

Kurs PK7. Forsøksmetodikk  
Institutt for plantekultur, NLH

F O R S Ø K S M E T O D I K K

av

Øivind Nissen og Kåre Ringlund

LANDBRUKSBOKHANDELEN

ISBN 82-557-0190-7

ÅS-NLH 1990

4. utg.

Kurs PK7. Forsøksmetodikk  
Institutt for plantekultur, NLH

F O R S Ø K S M E T O D I K K

av

Øivind Nissen og Kåre Ringlund

LANDBRUKSBOKHANDELEN

ISBN 82-557-0190-7

ÅS-NLH 1990

4. utg.

## FORORD

Dette manuskriptet bygger på notater fra to forelesningsrekker, Ringlunds forelesninger i forsøksmetodikk (PK7) for studentene i 3. og 4. årsklasse ved NLH, og Nissens forelesninger i forsøksmetodikk og databehandling for fagassistenter. Studentene har tidligere hatt et kurs i statistikk, mens de fleste av fagassistentene ikke har noe kjennskap til dette faget. Dette er årsaken til at vi starter med en elementær innføring i matematisk statistikk. Vi har prøvd å gjøre dette avsnittet så enkelt som mulig, og bruker mer ord og forklaringer enn matematiske formler.

Stoffet er ordnet i logisk rekkefølge, men dette er ikke alltid best pedagogisk. Forelesningene vil derfor ikke alltid bli gitt i samme rekkefølge.

Øvelsesoppgaver er gitt bakerst i heftet.

Vi takker professor Inge Helland for nyttige kommentarer og korreksjoner til en tidligere utgave.

Vi er fortsatt takknemlige for å bli gjort oppmerksom på eventuelle feil, og for forslag til forbedringer.

Ås, NLH januar 1990

Øivind Nissen

Kåre Ringlund

## INNHold

	side
I.    INNFORING I MATEMATISK STATISTIKK .....	1
1.    Uttak av prøver .....	1
2.    Frekvensfordeling .....	2
3.    Beregning av gjennomsnittet .....	5
4.    Beregning av standardavvik .....	6
5.    Sanne verdier for gjennomsnitt og standardavvik	7
6.    Normalfordeling .....	8
7.    Prøver fra en kjent normalfordeling .....	8
8.    Prøver fra en ukjent normalfordeling .....	11
9.    Konfidensgrenser .....	12
10.    Feilen på en differens .....	14
11.    Testing av en differens .....	15
12.    Univers. prøver og hypoteser .....	18
13.    Variansanalyse .....	18
14.    F.test .....	20
15.    Forutsetninger for testing av differenser .....	22
16.    Gjennomsnitt og feil på differenser .....	22
17.    Toveis variansanalyse .....	24
18.    t.test eller F-test .....	27
19.    Feilen på differenser mellom middeltall i en toveis analyse .....	27
20.    Minste signifikante forskjell (LSD) .....	27
21.    Variasjonskoeffisienten .....	28
22.    Flerveis variansanalyser .....	28
II    STATISTISKE MODELLER .....	33
1.    Fullstendig tilfeldig fordeling .....	33
2.    Blokkforsøk .....	35
3.    Latinsk kvadrat .....	36
4.    Faktorielle forsøk .....	36
III.   FORSØK ELLER SURVEY .....	39
1.    Survey .....	39
2.    Forsøk .....	41



	side
IV. FORSØKSLEDD/FORSØKSFAKTOR .....	43
1. Kvantitative og kvalitative forsøksfaktorer ...	44
2. Faktorielle forsøksspørsmål .....	46
V. FORSØKSENHET OG ANTALL GJENTAK .....	50
1. Forsøksenhet, størrelse og form .....	50
2. Gjentak og paralleller .....	53
VI. FORSØKSPLANER .....	54
1. Fullstendig tilfeldig fordeling .....	54
2. Blokkforsøk .....	55
3. Latinsk kvadrat .....	58
4. Split-plot .....	61
5. Ufullstendige blokker (konfundering) .....	67
VII. OPPDELING AV SS OG DF .....	72
1. Kontraster .....	73
2. Ortogonale polynomer .....	75
VIII. FORSØKSSERIER .....	79
1. Ettårige forsøksserier .....	79
2. Flerårige forsøksserier .....	82
3. Forsøk med gjentatt behandling på de samme ruter	84
IX. OPPGAVER .....	86
Oppgave nr. 1 .....	86
" " 2 .....	86
" " 3 .....	87
" " 4 .....	88
" " 5 .....	89
" " 6 .....	90
" " 7 .....	91
" " 8 .....	92
" " 9 .....	93
" " 10 .....	94
" " 11 .....	95
" " 12 .....	96

## I. INNFORING I MATEMATISK STATISTIKK

Faget forsøksmetodikk er anvendt statistikk. Derfor gir vi vår innføring ved hjelp av et praktisk eksempel.

Vi har to binger med poteter, en med sort A og en med sort B, og får spørsmål om hvilken bingetype som har de største potetene. Først må vi be om å få spørsmålet presisert. Hva menes med størst?

Er det vekt eller f.eks. lengde eller tykkelse? La oss gå ut fra at svaret er at det er knollengden vi er interessert i, og at det vi ønsker å få vite er hvilken bingetype som i gjennomsnitt har de lengste knollene. Ordet gjennomsnitt er da brukt i den vanlige dagligdagse mening, nemlig det tall vi ville komme til hvis vi målte samtlige knoller, la sammen lengdene og delte summen med antall knoller.

Hvis vi vil være ekstra forsiktig og sikre oss mot misforståelse, bruker vi i statistikken betegnelsen aritmetisk middel om dette gjennomsnittet. Det finnes nemlig andre måter å beregne middel på, og det hender at disse middeltall kan være "bedre" enn det aritmetiske middel.

I det valgte eksemplet var det teknisk mulig å bestemme gjennomsnittslengden i hver bingetype ved å måle samtlige knoller. Hvis vi ser bort fra direkte måle-, telle- og regnefeil, ville resultatet bli helt riktig, og det blir ingen tvil om hvilken bingetype som har de lengste knollene. For de fleste typer av målinger er det ikke mulig å måle hele "bingen". Vi er nødt til å uttale oss om den riktige verdien på grunnlag av en prøve.

### 1. Uttak av prøver

Ved å foreta målingene på en prøve innfører vi en feil eller usikkerhet. Da knollene i hver bingetype varierer i lengde, vil prøven aldri svare eksakt til bingen, og det er nettopp denne

variasjonen som gjør at vi i det hele tatt har bruk for statistikk. Hvis alle knollene i en binge var helt like var det jo tilstrekkelig å måle en knoll fra hver binge, og vårt svar ville blitt like riktig som om vi hadde målt alle knollene.

Den første betingelsen for at vi skal kunne slutte noe av en prøve er at prøven er tatt ut tilfeldig. Med dette mener vi at den er tatt ut slik at alle knollene i bingen får lik sjanse til å bli med. (La oss se bort fra at dette er meget vanskelig eller nærmest umulig å få til.) Hvis vi hadde plukket ut vesentlig store eller vesentlig små knoller, ville vi selvfølgelig fått et helt galt gjennomsnitt.

Hadde vi tatt ut en blanding av store og små knoller, kunne vi vårt heldige og fått et noenlunde riktig gjennomsnitt, men vi hadde kommet til å overvurdere variasjonen mellom knollene. Hvis vi hadde plukket ut bare middels knoller, kunne vi også vært heldige med gjennomsnittet, men da ville vi ha undervurdert variasjonen.

## 2. Frekvensfordeling

Anta at vi hadde målt 20 tilfeldige knoller fra binge A, og 100 fra binge B, og notert målene i hele cm. For å få en bedre oversikt kan vi telle opp hvor mange knoller fra hver binge som har ulike lengder, og stille disse tallene opp i en frekvenstabell.

Tabell I.1. Frekvensfordeling for knollengde i binge A og B.

	Knollengde i cm												Total		
	5	6	7	8	9	10	11	12	13	14	15	16	antall	Sum	Gj.sn.
A:Antall	2	6	3	3	2	2	1	1					20	152	7,60
B:Antall	4	3	8	20	18	17	15	10	3	0	1	1	100	941	9,41

Eksemplet er valgt for å vise at det ikke er nødvendig å ta ut like store prøver i begge bingene, men vi skal senere vise at vi hadde oppnådd en sikrere sammenligning ved å ta ut 60 poteter fra hver bing.

Frekvenstabeller kan vi også presentere i form av stolpediagrammer, eller histogrammer, se fig. I.1.

I stedet for å operere med absolutte antall kan vi regne om til relative frekvenser, og f.eks. i histogrammet angi % knoller i hver lengdeklasse. Fordelen er at man da kan sammenligne to histogrammer direkte, selv om de bygger på ulike antall.

Da det er tatt ut akkurat 100 knoller fra B, er det bare nødvendig å regne om til % for A, se det nederste histogram. Uregelmessighetene i histogrammet vil bli mindre hvis vi øker antallet i prøven. Da kan vi også minke klassebredden (måle nøyaktigere) og derved få mindre avstand mellom stolpene i histogrammet. Grensen for dette (når antallet øker mot uendelig og klassebredden minker mot 0) er en fordelingskurve.

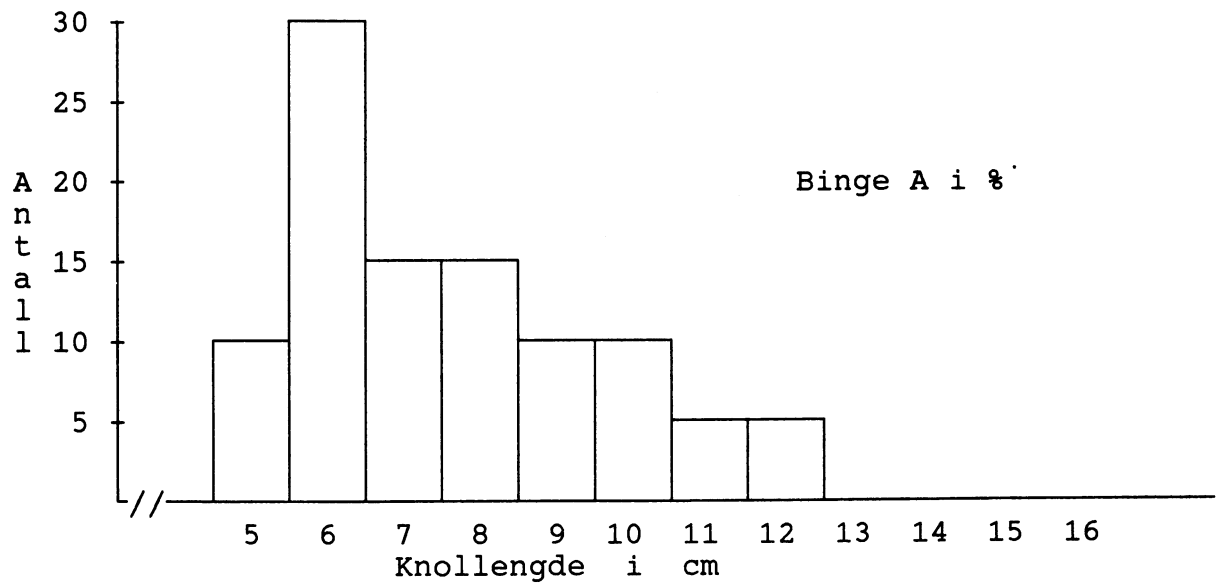
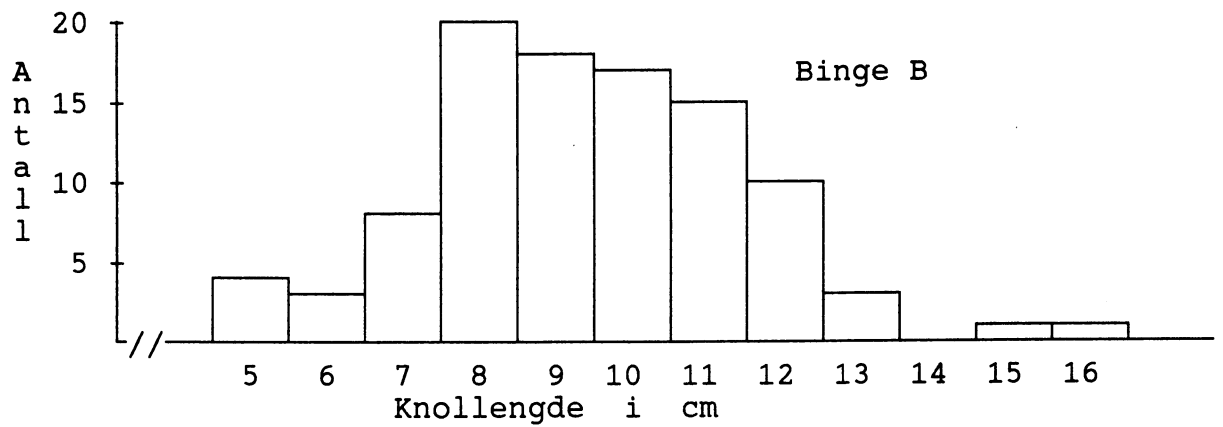
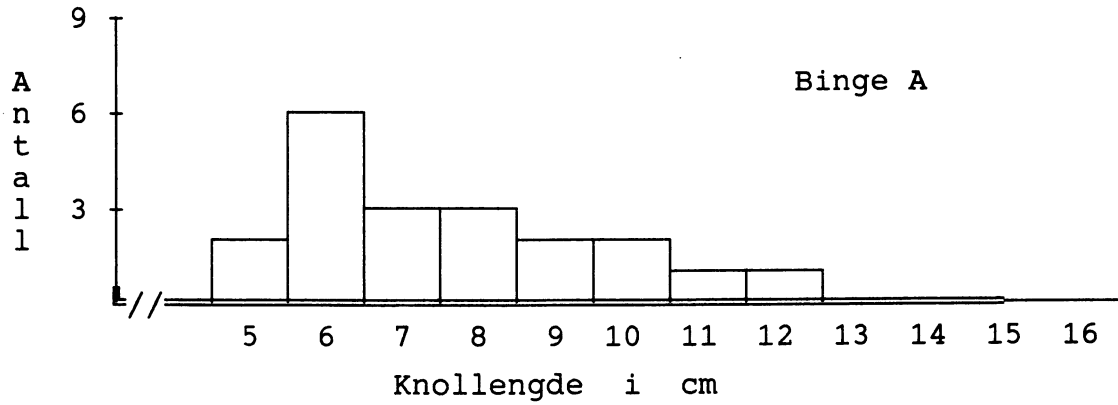
De histogrammene som er gitt i fig.I.1 er ganske karakteristiske for de fleste av de fordelinger som vi finner i praksis, med en mer eller mindre utpreget topp i midten, og avtakende verdier til begge sider. Av og til finner vi fordelinger med to eller flere topper. Dette vil da som oftest bety at vårt materiale er en blanding av to eller flere svært ulike typer.

Fordelingen kan være noenlunde symmetrisk som i eksemplet eller mer eller mindre skjev. Hvis vi f.eks. hadde bestemt vekten av hver enkelt knoll, ville vi sikkert fått en tydelig skjev fordeling. Hovedmengden hadde ligget et sted mellom 100 og 150 gram. Ingen poteter veier under 0 gram, men på den andre siden ville noen veid over 500 gram.

En entoppet og noenlunde symmetrisk fordeling er karakterisert ved hjelp av to størrelser, nemlig et mål for hvor på skalaen de fleste av observasjonene ligger, og et mål for bredden av forde-



Figur I.1 Histogram for frekvensfordelinger av potetprøver fra 2 ulike binger.



lingen Som mål for beliggenheten kan vi bruke det aritmetiske middel Som mål for bredden kunne vi brukt den absolutte variasjonsbredden, altså forskjellen mellom det største og det minste tallet. I vårt eksempel er variasjonsbredden 7 cm for A og 11 cm for B. Ulempen ved et slikt mål er for det ene at det er meget usikkert bestemt. Selv om vi har målt mange knoller er det jo bare to av målene som er med å bestemme variasjonsbredden. For det andre har variasjonsbredden den ulempen at den vil øke med økende antall i prøven. Jo flere knoller vi måler, jo større sjanse er det for at vi skal få med en riktig stor og en riktig liten knoll.

Av flere grunner bruker vi en størrelse som kalles standardavviket som breddemål.

### 3. Beregning av gjennomsnittet

For A er summen:

$$2 \times 5 + 6 \times 6 + 3 \times 7 \dots = 152, \text{ og Gj.snitt} = 152/20 = 7,60$$

og for B:

$$4 \times 5 + 3 \times 6 + 8 \times 7 \dots = 941, \text{ og Gj.snitt} = 941/100 = 9,41$$

I matematisk form bruker vi gjerne å angi de enkelte mål med  $x$ , og gjennomsnittet med  $\bar{x}$  ( $x$  med en strek over, uttales  $\bar{x}$ -bar), og vi får:

$$\bar{x} = \Sigma x/N, \text{ hvor } N \text{ er antallet.}$$

For en frekvensfordeling blir formelen:

$$\bar{x} = \Sigma x_i n_i / N$$

hvor  $n_i$  er antall observasjoner for hver klasse i frekvens tabellen,  $\Sigma n_i = N$

Vi har nå et anslag eller estimat av middellengden i A på 7,60 og i B på 9,41 cm. Vårt foreløbige svar ville være at binge B har de

lengste knollene. Hvor sikre kan vi være på at dette resultatet holder hvis vi måler alle knollene? Det er dette spørsmålet statistikken kan gi oss svar på.

#### 4. Beregning av standardavviket

Standardavviket ( $s$ ) beregnes slik:

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{(N-1)}}$$

Differensen mellom hver enkelt observasjon og det aritmetiske middel kvadreres, kvadratene summeres og summen divideres med antallet minus 1. Standardavviket er kvadratroten av kvotienten.

Standardavviket endrer seg ikke systematisk om det beregnes på grunnlag av mange eller få knoller i prøven. Forventningen for  $s^2$  er uavhengig av  $N$ .

Summen av de kvadrerte avvikene kalles for kvadratsummen (engelsk Sum of Squares, forkortet SS), mens  $(N-1)$  er antall frihetsgrader (engelsk Degrees of Freedom, forkortet DF).

Dividerer vi kvadratsummen med antall frihetsgrader får vi middelkvadratet (engelsk Mean Square, forkortet MS).

Formelen ovenfor er en definisjonsformel. I praksis brukes den sjelden til beregningene, både fordi den er tungvint (vi må regne ut alle differensene) og fordi den som oftest er unøyaktig (gjennomsnittet vil ha en avrundingsfeil som blir med i alle differenser og kvadrater). Ved hjelp av enkel algebra kan vi ut fra formelen for kvadratsummen utlede en enklere beregningsformel.

$$\begin{aligned} \sum(x-\bar{x})^2 &= \sum(x^2 - 2x\bar{x} + \bar{x}^2) \\ &= \sum x^2 - \frac{2(\sum x)(\sum x)}{N} + \frac{N(\sum x)^2}{N^2} \\ &= \sum x^2 - 2 \frac{(\sum x)^2}{N} + \frac{(\sum x)^2}{N} = \sum x^2 - \frac{(\sum x)^2}{N} \end{aligned}$$

Leddets  $(\sum x)^2/N$  kalles fradragsleddet, eller CT (for Correction Term).

I vårt eksempel kan vi beregne følgende resultater:

A:  $\sum x^2=1232$     CT=1155,20    SS= 76,80    MS=4,04    s=2,01

B:  $\sum x^2=9281$     CT=8854,81    SS= 426,19    MS=4,30    s=2,07

## 5. Sanne verdier for gjennomsnitt og standardavvik

Det vi i virkeligheten ønsker å bestemme er den sanne verdi av gjennomsnittet i bingene, den vi ville ha funnet hvis vi hadde målt alle knollene. En slik sann verdi blir gjerne betegnet med det greske bokstav  $\mu$  (my), og vår verdi  $\bar{x}$  er da et estimat av denne ukjente sanne verdien.

På tilsvarende måte har vi en ukjent sann verdi av standardavviket som oftest betegnet med det greske bokstav  $\sigma$  (sigma). Vår verdi  $s$  er et estimat av  $\sigma$ . Vi skal nå se på hvordan vi ut fra våre estimater  $\bar{x}$  og  $s$  kan si noe mer om  $\mu$ .

### Gjennomsnitt og standardavvik av gjennomsnitt

La oss først se på hvordan våre estimater av gjennomsnitt og standardavvik ville ha blitt om vi ikke hadde tatt ut en og en knoll, men f.eks. tatt ut en rekke prøver a 16 knoller, regnet gjennomsnittslengden for hver prøve, og så stilt disse gjennomsnittene sammen i et histogram, og regnet middel og standardavvik på gjennomsnittstallene.

