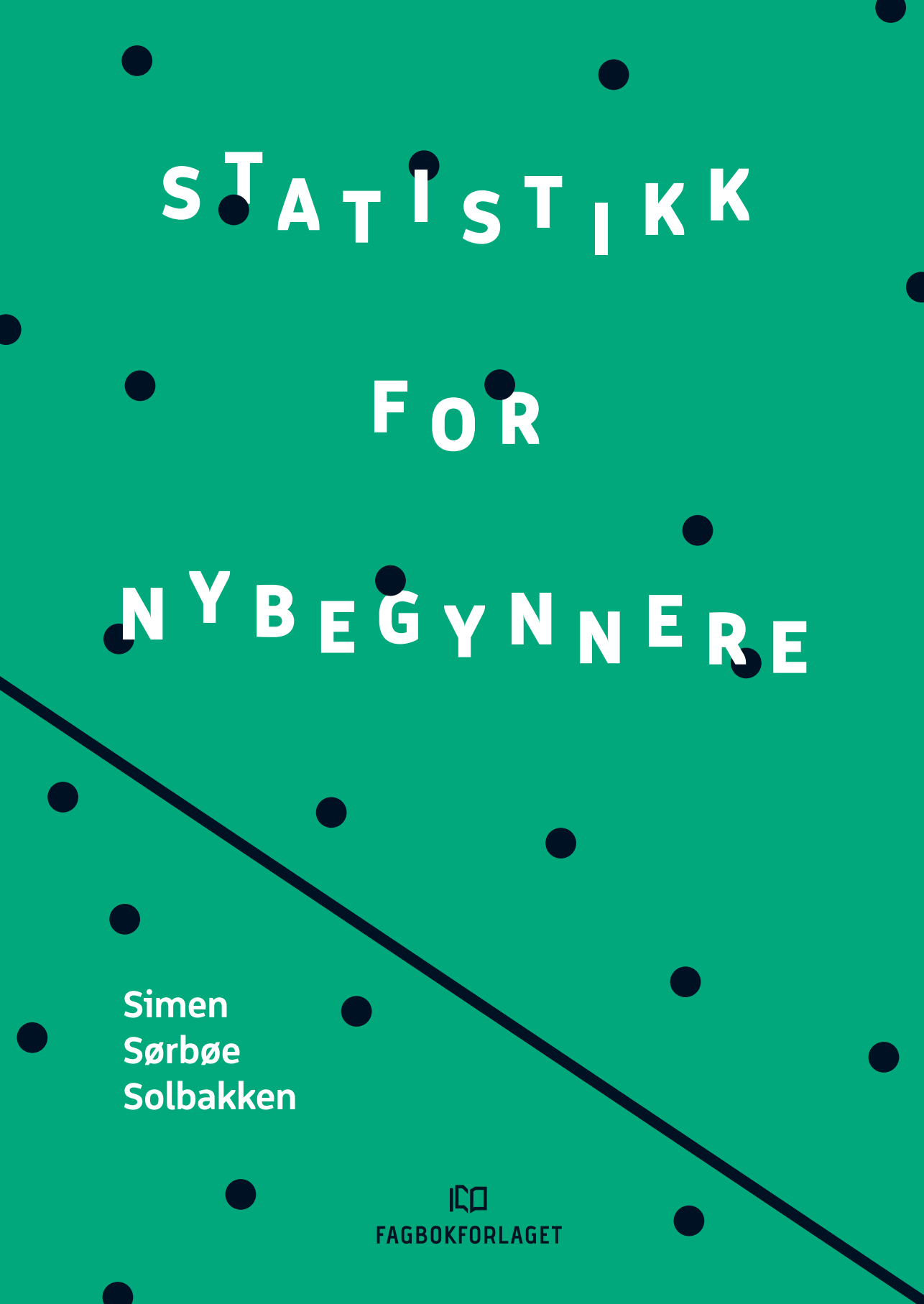


The background is a solid teal color. It is decorated with numerous black dots of varying sizes scattered across the surface. A thick, solid black diagonal line runs from the top-left corner towards the bottom-right corner, passing behind the title text.

