

Innhold

Del I Innledning

1 Hva er statistikk? 19

- 1.1 Bokas innhold 20
 - 1.1.1 Noen eksempler 20
 - 1.1.2 Historie 23
 - 1.1.3 Bokas oppbygning 25
- 1.2 Noen viktige begreper 26
 - 1.2.1 Populasjon og utvalg 26
 - 1.2.2 Variasjon 28
 - 1.2.3 Variabeltyper 29
- 1.3 Modeller 30
- 1.4 Innsamling av data 32
 - 1.4.1 Problemløsning krever data 32
 - 1.4.2 Generelt om datainnsamling 33
 - 1.4.3 Spørreundersøkelser 34
- 1.5 Oppsummering 36

2 Hva er fakta om utvalget? 37

- 2.1 Beskrivelse av kategoriske data 38
- 2.2 Beskrivelse av diskrete data 41
- 2.3 Beskrivelse av kontinuerlige data 42
 - 2.3.1 Frekvenstabell 42
 - 2.3.2 Histogram 44
 - 2.3.3 Stamme- og bladdiagram 46
 - 2.3.4 Spredningsplott 46
 - 2.3.5 Tidsserieplott 47
 - 2.3.6 Generelle diagramtips 48
 - 2.3.7 Formen på histogrammet 49

2.4	Sentralmål	50
2.4.1	Modus	50
2.4.2	Median	51
2.4.3	Gjennomsnitt	52
2.4.4	Sammenligning av modus, median og gjennomsnitt	55
2.5	Spredningsmål	56
2.5.1	Variasjonsbredde	56
2.5.2	Varians og standardavvik	57
2.5.3	Variasjonskoeffisient	59
2.6	Skjevhet og kurtosis*	60
2.7	Grupperte data*	60
2.8	Statistikk med Excel	62
2.8.1	Legge inn data	62
2.8.2	Vanlige diagrammer	63
2.8.3	Frekvenstabeller og histogrammer	64
2.8.4	Beskrivende statistikk	65
2.9	Statistikk med Minitab	66
2.9.1	Skjermbildet	66
2.9.2	Grafikk og statistikk	68
2.10	Statistikk med SPSS	69
2.11	Statistikk med R	70
2.11.1	Skjermbildet	70
2.11.2	Grafikk og statistikk	71
2.12	Oppsummering	72
	Oppgaver til kapittel 2	72

Del II

Sannsynlighetsteori

3

Sannsynlighetsregning 79

3.1	Utfallsrom og hendelse	80
3.2	Sannsynligheten for en hendelse	83
3.2.1	Uniform sannsynlighetsmodell	84
3.2.2	Sannsynlighet basert på relativ frekvens	85
3.2.3	Subjektiv sannsynlighet for en hendelse	89
3.2.4	Eksempel på tallfesting av sannsynligheter	90
3.3	Innføring i sannsynlighetsregning	92
3.3.1	Litt mengdelære	92
3.3.2	Sannsynlighetsregningens aksiomer	94
3.3.3	De viktigste regnereglene	94
3.4	Betingede sannsynligheter	98
3.4.1	Regler for betingede sannsynligheter	100
3.4.2	Litt om Bayes	103
3.4.3	Hendelsestre	106
3.5	Uavhengige hendelser	109

3.5.1	Er hendelsene uavhengige?	110
3.5.2	Regler for uavhengige hendelser	111
3.6	Telleregler – Kombinatorikk	114
3.6.1	Produktregelen	114
3.6.2	Urnemodellen	115
3.6.3	De viktigste tellereglene	116
3.6.4	Sannsynligheter basert på kombinatorikk	121
3.7	Bruk av datamaskin	123
3.8	Oppsummering	123
	Oppgaver til kapittel 3	124

4 Stokastiske variabler133

4.1	Stokastiske variabler	134
4.2	Diskrete sannsynlighetsmodeller	136
4.2.1	Sannsynlighetsfordeling	136
4.2.2	Forventningsverdi	143
4.2.3	Varians og standardavvik	147
4.3	Kontinuerlige sannsynlighetsmodeller	149
4.4	Flere variabler samtidig	155
4.4.1	Simultanfordeling	156
4.4.2	Kovarians og korrelasjon	159
4.4.3	Uavhengighet	162
4.4.4	Hendelsestre	164
4.5	Oppsummering	166
	Oppgaver til kapittel 4	167

5 Vanlige sannsynlighetsmodeller173

5.1	Innledning	174
5.2	Binomisk modell	176
5.3	Hypergeometrisk fordeling	181
5.4	Geometrisk fordeling	184
5.5	Poissonfordelingen	185
5.6	Ekspontialfordelingen	190
5.7	Normalfordelingen	193
5.7.1	Normalfordelingen	194
5.7.2	Sjekk av normalitet	200
5.8	Sentralgrenseteoremet	202
5.9	Andre modeller*	207
5.9.1	Kjikkvadratfordelingen	208
5.9.2	Student t -fordelingen	209
5.9.3	Fisher F -fordelingen	211
5.9.4	Multinomisk fordeling	212
5.9.5	Binormalfordelingen	214

