DET TEKNISK-NATURVITENSKAPELIGE FAKULTETET

MASTEROPPGAVE

Studieprogram/spesialisering: Lektorutdanning for 8-13 trinn i realfag.	Vårsemesteret, *2023* Åpen
Forfatter: Ine Sæbø Kaltveit	Ine Sæbø Kaltveit
Veileder: Jan Terje Kvaløy	
Tittel på oppgaven: Paradokser i sannsynlighetsregning og statistikk, og bruk av disse i undervisningssammenheng Engelsk tittel: Paradoxes in probability calculus and statistics, and how to implement them in teaching contexts	
Studiepoeng: 30	
Emneord: Paradokser, sannsynlighetsregning, statistikk, undervisning	Sidetall: 47 inkl. forside og referanseliste Stavanger, *14.06.2023*

Forord

Denne masteroppgaven markerer slutten på mine fem år som lektorstudent ved Universitetet i Stavanger. Det har vært fem lange og krevende år, men jeg sitter igjen med mest glede. I løpet av disse årene har jeg fått enormt med kunnskap, venner for livet og ikke minst et stort ønske om å undervise og videreformidle kunnskap.

Jeg vil først takke mine medstudenter som har hjulpet, støttet og generelt gjort studieløpet gøyere.

Videre skal en stor takk rettes til min veileder, Jan Terje Kvaløy, som både har veiledet meg i bacheloroppgaven min og nå i denne masteroppgaven. Takk for den gode veiledningen og oppfølgingen, samt takk for alt du har lært meg. Dette skal jeg ta med meg ut i arbeidslivet som lærer.

Familie og venner har vært gode ha i denne perioden med masterskriving. Takk for all støtte og gode ord.

Til slutt vil jeg takke min kjære samboer. Du er alltid der når jeg trenger deg som mest, og det setter jeg uendelig stor pris på.

Stavanger, 2023

Ine Sæbø Kaltveit

Sammendrag

I denne oppgaven presenteres ulike paradokser innenfor sannsynlighetsregning og statistikk. Et paradoks er en påstand som er sann, men som virker absurd eller selvmotsigende. Vi kan møte på mange typer paradokser i matematikkens verden og det kan være en fordel å ha kjennskap til disse. Vi bruker statistikk for å undersøke utvalg og deretter trekke konklusjoner om hele populasjonen. I slike undersøkelser kan man havne i ulike fallgruver og på grunn av det trekke gale konklusjoner. Denne oppgaven skal gjøre leseren mer bevisst på slike fallgruver i form av paradokser og se eksempler på hvordan disse løses. Noen av disse eksemplene har jeg utarbeidet selv, andre er hentet fra ulike kilder. Største del av oppgaven omhandler Simpson's paradoks og Berkson's paradoks. I tillegg forklares St.Petersburg paradokset og Monty Hall problemet. Avslutningsvis presenteres det ulike måter å implementere paradokser på i undervisningssammenhenger, samt en forklaring på hvordan Simpson's paradoks kan opptre i undersøkelser i skoleverket.

Innhold

1 Introduksjon

Et paradoks er en påstand som er sann, men som virker absurd eller selvmotsigende. Vi kan møte på flere og ulike typer paradokser i matematikkens verden og i denne oppgaven skal noen paradokser innenfor sannsynlighet og statistikk presenteres. Vi bruker statistikk for å undersøke utvalg og deretter trekke konklusjoner om hele populasjonen. Man kan for eksempel bruke statistiske metoder for å undersøke hvilken behandling som er den mest effektive mot en spesifikk sykdom, om det å bruke hjelm beskytter ATV-førere mot sykehusinnleggelse og om man får bedre karakterer ved å følge med i undervisningen. Men man må ikke stole blindt på all statistikk. Det finnes flere fallgruver å havne i når man arbeider med tall og data, og disse bør både de som utfører statistiske undersøkelser og de som leser dem være klar over.

Først skal jeg i denne oppgaven gjøre rede for Simpson's paradoks. Paradokset går ut på at det finnes forskjeller i undergrupper som lett kan bli oversett eller ikke bli tatt med i analysen og deretter føre til gale konklusjoner. Videre i oppgaven skal Berkson's paradoks diskuteres. Dette paradokset handler om at uavhengige hendelser kan bli avhengige dersom man ser på en delmengde av utfallsrommet og på den måten føre til at det oppstår falske sammenhenger i undersøkelser. Etterpå skal to paradokser som omhandler spill presenteres; St.Petersburg paradokset og Monty Hall problemet. Avslutningsvis presenteres det ulike måter å implementere de nevnte paradoksene på i undervisningssammenhenger, samt en forklaring på hvordan Simpson's paradoks kan opptre i undersøkelser i skoleverket.

2 Simpson's paradoks

Edward Simpson skrev et notat om krysstabeller da han var doktorgradstipendiat i matematisk statistikk i 1945-1947. I 1951 publiserte han notatet i artikkelen *The Interpretation of Interaction in Contingency Tables* [1]. I artikkelen beskrev han sitt teoretiske arbeid om et fenomen han hadde oppdaget [2]. Fenomenet omhandlet det at en effekt pekte i én retning hvis man så på to grupper hver for seg, men pekte i motsatt retning når man så på observasjonene samlet. Notatet inneholdt et hypotetisk eksempel på fenomenet, men paradokset som senere fikk navnet *Simpsons paradoks*, er høyst reelt [2].

2.1 Eksempel fiktive tall

I dette eksempelet er tall fra Simpson's originale artikkel brukt. Eksempelet omhandler sammenhengen mellom behandling og overlevelse. Tallene er fiktive.

Ut fra tabell 1 ser vi at behandling virker ineffektiv hvis man ser på den totale populasjonen. Det er 50 prosent sjanse for suksess/ikke suksess med behandling og uten behandling. Med suksess menes overlevelse [1].

	Total populasjonen (52)		
	Suksess	Ikke suksess	Suksessrate
Ikke behandling	6	6	6/12 = 50%
Behandling	20	20	20/40 = 50%

Tabell 1: Total populasjon.

I tabell 2 har dataene blitt delt inn i undergruppene kvinner og menn. Her ser vi at det er en positiv assosiasjon mellom behandling og overlevelse både blant menn og kvinner.

	Menn (20)			Kvinner (32)		
	Suksess	Ikke suksess	Suksessrate	Suksess	Ikke suksess	Suksessrate
Ikke behandling	4	3	4/7 = 57%	2	3	2/5 = 40%
Behandling	8	5	8/13 = 61,5%	12	15	12/27 = 44%

Tabell 2: Undergrupper.

Dersom dette hadde vært et reelt eksempel, ville det blitt feil å konkludere med at behandlingen var ineffektiv. Ut i fra suksessraten til hele populasjonen kan det se ut som at den er det. Men det viser seg, når man deler inn i undergrupper, at behandlingen faktisk har effekt både på kvinner og menn [1]. I dette eksempelet snus trenden når vi deler inn i undergrupper. Dette kan skyldes en skjevhet i fordelingen av variablene i de ulike undergruppene, som påvirker resultatet i ulik grad. I tabell 2 ser vi at det er ulikt antall i hver gruppe og at det er ulik suksessrate mellom kvinner og menn. Trenden snus og paradokset oppstår på grunn av denne skjevheten.

2.2 Eksempel reelle tall

I dette eksempelet er tallene hentet fra en medisinsk studie [3]. Studien sammenliknet to behandlinger, behandling 1 og behandling 2, for nyrestein. Tabell 3 viser sukessratene og antall behandlede både for små og store nyresteiner. Behandling 1 inkluderte åpne operasjoner og behandling 2 inkluderte lukkede operasjoner. Behandlingen ble regnet som suksessfull dersom nyresteinene ble fjernet eller redusert til mindre enn 2 mm etter tre måneder [3].

	Liten nyrestein	Stor nyrestein	Begge
Behandling 1	81/87 = 93%	192/263 = 73%	273/350 = 78%
Behandling 2	234/270 = 87%	55/80 = 69%	289/350 = 83%

Tabell 3: Medisinsk studie.

Ut i fra tabell 3 ser vi at behandling 1 er mer effektiv når det kommer til små nyresteiner, men også store nyresteiner. Behandling 2 er mest effektiv hvis vi ser på begge størrelsene samtidig. Simpson's paradoks kommer først til syne når vi skiller mellom størrelsene, altså deler inn i undergrupper. Uten å ta hensyn til størrelsen på steinene vil konklusjonen om hvilken behandling som er best bli feil. Hovedårsaken til at suksessraten snudde er fordi sannsynligheten for åpen operasjon, behandling 1, varierte i forhold til størrelsen på nyresteinene. Suksessraten er også forskjellig for små og store steiner, hvor suksessraten er større for små steiner [3]. Steinstørrelsen en er konfunderende variabel, dette kommer vi tilbake til i kapittel 2.6.3. Vi ser i dette eksempelet det samme som skjedde i eksempelet med fiktive tall - trenden snus når vi deler dataene i undergrupper og paradokset oppstår på grunn av skjevhet i variablene. Konklusjonen blir dermed at behandlig 1 virker best, ettersom den er mest effektiv når vi ser på størrelsene hver for seg.

2.3 Brøk

Simpson's paradoks kan forklares ved bruk av brøker. La oss tenke på to dørselgere, person 1 og person 2, som skal selge alarmsystemer. Person 1 gikk til ett hus første uken, men det ble ingen salg. Person 2 gikk til fem hus og fikk solgt et alarmsystem første uken. Uken etterpå gikk person 1 til fem hus og fikk solgt fire systemer, person 2 gikk til ett hus og fikk solgt ett system. Tallene er fremstilt i tabell 4 nedenfor.

	Uke 1	Uke 2
Person 1	**0/1**	**4/5**
Person 2	**1/5**	**1/1**

Tabell 4: Antall salg per uke.

Tabell 4 viser at person 2 hadde større andel salg hver uke. Legger vi totalen for de to ukene sammen, er det derimot motsatt. Ser vi på tallene på denne måten, vil vi se at person 1 har solgt en større andel alarmsystemer. Dette vises i tabell 5 og ved bruk av brøk og ulikheter nedenfor.

	Uke 1	Uke 2	Total
Person 1	**0/1**	**4/5**	**4/6**
Person 2	**1/5**	**1/1**	**2/6**

Tabell 5: Antall salg per uke og totalt.

I oppstillingen av brøkene under ser vi at ulikhetstegnet snur når vi ser på totalen, altså summerer sammen teller med teller og nevner med nevner. Person 1 har totalt solgt en større andel alarmsystemer.

$$\frac{0}{1} < \frac{1}{5}$$

$$\frac{4}{5} < \frac{1}{1}$$

$$\frac{0+4}{5+1} = \frac{4}{6} > \frac{2}{6} = \frac{1+1}{5+1}$$

I dette eksempelet, motsatt fra eksemplene i 2.1 og 2.2, snus trenden når vi ser på totalen. Hadde vi sett på salg per uke ville vi konkludert med at person 2 hadde størst andel salg, men konklusjonen er at person 1 har solgt størst andel alarmsystemer av de to.

