



مؤسسه ملی تحقیقات سلامت
جمهوری اسلامی ایران

پیمایش‌های ملی سلامت خانوار

اقدامات جهانی



زمستان ۱۴۰۱
دبیرخانه دیده‌بانی سلامت

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

درباره محتوی

اجرای پیمایش‌های مبتنی بر خانوار مستلزم رعایت دستورالعمل‌های متعددی است و تمام مراحل از برنامه‌ریزی تا اجرا و انتشار و سپس استفاده از داده‌ها باید مطابق دستورالعمل‌های مناسب و استاندارد پیش برده شود. با وجودی که اجرای این نوع پیمایش‌ها در دنیا قدمت طولانی دارد اما هنوز دستورالعمل‌های یکسانی که بتواند تمام مراحل پیمایش را دربرگیرد و قابل استفاده برای تمام مراکز مرتبط باشد، وجود ندارد. مجموعه‌ای با نام شبکه بین‌المللی پیمایش‌های خانوار تا حدودی این موارد را اجرایی و ارائه کرده است. این شبکه بین‌المللی، به عنوان سازوکاری جهت هماهنگی و تقویت همکاری بین سازمان‌ها و آژانس‌های بین‌المللی ایجاد شده است. در حال حاضر، عضویت در این شبکه محدود به سازمان‌ها و آژانس‌هایی است که برای اجرای برنامه‌های ملی و/یا بین‌المللی پیمایش در کشورهای در حال توسعه، پشتیبانی مالی و/یا فنی ارائه می‌دهند و سایر کشورها نمی‌توانند عضو شوند. رسالت این شبکه بهبود استفاده، دسترسی، و کیفیت داده‌های پیمایش در کشورهای در حال توسعه، و تشویق به تجزیه و تحلیل و استفاده از این داده‌ها توسط تصمیم‌گیران ملی و بین‌المللی، جامعه پژوهشی و سایر ذینفعان است. شبکه بین‌المللی پیمایش‌های خانوار به دنبال ارائه یا انتشار دستورالعمل‌ها، استانداردها و بهترین شیوه‌های مربوط به مراحل اجرای پیمایش است. این پروژه‌ها در پنج دسته سازماندهی می‌شوند، در متن گزارش حاضر به ترتیب ارائه شده است: برنامه‌ریزی و یکپارچگی پیمایش؛ روش‌ها و ارزیابی‌های پیمایش؛ گزینش و تجمیع داده‌ها؛ کاتالوگ کردن پیمایش؛ انتشار ریزداده‌ها. در ادامه، دستورالعمل‌ها و بهترین شیوه‌های مربوط به تمام مراحل اجرای پیمایش داده می‌شود. سپس، فهرستی از مجموعه مقاله‌های شبکه بین‌المللی پیمایش‌های خانوار برای تشویق تبادل نظر و بحث در مورد موضوعات مرتبط با طراحی و اجرای پیمایش‌های خانوار و تجزیه و تحلیل، انتشار و استفاده از داده‌های پیمایش ارائه شده است. در ادامه، در خصوص تیمی که فعالیت‌های شبکه بین‌المللی پیمایش‌های خانوار را در مورد داده‌ها و اطلاعات هماهنگ می‌کند، و نرم‌افزارهایی که برای تجزیه و تحلیل و خروجی‌های پیمایش در شبکه بین‌المللی پیمایش‌های خانوار ارائه شده است، پرداخته می‌شود.

مؤسسه ملی تحقیقات سلامت در ایران به‌عنوان یک مؤسسه پیشگام در برنامه‌ریزی و اجرای پیمایش‌های سطح خانوار در حوزه سلامت درصدد هست با بهره‌گیری از بهترین دستورالعمل‌های موجود در این زمینه اجرا این پیمایش‌ها را با بهترین استانداردها همگام سازد. این گزارش به همین منظور ارائه می‌گردد.

برای استفاده از مطالب این گزارش لطفاً از آدرس استنادی زیر استفاده شود:

گزارش پیمایش‌های ملی سلامت خانوار: اقدامات جهانی. مؤسسه ملی تحقیقات سلامت. جمهوری اسلامی ایران. زمستان ۱۴۰۱

فهرست محتوی

۱	معرفی شبکه بین‌المللی پیمایش‌های خانوار
۷	پروژه‌های شبکه بین‌المللی پیمایش‌های خانوار
۸	برنامه‌ریزی، هماهنگی و یکپارچگی پیمایش
۱۴	روش‌ها و ارزیابی‌های پیمایش
۳۶	تجزیه و تحلیل و استفاده داده
۴۹	گزینش و تجمیع داده‌ها
۶۲	کاتالوگ کردن پیمایش
۶۸	انتشار ریزداده‌ها
۷۲	دستورالعمل‌های اجرای پیمایش
۷۴	طراحی برنامه‌های پیمایش
۷۵	بودجه‌بندی پیمایش
۷۷	اجرای پیمایش‌ها
۱۰۵	ابزارهای یکپارچه‌سازی پیمایش
۱۰۸	بایگانی کردن و انتشار داده‌ها
۲۰۹	مقالات شبکه بین‌المللی پیمایش‌های خانوار
۲۱۳	پشتیبانی فنی برنامه تسریع داده (ADP)
۲۱۴	مقدمه‌ای بر Stata در تجزیه و تحلیل داده‌های پیمایش
۲۸۰	نرم‌افزار
۲۸۷	منابع



معرفی شبکه بین‌المللی پیمایش‌های خانوار

رسالت و اهداف

رسالت شبکه بین‌المللی پیمایش‌های خانوار^۱ بهبود استفاده، دسترسی، و کیفیت داده‌های پیمایش در کشورهای در حال توسعه، و تشویق به تجزیه و تحلیل و استفاده از این داده‌ها توسط تصمیم‌گیران ملی و بین‌المللی، جامعه پژوهشی و سایر ذینفعان است.

در این راستا، اهداف اصلی شبکه بین‌المللی پیمایش‌های خانوار عبارت هستند از:

- هماهنگی برنامه‌های پیمایش تحت حمایت بین‌المللی/ملی با تکیه بر زمان‌بندی، تعیین توالی، تکرار و مقرون به صرفه بودن آنها؛
- تدوین دستورالعمل‌های فنی و کاربردی برای تمامی مراحل چرخه عمر پیمایش؛
- تدوین کاتالوگ داده پیمایش برای اطلاع‌رسانی به کاربران در خصوص داده پیمایش و داده سرشماری؛
- تدوین استانداردها، ابزارها و دستورالعمل‌های مستندسازی، انتشار و حذف ریز داده‌ها؛
- بهبود همکاری بین تولیدکنندگان و کاربران داده.

فعالیت

فعالیت‌های شبکه بین‌المللی پیمایش‌های خانوار که توسط دبیرخانه این شبکه پیشنهاد شده و توسط گروه مدیریت تأیید شده است، شامل موارد زیر است:

- بهبود شیوه‌های جمع‌آوری داده از طریق ارزیابی و بهبود روش‌ها و برنامه‌های پیمایش. شبکه بین‌المللی پیمایش‌های خانوار، اقدامات کارشناسان و نیروهای تخصصی را که دستورالعمل‌ها و موارد مرجع را تولید می‌کنند، تسهیل می‌سازد. همچنین شبکه بین‌المللی پیمایش‌های خانوار، یک بانک سؤال، مخزن اصلی دستورالعمل‌های بین‌المللی پیمایش، تعریف مفاهیم و شاخص‌های مرتبط، دستورالعمل‌های پرسشگر، و طبقه‌بندی‌ها را ایجاد کرده و نگهداری می‌کند؛

1. International Household Survey Network (IHSN)



- ایجاد و نگهداری ابزار و دستورالعمل‌های جهت بهبود جمع‌آوری اسناد، نگهداری، حفظ حریم خصوصی اطلاعات، کاتالوگ کردن، انتشار، و بایگانی ریزداده‌های پیمایش و سرشماری؛
 - اطلاع‌رسانی به کاربران داده در خصوص داده‌های پیمایش‌ها و داده‌های سرشماری با تدوین کاتالوگ. شبکه بین‌المللی پیمایش‌های خانوار، ریزداده‌ها را توزیع نمی‌کند، اما با ارائه اسناد داده، کاتالوگ کردن، و انتشار ابزارها و توصیه‌هایی به سازمان‌هایی که دارای چنین داده‌هایی هستند، به دسترسی بیشتر این داده‌ها کمک می‌کند.
- شبکه بین‌المللی پیمایش‌های خانوار، پشتیبانی فنی یا مالی ارائه نمی‌دهد، اما فعالیت‌های خود را با برنامه تسریع داده^۱ هماهنگ می‌کند که این برنامه از اجرای بهترین شیوه‌های بین‌المللی طراحی پیمایش و بایگانی داده‌ها در کشورهای در حال توسعه پشتیبانی می‌کند.



اعضا و شرکا

شبکه بین‌المللی پیمایش‌های خانوار، به عنوان سازوکاری جهت هماهنگی و تقویت همکاری بین سازمان‌ها و آژانس‌های بین‌المللی ایجاد شد. در حال حاضر، عضویت در شبکه بین‌المللی پیمایش‌های خانوار محدود به سازمان‌ها و آژانس‌هایی است که برای اجرای برنامه‌های ملی و/یا بین‌المللی پیمایش در کشورهای در حال توسعه، پشتیبانی مالی و/یا فنی ارائه می‌دهند. گروه مدیریت شبکه بین‌المللی پیمایش‌های خانوار اعضای جدید را به عضویت می‌پذیرد.

عضویت به معنای هیچ‌گونه تعهد قانونی یا مشارکت اجباری نیست. با این حال، عضویت نشان‌دهنده تعهد اعضا برای رعایت اصول زیر است:

- اطلاعات مربوط به اقدامات برنامه‌های در حال انجام و برنامه‌ریزی شده پیمایش را با هدف هماهنگی زمان‌بندی و تعیین توالی پیمایش‌ها به اشتراک بگذارند؛
- به دنبال هماهنگی برنامه‌های پیمایش و اجرای فعالیت‌های مشترک باشند؛
- به دنبال استانداردهای پرسشنامه‌ها برای بهبود سازگاری بین‌المللی باشند؛
- فراداده‌های (شناسنامه داده) پیمایش را به اشتراک بگذارند و در نگهداری کاتالوگ مرکزی پیمایش مشارکت کنند؛
- از بهترین شیوه‌ها در جمع‌آوری، تجزیه و تحلیل و انتشار داده‌های پیمایش حمایت کنند؛
- با ارائه کمک‌های داوطلبانه غیرنقدی یا نقدی در هزینه فعالیت‌های شبکه بین‌المللی پیمایش‌های خانوار سهیم شوند.

اعضای شبکه بین‌المللی پیمایش‌های خانوار عبارتند از:

- بانک توسعه آفریقا^۱؛
- بانک توسعه آسیایی^۲؛
- پیمایش‌های جمعیت‌شناسی و سلامت^۳ توسط آژانس توسعه جهانی ایالات متحده^۴ و سایر حامیان مالی سرمایه‌گذاری می‌شود، و توسط شرکت خدمات حرفه‌ای بین‌المللی ICF اجرا می‌شود^۵؛

1. African Development Bank (AfDB)
2. Asian Development Bank (ADB)
3. Demographic and Health Surveys (DHS)
4. U.S. Agency for International Development (USAID)
5. Inner City Fund (ICF)



- دپارتمان توسعه بین‌المللی^۱، بریتانیا؛
- سازمان خواربار و کشاورزی ملل متحد (فائو)^۲؛
- بانک بین‌الملل توسعه ایالات متحده^۳؛
- سازمان بین‌المللی کار^۴؛
- اتحادیه بین‌المللی مخابرات^۵؛
- دیده‌بانی اقتصاد و آمار افریقای جنوبی^۶؛
- مشارکت در آمارهای توسعه قرن ۲۱ (سازمان همکاری و توسعه اقتصادی / PARIS21)^۷؛
- صندوق کودکان سازمان ملل متحد^۸؛
- برنامه پیشرفت و توسعه سازمان ملل متحد^۹؛
- کمیسیون اقتصادی و اجتماعی سازمان ملل متحد برای آسیا و اقیانوسیه^{۱۰}؛
- کمیسیون اقتصادی و اجتماعی سازمان ملل متحد برای آمریکای لاتین و کارائیب^{۱۱}؛
- کمیسیون اقتصادی و اجتماعی سازمان ملل متحد برای آسیای غربی^{۱۲}؛
- کمیسیون اقتصادی سازمان ملل متحد برای آفریقا^{۱۳}؛
- سازمان آموزشی، علمی و فرهنگی ملل متحد، موسسه آمار (یونسکو/UIS)^{۱۴}؛
- دفتر مقابله با مواد مخدر و جرم سازمان ملل متحد^{۱۵}؛

-
1. Department for International Development (DFID)
 2. Food and Agriculture Organization of the United Nations (FAO)
 3. Inter-American Development Bank (IDB)
 4. International Labour Organization (ILO)
 5. International Telecommunications Union (ITU)
 6. Observatoire Economique et Statistique d'Afrique Subsaharienne (AFRISTAT)
 7. Partnership in Statistics for Development in the 21st Century (PARIS21/OECD)
 8. United Nations Children's Fund (UNICEF)
 9. United Nations Development Programme (UNDP)
 10. United Nations Economic and Social Commission for Asia and the Pacific (ESCAP)
 11. United Nations Economic and Social Commission for Latin America and the Caribbean (ECLAC)
 12. United Nations Economic and Social Commission for Western Asia (ESCWA)
 13. United Nations Economic Commission for Africa (UNECA)
 14. United Nations Educational, Scientific and Cultural Organization, Institute for Statistics (UNESCO/UIS)
 15. United Nations Office on Drugs and Crime (UNODC)



- بخش آمار سازمان ملل متحد؛^۱
- اداره توسعه بین‌المللی آمریکا؛^۲
- برنامه نظارت مشترک سازمان جهانی بهداشت/ صندوق کودکان سازمان ملل متحد برای تأمین آب و سلامت همگانی؛^۳
- بانک جهانی، گروه توسعه داده‌ها؛^۴
- برنامه جهانی غذا؛^۵
- سازمان جهانی بهداشت.^۶

شبکه بین‌المللی پیمایش‌های خانوار همچنین از همکاری آژانس‌های غیر عضو منتفع می‌شود.

شبکه بین‌المللی پیمایش‌های خانوار با آژانس‌ها، انجمن‌ها یا سایر گروه‌هایی که در زمینه‌های خاص مربوط به جمع‌آوری، تجزیه و تحلیل داده‌ها یا انتشار داده‌های پیمایش پیش‌رو هستند، مشارکت می‌کند، از جمله:

- اتحاد ابتکار استناد داده؛^۷
- DataFirst، دانشگاه کیپ‌تاون؛^۸
- انجمن تحقیقات اقتصادی، قاهره؛^۹
- مجموعه ریزداده‌های استفاده عموم هماهنگ، دانشگاه مینه‌سوتا؛^{۱۰}
- شبکه بین‌المللی ارزیابی امکانات سلامتی؛^{۱۱}
- کنسرسیوم بین دانشگاهی برای تحقیقات سیاسی و اجتماعی؛^{۱۲}
- بنیاد داده باز.^{۱۳}

-
1. United Nations Statistics Division (UNSD)
 2. United States Agency for International Development (USAID)
 3. WHO/UNICEF Joint Monitoring Programme (JMP) for Water Supply and Sanitation
 4. World Bank, Development Data Group (WB-DECDG)
 5. World Food Programme (WFP)
 6. World Health Organization (WHO)
 7. Data Documentation Initiative (DDI) Alliance
 8. DataFirst, University of Cape Town
 9. Economic Research Forum (ERF), Cairo
 10. Integrated Public Use Microdata Series (IPUMS-International), University of Minnesota
 11. International Health Facility Assessment Network (IHFAN)
 12. Inter-university Consortium for Political and Social Research (ICPSR)
 13. Open Data Foundation (ODaF)



حامیان مالی

شبکه بین‌المللی پیمایش‌های خانوار از کمک‌های غیرنقدی آژانس‌های عضو خود و از حمایت مالی موارد زیر بهره‌مند می‌شود:

- دپارتمان توسعه بین‌المللی بریتانیا که این دپارتمان در سال ۲۰۰۸ میلادی از پرسنل اعزام شده دبیرخانه شبکه بین‌المللی پیمایش‌های خانوار حمایت کرد و صندوق امانت^۱ (TF071804) که توسط بانک جهانی اجرا می‌شود را حمایت می‌کند. این صندوق امانت که برای سایر حامیان مالی باز است، در ژانویه ۲۰۱۲ برای حمایت از فعالیت‌های شبکه بین‌المللی پیمایش‌های خانوار و برنامه تسریع داده، تأسیس شد؛
- تسهیلات کمک هزینه توسعه^۲ بانک جهانی که در سال ۲۰۰۶ تأسیس شد (DGF 06-401007 تا 06-401012). این کمک مالی بانک جهانی توسط دبیرخانه PARIS21 اداره می‌شود؛
- صندوق توسعه سیاست و منابع انسانی ژاپن^۳ که توسط بانک جهانی اداره می‌شود.

1. Trust fund

2. Development Grant facility

3. Policy and Human Resources Development (PHRD)



پروژه‌های شبکه بین‌المللی پیمایش‌های خانوار

این بخش اطلاعاتی در مورد پروژه‌های قبلی و فعلی اجرا شده توسط شبکه بین‌المللی پیمایش‌های خانوار را ارائه می‌دهد. شبکه بین‌المللی پیمایش‌های خانوار به دنبال ارائه یا انتشار دستورالعمل‌ها، استانداردها و بهترین شیوه‌های مربوط به تمام مراحل اجرای پیمایش است. این پروژه‌ها در ۵ دسته سازماندهی می‌شوند:

- برنامه‌ریزی و یکپارچگی پیمایش^۱؛
- روش‌ها و ارزیابی‌های پیمایش^۲؛
- گزینش و تجمیع داده‌ها^۳؛
- کاتالوگ کردن پیمایش^۴؛
- انتشار ریزداده‌ها^۵.

-
1. Survey planning and integration
 2. Survey methods and assessments
 3. Data curation
 4. Survey cataloguing
 5. Microdata dissemination



برنامه‌ریزی، هماهنگی و یکپارچگی پیمایش

کمبود منابع، فقدان استانداردهای ملی مشخص برای جمع‌آوری داده‌ها، وابستگی زیاد به بودجه خارجی ناپایدار، یا اولویت‌های متناقض، بسیاری از کشورها را از توسعه و اجرای یک برنامه پیمایش منسجم باز می‌دارد، اما بسیار مهم است که برنامه‌ریزی پیمایش خوب، در چارچوب یک استراتژی آماری ملی بنا می‌شود، پروژه‌های این بخش مستندسازی و توصیه‌هایی برای برنامه‌ریزی، هماهنگی و یکپارچگی برنامه‌های پیمایش‌ها و سرشماری ارائه می‌کند.

پروژه‌های برنامه‌ریزی پیمایش شبکه بین‌المللی پیمایش‌های خانوار شامل تهیه دستورالعمل‌ها یا گزارش‌هایی در مورد آنها می‌باشد:

- هماهنگی برنامه‌های پیمایش، به عبارت دیگر، تدوین یک کُد بین‌المللی اقدام برای پیمایش‌های خانوارها؛
- طراحی برنامه‌های پیمایش، به عبارت دیگر، ارائه یک برنامه پیمایش و سرشماری ملی با زمان‌بندی و توالی بهینه پیمایش‌ها؛
- بودجه‌بندی پیمایش‌های خانوارها؛
- یکپارچگی تولید آماری، به عبارت دیگر، پذیرش استانداردهای ملی برای طراحی پیمایش و سایر ابزارهای جمع‌آوری داده (شامل ماژول‌های استاندارد پرسشنامه، مفاهیم و طبقه‌بندی‌های استاندارد و غیره). یکی از اهداف کلیدی یکپارچگی آماری افزایش ثبات (قابلیت مقایسه) و قابلیت اطمینان داده‌های به‌دست آمده در منابع مختلف و در طول زمان است.



گد اقدام برای پیمایش خانوارها



عدم هماهنگی بین تهیه‌کنندگان و حامیان پیمایش منجر به ناکارآمدی (تکرار فعالیت‌ها، عدم پشتیبانی در برخی از کشورها) و سایر مسائل (مقایسه ضعیف داده‌ها، مشکلات مربوط به کیفیت، مشاوره فنی متناقض و غیره) شده است.

بهبود هماهنگی پیمایش یکی از اهداف شبکه بین‌المللی پیمایش‌های خانوار است. هماهنگی باید توسط مجموعه‌ای از قوانین و سازوکارهای اجرایی مرتبط، مورد توافق قرار گیرد. رعایت یک قانون مشترک توسط همه ذینفعان در اجرای برنامه‌های پیمایش ملی و بین‌المللی، سیستم‌های آماری منسجم‌تر و مؤثرتری را می‌پروراند که نیازهای دولت‌ها و جامعه را برآورده می‌کند.

یک گد اقدام، ضمانت کیفیت و در دسترس بودن داده‌ها نیست، اما رعایت آن به کارآمدتر و مقرون به‌صرفه‌تر شدن سیستم کمک می‌کند. مشکلات تأمین مالی برنامه‌های پیمایش، ارائه مشاوره‌های فنی متناقض، ظرفیت فنی پایین، یکپارچگی کم و غیره، مشکلاتی هستند که در چارچوبی که مجموعه‌ای اصلی از اصول و شیوه مورد توافق ارائه می‌دهد، بهتر حل می‌شوند.

وضعیت پروژه: باز

حامی (ان): صندوق دپارتمان توسعه بین‌المللی شماره TF۰۱۱۷۲۲ که توسط بانک جهانی اداره می‌شود.

اجرا شده توسط: بانک جهانی و آژانس‌های شریک شبکه بین‌المللی پیمایش‌های خانوار (با هماهنگی دبیرخانه شبکه بین‌المللی پیمایش‌های خانوار)

نوع خروجی: گد اقدام (سند آنلاین) که توسط آژانس‌های عضو شبکه بین‌المللی پیمایش‌های خانوار و سایر، با دستورالعمل‌های مربوطه تأیید شده است.

فعالیت‌ها: دبیرخانه شبکه بین‌المللی پیمایش‌های خانوار فهرستی آزمایشی از اصول اصلی و شیوه‌های مربوط به پیمایش را براساس مقایسه چندین گد موجود تهیه کرده است. این سند پیش‌نویس به اشتراک گذاشته شده است. دبیرخانه شبکه بین‌المللی پیمایش‌های خانوار از نظرات و پیشنهادات استقبال می‌کند.



A Code of Practice for Household Surveys (DRAFT)

IHSN Secretariat and World Bank Data Group - 1 October 2014

The document proposes a list of 15 core principles and associated practices that would form a Code of Practice for Household Surveys. The objective of such a Code would be to establish a better environment for coordination and coordination of survey sponsors and producers in developing countries. This document is a DRAFT for discussion.

[Download](#)



طراحی برنامه‌های پیمایش^۱



بسیاری از سازمان‌های آماری درگیر در برنامه تسریع داده در مورد بهبود برنامه‌ریزی در برنامه‌های پیمایش خود راهنمایی خواسته‌اند. هدف از پروژه طراحی برنامه‌های پیمایش، ارائه راهنمایی عمومی براساس دو مطالعه موردی است: هیأت هماهنگی ملی آمار فیلیپین^۲ و آژانس مرکزی آمار اتیوپی^۳ نحوه هماهنگی ذینفعان اصلی، تصمیم‌گیری، تطابق و مدیریت بازخورد کاربران را برای غلبه بر چالش‌های سازمانی، مالی و روش‌شناختی درگیر در طراحی و اجرای برنامه پیمایش ملی را مستند کرده‌اند. این اطلاعات به ایجاد مشاوره عمومی کمک می‌کند.

وضعیت پروژه: باز



حامی(ان): تسهیلات کمک هزینه توسعه بانک جهانی، کمک هزینه شماره ۴۰۰۱۰۱۲۰۶، تحت مدیریت دبیرخانه PARIS21 در سازمان همکاری و توسعه اقتصادی^۴



اجرا شده توسط: هیأت هماهنگی ملی آمار فیلیپین، آژانس مرکزی آمار اتیوپی، و مشاور شبکه بین‌المللی پیمایش‌های خانوار.

نوع خروجی: مقاله کاربردی^۵ با ماتریس مفهومی، دستورالعمل‌های عملی کوتاه، و دو مورد مطالعه کشور. دستورالعمل‌های مربوط به طراحی برنامه‌های پیمایش و منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

 [An Accelerated Program for Implementing the Marrakech Action Plan for Statistics in Selected Countries](#)  (1.24 MB)

 [Costing MDG surveys in Sub-Saharan Africa \(2006-2015\)](#)  (124 KB)

1. Designing survey programs
2. Philippines National Statistical Coordination Board (NSCB)
3. Ethiopian Central Statistical Agency (CSA)
4. Organisation for Economic Co-operation and Development (OECD)
5. Working paper



بودجه بندی پیمایش های خانوار^۱



زمان ایجاد بودجه پیمایش باید دقت زیادی به خرج داد. تجربه نشان می دهد که تمایل به نادیده گرفتن یا دست کم گرفتن هزینه برخی از فعالیت ها مانند آموزش کارکنان میدانی یا تجزیه و تحلیل، بایگانی، و انتشار ریزداده ها وجود دارد. بودجه پیمایش به شرایط محلی هر کشور و ویژگی هر پیمایش بستگی دارد. قبل از تهیه بودجه، باید اندازه نمونه، دوره جمع آوری داده ها و مدت مصاحبه را دانست. با این حال، راه هایی برای ساختار بودجه و زمان بندی وجود دارد که در همه عملیات پیمایش خانوار مشترک است. هدف این پروژه ارائه دستورالعمل های کلی برای ساختار بندی چنین بودجه هایی است.

وضعیت پروژه: باز

حامی (ان): بانک جهانی، تسهیلات کمک هزینه توسعه (کمک هزینه تحت مدیریت PARIS21)

اجرا شده توسط: مشاور فردی

نوع خروجی: خروجی شامل یک برنامه MS Excel با دستورالعمل ها و توصیه های فنی است.



یکپارچگی تولید داده^۱



این پروژه، یکپارچگی تولید داده‌ها، تجربیات آمار کانادا را در یکپارچگی پیمایش‌های مختلف و فعالیت‌های جمع‌آوری داده‌ها مستند کرده است. هدف از این پروژه تولید موارد حمایتی بود که پیچیدگی شکستن رویکرد "تولید سیلو"^۲ را در یک سازمان آماری نشان می‌داد. مسائلی که اداره آمار کانادا با آن روبه‌رو است تفاوت اساسی با مواردی که سازمان‌های آماری در کشورهای کم درآمد یا متوسط با آن روبه‌رو هستند، ندارد و شامل وابستگی به بودجه خارجی و موانع قدیمی، فرهنگی و فنی است.

وضعیت پروژه: بسته

حامی(ان): تسهیلات کمک هزینه توسعه بانک جهانی، کمک هزینه شماره ۰۶-۰۹-۱۰۰۹۴۰۰، تحت مدیریت دبیرخانه PARIS21 در سازمان همکاری و توسعه اقتصادی

اجرا شده توسط: Gordon Priest، مشاور شبکه بین‌المللی پیمایش‌های خانوار

نوع خروجی: مقاله کاربردی



The Struggle for Integration and Harmonization of Social Statistics in a Statistical Agency - A Case Study of Statistics Canada

Gordon Priest - June 2010

IHSN Working Paper No 004. This paper describes how external influences and a corporate strategy lead to partial coordination of surveys and data collection practices at Statistics Canada.

[Download](#)

دستورالعمل‌های مربوط به ادغام داده گذاشته شده است.

1. Integration of data production
2. "Silo production" approach



روش‌ها و ارزیابی‌های پیمایش

شبکه بین‌المللی پیمایش‌های خانوار نه تأمین مالی می‌کند و نه پشتیبانی فنی برای اجرای پیمایش انجام می‌دهد. نقش آن کمک به بهبود کیفیت، ارتباط، هزینه اثربخشی و دقت پیمایش‌ها با ارزیابی و ارائه روش‌های پیمایش است. این هدف با اجرای پروژه‌های مختلف حاصل می‌شود، شامل:

- ارائه بانک سؤال به عنوان مخزن اصلی دستورالعمل‌های بین‌المللی برای طراحی پیمایش؛
- انجام ارزیابی‌هایی از نحوه برخورد مناسب مسائل خاص در پیمایش‌های موجود. رویکرد شبکه بین‌المللی پیمایش‌های خانوار شامل: (۱) شناسایی کاربران اصلی داده‌ها؛ (۲) توصیف موارد استفاده آنها از داده‌ها و نیازهای خاص آنها از نظر ارتباط و قابلیت اطمینان؛ (۳) ترجمه این معیارهای کیفیت به چارچوب‌های ارزیابی خاص-بخش^۱؛ (۴) ارزیابی تعداد زیادی از پیمایش‌های موجود با پوشش جغرافیایی وسیع؛ و (۵) گزارش ارزیابی و تدوین توصیه‌ها. این موضوعات براساس ارتباط خاص آنها انتخاب می‌شوند. پروژه‌های قبلی و فعلی به سنجش‌های زیر می‌پردازند:

○ تحصیلات؛

○ مصارف خوراکی خانوار (مصارف خوراکی برای اهداف مختلفی از جمله تجزیه و تحلیل

فقر، طراحی برنامه‌های غنی‌سازی مواد غذایی، ارزیابی امنیت غذایی، تدوین حساب‌های ملی و شاخص‌های قیمت مصرف‌کننده و غیره)، جنسیت، تحصیلات و موارد دیگر)؛

○ مصارف غیرخوراکی خانوار؛

○ مسائل جنسیت.

- ارزیابی استفاده و ارتباط داده‌های پیمایش خانوار برای طراحی یا ارزیابی برنامه‌های غنی‌سازی مواد غذایی؛

• ارزیابی استفاده و ارتباط داده‌های پیمایش خانوار برای تدوین حساب‌های ملی؛

- ایجاد یک چارچوب ارزیابی کیفیت پیمایش در قالب یک وقایع‌نگار^۲ (لوگ بوک) برای مدیران پیمایش. این یک ابزار برای مدیران پیمایش است که در طول اجرای پیمایش و برای ارزیابی پس از پیمایش‌ها از آن استفاده می‌کنند. این چارچوب ارزیابی خاص-بخش را که قبلاً ذکر شد، تکمیل می‌کند.

1. Sector-specific assessment frameworks

2. Log book



ساخت بانک سؤال^۱



نیاز به هماهنگی دستورالعمل‌های پیمایش در جلسات متعدد سطح بالا در دهه گذشته تأکید شده است. برای تقویت هماهنگی و بهبود روش‌های جمع‌آوری داده‌ها، شبکه بین‌المللی پیمایش‌های خانوار از گروه عملیاتی و متخصصانی که بر روی ارزیابی کار می‌کنند و به بهبود روش‌های موجود کمک می‌کنند، پشتیبانی می‌کند. شبکه بین‌المللی پیمایش‌های خانوار همچنین دارای یک بانک سؤال است که مخزن اصلی شاخص‌ها، طبقه‌بندی‌ها، مفاهیم و سؤالات و مطالب مرجع است. این شبکه، دستورالعمل‌های بین‌المللی در مورد طراحی پیمایش و محاسبات شاخص از

منابع مختلف بین‌المللی را ارائه می‌دهد؛ شامل هیچ‌گونه داده یا پرسشنامه پیمایش واقعی نیست (برای پرسشنامه‌ها، به کاتالوگ پیمایش شبکه بین‌المللی پیمایش‌های خانوار رجوع کنید). برنامه نرم‌افزاری بانک سؤال توسط شبکه بین‌المللی پیمایش‌های خانوار ایجاد شده است؛ حفظ محتوای آن یک تلاش مشترک بین آژانس‌ها و شرکای عضو شبکه بین‌المللی پیمایش‌های خانوار است. محتوای بانک سؤال از دو منبع اصلی مشتق شده است:

- دستورالعمل‌های پیمایش موجود، تهیه شده توسط آژانس‌ها یا گروه‌های متخصص مختلف؛
- دستورالعمل‌های تهیه شده توسط یا برای شبکه بین‌المللی پیمایش‌های خانوار.

وضعیت پروژه: باز

حامی(ان): تسهیلات کمک هزینه توسعه بانک جهانی، کمک هزینه شماره ۰۶-۰۹-۲۰۰۹-۴۰۰۱، تحت مدیریت دبیرخانه PARIS21 در سازمان همکاری و توسعه اقتصادی و صندوق امانت دپارتمان توسعه بین‌المللی، تحت مدیریت بانک جهانی، گروه توسعه داده‌ها

اجرا شده توسط: دبیرخانه بانک جهانی، گروه توسعه داده‌ها و شبکه بین‌المللی پیمایش‌های خانوار؛ با مشاوران شبکه بین‌المللی پیمایش‌های خانوار برای حمایت از توسعه طرح XML و ویرایشگر بانک سؤال و رابط کاربری

نوع خروجی: نرم‌افزار منبع باز (شامل طرح XML، ویرایشگر تخصصی، و برنامه رابط کاربری)، و مخزن آنلاین دستورالعمل‌ها برای طراحی پرسشنامه‌های پیمایش (بانک سؤال شبکه بین‌المللی پیمایش‌های خانوار).

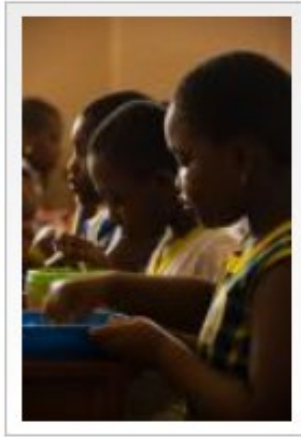


خروجی‌ها

- طرح LMX که از استانداردهای فراداده IDD و XMDS وام گرفته است برای بانک سؤال ایجاد شد؛
- یک برنامه نرم‌افزاری بانک سؤال در PHP با پایگاه داده LQSYM ایجاد شد. هدف، انتشار این برنامه به عنوان نرم‌افزار منبع باز است؛
- خود بانک سؤال شبکه بین‌المللی پیمایش‌های خانوار که مخزن اصلی و قابل جستجو از دستورالعمل‌های مربوط به طراحی پرسشنامه است، پرسشنامه‌های مُدل، سؤالات توصیه شده، تعاریف مفاهیم و شاخص‌ها، طبقه‌بندی‌ها و بسیاری مطالب دیگر را که از منابع متعدد به‌دست آمده است، ارائه می‌دهد.



سنجش آموزش



پیمایش‌های خانوار منبع مهمی از اطلاعات پیرامون سیستم‌های آموزشی هستند. برنامه‌های پیمایش بین‌المللی مانند پیمایش‌های جمعیت‌شناسی و سلامت و پیمایش‌های خوشه‌ای شاخص‌های چندگانه به منابع استاندارد پایگاه‌های اطلاعاتی بین‌المللی در زمینه آمار آموزش تبدیل شده‌اند. انواع دیگر پیمایش‌ها به ندرت مورد استفاده قرار گرفته و داده‌های آموزشی آنها گاهی به خوبی درک و استاندارد نمی‌شوند. این پروژه ارزیابی ۳۰ پیمایش خانوار را انجام داده و کیفیت فرآیند جمع‌آوری اطلاعات را تجزیه و تحلیل کرده است. توصیه‌های کلی را ارائه می‌دهد که باید برای افراد و آژانس‌های درگیر در طراحی پرسشنامه‌های آموزش برای پیمایش خانوارها و استفاده از داده‌های حاصله مفید باشند.

وضعیت پروژه: بسته

حامی(ان): تسهیلات کمک هزینه توسعه بانک جهانی، کمک هزینه شماره ۰۶-۰۹-۱۰۰۹-۴۰۰۱، تحت مدیریت دبیرخانه PARIS21 در سازمان همکاری و توسعه اقتصادی

اجرا شده توسط: آکادمی توسعه آموزشی، سیاست آموزش و مرکز داده

نوع خروجی: مقاله کاربردی

نمونه خروجی مقاله کاربردی به اشتراک گذاشته شده است:



How (well) is Education Measured in Household Surveys? A Comparative Analysis of the Education Modules in 30 Household Surveys from 1996-2005

Education Policy and Data Center - October 2009

This report studies 30 household surveys and analyzes how, and how well, their education information is collected. It provides a number of general recommendations useful to the individuals and agencies involved in the design of education questionnaires for household surveys and use of the resultant data.

[Download](#)



سنجش مصارف خوراکی^۱

اکثر کشورها با استفاده از پیمایش‌هایی مانند بودجه خانوار، هزینه درآمد و استانداردهای زندگی، داده‌های مربوط به مصرف یا هزینه‌های خانوار را جمع‌آوری می‌کنند. این پیمایش‌ها ممکن است برای ارائه اطلاعات حساب‌های ملی یا شاخص قیمت مصرف‌کننده^۲، اندازه‌گیری فقر یا دلایل دیگر انجام شود. اما داده‌ها اغلب برای اهدافی غیر از آنهایی که در ابتدا برای آنها در نظر گرفته شده بود، استفاده می‌شوند. به عنوان مثال، داده‌های مصارف خوراکی ممکن است برای ارزیابی امنیت غذا یا آسیب‌پذیری و تنوع رژیم غذایی، ارزیابی فرصت‌های تجاری یا به عنوان ورودی برای طراحی یا نظارت بر برنامه‌های غنی‌سازی مواد غذایی مورد استفاده قرار گیرند. جوامع مختلف کاربران داده نیازهای متفاوت و خاصی دارند؛ معیارهای ارتباط و قابلیت اطمینان داده‌های مصارف خوراکی بستگی به استفاده از داده‌ها دارد. مرتبط بودن و قابل اعتماد بودن پیمایش‌ها برای جوامع مختلف کاربرانی که داده‌ها را "مجدداً هدف گذاری"^۳ کرده‌اند، می‌تواند به‌طور بالقوه ارزش قابل توجهی به داده‌ها اضافه کند. پروژه سنجش مصارف خوراکی، استفاده‌ها و استفاده‌کنندگان از داده‌های مصارف خوراکی و نیازهای خاص آنها را برای ارزیابی ارتباط و قابلیت اطمینان مجموعه وسیعی از پیمایش‌ها (براساس مرور پرسشنامه‌های پیمایش) و ارائه توصیه‌هایی برای بهبود این پیمایش‌ها توصیف می‌کند.

وضعیت پروژه: بسته

حامی(ان): صندوق امانت دپارتمان توسعه بین‌المللی شماره TF011722 تحت مدیریت بانک جهانی، گروه توسعه داده‌ها، تسهیلات کمک هزینه توسعه بانک جهانی، کمک هزینه شماره ۰۶-۱۰۰۹-۴۰۰ تحت مدیریت دبیرخانه PARIS21 در سازمان همکاری و توسعه اقتصادی

اجرا شده توسط: Olivier Dupriez (گروه توسعه داده‌های بانک جهانی)، Lisa Smith (مشاور شبکه بین‌المللی پیمایش‌های خانوار)، Nathalie Troubat (سازمان غذا و دارو)

نوع خروجی: (۱) چارچوب ارزیابی برای سنجش داده‌های مصرف خوراک؛ (۲) گزارشی از ارزیابی جهانی پیمایش‌ها و راهنماهای فنی ناشی از آن؛ و (۳) پایگاه داده متا در مورد روش‌های پیمایش. پیش‌نویس کامل این خروجی‌ها در اوایل سال ۲۰۱۳ در دسترس بوده است.

1. Measuring food consumption
2. Consumer price index
3. Re-purpose



رویکرد: یک فرم ارزیابی برای جمع‌آوری اطلاعات در مورد طراحی ماژول‌های مصارف خوراکی یا هزینه پیمایش تهیه شد. این فرم، که در اینجا با دستورالعمل‌های مرتبط ارائه شده است، برای ارزیابی پرسشنامه‌های پیمایش از ۱۰۰ کشور استفاده شد. فراداده جمع‌آوری شده با استفاده از فرم، به Excel و Stata فرستاده شده و پایگاه داده‌های مورد استفاده برای تجزیه و تحلیل قابلیت اطمینان و ارتباط داده براساس معیارهای تعریف شده توسط کاربران زیر را ارائه می‌دهد:

- تحلیلگران فقر؛

- حسابداران ملی و گردآورندگان شاخص‌های قیمت مصرف‌کننده؛

- کارشناسان امنیت غذایی و گردآورنده ترازنامه‌های غذایی؛

- متخصصان تغذیه، به ویژه متخصصان غنی‌سازی مواد غذایی؛

- سایر موارد کاربرد و کاربران (بخش خصوصی، جامعه تحقیقاتی).

شرح نیازهای حسابداران ملی با اجرای پروژه مرتبط "پیمایش‌های خانوار و حساب‌های ملی" به‌دست آمده است.

پروژه مشابهی برای ارزیابی سنجش "مصرف یا مخارج غیر خوراک خانوار" در حال اجرا است.

خروجی میانی^۱: فرم "پیمایش‌های مصرف یا هزینه خوراک"، برای ارزیابی مورد استفاده قرار گرفته است. در مرحله تجزیه و تحلیل پروژه، نقاط ضعف در فرم مشخص شده است؛ نسخه تجدید نظر شده و دستورالعمل‌های استفاده و پیاده‌سازی آن در سال ۲۰۱۳ ارائه شده است. علاوه بر این، فایل‌های Stata و MS Excel، "ارزیابی پیمایش‌های مصرف خوراک"، یک پایگاه داده اطلاعات جمع‌آوری شده از پیمایش‌های ۱۰۰ کشور را ارائه می‌دهد. پایگاه داده با فرم ارزیابی مطابقت دارد و فرهنگ لغت داده‌های^۲ لازم را ارائه می‌دهد.

منابع مرتبط به اشتراک گذاشته شده است:

1. Intermediary output
2. Data dictionary



Food consumption or expenditure surveys - Assessment form (version 3.0)

IHSN - September 2012

This is a PDF form used for assessing the reliability and relevance of food consumption or expenditure modules in household surveys. This form was used to assess a large collection of surveys and generates a met database. This is version 3.0 (September 2012) of the form.



Assessment of the Reliability and Relevance of the Food Data Collected in National Household Consumption and Expenditure Surveys

Lisa C. Smith, Olivier Dupriez and Nathalie Troubat - February 2014

Food consumption data are collected in most countries through a variety of household surveys. The primary objective of these surveys is usually to measure poverty, to derive consumption patterns needed for the calculation of consumer price indices, or to provide input to the compilation of national accounts. Increasingly, these data are re-purposed and used to calculate food security indicators, to compile food balance sheets, to plan and monitor food-based nutrition interventions, to serve information needs of the private sector, and for other research work. This report is thus based on a desk review of survey questionnaires and methods, not on an assessment of the data themselves.

[Download](#)



Presentation of the Food Assessment report, New York 2014

Olivier Dupriez, World Bank - 28 February 2014

Powerpoint presentation of the Food Assessment report, made at a side event of the United Nations Statistics Commission on February 28, 2014.

[Download](#)

Related Resources

Documents

- [Assessment of the Reliability and Relevance of the Food Data Collected in National Household Consumption and Expenditure Surveys](#)
📄 (1.87 MB)

- [DRAFT - Household Surveys and the National Accounts - An overview of uses of household survey data in National Accounting](#)
📄 (255.63 KB)



سنجش هزینه‌های غیرخوراکی^۱



این ابتکار، پروژه‌ای که پیمایش‌های مصارف خوراکی خانوار را ارزیابی می‌کرد، تکمیل می‌سازد، و از رویکرد مشابهی برای ارزیابی داده‌های مصارف غیرخوراکی استفاده می‌کند. طراحی یک فرم ارزیابی برای سنجش ارتباط و قابلیت اطمینان داده‌های مصرف و هزینه غیرخوراکی، براساس نیازهای جوامع مختلف کاربران در دست انجام است. از این فرم برای بررسی همان ۱۰۰ پرسشنامه پیمایشی که در پروژه مصارف خوراکی ارزیابی شده بود، استفاده خواهد شد.

وضعیت پروژه: باز

حامی(ان): صندوق امانت دپارتمان توسعه بین‌المللی شماره TF011722 تحت مدیریت گروه داده بانک جهانی، و تسهیلات کمک هزینه توسعه بانک جهانی، کمک هزینه شماره ۰۶-۴۰۰۱۰۰۹، تحت مدیریت دبیرخانه PARIS21 در سازمان همکاری و توسعه اقتصادی

اجرا شده توسط: گروه توسعه داده بانک جهانی

نوع خروجی: (۱) چارچوب ارزیابی برای سنجش داده‌های مصرف یا هزینه غیرخوراکی؛ (۲) گزارشی از ارزیابی جهانی پیمایش‌ها و راهنماهای فنی ناشی از آن؛ و (۳) پایگاه متا داده در مورد روش‌های پیمایش. فراداده‌های گردآوری شده برای تجزیه و تحلیل با کاربران مختلف (از جمله مؤسسه آمار یونسکو و سازمان جهانی بهداشت) به اشتراک گذاشته شده است.

خروجی میانی: فرم تهیه شده و ارزیابی ۱۰۰ پیمایش که در اگوست ۲۰۱۴ آغاز شد و فراداده گردآوری شده در فایل Stata ۱۴ به اشتراک گذاشته شده است:

1. Measuring non-food expenditures



Non-food consumption or expenditure surveys - Assessment form (Final, version 1.0)

World Bank Data Group, Household Survey Development Team - 27 August 2014

This form is being used to assess the non-food component of the consumption or expenditure modules of household 100 surveys.

[Download](#)

The compiled metadata are available in the following Stata 14 file.



Non-food expenditure assessment: metadata (Stata file)

IHSN -

This Stata 14 data file contains the metadata on 100 surveys compiled using the assessment form developed by the IHSN.

[Download \(300K\)](#)



موضوع جنسیت در پیمایش‌ها^۱



در سال ۲۰۰۶، آژانس بین‌المللی و گروه متخصص آمار جنسیت^۲ برای تقویت همکاری و همیاری جهت ارائه آمار مربوط به جنسیت و تفکیک جنسیتی تشکیل شد. آژانس بین‌المللی و گروه متخصص آمار جنسیت، مجموعه‌ای از شاخص‌های جنسیت را انتخاب کرد که براساس پلتفرم اقدام پکن و اهداف توسعه هزاره^۳ بنا شده است.

از آن زمان، کارها و ابتکارات دیگری برای بهبود آمار جنسیت آغاز شده است. این موارد شامل تولید دستورالعمل‌ها و راهنماهایی مانند آمار سوئد^۴، آمار کمیسیون اقتصادی سازمان ملل متحد برای اروپا^۵ و آمار توسعه جنسیت بانک جهانی^۶، راهنمای آمار جنسیت بخش

آمار سازمان ملل متحد^۷ (آتی)، و دستورالعمل‌های بخش آمار سازمان ملل متحد برای تهیه آمار خشونت علیه زنان^۸ (آتی) است. علاوه بر این، ابتکار و شواهد و داده‌ها برای برابری جنسیتی^۹ که اخیراً راه‌اندازی شده است، در حال ارائه روش‌ها و دستورالعمل‌هایی برای کارآفرینی و داده‌های اموال و دارایی است.

برای تکمیل این ابتکارات، شبکه بین‌المللی پیمایش‌های خانوار پروژه‌ای را برای ارزیابی این‌که آیا و چگونه (به خوبی) برنامه‌های پیمایش در کشورهای در حال توسعه اطلاعات لازم را برای تولید شاخص‌های جنسیت و انجام تجزیه و تحلیل در مورد مسائل جنسیت ارائه می‌دهد، آغاز کرد. پروژه دو هدف را دنبال می‌کند: تهیه فهرست جامعی از اطلاعات موجود، شناسایی داده‌هایی که در دسترس نیستند^{۱۰} و اطلاع‌رسانی به کاربران در مورد داده‌های موجود؛ و شناسایی مناطقی که لازم است روش‌های پیمایش بهبود داده شوند.

وضعیت پروژه: باز

1. Gender issues in surveys
2. Inter-agency and Expert Group on Gender Statistics (IAEG-GS)
3. Millennium Development Goals (MDGs)
4. Statistics Sweden's Engendering Statistics
5. United Nations Economic Commission for Europe (UNECE)
6. World Bank's Developing Gender Statistics
7. United Nations Statistics Division's (UNSD) Gender Statistics Manual
8. UNSD's Guidelines for Producing Statistics on Violence against Women
9. Evidence and Data for Gender Equality (EDGE)
10. Data gaps



حامی(ان): صندوق امانت دپارتمان توسعه بین‌المللی بریتانیا شماره TF011722 تحت مدیریت گروه داده بانک جهانی

اجرا شده توسط: این پروژه توسط گروه توسعه داده‌های بانک جهانی و با همکاری هم‌تایان از چندین سازمان شریک (به‌عنوان مثال، سازمان خواربار و کشاورزی ملل متحد، سازمان بین‌المللی کار، سازمان آموزشی، علمی و فرهنگی ملل متحد، موسسه آمار، بخش آمار سازمان ملل متحد، زنان سازمان ملل، سازمان جهانی بهداشت، بانک جهانی) اجرا می‌شود.

رویکرد: ارزیابی نحوه پرداختن به مسائل جنسیت در پیمایش‌ها شامل پنج فعالیت زیر است:

- ۱. شناسایی مسائل اصلی جنسیت که به‌طور بالقوه توسط پیمایش‌های خانوار پوشش داده می‌شود. جنسیت یک موضوع مقطعی^۱ است، بنابراین تعیین محدوده ارزیابی بسیار مهم است. از فهرست شاخص‌های توصیه شده توسط برنامه ابتکار و شواهد و داده‌ها برای برابری جنسیت، مسائل جنسیت که در اسناد کلیدی توصیف شده است و مشورت با متخصصان جنسیت در چندین آژانس جهت تعیین محدوده ارزیابی، استفاده شده است. فهرست شاخص‌ها و اسناد کلیدی در ادامه در بخش "منابع مرتبط" ارائه شده است.



UNSD Gender Statistics Manual

United Nations Statistics Division - May 2013

The gender statistics manual published by the United Nations Statistics Division provided a framework for our assessment.

The manual covers four topics: 1) Users, uses and production of gender statistics; 2) Bringing gender issues into statistics; 3) Integrating a gender perspective into data collection; 4) Analysis and presentation of gender statistics.

[Download](#)


- ۲. ترجمه این داده‌ها به فهرستی از متغیرها و تهیه فرم ارزیابی نیاز دارد. برای ارزیابی ارتباط و پایایی پرسشنامه‌های پیمایش، باید سؤالات و متغیرهای به‌دست آمده از اجرای پیمایش را شناسایی کنیم. این مرحله، یک فرم ارزیابی، متشکل از پرسشنامه در مورد پرسشنامه‌های پیمایش^۲ را فراهم می‌کند. فرم ارزیابی جزء حیاتی پروژه است و توجه زیادی به بررسی و

1. Cross-cutting

2. Questionnaire on survey questionnaires



آزمایش آن توسط کارشناسان و آژانس‌های متعدد، معطوف شده است. دستورالعمل‌های تکمیل فرم نیز در این مرحله تهیه می‌شود. نسخه پیش‌نویس فرم به صورت PDF در دست تهیه است (در زیر ارائه شده است). نسخه نهایی شامل یک فرم مبتنی بر وب است که در اختیار سایر تیم‌ها/آژانس‌های علاقه‌مند به انجام پروژه‌های مشابه نیز قرار می‌گیرد و این امکان را فراهم می‌کند تا محدوده و پوشش پروژه گسترش داده شود.



Gender Assessment Form (version 1.1)
IHSN and World Bank - January 2013

This document provides a PDF version of the form used to conduct an inventory of variables related to gender issues in household surveys, with a short instruction manual. A web-based version of this form is being used by the IHSN to conduct the assessment. The form is provided for information. The development of this form has benefited from input from many agencies and experts.

[Download](#)

۳. ارزیابی مجموعه بزرگی از پیمایش‌ها و سرشماری‌های خانوار. پوشش جغرافیایی این تمرین شامل کلیه کشورهای با درآمد کم و متوسط (براساس طبقه‌بندی کشورها توسط بانک جهانی) است. در زمان و مکان موجود، این پروژه عملیات زیر را در این کشورها ارزیابی خواهد کرد:

- سرشماری نفوس و مسکن؛
- بررسی درآمد و هزینه خانوار یا معادل آن؛
- مطالعه اندازه‌گیری استانداردهای زندگی^۱؛
- پیمایش‌های پرسشنامه شاخص‌های اصلی رفاه^۲؛
- سایر پیمایش‌های نظارت اجتماعی - اقتصادی خاص کشور؛
- پیمایش‌های جمعیت‌شناسی و سلامت^۳؛
- پیمایش‌های خوشه‌ای شاخص‌های چندگانه^۴؛

1. Living Standards Measurement Study (LSMS)
2. Core Welfare Indicators Questionnaire (CWIQ)
3. Demographic and Health Surveys (DHS)
4. Multiple Indicator Cluster Surveys (MICS)



- سرشماری‌ها یا پیمایش‌های کشاورزی؛
- پیمایش‌های نیروی کار؛
- پیمایش‌های شمول مالی؛
- سایر پیمایش‌ها (آنهايي که پوشش سراسری و پیمایش‌های خاص جنسیتی دارند در اولویت هستند).

این پروژه جدیدترین پیمایش‌ها را برای ارزیابی، اولویت‌بندی می‌کند، اما در نهایت، حداکثر فعالیت‌های جمع‌آوری داده‌ها را از سال ۲۰۰۰ تا به امروز پوشش می‌دهد. فهرستی از پیمایش‌های شناسایی شده برای ارزیابی در زیر ارائه شده است. این فهرست مرتب برای افزودن پیمایش‌ها و سایر اطلاعات به‌روز می‌شود.



Gender assessment project - List of surveys reviewed


IHSN / World Bank - 24 February 2014

This MS-Excel file provides the list of surveys and censuses that the IHSN has assessed from a gender perspective.

[Download](#)

- ۴. تولید یک بانک اطلاعات متا. تمام اطلاعات جمع‌آوری شده با استفاده از فرم ارزیابی در قالب بانک اطلاعات ذخیره می‌شود و همچنین در قالب‌های MS Excel، Stata و SPSS برای تجزیه و تحلیل آسان در دسترس قرار می‌گیرد. یک رابط کاربری مبتنی بر وب ایجاد خواهد شد تا دسترسی عموم به این ابزار "کشف داده" امکان‌پذیر باشد. بانک اطلاعات در قالب‌های دیگری نیز در دسترس عموم قرار خواهد گرفت.





Gender Data Navigator (web query tool)

The Gender Data Navigator application is a searchable inventory of gender-related questions found in survey and census questionnaires from low- and middle-income countries. It provides a convenient data discovery tool to researchers interested in identifying surveys and censuses that collected data on a particular topic of interest.

[Download](#)

۵. تهیه گزارش یا کارنامه با موضوع "چقدر (به خوبی) مسائل جنسیتی در پیمایش‌ها و سرشماری‌ها پوشش داده می‌شود."؟ تهیه گزارش توسط شبکه بین‌المللی پیمایش‌های خانوار و بانک جهانی هماهنگ می‌شود، اما از سایر آژانس‌های علاقه‌مند دعوت می‌شود تا به عنوان نویسندگان یا داوران مشترک همکاری کنند.



How (well) are gender issues covered in surveys and censuses?

Gayatri Koolwal - July 2015


This report presents the findings from a desk review of national household survey and census questionnaires to understand the availability of different types of sex-disaggregated data. It also provides recommendations on questions that could be added in the near term to different survey types, to address substantial gender data gaps we observe across countries.

[Download](#)



منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Links

 Beijing Declaration and Platform for Action	 Download
 Inter-agency and Expert Group on Gender Statistics (IAEG-GS)	 Download
 UN Women website	 Download

Documents

 Developing Gender Statistics: A Practical Tool	 Download
 Engendering Statistics - A Tool for Change	 Download
 On Norms and Agency - Conversations about Gender Equality with Women and Men in 20 Countries	 Download
 Women's Empowerment in Agriculture Index (WEAI)	 Download



ماژول‌های استاندارد شده پیمایش سلامت^۱



در سرتاسر جهان، پیمایش‌های خانوار روش ضروری برای جمع‌آوری اطلاعات در مورد پیامدهای سلامت، تعیین‌کننده‌ها، بهره‌مندی از خدمات سلامت و پوشش مداخلات هستند. پیمایش‌های خانوار همچنین امکان تفکیک اطلاعات^۲ را توسط طبقه‌های مختلف ارزیابی عدالت^۳ برای اندازه‌گیری توزیع خطرات، نتایج و مداخلات فراهم می‌آورند. پیمایش‌ها روشی مقرون به‌صرفه و بسیار قابل اعتماد برای به‌دست آوردن بینش در مورد بسیاری از روندهای مختلف سلامت، تعیین‌کننده‌ها و نابرابری‌های موجود در کشورها ارائه

می‌دهند. در حقیقت، برای بسیاری از شاخص‌های خانوار تنها راه برای به‌دست آوردن اطلاعات هستند.

در دوره اهداف توسعه هزاره، بیشتر جمع‌آوری داده‌ها بر مداخلات و پیامدهای مربوط به بیماری‌های واگیر، کمبودهای تغذیه‌ای و سلامت باروری متمرکز بود. این دستور کار همچنان مهم خواهد بود، اما در عین حال اهمیت شرایط و آسیب‌های بیماری‌های غیرواگیر به سرعت در تمام دنیا افزایش یافته است. این انتقال سلامت بر بحث‌های مربوط به اهداف سلامت در دستور کار توسعه (اهداف توسعه پایدار) پس از سال ۲۰۱۵ میلادی تأثیر می‌گذارد، که باید به دستور کار ناتمام اهداف توسعه هزاره، و همچنین بیماری‌های مزمن، سلامت و آسیب‌های روان و پوشش همگانی سلامت^۴ بپردازد. این بدان معنا است که ابزارهای پیمایش، چه از طریق پیمایش‌های ویژه سلامت یا ماژول‌های سلامت در پیمایش‌های اقتصادی-اجتماعی، باید طیف وسیع‌تری از شاخص‌های سلامت را در دهه آینده در همه کشورها پوشش دهند. ماژول‌های استاندارد شده پیمایش خانوار باید مداخلات سلامت و عوامل خطر را برای همه شرایط سلامت اولویت‌دار پوشش دهد، و

1. Standardized health survey modules
2. Disaggregation of information
3. Equity stratifiers
4. Universal Health Coverage (UHC)



همچنین شامل ارزیابی‌های جامع مرگ‌ومیر، سلامت و پاسخگویی باشد. با پیشروی تدریجی کشورها به سمت تحقق پوشش همگانی سلامت، توافق بر سر مجموعه‌ای از مازول‌های پیمایش قابل استفاده توسط کشورها و برنامه‌های پیمایش منطقه‌ای و جهانی حائز اهمیت خواهد بود.



Monitoring progress towards universal health coverage at country and global levels - Framework, measures and targets

World Health Organization and The World Bank - May 2014

This paper proposes a framework for tracking country and global progress towards universal health coverage. Its aim is to inform and guide these discussions and assessments of both aggregate and equitable coverage of essential health services as well as financial protection.

[Download](#)



Global Reference List of 100 Core Health Indicators (Version 4)

World Health Organization - 03 October 2014

WHO has collaborated with international and multilateral partners and countries to move towards agreement on a global reference list of core health indicators that the global community prioritizes for the purposes of monitoring global progress, maintaining program support and advocating for resources and funding. This list was developed from existing recommended lists that have been proposed in the context of international governing bodies and fora, global and regional health initiatives, technical reference groups and programs.

[Download](#)

وضعیت پروژه: باز.

حامی(ان): صندوق امانت دپارتمان توسعه بین‌المللی شماره TF011722 تحت مدیریت گروه داده بانک جهانی، سازمان جهانی بهداشت و بانک جهانی

اجرا شده توسط: این پروژه که توسط سازمان جهانی بهداشت رهبری می‌شود، به‌طور مشترک



توسط سازمان جهانی بهداشت، بانک جهانی و سایر آژانس‌های عضو شبکه بین‌المللی پیمایش‌های خانوار با پشتیبانی مشاوران فردی اجرا می‌شود. گروهی از کارشناسان از چندین آژانس با شرکت در جلسه گروه کارشناسی به پروژه کمک می‌کنند.

نوع خروجی: گزارش ارزیابی ماژول‌های سلامت در پیمایش‌های موجود، و توصیه‌هایی برای بهبود ماژول‌های پیمایش (به همراه دستورالعمل پرسشگران).

اصول و مبنا: شبکه بین‌المللی پیمایش‌های خانوار قصد دارد با سازمان جهانی بهداشت و سایر شرکای خود در این دستور کار همکاری کند، زیرا پیمایش‌های خانوار منبع بسیاری از آمارهای سلامت پیرامون پیامدهای سلامت جمعیت، عوامل خطر و بهره‌مندی خدمات است و امکان تفکیک نتایج براساس موقعیت اقتصادی-اجتماعی، جنسیت و سایر متغیرهای مهم عدالت را فراهم می‌کند. ایجاد مجموعه‌ای مشترک از سؤالات برای به‌دست آوردن داده‌های قابل مقایسه و افزایش قابلیت اطمینان و اعتبار سؤالات پیمایش ضروری است. این امر، کیفیت داده‌ها را بهبود می‌بخشد و تولید سریع نتایج و گزارش‌ها را تسهیل می‌کند. علاوه‌بر این، مرور نظام‌مند سؤالات و ماژول‌ها به احتمال زیاد منجر به آزمایش سؤالات موجود و جدید و حذف سؤالاتی می‌شود که ارزش اطلاعاتی محدودی دارند یا دچار مشکلات اندازه‌گیری مداوم هستند. همچنین می‌توان نسخه‌های کوتاه و طولانی از یک ماژول را در نظر گرفت و درجه‌بندی قدرت شواهد برای پایایی و اعتبار سؤالات را ارزیابی کرد. این ممکن است عناصر متقابل فرهنگی خاصی داشته باشد و به این معنی است که سؤالات هماهنگ قبل از انجام پیمایش در کشورهای با زبان‌های گفتاری متنوع، نیاز به ارزیابی دارند.

ماژول‌های متمایز زیادی در پیمایش‌های سلامت برای تولید داده در جنبه‌های مختلف مرگ‌ومیر، سلامت و بیماری استفاده می‌شوند. مجموعه‌ای از ماژول‌های پیمایش که استانداردهای بین‌المللی را برآورده می‌کنند و فقط شامل سؤالات مصاحبه و تست‌های بیولوژیکی قابل اعتماد و معتبر می‌شوند، اساس برنامه پیمایش سلامت در هر نقطه از جهان را تشکیل می‌دهند. چنین مجموعه‌ای از ماژول‌ها باید تمام موضوعات سلامت در اولویت، بیماری‌های غیرواگیر و واگیر، شرایط حاد و مزمن و سلامت کودکان، بزرگسالان و افراد مسن را پوشش دهد. اگر کشوری به‌طور منظم، به عنوان مثال، برنامه



پیمایش سلامت دوسالانه داشته باشد، این برنامه می‌تواند وسیله‌ای برای اجرای انعطاف‌پذیر چنین ماژول‌هایی باشد، که تحت تأثیر نیاز داده‌های ملی و بین‌المللی و ملاحظات اندازه‌گیری قرار می‌گیرد. گنجاندن ماژول‌ها در پیمایش‌های اقتصادی و سایر پیمایش‌ها یک امکان بیشتر است اما با توجه به پیچیدگی پیمایش‌های سلامت، دامنه محدودی خواهد داشت.

ایجاد هنجارها و استانداردها برای ماژول‌ها و سؤالات پیمایش سلامت، کیفیت و قابلیت مقایسه را افزایش می‌دهد، اعتبار را نشان می‌دهد، استفاده از سؤالات و ماژول‌های تست شده ضعیف را به حداقل می‌رساند و اجرای انعطاف‌پذیر ماژول‌های پیمایش را برای برآوردن نیازهای کشور تسهیل می‌کند.

اسنادی که در منابع مرتبط زیر یافت می‌شوند، مطالعات پیش‌نویس ماژول‌های پیمایش موجود هستند که به عنوان ورودی برای بحث توسط گروهی از کارشناسان که در ۲۰-۱۹ نوامبر در واشنگتن دی سی تشکیل شد، تهیه شده است.

منابع مرتبط و لینک‌ها به اشتراک گذاشته شده است:

Related Resources

Documents

Module Adult and Maternal Mortality and Orphanhood 12 NOV 2014	(930.52 KB)
Module Antimicrobial Resistance 25 JUN 2014	(117.82 KB)
Module Child Health - Feeding Practices 17 NOV 2014	(172.03 KB)
Module Domestic Violence 12 OCT 2014	(135.24 KB)
Module Fertility and Contraceptive Use 17 NOV 2014	(206.37 KB)
Module Fruits Vegetables 13 OCT 2014	(260.7 KB)
Module Health Expenditure 30 SEPT 2014	(588.15 KB)
Module Health Insurance 10 OCT 2014	(481.33 KB)
Module HIV-AIDS 30 SEPT 2014	(107.04 KB)
Module Hypertension 13 OCT 2014	(251.73 KB)
Module Immunization Coverage 22 JUL 2014	(270.48 KB)
Module Malaria 5 SEPT 2014	(276.02 KB)
Module Maternal Newborn Health 8 OCT 2014	(1.86 MB)
Module Morbidity Self Assessed Health 5 SEPT 2014	(115.71 KB)
Module Risk Factors: Physical Activity 19 OCT 2014	(292.48 KB)
Module Risk Factors: Tobacco Use 13 SEPT 2014	(122.23 KB)
Module Tuberculosis 5 SEPT 2014	(81.58 KB)
Module Water and Sanitation 30 OCT 2014	(117.74 KB)
Suggestions to improve the collection of data on adult mortality in DHS surveys	(56.83 KB)



داده برای برنامه‌های غنی‌سازی مواد غذایی^۱



متخصصان تغذیه عموماً معتقدند که پیمایش‌های مصرف غذا، استاندارد طلایی برای جمع‌آوری داده‌های مصرف مواد غذایی هر فرد است. با این حال، پیمایش‌های مربوط به مصرف غذا گران هستند، انجام آنها دشوار است و با خطای اندازه‌گیری قابل توجهی مواجه می‌شوند. اکنون شناخت روزافزونی وجود دارد که پیمایش‌های درآمد و هزینه خانوار (شبکه بین‌المللی پیمایش‌های خانوار) جایگزینی است که می‌تواند معیارهای غیرمستقیم مصرف غذا را ارائه دهد. اگرچه تعدادی از ویژگی‌های ابزارهای شبکه بین‌المللی پیمایش‌های خانوار فعلی به عنوان کاستی‌های رایج در طراحی و ارزیابی برنامه‌های

غنی‌سازی شناخته شده‌اند، اما می‌توان این کاستی‌ها را با اصلاحات نسبتاً جزئی حذف کرد یا بهبود بخشید. هدف این پروژه ایجاد دستورالعمل‌هایی برای کمک به بهبود شبکه بین‌المللی پیمایش‌های خانوار به عنوان ابزاری برای طراحی و ارزیابی برنامه‌های غنی‌سازی بود.

وضعیت پروژه: بسته

حامی (ان): تسهیلات کمک هزینه توسعه بانک جهانی، کمک هزینه شماره ۰۶-۴۰۰۱۰۰۹-۰۶، تحت مدیریت دبیرخانه PARIS21 در سازمان همکاری و توسعه اقتصادی

اجرا شده توسط: Jack Fiedler، مشاور شبکه بین‌المللی پیمایش‌های خانوار

نوع خروجی: مقاله کاربردی



Strengthening Household Income and Expenditure Surveys as a Tool for Designing and Assessing Food Fortification Programs

John L. Fiedler - June 26, 2009

These guidelines are intended to help improve household income and expenditure surveys as a tool for designing and assessing fortification programs, and thereby aid in accelerating the development of more evidence-based fortification programs.

