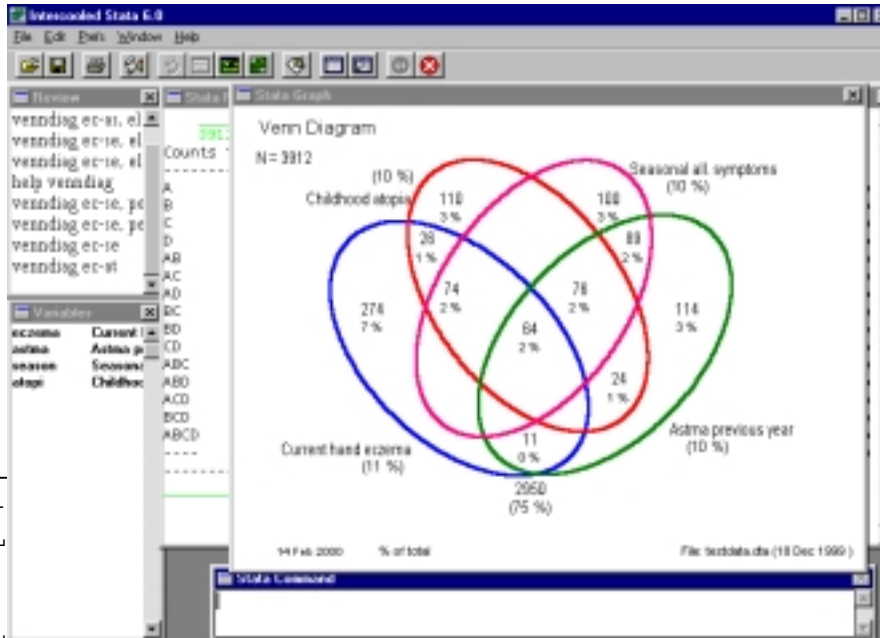


Stata-Introduktion



```

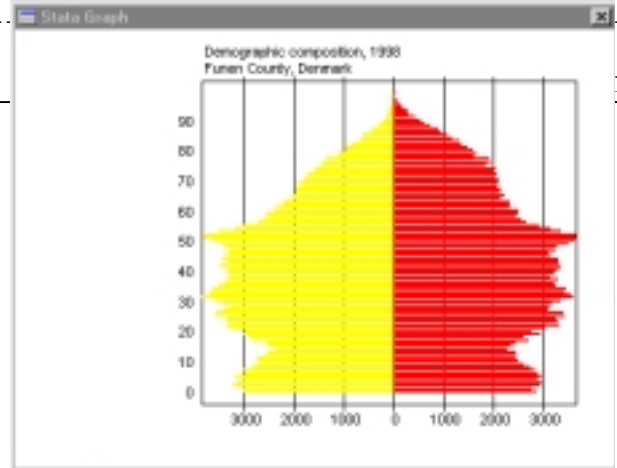
of obs = 240
chi2(1) = 14.89
chi2 = 0.0001
R2 = 0.0453
  
```

test	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
sr5a	.3386243	.0980107	-3.741	0.000	.1920208 .5971564

```

-> xi: logistic test sr5b if sportn == 1 , cluster(hold)
Logit estimates
Log likelihood = -126.28648
Number of obs = 192
Wald chi2(1) = 4.27
Prob > chi2 = 0.0387
Pseudo R2 = 0.0420
(standard errors adjusted for clustering on hold)
  
```

test	Odds Ratio	Robust Std. Err.
------	------------	------------------



Jens M. Lauritsen ©

- Forslag til arbejdsmåde
- Dokumentation
- Oparbejdning og afledte variable
- De mest anvendte analysefunktioner
- Nyttige hjælpeprogrammer
- Stata Technical Bulletin
- Tips til opsætning

• 3. udgave 2000 (13.sept 2000)

Stata - Introduktion .

© Jens M. Lauritsen

3.udgave, 2000 (kontrollér dato på forsiden) (i samme udgave nr rettes kun trykfejl)

ISBN-87-987843-0-7 (trykt udgave)

ISBN-87-987843-1-5 (Elektronisk distribueret udgave)

Den trykte udgave distribueres via Studenterbogladsen ved Syddansk Universitet.

E- Mail: studenter@boghandel.sdu.dk.

Materialet inklusive øvelsesfiler må ikke kopieres, udgives eller anvendes til undervisning uden forudgående aftale med forfatteren. Kopiering skal anmeldes til Copy-Dan. Materialet kan dog anvendes til personligt brug.

Der er udarbejdet et sæt øvelsesfiler til denne note. Kan downloades fra <http://www.bola.suite.dk> Se også side 12.

Forbehold: Da det er et meget stort arbejde at udarbejde noter, holde styr på figurer og undersøge om øvelserne fungerer må læserne bære over med uoverensstemmelser mellem de præsenterede skærbilleder og de billeder der kommer frem ved at udføre øvelserne. Fejl og mangler vil jeg dog gerne have oplysninger om. Den elektroniske form tilpasses efterhånden, mens der kan være – forhåbentlig mindre – uoverensstemmelser med indholdet heri.

Kommentarer til noten sendes til: JM.Lauritsen@dadlnet.dk.

Angiv venligst hvilken udgave af noten, som kommentaren knytter sig til (Se forsiden).

Nogle nyttige internet sider om Stata, biostatistik eller hjælpeprogrammer:

Statistik rutiner og uddybende forklaringer af forskellige analysetyper i Stata:

<http://www.oac.ucla.edu/training/stata/> og <http://www.stata.com/links/resources1.html>

Instruktionsrutiner og forklaring af analyse af forskellige datatyper, inklusive øvelsesdata. Er baseret på EpiInfo, men principperne er generelt gyldige uanset statistikprogram:

<http://www.sjsu.edu/faculty/gerstman/EpiInfo>

Enkle manualer

http://mkn.co.uk/help/extra/people/Brixton_Books

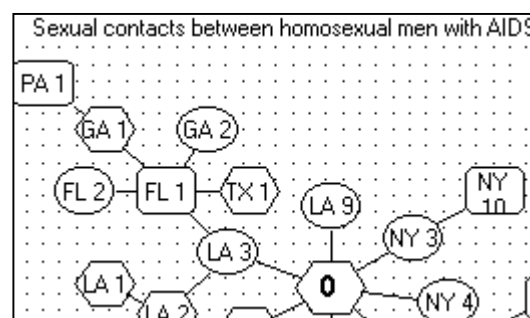
Gratis software til særlige formål:

EpiData til indtastning af data på en nem og pålidelige måde win95/98/NT/2000. Programmet kan downloades fra www.epidata.dk. Programmet skriver datafiler, der direkte kan analyseres i Stata. (+Dbase, Excel og EpiInfo). Value labels m.v. indgår i eksporterede filer til Stata.

Epicalc 2000 (Simple beregninger-nærmest lommeregner) og et simpelt program til at tegne flowcharts og diagrammer (**EpiGram**):




<http://www.myatt.demon.co.uk/>

Uddrag af diagram tegnet med EpiGram. Ved endelige brug af diagrammet kopieres prikkerne ikke.



Stata - introduktion.	6
Forord.....	6
Hvorfor Stata ?.....	6
Noten er skrevet ud fra følgende forudsætninger:	7
Hjælpeprogrammer	7
Manualer, Stata moduler og øvrige indkøb.	7
Yderligere kurser om Stata.....	7
Krav til dokumentation af data og forskning.....	8
A. Mulighed for at finde tilbage til originalmateriale.....	8
B. Kvalitetskontrol, dokumentation og arkivering.	8
Indsamling, oparbejdning og analyse af data.	8
Øvelsesdata – og hjælperutiner	10
Datasæt 1 – Bromaraton.....	10
Datasæt 2 - Euroqol.....	10
Datasæt 3 – Befolkningen i Fyns Amt pr 1.1.1998	11
Datasæt 4 – Simulerede data over forekomst af astma, høfeber og eksem.	11
Datasæt 5 – Data over varighed for brug af hoftebeskytter hos 42 ældre.....	11
Hent øvelsesfiler og læg dem ind på PC'en	12
Øvelse 1 Opret en mappe med navnet c:\data (eller d:\data).....	12
Øvelse 2 Hent øvelsesfilerne på internet	12
Øvelse 3 Pak filerne ud på PC'en.....	12
Øvelse 4 Installér Stata.....	12
Øvelse 5-s Opsæt PC'en på en god måde.....	12
Første start af Stata (Delvis med QUEST hjælpe/menusystemet).....	13
Øvelse 6 Start Stata	13
Øvelse 7 Læs en datafil	13
Øvelse 8 Aftegn en graf for de to variable "alder" og "dectime"	14
Øvelse 9 Start menusystemet quest	14
Øvelse 10 Tutorials – selvstudieprogrammer	15
Stata - hjælp.	16
Øvelse 11 Opslag i hjælpesystemet - 1.....	16
Øvelse 12-s Opslag i hjælpesystemet - 2.....	16
Øvelse 13-s Opslag i hjælpesystemet - 3.....	17
Øvelse 14 Opsætning af Stata skærm.	17
Øvelse 15 Gem kommandoer i en "do" fil	17
Anden start af Stata – kopi af resultater	18
Øvelse 16 Start en logfil ud fra menu systemet.	18
Øvelse 17 Frekvenstabel af køn og graf af løbstid og alder	18
Øvelse 18 Gem kommandoer i en "do" fil	19
Indhold i "do" filer : Opbygning af kommandoer i Stata	19
Øvelse 19 Kopiér grafen fra Stata til tekstbehandling.	20
Øvelse 20 Kopi'er en tabel direkte fra Stata til tekstbehandling.	20
Øvelse 21 Print logfil	21
Øvelse 22 Afslut logfil	21
Øvelse 23 Overfør log fil og graf til tekstbehandling.....	21
Øvelse 24 Installering af de rutiner der indgår i kursus materialet.	22
Øvelse 25-s Opslag PÅ internet direkte fra Stata 6	22
Tredie start af Stata – rettelse i "do" filer.....	23
Øvelse 26 Gentag analysen ud fra filen <i>first.do</i>	23
Øvelse 27 Kontrollér indholdet af first.do i "Do-file" editor	23
Øvelse 28 Udfør kommandoer direkte fra "Do-file" editor.....	23
Øvelse 29 Bestem selv over logfiler.....	24

Øvelse 30-s gentag analysen og få kopi til logfil automatisk – logfil overskrives.....	25
Øvelse 31-s Se på indholdet af first.log og sammenlign med den tidligere logfil.....	25
Øvelse 32-s Gentag analysen og få tilføjet til tidligere logfil	25
Fjerde start af Stata – labels og beregning af nye variable.	26
Øvelse 33 Strategi 1 : Labels ud fra menusystem.....	26
Øvelse 34 Strategi 2. Tildel labels med kommandoer	27
Øvelse 35 Alle labels og kodning af varighed for brug af hoftebeskytter.....	27
Øvelse 36 Materiale beskrivelse for hoftebeskytter.....	27
Øvelse 37 Beregn alder og aldersgrupper.....	28
Øvelse 38 Omsæt tiden fra kryptisk tid til decimaltid	28
Øvelse 39 Ret en værdi for en enkelt person.....	28
Øvelse 40 Afprøv den udregnede decimaltid	29
Øvelse 41 Materiale beskrivelse for bromaraton.....	29
Øvelse 42 Association alder-løbstid.....	29
Andre nyttige muligheder med STATA	30
Øvelse 44 Venn Diagram	31
Øvelse 45 Stata som regnemaskine. Statistik ved indtastning	31
Øvelse 46 Befolkningspyramider	32
Øvelse 47-s Meta analyser.....	32
Tabel fra Stata til tekstbehandling	33
Graf fra Stata til tekstbehandling.....	34
Konvertering af data.....	36
Øvelse 48-s Oversæt data til Stata format.	36
Samlet oparbejdning af Bromaraton data med Stata.....	37
Øvelse 49 Kig på tallene.	37
Opgave 1: Hvor mange personer og omfang af uoplyst i de enkelte variable ?.....	38
Øvelse 50 Start Stata og gør klar til dagens opgave	39
Øvelse 51 Indlæs data fra originalfilen.....	39
Øvelse 52 Gem kommandoer i en ”do” fil	39
Øvelse 53-s Gem kommandoer i en ”do” fil (alternativ strategi)	39
Øvelse 54 Kontrollér indholdet af oparbejd.do i ”Do-file” editor.....	39
Strategi 2 ”Ekstern Editor” Alternativ for erfarne brugere	40
Øvelse 55-s Kontrollér indholdet af oparbejd.do med brug af Win-Commander.....	40
Øvelse 56-s Se på resultatet fra logfilen	40
Øvelse 57 Insufficiente data for enkeltindivider ?.....	42
Beregn afledte variable.....	43
Øvelse 58 Kontrol af variabelen pnr	43
Øvelse 59 Stata som regnemaskine. Afprøv kommandoen ”display”	45
Øvelse 60 Rækkefølge af linier med rekodning !	46
Øvelse 61 Beregn afstand og aftegn en frekvenstabel over grupperet afstand.....	47
Øvelse 62 Beregn alder, aldersgrupper og decimaltid.....	48
Øvelse 63 Dokumentation, anvendte filer og beslutninger samles	50
Bortfaldsanalyse	50
Øvelse 64 Udfør en bortfaldsanalyse.....	50
Endelig besvarelse af hypoteser fra Bromaraton løbet	50
Øvelse 65 Belysning af de opstillede hypoteser	50
Nyhedsgruppe på internet for Stata og www.stata.com.	51
Øvelse 66 Opslag på Stata’s internetadresse (www.Stata.com).....	51
Tilmelding til nyhedsgruppen.....	51
Installering af Stata, Stat/Transfer og Quest.....	52
Øvelse 67 Installér Stata og Stat/Transfer	52
Øvelse 68-s Speciel opstart af Stata (ændring af menuen).	53
Opdatering og tilføjelser til STATA	54

Øvelse 69 Opdatering og tilføjelser til Stata version 6 direkte fra internet.....	54
Øvelse 70 Hent opdateringsfiler til Stata på internet.....	55
Øvelse 71 Hent Quest systemet.....	55
Øvelse 72 Installér den nye Wstata.bin, som er hentet fra internet.....	55
Øvelse 73 Installér ny udgave af ADO systemet, som er hentet fra internet.....	56
Øvelse 74 Installér <i>quest.zip</i> , som er hentet fra internet.....	56
Øvelse 75 Hent STB udvidelser som zip filer og installér fra harddisken.....	57
Tilføjelse af rutiner direkte fra internet.....	57
Øvelse 76 Tilføjelse af rutiner direkte fra internet.....	57
Eksempel på svar fra FAQ listen på www.Stata.com.....	58
What are some of the problems with stepwise regression?	58
Hjælpeprogrammer	59
Installér Hjælpeprogrammer.....	60
Øvelse 77-s Hent win-commander fra internettet:.....	60
Øvelse 78-s Installér Win-Commander	60
Øvelse 79-s Afprøv Win-Commander.....	61
Øvelse 80-s Opsætning af Win-Commander – 1 (grundlæggende).....	61
Øvelse 81-s Hent editor filerne fra internet:.....	62
Øvelse 82-s Installering af “Programmers file editor”	62
Øvelse 83-s Opsætning af Win-Commander – 2 (til brug af Pfe)	62
Øvelse 84-s Afprøv Programmers File Editor - Pfe	62
Øvelse 85-s Opsætning Pfe – 1 (stata filtyper vises).....	62
Øvelse 86-s Opsætning Pfe – 3 (Standard tilpasning).....	63
Øvelse 87-s Opsætning Pfe – 2 (tastatur).....	63
Øvelse 88-sa Opsætning Pfe – 4 (Avanceret tilpasning).....	63
Backup og kopi af data.....	64
Øvelse 89-s Pak en kopi af ”do” og datafiler til et arkiv på harddisken.....	64
Øvelse 90-s Pak en kopi af ”do” og datafiler til en diskette.....	64
Øvelse 91-s Kopiér fra zip filen tilbage til c:\data.....	65
Vejledning for udformning af forsøgsprotokoller og forsøgsrapporter, datadokumentation og opbevaring af data inden for sundhedsvidenskabelig basalforskning.....	66
Vejledning for udformning af undersøgelsesplaner, datadokumentation og opbevaring af data inden for klinisk og klinisk-epidemiologisk forskning.....	67
Euroqol 5-d.....	68
God opsætning af start for Stata og Stat/Transfer	70
Øvelse 92-s God opbygning af startmenuen (tast )	70
Øvelse 93-s Opbygning af STATA mappe på startmenuen (tast )	71
Øvelse 94-s Hjælpeprogrammer i startmenuen. ().....	75
Hypigt anvendte tastaturkombinationer i windows.....	76
General Commands in Stata	77

Stata - introduktion.

Forord

Hensigten med denne note er, at introducere brugen af statistik programmet Stata version 6. Noten giver anvisninger til den konkrete måde at arbejde på og kan **ikke** erstatte håndbøger, manualer og lærebøger i statistik og forskningsmetode. Tidligere udgaver af noten har angivet flere alternative metoder til arbejdet, hvilket i et vist omfang har forvirret "ikke erfarne" edb/statistik brugere. Denne udgave søger at fokusere på de muligheder der findes ved at bruge Stata, som programmet er udviklet. Dermed bliver noten også mere uafhængig af styresystem, idet Stata fungerer stort set ens uanset om du arbejder ud fra Windows, Linux eller Macintosh.

Forslag til hjælpeprogrammer er samlet i et særligt afsnit, herunder hvordan backup og lignende kan foretages.

Hvor meget edb erfaring er nødvendigt for at bruge Stata effektivt? Ikke så meget. Det er nødvendigt at mestre grundlæggende tekstbehandling, at kunne kopiere filer, oprette mapper og være i stand til at hente filer fra internet. Desuden bør du være i stand til at arbejde med pakkede filer, især hvis din PC ikke har adgang til internet.

Filen kan pakkes på forskellige måder, fx såkaldt zip, arj eller uar format. Når en fil "pakkes" sker der det, at unødigt tom plads fjernes. Fx kan en halv tom side i et dokument jo ændres til kommandoen "hop til næste side". Når en sådan fil pakkes ud igen indsættes den halve tomme side blot igen. Størrelsen på pakkede filer varierer fra 10-90 % af originalfilen. Det er nyttigt at arbejde med pakkede filer, hvis man skal sende over internet fordi transmissionstiden så nedsættes, men det forudsætter selvfølgelig at modtager kan pakke ud igen.

Der findes ingen universel **rigtig** måde at arbejde på. Find din egen og sørg for at blive god til at arbejde på den måde. Det er vigtigt, at du arbejder **på en reproducerbar måde**. Minimér brugen af "mus", minimér brug af menu'er og maksimer reproducerbarhed og dokumentation af analyser og arbejdsformer. Du sparer dig selv for megen statisk muskelbelastning ved at bruge tastaturkombinationer hyppigt, se side 76 og 77

Ved nyindkøb af computer foreslås følgende prioritering: 1. En bedre kvalitet af maskinen med langsommere processor fremfor en hurtigere processor på en dårligere generel kvalitet (pris er et godt udgangspunkt for kvalitet, 3 års garanti også udtryk for kvalitet). 2. Mindst 64 Mb Ram. 3. Skærnkvalitet (køb en dyrere 17" skærm, der har flere punkter pr enhed). 4. Et bedre skærnkort der har en hurtigere "refresh rate". 5. Lavest muligt støjniveau. 6. Hurtigere og bedre harddisk, evt. 4 Gb. 6. Optimalt backup udstyr. 7. Bedre stol, bord og lampe. 8. Arbejdstid til at blive god til at bruge systemet, fx kurser. 9. Hurtigere maskine, Mere Ram end 64Mb. osv..

Hvorfor Stata ?

Principielt er det ligeegyldigt hvilket statistikprogram der bruges, hvis blot det kan udføre de analyser der er brug for, på en måde der er til at finde ud af og at programmet kan købes til en rimelig pris. Det kan ikke betale sig at gøre en masse ud af hvorfor det ene program er bedre end det andet. Nogle af begrundelserne er udtryk for vane, smag og behag. For mig er der følgende gode grunde til at anvende Stata som statistikprogram fremfor andre (spss, sas, systat, ...).

- Det understøtter den allernyeste udvikling indenfor biostatistik. Det omhandler stort set alle analysetyper der er relevante i biomedicin. Dvs ét program er tilstrækkeligt.
- Det anvendes af en række gode biostatistikere og epidemiologer, som har skrevet uddybende rutiner på flere områder. (Bl.a. David Clayton, Michael Hills, Hosmer, Lemeshaw)¹ Det er veldokumenteret og der er særdeles fyldige og gode eksempler på de fleste biomedicinske analysetyper i manualerne. Nye rutiner dokumenteres og skrives af avancerede brugere og distribueres gratis. Der er muligt at foretage omfattende matrix beregninger. Det håndterer datoer på en effektiv måde.
- Det arbejder hurtigt og fylder lidt (Kan være på tre disketter). En komplet installation svarende til noten fylder kun 6-10 Mb afhængig af antal ekstra rutiner der anvendes.
- Det er særdeles velegnet til analyse og beskrivelse af follow-up data, som fx overlevelsedata. Både "simple" data med kun ind- og udgangstid og data med gentagne outcome, fx gentagne sygdomsepisoder.
- De fleste regressionsmetoder har indbygget mulighed for at tilføje en klynge (cluster) effekt. Samt anvende "robust methods".
- Når man har købt en licens er denne "uendelig". Dvs. at en typisk bruger har programmet til ejendom. I modsætning til en årlig afgift. Prisen er overkommelig (< 5000 kr) og specielt hvis man arbejder i forskningsmiljø (< 1000 kr). For køb se www.stata.com. Evt. www.sdu.dk/dou eller www.metrika.se.

¹ Clayton D, Hills M. Statistical Models in Epidemiology. Oxford, OUP, 1993. Analyser fra bogen er uddybet i: Clayton D, Hills M. Analysis of follow-up studies with Stata 5.0 STB Reprints Vol 7: 253-268. Clayton D, Hills M. Analysis of case-control and prevalence studies STB Reprints Vol 5: 227—233. Levy & Lemeshow Sampling of Populations: Methods and Applications. Wiley, 1999. Hosmer & Lemeshaw: "Applied Survival Analysis" 1998. Se desuden www.stata.com/bookstore

Der er selvfølgelig også nogle begrænsninger i Stata. Der kunne godt være flere eksakte test, fx i visse non-parametriske test.

Noten er skrevet ud fra følgende forudsætninger:

- Du kender til en metode at få gode data indtastet på. Dvs. data der er indtastet efter en god standard, fx dobbeltindtastede, kontrollerede og rensset for logiske fejl. Se fx noten om datadokumentation og dataindtastning²
- Du har adgang til internet og kan hente filer på internet.
- Du har licens til Stata og stat-transfer. Du har adgang til installations disketter eller CD'rom.
- Detaljer i noten henviser til styresystemet Windows-95/98/NT. men selve Stata delen kan anvendes til andre styresystemer (Mac, Dos, Unix. Linux). Stata fungerer ens under alle styresystemer.
- Du er ikke i tvivl om hvordan du: kopierer filer, opretter mapper (directory) på harddisken, finder en fil på harddisken som umiddelbart ikke synes at være der (dvs gemt et forkert sted). Henter filer på internet og at du har prøvet at arbejde med pakkede filer (zip filer). (Pakke filer ud og pakke filer)

Hjælpeprogrammer

Under oparbejdning af data, statistisk analyse, samt beskrivelse af resultater og udarbejdelse af dokumentation indgår flere programmer. Det er en smagssag hvilke programmer der er bedst. Der er fordele og ulemper ved alle programmer. I ansnittet om hjælpeprogrammer introduceres **Win-Commander** og **Pfe** (Programmernes File Editor) som begge er meget fleksible og betyder at du i vidt omfang selv kan definere arbejdsmåde³. Brugen af disse hjælpeprogrammer er ikke en forudsætning for at arbejde med Stata. Begge skal hentes fra internettet, pakkes ud og installeres.

Konventioner og typografi.


- **Taster:** du skal trykke på vises sådan F4, se side 76. Alt+a er det samme som Alt+A. Hvis du ikke har et windows-95 tastatur med tasten , bruges i stedet Ctrl+Esc. Enter/retur tasten angives: ↵.
- **Filnavne:** angives i kursiv med fed type: ***kursiv.fil***.
- **Øvelsesfiler** hentes fra www.bola.suite.dk (er indlagt som zip fil).
- **Mappe:** anvendes som betegnelse for Folder/Bibliotek/directory/indeks, i eksemplerne arbejdes altid: **c:\data**.
- **Øvelser:** som har tilføjet "-s" er supplerende.
- **Variabel:** Bruges som synonym for **parameter, kovariat ...**
- Kommandoer tastet i Stata vises indrammet. Hvis der står en prik "." først er det output fra Stata

table sex smoke, chi

Manualer, Stata moduler og øvrige indkøb.

Minimum indkøb⁴:

- Intercooled Stata (**ikke** studenter udgaven) , seneste udgave. "*Stata Getting Started*" følger med.
- Stat/Transfer, seneste udgave (med mindre du intaster data med programmet EpiData).

Desuden bør du have adgang til (kan deles med andre):

- Stata Users Guide og Stata reference guide (Er god at have pga særdeles gode eksempler på analyser)
- STB (Stata Technical Bulletin, - review volumes, samt seneste udgivelser)
- Bøger om særlige aspekter. Se gennemgangen i www.Stata.com/bookstore/

Yderligere kurser om Stata

Flere kursussteder afholder kurser om Stata – også internet baserede fjernkurser. Hvor man i en kortere periode, fx. 4-6 uger én gang om ugen får adgang til kursusmateriale og får stillet nogle opgaver. Opgaver og instruktionsmateriale kan så løses på egen arbejdsplads i vante omgivelser uden brug af tid på rejse mm. Omfanget svarer som regel til 2-4 timers arbejde pr uge. Deltagelse giver adgang til at stille spørgsmål til en instruktør. Da der arbejdes på egen arbejdsplads med øvelserne er der et godt grundlag for at fortsætte med at arbejde med egne data. Se <http://www.stata.com/netcourse>. Der findes både grundlæggende introducerende kurser og avancerede kurser i programmering, overlevelsesanalyse og Maximum Likelihood Estimation (avancerede regressions modeller).

² EpiData - datadokumentation og dataindtastning af JM.Lauritsen, seneste udgave. se www.epidata.dk

³ Programmerne Win-Commander og Pfe introduceres her i noten.

⁴ Personer der arbejder med forskning i Danmark kan få oplysning om indkøb og rabatordninger hos Hanne Lillelund på Syddansk Universitet (lillelund@dou.sdu.dk). Pris for det foreslåede er under 1000 kr for en permanent licens til forskning. Meget billigt for en permanent licens til et hovedværktøj til kvantitativ forskningsarbejde.

Krav til dokumentation af data og forskning.

Kravene til dokumentation er skærpet de senere år. I forbindelse med forskning fremgår de af de vejledende retningslinier fra "*Udvalget vedrørende Videnskabelig uredelighed*": Vejledning i god forskningspraksis "http://www.forsk.dk/publ/vejled_vid_praksis/". Retningslinierne er udformet i generelle henstillinger. Fx skrives - kvalitetskontrol - uden at specificere dette nærmere. Se side 66 og 67-. To krav kan nævnes:

A. Mulighed for at finde tilbage til originalmateriale.

Det skal være muligt for enhver oplysning (fx et punkt i en figur) at finde tilbage til originalmaterialet. Det vil sige, at der skal være knyttet et id nummer til alle observationer. Id nummeret følger med i alle udgaver af data og er **entydigt** forbundet med de originale observationer. Originalmaterialet skal opbevares i 10 år. Originalmaterialet omfatter notater (også håndskrevne retteblade mm), spørgeskemaer, analyseskemaer,

Der er en udefineret overgang fra observeret materiale til originalmateriale. Det konkrete valg af skillelinje bør fremgå af dokumentationen. Konsekvensen af dette krav er, at det ikke kan tillades at rette i originaldata, når der er tastet forkeret. Hvertfald ikke uden at gemme en kopi af den først indtastede datafil.

B. Kvalitetskontrol, dokumentation og arkivering.

Samtidig med ovenstående proces dokumenteres antal fundne fejl og de konsekvenser det har haft. Data arkiveres sammen med projektplaner, kopi af anvendte skemaer og den udarbejdede dokumentation. For større projekter med fordel i Dansk Data Arkiv/ERAS.

Ved arkivering i ERAS bevarer den person der har afleveret data den fulde kontrol over hvem der må få data udleveret igen. For data hvor der senere skal foretages opfølgning ud fra cpr nummer kan der afleveres en nøglefil til DDA/ERAS. Dermed slipper brugeren for at have sikkerhedssystemer mv. nødvendige for opbevaring af cpr nummer i perioden fra primær undersøgelse til opfølgning mange år senere. Registertilsynet har godkendt denne ordning.

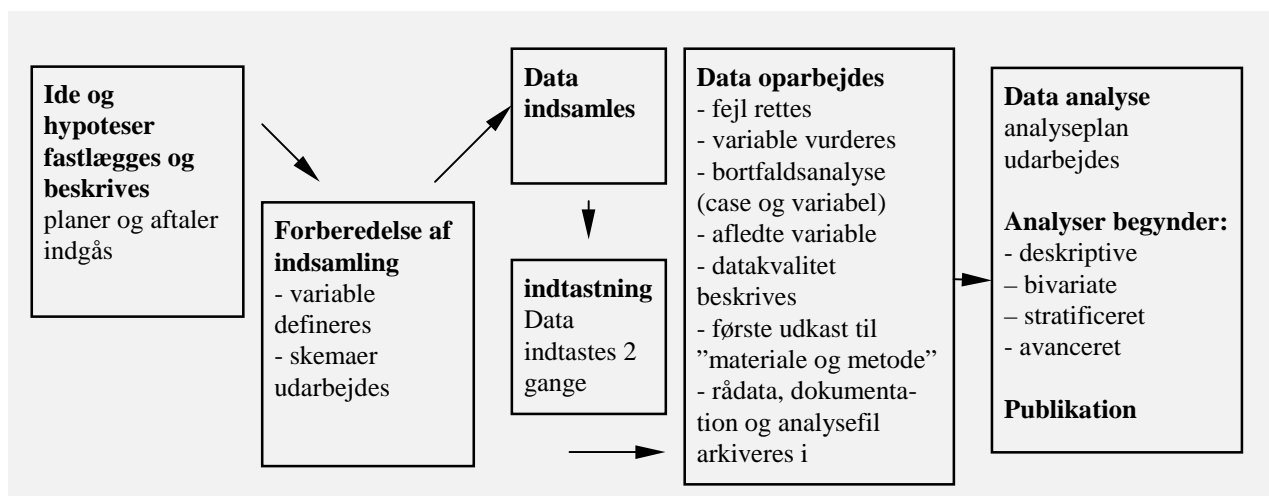
Kontakt ERAS for nærmere oplysning: "<http://www.dda.dk/eras>" eller "mail@dda.dk".

Indsamling, oparbejdning og analyse af data.

Når der indsamles kvantitative data (spørgeskema, kodede interviewspørgsmål ...) er det meget væsentligt, at kvaliteten af de registrerede data er i orden. Der findes en lang række af udarbejdede og afprøvede enkeltspørgsmål og skalaer, som belyser en bestemt problemstilling. Find disse skala'er eller formuleringer og anvend dem. Et af områderne er generelle helbredsstatussmål. Et af disse anvendes i øvelserne (EuroQol 5-d) og er gengivet bagest i noten. I Danmark vil et udmærket udgangspunkt være det standardspørgeskema, som Dansk Institut for Klinisk Epidemiologi (Nu Statens Institut for Folkesundhed "www.dike.dk") har udarbejdet. Dette giver mulighed for at sammenligne med andre undersøgelser.

