

Grafiske analysemetoder for ikke-geometriske design

Ranveig Niemi

Master i fysikk og matematikk
Oppgaven levert: Juni 2008
Hovedveileder: John Sølve Tyssedal, MATH

Oppgavetekst

Se på hvordan en grafisk kan analysere 12 forsøks Plackett and Burman design og to design med 16 forsøk på en ny måte. Dette gjøres ved hjelp av scatterplott, dendogram, analysering av aliasstrukturen til designa og ortogonalisering.

Oppgaven gitt: 15. januar 2008
Hovedveileder: John Sølve Tyssedal, MATH

Forord

Denne mastergraden er blitt til i løpet av vårsemesteret 2008. Studiebelastningen på oppgaven er på 30 studiepoeng. Høsten 2007 jobbet jeg med et prosjekt med samme tema som i vår. Denne oppgaven baserer seg derfor på prosjektoppgaven som jeg skrev i høst. Forskjellen mellom prosjekt og master er at jeg i høst skrev utelukkende om 12 enkeltforsøk av Plackett and Burman. I vår har jeg i tillegg sett på et par ulike design med 16 forsøk. For å få flyt i oppgaven er kapittel 2 om tonivåforsøk et noe redigert kapittel fra prosjektoppgaven.

Oppgaven er skrevet med tanke på at leseren skal ha den kunnskapen som forventes av folk som har studert matematikk og statistikk på masternivå. Det vil si en forståelse av grunnleggende statistikk og forståelse av andre deler av matematikken. For eksempel må leseren ha noe kunnskap om matriser. En god del er likevel forklart underveis.

I løpet av våren har jeg lært mye om å skrive oppgave og om forsøksplanlegging. Dette tar jeg med meg videre i livet.

Takk til John S. Tyssedal for gode idéer og god veiledning.

Trondheim, 7. juni 2008

Ranveig Niemi

Sammendrag

Den grafiske analyseringen av tonivåforsøk har lenge begrenset seg til Lenth's metode, normalplott og halvnormalplott. Disse plotta baserer seg på at de estimerte kontrastene enten representerer aktive effekter eller støy. For design som ikke tilhører 2^{k-p} familien, dvs. de ikke-geometriske designa, er ikke dette lenger oppfylt. I denne oppgaven viser vi hvordan en ved hjelp av scatterplott og dendrogram av absoluttverdien til hovedeffekter og tofaktorsamspill, sammen med aliasstrukturen til designa og ortogonalisering, kan konstruere plott som gjør det lettere å finne aktive effekter. Designa som blir undersøkt er 12 forsøks Plackett and Burman design og to design med 16 forsøk.

Nøkkelord: Plackett and Burman, Ortogonalisering, 16 forsøks design, Tonivåforsøk

Innhold

Forord	i
Sammendrag	iii
Innhold	v
1 Innledning	1
2 Tonivå faktorielle forsøk	3
2.1 Tonivåforsøk	3
2.2 Forsøk som tilhører 2^{k-p} familien	3
2.2.1 2^k -forsøk	3
2.2.2 Fraksjoner av faktorielle forsøk	5
2.3 Ikke-geometriske design	7
2.3.1 Plackett-Burman design	7
2.3.2 Aliasstrukturen til tonivådesign	9
3 16 forsøks design	13
3.1 Konfunderende mønster til 2_{IV}^{8-4}	13
3.2 Design med 16 enkeltforsøk	14
3.3 H_2	14
3.3.1 Aliasmatrise til H_2	17
3.4 H_3, H_4 og H_5 design	18
3.5 $H_6(16,8,3)$ screen	19
4 Ortogonalisering	23
4.1 Metode for ortogonaliseringen	25
4.1.1 12 forsøks eksempel	25
4.1.2 16 forsøks eksempel	26
5 Resultat	29

5.1	12 forsøks PB design	29
5.1.1	Algoritme for 12 forsøks PB design	29
5.1.2	Modell: $Y = 2A + 2B + BC + \varepsilon$	31
5.1.3	Modell $Y = A + BC + \varepsilon$	36
5.1.4	Modell $Y = 2A + 2C + AB + BC + \varepsilon$	42
5.2	16 forsøks design med bruk av H_2	47
5.2.1	Algoritme for 16 forsøks metode	47
5.2.2	Modell $Y = A + DJ$	50
5.2.3	Modell $Y = A + AJ + \varepsilon$	53
5.2.4	Modell $Y = A + B + DJ + \varepsilon$	56
5.2.5	Modell $Y = A + B + AJ + \varepsilon$	61
5.2.6	Modell: $Y = 2A + 2D - AE + AF + \varepsilon$	64
5.3	16 forsøks design med bruk av H_6	67
5.3.1	Modell: $Y = A + AC$	68
5.3.2	Modell: $Y = 2A + 3B + C + CD$	70
5.4	Kritiske kommentarer	73
6	Konklusjon	75
	Referanser	78
A	Ortogonale design med 16 forsøk	79

Kapittel 1

Innledning

I forsøksplanlegging har vi ofte situasjoner der vi har prosesser med mange faktorer på ulike nivå. Vi vil da gjerne finne hva slags effekt disse faktorene har på en eller flere responser. En av de vanligste måtene å begrense problemet på er å se på faktorer der man har to nivå, da kan man utføre noe som kalles tonivåforsøk. En av de vanligste formene for tonivåforsøk er 2^k forsøk. En annen type design er Plackett and Burman (PB) design. PB design har tradisjonelt sett bare vært brukt til å finne hovedeffekter fordi de egner seg veldig godt til dette. Men det viser seg at de egner seg godt til å se på tofaktorsamspill også. Også andre design kan konstrueres og i denne oppgaven ser vi på tre forskjellige design.

De tre forskjellige designa vi ser på er 12 forsøks PB designet og et par forskjellige 16 forsøks design som tar utgangspunkt i 2_{IV}^{8-4} designet. Det første av disse er et design som kalles H_2 . Dette er en naturlig utvidelse av 2_{IV}^{8-4} designet. H_2 har samme projektivitet som 2_{IV}^{8-4} designet, men har 12 faktorer i stedet for 8. Projektivitet defineres senere i oppgaven. Det andre designet er et design som kalles H_6 . Dette er et design som ligner på 2_{IV}^{8-4} designet ved at det har projektivitet 3 og det har 16 forsøk og 8 faktorer. Men der 2_{IV}^{8-4} designet har tofaktorsamspill som er konfunderte med tofaktorsamspill har ikke H_6 dette. Både H_2 og H_6 er design som ikke er så mye brukt og som kan være interessante å se på.

For å undersøke disse designa vil vi i denne oppgaven se på forskjellige modeller for alle de tre designa. Vi vil simulere data og se på hvordan det går an å tolke hovedeffekter og tofaktorsamspill fra disse simulerte dataene. Scatterplott og dendogram er to viktige redskap for å kunne se hvilke hovedeffekter og tofaktorsamspill som er aktive. For å kunne tolke plottene best mulig ser vi på en viktig egenskap ved designa, nemlig aliasstrukturen. For å lage enda bedre plott ser vi også på ortogonalisering av designa. Ved å ha ortogonale matriser kan vi finne forventningsrette

estimerer for effektene. På de forbedrede plottene er det da lettere å se de aktive faktorene fordi forventningsskjevheten som kom fra konfunderingen mellom hovedeffekter og tofaktorsamspill er borte.

Det som er viktig å huske på i denne type forsøk er at de er innledende undersøkelser. I alle tilfeller vil man etter å ha utført denne typen tester, som vi presenterer i denne oppgaven, gå videre med flere forsøk. Det som er fasinerende er hva man faktisk kan finne ut av fra disse 12 eller 16 forsøkene.

Opgaven er delt inn slik at det først er en del om tonivåforsøk, der ser vi på 2^k forsøk og på 12 forsøks Plackett and Burman design. Etter det ser vi på 16 forsøks design. Først generelt, men så inngående om både H_2 og H_6 . Resten av teoridelen handler om ortogonalisering og hvordan dette blir brukt i vårt tilfelle. Resultatdelen inneholder en del om 12 forsøks PB design, en del om 16 forsøks design ved bruk av H_2 og en del om 16 forsøks design ved bruk av H_6 . Til slutt i resultatdelen prøver vi å se litt kritisk på det vi har gjort og hva som er negativt med det vi har gjort. I appendiks har vi tatt med de tre 16 forsøksdesigna H_3 , H_4 og H_5 . Disse nevner vi i oppgaven, og det kan være interessant å se hvordan de ser ut i forhold til H_2 og H_6 .

Kapittel 2

Tonivå faktorielle forsøk

2.1 Tonivåforsøk

Tonivåforsøk er flerfaktorielle forsøk der hver faktor har to nivå. Et forsøk med k faktorer på to nivå kalles 2^k forsøk. Vi skal se på forsøk som tilhører 2^{k-p} familien, $p = 0, 1, \dots, k - 1$, dvs. $\frac{1}{2^p}$ fraksjoner av et 2^k forsøk. Deretter vil vi se på noen andre tonivå forsøk som ikke tilhører denne familien.

2.2 Forsøk som tilhører 2^{k-p} familien

2.2.1 2^k -forsøk

Faktorielle forsøk består av en responsvariabel (ofte kalt Y) og flere faktorer som det er mulig å forandre på (ofte kalt $A, B, C \dots$). I 2^k forsøk står k -en for antall faktorer. Så i et 2^3 forsøk har vi 3 faktorer (A, B og C). Og $2^3 = 8$ står for antall forsøk som blir gjort i et 2^k -forsøk. Disse forsøkene blir altså gjort med k faktorer på to nivå hver. Disse nivåene blir ofte betegnet som lavt (-1 eller -) eller høyt (+1 eller +). + og - blir brukt i tabeller, mens -1 og 1 blir brukt i matriser. Nivåene trenger ikke kvantitativt å være høye og lave. Det kan være to nivå som er kvalitativt ulike også. For eksempel om prosessen blir påvirket av noe (høyt), eller ikke (lavt). Hvis vi ser på et 2^3 -forsøk vil det da se ut som i tabell 2.1.

Her ser vi at i forsøk 1 er alle de tre faktorene på lavt nivå, mens i forsøk nr 2 er faktor A på høyt og de andre to på lavt nivå. Designet inneholder alle de 8 mulige kombinasjonene vi kan ha når vi har 3 faktorer. Sånn forsøket er satt opp i

<i>Forsøk</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>Resultat</i>
1	-	-	-	Y_1
2	+	-	-	Y_2
3	-	+	-	Y_3
4	+	+	-	Y_4
5	-	-	+	Y_5
6	+	-	+	Y_6
7	-	+	+	Y_7
8	+	+	+	Y_8

Tabell 2.1: Faktornivåer for et 2^3 forsøk.

tabell 2.1 er standardformen. Når vi utfører forsøkene, og stiller opp faktor A , B og C sånn som i tabell 2.1 får vi ut en respons Y_i , $i = 1, \dots, 8$ fra alle 8 forsøkene. Ved å bruke disse responsverdiene, og kunnskapen vår om oppsettet, kan vi finne ut noe om hvilke faktorer som er signifikante. En vanlig ting å regne på er kontraster. Kontraster defineres generelt som:

$$\sum c_i Y_i \quad (2.1)$$

Der c_i er forskjellig vektning på de ulike Y_i -ene og $\sum c_i = 0$. Hvis vi ser spesifikt på vår situasjon er det ofte en spesiell type kontraster som kalles hovedeffekter vi regner på. Hovedeffekten til A , A_{eff} , defineres som $E(\bar{Y}_{A+} - \bar{Y}_{A-})$. Dvs. forventningen til gjennomsnittet av responsen når A er på høyt nivå minus gjennomsnittet av responsen når A er på lavt nivå. Denne kan estimeres som:

$$\hat{A}_{\text{eff}} = \frac{1}{n/2} \sum_{j=1}^n Y_j x_j^A \quad (2.2)$$

Der x^A er kolonnen med -1 og $+1$ for faktor A . I vårt 2^3 -eksempel er $n = 8$. På samme måte finner vi de andre hovedeffektene. I tillegg til hovedeffekter er det ofte tofaktorsamspillet vi vil estimere. For å estimere disse må vi først lage en kolonne for tofaktorsamspillet, og deretter regne ut kontrasten på samme måte som i (2.2). Tofaktorsamspillskolonnene får vi ved å gange de to respektive faktorkolonnene sammen. Når vi i vårt eksempel bare har 3 faktorer, kan vi lett regne ut de 3 tofaktorsamspillene og det ene trefaktorsamspillet. I tabell 2.2 ser vi alle kombinasjonene vi har i dette eksempelet.

Ser vi på kolonnene i tabell 2.2 så ser vi at alle kolonnene er ortogonale. Dette betyr at vår utregning av effekt A ikke vil påvirke vår utregning av effekten til

<i>Forsøk</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>AB</i>	<i>AC</i>	<i>BC</i>	<i>ABC</i>
1	-	-	-	+	+	+	-
2	+	-	-	-	-	+	+
3	-	+	-	-	+	-	+
4	+	+	-	+	-	-	-
5	-	-	+	+	-	-	+
6	+	-	+	-	+	-	-
7	-	+	+	-	-	+	-
8	+	+	+	+	+	+	+

Tabell 2.2: Faktornivåer for et 2^3 forsøk

BC , eller noen av de andre effektene. For å regne med kontraster og hovedeffekter når vi får mange forsøk, er det lurt å definere en designmatrise. Denne består av faktor-kolonnene fra tabell 2.1 i tillegg til en kolonne med bare +1 som kommer foran faktorkolonnene. For 2^3 -forsøket ser det ut som i (2.3).

$$X = \begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad (2.3)$$

Designmatrisa i (2.3) kan vi bruke til å regne ut kontrastene vi trenger. F.eks. hovedeffekter og tofaktorsamspill. Se Tyssedal og Samset (1997).

2.2.2 Fraksjoner av faktorielle forsøk

Fraksjoner av faktorielle forsøk er forsøk der ikke alle mulige kombinasjoner er med, som i tabell 2.2, men bare deler av forsøkene. For å få noe ut av dette må vi spesifisere hva vi har lyst til å finne. Hvis vi baserer oss på effektparsomhet som Hamada og Wu (1992) skriver om, så vil antagelig ikke så mange effekter være signifikante. La oss se litt systematisk på det. Hvis vi har et 2^k forsøk er det ingen effekter som er konfunderte med hverandre. Det vil si at alle effekter er reelle. Hvis vi derimot har det halve antall forsøk, dvs. 2^{k-1} forsøk, så vil det være fullstendig konfundering mellom noen av effektene. Hvilke effekter dette er, kommer an på

hvordan designet til forsøket er laget. Hvis vi i tabell 2.2 kutter siste halvdel i forsøket så ser vi at $A = -AC$ og $B = -BC$ og $AB = -ABC$. Da får vi en tabell med forsøk som i tabell 2.3.

<i>Forsøk</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>AB</i>	<i>AC</i>	<i>BC</i>	<i>ABC</i>	<i>Y</i>
1	-	-	-	+	+	+	-	Y_1
2	+	-	-	-	-	+	+	Y_2
3	-	+	-	-	+	-	+	Y_3
4	+	+	-	+	-	-	-	Y_4

Tabell 2.3: Faktornivåer for et $2^{3-1} = 2^2$ forsøk

Siden vi får I (identitetskolonnen) hvis vi ganger to like kolonner med hverandre ser vi at vi har $I = -C$ fra $AA = -AAC$ og de to andre ligningene. I kalles den definerende relasjon. Rekken med bokstaver som danner I kalles ordet. Her er ordet C , mens det kan være ABC , eller enda lengre i andre tilfeller. Fra tabellen ser vi at vi bare har med forsøk der faktor C er på lavt nivå. Akkurat denne måten å dele forsøket på er kanskje ikke så lurt, men det illustrerer poenget med konfundering. Hvis vi heller tar forsøket der ABC er enten -1 eller 1 så får vi et litt enklere forsøk. Men da vil vi få at $A = \pm BC$ og $B = \pm AC$ og $C = \pm AB$. Dette betyr at vi ikke kan si om det er A eller BC , eller begge som gir effekten vi finner, hvis analysen vår sier at A er signifikant.

I dette lille eksempelet med bare 8 forsøk er det egentlig unødvendig å dele det opp. Hvis vi har et 2^5 -forsøk vil vi ha 5 hovedeffekter, 10 tofaktorsamspill, og flere trefaktor og firefaktor-samspill i tillegg til femfaktorsamspillet. I et sånt tilfelle, eller tilfeller med enda flere faktorer, er det mer hensiktsmessig å snakke om fraksjoner av faktorielle forsøk. Av de ulike faktorer og tofaktorsamspill er det mest sannsynlig ikke så mange som er signifikante. Og de som er signifikante er mest sannsynlig hovedeffekter, tofaktorsamspill og kanskje trefaktorsamspill. Derfor trenger vi ikke $2^5 = 32$ forsøk for å regne ut hvilke av disse laveregrads kontraster som er signifikante. Deler vi et 2^5 -forsøk i to trengs det bare $2^{5-1} = 16$ enkeltforsøk. Kostnadene med forsøket blir altså halvert, mens resultatene vi får vil i de fleste tilfeller si det samme. En mulig definerende relasjon det går an å se på er $I = ABCDE$. Da blir konfunderingene blant annet: $A = BCDE$, $AB = CDE$ og $ABC = DE$. Da kan vi med ganske god sannsynlighet si at det er hovedeffekten som er signifikant og ikke firefaktorsamspillet hvis A, B, C, D eller E ser ut til å være signifikante. Og tofaktorsamspillene er mer sannsynlig enn trefaktorsamspillene, så vi kan gå ut i fra at det er de som er signifikante, hvis f.eks. AB viser seg å være signifikant.

Det vi har sett på her er at 2^k -forsøk gir en fullstendig konfundering, hvis det

er konfundering mellom noen av faktorene eller samspillene. Det vi skal se på etterhvert er at for mange andre design er det ikke en sånn fullstendig konfundering. I stedet er det en delvis konfundering mellom de ulike effektene.

2.3 Ikke-geometriske design

Ikke-geometriske design er ortogonale tonivådesign som ikke er med i 2^{k-p} familien. Et av de kjente typene slike design heter Plackett and Burman (PB) design. Dette skal vi se nærmere på her, spesielt på PB design med 12 enkeltforsøk, siden det er dette vi har brukt i anvendelsene.

2.3.1 Plackett-Burman design

Plackett and Burman design ble utarbeidet av Plackett and Burman i 1946. Dette er også design med ortogonale kolonner. Fordelen med PB design er at du kan ha mange færre forsøk enn ved 2^k forsøk og likevel finne signifikante effekter. PB design er heller ikke begrenset til at det må være 2^k forsøk, men er begrenset til at det må være et multippel av 4 forsøk. Når 2^k bare kan ha 4, 8, 16, 32, 64 osv. forsøk, så kan PB design ha 12, 20, 24, 28, 36 osv. forsøk i tillegg til det antall 2^k forsøk kan ha. Så du har flere valgmuligheter når du skal velge antall forsøk du skal gjøre. Dette kan være viktig på grunn av økonomiske hensyn. De ulike designa deler seg i to grupper. Geometriske design, som 2^k forsøkene er og de ikke-geometriske designa. PB designet med 12 enkeltforsøk ser ut som i tabell 2.4. Dette er et syklisk design som konstrueres ved at en rad med $N - 1 = 11$ pluss- og minustegn blir brukt, og $N - 2 = 10$ nye rader blir konstruert ved å syklisk flytte alle pluss- og minustegn et hakk hver rad. Til slutt må en rad med bare minustegn legges til. Den første raden som vi starter med er gitt av Plackett and Burman (1946).

I utgangspunktet har PB design blitt brukt til å beregne hovedeffekter. Dette fordi dette designet kan estimere alle hovedeffekter hvis alle samspill er neglisjerbare. Det oppstår problem hvis alle samspill ikke er neglisjerbare. Da stemmer ikke de resultatene vi får, med det vi skulle forvente (Lin & Draper 1993). En bra egenskap til PB design er projeksjonsegenskapene til designet. Projeksjon handler om egenskapene til faktorielle forsøk når vi begrenser forsøket til noen av faktorene. Definisjonen Box og Tyssedal (1996) gir er:

Et $n \times k$ design D , med n observasjoner og k faktorer på to nivå, har projektivitet P hvis alle delmengder med P faktorer, ut av de k , inneholder et komplett 2^P faktorielt design. Designet blir da kalt et (n, k, P) forsøk (screen).

<i>Forsøk</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>J</i>	<i>K</i>	<i>L</i>
1	+	-	+	-	-	-	+	+	+	-	+
2	+	+	-	+	-	-	-	+	+	+	-
3	-	+	+	-	+	-	-	-	+	+	+
4	+	-	+	+	-	+	-	-	-	+	+
5	+	+	-	+	+	-	+	-	-	-	+
6	+	+	+	-	+	+	-	+	-	-	-
7	-	+	+	+	-	+	+	-	+	-	-
8	-	-	+	+	+	-	+	+	-	+	-
9	-	-	-	+	+	+	-	+	+	-	+
10	+	-	-	-	+	+	+	-	+	+	-
11	-	+	-	-	-	+	+	+	-	+	+
12	-	-	-	-	-	-	-	-	-	-	-

Tabell 2.4: Faktornivåer for 12 forsøks Plackett and Burman design

Et ortogonalt design med $k = n - 1$ blir kalt metta (saturated). Når noe er metta, betyr det at det er fylt opp. Når det her er et design som er metta, med antall faktorer en mindre enn antall forsøk, er det det fordi antall faktorer er lik antall ortogonale kolonner. En tabell over projektiviteten til metta PB design finnes i tabell 2.5 (Box & Tyssedal 1996, Tyssedal & Samset 1999).

<i>Projektivitet</i>	<i>Antall forsøk, n</i>
$P=2$	4, 8, 16, 32, 40, 56, 64, 88, 96
$P=3$	12, 20, 24, 28, 36, 44, 48, 52, 60, 76, 100
$P=4$	68, 72, 80, 84

Tabell 2.5: Projektiviteten til Plackett and Burman design.

I utgangspunktet kan det virke ganske utrolig at vi fra et forsøk med 11 faktorer og 12 forsøk skal kunne finne ut av noe i det hele tatt. Sammenligner vi med 2^k -forsøk med $k=5$ må vi jo ha 32 forsøk for å utføre hele forsøket. For å utføre halvparten må vi uansett gjøre 16, som også er mer enn 12. Men det er her projektiviteten til designet kommer inn. I tabell 2.5 ser vi at alle geometriske design har projektivitet 2, mens 12 forsøks PB har projektivitet 3. I et 12-forsøks PB design vil alle hovedeffektene være konfunderte med tofaktorsamspill, så det er ikke noe som forandrer på det. Men konfunderingen er ikke total, som i 2^k -forsøkene, så det går an å finne noe ut av forsøkene våre likevel, selv om høyereordens samspill er tilstede.

