
DST SIKKERHED OG ANVENDELSE

VERSION 2.4 – 2021 – KURSUS FOR MEDARBEJDER VED INSTITUT FOR FOLKESUNDHED VED AARHUS UNIVERSITET – UNDER AUTORISATION 50

DST SIKKERHED OG ANVENDELSE

- ▶ Introduktion til brug af DST
- ▶ Ansøgning om adgang til DST
- ▶ **DEN PERSONLIGE ADGANG**
- ▶ Hvordan kommer jeg på?
- ▶ Hvordan hjemtager jeg analyseresultater?
- ▶ **HVAD ER REGLERNE FOR NEDTAGNING AF INFORMATIONER?**
- ▶ Eksempler og værktøjer
- ▶ Kontrol af data før hjemtagning
- ▶ Brud, sanktioner og konsekvenser

INTRODUKTION TIL BRUG AF DST

- ▶ **Danmarks Statistik (DST)** - www.dst.dk
- ▶ Særlig adgang via tjenesten **DSTs Forskermaskiner** - <https://remote.dst.dk>
- ▶ Forskere kan her få adgang til informationer i personfølsomme registre hos DST:
Registre som: Befolkning (BEF), Uddannelse (UDDA), Landspatientregistret (LPR), Sygesikring (SSSY), Død (DOD), Indkomst (FAIK) m.f
- ▶ Forskere kan også få adgang til informationer fra andre registre via DST.
Registre som: Sundhedsdatastyrelsen (LSR, LMDB), Landspatientregisteret (LPR), De praktiserende lægers og regionernes fælles enhed for kvalitetsudvikling (DAK-E), Klinisk Epidemiologisk Afdeling (LABKA), Lægemiddelstatistikregisteret (LSR) m.f.
- ▶ Institut for folkesundhed AU og CIRRAU har selv projektdatabaser – *kan være en genvej!*
- ▶ Adgang til registrene kræver ansøgning og godkendelse

ANSØGNING OM ADGANG

- ▶ Har du ikke gennemført dette kursus, så er der ingen adgang!
- ▶ Institutlederen er ansvarlig for ALLE aftaler indgået med DST
- ▶ Projektbeskrivelse
- ▶ Tilladelser fra mindst Datatilsynet
- ▶ Derudover tilladelser fra evt. eksterne registre (ikke DST data)
- ▶ Ansøgninger til DST:
 - Datatilsynsnummer, Projektbeskrivelse, Protokol m.m.
 - <http://www.dst.dk/da/TilSalg/Forskningservice>
- ▶ Du kan få hjælp til at komme i gang hos Datamanagement Gruppen



Spørg
Marianne Pedersen
mp@ph.au.dk

DEN PERSONLIGE ADGANG

- ▶ **Når du har fået DST-adgang, så er den personlig!**
 - Ingen andre må bruge din adgang
 - Ingen andre må se dig over skulderen
 - Du må ikke vise dine data til andre
- ▶ **Datafortrolighed**

Husk!

Du på kun lave det med data, som du har søgt om tilladelse til!

**Individdata er fortrolige, jf.
Forvaltningslovens § 27, stk. 3
Straffelovens §152.
EU Dataforordning fra 25. maj 2018**

EU DATAFORORDNING

Den nye EU Dataforordning fra 25. maj 2018 (GDPR)

- ▶ Den nye danske persondatalov gælder også på DST
- ▶ Underretning om overtrædelser til DST
- ▶ Underretningspligt om overtrædelser til Datatilsynet
- ▶ Underretningspligt om overtrædelser til de omfattede personer
- ▶ Mulighed for bøder for overtrædelse – ikke fastsat endnu
- ▶ Erstatning til krænkede personer, ca. 1000 kr. pr. person
- ▶ Hertil sanktioner fra DST og evt. politianmeldelse.



Udover GDPR
har DST egne
regler for arbejde
med persondata!

