Statistisk forsøksplanlegging og analyse

for forbedring av produkter og prosesser

Innhold

Forord

Statistisk forsøksplanlegging (DOE) og analyse er et slagkraftig verktøy ved forbedring av produkter og prosesse, gjerne brukt i kombinasjon med statistisk prosesstyring (SPC). Dette er en kort elementær innføring i sentrale ideer og prinsipper i moderne forsøksplanlegging og analyse. Vi vil ikke forutsette spesielle forkunnskaper i statistisk teori av leseren, og vi vil også styre klar av tekniske detaljer knyttet til statistisk teori i presentasjonen av stoffet. Et eget avsnitt tar opp til diskusjon noen av de sentrale statistiske modellene innen feltet, dels for de som har noe forkunnskaper i statistisk teori, og som ønsker å reaktivere eller utvide disse. Vi gir også omfattende eksempler på utskrifter fra programvare (Minitab).

Fremstillingen dekker klassisk forsøksplanlegging i den anglo/amerikanske tradisjon, supplert med innslag fra den japanske tradisjon (Taguchi). En representant for den anglo/amerikanske tradisjonen er G.E.P.Box, som har gitt sentrale bidrag til feltet, både når det gjelder teori og tilrettelegging av metoder for praktisk bruk. Denne fremstilling er i stor grad inspirert av han. Merk at noen misvisende kaller hele feltet Taguchi-metoder.

Innen en del bransjer bl.a. næringsmiddel, har man de senere år tatt i bruk relativt avanserte multivariate statistiske metoder, ofte av eksplorativ og dataintensiv karakter. Slike er ikke tatt med her.

Kompendiet er skrevet ut fra behovet for materiale for opplæring i statistiske metoder som oppfyller kravene til kvalitetspersonell iht. PRS-ordningen administrert i Norge av Norsk Forening for Kvalitet og Lederskap, jfr. The EOQ-Personal Registration Scheme: Rules and Handbook. 13. Statistical Methods. Kompendiet inneholder nok materiale til å dekke 13.5 Design of experiments.

Bergen, 14 januar 2000

Jostein Lillestøl, Norges Handelshøyskole

Fagstyret for statistiske metoder,
Norsk Forening for Kvalitet og Lederskap

1. Innledning

Eksperimenter basert på velfunderte forsøksplaner kan bidra sterkt til forbedring av produkter og prosesser. Det gjelder mange aspekter av det vi oppfatter som kvalitet: ytelse, pålitelighet, brukervennlighet, vedlikeholdsfri osv. En rekke faktorer er bestemmende for kvalitet, så som produktdesign, maskindesign, produksjonsrutiner og råmaterialer.

I kvalitetsstyring og forbedring snakker man om å gjøre "de riktige tingene" og "å gjøre tingene rett". Eksperimenter basert på velfunderte forsøksplaner kan bidra til dette. Forsøksplanlegging kan kombineres med andre verktøy for styring og forbedring av kvalitet, ved kartlegging og fjerning av årsaker til problemer og ved forbedring av prosesser, f.eks i kombinasjon med statistisk prosesstyring. Med en prosess i statistisk kontroll har vi to muligheter mht. å bedre kvaliteten: endre kvalitetsnivået og redusere variasjonen.

Det vil som regel være slik at mange faktorer påvirker et resultat, men få er viktige. Det dreier seg ofte om å finne disse ("screening") og bestemme gode parameterverdier for faktorene (optimalisering). Dette kan skje ved å kartlegge den såkalte responsflaten. Det er også viktig å få mest mulig ut av begrensede ressurser. Et godt planlagt eksperiment kan gi mye mer informasjon med færre gjentak ("runs") enn et eksperiment uten plan eller planlagt i god tro. Fallgruvene er mange! Spesielt vil vi advare mot troen på at variasjon av en faktor om gangen er å foretrekke fremfor å variere flere. Snarere tvert i mot!

Statistisk eksperimentplanlegging og analyse kan bidra til å finne bedre kombinasjoner av faktorer enn de som er i bruk til nå. Det er ofte mulig å gruppere faktorer i henhold til om de i hovedsak kan påvirke nivå eller variasjon, og det kan være et spørsmål hvilke en bør gripe fatt i først, eller om de kan studeres samtidig. Mange foretrekker å angripe variabilitet først, og når denne er redusert studeres mulighet for forbedring i nivå. Vi vil i første omgang fokusere på endring av nivå.

2. En faktor: Sammenligning av to grupper

Den enkleste problemstilling i forsøksplanlegging er å sammenligne to grupper eller nivåer for en faktor. Det kan f.eks. være i produksjon der en sammenligner to metoder, to maskiner eller to materialer. Det kan også være testing av to produkter mot hverandre. Vi vil i denne sammenheng ta for oss to situasjoner:

- To uavhegige grupper (komplett randomisering)
- Parret sammenligning (randomiserte blokker)

Følgende eksempel vil kunne klargjøre begrepene:

Eksempel 1: A eller B best?
Et taxiselskap bruker dekktype A, men har fått tak i 8 dekk av en ny type B som de ønsker å sammenligne med A mht. slitestyrke. En ønsker å teste dekkene ved variert bruk, og slik at ingen av dekktypene blir favorisert. En vil derfor også bruke 8 dekk av type A.

Det fins mange ulike mulige forsøksplaner: Plan 1a er å montere de 16 dekkene på 4 biler, med 4 av samme type på hver, slik at to biler kjører med A.dekk og to biler kjører med B-dekk. Plan 1b er å bruke 16 biler og erstatte et av dekkene der med et av de 16 dekkene som er med i forsøket. Ved Plan 1b reduserer en risikoen for at eventuelle forskjeller i slitasje bare skyldes ulik kjørestil hos den (de) som har brukt de to bilene med samme type dekk. En mulig favorisering ved Plan 1a kan unngås ved å trekke lodd blant de 4 aktuelle biler (randomisering).
En helt annen mulighet er å montere dekk av begge typer på hver bil. Plan 2a bruker to A-dekk og to B-dekk på hver av de 4 bilene, mens plan 2b bruker ett A-dekk og ett B-dekk på hver av 8 biler.

Det kan gis argumenter for at Plan 1b og 2b er å foretrekke fremfor Plan 1a og 2a. Hvis en skal generalisere resultatet fra de benyttede biler til hele bilparken, er det viktig at dekkene blir prøvd ut under så varierte forholde som mulig. Verdieneav å observere to dekk av samme type på samme bil er derfor trolig liten, med mindre slitasjen varierer betydelig med hvor på bilen dekket er montert, foran/bak eller høyre/venstre. Skulle dette være tilfelle, må en ved bruk av Plan 1a og 2a randomisere plasseringen for å være upartisk.

La oss derfor se nærmere på Plan 1b og Plan 2b, og heretter kalle disse Plan 1 og Plan 2.

Plan 1: Komplett randomisering
Blant 16 busser velges <u>tilfeldig</u> 8, som får ett nytt A-dekk, mens de øvrige 8 busser får ett nytt B-dekk. Slitasjen måles så etter en forutbestemt kjørelengde.

Merknad. Her er det rimelig å montere alle i samme posisjon på bilen. Hvis dette ikke gjøres, må monteringen skje på en måte som ikke favoriserer, f.eks. ved at alle posisjoner er representert like ofte for begge typer dekk, eller ved at posisjonen er randomisert.

Plan 2: Randomiserte blokker
På hver av 8 biler monteres to dekk av hver type. Plassering må være upartisk, noe som sikres best ved å velge den tilfeldig. Slitasjen måles så etter en forutbestemt kjørelengde.

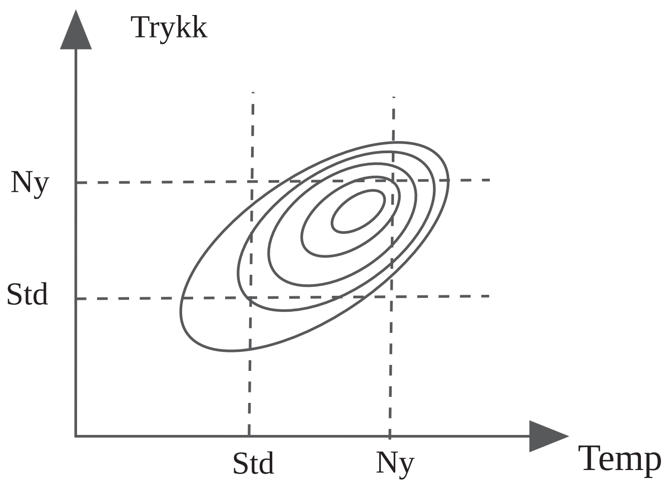
Merknad. Her kunne en nok tenke seg å plassere alle dekk enten foran eller bak, og evt. randomisere mellom de to mulige valgene for A-dekket og B-dekket, eller velge et systematisk og upartisk mønster. Imidlertid skal en vokte seg for systematikk som kanskje bare tilsynelatende er upartisk. Randomisering er en god leveregel dersom en er usikker og en ønsker å begrense seg til forsøksplaner basert på en faktor. Her vil posisjon kunne trekkes inn som en (to) ekstra faktor(er), men dette vil selvsagt komplisere analysen – trolig unødig. Merk også at ved Plan 2 er det ikke av samme betydning at de ulike bilene har eksakt samme kjørelengde når målingene foretas.

Vi vil komme tilbake til de statistiske modeller som vanligvis legges til grunn for sammenligning av to grupper i et eget avsnitt

3. Flere faktorer: En om gangen?

La oss kort illustrere dette tema med et eksempel:

Eksempel 2: Ytelse vs. temperatur og trykk
Anta at temperatur og trykk er bestemmende for ytelsen og at sammenhengen er som i Figur 1, der ytelsen er illustrert ved kotene i et kart, høyest er best. Vi må da tenke oss responsflaten som terrenget over (Temp,Trykk)-planet.



Figur 1 Responsflate

Vi ser at det er samspill/interaksjon mellom de to faktorene, og at vi for å få best mulig resultat må bevege oss i nord-østlig retning for å forbedre ytelsen i forhold dagens standard parameterverdier (markert med Std). Et forsøk på å se hva som skjer ved å variere temperatur, men beholde standardverdien for trykk, vil ikke gi oss den nødvendige innsikt. Heller ikke ved å variere trykk, men beholde standardverdien for temperatur. Hvis terrenget og utgangspunktet er som i figuren, ville vi ikke oppdage potensialet for forbedring, og kanskje slå oss til ro med litt større parameterverdier for både temperatur og trykk, men fortsatt være langt fra det optimale. I praksis kan terrenget være langt mer komplisert, og vi kan risikere å fjerne oss fra toppen pga. at kompasset er for dårlig og/eller vår bruk av det er ukyndig. Dette tema følges opp i et senere avsnitt.

I industriell praksis gjennomføres mange forsøk ved å variere en faktor om gangen. Dette kan skyldes uvitenhet om hvilke farer dette innebærer og hvilke alternativer som finnes. Blant de gode argumentene for forsøksplaner med variasjon av flere faktorer om gangen er:

- mindre ressurser (forsøk, tid, materiale) for å få samme informasjon
- betydningen av hver faktor kan anslås mer presist
- eventuelle samspill (interaksjon) mellom faktorer kan lettere oppdages
- mulighet for mer effektiv produkt- og prosess-optimalisering

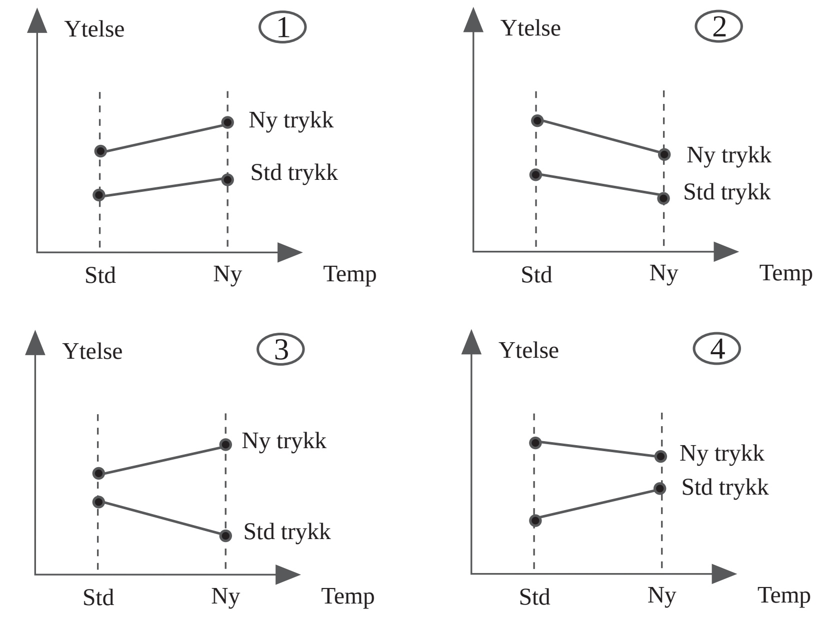
Selv etter å ha blitt gjort oppmerksom på dette, er det vanskelig å få gjennomslag for endret praksis. La oss belyse dette ytterligere med et eksempel som kan sees i sammenheng med Eksempel 2.

Eksempel 3: To faktorer - tre forsøk
Ved produksjon av elektronisk komponent benyttes standardverdier for trykk og temperatur. Det er hevdet at økt verdi av trykk og temperatur gir bedre resultat (ytelse) og at en bør gjøre et forsøk der effekten av en slik endring kan måles. En er enige om en alternativ innstilling av trykk og temperatur og at en vil kjøre forsøk med standardverdiene og ellers variere en faktor av gangen, først temperatur og så trykk. Dermed vil man greie seg med tre forsøkskombinasjoner (tror man). En slik gjennomføring kan skje på tre ulike måter mht. kombinasjonen (Temperatur, Trykk) i rekkefølge:

A Variere med standard som utgangspunkt: (Std, Std), (Ny, Std), (Std, Ny)
B Introdusere ny innstilling etter tur: (Std, Std), (Ny, Std), (Ny, Ny)
C La siste forsøk være bestemt av utfallet av de to første forsøk, alt ettersom hvilket alternativ som var best av de to første: (Std, Std), (Ny, Std), (?, Ny).

For disse tre kan en alternativt endre trykk først.

I Figur 2 er illustrert fire ulike situasjoner mht. målt ytelse (høy verdi er best)



Figur 2 Fire situasjoner

I situasjon 1 vil det neppe gjøre noen forskjell på hvilken av de tre gjennomføringer som ble valgt. Selv om den beste kombinasjonen (Ny, Ny) ikke er forsøkt ved 1A, vil en ut fra de separate effektene kunne slutte at denne trolig er best. Situasjon 2 er tilsvarende der (Std, Ny) er best, men ikke er forsøkt ved 2B, men oppdages likevel. I disse to situasjonene spiller det heller ingen rolle om trykket varieres først.
I situasjon 3 er den beste kombinasjonen (Ny, Ny) ikke forsøkt ved 3A, og siden den observerte negative effekten av økt temperatur er større enn den tilsvarende negative effekt av økt trykk, vil en konkludere med at (Std, Ny) er best og ikke oppdage at (Ny, Ny) er bedre. I situasjon 3B finner vi dette, mens i situasjon 3C gjør vi det ikke.
Diskuter selv hva som kan skje i situasjon 4, og tegn situasjonen som er illustrert i Figur 1 på samme vis som ovenfor (situasjon 5).