- 5.10 Bruk av datamaskinen 215
 - 5.10.1 Excel 215
 - 5.10.2 Minitab 216
 - 5.10.3 SPSS 216
 - 5.10.4 R 217
- 5.11 Oppsummering 217
 - Oppgaver til kapittel 5 218

Del III Grunnleggende statistikk

6 Estimering og hypotesetesting 227

- 6.1 Et innledningseksempel 228
 - 6.1.1 Hvilken vraksannsynlighet har det nye utstyret? 229
 - 6.1.2 Hvor sikkert er estimatet? 230
 - 6.1.3 Er det nye utstyret bedre enn det gamle? 231
 - 6.1.4 Bør AluProd kjøpe det nye utstyret? 234
- 6.2 Punktestimering 234
 - 6.2.1 Basisteori 234
 - 6.2.2 Punktestimat for forventningsverdien μ 237
 - 6.2.3 Punktestimat for standardavviket σ 240
 - 6.2.4 Punktestimat for sannsynligheten p 241
 - 6.2.5 Punktestimat for raten λ 242
 - 6.2.6 Stratifisering* 243
- 6.3 Konfidensintervaller 244
 - 6.3.1 Teori 245
 - 6.3.2 Konfidensintervall for μ når σ er kjent (Z -intervall) 247
 - 6.3.3 Konfidensintervall for μ når σ er ukjent (T -intervall) 248
 - 6.3.4 Konfidensintervall for σ 251
 - 6.3.5 Konfidensintervall for sannsynligheten p 252
 - 6.3.6 Konfidensintervall for raten λ 254
- 6.4 Hypotesetesting – metode og teori 255
 - 6.4.1 Etablering av modell og formulering av hypotesene 255
 - 6.4.2 Testobservator og forkastingsområdets form 257
 - 6.4.3 Valg av akseptable feilsannsynligheter 259
 - 6.4.4 Spesifikasjon av forkastingsområde 260
 - 6.4.5 Styrkefunksjon 262
 - 6.4.6 Nødvendig utvalgsstørrelse 265
 - 6.4.7 Testens konklusjon 266
 - 6.4.8 Signifikanssannsynlighet (p -verdi) 268
 - 6.4.9 Sammenheng med konfidensintervaller 269
- 6.5 Hypotesetesting i praksis 269
 - 6.5.1 Hypotesetest av μ når σ er kjent (Z -test) 269
 - 6.5.2 Hypotesetest av μ når σ er ukjent (T -test) 271
 - 6.5.3 Hypotesetest av sannsynligheten p 273

- 6.6 Empirisk fordelingsfunksjon* 275
- 6.7 Bruk av datamaskinen 277
 - 6.7.1 Excel 277
 - 6.7.2 Minitab 277
 - 6.7.3 SPSS 277
 - 6.7.4 R 278
- 6.8 Oppsummering 278
 - Oppgaver til kapittel 6 278

Del IV Statistisk analyse

7 Analyse av sammenhenger 287

- 7.1 Innledning 288
- 7.2 Korrelasjon 290
- 7.3 Enkel lineær regresjon 293
 - 7.3.1 Regresjonsmodellen 293
 - 7.3.2 Regresjonslinjen 294
 - 7.3.3 Hvor god er modellen? 297
 - 7.3.4 Noen statistiske egenskaper ved regresjonsmodeller 298
 - 7.3.5 Er det en sammenheng mellom x og y ? 299
 - 7.3.6 Hvilken respons kan vi forvente? 301
 - 7.3.7 Hvilke observasjoner er sannsynlige? 302
 - 7.3.8 Kontroll av forutsetningene 303
 - 7.3.9 Vektet regresjon* 305
- 7.4 Multippel lineær regresjon 306
- 7.5 Ikke-lineær regresjon 310
 - 7.5.1 Vanlig ikke-lineær regresjon 311
 - 7.5.2 Polynomisk regresjon 313
- 7.6 Tidsrekkeanalyse – prognoser 315
 - 7.6.1 Trend 317
 - 7.6.2 Sesongvariasjon 318
 - 7.6.3 Tilfeldig variasjon 319
 - 7.6.4 Prognose 319
 - 7.6.5 Mer avanserte modeller* 320
- 7.7 Bruk av datamaskinen 322
 - 7.7.1 Excel 322
 - 7.7.2 Minitab 323
 - 7.7.3 SPSS 324
 - 7.7.4 R 325
- 7.8 Oppsummering 325
 - Oppgaver til kapittel 7 326

