



En kort oversigt over indtastning af data ved hjælp af

EpiData

© Jens M. Lauritsen & Michael Bruus. EpiData Foreningen, Odense, Danmark. 2003-. For yderligere oplysning og information om seneste udgave: Se <http://www.epidata.dk>

Hvad er EpiData ?

EpiData er et program til indtastning og dokumentation af data.

Brug EpiData når du har indsamlet data som er nedskrevet på papir og du ønsker at udføre statistiske analyser eller tabulering data. Data kan både være spørgeskemaer, interviewskemaer eller systematiske optegnelser, fx. registreringsark fra en tælling. EpiData indeholder kun begrænsede funktioner til data analyse, men kan omforme data til direkte brug i statistikprogrammer.

Med EpiData kan du anvende ”**Kontrolleret indtastning**”. Kontrolleret betyder at EpiData kun tillader brugeren at indtaste data, som opfylder bestemte kriterier. fx. lovlige værdier med tilhørende såkaldte værdi labels (1 = Nej 2= Ja), interval check (kun aldre fra 20-100 er tilladt), lovlige værdier (fx 1,2,3 og 9) eller lovlige datoer (fx er 29 febr 2001 ikke gyldig).

EpiData kan bruges til enkle datasæt med et skema, men også ved flere sammenhængende skemaer eller kombinationer af datakilder. **EpiData** hentes fra [Http://www.epidata.dk](http://www.epidata.dk) og er gratis. Se på samme www side efter en versionsliste og en redegørelse for udviklingen.

Princippet i EpiData er udviklet i dos programmet EpiInfo, som har mange brugere rundt om i verden. Ideen er at du skriver simple tekst linier. Programmet konverterer linierne til en indtastnings formular, kaldet dataform. Når dataformen er klar er det let at indtaste direkte eller at definere særlige kontroller, som skal udføres under indtastningen.

EpiData fungerer uafhængigt af andre programmer.

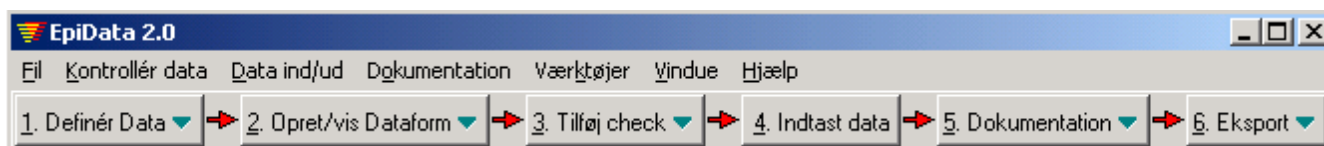
Det er et vigtigt princip for EpiData ikke at blande sig i opsætningen af din computer. EpiData består af en program fil og nogle få hjælpefiler. (Teknisk set betyder dette, at der ikke installeres eller læses i dll system filer udenfor selve EpiData mappen – opsætningen gemmes i ”registry”).

Begrænsninger

Der er ingen begrænsning i antal observationer. (testet med 250.000). Søgning med indeks i 80.000 observationer < 1 sek med Pentium I 200Mhz). Alle felter (variable) skal kunne være indenfor 999 tekstlinier.

Sådan arbejder du med EpiData

EpiData skærmen har ”standard” Windows layout med en menu linie og to værktøjslinier.



”Proces værktøjslinien” følger et projekt trinvis fra ”1. Definér data” til ”6. Eksport” af data til et statistisk analyse program. Hvert af disse omtales kort nedenfor.

