال ۱۳۸۴ مقدور نبود تا بر اساس آن‌ها آستانه‌ی برش و اندازه‌ی نمونه‌ی لازم در سال ۱۳۸۵ تعیین شود، لذا از آستانه‌ی برش و اندازه‌ی نمونه‌ی تعیین شده برای سال ۱۳۸۶ برای داده‌های سال ۱۳۸۵ استفاده شده است. در نهایت در یک مطالعه عمل‌کرد روش نمونه‌گیری برشی با رهیافت حذف و مدل‌یار مقایسه شدند. نتیجه‌ی این مطالعه نشان می‌دهد که رهیافت مدل‌یار تحت شرایطی بهتر از رهیافت حذف در این شیوه‌ی نمونه‌گیری عمل می‌کند. همچنین نتایج به دست آمده برای نمونه‌گیری برشی نوع سوم با رهیافت مدل‌یار با نتایج به دست آمده از نمونه‌گیری نسبتی مقایسه شد که نشان‌گر عملکرد بهتر نمونه‌گیری نسبتی می‌باشد.

از میان فعالیت‌های مختلفی که در کارگاه‌های صنعتی انجام می‌گیرد، کارگاه‌های صنعتی تولید آجر به دلیل تعداد زیاد کارگاه‌ها (۱۴۸۴ کارگاه) به عنوان جامعه‌ی هدف انتخاب شد. سپس با استفاده از نمونه‌گیری برشی و رهیافت‌های حذف و مدل‌یار و تولید ۱۰۰۰ نمونه، میانگین ارزش افزوده‌ی هر یک از رهیافت‌ها محاسبه و آریبی و میانگین توان‌های دوم خطای (MSE) آن‌ها محاسبه و مقایسه می‌شوند.

دلایل انتخاب ارزش افزوده به عنوان عامل مورد بررسی (۱) قابلیت دسترسی آسان و (۲) برای محاسبه‌ی قیمت نسبی و شاخص قیمت<sup>۳</sup> است. این اطلاعات را می‌توان از طریق فرمول‌های مربوط و اطلاعات پرسشنامه در بخش عملکرد کارگاه از دفترچه‌ی راهنمای بازیبن فنی استخراج نمود. ارزش افزوده، ثروت ایجاد شده توسط واحد تجاری است که این ثروت حاصل کار و تلاش گروهی تمامی افرادی است که به نوعی در واحد تجاری سهامی بوده‌اند. اما در شرایط کلی می‌توان گفت توسعه‌ی منابع و بهره‌گیری از تمامی منابع موجود در واحد تجاری، به شکلی که بازده‌ای قابل توجه و پایدار بیش از آنچه که خرج شده است برای این واحد به بار آورد، در نهایت از طریق ستانده منهای داده محاسبه می‌شود.

لازم به ذکر است برای محاسبه‌ی برآوردها و معیارهای مقایسه، از نرم‌افزار R استفاده شده است. اکنون به تشریح کامل مراحل نمونه‌گیری برشی نوع سوم که توسط آن برآوردها محاسبه شده‌اند، پرداخته می‌شود.

با توجه به توضیحاتی که برای نمونه‌گیری برشی در ابتدا داده شد اولین نکته در استفاده از این نوع روش نمونه‌گیری تعیین آستانه‌ی برش است که به طور مستقیم در آماره‌های آمارگیری مؤثر می‌باشد. در شروع کار چون هیچ اطلاعی درباره‌ی متغیر مورد بررسی در چارچوب وجود ندارد نمی‌توان با استفاده از چارچوب آستانه‌ی برش را تعیین نمود. در نتیجه می‌توان از عامل "تعداد کارکنان" به عنوان متغیر کمکی استفاده کرد. که با توجه به در دسترس بودن متغیر کمکی تعداد کارکنان و موجود بودن آن در تمامی چارچوب‌ها، به کارگیری این متغیر مطلوب‌تر به نظر می‌رسد.

با توجه به نوع روش نمونه‌گیری برشی باید دو آستانه‌ی برش برای جامعه‌ی مورد نظر به منظور تفکیک سه ناحیه با استفاده از متغیر کمکی تعداد کارکنان تعیین نمود (داده‌های مربوط به ارزش افزوده بر حسب میلیون ریال می‌باشند). در ادامه آستانه‌ی برش و اندازه‌ی نمونه‌ی لازم برای اجرای نمونه‌گیری برشی نوع سوم در سال ۱۳۸۶ با استفاده از داده‌های کارگاه‌های صنعتی ۱۰ نفر کارکن و بیش‌تر سال ۱۳۸۵ تعیین می‌شود.



