



تهیه و تدوین استانداردهای حفظ و کنترل محرمانگی آمارها در نظام آماری ایران

مجری

اشکان شباک

همکاران

عباس مرادی

ودود کرامتی

گروه پژوهشی پردازش داده‌ها و اطلاع‌رسانی

تیر ۱۳۹۹

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ

کد شناسه: RP-9916

عنوان: تهیه و تدوین استانداردهای حفظ و کنترل محرمانگی آمارها در نظام آماری ایران

مجری: اشکان شباک

همکاران: عباس مرادی، ودود کرامتی

گروه پژوهشی: پردازش داده‌ها و اطلاع‌رسانی

تاریخ انتشار: تیر ۱۳۹۹

نوبت انتشار: اول

طراح جلد: علیرضا رنجبر

صفحه‌آرا: طاهره امینی

❖ حق مالکیت معنوی این طرح پژوهشی متعلق به پژوهشکده‌ی آمار است و نقل مطالب فقط با ذکر مأخذ مجاز است.



پژوهشکده‌ی آمار



مرکز آمار ایران

تهران، خیابان دکتر فاطمی، خیابان باباطاهر، خیابان سرتیپ فکوری،

شماره‌ی ۱۴۵

۰۲۱ ۸۸۶۳۰۴۴۰ - ۳

www.srtc.ac.ir

پیش‌گفتار

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 این وجود باید توجه داشت که این اصل 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 برای سایر تولیدکنندگان آمار رسمی برای تولید آمار و 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 و استانداردها و قوانینی است که آمار و داده‌ها تولید شده به راحتی در اختیار کاربران قرار گیرد و در عین حال محرمانگی آنها نیز حفظ شود.

در این طرح پژوهشی برآنیم تا ضمن شناخت منابع افشای داده‌ها و علل آن، تبیین مفهوم حریم شخصی در آمار رسمی و تعیین ابعاد محرمانگی آمار مطابق آن، مجموعه‌ای از روش‌ها و ابزارها را در قالب استانداردهای مشخص و دست‌ورالعمل‌های عملیاتی معین برای کاربرد در نظام آماری ایران و به ویژه مرکز آمار ایران تدوین و شناسایی کنیم. همچنین علاوه بر استانداردهای حفظ محرمانگی نیاز است تا استانداردهایی برای کنترل و تضمین حفظ محرمانگی داده‌ها وجود داشته باشد که با ارزیابی مداوم بتواند روش‌ها و ابزارهای حفظ محرمانگی را به طور پیوسته بهبود بخشد. همچنین مشخص شود در چه مواردی موضوع محرمانگی داده‌ها می‌تواند مورد تفسیر قرار گرفته و با چه شرایط و تضامینی، ریزداده‌ها در اختیار کاربران خاص قرار گیرند.

این پژوهش در گروه پژوهشی پردازش داده‌ها و اطلاع‌رسانی پژوهشکده‌ی آمار با همکاری آقای دکتر اشکان شباک به‌عنوان مجری طرح و آقایان دکتر عباس مرادی و ودود کرامتی به‌عنوان همکاران اصلی صورت گرفته است. که بدین‌وسیله از همه این عزیزان صمیمانه تشکر و قدردانی می‌شود.

