

تبادل بین دسترسی کاربران به ریزداده‌ها و حفظ حریم خصوصی واحدهای آماری در آمار رسمی

عباس مرادی

پژوهشکده آمار

چکیده. یکی از وظایف مرکز آمار در دنیا انتشار ریزداده‌ها، داده‌ها و اطلاعات آماری به صورت عمومی و قابل دسترس برای تمام کاربران است. از آنجا که ریزداده‌ها حاوی اطلاعات خصوصی واحدهای آماری است، باید راه‌هایی برای حفظ حریم خصوصی این واحدها توسط مراکز آماری در نظر گرفته شود که این راه‌ها به محرمانگی معروف است. این محرمانگی باید به‌گونه‌ای باشد که ریزداده‌ها را بدون کم شدن از کلیت آن‌ها، با ناشناس‌سازی، در محیطی امن و جدا در اختیار کاربران و محققین مرتبط قرار گیرد. تبادل بین حریم خصوصی و انتشار ریزداده‌ها و اطلاعات آماری از طریق چند مولفه‌ی مهم ناشناس‌سازی و توافق‌نامه‌های اشتراک‌گذاری داده با دانشگاه‌ها، سازمان‌های دولتی و محققان و رده‌بندی کردن تمامی ریزداده‌های مرکز آمار ایران صورت می‌گیرد. زمان و مکان از این جهت بسیار حائز اهمیت است که همواره این متغیرها در حال تغییر هستند. بدین معنی که هرگونه گواهی‌نامه یا توافق‌نامه‌ای که به محقق حقیقی یا حقوقی برای یک مجموعه داده، اعطا شود، صرفاً محدود به زمان خاصی است و بنابراین تاریخ انقضا دارد. از این رو ممکن است ارزیابی مجدد حریم خصوصی داده‌هایی که در اختیار ایشان بوده است نیاز به بازنگری داشته باشد به گونه‌ای که حتی دیگر به ایشان اجازه‌ی دسترسی به آن ریزداده پس از زمان انقضای توافق‌نامه داده نشود. در این مقاله به معرفی روش‌های کنترل افشا و نرم‌افزارهای مرتبط با آن اشاره شده است. با استفاده از نرم‌افزار متن‌باز ARX یک مثال فرضی از داده‌های مربوط به ویروس کرونا با فرض حساسیت کشورها به میزان ابتلا و مرگ و میر ارائه شده است. همچنین میزان ریسک شناسایی کشورها پیش و پس از ناشناس‌سازی ارائه شده است. این بدان معنا است که روش‌های کنترل

دریافت: ۱۴۰۰/۸/۱۰، پذیرش: ۱۴۰۳/۹/۱۰.

افشا، نقطه‌ی تعادلی بین حریم خصوصی واحدهای آماری و دسترسی محققین به ریزداده‌ها را فراهم می‌نماید.

واژگان کلیدی: حریم خصوصی، محرمانگی، نقطه تعادل، واحدهای آماری.

۱- مقدمه

تاریخ گواه آن است که انسان‌ها تمایل به حفظ و نگهداری از دارایی‌های با ارزش خود داشته‌اند و بعضاً آن‌ها را دفن می‌کردند. یکی از مهمترین دارایی‌ها، اسرار و رمز و رازهای شخصی هر فردی است که تمایل به حفظ و نگهداری آن برای هر شخصی به درازای تاریخ بشر بر می‌گردد. به ویژه زمانی که زندگی اجتماعی بشر در قالب جوامع مدنی گسترش یافت، این مقوله اهمیت بیشتری پیدا کرد. با این وجود واژه‌ی حریم خصوصی به مفهوم قرن بیستمی خود، به تازگی در ادبیات حقوقی جهان به‌عنوان موضوعی مهم و یک حق بنیادین انسان‌ها به‌کار رفته و شناخته شده است به‌طوری‌که اصل ششم از اصول بنیادین آمار رسمی، به اصل محرمانگی داده‌ها اشاره دارد. طبق این اصل محرمانگی آمارها با هدف حفاظت از حریم خصوصی واحدهای آماری (اعم از حقیقی یا حقوقی) و جلوگیری از افشای داده‌ها و اطلاعات شخصی ایشان صورت می‌پذیرد. با توجه به نیاز روزافزون جوامع به داشتن آمارهای با کیفیت، سازمان‌های آماری هر روز بیش از پیش به افزایش اعتماد عمومی و سرمایه اجتماعی خود نیاز دارند. در همین راستا اطمینان‌بخشی به پاسخگویان در مورد احترام به حریم شخصی آنان و محرمانه ماندن اطلاعات آنان در نتیجه انجام فعالیت‌های آماری ضروری است. ازسوی دیگر تکلیف سازمان‌های آماری در ارائه‌ی سریع و قابل دسترس آمار مطابق استانداردهای جهانی و نیز ارائه‌ی فراداده‌ها و ریزداده‌ها، آنان را ملزم می‌کند تا بر اساس قواعد روشن و استانداردهای مناسب طوری عمل نمایند تا ضمن حفظ محرمانگی، اطلاعات آماری به بهترین شکل ممکن و با بیشترین وضوح در دسترس همگان قرار گیرند. با توجه به پیشرفت تکنولوژی و روش‌های جدید گردآوری داده‌ها مانند استفاده از مه‌داده‌ها و سایر منابع داده‌ای نیاز به بازنگری در روش‌ها و استانداردهای موجود برای حفظ محرمانگی آمار و روزآمد کردن آن‌ها بیشتر مطرح می‌شود. به ویژه زمانی که یک سازمان آماری با دسترسی به ثبت‌های اداری یا ریزداده‌های دیگر در پیوند، بخواهد با ترکیب آن‌ها اطلاعات آماری مورد نیاز خود را تولید

و سپس منتشر نماید استفاده از روش‌ها و ابزارهای متناسبی که بتواند حفظ محرمانگی اطلاعات را تضمین نماید از مهمترین چالش‌ها و نکاتی است که باید مورد توجه قرار گیرند. با این وجود باید توجه داشت که اصل محرمانگی نباید به مانعی برای مهمترین کارکرد تولیدکنندگان آمار رسمی یعنی تولید و ارایه آمارها منجر شود. امروزه موضوع استفاده از ریزداده‌های تولید شده توسط مراکز آماری، به ویژه توسط سایر تولیدکنندگان آمار رسمی با هدف تولید آمار توسط مراکز پژوهشی برای اهداف پژوهشی بسیار مهم است که وظیفه تولیدکنندگان آمار رسمی در اختیار گذاشتن این ریزداده‌ها با رعایت محرمانگی آن‌ها است. بنابراین وظیفه سازمان‌های تولیدکننده آمار رسمی پیدا کردن شیوه‌ها، استانداردها و قوانینی است که آمار و داده‌های تولید شده به راحتی در اختیار کاربران قرار گیرند و در عین حال محرمانگی آن‌ها نیز حفظ شود.

۲- مرور نوشتگان

موضوع محرمانگی اطلاعات آماری سابقه طولانی در آمار رسمی داشته و به موازات گسترش نظام‌های آمار رسمی توسعه یافته است. البته در اواخر دهه ۱۹۷۰ با انقلابی که در زمینه فناوری‌های رایانه‌ای و پیرو آن در زمینه علم اطلاعات و داده‌ها صورت پذیرفت حفظ و کنترل محرمانگی اطلاعات آماری در کانون توجه سازمان‌های کنشگر در نظام‌های آماری به ویژه مراکز ملی آمار قرار گرفت. با توجه به اوج گرفتن این مباحث مؤسسه بین‌المللی آمار، برای اولین بار اعلامیه اخلاق حرفه‌ای آمارگیری را در سال ۱۹۸۵ منتشر کرد که در آن به الزام رعایت حفظ محرمانگی اطلاعات آماری اشاره کرده بود. این مؤسسه در سال ۲۰۱۰ این اعلامیه را پس از بازنگری و تجدید نظر بازنشر کرد که در نسخه جدید نیز به موضوع محرمانگی با تفصیل و تأکید بیشتر و البته متناسب با دگرگونی‌های فناوری تولید و انتشار آمار پرداخته شده است [۵] و [۲]. با انتشار اصول بنیادین آمارهای رسمی در سال ۱۹۹۴ توسط سازمان ملل، نقطه عطفی در مطالعات و تحقیقات پژوهشگران آمار در ادبیات موضوع محرمانگی اطلاعات آماری ایجاد شد. مجلات آماری توجه بیشتری به انتشار مقاله‌های علمی در این خصوص نشان دادند. به عنوان مثال مجله معتبر آمار رسمی در سال ۱۹۹۸ یک شماره مخصوص را به انتشار مجموعه مقالات در این باره اختصاص داد [۷]. کمیسیون اروپا در سال ۲۰۰۰ در پی جلسات متعددی که

