

Projekt nr. 1:

Et datasæt vedrørende trafikuheld.

Fra Danmarks Statistik har vi (via et seminar på 2. år udført af Martin Kufall Beck og Rikke Garth Hansen) fået et datasæt vedrørende trafikuheld i Danmark i 2002. Datasættet omfatter oplysninger for hver tilskadekommen om køn, alder, køretøjets art, graden af skade (drab/alvorlig skade/lettere skade) og amt. Datasættet lægger op til adskillige analyser ved hjælp af modeller, som MØK'er gerne skulle kende i forvejen fra undervisningen på 2. og 3. år (log-lineære modeller og logistisk regression).

Opgavens formål er at finde de sammenhænge, der måtte være mellem de demografiske faktorer, transportmidlet og og risikoen for at komme til skade i trafikken. Er det for eksempel generelt sådan, at risikoen for at komme til skade er større i byen end på landet (eller omvendt), er børn særligt udsatte for bestemte kategorier af ulykker, kommer mænd mere eller alvorligere til skade end kvinder for bestemte køretøjstyper, osv. osv. Der er rigeligt med spørgsmål at tage fat på, og det kræves ikke at man besvarer dem alle. Men man bør give en oversigt over de væsentlige sammenhænge, og gå mere i dybden med nogle specielle problemstillinger, som man kan vælge selv.

Data foreligger i form af filen 04-1-dat.txt, som kan hentes fra kursets hjemmeside

www.mes.cbs.dk/~sttt/MPAS.html

Filen er på $4950 = 15 \times 2 \times 5 \times 11 \times 3$ linier plus en overskriftslinie. Overskriftslinien og de første otte datalinier er gengivet her:

AMT	KON	ALDER	TRANSP	TYPE	ANTAL	BEFOLK
1	1	1	1	1	0	48980
1	1	1	1	2	4	48980
1	1	1	1	3	1	48980
1	1	1	2	1	0	48980
1	1	1	2	2	0	48980
1	1	1	2	3	0	48980
1	1	1	3	1	0	48980
1	1	1	3	2	0	48980

...

Hver linie kan opfattes som de oplysninger der vedrører en bestemt celle i en $15 \times 2 \times 5 \times 11 \times 3$ -antalstabel. Faktorerne AMT, KON, ALDER, TRANSP og TYPE definerer cellen, ANTAL er naturligvis antal ulykker der hører til i den pågældende celle, og BEFOLK er det tilhørende befolkningsunderlag (dvs. antal personer boende i det pågældende amt af det pågældende køn og tilhørende den pågældende aldersgruppe). Betydningen af de 7 oplysninger i hver linie er mere præcist følgende:

AMT

Faktor på 15 niveauer.

Det amt hvor ulykken er sket.

1=Kbh og Frb

2=Københavns Amt

3=Frederiksborg Amt

4=Roskilde Amt

5=Vestsj. Amt

6=Storstrøms Amt

7=Bornholm

8=Fyns Amt

9=Sønderjyllands Amt

10=Ribe Amt

11=Vejle Amt

12=Ringkøbing Amt

13=Århus Amt

14=Viborg Amt

15=Nordjyllands Amt

KON

Faktor på 2 niveauer.

Offerets køn.

1=mand

2=kvinde

ALDER

Faktor på 5 niveauer.

Offerets alder.

1=0-17

2=18-24

3=25-44

4=45-64

5=65-

TRANSP

Faktor på 11 niveauer.

Offerets transportmiddel:

1=Personbil

2=Taxi

3=Udrykningskøretøj

4=Varevogn
5=Lastbil
6=Bus
7=Motorcykel
8=Knallert 45
9=Knallert
10=Cykel
11=Fodgænger

TYPE

Faktor på 3 niveauer.

Grad af tilskadekomst.

1=Dræbt

2=Alvorlig tilskadekommen

3=Lettere tilskadekommen

ANTAL

Antal ulykker i cellen

KON*ALDER*INDBLAND*AMT*UHELD

BEFOLK

Befolkningens størrelse defineret ved

AMT, KON og ALDER.

Eksempel. Af datafilens anden linie fremgår, at 4 drenge under 18 er kommet alvorligt til skade under transport med personbil i området Kbh.+Frb. Bemærk: Der er ingen sikkerhed for at de fire tilskadekomne rent faktisk er blandt de 48980 drenge under 18 som er bosat i området.

Bemærk, at medens data i det væsentlige er organiseret som en femsidede antalstabel, så er oplysningen BEFOLK i virkeligheden en tresidede tabel som er "strukket ud" (ved almindelig gentagelse) til en femsidede.

Projektbesvarelsen afleveres mandag d. 18. oktober.