### ۱-۳- تعیین آستانه‌ی برش میان ناحیه‌های حذف ( $U_E$ ) و مابقی جامعه ( $U_{CS}$ )

ابتدا تعداد کارکنان کارگاه‌ها به ترتیب صعودی مرتب می‌شوند، سپس جمع تجمعی تعداد کارکنان محاسبه شده و با  $S(l)$  نشان داده می‌شود. با توجه به فرمول ارائه شده در بخش تعیین آستانه‌ی برش مجموع کل برابر  $S(N) = 1/21585E + 6$  شده، با توجه به این که  $\varepsilon = 0/90$  (مقدار  $\varepsilon$  از طریق اطلاعات موجود در سال‌های قبل و نظر کارشناسان تعیین می‌گردد) در نظر گرفته می‌شود، در نهایت آستانه‌ی برش برای ناحیه‌ی حذف و نمونه‌گیری به صورت زیر محاسبه می‌شود:

$$S(l) \leq (1 - \varepsilon)S(N) \rightarrow S(l) \leq 0/1S(N) \rightarrow S(l) = 4265$$

$S(l)$  جمع تجمعی ارزش افزوده است که مقدار ارزش افزوده‌ی این مقدار برابر  $116/390$  میلیون ریال می‌باشد. از آن‌جا که داده‌ها با توجه به تعداد کارکنان کارگاه‌ها مرتب شده‌اند و میزان ارزش افزوده‌ی  $116/390$  متعلق به کارگاهی با ۱۵ کارکن است لذا کارگاه‌هایی با تعداد کارکنان ۱۵ نفر و کم‌تر در ناحیه‌ی حذف قرار می‌گیرند. بنا بر این کارگاه‌هایی که تعداد کارکنان آن‌ها بیش‌تر از ۱۵ نفر بوده است به‌عنوان ناحیه‌ی سرشماری و نمونه‌گیری در نظر گرفته می‌شوند (که تعداد کارگاه‌هایی که در ناحیه‌ی حذف قرار می‌گیرد ۳۳۳ کارگاه می‌باشد).

### ۲-۳- تعیین آستانه‌ی برش برای ناحیه‌های نمونه‌گیری ( $U_S$ ) و سرشماری ( $U_C$ )

برای تفکیک دو ناحیه‌ی نمونه‌گیری و سرشماری از دو نوع رهیافت استفاده می‌کنیم. اولی رهیافت حذف است که در آن عنصرهای ناحیه‌ی حذف در محاسبات شرکت داده نمی‌شوند و دیگری رهیافت مدل‌یار است که در آن به وسیله‌ی ایجاد رابطه بین عنصرهای ناحیه‌ی حذف، ناحیه‌ی نمونه‌گیری و ناحیه‌ی سرشماری به نحوی از عنصرهایی که در ناحیه‌ی حذف قرار می‌گیرند استفاده می‌شود.

همان طور که قبلاً توضیح داده شد، علاوه بر تعیین آستانه‌ی برش و اندازه‌ی نمونه‌ی لازم برای اجرای نمونه‌گیری برشی نوع سوم در آمارگیری از کارگاه‌های صنعتی دارای ۱۰ کارکن و بیش‌تر سال ۱۳۸۶ در نظر است میانگین ارزش افزوده‌ی کارگاه‌ها در این آمارگیری با دو رهیافت حذف و مدل‌بار نیز برآورد شده و عمل کرد آن‌ها مقایسه شود. از آن‌جا که دسترسی به داده‌های این آمارگیری در سال ۱۳۸۴ مقدور نبود تا بر اساس آن‌ها آستانه‌ی برش و اندازه‌ی نمونه‌ی لازم در سال ۱۳۸۵ تعیین شود، لذا از آستانه‌ی برش و اندازه‌ی نمونه‌ی تعیین‌شده برای سال ۱۳۸۶ برای داده‌های سال ۱۳۸۵ استفاده شده است.