Ved indtastning af data er der en række muligheder for at sikre kvaliteten. Alt eller udvalgte dele af data bør testes dobbelt. Variable skal forklares og fundne fejl dokumenteres. Et stort arbejde i starten, men erfaringen viser, at indsatsen giver en klarhed og dermed sikrere resultater.

Processen kan opdeles i disse trin:



Bemærk, at dataopbejdning afsluttes med at første udkast til materiale og evt. også metode skrives. Megen tid og præcision af analysearbejdet vil gå tabt hvis der ikke er god forståelse for styrker og svagheder samt logisk konsistens af data inden analysen påbegyndes.

I en afdeling kunne man fx. have den regel at "ingen abstract" eller mødeoplæg holdes før data er arkiverede i en dokumenteret form. Eller at vejledere i phd projekter først diskuterer analyse af data den dag, hvor data er arkiveret med fornødne oplysninger om kvaliteten af data.

En gennemgang af hele processen er beskrevet⁵ og yderligere vejledning om indtastning af data og kvalitetstestkontrol findes i en anden note "EpiData- Datadokumentation og Dataindtastning", som kan hentes fra www.bola.suite.dk i løbet af foråret 2000.

For at illustrere hvad denne dokumentation kunne gå ud på gennemgås i denne note en samlet opbejdning af et datasæt med tilhørende dokumentation. Se Øvelse 50 til Øvelse 66.

Flere publikationer og bøger er relevante⁶

⁵ Hansen JM, Lauritsen JM. Dataindsamling Og Analyse. Ugeskr Læger 1999 (juli)

⁶ McDowell I and C: Measuring Health: A guide to Rating Scales and Questionnaires. Oxford University Press. 2nd ed: 1995/96. Dirksen A, Christensen E, Jørgensen T, Kampmann JP, Kjær P. Klinisk forskningsmetode, en grundbog. København: Munksgaard; 1996:93-110. Jørgensen PH, Kyvik KO. Registrering og arkivering af sundhedsvidenskabelige data. Ugeskr Læger 1997; 159: 963-4. (www.dda.dk/eras)

Øvelsesdata – og hjælperutiner

Disse data skal belyse forskellige problemstillinger og ***må ikke anvendes til publikation***, anden undervisning eller lignende uden særlig aftale. Bl.a er der af pædagogiske hensyn foretaget visse simplificeringer. Alle øvelsesfiler findes i en pakket arkivfil med navnet *kursus.zip* eller *kursus.exe*. Se på internet på adressen www.bola.suite.dk. Kursusfilerne indeholder udover data også nogle ekstra hjælpefiler.

Datasæt 1 – Bromaraton.

Første data er fra det maratonløb som gik over den gamle og nye Lillebæltsbro i 1995.

Indhold	Kodning	Navn på variabel.
id	Dét nummer løberne var tildelt	id
postnummer	For bopæl. For udlændinge er angivet landekode og nr. Fx s-120	pnr
fødselsår	kalenderår, hvor personen er født.	born
køn	F=kvinde M=mand	sex
Løbstid	Timer,minutter og sekunder. Fx 40101 (4 timer, 1 minut og 1 sekund).	runtime

Data findes i filen: *bromraw.txt* (rådata) og desuden som oparbejdet fil, hvor løbstid er omsat til et decimaltal og fødselsåret omsat til alder. Den oparbejdede fil er gemt som *bromar.dta*. Den fil der er anvendt til omformningen er inkluderet som *bromar.do* og *brompost.do*.

Et uddrag af data er gemt som *bromraw.dta*.

Indhold	Kodning	Navn på variabel.
id	Dét nummer løberne var tildelt	id
fødselsår	kalenderår, hvor personen er født.	born
Løbstid	Timer,minutter og sekunder. Fx 40101 (4 timer, 1 minut og 1 sekund).	runtime

Datasæt 2 - Euroqol.

Et af de generelle sundhedsstatusmål hedder Euroqol-5d. Øvelsesdata er indsamlet på et møde, hvor deltagerne udfyldte skemaet ved mødets start og igen tre timer senere. En kopi af skemaet findes bagest i noten.

Indhold	Kodning	Navn på variabel.
	Indtastning af data fra mødets:	start slut
id	Identifikation af besvarelsen.	id id
alder	angivet antal år ved besvarelse	s2 t2
køn	1=mand 2=kvinde	s3 t3
rygning	1: nuværende ryger, 2: forhenværende ryger 3: aldrig ryger	s4 t4
uddannelse	Fortsatte du uddannelse efter folkeskolen 1:ja 2:nej	s7 t7
universitet	Har du en universitetsuddannelse 1:ja 2:nej	s8 t8
euroqol dimensioner	De seks euroqol spørgsmål, se bagest i noten	eq1-eq6 teq1-teq6
euroqol samlet	Aflæsning af visual analog skala (0-100)	eqg teqg

Data findes som *eqfirst.dta* og *eqsecond.dta*. Data fra første udfyldelse er tastet to gange og findes i epiinfo format: *eqfirst.rec* og *eqfirst1.rec*. Fra anden udfyldelse i filen: *eqsecond.rec*.

Datasæt 3 – Befolkningen i Fyns Amt pr 1.1.1998

Filen *fyn.dta* er udtrukket således at den indeholder antal personer i hver alder fra 0-99+ opdelt efter køn. Filen er i et regnearkslignende format med 5 linier og 102 variable..

Indhold	Kodning	Navn på variabel.
tekst	Angiver hvad denne linie indeholder, fx er det for linie 3=mænd, linie 4= kvinder og linie 5= total.	c1
c2,c3 ... c101	I alt 99 variable, som hver indeholder antal personer i aldrene fra 0-99+ år.	c2 c3 c101
c102	Total antal personer i den pgl. linie. I linie 3 er det derfor den samlede mandlige fynske befolkning.	c102

Filen anvendes i forbindelse med

Øvelse 47

Datasæt 4 – Simulerede data over forekomst af astma, høfeber og eksem.

Filen indeholder simulerede data og anvendes til at tegne Venn Diagrammer.

Indhold	Kodning	Navn på variabel.
håndeksem	Tilstedeværelse af håndeksem ja/nej	eczema
astma	Astma ja/nej	astma
høfeber	Høfeber ja/nej	season
atopi	Tegn på atopisk sygdom i barnealderen ja/nej	atopi

Variablene dannes med *venntest.do*. Se Øvelse 45

Datasæt 5 – Data over varighed for brug af hoftebeskytter hos 42 ældre.

Data er indsamlet ved en praktisk afprøvning af hoftebeskyttere i nogle kommuner i Fyns Amt. Indsamlingen skete som forberedelse til et udvidet forsøg med brug af hoftebeskyttere. Om hoftebeskyttere, se J.B.Lauritzen, M.M.Petersen, and B.Lund. Effect of external hip protectors on hip fractures. *Lancet* 341: 11-13, 1993.

Indhold	Kodning	Navn på variabel.
id	Identifikation af besvarelsen.	id
født den	fødsels dato	h1
Udleveret Dato	Dato hvor hoftebeskytteren blev udleveret.	h4
Bruges hoftebeskytter	Angivelse af hvorvidt hoftebeskytteren stadig bruges 1=Nej 2=Ja	h8
Status for brug	0= Ophørt div. begrundelser 1= Ophørt pga vandladningsbesvær ved brug. 4= Død (anvendte hoftebeskytter indtil død) 5= Bruger stadig.	h9
Dato for interview	Den dato hvor oplysningerne er indsamlet	hda

Filen findes som *hipraw.dta*. Men bemærk, at der udover variabelnavn ikke er nogen beskrivelse tilføjet. Filen *hipmake.do* indeholder beregning af datoer mm.

Supplerende rutiner.

Udover en standard installation af Stata anvendes nogle ekstra rutiner. Installering af disse indgår i Øvelse 24. Hvis du ikke kan udføre en bestemt øvelse kan det være at Øvelse 24 ikke er udført. Navnene på ekstra rutiner nævnes i filerne *kursado.pkg* og *kursus.pkg*.

Hent øvelsesfiler og læg dem ind på PC'en

Øvelse 1 Opret en mappe med navnet c:\data (eller d:\data)

Brug det program du plejer at bruge til dette.

Øvelse 2 Hent øvelsesfilerne på internet

Brug en internet browser⁷ og hent filerne fra www.bola.suite.dk⁸

Hent filen "Selv-udpakkende fil til windows" (kursus.exe) og læg den ind i c:\data

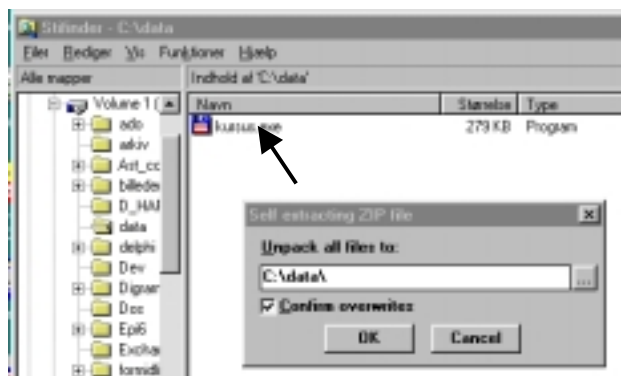
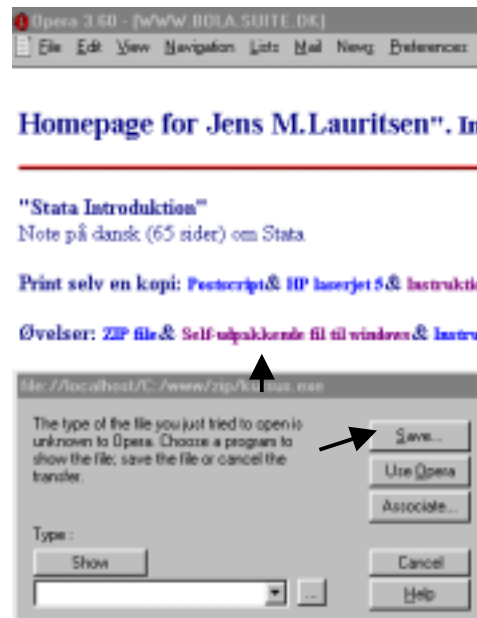
Øvelse 3 Pak filerne ud på PC'en

Du skal eksekvere ("kør" - "run") filen *c:\data\kursus.exe*. På figuren er det vist ud fra Stifinder (dobbelklik med musen ud for *kursus.exe*)

Udover nogle øvelsesfiler, indeholder filen også nogle ekstra rutiner, som er nødvendige for at udføre øvelserne i dette hæfte. Nærmere om dette senere.

Øvelse 4 Installér Stata

Inden du starter Stata første gang skal du installere programmet, se side 69. Du bør, hvis du er nybegynder – installere StataQuest. Husk at angive **c:\data** som det sted, hvor du opbevarer data.



For erfarne edb brugere:

Øvelse 5-s Opsæt PC'en på en god måde

Endelig kan det være en god ide, at opsætte PC'en så det bliver let at arbejde rationelt, se side 70. Bemærk, at der står **-s**. Det angiver at øvelsen er supplerende og især beregnet for erfarne brugere.


⁷ Opera (Opera findes på www.operasoftware.com)

⁸ Udseende af siden kan være ændret fra det viste

Første start af Stata (Delvis med QUEST hjælpe/menusystemet).

Inden du starter Stata første gang skal du udføre Øvelse 1 til Øvelse 4

Øvelse 6 Start Stata

Tryk på  og start Stata fra den menu/gruppe, hvor du har lagt programmet ind. Skærmen ser ud som vist nedenfor. Læg specielt mærke til nederste venstre hjørne af skærmen. Der **skal stå c:\data** ellers passer angivelser i denne note ikke.

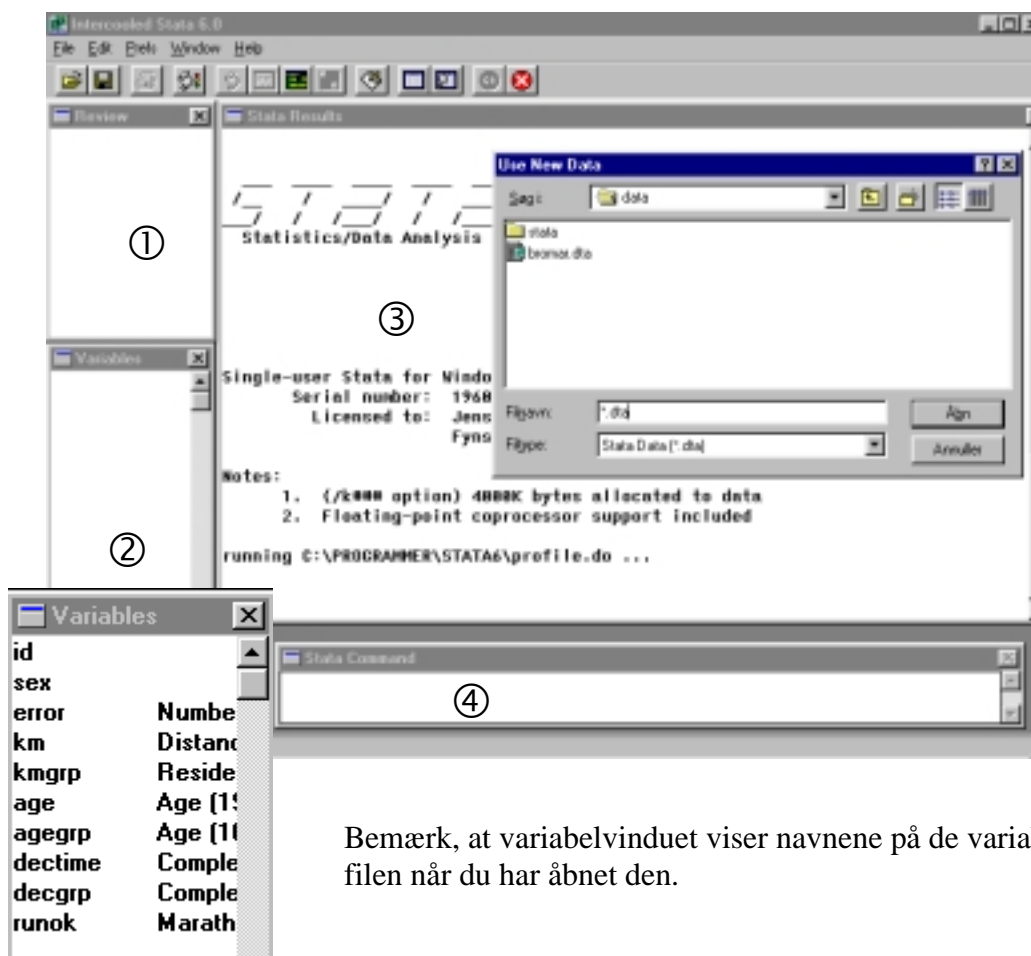
Hvis der ikke står c:\data kan du skrive: `cd \data` i kommandofeltet. Men du bør ændre opsætningen, således at du straks Stata er startet er det rigtige sted. Ellers kan det være vanskeligt at finde filer igen.

Skærmen har fire hovedelementer. Find dem på skærmen:

1. kopi af de kommandoer som er udført (review).
2. oplysninger om variable, når en fil er indlæst (variables).
3. resultater (results)
4. kommandofelt hvor kommandoer tastes (Command).

Øvelse 7 Læs en datafil

Du kan nu åbne en datafil ved at taste: **Alt+F** (øverste venstre menupunkt) og dernæst taste "o" for open og anvise filen bromar.dta (i c:\data). Du kunne også trykke på **Ctrl+o**



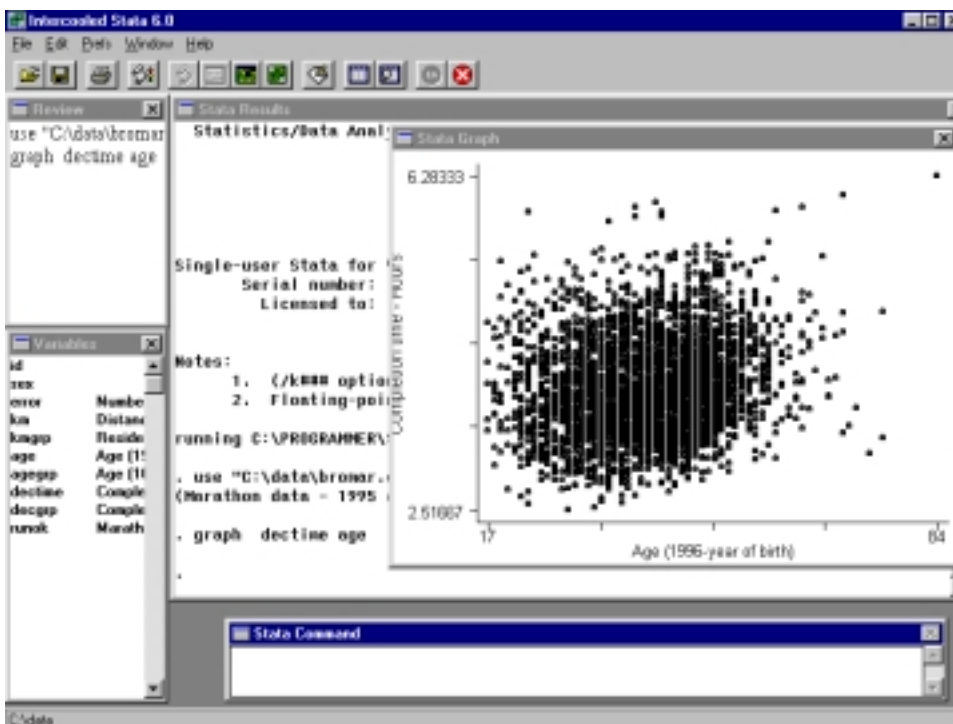
Bemærk, at variabelvinduet viser navnene på de variable, som findes i filen når du har åbnet den.

Øvelse 8 Aftegn en graf for de to variable "alder" og "dectime"

Du skriver blot **graph** i kommando feltet, og dernæst clicker du med musen på **age** og **dectime** i variabel vinduet, således at kommando feltet nu ser sådan ud:

```
graph dectime age
```

Hvis du nu trykker på "Enter" tasten ↵ vil skærmen se sådan ud:



Som et alternativ til selv at skrive kommandoen findes et menu system "Quest".

Øvelse 9 Start menusystemet quest

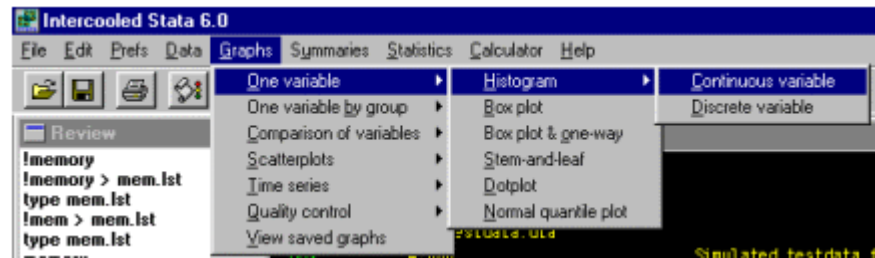
Hvis du har installeret dette se Øvelse 68-punkt 4, samt Øvelse 72 og Øvelse 75 kan du skrive `quest on` og slå menusystemet til, dernæst kan det slås fra med `quest off`: Prøv begge, i resultatvinduet skal der derefter stå således:

```
. quest on
StataQuest menus on. The menubar has been changed.
. quest off
StataQuest menus off. The menubar has been changed.
```

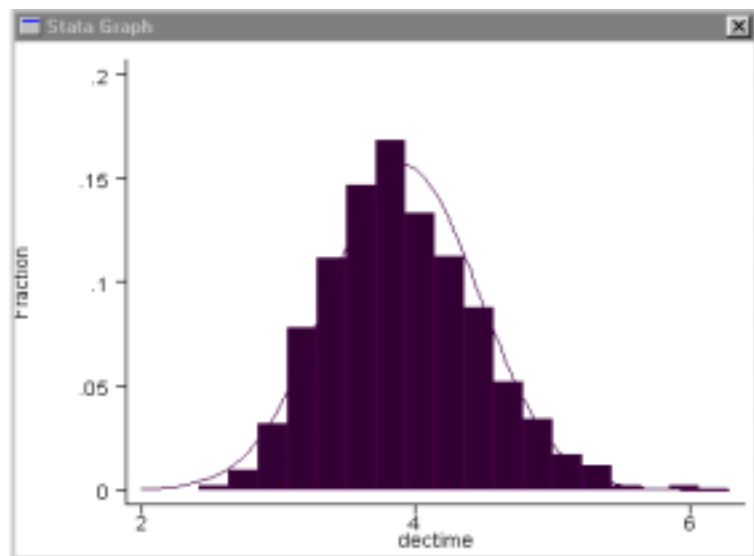
Slå systemet til igen **-quest on-** skærmen **ændrer sig**, og der er nu en anden menu linie øverst:



Prøv dernæst at trykke på **graph**, **One variable**, **Histogram**, **Continuous variable**

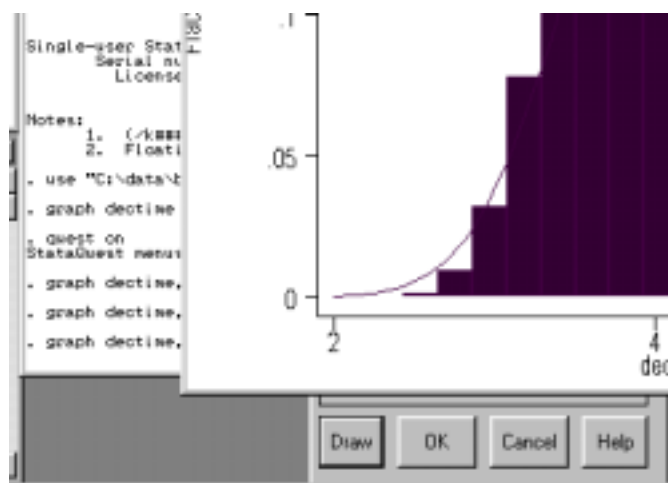


- Angiv variabelen **dectime** og du får vist et histogram, eventuelt med en normalkurve oveni. (bin angiver antal stave. Her er "bin" sat til 15)
- Prøv et par andre grafer også. Fx en afbildning af løbstid i forhold til alder.



Quest systemet kan være en hjælp til at lære nye kommandoer at kende. Bemærk, at der kommer en kopi af kommandoerne i (review vinduet), således at du næste gang blot kan skrive den pågældende kommando.

Når du bruger Quest systemet skal du være opmærksom på, at de dialog bokse, som Quest viser skal afsluttes inden du kan arbejde videre. De kan nemt gemmes bag en graf, og du kan så ikke forstå, hvorfor du ikke kan arbejde videre. Flyt derfor "dialog boksen" udenfor det sted, hvor grafen vises. Se th.



Hvis dit grafvindue var større ville du ikke se dialog boksen. Luk eller flyt grafvinduet og du vil se dialogboksen igen.

Stata har udarbejdet særlige introduktions programmer, Hvor mange "tutorials findes der ?":

Øvelse 10 Tutorials – selvstudieprogrammer

Du får en oversigt over dem således:

```
tutorial intro
tutorial contents
```

Stata - hjælp.

Når du skal have hjælp til at gøre et eller andet kan du starte med at slå op i hjælpesystemet som vist her⁹.

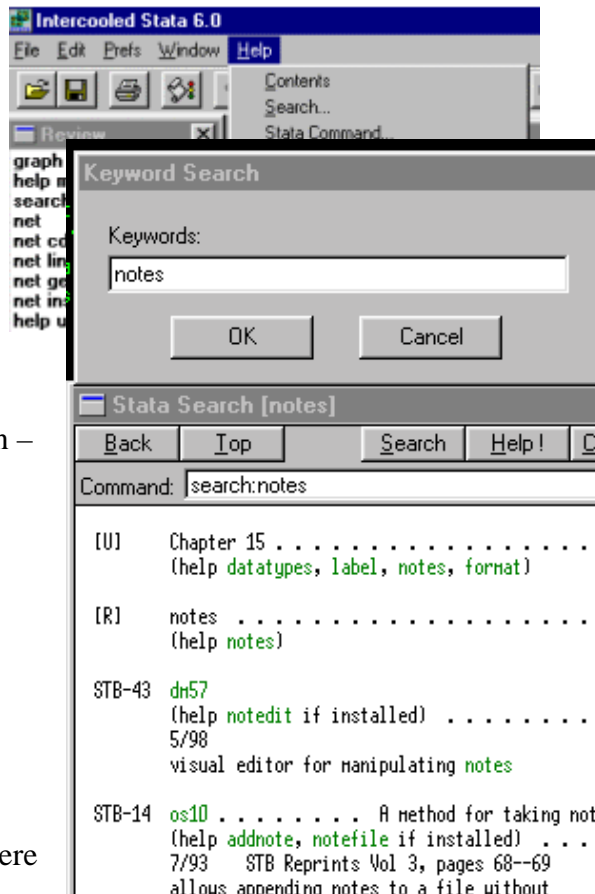
Øvelse 11 Opslag i hjælpesystemet - 1

Find ud af, hvad "notes" er: Tast **Alt+h**, Vælg "**Search**". Der kommer en boks frem. Skriv notes og tast OK.

Der fremkommer nu et nyt vindue på skærmen – hjælpevinduet hvor du kan bladere eller hoppe mellem termer vist i *grøn farve*, når du peger med musen.

Prøv forskellen mellem "Search" og "Stata Command" når du har tastet **Alt+h** eller **F1**

Det er ofte en god ide at kigge nederst i hjælpefiler, der står der som regel nogle helt konkrete eksempler, som man kan arbejde videre med.



Alternativ strategi:

Øvelse 12-s Opslag i hjælpesystemet - 2

Hjælp kan også vises i resultatvinduet
Skriv sådan i kommando vinduet:

```
help notes ↵
search notes ↵
```

Efter at have skrevet linie 1 eller 2 og trykket ↵ kommer indholdet af hjælpen frem i resultatvinduet. Indholdet af teksten er den samme som i øvelsen ovenfor. Du kan stoppe hjælpeteksten ved at trykke på **Ctrl+"Break tasten"** Eller på den røde firkant øverst til højre i menuen. lige under help.

Det er en skønssag om Du foretrækker hjælpeopslag efter Øvelse 11 eller Øvelse 12.

Men som du så, er indholdet af oplysningerne det samme. Efter princippet i Øvelse 11 kan du "hoppe" videre ud fra indholdet i hjælpesystemet. Mens du efter princippet i Øvelse 12 kan afbryde med **Ctrl+Break** .

⁹ Inspiration til at løse et bestemt problem kan du også få ved at læse i *Stata User Guide* og i *Stata Reference Guide* og i denne note. Endelig kan du melde dig til nyhedsgruppen på internet (se side 51) eller se i FAQ listen (Frequently Asked Questions) på www.stata.com. Samt se på relevante internetsider, se side 3.

Øvelse 13-s Opslag i hjælpesystemet - 3

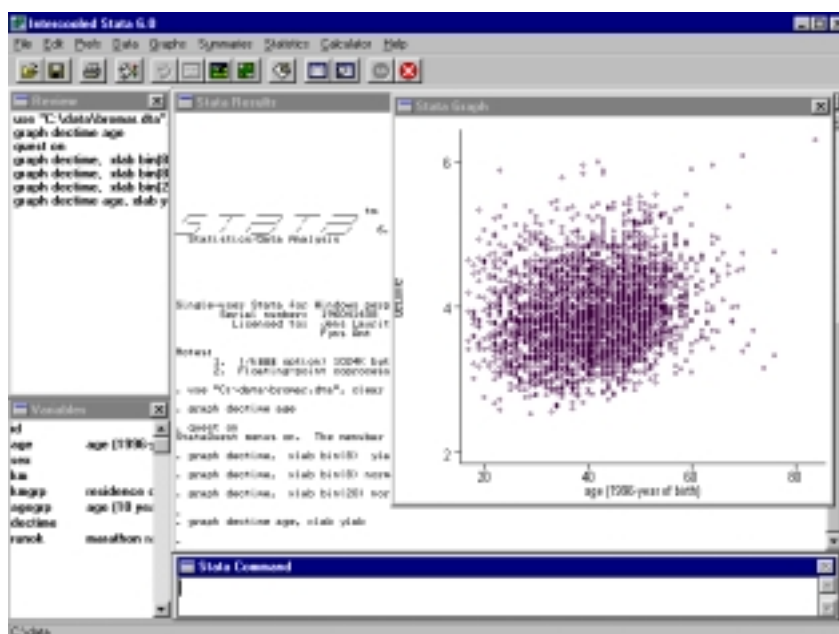
Prøv at finde en forklaring på spørgsmålet: *How do I join paired points on a graph?*

Ind imellem vil du "støde panden mod en mur", fordi der ikke er hjælp at få i de viste systemer. Prøv systematisk at følge rækkefølgen: help-search-faqlist-manualer- "sluk maskinen og gå en tur" – forfra –spørg en kollega - skriv et spørgsmål til support@stata.com. etc.

Øvelse 14 Opsætning af Stata skærm.

Brug musen til at ændre størrelsen på graf vinduet, så du kan se 1/3 af resultatvinduet til venstre for grafen. Ret også på de andre vinduer. Tast derefter **Alt+P** (Pref) og angiv "Save Windowing Preferences".

Prøv derefter at skjule graf vinduet ved at trykke **Ctrl+1** (Results) og få grafen frem igen enten ved **Ctrl+2** eller **Ctrl+W G** Standard opsætning (default windowing) kan gendannes med **Alt+P D**. Brug denne hvis det går helt galt. Baggrund og farver af resultatvinduet kan også ændres under **Alt+P e**



Supplerende om ændring af skrifttype

Du kan også prøve at ændre skrifttype i de forskellige vinduer ved at trykke på de små firkanter øverst i linien for hvert vindue, fx til venstre for "Stata Results". For at få æøå med i resultatfilen kan du bruge "courier"

Med eksemplerne til noten følger to skrifttyper *thinfix.fnt* og *boldfix.fnt*. Du skal flytte dem fra **c:\data** til dér hvor dine øvrige skriftfiler er (typisk: **c:\win32\fonts** eller **c:\windows\fonts**) for at kunne bruge dem. Du kunne fx angive *thinfix* eller *boldfix*, som også har danske bogstaver med, hvilket Stata's standard skrifttyper ikke har.

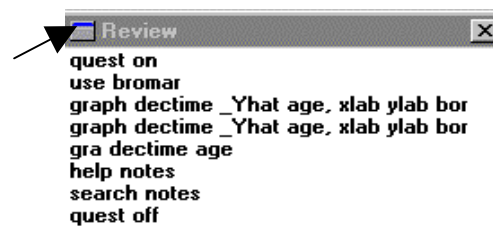
Hvis analysen skal udføres igen vil det være "smart" at slippe for at starte de samme menuliner i quest systemet igen. Dette er også nyttigt hvis du vil vise andre, hvordan du har udført analysen. Derfor skal kommando linierne i "review" vinduet gemmes i en fil til senere brug. Der følger senere en nærmere forklaring om "do" filer. Gør blot som vist i øvelsen.

Øvelse 15 Gem kommandoer i en "do" fil

Filen kaldes "do" fil, fordi den indeholder kommandoer


Peg med musen på den lille firkant til venstre for "Review"

Angiv *quest.do*, som navn og gem i mappen **c:\data**



Anden start af Stata – kopi af resultater

Ved at arbejde som vist i foregående eksempel mangler der en kopi af resultaterne, således at du kan vise dem til kolleger og andre. Stata kan gemme en kopi af både grafer og andre resultater. Resultater gemmes i såkaldte "log" filer, mens grafer gemmes i filer der hver kun indeholder en graf. **Stata starter kun en logfil når du selv beder om det !**. For at få en logfil skal du altså selv starte den.

