

Kjære alle!

Ikke alle er på facebook. Her er en oppsummering av de siste timenes diskusjoner i STK1000-gruppa.

- Spørsmål: Kan vi få utskrift fra Minitab?

Svar: Dere får R, ikke minitab. Minitab ble brukt i STK1000 frem til og med 2015. Fra og med 2016 har det vært R. Det er derfor jeg har lagt ut både ordinært eksamenssett fra i fjor, og konteeksamenssettet fra i fjor (selv om det ikke ligger en fasit til den ute), så dere har litt mer å se på. Det er uansett nyttig å prøve seg på noe uten fasit, for det blir sånn i morgen også.

- Hei! Raskt spørsmål: Man skal alltid korrigere for en confounder, men ikke korrigere for en collider. Men hva med en mediator? Er det lurt å korrigere for det, eller ser man an situasjonen? 😊😊

Det korte svaret er ja, du har rett. Det kommer an på problemstillingen.

Det lange svaret er ikke noe nytt i forhold til det dere har hørt på forelesninger/lest i boka, men nå er dere klare for det 😊:

1) Man skal alltid korrigere for confoundere, for de fører til bias i tallstørrelsen for sammenhengen du ønsker å studere. En confounder er definert som (jeg siterer fra min egen doktorgradsavhandling:

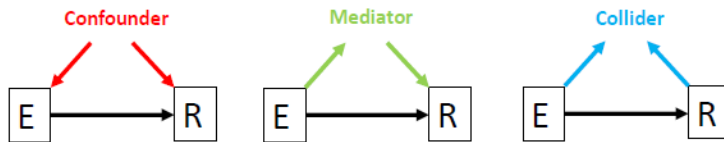
Confounding is bias of the estimated effect of an exposure on an outcome due to the presence of a common cause of the exposure and the outcome (1). Confounding is an important issue in observational designs, and may lead to underestimation, overestimation, or even change the sign of the estimated effect (17). A confounder is a variable that is associated with the outcome (either as a cause or a proxy for a cause, but not as an effect of the disease), associated with the exposure, and not an effect of the exposure (17;121). The definition of confounding may also include bias due to baseline differences in exposure groups in the risk factor for the outcome, although this may be considered as selection bias (1), as in this thesis.

2) Man kan korrigere for en mediator, hvis problemstillingen legger opp til det. Noen ganger kan det være riktig å korrigere, andre ganger ikke.

I eksemplet fra forelesningen der vi ønsker å studere sammenhengen mellom alvorlighet av skade, og stressreaksjoner (Post-traumatisk stress), kan man argumentere for at varighet av sykehusopphold er en mediator (en mellomliggende variabel) mellom de to: Skadeomfang påvirker lengden på sykehusopphold, og lengden på sykehusopphold påvirker stress. Hvis man ønsker å beregne totaleffekten av skaden på post-traumatisk stress, skal man ikke korrigere for lengden på sykehusoppholdet. Hvis man derimot ønsker å estimere den delen av effekten av skade som gir stress via andre mekanismer enn sykehusopphold (fjerne effekten av sykehusopphold), skal du korrigere.

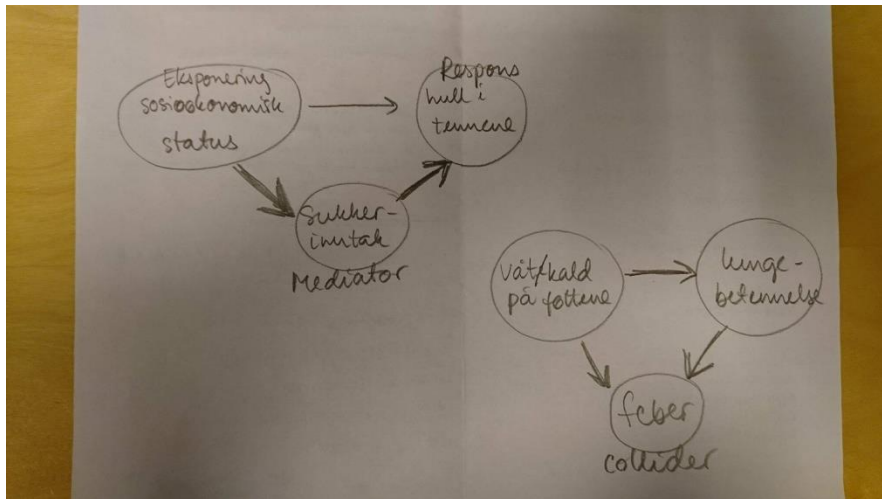
3) Hvis en variabel er en collider (en konsekvens av både eksponeringsvariabelen og responsvariabelen), vil du introdusere bias i sammenhengen hvis du korrigerer for den, og det er derfor feil å korrigere. Du skal aldri korrigere for collidere.