2.3.1 Generell forklaring med brøk

Vi kan ha

$$\frac{a_1}{b_1} < \frac{c_1}{d_1}$$

og

$$\frac{a_2}{b_2} < \frac{c_2}{d_2}$$

men samtidig

$$\frac{a_1 + a_2}{b_1 + b_2} > \frac{c_1 + c_2}{d_1 + d_2}$$

Det er viktig å poengtere at $\frac{a_1}{b_1} + \frac{a_2}{b_2} \neq \frac{a_1+a_2}{b_1+b_2}$ men blir slik:

$$\begin{array}{ccccc}
\dfrac{c_1 + c_2}{d_1 + d_2} = & \dfrac{c_1}{d_1} & \dfrac{d_1}{d_1 + d_2} + & \dfrac{c_2}{d_2} & \dfrac{d_2}{d_1 + d_2} \\
\wedge & \vee & & \vee & \\
\dfrac{a_1 + a_2}{b_1 + b_2} = & \dfrac{a_1}{b_1} & \dfrac{b_1}{b_1 + b_2} + & \dfrac{a_2}{b_2} & \dfrac{b_2}{b_1 + b_2}
\end{array}$$

For at ulikhetstegnene skal snu retning, må forholdstallene være ulike. Forholdstallene er i denne generelle forklaringen er følgende:

$$\frac{d_1}{d_1 + d_2}\ ,\ \frac{d_2}{d_1 + d_2},\ \frac{b_1}{b_1 + b_2} \text{ og } \frac{b_2}{b_1 + b_2},\ \text{hvor}$$

$$\frac{d_1}{d_1 + d_2} \neq \frac{b_1}{b_1 + b_2} \text{ og } \frac{d_2}{d_1 + d_2} \neq \frac{b_2}{b_1 + b_2}$$

2.3.2 Bruk av brøker i nyrestein-eksempelet

Som nevnt i 2.3.1 må forholdstallene være ulike for at effekten skal snu. Under blir tallene fra nyrestein-eksempelet fremstilt ved bruk av brøker:

$$\frac{234}{270} < \frac{81}{87}$$

$$\frac{55}{80} < \frac{192}{263}$$

$$\frac{234+55}{270+80} = \frac{289}{350} > \frac{81+192}{87+263} = \frac{273}{350}$$

$$\begin{array}{ccccccccc}
\frac{273}{350} & = & \frac{81}{87} & \times & \mathbf{\frac{87}{350}} & + & \frac{192}{263} & \times & \mathbf{\frac{263}{350}} \\
\wedge & & \vee & & & & \vee & & \\
\frac{289}{350} & = & \frac{234}{270} & \times & \mathbf{\frac{270}{350}} & + & \frac{55}{80} & \times & \mathbf{\frac{80}{350}}
\end{array}$$

Forholdstallene i dette eksempelet er markert med tykk skrift. Vi ser at disse er forskjellige, og det er derfor effekten og ulikhetstegnene snur retning når vi ser på totalen.

For ordens skyld er det verdt å gjenta at i nyrestein-eksempelet blir det tatt hensyn til undergruppene og den rette konklusjonen blir det man ser i disse.

2.3.3 Vektor

Det som skjer i eksemplene med brøk, kan også vises med vektorer. Her snur også ulikhetstegnet retning når vi summerer sammen totalen [4].

Vi kan ha fire vektorer, s står for strekning, hvor

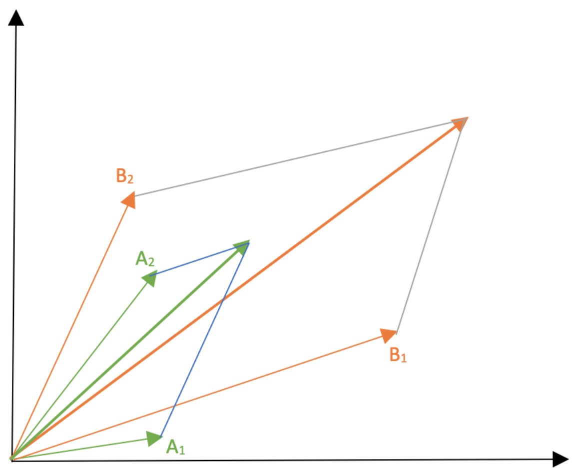
$$s(A_1) < s(B_1)$$

og

$$s(A_2) < s(b_2)$$

men samtidig

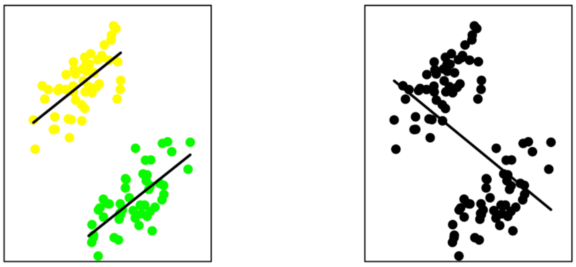
$$s(A_1 + A_2) > s(B_1 + B_2)$$



Figur 1: Simpson's paradoks fremstilt med vektorer.

Figur 1 viser hvordan Simpson's paradoks kan forekomme ved bruk av vektorer. Vektor A1 har en mindre helning enn vektor B1, samme gjelder for vektor A2 og B2. Likevel kan summen av vektorene A1 og A2 ha en større helning enn summen av vektorene B1 og B2 som vist i figur 1. For at dette skal kunne skje må en av de grønne vektorene ha en større helning enn en av de oransje vektorene. I dette tilfellet har den grønne vektoren A2 en større helning enn den oransje vektoren B1. I tillegg må vektorene strekke seg over ulike lengder langs x-aksen.

2.4 Eksempel med simulerte data



Figur 2: Simulerte data.

Figur 2 viser en grafisk fremstilling av simulerte data. Første bilde viser at når vi ser på gruppene hver for seg peker effekten én vei, mens det andre bildet viser at effekten peker motsatt vei når vi ser på dataene samlet [2].

2.5 Sannsynlighetsregning

I dette delkapitlet skal vi se på hvordan Simpson's paradoks kan forklares med sannsynlighetsregning.

Det er mulig å ha

$$P(A|B) < P(A|B^c) \tag{1}$$

og samtidig ha begge

$$P(A|B \cap C) > P(A|B^c \cap C) \tag{2}$$

og

$$P(A|B \cap C^c) > P(A|B^c \cap C^c) \tag{3}$$

I likning (1) ser man at sannsynligheten for A gitt B er mindre enn sannsynligheten for A gitt ikke B. I likning (2) og (3) snur ulikhetstegnene vei. Årsaken til dette er at B og C er ulike og avhengige, det ville ikke skjedd hvis de var like. Videre har vi loven om total sannsynlighet [5]:

$$
\begin{aligned}
P(A|B) =& P(A|B \cap C) * P(C|B) + P(A|B \cap C^c) * P(C^c|B) \\
\wedge \quad & \vee \qquad\qquad\qquad\qquad\qquad\qquad \vee \\
P(A|B^c) =& P(A|B^c \cap C) * P(C|B^c) + P(A|B^c \cap C^c) * P(C^c|B^c)
\end{aligned}
$$

Simpson's paradoks vil kun forekomme dersom C og B er avhengige [5]. Det vil si at sannsynligheten for hendelse C blir påvirket av om hendelse B inntreffer [6]. Dersom C og B hadde vært uavhengige slik at

$$
\begin{aligned}
P(C|B) &= P(C) \\
P(C|B^c) &= P(C)
\end{aligned}
$$

og

$$
\begin{aligned}
P(C^c|B) &= P(C^c) \\
P(C^c|B^c) &= P(C^c)
\end{aligned}
$$

kunne ikke paradokset oppstått.

2.5.1 Sannsynlighetsregning med nyrestein-eksempelet

Tall hentet fra eksempelet om nyrestein. A = suksess, B = behandling 1 og C = stor nyrestein. Da får vi disse estimerte sannsynlighetene:

$$P(A|B) = 0.78 < P(A|B^c) = 0.83$$

$$P(A|B \cap C) = 0.73 > P(A|B^c \cap C) = 0.69$$

$$P(A|B \cap C^c) = 0.93 > P(A|B^c \cap C^c) = 0.87$$

Loven om total sannsynlighet gir:

$$
\begin{aligned}
0.78 =& 0.93 * 0.25 + 0.73 * 0.75 \\
\wedge \quad & \vee \qquad\qquad\quad \vee \\
0.83 =& 0.87 * 0.77 + 0.69 * 0.23
\end{aligned}
$$

Dersom B og C hadde vært uavhengige hendelser, ville ikke paradokset oppstått. Dette på grunn av at paradokset oppstår ved interaksjon/avhengighet mellom B og C, som nevnt tidligere. Dette på lik linje som at forholdstallene i brøk, sett i delkapittel 2.3.1 og 2.3.2, må være ulike for at en effekt skal skifte retning.

I dette tilfellet betyr det at størrelsen på nyresteinen påvirker hvilken behandlingsform pasienten får. Vi ser at sannsynligheten for suksess gitt behandling 1 er mindre enn sannsynligheten for suksess gitt behandling 2. Dette endrer seg, som nevnt, når vi ser på steinstørrelsene hver for seg. Sannsynligheten for suksess både for liten nyrestein og for stor nyrestein er større hvis man benytter behandling 1. I dette tilfellet var C = stor nyrestein, og man konkluderer med at B = behandling 1 gir størst A = suksess.

2.6 Gruppenivå eller overordnet nivå?

Eksemplene i delkapitlene 2.1 og 2.2 har vist oss at retningen på effekten endrer seg når vi deler inn i undergrupper og at den rette konklusjonen blir det man ser i undergruppene. Eksempelet om dørselgere i delkapittel 2.3 viste oss at den rette konklusjonen blir det man ser når all data er samlet. Hvordan kan vi vite om det er gruppenivå eller overordnet nivå som gir oss den riktige konklusjonen? Dette skal vi se nærmere på i dette delkapitlet, hvor årsak og virkning trekkes inn.

2.6.1 Lav fødselsvekt paradoks

I 1950 fant forskere ut at dersom mødre røyker under svangerskapet vil barnet kunne få lav fødselsvekt [7]. Og i løpet av 1960 fant de også bevis på at barn med røykende mødre hadde høyere sannsynlighet for spedbarnsdødelighet. Men effekten morens røyking hadde på spedbarnsdød hadde en spesiell vri. Det viste seg nemlig at barn med lav fødselsvekt og røykende mødre hadde mindre sannsynlighet for spedbarnsdød enn barn med lav fødselsvekt og ikke-røykende mødre. Yerushalmy var en fremtredende epidemiolog som argumenterte for at røyking ikke var skadelig for barnet, basert på denne teorien. Med første øyekast kunne det se ut som at det var positivt at moren røykte [7].