[Download](#)



1. Data for food fortification programs



منابع مرتبط به اشتراک گذاشته شده است:

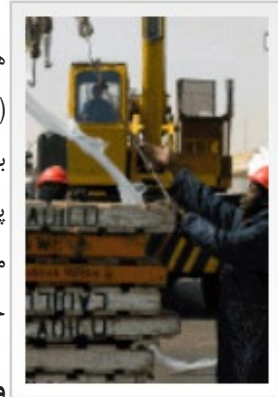
Related Resources

Documents

- [Household income and expenditure surveys: A tool for accelerating the development of evidence-based fortification programs](#)  Download
- [Still waiting for Godot? Improving Household Consumption and Expenditures Surveys \(HCES\) to enable more evidence-based nutrition policies](#)  Download

استفاده از پیمایش‌های خانوار در حساب‌های ملی^۱

هدف از این پروژه توصیف بسیاری از کاربردهای ممکن از پیمایش‌های خانوار (به عبارت دیگر، شبکه بین‌المللی پیمایش‌های خانوار) در تدوین حساب‌های ملی بود. اساساً، هدف اصلی، شناسایی معیارهایی برای قابلیت اطمینان و مرتبط بودن پیمایش‌های خانوار از دیدگاه تدوین‌کنندگان حساب‌های ملی و شاخص قیمت مصرف‌کننده^۲ بود تا در نهایت دستورالعمل‌هایی برای ارزیابی و بهبود پیمایش‌های خانوار ایجاد شود.



وضعیت پروژه: باز

حامی(ان): صندوق امانت دپارتمان توسعه بین‌المللی شماره TF011722 تحت مدیریت گروه داده بانک جهانی

اجرا شده توسط: Willem van den AnDEL، مشاور شبکه بین‌المللی پیمایش‌های خانوار

نوع خروجی: مقاله کاربردی

توجه: پیش‌نویس زیر تهیه شده است اما هنوز توسط دبیرخانه شبکه بین‌المللی پیمایش‌های خانوار ویرایش نشده است. قبل از انتشار به عنوان یک سند شبکه بین‌المللی پیمایش‌های خانوار، برای نظرات بیشتر منتشر می‌شود، ویرایش می‌شود و با ورودی‌های بیشتر تکمیل می‌شود.

1. Use of household surveys in national accounts
2. National accounts and CPI compilers



DRAFT - Household Surveys and the National Accounts - An overview of uses of household survey data in National Accounting

Willem van den Aniel - August 2012

This document describes the many possible uses of HIES in the compilation of the national accounts. Possible uses of the data in the national accounts depend on both the data contents of the available HIES as well as the level of detail at which the national accounts are compiled.

[Download](#)

This paper was produced as an input to two other projects that assessed how (well) household consumption/expenditure is measured in household surveys. One project focuses on [assessing the measurement of food consumption](#), the other on [assessing the measurement of non-food consumption/expenditure](#). For both projects, the approach consisted of describing the key uses and users of these data, specifying criteria and a framework for assessing the relevance and reliability of data from their particular perspective, and conducting a global assessment of many surveys with a view to deriving guidelines and recommendations.



چارچوب ارزیابی کیفیت پیمایش^۱



چک‌لیست چارچوب ارزیابی کیفیت پیمایش، مجموعه‌ای از سؤالات است که به مدیران پیمایش کمک می‌کند تا کیفیت پیمایش‌های خود را بهبود بخشند. در مورد جنبه‌های مختلف فرآیند پیمایش، سؤال می‌پرسد و تأیید، مستندسازی، و پیاده‌سازی سیستم‌ها را برای به حداقل رساندن خطاها و اطمینان از کامل بودن اطلاعات تشویق می‌کند.

وضعیت پروژه: باز

حامی(ان): تسهیلات کمک هزینه توسعه بانک جهانی، کمک هزینه شماره ۰۶-۴۰۰۱۰۰۹، تحت مدیریت دبیرخانه PARIS21 در سازمان همکاری و توسعه اقتصادی

اجرا شده توسط: دبیرخانه شبکه بین‌المللی پیمایش‌های خانوار؛ Sistemas Integrales (مشاوران)؛ مرکز خدمات آماری دانشگاه ری‌دینگ (مشاوران)

نوع خروجی: مقاله کاربردی؛ چارچوب ارزیابی

توجه: پیش‌نویس زیر توسط دبیرخانه شبکه بین‌المللی پیمایش‌های خانوار بازنگری خواهد شد، با مدل عمومی فرآیند کسب و کار آماری سازگارتر خواهد بود، و با عنوان مدیریت کیفیت پیمایش‌ها - کتاب گزارش مدیریت پیمایش مبتنی بر مدل عمومی فرآیند کسب و کار آماری^۲ خواهد بود.



DRAFT - Survey Quality Assessment Framework (SQAF)

Statistical Services Centre of the University of Reading (based on a document prepared by Sistemas Integrales and the IHSN Secretariat) - 2009

The SQAF checklist is a series of questions to help survey managers improve the quality of their surveys. It asks questions about multiple aspects of the survey process, and encourages checking, documentation, and implementation of systems to minimize the introduction of errors and ensure completeness of information

[Download](#)

1. Survey quality assessment framework (SQAF)
2. Managing Quality of Surveys – A GSBPM-based Survey Manager Log Book

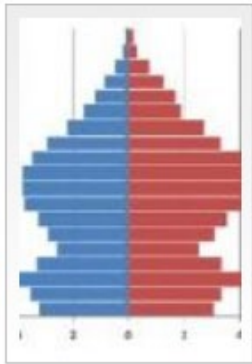


تجزیه و تحلیل و استفاده داده

شبکه بین‌المللی پیمایش‌های خانوار به ارائه و انتشار ابزارها و دستورالعمل‌هایی برای استفاده و تجزیه و تحلیل بهتر ریزداده‌های پیمایش کمک می‌کند.

اولین پروژه برای ساخت ابزارها و دستورالعمل‌های ارائه جمعیت‌های مصنوعی^۱ آغاز شد.

کالیبراسیون نمونه



در کشورهای صنعتی، میزان پاسخ‌دهی در پیمایش‌های خانوارها اغلب پایین است و حداقل برای دو دهه به‌طور پیوسته در حال کاهش بوده است. بنابراین، تکنیک‌های کالیبراسیون نمونه، با بهره‌گیری از اطلاعات جانبی برای تنظیم وزن نمونه، اجرا می‌شوند. در کشورهای در حال توسعه، میزان پاسخ معمولاً بالا است (معمولاً بالای ۹۰ درصد). بنابراین، تنظیمات پیچیده وزن نمونه برای عدم پاسخگویی، ضروری تلقی نمی‌شوند و به ندرت اجرا می‌شوند. اما نشانه‌هایی وجود دارد که وضعیت در برخی از این کشورها در حال تغییر است، که ممکن

است بر پایایی برآوردهای شاخص‌های اجتماعی-اقتصادی به‌دست آمده از پیمایش‌های نمونه تأثیر بگذارد.

پروژه کالیبراسیون نمونه (با بودجه مشترک صندوق امانت شبکه بین‌المللی پیمایش‌های خانوار و برنامه دانش بانک جهانی برای تغییر^۲) با هدف ارزیابی ارتباط تکنیک‌های کالیبراسیون نمونه در کشورهای با درآمد متوسط انجام شد. سه کشور، که در آن‌ها «ناهنجاری‌هایی^۳» در توزیع سنی جمعیت در پیمایش‌های خانوار شناسایی شد، برای یک کار آزمایشی انتخاب شده‌اند. تکنیک‌های کالیبراسیون نمونه برای مجموعه داده‌های مربوطه اعمال شد.

اهداف پروژه عبارت بودند از: (۱) ارزیابی این که آیا و چگونه تکنیک‌های کالیبراسیون نمونه کیفیت شاخص‌های اقتصادی-اجتماعی (به ویژه معیارهای فقر و نابرابری^۴) را بهبود می‌بخشد؛ (۲) تهیه دستورالعمل‌ها و مواد آموزشی برای اجرای تکنیک‌ها؛ و (۳) راه‌اندازی یک دوره آموزشی در مورد کالیبراسیون نمونه برای آماردانان رسمی.

1. Development of synthetic populations
2. Knowledge for Change Program of the World Bank
3. Anomalies
4. Inequality



این پروژه با یک نمونه تمرین کالیبراسیون و با تهیه دستورالعمل‌ها و مواد آموزشی مرتبط آغاز شد. پس از آن یک دوره آموزشی منطقه‌ای برگزار شد. این کارگاه، که توسط بانک جهانی سازماندهی شد توسط بانک توسعه آسیایی در مانیل برگزار شد. این کارگاه، با حضور کارشناسان آمار رسمی از فیلیپین، تایلند و ویتنام برگزار شد.

تمام مواد پروژه، به استثنای مجموعه داده‌های پیمایش کشور، در ادامه، آورده شده است.

وضعیت پروژه: بسته: پروژه بین ژوئن سال ۲۰۱۵ تا اوت سال ۲۰۱۶ میلادی اجرا شد

حامی (ان): برنامه دانش بانک جهانی برای تغییر، کمک هزینه شماره TFOA1077 و دپارتمان توسعه بین‌المللی بریتانیا (صندوق امانت شماره TF011722 که توسط گروه توسعه داده بانک جهانی اداره می‌شود)

اجرا شده توسط: این پروژه توسط گروه توسعه داده‌بانک جهانی، با پشتیبانی کارشناسی از اداره آمار ایتالیا اجرا می‌شود. بانک توسعه آسیایی نیز با میزبانی کارگاه آموزشی در مانیل (می ۲۰۱۶) کمک کرد.

نوع خروجی: گزارش مطالعات امکان‌سنجی، و مواد آموزشی

ابزارها: این پروژه از ReGenesees، یک بسته منبع باز R که توسط اداره آمار ایتالیا ارائه شده است، استفاده می‌کند.

خروجی

گزارش مطالعات امکان‌سنجی



A Feasibility Study on Calibration for Selected South-East Asia Household Surveys

Diego Zardetto - November 2015

Modern large-scale household surveys are expected to provide high quality estimates of population parameters, compatible across data sources. However, signals of possible bias have occasionally been detected for some South-East Asia countries. For instance, the World Bank's Household Survey Unit found significant discrepancies between survey-based estimates of age-sex and household size distributions and the corresponding Census counts for Vietnam, Thailand and the Philippines. A feasibility study has been carried out to investigate whether this issue could be solved through a preliminary calibration procedure. Three Proofs of Concept have been carried out using the ReGenesees open source package. The study showed that it was technically feasible to integrate a calibration procedure in the production workflow of all household surveys taken into account.

[Download \(1.4 Mb\)](#)

نرم‌افزار



ReGenesees Package

Diego Zardetto, Istat -

ReGenesees (R Evolved Generalized Software for Sampling Estimates and Errors in Surveys) is a full-fledged R (open source) software developed and disseminated by the Italian Statistics Office.

ReGenesees is a tool for design-based and model-assisted analysis of complex sample surveys. The package (and its graphical user interface package ReGenesees.GUI) run under Windows, Mac OS, Linux and most Unix-like operating systems.

[Download](#)



مواد آموزشی

برنامه آموزشی و ارائه‌ها



Calibration Training: Program, slides, and demos

Diego Zardetto - May 2016

This package (zip file) contains the course program, slides, and selected demos (R scripts) used for the 4-day training course on sample calibration organized by the World Bank and Asian Development Bank in Manila, May 2016. In addition to this package, a zip file containing exercises and solutions (R scripts) is made available.

[Download \(6.5 Mb\)](#)

تمرین‌ها و مجموعه داده‌های مرتبط



Calibration Training: Exercises (R / ReGenesees) and datasets

Diego Zardetto - May 2016


This package (zip file) contains the exercises and R scripts (with related CSV datasets) used for the 4-day training course on sample calibration organized by the World Bank and Asian Development Bank in Manila, May 2016. During the course, statisticians from 3 participating countries also worked on their own survey datasets. These datasets are the property of the respective national statistical agencies, and are not publicly available. Another package also available from this website provides the course program, slides, and selected demos (R scripts).

[Download \(1.5 Mb\)](#)

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Tools

 [ReGenesees Package](#)

 [Download](#)



تولید جمعیت مصنوعی^۱



شبکه بین‌المللی پیمایش‌های خانوار و گروه توسعه داده بانک جهانی در پروژه‌های با هدف ترویج استفاده از ریزداده سرشماری و پیمایش موجود برای شبیه‌سازی‌های مبتنی بر عامل^۲ و مدل‌سازی میکرو^۳ همکاری کردند. یکی از اجزای مهم این پروژه ارائه (یا بهبود) روش‌ها و ابزارهای تولید جمعیت‌های مصنوعی بود. منطق تولید چنین مجموعه‌های داده مصنوعی این است که (۱) داده‌های مصنوعی تا حد زیادی مسئله خطر افشای آماری را حل می‌کند، و (۲) روش اجازه ترکیب (تلفیقی) داده‌ها از منابع و انواع مختلف را می‌دهد، در نتیجه مجموعه داده‌ای با

ارتباط خاص برای مدل‌سازی و شبیه‌سازی ایجاد می‌کند. این پروژه با همکاری تیمی از کارشناسان در اتریش منتهی به انتشار simPop، یک بسته منبع باز R شد.



simPop : An Open Source R Package for Generating Synthetic Populations

Bernhard Meindl, Matthias Templ, Andreas Alfons, Alexander Kowarik, with contributions from Mathieu Ribatet - 18 November 2014

simPop is an R (open source) package for generating synthetic populations based on survey data and ancillary information. The package includes model-based methods, calibration and combinatorial optimization tools.

[Download](#)

وضعیت پروژه: بسته (اما فعالیت‌های بعدی در حال اجرا است)

فعالیت‌های در حال انجام یا برنامه‌ریزی شده مرتبط با بسته simPop R package برای تولید جمعیت‌های مصنوعی واقعی استفاده خواهد شد. گد و داده‌ها منتشر خواهد شد (عمدتاً به عنوان مواد آموزشی برای کاربران علاقه‌مند استفاده می‌شود). سایر فعالیت‌های مربوط به ریزشبه‌سازی و Geo referencing جمعیت‌های مصنوعی در حال انجام یا برنامه‌ریزی است.

حامی(ها): صندوق امانت دپارتمان توسعه بین‌المللی به شماره TF011722 با مدیریت بانک جهانی، گروه توسعه داده بانک جهانی و بانک جهانی

1. Generation of synthetic populations
2. Agent-based simulations
3. Micro modeling





اجرا شده توسط: گروه توسعه داده بانک جهانی و شرکت مشاوره

نوع خروجی: (۱) نرم افزار منبع باز (بسته R با دفترچه راهنما)؛ و (۲) دستورالعمل‌های فنی.

منابع مرتبط به اشتراک گذاشته شده است:

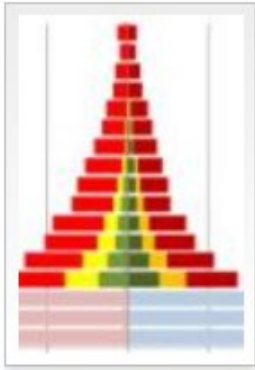
Related Resources

Tools

 [simPopulation: R package for the generation of synthetic populations](#)  [Download](#)



میکروشبیه‌سازی پویا برای پیش‌بینی‌های جمعیت^۱



این پروژه با هدف ابهام‌زدایی از استفاده از میکروشبیه‌سازی پویا برای پیش‌بینی‌های جمعیت و نشان دادن امکان‌سنجی و ارتباط آن در کشورهای در حال توسعه انجام شد. ابزارهای برنامه‌نویسی پیشرفته و رایگان در دسترس و بهبود در دسترسی و کیفیت ریزداده‌ها به امکان‌پذیر شدن میکروشبیه‌سازی پویا با هزینه‌ای معقول کمک کرده است. هسته اصلی این پروژه یک مثال کاربردی برای موریتانی است، که روش‌های آماری مورد استفاده در میکروشبیه‌سازی و تکنیک‌های برنامه‌نویسی میکروشبیه‌سازی را معرفی می‌کند، و می‌تواند برای

کشورهای دیگر تطبیق و تکرار شود. این مدل دارای یک رابط کاربری گرافیکی بصری است و بر روی یک کامپیوتر شخصی استاندارد اجرا می‌شود. کُد و فایل‌های تجزیه و تحلیل آماری آن در دسترس سازندگان مدل است، که می‌توانند از آنها به عنوان یک کتاب مرجع و جعبه ابزاری برای ارائه و پیاده‌سازی مدل میکروشبیه‌سازی استفاده کنند.

وضعیت پروژه: بسته (اما فعالیت‌های بعدی در حال اجرا است)

فعالیت‌های در حال انجام یا برنامه‌ریزی شده مرتبط: این پروژه با تمرکز جدید بر پیش‌بینی جمعیت بر اساس سطح
تحصیلات و Geo referencing

به جمعیت‌های پیش‌بینی شده، در حال گسترش به کشور دوم است.

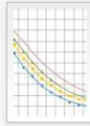
حامی(ها): صندوق امانت دپارتمان توسعه بین‌المللی به شماره TF011722 با مدیریت بانک جهانی؛ برنامه سوم دانش بانک جهانی برای تغییر کمک هزینه شماره TFOA1095 که توسط بانک جهانی اداره می‌شود؛ بودجه بانک جهانی

اجرا شده توسط: گروه توسعه داده بانک جهانی و مرکز تحلیل سیاسی موریتانی

نوع خروجی: مدل پیش‌بینی جمعیت تکرارپذیر و سازگار با میکروشبیه‌سازی پویا، توسعه‌یافته با استفاده از MOD-GEN (نرم افزار رایگان توسط Statistics Canada) و دستورالعمل‌های دقیق

خروجی

1. Dynamic micro simulation for population projections



Dynamic micro-simulation: two Powerpoint presentations

Martin Spielauer - 2017

This zip file contains two Powerpoint presentations by Martin Spielauer (World Bank consultant), dated February 2017 (presentation at University of Southampton) and May 2017 (presentation at the Mauritanian Center for Policy Analysis). The presentations cover the rationale and methods of dynamic micro-simulation, and describe a portable population projection application.

[Download \(1.2 MB\)](#)



DYNAMIS-POP: Presentation at IMA 6th World Congress

Martin Spielauer, Olivier Dupriez -

Powerpoint presentation, presented at the 6th World Congress of the International Microsimulation Association (Torino, Italy, 21-23 June 2017).

[Download \(1 MB\)](#)



A population projection model for Mauritania - Technical report (HTML)

May 2017 (version v01)

This report describes in detail the rationale, approach of micro-simulation applied to population projection. It provides all information needed to modelers interested in replicating or adapting the model developed in collaboration with the Mauritanian Center for Policy Analysis (CMAP) and Office of National Statistics (ONS). The model is named DYNAMIS-POP (for DYNAMIC Micro-Simulation - POPulation Projection). This report is related to version 1 of the model (DYNAMIS-POP-MRT-2017_v01). The source code of the model, the parameters tables, the Stata scripts and Excel files produced to generate the parameters tables and scenarios, are also made publicly available. The model is developed using MODGEN (MODEL GENERator, a freeware by Statistics Canada).

[Download](#)

A population projection model for Mauritania - Technical report (PDF)

May 2017 (version v01)

This report describes in detail the rationale, approach of micro-simulation applied to population projection. It provides all information needed to modelers interested in replicating or adapting the model developed in collaboration with the Mauritanian Center for Policy Analysis (CMAP) and Office of National Statistics (ONS). The model is named DYNAMIS-POP (for DYNAMIC Micro-Simulation - POPulation Projection). This report is related to version 1 of the model (DYNAMIS-POP-MRT-2017_v01). The source code of the model, the parameters tables, the Stata scripts and Excel files produced to generate the parameters tables and scenarios, are also made publicly available. The model is developed using MODGEN (MODEL GENERator, a freeware by Statistics Canada).

[Download \(17 MB\)](#)



DYNAMIS-POP-MRT_v01 - Replication files for model developers

May 2017

This file contains all Modgen files (model code and scenario) needed to replicate the production version of the DYNAMIS-POP-MRT-2017_v01. Another file is also made available on this web page, which provides a step-by-step version of the development of the model (in 19 steps). We also provide in this web page a USER version of the model (executable Modgen model), to be used by those interested in implementing the model with the parameters we provide, or with new scenarios/parameters.

[Download \(25 MB\)](#)



DYNAMIS-POP-MRT_v01 - Replication files for model users

May 2017

Executable version of the model DYNAMIS-POP-MRT-2017_v01. This collection of files allows replication of the model. A baseline population of 100,000 individuals, and all parameters and scenarios, are provided. Running the executable version of the model requires prior installation of the MODGEN Prerequisite software (freeware by Statistics Canada; link provided in this web page).

[Download \(3 MB\)](#)

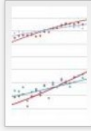


DYNAMIS-POP-MRT_v01 - Replication files for model developers (step-by-step version)

Martin Spielauer (World Bank consultant) - May 2017

This zip file contains all source code (Modgen language) used for the development of the DYNAMIS-POP model developed for Mauritania in 2016/2017. The model was developed in 19 steps, starting with the replication of a simple cohort-based population projection model, then progressively adding complexity showcasing the power of the micro-simulation approach. A technical report is also available, which provides a very detailed description of the modules, their purpose, and their design.

[Download \(260 MB\)](#)



DYNAMIS-POP-MRT-2017_v01: Parameters

May 2017

This file contains all parameters (Excel files) used to run the micro-simulation population projection model developed for Mauritania (DYNAMIS-POP-MRT-2017_v01). The Excel files can be used to modify the assumptions made to run the model. They also provide charts and tables describing the parameters.

[Download \(1.7 MB\)](#)



DYNAMIS-POP-MRT-2017_v01: Stata scripts

May 2017

This zip file contains a collection of Stata (version 14) do files used to analyze data from the Mauritania Population Census 2013 and Multiple Indicators Cluster Survey (MICS) 2015. The log files resulting from running the do files on the source datasets are also provided. This analysis of the census and survey data was conducted to generate a starting population and to derive parameters (core and relative risk tables) for the population projection model. The methods applied in these Stata scripts include simple tabulations, and various regression models (including proportional hazard regression models). NOTE: The source datasets for these scripts are the property of the Office of National Statistics (ONS) of Mauritania. The scripts that require access to the Population Census data were run by ONS staff. Other scripts apply to the MICS 2015 survey dataset, which is (or will be) made publicly available on-line by UNICEF (visit <http://mics.unicef.org/surveys>).

[Download \(90 K\)](#)



Modgen - A Micro-simulation modeling freeware from Statistics Canada

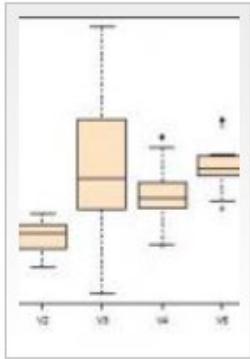
Statistics Canada -

The model DYNAMIS-POP-MRT was developed using Modgen 12 (Model Generator), a generic programming language (freeware) supporting the creation, maintenance and documentation of dynamic micro-simulation models. Modgen is developed and disseminated by Statistics Canada. Two components need to be downloaded to use Modgen: Modgen itself (to develop or edit models), and Modgen Prerequisites (to run models as executable files). Developing or adapting Modgen models require a C++ compiler; such as the free Microsoft Visual Studio Community Edition 2017.

[Download](#)



نقاط پرت در پیمایش‌های مصرف^۱



فقر و نابرابری براساس مصرف خانوار یا داده‌های هزینه درآمد، یا در برخی موارد براساس داده‌های درآمد خانوار اندازه‌گیری می‌شود. همان مجموعه داده‌های پیمایش برای اهداف متعدد دیگری مانند تجزیه و تحلیل و نظارت بر تغییرات در الگوهای مصرف خانوار استفاده می‌شوند.

این نوع داده‌های پیمایش پیچیده هستند و توسط مراکز ملی آمار با ظرفیت مالی و فنی محدود اجرا می‌شوند. بنابراین هنگام تجزیه و تحلیل داده‌ها، توجه خاص باید به کیفیت داده‌ها معطوف شود و باید تکنیک‌هایی نه تنها برای شناسایی مسائل، بلکه برای حل آنها تا حد امکان اجرا شود.

موضوع تشخیص و تثبیت نقاط پرت در داده‌های مصرف یا درآمد، تا حدی به دلیل فقدان راهنمایی یا ابزار روشن، توجه کافی را به خود جلب نکرده است. تحت چتر برنامه کاری شبکه بین‌المللی پیمایش‌های خانوار، گروه توسعه داده بانک جهانی در حال اجرای یک پروژه تحقیقاتی متشکل از یک ارزیابی در مقیاس بزرگ از این موضوع است.

وضعیت پروژه: باز

حامی(ها): صندوق امانت دپارتمان توسعه بین‌المللی به شماره TF011722 با مدیریت بانک جهانی، گروه توسعه داده بانک جهانی، و بودجه بانک جهانی

اجرا شده توسط: دانشگاه فناوری وین، اتریش

نوع خروجی: خروجی پروژه شامل موارد زیر خواهد بود:

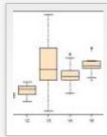
- ارزیابی دقیق الگوریتم‌های تشخیص نقاط پرت (گزارش طولانی، با یافته‌های کامل)؛
- یک سند کوتاه‌تر، ارائه راهکارهای عملی برای شناسایی و تثبیت موارد پرت در مجموعه داده‌های پیمایش هزینه‌ها؛
- برنامه‌ها و اسکریپت‌ها^۲ (بسته R یا اسکریپت‌های مبتنی بر بسته‌های موجود، و اسکریپت‌های Stata) که برای تشخیص و جانمایی^۳ موارد پرت استفاده می‌شوند.

1. Outliers in consumption surveys
2. Scripts
3. Imputation



خروجی

نسخه طولانی گزارش



Detecting outliers in household consumption survey data

Peter Filzmoser, Johannes Gussenbauer, Matthias Templ - April 26, 2016

To assess the relevance and impact of various outliers detection and imputation methods on household expenditure data in sample surveys, the World Bank and the IHSN commissioned a comprehensive review of existing algorithms, including an assessment of their implementation on actual survey datasets. This report presents the main findings of this work. A shorter version of this report, with practical guidelines and instructions for implementation in R and Stata, will also be produced.

[Download \(1.1 Mb\)](#)

دستورالعمل‌ها و اسکریپت‌های فنی

در دست تهیه است.

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

Detecting outliers in household consumption survey data

(1.1 MB)



Stata برای تجزیه و تحلیل داده‌های پیمایش: کتابچه راهنمای آموزشی و موارد مرتبط^۱



Stata یک بسته پرکاربرد برای تجزیه و تحلیل داده‌های پیمایش است. این کتابچه راهنما و موارد مرتبط (مجموعه داده آموزشی، مثل کُد، تمرینات) مطالبی را برای خودآموزی یا استفاده توسط مربیان فراهم می‌کند. تمام دستوراتی که معمولاً برای مدیریت داده‌ها، ارزیابی کیفیت، ایجاد متغیرها، جدول‌بندی، تولید نمودارها و تحلیل‌های پایه (به عنوان مثال، رگرسیون) استفاده می‌شوند، توسط کتابچه راهنمای کاربر پوشش داده می‌شوند.

وضعیت پروژه: باز:

حامی(ها): بانک جهانی و دپارتمان توسعه بین‌المللی بریتانیا، (صندوق امانت شماره TF011722 که توسط گروه توسعه داده بانک جهانی اداره می‌شود)

اجرا شده توسط: این پروژه توسط گروه توسعه داده بانک جهانی اجرا می‌شود. تهیه این کتابچه راهنمای کاربر از ورودی‌های ارائه شده توسط متخصصین آمار پیمایش از دبیرخانه جامعه اقیانوس آرام به مناسبت یک دوره آموزشی Stata در نوامبر، ژوئیه سال ۲۰۱۶ میلادی، استفاده می‌کند.

نوع خروجی: مواد آموزشی (راهنما، مجموعه داده‌ها، تمرین‌ها، اسکریپت‌های Stata).

1. Stata for survey data analysis: training manual and related materials
2. Pacific Community



گزینش و جمع‌داده‌ها

معمولاً، مراکزی که پیمایش‌ها را اجرا می‌کنند و حامیان بین‌المللی این‌گونه فعالیت‌ها، توجه کمی به مستندسازی، انتشار و حفظ ریزداده‌های پیمایش دارند که این امر، منجر به از بین رفتن یا استفاده سوء از داده‌های پیمایش و فراداده می‌شد. اما، هم تقاضا برای ریزداده‌های پیمایش از سوی جامعه پژوهشی و هم توجه به مستندسازی و به اشتراک‌گذاری ریزداده‌ها افزایش یافته است. استانداردها و شیوه‌های جدید، اغلب در جوامع دانشگاهی پدیدار شده است. مراکز ملی آمار و سایر تولیدکنندگان رسمی داده‌ها، قوانین و خدمات خود را به متقاضیان، مُدرن کرده‌اند، و اکنون توجه بیشتری به حفظ داده‌های پیمایش و مواد مرتبط، و انتشار ریزداده‌ها نشان می‌دهند.

شبکه بین‌المللی پیمایش‌های خانوار با همکاری نزدیک با مؤسسات و مراکز داده مختلف متخصص در نگهداری ریزداده‌ها، در ارائه یا تطبیق استانداردها، ابزارها، و دستورالعمل‌های استفاده مراکز ملی آمار کار می‌کند.

کار شبکه بین‌المللی پیمایش‌های خانوار در این زمینه شامل موارد زیر است:

- کمک به ارائه استاندارد فراداده طرح مستندسازی داده‌ها، به ویژه کتاب‌کُد طرح مستندسازی داده‌ها؛
 - پشتیبانی ارائه یک ویرایشگر فراداده تخصصی طرح مستندسازی داده‌ها (DDI) برای مستندسازی داده‌های پیمایش و سرشماری مطابق با استاندارد فراداده طرح مستندسازی داده‌ها (DDI)؛
 - پشتیبانی ارائه نرم‌افزار متن باز و رایگان و دستورالعمل‌هایی برای بی‌نام‌سازی (دیکدینگ) ریزداده‌ها؛
 - تهیه دستورالعمل‌هایی برای حفظ داده‌های دیجیتال و فراداده.
- علاوه بر این پروژه‌ها، شبکه بین‌المللی پیمایش‌های خانوار همچنین، ارائه ابزارهای کاتالوگ کردن ریزداده‌ها و دستورالعمل‌های خط‌مشی انتشار ریزداده‌ها را پشتیبانی می‌کند.



استاندارد فراداده طرح مستندسازی داده‌ها



طرح مستندسازی داده‌ها (DDI) یک استاندارد فراداده بین‌المللی است که برای توصیف پیمایش‌های اجتماعی-اقتصادی، سرشماری‌ها و سایر فعالیت‌های جمع‌آوری ریزداده‌ها طراحی شده است. استاندارد طرح مستندسازی داده‌ها (DDI) توسط پیمان طرح مستندسازی داده‌ها ارائه و نگهداری می‌شود، و شبکه بین‌المللی پیمایش‌های خانوار یک شریک فعال در ارائه آن است. براساس بازخورد آژانس‌های کشورهای با درآمد کم و متوسط که از ابزارهای شبکه بین‌المللی پیمایش‌های خانوار مبتنی بر طرح مستندسازی داده‌ها استفاده

می‌کنند، و از آژانس‌های منطقه‌ای و بین‌المللی مانند بانک جهانی، شبکه بین‌المللی پیمایش‌های خانوار درخواست ارتقای مشخصات طرح مستندسازی داده‌ها (DDI) را کرد که منجر به انتشار نسخه ۲٫۵ استاندارد کتاب‌گد طرح مستندسازی داده‌ها شد. شبکه بین‌المللی پیمایش‌های خانوار همچنین با پیمان طرح مستندسازی داده‌ها در صدور گواهینامه کتاب‌گد طرح مستندسازی داده‌ها (DDI) به عنوان استاندارد ISO کار می‌کند. علاوه بر این، شبکه بین‌المللی پیمایش‌های خانوار تهیه مقاله‌ای را سفارش داد که استانداردهای مختلف فراداده مربوط به مدیریت ریزداده را توصیف می‌کرد.

وضعیت پروژه: باز

حامی(ها): تسهیلات کمک هزینه توسعه بانک جهانی، کمک هزینه شماره ۴۰۰۱۰۱۲۰۶، تحت مدیریت دبیرخانه PARIS21 در سازمان همکاری و توسعه اقتصادی

اجرا شده توسط: DDI Alliance و مشاوران شبکه بین‌المللی پیمایش‌های خانوار

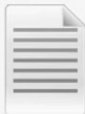
نوع خروجی: استاندارد بین‌المللی فراداده XML و یادداشت فنی در مورد استانداردهای فراداده



DDI Alliance website

The DDI Alliance, hosted by the University of Michigan, maintains the DDI metadata standard. Its website provides detailed information on the DDI Codebook and DDI Lifecycle specifications, and provides a catalog of tools that make use of the DDI standard.

[Download](#)



Metadata Standards - IHSN technical note (DRAFT)

Metadata Technology Ltd - October 2012

These technical notes describe the various metadata standards relevant for statistical data, especially microdata, documentation.

[Download](#)


For more information on the DDI and other metadata standards, see our [Guidelines on Metadata Standards](#).

The IHSN also supports the development of a [DDI Metadata Editor](#) and other DDI-compliant tools; see our [Software](#) section.

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

 [Metadata Standards - IHSN technical note \(DRAFT\)](#)

 (255.8 KB)



ویرایشگر فراداده طرح مستندسازی داده‌ها^۱



هدف از این پروژه ارائه یک ویرایشگر تخصصی فراداده رایگان، کاربرپسند و قدرتمند مطابق با استاندارد DDI به تولیدکنندگان و متصدیان داده است. شبکه بین‌المللی پیمایش‌های خانوار از ارائه نرم‌افزار رایگان توسعه یافته توسط آرشیو داده‌های علوم اجتماعی نروژ پشتیبانی می‌کند.

وضعیت پروژه: بسته

حامی(ها): تسهیلات کمک هزینه توسعه بانک جهانی، کمک هزینه شماره

۴۰۰۱۰۰۹-۰۶، ۴۰۰۱۰۱۰-۰۶، ۴۰۰۱۰۱۱-۰۶ تحت مدیریت دبیرخانه PARIS21 در سازمان همکاری و توسعه اقتصادی؛ صندوق امانی صندوق امانت دپارتمان توسعه بین‌المللی؛ صندوق توسعه سیاست و منابع انسانی ژاپن

اجرا شده توسط: آرشیو داده‌های علوم اجتماعی نروژ،^۲ ارائه نرم افزار؛ مشاوران بانک جهانی و شبکه بین‌المللی پیمایش‌های خانوار، ترجمه برنامه کاربردی و تولید دستورالعمل‌ها و مواد آموزشی

نوع خروجی: نرم‌افزار رایگان

The software and other related materials are freely available. See our [Software - DDI Metadata Editor](#) section.

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

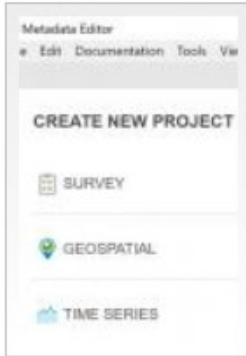
 Quick Reference Guide for Data Archivists

 [Download](#)

1. DDI metadata editor
2. Norwegian Social Science Data Archive (NSD)



ویرایشگر فراداده چنداستاندارد^۱



بانک جهانی در حال ارائه یک ویرایشگر فراداده جدید است تا جایگزین نرم‌افزار رایگان Nesstar Publisher شود. این ویرایشگر جدید فراداده با چندین استاندارد فراداده (DDI برای ریزداده، ISO 19115 برای داده‌های ژئوفضایی، Dublin Core) سازگار خواهد بود و به کاربران اجازه می‌دهد تا طرح‌ها و قالب‌های فراداده خود را بسازند. این نرم‌افزار مبتنی بر وب به عنوان نرم‌افزار متن باز منتشر خواهد شد.

وضعیت پروژه: باز

حامی(ها): بانک جهانی

اجرا شده توسط: گروه توسعه داده بانک جهانی و فرم ارائه نرم‌افزار

نوع خروجی: نرم افزار متن باز

موارد قابل تحویل: نسخه کامل بتا

1. Multi-standard metadata editor



کنترل افشای آماری (بی‌نام‌سازی) - ارائه نرم‌افزار^۱

برای پرداختن به موضوع حقوقی محرمانگی داده‌ها، شبکه بین‌المللی پیمایش‌های خانوار در حال تولید دستورالعمل‌ها و ابزارهای کنترل افشای آماری^۲، بی‌نام‌سازی ریزداده، با همکاری کارشناسان و متخصصان است. هدف، ارائه راه‌حل‌های عملی برای ارزیابی و کاهش خطر افشا در فایل‌های ریزداده است که به تولیدکنندگان داده اجازه می‌دهد فایل‌های ریزداده بی‌نام مفید اما ایمن برای انتشار را مطابق با قوانین ملی و سیاست‌های انتشار تولید کنند. شبکه بین‌المللی پیمایش‌های خانوار همچنین مطالعات موردی را توسط ادارات ملی آمار فیلیپین و اروگوئه سفارش



می‌دهد.

وضعیت پروژه: باز

حامی(ها): تسهیلات کمک هزینه توسعه بانک جهانی، کمک هزینه شماره ۴۰۰۱۰۱۲-۰۶، ۴۰۰۱۰۱۰-۰۶، ۴۰۰۱۰۱۱-۰۶ تحت مدیریت دبیرخانه PARIS21 در سازمان همکاری و توسعه اقتصادی

اجرا شده توسط: PARIS21 و بانک جهانی، گروه توسعه داده‌ها، با مشاوران شبکه بین‌المللی پیمایش‌های خانوار در مراحل مختلف پروژه

نوع خروجی: دستورالعمل‌های فنی، مطالعات موردی و نرم‌افزار متن باز

نرم افزار و دستورالعمل‌های کاربر به صورت رایگان در دسترس هستند؛ به بخش‌های دیگر پیرامون نرم‌افزار - کنترل افشای آماری و دستورالعمل‌های مربوط به بی‌نام‌سازی ریزداده‌ها رجوع کنید.

شرح پروژه

در ۲۰ سال گذشته، ابتکارات بسیاری برای توسعه دانش و به اشتراک‌گذاری تخصص و منابع در زمینه کنترل افشای آماری (SDC) شکوفا شده است. برخی از این ابتکارات صرفاً آکادمیک هستند، برخی دیگر توسط ادارات آمار ملی هدایت می‌شوند و برخی دیگر، جوامع علاقه‌مند به کنترل افشای آماری (SDC) را دخیل می‌کنند. شایان ذکر

1. Statistical disclosure control (anonymization) - Software Development
2. Statistical disclosure control (SDC)



است، چندین پروژه اروپایی در کنترل افشای آماری (SDC) به‌دست آمده‌اند، که از پروژه 4th Framework SDC (۱۹۹۶-۱۹۹۸) شروع شده و با پروژه 5th Framework CAS (2000-2003) CENEX (۲۰۰۶) و دو پروژه ESSnet (۲۰۰۸-۲۰۱۳) در مورد کنترل افشای آماری و دسترسی از راه دور به ریزداده‌ها در یک محیط امن ادامه یافته است. این جریان‌های کاری باعث ایجاد mu-argus شده‌اند، که برای مدت طولانی تنها نرم‌افزار موجود برای محافظت از ریزداده‌ها بوده است. علی‌رغم این ابتکارات، فقط راهنمایی‌ها و کمک‌های فنی محدودی در مورد کنترل افشای آماری (SDC) در اختیار مراکز قرار گرفته است.

رواج نرم‌افزارهای رایج تجزیه و تحلیل آماری در مراکز، شبکه بین‌المللی پیمایش‌های خانوار را به ایجاد برنامه‌های تخصصی مستندسازی شده برای SPSS، Stata و SAS سوق داد که به جلوگیری از مشکلات نگهداری و پشتیبانی طولانی‌مدت کمک کرد؛ اکثر کاربران و سازمان‌های آنها نمی‌خواهند یا نمی‌توانند روی آموزش نرم‌افزار جدید سرمایه‌گذاری کنند. علاوه بر این، استفاده از نرم‌افزار تخصصی کنترل افشای آماری (SDC) موجود، به دلیل (۱) نگرانی در مورد پایداری توسعه و پشتیبانی موجود؛ (۲) مستندسازی ضعیف؛ (۳) عدم سازگاری کاربر؛ و مهمتر از همه، (۴) مسائل مربوط به عملکرد و ارتباط با مجموعه داده‌های پیمایش بزرگ کاملاً رضایت بخش نبوده است. با حمایت و مشارکت کارشناسان مختلف، شبکه بین‌المللی پیمایش‌های خانوار مجموعه‌ای از افزاینده‌ها (Plug-ins) را برای ++C ارائه داد که عملکرد بهینه را پشتیبانی می‌کنند. Plug-ins با موفقیت در Stata ۸، ۹ و ۱۰، SPSS 16+ و Windows/Linux در خط فرمان (Command line) آزمایش شدند و برای تکنیک‌های بی‌نام‌سازی زیر توسعه یافته و بهینه شده‌اند، که به‌طور گسترده استفاده شده و در مطالعه شرح داده شده است:

● اندازه‌گیری خطر

- ۱. اندازه‌گیری خطر SUDA-DIS
- ۲. اندازه‌گیری خطر نمونه وزنی mu-argus (انفرادی و سلسله مراتبی)
- ۳. (k-بی‌نام) k-anonymity
- ۴. (I-تنوع) I-diversity

● کاهش خطر

- ۱. کُدگذاری مجدد محلی (براساس حداکثر الگوریتم تطبیق وزنی)
- ۲. k-بی‌نام (با استفاده از منحنی پُر کردن فضای هیلبرت)



- ۳. مبادله رتبه‌های عددی
 - ۴. اضافه صدا
 - ۵. MDAV (الگوریتم ریزتجمع با طول ثابت)
 - ۶. PRAM
 - ۷. نمونه‌گیری (اجرای دو روش نمونه‌گیری: نظام‌مند و متوازن؛ نمونه‌گیری را می‌توان برای ایجاد نمونه‌های فرعی با احتمال نمونه‌گیری بسته به متغیرهای عددی حساس یا خود اندازه‌گیری خطر استفاده کرد)
- در سال‌های اخیر، محیط نرم‌افزار آماری R در محافل آماری دانشگاهی و رسمی برای اهداف آماری پیشرفته، جامع‌تر و مرتبط‌تر شده است. در حال حاضر R پیش‌روترین نرم‌افزار آماری متن باز است و با افزایش محبوبیت، R در حال تبدیل شدن به یک زبان برنامه‌نویسی استاندارد در زمینه خود است. از آنجایی که فرض بر این است که این روند ادامه خواهد داشت، پیاده‌سازی پلاگین‌های ++C شبکه بین‌المللی پیمایش‌های خانوار در R دارای چندین مزیت است که عبارتند از:
- کُد ++C را می‌توان در یک محیط نرم‌افزار آماری رایگان و متن باز استفاده کرد؛
 - این روش‌های جدید در نرم‌افزارهای آماری محبوب ارائه می‌شوند؛
 - ادغام کُد ++C امکان محاسبات سریع در R را فراهم می‌کند.
- بسته R package sdcMicro مجموعه شناخته شده‌ای از روش‌های حفاظت از ریزداده‌ها است که توسط سازمان آمار اتریش توسعه یافته است و در حال حاضر در چندین سازمان آمار ملی استفاده می‌شود. sdcMicro در طول ۵ سال گذشته به یکی از ابزارهای استاندارد برای محافظت از ریزداده تبدیل شده است. شبکه بین‌المللی

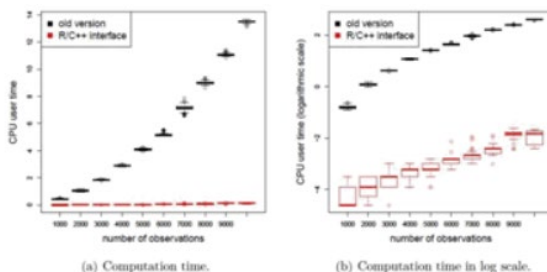


پیمایش‌های خانوار از توسعه بیشتر SdcMicro پشتیبانی می‌کند و با توسعه‌دهندگان آن برای موارد زیر همکاری کرده است:

- روش‌های مرتبط SdcMicro موجود در پلاگین‌های شبکه بین‌المللی پیمایش‌های خانوار را درج کند؛
- SdcMicro را روی مجموعه داده‌های واقعی آزمایش کند تا خروجی‌های آن را کالیبره کند و تفسیر آنها را تسهیل کند؛
- دستورالعمل‌های عملی برای پشتیبانی از استفاده از جعبه ابزار ایجاد کند و به کاربران کمک کند بین روش‌ها و الگوریتم‌های مرتبط حرکت کنند.

SdcMicro در حال حاضر شامل چندین روش محبوب برای بی‌نام‌سازی ریزداده‌ها می‌شود؛ برخی از این روش‌ها را می‌توان در پلاگین‌های ++C شبکه بین‌المللی پیمایش‌های خانوار نیز یافت. روش‌های همپوشانی با اجرای مشابه آنها در SdcMicro آزمایش و مقایسه شده‌اند. سه روش جدید (یا پیاده‌سازی‌های بهبودیافته) در SdcMi- cro گنجانده شده است: suda2 (به عبارت دیگر، یافتن نمونه‌های منحصر به فرد)، مبادله رتبه (به عبارت دیگر، تعویض رتبه عددی)، و mdav (به عبارت دیگر، ریزتجمع). از آنجایی که ++C حاوی ساختارهای کلاس خاصی بود و نیاز به چندین و گاهی اوقات files Header مختلف برای کامپایل کردن کد داشت، گنجاندن این روش‌های جدید در R کار پیچیده‌ای است. شکل ۱ در ادامه، افزایش بهره‌وری زمان محاسبات را بین اجرای قدیمی و جدید تعویض رتبه در SdcMicro، براساس ۱۰۰ اجرا بر روی یک مجموعه داده ۱۰ بعدی با تعداد مشاهدات متفاوت نشان می‌دهد.

شکل ۱ افزایش بهره‌وری زمانی محاسبات بین اجرای قدیمی و جدید الگوریتم‌های مبادله رتبه



نسخه ۴,۴,۰ بسته SdcMicro در شبکه جامع آرشیو R موجود است. دستورالعمل‌های موجود و راهنمای کاربر



SdcMicro در حال به‌روزرسانی هستند. یک آموزش خاص برای نشان دادن نحوه پیاده‌سازی این مفاهیم و الگوریتم‌ها بر روی مجموعه داده‌های واقعی در حال توسعه است. این آموزش با نمونه‌هایی از مجموعه داده SES اتحادیه اروپا تهیه شده است. شبکه بین‌المللی پیمایش‌های خانوار پذیرش SdcMicro و جعبه ابزار دستورالعمل‌های مرتبط را برای ایجاد فایل‌های استفاده عموم و فایل‌های استفاده علمی ترویج می‌کند. به صفحات نرم افزار - کنترل افشای آماری و دستورالعمل‌های مربوط به بی‌نام‌سازی ریزداده‌ها رجوع کنید.

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

- | | |
|---|-------------|
| sdcApp Reference Manual | Download |
| Statistical Disclosure Control for Microdata: A Practice Guide | Download |
| Statistical Disclosure Control for Microdata: A Practice Guide - Case Study Data and R Script | (856.35 KB) |
| Statistical Disclosure Control for Microdata: Theory | Download |

Links

- | | |
|--|----------|
| Statistical Disclosure Control for Micro-Data Using the R Package sdcMicro | Download |
|--|----------|

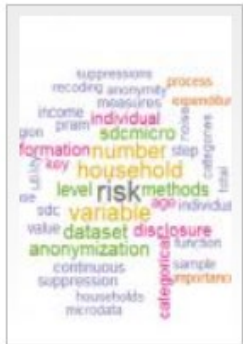


کنترل‌افشای آماری (بی‌نام‌سازی) - راهنمای عملی^۱

چالش انتشار ریزداده: ایجاد تعادل در حفاظت از داده و مطلوبیت داده^۲

وضعیت پروژه: باز

حامی(ها): صندوق امانت دپارتمان توسعه بین‌المللی به شماره TF011722 با مدیریت بانک جهانی، گروه توسعه داده بانک جهانی و بانک جهانی. برنامه دوم دانش بانک جهانی برای تغییر: چالش انتشار ریزداده‌ها؛ ایجاد تعادل در حفاظت از داده‌ها و مطلوبیت داده. Concept Note موجود است.



هزاران پیمایش از طریق برنامه‌های اجرا شده توسط بانک جهانی و شبکه بین‌المللی پیمایش‌های خانوار، مستند و فهرست‌بندی شده‌اند. تنها بخشی از این مجموعه داده‌ها به‌طور آشکار قابل دسترسی هستند. مسائل حقوقی و اخلاقی در مورد حفظ حریم خصوصی داده‌ها و حفاظت از هویت پاسخ‌دهنده هنوز بسیاری از تولیدکنندگان داده از جمله بانک جهانی را از انتشار عمومی برخی از ریزداده‌های خود باز می‌دارد. چندین آژانس با ایجاد فایل‌های استفاده عموم با استفاده از تکنیک‌های حفاظت از داده، این مشکل را برطرف می‌کنند. این تکنیک‌ها می‌توانند برای تولید مجموعه‌های داده «ایمن» بسیار مؤثر باشند، اما ممکن است منجر به از دست دادن اطلاعات قابل توجه و ایجاد داده‌هایی با کاربرد محدود شوند (به عنوان مثال، زمانی که متغیرهای مهم حذف می‌شوند یا در سطوحی جمع می‌شوند که تجزیه و تحلیل را دشوار می‌کند). در بسیاری از موارد، استفاده از روش‌های جایگزین کنترل‌افشا، تعادل بهتری بین حفاظت از داده‌ها و کاربرد داده‌ها را ممکن می‌سازد. برخی از این روش‌ها پیچیده هستند و تعداد متخصصان با تجربه در به کارگیری آنها هنوز اندک است. در حالی که مطالعات قابل توجهی در مورد روش‌های موجود وجود دارد، اسناد عملی گام به گام کمی در دسترس آژانس‌ها وجود دارد که روش‌ها و ابزارهای موجود برای اجرای آنها را به هم مرتبط می‌کند.

تقاضا برای راه‌حل‌های عملی و پشتیبانی فنی برای اعمال کنترل‌افشای آماری (SDC)، که به نام «بی‌نام‌سازی ریزداده‌ها» نیز شناخته می‌شود، وجود دارد. این تقاضا ناشی از نیاز یا الزام تولیدکنندگان داده به انتشار داده‌ها اما در عین حال رعایت مقررات حفاظت از حریم خصوصی است. ارائه راه‌حل‌های کافی و پشتیبانی فنی، این پتانسیل را دارد که تعداد زیادی از مجموعه‌های داده را «Unlock» کند.

1. Statistical disclosure control (anonymization) - Practice Guide
2. Data utility



اطمینان از در دسترس بودن راه حل متن باز رایگان برای آژانس‌ها (بسته توسعه یافته توسط شبکه بین‌المللی پیمایش‌های خانوار) گام مهمی به جلو بود، اما کافی نبود. همچنان دانش تلفیقی و گزارش شده محدودی در مورد تأثیر روش‌های کاهش خطر افشا بر مطلوبیت داده وجود دارد. این دسترسی محدود به دانش همراه با فقدان تجربه در استفاده از ابزارها و روش‌ها، اجرای راه‌حل‌های "بهینه" را برای بسیاری از آژانس‌ها دشوار می‌کند. منظور از بهینه، انجام تعهدات آنها در قبال حفاظت از حریم خصوصی است، اما همچنین تعهد آنها برای انتشار داده‌های مفید جهت نظارت و ارزیابی خط مشی است. این پروژه بانک جهانی تلاش می‌کند تا این شکاف مهم را (۱) با مستند کردن تحقیقات انجام شده در بانک جهانی از طریق ارزیابی مقیاس بزرگ از تکنیک‌های ناشناس‌سازی پُر کند و (۲) این نتایج را به دستورالعمل‌های عملی تبدیل کند.



Statistical Disclosure Control for Microdata: A Practice Guide

Thijs Benschop, Matthew Welch - June 2016

Releasing data in a safe way is required to protect the integrity of the statistical system, by ensuring agencies honor their commitment to respondents to protect their identity. Agencies do not widely share, in substantial detail, their knowledge and experience using SDC and the processes for creating safe data with other agencies. This makes it difficult for agencies new to the process to implement solutions. We consolidated knowledge from literature as well as from our own experience to inform our discussion of the processes and methods presented in this guide. This guide focuses on the implementation of methods and uses the free R based package sdcMicro for its examples. If you are interested in reading in detail about the theory behind the methods used, we suggest reading our accompanying guide: Statistical Disclosure Control for Microdata: Theory.

[Download](#)



Statistical Disclosure Control for Microdata: A Practice Guide - Case Study Data and R Script

June 2016

[Download \(1 MB\)](#)

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

- | | |
|--|--------------------------|
| sdcApp Reference Manual | Download |
| Statistical Disclosure Control for Microdata: A Practice Guide | Download |
| Statistical Disclosure Control for Microdata: Theory | Download |



حفظ ریزداده‌ها^۱

این پروژه در حال تولید راهنمایی برای کمک به تولیدکنندگان رسمی داده در تعریف و برآوردن الزامات و تعهدات حفظ دیجیتالی خود است. توصیه‌ها گام‌هایی را برای توسعه یک روش حفاظتی مؤثر متناسب با نیازها و الزامات تولیدکنندگان رسمی داده‌ها، و برای انطباق با استانداردها و بهترین عملکرد در حفاظت دیجیتالی مشخص می‌کند؛ که مورد دوم، متعهد به مدیریت طولانی‌مدت محتوای دیجیتالی از همه نوع است.

وضعیت پروژه: بسته

حامی(ها): تسهیلات کمک هزینه توسعه بانک جهانی، کمک هزینه شماره ۰۶-۱۰۱۰۱۰۴۰۰ تحت مدیریت دبیرخانه PARIS21 در سازمان همکاری و توسعه اقتصادی

اجرا شده توسط: کنسرسیوم بین دانشگاهی تحقیقات سیاسی و اجتماعی^۲

نوع خروجی: مقاله کاربردی



Principles and Good Practice for Preserving Data

Interuniversity Consortium for Political and Social Research (ICPSR) -
December 2009

This document provides basic guidance for managers in statistical agencies who are responsible for preserving data using the principles and best practices defined by the digital preservation community. The document defines the rationale for preserving data and the principles and standards applied to data preservation; documents the development of a digital preservation policy; and uses digital archive audit principles to suggest best practices for data.

[Download](#)

For more information and materials related to the preservation of digital data and metadata, see our [Guidelines on Preservation](#).

1. Preservation of microdata
2. Inter-university Consortium for Political and Social Research (ICPSR)



کاتالوگ کردن پیمایش

تعداد فزاینده‌ای از مجموعه داده‌های پیمایش، در دسترس جامعه پژوهشی قرار می‌گیرند. دسترسی عمومی به ریزداده‌ها می‌تواند به‌طور قابل توجهی استفاده و ارزش داده‌ها را افزایش دهد. تجربه نشان داده است که انتشار ریزداده‌های به خوبی مستند شده و با دسترسی آسان، اغلب منجر به تولید مقدار زیادی خروجی‌های تحلیلی می‌شود، بدون این که هیچ هزینه‌ای برای محقق اولیه یا حامی پیمایش داشته باشد. برای به حداکثر رساندن استفاده از این مجموعه داده‌ها، کاتالوگ کردن مناسب بسیار مهم است. مکان‌یابی^۱ مجموعه داده‌های مورد نظر اغلب مانعی برای استفاده از آنها است، زیرا ممکن است محققان از وجود یا ویژگی‌های داده‌ها آگاه نباشند.

شبکه بین‌المللی پیمایش‌های خانوار، داده‌ها را ندارد و موظف به توزیع داده‌ها از طرف اعضای خود یا سایر آژانس‌ها نیست. اما شبکه بین‌المللی پیمایش‌های خانوار موارد ذیل را انجام می‌دهد:

- کاتالوگ مرکزی^۲ از پیمایش‌ها و سرشماری‌ها را با فراداده‌های قابل جستجو (با همکاری بانک جهانی) حفظ می‌کند، و در صورت وجود، پیوندهایی به سایت‌هایی ارائه می‌دهد که می‌توان داده‌ها را در آنها به‌دست آورد؛
- کاتالوگی از اسنادها^۳ را حفظ می‌کند که فهرستی از خروجی تحلیلی تولید شده با استفاده از مجموعه داده‌های فهرست شده در کاتالوگ پیمایش مرکزی ارائه می‌دهد (در نتیجه ارزش و اهمیت اشتراک داده را نشان می‌دهد)؛
- دستورالعمل‌های کاتالوگ کردن ریزداده‌های متن باز را توسعه می‌دهد و نگهداری می‌کند، تا به تولیدکنندگان و نگهداری‌کنندگان^۴ داده اجازه دهد تا سیستم‌های پیشرفته انتشار داده را با هزینه کم ایجاد کنند؛
- ابزارهای ذخیره ریزداده متن باز را توسعه می‌دهد و نگهداری می‌کند تا امکان آرشیو داده‌ها برای مدیریت آسان اکتساب مجموعه داده‌ها را دهد.

1. Locating
2. Central catalog
3. Citations
4. Depositors



نگهداری کاتالوگ پیمایش^۱



کاتالوگ شبکه بین‌المللی پیمایش‌های خانوار توسط بانک جهانی با همکاری بسیاری از آژانس‌های بین‌المللی، منطقه‌ای و ملی نگهداری می‌شود. پیمایش‌ها و سرشماری‌ها با استفاده از ویرایشگر فراداده شبکه بین‌المللی پیمایش‌های خانوار برای تولید فراداده مطابق با استاندارد فراداده طرح مستندسازی داده‌ها (DDI) ثبت می‌شوند. سپس این فراداده‌ها در نرم‌افزار متن باز ابزار کاتالوگ کردن آرشیو داده ملی (NADA) که توسط شبکه بین‌المللی پیمایش‌های خانوار توسعه یافته، منتشر می‌شوند.

وضعیت پروژه: باز

حامی(ها): تسهیلات کمک هزینه توسعه بانک جهانی، کمک هزینه شماره ۰۶-۴۰۰۱۰۱۲-۰۶، ۰۶-۴۰۰۱۰۱۰-۰۶
 ۰۶-۴۰۰۱۰۱۱-۰۶ تحت مدیریت دبیرخانه PARIS21 در سازمان همکاری و توسعه اقتصادی؛ صندوق امانت دیپارتمان توسعه بین‌المللی با مدیریت گروه توسعه داده بانک جهانی
اجرا شده توسط: بانک جهانی و بسیاری از آژانس‌ها
نوع خروجی: پایگاه داده (کاتالوگ پیمایش) و وبسایت
پروژه مرتبط: نگهداری از پایگاه داده استناد^۲

Visit the [IHSN Central Catalog](#).

This catalog is built on the [DDI metadata standard](#) and various [free software applications](#). More information on the development and maintenance of a survey catalog can be found in our [guidelines on microdata archiving and dissemination](#).

1. Maintenance of a survey catalog
 2. Citations database



نگهداری کاتالوگ استناد



یکی از مزایای اصلی در دسترس تر و قابل استفاده تر کردن ریزداده‌ها، افزایش کمیت و تنوع تحقیقات است. تجربه نشان می‌دهد که وقتی داده‌های با کیفیت و مستند به راحتی در دسترس قرار می‌گیرند، خروجی تحلیلی زیادی تولید می‌شود. مستندسازی این موارد استفاده ریزداده مهم است. برای ارتقای دسترسی به ریزداده‌ها، پروژه کاتالوگ کردن استنادی شبکه بین‌المللی پیمایش‌های خانوار یک پایگاه داده از استنادها و Quotes از پیمایش‌های فهرست شده در کاتالوگ مرکزی پیمایش، ایجاد و نگهداری می‌کند. استنادها شامل مقالات یا فصل‌هایی از

مجلات دانشگاهی، کتاب‌ها، گزارش‌ها، مقالات کاربردی، پایان‌نامه‌ها و غیره است. برای ایجاد این پایگاه داده استنادها، شبکه بین‌المللی پیمایش‌های خانوار اساساً به Google Scholar (پایگاه‌های اطلاعاتی دیگر ممکن است در آینده استفاده شود) و برنامه کاتالوگ کردن پیمایش آرشیو داده ملی (NADA) که توسط شبکه بین‌المللی پیمایش‌های خانوار توسعه یافته است، متکی است.

کاتالوگ استناد به محققان و حامیان پیمایش امکان می‌دهد:

- تحقیقاتی را که قبلاً در یک منطقه، کشور یا در یک موضوع خاص انجام شده است، شناسایی کنند؛
- خطر تکرار کار تحقیقاتی را کاهش دهند؛
- نحوه استفاده از داده‌های پیمایش را اندازه‌گیری و ارزیابی کنند. این امر به‌ویژه در زمینه ابتکارات داده‌های باز، که استفاده مجدد از داده‌ها به عنوان یک استدلال اصلی برای آن ارائه می‌شود، مهم است. در یک پروژه آینده، و زمانی که کاتالوگ استناد به هدف ۱۵۰۰۰ استناد برسد، شبکه بین‌المللی پیمایش‌های خانوار می‌تواند تجزیه و تحلیل دقیقی از استنادها انجام دهد.

وضعیت پروژه: باز

حامی(ها): تسهیلات کمک هزینه توسعه بانک جهانی، کمک هزینه شماره ۰۶-۴۰۰۱۰۰۹، ۰۶-۴۰۰۱۰۱۰،



۰۶-۲۰۱۱-۴۰۰ تحت مدیریت دبیرخانه PARIS21 در سازمان همکاری و توسعه اقتصادی؛ صندوق امانت دپارتمان توسعه بین‌المللی با مدیریت گروه توسعه داده بانک جهانی

اجرا شده توسط: شبکه بین‌المللی پیمایش‌های خانوار به تیم‌هایی از دانشجویان متکی است که توسط کتابداران حرفه‌ای داده در DataFirst (دانشگاه Cape Town، آفریقای جنوبی) و دانشگاه تورنتو (کانادا) نظارت می‌شوند.

نوع خروجی: پایگاه داده و وبسایت

پروژه مرتبط: نگهداری کاتالوگ پیمایش

The citation catalog is not an independent database—it is part of the [IHSN central survey catalog](#). Users can search the catalog for surveys and view the related citations, or search the citation database for surveys quoted in a particular citation.

نرم‌افزار کاتالوگ پیمایش آرشیو داده ملی (NADA)^۱



کاربران باید به درستی از وجود و ویژگی‌های داده‌های پیمایش موجود مطلع شوند. برای ترویج استفاده از ریزداده‌ها، فراداده‌های دقیق باید در قالب یک کاتالوگ آنلاین قابل جستجو در دسترس باشند. برای پشتیبانی از این تلاش، شبکه بین‌المللی پیمایش‌های خانوار استفاده از استاندارد فراداده طرح مستندسازی داده‌ها (DDI) را ترویج می‌کند، و در حال توسعه یا حمایت از توسعه ابزارها و دستورالعمل‌های منطبق با طرح مستندسازی داده‌ها (DDI)، از جمله ویرایشگر تخصصی فراداده طرح مستندسازی داده‌ها (DDI) و آرشیو داده ملی (NADA)^۲، یک برنامه کاتالوگ کردن متن باز و با دسترسی آزاد است.

وضعیت پروژه: باز

1. NADA survey cataloging software
2. National Data Archive (NADA)



حامی(ها): تسهیلات کمک هزینه توسعه بانک جهانی، کمک هزینه شماره ۴۰۰۱۰۰۹-۰۶، ۴۰۰۱۰۱۰-۰۶، ۴۰۰۱۰۱۱-۰۶ تحت مدیریت دبیرخانه PARIS21 در سازمان همکاری و توسعه اقتصادی؛ صندوق امانت دپارتمان توسعه بین‌المللی با مدیریت گروه توسعه داده بانک جهانی

اجرا شده توسط: بانک جهانی، گروه توسعه داده‌ها، تیم توسعه پیمایش خانوار













نوع خروجی: نرم افزار متن باز

منابع مرتبط به اشتراک گذاشته شده است:

To download the NADA application and related documentation and materials, visit our [Software - Microdata Cataloguing Tool](#) page. For more information on survey cataloguing, see our [guidelines on cataloguing](#).

Related Resources

Documents

 Checklist Sécurité NADA à faire compléter et signer par les consultants	 (0 bytes)
 Directives pour la vérification de la qualité et de la fonctionnalité d'un catalogue NADA	 (26.14 KB)
 Guía para la revisión de la calidad y funcionalidad general del catálogo ANDA	 (26.14 KB)
 Lista de verificaci de la seguridad del ANDA y el servidor web	 (23.36 KB)
 NADA BSD 3-Clause License	 (1.48 KB)
 NADA Security Checklist	 (194.77 KB)



نرم‌افزار نگهداری داده^۱



اژانس‌ها و سازمان‌هایی که آرشیو داده‌ها را مدیریت می‌کنند ممکن است داده‌ها و فراداده‌ها را از چندین منبع دریافت کنند. سازماندهی این سپرده‌های داده برای موارد زیر مهم است:

- اطمینان حاصل کنند که تمام داده‌ها و فراداده‌های لازم ارائه شده است؛
- هزینه‌های دریافت و مستندسازی مجموعه داده‌ها برای انتشار را به حداقل برسانند؛
- اطلاعاتی در مورد آنچه که توسط چه کسی و چه زمانی نگهداری شده است را حفظ کنند؛
- به شخص یا اژانسی که داده‌ها را نگهداری می‌کنند، این فرصت را بدهند که بایگانی داده‌ها را در مورد محدودیت‌های دسترسی مطلع کنند؛
- به نگهداری کننده داده یک «رسید» ارائه دهند که نشان دهد چه چیزی نگهداری شده است.

ابزار نگهداری داده شبکه بین‌المللی پیمایش‌های خانوار یک نرم‌افزار متن باز و سازگار با طرح مستندسازی داده‌ها (DDI) است که می‌تواند به عنوان بخشی از بسته آرشیو داده ملی (NADA) یا به عنوان یک ابزار مستقل استفاده شود. برای اژانس‌هایی که می‌خواهند یک سیستم آنلاین ساده، کاربرپسند و در عین حال قدرتمند و ایمن برای دستیابی به مجموعه داده‌ها پیاده‌سازی کنند، مفید خواهد بود.

وضعیت پروژه: باز

حامی(ها): تسهیلات کمک هزینه توسعه بانک جهانی، کمک هزینه شماره ۰۶-۱۰۱۰۱۰۴۰۰ تحت مدیریت دبیرخانه PARIS21 در سازمان همکاری و توسعه اقتصادی؛ صندوق امانت دپارتمان توسعه بین‌المللی با مدیریت گروه توسعه داده بانک جهانی

اجرا شده توسط: بانک جهانی، گروه توسعه داده

نوع خروجی: نرم افزار منبع باز

Download

The IHSN intends to release this application in March 2013.



انتشار ریز داده‌ها

به‌طور سنتی، تولیدکنندگان رسمی داده‌ها، پیمایش‌های خانوار را انجام می‌دادند و نتایج را در قالب یک گزارش توصیفی یا تحلیلی و مجموعه جداول منتشر می‌کردند. اکنون این آژانس‌ها به‌طور فزاینده‌ای، ریز داده‌ها را در فایل‌های استفاده عموم یا «اکاربرد پژوهشی» منتشر می‌کنند.

فراهم کردن دسترسی جامعه پژوهشی به ریز داده‌ها مزایای عمده‌ای دارد و ارزش قابل توجهی به داده‌ها می‌افزاید. به همین دلیل، شبکه بین‌المللی پیمایش‌های خانوار یک مدافع قوی از انتشار ریز داده است. اما باید اذعان داشت که از دیدگاه تولیدکننده داده، اشتراک‌گذاری ریز داده مستلزم هزینه‌ها و خطراتی است. نه تنها مسائل فنی و مالی بلکه باید به مسائل حقوقی و اخلاقی نیز رسیدگی شود. انتشار ریز داده‌ها باید مطابق با قوانین ملی و براساس سیاست‌ها و رویه‌های روشن انجام شود. برای کمک به آژانس‌ها در تعریف خط‌مشی‌ها و پروتکل‌ها برای انتشار ریز داده‌ها، شبکه بین‌المللی پیمایش‌های خانوار:

- مقاله‌ای تهیه کرد که توصیه‌هایی را برای تدوین سیاست‌ها و شیوه‌های انتشار ریز داده ارائه می‌کند؛
- مقاله‌ای در مورد ابتکار آزادسازی داده‌های سازمان آمار کانادا^۱ ارائه کرد و توضیح داد که چرا و چگونه یک آژانس رسمی آماری از "محافظت از داده‌ها" به مُدِل "اشتراک‌گذاری داده‌ها" حرکت کرد؛
- مقاله‌ای در مورد تجربه مشابه آژانس مرکزی اتیوپی ارائه داد.