1. Definér Data

Definér data ved at skrive tre typer information for hver variabel:

- A.. Navn på feltet (En variabel læses fra et felt, fx v1 eller alder).
- B.. Tekst der beskriver variabelen. (fx køn eller "fødselsdato")
- C.. Definition, fx. ## for et to cifret tal.

```
Mit første indtastningsskema
id      (automatisk id nummer) <idnum >
v1      køn #
v2      Højde (meter) #.#
v3      Vægt(kilo) ###.#
bmi     Body Mass Index ##.#
v4      Fødselsdato <dd/mm/yyyy>
age     Alder i dag ###
s1      Bor i dette land _____
s2      By (nuværende adresse) <a >
t1      Dags Dato <Today-dmy>
```

Andre felt typer er "logiske" (ja-nej) eller lydkode (Soundex).

Variabel navne kan være på to måder:

- a. v1koen (8 første bogstaver i linien)
- b. v1 (første ord i linien).

2. Opret/vis dataform.

Efter at du har skrevet definitionen kan du enten vise dataformen eller oprette en datafil.

3. Tilføj Check – under indtastning

Der kan defineres omfattende **regler og beregninger under indtastning**.

- Sæt restriktioner, så kun bestemte værdier indtastes. Tilføj samtidig kodeværdier der beskriver disse talkoder.
- Angiv rækkefølge af indtastning eller at nogle felter springes over, fx for mænd.
- Foretag beregninger under indtastning. Fx. antal dage fra indlæggelse til et bestemt symptom eller alder ved undersøgelse i et ambulatorie. Beregn skalaer eller indeks.
- Særlige instruktioner eller udvidede definitioner afhængig af indtastede værdier, fx. hierarkiske datastrukturer eller udvidede if .. then ...endif strukturer.
(Se <http://www.epidata.dk/downloads/examples/first.htm>).

4. Indtast Data

Filen åbnes og du kan tilføje, rette eller søge i data. Farver kan defineres. Her er valgt hvid baggrund og gule felter.

Den blå forklarende tekst er tilføjet af EpiData ud fra værdier i de pgl. felter. Body mass index og alder er beregnede variable.

Filer der oprettes:

- A. Dataform definitions fil, fx. **first_dk.qes**
- B. Filen der indeholder data.fx **first_dk.rec**.
- C. Filen med definerede check. fx **first_dk.chk**
- D. Supplerende filer, fx **first_dk.not** med indtastningsnoter eller **first_dk.log** med dokumentation.

Mit første indtastningsskema

ID	(automatisk id nummer)	1
V1	køn	2 Female
V2	Højde (meter)	1.92
V3	Vægt(kilo)	122.0
BMI	Body Mass Index	33.09
V4	Fødselsdato	12/12/1978
AGE	Alder i dag	22
S1	Bopæl i (land)	Denmark
S2	By (nuværende adresse)	00 Odense
T1	Dags Dato	21/02/2001

1/5 * id= 1

5. Dokumentation af Data

Efter oprettelse af datafilen kan strukturen dokumenteres. Eksempel (del af *first_dk.rec*) er:

```
DATAFIL: C:\data\first_dk.rec
Fillabel: Min første datafil (eksempel)

Filstørrelse:      1016 bytes
Sidste ændring:   21. feb 2001 20:25
Antal felter:     10
Antal observationer: 2
Checks tilføjet:  Ja (Sidste ændring 21. feb 2001 20:08)

Felter i datafilen:
```

Nr.	Navn	Variabel label	Felttype	Bredde	Checks	Værdi labels
1	ID	(automatisk id nummer)	ID-number	6		
2	V1	køn	Integer	1		sex 1: Mand 2: Kvinde 9: Uoplyst
5	BMI	Body Mass Index	Fixed number	5:2	NoEnter	
8	S1	Bopæl i (land)	Text	14	Legal: Denmark,Cuba, Columbia,Nepal, China,Australia, South Africa	
10	T1	Dags Dato	Today date-dmy	10		

(øvrige felter udeladt her)

