

Simpsons Paradoks

Et emnearbejde om årsag og sammenhæng i
kvantitative undersøgelser.

Inge Henningsen
Afdeling for Anvendt Matematik og Statistik
Københavns Universitet

Sammenhæng er ikke det samme som årsag - et eksempel

”Fra en beholder med lopper tager den berømte videnskabsmand Van Dumholtz forsigtigt en enkelt loppe, river omhyggeligt bagbenene af den og kommanderer med høj røst ”spring”. Han noterer at det gør den ikke og prøver det samme med en ny loppe. Efter at have gentaget forsøget et passende antal gange, analyserer han omhyggeligt resultaterne og konkluderer, at lopper hører med bagbenene.”

(Paulos J.A. (1988) *Innumeracy* Penguin Books, London.)

Simpsons paradoks.

Et væsentligt problem i kvantitativ analyse er den type fejlslutninger, hvor man forveksler sammenhæng med kausalitet og f.eks. tror at finde en betydningsfuld (årsags)sammenhæng mellem to variable, hvor sammenhængen i virkeligheden skyldes at de begge påvirkes af den samme baggrundsvariabel. Simpsons paradoks går et skridt videre og viser, hvorledes inhomogene forsøgspopulationer kan føre til, at man vender sammenhængen i en tabel om. Som illustration til Simpsons paradoks indeholder materialet en række opdigtede avisnotitser. De strækker sig fra det let gennemskuelige (*skonummer* og *uddannelsesvalg*) til det knapt så indlysende (*alder* og *arbejdsløshed*), men fælles er, at de på grundlag af det man kunne kalde en *statistisk* sammenhæng postulerer en falsk *årsagssammenhæng*.

Simpsons Paradoks

-Et emnearbejde om årsag og sammenhæng i kvantitative undersøgelser.

Formål

At introducere/formalisere/diskutere begreber som kausalitet og konfundering (falsk sammenhæng). Uddybe forståelse af begreberne stokastisk uafhængighed og betinget uafhængighed. At klargøre forholdet mellem statistisk og kausal afhængighed.

Kompetence

De studerende skal være i stand til at vurdere situationer, hvor der sluttes om kausalitet på grundlag af observationer med usikkerhed. De skal have redskaber til på formel måde at afgøre, hvornår man kan se bort fra baggrundsvariable. De skal kunne kritisere og forklare misbrug/forkert brug af sammenhæng i tabeller. De skal kunne gøre rede for analyseresultater i et passende, utvetydigt sprog.

Forudsætninger: Sandsynlighed på endeligt udfaldsrum. Uafhængighed. Betinget fordeling. Stokastisk variabel. Uafhængighed og betinget uafhængighed af stokastiske variable. Binomialfordeling. Hypergeometrisk fordeling. Stikprøveudtagning med og uden tilbagelægning. Binomialfordeling (polynomialfordeling) som approksimation til hypergeometrisk fordeling. Test i statistiske modeller på endelige udfaldsrum. Frekvensfortolkning af sandsynlighed. Store tals lov i endelig version.

Videre overvejelser. Udgangspunktet for arbejdet er en række opdigtede avisnotitser. De er selvfølgelig opdigtede, men **realistiske** om end på en overdreven måde. Den stokastiske variation er taget væk for at gøre pointerne tydeligere. Ved at beskæftige sig med situationer, som de har erfaring med og derfor intuition omkring ledes eleverne frem til at operationalisere begreber som samvariation, fravær af samvariation, samt se disses sammenhæng med de teoretiske begreber uafhængighed, konfundering og betinget uafhængighed.

Materiale

1. Simpsons paradoks: Indledende problemstilling
2. Opdigtede avisnotitser
3. Diskussion af de opdigtede avisnotitser
4. Skema: Hvad skal man tage hensyn til.
5. Egne opdigtede avisnotitser
6. Operationalisering af sammenhæng mellem variable
7. Sammenhæng mellem variable (Teoretiske begreber, modellering)
8. Mål for afhængighed
9. Taleksempler: 2x2x2- tabeller
10. Hvornår kan man se bort fra baggrundsvariable?
(Operationalisering)
11. Hvornår kan man se bort fra baggrundsvariabel?
(Teoretiske begreber, modellering)
12. Eksempler fra aviser og andre medier.
13. Grafisk illustration af Simpsons paradoks.
14. Tabeller med usikkerhed
15. Regression
16. Videregående eksempel: "Unge studenter"
17. Statistisk sammenhæng og kausalitet

Simpsons Paradoks

Et emnearbejde om årsag og sammenhæng i kvantitative undersøgelser.

