LUNDS UNIVERSITET
Ekonomihögskolan
Statistiska institutionen

Klusteranalys på bas av konsumenters nyttovärden från en conjointstudie

Författare: Anders Nilsson
Kandidatuppsats
Nivå: 61-90 hp
Juni 2009

Handledare: Björn Holmquist och Antonio Marañon

Sammanfattning

Syftet med uppsatsen ligger i att genomföra en kundsegmentering med hjälp av klusteranalys på bas av kundpreferenser från en conjointanalys. Datamaterialet rör konsumentpreferenser med de två attributen öppning, med fem nivåer, och förpackning, om fyra nivåer. Klustermetoderna som används är K-mean, K-median och Fuzzy klustering. Segmenteringen baseras på att använda sig av standardiserade och ickestandardiserade part-worth värden och jämföra utfallen från de olika segmenten. Jämförelsen görs med Adjusted Rand Index efter Hubert & Arabie (1985) och Hit Ratio vilket används inom diskriminantanalys.

Utifrån att studera F-kvoten väljes fem kluster. Resultaten visar att nivåerna AC (öppning) och AM (förpackning) är de mest föredragna alternativen för både icke-standardiserade och standardiserade part-worth värden. Dessa alternativ visade sig också vara de mest föredragna alternativen efter modellvärdering. Något ytterligare segment är svårare att hitta. Eventuellt skulle BC (öppning) vara ett alternativ som föredras, utifrån K-means ickestandardiserade och standardiserade part-worth värden och för K-median med standardiserade nyttovärden. Vid borttagning av AC visar Fuzzy c-means klustermetod att öppningsalternativet DL är att föredra. Vid borttagning av förpackningsalternativet AM föredras istället BM. Överlag har segmenten högre part-worth värden för öppningsalterantiven än förpackningsalterantiven vilket antyder att segmenten väljer att prioritera typ av öppning före typ av förpackning. Dock finns det variationer bland segmenten där vissa segment föredrar öppning mer än andra segment.

Adjusted Rand Index och Hit Ratio visar på olika resultat för klustermetoderna. Adjusted Rand Index visar att Fuzzy c-means har högst överensstämmelse för grupperingar vid jämförelse av standardiserade och ostandardiserade nyttovärden. Enligt Hit Ratio har K-means högst andel korrekt klassificerade. K-median och Fuzzy c-means har en väldigt låg klassificeringsgrad. En standardisering medför enbart en liten förbättring i andelen korrekt klassificerade.

Nyckelord: Klusteranalys, K-median, K-mean, Fuzzy clustering, Conjointanalys

Abstract

The aim of this thesis is to conduct a customer segmentation with the help of cluster analysis with a dataset based on customer preferences from a conjoint study. The data material consists of consumer preferences concerning the attributes opening, with five levels, and the attribute package with four levels. The methods used in cluster analysis of use involve K-mean, K-median and Fuzzy clustering. The segmentation is based on comparing standardized as well as non-standardized part-worth values and comparing the segment structure. The comparison is done with Adjusted Rand Index presented by Hubert & Arabie (1985) and Hit Ratio which is used in discriminant analysis.

By looking at the F-quotient, five clusters are chosen. Results indicate that the levels AC (opening) and AM (package) are the most preferred alternatives, both for non-standardized and standardized part-worth values. These results are also found after the cluster groups have been validated. Despite that five clusters have been chosen, it is hard to find another segment. BC (opening) could be an alternative that is prefered, by looking at K-means non-standardized and standardized part-worth values and standardized part-worth values for K-median. When removing AC from the data set, results for Fuzzy c-means indicate that DL is preferred. Meanwhile BM is preferred when AM has been removed. On an overall, the segments have higher part-worth values for opening than for package, which would imply that the segments prefer type of opening to type of package. However, there are variations among the segments where some segments prefer type of opening more than other segments.

Adjusted Rand Index and Hit Ratio show different results for the cluster methods. Adjusted Rand Index indicates that Fuzzy c-means has the best agreement of groupings when comparing standardized and non-standardized part worth values. According to Hit Ratio, K-mean shows the best classification rate whereas K-median and Fuzzy c-means have a very low classification rate. According to Hit Ratio, a standardization did only show a very small improvement of the number of correct classified cases.

Keywords: Cluster analysis, K-median, K-mean, Fuzzy clustering, Conjoint analysis

Innehåll

1 Inledning

1.1 Bakgrund

Conjointanalys tillhör familjen multivariata tekniker som utvecklades av Green och Rao (1971) under början av 1970-talet och används som metod för att mäta och analysera konsumentpreferenser, Douglas & Green (1995). Conjointanalys har kommit att bli en statistisk metod som är vanligt förekommande inom marknadsföring; inom området för marknadsanalys. Tillämpningsområden för conjointanalys inom marknadsföring är exempelvis att effektivisera en produkts design, där informationen från en conjointanalys ger möjlighet att på förhand veta vilka attribut hos produkten konsumenten anser är viktiga eller oviktiga. Preferensen mäts i form av nyttovärden och ett antagande görs att varor/egenskaper med höga nyttovärden kommer att föredras framför de med låga värden, SPSS (1997). Resultaten från en conjointanalys går att tillämpa för vidare analys. En av dessa tillämpningsområden är att använda nyttovärdena för en marknadssegmentering, Hair et al (2005).

Vanligtvis har konsumenter olika preferensstrukturer och ett steg för att hitta rätt marknadsföringsstrategi är att genom en marknadssegmentering gruppera konsumenter med likartad karaktäristika. Vanliga segmenteringsgrunder är att gruppera utefter demografiska, geografiska, socioekonomiska, psykografiska och beteendemässiga variabler. Inom behovsbaserad segmentering, gruppering på bas av beteendemässiga konsumentvariabler, görs en segmentering utifrån vad konsumenten eftersträvar. Här är det vanligt att använda sig av conjointanalysens nyttovärden, de s.k. part-worths från de olika produktattributen. För denna gruppering av konsumenter används vanligtvis klusteranalys, Hair et al (2005). En fråga när det gäller klusteranalys rör huruvida datamängden bör standardiseras eller inte. En standardisering skulle innebära att variablerna omskalas till att ha en gemensam skalenhet.

1.2 Tidigare forskning

Vid segmentering används oftast klusteranalys som ett statistiskt hjälpmedel för att gruppera individer till olika segment. Innan en segmentering genomförs kan det vara önskvärt att standardisera datamängden, detta för att variablerna ska vara relativt jämförbara med varandra, för att på så sätt uppnå ett resultat som bygger på datamängd med en gemensam skalenhet. En konsekvens i användandet av ostandardiserade värden är att variabler med höga skalenheter kan ha alltför stor inverkan i resultaten. Exempelvis är det Euklidiska avståndsmåttet, som används inom klusteranalys, väldigt känsligt för storheten hos olika skalor, Milligan & Cooper (1988). Forskarna menar å andra sidan att standardisering inte alltid är önskvärd. En problematik är här att vald standardisering kan ge olika klustersammansättningar jämfört med en annan standardiseringsform.

Schaffer & Green (1996) pekar på att den vanliga z-standardiseringsformen, som visar på hur många standardavvikelser ett värde ligger, kan vara problematisk att använda beroende på vilken datamängd som används. Inom conjointanalys skulle en z-standardisering innebära att medelvärdet om värdet 0 och standardavvikelsen 1 skulle erhållas inom respektive attribut. Milligan & Cooper (1988) visar i sin studie att standardisering med

hjälp av att dividera variablerna med sin variationsvidd ger bättre informationsåtervinnande hos den underliggande klusterstrukturen. Schaffer & Greens studie om olika standardiseringsformer med data från en conjointstudie för analys med hjälp av K-means metoden efter MacQueen (1967) visar istället att standardisering med en trimmad variationsvidd, där 5% av de värden med högsta och lägsta värden tas bort, ger bäst resultat vid utvärdering med hjälp av Adjusted Rand Index och Hit Ratio. En standardisering med variationsvidden är här inte lika effektiv, inte heller den vanliga z-standardiseringsformen.

Inom behovsbaserad segmentering tillämpas nyttovärden från en conjointanalys, och praxis och vissa forskningsartiklar förlitar sig på att standardisera nyttovärdena innan man tillämpar klusteranalys. Denna standardiseringsform presenteras av Backhaus et al (2000) och är tänkt att användas för att bilda mer homogena klustergrupper. Däremot skiljer sig denna standardiseringsform något jämfört med presenterade standardiseringar av Schaffer & Green (1996) och Milligan & Cooper (1988). För uppsatsen görs därför en noggrannare utvärdering av klustersammansättningarna med standardiserade och icke-standardiserade nyttovärden.

1.3 Syfte

Uppsatsen har ett explorativt syfte att utforska resultaten från en redan befintlig conjointanalys. Syftet kan ses som tvådelat. Dels är syftet att hitta kundsegment och studera deras preferenser för mest föredragna produktattribut. Dels är syftet att med olika klustermetoder jämföra klustersammansättningarna med standardiserade och icke-standardiserade nyttovärden. Jämförelse kan göras utifrån Adjusted Rand Index och Hit Ratio.

1.4 Disposition

Inledande avsnitt klargjorde uppsatsens bakgrund, tidigare forskning och syfte. Nästkommande avsnitt, avsnitt två beskriver hur conjointanalys är uppbyggd, med koppling till den genomförda conjointanalys datamaterialet består av, samt att avsnittet förklarar hur standardisering av nyttovärden presenterad av Backhaus et al (2000) är utformad. Avsnitt tre beskriver tillhandahållet datamaterialet från GfK och hur denna datamängd har behandlats inför analysen. Klusteranalys med använda klustermetoder och vilka steg som har vidtagits i klusteranalysprocedurerna behandlas i avsnitt fyra. I avsnitt fem presenteras resultaten från klusteranlyserna med standardiserade och icke-standardiserade nyttovärden. Detta avsnitt presenterar också klustersammansättningarna utifrån Adjusted Rand Index och Hit Ratio. Avslutande avsnitt, avsnitt sex, innehåller en diskussion om resultaten och förslag till vidare forskning.

2 Conjointanalys

Conjointanalys är en multivariat analysmetod som studerar vilka preferenser respondenter har för olika typer av produkter, Hair et al (2005). En simpel modellspecifikation för conjointanalys är att den beroende variabeln kan uttryckas i form av metrisk eller icke-metrisk karaktär och att de oberoende variablerna kan uttryckas som icke-metriska. Olika sätt finns för att mäta preferensstrukturen för ett objekt. Conjointanalys är ett sätt och tillhör kategorin decompositional, vilket innebär att preferensstrukturen skattas genom att försökspersonen rangordnar och utvärderar presenterat stiumuli, Hair et al (2005).