دستورالعمل‌های خط مشی انتشار داده‌ها^۱



این پروژه راهنمایی ارائه می‌کند که مروری بر مسائل فنی، قانونی و اخلاقی مربوط به انتشار ریزداده‌ها ارائه می‌دهد. بسیاری از تولیدکنندگان داده به این موضوع علاقه‌مند هستند اما هنوز تمایلی به اشتراک‌گذاری ریزداده‌ها ندارند و به دنبال راهنمایی برای تدوین خط‌مشی‌ها و رویه‌های انتشار ریزداده هستند. این پروژه توصیه‌هایی را براساس تجربه و تخصص چندین آژانس ملی و بین‌المللی ارائه می‌دهد که ابتکارات انتشار ریزداده را آغاز کرده‌اند.

وضعیت پروژه: بسته

حامی(ها): تسهیلات کمک هزینه توسعه بانک جهانی، کمک هزینه شماره ۰۶-۱۰۱۰۴۰۰۱ تحت مدیریت دبیرخانه PARIS21 در سازمان همکاری و توسعه اقتصادی

اجرا شده توسط: Ernie Boyko (مشاور شبکه بین‌المللی پیمایش‌های خانوار) و Olivier Dupriez (بانک جهانی)

نوع خروجی: مقاله کاربردی



Dissemination of Microdata Files - Principles, Procedures and Practices

Olivier Dupriez and Ernie Boyko - August 2010

In all countries, data producers face growing demand for microdata. Determining the best way to disseminate these data is a challenge. Formal policies and procedures defining the conditions of access to microdata must be formulated. This guide provides an overview of such policies and procedures, and documents existing best practices.

[Download \(2.05 MB\)](#)



Diffusion des fichiers de microdonnées - Principes, procédures et pratiques

Olivier Dupriez and Ernie Boyko - August 2010

Les producteurs de données de tous pays sont confrontés à une demande croissante de microdonnées. Décider de la meilleure façon de diffuser ces données constitue un véritable défi. Cela implique l'établissement de politiques et de procédures définissant les conditions d'accès aux microdonnées. Le présent document contient une description générale de ces politiques et de ces procédures, et recense les bonnes pratiques en la matière.

[Download](#)



بث و نشر ملفات البيانات الجزئية - مبادئ وأنظمة وممارسات

أوليفيه دوبريز و إرنه بويكو - أغسطس 2010

في جميع الدول، منتجي البيانات يواجهون طلب متزايد على البيانات الجزئية. تحديد ووضع أفضل الطرق لنشر هذه البيانات يعتبر تحدي. السياسات والإجراءات الرسمية التي تحدد شروط الوصول والنفاذ للبيانات الجزئية والحصول عليها يجب أن تصاغ وتكون واضحة. يقدم هذا الدليل نظرة عامة عن هذه السياسات والإجراءات، ويحتوي هذا الدليل كذلك على الممارسات الجيدة للتعامل مع البيانات الجزئية.







[Download](#)

For more information and materials related to the dissemination of microdata, see our [Guidelines on Dissemination](#).

منابع مرتبط به اشتراك گذاشته شده است:

Related Resources

Documents

-  Diffusion des fichiers de microdonnées - Principes, procédures et pratiques  (1.42 MB)
-  Managing Statistical Confidentiality and Microdata Access - Principles and guidelines of Good Practice  [Download](#)
-  National Statistics Code of Practice - Protocol on Data Access and Confidentiality (UK)  [Download](#)



ابتکار آزادسازی داده‌های کانادا



هدف این پروژه مستندسازی چرایی و چگونگی تصمیم سازمان آمار کانادا برای شروع به اشتراک‌گذاری ریزداده‌ها با محققان بود. تجربه سازمان آمار کانادا سرشار از درس‌هایی برای آژانس‌های دیگر است. سازمان آمار کانادا با محدودیت‌های فنی، قانونی، اخلاقی و منابع مشابهی روبه‌رو بود که اکثر آژانس‌هایی که در حال حاضر ابتکارات مشابهی را آغاز می‌کنند با آن مواجه هستند. این مقاله کاربردی برای ارائه یک "سند مرجع حمایتی" جهت ترویج بیشتر سیاست‌های دسترسی ریزداده‌های باز ارائه شده است.

وضعیت پروژه: بسته

حامی(ها): تسهیلات کمک هزینه توسعه بانک جهانی، کمک هزینه شماره ۰۶-۱۰۱۰۱۰۴۰۰ تحت مدیریت دبیرخانه PARIS21 در سازمان همکاری و توسعه اقتصادی

اجرا شده توسط: Ernie Boyko و Wendy Watkins (مشاوران شبکه بین‌المللی پیمایش‌های خانوار)

نوع خروجی: مقاله کاربردی



The Canadian Data Liberation Initiative: An Idea worth Considering?

Ernie Boyko and Wendy Watkins - November 2011

The Data Liberation Initiative (DLI) is a Canadian program that provides Canadian post-secondary institutions affordable access to Statistics Canada data resources. This paper describes the project and key success factors so that other agencies may determine its applicability to their own situations.

[Download](#)

For more information and materials related to the dissemination of microdata, see our [Guidelines on Dissemination](#).



دستورالعمل‌های اجرای پیمایش

این بخش، دستورالعمل‌ها و بهترین شیوه‌های مربوط به تمام مراحل اجرای پیمایش را ارائه می‌دهد. برخی از آنها توسط یا برای شبکه بین‌المللی پیمایش‌های خانوار تولید شدند، برخی دیگر براساس کار آژانس‌های دیگر است. این بخش به عنوان یک پورتال عمل می‌کند و دستورالعمل‌های آن به موارد زیر می‌پردازد:

- **طراحی برنامه‌های پیمایش^۱**. در بسیاری از کشورهای در حال توسعه، برنامه‌های پیمایش هنوز موقتی^۲ هستند و به‌شدت به بودجه خارجی (و به طور کلی تأمین بودجه) وابسته هستند. در نتیجه، برنامه‌های پیمایش ملی اغلب فاقد ثبات و انسجام در زمان‌بندی، تعیین توالی، محدوده و روش‌های خود هستند. این دستورالعمل‌ها در درجه اول برای آژانس‌های (سازمان‌های) آماری ملی در نظر گرفته شده است.
- **ایجاد بودجه پیمایش^۳**
- **اجرای پیمایش‌ها^۴**. این دستورالعمل‌ها به کل چرخه عمر یک پیمایش مربوط می‌شوند، مطابق مدل فرآیند تجاری آماری عمومی^۵ سازماندهی شده‌اند و توصیه‌های کلی برای همه مراحل اجرای پیمایش ارائه می‌دهند، اما توصیه‌ای برای طراحی ابزارهای پیمایش خاصی ندارند.
- **یکپارچه‌سازی پیمایش‌ها^۶**. براساس طبقه‌بندی‌ها، مفاهیم و روش‌های سازگار با سایر عملیات جمع‌آوری داده‌ها مانند پیمایش‌ها، سرشماری‌ها و جمع‌آوری داده‌های اداری، چنین یکپارچه‌سازی برای ایجاد یک پایگاه دانش منسجم ضروری است، اما دستیابی به آن دشوار است.
- **بایگانی و انتشار ریزداده‌ها^۷**. اگرچه بایگانی و مستندسازی داده‌های پیمایش بخشی از مدل فرآیند تجاری آماری عمومی است، دستورالعمل‌هایی در بخش جداگانه ارائه داده می‌شود زیرا این زمینه‌ای است که شبکه بین‌المللی پیمایش‌های خانوار توجه و منابع زیادی به آن اختصاص داده است.

1. Designing survey programs
2. Ad-hoc
3. Creating survey budgets
4. Implementing surveys
5. Generic Statistical Business Process Model (GSBPM)
6. Integrating surveys
7. Archiving and dissemination of microdata



این بخش همچنین مجموعه‌ای از مقالات کاری شبکه بین‌المللی پیمایش‌های خانوار را ارائه می‌دهد. مجموعه مقالات کار شبکه بین‌المللی پیمایش‌های خانوار برای تشویق تبادل نظر و بحث در مورد موضوعات مربوط به طراحی و اجرای پیمایش خانوارها و تجزیه و تحلیل، انتشار و استفاده از داده‌های پیمایش است.



طراحی برنامه‌های پیمایش

در کشورهای در حال توسعه، پیمایش‌های خانوار اغلب به شکل غالب جمع‌آوری داده‌ها است و اجرای برنامه‌های پیمایش تا حد زیادی توسط حامیان خارجی تأمین می‌شود. این وابستگی زیاد به حمایت مالی خارجی، طراحی و اجرای برنامه‌های پیمایش ملی کاملاً منسجم را تقویت نمی‌کند. پیامدهای برنامه‌های پیمایش موقتی شامل ناسازگاری در پیمایش‌ها و در طول زمان، تکرار فعالیت‌ها، شکاف داده‌ها (موجود نبودن داده‌ها) و ناکارآمدی است. دستورالعمل‌های زیادی برای اجرای پیمایش‌ها وجود دارد، اما راهنمای کمی در مورد زمان‌بندی، تعیین توالی و یکپارچه‌گی برنامه‌های پیمایش در دسترس است. آژانس‌های (سازمان‌های) آماری ملی باید نحوه انجام پیمایش‌های ملی را به عنوان اجزای سیستم خود بهبود ببخشند.

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

 An Accelerated Program for Implementing the Marrakech Action Plan for Statistics in Selected Countries	 (1.24 MB)
 Costing MDG surveys in Sub-Saharan Africa (2006-2015)	 (124 KB)
 Household Sample Surveys in Developing and Transition Countries	 Download
 Household Surveys and the Millennium Development Goals	 Download



بودجه‌بندی پیمایش

بودجه به شرایط محلی در هر کشور بستگی دارد. سطح حقوق کارکنان متفاوت است و تا حد زیادی به وضعیت اقتصادی یک کشور بستگی دارد. با این حال، راه‌هایی برای تنظیم بودجه و زمان‌بندی وجود دارد که در همه عملیات پیمایش خانوار، صرف نظر از کشور، مشترک است.

دست‌کم گرفتن ارزش برخی اقلام، یا عدم پیش‌بینی نوسانات قیمت ابزار بین تدوین بودجه و خرید ابزار، می‌تواند موفقیت پروژه را به‌طور جدی به خطر اندازد. هنگام تدوین بودجه پروژه باید دقت زیادی به خرج داد. تجربه نشان می‌دهد که تمایل به نادیده گرفتن یا دست‌کم گرفتن هزینه برخی از فعالیت‌ها مانند آموزش کارکنان میدانی یا تجزیه و تحلیل، بایگانی و انتشار ریزداده‌ها وجود دارد.



تهیه بودجه

قبل از تهیه بودجه، باید اندازه نمونه، دوره جمع‌آوری داده‌ها و مدت مصاحبه را دانست. تنها در این صورت است که می‌توان تعداد تیم‌های میدانی را تخمین زد. این پارامترهای اساسی بر هزینه‌های کارکنان، هزینه‌های حمل‌ونقل و هزینه سفر، و تجهیزات مورد استفاده تأثیر می‌گذارد. اگر این پارامترها در ابتدا به وضوح تعریف نشده‌اند، توصیه می‌شود بودجه را براساس فرض‌های مختلف اندازه، در چندین نوع مختلف تهیه کنید.

تمام هزینه‌ها باید در بودجه لحاظ شوند، از جمله حقوق کارکنان، صرف نظر از این که آیا توسط نهاد مجری پرداخت می‌شود یا خیر. کمک‌های فنی ارائه شده توسط نهادهای بین‌المللی باید با هزینه واقعی آن ارزیابی شوند، حتی اگر مستقیماً از بودجه پرداخت نشوند. اگر نیازی به خرید وسایل نقلیه نیست زیرا می‌توان از وسایلی که برای پروژه دیگری خریداری شده‌اند، استفاده کرد، ارزش آنها هنوز باید تخمین زده شود و به عنوان هزینه‌ای که باید توسط نهاد مجری متحمل شود، وارد گردد، زیرا در غیر این صورت نهاد مجری از آنها برای اهداف تولیدی دیگری استفاده خواهد کرد. همین فلسفه برای فضای اداری نیز وجود دارد؛ اگر از اتاق‌های داخل نهاد مجری به جای اجاره محل استفاده می‌شود، ارزش محاسبه شده آنها باید در بودجه لحاظ شود.

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

 [Survey costs \(in Household Sample Surveys in Developing and Transition Countries\)](#)

 [Download](#)



اجرای پیمایش‌ها

این دستورالعمل‌ها به‌عنوان یک پورتال عمل می‌کنند و از ساختار مدل فرایند تجاری آماری عمومی (GSBPM) v4.0 پیروی می‌کنند که فرآیندهای آماری - مانند اجرای یک پیمایش - را به‌صورت منسجم توصیف می‌کند.

مدل فرایند تجاری آماری عمومی شامل ۹ مرحله است که هرکدام به فرآیندهای فرعی تقسیم می‌شوند. در این بخش، شرح مختصری از هر مرحله (اقتباس از شرح مدل فرایند تجاری آماری عمومی)، و منابع مرتبط توصیه شده (به‌عنوان مثال، اسناد، ابزارها و لینک‌ها) ارائه می‌شود. ۹ مرحله مدل فرایند تجاری آماری عمومی به شرح زیر است:

- ۱. تعیین نیازهای داده؛
- ۲. طراحی؛
- ۳. ساختن؛
- ۴. جمع‌آوری (شامل ورود داده)؛
- ۵. فرآیند (شامل ویرایش داده)؛
- ۶. تجزیه و تحلیل؛
- ۷. انتشار نتایج پیمایش؛
- ۸. بایگانی؛
- ۹. ارزیابی پیمایش.

علاوه بر این ۹ مرحله، مدل فرایند تجاری آماری عمومی شامل دو جزء اصلی است:

- ۱۰. مدیریت کیفیت؛
- ۱۱. مدیریت فراداده.

ساختار مدل فرایند تجاری آماری عمومی



مدیریت کیفیت

۱	۲	۳	۴	۵	۶	۷	۸	۹
تعیین نیازهای داده	طراحی	ساختن	جمع‌آوری (شامل ورود داده)	فرآیند (شامل ویرایش داده)	تجزیه و تحلیل	انتشار نتایج پیمایش	بایگانی	ارزیابی پیمایش

مدیریت فراداده

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Links

 [Generic Statistical Business Process Model \(GSBPM\)](#)

 [Download](#)



۱. تعیین نیازهای داده

مرحله اول یک پیمایش باید شامل ارزیابی دقیق تقاضا برای آمارهای مشخص شده و ظرفیت - مالی فنی و سازمانی - سازمان آماری برای تهیه آنها باشد.

این مرحله شامل ۶ فرایند فرعی است:

۱,۱ تعیین نیازهای اطلاعاتی - شامل در نظر گرفتن شیوه‌ها، به‌ویژه روش‌های پیمایش است که توسط سازمان‌های آماری ملی و بین‌المللی استفاده می‌شوند که داده‌های مشابهی را تولید می‌کنند و پیمایش‌های مشابه را انجام می‌دهند.

۲,۱ مشورت و تأیید نیازهای اطلاعاتی - درک درستی از نیازهای کاربر لازم است تا سازمان آماری نه تنها آنچه را که انتظار می‌رود، ارائه دهد، بلکه بداند که چه زمانی، چطور، و شاید مهمتر از همه، دلیل آن را بداند. این فرآیند فرعی سازمان آماری بر مشورت با ذینفعان متمرکز است. تکرارهای دوم و بعدی این مرحله بر تعیین این‌که آیا نیازهای شناسایی شده قبلی تغییر کرده است یا خیر متمرکز می‌شود. درک دقیق نیازهای کاربر بخش مهم این فرایند فرعی است.

۳,۱ تعیین اهداف خروجی - این فرایند فرعی خروجی‌های آماری مورد نیاز برای برآوردن نیازهای کاربر که در فرایند فرعی ۲,۱ مشخص شدند را شناسایی می‌کند، شامل ایجاد توافق با کاربران در مورد مناسب بودن خروجی‌های پیشنهادی است.

۴,۱ شناسایی مفاهیم - این فرآیند فرعی، مفاهیم مورد نیاز برای سنجش توسط پیمایش را از دیدگاه کاربر روشن می‌کند. در این مرحله، مفاهیم شناسایی شده ممکن است با استانداردهای آماری موجود همسو باشد. این همسویی و تعریف مفاهیم و متغیرهای آماری که استفاده می‌شوند در فرآیند ۲,۲ مورد بررسی قرار می‌گیرند.

۵,۱ بررسی در دسترس بودن داده - این فرآیند فرعی بررسی می‌کند که آیا منابع داده فعلی می‌توانند الزامات کاربر و شرایطی را که تحت آن در دسترس باشند، از جمله هرگونه محدودیت در استفاده از آنها را برآورده سازد. ارزیابی جایگزین‌های احتمالی معمولاً شامل تحقیق در مورد منابع احتمالی داده‌های اداری و روش‌های آنها برای تعیین این است که آیا آنها برای استفاده برای اهداف آماری مناسب هستند یا خیر. هنگامی که منابع موجود مورد ارزیابی قرار گرفتند، یک استراتژی برای پر کردن شکاف‌های باقی‌مانده (موجود نبودن داده‌های باقی‌مانده) در مورد



نیاز داده‌ها آماده می‌شود. این فرآیند فرعی همچنین شامل ارزیابی کلی چارچوب قانونی است که در آن داده‌ها جمع‌آوری و مورد استفاده قرار می‌گیرد، و بنابراین ممکن است تغییرات پیشنهادی در قوانین موجود یا معرفی یک چارچوب قانونی جدید را شناسایی کند.

۶.۱ تهیه پرونده تجاری^۱ - این فرآیند فرعی، یافته‌های سایر فرآیندهای فرعی را در قالب یک پرونده تجاری مستند می‌کند که برای اجرای فرآیند تجاری آماری^۲ جدید یا اصلاح شده به تأیید می‌رسند. چنین پرونده تجاری معمولاً شامل موارد زیر می‌شود:

- شرح فرآیند تجاری^۳ "همانطور که هست"^۴ (در صورت وجود)، با اطلاعاتی در مورد نحوه تولید آمار فعلی، و برجسته کردن هرگونه ناکارآمدی و موضوعاتی که باید برطرف شوند؛
- راه حل پیشنهادی "بودن"^۵ با جزئیات نحوه توسعه فرآیند تجاری آماری برای تولید آمار جدید یا تجدید نظر شده، ارائه می‌شود؛ و
- ارزیابی هزینه‌ها و مزایا، و همچنین هرگونه محدودیت خارجی.

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

 Household Sample Surveys in Developing and Transition Countries	 Download
 National Statistics Code of Practice - Protocol on Consultation Arrangements Between the National Statistician and UK Government Ministers (UK)	 Download
 National Statistics Code of Practice - Protocol on Customer Service and User Consultation (UK)	 Download

1. Prepare business case
2. Statistical business process
3. Business process
4. As-is
5. To-be



۲. طراحی

این مرحله شامل ۶ فرآیند فرعی است:

۱,۲ طراحی خروجی‌ها - این فرآیند فرعی شامل طراحی دقیق خروجی‌های آماری است که باید تولید شود، از جمله، کار توسعه مرتبط و آماده‌سازی سیستم‌ها و ابزارهای مورد استفاده در فاز ۷ (انتشار). در صورت امکان، خروجی‌ها باید مطابق استانداردهای موجود طراحی شوند. ورودی‌های این فرآیند ممکن است شامل فراداده از مجموعه‌های مشابه یا قبلی، استانداردهای بین‌المللی، و اطلاعاتی در مورد عملکردهای دیگر سازمان‌های آماری از فرآیند فرعی ۱,۱ (تعیین نیازهای اطلاعاتی) باشد.


۲,۲ طراحی توضیحات متغیر - این فرآیند فرعی متغیرهای آماری را که باید از طریق ابزار جمع‌آوری داده جمع‌آوری شود، و همچنین سایر متغیرهای حاصل از آنها در فرآیند فرعی ۵,۵ (استخراج متغیرها و واحدهای آماری جدید) و هر طبقه‌بندی مورد استفاده را تعریف می‌کند. انتظار می‌رود استانداردهای ملی و بین‌المللی موجود تا آنجا که ممکن است رعایت شود. این فرآیند فرعی ممکن است به موازات فرآیند فرعی ۳,۲، طراحی روش جمع‌آوری داده، نیاز باشد، زیرا تعریف متغیرهای جمع‌آوری شده و انتخاب ابزار جمع‌آوری داده ممکن است تا حدودی به هم وابسته باشد. تهیه توصیف‌های فراداده متغیرهای جمع‌آوری و مشتق شده و طبقه‌بندی‌ها، پیش شرط لازم برای فازهای بعدی است.

۳,۲ طراحی روش جمع‌آوری داده - این فرآیند فرعی مناسب‌ترین روش (ها) و ابزار (ها) جمع‌آوری داده را تعیین می‌کند. فعالیت‌های واقعی در این فرآیند فرعی با توجه به نوع ابزار جمع‌آوری مورد نیاز متفاوت خواهد بود، و می‌تواند شامل مصاحبه به کمک رایانه، پرسشنامه‌های کاغذی، رابط‌های داده اداری، و تکنیک‌های یکپارچگی داده باشد. این فرآیند فرعی شامل طراحی الگوهای پرسش و پاسخ همراه با متغیرها و طبقه‌بندی‌های طراحی شده در فرآیند ۲,۲، طراحی توضیحات متغیر است. این طرح همچنین شامل طراحی هرگونه توافق‌نامه‌های رسمی مربوط به تأمین داده، مانند تفاهم‌نامه‌ها، و تأیید مبنای قانونی برای جمع‌آوری داده می‌شود. این فرآیند فرعی توسط ابزارهایی مانند کتابخانه‌های سؤال فعال می‌شود تا استفاده مجدد از سؤال‌ها و ویژگی‌های مرتبط؛ ابزار پرسشنامه برای گردآوری سریع و آسان سؤال‌ها در قالب‌های مناسب برای آزمون شناختی؛ و الگوهای توافق‌نامه برای استانداردسازی شرایط و ضوابط را تسهیل کند. این فرآیند فرعی همچنین شامل طراحی سیستم‌های مدیریت ارائه‌دهنده خاص - فرآیند است.

۴,۲ چارچوب طراحی و روش نمونه‌گیری - این فرآیند فرعی، جمعیت‌های مورد نظر را شناسایی و مشخص می‌کند، چارچوب نمونه‌گیری (و، در صورت لزوم، ثبتي که از آن مشتق شده) را تعریف می‌کند، و مناسب‌ترین معیارها



و روش نمونه‌گیری را که ممکن است شامل سرشماری کامل باشد، تعیین می‌کند. منابع مشترک، ثبت‌های اداری و آماری، سرشماری‌ها، و نمونه پیمایش‌ها هستند. این فرآیند فرعی نحوه ترکیب این منابع در صورت نیاز را توضیح می‌دهد. تجزیه و تحلیل این که آیا چارچوب ارزیابی این که جمعیت هدف چگونه باید انجام شود و یک برنامه نمونه‌گیری تهیه شود را می‌کند: نمونه واقعی توسط فرآیند ۱,۴ (انتخاب نمونه)، با استفاده از روش مشخص شده در این فرآیند فرعی ایجاد می‌شود.



**Designing Household Survey Samples:
Practical Guidelines**

United Nations, Department of Economic and Social Affairs, Statistics
Division - 2008

The purpose of the handbook is to include in one publication sample survey design issues for convenient referral by practicing national statisticians, researchers, and analysts involved in sample survey work and activities. Methodologically sound techniques that are grounded in statistical theory are presented, implying the use of probability sampling at each stage of the sample selection process.

[Download](#)

۵,۲ طراحی روش پردازش آماری - این فرآیند فرعی، روش پردازش آماری را که در فاز ۵ (فرآیند) و فاز ۶ (تجزیه و تحلیل) اعمال می‌شود، طراحی می‌کند. این طراحی می‌تواند شامل مشخصات معمول برای کدگذاری، ویرایش، محاسبه، تخمین، یکپارچه‌سازی، اعتبارسنجی و نهایی شدن مجموعه داده باشد.

۶,۲ طراحی سیستم‌های تولید و جریان کار - این فرآیند فرعی، جریان کار را از جمع‌آوری داده تا بایگانی، ارائه یک نمای کلی از تمام فرآیندهای مورد نیاز در کل فرآیند تولید آماری و اطمینان دادن از این که به‌طور مؤثر در کنار هم بدون شکاف یا اضافات^۱ قرار می‌گیرند را تعیین می‌کند. در طول فرآیند به سیستم‌ها و پایگاه‌های داده مختلف نیاز است. از آنجا که اصل کلی، استفاده مجدد از فرآیندها و فناوری در بسیاری از فرآیندهای تجاری آماری است، ابتدا باید سیستم‌ها و پایگاه‌های داده موجود بررسی شوند تا مشخص شود که آیا برای این فرآیند خاص مناسب هستند یا خیر؛ سپس، در صورت شناسایی شدن هرگونه شکاف، باید راه‌حل‌های جدیدی طراحی شود. این فرآیند فرعی همچنین نحوه تعامل کارکنان با سیستم‌ها، و این که چه کسی و چه زمانی مسئول چه خواهد بود را در نظر می‌گیرد.

منابع مرتبط به اشتراک گذاشته شده است:



Related Resources

Documents

Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study (3 volumes)	
Designing Household Survey Samples: Practical Guidelines	
National Statistics Code of Practice - Protocol on Statistical Integration and Classification (UK)	
Survey Methods and Practices	

۳. ساختن

این فاز، سیستم‌های تولید را به‌طور کامل ساخته و آزمایش می‌کند که برای استفاده در محیط "زنده" آماده هستند. این فاز شامل ۶ فرآیند فرعی است:

۱,۳ ساختن ابزار جمع‌آوری داده - ابزار جمع‌آوری براساس مشخصات طراحی ایجاد شده در فاز ۲ (طراحی) ساخته می‌شود. این فرآیند فرعی همچنین شامل تهیه و آزمایش محتویات و عملکرد آن ابزار است (به‌عنوان مثال، آزمایش سؤال‌ها در یک پرسشنامه). ثبت معیارهای جمع‌آوری داده (به‌عبارت دیگر، پارادیتا) نیز در این فرآیند فرعی مهم است.

۲,۳ ساختن یا تقویت اجزای فرآیند - این فرآیند فرعی فعالیت‌های مورد نیاز برای ساختن و تقویت اجزای جدید نرم‌افزاری موجود برای فرآیند تجاری را شرح می‌دهد، همان‌طور که در فاز ۲ (طراحی) طراحی شده است. اجزاء می‌تواند شامل عملکردها و ویژگی‌های داشبورد، مخازن داده، ابزارهای تبدیل، اجزای چارچوب جریان کار، و ابزارهای مدیریت ارائه‌دهنده و فراداده باشد.

۳,۳ پیکربندی جریان‌های کار^۱ - این فرآیند فرعی، جریان کار، سیستم‌ها، و تحولات مورد استفاده در فرآیندهای تجاری آماری، از جمع‌آوری داده‌ها تا بایگانی خروجی‌های آماری نهایی را پیکربندی می‌کند. این اطمینان را می‌دهد که جریان کار مشخص شده در فرآیند ۲,۳ (سیستم پردازش و جریان کار) در عمل کار می‌کند.

۴,۳ آزمایش سیستم تولید - این فرآیند فرعی بر آزمایش سیستم‌ها و ابزارهای رایانه‌ای، از جمله آزمایش فنی،

1. Workflow



بستن برنامه‌ها و روال‌های جدید، و تأیید این‌که روال‌های موجود از سایر فرآیندهای تجاری آماری برای استفاده مناسب هستند، متمرکز می‌باشد.







۵,۳ آزمایش فرآیند تجاری آماری – این فرآیند فرعی، فعالیت‌ها را برای مدیریت یک آزمون یا پایلوت میدانی از فرآیند تجاری آماری توصیف می‌کند. به‌طور معمول، شامل جمع‌آوری داده در مقیاس کوچک برای آزمایش ابزار جمع‌آوری، و سپس پردازش و تجزیه و تحلیل داده جمع‌آوری شده است تا اطمینان حاصل شود که فرآیند تجاری آماری مطابق انتظار انجام می‌شود. پس از اجرای آزمایشی، ممکن است، لازم باشد به مرحله قبلی بازگشت و ابزارها، سیستم‌ها یا اجزاء را تنظیم کرد. برای یک فرآیند عمده تجاری آماری، به‌عنوان مثال، سرشماری جمعیت، ممکن است چندین بار تکرار شود تا این فرآیند به‌طور رضایت‌بخش کار کند.

۶,۳ نهایی‌سازی سیستم‌های تولید – این فرآیند فرعی شامل فعالیت‌هایی برای قرار دادن سیستم‌های گردش کار، اجزای تغییر یافته و تازه ساخته شده و سایر اجزاء به‌منظور تولید می‌باشد بنابراین آماده استفاده توسط حوزه‌های تجاری است. فعالیت‌های مرتبط شامل مستندسازی اجزای فرآیند، از جمله مستندسازی فنی و راهنمای کاربر، آموزش کاربران تجاری در مورد نحوه عملکرد فرآیند، و انتقال اجزای فرآیند به محیط تولید و اطمینان از عملکرد مطابق انتظار آنها در آن محیط است (این فعالیت همچنین ممکن است بخشی از فرآیند فرعی ۴,۳، آزمایش سیستم تولید، باشد).

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

- 
[Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study \(3 volumes\)](#)
 Download
- 
[National Statistics Code of Practice - Protocol on Managing Respondent Load \(UK\)](#)
 Download
- 
[Survey Methods and Practices](#)
 Download


۴. جمع‌آوری (شامل ورود داده)

این فاز با استفاده از حالت‌های مختلف جمع‌آوری (از جمله استخراج از ثبت‌ها و پایگاه‌های داده اداری و آماری)



همه داده لازم را جمع‌آوری کرده و آنها را در محیط داده مناسب بارگذاری می‌کند. شامل هیچ‌گونه تغییرات داده جمع‌آوری شده نیست، زیرا این موارد در فاز ۵ (فرآیند) مورد بررسی قرار می‌گیرند. این فاز شامل ۴ فرآیند فرعی است:

۱,۴ انتخاب نمونه - این فرآیند فرعی، چارچوب را ایجاد می‌کند و نمونه را برای این تکرار جمع‌آوری انتخاب می‌کند، همان‌طور که در فرآیند ۴,۲ (چارچوب طراحی و روش نمونه‌گیری) مشخص شده است. همچنین شامل هماهنگی نمونه‌ها بین موارد فرآیند تجاری آماری مشابه (به‌عنوان مثال، مدیریت همپوشانی یا پرخش)، و بین فرآیندهای مختلف با استفاده از یک چارچوب یا ثبت مشترک (به‌عنوان مثال، مدیریت همپوشانی یا توزیع بار پاسخ) است. تضمین کیفیت، تأیید، و نگهداری چارچوب و نمونه منتخب نیز در این فرآیند فرعی انجام می‌شوند، اگرچه نگهداری ثبت‌های زیر بنایی، که از آنها چارچوب‌های متعددی برای فرآیندهای تجاری آماری گرفته شده است، به‌عنوان یک فرآیند تجاری جداگانه تلقی می‌شوند. جنبه نمونه‌گیری از این فرآیند فرعی معمولاً برای فرآیندهایی که کاملاً مبتنی بر استفاده از منابع داده از قبل موجود (مانند داده اداری) هستند، مناسب نیست، زیرا چنین فرآیندهایی به‌طور کلی چارچوب‌هایی از داده موجود ایجاد می‌کنند و سپس از روش سرشماری پیروی می‌کنند.



**Designing Household Survey Samples:
Practical Guidelines**

United Nations, Department of Economic and Social Affairs, Statistics
Division - 2008

The purpose of the handbook is to include in one publication sample survey design issues for convenient referral by practicing national statisticians, researchers, and analysts involved in sample survey work and activities. Methodologically sound techniques that are grounded in statistical theory are presented, implying the use of probability sampling at each stage of the sample selection process.

[Download](#)

۲,۴ تنظیم جمع‌آوری - این فرآیند فرعی تضمین می‌کند که افراد، فرآیندها، و فناوری، آماده جمع‌آوری داده در همه حالت‌ها، همان‌طور که طراحی شده است، هستند. فرآیند فرعی در طول زمان انجام می‌شود، و شامل استراتژی،



برنامه‌ریزی، و فعالیت‌های آموزشی در آماده‌سازی برای نمونه خاص فرایند تجاری آماری است. در مواردی که این فرآیند به‌طور منظم تکرار می‌شود، ممکن است برخی یا همه این فعالیت‌ها به‌طور مشخص، برای هر تکرار مورد نیاز نباشند. برای فرآیندهای استثنایی و جدید، این فعالیت‌ها می‌تواند طولانی باشد. این فرآیند فرعی شامل:

- تهیه استراتژی جمع‌آوری؛
- آموزش کارکنان جمع‌آوری؛
- اطمینان از این‌که منابع جمع‌آوری در دسترس هستند، به‌عنوان مثال، لپ‌تاپ؛
- پیکربندی سیستم‌های جمع‌آوری برای درخواست و دریافت داده؛
- اطمینان از امنیت داده جمع‌آوری شده؛
- تهیه ابزار جمع‌آوری (به‌عنوان مثال، چاپ پرسشنامه، پر کردن آنها با داده موجود، بارگذاری پرسشنامه‌ها و داده بر روی رایانه مصاحبه‌کنندگان و غیره).

۳,۴ اجرای جمع‌آوری - این فرآیند فرعی است که در آن جمع‌آوری با ابزارهای مختلف برای جمع‌آوری داده اجرا می‌شود. شامل تماس اولیه با ارائه‌دهندگان و هرگونه پیگیری یا یادآوری بعدی است. زمان و نحوه تماس با ارائه‌دهندگان و پاسخ آنها را ثبت می‌کند. این فرآیند فرعی همچنین شامل مدیریت ارائه‌دهندگان درگیر در جمع‌آوری فعلی است، اطمینان حاصل می‌شود که رابطه بین سازمان آماری و ارائه‌دهندگان داده مثبت است، و نظرات، سؤال‌ها و شکایات را ثبت می‌کند و پاسخ می‌دهد.

۴,۴ نهایی کردن جمع‌آوری (ضبط داده) - این فرآیند فرعی شامل بارگذاری داده و فراداده جمع‌آوری شده در یک محیط الکترونیکی مناسب برای پردازش بیشتر در مرحله ۵ (فرآیند) است. ممکن است شامل دریافت خودکار داده باشد، به‌عنوان مثال، استفاده از ابزارهای تشخیص شخصیت نوری برای استخراج داده از پرسشنامه‌های کاغذی یا تبدیل فرمت فایل‌های داده دریافتی از سایر سازمان‌ها. در مواردی که ابزار جمع‌آوری فیزیکی داده وجود دارد که برای پردازش بیشتر مورد نیاز نیست، مانند پرسشنامه کاغذی، این فرآیند فرعی، بایگانی آن ابزار را مطابق با اصول تعیین شده در مرحله ۸ (بایگانی) مدیریت می‌کند.



Survey Solutions (CAPI application)

World Bank -

Survey Solutions is a Computer-Assisted Personal Interview technology developed by the World Bank. It assists governments, statistical offices and non-governmental organisations in conducting complex surveys with dynamic structures using tablet devices. The software can be tailored to the needs of the clients, allowing them to successfully complete simple and more sophisticated projects: from basic evaluation questionnaires to complicated multistage panel surveys. The software is offered free of charge, its development being co-financed by the World Bank, the Bill and Melinda Gates Foundation and the Food and Agriculture Organization of the United Nations. Surveys can be conducted on low-cost Android tablets.

[Download](#)



Census and Survey Processing System (CSPRO)

United States Census Bureau -

The Census and Survey Processing System (CSPRO) is a public domain software package used by hundreds of organizations and tens of thousands of individuals for entering, editing, tabulating, and disseminating census and survey data. CSPRO is user-friendly, yet powerful enough to handle the most complex applications. It can be used by a wide range of people, from non-technical staff assistants to senior demographers and programmers.

[Download](#)

منابع مرتبط به اشتراک گذاشته شده است:



Related Resources

Tools

- | | |
|---|--|
|  Census and Survey Processing System (CSPro) |  Download |
|  Survey Solutions (CAPI application) |  Download |

Documents

- | | |
|---|--|
|  Census Data Capture Methodology: Technical Report |  Download |
|  Comparative Assessment of Software Programs for the Development of Computer-Assisted Personal Interview (CAPI) Applications |  Download |
|  Designing Household Survey Samples: Practical Guidelines |  Download |
|  Household Sample Surveys in Developing and Transition Countries |  Download |
|  National Statistics Code of Practice - Protocol on Managing Respondent Load (UK) |  Download |
|  Survey Methods and Practices |  Download |
|  Tracking in Longitudinal Household Surveys |  Download |

Links

- | | |
|---|--|
|  Guidelines on the application of new technology to population data collection and capture (ESCAP) |  Download |
|---|--|



۵. فرآیند (شامل ویرایش داده)

این فاز، تمیز کردن سوابق داده^۱ و آماده‌سازی آنها برای تجزیه و تحلیل را توضیح می‌دهد. شامل فرآیندهای فرعی است که داده‌های جمع‌آوری شده را بررسی، تمیز و تبدیل می‌کند، و ممکن است چندین بار تکرار شود.

فازهای "فرآیند" و "تجزیه و تحلیل" می‌توانند تکراری و موازی باشد. تجزیه و تحلیل می‌تواند درک بیشتری از داده را آشکار کند، که ممکن است نشان دهد که نیاز به پردازش بیشتر است. ممکن است قبل از اتمام مرحله "جمع‌آوری" فعالیت‌های فاز "فرآیند" و "تجزیه و تحلیل" آغاز شود. این امر باعث گردآوری نتایج موقت می‌شود که در آن به موقع بودن یکی از دغدغه‌های مهم کاربران است، و زمان در دسترس، برای تجزیه و تحلیل را افزایش می‌دهد. تفاوت اصلی بین این فازها این است که "فرآیند" به تغییر ریزداده مربوط می‌شود، در حالی که "تجزیه و تحلیل" به تلقی بیشتر مجموعه‌های آماری^۲ مربوط می‌شود.

این فاز، شامل ۸ فرآیند فرعی است:

۱,۵ یکپارچه‌سازی داده - این فرآیند فرعی داده‌های یک یا چند منبع را یکپارچه می‌کند. داده ورودی می‌تواند از مخلوطی از منابع داده خارجی یا داخلی و انواع حالت‌های جمع‌آوری، از جمله اکسترنکت داده اداری، و در نتیجه مجموعه داده‌های هماهنگ باشد. یکپارچه‌سازی داده‌ها معمولاً شامل تطبیق و ثبت روال‌های پیوندی برای پیوند داده از منابع مختلف، جایی که این داده‌ها به یک واحد اشاره می‌کنند؛ و اولویت‌بندی، زمانی که دو یا چند منبع (با مقادیر بالقوه متفاوت) حاوی داده‌های یک متغیر باشند، است. یکپارچه‌سازی داده ممکن است در هر نقطه از این فاز، قبل یا بعد از هر فرآیند فرعی انجام شود. همچنین ممکن است موارد متعددی از یکپارچه‌سازی داده در هر فرآیند تجاری آماری وجود داشته باشد. پس از یکپارچه‌سازی، و بسته به الزامات حفاظت از داده، ممکن است داده‌ها ناشناس شوند، به عبارت دیگر، از شناسه‌هایی مانند نام و آدرس فاقد شوند، تا محرمانه بودن حفظ شود.

۲,۵ طبقه‌بندی و کُدگذاری - این فرآیند فرعی، داده ورودی را طبقه‌بندی و کُدگذاری می‌کند. به عنوان مثال، روال‌های برنامه‌نویسی اتومات یا دفتری ممکن است کُدهای عددی را به پاسخ‌های متنی طبق یک طرح طبقه‌بندی از پیش تعیین شده اختصاص دهند.

۳,۵ مرور، تأیید و ویرایش - این فرآیند فرعی در مورد ریز داده‌های جمع‌آوری شده اعمال می‌شود. برای

1. Cleaning of data records
2. Further treatment of statistical aggregates



شناسایی و در صورت لزوم، رفع مشکلات احتمالی، خطاها، و مغایرت‌هایی مانند موارد پرت، پاسخ داده نشده و کُدگذاری نادرست، به هر ثبتي نگاه می‌کند. همچنین به‌عنوان اعتبارسنجی داده‌های ورودی نیز شناخته می‌شود. ممکن است به‌صورت تکراری اجرا شود، و داده‌ها را براساس قوانین ویرایش از پیش تعیین شده، معمولاً به‌ترتیب مشخص، اعتبارسنجی کند. ممکن است ویرایش اتومات اعمال کند، یا هشدارهایی برای بازرسی دستی و تصحیح داده‌ها ایجاد کند. بازبینی، اعتبارسنجی، و ویرایش می‌تواند در مورد ثبت‌های واحد از پیمایش و منابع اداری، قبل و بعد از یکپارچه‌سازی اعمال شود. در موارد خاص، جانپي (فرآیند فرعی ۴,۵) ممکن است به‌عنوان نوعی ویرایش مورد استفاده قرار گیرد.

۴,۵ جانپي – در مواردی که داده‌ها مفقود یا غیرقابل اعتماد هستند، ممکن است اغلب با استفاده از رویکرد مبتنی بر قانون، برآوردها جانپي شوند (جایگزین کردن داده‌های گمشده با مقادیر جایگزین). معمولاً مراحل مشخص شامل موارد زیر است:

- شناسایی خطاها و شکاف‌های (داده‌هایی که در دسترس نیستند) احتمالی؛
- انتخاب داده‌ها برای شامل شدن یا حذف شدن از روال‌های جانپي؛
- جانپي با استفاده از یک یا چند روش از پیش تعریف شده مانند "بی‌درنگ" یا "بادرنگ"؛
- نوشتن بازگشت داده‌های جانپي شده به مجموعه داده و علامت‌گذاری آنها به‌عنوان جانپي شده؛ و
- تولید فراداده در فرآیند جانپي.


۵,۵ استخراج متغیرها و واحدهای آماری جدید – این فرآیند فرعی برای متغیرها و واحدهای آماری که به‌طور مشخص در مجموعه ارائه نشده‌اند، اما برای ارائه خروجی‌های مورد نیاز، ضروری هستند، مقادیر به‌دست می‌آورد. متغیرهای جدید را با استفاده از فرمول‌های حسابی^۳ برای یک یا چند متغیر موجود در مجموعه داده به‌دست می‌آورد. ممکن است نیاز به تکرار داشته باشد، زیرا برخی از متغیرهای مشتق شده ممکن است خود براساس سایر متغیرهای مشتق شده باشند. بنابراین مهم است که اطمینان حاصل شود که متغیرها به‌ترتیب صحیح مشتق شده‌اند. واحدهای آماری جدید ممکن است با تجمیع یا تقسیم داده‌ها برای واحدهای جمع‌آوری، یا با روش‌های مختلف

1. Hot-deck
2. Cold-deck
3. Arithmetic formulae



تخمین دیگر، به‌دست آیند. به‌عنوان مثال، خانوارهای مشتق شده‌ای^۱ که واحدهای جمع‌آوری اشخاص هستند، یا شرکت‌هایی که واحدهای جمع‌آوری واحدهای قانونی هستند.

۶,۵ محاسبه وزن‌ها – این فرآیند فرعی وزن‌هایی را برای ثبت‌های داده واحد^۲ با توجه به روش موجود در فرآیند فرعی ۵,۲ ایجاد می‌کند (طراحی روش پردازش آماری). این وزن‌ها را می‌توان برای "محاسبه ناخالص"^۳ نتایج نمونه پیمایش برای این که آنها را نماینده جمعیت مورد نظر قرار دهد، یا تعدیل عدم پاسخگویی در کل سرشماری‌ها استفاده کرد. برای اطلاعات و موارد مربوط به کالیبراسیون نمونه، به صفحه پروژه مربوطه مراجعه کنید.



ReGenesees Package
Diego Zardetto, Istat -

ReGenesees (R Evolved Generalized Software for Sampling Estimates and Errors in Surveys) is a full-fledged R (open source) software developed and disseminated by the Italian Statistics Office. ReGenesees is a tool for design-based and model-assisted analysis of complex sample surveys. The package (and its graphical user interface package ReGenesees.GUI) run under Windows, Mac OS, Linux and most Unix-like operating systems.

[Download](#)

۷,۵ محاسبه حاصل جمع (مجموعه‌ها)^۴ – این فرآیند فرعی، مجموعه داده^۵ و مجموع جمعیت^۶ را از ریزداده ایجاد می‌کند. شامل جمع‌بندی داده برای ثبت‌های به اشتراک گذاشته شده دارای ویژگی‌های خاص، تعیین معیارهای میانگین و پراکندگی، و اعمال وزن از فرآیند فرعی ۶,۵ برای نمونه‌گیری از داده پیمایش برای به‌دست آوردن مجموع جمعیت است.

۸,۵ نهایی کردن فایل‌های داده – این فرآیند فرعی، نتایج فرآیندهای فرعی دیگر در این فاز را برای تهیه یک فایل داده، معمولاً از کلان داده^۷، که به‌عنوان ورودی فاز ۶ (تجزیه و تحلیل) استفاده می‌شود، گردآوری می‌کند. گاهی اوقات می‌تواند یک فایل میانی باشد نه یک فایل نهایی، به‌ویژه برای فرآیندهای تجاری که فشارهای زمانی شدید و الزامی برای تولید برآوردهای اولیه و نهایی وجود دارد.

منابع مرتبط به اشتراک گذاشته شده است:

1. Deriving households
2. Unit data records
3. Gross-up
4. Aggregates
5. Aggregate data
6. Population totals
7. Macrodata



Related Resources

Documents

Anonymisation: managing data protection risk code of practice	Download
Comparative Assessment of Software Programs for the Development of Computer-Assisted Personal Interview (CAPI) Applications	Download
National Statistics Code of Practice - Protocol on Data Matching (UK)	Download
Statistical Data Editing - Volume No. 3 - Impact on Data Quality	Download
Survey Methods and Practices	Download
Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices	Download

Tools

Census and Survey Processing System (CSPro)	Download
ReGenesees Package	Download

۶. تجزیه و تحلیل

در این فاز، آمار تولید می‌شود، با جزئیات مورد بررسی قرار می‌گیرد، و برای انتشار آماده می‌شود. فاز تجزیه و تحلیل شامل ۵ فرآیند فرعی است:

۱,۶ تهیه پیش‌نویس خروجی‌ها - در این فرآیند فرعی، داده‌های جمع‌آوری شده به خروجی‌های آماری تبدیل می‌شوند؛ سنجش‌های بیشتر مانند شاخص‌ها، روندها، یا سری‌های تعدیل‌شده فصلی تولید می‌شوند؛ و ویژگی‌های کیفی ثبت می‌شوند.

۲,۶ اعتبار بخشیدن خروجی‌ها - در این فرآیند فرعی، متخصصین آمار، کیفیت خروجی‌های تولید شده را مطابق با چارچوب کلی کیفیت و انتظارات، اعتبار می‌بخشند. این فرآیند فرعی همچنین شامل فعالیت‌های مرتبط با جمع‌آوری اطلاعات و اثر تجمعی ایجاد مجموعه دانش پیرامون حوزه آماری خاص است. این دانش سپس برای جمع‌آوری داده‌های فعلی در محیط فعلی به کار گرفته می‌شود، تا هرگونه فاصله از انتظارات را شناسایی کرده و تجزیه و تحلیل آگاهانه را مجاز سازد. فعالیت‌های اعتبارسنجی می‌تواند شامل موارد زیر باشد:

- بررسی پوشش جمعیت لازم و میزان پاسخگویی؛
- مقایسه آمار با چرخه‌های قبلی (در صورت وجود)؛
- مقایسه آمار با سایر داده‌های مرتبط (داخلی و خارجی)؛



- بررسی ناهماهنگی در آمار؛
- انجام ویرایش کلان؛ و
- اعتبارسنجی آمارها در برابر انتظارات و حوزه اطلاعات.

۳,۶ موشکافی کردن و توضیح دادن - در این فرآیند فرعی، متخصصین آمار، درک عمیقی از خروجی‌ها به‌دست می‌آورند. آنها از این درک، جهت بررسی دقیق و توضیح آمار تولید شده برای این چرخه استفاده می‌کنند با ارزیابی این‌که با چه کیفیتی، آمار، انتظارات اولیه آنها را منعکس می‌کند، با مشاهده آمار از همه‌منظر با استفاده از ابزارها و رسانه‌های مختلف، و با انجام تجزیه‌وتحلیل آماری عمیق.

۴,۶ اعمال کنترل افساء - این فرآیند فرعی تضمین می‌کند که داده‌ها و فراداده‌هایی که باید منتشر شوند قوانین محرمانه را نقض نمی‌کنند که می‌تواند شامل بررسی افشای اولیه و ثانویه، همچنین اعمال کوچک‌سازی داده‌ها یا تکنیک‌های اختلال باشد.

۵,۶ نهایی‌سازی خروجی‌ها - این فرآیند فرعی تضمین می‌کند که آمار و اطلاعات مربوط، به هدف مورد نظر آنها مناسب بوده و سطح کیفی لازم را کسب کرده‌اند، و بنابراین آماده استفاده هستند. شامل:

- تکمیل بررسی‌های یکپارچگی؛
- تعیین سطح انتشار و اعمال هشدارها؛
- جمع‌آوری اطلاعات حمایتی، شامل تفسیر، توضیحات، اندازه‌گیری عدم قطعیت، و سایر فراداده‌های ضروری؛
- تهیه اسناد داخلی پشتیبان؛
- انجام بحث با کارشناسان موضوعات داخلی قبل از انتشار؛ و
- تأیید محتوای آماری برای انتشار.

منابع مرتبط به اشتراک گذاشته شده است:




Related Resources

Tools

 ADePT

 Download

Links

 Boston College Department of Economics - Statistical Software Components

 Download

Documents

 Household Sample Surveys in Developing and Transition Countries

 Download

 The Analysis of Household Surveys: A Microeconomic Approach to Development Policy

 Download

۷. انتشار نتایج پیمایش

اطلاعات موجود در این بخش کلی است و نمای کلی از اهمیت بایگانی را ارائه می‌دهد. برای اطلاعات بیشتر و دقیق‌تر در مورد بایگانی داده‌ها و فراداده‌ها، به دستورالعمل‌های انتشار ریزداده‌ها مراجعه کنید.

این فاز انتشار محصولات آماری را برای مشتریان مدیریت می‌کند، و شامل ۵ فرآیند فرعی است:

۱,۷ به‌روز رسانی سیستم‌های خروجی – این فرآیند فرعی، به‌روزرسانی سیستم‌هایی را که در آن داده‌ها و فراداده‌ها برای اهداف انتشار ذخیره می‌شوند، مدیریت می‌کند. قالب‌بندی، بارگذاری، و پیوند فراداده در فازهای اولیه انجام می‌شود؛ این فرآیند فرعی شامل تأیید وجود همه فراداده‌های لازم و آماده انتشار است. شامل:

- قالب‌بندی داده‌ها و فراداده‌هایی که آماده اضافه شدن به پایگاه‌های داده خروجی هستند؛
- بارگذاری داده‌ها و فراداده‌ها در پایگاه‌های داده خروجی؛ و
- اطمینان از این‌که داده‌ها به فراداده مربوطه پیوند می‌شوند.

۲,۷ تولید محصولات انتشار – این فرآیند فرعی محصولاتی را تولید می‌کند که در فرآیند فرعی ۱,۲ برای رفع نیازهای کاربر طراحی شده‌اند. محصولات می‌توانند اشکال مختلفی داشته باشند، از جمله نشریات چاپی، بیانیه‌های مطبوعاتی، و وبسایت‌ها. مراحل شامل:



- تهیه اجزای محصول، مانند متن توضیحی، جدول‌ها و نمودارها؛
- مونتاژ اجزا به محصولات؛ و
- ویرایش محصولات و بررسی این‌که آنها مطابق با استانداردهای انتشار هستند.

۳,۷ مدیریت انتشار محصولات انتشار - این فرآیند فرعی تضمین می‌کند که همه عناصر برای انتشار در جای خود قرار دارند و زمان انتشار را مدیریت می‌کنند. این گزارش به گروه‌های خاصی مانند مطبوعات یا وزرا می‌پردازد؛ محدودیت‌های قبل از انتشار را ترتیب می‌دهد؛ و شامل ارائه محصولات به مشترکین می‌شود.

۴,۷ پشتیبانی محصولات انتشار^۱ - اگرچه بازاریابی را می‌توان یک فرآیند فراگیر دانست، این فرآیند فرعی به‌طور فعال محصولات آماری تولید شده در یک فرآیند تجاری آماری خاص را حمایت می‌کند تا به آنها در دستیابی به گسترده‌ترین مخاطبان ممکن کمک کند. از ابزارهای مدیریت ارتباط با مشتری برای هدف‌گیری بهتر کاربران بالقوه محصول استفاده می‌کند، و از وب‌سایت‌ها، ویکی‌ها و وبلاگ‌ها برای تسهیل ارتباط اطلاعات آماری با کاربران استفاده می‌کند.

۵,۷ مدیریت پشتیبانی کاربر - این فرآیند فرعی تضمین می‌کند که پرسش‌های مشتری ثبت، و پاسخ‌ها در پارامترهای مهلت ارائه می‌شوند. این پرسش‌ها باید به‌طور مرتب مورد بازبینی قرار گیرند تا بتوانند اطلاعاتی را در فرآیند مدیریت کیفیت ارائه دهند، زیرا می‌توانند نیازهای کاربر جدید یا در حال تغییر را نشان دهند.



Related Resources

منابع مرتبط به اشتراک گذاشته شده است:

Documents

- 📄 Diffusion des fichiers de microdonnées - Principes, procédures et pratiques 📄 (1.42 MB)
- 📄 National Statistics Code of Practice - Protocol on Customer Service and User Consultation (UK) 📄 Download
- 📄 National Statistics Code of Practice - Protocol on Data Presentation, Dissemination and Pricing (UK) 📄 Download
- 📄 National Statistics Code of Practice - Protocol on Release Practices (UK) 📄 Download
- 📄 Survey Methods and Practices 📄 Download
- 📄 The Canadian Data Liberation Initiative: An Idea worth Considering? 📄 (683.32 KB)

Links

- 📄 Guidelines on the application of new information technology to population data dissemination (ESCAP) 📄 Download



۸. بایگانی

اطلاعات عمومی زیر یک نمای کلی از اهمیت بایگانی ارائه می‌دهد؛ برای اطلاعات بیشتر و دقیق‌تر در مورد بایگانی داده‌ها و فراداده‌ها، به دستورالعمل‌های بایگانی داده‌ها مراجعه کنید.

این فاز بایگانی داده‌های آماری و فراداده را مدیریت می‌کند و از ۴ فرآیند فرعی تشکیل شده است:

۱,۸ تعیین قوانین بایگانی – قوانین بایگانی داده‌های آماری و فراداده حاصل از یک فرآیند تجاری آماری، توسط این فرآیند فرعی تعیین می‌شود. الزام بایگانی خروجی‌های میانی، مانند فایل نمونه، داده‌های خام از فاز جمع‌آوری، و نتایج مراحل مختلف فرآیند، و تجزیه و تحلیل فازها نیز باید مورد توجه قرار گیرند. قوانین بایگانی برای یک فرآیند تجاری خاص آماری ممکن است به‌طور کامل یا تا حدی به سیاست بایگانی عمومی سازمان آماری یا، برای سازمان‌های ملی، به استانداردهای اعمال شده در بخش دولتی وابسته باشند. قوانین باید شامل در نظر گرفتن واسطه انتقال، محل بایگانی و الزامات نگهداری نسخه‌های تکراری باشد. همچنین باید شرایط (در صورت وجود) را در نظر بگیرند که تحت آن داده‌ها و فراداده‌ها باید کنار گذاشته شوند. حداقل برای اولین بار از یک فرآیند تجاری آماری، این فرآیند فرعی منطقی و قوی با فاز ۲ – طراحی، پیوند می‌شود.

۲,۸ مدیریت مخزن بایگانی – این فرآیند فرعی یک یا چند مخزن بایگانی را مدیریت می‌کند، از جمله، پایگاه‌های داده یا مکان‌های فیزیکی که کپی‌های داده یا فراداده در آن ذخیره می‌شوند. شامل:

- حفظ کاتالوگ‌های داده و بایگانی‌های فراداده، با اطلاعات کافی برای اطمینان دادن از این که داده‌های فردی یا مجموعه‌های فراداده به راحتی قابل بازیابی هستند؛
- آزمایش فرآیندهای بازیابی؛
- بررسی دوره‌ای یکپارچگی داده‌ها و فراداده‌های بایگانی شده؛ و
- ارتقاء فرمت‌های بایگانی ویژه نرم‌افزار هنگام تغییر نرم‌افزار.

این فرآیند فرعی ممکن است بسته به میزان استانداردسازی در سازمان، یک فرآیند تجاری آماری خاص یا گروهی از فرآیندها را پوشش دهد. در نهایت، اگر استانداردهای وسیع سازمانی وضع شوند، ممکن است فرآیندی فراگیر تلقی شود.



۳,۸ حفظ داده‌ها و فراداده‌های مرتبط - در این فرآیند فرعی، داده‌ها و فراداده‌های یک فرآیند تجاری آماری خاص بایگانی می‌شوند. شامل:

- شناسایی داده و فراداده برای بایگانی، مطابق با قوانین تعریف شده در ۱,۸؛
- قالب‌بندی آن داده‌ها و فراداده‌ها برای مخزن؛
- بارگذاری یا انتقال داده‌ها و فراداده‌ها به مخزن؛
- کاتالوگ کردن داده‌ها و فراداده‌های بایگانی شده؛ و
- بررسی این‌که داده‌ها و فراداده‌ها با موفقیت بایگانی شده‌اند.

۴,۸ کنارگذاشتن داده‌ها و فراداده‌های مرتبط - در این فرآیند فرعی، داده‌ها و فراداده‌های یک فرآیند تجاری آماری خاص کنار گذاشته می‌شوند. شامل:

- شناسایی داده و فراداده برای رد کردن، مطابق با قوانین تعریف شده در ۱,۸؛
- رد کردن آن داده‌ها و فراداده‌ها؛ و
- ثبت رد آن داده‌ها و فراداده‌ها.

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

 National Statistics Code of Practice - Protocol on Data Management, Documentation and Preservation (UK)	 Download
 Principles and Good Practice for Preserving Data	 (1.31 MB)



۹. ارزیابی پیمایش

این فاز، ارزیابی یک نمونه خاص از یک فرآیند تجاری آماری را برخلاف یک فرآیند فراگیرتر مدیریت کیفیت آماری توصیف شده در بخش ۶، مدیریت می‌کند. منطقاً، در پایان فرآیند اتفاق می‌افتد، اما متکی بر ورودی‌هایی است که در مراحل مختلف جمع‌آوری شدند. این فاز شامل ۳ فرآیند فرعی است:

۱,۹ جمع‌آوری ورودی‌های ارزیابی – مواد ارزیابی را می‌توان در هر فاز یا فرآیند فرعی تولید کرد. اشکال مختلفی می‌تواند داشته باشد، از جمله، بازخورد کاربران، فرآیند فراداده، سیستم معیارها، و پیشنهادات کارکنان. گزارشات پیشرفت برنامه عملیاتی که طی یک تکرار قبلی ایجاد شده است نیز ممکن است ورودی برای ارزیابی‌های تکرارهای بعدی باشد. این فرآیند فرعی همه این ورودی‌ها را جمع‌آوری کرده و در اختیار فرد یا تیم ارزیابی قرار می‌دهد.


۲,۹ انجام ارزیابی – این فرآیند فرعی، ورودی‌های ارزیابی را تجزیه و تحلیل کرده و آنها را در یک گزارش ارزیابی که هرگونه مسائل کیفی مربوط به این تکرار فرآیند تجاری آماری را ذکر می‌کند، ترکیب می‌کند و در صورت لزوم توصیه‌هایی برای تغییرات ارائه می‌دهد. توصیه‌ها می‌توانند تغییرات هر فاز یا فرآیند فرعی را برای تکرارهای بعدی فرآیند مورد بررسی قرار دهند، یا می‌توانند نشان دهند که این روند، تکرار نمی‌شود.

۳,۹ موافقت یک برنامه عملیاتی – این فرآیند فرعی، قدرت تصمیم‌گیری لازم را برای تدوین یک برنامه عملیاتی براساس گزارش ارزیابی، گردآوری می‌کند. باید شامل در نظر گرفتن سازوکاری برای نظارت بر تأثیر آن عملیات باشد، که ممکن است به نوبه خود، ارزیابی‌های تکرارهای بعدی فرآیند را تحت تأثیر قرار دهد.

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

 DRAFT - Survey Quality Assessment Framework (SQAF)

 (1.23 MB)



۱۰. مدیریت کیفیت

این فرآیند در سراسر مُدل وجود دارد. با فاز ۹ (ارزیابی پیمایش)، که نمونه‌های فردی یک فرآیند تجاری آماری را ارزیابی می‌کند، ارتباط نزدیکی پیدا می‌کند. با این حال، فرآیند مدیریت کیفیت گسترده‌تر، هم دامنه عمیق‌تر دارد و هم گسترده‌تر است. علاوه بر تکرارهای ارزیابی یک فرآیند، ارزیابی فازها و فرآیندهای فرعی جداگانه نیز، در حالت ایده‌آل، هر بار که مورد استفاده قرار می‌گیرند، ولی حداقل براساس برنامه توافق شده، ضروری است. فراداده‌های تولید شده توسط فرآیندهای فرعی مختلف نیز به‌عنوان ورودی فرآیند مدیریت کیفیت مورد توجه قرار می‌گیرند. این ارزیابی‌ها می‌توانند در یک فرآیند خاص یا در چندین فرآیند که از اجزای مشترک استفاده می‌کنند، اعمال شوند.

مدیریت کیفیت همچنین شامل ارزیابی گروه‌هایی از فرآیندهای تجاری آماری است، و بنابراین می‌تواند تکرارها یا شکاف‌های احتمالی را شناسایی کند. همه ارزیابی‌ها منجر به بازخورد می‌شوند، که باید برای بهبود فرآیند، فاز، یا فرآیند فرعی مربوطه و ایجاد یک حلقه کیفیت، مورد استفاده قرار گیرند.

مدیریت کیفیت می‌تواند انواع مختلفی داشته باشد، از جمله:

- جستجو و تجزیه و تحلیل بازخورد کاربر؛
- مرور عملیات و ثبت درس‌های آموخته شده؛
- بررسی فرآیند فراداده و دیگر معیارهای سیستم؛ و
- محک زدن یا بررسی فرآیندهای همکار با سایر سازمان‌ها.

ارزیابی به‌طور معمول در چارچوب کیفیت خاص یک سازمان صورت می‌گیرد، و بنابراین ممکن است اشکال متفاوتی داشته باشد و نتایج متفاوتی را در سازمان‌های مختلف ارائه دهد. با این حال، بین سازمان‌های آماری، توافق کلی وجود دارد که کیفیت باید مطابق استاندارد ISO 9000-2005 تعریف شود: "میزانی که مجموعه‌ای از ویژگی‌های ذاتی الزامات را برآورده می‌کند".

کیفیت، یک مفهوم چندوجهی^۱ و کاربر محور است. ابعاد کیفیت که مهمترین آنها در نظر گرفته می‌شوند به دیدگاه‌ها، نیازها و اولویت‌های کاربر بستگی دارد که بین فرآیندها و گروه‌های مختلف کاربران متفاوت است. چندین سازمان آماری، فهرست‌هایی از ابعاد کیفیت تهیه کرده‌اند، که برای سازمان‌های بین‌المللی، تحت رهبری کمیته



هماهنگی فعالیت‌های آماری^۱ هماهنگ می‌شوند.

تعدد چارچوب‌های کیفیت، اهمیت ارزیابی روش‌های محک‌زنی و بررسی دقیق را افزایش می‌دهند، و در حالی که بعید است این رویکردها برای تکرار هر بخش از هر فرآیند تجاری آماری امکان‌پذیر باشند، باید طبق روشی سیستماتیک، مطابق برنامه از پیش تعیین شده‌ای مورد استفاده قرار گیرند که امکان بررسی همه قسمت‌های اصلی فرآیند را در یک بازه زمانی مشخص فراهم کند.

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

 DRAFT - Survey Quality Assessment Framework (SQAF)	 (1.23 MB)
 National Statistics Code of Practice - Protocol on Quality Management (UK)	 Download
 Statistical Data Editing - Volume No. 3 - Impact on Data Quality	 Download
 Statistics Canada Quality Guidelines (Fifth Edition, October 2009)	 Download
 Statistics Canada's Quality Assurance Framework 2002	 Download
 Survey Methods and Practices	 Download

Links

 National Quality Assurance Frameworks	 Download
---	--

1. Committee for the Coordination of Statistical Activities (CCSA)



۱۱. مدیریت فراداده

اطلاعات عمومی زیر از مدل فرایند تجاری آماری عمومی گرفته شده است. دستورالعمل‌ها، ابزارها، و استانداردهای دقیق‌تر و کاربردی‌تر در مورد مدیریت فراداده در بخش‌های مستندسازی و نرم‌افزار در ادامه، ارائه شده است.

مدیریت فراداده خوب برای عملکرد کارآمد فرآیندهای تجاری آماری ضروری است. فراداده‌ها در هر فاز وجود دارند. چالش اصلی این است که اطمینان حاصل شود که این فراداده‌ها در اسرع وقت جمع‌آوری شده، و با داده‌های مرتبط با آنها از هر فاز به فاز دیگر ذخیره و منتقل می‌شوند. استراتژی و سیستم‌های مدیریت فراداده حیاتی هستند.

بخش الف چارچوب مشترک فراداده، ۱۶ اصل اساسی زیر را برای مدیریت فراداده شناسایی می‌کند، که همه آنها در فرآیند مدیریت فراداده مورد بررسی قرار گرفته و زمان تهیه چشم‌انداز سیستم فراداده^۱ و معماری سراسری، و هنگام اجرای سیستم فراداده به آنها اشاره می‌شود. این اصول را می‌توان به گروه‌های زیر طبقه‌بندی کرد:

- مدیریت فراداده^۲؛
- مرجع فراداده^۳؛
- ارتباط با چرخه آماری/فرآیندها^۴؛ و
- کاربران.

-
1. Statistical metadata system (SMS)
 2. Metadata handling
 3. Metadata Authority
 4. Relationship to Statistical Cycle / Processes



مدیریت فراداده

۱. **مدل فرآیند تجاری آماری:** فراداده‌ها را با تمرکز بر مُدل کلی فرآیند تجاری آماری مدیریت کنید؛
۲. **فعال، نه منفعل:** فراداده‌ها را تا حد ممکن فعال کنید تا از صحت و به‌روز بودن آنها اطمینان حاصل کنید. فراداده‌های فعال، فراداده‌هایی هستند که سایر فرآیندها و عملیات دیگر را هدایت می‌کنند؛
۳. **استفاده مجدد:** در صورت امکان از فراداده‌ها برای یکپارچگی آماری و همچنین کارآیی، استفاده مجدد کنید؛
۴. **نسخه‌ها:** با حفظ نسخه‌های قدیمی، تاریخچه فراداده را حفظ کنید.

مرجع فراداده

۱. **ثبت:** اطمینان حاصل کنید که ثبت فرآیند/جریان کار مرتبط با هر عنصر فراداده به خوبی مستند می‌شود، بنابراین شناسایی واضحی از مالکیت، وضعیت تأیید، تاریخ بهره‌برداری و غیره وجود دارد؛
۲. **منبع واحد:** اطمینان حاصل کنید که برای هر عنصر فراداده، یک منبع معتبر و واحد، به‌عبارت دیگر، "مرجع ثبت" وجود دارد؛
۳. **یک ورودی/به‌روز رسانی:** با یک‌بار وارد کردن و به‌روز رسانی در یک مکان، خطاها را به حداقل برسانید؛
۴. **تغییرات استانداردها:** اطمینان حاصل کنید که تغییرات استانداردها به دقت مدیریت، تأیید و مستند می‌شوند و قابل مشاهده هستند.

