INNHOLD

FORORD.......................................................................................................................... 5

DEL I
KOM I GANG.......................................................................................................................... 17

KAPITTEL 1
HVA ER R?.............................................................................................................................. 19
1.1 Et statistikkprogram ........................................................................................................ 19
1.1.1 Installere programmet ................................................................................................. 20
1.1.2 Installere R.................................................................................................................. 20
1.1.2.1 Installere og oppdatere RStudio ............................................................................. 20
1.2 Et programmeringsspråk ............................................................................................ 22

KAPITTEL 2
R SOM KALKULATOR.............................................................................................................. 25
2.1 Åpne RStudio for første gang ........................................................................................ 26
2.2 Datatyper i R .................................................................................................................. 30
2.2.1 Vektorer .................................................................................................................... 30
2.2.1.1 Indeksering ............................................................................................................ 31
2.2.1.2 Regneoperasjoner ................................................................................................. 32
2.2.1.3 FunkSJoner og argumenter ............................................................................... 33
2.2.1.4 Sortere vektorer .................................................................................................. 35
2.2.2 Matriser ..................................................................................................................... 36
2.2.2.1 Opprette en matrise .............................................................................................. 36
2.2.2.2 Opprette et datasett ............................................................................................. 37
2.2.2.3 Indeksering .......................................................................................................... 39
2.2.2.4 Målenivå ............................................................................................................... 40
2.2.3 Listen ......................................................................................................................... 43
2.2.3.1 Opprette en liste .................................................................................................. 44
2.2.3.2 Indeksering .......................................................................................................... 44
Oppgaver ............................................................................................................................... 46
KAPITTEL 3
R SOM TEGNEPROGRAM
3.1 Plassere punkter i et koordinatsystem ........................................ 47
3.2 Opprette tegneark for å legge inn informasjon .............................. 50
  3.2.1 Opprette og tilrettelegge tegnearket .................................... 50
  3.2.2 Legge til grafikk på eksisterende ark .................................. 51
    3.2.2.1 Legge til punkter, linjer og tekst .................................. 51
    3.2.2.2 Spesifisere farge og størrelse .................................... 53
    3.2.2.3 Gi tegnforklaringer .................................................. 54
Oppgaver .................................................................................. 56

KAPITTEL 4
INN OG UT AV R ........................................................................ 57
4.1 Arbeidsflyt ........................................................................... 58
  4.1.1 Lagre eget arbeid ............................................................... 58
    4.1.1.1 Organisere kodefiler .................................................. 58
    4.1.1.2 Lagre kodefiler .......................................................... 59
    4.1.1.3 Organisere arbeid i mapper ......................................... 59
  4.1.2 Bruke andre arbeid: R-pakker .......................................... 61
4.2 Import/eksport av datafiler .................................................... 64
  4.2.1 R på hjemmebane (Rda) ...................................................... 65
  4.2.2 Skrive data ut av R (txt og .csv) .......................................... 66
  4.2.3 Åpne data i Excel/OpenOffice ............................................ 67
  4.2.4 Åpne data i R ................................................................. 68
    4.2.4.1 Import fra tekstfiler (txt og .csv) ............................... 69
    4.2.4.2 Import fra andre filformater ....................................... 70
Oppgaver .................................................................................. 71

DEL II
UTFORSKE OG TILRETTELEGGE DATA ....................................... 73

KAPITTEL 5
BESKRIVENDE STATISTIKK ...................................................... 75
5.1 Datastrukturer ....................................................................... 76
5.2 Univariate fordelinger ........................................................... 77
  5.2.1 Kategoriske variabler ......................................................... 78
    5.2.1.1 Frekvenstabell ............................................................ 78
    5.2.1.2 Kake- og søylediagram .............................................. 79
  5.2.2 Kontinuerlige variabler ..................................................... 80
    5.2.2.1 Numeriske beskrivelser ............................................. 80
    5.2.2.2 Grafiske alternativer ................................................ 82
5.3 Bivariante sammenhenger ..................................................... 86
  5.3.1 To kategoriske variabler ................................................... 86
    5.3.1.1 Bivariat sammenheng med tall .................................. 86
    5.3.1.2 Bivariat sammenheng i tegning ................................. 89
INNHOLD

