

NORGES LANDBRUKSHØGSKOLE

Institutt for plantekultur

FORELESNINGER I FORSØKSMETODIKK

av

Kåre Ringlund og Øivind Nissen

F O R O R D

Disse forelesningene i forsøksmetodikk forutsetter grunnkunnskaper i matematisk statistikk. Leserne må være kjent med begrep som gjennomsnitt, varians, standardavvik, korrelasjon, t-test og F-test.

Forfatterne vil understreke at faget forsøksmetodikk læres best gjennom øvelser og ikke bare gjennom lesing. Det er utarbeidet et eget øvelseshefte for kurset PK7.

Ås - NLH , den 20.2.1980

Kåre Ringlund

Øivind Nissen

I N N H O L D

Kapitel		Side
I	Innledning	1
II	Numeriske data	2
III	Repetisjon av matematisk statistikk	4
	a. Univers og prøver	4
	b. Statistiske formler og terminologi	5
	c. Hypoteser og hypotesetesting	10
	1. Forskjeller mellom univers	13
	2. Forskjeller mellom prøver	13
	3. t-test	13
	4. Konfidensgrenser	14
	5. Testing av en nullhypotese	14
	6. Minste signifikante forskjell	14
	7. F-test	15
	8. Variasjonskoeffisienten	15
	9. Signifikansnivå	16
IV	Statistiske modeller	17
	a. Estimering av modellkomponenter	18
	b. Fixed og random effekter	21
V	Forsøksplaner	22
	a. Forsøksspørsmål	22
	b. Forsøksenhet	22
	c. Valg av forsøksplan	23
	d. Fullstendig tilfeldig fordeling	23
	e. Blokkforsøk	24
	f. Latinsk kvadrat	26
	g. Faktorielle forsøk	27
	1. Tilfeldig fordeling av alle kombinasjonene innen gjentak	29
	2. Split-plot forsøk	31
	3. Forsøksserier	33
VI	Orthogonale kontraster	38
	a. Orthogonale koeffisienter	40
	b. Orthogonale polynomer	42
	1. Utregning av orthogonale koeffisienter	43
	2. Beregning av lineære og kvadratiske effekter	45

Kapitel		Side
VII	Forutsetninger for variansanalyser	46
	a. Tilfeldig fordeling av forsøksfeilen	46
	b. Homogen varians	46
	c. Normal fordeling	47
	d. Additivitet	47
	e. Transformasjoner	48
VIII	Confounding, sammenblanding av effekter	49
IX	Variasjonsårsaker i markforsøk	50
	a. Rutestørrelse	50
	b. Rutenes form	51
	c. Antall gjentak	51

I. INNLEDNING

Det er to prinsipielt forskjellige måter å samle erfaringer på. Den enkleste, eldste og kanskje enda mest utbredte metoden er observasjoner av individer eller biologiske systemer i sitt naturlige miljø uten forsøk på påvirkning fra observatørens side. Det meste av vår viten innenfor naturvitenskapene er samlet på denne måten. Darwins utviklings teori bygger på observasjoner av planter og dyr i sitt naturlige miljø.

Den andre måten å skaffe seg erfaring på er ved å anlegge et forsøk, et eksperiment.

Mendel brukte eksperimentmetoden. Han skapte sitt eget plantemateriale ved å krysse sammen planter med ulike egenskaper for å studere spaltingsforholdet mellom disse egenskapene hos avkommet.

Mendel holdt seg til kvalitative karakterer. Det vil si karakterer som kan variere mellom to eller noen få alternative verdier. Dette kurset skal behandle metodikk som er laget for å studere kvantitative karakterer.

Direkte observasjoner i et naturlig miljø kan gi nøykatige informasjoner om sluttresultatet, men det kan være vanskelig å si noe om årsakssammenhengen. Det er f.eks. lett å se at to trær har ulik størrelse, men det er vanskeligere å slå fast om dette skyldes ulik alder eller ulik veksthastighet.

Ulik kornavling på to jordstykker er også lett å konstantere, men det er vanskelig uten videre å si om avlingsforskjellen skyldes jordart, gjødsling, vanntilgang, sort, såtid eller andre faktorer.

Når vi anlegger forsøk, eliminerer vi så godt vi kan de variasjonsårsakene som ikke inngår i forsøksspørsmålet, slik at resultatet gir et best mulig svar på hva det spørres om. Spørsmålene kan være begrenset til enkeltfaktorer som i sortsforsøk eller i gjødslingsforsøk med en bestemt gjødseltype, men ofte er effekten av en faktor avhengig av andre faktorer. Virkningen av nitrogen på grasavlinger er avhengig av både kalium- og fosfortilstanden i jorda. For å finne ut hva som er optimal N-gjødsling må forsøket derfor anlegges med varierende mengder av både N, P og K. Slike forsøk kalles faktorielle forsøk.

For å få et tall for effekten av de ulike forsøksledd, eller behandlinger ville det være nok med en rute eller en observasjon av hvert forsøksledd. Ved å anlegge flere ruter eller gjentak av hvert forsøksledd blir effekten av de ulike forsøksledd sikrere bestemt. Videre blir det mulig å regne ut et mål for variasjonen innen forsøksledd.

Variasjonen mellom forsøksledd i forhold til variasjonen innen forsøksledd gir et uttrykk for hvorvidt de observerte forskjellene kan skyldes tilfeldig variasjon eller om det virkelig er forskjell mellom forsøksleddene.

II. NUMERISKE DATA

Den første forutsetning for å kunne behandle forsøksdata med statistiske metoder er at de foreligger i en eller annen numerisk form. Observasjoner fra markforsøk er vanligvis numeriske observasjoner.

Det er imidlertid en del data som ikke er numeriske. Ett eksempel er sjukeangrep på en plante. Den enkleste numeriske skala som kan lages er en skala med to verdier, 0 for friske planter og 1 for syke planter. En slik skala ville være lite hensiktsmessig fordi den ikke skiller mellom en plante som så vidt er angrepet og en plante som er ødelagt av sykdommen. En skala med verdiene 1, 2 og 3 for henholdsvis litt angrepet, middels angrepet og sterkt angrepet, og 0 for friske planter ville være mer hensiktsmessig. Uansett hvor nøyaktig en slik skala bygges opp, vil sjelden verdiene stå i et lineært forhold til skaden sykdommen har gjort. På skalaen fra 0 - 3

er det vanskelig å få avstanden mellom intervallene lik slik at $3 - 2 = 1 - 0$. I praksis er likevel slike konstruerte numeriske skalaer et brukbart hjelpemiddel til å måle variasjon i en karakter.

En del karakterer registreres som prosenter eller relative tall. Disse må behandles med forsiktighet. Gjennomsnittet av prosenter og prosenter av gjennomsnittlige originalverdier kan være to forskjellige ting. Et lite eksempel vil illustrere dette.

Tabell II 1. Absolutte og relative skalaer.

Sted	Avling i kg korn		A i %	B i %
	Sort A	Sort B	av B	av A
I	200	300	67	150
II	300	200	150	67
\bar{x}	250	250	108,5	108,5

Sjøl om sortene A og B i gjennomsnitt har gitt akkurat like stor avling, har begge gitt 8,5% mer enn den andre hvis det regnes gjennomsnitt av relative verdier.

Data fra et forsøk er ofte gitt i kg rått materiale pr. rute og f.eks. dato for aksskyting og modning. Sluttresultatet skal presenteres som kg tørrstoff pr. dekar og dager fra såing eller fra veksttida startet fram til aksskyting og modning. De observerte data må derfor regnes om ved hjelp av dekarfaktor og vannbestemmelser og ved å legge til et fast antall dager til de observerte aksskytings- og modningsdata.

Det er av og til nødvendig å regne om, transformere, data for at de skal passe bedre til statistisk behandling. Dette skal vi komme tilbake til i et senere kapitel.

III REPITISJON AV MATEMATISK STATISTIKK

I faget forsøksmetodikk er matematisk statistikk et nødvendig hjelpemiddel. Det er særlig viktig å ha klart for seg begrep som univers og prøver eller utvalg, utregning av gjennomsnitt og varians, og testing av hypoteser. Vi skal her gi en kort repetisjon av disse områdene fra matematisk statistikk.

III a Univers og prøver

Univers betyr generelt "det hele", men i matematisk statistikk har univers heller betydningen "en helhet", - en samling observasjoner som har en eller annen felles karakteristikk. Etter denne definisjonen kan det eksistere flere univers, og et stort univers kan være delt opp i flere subunivers.

Vektene av potetene i en potetbinge med en bestemt sort kan være et univers, mens knollvektene i en binge med en annen potetsort kan være et annet univers. Tørrstoff avling pr. dekar på dyrket jord i Norge er et univers, mens tørrstoffavling av gras på dyrket jord i Øst-Norge er et subunivers av dette.

I en del tilfelle, som i eksemplet med potetbingene, er det mulig å måle hele universet. I andre tilfelle er universet uendelig eller så stort at det i praksis er umulig å bestemme alle verdiene.

Under planlegging av en forsøksserie er det viktig å være klar over hvilket

univers resultatene skal gjelde for. De enkelte forsøk eller forsøksruter må være en tilfeldig prøve av nettopp dette universet.

For at en prøve skal representere universet må den være tatt ut tilfeldig. I det ligger at de observerte verdiene er tatt ut tilfeldig blant alle mulige verdier i det universet forsøksresultatene skal gjelde for. Først hvis alle mulige verdier har like stor sjanse til å bli med i prøven er det mulig å si noe om universet på grunnlag av prøven. Et eksempel fra universet av poteter i en bingje vil illustrere dette.

En prøve av de største eller de minste potetene ville ikke være egnet til å karakterisere knollvekten i bingen. I begge tilfelle ville både gjennomsnittlig vekt og variasjon gi et feilaktig bilde av universet. En prøve av middels store knoller kunne gi et noenlunde riktig bilde av gjennomsnittlig vekt, men variasjonen ville bli undervurdert.

Bare en tilfeldig uttatt prøve hvor både store, middels og små knoller har like stor sjanse til å bli med, vil gi et riktig bilde av både gjennomsnitt og variasjon.

Selv for et så enkelt tilfelle som eksemplet ovenfor, er det vanskelig å ta ut en tilfeldig prøve. Enda vanskeligere er det å skaffe seg en tilfeldig prøve hvis universet er f.eks. dyrket jord på Vestlandet. I praksis er det så mange ting som har betydning for hvor et forsøk skal plasseres at regelen om en tilfeldig prøve nesten aldri er tilfredsstillt. Det er derfor viktig at den som utfører forsøkene er klar over dette forholdet og viser den nødvendige forsiktighet ved bruken av resultatene. Resultatene har nemlig gyldighet bare for det univers observasjonene kunne vært en tilfeldig prøve fra.

III b. Statistiske formler og terminologi

De viktigste statistiske parametrene for å beskrive en rekke tall er det aritmetriske gjennomsnittet, \bar{x} , og variansen, s^2 . Gjennomsnittet er summen av alle observasjoner dividert med antallet.

$$\bar{x} = (x_1 + x_2 + x_3 + \dots + x_n) / n = \sum x/n,$$

og variansen er kvadratsummen av differansen mellom de enkelte observasjonene og gjennomsnittet dividert med antallet minus 1.

$$s^2 = ((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2) / (n - 1)$$
$$= \sum (x - \bar{x})^2 / (n - 1)$$

Telleren i formelen for variansen kalles kvadratsummen og nevneren, $n - 1$, antall frihetsgrader. De engelske betegnelsene på henholdsvis varians, kvadratsum og frihetsgrader er Mean Square, Sum of Squares og Degrees of Freedom, som forkortes MS, SS og DF

$$MS = SS/DF$$

I engelske og i stor utstrekning også i norske forsøksmeldinger blir de engelske forkortelsene brukt, og de vil derfor også bli nyttet i dette kurset.