ویژه‌ی موضوع محرمانگی آمار تشکیل داده بود اداره آمار اتحادیه اروپا را موظف کرد تا قانونی را در این خصوص تهیه کند که در نهایت منجر به تهیه و تصویب قانون «حمایت از افراد در محافظت اطلاعات شخصی» در سال ۲۰۰۱ شد [۸] و [۹]. در سال‌های ۲۰۰۹ و ۲۰۱۳ نیز قوانین مکملی در همین باره برای حفظ محرمانگی آمار در استفاده از ریزداده‌ها و پژوهش‌های علمی تصویب شد [۱۲] و [۱۶]. همچنین در آخرین نسخه از دستورالعمل اجرایی اتحادیه اروپا در انجام فعالیت‌های آماری، محرمانگی اطلاعات آماری مورد توجه و تأکید قرار گرفته است [۱۴]. کمیسیون اروپایی آمار سازمان ملل نیز در ۵۱مین نشست سالانه خود در سال ۲۰۰۳ (کنفرانس آمارشناسان اروپایی) این موضوع را به طور تخصصی در دستور کار قرار داد که پیرو آن یک گروه کاری تشکیل شد [۱۵]. این گروه کاری در سال ۲۰۰۷ دستورالعملی را برای مدیریت دسترسی به ریزداده‌ها با حفظ محرمانگی ارائه کرد [۱۵]. در سال ۲۰۰۹ و ۲۰۱۱ نیز در کنفرانس ۱۵۹ام، بار دیگر این موضوع مورد بحث قرار گرفت و قواعد و دستورالعمل‌های موجود بازنگری شد [۱۳] و [۱۰]. این در حالی است که در همه استانداردهای مدرن کسب و کار آماری مانند $GSPBM^1$ ، $GSIM^2$ ، $CSPA^3$ ، محرمانگی اطلاعات آماری مورد تأکید قرار گرفته و یکی از مؤلفه‌های مهم تشکیل‌دهنده‌ی آن‌ها است. بخش آمار سازمان ملل نیز از سال ۲۰۱۱ موضوع بازنگری و بهنگام کردن اصول بنیادین آمارهای رسمی را در دستور کار نشست ۴۲ام قرار داد که در سال ۲۰۱۳ در نشست ۴۴ام نهایی شد و در سال ۲۰۱۴ پس از تصویب در مجمع عمومی به طور رسمی منتشر شد. علاوه بر همه آنچه گفته شد، موضوع حفظ محرمانگی اطلاعات آماری یکی از مؤلفه‌های مهم در چارچوب کیفیت آمارها است [۱۱]. به طور کلی مفهوم محرمانگی اطلاعات واحدهای آماری ابتدا توسط آمارشناسان قرن نوزدهم نظام آماری ایالت‌های متحده مطرح شد. روش‌های کنترل افشای اطلاعات واحد آماری، به‌طور رسمی با مقاله‌ی دالینیوس با عنوان «به‌سوی یک روش‌شناسی برای کنترل افشای آماری» در سال ۱۹۷۷ بیان شد. پس از امریکا موضوع محرمانگی اطلاعات واحدهای آماری و روش‌های کنترل افشای آن‌ها در کشورهای اروپایی و دیگر کشورها گسترش یافت. کشورهای امریکا، کانادا، هلند، انگلستان، ایتالیا، آلمان و اسپانیا از جمله کشورهای پیشرو در زمینه‌ی استفاده از روش‌های کنترل افشا هستند. در ایران نیز اقداماتی در این راستا انجام شده است، بردبار [۱]، رحمانی و

شیری [۳]، مهدوی پور و همکاران [۵]، نواب پور و همکاران [۷] و حق دوست و نواب پور [۲] روی این موضوع به تحقیق پرداخته‌اند.

۲-۱- حریم خصوصی، محرمانگی و افشا

منظور از حریم خصوصی یا حریم شخصی، حق یک فرد برای حفظ اطلاعات شخصی خود در برابر دیگران است و منظور از محرمانگی در آمار رسمی، محرمانه نگه داشتن اطلاعات شخصی واحدهای آماری (فردی و مؤسسه‌ای) خواهد بود [۱]. همواره اطلاعات واحدهای آماری باید برای همیشه محرمانه باقی بماند. بنا بر این دو نکته‌ی بسیار مهم از تفاوت محرمانگی و حریم خصوصی عبارت است از:

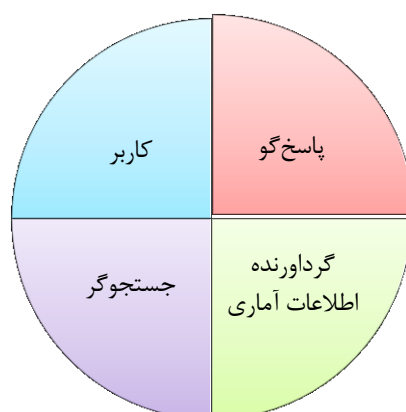
حریم خصوصی: حق مستقیم فرد (واحد آماری) برای حفظ اطلاعات شخصی خود.
محرمانگی: حق غیر مستقیم فرد (واحد آماری) است که به سازمان ملی آمار واگذار شده است.

منظور از افشا، کشف اطلاعاتی در خصوص هویت واحدهای آماری یا برخی ویژگی‌های محرمانه مربوط به واحدهای آماری است. افشا به طور کلی به دو دسته‌ی زیر تقسیم می‌شود:

افشای صفت: افشای صفت هنگامی رخ می‌دهد که اطلاعی محرمانه در مورد شخص به آسانی آشکار یا برآورد شود.

افشای هویت: افشای هویت در صورتی رخ می‌دهد که شخص ثالث بتواند سوژه یا مخاطب را از داده‌های منتشر شده شناسایی کند.

افشای هر کدام از این موارد برای سازمان‌های آماری و واحدهای آماری تبعاتی خواهد داشت. تبعات افشای هویت بسیار سنگین‌تر از تبعات افشای صفت برای واحدهای آماری خواهد بود. هرچند واحد آماری هیچگونه رضایتی از افشای صفت نیز نخواهد داشت. چهار عامل مهم در افشای اطلاعات آماری در شکل ۱ نمایش داده شده است.



شکل ۱ - ارکان افشای اطلاعات آماری

۳- انتشار انواع داده‌ها

تبادل ریزداده‌ها و اطلاعات آماری، در یک جامعه‌ی داده‌محور، پایه و اساس برنامه‌ریزی‌های منظم و دقیق است. از یک سو براساس اصل جریان یک‌طرفه ریزداده‌ها، رابطه تبادلی ریزداده‌ها بین سازمان ملی آمار و تولیدکنندگان ریزداده‌ها به صورت یک‌سویه به سمت سازمان ملی آمار است. این اصل تضمین‌کننده حفظ اعتماد عمومی به سازمان ملی آمار است. اما کوچکترین خلل در ریزداده‌ها و اطلاعات آماری، سنگ‌بنای برنامه‌ریزی‌ها را با مشکل مواجه خواهد کرد. از سوی دیگر حکمرانان و سیاست‌گذاران شایسته در هر جامعه‌ای داده محور با برنامه‌ریزی دقیق برگرفته‌شده از اطلاعات آماری درست، تصمیمات بهینه را خواهند گرفت و سیاست‌های کارآمدی را برای جامعه اتخاذ خواهند کرد. دولت جمهوری اسلامی ایران «نقشه راه توسعه دولت الکترونیک ایران» را در قالب دو مصوبه هماهنگ «آیین‌نامه توسعه خدمات الکترونیکی» مصوب شورای عالی اداری و «ضوابط فنی اجرایی توسعه دولت الکترونیکی» مصوب شورای عالی فناوری اطلاعات به منظور توسعه دولت الکترونیک تهیه و تصویب نموده است. در این نقشه راه، مرکز ملی IX و شبکه ملی اطلاعات به عنوان شبکه ارتباطی و انتقال اطلاعات اداری