I praksis bør forsøk gjennomføres i randomisert rekkefølge. Alternativet til å variere en faktor om gangen med i alt tre faktorkombinasjoner er å kjøre alle fire, dvs. et fullt faktorielt eksperiment. I utgangspunktet er dette 25% dyrere, men det er billig i forhold til informasjonsverdien. I praksis vil en som regel utføre flere forsøk pr. faktorkombinasjon. La oss si at vår bedrift har råd til 12 forsøk. Hva er da best: å kjøre 4 gjentak av de 3 en-om-gangen kombinasjonene, eller 3 gjentak av alle 4 kombinasjoner? Det viser seg nå at et fullt faktorielt eksperiment lønner seg uansett. Vi oppnår med fullt faktorielt eksperiment:

1. Anslag av hver effekt med tilhørende feilmargin som er 18% lavere med et like stort en faktor-om-gangen eksperiment.
2. Påvisning og anslag med feilmargin av eventuelle samspill/interaksjon mellom faktorene.
3. En har informasjon om et større variasjonsområde for faktorene som kan danne grunnlag for bedre planlegging av fortsatte forsøk med nye alternative verdikombinasjoner av faktorene utenfor området.

Merk også at selv om ytelsen for de observerte faktorkombinasjoner blir anslått med noe mindre feilmargin ved en-faktor-om-gangen eksperimentet enn ved full-faktor eksperimentet, så vil forsøk på å ekstrapolere til den faktorkombinasjon som ikke er observert, medføre et anslag som det heftes langt større usikkerhet ved. I gjennomsnitt vil de fire anslagene ha usikkerhet som er 13% større ved en-faktor-om-gangen eksperimentet.

4. Faktorielle eksperimenter

Det er ofte mange faktorer som kan påvirke et resultat, ytelse eller responsen. Dersom hver faktor ønskes observert på flere enn to nivåer, blir det i alt mange mulige faktorkombinasjoner, og med flere observasjoner pr. faktorkombinasjon, vil omfanget av eksperimentet bryte grenser for det som er praktisk og økonomisk forsvarlig å gjennomføre. En må da forsøke å inngå kompromisser mht. eksperimentplan, f.eks. ved å

- begrense seg til faktorer som anses mest vesentlige,
- redusere antall nivåer for hver faktor, f.eks. til 2,
- la være å observere enkelte faktorkombinasjoner,
- begrense seg til (høyst) en observasjon pr. faktorkombinasjon.

Ved alle disse forslag risikerer en å tape noe, og valg av eksperimentplan må skje med åpne øyne, slik at vesentlig informasjon har minst mulig sjanse for å unnslippe. Fjerning av faktorer uten nærmere undersøkelse kan medføre at uventede kvalitetspåvirkende effekter forblir uoppdaget. Ved reduksjon av antall nivåer til to, vil ikke-lineære effekter ikke kunne oppdages (i første omgang). Ved å la være å observere enkelte faktorkombinasjoner risikerer en at enkelte effekter ikke lar seg identifisere, med mindre en velger spesielle eksperimentplaner, og gjør tilleggsantakelser om effektene. Ved å observere hver faktorkombinasjon bare en gang, kan en vanskelig anslå usikkerhet, med mindre en gjør bestemte antakelser om fravær av mer kompliserte effekter (samspill).

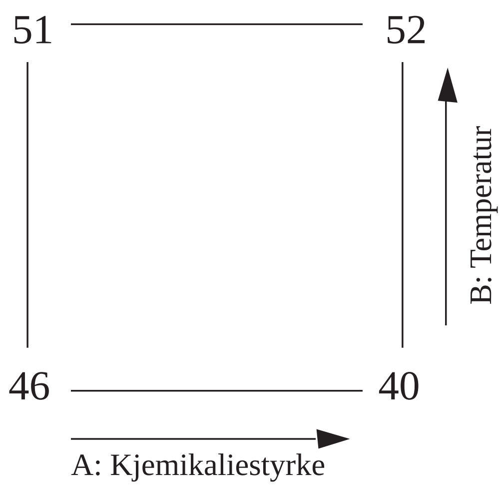
Det har i praksis vist seg at en kan lære mye, selv når en begrenser seg til to nivåer (f.eks. Lav/Høy) for hver faktor. Vi skal nedenfor gi en smakebit på fremgangsmåten i slike situasjoner.

Eksempel 3: Et 2 x 2-forsøk

Ved overflatebehandlingen av et produkt brukes et kjemikalium som vurderes tilsatt i to ulike styrkegrader (Faktor A: Lav/Høy), ved to ulike temperaturer (Faktor B: Lav/Høy). Kvaliteten av finishen måles (høyere tall jo bedre). Det er gjort 2 observasjoner for hver av de 2x2 = 4 faktorkombinasjonene, og resultatet ble

Faktor-kombinasjon	Faktor A	Faktor B	Observasjoner nr. 1	Observasjoner nr. 2	Gjennomsnitt	Standardavvik
1	–	–	45	47	$\overline{X}_1 = 46$	$S_1 = 1.4$
2	+	–	40	40	$\overline{X}_2 = 40$	$S_2 = 0.0$
3	–	+	53	49	$\overline{X}_3 = 51$	$S_3 = 2.8$
4	+	+	50	53	$\overline{X}_4 = 52$	$S_4 = 1.4$

Her er faktorkombinasjonene nummerert fortløpende og er beskrevet for hver faktor med koder -/+ for Lav/Høy. Observasjonene, deres gjennomsnitt og standardavvik er ført ut til høyre for hver faktorkombinasjon. Gjennomsnittsresultatene er i Figur 3 .



Figur 3 Resultat 2 x 2 forsøk

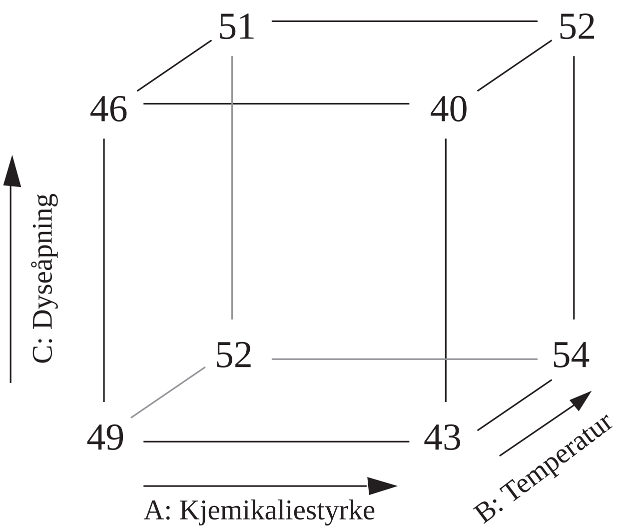
Av resultatene ser vi en klar temperatureffekt i favør av høy temperatur. Når det gjelder kjemikaliestyrke er bildet noe mer komplisert. Vi ser at lav kjemikaliestyrke gir høyest resultat sett i gjennomsnitt over de to temperaturer, men høy kjemikaliestyrke kombinert med høy temperatur gir den beste kombinasjonen (samspilleffekt).

Standardavvikene forteller noe om variasjonen i resultat ved gjentatt måling for samme faktorkombinasjon. Denne tilfeldige variasjon kan skyldes andre faktorer som er utenfor vår kontroll eller rett og slett måleusikkerhet. Gjennomsnittet av disse standardavvikene kan brukes som et mål på den tilfeldige variasjon ved en enkelt måling. Gjennomsnittene og standardavvikene i tabellen gir også mulighet for estimere hovedefekter og samspill med tilhørende mål for usikkerheten av disse (standardfeil), se eget avsnitt. En kan da vurdere om effektene er statistisk signifikante, dvs. ikke bare skyldes tilfeldighet ved de enkelte målinger. Denne usikkerheten vil kunne reduseres med mer enn to målinger pr. faktorkombinasjon, men tid og penger kan gjøre dette umulig. Kanskje er man nødt til å nøye seg med bare en måling pr. faktorkombinasjon. En må da være rimelig trygg på at den tilfeldige variasjon i enkeltmålingene er neglisjerbar i forhold til de effekter vi tar sikte på å avdekke.

Eksempel 4: Et 2 x 2 x 2- forsøk
Anta at vi i forrige eksempel kan tenkes å velge en av to dyseåpninger (Faktor C: Liten/Stor) i tillegg til faktorene A og B. Vi har da følgende skjema med resultater med to observasjoner pr. faktorkombinasjon.

Faktor-kombinasjon	Faktorer A	B	C	Gj.snitt $\overline{X}_i$	St.avvik S_i
1 •	–	–	–	49	1.0
2 o	+	–	–	43	1.7
3 o	–	+	–	52	1.7
4 •	+	+	–	54	2.4
5 o	–	–	+	46	1.4
6 •	+	–	+	40	0.0
7 •	–	+	+	51	2.8
8 o	+	+	+	52	1.4

Gjennomsnittsresultatene er illustrert i Figur 4.



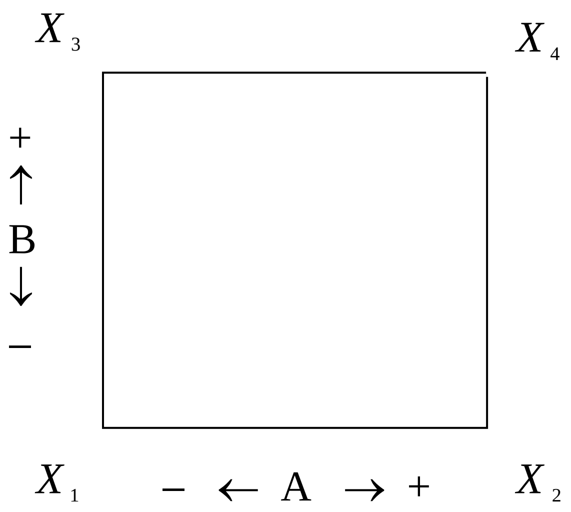
Figur 4 Resultat 2 x 2 x 2 forsøk

Observasjonene i forrige eksempel var for den store dyseåpningen, og gjenfinnes her på toppen av kuben. Vi ser igjen at temperatureffekten er klar i favør av den høyeste temperaturen, uansett kjemikaliestyrke og dyseåpning. Det ser også ut til at den minste dyseåpning kommer noe bedre ut, uansett kjemikaliestyrke og temperatur. Den lave kjemikaliestyrke kommer i gjennomsnitt noe bedre ut enn den høye, men vi ser at høy kjemikaliestyrke kommer best ut i kombinasjon med høy temperatur. Det kan altså se ut som det er et samspill mellom kjemikaliestyrke og temperatur. Som i forrige eksempel kan vi estimere de ulike effekter, og vurdere om de er statistisk signifikante, se senere avsnitt. Dette krever mer enn en observasjon pr. faktorkombinasjon, men med bare en observasjon pr. faktorkombinasjon må en nøye seg med en uformell vurdering av usikkerheten.

En forsøksserie må alltid planlegges slik at utenforliggende elementer i minst mulig grad kan påvirke resultatet, eller kort og godt eksperimentet må være rettferdig.
Et middel til å oppnå dette er å gjennomføre forsøkene i tilfeldig rekkefølge (såkalt randomisering). Noen ganger rekker en ikke å gjennomføre alle forsøk samme dag. Det er det isteden aktuelt å dele forsøksserien i to såkalte halvfraksjoner, som markert i tabellen ovenfor med hhv. hvit og sort sirkel. Dette kalles blokking. Vi kommer tilbake til temaene nedenfor.

5. Beregning av effekter

Betrakt situasjonen 2x2-situasjonen ovenfor med to faktorer A og B med to nivåer – og + for hver faktor. Anta at vi har en observasjon pr. faktorkombinasjon med notasjon som i Figur 5.



Figur 5   Beregning av effekter

Vi kan tale om observasjonenes nivå, hovedeffekten av A, hovedeffekten av B og samspillet mellom A og B. Vi kaller disse effektene hhv. I, A, B og AB. Vi har da

Nivå	I	$\frac{1}{4}\left(X_1 + X_2 + X_3 + X_4\right)$
Hovedeffekt	A	$\frac{1}{2}\left(\left(X_2 - X_1\right) + \left(X_4 - X_3\right)\right)$
Hovedeffekt	B	$\frac{1}{2}\left(\left(X_3 - X_1\right) + \left(X_4 - X_2\right)\right)$
Samspill	AB:	$\frac{1}{2}\left(\left(X_4 - X_3\right) - \left(X_2 - X_1\right)\right)$

Nivået er gjennomsnittet av alle observasjonene. Hovedeffekten A er den gjennomsnittlige forskjell i responsen fra å gå fra lavt til høyt nivå for A for de to nivåene for B. Hovedeffekten B er den gjennomsnittlige forskjell i responsen fra å gå fra lavt til høyt nivå for B for de to nivåene for A. Samspillet AB er forskjellen i endring fra lav til høy for A for hhv. lav og høy for B. Dette ka også skrives som forskjellen i endring fra lav til høy for B for hhv. lav og høy for A.

Beregningen kan illustreres i et enkelt skjema, der en bruker +1/-1 notasjonen for nivåene for A og B, og der samspillet AB representeres ved produktet av de respektive faktorene. Vi ser at de fire beregningene ovenfor fremkommer ved å multiplisere observasjonene med tallet i den aktuelle kolonne og summere, og deretter dividere med 2 (evt.4). I og for seg kan vi godt sløyfe ett-tallene i tabellen. Hvis vi har flere (men like mange) observasjoner pr. faktorkombinasjon, gjør vi beregningen av effektene ut fra gjennomsnittet for hver faktorkombinasjon.

Observert	Effekter			
	I	A	B	AB
X_1	+1	-1	-1	+1
X_2	+1	+1	-1	-1
X_3	+1	-1	+1	-1
X_4	+1	+1	+1	+1

I Eksempel 3 i forrige avsnitt blir effektene utregnet: A: -2.5, B:8.5 og AB: 3.5. Når man har funnet samspilleffekter, er det ikke særlig meningsfylt å bruke hovedeffektene til å karakterisere virkningen av faktorene A og B. Da er Figur 3 langt mer instruktiv.

Spørsmålet er nå om det beregnede samspill er så stort at det er av noensomhelst betydning. Det kan jo hende at det ikke er noe samspill mellom faktorene, dvs. samspillet er null, og at den observerte verdi skyldes ren tilfeldighet. I statistiske termer kalles det å teste om samspillet er statistisk signifikant. Vi trenger da å anslå hvor mye tilfeldighet det er i målingene. Det er mulig når vi har mer enn en observasjon for hver faktorkombinasjon. En kan beregne det gjennomsnittlige standardavvik S for de fire kombinasjonene, som her blir 1.72. Standardfeilen til hver effekt er da S dividert med kvadratroten av antall observasjoner m for hver faktorkombinasjon, her to. Dette gir oss en standardfeil lik 1.21, som gir feilmarginer med 95% garanti lik plass/minus 3.36.[1]   Det ser altså ut til at samspilleffekten er statistisk signifikant.

[1] Dette bygger på at ved normalfordelte observasjoner kan risikoen vurderes ut fra den såkalte t-fordeling med 4(m-1) frihetsgrader, i eksemplet 4. Dette gir en sikkerhetsfaktor på k=2.776, som standardfeilen må multipliseres med.

