

# طرح بررسی و تعیین روش جانمایی و برآورد اطلاعات کارگاه های صنعتی

قدرت اله روشنایی

محمدرضا اناری،

مرجان نورینی،

امیرکسیان،

گلنازه احسنی

کرده پژوهشی طرح های فنی و روش های آماری

پژوهشکده های آمار

تابستان ۱۳۹۳



## پیش‌گفتار

وجود خطا در داده‌های حاصل از آمارگیری‌ها امری اجتناب‌ناپذیر است. این خطاها به صورت اطلاعات گزارش نشده، اطلاعات غیر مجاز، اطلاعات ناسازگار و ... بروز می‌کنند. خطاها به هر دلیل که رخ دهند، باید در گام اول شناسایی و در گام بعدی به بهترین شیوه اصلاح شوند و هدف از جانمایی اصلاح این خطاها است. بی‌پاسخی، خطایی ناشی از عدم مشاهده است که بازتاب تلاشی ناموفق در به‌دست آوردن اطلاعات مورد نظر از یک واحد واجد شرایط است. به‌طور کلی، دو نوع خطای اصلی بی‌پاسخی وجود دارد: بی‌پاسخی واحد و بی‌پاسخی سؤال.

بی‌پاسخی واحد هنگامی رخ می‌دهد که هیچ پاسخی از واحد نمونه دریافت نشود اعم از این که واحد نمونه خانوار، فردی از افراد خانوار و یا یک کارگاه کسب و کار باشد.

در بسیاری از آمارگیری‌ها، به‌ویژه در آمارگیری از کارگاه‌های صنعتی، بی‌پاسخی‌ها مسئله عمومی و جدی هستند. مقادیر بی‌پاسخ در داده‌های آمارگیری منجر به اریبی در برآورد پارامترها می‌شود به‌طوری‌که نمی‌توان به‌درستی تصویر جامعه را نشان داد. روش‌های گوناگونی برای بررسی و جبران داده‌های بی‌پاسخ در آمارگیری وجود دارد:

- نمونه‌گیری جزئی<sup>۱</sup> در زمان اجرا
  - تعدیل وزن‌ها برای جبران اریبی برآوردهای حاصل از بی‌پاسخی در آمارگیری
  - برخی روش‌های جانمایی از قبیل جانمایی مدل مبنا (جانمایی رگرسیونی، نسبتی و میانگین) که هنوز هم برای جایگذاری مقادیر گم‌شده با مقادیر قابل قبول مورد استفاده قرار می‌گیرند.
- یکی از مشکلات آمارگیری‌های کارگاهی، به‌ویژه آمارگیری از کارگاه‌های صنعتی، بی‌پاسخی‌های کلی و جزئی است. در روش فعلی مورد استفاده برای بی‌پاسخی، بر اساس اطلاعات سال‌های قبل کارگاه‌های بی‌پاسخ که از نتایج طرح کارگاه‌های صنعتی بزرگ وزارت صنایع و معادن در سال ۱۳۸۱ و ۱۳۸۰ در سطح کدهای ۴ رقمی فعالیت به‌دست می‌آید، ضرایب افزایش یا کاهش محاسبه

---

<sup>1</sup> Subsampling

شده و در اطلاعات سال قبل کارگاه بی‌پاسخ ضرب شده و اطلاع در سال جدید به دست می‌آید. با توجه به این که با گذشت ۹ سال، خطا در ضرایب در نظر گرفته شده، افزایش پیدا کرده لذا لزوم بازنگری در روش فوق امری ضروری است. علاوه بر این به کارگیری روشی که در عین قابلیت کاربرد و به کارگیری بالاترین کارایی و بیشترین دقت را نیز داشته باشد حائز اهمیت است. برای شناسایی بهترین روش و کاربست‌ها در جهانی لازم است که مطالعات تطبیقی و الگو برداری صورت پذیرد. این گزارش شامل چهار فصل است.

در فصل اول به بررسی روش‌های مورد استفاده برای جبران بی‌پاسخی در طرح آمارگیری صنعتی یا طرح‌های مشابه در کشورها و سازمان‌های منتخب شامل کانادا، بریتانیا، فیلیپین، کمیته فدرال روش‌شناسی، تایلند، سوئد، هلند، ایتالیا، آمریکا، نیوزیلند و پروژه یورو ادیت پرداخته می‌شود. در فصل دوم روش‌های جهانی شرح داده می‌شود و فصل سوم به بررسی طرح‌های انجام شده مربوط به جهانی در پژوهشکده آمار و سوابق روش‌های مورد استفاده در مرکز آمار ایران در جهانی کارگاه‌های صنعتی آورده شده است. فصل چهارم به اعمال روش‌های جهانی منتخب از بین روش‌های جهانی مورد بررسی بر روی کارگاه‌های بی‌پاسخ طرح صنعت در سال ۱۳۸۸ پرداخته خواهد شد.

**گروه پژوهشی طرح‌های فنی و روش‌های آماری**

## فهرست مطالب

صفحه	عنوان
	فصل اول: مطالعات تطبیقی و الگوبرداری
۱-۱-۱-۱	مقدمه
۴-۲-۱-۱	کانادا
۴-۲-۱-۲-۱	آمارگیری بنگاهی یکپارچه شده
۷-۲-۲-۱	آمارگیری از دارایی‌های بنگاه‌های کوچک و متوسط
۱۰-۲-۳-۱	آمارگیری سالانه‌ی صنعت و برش و تولید چوب (ASML)
۱۸-۲-۴-۱	آمارگیری ماهانه از صنعت
۱۹-۳-۱	بریتانیا
۲۱-۴-۱	فیلیپین
۲۶-۵-۱	نیوزیلند
۲۶-۵-۱-۱	آمارگیری از عملکرد کارگاه‌ها BOS
۳۲-۵-۲-۱	آمارگیری اقتصادی فصلی از صنعت (QESM)
۳۳-۶-۱	کمیته‌ی فدرال روش‌شناسی آماری
۳۸-۷-۱	تایلند
۴۱-۸-۱	مرکز آمار فنلاند
۴۴-۹-۱	مرکز آمار سوئد
۴۵-۱۰-۱	مرکز آمار ایتالیا
۴۵-۱۱-۱	دفتر سرشماری امریکا
۴۶-۱۲-۱	مرکز آمار هلند
۴۸-۱۳-۱	نروژ
۵۶-۱۴-۱	پروژه یورو ادیت

## فصل دوم: روش‌های جانهی

۶۹	مقدمه
۷۲	۱-۲- جانهی استنتاجی
۷۳	۲-۲- جانهی با میانگین کل
۷۳	۳-۲- جانهی با میانگین‌های سلولی
۷۵	۴-۲- جانهی بادرنگ
۷۶	۵-۲- جانهی بی‌درنگ متوالی
۷۶	۶-۲- جانهی متوالی سلسله مراتبی
۷۷	۷-۲- جانهی بی‌درنگ قطعی
۷۸	۸-۲- جانهی بی‌درنگ تصادفی
۷۸	۹-۲- جانهی کلی تصادفی
۷۹	۱۰-۲- جانهی تصادفی درون رده‌ای
۸۰	۱۱-۲- جانهی بی‌درنگ نزدیک‌ترین همسایه دنباله‌ای (جانهی بی‌درنگ سستی)
۸۱	۱۲-۲- جانهی رگرسیون
۸۲	۱۳-۲- جانهی نزدیک‌ترین همسایه
۸۶	۱۴-۲- جانهی نسبتی
۸۷	۱۵-۲- روش‌های پیوندی یا نیمه‌پارامتری
۸۸	۱۶-۲- جانهی جورشدگی چندمتغیره
۸۸	۱۷-۲- جانهی تصادفی ساده
۸۹	۱۸-۲- جانهی رگرسیون مبتنی بر نزدیک‌ترین همسایه
۹۱	۱۹-۲- جانهی چندگانه
۹۲	۲۰-۲- جانهی انبوه
۹۹	۲۱-۲- جانهی رگرسیون پیش‌بینی‌کننده
۹۹	۲۲-۲- جانهی $k$ -همسایگی میانگین
۱۰۰	۲۳-۲- جانهی میانه

