

# HANDBOK I STATISTIK FÖR FÄLTFÖRSÖK

Johannes Forkman

Fältforsk  
Sveriges lantbruksuniversitet

2012-03-13

Fältforsk  
Sveriges lantbruksuniversitet, SLU, Uppsala

Handbok i statistik för fältförsök  
Johannes Forkman

# Innehåll

<b>1</b>	<b>Grundläggande statistiska begrepp</b>	<b>7</b>
1.1	Experimentenheter och replikat . . . . .	7
1.2	Randomisering . . . . .	11
1.3	Spridningsmått . . . . .	11
1.4	Normalfördelning . . . . .	14
1.5	Hypotesprövning . . . . .	14
<b>2</b>	<b>Randomiserade fullständiga blockförsök</b>	<b>19</b>
2.1	Utläggning i fält . . . . .	19
2.2	Statistisk analys och tolkning . . . . .	21
<b>3</b>	<b>Dimensionering</b>	<b>27</b>
<b>4</b>	<b>Blandade modeller</b>	<b>35</b>
4.1	Fältförsök har flera varianskomponenter . . . . .	35
4.2	Ett enkelt sätt att specificera modeller . . . . .	37
4.3	Slumpmässiga blockeffekter . . . . .	38
4.4	Romerska kvadrater . . . . .	39
<b>5</b>	<b>Fältförsök med ofullständiga block</b>	<b>43</b>
5.1	Gitterdesign . . . . .	43
5.2	Rektangulär gitterdesign . . . . .	46
5.3	Alfadesign . . . . .	48
5.4	Val av blockdesign . . . . .	51
<b>6</b>	<b>Fältförsök med flera faktorer</b>	<b>53</b>
6.1	Fördelar med flerfaktoriella försök . . . . .	53
6.2	Split-plot-försök . . . . .	54
6.3	Split-block-försök . . . . .	56
<b>7</b>	<b>Försöksserier</b>	<b>59</b>
7.1	Statistiska modeller för försöksserier . . . . .	60
7.2	Dimensionering av enkla försöksserier . . . . .	62



# Förord

Med Handbok i statistik för fältförsök hoppas jag kunna bidra till att reda ut en del grundläggande begrepp i statistikens förbryllande värld, och särskilt då de begrepp som man mest sannolikt träffar på i samband med experiment som utförs på jordbruksmark. Det handlar till exempel om medelfel, variationskoefficienter, sannolikhetsvärde och minsta signifikanta skillnaden. Jag vill också redogöra för och diskutera de vanligaste typerna av fältförsök: fullständiga randomiserade blockförsök, ofullständiga blockförsök, split-plot-försök och split-block-försök.

Jag tar upp traditionell variansanalys och modeller med flera varianskomponenter eftersom sådana är mycket användbara i fältforskning. Det nödvändiga antalet upprepningar och försök, samt modeller för statistisk analys av försöks-serier, diskuteras särskilt.

Handbok i statistik för fältförsök är tänkt att komplettera Försökshandboken som publiceras på Fältforsks hemsida<sup>1</sup>. På kursen Försöksstatistik fungerar Handbok i statistik för fältförsök som kursdokumentation.

Som förebild har jag använt Handbok för jordbruksförsök: statistisk planering och analys (Seeger, 1984). För den som önskar lite mer kött på benen kan jag även rekommendera Variansanalys och försöksplanering (Engstrand och Olsson, 2003).

Jag tar tacksamt emot förslag på förbättringar och upplysningar om felaktigheter och otydligheter, stort som smått.

Uppsala den 22 oktober 2010

Johannes Forkman

---

<sup>1</sup>[www.slu.se/sv/fakulteter/nl/om-fakulteten/ovriga-enheter/faltforsk/utbildning/forsoks-handbok/](http://www.slu.se/sv/fakulteter/nl/om-fakulteten/ovriga-enheter/faltforsk/utbildning/forsoks-handbok/)



# Kapitel 1

## Grundläggande statistiska begrepp

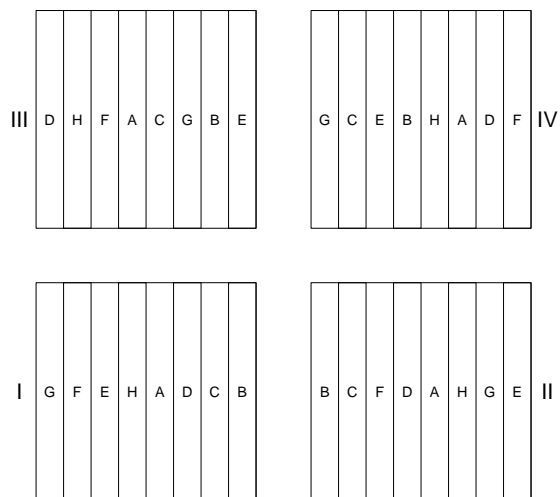
### 1.1 Experimentenheter och replikat

Med fältförsök avser vi experiment som utförs på jordbruksmark för att studera effekten av olika behandlingar. Behandlingarna kan till exempel vara olika metoder för jordbearbetning, olika växtskyddsbehandlingar eller olika sorter av någon gröda. Behandlingar kallas även led, och vi kommer att använda de båda orden omväxlande och synonymt. Många försök innehåller ett controlled, vilket är ett led utan aktiv behandling. Även controlledet är en behandling i statistisk mening, så om försöket omfattar åtta aktiva behandlingar och ett controlled ska vi säga att försöket omfattar nio behandlingar eller led.

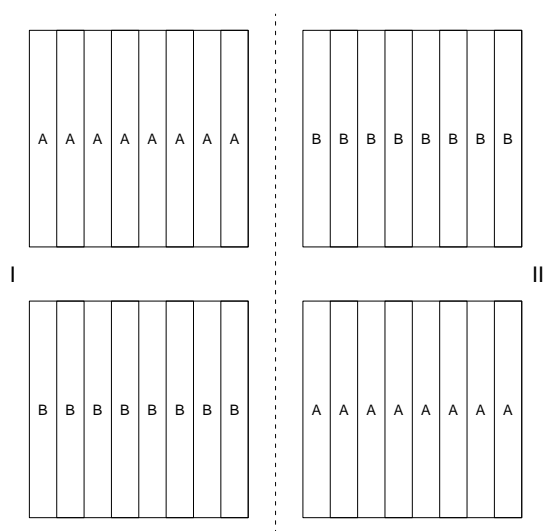
Man skiljer mellan jämförande och bestämmande försök. Fältförsök är vanligen jämförande: de syftar till att jämföra två eller flera behandlingar med varandra. Ibland görs dock bestämmande försök, i vilka syftet är att skatta någon storhet, till exempel skördens torrsvikt eller försöksplatsens variation, under en specifik behandling.

Ofta delar man in försöksfältet i delar, till exempel så som i figur 1.1. I exemplet är fältet indelat i fyra block (I–IV), och varje block innehåller åtta avlånga rutor. Om man har åtta behandlingar som ska jämföras med varandra, kan man slumpa ut de åtta behandlingarna på de åtta rutorna inom varje block, så att varje block kommer att innehålla samtliga åtta behandlingar en gång. När behandlingarna randomiseras till rutorna inom blocken är rutorna experimentenheter, och försöket är ett så kallat randomiserat blockförsök. Eftersom försöket omfattar fyra observationer av varje led säger man att försöket har fyra replikat eller upprepningar. I ett fullständigt blockförsök, som detta, förekommer varje behandling en gång i varje block. Då finns det lika många replikat som block, och orden replikat och block används synonymt.

På samma fält skulle man kunna jämföra två behandlingar enligt figur 1.2. Med den här planen omfattar försöket bara två block eller replikat. Blocken



Figur 1.1: Ett försök med åtta behandlingar (A–H), fyra block (I–IV) och åtta rutor per block. Det finns åtta experimentenheter per block.

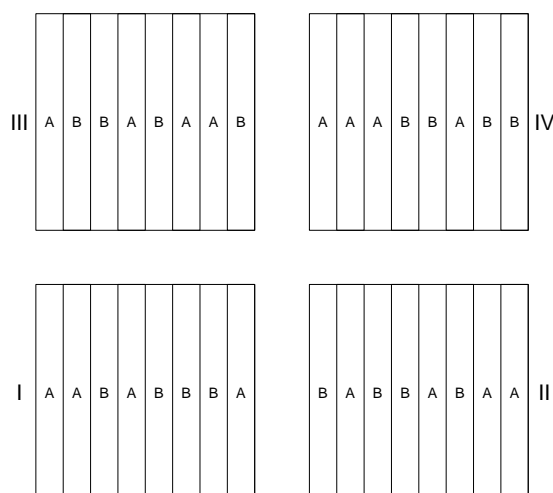


Figur 1.2: Ett försök med två behandlingar (A, B), två block (I och II), två stor-rutor per block och åtta smårutor per storruta. Det finns två experimentenheter per block.



består vardera av två storrutor, och varje storruta består av åtta smårutor. Om de båda behandlingarna läggs på storrutorna, så att samtliga smårutor i samma storruta får samma behandling är det storrutorna som är experimentenheterna. Då finns det totalt två replikat. Om vi gör mätningar i varje småruta finns det åtta observationsenheter i varje storruta. I så fall är observationsenheterna mindre än experimentenheterna.