ارتباط با چرخه آماری / فرآیندها

۱. **یکپارچگی:** کار مرتبط با فراداده را به یک بخش جدایی‌ناپذیر فرآیندهای تجاری در سراسر سازمان تبدیل کنید؛
۲. **تطابق فراداده:** اطمینان حاصل کنید که فراداده ارائه شده به کاربرانِ نهایی با فراداده‌ای که فرآیند تجاری را هدایت کرده یا در طول فرآیند ایجاد شده است، مطابقت دارد؛
۳. **شرح جریان:** جریان فراداده را با فرآیندهای آماری و تجاری، جریان داده‌ها، و منطق تجاری شرح دهید؛



۴. **ثبت در منبع:** فراداده‌ها را در منبع خود، ترجیحاً به صورت اتومات، به عنوان محصول جانبی سایر فرآیندها ثبت کنید؛

۵. **مبادله و استفاده:** فراداده‌ها را مبادله کرده و از آنها برای اطلاع‌رسانی فرآیندهای رایانه‌ای و تفسیر انسانی استفاده کنید. زیرساخت مبادله داده‌ها و فراداده‌های مرتبط باید براساس مؤلفه‌های جفت شده آزاد، با انتخاب زبان‌های مبادله استاندارد مانند XML باشد.

کاربران

۱. **شناسایی کاربران:** اطمینان حاصل کنید که کاربران به وضوح برای همه فرآیندهای فراداده شناسایی می‌شوند، و همه فراداده‌های ثبت شده برای آنها ارزش ایجاد می‌کنند؛





۲. **فرمت‌های مختلف:** فراداده متنوع است. نماهای مختلف مربوط به کاربردهای مختلف داده است؛ کاربران به سطوح مختلفی از جزئیات نیاز دارند؛ و فراداده‌ها بسته به فرآیندها و اهدافی که برای آنها تولید و استفاده می‌شوند، در قالب‌های مختلفی ظاهر می‌شوند؛

۳. **دسترسی:** اطمینان حاصل کنید که فراداده‌ها به آسانی در دسترس بوده و در زمینه نیازهای اطلاعاتی کاربران داخلی یا خارجی قابل استفاده هستند.





منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Links

- | | |
|---|--|
|  Metadata and the Statistical Business Process (Common metadata Framework, Part C) |  Download |
|  Metadata Concepts, Standards, Models and Registries (Common metadata Framework, Part B) |  Download |

Documents

- | | |
|---|--|
|  National Statistics Code of Practice - Protocol on Data Management, Documentation and Preservation (UK) |  Download |
|  Statistical Metadata in a Corporate Context: a guide for managers (Common metadata Framework, Part A) |  Download |



ابزارهای یکپارچه‌سازی پیمایش

ابزارهای یکپارچه‌سازی پیمایش به معنای منسجم و یکپارچه کردن آنها با یکدیگر و با سایر ابزارهای جمع‌آوری داده است که توسط سیستم آماری یک کشور مورد استفاده قرار می‌گیرد و شامل استفاده از سؤال‌ها، مفاهیم، طبقه‌بندی‌های قابل مقایسه و غیره است. اهداف یکپارچه‌سازی شامل ایجاد یک مجموعه منسجم داده، قابل مقایسه در طول زمان و در بین منابع در صورت مرتبط بودن، و اطمینان از هزینه اثربخشی تولید داده است.

چالش‌ها بسته به این که آیا کشوری دارای سابقه جمع‌آوری داده‌های اجتماعی از طریق سوابق اداری، سرشماری‌ها یا پیمایش‌های خانوار است، و این که آیا نسبتاً تازه، وارد این فعالیت‌ها شده است، متفاوت خواهد بود. در مورد اشاره شده اول، یعنی به تازگی وارد شدن به این مهم، آژانس ممکن است با چالش‌های یکپارچگی و هماهنگی روبه‌رو شود و در مورد اشاره شده دوم، یعنی داشتن سابقه، آژانس ممکن است کار را راحت‌تر ببیند، اگرچه عدم تجربه استفاده از پیمایش‌های خانوار به خودی خود یک چالش خواهد بود. با این وجود، هر دو نوع آژانس، برای توسعه پایگاه‌های داده متا جهت کمک به جمع‌آوری و انتشار داده‌ها، برای ایجاد استانداردهایی که در هر درجه ممکن با استانداردهای بین‌المللی مطابقت داشته باشند، برای یکپارچه کردن داده‌ها از منابع مختلف جهت اطمینان از گزارش متوازن و جامع و برای ارائه اطلاعات به جای داده‌های اولیه به چالش کشیده خواهند شد.



The Struggle for Integration and Harmonization of Social Statistics in a Statistical Agency - A Case Study of Statistics Canada

Gordon Priest - June 2010

IHSN Working Paper No 004. This paper describes how external influences and a corporate strategy lead to partial coordination of surveys and data collection practices at Statistics Canada.

[Download](#)

برخی از کلیدهای موفقیت در رویارویی با این چالش‌ها موارد زیر خواهد بود:

- **مراکز ملی آمار باید به دقت استقلال سیاسی خود را محافظت کنند.** این می‌تواند یک چالش باشد، اما اگر قرار است، خروجی‌های آماری دارای اعتبار بوده و در نتیجه مطلوبیت داشته باشند، آژانس‌ها باید اطمینان حاصل کنند که آنها عاری از مداخله سیاسی هستند. یکی



از بهترین راه‌ها برای اطمینان از استقلال آژانس‌های آماری، ایجاد مشارکت قوی با متقاضیان عمده، از جمله آژانس‌های بین‌المللی، ادارات دولتی، دانشگاهیان، سازمان‌های غیردولتی، مشاغل و رسانه‌ها است.

- **روابط با رسانه‌ها از اهمیت ویژه‌ای برخوردار است**، نه تنها از منظری که از گوش دادن به رسانه‌ها و برداشت‌های عمومی از راه‌های تحقیق مورد نیاز به دست می‌آید، بلکه از فرصتی برای ارائه نه تنها داده‌ها بلکه اطلاعات، به عبارت دیگر، خروجی‌های تحلیلی، به شیوه روزنامه‌نگاری نوشته شده که به راحتی قابل دستیابی است. اطلاعاتی که عموم می‌توانند با آنها ارتباط برقرار کنند، مانند مشخصات مناطق جغرافیایی یا گروه‌های جمعیتی خاص، مفید هستند. به عنوان مثال، مشخصات خانواده‌های تک سرپرست، سالمندان، کودکان و جوانان یا کارگران کشاورزی می‌تواند به ایجاد جمعیت آگاه کمک کند و اطلاعات مهمی را در اختیار تصمیم‌گیرندگان قرار دهد.
- بسیار مهم است که **داده‌های منابع مختلف در داخل آژانس یکدست باشند**. مفاهیم، تعاریف، قراردادهای نام‌گذاری، و طبقه‌بندی‌ها باید از یک منبع به منبع دیگر یکدست باشند. بنابراین، آژانس‌ها، نباید زمانی را برای تدوین سیاست استانداردهایی که تا حد امکان استانداردهای بین‌المللی را در بر می‌گیرد، تلف کنند. این سیاست باید تضمین کند که چارچوب‌های مفهومی برای تجمیع اطلاعات آماری استفاده می‌شوند؛ نام‌ها و تعاریف استاندارد برای گروه‌های جمعیت، واحدهای آماری، مفاهیم، متغیرها و طبقه‌بندی‌ها؛ و روش‌های منسجم جمع‌آوری و پردازش استفاده می‌شوند. سیاست باید در بالاترین سطح آژانس کنترل و اجرا شود.
- **یک پایگاه داده متا برای پشتیبانی از خط مشی باید ایجاد شود**. وظیفه اصلی مراکز ملی آمار، تولید و انتشار داده‌های آماری از شرایط اقتصادی و اجتماعی در آن کشور است. داده‌های آماری به صورت اعداد در فایل‌های داده، جدول‌های آماری، یا متونی مانند خبرها و مقالات به کار می‌روند. این اعداد به تنهایی قابل درک نیستند. اطلاعات توضیحی، فراداده نامیده می‌شوند، و برای درک و تفسیر داده‌های آماری ضروری هستند.

فراداده، حداقل، باید توصیف داده‌ها را پوشش دهد. استانداردی که توسط اداره آمار کانادا برای ساختاردهی و ارائه

1. Major clients
2. Single-parent families



این نوع فراداده استفاده می‌شود، ISO/IEC 11179 "فناوری اطلاعات - مشخصات و استانداردهای عناصر داده" است. در اصطلاحات آماری، عناصر داده معمولاً متغیر نامیده می‌شوند؛ بنابراین، این استاندارد، راهنمایی برای ساختاردهی و ارائه اطلاعات پایه در مورد متغیرها ارائه می‌دهد. با این حال، فرآیند ایجاد اطلاعات مطابق این استاندارد، در مفهوم‌سازی، نام‌گذاری، و سازماندهی متغیرهایی که داده‌ها برای آنها تولید می‌شود، سازگاری و دقت بیشتری را به همراه دارد. تجربه نشان می‌دهد که فراداده‌ها به بهترین وجه در ابتدای هر طرح پیمایش یا طراحی مجدد مستند می‌شوند، تا بعد از آن. هنگامی که آنها بخشی از فرآیند هستند، در واقع به آن کمک می‌کنند.

فراداده‌ها از سه فعالیت آماری ضروری پشتیبانی می‌کنند. اول، از طراحی و توسعه پیمایش‌های جدید یا توسعه مجدد پیمایش‌های موجود پشتیبانی می‌کنند و یک گزارش فوری از آنچه قبلاً انجام شده و ممکن است در آینده، مجدد مورد استفاده قرار گیرد، ارائه می‌دهند، در نتیجه، کارایی و صرفه‌جویی قابل توجهی در هزینه‌ها را به دنبال دارند. دوم، فراداده‌ها، چارچوبی را برای توسعه و حفظ استانداردها و بهترین شیوه‌ها فراهم می‌آورند که منجر به کارایی، صرفه‌جویی در هزینه‌ها، و بهبود محصول می‌شوند. در نهایت، فراداده‌ها استخراج مؤثر منابع داده‌های یک آژانس را ترویج و تشویق می‌کنند.

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

 National Statistics Code of Practice - Protocol on Statistical Integration and Classification (UK)  Download

 The Struggle for Integration and Harmonization of Social Statistics in a Statistical Agency - A Case Study of Statistics Canada  (552.09 KB)



بایگانی کردن و انتشار داده‌ها

ریز داده‌های پیمایش و سرشماری، منابع ارزشمندی برای ادارات دولتی و محققان دانشگاهی هستند. آنها دارایی‌های ارزشمند و غیرقابل جایگزینی هستند و باید به گونه‌ای مدیریت شوند که به وسیع‌ترین استفاده ممکن و استفاده مجدد از آنها تشویق کنند. در عین حال، حفاظت از حریم خصوصی پاسخ‌دهندگان پیمایش یکی از دغدغه‌های اصلی گردآورندگان داده^۱ است. رویه‌های قوی برای اطمینان دادن به ذینفعان ضروری است که ریز داده‌ها در حال حاضر و در آینده به شیوه‌ای بهینه به نفع کاربران و همه افراد جامعه، منتشر و استفاده می‌شوند. ایجاد و اجرای چنین رویه‌هایی به تخصص و منابع نیاز دارد. بنابراین توصیه می‌شود بایگانی‌های ریز داده‌های ملی یا بین‌المللی برای موارد زیر ایجاد شوند:

- ترویج کسب^۲، مستندسازی، انتشار و حفظ ریز داده‌ها، ضروری برای تولید آمار و تحقیقات ملی و آموزش در علوم اجتماعی؛
- ترویج استفاده مؤثر از داده‌های سرشماری و پیمایش موجود؛
- اطمینان از زنده ماندن و قابلیت استفاده ریز داده‌ها در حال حاضر و در آینده؛
- با حفظ محرمانه‌گی و پیروی از توصیه‌ها و بهترین شیوه‌های بین‌المللی، فراهم نمودن دسترسی عادلانه به این داده‌ها در چارچوب قوانین ملی به نفع همه شهروندان.

این بخش، راهنمایی و بهترین شیوه‌های مربوط به نقش‌های بایگانی داده‌ها را ارائه می‌دهد:

- ۱. مستندسازی؛
- ۲. بی‌نام‌سازی؛
- ۳. کاتالوگ کردن؛
- ۴. انتشار دادن؛
- ۵. محافظت.

همچنین برای اجرای چنین خدماتی موارد زیر را توضیح می‌دهد:

- ۶. ترتیبات سازمانی؛ و



- ۷. الزامات فنی.

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

Anonymisation: managing data protection risk code of practice	Download
Managing and sharing data - Best practice for researchers	Download
Managing Statistical Confidentiality and Microdata Access - Principles and guidelines of Good Practice	Download
National Statistics Code of Practice - Protocol on Data Management, Documentation and Preservation (UK)	Download

Links

The Data Seal of Approval (DSA)	Download
---------------------------------	----------



۱. مستندسازی ریزداده

"از دید بایگانی و کاربر نهایی، یک مجموعه داده "خوب" مجموعه‌ای است که استفاده از آن آسان است. مستندات آن واضح و قابل فهم است، دسترسی به داده‌ها هیچ چیز شگرفی نیست و کاربران می‌توانند با زمان شروع نسبتاً کمی به مجموعه داده‌ها دسترسی داشته باشند". (راهنمای آماده‌سازی و بایگانی داده‌های علوم اجتماعی، کنسرسیوم بین دانشگاهی تحقیقات سیاسی و اجتماعی (ICPSR))

مستندسازی داده‌ها یا فراداده‌ها به محقق کمک می‌کند تا:

- **داده‌های مورد علاقه خود را بیابند.** بدون نام، چکیده، کلمات کلیدی و سایر عناصر فراداده مهم، ممکن است برای یک محقق دشوار باشد که مجموعه داده‌ها و متغیرهای خاصی را پیدا کنند. هر سیستم کاتالوگ کردن و مکان‌یابی منابع - اعم از دستی یا دیجیتالی - براساس فراداده است.
- **دریابند که داده‌ها چه چیزی را اندازه‌گیری می‌کنند و چگونه داده‌ها ایجاد شده‌اند.** بدون توصیف مناسب از طرح پیمایش و روش‌های مورد استفاده در جمع‌آوری و پردازش داده‌ها، احتمال خطر سوء تفاهم و حتی سوء استفاده کاربر از آنها بالا است.
- **کیفیت داده‌ها را ارزیابی کنند.** برای دانستن این که آیا داده‌ها برای یک پروژه تحقیقاتی مفید هستند یا خیر، محققان به اطلاعاتی در مورد استانداردهای جمع‌آوری داده‌ها و همچنین هرگونه انحراف از استانداردهای برنامه‌ریزی شده نیاز دارند.
- **فراداده غنی همچنین بار تولیدکننده داده را کاهش می‌دهد،** زیرا نیاز به ارائه پشتیبانی منظم به کاربران داده‌ها را کاهش می‌دهد.

سابق بر این، تولیدکنندگان داده‌ها و بایگانی‌های داده‌ها، کتاب‌های کُد مبتنی بر متن را تولید می‌کردند. جایگزین امروزی، کتاب‌های کُد مبتنی بر XML است که مطابق با استانداردهای فراداده بین‌المللی مانند طرح مستندسازی داده‌ها (DDI) و Dublin Core تولید می‌شوند. برای تسهیل مستندسازی ریزداده‌ها، شبکه بین‌المللی پیمایش‌های خانوار، مجموعه ابزار مدیریت ریزداده‌ها را توزیع می‌کند و اتخاذ بهترین شیوه‌های بین‌المللی را ترویج می‌کند.

همان‌طور که در مُدل عمومی فرآیند کسب‌وکار آماری (GSBPM) توصیه می‌شود، فراداده‌ها باید "در زمان واقعی"



در کل چرخه عمر پیمایش جمع‌آوری شوند. مستندسازی طراحی یا اجرای یک پیمایش و داده‌های حاصله نباید به عنوان آخرین مرحله در اجرای پروژه تلقی شوند.

در این بخش، اطلاعات یا دستورالعمل‌هایی در مورد آن ارائه داده می‌شود:

- محدوده مورد انتظار مستندسازی پیمایش: چه نوع اطلاعاتی فراداده خوبی را ایجاد می‌کند؛
- استانداردها و مدل‌های فراداده: استانداردهای فراداده منتخب و مدل‌های مرتبط خاص برای مستندسازی ریزداده‌ها؛
- ابزارها: برنامه‌های کاربردی نرم‌افزار رایگان برای مستندسازی ریزداده‌های پیمایش.

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

- | | |
|--|----------|
| Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle (5th ed.) | Download |
| Managing and sharing data - Best practice for researchers | Download |
| National Statistics Code of Practice - Protocol on Data Management, Documentation and Preservation (UK) | Download |
| Quick Reference Guide for Data Archivists | Download |

Tools

- | | |
|--|----------|
| Nesstar Publisher (IHSN Metadata Editor) | Download |
|--|----------|

Links

- | | |
|---------------------------------|----------|
| The Data Seal of Approval (DSA) | Download |
|---------------------------------|----------|



۱.۱ محدوده مستندسازی

چه اطلاعاتی باید ارائه شود؟

اطلاعات زیر از شیوه‌های خوب در مستندسازی داده، بایگانی داده‌های انگلستان، دانشگاه Essex استخراج شده است. همچنین به "راهنمای مرجع سریع برای بایگانی‌کنندگان داده" در شبکه بین‌المللی پیمایش‌های خانوار مراجعه کنید.

سه نوع اصلی مستندسازی ایده‌آل برای مجموعه داده وجود دارد: ۱. مواد توضیحی؛ ۲. اطلاعات متنی؛ و ۳. کاتالوگ کردن.

۱. مواد توضیحی

نشان دهنده حداقل موادی است که باید ایجاد و حفظ شوند تا از ماندگاری و عملکرد طولانی‌مدت یک مجموعه داده و درک کامل مجموعه داده و محتویات آن اطمینان حاصل شود.

• اطلاعاتی در مورد روش‌های گردآوری داده‌ها

این بخش، فرایند جمع‌آوری داده‌ها را توصیف می‌کند، خواه این یک پیمایش، جمع‌آوری اطلاعات اداری، یا رونویسی از منبع سند باشد. باید ابزارهای مورد استفاده و روش‌های به‌کار گرفته شده و نحوه توسعه آنها را توضیح دهد. در صورت امکان، جزئیات طرح نمونه‌گیری و قالب‌های نمونه‌گیری باید درج شوند. همچنین گنجانیدن اطلاعات هر گونه فرآیند نظارتی که در طول جمع‌آوری داده‌ها انجام شده و نیز جزئیات کنترل‌های کیفیت، مطلوب و مفید است.

• اطلاعاتی در مورد ساختار مجموعه داده

کلید این اطلاعات یک سند مفصل است که ساختار مجموعه داده را توصیف می‌کند و شامل اطلاعاتی در مورد روابط بین فایل‌ها یا سوابق فردی در مطالعه است. برای مثال، باید متغیرهای کلیدی مورد نیاز برای شناسایی منحصر به فرد موضوعات در بین فایل‌ها، تعداد موارد و متغیرهای هر فایل، و تعداد فایل‌های مجموعه داده را شامل شود. برای مدل‌های رابطه‌ای، باید نموداری ایجاد شود که ساختار و روابط بین مجموعه داده‌ها و عناصر را نشان دهد.

• اطلاعات فنی



این اطلاعات به چارچوب فنی مربوط می‌شود و باید شامل سیستم رایانه‌ای مورد استفاده برای تولید فایل‌ها؛ بسته‌های نرم‌افزاری که فایل‌ها با آنها ساخته شده‌اند؛ رسانه‌ای که داده‌ها بر روی آن ذخیره شده‌اند، و یک لیست کامل از تمام فایل‌های داده موجود در مجموعه داده باشد.

• متغیرها و مقادیر، برنامه‌های کُدگذاری و طبقه‌بندی

مستندات باید حاوی لیست کاملی باشند که تمام متغیرها یا فیلدها را در مجموعه داده از جمله، توضیح کامل و جزئیات کامل در مورد کُدگذاری و طبقه‌بندی‌های مورد استفاده برای اطلاعات اختصاص داده شده به آن فیلدها را توصیف می‌کند. به خصوص مهم است که فیلدهای خالی و مفقود، توضیح داده شده و حساسی شوند. شناسایی متغیرهایی که طبقه‌بندی‌های استاندارد کُدگذاری برای آنها اعمال می‌شود و ثبت نسخه طرح طبقه‌بندی استفاده شده، ترجیحاً با ارجاع کتابنامه به آن کُد، مفید است.

• اطلاعات مربوط به متغیرهای مشتق شده

بسیاری از تولیدکنندگان داده‌ها، متغیرهای جدیدی را از داده‌های اصلی استخراج می‌کنند. این ممکن است به سادگی گروه‌بندی داده‌های سن خام (به سال) براساس گروه‌های سنی مناسب برای پیمایش باشد، یا ممکن است بسیار پیچیده‌تر باشد و نیاز به استفاده از الگوریتم‌های پیچیده داشته باشد. وقتی متغیرهای گروه‌بندی شده یا مشتق شده ایجاد می‌شوند، مهم است که منطق گروه‌بندی یا مشتق‌گیری روشن باشد. گروه‌بندی ساده، مانند سن، می‌تواند در فرهنگ لغت داده‌ها گنجانده شود. مشتقات پیچیده‌تر به ابزارهای دیگری برای ثبت اطلاعات نیاز دارند. بهترین روش برای توصیف این موارد، استفاده از نمودارهای جریان یا عبارات دقیق Boolean است. اطلاعات پشتیبان کافی باید ارائه شود تا امکان ارتباط آسان بین متغیرهای اصلی مورد استفاده و متغیرهای حاصله فراهم شود. علاوه بر این، الگوریتم‌های رایانه‌ای که برای ایجاد مشتقات استفاده می‌شوند باید همراه با اطلاعات روی نرم‌افزار ذخیره شوند.

• وزن‌گیری و ناخالصی

متغیرهای وزنی و ناخالص باید به طور کامل همراه با توضیحاتی در مورد ساختار متغیرها و نشانه‌هایی روشن از شرایطی که باید در آنها استفاده شوند، مستند شوند. که مورد دوم ذکر شده، به ویژه هنگامی اهمیت دارد که وزن‌های مختلف برای اهداف مختلف اعمال می‌شود.

• منبع داده



جزئیات مربوط به منبعی که داده‌ها از آن گرفته شده‌اند باید درج شود. برای مثال، وقتی منبع داده شامل پاسخ به پرسشنامه‌های پیمایش است، هر سؤال باید به دقت در مستند ثبت شود. در حالت ایده‌آل، متن شامل ارجاع به متغیر(های) تولید شده باشد. همچنین توضیح شرایطی که در آن سؤال مطرح می‌شود از جمله، در صورت امکان، مواردی که برای آن اعمال می‌شود و در حالت ایده‌آل، خلاصه‌ای از آمار پاسخگویی نیز مفید و مطلوب است.

• محرمانه‌گی و بی‌نام‌سازی

تعیین این که آیا داده‌ها حاوی هرگونه اطلاعات محرمانه در مورد افراد، خانواده‌ها، سازمان‌ها یا مؤسسات هستند یا خیر، مهم است. اگر چنین است، چنین اطلاعاتی باید همراه با هرگونه توافقی در مورد نحوه استفاده از داده‌ها، مانند پاسخ‌دهندگان پیمایش، ثبت شود. مسائل مربوط به محرمانه بودن ممکن است تجزیه و تحلیل‌هایی را که باید انجام شود یا نتایجی که باید منتشر شود را محدود کند، به ویژه اگر داده‌ها برای استفاده ثانویه در دسترس قرار گیرند. اگر داده‌ها برای جلوگیری از شناسایی، بی‌نام شدند، عاقلانه است که روش بی‌نام‌سازی و تأثیر آن بر داده‌ها ثبت شود، زیرا چنین تغییری ممکن است تجزیه و تحلیل بعدی را محدود کند.



۲. اطلاعات متنی

این موضوع، مطالبی را در مورد زمینه جمع‌آوری داده‌ها و نحوه استفاده از آنها در اختیار کاربران قرار می‌دهد. این اطلاعات به مستندات غنا و عمق می‌بخشد و کاربر ثانویه را قادر می‌سازد تا پیشینه و فرآیندهای پشت جمع‌آوری داده‌ها را به‌طور کامل درک کند. همچنین یک سابقه تاریخی حیاتی برای محققان آینده ایجاد می‌کند.

• شرح پروژه اولیه

جزئیات باید در مورد سابقه پروژه یا فرآیندی که مجموعه داده را تولید کرده است، شامل اطلاعات مربوط به چارچوب فکری و محتوایی ارائه شود. برای مثال، توضیحات می‌تواند موضوعاتی مانند ذیل را پوشش دهد:

- چرا جمع‌آوری داده‌ها ضروری بود؛
- اهداف پروژه؛
- چه کسی یا چه چیزی مورد مطالعه قرار گرفته است؛
- پوشش جغرافیایی و زمانی؛
- انتشارات یا تحولات سیاستی که پروژه به آنها کمک کرده است یا به دلیل آن پروژه به عنوان پاسخ به وجود آمده است؛ و
- سایر اطلاعات مربوط.

• منشأ مجموعه داده

این اطلاعات به جنبه‌هایی مانند سابقه فرایند جمع‌آوری داده‌ها، تغییرات و تحولاتی که در داده‌ها و روش‌شناسی رخ داده است، یا هرگونه تعدیل^۱ انجام شده مربوط می‌شود. موارد زیر را نیز می‌توان ارائه کرد:

- جزئیات خطاهای داده؛
- مشکلاتی که در فرایند جمع‌آوری اطلاعات، ورود، بررسی و پاک کردن داده‌ها وجود دارد؛
- تبدیل به نرم‌افزار یا سیستم عامل متفاوت؛
- ارجاعات کتابنامه به گزارش‌ها یا نشریات مرتبط با مطالعه؛ و
- سایر اطلاعات مفید در مورد چرخه عمر مجموعه داده.



- **مجموعه داده‌های سریال و سری زمانی، ویرایش‌های جدید**

برای مجموعه داده‌های مقطعی، پانل یا سری زمانی، بدست آوردن اطلاعات بیشتر در مورد تغییرات در متن سؤال، برچسب‌گذاری متغیر یا روش نمونه‌گیری یا سایر تغییرات مفید است.





۳. کاتالوگ کردن

این موضوع دو هدف را دنبال می‌کند. اول، ثبت یک کتابنامه از مجموعه داده است که اجازه می‌دهد تا مجموعه داده به درستی تایید شود و در نشریات مورد استناد قرار گیرد، و مطالب به یک ثبت رسمی برای اهداف حفاظت شدن تبدیل شوند. دوم، ابزار اساسی برای کشف منبع است، که به مجموعه داده اجازه می‌دهد تا با ارائه اطلاعاتی که به کاربران ثانویه کمک می‌کند مطالعه را برای آنها مفید تشخیص دهند، به صورت منحصر به فرد در مجموعه شناسایی شود.

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

 [Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle \(5th ed.\)](#)  Download

 [Managing and sharing data - Best practice for researchers](#)  Download

Tools

 [Nesstar Publisher \(IHSN Metadata Editor\)](#)  Download

۲,۱ استانداردها و مدل‌های فراداده

مجموعه‌ای از استانداردها و مدل‌های فراداده برای تسهیل ارتباط داده‌ها بین سازمان‌ها و سیستم‌های نرم‌افزاری و بهبود کیفیت مستندات آماری ارائه شده به کاربران داده، توسعه یافته است. این استانداردهای فراداده، چارچوبی ساختار یافته برای سازماندهی و انتشار اطلاعات در مورد محتوا و ساختار اطلاعات آماری ارائه می‌دهند.

استانداردهای فراداده XML

- استاندارد فراداده طرح مستندسازی داده‌ها (DDI)، که به‌طور خاص برای مستندسازی و



کاتالوگ کردن ریزداده‌ها ارائه شده است؛

- طرح فراداده^۱ Core Dublin (DCMI)
- استاندارد تبادل داده‌های آماری و فراداده^۲ (SDMX)، که برای مستندسازی و به اشتراک‌گذاری داده‌های سری زمانی ارائه شده است؛ به‌طور مستقیم برای ریزداده‌های پیمایش قابل استفاده نیست، اما برای شاخص‌های مشتق شده مناسب است.

سایر استانداردهای فراداده و مدل‌های فراداده (یا "چارچوب‌ها")

- ISO 11179
- مدل عمومی فرآیند کسب و کار آماری^۳ (GSBPM)؛
- مدل اطلاعات عمومی^۴ (GSIM).

XML	DDI	DCMI	ISO 11179	SDMX	GSBPM	GSIM
-----	-----	------	-----------	------	-------	------

زبان XML

زبان نشانه‌گذاری گسترش‌پذیر^۵ یا XML به عنوان یک ابزار مشترک جهت ساختاردهی اطلاعات برای اشتراک‌گذاری در وب و بین سیستم‌های نرم‌افزاری توسعه داده شد. XML روشی برای برچسب‌گذاری متن^۶ (تگ کردن متن) برای معنی به جای ظاهر است، به عبارت دیگر، XML می‌تواند محتوای متن را با برچسب‌گذاری (تگ کردن) اطلاعات معنی‌دار سازماندهی کند. اگرچه "برچسب‌ها" (تگ‌ها) از نظر سازماندهی با "فیلدهای" یک پایگاه داده یکسان هستند، اما تفاوت بین فایل‌های XML و پایگاه داده در این است که فایل‌های XML، فایل‌های متن معمولی هستند که با استفاده از هر ویرایشگر متن استاندارد قابل مشاهده و ویرایش می‌باشند. فایل را می‌توان مانند یک پایگاه داده معمولی با استفاده از ابزارهایی مانند Xpath یا Xquery جستجو و پرس‌وجو کرد و با استفاده از Xforms ویرایش کرد. (یک آموزش در وب در مورد این ابزارها در آدرس <http://www.w3schools.com/xml> دنبال کنید.) همان‌طور که محتوای یک پایگاه داده را می‌توان به

1. Dublin Core Metadata Initiative (DCMI)
2. Statistics Data and Metadata Exchange (SDMX)
3. Generic Statistical Business Process Model (GSBPM)
4. Generic Statistical Information Model (GSIM)
5. eXtensible Markup Language (XML)
6. Tagging text



گزارش تبدیل کرد، اسناد XML را می‌توان خواند و توسط سایر نرم‌افزارها به قالب‌های کاربرپسند مانند صفحات گسترده، فایل‌های PDF یا صفحات وب تبدیل کرد.

مثال زیر نشان می‌دهد که چگونه اطلاعات متنی درباره یک پیمایش در XML ارائه می‌شود.

همان اطلاعاتی که با استفاده از برچسب‌های (تگ‌های) DDI به XML تبدیل شده‌اند به این شکل است:

```
<tit1>Multiple Indicator Cluster Survey 2005</tit1>
<altTit1>MICS</altTit1>
<AuthEnty>National Statistics Office (NSO)</AuthEnty>
<fundAg abbr="UNICEF">United Nations Children Fund</fundAg>
<collDate date="2005-01" event="start"/>
<collDate date="2005-03" event="end"/>
<nation>Popstan</nation>
<geogCover>National</geogCover>
<sampProc>5,000 households, stratified two stages</sampProc>
<respRate>98 percent</respRate>
```




استفاده از برچسب‌ها (تگ‌ها) به ویژه هنگامی بسیار قدرتمند است که جامعه کاربری بر روی مجموعه‌ای از برچسب‌های (تگ‌های) مشترک (مانند استانداردهای طرح مستندسازی داده‌ها (DDI) یا طرح فراداده Core Dublin (DCMI)) توافق می‌کنند. تصویب مجموعه‌ای مشترک از برچسب‌های (تگ‌های) XML مزایای عمده‌ای در مستندسازی ریزداده‌ها از جمله، ایجاد یک "چک‌لیست" جامع از عناصر فراداده مفید؛ امکان ارزیابی محتویات فایل با تعیین این که آیا برچسب‌های خاصی در آن فایل هستند یا نیستند؛ ایجاد یک کاتالوگ مجموعه داده که می‌تواند برای عناصر فراداده مهم پرس‌وجو شود؛ و امکان تبدیل فایل به فرمت‌های کاربر پسندتر را دارد. فایل‌های XML را می‌توان با استفاده از XSL Transformations (تبدیل‌های XSL) به HTML، PDF یا سایر اسناد تبدیل کرد یا با استفاده از سرویس‌های وب یا SOAP بین شبکه‌ها یا اینترنت مبادله کرد. مثالی از کاربرد "تبدیل XSL" در فایل XML قبلی، صفحه وب HTML زیر است:

unicef Statistics		End Decade Assessment	
		Multiple Indicator Cluster Survey	
		POPSTAN	
		Multiple Indicators Cluster Survey (MICS)	
Data producer:	National Statistics Office (NSO)		
Funding:	United Nations Children Fund (UNICEF)		
Coverage:	National		
Sampling:	5,000 households, stratified two stages		
Response rate:	98 percent		
Data collected from:	Jan. 2005	to:	Mar. 2005

منابع مرتبط به اشتراک گذاشته شده است:



Links

DDI Alliance website	Download
Dublin Core Metadata Initiative (DCMI)	Download
Generic Statistical Business Process Model (GSBPM)	Download
Generic Statistical Information Model (GSIM)	Download
SDMX – Statistical Data and Metadata Exchange	Download


Documents

DDI and SDMX	Download
DDI and SDMX: Complementary, Not Competing, Standards	Download
Metadata Standards - IHSN technical note (DRAFT)	(255.8 KB)



طرح مستندسازی داده‌ها (DDI)

از گذشته، تولیدکنندگان داده‌ها کتاب‌های کُد مبتنی بر متن می‌نوشتند. برای استفاده کامل از فناوری وب، اکثر استانداردها در حال حاضر به زبان XML تعریف شده‌اند. طرح مستندسازی داده‌ها (DDI) استاندارد اختصاصی برای مستندات ریزداده است که امکان مستندسازی حتی پیچیده‌ترین فایل‌های ریزداده را به گونه‌ای فراهم می‌کند که همزمان انعطاف‌پذیر و دقیق باشند. این ابزار ساده‌ای برای ثبت و انتقال تمام ویژگی‌های بارز مجموعه‌های ریزداده است.



DDI Alliance website

The DDI Alliance, hosted by the University of Michigan, maintains the DDI metadata standard. Its website provides detailed information on the DDI Codebook and DDI Lifecycle specifications, and provides a catalog of tools that make use of the DDI standard.

[Download](#)

پیمان طرح مستندسازی داده‌ها (DDI Alliance) دو نسخه از مشخصات DDI را حفظ می‌کند: کتاب کُد DDI (که توسط شبکه بین‌المللی پیمایش‌های خانوار استفاده و توصیه می‌شود)، و چرخه عمر DDI (نسخه پیچیده‌تر مشخصات) است. کتاب کُد DDI تحول بزرگی از "کتاب کُد" الکترونیکی است که زمانی آشنا بود و مجموعه‌ای از قابلیت‌ها را حفظ می‌کند اما دامنه و دقت اطلاعات موجود در آن را تا حد زیادی افزایش می‌دهد.

مشخصات فراداده DDI از کنسرسیوم بین دانشگاهی تحقیقات سیاسی و اجتماعی (ICPSR)، سازمانی مبتنی بر عضویت با بیش از ۵۰۰ کالج و دانشگاه عضو در سراسر جهان، سرچشمه گرفته است. اکنون پروژه پیمان^۱ مؤسسات، در آمریکای شمالی و اروپا است. مؤسسات عضو، بسیاری از بزرگترین تولیدکنندگان داده و بایگانی‌های داده در جهان را شامل می‌شوند.

مشخصات DDI انواع داده‌های حاصل از پیمایش‌ها، سرشماری‌ها، سوابق اداری، آزمایش‌ها، مشاهده مستقیم و سایر روش‌های سیستماتیک برای ایجاد اندازه‌گیری‌های تجربی را مورد بررسی قرار می‌دهد. برای مثال، واحدهای تجزیه و تحلیل می‌توانند افراد، خانوارها، خانواده‌ها، مؤسسات تجاری، معاملات، کشورها، یا سایر موضوعات مورد

1. Alliance



علاقه علمی باشند. به‌طور مشابه، مشاهدات می‌توانند شامل اندازه‌گیری در یک نقطه از زمان در یک مکان واحد باشد، مانند نمونه‌ای از افراد یک کشور در طول یک هفته. یا ممکن است شامل مشاهدات مکرر در تنظیمات متعدد، از جمله، داده‌های مقطعی طولی و مکرر از بسیاری از کشورها، و همچنین سری‌های زمانی داده‌های جمع‌آوری شده باشد. مشخصات DDI همچنین توضیحات کاملی از روش مطالعه (برای مثال، نحوه جمع‌آوری داده‌ها، روش‌های نمونه‌گیری قابل اجرا، جهان، مناطق جغرافیایی مطالعه، سازمان و افراد مسئول و غیره) ارائه می‌دهد.

ساختار

مشخصات DDI اجازه می‌دهد تا همه جنبه‌های یک پیمایش به تفصیل شرح داده شود: روش، مسئولیت‌ها، فایل‌ها، و متغیرها. این فهرستی ساختاریافته و جامع از صدها عنصر و ویژگی است که ممکن است برای مستندسازی یک مجموعه داده مورد استفاده قرار گیرد، اگرچه بعید است که یک مطالعه از همه آنها استفاده کند. با این حال، برخی از عناصر، مانند "عنوان" اجباری هستند و باید منحصر به فرد باشند. سایر عناصر اختیاری هستند و می‌توانند تکرار شوند، مانند "نهاد نویسنده/محقق اصلی"، زیرا شامل اطلاعات مربوط به شخص(ها) و/یا سازمان(های) مسئول پیمایش است. عناصر کتاب‌کُد DDI (نسخه 2.0) در ۵ بخش سازماندهی شده است:

بخش ۱،۰: شرح سند

یک مطالعه (به عبارت دیگر، پیمایش، سرشماری یا سایر موارد) همیشه توسط همان آژانسی که داده‌ها را تهیه کرده است، مستند و منتشر نمی‌شود. بنابراین، ارائه اطلاعات (به عبارت دیگر، فراداده) نه تنها در مورد خود مطالعه، بلکه در مورد فرایند مستندسازی نیز مهم است. شرح سند شامل یک نمای کلی است - "فراداده درباره فراداده"^۱ XSL که توصیف سند XML سازگار با DDI است.

بخش ۲،۰: شرح مطالعه

شرح مطالعه، یک نمای کلی از مطالعه است و شامل اطلاعاتی درباره چگونگی استناد به مطالعه؛ چه کسی داده‌ها را جمع‌آوری، گردآوری و توزیع کرده است؛ خلاصه (چکیده) از محتوای داده؛ جزئیات روش‌های جمع‌آوری و پردازش داده‌ها؛ و غیره می‌باشد.

1. Authoring Entity/Primary Investigator
2. Metadata about metadata



بخش ۳، ۵: شرح فایل داده‌ها

این بخش محتوای هر فایل داده، تعداد ثبت و متغیرها، نسخه، تولیدکننده و غیره را توضیح می‌دهد.

بخش ۴، ۵: شرح متغیر

این بخش جزئیات هر متغیر، از جمله، متن پرسش تحت‌اللفظی، همگانی^۱ بودن، برچسب‌های متغیر و مقدار، روش‌های اشتقاق و جانمایی و غیره را ارائه می‌دهد.

بخش ۵، ۵: سایر موارد

این بخش امکان توضیح سایر مطالب مرتبط با مطالعه را فراهم می‌آورد. این موارد می‌تواند شامل اسنادی مانند پرسشنامه‌ها، اطلاعات کُذگذاری، گزارشات فنی و تحلیلی و راهنمای مصاحبه‌گران؛ پردازش داده‌ها و برنامه‌های تحلیلی؛ عکس‌ها؛ یا نقشه‌ها باشد.

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Links

DDI Alliance website	Download
Dublin Core Metadata Initiative (DCMI)	Download
Generic Statistical Business Process Model (GSBPM)	Download
Generic Statistical Information Model (GSIM)	Download
SDMX - Statistical Data and Metadata Exchange	Download


Documents

DDI and SDMX	Download
DDI and SDMX: Complementary, Not Competing, Standards	Download
Metadata Standards - IHSN technical note (DRAFT)	(255.8 KB)



مشخصات فراداده (DCMI) Core Dublin

مطالب زیر از وبسایت DCMI ([http //dublincore.org](http://dublincore.org)) گرفته شده است.



Dublin Core Metadata Initiative (DCMI)

The DCMI Metadata Element Set (ISO standard 15836), also known as the Dublin Core metadata standard, is a simple set of elements for describing digital resources, especially those resources related to microdata, such as questionnaires, reports, manuals, data processing scripts and programs, etc.

[Download](#)

مجموعه عناصر فراداده DCMI (استاندارد ISO 15836)، که به عنوان استاندارد فراداده Core Dublin نیز شناخته می‌شود، مجموعه ساده‌ای از عناصر برای توصیف منابع دیجیتال است. این استاندارد به ویژه در توصیف منابع مرتبط با ریزداده‌ها مانند پرسشنامه‌ها، گزارش‌ها، کتابچه‌های راهنما، اسکریپت‌ها و برنامه‌های پردازش داده و غیره مفید است. در سال ۱۹۹۵ توسط مرکز کتابخانه آنلاین کامپیوتر^۱ (OCLC) و مرکز ملی برنامه‌های ابررایانه^۲ (NCSA) در کارگاهی در Dublin، Ohio تأسیس شد. در طول سال‌ها، این استاندارد به پرکاربردترین استاندارد برای توصیف منابع دیجیتال در وب تبدیل شده است و به عنوان استاندارد ISO در سال ۲۰۰۳ تأیید شد. این استاندارد توسط DCMI، یک سازمان بین‌المللی که به ارتقای استانداردهای فراداده قابل همکاری اختصاص دارد، حفظ و توسعه می‌یابد.

دلیل اصلی موفقیت استاندارد فراداده Dublin Core سادگی آن است. از همان ابتدا، هدف طراحان این بوده است که مجموعه عناصر را تا حد امکان کوچک و ساده نگه دارند تا امکان استفاده از استاندارد توسط افراد غیرمتخصص فراهم شود. این استاندارد همچنین ایجاد سوابق توصیفی^۳ ساده برای منابع اطلاعاتی را آسان و ارزان می‌کند و در

1. Online Computer Library Center (OCLC)

2. National Center for Supercomputing Applications (NCSA)

3. Descriptive records



عین حال امکان بازیابی مؤثر آن منابع را در وب یا هر محیط شبکه‌ای مشابه فراهم می‌کند.

Dublin Core در ساده‌ترین شکل خود از ۱۵ عنصر فراداده زیر تشکیل شده است که همه آنها اختیاری و قابل تکرار هستند: عنوان؛ ارتباط؛ حقوق؛ موضوع؛ پوشش؛ تاریخ؛ شرح؛ ایجادکننده؛ قالب؛ نوع؛ ناشر؛ شناسه؛ منبع؛ مشارکت‌کننده؛ زبان.



منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Links

DDI Alliance website	Download
Dublin Core Metadata Initiative (DCMI)	Download
Generic Statistical Business Process Model (GSBPM)	Download
Generic Statistical Information Model (GSIM)	Download
SDMX - Statistical Data and Metadata Exchange	Download

Documents

DDI and SDMX	Download
DDI and SDMX: Complementary, Not Competing, Standards	Download
Metadata Standards - IHSN technical note (DRAFT)	(255.8 KB)

ISO 11179 - ثبت‌های فراداده- فناوری اطلاعات (MDR)

استاندارد بین‌المللی ISO/IEC 11179-1 توسط کمیته فنی مشترک 1 JTC ISO/IEC، فناوری اطلاعات، کمیته فرعی SC 32، خدمات مدیریت داده توسعه یافته است. ISO/IEC 11179، استانداردسازی و ثبت عناصر داده را برای قابلیت درک و اشتراک‌گذاری داده‌ها توصیف می‌کند. استانداردسازی و ثبت عناصر داده همان‌طور که در ISO/IEC 11179 توضیح داده شده است، امکان ایجاد یک محیط داده مشترک را در زمان بسیار کمتر و با تلاش بسیار کمتر از روش‌های متداول مدیریت داده‌ها، فراهم می‌کند. (منبع: ISO-IEC 1999، موجود در http://metadata-stds.org/11179-1/ISO-IEC_11179-1_1999_IS_E.pdf)

منابع مرتبط به اشتراک گذاشته شده است:



Related Resources

Links

DDI Alliance website	Download
Dublin Core Metadata Initiative (DCMI)	Download
Generic Statistical Business Process Model (GSBPM)	Download
Generic Statistical Information Model (GSIM)	Download
SDMX – Statistical Data and Metadata Exchange	Download

Documents

DDI and SDMX	Download
DDI and SDMX: Complementary, Not Competing, Standards	Download
Metadata Standards - IHSN technical note (DRAFT)	(255.8 KB)


تبادل داده‌های آماری و فراداده (SDMX)

SDMX با تمرکز بر سری‌های زمانی و شاخص‌ها، نتیجه تلاش مشترک بین بانک تسویه‌حساب‌های بین‌المللی^۱، بانک مرکزی اروپا^۲، اداره آمار اروپا^۳، صندوق بین‌المللی پول، سازمان همکاری و توسعه اقتصادی، سازمان ملل متحد و بانک جهانی برای ایجاد مشخصات XML جهت پشتیبانی از تبادل کل داده‌ها و فراداده‌ها است. SDMX سه نوع استاندارد فراداده آماری را ارائه می‌دهد: استانداردهایی برای قالب‌های داده، استانداردهایی برای فراداده‌ها، و معماری مبتنی بر رجیستری (registry) برای پیاده‌سازی این استانداردها و تبادل داده‌ها بین سیستم‌ها.

یکی از الزامات SDMX هماهنگی با سایر مشخصات فراداده مانند DDI بود. هر یک از فراداده‌های DDI، که بر فراداده‌ها و ریزداده‌های آرشیوی به جای داده‌های انبوه تأکید دارند، در قالب فراداده SDMX معادل، قابل تبادل هستند. این کار عملکرد متقابل فراداده را در فضای نام (علوم رایانه) تضمین می‌کند.

-
1. Bank for International Settlements
 2. European Central Bank (ECB)
 3. EUROSTAT





SDMX – Statistical Data and Metadata Exchange

SDMX stands for Statistical Data and Metadata Exchange—the electronic exchange of statistical information. Its goal is to explore e-standards that could allow us to gain efficiency and avoid duplication of effort in our own work and possibly in the work of others in the field of statistical information.

[Download](#)

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Links

DDI Alliance website	Download
Dublin Core Metadata Initiative (DCMI)	Download
Generic Statistical Business Process Model (GSBPM)	Download
Generic Statistical Information Model (GSIM)	Download
SDMX – Statistical Data and Metadata Exchange	Download

Documents

DDI and SDMX	Download
DDI and SDMX: Complementary, Not Competing, Standards	Download
Metadata Standards - IHSN technical note (DRAFT)	Download (255.8 KB)




مدل عمومی فرآیند کسب و کار آماری (GSBPM)

GSBPM فرآیندهای آماری مانند اجرای یک پیمایش را در ۹ مرحله توصیف می‌کند که هر مرحله به فرآیندهای

فرعی تقسیم می‌شود:

- مشخص نمودن نیازهای داده؛
- طراحی؛
- ساختار؛
- جمع‌آوری (شامل ورود داده‌ها)؛
- فرآیند (شامل ویرایش داده‌ها)؛
- تجزیه و تحلیل؛
- انتشار؛
- بایگانی؛
- ارزیابی.

علاوه بر این ۹ مرحله، GSBPM شامل دو جزء کلی است: مدیریت کیفیت و مدیریت فراداده.



Generic Statistical Business Process Model (GSBPM)

The GSBPM provides a framework to describe the statistical production process in terms of standard components (i.e., phases and sub-processes). It is intended to apply to all activities undertaken by producers of official statistics, at both the national and international levels, which result in data outputs. It is designed to be independent of the data source, so it can be used for the description and quality assessment of processes based on surveys, censuses, administrative records, and other non-statistical or mixed sources.

[Download](#)

منابع مرتبط به اشتراک گذاشته شده است:



Related Resources

Links

DDI Alliance website	Download
Dublin Core Metadata Initiative (DCMI)	Download
Generic Statistical Business Process Model (GSBPM)	Download
Generic Statistical Information Model (GSIM)	Download
SDMX - Statistical Data and Metadata Exchange	Download

Documents

DDI and SDMX	Download
DDI and SDMX: Complementary, Not Competing, Standards	Download
Metadata Standards - IHSN technical note (DRAFT)	Download (255.8 KB)



مدل اطلاعات عمومی آماری (GSIM)

GSIM یک چارچوب مرجع از تعاریف، ویژگی‌ها و روابط پذیرفته‌شده بین‌المللی است که اطلاعات مورد استفاده در تولید آمار رسمی و اشیاء اطلاعاتی را توصیف می‌کند. این چارچوب، توصیفات کلی از تعریف، مدیریت و استفاده از داده‌ها و فراداده‌ها را در سراسر فرآیند تولید آماری امکان‌پذیر می‌کند.



GSIM یک زبان مشترک برای توصیف اطلاعات ارائه می‌دهد که از کل فرآیند تولید آماری، از شناسایی نیازهای کاربر تا انتشار محصولات آماری را پشتیبانی می‌کند.

GSIM با استانداردهای مربوط به مدیریت داده و تبادل داده، مانند DDI و SDMX، هم‌راستا است، اما مستقیماً به آنها یا هیچ فناوری خاصی وابسته نیست.

GSIM نرم‌افزار و استاندارد فناوری اطلاعات (IT) نیست. این یک رویکرد استراتژیک و یک روش جدید از تفکر است که برای گرد هم آوردن آماردانان، روش‌شناسان و متخصصان فناوری اطلاعات به منظور مدرن‌سازی و ساده‌سازی تولید آمار رسمی طراحی شده است.

اطلاعات ذکر شده از بروشور GSIM که در وبسایت GSIM موجود است، استخراج شده است.

منابع مرتبط به اشتراک گذاشته شده است:



Related Resources

Links

DDI Alliance website	Download
Dublin Core Metadata Initiative (DCMI)	Download
Generic Statistical Business Process Model (GSBPM)	Download
Generic Statistical Information Model (GSIM)	Download
SDMX – Statistical Data and Metadata Exchange	Download

Documents

DDI and SDMX	Download
DDI and SDMX: Complementary, Not Competing, Standards	Download
Metadata Standards - IHSN technical note (DRAFT)	(255.8 KB)

۳.۱ ابزارهایی برای مستندسازی ریزداده‌ها

شبکه بین‌المللی پیمایش‌های خانوار، پذیرش استاندارد فراداده DDI را برای مستندسازی ریزداده‌های پیمایش توصیه می‌کند. DDI یک استاندارد XML است. بنابراین، مستندسازی پیمایش‌ها با استفاده از DDI، به یک ویرایشگر فراداده DDI نیاز دارد. یک ویرایشگر تخصصی رایگان، برنامه Nesstar Publisher، در دسترس است.

برای دسترسی به ویرایشگر فراداده، (Nesstar Publisher) DDI به بخش نرم‌افزار رجوع کنید.

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

Quick Reference Guide for Data Archivists	Download
---	----------

۲. بی‌نام‌سازی ریزداده

آژانس‌های آمار و سایر تولیدکنندگان داده‌ها به‌طور فزاینده‌ای ریزداده‌های به‌دست‌آمده از پیمایش‌های نمونه،



سرشماری‌ها، و سیستم‌های جمع‌آوری داده‌های اداری را منتشر می‌کنند. انتشار ریزداده‌ها به دلیل تقاضای زیاد جامعه پژوهشی، فشار برای شفافیت، و گاه تعهدات قانونی یا قراردادی ضروری است و باید به گونه‌ای انجام شود که محرمانه بودن اطلاعات ارائه شده توسط پاسخ‌دهندگان حفظ شود.

در این بخش موارد ذیل ارائه می‌شود:

- اصول مهم مرتبط با بی‌نام‌سازی ریزداده‌ها؛
- تکنیک‌های مختلف اندازه‌گیری خطر افشا؛
- روش‌های موجود برای کاهش خطر افشا؛
- روش‌های ارزیابی از دست دادن اطلاعات.

همچنین موارد ذیل نیز در ادامه ارائه شده است:

- ابزارهای موجود؛ و
- مجموعه‌ای از روش‌ها.

بی‌نام‌سازی معمولاً برای تولید فایل‌های استفاده عموم و تا حدی کمتر برای تولید فایل‌های دارای مجوز لازم است. اما بی‌نام‌سازی تنها یکی از راه‌حل‌های متعدد برای به حداقل رساندن خطر افشا در هنگام توزیع ریزداده است. سایر اقدامات قانونی و سازمانی نیز به این امر کمک می‌کند. برای مجموعه داده‌هایی که به کاربران منتخب با حسن نیت ارائه می‌شوند، توافق‌نامه قانونی می‌تواند سطح امنیتی بالاتری نسبت به بی‌نام‌سازی به تنهایی داشته باشد (به بخش تدوین خط‌مشی انتشار داده مراجعه کنید).



۳ راهنما تولید شده است:

- راهنمای تئوری که نمای کلی از روش‌های رایج و همچنین فرآیند کنترل افشای آماری (SDC) را ارائه می‌کند؛
- راهنمای تمرینی-که نحوه اعمال روش‌ها را با استفاده از واسط نوشتاری برای بسته R package sdcMicro توصیف می‌کند، و
- دفترچه راهنما برای sdcApp- یک رابط کاربری گرافیکی sdcMicro برای کاربرانی که در استفاده از R از واسط نوشتاری راحت نیستند.



Statistical Disclosure Control for Microdata: Theory

Thisj Benschop and Matthew Welch - 11/12/2019

This guide provides an introduction to the theory of Statistical Disclosure Control (SDC) for microdata. It includes an overview of the most commonly applied methods in SDC, a step-by-step overview of the complete SDC process and many examples from practice in National Statistics Offices (NSOs).

For guidance on the technical implementation of the theory mentioned in the guide, please refer to our guides:

- *Statistical Disclosure Control for Microdata: A Practice Guide for guidance on the application of methods and on using sdcMicro from the command-line*

- *sdcApp manual for guidance on the application of methods and on using the GUI sdcApp available for sdcMicro*

[Download](#)



Statistical Disclosure Control for Microdata: A Practice Guide

Thisj Benschop, Matthew Welch - June 2016

Releasing data in a safe way is required to protect the integrity of the statistical system, by ensuring agencies honor their commitment to respondents to protect their identity. Agencies do not widely share, in substantial detail, their knowledge and experience using SDC and the processes for creating safe data with other agencies. This makes it difficult for agencies new to the process to implement solutions. We consolidated knowledge from literature as well as from our own experience to inform our discussion of the processes and methods presented in this guide. This guide focuses on the implementation of methods and uses the free R based package sdcMicro for its examples. If you are interested in reading in detail about the theory behind the methods used, we suggest reading our accompanying guide: Statistical Disclosure Control for Microdata: Theory.

[Download](#)



Statistical Disclosure Control for Microdata: A Practice Guide - Case Study Data and R Script

June 2016

[Download \(1 MB\)](#)



sdcApp Reference Manual

Thisj Benschop, Matthew Welch - 2019-11-12

This is documentation and guidance for using sdcApp, a graphic user interface for the sdcMicro R package. sdcMicro provides tools for Statistical Disclosure Control (SDC) for microdata, also known as microdata anonymization. For an overview of the theory of SDC for microdata we suggest reading: Statistical Disclosure Control for Microdata: A Theory Guide.

[Download](#)

و همچنین یک مقاله کاربردی شبکه بین‌المللی پیمایش‌های خانوار:



Introduction to Statistical Disclosure Control (SDC)

Matthias Templ, Bernhard Meindl, Alexander Kowarik and Shuang Chen - August 2014

This guide, Introduction to Statistical Disclosure Control (SDC), discusses common SDC methods for microdata obtained from sample surveys, censuses and administrative sources.

[Download](#)

منابع مرتبط به اشتراک گذاشته شده است:



Related Resources

Documents

Anonymisation: managing data protection risk code of practice	Download
Handbook on Statistical Disclosure Control (Version 1.2)	Download
Introduction to Statistical Disclosure Control (SDC)	(752.9 KB)
Managing Statistical Confidentiality and Microdata Access - Principles and guidelines of Good Practice	Download
National Statistics Code of Practice - Protocol on Data Access and Confidentiality (UK)	Download
sdcApp Reference Manual	Download
Statistical Disclosure Control for Microdata: A Practice Guide	Download
Statistical Disclosure Control for Microdata: A Practice Guide - Case Study Data and R Script	(856.35 KB)
Statistical Disclosure Control for Microdata: Theory	Download

Tools

Package SdcMicro	Download
------------------	----------

۱،۲ اصول بی‌نام‌سازی

بی‌نام‌سازی یک مجموعه ریزداده شامل حذف یا اصلاح متغیرهای شناسایی آن است. "معمولاً یک متغیر شناسایی، متغیری است که ویژگی یک شخص را توصیف می‌کند که قابل مشاهده است و ثبت شده است (شماره شناسایی و غیره)، یا به‌طور کلی می‌تواند برای افراد دیگر شناخته شود." (راهنمای Argus μ)

متغیرهای شناسایی عبارتند از:

- **شناسه‌های مستقیم**، که متغیرهایی مانند نام، آدرس یا شماره کارت شناسایی هستند. این شناسه‌ها اجازه شناسایی مستقیم یک پاسخ‌دهنده را می‌دهند اما برای اهداف آماری یا تحقیقاتی مورد نیاز نیستند و بنابراین باید از مجموعه داده منتشر شده، حذف شوند؛
- **شناسه‌های غیرمستقیم**، ویژگی‌هایی هستند که ممکن است بین چندین پاسخ‌دهنده مشترک باشد و ترکیب آن‌ها می‌تواند منجر به شناسایی مجدد یکی از آنها شود. برای مثال، اگر تنها یک فرد از آن جنس، سن و حرفه خاص در آن منطقه زندگی کند، ترکیبی از متغیرهایی مانند منطقه محل سکونت، سن، جنس و حرفه، مشخص‌کننده خواهد بود. چنین متغیرهایی برای اهداف آماری مورد نیاز هستند و نباید از فایل‌های داده منتشر شده، حذف شوند. بی‌نام‌سازی داده‌ها شامل تعیین این‌که کدام متغیرها شناسه‌های بالقوه هستند (براساس



قضاوت شخصی) و اصلاح ویژگی این متغیرها برای کاهش خطر شناسایی مجدد تا سطح قابل قبولی است. چالش این است که امنیت را به حداکثر برسانیم و در عین حال از دست دادن اطلاعات را به حداقل برسانیم.

خطر افشا نه تنها به وجود متغیرهای شناسایی در مجموعه داده بستگی دارد، بلکه به موارد زیر نیز بستگی دارد:

- **وجود یک مُخِل (نفوذگر)^۱**، که به نوبه خود به مزایای بالقوه‌ای که این مُخِل (نفوذگر) از شناسایی مجدد به دست می‌آورد، بستگی دارد. برای برخی از انواع داده‌ها مانند داده‌های تجاری، انگیزه نفوذگر می‌تواند بالا باشد. برای انواع دیگر، مانند پیمایش‌های خانوار در کشورهای در حال توسعه، انگیزه معمولاً بسیار پایین‌تر است، زیرا در شناسایی مجدد پاسخ‌دهندگان، سود کمی وجود دارد؛
- **هزینه شناسایی مجدد. هر چه هزینه بالاتر باشد، سود کمتری برای یک مُخِل (نفوذگر) خواهد داشت.**

برای محاسبه این پارامترها، یک سناریوی افشا باید به عنوان اولین گام در فرآیند بی‌نام‌سازی تعریف شود. سناریوها را می‌توان در دو دسته طبقه‌بندی کرد:

- **سناریوهای همسایه مزاحم^۲**. در این سناریوها فرض می‌شود که نفوذگر، اطلاعات کافی در مورد یک واحد یا واحدها دارد؛ این اطلاعات از دانش شخصی او ناشی می‌شود. به عبارت دیگر، مزاحم متعلق به دایره آشنایان یک واحد آماری است؛
- **سناریوهای بایگانی خارجی**. چنین سناریوهایی بر این فرض استوار است که نفوذگر می‌تواند ثبت‌های متعلق به مجموعه داده توزیع شده را به ثبت‌هایی از مجموعه داده یا رجیستر موجود دیگری که حاوی شناسه‌های مستقیم است، پیوند دهد. نفوذگر این کار را با استفاده از متغیرهای شناسایی موجود در هر دو مجموعه داده به عنوان کلیدهای ادغام (به عبارت دیگر، تطبیق داده‌ها) انجام می‌دهد. مفروضات محافظه‌کارانه اغلب برای تعریف بدترین سناریو مطرح می‌شود.

هنگام تولید فایل‌های ریزداده، همیشه باید دیدگاه کاربر را در نظر داشت. این اساسی است که فایل منتشر شده

1. Intruder

2. Nosy neighbor scenarios



الزامات محقق را برآورده کند. هم محتوای اطلاعاتی و هم انتخاب روش‌های حفاظتی باید تا حد امکان بر روی نیازهای کاربر متمرکز شود. دانستن این که کاربر می‌خواهد چه تحلیل آماری انجام دهد به تصمیم‌گیری در مورد استراتژی بی‌نام‌سازی کمک می‌کند.



منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

Anonymisation: managing data protection risk code of practice	Download
Handbook on Statistical Disclosure Control (Version 1.2)	Download
National Statistics Code of Practice - Protocol on Data Access and Confidentiality (UK)	Download
National Statistics Code of Practice - Protocol on Data Matching (UK)	Download
Principles and Guidelines on Confidentiality Aspects of Data Integration Undertaken for Statistical or Related Research Purposes	Download
sdcApp Reference Manual	Download
Statistical Disclosure Control for Microdata: A Practice Guide	Download
Statistical Disclosure Control for Microdata: Theory	Download
μ-ARGUS Manual (version 4.2)	Download

Tools

Package SdcMicro	Download
------------------	----------

۲،۲ اندازه‌گیری خطر افشا

نقض محرمانگی زمانی رخ می‌دهد که یک واحد آماری دوباره شناسایی شود و مقادیر متغیرهای حساس فاش شوند. چندین روش برای اندازه‌گیری افشا به عبارت دیگر، خطر شناسایی مجدد پیشنهاد شده است، اما هیچ یک از آنها به عنوان بهترین روش پذیرفته نشده است.

با این حال، اندازه‌گیری کمی خطر ضروری است. از آنجایی که خطر افشا را نمی‌توان به صفر کاهش داد، چنین اقدامی به معنای اتخاذ یک قانون آستانه برای تعیین ایمن بودن انتشار یک مجموعه داده است.

معیارهای ریاضی خطر شناسایی مجدد را می‌توان به صورت زیر طبقه‌بندی کرد:



- **معیارهای فردی**، که خطر را در هر ثبت اندازه‌گیری می‌کند. معمولاً با استفاده از احتمال شناسایی مجدد صحیح یک واحد یا با استفاده از منحصر به فرد بودن و نادر بودن در نمونه یا جمعیت بیان می‌شود؛
- **معیارهای سراسری**، که خطر را برای کل فایل اندازه‌گیری می‌کند. معمولاً با استفاده از تعداد مورد انتظار شناسایی مجدد صحیح بیان می‌شود. معیارهای سراسری خطر را می‌توان با ترکیب معیارهای فردی به‌دست آورد.

مزیت معیار خطر فردی این است که فقط آن دسته از ثبت‌هایی که برای یک آستانه خطر معین، ناامن به‌نظر می‌رسند باید به صورت محلی محافظت شوند، در حالی که یک معیار سراسری شامل حفاظت از کل فایل است.

فرض کنید K تعداد ترکیبات در جامعه P باشد که با جدول‌بندی متقاطع مجموعه معینی از متغیرهای کلیدی به‌دست می‌آید. ترکیبی از مقادیر مشاهده شده بر روی یک واحد نمونه را با $K, k=1, \dots, K$ نشان دهید. هر ترکیب K ، خطر شناسایی مجدد خود را دارد. همه ثبت‌هایی که با ترکیب یکسان مشخص می‌شوند، خطر شناسایی مجدد یکسانی دارند. فرض کنید f_k تعداد فراوانی ثبت‌های موجود در نمونه‌ای باشد که همان ترکیب K از متغیرهای کلیدی را ارائه می‌کند، و اجازه دهید F_K تعداد فراوانی نسبت به همان ترکیب K در جامعه P باشد.

در مثال زیر، فرض می‌کنیم که ۳ متغیر، شناسه بالقوه هستند: جنسیت (M = مرد، F =زن)، سن، و وضعیت تأهل (M = متاهل؛ N = هرگز ازدواج نکرده). این فایل شامل ۲۵۰۰ مشاهده است.

	sex	age	marital	hhweight
1	M	40	M	4174.71
2	F	39	M	2466.90
3	F	17	N	2818.74
4	F	13	N	3227.81
5	F	5	.	2824.17
6	M	1	.	3509.14
7	F	47	N	2925.86
8	F	16	N	2818.74

فراوانی هر ترکیبی از این ۳ متغیر را محاسبه کرده و موارد زیر را به‌دست می‌آوریم:



	sex	age	marital	fk
1	M	10	N	34
2	M	11	N	30
3	M	12	N	24
4	M	13	N	24
5	M	13	M	1
6	M	14	N	32
7	M	15	N	30
8	M	16	N	27
9	M	17	N	27
10	M	17	M	1
11	M	18	N	28

برای مثال، متوجه شدیم که فایل شامل ۳۴ مورد از پسران ۱۰ ساله است که هرگز ازدواج نکرده‌اند (ترکیب 'M / N' / 10 / N' اما فقط یک پسر متأهل، ۱۳ ساله (ترکیب 'M / 13 / M' است. این ترکیب دارای فرکانس fk of 1 است و بنابراین یک نمونه منحصر به فرد است. از آنجایی که این اطلاعات از یک نمونه پیمایش به دست می‌آید، نمی‌دانیم که آیا این ترکیب در کل جمعیت نیز منحصر به فرد است یا خیر. به عبارت دیگر، FK مشخص نیست و باید تخمین زده شود.

در نمونه‌ای که قرار است منتشر شود، تنها یک زیرمجموعه از تعداد کل ترکیبات K مشاهده خواهد شد و تنها این زیرمجموعه، که مشخصاً $fk > 0$ برای آن وجود دارد، در معرض خطر افشا است.

- **نمونه منحصر به فرد.** یک ثبت به عنوان یک نمونه منحصر به فرد تعریف می‌شود اگر $fk = 1$ باشد، به عبارت دیگر، تنها یک ثبت در فایل وجود دارد که ترکیب K امتیازهای متغیرهای کلیدی را نشان می‌دهد؛
- **نمونه نادر.** اگر fk یک واحد در معرض خطر است؛
- **جمعیت منحصر به فرد.** یک ثبت به عنوان یک جمعیت منحصر به فرد تعریف می‌شود که $fk = 1$ باشد. برای داده‌های سرشماری، یا زمانی که یک ثبت اداری که کل جمعیت را پوشش می‌دهد، FK برای هر K شناخته می‌شود و می‌توان معیار خطر را محاسبه کرد.
- **جمعیت تخمینی منحصر به فرد.** برای پیمایش‌های نمونه، زمانی که FK برای هر K ناشناخته است، چندین معیار مبتنی بر رویه‌های استنتاجی (به عبارت دیگر، مدل‌های خطی لاگ) پیشنهاد شده است.
- **خطر فردی تخمینی.** چندین مدل احتمالی در سال‌های اخیر برای تخمین خطر شناسایی مجدد تحت یک سناریوی ثبت خارجی پیشنهاد شده است. این رویکرد شامل تخمین تعداد مورد انتظار شناسایی مجدد براساس وزن طراحی نمونه‌گیری است. با توجه به ویژگی‌هایشان،



این روش‌ها برای پیمایش‌های نمونه داده‌های اجتماعی زمانی مناسب هستند که متغیرهای کلیدی مربوط به متغیرهای طبقه‌بندی هستند که اغلب برای تعریف لایه‌ها در مرحله طراحی نمونه‌گیری استفاده می‌شوند.