### ۱-۲-۳- روش حذف

آستانه‌ی برش برای ناحیه‌ی حذف را می‌توان به وسیله‌ی دو روش تعیین نمود: اول بر اساس اندازه‌ی نمونه و دوم بر اساس اندازه‌ی دقت مورد نظر. چون یکی از اهداف یافتن اندازه‌ی نمونه می‌باشد، یعنی این که این تعداد در دسترس نمی‌باشد، در نتیجه آستانه‌ی برش بر اساس دقت مورد نظر تعیین می‌شود.

در ابتدا چون هیچ اطلاعی درباره‌ی واریانس و میانگین توان دوم خطای جامعه نداریم برای برآورد دقت مورد نظر در تعیین اندازه‌ی نمونه، نمونه‌ای به اندازه‌ی ۲۰ از ناحیه‌ی نمونه‌گیری و سرشماری به طور تصادفی انتخاب کرده و بر اساس آن MSE متغیر تعداد کارکنان  $X$  را برآورد می‌کنیم.

$$\text{var}(\hat{X}) = 237 / 3684 \quad \text{برآورد واریانس}$$

$$\text{Bias}(\hat{X}) = 8 / 259814 \quad \text{میزان اریبی}$$

$$\text{MSE}(\hat{X}) = 15 / 275390 \quad \text{در نهایت MSE}$$

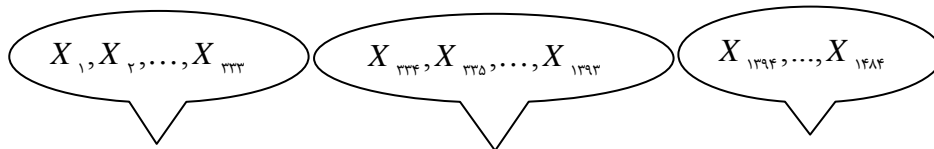
با استفاده از فرمول تعیین آستانه‌ی برش و همچنین محاسبه‌ی مقدار میانگین و واریانس ناحیه‌ی نمونه‌گیری برای رهیافت حذف مقدار مرزی بین ناحیه‌های نمونه‌گیری و سرشماری به صورت زیر تعیین می‌شود:

میانگین تعداد کارکنان ناحیه‌ی نمونه‌گیری و سرشماری  $\bar{X}_{CS} = 33/35938$

واریانس تعداد کارکنان ناحیه‌ی نمونه‌گیری و سرشماری  $S_{CS}^2 = 696/6283$

$$X_S^* = \bar{X}_{CS} + \sqrt{\frac{N^2}{N-1} \text{MSE} + S_{CS}^2} = 61/314 \quad \text{در نتیجه}$$

در این حالت کارگاه‌هایی که دارای ۶۲ کارکن و بیش‌تر می‌باشند در ناحیه‌ی سرشماری قرار گرفته و بقیه‌ی کارگاه‌ها (۱۵ کارکن تا ۶۱ کارکن) در ناحیه‌ی نمونه‌گیری قرار می‌گیرند. با این تقسیم‌بندی اندازه‌ی ناحیه‌ی سرشماری برابر ۹۱ ( $N_C$ ) و اندازه‌ی ناحیه‌ی نمونه‌گیری برابر ۱۰۶۰ ( $N_S$ ) به دست می‌آید.



ناحیه‌ی سرشماری = ۹۱      ناحیه‌ی نمونه‌گیری = ۱۰۶۰      ناحیه‌ی حذف = ۳۳۳

که در نتیجه وزن‌های مربوط به هر ناحیه به صورت زیر تعریف می‌شود:

$$W_E = \frac{333}{1484} \quad W_S = \frac{1060}{1484} \quad W_C = \frac{91}{1484}$$

### ۱-۲-۳- تعیین اندازه‌ی نمونه در رهیافت حذف

بعد از تعیین مرز بین ناحیه‌ها و تعداد اعضای ناحیه‌ها می‌توان تعداد اندازه‌ی نمونه را محاسبه نمود. در ناحیه‌ی نمونه‌گیری تعداد اندازه‌ی نمونه با استفاده از رابطه‌ی زیر تعیین می‌شود [چون هیچ اطلاعی در زمینه‌ی ارزش افزوده در دسترس نیست از اطلاعات مربوط به تعداد کارکنان (متغیر کمکی) استفاده می‌شود]

$$n \cong \frac{W_S^2 S_{X_S}^2}{\frac{W_S^2 S_{X_S}^2}{N_S} + (\text{MSE} - W_E^2 \bar{X}_E^2)} + N_C$$

