# To-nivå faktorielle forsøk og blokkdeling.

I regresjonsmodellen $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$ er designmatrisa $\mathbf{X}$ av avgjerande betydning for kor lett det er å finne ein god modell. Spesielt har vi sett (kapittel 12.7) at dersom kolonnene i design-matrisa, $1, \mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_k$, er ortogonale, så er vektoren av estimatorar for koeffisientane gitt ved:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 1/n & 0 & 0 & 0 \\ 0 & (\mathbf{x}_1'\mathbf{x}_1)^{-1} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & (\mathbf{x}_k'\mathbf{x}_k)^{-1} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \mathbf{x}_1'\mathbf{Y} \\ \vdots \\ \mathbf{x}_k'\mathbf{Y} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} \frac{Y_i}{n} \\ (\mathbf{x}_1'\mathbf{x}_1)^{-1}\mathbf{x}_1'\mathbf{Y} \\ \vdots \\ (\mathbf{x}_k'\mathbf{x}_k)^{-1}\mathbf{x}_k'\mathbf{Y} \end{bmatrix} \qquad (1)$$

Vi observerer at estimatoren for $\beta_j$ avheng berre av $\mathbf{x}_j$ og $\mathbf{Y}$. Det kan og visast at ortogonale kolonner, gjev minst varians i estimatorane for koeffisientane.

Når ein utfører forsøk står ein rimeleg fritt til å velge verdiar for forklaringsvariablane $x_1, x_2, \ldots, x_k$. Ein bør då velge desse slik at det er mest gunstig for estimeringa.

Eksempel 1

Samanhengen mellom utbyte av ein kjemisk prosess og dei to faktorane temperatur og konsentrasjon skulle undersøkjast. Det blei utført 4 forsøk der ein nytta 2 verdiar for kvar av faktorane temperatur og konsentrasjon. Dette gjev 4 mogelege nivåkombinasjonar av dei to faktorane til å teste ut utbytte. Forsøket er sett opp i tabellen nedanfor, der og det registrerte utbyte av responsen er gitt:

| Forsøksnr. | Temperatur | Konsentrasjon | Utbyte |
|---|---|---|---|
| 1 | 160 | 20 | 60 |
| 2 | 180 | 20 | 72 |
| 3 | 160 | 40 | 54 |
| 4 | 180 | 40 | 68 |

Ein modellsamanheng av typen $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$ kan då estimerast ut i frå dataane der dei 4 verdiane for utbyte er dei observerte responsverdiane og designmatrisa $\mathbf{X}$ består av ei kolonne av 1-tal, ei kolonne med verdiane for temperatur, ei med verdiane for konsentrasjon og ei kolonne med produktet av verdiane for temperatur og konsentrasjon.

$$\begin{bmatrix} 1 & 160 & 20 & 3200 \\ 1 & 180 & 20 & 3600 \\ 1 & 160 & 40 & 6400 \\ 1 & 180 & 40 & 7200 \end{bmatrix}$$

**Regression Analysis: utbyte versus temp; kons; temp*kons**

```
The regression equation is
utbyte = - 14,0 + 0,500 temp - 1,10 kons + 0,00500 temp*kons

Predictor        Coef
Constant     -14,0000
temp         0,500000
kons         -1,10000
temp*kon   0,00500000
```

I første omgang skal vi berre feste oss ved dei estimerte koeffisientane :

$$\begin{bmatrix} b_0 = -14 \\ b_1 = 0.5 \\ b_2 = -1.1 \\ b_{12} = 0.005 \end{bmatrix}$$

La oss no kode om faktorane ved å innføre nye faktorar $z_1 = \frac{x_1 - 170}{10}$ og $z_2 = \frac{x_2 - 30}{10}$. Verdiane for faktorane er altså sentrert, og i tillegg har vi delt ned på halvparten av avstanden mellom høgt og lågt nivå.

Design matrisa i $z_1, z_2$ og $z_{12}$ blir då:

$$\begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Legg merke til at den nye designmatrisa har berre ortogonale kolonner slik at om vi skal rekne ut koeffisientane får vi frå (1) at

$$b_0^* = \frac{60 + 72 + 54 + 68}{4} = 63.5$$
$$b_1^* = \frac{-60 + 72 - 54 + 68}{4} = 6.5$$
$$b_2^* = \frac{-60 - 72 + 54 + 68}{4} = -2.5$$
$$b_{12}^* = \frac{60 - 72 - 54 + 68}{4} = 0.5$$

Kontrollutrekning med MINITAB gjev no for omkoda verdiar av temperatur og konsentrasjon:

**Regression Analysis: utbyte versus temp; kons; temp*cons**

```
The regression equation is
utbyte = 63,5 + 6,50 otemp - 2,50 okons + 0,500 otemp*ocons

Predictor        Coef
Constant      63,5000
temp          6,50000
kons         -2,50000
temp*con     0,500000
```

For å sjå at vi har same modell kan vi sette inn for $z_1, z_2$ og $z_{12}$ som gjev:

$$\hat{y} = 63.5 + \frac{6.5(x_1 - 170)}{10} - 2.5\frac{(x_2 - 30)}{10} + 0.5\frac{(x_1 - 170)(x_2 - 30)}{10 \cdot 10}$$

$$= 63.5 - 110.5 + 7.5 + 25.5 + 0.65x_1 - 0.15x_1 - 0.25x_2 - 0.85x_2 + 0.005x_1x_2$$
$$= -14 + 0.5x_1 - 1.1x_2 + 0.005x_1x_2$$

Når vi skal analysere to-nivå forsøk, er det oftast mest praktisk å omkode faktorverdiane til 1 og –1 som ovanfor. Då får vi ortogonale designkolonner, og det er lett å rekne ut koeffisientane.
Eit anna argument er at vi ofte har kvalitative variablar som t. d: Vi skal teste ut 2 forskjellige merker eller ein vil teste ut kva som skjer med og utan ei behandling. I det siste tilfelle vil ein gjerne og ha eit mål på kva effekt behandlinga har på responsen.
La faktorverdien 1 svare til behandling og –1 til ikkje behandling. Gjennomsnittsverdi når faktoren er på høgt nivå (1) – gjennomsnittsverdi når faktoren er på lågt nivå (-1) er då av interesse.