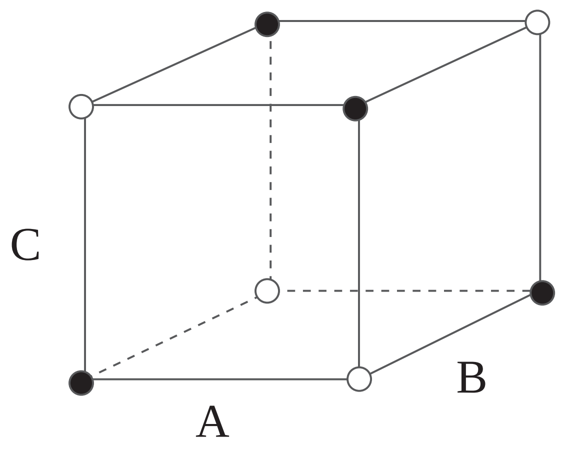
Disse betraktningene bygger på statistisk teori basert på en staokastisk modell for observasjonene, blant annet at disse er normalfodelte med samme varians, og der forventningene kan uttrykker lineært ved parametre som representerer hhv. nivå, faktor A, faktor B og samspillet AB. Denne modellen er beskrevet nærmere i et senere avsnitt.

Ideene ovenfor generaliseres til situasjoner med flere enn to faktorer, og gjør at vi i hvert fall slipper å streve med kompliserte formler for å beregne effekter, mens spørsmålet om statistisk signifikans krever typisk noe mer teoretisk innsikt, se f.eks. Lillestøl (1997). Imidlertid fins en uformell måte å vurdere om effekter er statistisk signifikante på, ved plotting av hver beregnet effekt på såkalt normalfordelingspapir Er effektene rent tilfeldige vil punktene ligge omtrent på linje, mens signifikante effekter vil tydelig avvike fra linjen.

6. Fraksjonelle planer

Med et moderat antall faktorer vil en, selv med bare to nivåer for hver faktor og ingen gjentak, ha mange kombinasjoner, eksempelvis gir 8 faktorer
2^8 = 256 forsøk. I praksis er det ofte bare et fåtall faktorer som påvirker responsen, ofte 2 - 3 (Paretoprinsippet). Det fins forsøksplaner, såkalte *fraksjonerte* planer}, der en ikke observerer alle faktorkombinasjoner, men likevel er i stand til å avdekke viktig informasjon, f.eks. hovedeffekter og evt. alle samspill mellom to faktorer, men ikke mer kompliserte samspill, som likevel forekommer sjelden i praksis.

La oss illustrere hva som skjer i eksemplet ovenfor med tre faktorer. Da kan en bruke en såkalt halv-fraksjonert plan, der bare 4 av de 8 mulige kombinasjonene observeres, enten de kombinasjonene markert med hvitt eller de med sort i Figur 6.



Figur 6 Halv-fraksjon

Vi ser at høy og lav for hver faktor forekommer like ofte i hver av halvfraksjonene. som for øvrig er de samme som ble markert i tabellen i Eksempel 4.

Med bare 4 observasjoner vil hovedeffekten for hver faktor ikke kunne skilles fra samspill-effekten mellom de to andre faktorene ved beregningen. En sier at de to effektene er konfundert ("confounded"). I dette tilfelle er det tre slike konfunderinger: A+BC, B+AC og C+AB. Konfunderte effekter blir også kalt "alias". Hvis en av andre grunner vet at det ikke er samspill, er saken grei. Hvis ikke, må en tolke resultatet i hvert enkelt tilfelle og eventuelt gjøre ekstra forsøk for å avklare hva en eventuell signifikant effekt er for noe.

I situasjoner med fire faktorer fins halvfraksjonert plan der ingen hovedeffekt er konfundert med to-faktor samspill, men bare med tre-faktor samspill, noe en ofte er villig til å anta ikke kan forekomme. På den annen side vil hvert to-faktor samspill være konfundert med sitt komplementære, dvs. at vi kan beregne effektene AB+CD, AC+BD og AD+BC, men ikke skille ut hver enkelt.

I situasjoner med flere enn fire faktorer er det ofte mulig å velge en fraksjonert plan der det kun er konfundering mellom effekter som en på forhånd vet (eller tror) ikke kan være tilstede.

Ved valg av faktoriell plan tar en omsyn til følgende momenter:

1. Fastlegge hvilke faktorer som har interesse.
2. Bestemme antall runs som kan tillates (tid og penger).
3. Identifisere eventuelle andre rammevilkår i valg av plan.
4. Undersøke muligheten for for opplegg med trinnvis læring.
5. Fastlegge hvilke effekter som er av interesse og hvilke som en kan anta er fraværende.

For det siste punktet finnes en systematikk, kalt planens resolusjon. Vi har

III	Ingen hovedeffekt er alias med annen hovedeffekt, men hovedeffekter er alias med to-faktor samspill, og disse er alias med hverandre.
IV	Ingen hovedeffekt er alias med annen hovedeffekt eller med to-faktor samspill, men disse er alias med hverandre.
V	Ingen hovedeffekt eller to-faktorsamspill er alias med annen hovedeffekt eller to-faktor samspill, men to-faktor samspill er alias med tre-faktor samspill. Osv.

For fraksjonelle faktorielle planer med to nivåer for hver faktor, har vi følgende tabell

Faktorielle planer (med resolusjon)									
	Antall faktorer								
Antall runs	2	3	4	5	6	7	8	9	10
4	Full	III							
8		Full	IV	III	III	III			
16			Full	V	IV	IV	IV	III	III
32				Full	VI	IV	IV	IV	IV

Vi ser at med 10 faktorer er det mulig å begrense seg til 16 runs dersom vi (i første omgang) bare er interessert i hovedeffekter, og kan nøye oss med resolusjon III. En spesiell klasse resolusjon III planer med fokus på hovedeffekter kalles **Plackett-Burman design** og er mye anvendt for en første screening av variable. Med en slik design er det for 10 faktorer (ja faktisk 11) nok med 12 runs!

En fraksjonell design kan "foldes" mhp. en valgt faktor. Dette er en måte å redusere konfundering på. Eksempelvis dersom en folder mhp. A, vil A og alle tofaktor-sampill som omfatter A kunne skilles fra alle andre faktorer og deres tofaktor-samspill. Hvis en folder mhp. alle faktorene, vil en kunne skille alle disse såvel som alle tofaktor-samspill. Folding av en plan med resolusjon III gir en plan med resulusjon IV. Denne ideen brukes ved når en vil planer som har visse ønskede egenskaper.

Ved en fraksjonell plan har en mulighet for å velge hvilken fraksjon en vil bruke. En halvfraksjon gir to mulige valg, mens en kvartfraksjon gir fire mulige valg. En av fraksjonene er pr. konvensjon kalt den prinsipale fraksjon, og programvare vil typisk presentere denne, med mindre man selv velger en annen. Det kan være situasjoner der det er upraktisk å gjennom en bestemt fraksjon, f.eks. en med alle faktorene på høyt nivå. En kan da bruke design-generatorene (tallene -1 og +1 som representerte nivåene for hver faktor) til å spesifisere den fraksjon en ønsker. Den prinsipale fraksjonen vil alltid omfatte en run med alle faktorene på høyt nivå (+1). Eksempel: Med tre faktorer A, B og C har en to halvfraksjoner, en der C=AB og en der C= - BC. Den første er den prinsipale fraksjon Med fem faktorer A, B, C, D og E har en fire kvartfraksjoner, svarende til de fire mulige valg av fortegn i D=+/- AB, E=+/-AC. Her representerer +-valget den prinsipale fraksjon. Hvis en vil unngå en plan med en run med alle faktorer på høyt nivå, velge en av de tre andre. Valgene listes vanligvis opp i såkalt standard (Yates) rekkfølge.

7. Randomisering og blokking

Det er to viktige prinsipper i forsøksplanlegging som tar sikte på å unngå at faktorer som ikke lar seg styre får påvirke resultatet. Disse er:

1. Randomisering
2. Blokking

Eksperimenter innebærer måling av prosesser eller produkter. For at målinger skal ha verdi som grunnlag for forbedringer, bør de prosesser som måles være i ”statistisk kontroll”, dvs. at dens variasjon ikke i vesentlig grad påvirkes av spesielle utenforliggende årsaker. Likeledes må selve målesituasjonen, dvs. måleutstyret, omgivelsene og den som måler, være i statistisk kontroll.

I praksis kan en sjelden være sikker på at situasjonen er i statistisk kontroll, selv om vi etter beste evne prøver å tilrettelegge for det. Utenforliggende årsaker som vi ikke tenkte på kan ofte påvirke målingene. Dersom målingene foregår over en viss tid, kan det hende at det er umulig å holde stabile forhold i hele det tidsrom som kreves for å gjennomføre forsøkene.

Randomisering vil si at forsøkene gjennomføres i tilfeldig rekkefølge. Hvis det f.eks. skulle være maskinslitasje og økende unøyaktighet før ny kalibrering neste morgen, vil randomisering sikre upartiskhet. Poenget er at en dermed unngår skjevheter av mulige årsaker som en ikke har oversikt over. Det fins en annen måte å sikre upartiskhet på, nemlig blokking.

Anta at vi i 2x2x2-forsøket vi studerte ovenfor har råd til 8 forsøk, dvs. en pr. faktorkombinasjon, men bare rekker 4 forsøk på en dag. De ikke-styrbare faktorene kan da være menneske (f.eks. ulik bemanning eller dagsform), maskin (f.eks. rekalibrering hver morgen), miljø (f.eks klimaforhold) osv. Hvis f.eks. alle forsøk med samme nivå for en av faktorene kommer på samme dag, kan eventuell forskjellig resultat for høy/lav skyldes noe annet enn faktoren selv. Et upartisk opplegg vil være å splitte de 8 forsøk i to blokker, nemlig de to halvfraksjonene markert med hhv. hvitt og sort i Figur 6. Med en slik balansert plan vil en systematisk forskjell mellom dagene oppveie hverandre ved beregningen av både hovedeffekter og samspill. Mange mener at en ikke bør foreta mer enn 8 forsøk i en faktoriell design uten å foreta en blokking.

Den prinsipielle ideen ved blokking er å fjerne heterogenitet som vi ikke er interessert i, det være seg mellom tidsperioder, ulike maskiner, partier eller skift. Med en velvalgt forsøksplan vil en kunne balansere eventuelle forskjeller mot hverandre, slik at de for vårt formål er fjernet. I praksis vil derfor effekter kunne oppdager lettere og bestemmes med større nøyaktighet. Uten blokking risikerer en at viktige effekter forblir uoppdaget, eller at det kreves mange flere forsøk for å oppdage dem. Det kan se ut som vi har oppnådd noe for ingenting. Det er ikke tilfelle. Blokking vil typisk innebære at høyere ordens effekter blir konfundert med den heterogenitet vi ønsker å fjerne. I 2x2x2 eksemplet ovenfor vil trefaktor-samspillet bli konfundrert med dag til dag effekten.

Det kan selvsagt hende at uventede forhold kan gjøre seg gjeldende innen samme blokk i et forsøksopplegg. Det kan vi løse ved å randomisere, her ved å utføre forsøkene innen hver blokk i tilfeldig rekkefølge. En god leveregel er:

Blokk det som er mulig, og randomiser det som ikke kan blokkes.

8. Responsflateplaner

Responsflatemetoder brukes ofte etter at de vitale få faktorer er identifisert, og en søker etter nivåer på disse som gir en ønsket forventet respons. Det kan være

- finne faktorkombinasjoner som gir størst forventet respons,
- finne faktorkombinasjoner som oppfyller gitte spesifikasjoner angående prosess/produkt,
- identifisere faktorkombinasjoner som gir en forbedring av nåværende situasjon,
- modellere relasjoner mellom faktorene og responsen.

Vi er interessert i hvordan en responsvariabel avhenger av inputvariable $X_1, X_2, ..., X_r$

$$Y = g(X_1, X_2, ..., X_r) + \varepsilon$$

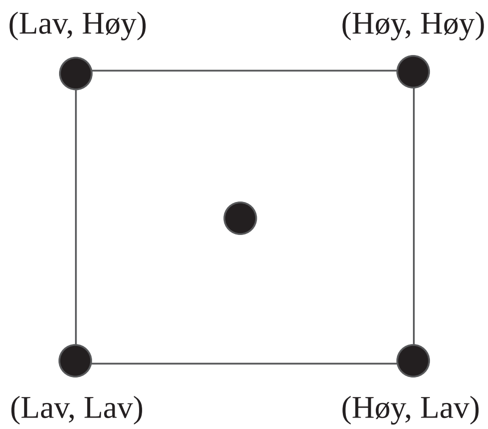
der g er responsfunksjonen, som grafisk kan oppfattes som en flate i et flerdimensjonalt rom. Denne uttrykker forventet respons for gitt input og ε er tilfeldig avvik med forventning null.

I praksis er responsfunksjonen helt eller delvis ukjent og må anslås ut fra forsøk. Vi har tidligere fastslått at dette ikke kan skje ved å variere en inputvariabel av ganger. Det foreligger da to muligheter:

1. Gjøre forsøk for tilstrekkelig mange kombinajoner av inputvariable for å få et godt bilde av responsflaten over hele det aktuelle mulighetsområdet for inputvariablene, for deretter å bestemme en kombinasjon som oppfyller ønskemålet.
2. Bestemme en tilnærming til responsflaten, basert på et mindre antall kombinasjoner av inputvariable innen en (lovende) del av mulighetsområdet. Nye forsøk utføres i den retning som blir pekt ut som lovende, f.eks. retningen med brattest stigning. Dette kan utføres i flere trinn, med lineær tilnærming i (de) første trinn og kvadratisk tilnærming ved (de) siste trinn. Et råd er å bytte til kvadratisk tilnærming når den lineære ikke lenger peker ut en klar retning for forbedring.

Erfaring har vist at en vanligvis vil finne optimum mer presist med færre forsøk totalt ved metode 2 enn ved metode 1. La oss se nærmere på noen av prinsippene for responsflateplaner.

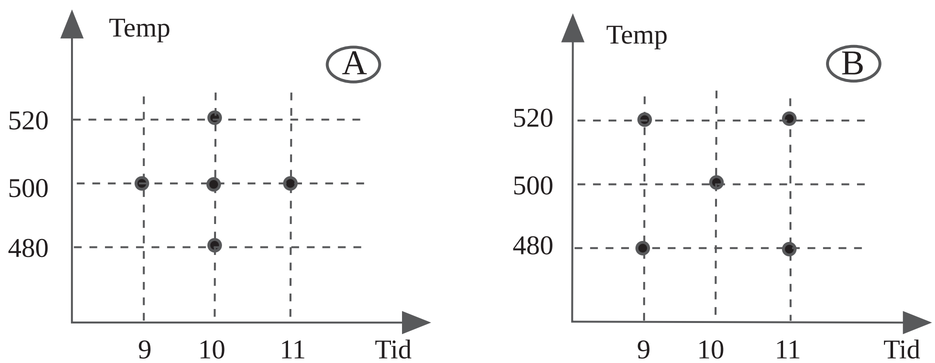
Vi så i Eksempel 2 at vi kunne ønske oss planer som best mulig kartla responsflaten, og spesielt finne den retning som er brattest. En mulighet er å inkludere en faktorkombinasjon svarende til et senterpunkt, som illustrert i Figur 7. Her svarer f. eks. Lav til Std ovenfor.



Figur 7 Forsøksplan med senterpunkt

Denne fremgangsmåten kan gi et inntrykk av krumningen av responsflaten og peke ut den retning en bør gå for eventuelt å finne enda bedre kombinasjoner av faktorene. En må da tenke seg responsen som en flate i rommet over figuren, der en ønsker å finne det høyeste punktet på flaten. Med råd til få ekstra forsøk kan det være aktuelt å bruke alle disse i senterpunktet for å anslå den naturlige variabilitet. Med råd til flere ekstra forsøk, kan det være aktuelt å fordele disse jevnere, f. eks. med 15 forsøk i alt fordele disse med 3 for hver av de 5 faktorkombinasjonene.