2.3.2 Aliasstrukturen til tonivådesign

Aliasstrukturen til ikke-geometriske design er annerledes enn til det tradisjonelle 2^k forsøket. Konfunderingen mellom kontrastene i 2^k forsøk er enten fullstendig eller null, mens den bare er delvis mellom kontrastene i de fleste ikke-geometriske PB design. Resten av dette avsnittet bygger på Lin og Draper (1993). De tar for seg alle PB design for N lik en multippel av 4, men med begrensningen $N \leq 100$ og $N \neq 92$. Videre argumenterer de for at hvis vi vet at alle to, eller høyereordens, samspill er neglisjerbare kan vi estimere hovedeffektene. Bruk av minste kvadraters metode (MKM) gir

$$\hat{\theta} = (X^T X)^{-1} X^T Y = \frac{1}{12} X^T Y \quad (2.4)$$

der X er designmatrisa til 12 forsøks PB design. La oss anta at vi vil tilpasse regresjonsmodellen

$$E(Y) = X\theta \quad (2.5)$$

med MKM. Da får vi at (2.4) er forventningsrett, dvs. $E(\hat{\theta}) = \theta$. Der θ er alle hovedeffektene delt med to. Men hvis modellen ikke er riktig vil estimatoren være skjev. Hvis den korrekte modellen er på formen

$$E(Y) = X\theta + X_1\theta_1 \quad (2.6)$$

blir $E(\hat{\theta}) = \theta + A\theta_1$ der $A = (X^T X)^{-1} X^T X_1$ kalles aliasmatrisen. Hvis f.eks. X_1 inneholder alle tofaktorsamspillskolonnene vil θ_1 være vektoren av tofaktorsamspill delt med to. A er da en $k \times k(k-1)/2$ matrise, der k er antall faktorer. Hovedeffektene $(1, \dots, k)$ knyttes til radene i A matrisa og samspillene $(1, 2), (1, 3), \dots, (k-1, k)$ knyttes til kolonnene. Den transponerte A^T kalles ofte aliasstabellen. Hvis vi ser på 12 forsøks PB så er konfunderingene begrenset til $\pm \frac{1}{3}$ og 0. Men hver hovedeffekt kan være konfundert med 45 tofaktorsamspill. Konfunderingen mellom A og tofaktorsamspillene med A er 0, pga ortogonaliteten mellom tofaktorsamspillskolonnene internt og mellom tofaktorsamspillskolonnene og hovedeffektkolonnen til A . Pga at konfunderingen er enten 0 eller i absoluttverdi lik $\frac{1}{3}$ vil endel av effektene samle seg i grupper, hvis vi ser på absoluttverdien til effektene. Siden de alltid vil bli konfunderte enten med $\frac{1}{3}$ eller ingenting. Dette er det mulig å bruke til å analysere dataene. La oss se på et konkret eksempel som kan forklare. Modellen $Y = A + AB$ er en enkel modell. Her er A ikke konfundert med A , siden A er med i tofaktorsamspillet. B er heller ikke konfundert med noen effekter. C, D, \dots, K vil

derimot være konfunderte med AB . Dette betyr at absoluttverdien av kontrastene for hovedeffektene mest sannsynlig vil dele seg i tre. A for seg, B for seg, men betydelig lavere enn A og C, D, \dots, K i en gruppe. Hvis vi da hadde fått et sånt plott fra noen data med støy, kunne vi ha konkludert med at det var noe spesielt med A og B og undersøkt disse nærmere. Så selv om det virker mer komplisert med konfunderinger som ikke er fullstendige kan det i tilfellet med PB design brukes i analysering av forsøkene.

Eksisterende metoder

Det eksisterer endel metoder for å analysere tofaktorforsøk. Noen av de mest vanlige er Lenth's metode (Lenth 1989) og Normal og Halv-Normal plott (Daniel 1959). I starten da PB design var ferske var det disse typen metoder som ble brukt til å analysere PB, på samme måte som 2^k forsøk. Fortsatt er dette det eneste tilbudet som finnes i mange programpakker. Dette går greit hvis vi vet at det bare er hovedeffekter som er signifikante. Men med en gang det er modeller med tofaktorsamspill, og vi prøver å bruke PB design fungerer ikke disse metodene så bra. De senere årene er det derfor flere som har jobbet med å lage metoder som tar for seg PB design spesifikt. Blant annet Hamada & Wu (1992) som har laget en effektbasert metode og Tyssedal & Samset (1997) som har en metode som er faktorbasert. La oss se litt nærmere på disse to metodene som er laget med tanke på PB design, før vi ser nærmere på hva som er nytt i denne oppgaven.

Hamada og Wu (1992) har laget en algoritme som bygger på de to basisprinsippene effektarvelighet og effektsparsomhet. Effektarvelighet går ut på at vi antar at et tofaktorsamspill er signifikant bare når minst en av de to faktorene sin hovedeffekt også er signifikant. Det kalles sterk arvelighet hvis begge faktorene er hovedeffekter og svak arvelighet hvis bare en av faktorene har en hovedeffekt. Effektsparsomhet går ut på at vi antar at det bare er et lite antall effekter som er viktige i forhold til resten. Begge disse antagelsene er nokså vanlig å ta utgangspunkt i når vi analyserer faktorielle forsøk. Algoritmen til Hamada og Wu er delt opp i tre skritt:

1. Betrakt alle hovedeffekter og samspill som er ortogonale til hovedeffektene. Bruk standard analysemetoder, dvs. variansanalyse og halv-normal plott, til å bestemme hvilke effekter som er signifikante.
2. Bruk effektarvelighet til å betrakte (i) effektene identifisert i forrige skritt, (ii) tofaktorsamspillene der minst en av faktorene er plukket ut som hovedeffekt i (i). Og se på (iii) tofaktorsamspill foreslått av de som vil ha forsøket gjennomført. Bruk forlengs regresjon til å identifisere effekter blant effektene i (i)-(iii). Gå til skritt 3.

3. Bruk forlengs regresjon til å identifisere signifikante effekter blant effektene funnet i forrige skritt i tillegg til hovedeffektene. Gå tilbake til skritt 2.

Gå mellom skritt 2 og 3 til modellen slutter å forandre seg mellom skrittene. Hamada og Wu (1992) konkluderer selv med at denne metoden fungerer bra hvis det bare er noen få av tofaktorsamspillene som er signifikante, og hvis disse ikke er så store. Men hvis størrelsen til tofaktorsamspillene blir like store eller større enn hovedeffektene får metoden deres problemer.

Tyssedal og Samset (1997) ser i motsetning til Hamada og Wu (1992) på faktorene i stedet for effektene. Tanken er først å finne ut hvilke faktorer som er aktive, for deretter å nærmere undersøke hvordan de påvirker responsen. Metoden bygger på antagelsen om faktorsparsomhet og projektivitetsegenskapene til designet. Konkret så ser de på estimat av variansen til feilleddet etter tilpasning av den faktoren eller faktorene som de antar er aktive. Hovedsaklig ser de på situasjonen med inntil tre aktive faktorer, men de kommenterer også situasjoner der det er flere aktive faktorer. Et problem med denne metoden er at den lett kan overtilpasse dataene. I tillegg til denne mer teoretiske metoden viser Tyssedal og Samset hvordan vi kan tolke halvnormalplott på en ny måte med kunnskap om aliasstrukturen til PB design. Her ser de etter grupperinger av effektene, som kan antyde antall tofaktorsamspill for eksempel.

Ny metode

I prosjektet som jeg skrev i høst introduserte vi en ny metode (Niemi 2007). Den bygger på metoden med tolkning av plott til Tyssedal og Samset (1997). Forskjellen er at det i den nye metoden blir brukt scatterplott av absoluttverdien til effektene i motsetning til normal eller halvnormalplott. I tillegg brukes dendogramplott for å skille effekter fra hverandre i den nye metoden, noe som det også er endel av i denne oppgaven. I denne oppgaven ser vi også på ortogonalisering for å finne forventningsrette hovedeffekter og tofaktorsamspill. Dette blir presentert senere i oppgaven i kapittel 4.

Kapittel 3

16 forsøks design

Tradisjonelt har man brukt 2_{IV}^{8-4} designet hvis man ville ha 16 forsøk og projektivitet 3. Dette tradisjonelle designet er et (16,8,3) design. I denne oppgaven skal vi se på andre design som i noen situasjoner kan være bedre enn dette. Vi skal se på et (16,12,3) design og et annet design som er et (16,8,3) design. Før vi presenterer disse gir vi en innføring i konstruksjonen og det konfunderende mønsteret til det tradisjonelle 2_{IV}^{8-4} designet.

3.1 Konfunderende mønster til 2_{IV}^{8-4}

Box, Hunter og Hunter (1978) skriver om generering av 2_{IV}^{8-4} design ved hjelp av foldover. Noen av de vanligste design vi har er enten resolusjon *III*, *IV* eller *V* design. Lengden på det korteste ordet i den definerende relasjonen til designet kalles resolusjonen til designet. Dette blir betegnet med romertall. Hvis det korteste ordet er *ABCD*, har vi et resolusjon *IV* design. Men hvis *ABC* er det korteste ordet, har vi et resolusjon *III* design. I resolusjon *III* design er hovedeffekter konfunderte med tofaktorsamspill. I resolusjon *V* design er hovedeffektene bare konfunderte med firefaktorsamspill. I tillegg er tofaktorsamspill konfunderte med trefaktorsamspill. Resolusjon *IV* design er en mellomting mellom disse. Hovedeffekter er konfunderte med trefaktorsamspill, og tofaktorsamspill er konfunderte med tofaktorsamspill. Det er altså mulig å skille hovedeffekter ut, men tofaktorsamspill må man skille inn i grupper med like tofaktorsamspill.

Konstruksjon av 2_{IV}^{8-4} ved hjelp av “foldover” gjør man ved å ha de første 8 radene som i 2_{III}^{7-4} design og i tillegg en 8. kolonne med bare pluss-tegn. De siste 8 radene er de første åtte radene, men med motsatt fortegn. Dvs. at rad nr 9 er rad 1 med

motsatt fortegn, osv. Sammenligner vi med tabell 3.1 ser vi at kolonne nummer fire, D -kolonnen, egentlig er kolonnen med halvparten pluss-tegn og halvparten minus-tegn. Alias-mønster til resolusjon IV design.

1. Gruppe med 8 forsøk, 2^{8-5} . $I_8 = D = ABF = ACG = BCH = ABCE$
2. Gruppe med 8 forsøk, 2^{8-5} . $I_8 = -D = -ABF = -ACG = -BCH = ABCE$

Alias-mønsteret til hele forsøket med 16 forsøk blir da:

$$\begin{aligned}
 I_{16} &= ABCE = ABDF = ACDG = BCDH = CDEF \\
 &= BDEG = ADEH = BCFG = ACFH = ABGH \\
 &= AEFH = BEFH = CEGH = DFGH = ABCDEFGH
 \end{aligned} \tag{3.1}$$

3.2 Design med 16 enkeltforsøk

I tabell 2.5 over projektiviteten til PB design ser vi at det metta PB designet med 16 enkeltforsøk har projektivitet 2. 2_{IV}^{8-4} design har projektivitet 3, og har derfor bedre projektivitetsegenskaper enn det metta forsøket. Box og Tyssedal (2001) introduserer endel andre 16 forsøks design med projektivitet 3 som vi skal se nærmere på her. Dette er design der de første åtte kolonnene er like som i 2_{IV}^{8-4} designet, men der det er flere kolonner lagt til. Bakgrunnen for dette er at Box og Tyssedal (1996) viser hvordan man kan plukke ut 8 kolonner fra 16 forsøks saturerte design som gir (16,8,3) projektivitet. Senere viser de at hvis man bruker andre ortogonale arrays, fra Hall (1961), kan man få design med 16 enkeltforsøk som er (16,12,3) design og (16,14,3) design (Box & Tyssedal 2001).

Det tradisjonelle designet (PB designet) for 16 forsøk finnes i tabell 3.1. I de første åtte kolonnene ser vi 2_{IV}^{8-4} designet. Den vertikale doble streken mellom H og J skiller disse åtte første kolonnene fra resten. (16,8,3)-designet gir et gjentatt 2^3 forsøk i alle de 56 valgene av $P = 3$ faktorer fra de $k=8$ faktorene som testes. De design som gir et (16,12,3) screen kalles H_2 , H_3 og H_4 . Og det som gir et (16,14,3) screen kalles H_5 . H_3 , H_4 og H_5 finnes i appendiks A.

3.3 H_2

I denne oppgaven ser vi i hovedsak på H_2 , se tabell 3.2. Grunnen til at vi heller vil bruke H_2 design enn H_1 design er på grunn av antall kolonner som gir et

	a	b	c	d	abc	abd	acd	bcd	ab	ac	ad	bc	bd	cd	abcd	
	<i>I</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>J</i>	<i>K</i>	<i>L</i>	<i>M</i>	<i>P</i>	<i>Q</i>	<i>R</i>
1	+	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+
2	+	+	-	-	-	+	+	+	-	-	-	-	+	+	+	-
3	+	-	+	-	-	+	+	-	+	-	+	+	-	-	+	-
4	+	+	+	-	-	-	-	+	+	+	-	-	-	-	-	+
5	+	-	-	+	-	+	-	+	+	+	-	+	-	+	-	-
6	+	+	-	+	-	-	+	-	+	-	+	-	-	+	-	+
7	+	-	+	+	-	-	+	+	-	-	-	+	+	-	-	+
8	+	+	+	+	-	+	-	-	-	+	+	-	+	-	-	-
9	+	-	-	-	+	-	+	+	+	+	+	-	+	-	-	-
10	+	+	-	-	+	+	-	-	+	-	-	+	+	-	-	+
11	+	-	+	-	+	+	-	+	-	-	+	-	-	+	-	+
12	+	+	+	-	+	-	+	-	-	+	-	+	-	+	-	-
13	+	-	-	+	+	+	+	-	-	+	-	-	-	-	+	+
14	+	+	-	+	+	-	-	+	-	-	+	+	-	-	+	-
15	+	-	+	+	+	-	-	-	+	-	-	-	+	+	+	-
16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Tabell 3.1: H_1 , der de første åtte kolonnene gir et 2_{IV}^{8-4} design.

	a	b	c	d	abc	abd	acd	bcd	ab	ac	bc					
	<i>I</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	J_2	K_2	L_2	M_2	P_2	Q_2	R_2
1	+	-	-	-	-	-	-	-	-	+	-	-	-	+	+	+
2	+	+	-	-	-	+	+	+	-	-	+	-	-	-	-	+
3	+	-	+	-	-	+	+	-	+	-	-	+	-	-	+	-
4	+	+	+	-	-	-	-	+	+	-	-	-	+	+	-	-
5	+	-	-	+	-	+	-	+	+	+	+	+	-	+	-	-
6	+	+	-	+	-	-	+	-	+	+	+	-	+	-	+	-
7	+	-	+	+	-	-	+	+	-	+	-	+	+	-	-	+
8	+	+	+	+	-	+	-	-	-	-	+	+	+	+	+	+
9	+	-	-	-	+	-	+	+	+	-	+	+	+	+	+	+
10	+	+	-	-	+	+	-	-	+	+	-	+	+	-	-	+
11	+	-	+	-	+	+	-	+	-	+	+	-	+	-	+	-
12	+	+	+	-	+	-	+	-	-	+	+	+	-	+	-	-
13	+	-	-	+	+	+	+	-	-	-	-	-	+	+	-	-
14	+	+	-	+	+	-	-	+	-	-	-	+	-	-	+	-
15	+	-	+	+	+	-	-	-	+	-	+	-	-	-	-	+
16	+	+	+	+	+	+	+	+	+	+	-	-	-	+	+	+

Tabell 3.2: H_2 , der de første åtte kolonnene gir et 2_{IV}^{8-4} design.

	ad	bd	cd	abcd	
J_2	+	+	-	+	$\times \frac{1}{2}$
K_2	-	+	-	-	
L_2	+	-	-	-	
M_2	-	-	-	+	

Tabell 3.3: Tabell for utregning av J_2, K_2, L_2 og M_2 til H_2 .

$P=3$ design. Fra H_1 kan vi lage et (16,8,3) screen, mens det fra H_2 er mulig å konstruere (16,12,3) screen. Kolonnene J_2, K_2, L_2 og M_2 i (16,12,3) screenet er funnet fra ligningene i tabell 3.3. Ser vi nærmere på J_2 ser vi at den kan skrives som i (3.2).

$$J_2 = \frac{1}{2}(ad + bd - cd + abcd) \quad (3.2)$$

Fra denne ligningen og de tre ligningene til K_2, L_2 og M_2 går det an å finne generatorer for H_2 . Ganger vi ligning (3.2) med J_2 på begge sider og de andre med respektive K_2, L_2 eller M_2 får vi fire generatorer. I tillegg får vi fire generatorer fra $E = ABC, F = ABD, G = ACD$ og $H = BCD$.

$$\begin{aligned}
I &= \frac{1}{2}(ADJ_2 + BDJ_2 - CDJ_2 + ABCDJ_2) \\
&= \frac{1}{2}(-ADK_2 + BDK_2 - CDK_2 - ABCDK_2) \\
&= \frac{1}{2}(ADL_2 - BDL_2 - CDL_2 - ABCDL_2) \\
&= \frac{1}{2}(-ADM_2 - BDM_2 - CDM_2 + ABCDM_2) \\
&= ABCE \\
&= ABDF \\
&= ACDG \\
&= BCDH
\end{aligned} \quad (3.3)$$

Fra de fire nederste ligningene ser vi allerede at flere tofaktorsamspill er fullstendig konfunderte med hverandre. Resten av tofaktorsamspillene som er konfunderte med hverandre finner vi ved å gange to og to eller tre og tre av generatorene sammen. Ved å gange $I = \frac{1}{2}(-ADK_2 + BDK_2 - CDK_2 - ABCDK_2)$ og $I = \frac{1}{2}(ADL_2 - BDL_2 - CDL_2 - ABCDL_2)$ sammen finner vi sammenhengen mellom A, B, L_2 og K_2 .

$$\begin{aligned}
I &= \frac{1}{4}(-ADK_2 + BDK_2 - CDK_2 - ABCDK_2)(ADL_2 - BDL_2 - CDL_2 - ABCDL_2) \\
&= \frac{1}{4}(4ABL_2K_2) = ABL_2K_2
\end{aligned} \tag{3.4}$$

Ganger vi andre generatorer sammen får vi andre nye generatorer. Her er en liste med noen flere generatorer med fire faktorer. Disse er alt du trenger for å finne alle tofaktorsamspill som er konfunderte med hverandre. I denne lista er totallet ikke med på J, K, L eller M .

$$\begin{aligned}
I &= ABCE = ABDF = ABGH = ABLK = ABJM = ACDG \\
&= ACFH = ACKM = -ACJL = ADEH = ACFG = AELM \\
&= -AEJK = BCDH = BCLM = BCFG = -BCJK = BDEG \\
&= BEKM = -BEJL = CDEF = CELK = CEJM = DFLK \\
&= DFJM = DGKM = -DGJL = -DHJK = DHLM
\end{aligned} \tag{3.5}$$

Senere skal vi også se at man kan regne ut aliasstrukturen til H_2 . Der vises også en mer oversiktlig tabell over hvilke tofaktorsamspill som er konfunderte med hvilke hovedeffekter.

Box og Tyssedal (2001) påpeker en interessant egenskap til $(16,12,3)$ screenet vi får fra H_2 . Vi deler de 12 kolonnene i tre grupper med fire kolonner i hver: $S_1 = (A, B, C, E)$, $S_2 = (D, F, G, H)$ og $S_3 = (J, K, L, M)$. Da vet vi allerede at S_1 og S_2 danner et 2_{IV}^{8-4} design. Men det viser seg at dette også stemmer for de åtte kolonnene som S_1 og S_3 gir og de åtte kolonnene som dannes fra S_2 og S_3 . $(16,12,3)$ screenet kan altså danne et 2_{IV}^{8-4} design fra hvilke som helst to av disse tre gruppene. Senere ser vi at vi får problemer hvis vi har modeller som kombinerer faktorer fra alle disse tre gruppene, men hvis vi velger faktorer fra to av de tre gruppene går det greit.

3.3.1 Aliasmatrise til H_2

Vi er her interessert i å finne aliasmatrisa til H_2 . Den er vi interessert i, fordi den kan si oss noe om hvilke hovedeffekter og tofaktorsamspill som påvirker hverandre. Vi vet allerede hvilke tofaktorsamspill som er konfunderte med hvilke andre tofaktorsamspill. Men vi vil gjerne vite noe om forholdet mellom hovedeffektene

og tofaktorsamspillene også. Informasjonen fra aliasmatrisa og informasjonen om tofaktorsamspillene vil vi senere bruke til å tolke scatterplott med hovedeffekt-kontraster og scatterplott med tofaktorsamspillskontraster.

Har man en korrekt modell som har både hovedeffekter og tofaktorsamspill kan man skrive den som $E(Y) = X\beta + X_1\beta_1$. Der X er designmatrisa og X_1 er matrisa med tofaktorsamspillene. Som forklart i kapittel 2 kan man da iflg. Lin og Draper (1993) finne aliasmatrisa. Den er gitt ved:

$$A = (X^T X)^{-1} X^T X_1 \quad (3.6)$$

Hvis X blir satt til å være designmatrisa med hovedeffekter og X_1 blir satt til å være kolonnene med de $\binom{12}{2}=66$ tofaktorsamspillene man får i et design med 12 hovedfaktorer. Aliasmatrisen viser konfunderende mønster mellom hovedeffekter og tofaktorsamspill.

I et 12 enkeltforsøks PB design så er alle hovedeffekter konfunderte med alle tofaktorsamspill, unntatt de tofaktorsamspillene som er involverte med hovedeffekten selv. I disse 16 forsøksdesignene som vi ser på nå er det ikke sånn. Her er det et litt enklere aliasystem. Alle tofaktorsamspill er konfunderte med 4 hovedeffekter og alle hovedeffekter er konfunderte med 16 tofaktorsamspill. En oversikt kan sees i tabell 3.4. Her ser vi hvilke tofaktorsamspill som “hører” til hvilke faktorer. Den nederste boksen viser hvilke tofaktorsamspill som ikke har noen faktorer de er konfunderte med. I tillegg er de tofaktorsamspillskolonnene som er på samme rad enten helt like eller bare med motsatte fortegn. F.eks. er $DJ = -FM = -GL = -HK$

Som for PB design med 12 enkeltforsøk er konfunderingen av samme størrelse for alle. I dette designet er den $\pm\frac{1}{2}$. Det betyr at de hovedeffektene som blir påvirket av et tofaktorsamspill vil legge seg sammen på et scatterplott av absoluttverdien av effekten til hovedfaktoren. Hvis for eksempel A , B , C og E legger seg sammen, indikerer dette at minst et av tofaktorsamspillene som hører til disse faktorene er aktive. De fire forskjellige tofaktorsamspillene (16 totalt, der fire og fire er like) som er konfunderte med den aktuelle gruppen hovedfaktorer går det da an å se nærmere på.

3.4 H_3 , H_4 og H_5 design

Når det gjelder de andre designene, dvs. H_3 , H_4 og H_5 design, vil aliasstrukturen i disse bli endel mer komplisert enn for H_2 . Designene er i appendiks A. Aliasstrukturen er ikke så enkel og systematisk som for H_2 . For disse designene er det ikke de

<i>Faktor</i>	<i>Tofaktorsamspill</i>
A, B, C, E	$DJ, FM, GL, HK,$ $DK, FL, GM, HJ,$ $DL, FK, GJ, HM,$ DM, FJ, GK, HL
D, F, G, H	$AJ, BM, CL, EK,$ $AK, BL, CM, EJ,$ $AL, BK, CJ, EM,$ AM, BJ, CK, EL
J, K, L, M	$AD, BF, CG, EH,$ $AF, BD, CH, EG,$ $AG, BH, CD, EF,$ AH, BG, CF, DE
	$AB, CE, DF, GH, JM, KL,$ $AC, BE, DG, FH, JL, KM,$ AE, BC, DH, FG, JK, LM

Tabell 3.4: Konfundering mellom faktorer og tofaktorsamspill i H_2 der alle tofaktorsamspill på samme rad er like eller bare har forskjellig fortegn.

samme tofaktorsamspillene som hører til de samme hovedeffektene sånn som for H_2 . Dette gjør at det ser mer komplisert ut, men kan også gjøre analysen lettere. Et eksempel på dette ser vi i analysen av H_6 designet.

3.5 $H_6(16,8,3)$ screen

Noe vi prøver å unngå når vi jobber med tonivåforsøk er konfunderinger. Har vi få forsøk vil det alltid være konfundering mellom et eller annet, men klarer vi å få fullstendige konfunderinger mellom hovedeffekter og tofaktorsamspill vekk er det en god ting. I disse 16 forsøks designene som vi ser på i denne oppgaven er det bare delvis konfundering mellom hovedeffekter og tofaktorsamspill. Slipper vi i tillegg fullstendige konfunderinger mellom tofaktorsamspill er det enda bedre. Hallgeir Grinde (2004) tok utgangspunkt i H_5 (se appendiks A.3) som er et $(16,14,3)$ screen, og laget fra dette et $(16,8,3)$ screen uten fullstendig konfundering mellom tofaktorsamspill. Dette ser vi i tabell 3.5. Her er de fire første hovedeffektene, A, B, C og D , like som for alle PB 16 design. De neste fire kolonnene, G_5, J_5, L_5 og P_5 er tatt fra H_5 på en slik måte at det ikke blir konfundering mellom tofaktorsamspill. Minuset med dette designet er at det blir færre faktorer vi ser på. Her er det bare 8 i motsetning til 12 i H_2 designet som vi har sett på tidligere. Men

fordelen ved å slippe konfundering mellom tofaktorsamspill er stor.