Definisjon hovedeffekt:

*For to-nivå forsøk definerer ein hovedeffekten av ein faktor som: Forventa gjennomsnittsrespons når faktoren er på høgt nivå – forventa gjennomsnittsrespons når faktoren er på lågt nivå.*

Det er naturleg at estimatet for denne effekten då blir $\bar{y}_H - \bar{y}_L$ som for temperatur blir :

$$\frac{72+68}{2} - \frac{60+54}{2} = 13 = 2b_1^*$$

og for konsentrasjon:

$$\frac{54+68}{2} - \frac{72+60}{2} = -5 = 2b_2^*$$

Estimert hovedeffekt av ein faktor vil alltid bli den tilhøyrande regresjonskoeffisienten multiplisert med 2 sidan ein hovedeffekt måler forandring i respons når vi går frå lågt (-1) til høgt nivå (1) som for vår omkoding er 2 einingar. Regresjonskoeffisienten på si side fortel om forandring i respons når faktoren endrar seg frå 0 til 1.

I vårt eksempel er koeffisienten føre $x_1x_2$ leddet, $b_{12}$, liten. Denne koeffisienten fortell oss om det er samspel mellom faktorane eller ikkje.

Definisjon:

*Samspelet mellom to faktorar er definert som: Halvdelen av hovedeffekten av ein faktor når den andre er på høgt nivå – halvdelen av hovedeffekten av faktoren når den andre er på lågt nivå.*

For å estimere samspelet mellom temperatur og konsentrasjon skal vi altså rekne ut:

Estimert hovedeffekt av temperatur når konsentrasjon er på høgt nivå gitt ved:

$$68 - 54 = 14$$

Estimert hovedeffekt av temperatur når konsentrasjon er på lågt nivå gitt ved:

$$72 - 60 = 12$$

Estimert samspel blir då $\frac{14}{2} - \frac{12}{2} = 1 = 2b_{12}^*$.

**Forteikn for å rekne ut kontrastane.**

For to-nivå forsøk gjeld: Alle kvantitative nivå kan kodast om til –1 og 1. Alle kvalitative nivå kan naturleg setjast til dette.
Om vi vedtek konvensjonen at høgt nivå av ein faktor svarar til 1 og lågt nivå til –1, ser vi ut i frå det som er gjort ovanfor at estimering av effektar kan gjerast ved å leggje saman responsverdiane med forteikn bestemt av forteikna i designmatrisa for deretter å dele ned på halvparten av antall observasjonar. Difor lagar vi oss ofte berre ei forteiknsmatrise der dei naudsynte forteikna for utrekning av effektar står. I vårt eksempel blir dette:

| Temp | Konsentrasjon | Temp*Kons |
|---|---|---|
| - | - | + |
| + | - | - |
| - | + | - |
| + | + | + |

**Andre notasjonar for faktorkombinasjonar.**

Høgt nivå markert med bokstaven for faktoren.
Lågt nivå markert med 1
1 er uteleten dersom andre bokstavar er brukt:

Eks:

| A | B | Nivåkode |
|---|---|---|
| - | - | 1 |
| + | - | a |
| - | + | b |
| + | + | ab |

### $2^3$ **forsøk.**

Eit $2^k$ forsøk har k faktorar, kvar på 2 nivå. I eksempel 1 såg vi på eit $2^2$ forsøk og korleis vi kunne estimere effektane i eit slikt forsøk. Eigenleg var det i dette forsøket med ein tredje kvalitativ faktor som var katalysator. Med tre faktorar kvar på to nivå er det mogleg å lage 8 mogelege nivåkombinasjonar av høg og låg eller + og – som vi gjerne omkodar til. La A vere Temperatur, B konsentrasjon og C katalysator. Då kan vi setje opp følgjande forteiknsmatrise utvida med nivåkode og responsverdiar.

| A | B | C | AB | AC | BC | ABC | Nivåkode | y |
|---|---|---|---|---|---|---|---|---|
| - | - | - | + | + | + | - | 1 | 60 |
| + | - | - | - | - | + | + | a | 72 |
| - | + | - | - | + | - | + | b | 54 |
| + | + | - | + | - | - | - | ab | 68 |
| - | - | + | + | - | - | + | c | 52 |
| + | - | + | - | + | - | - | ac | 83 |
| - | + | + | - | - | + | - | bc | 45 |
| + | + | + | + | + | + | + | abc | 80 |

Estimerte hovedeffektar blir:

$$\hat{A} = \frac{72+68+83+80}{4} - \frac{60+54+52+45}{4} = 23$$
$$\hat{B} = \frac{54+68+45+80}{4} - \frac{60+72+52+83}{4} = -5$$
$$\hat{C} = \frac{52+83+45+80}{4} - \frac{60+72+54+68}{4} = 1.5$$

