

ANALYSE AF
SPØRGESKEMADATA
MED SPSS

Jan Møller Jensen & Thorbjørn Knudsen

ANALYSE AF
SPØRGESKEMADATA
MED SPSS

TEORI, ANVENDELSE OG PRAKSIS

Syddansk Universitetsforlag 2014

© forfatterne og Syddansk Universitetsforlag
Trykt af Narayana Press, Gylling
Omslagsdesign af Anne Charlotte Mouret, UniSats
ISBN 978-87-7674-796-1

3. udgave, 1. oplag 2014

Forsidebillede: Corbis
Louis K. Meisel Gallery / Charles Bell

Mekanisk, fotografisk, elektronisk eller anden mangfoldiggørelse
af denne bog er kun tilladt med forlagets tilladelse eller ifølge
overenskomst med Copy-Dan.



Syddansk Universitetsforlag
Campusvej 55 . 5230 Odense M
Tlf. 66 15 79 99 . Fax 66 15 81 26
press@forlag.sdu.dk
www.universitypress.dk

Forord

Denne bog handler om analyse af spørgeskemadata med SPSS. Som undervisere og vejledere for studerende og virksomheder, der anvender markedsanalyser, er det vores erfaring, at man med kendskab til nogle få grundlæggende statistiske begreber og sammenhænge ret hurtigt bliver fortrolig med analyseteknikkerne. Især hvis man får tilbudt hjælp til selvhjælp med at anvende statistikprogrammet SPSS. Ud fra denne logik har vi udviklet en indføring i SPSS, der hjælper den uindviede med at komme godt i gang – og den øvede med at blive endnu bedre.

Vores indføring i SPSS er baseret på arbejdet med spørgeskemadata i praksis. Derfor har vores undervisning i kvantitativ markedsanalyse altid taget udgangspunkt i en praktisk problemstilling. De studerende har arbejdet med data, der er indsamlet for at løse denne problemstilling, og med denne bog indøvet færdigheder, der har gjort dem fortrolige med selv de mest avancerede analyseteknikker.

I bogen guider vi dig trin for trin igennem en lang række af de analyser, som kan udføres med SPSS. Alle analyser tager udgangspunkt i et praktisk eksempel, hvor vi anvender datafilen *E-turisme*. Denne fil kan downloades fra www.virksomhedsoekonomi.dk. Hvis du har adgang til SPSS, kan du sidde foran din computer og læse bogen, mens du efterprøver analyseeksemplerne.

Bogen henvender sig først og fremmest til studerende, som ønsker at indlære færdigheder i analyse af spørgeskemadata, men den vil sikkert også have interesse for andre, der arbejder med analyse af spørgeskemadata. Det er ingen forudsætning, men bestemt en fordel, hvis læseren på forhånd har erhvervet sig et vist kendskab statistik og dataanalyse. Skulle du føle behov for at opfriske dine statistikkundskaber, er der forslag til supplerende litteratur bagerst i bogen.

I denne tredje reviderede udgave af bogen er der foretaget en række rettelser og tilpasninger. De anvendte illustrationer fra SPSS er opdateret til version 21.0 og alle analyseeksempler gennemføres på en ny datafil, der er mere tidssvarende og indeholder flere besvarelser, end datafilen i bogens tidligere udgaver. Dataindsamlingen er foretaget som en online spørgeskemaundersøgelse i samarbejde med analysevirksomheden Wilke. Vi vil i den forbindelse gerne takke Wilke for, at vi må anvende datafilen til eksemplerne i denne bog.

Bogens opbygning og emner er uændret i forhold til tidligere udgaver, men i mange af kapitlerne, er der foretaget strukturændringer og tilføjelser. Af væsentlige tilføjelser kan nævnes en udvidet beskrivelse af de enkelte analyseteknikker, en uddybende diskussion af forudsætningerne bag de enkelte analyseteknikker samt en række gode råd til, hvad du skal være opmærksom på, når du fortolker analyseresultaterne.