VEJLEDNINGER PÅ DST

- ▶ <https://www.dst.dk/da/TilSalg/Forskningservice/brug-af-forskermaskiner>
- ▶ <https://www.dst.dk/da/TilSalg/Forskningservice/hjemtagelse-af-analyseresultater>

Brug af forskermaskiner

Her findes vejledninger til opsætning, logon og brug af Danmarks Statistiks forskermaskiner:

Danske vejledninger

- ↓ Forbindelse til forskerservere i Danmarks Statistik - Windows brugere
- ↓ Forbindelse til forskerservere i Danmarks Statistik - Mac brugere
- ↓ Forbindelse til Ministerieordningen i Danmarks Statistik - Windows brugere
- ↓ Forbindelse til Ministerieordningen i Danmarks Statistik - Mac brugere
- ↓ Vejledning til skift af password
- ↓ Vejledning til reset af password
- ↓ Nødvendige opdateringer af Microsoft RDP Klient på Windows7
- ↓ Forbindelse til Danmarks Statistiks forskerservere fra udlandet
- ↓ Arbejde på Danmarks Statistiks forskerservere
- ↓ Adgang til Statistikbanken og Danmarks Statistiks hjemmeside fra forskerservere
- ↓ Vejledning til brug af SAS formater i Danmarks Statistik
- ↓ Vejledning til upload af data
- ↓ Vejledning til hurtig genskabelse af filer
- ↓ Hardware og software på Danmarks Statistiks forskermaskiner

Hjemtagelse af analyseresultater

Her finder du relevante vejledninger, til når du skal hente resultater hjem, efter du har arbejdet med mikrodata på forskermaskinen.

Danske vejledninger

- ↓ Hjemsendelse af filer
- ↓ Hjemsendelse af filer - tilladte filtyper
- ↓ Retningslinjer for hjemsendelse af analyseresultater
- ↓ Diskretionering af rækker i aggregerede datasæt
- ↓ Fjernelse af diskretionerede rækker i aggregerede datasæt



HVORDAN KOMMER JEG PÅ DST (WINDOWS)?

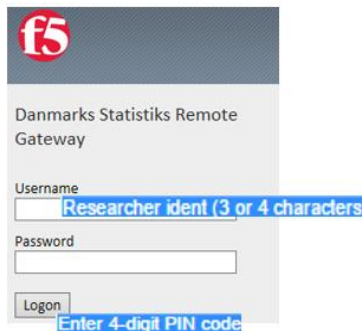
- ▶ Når tilladelserne er på plads, bliver du oprettet som bruger på **DST Forskerservice**
- ▶ Dette giver dig adgang til DSTs Forskermaskiner
 - men kun til dit eget projekt.

Linket til vejledningen er:

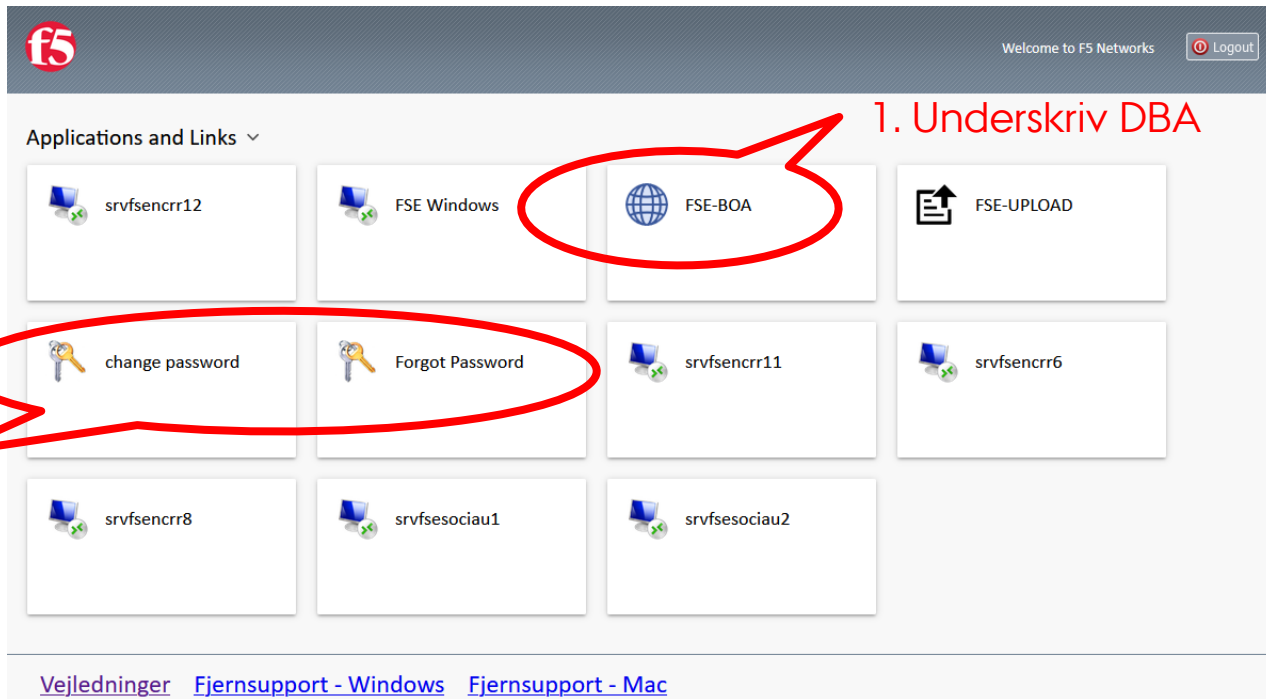
<http://dst.dk/ext/2115449284/0/forskning/Forbindelse-til-forskerservere-i-Danmarks-Statistik-Windows-brugere--pdf>

