

ADGANG TIL MIKRODATA

TABELLER, DATAKONFIDENTIALITET, MM.

ADGANG TIL MIKRODATA

- Tabeller med Peg-og-klik
 - Hvad er formålet?
 - Lav en tabel og send den hjem
 - Indhold, kilder, dokumentation
 - Formater
 - Lav selv en tabel og send den hjem
- Datafortrolighed
 - Hvad må (ikke) sendes ud
 - Mikrodata
 - Tabeller
 - Eksamen
- Brugerforum
 - Registrer + Subscribe

Tabeller med Peg-og-klik

Tabeller med Peg-og-klik

- Hvad er formålet med Peg-og-klik?
 - At gøre det muligt at trække oplysninger fra mikrodata på en nogenlunde nemt tilgængelig og fleksibel måde uden at have direkte adgang til mikrodata.
 - Flexibelt alternativ til DSTs Statistikbank og lignende
 - Erstatning for tidligere direkte SAS-adgang til mikrodata
 - Kompromis med DSTs Forskningservice

Tabeller med Peg-og-klik

- Lav en tabel, fx
 - Hvad er indkomsten i forskellige brancher i landets kommuner?
 - Hvad er den personlige indkomst for mænd og kvinder i forskellige aldersgrupper? ...Opdelt på pendlingsafstand?
 - Hvilke uddannelser er regionens indbyggere i gang med, opdelt på nuværende "højest fuldførte" uddannelse?
 - Hvor mange er flyttet mellem regionerne? ...Hvor mange beskæftigede?
 - Hvad er afkastningsgraden for regionens virksomheder i forskellige brancher? ...Opdelt på omsætning?

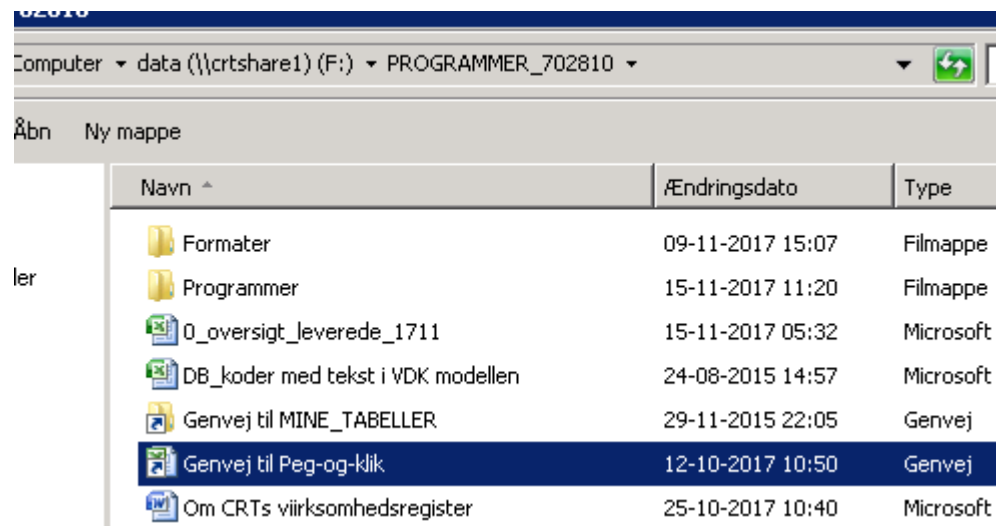
- Og send den hjem:



NB: Konfidentialitet!

Tabeller med Peg-og-klik

- Åbne Peg-og-klik
 - Åben browser: <https://remote.dst.dk>, log ind
 - Start Peg-og-klik (genvej), tillad evt. makroer



Computer > data (\\crtshare1) (F:) > PROGRAMMER_702810

Åbn Ny mappe

Navn ^	Ændringsdato	Type
Formater	09-11-2017 15:07	Filmappe
Programmer	15-11-2017 11:20	Filmappe
0_oversigt_leverede_1711	15-11-2017 05:32	Microsoft
DB_koder med tekst i VDK modellen	24-08-2015 14:57	Microsoft
Genvej til MINE_TABELLER	29-11-2015 22:05	Genvej
Genvej til Peg-og-klik	12-10-2017 10:50	Genvej
Om CRTs virksomhedsregister	25-10-2017 10:40	Microsoft

Tabeller med Peg-og-klik

- Personlige indkomst efter køn, alder og pendlingsafstand

2

The screenshot shows the 'Peg-og-klik' software interface with the following settings:

- Tabeller:** Lav en tabel | Særlige indstillinger | Noter
- Datasæt og population:** Persondata | 1. Standardbefolkning, alle personer
- Serie fra år:** Ej serie
- Sum i år:** 2015
- Sum af variabel:** PERINDKIALT_13
- Check:** Tynde celler | Dominans
- Akser og aggregeringer:**
 - Akse 1 (rækker):** 2015 | AFSTAND | AFSTAND_7a
 - Akse 2 (søjler):** 2015 | ALDER | ALDER_10A
 - Akse 3 (flere tabeller):** 2015 | KON | KON_3a
 - Akse 4 (flere tabeller):** (Empty)

Tabeller med Peg-og-klik

Afkastningsgraden for regionernes virksomheder i forskellige brancher, opdelt på omsætning

29

The screenshot shows the 'Peg-og-klik' software interface with the following configuration:

- Datasæt og population:** Virksomhedsdata (dropdown), 1. Virksomheder (dropdown)
- Serie fra år:** Ej serie (dropdown)
- Sum i år:** 2015 (dropdown)
- Sum af variabel (top row):** CRT_AKG (dropdown), [empty] (dropdown)
- Check:** Tynde celler, Dominans
- Sum af variabel (middle row):** [empty] (dropdown), [empty] (dropdown)
- Akser og aggregeringer:**
 - Akse 1 (rækker):** 2015 (dropdown), OMS (dropdown), OMS_9a (dropdown), [empty] (dropdown)
 - Akse 2 (søjler):** 2015 (dropdown), DB07 (dropdown), DB07_10a (dropdown), [empty] (dropdown)
 - Akse 3 (flere tabeller):** 2015 (dropdown), KOM (dropdown), KOM_5_regioner (dropdown), [empty] (dropdown)
 - Akse 4 (flere tabeller):** [empty] (dropdown), [empty] (dropdown), [empty] (dropdown), [empty] (dropdown)

Tabeller med Peg-og-klik

Datasæt og population

- Persondata
 - Hele befolkningen: Alle
 - RAS-befolkningen: Alle, der er i RAS, som opgøres november
- Virksomhedsdata
 - Virksomheder – defineret ved CVRNR
 - Arbejdssteder – defineret ved CVRNR*ARBNR
 - Ansatte (hovedbeskæftigede) – defineret ved CVRNR*ARBNR*PNR
 - NB: Kun "private byerhverv" dvs. "ekskl. landbrug, fiskeri, havne mv., pengeinstitutter, forsikring, almene boligselskaber, offentlig administration mv."

Tabeller med Peg-og-klik

Indstillinger

- Variabler: De valgte variabler summeres
 - Middelværdi er let at beregne; standardafvigelse, median, Gini-koefficient og andre mål tilføjes måske
- Akser bruges til at definere tabellernes struktur
 - Kan trækkes fra forskellige år, fx for at finde overgange eller betinge på tidligere år
 - Formater bruges til at bestemme aksekategorier

Tabeller med Peg-og-klik

Hvad indeholder grunddata?

- Persondata: Oplysninger om hele befolkningen (pr ultimo året)
 - Demografi: Køn, alder, etnicitet, bopæl, ...
 - Uddannelse: Højest fuldførte, igangværende
 - Beskæftigelse: Branche, social status, jobtype, arbejdssted, ...
 - Bolig: Størrelse, type, ejerforhold, ... (pt ikke fritidshuse)
 - Indkomst og formue: Mange variabler
 - Sundhed (kommer): Lægebesøg, indlæggelser
- Virksomhedsdata: Alle virksomheder og arbejdssteder i private byerhverv
 - Regnskabsoplysninger: Omsætning, investeringer, overskud, eksport, ...
- Generelt om geografi: Kommune og subkommunale (ikke adresser)
 - Postnummer, sogn, rode, storhed, dstnavn. Ghetto og kystnær tilføjes måske

Tabeller med Peg-og-klik

Hvor kommer grunddata fra? I hovedtræk:

- Persondata: Oplysninger om hele befolkningen (pr ultimo året)
 - Demografi: BEF, IEPE, DOD
 - Uddannelse: UDDA, INST
 - Beskæftigelse: RAS, LONN
 - Bolig: BBR, EJER, EJVK
 - Indkomst og formue: IND, AKM, REAL
 - Sundhed (kommer): SSSY, SYIN, måske LPRPOP
- Virksomhedsdata: Alle virksomheder og arbejdssteder i private byerhverv
 - Regnskabsoplysninger: FIRE, FIRM, FIRA (og udvalgte personoplysninger)
- Generelt om geografi: Kommune og subkommunale (ikke adresser)
 - BYSTRA, BYSTRB, BYSTRS

• http://www.dst.dk/da/TilSalg/Forskningservice/Data/Register_Variabeloversigter

Tabeller med Peg-og-klik

Dokumentation/Hvad indeholder variablerne?