Beregningsmessig er det tungvint å ta differansen mellom de enkelte observasjonene og gjennomsnittet og å kvadrere og summere for å finne kvadratsummen. Dessuten vil gjennomsnittet oftest være beheftet med en avrundingsfeil. Kvadratsummen kan uttrykkes på en annen måte som er mye lettere å beregne.

$$SS = \sum (x - \bar{x})^2$$
$$= \sum (x^2 - 2x\bar{x} + \bar{x}^2)$$
$$= \sum x^2 - 2\bar{x}\sum x + n\bar{x}^2$$
$$= \sum x^2 - \frac{2(\sum x)^2}{n} + n \left(\frac{\sum x}{n}\right)^2$$
$$= \sum x^2 - \frac{(\sum x)^2}{n} \text{ eller } \sum x^2 - n\bar{x}^2$$

Etter denne formelen består kvadratsummen av to ledd, nemlig summen av kvadratene av de enkelte observasjonene og kvadratet av summen dividert med antallet. Det siste leddet heter på engelsk Correction Term som forkortes CT. Også denne engelske forkortelsen vil bli brukt i dette kurset.

I en variansanalyse av et tallmateriale hvor alle observasjonene inngår i alle summeringer er CT den samme for alle kvadratsummene som regnes ut direkte.

Kvadraten av variansen kalles standardavviket og betegnes med s .

$$s = \sqrt{MS}$$

I de tilfellene at variansen er et mål for tilfeldig variasjon, er standardavviket et mål for feilen på en enkeltobservasjon.

Formålet med forsøk er blant annet å bestemme differansen mellom forsøksbehandlingene. Det er da av spesiell interesse å kjenne feilen på en differanse mellom to tall. Siden det er vanlig med flere observasjoner av hvert forsøksledd, er det også av interesse å kjenne feilen på gjennomsnittet, og endelig feilen på differansen mellom to gjennomsnitt.

La utgangspunktet være to tilfeldige tallrekker $x_1 x_2 \dots x_n$ og $y_1 y_2 \dots y_n$ og differansen $x - y$.

x_1	y_1	$x_1 - y_1$
x_2	y_2	$x_2 - y_2$
x_3	y_3	$x_3 - y_3$
.	.	.
.	.	.
.	.	.
.	.	.
x_n	y_n	$x_n - y_n$

Gjennomsnitt av differansen blir $\overline{(x - y)} = \frac{\sum(x - y)}{n} = \frac{\sum x - \sum y}{n} = \bar{x} - \bar{y}$

Kvadratsummen for differansen blir

$$\begin{aligned}
 SS &= \sum [(x - y) - (\bar{x} - \bar{y})]^2 = \sum (x - y - \bar{x} + \bar{y})^2 = \sum [(x - \bar{x}) - (y - \bar{y})]^2 \\
 &= \sum [(x - \bar{x})^2 - 2(x - \bar{x})(y - \bar{y}) + (y - \bar{y})^2] \\
 &= \sum (x - \bar{x})^2 - 2 \sum (x - \bar{x})(y - \bar{y}) + \sum (y - \bar{y})^2
 \end{aligned}$$

under forutsetning av at det ikke er noen sammenheng, korrelasjon, mellom x og y vil

$$2 \sum (x - \bar{x})(y - \bar{y}) = 0$$

og $SS_{(x-y)} = SS_x + SS_y$

Det følger av dette at

$$MS_{(x-y)} = MS_x + MS_y$$

og $s_{(x-y)} = \sqrt{MS_x + MS_y}$

Under forutsetning av at

$$MS_x = MS_y \text{ er}$$

$$s_{(x-y)} = s\sqrt{2}$$

Kvadratsummen og variansen for en differanse er altså summen av kvadratsummene og variansene for de variablene det er tatt differanser av, og standardavviket er kvadratroten av summen av enkeltvariansene. I en variansanalyse er en av forutsetningene at den tilfeldige variasjonen er den samme for de ulike forsøksledd. Hvis x og y er to forsøksledd skal altså MS_x være lik MS_y og standardavviket på differansen mellom x og y blir da et felles standardavvik på enkeltobservasjonene multiplisert med kvadratroten av 2.

På tilsvarende måte blir

$$\overline{(x + y)} = \bar{x} + \bar{y}$$

$$SS_{(x+y)} = SS_x + SS_y \text{ eller } 2 SS$$

$$MS_{(x+y)} = MS_x + MS_y \text{ eller } 2 MS$$

$$s_{(x+y)} = s\sqrt{2}$$

Gjennomsnittet av en sum er det samme som summen av gjennomsnittene, og variansen for en sum er summen av variansene for enkeltobservasjonene.

Vi skal nå undersøke hvordan multiplikasjon av observasjonene med en fast faktor virker på gjennomsnitt og varians.

Utgangspunktet er at en rekke observasjoner er multiplisert med en fast faktor, C.

$$Cx_1$$

$$Cx_2$$

$$Cx_3$$

.

.

.

.

$$Cx_n$$

$$\bar{Cx} = \frac{\sum Cx}{n} = C \frac{\sum x}{n} = C\bar{x}$$

$$\begin{aligned} SS_{Cx} &= \sum (Cx - \bar{Cx})^2 = \sum [C(x - \bar{x})]^2 \\ &= C^2 \sum (x - \bar{x})^2 = C^2 SS_x \end{aligned}$$

Gjennomsnitt av en tallrekke hvor alle observasjonene er multiplisert med en fast faktor, er den faste faktoren multiplisert med gjennomsnittet av observasjonene. Kvadratsummen av samme tallrekke er kvadratsummen for enkeltobservasjonene multiplisert med kvadratet av den faste faktoren. Siden antall frihetsgrader er det samme for originalobservasjonene og observasjonene multiplisert med den faste faktor er

$$MS_{Cx} = C^2 MS \text{ og } s_{Cx} = C s_x$$

Restvariansen i en enkel variansanalyse er et felles mål for feilvariansen for alle observasjonene i forsøket. Hvis forsøket er anlagt med r gjentak vil derfor feilen (standardavviket) for summen av forsøksleddene være

$$s_{\Sigma x} = s\sqrt{r}$$

og feilen på gjennomsnittet, middelavviket, er

$$s_{\bar{x}} = \frac{1}{r} s\sqrt{r} = s/\sqrt{r} = m$$

Endelig er feilen på en differens mellom to gjennomsnitt

$$m_{\bar{x}_1} - \bar{x}_2 = s\sqrt{2/r}$$

III c. Hypoteser og hypotesetesting.

Bakgrunnen for å anlegge forsøk er ønsket om å teste en bestemt hypotese. Om forsøket gjelder ulike arter og sorter eller om det gjelder gjødsling, jordarbeiding, vanning, grøfting eller sprøyting, er utgangspunktet en eller annen hypotese om hvordan forsøksbehandlingene vil virke på resultatet. Disse utgangshypotesene går gjerne ut på at det er forskjeller mellom behandlingene. Forsøket anlegges for å finne ut hvilken behandling som gir det beste resultatet.

En hypotese om at det er forskjeller mellom forsøksledd er vanskelig å teste fordi den også må inneholde eksakte utsagn om hva forskjellene går ut på. Som arbeidshypoteser brukes derfor nullhypoteser som går ut på at det ikke er noen forskjell mellom forsøksleddene. Dette er en eksakt formulert hypotese. De statistiske tester går ut på å finne ut om forsøksresultatene gir grunnlag for å godta eller for å forkaste nullhypotesen.

Fordelingsfunksjoner.

Et eksempel på et univers var observasjoner for vekten av potetene i en potetbinge. Disse observasjonene kan ordnes etter vekt og antall poteter i hver vektklasse kan telles opp. Dette gir en frekvensfordeling for knollvekt.

Anta at vi har veid 20 knoller fra bingje A og 100 knoller fra bingje B og funnet frekvensfordelingene som vist i tabell III 2.

Tabell III 2. Frekvensfordeling for knollvekt i to potetbingjer.

		Knollvekt i gram											
		0,1	20,1	40,1	60,1	80,1	100,1	120,1	140,1	160,1	180,1	200,1	
		-20,0	-40,0	-60,0	-80,0	-100,0	-120,0	140,0	160,0	180,0	200,0	220,0	
Bingje	\bar{x}	10	30	50	70	90	110	130	150	170	190	210	
A		2	6	3	3	2	2	1	1				
B		4	3	8	20	18	17	15	10	3	1	1	

Histogram for de to frekvensfordelingene er vist i figur III 1. Fordi prøvestørrelsen var 20 i bingje A og 100 i bingje B gir ikke histogrammene noe godt bilde av forskjellene mellom de to bingjene. Ved å uttrykke frekvensfordeling i bingje A i prosent blir fordelingene sammenlignbare. I vårt eksempel brukte vi klassebredder på 20 gram. Ved å minske klassebreddene og øke prøvestørrelsen vil histogrammene gå over til jevne kurver, og slike kurver kalles fordelingskurver.

En slik fordelingskurve som går fra minus uendelig til pluss uendelig, er symmetrisk og med en bestemt form kalles den normale fordelingskurve og kan beskrives ved hjelp av den normale fordelingsfunksjon. Den normale fordelingsfunksjon kan bestemmes ved hjelp av forventningen, eller gjennomsnittet, og standardavviket. Hvis forventningen, μ , for en normal fordelingsfunksjon er 0, vil 2,5% av observasjonene ligge utenfor grensene $\pm 1,96\sigma$ hvor σ er standardavviket. For samme fordeling ligger 32% av observasjonene utenfor grensene $\pm \sigma$. Fordi både μ og σ er nødvendige for å beskrive en bestemt normalfordeling, kan forskjeller mellom to ulike fordelinger skyldes at forventningene eller standardavviket eller begge to er forskjellige.

Observasjoner som vi gjør i et biologisk system passer aldri helt inn i den normale fordelingsfunksjon sjøl om vi tar en meget stor prøve. I en del tilfeller vil vi til og med finne en flertoppet fordelingskurve. Hvis vi i stedet for å sette opp en frekvensfordeling for enkeltobservasjoner setter opp en frekvensfordeling over gjennomsnitt av prøver på f.eks. 10 observasjoner, vil denne være tilnermet normalfordelingen sjøl om fordelingen for enkeltobservasjoner ikke er det.

Figur III 1. Histogram for frekvensfordelingen i tabell III 2.

Ant.

8

4

10 30 50 70 90 110 130 150 170 180 210

Binge A

28

24

20

16

12

8

4

10 30 50 70 90 110 130 150 170 190 210

Binge B

28

24

20

16

12

8

4

10 30 50 70 90 110 130 150 170 190 210

Binge A 1. 2

1. Forskjeller mellom univers.

I vårt eksempel med potetbingene er det umulig å måle hele universet, det vil si å veie alle knollene i begge bingene. Ut fra disse observasjonene kunne vi regne ut gjennomsnitt og standardavvik og vi kunne uttale oss med 100% sikkerhet om vekten av knollene i de to bingene.

De langt fleste univers er så store at det ikke er mulig å måle hele universet.