I browser indtast adressen

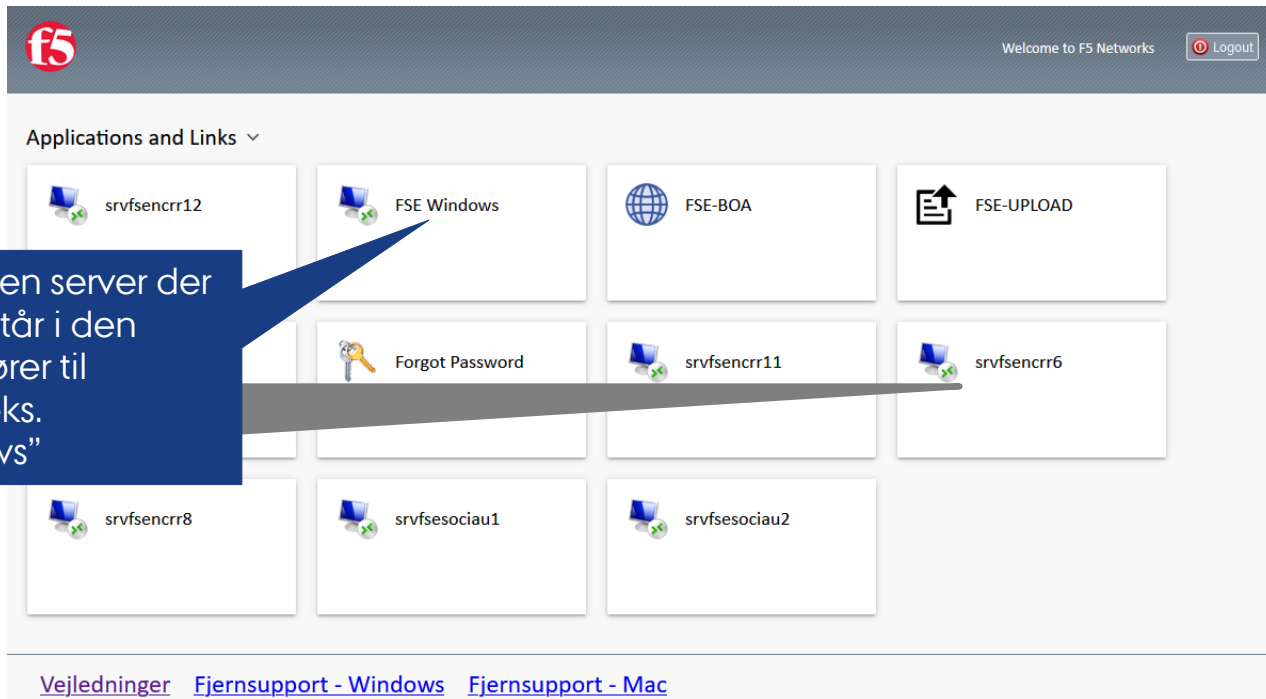
<https://remote.dst.dk>



HVORDAN KOMMER JEG PÅ DST (WINDOWS)?



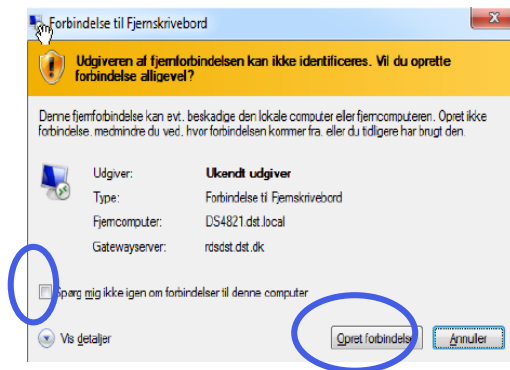
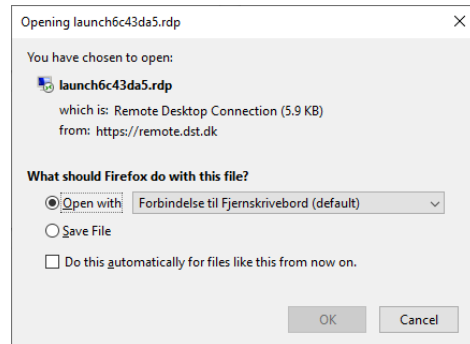
HVORDAN KOMMER JEG PÅ DST (WINDOWS)?



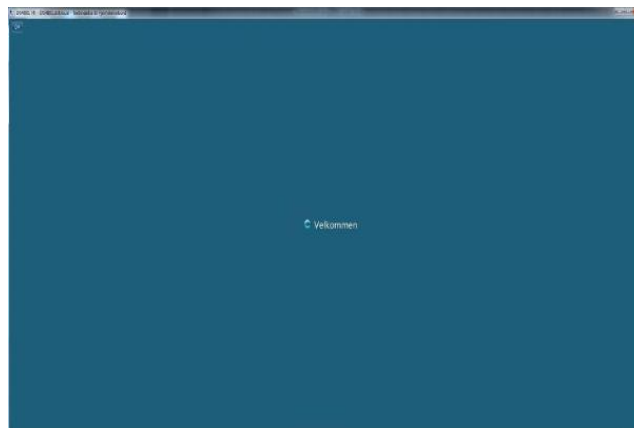
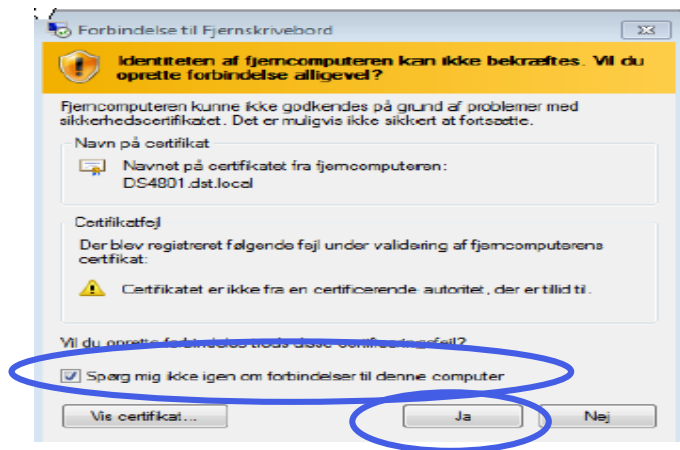
HVORDAN KOMMER JEG PÅ DST (WINDOWS)?