Bogens tilblivelse er, som nævnt ovenfor, sket med udgangspunkt i vores mangeårige undervisningserfaring i faget kvantitativ markedsanalyse og i høj grad baseret på det undervisningsmateriale, som vi løbende har udleveret til de studerende. Tak til kollegaer og studerende, som har læst tidligere udgaver af bogen og er kommet med mange værdifulde kommentarer. Også tak til SDU-studerende Luma Mosa og Charlotte Fich for gennemlæsning og kommentarer til manuskriptet til denne nye udgave. En særlig til vores kollega Judy Hermansen, som har stået for den endelige korrektur og til Lene Holbæk for arbejdet med bogens layout og tekst. Endelig skal Syddansk Universitetsforlag og Martin Westergaard i særdeleshed have stor tak for godt samarbejde og gode råd i forbindelse med udgivelsen.

December 2013

Jan Møller Jensen

Thorbjørn Knudsen

<i>Indholdsfortegnelse</i>	7
Forord	5
Indholdsfortegnelse	7
1 Indledning	8
DEL I	
Forberedelser	13
2 Datas egenskaber og kvalitet	14
3 Hypotesetest og inferens	31
4 Analyseplanen	41
DEL 2	
Datafilen	47
5 Opstilling af datafil	48
6 Datatransformationer	66
DEL 3	
Beskrivende analyser	73
7 Frekvenstabeller	74
8 Krydstabeller og <i>chi-square</i> test	84
9 Ikke-parametriske tests	100
DEL 4	
Middelværdianalyse	113
10 T-test	114
11 ANOVA	133
DEL 5	
Analyse af lineære sammenhænge	149
12 Korrelationsanalyse	150
13 Regression	164
14 Multipel regression	184
15 Diskriminantanalyse	204
DEL 6	
Strukturanalyser	223
16 Faktoranalyse	224
17 Reliabilitetsanalyse på sammensat skala	251
18 Klyngeanalyse	260
Appendix	280

Kapitel 1

Indledning

Denne bog handler om, hvordan man analyserer spørgeskemadata med SPSS. Bogen er skrevet med henblik på at gøre læseren kompetent til at anvende de mange muligheder, SPSS giver for at analysere spørgeskemadata. Som bogens undertitel antyder, omfatter en sådan kompetence både *teori*, *anvendelse* og *praksis*. Bogen er derfor skrevet i en form, der integrerer teori og anvendelse med udgangspunkt i praktiske eksempler på spørgeskemadata. Det gør den særdeles velegnet til selvstudium.

Bogens indhold og opbygning

Bogen indeholder, udover nærværende kapitel, 17 kapitler, fordelt over seks hovedafsnit.

Del I, som indeholder kapitel 2, 3 og 4, omhandler de mest grundlæggende forberedelser, der bør gå forud for enhver dataanalyse. Kapitel 2 belyser implikationer af datas egenskaber og kvalitet. Forskellige dataanalyser stiller forskellige krav til dine datas egenskaber, og nogle af analyserne er meget følsomme over for afvigelser i dataenes kvalitet. Det er derfor vigtigt, at du sætter dig ind i kapitel 2, inden du starter på dataanalyserne. Kapitel 3 præsenterer det mest elementære begrebsapparat og terminologi, som anvendes i forbindelse med hypotesetest og inferens, herunder de forudsætninger, som hypotesetest og inferens hviler på. Emnet udgør omdrejningspunktet for alle de statistiske analyser, der gennemgås i bogen. Kapitel 4 viser, hvordan du kan opstille en analyseplan forud for dataanalyserne. Analyseplanen er en slags 'køreplan', der sikrer, at dataanalyserne fører til, at de spørgsmål, du ønsker at belyse, også bliver besvaret på en korrekt måde. Den er derfor en vital del af forberedelserne.

Del II fokuserer på datafilen. I kapitel 5 lærer du at opstille en datafil, og vi viser dig, hvordan datafilen renses for fejl. Sidst i kapitlet viser vi dig,

hvordan du kan udvælge eller opdele respondenterne i datafilen forud for dataanalyserne. Kapitel 6 omhandler datatransformationer, dvs. hvordan du kan ændre eksisterende variable eller generere nye variable ud fra dem, der er i datafilen.