## فصل سوم: پیشینه موجود در کشور

مقدمه	۱۰۳
۱-۳- طرح جهانی بی‌پاسخی‌ها و اثرات آن در برآوردها و تجزیه و تحلیل داده‌ها	۱۰۵
۲-۳- روش‌های جهانی در طرح آمارگیری از کارگاه‌های صنعتی	۱۱۲
۱-۲-۳- بی‌پاسخی جزئی	۱۱۵
۱-۱-۲-۳- جهانی در بخش ۳ پرسش‌نامه	۱۱۵
۲-۱-۲-۳- جهانی در بخش‌های ۴ و ۵ پرسش‌نامه	۱۱۸
۳-۱-۲-۳- جهانی در بخش ۸ پرسش‌نامه	۱۱۹
۴-۱-۲-۳- جهانی در بخش ۹ پرسش‌نامه	۱۲۰
۲-۲-۳- بی‌پاسخی کلی	۱۲۱
۱-۲-۲-۳- جهانی اقلام بخش ۳ پرسش‌نامه	۱۲۴
۲-۲-۲-۳- جهانی اقلام بخش ۴ پرسش‌نامه	۱۲۵
۳-۲-۲-۳- جهانی اقلام بخش ۵ پرسش‌نامه	۱۲۶
۴-۲-۲-۳- جهانی در بخش ۶ پرسش‌نامه	۱۲۹
۵-۲-۲-۳- جهانی در بخش ۷ پرسش‌نامه	۱۲۹
۶-۲-۲-۳- جهانی در بخش ۸ پرسش‌نامه	۱۲۹
۷-۲-۲-۳- جهانی در بخش ۹ پرسش‌نامه	۱۳۰
۸-۲-۲-۳- جهانی در بخش ۱۰ پرسش‌نامه	۱۳۰
۹-۲-۲-۳- نتیجه‌گیری کلی	۱۳۱

## فصل چهارم: بررسی و اجرای روش‌های جهانی منتخب

مقدمه	۱۳۵
۱-۴- تعریف بی‌پاسخی کلی در طرح آمارگیری صنعت	۱۳۵
۲-۴- بررسی کیفیت داده‌ها	۱۳۶
۳-۴- خوشه‌بندی	۱۴۴
۱-۳-۴- متغیرهای مورد استفاده در خوشه‌بندی	۱۴۵
۲-۳-۴- روش مورد استفاده برای خوشه‌بندی	۱۴۶

- ۳-۳-۴- تعداد خوشه‌ها..... ۱۴۷
- ۴-۴- روش‌های مورد استفاده برای جانمایی اطلاعات کارگاه‌های بی‌پاسخ..... ۱۵۴
- ۴-۵- سناریو مورد استفاده برای جانمایی اطلاعات کارگاه‌های بی‌پاسخ..... ۱۵۸
- ۴-۶- ارزیابی روش‌های منتخب مورد استفاده در این طرح برای جانمایی و مقایسه آن‌ها با نتایج جانمایی مرکز آمار ایران..... ۱۳۵
- ۴-۷- جمع‌بندی، یافته‌ها و پیشنهادات..... ۱۷۴

#### پیوست‌ها

- پیوست یک: نمونه مقادیر جانمایی شده به روش‌های منتخب بر حسب سناریو و استان..... ۱۸۱
- پیوست دو: ماکروهای برنامه SAS برای جانمایی بخش‌های مختلف پرسش‌نامه با سناریوهای مختلف..... ۳۵۱
- پیوست سه: چکیده مطالعات تطبیقی..... ۳۶۹
- مرجع‌ها..... ۳۷۷



## فهرست جدول‌ها

صفحه	عنوان
۱۸	جدول ۱-۱: وضعیت فایل داده بعد از ادیت و جانپی در آمارگیری سالانه از کارگاه‌ها در فیلیپین
۱۹	جدول ۲-۱: روش‌های جانپی مورد استفاده برای بی‌پاسخی جزئی و کلی در آمارگیری کارگاه‌ها در فیلیپین
۲۲	جدول ۳-۱: سازمان‌های مشارکت کننده در آمارگیری از عملکرد کارگاه‌ها در نیوزیلند
۳۹	جدول ۴-۱: تعداد واحدهای جامعه و نمونه در هر طبقه در طرح آمارگیری کسب و کار نروژ
۴۳	جدول ۵-۱: نتایج شبیه‌سازی با روش جانپی تصادفی با یک مدل گروه‌های همگن پاسخ
۷۴	جدول ۱-۲: نرخ‌های بی‌پاسخی در NPSAS
۷۷	جدول ۲-۲: مقایسه توزیع‌های وزنی قبل و پس از جانپی در جانپی انبوه
۹۸	جدول ۱-۳: اقلام کلیدی برای جانپی بی‌پاسخی کلی به تفکیک بخش‌های پرسش‌نامه طرح صنعت
۱۱۰	جدول ۱-۴: فراوانی بی‌پاسخی کلی در طرح صنعت سال ۸۸ به تفکیک کد دورقمی
۱۱۱	جدول ۲-۴: فراوانی بی‌پاسخی کلی به تفکیک کدهای ۴ رقمی در طرح صنعت ۸۸ به ترتیب نزولی
۱۱۳	جدول ۳-۴: تعداد موارد دارای صفر ساختاری بر حسب متغیرها در ۴۳۰۶ کارگاه در سال ۱۳۸۸
۱۱۵	جدول ۴-۴: تعداد موارد دارای صفر ساختاری بر حسب متغیرهای برحسب کد دو رقمی در ۴۳۰۶ کارگاه در سال ۱۳۸۸
۱۲۱	جدول ۵-۴: توزیع فراوانی کدهای فعالیت در خوشه‌ها با تعداد خوشه ۵
۱۲۲	جدول ۶-۴: فراوانی کارگاه‌ها در هر یک از خوشه‌ها در کدهای ۱۵، ۱۶، ۱۷ و ۱۸
۱۲۷	جدول ۷-۴: اقلام مورد نظر برای جانپی در طرح صنعت به تفکیک بخش‌های پرسشنامه
۱۲۸	جدول ۸-۴: عناوین اقلام متناظر با متغیرهای مورد استفاده در جانپی
۱۳۵	جدول ۹-۴: متغیرهای وابسته و مستقل در روش رگرسیونی بر حسب بخش‌های پرسش‌نامه
۱۳۷	جدول ۱۰-۴: فراوانی کارگاه‌های موجود در بورس به تفکیک کد دورقمی
۱۳۷	جدول ۱۱-۴: فراوانی کارگاه‌های موجود در بورس دارای فراوانی بیشتر از ۸ به تفکیک کد چهار رقمی
۱۴۱	جدول ۱۲-۴: تعداد کدهای چهاررقمی دارای قدرمطلق تفاوت نسبی در میانگین کمتر از ۱۵ درصد بر حسب سناریو و
۱۴۶	جدول ۱۳-۴: روش‌های جانپی مورد استفاده در طرح صنعت



## فصل اول

### مطالعات تطبیقی و الگوبرداری

#### ۱-۱- مقدمه

بی پاسخی، خطایی ناشی از نبود مشاهده است که بازتاب تلاشی ناموفق در به دست آوردن اطلاعات مورد نظر از یک واحد آماری واجد شرایط است. موضوع خطای بی پاسخی و اریبی بی پاسخی در برآوردهای آمارگیری همواره به عنوان موضوعی مهم در پژوهش های آمار رسمی تلقی شده است. بی پاسخی کلی هنگامی رخ می دهد که هیچ پاسخی از واحد نمونه دریافت نشود اعم از این که واحد نمونه خانوار، فردی از افراد خانوار و یا یک کارگاه کسب و کار باشد. با وجود اعمال بیشترین تلاش ها در طی گردآوری داده ها، احتمال بروز سطوحی از بی پاسخی همواره وجود دارد. برای حداقل کردن اثر خطای بی پاسخی واحد، اغلب از روش های وزن دهی استفاده می شود. بی پاسخی پرسش نیز بر کیفیت داده ها اثر می گذارد. برای بهبود نرخ های پاسخ در پرسش های پرسش نامه به ویژه در مواقعی که اقلام آماری مهم مطرح است، می توان از تماس مجدد یا روش های پیگیری استفاده کرد. معیارهای این تلاش ها می توانند به عنوان نماگرهایی از کیفیت داده ها تلقی شوند. برای بقیه ی بی پاسخی های پرسش می توان روش های جانهی را مورد استفاده قرار داد تا بی پاسخی پرسش جبران شود. این روش ها داده های گم شده را به طور کامل جبران نمی کنند و به هر حال سطوح نامشخصی از خطای ناشی از بی پاسخی در داده ها باقی خواهد ماند.

توجه به این نکته مهم است که بی‌پاسخی کامل و پرسش به‌طور کامل متمایز نیستند. اگر پرسش‌نامه ی آمارگیری با پاسخ‌هایی فقط برای قلم‌ها، ولی نه قلم‌های مهم و اساسی، بازگردد، می‌توان آن را به عنوان بی‌پاسخی واحد تلقی کرد.