I praktisk veiledning er vi oftest interessert i å uttale oss om et univers som inkluderer kommende år. Uttalelser om universet må derfor bygge på en eller flere prøver.

2. Forskjeller mellom prøver.

Hvis vi har en prøve og ikke på forhånd vet hvor den er tatt ut, kan vi bruke sannsynlighetene fra den normale fordelingsfunksjon til å vurdere om prøven "med rimelighet" kan være tatt ut av en kjent fordeling. Vi har en normal fordeling med forventning 100 og standardavvik 10 og får spørsmål om en observasjon på 111 kan være tatt ut av denne fordelingen. Da avviket fra er bare 11 mens $\sigma = 10$ må vi svare ja til dette. Det er jo omtrent 30% sannsynlighet for at en enkeltobservasjon skal avvike så mye fra forventningen i en eller annen retning. Hadde observasjonen vært 70, ville avviket vært 30, eller 3σ . Heller ikke da kunne vi med 100% sikkerhet sagt at denne observasjonen ikke var tatt ut av fordelingen, men vi kunne sagt at det var lite rimelig. Det ville være mindre enn 1% sannsynlighet for å finne en så ekstrem verdi.

I praksis kjenner vi ikke μ og σ men må erstatte dem med \bar{x} og s . \bar{x} og s er våre estimat av μ og σ .

3. t-test.

Sjefsbryggeren ved det kjente engelske bryggeriet Guinness, som skrev avhandlinger under pseudonymet "Student", utarbeidet frekvensfordeling for $(x - \mu)/s$ hvor s er standardavviket for x . Denne størrelsen kalte han t . Fordelingsfunksjonen for t har fått navnet Student's t -fordeling. Den ligner den normale fordelingsfunksjon, men endrer seg med antall frihetsgrader for s . Når antall frihetsgrader er uendelig, faller t -fordelingen og normalfordelingen sammen. Når antall observasjoner avtar, blir bestemmelsen av \bar{x} og s mer unøyaktig og grensene for 5% nivået fjerner seg fra gjennomsnittet. For den normale fordelingsfunksjonen var grensen $\pm 1,96\sigma$. For t -fordelingen med 10 frihetsgrader er grensen $\pm 2,23\sigma$ og for 1 frihetsgrad er grensen $\pm 12,71\sigma$.

Argumentet på forrige side om hvordan vi kan bedømme om en observasjon er tatt ut av en kjent fordeling, er sjelden aktuelt fordi vi ikke kjenner μ og σ .

4. Konfidensgrenser.

Vi skal nå snu resonnementet. Ut fra en prøve skal vi beregne grenser som vi tror μ vil ligge innenfor. Dette kalles konfidensgrenser.

Vi går ut fra definisjonen av t som er

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{m}}}$$

Ut fra t-tabellen finner vi $t_{2,5\%}$ som betyr at det er 5% sjanse til å finne en t som enten er større enn $+t_{2,5\%}$ eller mindre enn $-t_{2,5\%}$.

Det er dermed 95% sannsynlighet for at

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

Løsningen av disse ulikhetene gir

$$\bar{x} - t_{2,5\%} \frac{s}{\sqrt{m}} \leq \mu \leq \bar{x} + t_{2,5\%} \frac{s}{\sqrt{m}}$$

Dette betyr at det er 95% sannsynlighet for at utsagnet om at μ ligger mellom grensene $\bar{x} \pm t_{2,5\%} \frac{s}{\sqrt{m}}$ er riktig.

5. Testing av en nullhypotese.

t-testet kan også brukes til å teste hypoteser om μ . En nullhypotese går normalt ut på at $\mu = 0$, og formelen for t forenkles da til

$$t = \frac{\bar{x}}{\frac{s}{\sqrt{m}}}$$

Hvis t-verdien vi finner etter denne formelen er større enn $t_{2,5\%}$ i t-tabellen, forkaster vi nullhypotesen. Vi sier at \bar{x} er signifikant forskjellig fra 0 på 5% nivået.

6. Minste signifikante forskjell.

t-testet kan også brukes til å regne ut den "minste signifikante forskjell" mellom to gjennomsnitt. Denne forkortes gjerne LSD etter det engelske "Least Significant Difference".

Utgangspunktet er en forskjell, $\bar{x}_1 - \bar{x}_2$, med tilhørende middelfeil, $m \cdot 2$.

$$t = (\bar{x}_1 - \bar{x}_2) / m \cdot \sqrt{2}$$

Hvis forskjellen skal være signifikant på 5% nivået, må $t \geq t_{2,5\%}$

$$(\bar{x}_1 - \bar{x}_2) / m \sqrt{2} \geq t_{2,5\%}$$

$$\bar{x}_1 - \bar{x}_2 \geq t_{2,5\%} \cdot m \cdot \sqrt{2}$$

Den minste signifikante forskjell er altså $t_{2,5\%} \cdot m \cdot \sqrt{2}$. $t_{2,5\%}$ tas ut i t-tabellen for det antall frihetsgrader som m har.

7. F-test.

Den mest brukte testmetoden i forsøksmetodikken er oppkalt etter den engelske statistikeren R.A. Fisher og har fått navnet F-test. F er definert som forholdet mellom to varianser.

$$F = MS_1 / MS_2$$

Under forutsetning av at de to variansene er tatt ut av samme populasjon er F fordelt etter en fordelingsfunksjon som er avhengig av antall frihetsgrader som ligger bak de to variansene. Der er utregnet F-tabeller med verdier som er slik at sannsynligheten for å finne en F som er større enn tabellverdien er f.eks. 0,05, 0,01 eller 0,001

8. Variasjonskoeffisienten

Et nyttig mål for å se på nøyaktigheten i et forsøk er variasjonskoeffisienten, CV%, som er kvadratroten av restvariansen i prosent av totalgjennomsnittet

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

9. Signifikansnivå.

Signifikansnivå velges etter forsøksmaterialet. I markforsøk er det oftest tilstrekkelig å teste på 5% nivået. Dette betyr at de slutningene som trekkes kan være feil i 5% av tilfellene. I mange tilfelle ville sikkert en praktiker ta i bruk en ny sort sjøl om det bare var 75%-sannsynlighet for at sorten var bedre enn gamle sorter.

I annen forskning, f.eks. innenfor medisinen, vil en kreve langt større sikkerhet. Signifikansnivå er derfor avhengig av hvilke konsekvenser en feilvurdering vil ha.

IV STATISTISKE MODELLER

Resultatene fra hver enkelt forsøksrute kan settes opp som en modell med komponenter for hovedgjennomsnittet, effekten av forsøksfaktorene, ulike samspill og en tilfeldig variasjon. For en forsøksplan med fullstendig tilfeldig fordeling av forsøksfaktorene vil modellen være

$$x_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

x_{ij} representerer forsøksresultatet, μ er forventningen for gjennomsnittet, α_i er effekten av det i -te leddet av forsøksfaktoren og ϵ_{ij} er et uttrykk for den tilfeldige variasjonen. Summen av α_i er 0 og forventningen for ϵ_{ij} er 0.

Med utgangspunkt i modellen kan de teoretiske forventningene for variansene i variansanalysen beregnes. For enveisgruppering blir de forventede variansene for forsøksledd og rest:

Variansårsak	DF	MS
Ledd	$k - 1$	$\sigma^2 + \frac{r}{k-1} \sum \alpha_i^2$
Rest	$(l-1)(r-1)$	σ^2

Den tilfeldige variable ϵ_{ij} har forventning 0 og varians σ^2 .

Det er denne variansen som estimeres ved MS_{rest} . Variansen for ledd,

MS_{ledd} , er sammensatt av et annet estimat for den tilfeldige variansen

pluss en komponent som er avhengig av variasjonen mellom forsøksleddene.

Da forsøksleddene vanligvis er fastsatt av den som anlegger forsøket, er ikke

forsøksledd en tilfeldig variabel og målet for variasjon mellom forsøksledd

er heller ikke en varians. Dette målet for variasjon kan likevel betraktes

som en varianskomponent i F-testet. Nullhypotesen går ut på at denne

variasjonskomponenten er 0 og under denne forutsetning er forventningen

for F lik 1. Den f-verdien som regnes ut i variansanalysen er:

$$F = MS_{ledd} / MS_{rest} = \left(\sigma^2 + \frac{k}{k-1} \sum \alpha_i^2 \right) / \sigma^2 = 1 + \frac{r \sum \alpha_i^2}{(k-1) \sigma^2}$$

En signifikant F betyr derfor at $\sum \alpha_i^2$ er forskjellig fra null.

IV a Estimering av modellkomponenter

For å få et mer praktisk grep på hva de ulike komponentene står for skal vi estimere dem fra forsøksdata. I et blokkforsøk med 5 sorter og 3 gjentak er modellen for den enkelte observasjon

$$x_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

hvor μ er forventningen for hovedgjennomsnittet, α_i er effekten av den enkelte sort og β_j er effekten av blokkene. ϵ_{ij} er som før et mål for den tilfeldige variasjonen eller feilen i forsøket. Estimatene av μ , α_i , β_j og ϵ_{ij} , er henholdsvis \bar{x} , A_i , B_j og E_{ij} .

Tabell IV 1. Kg tørrstoff pr. rute fra et forsøk med 5 potetsorter og 3 gjentak.

	Gjentak			\bar{x}
	j = I	j = II	j = III	
i = 1	6,40	6,43	6,40	6,41
i = 2	5,66	5,4	4,62	5,23
i = 3	6,84	6,52	6,19	6,52
i = 4	6,67	6,72	6,15	6,51
i = 5	5,12	5,58	5,00	5,23
\bar{x}	6,14	6,13	5,67	5,98

De enkelte observasjonene i tabell VI 1 er altså representert ved

$$x_{ij} = \bar{x} + A_i + B_j + E_{ij}$$

$$\bar{x} = 5,98$$

A_i -verdiene er avviket fra hovedgjennomsnittet for de enkelte sorter

$$A_1 = 6,41 - 5,98 = 0,43$$

$$A_2 = 5,23 - 5,98 = -0,75$$

$$A_3 = 6,52 - 5,98 = 0,54$$

$$A_4 = 6,51 - 5,98 = 0,53$$

$$A_5 = 5,23 - 5,98 = -0,75$$

B_j -verdiene er avviket fra hovedgjennomsnittet for de enkelte gjentak.

$$B_I = 6,14 - 5,98 = 0,16$$

$$B_{II} = 6,13 - 5,98 = 0,15$$

$$B_{III} = 5,67 - 5,98 = 0,31$$

E_{ij} er endelig avviket fra $\bar{x} + A_i + B_j$ for hver forsøksrute.

$$E_{ij} = x_{ij} - \bar{x} - A_i - B_j$$

i	j = I	j = II	j = III
1	-0,17	-0,13	+0,30
2	+0,27	+0,02	-0,30
3	+0,16	-0,15	-0,02
4	0,00	+0,06	-0,05
5	-0,27	+0,20	+0,08

Kvadratsummen for E_{ij} er kvadratsummen for rest i variansanalysen

$$\sum E_{ij}^2 = SS_{rest} = (-0,17^2 + \dots + 0,08^2) = 0,4730$$

Kvadratsummene for A_j og B_j er mål for variasjonen for henholdsvis forsøksledd og gjentak, men siden A_i og B_j er regnet ut som gjennomsnitt og ikke som summer, må kvadratsummene multipliseres med antallet som ligger bak hvert gjennomsnitt for å få kvadratsummen for tall på enkelt rutenivå. Et lite eksempel vil vise at dette er riktig. Forsøksleddsummene er $r \cdot A_i$. Da

$A_i = 0$ blir også CT lik 0. Kvadratsummen for forsøksledd etter den vanlige formel er

$$SS_{Ledd} = \sum (rA_i)^2 / r = r \sum A_i^2$$

For eksemplet er derfor

$$SS_{sort} = 3 \sum A_i^2 \quad \text{og}$$
$$SS_{gjentak} = 5 \sum B_j^2$$

Forskjellen mellom de teoretiske verdiene α_i og β_j i den opprinnelige modellen og estimatene A_i og B_j er at estimatene har i seg en komponent av tilfeldige variasjon som de teoretiske verdiene ikke har. Dette kan vises ved å bruke en del av Bastian Larsens blindforsøk. Et forsøk med 3 forsøksledd, 1, 2 og 3, og 3 gjentak, I, II og III, legges ut i første del av blindforsøket. Data ordnes etter forsøksledd og gjentak på samme måte som i øvingsoppgave 4.