گروه پژوهشی پردازش داده‌ها و اطلاع‌رسانی

پژوهشکده‌ی آمار

فهرست مطالب

۱	کلیات تحقیق.....
۱-۱-۱	مقدمه.....
۲-۱-۱	طرح مسئله و مرور نوشتگان.....
۳-۱-۱	اهمیت و ضرورت پژوهش.....
۴-۱-۱	اهداف تفصیلی پژوهش.....
۱-۴-۱-۱	اهداف تفصیلی طرح.....
۵-۱-۱	تعاریف و مفاهیم کلی مورد نیاز.....
۱۳	مروری بر ادبیات پژوهش.....
۱-۲-۱	مقدمه.....
۲-۲-۱	اصول بنیادین آمارهای رسمی.....
۳-۲-۱	محرمانگی و تولید آمار.....
۱-۳-۲-۱	ماده ۷ قانون مرکز آمار ایران مصوب بهمن ماه سال ۱۳۵۳.....
۲-۳-۲-۱	حریم خصوصی و محرمانگی.....
۴-۲-۱	افشا.....
۵-۲-۱	انتشار انواع داده‌ها.....
۲۳	روش‌های کنترل افشای داده‌های جدولی.....
۱-۳-۱	مقدمه.....
۲-۳-۱	داده‌های جدولی.....
۱-۲-۳-۱	جدول‌های K بعدی تکی.....
۲-۲-۳-۱	جدول‌های سلسله‌مراتبی.....
۳-۲-۳-۱	جدول‌های ربطی.....
۳-۳-۱	ضرورت بررسی داده‌های جدولی.....
۱-۳-۳-۱	قاعده‌ی آستانه‌ای برای جدول‌های فراوانی.....
۲-۳-۳-۱	قاعده‌ی تسلط برای جدول‌های مقداری.....
۳-۳-۳-۱	قاعده‌ی P درصدی برای جدول‌های مقداری.....
۴-۳-۱	سطح‌های حفاظتی.....
۱-۴-۳-۱	روش‌های محدودسازی افشای اطلاعات داده‌های جدولی.....
۳۹	روش‌های کنترل افشای ریزداده‌ها.....
۱-۴-۱	مقدمه.....
۲-۴-۱	مفاهیم ریزداده‌ها.....
۳-۴-۱	مراحل فرایند کنترل افشای ریزداده‌ها.....
۱-۳-۴-۱	حفاظت از محرمانگی.....

۴۴	۲-۳-۴- مشخصات و کاربردهای اصلی داده
۴۷	۳-۳-۴- خطر افشا و ارزیابی آن
۵۰	۴-۳-۴- روش‌های کنترل افشای اطلاعات
۵۶	۵-۳-۴- حالت تک‌متغیره
۵۸	۶-۳-۴- حالت چندمتغیره
۶۰	۷-۳-۴- پیاده‌سازی (اجرا)
۶۲	۴-۴- ارزیابی مخاطره‌ی افشا، نمادها و مفروضات
۶۵	۵-۴- ریسک افشا برای شبه‌شناساگرهای پیوسته
۶۶	۶-۴- برآورد ریسک شناسایی مجدد
۶۷	۱-۶-۴- ریسک فردی بر اساس نمونه: قانون آستانه
۶۸	۲-۶-۴- برآورد ریسک فردی با استفاده از وزن نمونه‌گیری
۷۰	۳-۶-۴- برآورد ریسک فردی با استفاده از مدل پواسون
۷۲	۴-۶-۴- سایر مدل‌ها که اطلاعات را از منابع دیگر قرض می‌گیرند
۷۲	۵-۶-۴- برآورد اکتشافی ریسک برای هر رکورد
۷۳	۶-۶-۴- ارزیابی ریسک از طریق پیوند رکورد
۷۵	معرفی نرم‌افزارهای ایمن‌سازی داده‌ها
۷۵	۱-۵- مقدمه
۷۵	۲-۵- برخی از نرم‌افزارهای مهم
۷۶	۱-۲-۵- نرم‌افزار آرگوس
۷۸	۲-۲-۵- بسته‌ی SdcMicro در R
۷۹	۳-۲-۵- بسته‌ی IVEware
۸۱	نتیجه‌گیری
۸۱	۱-۶- مقدمه
۸۱	۲-۶- نظام آمار ملی مدرن
۸۴	۱-۲-۶- رابطه بین استانداردها، دستورالعمل‌های بین‌المللی و قوانین آماری ملی در نظام آمار رسمی
۸۶	۲-۲-۶- اصل محرمانگی در نظام آمار ملی مدرن
	۳-۶- اصول و رهنمودهای مربوط به محرمانه بودن جنبه‌های یکپارچه‌سازی داده‌ها برای اهداف آماری یا
۸۹	تحقیقات آماری
۹۱	۴-۶- اصول و دستورالعمل‌ها
۹۴	۵-۶- اصول و دستورالعمل‌های محافظت فیزیکی و مجازی از داده‌ها
۹۵	۱-۵-۶- برخی از اصول محرمانگی و افشای اطلاعات
۹۶	۲-۵-۶- تسلط و افشای باقیمانده
۹۶	۳-۵-۶- محرمانگی و آمار خانوارها
۹۷	۴-۵-۶- رضایت افشای اطلاعات
۹۷	۵-۵-۶- افشای اجباری
۹۸	۶-۵-۶- محرمانگی منفعل
۹۸	۷-۵-۶- سازمان‌دهی تحقیق
۱۰۰	۶-۶- نمونه‌ای از یک فهرست کسب و کار آماری
۱۰۲	۷-۶- حکمرانی شایسته و محرمانگی
۱۰۳	۸-۶- بحث و نتیجه‌گیری
۱۰۹	مرجع‌ها