دو رهیافت کلی برای جبران بی‌پاسخی وجود دارد: تعدیل به عنوان بخشی از فرایند برآوردکردن (مانند تعدیل‌های وزنی) و برآورد مستقیم مقداری برای هر واحد بدون پاسخ که آن واحد در صورت پاسخ‌دادن ممکن بود گزارش کند (جانهی). در بسیاری از متون نظری بیان شده که تعدیل‌های وزنی فقط در بی‌پاسخی کلی مورد استفاده قرار می‌گیرند در حالی که شیوه‌های جانهی بیش‌تر برای جبران قلم‌های گم‌شده به‌کار می‌روند. برای برخی آمارگیری‌های دوره‌ای، شیوه‌های جانهی برای برآورد همه‌ی قلم‌های آمارگیری در مورد تک‌تک پاسخ‌گویان مورد استفاده قرار می‌گیرند.

تعدیل وزن بی‌پاسخی، مستلزم دسترسی به اطلاعاتی درباره‌ی بی‌پاسخ‌ها و پاسخ‌دهنده‌ها و همین‌طور فرض‌هایی درباره‌ی تفاوت‌های بین پاسخ‌دهندگان و بی‌پاسخ‌هاست. در بیشتر موارد غیر از اطلاعات موجود در چارچوب نمونه‌گیری، اطلاع اندکی از واحد بی‌پاسخ وجود دارد. هم‌چنین در مورد مطالعات پانلی می‌توان از داده‌ها و اطلاعات مربوط به مصاحبه‌ی اولیه برای انجام تعدیل‌های بی‌پاسخی استفاده کرد. داده‌های طولی حاصل از آمارگیری‌ها را برای بررسی ویژگی‌های بی‌پاسخ‌های پانلی و ارزیابی روش‌های تعدیل وزن بی‌پاسخی مورد استفاده قرار داده‌اند.

از آن‌جایی که وزن دادن برای بی‌پاسخی بر کیفیت برآوردهای آمارگیری تأثیرگذار است، تحقیق در زمینه‌ی اثر تعدیل‌های بی‌پاسخی بر برآوردهای آماری برای سازمان‌ها حائز اهمیت است. ارزیابی روش‌های تعدیل بی‌پاسخی، مستلزم وجود برخی منابع بیرونی از داده‌های مشابهی است که در مطالعات اریبی بی‌پاسخی مورد استفاده قرار می‌گیرد.

چون جانهی برای بی‌پاسخی قلم می‌تواند کیفیت برآوردهای حاصل از یک آمارگیری را تحت تأثیر قرار دهد، تحقیق درباره‌ی اثر این روش‌ها نیز برای سازمان‌های آماری مهم است. برای ارزیابی اثر جانهی بر داده‌ها، چندین روش اساسی وجود دارد. یک روش برای تعیین اثر جانهی آن است که

برآوردهایی که شامل مقادیر جهانی شده هستند با برآوردهایی که بر مبنای داده‌های گزارش شده به دست آمده‌اند، مقایسه شوند. مقایسه‌ی برآوردها با منابع بیرونی نیز از دیگر روش‌های تعیین اثرهای جهانی است.

کلی‌ترین تعریفی که از جهانی می‌توان ارائه کرد «اختصاص داده به یک فیلد به دلیل بی‌پاسخی یا برای تغییر یک داده‌ی ثبت شده که بر اساس مجموعه‌ای از قواعد ادیت، ناسازگار تشخیص داده شده است». به هر حال اصلاح داده‌ها به هر روشی که انجام شود باید به گونه‌ای باشد که کمترین تغییرات ممکن در داده‌های ثبت شده صورت گیرد، ناسازگاری‌های اطلاعات رفع شود، و برای اقلام نادرست یا گم‌شده با استفاده از سایر اطلاعات، مقدار مجازی درج شود. نتیجه‌ی جهانی، مجموعه‌ی داده‌ای حاوی رکوردهای قابل قبول و سازگار برای تمامی اقلام اطلاعاتی مربوط به واحدهای بی‌پاسخ است.

در این فصل به بررسی روش‌های مورد استفاده برای جبران بی‌پاسخی در طرح آمارگیری صنعتی یا طرح‌های مشابه در کشورها و سازمان‌های منتخب پرداخته می‌شود. شایان ذکر است در میان کشورهای مورد بررسی تنها در کشورهای کانادا، نیوزیلند، آمریکا، سوئد و فیلیپین طرح آمارگیری ویژه بخش صنعت انجام شده و در سایر کشورها از جمله تایلند، ایتالیا و انگلیس طرح آمارگیری کارگاهی که صنعت زیر مجموعه‌ای از آن است انجام می‌شود. طرح آمارگیری صنعتی در فیلیپین به صورت سالانه، در نیوزیلند به صورت فصلی و در کانادا به صورت ماهانه اجرا می‌شود. در پروژه یورو ادیت که کشور هلند در آن مشارکت داشته علاوه بر روش‌های جهانی مورد استفاده در دو طرح آمارگیری کارگاهی، شاخص‌های استاندارد ارزیابی روش‌های جهانی بیان شده است.

آنچه که مسلم است انطباق کامل بین طرح آمارگیری صنعتی در مرکز با طرح آمارگیری صنعتی در سایر کشورها امری ناممکن است و موجب محدود شدن مطالعات تطبیقی و الگوبرداری می‌شود. روش‌های جبران بی‌پاسخی افزون بر نوع متغیر به عوامل متعدد دیگری از جمله طرح نمونه‌گیری، سطح برآورد و ... بستگی دارد. هم‌چنین مرز بین بی‌پاسخی جزئی و کلی همواره از مسائل مورد بحث بوده است. علاوه بر این اطلاعات بیشتری نیز در خصوص بخش‌های پرسش‌نامه و

طبقه‌بندی‌های مورد استفاده در طرح‌های سایر کشورها در این فصل آورده شده که امکان مقایسه را بهتر فراهم می‌نماید.

## ۱-۲- کانادا

مرکز آمار کانادا با استفاده از ۷۰۰ پرسش‌نامه مجزا در حدود ۲۰۰ آمارگیری کارگاهی در سال اجرا می‌کند. این آمارگیری‌ها در حوزه‌هایی مانند کشاورزی، حمل و نقل و صنعت انجام و در یک برنامه آمارگیری اصلی که آمارگیری بنگاهی یکپارچه شده<sup>۱</sup> نامیده می‌شود، ادغام می‌شوند. در این طرح یکپارچه، داده‌های تفصیلی‌تر صنعت و کالا در سطح ایالتی گردآوری می‌شود.

### ۱-۲-۱- آمارگیری بنگاهی یکپارچه شده

این آمارگیری، یک آمارگیری سالانه و بخشی از پروژه بهبود آمار اقتصادی ایالتی محسوب می‌شود و در آن اطلاعات بنگاه‌ها گردآوری می‌شود. در این آمارگیری برآوردهای ایالتی با کیفیت بالا تولید می‌شود و از این برآوردها برای باز توزیع مالیات‌های ایالتی استفاده می‌شود. در این آمارگیری نتایج حاصل از بخش‌های فعالیت‌های اقتصادی (تولید، ساختمان، سرمایه‌گذاری و...) یکپارچه می‌شوند. طرح نمونه‌گیری، یک طرح نمونه‌گیری دو فازی است و از پرسش‌نامه‌های پستی و پیگیری‌های تلفنی در این آمارگیری استفاده می‌شود. جانهای کلی و جزئی برای جبران بی‌پاسخی مورد استفاده قرار می‌گیرد. برای بی‌پاسخی کلی، داده‌های مالیاتی برای اجرای قسمت عمده جانهای از طریق تطبیق به کار برده می‌شود. برای جانهای جزئی، جانهای نزدیک‌ترین همسایه مورد استفاده قرار می‌گیرد. در مورد بی‌پاسخی جزئی از سامانه<sup>۲</sup> GEIS برای تعیین فیلدها و اجرای جانهای نزدیک‌ترین همسایه استفاده می‌شود. در گام برآورد، از برآوردهای رگرسیونی تعمیم‌یافته برای تولید برآورد در سطح حوزه استفاده می‌شود. اخیراً دقت برآورد (واریانس) با فرض این که بی‌پاسخی وجود ندارد از

---

<sup>1</sup> Integrated Enterprise Survey

<sup>2</sup> Generalized Edit and Imputation System

طریق مجموعه داده کامل انجام می‌شود اما سامانه SIMVAR برای برآورد واریانس ناشی از جانهی در حال طراحی است.

