

BILAG 4

Korrespondanceanalyse

- *Teknisk gennemgang af analysen og uddybende resultater*

I det følgende vil jeg præsentere de forskellige trin i en korrespondanceanalyse. Der tages udgangspunkt i den simple korrespondanceanalyse, hvorefter denne generaliseres til MCA. De indledende rekodninger er udarbejdet i SPSS, og selve korrespondanceanalysen i SPAD (version 7.3) på baggrund af kurset *Korrespondanceanalyse* på Sociologistudiet, AAU efterår 2010, Brigitte le Roux and Philippe Bonnets SPAD guide *MCA of the taste Example using SPAD (version 7.3)*.

Der findes simpel og multipel korrespondanceanalyse (forkortet CA, Correspondance Analysis og MCA, Multiple Correspondance Analysis)¹. Forskellen mellem de to er, at simpel korrespondanceanalyse undersøger sammenhænge mellem to variable, og multipel korrespondanceanalyse undersøger sammenhænge mellem et stort antal variable. (Høyen 2010: 13)

Den multiple korrespondanceanalyse bygger på principperne for den simple. Da en gennemgang af CA kan fremstilles eksemplificeret med en overskuelig krydstabel tager præsentationer af metoden ofte udgangspunkt i CA (eksempelvis Høyen 2010, Skjøtt-Larsen 2008). Jeg vælger at gøre det samme her, eksemplificeret ved en krydstabel mellem borgerne i Region Midtjyllands indtag af henholdsvis frugt og fisk opdelt i sundt, middelsundt og usundt indtag. Udgangspunktet for CA er således en klassisk krydstabel med to kategoriske variable, som er gensidigt udelukkende (Høyen 2010:13).

Simpel korrespondanceanalyse

Indledningsvis omregnes de absolutte tal til såkaldte **kolonneprofiler**. En kolonneprofil er andelen af den samlede population (1) som tilhører en bestemt gruppe, dvs et tal mellem 0 og 1. Rækker og kolonner vægtes ligedan i CA, dermed kunne man i princippet lige så vel have udregnet rækkeprofiler, men i praksis udregnes oftest kolonner (Høyen 2010: 23). Eksempelvis er kolonneprofilen for mennesker som både har sunde kostvaner i forhold til fisk og frugt 0,097.

¹ Der findes desuden typen JCA, *Joint Correspondance Analysis*, som jeg dog ikke vil behandle yderligere. For en introducerende gennemgang af denne variant af metoden henvises til eksempelvis Høyen 2010

	Frugt_sund	Frugt_mellem	Frugt_usund	Rækkeproportion (frekvens)
Fisk_sund	0,284 (2773)	0,222 (3350)	0,164 (594)	0,236 (6717)
Fisk_mellem	0,599 (5836)	0,600 (9055)	0,521 (1891)	0,590 (16782)
Fisk_usund (frekvens)	0,117 (1138)	0,178 (2683)	0,316 (1148)	0,175 (4969)
Kolonneproportion (frekvens)	0,342 (9747)	0,530 (15088)	0,128 (3633)	1 (28468)

Hver kolonne svarer til en **dimension**, og således indebærer ovenstående eksempel at gruppen af mennesker som har sunde kostvaner i forhold til frugt skal bevæge sig 0,097 enheder ud af dimensionen for sunde kostvaner i forhold til fisk.

Kolonneprofilerne kan nu omregnes til **vektorer**. En vektor er en størrelse med en bestemt retning eller position (Skjøtt-Larsen årstal: 58). Således definerer vektoren et punkts placering i et rum og kan ansues som en slags koordinat, der angiver hvor mange enheder man skal bevæge sig ud af de forskellige dimensioner for at finde punktets placering. I eksemplet ovenfor vil vektoren for borgere med usunde kostvaner opgjort på indtag af fisk have koordinaterne [0,040; 0,094; 0,040] hvilket betyder at borgerne ligger 0,040 enheder ude af dimensionen for sundt frugtindtag, herefter 0,094 enheder ude af dimensionen for mellem frugtvaner og endelig 0,040 enheder ud af dimensionen for usundt indtag af frugt. Herefter er profilen for usundt fiskeindtag beskrevet.

Tilsammen danner vektorerne en **sky** af punkter i et flerdimensionelt **rum**. Et rum udgøres således af mindst tre dimensioner, mens to dimensioner tilsammen udgør et plan (Høyen 2010: 19).

Hver profil indeholder et bestemt antal borgere, eksempelvis er der 4969 borgere som har usundt fiskeindtag. Med henblik på at deldele den enkelte profil en repræsentativ vægt i forhold til det samlede antal borgere udregnes rækkeproportion eller rækkemassen, dvs. rækkens proportion af det samlede antal borgere. Således har de 4969 borgere i profilen for usundt fiskeindtag en rækkemasse på $4969/28468=0,175$

Rækkemassen betegnes også **gennemsnitsprofilen**, eller centroiden, som udgør midtpunktet i den enkelte dimension. Således er midtpunktet i dimensionen for usundt fiskeindtag 0,175, og det fremstår således at gruppen af borgere med usunde frugtvaner ligger over gennemsnittet på dimensionen for usunde fiskevaner, hvorimod gruppen af borgere som har sundt frugtindtag ligger under gennemsnittet på angivelse af usunde fiskevaner.

Som beskrevet tidligere er korrespondanceanalyse en geometrisk dataanalyse, og således omregnes de ovenstående forskelle mellem grupper af borgere til geometriske afstande. I korrespondanceanalysen vægter man afstande således at de kategorier, der optræder sjældent ifht. de øvrige, bidrager mere til beregningerne af afstande mellem punkterne, mens de kategorier, der optræder ofte, bidrager mindre (Høyen 2010: 27). Dette gøres for at opveje for forskelle i frekvenser i kategorierne. Måden man udregner vægtningen svarer til udregningen af chi2 forskelle mellem forventede værdier og faktiske værdier, derfor betegnes de vægtede afstande **chi2-afstande** (Høyen 2010: 28).