Del III indleder gennemgangen af analyserne. Denne del indeholder de mest simple analysetyper. Fælles for analyserne i dette afsnit, er, at de ikke stiller krav til variabelens skalaniveau. Kapitel 7 og 8 omhandler frekvens- og krydstabulering samt logikken og fremgangsmåden ved *chi-square* test. Kapitel 9 præsenterer test, der kan anvendes på ikke-metriske data, og gennemgår repræsentativitetstest som et vigtigt eksempel.

Del IV vedrører analyse af middelværdier. Denne del indeholder den første af mange analyser, der stiller krav til variabelens skalaniveau, og samtidig hviler på en række andre forudsætninger om variabelens egenskaber. Kapitel 10 gennemgår t-testen, som anvendes til analyse af forskelle mellem to middelværdier. Kapitel 11 omhandler ANOVA, der anvendes til at teste, om tre eller flere middelværdier er forskellige.

Del V fokuserer på test af lineære sammenhænge mellem to eller flere variable. Kapitel 12 indeholder den mest simple af disse analyser, korrelationsanalysen. Korrelationsanalyse tester styrken og retningen af sammenhængen mellem to variable, uden dog at antage nogen form for kausalitet mellem variablene. Kapitel 13 gennemgår regressionsanalyse på to variable, også kaldet bivariat eller simpel regression. Analysen minder meget om korrelationsanalysen, men bygger på en antagelse om, at den ene af variablene er årsag til ændringerne i den anden variabel. Sidst i kapitlet viser vi dig, hvordan du kan udvide regressionsanalysen med mediator- og moderator-effekter. Kapitel 14 omhandler multipel regressionsanalyse. Den væsentligste forskel i forhold til simpel regression er, at analysen inddrager to eller flere mulige uafhængige variable. Den sidste af analyserne i Del V, er diskriminantanalysen. Den adskiller sig fra regressionsanalyse på flere punkter. Bl.a. kan diskriminantanalysen anvendes til klassifikation af respondenter. Når vi alligevel har taget den med i Del V, er det fordi diskriminantanalyse, ligesom regression, ofte finder anvendelse som forklaringsmodel. Diskriminantanalysen kan anvendes i stedet for regressionsanalysen, hvis den variabel, der skal forklares, ikke opfylder sidstnævntes krav om intervallskalaniveau.

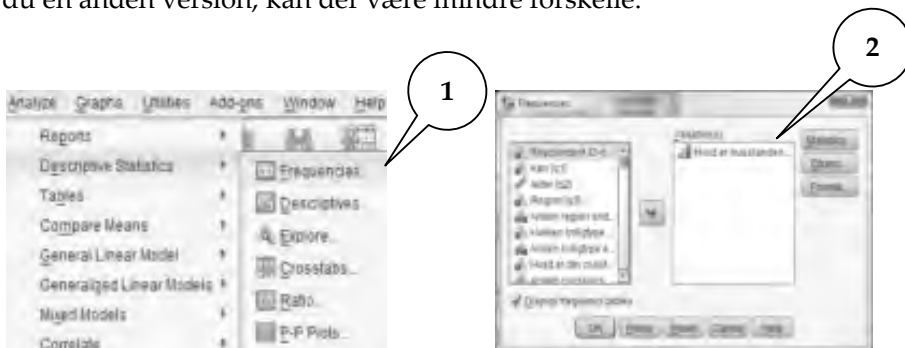
Del VI fokuserer på de såkaldte strukturanalyser, der, som navnet antyder, forsøger at afdække strukturer i datamaterialet. Kapitel 16 omhandler faktoranalysen, som anvendes til at reducere eller forenkle datamaterialet. Kapitel 17 præsenterer reliabilitetstest på sammensat skala. Kapitel 18 omhandler

klyngeanalysen, der, ligesom faktoranalysen, forsøger at skabe struktur og forenkling i datamaterialet. Men hvor faktoranalysen grupperer variablene i datasættet, forsøger klyngeanalysen at gruppere respondenterne.