	1	2	3	Σ
I	36	36	40	112
II	50	49	58	157
III	61	59	57	117
Σ	147	144	155	446
A_i	- 0,56	-1,56	2,12	$\bar{x} = 49,56$

En variansanalyse av disse data gir

	DF	SS	MS
Total	8	806,22	
Gjentak	2	738,89	
Ledd	2	21,56	10,78
Rest	4	45,78	11,44

MS_{Ledd} og MS_{Rest} er her to uavhengige estimat av den samme tilfeldige variasjonen.

Vi kan nå innføre reelle forskjeller mellom forsøksledd slik at differansen mellom A_i økes, f.eks. ved å gi A_2 et tillegg på -5 og A_3 et tillegg på +5. Tabellen over forsøksresultatene ville da se slik ut:

	1	2	3	Σ
I	36	31	45	112
II	50	44	63	157
III	61	54	62	177
Σ	147	129	170	446
A_i	-0,56	-6,56	7,12	

Variansanalyse av disse data gir

	DF	SS	MS
Total	8	1066,22	
Gjentak	2	738,89	
Ledd	2	281,56	140,78
Rest	4	45,77	11,44

Variasjonen mellom gjentak og restvariansen er uendret, mens den totale variasjonen og variasjonen mellom ledd har økt. MS_{Ledd} har fortsatt den opprinnelige tilfeldige variasjon, men har i tillegg en variasjonskomponent som skyldes reelle forskjeller mellom forsøksleddene.

IV b Fixed og random effekter.

I markforsøk er forsøksleddene oftest fastsatt av den som planlegger forsøket. De ulike trinn av forsøksfaktorene er ikke trukket ut tilfeldig, men valgt for at forsøksresultatene skal vise effektene av nettopp disse forsøksleddene. Forsøksledd er derfor oftest en såkalt "fixed" variabel.

Gjentakene i et forsøk er tilfeldige prøver av det universet forsøksresultatene skal gjelde for. Effektene av de enkelte gjentak har ingen interesse. Det samme gjelder forsøksted, eller felt, i en forsøksserie. Gjentak og felt er tilfeldig, eller "random", variabler.

Om en variabel er fixed eller random kan konstanteres ved et lite tankeeksperiment. Hvis utskifting av et trinn av en variabel med et annet trinn endrer forsøksspørsmålet, er vedkommende variabel en fixed variabel. Forsøksspørsmålet endres ikke om et gjentak skiftes ut med et annet gjentak.

Hvis en er interessert i å bestemme gjennomsnittene for hvert trinn av en forsøksfaktor, er vedkommende forsøksfaktor en fixed variabel. Hvis en derimot er interessert i å bestemme variasjonen mellom trinnene er det en random variabel. Variansen for en random variabel er alltid et estimat av en varianskomponent, eller av en sum av flere varianskomponenter. Variansen for en fixed variabel estimerer en eller flere varianskomponenter pluss et uttrykk for variasjonen mellom trinnene i den bestemte variabel. Hvordan de teoretiske forventningene for variansen endrer seg avhengig av om variasjonsårsakene er fixed eller random variabler, vil til en viss grad bli diskutert under behandling av forsøksserier.

V FORSØKSPLANER

Målet for et forsøk er å få så nøyaktige svar som mulig på forsøksspørsmålene. Det er også et mål å kunne sette opp tallmessige grenser, konfidensintervall, for hvor nøyaktige svarene er. Dessuten er det viktig å kunne si under hvilke betingelser svarene er gyldige, hvilket univers vi kan uttale oss om.

V a Forsøksspørsmål

Det er praktisk å skille mellom to ulike typer av forsøksspørsmål, kvalitative og kvantitative. For de kvalitative spørsmålene er forsøksleddene sideordnet som f.eks. ulike sorter, metoder, redskaptyper, sprøytemidler osv. For de kvantitative spørsmålene kan forsøksleddene rekkeordnes som f.eks. gjødselmengder, sprøytemengder, behandling med ulike temperaturer osv.

I enkelte tilfeller kan det være tvil om en forsøksfaktor er kvalitativ eller kvantitativ. Hvis vi vil undersøke avling hos sorter med ulik veksttid kan sortene ordnes etter tidlighet, som er en kvantitativ faktor.

I de fleste tilfelle er det ganske greit å vite om vi har sideordnete eller rekkeordnete forsøksledd, og dette er viktig for å bestemme hvor mange forsøksledd vi skal starte med.

Med sideordnete forsøksledd er det best å starte med flest mulig alternativer og heller lage mindre nøyaktige forsøk. Det hjelper lite å få vite med stor sikkerhet hvilken av sortene A, B, C eller D som er best, hvis sort F er den aller beste. Først prøves alle aktuelle sorter i et relativt unøyaktig forsøk, og i neste omgang utføres nøyaktig undersøkelser med de beste sortene.

For rekkeordnete forsøksledd er vi interesserte i å bestemme et optimum på en kurve. Hvis vi på forhånd vet lite om hvor dette optimum er, gjør vi først et forsøk med få forsøksledd som ligger langt fra hverandre. Vi er da noenlunde sikre på å få med forsøksledd som ligger på begge sider av optimum. I neste omgang utfører vi et forsøk med relativt mange forsøksledd rundt optimum.

V b Forsøksenhet

Forsøksenheten kan være enkeltplanter, potter, kasser lagringsbinger, analyseprøver osv. Vi skal legge mest vekt på markforsøk, og da er det de enkelte rutene som er

forsøksenheten. I et senere kapittel skal vi diskutere størrelse og form på forsøksruter. En viktig forutsetning for at forsøksresultatene skal kunne brukes er at forsøksenheten er representativ for det universet vi vil uttale oss om.

V c Valg av forsøksplan

Med forsøksplan mener vi de prinsippene vi bruker for å fordele forsøksleddene på forsøksenhetene. Hvilken forsøksplan som velges er avhengig av forsøksspørsmålene og forsøksstypen, -markforsøk, veksthusforsøk, laboratorieforsøk-, av antall forsøksledd, og særlig av hvilken variasjon det er mellom forsøksenhetene. Hvis det er systematisk variasjon mellom forsøksenhetene prøver vi å eliminere denne ved hjelp av forsøksplanen.

V d Fullstendig tilfeldig fordeling

Den enkleste forsøksplan er tilfeldig fordeling av forsøksleddene på forsøksenhetene.

En forsøksplan med 4 sorter A, B, C og D på 16 ruter konstrueres slik at alle sortene har lik sjanse til å bli trukket ut på alle rutene. Trekningen kan foregå på forskjellige måter. En måte er at rutene nummereres fra 1 til 16 og at et bestemt antall ruter trekkes ut til sort A, et annet sett ruter til sort B osv.

Tabell V 1. Forsøksplan og avlingsdata for et forsøk med 4 sorter à 4 ruter, tilfeldig fordeling.

D 100	C 50	A 80	D 90
A 85	B 70	B 65	D 95
D 75	C 60	B 75	A 87
A 90	C 55	B 75	C 58

I tabell IV 1 er det gjengitt en forsøksplan for 4 sorter hver med 4 ruter i tilfeldig fordeling. Avlingsdata for hver rute er ført direkte på planen. For analysen må avlingstallene ordnes i en tabell etter sort som vist i tabell IV 2.

Tabell V 2. Avlingsdata ordnet etter sort.

	A	B	C	D	
	85	70	50	100	
	90	65	60	75	
	80	75	55	90	
	87	75	58	93	
Σ	342	285	223	360	= 1210
\bar{x}	85,5	71,25	55,75	90,-	

På grunnlag av tabell IV 2 kan det settes opp en enveis variansanalyse for forsøket.

$$CT = 91506$$

Variasjonsårsak	DF	SS	MS	F
Total	15	3402		
Sort (Ledd)	3	2873	958	21,8
Rest	12	529	44	

Fullstendig tilfeldig fordeling brukes når det er liten systematisk variasjon mellom forsøksenhetene. All variasjon innen forsøksledd blir betraktet som tilfeldig feil. En fordel ved planen er at det ikke er nødvendig å ha like mange gjentak av de ulike forsøksledd.

V e. Blokk forsøk

I stedet for å fordele sortene tilfeldig på alle 16 rutene, kunne de vært samlet i blokker som vist i tabell IV 3. Alle sortene finnes da i hver av blokkene, men de er tilfeldig fordelt innen blokker.

Tabell V 3. Forsøksplan og avlingsdata for et forsøk med 4 sorter i et blokk-forsøk med 4 blokker.

Blokk				
I	A 90	B 75	C 60	D 100
II	B 75	D 90	C 50	A 85
III	A 87	C 55	D 95	B 70
IV	C 58	D 75	A 80	B 65

Avlingstallene er de samme som i forrige eksempel. For beregning samles data fra et blokkforsøk i en toveistabell som vist i tabell IV 4.

Tabell V 4. Avlingsdata ordnet etter sort og blokk.

Blokk	Sort				Σ
	A	B	C	D	
I	90	75	60	100	325
II	85	75	50	90	300
III	87	70	55	95	307
IV	80	65	58	75	278
Σ	342	285	223	360	1210
\bar{x}	85,5	71,25	55,75	90,-	

Blokkforsøk analyseres ved hjelp av en toveis variansanalyse.

$$CT = 91506$$

Variasjonsårsak	DF	SS	F	
Total	15	3402		
Blokk (Gjentak)	3	283		
Sort	3	2873	958	35,5
Rest	9	246	27	

I markforsøk er det oftest noe systematisk variasjon mellom forsøksrutene. Ved å lage et blokkforsøk som er slik at denne systematiske variasjonen blir mellom blokker, kan den elimineres.

Det er bare variasjonen mellom forsøksenhetene innen blokk som går inn i forsøksfeilen.

Ved blokkforsøk kan vi altså eliminere noe systematisk variasjon, men forsøksfeilen (restvariansen) blir mindre nøyaktig bestemt enn ved tilfeldig fordeling. I vårt eksempel koster det 3 frihetsgrader å bestemme blokkeffektene, og disse frihetsgradene må vi ta fra frihetsgradene for rest.

V f. Latinsk kvadrat.

I et latinsk kvadrat blir forsøksleddene fordelt slik at de forekommer en gang i hver rekke og en gang i hver kolonne. Først konstruerer vi et system hvor denne forutsetningen er tilfredsstilt.

	4	1	3	2
4	A	B	C	D
2	D	A	B	C
1	C	D	A	B
3	B	C	D	A

Videre fordeler vi både rekker og kolonner tilfeldig ved at de gis et tilfeldig nr som vist på figuren. Forsøksplanen settes deretter opp som vist i tabell IV 5. Der er også avlingstallene ført inn.