Inertibegrebet er et kombineret begreb som samlet beskriver både masse og afstande fra et midtpunkt; jo længere væk fra midtpunktet en masse befinder sig, desto højere inert. Inerti beregnes som masse gange kvadreret afstand

fra midtpunktet. Således kan stor afstand kompensere for lav masse etc. Den **totale inerti** udgøres af summen af inertidrag fra alle rækkeprofiler tilsammen. Det vil sige at jo større afstand mellem centroid og punkter, jo højere inertiværdier, og således kan høje inertiværdier tolkes som stor spredning af punkterne i et rum, hvorimod lav samlet inerti betyder at punkterne ikke er særlig godt spredte i rummet. (Høyen 2010: 28)

I CA – men ikke i MCA - er en skys samlede inerti derfor udtryk for graden af sammenhæng mellem de undersøgte variable og kategorier. I MCA er den samlede inerti udelukkende givet af det gennemsnitlige antal kategorier pr. variabel (Le Roux & Rouanet 2010: 38)

Efter således at have konstrueret punktskyen i et mangedimensionelt rum består næste trin i at reducere dimensionaliteten, fra de mange dimensioner indtil der tilbage er et tilpas antal dimensioner til, at disse er fortolkningsmæssigt meningsfulde samt at den grafiske afbildning er mulig. Samtidig må disse tilsammen opfange størst mulig andel af skyens inerti (Skjøtt-Larsen 2008: 62-63). Høyen (2010:34) beskriver det med, at vi leder efter et rum af lavere dimensionalitet, som så nøjagtigt som muligt afbilder chi2-afstandene mellem profilerne, hvilket tilsvarende det plan (af to dimensioner) som ligger tættest på alle punkterne.

Denne proces består i, at finde den **akse** som bedst repræsenterer punkterne, eller hvor den samlede afstand mellem akserne og de enkelte punkter er mindst mulig, således at der forekommer mindst mulig informationstab. Informationstabet henviser til at den konstruerede dimension er en konstruktion, eller en tilnærmelse, og således ikke kan repræsentere samtlige punkter 100 procent. Informationstabet eller mængden af inerti, som går tabt, når det multi-dimensionale rum med punktskyen reduceres betegnes **residual inerti** (Høyen 2010: 34).

Den første akse, som konstrueres, opfanger mest inerti, den næste akse næstmest inerti osv. Hvor stor en mængde inerti en akse opfanger angives med aksens egen værdi. Dvs., at den første akse opfanger de største modsætninger i datamaterialet, den næste opfanger en stor del af de modsætninger, som den første akse ikke opfangede osv. En tommelfingerregel siger, at man skal medtage akser indtil 80% af den samlede inerti er forklaret. Den mængde inerti som ikke opfanges af de akser, der inddrages i analysen benævnes **residual inerti**.

Således kan man sige at den totale inerti udgøres af inerti i planet + residual inerti, eller af summen af alle aksers egen værdier.

Aksernes midtpunkt er i **barycentret**.

I analysen af korrespondancekortene fra resultaterne af korrespondanceanalysen inddrages for overskuelighedens skyld to akser ad gangen. Det fortolkningsmæssige arbejde består i at finde frem til de underliggende sociologiske temaer, som hver akse repræsenterer, og som altså kan siges at strukturere borgerne.

Hver akse er særligt præget af enkelte variable som tilsammen danner et tema. Ved at analysere sig frem til temaerne for akserne kan man udlede, hvilke differentieringsprincipper, der er i spil langs en akse. På samme måde kan borgerne spredning langs en akse fortolkes som deres differentiering langs en dimension, i svarafgivelsen på spørgsmål, som repræsenterer et tema. Således har personer, som placerer sig i nærheden af hinanden langs en akse på kortet

afgivet relativt ens svar på de spørgsmål, som strukturerer dimensionen, hvorimod borgere, som er placeret langt fra hinanden har afgivet forskellige svar.

Når rummet således er dannet med respondenternes spredning langs de strukturerende akser, kan der indsættes en række supplementærpunkter, også kaldet passive variable. De kaldes passive fordi de ikke indgår i selve rumdannelsen (modsat de aktive variable). Således bidrager de ikke til strukturering af dimensionerne i rummet, men kan indplaceres således, at man kan se sammenhængen mellem dem og de aktive variable. I nærværende analyse udgøres de aktive variable af indikatorer på sundhedsadfærd og holdninger hertil, og de passive variable er sociale baggrundsfaktorer.

Multipel korrespondanceanalyse

I det følgende vil jeg kort generalisere CA til MCA, idet jeg i specialet benytter mig af MCA, hvorimod CA udelukkende er præsenteret 'af pædagogiske årsager'. I MCA er udgangspunktet stadig en tovejstabel, hvorfor det gælder om at omforme en mangedimensionel tabel til en tovejstabel, dvs. individer som rækker og kategoriserede variable som kolonner (Høyen 2010: 41)

Således omdannes hver enkelt variabel til en form for dummy-kategori, som kan antage 1 eller 0 afhængig af om borgeren har angivet pågældende kategori.

I MCA konstrueres både en individsky og en sky af modaliteter og der er korrespondance mellem de to skyer. Tolkninger af relevante akser i skyen for modaliteter kan således overføres til tilsvarende akser i individskyen og omvendt. I MCA lægges vægten traditionelt på fortolkning af individskyen (Skjøtt-Larsen 2008: 64).

I MCA er den totale inerti bestemt ved forholdet mellem antal kategorier og antal spørgsmål og således ikke – som i CA – af sammenhængen mellem de undersøgte modaliteter. (Skjøtt-Larsen 2008: 64)

Endelig adskiller MCA sig fra CA ved, at de angivne egenverdier for akserne i MCA undervurderes, hvorfor det anbefales at udregne modificerede egenverdier for akserne.

Ovenstående uddybes og konkretiseres i følgende afsnit om udarbejdelse af et rum for sundhedsadfærd. Ligeledes indeholder afsnittet retningslinier og fremgangsmåder for fortolkning i MCA.

Specifik MCA – rum for sundhedsadfærd

Indledningsvis vil jeg udarbejde udregninger af variansen i de kategori- og individskyer, som fremkommer på baggrund af analysen, samt udregne bidrag til denne varians fra de respektive spørgsmål og kategorier² i modellen. Efterfølgende

² Benævnes også modaliteter

de vil jeg udvælge det antal akser, som er nødvendigt for at beskrive mest muligt af variansen i skyen. Endelig vil jeg anskueliggøre, hvilke temaer disse akser repræsenterer, samt, hvorledes de indholdsmæssigt kan fortolkes.

VARIANSEN I SKYERNE

Som beskrevet i det foregående, fremkommer på baggrund af en MCA to skyer, en individsky og en kategorisky, hvori respondenterne er spredt så meget som muligt i en mangedimensionelt rum. Variansen i individ- og kategoriskyen den samme, og den kan udregnes som det gennemsnitlige antal kategorier pr. spørgsmål minus én. Således er skyens overordnede varians ikke afhængig af data, men af fordelingen mellem spørgsmål og kategorier.