Undervisningsplan

Materiale	Formål med arbejdet
Indledende problemstilling (Bilag 1)	Introduktion af problemstilling
De fiktive avisnotitser (Bilag 2)	Hvad er der galt? Hvilke begreber er der brug for, hvis man vil samtale uden misforståelser?
Den skrevne diskussion af de fiktive avisnotitser (Bilag 3)	Er man enig i pointerne? Forslag til tilføjelser eller ændringer. Kom der nye momenter i diskussionen?
Skema: Hvad skal man tage med? (Bilag 4)	Enighed om, hvad der skal/ikke skal tages hensyn til
Nye fiktive avisnotitser	Vælg eller opfind eksempler
"Sammenhæng mellem variable" i 2x2-tabel (operationalisering) (Bilag 5)	Forskellige former for sammenhæng. Hvilke begreber skal man have for utvetydigt at kunne tale om sammenhæng.
"Sammenhæng mellem variable". (Teoretiske begreber, modellering) (Bilag 6)	Hvordan kan uafhængighed hhv. samme betingede fordeling bruges til at modellere manglende sammenhæng?
Mål for afhængighed. (Bilag 7)	Indførelse af odds-ratio, som et mål for sammenhæng
Taleksempler: 2x2x2-tabeller (Bilag 8)	Brug af de teoretiske begreber.

Undervisningsplan (fortsat)

Materiale	Formål med arbejdet
Hvornår kan man se bort fra baggrundsvariable? (Operationalisering) (Bilag 9)	Hvad vil det sige at man kan se bort fra baggrundsvariabel
Hvornår kan man se bort fra baggrundsvariable? (Teoretiske begreber, modellering) (Bilag 10)	Hvordan kan begrebet betinget uafhængighed bruges til at modellere en situation hvor man kan se bort fra baggrundsvariabel
Find eksempler fra aviser.	Virkelige situationer er mere spegede. Hvornår skal man bruge samlede og hvornår opdelt tabeller?
Grafer der illustrerer konfundering. (Bilag 11)	
Tabeller med usikkerhed (Bilag 12)	Afklaring af forholdet mellem systematisk og tilfældig variation
Regression (Bilag 13)	Inhomogenitet kan også vende sammenhæng i regression
Unge studenter (Bilag 14)	Illustrerer de foregående. Indeholder også eksempler på selektion, der kan pege på ny problemstilling
Statistisk sammenhæng og kausalitet (Bilag 15)	Afklaring af forholdet mellem statistisk sammenhæng og kausalitet

Opdigtede avisnotitser til diskussion af forholdet mellem årsag og sammenhæng. (1)

Humanister går i små sko.

Ved en sammenligning af en gruppe naturvidenskabelige og en gruppe humanistiske studenter opdagede man en overraskende forskel. De naturvidenskabelige studenter var i gennemsnit næsten 3 cm højere end de humanisterne og brugte 1 1/2 nummer større i sko.

Mænd og kvinder vælger stadig forskellige uddannelser.

Ved optaget på de lange videregående uddannelser i 1998 var der knap 40% kvinder på de naturvidenskabelige uddannelser, mens kvindeproucenten var over 60% på de humanistiske uddannelser. Der er således stadig forskel på mænds og kvinders studievalg.

Hoftebrud medfører forhøjet kræftisiko.

Ved en undersøgelse på Rigshospitalet har det vist sig, at patienter indlagt for brud på lårhalsknoglen har en 10 gange så høj risiko for at få kræft inden for en periode på 5 år, som patienter indlagt med henblik på en meniskoperation. Lægerne undersøger nu, hvorfor en forskellig traumelokalisering giver en så markant forskel i risikoen for få kræft. På baggrund af undersøgelsen påpeger bandagist NN, at systematisk brug af firmaets nyudviklede underbukser med stødabsorberende indlæg til beskyttelse af hoftepartiet, vil kunne reducere den alt for høje danske kræftisiko.

Opdigtede avisnotitser til diskussion af forholdet mellem årsag og sammenhæng. (2)

Jørn Hjørtning skyld i ekstra kræfttilfælde.

