Digitala verktyg för riskbedömning och skyddsåtgärder i tabulära data med personuppgifter

README.docx

Datum: 2024-01-29  
Version: 1

Vänligen citera som:

Svensk Nationell Datatjänst. (2024). *Digitala verktyg för riskbedömning och skyddsåtgärder i tabulära data med personuppgifter.* Zenodo. https://doi.org/10.5281/zenodo.10777054

Det här verket är licensierat under en [Creative Commons Erkännande 4.0](https://creativecommons.org/licenses/by/4.0/deed.sv) licens (CC-BY 4.0)

**Titel:** Digitala verktyg för riskbedömning och skyddsåtgärder i tabulära data med personuppgifter.

**Beskrivning**:

Syftet med workshopen är att introducera deltagarna till grundläggande begrepp och praktiska verktyg för statistisk röjandekontroll (statistical disclosure control) i tabulära data. Under workshopen belyses såväl möjligheter som begränsningar i befintliga verktyg och metoder för pseudonymisering, med introduktion av R-paketet sdcMicro för att illustrera hur statistiska metoder kan tillämpas för att bedöma risk för röjande och bakvägsidentifiering av individer i tabulära dataset. Deltagarna får en grundläggande genomgång av verktygets principer och funktioner och kommer därefter själva att få möjlighet att tillämpa sina kunskaper på ett syntetiskt dataset.

Inga förkunskaper i R krävs.

**Typ av undervisning:** Workshop. Lämplig för utbildning på plats men skulle kunna anpassas för distans.

**Skaparen:** Svensk nationell datatjänst, SND

**Skapades**: 2024-01-29

**Version**: 1

**Språk**: svenska

**Licens:** CC-BY 4.0

**Mål**:

Efter genomförd workshop ska deltagarna:

* Ha en grundläggande överblick av statistiska koncept och metoder för att pseudonymisera tabulära data.
* Kunna identifiera vilka variabler som bör inkluderas i en riskbedömning och röjandekontroll.
* Vara bekanta med grundläggande tillvägagångssätt och funktioner för pseudonymisering och omkodning med sdcMicro genom webbgränssnittet sdcApp.
* Kunna tolka och bedöma resultat av pseudonymisering och omkodning av data samt dess inverkan på datasetets användbarhet.
* Ha en förståelse för hur statistiska metoder för pseudonymisering och omkodning samspelar och kan kombineras med andra typer av skyddsåtgärder.

**Målgrupp**: Forskare, DAU-personal, forskningsdatarådgivare

**Förkunskapskrav, instruktör:**

* Välbekant med konceptet pseudonymisering/anonymisering.
* Ha en grundläggande överblick av metoder som används inom statistisk röjandekontroll.
* Tidigare praktisk erfarenhet av R är en fördel, men inte ett krav.
* Erfarenhet av att installera R, RStudio och R-paketet sdcMicro, för att hjälpa deltagare som stöter på problem.
* Erfarenhet av att arbeta igenom övningen som används i workshopen.

**Pedagogisk struktur, sammanfattning:**

Workshoppen inleds med en presentation från instruktör och en genomgång av verktyget följt av en övning där deltagare följer skriftliga instruktioner och utforskar verktygen själv i egen takt, enskilt eller i par.

**Planerad längd** - 2,5 timmar.

**Förberedelser, deltagare:**

Avsätt 30 minuter för att läsa genom workshopmaterialet som skickats ut och 30 minuter till att installera R, RStudio och sdcMicro. Se förslag till meddelande under Förberedelser, instruktören.

**Förberedelser, instruktören:**

1. Modifiera avsnittet **Agenda/hålltider** i *Workshopprogram.docx* så att det indikerar klockslag snarare än varaktighet för varje moment, om så önskas.
2. Denna information bör skickas till deltagare i god tid innan workshoppen:

***!! För den här workshopen krävs att du tar med egen dator.***

Reservera 30 minuter för att läsa igenom workshopmaterial, och 30 minuter till för att installera R, RStudio och sdcMicro. Lämna inte installationen till sista minuten! Du behöver lämna tid för att lösa eventuella problem med installationen, om sådana skulle uppstå.

* Få en uppfattning om vad du ska göra:
  + Läs genom *Workshopprogram.docx*
* Installera R, RStudio och sdcMicro och se till att du kan starta sdcApp:
  + Följ de förkortade instruktionerna i *Installera sdcApp med RStudio.docx*
  + Är något oklart, följ de mer utförliga ”Fullständiga instruktioner” i filen.
* I Workshopprogram.docx finns också en lista över “Externt material som är bra att gå igenom innan workshopen” som kan vara till nytta!

1. Redigera slide 1 i *Presentation Digitala Verktyg.pptx* och inkludera ditt namn och anknytning om så önskas. Redigera slide 4 Hålltider på samma sätt som ovan. Byt presentationstema osv om så önskas.
2. Deltagare får ta del av filer under mappen*Digitala verktyg workshop deltagarfiler.* Dessutom rekommenderas att infoga *Presentation Digitala verktyg.pptx* eller en pdf-version.

**Paketets Innehåll:**

*README.docx* – Detta dokument!

Mappen*Digitala verktyg workshop deltagarfiler* som innehåller:

*Workshopprogram.docx* – Workshop program inklusive Agenda och uppgifter.

*Installera sdcApp med RStudio.docx –* sammanfattning för deltagare om hur de installera R, RStudio, sdcMicro och sdcApp i förhand.

*Installing sdcApp with RStudio.pdf –* (as above in English).

*Syntetiskt dataset for sdcApp.csv –* dataset för övningen.

*Genomgång sdcAppSteg-för-steg.docx* – instruktioner för instruktören om vad ska visas i sdcApp-gränssnittet innan deltagarna själva får prova på egen hand.

*Presentation Digitala verktyg.pptx* – workshoppresentation, med introduktion till riskbedömning, skyddsåtgärder och röjandekontroll.

**Acknowledgements**

Vi är tacksamma för följande resurser som används i presentationen:

The [image of the Guess Who game being played at Internationalen Spieltage SPIEL 2008](https://commons.wikimedia.org/wiki/Category:Guess_Who%3F#/media/File:Essen08_-_Kdo_jsem.jpg) is used under a CC-SA license. The current work is separable-from this resource and does not remix, transform, or build upon this image. This image is not covered by the current work’s CC-BY license.

The [Guess Who Characters](https://www.printablee.com/post_guess-who-game-sheets-printable_214348/) from [printablee.com](https://www.printablee.com/) may be “[used for free](https://www.printablee.com/pages/license/)”.

The [R logo](https://en.m.wikipedia.org/wiki/File:R_logo.svg) is used under a CC-SA license. The current work is separable-from this resource and does not remix, transform, or build upon this image.  This image is not covered by the current work’s CC-BY license.

Amnesia logo from <https://amnesia.openaire.eu/> is used under a CC-BY license.

The [ARX logo](https://github.com/arx-deidentifier/arx/blob/master/logo/logo_256.png) is used under the Apache License Version 2.0 <https://github.com/arx-deidentifier>. The current work is separable-from this resource and is not a “Derivative Work”. This image is not covered by the current work’s CC-BY license.

The infographic [“A visual Guide to Practical Data De-identification"](https://fpf.org/blog/a-visual-guide-to-practical-data-de-identification/) is used under the CC-BY license.

The Schwantz.nl “Brown Chicken” cartoon from the [Anonymization-reference-card-for-researchers.pdf](https://lcrdm.nl/wp-content/uploads/2023/03/Anonymization-reference-card-for-researchers.pdf) is used under a CC-BY license.