- Oversigter i F:\PROGRAMMER_702810
 - "Om virksomhedsdata" kort opsummering
 - "Om persondata" kommer
 - "Oversigt_grunddata" har labels og år
 - "Oversigt_leverede" viser hvilket DST-register variabelen kommer fra
- Danmarks Statistiks dokumentation
 - Høj kvalitetsdata og TIMES:
<http://www.dst.dk/da/TilSalg/Forskningservice/Dokumentation>
 - Google: "dst.dk <variabelnavn>"
- Formater på forskermaskine



Tabeller med Peg-og-klik

Brug af egne formater til akser

- Nogle gange er de eksisterende formater ikke optimale
 - En bestemt opdeling er vigtig (fx aldersgrupper, brancher)
 - Mange kategorier er irrelevante (fx kommuner udenfor egen region)
- Lav et nyt format som løser opgaven
 - Tag udgangspunkt i eksisterende format fra F:\PROGRAMMER_702810\Formater, eller D:\...\MIT_KATALOG\MINE_FORMATER
 - Giv formatet et passende navn og sørg for at indhold og filnavn matcher. Det virker bedst hvis formatnavn/filnavn bruger variabelens navn.
 - Placér "fmt_<navn>.sas"-filen i "D:\...\MIT_KATALOG\MINE_FORMATER"
 - Tilpas formatet (SAS-koden) så det aggregerer som ønsket
 - Start Peg-og-klik
 - Så skulle det gerne virke



Tabeller med Peg-og-klik

- Praktiske bemærkninger
 - Tabellerne dannes for hele befolkningen, så lav evt. flere tabeller hvor én har de relevante personer
 - Brug "Testkørsel" til at tilpasse tabellen, især ved persondata
 - Output-tabellerne ligger som .xls-html filer i ".../MIT_KATALOG/MINE_TABELLER"
 - Der kan være forskel på det du gerne vil vide og det du spørger om. Programmet tænker ikke, det laver bare tabeller
- Hvis noget ikke fungerer
 - Der ligger log-filer i ".../MIT_KATALOG/MINE_TABELLER/LOG"
 - Selv hvis log-filer ikke umiddelbart hjælper, vedhæft dem gerne da de viser hvad der er kørt (kontakt: jens@crt.dk)

Datafortrolighed

Datafortrolighed

Hvad må ikke sendes ud?

- Ingen mikrodata eller oplysninger, som kan henføres til enkeltpersoner eller –virksomheder.

Retningslinjer findes på dst.dk:

- <http://www.dst.dk/da/TilSalg/Forskningssservice/Dataadgang>
- <http://www.dst.dk/da/TilSalg/Forskningssservice/Vejledninger>
- Spørg Forskningssservice hvis du er i tvivl

Generelt: Du må ikke sende noget ud, som er forbudt

Datafortrolighed

Kun personer, som er godkendt til projektet, må have adgang

Mikrodata

- Mikrodata må aldrig sendes ud. Ikke en eneste record, eller bare et eneste løbenummer, heller ikke hvis det bruges i en program-fil.
- Kun de 3 administratorer af Overvågningsmodellens projektdatabase har adgang til mikrodata i databasen (og kan ikke hjemsende fra database-projektet)
- Brugere kan få adgang til mikrodata til detaljerede undersøgelser ved projektansøgning til Forskningservice

Datafortrolighed

Tabeller og aggregerede data

Hvad betyder "Ingen mikrodata eller oplysninger, som kan henføres til enkeltpersoner eller –virksomheder"?