Om däremot de båda behandlingarna slumpas ut på de små rutorna, så att det finns fyra replikat av varje behandling i varje storruta är smårutorna experimentenheter och storrutorna block (figur 1.3). I så fall omfattar försöket totalt 16 replikat. Skulle man skörda varje småruta för sig, och sedan före mätning blanda skördarna från samma storruta och behandling, skulle man ha två observationsenheter per storruta. I så fall skulle det per behandling finnas 16 experimentenheter men bara fyra observationsenheter. Experimentenheten kan alltså vara mindre än observationsenheten. I de flesta försök är dock observationsenheten identisk med experimentenheten.



Figur 1.3: Ett försök med två behandlingar (A, B), fyra block (I–IV) och åtta rutor per block. Det finns åtta experimentenheter per block.

När man planerar ett fältförsök, och när man analyserar det statistiskt, måste man alltid beakta att olika delar av fältet ger olika resultat. Även om man behandlar samtliga rutor lika kommer resultaten att variera, till exempel på grund av skillnader i jordmån och konkurrensförhållanden. Om man har delat in fältet i block måste man beakta att det kan finnas en variation mellan blocken. Likaledes om man har delat in fältet i stora rutor och små måste man ta hänsyn till att det kan finnas en variation dels mellan de stora rutorna, och

dels mellan de små. Blocken, storrutorna och smårutorna utgör försökets strata. Ibland är fältet indelat i rader och kolumner, och då är dessa försökets strata. Vanligen skattar man varianserna mellan blocken och smårutorna. I förekommande fall skattar man dessutom varianserna mellan storrutor, kolumner och rader. Eftersom de summerar till försökets totala varians säger man att man skattar försökets varianskomponenter.

I den statistiska analysen jämförs skillnaderna mellan behandlingarna med variansen mellan experimentenheterna. Försöket måste innehålla replikat för att variansen mellan experimentenheterna ska kunna skattas. Med andra ord bör varje behandling undersökas på flera experimentenheter.

I de fall observationsenheten är mindre än experimentenheten, dvs. när man har gjort flera observationer per experimentenhet, sägs observationerna vara så kallade pseudoreplikater. Flera observationer från samma experimentenhet får inte betraktas som äkta replikat. Om behandlingen inte har upprepats på flera experimentenheter finns alltså inga äkta replikat. Då går det inte att skatta spridningen i försöket, dvs. spridningen mellan experimentenheterna, och statistisk analys blir i det närmaste omöjlig.

I fältförsök får man pseudoreplikater om man mäter flera plantor i samma ruta, och behandlingarna slumpats ut på rutor och inte plantor. Det är naturligtvis inte fel att göra många mätningar per experimentenhet, men man måste beakta i den statistiska analysen att observationerna knappast är oberoende. I försök med flera observationer per experimentenhet kan man analysera medelvärdena beräknade per experimentenhet istället för de ursprungliga observationerna.

Oftast observerar man inte bara en enda variabel utan flera. I ett ogräsförsök, till exempel, observerar man inte bara totala ogräsvikten, utan även vikterna för ett antal olika ogräsarter. Man säger att man studerar flera responsvariabler.

Det är också vanligt att försöket omfattar flera förklarande variabler. Om man till exempel dels undersöker ett antal olika gödselmedel, och dels ett antal olika kornsorter omfattar försöket två faktorer: gödselmedel och kornsort. Om försöket omfattar tre gödselmedel säger man att faktorn gödselmedel har tre nivåer. Likaledes om försöket omfattar till exempel fem kornsorter säger man att faktorn kornsort har fem nivåer.

Antalet replikat bör vara så stort att alla för jämförelserna väsentliga varianskomponenter kan skattas tillräckligt väl. Det krävs upprepade observationer för att skatta varianserna. I statistiken är frihetsgraderna ett mått på hur många observationer som finns tillgängliga för skattning av varianser. Antalet frihetsgrader är lika med antalet observationer reducerat med antalet skattade parametrar (konstanter i modellen). Försöket bör alltså omfatta så många observationer som möjligt. Men ju mer komplicerad statistisk modell man använder, dvs. ju fler faktorer försöket omfattar och ju fler nivåer dessa har, desto fler frihetsgrader åtgår till skattningarna av modellens parametrar, och desto färre frihetsgrader finns tillgängliga för skattningarna av försökets varianser. Därför är det lämpligt att redan på planeringsstadiet förvissa sig om att antalet observationer räcker till att skatta samtliga intressanta parametrar och dessutom varianserna. I det följande kommer flera exempel ges på hur sådana frihetsgradsberäkningar görs. Med tio frihetsgrader eller fler brukar skattningen av en

variens bli tillförlitlig.

## 1.2 Randomisering

I ett vetenskapligt försök ska man inte medvetet gynna någon behandling framför någon annan. Alla behandlingar ska jämföras under lika förutsättningar. Detta kan förefalla svårt i ett fältförsök, för vanligen är somliga delar av fältet mer gynnsamma än andra. När någon ruta på fältet kan förmodas vara mer fördelaktig än de andra uppstår frågan vilken behandling den rutan ska få. Det led som blir utvalt blir gynnat, och då har leden har inte jämförts under lika förutsättningar.

Lösningen på detta dilemma heter randomisering. Det innebär att man låter slumpen avgöra vilken behandling rutan ska få. Alla led har i så fall lika stor chans att randomiseras till den gynnsamma rutan. På detta sätt överförs de systematiska skillnaderna mellan rutorna på fältet till slumpmässiga skillnader mellan leden. Detta är utmärkt, för systematiska skillnader kan man inte göra någonting åt, men slumpmässiga kan reduceras med fler observationer. De slumpmässiga felen kan också skattas med statistiska metoder.

Försöksplaner bör som regel alltid randomiseras. Man ska undvika systematiska planer, dvs. planer i vilken behandlingarna läggs i en naturlig ordningsföljd. Det kan nämligen inte uteslutas att ordningen sammanfaller med en gradient på fältet. Så kallade demonstrationsförsök, med ordnade led för visuell demonstration i fält, kan vara lämpliga för sitt syfte men är inte vetenskapliga. En plan som inte är randomiserad är subjektiv och strider följaktligen mot kravet att försöket ska vara objektivt.

Försöksplaner ska randomiseras för att undvika såväl avsiktliga som oavsiktliga systematiska fel. Man gör till exempel avsiktligt fel om man låter ogräsförekomsten styra vilken ruta som ska behandlas med vilket ogräspreparat. Oavsiktligt kan man göra fel till exempel genom att lägga behandlingarna i samma ordning i samtliga block med syftet att underlätta utläggningen eller skötseln av försöket.

I en försöksserie bör de ingående försöken randomiseras var för sig, så att behandlingarna inte läggs i samma ordning i försöken.

## 1.3 Spridningsmått

Ett datamaterial kan kortfattat beskrivas med medelvärdet och något mått på spridningen kring medelvärdet, till exempel variansen, standardavvikelsen eller variationskoefficienten.

Om det finns  $n$  observationer och dessa betecknas  $y_1, y_2, \dots, y_n$ , är medelvärdet

$$m = \frac{y_1 + y_2 + \dots + y_n}{n} .$$

Variansen är, nästan, den genomsnittliga kvadrerade avvikelsen till medelvärdet,

men man dividerar med  $n - 1$  istället för med  $n$ . Variansen är alltså

$$s^2 = \frac{(y_1 - m)^2 + (y_2 - m)^2 + \dots + (y_n - m)^2}{n - 1} \quad (1.1)$$

Standardavvikelsen är roten ur variansen:

$$s = \sqrt{\frac{(y_1 - m)^2 + (y_2 - m)^2 + \dots + (y_n - m)^2}{n - 1}} \quad (1.2)$$

Variationskoefficienten är  $s/m$ , och den uttrycks oftast i procent. Variationskoefficienten anger hur stor standardavvikelsen är i relation till medelvärdet. Den betecknas ofta CV.