En undersøgelse foretaget af Amtsrådsforeningen har vist, at personer der lytter til "De ringer, vi spiller" har en 3 gange så høj risiko for at få kræft inden for en periode på 5 år, end dem der jævnlige ser udsendelsen "Beat". Kræftrisikoen ved at høre Jørn Hjørtning er forhøjet med 2000% i forhold til at høre børneradio. Disse fund bør få konsekvenser for programlægningen.

Mascara beskytter mod testikelkræft.

En undersøgelse på Frederikssund Sygehus, der tidligere har haft en overhyppighed af testikelkræfttilfælde, har vist, at brug af mascara reducerer forekomsten af testikelkræft drastisk. I en 5-års periode har man i hele Frederiksborg Amt kun haft 1 tilfælde af testikelkræft blandt mascarabrugere, hvor man i en aldersmæssigt tilsvarende gruppe ville have forventet 37 tilfælde.

P-piller disponerer for rygning. Beskytter mod hoftebrud.

En undersøgelse i Glostrup af alle kvinder over 20 år har vist, at brug af p-piller medfører øget tendens til rygning. Til gengæld synes p-pillerne at give en vis beskyttelse mod hoftebrud.

Opdigtede avisnotitser til diskussion af forholdet mellem årsag og sammenhæng. (3)

Nedlægger intensiv afdeling.

Amtssygehuset i XX har besluttet at nedlægge den intensive afdeling, idet en undersøgelse har vist, at dødeligheden på denne afdeling ligger langt over hospitalets gennemsnit. I fremtiden vil alle alvorligt syge patienter blive indlagt på ortopædkirurgisk afdeling. Hospitalsdirektøren tror, at den nye organisering vil blive en væsentlig sundhedsmæssig gevinst for amtets borgere.

De helt unge har foruroligende stor arbejdsløshed.

Ved en opgørelse foretaget af arbejdsformidlingen i København har det vist sig, at arbejdsløsheden blandt de 16-18-årige er mere end dobbelt så høj som arbejdsløsheden i aldersgruppen fra 30-35.

Hvornår kan man se bort fra baggrundsvariable?

Det er vigtigt at inddrage relevante baggrundsvariable hvis man vil undersøge sammenhængen mellem to variable.
Men

- Hvad er relevante baggrundsvariable?
- Hvornår må man se bort fra en baggrundsvariabel?
- Hvordan inddrager man baggrundsvariable i den statistiske analyse?

Sammenhæng i de opdigtede avisnotitser:

Hvis vi vil undersøge sammenhæng mellem	Skal vi dele op efter	Behøver vi ikke at dele op efter
Skonummer og uddannelsesvalg	Køn	Fregner
Uddannelsesvalg og køn	-	Skonummer
Hoftebrud og kræft	Alder	-
Jørn Hjorting og hoftebrud	Alder	-
Hoftebrud og alder	-	Jørn Hjorting
Maskara og testikelkræft	Køn	Jørn Hjorting
P-piller og rygning	Alder	-
P-piller og hoftebrud	Alder	-
Intensiv afd. og dødelighed	Prognose	Køn
Alder og arbejdsløshed	Uddannelse	Gråt hår

Sammenhæng mellem variable

Årsagssammenhæng:

- Hvad påvirker uddannelsesvalg?
- Mascara beskytter mod testikelkræft
- Hoftebrud medfører forhøjet kræftrisiko.

Samvariation

- De helt unge har foruroligende stor arbejdsløshed.
- Større dødelighed på intensiv afdeling.

Det drejer sig i alle tilfælde om

- Empiriske/statistiske sammenhænge.
- Udsagn baseret på observationer.
- Induktive ræsonnementer.

Ikke tale om sikre sammenhænge, men om at nogle sammenhænge blev observeret oftere end andre. F.eks.

- Kvinder studerer oftere et humanistisk fag end mænd gør, men ikke alle humanister er kvinder.
- Personer indlagt med hoftebrud får oftere kræft end dem der er indlagt for en meniskskade. Men ikke alle med hoftebrud får kræft, og kræft kan også ramme unge mennesker.

At slutte fra observationer til de bagvedliggende årsagssammenhænge foregår gennem en **model**. Sådanne modeller bliver ikke altid formuleret eksplicit, men de er der alligevel.

Sammenhæng mellem 2 variable.