- ▶ Efter denne slide, så er vi næsten halvvejs i login proceduren....
 - dette kaldes **datasikkerhed**!

Windows



HVORDAN KOMMER JEG PÅ DST (WINDOWS)?



HVORDAN KOMMER JEG PÅ DST (WINDOWS)?

- ▶ Når du har afsluttet dit arbejde på serveren, logges der ud på denne ved at klikke på



HVOR ER MINE DATA?

På DST er der to slags data **RAW** og **WORK**:

\\70nnnn\RAW

- ▶ er de rådata man har fået udleveret.
- ▶ De er i SAS-format og er skrivebeskyttet

\\70nnnn\WORK er arbejdsdata.

- ▶ Dette er rådata konverteret fra SAS til e.g. Stata eller R
- ▶ Eller programkode, beregnede data, grafiske fremstillinger osv.
- ▶ Der må/kan kun hjemtages data fra WORK (efter kontrol fra Datamanagement)

HVORDAN HJEMTAGER JEG ANALYSERESULTATER?

- ▶ Hvordan hjemtager jeg **analyseresultater** fra DST?
- ▶ **Det er her alle problemerne kan opstå...**
- ▶ Og **NEJ**, du må ikke "bare" nedtage informationer!
- ▶ Det er derfor, at du arbejder på DSTs forskermaskiner.
- ▶ Det er derfor, at du skulle søge om alle tilladelserne.
- ▶ Det er derfor, at du er personligt ansvarlig.
- ▶ Der er derfor, at Datamanagement skal lave en kontrol



Spørg
Niels-Søren Bøgh
nsb@ph.au.dk

HVAD SIGER REGLERNE?

- ▶ Citeret fra DST "Hjemsendelse-af-filer-fra-forskermaskiner.pdf":
 - ▶ 1. Filerne må ikke indeholde **identificerbare data**, dvs. data der indeholder enkeltrekords, eller hvor virksomheder eller personer direkte eller på anden måde er identificerbare.
 - ▶ 2. Filen må derfor fx ikke indeholde **én variabel fra én observation** fra et rådatasæt, lige meget hvad denne variabel indeholder.
 - ▶ 3. Optællinger, tabeller, output, kørselslogs, programmer og grafik **filer**, må gerne hjemsendes. Alt sammen under forudsætning af, at de ikke indeholder identificerbare data.

HVAD SIGER REGLERNE?

Regel nr. 1. "Identificerbare data"

- ▶ Ingen hele records/rækker, altså **en observation**, må hjemtages fra et rådatasæt.
- ▶ Eller en reduceret record/række med informationer, som kan identificere en person eller et firma direkte eller indirekte.
- ▶ Eksempel: En overtrædelse kunne f.eks. være en information, der siger: "Et firma med en omsæt på over 20 mia. kr. i Bjerringbro". Her er to variable nok, **omsætning og by**. Så ved alle, at det i dette eksempel drejer sig om Grundfos.

HVAD SIGER REGLERNE?

Regel nr. 2. "*én variabel fra én observation*"

- ▶ Der står faktisk; at ikke så meget som én **celleværdi** fra et rådatasæt må hjemtages!
- ▶ Uanset, hvad værdien af cellen er.
- ▶ Dvs. bare ét tal fra ét rådatasæt.
- ▶ **Her skal man virkelig passe på!**
- ▶ Eksempel: I eksemplet med Grundfos, havde vi den ene variables værdi sat til "over 20 mia. kr.". Hvis vi nu havde det **præcise tal** for omsætning fra rådata, f.eks. 22,6 mia. kr., som omsætningen var i 2012, så ville dette ene tal være en overtrædelse af reglerne!