Inom coinjointanalys gör respondenterna avvägningar mellan olika karakteristika hos produktattribut från vilken egenskap hos attributet/faktorn de föredrar mest till den de föredrar allra minst för en vara eller tjänst. Metoden bygger således på rangordning. Varje faktor har därefter olika nivåer. Exempelvis kan en produkt utvärderas efter attributen färg och form, och där rött, grönt och blått är nivåer för attributet färg. Respondenten kan utvärdera ett flertal produktattribut. Vilka dessa attribut är och hur många, måste fördefinieras innan studien. Praktiska avgränsningar måste göras i förhållande till vilka attribut som är betydelsefulla för utvärdering, och att begränsat antal attribut kan tas med i studien beroende på vilken metod som tillämpas, SPSS (1997).

2.1 Full-profile

Vid utformning av en conjointanalys finns det flera olika modeller att välja mellan för att skatta preferenserna hos ett objekt. En väldigt vanligt förekommande modell är full-profile variaten, vars datamaterial denna uppsats bygger på. Vald conjointmetod är avgörande för hur många attribut som utvärderas. För full-profile metoden rekommenderas upp till tio olika attribut, Hair et al (2005).

Enligt full-profile metoden får respondenten rangordna samtliga attribut för en given produkt. Ett viktigt antagande görs här att de olika nyttovärdena som beräknas för attributens olika nivåer ska vara ortogonala, vilket innebär att skattningarna för parth-worths är oberoende från varandra. En begränsning finns i metoden att inte alla attribut kan utvärderas, utan endast ett visst utvalt antal attribut och nivåer kan väljas ut och presenteras med fullständiga produktprofiler. En svårighet ligger i att utvärdera ett objekt med flera attribut och många nivåer, en produktprofil med tre attribut om fyra nivåer skulle leda till en skattning av 72 olika möjliga kombinationer. Därför kommer endast ett visst antal utvalda profiler att väljas ut så att en skattning kan beräknas för respektive attributs olika nivåer. Full-profile modellen är endast additiv. En skattning beräknas endast för respektive faktor, dvs att en skattning för huvudeffekter beräknas och att hänsyn inte tas till eventuella interaktioner, Hair et al (2005).

2.2 Nyttovärden

Tanken med conjointanalys är att mäta preferensstrukturen för ett objekt. Denna preferensstruktur antas följa antingen ett linjärt samband, kvadratiskt eller av separata nyttovärden. Ett linjärt samband innebär att en koefficient skattas och därefter multipliceras för varje nivå. Ett kvadratiskt förhållande innebär ingen strikt linjäritet, sambanden kan skifta upp eller ner. Separata part-worths innebär olika skattningar för respektive nivå. Part-worths utgörs av den generella preferensen eller nyttan som associeras till varje nivå

för varje attribut. Den totala nyttan hos ett objekt beräknas genom att addera de olika kombinationerna med respektive attributs nivåer.

2.2.1 Beräkning av nyttovärden

Malhotra & Birks (2003) presenterar en enkel utformning av beräkningarna av conjointanalysmodellen, vilken följer enligt

$$U(X) = \sum_{i=1}^{m} \sum_{j=1}^{k_i} \alpha_{ij} x_{ij} \tag{1}$$

där funktionen $U(X)$ kännetecknas som det totala nyttovärdet för en produkt. Beteckningen α_{ij} är det part-worth värdet eller nyttovärdet som associeras med $j : te$ nivån för $i : te$ attributet. Vidare betecknas x_{ij} som en indikatorvariabel, med ett dikotomt utfall om värdena 0 och 1, vilket kommer att indikera huruvida attributet och nivån finns representerad hos respondenten. Vidare kännetecknas k_i av antalet nivåer för attribut i, och m kännetecknas som antalet attribut. Vikten hos varje attribut, I_i, kan definieras som spannet för de part-worths, α_{ij} , genom antal nivåer för det givna attributet, vilket kan definieras som

$$I_i = \{\max_i(\alpha_{ij}) - \min_j(\alpha_{ij})\} \tag{2}$$

Varje attributs viktighet kommer att normaliseras för att göra resultaten jämförbara med andra attribut. Detta görs genom att dividera varje enskilt attribut med summan av samtliga attribut enligt följande,

$$W_i = \frac{I_i}{\sum_{i=1}^{m} I_i} \tag{3}$$

viktighetsgraden, med summan över antalet nivåer för antalet attribut, ska vidare kunna summeras till 1 enligt

$$\sum_{i=1}^{m} W_i = 1 \tag{4}$$

2.3 Standardisering av nyttovärden

Generellt sätt standardiseras variabler i undersökningar för att göra variablerna mer jämförbara vid analys. En svårighet i en undersökning kan vara att variablerna har olika skalenheter. Problem som kan uppstå är att vissa av variablerna får missvisande resultat, exempelvis att en variabel kan få större inflytande i resultatet i förhållande till en annan variabel. En vidare anledning till standardisering är att göra analysmetodernas resultat mer jämförbara, exempelvis av mätning av preferensstrukturerna och särskilt om olika mätmetoder har använts.

En rad möjliga standardiseringsformer finns för beräkning av metrisk data. En möjlighet för standardisering av nyttovärden beskrivs av Backhaus et al (2000) och kan användas

som ett sätt för att bilda mer homogena grupper vid segmentering. Standardiseringen går till genom att först tilldela värde 0 för det faktorvärde inom respektive attribut som har lägsta part-worth värde. Därefter ska skillnaden mellan det enskilda part-worth värdet α_{ij}, och det lägsta part-worth värdet α_{imin} beräknas. Denna beräkning genomförs för att enbart få positiva nyttovärden. Beteckning för detta steg följer enligt:

$$\alpha_{ij}^{*} = \alpha_{ij} - \alpha_{imin} \tag{5}$$

där $\beta imin$ kan betecknas som

$$\alpha_{imin} = \underset{j}{min}\{\alpha_{ij}\} \tag{6}$$

därefter summeras det nyttovärde som har högst värde inom respektive attribut, β_{ij}^{*} för varje stimuli j, till ett totalt nyttovärde enligt följande

$$\sum_{j=1}^{J} \underset{j}{max}\left(\alpha_{ij}^{*}\right) \tag{7}$$

och den starkaste preferensen ges värdet 1. Slutligen erhålls de standardiserade partiella nyttovärdena $\widehat{\alpha}_{ij}$ genom att dividera föregående steg, steg (5) med steg (7) enligt följande:

$$\widehat{\alpha}_{ij} = \frac{\alpha_{ij}^{*}}{\sum_{j=1}^{J} \underset{j}{max}\left(\alpha_{ij}^{*}\right)} \tag{8}$$

vilket medför en standardisering med ett värde som ligger inom intervallet $[0, 1]$. Det högsta standardiserade part-worth värdet ger oss möjlighet att uttala oss om den relativa viktighetsgraden hos ett attribut. Vanligtvis multipliceras attributets värde med 100 för att tolka siffrorna i procent. En viktig notering är att standardisering innebär att ett likartat resultat uppnås vid beräkning av relativ viktighetsgrad. Likheter finns i beräkningen mellan täljaren i ekvation (3) och ekvation (7).

När resultaten aggregeras görs en sammanslagning av alla individuella resultat. Denna sammanslagning innebär att eventuell information går förlorad i jämförelse med om man endast skulle studera de individuella nyttonivåerna. Om den aggregerade nyttostrukturen är väldigt heterogen medför en aggregering en större informationsförlust. En tanke med att standardisera nyttovärdena är att forma mer homogena grupper, vilket kan genomföras med hjälp av klusteranalys.

3 Datamaterial

Datamaterialet är tillhandahållet av GfK Sverige (Growth from Knowledge) och består av resultat från en conjointanalys beträffande försökspersoners preferenser för juice- och mjölkförpackningar år 2008. Denna uppsats avser att studera preferenser för juiceförpackningar genomförda i Berlin, Tyskland, med totalt 153 försökspersoner. Statistisk programvara som används är SPSS version 17, Stata version 10, MATLAB version 7.4 och R version 2.9.0 med tilläggspaketet mclust.

3.1 Upplägg av GfKs studie

Totalt baseras undersökningen på 600 försökspersoner i Spanien och Tyskland, där 303 försökspersoner fick testa kombinationer av juice och mjölk i Tyskland och 304 försökspersoner i Spanien. I Spanien testades försökspersoner i Barcelona och Madrid. I vardera stad testades omkring 76 försökspersoner gällande juice och mjölk. För Tyskland fick Hamburg testa alternativen för mjölk och Berlin fick testa alternativen för juice. Försökspersonerna valdes ut på bas av ett flertal kriterier. Försökspersoner för mjölk valdes ut proportionellt uppdelat på kön, där 50% var mellan 25 och 45 år och resterande 50% var 46 år och äldre. Försökspersonerna skulle minst en gång i veckan köpa mjölk i kartong i familjepack om 1 liter. Försökspersonerna för juice valdes proportionellt ut utifrån bakgrundvariablerna kön, uppdelat på män och kvinnor, samt ålder, med unga och äldre och där uppdelning gjordes om att 50% av försökspersonerna skulle vara mellan 18 och 29 år och resterande 50% av försökspersonerna skulle vara mellan 30 och 40 år. Ett krav för att delta i undersökningen, beträffande juice, var att försökspersonerna minst en gång i veckan köpte portionsförpackad juice i kartong om 500 milliliter med fruktjuice.

Eftersom inte alla produktkombinationer var möjliga att undersöka med samtliga attributsnivåer, dvs att tre attribut med fyra till fem nivåer skulle ge 5*4*3 = 60 produktkombinationer, valdes 25 produktkombinationer som försökspersonerna fick rangordna från 1 till 25 från, vilket alternativ de mest föredrog till vilket alternativ de minst föredrog.

3.2 Beskrivning av datamaterialet

Conjointanalysen genomfördes i SPSS med tilläggspaketet Conjoint. Designen som användes vid beräkning är full-profile metoden. Datamaterialet består av attributet *öppning* med fem nivåer, kategoriserade som AC, BC, CT, DL och E, *förpackning* med fyra nivåer, kategoriserade som AS, BS, AM och BM, och attributet *pris* med tre linjära prisnivåer. Preferensstrukturen för öppning och förpackning antas följa separata nyttovärden, olika kategorier, medan pris följer ett linjärt samband. För attributet pris skattas en koefficient, vilken multipliceras med respektive nivå.

3.3 Undersökning av datamaterialet

Enligt Hair et al (2005) rekommenderas vid undersökning av datamaterialet att ställningstagande görs beträffande kontroll av saknad data, kontroll av multikolinjäritet, identifikation av outliers, standardisering av datamaterial och att man testar huruvida de underliggande antagandena hos de multivariata metoderna stämmer.