### سامانه ادیت و جانهی تعمیم یافته

این سامانه از اواسط دهه ۸۰ در اداره آمار کانادا طراحی و مورد استفاده قرار می‌گیرد. این سامانه در ابتدا برای تامین نیازهای جانهی و ادیت در آمارگیری‌های اقتصادی و برای متغیرهای پیوسته طراحی شد. این سامانه مبتنی بر Oracle است و در بستر کامپیوترهای بزرگ و Unix قابل اجراست. ویژگی‌های اصلی این سامانه عبارت‌اند از:

- ادیت

- موضع‌یابی خطا

- جانهی

اولین ماژول برای ادیت به‌منظور تعریف و بررسی قواعد ادیتی مورد استفاده قرار می‌گیرد. در GEIS قواعد ادیتی باید روابط برابری‌ها یا نابرابری‌های خطی باشند. تعریف رده‌های ادیت در هر سطحی با استفاده از توابعی از متغیرهای خاص امکان‌پذیر است. این سامانه هم‌چنین تعدادی از تشخیص‌ها مانند نقاط گرانگین به‌وسیله ادیت تعریف شده، ادیت ضمنی<sup>۱</sup>، ادیت مازاد<sup>۲</sup>، و نقاط پرت تولید می‌کند. هم‌چنین آمارهای مربوط به نرخ‌های ناموفق ادیت توسط سامانه تولید می‌شود.

دومین ماژول، تابع موضع‌یابی خطاست. این بخش از سامانه میزان تغییرات یک رکورد برای برقراری شرایط ادیت را تعیین می‌کند. سامانه مبتنی بر اصل حداقل تغییرات فلگی و هالت (۱۹۷۶) است. اصول فلگی و هالت عبارت‌اند از:

- اصل حداقل تغییرات

- سنجش دقت رسمی

- بهبود فرایند

---

<sup>1</sup> Implied edit

<sup>2</sup> Redundant Edit

- جانهی به عنوان فعالیت مدل سازی

- مکانیسم پاسخ شرطی

سومین بخش سامانه، مربوط به جانهی است. سه نوع جانهی، یعنی جانهی منطقی، جانهی پیشگویی و جانهی اهدا کننده در سامانه وجود دارد.

جانهی منطقی هنگامی مورد استفاده قرار می گیرد که تنها یک متغیر پس از ادیت موضوعی گم شده است و می تواند از طریق تفریق کردن از سایر متغیرها به دست آید.

روش های جانهی پیشگویی شامل تمام روش هایی است که یک مقدار به دست آمده از طریق تابعی از متغیرهای موجود در مجموعه پاسخ برای نمونه جانهی می شود. جانهی نسبتی، میانگین و مقدار قبلی را می توان به عنوان مثال هایی از این روش نام برد.

بالاخره، جانهی اهدا کننده، روشی است که در آن مقادیری از رکورد دیگر در رکورد دارای مقادیر گم شده جانهی می شوند. لازم به ذکر است که روش های جانهی بی درنگ و جانهی نزدیک ترین همسایه در سامانه GEIS موجود هستند. در جانهی به روش نزدیک ترین همسایه تعداد زیادی از «فیلدهای جور<sup>۱</sup>» می توانند به دست آیند. این فیلدها برای یافتن نزدیک ترین رکورد به رکورد گیرنده که نیارمند جانهی است مورد استفاده قرار می گیرند. فاصله مورد استفاده برای یافتن نزدیک ترین همسایه از طریق یک سری تبدیل ها از متغیرهای جور به دست می آیند. برای هر متغیر تبدیل ها به صورت زیر انجام می شوند:

- داده ها به صورت صعودی مرتب می شوند.

- یک رتبه به هر داده اختصاص می یابد.

- رتبه ها در مقیاس صفر تا یک استاندارد می شوند.

سپس نزدیک ترین همسایه برای رکورد دارای مقادیر گم شده، رکوردی است که حداقل مقدار کمیت زیر را در تمام واحدهای مجموعه پاسخ داشته باشد:

$$\text{MAX}\{ |Z_{D1} - Z_{R1}|, |Z_{D2} - Z_{R2}|, \dots, |Z_{DP} - Z_{RP}| \}$$

---

<sup>1</sup> Matching Fields



که در آن  $Z_{DP}$  متغیر کمکی  $P$  از اهداکننده و  $Z_{RP}$  متغیر کمکی  $P$  از گیرنده است.

## چارچوب جانھی

طی یک دوره زمانی، چارچوب‌هایی برای جانھی طراحی شده است. بسیار مهم است که بدانیم چه چارچوبی را هنگام انتخاب استراتژی جانھی برای درک خواص روش جانھی به کار می‌بریم. یک طبقه‌بندی ممکن برای این چارچوب‌ها عبارت است از:

- کارآزموده-مبنا<sup>1</sup>: کارشناسان جانھی بر اساس دانش خود انجام می‌دهند.
- توزیع-مبنا: توزیع‌ها برآورد می‌شوند و مقادیر جانھی از توزیع به دست می‌آیند.
- مدل-مبنا: مدلی ساخته و تایید می‌شود و برای تولید مقادیر جانھی مورد استفاده قرار می‌گیرد.
- فراوانی-مبنا: تحت یک مکانیسم پاسخ، مقادیر گم‌شده بدون استفاده از مدل، جانھی می‌شوند.
- تجربه-مبنا<sup>2</sup> (اهدا کننده-مبنا): یک اهداکننده پیدا می‌شود و مقادیر آن برای جانھی مورد استفاده قرار می‌گیرد. این چارچوب‌ها برای تدوین خواص نظری روش‌های جانھی بسیار سودمند هستند.

### ۱-۲-۲- آمارگیری از دارایی‌های بنگاه‌های کوچک و متوسط

این آمارگیری برای بررسی دارایی‌ها و سرمایه‌گذاری مورد استفاده در بنگاه‌های کوچک و متوسط و گردآوری اطلاعاتی در زمینه تلاش‌های اخیر برای به دست آوردن سرمایه‌گذاری‌های جدید طراحی شده است. این آمارگیری به صورت موردی انجام می‌شود و یک آمارگیری دوره‌ای نیست. آخرین بار این آمارگیری در سال ۲۰۰۷ انجام شده است.

---

<sup>1</sup> Experience-based

<sup>2</sup> Empirical-based

هدف از اجرای این آمارگیری گردآوری ویژگی‌های عمومی بنگاه‌های کوچک و ابتکارات سرمایه‌گذاری آن‌هاست. این آمارگیری اطلاعاتی درباره انواع بدهی، کرایه یا اجاره و بالابردن سرمایه از طریق فروش سهام نیز گردآوری می‌کند. اداره آمار کانادا از طرف اداره مالیه و صنعت کانادا این آمارگیری را به‌عنوان بخشی از یک برنامه پژوهشی بزرگتر درباره بنگاه‌های کوچک و بزرگ که اخیراً آغاز شده است انجام می‌دهد.

### جامعه هدف

جامعه هدف این آمارگیری تمام بنگاه‌های موجود در فایل جامع ثبت کارگاهی را شامل می‌شود. بنگاه‌های زیر از جامعه کنار گذاشته می‌شوند:

- ۱- بنگاه‌های دارای ۵۰۰ کارکن و بیشتر
- ۲- بنگاه‌های دارای درآمد ناخالص بیش از ۵۰ میلیون دلار
- ۳- بنگاه‌هایی که به‌صورت غیرانتفاعی کدگذاری می‌شوند (مدارس، بیمارستان‌ها، موسسات خیریه و ...)
- ۴- تعاونی‌ها
- ۵- سرمایه‌گذاری مشترک
- ۶- شهرداری‌ها/ دولت فدرال
- ۷- بنگاه‌های دارای فعالیت‌های اقتصادی خاص که به‌وسیله نظام طبقه‌بندی فعالیت‌های آمریکای شمالی<sup>۱</sup> (NAICS) تعیین شده‌اند. این بنگاه‌ها شامل تاسیسات (۲۲)، بیمه و مالیه (۵۲) مدیریت شرکت‌ها و بنگاه‌ها (۵۵)، خدمات آموزش (۶۱)، اداره عمومی (۹۱)، اجاره و کرایه تجهیزات ماشینی (۵۳۲۱)، مراکز مراقبت‌های سرپایی (۶۲۱۴)، اجاره و کرایه تجهیزات و ماشین‌های صنعتی و تجاری (۵۳۲۴)، آزمایشگاه‌های تشخیص طبی و پزشکی (۶۲۱۹)، بیمارستان‌های جراحی و پزشکی عمومی (۶۲۲۱)، بیمارستان‌های ترک سوء مصرف مواد و

---

<sup>1</sup> North American Industry Classification System

روانپزشکی (۶۲۲۲)، بیمارستان‌های خاص (به جز بیمارستان‌های ترک سوء مصرف مواد و روانپزشکی) (۶۲۲۳) خدمات اضطراری، مسکن و غذا و سایر خدمات امدادی (۶۲۴۲).  
توجه به این نکته ضروری است که سایر استثنائات مانند شرکت‌های تابعه براساس اطلاعات چارچوب قابل شناسایی نیست. چنین شرکت‌هایی در مرحله اجرا غربال می‌شوند.