- **ثبت رویه پیوند.** هنگامی که متغیرهای کلیدی پیوسته هستند، همان‌طور که در مورد ریزداده‌های کسب و کار وجود دارد، رویکرد کاملاً متفاوتی برای معیارهای خطر باید اتخاذ شود. در واقع، ایده ترکیب نمرات روی متغیرهای کلیدی بی معنی است. در این زمینه، هر مشاهده‌ای می‌تواند در تئوری، منحصر به فرد باشد. مفهوم نادر بودن در جمعیت را می‌توان به نادر بودن در همسایگی ثبت ترجمه کرد و ما باید به دنبال مشاهدات دور از دسترس باشیم. یکی از راه‌های اندازه‌گیری نادر بودن در همسایگی، از طریق پیوند ثبت است.

اگرچه همه این معیارها معیارهای فردی خطر هستند، اما می‌توان توابی را استخراج کرد تا آنها را به یک معیار سراسری خطر افشا برای کل فایل تبدیل کند، برای مثال، با استفاده از تعداد مورد انتظار شناسایی مجدد صحیح در فایل. تعداد مورد انتظار شناسایی مجدد و میزان شناسایی مجدد را می‌توان به ترتیب به عنوان معیارهای مطلق و نسبی افشا به دست آورد.

توجه به فایل‌های سلسله مراتبی^۱

پیمایش‌های اجتماعی اغلب اطلاعات یکسانی را برای هر یک از اعضای خانوار جمع‌آوری می‌کنند، و این اطلاعات معمولاً در ثبتی ذخیره می‌شود که به خانوار اشاره دارد. هنگام تعریف خطر شناسایی مجدد، مهم است که این وابستگی بین واحدها در نظر گرفته شود؛ در واقع، شناسایی مجدد یک فرد در یک گروه ممکن است بر احتمال افشای همه اعضای آن تأثیر بگذارد. بنابراین ما خطر خانوار را به عنوان احتمال شناسایی مجدد حداقل یک فرد در خانوار تعریف می‌کنیم.







منابع مرتبط به اشتراک گذاشته شده است:

1. Hierarchical files



Related Resources

Documents

- | | |
|--|--|
|  Disclosure risk assessment in statistical microdata protection via advanced record linkage |  Download |
|  Handbook on Statistical Disclosure Control (Version 1.2) |  Download |
|  μ-ARGUS Manual (version 4.2) |  Download |

Tools

- | | |
|--|--|
|  Package SdcMicro |  Download |
|--|--|



۳,۲ کاهش خطر افشا

روش‌های محدودیت افشای آماری را می‌توان در ۲ دسته طبقه‌بندی کرد:

- **روش‌های مبتنی بر کاهش داده‌ها.** چنین روش‌هایی تعداد افرادی را در نمونه/جمعیت که ویژگی‌های شناسایی یکسان یا مشابه ارائه شده توسط واحد آماری مورد بررسی را دارند، افزایش می‌دهد. این روش‌ها تمایل دارند از حضور افراد منحصر به فرد یا نادر قابل تشخیص اجتناب کنند؛
- **روش‌های مبتنی بر بهم‌ریختگی داده‌ها^۱.** چنین روش‌هایی از دو طریق به حفاظت داده‌ها دست می‌یابند. اول، اگر داده‌ها اصلاح شوند، شناسایی مجدد با استفاده از پیوند ثبت^۲ یا الگوریتم‌های تطبیق دشوارتر و نامطمئن‌تر است. دوم، حتی زمانی که یک نفوذگر می‌تواند یک واحد را دوباره شناسایی کند، نمی‌تواند مطمئن باشد که داده‌های افشا شده با داده‌های اصلی مطابقت دارند.

یک راه حل جایگزین، تولید ریزداده مصنوعی^۳ است.

ریزداده مصنوعی	بهم‌ریختگی داده	کاهش داده
----------------	-----------------	-----------

کاهش داده

حذف متغیرها

اولین کاربرد آشکار این روش حذف شناسه‌های مستقیم از فایل داده است. یک متغیر زمانی باید حذف شود که به شدت حساس شناسایی شود و هیچ روش حفاظتی دیگری نمی‌توان اعمال کرد. یک متغیر همچنین می‌تواند حذف شود زمانی که برای استفاده عموم بسیار حساس باشد یا برای اهداف تحلیلی نامربوط باشد. برای مثال، اطلاعات

1. Data perturbation
2. Record linkage
3. Synthetic microdata



مربوط به نژاد، مذهب، HIV و غیره ممکن است در یک فایل استفاده عموم منتشر نشوند، اما می‌تواند در یک فایل دارای مجوز منتشر شوند.

حذف سوابق

حذف سوابق می‌تواند به عنوان یک اقدام جدی برای حفاظت از داده‌ها، زمانی که واحد قابل شناسایی باشد، علی‌رغم استفاده از سایر تکنیک‌های حفاظتی، استفاده شود. برای مثال، در مجموعه داده‌های پیمایش سازمانی، یک شرکت معین ممکن است تنها شرکت متعلق به یک صنعت خاص باشد. در این مورد، ممکن است ترجیح داده شود که این سابقه خاص به جای حذف متغیر "صنعت" از همه سوابق حذف شود. از آنجایی که تا حد زیادی بر ویژگی‌های آماری داده‌های منتشر شده تأثیر می‌گذارد، در صورت امکان باید از حذف سوابق اجتناب شود.

هنگامی که سوابقی که باید حذف شوند براساس طراحی نمونه‌گیری انتخاب می‌شوند، این روش را نمونه‌گیری فرعی می‌نامند. زمانی که ماتریس اصلی نشان دهنده داده‌های سرشماری باشد، نمونه‌گیری نامیده می‌شود.

رمزگذاری مجدد سراسری

رمزگذاری مجدد سراسری شامل تجمیع مقادیر مشاهده شده در یک متغیر در دسته‌های از پیش تعریف شده است (برای مثال، رمزگذاری مجدد سن به گروه‌های سنی ۵ ساله، یا تعداد کارکنان در ۳ دسته: کوچک، متوسط و بزرگ). این روش برای متغیرهای عددی پیوسته یا گسسته کاربرد دارد. تمام ثبت‌های موجود در فایل داده را تحت تأثیر قرار می‌دهد.

هنگام کار با متغیرهای طبقه‌بندی شده (یا دسته‌بندی شده عددی)، روش رمزگذاری مجدد سراسری، دسته‌های مشابه یا مجاور را یکجا جمع می‌نماید.

برای مثال، متغیر «وضعیت تأهل» را در نظر بگیرید که اغلب در دسته‌های زیر مشاهده می‌شود: مجرد، متأهل، جدا شده، مطلقه، بیوه. فراوانی نمونه از دسته‌های طبقه‌بندی شده ممکن است کم باشد، به خصوص زمانی که با سایر متغیرها جدول بندی شود. دو دسته مجاور، جدا شده و مطلقه، را می‌توان در یک دسته واحد به نام 'جدا شده



یا مطلقه' ادغام کرد. فراوانی ترکیب‌ها در این دسته جدید بیشتر از نسبت جدا شده و مطلقه به‌طور جداگانه خواهد بود. دسته‌هایی که می‌توانند با هم ترکیب شوند به کاربرد داده‌ها و همچنین کنترل آماری فراوانی‌ها بستگی دارد. این روش را می‌توان برای متغیرهای کلیدی مانند کدهای جغرافیایی نیز به کار برد تا اثر شناسایی آنها کاهش یابد.

کُد نویسی بالا و پایین

کُدگذاری بالا و پایین یک مورد خاص از کُدگذاری مجدد سراسری است که می‌تواند برای متغیرهای دسته‌بندی عددی یا ترتیبی اعمال شود. متغیرهای «حقوق» و «سن» دو نمونه هستند. بالاترین مقادیر این متغیرها معمولاً بسیار نادر هستند و بنابراین قابل شناسایی هستند. کُدگذاری برتر در آستانه‌های خاص، دسته‌بندی‌های جدیدی مانند 'حقوق ماهانه بالاتر از ۶۰۰۰ دلار' یا 'سن بالای ۷۵ سال' را معرفی می‌کند و سایر مقادیر مشاهده شده را بدون تغییر می‌گذارد. همان استدلال اعمال شده برای مقادیر کوچکتر مشاهده شده، کُدگذاری پایین را تعریف می‌کند. هنگام برخورد با متغیرهای دسته‌بندی ترتیبی، یک دسته بالا یا پایین به ترتیب با جمع‌آوری دسته‌های «بالاترین» یا «کوچک‌ترین» تعریف می‌شود.

کوچک‌سازی موضعی

کوچک‌سازی موضعی شامل جایگزینی مقدار مشاهده شده یک یا چند متغیر در یک ثبت خاص با یک مقدار از دست رفته است. کوچک‌سازی موضعی مخصوصاً برای تنظیم متغیرهای کلیدی طبقه‌بندی شده و زمانی که ترکیبی از امتیازات در چنین متغیرهایی در خطر است، مناسب می‌باشد. در این مورد، کوچک‌سازی موضعی شامل جایگزینی یک مقدار مشاهده شده در ترکیب با یک مقدار از دست رفته است. این روش محتوای اطلاعاتی ترکیبات نادر را کاهش می‌دهد و در نتیجه، تعداد دفعات ثبت‌های حاوی همان ترکیب (اصلاح شده) افزایش می‌یابد. برای مثال، فرض کنید ترکیب «وضعیت تأهل=بیوه؛ سن=۱۷» یک جمعیت منحصر به فرد است. اگر اطلاعات مربوط به سن کوچک‌سازی شود، ترکیب 'وضعیت تأهل=بیوه؛ سن=مفقود' دیگر قابل شناسایی نخواهد بود. از طرف دیگر، می‌توان اطلاعات مربوط به وضعیت تأهل را نیز کوچک‌سازی کرد. بنابراین یک معیار برای تصمیم‌گیری این‌که کدام متغیر در یک ترکیب مخاطره‌آمیز باید به صورت موضعی کوچک‌سازی شود، ضروری است. بدیهی است که معیار اولیه، به حداقل رساندن تعداد کوچک‌سازی‌های موضعی است. به عنوان مثال، مقادیر متغیرهای کلیدی، 'جنس = زن، وضعیت تأهل = بیوه، سن = ۱۷، شغل = دانش آموز' را در یک واحد در نظر بگیرید. هر دو ترکیب 'وضعیت تأهل=بیوه؛ سن=۱۷' و 'جنس=زن؛ وضعیت تأهل=بیوه؛ شغل=دانشجو' واحد را مشخص می‌کنند و می‌تواند جمعیت



منحصر به فرد یعنی ترکیباتی در معرض خطر باشند. برای به حداقل رساندن تعداد کوچک‌سازی‌های موضعی، می‌توان متغیر «وضعیت تأهل» را با مقادیر از دست رفته جایگزین کرد، به طوری که هر دو ترکیب به طور همزمان با استفاده از یک کوچک‌سازی موضعی واحد محافظت می‌شوند. اگر متغیرها به طور مستقل در نظر گرفته شوند، دو کوچک‌سازی موضعی مورد نیاز است. معیار دیگری را می‌توان با توجه به اندازه‌گیری از دست دادن اطلاعات تعریف کرد (برای مثال، مقدار حداقل کردن یک شاخص آنتروپی-تن‌سنجی - می‌تواند برای کوچک‌سازی موضعی انتخاب شود). علاوه بر این، وزن‌های کوچک‌سازی را می‌توان به متغیرهای کلیدی اختصاص داد تا کوچک‌سازی موضعی را به سمت متغیرهای کمتر مهم هدایت کند. کوچک‌سازی موضعی همچنین به یک معیار انتخاب برای ثبت‌ها نیاز دارد. پیش‌تر به چندین قاعده برای تعریف ثبت در معرض خطر اشاره شد؛ کوچک‌سازی موضعی را می‌توان فقط برای ثبت‌های پرخطر اعمال کرد، به عنوان مثال، ثبت‌هایی که حاوی ترکیبات در معرض خطر هستند.

اغتشاش داده

ریزتجمیع^۱

ریزتجمیع یک تکنیک بهم‌ریختگی است که برای اولین بار توسط اداره آمار اروپا به عنوان یک روش افشای آماری برای متغیرهای عددی پیشنهاد شد. ایده این است که یک مقدار مشاهده شده با میانگین محاسبه شده بر روی گروه کوچکی از واحدها (تجمع کوچک یا ریزتجمیع)، از جمله واحد مورد بررسی، جایگزین شود. واحدهای متعلق به همان گروه در فایل منتشر شده با همان مقدار نشان داده می‌شوند. گروه‌ها دارای حداقل تعداد از پیش تعریف شده K واحد هستند. حداقل مقدار پذیرفته شده K عدد سه است. برای یک K معین، مسئله شامل تعیین پارتیشن کل مجموعه واحدها در گروه‌های حداقل K واحد (K پارتیشن)، به حداقل رساندن از دست دادن اطلاعات است که معمولاً به عنوان از دست دادن تغییرپذیری بیان می‌شود. بنابراین، گروه‌ها براساس معیار حداکثر شباهت بین واحدها ساخته می‌شوند. مکانیسم ریزتجمیع با اطمینان از وجود حداقل K واحد با همان مقدار در فایل داده، حفاظت از داده‌ها را به دست می‌آورد.

هنگامی که ریزتجمیع به طور مستقل برای مجموعه‌ای از متغیرها اعمال می‌شود، این روش رتبه‌بندی فردی نامیده می‌شود. هنگامی که همه متغیرها به طور همزمان برای هر گروه میانگین می‌شوند، این روش ریزتجمیع چند متغیره نامیده می‌شود.

1. Micro-aggregation



ساده‌ترین راه برای گروه‌بندی ثبت‌ها قبل از تجمیع آنها، مرتب کردن واحدها براساس شباهت آنها و مقادیر حاصل از این معیار، و تجمیع واحدهای متوالی در گروه‌های اندازه ثابت است. تنظیم اندازه در نهایت برای اولین یا آخرین گروه مورد نیاز است. برای ریزتجمع تک متغیره، معیار مرتب‌سازی ممکن است خود متغیر باشد.

برای ریزتجمع چند متغیره، شباهت را می‌توان به عنوان معیاری برای متغیرهای مشاهده شده استفاده کرد یا برای افزایش اثربخشی روش، می‌توان آن را ترکیبی از متغیرها تعریف کرد. برای مثال، اولین جزء اصلی یا مجموع مقادیر امتیاز Z در امتداد مجموعه متغیرها می‌تواند معیاری برای ریزتجمع با اندازه ثابت باشد.

ریزتجمع چند متغیره بسیار محافظت‌کننده‌تر از رتبه‌بندی فردی در نظر گرفته می‌شود، زیرا این روش تضمین می‌کند که حداقل واحدهای k در فایل یکسان هستند (همه متغیرها به‌طور هم‌زمان میانگین می‌شوند)، اما از دست دادن اطلاعات بیشتر است.

مبادله داده‌ها

مبادله داده‌ها در ابتدا به عنوان یک تکنیک بهم‌ریختگی برای ریزداده‌های طبقه‌بندی شده پیشنهاد شد و هدف آن حفاظت از جدول‌بندی ناشی از فایل ریزداده آشفته بود. مبادله داده‌ها شامل تغییر نسبتی از ثبت‌های یک فایل با تبادل مقادیر زیرمجموعه‌ای از متغیرها بین جفت‌های انتخاب شده یا جفت مبادله شده سوابق است.

سطح حفاظت از داده‌ها به سطح بهم‌ریختگی القا شده در داده‌ها بستگی دارد. برای تعیین این‌که کدام متغیرها و ثبت‌ها (نرخ مبادله) باید مبادله شوند، باید معیاری اعمال شود. برای داده‌های طبقه‌بندی، مبادله اغلب برای ثبت‌هایی اعمال می‌شود که نمونه منحصر به فرد یا نمونه نادر هستند، زیرا این رکوردها معمولاً خطرات بیشتری برای شناسایی مجدد دارند.

یافتن مبادله داده‌هایی که حفاظت کافی را در عین حفظ آمار دقیق پایگاه داده اصلی ارائه می‌کنند، عملی نیست. حتی زمانی که لحظه‌های تک متغیره حفظ می‌شوند، مبادله داده‌ها معمولاً داده‌ها را بیش از حد تغییر می‌دهد.

ارسال تصادفی

به عنوان یک تکنیک کنترل افشای آماری، ارسال تصادفی با مبادله آنها براساس یک مکانیسم احتمالی، عدم قطعیت را در مقادیر برخی از متغیرها ایجاد می‌کند. بنابراین ارسال تصادفی می‌تواند به عنوان یک نسخه تصادفی از مبادله داده در نظر گرفته شود. مانند مبادله داده‌ها، حفاظت از داده‌ها به این دلیل حاصل می‌شود که یک نفوذگر



نمی‌تواند مطمئن باشد که آیا مقدار مشخص منتشر شده درست است و بنابراین تطبیق ثبت با شناسه‌های خارجی می‌تواند منجر به عدم تطابق یا طبقه‌بندی نادرست ویژگی شود. این روش برای متغیرهای طبقه‌ای معرفی شده است، اما می‌توان آن را به متغیرهای عددی نیز تعمیم داد.

اضافه کردن پارازیت

افزودن پارازیت شامل افزودن یک مقدار تصادفی ϵ ، با میانگین صفر و واریانس از پیش تعریف شده σ^2 ، به تمام مقادیر متغیری است که باید محافظت شود. به‌طور کلی، روش‌های مبتنی بر اضافه کردن پارازیت، از نظر حفاظت از داده‌ها چندان مؤثر در نظر گرفته نمی‌شوند.

نمونه‌گیری مجدد

نمونه‌برداری مجدد یک روش حفاظتی برای ریزداده‌های عددی است که شامل ترسیم نمونه‌های جایگزین t با n مقدار از داده‌های اصلی، مرتب‌سازی نمونه و میانگین‌گیری مقادیر نمونه‌گیری شده است. سطح حفاظت از داده تضمین شده توسط این روش به‌طور کلی بسیار پایین در نظر گرفته می‌شود.

ریزداده مصنوعی

ریزداده مصنوعی یک روش جایگزین برای حفاظت از داده‌ها است و با استفاده از الگوریتم‌های شبیه‌سازی داده تولید می‌شود. منطق این روش این است که داده‌های مصنوعی با توجه به کنترل اِشْیای آماری مشکلی ایجاد نمی‌کنند زیرا حاوی داده‌های واقعی نیستند اما ویژگی‌های آماری خاصی را حفظ می‌کنند. در ابتدا، Rubin تولید داده‌های مصنوعی را از طریق جانهی‌های متعدد پیشنهاد کرد، در حالی که Feinberg با استفاده از روش‌های boot-strap پیشنهاد کرد. روش‌های بیشتر مانند جانهی چندگانه، نمونه‌گیری Latin hypercube، مدل‌سازی، و توزیع داده‌ها براساس احتمال پیشنهاد شده‌اند.

به‌طور کلی، کاربران با داده‌های مصنوعی راحت نیستند زیرا نمی‌توانند از نتایج تجزیه و تحلیل آماری خود اطمینان داشته باشند. با این حال، این روش می‌تواند به تولید «مجموعه‌های ریزداده آزمایشی» کمک کند، که در آن فایل‌های داده مصنوعی منتشر می‌شوند تا به کاربران اجازه دهد روش‌های آماری خود را برای دسترسی متوالی به ریزداده‌های «واقعی» در یک محفظه داده آزمایش کنند.

منابع مرتبط به اشتراک گذاشته شده است:



Related Resources

Documents

Achieving k-anonymity privacy protection using generalization and suppression	Download
An empirical evaluation of PRAM	Download
Data Privacy Through Optimal k-Anonymization	Download
Efficient multivariate data-oriented microaggregation	Download
Fast Generation of Accurate Synthetic Microdata	Download
Measuring Risk and Utility of Anonymized Data Using Information Theory	Download
Microaggregation for Database and Location Privacy	Download
On the Security of Noise Addition for Privacy in Statistical Databases	Download
Ordinal, Continuous and Heterogeneous k-Anonymity Through Microaggregation	Download
Producing Public-Use Microdata that are Analytically Valid and Confidential	Download
μ -ARGUS Manual (version 4.2)	Download

۴,۲ ارزیابی از دست دادن اطلاعات

بیشتر تکنیک‌های بی‌نام‌سازی شامل کاهش سطح جزئیات در اطلاعات ارائه شده، یا کوچک‌سازی اطلاعات است. بنابراین آنها معمولاً منجر به از دست دادن اطلاعات می‌شوند. چالش پیش‌روی آمارگیر ایجاد تعادل بین اهداف متضاد کاهش خطر افشا و به حداقل رساندن این از دست دادن است.

روش‌های مختلفی برای ارزیابی از دست دادن اطلاعات موجود است. برای داده‌های طبقه‌بندی، این روش‌ها شامل مقایسه مستقیم، مقایسه جداول احتمالی و معیارهای مبتنی بر آنتروپی است. برای داده‌های پیوسته، روش‌ها شامل مقایسه میانگین مربع، میانگین مطلق و میانگین تغییرات می‌شوند.

اطلاعات مربوط به این تکنیک‌ها از منابع زیر در دسترس است:



Related Resources

Documents

A Framework for Evaluating the Utility of Data Altered to Protect Confidentiality	Download
Comparing SDC Methods for Microdata on the Basis of Information Loss and Disclosure Risk	Download
Disclosure Control Methods and Information Loss for Microdata	Download
Distortion Measures for Categorical Data Swapping	Download
Measuring Risk and Utility of Anonymized Data Using Information Theory	Download
Producing Public-Use Microdata that are Analytically Valid and Confidential	Download
The Effect of Microaggregation Procedures on the Estimation of Linear Models: A Simulation Study	Download

۵,۲ ابزارهای کنترل افشای آماری

طی همکاری شبکه بین‌المللی پیمایش‌های خانوار، PARIS21 (سازمان همکاری و توسعه اقتصادی)، اداره آمار اتریش و دانشگاه فناوری Vienna، بانک جهانی به ارائه یک بسته نرم‌افزاری متن باز به نام `sdcMicro` برای کنترل افشای آماری کمک کرده است. این بسته برای استفاده با نرم‌افزار آماری متن باز R ارائه شده است: از شبکه جامع آرشیو R CRAN در <http://lib.stat.cmu.edu/R/CRAN/> در دسترس است. این بسته شامل الگوریتم‌های متعددی برای ارزیابی و کاهش خطر افشا در ریزداده‌ها است.

برای دسترسی به جعبه ابزار رایگان کنترل افشای آماری (SDCMicro)، به بخش نرم‌افزار مراجعه کنید.

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

Statistical Disclosure Control for Microdata: Theory	Download
--	----------



۶,۲ شیوه‌های بی‌نام‌سازی

هیچ استاندارد بین‌المللی، روش‌های بی‌نام کردن داده‌ها، سطوح قابل قبول خطر یا معیارهای توصیه شده از دست دادن اطلاعات را تعریف نمی‌کند. میزان و نوع حفاظت مورد نیاز برای هر مجموعه داده، بسته به حساسیت و «ارزش تجاری» محتوا، و هر محیط خاص قانونی و فرهنگی، ویژه است. بنابراین مستند کردن برخی از شیوه‌ها مفید است. با این حال، این کار آسانی نیست، زیرا آژانس‌هایی که مجموعه داده‌های خود را بی‌نام می‌سازند، در مورد روش‌های پیاده‌سازی شده و سطوح خطر در داده‌هایی که منتشر می‌کنند، ارتباط زیادی برقرار نمی‌کنند.

این دسترسی محدود به دانش همراه با فقدان تجربه در استفاده از ابزارها و روش‌ها، اجرای راه‌حل‌های "بهینه" را برای بسیاری از آژانس‌ها دشوار می‌کند. منظور از بهینه، عمل کردن به تعهدات خود در قبال حفاظت از حریم خصوصی و همچنین تعهد به انتشار داده‌های مفید برای نظارت و ارزیابی خط مشی است. به‌منظور پُر کردن این شکاف، در دستورالعمل‌های عملی، بانک جهانی پروژه‌ای را تکمیل کرد که توسط برنامه دوم دانش برای تغییر تأمین شده بود، که به دنبال ایجاد یک پایگاه دانش از طریق آزمایش بر روی مجموعه‌ای متنوع از ریزداده‌ها بود. سپس این دانش به یک راهنمای عملی برای انتشار عمومی ترجمه شد. راهنمای عملی، با مستند کردن تحقیقات انجام شده در بانک جهانی از طریق مقیاس بزرگ ارزیابی تکنیک‌های بی‌نام‌سازی، و با ترجمه این نتایج به دستورالعمل‌های عملی، شکافی مهم را پر می‌کند. این راهنمای عملی در سال ۲۰۱۵ منتشر شد. به‌منظور افزودن روش‌های جدید و همگام شدن با ویژگی‌های جدید موجود در بسته نرم‌افزار متن باز R sdcMicro که راهنماها برای مثال‌های عملی آن استفاده می‌کنند، مرتباً به‌روزرسانی می‌شود.













نسخه فعلی راهنما در <https://sdcpractice.readthedocs.io/en/latest/> موجود است.



منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

 Anonymisation: managing data protection risk code of practice	 Download
 NCHS Staff Manual on Confidentiality	 Download
 sdcApp Reference Manual	 Download
 Statistical Disclosure Control for Microdata: A Practice Guide	 Download
 Statistical Disclosure Control for Microdata: Theory	 Download
 Statistical Policy Working Paper 22 (Second version, 2005) - Report on Statistical Disclosure Limitation Methodology	 Download



۳. کاتالوگ کردن

کاربران علاقه‌مند باید به درستی از وجود و ویژگی‌های مجموعه داده‌های موجود مطلع شوند. بسیاری از کاربران بالقوه، اطلاعات بسیار کمی در مورد مجموعه داده‌های موجود دارند. فراداده خوب باید در دسترس باشد، ترجیحاً در قالب یک کاتالوگ آنلاین قابل جستجو باشد.

هدف کاتالوگ ریزداده فراهم کردن دسترسی آسان به داده‌ها و اسناد در قالبی که برای کاربران راحت‌تر است، می‌باشد. کاتالوگ پیمایش، ابزارهایی را برای موارد زیر فراهم می‌کند:

- یافتن مجموعه داده مناسب‌ترین با نیازهای کاربر و می‌تواند زمانی که تعداد فایل‌های ریزداده کم است، ساده باشد. اما، با افزایش تعداد فایل‌ها، ابزاری که بتواند فایل‌های داده را در سطح متغیر جستجو کند، ضروری می‌شود؛
- ارزیابی اطلاعاتی که برای اطمینان از سازگاری با نیازهای محقق شناسایی شده است، برای مثال، سراسری بودن، مفاهیم و تعاریف به کار رفته در پیمایش. این نقش توسط فراداده مورد استفاده برای مستندسازی فایل پشتیبانی می‌شود؛
- دسترسی به داده‌ها با استفاده از استخراج و/یا نوعی سیستم تحویل. معمولاً، چنین فایل‌هایی را می‌توان از طریق یک وب‌سایت/پورتال و یک سرور FTP تحویل داد یا از چنین ابزارهایی در NSO برای ساخت یک CD/DVD جهت تحویل داده‌ها استفاده کرد؛
- استفاده کردن از داده‌ها. چیزی به نام ابزار واحدی برای پژوهشگران جهت انجام کارهای تحلیلی خود وجود ندارد. محققان داده‌های موجود در قالب‌های مختلف را ترجیح می‌دهند تا بتوانند از ابزارهای دلخواه خود استفاده کنند. به‌طور معمول، این فرمت‌ها شامل فرمت‌های SAS، STATA، SPSS و ASCII می‌شوند.

در این بخش، ویژگی‌های اصلی یک سیستم کاتالوگ کردن خوب داده از دیدگاه کاربر داده و مدیر کاتالوگ شرح داده می‌شود و ابزارهای رایگان برای توسعه و مدیریت کاتالوگ پیمایش آنلاین ارائه می‌شود.

۱,۳ ویژگی‌های یک کاتالوگ خوب

یک کاتالوگ خوب چیست؟



دیدگاه کاربر

- **مطابق با استاندارد بین‌المللی فراداده است.** استانداردهای بین‌المللی فراداده XML مانند DDI و Dublin Core به‌طور قابل توجهی تولید و نگهداری چنین کاتالوگ‌هایی را تسهیل می‌کنند؛
- **در وب برای تسهیل یافتن آن است؛**
- **فراداده‌های غنی ارائه می‌دهد، از جمله، در سطح متغیر.** کاتالوگ‌های پیمایش زمانی مرتبط و قدرتمند می‌شوند که فراداده پیمایش نه تنها شرح مفصلی از خود پیمایش (با اطلاعاتِ عنوان، محقق اولیه، نمونه‌گیری، تاریخ جمع‌آوری داده‌ها، موضوعات، پوشش جغرافیایی، و غیره)، بلکه هر متغیر را با اطلاعاتی در مورد نام و برچسب متغیر، دسته‌بندی‌ها، سؤال تحت‌اللفظی، دستورالعمل‌های مصاحبه‌کننده، و تعاریف ارائه می‌دهد. با استفاده از استاندارد فراداده DDI و ابزارهای شبکه بین‌المللی پیمایش‌های خانوار، به ویژه جعبه ابزار مدیریت ریزداده شبکه بین‌المللی پیمایش‌های خانوار و برنامه رایگان آرشیو داده ملی (NADA)، می‌توان یک کاتالوگ سطح متغیر را نسبتاً آسان ایجاد کرد؛
- **قابل جستجو در تمام زمینه‌های مرتبط با مطالعه است.** در چارچوب DDI، بدین معنا است که کاتالوگ باید در هر دو قسمت شرح مطالعه (عنوان، سال، کشور و سازمان) و متغیر (نام متغیر، برچسب متغیر و برچسب ارزش متغیر) قابل جستجو باشد. کاتالوگ باید قابلیت‌های جستجوی متن کامل کاربر پسند را ارائه دهد؛
- **اطلاعات واضحی در مورد خط‌مشی‌ها و روش‌های دسترسی به داده‌ها ارائه می‌دهد؛**
- **فهرست و دسترسی مستقیم به مواد مرجع (پرسشنامه‌ها، راهنماها، گزارش‌ها) را فراهم می‌کند؛**
- **شامل "جستجو براساس موضوع" مطابق با طبقه‌بندی استاندارد موضوعات است.** برای تسهیل تبادل اطلاعات بین کاتالوگ‌ها، جامعه آرشیو داده‌ها، گنجینه‌ای (واژه‌نامه) را برای توصیف موضوعات تحت پوشش مجموعه داده‌های فهرست شده در کاتالوگ مربوطه خود ایجاد کرده است. واژه‌نامه، مجموعه‌ای از اصطلاحات یا مفاهیمی است که برای توصیف مواردی مانند مجموعه داده‌ها، متغیرها، کتاب‌ها و غیره استفاده می‌شود. اصطلاحات معمولاً به صورت سلسله‌مراتبی سازماندهی می‌شوند و اصطلاحات گسترده‌تر عبارت‌اند از والدین تا اصطلاحات جزئی‌تر. معمولاً یک واژه‌نامه شامل اصطلاحات و مترادف‌های موازی است که به کاربران امکان می‌دهد آنچه را که به دنبال آن هستند



پیدا کنند، حتی زمانی که از اصطلاحات ترجیحی استفاده نمی‌کنند. بسیاری از آرشيوها هنگام اضافه کردن کلمات کلیدی، در سطح مطالعه، یا مفاهیم، در سطح متغیر از واژه‌نامه استفاده می‌کنند. استفاده از واژه‌نامه با حصول اطمینان از این که عبارات یکسان در هنگام توصیف موارد یکسان انتخاب شده‌اند، باعث ایجاد ثبات می‌شود. علاوه بر این، اگر کاربران هنگام جستجوی داده‌ها به واژه‌نامه دسترسی داشته باشند، احتمال بیشتری وجود دارد که از اصطلاحات و مفاهیمی استفاده کنند که مرتبط‌ترین فهرست بازدیدها را برمی‌گرداند. نمونه‌ای از استفاده از واژه‌نامه، کاتالوگی است که توسط شورای آرشيو داده‌های علوم اجتماعی اروپا (CESSDA)، یک سازمان چتر برای آرشيو داده‌های علوم اجتماعی در سراسر اروپا، نگهداری می‌شود؛

- قادر به **نمایش سریع نتایج جستجوها**، حتی در کاتالوگ‌های بزرگ است. این به معنای یک سیستم نمایه‌سازی کارآمد است؛
- وسیله‌ای برای **مقایسه اقلام کاتالوگ** ارائه دهد که در مقایسه متغیرها در پیمایش‌های استاندارد شده یا پیمایش‌هایی که چندین نسخه از یک مطالعه برای آنها بارگذاری شده است، مفید می‌باشد؛
- **اطلاعاتی در مورد سیاست‌های دسترسی** که به راحتی قابل مشاهده است برای هر مطالعه ارائه دهد. برای مثال، نشان می‌دهد که آیا ریزداده‌ها در دسترس هستند یا خیر، و اگر چنین است، دستورالعمل‌های روشنی در مورد نحوه به‌دست آوردن آنها ارائه می‌دهد؛
- برای کاربران، **On-screen help** خوبی ارائه دهد؛
- ابزاری برای **پیوند اقلام کاتالوگ به منابع وبسایت خارجی** و همچنین امکان پیوست کردن اطلاعات بیشتر، مانند ارجاعات کتابنامه به انتشاراتی که از مطالعه استفاده کرده‌اند، ارائه دهد.



دیدگاه مدیر

- **محیطی امن** برای ذخیره و به اشتراک‌گذاری داده‌ها و فراداده‌ها فراهم می‌کند؛
- ابزارهایی را برای **مدیریت فرآیند دسترسی ریزداده فراهم می‌کند** که محدوده‌ای از تأیید اتومات ریزداده‌ها بدون محدودیت دسترسی، تا سیستم‌هایی برای مدیریت و پردازش برنامه‌هایی که بررسی آنها قبل از اعطای دسترسی لازم است، را شامل می‌شود؛
- راه حلی برای **به اشتراک‌گذاری فایل‌های استفاده عموم و فایل‌های دارای مجوز** ارائه می‌دهد؛
- **وسپله‌ای امن برای به اشتراک‌گذاری ریزداده‌ها و اسناد فراهم می‌کند**، بنابراین دسترسی کاربر نهایی را افزایش می‌دهد؛
- **اطلاعات مربوط به کاربردها و کاربران کاتالوگ را جمع‌آوری می‌کند**: داده‌های دانلود شده، و، در صورت لزوم، هدفی که داده‌ها برای آن استفاده می‌شوند. چنین سوابقی برای حامیان مطالعات مفید است، زیرا ابزاری برای سنجش استفاده از ریزداده‌ها فراهم می‌کند. چنین سوابقی برای کاربران نیز مفید است، زیرا اطمینان می‌دهند که کاربران هنگام انتشار نسخه‌های جدید داده‌ها یا تغییراتی در مطالعاتی که دانلود کرده‌اند، مطلع می‌شوند.

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Links

 [HASSET - The Humanities and Social Science Electronic Thesaurus \(UK Data Archive\)](#)  [Download](#)

۲,۳ ابزارهای کاتالوگ کردن ریزداده‌ها

شبکه بین‌المللی پیمایش‌های خانوار، اتخاذ استاندارد طرح مستندسازی داده‌ها (DDI)، کاتالوگ کردن و انتشار



ریز داده‌های پیمایش را توصیه می‌کند.

شبکه بین‌المللی پیمایش‌های خانوار و بانک جهانی یک برنامه رایگان، متن باز و مطابق با طرح مستندسازی داده‌ها (DDI) برای کاتالوگ کردن و انتشار ریز داده‌ها ارائه داده‌اند. این برنامه تکمیل کننده ویرایشگر فراداده DDI تخصصی، شبکه بین‌المللی پیمایش‌های خانوار است.

برای دسترسی به برنامه نرم‌افزار کاتالوگ کردن (آرشیو داده‌های ملی یا NADA)، به بخش نرم‌افزار مراجعه کنید.



۴. انتشار ریز داده‌ها

تولیدکنندگان داده با تقاضای روزافزون اطلاعات مرتبط و دقیق مواجه هستند. فراهم کردن دسترسی محققین به ریزداده مزایای بالقوه زیادی دارد: این امر تنوع و کیفیت تحلیل‌ها را تقویت می‌کند، استفاده از داده‌های موجود را گسترش می‌دهد و بازده سرمایه‌گذاری‌های جمع‌آوری داده‌ها را افزایش می‌دهد.



Dissemination of Microdata Files - Principles, Procedures and Practices

Olivier Dupriez and Ernie Boyko - August 2010

In all countries, data producers face growing demand for microdata. Determining the best way to disseminate these data is a challenge. Formal policies and procedures defining the conditions of access to microdata must be formulated. This guide provides an overview of such policies and procedures, and documents existing best practices.

[Download \(2.05 MB\)](#)



Diffusion des fichiers de microdonnées - Principes, procédures et pratiques

Olivier Dupriez and Ernie Boyko - August 2010

Les producteurs de données de tous pays sont confrontés à une demande croissante de microdonnées. Décider de la meilleure façon de diffuser ces données constitue un véritable défi. Cela implique l'établissement de politiques et de procédures définissant les conditions d'accès aux microdonnées. Le présent document contient une description générale de ces politiques et de ces procédures, et recense les bonnes pratiques en la matière.

[Download](#)



Difusión de archivos de microdatos - Principios, procedimientos y prácticas

Olivier Dupriez y Ernie Boyko - Agosto de 2010

Los productores de datos de todos los países se enfrentan a una creciente demanda de microdatos. Decidir cuál es la mejor forma de difundir estos datos constituye un verdadero desafío. En este contexto, la difusión de microdatos implica la creación de políticas y procedimientos formales que definan las condiciones de acceso a los microdatos. El presente documento contiene una descripción general de estas políticas y procedimientos e identifica buenas prácticas existentes en este ámbito.

[Download](#)



بث و نشر ملفات البيانات الجزئية - مبادئ وأنظمة وممارسات

أوليفيه دوبريز و إرنی بویکو - 2010 أغسطس / آب

في جميع الدول، منتجي البيانات يواجهون طلب متزايد على البيانات الجزئية. تحديد ووضع أفضل الطرق لنشر هذه البيانات يعتبر تحدي. السياسات والإجراءات الرسمية التي تحدد شروط الوصول والنفاذ للبيانات الجزئية والحصول عليها يجب أن تصاغ وتكون واضحة. يقدم هذا الدليل نظرة عامة عن هذه السياسات والإجراءات، ويحتوي هذا الدليل كذلك على الممارسات الجيدة للتعامل مع البيانات الجزئية.

[Download](#)

با این حال، انتشار ریزداده‌ها همچنین مستلزم هزینه‌ها و خطراتی است که بارزترین آنها خطر افشای اطلاعات محرمانه است. در حالی که تولیدکنندگان داده به خوبی از قدرت و اهمیت ریزداده آگاه هستند، باید این تقاضا را با تعهدات اخلاقی و قانونی برای محرمانه نگه داشتن اطلاعات پاسخگو متعادل کنند. در بیشتر موارد، این یک الزام قانون آمار و حریم خصوصی یک کشور است و توسط پاسخ‌دهندگان هنگام ارائه اطلاعات، فرض می‌شود. آژانس‌های آماری باید اطمینان دهند که اعتماد و اطمینان پاسخ‌دهندگان را حفظ می‌کنند. بدون این اعتماد، همکاری در پیمایش‌ها کاهش می‌یابد و کیفیت آمار آسیب می‌بیند.

تعیین بهترین راه برای انتشار انبوه اطلاعات، یکی از چالش‌های بسیاری است که تولیدکنندگان داده با آن روبه‌رو هستند. روش ارائه دسترسی به فایل‌های ریزداده بیش از ۴۰ سال است که وجود داشته و اخیراً مورد توجه آژانس‌های آماری بین‌المللی قرار گرفته است. پیشنهادات و توصیه‌های ارائه شده در وب‌سایت شبکه بین‌المللی پیمایش‌های خانوار براساس این تجربه است و هدف آن کمک به تولیدکنندگان داده در تدوین خط مشی انتشار و تعیین گزینه‌های فنی و عملی برای به اشتراک گذاشتن داده‌هایشان است. شامل موارد ذیل می‌شود:

- مزایای انتشار ریزداده‌ها؛
- هزینه‌ها و خطرات مرتبط با اشتراک‌گذاری ریزداده‌ها؛



- جایگزین‌های انتشار ریزداده‌ها؛
- مسائل حقوقی مرتبط با اشتراک‌گذاری ریزداده‌ها؛
- مسائل اخلاقی مرتبط با به اشتراک‌گذاری ریزداده‌ها؛
- سطوح مختلف دسترسی به ریزداده‌ها؛
- سیاست قیمت‌گذاری: آیا داده‌ها باید به صورت رایگان ارائه شوند یا با هزینه؟؛
- زمان انتشار داده‌ها: ریزداده‌ها چه زمانی باید منتشر شوند؟؛
- رهنمودهایی برای تدوین خط مشی انتشار.

دستورالعمل‌های ارائه‌شده در این بخش برای کمک به تولیدکنندگان و نگهداری‌کنندگان ریزداده‌ها در ارائه خط‌مشی‌ها و رویه‌های خود برای انتشار فایل‌های ریز داده است. مهم است که چنین سیاست‌ها و رویه‌هایی، رسمی و شفاف باشند. انتشار صحیح ریزداده‌ها نه تنها شامل ارائه داده‌ها و اسناد مربوطه است، بلکه شرایط اجباری برای استفاده کردن از داده‌ها را نیز شامل می‌شود.










در حالی که بیشتر این راهنما، عمومی است، اما اساساً برای استفاده تولیدکنندگان رسمی داده‌ها - ادارات آمار ملی و وزارتخانه‌های مرتبط^۱ - در کشورهای در حال توسعه است. اصطلاح "داده" به‌طور کلی به ریزداده‌های به‌دست آمده از پیمایش‌های نمونه، سرشماری‌ها، و سیستم‌های جمع‌آوری داده‌های اداری اشاره دارد.



منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

 Dissemination of Microdata Files - Principles, Procedures and Practices	 (2.05 MB)
 Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle (5th ed.)	 Download
 Managing and sharing data - Best practice for researchers	 Download
 Managing Statistical Confidentiality and Microdata Access - Principles and guidelines of Good Practice	 Download
 National Statistics Code of Practice - Protocol on Data Access and Confidentiality (UK)	 Download

۱۴ مزایای انتشار ریزداده‌ها

انتشار، مسئولیت کلیدی یک آژانس آمار است. مطالعه سیاست‌ها و تجربه انتشار بیانیه‌های مأموریت تولیدکنندگان داده در سراسر جهان، بر اهمیت و دلایل دسترسی به فایل‌های ریزداده تأکید می‌کند. همچنین موضوعات چندی برای مراکز و سایر ارائه‌دهندگان و گردآورندگان داده‌ها وجود دارد که باید هنگام تدوین و اجرای سیاست‌ها و برنامه‌های انتشار ریزداده‌ها در نظر بگیرند. موضوعات شامل هزینه‌ها و تخصص درگیر؛ سوالات کیفیت داده‌ها؛ سوءاستفاده احتمالی و سوءتفاهم از داده‌ها توسط کاربران؛ مسائل حقوقی و اخلاقی؛ و حفظ اعتماد و حمایت پاسخ‌دهندگان هستند.

حمایت از تحقیقات

دلیل اصلی - و اغلب تنها دلیل صریح - برای این‌که چرا تولیدکنندگان داده، ریزداده‌های خود را به اشتراک می‌گذارند، حمایت از تحقیقات است. پس از انجام یک پیمایش، آژانس‌هایی که داده‌ها را جمع‌آوری می‌کنند معمولاً طیف گسترده‌ای از خروجی‌های جدولی را تولید می‌کنند تا نکات برجسته و یک نمای کلی از نتایج را به کاربران



ارائه دهند. به سختی می‌توان انتظار داشت که تمام سوالات تحقیقی را که می‌توان با استفاده از این داده‌ها به آنها پرداخت، شناسایی کرد و برای آنها تأمین مالی شد. فایل‌های ریزداده انعطاف‌پذیری قابل توجهی را برای شناسایی روابط و تعامل بین پدیده‌های تحت پوشش در پیمایش به محققان ارائه می‌دهند و در نتیجه تنوع و کیفیت تحقیق را تقویت می‌کنند.

افزایش اعتبار آمارهای رسمی

دسترسی گسترده‌تر به ریزداده‌ها، اعتماد تولیدکنندگان به داده‌ها را با امکان تکرار یا تأیید آن‌ها توسط طرف‌های مستقل نشان می‌دهد.

بهبود قابلیت اطمینان^۱ و ارتباط داده‌ها

رابطه نزدیک‌تر بین ارائه‌دهندگان داده و کاربران آگاه می‌تواند مزایای دیگری را به همراه داشته باشد. استفاده از داده‌ها اغلب، بینش‌هایی را برای بهبودهای بالقوه مانند طراحی پیمایش و انتشار ریز داده‌ها فراهم می‌کند. بازخورد به یک اداره ملی آمار را می‌توان در فرآیند انتشار ریزداده‌ها گنجانده. برای مثال، اداره سرشماری ایالات متحده فرآیند بازخورد از محققان را رسمی کرد، و بازخورد کاربران می‌تواند منجر به بهبود پیمایش در طول زمان شود.

کاهش تکرار در جمع‌آوری داده‌ها

در دسترس قرار دادن فایل‌های ریزداده برای کاربران، اغلب آنها را از جمع‌آوری داده‌های مورد نیاز خود منصرف می‌کند. این امر بار پاسخ‌دهندگان را کاهش می‌دهد و خطر مطالعات ناسازگار در مورد همان موضوع را به حداقل می‌رساند.

افزایش بازده سرمایه‌گذاری

"اشتراک‌گذاری و دسترسی آزاد به داده‌های تحقیقاتی با بودجه عمومی نه تنها به حداکثر کردن پتانسیل تحقیقاتی فناوری‌ها و شبکه‌های دیجیتال جدید کمک می‌کند، بلکه بازده بیشتری از سرمایه‌گذاری عمومی در تحقیق را فراهم می‌کند. (...)" به‌طور مستمر، مقادیر رو به رشد داده‌ها توسط محققان و مؤسسات تحقیقاتی با بودجه عمومی جمع‌آوری می‌شود. این مجموعه داده‌های تحقیقاتی که به سرعت در حال گسترش است، هم سرمایه‌گذاری هنگفتی از بودجه عمومی و هم منبع بالقوه دانش مورد نیاز برای رسیدگی به چالش‌های بی‌شمار پیش روی ما را نشان می‌دهد. برای ارتقای بازده علمی و اجتماعی بهبودیافته در سرمایه‌گذاری عمومی در داده‌های تحقیقاتی، کشورهای عضو سازمان همکاری و توسعه اقتصادی (برای مثال) قوانین، سیاست‌ها و شیوه‌های مختلف دسترسی



به داده‌های تحقیقاتی در سطح ملی را وضع کرده‌اند. در این زمینه، دستورالعمل‌های بین‌المللی، کمک مهمی به تقویت تبادل جهانی و استفاده از داده‌های تحقیقاتی خواهد بود." (برگرفته از اصول و رهنمودهای سازمان همکاری و توسعه اقتصادی در دسترسی به داده‌های تحقیقاتی از بودجه عمومی).



OECD Principles and Guidelines for Access to Research Data from Public Funding

Organisation for Economic co-Operation and Development (OECD) - 2007

The Principles and Guidelines apply to research data gathered using public funds for the purposes of producing publicly accessible knowledge. The nature of "public funding" of research varies significantly from one country to the next, as do existing data access policies and practices at the national, disciplinary and institutional levels.

[Download](#)

استفاده از منابع مالی برای آمار

هرچه فایل‌های داده بیشتر منتشر و مورد استفاده قرار گیرند، برای نهادهای تأمین مالی ارزشمندتر به نظر می‌رسند. این درک می‌تواند آژانس‌های حامی را تشویق به تأمین مالی جمع‌آوری داده‌ها کند. در واقع، در برخی موارد، شواهد استفاده از الزامات حامیان مالی است. استفاده بهتر از داده‌ها به معنای بازدهی بهتر حامیان پیمایش است که در نتیجه تمایل بیشتری به حمایت از فعالیت‌های جمع‌آوری داده‌ها دارند. به‌طور فزاینده‌ای، تأمین مالی پیمایش‌ها توسط حامیان بین‌المللی، تابع انتشار مجموعه داده‌های حاصل می‌شود.

کاهش هزینه انتشار داده‌ها

یک مزیت نهایی می‌تواند از طریق بهبود کارایی^۱ به جمع‌آوردندگان داده عاید شود، زیرا ممکن است تعداد جدول‌های از پیش تعریف‌شده‌ای را که تولید می‌کنند، کاهش دهند و زمان بیشتری را برای ارائه نتایج تحلیلی سطح بالا

1. Improved efficiency



اختصاص دهند. برای مثال، برجسته کردن این موضوع از طریق بخش رسانه و آموزش، می‌تواند برای مخاطبان بیشتری جذاب باشد و حمایت از کار جمع‌آوران داده‌ها، مانند ادارات آمار ملی را تشویق کند. با این حال، زمانی که محققان عمیق‌تر تحقیق کنند، چنین جدول‌هایی کافی نخواهند بود. کارایی نیز باید در زمینه هزینه‌های تولید و انتشار فایل‌های ریزداده در نظر گرفته شود. این موضوعات در بخش بعدی بررسی می‌شوند.

انجام تعهدات قراردادی یا قانونی

در برخی کشورها، سازمان‌های دولتی موظف به انتشار برخی از ریزداده‌های خود هستند. جمع‌آوری داده‌ها اغلب توسط مالیات‌دهندگان تأمین می‌شود و بنابراین یک کالای عمومی در نظر گرفته می‌شود. در موارد دیگر، جمع‌آوری داده‌ها توسط سازمان‌های حامی مالی تأمین می‌شود که نیاز دارند داده‌های حاصل در دسترس محققان قرار گیرد. این تعهد به انتشار ریزداده با تعهد به حفظ محرمانه بودن و حفظ حریم خصوصی در تضاد نیست. مسئولیت تصمیم‌گیری درباره ریزداده‌هایی که باید منتشر شوند و رویه‌هایی که باید برای تولید فایل‌های استفاده عمومی اجرا شوند بر عهده آمارگیر ارشد سازمان یا کمیته انتشار داده‌هایی است که توسط سازمان ایجاد شده است.

ترویج توسعه ابزارهای جدید برای استفاده از داده‌ها

جنبش جدیدی که اغلب از آن به عنوان جنبش «داده‌های باز» یاد می‌شود، در سال‌های اخیر جایگاه زیادی به دست آورده است. در قلب آن، این مفهوم وجود دارد که داده‌های جمع‌آوری شده با استفاده از بودجه عمومی یا تحت نظارت یک سازمان دولتی یک «کالای عامه^۱» است. چندین دولت این دیدگاه را پذیرفته‌اند. برای مثال، طرح دولت باز^۲ در ایالات متحده (www.data.gov) یا معادل آن در بریتانیا (<http://data.gov.uk>) را ببینید. چنین حرکتی مردم را به افزودن ارزش به داده‌های موجود به چالش می‌کشد. با ارائه دسترسی نامحدود به ریزداده‌ها، چنین طرح‌هایی، توسعه نرم‌افزارهای کاربردی جدید، به ویژه برنامه‌های کاربردی نوآورانه فناوری Web 2 را ترویج می‌کنند.

منابع مرتبط به اشتراک گذاشته شده است:

1. Public good
2. Open Government



Related Resources

Documents

- 📄 Diffusion des fichiers de microdonnées - Principes, procédures et pratiques 📄 (1.42 MB)
- 📄 Dissemination of Microdata Files - Principles, Procedures and Practices 📄 (2.05 MB)
- 📄 Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle (5th ed.) 📄 Download
- 📄 OECD Principles and Guidelines for Access to Research Data from Public Funding 📄 Download

۲۰۴ خطرات و هزینه‌ها

خطرها و هزینه‌ها شامل تخصص درگیر؛ سؤالات مربوط به کیفیت داده‌ها، سوءاستفاده احتمالی، و درک نادرست داده‌ها توسط کاربران؛ مسائل حقوقی و اخلاقی؛ و حفظ اعتماد و حمایت پاسخ‌دهندگان است.

مسائل اخلاقی و حفظ اعتماد پاسخ‌دهندگان

هنگام جمع‌آوری داده‌ها از افراد، تأسیسات یا مؤسسات، آژانس‌های آماری و سایر تولیدکنندگان داده معمولاً به پاسخ‌دهندگان اطمینان می‌دهند که اطلاعاتی که ارائه می‌کنند فقط برای مقاصد آماری استفاده خواهد شد. این یک تعهد اخلاقی و قانونی است. برای موفقیت، آژانس آماری باید اعتماد پاسخ‌دهندگان را حفظ کند تا از ادامه همکاری در جمع‌آوری داده‌ها اطمینان حاصل کنند. حفاظت از محرمانه بودن کلید این اعتماد است. اگر پاسخ‌دهندگان درک کنند که یک آژانس آماری از محرمانه بودن داده‌های آنها محافظت نمی‌کند، احتمال کمتری برای همکاری یا ارائه داده‌های دقیق وجود دارد. یک اتفاق، به ویژه، اگر مورد توجه رسانه‌ها قرار گیرد، می‌تواند تأثیر قابل توجهی بر همکاری پاسخ‌دهندگان و در نتیجه بر کیفیت آمار رسمی داشته باشد. این موضوع از دیدگاه آژانس آماری غالب است، اما نگرانی‌های دیگری نیز وجود دارد؛ به بخش مسائل اخلاقی در ادامه، رجوع کنید.

مسائل حقوقی

از دیر باز، قوانین آماری بر مسائل حفظ حریم خصوصی و حفاظت از داده‌ها تأکید داشته و اغلب از انتشار ریزداده‌ها



جلوگیری کرده است. برخی از قوانین آماری برای امکان انتشار فایل‌های ریزداده بی‌نام انتخاب شده، مدرن شده‌اند. بسیاری از قوانین باقی مانده است که باید مدرن شوند، تا مجموعه‌های متنوع فزاینده‌ای از داده‌های تولید شده، مدیریت و منتشر شده توسط آژانس‌های آماری را بهتر در نظر بگیرند؛ به بخش مسائل حقوقی در ادامه، رجوع کنید.

قرار گرفتن در معرض انتقاد و تناقض

"برخی از آژانس‌های آماری نگران هستند که کیفیت ریزداده‌هایشان برای انتشار بیشتر به اندازه کافی خوب نباشد. در حالی که کیفیت ممکن است به اندازه کافی دقیق باشد تا آمار کل را پشتیبانی کند، می‌تواند این مورد برای تجزیه و تحلیل، بسیار دقیق نباشد. در برخی موارد، تنظیماتی برای جمع‌آوری آمار در مرحله ویرایش خروجی بدون اصلاح ریزداده انجام می‌شود. در نتیجه، ممکن است بین نتایج تحقیقات مبتنی بر ریزداده‌ها و داده‌های انبوه منتشر شده ناسازگاری وجود داشته باشد." (مدیریت محرمانه بودن آماری و دسترسی به ریزداده‌ها: اصول و دستورالعمل‌های عملکرد خوب کمیسیون اقتصادی سازمان ملل متحد برای اروپا [UNECE 2007]). اگر برخی از بخش‌های مجموعه داده‌ها بیش از حد غیرقابل اعتماد در نظر گرفته شوند، می‌توانند قبل از انتشار حذف شوند. تولیدکنندگان داده باید در مورد کیفیت، باز و شفاف باشند. نگرانی دیگر این است که ارائه ریزداده به محققان امکان انتشار نتایج را باز می‌کند، که می‌تواند با تخمین‌های تولیدکننده داده در تضاد باشد. هنگامی که تولیدکننده داده‌ها یک آژانس آمار رسمی است، ممکن است منجر به تضاد برآوردهای رسمی در مقابل غیررسمی شود و منجر به زیر سؤال بردن داده‌ها، با پیامدهای سیاسی احتمالی گردد.

ممکن است دلایل مختلفی برای تفاوت وجود داشته باشد. اول، ممکن است در برآوردهای رسمی اشتباهاتی وجود داشته باشد، که در این صورت بررسی‌های بیرونی یک مزیت است. دوم، تفاوت‌ها می‌توانند ناشی از استفاده از نسخه‌های مختلف داده‌ها باشند (برای مثال، فایل اصلی کامل در مقابل یک نسخه عمومی بی‌نام/کم شده، ویرایش بیشتر توسط محقق، و غیره). این تفاوت‌ها باید حاشیه‌ای باشد و به راحتی قابل توضیح باشند. سوم، این ممکن است نتیجه روش‌های مختلف مورد استفاده باشد و اغلب یک مسئله چالش برانگیز برای تولیدکنندگان داده است، زیرا عموم مردم همیشه قادر به درک توضیحات بسیار فنی نیستند.

برای تولیدکنندگان داده مهم است که بتوانند از برآوردهای خود دفاع کنند. این بدان معنی است که جمع‌آوری، پردازش و تجزیه و تحلیل داده‌ها باید کاملاً مستند باشد و این اطلاعات برای دسترسی آسان حفظ شود. در برخی موارد، نتایج منتشر شده ممکن است توسط یا با کمک کارشناسان خارجی که دیگر برای پاسخ به سؤالات در



دسترس نیستند، تولید شده باشد. تولیدکنندگان داده‌ها می‌توانند با اتخاذ و اجرای شیوه‌های دقیق مستندسازی و حفاظت مطابق با استاندارد تکرار^۱، از خود در برابر این خطر محافظت کنند. به‌طور خلاصه، استاندارد تکرار به این صورت تعریف می‌شود: " (...) تنها راه برای درک و ارزیابی کامل یک تحلیل تجربی، دانستن فرآیند دقیقی است که توسط آن داده‌ها تولید شده و تجزیه و تحلیل انجام شده است. (...) استاندارد تکرار معتقد است که اگر شخص ثالثی بتواند نتایج را بدون هیچ‌گونه اطلاعات بیشتر از نویسنده تکرار کند، اطلاعات کافی برای درک، ارزیابی، و استناد به کار قبلی وجود دارد. [Gary King . ۱۹۹۵ . همانندسازی^۲]



Replication replication

Gary King, Harvard University -

Political science is a community enterprise; the community of empirical political scientists needs access to the body of data necessary to replicate existing studies to understand, evaluate, and especially build on this work. Unfortunately, the norms we have in place now do not encourage, or in some cases even permit, this aim. The paper provides suggestions that would facilitate replication and are easy to implement- by teachers, students, dissertation writers, graduate programs, authors, reviewers, funding agencies, and journal and book editors.

[Download](#)

هزینه

آژانس‌های آماری ممکن است نگران هزینه‌ها نیز باشند. این هزینه‌ها نه تنها شامل هزینه‌های ایجاد و مستندسازی فایل‌های ریزداده‌ها، بلکه هزینه‌های ایجاد ابزارهای دسترسی و پادمان‌ها، و پشتیبانی و تأیید درخواست‌های انجام‌شده توسط جامعه پژوهشی می‌شود؛ کاربران جدید فایل‌های داده برای هدایت ساختارهای فایل پیچیده و تعاریف متغیر به کمک نیاز دارند. اگرچه هزینه‌ها توسط آژانس‌های آماری تقبل می‌شود، اما معمولاً برای انجام کارهای بیشتر، بودجه تکمیلی در اختیار آنها قرار نمی‌گیرد. و در کل، محققان بودجه لازم برای کمک به این هزینه‌ها را ندارند [UNECE 2007] بنابراین،

1. Replication standard
2. Replication



در صورت امکان، چنین هزینه‌هایی باید برای اطمینان از حداکثر استفاده از نتایج پیمایش، در بودجه پیمایش گنجانده شوند. این نفع عمومی است که بینش استفاده از داده‌ها برای اطلاع تصمیم‌گیرندگان و مردم در دسترس قرار گیرد. علاوه بر این، اگر داده‌های پیمایش به‌طور گسترده‌تری در این راه مورد استفاده قرار گیرند، سطح بیشتری از محافظت در برابر کاهش بودجه برنامه‌های آماری را فراهم می‌کنند. پیمایش‌هایی که دانش محدودی در حمایت از سیاست‌گذاری ارائه می‌دهند، در برابر حذف آسیب‌پذیرتر هستند.

از دست دادن انحصار

هنگام انتشار ریزداده، صاحبان داده‌ها حق انحصاری داده را برای خود از دست می‌دهند. این موضوع برای محققان دانشگاهی بیشتر از تولیدکنندگان رسمی است، اگرچه تولیدکنندگان رسمی داده یا برخی از کارکنان آنها گاهی از دسترسی انحصاری به داده‌ها برای ارائه خدمات مشاوره استفاده می‌کنند. بیشتر، حامیان پیمایش، یک دوره قانونی و «معقول» دسترسی انحصاری تولیدکننده به داده‌ها را تعریف می‌کنند و پس از آن داده‌ها باید برای سایر کاربران قابل دسترسی باشند.

ظرفیت فنی

ظرفیت فنی خاصی برای پشتیبانی از انتشار فایل‌های ریزداده مورد نیاز است. فایل‌ها باید به خوبی مستند شده باشند، ترجیحاً از استاندارد فراداده DDI استفاده کنند و حفظ شوند. علاوه بر این، فایل‌ها باید برای شناسایی خطر افشای اطلاعات فردی بررسی شوند و با استفاده از تکنیک‌های مختلف، این خطر کاهش یابد.

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

-  [Managing Statistical Confidentiality and Microdata Access - Principles and guidelines of Good Practice](#)  [Download](#)



آیا جایگزین‌هایی برای اشتراک‌گذاری فایل‌های ریزداده وجود دارد؟

انتشار ریزداده به این معنی است که به کاربر امکان دسترسی به فایل‌های ریزداده می‌شود. سایر اشکال دسترسی به ریزداده‌ها شامل ارسال داده از راه دور و دسترسی از راه دور به ریزداده‌ها، بدون اجازه دسترسی مستقیم به کاربران است. الزامات خط مشی برای چنین دسترسی توسط دستورالعمل‌های ما پوشش داده نمی‌شود، اما در اینجا خلاصه‌ای از سایر مراجع ارائه می‌شود. توجه داشته باشید که این رویکردها معمولاً پرهزینه و از نظر فنی سخت هستند.

ارسال کار^۱

یکی از روش‌هایی که به کاربران امکان می‌دهد تجزیه و تحلیل داده‌های محرمانه را انجام دهند، شامل ایجاد فرآیندی است که آنها را قادر می‌سازد برنامه‌های پردازش و تجزیه و تحلیل داده‌ها را از راه دور به نگهدارنده داده ارسال کنند. به کاربر، یک مجموعه داده مصنوعی داده می‌شود که ساختار و محتوای مجموعه داده‌های واقعی را تکرار می‌کند. این امر، محقق را قادر می‌سازد تا اقداماتی را با استفاده از نرم‌افزارهایی مانند SAS، SPSS یا Stata انجام دهد. سپس اقداماتی که کاربر انجام می‌دهد به کارکنان نگهدارنده داده منتقل می‌شوند که کار را براساس مجموعه داده واقعی اجرا می‌کنند. سپس نتایج برای افشا بررسی می‌شود و به محقق بازگردانده می‌شود.

دسترسی از راه دور به داده‌ها

دسترسی از راه دور به داده‌ها، دسترسی به نرم‌افزار تجزیه و تحلیل و جدول‌بندی داده‌ها در وب را برای کاربران فراهم می‌کند، بدون این که کاربران امکان دانلود مجموعه داده‌ها یا تولید جدول‌هایی را که تعداد ثبت‌های فردی یا کوچک^۲ را نشان می‌دهند، داشته باشند. نرم‌افزارهای مختلفی، عمدتاً تجاری (مانند Nesstar، Beyond، SuperCross، Redatam، PcAxis، 20/20 و غیره) در دسترس هستند. برخی از مراکز داده پیشرفته، برنامه‌های نرم‌افزاری خود را ارائه می‌دهند. برای مثال، آرشیو داده‌های بریتانیا^۳ (UKDA) در حال اجرای آزمایشی سرویس داده ایمن^۴ (SDS) است که "قصد دارد با ایجاد امکان دسترسی ایمن و امن از راه دور توسط محققان با حسن نیت به داده‌هایی که تا به حال آن‌قدر حساس، جزئی، محرمانه یا به‌طور بالقوه افشاءکننده تلقی می‌شدند که تحت ترتیبات استاندارد صدور مجوز و انتشار در دسترس نبودند، تعالی در تحقیقات را ارتقا بخشد." این رویکرد برای جدول‌بندی، به‌ویژه سرشماری نفوس و مسکن رضایت‌بخش است، اما برای تحلیل‌های پیشرفته، نیست.

1. Job Submission
2. Individual or small numbers of records
3. UK Data Archive (UKDA)
4. Secure Data Service (SDS)



جذب یک محقق به عنوان کارمند موقت

برخی از تولیدکنندگان داده با به کارگیری آنها به عنوان کارکنان موقت، امکان دسترسی به ریزداده‌ها را برای محققان فراهم می‌کنند. این امر آنها را مشمول مقررات امنیتی مشابه کارکنان دائمی آژانس تولیدکننده داده می‌کند. این ترتیبات باید محدود به مواردی باشد که محقق واقعاً به کار تولیدکننده داده کمک می‌کند. (کمیسیون اقتصادی سازمان ملل متحد برای اروپا (UNECE)، کنفرانس آماردانان اروپایی، ۲۰۰۷. "مدیریت محرمانه بودن آماری و دسترسی ریزداده‌ها: اصول و دستورالعمل‌های عملکرد خوب"^۱)

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

[Managing Statistical Confidentiality and Microdata Access - Principles and guidelines of Good Practice](#)

Links

[Secure Data Service \(SDS\), UK Data Archive](#)

۴'۴ مسائل حقوقی

آیا انتشار فایل‌های ریزداده برای یک تولیدکننده داده قانونی است؟ هیچ پاسخ یکسان و شمشخصی برای این سؤال وجود ندارد. قانونی که براساس آن یک آژانس تولیدکننده داده کار می‌کند، مختص هر کشور و در چارچوب برنامه خاص است. انتشار ریزداده در برخی موارد یک الزام قانونی است. اما در بیشتر موارد، قانون محدودیت‌هایی ایجاد می‌کند، نه تعهداتی، و سیاست‌های انتشار ریزداده‌های یک کشور براساس چارچوب قانونی آن شکل می‌گیرد. برای تولیدکنندگان داده بسیار مهم است که "اطمینان حاصل کنند که یک پایه قانونی و اخلاقی معتبر (همچنین ابزارهای فنی و روش‌شناختی) برای محافظت از محرمانگی وجود دارد. این مبنای حقوقی و اخلاقی، مستلزم ارزیابی متوازن بین منافع عمومی حفاظت از رازداری از یک‌سو و منافع عمومی تحقیق از سوی دیگر است. تصمیم در مورد ارائه یا عدم دسترسی می‌تواند به شایستگی پیشنهادهای تحقیقاتی خاص و اعتبار محقق بستگی داشته باشد و باید در ترتیبات قانونی برای این امر در نظر گرفته شود. (کمیسیون اقتصادی سازمان ملل متحد برای اروپا (UNECE)،

1. Managing Statistical Confidentiality and Microdata Access: Principles and Guidelines of Good Practice



کنفرانس آماردانان اروپایی، ۲۰۰۷. "مدیریت محرمانه بودن آماری و دسترسی ریزداده‌ها: اصول و دستورالعمل‌های عملکرد خوب"



Managing Statistical Confidentiality and Microdata Access - Principles and guidelines of Good Practice

Conference of European Statisticians (CES) and United Nations Economic Commission for Europe (UNECE) - 2007

These guidelines have been prepared at the request of the Conference of European Statisticians (CES) by a task force chaired by Dennis Trewin, the Australian statistician. The guidelines and core principles of confidentiality and microdata access were adopted by the CES plenary session in June 2006 and the CES Bureau in October 2006.