Før vi konkluderer med at røyking er positivt bør vi se nærmere på årsak og virkning. Simpson's paradoks kan i dette tilfellet forklares statistisk ved å avdekke retning på forholdet mellom røyking og de to andre variablene: lav fødselsvekt og spedbarnsdødelighet. Både lav fødselsvekt og spedbarnsdødelighet påvirkes uavhengig av røyking og andre ugunstige forhold.

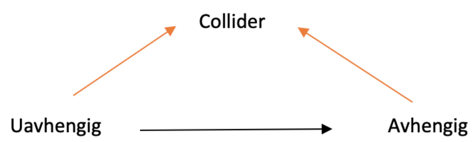
Fødselsvekten blir lavere og risiko for død blir høyere. Imidlertid påvirker ikke hver tilstand variablene i samme grad [7].

Hvis en mor røyker, er det morens handlinger som kan føre til at et barn blir født med lav vekt. Derfor kan friske barn, som ellers ville blitt født med høyere vekt hvis det ikke var for morens røyking, bli født undervektige. Disse barna vil ha lavere sannsynlighet for spedbarnsdødelighet enn barn med lav fødselsvekt av andre medisinske grunner [7].

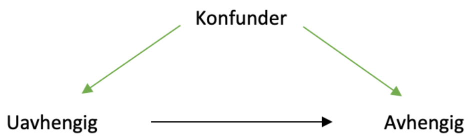
For å oppsummere kan vi si at røyking er skadelig i form av at det kan føre til lav fødselsvekt hos barn, som igjen fører til større sannsynlighet for spedbarnsdødelighet. Men andre tilfeller hvor barn blir født med lav fødselsvekt på grunn av medisinske årsaker er generelt sett enda mer farlig enn røyking [7].

2.6.2 Årsak-virkning

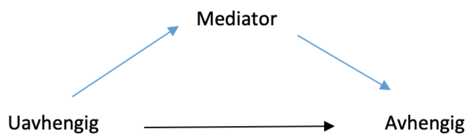
Når vi snakker om årsak-virkning forholder vi oss til to variabler. Én uavhengig variabel og én avhenging variabel. Én ting fører til at én annen ting skjer. Går vi fra to til tre variabler kan det oppstå problemer og alt kan skje [8]. Hva som blir det riktige å konkludere med avhenger av situasjonen. Den tredje variabelen kan ha en relasjon til de to første variablene. Dette kan skje på tre ulike måter: Den tredje variabelen kan være en avhengig variabel for begge, se figur 3. Den kan være en uavhengig variabel for begge, se figur 4. Og den kan være en avhengig variabel for den ene og uavhengig for den andre, se figur 5.



Figur 3: Collider.



Figur 4: Konfunder.



Figur 5: Mediator.

Hva som er rette analyser avhenger av hvilke vei pilene går, men dataene kan ikke vise hvilke vei de skal gå. Det må den som utfører analysen selv vite [8]. Derfor er det viktig å ha kjennskap til de nevnte situasjonene og disse begrepene: collider, konfunder og mediator.

I figur 3 kalles den tredje variabelen en collider, og den er som nevnt en avhengig variabel for de to andre. Den påvirker ikke sammenhengen mellom den uavhengige og den avhengige variabelen, og skal derfor ikke være med i den statistiske analysen [8]. En collider kan skape problemer hvis man ubevisst tar den med i analysen, og feilaktig justerer analysen med hensyn på den. Dette er en bekymring i for eksempel covid-19 studier [8].

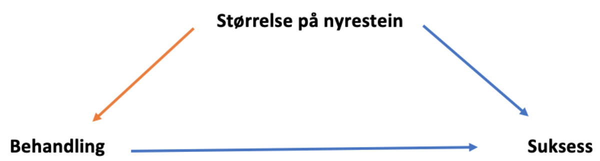
I figur 4 kalles den tredje variabelen for konfunder. En slik variabel påvirker den avhengige variabelen, samt assosieres med, men blir ikke påvirket av, den uavhengige variabelen. Når vi har et tilfelle med en konfunderende variabel er det nødvendig å justere analysen slik at vi ikke trekker gale konklusjoner [8].

I figur 5 kalles den tredje variabelen for en mediator. Når det oppstår en situasjon med mediering må vi tenke på formålet med analysen før en konklusjon trekkes. Vi må avklare om det er den direkte, indirekte eller totale effekten vi er på jakt etter [8].

Oppsummert kan vi si at dersom den tredje variabelen er en collider skal man ikke justere, er den en kollider må man justere og møter vi på en mediator vil rett analyse avhenge av hva vi vil med analysen [8].

2.6.3 Årsak-virkning i nyrestein-eksempelet

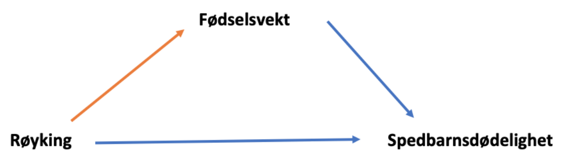
I nyrestein-eksempelet var størrelsen en konfunderende variabel. Størrelsen på nyresteinen påvirker suksessraten, og assosieres med type behandling, se figur 6. I dette tilfellet vil det være nødvendig å justere analysen med hensyn på størrelsen på nyresteinen. Derfor blir den rette analysen her å se på resultatene i hver gruppe, altså konklusjonen trekkes ut i fra det vi ser på gruppenivå.



Figur 6: Størrelse som konfunderende variabel.

2.6.4 Årsak-virkning i lav fødselsvekt paradokset

I lav fødselsvekt paradokset var lav fødselsvekt en mediator. Når det oppstår en situasjon med mediering må vi, som nevnt tidligere, tenke på formålet med analysen før en konklusjon trekkes [8]. Røyking påvirker spedbarnsdødelighet indirekte via lav fødselsvekt, samtidig som røyking påvirker spedbarnsdødelighet direkte, se figur 7. For å se den totale effekten vil man i dette tilfellet ikke justere for mediatoren. Den rette analysen for å vurdere den totale effekten av røyking vil derfor være å se på dataene fra de to fødselsvektgruppene slått sammen, altså konklusjonen trekkes ut i fra det vi ser på overordnet nivå.



Figur 7: Fødselsvekt som en mediator.

2.7 Oppsummering Simpson's paradoks

Vi har i eksemplene i dette kapitlet sett at man fort kan trekke gale konklusjoner hvis man ikke er klar over Simpson's paradoks. Det å utelate en viktig variabel kan resultere i et helt feil bilde og gale konklusjoner. Vi har også sett på viktigheten av å bruke tid på å forstå hvordan de ulike variablene i en undersøkelse henger sammen for å kunne gjøre rett analyse. For eksempel kunne personer med nyrestein fått dårligere behandling dersom det ikke ble tatt hensyn til steinstørrelse når man valgte behandling, eller man kunne konkludert med at røyking i svangerskapet er positivt for barnet. Å ha riktig kunnskap og gjøre riktige tolkninger er viktig når man arbeider med tall og data.

3 Berkson's paradoks

Joseph Berkson beskrev først paradokset som har fått navnet Berkson's paradoks i en artikkel i 1946 [9]. Paradokset handler om at to uavhengige hendelser A og B kan bli avhengige dersom vi bare ser på en delmengde av utfallsrommet. Det er verdt å nevne at Berkson's paradoks også handler om situasjoner hvor datainnsamlingen naturlig gir en skjevhet i utvalget. Paradokset blir ofte beskrevet innenfor medisinsk statistikk, slik som den opprinnelige beskrivelsen til Berkson, men kan også forekomme i hverdagssammenhenger.

3.1 Berkson's originale beskrivelse

Berkson beskrev en type skjevhet der ikke-relaterte sykdommer falskt blir assosiert sammen. I en medisinsk studie ble diabetes assosiert med kolecystitt (betennelse i galleblæren) blant innlagte sykehuspasienter [9]. Selv om de to sykdommene er uavhengige i samfunnet, ble det observert en sammenheng mellom disse blant pasienter innlagt på sykehus. Denne falske sammenhengen oppstod fordi pasienter med flere sykdommer har større sannsynlighet for å bli innlagt enn pasienter med bare én enkelt sykdom og pasienter uten noen av disse to sykdommene har mindre sannsynlighet for å bli innlagt enn de med minst én [9]. Prøver ble kun tatt fra innlagte pasienter på sykehus og ikke fra hele befolkningen, dette ga et ikke-representativt utvalg.

Før vi går videre vil jeg kort forklare hva representative og ikke-representative utvalg er. For at resultatet av en statistisk undersøkelse skal være riktig må de innsamlede dataene komme fra et representativt utvalg [10]. For å få et representativt utvalg må man trekke et tilfeldig utvalg fra populasjonen. Dersom utvalget er tilfeldig trukket fra populasjonen kan man trekke konklusjoner om hele populasjonen [10]. Har man et utvalg fra en bestemt delmengde av populasjonen, har man et ikke-representativt utvalg, og kan ikke trekke konklusjoner angående hele populasjonen [10].

Figur 8 nedenfor viser en tabell hvor A, B, C og D er antall personer i et stort representativt utvalg. A representerer antall personer som har både diabetes og kolecystitt. B representerer antall personer som ikke har diabetes, men har kolecystitt. C representerer antall personer som har diabetes, men ikke kolecystitt. Og D representerer antall personer som ikke har noen av sykdommene. Utfallsrommet er hele befolkningen.

	Har diabetes	Har ikke diabetes
Har kolecystitt	**A**	**B**
Har ikke kolecystitt	**C**	**D**

Figur 8: Hele utfallsrommet.

Vi får følgende:
Sannsynligheten for å ha kolecystitt blant de som har diabetes

$$\frac{A}{A+C}$$

Sannsynligheten for å ha kolecystitt blant de som ikke har diabetes

$$\frac{B}{B+D}$$

Dersom hendelsene er uavhengige får vi:

$$\frac{A}{A+C} = \frac{B}{B+D}$$

Det er ingen sammenheng mellom sykdommene og de påvirker ikke hverandre. Dette dersom utvalget er et tilfeldig utvalg fra hele befolkningen.