یعنی مقدار  $n_S$  برابر است با:

$$n_S \cong \frac{W_S^r S_{X_S}^r}{\frac{W_S^r}{N_S} S_{X_S}^r + (\text{MSE} - W_E^r \bar{X}_E^r)}$$

با استفاده از برآورد واریانس متغیر کمکی در ناحیه‌ی نمونه‌گیری ( $S_{X_S}^r = 99/6957$ ) مقدار اندازه‌ی نمونه برابر است با:  $n_S \cong 48/7623 \cong 49$  و در نتیجه تعداد کل کارگاه‌های مورد بررسی برابر است با:  $n = 49 + 91 = 140$ .  $49$  در این مرحله نمونه‌ای به اندازه‌ی  $49$  از تعداد کارکنان کارگاه‌های صنعتی که در ناحیه‌ی نمونه‌گیری قرار گرفته‌اند، سپس میانگین ارزش افزوده‌ی کارگاه‌هایی که در نمونه انتخاب شده‌اند و واریانس این آماره را برآورد می‌کنیم:

$$\begin{aligned} \bar{y}_S &= 1121/271 && \text{برآورد میانگین ارزش افزوده در ناحیه‌ی نمونه‌گیری} \\ s_S^r &= 1129583 && \text{برآورد واریانس میانگین ارزش افزوده در ناحیه‌ی نمونه‌گیری} \\ \bar{Y}_C &= 5398/813 && \text{میانگین ارزش افزوده‌ی ناحیه‌ی سرشماری} \end{aligned}$$

در نهایت برآوردگر برشی نوع سوم در رهیافت حذف به صورت زیر محاسبه می‌شود:

$$\begin{aligned} \hat{Y}_{\text{cut}} &= W_C \bar{Y}_C + W_S \bar{y}_S \rightarrow \hat{Y}_{\text{cut}} = 1131/967 \\ \text{var}(\hat{Y}_{\text{cut}}) &= W_C^r \text{var}(\bar{Y}_C) + W_S^r \text{var}(\bar{y}_S) \rightarrow \text{var}(\hat{Y}_{\text{cut}}) = 10171/12 \end{aligned}$$

در عبارت بالا، واریانس ناحیه‌ی سرشماری به علت سرشماری صفر می‌باشد. برآورد ارائه شده در بالا به علت حذف تعدادی از اعضای جامعه دارای اریبی است. برای محاسبه‌ی این اریبی عمل نمونه‌گیری با استفاده از نرم افزار R به تعداد ۱۰۰۰ بار، به گونه‌ای که در هر مرحله نمونه‌ای به اندازه‌ی  $49$  از ناحیه‌ی نمونه‌گیری گزینش شده و میانگین ارزش افزوده را در این ناحیه محاسبه نماید، انجام می‌شود. میانگین برآورد میانگین این ۱۰۰۰ نمونه عبارت است از:

$$\hat{E}(\bar{y}) = \hat{Y}_S = 1122/357$$

و برآورد میانگین نمونه‌گیری برشی نوع سوم به صورت زیر محاسبه می‌شود:

$$\hat{Y}_{\text{cut}} = \hat{E}(\bar{Y}_{\text{cut}}) = W_C E(\bar{Y}_C) + W_S \hat{E}(\bar{y}_S) = 1132/743.$$

از داده‌های مربوط به جامعه میانگین ارزش افزوده را برای هر کارگاه محاسبه می‌کنیم:

$$\bar{Y}_N = 1271/655.$$

قدرمطلق اریبی برآورد میانگین عبارت است از:

$$| \text{Bias}(\hat{Y}_{\text{cut}}) | = | W_C \bar{Y}_C + W_S \hat{Y}_S - \bar{Y}_N | = 438/4731,$$

در نهایت MSE برآورد میانگین به صورت زیر برآورد می‌شود:

$$\hat{\text{MSE}}(\hat{Y}_{\text{cut}}) = \frac{1}{(1000-1)} \sum_{i=1}^{1000} (\bar{y}_i - \hat{Y}_S)^2 + W_E \bar{Y}_E^2 = 51167/42.$$

### ۲-۳-۲ روش مدل‌یار

با حذف بخشی از جامعه (ناحیه‌ی حذف) اطلاعاتی که برای برآورد مورد نظر ممکن است بسیار تأثیرگذار باشند از بین می‌روند. برای جلوگیری از این امر می‌توان با به دست آوردن رابطه‌ای میان داده‌هایی که در ناحیه‌ی حذف و داده‌هایی که در مابقی جامعه قرار دارند به نحوی از اطلاعاتی که در ناحیه‌ی حذف قرار دارند استفاده کرد.

