

En kort innføring i dataregistrering med **EpiData**

© Jens M. Lauritsen, Michael Bruus & Mark Myatt. Fyns Amt, Danmark og Brixton Health, Wales UK. 2001. For nærmere informasjon og nedlasting av siste versjon, se: <http://www.epidata.dk>

Hva er EpiData ?

EpiData er et program for registrering og dokumentasjon av data.

Bruk EpiData som et verktøy til å registrere på datamaskin opplysninger fra spørreundersøkelser eller annen type systematiske registreringer. Du kan lage enkle utlistinger og frekvenstabeller, eller eksportere data for statistisk analyse med andre dataprogrammer.

Under inntasting kan data kvalitetssikres ved at du kan legge til **kontrollregler** for eksempel ved at programmet aksepterer inntasting av tall innenfor et intervall som du bestemmer, eller at du kun tillater inntasting av lovlige koder. Kodene kan ha forklarende tekst, "verdi- forklaring", for eksempel 1= Ja, 2=Nei. Registrering av koder kan også skje ved valg fra en meny på skjermen.

EpiData egner seg for enkle datakilder med bare ett registreringsskjema, og for mer kompliserte databaser med flere eller forgrenede registreringsskjemaer. Programmet er gratis, og kan lastes ned fra adressen øverst på siden. Det foregår stadig videreutvikling av programmet.

EpiData tar vare på prinsippene fra det populære statstikkprogrammet Epi Info for DOS som har mange tusen brukere over hele verden. Ideen er at i et tekstdokument, et *spørreskjema*, skriver du navn og koder for angivelse av type for *datafelt* (eller *variabler*). Programmet oppretter på bakgrunn av spørreskjemaet automatisk en datafil for lagring av data. Deretter kan du begynne å registrere data. Du kan også legge til egne kontrollregler eller sjekk-koder som forklart ovenfor. Videre er det mulig å kommentere og dokumentere dataregistreringen.

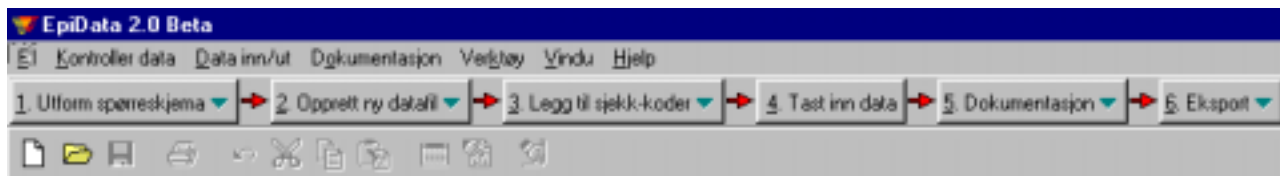
EpiData vil ikke forstyrre oppsettet på datamaskinen din

Det er et viktig prinsipp at EpiData ikke skal forstyrre oppsettet på datamaskinen. EpiData består av en enkel programfil og noen få hjelpefiler. EpiData er ikke avhengig av, forandrer eller erstatter noen DLL filer i systemkatalogen. Innstillingene du velger i programmet blir lagret i systemregisteret (I neste versjon: i en .ini fil i EpiData katalogen).

Begrensninger

Det er ingen grense for antall dataposter (testet med >250.000). Søking med index i 80.000 poster < 1 sekunder på Pentium I 200Mhz). Alle felter (variabler) må få plass på 999 tekstlinjer.

Slik arbeider du med EpiData



EpiData har "standard" Windows layout med en menylinje, og to verktøylinjer. Prosess-verktøylinjen leder deg gjennom den trinnvise arbeidsprosessen fra 1 (Utform spørreskjema) til 6 (Eksport av datafil for videre analyse). Hvert av arbeidstrinnene beskrives kort her:

1. Utform spørreskjema

Ved hjelp av spørreskjemaet definerer du datafilens struktur ved å skrive:

- Feltnavn (=variabelnavn, f. eks. "v1" eller "blodtrykk").
- Tekst som beskriver variabelen. (f. eks. Kjønn eller "Alder i hele år")
- En feltdefinisjonskode, for eksempel ## for numerisk verdi med to siffer.

Mitt første registrerings skjema (spørreskjema)			
id	(automatisk id nummer)	<idnum >	
V1	Kjønn	#	
V2	Høyde (meter)	#.##	
v3	Vekt (kilo)	###.#	
bmi	Body Mass Index	##.##	
v4	Fødselsdato	<dd/mm/yyyy>	
age	Alder i dag	###	
s1	Bosted (land)		
s2	By (Bostedsadresse)	<a >	
t1	Dagens dato	<Today-dmy>	

Andre feltyper er tekst, logisk (ja/nei), dato og SOUNDEX (automatisk "pseudonymisering" av navn).