Tabell V 5. Forsøksplan og avlingsdata for et forsøk med 4 sorter i et latinsk kvadrat.

Rekke	Kolonne				Σ
	1	2	3	4	
I	D 100	B 75	A 90	C 60	325
II	A 85	C 50	B 75	D 90	300
III	C 55	A 87	D 95	B 70	307
IV	B 65	D 75	C 58	A 80	278
Σ	305	287	310	300	1210

Data fra et latinsk kvadrat må summeres både for rekker, kolonner og forsøksledd. I tillegg til summeringen for rekker og kolonner i tabell IV 5 er derfor nødvendig med en tabell hvor data er ordnet etter sort. For det eksemplet som er vist her blir sortssummene de samme som i tabell IV 2 og tabell IV 4.

CT = 91506

Variasjonsårsak	DF		F
Total	15	3402	
Rekker	3	283	
Kolonner	3	123	
Sort	3	2873	958 : 45,6
Rest	6	123	21

Med latinsk kvadrat kan vi eliminere systematisk variasjon i to retninger. I markforsøk er det lett å tenke seg slik variasjon, men latinsk kvadrat kan også brukes i andre typer av undersøkelser. Hvis oppgaven var å undersøke effektiviteten av 4 arbeidsmetoder for tømmerhogst, og hver metode krevde en person i en dag, kunne vi engasjere 4 personer i fire dager og sette opp forsøksplanen slik at alle metoder ble prøvd hver dag og at alle personene arbeidet en dag med hver metode. Dette ville da være et latinsk kvadrat.

For latinsk kvadrat blir antall frihetsgrader ytterligere redusert i forhold til blokkforsøk. Antall forsøksledd og antall gjentak må være likt, og minimum er 3 forsøksledd. I markforsøk er det sjelden aktuelt å anlegge forsøk med mer enn 5-6 gjentak. Latinsk kvadrat er derfor begrenset til forsøk med fra 3 til 6 forsøksledd.

V g. Faktorielle forsøk

Fordelene ved faktorielle forsøk i forhold til enkeltfaktor forsøk, er at alle hovedeffektene blir sikrere bestemt ved samme forsøksinnsats og dessuten kan samspill mellom forsøksfaktorene bestemmes.

Forsøk med 3 sorter (A - C) og 3 gjentak

I	A	C	B
II	B	A	C
III	B	C	A

Forsøk med 3 gjødselmengder (1-3) og 3 gjentak.

I	1	3	2
II	1	2	3
III	3	2	1

2.

Forsøk med 3 sorter og 3 gjødselmengder og 2 gjentak.

A1	B2	C1	C2	A3	B1	A2	C3	B3
C3	A3	B1	B2	C2	B3	A1	C1	A2

3.

I forsøk 1 og 2 er det 9 ruter og i forsøk 3 er det 18 ruter. Arbeidsinnsatsen med å anlegge forsøket 1 og 2 er dermed tilnærmet lik arbeidsinnsatsen ved å anlegge forsøk 3. Forsøk 1 gir 3 observasjoner for hver av sortene og forsøk 2 gir tre observasjoner for hver av gjødselmengdene. Med enkeltfaktorforsøk er det ingen mulighet til å undersøke om det er samspill mellom de to faktorene. Forsøk 3 gir seks observasjoner av hver sort, og 6 observasjoner av hver gjødselmengde. Altså blir hovedfaktorene bestemt dobbelt så mange ganger i forsøk 3 som i forsøk 1 og 2. I tillegg er det i forsøk 3 mulig å bestemme samspillet mellom sort og gjødselmengder.

Ulempen med faktorielle forsøk er at gjentakene blir store og den tilfeldige variasjonen innen gjentak blir dermed også større enn for enkle forsøk.

En forsøksoppgave kan være å undersøke konkurranseevnen til en ny planteart i et jordbruksområde. En måte å angripe problemet på er å anlegge en serie enkeltforsøk for å få greie på såmengde, radavstand, sort gjødsling osv. For hvert av disse enkeltforsøkene måtte nivået av de andre faktorene bestemmes. Hvis det valgte nivå av gjødsel og radavstand viste seg å ligge langt fra de optimale ville heller ikke resultatene for såmengde eller sort ha almen gyldighet. Ved å anlegge et faktorielt forsøk med alle faktorene sammen ville hovedfaktorene bli langt sikrere bestemt ved samme innsats som i enkeltforsøkene, og dessuten ville det være mulig å bestemme de optimale kombinasjonene av de ulike faktorene.

Samspill mellom to faktorer betyr at effekten av ulike trinn av den ene faktoren er avhengig av nivået av den andre. I et gjødslingsforsøk er det kanskje ikke utslag for ulike mengder av nitrogen, hvis det ikke er nok fosfor tilstede, men

stort utslag for nitrogen, hvis det samtidig er gjødslet med fosfor. Samspill mellom flere enn to faktorer kan også forekomme, og 3-faktorsamspill kan i enkelte tilfelle ha mening. Samspill av høyere orden forekommer sjelden og om de forekommer er de vanligvis svært vanskelig å tolke, og enda vanskeligere å utnytte i praktisk veiledning.

1. Tilfeldig fordeling av alle kombinasjoner innen gjentak.

La oss se på planlegging og beregning av et forsøk med to faktorer med henholdsvis 3 og 4 trinn. Faktor A får verdiene A1, A2 og A3 og faktor B får verdiene B1, B2, B3 og B4. De to faktorene settes sammen slik at alle kombinasjonene blir representert.

A1 B1, A1 B2, A1 B3, A1 B4
A2 B1, A2 B2, A2 B3, A2 B4
A3 B1, A2 B2, A3 B3, A3 B4

Tilsammen blir dette 12 forsøksledd som fordeles tilfeldig innenfor hvert gjentak i en plan med f.eks. 3 gjentak.

For beregning av gjennomsnitt og for å finne de nødvendige summene for beregning av variansanalysen, må data ordnes i en 3-veis tabell. I tabell V 6 er det satt opp en slik tabell for vårt eksempel, men for en helt generell 3-veis tabell kan antallet av hver faktor erstattes med bokstavsymboler.

r = antall Rep. (gjentak) (3)
a = antall trinn av faktor A (3)
b = antall trinn av faktor B (4)
l = antall ledd (L) som er alle kombinasjonene av A og B (12)

Tabell V 6. Enkeltdata og summer i en 3-veis tabell

Rep.	B1	B2	B3	B4	\sum_b	
I	A1	x_{111}	x_{112}	x_{113}	x_{114}	$x_{11.}$
	A2	x_{121}	x_{122}	x_{123}	x_{124}	$x_{12.}$
	A3	x_{131}	x_{132}	x_{133}	x_{134}	$x_{13.}$
	\sum_a	$x_{1.1}$	$x_{1.2}$	$x_{1.3}$	$x_{1.4}$	$x_{1..}$
II	A1	x_{211}	x_{212}	x_{213}	x_{214}	$x_{21.}$
	A2	x_{221}	x_{222}	x_{223}	x_{224}	$x_{22.}$
	A3	x_{231}	x_{232}	x_{233}	x_{234}	$x_{23.}$
	\sum_a	$x_{2.1}$	$x_{2.2}$	$x_{2.3}$	$x_{2.4}$	$= x_{2..}$
III	A1	x_{311}	x_{312}	x_{313}	x_{314}	$x_{31.}$
	A2	x_{321}	x_{322}	x_{323}	x_{324}	$x_{32.}$
	A3	x_{331}	x_{332}	x_{333}	x_{334}	$x_{33.}$
	\sum_a	$x_{3.1}$	$x_{3.2}$	$x_{3.3}$	$x_{3.4}$	$x_{3..}$
\sum_r	A1	$x_{.11}$	$x_{.12}$	$x_{.13}$	$x_{.14}$	$x_{.1.}$
	A2	$x_{.21}$	$x_{.22}$	$x_{.23}$	$x_{.24}$	$x_{.2.}$
	A3	$x_{.31}$	$x_{.32}$	$x_{.33}$	$x_{.34}$	$x_{.3.}$
	$\sum_r \sum_a$	$x_{..1}$	$x_{..2}$	$x_{..3}$	$x_{..4}$	$x_{...}$

Fotskriftene på x-verdiene i tabell IV 6 står for henholdsvis Rep. faktor A, og faktor B. Når fotskriften er erstattet med punktum er verdien summert over vedkommende faktor. $x_{12.}$ = summen for rep I av trinn 2 av faktor A summert over B. For å sette opp en hvilken som helst x-verdi kan fotskriften erstattes med i, j og k, i går fra 1 til r, j fra 1 til a og k fra 1 til b.

For utregning av gjennomsnittene er det bare å ta hovedsummene for hvert trinn av faktorene og dividere på antall observasjoner som ligger bak hver sum.

Utregning av variansanalysen er i prinsippet det samme som før. Total kvadrat-

sum er kvadratene av alle enkeltobservasjonene summert minus CT. For utregning av kvadratsummen for gjentak brukes hovedsummen for hvert gjentak. Kvadratsummene for forsøksledd regnes ut på grunnlag av alle 12 kombinasjonene av A og B summert over Rep. Restkvadratsummen er

$$SS_{\text{Total}} - (SS_{\text{Rep.}} + SS_{\text{Ledd}}) \text{ som før.}$$

Det nye er at kvadratsummen for forsøksledd deles opp i hovedeffekter for faktorene A og B og i samspillet A x B. Før å regne ut de to hovedeffektene brukes totalsommene for henholdsvis A og B og samspillet regnet ut som en differens mellom SS_L og summen av SS_A og SS_B .

Variansanalyse for et 3x4x3 faktorielt forsøk.

Var. årsak	DF	SS
Total	35	$\sum x_{ijk}^2 - CT = SS_T$
Rep.	2	$\sum x_{i..}^2 / 12 - CT = SS_R$
Forsøksledd	11	$\sum x_{.jk}^2 / 3 - CT = SS_L$
A	2	$\sum x_{.j.}^2 / 12 - CT = SS_A$
B	3	$\sum x_{..k}^2 / 9 - CT = SS_B$
A x B	6	$SS_L - SS_A - SS_B$
Rest	22	$SS_T - SS_R - SS_L$

For dette tilfellet testes både hovedeffektene og samspillet mot den samme restvariansen.

2. Split-plot forsøk

Enkelte forsøksfaktorer krever store ruter eller store grensebelter mellom rutene mens andre faktorer kan bestemmes på små ruter uten grensebelter. I et vanlig faktorielt forsøk med tilfeldig fordeling innen gjentak må rutestørrelsen bestemmes etter den faktoren som tregner størst areal pr. rute.

I et split-plot forsøk legges en faktor ut på store ruter, som deles opp i mindre ruter for en annen forsøksfaktor. Faktoren på storruter må fordeles tilfeldig innen gjentak og faktoren på småruter må fordeles tilfeldig innen storruter.

Bestemmelsen av forsøksfaktoren på storruter blir minst presis, mens faktoren som er på småruter og samspillet mellom faktoren på storruter og faktoren på småruter blir nøyaktigere bestemt.

La oss sette opp en plan for 3 jordarbeidingsmetoder (A,B og C) og 3 fullgjødselmengder (1, 2 og 3) til korn. Det vil da være behov for store ruter for jordarbeiding, mens gjødselrutene kan være mindre. Forsøket planlegges med 3 gjentak.