Skyens varians tilsvarende summen af egenverdier for alle akserne:

$$V_{cloud} = \frac{K}{Q} - 1$$

Efter rensning³ er der, som beskrevet, 80 kategorier fra 25 aktive spørgsmål.

Således er den samlede models varians:

$$\frac{80}{25} - 1 = 2,2$$

Af den samlede varians i skyen kan man udregne bidrag fra både kategorier og spørgsmål, således at man kan afgøre, hvor stor en del af skyens varians disse hver især står for. Hvorledes dette udregnes gennemgås i det følgende.

Spørgsmålet, q 's bidrag, Ctr_q , til skyens varians udregnes efter formlen:

$$Ctr_q = \frac{K_q - 1}{K - Q}$$

Hvor K angiver antal kategorier i modellen og Q antal spørgsmål. Det vil sige, at spørgsmålenes bidrag udelukkende er afhængige af antal kategorier i de enkelte spørgsmål. Dette er årsagen til, at det tilrådes at vælge spørgsmål med et ensartet antal kategorier (Le Roux & Rouanet 2010: 38).

Dette har imidlertid ikke været muligt i denne analyse, idet de respektive spørgsmål ikke er udformet med ensartede antal kategorier, og det indholdsmæssigt ikke har givet mening af rekodde dem til at have et ensartet antal kategorier. Derfor bidrager spørgsmålene ikke lige meget til skyens varians.

Eksempelvis bidrager spørgsmålet om borgernes rygevaner med 5 kategorier således med:

³ Det vil sige frasortering af de kategorier, som jeg – i kraft af metoden specifik multipel korrespondanceanalyse – har valgt ikke at inddrage

$$Ctr_{rygevaner} = \frac{K_{rygevaner} - 1}{K - Q} = \frac{5 - 1}{80 - 25} = 0,073$$

Således bidrager spørgsmålet om rygevaner med 7,3% af modellens samlede varians. Derimod bidrager spørgsmålet om, hvorvidt borgerne ryger i hjemmet med tre kategorier:

$$Ctr_{hjem_ryge} = \frac{K_{hjem_ryge} - 1}{K - Q} = \frac{3 - 1}{80 - 25} = 0,036$$

Og spørgsmålet om rygning i hjemmet bidrager således kun med 3,6% af modellens samlede varians mod spørgsmålet om rygevaners 7,3% bidrag.

Omvendt afhænger bidraget fra kategorierne af individernes fordeling i datamaterialet, idet frekvensen af de enkelte kategorier indgår i beregningen af bidrag. Kategorien k 's bidrag, Ctr_k til variansen i den overordnede sky udregnes således:

$$Ctr_k = \frac{1 - f_k}{K - Q}$$

Hvor f_k angiver frekvensen eller vægten af kategorien. Dette betyder at en kategori som optræder sjældent (dvs. har en lav frekvens) bidrager mere til variansen i skyen end en kategori, som optræder ofte. Dette er grundlaget for tommelfingerreglen om udelukkende at inddrage kategorier, som indeholder 5 % af populationen. (Le Roux & Rouanet 2010: 38-39)

Eksempelvis kan udregnes de respektive bidrag fra kategorierne lejlighedsryger og aldrigryger i spørgsmålet om borgernes rygevaner:

$$Ctr_{lejlighedsryger} = \frac{1 - f_{lejlighedsryger}}{K - Q} = \frac{1 - 0,0282^4}{80 - 25} = 0,0177$$

Hvorimod kategorien aldrigryger bidrager med:

$$Ctr_{aldrigryger} = \frac{1 - f_{aldrigryger}}{K - Q} = \frac{1 - 0,4471^5}{80 - 25} = 0,0101$$

Således har kategorien lejlighedsryger mere end halvanden gang så stort bidrag til den samlede varians i skyen som kategorien aldrigryger, og dette forsøger man således at tage højde for med anbefalingen om udelukkende at inddrage kategorier, som indeholder 5% af populationen (hvilket jeg dog ikke har valgt at efterleve i nærværende MCA, idet den gennemførte analyse viser, at bidragene fra de lavfrekvente kategorier ikke 'fortegner billedet')

⁴ Udregnet på baggrund af N for variabelen Rygevaner, hvor N=28078, aflæst af Excel outputtet, Corem1, fra MCA'en, bilag 5

⁵ Aflæst fra Excel outputtet, Corem1 fra MCA'en

De øvrige kategorier i variabelen om rygevaner bidrager således: Storryger 0,0162, Moderat ryger 0,0164 og Eksryger 0,024. Samlet set bidrager kategorierne således med $(0,0177 + 0,0101 + 0,0164 + 0,0177 + 0,0124 = 0,0727)$ 7,3% af variansen i skyen, tilsvarende udregningen om de enkelte spørgsmåls bidrag tidligere.

Nedenfor fremgår en samlet oversigt over de 25 aktive spørgsmål og 80 kategorier med svarfrekvenser (i absolutte tal, n_k , og relativ andel, f_k) samt henholdsvis spørgsmålenes og kategoriernes bidrag til skyens varians (Ctr_q, Ctr_k).

Inden oversigten over spørgsmålene og kategoriernes respektive bidrag præsenteres en opgørelse over temaernes bidrag:

Tabel 1: Oversigt over temaers bidrag i MCA af et rum for sundhedsadfærd og -holdninger

Kost 20 %
Rygning 14,6 %
Alkohol 23,6 %
Motion 20 %
Helbred 7,3 %
Vægt 14,5 %

Tabellen er udarbejdet med udgangspunkt i Excel outputtet, Corem1, fra MCA'en

Som beskrevet i kapitel 5, *Operationalisering* har udvælgelsen af variable dels været styret af det tilgængelige antal variable til belysning af de enkelte områder af sundhedsadfærden, hvorfor det ikke har været muligt at inddrage lige mange variable og kategorier til belysning af de forskellige temaer. Imidlertid er der et nogenlunde jævnt bidrag fra de tre temaer *Kost*, *Alkohol* og *Motion*, hvorimod *Rygning* og *Vægt* bidrager en smule mindre, og endelig bidrager *Helbred* relativt mindst. Dette indebærer, at det i analysen er vigtigt at holde sig det for øje, i den forstand at fordelingen af variable og kategorier mellem de forskellige temaer også har indflydelse på bidrag langs akserne. Endvidere er der mulighed for at *Vægt* får stor betydning, idet *Vægt*, *Kost* og *Motion* som bekendt hænger sammen.