Kapitlernes opbygning

Vi har skrevet bogen i en form, der er velegnet til selvstudium. Alle kapitlerne, der gennemgår analyser, er derfor bygget op efter samme skabelon. Kapitlet indledes med en kort beskrivelse af den enkelte analyse. Derefter følger et afsnit, *Hvad bør jeg vide, før jeg går i gang?*, der gennemgår de forudsætninger, der skal være opfyldt for at gennemføre analysen, og giver gode råd til hvordan du skal fortolke dine analyser. Afsnittet indeholder desuden en liste med forklaring af vigtige begreber og statistikker samt en standard-procedure for gennemførelse af analyseprocessen.

Herefter følger afsnittene *Hvordan gør jeg?* og *Hvordan fortolker jeg resultaterne?* Trin for trin guider vi dig igennem de vigtigste analyser, som kan udføres med SPSS. Hvert trin er udførligt beskrevet og bliver yderligere illustreret med tilhørende henvisninger til skærmbilleder fra SPSS. Nedenfor er vist eksempler på, hvordan hvert trin i analysen og fortolkningsprocessen illustreres med henholdsvis hvide og grå 'talbobler' i de tilhørende skærmbilleder. Du skal være opmærksom på, at skærmbilledene er taget fra version 21.0 af SPSS. Har du en anden version, kan der være mindre forskelle.



1 *q8 Hvad er husstandens årlige bruttoindkomst?*

N	Valid	401
	Missing	76
Mean		7,66
Std. Error of Mean		,136
Median		8,00
Mode		9
Std. Deviation		2,726
Variance		7,430

2

Alle analyserne tager udgangspunkt i et praktisk eksempel, hvor vi anvender data fra datafilen *E-turisme*. Hele datafilen kan downloades fra www.virksomhedsoekonomi.dk. Har du adgang til SPSS, kan du således sidde foran din computer og læse bogen, mens du efterprøver eksemplerne.

Bogens case: E-turisme

Langt størstedelen af verdens turister har i dag adgang til digitale platforme og en voksende andel af turisterne anvender derfor også internettet i forbindelse med planlægning, gennemførelse og opfølgning på deres rejse. Men hvor mange danske ferierejsende anvender internettet før, under og efter deres ferierejser, og hvad bruger de internettet til? Viden herom har stor værdi for danske rejsebureauer og andre aktører i rejsebranchen, som ønsker at tiltrække og servicere danske rejsende. Med det formål for øje gennemførte en af bogens forfattere i efteråret 2010 en spørgeskemaundersøgelse, som blandt andet søgte svar på nedenstående spørgsmål.

1. Hvor udbredt er anvendelsen af internettet i forbindelse med rejseoplevelser - før, under og efter rejsen?
2. Hvad anvendes internettet til før, under og efter rejsen?
3. Hvem anvender internettet i forbindelse med rejseoplevelser?
4. Har turistens rejsemotiver indflydelse på, hvor vigtigt det er for turisten at komme på internettet før, under og efter rejsen?
5. Hvor vigtigt er det for turisterne, at de kan komme på internettet før, under og efter deres rejse?
6. Er det muligt at opdele turisterne i et antal segmenter, som adskiller sig i forhold til, hvor vigtigt det er for dem at komme på internettet før, under og efter rejsen?

Den del af spørgeskemaet, som blev designet til besvarelse af ovennævnte undersøgelsesspørgsmål, er vist i Appendix. Dataindsamlingen er foretaget som en online spørgeskemaundersøgelse (CAWI = *computer assisted web-interviews*) i samarbejde med analysevirksomheden Wilke. Invitation til deltagelse i undersøgelsen blev i november 2010 udsendt til et repræsentativt udsnit af Wilke's web-panel (læs mere herom på www.wilke.dk). I alt 500 respondenter deltog i undersøgelsen. 23 af respondenterne havde ikke været på ferierejse i perioden 2008-2010 og har derfor ikke kunnet svare på spørgeskemaets spørgsmål vedrørende denne periode (jf. spørgsmål 16-20 i Appendix). Datafilen, *E-turisme*, som anvendes i bogens eksempler, indeholder derfor kun de resterende 477 respondenter.