STATISTIKK FOR NYBEGYNNERE

Simen
Sørbøe
Solbakken



FAGBOKFORLAGET

STATISTIKK FOR NYBEGYNNERE

Til Daniel, Nathaniel og Jonathan

SIMEN SØRBØE SOLBAKKEN

STATISTIKK FOR NYBEGYNNERE



FAGBOKFORLAGET

Copyright © 2019 by
Vigmostad & Bjørke AS
All Rights Reserved

1. utgåva / 1. opplaget 2019

ISBN:978-82-450-2939-0
ISBN (trykt): 978-82-450-2351-0

Tilrettelagt for e-bok: John Grieg, Bergen
Forsidedesign ved Tsz Yan Tong
Sats ved forfatteren

Spørsmål om denne boka kan rettast til:
Fagbokforlaget
Kanalveien 51
5068 Bergen
Tlf.: 55 38 88 00 Faks: 55 38 88 01
e-post: fagbokforlaget@fagbokforlaget.no
www.fagbokforlaget.no

Materialet er verna etter åndsverklova.
Utan uttrykkjeleg samtykke er eksemplarframstilling
berre tillate når det er heimla i lov eller avtale med Kopinor.

Innhold

Forord	11
1 Introduksjon	13
1.1 Sentrale begreper	13
1.1.1 Vitenskapelig metode	13
1.1.2 Tema, problemstilling og hypoteser	14
1.1.3 Enheter	15
1.1.4 Populasjon	16
1.1.5 Utvalg	17
1.1.6 Variabler	18
1.1.7 Verdier	18
1.1.8 Univariat, bivariat og multivariat	20
1.1.9 Analyse	22
1.1.10 Observerte, estimerte og predikerte verdi	22
1.2 Matematikk	23
1.2.1 Matematisk uttrykk	24
1.2.2 Matematisk notasjon	24
1.2.3 Parenteser	25
1.2.4 Fortegnsregler	25
1.2.5 Brøker	26
1.2.6 Kvadratrot	26
1.2.7 Potens og eksponent	26
1.2.8 Absoluttverdien til et tall	27
1.2.9 Tilfeller (X_i)	27
1.2.10 Sigma (Σ)	29
1.2.11 Estimerte og predikerte verdier med hått eller k��ll	30
1.2.12 Prosent og proporsjon	31
1.2.13 Prosent og prosentpoeng	31

2	Data	33
2.1	Utvelging av enheter	34
2.1.1	Tilfeldige utvalgsfeil	34
2.1.2	Systematiske utvalgsfeil	35
2.1.3	Sannsynlighetsutvelging	36
2.1.4	Enkel tilfeldig utvelging	37
2.2	Definisjoner og operasjonalisering	37
2.3	Validitet	39
2.4	Ytre validitet	40
2.5	Indre validitet	41
2.6	Målevaliditet	41
2.6.1	Umiddelbar validitet	42
2.6.2	Definisjonsmessig validitet	43
2.6.3	Reliabilitet	44
2.7	Variabelens målenivå	46
2.7.1	Nominalnivå	48
2.7.2	Ordinalnivå	48
2.7.3	Metrisk nivå	50
2.7.4	Dikotome variabler: Variabler med to verdier	51
2.8	Kategoriske, diskrete og kontinuerlige variabler	52
3	Univariat analyse	55
3.1	Datamatrise	55
3.2	Fordelinger	56
3.2.1	Frekvens, prosent og sannsynlighet	57
3.2.2	Grafer	57
3.2.3	Fordelinger og skjevhet	63
3.2.4	Ulike fordelinger	63
3.3	Sentraltendens	66
3.3.1	Gjennomsnitt	67
3.3.2	Median	68
3.3.3	Modus	69
3.3.4	Valg av mål for sentraltendens	70
3.4	Spredning	75
3.4.1	Variasjonsbredde	77
3.4.2	Kvartildifferanse	78
3.4.3	Standardavvik og varians	79
3.4.4	Valg av spredningsmål	82