فهرست جدول‌ها

- جدول ۳-۱- تعداد درخواست‌های پذیرش دکتری دانشگاه‌های یک شهر بر اساس نوع آموزش و نوع دانشگاه... ۲۴
- جدول ۳-۲- تعداد دانشگاه‌های یک شهر بر اساس حضور در نوع آموزش و نوع دانشگاه... ۲۵
- جدول ۳-۳- تعداد کارمندان یک اداره برحسب متغیر رشته‌ی مدرک تحصیلی... ۲۵
- جدول ۳-۴- تعداد دانشجویان برحسب متغیر سلسله‌مراتبی محل تولد و جنسیت... ۲۶
- جدول ۳-۵- تعداد (فرضی) مجوز فعالیت اقتصادی مهد کودک‌های محله امیرآباد تهران... ۲۸
- جدول ۳-۶- سطح حفاظتی بالا برای قاعده‌های شناسایی... ۳۲
- جدول ۳-۷- تعداد دانشجویان یک شهر برحسب هزار نفر و بر اساس نوع آموزش... ۳۳
- جدول ۳-۸- تعداد دانشجویان یک شهر برحسب هزار نفر و بر اساس نوع آموزش- رویکرد بازگذاری... ۳۳
- جدول ۳-۹- تعداد دانشجویان یک شهر برحسب هزار نفر و بر اساس نوع آموزش (یک سلول حساس)... ۳۴
- جدول ۳-۱۰- تعداد دانشجویان یک شهر برحسب هزار نفر و بر اساس نوع آموزش- رویکرد پنهان‌سازی سلولی... ۳۴
- جدول ۳-۱۱- تعداد دانشجویان یک شهر برحسب هزار نفر و بر اساس نوع آموزش- رویکرد پنهان‌سازی سلولی
مکمل... ۳۴
- جدول ۳-۱۲- تعداد دانشجویان یک شهر برحسب هزار نفر و بر اساس نوع آموزش- محاسبه‌ی بازه‌ی برآورد... ۳۵
- جدول ۳-۱۳- تعداد کارکنان یک بنگاه اقتصادی بر اساس شغل و جنسیت... ۳۶
- جدول ۳-۱۴- تعداد کارکنان یک بنگاه اقتصادی بر اساس شغل و جنسیت- رویکرد گرد کردن... ۳۶
- جدول ۳-۱۵- تعداد کارکنان یک بنگاه اقتصادی بر اساس شغل و جنسیت- رویکرد گرد کردن تصادفی... ۳۶
- جدول ۳-۱۶- تعداد کارکنان یک بنگاه اقتصادی بر اساس شغل و جنسیت (پس از اجرای روش گرد کردن بر پایه‌ی
(۳) ... ۳۷
- جدول ۳-۱۷- تعداد کارکنان یک بنگاه اقتصادی بر اساس شغل و جنسیت (جدول تعدیل شده بر اساس سطح
حفاظتی پایین)... ۳۷
- جدول ۳-۱۸- تعداد کارکنان یک بنگاه اقتصادی بر اساس شغل و جنسیت (جدول تعدیل شده بر اساس سطح
حفاظتی بالا)... ۳۷
- جدول ۴-۱- مروری بر فرایند کنترل افشای آماری برای ریزداده‌ها... ۴۲
- جدول ۴-۲- ریزداده‌های مربوط به یک شهرستان (فرضی)... ۵۸
- جدول ۴-۳- ریزداده‌های مربوط به یک شهرستان مبادله شده (فرضی)... ۵۹
- جدول ۶-۱- انواع داده‌ها از نظر محرمانگی... ۱۰۶