Nyttovärden för fem observationer saknas, varför dessa observationer tas bort ur datamängden. Angående kontroll av multikolinjäritet användes samtliga variabler för öppning, förpackning och pris. Pris med tillhörande tre nivåer är ett attribut med linjära nivåer, där ett värde fås fram och sedan multipliceras med grad av nivå. En bivariat korrelationsplot med hjälp av Spearmans rangkorrelationskoefficient visar, se korrelationstabellen i appendix A, att prisnivåerna är korrelerade. Detta kan medföra problem för resultaten för klustersegmenten då prisnivåerna kan komma att dominera resultaten. Pris med tillhörande nivåer tas därför bort i undersökningen. Inga nämnvärt höga korrelationer har uppmätts mellan variablerna.

Standardisering av datamaterialet har gjorts i enlighet med avsnitt om standardisering i föregående kapitel. Argument för att genomföra en standardisering ligger i att uppnå mer homogena klustergrupper. Utfallet för de standardiserade klustergrupperna kommer att jämföras med de icke-standardiserade klustergrupperna.

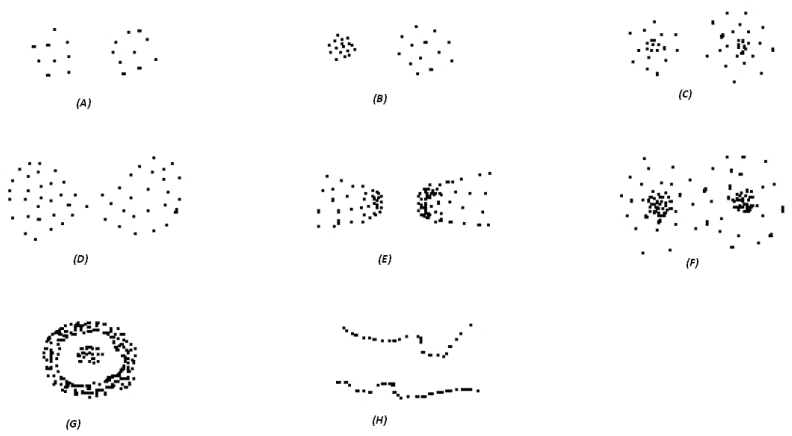
En kontroll för univariata outliers visar på eventuellt två outliers större än $z > 3.29$ om $p < 0.01$. Vid granskning av eventuella multivariat outliers visar Mahalanobis avståndet med $\chi^2(9)$ inga större avvikelser om $p < 0.001$. Inget större antal outliers befaras och eventuella outliers antas inte ha haft alltför stor inverkan för datamaterialet med 148 individer. En problematik med eventuell borttagning av eventuella outliers är att de kan ha inverkan på klustersammansättningarna och borttagning skulle innebära andra klustersammansättningar. Vidare kan outliers i sig utgöra en potentiell gruppering.

4 Klusteranalys

Klusternanalys kan ses som en explorativ teknik för att utforska ett ömsesidigt beroendeförhållande i ett givet datamaterial. Klusteranalys, klustring i allmänhet, kan ses som en metod för att klassificera objekt till grupper eller klasser. Beteckningen klassificeringsmetod ämnar sig dock för andra statistiska metoder exempelvis logistik regression och diskriminantanalys. En viktig distinktion är att vid klustring är grupperna inte fördefinerade, en klassificering sker efter vad som anses som en naturlig gruppering, Fielding (2007). Klusteralgoritmerna är tänkta att gruppera objekt som tillhör varandra nära i avstånd. En klusteranalys gör det därför möjligt att gruppera respondenter med likartade preferenser.

Klassificeringsmetoderna kan vara övervakade/kontrollerade medan klustringsproceduren inte är det. Klusteranalys ses mer som en datorbaserad metod baserad på algoritmer, snarare än en statistisk metod. En svårighet finns att göra en noggrannare precisering av resultaten från en icke övervakad/okontrollerad klassificeringsmetod. Resultaten måste utvärderas i förhållande till dess användbarhet, om resultaten från klusteranalysen är användbara och att det går att tolka resultaten på ett meningsfullt sätt, Fielding (2007).

En grafisk presentation åskådliggörs över olika klustergrupperingar som varierar med olika tätheter och avstånd till varandra, (Zahn 1971 genom Shihab 2000:33ff). Vi kan från (A) se att de två klustren har på ett ungefär likartad täthet. I (B) har de två klustrena olikartade tätheter. Klustrena representerade i (C) har tätheter som varierar proportionellt i förhållande till medelvärdet. I (D) finns en jämn varierad täthet, och separationen uppkommer närmast den punkt med störst täthet för de klasserna. I delfigur (E) och (F) är separationen mellan de två klustrena icke-existerande, då båda klustren rör vid varandra. Problematik finns för klusteralgoritmer att identifiera de två klustergrupperna som man med mänskliga ökat kan åskådliggöra i (G) och (H). I (G) kommer algoritmerna att vara beroende av längden hos varje string samt hur långt de olika typerna av string är ifrån varandra. I (G) visas kluster med en klass som omsluts av en annan. Den innersta grupperingen kan åstadkomma problem för flera algoritmer. Shihab (2000).



Figur 1: *Exempel på kluster*

4.1 Icke-hierarkisk klustring

Vid icke-hierarkisk klustring grupperas ett antal objekt till ett visst antal kluster som inte har någon inbördes relation. En gruppering ska inte gå att överlappa. Icke-hierarkisk klusteranlays kommer inte till skillnad från hierarkisk klusteranalys att reducera eller forma klustergrupper till att slutligen tillhöra en enda klustergrupp. För uppsatsen används de icke-hierarkiska klustermetoder, K-means, uträknat i SPSS version 17, och K-median, uträknat i Stata version 10.

4.1.1 K-means

MacQueens K-means klusteringmetod (1967) används för att gruppera individer till $k$antal partitioner genom villkoret att minimera inomgruppsvarians för klustergrupperna. För uppsatsen grupperas individer utifrån deras preferenser till en gemensam grupp. Proceduren beskrivs som iterativ, där i regel de två förstnämnda stegen upprepas.

1. Dela upp objekten i k st (initiala kluster).
2. Gå igenom varje objekt och tilldela objektet till det kluster vars center, mätt som medelvärde inom klustergruppen, ligger närmast.
3. Upprepa förfarandet i steg 2 till dess att ingen omgruppering av objekt sker.

4.1.2 K-median

K-median är en annan partitionerinsmetod som är snarlik proceduren för beräkning av k-means, men använder sig av medianen som spridningsmått för indelning av kluster. En fördel med att använda medianen vid beräkning av klustercentra är att den är mindre påverkad av outliers, som klustercentra baserade på medelvärde med K-means kan påverkas av, Fielding (2007).

4.2 Fuzzy klustering

Vanligtvis kan utfall betecknas som sanna eller falska. Utfallen kan uttryckas som binära och antar antingen värdet 0 för falskt eller 1 för sant. Fuzzy clustering grundar sig på Fuzzy logic som bygger på principen om att varje observation i varierande grad kan tillhöra ett utfall. Tidigare nämnda klustermetoder bygger på att ett objekt endast kan tillhöra ett kluster och därför kännetecknas som hårda/crisp. Fuzzy clustering bygger således på att observationer med varierande grad kan tillhöra olika kluster inom intervallet $[0, 1]$. Användning av Fuzzy c-means metoden vid klustring av data från en conjointstudie innebär att varje respondents nyttovärde för en nivå hos ett attribut kännetecknas som en datapunkt som kan tillhöra samtliga klustergrupper av olika tillhörighetsgrad. För att göra klusterresultaten från Fuzzy klustering jämförbara med tidigare klusterresultat från K-mean och K-median har varje observation placerats till endast en klustergrupp utifrån högsta tillhörighetsgrad. Beräkning för fuzzy clustering baseras på Matlabs, version 7.4.0, inbyggda Fuzzy c-means clustering algoritm (FCM) som är baserad på Bezdec (1981), MATLAB (2008).

4.3 Adjusted Rand Index

Adjusted Rand Index är ett mått som kan användas för att kolla huruvida två klusterpartitioner tillhör en och samma klustergruppsindelning. Schaffer & Green (1996) använder måttet för att jämföra utfallet av klusterresultaten med och utan standardisering.

Adjusted Rand Index kan ses som ett mått på överensstämmelse huruvida två partitioner har samma klustergruppsindelning, dvs om de två partitionerna kan sägas tillhöra en och samma partition. De två partitionerna skulle kännetecknas av datamängd med standardiserade och ostandardiserade nyttovärden för de olika klustermetoderna. Vi kan anta ha en total datamängd med tillhörande objekt, betecknade som $S = \{O_1, ..., O_n\}$. Den totala datamängden S innehar två klusterparitioner U och V, med standardiserad och ostandardiserad datamängd för respektive klustermetod.

Vi låter nu a beteckna antalet par av objekt som kan placeras i U och V. Låt b vara antalet par hos objekten som tillhör U men inte i samma kluster i V, c är antalet par hos objekten som tillhör V men inte i samma kluster i U. Slutligen låt d vara antalet par av objekten i olika kluster hos klusterpartitionerna. Kvantiteterna a och d kan uttryckas som antalet par som stämmer överens hos klustertilldelningarna, och c och d som par som inte stämmer överens för klustertilldelningarna, Yeung & Ruzzo (2001). Hubert & Arabie (1985) menade att föregående Rand index inte korrigerar för sannolikheten som är lika med noll för slumpmässiga partitioner som har samma antal objekt inom varje klass. Beräkning av Adjusted Rand Index följer enligt

$$\frac{2(ab - cd)}{((a + d)(d + b) + (a + b)(c + b))} \tag{9}$$

och ligger inom $[0, 1]$ där ett värde om 0 visar på en samstämmighet som skulle uppkomma genom slump, ett värde om 1 kännetecknas av att partitionerna har en hög grad av samstämmighet. Ett negativt värde kan uppkomma och indikerar att det inte finns en samstämmighet.

4.4 Hit Ratio

Hit Ratio är andelen objekt som korrekt klassificeras inom diskriminantanalys och används som ett sätt för att validera resultaten inom diskriminantanalys. Inom diskriminantanalys önskar man klassificera objekt till en tillhörande grupp. Vilken grupp objekten tillhör beror på deras erhållna diskriminantvärden och genom en lämplig beslutsregel. En lämplig klassificeringsgrad är att objekten klassificeras till en andel som är 25% högre än vad som erhålls av slumpen, Malhotra & Birks (2003). Utifrån tillämpning av Hit Ratio skulle vi kunna uttala oss om huruvida resultaten ifrån en klusteranalys ger möjlighet till rätt klassificering.