**Exempel 1.1.** Tabell 1.1 innehåller  $n = 4$  observationer. Medelvärdet  $m$  är 50. I tabellen finns också alla avvikelser till medelvärdet, och alla kvadrerade avvikelser till medelvärdet. Som mått på spridningen i materialet duger inte den genomsnittliga avvikelser, för den är alltid 0. Istället används summan av alla kvadratavvikelser. Denna summa är 120. Variansen är  $120/(n - 1) = 40$ . Standardavvikelsen är roten ur 40, dvs. 6,32. Variationskoefficienten är standardavvikelsen dividerad med medelvärdet, dvs.  $6,32/50 = 12,6$

Tabell 1.1: Fyra observationer samt deras avvikelser och kvadrerade avvikelser till medelvärdet  $m = 50$ . Den genomsnittliga avvikelser är 0, och den genomsnittliga kvadratavvikelsen är 120.

Obs.	$y$	$y - m$	$(y - m)^2$
1	48	-2	4
2	42	-8	64
3	54	4	16
4	56	6	36
Summa	200	0	120

Ju fler observationer man har, desto säkrare blir medelvärdet. Man säger att medelvärdet  $m$  närmar sig det förväntade värdet  $\mu$  när antalet observationer ökar. Medelvärdet  $m$  är en skattning av det okända medelvärdet  $\mu$ . På samma sätt är  $s^2$  en skattning av den okända variansen  $\sigma^2$ . De okända parametrarna betecknas i statistiken oftast med grekiska bokstäver.

Per definition är  $\sigma^2$  den förväntade kvadrerade skillnaden mellan en slumpmässig observation  $y$  och medelvärdet  $\mu$ . Stickprovsvariansen  $s^2$  är en skattning av  $\sigma^2$ . Egentligen skulle vi vilja använda  $\mu$  i beräkningen 1.1 av  $s^2$ , men eftersom vi inte känner  $\mu$  använder vi istället  $m$ . Därmed underskattar vi kvadratsumman i täljaren något. Man kan emellertid visa att denna underskattning exakt kompenseras av att vi dividerar med  $n - 1$  istället för  $n$ . Genom att dividera med  $n - 1$  istället för  $n$  får vi en skattning av  $\sigma^2$  som genomsnittligt träffar mitt i prick trots att vi räknar med  $m$  istället för  $\mu$ .

För varianser gäller enkla matematiska regler. Låt  $y_1, y_2, \dots, y_n$ , vara  $n$  observationer av en slumpmässig variabel  $Y$  som varierar med variansen  $\text{var}(Y) =$

Tabell 1.2: Spridningsmått för ett enkelt datamaterial ( $m$  betecknar ett medelvärde och  $n$  är antalet observationer vid beräkningen av ett medelvärde)

Spridningsmått	
Varians	$s^2$
Standardavvikelse	$s$
Variationskoefficient	$s/m$
Medelfel	$s/\sqrt{n}$
Medelfel för skillnad	$\sqrt{2s^2/n}$

$\sigma^2$  kring medelvärdet  $\mu$ . Då är variansen i medelvärdet lika med

$$\frac{\sigma^2}{n} . \quad (1.3)$$

Ju fler observationer vi har, dvs. ju större  $n$  är, desto mindre blir variansen i medelvärdet. Man skattar förstas (1.3) med  $s^2/n$ . Standardavvikelsen i medelvärdet skattas med det så kallade medelfelet:

$$\frac{s}{\sqrt{n}} . \quad (1.4)$$

Om  $Y_1$  och  $Y_2$  är två oberoende slumpmässiga variabler, är variansen för  $Y_1 + Y_2$  lika med variansen för  $Y_1$  plus variansen för  $Y_2$ :

$$\text{var}(Y_1 + Y_2) = \text{var}(Y_1) + \text{var}(Y_2) .$$

Likaså är variansen för skillnaden mellan två oberoende variabler lika med summan av varianserna:

$$\text{var}(Y_1 - Y_2) = \text{var}(Y_1) + \text{var}(Y_2) . \quad (1.5)$$

Detta har stor betydelse i fältforskning, eftersom man i fältförsök vanligen jämför två eller flera behandlingar med varandra. Variansen i skillnaden mellan behandlingarna är lika med summan av de båda behandlingarnas varianser. Variansen i en skillnad mellan två medelvärden, beräknade på  $n_1$  respektive  $n_2$  observationer, blir, enligt (1.3) och (1.5),

$$\frac{\text{var}(Y_1)}{n_1} + \frac{\text{var}(Y_2)}{n_2} . \quad (1.6)$$

Ofta har vi lika många observationer av varje led, så att vi kan skriva  $n_1 = n_2 = n$ , och ofta kan vi anta att variansen är lika stor i båda leden, så att  $\text{var}(Y_1) = \text{var}(Y_2) = \sigma^2$ . I så fall är variansen i skillnaden lika med

$$\frac{2\sigma^2}{n} , \quad (1.7)$$

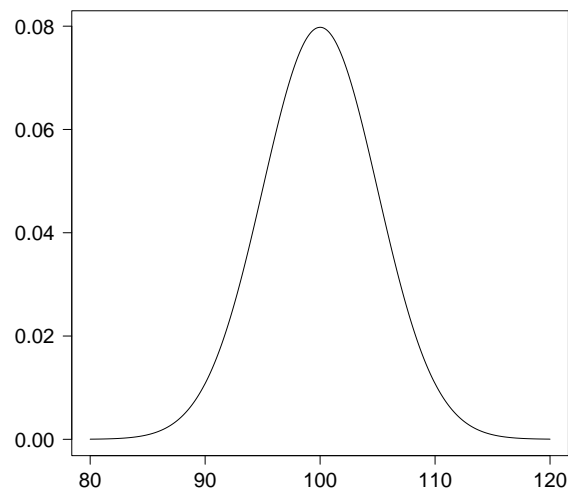
och om  $s^2$  är en skattning av den gemensamma variansen  $\sigma^2$  är "medelfelet för skillnaden" lika med

$$\sqrt{\frac{2s^2}{n}} .$$

Spridningsmått sammanfattas i tabell 1.2. Observera att formlerna i det här avsnittet bara gäller för det enklaste fallet. I randomiserade fullständiga blockförsök, till exempel, skattas  $s^2$  med medelkvadratsumman för residualerna, dvs. MSE, så som beskrivs i kapitel 2.

## 1.4 Normalfördelning

Ofta antar man att observationerna kommer från en normalfördelning. Figur 1.4 visar en normalfördelning med medelvärde 100 och standardavvikelse 5. Normalfördelningen är symmetrisk kring medelvärdet. Observationer nära medelvärdet är vanligare än observationer långt från medelvärdet. Sannolikheten är 5 % att en godtycklig observation avviker från medelvärdet mer än två standardavvikelser (egentligen 1,96 standardavvikelser). Givet normalfördelningen blir standardavvikelsen lätt att tolka: ungefär 95 % av observationerna avviker mindre än 2 standardavvikelser från medelvärdet.



Figur 1.4: Normalfördelning med medelvärde 100 och standardavvikelse 5. Cirka 95 % av fördelningen ligger i intervallet (90 , 110).

## 1.5 Hypotesprövning

Alla fältförsök har naturligtvis ett syfte, men syftet kan vara mer eller mindre tydligt. Oftast syftar försöket till att besvara många frågor. I jämförande försök

är syftet alltid att jämföra de olika leden. Det är dock nödvändigt att specificera syftet eller syftena ytterligare. Man måste förväg veta exakt vilken eller vilka egenskaper hos behandlingarna som ska jämföras. I ett sortförsök kan man till exempel vara intresserad av torrvikten per hektar, tusenkornvikten, kvalitén, stråstyrkan, något sjukdomsindex, koktiden, färgen eller smaken. Man ska klargöra redan innan försöket utförs vilken fysisk parameter som är den mest intressanta, dvs. vilken variabel som är den primära variabeln, och hur och när denna ska mätas. I försöksplanen bör anges vilken variabel som är den primära och vilka andra variabler som är sekundära. Ibland händer det att man iakttar skillnader mellan leden i någon variabel som det inte var tänkt att man skulle undersöka. Till exempel kan det visa sig att somliga led blivit angripna av en speciell sjukdom, men andra inte. Sådana iakttagelser är intressanta och värdefulla för framtida forskning. Men även om skillnaderna är statistiskt signifikanta i den variabeln, så är inte testet teoretiskt giltigt. Försöket kan bara besvara de frågor som ställts före försökets utförande. Ett nytt försök behövs för att undersöka om behandlingarna har betydelse för förekomsten av den speciella sjukdomen.