Eksempel 6: To faktorer - 6 forsøk?
En ønsker å kartlegge effekten av de to faktorene tid og temperatur på produksjon av en maskindel. En tenker seg muligheten av en ikke-lineær sammenheng, og trenger da minst tre tider og tre temperaturer. En vurderer først å variere en faktor av gangen, dvs. for en gitt tid gjøre forsøk med tre ulike temperaturer og deretter for en gitt temperatur gjøre forsøk med tre ulike tider. Dette er plan A illustrert i Figur 8A (med dublert senterpunkt). Et alternativ er plan B, en faktoriell plan med senterpunkt, som illustrert i Figur 8B. Dubleres senterpunktet, har vi like mange (seks) forsøk i de to planene.

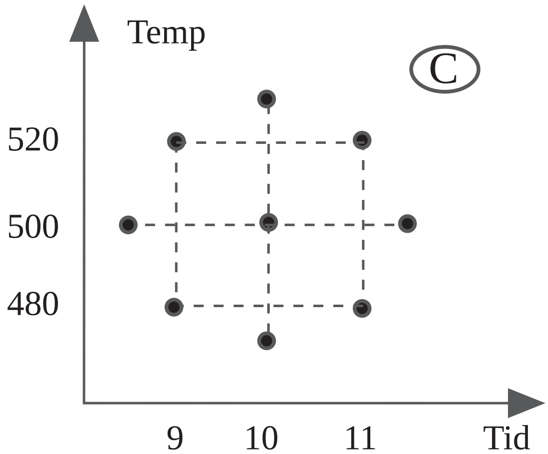


Figur 8 To situasjoner

Begge planene er i stand til å avdekke en mulig ikke-lineær sammenheng mellom respons og hver av faktorene, men det er bare plan A som setter oss i stand til å anslå størrelsen på krumningen i vertikal og horisontal retning, riktignok bare for de mellomste nivå for tid og temperatur. Begge planene med sine to senterpunkt gir anledning til å anslå den naturlige variabilitet. Plan B har imidlertid sine fordeler fremfor plan A:

- Effekten av hver faktor kan anslås på grunnlag av alle 6 observasjonene - og ikke bare 4 - og blir derfor mer pålitelig.
- Samspill/interaksjon kan avdekkes, noe som ikke er mulig med plan A.
- En får informasjon om et større variasjonsområde for faktorene som kan gi grunnlag for bedre planlegging av ytterligere forsøk med nye verdikombinasjoner utenfor området.

Hvis en også ønsker å anslå krumningen i responsen har en mulighet til å utvide plan B til en såkalt central-composite plan (plan C) som illustrert i Figur 9.



Figur 9 Central Composite plan

En CC-plan består av tre ulike typer design-punkter: Kubepunkter, aksialpunkter og senterpunkter. I plan C i figuren er det 4 kubepunkter, 4 aksialpunkter og ett senterpunkt, i alt 9 faktorkombinasjoner. I CC-planer blir typisk senterpunkter gjentatt.

Sammenlignet med plan A gir plan C mulighet for å anslå krumningen i en hvilket som helst retning, sett fra et vilkårlig punkt innenfor variasjonsområdet. Det er nødvendig hvis en tar sikte på å finne optimale verdier av parametrene.
Fordelen med plan A er selvsagt at denne er basert på bare 6 forsøk, mens plan C krever 9. En er imidlertid blitt rikelig belønnet ved en slik utvidelse. Dersom en har ressurser til å gjennomføre et større antall forsøk, f. eks. 18, vil det ikke være særlig lurt å bruke plan A med 3 gjentak. En burde i stedet bruke plan C med 2 gjentak. Med ressurser til 12 forsøk vil en typisk foretrekke plan C med 4 gjentak av senterpunktet fremfor plan A med 2 gjentak av alle.

Ideen med central-composite planer kan selvsagt utvides til mer enn to faktorer og kombineres med ideen om fraksjonerte planer og blokking. Slike planer er først og fremst aktuelle ved flere serier med forsøk, der resultatet av en serie er bestemmende for designen av den etterfølgende, for om mulig å finne nye lovende retninger å gå i optimaliseringsprosessen.

I situasjoner der en kjenner begrensningen av det aktuelle parameterområdet (f.eks. tekniske begrensninger), vil en typisk foretrekke å kjøre kun en serie med forsøk. Et aktuelt alternativ er da såkalte **Box-Behnken planer**. Disse tillater beregning av stigningstall i ulike retninger. Med det samme antall faktorer krever en slik plan færre runs enn en tilsvarende central-composite plan. I situasjonen med tre faktorer har en BB-design 15 runs. På kuben (se Figur 4) er disse 12 punkter på midten av alle sidekanter og 3 senterpunkter, dvs. ingen punkter i hjørnene, mens en tilsvaende CC-design har 20 runs. En kan derfor merke seg at en BB-plan aldri har runs der alle faktorer er satt til sitt ytternivå samtidig. Med en CC-plan kan det hende at aksialpunkter ligger utenfor det aktuelle og trygge parameterområdet.

En sammenligning av antall runs i en CC-plan og en tilsvarende BB-plan for gitt antall faktorer er gitt i følgende tabell:

Antall runs					
Plan	Antall faktorer				
	2	3	4	5	6
Central Composite	13	20	31	52	90
Box-Behnken		15	27	46	54

La oss avslutte dette avsnittet med å gi noen praktiske detaljer ved metoden med ”brattest stigning” for to faktorer (X_1, X_2) observert for de 5 nivåkombinasjonene i et kvadrat med senterpunkt og fire hjørner. En analyse som ikke avhenger av valgt måleskala for faktorene må baseres på de tilhørende kodetallene.

$$(K_1, K_2): \ (0, 0) \ \ (-1, -1) \ \ (+1, -1) \ \ (-1, +1) \ \ (+1, +1)$$

Anta at vi ut fra de tilgjengelige data anslår den lineære sammenhengen

$$Y = b_0 + b_1 K_1 + b_2 K_2$$

For ulike ”høyder” Y bestemmer dette nivålinjer (”koter”) i et kart i (K_1, K_2)-planet som vist i Figur 10 for b_1 og b_2 positive, slik at ”oppoverbakken” er mot nordøst i figuren.

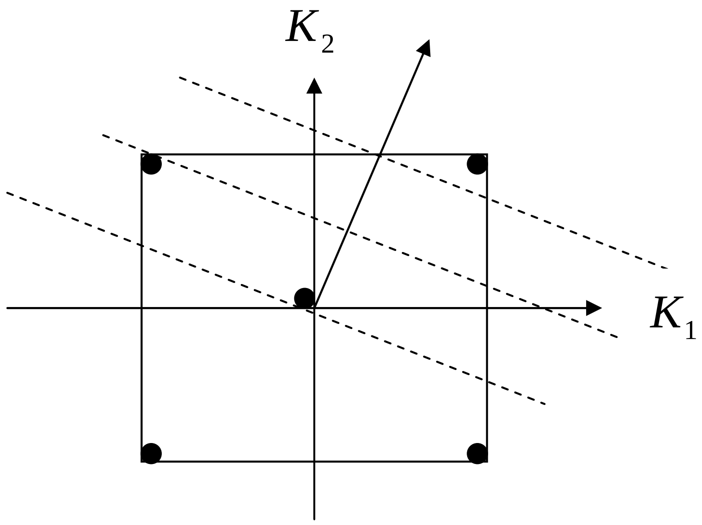


Figure 10   Metoden med brattest stigning

Retningen for brattest stigning er vinkelrett på disse linjene, som er linjer med vinkelkoeffisient b_2 / b_1. Med (0, 0) som utgangspunkt, må vi flytte oss langs pilen i figuren. Som nytt senterpunkt utefor kvadratet velger vi (k_1, k_2) der k_1 i tallverdi er større enn 1 og $k_2 = b_2 / b_1 k_1$. Uttrykt på den opprinnelige måleskala er dette å gå fra (x_1^0, x_2^0) til (x_1, x_2) der

$$x_1 = x_1^0 + k_1(x_1^+ - x_1^0) \text{ og } x_2 = x_2^0 + k_2(x_2^+ - x_2^0)$$

og toppskriften + svarer til + nivået for variablene.

9. Robust parameter design

Robust parameter design er en teknikk som samtidig betrakter nivå og variasjon. Typisk problemstilling er å finne parameterverdier slik at output er nær ønsket verdi med minst mulig variasjon. En vil derfor kunne akseptere et mindre systematisk avvik fra nominell verdi, dersom en med dette oppnår en mindre risiko for sporadisk store avvik. "Signal to noise ratio", forkortet S/N, er et forsøk på å gi et tallmessig mål som kombinerer nivå ("signal") og variasjon ("noise"), og er foreslått som målfunksjon ved optimalisering. S/N er ikke et entydig begrep, men defineres ulikt alt ettersom målet er å maksimere/minimere responsen eller minimere avstanden til nominell verdi. Vi vil ikke gi de formelle definisjoner her.

Et problem ved bruk av S/N-mål for optimalisering basert på forsøk, er at effekter som påvirker nivå og variasjon kan være konfundert, slik at en kan bli i villrede om hva som har gitt optimum. Det er derfor neppe lurt å ukritisk betrakte S/N-ratio alene. Slike mål har vunnet popularitet delvis gjennom japaneren Taguchi's anbefalinger. Andre mener at S/N er uten særlig verdi, og anbefaler å bruke mål for nivå og variasjon direkte, men se dem i sammenheng.

Vi vil her ta opp to problemstillinger knyttet til robust kvalitet.

En kan ønske seg produkter som er robuste overfor varierende bruksforhold og bruksmåte, dvs at kunden er tilfredsstilt også når produktet brukes i omgivelser som ikke er ideelle og/eller om bruksanvisningen ikke er fulgt til punkt og prikke. Det er foreslått eksperimentplaner, bl.a. av japaneren Taguchi, som ser såkalte *designfaktorer* i sammenheng med *bruksfaktore*r, med sikte på å velge designfaktorene slik at produktet virker tilfredsstillende under variasjon i bruksfaktorene. La oss illustrere dette med et eksempel:

Eksempel 7: Kakemiks
Vi vil markedsføre en ny kakemiks, der anbefalt steketemperatur (T) og steketid (t) skal stå på boksen. Det er ønskelig at mindre avvik i disse to bruksfaktorene ikke har avgjørende innflytelse på kundenes oppfatning av den ferdige kake. I denne sammenheng studeres tre designfaktorer kvantum mel (M), sukker (S) og eggpulver (E). Det tas utgangspunkt i en oppskrift som kundene anser som en god kake. En vil lage et eksperiment der en varierer alle 5 faktorer for å finne et alternativ som blir vurdert nesten like godt ved anbefalte steketider og steketemperaturer, og som helst ikke gir vesentlig dårligere resultat for mindre avvik fra anbefalingen.

Eksperimentet ble utført ved å tilby et smakspanel ulike kaker. I tillegg til standardoppskrift og bearbeiding (i tabellen markert med 0), kommer alle kombinasjoner der de ulike faktorene settes et hakk opp (+) eller et hakk ned (-). Dette gir i alt 45 ulike kaker. Hver smaker gir karakter på en skala fra 1 til 7, hvoretter gjennomsnittskarakterer beregnes. Resultatet ble:

	Design variable			Bruksvariable				
run				T 0	-	+	-	+
nr.	M	S	E	t 0	-	-	+	+
0	0	0	0	6.7	3.4	5.4	4.1	3.8
1	-	-	-	3.1	1.1	5.7	6.4	1.3
2	+	-	-	3.2	3.8	4.9	4.3	2.1
3	-	+	-	5.3	3.7	5.1	6.7	2.9
4	+	+	-	4.1	4.5	6.4	5.8	5.2
5	-	-	+	5.9	4.2	6.8	6,5	3.5
6	+	-	+	6.4	5.0	6.0	5.9	5.7
7	-	+	+	3.0	3.1	6.3	6.4	3.0
8	+	+	+	4.5	3.4	5.4	4.1	3.8

Vi ser at design nr.6 med M=+, S=- og E=+, noe mer mel og egg, og litt mindre sukker enn standardoppskriften gir vel så god kake ved samme anbefalte steketemperatur og steketid, og ikke mye dårligere kake dersom en fraviker anbefalingen. Ved flere av de andre kombinasjonene kan det se ut som at fravik i en eller flere retninger kan gi et vesentlig dårligere produkt.

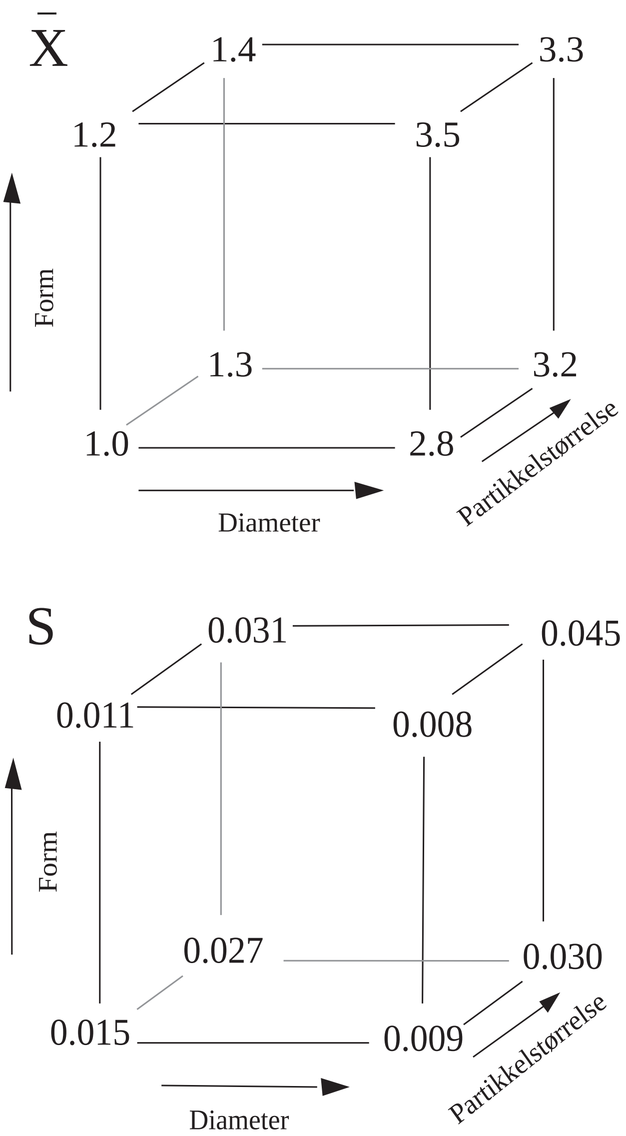
I praksis ønsker en nok å teste flere enn 2-3 design- og bruksfaktorer.
Det er da ikke praktisk å gjennomføre et eksperiment der alle kombinasjoner blir forsøkt, ei heller at en og samme smaker må forholde seg til og gradere så mange ulike kaker. En trenger derfor fraksjonerte eksperimentplaner, der en prøver ut et mindre antall kombinasjoner av faktorene, men der en i hovedsak er istand til å finne robuste kombinasjoner av enkeltfaktorene. Dette gir både en egnet blanding av designfaktorene i kakemiksen, samt en god anbefaling på boksen.