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	G_5	J_5	L_5	P_5
1	-	-	-	-	+	+	+	-
2	+	-	-	-	+	+	+	+
3	-	+	-	-	-	+	-	-
4	+	+	-	-	-	-	+	+
5	-	-	+	-	+	-	-	+
6	+	-	+	-	-	-	-	+
7	-	+	+	-	+	-	+	-
8	+	+	+	-	-	+	-	-
9	-	-	-	+	-	+	-	+
10	+	-	-	+	+	-	-	-
11	-	+	-	+	-	-	+	+
12	+	+	-	+	+	-	-	-
13	-	-	+	+	-	-	+	-
14	+	-	+	+	-	+	+	-
15	-	+	+	+	+	+	-	+
16	+	+	+	+	+	+	+	+

Tabell 3.5: H_6 , utvalgte kontraster fra H_5 for å danne et (16,8,3) screen

De definerende relasjonene i dette designet er:

$$\begin{aligned}
 I &= \frac{1}{2}(-ACG_5 + ADG_5 + BCG_5 + BDG_5) \\
 &= \frac{1}{2}(ACJ_5 + BCJ_5 + CDJ_5 - ABCDJ_5) \\
 &= \frac{1}{2}(CDL_5 + ACDL_5 - BCDL_5 + ABCDL_5) \\
 &= \frac{1}{2}(-ADP_5 + BDP_5 + ACDP_5 + BCDP_5)
 \end{aligned} \tag{3.7}$$

I tabell 3.6 ser vi en tabell over konfunderingene mellom hovedeffekter og tofaktorsamspill for H_6 . Her ser vi at det ikke er like mye system som vi ser i den samme tabellen for H_2 , tabell 3.4. Hver hovedeffekt har seks eller åtte tofaktorsamspill som den er konfundert med. Konfunderingen er $\pm\frac{1}{2}$ som for H_2 . Ser vi på det fra tofaktorsamspillet synspunkt har alle tofaktorsamspill to eller tre hovedeffekter som de er konfunderte med. Unntaket er AB som ikke er konfundert med noen hovedeffekter. Selv om dette er en mer komplisert aliasstruktur kan dette være fordelaktig ved analysen.

<i>Faktor</i>	<i>Tofaktorsamspill</i>
<i>A, B</i>	<i>CG, CJ, DG, DP, JL, LP</i>
<i>C</i>	<i>AG, AJ, BG, BJ, DJ, DL, GP, LP</i>
<i>D</i>	<i>AG, AP, BG, BP, CJ, CL, GL, JP</i>
<i>G</i>	<i>AC, AD, BC, BD, CP, DL, JL, JP</i>
<i>J</i>	<i>AC, AL, BC, BL, CD, DP, GL, GP</i>
<i>L</i>	<i>AJ, AP, BJ, BP, CD, CP, DG, GJ</i>
<i>P</i>	<i>AD, AL, BD, BL, CG, VL, DJ, GJ</i>
	<i>AB</i>

Tabell 3.6: Konfundering mellom faktorer og tofaktorsamspill i H_6 .

Kapittel 4

Ortogonalisering

Et problem som oppstår når vi ser på modeller som inneholder en eller to hovedeffekter og et eller to tofaktorsamspill er å tolke kontrastene. Problemet er at effektene vi regner ut ikke blir forventningsrette, og scatterplottene vi får blir vanskelige å tolke. En løsning på dette vil være å prøve å finne forventningsrette effekter. I dette kapitlet ser vi på en mulig måte å gjøre dette på.

La sammenhengen mellom Y og X_1 være gitt ved $Y = X_1\theta_1 + \varepsilon$, der Y , θ_1 og ε er vektorer og X_1 er en designmatrise med p_1 kolonner. Da vet vi fra regresjonsanalysen at vi kan regne ut:

$$\hat{\theta}_1 = (X_1^T X_1)^{-1} X_1^T Y \quad (4.1)$$

Men hvis modellen egentlig er $Y = X_1\theta_1 + X_2\theta_2 + \varepsilon$, med et ekstra ledd, X_2 , som ikke er ortogonal med X_1 , blir det vanskeligere. X_2 har p_2 kolonner. Dette kan f.eks. være tofaktorsamspill som er aktive i tillegg til hovedeffektene. $\hat{\theta}_1$, utregnet som (4.1), blir da skjev fordi vi ikke tar med X_2 i utregningen. Box og Draper (1987) har skrevet om dette, og store deler av utledningen i dette kapitlet er tatt derfra. Situasjonen de ser på er når $X_1^T X_2 \neq 0$. Dvs. når X_1 og X_2 ikke er ortogonale, som med vår X_1 og X_2 . Målet de har er å finne forventningsrette estimat θ_1 og θ_2 . Hvis $X_1^T X_2 = 0$, dvs. at X_1 og X_2 er ortogonale, kan vi se på $Y = X_1\theta_1$ og $Y = X_2\theta_2$ hver for seg. Da finner vi forventningsrette estimat for θ_1 og θ_2 ved å bruke MKM på X_1 og X_2 hver for seg.

La oss se på $Y = X_1\theta_1$. Residualene $Y - \hat{Y} = Y - X_1\hat{\theta}_1$ er ortogonale til alle kolonnene i X_1 . Kaller vi $Y - \hat{Y}$ for Y_1 er denne en komponent av Y som er ortogonal til et rom av p_1 dimensjoner.

$$\begin{aligned}
Y_{\cdot 1} &= Y - \hat{Y} = Y - X_1\theta_1 \\
&= Y - X_1(X_1^T X_1)^{-1}X_1^T Y \\
&= (1 - R_1)Y
\end{aligned} \tag{4.2}$$

Der $R_1 = X_1(X_1^T X_1)^{-1}X_1^T$.

På samme måte kan vi se på matriser. Da har vi f.eks. en matrise X_2 , i stedet for Y , som inneholder kolonner av ± 1 for forskjellige kontraster. Likt som tidligere har vi da en komponent $X_{2\cdot 1} = X_2 - \hat{X}_2$ som er en komponent av X_2 der alle p_2 kolonner er ortogonale til alle p_1 kolonner i X_1 . En videre utregning med X_2 og $X_{2\cdot 1}$ i stedet for Y og $Y_{\cdot 1}$ blir da:

$$\begin{aligned}
X_{2\cdot 1} &= (1 - R_1)X_2 \\
&= X_2 - R_1X_2 \\
&= X_2 - X_1A
\end{aligned} \tag{4.3}$$

der

$$A = (X_1^T X_1)^{-1}X_1^T X_2 \tag{4.4}$$

Ved å skrive om $Y = X_1\theta_1 + X_2\theta_2 + \varepsilon$ til:

$$\begin{aligned}
Y &= X_1(\theta_1 + A\theta_2) + (X_2 - X_1A)\theta_2 + \varepsilon \\
&= X_1\theta + X_{2\cdot 1}\theta_2 + \varepsilon
\end{aligned} \tag{4.5}$$

der $\theta = \theta_1 + A\theta_2$, får vi to X matriser som er ortogonale, X_1 og $X_{2\cdot 1}$. Da kan man regne ut $X_{2\cdot 1}$ og finne forventningsrett θ_2 . For å finne θ_1 kan man da enten bytte om på X_1 og X_2 når man regner ut A og $X_{2\cdot 1}$, eller regne det ut fra $Y = X_1(\theta_1 + A\theta_2) + \varepsilon$. For å bruke den siste metoden må man bruke MKM på $Y = X_1\theta$, der $\theta = \theta_1 + A\theta_2$. Deretter kan man regne ut θ_1 ved å sette inn forventningsrett $\hat{\theta}_2$, og de utregnede A og θ .

Problemet, eller utfordringen, forbundet med denne metoden å ortogonalisere på, er hva man skal velge som X_1 og X_2 . Dette kommer vi nærmere inn på i seksjon 4.1.

4.1 Metode for ortogonaliseringen

Vi har i starten av dette kapitlet sett på utfordringene ved å regne ut forventningen til θ hvis vi har to matriser som ikke er ortogonale. Vi vil her beskrive metoden som kan brukes for å bruke denne teorien i forhold til PB 12 forsøks design og 16 forsøks design. Som beskrevet tidligere lønner det seg å skrive ligningen som:

$$Y = X_1(\theta_1 + A\theta_2) + (X_2 - X_1A)\theta_2 + \varepsilon \quad (4.6)$$

$$= X_1\theta + X_{2.1}\theta_2 + \varepsilon \quad (4.7)$$

der $A = (X_1^T X_1)^{-1} X_1^T X_2$ og $X_{2.1} = X_2 - X_1 A$.

Siden vi velger X_1 og X_2 er det mulig å regne ut A . Videre vil vi gjerne regne ut $X_{2.1}$ og finne den forventningsrette θ_2 ved å regne ut $\theta_2 = (X_{2.1}^T X_{2.1})^{-1} X_{2.1}^T Y$.

Problemet vi da står overfor er hva vi skal velge som X_1 og X_2 ? Siden vi har design med enten 12 eller 16 forsøk kan vi ikke ha altfor mange kolonner før vi har brukt opp dimensjonen i matrisene. Et mål på om det er for mange kolonner i X_1 og X_2 får man ved å sjekke om $n_{X_1} + n_{X_2} \leq n - 1$. Men selv om dette stemmer kan det hende at det er andre problem med designet som gjør at du ikke får ut noe svar. Dessuten må de aktive hovedfaktorene og tofaktorsamspillene være med i enten X_1 eller X_2 for at vi skal klare å finne riktige kontraster. Men selv uten alle aktive effekter er det mulig å få endel hint om hvilke effekter som er viktige. Hvis vi ikke vet hvilke faktorer som spiller inn blir dette altså en litt tungvint måte å gjøre det på. Har vi derimot en god indikasjon fra scatterplott av hovedeffekter og tofaktorsamspill på hvilke effekter som kan være interessante å se mer på, er dette en god måte å forbedre plottene. Forbedringen går ut på at plottene blir enklere. Det vil si at de effektene som ikke er aktive havner rundt 0, og de effektene som er aktive havner på det nivået de skal havne på.

En viktig ting å merke seg med de utregnede θ_1 og θ_2 her er at de er halvparten så store som hovedeffekter og tofaktorsamspill som vi vanligvis regner ut. Det betyr at tallene vi får fra ortogonalisering ikke er direkte sammenlignbare med de tallene vi har fra tidligere. Dette er forklart i alle modellene der vi har brukt ortogonalisering.

4.1.1 12 forsøks eksempel

La oss si at vi har en modell

$$Y = A + B + BC + \varepsilon \quad (4.8)$$

Her har vi to hovedeffekter (A og B) og et tofaktorsamspill (BC) i tillegg til noe støy (ε) som kan være av forskjellig størrelse. Først vil vi fra scatterplott av hovedeffektene og aktuelle scatterplott av tofaktorsamspill bestemme hvilke kontraster som ser ut til å være aktuelle kandidater. La oss si at vi ser at A og B ser ut til å være aktive. Så lager vi tofaktorsamspillsplott for A og B og ser at BC er aktiv, men også noen andre tofaktorsamspill ser ut til å kunne være aktive. Da kan vi velge å sette $X_1 = [A, B]$ og så vil vi gjerne velge en X_2 sånn at vi kan finne en θ_2 som er forventningsrett og av interesse. Et problem vi har er at vi ikke kan ta med alle kolonnene vi bare vil i X_2 , vi må velge få nok, ellers fungerer ikke metoden. Det viktige da er at vi får med oss de tofaktorsamspillene som faktisk er aktive. Vi velger altså en X_2 som har få nok kolonner til at dimensjonen ikke blir brukt opp. Etter denne første testinga kan vi da med all sannsynlighet ta vekk noen av de tofaktorsamspillene som vi trodde var aktive fordi de viser seg å ikke være det. For å sjekke kontrastene til A og B kan vi da enkelt bytte rundt, sånn at vi har kolonnene med tofaktorsamspill som X_1 og $X_2 = [A, B]$. Da får vi ut kontraster til A og B som er nærmere de reelle, eller halvparten av de reelle, enn de kontrastene som er konfunderte og som vi har regnet ut tidligere. Det eneste som nå påvirker de er støyen. Så lenge denne støyen ikke er altfor stor vil vi få kontraster som er i nærheten av de reelle. Neste skritt blir da å se på flere av de andre tofaktorsamspillene, samtidig som vi beholder de tofaktorsamspillene som vi ser er aktive. Hvis vi har klart å få med oss de aktive hovedeffektene og tofaktorsamspillene, vil de neste tofaktorsamspillene som vi tar med komme ut ≈ 0 . Hvis vi derimot ikke har valgt riktig vil resultatene våre forandre seg. Da må utgangspunktet granskes på nytt for å vurdere om de riktige aktive hovedeffekter og tofaktorsamspill er med fra starten av.

4.1.2 16 forsøks eksempel

I disse noe større designa vil vi velge X_1 og X_2 på nokså samme måte som for 12 forsøks designet. Den største forskjellen er at det nå er 16 forsøk, og derfor 16 rader. Det betyr at det kan være litt flere kolonner i X_1 og X_2 før vi har brukt opp dimensjonen i matrisa.

I H_2 tilfellet er det i tillegg til tofaktorsamspill som er fullstendig konfunderte med hverandre et annet element som spiller inn. På grunn av generatorene vi får fra J_2, K_2, L_2 og M_2 viser det seg at vi får endel sammenhenger mellom tofaktorsamspill og hovedeffekter som kan være litt uheldig. Ligningene til J_2, K_2, L_2 og M_2

ser vi i ligning (3.3). Ser vi litt nærmere på disse generatorene og kombinerer de litt ser vi hva problemet er.

$$\begin{aligned} DJ - DM &= \frac{1}{2}A + \frac{1}{2}B - \frac{1}{2}C + \frac{1}{2}E - \left(-\frac{1}{2}A - \frac{1}{2}B - \frac{1}{2}C + \frac{1}{2}E\right) \\ &= A + B \end{aligned} \quad (4.9)$$

En liste over resten kan ses i tabell 4.1.

$$\begin{array}{l|l} DJ_2 - DM_2 = A + B & DJ_2 + DM_2 = -C + E \\ DJ_2 - DK_2 = A + E & DJ_2 + DK_2 = B - C \\ DJ_2 - DL_2 = B + E & DJ_2 + DL_2 = A - C \\ DK_2 - DL_2 = -A + B & DK_2 + DL_2 = -C - E \\ DK_2 - DM_2 = B - E & DK_2 + DM_2 = -A - C \\ DL_2 - DM_2 = A - E & DL_2 + DM_2 = -B - C \end{array}$$

Tabell 4.1

Her ser vi at vi får problemer med ortogonaliseringen. Det går nemlig ikke å ha A, B, DJ_2 og DM_2 samtidig i X_1 og X_2 . Eller noen av de andre kombinasjonene. Hvis man ganger med D på begge sider av likhetstegnet får man en del andre kombinasjoner som heller ikke fungerer. For eksempel er $J_2 - M_2 = AD + BD$.

Dette vil bare bli et problem hvis vi har aktive faktorer fra alle de tre gruppene S_1, S_2 og S_3 samtidig. I praksis betyr dette at vi må passe på dersom vi får et tilfelle der det ser ut som om det er aktive faktorer fra alle tre gruppene. Finnes det informasjon om faktorene på forhånd kan det være lurt å putte de 8 faktorene som man anser som viktigst i enten S_1 og S_2 , eller i to av de andre S -ene.

I H_6 er ikke tofaktorsamspill konfundert med hverandre, så vi slipper det problemet. På grunn av de litt mer avanserte definerende relasjonene til dette designet må det være flere faktorer samtidig i X_1 og X_2 før vi får problemer.

Kapittel 5

Resultat

5.1 12 forsøks PB design

I 12 forsøks PB design er konfunderingen mellom hovedeffekter og tofaktorsamspill enten 0 eller $\pm\frac{1}{3}$. Konfunderingen er bare 0 mellom en hovedeffekt og tofaktorsamspillene som inneholder denne hovedeffekten. Derfor vil alltid hovedeffektskontrastene dele seg i to, hvis det er en hovedeffekt og et tofaktorsamspill. Hovedeffekten som er aktiv for seg, og resten av hovedeffektene på samme nivå. (Med mindre man er veldig uheldig med størrelsen på hovedeffekt og tofaktorsamspill.) Videre vil det danne seg to grupperinger med hovedeffekter hvis man har to tofaktorsamspill. Derfor kan man si ganske mye om tofaktorsamspillene som kanskje er aktive selv når man bruker et 12 forsøks PB design. Nedenfor følger en detaljert algoritme for hvordan man kan analysere data der man har brukt 12 forsøks PB design.

5.1.1 Algoritme for 12 forsøks PB design

1. Finn hovedeffektskontrastene til faktorene ved å regne ut $\hat{\theta} = \frac{1}{6}X'Y$. Der X er fra $Y = X\theta$. Ta absoluttverdi av disse og lag scatterplott og dendogram.
2. Fra scatterplott og dendogram skill hovedeffektene som er aktive fra de som ikke er aktive. Dette gjøres ut i fra en tolkning av aliasmatrisen til PB. Hvis det er et eller to tofaktorsamspill skiller de ikke-aktive hovedeffektene seg fra de aktive i hhv. en eller to grupper. Hvis det er tre eller flere tofaktorsamspill, og da mange forskjellige grupperinger i dendogrammet, blir det vanskelig å

tolke hvilke hovedeffekter som er aktive og ikke. Da kan man være nødt til å bruke en annen metode.

3. Regn ut effektene til de tofaktorsamspillene som er interessante å se på.
4. Lag scatterplott og dendogram til absoluttverdien av kontrastene til de tofaktorsamspillene som er interessante.
5. Hvis scatterplott og dendogram av hovedeffektene, og tofaktorsamspillsplott og dendogram samsvarer med hverandre, konkluder med hvilke effekter som er interessante. Hvis ikke gjør flere undersøkelser av andre tofaktorsamspill.
6. Bruk ortogonaliseringsteorien og beregn nye kontraster for hovedfaktorene og de aktuelle tofaktorsamspillene.

Vi har sett på tre modeller der vi har brukt PB 12 forsøks design. Det er $Y = 2A + 2B + BC + \varepsilon$, $Y = A + BC + \varepsilon$ og $Y = 2A + 2C + AB + BC + \varepsilon$. Den første modellen er en modell med arvelighet, tre ledd og tre aktive faktorer. Den andre modellen er en enkel modell, men uten arvelighet og den tredje er en modell med fire ledd og tre aktive faktorer. Oversikt i tabell 5.1. Det å se på modeller uten arvelighet er egentlig nokså uvanlig tradisjonelt sett. Mange metoder bruker derfor arvelighet aktivt i metodene. Ser man derimot litt nærmere på det er det slett ikke sikkert at det alltid er logisk med arvelighet i modellene. Vi har f.eks. en god del eksempler i naturen der dette ikke gjelder. $E = mc$ er et eksempel. På grunn av dette kan det være greit med metoder der man ikke er avhengig av arvelighet for å få noe ut av metoden.

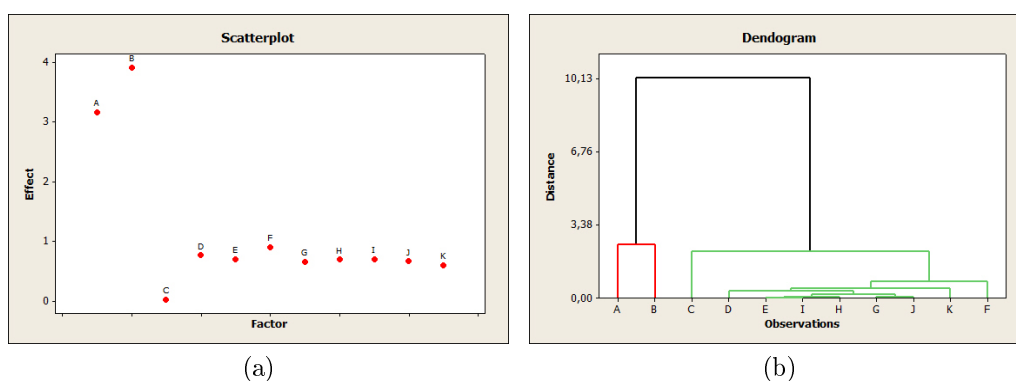
<i>Modell</i>	<i>#Hovedeffekter</i>	<i>#Tofaktorsamspill</i>
$Y = 2A + 2B + BC$	2	1
$Y = A + BC$	1	1
$Y = 2A + 2C + AB + BC$	2	2

Tabell 5.1: Modeller der 12 forsøks PB design er brukt.

I 12 forsøk PB modellene vi har sett på så er faktorene som er brukt $A, B, C, D, E, F, G, H, I, J$ og K . I er her en egen faktor og ikke identitetskolonnen.

5.1.2 Modell: $Y = 2A + 2B + BC + \varepsilon$

Denne modellen har to hovedeffekter, et tofaktorsamspill og tre aktive faktorer. I figur 5.1 ser vi scatterplott og dendrogram av hovedeffektene vi får ved å regne ut θ .



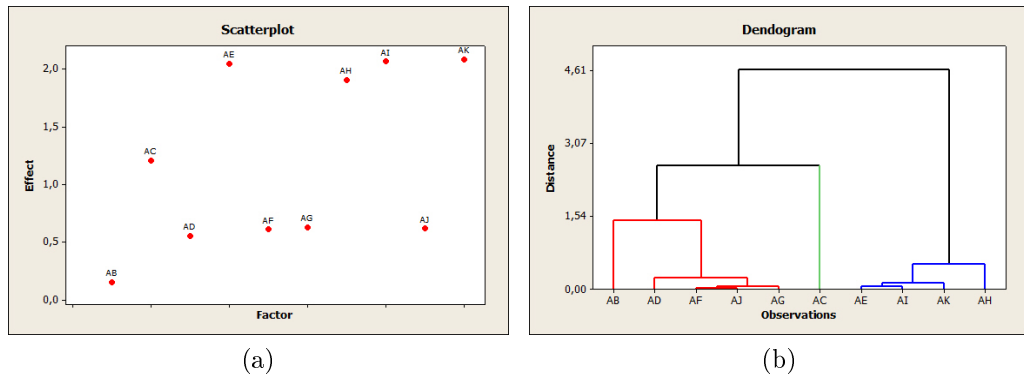
Figur 5.1: Scatterplott og dendrogram av hovedeffektene til modell $Y = 2A + 2B + BC + \varepsilon$, $\varepsilon \sim N(0, 0.25)$

Her ser vi at A og B skiller seg ut. De har en mye høyere kontrast enn resten av hovedeffektene. Det kan se ut som C skiller seg litt ut også, ved at den ikke ser ut til å være påvirket i samme grad som D, \dots, K . Dette kan tyde på at det er et tofaktorsamspill med C som er aktivt. Dette vil påvirke de andre hovedeffektene, men ikke C , siden C er ortogonal på alle tofaktorsamspill med C . Videre vil vi se på scatterplott av tofaktorsamspillene til A og B , fordi A og B er plukket ut som hovedeffekter, og scatterplott av tofaktorsamspillene til C , fordi vi tror det kan være noe med et av disse. Disse plottene kan sees i figur 5.2 til figur 5.4.

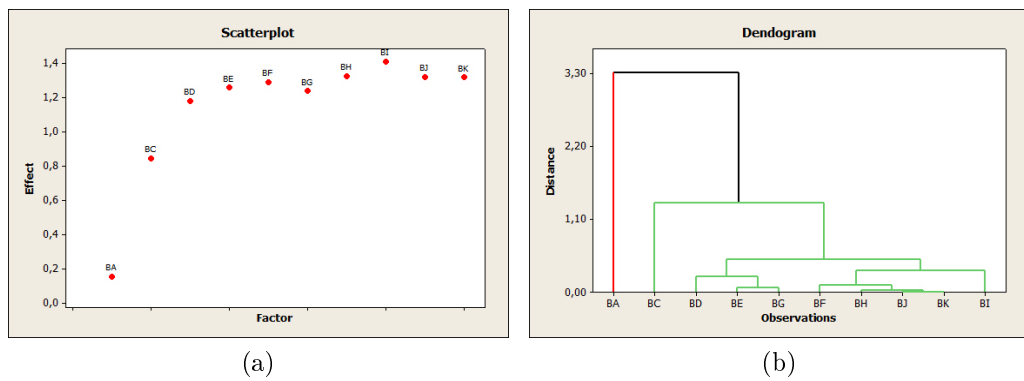
I disse figurene så ser vi at AB , AC og BC skiller seg ut. AB skiller seg ut fordi den er liten, AC og BC skiller seg ut ved å legge seg på et annet nivå enn resten av tofaktorsamspillene i de respektive plottene. AB er ikke konfundert med hovedeffektene A og B . Siden den kommer lavt ut er det grunn til å tro at AB samspillet er neglisjerbart og at det aktive tofaktorsamspillet ikke er konfundert med AB . Både AC og BC oppfyller dette.