Den vetenskapliga hypotesen eller teorin handlar ofta om fysiska orsakssamband som yttrar sig i skillnader mellan behandlingarna. Man förväntar sig att behandlingarna ska ge olika resultat. Men den statistiska hypotesen, den så kallade nollhypotesen, är nästan alltid att det inte finns någon skillnad mellan behandlingarna. Vanligen tror vi inte på nollhypotesen, utan önskar kunna förkasta den på grundval av försökets resultat. Om skillnaderna mellan behandlingarna är tillräckligt stora i förhållande till spridningen i försöket kan hypotesen förkastas och skillnader mellan behandlingarna påvisas. Men om skillnaderna är små i förhållande till spridningen kan hypotesen inte förkastas. Lägg märke till att man aldrig kan bevisa en statistisk hypotes. Man kan bara förkasta den eller inte förkasta den. Man kan aldrig visa att det inte finns några skillnader mellan behandlingarna, ty för detta skulle det krävas oändligt många observationer. Däremot kan man ofta påvisa att det finns skillnader, vilket är just vad man vanligen önskar. Det är också möjligt att utforma hypotesprövningen med syftet är att visa att skillnaderna inte är stora. Detta gör man i så kallade ekvivalensstudier, i vilka man vill visa att två eller flera behandlingar har likvärdiga effekter.

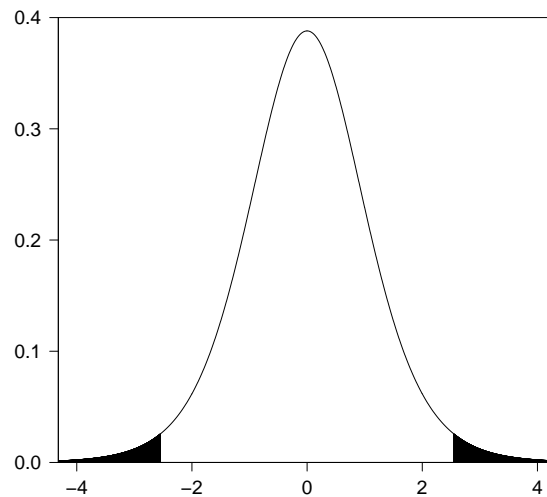
Vid statistisk hypotesprövning beräknas oftast värdet av en så kallad testvariabel. Vi ska helt enkelt kalla detta värde för försökets testvärde. Det är vanligen ett mått på hur stora skillnaderna mellan behandlingarna är i relation till försökets slumpmässiga variation. Exempelvis beräknas testvärdet för hypotesen att medelvärdet är noll så här:

$$t = \frac{m}{s/\sqrt{n}} . \quad (1.8)$$

I ekvation (1.8) är testvärdet  $t$  helt enkelt lika med medelvärdet i relation till medelfelet (se tabell 1.2). Den här testvariabeln använder man vid parvisa mätningar, dvs. när man i ett antal block har gjort mätningar på två rutor behandlade med olika behandlingar. För varje block beräknar man skillnaden mellan

de båda behandlingarna. I ekvation (1.8) är  $m$  den genomsnittliga skillnaden,  $n$  antalet observerade skillnader (dvs. antalet block) och  $s$  standardavvikelsen för skillnaderna, beräknad enligt ekvation (1.2).

Om hypotesen är sann, dvs. om det inte finns några skillnader mellan de båda behandlingarna, har testvariabeln en känd fördelning, förutsatt att skillnaderna är normalfördelade. I så fall är nämligen testvariabeln (1.8) t-fördelad med  $n - 1$  frihetsgrader. Denna fördelning, t-fördelningen, ser ut ungefär som normalfördelningen, men har lite större sannolikhetsmassa i ändarna. Det exakta utseendet beror på antalet frihetsgrader. Ju färre frihetsgrader, desto tyngre ändar. I figur 1.5 visas en t-fördelning med nio frihetsgrader. Det är alltså den t-fördelning man använder vid försök med två behandlingar i tio block.



Figur 1.5: En t-fördelning med 9 frihetsgrader. Sannolikheten att  $|t| > 2,55$  är 0,031.

Man beräknar hur sannolikt det är att av en slump få ett till beloppet så stort testvärde som man har fått, eller ett ännu större, under förutsättning att den statistiska hypotesen är sann. Om testvärdet (1.8) exempelvis är 2,55, beräknar man sannolikheten att av en slump erhålla testvärden större än 2,55, eller mindre än -2,55. Denna sannolikhet illustreras av den mörka massan i figur 1.5. I praktiken gör man beräkningen med hjälp av något datorprogram för statistisk analys. Sannolikheten är 0,031. Man säger att testes sannolikhetsvärde är 0,031. Sannolikhetsvärdet jämförs med testets signifikansnivå, som i teorin ska ha bestämts innan försöket utfördes. I fältforskning, liksom i flertalet andra vetenskaper, används oftast signifikansnivån 5 %. Man kan säga att signifikansnivån är gränsen mellan vad som anses sannolikt och vad som anses osannolikt.



annolikt. Testvärden med sannolikhetsvärde större än signifikansnivån är i så fall sannolika, medan testvärden med sannolikhetsvärde mindre än signifikansnivån är osannolika. Om sannolikhetsvärdet är mindre än signifikansnivån finns, under antagandet om normalfördelning, två möjligheter. Antingen är hypotesen sann och någonting osannolikt har inträffat, eller också är hypotesen falsk. Eftersom man inte gärna tror att någonting osannolikt har inträffat säger man att testet är signifikant (på den valda signifikansnivån) och att skillnader mellan behandlingarna har påvisats.

När försöket omfattar fler än två behandlingar gör man en variansanalys och testar med hjälp av ett så kallat F-test hypotesen att alla behandlingar är lika. Kapitel 2 beskriver variansanalysen för randomiserade fullständiga blockförsök. När F-testet indikerar att behandlingarna har olika effekter vill man som regel undersöka vilka par av behandlingar som ger signifikant olika effekt. Enklast utför man parvisa t-test med testvariabeln

$$t = \frac{m_1 - m_2}{\sqrt{2s^2/n}} \quad (1.9)$$

I ekvation 1.9 är  $m_1$  och  $m_2$  de skattade medelvärdena för de båda behandlingarna som ska jämföras. I nämnaren finns medelfelet för skillnaden. Lägg märke till att skillnaden  $m_1 - m_2$  helt enkelt relateras till sin standardavvikelse. Variansen  $s^2$  hämtas från variansanalysen. I det randomiserade fullständiga blockförsöket skattas  $s^2$  med medelkvadratfelet, så som beskrivs i kapitel 2, och testvärdet jämförs med en t-fördelning med  $(v - 1)(n - 1)$  frihetsgrader, där  $v$  är antalet behandlingar och  $n$  antalet block.

I vid mening inbegriper statistisk hypotesprövning av fältförsök följande moment:

1. Formulera en vetenskaplig teori
2. Bestäm försökets led
3. Specificera vilka variabler som ska observeras, och vilken av dessa som är den primära variabeln
4. Formulera en statistisk hypotes (eller flera)
5. Definiera experimentenheter, observationsenheter och eventuell blockstruktur
6. Specificera hur observationerna ska samlas in
7. Gör en försöksplan med tillräckligt många upprepningar
8. Ange vilka varianskomponenter som ska skattas och hur den statistiska hypotesen ska testas
9. Randomisera behandlingarna till försökets rutor
10. Utför experimentet

11. Beräkna hypotesens testvärde och sannolikhetsvärde
12. Jämför sannolikhetsvärdet med den i förväg bestämda signifikansnivån
13. Förkasta eller förkasta inte den statistiska hypotesen
14. Bedöm den vetenskapliga teorin utifrån resultatet av hypotestestet

## Kapitel 2

# Randomiserade fullständiga blockförsök

### 2.1 Utläggning i fält

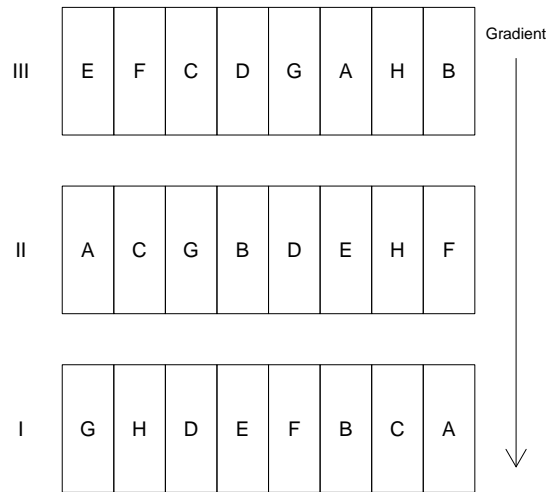
Många fältförsök är randomiserade blockförsök. I ett randomiserat blockförsök delas försöksfältet in i block som i sin tur delas in i rutor. Vanligen har varje block lika många rutor som försöket har behandlingar. Det blir därmed möjligt att jämföra samtliga behandlingar inom varje block. Figur 2.1 visar hur ett randomiserat blockförsök kan se ut. Leden har randomiserats till försökets rutor, inom varje block.