Eksperimentering kan brukes til å finne hvilke faktorer som kan bidra til å få en prosessindikator nærmere et mål og hvilke som kan bidra til at variabilitet reduseres. La oss ta et eksempel [2]

Eksempel 8: Vaskepulver
Ved fylling av vaskepulverbokser studeres tre faktorer: traktens form og diameter, samt pulverets partikkelstørrelse. Det er viktig å kunne holde fyllehastigheten på et valgt nivå med liten variasjon. Anta at hver faktor blir studert på to nivåer, og at resultatet ble som illustrert i Figur 11.

---

2 Eksemplet er hentet fra Box, Bisgaard & Fung (1988)}:



Figur 11 Minimere variasjon omkring mål

Vi ser at variasjonen fyllehastigheten (målt med standardavvik) i hovedsak bare er avhengig av partikkelstørrelsen, mens nivået ser ut til å kunne endres med diameteren av trakten, uten at slike endringer betyr noe for variasjonen. Hvis en f.eks. ønsker en fyllehastighet nærmest 3.0 med minst mulig standardavvik, bør en velge den minste partikkelstørrelsen av de to, for deretter å bruke traktdiameteren til å justere fyllehastigheten mot målet.

Dette er selvsagt et svært forenklet eksempel, der to av tre faktorer kunne skilles ut med en klar rolle. Situasjonen er ikke alltid så enkel. Vi har ofte mange flere faktorer, og det slett ikke sikkert at disse naturlig kan skilles i tre grupper: de som i hovedsak er knyttet til prosessnivå, som er knyttet til prosessvariasjon og de som ikke har særlig betydning overhodet. Det kan tenkes faktorer som er både nivå- og variasjonspåvirkende, og det krever en grundigere forståelse dersom slike skal brukes i målstyring av prosessen.

10. Oppsummering

Formål

A. Hva er et eksperiment?
- Målrettet generering av data
- under veldefinerte, helst optimale, betingelser
- med sikte på å avklare prioriterte spørsmål

B. Hva slags spørsmål kan eksperimenter avklare?
- hvor kommer variabilitet fra?
- hvorfor inntreffer feil og avvik?
- hva er optimal kombinasjon av faktorer?
- er våre hypoteser holdbare?
- kan vi få nye ideer?

C. Hvor passer det å gjøre eksperimenter?
- Overalt!

D. Hva er kostnadene og farene ved eksperimenter?
- materialer og utstyr
- arbeidskraft til planlegging og utføring
- dårlig design og gjennomføring
- feiltolkning av resultater

E. Hva kjennetegner et godt eksperiment?
- presist formål og grundig planlagt (design)
- godt gjennomørt (alle prosesser i kontroll)
- gir objektive tolkninger og prediksjoner
- god analysemetode med sjekk av forusetninger
- grundig dokumentert og godt presentert

Ved omhyggelig forsøksplanlegging vil en kunne unngå mange skjær i sjøen, som i verste fall kan medføre at arbeidet er bortkastet. Dersom prosjektet har lav prioritet innen organisasjonen, kan det være lurt å gjennomføre det sekvensielt. I en situasjon der prosjektet stoppes før det er ferdig, midlertidig eller for godt, vil de data som allerede er innhentet kunne brukes til noe. Dersom muligheten byr seg senere, kan en så fortsette der en slapp. Dette krever imidlertid omtanke ved planleggingen.

- Klargjør de aktuelle problemstilling, og formuler en målsetting for forsøkene, f.eks. identifikasjon av betydningsfulle faktorer eller optinalisering.
- Utvikl en forsøksplan i samsvar med målsettingen, som kan gi meningsfull og brukende informasjon innenfor de fysiske og økonomiske rammes som eksisterer.
- Forsikre deg om at prosessen og målesystemet er i statistisk kontroll, og at variabiliteten av dette ikke gjør det umulig å oppdage effekter av betydning.

Hva trengs av kunnskaper?

Eksperimenter vil for mange være nødvendig for å lykkes i internasjonal konkurranse, og bør da være en integrert del av kvalitetsledelse. En del forutsetninger må være oppfylt for å lykkes for å bruke eksperimenter:

- Ledelsen må forstå hva eksperimenter er, og når og hvordan det er et tjenlig redskap
- Ledelsen må treffe tiltak for at ressurser og ekspertise er tilgjengelig

A. Toppledelsens rolle:

1. Forme den overordnede kvalitetspolitikk
2. Forstå eksperimenters rolle og hvordan disse integreres med øvrige kvalitetsaktiviteter
3. Stimuler til bruk i organisasjonen
4. Overvåk bruken og nytten av eksperimenter

B. Mellomledelsens rolle:

1. Hovedmotivatorer for bruk i organisasjonen
2. Forstå grunnleggende ideer ved eksperiment og elementene i planlegging, gjennomføring og analyse
3. Forstå hvordan eksperimenter kan integreres med øvrige kvalitetsaktiviteter
4. Ha autoritet til å implementere resultater

C. Arbeidsledere:

1. Ha teknisk kompetanse i planlegging, gjennomføring og analyse
2. Motivere medarbeidere til å eksperimentere, gi tid til opplæring
3. Evne til å vurdere resultater og stille kritiske spørsmål
4. Vurdere når profesjonell ekspertise trengs

Hvordan kommer en i gang?

Ved egenutvikling - interopplæring - konsulenter - om nødvendig ansett en statistiker. En statistiker vil kunne drive opplæring, tilpasse verktøy til organisasjonens forutsetninger og arbeide med prioriterte oppgaver, samt fungere som rådgiver for ulike prosjektgrupper i organisasjonen.

11. Statistisk teori

Vi vil i dette avsnitt ta for noen av de statistiske modellene som ligger til grunn for dataanalyse i avsnittene ovenfor. Disse blir formulert i generelle vendinger

La oss først se på de statistiske modeller som ligger til grunn for analyse av forskjellen mellom dekktypene for de to planene i det innledende avsnitt.

To-utvalgsmodellen (for plan 1)

Gitt $n_1 + n_2$ observasjoner i to grupper

Gruppe 1: $X_{11}, X_{12}, \ldots, X_{1n_1}$
Gruppe 2: $X_{21}, X_{22}, \ldots, X_{2n_2}$

Observasjonene antas uavhengige normalfordelte der

$$\begin{aligned} \mathrm{E}X_{1i} &= \mu_1 \qquad \mathrm{var}\, X_{1i} = \sigma_1^2 \\ \mathrm{E}X_{2i} &= \mu_2 \qquad \mathrm{var}\, X_{2i} = \sigma_2^2 \end{aligned}$$

Vi er interessert i forskjellen i forventninger

$$\delta = \mu_2 - \mu_1$$

Denne forskjellen i forventninger anslår vi med forskjellen i de observerte gjennomsnitt

$$\hat{\delta} = \overline{X}_2 - \overline{X}_1$$

Anslag med feilmarginer etter standardavviksmetoden er gitt ved

$$\hat{\delta} \pm k \cdot \mathrm{SE}(\hat{\delta})$$

der $\mathrm{SE}(\hat{\delta})$ er den såkalte standardfeilen og *k* er en sikkerhetsfaktor, der

k = 1 svarer til ca. 68% sikkerhet og
k = 2 svarer til ca. 95% sikkerhet.

$\mathrm{SE}(\hat{\delta})$ blir vanligvis beregnet ved

$$\mathrm{SE}(\hat{\delta}) = S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

der

$$S^2 = \frac{1}{n_1+n_2-2}\left(\sum_{i=1}^{n_1}\left(X_{1i} - \overline{X_1}\right)^2 + \sum_{j=1}^{n_2}\left(X_{2j} - \overline{X_2}\right)^2\right) = \frac{1}{n_1+n_2-2}\left((n_1-1)S_1^2 + (n_2-1)S_2^2\right)$$

Dette er et anslag for variansen under forutsetning av at denne er lik for de to gruppene, dvs. $\sigma_1^2 = \sigma_2^2$. Dersom variansene er ulike, er formelen noe mer komplisert. Benytter vi statistisk programvare (som Minitab), får vi skrevet ut anslag og feilmarginer automatisk, uten at vi behøver å bry oss om formlene.

Vi ønsker ofte å teste hypotesen at forventningen er den samme i de to gruppene mot at den er større i gruppe 2, dvs.

$$H_0 : \delta = 0 \qquad \text{mot} \qquad H_A : \delta > 0$$

Den såkalte *t*-observatoren gitt ved

$$T = \frac{\hat{\delta} - 0}{\mathrm{SE}(\hat{\delta})}$$

Hypotesen H_0 forkastes dersom $T \geq k$, der *k* er en kritisk verdi bestemt ved den såkalte *t*-fordelingen med $n_1 + n_2 - 2$ frihetsgrader. Kritiske verdier kan finnes fra *t*-tabeller.

I programvare rapporteres ofte den såkalte *P*-verdi gitt ved

$$P = P_{H_0}(T \geq t_{\text{obs}})$$

der t_{obs} er beregnet verdi av T ut fra observasjonene. P er sannsynligheten for et resultat minst like ekstremt som det obseverte i favør av den alternative hypotesen H_A, beregnet under forutsetning av at H_0 er riktig. Dersom *P* er liten er det en indikasjon på at H_0 ikke kan være riktig og H_A påstås. Hvis vi er villig til å løpe en risiko lik α (f.eks. $\alpha = 0{,}05$) for å påstå H_A selv om H_0 er riktig, vil vi påstå H_A dersom $P \leq \alpha$, ellers ikke.

Modellen for parrede observasjoner (plan 2)

Gitt *n* observasjonspar

$$(X_{11}, X_{21}) \quad (X_{12}, X_{22}) \ldots (X_{1n}, X_{2n})$$

Betrakt $Z_j = X_{2j} - X_{1j} \qquad j = 1, 2, \ldots, n$. $Z_1, Z_2, \ldots, Z_n$ antas uavhengige normalfordelte

$$\mathrm{E}Z_j = \delta \qquad \mathrm{var}\, Z_j = \sigma^2$$

Merknad Modellen tillater ulike forventninger for hvert observasjonspar, bare forventet endring δ er konstant, dvs.

$$\mathrm{E}X_{1j} = \alpha_j \qquad \mathrm{E}X_{2j} = \alpha_j + \delta$$

Modellen kalles derfor ofte for *konstant-effekt* modellen.

Det er effekten δ som er i fokus, og denne kan anslås med gjennomsnittet av de observerte forskjeller innen hvert par, dvs.

$$\hat{\delta} = \overline{Z} = \overline{X_2} - \overline{X_1}$$

Anslag med feilmarginer etter standardavviksmetoden er gitt ved

$$\hat{\delta} \pm k \cdot \mathrm{SE}(\hat{\delta})$$

der standardfeilen SE er gitt ved

$$\mathrm{SE}(\hat{\delta}) = S/\sqrt{n}$$

der

$$S^2 = S_Z^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Z_i - \overline{Z}\right)^2$$

Vi ønsker å teste hypotesen om ingen effekt mot alternativet at positiv effekt for komponent nr.2 i forhold til komponent nr.1, dvs.

$$H_0 : \delta = 0 \qquad \text{mot} \qquad H_A : \delta > 0$$

t-observatoren er igjen gitt ved

$$T = \frac{\hat{\delta} - 0}{\mathrm{SE}(\hat{\delta})}$$

Hypotesen H_0 forkastes dersom $T \geq k$, der *k* er en kritisk verdi bestemt ved *t*-fordelingen med *n* – 1 frihetsgrader. Programvare gir typisk *P*-verdien gitt ved

$$P = P_{H_0}(T \geq t_{\mathrm{obs}})$$

som har den samme fortolkning som ovenfor.

Modellene ovenfor er spesialtilfeller av såkalte ANOVA-modeller ("Analysis of Variance").

ANOVA-modeller søker å forklare observasjonene lineært ved parametre som representerer de ulike faktorer som inngår. Disse parametrene kan estimeres ut fra data og interessante hypoteser testes. Spesielt for ANOVA er også et "regnskap", der den totale variasjon i data blir splittet opp iht. hver faktor som inngår, pluss den uforklarte variasjon. For å representere variasjon brukes ulike kvadratsummer (SS = "Sum of Squares"). Til hver kvadratsum hører et såkalt frihetsgradstall *f*, en slags skaleringsfaktor som forteller hvor stor en forventer at den tilhørende SS er, dersom ren tilfeldighet råder. MS = SS/*f* kalles "Mean Sum of Squares".

Dette vil bli klargjort gjennom et par viktige modeller.

ANOVA-modellen med en faktor

Hvis vi istedenfor to uavhengige grupper har *I* > 2 grupper (ofte kalt nivåer), er det tale om ANOVA med en faktor A. La oss si at vi har n_i observasjoner i gruppe nr. *i*, slik at $n = n_1 + n_2 + \ldots + n_I$ er totalt antall observasjoner. La videre X_{ik} være observasjon nr. *k* i gruppe nr. *i*. ANOVA – modellen antar at de *n* observasjonene er uavhengige normalfordelte med forventning

$$EX_{ik} = \lambda + \lambda_i^A \qquad \text{der} \qquad \sum \lambda_i^A = 0$$

og varians $\operatorname{var} X_{ik} = \sigma_i^2$, der $\sigma_i^2 = \sigma^2$ for alle *i*.

Her kan λ oppfattes som en referanseforventning som korrigeres med effektparameteren λ_i^A når observasjonen er for nivå nr. *i*. Man kan enten bruke en av gruppeene som referanse med effektparameter null og korrigere ut fra denne, eller la referansen λ være den gjennomsnittlige forventningen, slik at summen av λ_i^A over alle gruppene er null.

I tilfellet med to grupper med forventninger hhv. μ_1 og μ_2 kan vi skrive enten

$$\lambda = \mu_1 \qquad \lambda_1^A = 0 \qquad \lambda_2^A = \mu_2 - \mu_1 = \delta \qquad \text{eller}$$

$$\lambda = \frac{1}{2}(\mu_1 + \mu_2) \qquad \lambda_1^A = \frac{1}{2}(\mu_1 - \mu_2) \qquad \lambda_2^A = \frac{1}{2}(\mu_2 - \mu_1)$$

Vi vil i det følgende ikke omtale disse restriksjonene i modellene, men man må være oppmerksom når man leser resultatene fra utførte variansanalyser.

En interessant hypotese her er om forventningene er like i alle grupper, dvs.

$$H_A: \quad \lambda_i^A = 0 \qquad \text{for} \quad i = 1, 2, \ldots, I$$

For denne modellen kan vi skrive den totale variasjon SS_T som

$$SS_T = SS_A + SS_E$$

der SS_A er variasjon som tilskrives faktoren A og SS_E er uforklart variasjon (E = ”Error”). For å teste hypotesen H_A mot alternativet at minst to ulike forventninger kan en bruke testobservatoren

$$F_A = \frac{MS_A}{MS_E} = \frac{SS_A/(I-1)}{SS_E/(n-I)}$$

og en forkaster H_A når $F_A \geq k$, der k er en kritisk verdi bestemt av tabeller over den såkalte F-fordelingen, her med frihetsgradtall $(I-1, n-I)$. Alternativt beregnes den såkalte P-verdi.

For denne modellen har vi ANOVA-tabellen

	SS	f	MS	F	P
Faktor A	SS_A	$I-1$	MS_A	F_A	P_A
Uforklart	SS_E	$n-I$	MS_E	–	–
Total	SS_T	$n-1$	–	–	–

Merknad. Noen ANOVA-tabeller har med ett ledd til, som representerer kvadratet av det totale gjennomsnitt med 1 frihetsgrad.