HVAD SIGER REGLERNE?

Regel nr. 3. "Filtyper"

- ▶ Dette lyder som om, at de nævnte filtyper gerne må hjemsendes
- ▶ Men, her skal man også passe på!
- ▶ Tilladte filtyper til hjemtagning
- ▶ Forudsat at de ikke indeholder Overtrædelser af Regel nr. 1 og Regel nr. 2



Filtype	Filekstension
Excel	xls xlsx xslm xlm xml
Grafik	eps png wmf tif jpg gif emf jpeg svg bmp ppt pptx pptm tiff
HTML	htm html mht
Programkode	sas r do doh ado
Logs	Log
SAS	lst sas7bdat
SPSS	sps sav spv
Stata	dta smcl
Tekst	tab txt
Latex	Tex
PDF	pdf
Word	doc docx rtf

HVAD SIGER REGLERNE?

Eksempler på ulovligt indhold i filer:

- ▶ I en do-fil kan der f.eks. stå "***drop if pnr == 1234***", hvilket er en **databasenøgle værdi**
- ▶ **Tællinger** som viser at der indgår mindre end 5 personer
- ▶ Specifikke **datoer og lokationer** for events
- ▶ **Logfiler** fra Stata kan indeholde langt mere information, end man lige tænker over!
– og mere information, end man har set på skærmen!
- ▶ **Plot af en variabel** (signifikant værdier, "haler" med få personer, begge akser! osv)
- ▶ **Signifikant værdi** (en uden for plagemage værdier, indeholder identificerbar data!)
- ▶ **Tekstvariable** med unikke oplysninger

HVAD SIGER REGLERNE?

- ▶ Krav til **særlige variable**:
- ▶ Der må aldrig nedtages pnr-numre og lign.
 - De af identificerbare cpr-numre kaldet pnr, recnum, famid etc.
- ▶ Der må aldrig nedtages nøgle-numre
 - Dvs. databasenøgler (keys), som nøgleværdier eller observations indeks.
- ▶ **Det er med andre ord mere end vanskeligt at hjemtage informationer!**

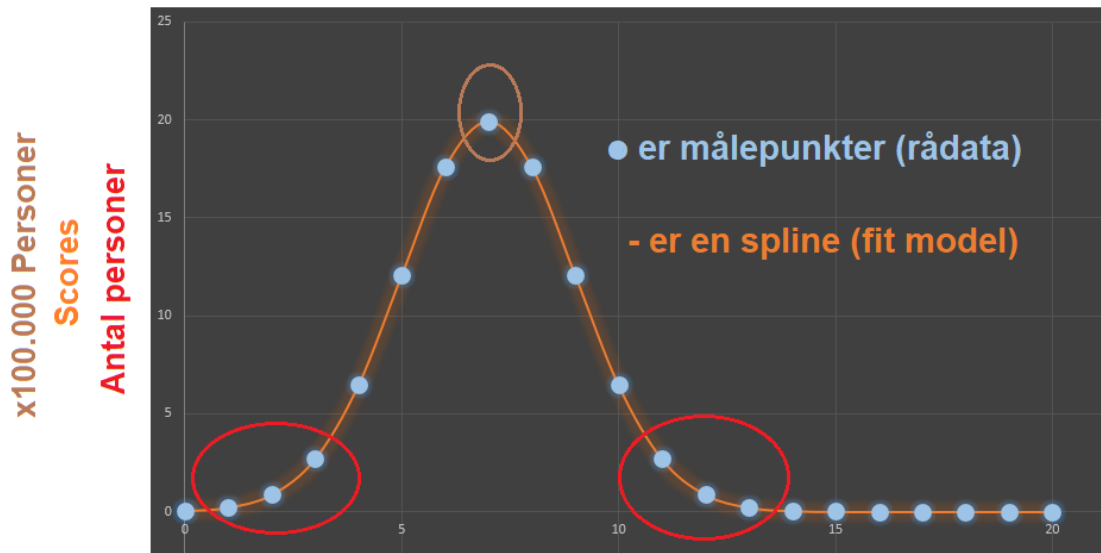
HVAD MÅ MAN SÅ?

