

Opgave om infektioner

I 18 familier bestående af far, mor og 3 børn (i veldefinerede aldersintervaller, med `child1` som det ældste barn og `child3` som det yngste) har man registreret antallet af infektioner over et bestemt tidsinterval.

Familierne er inddelt efter, hvor tæt, de bor sammen, angivet ved faktoren *crowding*, som har 3 niveauer, *uncrowded*, *crowded* og *overcrowded*.








Data ses i tabellen nedenfor:

Table 8.13. Numbers of swabs positive for pneumococcus during fixed periods

Crowding category	Family serial number	Family status					Total
		Father	Mother	Child			
				1	2	3	
Overcrowded	1	5	7	6	25	19	62
	2	11	8	11	33	35	98
	3	3	12	19	6	21	61
	4	3	19	12	17	17	68
	5	10	9	15	11	17	62
	6	9	0	6	9	5	29
		41	55	69	101	114	380
Crowded	7	11	7	7	15	13	53
	8	10	5	8	13	17	53
	9	5	4	3	18	10	40
	10	1	9	4	16	8	38
	11	5	5	10	16	20	56
	12	7	3	13	17	18	58
		39	33	45	95	86	298
Uncrowded	13	6	3	5	7	3	24
	14	9	6	6	14	10	45
	15	2	2	6	15	8	33
	16	0	2	10	16	21	49
	17	3	2	0	3	14	22
	18	6	2	4	7	20	39
		26	17	31	62	76	212
Total		106	105	145	258	276	890

og de ligger i det såkaldte *lange format* på hjemmesiden.

Vi ønsker at undersøge, om boligforholdene har betydning for antallet af infektioner, samt om visse familiemedlemmer er mere udsat end andre.

1. Indlæs data i det midlertidige datasæt **swabs**, og overvej strukturen i disse:
 - (a) Hvad er outcome? 
 - (b) Hvilke kovariater har vi? 
 - (c) - og hvordan skal effekten af disse beskrives i modellen?
dvs. hvilke er systematiske, og hvilke er tilfældige? 
2. Lav en (eller flere) arbejdstegning(er), der så vidt muligt indeholder al informationen i data.
3. Fit en passende varianskomponentmodel til disse data, dvs. en model, der specificerer samme korrelation mellem alle familiemedlemmer i samme familie.
 - (a) Hvad er estimatet for denne korrelation mellem familiemedlemmer i samme familie? 
 - (b) Giv et estimat (med konfidensinterval) for forskellen på **overcrowded** og **uncrowded**. 
 - (c) Giv også et estimat (med konfidensinterval) for forskellen på ældste og yngste barn. 
4. Er der evidens for forskellig effekt af trange boligforhold for de enkelte familienmedlemmer? 
5. Lav en illustration af de fittede værdier i den model, I ender med.
6. Hvilke betragtninger tror I, der ligger til grund for valget af design i denne undersøgelse? Et alternativ kunne jo være blot at undersøge et tilfældigt udpluk af personer.

Hvis I herefter har tid, skal I prøve at se, om I kan opnå nogle af ovenstående resultater med traditionelle analysemetoder:

7. Udregn gennemsnitligt antal infektioner for hver familie, f.eks. ved at skrive nedenstående kode (overvej ideen i denne, f.eks. ved at se på det resulterende datasæt **averages**)

```
proc sort data=swabs; by family;
run;
proc means noprint data=swabs; by family;
var swabs;
```

```
output out=averages mean=mswabs;  
id crowding;  
run;
```

8. Brug det nydannede datasæt til at vurdere effekten af boligforhold
 - (a) Hvilken type analyse er der tale om her?
Husk en figur til illustration.
 - (b) Giv igen et estimat (med konfidensinterval) for forskellen på **overcrowded** og **uncrowded**, og sammenlign med det ovenfor fundne (spm. 3b).
 - (c) Hvorfor kan vi ikke udfra denne analyse sige noget om forskelle på familiemedlemmer?
9. Nu ser vi et øjeblik væk fra faktoren *crowding* og betragter bare de 18 familier, som om de var tilfældige stikprøver fra samme population. Vi går altså tilbage til det oprindelige datasæt, **swabs**:
 - (a) Lav en tosidet variansanalyse med faktorerne **family** og **name**.
 - (b) Er der evidens for forskelle på de 5 familiemedlemmer?
 - (c) Giv igen et estimat (med konfidensinterval) for forskellen på ældste og yngste barn, og sammenlign med det tidligere fundne (spm. 3c).
10. Kan vi vurdere interaktionen mellem **crowding** og **name** uden at benytte varianskomponentmodellen?
11. Diskuter hvad der sker med de forskellige analysetyper, hvis der mangler nogle observationer.