

Projekt 9.5 Racefordomme i USA og Simpsons paradoks (B og A)

(Data er hentet fra M. Radelet, "Racial characteristics and imposition of death penalty", *American Sociological Review*, 46 (1981), pp 918-927)

I dette projekt vil vi undersøge *racefordomme* i USA: Bliver de sorte diskrimineret i forhold til de hvide? Fx har det været påstået at retssystemet ikke er så farveblindt som det måske burde være. For at undersøge dette har man kigget på dødstallene for 326 retssager, hvor den anklagede risikerede dødsstraf.

Øvelse 1. Den anklagedes hudfarve

Den følgende krydstabel viser sammenhængen mellem den anklagedes hudfarve og den dom, der blev fældet i retssagen:

Status\Anklaget	Hvid	Sort
Dømt til døden	19	17
Frifundet	141	149

- Udbyg tabellen med totalværdier, og oversæt også tabellen til en procenttabel, der viser hvor stor en andel af de anklagede, der dømmes til døden, den såkaldte *dødsrisiko*.
- Afbild tabellen i passende diagram og skriv en foreløbig konklusion som svar på spørgsmålet: Tyder data fra de amerikanske retssager på at sorte bliver diskrimineret i forhold til hvide?

Øvelse 2. Ofrets hudfarve

I denne øvelse inddrager vi endnu en variabel, nemlig ofrets hudfarve. Måske er juryerne påvirket af ofrets hudfarve i deres domfældelse? Det ville jo være en lige så klar racediskrimination. Datamaterialet opdeles derfor yderligere efter ofrets hudfarve. Det giver anledning til de følgende to deltabeller:

Ofret var hvid			Ofret var sort		
\Anklaget	Hvid	Sort	Status\Anklaget	Hvid	Sort
Dømt til døden	19	11	Dømt til døden	0	6
Frifundet	132	52	Frifundet	9	97

- Gennemfør nu den samme beregninger som i første øvelse for hver af de to krydstabeller: dvs. undersøg dødsrisikoerne i de forskellige tilfælde
- Skriv igen en ny foreløbig konklusion på spørgsmålet: Tyder data fra de amerikanske retssager på at sorte bliver diskrimineret i forhold til hvide?

Overvej om der er overensstemmelse mellem de to konklusioner fra første og anden akt? Hvis ikke, hvad kan da være grunden til at du når til to forskellige konklusioner?

Simpsons paradoks

I det foregående skulle du gerne have set et eksempel på at en konklusion kan vendes, når man slår to deltabeller sammen til en større tabel, dvs. når man ser bort fra en tredje variabels indflydelse på sagen – den såkaldte skjulte variabel. I denne sammenhæng betyder et paradoks at man er overrasket over noget, at resultatet af undersøgelsen strider mod ens umiddelbare forventning: Slår man to delundersøgelser sammen til en større undersøgelse burde det jo ikke ændre på konklusionen.

I første øvelse tog vi kun hensyn til den anklagedes hudfarve ud fra en forventning om at den var afgørende for dødsrisikoen. Det viste sig da at de hvide anklagede faktisk havde en højere dødsrisiko end de sorte anklagede, om end forskellen ikke var så markant. Men for at være sikker på konklusionen er vi nødt til at forudsætte 'alt andet lige' princippet. Der kunne jo være andre faktorer, der havde indflydelse på dødsrisikoen, skjulte variable som vi ikke har inddraget i undersøgelsen. Men hvis vi har sørget for at sammensætningen for de øvrige variable er den samme i de to typer retssager, dem hvor det er en hvid, der er anklaget og dem, hvor det er en sort, der er anklaget, så burde virkningen af disse skjulte variable være den samme i begge tilfælde, og en eventuel forskel burde derfor kunne tilskrives den anklagedes hudfarve. Konklusionen holder altså kun hvis sammensætningen af de to typer retssager alt andet lige er den samme for alle andre variable, som vi ikke har taget hensyn til (dvs. vi har udført variabelkontrol og holdt alle andre variable på samme niveau i de to typer).

Skulle det mod forventning vise sig at sammensætningen af de to typer retssager faktisk er meget forskellige med hensyn til en tredje skjult variabel, så står vi derimod meget dårligt i vores konklusion, for så kunne forskellen i domfældelserne jo lige så godt skyldes ændringen af den skjulte variabel.

I den anden øvelse inddrog vi netop en sådan skjult variabel, ofrets hudfarve, og nu viser der sig pludselig en markant forskel i sammensætningen af de to typer retssager: Der er stort set ingen sorte ofre i de sager, hvor de hvide er på anklagebænken. Hvorfor det er sådan kan også i sig selv vække grund til bekymring: Er det fx sådan at der bare ikke bliver rejst sag i de tilfælde, hvor en sort overfaldes af en hvid? Men det vil vi ikke se nærmere på her. Her holder vi os til data, og når vi inkluderer den skjulte variabel, så tyder data pludselig på at de sortes dødsrisiko er markant større end de hvides.

