

Eksempel 281:

Simpsons paradoks

– et emneforløb i statistik om sammenhænge og skjulte variable.

Formål:

At give eleverne et kritisk beredskab over for påstande om kausale sammenhænge gennem en undersøgelse og en diskussion af begrebet falsk sammenhæng (konfundering). At diskutere afhængighed og uafhængighed (henholdsvis rent stokastisk uafhængighed og betinget uafhængighed).

Indhold og forløb:

Simpsons paradoks fortæller den umiddelbart overraskende historie, at såfremt to hospitaler tilbyder deres patienter to forskellige slags behandlinger *A* og *B* af samme sygdom, og det for begge hospitaler gælder, at behandlingsform *A* er den mest effektive, så kan resultatet, når man lægger tallene sammen, godt være, at behandlingsform *B* er den bedste. Paradokset er dermed eksemplarisk til at sætte fokus på problemet med skjulte variable.

Der arbejdes i grupper med analyse af en række autentiske og opdigtede avisnotitser. Eleverne prøver også selv at opdigte nye.

Gennem arbejdet med eksemplerne søger eleverne efter svar på spørgsmål som:

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

I visse tilfælde kan almindelig sund fornuft fortælle én, at en bestemt baggrundsvariabel ikke kan influere på undersøgelsen, eller omvendt at en ikke medtaget variabel må have stor indflydelse. Men i enhver undersøgelse er der mange mulige baggrundsvariable, og man kan ikke altid forudse, hvad man kan se bort fra. Et samarbejde med andre fag kan give bedre grund under fødderne i nogle sammenhænge.

Der kan laves øvelser i at opstille skemaer over alle tænkelige baggrundsvariable, som man burde opdele en analyse af et materiale efter.

Dette kan evt. kobles sammen med en forberedelse til indsamling af et autentisk materiale via eksperimenter eller spørgeskemaer: Hvordan tilrettelægges sådanne undersøgelser? Hvorfor er det vigtigt at foretage sådanne analyser før materialet indsamles?

Er der stærkt belæg for en hypotese om sammenhæng, vil en regressionsanalyse af og til kunne bidrage til analysen og afsløre, at der er andet på spil. Men hvad dette andet er, vil ofte kræve en indsigt, andre fag kan bidrage med.

Efter de indledende diskussioner kan man også vælge at koncentrere sig om et beslægtet forløb: En undersøgelse af om det er korrekt, at nedsat fødselshyppighed fører til, at brødre og søstre bliver sjældne væsner?

Produktet:

Kan være elevernes egne opdigtede – eller i aviserne fundne – notitser om tilsyneladende sammenhænge. Præsenteres f.eks. via posters, sammen med kommentarer om kvaliteten af statistiske undersøgelser og konklusioner.

Materialer:

På fagets side på www.emu.dk ligger et notemateriale, som kan anvendes som grundlag for forløbet.