Gjennomsnittet ville ikke blitt systematisk forandret, det er fortsatt et estimat av det sanne gjennomsnittet. Standardavviket ville derimot blitt mindre enn før. Det kan vises at det bereg-



nete standardavviket ikke lenger er et estimat av  $\sigma$ , men av  $\sigma/\sqrt{n}$ , hvor  $n$  er antall observasjoner i hver enkeltprøve. Med 16 knoller i hver enkeltprøve ville vi få et standardavvik på gjennomsnittet som er  $1/4$  av standardavviket på enkeltobservasjonene. Standardavviket på et gjennomsnitt kalles middelfeilen (forkortes til  $m$ , engelsk Standard Error, forkortet SE), og vi får:

$$m = s/\sqrt{n}$$

## 6. Normalfordelingen

Histogrammet for gjennomsnittene ligger med tyngdepunktet på samme sted på skalaen som det opprinnelige histogrammet, men det er mye smalere. Samtidig, og dette er et svært viktig poeng, er histogrammet for gjennomsnittene mer symmetrisk enn det opprinnelige. Med økende  $n$  nærmer det seg en bestemt matematisk form, den såkalte normale fordelingskurven (se fig I.2).

Ordet normal må ikke oppfattes som at slike fordelinger er vanlige i naturen. Men det faktum at selv fordelinger som avviker sterkt fra den normale, flertoppete eller skjeve, nærmer seg den normale når vi tar gjennomsnittet av endel individer, har stor betydning. Dette er begrunnelsen for at vi kan bruke formler utledet fra den normale fordeling ved vurderingen av middeltall og differenser mellom middeltall i empiriske fordelinger, sjøl om disse er ganske langt fra normale.

Normalfordelingen har som nevnt en bestemt matematisk form. Hvis vi setter midtpunktet til 0 ( $\mu=0$ ), vil vi ha overgangen mellom konkav og konveks krumning ved  $+\sigma$  og ved  $-\sigma$ . Mellom  $-\sigma$  og  $+\sigma$  ligger 68% av arealet under kurven. Utenfor  $+$  eller  $-1,96\sigma$  ligger sammenlagt 5% (halvparten på hver side), og utenfor  $+$  eller  $-2,58\sigma$  ligger sammenlagt 1% av arealet, en halv prosent på hver side.

## 7. Prøver fra en kjent normalfordeling

Hvis vi har en kjent normalfordeling (d.v.s. at vi kjenner både  $\mu$  og  $\sigma$ ) og tar ut en tilfeldig prøve av denne, kan vi oppgi sannsynligheten for at gjennomsnittet av prøven skal komme til å falle utenfor visse grenser. Hvis prøven er på  $n$  individer, vil middelfeilen på gjennomsnittet,  $m$ , være lik  $s/\sqrt{n}$ , og vi vet da at sannsynligheten for at gjennomsnittet skal ligge mer enn  $1,96 m$  fra  $\mu$  er 5%, og at det er 1% sannsynlighet for at det skal ligge utenfor  $2,58 m$  fra  $\mu$ .

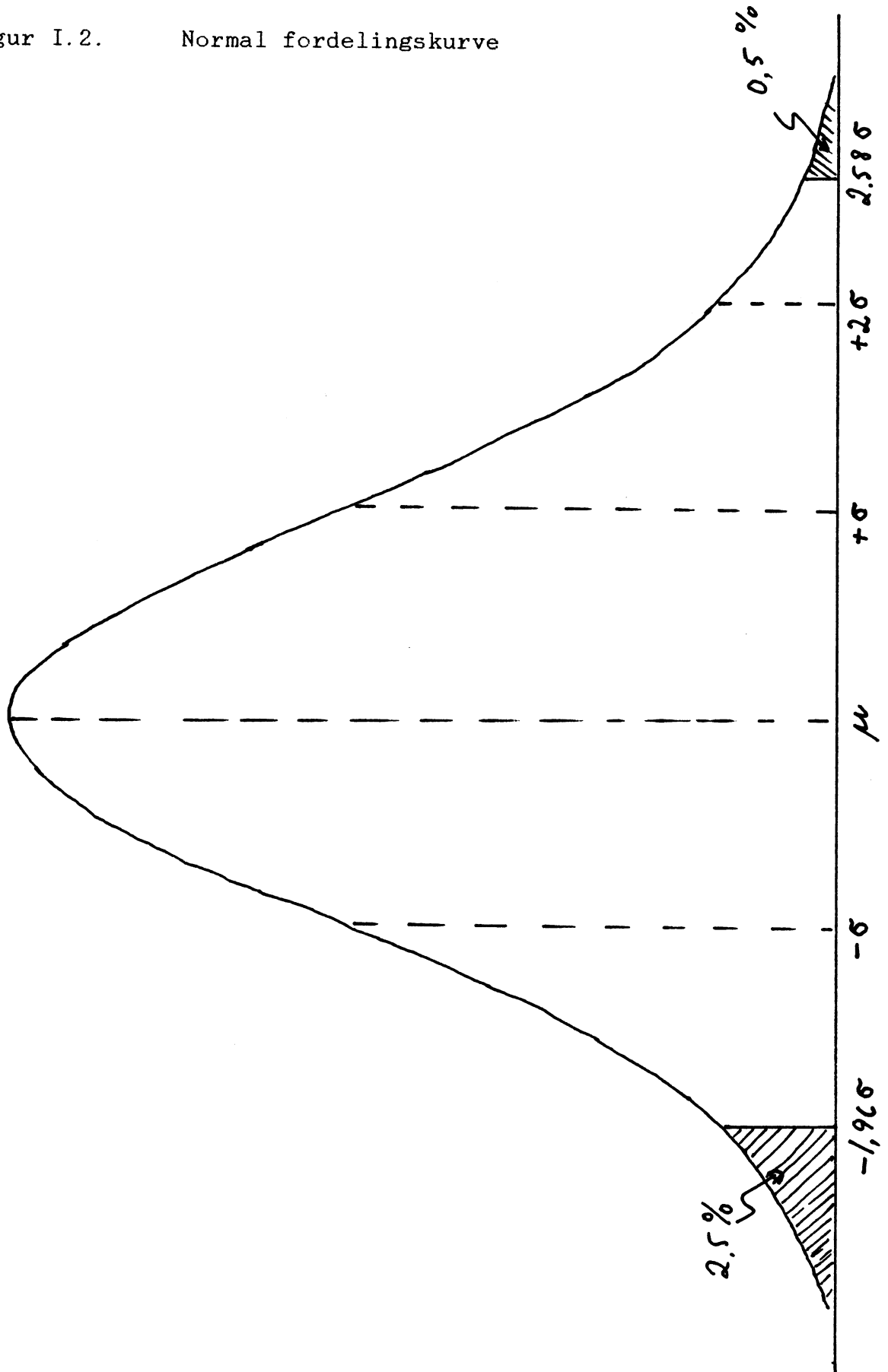
Hvis vi har en prøve og ikke på forhånd vet hvor den er tatt ut, kan vi bruke disse sannsynlighetene til å vurdere om prøven med rimelighet kan være tatt ut fra den kjente normalfordelingen.

### Talleksempel

Vi har en normalfordeling med  $\mu=100$  og  $\sigma=10$ , og får spørsmål om et enkeltindivid med verdi 111 kan være tatt ut fra denne fordelingen. Da avviket fra  $\mu$  (11) bare er litt større enn  $\sigma$  må vi svare ja til dette. Det er jo omtrent 30% sannsynlighet for at en enkeltprøve skal avvike minst så mye eller mer fra  $\mu$  i den ene eller i den andre retningen.

Hadde verdien vært 70, ville avviket vært 30, eller 3 ganger  $\sigma$ . Heller ikke da kunne vi med 100% sikkerhet sagt at prøven ikke var tatt ut av fordelingen, men vi kunne sagt at det var lite rimelig, det ville jo vært langt under 1% sannsynlighet for å få en så ekstrem verdi.

Figur I.2. Normal fordelingskurve



Som et tredje eksempel tenker vi oss en prøve på 16 individer med et gjennomsnitt på 108. Kan denne prøven være tatt ut (tilfeldig) av en fordeling med  $\mu=100$  og  $\sigma=10$  ? Her kan vi med ganske stor sikkerhet svare nei. For et gjennomsnitt av 16 har vi  $m=10/4=2,5$ , avviket fra 100 er 3,2 ganger dette, og et så stort avvik er det bare ca 0,1% sannsynlighet for.

Vi bruker vanligvis uttrykket at prøven avviker signifikant fra en oppgitte fordeling, og vi kan presisere dette ved å si "signifikant på 0,1% nivået". En slik sannsynlighet betegnes vanlig med P (engelsk Probability), og det er blitt vanlig å bruke uttrykket signifikant når P ligger mellom 5% og 1%, og meget signifikant når P er mindre enn 1%.

#### 8. Prøver fra en ukjent normalfordeling, t-test

I praksis kjenner vi ikke  $\sigma$ , men må bruke estimatet s. Dette fører da til en ekstra usikkerhet som gjør at vi må øke koeffisientene 1,96 og 2,58. Økningen må være større jo større usikkerhet vi har i estimeringen av  $\sigma$ , det vil igjen si jo færre frihetsgrader vi har i bestemmelsen av s. En britisk forsker, som publiserte under pseudonymet "Student", utarbeidet fordelingsfunksjon for

$$t = (\bar{x} - \mu) / m$$

t-fordelingen er identisk med standardnormalfordelingen når  $DF = \infty$ , og koeffisientene for de ulike sannsynlighetsnivåene øker med minkende antall frihetsgrader som er antall observasjoner minus 1 (tabell I.2.).



Tabell I.2. t-verdier ved ulike antall frihetsgrader.

P	Antall frihetsgrader (DF)										
	2	4	8	15	20	25	30	40	60	120	
10%	2,92	2,13	1,86	1,75	1,73	1,71	1,70	1,68	1,67	1,66	1,65
5%	4,30	2,78	2,31	2,13	2,09	2,06	2,04	2,02	2,00	1,98	1,96
1%	9,93	4,60	3,35	2,90	2,84	2,79	2,75	2,70	2,66	2,62	2,58
,1%	31,60	8,61	5,04	4,07	3,85	3,73	3,65	3,55	3,46	3,37	3,20

## 9. Konfidensgrenser

Vi kan nå snu resonnetet. Hvis vi av en ukjent samling har tatt ut en tilfeldig prøve på  $n$  stykker ( $n-1$  frihetsgrader) og har bestemt  $\bar{x}$ ,  $s$  og  $m$ , kan vi gi grenser for  $\mu$ . Disse betegnes som konfidensgrenser. Vi tar utgangspunkt i t-tabellens verdier for 2,5% sannsynlighet,  $t_{2,5\%}$ . Sannsynligheten for å finne en t-verdi som er større enn  $t_{2,5\%}$  er 2,5%, og sannsynligheten for å finne en t-verdi som er mindre enn  $-t_{2,5\%}$  er også 2,5%. Det er altså 95% sannsynlighet for å finne en t-verdi som ligger mellom  $-t_{2,5\%}$  og  $+t_{2,5\%}$ , det vil si:

$$-t_{2,5\%} \leq (\bar{x}-\mu)/m \leq t_{2,5\%}$$

Ved å løse disse to ulikhetene finner vi at:

$$(\bar{x}-\mu)/m \leq t_{2,5\%}$$

$$\bar{x}-\mu \leq mt_{2,5\%}$$

$$\bar{x}-mt_{2,5\%} \leq \mu$$

og at

$$(\bar{x}-\mu)/m \geq -t_{2,5\%}$$

$$\bar{x}-\mu \geq m(-t_{2,5\%})$$

$$\bar{x}+mt_{2,5\%} \geq \mu$$

eller at

$$\bar{x}-mt_{2,5\%} \leq \mu \leq \bar{x}+mt_{2,5\%}$$

Vi sier at vi har bestemt 95 % konfidensgrenser for  $\mu$  når vi fra  $\bar{x}$  trekker eller legger til produktet av tabellens t-verdi (for 2,5% og det antall frihetsgrader som vi har) og m.

La oss gå tilbake til poteteksemplet. Resultatene fra målingene av knollengder i bingje A og bingje B er stilt sammen i tabell I.3.

Tabell I.3.

	n	$\bar{x}$	SS	s	m	$t_{2,5\%}$	95% k.int.	$t_{1\%}$	99% k.int.
A	20	7,60	76,8	2,01	0,450	2,09	6,66-8,54	2,84	6,32-8,88
B	100	9,41	426,2	2,07	0,207	1,99	9,00-9,82	2,63	8,87-9,95

Vi bruker betegnelsen 95% konfidensintervall for området fra nedre 95% konfidensgrense til øvre 95% konfidensgrense, og tilsvarende for 99% intervallet.

Det er ikke riktig å si at det sanne middel for bingje A med 95% sannsynlighet ligger mellom 6,66 og 8,54. Det sanne middel er ingen tilfeldig variabel, det ligger (med 100% sikkerhet) et bestemt sted. Det som det er 95% sannsynlighet for er at vi skal treffe rett når vi angir et 95% konfidensintervall. Eller med andre ord: Hvis vi mange ganger regner ut et 95% konfidensintervall vil vi i det lange løp treffe riktig i 95% av tilfellene, og bomme i 5%.

Av tabellen foran ser vi at 99% konfidensintervallene for A og B såvidt overlapper. Det er da ikke rimelig å anta at de sanne gjennomsnitt i de to bingjene er like. Anta f.eks. at dette hadde vært 8,875 cm. Sannsynligheten for å få de to middeltall som vi har fått, 7,60 eller mindre og 9,41 eller større er da i begge tilfeller ca 1%. At dette hadde kommet i begge prøvene er selvfølgelig enda mindre rimelig.

## 10. Feilen på en differens

Vi skal nå se på utregning av feilen på en differens mellom to middeltall, og hvordan vi ved hjelp av denne kan bestemme sannsynligheten for ulike differenser mere nøyaktig.

La oss først tenke oss at vi tar ut to tilfeldige knoller fra samme bing, bestemmer differensen mellom knoll 1 og knoll 2, og så gjentar dette en rekke ganger. Differensene kan nå settes opp i en frekvenstabell, og vi kan regne middel og standardavvik på differensene. Hvis knollene er tatt ut tilfeldig blir differensen like ofte negativ som positiv. Fordelingen blir symmetrisk, og det sanne middel blir 0.

Standardavviket på differensene blir større enn på målingene av de enkelte knollene. "I gjennomsnitt" blir MS for differensrekken dobbelt så stor som MS for enkeltknollene, følgelig må standardavviket for enkeltknoller multipliseres med  $\sqrt{2}$  for å gi standardavviket på differensen.

Vi skal så se på hva resultatet blir hvis vi tar ut to prøver fra samme bing, den ene prøven på  $n_1$  (f.eks. 10) og den andre prøven med  $n_2$  (f.eks. 20) knoller. Vi beregner gjennomsnittet for hver prøve, og differensen mellom gjennomsnittene, og gjentar dette en rekke ganger. Også nå blir gjennomsnittet av differensene nær 0, og det kan vises at MS for differensrekken nærmer seg  $(\sigma^2/n_1 + \sigma^2/n_2)$ , hvor  $\sigma$  er standardavviket for enkeltprøvene.

I eksemplet med en prøve fra hver av potetbingene får vi det beste mål for et felles middelkvadrat innen bingene ved å legge sammen kvadratsummene og dividere med summen av frihetsgradene for prøven fra bing A og bing B.

$$MS = (SS_1 + SS_2) / (DF_1 + DF_2) = (76,8 + 426,2) / (19 + 99) = 4,26$$

Differensen mellom middeltallene i de to bingene er

$$d = \bar{x}_1 - \bar{x}_2 = 9,41 - 7,60 = 1,81$$

middelfeilen på differensen blir

$$m_d = \sqrt{MS/n_1 + MS/n_2} = \sqrt{4,26/20 + 4,26/100} = 0,50$$

### 11. Testing av en differens

For å teste om en differens er forskjellig fra 0, setter vi opp en hypotese om at den sanne differensen er lik 0. Under denne forutsetning er

$$t = (d - \mu) / m_d = 1,81 / 0,50 = 3,62$$

$DF_1 + DF_2 = 19 + 99 = 118$  frihetsgrader, og i t-tabellen finner vi da at P er mindre enn 0,1%. Det er derfor svært sannsynlig at den gjennomsnittlige knollengde er størst i bing B. Vi kan også regne ut konfidensgrensene for differensen. Hvis vi bruker tabellens t-verdi for 120 DF og  $P=0,1\%$  (3,29) og multipliserer denne med middelfeilen på differansen, 0,50, får vi grensene for 99,9% konfidensintervallet til å bli 0,16 og 3,46.

Dette intervallet omfatter ikke 0, og dette svarer da helt til at den P vi fant i t-testen var mindre enn 0,1%.



Uavhengige prøver

I poteteksemplet bestemte vi differensen i knollengde ved først å bestemme middellengden i hver bingje, og så ta differensen mellom disse. Vi kunne også bestemt den midlere differens ved å måle en tilfeldig knoll fra hver bingje, ta differensen mellom disse måle-  
ne, og så gjenta dette en rekke ganger.

Forventningene for middeltallet og middelfeilen er like store ved begge metoder forutsatt at prøvestørrelsene er like. I det ene tilfellet bestemmes middelkvadratet på differensen direkte, i det andre ved å legge sammen middelkvadratene for bingje A og B.

Ved å ta en prøve fra hver bingje ville middelfeilen blitt sikrere bestemt enn ved å måle differenser. Sikkerheten i bestemmelsen henger sammen med antall frihetsgrader. Hvis vi f.eks. måler 50 knoller fra hver bingje vil vi få  $49+49=98$  frihetsgrader for feilen, mens vi ved å ta parvise differenser bare får 49 frihetsgrader. For å bestemme differensen ved hjelp av parvise prøver er vi nødt til å måle like mange prøver fra hver bingje.

Hvis vi ikke har reelle parvise observasjoner, har det altså ingen hensikt å beregne middelfeilen på parvise differenser.

I alminnelighet er det mest effektivt å måle det samme antall fra hver bingje. Med de samme omkostninger vil vi få den nøyaktigste bestemmelse av differensen når prøvene er like store. Dette er lett å vise med et talleksempel.

Vi hadde målt 20 knoller i A og 100 i B og feilen på differensen blir:

$$m = \sqrt{MS/20 + MS/100} = \sqrt{0,06 MS} = 0,24 s$$

Hvis vi i stedet hadde målt 10 knoller fra A og 110 fra B, hadde vi fått :

$$m = \sqrt{MS/10 + MS/110} = \sqrt{0,11 MS} = 0,33 s,$$

altså adskillig større.

Minst feil hadde vi fått med det samme antallet i hver prøve:

$$m = \sqrt{MS/60 + MS/60} = \sqrt{0,033 MS} = 0,18 s$$

Lik prøvestørrelse har også andre fordeler som at testen er mer robust ovenfor forutsetningene.

### Parvise prøver

La oss ta et helt annet eksempel. Det sies at mennesker er litt høyere om morgenen enn om kvelden, at vi synker litt sammen i løpet av dagen.

For å bestemme denne differensen er det vel ingen som ville måle 100 mennesker om morgenen og 100 andre om kvelden, og å beregne differensen mellom gjennomsnittstallene morgen og kveld. Det er nokså selvfølgelig at vi ville få en mye nøyaktigere bestemmelse av differensen hvis vi målte de samme personer morgen og kveld. Hvor ligger egentlig forskjellen mellom disse to eksemplene?

I det første tilfellet er det ingen sammenheng mellom lengden av knoll nr 1 i bunge A og knoll nr 1 i bunge B, de er helt uavhengige av hverandre. Høydemålet for samme menneske morgen og kveld viser derimot en sterk sammenheng, de er sterkt korrelerte som det heter i statistikken. Hvis vi prikket målene inn på en figur med morgenmålet i den ene og kveldsmålet i den andre retningen, ville prikkene kommet til å ligge meget nær en rett linje med 45 graders stigning. Det er bare for helt ukorrelerte størrelser at middelkvadratet for en differens er summen av middelkvadratene for de to leddene. Hvis det er positiv sammenheng vil feilen på differensen bli mindre.

Hvis vi har observasjoner som naturlig opptrer parvis, skal vi derfor bestemme de parvise differenser, og så beregne middeltall og konfidensgrenser, eventuelt teste om differensen er forskjellig fra 0.

## 12. Univers, prøver og hypoteser

I statistikken brukes betegnelsen UNIVERS om en stor samling av individer. Universet kan være endelig, som i poteteksemplet. Vi har også uendelige univers, f.eks. avlingen (kg pr dekar) av en bestemt byggsort. Her er vi like interessert i fremtidige avlinger som i de avlinger som vi har hatt, og da har vi ikke noe endelig univers. Men i begge tilfelle er vi interessert i å estimere egenskaper i universet ved hjelp av prøver.

Vi har nå sett hvordan vi ut fra en tilfeldig prøve kan estimere middeltall og standardavvik i et univers, og videre bestemme middelfeilen og konfidensgrenser for forventningen ( $\mu$ ). Gjennomsnittet for to grupper kan sammenlignes ved hjelp av t-test, og ved beregning av konfidensgrenser for den forventede differensen.

### Null-hypotese

De hypotesene vi tester ved hjelp av t-test (og andre signifikanstester som vi skal komme inn på senere) kalles null-hypoteser.

I poteteksemplet er 0-hypotesen at lengdene på potetene i de to bingene er like store, eller for å være eksakt: At knoll-lengdene i de to bingene er tatt ut av identiske normale fordelinger.

Vi kan aldri bevise at en 0-hypotese er riktig. Det en signifikanstest eventuelt kan si oss er at det er svært lite rimelig at den er korrekt.

## 13. Variansanalyse

Ved hjelp av en teknikk kalt variansanalyse kan vi sammenligne flere grupper på en gang. Vi kan teste i hvilken grad forskjellen mellom gruppene sett under ett er signifikant.