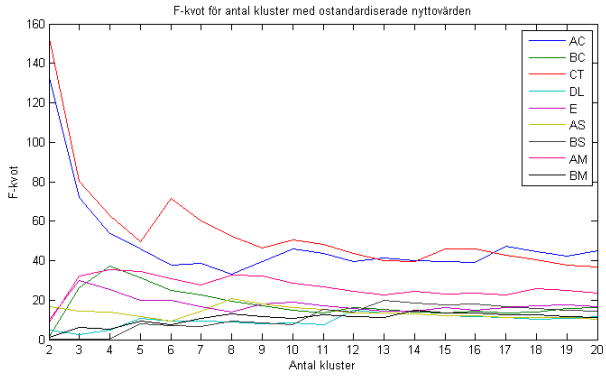
4.5 Antal kluster

En problematik med icke-hierarkiska klustermetoder är att välja en startpunkt för antal kluster som ska ingå i analysen. En intuitiv utgångspunkt är att välja antal kluster som är rimligt att forma utifrån datamängdens storlek, eller att det är förutbestämt hur många kundsegment som man önskar segmentera på bas av. Ett sätt att välja antal kluster är att studera F-kvoten, och välja antal kluster som maximerar mellangruppsvariansen i förhållande till inomgruppsvariansen, Johnson & Wichern (2007). F-kvoten definieras enligt

$$F = \frac{MST}{MSE} \tag{10}$$

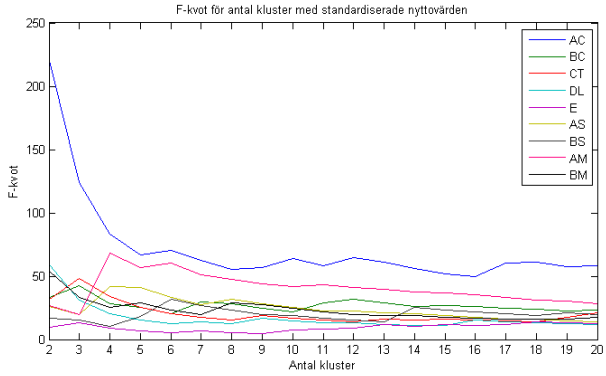
med $MST = \frac{SST-SSE}{a-1}$ som kännetecknas som mellangrupsvariationen, och $MSE = \frac{SSE}{N-a}$ som kännetecknas som inomgruppsvariationen med N totalt antal observationer och a antal parametrar som skattas i modellen.

Genom att studera F-kvoten kan man välja ut ett lämpligt antal kluster, vilket kan göras utefter det så kallade armbågskriteriet. Ett lämpligt antal kluster kan väljas där linjen för F-kvoten skärs av innan den börjar plana ut. Beräkning av F-kvoten genomförs genom beräkning av en envägs ANOVA i SPSS för de olika nivåerna för attributen öppning och förpackning. En grafisk representation för de ostandardiserade nyttovärdena visas i figuren nedan. Fem kluster kommer härefter att väljas. ANOVA tablån för färre en fem kluster visade sig ha flera icke signifikanta resultat för flertalet av nivåerna för *öppning* och *förpackning*. Först efter fem kluster uppvisade sig de olika nivåerna ha signifikanta F-kvoter.



Figur 2: *F-kvot för antal kluster med ostandardiserade nyttovärden*

F-kvoten för de standardiserade nyttovärdena presenteras i nedanstående figur. Figur 3 uppvisar liknande avtagande mönster för F-kvoten som för figur 2 ovan med ostandardiserade nyttovärden. Antal kluster för de standardiserade nyttovärderna väljs till fem kluster. Fem kluster väljs för att möjliggöra eventuell jämförelse med ostandardiserade nyttovärden med fem kluster. Samtliga F-kvoter visade sig vara signifikanta för antal valda klusterkörningar.

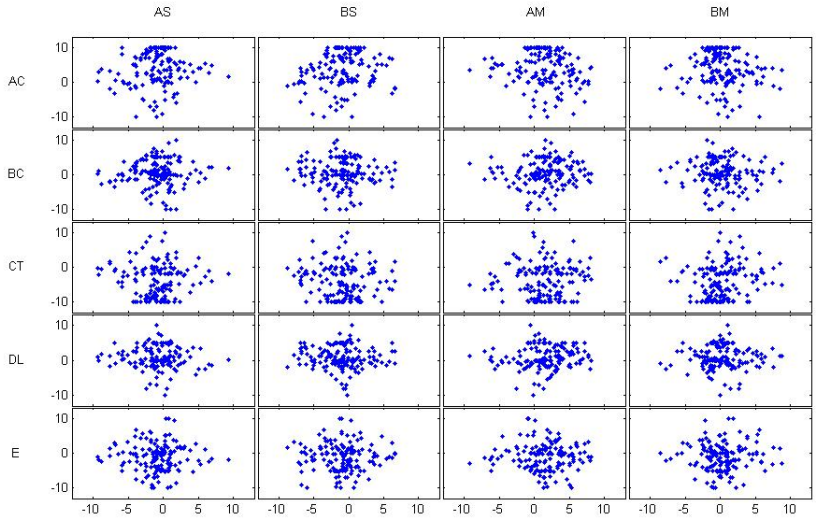


Figur 3: *F-kvot för antal kluster med standardiserade nyttovärden*

5 Resultat

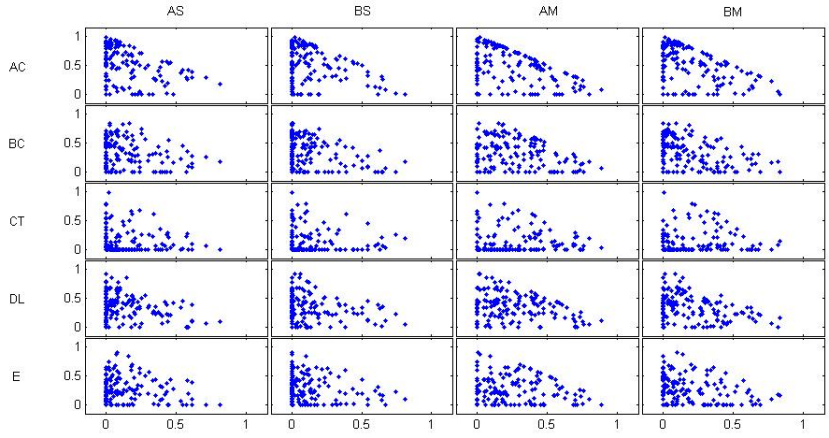
5.1 Ostandardiserade och standardiserade nyttovärden

Vid ostandardiserade nyttovärden plottar vi de parvisa nivåerna för attributen *förpackning* (horistontell axel) mot *öppning* (vertikal axel). Attributet *öppning* har nivåerna AC, BC, CT, DL och E, och attributet *förpackning* har nivåerna AS, BS, AM och BM. Enligt figuren nedan kan det vara svårt att exakt se något tydligt mönster då nyttovärdena är relativt spridda.



Figur 4: *Ostandardiserade nyttovärden för parvisa nivåer för öppning och förpackning*

Vi går vidare med att plotta de standardiserade nyttovärdena för nivåerna för respektive variabel *öppning* (vertikal axel) mot *förpackning* (horistonell axel). Ett eventuellt tydligt mönster går att tyda utifrån standardisering, då ett flertal av figurerna visar en centrering antingen i övre vänstra eller nedre vänstra delen av figurerna. Det skulle därför kunna antydas att respondenterna visar starkare preferenser för *öppning* än för *förpackning*.



Figur 5: *Standardiserade nyttovärden för parvisa nivåer för öppning och förpackning*

5.2 Klustersammansättningar

Inledningsvis redogörs för en deskriptiv presentation av datamängden, genom att endast summera de produktattribut som flest antal respondenter föredrar, benämnt som top choice i tabell 5. Genom att summera de mest föredragna produktattributen, kommer flest antal respondenter att föredra alternativen AC (öppning) och AM (förpackning).

Vidare presenteras preferensstrukturerna för ostandardiserade nyttovärden med hjälp av olika klustermetoder. Tabell 6 med K-means klustermetod med ostandardiserade nyttovärden visar att det alternativ som föredras mest bland öppningsalternativen är AC , segment 3 och 5, och för förpackningalternativ så är det AM, segment 1, 2 och 5. De andra öppningstyperna som föredras är BC, enligt segment 1, samt DL, som segment 2 föredrar, och CT, enligt segment 4. Segment 1, 3 och 5 föredrar valt öppningsalternativ högre än vald typ av förpackning. Vice versa gäller för segment 3 och 4. Tabell 7 med K-median klustermetod med ostandardiserade nyttovärden visar att samtliga segment föredrar AC (öppning), framförallt segment 3. Den mest föredragna förpackningstypen är BS, för segment 1, 3 och 4. Segment 2 och 5 föredrar AM. Samtliga segment uppvisar högre nyttovärden för öppning än för förpackning. Utifrån tabell 8 med Fuzzy c-means med ostandardiserade nyttovärden kan vi se att samtliga segment föredrar AC (öppning) och AM (förpackning). De flesta segment föredrar öppning före förpackning, däremot har segment 1 och 5 relativt låga preferenser för öppning.

Vidare studeras preferensstrukturen med standardiserade nyttovärden. I tabell 9 med K-mean och standardiserade nyttovärden, föredrar samtliga segment, förutom segment 2, alternativet AC (öppning). Segment 2 föredrar öppningsformen BC. När det gäller typ av förpackningsform, föredrar samtliga segment, förutom segment 3, förpackningsformen AM. Segment 3 föredrar BS. Segment 3 och 5 föredrar vald förpackningsform mer än sitt valda öppningsalternativ. För tabell 10 med K-median och standardiserade nyttovärden föredrar samtliga segment, förutom segment 1, öppningsalternativet AC. Segment 1 föredrar BC. Segment 1,2 och 4 föredrar förpackningsalternativet AM. Segment 3 föredrar BM, och segment 5 föredrar BS.

Dessa två sistnämnda segmenten föredrar valt förpackningsalternativ högre än valt öppningsalternativ. För tabell 11 med Fuzzy c-means klustermetod och med standardiserade nyttovärden föredrar samtliga segment alternativen AC (öppning) och AM (förpackning).

Vi kan för de valda klustermetoderna för dels ostandardiserade nyttovärden, tabell 6 till 8, och dels standardiserade nyttovärden, tabell 9 till 12, se att de mest förekommande produktattributen är AC (öppning) och AM (förpackning). Nyttovärdena för öppning verkar vara högre än för förpackning, vilket indikerar på att respondenterna föredrar öppningsalternativen. En viss variation finns mellan segmenten, vilket kan indikera på att vissa segment föredrar förpackning mer än andra som kan uppvisa negativa preferenser eller väldigt låga värden, vilket kan indikera att respondenterna inte alls föredrar förpackningsalternativet eller är näst intill indifferenta.