Det randomiserade blockförsöket är lämpligt när det visserligen finns stora variationer över fältet, men det är möjligt att dela in fältet i block inom vilka variationerna är små. Det gäller alltså att definiera block på fältet som är så jämna som möjligt. Blocken ska vara jämna i den meningen att om man behandlade samtliga rutor i ett block på samma sätt skulle rutorna ge snarlika observationer. Vid samma behandling skulle det inte vara stor variation inom blocket, utan liten.

Om man lyckas hitta block som är jämna i detta avseende är de väl ägnade som försöksytor i experimentet. På dessa försöksytor, dvs. inom dessa block, är det lämpligt att jämföra försökets olika behandlingar. Eftersom den slumpmässiga variationen är liten mellan rutorna inom blocken blir det lätt att se skillnaderna mellan behandlingarna, även om dessa skillnader är små.

När försöket omfattar många behandlingar som jämförs på rutor som ligger sida vid sida blir blocken ofta långa. Om det finns en känd jordmånsgradient på fältet bör de långsmala blocken läggas vinkelrätt mot gradienten, så som visas i figur 2.1. På detta sätt minimeras den slumpmässiga spridningen inom blocken, vilket underlättar jämförelserna mellan behandlingarna. Om blockförsöket är fullständigt, så att varje block innehåller exakt en ruta med varje behandling, jämförs samtliga behandlingar såväl på sämre delar av fältet som på bättre.

Rutorna inom blocken ska vara så lika som möjligt i alla tänkbara avseen-



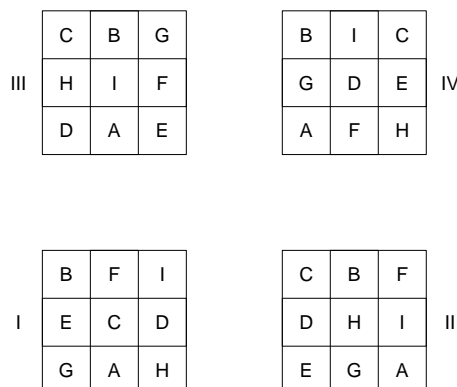
Figur 2.1: Ett randomiserat fullständigt blockförsök med tre block (I–III) och åtta behandlingar (A–H) per block. Blocken ligger vinkelrätt mot jordmånsgradienten för att den slumpmässiga variationen inom blocken ska bli så liten som möjligt.

den. Därför bör man inte bara beakta eventuella jordmånsgradienter när man bestämmer var blocken ska ligga på fältet, utan även fältets lutning och tillgången till vatten och näring. Man bör även under försökets genomförande behandla rutor från samma block så identiskt som möjligt. I Försökshandboken<sup>1</sup> ges fler råd om valet av försöksplats och vad man bör tänka på vid utläggning av försök.

När fältet har två vinkelräta gradienter, snarare än en enda rät, bör blocken läggas vinkelrätt mot den starkaste gradienten. Skulle gradienterna vara ungefär lika stora skulle kvadratiska block vara att föredra framför rektangulära. Istället för att lägga behandlingarna på en enda lång rad i ett avlångt block skulle man kunna placera de  $v$  behandlingarna i  $k$  rader och  $s$  kolumner, förutsatt att  $v = ks$ . På det sättet minimeras blockens utsträckning över fältet, och variationen mellan rutorna inom blocken blir så liten som möjligt. Figur 2.2 visar hur nio behandlingar skulle kunna randomiseras till fyra kvadratiska block. I praktiken är det naturligtvis många gånger svårt att behandla och sköta om rutor i en kvadrat utan några gånger mellan raderna eller kolumnerna. Detta är ett skäl till att enkelradiga block är vanligare än block med flera rader.

Vanligen gör man blocken så stora att de omfattar en ruta per led, men det är förstås möjligt att definiera större block och undersöka något led på två eller flera rutor i varje block. När en av behandlingarna är ett kontrollled kan det

<sup>1</sup>[www.slu.se/sv/fakulteter/nl/om-fakulteten/ovriga-enheter/faltforsk/utbildning/forsokshandbok/](http://www.slu.se/sv/fakulteter/nl/om-fakulteten/ovriga-enheter/faltforsk/utbildning/forsokshandbok/)



Figur 2.2: Fyra randomiserade fullständiga block (I–IV) med nio behandlingar (A–I). Blocken är kvadratiska för att minimera variationen inom blocken.

vara motiverat att undersöka kontrolledet på flera rutor i varje block, nämligen om försöket syftar till att jämföra de övriga leden med kontrolledet snarare än samtliga led med varandra. En behandling bör dock alltid, förutsatt att alla block är lika stora, undersökas på lika många rutor i varje block. I så fall kan nämligen behandlingseffekterna skiljas tydligt från blockeffekterna. Exempelvis kan behandlingen A randomiseras till två rutor per block, och behandlingarna B, C, D och E till en ruta per block. Nackdelen är förstås att blocken blir större om något eller några led undersöks på flera rutor per block än om alla led undersöks på bara en ruta per block. Vid olikstora block bör man, för att kunna särskilja ledeffekter från blockeffekter, planera så att kvoterna mellan behandlingarnas replikatantal är lika stora i samtliga block. I ett försök med två block och tre behandlingar är detta till exempel uppfyllt om led A ligger på tre rutor i första blocket och sex rutor i andra blocket, led B ligger på två rutor i ena blocket och fyra rutor i andra blocket, och led C på en ruta i första blocket och två rutor i andra blocket. I båda blocken undersöks i så fall led A på tre gånger så många rutor som led C, och led B på dubbelt så många rutor som led C.

## 2.2 Statistisk analys och tolkning

Exempel 2.1 illustrerar hur randomiserade fullständiga blockförsök kan analyseras statistiskt.

*Exempel 2.1.* I ett randomiserat fullständigt blockförsök med fyra block jämfördes 13 olika kombinationer av preparat mot svampsjukdomar, applicerade vid olika tillfällen. Försöket inkluderade också ett obehandlat kontrollad, så totalt omfattade försöket 14 led. De olika leden och de observerade medelskördarna (avrundade till tiotal) redovisas i tabell 2.1.

Tabell 2.1: Observerade medelvärden för de 14 leden i Exempel 2.1

Led	Medel kg/ha
A. Obehandlat	8030
B. Proline + Comet och Proline, I	8630
C. Juventus + Comet och Juventus, I	8580
D. Proline + Tilt Top + Comet och Proline + Tilt Top	8260
E. Proline och Proline	8550
F. Proline + Comet och Proline, II	8310
G. Juventus + Comet och Juventus, II	8370
H. Proline + Comet och Proline, III	8330
I. Bravo och Proline+ Comet och Proline	8780
J. Bravo + Proline + Comet och Proline	8460
K. Proline och Proline + Comet	8590
L. Proline + Comet	8520
M. Stereo och Proline + Comet (sortfösk.)	8560
N. Tilt Top och Proline + Comet och Proline	8520

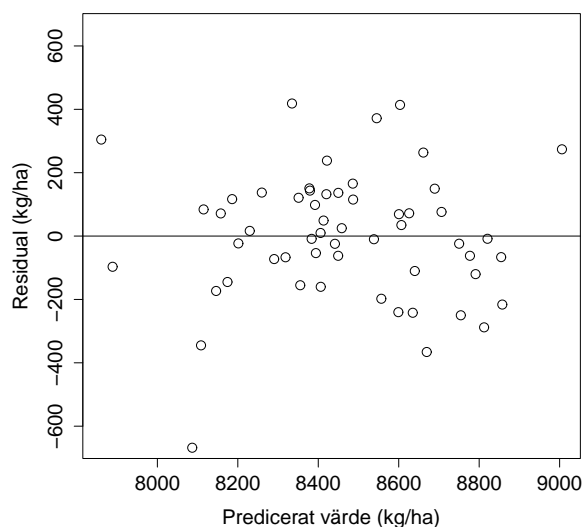
Till att börja med räknar vi ut hur många frihetsgrader vi har för jämförelser av leden. Försöket omfattar totalt 56 observationer, vilket betyder att det totalt finns  $56 - 1 = 55$  frihetsgrader i försöket, för en frihetsgrad går åt till att skatta totala medelvärdet. Det krävs  $4 - 1 = 3$  frihetsgrader för att skatta effekterna av blocken, eftersom det finns fyra block. Kvar finns 52 frihetsgrader för jämförelser av leden. Skattningen av ledens effekter kräver  $14 - 1 = 13$  frihetsgrader, eftersom det finns 14 led. Således kan försöksfelet skattas med 39 frihetsgrader, vilket är mer än 10 och därför rimligen tillräckligt (se avsnitt 1.1). Tabell 2.2 sammanfattar frihetsgradsberäkningarna.