I den medisinske studien Berkson analyserte ble det som nevnt observert sammenheng mellom sykdommene blant pasienter innlagt på sykehus. Studien baserte seg på tall fra en delmengde av befolkningen. Under vises en ny figur hvor antallet D har blitt erstattet med antallet D*. D* vil si pasienter innlagt på sykehus uten noen av sykdommene. D* vil være mindre enn D, ettersom D* kun gjelder innlagte pasienter og ikke personer fra hele befolkningen. For illustrasjonens skyld antar vi at A, B og C er som før, det vil si at vi har samlet et ca. like stort utvalg som før. I praksis kan det være flere ulike ting som skjer, men her blir det tatt utgangspunkt i D.

	Har diabetes	Har ikke diabetes
Har kolecystitt	**A**	**B**
Har ikke kolecystitt	**C**	**D***

Figur 9: Delmengde av utfallsrommet.

Hvis D $\rightarrow$ D* og D* $<$ D får vi:

$$\frac{A}{A+C} < \frac{B}{B+D^*}$$

Sannsynligheten for å ha kolecystitt blant de som ikke har diabetes blir større enn sannsynligheten for å ha kolecystitt blant de med diabetes dersom D $\rightarrow$ D* $<$ D. Det vil dermed, feilaktig, se ut som at å ha diabetes beskytter mot kolecystitt, selv om det i realiteten ikke finnes noen sammenheng mellom de to sykdommene.

3.2 Berkson's paradoks ATV-ulykker

Woodfine og Redelmeier [9] beskrev et annet eksempel på Berkson's paradoks innenfor medisinsk statistikk i 2015. Deres studie omhandlet barn og unge som fikk akutt medisinsk hjelp etter ATV-ulykker mellom 1. april 2002 og 31. mars 2014 i Canada. Studien omfattet 35 202 barn og unge. 28 prosent av dem brukte hjelm på skadetidspunktet, mens de resterende 72 prosentene brukte ikke hjelm [9]. Det kom frem i studien at bruk av hjelm var assosiert med større skadealvorlighet, og at det hadde gått bedre med dem som ikke brukte hjelm. Berkson's paradoks er en sannsynlig forklaring på disse funnene. Dataene fokuserte på barn og unge som fikk akutt hjelp etter en ulykke, men manglet de andre tilfellene hvor hjelmen beskyttet så godt at det ikke var nødvendig med medisinsk hjelp eller tilfellene hvor noen døde på stedet fordi de ikke hadde hjelm på [9]. Tallene brukt i figurene under er hentet fra Woodfine og Redelmeier sin artikkel [9].

	Hjelm	Ikke hjelm
Innlagt på sykehus	**945**	**1652**
Ikke innlagt på sykehus	**8917**	**23688**

Figur 10: Observerte ulykker.

I figur 10 ser vi at andelen av de som bruker hjelm og som blir innlagt på sykehus er større enn de som ikke bruker hjelm og som blir innlagt.

$$\text{Med hjelm: } \frac{945}{945 + 8917} = \frac{945}{9862} = 0.096$$

$$\text{Uten hjelm: } \frac{1652}{1652 + 23688} = \frac{1652}{25340} = 0.065$$

Tallene viser at 9,6 prosent av dem med hjelm ble innlagt på sykehus, mens bare 6,5 prosent av dem som ikke brukte hjelm ble innlagt. Det kan se ut som at flere unngår alvorlige ulykker dersom de ikke bruker hjelm.

	Hjelm	Ikke hjelm
Innlagt på sykehus	**0**	**0**
Ikke innlagt på sykehus	**8917**	**0**

Figur 11: Ikke observerte ulykker.

Det kan være rimelig å tenke at antall ulykker, hvor personer har brukt hjelm og det har gått så bra at de ikke trenger medisinsk hjelp, er vesentlig høyere enn antallet i figur 10. I figur 11 vises det et fiktivt tenkt tall på uobserverte ulykker hvor noen har brukt hjelm, men som ikke har hatt behov for akutt hjelp. Dersom vi legger til disse fører det til at antallet ikke innlagt blant de med hjelm dobles i totalen, som man ser i figur 12 nedenfor.

	Hjelm	Ikke hjelm
Innlagt på sykehus	**945**	**1652**
Ikke innlagt på sykehus	**17 834**	**23688**

Figur 12: Total ulykker.

I figur 12 vises den hypotetiske totalen av antall ulykker, både observerte og uobserverte. Antallet personer som bruker hjelm og ikke blir innlagt har doblet seg når vi legger observerte og ikke observerte ulykker sammen. Dette gir trolig et mer realistisk bilde. Med disse tallene blir andeler innlagt:

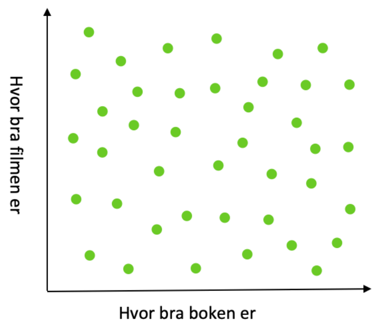
$$\text{Med hjelm: } \frac{945}{945 + 17834} = \frac{945}{18779} = 0.05$$

$$\text{Uten hjelm: } \frac{1652}{1652 + 23688} = \frac{1652}{25340} = 0.065$$

Ved å gjøre en justering av tallene og ta med uobserverte ulykker får vi at 5 prosent av dem som brukte hjelm ble innlagt på sykehus, mens 6,5 prosent av dem uten hjelm ble innlagt. Mangel på kunnskap om Berkson's paradoks i dette tilfellet kan føre til gale konklusjoner og potensielt farlige situasjoner. En undersøkelse som viser at det er sikrere å kjøre ATV uten hjelm kan avskrekke førere fra å bruke hjelm [9]. Derfor er det viktig å behandle data på riktig måte, spesielt i medisinsk statistikk.

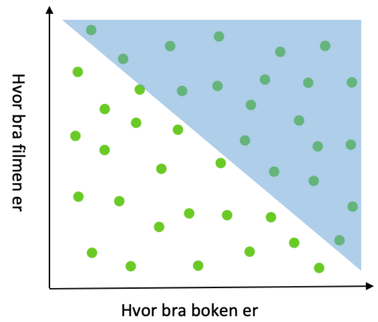
3.3 Berkson's paradoks i hverdagssammenheng

Så langt i kapitlet har vi sett på Berkson's paradoks i medisinsk statistikk. Berkson's paradoks dukker også opp i mindre formelle sammenhenger, som for eksempel i hverdagen vår. I figur 13 vises en fiktiv fremstilling av hva ulike personer synes om ulike bøker og filmer for situasjoner hvor det er skrevet bok og laget film om samme historie. Dette eksempelet er inspirert av en video funnet på nett [11]. La oss anta at det ikke er sammenheng mellom de to variablene. Da vil de grønne prikkene i figuren være fordelt tilfeldig og over alt.



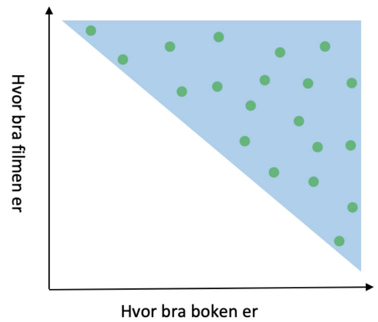
Figur 13: Tilfeldig fordeling.

I figur 14 nedenfor er bra filmer, bra bøker eller begge markert med blått. Dette er en delmengde av alle filmene og bøkene som eksisterer i verden. Hvorfor markeres disse? Jo, fordi det som oftest er disse filmene og bøkene som dukker opp i bevisstheten vår når vi tenker på filmer og bøker, pluss det er også disse som får mest publisitet og som det dermed er størst sjans for at vi velger å lese eller se. De grønne prikkene under den blå trekanten representerer bøker og filmer (basert på bøker) som er så dårlige at vi ikke har hørt om dem eller ikke har hatt lyst til å lese eller se.



Figur 14: Delmengde av filmer og bøker.

I figur 15 er de grønne prikkene under den blå trekanten tatt bort. Det kan nå se ut som at det er en negativ korrelasjon mellom de to variablene, altså at hvor bra en bok er er negativt relatert til hvor bra en film er.



Figur 15: Negativ korrelasjon.

Her dukker Berkson's paradoks opp. Det har blitt en feilaktig sammenheng mellom to variabler på grunn av at man ser på en delmengde av alle filmer og bøker som eksisterer i verden. Det er nødvendigvis ikke sant at det blir laget dårlige filmer av gode bøker.

3.4 Berkson's paradoks i skolesammenheng

I dette delkapitlet skal vi se på et tenkt eksempel som omhandler skolen. Eksempelet er inspirert av en video hentet på nett [12].

Er det slik at elever som ikke følger med eller ikke jobber med faget i timen får bedre karakterer? I realiteten vil nok elever som ikke følger med i undervisningen få dårligere karakterer enn de som følger med, men dette eksempelet er laget for å vise at denne tanken eller konklusjonen kan oppstå dersom man ikke er klar over Berkson's paradoks.

La oss tenke at man skal gjennomføre en spørreundersøkelse hvor elevene skal avgi sine svar på skolens digitale arena og at denne undersøkelsen ikke er obligatorisk å svare på. Spørsmålene i undersøkelsen skal dreie seg om motivasjon, om elevene følger med i timen og om de får gode eller dårlige karakterer.

Elevene som både er umotiverte og uengasjerte vil mest sannsynlig ikke svare på undersøkelsen, og vil dermed ikke være representert i undersøkelsen. Velger noen av disse å svare så vil de i så fall være representert i svært liten grad. Dette fører til at de elevene som får gode karakterer, men som ikke følger med i timen vil være overrepresentert i og med at noen av disse sannsynligvis vil svare på undersøkelsen. Tar man utgangspunkt i dette, uten å ha kjennskap til Berkson's paradoks, vil konklusjonen kunne bli at det å ikke følge med i timen gir gode karakterer.

