

## Projekt 9.7 Testet positiv – men er man syg? (B og A)

Dette projekt har hentet materiale fra Susanne Ditlevsens film, og har en vis fællesmængde med den indledende fortælling i kapitel 9. Vælger man at gennemføre et egentligt forløb om betingede sandsynligheder, kan materialet her indgå som en slags paradigmatisk eksempel.

I Susanne Ditlevsens film præsenteres en case, hvor der er udviklet en bestemt test til at afsløre om man har en given sygdom. Den nye metode er åbenbart rigtig god, idet den faktisk fanger langt hovedparten af de der er syge.

Antalstabeller kan hjælpe med til at give overblik over nogle oplysninger og spørgsmål, der i første omgang forekommer lidt udviklede. Så for at få overblik over tallene kan det være en fordel at opstille de oplysninger vi har i en antalstabel:

test \ tilstand	syg	rask	I alt
positivt udslag	a	c	a+c
ikke positivt udslag	b	d	b+d
I alt	a+b	c+d	

### Øvelse 8.1

- Se nu den del af filmen igen og noter dig hvilke oplysninger du får. Hvor i tabellen kan du indsætte disse oplysninger?
- Læg mærke til, at det er procent-oplysninger. De enkelte procenttal der nævnes må jo være procent af noget. Men af hvad?

### Øvelse 8.2

Det spørgsmål vi gerne vil have svar på er: Hvis jeg bliver testet positiv, hvad er så sandsynligheden for at jeg faktisk er syg. Hvilke tal i tabellen skal man kende, for at kunne svare på det?

Se nu igen den del af filmen, hvor Susanne diskuterer med de studerende i auditoriet. Der kommer to svar, som vi åbenbart kan bruge. Noter disse svar. Et af dem taler om et begreb, der hedder *falsk positiv*.

### Øvelse 8.3. Begrebet falsk positiv

- Hvad betyder dette begreb, *falsk positiv*?
- (*Forudsætter du har arbejdet med hypotesetest*) Når vi arbejder med hypotesetest udregnes en p-værdi på grundlag af en nulhypotese. p-værdien sammenlignes med et på forhånd fastlagt signifikansniveau, og størrelsen af p-værdien afgør om vi forkaster eller accepterer nulhypotesen. Der er således indbygget en subjektivt fastlagt grænse mellem accept og forkastelse i hypotesetest. Dette betyder også, at vi kan begå fejl to typer af fejl: Forkaste noget sandt eller acceptere noget falsk. Læs evt. afsnittet om retsagsmetaforen i *Hvad er matematik?* B og svar på, hvad sammenhængen er mellem *falsk positiv*, *falsk negativ* og de to fejltyper i hypotesetest.

- c) Argumenter for, at hvis vi for et øjeblik glemmer selve tallene, og giver en rent kvalitativ beskrivelse af de enkelte rubrikker i tabellen, så kan det gøres således:

test \ tilstand	syg	rask	I alt
positivt udslag	sand positiv (a)	falsk positiv (c)	a+c
ikke positivt udslag	falsk negativ (b)	sand negativ (d)	b+d
I alt	a+b	c+d	

Giv med dine egne ord en beskrivelse af, hvad der dækker sig bag de andre tre begreber.

#### Øvelse 8.4

Der kom yderligere et svar fra en af de studerende. Hvor i tabellen er vi henne med dette svar?

#### Øvelse 8.5

I Susannes opsummering på casen trækkes nogle absolutte tal ind på scenen. Hvor i antalstabellen hører disse til? Overvej hvorfor vi med de nye oplysninger faktisk kan svare på spørgsmålet.

#### Øvelse 8.6. Case om betydningen af antallet af smittede

Løs nu selv følgende øvelse med brug af antalstabeller som ovenfor.

Antag vi har en test for HIV, der er meget effektiv, idet den fanger alle der er smittede. Testen har en falsk positiv rate på 5%, og en falsk negativ rate på 0%.

Vi har en samlet population på 1000, og ser på to forskellige situationer:

A) 40% er smittede                      B) 2% er smittede

Du testes positiv. Hvad er sandsynligheden for at du faktisk er smittet i henholdsvis tilfælde A og tilfælde B?

Der findes mange beslægtede problemstillinger, hvor vi så at sige gerne vil regne baglæns:

- fra en viden om hvor mange der testes positivt til en sandsynlighed for, hvor mange der faktisk er syge
- fra en viden om et barns og en potentiel faders blodtyper, til en sandsynlighed for at han faktisk er far til barnet.
- fra en viden om et dna-materiale fundet på et gerningssted og om en mistænkt persons dna-profil, til sandsynligheden for at det fundne materiale faktisk stammer fra den mistænkte.

De sandsynligheder vi her taler om kaldes *betingede sandsynligheder*. Som udgangspunkt har vi stadig et *frekventielt* grundlag for opstilling af sandsynligheder: Sandsynligheden for at en tilfældig nyfødt i Danmark er en dreng er 49,1%, fordi 49,1% af alle nyfødte er drenge, dvs frekvensen af drenge er 49,1%.

Ideerne og metoderne inden for denne del af sandsynlighedsregning behandles i *projekt 9.8*, hvor vi også giver en kort introduktion til *Bayesiansk statistik*. Det er en gren af statistikken, nogle vil sige en skole inden for statistikken, hvor man populært sagt forlader det frekventielle grundlag og i stedet *tager udgangspunkt i opstilling af betingede sandsynligheder*.