5.3.2 Én kategorisk og én kontinuerlig variabel ......................................................... 90
5.3.3 To kontinuerlige variabler ................................................................................. 93
  5.3.3.1 Numeriske beskrivelser .............................................................................. 94
  5.3.3.2 Grafiske alternativer .................................................................................. 96
Oppgaver .......................................................................................................................... 100

KAPITTEL 6
TILRETTELEGGE DATA ............................................................................................... 101
6.1 Omkode eksisterende variabler ............................................................................. 101
  6.1.1 Kontinuerlige variabler .................................................................................. 102
    6.1.1.1 Additive indekser .................................................................................... 102
    6.1.1.2 Snu skalarretnin ..................................................................................... 105
    6.1.1.3 Endre skalaenheter .................................................................................. 106
  6.1.2 Kategoriske variabler (betinget omkoding) ...................................................... 106
    6.1.2.1 Omkode ved hjelp av indeksering .......................................................... 107
    6.1.2.2 Omkode med regelmessige uttrykk (regeek) ............................................. 109
    6.1.2.3 Automatiske varianter .............................................................................. 112
  6.2 Sla sammen data og ende verdier ....................................................................... 113
    6.2.1 Forprosesserings – tilrettelegge for sammenslåing ...................................... 114
      6.2.1.1 Aggregere data når observasjonsnivåene er ulike ................................. 114
      6.2.1.2 Nærerivariabler for observasjoner som hører sammen ....................... 115
    6.2.2 Data fra andre kilder ..................................................................................... 116
      6.2.2.1 Legge til enkeltevariabler (match) .......................................................... 117
      6.2.2.2 Sla sammen hele datasett (merge) .......................................................... 117
Oppgaver .......................................................................................................................... 119

DEL III
REGRESJONSANALYSER .............................................................................................. 121

KAPITTEL 7
ESTIMERE OG TOLKE EN LINEÆR MODELL (OLS) ..................................................... 123
7.1 Introduksjon til OLS .............................................................................................. 124
  7.1.1 Estimere en lineær modell .............................................................................. 125
  7.1.2 Lese modellsammendraget ............................................................................. 127
  7.2 Tolke resultatene .................................................................................................. 128
    7.2.1 Tolke marginaleffektene .............................................................................. 129
    7.2.2 Tolke med konfidensintervall ...................................................................... 130
    7.2.3 Tolke med predikerte verdier ...................................................................... 131
      7.2.3.1 Beregn predikerte verdier ..................................................................... 131
      7.2.3.2 Tegne predikerte verdier ..................................................................... 132
      7.2.3.3 Beregn nøyaktighet ............................................................................. 133
      7.2.3.4 Tegne usikkerhet .................................................................................. 134
      7.2.3.5 Legge til observerte verdier .................................................................. 135
Oppgaver .......................................................................................................................... 138
KAPITTEL 8
MULTIPPEL REGRESJON: FLERE VARIABLER
OG IKKE-LINEØRE EFFEKTER

8.1 Effekten av å kontrollere for andre variabler ................................................. 139
8.1.1 Estimer en multippel lineær modell .......................................................... 140
8.1.2 Tolke kontrollerte effekter i lineær modell ................................................ 143
8.1.3 Sammenlikne effekter ................................................................................. 144
8.1.3.1 Sammenlikning mellom modeller krever et likt utvalg ....................... 145
8.1.3.2 Sammenlikning i samme modell krever standardiserte variabler .... 145
8.1.4 Sammenlikne konfidensintervaller (<coefplots>) ...................................... 147