5.3 Adjusted Rand Index

Resultat för Hubert & Arabie Adjusted Rand Index för de olika klustermetoderna redovisas i tabell 1. Notera att ett värde om noll indikerar på en samstämmighet som skulle uppkomma genom slump och ett negativt värde indikerar att det inte finns en samstämmighet. Tabellvärdena visar att bäst överensstämmelse mellan standardiserade och ostandardiserade klustersammansättningar hade Fuzzy c-means metoden, och sämst hade K-means. En jämförelse klustermetoderna sinsemellan visar att K-mean med standardiserade värden skiljer sig rent klassificeringsmässigt från K-median och Fuzzy c-mean med ostandardiserade värden. K-median och Fuzzy c-means uppvisar en mild samstämmighet för både standardiserade och ostandardiserade nyttovärden.

Tabell 1: *Adjusted Rand Index jämfört mellan datamängder för klustermetoder*

		Standardiserat		
		K-means	K-median	Fuzzy c-means
Ostandardiserat	K-means	0,219	-0,007	0,012
	K-median	-0,022	0,244	0,262
	Fuzzy c-means	-0,016	0,364	0,316

Tabell 2 nedan jämför de ostandardiserade och standardiserade klusterutfallens samstämmighet inbördes inom respektive datamängd med ostandardiserade och standardiserade nyttovärden. Från tabell 2 går det att utläsa att Fuzzy c-means och K-median har en relativt hög samstämmighet vid klassificeringarna jämfört med K-mean. Vi kan från nedanstående tabell se att standardiserad datamängd inte medför en väsentligt högre andel överensstämmande grupperingar för de olika klustermetoderna.

Tabell 2: *Adjusted Rand Index för inbördes jämförelse av datamängd*

	Ostandardiserat	Standardiserat
K-means och K-median	-0,016	-0,006
K-mean och Fuzzy c-means	-0,006	-0,002
Fuzzy c-mean och K-median	0,413	0,463

Vad som kan noteras när det gäller Adjusted Rand Index är att det är ett mått som studerar huruvida två partitioner har gemensamma klustergrupperingar. Vad dessa klus-

tergrupperingarär utgörs av måste bedömmas för respektive partition för sig och sedan jämföras med den andra partitionen. Tabell 7 och 8 visar för ostandardiserade nyttovärden för K-median och Fuzzy c-means att samtliga segment föredrar öppningsalternativet AC. Däremot skiljer sig något segmenten i preferenser om förpackningstyp, men flertalet av segmenten föredrar alternativet AM. Likaså visar tabell 10 och 11 för standardiserade nyttovärden för K-median och Fuzzy c-means att flertalet av segmenten föredrar öppningsalternativet AC och förpackningsalternativet AM.

5.4 Hit Ratio

Beräkning av Hit Ratio i nedanstående tabell visar att K-means har högst korrekt andel klassificerade objekt till respektive kluster. K-median och och Fuzzy c-means uppvisar en låg andel korrekt klassificerade objekt. Vid beräkning av Hit Ratio med standardiserade nyttovärden beräknas Hit Ratio vara något högre än med ostandardiserade nyttovärden. Dock visar både K-median och Fuzzy c-means för ostandardiserade och standardiserade nyttovärden en sämre klassificering än vad som skulle ha uppkommit genom slump, vilken genom Malhotra & Birks (2003) tumregel för fem grupper beräknas till $\frac{1}{5} + 25\% = 45\%$.

Tabell 3: *Hit Ratio*

	Ostandardiserat		Standardiserat	
K-mean	90,54%	(134/148)	93,92%	(139/148)
K-median	23,65%	(35/148)	35,81%	(53/148)
Fuzzy c-means	35,81%	(53/148)	36,49%	(54/148)

5.5 Utvärdering av klustergrupper

Efter att framtagna resultat har rapporterats är det senare viktigt att kontrollera dessa. Ovanstående Adjusted Rand Index och Hit Ratio kan ses fungera som valideringsmått för klustersammansättningarna. Ytterligare formella krav för utvärdering av klustergrupper anser Malhotra & Birks som komplexa och inte helt utan kritik. Exempel på mått är att beräkna homogeniteten inom grupper och heterogeniteten mellan grupper, Klastorin (1983). Malhotra och Birks (2003) ger sina förslag på kriterier för utvärdering av klustresultat och för uppsatsen utvärderas de tre sistnämnda kriterierna som presenteras enligt nedan:

1. Använda olika avståndsmått för datamängden och jämföra resultaten för att se hur stabila resultaten är.

2. Använda olika klustermetoder, sequential threshold, parallel threshold eller optimising partitioning.

3. Dela datamängden i olika halvor och genomföra klusteranalys på de olika datamängderna och sedan.jämföra centroiderna.

4. Slumpmässigt ta bort variabler och genomföra klusteranalys på reducerad variabelmängd och jämföra med alla variabler.

5. Inom icke-hierarkisk klusteranalys kan lösningen bero på i vilken ordning observerationerna kommer. Genomför därför flera körningar med olika positioner för observationerna tills resultaten är stabila.

5.5.1 Uppdelning av datamaterialet

För det tredje kriteriet delades datamängden upp i två separata mängder om 74 individer i vardera del. Resultatet finns sammanställt i tabell 12-17 i appendix. Resultaten för de ostandardiserade och standardiserade nyttovärdena utgörs till större delen av kombinationen AC och AM. Denna kombination var ännu mer utpekande för de standardiserade nyttovärdena. Den första halvan av datamaterialet verkar uppvisa större variationer än den andra halvan, varför den första halvan av respondenter uppvisar större variation i preferenser. Mest spridda resultat verkar K-means metoden uppvisa jämfört med K-median och framförallt Fuzzy c-means.

5.5.2 Borttagning av variabler

För det fjärde kriteriet togs dels variablerna CT och BM bort, och också AC och AM togs bort. En ny standardisering gjordes för att erhålla standardiserade nyttovärden inom intervallet $[0, 1]$. Resultatet finns sammanställt i tabell 18-23. Vid borttagning av CT och BM visar resultaten att AC med AM är den mest förekommande produktsammansättningen av öppning och förpackning. Detta gäller även för ostandardiserade och standardiserade nyttovärden. För Fuzzy c-means visar AC och AM vara det enda egentliga segmentet. En större variation i sammansättningar finns det för K-median och ännu mer för K-means.

Vid borttagning av AC och AM är det svårare att uttala sig, utifrån klustermetoderna K-means och K-median, om vilket det mest föredragna öppningsalternativet är. Fuzzy c-means med ostandardiserade nyttovärden pekar på att DL är det mest föredragna öppningsalternativet, medan med standardiserad datamängd visar Fuzzy c-means att öppningsalternativet DL föredras. För förpackningslaternativen med både ostandardiserade och standardiserade nyttovärden visar K-means och K-median att BS och BM är de alternativ som föredras mest. Fuzzy c-means visar att BM är det förpackningsalternativ segmenten föredrar mest, och detta gäller både med ostandardiserade och standardiserade nyttovärden.

5.5.3 Randomisering av observationer

För det femte kriteriet gjordes en randomisering för samtliga observationer. Resultatet finns sammanställt i tabell 24-29. Resultaten indikerar även här att AC och AM är den mest förekommande produktkombinationen, både för ostandardiserade och standardiserade nyttovärden för de olika klustermetoderna.

6 Diskussion

Syftet med denna uppsats kan ses som tvådelat. I den första delen gällde det att hitta kundsegment och studera deras preferenser för de mest föredragna produktattributen för förpackning och öppning. Resultaten med K-means, K-median och Fuzzy c-means indikerar att kombinationen AC (öppning) och AM (förpackning) utgörs av ett segment. Detta resultat kan ses som en fördel då metoderna uppvisar sammanhängande resultat, men nackdelen är att det verkar vara det enda riktigt säkerställda segmentet, trots att fem kluster har valts. Vid endast en enkel additiv beräkning av det mest föredragna alternativet för öppnings- och förpackningstyp visar det sig att AC och AM är det mest föredragna alternativet. Variationer bland klustersammansättningarna tyder på att preferenserna bland kundsegmenten varierar. Vissa segment uppvisar ett högre nyttovärde för öppning, än förpackning, vilket antyder att typ av öppning föredras mer än typ av förpackning.

Resultaten var samstämmiga både för ostandardiserade och standardiserade nyttovärden. Vid validering av resultaten visade det sig också att AC och AM var de mest föredragna produktalternativen. Vid borttagning av AC och AM var resultaten däremot svårare att tolka om vilka produktsammasättningar som föredras mest. Förutom alternativet AC går det inte att klargöra vilket det andra öppningsalternativet skulle vara som föredras. Bland förpackning antyder resultaten i tabell 18-23 att BM skulle vara det mest föredragna alternativet efter att AM har tagits bort ur datamängden.

I den andra delen av uppsatsen gällde det att med olika klustermetoder jämföra klustersammansättningarna för standardiserade och icke-standardiserade nyttovärden, dels med Adjusted Rand Index och dels med Hit Ratio. Validering med Adjusted Rand Index och Hit Ratio ger olika resultat. Resultaten för Adjusted Rand Index ger intrycket att K-median och Fuzzy c-means visar på mer sammanhängande resultat, både för standardiserade och ostandardiserade nyttovärden. Hit Ratio visar däremot att K-mean metoden har en hög andel korrekt klassificerade, både för standardiserade och ostandardiserade nyttovärden. En oklarhet finns därför vilken klustermetod som ger mest samstämmiga resultat. Standardisering av datamaterialet visar sig, enligt resultaten för Hit Ratio i tabell 3 inte utgöra en väsentligt bättre klassificeringsgrad för de olika klustermetoderna. För K-median blev klassificeringsgraden ca 12 procentenheter högre med standardiserad datamaterial till 35,81%, som dock är lägre än bestämd tilldelning om 45% enligt fastställd tumregel av Malhotra & Birks (2003).

Eventuell problematik för uppsatsen var åtkomsten och förståelsen gällande informationsmaterialet om standardiseringsformen presenterad av Backhaus et al (2000) och tillhörande referenserlitteratur. Litteraturen och referenslitteraturen var svårtillgängliga genom databaser och via lån, och dessutom fanns begränsning att tillgodogöra sig innehållet då informationen presenterades på tyska.