[Download](#)



"ترتیبات دسترسی به داده‌ها باید به حقوق قانونی و منافع مشروع همه ذینفعان در شرکت تحقیقاتی عمومی احترام بگذارد. دسترسی و استفاده از داده‌های تحقیقاتی خاص لزوماً توسط انواع مختلفی از الزامات قانونی محدود می‌شود، که ممکن است شامل محدودیت‌هایی به دلایل زیر باشد:

- **امنیت ملی:** داده‌های مربوط به اطلاعات، فعالیت‌های نظامی یا تصمیم‌گیری سیاسی می‌توانند طبقه‌بندی شده و در نتیجه دسترسی محدود شود؛
- حریم خصوصی و محرمانه بودن: داده‌های مربوط به افراد و سایر داده‌های شخصی تحت قوانین و سیاست‌های ملی برای محافظت از محرمانگی و حریم خصوصی، در معرض دسترسی محدود هستند. با این حال، رویه‌های بی‌نام یا محرمانه‌سازی که سطح رضایت‌بخشی از محرمانگی را تضمین می‌کنند باید توسط متولیان چنین داده‌هایی در نظر گرفته شوند تا حداکثر کاربرد داده‌ها را برای محققان حفظ کنند؛
- اسرار تجاری و حقوق مالکیت معنوی: داده‌های مربوط به، یا از، کسب‌وکارها یا سایر طرف‌هایی که حاوی اطلاعات محرمانه هستند ممکن است برای تحقیق در دسترس نباشد. (...)"

مرکز ملی آمار سلامت (NCHS). ۲۰۰۲. "سیاست انتشار ریزداده‌ها"



Policy on Micro-data Dissemination (NCHS)

National Center for Health Statistics - July 2002

This policy addresses when, to whom, and in what form the National Center for Health Statistics (NCHS) disseminates data specific to individuals, households, establishments, or events—i.e., micro-data—and also outlines dissemination procedures.

[Download](#)

اصول بنیادی آمار رسمی سازمان ملل متحد

از آنجایی که بسیاری از کشورها هنگام تنظیم قوانین خود به اصول بنیادی آمار رسمی سازمان ملل متحد اشاره می‌کنند، بررسی این اصول از آنجایی که مربوط به محرمانه بودن آماری است، مفید می‌باشد.



متن زیر از کمیسیون اقتصادی سازمان ملل متحد برای اروپا (UNECE)، کنفرانس متخصصان آمار اروپا ۲۰۰۷ گرفته شده است. "مدیریت محرمانه بودن آماری و دسترسی ریزداده‌ها: اصول و دستورالعمل‌های عملکرد خوب."

اصل ششم حاکم بر فعالیت‌های آماری بین‌المللی می‌گوید: "داده‌های فردی جمع‌آوری شده توسط آژانس‌های آمار برای جمع‌آوری آمار، خواه به اشخاص حقیقی یا حقوقی مربوط باشد یا نه، باید کاملاً محرمانه بوده و منحصرأً برای مقاصد آماری استفاده شود."

هر اصولی برای دسترسی به ریزداده باید با این اصل توصیه شده یا اصول مندرج در قانون آژانس‌های آماری سازگار باشد. برای مدیریت محرمانه بودن ریزداده‌ها باید نکات زیر در نظر گرفته شود.

اصل ۱: استفاده مناسب ریزداده‌ها

"مطلوب است که ریزداده‌های جمع‌آوری شده برای مقاصد آمار رسمی جهت تجزیه و تحلیل آماری به منظور پشتیبانی از تحقیقات تا زمانی که محرمانه بودن، محفوظ است، استفاده شود. (...)

در دسترس قرار دادن ریزداده‌ها برای تحقیق، تا زمانی که داده‌ها شناسایی یک فرد را غیرممکن می‌کند، با اصل ششم بنیادی سازمان ملل منافات ندارد. اصل یک تعهدی برای ارائه ریزداده‌ها نیست. مرکز باید تصمیم بگیرد که آیا ریزداده ارائه دهد یا خیر.

ممکن است نگرانی‌های دیگری (به عنوان مثال، کیفیت) وجود داشته باشد که ارائه دسترسی به ریزداده‌ها را نامناسب می‌سازد. یا ممکن است افراد یا مؤسسات خاصی وجود داشته باشند که ارائه ریزداده به آنها نامناسب باشد."

اصل ۲: ریزداده‌ها فقط باید برای اهداف آماری در دسترس قرار گیرند

"برای اصل دو، باید بین استفاده‌های آماری یا تحلیلی و کاربری‌های اداری تمایز قائل شد. در مورد استفاده آماری یا تحلیلی، هدف، استخراج آماری است که به یک گروه (اعم از اشخاص یا اشخاص حقوقی) اشاره دارد. در مورد استفاده اداری، هدف، به دست آوردن اطلاعات در مورد یک شخص یا شخص حقوقی خاص برای اتخاذ تصمیمی است که ممکن است برای فرد، منفعت یا ضرر داشته باشد. برای مثال، برخی از درخواست‌های داده ممکن است قانونی (حکم قضایی) باشد اما با این اصل مغایرت داشته باشد. رد این درخواست‌ها به نفع اعتماد عمومی به سیستم آماری رسمی است. اگر استفاده از ریزداده با اهداف آماری یا تحلیلی ناسازگار باشد، در این صورت دسترسی ریزداده



نباید فراهم شود. کمیته‌های اخلاقی یا ترتیب مشابه^۱ ممکن است در شرایطی که عدم اطمینان وجود دارد که آیا امکان دسترسی وجود دارد یا خیر، کمک کنند.

محققان برای اهداف تحقیقاتی به ریزداده‌ها دسترسی دارند، اما برای حمایت از این تحقیق، ممکن است نیاز به جمع‌آوری آماری از اشکال مختلف، گردآوری توزیع‌های آماری، مدل‌های آماری مناسب، یا تجزیه و تحلیل تفاوت‌های آماری بین جمعیت‌های فرعی داشته باشند. این استفاده‌ها با اهداف آماری سازگار خواهد بود. تا جایی که از ریزداده‌ها به این صورت استفاده می‌شود، می‌توان گفت که از اهداف تحقیقاتی نیز پشتیبانی می‌کند."

اصل ۳: ارائه ریزداده باید مطابق با مقررات قانونی و سایر ترتیبات ضروری باشد که تضمین می‌کند محرمانه بودن ریزداده‌های منتشر شده محافظت می‌شود.

"با توجه به اصل سه، ترتیبات قانونی برای محافظت از محرمانه بودن باید قبل از انتشار هرگونه ریزداده وجود داشته باشد. با این حال، ترتیبات قانونی باید با اقدامات اداری و فنی تکمیل شود تا دسترسی به ریزداده‌ها را تنظیم کند و اطمینان حاصل شود که داده‌های فردی نمی‌توانند فاش شوند. وجود و قابل مشاهده بودن چنین ترتیباتی (اعم از قانون یا مقررات تکمیلی، احکام و غیره) برای افزایش اعتماد عمومی به استفاده مناسب از ریزداده‌ها ضروری است. ترتیبات قانونی به وضوح، ارجحیت دارند، اما در برخی کشورها ممکن است این امکان‌پذیر نباشد و شکل دیگری از ترتیبات اداری باید ایجاد شود. مقررات قانونی (یا سایر ترتیبات) نیز باید با مقامات حفظ حریم خصوصی کشورهایی که در آنجا وجود دارند، قبل از این که توسط قانون ایجاد شوند، تأیید شود. اگر چنین مقاماتی وجود نداشته باشند، ممکن است سازمان‌های غیردولتی وجود داشته باشند که نقش «دیده‌بان» در مسائل مربوط به حریم خصوصی را داشته باشند. معقول است که حمایت آنها برای هرگونه ترتیبات قانونی یا دیگر ترتیبات، یا حداقل رسیدگی به هرگونه نگرانی جدی که ممکن است داشته باشند، دریافت شود.

در برخی کشورها، قانون مجاز وجود ندارد. حداقل، انتشار ریزداده‌ها باید توسط برخی از مقامات پشتیبانی شود. با این حال، یک قانون مجاز، یک رویکرد ارجح است."

اصل ۴: رویه‌های دسترسی محقق به ریزداده، و همچنین استفاده‌ها و استفاده‌کنندگان ریزداده‌ها باید شفاف و در دسترس عموم باشد.

"اصل چهار برای افزایش اعتماد عمومی به استفاده مناسب از ریزداده‌ها و نشان دادن این که تصمیمات در مورد انتشار







1. Similar arrangement



ریز داده‌ها بر مبنای عینی گرفته می‌شود، مهم است و به مرکز بستگی دارد که تصمیم بگیرد آیا، چگونه و برای چه کسی می‌توان ریز داده‌ها را منتشر کرد. اما تصمیمات آنها باید شفاف باشد. وبسایت مرکز روشی مؤثر برای اطمینان از انطباق و همچنین برای ارائه اطلاعات در مورد نحوه دسترسی به گزارش‌های تحقیقاتی براساس ریز داده‌های منتشر شده است." منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

 Managing Statistical Confidentiality and Microdata Access - Principles and guidelines of Good Practice	 Download
 NCHS Staff Manual on Confidentiality	 Download
 Policy on Micro-data Dissemination (NCHS)	 Download

۵'۴ مسائل اخلاقی

هنگام جمع‌آوری داده‌ها از افراد، تأسیسات یا مؤسسات، آژانس‌های آماری و سایر تولیدکنندگان داده معمولاً به پاسخ‌دهندگان این اطمینان را می‌دهند که اطلاعاتی که ارائه می‌کنند فقط برای مقاصد آماری استفاده خواهد شد. این یک تعهد اخلاقی و قانونی است.

برای موفقیت، "سازمان‌های غیردولتی باید اعتماد پاسخ‌دهندگان را حفظ کنند تا در جمع‌آوری داده‌های خود به همکاری ادامه دهند. حفاظت از محرمانه بودن، عنصر کلیدی این اعتماد است. اگر پاسخ‌دهندگان بر این باور یا درک باشند که یک مرکز از محرمانه بودن داده‌های آنها محافظت نمی‌کند، احتمال کمتری دارد که همکاری کنند یا داده‌های دقیق ارائه کنند. یک اتفاق، به ویژه اگر مورد توجه رسانه‌ها قرار گیرد، می‌تواند تأثیر قابل توجهی بر همکاری پاسخ‌دهندگان و در نتیجه بر کیفیت آمار رسمی داشته باشد. این موضوع از دیدگاه مراکز غالب است اما نگرانی‌های دیگری نیز وجود دارد. یکی از موارد کلیدی این است که آیا آنها از اختیارات کافی برای حمایت از دسترسی محقق به ریز داده‌ها، چه از طریق یک دستور قانونی یا هر نوع مجوز دیگری برخوردار هستند یا خیر. (UNECE)، کنفرانس آماردانان اروپایی ۲۰۰۷. "مدیریت محرمانه بودن آماری و دسترسی به ریز داده‌ها: اصول و



دستورالعمل‌های عملکرد خوب.^۱)

برای به حداکثر رساندن استفاده از ریزداده‌ها، مراکز و سایر ارائه‌دهندگان داده باید بین نیاز به محرمانه بودن و پیش‌بینی دسترسی، تعادل برقرار کنند و می‌تواند با استفاده از انواع مختلف ریزداده، همان‌طور که قبلاً بحث شد، به‌دست آید. گزینه دیگر، اگرچه کاربرد محدودی دارد، کسب رضایت رسمی برای به اشتراک گذاشتن داده‌های جمع‌آوری شده از هر پاسخ‌دهنده به پیمایش است.

کسب رضایت از افراد

رضایت ممکن است از طریق امضا یا ساخت‌وساز^۱ به‌دست آید. در وهله اول، ممکن است توضیحی کتبی در مورد اطلاعاتی که از شخص خواسته می‌شود به پاسخ‌دهنده ارائه شود و از او خواسته شود نام خود را امضا کند، بنابراین اجازه استفاده از آن اطلاعات را همان‌طور که توضیح داده شد، نشان می‌دهد. اما در برخی موارد، این اطلاعات به صورت کتبی یا شفاهی به او داده می‌شود. اگر پاسخ‌دهنده داده‌های درخواستی را ارائه کند، [محقق]^۲ «تعبیر» می‌کند که پاسخ‌دهنده با استفاده‌های مورد نظر و به اشتراک‌گذاری داده‌ها با طرف‌هایی که خوانده یا درباره آنها گفته شده موافقت می‌کند. سپس [محقق] می‌تواند از داده‌هایی که برای پاسخ‌دهنده توضیح داده شده است استفاده کند، اما هیچ استفاده دیگری از داده‌ها ممکن نیست. [*]

اخذ رضایت از یک مؤسسه

"در مورد مؤسسات، رویکرد تا حدی به این بستگی دارد که درخواست اطلاعات در مصاحبه شخصی یا از طریق پست انجام شود.

الف. اگر درخواست اطلاعات شخصاً توسط یکی از کارکنان یا عامل [محقق] انجام شود، فرد تماس‌گیرنده ابتدا از این که چه کسی مجاز به ارائه داده‌های درخواستی از طرف مؤسسه است، سؤال می‌کند. هنگامی که چنین شخص مجاز از کاربردهایی که قرار است از داده‌ها استفاده شود مطلع می‌شود و سپس داده‌ها را ارائه می‌دهد، کارکنان [آژانس تحقیق] تعبیر می‌کنند که مؤسسه با استفاده از داده‌ها همان‌طور که مشخص شده است موافقت کرده است.


ب. هنگامی که داده‌ها از طریق پست از یک مؤسسه درخواست می‌شوند، این درخواست، ممکن است به خود مؤسسه، به مدیر مؤسسه یا شخص دیگری که همان‌طور که [آژانس تحقیق] قبلاً تأیید کرده است، مجاز است داده‌های

1. Construction
2. Investigator



درخواستی را از طرف مؤسسه ارائه دهد، ارسال شود. نامه‌ای که درخواست را ارسال می‌کند، موارد استفاده از داده‌ها را توضیح می‌دهد. هنگامی که کارکنان [آژانس تحقیق] داده‌های درخواستی را از مؤسسه دریافت می‌کنند، به این معنی است که مؤسسه با استفاده‌هایی که از آن مطلع شده است موافقت کرده است. [*]

[*] منبع: مرکز ملی آمار سلامت - راهنمای کارکنان در مورد محرمانگی (۲۰۰۴)



NCHS Staff Manual on Confidentiality

National Center for Health Statistics (NCHS) - 2004

The confidentiality of records is a matter of primary concern to the NCHS. To elicit health information from the American people and from the health care providers through our surveys, the NCHS must be able to assure them that this information will be protected from all unauthorized persons. This means that NCHS must have strong laws to protect these records, and must establish and follow established procedures. This manual outlines the NCHS's policies that implement federal law and ensure that all confidential information will be fully protected. It should be viewed in unison with the NCHS data release policies addressing access to data and NCHS Research Ethics Review Board Requirements.

[Download](#)

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

- Managing Statistical Confidentiality and Microdata Access - Principles and guidelines of Good Practice
 [Download](#)
- NCHS Staff Manual on Confidentiality
 [Download](#)

۶'۴ دسترسی به گزینه‌ها



فایل‌های ریزداده برای انتشار تقریباً همیشه با مواردی که صرفاً توسط کارکنان آژانس‌های تولیدکننده داده استفاده می‌شود، متفاوت است. آماده‌سازی فایل‌های ریزداده خام برای انتشار شامل فرآیندهایی است که ممکن است محتوا و/یا تعداد ثبت‌ها را تنظیم کند. محتوای سوابق (ثبت‌های) موجود در فایل‌های ریزداده برای انتشار با حذف اطلاعات از شناسه‌های مستقیم و غیرمستقیم جهت محافظت از بی‌نام بودن پاسخ‌دهندگان ویرایش می‌شود. اما کوچک‌سازی اطلاعات لزوماً به معنای حذف متغیرها نیست. در برخی موارد، کدگذاری مجدد متغیرها به دسته‌های با جزئیات کمتر برای این‌که اطلاعات کمتری داشته باشند، کافی است. گاهی اوقات این نیز مستلزم کوتاه کردن تعداد سوابق موجود در یک فایل ریزداده منتشر شده است، به ویژه در مورد داده‌های سرشماری جمعیت. فرآیندهای حفاظت از هویت پاسخ‌دهندگان در مجموع به عنوان کنترل افشای آماری (SDC) یا بی‌نام‌سازی نامیده می‌شود. در صورتی که بتوان از محرمانگی پاسخ‌دهنده به اندازه کافی محافظت کرد، فایل‌های ریزداده جمع‌آوری شده برای آمار رسمی باید منتشر شوند. در نظر گرفتن سه نوع فایل در هنگام ایجاد خط مشی انتشار توصیه می‌شود: فایل‌های استفاده عمومی، فایل‌های دارای مجوز، و محفظه‌های داده. این فایل‌ها از نظر سطح دسترسی برای کاربران و میزان بی‌نام بودن آنها متفاوت هستند.

"هیچ فردی (...) نمی‌تواند به واسطه شغل خود، حق دریافت یا دسترسی به داده‌های قابل شناسایی (...) را داشته باشد. دسترسی به داده‌های قابل شناسایی صرفاً براساس وضعیت شغلی، وابستگی سازمانی یا تعهد مالی تعیین نمی‌شود. مهمتر، نیاز به داده‌های قابل شناسایی، استفاده از داده‌ها و نقش و مسئولیت درخواست‌کننده در رابطه با فعالیت جمع‌آوری داده‌ها است. از آنجایی که هرگونه دسترسی به داده‌های قابل شناسایی خطرآفرین است، دسترسی به این داده‌ها پس از اعطای دسترسی به دقت ارزیابی و ردیابی می‌شود." [مرکز ملی آمار سلامت (NCHS)، ۲۰۰۲].

"سیاست انتشار ریزداده‌ها[۱]."

فایل‌های قابل استفاده برای عموم^۱ (PUF)

PUFS برای هر کسی که موافقت می‌کند با گرووی از مجموعه‌ای از شرایط ساده، تعیین می‌کند با داده‌ها چه کاری می‌توان انجام داد (به عنوان مثال، داده‌ها نمی‌توانند فروخته شوند). در برخی موارد، PUFS بدون قید و شرط منتشر می‌شوند و اغلب به صورت آنلاین در دسترس هستند. این داده‌ها به راحتی قابل دسترسی هستند زیرا خطر شناسایی تک تک پاسخ‌دهندگان در حداقل است. به حداقل رساندن خطر افشا، شامل حذف تمام محتوایی است که می‌تواند مستقیماً پاسخ‌دهندگان را شناسایی کند - به عنوان مثال، نام، آدرس، و شماره تلفن. علاوه بر این، نیاز به پاک‌سازی

1. Public Use Files (PUF)



شناسه‌های غیرمستقیم مربوطه از فایل ریزداده دارد. اینها در طرح‌های پیمایش متفاوت است، اما معمولاً شناسه‌های غیرمستقیم کوچک‌سازی‌شده شامل اطلاعات جغرافیایی زیر سطح فرو ملی است که نمونه در آن نماینده است. گاهی اوقات، برخی ثبت‌ها ممکن است از PUFs، و همچنین متغیرهایی که با توزیع بسیار ناهنجار یا نقاط پرت مشخص می‌شوند، حذف شوند. با این حال، به جای حذف کل ثبت‌ها یا متغیرها از فایل‌های ریزداده، روش‌های جایگزین SDC می‌توانند خطر افشا را به حداقل و در عین حال محتوای اطلاعات را به حداکثر برسانند. چنین روش‌هایی شامل کدگذاری بالا و پایین، کوچک‌سازی موضعی یا تکنیک‌های بهم‌ریختگی داده می‌شود.

PUFs معمولاً از فایل‌های داده‌های سرشماری - زیرمجموعه‌ای از سوابق به جای کل فایل - و پیمایش‌های خانوار تولید می‌شوند. در حالی که از نظر فنی ایجاد PUFs برای پیمایش‌های تجاری امکان‌پذیر است، این امر مجموعه‌ای از چالش‌ها را ارائه می‌دهد که به‌طور جداگانه مورد بررسی قرار خواهند گرفت.

PUFs باید تا حد امکان حاوی اطلاعات مفید باشند. همان‌طور که توسط مرکز ملی آمار سلامت ایالات متحده (NCHS) در سال ۲۰۰۲ بیان شد، "هدف این است که ریزداده‌ها را به‌طور گسترده و در دقیق‌ترین شکل ممکن در دسترس قرار دهیم، تنها مشروط به محدودیت‌های تحمیل شده توسط منابع، کیفیت داده‌ها، فناوری و نیاز به حفاظت از محرمانه بودن".



شرایط دسترسی و استفاده PUFs

۱. داده‌ها و سایر مطالب ارائه شده توسط مرکز بدون موافقت کتبی به افراد، مؤسسات یا سازمان‌های دیگر توزیع یا فروخته/ارائه نخواهد شد.

۲. داده‌ها فقط برای مقاصد آماری و تحقیقات علمی مورد استفاده قرار خواهند گرفت. آنها صرفاً برای گزارش دادن اطلاعات جمع‌آوری شده، از جمله مدل‌سازی، و نه برای تحقیق در مورد افراد یا سازمان‌های خاص استفاده خواهند شد.

۳. هیچ تلاشی برای شناسایی مجدد پاسخ‌دهندگان انجام نخواهد شد و از هویت هر شخص یا مؤسسه‌ای که به‌طور سهوی شناسایی شده باشد استفاده نخواهد شد. هرگونه شناسایی بلافاصله به مرکز گزارش خواهد شد.

۴. هیچ تلاشی برای ایجاد ارتباط (لینک) بین مجموعه داده‌های ارائه شده توسط مرکز یا بین داده‌های مرکز و سایر مجموعه داده‌هایی که بتواند افراد یا سازمان‌ها را شناسایی کند، انجام نخواهد شد.

۵. هر کتاب، مقاله، مقاله کنفرانس، پایان نامه، رساله، گزارش، یا سایر انتشاراتی که از داده‌های به‌دست آمده از مرکز استفاده می‌کنند، منبع را با شرط استناد ارائه شده به همراه مجموعه داده و تشکر از مرکز ارائه شونده ذکر می‌کنند.

۶. یک نسخه الکترونیکی از کلیه نشریات براساس داده‌های درخواستی برای مرکز ارسال خواهد شد.

۷. گردآورنده اصلی داده‌ها، مرکز، و آژانس‌های تامین مالی مربوطه هیچ مسئولیتی در قبال استفاده یا تفسیر یا استنتاج داده‌ها براساس آن ندارند.

توجه: موارد ۳ و ۶ در لیست، مستلزم آن است که به کاربران یک راه آسان برای برقراری ارتباط با ارائه‌دهنده داده، ارائه شود. ارائه شماره تماس، آدرس ایمیل، و احتمالاً یک سیستم "ارائه بازخورد" آنلاین، روش خوبی است.



فایل‌های دارای مجوز

فایل‌های دارای مجوز - که فایل‌های تحقیقاتی نیز نامیده می‌شوند - از PUFs متمایز هستند: انتشار آن‌ها به کاربرانی محدود می‌شود که پس از ارسال یک درخواست مستند و امضای توافقنامه‌ای/تفاهم‌نامه‌ای که بر استفاده از داده‌ها حاکم است، مجوز دسترسی دریافت کرده‌اند. در حالی که فایل‌های دارای مجوز معمولاً برای به حداقل رساندن خطر شناسایی افراد هنگام استفاده مجزا، بی‌نام می‌شوند، اگر با فایل‌های داده دیگر مرتبط شوند، ممکن است حاوی داده‌های بالقوه قابل شناسایی باشند.

شناسه‌های مستقیم مانند نام پاسخ‌دهندگان باید از مجموعه داده‌های دارای مجوز حذف شوند. با این حال، فایل‌های داده ممکن است همچنان حاوی متغیرهای غیرمستقیم باشند که می‌توانند پاسخ‌دهندگان را با تطبیق آنها با فایل‌های داده دیگر مانند فهرست‌های رأی‌دهندگان، ثبت اراضی، یا سوابق مدرسه شناسایی کنند.

هنگام انتشار فایل‌های دارای مجوز، ایجاد و امضای توافق‌نامه بین تولیدکننده داده و کاربران با حسن نیت - کاربران قابل اعتماد با نیاز قانونی به دسترسی به داده‌ها - توصیه می‌شود. این توافقنامه باید بر دسترسی و استفاده از چنین فایل‌های ریزداده نظارت کند. گاهی اوقات، قراردادهای مجوز فقط با کاربران وابسته به یک مؤسسه حامی مناسب، مانند مراکز تحقیقاتی، دانشگاه‌ها یا شرکای همان مرکز منعقد می‌شوند.

همچنین توصیه می‌شود که قبل از انعقاد قرارداد دسترسی و استفاده از داده‌ها، تولیدکننده داده از کاربران بالقوه بخواهد فرم درخواستی را تکمیل کنند تا نیاز به استفاده از یک فایل دارای مجوز به جای نسخه PUF را در صورت موجود بودن، برای یک آمار اعلام شده یا هدف تحقیق نشان دهند. فرم‌های درخواست و توافقنامه‌های فایل‌های دارای مجوز الگو در ادامه ارائه شده‌اند که شرایطی را که تحت آن باید به فایل‌های ریزداده دسترسی داشت را مورد بحث قرار می‌دهند.

شرایط دسترسی و استفاده از فایل‌های داده دارای مجوز

نکته: موارد ۱ تا ۸ مشابه شرایط استفاده از فایل‌های استفاده عمومی می‌باشد. موارد ۹ و ۱۰ باید در مورد یک توافق عمومی تطبیق داده شوند.

۱. داده‌ها و سایر مطالب ارائه شده توسط مرکز بدون اجازه کتبی مرکز به افراد، مؤسسات یا سازمان‌های دیگر توزیع



یا فروخته/ ارائه نخواهد شد.

۲. داده‌ها فقط برای مقاصد آماری و تحقیقات علمی مورد استفاده قرار خواهند گرفت. آنها صرفاً برای گزارش دادن اطلاعات جمع‌آوری شده، از جمله مدل‌سازی، و نه برای تحقیق در مورد افراد یا سازمان‌های خاص استفاده خواهند شد.

۳. هیچ تلاشی برای شناسایی مجدد پاسخ‌دهندگان انجام نخواهد شد و از هویت هر شخص یا مؤسسه‌ای که به‌طور سهوی شناسایی شده باشد، استفاده نخواهد شد. هر نوع شناسایی مشابه این موضوع بلافاصله به مرکز گزارش خواهد شد.

۴. هیچ تلاشی برای ایجاد ارتباط (لینک) بین مجموعه داده‌های ارائه شده توسط مرکز یا بین داده‌های مرکز و سایر مجموعه داده‌هایی که بتواند افراد یا سازمان‌ها را شناسایی کند، انجام نخواهد شد.

۵. هر کتاب، مقاله، مقاله کنفرانس، پایان‌نامه، رساله، گزارش، یا سایر انتشاراتی که از داده‌های به‌دست آمده از مرکز استفاده می‌کنند، مطابق با الزامات استنادی ارائه شده به مجموعه داده، منبع را ذکر خواهند کرد.

۶. یک نسخه الکترونیکی از کلیه نشریات براساس داده‌های درخواستی برای مرکز ارسال خواهد شد.

۷. مرکز و آژانس‌های تأمین مالی مربوطه هیچ مسئولیتی در مورد استفاده از داده‌ها یا تفسیر یا استنباط براساس آن ندارند.

۸. یک نسخه الکترونیکی از کلیه نشریات براساس داده‌های درخواستی برای مرکز ارسال خواهد شد.

۹. سازمان محقق، مدیر ارشد و سایر محققین دخیل در استفاده از داده‌ها باید شناسایی شوند. محقق اصلی باید مجوز را از طرف سازمان امضا کند. اگر مدیر مجاز به امضاء از طرف سازمان پذیرنده نباشد، باید یک نماینده مناسب مشخص شود.

۱۰. استفاده مورد نظر از داده‌ها، شامل فهرستی از خروجی‌های مورد انتظار و خط مشی انتشار سازمان باید مشخص شود.

توجه: شرایط ۹ تا ۱۱ ممکن است برای مؤسسات آموزشی لغو شود.



توافق بین [آژانس ارائه‌دهنده] و [آژانس دریافت‌کننده] در مورد نگهداری و استفاده از ریزداده‌ها

الف- این توافقنامه مربوط به مجموعه ریزداده‌های زیر است:

۱.
۲.
۳.
۴.
۵.

ب- شرایط قرارداد:

نماینده [آژانس ارائه‌دهنده] به عنوان مالک حق نسخه‌برداری در مواد فهرست شده در بخش الف، یا طبق مجوز قانونی صاحب حق نسخه‌برداری در مواد، به [آژانس دریافت‌کننده] اجازه مجموعه داده‌های فهرست شده در بخش الف را می‌دهد. توسط کارمندان [آژانس پذیرنده] با رعایت شرایط زیر استفاده شود:

۱. ریزداده‌ها (شامل زیرمجموعه‌های مجموعه داده‌ها) و مطالب دارای حق نسخه‌برداری ارائه‌شده توسط [آژانس ارائه‌دهنده] بدون اجازه کتبی [آژانس ارائه‌دهنده] به افراد، مؤسسات یا سازمان‌های دیگر توزیع یا فروخته/ارائه نمی‌شوند. مطالبی که حق نسخه‌برداری (کپی رایت) ندارند و حاوی ریزداده نیستند (مانند پرسشنامه‌های پیمایش، راهنماها، کتاب‌های کد یا فرهنگ لغت داده‌ها) ممکن است بدون مجوز بیشتر توزیع شوند. مالکیت کلیه مواد ارائه شده توسط [آژانس ارائه‌دهنده] در اختیار [آژانس ارائه‌دهنده] باقی می‌ماند.

۲. داده‌ها فقط برای مقاصد آماری و تحقیقات علمی مورد استفاده قرار خواهند گرفت و صرفاً برای گزارش اطلاعات جمع‌آوری شده، از جمله مدل‌سازی، و نه برای تحقیق در مورد افراد یا سازمان‌های خاص استفاده خواهند شد.



۳. هیچ تلاشی برای شناسایی مجدد پاسخ‌دهندگان انجام نخواهد شد و از هویت هر شخص یا مؤسسه‌ای که به‌طور شناسایی شده باشد، استفاده نخواهد شد. هرگونه شناسایی بلافاصله به [آژانس ارائه‌دهنده] گزارش خواهد شد.
۴. هیچ تلاشی برای ایجاد ارتباط (لینک) بین مجموعه داده‌های ارائه شده توسط [آژانس ارائه‌دهنده] یا بین داده‌های [آژانس ارائه‌دهنده] و سایر مجموعه داده‌هایی که می‌تواند افراد یا سازمان‌ها را شناسایی کند، انجام نخواهد شد.
۵. هر کتاب، مقاله، مقاله کنفرانس، پایان نامه، رساله، گزارش، یا سایر انتشاراتی که از داده‌های به‌دست آمده از [آژانس ارائه‌دهنده] استفاده می‌کند، مطابق با الزامات استنادی ارائه شده با مجموعه داده، منبع را ذکر خواهد کرد.
۶. یک نسخه الکترونیکی از کلیه نشریات براساس داده‌های درخواستی به [آژانس ارائه‌دهنده] ارسال خواهد شد.
۷. [آژانس ارائه‌دهنده] و آژانس‌های تأمین مالی مربوطه هیچ مسئولیتی در مورد استفاده از داده‌ها یا تفسیر یا استنباط براساس آن ندارند.
۸. داده‌ها در یک محیط امن، با محدودیت‌های دسترسی کافی، ذخیره خواهند شد. [آژانس ارائه‌دهنده] می‌تواند در هر زمان اطلاعاتی در مورد امکانات ذخیره‌سازی و انتشار در [آژانس گیرنده] درخواست کند.
۹. [آژانس گیرنده] گزارشی سالانه در مورد استفاده‌ها و کاربران مجموعه‌های ریزداده فهرست شده با اطلاعاتی در مورد تعداد محققینی که به هر مجموعه داده دسترسی داشته‌اند و خروجی این تحقیق را به [سازمان ارائه‌دهنده] ارائه خواهد کرد.
۱۰. این دسترسی برای یک دوره اعطا می‌شود [اطلاعات مربوط به این دوره را ارائه دهید، یا اعلام کنید که توافقنامه پایان ناپذیر است].



ج- ارتباطات

[سازمان دریافت‌کننده] یک فرد در تماس را منصوب خواهد کرد که به عنوان شخص کانونی منحصر به فرد برای این توافقنامه عمل خواهد کرد. در صورتی که فرد کانونی جایگزین شود، [اژانس گیرنده] فوراً، نام و مختصات فرد در تماس جدید را به [اژانس ارائه‌دهنده] اطلاع می‌دهد. ارتباطات برای اهداف اداری و رویه‌ای ممکن است از طریق ایمیل، فکس یا پست پستی به شرح زیر انجام شود:

ارتباطات ایجاد شده توسط [اژانس ارائه‌دهنده] به [اژانس گیرنده] به این موارد هدایت می‌شود:

نام مسئول تماس:

عنوان مسئول تماس:

آدرس اژانس گیرنده:

پست الکترونیک:

تلفن:

فکس:

ارتباطات انجام شده توسط [اژانس گیرنده] به [اژانس سپرده‌گذار] به این آدرس هدایت می‌شود:

نام مسئول تماس:

عنوان مسئول تماس:

آدرس اژانس گیرنده:

پست الکترونیک:

تلفن:

فکس:



د- امضاکنندگان

امضاکنندگان زیر توافقنامه را که در بالا ارائه شد، خوانده و با آن موافقت کرده‌اند:

نماینده [آژانس ارائه‌دهنده]

..... نام

..... امضاء تاریخ

نماینده [آژانس گیرنده]

..... نام

..... امضاء تاریخ



فایل‌های قابل دسترسی در محفظه داده

برخی از فایل‌ها ممکن است تحت شرایط سخت در یک محصور داده به کاربران ارائه شوند. این مرکز مجهز به رایانه‌هایی است که به اینترنت یا شبکه خارجی متصل نیستند و هیچ اطلاعاتی از طریق پورت‌های USB، سی دی، دی وی دی، یا درایوهای دیگر قابل دانلود نیست. محفظه‌های داده حاوی داده‌هایی هستند که به ویژه حساس هستند یا امکان شناسایی مستقیم یا آسان پاسخ‌دهندگان را فراهم می‌کنند. نمونه‌ها شامل مجموعه داده‌های سرشماری کامل جمعیت، پیمایش‌های سازمانی، و مجموعه داده‌های مرتبط با سلامتی که حاوی اطلاعات بسیار محرمانه هستند، می‌شود.

کاربرانی که علاقه‌مند به دسترسی به یک محفظه داده هستند، ممکن است به مجموعه داده کامل دسترسی نداشته باشند، اما فقط به زیرمجموعه داده خاصی که نیاز دارند دسترسی داشته باشند. از آنها خواسته می‌شود فرم درخواستی را تکمیل کنند که نشان‌دهنده نیاز مشروع به دسترسی به این داده‌ها برای تحقق یک هدف آماری یا تحقیقاتی اعلام شده است (مثال آن در بخش‌های بعدی آورده شده است). خروجی‌های تولید شده باید قبل از انتشار از طریق بازبینی افشای کامل مورد بررسی قرار گیرند.

بهره‌برداری از یک محفظه داده ممکن است گران باشد. برای اطمینان از عدم وجود خطر افشا، به مکان‌های ویژه و تجهیزات رایانه‌ای و کارکنانی با مهارت‌ها و زمان برای بررسی خروجی‌ها قبل از حذف آنها از محفظه داده‌ها نیاز است. چنین کارکنانی باید با تجزیه و تحلیل داده‌ها آشنا باشند و بتوانند فرآیند درخواست را بررسی کرده و سرورهای فایل را مدیریت کنند.

به دلیل هزینه‌های عملیاتی قابل توجه و مهارت‌های فنی مورد نیاز، برخی از آژانس‌های آماری یا سایر تولیدکنندگان رسمی داده‌ها ترجیح می‌دهند با مؤسسات دانشگاهی یا مراکز تحقیقاتی برای ایجاد و مدیریت پایگاه‌های داده همکاری کنند. نمونه‌هایی از محفظه‌های داده با وبسایت‌های آموزنده عبارتند از: مرکز داده‌های تحقیقاتی سرشماری میشیگان^۱ (MCRDC)، پروژه مشترک اداره سرشماری ایالات متحده و دانشگاه میشیگان (www.isr.umich.edu/)، مرکز ملی تحقیقات افکارسنجی^۲ (NORC) در دانشگاه شیکاگو (www.norc.org/Da-)، برنامه مراکز داده تحقیقاتی^۳ (RDC) آمار کانادا (www.statcan.gc.ca/rdc cdr/index)؛ و مرکز داده تحقیقاتی NCHS ایالات متحده (<http://www.cdc.gov/nchs>).

1. Michigan Census Research Data Center (MCRDC)

2. National Opinion Research Center (NORC)

3. Research Data Centres (RDC)



منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

 Policy on Micro-data Dissemination (NCHS)

 Download

۷'۴ قیمت گذاری

آیا ریزداده باید فروخته شود یا رایگان ارائه شود؟

اگر تولید ریزداده برای انتشار بخشی از بودجه پیمایش نباشد، مرکز ممکن است انتخاب کمی داشته باشد جز این که تلاش برای بازبایی هزینه آن کند. یکی از روش‌هایی که تولیدکنندگان داده ممکن است در نظر بگیرند، فروش محصولات ریزداده است. از آنجایی که ادارات آمار مستقل تر شده‌اند، اغلب با الزام به شناسایی و اجرای فعالیت‌های درآمدزا مواجه هستند. ریزداده‌ها یکی از محصولات با ارزش بالا هستند. در حالی که در دسترس بودن فایل‌های ریزداده ارزش یک پیمایش را افزایش می‌دهد، هزینه‌هایی در فرآیند مستندسازی و بی‌نام‌سازی متحمل می‌شود.

یک سرویس کامل انتشار ریزداده شامل هزینه‌های دیگری مانند فرآیند مطالعه برای انتشار فایل‌ها، صدور مجوز فایل‌های داده، راه‌اندازی یک محفظه داده، پشتیبانی از کاربران و حفظ زیرساخت است. به‌طور کلی این نیاز به کارکنان و زیرساخت‌های اختصاصی دارد. اگر اینها در بودجه آژانس نباشد یا نتوان مسئولیت را به بخشی از سازمان واگذار کرد، ممکن است ارائه خدمات پایدار و با کیفیت دشوار باشد. فروش داده یکی از راه‌های متحمل شدن برخی از این هزینه‌ها توسط ذینفعان سرویس داده است.

رایگان یا با هزینه؟

هیچ پاسخ قطعی برای این سؤال وجود ندارد. دلایل زیادی برای به حداقل رساندن هزینه‌های دسترسی به فایل‌های ریزداده وجود دارد. این بدان معنا نیست که اصلاً نباید هزینه‌ای وجود داشته باشد.



رایگان

استدلال اصلی برای عدم دریافت هزینه این است که در بسیاری از موارد، مراکز در حال تقویت شیوه‌های انتشار داده‌های خود برای فایل‌های ریزداده هستند. آنها می‌ترسند که قیمت بسیار بالا مانعی برای جذب کاربران باشد.

- فروش داده‌ها به‌طور قابل ملاحظه‌ای تعداد کاربران بالقوه و در نتیجه ارزش واقعی داده‌ها را کاهش می‌دهد.
- به ویژه در کشورهای در حال توسعه، ممکن است مانعی برای کاربرانی باشد که بیشترین علاقه را به داده‌ها دارند، مانند دانشجویان، مراکز تحقیقاتی محلی، و دانشگاه‌ها.

یکی دیگر از استدلال‌ها علیه هزینه‌ها، هزینه جمع‌آوری آنها است. همچنین مهم است که هزینه‌ها به مراکز یا یک آژانس مرکزی بازگردد. این موضوع بر انگیزه‌های کارکنان برای بازیابی هزینه‌ها تأثیر می‌گذارد. برای کارکرد صحیح، سیستم باید کارآمد باشد.

- فروش، تعهد کیفیت و خدمات را تحمیل می‌کند.
- فروش، حداقل درآمد را ایجاد می‌کند. بیشتر تقاضاها از دانشگاه است، که منابع محدودی دارد و امکان انجام تحقیقات خود را در جای دیگر با داده‌های دیگر دارد.

در حالی که تجربه در کشورهای دیگر - اغلب توسعه‌یافته - نشان می‌دهد که تا حدودی، بازیابی هزینه ممکن می‌باشد، بعید است که تمام هزینه‌های بیشتر انتشار ریزداده‌ها قابل بازیابی باشد. نشانه‌هایی وجود دارد که بازیابی هزینه‌ها، استفاده از فایل‌های ریزداده را منع می‌کند و در دراز مدت، ارزش بالقوه پیمایش را کاهش می‌دهد.

مطلوب‌ترین رویکرد برای یک مرکز این است که تمام هزینه‌ها در بودجه پیمایش پوشش داده شود، که دسترسی را به حداکثر می‌رساند. چنین هزینه‌هایی می‌تواند توسط حامیان مالی به عنوان راهی برای به حداکثر رساندن مزایای پیمایش متحمل شود. این امر به ویژه در کشورهایی که محققان دارای منابع مالی محدود و تولیدکنندگان دارای منابع کمی برای تجزیه و تحلیل هستند، اهمیت دارد.

هزینه

این احتمال زیاد است که شارژ برای فایل‌های داده مقداری درآمد ایجاد کند. اما هنگام تصمیم‌گیری در مورد شارژ



یا عدم شارژ، عوامل دیگری مانند نکات زیر نیز باید در نظر گرفته شوند:

- آیا مرکز حق قانونی دریافت هزینه برای محصولات خود را دارد؟
- مرکز مایل است کدام هزینه‌ها را جبران کند؟
- آیا این هزینه‌ها قابل شناسایی هستند، یعنی کاربران آنها را درک کرده و می‌پذیرند؟
- آیا کاربران می‌توانند هزینه پرداخت کنند؟ آیا می‌توان یک کنسرسیوم کاربری برای بازیابی هزینه‌ها از قبل تشکیل داد؟ این مستلزم شناسایی هزینه‌هایی است که باید بازیابی شوند و آنها را بین سازمان‌های کاربر تقسیم می‌کنند.
- آیا می‌توان هزینه‌ها را به‌طور مؤثر جمع‌آوری کرد؟
- فایل‌های پیچیده‌ای که به صورت رایگان در یک وبسایت در دسترس هستند ممکن است منجر به دسترسی افرادی شود که توانایی استفاده از ریزداده را ندارند که می‌تواند منجر به افزایش تقاضا برای پشتیبانی شود.

یکی دیگر از ملاحظات مهم در هنگام ارائه یک استراتژی قیمت‌گذاری، همسویی آن با سیاست/فلسفه قیمت‌گذاری برای سایر محصولات مانند انتشارات کاغذی و دسترسی به اینترنت است. اکثر وبسایت‌های مرکز به‌طور رایگان در دسترس کاربران هستند، از آنجایی که هزینه‌های فزاینده^۱ کمی وجود دارد. با این حال، این موضوع در مورد تولیدات کاغذی^۲ صادق نیست. اگر قیمت‌گذاری انتشار براساس پوشش هزینه‌های فزاینده تولید و ارسال نسخه‌های بیشتر باشد، همین اصل می‌تواند برای فایل‌های ریزداده و هزینه بیشتر پشتیبانی یا خدمات به کاربران دیگر اعمال شود.

تجربه برخی کشورها

سابقه طولانی بازیابی هزینه توسط مراکز در سراسر جهان وجود دارد. در پاسخ به کاهش بودجه و فشار ادارات مرکزی، در دهه ۱۹۸۰، انتقال بار هزینه آمار از مالیات‌دهندگان به کاربر رایج شد. در حالی که بررسی کامل فعالیت‌های بازیابی هزینه مراکز خارج از محدوده این سند است، در اینجا چند مشاهده وجود دارد:

- **اداره آمار نیوزلند^۳ (SNZ)** متعهد شد ۲۵ درصد از کل بودجه خود را از فروش محصولات و خدمات بازیابی کند. این در زمانی اتفاق افتاد که نیوزلند تحت یک بررسی اداری دقیق

1. Incremental costs

2. Paper products

3. Statistics New Zealand (SNZ)



قرار گرفت. در حالی که این تعهد مانع کاهش برنامه شد، نتوانست به هدف خود برسد و در نهایت انجام این کار نامناسب تشخیص داده شد. SNZ به جای تلاش برای بازیابی هزینه‌ها، یک استراتژی کاهش هزینه را در پیش گرفته است. وبسایت SNZ نشان می‌دهد که حدود ۹۰ درصد از اطلاعات آنها به صورت رایگان در دسترس است؛ هزینه‌هایی برای جدول‌های سفارشی و جداول دقیق اخذ می‌شود. برای دسترسی به فایل‌های ریزداده هزینه‌هایی وجود دارد، اما این هزینه‌ها در وبسایت مشخص نشده است.

- **اداره آمار کانادا** برنامه جامع بازیابی هزینه‌ها را در دهه ۱۹۸۰ در پاسخ به فشارهای بودجه‌ای و سیاسی اجرا کرد. علی‌رغم مخالفت شدید کاربران، این چالش را مثبت پذیرفت. این برنامه با موفقیت ناموزون روبه‌رو شد و بسیاری از بخش‌های آن حذف شده یا بسیار کاهش یافته است. هنگامی که هزینه محصولات ریزداده به میزان قابل توجهی افزایش یافت، محققان کانادایی شروع به استفاده از داده‌های ایالات متحده کردند، که داده‌های آن به راحتی از طریق کنسرسيوم بین دانشگاهی برای تحقیقات سیاسی و اجتماعی (ICPSR) در دسترس بود؛ ابتکار آزادسازی داده (DLI) زمانی خلق شد که مشخص شد این امر در خدمت اهداف کانادا نیست. نقل قول زیر انگیزه DLI را فراهم کرد: "... اجرای واقعی دموکراسی بیشتر مستلزم آن است که شهروندان به اطلاعات پیچیده دسترسی داشته باشند و مهارت‌های لازم برای درک آن را دارا باشند." در حالی که فشارهایی بر اداره آمار کانادا برای کاهش هزینه‌ها و افزایش درآمد وجود دارد، احساس می‌کند که نتیجه آن محدودیت ... "دسترسی به اطلاعات فقط برای گروه‌هایی است که توانایی کامل پرداخت را دارند Bernard". احساس می‌کند که این ممکن است ... "مشارکت در بحث‌های عمومی گروه‌هایی را که مشارکت آنها با پول زیادی پشتیبانی نمی‌شود" و همچنین "افرادی که هیچ چشم‌اندازی برای کسب سود یا به‌دست آوردن منافع ملموس و نسبتاً فوری از استفاده ندارند، مختل کند." به گفته او، این "... احتمالاً در درازمدت به توسعه نامطلوب دموکراسی ناکامل منجر خواهد شد." این منجر به ایجاد شبکه‌ای از ۷۴ مؤسسه آموزشی با دسترسی به تمام داده‌های عمومی از آمار کانادا شد. این داده‌ها شامل تقریباً ۳۰۰ فایل ریزداده استفاده عمومی و هزاران فایل، پایگاه داده و فایل‌های جغرافیایی دیگر است. هزینه اشتراک، هزینه‌های پشتیبانی و همچنین ارائه یک زیرساخت فنی پیشرفته را پوشش می‌دهد که به مشترکین و همچنین آژانس، خدمات‌رسانی می‌کند. برای پوشش، هزینه داده‌ها در نظر گرفته نشده است. پورتال دسترسی



DLI برای ادارات دولتی در دسترس نیست، بسیاری از آنها توافقاتی با اداره آمار کانادا برای تقسیم هزینه‌های برخی پیمایش‌ها دارند. به‌طور کلی، دسترسی آنها به ریزداده‌های تحت پوشش چنین قراردادهایی است و بنابراین ممکن است به این سرویس نیازی نداشته باشند.



The Canadian Data Liberation Initiative: An Idea worth Considering?

Ernie Boyko and Wendy Watkins - November 2011

The Data Liberation Initiative (DLI) is a Canadian program that provides Canadian post-secondary institutions affordable access to Statistics Canada data resources. This paper describes the project and key success factors so that other agencies may determine its applicability to their own situations.

[Download](#)

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

 National Statistics Code of Practice - Protocol on Data Presentation, Dissemination and Pricing (UK)

 [Download](#)

۸'۴ زمان انتشار داده‌ها

فایل‌های ریزداده چه زمانی در چرخه انتشار باید منتشر شوند؟

تولید و انتشار فایل‌های ریزداده باید بخشی از چرخه انتشار باشد. یک استدلال قوی وجود دارد که از انتشار



اطلاعاتی که ابتدا مخاطبان گسترده را هدف قرار می‌دهد، حمایت می‌کند تا مرکز بتواند به اهداف فوری خود دست یابد و بازخورد عمومی را ارائه دهد. چنین اطلاعاتی شامل گزارش‌های پیمایشی توصیفی و تجزیه‌وتحلیل توسط تولیدکننده داده است. برای تولیدکنندگان رسمی داده، ایجاد این نتایج رسمی و انتشار آنها در ابتدای چرخه انتشار مهم است.







تولید یک فایل ریزداده به زمان بیشتری نیاز دارد، از منابع تخصصی استفاده می‌کند که باید برنامه‌ریزی شود، و به نوعی بررسی نیاز دارد. مرکز ممکن است نیاز داشته باشد که اهداف تحقیقاتی/تحلیلی داخلی یا خارجی خاصی را برآورده کند و بنابراین ممکن است ایجاد یک فایل ریزداده را برای چند ماه پس از انتشار رسمی نتایج پیمایش به تأخیر بیندازد. صرف نظر از جدول زمانی، محققان دوست دارند از تاریخ‌های انتشار برنامه‌ریزی شده مطلع شوند تا بتوانند کار خود را برنامه‌ریزی کنند. تأخیر باید معقول باشد؛ نتایج چندین ساله باعث می‌شود که نتایج بسیار کمتر مرتبط باشند.

«حتی زمانی که ریزداده‌ها تا حد امکان به سرعت منتشر می‌شوند، موقعیت‌هایی وجود دارد که انتشار بخشی از ریزداده‌ها یا داده‌های جمع‌آوری شده قبل از زمانی که مجموعه کامل ریزداده‌ها می‌تواند در دسترس قرار گیرد، مفید است. چنین الزاماتی [باید] در مراحل برنامه‌ریزی گنجانده شود یا به محض آشکار شدن نیاز آنها مطرح شود.» (NCHS، ۲۰۰۲. "سیاست انتشار ریزداده‌ها.")

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

- 
[National Statistics Code of Practice - Protocol on Data Presentation, Dissemination and Pricing \(UK\)](#)
 Download
- 
[National Statistics Code of Practice - Protocol on Release Practices \(UK\)](#)
 Download
- 
[Policy on Micro-data Dissemination \(NCHS\)](#)
 Download




۹'۴ الزام استناد

برای ارائه اعتبار مناسب به تولیدکنندگان داده برای ورودی آنها در خروجی های تجزیه و تحلیل داده‌ها، و اطمینان از شفافیت لازم برای تضمین "تکرارپذیری" کار تحلیلی، ضروری است که داده‌های استفاده شده توسط محققان به درستی استناد شوند. این همچنین به مخازن داده اجازه می‌دهد تا کاتالوگ‌هایی از اسنادها را نگهداری کنند (مانند فهرستی که توسط شبکه بین‌المللی پیمایش‌های خانوار نگهداری می‌شود)، که برای اندازه‌گیری و ارزیابی استفاده از داده‌های موجود مفید است، و به نوبه خود می‌تواند توسط تولیدکنندگان داده برای توجیه سرمایه‌گذاری در جمع‌آوری داده‌ها استفاده شود.

اسنادها باید اطلاعاتی در مورد منبع و نسخه مجموعه داده ارائه دهند. "نیاز به استناد" یکی از عناصر استاندارد فراداده طرح مستندسازی داده‌ها است. اسناد باید حداقل شامل اطلاعات مربوط به محقق اولیه، عنوان و مخفف مجموعه داده، سال(های مرجع)، شماره نسخه و شماره شناسه مجموعه داده باشد. برای داده‌های به‌دست‌آمده به‌صورت آنلاین، می‌تواند شامل نشانی اینترنتی وبسایتی که داده‌ها از آن دانلود شده‌اند و تاریخ دانلود نیز باشد.

برای شناسه مجموعه داده، استفاده از شناسه دیجیتال شیء^۱ (DOI) توصیه می‌شود.



DataCite

DataCite is a not-for-profit organization formed in London on 1 December 2009 to: (i) establish easier access to research data on the Internet, (ii) increase acceptance of research data as legitimate, citable contributions to the scholarly record; and (iii) support data archiving that will permit results to be verified and re-purposed for future study.

[Download](#)

۱۰'۴ تدوین خط مشی

در همه کشورها، تولیدکنندگان داده با تقاضای رو به رشد برای ریزداده مواجه هستند. تعیین بهترین راه برای انتشار این داده‌ها یک چالش است. چالش، فنی است، زیرا تولیدکنندگان داده باید رویه‌هایی را برای مستندسازی، کاتالوگ کردن و انتشار داده‌ها اجرا کنند. برخی چالش نیز قانونی و اخلاقی است. در حالی که تولیدکنندگان داده اغلب به

1. Digital Object Identifier (DOI)



خوبی از قدرت و اهمیت ریزداده آگاه هستند، باید این تقاضا را با نیاز به محرمانه نگه داشتن اطلاعات پاسخگو متعادل کنند. این یک الزام قانون آماری و حریم خصوصی یک کشور است و اغلب تعهدی است که هنگام جمع‌آوری اطلاعات به پاسخ‌دهندگان داده می‌شود. این امر ایجاد سیاست‌ها و رویه‌های رسمی را برای تعیین شرایط دسترسی به ریزداده‌ها را تحمیل می‌کند.



Dissemination of Microdata Files - Principles, Procedures and Practices

Olivier Dupriez and Ernie Boyko - August 2010

In all countries, data producers face growing demand for microdata. Determining the best way to disseminate these data is a challenge. Formal policies and procedures defining the conditions of access to microdata must be formulated. This guide provides an overview of such policies and procedures, and documents existing best practices.
[Download \(2.05 MB\)](#)



Diffusion des fichiers de microdonnées - Principes, procédures et pratiques

Olivier Dupriez and Ernie Boyko - August 2010









Les producteurs de données de tous pays sont confrontés à une demande croissante de microdonnées. Décider de la meilleure façon de diffuser ces données constitue un véritable défi. Cela implique l'établissement de politiques et de procédures définissant les conditions d'accès aux microdonnées. Le présent document contient une description générale de ces politiques et de ces procédures, et recense les bonnes pratiques en la matière.
[Download](#)

منابع مرتبط به اشتراک گذاشته شده است:



Related Resources

Documents

 Diffusion des fichiers de microdonnées - Principes, procédures et pratiques	 (1.42 MB)
 Difusión de archivos de microdatos - Principios, procedimientos y prácticas	 (1.17 MB)
 Dissemination of Microdata Files - Principles, Procedures and Practices	 (2.05 MB)
 Managing Statistical Confidentiality and Microdata Access - Principles and guidelines of Good Practice	 Download
 National Statistics Code of Practice - Protocol on Data Presentation, Dissemination and Pricing (UK)	 Download
 NCHS Staff Manual on Confidentiality	 Download
 Policy on Micro-data Dissemination (NCHS)	 Download
 بتّ و نشر ملفات البيانات الجزئية - مبادئ وأنظمة وممارسات	 (757.68 KB)



۵. محافظت

تولیدکنندگان داده‌های رسمی (سازمان‌های آماری ملی و وزارتخانه‌های خطی) هر ساله حجم وسیعی از داده‌ها و اطلاعات را به صورت دیجیتال از جمله، ریزداده‌ها از پیمایش‌ها و سرشماری‌ها، محتوای سیستم‌های ثبت اداری، پایگاه‌های اطلاعاتی شاخص‌ها، چارچوب‌های نمونه، ثبت‌ها، گزارش‌های روش‌شناسی و تحلیلی، نشریات، نقشه‌ها، و انواع دیگر محتوا تولید می‌کنند. این دارایی‌های دیجیتال نشان‌دهنده سرمایه‌گذاری قابل توجه تولیدکنندگان است، ارزش قابل توجهی برای کاربران فعلی و آینده دارد و باید حفظ شوند. متأسفانه، نمونه‌های متعددی از داده‌های مهم وجود دارد که در غیاب یک روش حفاظتی مؤثر از بین رفته‌اند. به عنوان مثال، داده‌ها بر روی نوارهای منسوخ، ناخوانا و سایر فناوری‌های منسوخ ذخیره شده‌اند. یک برنامه محافظت پایدار به مسائل سازمانی، نگرانی‌های تکنولوژیکی، و سؤالات تأمین مالی می‌پردازد. >P

- **زیرساخت‌های سازمانی** شامل سیاست‌ها، رویه‌ها، شیوه‌ها و افراد است - عناصری که هر برنامه‌ای برای پیشرفت به آن‌ها نیاز دارد، اما آنها باید برای رسیدگی به الزامات محافظت دیجیتالی، تخصصی باشند. زیرساخت به این سؤال کلیدی توسعه می‌پردازد: الزامات و پارامترهای برنامه محافظت دیجیتال سازمان چیست؟
- **زیرساخت‌های فناوری** شامل تجهیزات، نرم‌افزار، سخت‌افزار، محیط امن و مهارت‌های لازم برای ایجاد و نگهداری برنامه محافظت دیجیتال است. فناوری در حال تغییر را پیش‌بینی می‌کند و عاقلانه به آن پاسخ می‌دهد. به این سؤال کلیدی توسعه می‌پردازد: سازمان چگونه الزامات تعریف شده محافظت دیجیتال را برآورده می‌کند؟
- **چارچوب منابع** به بودجه لازم برای راه‌اندازی، مستمر، و اضطراری برای فعال کردن و حفظ برنامه محافظت دیجیتال می‌پردازد. به این سؤال کلیدی توسعه می‌پردازد: چه منابعی برای توسعه و حفظ برنامه کحفاظت دیجیتال مورد نیاز است؟

شبکه بین‌المللی پیمایش‌های خانوار، دستور تهیه راهنمای زیر را با عنوان "اصول و روش خوب برای حفظ داده‌ها" در مورد محافظت داده‌های دیجیتالی داده است.



Principles and Good Practice for Preserving Data

Interuniversity Consortium for Political and Social Research (ICPSR) -
December 2009

This document provides basic guidance for managers in statistical agencies who are responsible for preserving data using the principles and best practices defined by the digital preservation community. The document defines the rationale for preserving data and the principles and standards applied to data preservation; documents the development of a digital preservation policy; and uses digital archive audit principles to suggest best practices for data.

[Download](#)

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

- | | |
|---|--|
|  National Statistics Code of Practice - Protocol on Data Management, Documentation and Preservation (UK) |  Download |
|  Principles and Good Practice for Preserving Data |  (1.31 MB) |

Links

- | | |
|---|--|
|  The Data Seal of Approval (DSA) |  Download |
|---|--|



۶. ترتیبات سازمانی

انتشار فایل‌های ریزداده ممکن است برای بسیاری از مراکز فعالیت جدیدی باشد و می‌تواند منجر به استفاده‌های جدید و مهم از داده‌های آنها شود. سازمان همکاری و توسعه اقتصادی (OECD) در انتشار خود، اصول و دستورالعمل‌های دسترسی به داده‌های تحقیقاتی از بودجه عمومی، مجموعه‌ای از اصولی را تعریف می‌کند که ارائه‌دهندگان داده باید به آن پایبند باشند:

- **باز بودن^۱**

باز بودن به معنای دسترسی با شرایط مساوی برای جامعه تحقیقاتی بین‌المللی با کمترین هزینه ممکن است، ترجیحاً با هزینه نهایی انتشار. (...)

- **انعطاف‌پذیری**

انعطاف‌پذیری مستلزم در نظر گرفتن تغییرات سریع و اغلب غیرقابل پیش‌بینی در فناوری‌های اطلاعات، (...)، نظام‌های حقوقی و فرهنگ‌های هر (... کشور است. (...)

- **شفافیت**

اطلاعات مربوط به داده‌های پژوهشی و سازمان‌های تولیدکننده داده، اسناد داده‌ها و مشخصات شرایط مرتبط با استفاده از این داده‌ها باید در سطح بین‌المللی به روشی شفاف و ایده‌آل از طریق اینترنت در دسترس باشد. (...)

- **مطابقت قانونی**

ترتیبات دسترسی به داده‌ها باید به حقوق قانونی و منافع مشروع همه ذینفعان احترام بگذارد. (...)

- **حفاظت از مالکیت معنوی**

ترتیبات دسترسی به داده‌ها باید قابلیت اعمال حق نسخه‌برداری یا سایر قوانین مالکیت معنوی را که ممکن است مربوط به پایگاه‌های اطلاعاتی تحقیقاتی با بودجه عمومی باشد، در نظر بگیرند. (...)

- **مسئولیت رسمی**

1. Openness



ترتیبات دسترسی باید شیوه‌های سازمانی رسمی و صریح، مانند تدوین قوانین و مقررات، در رابطه با مسئولیت‌های طرف‌های مختلف درگیر در فعالیت‌های مرتبط با داده‌ها را ترویج دهند. این شیوه‌ها باید به تألیف، اعتبار تولیدکننده، مالکیت، انتشار، محدودیت‌های استفاده، ترتیبات مالی، قوانین اخلاقی، شرایط مجوز، مسئولیت، و بایگانی پایدار مربوط باشد. (...)

- حرفه‌ای‌گری^۱

ترتیبات سازمانی برای مدیریت داده‌های پژوهشی باید براساس استانداردها و ارزش‌های حرفه‌ای مرتبط باشد که در آیین‌نامه‌های رفتاری جوامع علمی درگیر گنجانده شده است. (...)



• قابلیت همکاری

قابلیت همکاری فن‌آوری و معنایی یک ملاحظه کلیدی در ایجاد و ارتقای دسترسی بین‌المللی و بین‌رشته‌ای به استفاده از داده‌های تحقیقاتی است. ترتیبات دسترسی، باید توجه لازم را به استانداردهای اسناد بین‌المللی داده‌های مربوطه داشته باشد. (...)

• کیفیت

ارزش و کاربرد داده‌های تحقیق تا حد زیادی به کیفیت خود داده‌ها بستگی دارد. مدیران داده‌ها و سازمان‌های جمع‌آوری داده‌ها باید توجه ویژه‌ای به اطمینان از انطباق با استانداردهای صریح کیفیت داشته باشند. (...)

• امنیت

توجه ویژه باید به حمایت استفاده از تکنیک‌ها و ابزارها برای تضمین یکپارچگی و امنیت (... داده شود. (...)

• بهره‌وری

یکی از اهداف اصلی ارتقای دسترسی و اشتراک‌گذاری داده‌ها، بهبود کارایی کلی منابع مالی عمومی (جمع‌آوری داده‌ها) برای جلوگیری از تکرار پرهزینه و غیرضروری تلاش‌های جمع‌آوری داده‌ها است. (...)

• مسئولیت

عملکرد ترتیبات دسترسی به داده‌ها باید در معرض ارزیابی دوره‌ای گروه‌های کاربر، مؤسسات مسئول و (... آژانس‌های تأمین مالی باشد. (...)



OECD Principles and Guidelines for Access to Research Data from Public Funding

Organisation for Economic co-Operation and Development (OECD) - 2007

The Principles and Guidelines apply to research data gathered using public funds for the purposes of producing publicly accessible knowledge. The nature of "public funding" of research varies significantly from one country to the next, as do existing data access policies and practices at the national, disciplinary and institutional levels.

[Download](#)

این احتمال وجود دارد که رعایت این اصول، نیازمند رویه‌ها و اهداف جدید باشد. سازمان همکاری و توسعه اقتصادی (OECD) در دستورالعمل‌های خود، مسائل اصلی ذاتی در ارائه دسترسی به داده‌ها را به شرح زیر تعریف می‌کند (این مسائل در مورد ریزداده‌های آماری تولید شده برای اهداف آماری توسط تولیدکنندگان رسمی داده اعمال می‌شود):

- **نهادی و مدیریتی:** در حالی که افزایش دسترسی برای همه جوامع علمی مهم است، تنوع شرکت‌های علمی نشان می‌دهد که انواع مدل‌های سازمانی و روش‌های مدیریت داده‌های متناسب، در برآوردن نیازهای محققان مؤثرترین هستند.
- **مالی و بودجه‌ای:** زیرساخت داده‌های علمی مستلزم برنامه‌ریزی بودجه‌ای مستمر و اختصاصی و حمایت مالی مناسب است. در صورتی که هزینه‌های دسترسی، مدیریت، و حفاظت و نگهداری در پروژه‌های تحقیقاتی اضافه شده یا پس از آن در نظر گرفته شود، استفاده از داده‌های تحقیقاتی به حداکثر نمی‌رسد. البته ذکر این نکته ضروری است که هزینه ذخیره و مدیریت داده‌ها در سال‌های اخیر به شدت کاهش یافته است و عدم آگاهی در مورد چنین تغییراتی به خودی خود می‌تواند مانعی برای پیشرفت باشد.
- **حقوقی و سیاست:** قوانین ملی و موافقت‌نامه‌های بین‌المللی، به ویژه در زمینه‌هایی مانند حقوق مالکیت معنوی و حفاظت از حریم خصوصی، مستقیماً بر روی دسترسی به داده‌ها و شیوه‌های اشتراک‌گذاری تأثیر می‌گذارند و باید در طراحی ترتیبات دسترسی به داده‌ها کاملاً در نظر گرفته شوند.
- **فرهنگی و رفتاری:** ساختارهای آموزشی و پاداش مناسب یک جزء ضروری برای ترویج دسترسی



به داده‌ها و شیوه‌های اشتراک‌گذاری است. "این ملاحظات برای کسانی که داده‌ها را تأمین مالی، تولید، مدیریت و استفاده می‌کنند، اعمال می‌شود. (... مسئولیت جنبه‌های مختلف دسترسی و مدیریت داده‌ها باید در اسناد مربوطه، مانند شرح وظایف رسمی مؤسسات، درخواست‌های کمک مالی، قراردادهای تحقیقاتی، موافقت‌نامه‌های انتشار و مجوزها مشخص شود".

(اصول و رهنمودهای سازمان همکاری و توسعه اقتصادی (OECD) برای دسترسی به داده‌های تحقیقاتی از بودجه عمومی. ۲۰۰۷)

یک جایگزین: مخازن داده‌های قابل اعتماد

برای برخی از تولیدکنندگان داده، ایجاد و نگهداری یک آرشیو داده و سرویس انتشار ممکن است به دلایل بودجه‌ای، قانونی یا هدف دیگر غیرواقعی باشد. یک جایگزین این است که به یک آرشیو داده‌های موجود مانند بایگانی داده‌های بریتانیا واقع در دانشگاه Essex اعتماد کنید. این سازمان، داده‌های آژانس‌های آماری، سازمان‌های تحقیقاتی و خود محققین را مدیریت و منتشر می‌کند. مثال دیگر، کنسرسیوم بین دانشگاهی برای تحقیقات سیاسی و اجتماعی (ICPSR) در دانشگاه میشیگان است که عملکرد مشابهی را در ایالات متحده انجام می‌دهد. این آرشیو داده‌ها نه تنها مدیریت مجوزهای شایسته را ارائه می‌کنند، بلکه در مدیریت داده‌ها و نوآوری نیز پیشرو هستند. به عنوان مثال، مجموعه جدید صفحات وب آرشیو داده‌های بریتانیا است که راهنمایی‌هایی در مورد مدیریت و اشتراک‌گذاری داده‌ها ارائه می‌دهد. اینها بهترین راهبردها و روش‌ها را برای ایجاد، آماده‌سازی و ذخیره مجموعه داده‌های قابل اشتراک‌گذاری در اختیار سازندگان داده، مدیران داده و متصدیان داده قرار می‌دهند.

منابع مرتبط به اشتراک گذاشته شده است:



Related Resources

Documents

-  OECD Principles and Guidelines for Access to Research Data from Public Funding  [Download](#)
-

Links

-  The Data Seal of Approval (DSA)  [Download](#)
-



۷. الزامات فنی

"دسترسی به [ریز داده‌ها] و بهره‌برداری بهینه از آن‌ها نیازمند زیرساخت‌های فناوری طراحی شده مناسب، توافق‌نامه بین‌المللی گسترده در مورد قابلیت همکاری، و کنترل‌های مؤثر کیفیت داده است. (...)

پایداری درازمدت زیرساخت مورد نیاز برای دسترسی به داده‌ها از اهمیت ویژه‌ای برخوردار است. مؤسسات تحقیقاتی و سازمان‌های دولتی باید مسئولیت رسمی جهت اطمینان از این‌که (... داده‌ها به‌طور مؤثر، حفظ، مدیریت و در دسترس به منظور استفاده کارآمد و مناسب در طولانی‌مدت باشند، به عهده بگیرند. (...)