Vi vil ved hjelp av et eksempel vise beregningsmåten først, og komme tilbake til teorien etterpå.

Anta at vi har spurt 27 gårdbrukere i Ås hvilken potetsort de dyrket siste år, og hvilken avling de fikk, og at vi har fått som svar de data som er gitt i tabell I.4.

Tabell I.4. potetavlinger i tonn pr dekar.

Sort	Sum	n	$\bar{x}$
A 3,0 3,7 2,1 2,3 2,8 3,5 3,1 2,4 2,7 2,4 2,6 2,3	32,9	12	2,74
B 3,6 3,9 3,1 3,3 3,5 3,2 4,0 2,8 3,0 3,1	33,5	10	3,35
C 3,2 3,5 2,8 2,9 3,0	15,4	5	3,08
Total	81,8	27	3,03

Vi starter variansanalysen ved å beregne fradragsledd og kvadratsum for hele materialet:

$$CT = 81,8^2/27 = 247,82$$

$$SS_{\text{total}} = 3,0^2 + 3,7^2 + \dots - CT = 254,30 - 247,82 = 6,48$$

Dette tall kan vi nå føre inn på første linje i variansanalysen.

Antall frihetsgrader, ialt 26, kan deles opp i to deler, frihetsgrader mellom sorter og frihetsgrader innen sorter. Da vi har 3 sorter får vi 2 frihetsgrader mellom sorter, og resten, 24, innen sorter. Dette siste tallet får vi jo også ved å legge sammen frihetsgrader for hver av de tre sortene. Da vi har henholdsvis 12, 10 og 5 svar, får vi  $11 + 9 + 4 = 24$  frihetsgrader innen sorter.

Beregning av SS mellom sorter blir:

$$\begin{aligned} SS_{\text{sort}} &= 32,9^2/12 + 33,5^2/10 + 15,4^2/5 - CT \\ &= 90,20 + 112,22 + 47,43 - 247,82 = 2,03 \end{aligned}$$

Etter at dette tallet er ført inn i variansanalysen finner vi SS innen sorter ved subtraksjon :  $6,48 - 2,03 = 4,45$ . For kontroll kan vi også regne ut SS innen hver sort, og vil få 2,75 for A, 1,39 for B og 0,31 for C. Summen av disse blir 4,45.

Til slutt regner vi ut MS for hver linje i tabellen ved å dividere SS med DF.

Variansanalyse:

Variasjonsårsak	DF	SS	MS
Total	26	6,48	0,249
Mellom sorter	2	2,03	1,015
Innen sorter=Rest	24	4,45	0,185

#### 14. F-test

Forholdet mellom MS for sort og MS for rest betegnes med F, altså her  $F = 1,015/0,185 = 5,49$ , og denne verdi brukes til å teste om det er noen reell forskjell på avlingstallene for de tre sortene.

Det kan vises at hvis det i k k e er noen forskjell, eller for å være eksakt: hvis de tre grupper av avlingstall er tatt ut tilfeldig av en og samme fordeling, vil de tre MS vi har regnet ut alle være estimater av sigma i denne fordeling. Tallene behøver ikke å bli like, de har jo alle sine tilfeldige feil, og feilen er særlig stor hvis antall frihetsgrader er lite, som her for MS mellom sorter.

Under forutsetning av at denne felles fordelingen er normal, er det regnet ut tabeller som viser hvor stor  $F$  kan bli med ulik sannsynlighet ( $P$ ). Disse  $F$ -tabellene må ha tre innganger:

1. Sannsynligheten ( $P$ )
2. Antall DF for MS i telleren i  $F$
3. Antall DF for MS i nevneren i  $F$

Slår vi opp i en  $F$ -tabell for  $P=5\%$ , vil vi for 2 frihetsgrader og 24 frihetsgrader finne tallet 3,40. Dette vil si at hvis vi i en normal fordeling hadde bestemt to uavhengige estimater av  $\sigma^2$ , er det 5% sannsynlighet for at forholdet mellom et estimat med 2 frihetsgrader og et med 24 frihetsgrader er 3,40 eller større. Det tilsvarende tall i  $F$ -tabellen for  $P=1\%$  er 5,61, og for  $P=0,1\%$  6,95. I vårt eksempel var den beregnete  $F=5,49$ . Og det var derfor bare litt over 1% sannsynlighet for å få en så stor verdi hvis avlingstallene for de tre sortene var trukket tilfeldig ut fra en og samme normale fordeling.

Vår null-hypotese er at de tre sortene gir like stor avling. Når  $F$ -verdien er stor, altså  $P$  liten, sier vi at vi har avvist nullhypotesen, og da må vi anta at den motsatte hypotese, avlingstallene er forskjellige, er den riktige. Vår konklusjon blir at sortene har gitt forskjellig avling. Vi sier at forskjellene mellom sortene er signifikant på 1%-nivået, eller at forskjellen er meget signifikant.

Vi vet på forhånd at null-hypotesen ikke holder. De tre sortene kan jo ikke være absolutt like. I virkeligheten er vi heller ikke interessert i å få vite at de er ulike. Det vi ønsker å få vite er hvilken sort som gir størst avling, eventuelt hvor stor forskjellene er, og konfidensgrensene for denne forskjellen.

Null-hypotesen og  $F$ -testen er et teknisk hjelpemiddel. Hvis vi ikke får signifikans ved  $F$ -testen, kan det enten bety at forskjellene er små og kanskje uten praktisk betydning, eller det kan bety at vårt tallmateriale er for dårlig.

## 15. Forutsetninger for testing av differenser

Hvis vi får en stor F-verdi og derfor liten P, vet vi at det er liten sannsynlighet for å få et slikt resultat hvis prøvene var tatt ut tilfeldig av en normalfordeling. Det ligger da nær å spørre hvordan forholdet hadde vært hvis vi hadde hatt en unormal, f.eks. en meget skjev eller kanskje totoppet fordeling.

Flere undersøkelser viser at dette har liten betydning for signifikansnivået. I de fleste tilfelle betyr det bare at det som vi har angitt som en 5% signifikansgrense kanskje er en 4% eller 6% grense. Følsomheten, eller styrken, av testen kan derimot bli nokså mye dårligere for observasjoner som ikke er normalfordelte.

Vi må være klar over at de grenser for P som vi pleier å bruke for å angi signifikans, 5%, 1% eller 0,1%, ikke er noen magiske grenser, men at de er vilkårlig valgt.

En annen forutsetning for testing av differenser er at observasjonene skal ha lik tilfeldig feil. Enkelte observasjoner med store feil kan endre nivået på testen vesentlig.

I noen tilfeller vet vi på forhånd, eller forsøket viser klart, at forsøksleddene har ulik variabilitet. Et eksempel kan være at vi vil sammenligne 4 ugrasmidler, og så i tillegg har med et ubehandlet ledd. Hvis alle midlene er noenlunde effektive, og hvis det er mye ugras tilstede, vil variasjonen mellom parallellrutene av ubehandlet være mye større enn variasjonen mellom de behandlede rutene. MS-rest i variansanalysen blir derfor altfor stor når vi vil sammenligne ugrasmidlene innbyrdes, og for liten når vi vil sammenligne ubehandlet med behandlet.

## 16. Gjennomsnitt og feil på differenser

Variansanalysen sier oss at det var forskjeller mellom de tre sortene. Den sier ikke noe om hvilken sort som gir størst avling, eller om hvilke sorter som er signifikant forskjellige. For å få

greie på det må vi se på gjennomsnittstallene, differensene mellom disse, og feilene på disse differensene.

I vårt eksempel (s 19) var det de to sortene (A og B) som hadde flest svar som samtidig hadde det høyeste og det laveste gjennomsnitt. Et signifikant resultat av variansanalysen kan da uten videre tolkes slik at det er signifikant forskjell på A og B. Om C er signifikant forskjellig fra A og/eller B sier F-testen ikke noe om.

Vi bruker  $MS_{rest}$  som et estimat for  $\sigma^2$ , idet vi går ut fra at det er samme variasjonen innen de ulike gruppene.

I eksemplet kan vi regne ut tre forskjellige differenser, og middelfeil og t-test for hver av disse (tabell I.5.).

Tabell I.5. Differenser og feil på differenser i avling mellom 3 potetsorter.

Sort	diff.	$m_d = \sqrt{MS_{rest}/n_1 + MS_{rest}/n_2}$	t
B-A	0,61	0,18	3,4
C-A	0,34	0,23	1,5
B-C	0,27	0,24	1,1

Det er bare differensen mellom B og A som er signifikant. Det kan ikke sies noe bestemt om hvor C ligger i forhold til A eller B.

Vi skal senere komme inn på de innvendinger som kan reises mot å teste alle differensene mot en feil beregnet ved hjelp av en felles  $MS_{rest}$ .



## 17. Toveis variansanalyse

Den variansanalysen vi har sett på hittil er en enveis analyse. Materialet kan grupperes i en retning, og den totale variasjonen kan deles i to deler, en del mellom grupper og en del innen grupper. Vi har også eksempler på at materialet kan grupperes i to eller flere retninger samtidig, og da må vi bruke to eller flerveis variansanalyser.

Tabell I.6. Avlingsdata for 3 potetsorter fra 5 ulike gårder.

Gård	S o r t			Sum	$\bar{x}$
	A	B	C		
1	3,0	3,6	3,2	9,8	3,27
2	3,7	3,9	3,5	11,1	3,70
3	2,1	3,1	2,8	8,0	2,67
4	2,3	3,3	2,9	8,5	2,83
5	2,8	3,5	3,0	9,3	3,10
Sum	13,9	17,4	15,4	46,7	
$\bar{x}$	2,78	3,48	3,08		3,11

Som eksempel på toveis variansanalyse kan vi bruke en del av de samme tall for potetavlinger som vi har brukt før (s 19). La oss forutsette at det ikke var avlingstall fra 27 ulike gårder, men bare fra 12, idet hver kollonne i tabellen var fra samme gård. De første 5 gårdene hadde dyrket alle 3 sorter, de neste 5 gårder sort A og B, og de 2 siste gårdene bare sort A.

Vi skal analysere tallene fra de 5 gårdene som hadde dyrket alle tre sortene. Avlingstallene er satt opp i tabell I.6, og det er regnet ut summer og middeltall for hver gård.

Som før begynner vi med å regne ut fradragsleddet og kvadratsum for hele materialet:

$$CT = 46,7^2/15 = 145,39 \text{ og}$$

$$SS_{\text{tot}} = 3,0^2 + 3,7^2 + \dots - CT = 148,89 - 145,39 = 3,50$$

Denne totale kvadratsum med 14 frihetsgrader kan nå deles opp i tre deler, en del mellom gårder med 4 frihetsgrader, en del mellom sorter med 2 frihetsgrader, og en rest med 8 frihetsgrader.

SS mellom gårder beregnes slik:

$$(9,8^2 + 11,1^2 + 8,0^2 + 8,5^2 + 9,3^2)/3 - CT = 1,94$$

På tilsvarende måte er SS mellom sorter:

$$(13,9^2 + 17,4^2 + 15,4^2)/5 - CT = 1,24$$

Det er samme framgangsmåte som ved beregning av SS mellom grupper i en enveis variansanalyse, men da vi her har det samme antall i alle grupper, kan vi addere kvadratene før vi dividerer med antallet. Vi sparer arbeide og får mindre avrundingsfeil. Resultatene samles i en variansanalysetabell,  $SS_{\text{rest}}$  regnes ut ved subtraksjon, og  $MS=SS/DF$ .

Variasjonsårsak	DF	SS	MS	F
Total	14	3,50		
Mellom gårder	4	1,94	0,485	
Mellom sorter	2	1,24	0,620	15,5**
REST(=samspill gård x sort)	8	0,32	0,040	

Til slutt beregnes en F-verdi mellom sorter, med  $MS_{\text{rest}}$  som nevner. F-verdien ligger godt over F-tabellens verdi for  $P=1\%$ , og nær  $0,1\%$  verdien. Dette er angitt ved at det er satt to stjerner bak F-verdien. (En stjerne for  $5\% > P > 1\%$ , to stjerner for  $1\% > P > 0,1\%$  og tre stjerner for  $P < 0,1\%$ ).

I vårt eksempel er det meget signifikant forskjell mellom de tre sortene.

Resten i variansanalysetabellen er også kalt samspillet mellom de to grupperinger. Hvis forskjellen mellom SORTENE var den samme på alle gårdene ville  $SS_{rest}$  bli 0.

I praksis vil aldri  $SS_{rest}$  bli 0 på grunn av tilfeldig variasjon som skyldes jordvariasjon, målefeil, skrivefeil, ulike angrep av plantesjukdommer og insekter osv. Det kan også tenkes at sortene virkelig reagerer forskjellig på vekstvilkårene på de 5 gårdene.

La oss se hvordan analysen av de samme 15 avlingstallene hadde blitt hvis vi ikke hadde fått den nye opplysningen om at tallene stammet fra 5 forskjellige gårder og ikke fra 15 forskjellige. Uten denne opplysningen hadde vi ikke kunnet trekke fra variasjonen mellom gårder. Resten hadde følgelig hatt 12 DF, med

$$SS_{rest} = 1,94 + 0,32 = 2,26,$$

$$MS_{rest} = 2,26/12 = 0,188 \text{ og}$$

$$F = 0,62/0,188 = 3,30.$$

Vi har da flere frihetsgrader for feilen, men F-verdien er nå ikke signifikant engang på 5 % nivået. Middelfeilen på gjennomsnittene blir over dobbelt så stor som etter en toveis analyse. I dette eksemplet har vi en reell forskjell på avlingsnivået på de ulike gårdene. Denne variasjonen fører ikke til noen øket usikkerhet for sortssammenligningene når vi sammenligner sortene på samme gård.

Forskjellen mellom en enveis og en toveis variansanalyse svarer til forskjellen mellom ikke parvis og parvis t-test.

## 18. t-test eller F-test

Hvis vi bare har to grupper, kan vi velge om vi vil bruke t-test eller variansanalyse, de vil gi nøyaktig samme signifikans. I variansanalysen får vi 1 frihetsgrad mellom grupper, og den beregnede F-verdien blir kvadratet av den beregnede t-verdien i t-testen. Hvis vi sammenligner en F og en t-tabell, vil vi finne at tabellens F-verdi med 1 frihetsgrad for MS i telleren er lik kvadratet av t-verdien for samme P.

## 19. Feilen på differenser mellom middeltall i toveis analyse

I en toveis variansanalyse har vi det samme antall i hver gruppe. Derfor får vi også den samme feil på alle differenser. I eksemplet blir middelfeilen på en differens mellom to middeltall

$$m_d = \sqrt{0,040/5 + 0,040/5} = 0,13 = \sqrt{2MS_{rest}/r}$$

hvor r er antall gjentak (i vårt eksempel antall gårder).

## 20. Minste signifikante forskjell (LSD)

Vi kan regne ut hvor stor en differens må være for å være signifikant. Denne størrelsen blir kalt LSD, Least Significant Difference, og den beregnes ved å multiplisere middelfeilen på differensen med t-tabellens tall for vedkommende signifikansnivå og det antall DF som feilen er bestemt med. I vårt eksempel har vi 8 DF for feilen,  $MS_{rest}$ . Tabellens t-verdi for P=5% er da 2,31 og for P=1% 3,36. Disse tallene multipliseres med middelfeilen (0,13) og vi får:

$$LSD_{5\%} = 0,32 \text{ og } LSD_{1\%} = 0,43$$

De tre differensene er:  $B-C=3,48-3,08= 0,40$   
 $C-A=3,08-2,78= 0,30$   
 $B-A=3,48-2,78= 0,70$

Etter dette er altså forskjellen mellom B og A meget signifikant, forskjellen mellom B og C signifikant, mens forskjellen C A er knapt signifikant.

## 21. Variasjonskoeffisienten

Som et mål på hvor nøyaktig et forsøk er, beregner vi ofte variasjonskoeffisienten som er standardavviket i prosent av gjennomsnittet. På engelsk heter denne størrelsen Coefficient of Variation, og forkortes CV%

$$CV\% = \frac{\sqrt{MS_{rest}} \cdot 100}{\bar{x}}$$

Variasjonskoeffisienten varierer med ulike typer av observasjoner og er forskjellig for f.eks. avlingstall, legdeprosjenter, sjukdomsangrep osv. For å ha noen nytte av variasjonskoeffisienten må vi derfor ha et visst erfaringsmateriale og sammenligne med. Hvis  $\bar{x}$  er nær 0 blir CV% svært stor. I slike tilfeller har det ingen hensikt å beregne CV%.

## 22. Flerveis variansanalyser

Avlingstallene fra det faktorielle forsøket som er diskutert på side 47 danner en tre-veistabell. Tabellen inneholder alle kombinasjoner av 4 blokker, 3 sorter og 2 gjødslinger.

For å gjøre fremstillingen mer generell, kaller vi de tre faktorene A, B og C, og sier at vi har a trinn av A, b trinn av B og c trinn av C. I alt har vi altså a.b.c tall, og disse kan vi sette opp i en tre-veistabell på 3 ulike måter. I tabell I.7. har vi laget en toveistabell over A og B for hver verdi av C (C1 og C2). Vi kunne selvfølgelig like godt hatt en toveistabell over A og C for hver verdi av B, eller for B og C for hver verdi av A.

Tabell I.7. Tre-veistabell for faktorene A, B og C.

	C <sub>1</sub>				Sum over		C <sub>2</sub>				Sum over
	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>	A		A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>	A
B <sub>1</sub>	x <sub>111</sub>	x <sub>211</sub>	x <sub>311</sub>	x <sub>411</sub>	x <sub>.11</sub>	B <sub>1</sub>	x <sub>112</sub>	x <sub>212</sub>	x <sub>312</sub>	x <sub>412</sub>	x <sub>.12</sub>
B <sub>2</sub>	x <sub>121</sub>	x <sub>221</sub>	x <sub>321</sub>	x <sub>421</sub>	x <sub>.21</sub>	B <sub>2</sub>	x <sub>122</sub>	x <sub>222</sub>	x <sub>322</sub>	x <sub>422</sub>	x <sub>.22</sub>
B <sub>3</sub>	x <sub>131</sub>	x <sub>231</sub>	x <sub>331</sub>	x <sub>431</sub>	x <sub>.31</sub>	B <sub>3</sub>	x <sub>132</sub>	x <sub>232</sub>	x <sub>332</sub>	x <sub>432</sub>	x <sub>.32</sub>
Sum over	x <sub>1..1</sub>	x <sub>2..1</sub>	x <sub>3..1</sub>	x <sub>4..1</sub>	x <sub>..1</sub>		x <sub>1..2</sub>	x <sub>2..2</sub>	x <sub>3..2</sub>	x <sub>4..2</sub>	x <sub>..2</sub>

Tallene inne i tabellen, f.eks.  $x_{231}$  står for avlingen av det 2. trinn av A, det 3. trinn av B og det 1. trinn av C. Generelt får et enkelt tall betegnelsen  $x_{ijk}$ . Tabellene er summert begge veier. Vi bruker vi en . istedenfor indeksen for å symbolisere sum over vedkommende faktor. De to sumkolonnene danner sammen en toveistabell for B og C, mens de to sumrekkene danner en toveistabell for A og C. Summene i nedre høyre hjørne ( $x_{..1}$  og  $x_{..2}$ ) er enveistabellen for C. Endelig kan vi summere tabellene for de ulike verdier av C, og får da tabell I.8.

Tabell I.8. Toveistabell for A og B summert over C.

	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>	Sum over C og A
B <sub>1</sub>	x <sub>11.</sub>	x <sub>21.</sub>	x <sub>31.</sub>	x <sub>41.</sub>	x <sub>.1.</sub>
B <sub>2</sub>	x <sub>12.</sub>	x <sub>22.</sub>	x <sub>32.</sub>	x <sub>42.</sub>	x <sub>.2.</sub>
B <sub>3</sub>	x <sub>13.</sub>	x <sub>23.</sub>	x <sub>33.</sub>	x <sub>43.</sub>	x <sub>.3.</sub>
Sum over C og B	x <sub>1..</sub>	x <sub>2..</sub>	x <sub>3..</sub>	x <sub>4..</sub>	x <sub>...</sub>

Marginalsummene i denne tabellen er enveistabellene for A og B.  $x_{...}$  er totalsummen.

Det er tidligere vist (s 25) at variasjonen i en toveistabell over A og B kan deles i tre deler, en hovedeffekt av A, en hovedeffekt av B, og en rest, som også kan kalles samspillet AB. På lignende måte kan variasjonen i en treveistabell deles i 7 deler, de tre hovedeffektene A, B og C, tre 2-faktorsamspill AB, AC og BC, og en rest som da blir 3-faktorsamspillet ABC.