4	Bivariat analyse	85
4.1	Eksempel på sammenheng mellom kjønn og inntekt	86
4.2	Modeller	87
4.3	Tolkning av spredningsplott	87
4.4	Sammenhengens egenskaper	92
4.4.1	Sammenhengens styrke	92
4.4.2	Sammenhengens retning	95
4.4.3	Sammenhengens styrke og retning	99
4.4.4	Sammenhengens form	101
4.5	Pearsons r	105
4.5.1	Ingen sammenheng	106
4.5.2	Sammenhengens retning	107
4.5.3	Deterministisk sammenheng	109
4.5.4	Stokastisk sammenheng	111
4.5.5	Kurvelineære sammenhenger	111
4.5.6	Pearsons r og målenivå	115
4.6	Skalaretning	115
4.6.1	Intuitiv og ikke-intuitiv skalaretning	120
4.6.2	Variabler uten en intuitiv skalaretning	120
5	Slutningsstatistikk	123
5.1	Hypoteser	124
5.1.1	Formulering av hypoteser	124
5.1.2	Nullhypotesen og den alternative hypotesen	125
5.1.3	Type I-feil og type II-feil	127
5.2	Feilmarginer og konfidensintervall	129
5.2.1	Sannsynlighetsfordelingen	132
5.3	Advarsel om språklig forvirring	135
5.4	Signifikans og konfidens	135
5.5	Sikkerhet, usikkerhet og presisjon	138
5.6	Z-fordelingen og t -fordelingen	138
5.6.1	t -tabellen	141
5.7	Beregning av resultatenes presisjon	144
5.7.1	Velge signifikansnivå	144
5.7.2	Finne antall frihetsgrader	144
5.7.3	Finne riktig t -verdi i t -tabellen	145
5.7.4	Finne standardfeilen i utvalget	145
5.7.5	Beregning av feilmarginer	147
5.7.6	Beregning av konfidensintervallet (nedre og øvre grense)	147
5.8	Tolke feilmarginer og konfidensintervaller (nedre og øvre grense)	150

5.9	Høyere presisjon med mindre feilmarginer og smalere konfidensintervall	150
5.9.1	Øke signifikansnivået	151
5.9.2	Øke størrelsen på utvalget	152
5.9.3	Når størrelsen på utvalget nærmer seg størrelsen på populasjonen	153
5.10	Hypotesetesting	154
5.10.1	Hypotesetest med konfidensintervall	155
5.10.2	Signifikanstesting	161
5.10.3	Signifikanstest med p -verdier	163
5.10.4	Signifikanstestene t -test og Z -test	165
5.10.5	Tommelfingerregel for t -test	170
5.10.6	Når regresjonskoeffisienten (b_i) er «signifikant» eller «signifikant forskjellig fra 0»	170
6	Regresjonsanalyse: Introduksjon	171
6.1	Regresjonslinjen	173
6.2	Avhengig og uavhengig variabel	173
6.3	Regresjonsmodeller	175
6.4	Residual (e)	177
6.5	Valg av regresjonsmodell	179
6.6	Valg av regresjonslinje	180
6.6.1	Minste absolutte avvik (LAD)	181
6.6.2	Minste kvadraters metode (OLS)	182
6.7	Konstantleddet (b_0) til regresjonslinjen	185
6.8	Stigningstallet til regresjonslinjen: Regresjonskoeffisienten (b_1)	187
6.9	Forventet Y med én uavhengig variabel (X)	190
6.9.1	Meningsløse og uinteressante verdier	193
6.9.2	Grunnløse verdier	193
6.10	Observert Y for en faktisk enhet	194
6.11	Standardfeilen $SE(b_i)$ til en regresjonskoeffisient	196
6.12	Beregning av feilmarginer og konfidensintervaller (nedre og øvre grense)	199
6.13	Hypotesetesting og signifikantesting i regresjonsanalyse	201
7	Multivariat analyse: Stianalyse	203
7.1	Stimodeller	203
7.2	Bakenforliggende og mellomliggende variabler	208
7.3	Forskjell mellom <i>sammenheng</i> og <i>effekt</i>	210

7.4	Forskjellen mellom <i>bivariat sammenheng</i> og <i>direkte effekt</i> . . .	211
7.5	Effektkomponenter	212
7.5.1	Direkte effekt	214
7.5.2	Indirekte effekt	215
7.5.3	Spuriøs effekt	220
7.5.4	Total effekt	229
7.5.5	Delvis direkte, indirekte og spuriøse effekter	229
7.6	Stianalyse	231
7.7	Betydningen av stianalyse	233
7.8	Utelatte variabler	234
7.9	Kausalitet og eksperimenter i samfunnsvitenskap	236
8	Multipel regresjonsanalyse	241
8.1	Multiple regresjonsmodeller	241
8.2	Spredningsplott i multipel regresjon	248
8.2.1	Spredningsplott med tre variabler	248
8.3	Minste kvadraters metode (OLS) i multipel regresjon	249
8.4	Konstantleddet (b_0) i multipel regresjon	251
8.4.1	Grafisk tolkning av konstantleddet i multipel regresjon	253
8.5	Stigningstall i multipel regresjon	255
8.5.1	Repetisjon av regresjonsmodeller med én X	257
8.5.2	Regresjonskoeffisienter (b_i) som direkte effekt av X_i	261
8.5.3	Kontrollert effekt	267
8.6	Regresjonstabeller	270
8.6.1	Tolkning av stjernene (***) i en regresjonstabell	272
8.7	Forklart varians (R^2)	272
8.7.1	Endring i R^2	274
8.7.2	R^2 , korrelasjon og kausalitet	275
8.7.3	R^2 og Pearsons r	276
8.8	Standardiserte regresjonskoeffisienter (β)	276
8.8.1	Standardisert konstantledd (β_0)	279
8.9	Regresjonsanalyse og målenivå	279
8.9.1	Variabler mellom ordinalnivå og metrisk nivå	280
8.10	Dummyvariabler	282
8.10.1	Tolkning av b_1 til dummyvariabler	287
8.11	Samspill	288
8.11.1	Fullstendig betingede effekter	289
8.11.2	Delvis betingede effekter	294
8.11.3	Samspill i regresjonstabeller	296
8.11.4	Forskjell mellom samspill og indirekte effekt	298