Tryk på  og start Stata, således at du arbejder i "c:\data". Menu'en ser sådan ud:



Øvelse 16 Start en logfil ud fra menu systemet.

En logfil indeholder en kopi af alt det output, som Stata producerer (undtagen grafer).

Peg på symbol nr 4 fra venstre (ved pilen). Angiv derefter et passende navn, fx "first.log" eller "Øvelse 16-Øvelse 22.log" og tryk på "Åbn/Open"

Når du har gjort dette kommer en hvid firkant med logfilen frem på skærmen. Denne kan du gemme bagved resultatvinduet ved at gå ind i **Window** (**Alt+W**) og vælge **Results** (**R**) eller ved at trykke på **Ctrl+1**. Alt som herefter kommer i **Results** vil også komme i logfilen, bortset fra opslag i hjælpesystemet.



Øvelse 17 Frekvenstabel af køn og graf af løbstid og alder

Skriv nu følgende linier (NB: **små** bogstaver !!) i kommandofeltet én ad gangen og tryk (↵) efter hver linie. Læg mærke til hvordan skærbilledet ændres Du skal **kun** skrive det som står til venstre indtil tegnet ↵: (Hver linie er forklaret i /**/).

<code>use bromar</code>	↵	<code>/*indlæs data fra filen c:\data\bromar.dta */</code>
<code><u>d</u>escribe</code>	↵	<code>/*vis variable i filen */</code>
<code><u>l</u>ist in 1/5</code>	↵	<code>/*list de første 5 personer i filen */</code>
<code><u>t</u>abulate sex</code>	↵	<code>/*optæl en frekvenstabel for variabelen "sex" */</code>
<code><u>g</u>raph dectime age</code>	↵	<code>/*tegn en simpel graf af løbstid og alder */</code>
<code><u>g</u>raph dectime age, t1("Bromaraton 1995") xlab(20,40,60,80)</code>	↵	<code>/*udbygget graf */</code>
<code>gphprint, saving("figur1.wmf",replace) nologo</code>	↵	<code>/*gem grafen*/</code>

Alle kommandoer kan forkortes. Her er det vist med **fed**, fx **d** for describe **gra** for graph. Det er et **komma**, der står efter age!.

Linierne i firkanten ovenfor indeholder alt hvad vi har brug for i analysearbejde:

- 1 læs data (**use**).
- 2 Vis hvilke variable vi har (**describe**) og et udpluk af filen (**list**)
3. Udfør analyse (**tabulate graph**)
- 4 Gem grafen i en fil (**gphprint**)
5. Gem output i en fil. Derfor åbnede du en logfil i Øvelse 16

Med punkt 1 til 4 overholdes kravet fra udvalget vedr. videnskabelig uredelighed: *Data skal kunne følges fra datafiler til endelige tabeller*¹⁰.

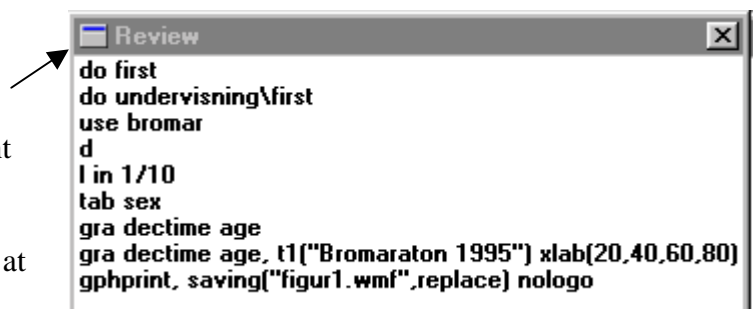
Hvis analysen skal udføres igen vil det være "smart" at slippe for at skrive linierne igen.. Derfor skal kommando linierne i "review" vinduet gemmes i en fil til senere brug.

Øvelse 18 Gem kommandoer i en "do" fil

Filen kaldes "do" fil, fordi den indeholder kommandoer, som kan udføre noget.

Peg med musen på den lille firkant til venstre for "Review"

Angiv *first.do*, som navn og husk at gemme i mappen *c:\data*



```

do first
do undervisning\first
use bromar
d
l in 1/10
tab sex
gra dectime age
gra dectime age, t1("Bromaraton 1995") xlab(20,40,60,80)
gphprint, saving("figur1.wmf",replace) nologo
  
```

Bemærk, at den gemte "do" fil **næsten altid indeholder for mange linier** !. Ideen er at gemme en midlertidig "do" fil. Derefter rettes de forkerte linier og overflødige linier slettes, således at kun de nødvendige linier bevares. Nærmere følger.

Indhold i "do" filer : Opbygning af kommandoer i Stata

Do filerne indeholder kommandoer, som Stata kan udføre. Disse har en bestemt struktur der tilstræbes opbygget ens uanset formål. Bagest i denne note ser du en liste med de mest almindelige kommandoer. De fleste projekter kan analyseres med mindre end 20 forskellige kommandoer (måske en påstand, men prøv). Kommandoerne er bygget op på denne måde:

kommando variable if/in , options

kommando	navnet på den procedure der skal udføres, fx graph
variable	navn på de variable, som kommandoen skal udføres på
if/in	hvis analysen kun skal udføres på et udpluk af data anføres det her Fx: graph x y if sex == "K" /*x og y skal afbildes <u>kun</u> for kvinder*/
,	et komma deler linien før evt. options (kan være vanskelig at se i noten)
options	særlig specifikation af kommandoen. Fx: graph x y, by(sex) /*x og y skal afbildes særskilt for M og K*/

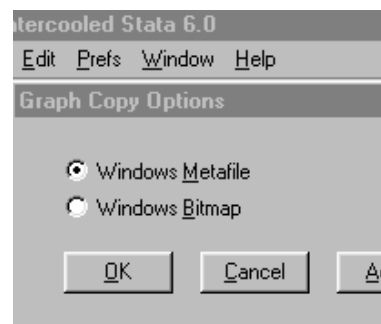
Analyse har kun mening hvis den formidles. Derfor skal resultater i form af grafer og tabeller kunne overføres til tekstbehandling eller diasprogram.

¹⁰ Forudsat, at vi har beskrevet hvordan data blev indsamlet og gemt i filen *c:\data\bromar.dta*

Øvelse 19 Kopiér grafen fra Stata til tekstbehandling.

For at gøre dette skal du nu i Stata producere en graf. Se fx i Øvelse 8.¹¹ Dernæst skal du opsætte Stata til at kopiere grafer på en god måde.

Menupunktet **E**dit har et underpunkt som hedder Graph Copy Options. Sæt punktet til metafile, som vist i figuren. Herefter kan du trykke på **Ctrl+C** og du har kopieret grafen til det såkaldte klippebord, som Windows anvender til at overflytte tekst og billeder mellem programmer.



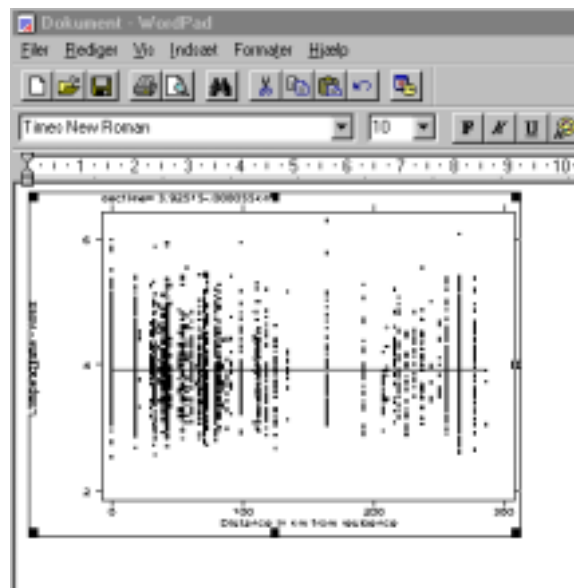
Start nu et tekstbehandlingsprogram, fx. Wordpad (lad Stata være åben) og mens du er i tekstbehandlingen, så tryk på **Ctrl+V**.

Nu får du grafen direkte over fra klippebordet.

Hvis dit tekstbehandlingsprogram kan rette i "wmf" objekter kan du nu rette grafen til.

Dette er for programmet Word97 forklaret nærmere på side 34.

Lad tekstbehandlingen være åben og skift til Stata ved at trykke på **Alt+Tab**



Øvelse 20 Kopi'er en tabel direkte fra Stata til tekstbehandling.

Start quest systemet (**quest on**) og se på en krydstabel af de to variable **sex** og **decgrp**. Du kan også skrive den tilsvarende kommando (**tab sex decgrp**). Tabellen vises i begge tilfælde i resultatvinduet. Gør nu følgende:

1. Aftegn på skærmen tabellen med musen.
2. Tryk **Ctrl+C**
3. Skift til tekstbehandlingen **Alt+Tab** og tryk på **Ctrl+V**.

```
. tab sex decgrp
```

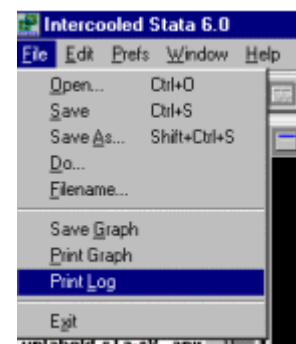
sex	Completion time - Hours				Total
	2.5-3.4	3.5-3.7	3.8-4.1	4.2-6.2	
K	36	76	122	203	437
M	853	850	751	665	3119
Total	889	926	873	868	3556

4. Ret skrifttypen til en "ikke proportional", fx Courier og tabellen ser pæn ud igen. Du har nu fået både tabel og graf over i tekstbehandlingen. Det fungerer udmærket til en enkelt graf og en enkelt tabel. Ved flere forskellige tabeller er det nyttigt at arbejde via logfiler og grafer, som er gemt i filer.

¹¹ Evt. kan du eksperimentere lidt med titler og undertekster i grafen. Fx: b1(undertekst) b2(undertekst2) r(tekst til højre) r2(tekst til højre 2) l1(tekst til venstre) l2(tekst til venstre 2) t1() og t2() tilsvarende. Altså bogstaverne t (top) l (left) r(right) og b(bottom) efterfulgt af tallet 1 (ét) og 2 (to).

Øvelse 21 Print logfil

Logfilen er åben og kan derfor printes, hvis du har en printer tilsluttet. Dette gøres enten ved at aktivere 3' die icon fra venstre eller ved at trykke **alt+f** og aktivere punktet "Print Log"

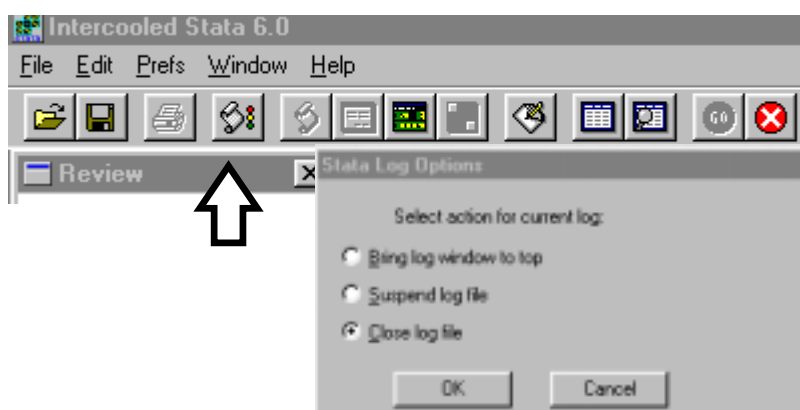


Øvelse 22 Afslut logfil

Logfilen skal nu lukkes, så vi kan afslutte.

Tryk på symbolet for logfil igen:

En ny valgboks kommer frem.



Vælg **C**lose og "OK"

Øvelse 23 Overfør log fil og graf til tekstbehandling

Fortsæt med at arbejde i tekstbehandlingsprogrammet som i foregående øvelse (skift dertil ved at trykke på **Alt+TAB** (Hold Alt tasten nede med én finger og tryk på tabulatortasten én gang med en anden finger. Du får nu en oversigt over de programmer som er aktive for tiden. Tryk på Tabulatortasten med enkeltryk indtil du er ovre i tekstbehandlingen.

Slet grafen samt tabellen fra øvelserne ovenfor og hent logfilen fra

Øvelse 22 (*first.log* eller hvad du nu kaldte den) og indsæt dernæst figurfilen som du gjorde i Øvelse 19. Hvis du bruger et mere kompliceret tekstbehandlingsprogram end Wordpad kan du i stedet indsætte filen *figur1.wmf*. (I de fleste programmer er der en "insert picture from file" mulighed)

Gem dette i et nyt dokument *first.doc* hvori du har tilføjet en kort beskrivelse (få linier) af hvor stort materialet er, hvilken kønsfordelingen og et umiddelbart bud på spørgsmålet: Øges løbstiden med alderen ?

Hermed har du i princippet udført en komplet dataanalyse. Dvs. læst data, udført analyse (tabel + graf), overført dette til tekstbehandling og beskrevet materiale og et analyseresultat.

I Stata har vi nu arbejdet med de fire grundlæggende filtyper:

1. En fil med data **datafilen** *bromar.dta*
2. En fil med resultater i **log filen**, som vi kaldte "**first.log**"
3. Grafer, *figur1.wmf*, gemt med **gphprint**¹² eller overført direkte: **Ctrl+C** **Ctrl+V**
4. En fil med kommandoer i **do filen**, som vi kaldte *first.do*

¹² Du kunne også have gemt filen ved at trykke **alt+f** og dernæst "save graph".

En andet vigtigt aspekt af arbejdet med Stata er at kunne installere og hente tillægsfiler, som andre personer udarbejder og som enten findes som vist i øvelsen side 22 eller hentes via nyhedsgruppen eller STB tillægssystemet, se side Øvelse 70, Øvelse 71 og Øvelse 74.

Øvelse 24 Installering af de rutiner der indgår i kursus materialet.

Udfør kommandoerne som vist i rammen. Øvelserne på side 12 skal være udført først.

*** først fortæller vi hvorfra de ekstra rutiner kommer:**

```
net from c:\data
```

*** dernæst vil vi installere eksempel filerne:**

```
net describe kursus  
net install kursus, replace  
net get kursus, replace
```

*** dernæst vil vi installere de nødvendige tillægsrutiner:**

```
net install kursado, replace  
net get kursado, replace
```

Bemærk, at filer der eksisterer og er uændrede ikke overskrives. Hvis option ”replace” udelades vil programmet ikke overskrive overhovedet, men stoppe.

Udvidet strategi:

Hvis du kender navnet på en rutine, men ikke ved hvor den findes kan du lede efter den på internet (forudsat du er koblet til internettet) enten via hjælpesystemet (**alt+h**), eller via kommandoen **webseek**¹³. Se også internet stederne på side 3.

Øvelse 25-s Opslag PÅ internet direkte fra Stata 6

Hvor mange steder på internet og hvilke steder kan du hente en rutine med navnet **venndiag**. ?

Prøv:

```
webseek venndiag
```

Afslut alle programmer med **Alt+F4** og hold en pause.

¹³ Hvis du får fejlmeddelelse ved at skrive **webseek** i kommando vinduet skal du opdatere din udgave af Stata. Det er gratis.

Tredie start af Stata – rettelse i "do" filer

Problemet med at arbejde som i "Første og delvis anden start af Stata" var, at det er vanskeligt at reproducere præcist hvad der er udført. Derfor introduceres nu en arbejdsmåde ud fra "do" filer ("gør noget filer"). Do filer indeholder kommander, som kan udføres. Do filer har den fordel at du præcist ved hvad du har udført. Går det galt kan du sende do filerne til en kollega eller Stata og spørge om de kan finde fejlen. Hvis der findes fejl i data eller der suppleres med ekstra observationer og analysen skal gentages kan det gøres på meget kort tid (sekunder-få minutter).

Tryk på  og start Stata, således at du arbejder i "c:\data" .

Øvelse 26 Gentag analysen ud fra filen *first.do*

Du gemte i Øvelse 18 en kopi af de kommandoer, der fandtes i "review" vinduet. Nu skal du prøve at udføre analyserne igen ved at få stata til at læse kommandoer fra *first.do*. Du skriver blot:

```
do first /* alle linier i filen first.do udføres én ad gangen */
```

Dette betyder for de fleste, at du i resultatvinduet får en "rød" fejlmeddelelse. Det sker når "do" filen indeholder en kommando med fejl i. Et andet typisk kendetegn er at de samme kommandoer gentages blot med lidt forskellige "options", fx. grafer med forskellige titler. Fejl og gentagelser må derfor rettes i "do" filen "*first.do*". Dette gøres i Stata med en såkaldt "Do file editor".

Øvelse 27 Kontrollér indholdet af *first.do* i "Do-file" editor

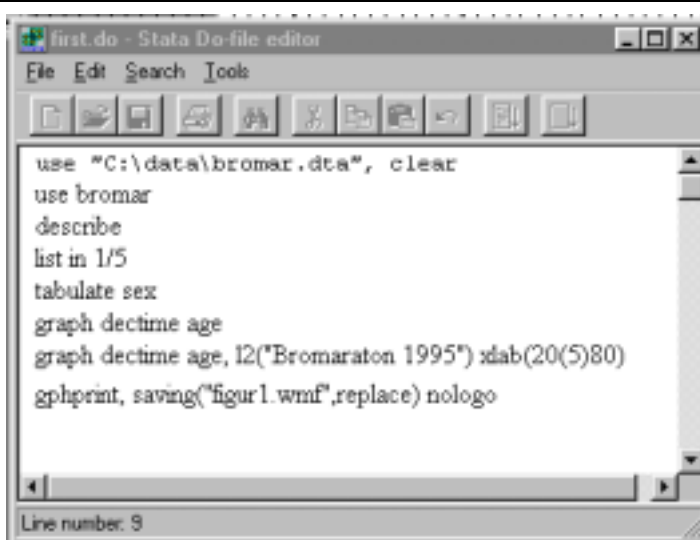
Skriv i kommandofeltet: **doedit first.do** eller alternativt click med musen på den øverste del af Stata Skærbilledet, femte icon fra højre og sørg derefter for at åbne (Ctrl+o) *first.do*

```
. doedit first.do
```

På skærmen åbnes nu en editor som er indbygget i Stata Version 6. Den er simpel. Det vil sige der kun er meget begrænsede muligheder for at tilpasse brugen efter eget ønske.

Til gengæld kan du udføre dele af "do" filer inde fra editoren.

Slet linier med fejl eller ret fejlene.

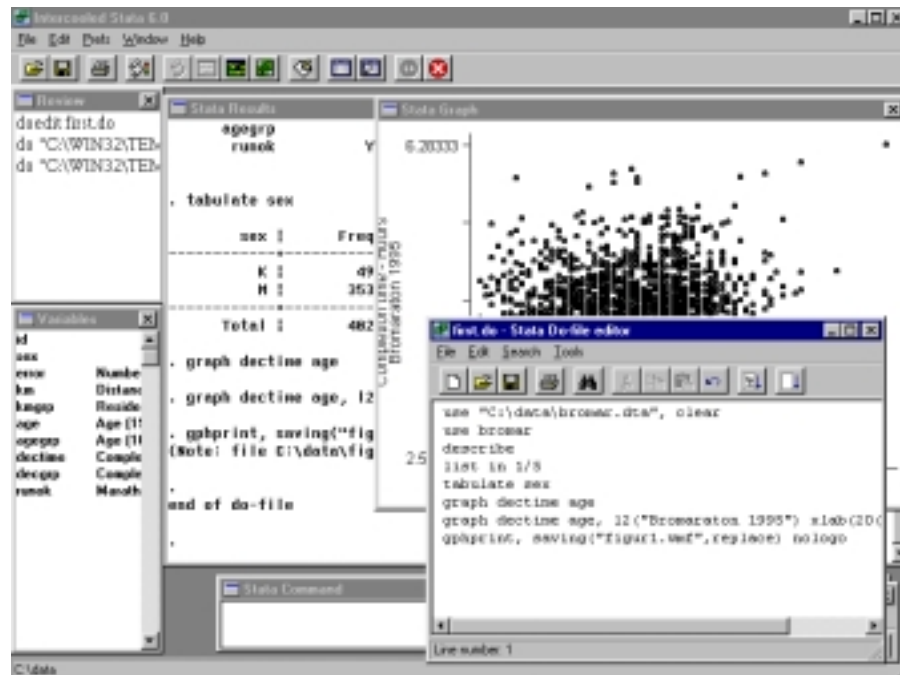


Øvelse 28 Udfør kommandoer direkte fra "Do-file" editor

Du kan nu i do-file editoren ved at markere en blok af din "do-fil" og dernæst med **Alt+T S** få kommandoerne i den pgl. blok udført i stata.

Du vil sikkert undre dig over, at der tilsyneladende ikke sker noget, men det er fordi det sker så hurtigt. Derfor kan det betale sig at opsætte skærmen som vist i figuren nedenfor.

Så kan du hele tiden følge med i hvad der sker i Stata.



Bemærk, at det nu kun er en kryptisk linie "`C:\win.....`" der kommer ind i review vinduet. Ikke de enkelte linier der er udført.

Måske bemærkede du, at kommandoerne blev udført uden at der var en logfil aktiv. Dette kunne ændres ved at åbne en logfil, som vist i Øvelse 16

Øvelse 22. men der er en lettere måde, hvor du får en logfil åbnet og lukket igen:

Øvelse 29 Bestem selv over logfiler

Du kan selv åbne og lukke logfiler ud fra kommandoer: Se i hjælpesystemet om logfiler. Du skriver blot:

```
log using first.log, replace
do first
log close
```

Hvis det giver fejl at åbne en logfil fordi logfilen er åben starter du blot med "log close"

Repetition: De vigtigste elementer i arbejdet med Stata:

1. Data er gemt i datafiler, som kan læses af Stata, fx *bromar.dta*.
2. **log** filer indeholder kopi af output. Disse kan startes, printes og kontrolleres enten ud fra menusystemet eller automatisk.
3. **do** filer indeholder kommandoer der kan udføres. Du har også set hvordan de oprettes, rettes med "do-file-editor" og udføres igen.
4. **graph** filer indeholder grafer der er udarbejdet i analysen, således at de kan overføres til tekstbehandling (eller printes ud direkte).
5. **skift** mellem flere programmer, der arbejder samtidig kan ske med **Alt+Tab**
6. **Opsætning af skærm** og skift mellem de forskellige Stata vinduer
7. Tilføjelse af ekstra rutiner.

Kommandoer har formen: **kommando variable if/in, options**

På sidste side i noten kan du se en liste over de mest anvendte kommandoer i Stata.

Alternativ strategi (for erfarne edb brugere) (Editor)

Alternativ til at bruge "do-file-editor" er at anvende en anden editor. Der findes en række af disse¹⁴. Den simpleste er nok metapad, men det kunne også være pfe, der omtales under hjælpeprogrammer. Se Øvelse 78- Øvelse 84. Metapad kan kun arbejde med én fil ad gangen, men er ekstrem hurtig til at starte. Pfe er knap så hurtig og også gratis. Notetab er mere avanceret og kan arbejde med html kode.

Når du har installeret en ekstern editor kan den tilføjes menustrukturen i Stata ved at tilføje en linie til filen *profile.do*, se Øvelse 69. Her vist for pfe og metapad:

```
window menu append string "Edit" "&F pfe" "winexec \program\pfe\pfe32.exe"
window menu append string "Edit" "&M metapad" "winexec \program\meta\metapad20.exe"
```

Når linien tilføjes *profile.do* ændres menuen hver gang du starter Stata. Andre opsætninger eller standarder du ønsker ved start af Stata kan også tilføjes *profile.do*. Det kunne fx. være større hukommelse. "set memory 8000" vil øge arbejdshukommelsen til 8Mb. se Øvelse 69:

Desuden kan du kombinere Win-Commander og pfe eller metapad, se Øvelse 84.

Alternativ strategi (for erfarne edb brugere) (log filer)

Øvelse 30-s gentag analysen og få kopi til logfil automatisk – logfil overskrives.

Udfør analysen igen ved at starte *first.do* med kommandoen **now** Du skriver blot:

```
now first ↵ /* filen first.do udføres og resultatet findes i first.log*/
```

Det **now** gør er først at åbne en ny logfil med samme navn som do filen. Dernæst tilføjes en linie med navnet på do filen, samt dato og tidspunkt. Derefter udføres kommandoerne i do filen og til sidst lukkes log filen igen. Logfilen kan derefter undersøges. **Brug af "now" forudsætter at "do" filen læser datafilen forfra (typisk use xxxx.dta).**

Øvelse 31-s Se på indholdet af first.log og sammenlign med den tidligere logfil.

Brug tekstbehandlingsprogram eller hjælpeprogram til at se dette. Du vil se, at der er tilføjet en linie i starten. Brug pfe, notepad, metapad¹⁴ eller wordpad til dette. Gerne i kombination med Win-Commander.

```
Analysis do-file first run on 2 Jul 1999 at 13:53:17
```

Øvelse 32-s Gentag analysen og få tilføjet til tidligere logfil

Udfør analysen ved at starte *first.do* med kommandoen **now**. Du skriver blot:

```
now first a ↵ /* filen first.do udføres og resultatet tilføjes first.log*/
```

¹⁴ pfe (www.lancs.ac.uk/people/cpaap/pfe), metapad (<http://welcome.to/metapad>), notetab (www.notetab.com). Disse editorer (små tekstbehandlingsprogrammer) starter med at være små hurtige programmer, som kun kan det mest nødvendige, men udviklerne ender med at falde for fristelsen til at inkludere flere muligheder.

Fjerde start af Stata – labels og beregning af nye variable.

En væsentlig del af al oparbejdning af data og analyse af data er at bringe oplysningerne på en form, der passer til den analyse, som skal udføres. Fx kan alder beregnes på en bestemt dato, aldre opdeles i 10 års grupper og datoer omsættes til varighed.

Tildeling af labels

Det er god praksis at variable tildeles titler og beskrivelser, sådan at tabeller altid læses med de rigtige betegnelser. Se på de to tabeller: Frekvenstabel for variabelen v1 som er tildelt en variabel label "sex" og en value label (1=male and 2=female). Tabellen med og uden labels:

uden labels				efter labels			
v1	Freq	Percent	Cum.	sex	Freq	Percent	Cum.
1	70	43.8%	43.8%	male	70	43.8%	43.8%
2	90	56.3%	100.0%	female	90	56.3%	100.0%
Total	160	100.0%		Total	160	100.0%	

Hoftebeskyttere:

Data i øvelsessæt nr fem er indsamlet i forbindelse med et pilotprojekt om brug af hoftebeskyttere i Fyns Amt. Ideen med hoftebeskyttere er, at ældre som har skrøbelige knogler kan tåle at falde uden at få hoftenære brud hvis noget af faldenergien optages af et stødabsorberende materiale. Hoftebeskytterne er i et studie vist at forebygge hoftebrud¹⁵. Efter indtastning af data to gange og kontrol af fejltastning er filen oversat til Stata. Der skal nu tilføjes labels. Datastrukturen er vist på side 11.

Øvelse 33 Strategi 1 : Labels ud fra menusystem

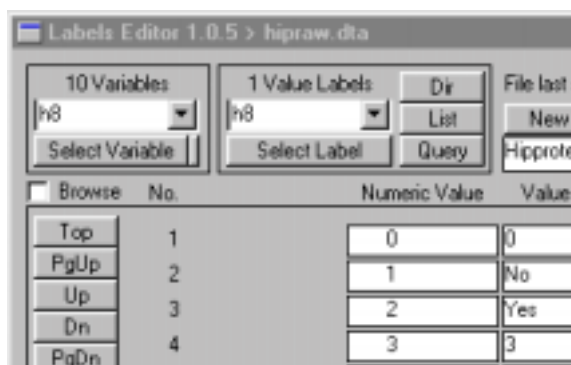
Tildeling af labels kan gøres med en kommando, der hedder **labedit**. (indgår i øvelsesmaterialet, se Øvelse 24)

Rettelser/definitioner **skal aktiveres** (der savnes en "apply" tast) og dette gøres med en af de seks "scroll" taster til venstre nederst (Top, PgUp etc). **Prøv dig lidt frem.**

Husk at gemme filen efterfølgende, således at de samlede kommandoer bliver:

```
use hipraw, clear
labedit
save hipx
```

Skærmen ser ud på to måder afhængig om du har "select variable" eller "select label". For at tildele labels skriver man ud for de forskellige variable. Prøv fx. at starte med at tildele variabel labels til alle variable. Definér derefter kodeværdi labels (label define) og tildel disse til variablene.



¹⁵ J.B.Lauritzen, M.M.Petersen, and B.Lund. Effect of external hip protectors on hip fractures. *Lancet* 341: 11-13, 1993.

Øvelse 34 Strategi 2. Tildel labels med kommandoer

Alternativ til labedit kan du i stedet oprette en "do" fil, der definerer labels. Filen nedenfor viser kommandoerne for **h9**. Se side 11 for de øvrige variable. Tilføj både variable labels (label var), kodeværdi labels (label value) og datasæt label (label data), samt notes. Gem kommandoerne i en do fil **hip.do**. Hver kommando kan forkortes til det understregede.

```
use hipraw, clear
describe
label var h9
label define h9 0 "Stopped-Variouse reasons" 1 "Stopped-continance"
label define h9 "Used until Death" 5 "Still Using", add
label value h9 h9
label data "Hip Protector Data Funen County"
note: Data hipprotector field test
note h9: Defined at interview by nurse
note
tab1 h9-h10
save hip, replace
label list
```

Bemærk:

- at kodeværdier tildeles i to omgange. Først defineres en kodeværdilabel med navnet h9 (label define) og dernæst tildeles h9 til en eller flere variabler (label value). Navnet h9 vælger du selv. Du kunne også kalde den pgl. label for "olsen" "eller"s".
- `label define h9 0 "Stopped-Variouse reasons" 1 "Stopped-continance"`
`label define h9 "Used until Dead" 5 "Still Using", add`
 Kan i stedet skrives: (alt mellem /* og */ i en do fil opfattes som en kommentar.):
`label define h9 0 "Stopped-Variouse reasons" 1 "Stopped-continance" /*
 */ 4 "Used until Dead" 5 "Still Using"`
- notes bruges til yderligere dokumentation. Her kunne også noteres oplysninger om fejl. Fx. note v1: Check kode 99 for id==12
- tabulering af flere variable ad gangen gøres let ved at skrive **tab1** (ét, ikke l).

Øvelse 35 Alle labels og kodning af varighed for brug af hoftebeskytter.

Se på indholdet af **hipmake.do**, som indgår i øvelsesfilerne. I **hipmake.do** tildeles labels, desuden beregnes varighed af hoftebeskytter brug ud fra datoer, tilstand (død) mm. Når du har forstået alle linierne, kan du udføre filen og se på fordelingen af hoftebeskytter anvendelse.

```
do hipmake
```

Øvelse 36 Materiale beskrivelse for hoftebeskytter.

Beskriv nu materialets sammensætning på en relevant og publiceringsorienteret måde. Du får bl.a. brug for kommandoerne: **table**, **ltable**, **egen** eller **generate**.

1. Angiv antal ældre og et relevant mål for alder i de fire grupper ud fra variabelen **h9**.
2. Hvilken proportion af de ældre har ønsket at afprøve hoftebeskytterne overhovedet ?
3. Hvordan er den tidsmæssige udvikling i brugen af hoftebeskyttere (compliance) vist som graf med relevante akser ?. Grafen skal kunne anvendes på en engelsksproget poster.

4. Beregn en variabel som angiver dato for ophør af brugen af hoftebeskytter eller censurering ved død eller observationsperiodens udløb. Samt en variabel der angiver alder ved udlevering af hoftebeskytteren.