توجه ویژه (همچنین) باید به حمایت از استفاده از تکنیک‌ها و ابزارهایی برای تضمین یکپارچگی و امنیت داده‌های تحقیق اختصاص داده شود. با توجه به تضمین یکپارچگی یک مجموعه داده، باید تمام تلاش خود را برای اطمینان از کامل بودن داده‌ها و عدم وجود خطا انجام داد. با توجه به امنیت، داده‌ها همراه با متا داده‌ها و توضیحات مربوطه باید در برابر از دست دادن عمدی یا غیرعمدی، تخریب، تغییر و دسترسی غیرمجاز مطابق با پروتکل‌های امنیتی صریح، محافظت شوند."

یک زیرساخت فن‌آوری مناسب باید برای اجزای مختلف آرشیو ریز داده‌ها، یعنی مستندسازی داده‌ها، کاتالوگ کردن، انتشار، بی‌نام‌سازی و حفاظت ایجاد شود.

مستندات ریز داده

استانداردهای بین‌المللی فراداده برای مستندسازی ریز داده‌ها و منابع مرتبط ایجاد شده است. طرح مستندسازی داده‌ها (DDI) و استانداردهای Dublin Core یک راه حل عملی ارائه می‌دهند. مستندسازی مجموعه داده‌ها مطابق با این استانداردها با در دسترس بودن ویرایشگرهای فراداده تخصصی مانند نرم‌افزار (IHSN Microdata Management Toolkit Nesstar Publisher) آسان می‌شود.

کاتالوگ کردن و انتشار ریز داده‌ها

کاربران علاقه‌مند باید از وجود و ویژگی‌های مجموعه داده‌های موجود مطلع شوند. بسیاری از کاربران بالقوه، اطلاعات بسیار کمی در مورد مجموعه داده‌های موجود خواهند داشت. فراداده خوب باید در دسترس باشد، ترجیحاً در قالب یک کاتالوگ آنلاین قابل جستجو باشد. هدف کاتالوگ ریز داده، فراهم کردن دسترسی آسان به داده‌ها و اسناد در قالبی مناسب برای کاربران است. کاتالوگ کردن مجموعه داده‌ها مطابق با استانداردهای بین‌المللی فراداده با در دسترس بودن نرم‌افزار متن باز ارائه شده توسط شبکه بین‌المللی پیمایش‌های خانوار آسان شده است.



بی‌نام‌سازی ریز داده‌ها

بی‌نام‌سازی داده‌ها مستلزم آن است که کارکنان در مورد آمار و بسته‌های نرم‌افزاری مانند Stata یا SPSS آگاهی داشته باشند. برخی از نرم‌افزارهای تخصصی برای اندازه‌گیری یا کاهش خطر افشا در دسترس هستند، اما هیچ یک از این برنامه‌ها راه حل یکپارچه و رضایت بخشی را برای فایل‌های داده سلسله مراتبی پیچیده ارائه نمی‌دهند. بی‌نام‌سازی ریزداده‌ها شامل دو مرحله است: تشخیص موارد احتمالی خطر افشا و نوعی کاهش یا بهم‌ریختگی داده‌ها برای کاهش خطر. این مرحله اخیر، مستلزم ورودی، از شخصی با دانش موضوعی است که می‌تواند کاهش داده^۱ (داده‌کاوی) را توصیه کند که کمترین آسیب را برای محققانی که از فایل‌ها استفاده می‌کنند، داشته باشد. نرم‌افزار و دستورالعمل‌های تخصصی موجود است.

حفظ ریزداده (و فراداده)

داده‌های دیجیتال و فراداده‌ها در برابر منسوخ شدن نرم‌افزار، فرسودگی سخت‌افزار و رسانه، تهدیدات فیزیکی و خطاهای انسانی آسیب‌پذیر هستند. بنابراین حفظ طولانی‌مدت داده‌ها و فراداده‌ها نیازمند رویه‌ها و زیرساخت‌های مناسب است. دستورالعمل‌های شبکه بین‌المللی پیمایش‌های خانوار را ببینید.

منابع مرتبط به اشتراک گذاشته شده است:

Related Resources

Documents

 OECD Principles and Guidelines for Access to Research Data from Public Funding  [Download](#)



مقالات شبکه بین‌المللی پیمایش‌های خانوار

مجموعه مقاله شبکه بین‌المللی پیمایش‌های خانوار برای تشویق تبادل نظر و بحث در مورد موضوعات مرتبط با طراحی و اجرای پیمایش‌های خانوار و تجزیه و تحلیل، انتشار و استفاده از داده‌های پیمایش در نظر گرفته شده است.



Assessment of the Reliability and Relevance of the Food Data Collected in National Household Consumption and Expenditure Surveys

Lisa C. Smith, Olivier Dupriez and Nathalie Troubat - February 2014

Food consumption data are collected in most countries through a variety of household surveys. The primary objective of these surveys is usually to measure poverty, to derive consumption patterns needed for the calculation of consumer price indices, or to provide input to the compilation of national accounts. Increasingly, these data are repurposed and used to calculate food security indicators, to compile food balance sheets, to plan and monitor food-based nutrition interventions, to serve information needs of the private sector, and for other research work. This report is thus based on a desk review of survey questionnaires and methods, not on an assessment of the data themselves.

[Download](#)



Introduction to Statistical Disclosure Control (SDC)

Matthias Templ, Bernhard Meindl, Alexander Kowarik and Shuang Chen - August 2014

This guide, Introduction to Statistical Disclosure Control (SDC), discusses common SDC methods for microdata obtained from sample surveys, censuses and administrative sources.

[Download](#)



The Canadian Data Liberation Initiative: An Idea worth Considering?

Ernie Boyko and Wendy Watkins - November 2011

The Data Liberation Initiative (DLI) is a Canadian program that provides Canadian post-secondary institutions affordable access to Statistics Canada data resources. This paper describes the project and key success factors so that other agencies may determine its applicability to their own situations.

[Download](#)



Dissemination of Microdata Files - Principles, Procedures and Practices

Olivier Dupriez and Ernie Boyko - August 2010

In all countries, data producers face growing demand for microdata. Determining the best way to disseminate these data is a challenge. Formal policies and procedures defining the conditions of access to microdata must be formulated. This guide provides an overview of such policies and procedures, and documents existing best practices.

[Download \(2.05 MB\)](#)



Diffusion des fichiers de microdonnées - Principes, procédures et pratiques

Olivier Dupriez and Ernie Boyko - August 2010

Les producteurs de données de tous pays sont confrontés à une demande croissante de microdonnées. Décider de la meilleure façon de diffuser ces données constitue un véritable défi. Cela implique l'établissement de politiques et de procédures définissant les conditions d'accès aux microdonnées. Le présent document contient une description générale de ces politiques et de ces procédures, et recense les bonnes pratiques en la matière.

[Download](#)



Difusión de archivos de microdatos - Principios, procedimientos y prácticas

Olivier Dupriez y Ernie Boyko - Agosto de 2010

Los productores de datos de todos los países se enfrentan a una creciente demanda de microdatos. Decidir cuál es la mejor forma de difundir estos datos constituye un verdadero desafío. En este contexto, la difusión de microdatos implica la creación de políticas y procedimientos formales que definan las condiciones de acceso a los microdatos. El presente documento contiene una descripción general de estas políticas y procedimientos e identifica buenas prácticas existentes en este ámbito.

[Download](#)



بیّن و نشر ملفات البيانات الجزئية - مبادئ و أنظمة و ممارسات

أوليفيه دوبريز و إيرني بويكو - أغسطس / آب 2010

في جميع الدول، منتجي البيانات يواجهون طلب متزايد على البيانات الجزئية. تحديد ووضع أفضل الطرق لنشر هذه البيانات يعتبر تحدي. السياسات والإجراءات الرسمية التي تحدد شروط الوصول والنفاد للبيانات الجزئية والحصول عليها يجب أن تصاغ وتكون واضحة. يقدم هذا الدليل نظرة عامة عن هذه السياسات والإجراءات، ويحتوي هذا الدليل كذلك على الممارسات الجيدة للتعامل مع البيانات الجزئية.

[Download](#)



The Struggle for Integration and Harmonization of Social Statistics in a Statistical Agency - A Case Study of Statistics Canada

Gordon Priest - June 2010

IHSN Working Paper No 004. This paper describes how external influences and a corporate strategy lead to partial coordination of surveys and data collection practices at Statistics Canada.

[Download](#)



Principles and Good Practice for Preserving Data

Interuniversity Consortium for Political and Social Research (ICPSR) - December 2009

This document provides basic guidance for managers in statistical agencies who are responsible for preserving data using the principles and best practices defined by the digital preservation community. The document defines the rationale for preserving data and the principles and standards applied to data preservation; documents the development of a digital preservation policy; and uses digital archive audit principles to suggest best practices for data.

[Download](#)



How (well) is Education Measured in Household Surveys? A Comparative Analysis of the Education Modules in 30 Household Surveys from 1996–2005

Education Policy and Data Center - October 2009

This report studies 30 household surveys and analyzes how, and how well, their education information is collected. It provides a number of general recommendations useful to the individuals and agencies involved in the design of education questionnaires for household surveys and use of the resultant data.

[Download](#)



Strengthening Household Income and Expenditure Surveys as a Tool for Designing and Assessing Food Fortification Programs

John L. Fiedler - June 26, 2009

These guidelines are intended to help improve household income and expenditure surveys as a tool for designing and assessing fortification programs, and thereby aid in accelerating the development of more evidence-based fortification programs.

[Download](#)



پشتیبانی فنی برنامه تسریع داده (ADP)

شبکه بین‌المللی پیمایش‌های خانوار، پشتیبانی فنی ارائه نمی‌دهد. با این حال، تیمی که فعالیت‌های شبکه بین‌المللی پیمایش‌های خانوار را هماهنگ می‌کند، اجرای یک برنامه مرتبط، برنامه تسریع داده (ADP) را نیز هماهنگ می‌کند.

ADP برای کمک به کشورها در بهبود برنامه‌های پیمایش و افزایش استفاده و ارزش داده‌های پیمایش راه‌اندازی شد. ADP پشتیبانی فنی و مالی را برای مستندسازی و انتشار داده‌های پیمایش و بهبود روش‌های پیمایش فراهم می‌کند. خروجی‌های کلیدی شامل ایجاد بانک‌های داده پیمایش ملی و ایجاد استانداردهای ملی جمع‌آوری داده‌ها برای تقویت مقایسه‌پذیری داده‌ها در بین منابع است.

ADP عمدتاً توسط تسهیلات کمک مالی توسعه بانک جهانی، با حمایت مالی بیشتر از بانک توسعه آفریقا، دپارتمان توسعه بین‌المللی بریتانیا (DFID) و سایرین تامین می‌شود. فعالیت‌های آن شامل بسیاری از آژانس‌های ملی، منطقه‌ای و بین‌المللی است و به‌طور مشترک توسط بانک جهانی و دبیرخانه PARIS۲۱ هماهنگ می‌شود.

برای اطلاعات در مورد پشتیبانی ارائه شده توسط ADP و سایر اطلاعات در مورد برنامه، به وبسایت ADP مراجعه کنید.



مقدمه‌ای بر Stata در تجزیه و تحلیل داده‌های پیمایش

Powerpoint presentation - http://ihsn.org/sites/default/files/stata_for_survey_analysis.pptx

Example Data - http://ihsn.org/sites/default/files/dataset_data_for_survey_analysis.zip

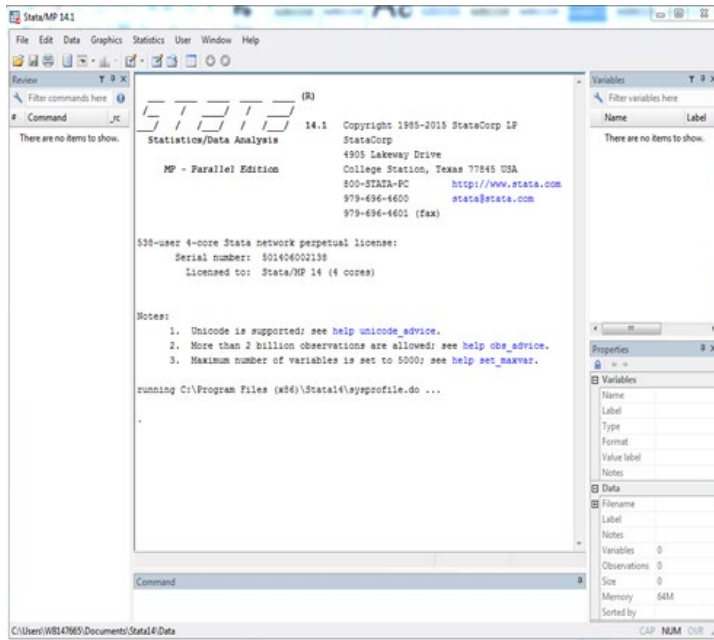
در ذیل، متن ارائه این پاورپوینت ارائه شده است:



An Introduction to Stata for Survey Data Analysis

Olivier Dupriez, World Bank
2018

۱- وقتی برنامه Stata را شروع می‌کنید (Launch)...

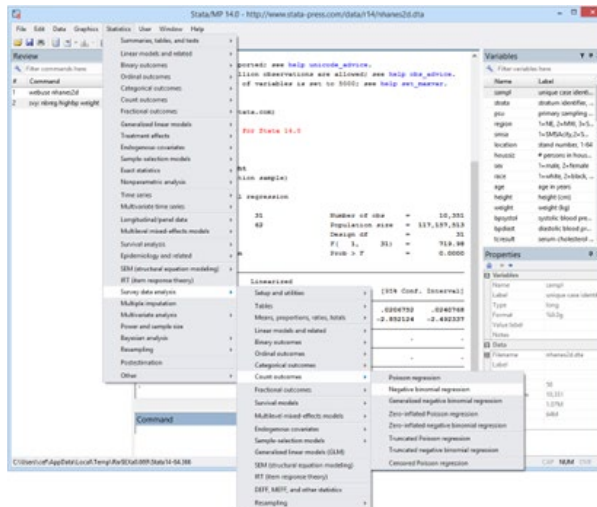


۲-۳ راه برای اجرای دستورات Stata وجود دارد:

- منوها و گنتگوها (رابط گرافیکی کاربر):
- خط فرمان (Command line):
- نوشتن برنامه‌ها (انجام فایل‌ها).

۳- گزینه ۱: رابط گرافیکی کاربری Stata

رابط کاربری اجازه می‌دهد تا بسیاری از وظایف منو محور و محاوره محور را انجام دهید، اما این روش استفاده حرفه‌ای از Stata نیست.



۴- گزینه ۲: خط فرمان (Command line)

دستورات برای اجرای فوری در پنجره "Command" تایپ می‌شوند. برای اجرای یک دستور، آن را در خط فرمان تایپ کرده و Enter را فشار دهید.



۵- گزینه ۳: نوشتن برنامه‌ها (انجام فایل‌ها)

• حرفه‌ای‌ها:

○ برنامه‌ها را می‌نویسند (فایل‌ها را انجام می‌دهند)، از گزینه‌های menu-driven یا گزینه‌های خط فرمان (command line) استفاده نمی‌کنند؛

○ در صورت لزوم، می‌نویسند یا از برنامه‌های ado (بسته‌های کمکی تخصصی) استفاده می‌کنند.

• چرا؟

○ قادر به حفظ، تکثیر، اشتراک‌گذاری، به‌روز رسانی، ساختن، استفاده مجدد و هدفمندی مجدد از



تجزیه و تحلیل خود باشند؛

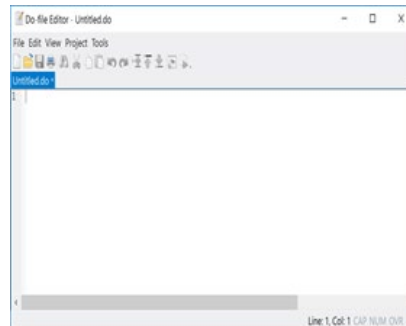
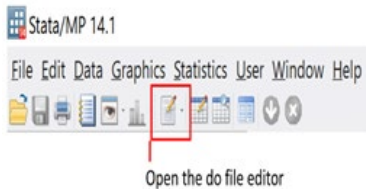
○ برای مستندسازی فرآیند تحلیلی؛

○ برای اتومات کردن برخی کارها.

توجه: گزینه menu-driven برای نوشتن برنامه‌ها مفید است، زیرا به‌طور اتومات انتخاب‌های شما را به دستوری تبدیل می‌کند که می‌توانید آن را کپی و در فایل‌های do خود جای‌گذاری کنید. برای مبتدیان Stata، این می‌تواند کمک کننده باشد.

۶- دسترسی به ویرایشگر do-file

- Do-files فایل‌های متنی (با پسوند do) هستند که می‌توانند با استفاده از هر ویرایشگر متنی تولید شوند؛
- توصیه: از ویرایشگر Stata do-file استفاده کنید.



۷- اجرای دستورات از ویرایشگر do-file

برنامه خود را در ویرایشگر do-file تایپ کنید.



```

Do-file Editor - F5M_HIES2013_prep_files.do
File Edit View Project Tools
127 gen age0_5=0
128 replace age0_5=1 if inlist(q10104,25,30,35,40,45,50,55,60)
129 gen age23_62=0
130 replace age23_62=1 if inrange(q10104,23,62)
131 collapse (sum) age0_5 age23_62, by(id01)
132 gen whipple= (age0_5 * 100) / (age23_62 / 5)
133 table i001, c(mean whipple)
134
135 * Each household should have one and only one head
136 restore, preserve
137 gen head=q10106==1
138 bysort hhid: egen abheads=total(head)
139 assert abheads==1
140
141 * If relationship is "Spouse", must be married or in union
142 restore
143 tab q10109 if q10106==2 & !inlist(q10109,2,3)

```

دستوراتی را که می‌خواهید اجرا کنید انتخاب کنید (هاپلایت کنید).

روی آیکن EXECUTE کلیک کنید.





۸- فایل‌های Ado (بسته‌های کمکی)

۹- فایل‌های Ado

- فایل‌های ADO بسته‌هایی هستند که توسط کاربر ارائه می‌شود و می‌توان آنها را در Stata نصب کرد تا قابلیت‌های تخصصی را به Stata اضافه کنند؛
- مجموعه بزرگی از بسته‌های ado به صورت آنلاین در دسترس است؛
- آنها را می‌توان با استفاده از دستور findit در Stata پیدا کرد؛

○ مثال، برای یافتن برنامه‌هایی برای تجزیه و تحلیل نابرابری: findit inequality

- آنها همچنین می‌توانند از داخل Stata با استفاده از "ssc install" نصب شوند.

○ مثال:

```
ssc install inequal7
```

```
ssc install poverty
```

۱۰- چند فایل ado مفید

- برای تولید جدول‌ها (علاوه بر دستورات جدول‌بندی Stata)
 - oTabout (beta version at <http://tabout.net.au/docs/home.php>)
- برای تهیه نقشه (به بخش نقشه‌ها در این ارائه مراجعه کنید)
 - oshprdata, spmap
- برای تحلیل فقر و نابرابری
 - glorenz ,opovdeco, poverty, ineqdeco, inequaly
- برای شما؟
 - با استفاده از Findit پیدا کنید.

۱۱- قبل از شروع...

۱۲- روش خوب برای تجزیه و تحلیل داده‌ها



برخی از قوانین مهمی که باید رعایت شوند:

- قبل از تجزیه و تحلیل داده‌های خود، آنها را درک کنید؛
- مجموعه داده خود را مستند کنید؛
- از داده‌های خود محافظت کنید - روی یک کپی کار کنید، نه روی مجموعه داده اصلی؛
- همه چیز را قابل برگشت و تکرار کنید؛
- برنامه‌های Stata خود را مستند کنید.



۱۳- برخی اطلاعات اساسی

- نام متغیرها می‌تواند حداکثر ۳۲ کاراکتر باشد؛
 - متغیرهای موجود در فایل Stata می‌توانند عددی یا الفبایی (متغیر رشته) باشند؛
 - Stata به حروف کوچک و بزرگ حساس است (برای دستورات، نام متغیرها و غیره)؛
- o دستورات باید با حروف کوچک تایپ شوند (مثال: use یک دستور معتبر است؛ اما اگر USE را تایپ کنید کار نخواهد کرد)؛
- o متغیری به نام Age با متغیری به نام age یکی نیست.

۱۴- کمک گرفتن

- Stata تعداد بسیار زیادی دستور دارد. هر دستور دارای یک syntax است و اغلب گزینه‌های متعددی را ارائه می‌دهد؛
- کاربران اغلب برای یافتن نحوه اجرای یک دستور به راهنمای آنلاین (on-line Help) تکیه می‌کنند؛
- دستور Stata برای دریافت help در مورد یک دستور، کمکی است که به دنبال آن نام دستور می‌آید، مثال: help merge؛
- درک نحوه خواندن syntax یک دستور بسیار مهم است؛
- اگر نام دستور را نمی‌دانید، از search استفاده کنید.

۱۵- دستورات Syntax

با چند استثنا، syntax اصلی زبان Stata:

[by varlist:] command [varlist=exp] [if exp] [in range] [weight] [, options]

جایی که:

- براکت‌های مربع، واجد شرایط و گزینه‌های اختیاری را از موارد ضروری متمایز می‌کند؛
- varlist نشان‌دهنده فهرستی از نام‌های متغیر، command نشان‌دهنده یک دستور Stata، exp نشان‌دهنده یک عبارت جبری، range نشان‌دهنده یک محدوده مشاهده،



weight نشان‌دهنده یک عبارت وزنی، و **options** نشان‌دهنده فهرستی از گزینه‌ها است.

۱۶- نمونه‌ای از Syntax

help summarize را در خط دستور تایپ کنید. دستور summarize انواع آمار خلاصه تک متغیره را محاسبه و نمایش می‌دهد. syntax ما این است:

```
summarize [varlist] [if] [in] [weight] [, options]
```

Options	Description
detail	display additional statistics
<u>meanonly</u>	suppress the display; calculate only the mean; programmer's option
format	use variable's display format
separator(#)	draw separator line after every # variables; default is separator(5)
<u>display_options</u>	control spacing, line width, and base and empty cells

۱۷- نام کوتاه و مختصر دستورات

- نام دستورات (و متغیرها) به‌طور کلی می‌تواند برای ذخیره تایپ به اختصار باشد؛
- به عنوان یک قاعده کلی، نام‌های option، command، و variable ممکن است به کوتاه‌ترین رشته از کاراکترهایی که به‌طور منحصر به فرد آنها را مشخص می‌کند، مخفف شوند؛
- به عنوان مثال، تایپ su (یا summ) به جای summarize کار خواهد کرد؛
- اگر command یا option کاری را انجام دهد که به راحتی قابل برگشت نباشد، این قانون نقض می‌شود. سپس دستور باید به‌طور کامل بیان شود؛
- syntax بر حداقل مجموعه کاراکترهای مورد نیاز تاکید می‌کند.

۱۸- مثال:



Syntax

`regress depvar [indepvars] [if] [in] [weight] [, options]`



Describe data in memory

`_describe [varlist] [, memory_options]`



`recode varlist (rule) [(rule) ...] [, generate(newvar)]`



۱۹- تجزیه و تحلیل داده‌های پیمایش نمونه: طراحی پیمایش، وزن نمونه و دستورات SVY

۲۰- یک یادآوری مختصر در مورد طراحی نمونه

- ما به دنبال استفاده از Stata برای تجزیه و تحلیل داده‌های پیمایش هستیم؛
- داده‌های پیمایش از نمونه‌ای از جامعه مورد نظر جمع‌آوری می‌شوند؛
- هر مشاهده در مجموعه داده نشان‌دهنده مشاهدات متعدد در کل جمعیت است؛
- نمونه را می‌توان به چند روش رسم کرد: تصادفی ساده، طبقه‌بندی شده و غیره؛
- به عنوان مثال: به‌طور تصادفی ابتدا N روستا در هر استان، و سپس ۱۵ خانوار در هر روستا را انتخاب کنید؛
- وزن نمونه، متغیرهایی هستند که نشان می‌دهند هر مشاهده چند واحد در جامعه را نشان می‌دهد.

۲۱- وزن‌های نمونه‌گیری

- وزن نمونه معمولاً معکوس احتمال مشاهده انتخاب شده است؛
- مثال: در یک انتخاب تصادفی ساده، اگر کل جمعیت ۱۰۰۰۰۰۰ خانوار داشته باشد و نمونه ۵۰۰۰ را ترسیم کنیم؛
- احتمال انتخاب شدن $0.0005 = 5000 / 1000000$ است؛
- وزن نمونه هر خانوار $200 = 1000000 / 5000$ خواهد بود؛



- در طرح‌های نمونه پیچیده‌تر، وزن نمونه برای هر منطقه یا ناحیه شمارش و غیره متفاوت خواهد بود؛
- هنگامی که ما تخمین‌هایی را ایجاد می‌کنیم (مجموع، میانگین، نسبت‌ها، و غیره) باید این وزن‌ها را اعمال کنیم تا تخمین‌هایی داشته باشیم که نشان‌دهنده جامعه باشد و نه نمونه (یعنی ما به "برآوردهای وزنی" نیاز داریم).

۲۲- کار بر روی فایل‌های داده

۲۳- ساختار فایل داده Stata

Variables

	iid	sex	age	relation	marital	rururb	wta_hh	PROV	DIST	
Observations →	1	05	2	12	3	2	1	2507.01	001	006
→	2	08	2	2	5	2	1	2507.01	001	006
→	3	02	2	42	2	1	1	2507.01	001	006
→	4	01	1	54	1	1	1	2507.01	001	006
→	5	03	2	6	3	2	1	2507.01	001	006
→	6	04	2	10	3	2	1	2507.01	001	006
→	7	06	1	32	4	1	1	2507.01	001	006
→	8	07	2	22	3	1	1	2507.01	001	006
→	9	02	2	37	2	1	1	2507.01	001	006

۲۴- باز کردن یک فایل داده

Syntax:

`use filename, clear`

If no path is specified, Stata will look in the default directory. You can find what is the default data directory by typing "cd" or "pwd" in the command line. You can change the directory by typing cd "path".

Example:

```
use "C:\Stata_Fiji\Data\household.dta", clear
or
cd "C:\Stata_Fiji\Data"
use "household.dta", clear
```



۲۵- مرتب‌سازی یک `data file - sort`

Syntax:

```
sort varlist
```

Example:

```
sort hhid totexp
```

۲۶- مرتب‌سازی یک `data file - gsort`

- دستور `sort` براساس مقدار صعودی متغیر(های) انتخاب شده مرتب می‌شود؛
 - برای `sort` به ترتیب نزولی از دستور `gsort` استفاده کنید؛
 - Syntax:
- ```
gsort [+|-] varname [[+|-] varname...] [, generate(newvar) mfirst]
```
- گزینه‌ها به شما امکان می‌دهند، در میان چیزهای دیگر، یک متغیر با تعداد متوالی رکوردهای مرتب شده تولید کنید؛
  - به عنوان مثال: برای `sort` یک فایل داده با کاهش ترتیب `variable income`:
- ```
gsort -tot_exp hhid
```

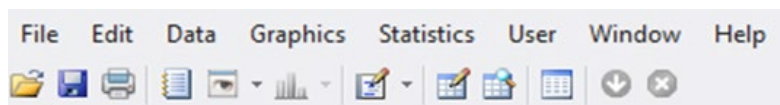
۲۷- فشرده‌سازی (Compressing) و ذخیره فایل‌های داده

- فشرده‌سازی (Compressing)
- `compress` تلاش می‌کند تا مقدار حافظه استفاده شده توسط داده‌های شما را کاهش دهد؛
- هرگز منجر به از دست دادن دقت نمی‌شود؛
- توجه: این همان فایل فشرده نیست.
- ذخیره فایل‌های داده Stata
- `save [filename] [, save_options]`
- E.g., `save "household.dta", replace`
- فایل‌های ذخیره شده در Stata ۱۴ با نسخه‌های قبلی نرم‌افزار، قابل خواندن نخواهند بود.



اگر نیاز به ذخیره داده‌ها در قالب قدیمی دارید، از گزینه saveold استفاده کنید.

۲۸- مرور (مشاهده) داده‌ها



	hhid	province	district	ea	urbrur	hhno	hheize	dwelling
1	532	West	11	104	Urban	0532	4	Detached House
2	380	East	8	74	Urban	0380	3	Detached House
3	950	South	20	188	Urban	0950	6	Detached House
4	533	West	11	104	Urban	0533	1	Detached House
5	568	West	12	112	Urban	0568	3	Detached House
6	865	South	18	171	Urban	0865	7	Detached House
7	746	South	16	148	Urban	0746	7	Detached House
8	806	South	17	160	Urban	0806	4	Detached House



۲۹- ارزیابی فایل‌های داده - توضیحات فایل

describe produces a summary of the dataset in memory

`describe [varlist] [, memory_options]`

```
Contains data from C:\Stata_Manual\Data\household.dta
obs:      1,000      Data file distributed with the Stata m
vars:      29       17 Mar 2017 11:49
size:     75,000    (_dta has notes)
```

variable name	storage type	display format	value label	variable label
hhid	int	%9.0g		Household ID
province	byte	%10.0g	province	Province
district	byte	%9.0g		District
ea	int	%9.0g		Enumeration area
urbrur	byte	%9.0g	urbrur	Rural/urban
hhno	str4	%9s		Household number
hhsz	byte	%16.0g		Household size
dwelling	byte	%23.0g	dwelling	Type of dwelling
water	byte	%26.0g	water	Main source of drinking water
toilet	byte	%16.0g	toilet	Toilet facility
wall	byte	%16.0g	wall	Main wall material

۳۰- ارزیابی فایل‌های داده - آمار خلاصه

Summarize محاسبه و نمایش انواع آمار خلاصه تک متغیره است. اگر varlist مشخص نشده باشد، آمار خلاصه برای همه متغیرهای مجموعه داده محاسبه می‌شود.

`summarize [varlist] [if] [in] [weight] [, options]`

مثال‌ها:

`summarize`

`summarize [weight=hhwgt]`

`summarize [weight=hhwgt] if province==1`

۳۱- ارزیابی فایل‌های داده - شمارش رکوردها (ثبت‌ها)

count تعداد مشاهداتی را که شرایط مشخص شده را برآورده می‌کند، می‌شمارد. اگر هیچ شرایطی مشخص نشده باشد، count تعداد مشاهدات را در داده‌ها نمایش می‌دهد.

`count [if] [in]`

مثال‌ها:



```
use "C:\Stata_Fiji\individual.dta", clear
count // Counting all observations in data file
count if sex == 1 // Counting males
count if sex == 2 & age > 12 & age < . // Counting females aged 12 +
```



۳۲- ارزیابی فایل‌های داده - فهرست مشاهدات

لیست به شما امکان می‌دهد مقادیر را در مشاهدات انتخاب شده مشاهده کنید

`list [varlist] [if] [in] [, options]`

مثال‌ها:

لیست ۵ مشاهدات برتر:

`list in 1/5`

نمایش شناسه (ID)، استان و جنسیت برای افراد ۲۵ یا ۳۰ ساله:

`list hhid province sex if age == 25 | age == 30`

۳۳- ارزیابی فایل‌های داده - دستور ارزیابی

دستور `inspect` خلاصه‌ای سریع از یک متغیر عددی متفاوت از دستور `summarize` را ارائه می‌کند.

`inspect [varlist] [if] [in]`

مثال: از دواج را بررسی کنید (وضعیت تأهل)

```
marital: Marital status
# #
# #
# #
# #
# # . .
1 4
(4 unique values)
```

	Number of Observations		
	Total	Integers	Nonintegers
Negative	-	-	-
Zero	-	-	-
Positive	3,926	3,926	-
Total	3,926	3,926	-
Missing	-	-	-
	3,926		

۳۴- ارزیابی فایل‌های داده - تولید یک کتاب کُد

کتاب کُد نام متغیرها، برچسب‌ها و داده‌ها را بررسی می‌کند تا یک کتاب کُد برای توصیف مجموعه داده تولید کند.

`codebook [varlist] [if] [in] [, options]`

مثال‌ها:



```
codebook // all variables in data file
codebook sex-literate // variables sex to literate
codebook hh* // all variables with name starting with hh
```

۳۵- افزودن (ضمیمه کردن) فایل‌های داده

append مجموعه داده‌های Stata-format ذخیره شده روی دیسک را به انتهای مجموعه داده در حافظه اضافه می‌کند.

```
append using filename [filename...] [, options]
```



۳۶- ساختار سلسله مراتبی مجموعه داده‌های پیمایش

- مجموعه داده‌های پیمایش معمولاً از چندین فایل داده مرتبط ساخته می‌شوند؛
- به عنوان مثال، در یک پیمایش خانوار، یک فایل ممکن است شامل موارد زیر باشد:
 - اطلاعات دموگرافیک (۱ مشاهده برای هر نفر)؛
 - داده‌های مربوط به تحصیلات (۱ مشاهده به ازای هر فرد ۴ سال به بالا)؛
 - داده‌های مربوط به اشتغال (۱ مشاهده به ازای هر فرد بالای ۱۵ سال)؛
 - داده‌های مربوط به تولد (۱ مشاهده به ازای هر زن ۱۲ تا ۴۹ ساله)؛
 - داده‌های مربوط به ویژگی‌های مسکن (۱ مشاهده در هر خانوار)؛
 - داده‌های مربوط به هزینه‌ها (۱ مشاهده به ازای هر محصول/خدمات به ازای هر خانوار)؛
 - غیره.
- برای ادغام این فایل‌ها به "keys" (متغیرهای مشترک) نیاز داریم.

۳۷- ساختار سلسله مراتبی و کلیدها

Individual-level data (1 or > rows per household (1 per household member))

PROVINCE	DISTRICT	VILLAGE	EA	HH_NUM	PERSON_NO	AGE	SEX	RELATION	MARITAL	EDUCATION	OCCUPATION
01	01	005	10	008	01	45	1	1	2	4	21
01	01	005	10	008	02	42	2	2	2	2	10
01	01	005	10	009	01	38	1	1	2	3	03
01	01	005	10	009	02	34	2	2	2	3	03
01	01	005	10	009	03	8	2	3	1	1	.
01	01	005	10	009	04	5	1	3	1	0	.
02	01	04	05	103	01	75	1	1	2	0	15
02	01	04	05	103	02	70	2	2	2	1	10

Household-level data (1 row per household)

PROVINCE	DISTRICT	VILLAGE	EA	HH_NUM	ROOF	WALLS	WATER
01	01	005	10	008	1	3	2
01	01	005	10	009	2	3	1
02	01	04	05	103	4	2	2



۳۸- ادغام فایل‌های داده

Individual level data (1 or > men per household [1 per household member])

HOUSEHOLD	PERSONID	SEX	RELATION	MARITAL	EDUCATION	OCCUPATION
01	01	M	01	01	01	01
01	02	M	02	02	02	02
01	03	M	03	03	03	03
01	04	M	04	04	04	04
01	05	M	05	05	05	05
01	06	M	06	06	06	06
01	07	M	07	07	07	07
01	08	M	08	08	08	08
01	09	M	09	09	09	09
01	10	M	10	10	10	10
01	11	M	11	11	11	11
01	12	M	12	12	12	12
01	13	M	13	13	13	13
01	14	M	14	14	14	14
01	15	M	15	15	15	15
01	16	M	16	16	16	16
01	17	M	17	17	17	17
01	18	M	18	18	18	18
01	19	M	19	19	19	19
01	20	M	20	20	20	20
01	21	M	21	21	21	21
01	22	M	22	22	22	22
01	23	M	23	23	23	23
01	24	M	24	24	24	24
01	25	M	25	25	25	25
01	26	M	26	26	26	26
01	27	M	27	27	27	27
01	28	M	28	28	28	28
01	29	M	29	29	29	29
01	30	M	30	30	30	30
01	31	M	31	31	31	31
01	32	M	32	32	32	32
01	33	M	33	33	33	33
01	34	M	34	34	34	34
01	35	M	35	35	35	35
01	36	M	36	36	36	36
01	37	M	37	37	37	37
01	38	M	38	38	38	38
01	39	M	39	39	39	39
01	40	M	40	40	40	40
01	41	M	41	41	41	41
01	42	M	42	42	42	42
01	43	M	43	43	43	43
01	44	M	44	44	44	44
01	45	M	45	45	45	45
01	46	M	46	46	46	46
01	47	M	47	47	47	47
01	48	M	48	48	48	48
01	49	M	49	49	49	49
01	50	M	50	50	50	50
01	51	M	51	51	51	51
01	52	M	52	52	52	52
01	53	M	53	53	53	53
01	54	M	54	54	54	54
01	55	M	55	55	55	55
01	56	M	56	56	56	56
01	57	M	57	57	57	57
01	58	M	58	58	58	58
01	59	M	59	59	59	59
01	60	M	60	60	60	60
01	61	M	61	61	61	61
01	62	M	62	62	62	62
01	63	M	63	63	63	63
01	64	M	64	64	64	64
01	65	M	65	65	65	65
01	66	M	66	66	66	66
01	67	M	67	67	67	67
01	68	M	68	68	68	68
01	69	M	69	69	69	69
01	70	M	70	70	70	70
01	71	M	71	71	71	71
01	72	M	72	72	72	72
01	73	M	73	73	73	73
01	74	M	74	74	74	74
01	75	M	75	75	75	75
01	76	M	76	76	76	76
01	77	M	77	77	77	77
01	78	M	78	78	78	78
01	79	M	79	79	79	79
01	80	M	80	80	80	80
01	81	M	81	81	81	81
01	82	M	82	82	82	82
01	83	M	83	83	83	83
01	84	M	84	84	84	84
01	85	M	85	85	85	85
01	86	M	86	86	86	86
01	87	M	87	87	87	87
01	88	M	88	88	88	88
01	89	M	89	89	89	89
01	90	M	90	90	90	90
01	91	M	91	91	91	91
01	92	M	92	92	92	92
01	93	M	93	93	93	93
01	94	M	94	94	94	94
01	95	M	95	95	95	95
01	96	M	96	96	96	96
01	97	M	97	97	97	97
01	98	M	98	98	98	98
01	99	M	99	99	99	99
01	100	M	100	100	100	100

Household level data (1 row per household)

HOUSEHOLD	PERSONID	SEX	RELATION	MARITAL	EDUCATION	OCCUPATION
01	01	M	01	01	01	01
01	02	M	02	02	02	02
01	03	M	03	03	03	03
01	04	M	04	04	04	04
01	05	M	05	05	05	05
01	06	M	06	06	06	06
01	07	M	07	07	07	07
01	08	M	08	08	08	08
01	09	M	09	09	09	09
01	10	M	10	10	10	10
01	11	M	11	11	11	11
01	12	M	12	12	12	12
01	13	M	13	13	13	13
01	14	M	14	14	14	14
01	15	M	15	15	15	15
01	16	M	16	16	16	16
01	17	M	17	17	17	17
01	18	M	18	18	18	18
01	19	M	19	19	19	19
01	20	M	20	20	20	20
01	21	M	21	21	21	21
01	22	M	22	22	22	22
01	23	M	23	23	23	23
01	24	M	24	24	24	24
01	25	M	25	25	25	25
01	26	M	26	26	26	26
01	27	M	27	27	27	27
01	28	M	28	28	28	28
01	29	M	29	29	29	29
01	30	M	30	30	30	30
01	31	M	31	31	31	31
01	32	M	32	32	32	32
01	33	M	33	33	33	33
01	34	M	34	34	34	34
01	35	M	35	35	35	35
01	36	M	36	36	36	36
01	37	M	37	37	37	37
01	38	M	38	38	38	38
01	39	M	39	39	39	39
01	40	M	40	40	40	40
01	41	M	41	41	41	41
01	42	M	42	42	42	42
01	43	M	43	43	43	43
01	44	M	44	44	44	44
01	45	M	45	45	45	45
01	46	M	46	46	46	46
01	47	M	47	47	47	47
01	48	M	48	48	48	48
01	49	M	49	49	49	49
01	50	M	50	50	50	50
01	51	M	51	51	51	51
01	52	M	52	52	52	52
01	53	M	53	53	53	53
01	54	M	54	54	54	54
01	55	M	55	55	55	55
01	56	M	56	56	56	56
01	57	M	57	57	57	57
01	58	M	58	58	58	58
01	59	M	59	59	59	59
01	60	M	60	60	60	60
01	61	M	61	61	61	61
01	62	M	62	62	62	62
01	63	M	63	63	63	63
01	64	M	64	64	64	64
01	65	M	65	65	65	65
01	66	M	66	66	66	66
01	67	M	67	67	67	67
01	68	M	68	68	68	68
01	69	M	69	69	69	69
01	70	M	70	70	70	70
01	71	M	71	71	71	71
01	72	M	72	72	72	72
01	73	M	73	73	73	73
01	74	M	74	74	74	74
01	75	M	75	75	75	75
01	76	M	76	76	76	76
01	77	M	77	77	77	77
01	78	M	78	78	78	78
01	79	M	79	79	79	79
01	80	M	80	80	80	80
01	81	M	81	81	81	81
01	82	M	82	82	82	82
01	83	M	83	83	83	83
01	84	M	84	84	84	84
01	85	M	85	85	85	85
01	86	M	86	86	86	86
01	87	M	87	87	87	87
01	88	M	88	88	88	88
01	89	M	89	89	89	89
01	90	M	90	90	90	90
01	91	M	91	91	91	91
01	92	M	92	92	92	92
01	93	M	93	93	93	93
01	94	M	94	94	94	94
01	95	M	95	95	95	95
01	96	M	96	96	96	96
01	97	M	97	97	97	97
01	98	M	98	98	98	98
01	99	M	99	99	99	99
01	100	M	100	100	100	100

Data on household consumption by product (1 row per product per household)

HOUSEHOLD	PERSONID	SEX	RELATION	MARITAL	EDUCATION	OCCUPATION	PRODUCT	PRICE	QUANTITY	UNIT
01	01	M	01	01	01	01	01	1.00	100	KG
01	02	M	02	02	02	02	02	2.00	200	KG
01	03	M	03	03	03	03	03	3.00	300	KG
01	04	M	04	04	04	04	04	4.00	400	KG
01	05	M	05	05	05	05	05	5.00	500	KG
01	06	M	06	06	06	06	06	6.00	600	KG
01	07	M	07	07	07	07	07	7.00	700	KG
01	08	M	08	08	08	08	08	8.00	800	KG
01	09	M	09	09	09	09	09	9.00	900	KG
01	10	M	10	10	10	10	10	10.00	1000	KG
01	11	M	11	11	11	11	11	11.00	1100	KG
01	12	M	12	12	12	12	12	12.00	1200	KG
01	13	M	13	13	13	13	13	13.00	1300	KG
01	14	M	14	14	14	14	14	14.00	1400	KG
01	15	M	15	15	15	15	15	15.00	1500	KG
01	16	M	16	16	16	16	16	16.00	1600	KG
01	17	M	17	17	17	17	17	17.00	1700	KG



- در فایل‌های داده‌های تمرین، هر خانوار دارای یک شناسه منحصر به فرد (متغیر hhid) است و هر یک از اعضای خانواده به‌طور منحصر به فرد با ترکیبی از دو متغیر شناسایی می‌شوند: hhid (که هویت خانوار را مشخص می‌کند) و indid که هویت فرد درون خانوار را نشان می‌دهد؛
- اصولاً، hhid برای هر خانوار در فایل household-level منحصر به فرد است و ترکیب hhid و indid برای هر فرد در فایل person-level منحصر به فرد است؛
- اگر این‌طور نباشد، ادغام موفقیت‌آمیز نخواهد بود.

۴۱- ادغام فایل‌های داده - The syntax

- ادغام یک به یک روی متغیرهای کلیدی مشخص شده
- merge 1:1 varlist using filename [, options]

- ادغام چند به یک روی متغیرهای کلیدی مشخص شده

merge m:1 varlist using filename [, options]

مهم: فایل‌های داده باید براساس متغیرهای کلیدی مرتب شوند تا ادغام کار کند. اگر داده‌ها مرتب نشده باشند، یک پیغام خطا دریافت خواهید کرد.

۴۲- ادغام فایل‌های داده - The _merge variable

دستور merge، یک متغیر جدید به نام _merge تولید می‌کند که نتیجه ادغام را گزارش می‌دهد. متغیر می‌تواند ۵ مقدار ممکن را بگیرد. مقادیر ۱ تا ۳ به ویژه مرتبط هستند:

۱ مشاهده فقط در فایل اصلی ظاهر شد

۲ مشاهده فقط در فایل "using" ظاهر شد

۳ match: مشاهده در هر دو فایل داده ظاهر شد

۴۳- بررسی یکپارچگی کلید(ها)

- با استفاده از دستور isid می‌توانیم به راحتی بررسی کنیم که متغیر(های) کلیدی یک شناسایی منحصر به فرد برای هر مشاهده ارائه می‌دهند.

isid varlist

- اگر موارد تکراری وجود داشته باشد، به این معنی است که شما متغیرهای مناسب را به عنوان



کلید شناسایی نکرده‌اید یا در فایل‌های داده مشکلاتی وجود دارد.

- با استفاده از دستور `duplicates` می‌توان موارد تکراری را شناسایی و فهرست کرد.

۴۴- برچسب‌گذاری موارد تکراری (`duplicates`) (یک مثال)

برای یافتن `duplicates` از گزینه "tag" دستور `duplicates` استفاده کنید

```
duplicates tag [varlist] [if] [in] , generate(newvar)
```

مثال: `duplicates tag hhid indid, generate(isdup) tabulate isdup`

۴۵- ادغام فایل‌های داده-مثال‌ها

- ادغام یک به یک، روی متغیرهای کلیدی مشخص شده (فایل‌های داده FSM HIES 2013)

```
use "household.dta", clear
```

```
merge 1:1 hhid using "dwelling.dta"
```

```
tab _merge
```

- ادغام چند به یک، روی متغیرهای کلیدی مشخص شده

```
use "individual.dta", clear
```

```
merge m:1 hhid using "household.dta"
```

```
tab _merge
```

۴۶- کار با متغیرها

۴۷- متغیرها - اصول اولیه

- نام متغیرها می‌تواند حداکثر ۳۲ کاراکتر باشد؛
- Stata به حروف کوچک و بزرگ حساس است؛
- متغیرهای موجود در یک فایل Stata می‌توانند عددی یا الفبایی (رشته‌ای) باشند؛
- نام متغیرها را می‌توان مخفف کرد (مانند دستورات)؛
- استفاده از * و ?؛
- لیست متغیرها: `v7 v3`.



۴۸- برچسب‌گذاری متغیرها و مقادیر

متغیرها باید مستند شوند.

- همه متغیرها باید دارای یک برچسب باشند. برچسب متغیر یک توضیح (حداکثر ۸۰ کاراکتر) از متغیر است؛
- همه متغیرهای طبقه‌بندی باید دارای برچسب‌های عددی باشند. برچسب‌های عددی^۱ توصیف‌دهایی هستند که در متغیرهای طبقه‌بندی استفاده می‌شوند (مثلاً برای جنسیت متغیر، ۱ = "مذکر" و ۲ = "مونث")؛
- برچسب‌ها به شما کمک می‌کنند متغیرها را شناسایی کنید و زمانی که جدول‌ها یا سایر خروجی‌ها تولید می‌گردند توسط Stata استفاده می‌شوند.

۴۹- برچسب‌گذاری متغیرها

برای افزودن یک برچسب به یک متغیر:

```
label variable varname ["label"]
```

برای تغییر یا اصلاح یک برچسب متغیر: همان دستور (برچسب موجود را بازنویسی (overwrite) می‌کند)

۵۰- برچسب‌گذاری عددی

افزودن برچسب‌های عددی یک فرآیند دو مرحله‌ای است: ابتدا مجموعه‌ای از برچسب‌ها را تعریف می‌کنیم (label define)، سپس آن را به یک متغیر متصل می‌کنیم (label values).

یک مجموعه مشابه را می‌توان برای چندین متغیر استفاده کرد.

برای مثال:

```
label variable sex "Sex"
```

```
label define gender 1 "Male" 2 "Female"
```

```
label value sex gender
```

۵۱- اصلاح و حذف برچسب‌های عددی

برای افزودن یا اصلاح برچسب‌های عددی:



```
label define lblname # "label" [# "label"...] [, add modify replace]
```

مثال:

```
label define sex 1 "Male" 2 "Girl"
```

```
label define sex 2 "Female", modify
```

```
label define sex 3 "Unknown", add
```

برای حذف برچسب‌های عددی:

```
label drop {lblname [lblname...] | _all}
```

مثال:

```
label drop sex
```

۵۲- جدول‌بندی عددهای برچسب زده شده به یک متغیر

توجه: در ادامه نحوه تولید جدول‌های متقاطع آمار خلاصه را خواهیم دید.

```
tabulate varname [if] [in] [weight] [, tabulate1_options]
```

یک گزینه مفید "nol" است (بدون برچسب)

مثال‌ها:

```
use "individual.dta", clear
```

```
tabulate marital
```

```
tabulate marital, nol
```

```
tabulate marital, sort
```

```
tabulate marital if sex == 1
```

۵۳- ایجاد متغیرهای عددی جدید

• در Stata می‌توانید با استفاده از دستور generate یک متغیر جدید تولید کنید. Syntax

کلی عبارت است از:

```
generate newvarname = expression
```

• اگر متغیری با همین نام از قبل وجود داشته باشد، نمی‌توانید متغیری ایجاد کنید؛



- برای تخصیص برچسب‌های عددی جدید به یک متغیر موجود از دستور replace استفاده کنید.

۵۴ اپراتورها

Relational operators	Logical operators	Mathematical operators
< (less than)	(or)	+
> (greater than)	& (and)	-
== (equal)	~ (not)	*
<= (less than or equal)		/
>= (greater than or equal)		^
!= or ~= (not equal)		

۵۵ توابع ریاضی

اگر X یک متغیر عددی باشد:

abs(x)	the absolute value of x
exp(x)	The exponential function of x
int(x)	the integer obtained by truncating x toward 0 (thus, int(5.2) = 5 and int(-5.8) = -5)
ln(x) or log(x)	the natural logarithm, ln(x)
max(x1,x2,::,xn)	the maximum value of x1; x2; ::: ; xn
min(x1,x2,::,xn)	the minimum value of x1; x2; ::: ; xn
mod(x,y)	the modulus of x with respect to y
round(x,y) or round(x)	x rounded in units of y or x rounded to the nearest integer if the argument y is omitted
sqrt(x)	the square root of x
sum(x)	the running sum of x, treating missing values as zero

۵۶ مقادیر گمشده (Missing values)

- مقادیر گمشده در Stata با یک نقطه نشان داده می‌شوند (.)؛
 - Stata امکان ایجاد انواع مختلفی از مقادیر گمشده را دارد
- o. /a /b / etc. until.z
- o By default, the simple dot is used (.)
- مهم: (.) توسط Stata به عنوان بزرگترین مقدار مثبت (بی نهایت) در نظر گرفته می‌شود.



این بدان معنی است که "ارزش" (۰) از هر عددی بزرگتر است؛

- وقتی با متغیرها کار می‌کنیم، این پیامدهای مهمی دارد:

۰ برای شمارش تعداد مشاهداتی که سن متغیر برای آنها وجود ندارد، تایپ کنید:

count if age >=.

برای ایجاد یک متغیر جدید و اختصاص مقدار ۱ اگر سن بزرگتر از ۶۵ سال است، تایپ کنید:

generate elderly = 1 if age > 65 & age <.

۵۷ ایجاد متغیرها - چند مثال

generate X = 1

generate X = age if age > 20

generate X = ln(tot_exp)

generate X =.

generate X = "Fiji" (create a string variable)

توجه: اگر یکی از اجزای عملیات گم شده باشد، نتیجه، از دست رفته است (به عنوان مثال، ۱ + =).

میانبر برای ایجاد یک متغیر ساختگی (ارزش‌های ۰ و ۱):

generate poor = pcexp > povline

→ Will have value 1 if pcexp > povline , and 0 otherwise

این کار همان را انجام می‌دهد:

generate poor = 0

replace poor = 1 if pcexp > povline



۵۸ کدگذاری مجدد متغیرها

Syntax:

`recode varlist (rule) [(rule) ...] [, generate(newvar)]`

rule	Example	Meaning
<code># = #</code>	<code>3 = 1</code>	3 recoded to 1
<code># # = #</code>	<code>2 . = 9</code>	2 and . recoded to 9
<code>#/ # = #</code>	<code>1/5 = 4</code>	1 through 5 recoded to 4
<code><u>nonmissing</u> = #</code>	<code>nonmiss = 8</code>	all other nonmissing to 8
<code><u>missing</u> = #</code>	<code>miss = 9</code>	all other missings to 9

۵۹- کدگذاری مجدد متغیرها-مثال

ایجاد گروه‌های سنی با کدگذاری مجدد سن

`recode age (0/4 = 0) (5/9 = 5) (10/14 = 14) ... (90/max=90), generate(age-group)`

۶۰- دستورات `decode` و `encode`

- از `encode` برای تبدیل رشته‌ها به متغیرهای عددی استفاده کنید. Stata با اختصاص اتومات کدهای عددی یک متغیر جدید (عددی) و برچسب‌های ارزش مربوط را ایجاد می‌کند.

مثال:

`encode prov, generate(province)`

برای انجام برعکس، از `decode` استفاده کنید. Stata یک متغیر جدید (رشته‌ای) حاوی برچسب متغیر عددی تولید می‌کند.

مثال:

`decode sex , generate(gender)`

۶۱- `inrange` و `inlist`



inlist () و inrange () توابع برنامه نویسی مفیدی هستند که با دستوراتی مرتبط هستند که اغلب استفاده می‌شوند.

مثال‌های استفاده:

```
generate region = 1 if inlist(province,3,4,7)
```

```
generate reprodw = 1 if inrange(age,12,49) & sex==2
```

۶۲- عملیات روی متغیرهای غیر عددی (رشته‌ای)

- در برخی موارد، متغیر عددی ممکن است به عنوان متغیرهای رشته‌ای وارد شده باشد (به عنوان مثال، ۱ به عنوان ارزش ۱ در نظر گرفته نمی‌شود، بلکه به عنوان یک کاراکتر الفبایی عددی در نظر گرفته می‌شود).

○ شما نمی‌توانید عملیات ریاضی را روی متغیرهای رشته‌ای انجام دهید.

○ توجه: در مرورگر Stata، متغیرهای رشته‌ای به رنگ قرمز نمایش داده می‌شوند.

- با استفاده از دستور `destring [variablename]` می‌توانید یک متغیر را از رشته‌ای به نوع عددی تبدیل کنید. این تنها در صورتی کار خواهد کرد که متغیر فقط شامل اعداد باشد نه حروف.

- Stata توابع زیادی را برای کار با متغیرهای رشته‌ای (از جمله توابع به رشته‌های زیرمجموعه، الحاق ' و غیره) فراهم می‌کند.

۶۳- عملیات روی متغیرهای رشته‌ای - برخی از توابع

- `abbrev (s, n)` → returns s (=text) abbreviated to a length of n
- `substr (s, n1, n2)` → returns the substring of s, starting at position n1, for a length of n2
- `strlower (s)` / `strupper (s)` → converts to lower (upper) case
- Functions can be combined (nested) into one command
- Strings can be combined using " + "
- Example:


```
generate staff = "Pierre"
generate staff2 = strupper(substr(staff,1,4)) + "."
// → staff2 = PIER.
```



۶۴- تغییر نام متغیرها

rename، نام متغیر موجود را تغییر می‌دهد.

مثال:

```
rename age age_years
```

Stata برخی از توابع را برای تغییر نام گروهی از متغیرها فراهم می‌کند؛ به `help rename group` مراجعه کنید.

۶۵- حذف (یا حفظ) متغیرها

- `drop` متغیرها را از فایل داده در حافظه حذف می‌کند؛
- `Keep` مانند `drop` عمل می‌کند، با این تفاوت که شما متغیرهایی را که باید نگه داشته شوند، به جای متغیرهایی که باید حذف شوند، مشخص می‌کنید؛
- هشدار: `drop` و `keep` قابل برگشت نیستند ("`undo`" وجود ندارد). هنگامی که متغیرها را حذف کردید، نمی‌توانید دوباره آنها را بخوانید. شما باید به مجموعه داده اصلی برگردید و دوباره آن را بخوانید.
- مثال‌ها:

```
o drop _merge
```

```
o keep hhid q1*
```

۶۶- حذف (یا حفظ) مشاهدات

- از همان دستورات `drop` و `keep` می‌توان برای انتخاب مشاهدات استفاده کرد؛
- `drop` مشاهده را حذف می‌کند؛ `keep` اثر مشابه `drop` دارد، با این تفاوت که مشاهداتی را که باید حفظ شوند به جای مواردی که باید حذف شوند را مشخص می‌کنید؛
- هشدار: `drop` و `keep` برگشت‌پذیر نیستند. وقتی مشاهدات را حذف کردید، نمی‌توانید دوباره آنها را بخوانید. شما باید به مجموعه داده اصلی برگردید و دوباره آن را بخوانید؛

• مثال‌ها: `drop if age == .`

`keep if age < .`

۶۷- مرتب‌سازی متغیرها



order ترتیبی را تغییر می‌دهد که در آن متغیرها در یک فایل داده فهرست می‌شوند. ارزش داده‌ها را تغییر نمی‌دهد. این معمولاً برای اطمینان از این که برخی از متغیرهای کلیدی در بالای لیست نمایش داده می‌شوند، انجام می‌شود. شما فقط باید ابتدا متغیرهایی را که می‌خواهید نمایش داده شوند، لیست کنید. مثلاً:

```
use "individual.dta", clear
describe
order hhid indid ea
describe
```

hhid	int	۹۹.0g		Household ID	hhid	int	۹۹.0g	Household ID
indid	byte	۹۱۶.0g		Individual ID	indid	byte	۹۱۶.0g	Individual ID
province	byte	۹۱۰.0g	province	Province	ea	int	۹۹.0g	Enumeration area
district	byte	۹۹.0g		District	province	byte	۹۱۰.0g	Province
ea	int	۹۹.0g		Enumeration area	district	byte	۹۹.0g	District
urbrur	byte	۹۹.0g	urbrur	Rural/urban	urbrur	byte	۹۹.0g	Rural/urban

۶۸- ایجاد متغیرهای جدید با egen

- egen متغیرهای جدیدی را ایجاد می‌کند که آمار خلاصه را نشان می‌دهد (محاسبه شده در ردیف یا ستون)؛
- egen از توابعی که به‌طور خاص برای آن نوشته شده است استفاده می‌کند؛
- syntax عبارت است از:

egen [type] newvar = fcn(arguments) [if] [in] [, options]

- توابع شامل count (), iqr (), min (), max (), mean (), median (), mode (), rank (), pctile (), sd (), و total () است؛
- این توابع یک by..... می‌گیرند: پیشوندی که امکان محاسبه آمار خلاصه در هر by-group را فراهم می‌کند.



۶۹- استفاده از egen - چند مثال

```
use "individual.dta", clear
```

```
* Add a variable with the age of the oldest hhld member for each hhld
egen oldest = max(age), by(hhid)
```

```
* Add the number of members declared as "spouse"
```

```
generate spouse= 1 if relat == 2
egen numsp = sum(spouse), by(hhid)
tabulate numsp
```

۷۰

`egen = rank()` creates a variable assigning the rank of a variable. For example, with a variable `tot_exp`:

- `egen rank0 = rank(tot_exp), field` → assigns rank = 1 to the highest income, etc (no correction for ties; if 2 observations have the same income, they will have the same rank)
- `egen rank1 = rank(tot_exp), track` → assigns rank = 1 to the lowest income, with no correction for ties)
- `egen rank2 = rank(tot_exp), unique` → assigns rank = 1 to the lowest income; all observations have a different rank (random allocation in case of ties)

۷۱ تولید دهک یا پنجک با استفاده از xtile

- دستور `xtile` برای مثال برای تولید پنجک یا دهک براساس مقادیر یک متغیر استفاده می‌شود (به عنوان مثال، پنجک‌های هزینه سرانه - pce)

```
xtile newvar = exp [if] [in] [weight] [, xtile_options]
```

- بسته به وزنی که در پیمایش خانوار استفاده می‌کنیم، پنجک‌های خانوار (۲۰ درصد خانوارها در هر پنجک) یا پنجک‌های جمعیت (۲۰ درصد افراد در هر پنجک) را تولید می‌کنیم.
 - از وزن نمونه خانوار برای پنجک‌های خانوار استفاده کنید؛
 - برای به‌دست آوردن پنجک جمعیت، وزن جمعیت = وزن خانوار X اندازه خانوار (ضرب) ایجاد کنید.



۷۲- محاسبه پنجگ هزینه سرانه

```
use "household.dta", clear
* To have population (not hhld) quintiles, we use population weight
generate pcexp = tot_exp / hhsize
generate popweight = hhwt * hhsize
xtile quinpop= pcexp [pweight= popweight], nq(5)
* Check
tab quinpop [aweight = popweight]
```

5 quantiles of pcexp	Freq.	Percent	Cum.
1	200.518997	20.05	20.05
2	200.135679	20.01	40.07
3	199.604612	19.96	60.03
4	201.964547	20.20	80.22
5	197.776166	19.78	100.00
Total	1,000	100.00	

73

۷۳- متغیرهای در حال جمع شدن (Collapsing)

- collapse مجموعه داده موجود در حافظه را به مجموعه داده‌ای از میانگین‌ها، مجموع، میانه‌ها و غیره تبدیل می‌کند.

collapse clist [if] [in] [weight] [, options]

- Collapsing فایل‌های داده ابزار بسیار مفیدی است که باید به خوبی درک شود.
- برای مثال برای تولید فایل‌های داده در سطح خانوار از فایل‌های داده در سطح فردی استفاده می‌شود.

۷۴- استفاده از دستور collapse: مثال‌ها

```
* Calculating household size and max/mean age from demographic data
use "individual.dta", clear // Data with demographic information
collapse (count) hh_size = sex /// Or any variable with no missing
          (mean) mean_age = age /// q10104 = age in years
          (max) max_age = age, by(hhid)

* Producing a file with number of hhlds and population by province
* Sum hhld and population sampling weights, by province
use "household.dta", clear // A file with observations at hhld level
generate popweight = hhwt * hhsize
collapse (sum) hhwght popweight, by(province)
```

توجه: از egen می‌توان برای تولید همان متغیرها بدون تولید فایل‌های جدید استفاده کرد.



۷۵- drop duplicates استفاده از

یکی از راه‌های حفظ تنها یک مشاهده در هر گروه (به عنوان مثال، در هر خانوار) استفاده از collapse است. راه دیگر این است که همه موارد duplicates متغیرهای کلیدی را با استفاده از دستور drop duplicates حذف کنید.

drop duplicates varlist [if] [in], force

۷۶- dummy تولید متغیرهای

- متغیرهای Dummy متغیرهایی با مقادیر ۰ (نادرست) و ۱ (درست) هستند. قبلاً دیدیم که چگونه یک متغیر Dummy با استفاده از دستور generate تولید کنیم، به عنوان مثال.
O راه طولانی:

generate male = 0

replace male = 1 if sex == 1

O راه کوتاه:

generate male = sex==1

- وقتی چندین دسته دارید، این روش خسته کننده است. به جای آن می‌توانید از دستور tab-ulate استفاده کنید. مثلاً:

tabulate province, gen(prov)

این متغیرهای ساختگی، prov1، prov2، prov3،، provN را ایجاد می‌کند (یک Dummy برای هر استان)

- یک گزینه بیشتر، استفاده از دستور xi است (به اسلایدهای مربوط به رگرسیون مراجعه کنید).

۷۷- تولید جداول

۷۸- جدول بندی

- ما پیش‌تر دیدیم که جداول فراوانی را می‌توان به راحتی با استفاده از دستور tabulate تولید کرد (همچنین به tab1 و tab2 مراجعه کنید).
- برای تولید جداول چند بُعدی با آمار خلاصه از دستورات table استفاده می‌کنیم.



- Stata همچنین دستور `tabstat` را برای تولید جداول با آمار خلاصه برای یک سری از متغیرهای عددی ارائه می‌دهد.
- یک بسته ارائه شده توسط کاربر^۱ (فایل `ado`) به نام `tabout` مکمل دستورات جدول‌بندی Stata است.

۷۹- یادداشتی در مورد `copy/pasting` جداول

- برای `copy` و `paste` جداول از پنجره نتایج Stata، از گزینه `copy table` استفاده کنید نه `copy`. سپس قالب‌بندی جدول حفظ می‌شود و `cells` هنگام چسباندن به اکسل به درستی متمایز می‌شوند.

۸۰- تولید جداول با استفاده از دستور `tabulate`

`tabulate` جداول یک طرفه یا دو طرفه تولید می‌کند. می‌توان از آن برای تولید جداولی که فراوانی‌ها را به درصد نشان می‌دهد استفاده کرد. `tab1` و `tab2` جداول یک طرفه و دو طرفه را برای چندین متغیر در یک دسته تولید می‌کنند (`tab2` جداول را برای همه ترکیب‌های متغیرهای مشخص شده تولید می‌کند).

`tabulate varname1 varname2 if in weight , options`

Example:

```
use "individual.dta", clear
tabulate province marital [aweight=hhwt], row nofreq
tabulate province marital [aweight=hhwt], column nofreq
tabulate province marital [aweight=hhwt], cell nofreq
tab1 province sex relat marital
tab2 sex relat marital // Produces 3 tables: sex by relat, sex by marital, relat by marital
```



۸۱- تولید جداول با استفاده از دستور table

table, جداول آمار خلاصه را محاسبه و نمایش می‌دهد.

table rowvar [colvar [supercolvar]] [if] [in] [weight] [, options]

مثال:

```
use "individual.dta", clear
table province marital [pweight=hhwt], row col format(%9.0f)
table province marital [pweight=hhwt], row col format(%9.0f)
table province marital [pweight=hhwt], c(mean age) row col format(%9.2f)
```

۸۲- تولید جداول با استفاده از دستور tabstat

مثال: جداول آمار خلاصه برای دو متغیر

```
use "household.dta", clear
tabstat tot_food tot_exp, by(province) stat(mean sd min max) nototal long

* Put the variables in row and the statistics in column
tabstat tot_food tot_exp, by(province) stat(mean sd min max) nototal col(stat)
```

province	stats	tot_food	tot_exp	province	variable	mean	sd	min	max
North	mean	71700.21	115795.2	North	tot_food	71700.21	40271.75	9592.185	232972.3
	sd	40271.75	83606		tot_exp	115795.2	83606	18749.05	554781.9
	max	232972.3	554781.9						
East	mean	56311.03	101179.2	East	tot_food	56311.03	35104.43	5872.188	221348.2
	sd	35104.43	87828.74		tot_exp	101179.2	87828.74	10445.87	904444.3

۸۳- تولید جداول با استفاده از بسته "tabout"

```
use "C:\Stata_Manual\Data\individual.dta", clear

* Recode age into age group
recode age (0/9=1 "0 - 9 years") (10/19=2 "10 - 19 years") ///
(20/29=3 "20 - 29 years") (30/39=4 "30 - 39 years") ///
(40/49=5 "40 - 49 years") (50/59=6 "50 - 59 years") ///
(60/69=7 "60 - 69 years") (70/79=8 "70 - 79 years")
(80/max=9 "80 and above"), generate(agegroup)

label variable agegroup "Age group"

tabout agegroup urbrur using "C:\Stata_Manual\Data\tabl1.xls", ///
c(col) f(1) clab(Col_%) npos(col) style(xls) replace
```

	A	B	C	D	E
1		Rural	Urban	Total	N
2	Age group	Urban	Rural	Total	
3		Col %	Col %	Col %	
4	0 - 9 years	19.7	21.2	20.5	805
5	10 - 19 yea	17.9	20.1	19.1	751
6	20 - 29 yea	15	13.2	14	550
7	30 - 39 yea	16.9	16.3	16.6	651
8	40 - 49 yea	14.4	13.1	12.6	494
9	50 - 59 yea	8.1	9.3	8.8	345
10	60 - 69 yea	5	5	5	197
11	70 - 79 yea	2.2	2.2	2.5	98
12	80 and abo	0.8	1	0.9	35
13	Total	100	100	100	3,926



۸۴- تولید نمودار

۸۵- نمودارها

Stata دارای قابلیت گراف قدرتمندی است.