فهرست شکل‌ها

شکل ۱-۲- فرایندها و زیرفرایندهای مدل عمومی کسب و کاری آماری (GSBPM).....	۱۴
شکل ۲-۲- تعادل بین کنترل افشای اطلاعات و انتشار ریزداده‌ها (تعادل بین کاربران و پاسخ‌گویان).....	۱۸
شکل ۳-۲- ارکان افشای اطلاعات آماری.....	۲۰
شکل ۱-۵- روش‌های نرم‌افزار آرگوس برای ایمن‌سازی داده‌ها.....	۷۶
شکل ۲-۵- مثالی از تجمیع ریز در بسته SDCMICRO در R برای کنترل افشا.....	۷۹
شکل ۱-۶- مدل مفهومی MMM.....	۸۴
شکل ۲-۶- مدل مفهومی گردآوری ریزداده و دسترسی به کاربران برای اهداف آماری و تحقیقی مرتبط با رعایت اصول محرمانگی.....	۱۰۶

۱

کلیات تحقیق

۱-۱- مقدمه

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 اطمینان بخشی 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 ارائه سریع و قابل دسترس آمارمطابق استانداردهای جهانی و نیز ارائه 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.

تمایل انسان به حفظ اسرار و رازهای خود را شاید بتوان به درازای پیدایش زندگی گروهی انسان‌ها دانست. به ویژه از زمانی که زندگی اجتماعی بشر در قالب جوامع مدنی گسترش می‌یافت این مقوله نیز اهمیت بیشتری پیدا کرد. با این وجود واژه حریم خصوصی^۱ 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. ماده ۱۲ اعلامیه جهانی حقوق بشر سازمان ملل به وضوح بر این حق اشاره کرده و آن را یک حق مسلم و غیر قابل انکار برای هر فرد دانسته است که باید به طور قانونی محافظت شود. البته شایان گفتن است که حریم خصوصی پیش از آنکه در غرب و کنوانسیون‌ها و اسناد بین‌المللی حقوق بشر پیش‌بینی شود، با تأکید بسیار در منابع اسلامی مورد توجه بوده و در آیات متعددی از قرآن مجید و نیز روایات اسلامی بر لزوم رعایت حریم خصوصی اشخاص تأکید شده است. با این وجود همیشه در خصوص تعریف استاندارد از حریم خصوصی و مرزهای آن، به ویژه از اواخر قرن گذشته که با پیشرفت تکنولوژی، مقوله اطلاعات و جریان آن رشد و گسترش انفجارگونه‌ای داشته است، هم‌داستانی وجود نداشته است. در مقوله آمار رسمی نیز از آنجایی که افراد باید آماری را در خصوص خود یا افراد یا اشیای دیگری ارائه دهند این چالش به طور جدی وجود داشته است. فلگی در سال ۱۹۷۲ تعریف زیر را از قول یکی از اساتید برجسته علم ارتباطات در آن زمان ارائه می‌دهد [۲]:

«حریم شخصی^۲ عبارت است از حقی که افراد دارند تا تصمیم بگیرند چه اطلاعاتی از خود را با دیگران به اشتراک بگذارند».

از طرفی واژه‌ی محرمانگی^۳ به اموری دلالت دارد که در رعایت کردن آن امور منجر به حفظ حریم شخصی خواهد شد. با توجه به تعاریف و تعابیر مختلفی که دو واژه حریم شخصی و محرمانگی در موضوعات و زمینه‌های مختلف دارند در اینجا به موضوع اطلاعات تمرکز می‌کنیم:

از نظر معنی هر دو عبارت به یک مفهوم اشاره می‌کنند البته حریم خصوصی به مفهوم حقوقی موضوع اشاره می‌کند و آن اینکه حق قانونی است که شخص در قبال داده‌هایی که از وی (به هر منظوری) گردآوری می‌شود دارد و محرمانگی نیز در این مفهومی به اقداماتی دلالت دارد که توسط ارگان‌ها و یا مؤسسات گردآورنده اطلاعات برای حفظ و تضمین حریم شخصی باید انجام شود^۴.