Beregning af nye variable.

Der findes mange funktioner i Stata til dette. Se under **"help functions"** **"help cut1"**.

I øvelsen anvendes bromaration data, filen **bromraw.dta**. I øvelsen nævnes ikke noget om logfiler, det må du selv sørge for.

Øvelse 37 Beregn alder og aldersgrupper

Gentag dette og sammenlign agegrp med agegrp2, hvad er forskellen ?. Der står cut1 (ét tal)

```
use bromraw, clear
generate age = 96-born
egen agegrp = cut1(age) , at(10(10)100) interval
egen agegrp2 = cut1(age), at(10(10)100) label
* bemærk, at der er to == efter age i næste linie (se "help operators")
gen agemis = 1 if age == .
recode agemis . = 0
codebook age agegrp agegrp2 agemis
tabulate agegrp agegrp2
```

Gem linierne fra øvelse 33 i en do fil med navnet **brom1.do**.

Øvelse 38 Omsæt tiden fra kryptisk tid til decimaltid.

Hvorfor ser grafen underlig ud, hvis du aftegner alder i forhold til runtime ?

```
gra runtime age, t2(Underlig graf, hvorfor ?) r2(hvornår er klokken 6.78 ?!)
* tiden skal ændres. Undersøg med "help functions" om du kan gennemskue dette.
gen time=int(runtime/10000)
gen minute=int((runtime-time*10000)/100)
list id-runtime time minute in 1/10
gen dectime = time + minute/60
egen decgrp = cut1(dectime), group(4) interval
drop runtime born time minut
*endelig beregnes en variabel der angiver gennemførelse af løbet
gen runok = cond(dectime != . , 1,0)
save bromshort, replace
```

Slå op i hjælpesystemet under **"Help operators"** for at forstå **"!="** i næstsidste linie. Gem linierne fra øvelse 34 i en do fil med navnet **brom2.do**.

Øvelse 39 Ret en værdi for en enkelt person.

Det viser sig nu, at der er sket en fejl i registreringen af data. Du får en meddelelse om at løbstiden for id = 2374 er "35557". Derfor skal beregningen i øvelse 33-34 ændres. Du skal **starte** dofile editoren og tilføje følgende:

```
use bromraw.dta, clear
replace runtime = 35557 if id == 2374
* herefter tilføjer du brom1.do og brom2.do (se under files "insert file") og retter til
```

Når du har rettet alle linierne og udført filen igen har du gemt en ny bromshort.dta med den rigtige værdi. Det noteres i dokumentationen, at fejlen for det pgl. id var fundet og rettet.

Øvelse 40 Afprøv den udregnede decimaltid.

Er de nye variable gemt i filen *bromshort.dta* ? Prøv at se efter:

```
use bromshort, clear
graph dectime age, b2(Bemærk dine nye variable) xlab ylab
```

Du vil se, at filen nu indeholder både de nye variable og de tidligere.

Øvelse 41 Tabeller med og uden labels. Med og uden missing.

Hvilken forskel i output fremkommer med option "nolabel" og "missing" når du udskriver en krydstabel med "**tabulate**". (sammenlignet med at undlade dem).

```
tabulate agegrp decgrp, gamma chi nolabel
tabulate agegrp decgrp, gamma chi missing
```

Øvelse 42 Materiale beskrivelse for bromaraton.

Besvar nu materialets sammensætning mht alder, køn og gennemførelse af løbet. Bestem om du vil bruge median eller middelværdi, percentiler eller andet. (Kommandoerne **table** og **graph**, samt **tabulate** og/eller **summarize** er nødvendige). Variablen køn findes i filen *bromar.dta* og kan tilføjes med kommandoen **mmerge** eller **merge**. Alternativt kan du vælge at bruge filen *bromar.dta*.

Øvelse 43 Association alder-løbstid.

Da grupperet alder og løbstid er ordinale variable kan associationen mellem dem beskrives ud fra gamma koefficienten. Udfør kommandoerne vist nedenfor og besvar ud spørgsmålene:

1. Hvad indeholder output fra rutinen "**partgam**"?
2. Angiver "**partgam**" samme gamma koefficient som der kommer fra "**tabulate**"?
3. Angiv gammakoefficienten inkl. konfidensinterval og p-værdi for associationen alder og løbstid.
4. Vil du i en publikation angive den partielle gamma koefficient eller den ujusterede gammakoefficient ?
5. Er relationen mellem alder og løbstid uafhængig af køn (ens blandt mænd og kvinder)?

```
partgam agegrp decgrp, by(sex)
partgam agegrp decgrp, by(sex) test
partgam agegrp decgrp, by(sex) subpct
partgam agegrp decgrp, by(sex) table
```

Andre nyttige muligheder med STATA

Datasæt er nærmere beskrevet under øvelsesdata på side 10.

Øvelse 44 Sammenkobling af filer, der indeholder dele af informationen i et projekt.

1. Brug euroqol data fra øvelsesdatasæt nr 2. *eqfirst.dta eqsecond.dta*
2. Læs i hjælpesystemet og Users Guide om kobling af filer. (**help merge**
help mmerge)
3. Aftegn et scatterplot for helbredstilstanden for variablene **egg** målt ved første og anden dataindsamling.

Er der nogen personer, som gik mellem første og anden dataindsamling, eller først kom efter starten ?

```
. use eqfirst
. mmerge id using eqsecond
```

matching type	auto
mv's on match vars	none
unmatched obs from	both

```
-----
```

master	file	eqfirst.dta
	obs	48
	vars	13
match vars	id	(key)

```
-----
```

using	file	eqsecond.dta
	obs	47
	vars	13
match vars	id	(key)

```
-----
```

common variables	none
------------------	------

```
-----
```

result	file	eqfirst.dta
	obs	50
	vars	26 (including _merge)

```
-----
```

_merge	code	freq
only in master data	1	3
only in using data	2	2
both in master and using data	3	45
Total		50

```
-----
```

```
. which mmerge
c:\ado\stbplus\m\mmerge.ado
*! 2.3.3 10jan2000 Jeroen Weesie/ICS

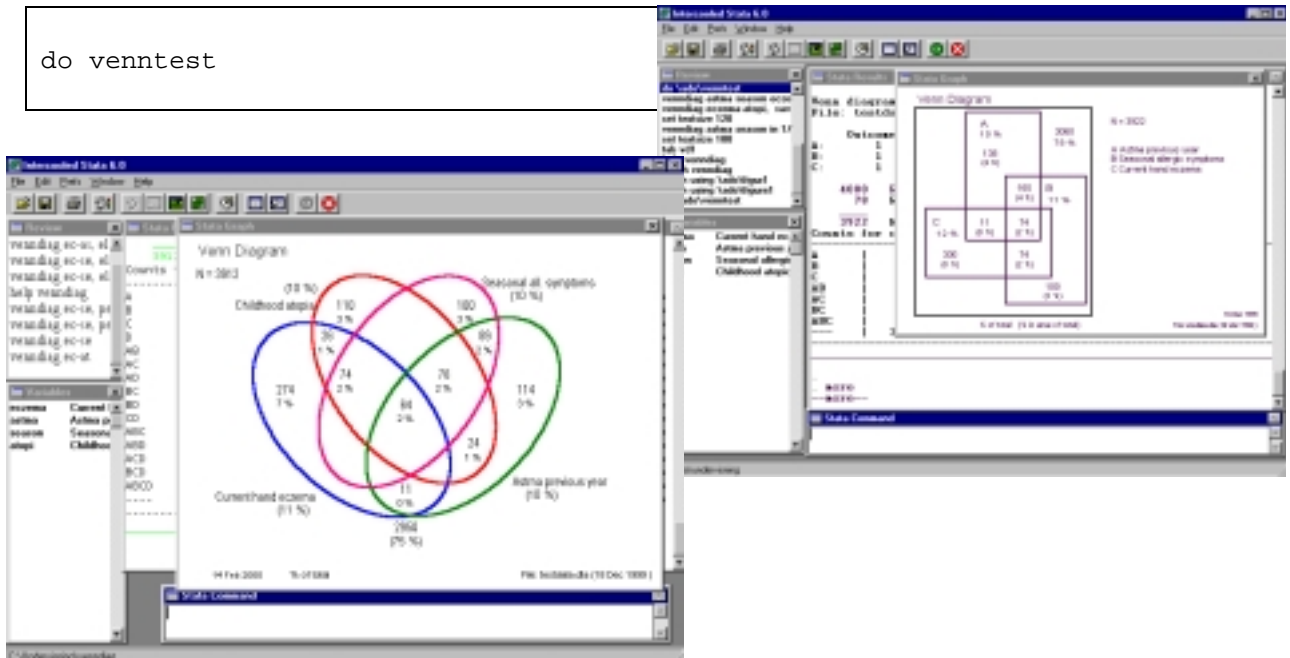
. update
Stata executable ..... currently installed: 20 Mar 2000
Ado-file updates ..... currently installed: 05 Apr 2000

. graph egg teqg
```

Hvis mmerge ikke virker skal du sikre, at både Stata program og ado system er opdateret. Du kan ovenfor se, hvilke udgaver der blev anvendt ved udarbejdelse af noten.

Øvelse 45 Venn Diagram

Tegn et Venn Diagram for overlap af symptomerne eksem, høfeber og astma. Brug til dette den "do fil", som er beskrevet under øvelsesdata med navnet "venntest.do". Hvor stor en andel af observationerne har **ingen** af de fire symptomer ?



Prøv desuden nogle options, fx "missing" og gen(). Se "help venndiag".

Øvelse 46 Stata som regnemaskine. Statistik ved indtastning

Der findes en række kommandoer der kan bruges til at få et umiddelbart test eller andet. Her er fx beregnet confidensinterval for et poissonfordelt udfald og for en proportion.

```
. cii 2 35, poisson
Variable | Exposure      Mean      Std. Err.      -- Poisson Exact --
          |                [95% Conf. Interval]
-----+-----
          |           2      17.5      2.95804      12.19272      24.33801
. cii 100 5
Variable | Obs           Mean      Std. Err.      -- Binomial Exact --
          |                [95% Conf. Interval]
-----+-----
          |          100      .05      .0217945      .0164307      .1128395
. tabi 30 18 38 \ 13 7 22, chi2 exact
   row |           1           2           3 | Total
-----+-----
     1 |           30           18           38 |      86
     2 |           13           7           22 |      42
-----+-----
 Total |           43           25           60 |     128
Pearson chi2(2) = 0.7967   Pr = 0.671
Fisher's exact = .707
```

Slå op i hjælpesystemet under **cii**, **tabi**

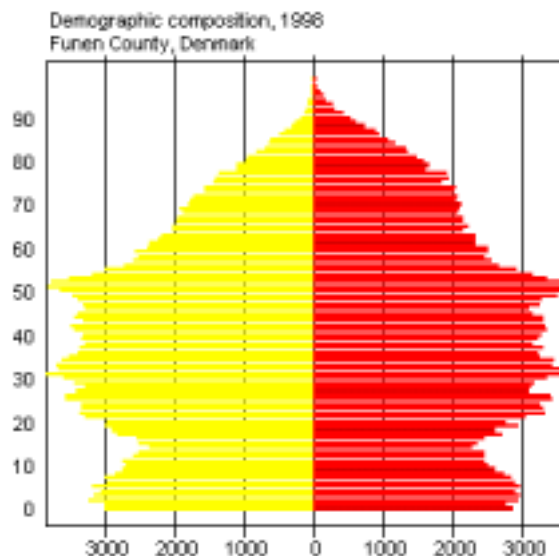
Øvelse 47 Befolkningspyramider

Tegn en befolkningspyramide for Fyns Amt. Samt en tilsvarende pyramide, som viser forskel i antal mænd og kvinder i de forskellige aldre. Dette gør du ved at gennemgå den tutorial, som hedder pyramid.

Denne startes således :

```
tutorial pyramid
```

Når du har gennemgået tutorial pyramid kan du udføre den tilsvarende *pyramid.do*

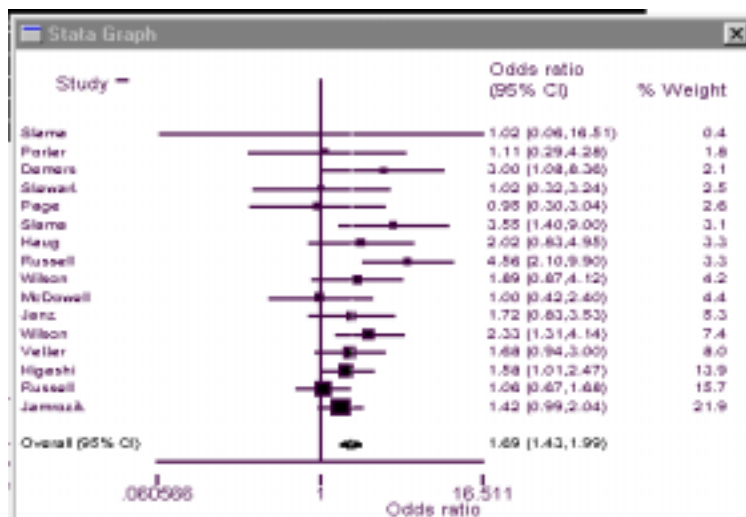


Øvelse 48-s Meta analyser

Find ud af hvordan dette gøres i Stata. Filen *meta.do* indgår i kursusmaterialet.

Søg i hjælpesystemet med ”*search metabias*” eller ”*search meta*” .

Hvis de ikke er på din Pc, så hent metan.ado og example1.dta fra STB-44 og STB-45, hvorefter du kan udføre denne graf:

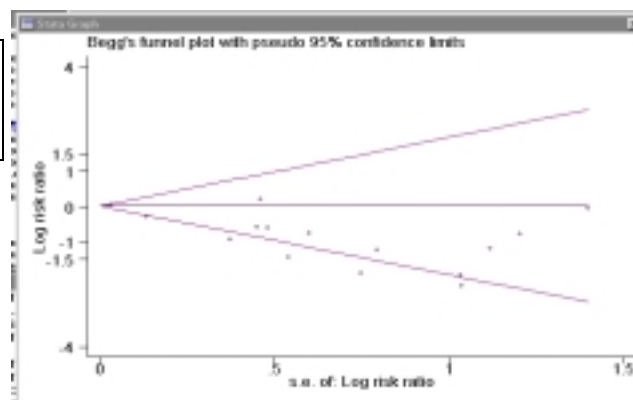


Bemærk, at meta analyser er en omfattende disciplin i sig selv. Og ikke kun et spørgsmål om at tegne grafer. Udvalgte dele findes i filen meta.do der udføres ved at skrive:

```
do meta
```

Nedenfor ses et såkaldt **funnel plot**.

Hvis der ikke er publikationsbias vil punktestimaterne i de publicerede studier være symmetrisk fordelt omkring det samlede estimat.



Eksemplet er en del af *meta.do*.

Tabel fra Stata til tekstbehandling

Når den endelige tabel ønskes publiceret skal den overføres til tekstbehandlingen **uden at tal tastes igen**. Her ønskes summeret antal hændelser i to variable v7a2 (alle hændelser) og v712 (lægebehandlede hændelser), samt risikotiden (risk2)

```
table v2f2grp5, c(n v7a2 sum v7a2 sum v712 sum risk2) col row
```

	N(v7a2)	sum(v7a2)	sum(v712)	sum(risk2)
Owner 25				
20-24	2	2	1	4056
25-29	16	8	2	39496
30-34	35	23	4	96498
35-39	44	42	10	123420
40-44	51	62	9	151732
45-49	45	21	5	120260
50-59	107	48	5	296400
60-69	38	9	2	83954
70+	8	4	1	21996
Total	346	219	39	937812

Tabellen markeres som en blok i resultatvinduet eller logfilen (dvs uden kommandoen), tryk dernæst **Ctrl+C**. Dernæst skiftes til tekstbehandling (her Word 97) **Alt+Tab** og når tekstbehandlingen er aktiv kan du tilføje selve tabellen som tekst ved at trykke på **Ctrl+V**

Her er den samme tabel omsat til en pænere tabel vhj.a “convert txt to table funktionen” i word (findes under ”Table”). Gør sådan:

- Markér tabellen som en blok i resultatvinduet eller logfilen. Tryk på **Ctrl+C**
- Sæt tabellen ind i tekstbehandlingen i et tomt dokument: Tryk på **Ctrl+V**
- Brug en søg og erstat funktion og ret alle “ ” (to mellemrum) til ét ” ” indtil der ikke er flere end højst ét ad gangen.
- Ret alle mellemrum til “;”
- Aktivér ”Convert text to table”, som findes under ”Tools” Angiv som kolonneseparator “;” ved omformningen. Sæt evt. højre justering af cellerne.

Med sletning af et par tomme celler og lidt tilpasning af kolonnetekster kan tabellen publiceres. Omformningen fra teksttabellen ovenfor til dette tog kun kort tid (få minutter) . Og **bemærk** at data er kommet direkte fra rådata. Dvs, at der ikke er sket fejl i tallene pga fejlindtastning.

Table 1: Age of owner and number of accidents.

	Persons	All Injuries (Count)	Hospital treated injuries	Time at Risk (Years)
20-24	2	2	1	4056
25-29	16	8	2	39496
30-34	35	23	4	96498
35-39	44	42	10	123420
40-44	51	62	9	151732
45-49	45	21	5	120260
50-59	107	48	5	296400
60-69	38	9	2	83954
70+	8	4	1	21996
Total	346	219	39	937812

Der findes også en makro til word. Indgår i filen hinst.zip, som er en del af kursus.zip). Se desuden kommandoen **outreg**

Graf fra Stata til tekstbehandling

Eksemplet er vist ud fra Word97 som tekstbehandling. Data er fra et studie, hvor Short Form 36 helbredsstatus målet blev anvendt til at udtrykke egen vurdering af helbred. De konkrete mål PCS og MCS er udregnet i en Stata fil til formålet. Prøv kommandoen hvis du vil vide mere om SF36 og Stata. Se også Øvelse 19, side 20.

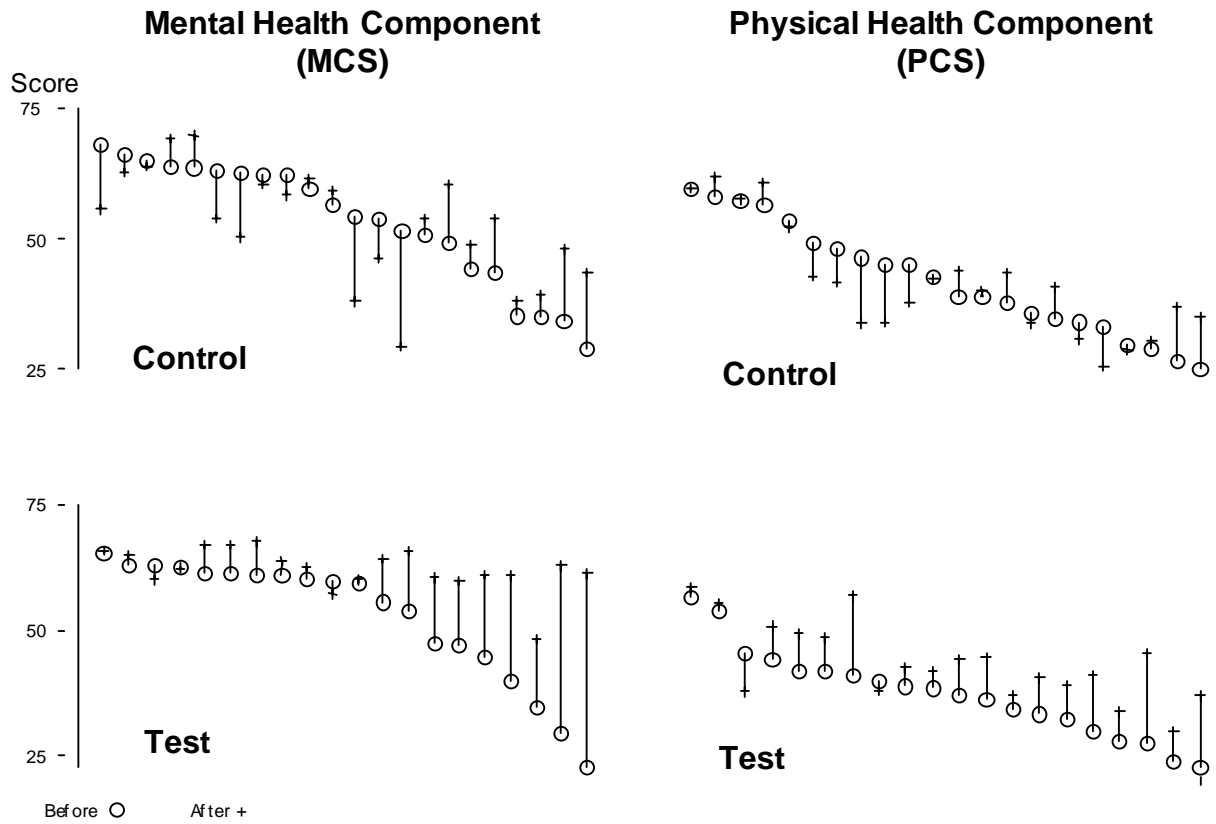
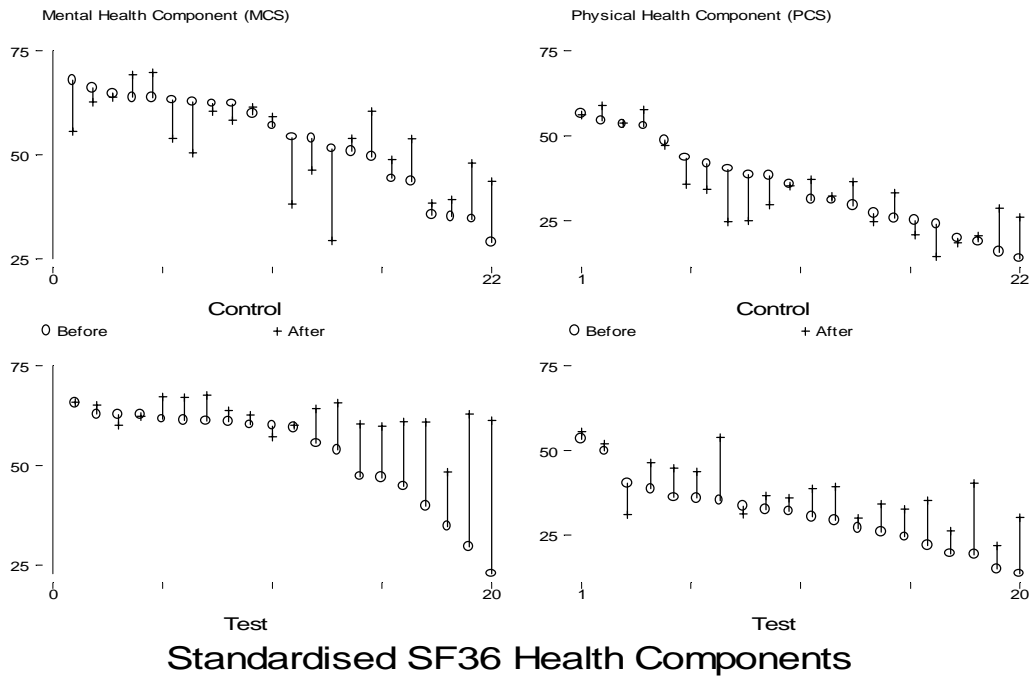
```
webseek sf36
```

Følg denne punktliste, som er en kopi af et indlæg jeg sendte til Stata-List diskussionslisten.

- 0a. Make sure you have the wmf import filter installed in word97.
- 0b. Make sure graph options in Stata are set
(copy format wmf, with color/black white, lines etc)
1. Make graph as final as possible in Stata.
2. Copy graph or Export to wmf file using gphprint, e.g.:
gphprint, nologo saving(figur1.wmf,replace)
3. Start Word
- 3a. Import into word as wmf from file (menu: insert . Picture . from file)
- 3b. Copy from clipboard (Ctrl+V)
4. Double click on the picture and you are in an "edit picture" mode.
5. Open the "drawing" toolbar
6. Put your mouse somewhere in the middle of the graph and "un-group" the whole content. (toolbar left "draw", submenu group . ungroup).
(Can be done with right mouse button click as well)
7. Click somewhere else to unselect all the objects which were auto selected by the ungroup procedure.
8. Encircle with the pointing device the objects you want to change,move or delete (drawing toolbar. Arrow number to from the left)
9. Make the changes. (change font, change line thickness etc). Add lines add text or whatever you prefer.

The tricky part is to make the objects ungrouped. Otherwise you cannot change fonts or line thickness for single points or lines etc.

Examples raw figure from stata and revised one available on next page.



Konvertering af data

Når data skal udveksles med andre kan det være nødvendigt at oversætte til et andet format. Der er ingen universel gyldigt format, men hvis du skal modtage data og afsender ikke kan sende dem som Stata filer, så vælg én af disse, hvis det er ét ældre system:

Forskellen mellem formater er især navngivning af variable (længden af variable), samt antallet af variable og muligheden for at navngive værdier i de enkelte variable.

Rene databaseformater (dvs kun med navn på variable)

- **Dbase 3** eller Dbase IV
- **Lotus 1-2-3, type Wks**
- Andre typer : Paradox rel 5, Access

Statistik program dataformater vil oftest indeholde både variabelnavne, definition (labels) på de enkelte værdier og særlige koder for uoplyst. Fx:

- Stata, SPSS eller SAS

EpiData

- Skriver data i samme format som EpiInfo, men på en måde hvor labels også kan defineres. Data kan eksporteres direkte til Stata med labels. Første udgave af **EpiData** der kan eksportere EpiInfo filer kan rekvireres fra medio april 2000 (Tro dog aldrig på sådanne tidsangivelser).

Rene talfiler (Ascii filer) indeholder kun tal. Se fx. Øvelse 50. Dette kan være uproblematisk, hvis der følger en grundig dokumentation med filerne, men uden en meget grundig beskrivelse kan Du risikere at læse data forkert.

Når data er modtaget kopierer du til harddisken og opbevarer en kopi af de originalt modtagne data udenfor Pc'en, sammen med den modtagne dokumentation. Husk at kontrollere for virus, hvis data kommer på en diskette eller på en form, der kan indeholde virus. Oversættelse til Stata format med programmet Stat/Transfer er let. Det kan konvertere stort set alle formater, også paradox, som ellers ofte kan være et problem.

Øvelse 49-s Oversæt data til Stata format.

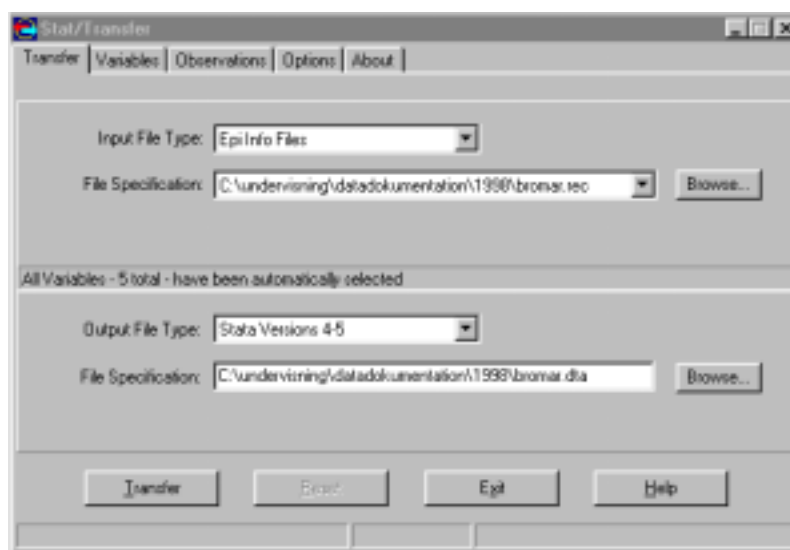
Brug **Stat/Transfer** eller **EpiData** til at oversætte filerne *eqfirst.rec*¹⁶ og *eqsecond.rec* til ”Stata format. (skærbilledet viser navnet på en anden fil)

Du angiver blot

Input File Type **Filnavn**

1. Output File Type.

Under ”Variables” eller ”Observations” kan du evt fravælge enkelte variable eller uddrag af datafilerne. Det er vigtigt, at der under ”Options” angives, at variabelnavne skal skrives med småt. Stata skelner x fra X !!.



¹⁶ Hvis epi-info ikke vises som en mulighed under ”Input File Type” kan du enten købe en nyere udgave eller alternativt kan du i EpiInfo /EpiData oversætte til dbase3 format.

Samlet oparbejdning af Bromaraton data med Stata

I de næste øvelser vises et samlet eksempel på oparbejdning af et datasæt til analyser.

Inden oparbejdningen planlægges er det vigtigt at se på hvilke analyser der skal udføres. Oparbejdningen udføres dels for at kontrollere data, men også for at sikre, at analyserne giver entydige svar. Forstået på den måde, at der ikke må være tvivl om hvordan de forskellige variable er opstået og defineret. Se afsnittet side 8.

Ofte indsamles data uden at de konkrete analyser er overvejet præcist, hvilket bl.a. kan betyde at der går unødigt megen tid med at få omformet data til en hensigtsmæssig struktur. Et af de vigtigste punkter er her at beslutte hvilken analyseenhed, der skal være¹⁷.

Vi vender tilbage til øvelsesdata 1 fra bromaraton løbet. Den overordnede interesse for analyserne er sammenhængen mellem alder, køn og dels gennemførelse af løbet, men også hvor hurtigt løbet blev gennemført. Endelig om dette er uafhængigt af hvor løberne bor.:

a. Gennemføres løbet overhovedet:

H⁰: Tendensen til at gennemføre et maratonløb er lige stor uanset hvor løberne bor.

H^a: Personer med bopæl langt fra løbet har større tendens til at gennemføre.

b. Hvor lang tid tager det at gennemføre løbet:

H⁰: Løbet gennemføres lige hurtigt uanset bopæl. (~ løbstid er uafhængig af bopæl)

H^a: Personer med bopæl langt fra løbet gennemførte hurtigst.

Fra løbsledelsen blev data fremsendt på en diskette¹⁸. Efter modtagelsen skal der foretages kvalitetskontrol og beregnes afledte variable. Dette gælder både for disse data, og data du selv modtager fra andre personer eller udlæser fra analyseudstyr, som opsamler data. Det første du gør når du modtager data er visuelt at få et indtryk af form, kvalitet og ensartethed.

I resten af noten gentages nogle af arbejdsformerne. Du kan selvfølgelig frit vælge ud fra de metoder der er vist i de foregående afsnit.

```
"0695", "5270", "39", "M", 41619
"0697", "7080", "78", "M",
"0698", "7080", "51", "M", 35550
"0699", "d-4000", "59", "M",
"0701", "8860", "42", "M", 32034
"0702", , "43", "F", 34847
"0703", , "46", "F", 41301
```

Øvelse 50 Kig på tallene.

Få et visuelt billede frem af *bromar.txt* frem. Brug

en editor eller *Win-Commander* til dette. ① *Hvad ser du?* ② *Er der de rigtige variable?* ③ *Hvilke typer variable er der?* (Tal, binær, ordinal, bogstav, kategorisk). (angiv i skemaet) †
Findes alle løbsnumre?

Indhold	Kodning	Type
id	Dét nummer løberne var tildelt	
postnummer	For bopæl.	
	For udlændinge er angivet landekode og nr. Fx s-120	
fødselsår	kalenderår, hvor personen er født.	
køn	F=kvinde M=mand	
Løbstid	Timer, minutter og sekunder.	
	Fx 40101 (4 timer, 1 minut og 1 sekund).	

Variable for data i filen: *bromraw.txt* (rådata ascii fil)

Det er flere forhold at lægge mærke til:

1. Nogle personer mangler postnummer. 2. Nogle personer har et bogstav i postnummeret . 3.

¹⁷ Se fx EpiData - Datadokumentation og dataindtastning af J.M.Lauritsen, seneste udgave.