کشور و نیز ۱۲ پایگاه اطلاعات و خدمات پایه دولت مشخص شده است. شبکه ملی آماری در بستر این شبکه تعریف شده است. در نظام ملی آماری ایران تنهاترین و مهم‌ترین هدف قرارگرفتن ریزداده‌های مربوط به تمام واحدهای آماری در شبکه ملی آمار ایران، تولید آمار رسمی از آن‌هاست. این ریزداده‌ها و هرگونه اطلاعات آماری استخراج شده از آن، جهت حفظ حریم خصوصی واحدهای آماری به شدت محرمانه تلقی می‌شوند. هرگاه خطر افشای واحد آماری در جدول‌های پیشابندی بالا باشد، باید با روش‌های علمی و با استفاده از قواعد کنترل خطر افشا، تعدیل شود. بنا بر این اصول و قواعد محرمانگی برای تولید آمار رسمی در یک نظام آماری مدرن بسیار حیاتی است.

جدول ۱- استراتژی کلاسیک برای استفاده از ریزداده‌ها و رفتار پاسخگویان

	شخص سوم با حسن نیت از ریزداده‌ها استفاده می‌کند.	شخص سوم با سوء نیت از ریزداده‌ها استفاده می‌کند.
تحلیل‌های خصوصی صورت می‌گیرد، تصمیمی ممکن است صورت نگیرد.	تحلیل‌های درست انجام می‌شود، تصمیم درست اتخاذ می‌شود.	واحد آماری پاسخگو است.
تحلیل نادرست صورت می‌گیرد و تصمیم نادرست اتخاذ خواهد شد.	تحلیل‌های نادرست صورت می‌گیرد، طبیعتاً تصمیم‌های غلط صورت خواهد گرفت.	واحد آماری پاسخگو نیست.

واحد آماری هرگاه بداند که با پاسخگویی و مشارکت خویش در آمارگیری‌ها می‌تواند در تصمیمات و سیاست‌گذاری کلان موثر باشد و به نوعی خودش نیز ذی‌نفع خواهد شد، به شرط اینکه حریم خصوصی وی رعایت گردد، استراتژی وی همواره پاسخگویی و همکاری هرچه بیشتر خواهد بود، این امر بدان معنی است که مرکز آمار ریزداده‌ها را محرمانه کرده و در اختیار محققین آماری مجاز برای اهداف آماری و برنامه‌ریزی جهت سیاست‌گذاری

قرار خواهد داد. این کار که اطلاعات واحدهای آماری با رعایت محرمانگی در اختیار شخص سومی قرار می‌گیرد، به نفع خود پاسخگو نیز خواهد بود. مرکز آمار با رعایت محرمانگی خیال پاسخگویان را راحت کرده و انگیزه‌ی همکاری آن‌ها را برای کارهای آتی نیز بیشتر خواهد کرد. اما اگر مرکز آمار نتواند با رعایت محرمانگی ریزداده‌ها را در اختیار محققین برای تصمیمات بهتر و سیاست‌گذاری بهینه قرار دهد یا اگر حریم خصوصی واحدهای آماری به هر طریقی توسط شخص سوم نقض شود، تبعات سنگینی برای مرکز آمار ایران و سیاست‌گذاری خواهد داشت زیرا مرکز آمار با بار پاسخگویی بیشتر رو به رو خواهد شد و حتی ممکن است که واحدهای آماری را دلسرد نماید و آن‌ها همکاری نداشته باشند. در این صورت کیفیت ریزداده‌ها پایین خواهد آمد و لذا سیاست‌گذاری و برنامه‌ریزی‌های کلان کشور با مشکل مواجه خواهد شد.

در نظریه بازی‌ها، به بازی‌ای که در آن برخی از بازیکنان در مورد منفعت^۴ بازیکنان اطلاعی ندارند بازی اطلاعات ناقص گفته می‌شود. بیشتر بازی‌ها، نقص اطلاعاتی دارند. فرض اطلاعات کامل در مورد بازی‌ها معمولاً یک فرض برای ساده‌سازی است که البته برای برخی کاربردها فرض قابل قبولی است. در این بازی واحدهای آماری اطلاعی در مورد شخص سوم ندارند که چگونه با ریزداده‌هایی که حاوی اطلاعات شخصی آن‌ها است برخورد خواهد کرد. در این صورت هرگاه مرکز آمار نتواند این اطمینان را به واحدهای آماری دهد بازی زیر رخ خواهد داد:

واحد آماری پاسخگو نخواهد بود یا اطلاعات غیر واقعی در اختیار مرکز آمار قرار می‌دهد، این امر هم به ضرر مرکز آمار، سیاست‌گذار و برنامه‌ریز، محقق به عنوان شخص سوم و خود پاسخگو خواهد بود. در این صورت تمامی بازیگران متضرر خواهند شد.

اما هرگاه مرکز آمار بتواند با رعایت محرمانگی و ایجاد روش‌های علمی و دقیق و رعایت اصول و استانداردهای محرمانگی اطلاعات ناشناس‌سازی شده افراد را منتشر کند که به هیچ وجه واحدهای آماری قابل شناسایی نباشد، در این صورت واحدهای آماری به عنوان بازیگرانی که به مرکز آمار اعتماد کرده‌اند بازی عقلایی خود را انجام خواهند داد یعنی اطلاعات کامل و دقیق خود را به طرح‌های آمارگیری مرکز آمار ارائه خواهند کرد. مرکز آمار نیز با استفاده از روش‌های دقیق و علمی خود با ناشناس‌سازی ریزداده‌ها می‌تواند آن اطلاعات را در اختیار کاربران و برنامه‌ریزان قرار دهد در این صورت شخص سوم حتی اگر سوء نیت داشته باشد نمی‌تواند به اطلاعات خصوصی واحدهای آماری دست پیدا کند.

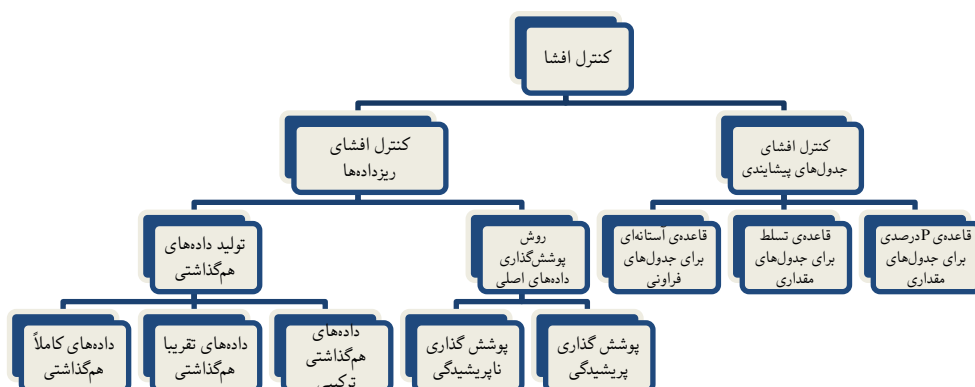
بنابراین بهترین خروجی از تحقیقات شخص سوم فقط نتایجی خواهد بود که وی با حسن نیت انجام داده است.