Når det gjeld samspelet mellom A og B, AB, skal vi finne hovedeffekten av A når B er på høgt nivå og trekkje frå hovedeffekten av A når B er på lågt nivå for deretter å dele på 2. Når vi estimerer effekten skal vi altså der B er på høgt nivå, leggje saman responsverdiane med forteikna i kolonna til A og der B er på lågt nivå skal vi leggje dei saman med motsette forteikn av det som står i kolonna til A. Men dette svarar til å bruke forteikna i ei kolonne der forteikna i kolonne A og kolonne B er multiplisert saman det vil sei forteikna i kolonna for AB. Tilsvarande får ein forteikna i dei andre samspelskolonnene. Forteikna i kolonnene for trefaktorsamspelet ABC får ein ved å multiplisere saman forteikna i kolonnene for A, B og C. Dette gjev:

$$\hat{AB} = \frac{60+68+52+80}{4} - \frac{72+54+83+45}{4} = 1.5$$
$$\hat{AC} = \frac{60+54+83+80}{4} - \frac{72+68+52+45}{4} = 10$$
$$\hat{BC} = \frac{45+80+60+72}{4} - \frac{83+52++68+54}{4} = 0$$
$$\hat{ABC} = \frac{80+52+54+72}{4} - \frac{83+45+60+68}{4} = 0.5$$

## Vurdering av signifikans av effektar

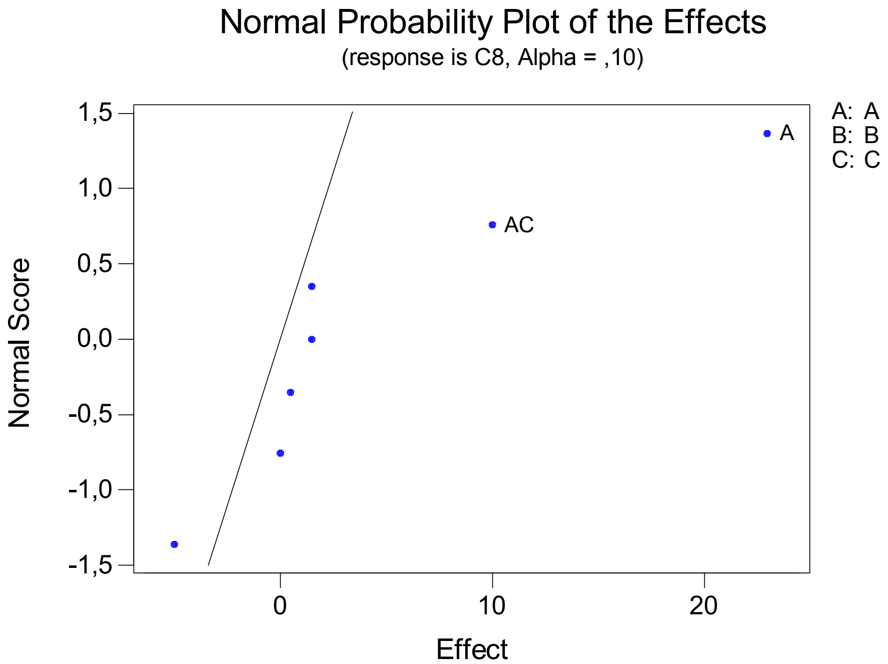
Estimatorane for effektane er gitt ved følgjande formel:

$\hat{\text{Effekt}} = \frac{\sum_{i=1}^{n} \partial_i Y_i}{n/2}$ der n er talet på observasjonar og $\delta_i$ er anten 1 eller –1 avhengig av forteikna i kolonna for den effekten vi reknar ut. Sidan alle $Y_i$ er uavhengige får vi:

$$\text{Var}(\hat{\text{Effekt}}) = \sigma^2_{\text{effekt}} = \frac{\sum_{i=1}^{n} \delta_i^2 \text{Var}(Y_i)}{n^2/4} = \frac{4\sum_{i=1}^{n} \sigma^2}{n^2} = \frac{4\sigma^2}{n}\ .$$

Dersom alle effektane er 0 og dataane er normalfordelte med same varians, $\sigma^2$, vil alle estimatorane for effektane vere $N(0, \frac{4\sigma^2}{n})$. I eit normalplott bør dei tilhøyrande estimata bli liggjande på ei rett linje. Tilhøyrande effektar til dei som fell av linja, kan vi rekne som signifikante.

Eit normalplott for $2^3$ forsøket med temperatur, katalysator og konsentrasjon er synt nedanfor.




Her ser ein at estimerte effektar av temperatur og samspelet mellom temperatur og katalysator klart skil seg frå dei andre.

Ofte kan ein gå ut ifrå at trefaktor- og høgare ordens samspel er 0. Desse kan då brukast til å estimere variansen til effektane.

## Bruk av effektar for å rekne ut signifikans

$2^4$ forsøk.

Ved å ta gjennomsnittet av dei 5 kvadrerte trefaktor- og firefaktorsamspela får vi eit estimat for variansen til effektane med 5 fridomsgrader. Dette kan brukast til å undersøke signifikansen til effektane. Til dømes er faktor A signifikant dersom:

$$\left|\frac{\hat{A}}{s_{\hat{A}}}\right| \geq t_{\frac{\alpha}{2},5}\,.$$

**Eksempel**

Eit $2^4$ forsøk i dei 4 faktorane A =katalysator ladning, B = temperatur, C = trykk og D = konsentrasjon blei utført i eit prosessutviklings-studie.