ANOVA-modellen med to faktorer

Hvis vi istedenfor parrede observasjoner, dvs. blokker med to grupper i hver blokk, har *I* > 2 grupper i hver av *J* blokker, er det tale om ANOVA med to faktorer A og B, der A er faktoren av primær interesse og B er blokkfaktoren.

La X_{ij} være observasjonen i gruppe nr. *i* innen blokk nr. *j*. Det totale antall observasjoner er da $n = I \cdot J$.

En ANOVA modell med to faktorer A og B antar at alle *n* observasjoner er uavhengige normalfordelte med forventninger

$$EX_{ij} = \lambda + \lambda_i^A + \lambda_j^B \qquad \text{der} \qquad \sum \lambda_i^A = \sum \lambda_j^B = 0$$

og varians $\operatorname{var} X_{ij} = \sigma^2$ for alle *i*, *j*.

Modellen for parrede observasjoner ovenfor svarer til $I = 2$ og $J = n$.

$$E(X_{2j} - X_{1j}) = \lambda + \lambda_2^A + \lambda_j^B - (\lambda + \lambda_1^A + \lambda_j^B) = \lambda_2^A - \lambda_1^A = \delta$$

I ANOVA-modellen med to faktorer kan vi teste to hypoteser

$$H_A: \ \lambda_i^A = 0 \quad \text{for} \quad i = 1, 2, \ldots, I$$

dvs. faktor A er uten betydning

$$H_B: \ \lambda_i^B = 0 \quad \text{for} \quad j = 1, 2, \ldots, J$$

dvs. at blokkfaktoren B er uten betydning.

For denne modellen har vi ANOVA-tabellen

	SS	f	MS	F	P
Faktor A	SS_A	$I-1$	MS_A	F_A	P_A
Blokk	SS_B	$J-1$	MS_B	F_B	P_B
Uforklart	SS_E	$(I-1)(J-1)$	MS_E	–	–
Total	SS_T	$IJ-1$	–		

Som før er $MS = SS/f$ og $F = MS/MS_E$.

For å redusere usikkerheten kan det være aktuelt med gjentatt observasjon for hver kombinasjon ij. Et eksempel er den før beskrevne situasjon med slitasje på bildekk, der en monterte to dekk av hver type på hver av bilene. I vår notasjon trengs en fotskrift til for å markere observasjon nr. k for kombinasjonen ij. Frihetsgradene i tabellen ovenfor vil da også avhenge av hvor mane gjentak det er.

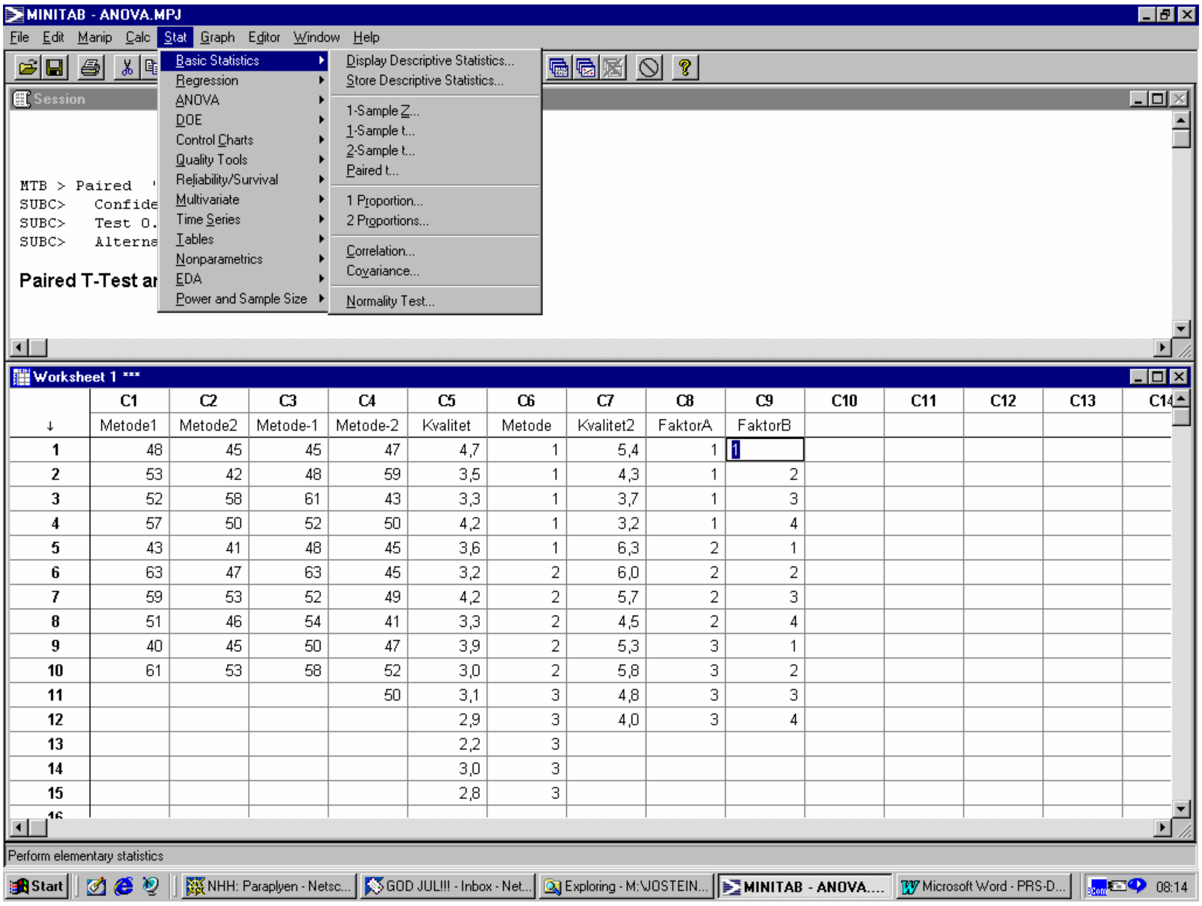
Mer generelt kan vi ha situasjoner med to faktorer A og B, der faktor A observeres på I nivåer og faktor B på J nivåer, der begge faktorene er av interesse, og B ikke bare er en blokkfaktor som vi vil korrigere for. Dersom det foreligger mulighet for samspill mellom faktorene, trengs mer enn en observasjon pr. faktorkombinasjon. La X_{ijk} være observasjon nr. k for faktorkombinasjon (i, j), $k = 1, 2, \ldots, K_{ij}$. Situasjonen kalles balansert når det er like mange observasjoner for hver faktorkombinasjon $K_{ij} = K$. En ANOVA-modell ser da slik ut:

$$EX_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$$

der hver av effektparametrene summeres til 0.

I dette avsnittet viser vi noen utvalgte eksempler på utskrifter fra Minitab. Dette er en brukervennlig statistisk programpakke med høy faglig kvalitet. I tillegg til de vanlige enkle statistiske analysemetode, omfatter den en rekke ting av spesiell interesse i kvalitetsstyring, bl.a. statistisk prosesstyring og forsøksplanlegging og analyse.

Nedenfor ser vi et Miniab-arbeidsområde med to viduer: et datavindu, der de data som skal analyseres skrives eller leses inn (fra ekstern fil) og et sesjonsvindu, der en får resultater. Mintab er menybasert, men tar også kommandoer i sesjonsvinduet. Vi ser nedenfor at valget **Stat** på hovedmenyen gir oss en rullegardinmeny med stikkord for grupper av ulike statistiske metoder. Her er valgt **Basic Statistics**, som i sin tur gir valg mellom ulike enkle statistiske beregninger og tester. Vi ser at **DOE** (Design of Experiments) er et av de andre valgene på rullegardinmenyen.



Minitab har gode interne hjelpemuligheter, og en kan faktisk lære en god del om forsøksplanlegging vel å utnytte disse mulighetene.

Vi vil først ta for oss eksempler på teorien i forrige avsnitt. Data for hvert eksempel ligger som kolonner i arbeidsområdet

Eksempel 1: Modellen for parrede observasjoner

En arbeidsoperasjon kan utføres ved to metoder, og en ønsker å finne ut om Metode 1 medfører lengre forventet tidsforbruk enn Metode 2 Ti medarbeidere har prøvd seg med hver av de to metodene. Resultatet ble (i sekunder):

Metode 1:	48	53	52	57	43	63	59	51	40	61
Metode 2:	45	42	58	50	41	47	53	46	45	53

```
MTB > Paired  'Metode1' 'Metode2';
SUBC>   Confidence 95.0;
SUBC>   Test 0.0;
SUBC>   Alternative 1.
```

Paired T-Test and Confidence Interval

```
Paired T for Metode1 - Metode2

                  N      Mean     StDev   SE Mean
Metode1          10     52.70      7.56      2.39
Metode2          10     48.00      5.40      1.71
Difference       10      4.70      6.70      2.12

95% CI for mean difference: (-0.09; 9.49)
T-Test of mean difference = 0 (vs > 0): T-Value = 2.22  P-Value = 0.027
```

Utskriften gir gjennomsnitt (Mean) og standardavvik (StDev) for tidsforbruket for hver av de to metodene, samt for differensen i tidsforbruk. Standardfeilen til gjennomsnittet (SEMean = StDev/sqrt(N)) bestemmer feilmarginen ved gjennomsnittet som anslag for de ukjente forventede (forskjeller) i tidsforbruk. Utskriften gir også et konfidensintervall (CI) for forventet forskjell i tidsforbruk. Konfidensgraden er 95%, men kan om ønskelig velges selv. En har her valgt å teste nullhypotesen ingen forventet forskjell, mot alternativet at Metode 1 gir lengre forventet tidsforbruk. T-testen gir en P-verdi på 2.7%, slik at nullhypotesen forkastes til fordel for alternativet ved valgt signifikansnivå på 5%, men ikke ved 1%. Ved tosidisg test hadde P-verdien vært det dobbelte 5.4%. Dette samsvarer med at det tosidige 95%-konfidensintervallet så vidt omslutter forventet forskjell null.

Merknad. Signifikansnivå er risikoen for å forkaste nullhypotesen når denne er riktig (ofte kalt feil av type 1). Et ønske om at denne skal være liten må holdes opp mot sannsynligheten for å ikke oppdage forskjeller når slike er tilstede (feil av type 2). Med et gitt antall observasjoner er disse to ønsker konkurrerende. P-verdi er sannsynligheten for minst like påfallende resultat som det observerte i favør av alternativet til nullhypotesen, beregnet under forutsetning av at denne er riktig. En liten P-verdi taler derfor i favør av altenativet til nullhypotesen, og kan sammenholdes med det valgte signifikansnivå.

Eksempel 2: To-utvalgsmodellen

En arbeidsoperasjon kan utføres ved to metoder, og en ønsker å finne ut om Metode 1 medfører lengre forventet tidsforbruk enn Metode 2 Av 21 medarbeidere ble 10 valgt ut tilfeldig til Metode 1, mens de øvrige brukte Metode 2. Resultatet ble (i sekunder):

Metode 1:	45 48 61 52 48 63 52 54 50 58
Metode 2:	47 59 43 50 45 45 49 41 47 52 50

```
MTB > TwoSample 95.0 'Metode-1' 'Metode-2';
SUBC>   Alternative 1;
SUBC>   Pooled.
```

Two Sample T-Test and Confidence Interval

```
Two sample T for Metode-1 vs Metode-2

           N      Mean     StDev   SE Mean
Metode-1  10     53.10      5.92       1.9
Metode-2  11     48.00      4.90       1.5

95% CI for mu Metode-1 - mu Metode-2: ( 0.2;  10.0)
T-Test mu Metode-1 = mu Metode-2 (vs >): T = 2.16  P = 0.022  DF = 19
Both use Pooled StDev = 5.40
```

Utskriften gir gjennomsnitt (Mean) og standardavvik (StDev) for tidsforbruket for hver av de to metodene, samt standardfeilen til gjennomsnittet (SEMean). Videre har vi et 95% konfidensintervall (CI) for forventet forskjell i tidsforbruk, samt T-testen for hypotesen ingen forventet forskjell, mot alternativet at Metode 1 gir lengre forventet tidsforbruk enn Metode 2. Vi ser at den beregnede P-verdien er 2.2%, slik at nullhypotesen forkastes til fordel for alternativet ved valgt signifikansnivå på 5%, men ikke ved 1%.

Eksempel 3: ANOVA med en faktor

En bedrift vurderer tre ulike produksjonsmetoder med sikte på å finne den beste, målt med en utvalgt nøkkelkarakteristikk. Det er prøveprodusert 5 enheter med hver metode. Resultatet ble (jo større tall desto bedre):

Metode 1:	4.7	3.5	3.3	4.2	3.6
Metode 2:	3.2	4.2	3.3	3.9	3.0
Metode 3:	3.1	2.9	2.2	3.0	2.8

Det er mest hensiktsmessig å organiserte disse data i en og samme kolonne ,med en tilhørende kolonne som indikerer produksjonsmetoden (kodetall 1, 2, 3)

```
MTB > Oneway 'Kvalitet' 'Metode'.
```

One-way Analysis of Variance

```
Analysis of Variance for Kvalitet
Source     DF        SS        MS        F        P
Metode      2     2.929     1.465     6.15    0.015
Error      12     2.860     0.238
Total      14     5.789
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level       N      Mean     StDev  --+---------+---------+---------+----
1           5    3.8600    0.5771                    (-------*-------)
2           5    3.5200    0.5070               (-------*-------)
3           5    2.8000    0.3536   (-------*-------)
                                   --+---------+---------+---------+----
Pooled StDev =   0.4882            2.40      3.00      3.60      4.20
```

Utskriften gir oss først ANOVA-tabellen med bl.a. F-testen for hypotesen at de tre produksjonsmetodene gir likt forventet resultat. Vi ser at denne hypotesen blir forkastet ved valgt 5% signifikansnivå, men ikke ved 1% signifikansnivå. Utskriften gir videre gjennomsnitt og standardavvik for hver metode, samt 95% konfidensintervaller for forventningene basert på antakelsen om like standardavvik. Vi er tilbøyelige til å rangere metodene i denne rekkefølge: Metode 1 > Metode 2 > Metode 3, men tallmaterialet gir egntlig bare holdepunkt for å påstå at Metode 1 er bedre enn Metode 3. På Minitab-mmenyen fins muligheter for mer detaljert informasjon vedrørende slike såkalte multiple sammenligninger.