Beregningen av kvadratsummene for de enkelte deler er slik:

$$\text{Korreksjonsledd} = CT = X^2_{\dots} / abc$$

$$\text{Kvadratsum for A} = SS_A = \Sigma X^2_{i..} / bc$$

$$\text{Kvadratsum for B} = SS_B = \Sigma X^2_{.j.} / ac - CT$$

$$\text{Kvadratsum for C} = SS_C = \Sigma X^2_{..k} / ab - CT$$

$$\text{Kv.sum for komb. A og B} = SS_{A\&B} = \Sigma X^2_{ij.} / c - CT$$

$$\text{Kv.sum for komb. A og C} = SS_{A\&C} = \Sigma X^2_{i.k} / b - CT$$

$$\text{Kv.sum for komb. B og C} = SS_{B\&C} = \Sigma X^2_{.jk} / a - CT$$

$$\text{Kv.sum for samspill AB} = SS_{AB} = SS_{A\&B} - SS_A - SS_B$$

$$\text{Kv.sum for samspill AC} = SS_{AC} = SS_{A\&C} - SS_A - SS_C$$

$$\text{Kv.sum for samspill BC} = SS_{BC} = SS_{B\&C} - SS_B - SS_C$$

$$\text{Total kvadratsum} = SS_{\text{total}} = \Sigma X^2_{ijk} - CT$$

$$\text{Restkvadratsum} = SS_{\text{rest}} = SS_{\text{total}} - SS_A - SS_B - SS_C - SS_{AB} - SS_{AC} - SS_{BC}$$

For tabeller med større dimensjon enn 3 (4-veis, 5-veis o.s.v.) blir prinsippet for oppdelingen av variasjonen den samme. En 4-veistabell gir fire hovedeffekter, seks 2-faktorsamspill, fire 3-faktorsamspill og et 4-faktorsamspill, 15 deler ialt, mens en 5-veistabell gir fem hovedeffekter, ti 2-faktor, ti 3-faktor, fem 4-faktor og et 5-faktorsamspill, 31 deler ialt.

## Variasjonsanalyse for en tre-veistabell

Variasjonsårsak	DF	SS
A	a-1	$SS_A$
B	b-1	$SS_B$
C	c-1	$SS_C$
AB	(a-1)(b-1)	$SS_{AB}$
AC	(a-1)(c-1)	$SS_{AC}$
BC	(b-1)(c-1)	$SS_{BC}$
ABC	(a-1)(b-1)(c-1)	$SS_{rest}$

Hvis vår 3-veistabell stammer fra et blokkforsøk med 2 forsøksfaktorer med alle kombinasjoner tilfeldig fordelt i hver blokk, går vi i almindelighet ut fra at de tre samsplillene med blokk bare er uttrykk for den tilfeldige variasjon, og vi slår dem derfor sammen til en felles feil. For eksemplet på side 47 får vi da:

Var.årsak	DF
Blokker (A)	3
Ledd (B&C)	5
som deles opp:	
Sorter (B)	2
Gjødsling (C)	1
Sort x Gj (BC)	2
Rest (feil)	15

Resten består her av de 3 samspillene, Blokk x Sort, Blokk x Gjødsling og Blokk x Sort x Gjødsling, med henholdsvis 6, 3 og 6 DF, og vi tester begge hovedeffektene og samspillet mot denne feilen.



I noen tilfeller er det grunn til å tro at det er et reelt samspill mellom en av forsøksfaktorene og blokker. Hvis f.eks. blokkene i det foregående eksempel var svært forskjellige med hensyn til næringsinnhold var det rimelig å tro at gjødsel-effekten varierte fra blokk til blokk. I så fall vil samspillet mellom gjødsling og blokk være større enn de to andre samspillene med blokk, og vi bør dele opp feilen i to deler, en feil med 3 DF på gjødslingen, og en feil (resten) med 12 DF på sorter og samspill sorter x gjødsling.

## II. STATISTISKE MODELLER

I dette avsnittet skal vi diskutere modeller og varianskomponenter for en del forsøksplaner, og begrepene faste (fixed) og tilfeldige (random) variabler.

### 1. Fullstendig tilfeldig fordeling

For den aller enkleste forsøksplanen, fullstendig tilfeldig fordeling, er modellen for en enkeltobservasjon:

$$X_{ij} = \mu + a_i + E_{ij}$$

$X_{ij}$  er den observerte verdi for forsøksledd nr  $i$  og for gjentak, eller rettere, parallell nr  $j$ .  $i$  har verdiene 1, 2 .. opp til  $t$  (antall forsøksledd), mens  $j$  har verdiene 1, 2 .. til  $r$  (hvis det er  $r$  paralleller for alle forsøksledd, eller til  $r_i$  hvis antall paralleller varierer. Modellen sier altså at den observerte verdi kan oppfattes som en sum av 3 ledd, nemlig et totalt middel ( $\mu$ ), et avvik ( $a_i$ ) som skyldes vedkommende forsøksledd, og en tilfeldig variabel  $E_{ij}$ . Hvis vi definerer  $\mu$  som gjennomsnittet av forsøksleddenes sanne verdier, vil summen av  $a$ -ene bli null. Både  $\mu$  og alle  $a$ -ene er det som betegnes som faste (fixed) variabler, mens  $E$ -ene er tilfeldige (random) variabler.

### Faste og tilfeldige variabler

Det er viktig å forstå forskjellen mellom faste og tilfeldige variabler, og vi skal prøve å forklare den med noen eksempler. Hvis vi er interessert i å bestemme forskjellen mellom de enkelte trinn av en variabel er den fast. Hvis vi derimot bare er interessert i variasjonen mellom trinnene er den tilfeldig. Forsøksledd er i almindelighet en fast variabel. Vi er enten interesserte i å bestemme forskjeller mellom leddene (kvalitative faktorer) eller i å finne det optimale nivået av en forsøksfaktor. I begge tilfeller har vi sjøl bestemt forsøksleddene, de er ikke trukket tilfeldig ut av et univers.

Et eksempel på forsøksledd som tilfeldig variabel er at vi i et forsøk sammenligner en rekke genotyper som er tatt ut tilfeldig av en krysningspopulasjon. Det vi da er interesserte i er ikke å bestemme forskjellen mellom de enkelte genotypene, men å bestemme variasjonen mellom dem for å kunne si noe om hele populasjonen. Genotypene er da tilfeldige utvalg av populasjoner.

Blokker er alltid en tilfeldig variabel. Vi er ikke interesserte i å bestemme forskjeller mellom blokkene. Vi er interesserte i å bestemme effektene av forsøksledd på et område, og blokkene skal være et tilfeldig utvalg av dette området. Hvis vi har forsøk på en rekke gårder, er gårder også vanligvis en tilfeldig variabel. Vi er ikke interesserte i å bestemme effekten av forsøksledd på de bestemte gårdene hvor forsøkene er plassert. Gårdene må betraktes som et tilfeldig utvalg av gårder i et område, og det er for dette området vi er interesserte i å bestemme effekten av forsøksledd.

I modellen vil vi bruke store bokstaver for tilfeldige variabler, og små for faste. På grunn av vår definisjon av  $\mu$  er gjennomsnittet av en fast variabel eksakt 0. En tilfeldig variabel må tenkes tatt ut fra en populasjon med middel lik 0. Forventningen av en tilfeldig variabel er derfor 0, men gjennomsnittet er ikke eksakt lik 0. Fordelingen av en tilfeldig variabel har en varians som vi er interesserte i å estimere. En fast variabel har ikke noen slik varians, men vi kan definere en lignende verdi som vi kaller kapp (K):

$$(K)^2 = \Sigma a_i^2 / (t-1)$$

### Varianskomponenter

I variansanalysen for et forsøk med fullstendig tilfeldig fordeling, kan det vises at forventningen av de to MS er:

Var.årsak	DF	MS	Forventning
Total	tr-1	--	
Ledd	t-1	MS1	$\sigma^2 + r(K)^2$
Rest	t(r-1)	MS2	$\sigma^2$

Estimatet av  $\sigma^2$  er MS2, mens estimatet av  $(K)^2$  er  $(MS1 - MS2)/r$ .  $\sigma^2$  og  $r(K)^2$  er komponenter av MS1.

F-testen  $MS1/MS2$  blir en test på om  $(K)$  er forskjellig fra 0, det vil si om noen av  $a$ -ene er forskjellig fra 0. Hvis MS1 er mindre enn MS2 får vi et negativt estimat av  $(K)^2$ , og dette må selvfølgelig skyldes tilfeldig feil. Summen av kvadrattall kan aldri bli negativ!

Hvis forsøksledd er en tilfeldig variabel kan vi estimere standardavviket i denne fordelingen ved å erstatte  $\kappa$  i ligningene ovenfor med standardavviket.

## 2. Blokkforsøk

Modellen for en enkeltobservasjon i et blokkforsøk er:

$$X_{ij} = \mu + a_i + B_j + E_{ij}$$

Det nye her er leddet  $B_j$ , effekten av blokk nr  $j$ . I variansanalysen får vi følgende forventninger for de tre MS:

Var.årsak	DF	MS	Forventning
Total	$tr-1$	--	
Blokk	$r-1$	MS1	$\sigma^2 + t\sigma_B^2$
Ledd	$t-1$	MS2	$\sigma^2 + r(K)_a^2$
Rest	$(t-1)(r-1)$	MS3	$\sigma^2$

F-testen  $MS2/MS3$  er en test på om leddene er forskjellige, det vil si om  $(K)_a$  er forskjellig fra 0. At  $(K)_a$  er forskjellig fra 0 betyr at  $a_i$  er forskjellig fra 0, eller med andre ord at forsøksleddene er forskjellige.

Vi kan også gjøre F-testen  $MS1/MS3$  for å teste om blokkene er forskjellige, men som nevnt foran er vi vanligvis ikke interessert i dette spørsmålet. Variansanalysemodellen illustrerer godt

effekten av blokker. Hvis det ikke er noen forskjell mellom blokkene, det vil si, hvis  $\sigma_B = 0$ , oppnår vi bare å redusere antall frihetsgrader for rest ved å anlegge forsøket som blokkforsøk i stedet for fullstendig tilfeldig fordeling. Er det derimot forskjeller mellom blokkene,  $\sigma_B > 0$ , vil vi oppnå en reduksjon av restkvadratsummen.

Hvis MS1 eller MS2 er mindre enn MS3 må det skyldes tilfeldigheter.

### 3. Latinsk kvadrat

Modellen for en enkeltobservasjon i et Latinsk kvadrat er:

$$X_{ijk} = \mu + a_i + R_j + K_k + E_{ijk}$$

Her er  $a_i$  effekten av forsøksledd,  $R_j$  og  $K_k$  effekter av rekker og kolonner. Både  $i$ ,  $j$  og  $k$  går fra 1 til  $t$ . Variansanalysen blir:

Var.årsak	DF	MS	Forventning
Total	$t^2 - 1$	---	
Rekker	$t - 1$	MS1	$\sigma^2 + t\sigma_R^2$
Kolonner	$t - 1$	MS2	$\sigma^2 + t\sigma_K^2$
Ledd	$t - 1$	MS3	$\sigma^2 + t(K)_a^2$
Rest	$(t-1)(t-2)$	MS4	$\sigma^2$

Her har vi ført inn to tilfeldige variabler,  $R$  for rekker og  $K$  for kolonner. F-testen MS3/MS4 er en test på om forsøksleddene er forskjellige.

### 4. Faktorielle forsøk

Modellen og forventningene for de ulike MS i variansanalysen vil være forskjellig om faktorene betraktes som faste eller tilfeldige variabler. La oss som eksempel ta et forsøk med tre ulike

gjødslingsstyrker og 4 sorter. Gjødslingene er uten tvil en fast variabel, de 3 trinn kan ikke betraktes som et tilfeldig utvalg av alle mulige gjødslinger. Sortene derimot kan betraktes på begge måter. De er faste variabler hvis vi er interessert i å få greie på forskjellen mellom disse 4 sortene, og gjødslingseffekten til hver av dem. Men sortene kan også betraktes som et utvalg av alle mulige sorter, og at hensikten med forsøket er å få greie på gjødslingseffekten til arten i sin alminnelighet. Variasjonen mellom gjødslingseffekten til de 4 sortene blir da en del av feilen på gjødsling.

Hvis vi har  $r$  blokker, hver med  $mn$  forsøksledd ( $m$  gjødslings-trin,  $n$  sorter), og vi regner med fast sortseffekt, bli modellen:

$$X_{ijk} = \mu + B_i + g_j + s_k + (gs)_{jk} + E_{ijk}$$

$B$  er som før blokkeeffekten, med  $i$  fra 1 til  $r$ ,  $g$  er gjødslingseffekten, med  $j$  fra 1 til  $m$ , og  $s$  er sortseffekten, med  $k$  fra 1 til  $n$ . Det nye i modellen er samspillet  $(gs)$ , den verdi som modellen må korrigeres med fordi gjødslingseffekten ikke er den samme for alle sorter. Tilslutt kommer så en tilfeldig variabel  $E$ , en verdi for hver rute.

Hvis vi regner med tilfeldig sorteffekt blir modellen tilsvarende. Vi bare erstatter  $s$  med  $S$ . Det er først når vi kommer til forventningene for de ulike MS i variansanalysen at forskjellen mellom de to modellene viser seg.

Hvis sortene betraktes som tilfeldige, må gjødslingseffekten (MS2) testes mot samspillet gjødsling x sort (MS4). Dette har to komponenter, nemlig det sanne samspill  $\sigma_{gs}^2$  og  $\sigma^2$ , den tilfeldige feilen. Hvis sortene er faste forsvinner samspillet i forventningen for MS for gjødsling. Under denne forutsetningen må alle tre forsøkseffektene, gjødsling, sort og samspillet gjødsling x sort testes mot restvariansen, MS5.

Forventning for MS:

Var.

årsak	DF	MS	Sort tilfeldig	Sort fast
Bl.	r-1	MS1	$\sigma^2 + nm \sigma_B^2$	$\sigma^2 + nm \sigma_B^2$
Gj.	m-1	MS2	$\sigma^2 + r \sigma_{g_s}^2 + rn(K)_g^2$	$\sigma^2 + rn(K)_g^2$
Sort	n-1	MS3	$\sigma^2 + rm \sigma_s^2$	$\sigma^2 + rm(K)_s^2$
GxS	(m-1)(n-1)	MS4	$\sigma^2 + r \sigma_{g_s}^2$	$\sigma^2 + r(K)_{g_s}^2$
Rest	rest	MS5	$\sigma^2$	$\sigma^2$

### Split plot

I et split plot forsøk regner vi med å ha to ulike tilfeldige feil, en feil på storruter og en på småruter. Hvis vi med det samme eksemplet som foran har gjødslingsalternativene på storruter og sortene på småruter blir modellen:

$$x_{ijk} = \mu + B_i + g_j + ES_{ij} + s_k + (gs)_{jk} + Es_{ijk}$$

ES er feilen på storruter og Es på småruter, og vi regner sort som en fast variabel. Variansanalysen blir da:

Var.årsak	DF	MS	Forventning
Blokker	r-1	MS1	$\sigma_{E_s}^2 + mn \sigma_B^2$
Gjødsel	(m-1)	MS2	$\sigma_{E_s}^2 + rn(K)_g^2$
feil(a)	(r-1)(m-1)	MS3	$\sigma_{E_s}^2$
Sort	(n-1)	MS4	$\sigma_{E_s}^2 + rm(K)_s^2$
GjxSo	(m-1)(n-1)	MS5	$\sigma_{E_s}^2 + r(K)_{g_s}^2$
feil(b)	m(r-1)(n-1)	MS6	$\sigma_{E_s}^2$

Gjødsling må testes mot feil(a), sort og samspill mot feil(b).

## III FORSØK ELLER SURVEY

Det er i hovedsaken to systematiske måter å samle erfaringer på:

- 1) Innsamlings- eller survey metoden.
- 2) Forsøks- eller eksperiment metoden.

### 1. Survey

I en survey samler vi inn opplysninger i et materiale som allerede finnes, og vi prøver så ved statistiske metoder å finne sammenhengen mellom de ulike variablene. Som oftest betrakter vi noen av variablene som årsak og andre som virkning.

Eksemplene og oppgavene som er gjennomgått i forrige kapittel er alle av surveytypen. Visse spørsmål er det umulig - eller i alle fall svært vanskelig - å belyse gjennom forsøk. Da må vi bruke surveymetoden. For spørsmål som lar seg løse gjennom forsøk er det å foretrekke å bruke forsøk.

I eksemplet med potetsorter i Ås hadde vi kommet til at det var signifikant forskjell på sortene, og at sort B ga større avling enn A. Er dette et tilstrekkelig grunnlag for å anbefale dyrkning av sort B fremfor sort A i Ås? (Vi ser bort fra alle andre egenskaper enn avlingen, og at tallene bare var fra et år, m.m.).

Sort B kan være en ny sort som bare de fremste bøndene har fått tak i. Disse bøndene gjødsler sterkere og setter tidligere enn andre. Årsaken til at sort B har gitt større avlinger enn sort A kan være forskjeller i dyrkingsteknikk og ikke reell sortsforskjell.

Selv om A og B er dyrket på samme gård (eksemplet på toveis variansanalyse), kan det tenkes at det er andre årsaker til avlingsforskjellen enn selve sortforskjellen.

Det kan f.eks. være kjent, eller antatt, at sort B tåler sterkere gjødsling enn sort A. Derfor kan det på alle de 5 gårdene



være brukt større gjødselmengder til B enn til A, og dette kan være årsaken til forskjellen i avling. Eller sort A kan være litt tidligere enn B. Den er blitt tatt opp først, og har av den grunn gitt minst avling.

Hvis vi sammenligner sortene i et forsøk, vil vi sørge for at andre faktorer ikke virker inn. Eventuelt sammenligner vi sortene både ved sterk og ved svak gjødsling, og kanskje med ulike høstetider for alle sortene.

Rent generelt gjelder det at vi i et survey-materiale ikke med sikkerhet kan vite hva som er årsak og hva som er virkning. Det kan være at det vi tar for en årsakssammenheng bare skyldes at begge faktorer henger sammen med en tredje faktor (gjødsling, høstetid).

I et skikkelig planlagt og utført forsøk er det aldri noen tvil om årsak og virkning. I en survey kan vi selvfølgelig notere gjødsling, høstetid etc. og prøve å ta dette med i vurderingene. For det ene vil dette komplisere arbeidet sterkt, og for det andre kan vi aldri få eliminert virkningen av forhold som vi ikke har vært oppmerksomme på, og som vi derfor ikke har notert noe om.

Som nevnt er det visse spørsmål som det vanskelig kan utføres forsøk med. Hvis vi vil vite om det er nyttig for vordende bønder å gå på landbruksskole, kan vi ikke plukke ut en gruppe vordende bønder, la halvparten gå på landbruksskole og halvparten ikke, og så sammenligne resultatet av deres drift senere. Vi kan imidlertid undersøke drifta hos bønder med landbruksskole og bønder uten landbruksskole.

En fordel med surveymetoden er at vi med den i visse fall kan få med oss opplysninger bakover i tid. I skogbruksforskningen kan vi f.eks. ved å måle årringer og avstanden mellom kvistkransene få greie på hvordan trærne har vokst, kanskje i de 50-100 siste årene. Den tidligere tettheten kan vi få et mål på ved å telle opp stubbene, og ved årringstudier kan vi finne ut når de enkelte trær var felt, i alle fall en del år bakover.

Når vi har opplysninger om klimaet kan vi ved hjelp av survey allerede i dag få greie på langsiktige sammenhenger som vi kanskje måtte vente 50 år på å få hvis vi var henvist til å studere dette gjennom forsøk.

## 2. Forsøk

I et forsøk varierer vi selv årsaksfaktorene på en måte som vi har bestemt på forhånd, og vi registrerer virkningen på ulike måter.

Målet for et godt forsøk er å få svar på de spørsmålene vi er interesserte i, og at disse svarene skal bli så nøyaktige som mulig. Samtidig bør vi kunne vurdere hvor nøyaktige de er, helst sette opp tallmessige grenser - konfidensgrenser - for svarene, og vi bør også kunne si under hvilke betingelser svarene er gyldige. Endelig bør omkostningene ved forsøket være så små som mulig.

Det er mange forhold som bidrar til nøyaktige forsøksresultater. Noe av det viktigste er å ha minst mulig ukontrollert variasjon i forsøksmaterialet. Dette vil bli drøftet nærmere i kapitlet om forsøksplaner.

Det er viktig å unngå grove feil slik som ombytting av behandlinger eller av resultater eller grove noteringsfeil. Videre må vi arbeide nøyaktig og bruke teknisk gode metoder. Hvis de beste metodene, de som gir minst feil, samtidig er dyrere eller mer arbeidskrevende enn andre mindre nøyaktige metoder, blir det en vurderingssak hvor stor vekt vi skal legge på dette punktet i forhold til andre måter å redusere feilen på.

Det er selvfølgelig at vi må ta hensyn til omkostningene når vi vil sammenligne ulike opplegg for et forsøk. Hvis to alternativer koster det samme, er det beste alternativet det som gir de nøyaktigste resultatene. Hvis de er like nøyaktige, er det billigste alternativet best.

Korrekt feilberegning forutsetter at feilene er tilfeldige. Den variasjon vi støter på i et forsøksmateriale er sjelden eller aldri tilfeldig. I et markforsøk vil fruktbarheten ofte forandre seg systematisk fra den ene kanten av et forsøksfelt til den andre, eller vi har flekker eller striper som er anderledes enn jorda omkring. Vi gjør slik systematisk variasjon om til tilfeldig variasjon ved å fordele forsøksleddene tilfeldig.