¹⁸ Løbsdata findes nu fra mange løb på internettet, hvis du vil prøve at analysere andre løbsdata.

Nogle personer mangler løbstid. 4. Køn er angivet med bogstaver i stedet for tal 5. Nogle løbsnumre mangler tilsyneladende, kan du finde et eksempel ?

Vi kunne rette op på de inkonsistente forhold ét ad gangen i et tekstbehandlingsprogram indtil data var systematiske, men dette ville IKKE være reproducerbart. Jvnf. reglen fra udvalget vedr. videnskabelig uredelighed (www.forskraad.dk), at data skal kunne følges fra originalbilag til endelig publikation. Derfor bør alt arbejde udføres, således at det er klart for en selv og andre, hvad der er sket. Dette kaldes oparbejdning, se side 8.

Ved oparbejdningen her skal vi sikre:

- at der er data for alle personer, hvor der skulle være data.
- om der er personer der mangler så store dele af data, at de reelt ikke bidrager til analysen.
- at der oprettes de nødvendige variable til de ønskede analyser.
- at data er grundigt beskrevet, så vi ikke tager fejl i tolkningen (betyder 0 ja eller nej)
- at ingen person er indtastet to gange.
- at der udarbejdes en samlet dokumentation, som beskriver indtastningsrutiner, datakontrol og kriterier for kodning af variable.

For andre typer data er det desuden nødvendigt at undersøge om data logisk er konsistente. Hvis man angiver 3 børn og ønsker alder for disse må der ikke være angivet alder for fire børn. En mormor kan ikke være 24 år gammel. En person med hæmoglobin på 3.5 mM/l kan ikke løbe omkring. Det er selvfølgelig ikke meningen, at man skal fjerne usandsynlige sammenhænge, som ofte er en del af forskningsprocessen. Men indholdsmæssigt eller logisk fejlagtige sammenhænge i data bidrager ikke til belysning af de opstillede hypoteser. Tværtimod er de snarere et tegn på den omhu der indgik i formulering af spørgsmål til de indsamlede data og i den grundighed målgruppen har anvendt til at besvare spørgsmålene. En række andre forbehold kunne nævnes, men dette er de mest grundlæggende.

Løbstiden har en uhensigtsmæssig skala Fx betyder 41826 at løbet er gennemført på 4 timer 18 minutter 26 sekunder. Værdier fra fx 46001 til 49999 kan slet ikke forekomme. Tiden skal altså omregnes og postnummer skal omsættes til afstand fra bopæl til løbet.

Vi ønsker denne kodning:

Indhold	Kodning	Navn på variabel
id	Dét nummer løberne var tildelt	id
afstand	Afstand fra bopæl til løbsstedet. – tilnærmet kontinuert - grupperet 0-25,26-50, 51-100,101-200,200+	km kmgrp
fødselsår	kalenderår, hvor personen er født.	born
køn	0=kvinde 1=mand	sex
Løbstid	Numerisk. Løbstid i timer + decimaldel af timer (Fx 4½ time= 4.50)	dectime
Gennemført	Binær variabel, som angiver om personerne gennemførte løbet	runok


Data med disse variable skal til sidst gemmes i filen: **bromar.dta** (Stata format).

Gradvist vil vi nu opbygge en ”do” fil, som kan læse rådata og beregne de nødvendige afledte variable.

Opgave 1: Hvor mange personer og omfang af uoplyst i de enkelte variable ?

Læs data og find ud af hvor mange personer der er og hvor mange personer, der har uoplyst i de forskellige variable.

Øvelse 51 Start Stata og gør klar til dagens opgave

Start Stata , og se efter i nederste venstre hjørne, at du er under det rigtige projekt (her forudsættes *c:\data*)

Øvelse 52 Indlæs data fra originalfilen

Indlæsning af data er grundigt beskrevet i Stata Users Guide (**help infile**). Læs dette senere og skriv disse linier direkte ind én ad gangen:

```
* Det er en god ide, at inkludere mange linier med kommentarer
insheet id pnr born sex runtime using bromar.txt, clear
describe
list in 1/15
codebook
summarize runtime
```

Øvelse 53 Gem kommandoer i en "do" fil

Filen kaldes "do" fil, fordi den indeholder kommandoer, som kan udføre

noget. Gem linierne som vist i Øvelse 18. Peg med musen på den lille firkant til venstre for "Review" Angiv *oparbejd.do* som navn¹⁹ og husk at gemme i mappen *c:\data*.

**Øvelse 54-s Gem kommandoer i en "do" fil (alternativ strategi)**

I stedet for foregående øvelse kan du gøre sådan her. (En eventuel åben logfil skal lukkes med `log close` først) Gør sådan her :

```
log using "oparbejd1.do"
#review 50
log close
```

Øvelse 55 Kontrollér indholdet af oparbejd.do i "Do-file" editor

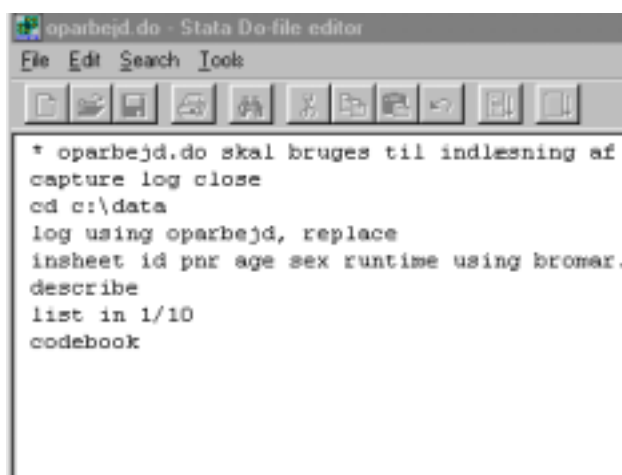
Skift til Stata (**Alt+Tab**) og skriv i kommandofeltet: **doedit oparbejd.do** eller alternativt click med musen på den øverste del af Stata Skærbilledet, femte icon fra højre og sørg derefter for at åbne (**Ctrl+o**) *oparbejd.do*

```
. doedit oparbejd.do
```

På skærmen åbnes nu den editor som er indbygget i Stata (fra Version 6).

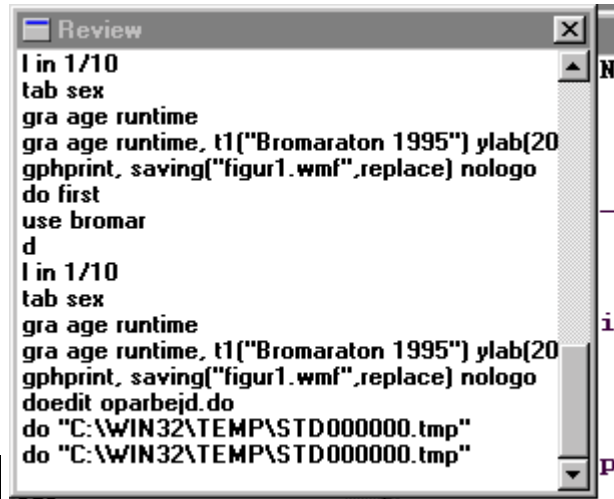
(På billedet til højre kan du se, at jeg har valgt at tilføje linie 2-4 , der åbner en logfil - dette er valgfrit. Logfilen kunne også åbnes ud fra "toolbar")

Du kan nu i do-file editoren rette linier, som er forkerte og dernæst markere en



¹⁹ Eller et andet navn. Hvis du vil have mellemrum i navnet kan det være nødvendigt at sætte et " " i begge ender" af hele navnet.

blok af din "do-fil" og taste **Alt+T S** (menuen tools i dofile editor) for at få kommandoerne i den pgl. blok udført i stata.



Bemærk, at det nu kun er en kryptisk linie "C:\win....." der kommer ind i review vinduet.

Strategi 2 "Ekstern Editor" Alternativ for erfarne brugere.

Prøv at bruge den eksterne editor **pfe** eller **notepad**, der kan bruges som alternativ til do file editor. (Forudsætter at du har udført Øvelse 78- Øvelse 81 , samt evt. Øvelse 83 og Øvelse 84. En fordel ved **pfe** er, at selv meget store filer kan indlæses. De fleste mindre programmer, såsom "do-file-editor" eller "notepad" kan højst læse filer på ca 64 kb.

Øvelse 56-s Kontrollér indholdet af oparbejd.do med brug af Win-Commander

Start Win-Commander og skift dertil med (**Alt+Tab**), peg på filen *oparbejd.do* og tryk på **F4**. Kommandofilen *oparbejd.do* indlæses nu og du kan se indholdet.

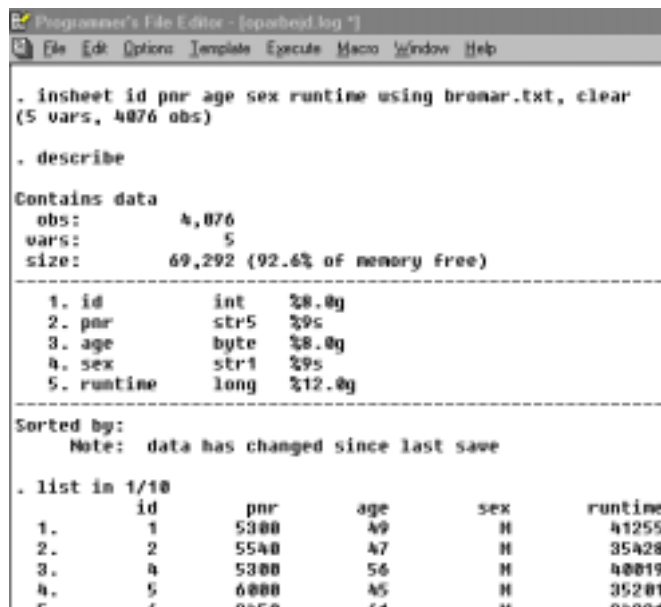
Øvelse 57-s Se på resultatet fra logfilen

Skift til Win-Commander (**Alt+Tab**), peg på filen *oparbejd.log* og tryk på **F4**. Log filen indlæses nu og du kan besvare spørgsmålene i opgave 1.

Se i logfilen, at der er følgende:

Antal personer er 4076
 5 variable
 type
 id int (integer)
 pnr str5 (5 bogstaver)
 age byte(kompakt tal)
 sex str1 (1 bogstav)
 runtime long (langt tal)

Antal uoplyste (ikke vist) er:
 id 0
 pnr 422
 age 260
 sex 0
 runtime 478



Vi vil kigge ekstra grundigt i logfilen under variabelen id:

```
. codebook
id ----- (unlabeled)
      type: numeric (int)
      range: [1,4255]          units: 1
      unique values: 4076      coded missing: 0 / 4076
```

Der er ingen personer, som er med to gange (Unique values 4076 ~ alle 4076)

Der mangler nogle løbsnumre !. Sidste løbsnummer er 4255, første nr 1 og $4255-4076=179$.

Dvs der er slet ingen data for 179 numre. Kun ved at tale med løbsledelsen kan vi vide om disse 179 er nogle der har meldt fra igen eller om rutinen var sådan, at alle lige efter start blev registreret. Det oplyses, at de manglende numre ikke har været brugt.

Dette er en meget god strategi. Det bør **aldrig** ske, at tildelte id numre (identitets numre) genanvendes til en ny person. Det vil uvægerligt give fejl senere i analyser. Hvis der i en undersøgelse fx er både et spørge- og et journalregistreringsskema. Så kan man risikere at to skemaer fra to forskellige personer har samme id, hvis numre har været genanvendt

Antallet af personer, som ikke gennemførte løbet er derfor 478 (dem uden løbstid).

Vi kan på nuværende tidspunkt skrive første del af et materialeafsnit eller datadokumentation, fx således:

Data til undersøgelsen er modtaget direkte fra løbsledelsen på diskette. Registreringsproceduren var udformet således, at alle der startede på løbet var med i de leverede data, som bestod af 5 variable (anonymt id nr., postnr, alder, køn og løbstid) for 4076 personer. Hvis løbet ikke blev gennemført var løbstiden kodet til 0 (n=478). Antal uoplyste var: Køn (0), alder (n=260) og postnummer (422). Løbstiden blev automatisk registreret ved at løberne ved målet afleverede en stregkode med løbsnummeret (id). Som en ekstra sikring af fejlaflæsning blev afleverede løbsnumre gemt i kontinuerte bundter, således at en manglende aflæsning af tiden kunne interpoleres mellem de nærmest liggende afleverede løbsnumre. Det præcise antal, hvor dette er sket kan ikke angives, men er oplyst til at være under 1 %.

Resume.

Nu er alle vigtige dele af selve arbejdsformen præsenteret og gentaget.

Bemærk at dokumentationen skrives undervejs. Det gør det lettere at arkivere data.

Både i egne arkiver og i Dansk Data Arkiv/Eras (se www.dda.dk/eras).

Du skal selv vælge hvordan du vil arbejde. I øvelserne er præsenteret både en standardmåde med do-file-editor og for de erfarne et alternativ med ekstern editor og i et vist omfang Win-Commander. Desuden menu systemet Quest.

Hvis du er i tvivl om nogle aspekter ved arbejdet nu. Så vend tilbage til de tidligere kapitler og gentag nogle af øvelserne.

Udeluk personer og variable med for mange uoplyste.

Skal nogle personer udgå fordi de mangler for mange oplysninger (fx de der mangler postnummer, alder og ikke har fuldført).

Stata har i datasæt en standardværdi for uoplyst, som er "." (punktum) for numeriske variable og "" (ingenting) for streng variable.

Operatorer ved logiske sammenligninger er:

== (to lighedstegn) betyder lig med
 != (udråbstegn og lighedstegn) betyder forskellig (~= kan også bruges)

Se flere muligheder ved at udføre ordren: **help operators**

Øvelse 58 Insufficente data for enkeltindivider ?

Beregn en variabel, som angiver hvor mange variable personen har uoplyst for.

```
* Ny variabel "fejl" angiver antal uoplyste for hver person
count if age == .
assert sex != ""
assert age != .
generate fejl = 0
replace fejl = 1 if age == .
replace fejl = fejl+1 if pnr == ""
list if fejl > 1 in 1/100
list if fejl == 2
tabulate fejl sex
```

Linien med "count" kan udelades, men count er god hvis man vil afprøve en logisk sætning, her er svaret 260. Hvis du vil sikre, at en given betingelse er opfyldt for alle data skal du i stedet bruge **assert age != .** **Assert** sikrer, at du ikke kan arbejde videre før du har taget stilling til hvad der skal ske, hvis betingelsen ikke er opfyldt. Hvis du vil undgå at Stata stopper kan du indsætte **capture noisily** foran **assert** (Prøv begge muligheder i en "do" fil).

Efter disse analyser skal vi beslutte om nogle personer skal udelukkes pga manglende data i for stort omfang. Vi kigger derfor i logfilen igen.

Beslut nu om du ud fra resultatet vil udelukke nogen personer, fordi de reelt ikke kan bidrage til analysen. Vi kan udelukke ingen, 19 eller 644

```
. list if fejl > 1
      id      pnr      age      runtime      sex      fejl
1979.  2008                .      25828      M        2
..... output udeladt .....
3857.  3892                .      30745      M        2
. tabulate fejl sex
      fejl | sex      K      M | Total
-----+-----+-----
      0 |      397     3016 |     3413
      1 |      94      550 |      644
      2 |       2      17  |      19
-----+-----+-----
Total |     493     3583 |     4076
```

Vi vælger at udelukke de 19 personer fordi alder og afstand er meget væsentlige i vores analyser og tilføjer derfor til materialeafsnittet (Er du enig?):

.....gende afleverede løbsnumre. Det præcise antal, hvor dette er sket kan ikke angives, men er oplyst til at være under 1 %. Nitten personer (17 mænd og 2 kvinder) havde ikke oplyst alder og postnummer for bopæl, disse udgår og analyserne er derfor udført på grundlag af de resterende 4057 personer

Vi skal derfor udelukke de 19 personer fra data. Tilføj en linie til *oparbejd.do*:

```
* De 19 personer udelukkes
drop if fejl == 2 /* De slettes, men findes i bromar.txt */
```

Når de 19 personer slettes med kommandoen `drop` betyder det blot, at de udgår af den **kopi** af data, som er indlæst i hukommelsen på PC'en. De findes stadig i rådata filen. Med **drop if fejl == 2** smides alle observationer ud, som opfylder betingelsen, at de har værdien 2 i variabelen **fejl**. Hvis vi i stedet havde skrevet **drop fejl** var det variabelen **fejl** der var slettet for alle personer. Hvis mange variable skal slettes er det lettere at skrive **keep** og angive de variable, som skal blive tilbage.

Variable kan angives i forkortet form. Hvis variable i en fil er: c1 c2 c3 c4 c5 c25 kan du spare tastearbejde ved at angive variable på kort form:

drop c1-c15 /* alle variable fra c1 til og med c15 slettes */

list age-fejl /* fra eksemplet ovenfor. Variablene age sex runtime fejl listes */

Endnu kortere kunne dette skrives sådan fordi der kun er én variabel med **a** og én med **f**, **list** kan forkortes til **l**, dvs at vi kan nøjes med at skrive **"l a-f"**:

```
. l a-f
      age      sex      runtime      fejl
1.    62       M      22424         0
2.    67       M      23124 ..... resten af output udeladt
```

Beregn afledte variable.

Løbstiden skal ændres, således at timer omregnes til et kontinuert mål. Desuden skal vi beregne en variabel som angiver om løbet var gennemført. Samt omsætte postnummer til afstand fra løbet.

Der findes en lang række funktioner indbygget til formålet. Kig i "Stata User Manual", samt i hjælpesystemet **"help generate"** **"help replace"** **"help recode"** **"help egen"** **"help functions"** **"help cut1"** **"help cut"**.

Beregning af afledte variable er en erfaringssag. For hver afledt variabel vil der indgå et element af kontrol af data. I eksemplerne nedenfor kan du se, hvordan jeg ville arbejde med at få sikkerhed for at variabelen **pnr** faktisk kunne bruges til at beregne afstande.

Øvelse 59 Kontrol af variabelen pnr

Læs nedenstående grundigt og gentag nogle af de søgninger jeg viser i Stata.

Del 1. Tilretning af afstand til løbsstedet.

Da vi kiggede på data fra filen *bromraw.txt*, se side 37. Bemærkede vi, at der i nogen postnumre var bogstaver og nogle havde længden 5, mens andre var fire karakterer i bredden. Vi skal derfor nu undersøge variabelen **pnr** nærmere.

For at få en ide om variabelen **pnr** vil vi først se på en frekvenstabel af denne:

```
. tab pnr
```

pnr	Freq.	Percent	Cum.
	1	0.03	0.03
500	1	0.03	0.05
--	1	0.03	0.08
0812	1	0.03	0.11
1138	2	0.05	0.16
12467	1	output udeladt

Allerede her ser vi, at en person har blankt postnr, én har værdien "--". Værdierne "500" og "0812" er ikke lovlige postnumre. Desuden er værdien "12467" for lang, men vi ved at postnumre i Tyskland er fem cifrede i nogle områder. Vi kunne starte med at se på længden af variabelen **pnr**. Funktionen **length(pnr)** viser længden af variabelen **pnr**, som jo er en streng variabel (se evt. **help datatypes**).

```
. generate le = length(pnr)
. tab le
```

le	Freq.	Percent	Cum.
0	422	10.35	10.35
1	1	0.02	10.38
4	3622	88.86	99.24
5	31	0.76	100.00
Total	4076	100.00	

Stemmer dette overens med frekvenstabellen ovenfor. Ja og Nej. Vi må undre os over at der ikke er nogen med længden 3. I tabellen ovenfor var der jo en med værdien "500", men da mellemrummet står i starten er længden af denne i alt 4. Lad os se på de observationer, som indeholder mellemrum " ":

```
. list if index(pnr," ") > 0
```

	id	pnr	age	sex	runtime
1826.	1851	pl 58	48	M	33030
3618.	3653	500	59	M	35405
3930.	3965		59	M	33529

Desuden kunne det være en god ide at se nærmere på de der har kort længde. Med funktionen **trim** fjernes eventuelle mellemrum i starten af **pnr** og med den anden betingelse: **length(pnr) > 0** sikrer vi at alle med uoplyst **pnr** slet ikke kommer med. De to betingelser bindes sammen af **&**. Hvis vi skulle sige "eller" bruges **|** (en lodret streg), som på de fleste tastatur er nr to til højre for tasten med 0 (nul). For at få **|** frem skal du også trykke på **AltGr**

```
list if length(trim(pnr)) < 4 & length(pnr) > 0
```

	id	pnr	age	sex	runtime
3078.	3111	--	55	M	33809
3618.	3653	500	59	M	35405
3930.	3965		59	M	33529

Ved at se dette bliver jeg nysgerrig efter at kontrollere hvorfor længden af variabelen **pnr** for **id=3965** er større end nul, umiddelbart ser den ud til at være helt uoplyst.

Derfor kigger jeg i den originale datafil med Win-Commander og ser dette billede:



Aha – forklaringen må være, at der står et ekstra mellemrum ud for denne person²⁰ inden komma nr 2. Prøv at sammenligne med id nr 3963 eller 3964.

Eksemplet viser, at der er vigtigt at være minutios i oparbejdning af data. Man skal ikke kontrollere alt muligt, men gå systematisk til værks for hver enkelt af de variable, som indgår i de hypoteser man ønsker at arbejde med. Vær opmærksom på at data indsamlet med automatisk dataregistrerings udstyr, fx lungefunktionsmåling også skal kontrolleres grundigt. I et bestemt projekt, hvor der skulle udskrives tre målinger pr patient, var der fejl i et ikke ringe antal ud af 700. Nogle havde kun 2 målinger. Hvis vi havde accepteret data kritiskløst ville det hurtigt gå galt, fordi det der så læstes som måling nr 3 for en given person faktisk var måling nr 1 for den næste person.

Øvelse 60 Stata som regnemaskine. Afprøv kommandoen ”display”

Prøv at skrive sådan i kommandofeltet i Stata:

```
. display index("5000","-")    " " length("5000")
0  4
. display index(" 500","-")    " " length(" 500")
0  4
. display index("D-5000","-")  " " length("D-5000")
6
. display 375/25

* også andre funktioner kan bruges, fx Chi2 tabeller:
* Hvad er P-værdien for Chi2= 3.85 med 1 frihedsgrad ?

. di chiprob(1,3.85)
.049746
```

Brug anledningen til at slå op i hjælpesystemet under **functions** og **display**

Efter at have prøvet lidt forskelligt for at se kritisk på **pnr** ved jeg at 30 observationer er fra personer med et udenlandsk post nummer. Desuden er der tre, som vi ikke kan afgøre om de er udenlandske eller blot nedskrevet forkert og vælger derfor at omkode det forkerte postnummer til uoplyst. Da køn var oplyst for disse tre skal de ikke udgå.

Jeg vælger at arbejde videre med en ”do” fil, og tilføjer disse linier til *oparbejd.do*:

```
generate str5 pnr1 = pnr
* foreign and invalid postal numbers are identified:
replace pnr = "9998" if (index(pnr,"pl") > 0)
```

²⁰ Umiddelbart undrer jeg mig over at filen læses på den pågældende måde og kunne finde på at prøve om jeg i en kopi af filen kan introducere den samme ”fejl” i et andet **id** nr. En mail til Stata listen blev sendt. Se spørgsmål og et svar i filen *mail.txt* i kursusmaterialet.

```

replace pnr = "9998" if length(pnr) > 4
replace pnr = "" if length(trim(pnr)) < 4
replace pnr = "9998" if (index(pnr,"-") > 0)
replace pnr = "" if index(pnr," ") > 0 /* embedded blank */
count if pnr1 != pnr /* count changed observations */
count if pnr == "9998" /* foreign postal codes */
list pnr1 pnr if pnr != pnr1 /*list all changed observations */
drop if pnr == "9998" /* foreign postal codes */
drop if fejl > 1 /* sex and postal code missing */

* nu er pnr i orden og vi danner en talvariabel i stedet.
* Funktionen real("123") omsætter indhold fra "123" i pnr til 123 som tal i postnr
generate postnr = real(pnr)

```

Øvelse 61 Rækkefølge af linier med rekodning !

Der vil være forskel på antal gyldige observation med linierne ovenfor og dem der er vist her nedenfor :

```

replace pnr = "9998" if (index(pnr,"pl") > 0)
replace pnr = "9998" if length(pnr) > 4
replace pnr = "9998" if (index(pnr,"-") > 0)
replace pnr = "" if length(trim(pnr)) < 4
replace pnr = "" if index(pnr," ") > 0 /* embedded blank */

```

I den ene situation bliver der 4027 observationer i analysen og i den anden 4026. Hvilket **id** gør forskellen ?

Da vi nu har taget nogle beslutninger der er vigtige, skal der tilføjes et afsnit til materialebeskrivelsen (data dokumentationen):

Fra:

..... Nitten personer (17 mænd og 2 kvinder) havde ikke oplyst alder og postnummer for bopæl, disse udgår og analyserne er derfor udført på grundlag af de resterende 4057 personer.

Til:

..... Nitten personer (17 mænd og 2 kvinder) havde hverken oplyst alder eller postnummer for bopælen, desuden havde 30 personer oplyst bopæl i udlandet. Disse udgår og analyserne er derfor udført på grundlag af de resterende 4027 personer. Blandt disse havde to personer oplyst et ugyldigt postnummer, dette blev rekodet til uoplyst postnummer.

Og da vi skal have omsat postnummeret til en afstand i kilometer yderligere.

....., dette blev rekodet til uoplyst postnummer
For hver af de 4027 personer blev der dannet en variabel med afstand til løbsstedet. Afstanden blev udmålt som antal centimeter på Post Danmarks kort over postnumre fra midtpunktet mellem Fredericia og Middelfart til det sted på kortet, hvor midtpunktet af det pågældende postnummer var. Da løbet blev gennemført mens der skulle sejles fra Nyborg til Korsør er der for postnumre øst for Storebælt tillagt en afstand således at denne svarer til en ekstra rejsetid på en time (ca 90 km).

For at udføre dette skal der nu beregnes en afstandsvariabel ud fra postnummeret. Du kan finde en kopi af disse linier i den fil, der hedder *postnr.do*

```
generate km = postnr
recode km 1138/2770 = 28.5
recode km 2791/2990 = 28.5
..... etc for alle de øvrige postnumre
```

Herefter kan vi kontrollere (kun et uddrag er vist)

```
. list id km pnr postnr
      id      km      pnr      postnr
  1.   2880      0      5500      5500
  2.   .....
  5.   3561      .
  8.   1053      0      ..... output udeladt
. tab km
      km |      Freq.      Percent      Cum.
-----+-----
      0 |      403      11.13      11.13
      4953 |      1      0.03      100.00
-----+-----
      Total |      3621      100.00

. set display 1 129
. list id pnr postnr km if km > 500, nodisplay
      id      pnr      postnr      km
  5.   3561      .
2664.  1028      0812      812      4953
3623.  1014      .
```

Fra det nederste ser du to ting. For det første opfattes uoplyst (værdien punktum .) som et stort tal ellers ville disse personer jo ikke blive listet. og for det andet er der desværre endnu et problematisk pnr, nemlig for **id** nr 1028. Dette må rettes til uoplyst med kommandoen:

```
replace km = . if postnr == 812
replace kmgrp = . if postnr == 812
```

Vi må så rette i dokumentationen, således at der er tre personer, som har fået rekodet til uoplyst. Hvilke **id** numre er det ?:

..... Nitten personer (17 mænd og 2 kvinder) havde hverken oplyst alder eller postnummer for bopælen, desuden havde 30 personer oplyst bopæl i udlandet. Disse udgår og analyserne er derfor udført på grundlag af de resterende 4027 personer. Blandt disse havde tre personer oplyst et ugyldigt postnummer, dette blev rekodet til uoplyst postnummer.

Øvelse 62 Beregn afstand og aftegn en frekvenstabel over grupperet afstand.

Se på den ”do” fil, som hedder *postnr.do*. Udfør den sammen med den du selv har udarbejdet ovenfor og reproducer denne tabel, prøv både med og uden ”missing”:

tab kmgrp				tabulate kmgrp, missing			
Residence of runner	Freq.	Percent	Cum	Residence of runner	Freq.	Percent	Cum.
0- 25 km	675	18.65	18.65	0- 25 km	675	16.76	16.76
26-62 km	911	25.17	43.81	26-62 km	911	22.62	39.38
63 -120 km	954	26.35	70.17	63 -120 km	954	23.69	63.07
120+ km	1080	29.83	100.00	120+ km	1080	26.82	89.89
				.	407	10.11	100.00

Total 3620 100.00	-----+----- Total 4027 100.00
---------------------	------------------------------------

Øvelse 63 Beregn alder, aldersgrupper og decimaltid

På tilsvarende vis vil vi nu oparbejde de resterende variable, nemlig køn, alder og løbstiden. Og til sidst gemme datafilen med **save**. (disse linier findes i filen **bromar.do**)

```

generate age = 96-age
egen agegrp = cutl(age) break(10(10)100) interval
label var agegrp "Age (10 year intervals)"
label var age "Age (1996-year of birth)"
notes age: Age calculated as 96-the year given as year of birth.
codebook age agegrp
notes
* tiden skal ændres
gen time=int(runtime/10000)
gen minute=int((runtime-time*10000)/100)
list id-runtime time minute in 1/10
gen dectime = time + minute/60
egen decgrp = cutl(dectime) group(4) interval
gra age dectime
drop runtime pnr postnr born time minut
*endelig beregnes en variabel der angiver gennemførelse af løbet
generate runok = cond(runtime != . , 1,0)
label var runok "Marathon race completed ?"
label define yn 0 "No" 1 "Yes"
label value runok yn
compress
label data "Marathon data - 1995 across bridges btw Funen and West-Jutland"
describe
save bromaraton, replace

```

Dokumentationen færdiggøres.

Nitten personer (17 mænd og 2 kvinder) havde hverken oplyst alder eller postnummer for bopælen, desuden havde 30 personer oplyst bopæl i udlandet. Disse 17+30 udgår og analyserne er derfor udført på grundlag af de resterende 4027 personer. Blandt disse havde tre personer oplyst et ugyldigt postnummer, dette blev rekodet til uoplyst postnummer. For to personer fandtes alderen udenfor det oplyste fra løbsledelsen og disse to blev omkodet til uoplyst (id 3404 og id 1605)

Data er gemt i disse filer:

1. rådata : bromraw.txt (xxx linier), dato, størrelse

oparbejdet fil: bromar.dta (N= xxxx) (Størrelse: yyyy) Dato(ddddddd)

anvendte do filer: postnr.do (dato og størrelse), bromar.do (dato og størrelse)

Oparbejdningen foretaget af: NN og MN

Anvendt software: Stata version 6.0 (opdateret til 21.1.2000). EpiData v. 1.x.

Originaldokumenter skal arkiveres indtil dddd mm aaaa. Evt. henvendelse forinden til : xxxxx

Data er afleveret i kopi til Dansk Data Arkiv/Enheden for registreringen af Sundhedsvidenskabelige Data under arkiv nr yyyyyy. Data kan genudleveres med disse restriktioner: osv.

Forskning og publicering: Ingen, forudsat der indgås samarbejdsaftale.

Undervisning: Ingen restriktioner

Eventuel henvendelse om udlevering skal ske til: NN

Underskrevet d_____ nn:_____

Øvelse 64 Dokumentation, anvendte filer og beslutninger samles

Du er nu klar til at arkivere den oparbejdede fil og dokumentationen. Saml følgende sammen

1. En zip fil med følgende filer: rå data fil, oparbejdet fil og anvendte do filer, samt oparbejdningensdokumentet
2. En zip fil med dokumenter i øvrigt: publiceringsaftaler, samarbejdsaftaler, projektbeskrivelse, etisk komite ansøgning, budgetter etc.
3. Projektbeskrivelse i papirkopi og oparbejdningensdokumentet i papirkopi.