Vi vil forbedre tofaktorsamspillsplott og hovedeffektsplottet sånn at det er tydeligere akkurat hva som er signifikant, og sånn at vi kan finne de forventningsrette effekter. For å gjøre dette vil vi ortogonalisere og regne ut nye kontraster. Bruker teorien i kapittel 4.

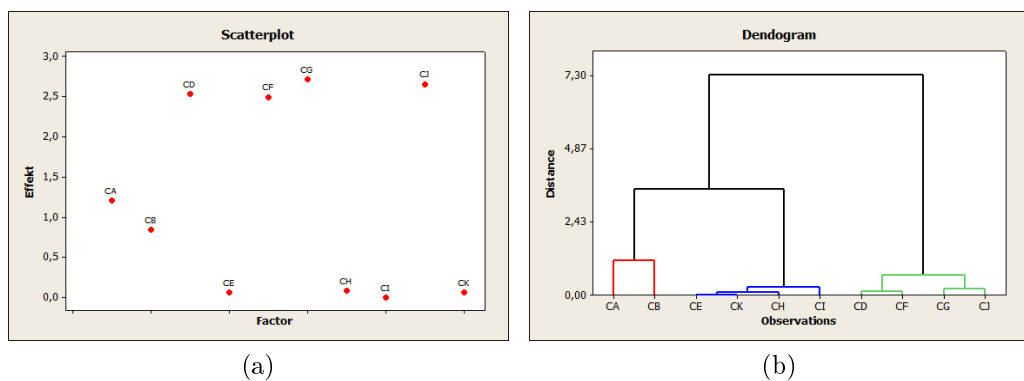
Vi må velge hva vi skal sette som X_1 og X_2 . X_1 velger vi til å være $X_1 = [A, B]$.



Figur 5.2: Scatterplott og dendrogram av tofaktorsamspill med A til modell $Y = 2A + 2B + BC + \varepsilon$, $\varepsilon \sim N(0, 0.25)$



Figur 5.3: Scatterplott og dendrogram av tofaktorsamspill med B til modell $Y = 2A + 2B + BC + \varepsilon$, $\varepsilon \sim N(0, 0.25)$



Figur 5.4: Scatterplott og dendrogram av tofaktorsamspill med C til modell $Y = 2A + 2B + BC + \varepsilon$, $\varepsilon \sim N(0, 0.25)$

Dette fordi vi tror at dette er de aktive hovedeffektene. Siden vi tror at vi har et tofaktorsamspill med C som er aktivt vil vi se hvordan disse fordeler seg, når vi har tatt vekk skjevheten som kommer fra hovedeffektene. Vi vil i grunnen ha en $X_2 = [AC, BC, CD, CE, CF, CG, CH, CI, CJ, CK]$. Men på grunn av begrensningen siden vi bare har 12 forsøk kan vi ikke ha så mange kolonner i X_2 . Derfor vil vi dele det opp i to deler. Først ser vi på $X_2 = [AC, BC, CD, CE, CF]$ og så tar vi resten etterpå. Hvis noen av de første kolonnene viser seg å være viktig, f.eks. AC eller BC , flytter vi denne kolonnen over i X_1 og tar med oss videre. For at dette skal fungere best mulig må nemlig de hovedeffektene og tofaktorsamspillene som er aktive være med i X_1 eller X_2 .

	<i>Før ortogonalisering</i>	<i>Etter 1</i>	<i>Etter 2</i>
AC	-1,21	0,08	
BC	0,84	1,08	1,07
CD	-2,54	0,08	
CE	-0,07	-0,06	
CF	2,50	-0,10	
CG	-2,72		-0,10
CH	-0,09		-0,082
CI	0,01		0,039
CJ	-2,66		-0,072
CK	-0,06		-0,068

Tabell 5.2: Kontraster fra $Y = 2A + 2B + BC + \varepsilon$ før ortogonalisering og forventningsrette effekter etter ortogonalisering.

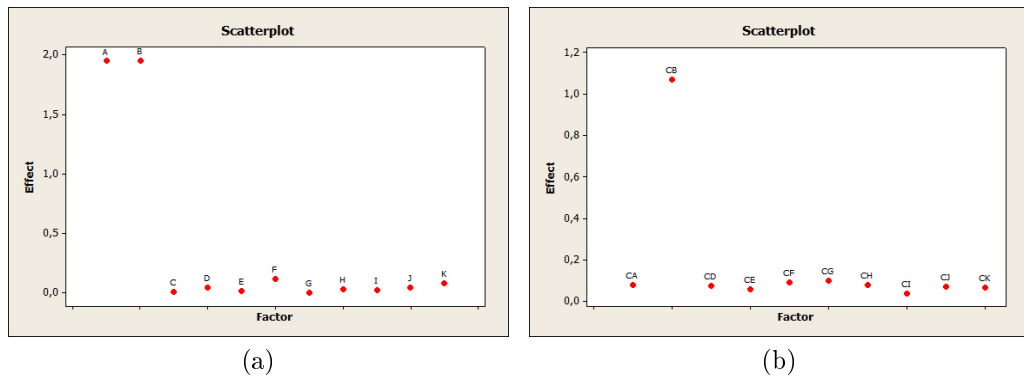
I tabell 5.2 ser vi opprinnelige kontrastverdier utregnet med MKM og de nye verdiene utregnet med ortogonale X matriser. Her ser vi at kontrastene kommer ut enten små, maks -0.10 eller rundt 1. Dette tyder på at vi har funnet det tofaktorsamspillet som er aktivt. Hvis vi ikke hadde kunnet konkludere ut i fra disse tofaktorsamspillene måtte vi sett på andre. Da ville det vært naturlig å se på tofaktorsamspillene til A og B , siden vi ser at disse er hovedeffekter i dette tilfellet.

Vi vil i tillegg til forbedrede tofaktorsamspill med C finne forbedrede hovedeffekter. Så når vi etter denne utregningen ser at det er A , B og BC som er aktive effekter kan vi sette disse i X_1 og ha $X_2 = [C, D, E, F, G, H, I, J, K]$. Her må også hovedeffektene tas i to deler for å ikke bruke opp dimensjonen til matrisa.

I tabell 5.3 ser vi disse forbedrede hovedeffektene. Etter å ha regnet dette ut er det mulig å lage to nye plott. Et for hovedeffektene og et for tofaktorsamspillene med C . Disse er gitt i figur 5.5. Her har vi brukt utregnede tall eller gjennomsnitt av flere utregnede tall der det er aktuelt.

	<i>Før Ortogonalisering</i>	<i>Etter 1</i>	<i>Etter 2</i>
<i>A</i>	3,16	1,91	1,94
<i>B</i>	3,91	1,95	1,83
<i>C</i>	0,03	0,01	
<i>D</i>	-0,77	-0,05	
<i>E</i>	0,71	0,02	
<i>F</i>	0,90	0,12	
<i>G</i>	-0,66	0,00	
<i>H</i>	-0,69		0,03
<i>I</i>	0,71		-0,02
<i>J</i>	-0,67		0,05
<i>K</i>	-0,60		0,08

Tabell 5.3: Kontraster fra $Y = 2A + 2B + BC + \varepsilon$ før ortogonalisering og forventningsrette effekter etter ortogonalisering.

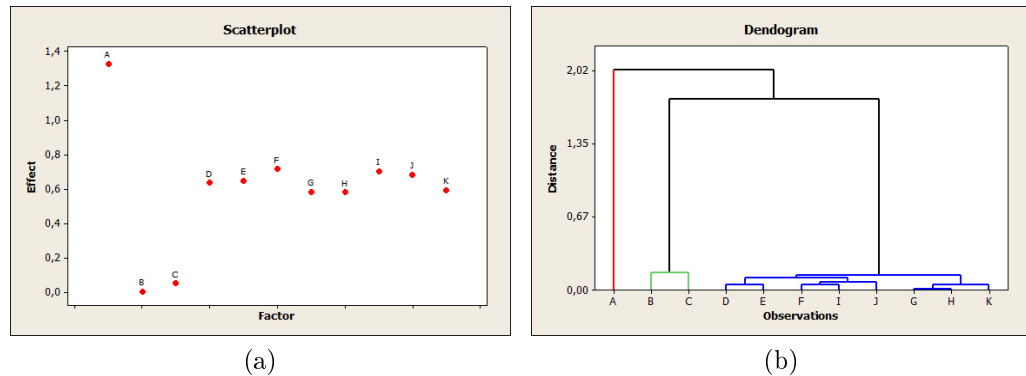


Figur 5.5: Scatterplott av hovedeffektene og tofaktorsamspill med C til modell $Y = 2A + 2B + BC + \varepsilon$, $\varepsilon \sim N(0, 0.25)$ etter ortogonalisering.

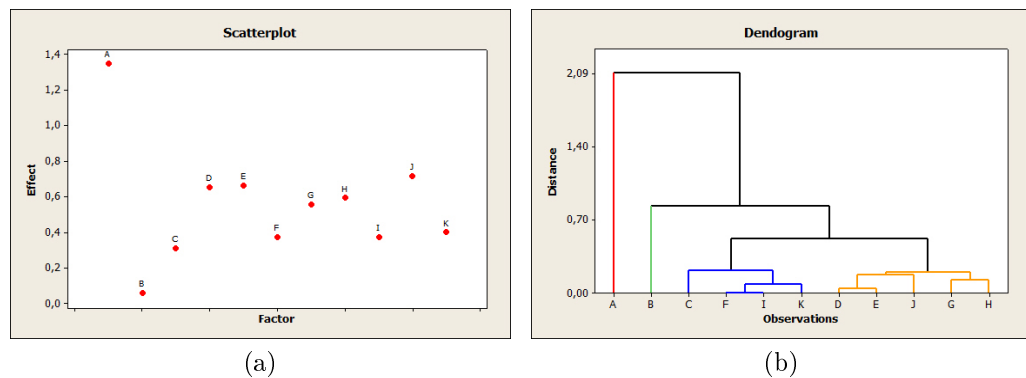
Her ser vi at verdiene til de aktive hovedeffektene og tofaktorsamspillene med C er halvparten så store som vi vanligvis får ved utregning av effekter. Som beskrevet i kapittel 4 kommer dette av måten θ_1 og θ_2 blir regnet ut.

5.1.3 Modell $Y = A + BC + \varepsilon$

Dette er en modell med en hovedeffekt og et tofaktorsamspill, uten arvelighet. Det er tre aktive faktorer i modellen. Støyen varierer vi her mellom $\varepsilon \sim N(0, 0.1)$, $\varepsilon \sim N(0, 0.25)$ og $\varepsilon \sim N(0, 0.5)$. Det minste nivået er et ganske beskjedent støynivå, mens vi skal se at det største kanskje er litt for høyt støynivå.

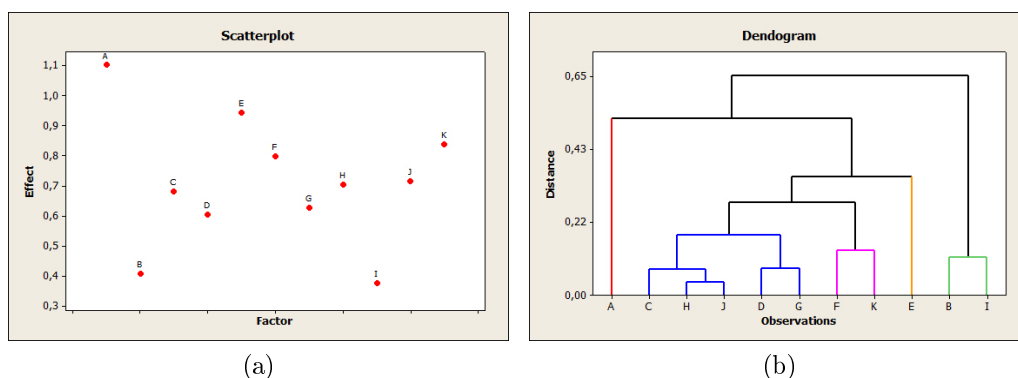


Figur 5.6: Scatterplott og dendrogram av hovedeffektene til modell $Y = A + BC + \varepsilon$, $\varepsilon \sim N(0, 0.1)$



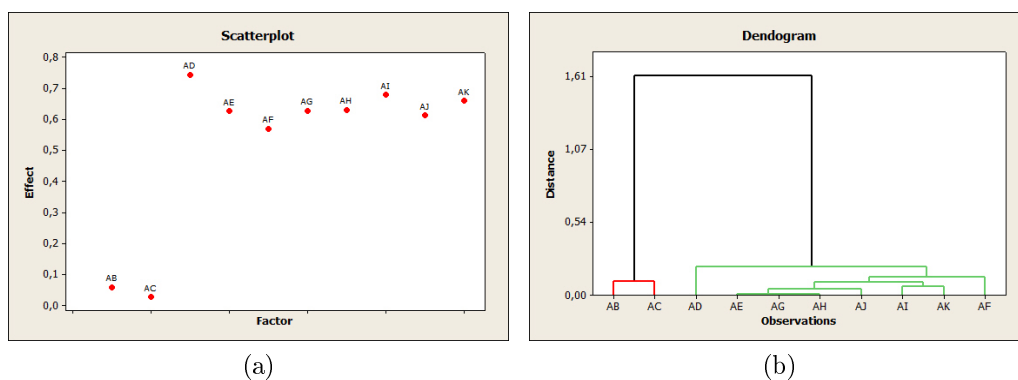
Figur 5.7: Scatterplott og dendrogram av hovedeffektene til modell $Y = A + BC + \varepsilon$, $\varepsilon \sim N(0, 0.25)$

I figur 5.6, figur 5.7 og figur 5.8 ser vi scatterplott og dendrogram av hovedeffektene med de tre ulike støynivåene. På det første plottet med minst støy ser vi tydelig at det er noe spesielt med A. Vi ser også at det er noe spesielt med B og C, siden disse er for seg selv og resten av faktorene samler seg i en klynge mellom A og B og C. Dette kan tyde på at det er et tofaktorsamspill som er aktivt, som påvirker



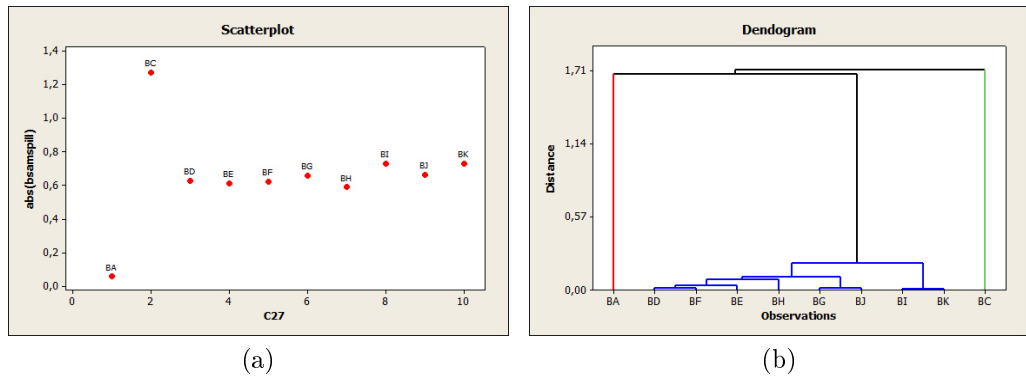
Figur 5.8: Scatterplott og dendrogram av hovedeffektene til modell $Y = A + BC + \varepsilon$, $\varepsilon \sim N(0, 0.5)$

D, \dots, K , men ikke påvirker B og C . I de to andre plottene ser vi også at det er noe spesielt med A , men det er ikke like lett å bedømme hvilke andre faktorer det er noe spesielt med. For alle de tre tilfellene vil vi uansett se nærmere på tofaktorsamspillene med A , fordi denne peker seg ut, og det alltid kan være greit å sjekke tofaktorsamspill med hovedeffekten. Hvis figur 5.6 var det plottet vi hadde fått ville vi valgt å se på tofaktorsamspillene til B og C i tillegg til A . Plottene for modellene med $\varepsilon \sim N(0, 0.1)$ finnes i figur 5.9 til figur 5.11.

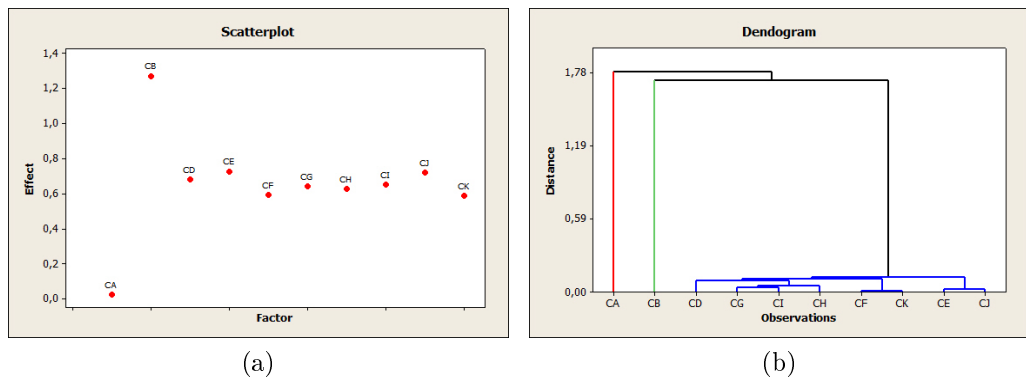


Figur 5.9: Scatterplott og dendrogram av tofaktorsamspill med A til modell $Y = A + BC + \varepsilon$, $\varepsilon \sim N(0, 0.1)$

I disse plottene så ser vi at BC skiller seg klart ut i både scatterplottet med tofaktorsamspill med B og med C . I scatterplottet med tofaktorsamspillene til A ser vi også at det er noe spesielt med AB og AC , men her er det at kontrastene er lave i motsetning til høye. Det tyder på at de ikke er påvirket av noe tofaktorsam-



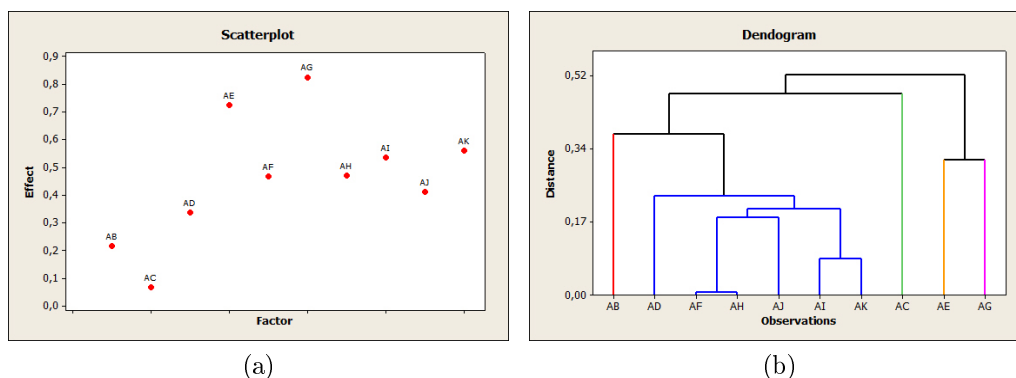
Figur 5.10: Scatterplott og dendrogram av tofaktorsamspill med B til modell $Y = A + BC + \varepsilon$, $\varepsilon \sim N(0, 0.1)$



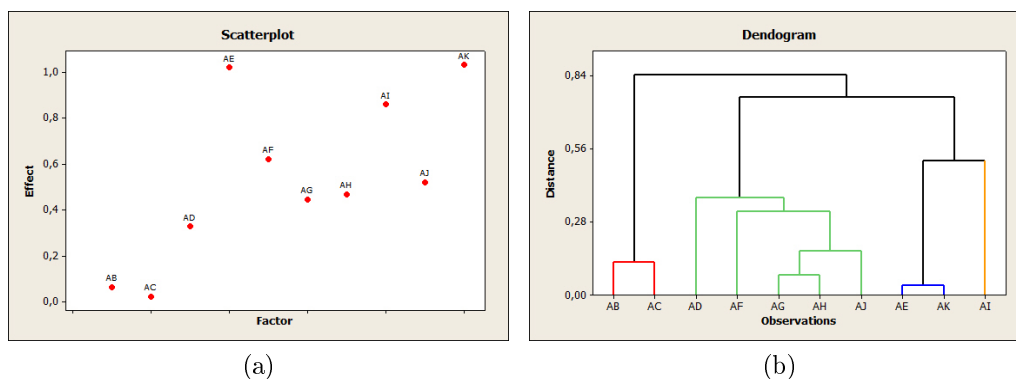
Figur 5.11: Scatterplott og dendrogram av tofaktorsamspill med C til modell $Y = A + BC + \varepsilon$, $\varepsilon \sim N(0, 0.1)$

spill, mens $AD, AE, \dots AK$ tydelig er påvirket av et tofaktorsamspill. Alle de tre scatterplottene peker altså mot at det er noe spesielt med BC .

Hvis situasjonen derimot var litt mer komplisert, som vi kan se at den er i de to tilfellene med støy lik $\varepsilon \sim N(0, 0.25)$ og $\varepsilon \sim N(0, 0.5)$ ville vi først se på scatterplott og dendrogram av tofaktorsamspill med A . Disse kan ses i figur 5.12 og figur 5.13.



Figur 5.12: Scatterplott og dendrogram av tofaktorsamspillene med A til modell $Y = A + BC + \varepsilon$, $\varepsilon \sim N(0, 0.25)$



Figur 5.13: Scatterplott og dendrogram av tofaktorsamspillene med A til modell $Y = A + BC + \varepsilon$, $\varepsilon \sim N(0, 0.5)$

I disse figurene ser vi ikke umiddelbart noen indikasjon på at det er tofaktorsamspillet BC vi er ute etter. Det vi kan se er en indikasjon på at noe påvirker AB og AC slik at de ikke er så høye som de andre i noen av plottene. Og enkleste forklaring på dette er at BC er aktiv.

Når det gjelder tilfellet med minst støy ($\varepsilon \sim N(0, 0.1)$) antar vi at det er A og BC som er interessante. Ved å bruke ortogonalisering kan vi finne enda enklere plott for hovedeffektene og de aktuelle tofaktorsamspillene. Vi bruker derfor $X_1 = [A, BC]$ og vil prøve med tofaktorsamspill til enten B eller C . Vi velger å se på tofaktorsamspillene til C , og $X_2 = [AC, CD, CE, CF, CG, CH, CI, CJ, CK]$ for å finne ut hva slags nivå A og BC egentlig ligger på. Finner da de resultat vi kan se i tabell 5.4. Som i forrige modell bør man gå tilbake til andre tofaktorsamspill, hvis det ser ut som om de nye plottene ikke er bedre enn de man får ved MKM. Aktuelle tofaktorsamspill å se på er da tofaktorsamspillene til A , siden A er en hovedeffekt.

	<i>Før Ortogonalisering</i>	<i>Etter 1</i>	<i>Etter 2</i>
AC	0.03	0.01	
BC	1.27	0.96	0.97
CD	-0.68	-0.02	
CE	-0.73	-0.04	
CF	0.59	-0.03	
CG	-0.64		0.01
CH	0.63		-0.02
CI	-0.66		0.01
CJ	-0.72		-0.03
CK	0.59		-0.04

Tabell 5.4: Kontraster fra $Y = A + BC + \varepsilon$ før ortogonalisering og forventningsrette effekter etter ortogonalisering.