## IV FORSØKSLEDD/FORSØKSFAKTOR

Når vi vil finne ut noe om et biologisk, økonomisk eller teknisk fenomen ved hjelp av forsøk, må vi definere de forsøksleddene vi skal teste. Dette kan være mer eller mindre vanskelig avhengig av det problemkomplekset vi vil lære noe om. Det kan også lett oppstå noe begrepsforvirring, og vi har derfor valgt å ta med en diskusjon av begrepene forsøksledd, forsøksfaktor, forsøkstrinn og forsøksspørsmål.

Ofte blir forsøksledd og forsøksspørsmål brukt som synonyme begreper, men forsøksledd er den mest konkrete betegnelsen på det materialet vi anlegger forsøk med. Forsøksspørsmål blir også brukt mer abstrakt om det fenomenet vi ønsker å finne ut noe om.

Det er lettest å forklare begrepene ved hjelp av eksempler. I et sortsforsøk er forsøksspørsmålet hvilken sort som egner seg best i et visst område, eller i et bestemt dyrkingssystem. Forsøksleddene er de ulike sortene som er med i forsøket.

Hvis vi vil undersøke flere ting samtidig, må vi innføre begrepet forsøksfaktor. I et kombinert forsøk med sorter og gjødslinger er sorter og gjødsling forsøksfaktorer. De ulike sortene er trinn av faktoren sort, og de ulike mengdene er trinn av faktoren gjødsling. Kombinasjonene av gjødseltrinn og sorter er forsøksledd.

Begrepet trinn er språklig sett mest korrekt brukt på forsøksfaktorer som varierer i mengde eller konsentrasjon. Hvis forsøksfaktoren er gjødsling eller sprøyting, kan trinnene være mengder eller konsentrasjoner, men trinnene kan også være ulike tidspunkter for gjødsling eller sprøyting, eller ulike gjødselslag og ulike plantevernmidler.

I forsøk med bare en faktor er forsøksledd og trinn det samme.

Valg av forsøksspørsmål er et emne som må avgjøres innen de enkelte fag, men vi kan gjerne understreke betydningen av å

stille fornuftige og vesentlige spørsmål.

Vi bør tenke grundig igjennom om de forsøksleddene vi tester virkelig svarer til de spørsmålene vi er interesserte i.

I et sortsforsøk er vi f.eks. interesserte i å undersøke forskjeller mellom sorter. Forsøksspørsmålet er altså om sortene genetisk sett gir ulik avling. Forsøksleddene er ulike såkornpartier eller plantematerialer av de sortene som skal testes. Hvis vi f.eks. i et sortsforsøk med poteter har noen sorter som er smittet med virus og andre som er friske, vil ikke vår forsøksfaktor gi et riktig svar på det forsøksspørsmålet vi egentlig stiller.

Vi bør ikke stille spørsmål hvis vi kjenner svaret på forhånd. Dette er i beste fall bortkastet arbeide. Denne regel syndes det ofte imot. Unnskyldningen er gjerne at man vil bruke forsøket til å demonstrere noen effekter som man allerede kjenner til. Demonstrasjon av kjente effekter og forsøk for å lære noe nytt bør ikke blandes sammen, det fører ofte til at vi bruker dårlig forsøksmetodikk. Kjente effekter kan demonstreres på enkelt-ruter. En annen ting er at forsøk ofte er nyttige til demonstrasjon.

## 1. Kvantitative og kvalitative forsøksfaktorer

Det er praktisk å skille mellom to typer av forsøksfaktorer, kvalitative og kvantitative. Eksempler på en kvalitativ faktor, eller forsøksledd som er sideordnede, er ulike sorter, ulike metoder eller ulike typer av redskaper. De enkelte ledd kan ikke på forhånd stilles opp i stigende eller fallende rekkefølge. For kvalitative forsøksfaktorer er vi interessert i å bestemme differenser mellom de ulike ledd, og konfidensgrensene for disse differensene.

For kvantitative, eller rekkeordnede forsøksfaktorer er forsøksleddene ulike gjødselmengder, ulike temperaturer til ulike tider

eller ulike konsentrasjoner. For kvantitative faktorer er vi først og fremst interessert i å bestemme forløpet av en kurve, og beliggenheten av et maksimum eller et optimum. Differensen mellom to ledd som ligger ved siden av hverandre i rekken av forsøksledd er vi lite interesserte i. Vi kan jo få den så stor vi vil ved å velge store forskjeller mellom de enkelte trinn av den rekkeordnete forsøksfaktoren.

Vi vil alltid finne noen eksempler som det er vanskelig å plassere i systemet. Hvis vi f.eks. vil undersøke virkningen av en eller annen faktor på sorter med ulik veksttid, kan sortene ordnes etter tidlighet, og dermed kan sort bli en kvantitativ faktor.

I de fleste tilfelle er det nokså greit å vite om en faktor er kvalitativ eller kvantitativ. For å bestemme hvor mange alternativer eller trinn det skal være av en forsøksfaktor, kan skillet mellom sideordnete og rekkeordnete faktorer være nyttig.

For sideordnete forsøksledd er det ofte grunn til å ta med mange alternativer i mindre nøyaktige forsøk, enn noen få alternativer i nøyaktige forsøk. Det hjelper lite å få vite med stor sikkerhet hvilken av sortene A, B, C og D som er best, hvis sort E, som kanskje er den aller beste, ikke har vært prøvd i det hele tatt. Det beste er at forsøksspørsmålet løses i to trinn. Først prøves alle aktuelle sorter i et relativt unøyaktig forsøk, og i neste omgang prøves de sortene som ga best resultat i det første forsøket mere grundig.

Med rekkeordnete forsøksledd er det omvendt. Her er vi interesserte i å bestemme et maksimum eller et optimum på en kurve. Hvis vi på forhånd vet lite om hvor dette ligger, er det praktisk å først anlegge et forsøk med noen få trinn med både riktig små og riktig store mengder. Dermed er vi sikre på at vi har fått med mengder på begge sider av det maksimum vi søker. Når vi har fått en antydning om hvor dette ligger, lager vi et nytt forsøk med mengder omkring maksimum, og gjerne med flere mengder for å få kurven bedre bestemt i dette området.

## 2. Faktorielle forsøksspørsmål

Det er vanligvis svært lønnsomt å anlegge forsøk med flere forsøksfaktorer samtidig. To forsøksoppgaver kan være å sammenligne de tre sortene A, B og C, og å måle virkningen av to ulike gjødslingsnivå, 1 og 2. Disse oppgavene kan løses på to måter. En måte er å anlegge ett sortsforsøk og ett gjødslingsforsøk som vist i tabell III.1. Begge forsøkene er anlagt med 5 gjentak og med tilfeldig fordeling av forsøksleddene innen gjentak.

Tabell IV.1. To en-faktor forsøk med sorter og gjødselmengder.

Sortforsøk				Gjødslingsforsøk		
Blokk	Sort			Blokk	Gjødselnivå	
I	A	B	C	I	1	2
II	B	A	C	II	2	1
III	A	C	B	III	2	1
IV	B	C	A	IV	1	2
V	C	A	B	V	1	2

En annen måte å løse oppgavene på er å anlegge ett faktorielt forsøk med sorter og gjødselmengder i det samme forsøket som vist i tabell IV.2. Forsøksleddene blir her alle 6 kombinasjonene av 3 sorter og 2 gjødselmengder, og disse forsøksleddene er fordelt tilfeldig innen hver blokk.

I det første alternativet har vi brukt 25 ruter ialt, i det andre 24. De to måtene koster derfor omtrent det samme. Nøaktigheten av våre estimer er avhengig av hvor mange ganger vi kan bestemme dem. Jo flere svar vi får for de ulike effektene, jo sikrere blir gjennomsnittet. I alternativ 1 får vi ett svar på forskjellen mellom sortene (f.eks. A - B) i hver blokk, altså 5

ialt. I alternativ 2 får vi 2 svar på sortforskjellene i hver blokk (både A,1 - B,1 og A,2 - B,2) altså 8 svar ialt.

Tabell IV.2. Faktorielt forsøk med 3 sorter og 2 gjødseltrinn.

Blokk	Sort og gjødselnivå					
I	A,1	B,2	B,1	C,2	C,1	A,2
II	B,2	C,2	A,1	C,1	A,2	B,1
III	B,1	C,2	C,1	A,2	A,1	B,2
IV	C,2	C,1	B,2	A,1	A,2	B,1

Gjødslingseffekten får vi 5 svar på i alternativ 1, mens alternativ 2 gir oss hele 12 svar, 3 i hver blokk. A,2 - A,1, B,2 - B,1 og C,2 - C,1 er alle gjødslingseffekter.

Vi får flere svar på begge hovedeffektene (sortsforskjeller og gjødslingseffekt) ved å bruke en faktoriell plan enn ved å bruke to en-faktor forsøk. Dessuten får vi i mulighet for å bestemme samspill når vi bruker en faktoriell plan. Med samspill mener vi at utslaget for den ene faktoren er forskjellig for de ulike trinn av den andre faktoren. I eksemplet blir det at forskjellen mellom sortene ikke er den samme ved svak og ved sterk gjødsling. Vi kan også si at utslaget for gjødsling varierer fra sort til sort. (A,2 - A,1) - (B,2 - B,1) er en del av et slikt samspill. Skrevet på denne måten leser vi det som: Gjødslingseffekten for sort A sammenlignet med gjødslingseffekten for sort B. Uttrykket kan omskrives til (A,2 - B,2) - (A,1 - B,1), som leses: Forskjellen mellom sort A og sort B ved sterk og ved svak gjødsling.

I det faktorielle forsøket får vi et svar på slike samspill i hver blokk, altså 4 svar ialt, mens to en-faktor forsøk ikke gir oss noen opplysning om samspillet.

For en-faktor forsøkene må vi velge et nivå av den andre faktoren før vi setter i gang forsøket. Vi må bestemme oss for gjøds-



lingsnivå for sortsforsøket, og vi må bestemme oss for sort for gjødslingsforsøket. Hvis det senere viser seg at vi har valgt feil sort eller feil gjødslingsnivå, og at det er samspill mellom sort og gjødsling, har resultatene fra en-faktor forsøkene svært begrenset verdi.

Faktorielle planer kommer bare fullt til sin rett hvis de er fullstendige, det vil si at vi tar med alle kombinasjoner av faktorene. La oss anta at vi i den faktorielle planen hadde sløffet C,2, enten fordi vi mente å vite at sort C ikke tålte sterk gjødsling, eller fordi vi hadde bestemt oss for å bruke en forsøksplan som bare kan ha 5 ledd. Til gjengjeld kunne vi altså ha 5 gjentak for at forsøket skulle koste det samme. Hvor mange svar får vi nå?

Forskjellen mellom A og B blir bestemt 2 ganger i hvert gjentak, altså 10 ganger ialt, mens vi bare får 5 svar på A - C og B - C. Gjennomsnittlig blir dette  $6 \frac{2}{3}$  svar på sortforskjellene, mot 8 i den fullstendige planen. Gjødslingseffekten får vi 2 svar på i hvert gjentak, altså 10 ialt, mot 12 i den fullstendige planen. Når det gjelder samspillet får vi 5 svar på om A og B reagerer ulikt på gjødsling, men ingen svar på forskjellen A - C eller B - C, altså i middel  $1 \frac{2}{3}$  svar mot 4 ved fullstendig plan.

Vi har eksempler på at det kan være vanskelig eller nesten umulig å få til visse kombinasjoner i forsøket, og enda værre i praksis. Et eksempel er at vil prøve ulike såmengder (pr dekar) og ulike radavstander. Med de maskiner vi har, kan det være umulig å få til riktig store såmengder kombinert med stor radavstand, eller riktig små mengder med liten radavstand.

Et alternativ kan da være at vi kombinerer ulike radavstander med ulike såmengder pr. m rad, da blir det ingen vanskelighet med å få til alle kombinasjoner.

I svært mange tilfeller hvor vi finner ufullstendige planer i bruk, skyldes dette ikke rasjonelle overveielser, men bare at den som la opp forsøket har ment at det ble for mange kombinasjoner.

Derfor har han sløffet en del, mer eller mindre tilfeldig. Fullstendige planer bør være regelen. Ufullstendige planer er unntak som må begrunnes!

Hvis vi har mange faktorer og/eller mange trinn av hver faktor, og derfor får svært mange forsøksledd blir det kostbart å ha mange gjentak. Det finnes metoder (split plot, konfundering) som gjør at vi likevel kan få relativt små feil på de sammenligninger vi er mest interesserte i. Vi skal komme tilbake til dette senere.

## V. FORSØKSENHET, GJENTAK OG PARALLELLER:

Forsøksenheten er den enheten vi tar enkeltobservasjonene på. Det kan være en rute i et markforsøk, en kasse eller en potte i et veksthusforsøk, et dyr eller en gruppe av dyr i husdyrforsøk, eller et tre eller en busk i frukt og bærforsøk. Forsøksenheten kan også være en enkelt plante eller deler av en plante i spesielle tilfeller. I økonomiske undersøkelser kan forsøksenheten være en driftsenhet eller et driftssystem.

Det er viktig at forsøksenheten virkelig representerer det vi er interesserte i å måle. Deler av et blad kan være representativt for å måle resistensegenskaper hos en plante, mens agronomiske egenskaper som avling, tidlighet og stråstyrke i korn og gras må bestemmes på en forsøksenhet som er stor nok til å representere et praktisk bestand.

Feilen på enkeltobservasjonene i et forsøk er blant annet avhengig av forsøksenheten. Feilen på gjennomsnittene av de enkelte forsøksledd er dessuten avhengig av antall observasjoner pr. forsøksledd.

### 1. Forsøksenhet, størrelse og form

Den vanligste forsøksenheten i markforsøk er en rute på 5-20m<sup>2</sup>. Normalt er det bare avlingen på hele ruten og ikke for den enkelte planten som blir bestemt, og det er også rutene som er enheten i den statistiske analysen. Forsøksfeilen går vanligvis ned når størrelsen på enheten øker. I et markforsøk vil alle måle- og veiefeil ha relativt større betydning på små enn på store ruter. Tilfeldig variasjon innen rutene vil også utjevnes når rutestørrelsen økes.

Det er vanskelig å si noe generelt om hva som er den beste størrelsen av en forsøksenhet. Omkostningene i et forsøk øker med rutestørrelsen, og hvis det ikke er noen kant- eller nabo-virkning, er det riktig å bruke små enheter. I markforsøk som er

anlagt for å gi veiledning om valg av sorter, gjødselnivå, sprøyting, jordarbeiding osv., er minimumskravet at rutene representerer et praktisk plantebestand.

Det er alltid en minste størrelse som det ikke går an å komme under. I forsøk med poteter er et ris et slikt absolutt minimum, i foringsforsøk med gris er bingen den minste praktiske enhet. Ved individuell foring kan "grisen" brukes som minste enhet.

### Nabovirkning

I gjødslingsforsøk vil plantene på de ugjødslete eller svakt gjødslete rutene ha nytte av gjødslinga på naborutene, både fordi det ikke er mulig å hindre at noe av gjødsla kommer inn på naborutene, og fordi røttene vokser inn på de gjødslete rutene. Den eneste måten å unngå nabovirkning på er å ha tilstrekkelig brede grensebelter. Avlinga bestemmes da bare på den midtre delen av ruta. Jo mindre rutene er, jo større andel vil gå bort i grensebelter. Den optimale brutto rutestørrelsen vil derfor øke når det er nødvendig å bruke grensebelter.

Også i sortsforsøk kan det være nabovirkning, især når sortene som sammenlignes er svært forskjellige. Hvis en sort starter veksten tidligere enn de andre vil den ta vann, næring og lys fra naborutene, og derved trykke veksten på disse. For samme rute-størrelse vil nabovirkningen bli større jo smalere rutene er, og den vil bli minst ved bruk av kvadratiske ruter.

Det er likevel vanlig å bruke relativt lange og smale ruter, og i sortsforsøk blir det normalt ikke brukt grensebelter. Årsaken til dette er for det første at de sorter som skal sammenlignes oftest er nokså like, og at det derfor ikke er særlig stor nabovirkning. Særlig ved maskinell høsting er det arbeidsbesparende med lange og smale ruter, og de sortene som skal sammenlignes kommer da da nærmere hverandre, og får dermed mer lik jord.

Også i andre typer av forsøk er det kant- eller nabovirkninger. Et eksempel er forsøk med forskjellige forslag til ku. For ikke

å få med forskjellen mellom kuene som en forstyrrende faktor, kan de forskjellige forslagene prøves på samme ku med en ukes mellomrom. Forsøksenheten blir da ikke ku men ku i ei uke. Det vil være en ettervirkning av hvert forslag som omtrent svarer til kantvirkningen i et markforsøk. Etter at vi har skiftet forslag må vi vente en tid før vi måler effekten, og lengden av denne tiden (=grensebeltet) vil være avhengig av hva vi vil måle. Forandringer i det kjemiske innholdet i blodet kommer nok før forandringer i mjølkas sammensetning, mens det antakelig tar lengre tid før vi får forandringer i mjølkemengden.

Et annet eksempel er at vi vil undersøke tidsforbruket for ulike metoder for et manuelt arbeide, f.eks. bruk av ulike hakker ved tynning av rotvekster. Hvis det ikke var nødvendig med opplæringstid, kunne samme person bruke de ulike hakkene i 5 minutter hver, og vi kunne måle prestasjonene i hver av disse periodene. Imidlertid må vi regne med at vi trenger en opplærings- eller tilvenningstid før vi kan få riktige og sammenlignbare tall for prestasjonene. Dette svarer da til kantbeltene i et markforsøk. Hvis metodene ikke var like anstrengende kunne vi komme helt galt ut ved å bruke svært korte tidsperioder. I ekstreme fall kan personene venne seg til en bestemt takt, og derfor bruke like lang tid på alle metodene, idet de hviler ut på de lette metodene og blir anstrengt på de andre. Overført til et gjødslingsforsøk ville det svare til at man brukte så små ruter at alle plantene, takket være røttens utbredelse i jorden, fikk like stor nærings-tilgang.

Ved planleggingen av et forsøk er det viktig å tenke nøye over forsøksspørsmålet for å finne den beste størrelse og form på forsøksenheten. Hvis vi venter store nabo- eller kantvirkninger, må rutene være tilsvarende store. Rutene må alltid være så store at forholdene i forsøket svarer så godt som mulig til forholdene i praksis. På den andre sida vil små og smale ruter gjøre at de enkelte ledd kommer nærmere hverandre og dermed får mere like vilkår. Små ruter vil også være billigere enn store ruter. For de samme omkostningene kan vi derfor ha flere gjentak, og dermed redusere feilen på gjennomsnittene.

## 2. Gjentak og paralleller.

Med økende antall observasjoner pr. forsøksledd vil feilen på forsøksleddsgjennomsnittene minke. Feilen er omvendt proporsjonal med kvadratrotten av antallet. Det er altså den første økningen, fra 2 til 3, eller 4 observasjoner som har størst effekt. Senere flater nedgangen ut. Økning av antall observasjoner har også en annen fordel. Det blir flere frihetsgrader for feilen som dermed blir sikrere bestemt. Ut fra en t- eller F-tabell ser vi at det er lite å vinne ved å ha mer enn 20-30 frihetsgrader for feilen. Av denne grunn er det mest å vinne på å ha flere observasjoner i forsøk med få forsøksledd. Har forsøket riktig mange ledd blir det likevel tilstrekkelig mange frihetsgrader for feilen.

Hvis vi vet omtrent hvor store forsøksfeil vi kan vente å få, og hvor små differenser vi er interesserte i å kunne fastslå, kan vi gjøre et anslag over hvor mange observasjoner vi trenger i et planlagt forsøk.

Observasjonene må være et tilfeldig utvalg av det universet vi vil uttale oss om. Det hjelper ikke å ha mange observasjoner på et enkelt jorde i ett enkelt år hvis resultatene skal brukes som veiledning for et område for neste år.

I en enveisgruppering kaller vi de ulike observasjonene for samme forsøksledd paralleller. Observasjon nr. 1 for ledd 1 og ledd 2 har da ingenting felles. Et eksempel på paralleller var observasjonene over 3 potetsorter på 27 ulike gårder.

I toveiseksemplet med 3 potetsorter på 5 gårder er gårdene gjentak. For å kunne bruke betegnelsen gjentak, må det være ett eller annet som er felles for forsøksleddene i gjentak 1, forsøksleddene i gjentak 2 osv. Som nevnt kunne gårder i dette eksemplet være dårlige gjentak da vi ikke har noen garanti for at de 3 sortene var behandlet likt. I et forsøk skal det være minst mulig variasjon mellom forsøksenhetene innen gjentak.

Vi kommer tilbake til ulike typer gjentak senere i kurset.

## VI. FORSØKSPLANER

Vi deler inn forsøksplanene etter de prinsippene som er brukt for å fordele forsøksleddene på de enkelte forsøksenhetene. Vi skal gå gjennom:

- Fullstendig tilfeldig fordeling
- Blokkforsøk (fullstendige blokker)
- Latinsk kvadrat
- Split plot / split blokk
- Ufullstendige blokker (konfundering)

Split plot og konfundering forutsetter faktorielle forsøksspørsmål, men vi understreker at det ofte er hensiktsmessig å bruke fullstendige blokkforsøk også for faktorielle spørsmål.