Der må hjemtages **aggregeret data**:

- Dvs. at tal i **datatabeller** mindst skal være et gennemsnit af 5 celler i rådata.
- Dvs. at **grafiske** punkter ikke må være identificerbare ned til mindre end gennemsnit af 5 celler i rådata.
- **Datoer** skal angives som måned, ugenummer, relativ tid etc.
- **Do-filer** skal være rensed for data, der er identificerbare ned til mindre end gennemsnit af 5 celler i rådata.
- **Log-filer** skal være rensed for data, der er identificerbare ned til mindre end gennemsnit af 5 celler i rådata.
- Eller at data ikke **på anden måde** kan identificere en person eller en virksomhed.
- **Kaplan-Meier Plots** skal bruge en relativ tid, f.eks. "antal dage/uger sinde start" som x-akse.

HAV STYR PÅ DINE PLOTS!

Vi har en vinder!



Mulige udfald

Individuelle Personer

Scores

Alle værdier kommer
fra 1 person !!!



AARHUS
UNIVERSITET

NIELS-SØREN BØGH
DST SIKKERHED OG ANVENDELSE

DST SIKKERHED OG ANVENDELSE
28. JANUAR 2020

HVAD MÅ MAN SÅ?

- ▶ Må jeg hjemtage **Min, Max, Percentiler og Median** værdier?

Nej, det må du ikke!
(Skriftlig svar fra DST)

- ▶ I særlige tilfælde accepteres et gennemsnit af 3 celleværdier (i stedet for 5), hvis:
Alle 3 værdier hver i sær udgør mindre end 80% af summen af de 3 værdier?

I særlige tilfælde ja – Men, vi har valgt mindre end 5 som udgangspunkt!
- dette skyldes, at vi ellers skal kontrollere ovenstående ibm. hver eneste værdi.

PSEUDO PERCENTILES

- ▶ Niels Henrik Bruun har lavet nogle værktøjer, som ligger på **DST og VDI**
- ▶ Her regnes **min, max, medianer, persentiler...** som et gennemsnit af 5 værdier omkring den oprindelige værdi
- ▶ Ligeledes ligger der værktøjer til at lave base-table, hvor værdier der er mindre end gennemsnittet af 5 automatisk er fjernet.

HVAD MÅ MAN SÅ?

- ▶ Krav til **særlige værdier**:
- ▶ **Missings** [.]
- ▶ Missings, altså manglede værdier, bliver på DST også betragtet som en værdi
- ▶ Der må derfor heller ikke rapporteres et antal missings, mindre end 5
- ▶ Hvis antallet af rækker med missing er <5 skal disse tages ud af undersøgelsen
- ▶ *Tip: Nogle vælger derfor at erstatte få missings med gennemsnittet eller medianen af de øvrige værdier.*

PAS PÅ MED LOG-FILER

- ▶ Hvis du f.eks. bruger STATA, så pas på med kommandoer som:

```
. codebook, . codebook, compact  
. summarize  
. list
```

Disse kommandoer kan skrive **enkelte celleværdier, min og max** ned i logfilen!

- ▶ Stop Log: `. log close`
- ▶ Start Log: `. Log using mylogfile.log`

ANDRE PROBLEMATIKKER

Andre problematikker:

- ▶ Hvis f.eks. din vejleder ikke er godkendt på projektet, så må de ikke kikke med!
- ▶ Lad være med at lave arbejdskopier eller medbringe identificerbare data til møder.
- ▶ Arbejd kun inde i den lukkede verden på DSTs Forskermaskiner.
- ▶ Du må aldrig f.eks. zippe det, du nedtager.
- ▶ Du må aldrig nedtage "store mængder" filer af gangen.