8.11.5	Samspill og skalaretning	299
8.11.6	Samspill uten dikotome variabler	301
8.11.7	Samspill med tre eller flere variabler	301
9	Forutsetninger for å benytte OLS	303
9.1	Hvorfor OLS?	303
9.1.1	(1) OLS er enkelt å forstå	303
9.1.2	(2) OLS gir oss de beste resultatene	304
9.2	Fire forutsetninger for å benytte OLS	309
9.3	Tilnærmet normalfordelte residualer	310
9.3.1	Er residualene normalfordelte?	310
9.3.2	Konsekvenser av ikke-normalfordelte residualer	312
9.3.3	Tiltak mot ikke-normalfordelte residualer	313
9.4	Homoskedastiske residualer	313
9.4.1	Konsekvenser av heteroskedastiske residualer	315
9.4.2	Tiltak mot heteroskedastisitet	317
9.5	Uavhengige residualer	317
9.5.1	Residualene korrelerer med en av de uavhengige variablene (X_i) i modellen	317
9.5.2	Residualene korrelerer med hverandre (<i>autokorrelasjon</i>)	320
9.6	Kolinearitet og multikolinearitet mellom de uavhengige variablene (X_i)	322
9.6.1	Er det kolinearitet mellom variablene?	325
9.6.2	Konsekvenser av <i>perfekt</i> kolinearitet	327
9.6.3	Konsekvenser av <i>sterk</i> tendens til kolinearitet	327
9.6.4	Tiltak mot <i>sterk</i> tendens til kolinearitet	328
9.6.5	Forskjellen mellom kolinearitet, indirekte effekt og spuriøs effekt	328
9.6.6	Samspill og sterk tendens til kolinearitet	329
9.6.7	Multikolinearitet	329
9.6.8	Dummyvariabler og multikolinearitet	332
9.7	Alternativer til <i>BLUE</i>	333
9.8	<i>BLUE</i> er ikke alltid relevant	333
	Sluttord	335
	Register	337

Forord

De fleste barn lærer å krabbe før de lærer å gå. Mitt mål er at du – gjennom å lese denne boken – skal krabbe inn i statistikkens verden, slik at du på et senere tidspunkt kan lære å gå eller attpåtil løpe. En dag løper du kanskje maraton, det vil si kjører en rangert multinomisk GLM-modell med panelkorrigerte standardfeil.

Albert Einstein skal ha sagt at «alt bør gjøres så enkelt som mulig, men ikke enklere enn det». Jeg er enig med Einstein. I denne boken har jeg derfor valgt å forenkle så mye som mulig. Når man bytter ut kompliserte fremmedord og definisjoner med enkle begreper og forklaringer, kan nyanser gå tapt. Disse nyansene kan du lære andre steder når de grunnleggende ferdighetene er på plass. Du vil derfor ikke finne en eneste *vitenskapelig definisjon* i denne boken. I stedet vil du finne enkle forklaringer *for nybegynnere*. Forenklinger kan være ubehagelige for dem som allerede forstår. For alle andre er forenklinger en befrielse. Denne boken er skrevet for alle andre, ikke for dem som allerede forstår.