به عبارت دیگر و به زبان نظریه‌ی بازی‌ها، وظیفه‌ی مرکز آمار این است که با ایجاد علامت‌های مناسب برای طرفین بازی، بازیگران بازی با اطلاعات ناقص^۶ را راهنمایی کند تا به نقطه‌ی تعادل بهتر در یک بازی با اطلاعات کامل‌تر برسند. بدین منظور مرکز آمار با ایجاد علامت مناسب به پاسخگویان آن‌ها را قانع می‌کند که مشارکت خود را به حداکثر مقدار ممکن برسانند علامت اختصاص یافته به پاسخگویان حفظ حریم خصوصی آنان توسط مرکز آمار است. از سوی دیگر برنامه‌ریزان، سیاست‌گذاران و محققین برای اهداف آماری مرتبط به عنوان نقطه مقابل پاسخگویان، بازیگرانی هستند که نیازمند ریزداده‌های گردآوری شده توسط مرکز آمار هستند. این بازیگران خواهان ریزداده‌ها هستند. مرکز آمار برای این بازیگران علامت دیگری اختصاص می‌دهد به‌طوری‌که آن‌ها به اطلاعات کاملی دسترسی دارند و در عین حال هویت پاسخگویان افشا نمی‌شود، این علامت تحت عنوان دستورالعمل‌های محرمانگی خواهد بود. بنابراین بازیگران در چنین بازی علامت‌داری با آرامش خاطر مشارکت حداکثری خواهند داشت.

همانطور که پیشتر اشاره شد اصل ششم از اصول بنیادین آمار رسمی، به اصل محرمانگی داده‌ها اشاره دارد. مرکز آمار با علامت‌دهی به واحدهای آماری از یک سو و تمامی کاربران و محققین آمار از سوی دیگر، بازی را به‌گونه‌ای تغییر دهد که واحد آماری پس از دریافت علامت، با اطمینان خاطر به تمامی سوالات پرسشنامه جواب‌های صحیح و درست را بدهد و همکاری خود را بیشینه نماید زیرا با دریافت علامت متوجه خواهد شد که حریم خصوصی وی کاملاً رعایت می‌شود و از طرف دیگر ایشان با مشارکت و همکاری در آمارگیری‌ها می‌تواند به برنامه‌ریزی دقیق‌تر کمک نماید و خود نیز از این برنامه‌ریزی دقیق منتفع خواهد شد. از سوی دیگر تمامی کاربران و محققین آمار برای اهداف آماری مربوطه با دریافت علامت از مرکز آمار، مطمئن خواهند شد که شناساگرهای ریزداده‌های منتشرشده حذف شده است. آن‌ها نخواهند توانست به اطلاعات واحدهای آماری دسترسی پیدا کنند. ممکن است این علامت از سوی مرکز آمار فراتر از یک علامت ساده باشد و بسیار واضح و سختگیرانه به محققین آمار علامت‌دهی شود بدین معنی که مرکز آمار اجازه دسترسی به برخی از ریزداده‌ها را در یک بستر ایزوله و امن تحت ضمانت اجرایی سنگین، در بازه‌ی زمانی قابل انقضا و در یک مکان مشخص مانند خود مرکز آمار یا پژوهشکده

آمار در اختیار قرار خواهد داد. تمامی این علامت‌دهی‌ها در قالب اصل ششم یعنی اصل حفظ محرمانگی است.

روش‌های کنترل افشا به ماهیت داده‌ها و اطلاعات آماری بر می‌گردد و به طور کلی به دو دسته تقسیم بندی می‌شود. دسته اول به کنترل افشای جدول‌های پیش‌بینی و دسته دوم به کنترل افشای ریز داده‌ها می‌پردازد. جزئیات بیشتر در شکل ۲ نمایش داده شده است.



شکل ۲- روش‌های کنترل افشا

شایان گفتن است که روش‌های پوشش‌گذاری پریشیدگی و ناپریشیدگی هر کدام شامل روش‌های مختلف زیر می‌باشند.

روش‌های پوشش‌گذاری پریشیدگی:

- تعدیل جدول کنترل شده
- گرد کردن
- داده‌های هم‌گذاشتی
- نوفه‌ی جمعی
- مبادله‌ی داده‌ای
- مبادله‌ی رتبه‌ای
- تجميع ریز

- باز نمونه‌گیری

روش‌های پوشش گذاری ناپریشیدگی:

- بازکدگذاری سراسری^۷
- بازکدگذاری موضعی
- پنهان سازی موضعی^۸
- کدگذاری بالا و پایین
- نمونه‌گیری

برای شرح جزئیات هر کدام از این روش‌ها به منابع [۷ و ۹] مراجعه شود. این روش‌های علمی برای کنترل افشا است. در بخش‌های بعد نرم‌افزارهای مرتبط با محرمانگی معرفی می‌شوند.

۴- سلول‌های حساس و ارزیابی ریسک افشا

روش‌هایی برای شناسایی سلول‌های حساس وجود دارد. این روش‌ها به نوع جدول^۹ وابسته‌اند. در واقع قاعده‌های شناسایی برای تشخیص سلول‌های حساس جدول‌ها عبارت‌اند از:

- در جدول‌های فراوانی
 - قاعده‌ی آستانه‌ای^{۱۰}
- در جدول‌های مقداری
 - قاعده‌ی تسلط^{۱۱}
 - قاعده‌ی P درصدی^{۱۲}

۴-۱- قاعده‌ی آستانه‌ای

برای شناسایی سلول‌های حساس در جدول‌های فراوانی از قاعده‌ی آستانه‌ای استفاده می‌شود. این قاعده بیان می‌کند که اگر تعداد پاسخ‌گویان سلول از مقدار آستانه‌ای (حداقل ۳) کم‌تر باشد، آن سلول جدول، حساس به‌شمار می‌آید زیرا امکان شناسایی پاسخ‌گویان

..... مجله‌ی بررسی‌های آمار رسمی ایران، سال ۳۲، شماره‌ی ۱، بهار و تابستان ۱۴۰۰، صص ۱۲۵ - ۱۵۰.....

مربوط به این خانه و افشای اطلاعات محرمانه‌ی آن وجود دارد [۷]. به عنوان مثال جدول فرضی زیر را که تعداد مجوز فعالیت اقتصادی مهدکودک‌های محله امیرآباد تهران را نشان می‌دهد، در نظر بگیرید.

جدول ۲- تعداد (فرضی) مجوز فعالیت اقتصادی مهدکودک‌های محله امیرآباد تهران

مهدکودک	مجوز فعالیت		
	دارد	ندارد	کل
خصوصی	۴	۱	۵
تحت پوشش بهزیستی	۷	۳	۱۰

با توجه به جدول ۲ تنها یک مهدکودک خصوصی مجوز فعالیت ندارد، از این رو این سلول یک سلول حساس به شمار می‌آید. همچنین ۷ مهدکودک تحت پوشش بهزیستی مجوز فعالیت دارند و ۳ مهدکودک تحت پوشش بهزیستی مجوز فعالیت ندارند. بنا بر این جستجوگر با احتمال ۳۰٪ می‌تواند هر یک از مهدکودک‌های تحت پوشش بهزیستی که مجوز فعالیت ندارند را شناسایی کند.

واحدهای پرمخاطره^{۱۳}: به واحدهایی که مقدار آن‌ها از مقدار آستانه‌ای بیشتر باشد، واحدهای پرمخاطره می‌گویند.

۲-۴- قاعده‌ی تسلط برای جدول‌های مقداری

این قاعده دارای دو پارامتر n و k است. مقدار این پارامترها توسط سازمان‌های ملی آماری تعیین می‌شوند. پارامتر n عددی صحیح و نامنفی (مقدار آن به طور معمول برابر ۳) و پارامتر k حداقل برابر ۰ و حداکثر برابر ۱۰۰ است [۷]. اغلب مقدار k برابر ۷۵ در نظر گرفته می‌شود. این قاعده سبب می‌شود که $n - ۱$ پاسخ‌گو با جمع مقادیرهای مربوط به خود و مقایسه‌ی آن با مقدار کل سلول مربوطه نتوانند برآوردی برای مقدار مربوط به پاسخ‌گوی n م به دست آورند. هرچه n بزرگ‌تر باشد، احتمال کم‌تری وجود دارد که $n - ۱$ پاسخ‌گو با یکدیگر همکاری کرده و خواهان برآورد مقدار مربوط به پاسخ‌گوی n م باشند. لذا پارامتر n باید از تعداد پاسخ‌گویانی که متحد هستند، بزرگ‌تر باشد. بر

..... مجله‌ی بررسی‌های آمار رسمی ایران، سال ۳۲، شماره‌ی ۱، بهار و تابستان ۱۴۰۰، صص ۱۲۵ - ۱۵۰.....