Af tabellen nedenfor fremgår spørgsmål og kategoriernes bidrag til den samlede varians:

KOST			
<i>Sundfisk</i>	n_k	f_k i %	Ctr_k i %
sund_fisk	6711,00	23,50	1,39
middelsund_fisk	16869,00	59,07	0,74
usund_fisk	4980,00	17,44	1,50
I alt	28560,00	100,00	3,64
<i>Sundfrugt</i>	n_k	f_k i %	Ctr_k i %
sund_frugt	9715,00	34,25	1,20
middelsund_frugt	15033,00	53,00	0,85
usund_frugt	3617,00	12,75	1,59
I alt	28365,00	100,00	3,64
<i>Sundfedt</i>	n_k	f_k i %	Ctr_k i %

sund_fedt	5372,00	19,10	1,47
middelsund_fedt	16531,00	58,76	0,75
usund_fedt	6230,00	22,14	1,42
I alt	28133,00	100,00	3,64
<i>Sundgrønt</i>	<i>n_k</i>	<i>f_k i %</i>	<i>Ctr_k i %</i>
sund_grønt	7509,00	26,29	1,34
middelsund_grønt	18870,00	66,07	0,62
usund_grønt	2181,00	7,64	1,68
I alt	28560,00	100,00	3,64
<i>Vurdering af kost</i>	<i>n_k</i>	<i>f_k i %</i>	<i>Ctr_k i %</i>
kostvur_sund	12409,00	43,63	1,02
kostvur_middelsund	15122,00	53,17	0,85
kostvur_usund	909,00	3,20	1,76
I alt	28440,00	100,00	3,64
<i>Ønske om at spise sundere</i>	<i>n_k</i>	<i>f_k i %</i>	<i>Ctr_k i %</i>
sunderespis_ja	15192,00	64,71	0,64
sunderespis_nej	8284,00	35,29	1,18
I alt	23476,00	100,00	1,82
Samlet bidrag <i>KOST</i>			20,00
RYGNING			
<i>Rygevaner</i>	<i>n_k</i>	<i>f_k i %</i>	<i>Ctr_k i %</i>
Storryger	3124,00	11,13	1,62
moderat ryger	2690,00	9,58	1,64
Lejlighedsryger	792,00	2,82	1,77
Eksryger	8917,00	31,76	1,24
aldrig ryger	12555,00	44,71	1,01
I alt	28078,00	100,00	7,27
<i>ønske om rygestop</i>	<i>n_k</i>	<i>f_k i %</i>	<i>Ctr_k i %</i>
ønsker rygestop	4631,00	16,67	1,52
ønsker ikke rygestop	1831,00	6,59	1,70
Ikkeryger	21321,00	76,74	0,42
I alt	27783,00	100,00	3,64
<i>rygning i hjemmet</i>	<i>n_k</i>	<i>f_k i %</i>	<i>Ctr_k i %</i>
Dagligt	5310,00	18,86	1,48
ugentligt eller sjældnere	2684,00	9,54	1,64
Aldrig	20154,00	71,60	0,52
I alt	28148,00	100,00	3,64
Samlet bidrag <i>RYGNING</i>			14,55
ALKOHOL			
<i>Alkoholforbrug</i>	<i>n_k</i>	<i>f_k i %</i>	<i>Ctr_k i %</i>
intet alkoholforbrug	2355,00	8,25	1,67

lavrisiko forbrug	19243,00	67,40	0,59
middelrisiko forbrug	3486,00	12,21	1,60
højrisiko forbrug	3465,00	12,14	1,60
I alt	28549,00	100,00	5,45
<i>binge drinking</i>	n_k	$f_k i \%$	$Ctr_k i \%$
Dagligt	328,00	1,59	1,79
Ugentligt	2259,00	10,96	1,62
Månedligt	4231,00	20,53	1,44
Aldrig	13794,00	66,92	0,60
I alt	20612,00	100,00	5,45
<i>dage med alkoholindtag/uge</i>	n_k	$f_k i \%$	$Ctr_k i \%$
0-2 dage	17016,00	66,71	0,61
3-4 dage	4089,00	16,03	1,53
5-7 dage	4403,00	17,26	1,50
I alt	25508,00	100,00	3,64
<i>alkohol udenfor måltiderne</i>	n_k	$f_k i \%$	$Ctr_k i \%$
alkohol udenfor måltider	4830,00	18,76	1,48
ikke alkohol udenfor måltider	20919,00	81,24	0,34
I alt	25749,00	100,00	1,82
<i>Alkoslags</i>	n_k	$f_k i \%$	$Ctr_k i \%$
mest øl	6189,00	26,26	1,34
mest spiritus	13184,00	55,95	0,80
mest vin	1099,00	4,66	1,73
blandet alkoholindtag	3093,00	13,13	1,58
I alt	23565,00	100,00	5,45
<i>ønske om at nedsætte alkoholforbrug</i>	n_k	$f_k i \%$	$Ctr_k i \%$
alkostop_ja	1661,00	6,99	1,69
alkostop_nej	22087,00	93,01	0,13
I alt	23748,00	100,00	1,82
Samlet bidrag alkohol			23,64
MOTION			
<i>Dage med mindst 30 min fysisk aktivitet</i>	n_k	$f_k i \%$	$Ctr_k i \%$
0-1 dag	4710,00	16,63	1,52
2-3 dage	7539,00	26,62	1,33
4-5 dage	6590,00	23,27	1,40
6-7 dage	9480,00	33,48	1,21
I alt	28319,00	100,00	5,45
<i>idræt i fritiden</i>	n_k	$f_k i \%$	$Ctr_k i \%$
idræt nej	13406,00	47,55	0,95
idræt ja	14790,00	52,45	0,86