I		II		III	
1		2		1	
3	A	1	B	2	B
2		3		3	
1		3		2	
3	C	2	A	1	C
2		1		3	
3		1		2	
1	B	2	C	1	A
2		3		3	

Det ville ikke være en korrekt split-plot plan hvis jordarbeidingen var utført gjennomgående over alle gjentak. Da ville forutsetningen om tilfeldig fordeling innen gjentak være brutt.

Variansanalysen for dette forsøket ville bli følgende:

	DF
Total	26
Rep.	2
Jorarbeiding	2
Feil A (Rep. x Jordarb.)	4

Gjødsling	2
Gjødsl. x Jordarb.	4
Feil B (Rep x Gjødsl x Rep x Gjødsl x Jordarb)	12

For å finne gjennomsnitt, samspillseffekter og kvadratsummer settes data opp i en treveistabell som vist på side 44. Feilen på storruter er forskjellig fra feilen på småruter. Effekten av jordarbeiding testes derfor mot feil A, og effektene av gjødsling og samspillet mellom jordarbeiding og gjødsling testes mot feil B.

3. Forsøksserier.

Veiledning om f.eks. sortsutvalg i et distrikt må bygge på forsøksmateriale fra et tilfeldig utvalg av steder i dette distriktet gjennom flere år.

Vi skal ta for oss analysen av kg korn pr. dekar fra en forsøksserie med 16 vårhvetesorter, som var prøvd i 17 felt hvert med 2 gjentak. Dette materialet stammer bare fra ett år. Analyse av materiale fra flere år blir diskutert senere.

For en total vurdering av sortenes dyrkingsverdi må mange karakterer undersøkes, men vurderingen av hvilken vekt en skal legge på de ulike karakterene er ikke et statistisk problem. Som en del av forsøksmetodikken er likevel denne vurderingen viktig. De ulike karakterene må i praksis gis vekt etter sin økonomiske verdi. For en del karakterer, som f.eks. tidlighet hos korn, overvintringsevne hos flerårige forvekster og karakterer som skallfarge hos matpotet og matkålrot må det settes absolutte grenser.

Hvis en karakter har svært liten variasjon, mister den sin verdi som utvalgs-kriterium sjøl om den har stor økonomisk betydning. Variansanalysen for en forsøksserie blir i prinsippet som for et faktorielt forsøk. Faktorene i eksemplet er: sort med 16 trinn felt med 17 trinn og gjentak med 2 trinn. Dette gir totalt 544 observasjoner og variansanalysen blir følgende:

Variasjonsårsak	DF	MS	estimat av MS
Total	543		
Felt	16		
Gjentak innen felt	17		
Sort	15	MS_S	$\sigma^2 + 2 \sigma_{SxF}^2 + 34 \sigma_{Sort}^2$
Sort x Felt	240	$MS_{S \times F}$	$\sigma^2 + 2 \sigma_{SxF}^2$
Rest (Sort x Gjentak innen felt)	255	MS_R	σ^2

Modellene for MS er satt opp under forutsetning av at Felt er en random variabel og Sort en fixed variabel. Gjentak er som alltid en random variabel. Den korrekte test for Sort er derfor

$$F = \frac{MS_{Sort}}{MS_{SxF}}$$

Samspillet Sort x felt testes mot restvariansen

$$F = \frac{MS_{SxF}}{MS_{Rest}}$$

Restvariansen representerer sort x gjentak innen felt.

Det er ingen sammenheng mellom gjentak nr. 1 på felt nr. 1 og gjentak på felt nr. 2. Det er derfor ikke mulig å ta ut en felles effekt av gjentak. Den gjennomsnittlige forskjellen mellom gjentak 1 og gjentak 2 skyldes bare tilfeldig variasjon. For hvert felt derimot er det reelle forskjeller mellom gjentak, og denne variasjonsårsaken kan elimineres. Hvert enkelt felt vil vi ha 1 frihetsgrad for gjentak. For 17 felt blir dette tilsammen 17 frihetsgrader og det er disse frihetsgradene som er samlet under variasjonsårsaken "Gjentak innen felt". Kvadratsummen for gjentak innen felt kan regnes ut som summen av kvadratsummene for gjentak for hvert felt. Fra en toveistabell over felt og gjentak, kan kvadratsummen for gjentak innen felt regnes ut som total-kvadratsummen minus kvadratsum for felt.

Felt	Gjentak		Σ
	I	II	
1	x	x	xx
2	x	x	xx
3	x	x	xx
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
17	x	x	xx

Variansanalyse

Variasjonsårsak	DF
Total	33
Felt	16
Gjentak innen felt	17

Hvis det hadde vært en sammenheng mellom gjentak 1 fra felt til felt og gjentak 2 fra felt til felt, ville gjentak innen felt vært sammensatt av gjentak med 1 frihetsgrad og gjentak x felt med 16 frihetsgrader. Sort x gjentak innen felt ville da vært sammensatt av sort x gjentak med 15 frihetsgrader og sort x gjentak x felt med 240 frihetsgrader.

Samspeillet sort x felt kan testes mot variansen for sort x gjentak innen felt. Hvis det kan påvises samspill mellom sorter og felt kan det være grunn til å dele opp materiale med tanke på å gi spesielle sortsanbefalinger for mindre distrikter. Hvis det på forhånd er klart at det av andre grunner må gis en felles sortsanbefaling for hele distriktet, har det liten hensikt å teste sort x felt samspeillet. En analyse av sortene kunne da like gjerne vært utført på gjennomsnittstall fra hvert felt. I eksemplet ville da variansene blitt halvparten så store fordi det er to gjentak pr. felt, men testen av sort mot sort x felt gir samme resultat.

Årsvariasjon

For å gi veiledning for "neste år" er det nødvendig å ha data for flere år. Hvis det er samspill mellom år og forsøksledd er den riktige feilen sort x år. Denne testen vil ha få frihetsgrader og det er nødvendig å ha data fra et stort

antall år for å kunne bestemme små forskjeller mellom forsøksledd. Samspillet mellom år og forsøksledd kan jo aldri utnyttes da vi ikke på forhånd kjenner til hvordan neste års vekstbetingelser blir.

Hvis samspillet mellom forsøksledd og år er utbetydelig, kan alle felt fra forskjellige år behandles under ett. I det tilfellet blir analysen nøyaktig som vist foran.

Antall felt.

I tillegg til å bruke variansanalysen for å finne ut om det er forskjeller mellom forsøksledd, kan estimatet av feilvariansen brukes til å regne ut hvor mange gjentak som må til for å bestemme differanser av en viss størrelse. Siden MS_{rest} bare er et estimat av den riktige feilvariansen, er også det antall gjentak som regnes ut et estimat av det "riktige" antall gjentak.

Middelfeilen på et gjennomsnitt, $m = s/\sqrt{r}$ eller $\sqrt{MS_{Rest}/r}$. Middelfeilen på en differanse er $\sqrt{2}m$. Et t-test for sammenligning av differansen mellom to gjennomsnitt blir derfor

$$t = (\bar{x}_1 - \bar{x}_2) / (\sqrt{2} \cdot s / \sqrt{r})$$

For at differansen skal være statistisk sikker på 5% nivået må

$$t \geq t_{5\%}$$

For å forenkle uttrykket settes $\bar{x}_1 - \bar{x}_2 = d$ og ulikheten blir

$$t_{5\%} \leq d \sqrt{r} / s \sqrt{2}$$

eller

$$\sqrt{r} \geq t_{5\%} \cdot s \cdot \sqrt{2} / d$$
$$r \geq t_{5\%}^2 \cdot s^2 \cdot 2 / d^2$$

$t_{5\%}$ er avhengig av r , men når antall gjentak er relativt stort er $t_{5\%}$ tilnærmet lik 2. For et grovt estimat av r kan derfor $t_{5\%}$ settes lik 2 og

$$r \geq 8 s^2 / d^2.$$

I forsøksserien med 16 råhvetesorter på 17 felt med 2 gjentak var $MS_{SxF} = 3140$. For at avlingsdifferanser på 10 kg skulle være signifikante på 5% nivået måtte

$$r \geq 8 \cdot 3140 / 100$$

$$r \geq 251$$

Siden r her er utregnet på grunnlag av forsøk med 2 gjentak måtte det anlegges 125 slike felt for å bestemme avlingsdifferanser på 10 kg pr. dekar.

Hvis det er tilstrekkelig å bestemme differansen på 20 kg, ville det klare seg med langt færre felt

$$r \geq 8 \cdot 3140 / 400$$

$$r \geq 63$$

eller 32 felt med 2 gjentak.

VI. ORTHOGONALE KONTRASTER

I et forsøk med flere forsøksledd kan det lages like mange uavhengige enkelt-sammenlikninger som det er frihetsgrader for forsøksledd. Slike uavhengige enkelt-sammenlikninger kalles orthogonale kontraster. I et forsøk med fire ledd er det tre frihetsgrader for ledd og det er tre mulige orthogonale kontraster. Hvis forsøksleddene er nummerert fra 1-4 er det mulig å sammenlikne f.eks. ledd 1 med 2 altså kontrasten 1-2, det er mulig å sammenlikne ledd 3 og ledd 4 altså kontrasten 3-4 og endelig er det mulig å sammenlikne 1+ 2 mot 3+4. Disse tre kontrastene er orthogonale. En forutsetning for å teste slike kontraster er at de er satt opp før resultatet av forsøket var kjent, altså at det er a priori sammenlikninger. Kvadratsummen for forsøksledd kan således deles opp med en kvadratsum for hver av de orthogonale kontrastene. Siden hver sammenlikning har en frihetsgrad blir kvadratsum og varians identiske. Disse variansene kan testes mot rest-variansen i variansanalysen på vanlig måte.

Det er ingen forutsetning at antall enkelt observasjoner som inngår på hver side i en kontrast må være like. I forsøket med tidligpotet i øvelsesoppgave nr 2 kunne f.eks. sort nr. 1, 3 og 4 være norske og sort 2 og 5 utenlandske sorter. Kontrasten norske mot utenlandske sorter kunne da testes. For å få en riktig sammenlikning måtte selvfølgelig gjennomsnittene av de to gruppene brukes slik at summen av de norske sortene ble dividert på 3 og summen av de utenlandske sortene ble dividert på 2. Innen de norske sortene er det to mulige orthogonale sammenlikninger. Disse to sammenlikningene kan godt slås sammen til en sammenlikning innen norske sorter. Innen utenlandske sorter er det bare en mulig sammenlikning.

Sum tørrstoffavling for tre gjentak for hver av de fem sortene er satt opp i tabell VI. 1. Variansanalysen for hele forsøket er gjengitt i tabell VI. 2.

Tabell VI.1. Tørrstoffavling av 5 potetsorter, sum av 3 enkeltruter.

Sort	Tørrstoffavling
1 Ostara	19.26
2 Alcmaria	15.69
3 Jonsok	19.57
4 Jaerla	19.57
5 Meteor	15.72

Tabell VI.2. Variansanalyse for tørrstoffavling hos 5 tidligpotetsorter.