تولید نمودارهای ساده بسیار آسان است. اما Stata گزینه‌های زیادی را ارائه می‌دهد که به شما امکان می‌دهد موارد پیچیده ایجاد کنید و هر جنبه‌ای از نمودارهای خود را سفارشی کنید. یک کتابچه راهنمای کامل به آن اختصاص داده شده است.

نکته: از ابزارهای منو محور استفاده کنید که کُد را برای شما تولید می‌کند.

ما در اینجا فقط برخی از دستورات اساسی و رایج را نشان می‌دهیم. پس از تسلط بر این دستورات، راهنمای Stata را برای اطلاعات بیشتر بخوانید. یا به Stata's در خط "Visual overview for creating graphs" مراجعه کنید در:

<http://www.stata.com/support/faqs/graphics/gph/stata-graphs/>

۸۶- نمودارهای میله‌ای

نمودارهای میله‌ای مقادیر را در دسته‌های مختلف یک متغیر مقایسه می‌کنند.

`graph bar yvars [if] [in] [weight] [, options]`

که در آن `yvars` لیستی از متغیرها است.

این دستور دارای گزینه‌های زیادی است و همچنین امکان ترسیم آمار خلاصه متغیرها (میانگین، میانه، صدک، حداقل، حداکثر و غیره) را فراهم می‌کند.

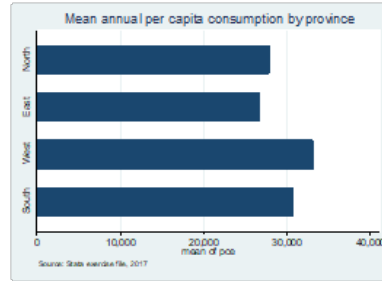
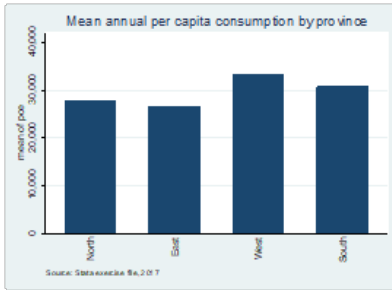
۸۷- نمودارهای میله‌ای - یک مثال

میانگین مصرف سرانه به تفکیک استان

```
use "household.dta", clear
generate pce = tot_exp / hhsize
generate popweight = hawght * hhsize
graph bar (mean) pce if pce < . [pweight = wgtpop],
over(province, label(angle(ninety))) title("Mean annual per
capita consumption by province") note("Source: Stata exercise
file, 2017")
* Use "hbar" instead of "bar" for horizontal chart
```



۸۸- نمودارهای میله‌ای - یک مثال



همان دستور، اما با `hbar` به جای `bar`

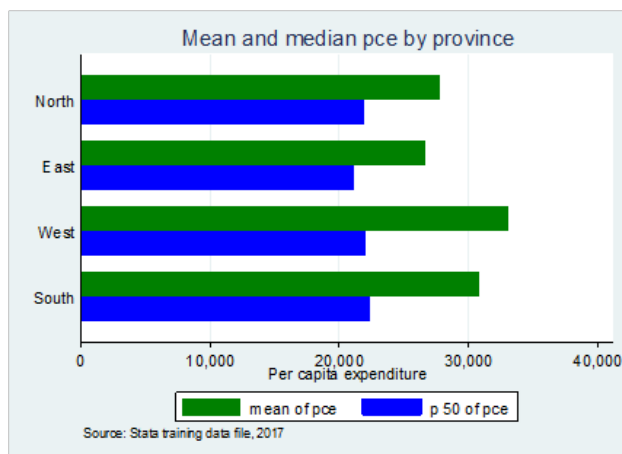
۸۹- نمودارهای میله‌ای - مثال دیگر

میانگین و میانه مصرف سرانه براساس State (۲ متغیر)

```
use "household.dta", clear
generate pce = tot_exp / hhsz
graph hbar (mean) pce (median) pce [pweight = wtppop],
over(province) title("Mean and median pce by province")
ytitle("Per capita expenditure") note("Source: Stata
training data file, 2017") bar(1, color(green)) bar(2,
color(blue))
```



۹۰ نمودارهای میله‌ای - مثال دیگر



۹۱- نمودارهای دایره‌ای

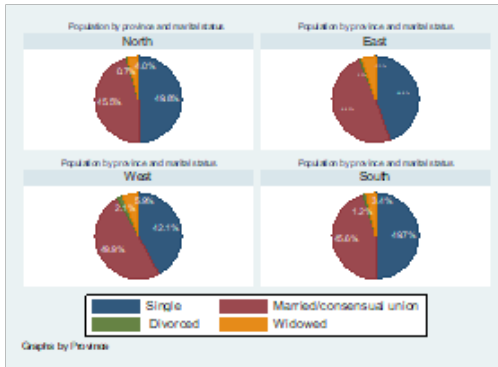
Syntax:

```
graph pie varlist [if] [in] [weight] [, options]
```

مثال:

use "individual.dta", clear

```
graph pie [pweight = hhwgt], over(marital) plabel(_all percent, color(white) format(%9.1f)) cw by(, legend(on)) by(province, title(Population by province and marital status))
```



۹۲- نمودارهای دایره‌ای

توجه داشته باشید که عنوان در بالای هر نمودار تکرار می‌شود. این عنوان بهتر است فقط یک بار در بالای همه نمودارهای دایره‌ای نمایش داده شود، زیرا برای همه صدق می‌کند.

این را می‌توان به سادگی با گنجانیدن دستورالعمل عنوان در گزینه "by" انجام داد. به مثال ارائه شده در اسلایدهای بعدی برای "نمودار نقطه‌ای" مراجعه کنید.

۹۳- نمودارهای خطی

مثال:

```
use "household.dta", clear
```

```
generate pce = tot_exp / hhsz
```

```
cumul pce, gen(cum)
```

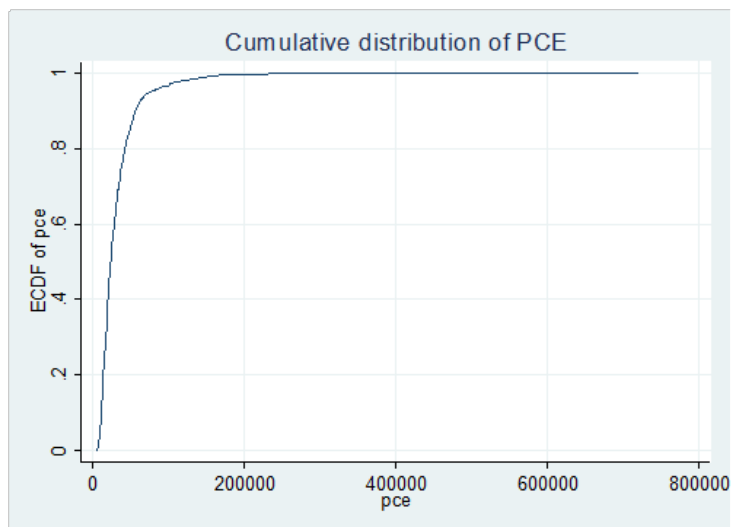
```
sort cum
```

```
line cum pce, ylab(, grid) xlab(, grid) title("Cumulative distribution of PCE")
```

توجه: cumul یک متغیر جدید ایجاد می‌کند که به عنوان تابع توزیع تجمعی تجربی یک متغیر عددی تعریف می‌شود.



۹۴- نمودارهای خطی

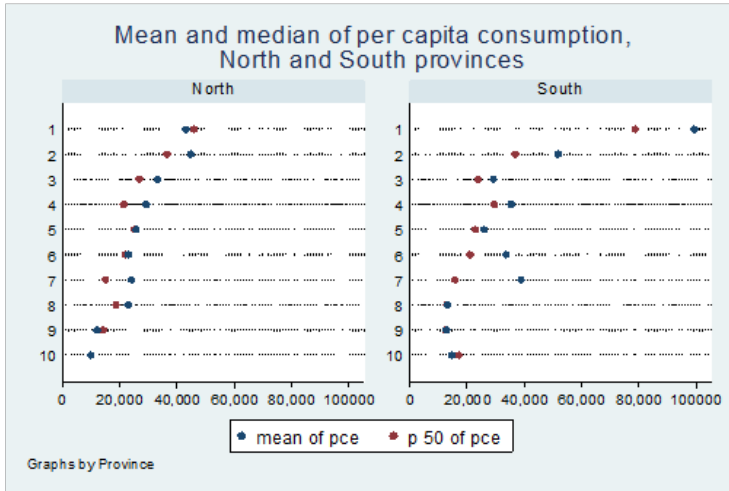


۹۵- نمودارهای نقطه‌ای

```
use "household.dta", clear
generate pce = tot_exp / hsize
recode hsize (10/max=10), generate(hsize2)
keep if province == 1 | province == 4 // We keep only two provinces
graph dot (mean) pce (p50) pce [pweight=wgtpop], over(hsize2)
by(province, title("Mean and median of per capita consumption," "North
and South provinces")) // The title is within the 'by' option
```



۹۶- نمودارهای نقطه‌ای



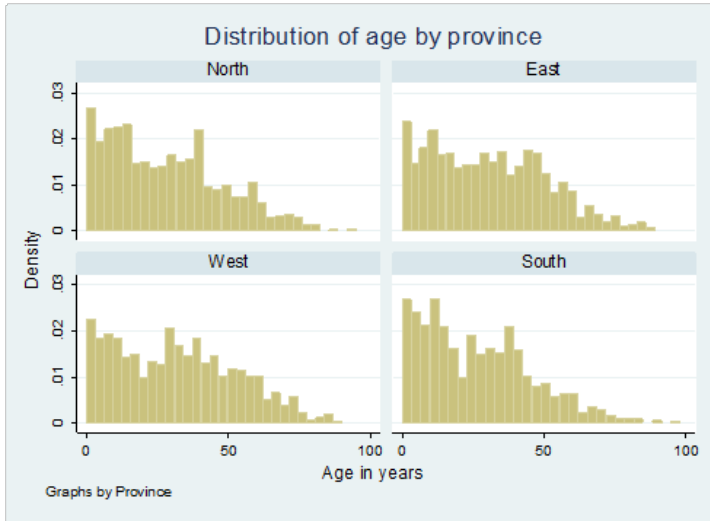
۹۷- هیستوگرام‌ها

use "individual.dta", clear

```
twoway histogram age, by(province, ///
title("Distribution of age by province"))
```



۹۸- هیستوگرام‌ها



۹۹- نمودارهای جعبه‌ای (Box)

نمودار جعبه‌ای (معروف به نمودار box and whisker) یک روش استاندارد برای نمایش توزیع داده‌ها براساس خلاصه پنج عددی است: حداقل، چارک اول، میانه، چارک سوم و حداکثر. در ساده‌ترین نمودار باکس، مستطیل مرکزی از ربع اول تا ربع سوم (محدوده بین چارکی یا IQR) را در بر می‌گیرد. قطعه‌ای در داخل مستطیل، میانه را نشان می‌دهد و "whiskers" در بالا و پایین کادر، مکان‌های حداقل و حداکثر را نشان می‌دهد.

۱۰۰- نمودارهای جعبه‌ای (Box) - مثال از FSM HIES 2013-14

```
use "household.dta", clear
```

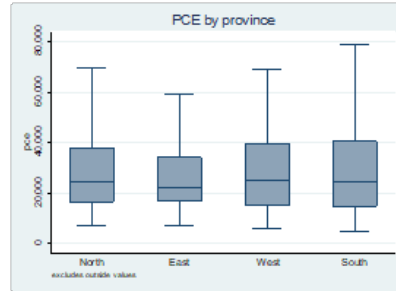
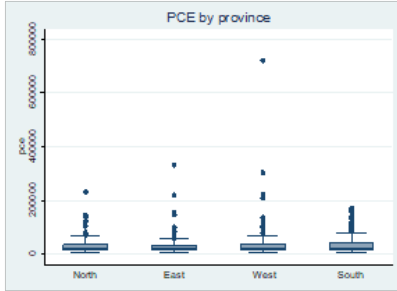
```
generate pce = tot_exp / hsize
```

```
graph box pce, over(province) title("PCE by province")
```

```
graph box pce, over(province) title("PCE by province") nooutsides
```



۱۰۱- نمودارهای جعبه‌ای (Box) - PCE in FSM, HIES 2013-14



۱۰۲- تجزیه و تحلیل آماری - رگرسیون‌ها

۱۰۳- رگرسیون‌ها در Stata

- Stata دستوراتی را برای اجرای بسیاری از انواع رگرسیون‌ها (خطی، غیرخطی، لجستیک، پروبیت، کمیت و غیره) ارائه می‌دهد.
 - متداول‌ترین انواع، مدل‌های خطی و لجستیکی هستند.
- مدل خطی برای پیش‌بینی مقدار یک متغیر پیوسته براساس مقدار یک یا چند متغیر مستقل استفاده می‌شود.
- مدل لجستیکی که برای پیش‌بینی مقدار یک متغیر باینری (مثلاً ضعیف / غیرضعیف) یا یک متغیر طبقه‌ای با بیش از ۲ دسته (رگرسیون چندجمله‌ای) استفاده می‌شود.
- برخی از دستورات خاص اجازه می‌دهد تا طرح‌های پیمایش پیچیده را در نظر بگیریم (دستور `(svyreg)`).



۱۰۴- نگاهی اجمالی به داده‌ها قبل از regressing: مقادیر پرت

قبل از اجرای یک رگرسیون، مطمئن شوید که داده‌های شما دارای مقادیر پرت، مقادیر نامعتبر یا تعداد زیادی موارد گمشده نیستند. شما می‌توانید این کار را با تولید انواع جداول و نمودار انجام دهید. به عنوان مثال، قبل از رگرسیون اجاره بهای خصوصیات مسکن، می‌توانید نمودارهای جعبه‌ای از برخی متغیرها را تولید کنید.

plots of some variables.

```
use "expenditure.dta", clear
keep if itemcode==44
graph box cons_purch
```



۱۰۵- نگاهی اجمالی به داده‌ها قبل از regressing: همبستگی بین متغیرها

همچنین می‌توانید با استفاده از دستور correlate به همبستگی متغیرهایی که قصد دارید در مدل رگرسیون استفاده کنید، نگاه کنید.

Syntax:

```
correlate [varlist] [if] [in] [weight] [, correlate_options]
```

مثال: همبستگی بین هزینه‌های سرانه، تعداد اتاق‌های مسکن و اندازه خانوار

```
. correlate pce rooms hhsz
(obs=1,000)
```

	pce	rooms	hhsz
pce	1.0000		
rooms	0.4460	1.0000	
hhsz	-0.2296	-0.2528	1.0000



۱۰۶- مدل رگرسیون خطی

- همه متغیرهای استفاده شده در مدل باید عددی باشند (بدون متغیر رشته‌ای).
- متغیر وابسته باید یک متغیر عدد واقعی باشد (یک متغیر پیوسته، برای مثال «درآمد خانوار» یا «مقدار اجاره»).
- متغیرهای مستقل می‌توانند متغیرهای پیوسته یا طبقه‌ای باشند. قبل از استفاده در مدل رگرسیون خطی، متغیرها می‌توانند و در برخی موارد باید تبدیل شوند، به عنوان مثال:
 - می‌توان از مقدار \log متغیرهای پیوسته به جای مقدار اصلی استفاده کرد (برای متغیرهای وابسته و پیش‌بینی کننده‌ها)
 - متغیرهای طبقه‌ای که به عنوان پیش‌بینی کننده استفاده می‌شوند باید به متغیرهای dum- my تبدیل شوند.

۱۰۷- رگرسیون خطی: regress, predict

- regress رگرسیون خطی حداقل مربعات معمولی را انجام می‌دهد.
 - Syntax:
- `regress depvar [indepvars] [if] [in] [weight] [, options]`
- هنگامی که یک مدل با استفاده از دستور رگرسیون مناسب شد، می‌توان آن را برای پیش‌بینی مقادیر متغیر وابسته با استفاده از دستور پیش‌بینی بر روی داده‌ها اعمال کرد. این دستور با استفاده از آخرین مدل رگرسیون اجرا شده توسط Stata، پیش‌بینی را انجام می‌دهد.

برای یک مدل معادله واحد، Syntax عبارت است از:

`predict [type] newvar [if] [in] [, options]`



۱۰۸- ایجاد dummies برای متغیرهای طبقه‌بندی شده

- بهترین گزینه برای تبدیل مقادیر طبقه‌بندی شده به dummies استفاده از دستور xi است. دستور فقط به انتخاب یک پیشوند برای نشان دادن نسخه dummy متغیرهایی که باید تبدیل شوند، نیاز دارد. به عنوان مثال، برای تبدیل متغیرهای استان و جنسیت به dum-mies، با پیشوند "i"، به سادگی باید تایپ کرد:

xi: i.province i.sex

- دستور xi و دستور رگرسیون را می‌توان به راحتی در یک دستور ترکیب کرد، به سادگی با پیش‌فرض کردن دستور regress با xi همان‌طور که در مثال کد زیر نشان داده شده است.

۱۰۹- مدل رگرسیون خطی - یک مثال

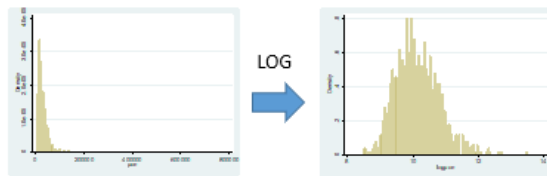
- در این مثال ما (log) هزینه سرانه (pce) را براساس متغیرهای متعدد پیش‌بینی می‌کنیم:

o Categorical: province, dwelling, water, toilet, wall, roof, floor, electricity, car

o hsize, rooms

- توزیع pce دارای چوله^۱ است. بنابراین، مدلی را برای پیش‌بینی مقدار log آن برآزش می‌دهیم که توزیعی شبه نرمال^۲ دارد. پس از پیش‌بینی $\log(pce)$ ، با استفاده از exp به مقادیر pce برمی‌گردیم.

```
histogram pce, bin(100)
gen logpce = log(pce)
histogram logpce, bin(100)
```



1. Skewed
2. Quasi-normal



۱۱۰- مدل رگرسیون خطی - یک مثال

```
use "C:\Stata_Fiji\Data\household.dta", clear
```

```
generate pce = tot_exp/hhsize
```

```
generate logpce = log(pce)
```

```
xi: regress logpce hhsize rooms    ///
```

```
    i.province i.dwelling i.water i.toilet  ///
```

```
    i.wall i.roof i.floor i.electricity i.car  ///
```

```
    [weight=hhwgt]
```

```
predict pred_logpce
```

```
generate pred_pce = exp(pred_logpce)
```

```
summarize pce pred_pce
```

۱۱۱- نتایج رگرسیون (۳/۱)

```
. xi: regress logpce hhsize rooms    ///
```

```
>    i.province i.dwelling i.water i.toilet  ///
```

```
>    i.wall i.roof i.floor i.electricity i.car  ///
```

```
>    [weight=hhwgt]
```

```
i.province    _Iprovince_1-4    (naturally coded; _Iprovince_1 omitted)
```

```
i.dwelling    _Idwelling_1-5    (naturally coded; _Idwelling_1 omitted)
```

```
i.water       _Iwater_1-5       (naturally coded; _Iwater_1 omitted)
```

```
i.toilet      _Itoilet_1-4      (naturally coded; _Itoilet_1 omitted)
```

```
i.wall        _Iwall_1-4        (naturally coded; _Iwall_1 omitted)
```

```
i.roof        _Iroof_1-7        (naturally coded; _Iroof_1 omitted)
```

```
i.floor       _Ifloor_1-6       (naturally coded; _Ifloor_1 omitted)
```

```
i.electricity _Ielectrici_0-1    (naturally coded; _Ielectrici_0 omitted)
```

```
i.car         _Icar_1-2         (naturally coded; _Icar_1 omitted)
```

```
(analytic weights assumed)
```

```
(sum of wgt is 6.1871e+07)
```

Source	SS	df	MS	Number of obs	=	1,000
Model	350.330071	32	10.9478147	F(32, 967)	=	111.71
Residual	94.7644786	967	.097998427	Prob > F	=	0.0000
				R-squared	=	0.7871
				Adj R-squared	=	0.7800
Total	445.09455	999	.44554009	Root MSE	=	.31305



۱۱۲ نتایج رگرسیون (۳/۲)

logpce	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
hhsz	-.0574795	.0067061	-8.57	0.000	-.0706396	-.0443193
rooms	.0114693	.0122005	0.94	0.347	-.0124732	.0354118
_Iprovince_2	-.049774	.0336565	-1.48	0.139	-.1158221	.0162742
_Iprovince_3	.0151265	.0341893	0.44	0.658	-.0519673	.0822203
_Iprovince_4	.0163434	.0374566	0.44	0.663	-.0571623	.089849
_Idwelling_2	-.5513828	.0369208	-14.93	0.000	-.6238369	-.4789287
_Idwelling_3	-.906243	.0574212	-15.78	0.000	-1.018927	-.7935586
_Idwelling_4	-1.238741	.0839349	-14.76	0.000	-1.403457	-1.074026
_Idwelling_5	-1.608476	.116596	-13.80	0.000	-1.837286	-1.379665
_Iwater_2	-.0953891	.0358235	-2.66	0.008	-.1656899	-.0250883
_Iwater_3	-.1609466	.0286203	-5.62	0.000	-.2171117	-.1047815
_Iwater_4	-.1803564	.0357671	-5.04	0.000	-.2505465	-.1101663
_Iwater_5	-.0579707	.0690435	-0.84	0.401	-.1934631	.0775217
_Itoilet_2	-.0768373	.0331723	-2.32	0.021	-.1419352	-.0117393
_Itoilet_3	-.0755131	.0606584	-1.24	0.213	-.1945503	.0435241
_Itoilet_4	-.0602838	.0284143	-2.12	0.034	-.1160447	-.0045229
_Iwall_2	-.0910194	.0301748	-3.02	0.003	-.150235	-.0318037
_Iwall_3	-.0649203	.0537421	-1.21	0.227	-.1703849	.0405443
_Iwall_4	-.1461342	.1098735	-1.33	0.184	-.3617521	.0694837
_Iroof_2	-.1539876	.0632853	-2.43	0.015	-.27818	-.0297953
_Iroof_3	.0410972	.1108051	0.37	0.711	-.1763489	.2585433
_Iroof_4	-.0430029	.06669	-0.64	0.519	-.1738768	.087871
_Iroof_5	.011435	.0711075	0.16	0.872	-.1281078	.1509777
_Iroof_6	-.033504	.1165107	-0.29	0.774	-.262147	.195139
_Iroof_7	.0396472	.1047088	0.38	0.705	-.1658354	.2451298
_Ifloor_2	-.0153205	.0419005	-0.37	0.715	-.0975468	.0669058
_Ifloor_3	-.0575531	.0283218	-2.03	0.042	-.1131324	-.0019738
_Ifloor_4	-.0473296	.0394306	-1.20	0.230	-.124709	.0300499
_Ifloor_5	-.0103842	.052756	-0.20	0.844	-.1139136	.0931452
_Ifloor_6	-.0647265	.0966009	-0.67	0.503	-.2542981	.124845
_Ielectrici_1	.0256608	.0379572	0.68	0.499	-.0488272	.1001488
_Icar_2	-.5110828	.0445085	-11.48	0.000	-.5984271	-.4237384
_cons	11.49543	.123861	92.81	0.000	11.25236	11.73849

۱۱۳ - نتایج رگرسیون (۳/۳)

summarize pce pred_pce

Variable	Obs	Mean	Std. Dev.	Min	Max
pce	1,000	32678.03	36303.18	4906.748	719784.7
pred_pce	1,000	29989.36	16858.07	5151.454	92220.89



۱۱۴- رگرسیون لجستیک، با نام مستعار مدل logit

- Logistic regression متغیرهای دوگانه را پیش‌بینی می‌کند، یعنی متغیر وابسته باینری است (درست/نادرست، بله/خیر، ضعیف/غیر ضعیف و غیره)
- جایگزین: رگرسیون probit
- دو دستور در Stata: logit و logistic (یکسان، با این تفاوت که logistic تخمین‌ها را به عنوان نسبت شانس^۱ نمایش می‌دهد)
- Syntax (برای جزئیات بیشتر در مورد گزینه‌ها به راهنمای Stata مراجعه کنید):

logit depvar [indepvars] [if] [in] [weight] [, options]

یا

logistic depvar indepvars [if] [in] [weight] [, options]

۱۱۵- مدل رگرسیون خطی - یک مثال

```
* We predict the poverty status of the households based on a few variables
use "C:\Stata_Fiji\Data\household.dta", clear
generate pce = tot_exp / hhszsize

* We create a variable poor (1) - non poor (0) using a poverty line = 18000
generate poor = pce < 18000
xi: logit poor hhszsize rooms i.province i.water i.toilet ///
    i.wall i.electricity i.car [pweight=hhwt]

predict poor_pred // We apply the logistic regression model

gen poor2 = poor_pred > 0.5 // If probability > 0.5 → poor, otherwise not
table poor poor2 // Show the confusion matrix
```



۱۱۶- مدل رگرسیون خطی - نتایج

```

Logistic regression              Number of obs   =    944
                                Wald chi2(15)    =   46.71
                                Prob > chi2         =  0.0000
Log pseudolikelihood = -7547461.9  Pseudo R2       =  0.7783

```

poor	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]
_b[base]	-.1797348	.1202888	1.49	0.135	-.056027 - .4154946
_b[rooms]	-2.979132	.4928465	-6.04	0.000	-3.945093 -2.01317
_b[province_2]	-.2224959	.4377789	-0.53	0.594	-1.041327 .5963357
_b[province_3]	-.6037842	.4694361	1.29	0.198	-.1362936 1.523862
_b[province_4]	-.3290973	.4433666	0.72	0.470	-.5492513 1.191066
_b[water_2]	-.3796244	.4481915	0.58	0.562	-.0996073 1.447256
_b[water_3]	-.9454424	.4927785	1.91	0.056	-.1257575 1.954683
_b[water_4]	-.8481463	.4299504	1.40	0.162	-.1488342 2.085187
_b[water_5]	2.444599	1.008224	2.42	0.015	.4480563 4.440741
_b[toilet_2]	-.454586	.473035	-0.96	0.337	-1.381228 .4725355
_b[toilet_3]	1.354784	.7592227	1.79	0.074	-.1322629 2.844835
_b[toilet_4]	-.0703007	.4377907	0.14	0.891	-.7481024 .6078955
_b[wall_2]	-.0071232	.3788574	0.02	0.985	-.7354257 .7496681
_b[wall_3]	0	[omitted]			
_b[wall_4]	0	[omitted]			
_b[electricity_1]	-.3253504	.4432731	-0.73	0.463	-1.19415 - .5434488
_b[car_2]	2.791787	.7166782	3.90	0.000	1.388304 4.195271
_b[_cons]	1.792926	2.039737	3.42	0.000	3.795106 -1.79973

Confusion matrix:

		poor2	
		0	1
poor	0	656	31
	1	26	287

نکته: روش مناسب بررسی یک مدل رگرسیون و جلوگیری از برازش بیش از حد، اندازه‌گیری عملکرد «خارج از نمونه» آن با ایجاد یک مجموعه آموزشی و یک مجموعه آزمایشی است.

۱۱۷- برنامه‌نویسی در Stata

۱۱۸- برنامه‌نویسی

- گنجانیدن نظرات (comments) در برنامه‌های شما بسیار مهم است؛
- از دستورات می‌توان برای توصیف برنامه، توضیح هدف برخی از مؤلفه‌ها و غیره استفاده کرد؛
- ۴ راه برای درج نظرات در فایل do وجود دارد.

○ خط را با '*' شروع کنید. Stata چنین خطوطی را نادیده می‌گیرد؛

○ نظر را در /* ... */ جداکننده‌ها قرار دهید؛

○ نظر را بعد از دو اسلش رو به جلو، یعنی // قرار دهید. هر چیزی که بعد از // تا انتهای خط جاری باشد یک نظر در نظر گرفته می‌شود؛

○ نظر را بعد از سه اسلش رو به جلو، یعنی /// قرار دهید. هر چیزی که بعد از /// تا انتهای خط جاری باشد یک نظر در نظر گرفته می‌شود.



۱۱۹- سربرگ (Header)

- اکیداً توصیه می‌شود که یک سربرگ (به عنوان "comment") در همه برنامه‌های خود قرار دهید که نویسنده، هدف، تاریخ، ورودی‌های ضروری و خروجی‌های برنامه را توضیح دهد.
- مثال:

```
*****
* Stata program for poverty analysis using test dataset
* Author: Olivier Dupriez, World Bank
* Date: ...
* Input files : ...
*****
```

۱۲۰- نسخه (version)، و تنظیم بیشتر (set more off)

- اولین دستوراتی که در برنامه‌های خود قرار می‌دهید اغلب نسخه و تنظیم فراتر هستند؛
- نسخه نشان می‌دهد که برنامه را برای کدام نسخه از Stata می‌نویسید (Stata تکامل می‌یابد و برخی از دستورات می‌توانند تغییر کنند)؛
- set more off پارامتری است که نمایش نتایج را کنترل می‌کند؛
- مثال:

version 14

set more off

۱۲۱- ثبت خروجی (Logging the output)

- در برخی موارد، ممکن است بخواهید log از نتایج تهیه کنید؛
- log را می‌توان به صورت یک فایل متنی یا به صورت فایل Stata فرمت شده تولید کرد؛
- باید در برنامه خود نام فایل و محل ذخیره log را وارد کنید؛
- در ابتدای برنامه، فایل log را 'open' خواهید کرد. در پایان آن را می‌بندید (توجه داشته باشید: اگر نمی‌خواهید همه نتایج را log کنید، می‌توانید ورود و خروج (log on and off) را در برنامه‌ها تنظیم کنید)؛
- می‌توانید هر بار فقط یک فایل log را باز کنید؛



- می‌توانید محتوای یک فایل log موجود را جایگزین کنید یا به آن اضافه کنید.

۱۲۲- ثبت خروجی - Syntax و مثال

- Syntax برای باز کردن یک log:

log using filename [, append replace [text|smcl] name(logname)]

○ مثال:

log using "C:/STATA_TRAINING/Exercise_01.txt", replace text

- Syntax برای بستن یک log:

log close

- Syntax برای تعلیق موقت logging یا از سرگیری ثبت logging:

log [off|on]

۱۲۳- دستورات طولانی - خط ادامه (continuation line)

- برخی از دستورات شما بیش از حد طولانی هستند که در یک خط قرار نمی‌گیرند؛
- به‌طور پیش‌فرض، Stata در نظر می‌گیرد که هر خط حاوی یک دستور است؛
- اگر دستوری در بیش از یک خط ارائه شده است، باید Stata را در مورد آن مطلع کنید. این می‌تواند توسط:

○ استفاده از یک کاراکتر ویژه برای اطلاع دادن به Stata که انتهای دستور #delimit است

(با استفاده از #delimit cr به حالت پیش‌فرض برگردید)؛

○ تایپ /// در انتهای هر خط (به جز آخرین).



۱۲۴- دستورات طولانی - مثال

```
#delimit ;
recode province (17=13) (5=14) (11=15) (16=16) (7=17)
(12=18) (3=1) (6=2) (4=3) (2=4) (14=5) (13=6) (10=7) (1=8) (8=9)
(15=11) (9=12), gen(prov) ;
#delimit cr
```

OR

```
recode province (17=13) (5=14) (11=15) (16=16) (7=17) ///
(12=18) (3=1) (6=2) (4=3) (2=4) (14=5) (13=6) (10=7) (1=8) (8=9) ///
(15=11) (9=12), gen(prov)
```

۱۲۵- Record تعداد و تعداد records

- Stata دارای دو متغیر ماکرو است که می‌توانید هر زمان که بخواهید در برنامه‌های خود از آنها استفاده کنید؛
- یکی N-نام دارد و تعداد کل مشاهدات موجود در فایل را نشان می‌دهد؛
- دیگری شماره ترتیبی هر مشاهده را در فایل داده نشان می‌دهد و n-نام دارد.

۱۲۶- (ماکروها) Macros

- در بسیاری از برنامه‌های Stata، شما از متغیرهای ماکرو استفاده خواهید کرد. اینها متغیرهایی هستند که در فایل‌های داده ذخیره نمی‌شوند، اما برای استفاده در طول اجرای برنامه‌های شما در دسترس هستند؛
- ماکروها می‌توانند محلی باشند (در این صورت فقط در یک فایل do خاص وجود دارند) یا سراسری (در این صورت می‌توانند در بین برنامه‌ها استفاده شوند)؛
- شما یک متغیر کلان را به سادگی با اعلام نوع آن و دادن یک مقدار (عددی یا رشته‌ای) به آن ایجاد می‌کنید، به عنوان مثال،

o local i = 1

o global myfolder = "C:\Stata_Fiji"



۱۲۷- (ماکروها) Macros

- هنگامی که یک ماکرو ایجاد شد و حاوی مقداری یا متنی بود، می‌توانید از آن در برنامه‌های خود استفاده کنید؛
- برای ارجاع به متغیرهای محلی در برنامه، نام ماکرو را بین نقل قول‌ها به صورت «mac-roname» قرار دهید. برای ماکروهای کلی، کاراکتر \$ را قبل از نام قرار دهید (به عنوان مثال، \$macroname)؛
- مثال:

```
local i = 10
```

```
display "The value of my local macro is "`i`"
```

```
global myfolder = "C:\Stata_Fiji"
```

```
display "The content of my global macro is " $myfolder
```

۱۲۸- فایل‌های موقت

- در برخی از برنامه‌ها، ممکن است بخواهید فایل‌های داده‌ای تولید کنید که فقط برای اجرای آن برنامه مورد نیاز هستند. با استفاده از دستور tempfile می‌توانید چنین فایل‌های موقتی ایجاد کنید. فایل‌های موقت به‌طور اتومات در پایان اجرای برنامه پاک می‌شوند؛
- می‌توانید چندین فایل موقت در یک برنامه ایجاد کنید؛
- قبل از قرار دادن محتوا در آنها، آنها را با نام‌گذاری‌شان ایجاد می‌کنید؛

مثال: برای ایجاد ۲ فایل موقت به نام‌های t0 و t1، تایپ کنید: tempfile t0 t1

- دستور tempfile را می‌توان در هر جایی از برنامه قرار داد؛
- برای ارجاع به یک tempfile، نام آن را در یک نقل قول (مانند ماکروهای محلی) قرار دهید.

مثال: save `t0', replace



۱۲۹- متغیرهای موقت

- همچنین می‌توانید متغیرهای موقت (همان روشی که می‌توانید فایل‌های داده موقت ایجاد کنید) در برنامه‌های Stata خود ایجاد کنید. این متغیرها ذخیره نمی‌شوند. آنها به‌طور اتومات در پایان اجرای برنامه حذف می‌شوند؛
- شما متغیرهای موقت را با استفاده از دستور `tempvar` راه‌اندازی می‌کنید. به عنوان مثال:
`tempvar tv1 tv2 tv3`
- در برنامه خود، شما به این متغیرها با محصور کردن آنها در نقل قول مانند کاری که با یک ماکرو محلی انجام می‌دهید، اشاره می‌کنید. به عنوان مثال: `gen `tv1' = income * 12`

۱۳۰- نتایج ذخیره شده

دستوراتی که خروجی را برمی‌گرداند اغلب نتایج را در حافظه ذخیره می‌کند که می‌تواند در برنامه‌ها استفاده شود؛ به عنوان مثال، دستور `summary` علاوه‌بر نمایش آمار خلاصه بر روی صفحه نمایش، نتایج زیر را ذخیره می‌کند:

<code>r(N)</code>	number of observations	<code>r(p1)</code>	1st percentile (detail only)
<code>r(mean)</code>	mean	<code>r(p5)</code>	5th percentile (detail only)
<code>r(skewness)</code>	skewness (detail only)	<code>r(p10)</code>	10th percentile (detail only)
<code>r(min)</code>	minimum	<code>r(p25)</code>	25th percentile (detail only)
<code>r(max)</code>	maximum	<code>r(p50)</code>	50th percentile (detail only)
<code>r(sum_w)</code>	sum of the weights	<code>r(p75)</code>	75th percentile (detail only)
<code>r(Var)</code>	variance	<code>r(p90)</code>	90th percentile (detail only)
<code>r(kurtosis)</code>	kurtosis (detail only)	<code>r(p95)</code>	95th percentile (detail only)
<code>r(sum)</code>	sum of variable	<code>r(p99)</code>	99th percentile (detail only)
<code>r(sd)</code>	standard deviation		

دستور میانگین نتایج را در `macros / scalars / matrices` مختلف (e) ذخیره می‌کند (به `help of mean command` مراجعه کنید)؛

توجه: برخی از بسته‌ها (به عنوان مثال، فقر) نتایج را در متغیرهای کلان سراسری^۱ ذخیره می‌کنند.

1. Global macro variables



۱۳۱- استفاده از نتایج ذخیره شده: یک مثال

دستوراتی که خروجی را برمی‌گرداند اغلب نتایج را در حافظه ذخیره می‌کند که می‌تواند در برنامه‌ها استفاده شود؛ راهنمای فرمان را برای فهرستی از نتایج ذخیره شده (در صورت موجود بودن) ببینید؛ به عنوان مثال، دستور summarize علاوه بر نمایش آمار خلاصه بر روی صفحه نمایش، نتایج زیر را ذخیره می‌کند

r(N)	number of observations	r(p1)	1st percentile (detail only)
r(mean)	mean	r(p5)	5th percentile (detail only)
r(skewness)	skewness (detail only)	r(p10)	10th percentile (detail only)
r(min)	minimum	r(p25)	25th percentile (detail only)
r(max)	maximum	r(p50)	50th percentile (detail only)
r(sum_w)	sum of the weights	r(p75)	75th percentile (detail only)
r(Var)	variance	r(p90)	90th percentile (detail only)
r(kurtosis)	kurtosis (detail only)	r(p95)	95th percentile (detail only)
r(sum)	sum of variable	r(p99)	99th percentile (detail only)
r(sd)	standard deviation		

۱۳۲- دستور display

Display رشته‌ها و مقادیر عبارات scalar را نمایش می‌دهد. از برنامه‌هایی که می‌نویسید خروجی تولید می‌کند. برای مثال می‌توان از آن برای نمایش نتیجه یک دستور یا مقدار یک ماکرو استفاده کرد؛ مثال ۱:

```
summarize hhsz // Produce summary stats of variable hhsz
display "Variable hhsz has a mean of " r(mean) " and a max of " r(max)
```

مثال ۲:

```
display "Today is the: " c(current_date) // c(current_date) = the system
date
```

۱۳۳- حلقه‌ها (Loops)

- بسیاری از برنامه‌ها حاوی دستورات یا مجموعه‌ای از دستورات هستند که باید تکرار شوند (به عنوان مثال، ممکن است لازم باشد مقادیر هر سال را در بازه‌ای از سال‌ها محاسبه کنید)؛
- Stata روش‌های مختلفی را برای looping یا تکرار دستورات در یک فایل do ارائه می‌دهد؛
- بسته به هدف حلقه، ممکن است بخواهید یکی از روش‌ها را نسبت به روش دیگر انتخاب کنید (در برخی موارد، بیش از یک روش ممکن است به همان نتیجه برسد، اما ممکن است یک روش از روش‌های دیگر 'موزون‌تر' یا کارآمدتر باشد).



۱۳۴- حلقه‌ها با استفاده از 'while'

اولین گزینه برای ایجاد حلقه در فایل do استفاده از دستور while است؛ Stata دستورات مشخص شده در حلقه را تا زمانی که شرط while برآورده شود تکرار می‌کند؛ به‌طور معمول، زمانی که مجموعه دستورات باید چندین بار تکرار شوند (مثلاً ۵ حلقه) استفاده می‌شود.

۱۳۵- حلقه‌ها با استفاده از 'while' - مثال

دستوری را اجرا می‌کنیم که ارزش سال تقویمی از ۲۰۰۰ تا ۲۰۲۰ را با افزایش ۵ نشان می‌دهد.

```
local year = 2000
while `year' <= 2020 {
    display "Calendar year is now: " `year'
    local year = `year' + 5
}
```

۱۳۶- حلقه‌ها با استفاده از 'forvalues'

راه دیگر برای دستیابی به یک حلقه از طریق مقادیر عددی، بیشترین استفاده از 'forvalues' است.

```
forvalues lname = range {
    commands referring to 'lname'
}
```

جایی که محدوده هست:

- #1(#d)#2 meaning #1 to #2 in steps of #d
- #1/#2 meaning #1 to #2 in steps of 1
- #1 #t to #2 meaning #1 to #2 in steps of #t - #1
- #1 #t : #2 meaning #1 to #2 in steps of #t - #1



۱۳۷- حلقه‌ها با استفاده از 'foreach'

foreach همراه با رشته‌ها استفاده می‌شود.

```
foreach country in KIR FSM FJI {
    display "The selected country is "`country`"
}
```

این دستور را می‌توان با نام متغیرها، اعداد یا هر رشته‌ای از متن استفاده کرد.

۱۳۸- حلقه‌ها با استفاده از 'levelsof'

levelsof لیست مرتب شده‌ای از مقادیر متمایز یک متغیر طبقه‌بندی شده را نمایش می‌دهد. با استفاده از این دستور، می‌توانید یک ماکرو حاوی لیستی از این مقادیر ایجاد کنید و از این لیست برای حلقه زدن مقادیر استفاده کنید.
مثال:

```
levelsof ethnicgrp, local(ethnic)
foreach l of local ethnic {
    ... some commands to be run for each value of ethnic
}
```

۱۳۹- انشعاب (Branching)

ممکن است بخواهیم برخی از دستورات را زمانی که یک شرط خاص برآورده شده است، و مجموعه دیگری از دستورات را زمانی که شرط برآورده نشده است، اجرا کنیم. این کار با "branching" با استفاده از دستورات 'if' و 'else' انجام می‌شود. پیاده‌سازی در Stata به شرح زیر است:

```
if [condition] {
    ... execute these commands ...
}
else {
    ... execute these other commands ...
}
```

به استفاده از براکت‌های مجعد { } و { توجه کنید. مجموعه دستوراتی که باید تحت هر شرایطی اجرا شوند باید در مجموعه‌ای از براکت‌ها فهرست شوند.



۱۴۰- حفظ (Preserving) و بازیابی (restoring) اطلاعات در حافظه

- مشکل برنامه نویسی را `preserve` و `restore` کنید که در آن داده‌های کاربر باید تغییر کند تا به نتیجه مطلوب برسد، اما وقتی برنامه به پایان رسید، برنامه‌نویس می‌خواهد آسیب وارد شده به داده‌ها را جبران کند؛
- هنگامی که `preserve` نشر می‌شود، داده‌های کاربر حفظ می‌شود. داده‌های موجود در حافظه بدون تغییر باقی می‌مانند. هنگامی که برنامه یا فایل `do` به پایان می‌رسد، داده‌های کاربر به‌طور خودکار بازیابی می‌شوند؛
- پس از `preserve`، برنامه‌نویس همچنین می‌تواند به `Stata` دستور دهد که اکنون داده‌ها را با دستور `restore` بازیابی کند. این زمانی مفید است که برنامه‌نویس به داده‌های اصلی نیاز داشته باشد و بداند که آسیب بیشتری به داده‌ها وارد نخواهد شد؛
- هنگامی که برنامه‌نویس به داده‌ها نیاز دارد اما قصد آسیب بیشتر دارد، می‌توان از `restore`، `save` استفاده کرد. هنگامی که برنامه‌نویس می‌خواهد ذخیره قبلی را لغو کند و داده‌های موجود در حافظه فعلی را به کاربر بازگرداند، نمی‌تواند استفاده شود.
- (توضیحات استخراج شده از راهنمای `Stata`)

۱۴۱- اجرای بی سر و صدا یا پرسر و صدای دستورات

در برخی موارد، ممکن است بخواهید دستوری را اجرا کنید اما خروجی ترمینال را نشان ندهید. این را می‌توان با استفاده از دستور `quietly` انجام داد.

Syntax: `quietly [:] command`

مثال:

`quietly regress pce province industry hhsz`

→ موجود است $e()$ هیچ خروجی ارائه نشده است، اما نتایج

توجه: می‌توانید بی‌صدا با `{ }` ترکیب کنید تا یک بلوک از دستورات را بی‌صدا اجرا کنید (و در صورت نیاز از دستوری در این بلوک به‌طور بدون صدا استفاده کنید).



۱۴۲- اشکال زدایی (Debugging) یک برنامه

ممکن است برنامه شما در نیمه راه به دلایلی از کار بیفتد. به عنوان مثال، اگر در حال تلاش برای ایجاد یک متغیر جدید به نام age هستید اما در حال حاضر متغیری به نام age وجود دارد.

```
use "individual.dta", clear
generate age = 10
variable age already defined
```

وقتی برنامه ساده باشد، تشخیص علت مشکل آسان است. با برنامه‌های پیچیده، همیشه آن قدر واضح نیست. دستور set trace که اجرای برنامه را خط به خط ردیابی می‌کند، ممکن است به شناسایی مشکل کمک کند.

۱۴۳- کار با فایل‌های CSV و Excel

۱۴۴- وارد کردن داده‌ها از یک فایل CSV

برای وارد کردن داده‌ها از یک فایل CSV، از import delimited استفاده کنید. شما این امکان را دارید که ردیف اول داده‌های CSV را به عنوان نام متغیر Stata در نظر بگیرید و محدوده خاصی از سطرها/ستون‌ها را انتخاب کنید.

Syntax: import delimited [varlist] filename [, options]

مثال:

Example:

```
* Importing a CSV file, where the first row contains variable names
import delimited "household.csv", clear
* We do the same, but for a selection of columns and rows of the CSV file
* (we keep the first 5 variables, and the top 50 observations)
import delimited "household.csv", rowrange(1:50) colrange(1:5) clear
```



۱۴۵- وارد کردن داده‌ها از یک کاربرگ اکسل

برای وارد کردن هر کاربرگ (یا محدوده سلولی^۱) از فایل XLS یا XLSX از `import excel` استفاده کنید. شما این امکان را دارید که ردیف اول داده‌های اکسل را به عنوان نام متغیر `Stata` در نظر بگیرید.

Syntax:

```
import excel [using] filename [, import_excel_options]
```

Example: `import excel "household.csv", clear`

(برای گزینه‌های بیشتر به کتابچه راهنمای `Stata` مراجعه کنید)

۱۴۶- خواندن سلول‌های خاص از یک کاربرگ اکسل

می‌توانید سلول‌های خاصی را از یک کاربرگ اکسل بخوانید و مقادیر را به عنوان متغیرهای کلان برای استفاده در برنامه‌های `Stata` ذخیره کنید. مثلاً:

```
import excel using "C:\poverty_lines.xlsx", cellrange(B1:C1) clear
```

```
local ctry = B
```

```
local year = C
```

۱۴۷- ذخیره یک فایل داده `Stata` با فرمت اکسل

برای ذخیره فایل داده `Stata` (همه متغیرها یا یک زیرمجموعه) در یک برگه `Excel` از `export excel` استفاده کنید. این امکان را دارید که کل یک کتاب کار را جایگزین کنید یا داده‌ها را به عنوان یک کاربرگ جدید در یک کتاب کار موجود ذخیره کنید. می‌توانید نام متغیر `Stata` یا برچسب متغیر را به عنوان ردیف اول کاربرگ ذخیره کنید. می‌توانید مقادیر یا برچسب‌های ارزش مربوطه را `export` کنید.

Syntax:

```
export excel [using] filename [if] [in] [, export_excel_options]
```

یا (برای `export` فقط زیرمجموعه‌ای از متغیرها)

1. Custom cell range



۱۴۸- ذخیره مقادیر در برگه‌های اکسل

برای ذخیره نتایج محاسبات Stata در سلول‌های خاص یک فایل اکسل، از putexcel استفاده می‌کنید. دستور putexcel set putexcel مورد استفاده و برخی از گزینه‌های قالب بندی را نشان می‌دهد. دستور putexcel مقادیر (از یک ماکرو یا ماتریس Stata) را در فایل اکسل می‌نویسد.
برای مثال:

```
putexcel set "poverty_lines", sheet("Sheet1") modify keepcellformat
putexcel B27 = matrix(WI) // B27 = top right corner of matrix
putexcel F13 = ("$_DATE")
putexcel F14 = ("$_TIME")
putexcel K20 = (`poverty_headcount')
```

۱۴۹- تعامل با اکسل: یک مثال

در این مثال، مقدار خطوط فقر مختلف را از یک برگه اکسل استخراج می‌کنیم، از این خطوط فقر برای محاسبه شاخص‌های فقر استفاده می‌کنیم و شاخص‌های فقر انتخاب شده را در فایل XLSX ذخیره می‌کنیم.
نکته ۱: برای اجرای مثال نشان داده شده در اسلاید بعدی، بسته poverty باید نصب شده باشد (ssc install poverty).
نکته ۲: بسته فقر نتایج مختلفی را که در ماکروهای جهانی به نام S_1 تا S_27 ایجاد می‌کند، ذخیره می‌کند (به فایل راهنمای بسته مراجعه کنید). ماکروهای جهانی در برنامه‌های Stata با نام S_1\$ تا S_27\$ شناخته می‌شوند.



۱۵۰- تعامل با اکسل: یک مثال

```
* We read values of poverty lines in Excel, calculate poverty in Stata, and save results in Excel.
set more off
cd "C:\Stata_Fiji\Data"
local myxls = "Test_poverty_lines.xlsx" // Excel file containing poverty lines
putexcel set "myxls", modify // Will save results in that same file

forvalues i = 10(1)18 { // Poverty lines are stored in cells B10 to B18
    import excel using "`myxls'", cellrange(B`i') clear // Read poverty line value
    local pline = B // Store it in a macro
    use "household.dta", clear
    gen pcc = tot_exp/ haise
    poverty pcc [aweight = wgtpop], line(`pline') all // Calculate poverty indic.
    putexcel C`i' = ($S_6) // Package poverty saves output in global macros
    putexcel D`i' = ($S_8) // We save two of the results in Excel (cols C and D)
}
putexcel C6 = (`$S_DATE') // We save the date in cell C6
```

۱۵۱- تعامل با سیستم عامل

۱۵۲- تعامل با سیستم عامل

در برخی از برنامه‌ها، ممکن است بخواهید برخی از دستورات را از خط فرمان عملیاتی اجرا کنید، به عنوان مثال برای پاک کردن یک فایل یا به دست آوردن لیستی از فایل‌ها در یک فهرست. می‌توانید چنین دستورات py را قبل از آنها با یک ! اجرا کنید. مثال‌ها:

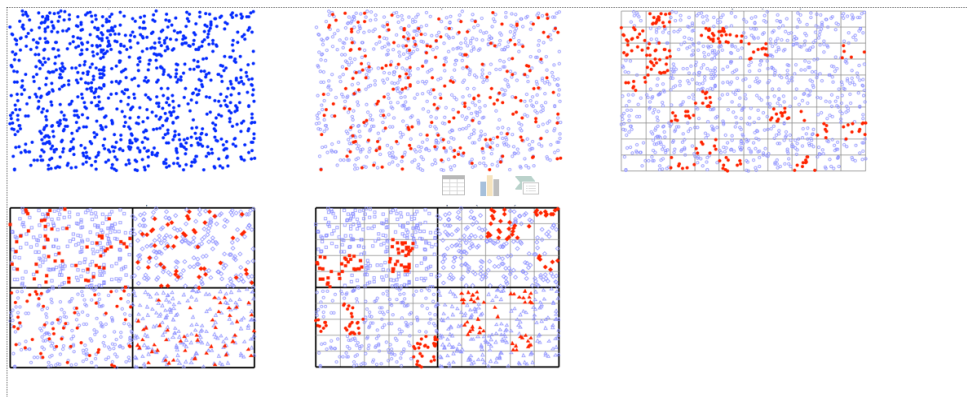
```
!dir C:\FSM
```

```
!erase "C:\FSM\temporary_file.dta"
```

۱۵۳- دستورات خاص برای جدول‌بندی و تجزیه و تحلیل داده‌های پیمایش



۱۵۴- چند مثال از نمونه طرح‌ها



Source: Jeff Pitblado, Associate Director, Statistical Software at StataCorp LP. 2009 Canadian Stata Users Group Meeting. Available at http://www.stata.com/meeting/canada09/ca09_pitblado_handout.pdf

155

۱۵۵- تعریف طرح پیمایش

- طراحی نمونه می‌تواند بر خطاهای استاندارد حاصل از نتایج تحلیل‌های آماری تأثیر بگذارد. تجزیه و تحلیل باید ویژگی‌های طراحی پیمایش را در نظر بگیرد؛
- برای انجام این کار، باید دستور `svyset` را انجام دهیم تا به `Stata` در مورد طرح نمونه بگوییم. شما از `svyset` برای تعیین متغیرهایی استفاده می‌کنید که حاوی اطلاعاتی در مورد طرح پیمایش هستند، مانند واحدهای نمونه‌گیری و وزن؛
- هنگامی که این دستور ایجاد شد، می‌توانید از پیشوند `svy`: قبل از هر دستور استفاده کنید.

۱۵۶- تعریف طراحی پیمایش - Syntax

برای طراحی تک مرحله‌ای:

```
svyset [psu] [weight] [, design_options options]
```

برای طراحی چند مرحله‌ای:

```
svyset psu [weight] [, design_options] [|| ssu , design_options]... [options]
```



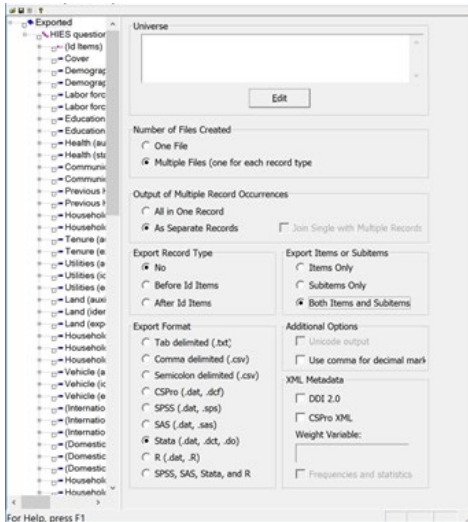
۱۵۷ استفاده از svy: دستورات

- بعد از svyset می‌توانید از دستورات زیادی با پیشوند svy استفاده کنید و نتایج دقیق‌تری دریافت خواهید کرد.
- برخی از دستوراتی که می‌توانند از svy استفاده کنند:

- آمار توصیفی: میانگین (mean)
- تخمین یعنی نسبت: نسبت (proportion)
- تخمین تناسب نسبت: نسبت (ratio)
- تخمین نسبت کل: کل (total)
- رگرسیون خطی: پسرفت (regress)

۱۵۸- وارد کردن داده‌ها از CsPro

۱۵۹- برای نشر یک مجموعه داده CsPro به Stata



یک پوشه جدید ایجاد کنید که در آن مواد انجام شده را ذخیره کنید؛

داده‌های CsPro مربوط به فایل‌هایی که باید روی آن کار انجام شود را باز کنید، سپس **Tools > Export Data** را انتخاب کنید.

کادر محاوره‌ای Export باز خواهد شد.

همان‌طور که در اسلاید بعدی نشان داده شده است گزینه‌ها را وارد کنید.

CsPro مجموعه‌ای از فایل‌ها را تولید می‌کند (برای ذخیره در پوشه جدید). این فایل‌ها حاوی مواد مورد نیاز

برای تولید فایل‌های داده Stata هستند (نه خود فایل‌های داده Stata).

برای تولید فایل‌های داده در فرمت Stata و ذخیره آنها باید همه فایل‌های [do] را اجرا کنید.



۱۶۰- گزینه‌های اکسپورت در CsPro

گزینه‌ها را به صورت زیر انتخاب کنید:

Number of Files Created <input type="radio"/> One File <input checked="" type="radio"/> Multiple Files (one for each record type)	
Output of Multiple Record Occurrences <input type="radio"/> All in One Record <input checked="" type="radio"/> As Separate Records <input type="checkbox"/> Join Single with Multiple Records	
Export Record Type <input checked="" type="radio"/> No <input type="radio"/> Before Id Items <input type="radio"/> After Id Items	Export Items or Subitems <input type="radio"/> Items Only <input type="radio"/> Subitems Only <input checked="" type="radio"/> Both Items and Subitems

Export Format

Tab delimited (.txt)
 Comma delimited (.csv)
 Semicolon delimited (.csv)
 CsPro (.dat, .dcf)
 SPSS (.dat, .sps)
 SAS (.dat, .sas)
 Stata (.dat, .dct, .do)
 R (.dat, .R)
 SPSS, SAS, Stata, and R

۱۶۱- اکسپورت CsPro به Stata

- اکسپورت CsPro به Stata برای هر نوع رکورد در CsPro ایجاد می‌کند:
 - فایل One do (پسوندها DO)
 - یک فایل فرهنگ (پسوندها DCT)
 - یک فایل داده (پسوندها DAT)
- CsPro فایل‌های داده Stata را تولید نمی‌کند. مواد مورد نیاز برای تولید فایل‌های داده Stata را تولید می‌کند.
- این می‌تواند شامل اجرای بسیاری از فایل‌های do (یکی برای هر نوع رکورد) باشد.
- می‌توان آنها را یکی یکی اجرا کرد یا یک فایل do تولید کرد تا آنها را در یک دسته اجرا کند.



۱۶۲- اجرای فایل‌های do جداگانه

- برای هر نوع رکورد در مجموعه داده CsPro، CsPro یک فایل DAT (یک فایل با فرمت ثابت ASCII حاوی داده‌ها برای هر نوع رکورد خاص)، یک فایل DCT که حاوی اطلاعات موقعیت هر متغیر در فایل DAT است، تولید می‌کند، و برچسب‌های متغیر و مقدار، و یک فایل DO که اطلاعات DCT را روی فایل DAT اعمال می‌کند.
- برای هر فایل do، باید کد زیر را (در Stata) اجرا کنید:

```
clear
do filename.DO
compress
save "filename.dta", replace
```

۱۶۳- اجرای همه فایل‌های do به صورت دسته‌ای با استفاده از یک فایل do

```
clear *
set more off
cd "C:\FSM_HIES_2013\CsPro" // Where the CsPro export files (DCT, DAT, DO) are stored

local outdir = "C:\FSM_HIES_2013\Stata" // Where we will save the Stata data files

capture !erase listDCT.txt // Delete the list of CsPro .DCT files if it exists
!dir *.dct /B -> listDCT.txt // Create a text file containing the list of CsPro .DCT files

file open Recs using listDCT.txt, read // Open that text file containing the list of files
file read Recs line

while r(eof)==0 { // We will read the lines one by one, until we reach the end of file (eof)
    local filenm = substr("`line'",1,length("`line')-4) // Remove ".DCT" to keep only the file name
    clear
    do " `filenm'.do" // Run the do file (convert data from ASCII to Stata, and add labels)
    compress
    save " `outdir'\`filenm'.dta", replace
    file read Recs line // If not last line of the text file, read next line
}

file close Recs // Job completed; we can close the text file
```




DDI Metadata Editor (Nesstar Publisher 4.0.10)



SCREENSHOTS





ویرایشگر فراداده شبکه بین‌المللی پیمایش‌های خانوار، همچنین به عنوان ناشر Nesstar شناخته می‌شود، یک ویرایشگر غنی از ویژگی‌ها برای آماده‌سازی فراداده‌ها و داده‌ها برای انتشار در یک کاتالوگ آنلاین، مانند آرشیو داده ملی (NADA) توسعه یافته شبکه بین‌المللی پیمایش‌های خانوار است. فراداده تولید شده توسط ویرایشگر با طرح مستندسازی داده (DDI) n.2 و استانداردهای فراداده Dublin Core XML مطابقت دارد. این نرم‌افزار توسط Nesstar در آرشیو داده‌های علوم اجتماعی نروژ (NSD) توسعه یافته و به عنوان نرم‌افزار رایگان توزیع شده است.

ویژگی

ابزارهایی برای محاسبه/بازنویسی/برچسب‌گذاری متغیرهای جدید یا موجود که باید قبل از انتشار به مجموعه داده اضافه شوند.	ویرایش/ایجاد و نشر آسان مجموعه داده‌های مستند DDI با تجربه XML مورد نیاز
امکان وارد کردن و نشر داده‌ها به رایج‌ترین فرمت‌های آماری، از جمله فایل‌های محدود شده.	ابزارهایی برای اعتبارسنجی متادیتا و متغیرها
چند زبانه: عربی، چینی، انگلیسی، فرانسوی، پرتغالی، روسی و اسپانیایی	امکان گنجاندن فراوانی تولید شده اتومات و آمار خلاصه برای هر متغیر

نیازهای سیستم

ویرایشگر فراداده / Nesstar Publisher به رایانه‌ای نیاز دارد که مایکروسافت ویندوز را اجرا می‌کند.

قرارداد مجوز

Nesstar Publisher به عنوان نرم‌افزار رایگان توسط NSD توزیع شده است. قرارداد مجوز در منوی Help برنامه ارائه شده است.

دانلودها

ویرایشگر فراداده / Nesstar Publisher را می‌توان به صورت رایگان دانلود کرد. علاوه بر این، کاربرانی که قصد استفاده از برنامه کاتالوگ کردن آنلاین NADA را دارند، باید الگوهای شبکه بین‌المللی پیمایش‌های خانوار موجود در زیر بخش زیر، قالب‌های شبکه بین‌المللی پیمایش‌های خانوار را نیز دانلود و نصب کنند.

نرم‌افزار




الگوهای شبکه بین‌المللی پیمایش‌های خانوار

ویرایشگر/ناشر فراداده از "الگوها" استفاده می‌کند، که مشخص می‌کند کدام عناصر DDI و Dublin Core باید برای مستندسازی مجموعه‌های داده استفاده شوند. نرم‌افزار با قالب پیش فرض ارائه شده است. کاربرانی که می‌خواهند از سایر برنامه‌های شبکه بین‌المللی پیمایش‌های خانوار استفاده کنند، به ویژه برنامه کاتالوگ کردن NADA، باید از الگوهای شبکه بین‌المللی پیمایش‌های خانوار استفاده کنند. الگوهای شبکه بین‌المللی پیمایش‌های خانوار به چندین زبان در دسترس هستند. نسخه فعلی نسخه ۱.۶ است.

افزونه‌ها^۱ و ابزارهای بیشتر


پلاگین مرکز گزارش شبکه بین‌المللی پیمایش‌های خانوار، که به ویژه حاوی ابزاری برای تبدیل DDI به یک سند PDF بود، متوقف شده است. آخرین نسخه Publisher / Metadata Editor شامل یک PDF Generator-ator است. نسخه ۴.۰ برنامه NADA همچنین یک PDF Generator را ارائه می‌دهد. ابزارهایی برای ترجمه Nesstar Publisher به زبان‌های دیگر:



Nesstar Translation Tool

A tool used for Translating the Nesstar Publisher interface into other languages.

[Download\(1.7MB\)](#)



Nesstar Translation Manager

A tool to manage the language translation files generated by the Nesstar Translation Tool and used to export the final language files for use with the Nesstar Publisher.

[Download\(1.6MB\)](#)

بهبودهای برنامه‌ریزی شده برای ناشر Nesstar

- DDI 2.5 compatibility
- UTF8 compliance



منابع مرتبط به اشتراک گذاشته شده است:

Statistical Disclosure Control (sdcmicro)

SDCMicro یک بسته متن باز مبتنی بر R برای تولید ریزداده‌های محافظت شده برای محققان و استفاده عموم است. داده‌های آژانس‌های آماری و سایر مؤسسات عمدتاً محرمانه هستند. این بسته می‌تواند برای تولید داده‌های بی‌نام (ریز داده) به عبارت دیگر، برای ایجاد فایل‌های عمومی و علمی استفاده شود. علاوه بر این، روش‌های مختلف برآورد خطر گنجانده شده است. بسته مرتبط sdcmicroGUI شامل یک رابط کاربری گرافیکی برای روش‌های مختلف در بسته sdcmicro است.

نویسندگان

Templ M, Kowarik A, Meindl B (2015). "Statistical Disclosure Control for Micro-Data Using the R Package sdcmicro." *Journal of Statistical Software*, 67(4), 1–36. doi: 10.18637/jss.v067.i04

نگهدارنده

Matthias Templ (matthias.templ at gmail.com)

ویژگی

رابط کاربری گرافیکی برای کاربران با دانش محدود از زبان R در دسترس است.	بسته متن باز رایگان مبتنی بر R که تحت مجوز GPL 2.0 توزیع شده است.
چندین روش برای ارزیابی خطر شناسایی مجدد (k بی‌نام بودن، فردی و سراسری خطر شناسایی مجدد).	گزینه‌های متعدد برای کاهش خطر افشا و برای ارزیابی از دست دادن اطلاعات.

نیازهای سیستم



MS Windows XP، Windows Vista، Windows 7، دارای، که نیاز دارد که رایانه‌ای نیاز دارد که دارای، MS Windows XP، Windows Vista، Windows 7، Windows 8، Mac OS X یا Unix/Linux باشد.

پیشرفت‌های اخیر

بودجه بانک جهانی و دپارتمان توسعه بین‌المللی بریتانیا برای توسعه یک رابط کاربری گرافیکی جدید برای بسته متن باز R sdcMicro استفاده شده است. برای کاربرانی که با استفاده از R راحت هستند، بسته sdcMicro ابزاری را برای استفاده از مجموعه جامعی از روش‌هایی که معمولاً در متون مربوط به کنترل افشا استفاده می‌شوند و شرح داده می‌شوند، ارائه می‌کند. کاربرانی که با R آشنا نیستند، اما نیاز فوری به ابزارهایی برای بی‌نام کردن ریز داده‌ها دارند، از یک رابط کاربری گرافیکی^۱ برای بسته sdcMicro بهره‌مند خواهند شد. برای ارائه یک محیط رابط کاربری گرافیکی برای کاربر غیر R، یک برنامه کاربردی Shiny به نام sdcApp به بسته sdcMicro اضافه شده است. کاربران sdcApp می‌توانند پرکاربردترین روش‌های بی‌نام‌سازی موجود در بسته sdcMicro را بدون نیاز به دانش عمیق R پیاده‌سازی کنند. علاوه بر روش‌های بی‌نام‌سازی پیاده‌سازی شده در بسته sdcMi- sdcApp، مجموعه‌ای جامع از اقدامات خطر و بهره‌وری را ارائه می‌دهد. این شامل توابعی برای اندازه‌گیری، تجسم و مقایسه خطر و بهره‌وری در طول فرآیند بی‌نام‌سازی است. sdcApp همچنین با تهیه گزارش‌هایی در مورد روش‌های مورد استفاده در قالب‌های مناسب برای مخاطبان داخلی و خارجی به کاربران کمک می‌کند و کد underlying R را برای اطمینان از تکرارپذیری ذخیره می‌کند. برای کاربران سایر بسته‌های آماری، sdcApp از وارد کردن و صادر کردن ریز داده‌ها در چندین فرمت (R، CSV، SPSS، SAS، STATA) پشتیبانی می‌کند. مانند sdcMicro، متن باز است و در مخازن CRAN و در GitHub موجود است.

دانلود

Version: 5.4.1

Date: 2018-04-13

License: GPL 2.0

Development URL: <https://github.com/sdcTools/sdcMicro>

The software and its documentation can be downloaded from the follow-

ing link: <http://cran.r-project.org/web/packages/sdcMicro/index.html>.

1. Graphic User Interface (GUI)



منابع مرتبط به اشتراک گذاشته شده است:

Microdata Cataloging Tool (NADA)

کاتالوگ داده NADA یک نرم‌افزار متن باز طراحی شده برای محققان برای مرور، جستجو، مقایسه، درخواست دسترسی و دانلود داده‌های تحقیق است.

نیازهای سیستم

NADA به PHP و پایگاه داده MySQL نیاز دارد.

نیازهای PHP

حداقل نسخه مورد نیاز PHP 7.0 است اما ما قویاً توصیه می‌کنیم از PHP 7.2 یا بالاتر استفاده کنید. پسوند‌های PHP زیر برای NADA مورد نیاز است:

xsl

xml

mbstring

mysqli



الزامات وب سرور

Apache 2.4 or later

Internet information services (IIS) 7 or later

NGINX

پایگاه‌های داده پشتیبانی شده

MySQL 5.x (recommended 5.7)

Microsoft SQL Server with fulltext support



منابع

<https://www.ihnsn.org/>