8.2 Ikke-lineære sammenhenger ............................................................................ 148
8.2.1 Samspillsledd ............................................................................................. 148
8.2.1.1 Estimere samspill ................................................................................ 149
8.2.1.2 Tolke et samspill .................................................................................. 150
8.2.2 Kurvilinearøre sammenhenger ................................................................. 155
8.2.2.1 Annengradsledd .................................................................................. 155
8.2.2.2 Eksponentielle sammenhenger .............................................................. 159

Oppgaver ............................................................................................................. 164

KAPITTEL 9
MODELLVURDERING ............................................................................................. 165

9.1 Hvor godt predikere modellen? ..................................................................... 166
9.1.1 Fordelingen til predikerte verdier ............................................................... 167
9.1.2 Korrelasjon mellom predikerte og observerte verdier ............................. 168
9.1.3 Prediksjon for undergrupper av data ......................................................... 169

9.2 Residualenes fordeling ................................................................................... 170
9.2.1 Er residualene normalfordelte? ................................................................. 170
9.2.1.1 Beregn residualene .............................................................................. 171
9.2.1.2 Kvantitative mål på skjævhet og kurtose ......................................... 172
9.2.1.3 Histogram med hjelpelinje .................................................................. 173
9.2.1.4 Kvantil – kvantilgrafikk ...................................................................... 174
9.2.1.5 Hva gjør vi hvis residualene ikke er normalfordelte? ......................... 174

9.2.2 Er residualene likt fordelt over avhengig variabel? ............................... 175

9.3 Uteliggere og innflytelsesrike observasjoner ................................................. 177
9.3.1 Innflytelse – hatverdier ............................................................................ 177
9.3.1.1 Beregn hatverdier ................................................................................ 178
9.3.1.2 Identifiser kritiske hatverdier ............................................................. 179
9.3.2 Uteliggere – t-standardiserte residualer .................................................... 182
9.3.2.1 Estimer t-standardiserte residualer ...................................................... 183
9.3.2.2 Identifiser kritiske verdier ................................................................. 184
9.3.3 Uteliggere med innflytelse – Cook’s D ....................................................... 185
9.3.3.1 Beregn Cook’s D .................................................................................. 186
9.3.3.2 Identifiser kritiske verdier ................................................................. 186
9.3.4 Hva gjør vi med innflytelsesrike uteliggere? .............................................. 190
9.4 Kolinearitet ................................................................................................................. 192
  9.4.1 Tegn på kolinearitet .......................................................................................... 192
  9.4.2 Løsninger på kolinearitet .................................................................................. 193
Oppgaver ............................................................................................................................. 194

KAPITTEL 10
BINOMISK LOGISTISK REGRESJON .................................................................................. 195
10.1 Estimere en binomisk logistisk modell ..................................................................... 196
10.2 Tolke en binomisk logistisk modell ......................................................................... 198
  10.2.1 Tolke marginaleffektene ................................................................................... 199
    10.2.1.1 Retning og signifikans (endring i logodds) ................................................. 199
    10.2.1.2 Relativ endring (endring i odds, oddsratio) ............................................... 199
  10.2.2 Predikerte sannsynligheter .............................................................................. 201
    10.2.2.1 Regne ut predikerte sannsynligheter ......................................................... 202
    10.2.2.2 Regne ut konfidenintervall til predikerte verdier ..................................... 203
    10.2.2.3 Tegne predikerte verdier .......................................................................... 204
10.3 Vurdere en binomisk logistisk modell ..................................................................... 210
  10.3.1 Hvor ofte predikerer modellen riktig? ............................................................... 210
    10.3.1.1 Definere kuttspunkt og predikere utfall ..................................................... 210
    10.3.1.2 Sammenlikne predikerte og observerte utfall ......................................... 211
  10.3.2 Avansert modellstatistikks .............................................................................. 214
    10.3.2.1 ROC-kurve ................................................................................................. 214
    10.3.2.2 Hosmer-Lemeshows «Goodness-of-fit» ..................................................... 215
Oppgaver ............................................................................................................................. 216

BIBLIOGRAFI ....................................................................................................................... 217

REGISTER FOR R-KOMMANDOER ..................................................................................... 219