Einstein skal ha hevdet at han ikke var spesielt flink i matematikk. For Einstein var *visualisering* nøkkelen til å forstå verden. Mange syntes, som Einstein, at *figurer* er enklere å forstå enn *formler*. Gjennom boken vil du derfor møte over 100 figurer. Forståelse er prioritert på bekostning av matematisk presisjon. Figurene må derfor anses som forenklinger, fremstilt med den hensikt å øke din forståelse. For å gjøre eksemplene så enkle som mulig, er alle tall og undersøkelser oppkonstruerte – med mindre noe annet er eksplisitt nevnt.

Statistikk handler om mye mer enn tall og formler. Det viktigste jeg selv har lært gjennom å studere statistikk, er hvor ufattelig komplisert verden er, og hvor lite vi egentlig vet *helt sikkert*. Statistikk er et fantastisk verktøy for å forstå verden. Men det kan også misbrukes – bevisst eller ubevisst – i søken etter sannhet og makt. Jeg tror verden vil bli et bedre sted dersom flere behersker statistikk, både i søken etter sannhet, og gjennom kritisk tenkning. Boken er således også et lite stykke dannelses.

Før jeg begynte på ungdomsskolen, spilte læreren min, Sylvi Bruhn, *Tenke sjæl* av Trond-Viggo Torgersen for hele klassen. Refrenget går som følger:

*Tenke sjæl og mene, måtte stå for det du sa
Ikke vri deg unna, ikke være likeglad
Ikke late som du ikke mente det du sa
Ikke si som andre
Du må tenke sjæl*

Jeg håper at denne boken kan bidra til at du forstår statistikk, og at du blir mer kritisk til alt du hører og leser. *Du må tenke sjæl.*

Denne boken ble ikke til i et vakuum. Terje Wilhelm Bardum, den flinkeste læreren jeg noensinne har hatt, lærte meg verdien av å forenkle kompliserte begreper og forklaringer. Gjennom Terjes undervisning lærte jeg at fullstendige forklaringer er ubrukelige hvis ingen forstår dem.

Jeg vil også takke Jon Hovi, Francesca Refsum Jensenius, Erlend Langørgen og May Helene Solberg i Fagbokforlaget for gode tilbakemeldinger. En spesielt stor takk må rettes til Knut-Andreas Christophersen, som har bidratt med uvurderlige tilbakemeldinger. Boken ville ikke vært den samme uten. Videre fortjener alle studentene jeg har vært så heldig å undervise en stor takk. Denne boken ble til på grunn av dere.

Til slutt vil jeg takke samboeren min, Eirin, for at hun lærte meg at pedagogikk er like viktig i høyere utdanning som i grunnskolen. Få studenter på Institutt for statsvitenskap ved Universitetet i Oslo kjenner til den innsatsen Eirin har gjort for undervisningskvaliteten ved instituttet. Eirins entusiasme for pedagogikk har smittet over på meg, og det har preget hele denne boken. Takk, Eirin.

Oslo, desember 2018

Simen Sørboe Solbakken

Kapittel 1

Introduksjon

1.1 Sentrale begreper

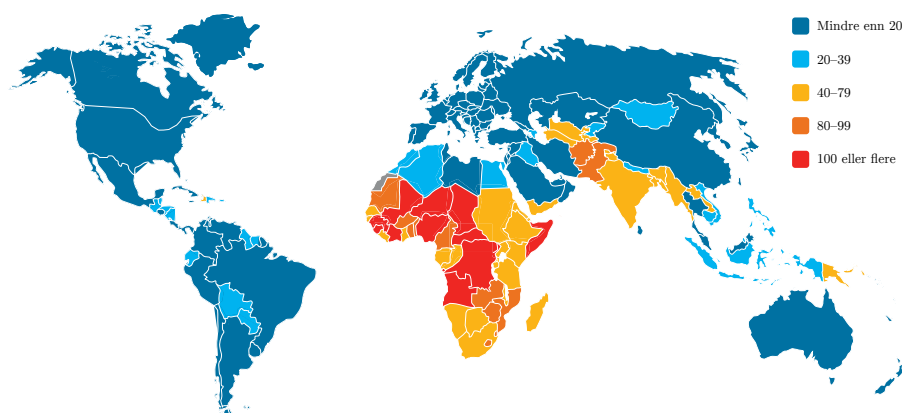
Når man skal lære seg statistikk, er man nødt til å forstå noen grunnleggende begreper. For å få mest mulig ut av resten av boken, må du både huske og forstå disse begrepene. Vitenskapelig metode, hypoteser, enheter, populasjon, utvalg, variabler, verdier, univariat, bivariat, multivariat og analyse er alle begreper du må forstå for å beherske statistikk. Heldigvis er ordene mindre kompliserte enn ordlyden gir inntrykk av.