در این روش در ابتدا از اطلاعات سرشماری سال قبل یعنی سال ۱۳۸۴ استفاده کرده و بر اساس تفکیک‌پذیری تقریبی جامعه به سه ناحیه‌ی حذف، نمونه‌گیری و سرشماری با توجه به نظر گروه کارشناسی و یا اطلاعات در دست (در این طرح از تقسیم‌بندی موجود به وسیله‌ی رهیافت حذف استفاده شده است و همچنین همان‌طور که در ابتدا توضیح دادیم به دلیل عدم دسترسی به اطلاعات سال ۸۴ از اطلاعات مربوط به سال ۸۵ استفاده می‌کنیم) رابطه‌ای میان عنصرهای ناحیه‌ها به دست می‌آوریم. این رابطه به صورت زیر تعریف می‌شود:

$$\delta = \frac{N_E \bar{Y}_E}{N_C \bar{Y}_C + N_S \bar{Y}_S} = 0/1338388$$

میانگین‌های ناحیه‌ی حذف، نمونه‌گیری و سرشماری متغیر ارزش افزوده‌ی مربوط به سرشماری کارگاه‌های با ۱۰ نفر کارکن و بیش‌تر سال ۱۳۸۵ به‌صورت زیر به دست می‌آیند:

$$\bar{Y}_E = ۶۲۰ / ۷۸۷۱۲۳$$

$$\bar{Y}_S = ۱۱۲۱ / ۸۱۲۸۳۴$$

$$\bar{Y}_C = ۴۴۹۹ / ۲۲۷۸۴۴$$

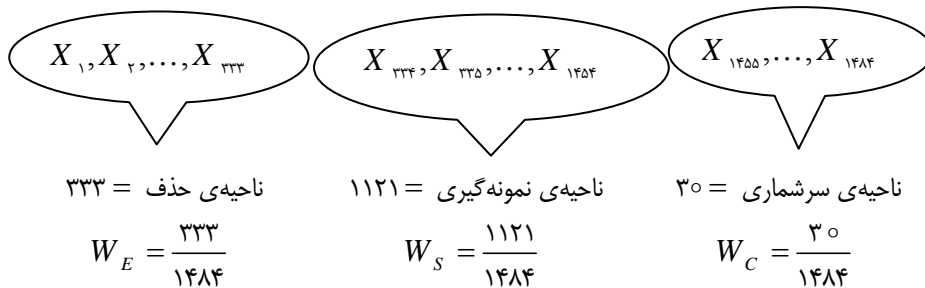
همچنین  $\tilde{\delta}$  که رابطه‌ی میان سه ناحیه برای متغیر کمکی تعداد کارکنان را تعریف می‌کند به‌صورت زیر محاسبه می‌شود:

$$\tilde{\delta} = \frac{N_E \bar{X}_E}{N_C \bar{X}_C + N_S \bar{X}_S} = ۰ / ۱۲۹۳۱۲$$

که میزان  $\tilde{\delta}$  با توجه به اطلاعات موجود در چارچوب سال ۱۳۸۵ برای متغیر کمکی تعداد کارکنان به دست می‌آید. ( چون اطلاعات متغیر کمکی تعداد کارکنان در چارچوب سال ۱۳۸۵ موجود می‌باشد از این اطلاعات استفاده شده است). در این حالت از فرمول تعیین آستانه‌ی برش برای رهیافت مدل‌بار استفاده کرده تا آستانه‌ی برش میان ناحیه‌ی نمونه‌گیری و سرشماری تعیین شود:

$$X_S^* = \bar{X}_{CS} + \sqrt{\frac{N^2}{N - N_E} \left( \frac{\text{MSE} - (\tilde{\delta} - \delta)^2 (\bar{X}_N - W_E \bar{X}_E)^2}{(1 + \tilde{\delta})^2} \right)} + S_{CS}^2 = ۱۰۶ / ۶۰۶۸$$