### 1. Fullstendig tilfeldig fordeling

Fullstendig tilfeldig fordeling er den enkleste forsøksplan. Som det går fram av navnet, fordeler vi forsøksleddene tilfeldig på forsøksenhetene. Det behøver ikke å være det samme antall observasjoner for de forskjellige forsøksledd. All variasjon mellom forsøksenhetene vil bidra til forsøksfeilen, og fullstendig tilfeldig fordeling blir derfor vanligvis bare brukt for et ensartet forsøksmateriale.

Et eksempel er karrforsøk med gjødsling. Hvis vi f.eks. skal sammenligne 10 ulike gjødslingsalternativer med 3 karr av hvert alternativ, blander vi omhyggelig en jordmengde som rekker til alle 30 karr, og fordeler denne på karrene. Etterpå blir gjødsla blandet inn i hvert enkelt karr.

Den statistiske analysen av et forsøk med fullstendig tilfeldig fordeling er en enveis variansanalyse (se s 20).

## 2. Blokkforsøk

Det er ofte vanskelig å skaffe et tilstrekkelig ensartet forsøksmateriale. Feilen i et forsøk med fullstendig tilfeldig fordeling blir da stor. Hensikten med et blokkforsøk er nettop å minske denne feilen.

Vi deler forsøksenhetene i så mange grupper, blokker, som vi har gjentak. Hver blokk skal inneholde hvert forsøksledd en gang. Antall gjentak betegnes med  $r$  (forkortelse for replication) og antall forsøksledd med  $t$  (forkortelse for treatment). Et blokkforsøk består av  $r$  blokker med  $t$  enheter i hver. Ved den praktiske gjennomføringen av forsøket arrangeres forsøksenhetene slik at det blir minst mulig variasjon mellom forsøksenhetene innen hver blokk. Siden alle forsøksledd kan sammenlignes i hver blokk, er det bare variasjonen innen blokkene som bidrar til forsøksfeilen.

Det vanligste eksempel på et blokkforsøk er et markforsøk. Hvis vi på forhånd ikke vet noe om jordvariasjonen på feltet, lager vi blokkene så kompakte som mulig, det vil si så nær kvadratiske som vi kan få til. Ofte vet vi ganske mye om jordvariasjonen før vi anlegger feltet. Ligger feltet i en skråning, vil det som regel være dypere jord nederst enn lenger oppe. Det blir da de jevneste forhold innen blokker hvis blokkene lages langstrakte og med lengderetningen på tvers av heldningen.

Det er vanlig å finne stripeformet variasjon i kulturjord, f.eks. etter grøfting, tidligere gjødsling eller jordarbeiding (teigfårer eller teigrygger). Blokkene må da legges langs stripene, med rutene innen blokkene på tvers av stripene. På den måten vil rutene i samme blokk bli så like som mulig.

Hvis vi har anledning til å se veksten på et felt året før forsøket skal anlegges der, er det ofte mulig å ta ut jevne blokker. Også under opptørkingen om våren er det forholdsvis lett å observere ujevnheter, og unngå disse ved oppdelingen i blokker.



I markforsøk er det vanlig å lage både blokkene og hele feltet kompakt, det vil si at rutene i hver blokk ligger samlet, og at blokkene ligger inntil hverandre. Ingen av delene er nødvendig. Hvis det er mulig å få jevnere blokker ved at blokkene eller rutene innen blokk legges et stykke fra hverandre, bør vi gjøre det. Hvis det er ujevne partier inne i det arealet som skal brukes, prøver vi å legge blokkene utenom disse partiene.

I andre typer forsøk må vi bruke helt andre kriterier for blokk-inndelingen, men målet er alltid å få minst mulig variasjon innen blokkene. Hvis det i et foringsforsøk med gris ikke er mer enn 4-5 forsøksledd, kan blokkene bestå av griser fra samme kull og av samme kjønn. Ved sammenligningen av forsøksleddene vil da all variasjon som skyldes ulik alder og ulikt kjønn, og det meste av den arvelige variasjonen være eliminert.

I et forsøk med sammenligning av ulike metoder for et manuelt arbeide kan forsøkspersonene danne blokker. Hvis alle personene som deltar i forsøket prøver alle metodene, vil forskjeller mellom personene ikke påvirke forskjellen mellom metodene.

Selv om det ikke er noen klare kriterier for å dele forsøksmaterialet opp i ensartete blokker, er det ofte nyttig å bruke blokker. Det kan f.eks. hende at såingen av et markforsøk må avbrytes på grunn av regnvær før hele feltet er ferdig sådd. Hvis avbruddet blir gjort mellom to blokker, vil det ha liten innflydelse på forsøksresultatene. Innhøsting av forsøksfelt kan ta lang tid. Hvis alle rutene i samme blokk blir høstet omtrent samtidig, kan gjerne ulike blokker høstes på forskjellige dager.

Kvalitetsanalysering av forsøksmaterialet bør også skje blokk for blokk og absolutt ikke ledd for ledd. Analyseresultatene kan forandres, f.eks. på grunn av prøvenes alder, eller på grunn av forandringer i temperatur, lysforhold, strømstyrke o.l. Disse forandringer vil da først og fremst virke mellom blokker, og bare i liten grad påvirke forskjellen mellom forsøksleddene.

En god blokkinnndeling er ofte både den billigste og mest effektive måte å lage nøyaktige forsøk på. Hvis blokkinnndelingen fører

til at restkvadratsummen i variansanalysen reduseres til halvparten i forhold til å bruke fullstendig tilfeldig fordeling, har oppdelingen i blokker hatt samme effekt som en dobling av antall gjentak.

Sammenlignet med et forsøk med fullstendig tilfeldig fordeling har et blokkforsøk to svakheter.

- 1) Vi må ha det samme antall gjentak for alle forsøksledd. Av og til kan vi ønske å variere antall gjentak, f.eks. for å utnytte alt tilgjengelig materiale, og dette er vanskelig å få til i et blokkforsøk.
- 2) Antall frihetsgrader for feilen blir litt mindre i et blokkforsøk enn i et forsøk med fullstendig tilfeldig fordeling,  $(r-1)(t-1)$  istedenfor  $(r-1)t$ . Dette kan bety endel i ganske små forsøk med få forsøksledd og få gjentak. Hvis vi har et forsøk med 3 forsøksledd og 3 gjentak, vil vi få 4 frihetsgrader for feilen i et blokkforsøk mot 6 i et forsøk med fullstendig tilfeldig fordeling. Ser vi etter i en t-tabell, finner vi at 5% verdiene er henholdsvis ca 3,5 og 3,0, og hvis feilen ikke går vesentlig ned ved blokkingen, kan det være lettere å påvise forskjeller hvis vi bruker fullstendig tilfeldig fordeling.

Dette gjelder bare hvis vi ønsker å trekke konklusjoner av et enkelt forsøk. Hvis forsøket er ledd i en serie forsøk, er vi lite interessert i å bestemme feilen i det enkelte forsøk, det gjelder bare å få den minst mulig, og da bruker vi blokkforsøk. Vi kommer nærmere inn på dette under forsøksserier.

Hvis vi på forhånd er klar over at relativt mange av forsøksenhetene kommer til å måtte forkastes før oppgjøret, f.eks. at mange planter eller dyr dør av årsaker som ikke har noe med forsøket å gjøre, har vi liten nytte av blokkingen, og da kan vi like godt anlegge forsøket med fullstendig tilfeldig fordeling.

### 3. Latinsk kvadrat

I et blokkforsøk eliminerer vi virkningen av systematisk variasjon i en retning, forskjellen mellom blokkene, fra restvariasjonen i forsøket. I et latinsk kvadrat eliminerer vi virkningen av systematisk variasjon i to retninger.

Hvis vi vil sammenligne 4 forsøksledd, deler vi forsøksarealet opp i 4 rekker med 4 ruter i hver rekke. Forsøksleddene fordeles slik at hvert forsøksledd kommer en gang i hver rekke og en gang i hver kolonne (se tabell VI.1).

Tabell VI.1. Fordeling av forsøksleddene A-D i 4 rekker (I-IV) og 4 kolonner (1-4).

Rekke	Kolonne			
	1 (2)	2 (4)	3 (1)	4 (3)
I (III)	A	B	C	D
II (I)	D	A	B	C
III (IV)	C	D	A	B
IV (II)	B	C	D	A

For å fordele forsøksleddene tilfeldig på forsøksrutene settes nye rekkenummer og kolonnennummer på i tilfeldig rekkefølge (tallene i parentes), og det endelige latinske kvadratet blir som i tabell VI.2.

Rekkenummer og kolonnennummer i tabell VI.2 er satt i parentes for å vise at dette er de tilfeldige rekke- og kolonnennummer fra tabell VI.1.

Tabell VI.2. Fordeling av forsøksleddene A-D i 4 rekker og 4 kollar etter tilfeldig fordeling i begge retninger.

Rekke	Kolonne			
	(1)	(2)	(3)	(4)
(I)	B	D	C	A
(II)	D	B	A	C
(III)	C	A	D	B
(IV)	A	C	B	D

Ved den statistiske analysen av et latinsk kvadrat deles den totale variasjonen opp i 4 deler: en del mellom rekker, en del mellom kolonner, en del mellom forsøksledd og en rest. Resten brukes som feil for sammenligningen mellom forsøksledd.

$CT = (\Sigma x)^2 / t^2$  hvor t er antall forsøksledd

$$SS_{total} = \Sigma x^2 - CT$$

$$SS_{rekke} = ((\Sigma I)^2 + (\Sigma II)^2 + \dots) / t - CT$$

$$SS_{kol.} = ((\Sigma 1)^2 + (\Sigma 2)^2 + \dots) / t - CT$$

$$SS_{ledd} = ((\Sigma A)^2 + (\Sigma B)^2 + \dots) / t - CT$$

$$SS_{rest} = SS_{total} - SS_{rekke} - SS_{kol.} - SS_{ledd}$$

Forutsatt at vi har to variasjonsårsaker som vi ønsker å eliminere, vil et latinsk kvadrat gi mindre feil enn et tilsvarende blokkforsøk.

I markforsøk hender det ofte at det er jordvariasjon i to retninger. Det kan f.eks. være striper på grunn av jordarbeiding den ene veien og gjødslingsstriper den andre veien. Eller det kan være en skråning med grunn jord på toppen og djupere jord lenger nede i bakken, og teigrygger og teigfårer langs fallet.

I et latinsk kvadrat passer det best med noenlunde kvadratiske ruter. Da blir også hele feltet nær kvadratisk, og rekker og kolonner like lange. Hvis rutene er lange og smale, vil også rekkene (eller kolonnene) bli lange og smale, og da skal det godt gjøres at stripene i jorden følger rekkene i hele lengden. Latinsk kvadrat er mest effektivt til å redusere feilen hvis forskjellen mellom kolonnene er lik for alle rekker, i et felt med variasjon på skrå i forhold til rutene vil effekten være mindre.

Latinsk kvadrat kan også brukes i andre typer av forsøk enn markforsøk. Hvis vi ønsker å sammenligne 4 ulike arbeidsmetoder, kan vi bruke 4 personer, som alle skal prøve alle metodene. Samtidig vet vi at været spiller en stor rolle for resultatet. For å få eliminert virkningen av dette, må vi sørge for at de ulike metodene blir prøvd samtidig. Dette kan vi få til hvis vi setter opp et 4x4 latinsk kvadrat med personene som rekker og rekkefølgen (f.eks. dag nr.) som kolonner. På denne måten blir hver metode prøvd en gang for hver person og en gang for hver dag. Hverken effekten av person eller effekten av dag vil da være med når vi sammenligner gjennomsnittene for de 4 metodene.

Ulempene ved et latinsk kvadrat (sammenlignet med et blokkforsøk) er for det første at vi er bundet til å ha like mange gjentak som forsøksledd, og at vi får enda færre frihetsgrader for feilen enn i blokkforsøket. I et 3x3 latinsk kvadrat vil vi bare ha 2 frihetsgrader for feilen. Av denne grunn er et 4x4 latinsk kvadrat det minste som kan brukes hvis vi vil trekke noen konklusjoner av et enkelt forsøk.

Det blir oftest for kostbart å bruke mer enn 6 gjentak. Latinske kvadrater brukes derfor fortrinnsvis for 4, 5 eller 6 forsøksledd. For 3 forsøksledd kan vi bruke en plan med 2 eller 3 kvadrater. Antall gjentak blir da 2 eller 3 ganger antall forsøksledd, og vi får et tilstrekkelig antall frihetsgrader for feilen.



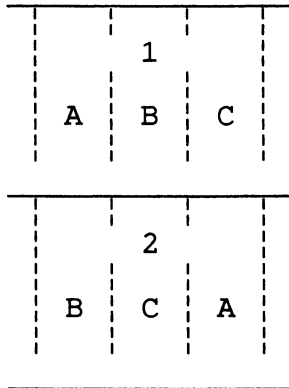
Gjødslingsrutene kommer lenger fra hverandre enn sortsrutene. Vi må derfor vente større tilfeldig variasjon for sammenligningen mellom gjødslingene enn for sammenligningen mellom sortene. I et split plot forsøk kalles feilen på storruter feil(a) og feilen for smårutene feil(b).

Sammenligningen av de 2 gjødslingsstyrkene blir dårligere enn sammenligningen av de 3 sortene, av to årsaker. For det første er feilen på storruter som oftest større enn feilen på småruter. For det andre er det færre frihetsgrader for feil(a). Når vi skal undersøke samspill, f.eks. om gjødseleffekten er forskjellig for sort A og sort B, skal vi bruke feil(b).

Antall frihetsgrader i variansanalysen blir:

Var.årsak	DF
Total	23
Rep	3
Gjødsling	1
feil(a)	3
Sorter	2
Gj.x Sorter	2
feil(b)	12

Regelen om at feil(b) er mindre enn feil(a) gjelder først og fremst hvis splittingsen av storrutene går i samme retning som oppdelingen av gjentakene i storruter, slik som i eksemplet foran. Hvis splittingsen går på tvers, slik som det er antydnet for et gjentak i figuren nedenfor, vil forskjellen på feil(a) og feil(b) være avhengig av i hvilken retning vi har størst jordvariasjon. Hvis vi har striper i samme retning som smårutene, kan feil(b) bli større enn feil(a).



Bruk av split plot er ikke begrenset til markforsøk. Sett at vi vil prøve 3 ulike oppskrifter på formkake og steking ved 2 forskjellige temperaturer, og at vi kan få 3 former inn i stekeovnen samtidig. Stekeovnen danner da storruiter, og de enkelte formene med hver sin oppskrift smårutene. Feil i temperaturreguleringen fra gang til gang vil virke inn på feil(a), mens feil(b) bare er påvirket av variasjoner i råvarene og av den tilfeldige feilen.

### Split split plot.

Hvis vi har 3 forsøksfaktorer som krever ulike rutestørrelser, kan det være aktuelt med 3 ulike rutestørrelser. Vi vil f.eks. sammenligne høstpløying (P) med bare stubbharving (H), og 3 ulike gjødslingsstyrker til de 3 vårkornartene havre, bygg og hvete. Jordarbeidingen krever svært store ruter, gjødslingen krever også ganske store ruter med grensebelter, mens artene kan sammenlignes på smale ruter uten grensebelter. Et gjentak av et slikt forsøksfelt kan da være slik:



Art	Gjødsling	Jordarbeiding
Havre	1	
Bygg	1	P
Hvete	1	l
		ø
Bygg	3	y
Hvete	3	e
Havre	3	t
Bygg	2	
Havre	2	
Hvete	2	
Hvete	2	
Havre	2	
Bygg	2	H
		a
Bygg	3	r
Havre	3	v
Hvete	3	e
		t
Hvete	1	
Bygg	1	
Havre	1	

Hvis vi har 3 gjentak blir variansanalysen for et slikt split split plot forsøk:

Var.årsak	DF
Total	53
Gjentak	2
Jorarbeiding	1
feil(a)	2
Gjødsling	2
Jordarb x Gjøds.	2
feil(b)	8
Arter	2
Jordarb x Art	2
Gjøds. x Art	4
Jor.x Gjø.x Art	4
feil(c)	24

Her er feil (a) samspillet mellom gjentak og jordarbeiding, feil (b) består av samspillene gjentak x gjødsling og gjentak x jordarbeiding x gjødsling, mens feil (c) består av 4 deler, nemlig samspillene gjentak x art, gjentak x jordarbeiding x art, gjentak x gjødsling x art og gjentak x jordarbeiding x gjødsling x art.

Alle hovedeffektene og samspillene mellom forsøksfaktorene skal testes mot den første feilen nedenfor i variansanalysen.

Vanligvis er MS for feil(a) størst og MS for feil(c) minst, mens feil(b) kommer i en mellomstilling. Da vi dessuten får svært få frihetsgrader for feilen på storrutene bør vi ikke bruke en sånn plan hvis det ikke er nødvendig av praktiske grunner.

Split split plot kan også brukes på andre typer forsøk. Vi vil undersøke virkningen av ulik luftfuktighet, ulik høydeplassing og ulik foring for burhøns. Rutene for klimaet må nødvendigvis være hele rom, for høydeplassing en etasje, mens foringen kan prøves på enkeltbur.

### Split blokk

Hvis vi har to faktorer som begge (av tekniske årsaker) trenger lange ruter, kan vi dele hvert gjentak opp i ruter i to retninger på tvers av hverandre for de to faktorene.

La oss anta at vi vil prøve 3 ulike pløedybder (1, 2 og 3) sammen med 4 forskjellige harvtyper (A, B, C og D). Det kan da være praktisk å bruke en split blokk plan. Et gjentak av en split blokk plan blir:

		Harvtype			
		D	A	C	B
Pløedybde	2				
	3				
	1				

Med 4 slike gjentak, hver med ny randomisering av begge faktorer, vil det riktige mønster for variansanalysen være.

Var. årsak	DF
Total	47
Rep	3
Dybde	2
feil(a) (rep x Dybde)	6
Harvtype	3
feil(b) (rep x Htype)	9
Dybde x Htype	6
Feil(c) (3-faktorsamspill)	18

Begge faktorer er her på storruter, og hvilken av feil(a) eller feil(b) som er størst vil være avhengig av i hvilken retning vi har størst jordvariasjon. Variasjon i begge retninger vil være eliminert fra samspillet og fra feil(c), som dermed blir mindre enn de to andre feilene. Feil(c) har dessuten flere frihetsgrader. Ved denne planen blir derfor samspillet best bestemt.

#### Når skal vi bruke split plot forsøk?

Sammenligning av variansanalysen for et split-plot (eller split blokk) forsøk med variansanalysen for et forsøk med de samme leddene i fullstendige blokker viser at kvadratsummen for feilen i blokk forsøket er summen av kvadratsummene for feilene i split plot forsøket. Hvis feil(b) i split plot forsøket er redusert, er samtidig feil(a) øket i forhold til den feilen vi ville hatt i blokkforsøket. Samtidig er det færre frihetsgrader både for feil(a) og for feil(b).

Vi bør derfor ikke bruke split plot, og enda mindre split split plot eller split blokk, ukritisk. Disse planene kan brukes enten når det er store praktiske fordeler, eller når vi er lite interesserte i hovedeffekten av den faktoren vi har på storruter. Split blokk bør bare brukes når det er nødvendig av tekniske årsaker, eller når det virkelig er samspill vi er interessert i å bestemme, og at begge hovedeffektene er nokså likegyldige.

## 5. Ufullstendige blokker (konfundering)

Konfundering betyr sammenblanding. For å redusere den totale variasjonen innen et gjentak, kan vi dele opp gjentakene i mindre blokker. Forsøksleddene fordeles på forsøksenhetene på en slik måte at forskjellene mellom, eller effekten av, småblokkene er det samme som effektene av samspill som vi er mindre interesserte i. Vi konfunderer altså visse samspillseffekter med småblokk-effektene.

Konfundering er mest aktuelt når forsøket har 3 eller flere faktorer og bare 2 trinn av hver faktor.

Med 3 faktorer, A, B og C, og med trinn 0 og 1 av hver faktor, har vi følgende 8 kombinasjoner:

A0B0C0	A1B0C0	A0B1C0	A1B1C0	A0B0C1	A1B0C1	A0B1C1	A1B1C1
1	a	b	ab	c	ac	bc	abc

Kombinasjonen med de lågeste verdier av alle faktorer får betegnelsen 1, mens de øvrige kombinasjoner får navn (små bokstaver) etter de faktorer som er i høyeste trinn.

Hovedeffekt A kan i hvert gjentak bestemmes som 4 differenser, nemlig  $(a-1)$ ,  $(ab-b)$ ,  $(ac-c)$  og  $(abc-bc)$ . I gjennomsnitt blir

$$\begin{aligned} A &= 1/4((a-1)+(ab-b)+(ac-c)+(abc-bc)) \\ &= 1/4(a+ab+ac+abc-1-b-c-bc) \end{aligned}$$

Alle ledd med a får altså +, alle ledd uten a -. På tilsvarende måte blir hovedeffektene for B og C beregnet, + for ledd med, og - for ledd uten vedkommende bokstav.

Samspillet mellom A og B er forskjellen mellom effekten av A når B er tilstede og når B ikke er tilstede, og (for å få samme feil på hovedeffekter og samspill) halvparten av dette. I hvert gjentak får vi to bestemmelser av AB-samspillet, et uten C og et med C.