Dei 16 forsøka oppsett på standardform er synt nedanfor. Her er og alle samspelskolonnene tekne med.

| Row | A | B | C | D | AB | AC | AD | BC | BD | CD | ABC | ABD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | -1 | -1 | -1 | -1 | 1 | 1 | 1 | 1 | 1 | 1 | -1 | -1 |
| 2 | 1 | -1 | -1 | -1 | -1 | -1 | -1 | 1 | 1 | 1 | 1 | 1 |
| 3 | -1 | 1 | -1 | -1 | -1 | 1 | 1 | -1 | -1 | 1 | 1 | 1 |
| 4 | 1 | 1 | -1 | -1 | 1 | -1 | -1 | -1 | -1 | 1 | -1 | -1 |
| 5 | -1 | -1 | 1 | -1 | 1 | -1 | 1 | -1 | 1 | -1 | 1 | -1 |
| 6 | 1 | -1 | 1 | -1 | -1 | 1 | -1 | -1 | 1 | -1 | -1 | 1 |
| 7 | -1 | 1 | 1 | -1 | -1 | -1 | 1 | 1 | -1 | -1 | -1 | 1 |
| 8 | 1 | 1 | 1 | -1 | 1 | 1 | -1 | 1 | -1 | -1 | 1 | -1 |
| 9 | -1 | -1 | -1 | 1 | 1 | 1 | -1 | 1 | -1 | -1 | -1 | 1 |
| 10 | 1 | -1 | -1 | 1 | -1 | -1 | 1 | 1 | -1 | -1 | 1 | -1 |
| 11 | -1 | 1 | -1 | 1 | -1 | 1 | -1 | -1 | 1 | -1 | 1 | -1 |
| 12 | 1 | 1 | -1 | 1 | 1 | -1 | 1 | -1 | 1 | -1 | -1 | 1 |
| 13 | -1 | -1 | 1 | 1 | 1 | -1 | -1 | -1 | -1 | 1 | 1 | 1 |
| 14 | 1 | -1 | 1 | 1 | -1 | 1 | 1 | -1 | -1 | 1 | -1 | -1 |
| 15 | -1 | 1 | 1 | 1 | -1 | -1 | -1 | 1 | 1 | 1 | -1 | -1 |
| 16 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

| Row | ACD | BCD | ABCD |
|---|---|---|---|
| 1 | -1 | -1 | 1 |
| 2 | 1 | -1 | -1 |
| 3 | -1 | 1 | -1 |
| 4 | 1 | 1 | 1 |
| 5 | 1 | 1 | -1 |
| 6 | -1 | 1 | 1 |
| 7 | 1 | -1 | 1 |
| 8 | -1 | -1 | -1 |
| 9 | 1 | 1 | -1 |
| 10 | -1 | 1 | 1 |
| 11 | 1 | -1 | 1 |
| 12 | -1 | -1 | -1 |
| 13 | -1 | -1 | 1 |
| 14 | 1 | -1 | -1 |
| 15 | -1 | 1 | -1 |
| 16 | 1 | 1 | 1 |

Dei 16 responsverdiane for prosentvis omdanning var:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 71 | 61 | 90 | 82 | 68 | 61 | 87 | 80 | 61 | 50 |
| 89 | 83 | 59 | 51 | 85 | 78 | | | | |

**Fractional Factorial Fit: % Conversion versus A; B; C; D**

Estimated Effects and Coefficients for %conversion (coded units)

| Term | Effect | Coef |
|---|---|---|
| Constant | | 72,250 |
| A | -8,000 | -4,000 |
| B | 24,000 | 12,000 |
| C | -2,250 | -1,125 |
| D | -5,500 | -2,750 |
| A*B | 1,000 | 0,500 |
| A*C | 0,750 | 0,375 |
| A*D | -0,000 | -0,000 |
| B*C | -1,250 | -0,625 |
| B*D | 4,500 | 2,250 |
| C*D | -0,250 | -0,125 |
| A*B*C | -0,750 | -0,375 |
| A*B*D | 0,500 | 0,250 |

```
A*C*D         -0,250    -0,125
B*C*D         -0,750    -0,375
A*B*C*D       -0,250    -0,125
```

```
Nscores -
         -                                             *   B
         -
      1,2+                    *    BD
         -               *    AB
         -               *    AC
         -               *    ABD
         -              2    AD,CD
      0,0+              *    ACD
         -             **    ABC,ABCD
         -             *    BCD
         -            *    BC
         -           *    C
     -1,2+      *    D
         -
         -   *    A
         -
           ----+---------+---------+---------+---------+--Effects
             -8,0      0,0       8,0      16,0      24,0

A = A         B = B         C = C         D = D
```

Frå normalplottet kan det verke som hovedeffekten av katalysator ladning, temperatur og konsentrasjon samt samspelet mellom temperatur og konsentrasjon er signifikante.

Vi ser at alle dei estimerte trefaktor- og firefaktorsamspela er små. Dersom dei sanne verdiane for desse er 0, vil estimatorane for desse ha forventing 0 og varians $\sigma^2_{\text{effekt}}$.

Dermed er eit estimator for $\sigma^2_{\text{effekt}}$ gitt ved: $\dfrac{\widehat{ABC}^2 + \widehat{ABD}^2 + \widehat{ACD}^2 + \widehat{BCD}^2 + \widehat{ABCD}^2}{5}$

For våre data blei estimatet: $s^2_{\text{effekt}} = \dfrac{(-0.75)^2 + (0.5)^2 + (-0.25)^2 + (-0.75)^2 + (-0.25)^2}{5} = 0.3$

og standardavviket til effektane blir estimert til $s_{\text{effekt}} = \sqrt{0.3} = 0.55$.