1.1.1 Vitenskapelig metode

Vi mennesker har alltid vært opptatt av å forstå verden rundt oss. Hvorfor blir det krig? Hvordan kan fugler fly? Kan vi kurere kreft? Hva var det som førte til at Tyskland vant VM i fotball i 2014? Vi *tror* ofte at vi forstår verden. Men når vi skal finne svaret på slike spørsmål, har vi en tendens til å svare feil. For å redusere risikoen for å svare feil, kan vi bruke *vitenskapelig metode*.

En snekker har ulike verktøy som hjelper henne å bygge et hus, for eksempel hammer og sag. Hvis hun benyttet alternative verktøy til å bygge huset, slik som telepati, ville hun fort merket at det ikke fungerte. Vitenskapelig metode er de verktøyene vi har til rådighet når vi skal besvare spørsmål om verden. Gjennom flere hundre år har mennesker testet ut mange alternative metoder, slik som spåkuler. Etter mye prøving og feiling, har vi etter hvert kommet frem til de verktøyene vi i dag benytter, og som vi kaller *vitenskapelig metode*.

Vi skiller gjerne mellom *kvalitative* og *kvantitative* metoder. Disse to kategoriene tilbyr oss ulike verktøy som vi kan bruke for å forstå verden. Skillet



Figur 1.1: Antallet barn som døde før fylte 5 år per 1000 fødte barn i 2013. Kilde: statista.com og UN Inter-agency Group for Child Mortality Estimation (UN IGME).

mellom kvalitative og kvantitative metoder dreier seg i hovedsak om hva slags informasjon vi har. Hvis informasjonen er *tallfestet*, benytter vi kvantitative metoder.inntekt er et eksempel på slik informasjon. Kvantitative metoder kan også brukes til å studere fenomener som i utgangspunktet ikke er tall, for eksempel følelser. Dette forutsetter at vi tallfester følelsene, eksempelvis ved å spørre folk om å svare på spørsmålet «Hvor lykkelig er du på en skala fra 1 til 5?». Dersom vi i stedet ber folk skrive en setning om følelsene sine, er ikke lenger informasjonen tallfestet. Hvis informasjonen er noe annet enn tall, for eksempel tekst, benytter vi kvalitative metoder. *Statistikk* er en del av den kvantitative verktøykassen som forutsetter at informasjonen er tallfestet.

1.1.2 Tema, problemstilling og hypoteser

Figur 1.1 viser at det er store forskjeller mellom ulike land og verdensdeler når det gjelder antall barn som dør før fylte 5 år. La oss si at vi ønsker å se nærmere på dette. Kartet berører ulike ting: barnedødelighet, økonomiske forskjeller, urettferdighet, kulturelle forskjeller, tilgang på medisiner, osv. Vi må derfor først bestemme oss for et *tema* vi ønsker å studere videre. Vi velger oss temaet *barnedødelighet*.

Når vi har bestemt oss for et tema, må vi bestemme oss for hvilket spørsmål vi skal besvare. Spørsmål som skal besvares med vitenskapelige metoder, kaller vi *problemstilling*. Det er viktig at problemstillingen er både (1) *tilstrek-*

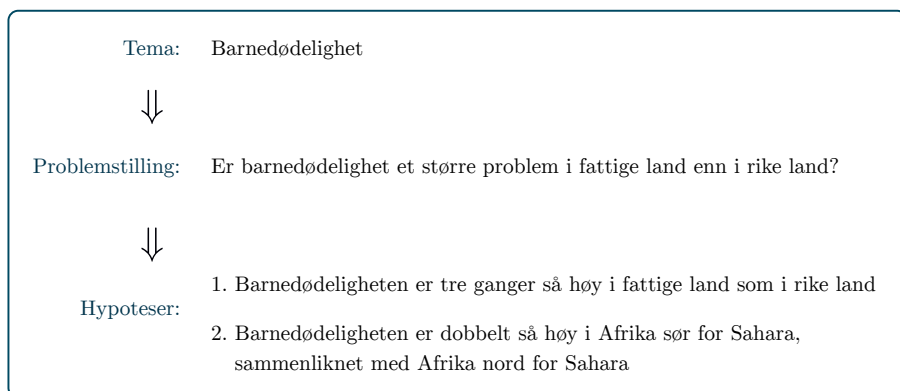
kelig presis og (2) *tilstrekkelig avgrenset*. Problemstillingen «Er barnedødeligheten mye høyere i Afrika enn i resten av verden?» er for upresis. Hva mener man egentlig med «mye høyere»? Dobbelte så høy? Tre ganger så høy? Hundre ganger så høy? Problemstillingen «Hvorfor er barnedødeligheten i Afrika høyere enn i resten av verden?» er på sin side utilstrekkelig avgrenset. Det finnes svært mange årsaker til barnedødelighet: økonomiske forskjeller, kulturelle forskjeller, tilgang på medisiner, osv. Å besvare denne problemstillingen i én undersøkelse er derfor umulig. Problemstillingen «Er barnedødelighet et større problem i fattige land enn i rike land?» er både tilstrekkelig presis og tilstrekkelig avgrenset. Vi kan derfor gå videre med denne problemstillingen.