	I alt	28196,00	100,00	1,82
<i>karakteristik af fritid</i>				
		n_k	$f_k i \%$	$Ctr_k i \%$
tit moderat til hård motion		7006,00	25,05	1,36
lidt moderat motion		16787,00	60,02	0,73
stillesiddende aktiviteter		4178,00	14,94	1,55
	I alt	27971,00	100,00	3,64
<i>Cykle</i>				
		n_k	$f_k i \%$	$Ctr_k i \%$
cykler året rundt		5601,00	21,08	1,43
cykler om sommeren		7079,00	26,65	1,33
cykler aldrig		13884,00	52,27	0,87
	I alt	26564,00	100,00	3,64
<i>egen vurdering af form</i>				
		n_k	$f_k i \%$	$Ctr_k i \%$
god form		9270,00	32,82	1,22
nogenlunde form		12926,00	45,76	0,99
dårlig form		6053,00	21,43	1,43
	I alt	28249,00	100,00	3,64
<i>ønske om at være mere aktiv</i>				
		n_k	$f_k i \%$	$Ctr_k i \%$
mere aktiv, ja		18887,00	76,01	0,44
mere aktiv, nej		5962,00	23,99	1,38
	I alt	24849,00	100,00	1,82
Samlet bidrag <i>MOTION</i>				20,00
HELBRED				
<i>egen vurdering af generelt helbred</i>				
		n_k	$f_k i \%$	$Ctr_k i \%$
fremragende helbred		2964,00	10,46	1,63
vældig godt helbred		11008,00	38,84	1,11
godt helbred		10763,00	37,97	1,13
mindre godt helbred		3159,00	11,14	1,62
dårlig helbred		451,00	1,59	1,79
	I alt	28345,00	100,00	7,27
Samlet bidrag <i>HELBRED</i>				7,27
VÆGT				
<i>BMI</i>				
		n_k	$f_k i \%$	$Ctr_k i \%$
Undervægtig		376,00	1,34	1,79
Normalvægtig		12836,00	45,67	0,99
Overvægtig		10438,00	37,13	1,14
svært overvægtig		4459,00	15,86	1,53
	I alt	28109,00	100,00	5,45
<i>egen vurdering af vægt</i>				
		n_k	$f_k i \%$	$Ctr_k i \%$
lidt lav vægt		991,00	3,50	1,75
tilpas vægt		9472,00	33,48	1,21

lidt for høj vægt	13693,00	48,40	0,94
alt for høj vægt	4138,00	14,63	1,55
I alt	28294,00	100,00	5,45
<i>ønske om at tabe vægt</i>	<i>n_k</i>	<i>f_k i %</i>	<i>Ctr_k i %</i>
ønsker vægttab	5602,00	20,72	1,44
ønsker muligvis vægttab	12227,00	45,21	1,00
ønsker ikke vægttab	9213,00	34,07	1,20
I alt	27042,00	100,00	3,64
Samlet bidrag <i>VÆGT</i>			14,55

Ovenstående variable og kategoriers bidrag er til den samlede varians. Nu rettes blikket mod, hvor mange akser, der må medtages i analysen med henblik på at opsummere så meget som muligt af denne i et rum af lavere dimensionali-tet.

Den maksimale dimensionalitet af en sky er $K - Q = 55$. Det vil sige, at skyens samlede varians kan opsummeres langs højst 55 akser. En tommelfingerregel er dog at analysere akser indtil 80 % af variansen i skyen er forklaret.

Variansen i en sky langs en given akse betegnes aksens eigenværdi, λ_l (Le Roux & Rouanet 2010: 39).

Den første principalakse opsummerer den største andel af variansen (dvs. har den største eigenværdi), den næste akse den næststørste etc (Le Roux & Rouanet 2010: 39). Med henblik på at opnå et antal akser, der gør det muligt at lave en meningsfuld fortolkning af rummet, er målet således at opfange så meget af den samlede varians på så få akser som muligt.

Variansraten af en akse, dvs. andelen af skyens varians, som opsamles af akserne, l , er

$$\tau_l = \frac{\lambda_l}{V_{cloud}} = \frac{\lambda_l}{\frac{K}{Q} - 1}$$

Skyens samlede varians tilsvarende som tidligere nævnt summen af eigenværdier (dvs. varianser på de enkelte akser), hvorfor aksernes gennemsnitlige eigenværdi, $\bar{\lambda}_l$, er:

$$\bar{\lambda}_l = \frac{1}{Q}$$

(Le Roux & Rouanet 2010: 46)

Det vil sige, at den gennemsnitlige eigenværdi for akserne i denne model er $1/25=0,04$

I MCA undervurderes den andel af den samlede varians, som opsummeres langs de første akser, hvorfor Le Roux & Rouanet (2010: 39) med henvisning til Benzecri (1992: 412) anbefaler at der udregnes modificerede eigenværdier.

Dette gøres kun for de akser, som har eigenverdier der overstiger gennemsnittet, dvs. hvor $\lambda_l > \bar{\lambda}$. I dette tilfælde gælder det for 20 akser⁶.

Beregninger foretages gennem to skridt (Le Roux & Rouanet 2010: 39)⁷:

En pseudo-eigenverdi, λ_l ud fra formlen:

$$\lambda_l = \left(\frac{Q}{Q-1} \right)^2 (\lambda_l - \bar{\lambda})^2$$

En sum, S , som er:

$$S = \sum_{l=1}^{l_{max}} \lambda_l$$

Hvorefter det gælder for $l \leq l_{max}$ at de modificerede rater er lig med $\tau_l = \frac{\lambda_l}{S}$

Dette betyder, at den modificerede eigenverdi for den første akse, med en eigenverdi på 0,1568, beregnes således:

$$\lambda_l = (25/25 - 1)^2 (0,1568 - (1/25))^2 = 0,01481$$

Hvorefter denne værdi divideres med summen af de modificerede eigenverdier, således at der opnås en rate dvs.:

$$\tau_l = \frac{0,01481}{0,0312} = 0,4745$$

Modificerede rater for de 19 akser fremgår af nedenstående tabel:

Tabel 2: Aksernes varians, variansrater og modificerede variansrater

Aksernes gennemsnitlige eigenverdi = 0,04

Akse, l	eigenverdi, λ_l	Variansrater	Modificeret eigenverdi	Modificeret variansrate	Kumuleret modificeret rate
1	0,1568	0,126	0,014813679	0,474534645	0,474534645
2	0,1315	0,106	0,009064335	0,290362769	0,764897414
3	0,0913	0,073	0,002845759	0,091159747	0,856057161
4	0,0794	0,064	0,001677371	0,053732138	0,909789299

⁶ aflæst af Excel outputtet fra MCA'en, bilag 5

⁷ den nærmere udregning af det nedenstående er desuden gennemgået med udgangspunkt i Le Roux & Bonnets SPAD-guide pp 12-13.