La S = svarer på undersøkelsen, F = følger med i timen, G = gode karakterer. Etter undersøkelsen får vi tilgang til dataene og kan finne antall svar fra elevene innenfor de ulike gruppene, la oss anta at det fordeler seg slik:

$$
\begin{aligned}
\#(F \cap G) &= 40 \\
\#(F^c \cap G) &= 6 \\
\#(F \cap G^c) &= 12 \\
\#(F^c \cap G^c) &= 1
\end{aligned}
$$

Vi ser ut i fra dette at det er totalt 59 elever som har svart på undersøkelsen, og får følgende:

$$
\begin{aligned}
P(F \cap G) &= \frac{40}{59} = 0.678 \\
P(F^c \cap G) &= \frac{6}{59} = 0.102 \\
P(F \cap G^c) &= \frac{12}{59} = 0.203 \\
P(F^c \cap G^c) &= \frac{1}{59} = 0.017
\end{aligned}
$$

Vi kan med denne informasjonen også finne ut sannsynligheten for F - følger med i timen, og F^c - følger ikke med i timen:

$$
P(F) = P(F \cap G) + P(F \cap G^c) = 0.678 + 0.203 = 0.88
$$
$$
P(F^c) = 1 - 0.88 = 0.12
$$

Til slutt kan vi bruke dette for å finne sannsynlighetene for $P(G|F)$ - gode karakterer gitt at man følger med i timen, og $P(G|F^c)$ - gode karakterer gitt at man ikke følger med i timen:

$$
P(G|F) = \frac{P(F \cap G)}{P(F)} = \frac{0.678}{0.88} = 0.77
$$

$$
P(G|F^c) = \frac{P(F^c \cap G)}{P(F^c)} = \frac{0.102}{0.12} = 0.85
$$

Sannsynlighetene forteller oss at det er anslått å være 85 prosent sjanse for å få gode karakterer dersom man ikke følger med i undervisningen, mot 77 prosent sjanse for å få gode karakterer dersom man følger med i undervisningen. Dersom en elev hadde kommet over denne undersøkelsen med dette resultatet hadde nok eleven tenkt at det er lurt å ikke følge med. Men her skal vi trekke inn Berkson's paradoks: Elevene som ikke følger med i timen og får dårlige karakterer er ikke representert eller i svært liten grad representert i dataene. I tallene vist ovenfor har det ikke blitt tatt høyde for de ikke- eller underrepresenterte elevene og sannsynlighetene blir feil.

Vi kan anslå den riktige totalen/konklusjonen ved å estimere andel svar innenfor de ulike gruppene. Før undersøkelsen estimerer vi at sannsynlighetene vil fordele seg slik:

$$
\begin{aligned}
P(S|F \cap G) &= 0.8 \\
P(S|F^c \cap G) &= 0.4 \\
P(S|F \cap G^c) &= 0.6 \\
P(S|F^c \cap G^c) &= 0.2
\end{aligned}
$$

Vi kan nå anslå det riktige antallet elever i hver gruppe (totalen) ved å ta utgangspunkt i de estimerte sannsynlighetene og antall svar fra elevene. Dette gjøres ved å ta antall svar delt på den estimerte sannsynligheten, første riktige antall blir da: $\frac{40}{0.8} = 50$. Vi får følgende:

$$
\begin{aligned}
\#(F \cap G) &= 50 \\
\#(F^c \cap G) &= 15 \\
\#(F \cap G^c) &= 20 \\
\#(F^c \cap G^c) &= 5
\end{aligned}
$$

Vi ser ut i fra disse tallene at det er totalt 90 elever som har fått tilgang til undersøkelsen, men bare 59 valgte å svare. Ut i fra denne informasjonen får vi:

$$
\begin{aligned}
P(F \cap G) &= \frac{50}{90} = 0.55 \\
P(F^c \cap G) &= \frac{15}{90} = 0.167 \\
P(F \cap G^c) &= \frac{20}{90} = 0.22 \\
P(F^c \cap G^c) &= \frac{5}{90} = 0.056
\end{aligned}
$$

Vi finner sannsynligheten for F - følger med i timen, og F^c - følger ikke med i timen:

$$
\begin{aligned}
&P(F) = P(F \cap G) + P(F \cap G^c) = 0.55 + 0.22 = 0.77 \\
&P(F^c) = 1 - 0.77 = 0.33
\end{aligned}
$$

Til slutt kan vi bruke dette for å få de riktige sannsynlighetene for $P(G|F)$ - gode karakterer gitt at man følger med i timen, og $P(G|F^c)$ - gode karakterer gitt at man ikke følger med i timen:

$$P(G|F) = \frac{P(F \cap G)}{P(F)} = \frac{0.55}{0.77} = 0.71$$

$$P(G|F^c) = \frac{P(F^c \cap G)}{P(F^c)} = \frac{0.167}{0.33} = 0.51$$

Her ser man at sannsynligheten for å få gode karakterer gitt at man følger med i timen er 71 prosent, og at sannsynligheten for å få gode karakterer gitt at man ikke følger med i timen er 51 prosent. Ut i fra justeringen som har blitt gjort, ser vi at det lønner seg for elever å følge med i timene. Dette eksempelet har illustrert at undersøkelser kan gi feil informasjon og viktigheten med representative data for å få riktig analyse.

3.5 Berkson's paradoks med sannsynlighetsregning

Vi kan illustrere eksemplene i delkapittel 3.1 og 3.2 ved bruk av sannsynlighetsregning og venndiagram.

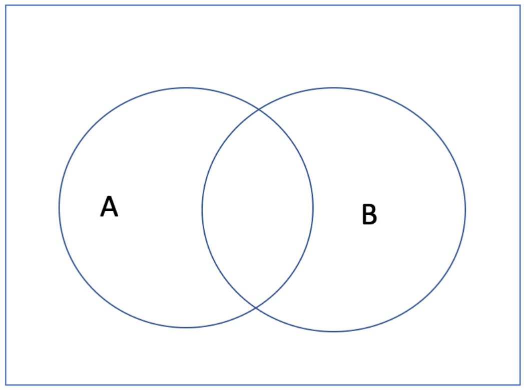
Generelt, hvis A og B er uavhengige hendelser vil sannsynligheten for disse være slik:

$$P(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Vi kan også illustrere dette ved bruk av venndiagram, se figur 16. Illustrert med venndiagram kan vi tenke oss at dette betyr at

$$\frac{\text{areal på A}}{\text{areal på utfallsrom}} = \frac{\text{areal på } A \cap B}{\text{areal på B}}$$

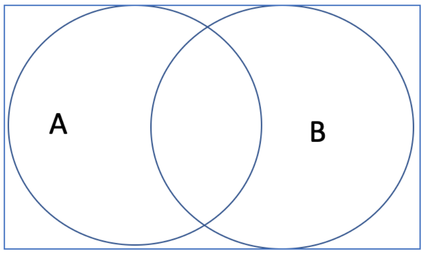
Det vil si, A utgjør en like stor andel av hele utfallsrommet som $A \cap B$ gjør av B.



Figur 16: Venndiagram uavhengige hendelser.

Men, dersom den delen av utfallsrommet som ikke inkluderer A og B blir mindre, vil $P(A)$ øke og $P(A|B)$ være uendret, se venndiagram i figur 17:

$$P(A) > P(A|B)$$



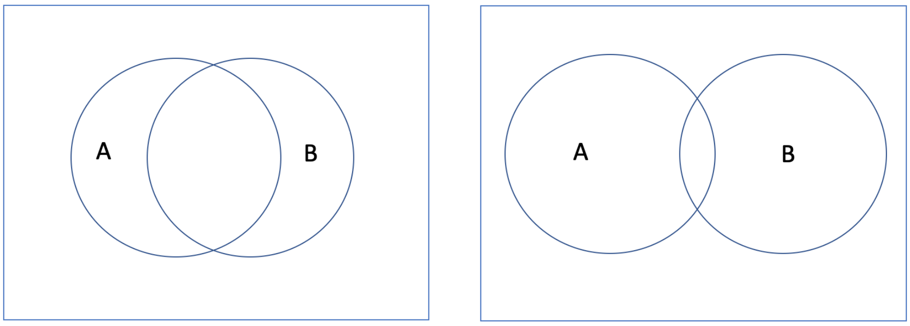
Figur 17: Mindre $(A \cup B)^c$.

I eksempelet i delkapittel 3.1 så vi hvordan to uavhengige sykdommer feilaktig ble avhengige når det ble sett på en delmengde av befolkningen. Figur 17 illustrerer hva som skjedde i eksempelet fra 3.1. Størrelsen på delen av utfallsrommet hvor ingen av sykdommene opptrådte var for lite, ettersom undersøkelsen kun var basert på tall fra innlagte pasienter og ikke fra hele befolkningen.

Vi kan også møte på tilfeller hvor $A \cap B$ øker. Dette fører til at $P(A|B)$ øker og $P(A)$ er uendret.

$$P(A) < P(A|B)$$

Dette kan også illustreres med et venndiagram, hvor $A \cap B$ er stor. I eksempelet i delkapittel 3.2 kunne det først se ut som at det å ikke bruke hjelm når man kjører ATV ga mindre sykehusinnleggelser enn hvis man brukte hjelm på skadetidspunktet. Anta at figur 16 illustrerer dette. Etter en justering av tallene ble konklusjonen at det å bruke hjelm ga mindre sykehusinnleggelser enn det å ikke bruke hjelm. Hvis A = ikke innlagt på sykhus, og B = hjelm, vil figur 18 med stort snitt, $A \cap B$, være den riktige illustrasjonen. For syns skyld har det blitt lagt til et venndiagram hvor $A \cap B$ er liten, men som ellers er lik, som illustrerer dataene man får når man bare baserer seg på tall fra de som har vært i kontakt med helsevesenet, se figur 18:



Figur 18: Endring i $A \cap B$.

3.6 Oppsummering Berkson's paradoks

Berkson's paradoks beskriver situasjoner hvor falske sammenhenger kan oppstå dersom undersøkelser baseres på tall fra en delmengde av populasjonen. Vi har i dette kapitlet sett på flere eksempel med slike falske sammenhenger, og forklaringer på hvordan denne sammenhengen oppstår. Vi har også sett på hva som kan gjøres for å få riktige konklusjoner ved å justere dataene slik at utvalget blir representativt.

4 St.Petersburg paradoks

Nicolaus Bernoulli introduserte St.Petersburg paradoks så tidlig som i 1713, men paradokset fortsetter den dag i dag å være en god kilde til nye matematiske gåter [13]. Paradokset er oppkalt etter et av de største vitenskapelige tidskriftene i det attende århundre. I dette tidskriftet ble det publisert en artikkel fra Daniel Bernoulli, søskenbarnet til Nicolaus Bernoulli, hvor han presenterte ulike løsninger på paradokset. Artikkelen med den oversatte tittelen «Utstilling av en ny teori om måling av risiko» ble publisert i 1738 [13].