Dette kan afleveres til samarbejdspartnere og evt. arkivering. Herefter kan første abstrakt skrives ud fra analyserne.

Bortfaldsanalyse

Der er nu en færdig analysefil. Første del af analysen vil derfor være at vurdere eventuel selektionsbias i data. Dette kan gøres ved at se på løbstiden blandt de der har oplyst alder i forhold til de der ikke har. Tilsvarende løbstid hos de der har oplyst adresse i forhold til resten. På samme måde kan du se på aldersfordeling blandt de der har oplyst postnummer i forhold til aldersfordeling blandt de der ikke har oplyst dette.

Øvelse 65 Udfør en bortfaldsanalyse

Gem de kommandoer du vælger i en fil med navnet *bortfald.do* Du bør kigge nærmere på nogle scatter og boxplot, stratificerede plot, krydstabeller og teste forskelle med et parametrisk (middelværdi) eller non-parametrisk test (rangordning).

Endelig besvarelse af hypoteser fra Bromaraton løbet**Øvelse 66 Belysning af de opstillede hypoteser**

Gem de kommandoer du vælger i en fil med navnet *analyse.do* Du bør kigge nærmere på nogle scatter og boxplot, stratificerede plot, krydstabeller og teste forskelle. Bestem ud fra graferne hvilke former for test der er relevante. (Parametrisk (middelværdi) eller non-parametrisk test (rangordning), eller skal der mere komplicerede metoder til efterfølgende (regression).

Nyhedsgruppe på internet for Stata og www.stata.com.

Der findes en nyhedsgruppe på internet, hvor forskellige spørgsmål om Stata diskuteres. Det kan være nyttigt at læse de kommentarer som gives. Fx har der været diskussioner om hvordan der i analyserne bedst skelnes mellem confounding og effect modifikation, samt hvilke analysestrategier der er velegnede i bestemte situationer. Niveaueet af korrespondancen er meget forskellig. Lige fra helt simple spørgsmål som ”hvorfør vil dette program ikke gøre” til ”hvordan skal jeg tolke og vælge mellem de estimerer, som denne xxxx statistiske model har givet i forhold til denne yyyy statistiske model.

Jævnligt sender brugere udbygninger af Stata til nyhedsgruppen. Fx et særligt statistisk test eller en ny statistisk model, der kan estimeres. De bedste – mest generelle - af disse udgives i det delvis uafhængige tidsskrift STB – Stata Technical Bulletin, som man kan abonnere på eller købe hver gang de er samlet til et hæfte (ca 1 gang hvert andet år). I disse findes en gennemgang af en række analyser, hvor både eksempler, det konkrete output og forbehold, samt referencer for en given analyse er nævnt.

Øvelse 67 Opslag på Stata’s internetadresse (www.Stata.com).

Kig på FAQ listen

Hvornår er den seneste udgave af STB udkommet ?. Hvad er STB ?

Tip: start en web browser (netscape, internet explorer, opera) og kig på www.Stata.com for at finde svarene (brug evt. søgefunktionerne). Bøger findes under *bookshop*. Hjælp under *User Support*.

Tilmelding til nyhedsgruppen

Sker ved at sende et almindeligt e-post brev med én linie i. Modtager skal være :

majordomo@hsphsun2.harvard.edu og i selve brevet - ikke “emne/subject”:

subscribe Statalist-digest ”email address”

Skriv din egen e-post adresse i stedet for “email adress”. Du modtager herefter et automatisk brev med svar om hvordan du melder dig fra listen igen (der kommer et brev pr dag !!).

Hvis du ikke er tilmeldt kan du søge i de gamle meddelelser. Find arkiverne på www.stata.com under Stata Listserver, hvor følgende arkiver nævnes:

Statalist Archives

- [Archives at Harvard University](#), providers of Statalist
- [Archives provided by findMail Communications](#)
- [Archives in Portugal](#), provided by Carlos Ramalheira, M.D.

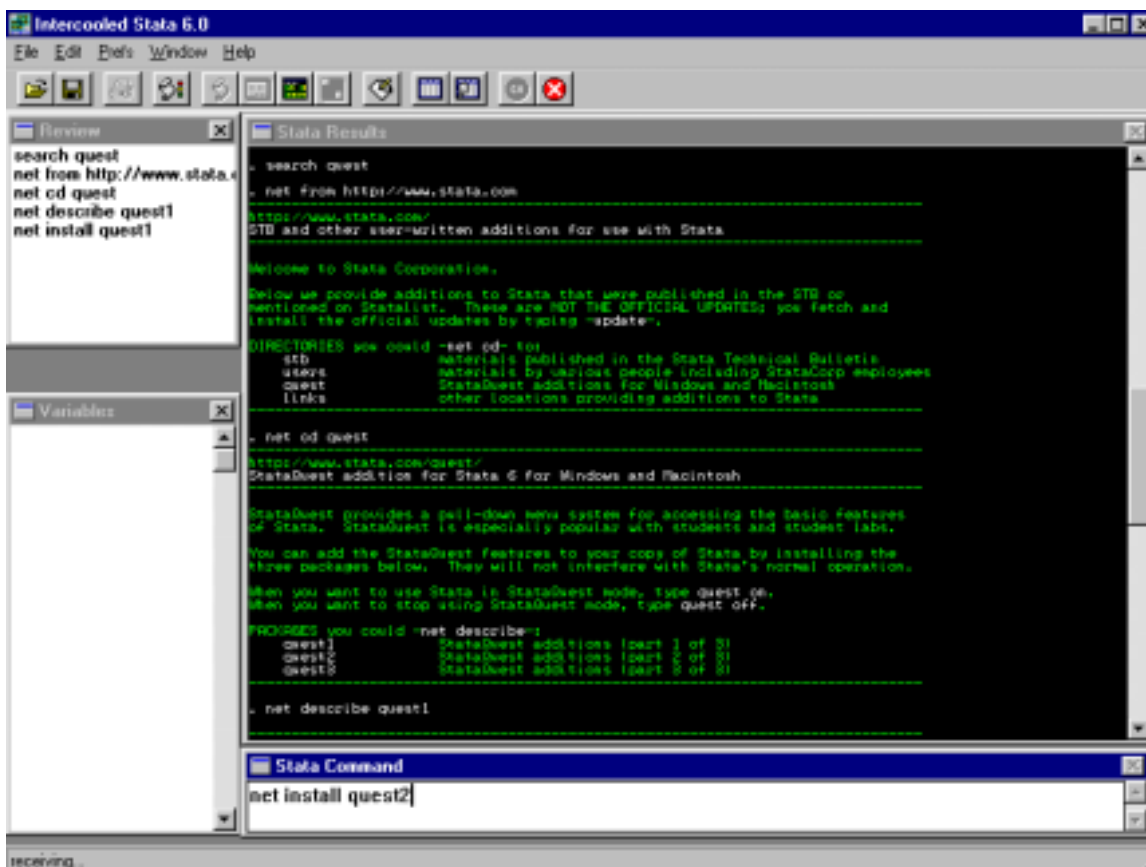
These sites provide archives of all postings to Statalist. The Harvard and FINDMAIL sites allow direct access to individual postings and can be ordered by date, author, or subject.

Installering af Stata, Stat/Transfer og Quest

Øvelse 68 Installér Stata og Stat/Transfer

På din egen PC bør du selv kunne installere Stata forfra. Ellers vil du komme i tvivl om dette på et tidspunkt. Se i Stata Users Guide under installering, samt side 70.

1. Anskaf en Stata licens og installér programmet ud fra den dokumentation der følger med. Programmet opretter en gruppe i din windows programflade, der sædvanligvis indeholder "intercooled Stata". Første gang du starter programmet bliver du bedt om indtaste licensnummer og en særlig kode.
2. Installér Stat/Transfer, således at du kan oversætte data til og fra forskellige dataformater
3. Det er en god ide, at opsætte windows-95 menu systemet, således at det er let at arbejde med forskellige projekter/artikler. Se side 70.
4. Som nybegynder er det en god ide, også at installere quest, som er en menuoverbygning på Stata. Hvis din PC er koblet til internet kan Stata version 6 installere quest direkte fra www.Stata.com ved at du udfører disse kommandoer²¹:



```

Intercooled Stata 6.0
File Edit Preferences Window Help
[Icons]
[Review] search quest
net from http://www.stata.com
net od quest
net describe quest1
net install quest1

[Variables]

[Stata Results]
. search quest
. net from http://www.stata.com
-----
http://www.stata.com/
STB and other user-written additions for use with Stata
-----
Welcome to Stata Corporation.

Below we provide additions to Stata that were published in the STB or
mentioned on Statalist. These are NOT THE OFFICIAL UPDATES; you fetch and
install the official updates by typing "update".

DIRECTORIES you could -net od- to:
stb      materials published in the Stata Technical Bulletin
users   materials by various people including StataCorp employees
quest   StataQuest additions for Windows and Macintosh
links   other locations providing additions to Stata
-----
. net od quest
-----
http://www.stata.com/quest/
StataQuest addition for Stata 6 for Windows and Macintosh
-----
StataQuest provides a pull-down menu system for accessing the basic features
of Stata. StataQuest is especially popular with students and student labs.

You can add the StataQuest features to your copy of Stata by installing the
three packages below. They will not interfere with Stata's normal operation.

When you want to use Stata in StataQuest mode, type quest on.
When you want to stop using StataQuest mode, type quest off.

PROGRESS you could -net describe-:
quest1  StataQuest additions (part 1) of 31
quest2  StataQuest additions (part 2) of 31
quest3  StataQuest additions (part 3) of 31
-----
. net describe quest1

[Stata Command]
net install quest2

```

```

net install quest1
net install quest2
net install quest3.

```

Se nærmere om opdatering ved at slå op i hjælpesystemet med **help update** og **help net**

²¹ Se i næste afsnit hvis du ikke er koblet til internet fra din PC

Supplerende opsætning: Tilrettet opstart af Stata og egenskaber.

Øvelse 69-s Speciel opstart af Stata (ændring af menuen).

Når du er erfaren kan det være en god ide, at opbygge egne menuer. Hvis du gemmer en fil med navnet *profile.do* og angiver den ved opstarten af Stata vil alle kommandoer deri blive udført. Du bestemmer selv indholdet, her har jeg ønsket at definere:

- Tilføj editoren **metapad** som et menupunkt under "edit"
- Funktionstasten **F2** skal udføre ordren **discard**
- Quest on** starter quest systemet fra starten.
- linielængden når der ses resultater i outputvinduet skal være 120 (set d l 120)
- linielængden i logfiler skal være 120 (set log l 120)
- med **set matsize** kan visse programmer være større end standard
- Med linien "**qui{**" først og afsluttende **}"** undgår jeg at se udførelse af linierne når Stata startes.

Indhold af profile.do:

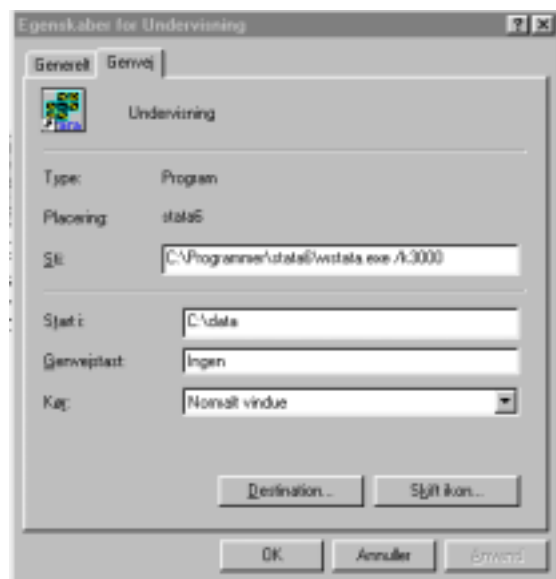
```
qui {
window menu append string "Edit" "&M metapad" "winexec \program\meta\metapad20.exe"
global F2 discard
set d l 120
set log l 120
set matsize 150
}
```

Menuen ser sådan ud:

Sæt egenskaber sådan i din opstarts kommando for Stata: (den ikon du bruger). Hvis du er i tvivl om hvad dette betyder, så læs nærmere i Windows hjælpefil om opsætning af programmer.

med **/k3000** gives programmet som standard 3Mb ram at arbejde med.

Her er det valgt, at der som standard skal arbejdes i **c:\data**



Opdatering og tilføjelser til STATA

De grundlæggende dele af dette blev gennemgået i Øvelse 24. Men her gengives en udvidet udgave.

Nye versioner af stata udgives med cirka 2 – 2½ års mellemrum. Hvis den version du har opfylder dine formål til analyser er der **ingen** grund til at få en ny version. Men så længe din nuværende version er den aktuelt nyeste er det en god ide 3-4 gange årligt at hente den nyeste udgave. Udgave 7 udkommer nok ikke før medio 2001 eller 2002.

Der er to måder at opdatere på afhængig af om du er koblet til internet eller ej:

Hvis du er koblet til internet

Øvelse 70 Opdatering og tilføjelser til Stata version 6 direkte fra internet.

Når du skal se status på din egen installation af stata, skriver du blot `update` hvorefter du får en slags statusrapport. Derefter kan du bestemme om du ønsker at opdatere hele systemet, ado systemet eller udvalgte dele.

```
update
```

Stata executable

```
folder:      C:\PROGRAMMER\STATA6\  
name of file:  wstata.exe  
currently installed: 19 Jan 1999
```

Ado-file updates

```
folder:      C:\PROGRAMMER\STATA6\ado\updates\  
names of files: (various)  
currently installed: 12 Jan 1999
```

Recommendation

```
Type -update query- to compare these dates with what is available from  
http://www.stata.com.
```

Prøv også ”news” og ”help whatsnew”

Hvis du ikke er på internet henter du nogle filer fra internet på en anden PC end den daglige, sådan som det er vist på næste side.

Øvelse 71 Hent opdateringsfiler til Stata på internet.

Er du ikke koblet til internet gør du sådan her:

Find en PC, der er koblet til internet og start en internet browser. Kig på "www.stata.com/support".

Her vil du finde relevant information om opdatering og yderligere hjælpe muligheder.



Du skal nu hente to filer:

1. Peg på <http://www.stata.com/support/updates/> "Stata executables" og download *wstata.bin*. Print installeringsvejledning ud eller gem den som en fil.
2. Peg dernæst på "Stata ado-files" og download *ado.zip*. Print installeringsvejledning ud eller gem den som en fil.

This rest of this page is intended for users of Stata 6 for Macintosh 68K, Windows 3.1 connections on their machines. If you are using Stata 6 for Power Macintosh, Windows direct Internet connection, type **update query** in Stata and follow the recommended Reference Manual for more information.

Update	Platforms	Date
Stata executables	Stata 6 for Windows 50/50/NT	22 Jun 1999
	Stata 6 for Windows 3.1	22 Jun 1999
	Stata 6 for Macintosh	22 Jun 1999
	Stata 6 for Unix	22 Jun 1999
Stata ado-files	All operating systems	28 Jun 1999

There are two components to Stata:

- the Stata executable
- the Stata ado-files

Øvelse 72 Hent Quest systemet.

1. Peg på "http://www.stata.com/support/quest/"
2. Find "StataQuest for Stata 6" og punktet "Manual download". Download *quest.zip*.

Kopier de tre filer *wstata.bin*, *ado.zip* og *quest.zip* til disketter og tag disse med hen til PC'en uden internetforbindelse.

Platform	Filename	Size	Instructions
Windows	ado.zip	611k	Windows Installation Instructions

The type of the file you just tried to open is unknown to Opera. Choose a program to show the file, save the file or cancel the transfer.

Type: application/zip

Show

Save... Use Opera Associate... Cancel Help

Øvelse 73 Installér den nye Wstata.bin, som er hentet fra internet.

Wstata.bin installeres på en bestemt måde, gør præcis sådan her:

- A. Kopiér filen *wstata.bin* til der hvor Stata er installeret: (fx *c:\stata* eller *c:\program\stata*)
- B. Omdøb filen *Wstata.exe* til *wstata.old* (slet en evt. eksisterende *wstata.old*)
- C. Omdøb filen *Wstata.bin* til *wstata.exe*
- D. Afprøv stata ved at starte det igen

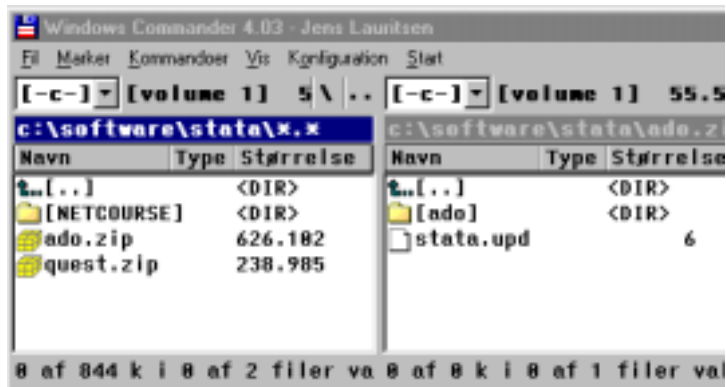
Øvelse 74 Installér ny udgave af ADO systemet, som er hentet fra internet.

De nyeste **ado** filer og **quest** installeres på en anden måde (her vist for **ado.zip**)

Det første du skal gøre er at pakke **ado.zip** ud på den rigtige måde, sådan at den kan installeres i Stata. Det vises her med **Win-Commander**, men kunne også udføres med et andet "Unzip" program.

Ved udpakning **skal** en mappe struktur (path i zip filen) bevares !! . Her kan du se, at der er én "undermappe" **ado**

Her vises de downloadede filer til venstre og indholdet i **ado.zip** til højre

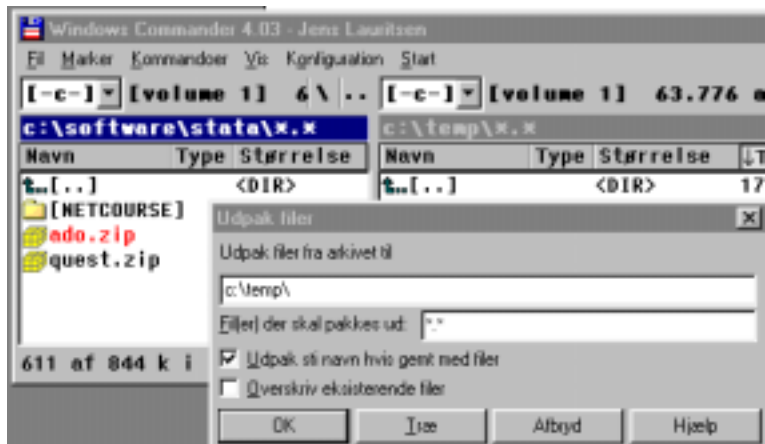


Gør nu følgende:

a. Skift højre skærm til fx. **c:\temp**. Peg på **ado.zip** til **venstre** således at filen er markeret og dernæst på **Alt+F U** (fil, udpak/file, unpack) og skærmen vil nu se sådan ud:

b. For at bevare mappe strukturen er det afgørende, at der markeret i "sti/path names", ellers pakkes det hele ud i en bunke. Og det vil ikke fungere.". Når du trykker "Ok" pakkes filerne ud.

c. Du er nu klar til at starte Stata og **opdatere** fra **c:\temp**. Prøv følgende kommandoer:



update from c:\temp
 update ado
 update query

d. Når det er gået godt kan du slette filerne i **c:\temp** og **ado.zip**.

Øvelse 75 Installér quest.zip, som er hentet fra internet.

7. Quest systemet. Pak **quest.zip** ud på samme. Bemærk også her, at zip filen har en mappestruktur.

Quest systemet er **ikke** en opdatering, men et tillægsmodul derfor skal du i stedet for "update from" skrive "net from":

net from c:\temp
 net install quest1
 net install quest2
 net install quest2
 quest on



Når det er gået godt kan du slette filerne i **c:\temp** og **quest.zip**.

Øvelse 76 Hent STB udvidelser som zip filer og installér fra harddisken.

STB tillægsmoduler findes via:

<http://www.stata.com/info/products/stb/stbftp.html>

Du skal finde det punkt, som hedder:

Installing the software from other Internet sources

Du ledes frem til ét af flere steder, hvorfra du kan hente stbxxvy.zip filer. xx står for nummeret og y for version af stata. Man kan altid hente og installere ældre versioner end den man har, men ikke nyere.

Se fx i figuren, hvordan stb49 er hentet.

Når zip filen er hentet gør du præcis som vist i

Tilføjelse af rutiner direkte fra internet.

I modsætning til øvelserne ovenfor kan du hente både tillægsmoduler fra STB systemet og opdateringer direkte fra Stata, hvis du er koblet til internet. Er du koblet direkte kan du skrive nogle kommander, der som den første starter med **net** (herefter vil din PC automatisk slå op med hjælp af Stata på www.stata.com inde fra Stata). Du guides nu rundt i et system, hvor man kan vælge videre med fx **net cd stb net link users**.

Øvelse 77 Tilføjelse af rutiner direkte fra internet.

En anden mulighed er at gå de samme veje direkte fra hjælpesystemet. Hvis du fx. i

hjælpesystemet skriver **search etab**, vises følgende: Beskeden fra "search etab" er, at der i STB-10 (Stata **T**echnical **B**ulletin) i mappen **sg12** findes to rutiner, hvoraf den ene hedder **etab**.

Hvis du vil læse mere om rutinen skal du have fat i **STB Reprints Vol 2, 140-41**.

Uanset om du er på internet eller ej er det første hjælpebillede det samme. Men er du på internet direkte kan du nu **click'e** på

"**click here to install**", hvorefter rutinen installeres. Ellers må du hente en zip fil, som forklaret ovenfor. STB Reprints er nødvendige for at få præcis besked om forbehold mm. om en bestemt rutine.



The best way for you to access STB programs is to use th

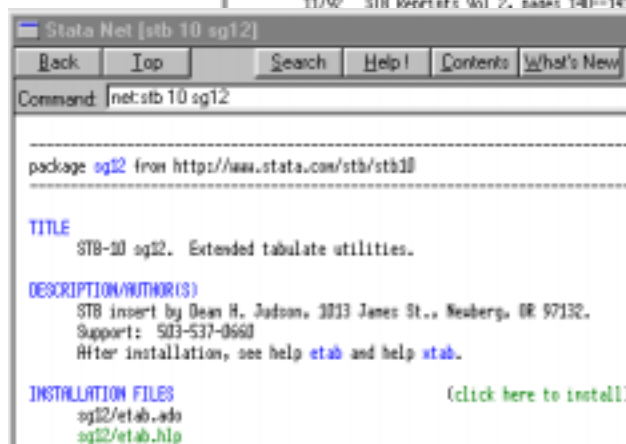
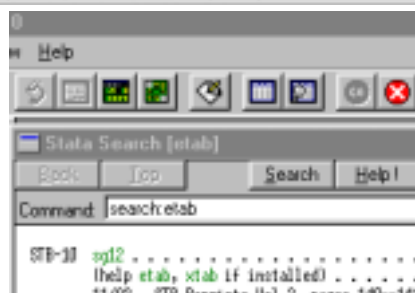
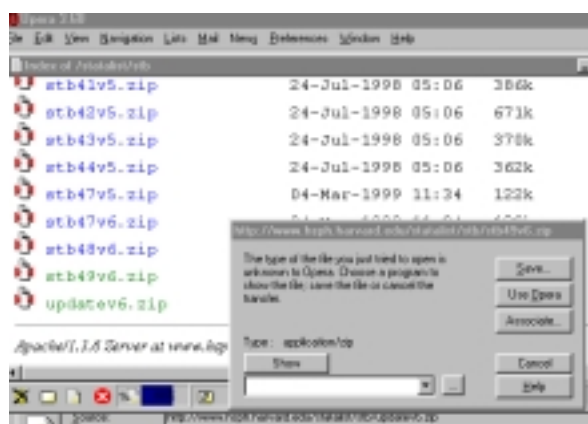
```
. net from http://www.stata.com
. net cd stb
```

and follow the instructions. Or,

1. In Stata, Pull down **Help** and select **STB and User-**
2. Click on <http://www.stata.com>.
3. Click on **stb**.

See [Installing the software from www.stata.com](#) in the ST see [Installing the software from other Internet sources](#).

If you use Stata 5 (an old release), click [here](#).



Eksempel på svar fra FAQ listen på www.Stata.com

Faq: Frequently Asked Questions (Uddrag af en enkelt)

What are some of the problems with stepwise regression?

Author Bill Sribney, Stata Corporation. Date May 1998

Frank Harrell's comments:

Here are some of the problems with stepwise variable selection. It yields R-squared values that are badly biased high. The F and chi-squared tests quoted next to each variable on the printout do not have the claimed distribution. The method yields confidence intervals for effects and predicted values that are falsely narrow (See Altman and Anderson, Statistics in Medicine). It yields P-values that do not have the proper meaning and the proper correction for them is a very difficult problem. It gives biased regression coefficients that need shrinkage (the coefficients for remaining variables are too large; see Tibshirani, 1996). It has severe problems in the presence of collinearity. It is based on methods (e.g. F tests for nested models) that were intended to be used to test prespecified hypotheses. Increasing the sample size doesn't help very much (see Derksen and Keselman). It allows us to not think about the problem. It uses a lot of paper. Note that "all possible subsets" regression does not solve any of these problems.

References

Altman, D. G. and P. K. Andersen. 1989. Bootstrap investigation of the stability of a Cox regression model. *Statistics in Medicine* 8: 771-783.

Shows that stepwise methods yields confidence limits that are far too narrow.

Derksen, S. and H. J. Keselman. 1992. Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology* 45: 265-282.

Conclusions

"The degree of correlation between the predictor variables affected the frequency with which authentic predictor variables found their way into the final model." "The number of candidate predictor variables affected the number of noise variables that gained entry to the model."

"The size of the sample was of little practical importance in determining the number of authentic variables contained in the final model." "The population multiple coefficient of determination could be faithfully estimated by adopting a statistic that is adjusted by the total number of candidate predictor variables rather than the number of variables in the final model."

Roecker, Ellen B. 1991. Prediction error and its estimation for subset--selected models. *Technometrics* 33: 459-468.

Shows that all-possible regression can yield models that are "too small".

Mantel, Nathan. 1970. Why stepdown procedures in variable selection. *Technometrics* 12: 621-625.

Hurvich, C. M. and C. L. Tsai. 1990. The impact of model selection on inference in linear regression. *American Statistician* 44: 214-217.

Copas, J. B. 1983. Regression, prediction and shrinkage (with discussion). *Journal of the Royal Statistical Society B* 45: 311-354.

Tibshirani, Robert. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B* 58: 267-288.

Ronan Conroy's comments: Personally, I would no more let an automatic routine select my model than I would let some best-fit procedure pack my suitcase.

Se resten på internet FAQ listen.

Hjælpeprogrammer

Det er min erfaring²², at mange personer har store problemer med at finde en god og konkret arbejdsform ved dataanalyse og håndtering af data. Og at en del af de samme personer har lige så store problemer med helt konkrete forhold såsom at oprette en god projektstruktur på harddisken, kopiere filer til disketter, installere nye programmer, pakke filer sammen som er for store til en diskette mm. Derfor er der en tendens til at skælde ud på computer, programmer og instruktionsmateriale i stedet for at afsætte den fornødne tid til at lære at bruge disse som værktøj. Nogle af disse problemer skyldes utilstrækkelig erfaring, men også at de nyere edb styresystemer prøver at overtage styringen, således at du ikke ved, hvad maskinen egentlig gør²³.

I den arbejdsmåde jeg foreslår indgår følgende programmer (udover Stata, som udfører resten), der tjener hvert sit formål:

navn	funktion	begrundelse for at bruge det ?
Win Commander (alternativ til stifinder/explorer)	Kopier filer, pak filer sammen, pak filer ud mm	Forenkler arbejdet med kopiering af filer, pakning af filer, find filer mm. Gør det lettere at undvære mus til arbejdet.
Stat/Transfer	konvertering af data fra database format eller indtastningsprogram	Er nyttigt især ved samarbejde med andre som anvender andre data formater
Tekstbehandlings- program	udarbejde publikationer	Her er det nemt at formattere tekst og tabeller på en brugbar umiddelbart publicerbar måde.
Programmers file editor (Pfe)	Rette i log filer og mere komplekse”do” filer, print af resultater	Hvorfor flyve til København fra Odense i en Boeing 747, når kun én person skal flyttes (Iflg. M.Hills) Brug et simpelt program til output og rutiner.
Andre nyttige programmer	dias fremstilling, litteraturstyring mm	Omtales ikke i denne note, men indgår også i arbejdet.

For at bruge hjælpeprogrammerne *Win Commander* og *Programmers File Editor* skal de først installeres på PC'en. Udfør derfor nu Øvelse 78 til Øvelse 81

Du kan evt. også udføre Øvelse 83 og Øvelse 84.

²² 8-10 års undervisning i dataanalyse.

²³ Brugen af disse hjælpeprogrammer er **ikke** en forudsætning for at anvende Stata. Prøv som vist i øvelserne og beslut dig derefter hvordan du vil arbejde.

Installér Hjælpeprogrammer

Som nævnt ovenfor kan det være nyttigt at have godt kendskab til nogle få hjælpeprogrammer. Se side 59.

Første hjælpeprogram: Win-Commander. Dette program erstatter eller supplerer stifinder (engelsk windows: explorer) og giver efter min mening en lettere måde at arbejde med med pakkede filer.

Øvelse 78-s Hent win-commander fra internettet:

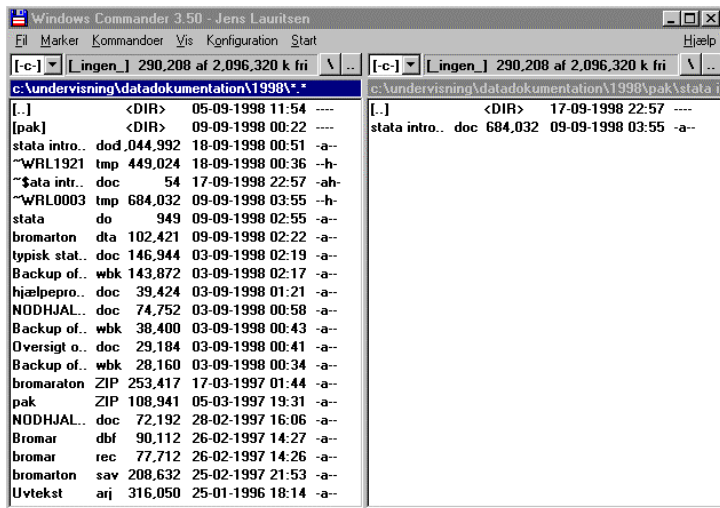
1. Opret en mappe på pc'en til at lægge filerne i. Hvis du vil gøre det inde fra Stata, så skriv: `mkdir c:\software` ↵ og `mkdir c:\software\wincmd` ↵
2. Start en internet browser²⁴ og find: www.ghisler.com
 - 2.1. Hent programmet wincommander (wincmd) i seneste 32 bits udgave
wincmd352.zip²⁵ gemmes i c:\software\wincmd
3. Undersøg om du har et udpakningsprogram på PC'en ved at pege på den modtagne fil i windows stifinder/explorer. Hvis du ikke har, kan du kigge i www.ghisler.com under "Add Ons" og hente et program, der kan pakke ud (Zip/unzip).


Øvelse 79-s Installér Win-Commander

Udpakning. Pak wincmd352.zip ud i c:\software\wincmd\

Installér herefter win commander ved at køre: c:\software\wincmd\install.