Ved vurdering av signifikans skal absolutt verdien av dei estimerte effektane samanliknast med: $s_{\text{effekt}} \cdot t_{0.025,5} = 0.55 \cdot 2.571 = 1.41$. Såleis er det rom for å undrast om hovedeffekten av trykk og er signifikant.

**Estimering av varians ved gjentak.**

Estimering av varians krev normalt at ein gjer gjentak av forsøket. Dersom eit gjentak er gjort, har ein to responsverdiar for kvar nivåkombinasjon som begge har same forventing.

La $y_{11}$ og $y_{12}$ vere dei observerte responsverdiane for nivåkombinasjon 1. Eit estimat for variansen til observasjonane, $\sigma^2$, er då gitt ved:

$$\sum_{j=1}^{2}(y_{1j}-\bar{y}_1)^2=(y_{11}-\frac{y_{11}+y_{12}}{2})^2+(y_{12}-\frac{y_{11}+y_{12}}{2})^2=$$

$$(\frac{y_{11}-y_{12}}{2})^2+(\frac{-y_{11}+y_{12}}{2})^2=\frac{(y_{11}-y_{12})^2}{2}$$

Normalt får ein $2^k$ slike estimat som kan brukast til å estimere $\sigma^2$ ved å ta gjennomsnittet av dei.

Eksempel. $2^3$ forsøk med gjentak

| A | B | C | $y_{i1}$ | $y_{i2}$ | $y_{i1}-y_{i2}$ | $\frac{(y_{i2}-y_{i1})^2}{2}$ |
|---|---|---|---|---|---|---|
| - | - | - | 59 | 61 | -2 | 2 |
| + | - | - | 74 | 70 | 4 | 8 |
| - | + | - | 50 | 58 | -8 | 32 |
| + | + | - | 69 | 67 | 2 | 2 |
| - | - | + | 50 | 54 | -4 | 8 |
| + | - | + | 81 | 85 | -4 | 8 |
| - | + | + | 46 | 44 | 2 | 2 |
| + | + | + | 79 | 81 | -2 | 2 |
| Totalt | | | | | | 64 |

Estimatet for $\sigma^2$ blir då: $s^2=\frac{64}{8}=8$.

$$\sigma^2_{\text{effekt}}=\frac{4\sigma^2}{n}\Rightarrow s^2_{\text{effekt}}=\frac{4\cdot s^2}{16}=\frac{4\cdot 8}{16}=2$$

Ved (m-1) gjentak (totalt m verdiar for kvar nivåkombinasjon) er $\sum_{j=1}^{m}\frac{(Y_{ij}-\overline{Y}_i)^2}{m-1}$ ein estimator for $\sigma^2$ for kvar i. Ved å midle over desse får vi ein estimator for variansen. Denne har $(m-1)2^k$ fridomsgrader.

## Tolkning av effektar

Dersom ein faktor ikkje har samspel tolkar vi estimert hovedeffekt som estimert forandring i forventa respons når vi går frå lågt til høgt nivå av faktoren. For faktorar som har samspel med andre faktorar blir tolkninga gjort ved hjelp av sampelsplott.

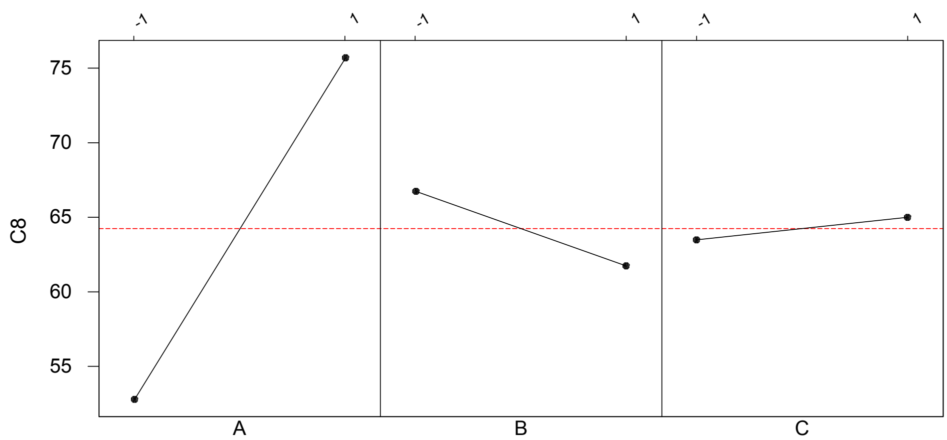
**Eksempel $2^3$ forsøk for kjemisk utbyte.**

Her var estimatet for hovedeffekten av temperatur 23 og samspelet mellom temperatur og katalysator blei estimert til 10. Vi kan no lage oss ein tabell der vi studerer kva som skjer for dei 4 nivåkombinasjonane av A og C. Ei grafisk framstilling av denne tabellen blir kalla eit samspelsplott.