یعنی کارگاه‌هایی که دارای ۱۰۶ کارکن و بیش‌تر می‌باشند در ناحیه‌ی سرشماری قرار گرفته و بقیه‌ی کارگاه‌ها (۱۵ کارکن تا ۱۰۵ کارکن) در ناحیه‌ی نمونه‌گیری قرار می‌گیرند. با این تقسیم‌بندی اندازه‌ی ناحیه‌ی سرشماری برابر  $30 (N_C)$  و اندازه‌ی ناحیه‌ی نمونه‌گیری برابر  $۱۱۲۱ (N_S)$  به دست می‌آید.



### ۱-۲-۳- تعیین اندازه‌ی نمونه در رهیافت مدل‌یار

در این مرحله که ناحیه‌ها مشخص و  $\tilde{\delta}$  محاسبه شده است، با استفاده از رابطه‌ی زیر می‌توان اندازه‌ی نمونه ( $n_s$ ) را محاسبه کرد:

$$n_s = \frac{W_S^r S_S^r}{\frac{W_S^r S_S^r}{N_S} + \psi} = 26 / 3826 \cong 27$$

برآورد واریانس متغیر کمکی در ناحیه‌ی نمونه‌گیری عبارت است از:

$$S_S^r = 239 / 4546$$

با توجه به مقدارهای  $\tilde{\delta}$  و  $\delta$  مقدار  $\psi$  به صورت زیر محاسبه می‌شود:

$$\psi = \frac{\text{MSE} - (\tilde{\delta} - \delta)^r (W_C \bar{Y}_C + W_S \bar{Y}_S)^r}{(1 + \tilde{\delta})^r} = 4 / 548351$$

و در نتیجه تعداد کل کارگاه‌های گزینش شده به صورت زیر محاسبه می‌شود:

$$n = \frac{W_S^r S_S^r}{\frac{W_S^r S_S^r}{N_S} + \psi} + N_C = 27 + 30 = 57$$

در ادامه نمونه‌ای به اندازه‌ی ۲۷ از ناحیه‌ی نمونه‌گیری به دست آورده و بر اساس مقادیر ارزش افزوده‌ی نمونه‌های به دست آمده برآوردگر برشی نوع سوم با رهیافت مدل‌یار به صورت زیر به دست می‌آید:

$$\begin{aligned} \bar{Y}_E &= 620 / 787123 && \text{میانگین ناحیه‌ی حذف} \\ \bar{y}_S &= 1570 / 834047 && \text{میانگین ناحیه‌ی نمونه‌گیری} \\ \bar{Y}_C &= 7874 / 445123 && \text{میانگین ناحیه‌ی سرشماری} \\ \hat{Y}_{cut} &= (1 + \tilde{\delta})(W_C \bar{Y}_C + W_S \bar{y}_S) = 1812 / 260612 && \text{و در نهایت برآوردگر برشی نوع سوم} \end{aligned}$$

واریانس برآوردگر برشی نوع سوم در رهیافت مدل‌یار به صورت زیر برآورد می‌شود:

$$\text{var}(\hat{Y}_{cut}) = (1 + \tilde{\delta})^2 W_S^2 (1 - f_S) \frac{s_S^2}{n_S} = 54542 / 7$$

که در آن  $s_S^2 = 2619207$ .

همانند قبل برای یافتن آریبی عمل نمونه‌گیری به اندازه‌ی ۲۷ را ۱۰۰۰ بار تکرار کردیم که در نهایت نتایج زیر حاصل شد:

$$\begin{aligned} \hat{E}(\hat{y}) &= \hat{Y}_S = 1296 / 183 \\ \hat{Y}_{cut} &= E(\hat{Y}_{cut}) = (1 + \tilde{\delta})(W_C \bar{Y}_C + W_S \hat{Y}_S) = 1591 / 108 \\ \bar{Y}_N &= 1271 / 655 \end{aligned}$$

و بالاخره قدر مطلق آریبی عبارت است از:

$$|\text{Bias}(\hat{Y}_{cut})| = |E(\hat{Y}_{cut}) - \bar{Y}_N| = |(\tilde{\delta} - \delta)(W_C \bar{Y}_C + W_S \hat{Y}_S)| = 216 / 0778$$