Variasjonsårsak	DF	SS	MS	F
Total	14	6,6855		
Gjentak	2	0,7084		
Sort	4	5,6816	1,4204	23,9
Norske-Utenlandske	1	5,6600	5,6600	95,3
Innen Norske	2	0,0214	0,0107	
" Utenlandske	1	0,0002	0,0002	
Rest	8	0,4755	0,0594	

Summen av de norske sortene er 58,40 og summen av de utenlandske er 31,41. Totalsummen for forsøket er 89,81. I variansanalysen kan de fire frihetsgradene for sort, med tilhørende kvadratsum deles opp i sammenlikningen mellom norske og utenlandske sorter, og sammenlikning innen de to gruppene som vist i tabell VI.2. Kvadratsummen for sammenlikningen norske - utenlandske er

$$58,4^2/9 + 31,41^2/6 - 89,81^2/15 = 5,6600$$

Divisor i de tre brøkene er det antall enkeltobservasjoner som ligger bak hver av summene i telleren. Kvadratsummen for sammenlikningen innen norske sorter er

$$(19,26^2 + 19,57^2 + 19,57^2)/3 - 58,4^2/9 = 0,0214$$

Kvadratsummen for sammenlikning innen utenlandske sorter er

$$(15,69^2 + 15,72^2)/3 - 31,41^2/6 = 0,0002$$

Summen av disse tre kvadratsummene = 5,6816 som er det samme som kvadratsummen for sort. En av de tre kvadratsummene kunne derfor vært regnet ut som en differanse.

VI.a Orthogonale koeffisienter

I et enkelt tilfelle hvor a, b og c er f.eks sum avling av r gjentak for 3 forsøksledd, og hvor det på forhånd er bestemt at kontrasten a mot b skal undersøkes, er det relativt lett å regne ut hvilken annen kontrast som er orthogonal med sammenlikningen a mot b. Dette kan gjøres ved å ta utgangspunkt i den totale kvadratsummen for forsøksledd som er

$$(a^2 + b^2 + c^2)/r - (a + b + c)^2/3r \quad (1)$$

og trekke fra kvadratsummen for kontrasten ab som er

$$(a^2 + b^2)/r - (a + b)^2/2r = (a-b)^2/2r \quad (2)$$

Differansen mellom disse to uttrykkene er

$$(a + b - 2c)^2/6r \quad (3)$$

Uttrykket inne i parenteser, kontrasten a + b - 2c, er den kontrasten som er orthogonal til sammenlikningen a-b. Ved en enkel omskriving er den andre kontrasten gjennomsnittet av a og b mot c,

$$(a + b)/2 - c$$

For beregning av kvadratsummer og effekter, og som et hjelpemiddel for å undersøke om kontrastene virkelig er orthogonale, er det nyttig å sette opp koeffisienter for de ulike kontrastene. Koeffisientene for kontrasten a-b er selvsagt, men de går også fram av koeffisientene for a og b i siste del av likning (2). Koeffisientene for den andre kontrasten går fram av likning (3).

Definisjonen på en kontrast er en sammenlikning hvor summen av koeffisienten er lik 0.

Tabell VI. 3. Orthogonale koeffisienter for to kontraster fra tre forsøksledd.

Forsøks- ledd	\sum_x^r	(a - b)	(a+b - 2c)	
		k_1	k_2	$k_1 k_2$
A	a	+1	+1	+1
B	b	-1	+1	-1
C	c	0	-2	0
$\sum k$		0	0	0
$\sum k^2$		2	6	

For at to kontraster skal være orthogonale må summen av produktene av to og to sett orthogonale koeffisienter være 0.

I tabell VI.3. er koeffisientene for de to kontrastene i eksemplet foran satt opp. I tillegg er $\sum k_1^2$, og $\sum k_2^2$ reknet ut. Kvadratsummen, SS, for en orthogonal kontrast er

$$SS = \frac{[\sum k(\epsilon_x)]^2}{\sum k^2 r}$$

For eksemplet ovenfor er effektene av de to kontrastene

$$(a - b)/r \text{ og}$$

$$(a + b - 2c)/2r$$

og kvadratsummene er

$$(a - b)^2/2r \quad \text{og}$$

$$(a + b - 2c)^2/6r$$

som stemmer med likning (2) og (3).

VI.6.Orthogonale polynomer

For forsøk med flere forsøksledd er det mulig å sette opp mange forskjellige orthogonale sammenlikninger. Problemet er ofte at det ikke på forhånd er gitt hvilke kontraster som er mest interessante å undersøke. Et eksempel kan være forsøk med fire gjødselmengder. Det er ikke på forhånd mulig å si at sammenlikning mellom de to laveste mengdene og de to høyeste mengdene er mer interessant enn sammenlikning mellom de tre laveste og den høyeste. For forsøk med rekkeordnede forsøksledd er det mulig å lage generelle orthogonale sammenlikninger for å finne ut om det er lineær eller kvadratisk sammenheng mellom forsøksbehandlingene og forsøksresultatene. Helt generelt kan det settes opp et polynom

$$x = a + bt + ct^2 + dt^3 \dots \dots \dots qt^n$$

I denne likningen står t for forsøksbehandlingene og x for forsøksresultatet. a, b, c, d, osv. er koeffisienter for de enkelte ledd. Likningen uttrykker altså forsøksresultatet som en funksjon av forsøksbehandlingene. Det er mulig å sette opp ett sett orthogonale koeffisienter for hvert enkelt ledd i polynomet. I likhet med de andre orthogonale sammenlikningene som er behandlet foran, settes også disse koeffisientene opp slik at sammenlikningene er uavhengig av hverandre. Kvadratsummene for de ulike sammenlikningene adderer opp til den totale kvadratsummen for forsøksledd. Som før har hver orthogonal sammenlikning en frihetsgrad.

Hvert ledd i polynomet kan også knyttes sammen med en null-hypotese. Den generelle null-hypotesen går ut på at det ikke er noen forskjell mellom forsøksledd, dvs. at $x = a$. Hvis den null-hypotesen må forkastes kan det settes opp nye nullhypoteser i det et nytt ledd tas med for hver gang. I uttrykket

$$x = a + bt$$

er x en lineær funksjon av t og null-hypotesen går ut på at $b = 0$. Neste funksjon er

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

Dette er en annengradsfunksjon og null-hypotesen går nå ut på at det siste leddet er 0. En annen måte å uttrykke nullhypotesen på er at det ikke er noen kvadratisk effekt av t på x . Antall ledd i polynomet som er mulig å teste er avhengig av hvor mange trinn vi har for forsøksfaktoren. Med tre trinn, dvs. to frihetsgrader, kan vi teste for lineær og kvadratisk effekt. Med fire trinn kan vi teste for lineær kvadratisk og kubisk effekt. I spesielle tilfeller kan en tredjegrads funksjon ha en fornuftig tolkning, men funksjoner av høyere grad er vanligvis meget vanskelige å tolke i biologisk sammenheng.

1. Utregning av orthogonale koeffisienter

Det er mulig å sette opp orthogonale koeffisienter for lineær, kvadratisk, kubisk effekt, osv., uansett hvordan nivåene i forsøksfaktoren varierer innbyrdes. Koeffisientene blir imidlertid mye enklere hvis det forutsettes lik avstand mellom trinnene i forsøksfaktoren.

Den enkleste mulige skala for en forsøksfaktor som varierer i 4 like trinn er en skala som går fra 0 til 3. t har altså verdiene 0, 1, 2 og 3. En lineær effekt av t må ha formen $k = a + bt$. Den lineære sammenhengen endres ikke om vi setter $b = 1$ og $k_1 = a + t$. På samme måte kan vi sette opp en kvadratisk effekt som er $k_2 = a + bt + t^2$. Utregningen av koeffisientene er vist i tabelle VI a og VI 4b. Etter definisjoner på en kontrast må summen av koeffisientene være 0.

$$4a + 6 = 0$$

$$a = 3/2$$

Ved å sette inn verdien for a i likning I for hvert forsøksledd i tabell VI 4a får vi en rekke koeffisienter som er merket k'_1 . Dette er et sett orthogonale koeffisienter, men da det er enklere å operere med hele tall multipliseres disse koeffisientene med 2 og vi oppnår rekken av koeffisienter, k_1 , som er satt opp i tabellen. Forutsetningen for at $\sum k=0$ holder også for disse koeffisientene, sjøl om de er dobbelt så store som de koeffisientene vi først kom fram til.

For å utlede koeffisientene for kvadratisk effekt er det nødvendig i tillegg til forutsetningen om at summen av koeffisientene skal være lik 0, også å ta inn forutsetningen om at summen av produktet av k_1 og k_2 være lik 0. Dette er vist i tabell VI.4.b.

Tabell VI.4a. Utledning av orthogonale koeffisienter for lineær effekt av en forsøksfaktor i 4 trinn

Forsøksledd	I	k_1	k_1
t	a + t		
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

Tabell VI.4b Utledning av orthogonale koeffisienter for kvadratisk effekt av en forsøksfaktor med 4 trinn.

Forsøksledd	II	$k_1 \cdot II$	k_2
t	a + bt + t ²		
0	a	- 3a	1
1	a + b + 1	-a - b - 1	-1
2	a + 2b + 4	a + 2b + 4	-1
3	a + 3b + 9	3a + 9b + 27	+1
Sum	4a + 6b + 14	10b + 30	0

På samme måte som for den lineære effekten settes de kvadratiske effektene opp for hvert forsøksledd og ved summering finner vi at summen av effektene blir $4a + 6b + 14$. Summen av produktet $k_j \cdot II$ blir $10b + 30$. Ved å sette disse to summene lik 0, løser vi ligningene både med hensyn på b og a. summene lik 0, løser vi ligningene både med hensyn på b og a.

$$10b + 30 = 0$$

$$b = - 3$$

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

$$a = 1$$

Endelig ved å sette disse verdiene inn i likning II finnes koeffisientene k_2 .

2. Beregning av lineære og kvadratiske effekter.

Det er vist foran hvordan orthogonale koeffisienter kan brukes til å lette regnearbeidet med å finne kvadratsummer for orthogonale kontraster. I tillegg kan de orthogonale koeffisientene brukes til å regne ut en lineær og en kvadratisk regresjonsfunksjon av t på x . x er fortsatt et mål for forsøksresultatet og t et mål for forsøksfaktoren. Utgangspunktet for beregningen av x -verdiene er gjennomsnittet av x .

$$\bar{x} = \sum x/n$$

Den gjennomsnittlige lineære effekten er

$$\bar{L} = \sum (k_1 \sum x) / r \sum k_1^2$$

og den gjennomsnittlige kvadratiske effekten er

$$\bar{K} = \sum (k_2 \sum x) / r \sum k_2^2$$

For hver verdi av t regner vi så ut den lineære effekten

$$x_L = \bar{x} + k_1 \bar{L}$$

og den kvadratiske effekten

$$x_k = \bar{x} + k_1 \bar{L} + k_2 \bar{K}$$

For grafisk framstilling plottes så disse x -verdiene mot t i et to-veis diagram.

VII FORUTSETNINGER FOR VARIANSANALYSEN

Hittil er det tatt for gitt at hvis resultatet av variansanalysen er en signifikant F , forkastes en nullhypotese som går ut på at gjennomsnittene for forsøksleddene er like. En signifikant F -verdi kan også bety at endel av forutsetningene for variansanalysen ikke er oppfylt, men oftest er det feilen som blir overestimert når forutsetningene ikke er oppfylt. Dette fører til at reelle forskjeller mellom forsøksleddene ikke kan påvises. De viktigste forutsetningene er:

Tilfeldig fordeling av forsøksfeilen

Homogen varians

Normal fordeling

Additivitet

VII a Tilfeldig fordeling av forsøksfeilen

Det er innlysende at hvis noen forsøksledd bare har positive avvik og andre forsøksledd negative avvik vil dette oppfattes som forskjeller mellom forsøksledd. Det er derfor viktig at forsøksleddene blir tilfeldig fordelt på forsøksrutene. I markforsøk er det ofte systematiske endringer i vekstvilkårene fra en ende av forsøket til den andre. Hvis et forsøk med stigende gjødselmengder anlegges uten tilfeldig fordeling av forsøksleddene innbyrdes vil derfor forsøksresultatene bli påvirket i en eller annen retning av jordvariasjonen.