Eksempel 4: ANOVA med to faktorer

Kvaliteten av et produkt kan avhenge av produksjonsmetode (faktor A) og råstoff (faktor B) . Det er tre produksjonsmetoder og fire råstoffkvaliteter, og en har produsert en enhet for hver av de 12 kombinasjonene metode og råstoff. Resultatet ble:

		Råstoff			
		1	2	3	4
	1	5.4	4.3	3.7	3.2
Metode	2	6.3	6.0	5.7	4.5
	3	5.3	5.8	4.8	4.0

Det er her hensiktsmessig å organisere målingene i en datakolonne med 12 elementer, med to supplerende kolonner som angir nivåene på hver av faktorene (kodetall 1, 2, 3 for faktor A og kodetall 1, 2, 3, 4 for faktor B) .

```
MTB > Twoway 'Kvalitet2' 'FaktorA' 'FaktorB';
SUBC>   Additive;
SUBC>   Means 'FaktorA' 'FaktorB'.
```

Two-way Analysis of Variance

```
Analysis of Variance for Kvalitet
Source        DF        SS        MS        F        P
FaktorA        2     4.372     2.186    15.83    0.004
FaktorB        3     5.497     1.832    13.27    0.005
Error          6     0.828     0.138
Total         11    10.697

                       Individual 95% CI
FaktorA         Mean   ---------+---------+---------+---------+--
1               4.15    (------*-------)
2               5.63                          (-------*------)
3               4.98                (-------*------)
                       ---------+---------+---------+---------+--
                             4.20      4.80      5.40      6.00

                       Individual 95% CI
FaktorB         Mean   --+---------+---------+---------+---------
1               5.67                          (-------*------)
2               5.37                      (-------*------)
3               4.73             (-------*------)
4               3.90   (-------*------)
                       --+---------+---------+---------+---------
                      3.50      4.20      4.90      5.60
```

Eksempel 5 Full-faktorielt 2x2x2-forsøk

En fruktdrikk leveres i pappemballasje, og en vurderer tre ulike faktorer ved produksjonen: Vokslag (A), luftrom (B) og tilsetning (C). Et full-faktorielt forsøk ble utført, der hver faktor ble variert på to nivåer med et gjentak. Smaken av produktet ble vurdert på en skala fra 1 til 9 av en smaksdommer. Resultatet ble:

Vokslag (A)	Luftrom (B)	Tilsetning (C)	Vurdering	
Tynt	Intet	Lav	6	8
Tynt	Intet	Høy	8	7
Tynt	Lite	Lav	9	9
Tynt	Lite	Høy	1	2
Tykt	Intet	Lav	6	7
Tykt	Intet	Høy	6	8
Tykt	Lite	Lav	9	8
Tykt	Lite	Høy	2	3

Disse data organiseres i en fil med en kolonne for 'Vurdering' (C1) og en kolonne med kodetall (-1 eller +1) for hver av faktorene.(C2, C3 og C4). Merk at dersom gjentaket var med to ulike dommere, bør vi definere hver som en blokk, og markere dette med kodetall i en kolonne (C5) og spesifisere dette ved analysen.

```
MTB > FFactorial 'Vurdering' = C2 C3 C4 C2*C3 C2*C4 C3*C4;
```

Factorial Fit

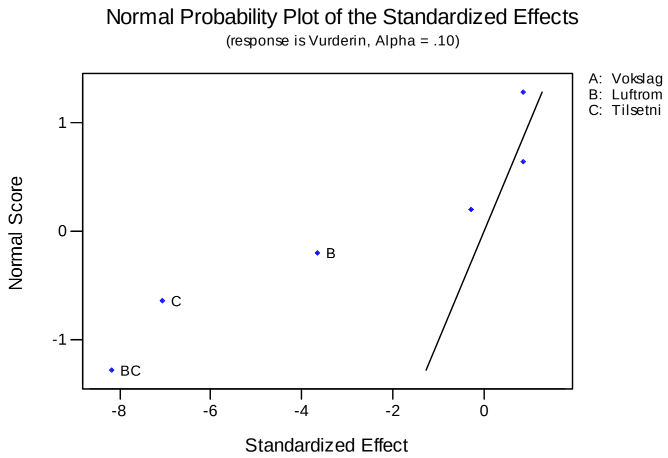
```
Estimated Effects and Coefficients for Vurdering

Term                  Effect      Coef  StDev Coef       T      P
Constant                         6.188      0.2215   27.94  0.000
Vokslag               -0.125    -0.062      0.2215   -0.28  0.784
Luftrom               -1.625    -0.812      0.2215   -3.67  0.005
Tilsetning            -3.125    -1.563      0.2215   -7.06  0.000
Vokslag*Luftrom        0.375     0.187      0.2215    0.85  0.419
Vokslag*Tilsetning     0.375     0.188      0.2215    0.85  0.419
Luftrom*Tilsetning    -3.625    -1.813      0.2215   -8.18  0.000

Analysis of Variance for Vurdering

Source             DF      Seq SS     Adj SS     Adj MS      F      P
Main Effects        3      49.688    49.6875    16.5625  21.11  0.000
2-Way Interactions  3      53.687    53.6875    17.8958  22.81  0.000
Residual Error      9       7.063     7.0625     0.7847
  Lack of Fit       1       0.563     0.5625     0.5625   0.69  0.430
  Pure Error        8       6.500     6.5000     0.8125
Total              15     110.437
```

Ut fra P-verdiene ser det ser ut til at 'Vokslag' ikke har betydning for smaken, men at 'Tilsetning' og 'Luftrom' har det. Vi ser også at samspilleffekten mellom disse to er statistisk signifikant. Disse funn kommer også til uttrykk i plottet nedenfor.



I dette såkalte normalplottet av effektene er hver effekt er markert med et punkt, og det er inntegnet en linje som gir en uformell mulighet for å vurdere om de enkelte effektene er av betydning. Betydningsløse effekter har punkter nær linjen, mens effektene øker i betydning ettersom punktene fjerner seg fra linjen. De mest betydningsfulle er markert med sin bokstavkombinasjon i plottet. En annen mulighet er det såkalte Paretoplottet, der faktorene representeres ved søyler, og der de vitale få faktorene fremtrer som størst.

Eksempel 6 Fraksjonell design

Et produktutviklingsteam har ved idedugnad kommet frem til 6 faktorer A, B, C, D, E og F som en tror kan påvirke slitestyrken av et tekstil. Man ønsker å bestemme hvilke av disse som er viktigst, for å prioritere denne/disse i det videre arbeid. En regner med muligheter for tofaktor samspill mellom faktorene, men utelukker høyere ordens samspill. En faktoriell design med resolusjon IV er derfor nok. Denne resolution IV faktoriell design, som i tilfellet med 6 faktorer er 1/4 fraksjon med 16 runs, er derfor nok. Minitab gir følgende:

Factorial Design

```
Fractional Factorial Design

Factors:   6    Base Design:      6, 16     Resolution: IV
Runs:     16    Replicates:           1     Fraction: ¼
Blocks: none    Center pts (total):   0

Design Generators:  E = ABC  F = BCD

Defining Relation:  I = ABCE = BCDF = ADEF

Alias Structure

I + ABCE + ADEF + BCDF
A + BCE + DEF + ABCDF
B + ACE + CDF + ABDEF
C + ABE + BDF + ACDEF
D + AEF + BCF + ABCDE
E + ABC + ADF + BCDEF
F + ADE + BCD + ABCEF
AB + CE + ACDF + BDEF
AC + BE + ABDF + CDEF
AD + EF + ABCF + BCDE
AE + BC + DF + ABCDEF
AF + DE + ABCD + BCEF
BD + CF + ABEF + ACDE
BF + CD + ABDE + ACEF
ABD + ACF + BEF + CDE
ABF + ACD + BDE + CEF

Data Matrix (randomized)

 Run   A  B  C  D  E  F
   1   +  +  +  -  +  -
   2   -  -  -  +  -  +
   3   +  -  +  +  -  -
   4   +  -  +  -  -  +
   5   -  +  -  -  +  +
   6   -  +  +  +  -  +
   7   -  -  +  -  +  +
   8   -  +  -  +  +  -
   9   +  +  -  -  -  +
  10   +  +  -  +  -  -
  11   -  +  +  -  -  -
  12   -  -  -  -  -  -
  13   +  -  -  +  +  +
  14   +  +  +  +  +  +
  15   +  -  -  -  +  -
  16   -  -  +  +  +  -
```

Eksempel 7 Plackett-Burman Screening design

Et produktutviklingsteam har ved idedugnad kommet frem til 11 faktorer A - L som en tror kan påvirke slitestyrken av et tekstil. En regner med at 2-3 av disse betyr langt mer en de øvrige. For å unngå å bruke ressurser til mange forsøk innledningsvis, velger en å se bort fra mulige samspill, og bruker en Plackett-Burman design, som med 11 faktorer har kun12 runs. Eventuelle samspill mellom de viktige faktorer som avdekkes tas omsyn til ved planleggingen av nye forsøk. Minitab gir følgende:

Factorial Design

```
Plackett-Burman Design

Factors:  11    Replicates:           1    Design:  12
Runs:     12    Center pts (total):   0

Data Matrix (randomized)
```

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

Eksempel 8 Responsflate[3]

I et laboratorieforsøk ønsker en å finne en kombinasjon av temperatur (T) og tid (t) som kan gi bedre resultat enn den kombinasjon som brukes i dag. Disse er T=130 C og t=75 minutter. En vil gå fram trinnvis og starter med en enkel 2x2-design med et senterpunkt gjentatt tre ganger, i alt 7 runs, som vist i tabellen

	Faktorer		Faktorer (kodet)		Respons
run nr.	tid (minutt)	Temp (i C)	x1	x2	y
1	70	127.5	-1	-1	54.3
2	80	127.5	+1	-1	60.3
3	70	132.5	-1	+1	64.6
4	80	132.5	+1	+1	68.0
5	75	130.0	0	0	60.3
6	75	130.0	0	0	64.3
7	75	130.0	0	0	62.3

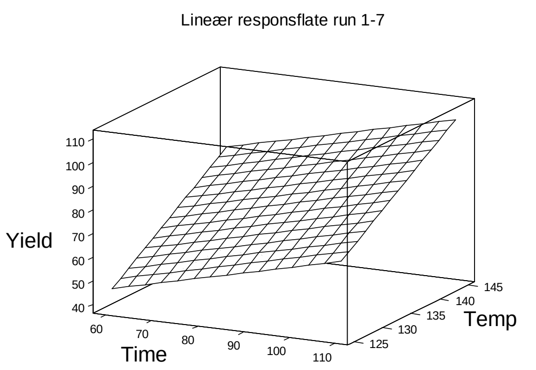
Response Surface Regression (1)

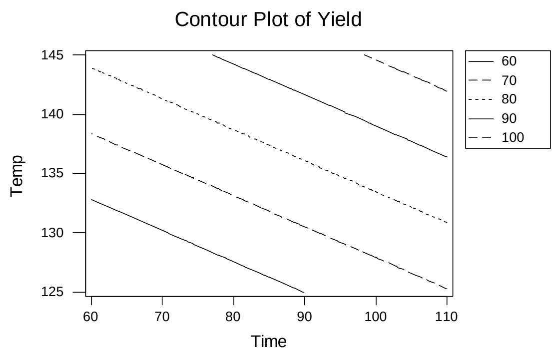
```
The analysis was done using coded units.

Estimated Regression Coefficients for Yield

Term            Coef      StDev         T      P
Constant      62.014     0.6011   103.160  0.000
Time           2.350     0.7952     2.955  0.042
Temp           4.500     0.7952     5.659  0.005

S = 1.590      R-Sq = 91.1%     R-Sq(adj) = 86.6%
```





3 Etter Box, Hunter & Hunter (1978)

Vi ser at responsen ”stiger mot nordøst”, og at en lineær tilnærming rundt senterpunktet er gitt ved

$$Y = 62.0 + 2.35K_1 + 4.50K_2$$

Dette betyr at retningen med brattest stigning er gitt ved vinkelkoeffisienten 4.50/2.35=1.91, slik at dersom en går k1 kodede enheter i t-retningen, må en gå k2=1.91k1 kodede enheter i T-retningen. Fem aktuelle kombinasjoner i denne retning er gitt i følgende tabell, der vi også har beregnet hvilke nivåer dette er på faktorenes naturlige måleskala.

	Faktorer (kodet)		Faktorer			Respons
	k1	k2	tid (t)	Temp (T)	run. nr.	y
center	0	0	75	130.0	5,6,7	snitt: 62.3
retning brattest stigning	1	1.91	80	134.8	8	73.3
	2	3.83	85	139.8		
	3	5.74	90	144.4	10	86.8
	4	7.66	95	149.1		
	5	9.57	100	153.9	9	58.2

Her har en valgt å kjøre tre runs i denne retningen (run 8, 9 og 10). Vi ser at run nr. 10 ligger halvveis mellom de to ytterpunktene 8 og 9 og ga det desidert beste resultatet. Ved run nr.9 er vi tydeligvis over bakketoppen,.og nede i neste dal. Dette gir oss beskjed om at en bør prøve en ny design med (90,145) som senterpunkt. Dette er run nr.11-16 nedenfor. En analyse av resultatet fra disse følger.

	Faktorer		Faktorer (kodet)		Respons
run nr.	tid (minutt)	Temp (i C)	x1	x2	y
11	80	140.0	-1	-1	78.8
12	100	140.0	+1	-1	84.5
13	80	150.0	-1	+1	91.2
14	100	150.0	+1	+1	77.4
15	90	145.0	0	0	89.7
16	90	145.0	0	0	86.8
Run i tillegg som til sammen gir en CC-design					
17	76	145.0	$-\sqrt{2}$	0	83.3
18	104	145.0	$\sqrt{2}$	0	81.2
19	90	138.0	0	$-\sqrt{2}$	81.2
20	90	152.0	0	$\sqrt{2}$	79.5
21	90	145.0	0	0	87.0
22	90	145.0	0	0	86.0

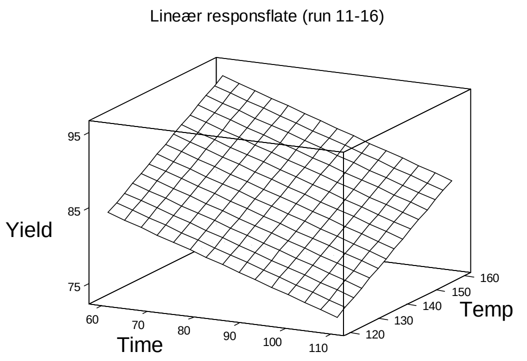
Response Surface Regression (2)

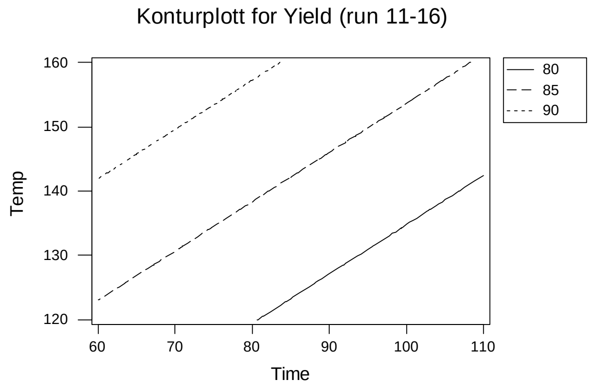
```
The analysis was done using coded units.

Estimated Regression Coefficients for Yield

Term            Coef       StDev          T       P
Constant      84.733       2.752     30.785   0.000
Time          -2.025       3.371     -0.601   0.590
Temp           1.325       3.371      0.393   0.721

S = 6.742       R-Sq = 14.7%     R-Sq(adj) = 0.0%
```





Plottene antyder at herfra går en ny retning for forbedring, nemlig mot nordvest. Imidlertid viser regresjonsanalysen en liten forklaringsgrad (bare 14.7%) og stor variasjon omkring regresjonsplanet (S=6.742), slik at terrenget er ulendt og tildels udefinerbart ut fra lineære betraktninger. En utvider derfor designen til en Central-Composite design, med 4 aksialpunkter og 2 nye gjentak av senterpunktet (run nr. 17-22). Merk at aksialpunktene ligger like langt fra senterpunktet som hjørnepunktene.

En analyse med full kvadratisk tilnærming basert på 12 runs (run nr.11-22) gir resultat nedenfor, og gir et godt inntrykk av terrenget "ved toppen". Full kvadratisk tilnærming betyr at det er både kvadratledd og produktledd (samspill) i regresjonen. Det er også mulighet for å forsøke en modell der enten kvadratleddene eller produktleddet er utelatt. Programvaren gir da mulighete for å vurdere hvor god tilnærmingen er for de ulike modellene. Fordelen med en enklest mulig modell er at vi får bedre anslag på usikkerheten, idet vi får færre ukjente regresjonskoeffisienter som må anslås.