Hvordan kan det nu være at en sådan skjult variabel kan vende billedet? For at forstå hvordan paradokset kan opstå kan det være en fordel at indføre en simpel model til at forklare hvad der foregår. Vi vil da give to forskellige modeller, en simpel, der viser hvordan man kan konstruere paradokset, og en lidt mere detaljeret, hvor vi både forsøger at forstå oprindelsen til paradokset og få et endeligt svar på spørgsmålet: *Tyder data fra de amerikanske retssager på at sorte bliver diskrimineret i forhold til hvide?*

Fejlslutningen

Vi vender tilbage til tabellerne hvor vi har yderligere opdelt retssagerne efter ofrenes hudfarve

Ofret var hvid				Ofret var sort			
Status\Anklaget	Hvid	Sort	I alt	Status\Anklaget	Hvid	Sort	I alt
Dømt til døden	19	11	30	Dømt til døden	0	6	6
Frifundet	132	52	184	Frifundet	9	97	106
I alt	151	63	214	I alt	9	103	112

Hvidt offer: I følge den første tabel er dødsrisikoen for hvid givet ved $19/151 = 12.6\%$, mens den for sort er givet ved $11/63 = 17.5\%$, dvs. sorts dødsrisiko er størst:

$$p_{\text{sort}|\text{ofret er hvid}} = \frac{11}{63} > \frac{19}{151} = p_{\text{hvid}|\text{ofret er hvid}}$$

Sort offer: Ifølge den anden tabel er dødsrisikoen for hvid givet ved $0/9 = 0\%$, mens den for sort er givet ved $6/103 = 5.8\%$, dvs. igen er sorts dødsrisiko størst:

$$p_{\text{sort}|\text{ofret er sort}} = \frac{6}{103} > \frac{0}{9} = p_{\text{hvid}|\text{ofret er sort}}$$

Men hvorfor kan vi så ikke slutte at det samme så også må gælde for den kombinerede dødsrisiko? Hvis vi fx lagde brøkerne sammen er det jo klart at summen af de to største brøker er større end summen af de to mindste brøker. Men vi lægger ikke brøkerne sammen, når vi kombinerer de to tabeller: Vi lægger tællerne sammen og nævnerne sammen – og det er noget helt andet:

Status\Anklaget	Hvid	Sort	I alt
Dømt til døden	$19 = 19 + 0$	$17 = 11 + 6$	$36 = 30 + 6$
Frifundet	$141 = 132 + 9$	$149 = 52 + 97$	$290 = 184 + 106$
I alt	$160 = 151 + 9$	$166 = 63 + 103$	$326 = 214 + 112$

$$p_{\text{hvid}} = \frac{19+0}{151+9} = \frac{19}{160} > \frac{17}{166} = \frac{11+6}{63+103} = p_{\text{sort}}$$

Øvelse 3. Geometrisk illustration

Dødsrisikoen, dvs. brøken $\frac{\text{antal dømte}}{\text{antal anklagede}}$ kan opfattes som hældningen for den rette linje gennem *Origo* (0,0) og (antal anklagede, antal dømte).

- a) Afsæt i et koordinatsystem antal anklagede ud af førsteaksen og antal dømte op af andenaksen. Afsæt heri punkterne hørende til $Hvid_{\text{samlet}}$ og $Sort_{\text{samlet}}$ fra tabellen nedenfor, dvs afsæt punkterne (160,19) og (166,17)

Status\Anklaget	Hvid	Sort
Dømt til døden	19	17
Frifundet	141	149
I alt	160	166

- b) Hvad er hældningerne for de linjestykkerne, der forbinder *Origo* (0,0) med de afsatte punkter?
- c) Gør det samme med de to deltabeller, hvor den første giver anledning til punkterne $Hvid_{\text{hvidt offer}}$ og $Hvid_{\text{sort offer}}$, mens den anden giver anledning til punkterne $Sort_{\text{hvidt offer}}$ og $Sort_{\text{sort offer}}$.
- d) Hvilken figur er knyttet til de fire punkter *Origo*, $Hvid_{\text{hvidt offer}}$, $Hvid_{\text{sort offer}}$ og $Hvid_{\text{samlet}}$? træk figuren op og farv det indre lysegråt (for hvid).
- e) Samme spørgsmål til de sorte punkter, hvor figuren farves mørkegrå (for sort)!

f) Hvad fortæller ulighederne $P_{\text{sort}|\text{ofret er hvid}} = \frac{11}{63} > \frac{19}{151} = P_{\text{hvid}|\text{ofret er hvid}}$

og $P_{\text{sort}|\text{ofret er sort}} = \frac{6}{103} > \frac{0}{9} = P_{\text{hvid}|\text{ofret er sort}}$

dig om den 'hvide' og den 'sorte' figur?