4.1 St.Petersburg spillet

Den mest sentrale versjonen av St.Petersburg paradoks stammer fra St.Petersburg spillet [13]. Det er verdt å nevne at for å spille dette spillet må man betale en avgift. Spillet fungerer på følgende måte:

Man skal kaste et ekte kronestykke som har to forskjellige sider: mynt og krone. Kronestykket skal kastes inntil siden med krone kommer opp første gang. Hvis krone-siden kommer frem etter første kast vinner man 2 kroner, kommer den på andre kast vinner man 4 kroner osv. Dette kan beskrives med følgende: Gevinsten er 2^k hvor k står for antall kast [13]. Fordelingen for antall kast til første krone er geometrisk med $p = \frac{1}{2}$. Sannsynligheten for at antall kast blir k, blir da: $\frac{1}{2^k}$. Eksempelvis er sannsynligheten for å vinne 32 kr, det vil si $k = 5$:

$$\frac{1}{2^k} = \frac{1}{2^5} = 0.03 = 3\%$$

Vi får at sannsynlighetene for de ulike utfallene er: $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, ...$
På grunnlag av dette kan den forventede gevinsten beskrives slik:

$$E(X) = \frac{1}{2} \times 2 + \frac{1}{4} \times 4 + \frac{1}{8} \times 8 + ... = 1 + 1 + 1 + ... = \sum_{k=1}^{\infty} (\frac{1}{2})^k \times 2^k = \infty$$

Vi ser at den forventede gevinsten, forventningsverdien, er uendelig. Hvis man vet at man kan vinne uendelig med penger på å spille dette spillet, hvor mye skal man da være villig til å betale i avgift for å spille? Det vil være logisk å tenke at enhver person vil være villig til å betale en god del penger, kanskje alle pengene de har, i avgift for å spille ettersom man kan vinne uendelig mye penger. Matematisk gir det mening å betale hvilken som helst sum som er mindre enn uendelig for å spille. Men, mest sannsynlig vil ingen risikere alle pengene de har for å spille. Forventningsverdien er uendelig stor, men sjansene for å vinne stort er likevel minimale. Sannsynligheten for å ikke vinne mer en 2 kr er $\frac{1}{2} = 0.50 = 50\%$ og sannsynligheten for å ikke vinne mer enn 4 kr er $\frac{3}{4} = 0.75 = 75\%$ [13]. Dette kan virke forvirrende fordi det å beregne forventningsverdien er en matematisk måte å finne utav om det lønner seg å spille eller ikke. I dette tilfellet et forventningsverdien uendelig, og dersom man stoler blindt på matematikken vil man konkludere med at det lønner seg å spille spillet.

4.2 Løsninger på St.Petersburg paradokset

Paradokset er i dette tilfellet at få er villige til å betale mye penger for å spille St.Petersburg spillet. Paradokset kan bli forklart ved å ta hensyn til uendelighet. For det første har ikke en bank uendelig med penger å dele ut i premie, og for det andre har ingen uendelig med penger til å spille uendelig mange ganger. Det har blitt utarbeidet flere forslag til hvordan paradokset kan løses. I denne oppgaven skal vi se på to av dem. Den ene løsningen er å begrense midlene banken har, og den andre løsningen er å begrense midlene spilleren har.

4.2.1 Bergrense banken

En forutsetning i St.Petersburg spillet er at banken har uendelige midler, noe den i realiteten ikke har. Dersom banken har begrensede midler, må spillet avsluttes når midlene er brukt opp. Forventningsverdien får derfor en endelig sum. La oss anta at totalen av bankens midler er W, og L er det maksimale antallet banken kan spille før den ikke kan dekke neste innsats. Vi får da gulvfunksjonen $L = \lfloor log_2(W) \rfloor$ [14]. $\lfloor\ \rfloor$ brukes for å vise at man runder ned til nærmeste heltall. Forventningsverdien blir følgende [14]:

$$E = \sum_{k=1}^{L} \frac{1}{2^k} \times 2^k + \sum_{k=L+1}^{\infty} \frac{1}{2^k} \times W = L + \frac{W}{2^L}$$

Hvis bankens totale midler er 1 million, vil: $L = \lfloor log_2(10^6) \rfloor \approx 19kr$ og forventingsverdien blir $E = 19 + \frac{10^6}{2^{19}} = 20.9$. Det vil si, for å i det lange løp ikke tape penger vil man i denne situasjonen ikke være villig til å betale mer enn 20 kr per spill.

4.2.2 Begrense spilleren

Løsningen om å begrense spillerens midler ble introdusert av Daniel Bernoulli. Bernoulli foreslo en logaritmefunksjon $U(w) = ln(w)$ hvor w står for spillerens totale midler [14]. Ut i fra dette får vi en nyttefunksjon som returnerer en positiv eller negativ verdi som indikerer om innsatsen gir grunnlag for om det lønner seg å spille eller ikke spille spillet. I funksjonen under står w som nevnt for spillerens totale midler, og c for størrelsen på avgiften som må betales for å spille. For hver hendelse vil endringen mellom nytten $ln(w+2^{k-1}-c)$ (spillerens midler etter hendelsen) minus $ln(w)$ (spillerens midler før hendelsen) vektes av sannsynligheten for at hendelsen inntreffer [14].

$$E = \sum_{k=1}^{\infty} \frac{(ln(w + 2^{k-1} - c) - ln(w))}{2^k} < \infty$$

Funksjonen beskriver forholdet mellom spillerens midler og hvor mye den skal være villig til å betale i avgift for å spille. Det er mulig å velge en c som gjør at forventningsverdien blir positiv og konvergerer mot en endelig sum [14].

4.3 Oppsummering St.Petersburg paradoks

St.Petersburg paradokset omhandler et spill hvor man tilsynelatende kan vinne uendelige pengepremier. I realiteten er ikke dette mulig ettersom hverken enkeltpersoner eller banker har uendelig med midler, og paradokset vil derfor ikke bli et problem hvis vi ser bort i fra matematikkens verden.

5 Monty Hall problemet

Monty Hall problemet er et matematisk problem basert på et TV-program med navn "*Let's make a deal*", som ble ledet av programlederen Monty Hall [15]. Problemet går ut på at tre dører presenteres for en deltaker. Bak én av dørene befinner det seg en gevinst i form av en bil, mens bak de to andre dørene står det en geit. Deltakeren skal velge én av dørene, men får ikke se hva som er bak. Etterpå åpner programlederen en av dørene. Deltakeren får nå valget om å bytte eller beholde den døren som den har valgt. Spørsmålet blir da om det lønner seg å bytte dør.

Problemet regnes som et paradoks fordi man under vanlige antakelser tilsynelatende skulle tro at sannsynligheten skal være lik, men det viser seg at man dobler sjansen for gevinst ved å bytte dør.

5.1 Løsning på problemet

Monty Hall problemet er svært omdiskutert og mange har ytret sine meninger om problemet. I denne oppgaven skal det fokuseres på løsningen som ble presentert av Marilyn von Savant som tar utgangspunkt i at programlederen vet hvor gevinsten befinner seg og bevisst åpner en dør med en geit bak [15].

I dette problemet kan vi møte på tre forskjellige scenario i første omgang:

1. At deltakeren velger bilen på første forsøk.
2. At deltakeren velger geit 1 på første forsøk.
3. At deltakeren velger geit 2 på første forsøk.

For alle disse scenarioene vil sannsynligheten være $\frac{1}{3}$ ettersom deltakeren skal velge én av tre dører. Hva deltakeren bør gjøre når den får tilbudet om å bytte dør avhenger av hva programlederen gjør. Vi tar utgangspunkt i at programlederen vet hvor premien befinner seg og åpner bevisst en dør med en geit bak. Deretter tilbyr han deltakeren å bytte dør.

Dør 1 (døren deltakeren velger i første omgang)	Dør 2	Dør 3	Går for førstevalget	Bytter dør
Bil	Geit	Geit	Vinner bil	Får en geit
Geit	Bil	Geit	Får en geit	Vinner bil
Geit	Geit	Bil	Får en geit	Vinner bil

Figur 19: Ulike utfall.

I figur 19 er de ulike utfallene fremstilt i en tabell. Ut i fra tabellen ser man at å bytte dør gir gevinst i form av en bil to av tre ganger. Dette betyr at sannsynligheten for å få gevinst dersom man beholder førstevalget er $\frac{1}{3} = 33\%$, mens sannsynligheten for gevinst dersom man bytter dør er $\frac{2}{3} = 66,67\%$.

Flere vil nok trekke konklusjoner om at det ikke spiller noen rolle å bytte dør etter at programlederen har åpnet en dør med geit bak. Dette fordi det nå gjenstår to dører, én med bil og én med geit. Det er dermed naturlig å tenke at det er $\frac{1}{2} = 50\%$ sjanse for å vinne bilen. Denne konklusjonen hadde vært riktig dersom programlederen hadde åpnet en tilfeldig dør, men det gjør han ikke. Valget til programlederen avhenger av deltakerens første valg, og programlederen gir dermed deltakeren mer informasjon.

For å forklare dette nærmere skal vi bruke betinget sannsynlighet. Betinget sannsynlighet handler om at en hendelse A blir påvirket av om en annen hendelse B inntreffer. Vi lar $P(\text{bil dør 1})$ være sannsynligheten for at deltakeren velger bil på første forsøk. *Dør 1* brukes altså for å betegne første valgte dør, *dør 2* brukes for å betegne tilbudt dør [15]. Vi får følgende:

$$P(\text{bil dør 1}) = \frac{1}{3} = 0.33 = 33\%$$

da blir:

$$P(\text{geit dør 1}) = 1 - \frac{1}{3} = \frac{2}{3} = 0.67 = 67\%$$

Hvis det i realiteten er en bil bak første valgte dør vil:

$$P(\text{geit dør 2}|\text{bil dør 1}) = 1$$
$$P(\text{bil dør 2}|\text{bil dør 1}) = 0$$

Hvis det er en geit bak første valgte dør vil sannsynlighetene bli følgende ettersom programlederen allerede har åpnet en dør med geit bak:

$$P(\text{geit dør 2}|\text{geit dør 1}) = 0$$
$$P(\text{bil dør 2}|\text{geit dør 1}) = 1$$

Videre gir loven om total sannsynlighet, også brukt i kapittel 2.5, følgende:

$$\begin{aligned} P(\text{bil dør 2}) &= P(\text{bil dør 2}|\text{bil dør 1}) \times P(\text{bil dør 1}) + P(\text{bil dør 2}|\text{geit dør 1}) \times P(\text{geit dør 1}) \\ &= 0 \times 0.33 + 1 \times 0.67 = 0.67 \end{aligned}$$

Ut i fra den totale sannsynligheten $P(\text{bil dør 2})$ er det 67 prosent sjanse for at bilen er bak den tilbudte døren, og som nevnt ovenfor er det 33 prosent sjanse for at bilen er bak første valgte dør. Konklusjonen blir at det lønner seg å bytte dør i det lange løp [15].

Vi skal se på en annen variant som kanskje tydeliggjør denne løsningen bedre. La oss tenke at deltakeren får valget mellom 100 dører. Da er det $\frac{1}{100}$ sjanse for at deltakeren velger døren med gevinst på første forsøk. Det er dermed $\frac{99}{100}$ dører igjen som kan inneholde gevinsten. Programlederen åpner alle de gjenværende dørene (p = N-2), som ikke inneholder gevinst, utenom én dør. Sannsynligheten blir da $\frac{98}{100}$ for at gevinsten befinner seg bak døren programlederen ikke åpnet, og det er tydelig at det lønner seg å bytte.