در این حالت با استفاده از برآورد واریانس برآوردگر و میزان آریبی، MSE به صورت زیر برآورد می‌شود:



$$\begin{aligned} \text{MSE}(\hat{Y}_{\text{cut}}) &= \frac{1}{(1000-1)} \sum_{i=1}^{1000} (\bar{y}_i - \hat{Y}_S)^2 \\ &+ (\bar{\delta} - \delta)^2 (W_C \bar{Y}_C + W_S \hat{Y}_S)^2 = 49355/48 \end{aligned}$$

با توجه به نتیجه‌های به دست آمده میزان اهمیت تعیین آستانه‌ی برش و نیز اندازه‌ی نمونه‌ی مورد نظر در این طرح به وضوح قابل مشاهده است که چگونه با تغییر دادن آستانه‌های برش برای ناحیه‌های مختلف نتیجه‌های به دست آمده تغییر می‌کنند. همچنین چگونگی تغییر نتیجه‌ها با استفاده از دو نوع رهیافت حذف و مدل‌یار نیز ملاحظه شد. در این بخش هدف این است که با در نظر گرفتن شرایط مورد نظر (محدودیت زمانی یا هزینه و...) تعیین شود که استفاده از کدام روش نتیجه‌ی دقیق‌تری را حاصل می‌نماید؟

با استفاده از نمونه‌گیری برشی نوع سوم و رهیافت مدل‌یار، جامعه به سه قسمت تقسیم می‌شود که نمونه‌ای به اندازه‌ی ۲۷ از ناحیه‌ی نمونه‌گیری به طور تصادفی انتخاب کرده، همراه با ۳۰ کارگاه که در ناحیه‌ی سرشماری وجود دارد، مورد پرسش قرار می‌گیرند، یعنی جمعاً باید به ۵۷ کارگاه مراجعه کرد به جای این که به ۱۴۸۴ کارگاه تولید آجر مراجعه شود. که در نهایت میانگین ارزش افزوده‌ی آن‌ها برابر ۱۸۱۲/۲۶۰۶۱۲ میلیون ریال می‌شود. در حالتی دیگر اگر نتوان رابطه‌ای میان ناحیه‌های حذف، نمونه‌گیری و سرشماری پیدا کرد و یا دستیابی به این رابطه مبتنی بر صرف زمان و هزینه‌ی بیش‌تری باشد از رهیافت حذف استفاده می‌شود. در مطالعه‌ی شبیه‌سازی نمونه‌ای به اندازه‌ی ۴۹ از ناحیه‌ی نمونه‌گیری گرفته به همراه ۹۱ کارگاه که در ناحیه‌ی سرشماری قرار داشتند، یعنی به طور کل از ۱۴۰ کارگاه پرسش به عمل می‌آید و در نهایت میانگین ارزش افزوده برای این روش برابر ۱۱۳۱/۱۷۳ میلیون ریال می‌شود.

برای مقایسه‌ی دو روش ذکر شده، همان‌طور که در فصل قبل توضیح داده شد اگر  $0 \leq \frac{\bar{\delta}}{\delta} \leq 2$  و یا  $\bar{\delta} \cong 0$  در این صورت اندازه‌ی میانگین توان دوم خطای برآوردگر مورد نظر در رهیافت مدل‌یار کم‌تر از رهیافت حذف است و در نتیجه این روش دقیق‌تر می‌باشد. در این طرح با توجه به اینکه نسبت