Variabelnavn kan ha to former:

- v3vekt (8 første tegn i setningen)
- v3 (første ord i setningen).

2. Opprett ny datafil.

Når du har laget spørreskjemaet ferdig, kan du forhåndsvisse på skjermen hvordan dataregistreringsvinduet vil se ut, og så oppretter du datafilen som du skal registrere i.

3. Legg til sjekk-koder

Med EpiData kan du **definere logiske regler og beregninger** under inntastningen, noe som reduserer muligheten for feiltasting. Du kan:

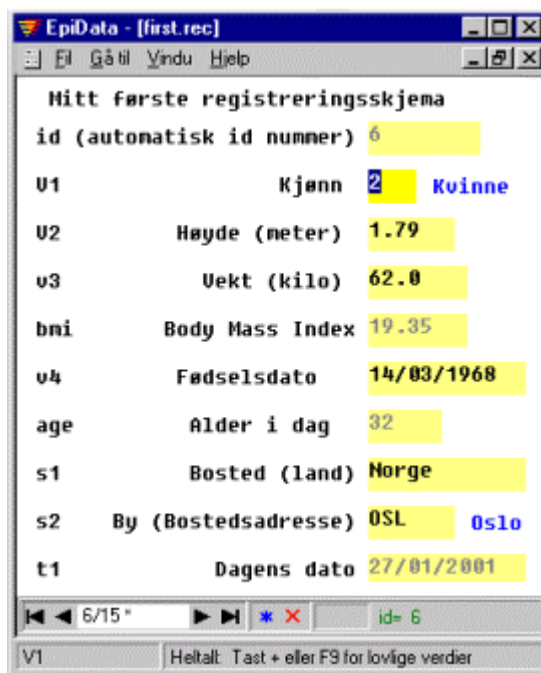
- Sette restriksjoner slik at kun forhåndsbestemte verdier kan tastes inn. Dette kan være koder som det knyttes beskrivende tekst til, eller minimum og maksimum for tallverdier.
- Angi inntastingsrekkefølge og betingede hopp i skjema (f.eks. hvis kjønn=mann, hoppes over felter som bare er relevante for kvinner)
- Foreta beregninger under inntasting. Alder kan f.eks. beregnes fra fødselsdato og intervjudato
- Hjelpetekster på skjermen, eller mer avanserte beregninger, if .. then ...endif sløyfer, osv (Se eksempler, hent *first.chk* fra <http://www.epidata.dk/downloads/examples/first.htm>).

4. Tast inn data

Åpne datafilen og tast inn, endre på, og søk fram data. Farger på tekst og bakgrunn kan endres i "Innstillinger". Her vises hvit bakgrunn og gule inntastingsfelter. De blå forklaringsstekstene ved siden av inntastingsfeltene legges til av EpiData på bakgrunn av definisjoner i sjekkfilen. Body mass index og alder beregnes automatisk.

Filer som blir opprettet:

- Registrerings skjema med feltdefinisjoner f.eks. **first.qes**
- Selve datafilen som inneholder alle observasjonene (datapostene) f.eks. **first.rec**
- En fil med alle inntastningsreglene f.eks. **first.chk**
- Tilleggsfiler, f. eks. **first.not** med fritekstnotater under inntastingen, og **first.log** med dokumentasjon.



5. Dokumentasjon

Når du har opprettet datafilen kan du dokumentere filstrukturen. Eksempel (del av *first.rec*):

```
DATAFIL: U:\EPIDATA TRANSLATION PACK\Norsk\first.rec
Filbeskrivelse: Mitt første spørreskjema

Filstørrelse:          1956 bytes
Siste endring:        20. jun 2001 09:53
Antall felter:        10
Antall dataposter:    15
Sjekk-koder tilføyd:  Ja (Siste endring 20. jun 2001 09:41)

Felter i datafilen:

Nr. Navn          Variabelbeskrivelse  Feltype          Bredde  Sjekk-koder      Verdi beskrivelse
-----
1 ID              (automatisk id      ID-number        6
nummer)
2 V1              Kjønn                Integer          1
                                                sex
                                                1: Mann
                                                2: Kvinne
                                                9: Ukjent
5 BMI            Body Mass Index      Fixed number     5:2     NoEnter
8 S1              Bosted (land)        Text             14      Legal: Norge,
                                                Danmark,Cuba,
                                                Colombia,Nepal,Kina,
                                                Australia

(resten av feltene er utelatt her)
```