5.1.1 Andre løsninger på problemet

Dersom programlederen ikke handler slik som i delkapittel 5.1, at han vet hvor gevinsten befinner seg og åpner en dør med geit bak og tilbyr bytting etterpå, finnes det andre mulige løsninger på problemet. Dersom programlederen kun tilbyr bytting hvis deltakeren har valgt døren med gevinst, vil bytting føre til tap. Programlederen kan også tilby bytting kun dersom deltakeren har valgt feil dør, og da vil bytting føre til gevinst. Hvis programlederen ikke har noe informasjon, og åpner tilfeldig en dør med geit bak, vil sannsynligheten for gevinst ved bytting være $\frac{1}{2}$.

5.2 Oppsummering Monty Hall problemet

Oppsummert kan vi si at i Monty Hall problemet vil det for deltakeren lønne seg å bytte dør i det lange løp hvis man vet at programlederen vet hvor gevinsten befinner seg og åpner en dør med geit bak og tilbyr bytting. Problemet er som nevnt et omdiskutert tema, og det finnes mange ulike teorier og løsninger på problemet. I kapittel 5 har vi sett på en overbevisende teori hvor det ble tatt i bruk betinget sannsynlighet.

6 Paradokser i undervisningssammenheng og i skoleverket

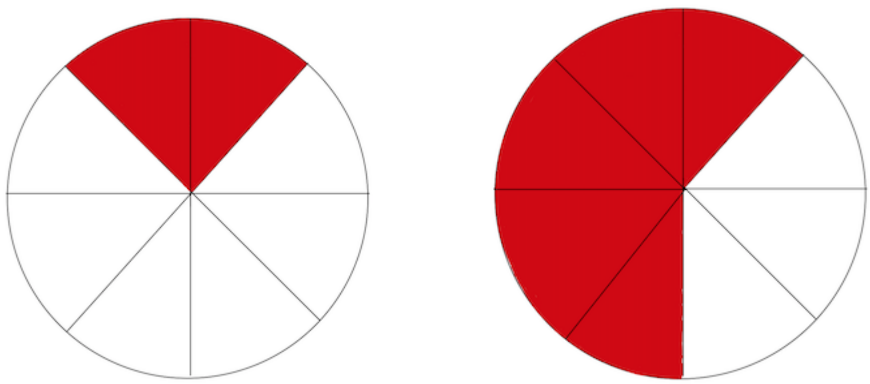
Paradoksene beskrevet i denne oppgaven kan introduseres i skolens matematikkundervisning. Vi har vært innom flere matematiske tema som er relevante både i grunnskolen og på videregående skole. Jeg skal i dette kapitlet foreslå hvordan man kan bruke paradokser i passende undervisningssituasjoner. Jeg har valgt å fokusere på matematikk i den videregående skolen, ettersom jeg mener det er mest passende å introdusere de ulike paradoksene her på grunn av vanskelighetsgrad og hvilke kunnskaper elevene trenger for å forstå dem. Jeg har likevel valgt å ta med et lett forståelig eksempel som kan passe både på barne- og ungdomstrinnet i delkapittel 6.1.1.

6.1 Simpson's paradoks i brøkundervisning

Brøk er et tema som dukker opp allerede på barnetrinnet og følger elevene videre i skoleløpet på ungdomstrinnet og i videregående skole. Simpson's paradoks kan forklares ved bruk av brøker, og derfor kan paradokset passe inn i dette teamet på skolen. Skal Simpson's paradoks introduseres i brøkundervisning er det også hensiktsmessig at elevene forstår bruk av ulikheter. Simpson's paradoks kan for eksempel trekkes inn når elever arbeider med addisjon av brøker. Det er verdt å nevne at selve paradokset ikke er relevant i brøkregning, men siden brøk er et element i paradokset kan det nevnes i temaet brøk. En vanlig misoppfatning knyttet til addisjon av brøker er at elevene ser på teller og nevner som uavhengige hele tall [16]. Disse elevene adderer to tellere og nevnere uavhengig av hverandre på denne måten: $\frac{1}{3}+\frac{4}{5}=\frac{5}{8}$ fordi de tenker at $1+4=5$ og $3+5=8$. Feilen her ligger i at elevene behandler brøkene som to ulike helheter og ikke to brøker av samme helhet [16]. Å addere sammen brøker på denne måten gjør vi når vi skal gjøre en samlet analyse av data fra to grupper, slik som vi gjør i Simpson's paradoks. Det er nettopp på grunn av at vi adderer sammen brøkene på denne uriktige måten at paradokset kan oppstå. Jeg har laget et eksempel som kan vises i undervisningen for å hjelpe elevene å forstå hvorfor man ikke kan addere brøker på denne måten til vanlig.

6.1.1 Lett forståelig brøk-eksempel

Per og Lise spiser pizza. Lise spiser $\frac{2}{8}$ *av pizzaen, mens Per spiser* $\frac{5}{8}$ *av pizzaen. Hvor mye av pizzaen spiser de til sammen?*



Figur 20: Pizzastykker som blir spist.

Pizzastykkene tilhører samme helhelt - én pizza. Hvis vi adderer på følgende måte: $\frac{2}{8} + \frac{5}{8} = \frac{2+5}{8+8}$ får vi at Lise og Per har spist $\frac{7}{16}$ stykker. Dette stemmer ikke fordi én pizza har totalt 8 stykker, ikke 16. $\frac{7}{16}$ er mindre enn en halv pizza, men vi ser ut i fra figuren og fra brøken at Per alene har spist mer enn en halv pizza. Riktig måte å addere brøker på er å finne fellesnevner og deretter legge dem sammen. I dette tilfellet har brøkene felles nevner og vi kan legge sammen tellerne : $\frac{2}{8} + \frac{5}{8} = \frac{2+5}{8} = \frac{7}{8}$. Lise og Per spiser til sammen $\frac{7}{8}$ av hele pizzaen. Hadde vi derimot skulle sett på en samlet analyse av data fra to grupper måtte vi ha addert på den andre måten. F.eks hvis Lise og Per spiste hver sin pizza og vi skulle sett på totalt antall spiste pizzastykker. Da hadde det blitt riktig å sagt at de spise 7 av 16 pizzastykker ettersom de to pizzaene er to ulike helheter.

6.1.2 Nyrestein-eksempelet i undervisningen

Nyrestein-eksempelet vist i delkapittel 2.2 og 2.3.2 kan introduseres for elever som har kjennskap til brøk og ulikheter. I læreplan for matematikk 2P står følgende to kompetansemål [17]:

"Mål for opplæringen er at eleven skal kunne utforske strategier for å løse ligninger, ligningsystemer og ulikheter og argumentere for tenkemåtene sine"

og

"Mål for opplæringen er at eleven skal kunne analysere og presentere funn i datasett fra lokalsamfunn og media."

Disse kompetansemålene forutsetter at elevene har kunnskap om ulikheter, men også analysering av datasett. Fra matematikk 1P, faget man har før matematikk 2P, har elevene også kjennskap til brøk. Derfor vil nyrestein-eksempelet være et fint eksempel å vise for elevene. Eksempelet kan bidra til økt forståelse for hvordan man skal tolke datasett, samt gi en bedre forståelse av brøker og hvordan man kan bruke ulikheter. Nyrestein-eksempelet viser elevene at man ikke må trekke konklusjoner før man har studert dataene nøye. I tillegg får elevene en oppfriskning i brøkregning. I nyrestein-eksempelet hadde vi følgende brøker:

$$\frac{234}{270} < \frac{81}{87}$$

$$\frac{55}{80} < \frac{192}{263}$$

$$\frac{234+55}{270+80} = \frac{289}{350} > \frac{81+192}{87+263} = \frac{273}{350}$$

Ut fra disse brøkene vil nok elevene identifisere at dette ikke er riktig måte å addere sammen brøker på. Som lærer kan man da forklare dette med brøk og helheter og at i Simpson's paradoks ser vi på en samlet analyse fra to grupper og derfor adderer på denne måten. Jeg mener at å introdusere et slik eksempel vil være med på å styrke elevenes kompetanse og fremme dybdelæring.

6.2 Bruk av Simpson's og Berkson's paradoks i undervisningen

Simpson's paradoks kan også trekkes inn i matematikk-fag hvor læreplanen inneholder kompetansemål som gjelder sannsynlighetsregning og statistikk. Det samme gjelder for Berkson's paradoks. På videregående skole møter vi slike kompetansemål i fagene matematikk S1 og S2. Paradoksene kan være høyst reelle for å gi elevene kunnskap om ulike misoppfatninger og gale konklusjoner vi ser i ulike statistiske undersøkelser. Kompetansemålene i læreplanen oppfordrer elevene til utforskning og derfor mener jeg at paradoksene kan trekkes inn i undervisningen. Et av kompetansemålene i læreplanen til matematikk S1 lyder som følger [18]:

"*Mål for opplæringen er at eleven skal kunne planlegge og gjennomføre et selvstendig arbeid med reelle datasett knyttet til samfunnsøkonomiske temaer og forhold, og analysere og presentere funn.*"

Dette kompetansemålet innebærer at elevene har kunnskaper om hva som er korrekte analyser og riktige konklusjoner. Simpson's paradoks og Berkson's paradoks er fenomen man bør være obs på når man arbeider med datasett slik at man kan trekke riktige konklusjoner.

Jeg har tidligere i oppgaven utformet eksempel om Simpson's paradoks og om Berkson's paradoks som med fordel kan brukes i undervisningen. Målet ved å ta i bruk disse eksempelene er at elevene skal få innsikt i ulike ting som kan gå galt når man arbeider med data og hvordan man kan avdekke eventuelle gale konklusjoner.

Jeg ville i undervisningen vist til eksempelet med tall fra Simpson's originale artikkel, skrevet om i delkapittel 2.1, for å vise hvor galt det kan gå hvis man ikke ser på de riktige dataene. I eksempelet ser vi på sammenhengen mellom behandling og overlevelse, hvor det ved første øyekast kan se ut som at behandling er ineffektiv for den totale populasjonen og at det er 50/50 sjanse for overlevelse med og uten behandling. Ser vi på undergruppene kvinner og menn finner vi en positiv assosiasjon mellom behandling og overlevese i disse undergruppene. Dette eksempelet viser elevene at dersom man ikke gjør rett analyse kan det gå fryktelig galt.

For å vise elevene viktigheten med representative data og representative utvalg ville jeg introdusert elevene for Berkson's paradoks. Eksempelet om ATV-ulykker vist i delkapittel 3.2 mener jeg kan være et fint eksempel å trekke frem i undervisningen når elevene skal arbeide med datasett og utvalg. Eksempelet omhandler, som nevnt tidligere, en undersøkelse som viste at bruk av hjelm var assosiert med større skadealvorlighet i ATV-ulykker. Woodfine og Redelmeier avdekket at Berkson's paradoks var en sannsynlig forklaring på funnene. Dataene i undersøkelsen var ikke representative og derfor oppstod det en falsk sammenheng mellom hjelm og sykehusinnleggelse.