I middel blir derfor

$$\begin{aligned} AB &= 1/2( 1/2((ab-b)-(a-1)) + 1/2((abc-bc)-(ac-c)) \\ &= 1/4(+1-a-b+ab+c-ac-bc+abc) \end{aligned}$$

Vi får + for de ledd hvor enten ingen eller begge av de to bokstavene er med, og - hvis bare en av bokstavene a og b er med. Det er vel verdt å kontrollere at vi kommer frem til det samme uttrykk hvis vi definerer samspillet AB som forskjellen mellom effekten av B når A er tilstede og når A ikke er tilstede.

Trefaktorsamspillet ABC er forskjellen mellom AB-samspillet når C er tilstede og når C ikke er tilstede, og nå med faktoren 1/4. Etter oppløsning av parentesene og ordning :

$$ABC=1/4(-1+a+b-ab+c-ac-bc+abc)$$

For et trefaktorsamspill får vi + hvis enten 1 eller alle 3 av faktorene er tilstede, og - hvis enten 0 eller 2 er med. Resultatene er samlet i tabell VI.5.

Tabell VI.5. Hovedeffekter og samspill for forsøk med 3 faktorer hver med 2 trinn.

Forsøks- ledd	Effekter						
	A	B	AB	C	AC	BC	ABC
1	-	-	+	-	+	+	-
a	+	-	-	-	-	+	+
b	-	+	-	-	+	-	+
ab	+	+	+	-	-	-	-
c	-	-	+	+	-	-	+
ac	+	-	-	+	+	-	-
bc	-	+	-	+	-	+	-
abc	+	+	+	+	+	+	+

En slik tabell er lett å sette opp, uten at det er nødvendig å huske hele tabellen. Først gir vi fortegnene for hovedeffektene (A, B og C) ved å sette + for de forsøksleddene der bokstaven finnes, og - for de andre forsøksleddene. I kolonnene for tofaktorsamspillene og multipliserer vi for tegnene for de to hovedeffektene. To pluss eller to minus gir +, et pluss og et minus gir -. Fortegnene for 3-faktorsamspillet får vi ved å multiplisere fortegnet for et 2-faktorsamspill med fortegnet for den tredje hovedeffekten, f.eks. AB og C.

Det er lett å utvide tabellen til flere enn 3 faktorer. Har vi også faktoren D tilføyer vi linjene d, ad, bd, abd, cd, acd, bcd og abcd, og kolonnene D, AD, BD, ABD, CD, ACD, BCD og ABCD. Vi bruker nøyaktig de samme reglene for å sette inn fortegnene.

I forsøket med 3 faktorer og 2 trinn av hver har vi 8 ledd. Vi deler opp hvert gjentak i to blokker med 4 ruter i hver blokk, og fordeler leddene slik at 1,ab,ac og bc kommer i den ene blokken, og a,b,c og abc i den andre. I tabellen foran ser vi at både de tre hovedeffektene og de tre tofaktorsamspillene har to + og to - i hver blokk. En eventuell forskjell mellom blokkene vil ikke virke inn på disse effektene. Trefaktorsamspillet ABC er akkurat lik forskjellen mellom de to blokkene, og vi sier at trefaktorsamspillet er konfundert (blandet sammen) med blokkforskjellene. Hvis vi har 3 gjentak (6 blokker) blir variansanalysen:

<u>Var.årsak</u>	<u>DF</u>
Total	23
Blokker	5
A	1
B	1
AB	1
C	1
AC	1
BC	1
<u>feil=rest</u>	<u>12</u>

Med fullstendige blokker ville vi hatt 2 DF for rep, 7 for ledd og en rest på 14 for feil. Variasjonen innen blokker på 4 enheter vil som oftest være mindre enn variasjonen innen 8-ruters blokker. Ved konfundering oppnår vi derfor å få mindre feil på hovedeffektene og på tofaktorsamspillene. På den andre siden får vi ikke bestemt trefaktorsamspillet, og feilen blir litt dårligere bestemt.

Trefaktorsamspill er som oftest små og vanskelige å tolke og utnytte i praktisk veiledning. Konfundering av dette samspillet kan derfor være nyttig. Høyere ordens samspill (4- eller 5-faktors) er av enda mindre interesse og kan trygt ofres. I 4-faktor forsøk, 16 ledd, kan hvert gjentak deles i to blokker, med 1,ab,ac,bc,ad,bd,cd og abcd (leddene med 0, 2 eller 4 bokstaver) i den ene blokken og a,b,c,abc,d,abd,acd og bcd (leddene med 1 eller 3 bokstaver) i den andre blokken.

Med 2 gjentak får vi 32 ruter fordelt på 4 blokker, og i variansanalysen blir det 1 DF for hver av 4 hovedeffekter, 6 tofaktorsamspill og 4 trefaktorsamspill, og 3 DF for blokker. Feilen blir bestemt med 14 DF mot 15 i et like stort forsøk med fullstendige blokker. Da variasjonen innen 8-ruters blokker sannsynligvis er mindre enn innen 16-ruters blokker, vil alle effekter og samspill av betydning bli best bestemt ved konfundering.

#### Sammenligning mellom konfundering og split plot

I eksemplet foran har vi konfundert trefaktorsamspillet ABC med blokker. Vi kunne også konfundert et av de andre samspillene ved å ta de 4 ledd med + for dette samspill i den ene blokken, og de 4 med - i den andre. Hadde vi f.eks. hatt 1,ab,c og abc i den ene blokken og a,b,ac og bc i den andre, ville vi konfundert AB med blokker, mens hovedeffektene og samspillene AC, BC og ABC ville vært bestemt innen blokker. Når vi deler gjentak i to blokker velger vi som oftest å konfundere det høyeste samspill, da det er dette vi er minst interesserte i.

Hva skjer hvis vi konfunderer en av hovedeffektene med blokker? Hvis vi f.eks. tar leddene med + for A i den ene og de med - for A i den andre. Jo, da blir resultatet et split plot forsøk med A på storuter. Da vi som oftest er mer interessert i hovedeffekter enn i samspill er dette en dårlig måte å gjøre det på hvis ikke rene praktiske årsaker gjør at vi vil ha store ruter for A. En forskjell mellom split plot og konfundering er at vi i split plot forsøk tester hovedeffekten mot feil(a), mens vi ved konfundering som oftest ikke bryr oss om å teste det samspillet som er konfundert. Vi kan dele opp variasjonen mellom blokker (i eksemplet med 5 DF) i tre deler, en del mellom gjentak (2 DF), en del for 3-faktor samspillet (1 DF) og en rest (2 DF), som da er feilen på 3-faktor samspillet.



## VII. OPPDELING AV SS OG DF

I avsnittet om faktorielle forsøksplaner har vi nevnt at variasjonen mellom forsøksledd kan deles opp i hovedeffekter og samspill. Også i enfaktorforsøk kan det ofte være nyttig å dele variasjonen opp i ulike deler som kan testes hver for seg.

Hvis de 5 byggsortene i oppgave 4 var to 2-radssorter og tre 6-radssorter, kan SS mellom sorter, med 4 DF, deles opp i en del mellom 2- og 6-rads sorter (1 DF), en del innen 2-rads sortene (1 DF), og en del innen 6-radssortene (2 DF).

I oppgave 4 fant vi at de ulike sortene ga følgende avlinger i sum over 3 gjentak:

Sort	Avling	Sum 2-rads	Sum 6-rads
A	10,44		
B	10,47	20,91	
C	7,99		
D	9,46		
E	9,18		26,63

Hvis vi forutsetter at sortene A og B er 2-radssorter og C, D og E er 6-radssorter, er summen av 2-radssortene 20,91 kg og summen av 6-radssortene 26,63 kg.

Beregningen av kvadratsummen for de tre sammenligningene blir:

$$SS_{2r-6r} : 20,91^2/6+26,63^2/9-CT = 0,99645$$

$$SS_{\text{innen } 2r} : (10,44^2+10,47^2)/3-20,91^2/6 = 0,00015$$

$$SS_{\text{innen } 6r} : (7,99^2+9,46^2+9,18^2)/3-26,63^2/9 = 0,40616$$

Summen av disse tre kvadratsummene er 1,40276 som stemmer med  $SS_{\text{sort}}$  i oppgave 4.

## 1. Kontraster

Vi skal se nærmere på definisjon og testing av kontraster, som er sammenligninger med 1 DF. En kontrast defineres som en kombinasjon av forsøksleddene, hvor summen av koeffisientene er null. Vi vil også her starte med et eksempel:

La oss kalle avlingssummene for de 5 byggsortene i oppgave 4 for a, b, c, d og e. a og b er avlingene for 2-radssortene og c, d og e er avlingene for 6-radssortene. Forskjellen mellom gjennomsnittet av de 2 gruppene er  $1/2(a+b) - 1/3(c+d+e)$ , eller hvis vi multipliserer med 6 for å unngå brøker:  $3a+3b-2c-2d-2e$ . Dette er en kontrast da summen av koeffisientene er 0.

Forskjellen mellom a og b, eller  $a - b$  er også en kontrast. Når forsøket har mer enn to forsøksledd kan det lages et ubegrenset antall kontraster. En spesiell type kontraster er uavhengige eller ortogonale kontraster. To kontraster er ortogonale hvis summen av produktene av parvise koeffisienter er 0. For eksemplet ovenfor er koeffisientene:

	Forsøksledd				
	A	B	C	D	E
Koef. for kontrast 1	+3	+3	-2	-2	-2
Koef. for kontrast 2	+1	-1	0	0	0
Produkt	+3	-3	0	0	0

Ialt kan vi lage like mange innbyrdes ortogonale kontraster som antall frihetsgrader for forsøksledd. Kontrasten  $a - c$  er f.eks. ikke ortogonal til de to første. Kontroller selv at summen av produktene ikke blir 0. Hvis vi som den tredje kontrasten velger  $c - d$  (som er ortogonal til 1. og 2.), så er den fjerde ortogonale kontrasten gitt, nemlig  $c + d - 2e$ .

Det går an å lage et ubegrenset antall sett med innbyrdes ortogonale kontraster, men alle er ikke like meningsfylte. I eksemplet er kontrast 1 forskjellen mellom 2-rads og 6-rads sorter, kontrast 2 er forskjellen innen 2-radssortene, mens kontrast 3 og 4 tilsammen er forskjellen innen 6-radssortene.

Hvis en kontrast (K) er regnet ut på forsøksleddsummene har vi denne formelen for kvadratsummen:

$$SS = K^2 / r \sum k^2$$

hvor r er antall gjentak og k-ene er de enkelte koeffisientene.

I oppgave 4 har vi følgende forsøksleddsummer:

$$a = 10,44 \quad b = 10,47 \quad c = 7,99 \quad d = 9,46 \quad e = 9,18$$

og de 4 ortogonale kontrastene blir:

1	$3a+3b-2c-2d-2e = 9,47$	med divisor	$3(9+9+4+4+4)=90$	$SS=0,99645$
2	$1a-1b = -0,03$	" "	$3(1+1) = 6$	$SS=0,00015$
3	$+1c-1d = -1,47$	" "	$3(1+1) = 6$	$SS=0,36019$
4	$+1c+1d-2e = -0,91$	" "	$3(1+1+4) = 18$	$SS=0,04601$
			Sum	$SS=1,4028$

Sum SS stemmer med SS i oppgave 4.

Hvis vi legger sammen SS for kontrast 3 og 4 får vi SS for sammenligningen av de tre 6-radssortene  $=0,4062$ , og  $MS=0,2031$ . Alle kontrastene og MS for de 3 6-radssortene kan testes mot  $MS_{rest} = 0,0436$ .

Konklusjonen blir at gjennomsnittsavlinga av de to 2-radssortene var signifikant større enn gjennomsnittsavlinga av de tre 6-radssortene. Det kan ikke påvises noen forskjell mellom 2-radssortene innbyrdes, og det er signifikant forskjell mellom 6-radssortene C og D.

## 2. Ortogonale polynomer

I forsøk med rekkeordnede forsøksledd, f.eks. stigende gjødselmengder, bruker vi de ortogonale kontrastene på en spesiell måte. Som før kan vi sette opp like mange ortogonale kontraster som vi har frihetsgrader for forsøksledd. La oss starte med et polynom hvor forsøksresultatet ( $x$ ) er uttrykt som en funksjon av forsøksledd ( $t$ ).

$$x = a + bt + ct^2$$

Koeffisientene for den første kontrasten settes opp slik at kontrasten er et uttrykk for den lineære regresjonen av forsøksresultat med hensyn på forsøksledd. Koeffisientene for den andre kontrasten settes opp slik at kontrasten er et uttrykk for 2. grads avvik fra den lineære regresjonen. Den tredje kontrasten skal være et uttrykk for 3. grads avvik fra 2. grads regresjonen osv.

I data fra markforsøk er det sjelden signifikante avvik fra 2. grads regresjon, og vi skal derfor nøye oss med å diskutere 1. grads (lineær) og 2. grads sammenhenger mellom forsøksledd og forsøksresultat.

For å komme fram til enklest mulig koeffisienter, forutsetter vi at det er lik avstand mellom forsøksleddene. I et gjødslingsforsøk med  $N$  forutsettes det f.eks. at leddene er 8, 10, 12 og 14 kg N pr. dekar, eller at avstanden mellom leddene er 2 kg N pr. dekar.

For å beregne den første kontrasten,  $K_1$ , som uttrykker  $x$  som en lineær funksjon av  $t$ , kan vi sette opp koeffisientene,  $k_{1i}$ , som som den enklest mulige lineære funksjon av  $t$ :

$$k_{1i} = a + t_i$$

Det eneste kravet vi setter til denne hjelpeligningen er at den representerer en lineær sammenheng mellom koeffisientene og forsøksledd.

$t_i$	$k'_{1i}$	$k''_{1i}$	$k_{1i}$
0	a	-3/2	-3
1	a+1	-1/2	-1
2	a+2	1/2	1
3	a+3	3/2	3
Sum	4a+6	0	0

I vårt eksempel varierer  $t$  fra 0 til 3 med like intervall på 1. Vi forutsetter at den kombinasjon av forsøksresultatene som koeffisientene gir skal være en kontrast med 1 DF, nemlig at summen av koeffisientene skal være 0.

$$4a + 6 = 0, a = -3/2$$

Verdien  $-3/2$  settes så inn for  $a$  i  $k'_{1i}$  og vi får koeffisientene  $k''_{1i}$ . For å forenkle disse multipliserer vi alle  $k''_{1i}$  med 2 og får rekken  $k_{1i}$ . Når disse koeffisientene brukes på forsøksresultater fra et forsøk med 4 forsøksledd med lik avstand mellom leddene, er kontrasten  $K_1 = \Sigma(k_{1i}x_i)$  et uttrykk for den lineære effekten av  $t_i$  på  $x_i$ .

På tilsvarende måte kan vi sette opp koeffisienter for den kvadratiske effekten av  $t$  på  $x$ .

Den enkleste 2. grads ligning vi kan sette opp er

$$k_{2i} = a + b + t_i^2$$

I tillegg til forutsetningen om at summen av koeffisientene skal være 0 må vi nå føre inn forutsetningen om at summen av produktet av to sett koeffisienter skal være 0.

$t_i$	$k_{1i}$	$k'_{2i}$	$k_{1i}k'_{2i}$	$k_{2i}$
0	-3	a	-3a	1
1	-1	a + b + 1	-a - b - 1	-1
2	1	a + 2b + 4	a + 2b + 4	-1
3	3	a + 3b + 9	3a + 9b + 27	1
Sum	0	4a + 6b + 14	10b + 30	0

$$10b + 30 = 0$$

$$b = -3$$

$$4a + 6b + 14 = 0$$

$$4a = 18 - 14 = 4$$

$$a = 1$$

Ved å sette verdiene for a og b inn i  $k'_{2i}$ , får vi koeffisientene  $k_{2i}$ . Ved hjelp av disse koeffisientene og forsøksresultatene, kan vi regne ut kontrasten  $K_2 = \sum k_{2i} x_i$ , som er et uttrykk for kvadratisk avvik fra den lineære regresjonen av x på t.

Koeffisientene for ortogonale polynomer er samlet i tabellverk. Et lite utdrag av en slik tabell er gjengitt i tabell VII.1.

#### Estimering av regresjoner

Ved hjelp av koeffisientene for lineære og kvadratiske kontraster kan vi beregne koordinatpunkter for å framstille 1. og 2. grads regresjonen av x med hensyn på t grafisk. Vi beregner først en gjennomsnittlig lineær effekt ( $\bar{L}$ ) og en gjennomsnittlig kvadratisk effekt ( $\bar{K}$ ).

$$\bar{L} = K_1 / r \sum k_{1i}^2$$

$$\bar{K} = K_2 / r \sum k_{2i}^2$$

Tabell VII.1. Koeffisienter for 1. og 2. gradsfunksjoner for forsøk med opptil 7 trinn.

Grad	A n t a l l t r i n n									
	3		4		5		6		7	
	1.	2.	1.	2.	1.	2.	1.	2.	1.	2.
	-1	+1	-3	+1	-2	+2	-5	+5	-3	+5
	0	-2	-1	-1	-1	-1	-3	-1	-2	0
	+1	+1	+1	-1	0	-2	-1	-4	-1	-3
			+3	+1	+1	-1	+1	-4	0	-4
					+2	+2	+3	-1	+1	-3
							+5	+5	+2	0
									+3	+5
sum k <sup>2</sup>	2	6	20	4	10	14	70	84	28	84

I dette tilfelle kan vi betrakte  $r$  som en skaleringsfaktor. I beregningen av  $SS$  må vi alltid dividere oss tilbake til samme grunnenhet, f.eks. kg avling pr. rute. Kontrastene regner vi ut på forsøksleddsummene og  $r$  er da det antall observasjoner som ligger bak hver forsøksleddsum. For grafisk framstilling av resultatene vil vi oftest bruke en annen benevning, f.eks. kg avling pr. dekar. Det enkleste er da å regne ut kontrastene i kg pr. dekar og dermed blir  $r=1$ . Formlene for beregning av koordinatpunktene er:

$$X_{L_i} = \bar{X} + k_{1_i} \bar{L}$$

$$X_{K_i} = X_{L_i} + k_{2_i} \bar{K}$$

hvor  $X_{L_i}$  er punkter på en 1. grads linje og  $X_{K_i}$  er punkter på en 2. grads linje.

## VIII. FORSØKSSERIER

Unntaksvis kan et enkelt godt forsøk være tilstrekkelig til å gi oss svar på det spørsmålet vi har stilt. Men som oftest, og særlig når det gjelder forsøk som skal danne grunnlaget for en veiledning til praktikerne, kan vi ikke nøye oss med ett forsøk. Vi må utføre en rekke forsøk, en forsøksserie. Årsaken er at resultatene vil variere fra forsøk til forsøk. Mest tydelig er dette når det dreier seg om forsøk med vekster dyrket på friland.

På grunn av ulikheter i jordbunnsforhold m.m., vil resultatene være forskjellig på ulike steder, og særlig på grunn av forskjell i været vil resultatene variere fra år til år. Statistisk kan det uttrykkes slik: Vi har samspill mellom forsøksfaktorene på den ene siden og jordbunnsforholdene og klimaet på den andre.

I forsøk med husdyr er ikke dette fullt så tydelig, bl.a. fordi de som oftest lever i et kunstig klima. Men også for husdyrforsøk må vi vente forskjeller fra besetning til besetning, både p.g.a. genetisk variasjon, og variasjon i miljøforholdene.

Vi bygger derfor nesten aldri vår praktiske veiledning på resultatene fra et enkelt forsøk, men venter til vi har gjentatt forsøket på flere steder og i flere år.

### 1. Ettårige forsøksserier

Vi skal først se på hvordan modellen og forventningene for MS i variansanalysen blir hvis vi har utført  $n$  forsøk med  $r$  gjentak og  $t$  sorter i et enkelt år. Modellen for enkeltobservasjonene er da:

$$x_{ijk} = \mu + F_i + (BiF)_{ij} + s_k + (Fs)_{ik} + E_{ijk}$$

Her er  $F$  effekten av felt (tilfeldig),  $(BiF)$  er effekten av blokker innen felt (tilfeldig),  $s$  er sorteffekt (fast),  $(Fs)$  er samspillet sort  $\times$  felt, mens  $E$  er den tilfeldige feil på hver enkelt rute (kan også betraktes som samspillet sort  $\times$  blokk innen felt).



Variansanalysen med forventningene for MS blir:

Var. årsak	DF	MS	Forventning
Total	$nrt-1$	--	
Felt	$n-1$	--	
Blokker innen felt	$n(r-1)$	--	
Sorter	$t-1$	MS1	$\sigma^2 + r \sigma_{F_s}^2 + rn(K)_s^2$
Felt x Sort	$(n-1)(t-1)$	MS2	$\sigma^2 + r \sigma_{F_s}^2$
Feil innen felt	rest	MS3	$\sigma^2$

Feilen på sortsammenligningene er MS2, samspillet mellom sorter og felt. F-test for sorter blir MS1/MS2. Forutsetningen for denne testen er at vi vil gi et råd om sortvalg for det området feltene dekker, og at feltene kan betraktes som et tilfeldig utvalg av dette området. Vi ville fått den samme feilen og den samme F-testen om vi ikke hadde brukt tallene for enkeltrutene i analysen, men gjennomsnittene for hver sort på hvert felt. Ved å utføre analysen på enkeltrutene kan vi oppnå to ting. Ved F-testen MS2/MS3 får vi et uttrykk for samspill mellom sorter og steder. Hvis denne F-testen er signifikant, kan det være grunn til å undersøke om det går an å dele området opp i underområder som bør ha forskjellige råd om sortvalg. Videre kan den fullstendige variansanalysen brukes til å estimere varianskomponentene  $\sigma^2$  og  $\sigma_{F_s}^2$ . Disse kan så i sin tur brukes til å estimere hvor mange felt vi bør ha, og hvor mange gjentak vi bør ha på de enkelte felt.