Tabell 2.2: Försöket omfattar två varianskomponenter. Jämförelsen av leden görs på rutorna inom blocken. Variansen mellan rutorna skattas med 39 frihetsgrader

Varianskomponent	Variationsorsak	Frihetsgrader
Block		3
Rutor	Led	13
	Residual	39
	Totalt	52
Totalt		55

Det randomiserade fullständiga blockförsöket brukar analyseras med en linjär modell med block och led som faktorer. De slumpmässiga avvikelserna från

modellen kallas residualer. För att hypotestesten ska vara giltiga krävs att residualerna är oberoende och normalfördelade, samt att de har lika varians. Man bör alltid undersöka residualerna genom att plotta dem mot predicerade värden, så som i figur 2.3. I det här fallet är variansen ungefär lika stor för små predicerade värden som för stora. Residualerna är approximativt normalfördelade kring medelvärdet 0, och det finns inga kraftigt avvikande värden (möjligen med ett undantag). Det finns inget mönster i grafen som indikerar beroende mellan residualerna. Förutsättningarna för analysen kan därmed anses vara uppfyllda.

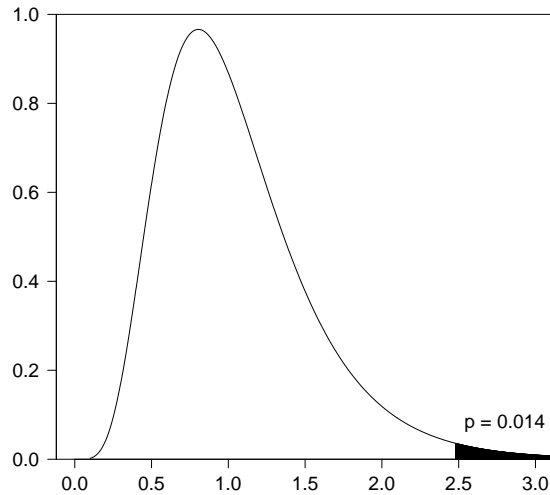


Figur 2.3: Residualerna följer inget mönster, och variansen är ungefär lika stor överallt.

Resultatet av den statistiska analysen av ett randomiserat fullständigt blockförsök kan presenteras i en variansanalystabell som i tabell 2.3. De engelska förkortningarna i tabellen är vanliga.

Tabell 2.3: Variansanalystabell för försöket i exempel 2.1. Frihetsgrader (DF), kvadratsummor (SS), medelkvadratsummor (MS), F-testvärde (F) och sannolikhetsvärde (P)

	DF	SS	MS	F	P
Block	3	1506139	502046	8,82	0,000
Led	13	1835411	141185	2,48	0,014
Residual	39	2220266	56930		
Totalt	55	5561817			



Figur 2.4: F-fördelning med 13 frihetsgrader i täljaren och 39 frihetsgrader i nämnaren. Sannolikheten att  $F > 2,48$  är 0,014

Blockens kvadratsumma är lika med de kvadrerade skillnaderna mellan blockmedelvärdet och totala medelvärdet, summerade över samtliga observationer. På samma sätt är ledens kvadratsumma lika med summan av de kvadrerade skillnaderna mellan ledmedelvärdena och totala medelvärdet. Totala kvadratsumman är summan av alla observationers kvadrerade avvikelser till totala medelvärdet. Residualkvadratsumman kan beräknas utifrån de övriga. Medelkvadratsummorna fås genom division med frihetsgraderna.

Medelkvadratsumman för residualerna är en skattning av försökets varians. Den betecknas ofta MSE (mean squared error). Variansen mellan rutor inom block skattas alltså till 56 930. Således är standardavvikelsen för försöksfelet 238,6 kg/ha. Denna standardavvikelse kallas även residualspridningen. Eftersom totala medelvärdet är 8 460 kg/ha, är variationskoefficienten (CV) lika med 2,8 %.

Variationskoefficienten anger hur stor residualspridningen är i relation till mätvärdenas nivå. En erfaren forskare vet hur stor variationskoefficienten brukar vara och kan därför avgöra om försöket varit jämnt eller inte. En ovanligt hög variationskoefficient kan tyda på att någonting har gått fel under försökets utförande. När variationskoefficienten är hög är det svårare att se skillnaderna mellan leden. Men det är de statistiska testen, inte variationskoefficienten, som avgör om skillnaderna mellan leden är så stora att de går att urskilja eller inte.

F-värdena är medelkvadratsummorna (MS) för block respektive led dividerade med variansen (MSE). F-värdet för block jämförs med en F-fördelning med



3 frihetsgrader i täljaren och 39 i nämnaren, och F-värdet för led jämförs med en F-fördelning med 13 frihetsgrader i täljaren och 39 i nämnaren. Sannolikheten är 0,014 att ett F-värde från en sådan fördelning är större än 2,48, vilket illustreras i figur 2.4. Eftersom sannolikhetsvärdet (P) är mindre än 0,05 förkastas hypotesen att det inte finns några skillnader mellan leden.

När det finns signifikanta skillnader mellan leden undrar man förstas vilka led som skiljer sig åt. Man vill utföra parvisa test. För att skydda sig mot falska signifikanser är det klokt att hålla sig till regeln att bara göra parvisa test när F-testet är signifikant, som i det här fallet. Om försöket omfattar  $v$  led finns nämligen  $v(v-1)/2$  par av led, och när man utför så många hypotestest är det stor risk att något eller några test ger signifikanser bara av en slump. Endast när vi påvisat att det finns signifikanta skillnader mellan leden bör vi därför undersöka vilka led som är olika. Alternativt, och mer effektivt, kan man skydda sig mot falska signifikanser vid parvisa jämförelser genom att använda någon särskild metod för multipla jämförelser (t.ex. Tukeys metod).

I det randomiserade fullständiga blockförsöket kan minsta signifikanta skillnaden, som vanligen betecknas LSD efter engelskans least significant difference, användas för att redovisa vilka par av medelvärden som är signifikant olika. Om antalet replikat är  $n$  och antalet behandlingar är  $v$ , beräknas LSD så här:

$$\text{LSD} = t \sqrt{\frac{2s^2}{n}},$$

där  $s^2$  är variansen i försöket, så att  $s^2 = \text{MSE}$ ,  $n$  är antalet replikat, och  $t$  är den 97,5:te percentilen i en t-fördelning med  $(n-1)(v-1)$  frihetsgrader. När skillnaden mellan två behandlingars medelvärden är större än LSD är skillnaden mellan behandlingarna signifikant på 5 % -nivån. I exemplet är  $n = 4$ ,  $v = 14$ ,  $s^2 = 56\,930$  och  $t = 2,02$ , så  $\text{LSD} = 341$  kg/ha. Med den uppgiften är det lätt att se i tabell 2.1 till exempel att leden A och B givit signifikant olika skörd, men inte leden B och C.

Med hjälp av LSD är det enkelt att beräkna ett 95 % konfidensintervall för skillnaden mellan två behandlingar. Konfidensintervallet är lika med skillnaden mellan medelvärdena  $\pm$  LSD. Till exempel skattas skillnaden mellan ledan B och A till 600 kg/ha (tabell 2.1). Konfidensintervallet är, avrundat till jämna tiotal, (260, 940) kg/ha. Intervallet anger säkerheten i skattningen 600 kg/ha. Eftersom intervallet inte innehåller 0 är skillnaden mellan leden B och A signifikant på 5 % -nivån. Det är rimligt att tro att den sanna effekten av preparatet B är mellan 260 och 940 kg/ha.

Signifikanta skillnader mellan blocken är inget problem. Snarare indikerar det att indelningen i block var framgångsrik. Ibland påstås det att skillnaderna mellan blocken var så stora att det inte gick att se några skillnader mellan behandlingarna. I själva verket kan skillnaderna mellan blocken vara hur stora som helst, i den meningen att somliga block ger höga värden och andra låga, utan att F-testet påverkas. Behandlingarna jämförs ju inom blocken. Ett icke signifikant F-test kan därför inte skyllas på stora nivåskillnader mellan blocken. Om det däremot har funnits ett samspel mellan blocken och leden, så att skill-

naderna mellan leden har varierat mellan blocken kan det vara en förklaring till ett icke signifikant resultat.

## Kapitel 3

# Dimensionering

I fältförsök brukar medelvärden beräknas. Ju fler observationer, desto säkrare blir bestämningarna av medelvärdena. Innan försöket utförs bör man förvissa sig om att det planerade antalet observationer är tillräckligt stort.