5	0,0680	0,055	0,000848738	0,027188078	0,936977378
6	0,0639	0,051	0,00061974	0,019852479	0,956829856
7	0,0608	0,049	0,000466383	0,014939914	0,97176977
8	0,0564	0,045	0,00029095	0,00932015	0,981089921
9	0,0547	0,044	0,000232367	0,007443542	0,988533462
10	0,0516	0,041	0,000146103	0,004680188	0,99321365
11	0,0489	0,039	8,53081E-05	0,002732722	0,995946372
12	0,0476	0,038	6,25242E-05	0,002002871	0,997949244
13	0,0458	0,037	3,65949E-05	0,001172263	0,999121506
14	0,0440	0,035	1,69253E-05	0,000542176	0,999663682
15	0,0427	0,034	7,75863E-06	0,000248536	0,999912219
16	0,0412	0,033	1,67408E-06	5,36268E-05	0,999965846
17	0,0408	0,033	7,27978E-07	2,33197E-05	0,999989165
18	0,0405	0,033	3,22324E-07	1,03252E-05	0,999999491
19	0,0401	0,032	1,17436E-08	3,7619E-07	0,999999867
20	0,0401	0,032	4,15767E-09	1,33185E-07	1
I alt	1,2462	1	0,0312	1	

Værdierne er udarbejdet med udgangspunkt i Exceloutputtet, Corem4 fra MCA'en

En tommelfingerregel i MCA siger, at man i analysen skal medtage dimensioner indtil 80% af skyens samlede varians er forklaret, eller så mange akser, som det fortolkningsmæssigt er meningsfuldt at inddrage (Le Roux & Rouanet 2010: 51-52). I ovenstående tabel kan det af rækken med de kumulerede modificerede rater aflæses, at de første tre akser skal inddrages i analysen med henblik på at forklare 80% af variansen i skyen. I analysen i kapitel 6 har jeg imidlertid af omfangsmæssige hensyn valgt at koncentrere mig om akse 1 og 2. Disse to beskrives i kapitel 6, akse 3 i det følgende, hvormed 80 % af inertien er beskrevet.

Efter således at have dannet et indtryk af det, man kan kalde analysens forklaringskraft, undersøges hvilke variable eller grupper af variable, der er med til at skabe variansen langs den enkelte akse.

Hvert af de aktive spørgsmål i analysen har deres gennemsnittpunkt i barycentret – og har kategorier, der afviger herfra til begge sider. Kategorier placeret på samme side af barycentret vil ofte optræde sammen (hos samme individer), mens kategorier på modsat side sjældent vil optræde sammen. Kategorier med en lille vægt, dvs. få individer placerer sig langt fra barycentret i forhold til kategorier med mange individer. Ikke alle kategorier er lige godt repræsenteret langs en akse; man betegner en variabel eller kategori som forklarende for en akse, hvis dens bidragsværdi er

større end gennemsnittet af bidragene fra alle variablene og kategorierne (Høyen 2010: 56, Le Roux & Rouanet 2010: 52).

En kategoris bidragsværdi kan udtrykkes med formlen:

$$ctr_k = \left(\frac{f_k}{Q} (y_k)^2 \right) / \lambda$$

Hvor f_k angiver kategoriens vægt i procent, y_k kategoriens position på akse med λ varians. Bidragsværdien fra en kategori afgøres således både af afstand til barycentret og af vægt.

Det gennemsnitlige variabelbidrag, $100/Q = 100/25 = 4$, og det gennemsnitlige kategoribidrag, $100/K = 100/80 = 1,25^8$.

Fortolkningen af en akse tager udgangspunkt i de variable, som bidrager over gennemsnittet til variansen langs akse. Ved at undersøge disse kan det afgøres, hvilke temaer akse repræsenterer. Desuden fokuseres på modaliteter og disses orientering i forhold til barycentret med henblik på at forstå, hvilke positioner og ligheder der er i spil.

I det følgende vil jeg fortolke 3 med udgangspunkt i variable og modaliteter med bidrag over gennemsnittet. Akserne 1 og 2 er præsenterede i kapitel 6.

AKSE 3:

De variable, som bidrager mere end gennemsnitligt til variansen langs akse 3, fremgår af nedenstående tabel:

Tabel 5: Tabel til fortolkning af akse 3

Variabel	Orientering	
	Positiv	Negativ
RYGNING (2,69%)		
Rygevaner (3,16)	(Moderat ryger (1,25))	Eksryger (1,44)
ALKOHOL (56,23%)		
Alkoholforbrug (12,89)	Intet forbrug (1,75)	Middelrisiko forbrug (3,79)
	Lavrisiko forbrug (1,85)	Højrisiko forbrug (5,5)
Binge drinking (11,21)	(Aldrig (0,8))	Månedlig (1,39)
		Ugentlig (7,58)
		Daglig (1,44)
Dage med alkoholindtag (11,9)	0-2 dage (2,58)	3-4 dage (2,86)
		5-7 dage (6,46)
Alkohol udenfor måltider (10,04)	(Nej (1))	Ja (9,03)
Ønske at nedsætte forbrug (10,21)	(Nej (0,56))	Ja (9,64)
VÆGT (24,28%)		
BMI (6,1)	Normalvægt (1,76)	Overvægt (3,36)
Egen vægtvurdering (10,54)	Tilpas (1,76)	Lidt for høj (5,18)

⁸ Hvilke variabel- og kategoribidrag, der overstiger gennemsnittet er aflæst af Exceloutputtet fra MCA'en.

Ønske vægttab (9,88)

Nej (5,38)

Muligvis (4,5)

Samlet bidrag: 21,03 (positiv) + 62,17 (negativ) =83,2

Det fremgår af tabel 5, at variansen langs akse 3 er struktureret omkring variable indenfor temaerne *Rygning*, *Alkohol*, *Motion* og *Vægt*. Temaet *Alkohol* bidrager med over halvdelen af variansen langs akse 3 (54,85%), *Vægt* bidrager med 26,45% af aksens varians, og *Rygnig* og *Motion* bidrager meget lidt (henholdsvis 2,3% for *Rygnig* og 1,59% for *Vægt*). Samlet set er 85,19% af variansen langs akse 3 beskrevet ved kategorierne i tabel 5, hvoraf de 62,33% af variansen er negativt orienteret og 22,97% positivt orienteret. Dermed er det kategorierne med negativ orientering som først og fremmest skaber variansen langs akse 3.

Kategorierne med negativ orientering er først og fremmest indikatorer på højt alkoholforbrug, det vil sige kategorierne for et alkoholforbrug, som ifølge Sundhedsstyrelsen medfører middel eller høj risiko for udvikling af alkoholrelaterede sygdomme. Dernæst er det kategorierne for ugentlig og månedlig binge drinking, indtag af alkohol mindst hver anden dag, og udenfor måltiderne. Kategorien for at ønske at nedsætte sit alkoholforbrug har ligeledes negativ orientering. Hvor det på akse 2 er kategorierne for et højrisiko alkoholforbrug, for daglig binge drinking etc., som er repræsenterede, er det altså på akse 3 kategorier for et højt forbrug, men alligevel et mere moderat forbrug, end de, som er repræsenterede langs akse 2.