Installér fx til c:\programmer\wincmd



Vælg dansk som sprog og gerne også engelsk. Når win commander er installeret er der også oprettet en gruppe med Wincmd. Find den ved at trykke på  efterfulgt af "p" og start programmet. Du vil nu se et skærmbillede, der *delvist* ser ud som på billedet²⁶. Opsætningen på figuren er ændret, så der er et simpelt skærmbillede uden ikoner, blokke mm.


Du har nu et hjælpeprogram, der kan kopiere filer, vise indhold, pakke og udpakke filarkiver ("Zip filer"). Zip filer vises på samme måde som en almindelig mappe. Der er et vindue til højre og venstre, hvor der i hvert gemmer sig indholdet af en mappe.

²⁴ Netscape, Internet Explorer eller Opera (Opera findes på www.operasoftware.com og fylder kun én diskette !)

²⁵ Navn på filen følger de forskellige versioner. 352 betyder version 3.52.

²⁶ Hvis du vil blive ved med at bruge programmet kan du registrere programmet og får derefter tilsendt et licensnummer. Hvilket betyder, at dit navn vises øverst, samt at du slipper for at trykke på 1-2 eller 3 ved starten af programmet.

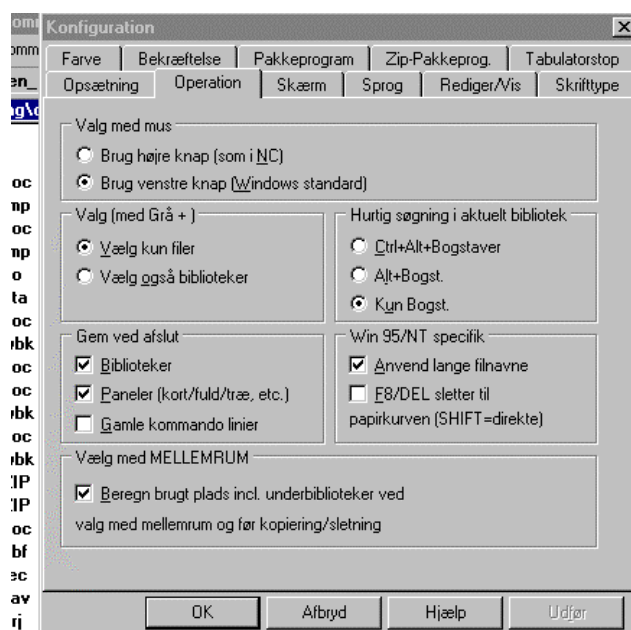
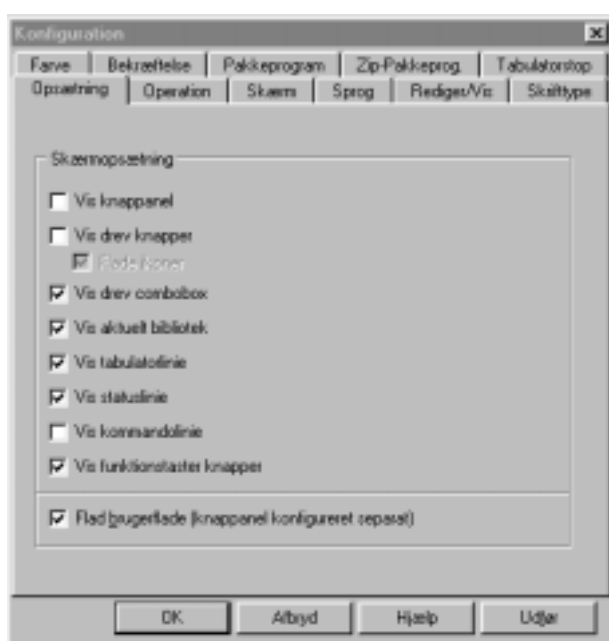
Øvelse 80-s Afprøv Win-Commander

Start programmet (tryk  og find mappen Win-Commander). Prøv de forskellige muligheder under **Alt+v** (vis). Skift mellem forskellige drev A/C/S²⁷ eller hvad du har.

- Find filen *wincmd32.exe* ved at søge på C-drevet (Find: **Alt+F7** eller **Alt+K S** (kommando) og indsæt i startmenuen, som vist i Øvelse 93.
- Hvor meget ledig plads er der på C-drevet på din PC ?

Øvelse 81-s Opsætning af Win-Commander – 1 (grundlæggende).

Aktivér nu opsætning af wincommander (konfiguration **Alt+o**) og sæt opsætningen



som vist nedenfor²⁸. Giver dette et andet skærbillede end vist i foregående figur ?
Hvad blev ændret i forhold til udseendet, da du startede programmet første gang ?

Tast derefter **Udfør** og **OK**. Du kan nu under **Alt+v** (vis) prøve at angive kort, fuld, bibliotekstræ visning indtil du får det udseende, du gerne vil have. Det er en god ide, at sortere filerne i omvendt kronologisk rækkefølge, så har du altid vist de nyeste filer øverst. Sorteringsorden finder du under **Alt+v** (vis). Kronologisk kan angives direkte ved **Ctrl+F5**

En anden nyttig tastaturkombination er **Ctrl+R**, som betyder ”læs igen”. Hvis du fx skal finde en fil på en diskette og du har filen på én af tre disketter. Sæt disketterne i A drevet én ad gangen. Tryk på **Alt+F2** A, er filen ikke på den første, sæt den næste diskette i og tryk **Ctrl+R**

Andet hjælpeprogram: Editor er alternativer til ”do file editor” og ”windows notepad”. Det kan være nyttigt at anvende en ekstra Editor som er mere fleksibel og kan arbejde med store log filer eller ascii data filer. PFE (programmers file editor) kan læse filer på både 2 og 4 megabyte. Metapad er en anden editor, som er meget lille og hurtigere end notepad og pfe. Se <http://welcome.to/metapad/> (omtales ikke yderligere)

²⁷ Tastaturkombinationen for drev i venstre box er: **Alt+F1** for højre box: **Alt+F2**

²⁸ Vælg ”konfiguration” og ”no bar” under knappanel, hvis du vil undgå ”knappanel”

Øvelse 82-s Hent editor filerne fra internet:

1. Opret en mappe på pc'en til at lægge editorfilerne i. Hvis du vil gøre det inde fra Stata, så skriv: `mkdir c:\software\editor`
2. Hent filen pfe (programmers file editor) fra : www.lancs.ac.uk/people/cpaap/pfe Det skal være en 32 bit udgave. *pfe602i.zip*.
3. Du kan eventuelt også installere et skærmkopi program. Næsten alle skærmudsnit i disse noter er dannet med programmet "Otto's PrintScr 32" - freeware. Kan hentes fra ftp.gamma.dou i biblioteket pub\win95\utils\prtscr3215.zip.

Øvelse 83-s Installering af "Programmers file editor"²⁹

"Programmers file editor" skal blot pakkes ud, brug Win-Commander til dette:

- 1: Opret mappen: `c:\programmer\pfe` (peg på `c:\programmer` og tryk på **F7** og angiv `pfe`)
- 2: Peg på filerne fra *pfe602i.zip* (Første gang du starter win-commander og ser på en zip fil, kommer nogle spørgsmål. Svar blot ok til dette.) og pak dem ud enten via **F5** eller via **Alt+f** udpak filer til : `c:\programmer\pfe`

Øvelse 84-s Opsætning af Win-Commander – 2 (til brug af Pfe)

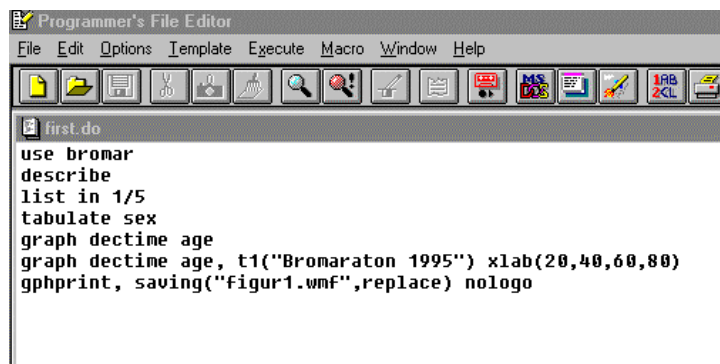
Nu skal du tilpasse win commander til at bruge **Pfe** . Aktivér punktet "konfiguration", - indstillinger. Vælg fanen "rediger/vis" og angiv nederst ved "**F4**" :

`\programmer\pfe\pfe32.exe`

Øvelse 85-s Afprøv Programmers File Editor - Pfe

I Win-Commander skal du nu skifte én af de to sider til `C:\data`. Dernæst skal du pege på filen *first.do*, som du gemte i

Øvelse 18. Hvis du trykker på tasten **F4**, mens markøren/cursoren står ud for *first.do* startes Pfe (pga tilpasningen af Win-Commander i Øvelse 84) og skærmen ser ud som vist.



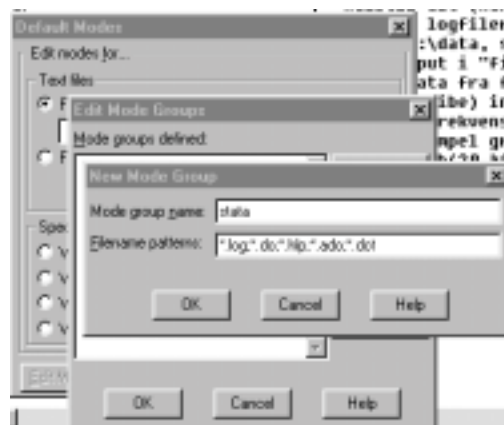
```

Programmer's File Editor
File Edit Options Template Execute Macro Window Help
first.do
use bromar
describe
list in 1/5
tabulate sex
graph dectime age
graph dectime age, t1("Bromaraton 1995") xlab(20,40,60,80)
gphprint, saving("figur1.wmf",replace) nologo

```

Hent på samme måde "*first.log*" ved at skifte til Win-Commander **Alt+tab**, flyt cursoren til filen "*first.log*" og tryk på **F4**. Herefter kan du i **Pfe**, se begge filer.

Det er nu let at rette i *first.do*, gemme filen (**Ctrl+S**), vende tilbage til Stata med **Alt+tab**, udføre den gemte fil ved at skrive **do first**, og dernæst læse resultatet i en graf eller en logfil.

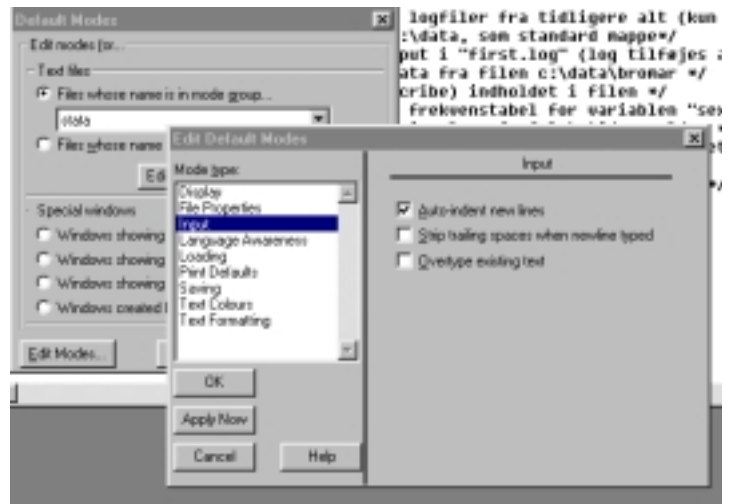
Øvelse 86-s Opsætning Pfe – 1 (stata filtyper vises)

²⁹ Du kan på samme måde installere skærbilledprogrammet printscr32, som er brugt til skærbillederne i denne note.

Under default modes skal du oprette en mode group med navnet stata, som vist til venstre. (Herved får du altid vist, de filer som hører til Stata, som kan rettes med Pfe)

Øvelse 87-s Opsætning Pfe – 3 (Standard tilpasning)

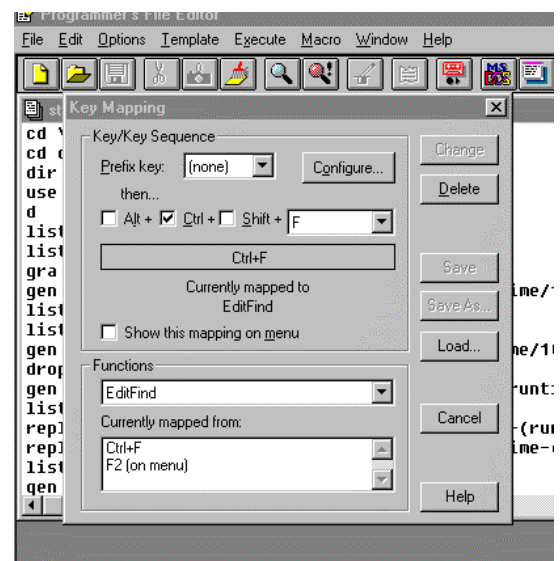
Under Edit Modes skal du sætte to kryds (mindst), ”Auto-indent New Lines” under ”Input” og ”Print page headers” under ”Print Defaults”



Øvelse 88-s Opsætning Pfe – 2 (tastatur)

Når pfe er startet kan du se, at der også her er en række muligheder foroven.

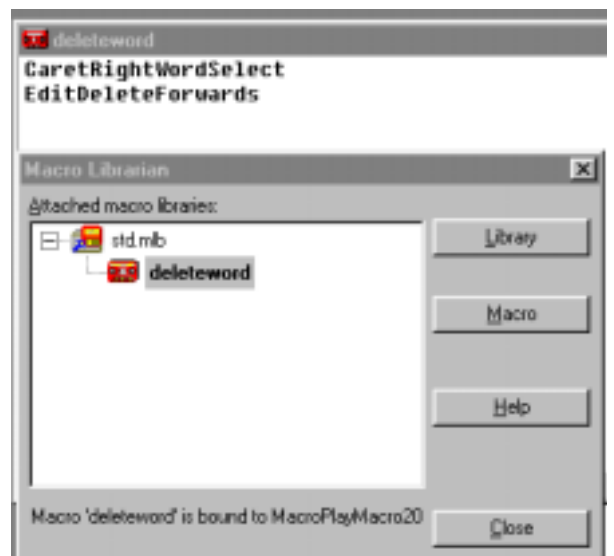
Under ”options” og underpunkt ”Keymapping” kan du definere egne keyboard shortcuts. Fx gør du sådan her for at få ”Søg” til at være knyttet til **Ctrl+f**. Sæt et kryds foran **Ctrl**. Dernæst finder du under ”functions” funktionen ”EditFind” ved at blade med ↓. Tilsvarende vil **Ctrl+H** kunne sættes til ”Søg og Erstat”, hvor du under functions angiver EditReplace



Øvelse 89-sa Opsætning Pfe – 4 (Avanceret tilpasning)

Hvis du vil have tasterne Ctrl+del til at slette et ord, skal du gøre følgende: Opret et macro bibliotek (Librarian, tast Alt+M) her gemt som **c:\programmer\pfe\std.mlb**

1. Opret en macro, her **deleteword** med to linier i som vist
2. Udfør under Macro et Bind til fx MacroPlayMacro20.
3. Knyt MacroPlayMacro20 til **Ctrl+Del** på samme måde som du brugte ved Ctrl+F i øvelsen ovenfor.



Backup og kopi af data

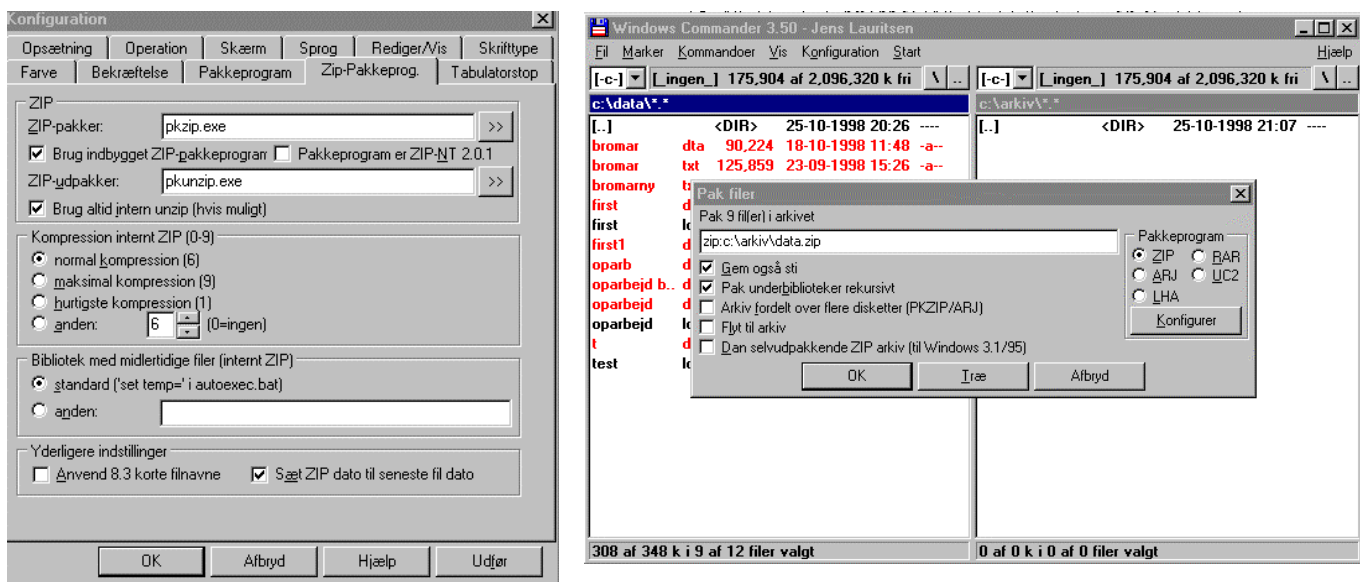
Regelmæssigt, fx én gang pr dag er det en god idé at tage en kopi af de nyeste arbejdsfiler, dvs ”do” og datafiler til diskette. Logfiler kan altid genskabes, hvis datafiler og ”do” filer er gemt.

I arbejdsintensive perioder er det desuden en god idé, at kopiere alle ”do” og datafiler til et reservested på PC’en regelmæssigt. Se foregående side inden disse øvelser udføres

Øvelse 90-s Pak en kopi af ”do” og datafiler til et arkiv på harddisken

Start Win-Commander og gør følgende:

1. Opret en mappe, der hedder `c:\arkiv` og skift højre side til dette.
2. Sørg for at venstre side er `c:\data`
3. Du skal nu vælge hvilke filer der skal med på én eller flere af disse måder:
 - 3.1. Peg på de ønskede filer og tryk på **Mellemrum**'s Tangenten.
 - 3.2. Peg på de ønskede filer og tryk på **Ins** Tasten.
 - 3.3. Alle ”do” filer: Tryk **Alt+M V** (eller **+** til højre) (Gruppe), skriv ”*.do”.
 - 3.4. Alle ”data” filer: Tryk **Alt+M V** (eller **+** til højre) (Gruppe), skriv ”*.dta”.
 - 3.5. Alle filer. Tryk **Alt+M G** (Alle filer) (eller tryk på **Ctrl+** til højre).
 - 3.6. En hel mappe kan vælges (Skift venstre side ét niveau op (**Home** ↵), så du til venstre har `c:\`, sæt cursor ud for `c:\data` og tryk på **Mellemrum**)
4. Tryk dernæst på **Alt+F P** og Win-Commander vil se ud som figuren til højre. Det er en god idé, at vælge konfiguration af Zip-Pakkeprog. se Øvelse 84, som vist til venstre.



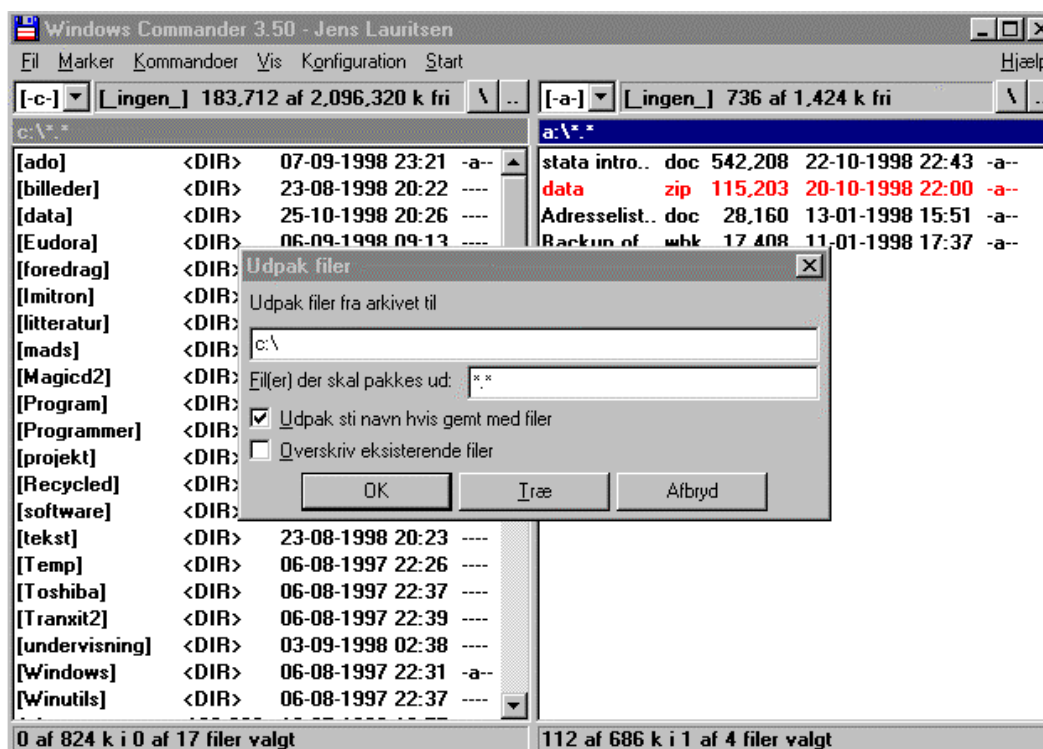
Øvelse 91-s Pak en kopi af ”do” og datafiler til en diskette

Gør som I foregående øvelse, men kopiér til diskette.

Øvelse 92-s Kopiér fra zip filen tilbage til c:\data

Skift på højre- og venstre side, så det ser ud som på figuren nedenfor.

- Peg på *data.zip* til højre og vis til venstre alle mapper i c:\ .
- Udpak filerne fra arkivet til c:\ og husk, at markere ”Udpak sti navn ...”



Hvad gør man når en pakket fil bliver for stor til en diskette ?

Man kan anskaffe en CD-rom brænder, et eksternt harddisk drev eller man kan dele filerne i flere stumper med Win-Commander eller et pakkeprogram. I Win-Commander er det kommandoen Split file (Opdel fil)

Uddrag af hjælpefilen er:

Split file

This function allows to split one (big) file into multiple smaller files (usually of the same size), e.g. for the transport over the Internet or with diskettes. The partial files can be written either directly on diskettes, or into a directory on the harddisk.

On the destination computer, the files can be combined either with Windows Commander (Command Combine files), or e.g. with the DOS command copy. The command line would be:

```
copy /b part.001 + part.002 + part.003 + part.004 original.dat
```

This combines parts 1 to 4 to the file original.dat.

After splitting Windows Commander creates a file with extension .crc. This file contains the name and size of the original file, as well as the CRC checksum. You don't need this CRC file to combine the files with a different program. However, Windows Commander can test with this file if the original file was recreated correctly.

Vejledning for udformning af forsøgsprotokoller og forsøgsrapporter, datadokumentation og opbevaring af data inden for sundhedsvidenskabelig basalforskning

Det er væsentligt for ethvert projekts gennemførelse, at forsker, vejleder og evt. andre medvirkende har gensidig informationsforpligtelse vedrørende de originale forsøgsresultater, deres bearbejdning og fortolkning. En hensigtsmæssig udformning og opbevaring af forsøgsprotokoller m.m. er derfor af afgørende betydning.

1. Forsøgsprotokollers overordnede afsnit bør skrives inden forsøgene udføres, således at der bliver tid til at forberede arbejdet. Forsøgsrapporten bør skrives færdig snarest muligt efter forsøgets afslutning og skal indeholde oplysninger om beregninger, foretagne korrektioner og deres forudsætninger i den udstrækning, det er nødvendigt for forståelsen af de opnåede resultater.
2. Forsøgsprotokoller skal føres med angivelse af dato og identifikation af den, der har ansvar for forsøgenes udførelse. Det er forfatterens ansvar, at protokoller, rapporter og bilag (også fra de forsøg, som ikke publiceres) opbevares forsvarligt i en for alle medvirkende umiddelbart tilgængelig form i mindst 10 år. Indeks skal regelmæssigt ajourføres. Den projektansvarlige skal have adgang til at medtage kopi af data ved ansættelsens ophør.
3. Forsøgsprotokoller skal være overskuelige og utvetydige for alle implicerede parter, ikke alene for dem, der planlægger og udfører forsøgene, men også for dem, der eventuelt senere skal vurdere resultaterne. Det kan derfor være en fordel at anvende en standardiseret opstilling af forsøgets titel, formål, materialer, metoder, tidsplaner, rådata og beregninger som disposition for hver forsøgsprotokol.
4. Forsøgsprotokoller skal udformes, så forsøgsomstændighederne kan reproducere, selv flere år senere eller i andre laboratorier. Derfor kan det være nødvendigt at beskrive nye forsøgsobjekter, apparater, kemikalier, isotoper, etc. når de anvendes for første gang.
5. Forsøgsrapporter skal indeholde tilstrækkelige oplysninger om eventuelle fejl og afvigelser fra den planlagte arbejdsgang og de anvendte materialer. Dette kan blive afgørende for, om visse data skal udgå af en opgørelse, hvortil kommer, at sådanne ændringer i forsøgsbetingelserne kan belyse nye aspekter, og derfor kan være af videnskabelig værdi. Hvis der foretages rettelser, skal det kunne ses, hvad der oprindeligt har stået.
6. Forsøgsrapporters data skal i så stor udstrækning som muligt omfatte originale rådata i en letlæselig form, fx som papirdokumentation af elektroniske data, opklæbte strimler fra printere, kurveskrivere, automatiske vægte, tællere, autoanalytatorer og regnemaskiner. Kopi af væsentlige elektroniske data bør snarest muligt arkiveres i en fælles database, som skal bero i institutionen. De deltagende forskere kan disponere over egne kopier.
7. Forsøgsbilag skal muliggøre en kvalitetsvurdering, dvs. en bedømmelse af resultaternes statistiske variation, reproducerbarhed og evt. systematiske fejl, såvel inden for det enkelte forsøg som mellem flere forskellige. Det skal være muligt at identificere de originale observationer, som indgår i de publicerede data (fx ved protokolnummer).
8. Det påhviler forskningsinstitutionens ledelse og projektvejledere at gøre ovenstående retningslinier bekendt for alle involverede parter, enten i den foreliggende form eller som en af institutionen udarbejdet vejledning, baseret på lignende principper.

Vejledning for udformning af undersøgelsesplaner, datadokumentation og opbevaring af data inden for klinisk og klinisk-epidemiologisk forskning

Det er væsentligt for ethvert projekts gennemførelse, at forsker, vejleder og evt. andre medvirkende har gensidig informationsforpligtelse vedrørende de originale forsøgsresultater, deres bearbejdning og fortolkning. En hensigtsmæssig udformning og opbevaring af undersøgelsesplaner m.m. er derfor af afgørende betydning

1. Undersøgelsesplaner, spørgeskemaer, personbilag (case report forms) og andre bilag skal være overskuelige og utvetydige for alle implicerede parter, ikke alene for dem der planlægger og udfører forskningen, men også for dem, der eventuelt senere skal vurdere resultaterne. Brug derfor skrivemaskine eller PC og en standardiseret opstilling af undersøgelsens titel, formål, materialer, procedurer, rådata og beregninger, som disposition for hver undersøgelsesplan.
2. Undersøgelsesplaner skal udformes i så god tid, at der bliver tid til at teste personbilagenes praktiske anvendelighed, indhente videnskabetisk accept, m.m.
3. Undersøgelsesplaner og -bilag skal være så detaljerede, at det bliver muligt at vurdere, hvorvidt en stikprøve er repræsentativ i forhold til den population, den stammer fra. Der skal derfor være præcise inklusionskriterier for undersøgte personer og beskrivelse af indgang til undersøgelsen, fx om den er planlagt at være konsekutiv, ligesom betingelser for udgang af undersøgelsen før tiden (drop outs) skal omtales.
4. Personrelaterede data fra klinisk-videnskabelige undersøgelser skal kunne identificeres sikkert, og de skal dateres og signeres. Rubrikker, der ikke udfyldes, udstreges.
5. Personbilagenes data skal i så stor udstrækning som muligt omfatte originale rådata i letlæselig form, fx som papirdokumentation af elektroniske data, opklæbde strimler fra printere, kurveskrivere, automatiske vægte, tællere, autoanalyser og regnemaskiner. Kopi af væsentlige elektroniske originaldata bør snarest muligt efter undersøgelsens afslutning arkiveres i en fælles database, som skal bero i institutionen. De deltagende forskere kan disponere over egne kopier.
6. Bilag til undersøgelsesrapporten skal indeholde udførte beregninger, herunder observationsberegninger, korrektioner og disses forudsætninger, som nødvendig dokumentation og for at lette forståelsen af de opnåede resultater.
7. Der skal foreligge oplysninger om kvalitetskontrol af væsentlige data, og det skal anføres hvilke statistiske metoder og edb-programmer, der er anvendt.
8. Det skal være muligt ud fra undersøgelsesbilag og spørgeskemaer at identificere de originale observationer, som indgår i publicerede tabeller og figurer.
9. Indhentede tilladelser fra det videnskabetiske komitéssystem, Registertilsynet, Strålehygiejnisk Laboratorium, Sundhedsstyrelsen og eventuelt andre berørte instanser, samt samtykkeerklæringer fra personer, der indgår i undersøgelsen, skal gemmes i henhold til de forordninger, der fremgår af gældende lovgivning, og under hensyn til eventuelle opfølgende undersøgelser. Det samme gælder interview- eller spørgeskemaer og personbilag. Alt sådant materiale skal opbevares separat, ikke i journaler, og gemmes i mindst 10 år. Dansk Data Arkiv er velegnet til opbevaring af data, specielt fra samfundsmedicinsk forskning. Anvendte koder til anonymisering skal ligeledes gemmes, i den udstrækning lovgivningen tillader det.
10. Det påhviler forskningsinstitutionens ledelse og projektvejledere at gøre ovenstående retningslinier bekendt for alle involverede parter, enten i den foreliggende form eller som en af institutionen udarbejdet vejledning, baseret på lignende principper

Euroqol 5-d³⁰

Angiv, ved at sætte kryds (sådan) i én af kasserne i hver gruppe, hvilke udsagn, der bedst beskriver din helbredstilstand i dag.

Bevægelighed

- Jeg har ingen problemer med at gå omkring
 Jeg har nogle problemer med at gå omkring
 Jeg er bundet til sengen

Personlig pleje

- Jeg har ingen problemer med min personlige pleje
 Jeg har nogle problemer med at vaske mig eller klæde mig på
 Jeg kan ikke vaske mig eller klæde mig på.....