Efter indtastning af data kan værdierne vises.

```
Observation 1
```

id	1	v1	Mand	v2	1.92
v4	12/12/1945	s1	denmark	s2	Copenhagen
t1	28/10/2000				

En "kodebog" kan udskrives med samlede tabeller. (eksemplet er ikke fra *first_dk.rec*)

```
v2 ----- Sex
      type: Integer
      værdi labels: sex
      range/legal: 1-2,2
      uoplyst: 0/25
      interval: [1,2]
      unikke værdier: 2
      tabel:  Frekv.  Pct.  Værdi  Label
              11    44.0    1    Male
              14    56.0    2    Female
v3 ----- Temp
      type: Floating point
      range/legal: 36.00-40.00
      uoplyst: 0/25
      interval: [36.00,37.50]
      unikke værdier: 12
      Middelværdi: 36,84
      std.afv.: 0,37
```

6. Eksport til statistisk analyse og sikring af data.

Med sikkerhedskopiering kopieres alle filer der hører til en bestemt datafil til et andet sted, fx. diskette eller netværksdrev. Det valgte drev huskes fra gang til gang.

Værktøj og andet

Epidata har flere funktioner end de nævnte, fx. kan filer der er dobbelttastede sammenlignes. Evt. forskelle udskrives sammen med id nummer. Filstrukturen kan tilrettes uden at miste indtastede data. Data kan indtastes efter et hierarkisk kodeprincip. Med komplet udvikling af EpiData forventes også relationel indtastning og udvidet kontrol af logisk sammenhæng i data efter indtastning. Desuden påtænkes funktioner der gør det muligt at opfylde datatilsynets krav om brug af "logfiler" og kryptering ved indtastning af data der kan henføres til enkelt personer, samt mulighed for at brugerne kan tilføje funktioner til check fil kommando sproget.

Analyse

Datafiler som skrives af EpiData kan umiddelbart analyseres i EpiInfo. Til andre programmer kan data eksporteres i simpel tekst form (data og variable: komma separerede ascii/ansi filer, dbaseIII eller Excel) eller i komplet form (data samt variable og labels: Stata version 4-8, SPSS, SAS). Desuden kan der importeres data.

Support

Hjælpefilen samt supplerende dokumenter kan hentes gratis fra <http://www.epidata.dk>.

Om EpiData

EpiData er et Windows 95/98/NT/2000 baseret program (32 bit) til indtastning og dokumentation af kvantitative data. Programmet er gratis og udgives af EpiData Foreningen. Program design Jens M. Lauritsen og Michael Bruus, EpiData Foreningen. Programmering udføres af M.Bruus.

Fonds støtte og taksigelser.

Mark Myatt, Brixtonhealth UK var meget aktiv i starten.

Følgende har støttet udviklingen af **EpiData**:

Fyns Amt, Danmark. Brixton Health, Wales UK. Dansk Data Arkiv/ERAS, Danmark, Syddansk Universitet - Sundhedsvidenskab, Århus Universitet - Sundhedsvidenskabeligt fakultet. Valid International. London UK. London School of Hygiene & Tropical Medicine, UK. International Centre for Eye Health, London UK. WHO-TDR Geneve, IUATLD, Paris m.fl. se <http://www.epidata.dk/funding.htm>

EpiData er udviklet for at fastholde qes-chk-rec fil strukturen som anvendt i Epi Info V6. Samt for at udvide denne strategi med principper for god datadokumentation. En komplet liste over "acknowledgements/taksigelser" findes i hjælpefilen.

Internationale udgaver findes på forskellige sprog, bl.a. Engelsk, Hollandsk, Spansk, Fransk, Slovensk og senere Portugisisk. Se [Http://www.epidata.dk](http://www.epidata.dk)

For bidrag eller donationer til yderligere udvikling kontakt, send en e-mail til: info@epidata.dk

Begrænset Ansvar

EpiData er udviklet og testet således at fejl under indtastning og dokumentation af data minimeres. Vi har gjort hvad der var os muligt for at udvikle et program der minimerer eller undgår fejl, men kan under ingen omstændigheder gøres ansvarlige for fejl, mistede eller slettede data, tab af arbejdstid og eller andre tab der måtte opstå i forbindelse med brug af programmet.