För att göra resultaten mer användbara rekommenderas att klustergruppernas preferensstruktur studeras ytterligare samt att en bakgrundsprofil bildas för respektive klustergrupp/kundsegment. Resultaten inom klusteranalys måste vara användbara och kunna tolkas på ett meningsfullt sätt. Även om produktkombinationen AC och AM dominerar klusterresultaten för de olika klustermetoderna behöver det inte utesluta andra potentiella kundsegment, som föredrar öppningsalternativen BC eller DL, kan finnas representerade. En möjlighet att finna segmenten meningsfulla är att skapa en bakgrundsprofil för de olika klustergrupperna. Detta kan göras genom att hitta skillnader i form av demo-

grafiska, socioekonomiska och geografiska bakgrundsvariabler. Genom att hitta distinka kundsegment gör man det möjligt att utforma lämpliga marknadsstrategier för att nå dessa kundsegment.

Förslag om vidare forskning vore att jämföra standardiseringteknik presenterad av Backhaus et al (2000) med standardiseringsalternativ använda av exempelvis Schaffer & Green (1996) och Milligan & Cooper (1988). Ytterligare former av valideringsformer inom klusteranalys av standardiserade nyttovärden jämfört med ickestandardiserade nyttvärden är också att beakta. En form av validering vore att inkludera holdout, en delmängd som vals ut på förhand och som slutligen används för att validera huvudresultaten. Studien begränsades med för få antal observationer för att en validering av klustersammansättningarna med hjälp av en holdout skulle vara meningsfullt att genomföra. En kommande undersökning måste därför ta hänsyn till att ha en tillräckligt stor datamängd. Detta kan dock vara ett praktiskt bekymmer vid genomförande av conjointanalys eftersom det är en kostsam datainsamlingsmetod.

Referenser

[1] Backhaus K, Erichson B, Plinke W & Weiber R, (2000) *Multivariate Analysemethoden: Eine anwendungsorientierte Einfuhrung*, Springer, 9e upplagan, Berlin.

[2] Bezdec J C, (1981), *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, 1a upplagan, New York.

[3] Douglas J D & Green P E (1995), "Psychometric Methods in Marketing Research: Part I, Conjoint Analysis", *Journal of Marketing Research*, Guest Editorial, 32, pp 385-391.

[4] Fielding A H, (2007), *Cluster and Classification Techniques for the Bioscienes*, 1a upplagan, Cambridge University Press, Cambridge.

[5] Green P E, Rao V R, (1971), "Conjoint measurement for quantifying judgmental data", *Journal of Marketing Research*, 8, pp 355-363.

[6] Hair J F, Black B, Babin B, Anderson R E & Tatham R L, (2005), *Multivariate Data Analysis*, 6e upplagan, Prentice Hall, New Jersey.

[7] Hubert L & Arabie P, (1985), "Comparing partitions", *Journal of Classification,* 2(1), pp 193-218.

[8] Johnson R A & Wichern D W, (2007), *Applied Multivariate Statistical Analysis*, 6e upplagan, Prentice Hall, New Jersey.

[9] Klastorin T D, (1983), "Assessing Cluster Analysis Results", *Journal of Marketing Research*, 20, pp 92-98.

[10] MacQueen J B, (1967) "Some Methods for classification and Analysis of Multivariate Observations", *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, 1, pp 281-297.

[11] Malhotra N K & Birks DF, (2003), *Marketing Research*, 2a upplagan, Prentice Hall, Harlow.

[12] MATLAB, (2008), *Fuzzy Logic Toolbox™ 2*, The MathWorks Inc.

[13] Milligan G W & Cooper M C, (1988), "A Study of Standardization of Variables in Cluster Analysis", *Journal of Classification*, 5, pp 181-204.

[14] Schaffer C M & Green P E, (1996), "An Empirical Comparison of Variable Standardization Methods in Cluster Analysis", 31, pp 149-169.

[15] Shihab A I, (2000) *Fuzzy clustering algorithms and their application to medical image analysis*, Dissertation PhD, University of London.

[16] SPSS, (1997), *Conjoint™ 8.0*, SPSS Inc.

[17] Yeung K Y, & Ruzzo W L, (2001) "Principal component analysis for clustering gene expression data", *Bioinformatics*, 17(9), pp 763-774.

[18] Zahn C T, (1971), "Graph-Theoretical Methods for Detecting and Describing Gestalt Clusters". *IEEE Transactions on Computers*, C-20(1), pp 68–86.

A Korrelationstabell

Tabell 4: *Spearmans rangordnings korrelationskoefficient för samtliga variabler*

	AC	BC	CT	DL	E	AS	BS	AM	BM	Pris1	Pris2	Pris3
AC	1,000											
BC	0,044	1,000										
CT	-0,661**	-0,190*	1,000									
DL	-0,007	-0,168*	-0,359**	1,000								
E	-0,452**	-0,522**	0,141	-0,156	1,000							
AS	0,061	0,009	-0,054	-0,042	0,004	1,000						
BS	0,192*	-0,043	-0,096	-0,104	-0,029	-0,304**	1,000					
AM	-0,172*	0,099	0,033	0,138	0,000	-0,486**	-0,238*	1,000				
BM	-0,137	-0,070	0,170*	0,029	0,002	-0,095	-0,458**	-0,347**	1,000			
Pris1	0,114	0,084	-0,106	0,134	-0,062	0,013	-0,206	0,119	0,039	1,000		
Pris2	0,114	0,084	-0,106	0,134	-0,062	0,013	-0,206	0,119	0,039	1,000**	1,000	
Pris3	0,114	0,084	-0,106	0,134	-0,062	0,013	-0,206	0,119	0,039	1,000**	1,000**	1,000

*Not * $p<0.05$ sign, ** $p<0.01$ sign*

B Tabellförteckning

Tabell 5: *Top choice av mest föredragna alternativ*

Attribut	Antal	Kombination	Antal
AC	87	ACAM	41
BC	19	ACBM	19
CT	11	ACBS	14
DL	13	ACAS	11
E	21	EAM	11
AS	22	BCAS	7
BS	22	DLAM	7
AM	72	EBM	7
BM	37	BCBM	5
		BCAM	5
		CTBM	4
		CTAM	3
		DLBS	3
		CTBS	2
		CTAS	2
		DLAS	2
		EAS	2
		BCBS	1
		DLBM	1
		ACBM/CTBS	1

Tabell 6: *K-means med ostandardiserade nyttovärden*

Variabler	K-means				
Segment	1	2	3	4	5
AC	2,59	0,76	**2,88**	**-**4,98	**6,99**
BC	**3,12**	0,16	1,80	0,85	-0,50
CT	1,04	-2,06	-3,33	**1,85**	-6,97
DL	-1,54	**1,07**	0,52	0,23	1,40
E	-5,20	0,10	-1,88	2,06	-0,91
AS	-0,75	-2,99	**2,67**	-2,44	-1,16
BS	-1,88	2,28	-1,47	-3,43	-1,13
AM	**2,19**	**3,44**	-2,00	2,83	**1,82**
BM	0,44	-2,73	0,80	**3,03**	0,47

Tabell 7: *K-median med ostandardiserade nyttovärden*

Variabler	K-median				
Segment	1	2	3	4	5
AC	**2,49**	**6,75**	**9,50**	**5,40**	**6,20**
BC	0,33	3,78	4,94	4,89	3,26
CT	0,13	0,01	-0,25	4,06	-0,43
DL	-0,98	-7,71	-4,46	-6,18	-2,42
E	-1,96	-2,84	-9,74	-8,18	-6,61
AS	0,09	-0,70	0,04	-0,36	-0,26
BS	**0,79**	0,16	**0,10**	**0,58**	-0,26
AM	-0,09	**0,53**	-0,36	-0,45	**0,60**
BM	-0,80	0,00	-0,22	0,23	0,23

Tabell 8: *Fuzzy c-means med ostandardiserade nyttovärden*

Variabler	Fuzzy c-means				
Segment	1	2	3	4	5
AC	**1,69**	**5,18**	**1,69**	**1,71**	**8,37**
BC	-0,01	-0,10	-0,01	-0,01	2.89
CT	-1,93	-5,30	-1,93	-1,95	-8,06
DL	0,46	1,15	0,46	0,46	0,83
E	-0,20	-0,93	-0,20	-0,21	-4,03
AS	-0,85	-0,75	-0,85	-0,85	-0,65
BS	-1,25	-0,78	-1,25	-1,25	-0,47
AM	**1,56**	**1,39**	**1,56**	**1,56**	**1,40**
BM	0,54	0,14	0,54	0,54	-0,28

Tabell 9: *K-mean med standardiserade nyttovärden*

Variabler	K-means				
Segment	1	2	3	4	5
AC	**0,81**	0,27	**0,30**	**0,53**	**0,25**
BC	0,48	**0,50**	0,16	0,16	0,17
CT	0,04	0,19	0,12	0,30	0,08
DL	0,47	0,47	0,25	0,22	0,18
E	0,25	0,40	0,13	0,42	0,12
AS	0,05	0,13	0,31	0,12	0,25
BS	0,07	0,08	**0,45**	0,10	0,13
AM	**0,16**	**0,31**	0,20	**0,21**	**0,55**
BM	0,06	0,16	0,28	0,18	0,48

Tabell 10: *K-median med standardiserade nyttovärden*

Variabler	K-median				
Segment	1	2	3	4	5
AC	0,15	**0,66**	**0,30**	**0,79**	**0,31**
BC	**0,51**	0,21	0,20	0,54	0,12
CT	049	0,06	0,14	0,05	0,11
DL	0,31	0,48	0,20	0,45	0,22
E	0,40	0,53	0,13	0,21	0,20
AS	0,08	0,12	0,32	0,05	0,14
BS	0,06	0,06	0,08	0,08	**0,49**
AM	**0,21**	**0,20**	0,37	**0,16**	0,44
BM	0,18	0,13	**0,46**	0,07	0,15

Tabell 11: *Fuzzy c-means med standardiserade nyttovärden*

Variabler	Fuzzy c-means				
Segment	1	2	3	4	5
AC	**0,45**	**0,37**	**0,80**	**0,44**	**0,37**
BC	0,29	0,26	0,52	0,28	0,26
CT	0,15	0,16	0,03	0,15	0,16
DL	0,32	0,29	0,47	0,32	0,29
E	0,28	0,26	0,25	0,28	0,25
AS	0,16	0,19	0,06	0,17	0,19
BS	0,15	0,17	0,06	0,15	0,17
AM	**0,29**	**0,33**	**0,15**	**0,29**	**0,33**
BM	0,22	0,27	0,07	0,23	0,27