اساس قاعده‌ی تسلط، یک سلول جدول در صورتی حساس به شمار می‌آید که جمع مقادیرهای مربوط به n پاسخ‌گو بیش‌تر از k درصد مقدار کل سلول مربوط به این پاسخ‌گویان باشد. فرض کنید در یک جدول، مقدار کل سلول‌های تحت بررسی با X ، تعداد پاسخ‌گویان آن سلول با N_X و مقدار پاسخ مربوط به پاسخ‌گوی i ام در سلول تحت بررسی با x_i نشان داده شود. همچنین بین مقدار مربوط به پاسخ‌گویان رابطه‌ی زیر برقرار باشد.

$$x_1 \geq x_2 \geq \dots \geq x_{N_X}$$

$$\sum_{i=1}^{N_X} x_i = X$$

یعنی بیش‌ترین مقدار متغیر تحت بررسی مربوط به پاسخ‌گوی اول است. برای این‌که جستجوگر نتواند مقدار x_i ها را به‌دست آورد. در حالت کلی می‌توان قاعده‌ی تسلط را به‌صورت زیر تعریف کرد:

$$T(x) > \frac{k}{100} X ,$$

$$T(x) = \sum_{i=1}^n x_i$$

که در آن $T(x)$ برابر جمع مقادیرهای مربوط به n پاسخ‌گو است، می‌توان نوشت:

$$x_1 + x_2 + \dots + x_n > \frac{k}{100} X$$

$$(۱) \quad X = x_1 + x_2 + \dots + x_{N_X} , \quad n < N_X$$

که در آن N_X تعداد کل افراد در سلول مورد نظر، X مقدار کل سلول تحت بررسی، و x_i مقدار مربوط به پاسخ‌گوی i ام در سلول تحت بررسی ($i = 1, \dots, N_X$) است.

اگر رابطه‌ی (۱) برقرار باشد، سلول مورد نظر حساس به‌شمار می‌آید. برای استفاده از قاعده‌ی تسلط باید مقداری از کل خانه را که متعلق به n پاسخ‌گو است، بدانیم. به‌عنوان مثال فرض کنید، $n = 3$ ، $k = 75$ و مقدار حقوق کارمندان یک شرکت بر حسب میلیون تومان برابر ۲،۷،۵ به ترتیب برای کارمندان به شماره ۱،۲،۳ و سایر کارمندان ۶ میلیون تومان و مقدار کل دریافتی ماه این شرکت ۲۰ میلیون تومان است. در این صورت با توجه به رابطه‌ی (۱) می‌توان نوشت:

$$T(x) > \frac{k}{100} X$$

$$T(x) = 7 + 5 + 2 = 14$$

جمع مقدارهای مربوط به سه پاسخ‌گوی اول برابر ۱۴ است. یعنی این مقدار ۷۰٪ از مقدار کل خانه را شامل می‌شود. از آنجایی که این مقدار بیش‌تر از ۷۵٪ مقدار کل خانه را شامل نمی‌شود، لذا این سلول جدول حساس به‌شمار نمی‌آید. به‌عبارت دیگر دو کارمند نمی‌توانند مقدار حقوق سومین کارمند دیگر را برآورد کنند.

بنا بر این در این حالت افشا رخ نمی‌دهد و خانه ایمن است. اما فرض کنید مقدار حقوق کارمندان آن شرکت بر حسب میلیون تومان برابر ۵، ۶، ۷ به ترتیب برای کارمندان به شماره ۱، ۲، ۳ و سایر کارمندان ۲ میلیون تومان و مقدار کل دریافتی ماه این شرکت ۲۰ میلیون تومان است.

در این صورت با توجه به رابطه‌ی (۱-۳) می‌توان نوشت:

$$T(x) > \frac{k}{100} X$$

$$T(x) = 5 + 6 + 7 = 18$$

جمع مقدارهای مربوط به سه پاسخ‌گوی اول برابر ۱۸ بوده و ۹۰٪ از مقدار کل خانه را شامل می‌شود. از آنجایی که این مقدار بیش‌تر از ۷۵٪ مقدار کل خانه را شامل می‌شود، لذا این سلول جدول حساس به‌شمار می‌آید و دو کارمند می‌توانند در صورت متحد شدن برآوردی نزدیک به حقوق سومین کارمند را به‌دست آورند [۷].

۴-۳- قاعدهی P درصدی برای جدولهای مقداری^{۱۴}

بر اساس قاعدهی P درصدی، سلول حساس است اگر پاسخ‌گویی بتواند سهم پاسخ‌گوی دیگر را با دقت P درصد برآورد کند. رایج‌ترین حالت هنگامی است که پاسخ‌گوی دوم بتواند مقدار مربوط به پاسخ‌گوی اول را با اختلاف کم‌تر یا برابر P درصد مقدار واقعی آن برآورد کند. رابطه‌ی مربوط به این قاعده به صورت زیر است:

$$(۲) \quad X - x_p < \left(1 + \frac{P}{100}\right)x_1$$

که

$$X = x_1 + x_p + \dots + x_{N_x}$$

رابطه‌ی (۲) را می‌توان به صورت رابطه‌ی زیر نیز نوشت:

$$(۳) \quad X - (x_1 + x_p) < \left(\frac{P}{100}\right)x_1$$

$$x_1 \geq x_p \geq \dots \geq x_{N_x} \quad \text{به طوری که:}$$

که در آن X مقدار کل سلول تحت بررسی، و x_i مقدار مربوط به پاسخ‌گوی i ام در سلول تحت بررسی ($i=1,2$) است.

با توجه به رابطه‌ی (۳) سلول تحت بررسی حساس است، اگر اختلاف جمع مقدارهای مربوط به پاسخ‌گویان اول و دوم از مقدار کل سلول مورد نظر کم‌تر از P درصد مقدار مربوط به پاسخ‌گوی اول باشد.

به عنوان مثال فرض کنید، مقدار کل سلول تحت بررسی برابر 120 و تعداد پاسخ‌گویان این سلول برابر 4 باشد. مقدار مربوط به این پاسخ‌گویان به ترتیب برابر 50 ، 35 ، 27 و 8 است. پاسخ‌گوی دوم می‌داند که مقدار مربوط به پاسخ‌گوی اول حداکثر برابر 85

(۳۵-۱۲۰) است. اگر مقدار P برابر ۳۰ در نظر گرفته شده باشد، در این صورت طبق رابطه‌ی (۲) می‌توان نوشت:

$$\left(1 + \frac{P}{100}\right)x_1 = \left(1 + \frac{30}{100}\right)50 = 65$$

از آنجایی که مقدار به‌دست آمده از ۸۵ کمتر است، بنا بر این سلول مورد نظر حساس به‌شمار نمی‌آید. حال اگر مقدار P برابر ۷۵ در نظر گرفته شود، در این صورت:

$$\left(1 + \frac{P}{100}\right)x_1 = \left(1 + \frac{75}{100}\right)50 = 87.5$$

که این مقدار از ۸۵ بیش‌تر است. بنا بر این با در نظر گرفتن مقدار ۷۵ برای P ، سلول تحت بررسی حساس به‌شمار می‌آید.

در صورتی که تمام پاسخ‌گویان سلول تحت بررسی به‌دنبال برآورد مقدار مربوط به پاسخ‌گوی اول باشند، رابطه‌ی (۳) به‌صورت زیر تغییر خواهد کرد:

$$(۴) \quad X - x_1 - \dots - x_{N_x} < \left(1 + \frac{P}{100}\right)x_1$$

که در آن:

X مقدار کل سلول تحت بررسی، و x_i مقدار مربوط به پاسخ‌گوی i ام در سلول تحت بررسی ($i = 1, \dots, N_x$) است.

به‌طور کلی سلول که از اجتماع دو سلول غیر حساس حاصل شود، غیر حساس به‌شمار می‌آید و نیازی به بررسی مجدد این خانه نیست. این ویژگی، ویژگی زیرجمعی نامیده می‌شود.