Når vi er inne på bruk av Berkson's paradoks i undervisningen vil jeg også trekke frem eksempelet fra delkapittel 3.3. Her ser vi på hvordan Berkson's paradoks kan forekomme i mer hverdagslige situasjoner enn statistiske undersøkelser. Eksempelet om film og bøker tror jeg kan være en «øyeåpner» for elevene, og kanskje føre til at de blir litt mer årvåkne på ulike sammenhenger de hører om i hverdagen.

6.3 Bruk av St.Petersburg paradoks i undervisningen

St.Petersburg paradokset kan introduseres til elevene som en «gruble» oppgave. Jeg mener paradokset er mest passende i matematikk S2 ettersom at elevene da har opparbeidet seg grunnleggende kunnskap om sannsynlighetsregning og statistikk, samt at temaene inngår i kompetansemålene.

Paradokset kan med fordel prøves i praksis, i klasserommet, men uten satsing av penger. På denne måten kan elevene få utforske hvor mange kast man bruker før den første kronen inntreffer. I læreplanen til matematikk S2 lyder to av kompetansemålene som følger [18]:

"Mål for opplæringen er at eleven skal kunne forstå begrepene forventningsverdi, varians og standardavvik, og bruke disse størrelsene til å tolke stokastiske variabler."

og

"Mål for opplæringen er at eleven skal kunne utforske egenskaper ved ulike rekker og gjøre rede for praktiske anvendelser av egenskaper ved rekker."

I St.Petersburg paradokset regner man ut forventningsverdi for å beregne forventet gevinst. Som lærer kan man utfordre elevene til å regne ut sannsynligheten for å vinne ulike pengepremier og forventningsverdi. På denne måten vil paradokset være relevant til de to nevnte kompetansemålene og paradokset kan brukes som lærestoff.

6.4 Bruk av Monty Hall problemet i undervisingen

Jeg tror bruk av Monty Hall problemet i undervisningen kan være med på å skape en faglig diskusjon i klasserommet. På lik linje som St.Petersburg paradokset kan det prøves ut i praksis. Man kan lage et game-show i klasserommet hvor læreren tar rollen som programleder og elevene er deltakere. Etter å ha spilt et par runder kan man starte diskusjonen om det lønner seg å bytte dør eller beholde først valgte dør.

6.5 Oppsummering paradokser i undervisningssammenheng

I kapittel 6 har vi sett på hvordan paradoksene forklart i oppgaven kan trekkes inn i undervisningen. Jeg mener at disse paradoksene kan bidra til å styrke elevenes kompetanse og fremme dybdelæring. I tillegg kan de være med på skape faglige diskusjoner i klasserommet, og kanskje vekke en nysgjerrighet hos enkelte elever. Det å bruke paradoksene i undervisningen kan også øke elevenes kunnskap om statistikk, og gjerne føre til at de blir mer bevisste på at man ikke kan stole på alt man leser og hører.

6.6 Simpson's paradoks i forbindelse med tolkning av resultater fra nasjonale prøver

Nasjonale prøver er obligatoriske prøver som blir gjennomført på 5., 8. og 9. trinn. Formålet med prøvene er å kartlegge elevenes grunnleggende ferdigheter i regning, engelsk og lesing. Resultatene skal danne grunnlag for underveisvurdering og kvalitetsutvikling på alle nivå i skolesystemet [19]. På 8.trinn blir resultatene målt i mestringsnivå, hvor nivå 1 er lavest og nivå 5 er høyest [19].

Når man tolker resultateter fra nasjonale prøver må man også være obs på Simpson's paradoks. Jeg har laget et fiktivt eksempel som skal illustrere et tenkt problem som kan oppstå ved tolkningen av resultater på nasjonale prøver.

En ungdomsskole er interessert i å se om det har blitt flere eller færre elever på mestringsnivå 5 fra 2021 til 2022 på 8.trinn. De undersøker resultatene fra nasjonale prøver og ser at disse fordeler seg slik:

Mestringsnivå 5	
2021	2022
21/90 = 23,3%	22/105 = 21%

Figur 21: Elever på mestringsnivå 5 i 2021 og 2022.

Resultatene fra tabellen i figur 21 viser at det har vært en prosentvis nedgang i antall elever på mestringsnivå 5 og skolen konkluderer med at 8.trinn gjorde det bedre i 2021 enn i 2022. Skolen ønsker også å se på hvordan resultatene fordeler seg mellom jenter og gutter og får følgende:

	Mestringsnivå 5	
	2021	2022
Jenter	12/40 = 30%	20/45 = 44,4%
Gutter	9/50 = 18%	12/60 = 20%
Total	21/90 = 23,3%	22/105 = 21%

Figur 22: Jenter og gutter på mestringsnivå 5 i 2021 og 2022.

I figur 22 er dataene delt inn i undergruppene jenter og gutter. Her ser skolen at det er en økning i både jenter og gutter på mestringsnivå 5 fra 2021 til 2022, og konklusjonen de trakk i første omgang er feil. Det har altså vært en prosentvis økning i elever fra 2021 til 2022.

6.7 Oppsummering Simpson's paradoks i skoleverket

Eksempelet i delkapittel 6.6 viser som sagt et tenkt problem som kan oppstå ved tolkningen av resultater på nasjonale prøver. Det finnes andre mulige og lignende problem som kan oppstå. For eksempel dersom to skoler, to trinn eller to kommuner sammenligner sine resultater. Når det blir gjennomført slike sammenligninger er det viktig å være klar over at det kan finnes forskjeller i undergrupper, og man bør ta høyde for disse før man konkluderer.

7 Avsluttende ord

I denne oppgaven har det blitt presentert ulike eksempeler og løsninger på paradokser i sannsynlighetsregning og statistikk. Oppgaven har vist viktigheten av å ha tilstrekkelig kunnskap om ulike fenomen for å unngå å trekke gale konklusjoner. For å øke bevisstheten om disse paradoksene blant elever på videregående skole har jeg i siste del av oppgaven presentert ulike måter å implementere disse på i undervisningen. Dette for å fremme læring hos elevene, men også for å gjøre dem bevisste på ulike fallgruver som finnes i arbeid med tall og data. Til slutt har jeg i denne oppgaven illustrert hvordan Simpson's paradoks kan opptre i undersøkelser i skoleverket.

Referanser

[1] E. H. Simpson, «The Interpretation of Interaction in Contingency Tables,» *Journal of the Royal Statistical Society. Series B*, årg. 13, s. 238–241, 1951.

[2] J. Røislien og J. T. Kvaløy, «Simpsons paradoks - når pluss og pluss blir minus,» *Tidsskriftet, Den Norske Legeforening*, årg. 6, 2021. adresse: https://tidsskriftet.no/2021/03/medisin-og-tall/simpsons-paradoks-nar-pluss-og-pluss-blir-minus.

[3] C. Chirag, D. Webb, S. Payne og J. Wickham, «Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy, and extracorporeal shockwave lithotripsy,» *British medical journal*, årg. 292, s. 879–882, 1986. adresse: https://www.bmj.com/content/292/6524/879.

[4] J. Kocik, «Proof without Words: Simpson's Paradox,» *Mathematics Magazine*, årg. 74, s. 399, 2001. adresse: http://www.jstor.org/stable/2691038.

[5] J. T. Kvaløy, *Pitfalls and paradoxes in data analysis*, upublisert seminarnotat.

[6] G. G.Løvås, *Statistikk for universiteter og høgskoler.* Universitetsforlaget, 2013.

[7] A. J. Wilcox, «On the importance - and the unimportance - of birthweight,» *International Journal og Epidemiology*, årg. 30, s. 1233–1241, 2001. adresse: https://doi.org/10.1093/ije/30.6.1233.

[8] J. Røislien og J. T. Kvaløy, «Trekantdrama,» *Tidsskriftet, Den Norske Legeforening*, årg. 7, 2021. adresse: https://tidsskriftet.no/2021/04/medisin-og-tall/trekantdrama.

[9] J. Woodfine og D. Redelmeier, «Berkson's paradox in medical care,» *Journal of Internal Medicine*, årg. 278, s. 424–426, 2015. adresse: https://doi.org/10.1111/joim.12363.

[10] J. T. Kvaløy, *Bruk statistikk riktig!* Upublisert notat, 2008. adresse: https://www.ux.uis.no/~jtk/Rettbruk.pdf.

[11] Numberphile, *Does Hollywood ruin books*, Youtube, 2019. adresse: https://www.youtube.com/watch?v=FUD8h9JpEVQ.

[12] N. Fenton, *Collider Bias (Berkson's paradox): how 'censored' data leads to flawed conclusions*, Youtube, 2021. adresse: https://www.youtube.com/watch?v=eJNPUfO-Raw&t=77s.

[13] M. Peterson, «The St. Petersburg Paradox,» i *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, red., Summer 2022, Metaphysics Research Lab, Stanford University, 2022.

[14] Wikipedia, *St. Petersburg paradox*, Sist besøkt 13.06.23. adresse: `https://en.wikipedia.org/wiki/St._Petersburg_paradox#cite_note-19`.

[15] S. Sæbø, *Fra spøkefull Monty Hall til alvorsfull kreftdiagnostikk*, Universitet for miljø- og biovitenskap, 2009. adresse: `http://www.umb.no/statisk/skole/matematiskstatistikksevu09.pdf`.

[16] Matematikksenteret, *Andre problemer knyttet til brøk*, Sist besøkt 11.06.23. adresse: `https://www.matematikksenteret.no/kartlegging-i-matematikk/misoppfatninger-i-matematikk/andre-problemer-knyttet-til-br%C3%B8k`.

[17] Utdanningdirektoratet, *Læreplan i matematikk fellesfag 2P*, Læreplan gyldig fra 01.08.2021, 2021. adresse: `https://data.udir.no/kl06/v201906/laereplaner-lk20/MAT05-04.pdf?lang=nob`.

[18] Utdanningdirektoratet, *Læreplan i matematikk for samfunnsfag (matematikk S)*, Læreplan gyldig fra 01.08.21, 2021. adresse: `https://data.udir.no/kl06/v201906/laereplaner-lk20/MAT04-02.pdf?lang=nob`.

[19] Utdanningsdirektoratet, *Rammeverk for nasjonale prøver*, Sist endret 19.08.22, 2022. adresse: `https://www.udir.no/eksamen-og-prover/prover/rammeverk-for-nasjonale-prover2/gjennomforing-og-resultater/`.