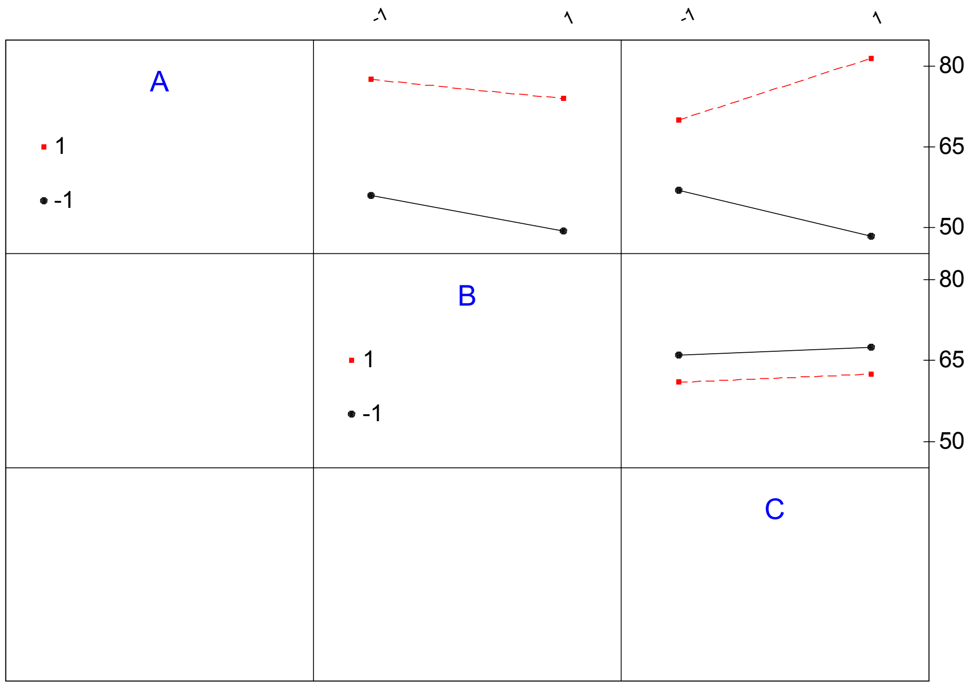
| | C | |
|---|---|---|
| A | - | + |
| - | $\frac{60+54}{2}=57$ | $\frac{52+45}{2}=48.5$ |
| + | $\frac{72+68}{2}=70$ | $\frac{83+80}{2}=81.5$ |

Hovedeffektsplott og samspelsplott frå MINITAB er synt nedanfor. Dersom det ikkje er samspel mellom to faktorar er effekten av ein faktor den same uavhengig av nivået til den andre faktoren. Linjene i tofaktorsamspelsplotta skal då bli parallelle. Dette ser vi langt på veg er tilfelle for faktorane A og B og for B og C, men ikkje for A og C.

Main Effects Plot (data means) for C8



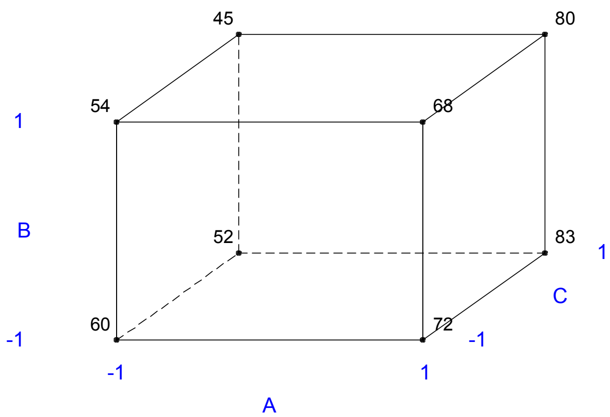
Interaction Plot (data means) for C8




Tolkninga av analysen blir då at effekten av katalysator er negativ når temperaturen er på lågt niva, men positiv når temperaturen er på høgt nivå. Best utbyte får vi når temperatur og katalysator begge er på høgt nivå.

Eit kubeplott illustrerar godt kva nivåkombinasjonar som er gunstige.

Cube Plot (data means) for C8



## Blokkdeling i $2^k$ forsøk

Når ein utfører eit forsøk skal ein alltid utføre det i randomisert rekkjefølgje. Randomisering er vår beste garanti for uavhengige observasjonar og gjer at utanforliggande faktorar får mindre sjanse for å påverke responsen slik at vi trekkjer feilaktige konklusjonar. Det er og viktig å stille inn alle nivåkombinasjonar på nytt mellom kvart forsøk. Dette for å sikre at alle observasjonane får mest mogeleg lik varians.

Dersom vi skal gjere mange forsøk vil det ofte vere slik at ytre forhold forandrar seg frå vi startar forsøket til vi er ferdig. Forandringar i ytre forhold kan påverke responsverdiane og gjere at vi feilestimerer effektane. For å motverke dette, kan vi blokkdele forsøket. Stundom er det andre føringar, mangel på råmateriale t. d som gjer at vi ynskjer å blokkdele eit forsøk. Når vi blokkdeler forsøket, skal randomiseringa skje i kvar blokk.

### $2^3$ forsøk: 2 blokker kvar på 4 forsøk:

Gå ut i frå at vi gjer forsøka der tre-faktor samspelet har – i blokk 1 og dei resterande i blokk 2.

| Forsøk | A | B | C | AB | AC | BC | | ABC |
|---|---|---|---|---|---|---|---|---|
| 1 | - | - | - | + | + | + | | - |
| 4 | + | + | - | + | - | - | Blokk 1 | - |
| 6 | + | - | + | - | + | - | | - |
| 7 | - | + | + | - | - | + | | - |
| | | | | | | | | |
| 2 | + | - | - | - | - | + | | + |
| 3 | - | + | - | - | + | - | Blokk 2 | + |
| 5 | - | - | + | + | - | - | | + |
| 8 | + | + | + | + | + | + | | + |

Vi observerer at om alle enkeltforsøka i blokk 2 får eit tillegg h, så vil utrekning av hovedeffektane og tofaktorsamspela vere upåverka fordi det er like mange – som + i kvar blokk. Dette gjeld ikkje for tre-faktor samspelet som blir konfundert med blokkeffekten.