- g) Prøv nu at forklar med dine egen ord, i hvilken forstand figuren fremstiller Simpsons paradoks.

Bemærkning: Hvis du i stedet opretter de 6 punkter $Hvid_{\text{hvidt offer}}$, $Hvid_{\text{sort offer}}$ og $Hvid_{\text{samlet}}$, $Sort_{\text{hvidt offer}}$, $Sort_{\text{sort offer}}$ og $Sort_{\text{samlet}}$ som frie gitterpunkter med heltallige koordinater har du nu en 'maskine' der hurtigt og nemt kan frembringe andre eksempler på antalstabeller, der illustrerer Simpsons paradoks.

Den skjulte variables rolle

Vi vil nu endelig prøve at forstå den skjulte variables rolle. Det sker igen ved hjælp af en simpel model. Den skjulte variabel handler om ofrenes sammensætning ved de involverede forbrydelser. Det er nemmest at indføre en enkelt variabel x , der netop repræsenterer den brøkdel som de hvide ofre udgør. Hvis der slet ingen hvide ofre er har den værdien 0, hvis der er lige mange hvide ofre og sorte ofre har den værdien $\frac{1}{2}$, hvis der kun er hvide ofre har den værdien 1. Variablen x antager altså værdier mellem 0 og 1.

Øvelse 4

- Hvilken værdi har variabelen x i tilfældet med de hvide anklagede? Denne værdi kaldes x_{hvid} .
- Hvilken værdi har variabelen x i tilfældet med de sorte anklagede? Denne værdi kaldes x_{sort} .

Dernæst har vi dødsrisikoen, som vi vil betegne med p . Den afhænger af ofrenes sammensætning. Hvis vi går ud fra de to deltabeller:

Ofret var hvid				Ofret var sort			
Status\Anklaget	Hvid	Sort	I alt	Status\Anklaget	Hvid	Sort	I alt
Dømt til døden	19	11	30	Dømt til døden	0	6	6
Frifundet	132	52	184	Frifundet	9	97	106
I alt	151	63	214	I alt	9	103	112

kan vi omstrukturere disse til en ny krydstabel, der viser forholdene for de hvide anklagede.

Retssager med en hvid på anklagebænken			
	Hvidt offer	Sort offer	I alt
Dømt til døden	19	0	19
Frifundet	132	9	141
I alt	151	9	160

Vi ser da at dødsrisikoen er $0/9$, hvis der kun er sorte ofre svarende til $x = 0$ og tilsvarende er den $19/151$, hvis der kun er hvide ofre svarende til $x = 1$. Vi kan altså udfylde den følgende tabel over sammenhængen mellem offerbrøken x og dødsrisikoen p :

x = offerbrøken	0 = startværdi (ingen hvide)	1 = slutværdi (kun hvide)
p = dødsrisikoen	$0/9$	$19/151$

Øvelse 5. Sammenligning af dødsrisiko for hvid og sort

- a) Tegn den lineære sammenhæng, der hører til denne tabel i et koordinatsystem, hvor du afsætter offerbrøken x (den skjulte variabel!) ud af første akse og dødsrisikoen p op af andenaksen. Afsæt også den faktiske offerbrøk $x_{\text{hvid}} = 151/160$ på x -aksen og find den tilhørende p -værdi grafisk. Kontroller at den stemmer overens med dine tidligere fundne resultater.
- b) Gentag derefter den samme øvelse med de retssager, hvor der er en sort på anklagebænken.
- c) Du har nu frembragt to lineære sammenhænge i det samme koordinatsystem, hvoraf den ene viser hvordan dødsrisikoen for en hvid anklaget afhænger af offerbrøken, mens den anden viser hvordan dødsrisikoen for en sort anklaget afhænger af offerbrøken.
- d) Brug diagrammet til at forklare Simpsons paradoks: Hvordan er det muligt at hvids dødsrisiko kan ligge højere end sorts dødsrisiko på trods af at hvids graf ligger under sorts graf?
- e) Hvis du korrigerer for de forskellige offerbrøker kan du nu også besvare det følgende spørgsmål ud fra grafen: Hvor meget højere er dødsrisikoen for en sort anklaget end dødsrisikoen for en hvid anklaget – hvis alt andet er lige?