Sædvanlige aktiviteter (fx. arbejde, studie, husarbejde, familie- eller fritidsaktiviteter)

- Jeg har ingen problemer med at udføre mine sædvanlige aktiviteter
 Jeg har nogle problemer med at udføre mine sædvanlige aktiviteter
 Jeg kan ikke udføre mine sædvanlige aktiviteter

Smertes/ubehag

- Jeg har ingen smerter eller ubehag.....
 Jeg har moderate smerter eller ubehag
 Jeg har ekstreme smerter eller ubehag

Angst/depression

- Jeg er ikke ængstelig eller deprimeret
 Jeg er moderat ængstelig eller deprimeret
 Jeg er ekstremt ængstelig eller deprimeret

Sammenlignet med min helbredstilstand gennem de seneste 12 måneder, er min helbredstilstand i dag:

- Sæt kryds i
én kasse
- Bedre
 Stort set den samme
 Værre

³⁰ For mere information om Euroqol henvises til: Frank de Charro Erasmus University Rotterdam telephone: +31 10-408 1545 E-mail: deCharro@gbr.frg.eur.nl.

“Nu vil jeg gerne beskrive din helbredstilstand i dag med et enkelt tal.”

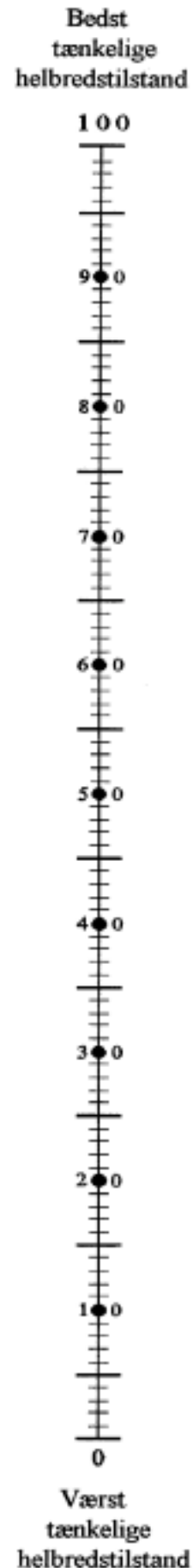
For at sige hvor god eller dårlig din helbredstilstand er i dag skal du forestille dig, at vi kunne måle den ligesom med et termometer. Hvor den bedste helbredstilstand du kan forestille dig er 100 og den værste helbredstilstand du kan forestille dig er 0”

“På en sådan skala fra 0 til 100, hvor godt eller dårligt mener du dit eget helbred er i dag ?”

“Sig et tal fra 0 til 100” (Tegn tallet ind i figuren)

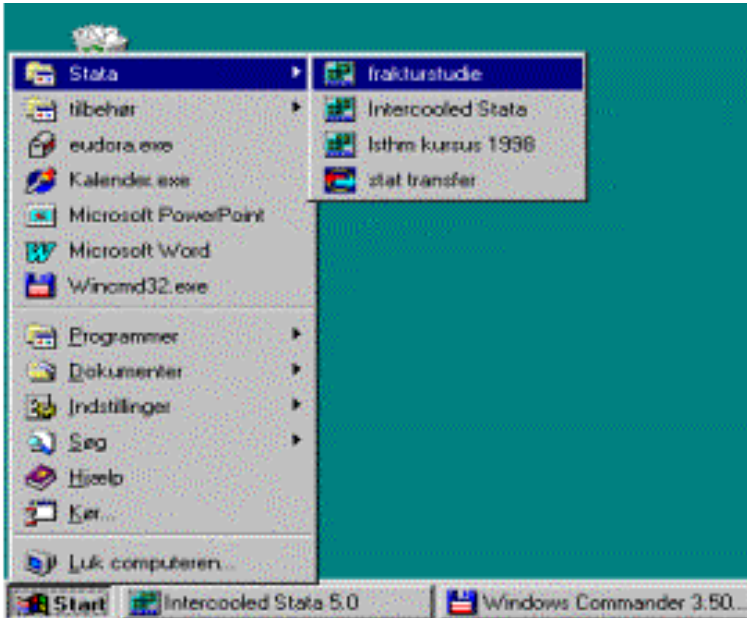
Tegn en streg her fra og til det tal, som bedst svarer til din helbredstilstand i dag

OBS, den originale form skal anvendes til konkrete formål, dette er blot en tilnærmet !!



God opsætning af start for Stata og Stat/Transfer

Hvis følgende er volapyk for dig, så må du få hjælp eller læse i windows hjælpefilerne om opsætning af menuer og skrivebord.



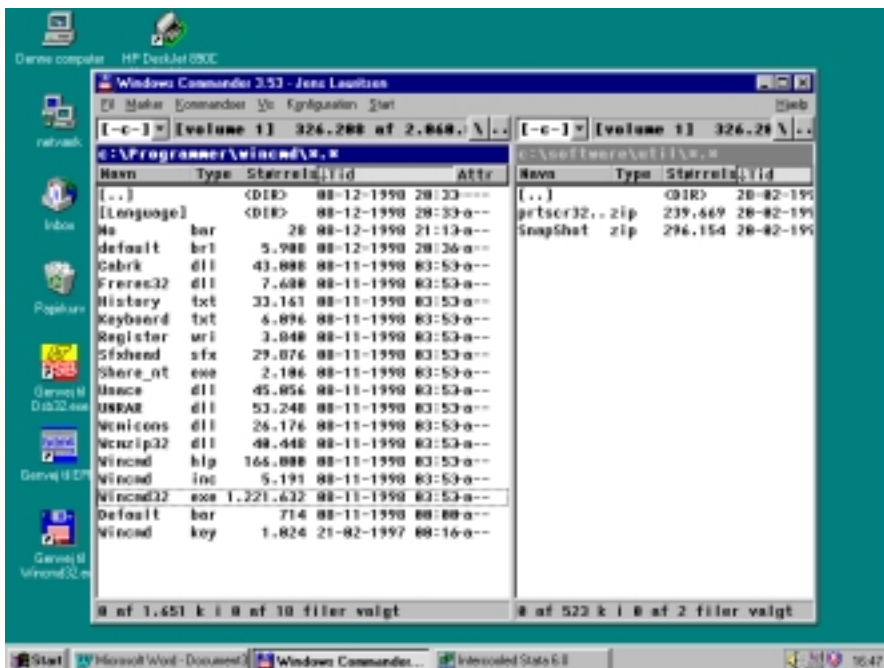
For at arbejde effektivt skal du tilføje et menupunkt for hvert projekt du arbejder på.

Ideen er, at du under startgruppen opretter en undergruppe, der hedder Stata.

Heri findes en linie for hver projekt eller artikel du arbejder på (hvor data er i hver sin mappe).


Øvelse 93-s God opbygning af startmenuen (tast **Alt+F7**)

Find (Søg **Alt+F7**) efter *wincmd32.exe*, peg på filen med musen og hold venstre musetast nede, mens du flytter pilmarkøren ned til Startikonen nederst til venstre op skærmen.



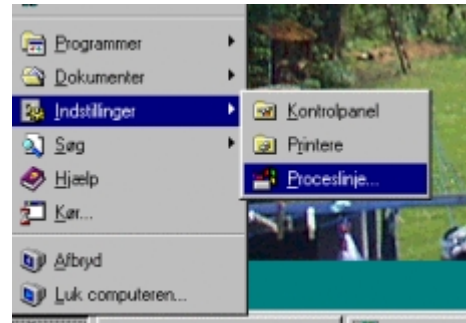
Herefter vil *wincmd32.exe* findes på startmenuen.

Øvelse 94-s Opbygning af STATA mappe på startmenuen (tast )

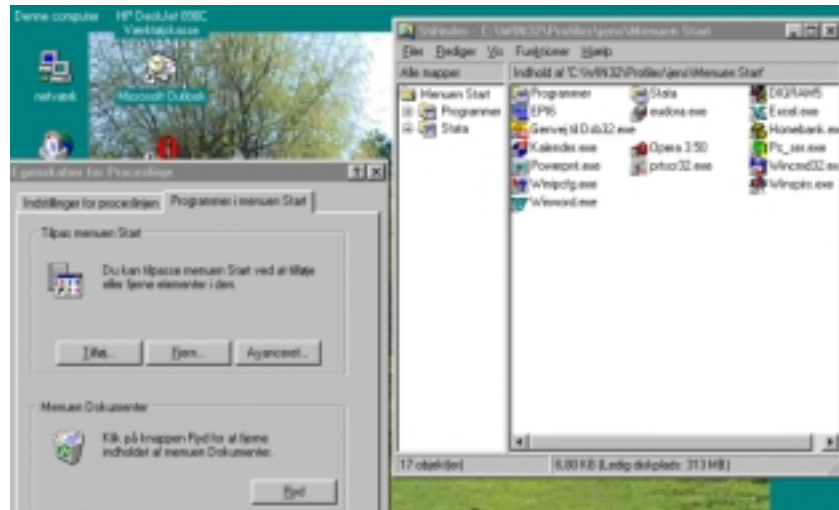
Nu skal du opbygge en startmenu, så det er let at få Stata startet direkte med få tastetryk.. Det vi ønsker er, at få Stata frem som en gruppe, når der trykkes på  (eller Ctrl+Esc) uden videre. Under Stata opbygger vi en linie for hvert projekt, der arbejdes på for tiden. På den måde undgår du at lede efter filer det forkerte sted.

Punkt 1: - Tryk  (eller Ctrl+Esc) og gå ind i ”Indstillinger” og dernæst ”Proceslinje”:

Den grå menu du ser til venstre kaldes startmenuen.



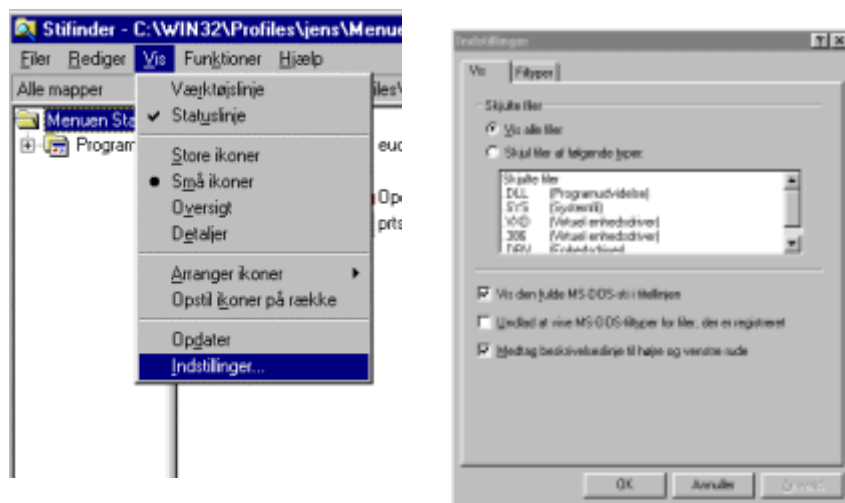
Punkt 2:
- Tryk på ”Avanced” og du får en firkant frem som vist til højre i figuren. (men med et andet indhold end her)



Punkt 3: Gå ind i ”Vis” og ”Indstillinger” og sæt systemet som vist til højre

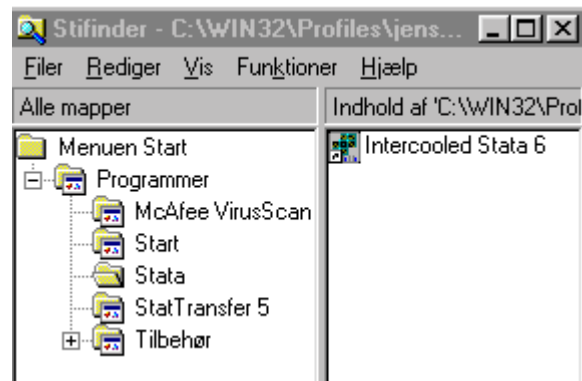
(Ellers kan det være svært at se adskille forskellige filtyper)

Tryk dernæst på ”Anvend” og ”OK”

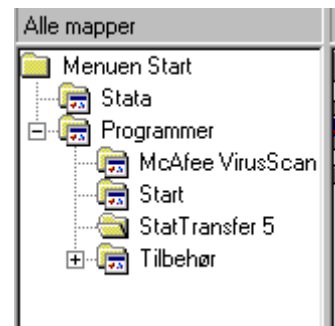


Punkt 4.

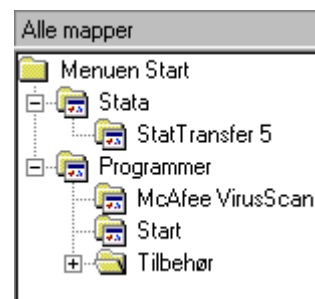
Hvis du nu med musen peger på ”Programmer” i feltet til venstre og dernæst på den mappe, der hedder ”Stata” får du dette billede frem:




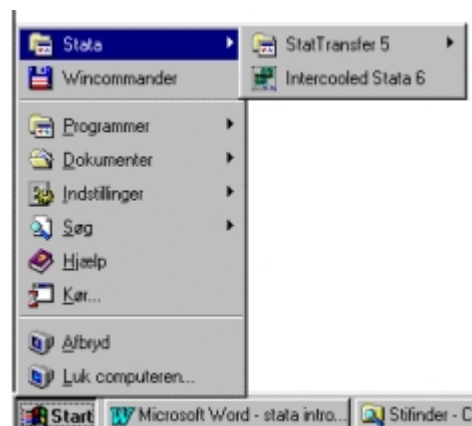
Punkt 5 Peg nu med musen på ”Stata” og hold venstre musetast nede, du kan nu flytte mappen ”Stata” ovenpå ”Menuen Start”, således at billedet derefter ser sådan ud:



Punkt 6. Endelig kan du flytte ”Stat Transfer 5” op som et underniveau til Stata. Hvis du derefter clicker på det ”+”, som er ud for Stata vil skærbilledet se sådan ud:



Som en kontrol af dette kan du følge med i ændringerne i start menuen ved at trykke på  og se, at skærmen nu ser sådan ud:



Punkt 7.

Peg nu med musen på ”Stata” mappen og til højre på Intercooled Stata 6. Tryk dernæst:

- Ctrl+C
- Ctrl +V
- Ctrl +V (her opbygges i alt tre kopier)
- Ctrl +V



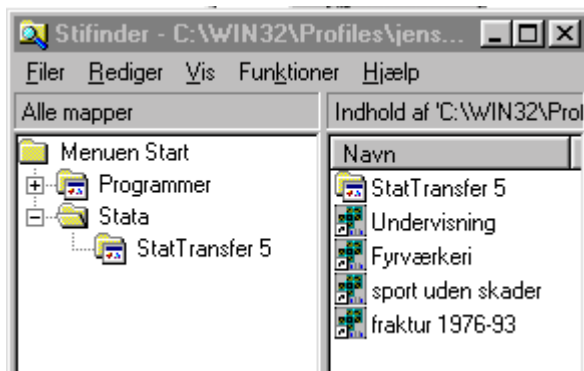
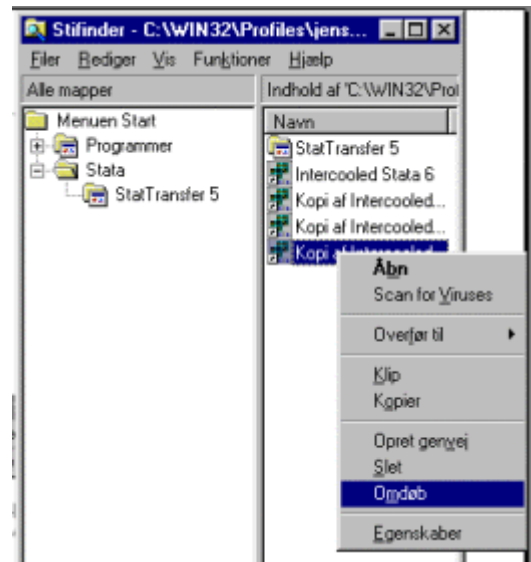
Hver gang du trykker på Ctrl +V tilføjes en kopi nedenunder ”Intercooled Stata 6”, således at det efter pkt. 7 ser sådan ud:



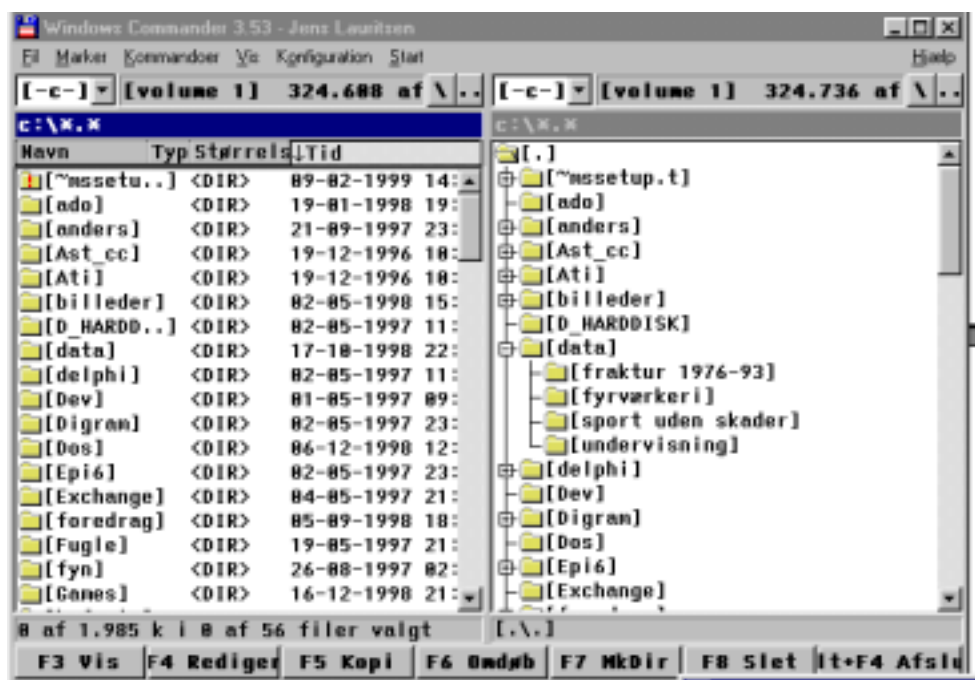
Punkt 8: Nu skal hver af kopierne tilpasses til et bestemt projekt. Først gives hver af de fire kopier navne. Dette gøres ved at du med højre musetast (eller tasten der aktiverer menu – på de fleste win-95 tastaturer er det tre tast til højre for mellemrumstangenten) clicker på en linie og dernæst vælger ”Omdøb”

Tænk på fire mulige projekter og giv de fire kopier navn, som de fire projekter. Her bruges:

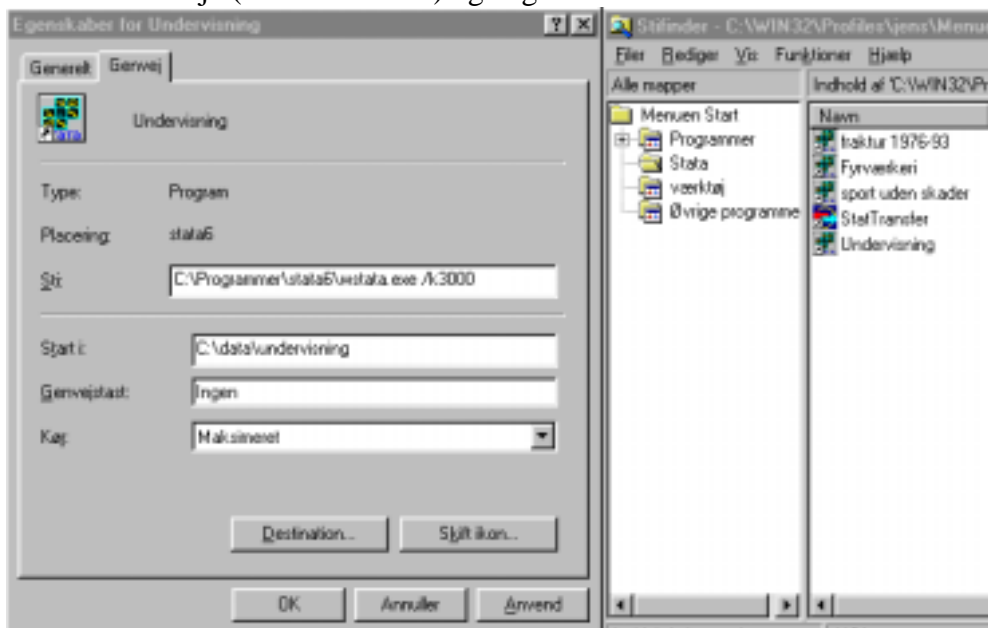
1. Fyrværkeri
2. Undervisning
3. Sport uden skader
4. Fraktur 1976-93



Punkt 8: Opret mapper på harddisken til projekterne. Dette kan gøres med Win-Commander, se billedet nedenfor (Mapper oprettes med F7 mkdir). Her er der i højre vindue valgt at vise (Vis) strukturen på harddisken (Ctrl+F8: bibliotekstræ/tree)



Pkt 9. Nu skal hver af af de fire opstartspunkter til Stata tilpasses. Du skal først skifte tilbage til stifinderen med Alt+TAB og dernæst få undermenuen frem, som vist i figuren ved punkt 8 ovenfor. I stedet for at omdøbe som i punkt 8 skal du aktivere punktet "Egenskaber". (Kan også gøres uden menuen ved at trykke på Alt+Enter) Når du har gjort det ser skærmen ud som vist nedenfor. Du skal aktivere den del der hedder "Genvej" (en såkaldt fane) og angive som vist:



Sti: Ved at skrive som vist /k3000 efter *wstata.exe* får Stata rådighed over 3Mb hukommelse. Størrelsen skal cirka være 1½ gange størrelsen af din datafil, se side 23 om hukommelse. Hvis du ikke skriver /k vælges standard størrelsen 1Mb. **Start i mappen:** "c:\data\undervisning" **Kør:** **Maksimeret** vælges, så Stata fylder hele skærmen.

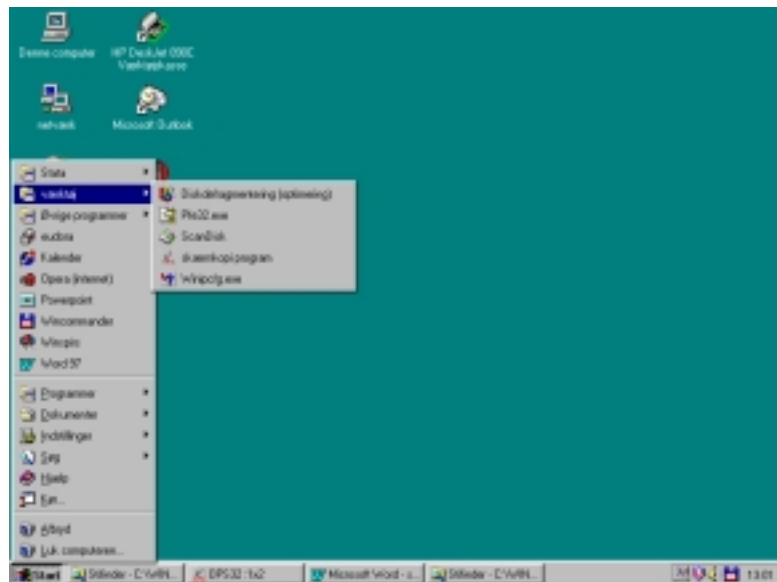
Øvelse 95-s Hjælpeprogrammer i startmenuen. (🗑)

På samme måde som ovenfor kan du opsætte en mappe med værktøj. Denne bør indeholde de mest anvendte, fx Diskdefragmentering, Backup og Scandisk. Disse kan du finde under systemværktøj, hvis de er installeret.

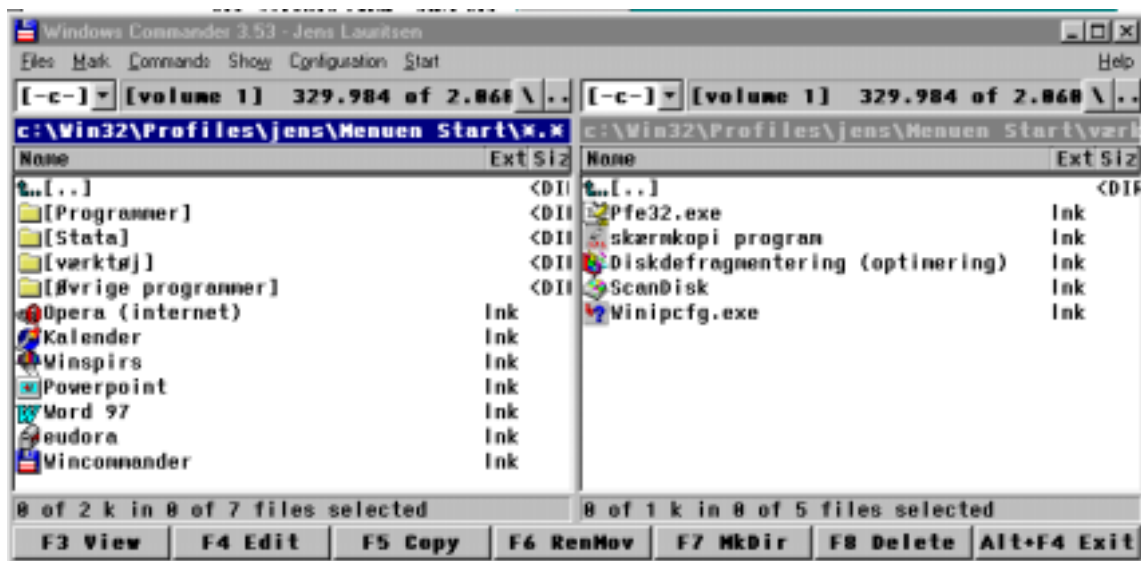
Hvis du gør som ovenfor kan du opbygge en værktøjsgruppe som vist til højre.



Når du herefter trykker på **Ctrl+Esc** eller **🗑** vil du få dette billede:




Der er ikke noget mystisk i denne start menu, prøv at kigge på den samme opbygning med Win-Commander. Du skal finde den mappe, som hedder "Menuen Start", hvis der er flere brugere af PC'en vil den findes under mappen



"Profiles"

Hyppigt anvendte tastaturkombinationer i windows.

Mellem programmer eller menu'er

TAB	Hop forlæns i en "dialogbox", fx når du skal printe
Shift+TAB	Hop baglæns i en "dialogbox", fx når du skal printe
Alt + TAB	Skift mellem aktive applikationer (en applikation er et program, fx Stata)
Ctrl + ESC eller 	Vis startmenuen
Ctrl+Alt+Del	Vær lidt varsom med denne kombination (men ikke bange for) ! Det kan være sidste udvej, hvis et bestemt program "hænger". Peg på det pågældende program og anfør " afslut ". Det pågældende program og kun det afsluttes. Resten af windows kører videre (burde). Hvis fejlen gentager sig må du overveje om det skyldes noget andet eller du evt. har arbejdet for længe.

Indenfor et enkelt program / applikation. Fx filhåndtering, Stata, word eller wordperfect.

Ctrl + F6	Skift mellem aktive vinduer i den aktuelle applikation.
ESC	Fortryder valg af en menu, opsætning eller andet. (Prøv denne tast, hvis der tilsyneladende ikke sker noget).
F10	Aktiver menuen øverst på skærmen.
Alt+f	Aktiver fil menu punktet, som er adgangen til at åbne filer, lukke filer, printe, vælge printer mm.
Alt+mellemlinjestangent	Gør det muligt at ændre den samlede størrelse af det hovedvindue, som det aktuelle program fylder. Tryk derefter på maksimer for at få aktuelle applikation til at fylde hele skærmen.
Alt+F4	Afslut programmet.

Indenfor et enkelt vindue i en applikation. Fx syntaksvinduet i Stata. Et bibliotekstræ i filhåndteringen eller et dokument i tekstbehandlingen.

end	Hop til slutningen af aktuelle linie eller til sidste fil i filhåndteringen
home	Hop til starten af aktuelle line eller første fil/første bibliotek i filhåndteringen.
Ctrl + a	Marker hele det aktive vindue som en blok.
Del	Slet det tegn cursoren står ved. Hvis en blok er markeret, slettes hele blokken. (En god måde at rense output vinduet på).
Shift+piltast	Marker fra aktuelle cursor sted og i den retning, hvor pilen flyttes
Ctrl+c	Tag en kopi af den aktuelt markerede blok og gem kopien på det såkaldte klippebord.
Ctrl+x	Tag en kopi af den aktuelt markerede blok og gem kopien på det såkaldte klippebord. Slet derefter blokken i det aktuelle vindue.
Ctrl+v	Hent en kopi fra klippebordet og indsæt den ved cursoren. (Henter det, som sidst er kopieret med Ctrl+c eller Ctrl+x).
Ctrl+ →	Hop et ord til højre
Ctrl+ ←	Hop et ord til venstre
Ctrl + end	Flyt til slutningen af aktuelt dokument (syntaksfil eller andet).
Ctrl + home	Flyt til starten af aktuelt dokument

Ctrl+a betyder: tryk først på Ctrl tasten med en finger og hold denne tast nede, mens du trykker på "a" med en anden finger.

General Commands in Stata

clear	Clear data from memory
display	Display values from functions
do	Execute commands
exit	Exit Stata
help	Find help
log	Echo output into a file
save	Save data for future use
use	Read a copy of a Stata file with data into memory
now	Start do file with automatic log (part of <i>kursus.zip</i>)
webseek	Find material on internet

Data manipulation and Management

collapse	aggregate data into a table
count	count number of observations
describe	which data are in memory
destring	change string variable to numerical
drop	eliminate variables or observations
egen	Group continous variable.
v=cut1(v2)	Label both ends of interval.
encode	create numeric variable from string
expand	duplicate observations
format	specify display format
generate	create new variable
infile	read fixed format data
input	add data via keyboard
label	descriptive text (values, vars., files)
list	list some variables
listjl	list id variable (also by group)
mmerge	combine files
move	change sequence of variables in file
mvdecode	recode to missing (e.g. from 9 to .)
mvencode	recode from "." to value
notes	add documentation to data + vars.
order	change sequence of variables
outfile	write data in fixed format
pattern	indicate patterns of missing
query	show set options
quietly	execute command, don't show output
recode	recode numeric
rename	change name of variable
replace	change contents of variable
set	set general options
sort	sort observations by variables.

Special Graphs

cdf	Cumulative plot of continous var.
graph	graph Y ₁ Y ₂ Y ₃ X , xlabel ylabel t1(" ") b1(" ") r2(" ") Saving(" ")
hbar	Horisontal bar charts,Popul.Pyramids
qnorm	Q normal probability plot
venndiag	Venn diagrams+creation of variables
violin	Violin distribution plot

Descriptive Statistics

codebook	Simple standard tables
etab	One variable against many
graph	graph Y ₁ Y ₂ Y ₃ X , xlabel ylabel xscale t1(" ") b1(" ") title(" ") Saving(" ")
summarize	display summary statistics
tab1	oneway tables for many vars.
table	Summary statistics + tables
tabulate	one and two way frequency

General Statistics

anova	Analysis of Variance
correlate	Correlation
oneway	One-way analysis of variance
ranksum	Wilcoxon ranksum test
tabulate	Crosstables incl. tests of homogeneity, gamma, exact r*c test
ttest	Mean comparison

Cohort / survival analysis

stmh	Manten-Haenzel rates
ltable	life table and KM plots
strate	table of rates
stset	define survival data
staalen	Aalen cumulative hazards
stcox	Cox regression (+stphtest)
stsplit	expand data according to a lexis diagram
sts graph	Kaplan Meyer survival functions in graph
sts test	Logrank test

Case Control / Cross Sectional analysis

mhods	Mantel Haenzel Odds Ratios
partgam	Partial Gamma Coefficient
tabodds	Table of Odds

Regression Models

clogit	conditional logistic regression
logit	logistic regression (output on log scale)
logistic	logistic regression (output on antilog scale)
lrtest	likelihood ratio test
outreg	Regression output to publication
poisson	poisson regression
predict	prediction + indicator of fit
regress	linear regression
Categorical variables- "dummy" indicator:	

sort X	/* stratify on X */
by X:	command variables, options
for var variables:	command X