Hvordan **operationaliseres** sammenhæng mellem f.eks. *køn* og *holdning til prisstigninger på offentlig befordring*.

Spørgsmål	Svar
<i>Hvornår vil man sige, at der er sammenhæng mellem køn og holdning til prisen på offentlig befordring?</i>	<p><i>Hvis der er en (væsentlig) forskel på den andel blandt kvinderne og den andel blandt mændene, der er mod prisstigninger.</i></p> <p><i>Hvis kendskab til en persons køn gør det lettere at gætte vedkommendes holdning til prisstigninger.</i></p>
<i>Hvornår vil man sige, at der ikke er sammenhæng mellem køn og holdning til prisen på offentlig befordring?</i>	<p><i>Hvis der er (næsten) samme andel blandt kvinderne som blandt mændene, der er mod prisstigninger.</i></p> <p><i>Hvis kendskab til en persons køn (næsten) ikke fortæller noget om vedkommendes holdning til prisstigninger.</i></p>

I hvilken forstand er dette det samme spørgsmål?

Undersøgelse af sammenhæng mellem 2 variable.(1)

Hvordan **undersøges** sammenhæng mellem f.eks. *køn* og *holdning til prisstigninger på offentlig befordring* empirisk?

Man kan

- *tilfældigt vælge et antal mænd og et antal kvinder, f.eks. 500 af hvert køn, og spørge om deres holdning til prisstigning på offentlig befordring*

Ser man bort fra problemer med at alle ikke kan træffes og at alle ikke kan eller vil svare på, om de er for eller imod, får man en tabel af følgende type, hvor de antal man selv har bestemt er skrevet med fed skrift.

<i>Køn</i>	<i>Prisstigning på offentlig befordring</i>		
	<i>Imod</i>	<i>For</i>	<i>I alt</i>
<i>Kvinde</i>	397	103	500
<i>Mand</i>	308	192	500
<i>I alt</i>	705	295	1000

Statistisk analyse

Dette kan modelleres ved to uafhængige binomialfordelinger og sammenligningen foretages med et χ^2 -test.

Undersøgelse af sammenhæng mellem 2 variable.(2)

Hvordan **undersøges** sammenhæng mellem f.eks. *køn* og *holdning til prisstigninger på offentlig befordring* empirisk?

Man kan også

- *tilfældigt udvælge et antal personer, f.eks. 1000, og spørge dem om køn og holdning til prisstigning på offentlig befordring*

Ser man bort fra problemer med at alle ikke kan træffes og at alle ikke kan eller vil svare på, om de er for eller imod, får man en tabel af følgende type, hvor de antal man selv har bestemt er skrevet med fed skrift.

<i>Køn</i>	<i>Prisstigning på offentlig befordring</i>		
	<i>Imod</i>	<i>For</i>	<i>I alt</i>
<i>Kvinde</i>	419	117	536
<i>Mand</i>	288	176	464
<i>I alt</i>	707	293	1000

Statistisk analyse

Dette kan modelleres ved uafhængighed mellem de to inddelingskriterier og analysen udføres med et χ^2 -test. Testet er i praksis identisk med det foregående test og modellen i 1) kan fås som betinget model i 2).

Falske og/eller misvisende sammenhænge

Simpsons paradoks og de opdigtede avisnotitser viste hvordan man kan forveksle sammenhæng med kausalitet og se betydningsfulde (årsags)sammenhænge mellem to variable, hvor sammenhængen i virkeligheden skyldes at de begge påvirkes af den samme baggrundsvariabel.

For at talmæssigt illustrere hvad der kan ske, når man ser bort fra en baggrundsvariabel, betragtes en (konstrueret) undersøgelse af 5 års-overlevelse ved to medicinske behandlinger. Her antages det, at patienterne enten tilhører en gruppe relativt raske personer med god udsigt til at overleve, betegnet som personer med god prognose, eller en anden, sygere gruppe betegnet som personer med dårlig prognose. Hver behandling gives til et antal patienter, og de to patientgrupper er på forskellig måde sammensat af et antal personer med god henholdsvis dårlig prognose.

Eksemplerne viser, at hvis man i sin analyse ikke tager højde for patientsammensætningen i grupperne, så kan man komme frem til nogle meget misvisende resultater. Det kan f. eks. ske, at en behandling, der fungerer bedst for begge patientgrupper, i en samlet analyse kommer ud som dårligere. Samtidig vises, at det under visse betingelser ikke er nødvendigt at dele op efter prognose. I tabellerne er der for nemheds skyld set bort fra den statistiske variation.