### $2^3$ forsøk i 4 blokker, kvar på 2 forsøk

Gå ut i frå at vi deler inn i blokker etter tofaktorsamspela AB og BC etter følgjande mønster

| Blokk 1<br>(- -) | Blokk 2<br>(- +) | Blokk 3<br>(+ -) | Blokk 4<br>(+ +) |
|---|---|---|---|

Dette gjev følgjande blokkdeling:

| Blokk | A | B | C | Forsøk | AB | BC | AC | ABC |
|---|---|---|---|---|---|---|---|---|
| Blokk 1 | - | + | - | 3 | - | - | + | + |
| | + | - | + | 6 | - | - | + | - |
| Blokk 2 | + | - | - | 2 | - | + | - | + |
| | - | + | + | 7 | - | + | - | - |
| Blokk 3 | + | + | - | 4 | + | - | - | - |
| | - | - | + | 5 | + | - | - | + |
| Blokk 4 | - | - | - | 1 | + | + | + | - |
| | + | + | + | 8 | + | + | + | + |

AB, BC og AC blir konfunderte med blokkeffekten.

<u>Korleis avgjere kva effektar vi kan blokkdele etter?</u>

Utgangspunkt: Prøver å få estimert hovedeffekten og lågfaktorsamspela. La I vere ei kolonne med berre + teikn. Vi har:

$$I = AA=BB=CC$$

Gå ut i frå at vi blokkdeler eit $2^3$ forsøk etter D=ABC og E=AC. Samspelet mellom D og E, DE=ABCAC=B, det vil sei ein hovedeffektane som då blir konfundert med blokkeffekten i tillegg til ABC og BC.

<u>Generalisering</u>

Gå ut i frå at vi skal dele eit $2^6$ forsøk opp i 8 blokker etter blokkfaktorane $B_1 = ACE$ $B_2 = ABEF$ og $B_3 = ABCD$. Blokkinndelinga følgjer då følgjande mønster:

| Blokk 1 | Blokk 2 | Blokk 3 | Blokk 4 | Blokk 5 | Blokk 6 | Blokk 7 | Blokk 8 |
|---|---|---|---|---|---|---|---|
| (- - -) | (+ - -) | (- + -) | (+ + -) | (- - +) | (+ - +) | (- + +) | (+ + +) |

Vi får:

$$B_1B_2 = ACEABEF = BCF$$
$$B_1B_3 = ACEABCD = BDE$$
$$B_2B_3 = ABEFABCD = CDEF$$
$$B_1B_2B_3 = ACEABEFABCD = ADF$$

Som saman med $B_1 = ACE$, $B_2 = ABEF$ og $B_3 = ABCD$ blir konfundert med blokkeffekten.

## Variansanalysetabellen ved $2^k$ forsøk.

Sidan alle kolonner i designmatrisa er ortogonale har vi frå kapittel 12.7 i læreboka at:

$$SS_R = b_1^2 \sum_{i=1}^{n} x_{1i}^2 + b_2^2 \sum_{i=1}^{n} x_{2i}^2 + \cdots + b_p^2 \sum_{i=1}^{n} x_{pi}^2 \qquad (2)$$

Brukt på to-nivå design der vi har $2^k - 1$ effektkolonner blir p lik $2^k - 1$. Set vi inn at ein koeffisient er lik den tilsvarande effekten delt på 2 og brukar at for to–nivå forsøk er $\sum_{i=1}^{n} x_{1i}^2 = n$, får vi :

$$SS_R = \frac{\hat{A}^2 \cdot n}{4} + \frac{\hat{B}^2 \cdot n}{4} + \cdots + \frac{\hat{A}BC....^2 \cdot n}{4}$$

Dei enkelte ledda ovanfor gjev oss kvadratsummane for effektane:

For eit $2^3$ forsøk blir denne:

| Kjelder | SS | F.G |
|---|---|---|
| A | $SS_A = 2\hat{A}^2$ | 1 |
| B | $SS_B = 2\hat{B}^2$ | 1 |
| C | $SS_C = 2\hat{C}^2$ | 1 |
| AB | $SS_{AB} = 2\hat{A}B^2$ | 1 |
| AC | $SS_{AC} = 2\hat{A}C^2$ | 1 |
| BC | $SS_{BC} = 2\hat{B}C^2$ | 1 |
| ABC | $SS_{ABC} = 2\hat{A}BC^2$ | 1 |
| Total | $SS_T = SS_R = \sum_{i=1}^{n} (y_i - \bar{y})^2$ | 7 |

Legg merke til at her er $SS_R = SS_T$. Det skuldast at når ein tilpassar eit konstantledd og 7 effektar til 8 observasjonar, blir alle residuala 0.

