

18. december 2023

Eksempel på diskretionering af antalstabeller for personstatistik

Vi betragter en population afgrænset ved et givent geografisk område, fx et boligområde. Personerne i populationen tilkøbes oplysninger om herkomst og tilknytning til arbejdsmarkedet. Diskretioneringskravet opfyldes således, såfremt der er minimum 3 eller ingen observationer i det givne geografiske område i alle kombinationer af variablene herkomst og tilknytning til arbejdsmarkedet.

Er ovenstående ikke opfyldt, er der et diskretioneringsproblem, som skal løses ved enten sletning af de pågældende oplysninger eller sammenlægning af undergrupper. Tallene i tabellen er fiktive.

Table 1: Ikke diskretionerede værdier

	Dansk oprindelse	Indvandrere	Efterkommere	Total
Total	124	69	42	235
1 Erhvervsaktive	88	47	33	168
11 Selvstændige	23	7	11	41
12 Medarbejdende ægtefælle	8	5	3	16
13 Lønmodtagere	57	35	19	111
2 Midlertidigt ikke erhvervsaktive	17	9	2	28
21 Arbejdsløs mindst halvdelen af året	12	7	1	20
22 Sygedagpenge, orlov mv.	5	2	1	8
3 Ikke erhvervsaktive	19	13	7	39
31 Under uddannelse	14	7	4	25
32 Pensionist/efterløn	4	3	1	8
33 Kontanthjælp	1	3	2	6

Tabellen ovenfor viser de ikke-diskretionerede værdier, hvor de problematiske værdier er markeret med rødt (celler med under 3 observationer). Variablen for tilknytning til arbejdsmarkedet har en hierarkisk struktur, hvor 11 Selvstændige, 12 Medarbejdende ægtefælle og 13 Lønmodtagere tilsammen giver kategorien 1 Erhvervsaktive, der er markeret med fed. Ligeledes summerer 21 Arbejdsløs mindst halvdelen af året og 22 sygedagpenge, orlov mv. til 2 Midlertidigt ikke erhvervsaktive. Slutteligt udgøres 3 Ikke erhvervsaktive af grupperne 31 Under uddannelse, 32 Pensionist/efterløn og 33 Kontanthjælp. I tabellen nedenfor er de diskretionerede værdier erstattet med to punktummer ("..").

Tabel 2: Diskretionerede værdier

	Dansk oprindelse	Indvandrere	Efterkommere	Total
Total	124	69	42	235
1 Erhvervsaktive	88	47	33	168
11 Selvstændige	23	7	11	41
12 Medarbejdende ægtefælle	8	5	3	16
13 Lønmodtagere	57	35	19	111
2 Midlertidigt ikke erhvervsaktive	..	9	..	28
21 Arbejdsløs mindst halvdelen af året	12	20
22 Sygedagpenge, orlov mv.	8
3 Ikke erhvervsaktive	..	13	..	39
31 Under uddannelse	14	7	4	25
32 Pensionist/efterløn	4	8
33 Kontanthjælp	6

Cellerne, der er markeret med rødt i tabellen ovenfor, er de celler, som er diskretioneret grundet primær diskretionering, fordi de ikke lever op til kravet om mindst 3 i cellen. De orange celler er diskretioneret grundet sekundær diskretionering (også kaldet følgediskretionering) for at sikre, at der ikke kan regnes baglæns og afsløre en værdi, der er diskretioneret grundet primær diskretionering.¹ Således har det været nødvendigt at diskretionere nogle celler, der ellers lever op til kravet om at have en værdi på mindst 3, men dette sker for at beskytte cellen, der har en værdi under 3.

I tabellen oplyses også totaler, hvilket gøres for at bevare mest mulig information til trods for diskretioneringen. Eksempelvis kan antallet af indvandrere, som er enten arbejdsløse eller på sygedagpenge, orlov mv., ikke oplyses. I stedet kan det oplyses, at 9 indvandrere er midlertidigt ikke erhvervsaktive, så det vil sige, at 9 indvandrere fordeles mellem arbejdsløse og sygedagpenge, orlov mv.

For personer med dansk oprindelse og efterkommere kan der af diskretioneringshensyn ikke oplyses, hvor mange der er midlertidigt ikke erhvervsaktive. Dog kan totalen på tværs af herkomstgrupper anvendes. Der er 28 midlertidigt ikke-erhvervsaktive i alt. Vi ved, at 9 af dem er indvandrere, derfor må der være 19 midlertidigt ikke-erhvervsaktive, som fordeler sig mellem dansk oprindelse og efterkommere. På denne måde kan der stadig udledes en vis information fra tabellen ved brug af totalerne, dog med en usikkerhed, så der ikke oplyses informationer om enkeltpersoner.

¹ Farverne på cellerne medtages her udelukkende som uddybende illustration af eksemplet. I publicerede tabeller må celler underlagt diskretion ikke være farvet eller have anden indikation af om cellen er underlagt primær eller sekundær diskretionering, da denne information kan udnyttes til at udlede diskretioneringsbelagte informationer.