Säkerheten i medelvärdet, eller egentligen osäkerheten, anges med medelvärdets varians (1.3). Enligt (1.3) minskar variansen med antalet observationer. Försöket bör ge så många observationer att variansen blir så liten att precisionen blir tillräckligt stor. Vid normalfördelning ligger medelvärdet inom  $\pm$  två medelfel från sitt förväntade värde med sannolikheten 95 % (se avsnitt 1.4). Detta kan användas för att bestämma det nödvändiga antalet observationer.

*Exempel 3.1.* Blastillväxten ska studeras i ett potatisförsök. Valet står mellan att låta försöket omfatta 12 eller 24 plantor. Medelvärdet ska kunna anges med ungefärliga säkerheten  $\pm 10\%$ . Baserat på tidigare erfarenheter förväntas ca 20 cm tillväxt i genomsnitt, och standardavvikelsen förväntas bli högst 5 cm. Enligt (1.3) är standardavvikelsen i medelvärdet 1,44 cm vid 12 observationer och 1,02 cm vid 24 observationer. Medelvärdet  $\pm t$  medelfel, där  $t \approx 2$ , utgör ett 95 % konfidensintervall för medelvärdet. Eftersom  $2 \cdot 1,44 \text{ cm} = 2,88 \text{ cm} > 10\% \cdot 20 \text{ cm}$  räcker inte 12 plantor. Däremot är 24 plantor ett rimligt antal, för  $2 \cdot 1,02 \text{ cm} = 2,04 \text{ cm} \approx 10\% \cdot 20 \text{ cm}$ .

Observera att man måste använda en maximal standardavvikelse i beräkningen. I exempel 3.1 förmodades standardavvikelsen bli lägre än 5 cm, och därför användes detta värde. Om man räknar med en förväntad standardavvikelse är det stor risk att standardavvikelsen blir högre i försöket, och då blir inte konfidensintervallet tillräckligt litet. Kupper och Hafner (1989) visade att stickprovsstorleken blir alldeles för liten om man dimensionerar försöket utifrån en förväntad standardavvikelse. De angav också i tabeller hur stickprovsstorlekar som beräknats baserat på förväntade standardavvikelser kan justeras uppåt så att konfidensintervallet sannolikt blir tillräckligt smalt.

Försöket i exempel 3.1 är ett bestämmande försök med syftet att bestämma

blasttillväxtens storlek. Jämförande försök, i vilka syftet är att jämföra två eller flera behandlingar, är vanligare. I jämförande försök måste variansen för skillnaden minimeras. Denna varians anges av (1.7), eller vid lika många replikat av båda behandlingarna, av (1.6).

*Exempel 3.2.* Effekten av en näringslösning ska undersökas i ett försök. Ledet med näringslösning förmodas ge 20 % större avkastning än ledet utan näringslösning. Vi vill beräkna ett 95 % konfidensintervall för skillningen av effektens storlek, dvs. för skillnaden mellan de båda ledens medelvärden, och önskar att detta konfidensintervall är mindre än  $\pm 5\%$ . Vi räknar med att variationskoefficienten (CV) blir högst 6 %. Konfidensintervallet är, uttryckt i procentenheter,

$$\text{Skillnaden}(\%) \pm t \sqrt{\frac{2 \text{CV}(\%)^2}{n}}. \quad (3.1)$$

Räknat med  $t \approx 2$  och  $\text{CV} = 6\%$  har vi ekvationen

$$2 \sqrt{\frac{2 \cdot 6^2}{n}},$$

vilket ger  $n = 11,52$ . Således behövs minst  $n = 12$  replikat. Avkastningens storlek har ingen betydelse för beräkningen.

I exempel 3.1 och 3.2 räknade vi med att t-fördelningens 97,5:te percentil var ungefär lika med 2. Detta är sant bara om antalet frihetsgrader är stort. Tabell 3.1 anger t-fördelningens 97,5:te percentil vid olika frihetsgradsantal. För mer noggranna beräkningar kan dessa värden användas.

*Exempel 3.3.* Minsta signifikanta skillnaden (LSD) är halva konfidensintervallets bredd, så

$$\text{LSD}(\%) = t \sqrt{\frac{2 \text{CV}(\%)^2}{n}}, \quad (3.2)$$

enligt (3.1). Antag att  $\text{CV} = 5\%$ . Med  $n = 4$  replikat blir  $\text{LSD} \approx 7\%$ , förutsatt att vi kan räkna med att  $t \approx 2$ , vilket är uppfyllt om antalet frihetsgrader är stort. I ett randomiserat fullständigt blockförsök är antalet frihetsgrader  $(v - 1)(n - 1)$ , där  $v$  är antalet led. Om försöket bara omfattar tre behandlingar är därför  $t = 2,45$  vid fyra replikat, enligt tabell 3.1. Då blir  $\text{LSD} = 8,7\%$  vid  $\text{CV} = 5\%$ . Vid fem replikat blir  $\text{LSD} = 7,3\%$ , och vid sex replikat blir  $\text{LSD} = 6,4\%$ .

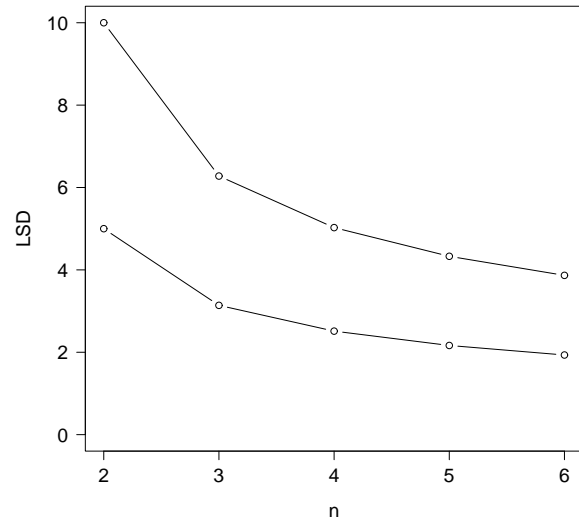
Som exempel 3.3 visar är det lätt att uppskatta vad LSD skulle ha varit vid ett större antal replikat om man vet vad LSD brukar vara vid ett mindre antal replikat. Den exakta beräkningen beror dock på hur många led försöket omfattar. Figur 3.1 illustrerar hur LSD förändras med antalet replikat i försök med fyra behandlingar. Om exempelvis LSD är 10 % vid två replikat skulle LSD ha varit knappt 4 % vid sex replikat. Om LSD är 5 % vid två replikat skulle

Tabell 3.1: t-fördelningens 97,5:te percentil ( $t$ ) vid 16 olika frihetsgradsantal (Fg) samt vid oändligt många frihetsgrader

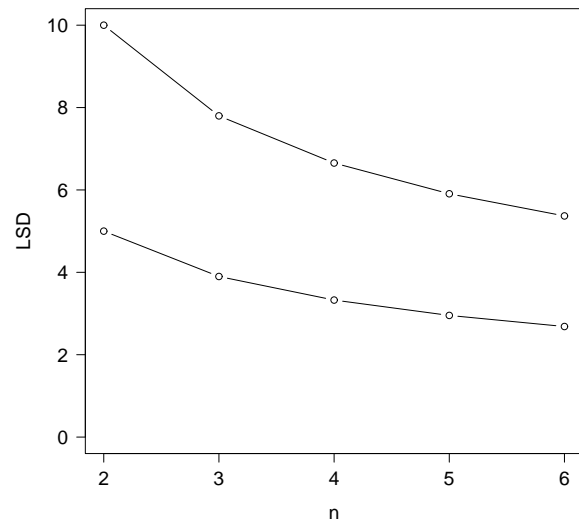
Fg	$t$
1	12.71
2	4.30
3	3.18
4	2.78
5	2.57
6	2.45
7	2.36
8	2.31
9	2.26
10	2.23
12	2.18
15	2.13
20	2.09
30	2.04
60	2.00
100	1.98
$\infty$	1.96

LSD ha varit 2 % i ett försök med sex replikat, förutsatt att försökets varians varit densamma. Figur 3.2 visar på samma sätt hur LSD förändras med antalet replikat i ett försök med 15 led.