### نمونه‌گیری

این نمونه‌گیری، یک نمونه‌گیری مقطعی<sup>۱</sup> است. ثبت کارگاهی اداره آمار کانادا به عنوان چارچوب نمونه‌گیری برای جامعه هدف تمام بخش خصوصی، بنگاه‌های انتفاعی با کمتر از ۵۰۰ کارکن و درآمد کمتر از ۵۰ میلیون دلار را مورد استفاده قرار می‌دهد. چارچوب نمونه‌گیری شامل ۱۹۹۹۰۰۰ بنگاه است. طبقه‌بندی اولیه بر اساس منطقه، اندازه و طول مدت فعالیت کارگاه و مشارکت در برنامه تضمین قانون سرمایه‌گذاری (مالیه) بنگاه‌های کوچک کانادا انجام می‌شود. تعداد کارکن بنگاه، برای تعریف اندازه بنگاه مورد استفاده قرار می‌گیرد و طول مدت فعالیت کارگاه با استفاده از تاریخی که بنگاه در ثبت کارگاهی وارد شده است برآورد می‌شود. نمونه‌ای از بنگاه‌ها مبتنی بر طبقه‌بندی از ثبت کارگاهی به تصادف انتخاب می‌شوند. اندازه نمونه شامل ۳۷۰۵۸ کارگاه بوده است.

پاسخ‌گویی در این آمارگیری اختیاری (داوطلبانه) است. داده‌ها به‌طور مستقیم از پاسخ‌گویان دریافت می‌شوند. گردآوری داده‌ها از طریق روش‌های مصاحبه تلفنی رایانه یار (CATI)<sup>۲</sup> انجام می‌شود.

### جانهی

سامانه جانهی نزدیک‌ترین همسایه<sup>۳</sup> در این آمارگیری اجرا می‌شود. این روش با جستجوی یک اهدا کننده<sup>۴</sup> و یک گیرنده با اندازه و صفات مشابه سروکار دارد. مقادیر داده‌ای برای متغیرهای ناکامل و گم شده در رکورد گیرنده از اهدا کننده جایگزین می‌شود.

---

<sup>1</sup> Cross Sectional survey

<sup>2</sup> Computer Assisted Telephone Interviewing

<sup>3</sup> Nearest Neighborhood Imputation

<sup>4</sup> Donor

## برآورد

پارامترهای مورد نظر از طریق سامانه برآورد تعمیم‌یافته اداره آمار برآورد می‌شوند. وزن‌های نمونه اولیه با در نظر گرفتن امتناع از پاسخ‌گویی‌ها و سایر بی‌پاسخی‌ها تعدیل می‌شوند. وزن‌ها با در نظر گرفتن واحدهایی که فعال هستند اما تماس یا پیدا کردن آن‌ها امکان‌پذیر نیست نیز تعدیل می‌شوند. برآوردها در بیش از ۲۰۰ حوزه مورد بررسی بر اساس متغیرهای طبقه‌بندی (مانند منطقه و فعالیت) و علاوه بر آن متغیرهای پرسش‌نامه (مانند تعداد کارکن، نوع بستانکاری) به دست می‌آیند. یک برآوردگر پس‌طبقه‌بندی<sup>۱</sup> برای کالیبره کردن تعداد کل بنگاه‌ها در هر طبقه تعداد کارکن مورد استفاده قرار می‌گیرد. تعداد کل بنگاه‌ها از ثبت کارگاهی با در نظر گرفتن نرخ‌های خارج از حوزه به دست می‌آیند.

### ۱-۲-۳- آمارگیری سالانه‌ی صنعت و برش و تولید چوب<sup>۲</sup> (ASML)

آمارگیری سالانه‌ی صنایع که شامل ۲۳ آمارگیری است به عنوان بخشی از برنامه آمارگیری متحد بنگاه‌های اقتصادی (UES)<sup>۳</sup> اداره‌ی آمار کانادا، به‌طور سالانه به گردآوری اطلاعاتی درباره‌ی بخش صنعت و رشته فعالیت برش و تولید چوب در کانادا می‌پردازد. این آمارگیری همه‌ی کارگاه‌هایی که به فعالیت‌های صنعتی و برش و تولید چوب می‌پردازند و همچنین دفاتر فروش و انبارهایی که این کارگاه‌ها را حمایت می‌کنند، را پوشش می‌دهد. اطلاعات گردآوری شده شامل آمارهای اصلی صنعت از جمله درآمد، اشتغال، هزینه‌ی مواد اولیه و ملزومات مورد استفاده، هزینه‌ی انرژی، آب و تأسیسات، موجودی‌ها، فروش کالا و همچنین تولید و مصرف کالا است. جامعه‌ی هدف آمارگیری سالانه‌ی صنعت، شامل همه‌ی کارگاه‌هایی است که بیشتر به فعالیت‌های صنعت و برش و تولید چوب می‌پردازند.

---

<sup>۱</sup> Post Stratification

<sup>۲</sup> Annual Survey of Manufactures and logging

<sup>۳</sup> United Enterprise Survey

داده‌ها براساس سامانه طبقه‌بندی رشته فعالیت‌های آمریکای شمالی، طبقه‌بندی شده‌اند. تحت این سامانه طبقه‌بندی، کارگاه‌های برش و تولید چوب متعلق به رشته فعالیت ۱۱۳۳ و کارگاه‌های صنعتی متعلق به رشته فعالیت بخش‌های ۳۱، ۳۲ و ۳۳ هستند. این طرح یک آمارگیری نمونه‌ای مقطعی است. چارچوب مورد استفاده برای این طرح بر اساس ثبت‌های کل کسب و کارهای اداره‌ی آمار کانادا<sup>۱</sup> است و واحد آماری، کارگاه است.

به منظور کاهش بار پاسخ‌گویی و صرفه‌جویی در هزینه‌های آمارگیری، جامعه‌ی هدف ASM به دو بخش اصلی با استفاده از آستانه‌های تحمل از قبل تعیین شده تقسیم شده است:

۱- بخش جامعه‌ی آمارگیری: شامل واحدهایی است که نمونه‌گیری می‌شوند و اطلاعات و داده‌های مورد نیاز برای این کارگاه‌ها، از طریق پرسش‌نامه جمع‌آوری می‌شود.

۲- بخشی از جامعه‌ی هدف که در جامعه‌ی آمارگیری نیست: که شامل واحدهایی است که برای دریافت پرسش‌نامه انتخاب نشده‌اند.

جامعه آمارگیری به ۴ بخش، Must-Take، Take-All، Take-Some1، Take-Some2 تقسیم می‌شوند. کارگاه‌هایی که در طبقه‌ی Must-Take قرار دارند، بر اساس پیچیدگی و اندازه‌ی آن کارگاه‌ها توسط تحلیل‌گران موضوعی از قبل مشخص می‌شوند. محدوده‌های طبقات Take-All، Take-Some1 و

Take-Some2 به وسیله‌ی یک متغیر که ماکسیمم بین درآمد ناخالص واحد کسب و کار (GBI)<sup>۲</sup> و کالاها است، استخراج می‌شوند. همه‌ی واحدها در طبقات Must-Take و Take-All از آمارگیری انتخاب می‌شوند در حالی که تنها کسری از واحدها در دو طبقه‌ی Take-Some انتخاب می‌شوند. علاقمندان به کسب اطلاعات بیش‌تر به (لبراسیر و ترامله، ۲۰۰۷) مراجعه کنند.

تشخیص خطا همواره بخشی جدایی‌ناپذیر در مراحل گردآوری و پردازش داده است. ادیت خودکار به منظور شناسایی خطاها در دو سطح، برای داده‌های ثبت شده در حین گردآوری داده‌ها به کار

<sup>۴</sup> Statistics Canada Business Register

<sup>۱</sup> Gross Business Income

می‌رود. سطح اول ادیت، بلافاصله پس از ورود داده‌ها برای همه‌ی داده‌های گردآوری شده اعمال می‌شود که بیشتر بدون تماس با پاسخ‌دهندگان رفع می‌شود. این ادیت‌ها، خطای بالقوه را بر اساس تغییرات متغیرهای کلیدی کل و نسبت (به عنوان مثال کل درآمد به هزینه) طی سال‌ها که از یک آستانه‌ی تحملی بالاتر باشد را تشخیص می‌دهد. همچنین ناسازگاری‌های موجود در داده‌ها (به عنوان مثال هزینه‌های کل با جمع هزینه‌های جزیی برابر نباشد) را تشخیص می‌دهد. این نوع ادیت مشکلات کدگذاری محصولات توسط پاسخگویان را تشخیص می‌دهد. سطح دوم ادیت مشابه سطح اول است با این تفاوت که آستانه‌های تحمل برای خطاها، خیلی دقیق‌تر تعریف شده است و تعداد قواعد ادیت بیش‌تر است. این ادیت تنها برای رکوردهایی اعمال می‌شود که برای رسیدن به سطح پوشش مورد نظر، لازم است. در حالت کلی این رکوردها اثر زیادی روی برآورد نهایی می‌گذارند.