بر اساس قاعده‌ی P درصدی، سلول تحت بررسی حساس به‌شمار می‌آید، اگر رابطه‌ی زیر برقرار باشد:

$$(۵) \quad \hat{X} - x_p^s < (1 + \frac{P}{100})x_1^s$$

که در آن:

$\hat{X}$ برآورد مقدار کل سلول تحت بررسی، و x_i^s مقدار نمونه‌ای مربوط به پاسخ‌گوی i ام در سلول تحت بررسی ($i=1,2$) است. هنگامی که از داده‌های حاصل از آمارگیری نمونه‌ای استفاده می‌کنیم، وزن‌های نمونه‌گیری باید به‌طور محرمانه حفظ شوند، در غیر این صورت رابطه‌ی (۵) به‌صورت زیر تغییر می‌کند:

$$(۶) \quad \hat{X} - w_p x_p^s < (w_1 + \frac{P}{100})x_1^s$$

که در آن $\hat{X}$ برآورد مقدار کل سلول تحت بررسی، x_i^s مقدار نمونه‌ای مربوط به پاسخ‌گوی i ام در سلول مورد نظر ($i=1,2$)، و w_i وزن نمونه‌گیری ($i=1,2$) است. حال اگر داده‌ها حاصل آمارگیری‌های نمونه‌ای باشند، یک خانه بر اساس قاعده‌ی تسلط حساس به‌شمار می‌آید اگر رابطه‌ی زیر برقرار باشد:

$$(۷) \quad x_1^s + \dots + x_n^s > \frac{k}{100} \hat{X}$$

که در آن $\hat{X}$ برآورد مقدار کل سلول تحت بررسی، و x_i^s مقدار نمونه‌ای مربوط به پاسخ‌گوی i ام در سلول تحت بررسی ($i = 1, \dots, N_X$) است. قاعده‌ی تسلط برای $k=100$ و قاعده‌ی P درصدی برای $P=0$ تعریف نشده هستند. زیرا وقتی k به سمت 100 و P به سمت 0 میل می‌کند، هر دو قاعده شبیه قاعده‌ی آستانه‌ای که برای جدول‌های فراوانی به کار می‌رود، هستند. شناسایی سلول‌های حساس یک جدول از هر یک از روش‌های بالا امکان‌پذیر است.

۵- نرم‌افزارهای ایمن‌سازی داده‌ها

بسیاری از محاسبات مربوط به شناسایی داده‌های حساس و ناایمن و ایمن‌سازی آن‌ها توسط نرم‌افزارها یا برنامه‌های کامپیوتری معمول قابل انجام نیست. به همین منظور یک نرم‌افزار خاص محدودسازی افشای داده‌ها به نام ARGUS توسط اداره‌ی آمار هلند

توسعه داده شده است. همچنین در سال‌های اخیر بسته‌ی sdcMicro در نرم‌افزار R نیز تهیه و به‌طور رایگان در اختیار افراد و سازمان‌های آماری قرار گرفته است. این بسته شامل روش‌هایی برای محدودسازی افشای ریزداده‌ها است. در واقع می‌تواند جایگزین نرم‌افزار μ -ARGUS شود. از ویژگی‌های این بسته نسبت به نرم‌افزار μ -ARGUS می‌توان به افزوده شدن روش‌های جدیدتر، امکان دیدن کدهای هر برنامه، وجود رابط گرافیکی مناسب و امکان ورود داده‌ها با فرمت‌های متنوع‌تر و نیاز نداشتن به تعریف فراداده در این بسته اشاره کرد. در واقع در بسته sdcMicro بسته همراهی به نام sdcMicroGUI به‌عنوان رابط کاربری گرافیکی برای روش‌های مختلف قرار داده شده است. بسته‌ی IVEware در سال ۲۰۱۱ برای جابجایی چندگانه توسط محققان برنامه‌نویسی روش‌شناسی آمارگیری، مرکز تحقیقات آمارگیری، انستیتو تحقیقات اجتماعی و دانشگاه میشیگان ساخته شده است. ماژول SYNTHESIZE ماژولی است که جابجایی چندگانه رگرسیونی را برای کنترل افشای آماری پیاده‌سازی و اجرا می‌کند. یک نرم‌افزار منبع باز جامع دیگر به نام ARX برای ناشناس ماندن اطلاعات حساس واحدهای آماری وجود دارد که دارای مدل‌ها و روش‌های زیر است:

- مدل‌های ریسک و حریم خصوصی
- روش‌های تبدیل داده‌ها
- روش‌های تجزیه و تحلیل میزان سودمندی داده‌های خروجی
- از این نرم‌افزار در زمینه‌های مختلفی از جمله:
- سیستم عامل‌های تجزیه و تحلیل مه‌داده‌های تجاری
- پروژه‌های تحقیقاتی
- به اشتراک گذاری داده‌های آزمایش بالینی
- اهداف آموزشی

استفاده می‌شود.

نه‌تنها، نرم‌افزار ARX قادر به مدیریت و ناشناس‌سازی مه‌داده‌ها است بلکه به‌گونه‌ای طراحی شده است که از رابط کاربری گرافیکی قابل قبولی برخوردار است.

۶- مثال کاربردی از ایمن سازی با استفاده از نرم افزار ARX

در این نرم افزار در ابتدا، باید متغیرهایی که با خطرات افشای اطلاعات مواجه اند شناسایی شود و پس از شناسایی اقدامات لازم برای محافظت از آنها انجام شود. بنابراین در اینجا متغیرها به چهار رده‌ی مختلف زیر رده‌بندی می‌شوند:

متغیرهای شناسایی^{۱۵}: این متغیرها بسیار پرخطر هستند و منجر به شناسایی واحد آماری خواهند شد. بنابراین حذف یا تغییرات آن‌ها تاثیری بر کلیت مساله ندارد از این رو می‌توان آن‌ها را از مجموعه داده حذف یا با تغییرات آن‌ها را غیرقابل شناسایی کرد. متغیرهایی مانند نام و نام خانوادگی، کد ملی، شماره تأمین اجتماعی و ... نمونه‌هایی از این نوع متغیرها هستند.

متغیرهای شبه‌شناسایی^{۱۶}: جستجوگران از این متغیرها به صورت ترکیبی برای شناسایی مجدد استفاده می‌کنند. باید آن‌ها را با تبدیلات کاربردی تغییر داد که حتی ترکیب آن‌ها نتواند اطلاعات واحد را افشا کند. برخی از متغیرهای متداول شبه‌شناساگرها جنسیت، تاریخ تولد و کد پستی می‌باشند.

متغیرهای حساس^{۱۷}: این متغیرها حاوی ویژگی‌های بسیار مهمی هستند که واحد آماری تمایلی ندارد که این ویژگی‌ها به آن‌ها نسبت داده شود، در واقع هیچ‌گونه تمایلی برای در اختیار قرار گرفتن این اطلاعات به سایرین ندارد. از این رو، ممکن است مورد توجه مهاجمین قرار گیرد و در صورت افشا، می‌تواند به واحد آماری آسیب جدی برساند. باید روی این متغیرها فنون کنترل افشا به صورت دقیق و منظم صورت گیرد.