موضوع حریم شخصی افراد و حفظ محرمانگی اطلاعات آماری در آمار رسمی نیز بسیار مهم تلقی شده و مورد تأکید قرار گرفته است. تا آنجا که در اصل ششم از اصول اساسی آمار سازمان ملل تصریح شده است [۴]. در ایران نیز از همان آغاز پی‌ریزی نظام آماری کشور به این مهم و محرمانه بودن آمارها توجه شده است. ماده ۱۶ قانون تأسیس مرکز آمار و سپس ماده ۷ قانون مصوب مرکز آمار ایران و نیز برنامه‌های ملی آمار تدوین شده بر محرمانگی آمارها اشاره مؤکد شده است. اما همانگونه که گفته شد، مفهوم، تعبیر و گستره حریم شخصی و پیرو آن محرمانگی آمار و اطلاعات نیاز به تبیین بیشتری دارد. به ویژه در دو دهه گذشته که تهیه و انتشار آمارهای

¹ Privacy

² Privacy

³ Confidentiality

^۴ برای مشاهده‌ی جزئیات بیشتر به آدرس زیر از دانشگاه کالیفرنیا مراجعه شود.

رسمی با توجه به فناوری‌های مدرن و پیشرفته تغییر شگرفی کرده است این مقوله اهمیت دوچندانی یافته است. از سوی دیگر با تغییر رویکردها و شیوه‌های تولید آمار رسمی و مدیریت جریان داده‌ها درون نظام‌های آماری و مدرنیزه شدن این نظام‌ها، فرایند تهیه و انتشار آمارهای رسمی به‌عنوان یک مدل پیشرفته کسب و کار مطرح است که در آن وجود چارچوب مشخص کیفیت برای مدیریت کیفیت آمارهای تهیه و منتشر شده یک امر حیاتی و زیربنایی است. در این راستا اعتماد مردم به یک نظام آماری و در مرکز آن یک سازمان ملی آمار مرجع، در افزایش درستی و دقت آمارها نقش کلیدی و مستقیم دارد. یکی از مؤثرترین مؤلفه‌های شکل‌گیری و بیشتر شدن پیوسته‌ی این اعتماد عمومی، به‌عنوان سرمایه‌ی اجتماعی برای هر نظام آماری، باور مردم به حفظ اطلاعات شخصی آن‌ها و به عبارت دیگر حفظ محرمانگی آمارها در یک نظام آماری است. یکی از مؤلفه‌های مهم در کیفیت آمارها دسترس‌پذیری و وضوح آن‌ها برای همه کاربران آمار است. بنا بر این در دسترس قرار دادن آسان و هرچه بیشتر آمارها و نیز حفظ محرمانگی آن‌ها یک وظیفه و در عین حال چالش بزرگی برای سازمان‌های کنش‌گر در نظام‌های آمار رسمی به ویژه سازمان‌های ملی آمار به‌عنوان مرجع و هماهنگ‌کننده‌ی نظام‌های آمار رسمی کشورها است. برای انجام هر چه بهتر این مهم، علاوه بر روشن شدن و تبیین گستره و دامنه شمول حریم شخصی، باید روش‌های استاندارد و نیز ابزارهای مشخصی برای حفظ محرمانگی آمار تدوین و تهیه کرد. بدیهی است که این روش‌ها و ابزارها، از تدوین قوانین معین تا ارائه روش‌های انتشار و ذخیره‌سازی داده‌ها را در بر می‌گیرد. در این پژوهش برآنیم تا ضمن شناخت منابع افشای داده‌ها و علل آن، تبیین مفهوم حریم شخصی در آمار رسمی و تعیین ابعاد محرمانگی آمار مطابق آن، مجموعه‌ای از روش‌ها و ابزارها را در قالب استانداردهای مشخص و دستورالعمل‌های عملیاتی معین برای کاربرد در نظام آماری ایران و به ویژه مرکز آمار ایران تدوین و شناسایی کنیم. همچنین علاوه بر استانداردهای حفظ محرمانگی نیاز است تا استانداردهایی برای کنترل و تضمین حفظ محرمانگی داده‌ها وجود داشته باشد که با ارزیابی مداوم بتواند روش‌ها و ابزارهای حفظ محرمانگی را به طور پیوسته بهبود بخشد.

در فصل حاضر که فصل آغازین این پژوهش بوده به تشریح بیان مسئله، اهمیت و ضرورت موضوع، اهداف تحقیق، فرضیه تحقیق، روش تحقیق و مفاهیم به کار رفته در تحقیق پرداخته شده است.

۱-۲- طرح مسئله و 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 [۶]. با انتشار اصول بنیادین آمارهای رسمی در سال ۱۹۹۴ توسط سازمان ملل، نقطه عطفی در مطالعات و تحقیقات پژوهشگران آمار در