قبل از جانهی، کارشناسان موضوعی از ابزار متنوعی برای شناسایی و برخورد با داده‌های پرت بر اساس تغییرات سال‌های مختلف و ارتباط بین متغیرهای مالی در داده‌های آمارگیری و داده‌های اداری استفاده می‌کنند. این مرحله با بررسی گزینشی داده‌های خرد شروع می‌شود به گونه‌ای که برخی از رکوردهایی که ممکن است اثر معنی‌داری روی برآوردهای نهایی آمارگیری به دلیل زیاد بودن سهم آن کارگاه صنعتی در یک رشته فعالیت یا در یک استان و یا به دلیل استفاده از آن در جانهی داشته باشند، انتخاب و بررسی می‌شوند. برای رفع مشکلات و ایرادهای داده‌های مالی، از داده‌های ثبتی و اطلاعات قبلی کارگاه مورد بررسی در سال‌های قبل، اطلاعات موجود از آمارگیری‌های مربوط دیگر، اطلاعات خاص کسب و کار از منابع خارجی (گزارش‌های سالانه، انتشارات بازرگانی) و متوسط متغیر مورد بررسی در زمان حال و یا سال‌های قبل برای کلیه‌ی کسب و کارهای مربوط در رشته فعالیت‌های مشابه، استفاده می‌شود. برای اطلاعات مربوط به بخش کالا، قبل از انجام جانهی، از کدگذاری صحیح کالاهایی که دارای ارزش بالایی هستند، اطمینان حاصل می‌شود.

پس از جانهی، تحلیل گران موضوعی، به مقابله‌ی برآوردهای مالی از طرح ASML و برآوردهای مالی آمارگیری ماهانه‌ی صنعت (MSM)<sup>۱</sup> می‌پردازند. در صورت مشاهده‌ی مغایرت، داده‌های مربوط به همه‌ی واحدهای مشابه در آمارگیری ASML و MSM مورد بررسی قرار می‌گیرند و اطمینان از سازگاری همه‌ی کسب و کارهایی موجود در هر دو آمارگیری حاصل شود و همچنین تغییرات سال به سال نقاط دور افتاده در برآوردهای مربوط به متغیرهای کلیدی و اصلی و سهم این نقاط دور افتاده در متغیرهای انتخاب شده در سطح کد ۶ رقمی NAICS در سطح استان مشخص می‌شود.

### سامانه جانهی خودکار در طرح ASML

همان‌طور که قبل از این اشاره شد، جانهی برای تعیین مقادیر قابل قبول همه‌ی متغیرهایی که گم‌شده هستند و یا تعدیل متغیرهایی که بین مقادیر آنها در داده‌های گردآوری شده و داده‌های ثبتی ناسازگاری وجود دارد، استفاده می‌شوند. روش‌های آماری متعددی برای این هدف با استفاده از داده‌های گردآوری شده در آمارگیری جاری و منابع اطلاعاتی کمکی استفاده شده است. منابع اطلاعات کمکی شامل داده‌های آمارگیری قبلی و داده‌های ثبتی است. در این آمارگیری، جانهی متغیرهای مالی یا ارزشی و متغیرهای کالایی به‌طور مجزا صورت می‌گیرد.

برخی از روش‌های مختلف که برای جانهی داده‌های گم‌شده و یا ناسازگار در متغیرهای مقدار برای آمارگیری ASML مورد استفاده قرار می‌گیرند به شرح زیر است. ساده‌ترین روش، استفاده از گروهی از قوانین قطعی و منسجم است که تمام روابط و روابط قابل قبول بین متغیرها را بیان می‌کند و مقادیر گم‌شده باقی‌مانده را استخراج می‌کند. مقادیر گم‌شده با اطلاعات و داده‌های ثبت‌شده‌ی جزئی‌تر با استفاده از نسبت بین متغیرها و روندهای قبلی مشاهده شده در داده‌ها جانهی می‌شوند. روش جانهی مورد استفاده دیگر، جایگذاری مستقیم متغیرهای گم‌شده با استفاده از داده‌های ثبتی و داده‌های قبلی است. آخرین روش جانهی، جانهی اهداکننده یا دهنده<sup>۲</sup> است که در

---

<sup>1</sup> Monthly Survey of Manufacturing

<sup>2</sup> Donor Imputation

این روش یک واحد پاسخ‌گو که مشابه واحد بی‌پاسخ است بر اساس اطلاعاتی که برای هر دو واحد کسب و کار موجود است، شناسایی می‌شود و اطلاعات موجود از واحد پاسخ‌گو برای واحد بی‌پاسخ در نظر گرفته می‌شود.

جانهی بر اساس روندهای قبلی و جانهی دهنده (برای واحدهای جدید در ASML) برای استخراج اطلاعات گم‌شده کالایی، استفاده می‌شوند. هیچ اطلاعات ثبتی برای استفاده در جانهی داده‌های کالایی وجود ندارد.

روش‌های جانهی نسبتی به‌طور وسیع برای جانهی آمارگیری سالانه‌ی صنعت کانادا استفاده می‌شود. به‌عنوان مثال، یک روش شامل به‌دست آوردن نسبت بین متغیر مورد نیاز برای جانهی و یک متغیر کمکی، با استفاده از گروهی از واحدهای واجد شرایط در جانهی است. این نسبت برای واحدهایی که نیازمند جانهی هستند برای تولید مقدار جانهی شده استفاده می‌شود. به‌منظور بهبود کیفیت داده‌های جانهی شده، علاوه بر طبقه‌بندی رشته فعالیت و مکان جغرافیایی، یک متغیر اندازه‌ی نمونه برای ساخت طبقات جانهی معرفی می‌شود به گونه‌ای که واحدهای با اندازه‌های مختلف بتوانند به‌طور مناسب معرفی شوند.

جانهی نسبتی حالت خاصی از جانهی رگرسیونی است که به‌طور وسیع در آمارگیری‌های کسب و کار به دلیل متغیرهای اقتصادی که به شدت با هم وابستگی دارند مورد استفاده قرار می‌گیرد اما به دلیل این‌که داده‌های کسب و کار و تجارت به شدت چاوله هستند، کیفیت جانهی نسبتی نسبت به انتخاب نادرست مدل، آسیب‌پذیر است. برای رفع این مشکل، واحدها اغلب به گروه‌های همگن که کلاس‌ها یا طبقات جانهی نامیده می‌شوند، تقسیم می‌شوند به گونه‌ای که بتوان مدل‌های مختلفی را به‌طور مستقل به کلاس‌ها و طبقات مختلف اختصاص داد.

در ادامه بیان می‌شود که چگونه متغیر اندازه می‌تواند برای ساخت و ایجاد رده و طبقات جانهی برای جانهی آمارگیری سالانه‌ی صنعت آمار کانادا مورد استفاده قرار گیرد.



پس از دریافت پرسش‌نامه‌ی همه‌ی کارگاه‌های انتخاب شده و گردآوری داده‌ها و برخی ادیت‌های اولیه، داده‌های سالانه صنعت کانادا توسط نرم افزار BANFF<sup>۱</sup> پردازش می‌شود.

در آمارگیری سالانه‌ی صنعت، ۹۰ متغیر، ادیت و جانهی می‌شوند. این سامانه جانهی و ادیت خودکار توسط فراداده‌ها به وجود آمده است. در این سامانه از روش‌های مختلف جانهی، شامل جایگذاری مستقیم از طریق داده‌های ثبتي (داده‌های مالیات) یا داده‌های آمارگیری ماهانه‌ی صنعت (MSM)، جانهی تاریخی (جانهی اولیه)، جانهی نسبتی سال جاری و جانهی دهنده استفاده می‌شود. ۷ متغیر کلیدی و اصلی درآمد کل<sup>۲</sup>، هزینه‌های کل<sup>۳</sup>، کل درآمد عملیاتی<sup>۴</sup>، کل درآمدهای دیگر، کل خرید<sup>۵</sup>، کل موجودی انبار در اول سال<sup>۶</sup> و کل موجودی انبار آخر سال<sup>۷</sup> است. این متغیرها ممکن است به‌طور مستقیم به فایل داده‌های مالیاتی مرتبط شوند و از این طریق جانهی بوسیله‌ی جایگزینی مستقیم داده‌های مالیات امکان‌پذیر می‌شود. داده‌های مالیاتی برای ۸۳ متغیر جزئی‌تر دیگر در دسترس نیست. استراتژی جانهی کل این است که ابتدا جانهی ۷ متغیر کلیدی و اصلی صورت می‌پذیرد و سپس ۸۳ متغیر جزئی دیگر جانهی می‌شود.