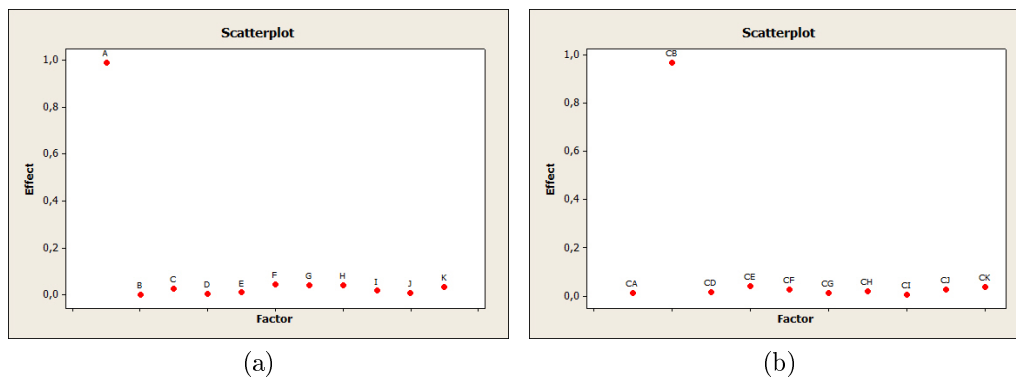
I tillegg kan vi finne hovedeffekter som ikke er påvirket av tofaktorsamspill ved å la $X_1 = [A, BC]$ og $X_2 = [B, C, D, E, F, G, H, I, J, K]$. Disse finner vi i tabell 5.5.

Nye plott for både hovedeffektene og tofaktorsamspillene med C finnes i figur 5.14.

Her ser vi at verdiene er halvparten så store som effekter vanligvis er. Dette kommer av at det er θ_1 og θ_2 som beskrevet i kapittel 4.

	<i>Før Ortogonalisering</i>	<i>Etter 1</i>	<i>Etter 2</i>
<i>A</i>	1.33	0.98	1.00
<i>B</i>	0.00	0.00	
<i>C</i>	0.05	0.03	
<i>D</i>	-0.64	-0.00	
<i>E</i>	0.65	0.01	
<i>F</i>	0.72	0.04	
<i>G</i>	-0.58		0.04
<i>H</i>	-0.58		0.04
<i>I</i>	0.70		0.02
<i>J</i>	-0.68		-0.01
<i>K</i>	-0.60		0.04

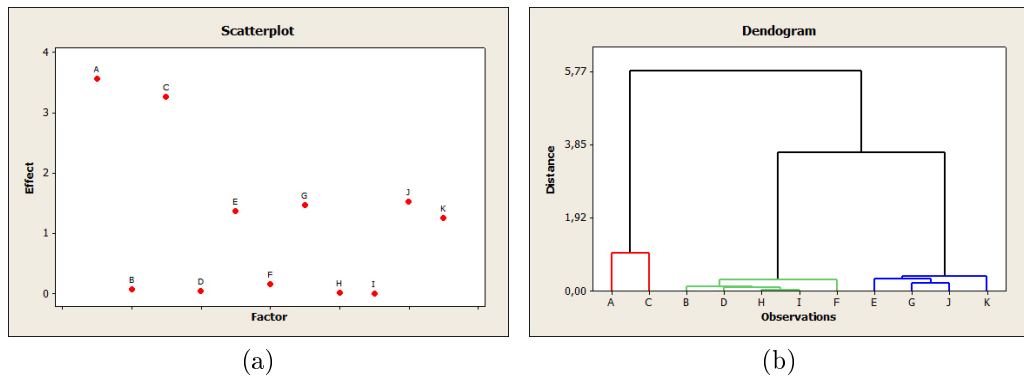
Tabell 5.5: Kontraster fra $Y = A + BC + \varepsilon$ før ortogonalisering og forventningsrette effekter etter ortogonalisering.



Figur 5.14: Scatterplott av hovedeffektene og tofaktorsamspillene med C til modell $Y = A + BC + \varepsilon$, $\varepsilon \sim N(0, 0.1)$ etter ortogonalisering.

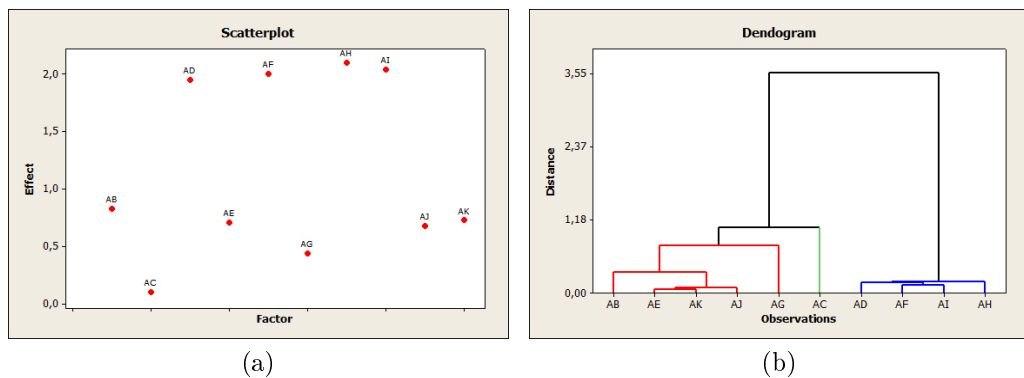
5.1.4 Modell $Y = 2A + 2C + AB + BC + \varepsilon$

Dette er en modell med to hovedeffekter, to tofaktorsamspill og tre aktive faktorer. Støyen er $\varepsilon \sim N(0, 0.25)$. Se scatterplott og dendrogram av hovedeffektene i figur 5.15.



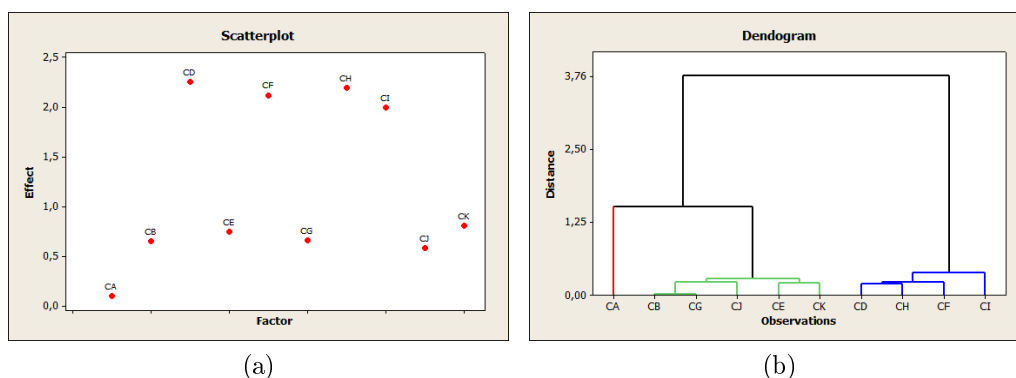
Figur 5.15: Scatterplott og dendrogram av hovedeffektene til modell $Y = 2A + 2C + AB + BC + \varepsilon$, $\varepsilon \sim N(0, 0.25)$

Det vi kan se i figur 5.15 er at A og C kommer ut store, og at resten av faktorene samler seg i to grupper. Dette kan tyde på at to tofaktorsamspill er aktive. Neste skritt blir da å se på tofaktorsamspillene til A og C for å se om vi har en modell med arvelighet, noe som er det vanligste å se på. I figur 5.16 og i figur 5.17 kan vi se scatterplott og dendrogram fra disse tofaktorsamspillene.



Figur 5.16: Scatterplott og dendrogram av tofaktorsamspillene med A til modell $Y = 2A + 2C + AB + BC + \varepsilon$, $\varepsilon \sim N(0, 0.25)$

Ser vi på figur 5.16 og figur 5.17 er det ikke så lett å se hvilke tofaktorsamspill som er aktive. I begge plottene skiller AC seg ut ved å være veldig lav, noe som ikke



Figur 5.17: Scatterplott og dendrogram av tofaktorsamspillene med C til modell $Y = 2A + 2C + AB + BC + \varepsilon$, $\varepsilon \sim N(0, 0.25)$

er så rart siden A og C er aktive hovedeffekter og AC dermed ikke blir påvirket av disse. Men resten av tofaktorsamspillene samler seg i to grupper. Hvilke er da aktive? Det at bare AC har havnet for seg selv kan gi oss et hint. Hadde det vært et eller to tofaktorsamspill som påvirket AC som var aktive hadde den ikke havnet så lavt. Det kan altså gi oss et hint om at de aktive tofaktorsamspillene er involverte med A eller C . Dette er også i tråd med tankegangen om arvelighet. For å finne ut hvilke tofaktorsamspill det kan være kan vi bruke ortogonaliseringsteorien i kapittel 4. Det vi vil gjøre er å ha $X_1 = [A, C]$ fra starten av. Deretter vil vi se på tofaktorsamspill med A i X_2 og finne ut om det er et av disse som skiller seg ut. Hvis det er det vil vi flytte dette tofaktorsamspillet inn i X_1 . Deretter vil vi se på tofaktorsamspillene til C i X_2 og finne de/det som skiller seg ut der. I tabell 5.6 ser vi kontrastene som kommer ut når vi ser på tofaktorsamspillene til A . Og i tabell 5.7 ser vi kontrastene som kommer ut når vi ser på tofaktorsamspillene til C .

Ser vi på de verdiene som vi har kalt “Før Ortogonalisering”, som er de kontrastene vi får ved å regne ut $\frac{1}{6}X'AY$, ser vi at absoluttverdien av disse gir en gruppering rundt 2 og en rundt 0,6. Dette har vi allerede sett i figur 5.16 og figur 5.17. “Etter 1” er i tabellen med tofaktorsamspillene til A fått ved å først ha $X_1 = [A, C]$ og $X_2 = [AB, AC, AD, AE, AF, AG]$. Deretter ble AB flyttet inn i X_1 og X_2 satt til $[AH, AI, AJ, AK]$. Deretter ble $X_2 = [AC, BC, CD, CE, CF]$ brukt til å finne flere C -samspill. Etter dette ser vi at BC også ser aktiv ut og denne blir flyttet inn i X_1 . Til slutt blir X_2 satt til å være $X_2 = [CG, CH, CI, CJ, CK]$, og tilhørende kontraster utregnet.

Ser vi litt nøyere på tallene i “Etter 1” i tabell 5.6 og tallene i “Etter” i tabell 5.7 ser vi at det i den siste tabellen er mye “finere” tall, dvs. tall nærmere 0 for

	<i>Før Ortogonalisering</i>	<i>Etter 1</i>	<i>Etter 2</i>
<i>AB</i>	0.83	1.11	
<i>AC</i>	0.11	0.05	0.05
<i>AD</i>	-1.95	-0.28	0.06
<i>AE</i>	-0.71	0.34	-0.00
<i>AF</i>	2.00	0.31	-0.03
<i>AG</i>	-0.44	0.47	0.13
<i>AH</i>	2.10	0.41	0.06
<i>AI</i>	-2.04	0.38	-0.03
<i>AJ</i>	-0.68	0.30	-0.05
<i>AK</i>	0.73	0.27	0.08

Tabell 5.6: Kontraster fra $Y = 2A + 2C + AB + BC + \varepsilon$ før ortogonalisering og forventningsrette effekter etter ortogonalisering.

	<i>Før Ortogonalisering</i>	<i>Etter</i>
<i>AC</i>	0.11	0.05
<i>BC</i>	0.66	1.02
<i>CD</i>	-2.26	-0.08
<i>CE</i>	-0.75	-0.03
<i>CF</i>	2.12	0.01
<i>CG</i>	-0.66	0.04
<i>CH</i>	2.19	0.00
<i>CI</i>	-2.00	0.09
<i>CJ</i>	-0.58	0.08
<i>CK</i>	0.81	0.03

Tabell 5.7: Kontraster fra $Y = 2A + 2C + AB + BC + \varepsilon$ før ortogonalisering og forventningsrette effekter etter ortogonalisering.

tofaktorsamspill vi vet ikke er aktive. Grunnen til at tallene i den første tabellen ikke er så fine er at tofaktorsamspillet BC ikke var med i hverken X_1 eller X_2 da vi regnet på disse kontrastene. I “Etter 2” i tabell 5.6 er derfor $X_1 = [A, C, AB, BC]$, $X_2 = [AC, AD, AE, AF, AG]$ og $X_3 = [AH, AI, AJ, AK]$ brukt til å regne ut resten av tofaktorsamspillene med A .

Det siste vi vil se på er å forbedre hovedeffektene. Som tidligere vil vi da bruke $X_1 = [A, C, AB, BC]$ og plassere de andre hovedeffektene i X_2 . Se tabell 5.8 for kontraster utregnet ved MKM og kontraster utregnet ved ortogonalisering. Utregnede tall for A og C er fått ved å snu rundt på hvilken matrise som er X_1 og hvilken som er X_2 .

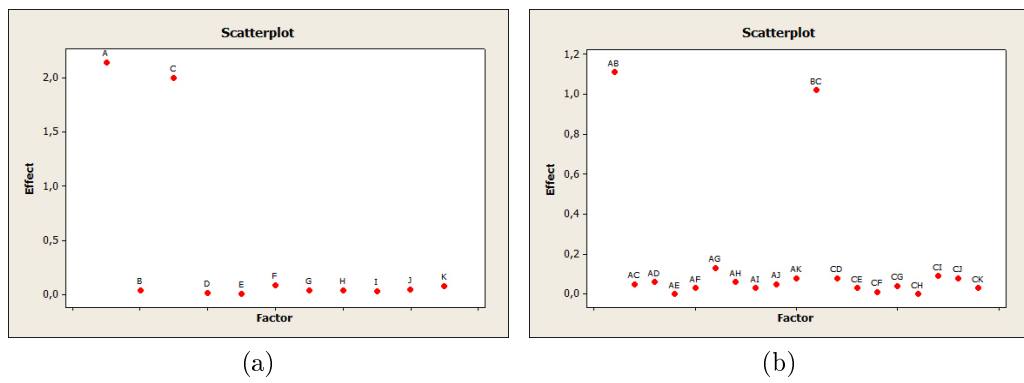
	<i>Før Ortogonalisering</i>	<i>Etter</i>
<i>A</i>	3,57	2,14
<i>B</i>	-0,08	-0,04
<i>C</i>	3,28	2,00
<i>D</i>	-0,05	-0,02
<i>E</i>	1,38	-0,01
<i>F</i>	-0,17	-0,09
<i>G</i>	-1,48	-0,04
<i>H</i>	-0,02	-0,04
<i>I</i>	-0,01	0,03
<i>J</i>	-1,54	-0,05
<i>K</i>	-1,26	0,08

Tabell 5.8: Kontraster fra $Y = 2A + 2C + AB + BC + \varepsilon$ før ortogonalisering og forventningsrette effekter etter ortogonalisering.

Ved ortogonalisering har vi her fått bestemt hovedeffektene og tofaktorsamspillene som ser ut til å være aktive. Resultatene vi får er veldig gode i forhold til vår modell, som var $Y = 2A + 2C + AB + BC + \varepsilon$.

Plott av de nye effektene i figur 5.18.

Her ser vi at verdiene er ca. halvparten av det man ville forventet av hovedeffekter og tofaktorsamspill. Dette kommer av måten de blir utregnet på etter ortogonaliseringen.



Figur 5.18: Scatterplott av hovedeffektene og tofaktorsamspillene med A og C til modell $Y = 2A + 2C + AB + BC + \varepsilon$, $\varepsilon \sim N(0, 0.25)$ etter ortogonalisering.

5.2 16 forsøks design med bruk av H_2

Konfunderingsmønsteret i 16 forsøks PB design er noe mer komplisert enn i 12 forsøks PB design. Som skrevet i kapittel 3.2 er det her konfunderinger på $\pm\frac{1}{2}$ mellom noen hovedeffekter og tofaktorsamspill. Tabell 3.4 viser hvilke tofaktorsamspill som er konfunderte med hvilke hovedeffekter. Etter at man har funnet ut hvilken gruppe tofaktorsamspill som er aktive må man altså bruke andre metoder, for eksempel oppfølgingsforsøk, til å finne ut akkurat hvilket tofaktorsamspill som er aktivt. Se tidligere i kapittel 3.2 for å finne ut mer om hvordan 16 forsøks design kan tolkes.

5.2.1 Algoritme for 16 forsøks metode

1. Finn hovedeffektskontrastene til faktorene ved å regne ut $\hat{\theta} = \frac{1}{8}X'Y$. Der X er fra $Y = X\theta$. Ta absoluttverdi av disse og lag scatterplott og dendogram.
2. Fra scatterplott og dendogram skill ut de hovedeffektene som er aktive, fra de som er inaktive. Dette gjøres ut i fra en tolkning av aliasmatrisen til PB. En gruppe hovedeffekter kan tyde på et aktivt tofaktorsamspill. F.eks. er A , B , C og E konfunderte med de samme 16 tofaktorsamspillene.
3. Regn ut effektene til de tofaktorsamspillene som antas å være aktive. Her vil vi som oftest finne at vi bare kan si at en gruppe på fire eller seks tofaktorsamspill er signifikante.
4. Lag scatterplott og dendogram til absoluttverdien av kontrastene til de tofaktorsamspillene som er interessante.
5. Lag scatterplott og dendogram til absoluttverdien av kontrastene til de tofaktorsamspillene som ikke er konfunderte med noen av hovedfaktorene for å utelukke at det er noen av disse som er aktive.
6. Hvis scatterplott og dendogram av hovedeffektene, og tofaktorsamspillsplott og dendogram samsvarer med hverandre, konkluder med hvilke effekter som er interessante. Hvis ikke gjør flere undersøkelser av andre tofaktorsamspill.
7. Bruk ortogonaliseringsteorien og beregn nye kontraster for hovedfaktorene og de aktuelle tofaktorsamspillene.

Vi vil her se på modeller både med og uten støy. På noen av modellene ser vi først på tilfellet uten støy og deretter med støy. Vi startet med å se på enkle modeller med en hovedeffekt og et tofaktorsamspill. I tillegg har vi sett på modeller med to hovedeffekter og et tofaktorsamspill. Vi har prøvd å få med de fleste forskjellige

kombinasjoner av konfunderinger mellom de ulike leddene. I et 16 forsøks design er det ikke alle tofaktorsamspill som er konfunderte med hovedeffektene. Det er derfor interessant å se på en situasjon der vi har én hovedeffekt og et tofaktorsamspill som er konfunderte, og på en situasjon der vi har én hovedeffekt og et tofaktorsamspill som ikke er konfunderte. Alle modellene finnes i tabell 5.9.

<i>Modell</i>	<i>#Hovedeffekter</i>	<i>#Tofaktorsamspill</i>
$Y = A + DJ$	1	1
$Y = A + AJ$	1	1
$Y = A + B + DJ$	2	1
$Y = A + B + AJ$	2	1
$Y = 2A + 2D - AE + AF$	2	2

Tabell 5.9: Modeller der 16 forsøks H_2 design er brukt.

I modellene der vi har tatt utgangspunkt i 16 forsøks designa er faktorene som er brukt $A-M$ unntatt I . I motsetning til 12 forsøks PB er altså ikke I med i disse modellene.

For enkelhetsskyld er tabell 3.4 gjengitt nedenfor i tabell 5.10. I tillegg til 16 tofaktorsamspill på hver av de 12 hovedfaktorene er det 18 tofaktorsamspill som ikke er konfunderte med noen hovedfaktor, bare seks og seks med hverandre.

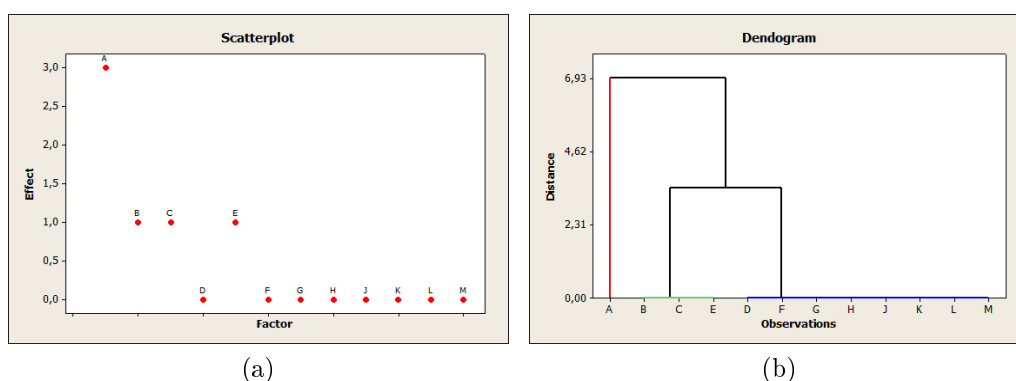
<i>Faktor</i>	<i>Tofaktorsamspill</i>
<i>A, B, C, E</i>	<i>DJ, FM, GL, HK, DK, FL, GM, HJ, DL, FK, GJ, HM, DM, FJ, GK, HL</i>
<i>D, F, G, H</i>	<i>AJ, BM, CL, EK, AK, BL, CM, EJ, AL, BK, CJ, EM, AM, BJ, CK, EL</i>
<i>J, K, L, M</i>	<i>AD, BF, CG, EH, AF, BD, CH, EG, AG, BH, CD, EF, AH, BG, CF, DE</i>
	<i>AB, CE, DF, GH, JM, KL, AC, BE, DG, FH, JL, KM, AE, BC, DH, FG, JK, LM</i>

Tabell 5.10: Konfundering mellom faktorer og tofaktorsamspill i H_2 der alle tofaktorsamspill på samme rad er like eller bare har forskjellig fortegn.

5.2.2 Modell $Y = A + DJ$

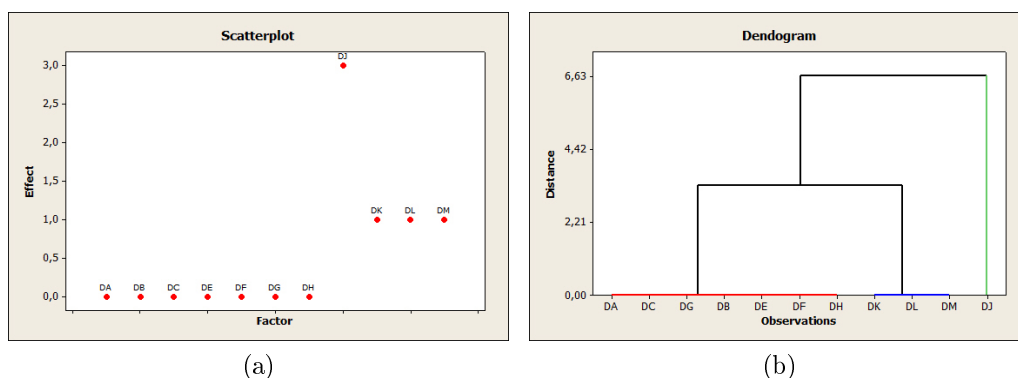
Dette er en enkel modell med en hovedeffekt og et tofaktorsamspill. Tofaktorsamspillet er her konfundert med hovedeffekten og det er ingen arvelighet i modellen. Tradisjonelt sett er dette en vanskeligere modell å jobbe med, enn en modell med arvelighet. En annen vanskelig ting med denne modellen er at vi har tre ulike faktorer og de er fra tre forskjellige grupper. Både fra S_1 , S_2 og S_3 . Denne modellen ser vi på uten støy.

Først regner vi ut kontrastene til faktorene, også kalt hovedeffektene. Disse kan du se i både scatterplott og dendrogram i figur 5.19.