Det, du nedtager, bliver kontrolleret af automatiske scanner – samt stikprøver!

POLITIK OM KONTROL AF HJEMTAGELSER

Inden du hjemsender informationer på IFS

- ▶ Skal du henvende dig til Datamanagement, så vi kan lave en kontrol af dine information!
 - ▶ Derefter kan den projektansvarlige lave selve hjemtagningen
 - ▶ Mange brugere har ikke en download adgang og kan derfor ikke hjemtage
 - ▶ Tag kun den information hjem du vil publicere
 - ▶ Hjemtagning er altid en risiko!
- \\.\Projektmappe\Hjemtagning\2017-09-27\mine_filer

Niels-Søren Bøgh
nsb@ph.au.dk
30360777

http://ph.medarbejdere.au.dk/fileadmin/ph/Interne_politikker/ProcedurerForskerservice_040915.pdf

HVAD MÅ MAN SÅ?



Med andre ord, må du kun nedtage færdigbehandlede og kontrollerede

Aggregerede resultater og dokumentation

- såfremt at de overholder førnævnte krav!

Vi hjemtager aldrig mere end højest nødvendigt!

NOTE

NOTE – RULES on DATA REPATRIATION from VDI

It's NOT allowed to:

- save data to other locations than your CryptShare
- e-mail, upload or transfer data out from the VDI
- copy, cut or paste data from VDI
- print, make screen dumps or take photos of the screen
- make hand copies of data from the screen
- take out any microdata - like rows, cell- or key values

Data repartitions **MUST** be checked and allowed by:

- Niels-Søren Bøgh nsb@ph.au.dk
- Lone Fredslund (BSIG data) lf@ph.au.dk

Data can leave the system in aggregated form **ONLY**:

- at least the average of 5 persons
- no min, max, medias or percentiles
- no significant points in plots
- no micro data in axis in plots
- no missing values below at least 5 missing values per variable

Put data for repartitions in the folder:

..\hjemtagning\<Date (yyyy_MM_dd)>\<your files>

GOD SOFTWARE SKIK

Vi er fælles om at dele DST-Serverens Ressourcer!

- ▶ STATA allokerer meget RAM
- ▶ SAS allokerer meget Disk Plads
- ▶ Starter du programmerne mere end en gang, så bruger du flere licenser!
- ▶ Starter du programmerne mere end en gang, så bruger du flere ressourcer!
- ▶ Husk at lukke programmerne!
- ▶ Midlertidige-filer slettes først når du logger af!

Så kan de andre arbejde bedre, når du ikke er på!

BRUD, SANKTIONER OG KONSEKVENSER

- ▶ **ÉN** forseelse lukker **ALLE** på instituttet ned på ubestemt tid
 - en genåbning kræver et direktionsmøde i DST
- ▶ Derudover får **den enkelte**, der var årsagen:
 - 1. gang: 3 mdr. karantæne
 - 2. gang: 12 mdr. karantæne
 - 3. gang: Permanent udelukkelse via PH.
- ▶ Bevidste handlinger for at tilegne sig data:
 - **Politianmeldelse, jf. Forvaltningslovens § 27, stk. 3 og Straffelovens §152**

TAK FOR JERES DELTAGELSE

Er du i tvivl, så kontakt:

▶ **DST Forskerservice** - 39 17 31 30 eller forskningsservice@dst.dk

▶ Eller få vejledning hos **Datamanagement på IFS**

Projekter Marianne Pedersen, mp@ph.au.dk

Hjemtagninger Niels-Søren Bøgh, nsb@ph.au.dk



AARHUS
UNIVERSITET