
Indhold

Tabeller	9
Figurer	10
Forord	13
I Kom i gang!	15
1 Hvad er R?	17
1.1 Et statistikprogram	17
Installering af programmet	18
Installering af R	18
Installering og opdatering af RStudio	18
1.2 Et programmeringssprog	19
2 R som regnemaskine	23
2.1 Når RStudio åbnes første gang	24
2.2 Datatyper i R	27
Vektorer	28
Indeksring	29
Funktioner og argumenter	29
Sortering af vektorer	31
Matricer	32
Oprettelse af en matrice	33
Oprettelse af et datasæt	34
Indeksring	35
Måleniveau	37
Lister	40
Oprettelse af en liste	40
Indeksring	40
3 R som tegneprogram	45
3.1 Placering af punkter i et koordinatsystem	45
3.2 Oprettelse af tegneark og tilføjelse af information	46
Oprettelse og tilrettelæggelse af tegnearket	46

Tilføjelse af grafik på eksisterende ark	49
Tilføjelse af punkter, linjer og tekst	49
Specificering af farve og størrelse	50
Angivelse af tegnforklaringer	52
4 Ind og ud af R	55
4.1 Arbejdsprocessen	56
Gemme arbejdet	56
Organisering af kodefiler	56
Gemme kodefiler	57
Organiser arbejdet i mapper	57
Brug af andres arbejde: R-pakker	58
4.2 Import/eksport af datafiler	61
R på hjemmebane (.rda):	61
R på tværs af programmer: data i tekstformat	62
Udskrivning af data fra R (.txt og .csv)	62
Åbn data i Excel/Open Office	64
Åbn data i R	64
Import fra tekstfiler (.txt og .csv)	65
Import fra andre filformater	66
 II Undersøgelse og tilrettelæggelse af data	 69
5 Beskrivende statistik	71
5.1 Datastrukturen	72
5.2 Univariate fordelinger	73
Kategoriske variabler	75
Frekvenstabel	75
Lagkage- og søjlediagrammer	76
Kontinuerte variabler	78
Numeriske beskrivelser	78
Grafiske alternativer	79
5.3 Bivariate sammenhænge	82
To kategoriske variabler	82
Bivariat sammenhæng med tal	82
Bivariat sammenhæng i tegning	85
En kategorisk og en kontinuert variabel	87
To kontinuerte variabler	91
Numeriske beskrivelser	91
Grafiske alternativer	93
 6 Tilrettelæggelse af data	 99
6.1 Omkodning af eksisterende variabler	99
Kontinuerte variabler	100
Additive indekser	100
Omvending af skalarretningen	103
Ændring af skalaenheder	104

Kategoriske variabler (betinget omkodning)	104
Omkodning ved hjælp af indeksering	105
Omkodning med regelmæssige udtryk ("regex")	107
Automatiske varianter	110
6.2 Fletning af data og ændring af værdier	111
Forprocessering - tilrettelæggelse af fletning	112
Aggregering af data, når observationsniveauerne er forskellige	112
Nøglevariabler for observationer, som hører sammen	113
Data fra andre kilder	114
Tilføjelse af enkeltvariabler ("match")	115
Fletning af hele datasæt ("merge")	116
 III Regressionsanalyser	 119
7 Estimere og tolke OLS	121
7.1 Introduktion til OLS	122
Estimering af en lineær model	122
Læsning af modelsammenfatningen	123
7.2 Fortolkning af resultaterne	126
Fortolkning af marginaleffekterne	126
Fortolkning med konfidensinterval	127
Fortolkning med prædikterede værdier	128
Beregning af prædikterede værdier	128
Tegning af prædikterede værdier	129
Beregning af nøjagtighed	130
Tegning af usikkerhed	131
Tilføjelse af observerede værdier	131
 8 Multipel regression	 137
8.1 Kontrollere for andre variabler	138
Estimering af en multipel lineær model	139
Fortolkning af kontrollerede effekter i lineær model	140
Sammenligning af effekter	142
Sammenligning mellem modeller kræver et ensartet udvalg	143
Sammenligning i samme model kræver standardiserede vari-	
abler	143
Sammenligning af konfidensintervaller ("coefplot")	144
8.2 Ikke-lineære sammenhænge	145
Samspilsled	146
Estimering af samspil	147
Fortolkning af et samspil	148
Kurvilineære sammenhænge	153
Andengradsled	154
Ekspontielle sammenhænge	158
 9 Modelvurdering	 163
9.1 Hvor godt prædikterer modellen?	164

Fordelingen af prædikterede værdier	165
Korrelation mellem prædikterede og observerede værdier	165
Prædiktion for undergrupper af data	167
9.2 Residualernes fordeling	168
Er residualerne normalfordelte?	169
Beregning af residualerne	169
Kvantitative mål på skævhed og kurtose	170
Histogram med hjælpelinje	171
Kvantil-kvantilgrafikk	171
Hvad gør vi, hvis residualerne ikke er normalfordelte?	173
Er residualerne ligeligt fordelt i den afhængige variabel?	173
9.3 Afvigere og observationer med stor indflydelse	175
Indflydelse - hatværdier	177
Beregning af hatværdier	177
Identifikation af kritiske hatværdier	178
Afgivere - t-standardiserede residualer	181
Estimering af t-standardiserede residualer.	181
Identifikation af kritiske værdier	184
Afgivere med indflydelse - Cook's D	186
Beregning af Cook's D	186
Identifikation af kritiske værdier	186
Hvad gør vi med indflydelsesrige afgivere?	190
9.4 Kolinearitet	191
Tegn på kolinearitet	191
Løsninger på kolinearitet	193
10 Binomisk logistisk regression	195
10.1 Estimering af en binomisk logistisk model	197
10.2 Fortolkning af en binomisk logistisk model	198
Fortolkning af marginaleffekterne	199
Retning og signifikans (ændring af logodds)	199
Relativ ændring (ændring i odds, oddsratio)	199
Prædikterede sandsynligheder	201
Udregning af prædikterede sandsynligheder	201
Udregning af konfidensintervallet for prædikterede værdier	202
Tegning af prædikterede værdier	203
10.3 Vurdering af en binomisk logistisk model	209
Hvor ofte prædikterer modellen rigtigt?	209
Definition af skæringspunkt og prædiktion af udfald	209
Sammenligning af prædikterede og observerede udfald	210
Avanceret modelstatistik	212
ROC-kurve	212
Hosmer-Lemeshows Goodness-of-fit	213
Register for R-kommandoer	217
Litteratur	221