Når vi har identifisert en problemstilling som vi ønsker å besvare, må vi formulere én eller flere *påstander* basert på problemstillingen. I statistikk kaller vi disse påstandene for *hypoteser*. En hypotese er en presis påstand om virkeligheten som vi kan teste. Et eksempel på en hypotese basert på problemstillingen over, er «Barnedødeligheten er tre ganger så høy i fattige land som i rike land». Denne hypotesen kan vi teste.

Hvis vi finner tilstrekkelig bevis for at hypotesen er sann – det vil si at den er riktig – beholder vi hypotesen. Vi kan da konkludere med at barnedødeligheten er tre ganger så høy i fattige land som i rike land, slik hypotesen vår påstod. Dersom vi ikke finner tilstrekkelig bevis for at hypotesen er sann – det vil si at den er feil – forkaster vi hypotesen. I så fall kan vi ikke konkludere med at barnedødeligheten er tre ganger så høy i fattige land som i rike land. I kapittel 5 ser vi nærmere på hvordan vi kan teste hypoteser med statistikk. Figur 1.2 illustrerer sammenhengen mellom *tema*, *problemstilling* og *hypoteser*.

1.1.3 Enheter

Enhetene i en undersøkelse er det eller de vi undersøker. I hypotesen «menn tjener mer enn kvinner» er alle enhetene *personer*. I hypotesen «det er mer krig i fattige stater enn i rike stater» er alle enhetene *stater*. Og med hypotesen «små fugler lever lenger enn store fugler» er alle enhetene *fugler*. Enhetene er med andre ord *hva* eller *hvem* som undersøkes, og hele undersøkelsen har kun én *type* enhet, eksempelvis personer, stater eller fugler. Når vi snakker om *én bestemt enhet*, sikter vi til én av enhetene, eksempelvis én av personene, statene eller fuglene.



Figur 1.2: Illustrasjon av sammenhengen mellom tema, problemstilling og hypoteser.

Enhet for nybegynnere

Enhetene i en undersøkelse er det vi undersøker. En undersøkelse har kun én *type* enhet, men består av mange enheter.

Nærliggende ord og uttrykk for *enhet*

- Observasjon
- Enkeltobservasjon
- Respondent
- Engelsk: *unit* eller *observation*

1.1.4 Populasjon

Populasjonen er alle enhetene vi ønsker å si noe om. Med hypotesen «menn tjener mer enn kvinner» ønsker vi å si noe om alle kvinner og menn. Populasjonen er dermed alle kvinner og menn. Med hypotesen «ungdomsskoleelever i Buskerud bruker i snitt én time på lekser hver dag», er populasjonen alle ungdomsskoleelever i Buskerud.

Populasjon for nybegynnere

Populasjonen er alle enhetene vi ønsker å si noe om i undersøkelsen.

Nærliggende ord og uttrykk for *populasjon*

- Univers
- Teoretisk univers

1.1.5 Utvalg

Når vi skal gjennomføre en undersøkelse, må vi hente inn informasjon om det eller de vi skal undersøke. Hvis hypotesen vår er at «norske menn bruker mer enn fire timer på husarbeid hver uke», må vi samle inn informasjon om hvor mye tid norske menn bruker på husarbeid per uke. Dersom vi skulle spurt hele populasjonen – det vil si alle norske menn – ville undersøkelsen blitt både dyr og tidkrevende. Vi må derfor ofte klare oss med et *utvalg*. Et utvalg er de enhetene som vi faktisk undersøker. I eksempelet over ville utvalget vært de mennene vi faktisk intervjuet om husarbeid. Figurene 1.3 og 1.4 gir en fremstilling av populasjon og utvalg.