- Der er ingen fast definition af hvad der kan "henføres til enkeltpersoner eller –virksomheder"
- Undgå celler med meget få observationer og, i hvert fald for virksomhedsdata, celler hvor meget få observationer er dominerende
- Man må heller ikke kunne "regne tilbage" til de "tynde" celler
- Nogle data "værre" end andre; Statistikbanken har mange tabeller med enkeltobservationer (f.eks. FAM100N)
- Der er forskel på at sige, at der f.eks. i København findes en 71-årig mandlig jurist med stor negativ formue, og sige det samme om et lille geografisk område hvor personen kan identificeres

Peg-og-klik kan

- Skjule indholdet i celler med meget få observationer ("Tynde celler")
- Skjule indholdet i celler hvor de to største observationer udgør 80% eller mere af værdien ("Dominans")
- Dette er ingen garanti for at kravene til datafortrolighed er overholdt.
- Al hjemsendelse sker på eget ansvar.
- Tjek altid filens indhold før du hjemsender.

Datafortrolighed

Konsekvenser ved brud på datafortrolighed

- DST lukker adgang for alle forskere ved den institution, der ejer projektet, hvor databrudet har fundet sted
 - Også hvis der er en tilknyttet forsker, der har begået bruddet
- DST kontakter institutionen (forsker og dataansvarlig)
 - Redegørelse for databrudet
 - Plan for hvordan institutionen fremadrettet vil undgå nye brud
- DSTs direktion orienteres og tager stilling til sagen
- Sanktion meldes ud til institutionen
(antal måneders nedlukning, flere ved gentagelse, evt permanent)

Datafortrolighed - Eksamen

Datafortrolighed - adgang

1. Du er logget ind på dit projekt på forskermaskinen og sidder sammen med en kollega og arbejder på jeres projekt. Du skal til møde og bliver derfor nødt til at gå en time. Kollegaen er autoriseret bruger på samme projekt, så I beslutter at han overtager tastaturet og arbejder videre. Er det en overskridelse af reglerne?

A. Ja

B. Nej

Datafortrolighed - adgang

2. Du er ny bruger og har endnu ikke så megen erfaring med at arbejde med data. Derfor har du bedt dei erfarne kollega om hjælp. Så I har aftalt at han kommer forbi og sidder ved siden af dig og ser på skærmen. Er det i overensstemmelse med sikkerhedsreglerne?

- A. Hvis kollegaen er autoriseret bruger på samme projekt?
- B. Hvis kollegaen har en autorisationsaftale, men kollegaen behøver ikke være autoriseret bruger på det specifikke projekt?
- C. Hvis kollegaen alene skal hjælpe med programmering behøver han ikke være autoriseret bruger?

Datafortrolighed - adgang

3. Du sidder alene ved skærmen og arbejder, men skal lige hente kaffe. Du forlader derfor computeren og falder i snak ved kaffemaskinen og kommer først tilbage 15 minutter senere. Er det i overensstemmelse med sikkerhedsreglerne?

- A. Det er OK, hvis du låser skærmen.
- B. Nej – du skal afbryde forbindelsen til serveren eller logge helt af serveren, hvis du forlader computeren mere end et par minutter.

Datafortrolighed - adgang

4. Du sidder i et tog (eller på en café) og har din bærbare computer med. Er det tilladt at logge ind og sidde og arbejde på dit projekt?

- A. Ja – hvis jeg er alene og sikrer at ingen andre kan se med.
- B. Nej – det er ikke tilladt at logge ind på projekter under FSE fra andre steder end din arbejdsplads eller din hjemmearbejdsplads.

Datafortrolighed - mikrodata

5. Er det OK at udskrive mikrodata på skærmen for at validere data?

- A. Ja
- B. Nej

6. Du har fået leveret mikrodata om uddannelse og ønsker at analysere frafald på ungdomsuddannelserne. For at validere data forsøger du at identificere dig selv i data og se hvordan din egen uddannelseshistorik er registreret i registrene. Er det OK?

- A. Ja – så længe du alene forsøger at identificere dig selv.
- B. Ja – det er ok så længe du ikke har adgang til CPR-numre, men alene forsøger at identificere dig selv ud fra andre oplysninger.
- C. Nej, det er ikke tilladt at forsøge at identificere enkeltpersoner, heller ikke dig selv.

Datafortrolighed - hjemsendelse

7. Jeg har indsamlet en survey, som jeg sidder og arbejder med lokalt. Jeg har søgt Datatilsynets tilladelserne er i orden. Jeg har også sendt surveyen til DST for at have muligheden for at koble surveyen med registerdata.