Tabell 12: *Split half för k-mean med ostandardiserade nyttovärden*

	Datamängd 1					Datamängd 2				
Variabler	K-means					K-means				
Segment	1	2	3	4	5	1	2	3	4	5
AC	1,08	0,39	**2,60**	-4,38	**5,43**	0,75	**6,58**	**4,25**	**8,96**	**8,15**
BC	2,20	-0,67	0,45	0,52	-2,49	-0,72	4,65	2,82	2,48	-3,49
CT	**2,50**	-0,84	-1,82	1,77	-6,34	-1,97	-3,75	-8,27	-8,14	-8,64
DL	-2,08	**1,33**	0,67	0,14	1,31	0,65	0,45	2,20	0,36	2,51
E	-3,70	-0,20	-1,91	**1,95**	2,09	**1,28**	-7,93	-1,00	-3,66	1,48
AS	-0,36	-4,72	**2,89**	-3,20	-0,97	0,43	-2,17	-0,20	-1,11	-0,94
BS	-1,64	2,14	-2,19	-3,76	-0,33	-2,03	-1,26	-1,74	0,72	-1,35
AM	**1,69**	**4,44**	-1,71	3,42	-0,27	**1,42**	**3,64**	**3,83**	**0,81**	**1,59**
BM	0,31	-1,86	1,01	**3,54**	**1,58**	0,19	-0,21	-1,89	-0,41	0,71

Tabell 13: *Split half för k-mean med standardiserade nyttovärden*

	Datamängd 1					Datamängd 2				
Variabler	K-means					K-means				
Segment	1	2	3	4	5	1	2	3	4	5
AC	**0,71**	0,16	**0,26**	**0,54**	**0,23**	**0,61**	**0,80**	**0,76**	**0,48**	0,14
BC	0,48	**0,54**	0,14	0,11	0,17	0,12	0,40	0,57	0,22	0,29
CT	0,26	0,29	0,14	0,42	0,08	0,02	0,00	0,03	0,05	0,17
DL	0,23	0,50	0,24	0,32	0,17	0,35	0,67	0,38	0,24	**0,38**
E	0,15	0,40	0,12	0,47	0,11	0,49	0,29	0,27	0,23	0,24
AS	0,04	0,10	0,35	0,11	0,26	0,20	0,03	0,07	0,09	0,29
BS	0,08	0,06	**0,45**	0,07	0,15	0,03	0,06	0,09	0,27	0,00
AM	0,10	**0,26**	0,21	**0,20**	**0,57**	0,19	**0,13**	**0,19**	**0,42**	**0,48**
BM	**0,23**	0,17	0,30	0,11	0,51	**0,21**	0,08	0,06	0,12	0,42

Tabell 14: *Split half för k-median med ostandardiserade nyttovärden*

	Datamängd 1					Datamängd 2				
Variabler	K-median					K-median				
Segment	1	2	3	4	5	1	2	3	4	5
AC	**9,17**	-0,31	-0,71	**5,20**	**7,30**	**9,12**	**0,95**	0,06	**7,14**	**7,04**
BC	3,60	1,80	1,22	1,80	-2,90	4,64	-0,41	-0,17	0,98	-4,27
CT	-8,82	-5,05	**1,89**	-6,11	-6,93	-7,08	-0,88	-0,74	-7,40	-6,36
DL	0,78	**3,03**	-0,61	0,29	1,63	-1,32	-0,36	**0,69**	3,74	1,76
E	-4,73	0,53	-1,79	-1,17	0,90	-5,36	0,71	0,17	-4,46	1,83
AS	0,08	0,26	0,00	0,86	-0,02	-1,48	-1,33	-6,91	-3,23	-0,07
BS	-1,04	-3,12	-2,93	**3,29**	-1,60	0,67	-1,47	4,05	1,12	-1,08
AM	**1,15**	**3,87**	1,26	-6,11	**2,16**	**2,27**	**1,55**	**5,88**	**3,22**	0,15
BM	-0,19	-1,00	**1,67**	1,97	-0,54	-1,46	1,26	-3,01	-1,11	**1,00**

Tabell 15: *Split half för k-median med standardiserade nyttovärden*

	Datamängd 1					Datamängd 2				
Variabler	K-median					K-median				
Segment	1	2	3	4	5	1	2	3	4	5
AC	**0,84**	0,06	0,23	**0,44**	**0,65**	**0,76**	**0,41**	**0,29**	**0,76**	**0,67**
BC	0,62	**0,57**	**0,24**	0,25	0,34	0,49	0,09	0,12	0,48	0,19
CT	0,04	0,45	0,16	0,14	0,11	0,02	0,05	0,12	0,15	0,12
DL	0,49	0,41	0,23	0,25	0,42	0,40	0,17	0,21	0,43	0,51
E	0,22	0,33	0,11	0,18	0,44	0,29	0,25	0,23	0,02	0,62
AS	0,07	0,09	0,37	0,30	0,16	0,03	0,41	0,05	0,01	0,06
BS	0,03	0,06	0,07	**0,38**	0,04	0,10	0,13	0,32	0,12	0,05
AM	**0,12**	**0,21**	**0,47**	0,06	**0,26**	**0,21**	0,09	**0,55**	**0,19**	**0,13**
BM	0,06	0,14	0,41	0,34	0,11	0,07	**0,42**	0,26	0,13	0,09

Tabell 16: *Split half för Fuzzy c-means med ostandardiserade nyttovärden*

	Datamängd 1					Datamängd 2				
Variabler	Fuzzy c-means					Fuzzy c-means				
Segment	1	2	3	4	5	1	2	3	4	5
AC	**3,95**	0,53	0,41	0,52	**8,68**	**1,14**	**8,43**	**3,32**	**3,26**	**7,50**
BC	1,17	**1,21**	**1,21**	**1,21**	3,56	-0,85	2,89	-0,66	-0,64	-4,41
CT	-4,56	-1,29	-1,23	-1,29	-8,89	-1,18	-7,23	-2,93	-2,90	-7,20
DL	0,79	0,58	0,59	0,58	0,85	-0,05	0,15	0,42	0,46	2,97
E	-1,34	-1,02	-0,98	-1,02	-4,19	0,95	-4,24	-0,16	-0,18	1,14
AS	0,33	0,13	0,10	0,12	0,08	-1,16	-1,81	-2,03	-2,20	-1,08
BS	-1,36	-2,30	-2,32	-2,30	-1,11	-0,02	0,54	-0,57	-0,52	-0,46
AM	**0,92**	**1,37**	**1,40**	**1,37**	**0,90**	**1,58**	**2,37**	**1,92**	**2,08**	**1,08**
BM	0,11	0,81	0,82	0,81	0,13	-0,39	-1,10	0,68	0,64	0,46

Tabell 17: *Split half för Fuzzy c-means med standardiserade nyttovärden*

	Datamängd 1					Datamängd 2				
Variabler	Fuzzy c-means					Fuzzy c-means				
Segment	1	2	3	4	5	1	2	3	4	5
AC	**0,36**	0,34	**0,81**	0,34	0,34	**0,46**	**0,73**	**0,22**	**0,46**	**0,77**
BC	0,35	**0,34***	0,57	**0,34***	**0,34***	0,20	0,22	0,11	0,20	0,55
CT	0,21	0,21	0,02	0,21	0,21	0,12	0,06	0,11	0,12	0,05
DL	0,32	0,31	0,48	0,31	0,31	0,27	0,50	0,18	0,27	0,40
E	0,24	0,23	0,24	0,23	0,23	0,30	0,45	0,17	0,30	0,21
AS	0,24	0,25	0,08	0,25	0,25	0,13	0,06	0,06	0,13	0,04
BS	0,12	0,12	0,03	0,12	0,12	0,15	0,08	0,45	0,15	0,11
AM	**0,28**	**0,29**	**0,13**	**0,29**	**0,29**	**0,30**	**0,16**	**0,61**	**0,30**	**0,19**
BM	0,26	0,27	0,08	0,27	0,27	0,24	0,11	0,30	0,24	0,06

*Not: * Med tre decimalers noggrannhet*

Tabell 18: *K-means med ostandardiserade nyttovärden med borttagna variabler*

Variabler			K-means			*Variabler*			K-means		
Segment	1	2	3	4	5	Segment	1	2	3	4	5
AC	**2,76**	-5,82	-1,15	**6,47**	**6,90**	BC	-0,59	**1,88**	-0,43	-4,71	1,60
BC	0,95	4,31	-0,25	2,30	-4,92	CT	**0,40**	-5,07	**3,74**	-8,23	-4,53
DL	0,81	**4,64**	-0,77	0,25	2,52	DL	-0,60	0,70	-3,31	2,72	**1,80**
E	-1,32	-1,84	**2,69**	-3,66	1,20	E	-1,38	-1,86	1,17	**3,66**	-2,94
AS	**3,58**	-1,21	-1,65	-1,75	-0,12	AS	**1,67**	-1,82	-2,15	-0,42	-0,04
BS	-2,47	-1,24	-0,92	-0,59	-1,06	BS	-2,43	**1,24**	-1,03	-1,00	-3,70
AM	-2,34	**3,09**	**2,00**	**2,28**	**0,91**	BM	-1,57	-1,25	**2,68**	**0,49**	**2,45**

Tabell 19: *K-median med ostandardiserade nyttovärden med borttagna variabler*

Variabler			K-median			*Variabler*			K-median		
Segment	1	2	3	4	5	Segment	1	2	3	4	5
AC	**7,03**	**8,16**	**3,86**	**8,36**	-0,90	BC	-7,15	0,76	**2,23**	**4,71**	-0,80
BC	-6,96	-0,35	0,29	4,54	0,76	CT	-7,36	-8,42	-1,34	-7,50	0,09
DL	2,66	4,03	-0,38	-1,17	**0,80**	DL	2,65	**2,94**	1,98	-0,69	-0,93
E	4,17	-4,72	-0,90	-4,90	0,51	E	**4,91**	-1,30	-2,66	-5,27	**0,29**
AS	-0,27	-0,73	**1,58**	-1,00	-2,23	AS	-0,25	-0,82	0,24	-1,02	-1,32
BS	-0,91	-0,87	-0,75	-0,62	-1,39	BS	-0,80	**-0,16**	-5,14	**-0,08**	**-0,05**
AM	**0,56**	**1,98**	-2,53	**2,14**	**3,54**	BM	**0,73**	-0,21	**2,90**	-0,92	-0,12

Tabell 20: *Fuzzy c-means med ostandardiserade nyttovärden med borttagna variabler*