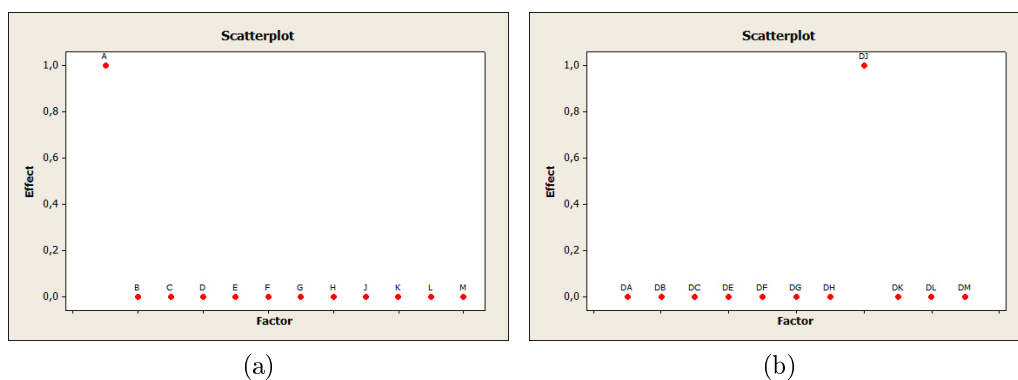
Figur 5.19: Scatterplott og dendrogram av hovedeffektene til modell $Y = A + DJ$

Her kan vi se at det er fire faktorer som skiller seg ut. A , B , C og E . Disse fire ligger i samme gruppe i tabell 5.10. Dette kan tyde på at det er et tofaktorsamspill som er konfundert med disse fire faktorene som er aktivt. I tillegg virker det som om det er A som er den aktive hovedfaktoren, siden den har en enda større kontrast enn de andre hovedeffektene. Ser man på tofaktorsamspillene som hører til A , B , C og E så er det fire forskjellige tofaktorsamspill. Disse fire har hver tre andre tofaktorsamspill som er like, eller evt. bare har motsatt fortegn. Å regne ut disse fire forskjellige kontrastene blir derfor neste skritt. Da vil vi forhåpentligvis finne et tofaktorsamspill som er aktivt, eller i realiteten fire tofaktorsamspill der vi vet at minst et er aktivt. Vi klarer altså ikke å konkludere bedre enn at minst et av fire samspill er aktivt. Men hvis vi har mer kunnskap, eller har mulighet til å gjøre flere forsøk kan vi kanskje klare å eliminere enda flere. Ser vi på tabell 5.10 ser vi at vi kan se på de ulike tofaktorsamspillene som er konfunderte med A, B, C og E ved å se på tofaktorsamspillene med D . Da får vi i tillegg med de tre siste radene i tabell 5.10. Disse inneholder tofaktorsamspill som ikke er konfunderte med noen hovedeffekter. I figur 5.20 ser vi scatterplott av disse.



Figur 5.20: Scatterplott og dendrogram av tofaktorsamspillene med D til modell $Y = A + DJ$

Her ser vi at DJ kommer sterkt fram. I tillegg har DK , DL , og DM kontraster forskjellig fra null. Vi vil derfor se på ortogonaliseringen av tofaktorsamspillene med D . Da vil vi bruk $X_1 = [A]$ og X_2 litt forskjellig, men passer på å ha med DJ , siden det virker som om denne er aktiv. For denne modellen så er $DJ - DK = A + E$. Dette skaper problemer når vi skal regne ut kontrastene. Siden DJ allerede ser viktig ut vil vi her ta med denne fra starten av. Så lenge vi ikke har med DK og E samtidig med DJ klarer vi dermed å bestemme at denne skal være med, og flytter DJ over i X_1 . For å finne ut noe om DK går det an å se på $X_1 = [A, DJ]$ med $X_2 = DK$.



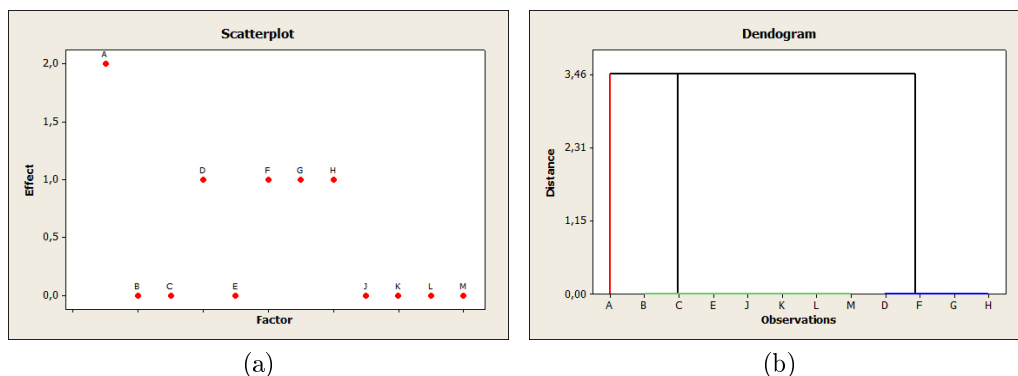
Figur 5.21: Scatterplott av hovedeffektene og tofaktorsamspillene med D til modell $Y = A + DJ$ etter ortogonalisering.

I figur 5.21 ser vi scatterplott av hovedeffektene og tofaktorsamspillene med D etter ortogonalisering. Her ser vi at vi får ut A og DJ med verdi 1. Dette er halvparten

så masse som det vi får hvis vi bruker tradisjonell måte å regne ut effekter på. Men når vi bruker ortogonaliseringsteorien som vist i kapittel 4 kommer verdiene ut som størrelsene i modellen i stedet for effekter.

5.2.3 Modell $Y = A + AJ + \varepsilon$

Denne modellen er også enkel, men med arvelighet. AJ er altså ikke konfundert med A , men AJ kan påvirke andre hovedeffekter. Vi forventer altså at A skal skille seg fra de andre hovedeffektene, og at vi får en gruppe med andre hovedeffekter som samler seg. For å finne ut av dette lager vi scatterplott og dendogram med hovedeffektene. Disse finnes i figur 5.22.



Figur 5.22: Scatterplott og dendogram av hovedeffektene til modell $Y = A + AJ$

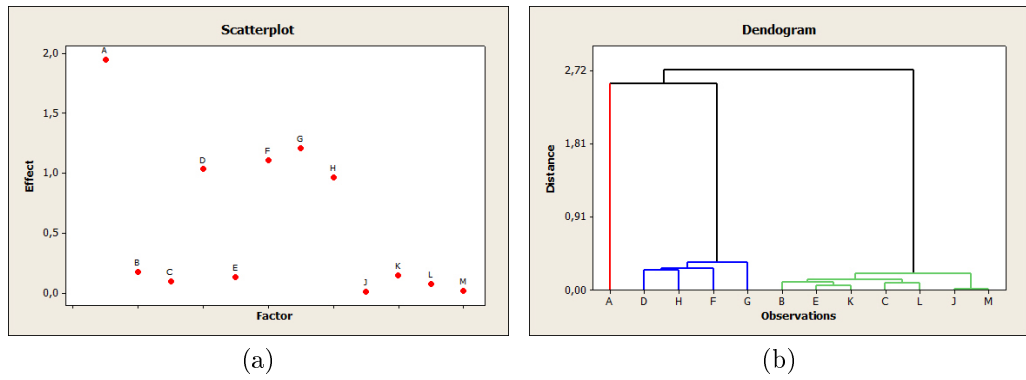
Her ser vi at D , F , G og H samler seg av hovedeffektene. Dette kan tyde på at det er et av tofaktorsamspillene som hører til den gruppa som er aktivt. Her går det heller ikke å finne nøyaktig det samspillet som er aktivt, men en gruppe på fire aktuelle kandidater.

For å se hvordan dette kan se ut med støy, følger derfor samme tankegang med $Y = A + AJ + \varepsilon$, $\varepsilon \sim N(0, 0.25)$. Scatterplott og dendogram til denne modellen kan man se i figur 5.23. Her ser vi det samme som i figur 5.22, at A får en høy kontrast og at D , F , G og H samler seg.

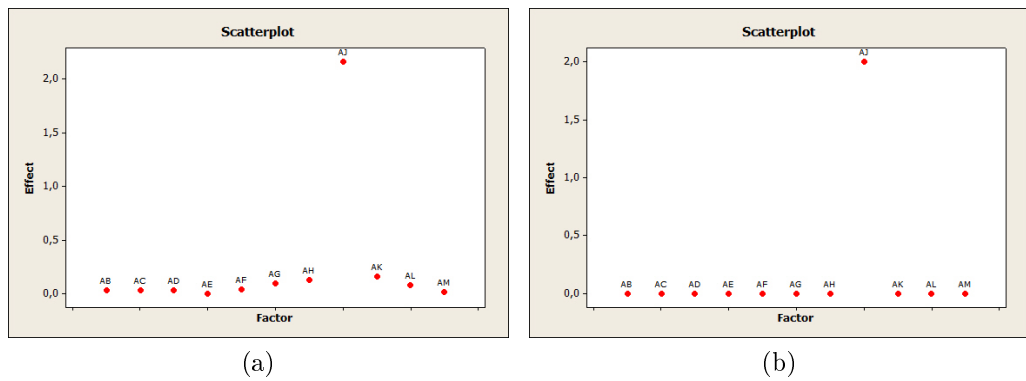
Siden D , F , G og H er de hovedeffektene som blir påvirket av tofaktorsamspill ser vi fra tabell 5.10 at ved å se på tofaktorsamspillene til f.eks. A , kan vi finne ut hvilket samspill som er aktivt. Da får vi med de tre siste radene med tofaktorsamspill i tabell 5.10 også, disse er ikke konfunderte med noen hovedeffekter.

I figur 5.24 ser vi scatterplott av absoluttverdien til kontrasten til tofaktorsamspillene med A med og uten støy. A sitt scatterplott er veldig oversiktlig og greit. AJ er det eneste tofaktorsamspillet som får noen kontrast. Dette ser man tydelig i plottet med støy også.

I denne modellen er det ikke noe vits å regne på ortogonalisering når det gjelder tofaktorsamspill. Siden A og AJ er de effektene som er aktive, og de er ortogonale.



Figur 5.23: Scatterplott og dendrogram av hovedeffektene til modell $Y = A + AJ + \varepsilon$, $\varepsilon \sim N(0, 0.25)$



Figur 5.24: Scatterplott av tofaktorsamspillene med A til modell $Y = A + AJ + \varepsilon$, med støy (a), $\varepsilon \sim N(0, 0.25)$, og uten støy (b).

Men vi kan se på ortogonalisering av hovedeffektene. For å regne ut disse settes $X_1 = [A, AJ]$ og X_2 settes til å være de andre hovedeffektene i to omganger. Se tabell 5.11 for kontrastene i figur 5.22 og utregnede kontraster med ortogonalisering for tilfellet uten støy.

Utrekning gir:

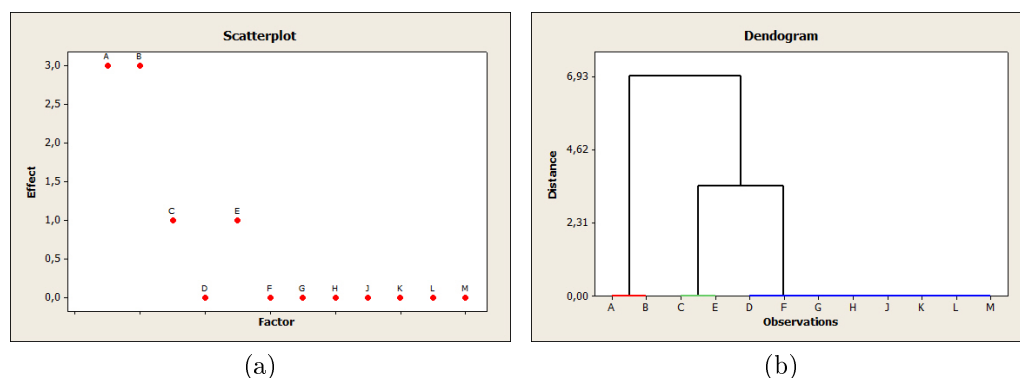
<i>Faktor</i>	<i>Før ortogonalisering</i>	<i>Etter</i>
<i>A</i>	2	1
<i>B</i>	0	0
<i>C</i>	0	0
<i>D</i>	1	0
<i>E</i>	0	0
<i>F</i>	1	0
<i>G</i>	1	0
<i>H</i>	1	0
<i>J</i>	0	0
<i>K</i>	0	0
<i>L</i>	0	0
<i>M</i>	0	0

Tabell 5.11: Kontraster fra $Y = A + AJ$ før ortogonalisering og forventningsrette effekter etter ortogonalisering.

Ved utregning av θ_1 og θ_2 etter ortogonalisering blir verdiene halvparten så store som hovedeffektene blir ved tradisjonell utregning. Her blir verdien til A lik 1, mens den ble to når vi regnet den ut på vanlig måte. Det som vi ser fra tabell 5.11 er at verdiene til D , F , G og H blir 0 etter ortogonalisering, mens de var 1 før. Vi har altså vist at verdiene disse fire hadde kom fra konfunderingen med AJ . I ortogonaliseringen har vi brukt modellen uten støy, men den samme metoden kan brukes på modellen med støy også.

5.2.4 Modell $Y = A + B + DJ + \varepsilon$

Dette er en modell med fire aktive faktorer, uten arvelighet. A og B ligger i samme gruppe med hovedfaktorer, dette kan vi se i tabell 5.10 og DJ er konfundert med både A og B . I tillegg er dette, som modell $Y = A + DJ$, en modell med faktorer fra alle de tre gruppene S_1 , S_2 og S_3 . Dette gjør det mer komplisert. I figur 5.25 ser vi scatterplott og dendrogram av hovedeffektene til modellen. Der ser vi at A og B samler seg, at C og E samler seg, og at D, F, \dots, K ligger på 0. Dette kan tyde på at et tofaktorsamspill som er konfundert med A, B, C og E er aktivt.



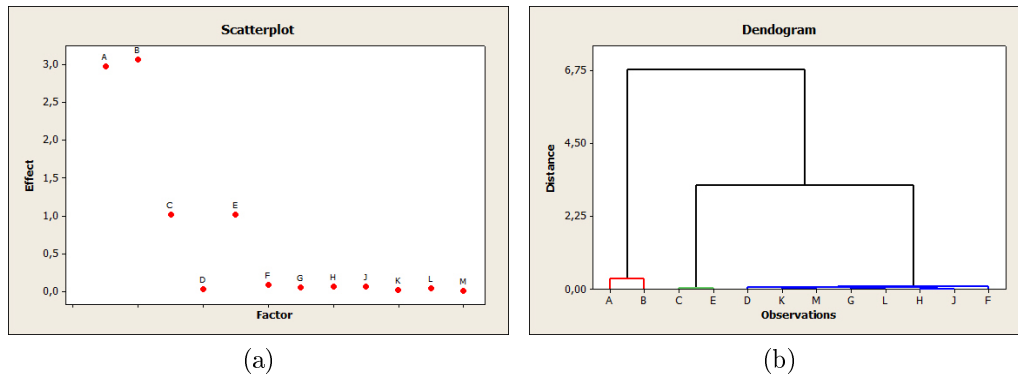
Figur 5.25: Scatterplott og dendrogram av hovedeffektene til modell $Y = A+B+DJ$

I figur 5.26, figur 5.27 og figur 5.28 ser vi scatterplott og dendrogram av hovedeffektene til den samme modellen med støy på tre forskjellige nivå. $N(0, 0.1)$, $N(0, 0.25)$ og $N(0, 0.5)$. Til og med i modellen med støy på nivå $\varepsilon \sim N(0, 0.5)$ klarer vi å skille hvilke hovedeffekter som er aktive. På de lavere nivåene er det enda enklere.

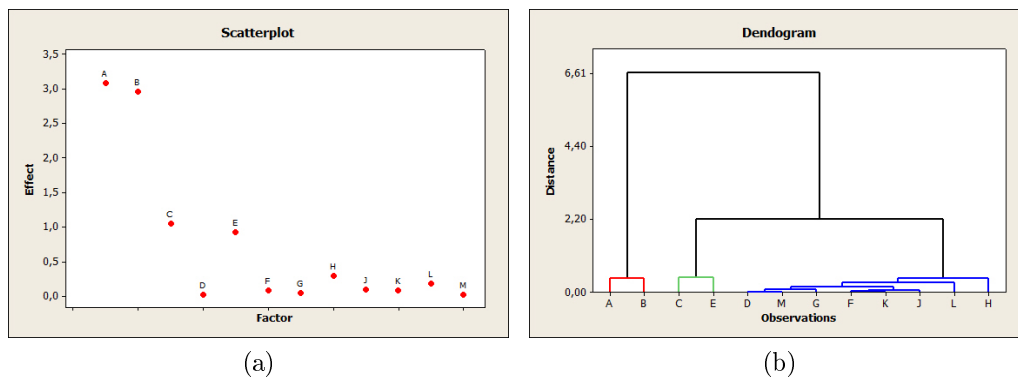
I tabell 5.10 ser vi at vi kan se på tofaktorsamspillene med D for å finne den aktuelle gruppa med tofaktorsamspill som er konfunderte med A, B, C og E . I figur 5.29 ser vi scatterplott og dendrogram til tofaktorsamspillene med D .

Her ser vi tydelig at DJ skiller seg ut fra de andre tofaktorsamspillene. I tillegg får DM også en kontrast. La oss prøve med litt støy, for å se om vi konkluderer med det samme da. Figur 5.30 og figur 5.31 viser scatterplott og dendrogram for $\varepsilon \sim N(0, 0.25)$ og $\varepsilon \sim N(0, 0.5)$.

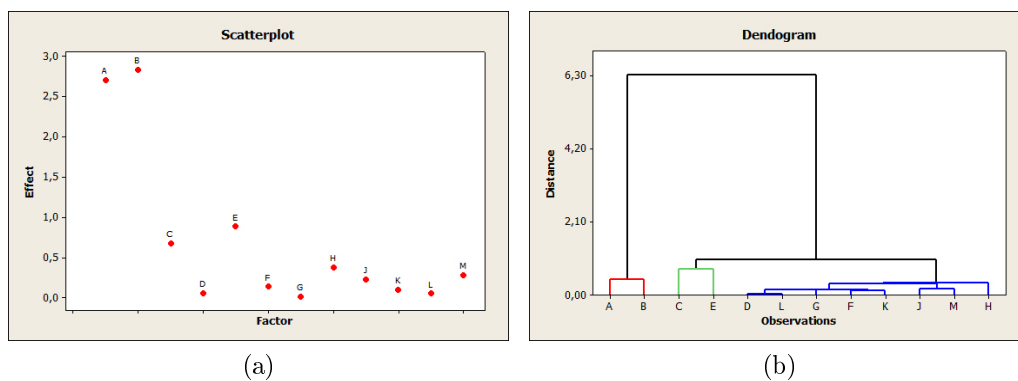
Her ser vi at DJ og DM kommer ut med kontraster som ligger langt fra de andre tofaktorsamspillene med D . Vi vet jo at det er DJ som er aktivt, men for å vise dette kan vi bruke ortogonaliseringsteorien fra kapittel 4. Da kan vi få vekk skjevheten i θ_2 og finne forventningsrett θ_2 . Ved å ha både A og B i X_1 og resten av D -ene i X_2 i et par omganger kan vi finne ut forventningsrett $\hat{\theta}$ for hovedfaktorene



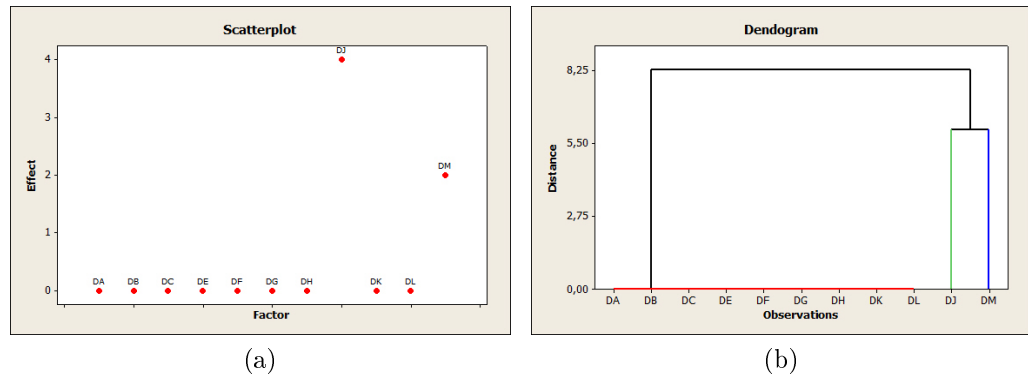
Figur 5.26: Scatterplott og dendrogram av hovedeffektene til modell $Y = A + B + DJ + \varepsilon$, $\varepsilon \sim N(0, 0.1)$



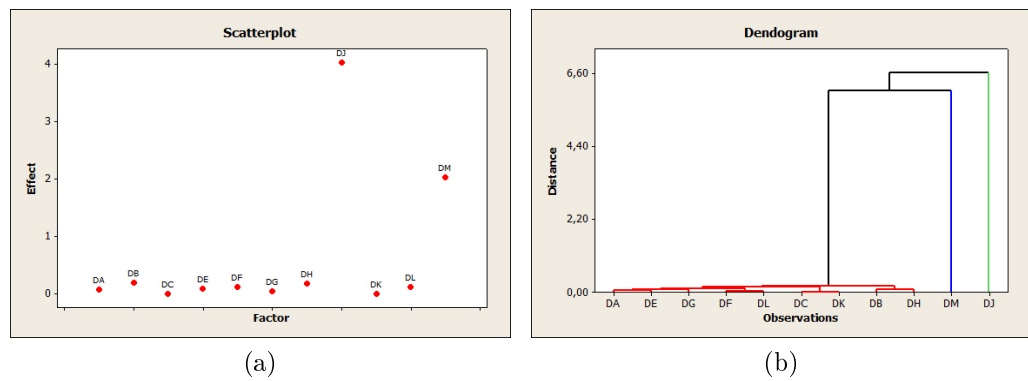
Figur 5.27: Scatterplott og dendrogram av hovedeffektene til modell $Y = A + B + DJ + \varepsilon$, $\varepsilon \sim N(0, 0.25)$



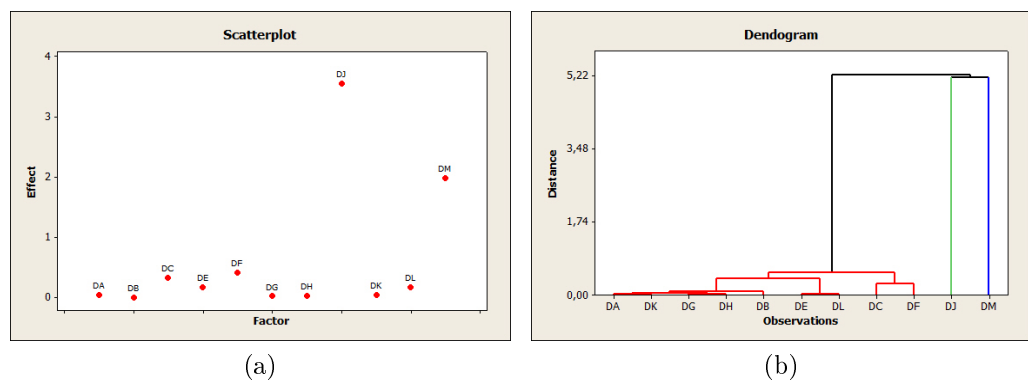
Figur 5.28: Scatterplott og dendrogram av hovedeffektene til modell $Y = A + B + DJ + \varepsilon$, $\varepsilon \sim N(0, 0.5)$



Figur 5.29: Scatterplott og dendrogram av tofaktorsamspillene med D til modell $Y = A + B + DJ$



Figur 5.30: Scatterplott og dendrogram av tofaktorsamspillene med D til modell $Y = A + B + DJ + \varepsilon$, $\varepsilon \sim N(0, 0.25)$



Figur 5.31: Scatterplott og dendrogram av tofaktorsamspillene med D til modell $Y = A + B + DJ + \varepsilon$, $\varepsilon \sim N(0, 0.5)$

og tofaktorsamspillene til D . Siden kontrasten til DJ er større enn DM sin kontrast prøver vi først å se på $X_1 = [A, B]$ og $X_2 = [AD, BD, CD, DE, DF, DG, DH, DJ]$. Da ser vi at DJ sin kontrast er mye større enn de andre. Å finne kontrastene til DK, DL og DM viser seg å være litt vanskeligere. Grunnen til at det oppstår problemer er at $DJ - DM = A + B$ og $DK - DL = -A + B$ for H_2 . En måte å løse det på er å ta en og en av de siste tofaktorsamspillene i X_2 og beholde $X_1 = [A, B]$. Da viser det seg at kontrastene for DJ og DM blir like store. Selv om vi i figur 5.31 har sett at DJ er større enn DM og det dermed er mer sannsynlig at den har en verdi. Verdiene vi finner til A og B blir tilnærmet lik 1 og verdiene til DJ og DM blir også tilnærmet lik 1. Siden vi vet at verdiene vi har regnet ut er halvparten så store som når vi regner ut hovedeffekter er det god grunn til å tro at modellen vår er lik $Y = k_1A + k_2B + k_3DJ/DM$ med $k_1 = 1, k_2 = 1$ og $k_3 = 1$.