در میان روش‌های مختلف جانهی، جانهی نسبتی سال جاری، روشی است که به‌طور وسیع در آمارگیری سالانه‌ی صنعت کانادا به‌خصوص برای متغیرهای جزئی استفاده می‌شود چرا که داده‌های مالیاتی و داده‌های آمارگیری ماهانه‌ی صنعت کانادا برای این متغیرها در دسترس نیست و برای بسیاری از واحدها، اطلاعات قبلی و پیشینه‌ی آنها گزارش نشده است. روش نسبتی سال جاری، شامل محاسبه‌ی نسبت بین متغیری که نیاز به جانهی دارد (Y) و متغیر کمکی (X) با استفاده از گروهی از واحدهای واجد شرایط در یک کلاس جانهی می‌باشد. این نسبت دوباره برای رکوردهایی که نیاز به جانهی دارد، برای تولید یک مقدار جانهی شده به‌کار می‌رود. این روش در BANFF،

---

<sup>۱</sup> سامانه جانهی و ادیت خودکار

<sup>۲</sup> Total Revenue

<sup>۳</sup> Total Cost/ expenses

<sup>۴</sup> Total Operational Revenue

<sup>۵</sup> Total Purchase

<sup>۶</sup> Total Opening Inventories

<sup>۷</sup> Total Closing

CURRATIO نامیده می‌شود. برای اجتناب از نسبت‌های برآورد شده‌ای که ناپایدار هستند، نیاز است که حداقل ۵ واحد واجد شرایط در کلاس جهانی قرار داشته باشد. واحدهای با مقادیر گم شده برای یک یا هر دو متغیر از محاسبه‌ی نسبت‌ها خارج می‌شوند. همچنین نقاط پرت نیز از داده‌ها خارج می‌شوند. طبقات و کلاس‌های با واحدهای واجد شرایط ناکافی، با هم ادغام می‌شوند تا واحدهای واجد شرایط را توسعه دهند.

در روش آمارگیری، طبقات جهانی اغلب به وسیله‌ی متغیرهای طبقه‌بندی و یا متغیرهای مورد استفاده برای سطوح انتشاراتی، تعیین می‌شوند. در آمارگیری سالانه صنعت کانادا، کلاس‌های جهانی اولیه که برای روش CURRATIO استفاده می‌شوند، به وسیله طبقه‌بندی رشته فعالیت و مناطق جغرافیایی تعیین می‌شوند. در این آمارگیری و آمارگیری‌های کسب و کار، این واقعیت وجود دارد که واحدهای بزرگ دارای ویژگی‌هایی متفاوت از واحدهای کوچک است. بنابراین استفاده از نسبت‌ها براساس واحدهای درون کلاس‌ها و طبقات جهانی که شامل ترکیبی از واحدهای بزرگ و کوچک باشد، منجر به برآورد نامناسب می‌شود. بنابراین به منظور بهبود کیفیت داده‌های جهانی، علاوه بر طبقه‌بندی رشته فعالیت و مناطق جغرافیایی، متغیر مقدار اندازه<sup>۱</sup>، برای ساخت کلاس‌های جهانی معرفی می‌شود به گونه‌ای که بیشتر معرف واحدهای با اندازه‌های مختلف بتواند باشند. متغیر مقدار اندازه‌ای که باید استفاده شود، باید به گونه‌ای باشد که برای همه‌ی واحدها، شامل واحدهای پاسخ‌گو و بی‌پاسخ موجود باشد. این متغیر همچنین باید یک متغیر قابل اعتماد برای تشخیص بین واحدهای کوچک و بزرگ باشد. در چارچوب نمونه‌گیری کارگاه‌های صنعتی کانادا، سه متغیر می‌تواند به عنوان متغیر مورد نظر انتخاب شود. این سه متغیر شامل متغیر ماکسیمم بین درآمد ناخالص کسب و کار (GBI) و کالا، متغیر طبقه‌بندی نمونه‌گیری، STRATA و وزن طرح نمونه-گیری است. در فایل کمکی نیز متغیر کل سود در نظر گرفته شده است. متغیر STRATA برای هدف طرح نمونه‌گیری از قبل ایجاد شده است که این متغیر دارای ۴ مقدار متناظر با ۴ طبقه‌ای که قبلاً معرفی شده است، می‌باشد. متغیر STRATA به عنوان یک متغیر اندازه، به دلیل این که در مواد

---

<sup>1</sup> Size measure variable

خاص بین کارگاه‌های کوچک و بزرگ نمی‌تواند تمایز قایل شود، نامناسب تشخیص داده شده است. به‌عنوان مثال، واحدهای با حجم کوچک در طبقه‌ی Must-Take ممکن است به‌عنوان Take-All در نظر گرفته شود. وزن‌های طرح نمونه‌گیری به‌طور ذاتی مشابه متغیر طبقه‌بندی نمونه‌گیری است به این مفهوم که واحدهای با اندازه‌های کوچک در طبقه‌ی Must-Take ممکن است با واحدهای با اندازه‌های بزرگ در طبقه‌ی Take-All ترکیب شود که هر دوی این گروه‌ها که با هم ترکیب شده‌اند دارای وزن طرح نمونه‌گیری ۱ می‌باشند. بنابراین این متغیر نیز به‌عنوان متغیر اندازه، قابل قبول در نظر گرفته نشده است.

برخلاف متغیر ماکسیمم بین درآمد ناخالص کسب و کار و کالا که برای تعیین محدوده‌ی طبقات نمونه‌گیری استفاده می‌شوند، متغیر بازده و سود کل از فایل کمکی که پس از گردآوری داده‌ها ساخته می‌شود، تکمیل می‌شود. این متغیر بیشتر براساس داده‌های مالیاتی و داده‌های قبلی کارگاه است. بعد از یک سری تحلیل‌ها، مشخص شد که این متغیر به‌عنوان متغیر مورد نظر، مناسب است. برای کسب اطلاعات بیشتر برای آشنایی با روش‌شناسی تعیین گروه‌ها و حدود متغیر مناسب برای تشکیل کلاس‌های جانمایی به مقاله‌ی لی از اداره‌ی آمار کانادا مراجعه شود که در ۵ مرحله با استفاده از رویه FASTCLUS<sup>۱</sup> در نرم‌افزار SAS به خوشه‌بندی کارگاه‌ها پرداخته است.

---

<sup>۱</sup> رویه FASTCLUS یک روش موثر برای یافتن خوشه‌های اولیه را با الگوریتم‌های تکراری استاندارد برای حداقل کردن مربع فاصله از میانگین خوشه‌ها تلفیق می‌کند.

## ۱-۲-۴- آمارگیری ماهانه از صنعت

هدف از این آمارگیری انتشار ارزش فروش کالاهای صنعتی، موجودی کالا و سفارشات است. داده‌های این طرح به‌عنوان نماگرهای شرایط اقتصادی بخش صنعت، به‌عنوان ورودی محصول ناخالص داخلی، به‌عنوان مؤلفه‌های نماگر مرکب اداره آمار کانادا و به‌عنوان ورودی تحلیل‌های اقتصادی و اقتصادسنجی مورد استفاده قرار می‌گیرد. نتایج این طرح هم در بخش‌های عمومی و خصوصی، بانک کانادا، صنعت کانادا، سامانه حساب‌های ملی و ... مورد استفاده قرار می‌گیرد. جامعه هدف این طرح شامل کارگاه‌های ثبت شده و ثبت نشده متعهد به فعالیت صنعتی است. طرح نمونه‌گیری در این طرح، نمونه‌گیری مقطعی است. ثبت‌های کارگاهی کانادا چارچوب نمونه‌گیری این طرح را فراهم می‌کند. چارچوب نمونه‌گیری شامل تمام کارگاه‌هایی است که در ثبت کارگاهی در بخش صنعت قرار گرفته‌اند. برای کاهش هزینه‌ها و بار پاسخ‌گویی واحدهای کوچک چارچوب تقریباً ۵ درصد (بر اساس ارزش فروش کالاهای صنعتی) در هر ایالت از آمارگیری کنار گذاشته شده‌اند. این راهبرد سبب می‌شود تا از ۹۸۰۰۰ کارگاه صنعتی فقط در حدود ۲۷۰۰۰ واحد امکان انتخاب در آمارگیری را داشته باشند. به‌طور تقریبی ۱۰۵۰۰ واحد از ۲۷۰۰۰ کارگاه چارچوب نمونه‌گیری به‌عنوان نمونه انتخاب می‌شوند.