$$\frac{\tilde{\delta}}{\delta} = \frac{0/129312}{0/13382388} = 0/9662,$$

مقداری بین صفر و ۲ است. رهیافت مدل یار نتایج بهتری را ارائه کرد. یعنی با توجه به محاسبات انجام شده میانگین توان دوم خطای رهیافت مدل یار ۴۹۳۵۵/۴۸ است که در مقایسه با میانگین توان دوم خطای رهیافت حذف (۵۱۱۶۷/۴۲) مقداری کم تر می باشد. در نتیجه در مطالعه‌ی شبیه سازی رهیافت مدل یار بهتر عمل کرده است. میزان قدر مطلق اریبی رهیافت مدل یار برابر ۲۱۶/۰۷۷۸ شده که در مقایسه با قدر مطلق اریبی در رهیافت حذف (۴۳۸/۴۷۳۱) کم تر می باشد. در نهایت می توان گفت در صورت استفاده از رهیافت مدل یار چون از داده های ناحیه ی برش داده شده نیز استفاده می شود برآوردهای حاصل دارای میانگین توان دوم خطا و همچنین اریبی کم تری از رهیافت حذف می باشند. البته لازم به ذکر است که این نتیجه تحت شرایط  $2 \leq \frac{\tilde{\delta}}{\delta} \leq 0$  یا  $\tilde{\delta} \equiv 0$  حاصل می شود.

### ۳-۳- بحث و نتیجه گیری

در نهایت برای ارزیابی نتایج به دست آمده از روش نمونه گیری برشی نوع سوم بر اساس رهیافت مدل یار، نتایج به دست آمده از این روش را با برآورد نسبتی از روش نمونه گیری تصادفی مقایسه می کنیم. به این ترتیب نمونه ای به اندازه ی ۵۷ از جامعه به روش تصادفی ساده بدون جایگذاری می گیریم و بر اساس این نمونه میانگین و واریانس آن را برآورد می کنیم و این عمل را ۱۰۰۰ بار تکرار می کنیم در نهایت برآورد نسبتی زیر حاصل می شود:

$$\hat{Y} = 1270/373$$

$$\text{var}(\bar{Y}) = 922/5216$$

$$|\text{Bias}(\hat{Y})| = 129/49038$$

$$\text{MSE}(\bar{Y}) = 17563/452$$

که در مقایسه با مقدار اریبی که برای نمونه گیری برشی نوع سوم داشتیم مقدار کم تری است.

..... گزیده مطالب آماری، سال ۲۰، شماره ی ۱، بهار و تابستان ۱۳۸۸، صص ۹۵-۱۲۲ .....

همچنین مقدار MSE برآوردگر نسبتی (۱۷۵۶۳/۴۵۲) نیز مقداری کم‌تر از MSE برآوردگر برشی نوع سوم بر اساس مدل‌یار (۴۹۳۵۵/۴۸) می‌باشد. در نهایت می‌توان گفت با توجه به نمونه‌ی گرفته شده، نمونه‌گیری برشی نتایج ضعیف‌تری به دست داده است.

### توضیحات

۱. شاخص قیمت تولیدی بخش صنعت در سطح هر یک از کدهای چهار رقمی بخش صنعت
۲. سطح جغرافیایی برآوردها، کل کشور است.
۳. شاخص قیمت عبارت است از متوسط تغییرات متناسب قیمت مجموعه مشخصی از کالاها و خدمات که بین دو دوره‌ی زمانی بوجود می‌آید.

### مرجع‌ها

- [۱] پورمحمود، مریم (۱۳۸۷). نمونه‌گیری برشی با یک کاربرد. پایان‌نامه‌ی کارشناسی ارشد، دانشگاه علامه طباطبائی.
- [2] Bee, M.; Benedetti, R., and Espa, G. (2007). A framework for cut off sampling in business survey design.
- [3] Haan, J. De.; Opperdoes, E. and Schut, C. M. (1999). Item selection in the consumer price index: Cut-off versus probability sampling. *Survey methodology*, **25**, 31-41.
- [4] Hidiroglou, M. A. (1986), The construction of a self representing stratum of large units in survey design. *The American Statistician*, **40**, 27-31.
- [5] Knaub, J. R., Jr. (2007). Cut off sampling. *Encyclopedia of Survey Research Methods*.
- [6] Sarndal, C. E.; Swensson, B. and Wretman, J. (1997). *Model Assisted Survey Sampling*. Springer, New york.

**مریم پور محمود**

کارشناس ارشد آمار

تهران، خیابان شهید بهشتی، نبش احمد قصیر، دانشکده‌ی اقتصاد دانشگاه علامه طباطبائی، گروه آمار.

رایانشانی: poormahmoud@yahoo.com

**حمیدرضا نواب پور**

دکتر آمار

تهران، خیابان شهید بهشتی، نبش احمد قصیر، دانشکده‌ی اقتصاد دانشگاه علامه طباطبائی، گروه آمار

رایانشانی: h.navvabpour@srta.ac.ir