8	Sammenligning av grupper.....	333
8.1	Innledning	334
8.1.1	Hvor mange grupper og hvor mange observasjoner?	334
8.1.2	Randomisering	334
8.1.3	Paring og blokkdeling	335
8.1.4	Klyngeutvalg	337
8.1.5	Hva slags data har vi?	337
8.2	Litt sammenligningsfilosofi	338
8.3	<i>T</i> -test for to grupper	340
8.3.1	Uparet <i>T</i> -test	340
8.3.2	Paret <i>T</i> -test	344
8.3.3	Vurdering av paret og uparet test*	346
8.4	Variansanalyse for flere grupper	346
8.4.1	Noen illustrasjoner	346
8.4.2	Enveis variansanalyse	348
8.4.3	Samtidige konfidensintervaller	351
8.4.4	Kontroll av forutsetninger	351
8.4.5	Toveis variansanalyse*	352
8.4.6	Litt forsøksplanlegging*	356
8.5	Ikke-parametriske tester	357
8.5.1	Uparet Mann-Whitney-Wilcoxon-test	357
8.5.2	Paret Wilcoxon-test	360
8.5.3	Paret fortegnstest	362
8.5.4	Kruskal-Wallis' test av flere grupper*	363
8.6	Analyse av kategoriske krysstabeller	364
8.6.1	Kjikkvadrattest	365
8.6.2	Simpsons paradoks	367
8.6.3	Modelltest	368
8.7	Bruk av datamaskinen	370
8.7.1	Excel	370
8.7.2	Minitab	371
8.7.3	SPSS	372
8.7.4	R	373
8.8	Oppsummering	374
	Oppgaver til kapittel 8	374

9 Spørreundersøkelse og datainnsamling..... 381

- 9.1 Innledning 382
- 9.2 Utforming av spørreskjema 383
- 9.3 Opplegg for innsamling 385
- 9.4 Organisering av data 389
- 9.5 Analyse av data og presentasjon av resultater 389
- 9.6 Praktisk datainnsamling og analyse 393
 - 9.6.1 Spørreundersøkelse om arealbruk 393
 - 9.6.2 Analyse av leiepriser i boligmarkedet 395
- 9.7 Fallgruver 399
- 9.8 Oppsummering 400
 - Oppgaver til kapittel 9 401

10 Big Data og digitalisering..... 403

- 10.1 Innledning 404
 - 10.1.1 Dataforskeren 404
 - 10.1.2 Litt om maskinlæring 405
 - 10.1.3 Maskinlæring som konkurransesport 410
- 10.2 Organisering av data 411
 - 10.2.1 Behandling av store datamengder 411
 - 10.2.2 Tilrettelegging for analyse 413
- 10.3 Utforskning og visualisering av data 414
- 10.4 Maskinlæring og dataanalyse 419
 - 10.4.1 Regresjon 419
 - 10.4.2 Klassifikasjon 423
 - 10.4.3 Gruppering (Clustering) 426
- 10.5 Fallgruver 428
- 10.6 Oppsummering 430
 - Oppgaver til kapittel 10 431

Del V Andre anvendelser

11	Statistisk kvalitetsstyring	435
11.1	Kontrolldiagrammer 436	
11.1.1	Shewharts \bar{x} -diagram 439	
11.1.2	s -diagram for spredning 445	
11.1.3	p -diagram for defektandel 447	
11.1.4	Andre kontrolldiagrammer 448	
11.2	Kapabilitet 449	
11.2.1	Toleransegrenser og kapabilitetsindeks 449	
11.2.2	Fortolkning av kapabilitetsindeks 451	
11.3	Akseptanskontroll 453	
11.3.1	Kontrollens karakteristikkk 454	
11.3.2	Akseptanskontroll basert på defektandel 456	
11.3.3	Akseptanskontroll basert på målinger 457	
11.4	Bruk av datamaskinen 460	
11.4.1	Excel 460	
11.4.2	Minitab 460	
11.4.3	SPSS 461	
11.4.4	R 461	
11.5	Oppsummering 461	
	Oppgaver til kapittel 11 462	
12	Simulering	465
12.1	Simulering av tilfeldige tall 466	
12.2	Simulering av en stokastisk variabel 468	
12.3	Simulering med datamaskinen 471	
12.3.1	Excel 471	
12.3.2	Minitab 471	
12.3.3	SPSS 472	
12.3.4	R 472	
12.3.5	Noen eksempler 473	
12.4	Simulering av flere stokastiske variabler 475	
12.4.1	Uavhengige variabler 476	
12.4.2	Avhengige variabler 477	
12.5	Simulering av funksjoner 479	
12.6	Simulering av integral* 483	
12.7	Oppsummering 485	
	Oppgaver til kapittel 12 486	

A	Eksamensoppgaver	493
B	Ekstra bevis	525
	B.1 Bevis til kapittel 3	525
	B.2 Bevis til kapittel 4	526
	B.3 Bevis til kapittel 5	528
	B.4 Bevis til kapittel 6	533
	B.5 Bevis til kapittel 7	534
	B.6 Bevis til kapittel 8	536
	B.7 Bevis til kapittel 12	537
C	Standard statistikkfunksjoner i Excel.....	538
	C.1 Standard statistikkfunksjoner i Excel	538
	C.2 Statistikkfunksjoner under dataanalyseverktøyet i Excel	542
D	Fasit	545
E	Statistiske tabeller.....	574
	E.1 Kumulativ binomisk sannsynlighet	575
	E.2 Kumulativ poissonfordeling	576
	E.3 Kumulativ standardnormalfordeling	577
	E.4 Standardnormalfordelingens kvantiltabell	578
	E.5 t -fordelingens kvantiltabell	579
	E.6 Kvikvadratfordelingens kvantiltabell	580
F	Stikkord	581