I samme ende af akse 3 er kategorien for at være stoppet med at ryge. I forhold til temaerne *Vægt* og *Motion*, er kategorierne for ofte at dyrke moderat til motion ligeledes orienteret negativt på akse 3, men samtidig er det i denne ende at kategorierne for at være overvægtig, selv at vurdere sin vægt til at være lidt for høj, og for muligvis at ønske et vægttab. Man kan således sige, at den negativt orienterede ende af akse 3 repræsenterer borgere med moderate alkohol- eller vægtproblemer (ifølge Sundhedsstyrelsens normativer for sund livsstil), hvor akse 2 tegner billedet af noget mere ekstremt.

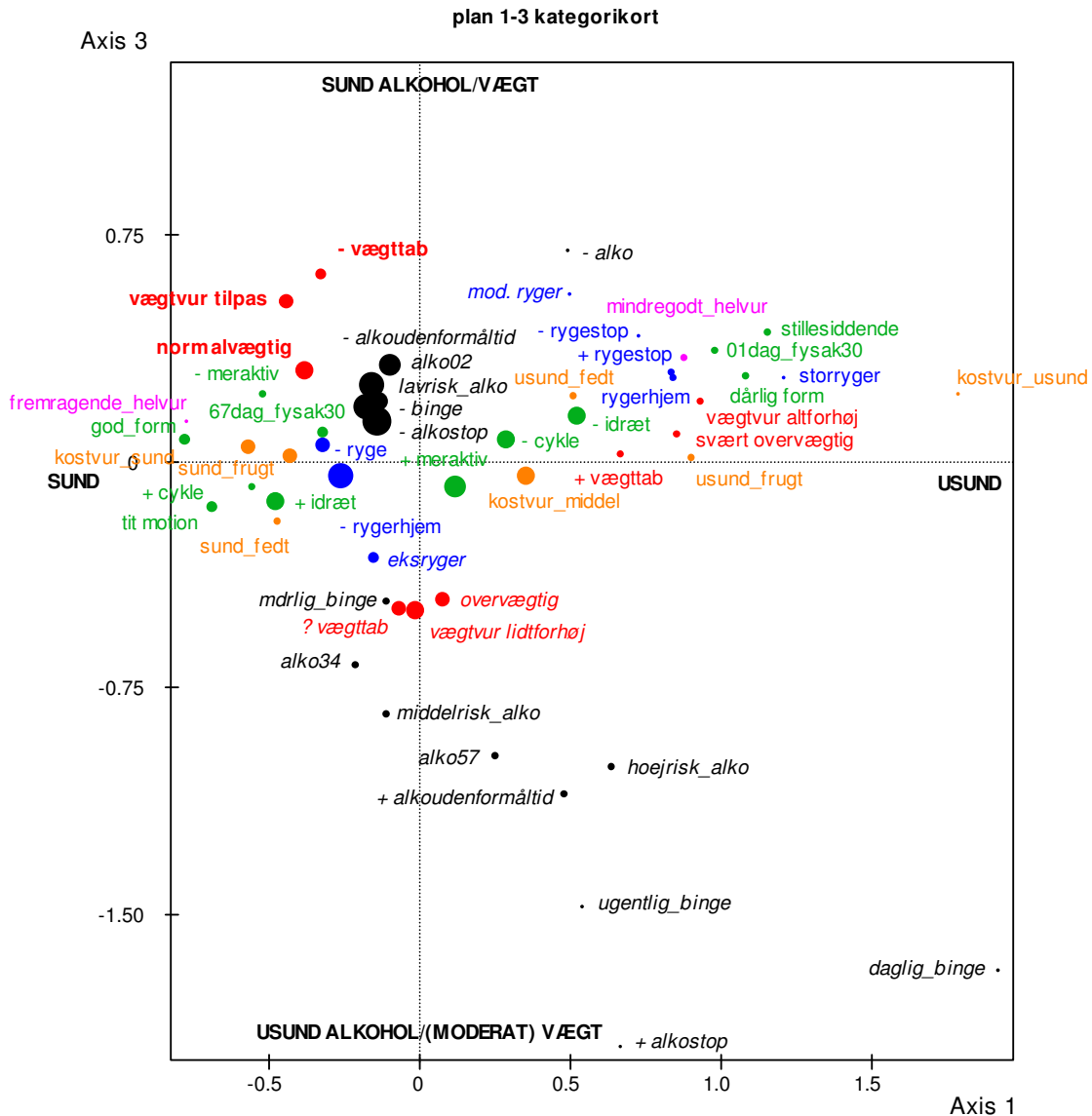
I den positive ende af akse 3 placerer sig kategorierne for ikke at indtage alkohol, eller at have et lavrisikoforbrug, og at drikke alkohol ingen eller få dage om ugen. I forhold til temaet *Motion* er kategorien for at betegne stillesiddende aktiviteter som karakteristisk for fritiden. Dog placerer denne kategori sig sammen med kategorier for at være normalvægtig og selv vurderer sin vægt som tilpas – og i tråd hermed ikke ønske vægttab.

Det er bemærkelsesværdigt, at kategorien for at dyrke meget motion placerer sig sammen med kategorierne for overvægt, og omvendt. For så vidt at kategorien for ønsket om vægttab placerer sig sammen med disse, kan motionen imidlertid være et udtryk netop for at arbejde på at reducere vægten.

Samlet set tegner akse 3 en opposition mellem et moderat højt alkoholindtag og moderat overvægt i den ene ende, og intet eller et lavt alkoholforbrug sammen med normal vægt i den anden ende.

I kapitel 6 har jeg beskrevet planet, som skabes af akse 1-2. Der er nogle aspekter omkring alkohol og vægt, som ikke opfanges i akse 2. Først og fremmest beskrives i akse 2 den ekstreme usunde adfærd angående alkohol og vægt, hvor-

imod akse 3 opfanger de mere moderate former for stort alkoholforbrug og overvægt. Desuden kan forandringsparatheden i forhold til ændring af alkoholvaner beskrives ved at se på akse 3. Derfor beskrives planet 1-3 kort nedenfor efter præsentationen af kategorikortet for planet:



Af figuren med kategorikortet for plan 1-3 fremgår det at forandringsparatheden i forhold til vægt er således, at de moderat overvægtige (som selv vurderer deres vægt som 'lidt for høj') angiver, at de muligvis ønsker et vægttab, hvorimod de svært overvægtige er mere klare i mælet, idet de entydigt ytrer ønske om et vægttab. Dette kan tolkes som et udtryk for, at forandringsparatheden øges i takt med øget opfattelse af, at vægten udgør et problem, eller at der er tale om en anomali i forhold til sundhedsdiskursen og de vægtidealer som italesættes med denne.