I STK1000 har vi tegnet confounder, mediator og collider veldig enkelt,



Note: The presence of either of these may affect the association of interest. Hence, including either of these in regression analyses may change effect estimates.

og her er to gode eksempler som Thora kom med.



- Konteeksamen 2017, 2b) skal vi her bruke t eller 1.96 i formelen for utregning av konfidensintervallet?

Spørsmålet dreier seg om sammenligning av to små grupper med normalfordelte data, der vi estimerer sigma med sd. Da er det t-test vi bruker. Z-test (1.96) er aktuelt 1) når gruppene er små, normalfordelte og vi kjenner sigma (skjer bare i lærebøker) og 2) når gruppene er store, enten de er normalfordelte eller ikke (da er det ikke viktig om vi kjenner sigma eller ikke, men det gjør vi som sagt bare i lærebøkene).

- BMI forklarer både blodsukker nivå og Fødselsvekt og er derfor en konfunderende variabel.
Er da beinbrudd konfunderende når den er et resultat av alder og beintetthet?
Eller gjelder det kun variabler som forklarer andre variabler?

Svar: Se det lange svaret på forrige side, og tegn opp sammenhengene i en DAG.

- Har et spørsmål til oppgave 1b på konteeksamen 2016. I formelen for 95% KI, hvilken n er det man bruker? Er det antall med høyere utdanning (47 stk) eller er det det totale antallet (95 stk)?

Spørsmålet dreier seg om estimering av p i en binomisk fordeling. Da bruker vi totaltallet som n. Se Ch 5.2 i læreboka, og side 10-13 i forelesningsnotatene

<http://www.uio.no/studier/emner/matnat/math/STK1000/h17/forelesningsnotater-uke-39-og-40.pdf>

Og side 3 her:

<http://www.uio.no/studier/emner/matnat/math/STK1000/h17/bonus-2-fredag-1.12.pdf>

- Har jeg rett når jeg tenker at oppg. 1, spes. b) og c) i dette eksamenssettet ikke er veldig relevant for oss???

Samme svar som på forrige spørsmål

- I eksamensoppgaven 2015 3 b) står det : "Sett opp modellen som regresjonsanalysen er basert på." I fasiten står svaret som: " Yi er uavhengige og normalfordelte $N(\beta_0 + \beta_1 x_i, \sigma)$ ".
Jeg har nok falt litt av lasset når det gjelder å definere en "modell", og lurte på hjelp i form av noen referanser til boka eller forelesninger der modell-begrepet presenteres.

Dette ble utdypet på bonus-forelesningene.

- Hei, hva er forskjellen på ensidig og tosidig hypotese? Hva er kriteriene for de to? 😊)

Svar fra medstudent: Kort oppsummert fra det de nevner i videoen: ensidig er f.eks "Denne pillen minker feber" en to sidig kan være "Denne pillen minker/øker feber" altså den har en effekt i to retninger mens ensidig er at den har en effekt i enten positiv eller negativ retning

- Trenger vi å kunne bruke regresjonsligninger for flere enn to variabler til eksamen?
For å for eksempel regne ut 95% KI for tre variabler?

Svar: Det er strengt tatt innenfor pensum for Ch 11, men hvis du klarer å regne ut KI nr modellen har to kovariater (tre regresjonsparametere, og $t_{(n-3)}$ som i fjorårets eksamen) er det rett frem å generalisere til tre kovariater (fire regresjonsparametere, og $t_{(n-4)}$.) men hvis det er effektstørrelsen du er ute etter (ikke prediksjon), vil du måtte vurdere hvorvidt begge de to ekstra kovariatene virkelig er confoundere (eller mediatorer eller collidere).

Følgespørsmål: Er det n minus antall kovariater som angir hvor mange frihetsgrader som skal brukes for T? Trodde det var n minus 1 (der man bruker den minste n dersom det er to grupper)

Antall frihetsgrader (i t-fordelingen) beregnes forskjellig avhengig av hvilken modell som ligger bak (og hvor mange parametere som må estimeres). I en ett-utvalgs t-test er det n-1, i en to-utvalgs t-test avhenger det av hvilke antakelser du legger til grunn når du estimerer standardfeilen til differansen (de ulike versjonene står i boka), og i regresjonsanalyser er det som jeg skriver n-(antall regresjonsparametere). (Ikke n-(antall kovariater), for du har nesten alltid en parameter for konstantleddet β_0 i ligningen også.)

LYKKE TIL!!!