I tillegg til at forskjellen mellom forsøksleddene ikke blir riktig, blir variasjonen innen forsøksledd (feilen) underestimert.

VII b. Homogen varians

En forutsetning for variansanalysen er at variansen er lik for alle forsøksledd. Hvis variansen er avhengig av forsøksledd d.v.s. at noen forsøksledd har større innbyrdes variasjon enn andre, vil variansanalysen normalt ikke gi en signifikant F -verdi sjøl om det er forskjeller mellom gjennomsnittene for forsøksledd.

Det er utarbeidet en egen test, Bartlett's test, for å undersøke om variansene er homogene. En ulempe ved denne testen er at den er svært avhengig av at forsøksfeilen er normalt fordelt.

Forutsetningen om homogen varians har størst betydning hvis antal frihetsgrader er lite.

VII c. Normal fordeling

Strengt tatt forutsettes også at forsøksfeilen er normalt fordelt. Denne forutsetningen er ikke særlig viktig. Store avvik fra normalitet fører bare til små endringer i signifikansnivå.

VII d. Additivitet

En viktig forutsetning i variansanalysen er at effektene av forsøksfaktorer og gjentak er additive. Den enkelte observasjon i et blokkforsøk kan da settes opp som

$$x_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

Et talleksempel vil illustrere nærmere hva additivitet betyr.

Tabell VII 1. Additiv effekt av α_i og β_j

	$\alpha_1 = 1$	$\alpha_2 = 2$	$\alpha_3 = 3$	Σ
$\beta_1 = 1$	2	3	4	9
$\beta_2 = 5$	6	7	8	21
Σ	8	10	12	30

Tabell VII 2. Multiplikativ effekt av α_i og β_j

	$\alpha_1 = 1$	$\alpha_2 = 2$	$\alpha_3 = 3$	Σ
$\beta_1 = 1$	1	2	3	6
$\beta_2 = 5$	5	10	15	30
Σ	6	12	18	36

Variansanalyse

Variasjonsårsak	DF	SS ADDITIV	SS MULT.
Total	5	28	148
α	2	4	36
β	1	24	96
$\alpha \cdot \beta$	2	0	16

I tabell VII 1 og 2 er det satt opp henholdsvis additive og multiplikative effekter av α og β . Det er ikke tatt med ε i denne tabellen. Vi forutsetter altså feilfrie data. I variansanalysen for additive data blir da kvadratsummen for samspillet lik 0, mens de multiplikative data gir en kvadratsum for samspillet på 16.

I et reelt tilfelle, hvor restvariansen brukes som et estimat av forsøksfeilen, vil multiplikative data føre til overestimering av feilen.

VII e Transformasjoner

Hvis forutsetningene for variansanalysen ikke er oppfylt kan dette ofte rettes på ved å transformere data til en annen skala. For eksempel i tabell VII 2 kan effektene gjøres additive ved en logaritmetransformasjon. Dette er vist i tabell VII 3.

Tabell VII 3. Logaritmetransformasjon av data i tabell VII 2.

	$\alpha_1 = 0$	$\alpha_2 = 0,30$	$\alpha_3 = 0,48$	Σ
$\beta_1 = 0$	0	0,30	0,48	0,78
$\beta_2 = 0,7$	0,70	1,00	1,18	2,88
Σ	0,70	1,30	1,66	3,66

Variansanalyse

Variasjonsårsak	DF	SS _{LOG}
Total	5	0,9702
α	2	0,2352
β	1	0,7350
$\alpha \cdot \beta$	2	0,0000

Ved å transformere til logaritmer blir effektene igjen additive og samspillet i variansanalysen blir igjen 0.

Transformasjoner kan også eliminere avhengighet mellom varians og gjennomsnitt og gjøre data mer tilnærmet normalt fordelt.

VIII CONFOUNDING, SAMMENBLANDING AV EFFEKTER

Variasjonen mellom to forsøksruter blir større jo lengre rutene ligger fra hverandre. Jo flere forsøksledd, jo større blir avstanden mellom rutene innen samme gjentak. Hvis variasjonen innen gjentak blir stor, går det an å dele opp gjentakene i mindre blokker. Variasjonen mellom blokker kan da elimineres. For å få til dette må forsøksplanen lages slik at forskjellen mellom blokker er "blandet sammen" med et samspill som har liten interesse.

Effekten av forsøksledd kan deles opp i så mange orthogonale kontraster som det er antall frihetsgrader. I et faktorielt forsøk med tre faktorer, hver med to trinn, kan disse kontrastene beskrives med orthogonale koefisienter som alle er +1 eller -1.

I et forsøk med N, P og K hvor hver faktor har to trinn, betrakter vi det forsøksleddet som har minste mengde av alle gjødselslagene med (1), og de øvrige forsøksleddene med n, p, k osv. som vist i tabell VIII 1. Forsøksleddet n har da minste mengde av P og K og største mengde av N. Leddet np har minste mengde av K og største mengde av N og P. De orthogonale koefisientene for hovedeffekten og samspillseffekten er også vist i tabell VIII 1.

Tabell VIII 1. Orthogonale koefisienter for hovedeffekten og samspills-effekter for et forsøk med tre faktorer i to trinn.

Forsøkseffekter							
Forsøksledd	N	P	K	NP	NK	PK	NPK
(1)	-1	-1	-1	+1	+1	+1	-1
n	+1	-1	-1	-1	-1	+1	+1
p	-1	+1	-1	-1	+1	-1	+1
np	+1	+1	-1	+1	-1	-1	-1
k	-1	-1	+1	+1	-1	-1	+1
nk	+1	-1	+1	-1	+1	-1	-1
pk	-1	+1	+1	-1	-1	+1	-1
npk	+1	+1	+1	+1	+1	+1	+1

Hvert gjentak kan deles i 2 blokker à 4 ruter slik at effekten av trefaktor-samspillet og blokker blandes sammen. Forsøksledd med +1 for trefaktorsamspillet plasseres i en blokk og forsøksledd med -1 for trefaktorsamspillet i den andre.

De andre effektene har to ledd med + og to ledd med - innen hver blokk. Derfor vil ikke forskjellen mellom blokker virke på disse.

Forsøksplanen vil for hvert gjentak vil da se slik ut:

O	N
NP	P
NK	K
PK	NPK

Innen disse blokkene kan så forsøksleddene fordeles tilfeldig. Den tilfeldige feilen i forsøket reduseres ved at variasjonen mellom blokker blir eliminert, og dette oppnås på bekostning av at trefaktorsamspillet ikke blir bestemt innen gjentak.

IX VARIASJONSÅRSAKER I MARKFORSØK

Forskjellene mellom de enkelte observasjonene fra et markforsøk skyldes tre prinsipielt ulike variasjonsårsaker. For det første er det forskjeller som skyldes forsøksbehandlingene. Formålet med forsøk er å bestemme disse forskjellene så sikkert som mulig.

Den andre typen av variasjonsårsaker er de tilfeldige feil i forsøket. Dette kan være jordvariasjon, variasjon i plantematerialet, arbeidsfeil under anlegg, stell, høsting og oppveing av forsøket, lese og skrivefeil under behandlingen av forsøksresultatene m.m. Jo mindre summen av de tilfeldige feilene er, jo sikrere blir eventuelle forskjeller mellom forsøksleddene bestemt.

Den tredje typen av variasjonsårsaker er de systematiske. Jordvariasjon kan f.eks. være systematisk ved at næringstilstanden endrer seg gradvis etter en bestemt retning på et jordstykke. Systematisk variasjon søkes eliminert ved hjelp av forsøksplanen. Gjentakene og forsøksrutene innen gjentak legges slik at det blir minst mulig variasjon mellom forsøksledd innen gjentak. Forskjeller mellom gjentak har ingen betydning for hovedeffekten av forsøksledd. Vi har også diskutert forsøksplaner som eliminerer systematisk jordvariasjon i to retninger.

Eventuell variasjon mellom forsøksruter innen gjentak kan gjøres tilfeldig ved at forsøksleddene fordeles tilfeldig på de ulike forsøksrutene.

IX a Rutestørrelse

Resultater av forsøk med enkeltplanter kan bare i liten grad gi opplysninger om hvordan disse plantene reagerer i praktisk dyrking. Det første kravet til rute-

størrelse er derfor at rutene må være store nok til å representere et bestand. Store forsøksruter er også en måte å redusere mange tilfeldige feil på. Små veiefeil og avrundingsfeil får relativt sett mindre betydning jo større mengder som behandles.

På den andre sida vil store forsøksruter føre til at de ulike forsøksleddene blir liggende lenger fra hverandre, og jo større jordstykker vi bruker til ett forsøk, jo større blir den totale variasjonen innen dette jordstykket. Dessuten vil hver forsøksrute koste mer jo større den er. En må derfor velge rutestørrelse slik at en får mest mulig nøyaktig data ut av et bestemt antall kroner.

Rutestørrelsen er også avhengig av hva slags forsøk det er tale om. Gjødslingsforsøk krever større ruter enn sortsforsøk på grunn av større kanteffekter. Grøftforsøk og vanningsforsøk må av praktiske grunner ha meget store ruter, mens utprøving av nytt sortsmateriale ofte foregår på små ruter på grunn av lite såfrø eller plantemateriale av hver sort.

I sorts- og gjødslingsforsøk her i landet brukes det vanligvis anleggsruter på 10-20 m². Høsterutene blir noe mindre på grunn av grensebelter som fjernes før høsting.

IX b Rutenes form

I gjødslingsforsøk vil plantene på de ugjødslete eller svakt gjødslete rutene ha nytte av gjødslingen på naborutene. Også i sortsforsøk har vi nabovirkning, særlig fordi sortene med raskere eller kraftigere vegetativ vekst tar vann og næring fra naboruter.

For samme rutestørrelse vil nabovirkningen bli større jo smalere rutene er. Den blir minst ved bruk av kvadratiske ruter. Den eneste sikre måten å unngå nabovirkning på er å ha tilstrekkelig brede grensebelter mellom rutene.

På grunn av tilpassning til teknisk utstyr og også fordi vi er mere interesserte i et kvadratisk gjentak enn i en kvadratisk rute blir det brukt rektangulære ruter i vanlige blokkforsøk.

IX c Antall gjentak

Variansen for et gjennomsnitt er lik variansen for enkeltobservasjonene dividert med antallet som ligger bak hvert gjennomsnitt.

$$MS_x = MS_x/n$$

Vi kan derfor redusere den tilfeldige variasjonen mellom gjennomsnittet for ulike forsøksledd ved å anlegge flere ruter av samme forsøksledd.

Nedgangen i middelavviket er omvendt proporsjonal med kvadratroten av antallet. Økningen fra 2 til 3, 4 og 5 har derfor størst effekt, og en videre økning av antall gjentak får mindre betydning. Ved å øke antall gjentak får vi 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. Det har derfor størst betydning å ha mange gjentak for forsøk med få forsøksledd.

I en forsøksserie som er anlagt for å gi veiledning for et distrikt er det viktigere å ha mange felt med få gjentak enn å ha mange gjentak på få felt.