Data kan opskrives i 2x2x2-tabeller.

Taleksempler i 2x2x2-tabeller (1)

Konstrueret tabel for behandling, prognose og overlevelse.

Behandling 1

<i>Prognose</i>	<i>Overlevende</i>		<i>Antal i alt</i>
	<i>Antal</i>	<i>Procent</i>	
<i>God</i>	80	80%	100
<i>Dårlig</i>	4	40%	10
<i>Alle</i>	84	76%	110

Behandling 2

<i>Prognose</i>	<i>Overlevende</i>		<i>Antal i alt</i>
	<i>Antal</i>	<i>Procent</i>	
<i>God</i>	8	80%	10
<i>Dårlig</i>	40	40%	100
<i>Alle</i>	48	44%	110

Her har behandling 1 den bedste overlevelse for den samlede patientgruppe. Når man som her har delt op efter prognose, har de to behandlinger imidlertid samme overlevelseshyppighed i begge prognosegrupper.

Sammenligning af behandlinger

<i>Behand- ling</i>	<i>Overlevende</i>		<i>Antal i alt</i>
	<i>Antal</i>	<i>Procent</i>	
1	84	76%	110
2	48	44%	110

Taleksemppler i 2x2x2-tabeller (2)

Konstrueret tabel for behandling, prognose og overlevelse.

Behandling 1

<i>Prognose</i>	<i>Overlevende</i>		<i>Antal i alt</i>
	<i>Antal</i>	<i>Procent</i>	
<i>God</i>	60	60%	100
<i>Dårlig</i>	4	40%	10
<i>Alle</i>	64	58%	110

Behandling 2

<i>Prognose</i>	<i>Overlevende</i>		<i>Antal i alt</i>
	<i>Antal</i>	<i>Procent</i>	
<i>God</i>	8	80%	10
<i>Dårlig</i>	50	50%	100
<i>Alle</i>	58	53%	110

Her har behandling 2 den bedste overlevelse i begge prognosegrupper, men behandling 1 har den bedste overlevelse for alle patienter samlet, fordi der er relativt flere "raske" patienter der får behandling 1.

Sammenligning af behandlinger

<i>Behand- ling</i>	<i>Overlevende</i>		<i>Antal i alt</i>
	<i>Antal</i>	<i>Procent</i>	
1	64	58%	110
2	58	53%	110

Taleksempler i 2x2x2-tabeller (3)

Konstrueret tabel for behandling, prognose og overlevelse.

Behandling 1

<i>Prognose</i>	<i>Overlevende</i>		<i>Antal i alt</i>
	<i>Antal</i>	<i>Procent</i>	
<i>God</i>	80	80%	100
<i>Dårlig</i>	20	40%	50
<i>Alle</i>	100	67%	150

Behandling 2

<i>Prognose</i>	<i>Overlevende</i>		<i>Antal i alt</i>
	<i>Antal</i>	<i>Procent</i>	
<i>God</i>	40	80%	50
<i>Dårlig</i>	10	40%	25
<i>Alle</i>	50	67%	75

Her virker de to behandlinger virker lige godt i hver prognosegruppe, og de virker også lige godt ved den samlede sammenligning, idet forholdet mellem patienter i de to prognosegrupper er det samme i begge tabeller.

Sammenligning af behandlinger

<i>Behand- ling</i>	<i>Overlevende</i>		<i>Antal i alt</i>
	<i>Antal</i>	<i>Procent</i>	
1	100	67%	150
2	50	67%	75

Taleksemppler i 2x2x2-tabeller (4)

Konstrueret tabel for behandling, fregner og overlevelse.

Behandling 1

<i>Fregner</i>	<i>Overlevende</i>		<i>Antal i alt</i>
	<i>Antal</i>	<i>Procent</i>	
<i>Ja</i>	60	60%	100
<i>Nej</i>	6	60%	10
<i>Alle</i>	66	60%	110

Behandling 2

<i>Fregner</i>	<i>Overlevende</i>		<i>Antal i alt</i>
	<i>Antal</i>	<i>Procent</i>	
<i>Ja</i>	8	80%	10
<i>Nej</i>	80	80%	100
<i>Alle</i>	88	80%	110

Da fregner ikke har nogen sammenhæng med overlevelse, så er overlevelsen den samme i den samlede tabel som i de opdelte tabeller og sammenligning mellem de to behandlinger kan foretages i den samlede tabel

Sammenligning af behandlinger

<i>Behand- ling</i>	<i>Overlevende</i>		<i>Antal i alt</i>
	<i>Antal</i>	<i>Procent</i>	
<i>1</i>	66	60%	110
<i>2</i>	88	80%	110

Undersøgelse af sammenhæng mellem flere variable.