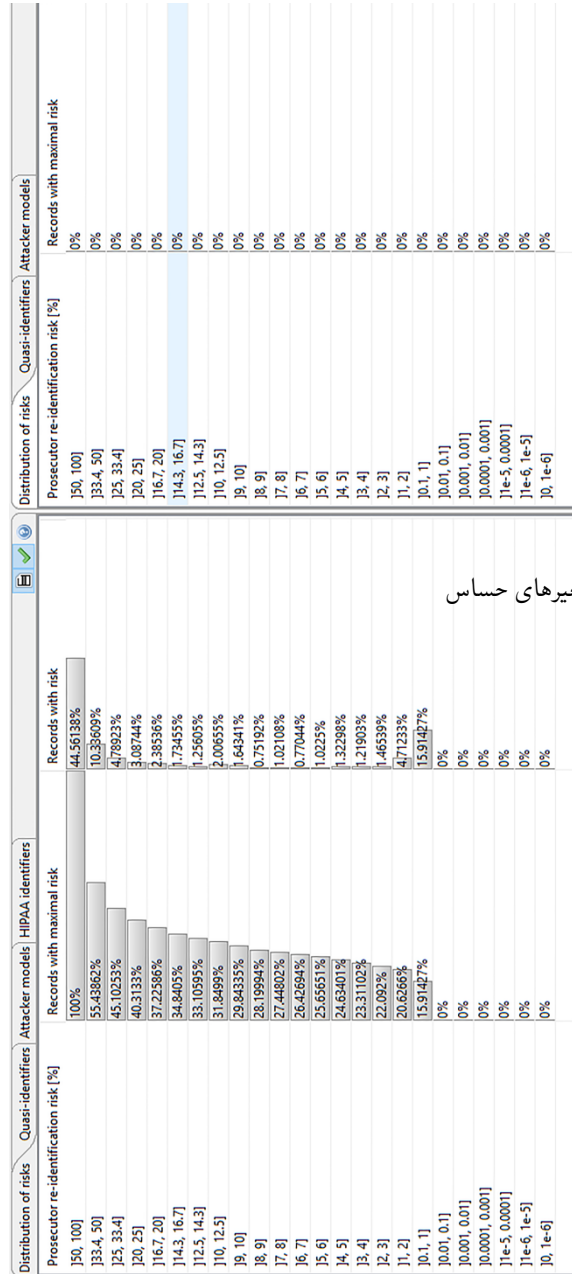
متغیرهای غیرحساس^{۱۸}: متغیرهایی که با خطرات افشا و حفظ حریم خصوصی ارتباط ندارند، متغیرهای غیرحساس نام دارند. آن‌ها بدون هیچ‌گونه تغییری در فایل داده‌ها نگهداری می‌شوند.

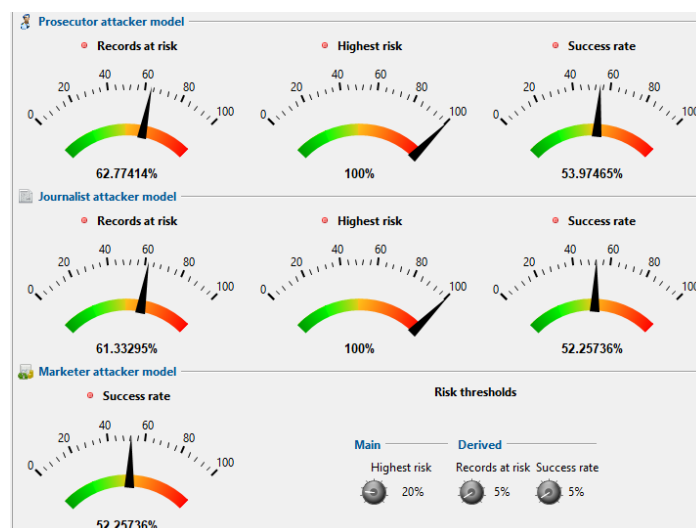
در شکل ۳ شمای کلی از برخی از متغیرهای مربوط به داده‌های ویروس کرونا برای کل کشورها شامل تعداد ابتلای جدید روزانه، نرخ مرگ و میر روزانه و تعداد بهبودیافته‌های روزانه و سایر اطلاعات مهم نشان داده شده است.

1	iso	co	location	date	tot	ne	new	total_dea	new_de	new_de	total_cases_p
2	AFG	Asia	Afghanistan	2/24/2020	1	1					0.026
3	AFG	Asia	Afghanistan	2/25/2020	1	0					0.026
4	AFG	Asia	Afghanistan	2/26/2020	1	0					0.026
5	AFG	Asia	Afghanistan	2/27/2020	1	0					0.026
6	AFG	Asia	Afghanistan	2/28/2020	1	0					0.026
7	AFG	Asia	Afghanistan	2/29/2020	1	0	0.143			0	0.026
8	AFG	Asia	Afghanistan	3/1/2020	1	0	0.143			0	0.026
9	AFG	Asia	Afghanistan	3/2/2020	1	0	0			0	0.026
10	AFG	Asia	Afghanistan	3/3/2020	2	1	0.143			0	0.051
11	AFG	Asia	Afghanistan	3/4/2020	4	2	0.429			0	0.103
12	AFG	Asia	Afghanistan	3/5/2020	4	0	0.429			0	0.103
13	AFG	Asia	Afghanistan	3/6/2020	4	0	0.429			0	0.103
14	AFG	Asia	Afghanistan	3/7/2020	4	0	0.429			0	0.103
15	AFG	Asia	Afghanistan	3/8/2020	5	1	0.571			0	0.128
16	AFG	Asia	Afghanistan	3/9/2020	7	2	0.857			0	0.18
17	AFG	Asia	Afghanistan	3/10/2020	8	1	0.857			0	0.206
18	AFG	Asia	Afghanistan	3/11/2020	11	3	1			0	0.283
19	AFG	Asia	Afghanistan	3/12/2020	12	1	1.143			0	0.308
20	AFG	Asia	Afghanistan	3/13/2020	13	1	1.286			0	0.334

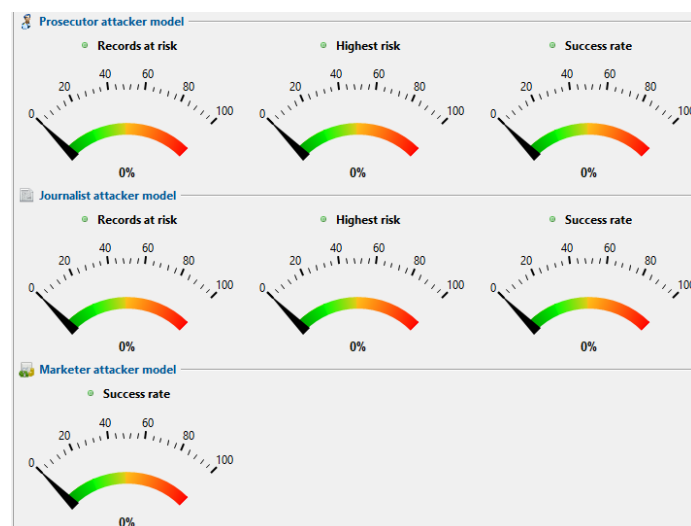
شکل ۳- داده‌های مربوط به ابتلا، مرگ و بهبود روزانه ویروس کرونا به تفکیک کشورها

در این بخش با یک مثال فرضی روی این داده‌های واقعی کار می‌شود: فرض شود برای کشورها این داده‌ها محرمانه باشد و نیاز دارند که اطلاعات مربوط به آن‌ها ناشناس‌سازی شود. عملیات ناشناس‌سازی با استفاده از روش‌های کنترل افشای روی ریزداده‌ها انجام می‌شود تا اطلاعات مربوط به کشورها به گونه‌ای انتشار یابد که داده‌ها نشانگر آن کشور نباشد، در شکل‌های ۴-۷ میزان ریسک آن‌ها پیش و پس از ناشناس‌سازی نشان داده می‌شود:

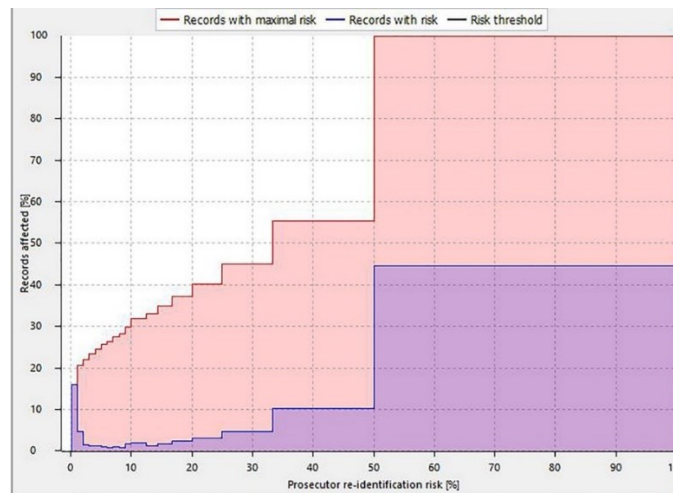




شکل ۵- میزان ریسک در مدل‌های متفاوت جستجوگران، پیش از ناشناس‌سازی



شکل ۶- میزان ریسک در مدل‌های متفاوت جستجوگران، پس از ناشناس‌سازی



شکل ۷- کاهش میزان ریسک شناسایی مجدد، پس از اعمال قواعد ناشناس سازی

۷- بحث و نتیجه گیری

بازیگران عقلایی هر بازی همواره علاقمند هستند که سود خود را بیشینه کنند. در یک بازی بین واحد آماری و کاربران، همواره نگرانی‌هایی وجود دارد. از یک سو واحد آماری نگران افشای اطلاعات خصوصی خویش است، به عبارتی تمایلی ندارد که حریم خصوصی وی نقض شود. از سوی دیگر کاربران و محققین آمار برای راهنمایی سیاست‌گذاران نیازمند ریزداده‌های دقیق و صحیح هستند. بنابراین واحدهای آماری و کاربران در یک بازی با اطلاعات ناقص بازی می‌کنند. سازمان آماری با استفاده از روش‌های نظری کنترل افشای اطلاعات، دانش و تجربیات خویش در این زمینه و با به‌کارگیری نرم‌افزارهای به‌روز در این زمینه، به عنوان یک رهبر در این بازی با اطلاعات ناقص ورود می‌کند. این روش‌ها را به کار می‌گیرد و ریزداده‌ها را بدون کم‌شدن از کلیت آن‌ها در اختیار کاربران تحت شرایطی که به حفظ محرمانگی می‌انجامد، قرار خواهد داد. در این صورت بازیگران علامت ارسالی را دریافت خواهند کرد و بازی اطلاعات ناقص آن‌ها به یک بازی بهتر با نقطه‌ی تعادل بالاتری تبدیل خواهد شد. در این صورت واحد آماری همکاری خویش را ارتقا خواهد داد زیرا با دریافت علامت مطمئن است که دیگر حریم خصوصی

وی نقض نخواهد شد، علاوه بر آن از برنامه‌ریزی دقیق و درست نیز منتفع خواهد شد. از سوی دیگر کاربران و محققین آماری می‌دانند ریزداده‌هایی که دریافت می‌کنند هم حاوی تمامی اطلاعات مربوط به نمونه یا جامعه است هم این که شرایط افشای اطلاعات آن با استفاده از دانش نظری و نرم‌افزارهایی به‌روز توسط مرکز آمار کنترل شده است و غیر قابل نفوذ خواهد بود و حتی در بدترین حالت ممکن برای داده‌هایی که با احتمال خیلی کم آسیب‌پذیر هستند، مرکز آمار تضامینی سنگین از آن‌ها دریافت می‌کند و برای مدتی معلوم در مکانی امن و جدا به آن‌ها اجازه‌ی دسترسی خواهد داد. در این صورت تمامی بازیگران کسب و کار آماری علامت‌های مهم را از رهبر بازی دریافت می‌کنند و بر اساس آن بازی درست را انجام خواهند داد و نقطه‌ی تعادل این بازی، همان همکاری کامل واحدهای آماری با مرکز آمار و فعالیت کاربران و محققین با حسن‌نیت خواهد بود. در مثال آخر نشان داده شد با فرض حساسیت کشورها به میزان ابتلا و مرگ و میر ویروس کرونا، می‌توان با روش‌های کنترل افشا، آن‌ها را به روش‌های علمی اقناع کرد. این مثال بیانگر آن است که با استفاده از روش‌های کنترل افشا می‌توان ریزداده‌های مربوط به ویروس کرونای کشور ایران در سطح استان و حتی در سطح شهر را انتشار داد. این مقاله بیان می‌دارد که ریزداده‌ها می‌توانند پس از کنترل افشا و رعایت روش‌ها و آئین‌نامه‌های محرمانگی در اختیار محققین قرار گیرند، این مهم ایجادکننده‌ی نقطه‌ی تعادل بین حریم خصوصی و انتشار ریزداده‌های آماری است.

توضیحات

1. Generic Statistical Business Process Model (GSBPM)
2. Generic Statistical Information Model (GSIM)
3. Common Statistical Production Architecture (CSPA)
4. Payoff
5. Signal

۶. در نظریه بازی‌ها، به بازی‌ای که در آن برخی از بازیکنان در مورد منفعت بازیکنان اطلاعاتی ندارند بازی اطلاعات ناقص گفته می‌شود.

7. Global Recording
8. Local suppression

۹. جداول دو نوع فراوانی یا مقداری هستند و تشخیص سلول حساس به نوع جدول وابسته است.

10. Threshold Rule

11. Dominance Rule

12. P% Rule

13. High-Risk

۱۴. میزان درصد p به حساسیت مراکز آماری بستگی دارد.

15. Identifying attributes

16. Quasi-identifying attributes

17. Sensitive attributes

18. Insensitive attributes

مرجع‌ها

- [۱] بردبارعشرت آبادی، محمد (۱۳۸۲). فنون محدودسازی افشای داده‌های آماری و کاربردهای آن‌ها. پایان‌نامه ارشد، علامه طباطبایی.
- [۲] حق‌دوست، سیما؛ نواب‌پور، حمیدرضا (۱۳۹۳). محدودسازی افشای اطلاعات واحد آماری در داده‌های خرد با کاربردی در داده‌های حاصل از کارگاه‌های صنعتی. مجله‌ی بررسی‌های آمار رسمی ایران، ۲۶(۱)، صص ۶۵-۸۸.
- [۳] رحمانی، علی؛ بشیری، نازنین (۱۳۹۳). مروری بر ادبیات افشای اطلاعات در ایران. پژوهش حسابداری، ۴(۳)، صص ۳۵-۶۷.
- [۴] کیانی، کاوه؛ شباک، اشکان؛ نورمحمدی، محمد (۱۳۹۸). طرح پژوهشی طراحی راهبردهای نوین‌سازی نظام ملی آماری ایران. پژوهشکده آمار ایران.
- [۵] مهدوی‌پور، علی؛ موسوی شیرینی، سیدمحمود؛ کریمی ریابی، علیرضا (۱۳۸۹). عوامل موثر بر افشای اطلاعات مالی از طریق اینترنت در وب سایت شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران. تحقیقات حسابداری و حسابرسی (تحقیقات حسابداری)، ۲(۵)، صص ۱۴۲-۱۶۱.
- [۶] نواب‌پور، حمیدرضا؛ محمدزاده، محسن؛ محمدزاده مومنی، سعید (۱۳۹۲). محدودسازی افشای اطلاعات واحدهای آماری. وزارت علوم، تحقیقات و فناوری - دانشگاه علامه طباطبایی - دانشکده اقتصاد.

- [۷] نواب‌پور، حمیدرضا؛ صفاکیش، محدثه؛ ایزدی، غلام‌رضا (۱۳۹۷). آمار رسمی. پژوهشکده آمار، تهران.
- [8] Bruengger, H. (2008). *How Should a Modern National System of Official Statistics Look*. UNECE, Statistical Division.
- [9] Domingo-Ferrer, J., Torra, V. (2001). Disclosure Control Methods and Information Loss for Microdata. Confidentiality, Disclosure, and Data Access, *Theory and Practical Applications for Statistical Agencies*, 91-110.
- [10] Eurostat. (2001). Regulation (EC) No 45/2001 of the European Parliament and of the Council of 18 December 2000. *Official Journal of the European Communities*, 14.
- [11] Eurostat. (2002). Commission regulation (ec) no 831/2002 concerning access to confidential data for scientific purposes.
- [12] Fellegi, I.P. (1972). On the Question of Statistical Confidentiality. *Journal of the American Statistical Association*. Mar 1, 67, 7-18.
- [13] Fienberg, E.S. (1998). Introduction to the Special Issue: Disclosure Limitation Methods for Protecting the Confidentiality of Statistical Data. *Journal of Official Statistics*, 14, 337-345.
- [14] ISI. (1985) Declaration on Professional Ethics. <http://www.isi-web.org/about-isi/professionalethics/43-about/about/151-ethics1985>.
- [15] ISI. (2010) Declaration on Professional Ethics. <http://www.isi-web.org/images/about/Declaration-EN2010.pdf>.
- [16] UNSD. (2013). Fundamental Principles of Official Statistics. (E/RES/2013/21).

عباس مرادی

دکترای ریاضی

تهران، خیابان فاطمی، خیابان بابا طاهر، خیابان سرتیپ فکوری، شماره ۱۴۵، پژوهشکده آمار
 رایانشانی: amoradi@src.ac.ir