Dersom ein blokkdeler forsøket etter trefaktorsamspelet ABC, kan ein sjå på det som å innføre ein blokkfaktor i modellen istadenfor trefaktorsampelet. Denne blokkfaktoren har to nivå, eit for kvar blokk som vi kan kode til –1 og +1. Variansanalysetabellen blir difor.

| Kjelder | SS | F.G |
|---|---|---|
| A | $SS_A = 2\hat{A}^2$ | 1 |
| B | $SS_B = 2\hat{B}^2$ | 1 |
| C | $SS_C = 2\hat{C}^2$ | 1 |
| AB | $SS_{AB} = 2\hat{A}B^2$ | 1 |
| AC | $SS_{AC} = 2\hat{A}C^2$ | 1 |
| BC | $SS_{BC} = 2\hat{B}C^2$ | 1 |
| Blokk | $SS_{blokk} = 2\hat{A}BC^2$ | 1 |
| Total | $SS_T = \sum_{i=1}^{n}(y_i - \bar{y})^2$ | 7 |

For å dele eit forsøk inn i 4 blokker, kan ein tenkje seg å innføre ein faktor som har 4 nivå. Men 4 nivå kan representerast ved 2 faktorar som begge har 2 nivå. Såleis kan vi plukke ut to effektkolonner og la desse to bestemme dei 4 nivåa for blokkene. Samspelet mellom desse kolonnene vil og gå inn i blokkeffekten.

For eit $2^3$ forsøk i 4 blokker delt inn etter AB og BC (AC) samspelet, får vi denne variansanalysetabellen

| Kjelder | SS | F.G |
|---|---|---|
| A | $SS_A = 2\hat{A}^2$ | 1 |
| B | $SS_B = 2\hat{B}^2$ | 1 |
| C | $SS_C = 2\hat{C}^2$ | 1 |
| Blokk | $SS_{blokk} = 2\hat{A}B^2 + 2\hat{B}C^2 + 2\hat{A}C^2$ | 3 |
| ABC | $SS_{ABC} = 2\hat{A}BC^2$ | 1 |
| Total | $SS_T = \sum_{i=1}^{n}(y_i - \bar{y})^2$ | 7 |

La oss no gå ut i frå at det blir gjort eit gjentak av eit $2^3$ forsøk, og at heile forsøket blir gjort i 4 blokker. Då kan ein bruke ei effektkolonne (kolonna for ABC samspelet) til å blokkdele etter i kvart av dei to gjentaka. Den andre blokkfaktoren vil ha –1 for dei 8 første forsøka og +1 for dei 8 siste. Begge blokkfaktorkolonnene er ortogonale på dei andre kolonnene. Det vil og samspelskolonna for blokkfaktorane vere.

La gjennomsnitta i dei 4 blokkene vere gitt ved: $\bar{y}_1, \bar{y}_2, \bar{y}_3$ og $\bar{y}_4$. Frå (2) blir då blokkeffekten:

$$SS_{blokk} = 16(\frac{\bar{y}_2 + \bar{y}_4}{4} - \frac{\bar{y}_1 + \bar{y}_3}{4})^2 + 16(\frac{\bar{y}_3 + \bar{y}_4}{4} - \frac{\bar{y}_1 + \bar{y}_2}{4})^2 + 16(\frac{\bar{y}_1 + \bar{y}_4}{4} - \frac{\bar{y}_2 + \bar{y}_3}{4})^2$$

Variansanalysetabellen blir:

| Kjelder | SS | F.G |
|---|---|---|
| A | $SS_A = 4\hat{A}^2$ | 1 |
| B | $SS_B = 4\hat{B}^2$ | 1 |
| C | $SS_C = 4\hat{C}^2$ | 1 |
| AB | $SS_{AB} = 4\hat{A}\hat{B}^2$ | 1 |
| AC | $SS_{AC} = 4\hat{A}C^2$ | 1 |
| BC | $SS_{BC} = 4\hat{B}C^2$ | 1 |
| Blokk | $SS_{blokk}$ | 3 |
| Feil | $SS_E = SS_T - (SS_A + SS_B + \cdots + SS_{blokk})$ | 6 |
| Total | $SS_T = \sum_{i=1}^{n} (y_i - \bar{y})^2$ | 15 |

**Partiell konfundering**

Ved gjentak kan ein dele inn i blokker etter forskjellige samspel kvar gong. Dette kallast partiell konfundering. Brukar ein ABC samspelet ein gong og AB samspelet neste gong, vil ein 1. gongen kunne estimere AB samspelet t. d. og andre gongen ABC samspelet o.s.|b.

Blokkeffekten bør no reknast ut etter den generelle formelen:

$$SS_{blokk} = k\sum_{i=1}^{m}\sum_{b=1}^{s} (\bar{y}_{ib} - \bar{y}_{...})^2$$

der k er talet på observasjonar i kvar blokk, m er talet på gjentak, s er talet på blokker i kvart gjentak, $\bar{y}_{ib}$ er gjennomsnittet i blokk b i gjentak i og $\bar{y}_{...}$ er gjennomsnittet av alle observasjonane.

Variansanalysetabellen for eit gjentatt $2^3$ forsøk blir:

| Kjelder | SS | F.G |
|---|---|---|
| A | $SS_A = 4\hat{A}^2$ | 1 |
| B | $SS_B = 4\hat{B}^2$ | 1 |
| C | $SS_C = 4\hat{C}^2$ | 1 |
| AB | $SS_{AB} = 2\hat{A}B^2$ | 1 |
| AC | $SS_{AC} = 4\hat{A}C^2$ | 1 |
| BC | $SS_{BC} = 4\hat{B}C^2$ | 1 |
| ABC | $SS_{ABC} = 2\hat{A}BC^2$ | 1 |
| Blokk | $SS_{blokk}$ | 3 |
| Feil | $SS_E = SS_T - (SS_A + SS_B + \cdots + SS_{blokk})$ | 5 |
| Total | $SS_T = \sum_{i=1}^{n}(y_i - \bar{y})^2$ | 15 |