قبل از انتخاب نمونه، چارچوب نمونه‌گیری به گروه‌های صنعت-ایالت تقسیم می‌شود. گروه‌های صنعت بر اساس نظام طبقه‌بندی امریکای شمالی (NAICS)<sup>۱</sup> ایجاد می‌شوند. بسته به تعداد کارگاه‌ها در هر سلول یا گروه، تقسیمات ریزتری بر اساس اندازه کارگاه انجام می‌شود که طبقه نامیده می‌شوند. اندازه کارگاه براساس ارزش سالانه اخیر فروش کالاهای صنعتی است.

جانهی برای هر صفت در پانل ثابت نمونه که داده‌ای برای آن‌ها گزارش نشده، اجباری است. موجودی کالا و سفارشات ثبت نشده در تمام رکوردهایی که در آن‌ها فروش کالاهای صنعتی از رکوردهای GST موجود است جانهی می‌شوند. برای سایر موارد جانهی‌ها به صورت خودکار با

---

<sup>1</sup> North American Industry Classification System

قیود خاص با استفاده از مقادیر ماه قبل، تغییرات ماه به ماه، سال به سال در واحدهای پاسخ‌گوی مشابه انجام می‌شود (Statistics Canada, 2011; Rancourt, 1999; Li, 2002).

### ۱-۳- بریتانیا

استعلام سالانه از کارگاه‌ها<sup>۱</sup>، آمارگیری است که در آن اطلاعات ساختار مالی کارگاه‌ها به صورت سالانه فراهم می‌شود. این آمارگیری در سال ۱۹۹۸ با ادغام چندین طرح سالانه که قبلاً در بخش‌های مختلف اقتصادی توسط اداره آمار انگلیس یا پیش از آن انجام می‌شد، شکل گرفت. این آمارگیری برآوردهای کل برای هر بخش اقتصادی با پوشش بخش‌های A تا O از طبقه‌بندی استاندارد فعالیت‌های اقتصادی (SIC) برحسب متغیرهای فهرست شده در زیر را ارائه می‌دهد:

- ۱- تعداد بنگاه‌ها
- ۲- کل عایدی یا درآمد
- ۳- ارزش افزوده ناخالص تقریبی به قیمت‌های پایه
- ۴- کل خدمات، مواد، خرید کالاها
- ۵- اشتغال کل
- ۶- اشتغال کل، متوسط در سال
- ۷- کل هزینه‌های اشتغال
- ۸- کل مخارج سرمایه‌ای خالص
- ۹- کل مخارج سرمایه‌ای خالص، تملک‌ها
- ۱۰- کل مخارج سرمایه‌ای خالص، در اختیار
- ۱۱- کل کالاهای ذخیره و در حال ساخت- ارزش در پایان سال
- ۱۲- کل کالاهای ذخیره و در حال ساخت- ارزش در ابتدای سال
- ۱۳- کل کالاهای ذخیره و در حال ساخت- افزایش در طول سال

---

<sup>1</sup> Annual Business Inquiry

نتایج در اینترنت با استفاده از جداولی که سری‌های زمانی را برای هر بخش اقتصادی و هر بخش انتخابی SIC<sup>۱</sup> ارائه می‌کند، منتشر می‌شود. برای گردآوری داده‌ها در سال ۲۰۰۶، ۴۲ پرسش‌نامه مختلف مورد استفاده قرار گرفت. این پرسش‌نامه‌ها برای هر بخش اقتصادی متفاوت هستند و برای چندین بخش (مانند خرده‌فروشی، خواربارفروشی و غیره) دو پرسش‌نامه وجود دارد: پرسش‌نامه بلند و پرسش‌نامه تفضیلی‌تر که برای گردآوری داده‌های کارگاه‌های بزرگ و نمونه‌ای از کارگاه‌های کوچک‌تر استفاده می‌شود و یک پرسش‌نامه کوتاه، که در آن اطلاعات کمتر تفضیلی از کارگاه‌های نمونه‌ای که پرسش‌نامه بلند را دریافت نکرده‌اند پرسیده می‌شود. طرح نمونه‌گیری نیز یک طرح نمونه‌گیری تصادفی ساده با طبقه‌بندی است. طبقه‌ها با ترکیبی از این سه متغیر ایجاد می‌شوند:

- منطقه (انگلیس، اسکاتلند و ولز)

- کدهای فعالیت SIC (۳ یا ۴ رقمی بسته به منطقه)

- رده اندازه اشتغال

تخصیص نمونه به شدت نامتناسب است. تمام کارگاه‌های بزرگ (کارگاه‌های دارای بیش از ۲۵۰ کارکن بر طبق تفکیک چارچوب آمارگیری به‌وسیله‌ی IDBR<sup>۲</sup>) با قطعیت در نمونه لحاظ می‌شوند و کسر نمونه‌گیری متناسب با اندازه (تعداد کارکن) کارگاه افزایش می‌یابد. اندازه نمونه کل در آمارگیری سال ۲۰۰۵، ۷۳۹۵۵ کارگاه با ۵۴۱۲۳ پاسخ‌گو بود. در نتیجه نرخ کل بی‌پاسخی ۲۶/۸ درصد بود. شیوه اصلی گردآوری داده‌ها از طریق پرسش‌نامه‌ی کاغذی بود.

این آمارگیری از سال ۲۰۰۸ با آمارگیری سالانه از کارگاه‌ها<sup>۳</sup> تغییر نام داد. روش نمونه‌گیری در این طرح یک نمونه‌گیری طبقه‌بندی تصادفی با استفاده از SIC2007 و تعداد کارکن به‌عنوان متغیرهای طبقه‌بندی است. چارچوب این طرح شامل ۲/۱ میلیون کارگاه است که پوشش تقریبی ۹۹ درصدی از فعالیت‌های اقتصادی بریتانیا را دارد. جامعه طرح ABS، ۱/۸ میلیون از کارگاه‌های فوق را پوشش می‌دهد. در سال ۲۰۱۰ تعداد نمونه‌های این طرح ۶۰۰۰۰ کارگاه بوده است. در این طرح جانمایی

---

<sup>1</sup> Standard Industrial Classification

<sup>2</sup> Inter-Depart Business Register

<sup>3</sup> Annual Business Survey

خودکار با استفاده از تغییرات دوره به دوره برای کارگاه‌های بی‌پاسخ در نمونه با بیش از ۱۰۰ کارکن انجام می‌شود. در موارد استثنایی که جانهی به‌نظر غیر قابل پذیرش است، جانهی به‌صورت دستی انجام می‌شود (Al-Hamad, 2008).

## ۱-۴ - فیلیپین

### درمان داده‌های گم‌شده در آمارگیری سالانه از کارگاه‌ها<sup>۱</sup>

مورد بی‌پاسخی، یا هر نوع داده گم‌شده به‌خاطر آن، یک رخداد عمومی در تمام آمارگیری‌هاست. پس مدیریت بی‌پاسخی ضروری است و در دو مرحله انجام می‌شود: پیش‌گیری و درمان. در مرحله پیش‌گیری، تمام راهبردها و کوشش‌ها برای رسیدن به نرخ پاسخ بالا و داده‌های با کیفیت، با عملیاتی مانند آموزش مناسب آمارگیران، اصرار بر پیگیری از طریق تماس‌ها، حضوری و تلفنی، ارسال نامه‌های اخطاری و تذکرآمیز، گردآوری داده‌ها از منابع ثانویه، مسابقه‌های تبلیغی و ادیت است. مرحله درمان وقتی که تمام تلاش‌ها برای تولید یک مجموعه داده کامل به پایان می‌رسد و هنوز بی‌پاسخی وجود دارد، آغاز می‌شود. زیرا آمارشناسان مشکل چگونگی برآورد با نمونه‌های دارای داده‌های گم‌شده را برعهده دارند.

بی‌پاسخی در آمارگیری‌های سالانه به سه رده بی‌پاسخی جزیی، بی‌پاسخی کلی و گزارش‌های غیرقابل استفاده یا ناسازگار (واحدهای خارج از حوزه یا پرسش بین طبقات) تقسیم می‌شود.

### فرایند درمان

درمان داده‌های گم‌شده در آمارگیری سالانه از کارگاه‌ها بر حسب طبقات متفاوت و کاربست‌های رایج که اخیراً اتخاذ شده به شرح زیر است:

### ۱- تعدیل وزن

این روش فقط برای بی‌پاسخی کلی در طبقه غیرحتمی<sup>۱</sup> قابل استفاده است. در این روش، وزن‌های  $W=N/n$  (وارون احتمال انتخاب) برای واحدهای پاسخ‌گو در آن طبقه در فرایند

---

<sup>1</sup> Survey of Establishment