Ligeledes fremstår det at forandringsparathed i forhold til alkoholvaner er placeret i nærheden af modaliteterne for at have et højt alkoholforbrug i nederste højre hjørne med ugentlig eller daglig binge drinking. Med inddragelse af akse 3 i planet er modaliteterne for at binge drikke månedligt og at drikke alkohol ca. hver anden dag, ligeledes beskrevet. Disse kategorier placerer sig i nærheden af hinanden, og i nærheden af modaliteterne for at være moderat overvægtig, og ligger omtrent midt mellem indikatorerne for et henholdsvis højt og lavt alkoholforbrug.

Endelig, i forhold til forandringsparathed fremgår det af kategorikortet for dette plan, at forandringsparathed i forhold til motion er placeret hos de middelaktive.

Overordnet set kan man sige, at planet 1-3 først og fremmest skaber oppositionen mellem det sunde overfor det usunde og en opposition mellem at drikke og ikke drikke alkohol, men desuden forklarer nogle aspekter omkring forandringsparathed.

Supplementærpunkter

Som beskrevet har jeg indsat indikatorer på sociale baggrundsfaktorer som supplementære eller passive variable i MCA'en. Dette indebærer, at de kan undersøges i rummet for sundhedsadfærd, men at de ikke har bidraget til variansen langs akserne i rummet. Teknisk set tildeles disse kategorier ingen masse, og bidrager derfor ikke til inertien i dimensionerne (Jæger 2003: 9).

Koordinaterne af en kategori, y^k , kan udregnes som gennemsnittet af koordinaterne, y^i af punkter for individer, n_k , som har valgt kategorien, divideret med kvadratroden af aksens egen værdi, $\sqrt{\lambda}$. Formlen er således:

$$y^k = \frac{1}{\sqrt{\lambda}} \sum_{i \in I_k} y^i / n_k$$

(Le Roux & Rouanet 2010: 41)

Når man skal afgøre, hvorvidt der er store afvigelser mellem kategorierne fra supplementære variable - altså hvorvidt individerne adskiller sig i forhold til variabelen langs en akse – angiver en tommelfingerregel, at standardafvigelser over 1 mellem kategorierne kan betegnes som store, afvigelser mellem 0,5 og 1 kan betegnes som moderate, og afvigelser under 0,5 kan betegnes som små (Le Roux & Rouanet 2010: 59)

Fortolkningen af de supplementære variable tager således udgangspunkt i denne værdi, samt i, hvorledes kategorierne fra variablene er placeret i kategorikortene. I nedenstående tabel fremgår de variable, som afviger med mindst 0,5 standardafvigelse langs de tre akser (præsenteres hver for sig). Følgende variables kategorier afviger uvæsentligt langs alle tre akser:

Baggrundsfaktorer: Køn, Alder i 1+års intervaller, Ejerskab af bolig, Ejerskab af bil, Civilstand, Egne børn, Hjemmeboende børn, Bymæssighed, Bopælskommune

Sundhedspolitiske holdningsspørgsmål: Rygeforbud (for/mod), Hævede afgifter på alkohol (for/mod), Hævede afgifter på slik og sodavand (for/mod), ansvar for sundhed (personligt/samfundets)

Helbred og trivsel:

Dagligdagens stress: Bragt ud af ligevægt over noget uventet (hyppighed), Følt sig nervøs og stresset (hyppighed), Følt sig i stand til at klare personlige problemer (hyppighed), Følt sig i stand til at håndtere dagligdagens udfordringer (hyppighed), Blevet vred pga. ting, man ikke selv er herre over (hyppighed)

Belastninger: Arbejdssituation, Forhold til partner, Partners/families/venners sygdom, Dødsfald blandt de nærmeste

Følgende kategorier har moderat – store standardafvigelse langs akse 1:

Variabel	Positiv orientering	Negativ orientering	Standardafvigelse
Baggrundsfaktorer			
Uddannelse (dunkode)	grundskole	forskerniv	0,95
Beskæftigelsesbranche (branche2)	industri/energi/land	finans/forsikring/ejendom	0,62
Personlig indkomst (egenink_10)	100000 - 149000	850000+	0,63
Husstandsindkomst (husindk_10)	100000-149000	850000+	0,72
Civilstand (civilstand)	separeret alene	gift/reg. partner	0,48
Egne børn (egneboern)	nej	ja	0,16
Hjemmeboende børn (boernhj)	5+	3	0,49
Bopælskommune (kommune)	Lemvig	århus	0,25
Helbrd og trivsel			
Smerter (smerter)	meget stærke smerter	ingen smerter	1,09
Lanvarige sygdomme (langsyg1)	ja	nej	0,51
Dagligdagens stress:			
Følt sig i stand til at kontrollere vigtige ting i livet (ukontrol1)	ofte	aldrig	0,54
Følt at tilværelsen formede sig efter hovedet (forme1)	aldrig	ofte	0,66
Oplevet ikke at kunne overkomme de ting, man skal (overkom1)	ofte	aldrig	0,51
Følelse af at have styr på tingene (harstyr1)	aldrig	ofte	0,72
Følt at problemerne hober sig op, så man ikke magter dem (ejmagte1)	ofte	aldrig	0,55
Belastninger:			
Økonomi (oekk1)	meget	nej	0,83
Boligsituation (bolig1)	meget	nej	0,59
Forhold til familie og venner (famfo1)	meget	nej	0,63
Egen sygdom (sygselv1)	meget	nej	1,04

Følgende kategorier har moderate – store standardafvigelser langs akse 2:

Variabel	Positiv orientering	Negativ orientering	Standardafvigelse
Baggrundsfaktorer			
Beskæftigelsesbranche (branche2)	sundheds- /socialvæsen	bygge/anlæg	0,65
Fysisk belastning i arbejdet (jobbelastning)	Tungt eller hurtigt arbejde som er anstrengende	hovedsageligt stillesiddende arbejde, som ikke kræver legemlig anstrengelse	0,51

Følgende kategorier har moderat – store standardafvigelser langs akse 3:

Akse 3

Variabel	Positiv orientering	Negativ orientering	Standardafvigelse
Baggrundsfaktorer			
Personlig indkomst (egenink_10)	0-99000	850000+	0,65
Helbrd og trivsel			
Smerter (smerter)	meget stærke smerter	meget lette smerter	0,5