Utvalg for nybegynnere

Utvalget er de enhetene som vi faktisk undersøker.

Nærliggende ord og uttrykk for *utvalg*

- Engelsk: sample



Figur 1.3: Populasjonen er alle enhetene vi ønsker å si noe om. I en meningsmåling ønsker vi å si noe om alle norske velgere.



Figur 1.4: Utvalget er de enhetene vi faktisk undersøker. I en meningsmåling er utvalget alle personene vi faktisk intervjuer.

1.1.6 Variabler

Variabler er egenskaper ved enhetene (det vi undersøker) som er relevante for undersøkelsen. For å teste hypotesen «menn tjener mer enn kvinner», må vi undersøke enhetenes kjønn og inntekt. Hypotesen inneholder med andre ord variablene **Kjønn** og **Inntekt**. For å gjøre det enklere å se hva som er variabler, vil vi bruke en liten boks rundt variabelnavn, slik som dette: **Variable**.

Variabler for nybegynnere

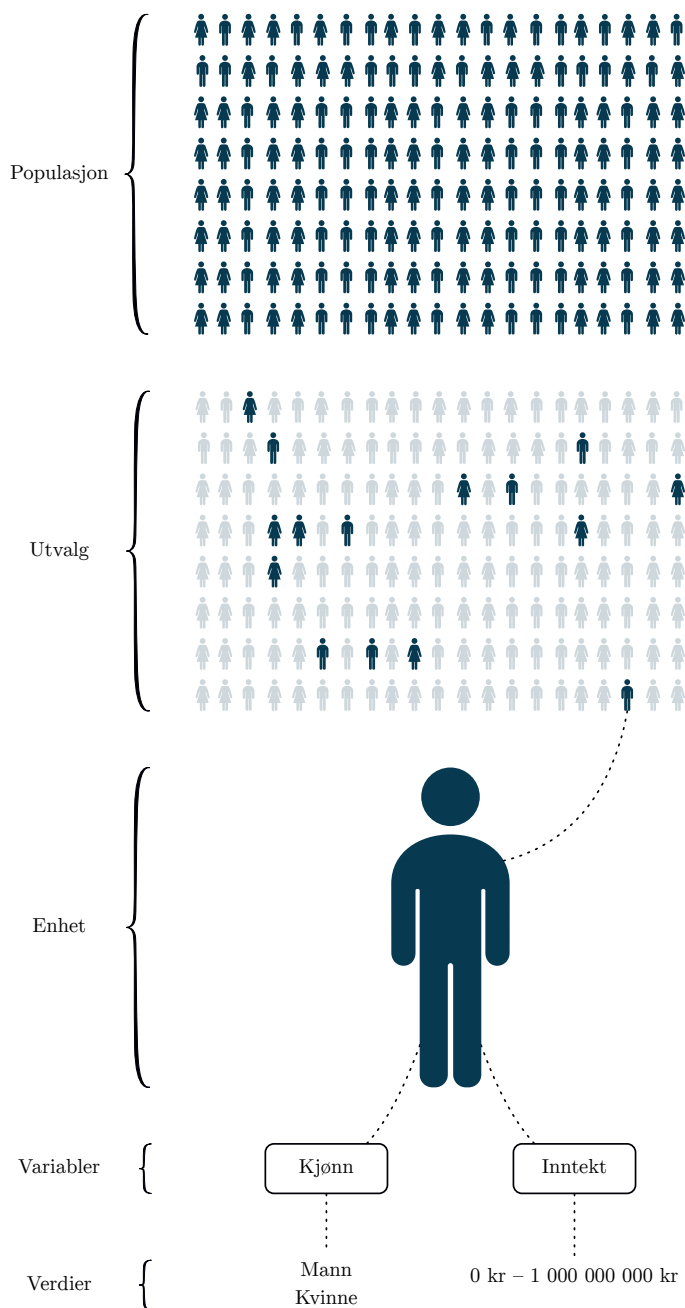
Variabler er egenskaper ved enhetene som er relevante for undersøkelsen.

Nærliggende ord og uttrykk for *variabel*

- Faktor

1.1.7 Verdier

En variabel har alltid to eller flere mulige *verdier*. Variabelen **Kjønn** kan eksempelvis ha to mulige verdier: mann og kvinne. Som du ser, har vi skrevet



Figur 1.5: En grafisk fremstilling av sammenhengen mellom *populasjon*, *utvalg*, *enhet*, *variabler* og *verdier*.