Variabler			Fuzzy c-means			*Variabler*			Fuzzy c-means		
Segment	1	2	3	4	5	Segment	1	2	3	4	5
AC	**5,02**	**1,32**	**8,37**	**2,84**	**1,70**	BC	-0,32	0,17	-0,32	-0,31	**3,51**
BC	-0,60	-0,01	3,73	-0,13	-0,01	CT	-2,18	-5,23	-2,18	-2,29	-7,84
DL	1,16	0,68	0,05	0,76	0,75	DL	**0,35**	**1,42**	**0,36**	**0,40**	0,76
E	-1,04	0,36	-4,68	-0,35	0,12	E	-0,29	-0,91	-0,29	-0,30	-4,00
AS	-0,81	-0,88	-0,61	-0,74	-0,92	AS	-0,82	-0,83	-0,82	-0,82	-0,62
BS	-0,88	-1,23	-0,52	-1,09	-1,22	BS	-1,15	-0,85	-1,15	-1,17	-0,62
AM	**1,29**	**1,78**	**1,36**	**1,38**	**1,75**	BM	**0,56**	**0,12**	**0,56**	**0,55**	**-0,31**

Tabell 21: *K-means med standardiserade nyttovärden med borttagna variabler*

Variabler			K-means			*Variabler*			K-means		
Segment	1	2	3	4	5	Segment	1	2	3	4	5
AC	**0,77**	0,83	0,52	**0,22***	**0,32**	BC	0,37	0,20	**0,67**	0,13	0,26
BC	0,53	**1,05**	0,15	0,13	0,27	CT	0,04	0,08	0,12	0,13	**0,73**
DL	0,35	1,00	0,48	0,22	0,26	DL	0,59	**0,23**	0,55	0,18	0,16
E	0,10	0,91	**0,59**	0,16	0,14	E	**0,65**	0,16	0,20	**0,19**	0,45
AS	0,05	0,51	0,08	0,02	**0,42**	AS	0,12	0,30	0,07	0,18	0,06
BS	0,08	0,00	0,05	0,39	0,14	BS	**0,12**	0,04	0,10	**0,53**	0,03
AM	**0,18**	**0,79**	**0,21**	**0,62**	0,26	BM	0,09	**0,53**	**0,11**	0,14	**0,24**

*Not: * Med tre decimalers noggrannhet*

Tabell 22: *K-median med standardiserade nyttovärden med borttagna variabler*

Variabler	K-median					*Variabler*	K-median				
Segment	1	2	3	4	5	Segment	1	2	3	4	5
AC	**0,62**	**0,71**	**0,41**	**0,77**	0,17	BC	0,21	0,25	0,22	**0,73**	0,15
BC	0,53	0,01	0,22	0,54	0,22	CT	**0,71**	0,06	0,09	0,09	0,10
DL	0,00	0,52	0,22	0,47	**0,30**	DL	0,21	0,52	**0,23**	0,60	**0,26**
E	0,49	0,55	0,12	0,02	0,23	E	0,35	**0,66**	0,17	0,28	0,17
AS	0,05	0,10	**0,51**	0,07	0,11	AS	0,08	**0,13**	0,31	0,07	0,09
BS	0,11	0,04	0,22	0,07	0,23	BS	0,04	0,12	0,05	**0,10**	**0,60**
AM	**0,16**	**0,19**	0,12	**0,20**	**0,54**	BM	**0,24**	0,10	**0,47**	0,09	0,16

Tabell 23: *Fuzzy c-means med standardiserade nyttovärden med borttagna variabler*

Variabler	Fuzzy c-means					*Variabler*	Fuzzy c-means				
Segment	1	2	3	4	5	Segment	1	2	3	4	5
AC	**0,31**	**0,80**	**0,46**	**0,46**	**0,51**	BC	0,26	0,25	0,45	0,26	**0,72**
BC	0,23	0,55	0,27	0,27	0,25	CT	0,18	0,16	0,09	0,18	0,05
DL	0,28	0,45	0,30	0,30,	0,31	DL	**0,32**	**0,29**	**0,53**	**0,31**	0,59
E	0,25	0,04	0,28	0,28	0,32	E	0,30	0,25	0,42	0,28	0,31
AS	0,17	0,06	0,18	0,17	0,16	AS	0,18	0,20	0,12	0,18	0,08
BS	0,19	0,06	0,16	0,15	0,14	BS	0,17	0,14	0,13	0,16	0,09
AM	**0,43**	**0,15**	**0,30**	**0,29**	**0,27**	BM	**0,26**	**0,34**	**0,14**	**0,29**	**0,09***

*Not: * Med tre decimalers noggrannhet*

Tabell 24: *K-means med ostandardiserade nyttovärden i randomiserad ordning*

Variabler	K-means				
Segment	1	2	3	4	5
AC	**5,38**	-6,91	**4,01**	-2,83	**5,97**
BC	2,07	4,34	0,90	0,03	-5,79
CT	-5,18	0,37	-4,01	**6,43**	-6,03
DL	-0,23	**4,74**	1,13	-5,93	2,02
E	-2,05	-2,54	-2,03	2,30	3,83
AS	**1,56**	-0,32	-2,18	-1,73	-0,13
BS	-0,31	-2,52	-1,24	-1,21	-0,79
AM	-0,75	**2,68**	**3,02**	0,29	0,03
BM	-0,50	0,16	0,40	**2,65**	**0,89**

Tabell 25: *K-median med ostandardiserade nyttovärden i randomiserad ordning*

Variabler	K-median				
Segment	1	2	3	4	5
AC	**9,05**	**7,95**	-3,70	**0,94**	**3,96**
BC	-3,13	4,22	1,80	-0,85	0,13
CT	-8,03	-7,87	-0,02	0,08	-3,92
DL	3,93	-0,30	**1,86**	-0,30	-0,07
E	-1,81	-3,99	0,05	0,13	-0,11
AS	-0,90	-1,08	-2,03	-2,12	**1,65**
BS	-0,58	-0,59	-4,26	0,63	-1,24
AM	**1,39**	**2,48**	**3,85**	**2,66**	-2,04
BM	0,09	-0,82	2,45	-1,17	1,63

Tabell 26: *Fuzzy c-means med ostandardiserade nyttovärden i randomiserad ordning*

Variabler	Fuzzy c-means				
Segment	1	2	3	4	5
AC	**1,67**	**1,75**	**1,67**	**8,37**	**5,17**
BC	-0,01	-0,01	-0,01	2,89	-0,10
CT	-1,91	-1,99	-1,91	-8,06	-5,30
DL	0,45	0,47	0,45	0,84	1,15
E	-0,19	-0,23	-0,19	-4,03	-0,93
AS	-0,85	-0,84	-0,85	-0,65	-0,75
BS	-1,25	-1,26	-1,25	-0,47	-0,78
AM	**1,57**	**1,56**	**1,57**	**1,40**	**1,39**
BM	0,53	0,55	0,53	-0,28	0,14

Tabell 27: *K-means med standardiserade nyttovärden i randomiserad ordning*

Variabler	K-means				
Segment	1	2	3	4	5
AC	**0,74**	0,31	0,29	**0,29**	**0,27**
BC	0,38	**0,46**	**0,63**	0,17	0,21
CT	0,05	0,61	0,31	0,13	0,11
DL	0,43	0,08	0,65	0,22	0,23
E	0,34	0,42	0,20	0,12	0,18
AS	0,08	0,06	0,06	0,34	0,23
BS	0,08	0,08	0,09	**0,47**	0,14
AM	**0,17**	0,12	**0,25**	0,17	**0,51**
BM	0,09	**0,26**	0,10	0,30	0,38

Tabell 28: *K-median med standardiserade nyttovärden i randomiserad ordning*

Variabler	K-median				
Segment	1	2	3	4	5
AC	**0,79**	**0,67**	0,15	**0,27**	**0,42**
BC	0,55	0,22	**0,49**	0,14	0,21
CT	0,04	0,06	0,46	0,13	0,10
DL	0,47	0,49	0,29	0,23	0,19
E	0,21	0,53	0,40	0,17	0,13
AS	0,05	0,11	0,12	0,13	0,41
BS	0,07	0,06	0,06	0,27	0,20
AM	**0,17**	**0,20**	**0,23**	**0,57**	0,09
BM	0,06	0,13	0,19	0,31	**0,39**

Tabell 29: *Fuzzy c-means med standardiserade nyttovärden i randomiserad ordning*

Variabler	Fuzzy c-means				
Segment	1	2	3	4	5
AC	**0,61**	**0,79**	**0,35**	**0,35**	**0,35**
BC	0,34	0,53	0,25	0,25	0,25
CT	0,09	0,04	0,17	0,17	0,17
DL	0,40	0,46	0,28	0,28	0,28
E	0,33	0,24	0,25	0,25	0,25
AS	0,11	0,06	0,19	0,19	0,19
BS	0,11	0,07	0,17	0,17	0,17
AM	**0,23**	**0,15**	**0,34**	**0,34**	**0,34**
BM	0,15	0,08	0,28	0,28	0,28

C Datakod

Matlab M-kod Fuzzy c-means clustering

```
load data.dat
fcmdata=data
[center,U,objFcn] = fcm(fcmdata,5);
figure
plot(objFcn)
title('Objective Function Values')
xlabel('Iteration Count')
ylabel('Objective Function Value')
maxU = max(U);
index1 = find(U(1, :) == maxU);
index2 = find(U(2, :) == maxU);
index3 = find(U(3, :) == maxU);
index4 = find(U(4, :) == maxU);
index5 = find(U(5, :) == maxU);
```

Stata kommandon K-median

cluster kmedian var1 var2 var3 var4 var5 var6 var7 var8 var9, k(5) name(cluster) gen(group)

tabstat var1 var2 var3 var4 var5 var6 var7 var8 var9, by(cluster)

Kommandon i R för Adjusted Rand Index

kmeanostandardiserat<-read.csv("file=datakmean.csv")

kmedianostandardiserat<-read.csv("file=datakmedian.csv")

fuzzyostandardiserat<-read.csv("file=datafuzzy.csv")

kmeanstandardiserat<-read.csv("file=datakmeans.csv")

kmedianstandardiserat<-read.csv("file=datakmedians.csv")

fuzzystandardiserat<-read.csv("file=datafuzzys.csv")

adjustedRandindex(kmeanostandardiserat,kmeanstandardiserat)

adjustedRandindex(kmedianostandardiserat,kmedianstandardiserat)

adjustedRandindex(fuzzyostandardiserat, fuzzystandardiserat)

adjustedRandIndex(kmeanostandardiserat,kmedianostandardiserat)

adjustedRandIndex(kmedianostandardiserat,fuzzyostandardiserat)

adjustedRandIndex(kmeanostandardiserat,fuzzyostandardiserat)

adjustedRandIndex(kmeanstandardiserat,kmedianstandardiserat)

adjustedRandIndex(kmeanstandardiserat,fuzzystandardiserat)

adjustedRandIndex(kmedianstandardiserat,fuzzystandardiserat)