Vi har altså to tofaktorsamspill som ser ut som om de er aktive og like stor, enten DJ eller DM . Men fra scatterplott av hovedeffektene har vi sett at det mest sannsynlig bare er et tofaktorsamspill som er aktivt, så begge kan ikke være det. En mulighet vi har er å se på forskjellen mellom de verdiene på Y vi har, og de verdiene som ville komme fra den ideelle modellen, $\hat{Y}_1 = A + B + DJ$ eller $\hat{Y}_2 = A + B + DM$. Ser vi på $\hat{\epsilon} = Y - \hat{Y}$ bør altså den rette \hat{Y} -en gi en kolonne med ϵ som ser ut til å være $\epsilon \sim N(0, \sigma)$. For da er $\epsilon = A + B + DJ + \epsilon - (A + B + DJ) = \epsilon$, hvis DJ er det riktige tofaktorsamspillet. Mens modellen som er feil bør gi en ϵ som er lik $\epsilon = A + B + DJ + \epsilon - (A + B + DM) = DJ - DM + \epsilon$, evt. med motsatt fortegn. Dette forutsetter at vi har valgt riktige k -er foran hovedeffektene og tofaktorsamspillene.

I tabell 5.12 ser vi Y fra de simulerte dataene våre med $\epsilon \sim N(0, 0.25)$, de to kolonnene med $\hat{\epsilon}$ og $DJ - DM$. Her ser vi at kolonne nr. 2 er den kolonnen som inneholder ϵ og ikke $DJ - DM + \epsilon$. Sammenligner vi kolonne nr. 3 og 4 ser vi at disse er nokså like, bortsett fra litt støy. Den rette modellen ser derfor ut til å være $Y = A + B + DJ + \epsilon$.

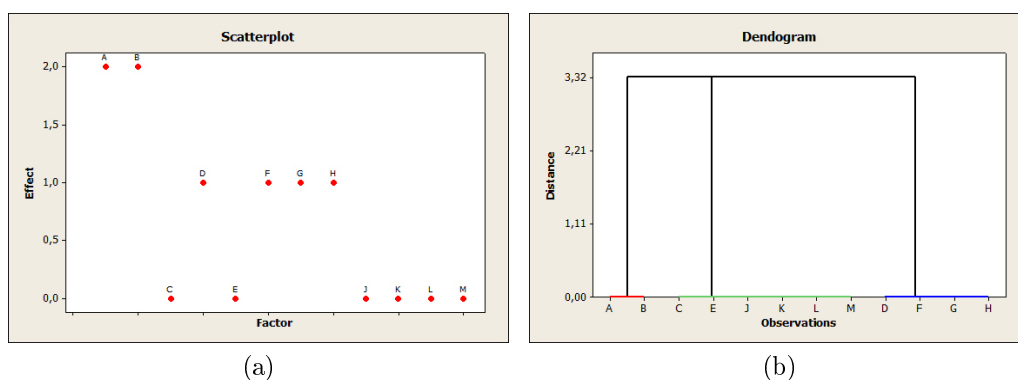
Som avsluttende kommentar må vi si at dette er en veldig idealisert situasjon. At vi finner koeffisientene som skal være foran effektene så nøyaktig som vi gjør er veldig heldig. Og det er selvfølgelig langt lettere å finne riktig modell når man faktisk vet hva slags modell dataene kom fra.

Y	$\hat{\epsilon}_1$	$\hat{\epsilon}_2$	$DJ - DM$
-2.60	0.40	- 1.60	-2
1.10	0.10	0.10	0
0.68	-0.32	- 0.32	0
3.10	0.10	2.10	2
-3.04	- 0.04	- 2.04	-2
-0.83	0.17	0.17	0
-1.20	- 0.20	- 0.20	0
3.10	0.10	2.10	2
-3.23	- 0.23	- 2.23	-2
0.76	- 0.24	- 0.24	0
1.40	0.40	0.40	0
3.42	0.42	2.42	2
-2.80	0.20	- 1.80	-2
-0.82	0.18	0.18	0
-1.15	- 0.15	- 0.15	0
2.91	- 0.09	1.91	2

Tabell 5.12: Tabell med Y og residualene $\hat{\epsilon}_1$ og $\hat{\epsilon}_2$.

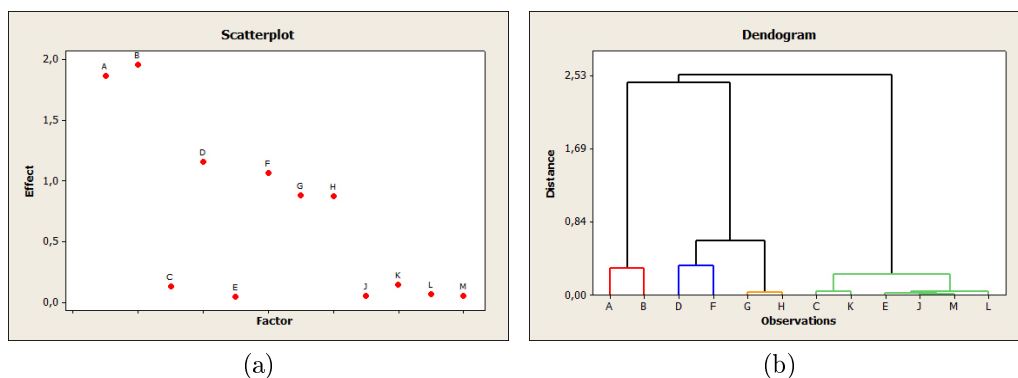
5.2.5 Modell $Y = A + B + AJ + \varepsilon$

Dette er en modell med to hovedeffekter, et tofaktorsamspill og tre aktive faktorer. A og B ligger i samme gruppe med hovedeffekter, dette kan vi se i tabell 5.10 og AJ er ikke konfundert med A eller B . I figur 5.32 ser vi at A og B samler seg, og at D , F , G og H samler seg. Resten ligger på 0. Dette kan tyde på at det er et tofaktorsamspill som er konfundert med D , F , G og H som er aktivt. Vi vet at AJ er det aktive tofaktorsamspillet, og utifra tabell 5.10 er AJ konfundert med D , F , G og H . Derfor virker dette lovende.



Figur 5.32: Scatterplott og dendrogram av hovedeffektene til modell $Y = A + B + AJ$

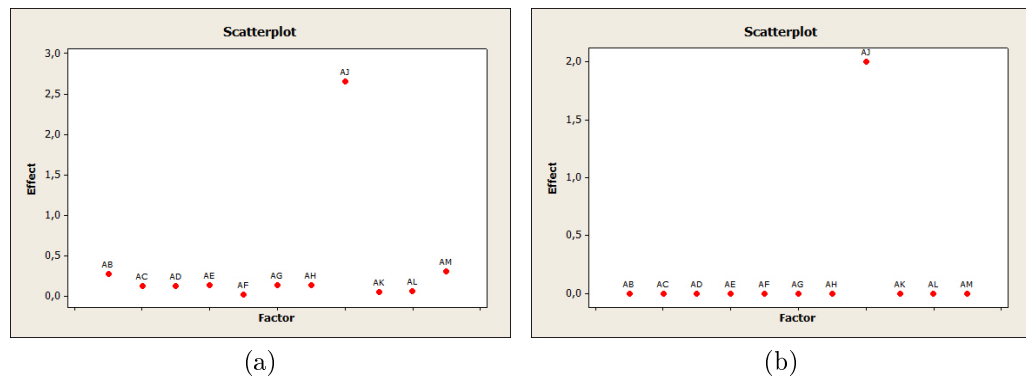
Legger vi på litt støy blir den tilsvarende figuren som figur 5.32 lik figur 5.33. Her ser vi også at D , F , G og H skiller seg ut fra de andre hovedeffektene.



Figur 5.33: Scatterplott og dendrogram av hovedeffektene til modell $Y = A + B + AJ + \varepsilon$, $\varepsilon \sim N(0, 0.25)$.

For å finne ut hvilket tofaktorsamspill som er aktivt kan vi se på tofaktorsamspillene med A . Da vil vi kunne finne ut om AJ , AK , AL eller AM er aktive og si

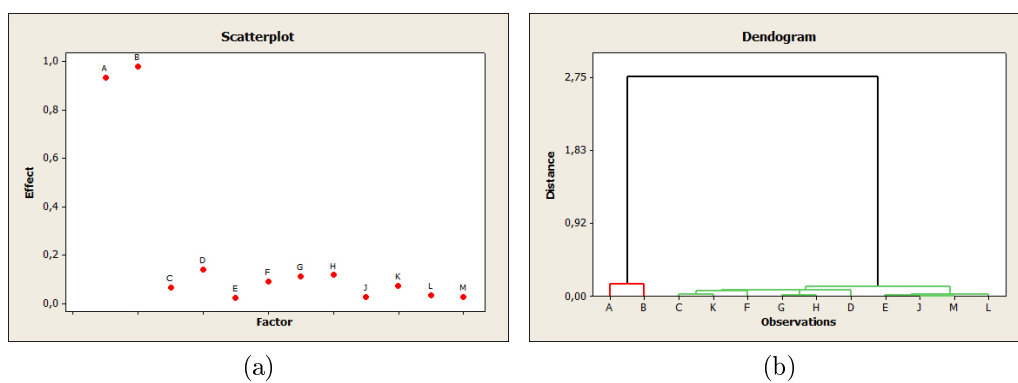
hvilken av de fire gruppene med like tofaktorsamspill som er aktive. Da finner vi også ut om AB , AC eller AE , som ikke er konfunderte med noen hovedfaktor, er aktive. I figur 5.34 ser vi scatterplott av tofaktorsamspillene med A med og uten støy. Her ser vi tydelig at det bare er AJ av tofaktorsamspillene med A som er aktivt.



Figur 5.34: Scatterplott av tofaktorsamspillene med A for modell $Y = A + B + AJ + \varepsilon$, med støy (a), $\varepsilon \sim N(0, 0.25)$ og uten støy (b).

Ved å bruke ortogonaliseringsteorien i kapittel 4 kan vi bestemme hvilke tofaktorsamspill som virkelig er aktive. Fra figur 5.33 ser vi at det er A og B som er aktive hovedeffekter. Disse kan vi derfor plassere i X_1 . Til X_2 kan vi prøve alle tofaktorsamspillene til A først, og se om det fungerer. Dette fører til at $X_1^T X_2 = 0$ så vi ikke trenger å ta hensyn til X_2 når vi skal bestemme θ_1 . Tofaktorsamspillskontrastene i figur 5.34 er derfor riktige. Men i forhold til hovedeffektene kan det være aktuelt å bruke ortogonaliseringsteorien.

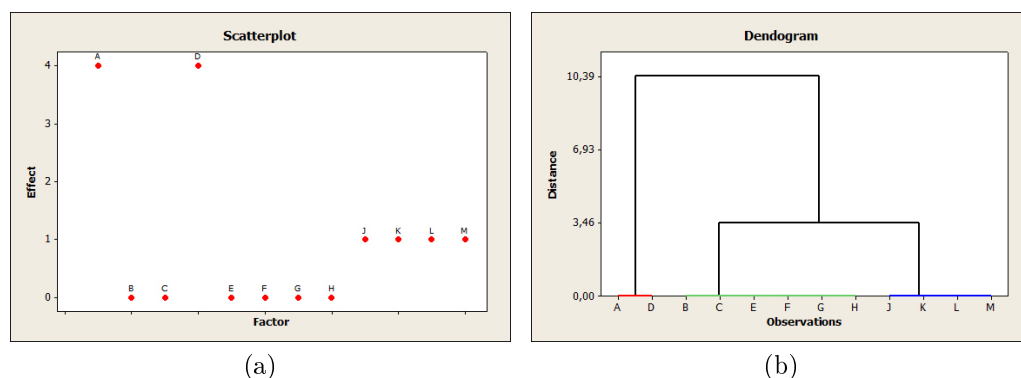
Vi vet altså at A , B og AJ er de aktuelle effektene, men vi vil finne forventningsrette effekter til disse og lage et forventningsrett plott av hovedeffektene. Vi lar derfor $X_1 = [A, B, AJ]$ og $X_2 = [C, D, E, F, G, H, J, K, L, M]$. Da kan vi lage scatterplott og dendrogram av de forbedrede hovedeffektene. I figur 5.35 ser vi at A og B legger seg rundt 1, som forventet, siden dette er halvparten av hva effektene burde vært. Resten av hovedeffektene legger seg rundt 0, i motsetning til at D , F , G og H ligger i en klynge for seg selv. Dermed ser vi at effekten disse hadde kom fra AJ akkurat som i modellen $Y = A + AJ$.



Figur 5.35: Scatterplott og dendrogram av hovedeffektene til modell $Y = A + B + AJ + \varepsilon$, $\varepsilon \sim N(0, 0.25)$ etter ortogonalisering.

5.2.6 Modell: $Y = 2A + 2D - AE + AF + \varepsilon$

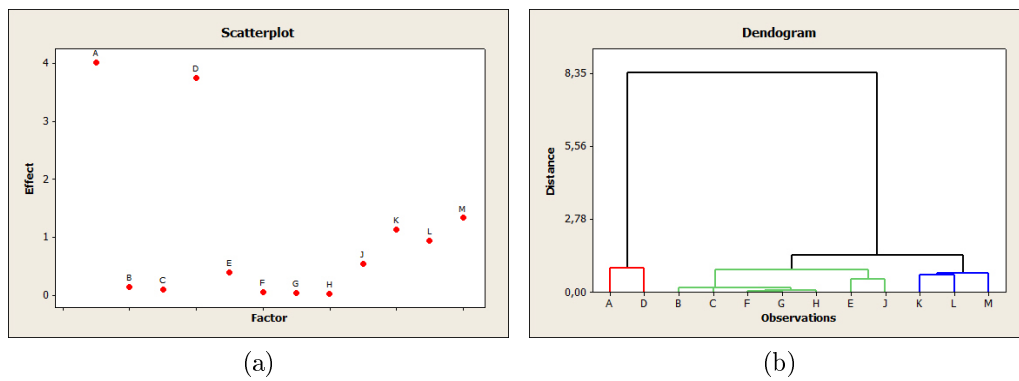
Her ser vi på en modell med to hovedeffekter og to tofaktorsamspill med arvelighet og fire aktive faktorer. I figur 5.36 ser vi scatterplott og dendrogram av hovedeffektene uten støy. Her kan vi tydelig se at A og D er aktive. I tillegg ser vi at J , K , L og M får en kontrast på 1. Dette tyder på at det er et tofaktorsamspill som er konfundert med disse hovedeffektene som er aktivt. I tabell 5.10 ser vi hvilke tofaktorsamspill dette er.



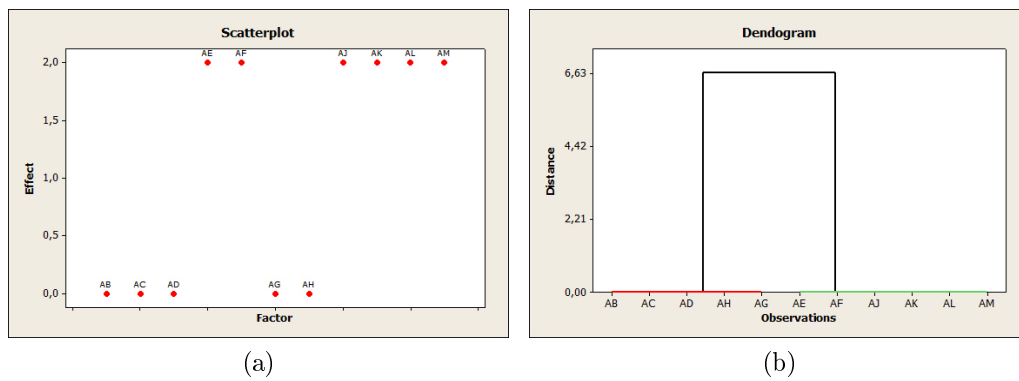
Figur 5.36: Scatterplott og dendrogram av hovedeffektene til modell $Y = 2A + 2D - AE + AF$

I figur 5.37 ser vi samme type scatterplott og dendrogram til modellen med støy på $\varepsilon \sim N(0, 0.5)$. Her ser vi at A og D kommer tydelig fram som i tilfellet uten støy. Når det gjelder de andre hovedeffektene er det litt vanskeligere å se hvilke som er aktive. Men i dendrogrammet ser det ut som om K , L og M samler seg i en klynge. Sammenligner vi med tabellen i tabell 5.10 ser vi da at J , K , L og M er en gruppe med de samme tofaktorsamspillene konfunderte med seg.

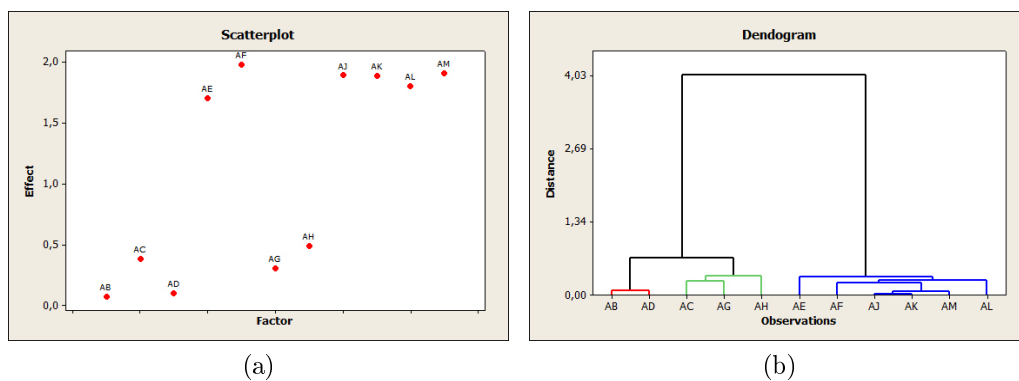
For å finne ut hvilket tofaktorsamspill det kan være som er konfundert med J , K , L og M kan vi se på tofaktorsamspillene til A . Da får vi med oss alle tofaktorsamspill som kan være aktuelle. Både de som er konfunderte med J , K , L og M og de som ikke er konfunderte med noen hovedeffekter. I figur 5.38 ser vi scatterplott og dendrogram til tofaktorsamspillene til A uten støy. Og i figur 5.39 ser vi scatterplott og dendrogram til tofaktorsamspillene til A med støy. Her ser vi at det er mange av tofaktorsamspillene som skiller seg ut. AE , AF , AJ , AK , AL og AM samler seg på et nivå lik, eller rundt 2. Siden vi allerede vet at A og D er aktive hovedeffekter kan vi i utgangspunktet, før vi undersøker mer, si at AJ , AK , AL og AM antageligvis får sin kontrast fra hovedeffekt D . Siden D er konfundert med disse tofaktorsamspillene, er det AE og AF som må undersøkes nærmere først.



Figur 5.37: Scatterplott og dendrogram av hovedeffektene til modell $Y = 2A + 2D - AE + AF + \varepsilon$, $\varepsilon \sim N(0, 0.5)$



Figur 5.38: Scatterplott og dendrogram av tofaktorsamspillene med A for modell $Y = 2A + 2D - AE + AF$



Figur 5.39: Scatterplott og dendrogram av tofaktorsamspillene med A for modell $Y = 2A + 2D - AE + AF + \varepsilon$, $\varepsilon \sim N(0, 0.5)$

For å finne ut av hva effektene til AE og AF er uten påvirkning fra andre effekter bruker vi ortogonalisering. Her ser vi på tilfellet uten støy, men det blir på samme måte med tilfellet med støy. Lar vi $X_1 = [A, D]$ og $X_2 = [AB, AC, AD, AE, AF, AG, AH, AJ, AK, AL, AM]$ får vi ikke ut noe bedre tall, fordi da er $X_{2.1} = 0$. Ser vi derimot på $X_1 = [A, D]$ og $X_2 = [AB, AC, AD, AE, AF, AG, AH, AJ]$ og deretter en og en av AK, AL og AM sammen med AE og AF finner vi at θ_1 og θ_2 er som i ligning (5.1).

$$\theta_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \text{ og } \theta_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (5.1)$$

Som nevnt tidligere blir verdiene θ_1 og θ_2 halvparten så store som utregningen av hovedeffekter. I stedet for at effekten til A blir lik fire, dobbelt så stor som to, blir den her lik to. AJ, AK, AL og AM har alle kommet ned på 0. Dette stemmer helt med modellen vi tok utgangspunkt i og viser at den kontrasten som disse tofaktorsamspillene hadde egentlig kom fra hovedeffektene som de er konfunderte med.

5.3 16 forsøks design med bruk av H_6

I kapittelet om 16 forsøks design skriver vi litt om H_6 designet. Et $(16,8,3)$ design der ingen tofaktorsamspill er konfunderte med andre tofaktorsamspill. Dette er alltid positivt, men spesielt når du har delvis konfundering mellom hovedeffekter og tofaktorsamspill. Siden vi vet at det ikke er noen konfundering mellom tofaktorsamspillene vet vi med en gang at det tofaktorsamspillet vi tror er aktivt er aktivt. Og ikke et av fire som tilfellet er for H_2 .

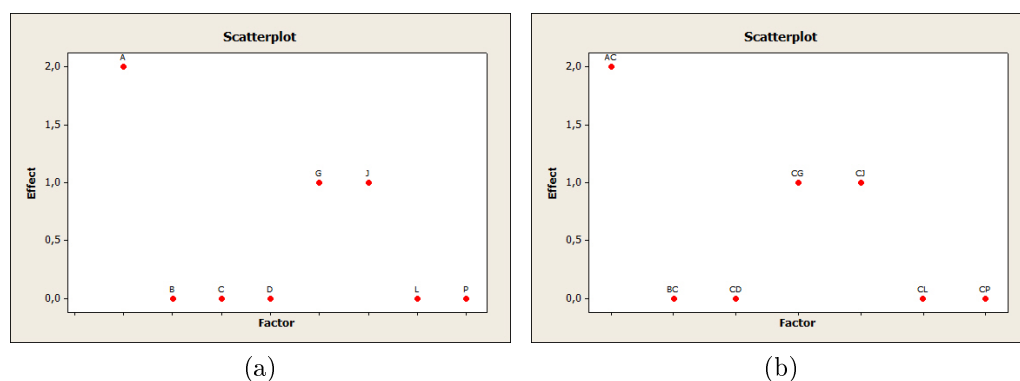
Når vi skal tolke hovedeffektsploTT for denne type design, må vi ta med i betraktning konfunderingene mellom hovedeffektene og tofaktorsamspill. I tabell 3.6 er det en oversikt over hvilke tofaktorsamspill som hører til hvilke hovedeffekter. Det kan være greit å ha en oversikt over motsatt sammenheng, altså hvilke hovedeffekter som hører til hvilke tofaktorsamspill. Dette finnes i tabell 5.13.

<i>Faktor</i>	<i>Tofaktorsamspill</i>
	<i>AB</i>
<i>A, B, C</i>	<i>LP</i>
<i>A, B, D</i>	<i>CJ</i>
<i>A, B, G</i>	<i>JL</i>
<i>A, B, J</i>	<i>DP</i>
<i>A, B, L</i>	<i>DG</i>
<i>A, B, P</i>	<i>CG</i>
<i>C, D</i>	<i>AG, BG</i>
<i>C, G</i>	<i>DL</i>
<i>C, J</i>	<i>GP</i>
<i>C, L</i>	<i>AJ, BJ</i>
<i>C, P</i>	<i>DJ</i>
<i>D, G</i>	<i>JP</i>
<i>D, J</i>	<i>GL</i>
<i>D, L</i>	<i>AP, BP</i>
<i>D, P</i>	<i>CL</i>
<i>G, J</i>	<i>AC, BC</i>
<i>G, L</i>	<i>CP</i>
<i>G, P</i>	<i>AD, BD</i>
<i>J, L</i>	<i>CD</i>
<i>J, P</i>	<i>AL, BL</i>
<i>L, P</i>	<i>GJ</i>

Tabell 5.13: Konfundering mellom hovedeffekter og tofaktorsamspill ved bruk av H_6 .

5.3.1 Modell: $Y = A + AC$

For å prøve en enkel modell ser vi på den enkle modellen $Y = A + AC$. Dette er en modell med arvelighet og bare to aktive faktorer. Først lager vi et scatterplott av hovedeffektene. Uten støy ser det ut som i figur 5.40a.