Og etter inntasting kan du liste ut verdiene for alle eller noen av datapostene:

```
Nummer 1 (datapost nr. 1)

ID          1          V1          Kvinne          V2          1.75
V3          67.0         BMI         21.88          V4          12/12/1956
AGE         44          S1          Denmark         S2          Odense
T1          27/01/2001
```

En “kodebok” kan inneholde enkle frekvenstabeller og gjennomsnitt:

```
V1 ----- Kjønn
      type: Integer
      verdikoder: sex

      data mangler: 1/15
      intervall: [1,2]
      unike verdier: 3

      tabell: Frekv.   Pst.   Verdi   Beskrivelse
              1     6.7    .   Ukjent
              6    40.0    1   Mann
              8    53.3    2   Kvinne

V2 ----- Høyde (meter)
      type: Floating point
      range/legal: 0.0-2.30,9

      data mangler: 11/15
      intervall: [1.75,1.83]
      unike verdier: 4

      middelvei: 1,78
      std.avvik.: 0,03
```

6. Eksport for videre analyse og for sikkerhetskopiering av data.

Sikkerhetskopieringsrutinen kopierer datafilen og alle hjelpefilene som hører til, for eksempel til diskett eller nettverksserver.

Verktøy og diverse

EpiData har flere hjelpeverktøy, for eksempel kan en datafil som er tastet inn to ganger sammenlignes og dataposter som er ulike kan listes ut. Datafilens struktur kan revideres uten at data går tapt. Hierarkisk koding er mulig. Planlagt videreutvikling av programmet omfatter støtte for relaterte filer, utvidet kontroll av logiske sammenhenger etter inntasting, loggfiler, kryptering av data og mulighet for brukerdefinerte funksjoner i sjekk-filen og i menyene.

Analyse

Filer registrert med EpiData kan analyseres direkte med Epi Info (www.cdc.gov/epiinfo). For bruk med andre programmer, kan filene eksporteres i forenklet tekstform (data og variabelnavn: kommaseparert ascii fil for dbaseIII eller Excel) eller i komplett eksportformat (data og variabelnavn, pluss tekstforklaringer: Stata versjon 4-6 og senere SPSS).

Brukerstøtte

Hjelpefiler og diverse tilleggsdokumentasjon er tilgjengelig gratis fra: <http://www.epidata.dk>.

Om EpiData

EpiData er et Windows 95/98/NT/2000 basert program (32 bit) for dataregistrering og dokumentasjon av kvantitative data.

Program design: Jens M. Lauritsen, Fyns Amt, Danmark., Michael Bruus, Danmark
Mark Myatt, Brixton Health, Wales og The Institute of Ophthalmology, London, UK.

Programmering: Michael Bruus, Danmark.

Finansiering og takk

Vi takker for økonomisk støtte til utviklingen av EpiData fra:

Fyns Amt, Danmark. Brixton Health, Wales UK. Dansk Data Arkiv/ERAS, Danmark. Syddansk Universitet – Odense Universitet, Sundhedsvitenskabeligt fakultet. Valid International, London UK. London School of Hygiene & Tropical Medicine, UK. International Centre for Eye Health, London UK.

EpiData er laget for å opprettholde qes-chk-rec filstrukturen som brukes i Epi Info V6. Epi Info 2000 har valgt en annen strategi. Du kan lese mer om dette på <http://www.cdc.gov/epiinfo>. For en fullstendig oversikt over takk, se hjelpefilen. Vi vil særlig takke alle som har utviklet prinsippene bak EpiInfo v.6. Internasjonale utgaver finnes på mange språk. Se <http://www.epidata.dk> for detaljer.

For bidrag eller økonomisk støtte til videre utvikling, kontakt: info@epidata.dk

Anbefalt referanse:

Lauritsen JM, Bruus M, Myatt M. EpiData, Et verktøy for kvalitetssikret registrering og dokumentasjon av data. Fyns Amt, Danmark & Brixton Health UK. 2000. (versjon x.y). (Norsk oversettelse av Vegard Høgli.)

Juridisk forbehold

EpiData er utviklet og testet for best mulig å sikre feilfri registrering og dokumentasjon av data. Vi har gjort hva som har vært mulig for oss for å sikre at programmet skal forhindre feil, men vi kan under ingen omstendighet gjøres ansvarlig for feil, tap av data, tap av arbeidstid eller andre tap som måtte oppstå i forbindelse med bruk av programmet.