- A. Jeg må gerne hjemsende mikrodata fra den survey jeg selv har fået lagt ind på forskermaskinen, og som jeg har tilladelse til at arbejde med lokalt.
- B. Det er kun mikrodata, som DST har leveret til mit projekt, jeg ikke må hjemsende.
- C. Det er under ingen omstændigheder tilladt at hjemsende mikrodata uanset om data er leveret af DST eller det er mine egne data.

Datafortrolighed - hjemsendelse

8. Det er ikke tilladt at hjemsende tabeller, der indeholder oplysninger om medianer, maksimum og minimum, da disse reelt er observationer for enkeltpersoner. Er det korrekt?

A. Ja

B. Nej

Datafortrolighed - hjemsendelse

9. Du har siddet og arbejdet på dit projekt i flere dage og har gemt 20-30 filer med tabeller og andet output. Du er ikke helt sikker på indholdet i alle filer, og du har travlt. Hvad gør du?

- A. Bruger DSTs nye skanningssystem til at filtrere data. De filer der kommer advarsler på tjekker du manuelt.
- B. Du tager den tid det tager at gennemse alle filerne inden du hjemsender.
- C. Andet?

Datafortrolighed - hjemsendelse

10. Det er sidst på dagen og du opdager at du ved en fejl har hjemsendt personhenførbare data. Hvad gør du?

- A. Du gør ingenting og satser på ikke at blive opdaget.
- B. Du kontakter den autorisationsansvarlige på din institution og afventer hvad hun beslutter at der skal gøres.
- C. Du kontakter straks DST og melder dig selv.

Datafortrolighed - hjemsendelse

11. Du har udtrukket nedenstående tabel. Hvad skal du være opmærksom på?

Fødsler i Gladsaxe kommune, efter moderens alder og oprindelse		
Alder	Dansker	Indvandrer
18 år	0	1
19 år	1	1
20 år	1	0
21 år	8	2
22 år	6	1
23 år	12	3

- A. Der er færre end 3 observationer i nogle af cellerne, så det er ikke tilladt at hjemsende tabellen
- B. Det nye sikkerhedsværktøj har skannet tabellen og giver ingen advarsel, så jeg kan trygt hjemsende tabellen.

Datafortrolighed - hjemsendelse

12. Det er muligt at arbejde med geografiske data og trække geografiske tabeller. Hvad skal du være særligt opmærksom på, når du arbejder med geografiske data som sogne eller områder med en bestemt storhed i en kommune?

Datafortrolighed

1: A	5: A	9: B
2: A	6: C	10: C
3: B	7: C	11: A
4: B	8: B	

Brugerforum

Brugerforum

Bruges primært til bred kommunikation/meddelelser fra CRT til brugere, og til feedback og spørgsmål fra brugere til CRT.

Det kunne godt bruges mere til vidensdeling blandt brugere.

Fordelen frem for direkte mails er, at indlæg ikke forsvinder
Indlæg skal dog helst også være relevante for andre

I praksis er der ikke mange indlæg, så man bliver ikke spammet ved at subscribe til alt

- Registrer på: <http://brugerforum.crt.dk/>
- Subscribe til alle kategorier for at modtage mails om nye tråde, kan derefter subscribe den nye tråd efter behov.

Generelt



17 Discussions

32 Comments

Most recent: Subkommunale arbejdsstedsvariabler Started 21 hours ago by Jens

Vidensdeling



0 Discussions

0 Comments

Spørgsmål og fejl



35 Discussions

89 Comments

Most recent: it-investeringer i datainput Started 3 weeks ago by vivi schmidt

Forslag



3 Discussions

5 Comments

Most recent: Ny vækstvariabel defineret ud fra omsætning ? Started 21 august 2015 by vivi schmidt

FAQ



5 Discussions

10 Comments

Most recent: turisme - findes business i samkline Started 11 februar 2015 by vivi schmidt

Latest Discussions

2

3

Subkommunale arbejdsstedsvariabler



Change Language

Danish (Denmark) ▼

Categories

All Categories

Generelt

Vidensdeling

Spørgsmål og fejl

Forslag

FAQ

Search

Go

Popular Tags

peg-og-klik 9

virksomhedsregister 9 ras 3

turisme 3

vækstvirksomheder 2

TAK FOR OPMÆRKSOMHEDEN

Jeg håber alle nu er klar til selv at udtrække tabeller med Peg-og-klik og har fået en generel ide om, hvilke krav der stilles til diskretion.

Har du spørgsmål, kan du altid henvende dig til mig.

Jens Clausen (jens@crt.dk)