Ud fra taleksemplerne og den forforståelse der er frembragt gennem de opdigtede eksempler formuleres betingelser for

- Hvad er relevante baggrundsvariable?
- Hvornår må man se bort fra en baggrundsvariabel?

I en videregående analyse kan man behandle

- Egentlig modellering. (Betinget uafhængighed)
- Hvordan man inddrager baggrundsvariable i den statistiske analyse.

Design af undervisningsmaterialer i matematik

Matematik fremstilles traditionelt i et deduktivt system, og lærebøgerne er oftest organiseret efter princippet

- Først teori og begreber; derefter eksempler der uddyber det teoretiske stof.

I mit foredrag og i min materialeoversigt har jeg valgt en anden vej

- Fra eksempler gennem operationaliseringer til teoretiske begreber.

(Begge organisationsmåder indeholder dog en spiralformet bevægelse, idet de gentages på højere niveauer byggende på både teori og eksempler fra den forudgående undervisning.)

Materialet om de tredimensionale tabeller er åbent for andre læsninger end den jeg har foretaget. F.eks. ved

- at begynde med at definere betinget uafhængighed og konfundering for derefter at opfatte de opdigtede avisnotitser som illustrationer af de gennemgåede principper.

- at begynde med taleksemplerne

- at fokusere på selektionsproblemer (som kan illustreres ud fra de valgte eksempler)

- ??

Rhizomet som model

I en traditionel lærebog ville bilagene komme efter hinanden med henvisninger bagud. Med internettet som et muligt medium kunne de imidlertid organiseres anderledes, således at de enkelte bilag var selvstændige og kunne læses i forskellig rækkefølge afhængigt af, hvilke erkendelsesinteresser der stod i fokus og hvilken tilgang, der passede bedst til den enkelte elevs læringsmønstre.

Dette kunne lægge op til det som f.eks. Deleuze og Guattari betegner som en *rhizomatisk læsning*.

Rhizom er et ord hentet fra botanikken, hvor det betegner en *rodstok*. Botanisk har et rhizom nogle særlig kvaliteter og kendetegn, der gør det velegnet som figur for en ny måde at tænke matematikundervisning på. Det er stærkt forgrenet, det skyder vildt og uordnet, det er et sammenhængende, men indfiltret netværk, hvor ethvert punkt kan og skal forbindes til alle andre. Det dør ikke, hvis der bliver hakket i det, tværtimod frembringes der nye skud fra de nye ender, og hvis det hakkes helt over, danner det nye rhizomer. Et brud er således ikke ødelæggende for vækst i den måde rhizomet er opbygget på. Bruddets betydning bliver at være skabende snarere end dræbende.

Deleuze og Guattari ser rhizomet som en modsætning til træet som mønster, og de fremhæver, at træmodellens fokusering på orden og kausalitet mellem hierarkisk forbundne punkter og positioner hæmmer kreativiteten, mens rhizomet fremmer denne.

Materiale

1. Simpsons paradoks: Indledende problemstilling
2. Opdigtede avisnotitser
3. Diskussion af de opdigtede avisnotitser
4. Skema: Hvad skal man tage hensyn til.
5. Egne opdigtede avisnotitser
6. Operationalisering af sammenhæng mellem variable
7. Sammenhæng mellem variable (Teoretiske begreber, modellering)
8. Mål for afhængighed
9. Taleksempler: 2x2x2- tabeller
10. Hvornår kan man se bort fra baggrundsvariable?
(Operationalisering)
11. Hvornår kan man se bort fra baggrundsvariabel?
(Teoretiske begreber, modellering)
12. Eksempler fra aviser og andre medier.
13. Grafisk illustration af Simpsons paradoks.
14. Tabeller med usikkerhed
15. Regression
16. Videregående eksempel: "Unge studenter"
17. Statistisk sammenhæng og kausalitet