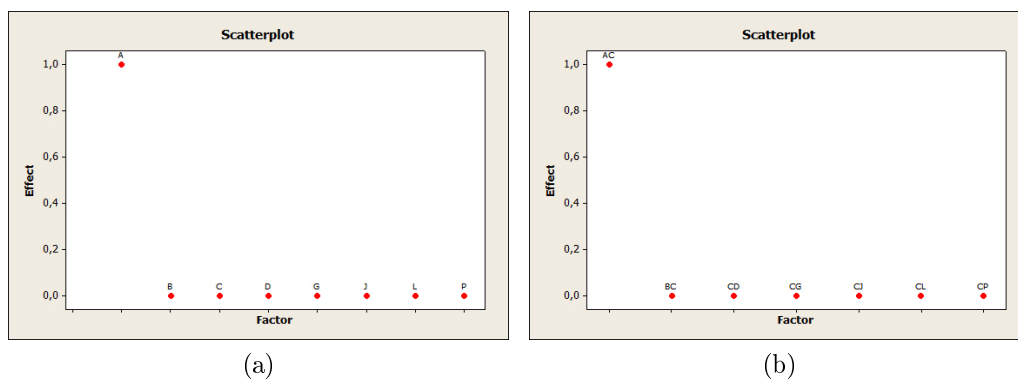


Figur 5.40: Scatterplott av hovedeffektene og tofaktorsamspillene med C for modell $Y = A + AC$

Her ser vi at hovedeffekten A skiller seg mest ut, G og J samler seg og resten av hovedeffektene havner på 0. Ser vi i tabell 5.13 hvilke tofaktorsamspill som er konfunderte med akkurat disse to hovedeffektene ser vi at det er AC og BC . For å undersøke disse nærmere kan vi se på scatterplott av tofaktorsamspill med C . Dette scatterplottet finnes i figur 5.40b. Her ser vi at av AC og BC er det AC som får en kontrast. CG og CJ får en kontrast på 1.0 og resten av tofaktorsamspillskontrastene med C blir null. Vi vil altså konkludere med at modellen vi søker er $Y = A + AC$.

I dette tilfellet går det også an å prøve ortogonalisering for å finne effekter som ikke er påvirket av konfunderinger. Først kan vi se på $X_1 = [A, AC]$ og $X_2 = [B, C, D, G, J, L, P]$ for å finne hovedeffektene. Etterpå kan vi se på tofaktorsamspillene med C . Da beholder vi samme X_1 , men setter $X_2 = [BC, CD, CG, CJ, CL, CP]$. Plott med de nye effektene finner vi i figur 5.41.

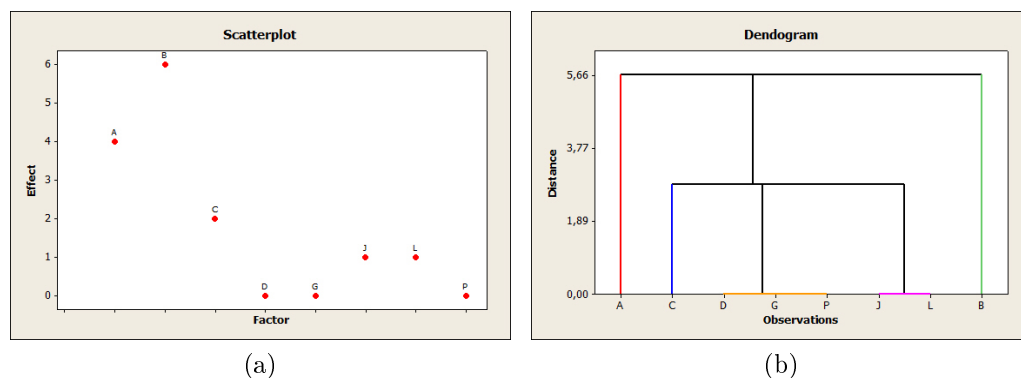
Her ser vi tydelig at det er A og AC som er de aktive effektene. Verdene disse legger seg på er halvparten så store som effektene vanligvis blir, i tråd med teorien i kapittel 4.



Figur 5.41: Scatterplott av hovedeffektene og tofaktorsamspillene med C for modell $Y = A + AC$ etter ortogonalisering.

5.3.2 Modell: $Y = 2A + 3B + C + CD$

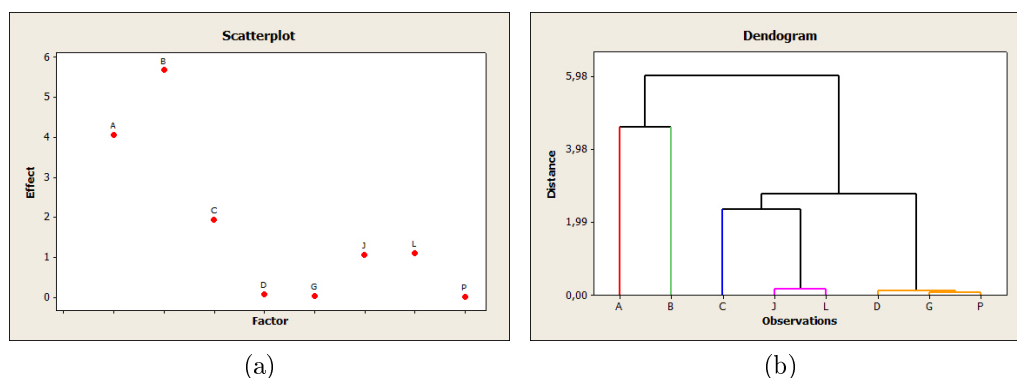
For å prøve ut designet enda mer ser vi her på en modell med fire faktorer med delvis arvelighet. Dette er en modell som det vanlige 2_{IV}^{8-4} designet vil ha problemer med. Dette på grunn av konfunderinger mellom tofaktorsamspill. I figur 5.42 ser vi scatterplott og dendogram for hovedeffektene.



Figur 5.42: Scatterplott og dendogram av hovedeffektene til modell $Y = 2A + 3B + C + CD$

Det vi ser her er at A og B havner på høye nivå. C havner litt lenger ned, J og L samler seg i midten og de tre siste hovedfaktorene havner på 0. Hvis vi vet at det ikke er noe støy i modellen kan vi anta at det er et tofaktorsamspill som er konfundert med J og L som er aktivt. I tabell 5.13 ser vi at det må være CD . Da kan vi konkludere med at $Y = k_1A + k_2B + k_3C + k_4CD$ er et godt forslag til modell. Hvis vi derimot ser eller vet at det er støy på modellen er det ikke så lett å konkludere med en så enkel modell med en gang. I figur 5.43 ser vi scatterplott og dendogram til hovedeffektene fra samme modell med støy ε , $\varepsilon \sim N(0, 0.25)$.

Det vi ser her er at det ikke er så stor forskjell på figur 5.43 og figur 5.42. I begge dendrogrammene ser vi tydelig at D , G og P samler seg, og i scatterplottene kan vi se at dette er på et lavt nivå. Disse hovedeffektene er altså ikke påvirket av noe, eller evt. påvirket positivt av et bidrag og negativt av et annet. La oss anta at de ikke er aktive. J og L ligger også på samme nivå. Men kontrastene til disse to hovedeffektene er ikke så store. Enten betyr det at disse er små effekter, eller så er det et tofaktorsamspill som påvirker disse to hovedeffektene. De siste hovedeffektene som vi ikke har kommentert er de som er størst, A , B og C . De er så store at det ser ut som om de er hovedeffekter. I hvertfall A og B , litt mer usikkert med C . Hvis vi argumenterer med at C sikkert er en hovedeffekt finner vi det samme som vi fant i tilfellet uten støy. Modellen vi vil velge er: $Y = k_1A + k_2B +$

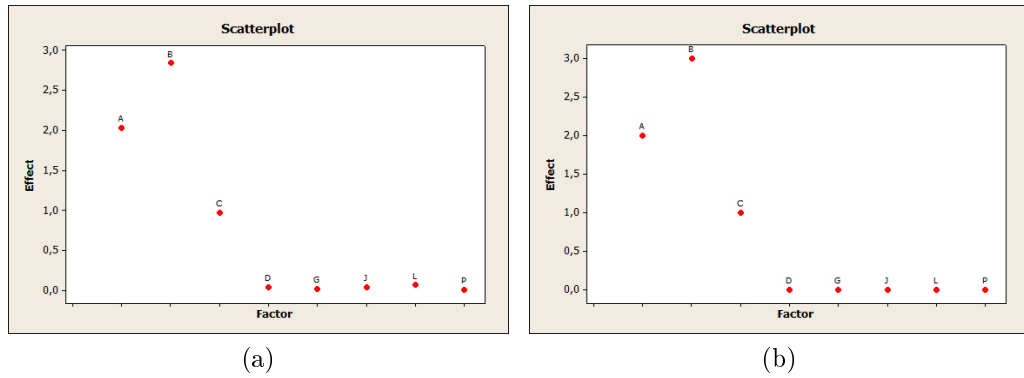


Figur 5.43: Scatterplott og dendrogram av hovedeffektene til modell $Y = 2A + 3B + C + CD + \varepsilon$, $\varepsilon \sim N(0, 0.25)$

$k_3C + k_4CD$. Er vi derimot litt forsiktige kan vi tolke det som om C er påvirket av et tofaktorsamspill og ikke har en effekt selv. Det betyr at C , J og L kan være påvirket av et tofaktorsamspill. Fra tabell 5.13 ser vi at det ikke er noe tofaktorsamspill som er konfundert med alle disse tre hovedeffektene. Dermed må det være to av disse hovedeffektene som er påvirket av et tofaktorsamspill. Den tredje har enten en effekt selv, eller er sterkt påvirket av støy. De mulige kombinasjonene er: C og J , som er konfunderte med GP , C og L , som er konfunderte med AJ og BJ og J og L , som er konfunderte med CD .

Siden vi har tofaktorsamspill som er konfundert med både C , G og J her kan det være lettere å gå rett på ortogonaliseringen i stedet for å regnet ut alle de tre tofaktorsamspillene. Vi ser tydelig at A og B er aktive, så disse kan få plass i X_1 . I X_2 vil vi gjerne ha de tofaktorsamspillene som er aktuelle, dvs. GP , AJ , BJ og CD . Setter vi disse i X_2 ser vi at $X_1'X_2 = 0$ men tar vi med litt flere tofaktorsamspill i X_2 unngår vi dette. For eksempel kan man sette $X_2 = [GP, AJ, BJ, CD, DL, DP, DJ]$. Da blir $\theta_2 = [0.08, 0.48, 0.35, 1.01, 0.13, 0.03, 0.03]$ og vi ser at CD er det tofaktorsamspillet som kommer ut aktivt, 1.01 i motsetning til maks 0.48. Går vi tilbake til tabell 5.13 ser vi at det er J og L som blir påvirket av CD , og vi kan konkludere med at C har verdi i seg selv og ikke er konfundert eller bare påvirket av støy. Da kan vi lage nye plott etter at vi regner ut nye verdier for hovedeffektene. I figur 5.44 ser vi scatterplott av hovedeffektene både med og uten støy. Her ser vi at verdiene er halvparten så store som effektene vanligvis er, som kommer av sånn de blir regnet ut.

Selv med en modell med fire aktive faktorer, av åtte, klarer vi å finne ut ganske masse. Støyen vi har brukt her på $\varepsilon \sim N(0, 0.25)$ er ganske beskjeden, men viser at det går an å finne fram til rett modell med litt støy. Vi har ikke prøvd flere



Figur 5.44: Scatterplott av hovedeffektene til modell $Y = 2A + 3B + C + CD + \varepsilon$, med støy (a), $\varepsilon \sim N(0, 0.25)$ og uten støy (b), etter orthogonalisering.

støynivå her, så vi vet ikke hvor grensen går.

5.4 Kritiske kommentarer

I denne oppgaven har vi sett på ideelle situasjoner der vi som har analysert dataene også har bestemt modell, støynivå og alt annet. Metodene vi bruker er ikke prøvd ut på reelle datasett som har vært undersøkt tidligere og vi vet derfor ikke hvordan metodene fungerer i forhold til tidligere brukte metoder. I tillegg har vi sett på tre forskjellige design som ikke er så lette å sammenligne med hverandre. Vi har sett på et design med 12 forsøk og to design med 16 forsøk. De to designa med 16 forsøk er også forskjellige ved at de ser på enten 8 eller 12 faktorer. Det er altså ikke så lett å bestemme hvilken av de designa vi har sett på som er best. Det vi kan si er at det definitivt er et pluss for de designa der tofaktorsamspill ikke er konfunderte med tofaktorsamspill. Dvs. at 12 forsøks PB design og H_6 er bedre, når det gjelder dette, enn H_2 . H_2 er bedre enn de andre hvis vi har flere faktorer og vet at det er en enkel modell vi skal finne.

En annen forutsetning vi gjør er at vi antar at støy skapt av utenforliggende årsaker er stabil. Det vil si at vi satser på at denne ikke forandrer seg over tid eller på grunn av oppsettet. I våre modeller som vi lager selv er dette lett å fikse. Men dette kan vi selvfølgelig ikke vite på forhånd når vi gjør reelle forsøk. Det er allikvel en forutsetning for at vi skal kunne analysere dataene våre. For å detektere evt ustabilitet i støy bruker man i 2^k blokkdeling av forsøket. Dette har vi ikke diskutert for de tre designa vi ser på i denne oppgaven. Blokkdeling fungerer veldig fint i 2^k forsøk, men om det fungerer her har vi ikke sett på.

I tillegg så har vi ikke undersøkt hvor mye støy metoden tåler. I de modellene der vi har sett på ulike støynivå får vi et visst innblikk i hvor masse støy de tåler. Men noe systematisk analyse av det har vi ikke gjennomført. Dette er noe som bør undersøkes nærmere før man bruker metoden aktivt.

Kapittel 6

Konklusjon

I denne oppgaven ser vi på 12 forsøks PB design og forskjellige typer 16 forsøks design. Vi bruker kunnskap om aliasstrukturen til designa for å tolke scatterplott av hovedeffekter og tofaktorsamspill. I tillegg blir klyngeanalyse brukt til å lage dendogram. Disse gjør det enklere å tolke scatterplottene ved å samle de effektene som er like eller tilnærmet like. Etter at vi har funnet effektene som vi tror er aktive bruker vi ortogonalisering for å finne forventningsrette koeffisienter. Vi bruker prinsippet om effektsparsomhet og forutsetter at støyen som påvirker dataene ikke er for stor.

Det å bruke aliasstrukturen til designa, for å tolke resultatene vi får, gir veldig gode resultater. Når vi har kunnskap om hvordan effektene oppfører seg, som resultat av designet, er det mye lettere å finne ut hva som er de faktiske aktive faktorene.

12 forsøks PB design er et design som vanligvis brukes for å finne hovedeffekter. I denne oppgaven, og i prosjektoppgaven skrevet i høst (Niemi 2007), ser vi at det fint går an å finne ut endel om tofaktorsamspill også. Dette på grunn av den delvise konfunderingen mellom hovedeffekter og tofaktorsamspill som er på $\pm\frac{1}{3}$ for alle tofaktorsamspill som ikke er involvert med hovedeffekten. Dette er et bra design å bruke fordi tofaktorsamspill ikke er fullstendig konfunderte med tofaktorsamspill.

16 forsøks design H_2 er et (16,12,3) screen som nesten ikke er brukt i det hele tatt. Aliasstrukturen i dette designet er nokså ryddig, men tofaktorsamspill er fullstendig konfunderte med tofaktorsamspill. Dette gjør at det ikke er mulig å se hvilket av fire eller seks tofaktorsamspillene som egentlig er aktive.

16 forsøks design H_6 er et (16,8,3) design som Grinde (2004) skriver om. Her er det tatt utgangspunkt i de fire første kolonnene i H_5 og deretter fire utvalgte kolonner fra samme design. Disse er utvalgt for å få et design der tofaktorsamspill ikke er

konfunderte med tofaktorsamspill.

Ser vi på de forskjellige modellene som vi har sett på, så ser vi at så lenge vi har inntill tre faktorer som er aktive fungerer det bra med alle tre designa. Det eneste som kan ødelegge er hvis vi i H_2 har faktorer fra S_1 , S_2 og S_3 samtidig. For eksempel sammenhengen mellom DJ , DM , A og B som er $DJ - DM = A + B$. Men selv med fire faktorer ser vi at vi kan finne ut en del fra $2A + 2D - AE + AF$ som er prøvd ut med H_2 . Av de modellene vi ser på så virker det som om ortogonalisering fungerer bra for alle. Det som kompliserer ting i H_2 er hvis vi har faktorer fra de tre gruppene samtidig og den fullstendig konfunderingen mellom tofaktorsamspill. Men alt i alt så finner vi bedre plott ved bruk av ortogonalisering.

Videre arbeid på dette område virker veldig spennende. En mulighet er å se nærmere på H_3 , H_4 og H_5 . Dette er design som ligner på H_2 , men som har mer komplisert aliasstruktur. Den kompliserte aliasstrukturen gjør det både vanskeligere å skjønne hva som er konfundert med hva, men som vi har sett i tilfellet med H_6 kan det også være en fordel fordi den kompliserte aliasstrukturen begrenser hva som kan være aktivt. Å se nærmere på H_6 , og evt. andre design med den samme egenskapen at tofaktorsamspill ikke er konfunderte med tofaktorsamspill, virker også spennende.

Bibliografi

- Box, G. E. & Draper, N. R. (1987), *Empirical Model - building and response surfaces*, John Wiley and Sons, Inc, New York.
- Box, G. E., Hunter, W. G. & Hunter, J. S. (1978), *Statistics for Experimenters, An Introduction to Design, Data Analysis and Model Building*, John Wiley and Sons, New York.
- Box, G. E. P. & Tyssedal, J. S. (1996), 'Projective Properties of certain Orthogonal Arrays', *Biometrika* **83**(4), 950–955.
- Box, G. E. & Tyssedal, J. S. (2001), 'Sixteen Run Designs of High Projectivity for Factor Screening', *Communications in statistics - Simulation and computation* **30**(2), 217–228.
- Daniel, C. (1959), 'Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments', *Technometrics* **1**(4), 311–341.
- Grinde, H. (2004), 'Screening av 7 til 14 faktorer i 16 eller færre forsøk'. Hovedoppgave ved Fakultet for informasjonsteknologi, matematikk og elektronikk, NTNU.
- Hall, M. J. (1961), 'Hadamard Matrices of Order 16', *Jet Propulsion Laboratory* **1**(36-10), 21–26.
- Hamada, M. & Wu, C. (1992), 'Analysis of Designed Experiments with Complex Aliasing', *Journal of Quality Technology* **24**(3), 130–137.
- Lenth, R. V. (1989), 'Quick and easy analysis of unreplicated factorials', *Technometrics* **31**(4), 469–473.
- Lin, D. K. & Draper, N. R. (1993), 'Generating alias relationships for two-level Plackett and Burman designs', *Computational Statistic & Data Analysis* **15**.

- Niemi, R. (2007), 'Analyse av Plackett and Burman design med 12 enkeltforsøk ved hjelp av klyngeanalyse'. Prosjektoppgave ved Fakultet for informasjonsteknologi, matematikk og elektronikk, NTNU.
- Plackett, R. L. & Burman, J. P. (1946), 'The Design of Optimum Multifactorial Experiments', *Biometrika* **33**(4), 305–325.
- Tyssedal, J. S. & Samset, O. (1997), 'Analysis of the 12 run Plackett and Burman design'. Preprint, Statistics No. 8/1997, Norwegian University of Science and Technology, Norway.
- Tyssedal, J. S. & Samset, O. (1999), 'Two-Level Designs with Good Projection Properties'. Preprint, Statistics No. 12/1999, Norwegian University of Science and Technology, Norway.

Tillegg A

Ortogonale design med 16 forsøk

	a	b	c	d	abc	abd	acd	bcd					ab		
	A	B	C	D	E	F	G	H	J_3	K_3	L_3	M_3	P_3	Q_3	R_3
1	-	-	-	-	-	-	-	-	+	-	-	-	+	+	+
2	+	-	-	-	+	+	+	-	-	+	-	-	-	+	-
3	-	+	-	-	+	+	-	+	-	-	+	-	+	-	-
4	+	+	-	-	-	-	+	+	-	-	-	+	-	-	+
5	-	-	+	-	+	-	+	+	+	+	+	-	-	-	+
6	+	-	+	-	-	+	-	+	+	-	+	+	-	+	-
7	-	+	+	-	-	+	+	-	+	+	-	+	+	-	-
8	+	+	+	-	+	-	-	-	-	+	+	+	+	+	+
9	-	-	-	+	-	+	+	+	-	+	+	+	+	+	+
10	+	-	-	+	+	-	-	+	+	+	-	+	+	-	-
11	-	+	-	+	+	-	+	-	+	-	+	+	-	+	-
12	+	+	-	+	-	+	-	-	+	+	+	-	-	-	+
13	-	-	+	+	+	+	-	-	-	-	-	+	-	-	+
14	+	-	+	+	-	-	+	-	-	-	+	-	+	-	-
15	-	+	+	+	-	-	-	+	-	+	-	-	-	+	+
16	+	+	+	+	+	+	+	+	+	-	-	-	+	+	+

Tabell A.1: H_3 , der de første åtte kolonnene gir et 2_{IV}^{8-4} design.

	a	b	c	d	abc	abd	acd	bcd							
	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>J</i> ₄	<i>K</i> ₄	<i>L</i> ₄	<i>M</i> ₄	<i>P</i> ₄	<i>Q</i> ₄	<i>R</i> ₄
1	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+
2	+	-	-	-	+	+	+	-	+	-	-	+	-	-	+
3	-	+	-	-	+	+	-	+	-	-	+	+	+	-	-
4	+	+	-	-	-	-	+	+	+	-	+	-	-	+	-
5	-	-	+	-	+	-	+	+	-	-	-	-	+	+	+
6	+	-	+	-	-	+	-	+	-	+	-	+	-	+	-
7	-	+	+	-	-	+	+	-	-	+	+	-	-	-	+
8	+	+	+	-	+	-	-	-	+	+	-	-	+	-	-
9	-	-	-	+	-	+	+	+	+	+	-	-	+	-	-
10	+	-	-	+	+	-	-	+	-	+	+	-	-	-	+
11	-	+	-	+	+	-	+	-	-	+	-	+	-	+	-
12	+	+	-	+	-	+	-	-	-	-	-	-	+	+	+
13	-	-	+	+	+	+	-	-	+	-	+	-	-	+	-
14	+	-	+	+	-	-	+	-	-	-	+	+	+	-	-
15	-	+	+	+	-	-	-	+	+	-	-	+	-	-	+
16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Tabell A.2: H_4 , der de første åtte kolonnene gir et 2_{IV}^{8-4} design.

	a	b	c	d	abc	abd									ab
	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i> ₅	<i>H</i> ₅	<i>J</i> ₅	<i>K</i> ₅	<i>L</i> ₅	<i>M</i> ₅	<i>P</i> ₅	<i>Q</i> ₅	<i>R</i> ₅
1	-	-	-	-	-	-	+	+	+	+	+	+	-	-	+
2	+	-	-	-	+	+	+	-	+	-	+	-	+	-	-
3	-	+	-	-	+	+	-	+	+	-	-	+	-	+	-
4	+	+	-	-	-	-	-	-	-	-	+	+	+	+	+
5	-	-	+	-	+	-	+	+	-	-	-	-	+	+	+
6	+	-	+	-	-	+	-	+	-	+	-	+	+	-	-
7	-	+	+	-	-	+	+	-	-	+	+	-	-	+	-
8	+	+	+	-	+	-	-	-	+	+	-	-	-	-	+
9	-	-	-	+	-	+	-	-	+	+	-	-	+	+	+
10	+	-	-	+	+	-	+	-	-	+	-	+	-	+	-
11	-	+	-	+	+	-	-	+	-	+	+	-	+	-	-
12	+	+	-	+	-	+	+	+	-	-	-	-	-	-	+
13	-	-	+	+	+	+	-	-	-	-	+	+	-	-	+
14	+	-	+	+	-	-	-	+	+	-	+	-	-	+	-
15	-	+	+	+	-	-	+	-	+	-	-	+	+	-	-
16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Tabell A.3: H_5 er arrangert sånn at et $(16,14,3)$ design fåes av de 14 første kolonnene.