Vi bør bruke den kombinasjon av antall felt og antall gjentak pr. felt som med gitte omkostninger gir minst mulig feil på sammenligningen av sortene.

Feilkvadratet på et sortgjennomsnitt er  $MS2/nr$ , eller

$$\sigma^2/nr + \sigma_{fs}^2/n$$

Hvis omkostningene ved forsøket var proporsjonale med antallruter vil det være riktig å sammenligne planer med det samme antall ruter ialt, altså konstant verdi av  $nr$ , og det er da klart at man ville få minst samlet feil om  $n$  var størst mulig og  $r$  liten. Samtidig ville antall frihetsgrader for feilen øke. Aller minst feil ville det bli om  $n = nr$ , altså  $r = 1$ , eller enkeltfelter uten gjentak.

Forutsetningen om konstante omkostninger pr. rute vil som oftest være feil. Endel av omkostningene ved et forsøk er pr. felt, en annen del pr. rute. Hvis vi kan skaffe oss et anslag over disse to delene av omkostningene og samtidig har estimert de to varianskomponentene, kan vi prøve oss fram med forskjellige kombinasjoner av antall felt og antall gjentak pr. felt for å finne den kombinasjon som med samme pris gir minst feil. Vi må samtidig huske på at feilen,  $MS2$ , blir best bestemt (flest DF) jo større  $n$  er, altså jo mindre  $r$  er.

De tre faktorene som spiller størst rolle i vurderingen av hvor mange gjentak vi skal ha pr. felt i en forsøksserie er:

1. Omkostningene pr. felt i forhold til pr. rute. Jo større dette forhold er, jo flere gjentak lønner det seg å ha i hvert forsøk. Hvis tilgjengelig forsøksmateriale, f.eks. såkorn av nye sorter, setter en absolutt grense for antall ruter ialt, er det riktig å bruke få gjentak, en eller to. (Samtidig bør vi bruke små ruter.)
2. Feilen i de enkelte forsøk i forhold til samspillet mellom forsøksledd og felt. Jo større samspill, jo færre gjentak i hvert forsøk.
3. Antall forsøksledd. Hvis forsøket har mange ledd, kan vi redusere antall gjentak. (De faste omkostninger pr. felt spiller da mindre rolle.)

Som regel bør vi prøve å få flest mulig felt i en forsøksserie, og heller redusere antall gjentak i hvert felt. Bare hvis de faste omkostningene pr. felt er svært store, kan det være riktig å bruke mer enn 3 gjentak pr. felt. To gjentak pr. felt vil oftest være tilstrekkelig, og i mange tilfeller vil vi få minst feil på gjennomsnittene ved å anlegge de enkelte felt uten gjentak. Da får vi ikke beregnet noen feil innen felt, og det er heller ikke mulig å estimere samspillet mellom forsøksledd og felt.

## 2. Flerårige forsøksserier

Hvis det er grunn til å anta at forsøksresultatene varierer fra år til år (i markforsøk vil dette alltid være tilfelle), er det ikke tilstrekkelig å ha forsøk i ett år. Aldri så mange og gode forsøk, spredt tilfeldig over et område, gir bare svar på hvordan resultatene ville vært det året. Som veiledning for praktikerne er dette av forholdsvis liten interesse. Det vi ønsker å få svar på er hvordan resultatene sannsynligvis vil bli i de kommende år. For å få statistisk holdbare svar på dette må vi utføre forsøkene i flere år, og årene må kunne betraktes som et tilfeldig utvalg av kommende år. Allerede her støter vi ofte på en stor vanskelighet. Sjøl to-tre år som følger etter hverandre kan være nokså ekstreme i klimatisk henseende, og derfor gi usikre svar på framtidige resultater. Dessuten har vi ofte en langsiktig trend som gjør at våre spådommer om framtida blir usikre.

Vi kan ha trend i de klimatiske faktorer, men vi har enda tydeligere trend i tekniske og økonomiske faktorer. F. eks. vil en overgang til sterkere gjødsling virke inn på sortsvalget. Og en øket bruk av mer stråstive kornsorter vil kunne forandre resultatene av gjødslingsforsøk. Det er derfor ikke gitt at resultatene fra en forsøksserie gjennom de siste 20 år gir et bedre svar på neste års forhold, enn om vi bare tar med de siste 10 år.

Vi skal foreløpig se bort fra disse ekstra vanskeligheter, og regne med at forsøksår er et tilfeldig utvalg. Hvis vi har en rekke felt hvert år, er det igjen to muligheter: a) Vi har felter på de samme steder (f.eks. gårder) hvert år, eller b) Vi har et nytt tilfeldig utvalg av steder hvert år. I praksis blir det gjerne en blanding av disse alternativene, noen nye og noen gamle steder hvert år. Vi skal her bare presentere modellen for enkeltobservasjonene for ulike forsøkssteder hvert år.

$$X_{ijkl} = \mu + \bar{A}_i + (Si\bar{A})_{ij} + (BiSi\bar{A})_{ijk} + s_l + (\bar{A}s)_{il} + (Si\bar{A}s)_{ijl} + E_{ijkl}$$

Vi har ingen hovedeffekt av steder eller samspill mellom steder og år, men derimot steder innen år ( $Si\bar{A}$ ). Vi har også tilsvarende for samspillet med sorter. Den interessante del av variansanalysen blir nå:

---

Var. årsak

---

sort	t-1	MS1	$\sigma^2 + r \sigma_{Si\bar{A}s}^2 + nr \sigma_{\bar{A}s}^2 + rnm(K)_s^2$
ÅR.sort	(m-1)(t-1)	MS2	$\sigma^2 + r \sigma_{Si\bar{A}s}^2 + rn \sigma_{\bar{A}s}^2$
StediÅr.sort	m(n-1)(t-1)	MS3	$\sigma^2 + r \sigma_{Si\bar{A}s}^2$
Feil innen felt	rest	MS4	$\sigma^2$

---

F-testen for sorter er MS1/MS2. Også hvis vi har felter på noen steder i flere år, må vi bruke denne modellen for variansanalysen. Vi behøver ikke å ha det samme antall felt hvert år. Regnearbeidet blir litt mer innviklet, særlig beregningen av varianskomponentene, da n ikke er konstant, men F-testen blir det samme.

Ved en slik analyse hvor vi tar med alle felter og år, vil resultatene i de enkelte år få en vekt som svarer til antall felter hvert år. Særlig hvis årene har vært svært forskjellige kan det diskuteres om dette er riktig, om ikke de enkelte år bør ha like stor vekt.

Hvis vi ønsker å gi årene lik vekt, regner vi først ut gjennomsnittene for hver sort hvert år, og foretar en toveis analyse av disse tallene over sort og år.

Ofte brukes også en annen beregningsmåte. Alle feltene beregnes ved hjelp av en toveis variansanalyse uten hensyn til hvilket årde enkelte felt stammer fra. Middeltallene for sorter blir da de samme som ved den veiete analysen foran, og MS for feil blir som om vi slår sammen DF og SS for MS2 og MS3. Hvis MS2 er tydelig større enn MS3 er dette opplagt galt. Vi har samtidig både redusert den virkelige feilen og øket antall DF for den.

### 3. Forsøk med gjentatt behandling på de samme ruter

Disse forsøk er av flere typer, men de følger stort sett den samme statistiske modell og analysemetode. Vi kan foreta behandlingen en gang, og måle resultatene en rekke ganger etter hverandre (eks. forsøk med engfrøblandinger, tynningsforsøk i skog). Vi kan også gjenta både behandlingen og målingen av resultatene en rekke ganger (eks. flerårige gjødslingsforsøk). Endelig har vi eksempler på at behandlingen blir gjentatt en rekke ganger, men at vi bare er interesserte i sluttresultatet (eks. sammenligning av ulike vaskemidler, kumulativ virkning på renhet eller slitasje).

Hvis resultatet bare blir målt en gang, blir analysen nøyaktig som for et engangsforsøk. Hvis resultatet blir målt flere ganger, kan vi utføre en flerveis variansanalyse, med gang (eller år) som en ekstra faktor.

Det er en viktig forskjell mellom denne analysen og analysen av en flerårig forsøksserie. For det første er ikke årene uav-

hengige. Samme behandling har vært utført på de samme ruter hvert år. Sjøl om vi får de samme resultatene hvert år, betyr ikke dette at det er liten feil på gjennomsnittene. Feilen blir her som ellers målt ved variasjonen mellom gjentakene. For det andre betyr en stor variasjon i resultatene fra år til år (et stort samspill mellom ledd og år), ikke nødvendigvis at resultatene er usikre. Ta som eksempel en sammenligning av ulike grasarter, hvor en art gir størst avling 1.år og minkende avling senere, mens en annen art har økende avling i de første årene. Det vi er interesserte i er sum avling over en rekke år, og variasjonen fra år til år spiller da mindre rolle. Statistisk sett er år (engår) ikke en tilfeldig, men en fast variabel. Vi er ikke interessert i variasjonen fra år til år, men i summen av avlingen over en rekke år.

I stedet for å ta år med som en faktor i analysen, kan vi derfor nøye oss med å analysere summen over alle år. I et flerårig forsøk er det ofte litt større interesse for resultatet i de første årene enn senere, bl.a. fordi vi i praksis ikke er sikre på å få nyttet avlingene i alle årene. En eng eller en jordbæråker kan bli pløyd opp etter to år selv om planen var at den skulle ligge i 4 år. Det kan derfor være grunn til å analysere avlingen 1. året, avlingen i sum for de 2 første årene, i sum for de 3 første årene o.s.v.

## IX. OPPGAVER

## Oppgave 1.

Les både a) og b) før du setter opp data.

- a) Bestem den gjennomsnittlige ordlengden (antall bokstaver pr. ord) i dette heftet. Tell opp bokstavene i 50 ord tatt ut tilfeldig, sett opp frekvensfordeling og histogram, og beregn gjennomsnitt og standardavvik.
- b) Beregn gjennomsnittet av de 10 første ordene, og de 10 neste osv. Beregn nå gjennomsnitt og standardavvik for disse 5 gjennomsnittene og sammenlign resultatene av de to beregningene.

## Oppgave 2.

Hvilke ord har flest bokstaver, de som begynner med en konsonant eller de som begynner med en vokal?

- a) Av en eller annen bok tar du ut et tilfeldig utvalg (hvordan er dette gjort?) av ord, minst 50 ialt, og teller antall bokstaver i konsonantord og vokalord hver for seg. Sett opp frekvenstabellene, beregn middeltall, standardavvik og konfidensgrenser for hver type og foreta en t-test på differensen.
- b) Differensen kan også bestemmes på en annen måte: Ta ut et tilfeldig konsonantord og det første vokalordet etter dette. Bestem differensen i ordlengde. Gjenta dette 25 ganger, og regn ut gjennomsnitt for differensene, standardavviket, middelfeil og konfidensgrenser.

Hvordan forholder standardavviket på enkeltobservasjonene (i del a) seg til standardavviket på differensene (i del b)?

## Oppgave 3.

I karrforsøk med ulike kalktyper og kalkmengder ved Institutt for Jordkultur, N.L.H., var det i alt 9 ulike forsøksledd med kalk og et kontrollledd (0) uten kalk. Forsøksveksten var korn, og totalavlingen av korn og halm (lo) i 0,1g pr. karr ble som vist i tabellen.

F o r s ø k s l e d d									
0	1	2	3	4	5	6	7	8	9
598	623	614	683	591	631	570	591	588	608
563	621	620	619	633	658	607	590	628	642
465	627	635	632	643	638	620	541	595	627
$\Sigma$ 1626	1871	1869	1934	1867	1927	1797	1722	1811	1877

- Foreta en enveis variansanalyse av alle ledd (0-9), og av de kalkete ledd (1-9).
- Sammenlign restvariansen i de to analysene og
- Kommenter resultatet.



## Oppgave 4.

Et sortforsøk i bygg med 3 gjentak ( $r=3$ ), I, II og III, og med 5 sorter ( $t=5$ ), A, B, C, D og E ga følgende avlinger i kg pr. rute a 10 m<sup>2</sup>.

B l o k k					
I		II		III	
sort	kg	sort	kg	sort	kg
A	3,46	C	2,86	B	3,28
C	2,71	B	3,51	E	3,27
D	3,14	E	2,85	C	2,42
B	3,68	A	3,56	D	2,86
E	3,06	D	3,46	A	3,42

- Utfør variansanalyse på kornavlingene.
- Sett opp en tabell over gjennomsnittsavlingene for hver sort.
- Regn ut  $LSD_{5\%}$  og CV% og diskuter resultatet.

## Oppgave 5.

Et gjødslingsforsøk med poteter ble anlagt som latinsk kvadrat. På kartet nedenfor er det på hver rute satt inn kg fullgjødning pr. dekar og kg knollavling pr. rute a 20 m<sup>2</sup>.

## K o l l o n n e

Rekke		A	B	C	D	E
1	Kg fullgj./dekar	0	10	30	40	20
	Kg knoller/rute	63	64	77	64	83
2	Kg fullgj./dekar	20	0	40	10	30
	Kg knoller/rute	72	65	70	59	83
3	Kg fullgj./dekar	30	20	10	0	40
	Kg knoller/rute	76	88	74	58	81
4	Kg fullgj./dekar	40	30	0	20	10
	Kg knoller/rute	78	83	65	53	75
5	Kg fullgj./dekar	10	40	20	30	0
	Kg knoller/rute	84	85	74	66	64

Utfør variansanalyse og beregn variasjonskoeffisienten. Beregn avlingsresultatene i kg/dekar. Sett resultatene opp i en figur og gi en tolking av dem.



## Oppgave 7.

I et forsøk med ulike ugrasmidler ble det telt opp følgende antall ugrasplanter pr. m<sup>2</sup>.

	Ledd	G j e n t a k					Sum
		I	II	III	IV	V	
Kontroll	a	350	200	470	310	260	1590
Hormon	b	12	15	10	14	14	65
---''---	c	18	12	20	16	14	80
Nitro	d	75	50	70	60	55	310
---''-	e	60	50	50	55	45	260

1. Sett opp en variansanalyse for forsøket.
2. Del opp kvadratsummen for forsøksledd etter følgende kontraster:
  - a. Kontroll mot behandling.
  - b. Hormon mot nitro.
  - c. Innen hormon.
  - d. Innen nitro.
  - e. Undersøk om disse kontrastene er ortogonale.
3. Del opp restkvadratsummen i:
  - a. En restkvadratsum for sammenligningen mellom kontroll og behandling.
  - b. En restkvadratsum for sammenligninger innen behandling.
4. Test de ulike effektene mot sine respektive restvarianser og diskuter resultatet.

## Oppgave 8.

Se kalkingsforsøket i oppgave 3. Leddene 1-9 er kalket etter denne tabellen:

Kalk- mengde	K a l k t y p e		
	Pulverisert	Finknust	Granulert
1	1	4	7
2	2	5	8
3	3	6	9

- Sett opp variansanalyse for de to faktorene kalkmengde og kalktype.
- Beregn og test noen interessante kontraster for hovedeffektene.
- Kommenter resultatene.

## Oppgave 9.

Kombinert gjødslings- og sortsforsøk i vårhvete.

3 Gjødslinger: A = uten salpeter  
 B = 20 kg salpeter pr. dekar  
 C = 40 ----- " -----

5 sorter: 1-5.

4 gjentak: I-IV

Forsøksplan: Split-plot med gjødsling på storruter.

Kart over feltet viser forsøksleddene og avling i kg korn pr. rute. Rutestørrelsen var 36 m<sup>2</sup>.

---

	I	II	III	IV
B 5	11,0	C 3 13,0	A 4 11,3	C 3 12,2
B 4	11,6	C 1 15,2	A 2 14,6	C 2 14,2
B 1	14,3	C 5 11,3	A 3 12,5	C 5 10,0
B 2	15,9	C 4 12,0	A 5 10,3	C 1 14,9
B 3	12,0	C 2 16,0	A 1 15,7	C 4 11,3
C 5	11,9	A 2 12,6	B 3 11,9	A 5 7,9
C 1	14,1	A 3 9,8	B 2 15,1	A 3 10,0
C 2	16,2	A 5 8,8	B 4 11,8	A 1 12,5
C 4	12,7	A 1 13,4	B 5 10,8	A 2 11,8
C 3	13,9	A 4 10,3	B 1 15,5	A 4 11,0
A 2	14,1	B 5 8,9	C 3 13,2	B 1 14,3
A 4	10,7	B 1 13,5	C 5 11,3	B 3 11,1
A 5	8,7	B 4 10,8	C 1 15,8	B 2 11,6
A 1	12,4	B 3 12,2	C 2 16,1	B 4 10,0
A 3	10,4	B 2 13,6	C 4 12,8	B 5 8,0

---

- Sett opp en variansanalyse for forsøket.
- Regn ut  $m$  og  $LSD_{5\%}$  for sammenligningen mellom sorter.
- Del opp variansen for gjødsling i linær og kvadratisk effekt og sett opp resultatene grafisk.
- Sett opp en tabell over gjennomsnitt for de signifikante effektene og kommenter resultatene.

## Oppgave 10.

I en forsøksserie med vårhvete, 5 sorter, 10 felt, 2 gjentak pr. felt, ble det registrert følgende kornavlinger i kg pr. dekar.

Sort	Gjentak	F e l t n r.									
		1	2	3	4	5	6	7	8	9	10
1 Runar	I	278	431	344	362	384	529	327	448	359	256
	II	193	392	254	335	457	468	308	497	372	283
2 Reno	I	259	531	341	355	463	527	330	423	368	263
	II	155	445	270	404	478	516	343	448	338	273
3 T68038	I	322	526	417	394	482	583	409	530	444	311
	II	208	447	308	387	420	520	319	595	396	269
4 T69027	I	325	523	397	387	434	600	371	529	432	254
	II	192	524	446	427	446	552	362	567	417	364
5 Sv70505	I	257	470	394	383	450	536	386	493	415	334
	II	205	499	372	414	411	496	321	513	425	429

1. Utfør en variansanalyse av forsøksserien og kommenter resultatene. Regn ut  $LSD_{5\%}$  for forskjellen mellom sorter.
2. Hvor mange felt med to gjentak måtte vi hatt for at avlingsdifferenser på 20 kg/dekar skulle blitt signifikante ( $P=5\%$ ).
3. Hvor stor ville LSD anslagsvis blitt om vi hadde lagt an 20 forsøk uten gjentak, istedenfor 10 forsøk med 2 gjentak?

## Oppgave 11.

I en forsøksserie med 4 grasarter og 2 såmengder over 2 engår og 4 forsøksfelt ble det registrert følgende avlinger i kg tørrstoff pr. dekar. Hvert enkelt felt var anlagt med såmengder på stor-ruter og arter på småruter.

Felt	Art	1. engår			2. engår			Feltsum
		såmengde		Sum	såmengde		Sum	
		1	2			1		2
1	Raigras	913	971		996	893		
	Timotei	846	723		746	690		
	Engsvingel	558	742		994	990		
	Hundegras	860	700		1327	1192		
	Sum	3177	3136	6313	4063	3765	7828	14141
2	Raigras	1162	1200		1451	1667		
	Timotei	832	1043		1696	1545		
	Engsvingel	830	922		1500	1504		
	Hundegras	1236	1279		2021	1665		
	Sum	4060	4444	8504	6668	6381	13049	21553
3	Raigras	1618	1452		829	819		
	Timotei	1330	1319		991	1011		
	Engsvingel	1353	1243		959	830		
	Hundegras	1612	1791		1251	1050		
	Sum	5913	5805	11718	4030	3710	7740	19458
4	Raigras	1066	1254		464	626		
	Timotei	1129	1174		639	733		
	Engsvingel	996	1131		845	814		
	Hundegras	1365	1281		921	1204		
	Sum	4556	4840	9396	2869	3377	6246	15642
Sum	Raigras	4759	4877	9636	3740	4005	7745	
	Timotei	4137	4259	8396	4072	3979	8051	
	Engsvingel	3737	4038	7775	4298	4138	8436	
	Hundegras	5073	5051	10124	5520	5111	10631	
	Sum	17706	18225	35931	17630	17233	34863	
Sum over 2 engår:		1. såmengde		2. såmengde	Artsum			
	Raigras	8499		8882	17381			
	Timotei	8209		8238	16447			
Sum	Engsvingel	8035		8176	16211			
	Hundegras	10593		10162	20755			
	Sum	35336		35458	70794			

a. Utfør en variansanalyse for forsøksserien.

b. Sett opp tabeller for og kommenter de signifikante effektene.



**Oppgave 12.**

Velg en vekst du kjenner og sett opp en forsøksplan for følgende forsøksspørsmål:

A: 4 sorter

B: 2 såmengder eller planteavstander

C: 3 gjødseltrinn

- a. Tegn forsøksplan. Angi rutestørrelser og mengder/avstander av faktor B og C.
- b. Sett opp variansanalysen med antall frihetsgrader og forslag til testing av hovedeffekter og samspill.