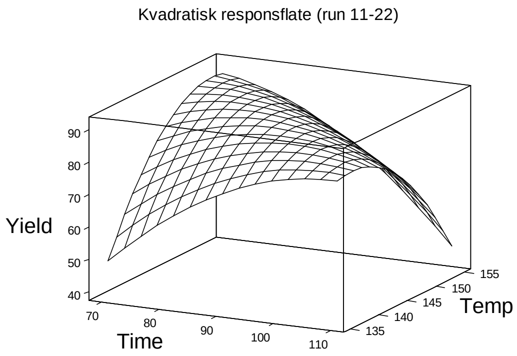
Muligheten er nå til stede for ytterligere forbedring ved å gå langs den funne stigende "åskam.".

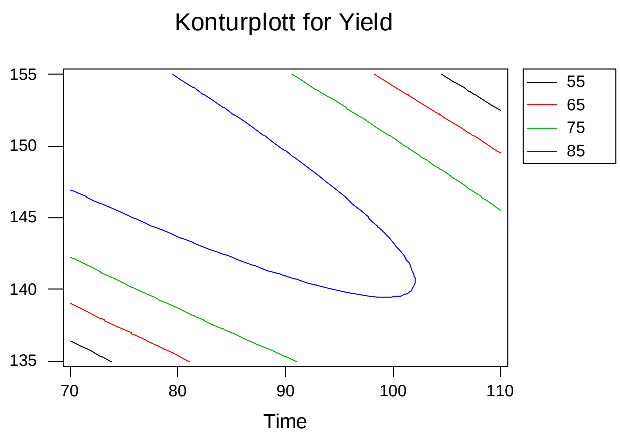
Response Surface Regression (3)

Estimated Regression Coefficients for Yield

Term	Coef	StDev	T	P
Constant	87.357	1.015	86.089	0.000
Time	-1.952	1.010	-1.933	0.101
Temp	0.516	1.010	0.511	0.627
Time*Time	-4.208	1.595	-2.639	0.039
Temp*Temp	-6.108	1.595	-3.830	0.009
Time*Temp	-9.555	1.989	-4.804	0.003

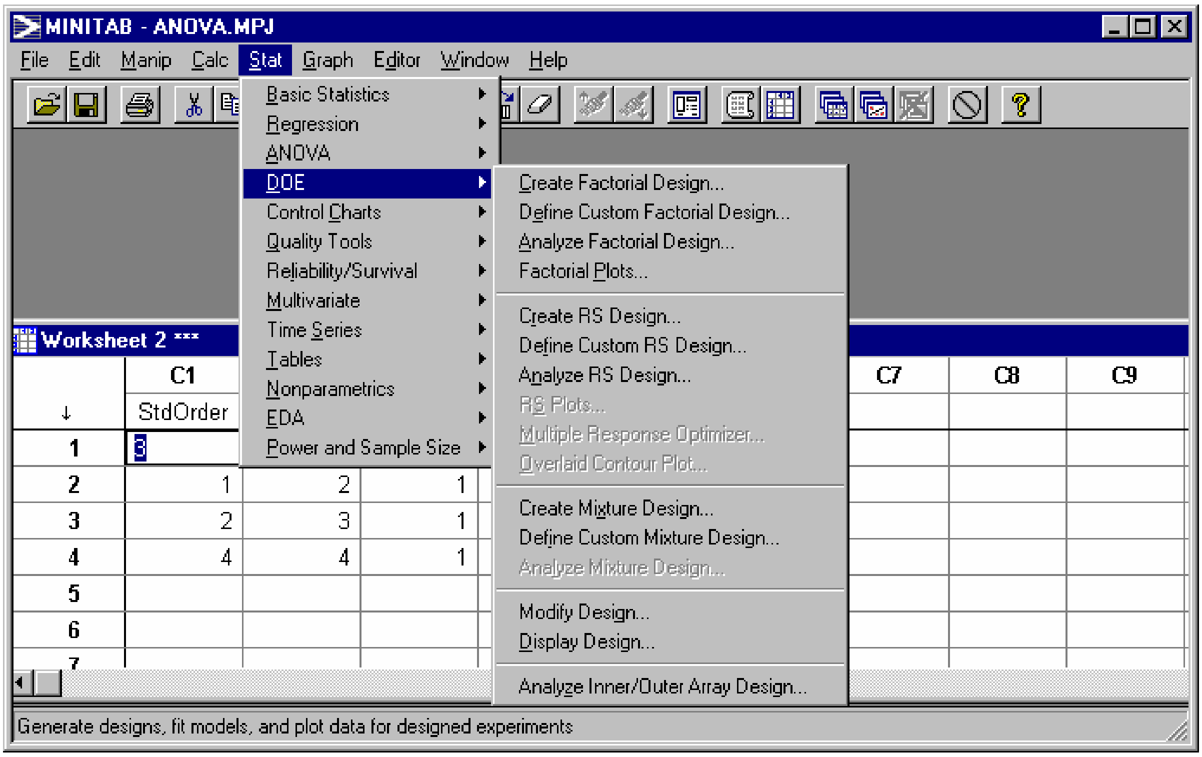
S = 2.030 R-Sq = 88.4% R-Sq(adj) = 78.7%

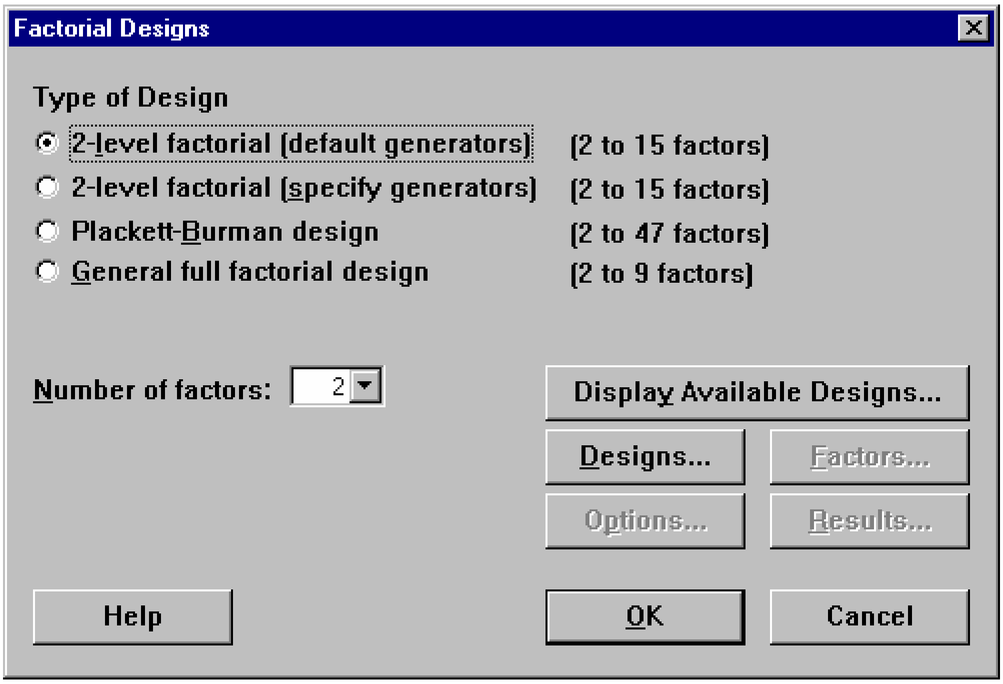


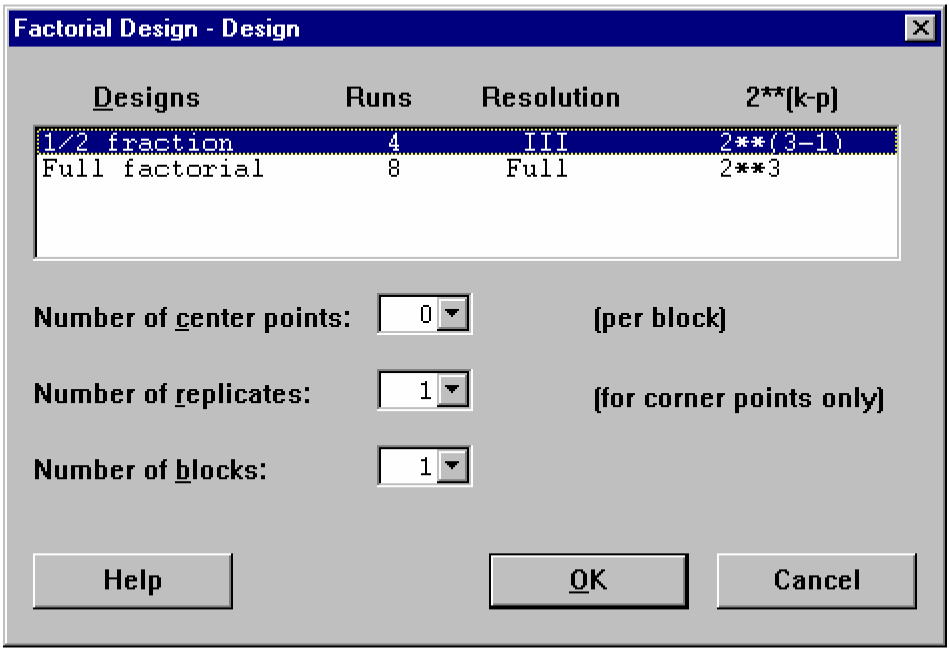


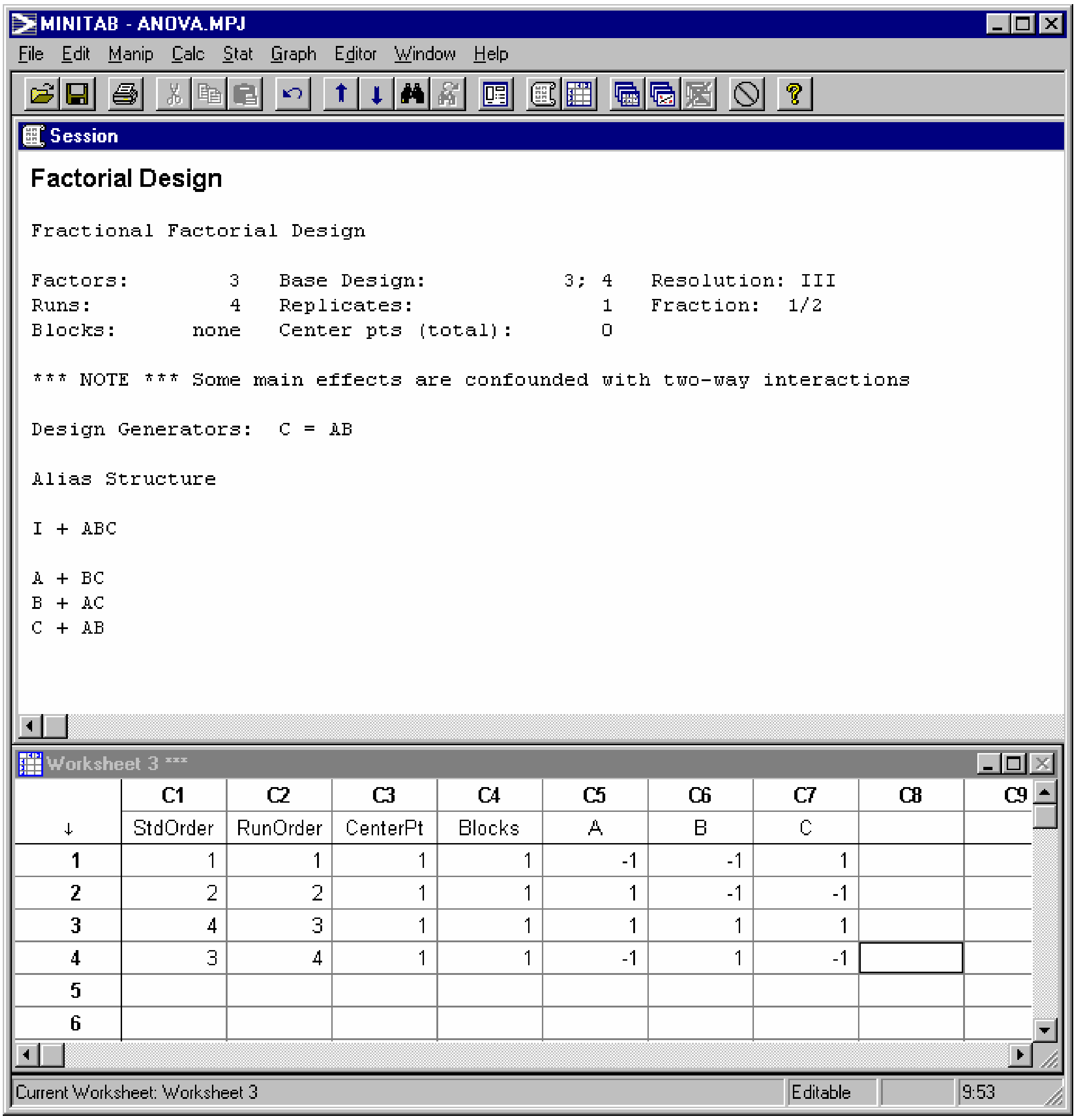
DOE-paneler i Minitab

Velger vi **DOE** på rullegardinmenyen i Minitab, får vi en rekke valg, bl.a. **Create factorial design.** Velges dette får vi panelene nedenunder for utfylling, og velger vi en halv-fraksjon, får vi resultat som vist til slutt.






Factorial Design - Design
Designs
Runs
Resolution
2**(k-p)
1/2 fraction 4 III 2**(3-1)
Full factorial 8 Full 2**3
Number of center points: 0
(per block)
Number of replicates: 1
(for corner points only)
Number of blocks: 1
Help
OK
Cancel


MINITAB - ANOVA.MPJ
File Edit Manip Calc Stat Graph Editor Window Help
Session
Factorial Design
Fractional Factorial Design
Factors: 3 Base Design: 3; 4 Resolution: III
Runs: 4 Replicates: 1 Fraction: 1/2
Blocks: none Center pts (total): 0
*** NOTE *** Some main effects are confounded with two-way interactions
Design Generators: C = AB
Alias Structure
I + ABC
A + BC
B + AC
C + AB
Worksheet 3 ***
C1 C2 C3 C4 C5 C6 C7 C8 C9
StdOrder RunOrder CenterPt Blocks A B C
1 1 1 1 1 -1 -1 1
2 2 2 1 1 1 -1 -1
3 4 3 1 1 1 1 1
4 3 4 1 1 -1 1 -1
5
6
Current Worksheet: Worksheet 3
Editable
9:53

.

13. Litteratur

Bergman, B.	Industriell Forøksplanering och Robust Konstruktion. Studentlitteratur, Lund 1992.
Box, G.E.P. & Draper, N.R.	Empirical Model-Building and Response Surfaces, Wiley, New York 1989
Box, G.E.P., Hunter, W.G.& Hunter, J.S.	Statistics for Experimenters. An introduction to Design, Data Analysis and Model Building Wiley, New York 1978.
Hansson,M & Bodizs, R.	Kvalitetsforbettring med Forsøksplanering. Studentlitteratur, Lund 1999.
Lillestøl, J.	Sannsynlighetsregning og Statistikk med anvendelser. Cappelen Akademisk Forlag, Oslo 1997.
Montgomery , D.C.	Design and Analysis of Experiments, Wiley, New York 1991.