Om försöksfelet kan skattas med minst tio frihetsgrader är  $t$  ungefär lika med 2, eller närmare bestämt mindre än 2,3. Detta är skälet till att man helst vill ha tio frihetsgrader eller fler för skattningen av försöksfelets varians. I små randomiserade blockförsök kan det hända att antalet frihetsgrader blir mindre än tio. I tabell 3.2 listas nio små blockförsök med få frihetsgrader. Antalet frihetsgrader är  $(v-1)(n-1)$  om  $v$  är antalet led och  $n$  är antalet block. Frihetsgradsantalet kan ökas antingen genom att inkludera fler block eller genom att inkludera fler led i försöket. Om man gör ett fältförsök större finns dock risken att variansen ökar på grund av att en större försöksyta måste användas, eftersom rutor som ligger långt från varandra på fältet ofta ger mer olika observationer än rutor som ligger nära varandra. Alternativt kan man överväga att använda en fullständigt randomiserad design, dvs. en design utan block. I en sådan randomiserad behandlingarna till experimentenheter utan några restriktioner i randomiseringen. Då åtgår inga frihetsgrader till skattningar av skillnader mellan block. I en fullständigt randomiserat försök skattas försöksfelet med  $v(n-1)$  frihetsgrader, där  $v$  är antalet led och  $n$  är antalet replikat per led. Om det finns en spridning mellan blocken blir dock skattningarna av skillnaderna mellan leden säkrare om leden jämförs inom block i ett randomiserat blockförsök än om leden jämförs på hela försöksytan i ett fullständigt randomiserat försök. Så om det lilla försöket syftar till att skatta skillnader som man redan vet finns är det ran-



Figur 3.1: LSD som funktion av antalet replikat i ett randomiserat fullständigt blockförsök med fyra behandlingar.



Figur 3.2: LSD som funktion av antalet replikat i ett randomiserat fullständigt blockförsök med 15 behandlingar.

domiserade blockförsöket att föredra, förutsatt att det finns en variation mellan blocken. Men om det lilla försökets syfte är att påvisa skillnader med hjälp av ett signifikanstest kan det fullständigt randomiserade försöket vara att föredra, eftersom det ger fler frihetsgrader till F-testets nämnare.

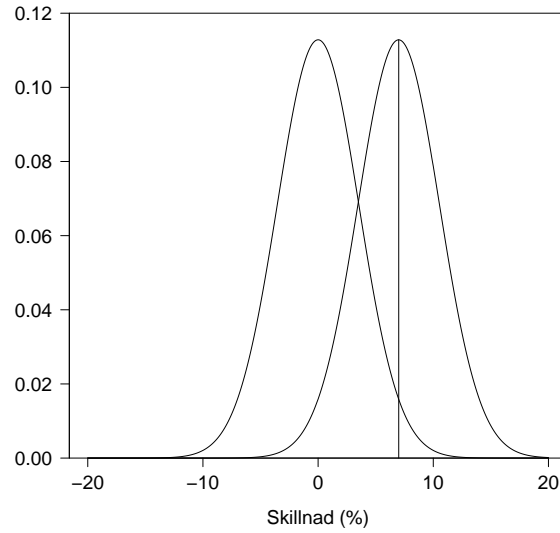
Tabell 3.2: Antal frihetsgrader (Fg) och t-fördelningens 97,5:te percentil ( $t$ ) i nio små fullständiga blockförsök

Antal led	Antal led	Fg	$t$
2	2	1	12.7
2	3	2	4.3
2	4	3	3.2
2	5	4	2.8
2	6	5	2.6
3	2	2	4.3
4	2	3	3.2
5	2	4	2.8

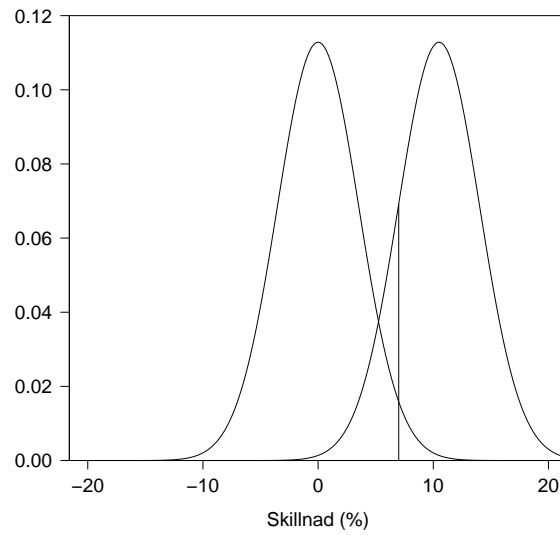
Vid planering av försök bör man även beakta hypotestestets styrka. Testets styrka är sannolikheten att förkasta en hypotes som är falsk. Styrkan bör vara förstås vara stor, och den ökar med antalet replikat. Man vill gärna att styrkan ska vara minst 80 % eller 90 %. Styrkan beror på hur felaktig hypotesen är.

*Exempel 3.4.* Antag att  $CV = 5\%$ . Enligt ekvation (3.2) är  $LSD \approx 7\%$  vid fyra replikat, förutsatt att försöket omfattar så många led att  $t \approx 2$ . Hypotesen att det inte finns någon skillnad mellan två givna behandlingar av intresse förkastas om skillnaden mellan behandlingarnas medelvärden är större än  $LSD$ , dvs. om skillnaden mellan behandlingarna uppmäts vara större än större än 7 %. Det betyder att om den sanna okända skillnaden mellan de båda behandlingarna är 7 %, så är sannolikheten 50 % att hypotesen förkastas. Då är styrkan 50 %. Detta illustreras i figur 3.3. Styrkans beroende av hur felaktig hypotesen är visas i figurerna 3.4 och 3.5. I figur 3.4 är sanna skillnaden 10,5 %. Då är styrkan 84 %. I figur 3.5 är sanna skillnaden 14 %. Vid en så stor skillnad mellan behandlingarna är sannolikheten 97,5 % att hypotesen förkastas.

Det finns inte någon explicit formel för hur många replikat som behövs för att få en given styrka i det statistiska testet. Om man vill räkna exakt får man prova sig fram genom att beräkna styrkan för olika tänkbara antal replikat. Alternativt kan man använda följande approximativa formel.

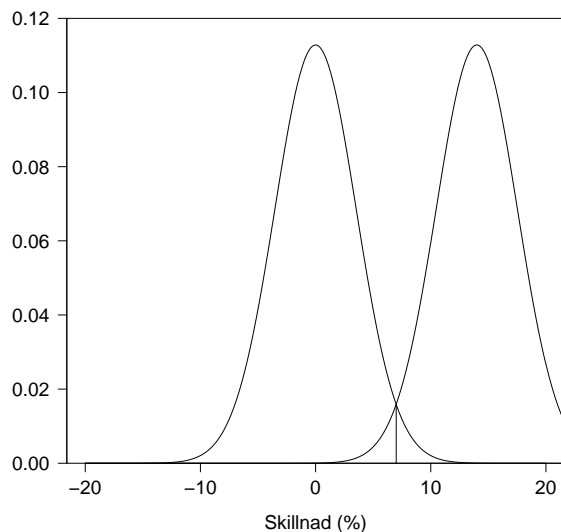


Figur 3.3: Om den sanna skillnaden är 7 % och LSD är 7 % är styrkan 50 %.



Figur 3.4: Om den sanna skillnaden är 10,5 % och LSD är 7 % är styrkan 84 %.





Figur 3.5: Om den sanna skillnaden är 14 % och LSD är 7 % är styrkan 97,5 %.

*Approximativ formel:* Om den sanna skillnaden mellan behandlingarna är  $\Delta$  behövs, för ett hypotestest med styrkan  $1 - \beta$  på  $\alpha$ -nivån,

$$n = \frac{2\sigma^2(z_{1-\alpha/2} + z_{1-\beta})^2}{\Delta^2} \quad (3.3)$$

replikat, där  $\sigma^2$  är variansen och  $z_p$  är den  $p$ :te percentilen i en normalfördelning med väntevärde noll och standardavvikelse 1.

*Exempel 3.6.* Antag att  $\sigma = 5$  och  $\Delta = 10,5$ . För ett test på 5 % -nivån krävs, eftersom  $z_{1-0.05/2} = 1,96$  och  $z_{1-0,15} = 1,04$ , enligt (3.3) minst  $n > 4,1$  replikat per led för att styrkan i testet ska vara minst 85 %. I praktiken får man välja mellan att använda fyra replikat per led och acceptera en lite sämre styrka än 85 % eller att använda fem replikat per led.

Det är ännu enklare att använda följande tumregel.

*Tumregel:* För att detektera en skillnad  $\Delta$  krävs

$$n = 18 \left( \frac{CV(\%)}{\Delta(\%)} \right)^2$$

replikat.

*Exempel 3.7.* Om  $CV = 5\%$  och  $\Delta = 10,5\%$  krävs enligt tumregeln  $n \approx 4$  replikat. Detta är just den situation som illustreras i figur 3.4. Det gäller generellt att tumregeln ger en styrka på ca 85 %.

## Kapitel 4

# Blandade modeller

### 4.1 Fältförsök har flera varianskomponenter

I den ursprungliga variansanalysen finns det bara en enda slumpmässig felterm. Observationernas avvikelser från det totala medelvärdet förmodas vara en summa av behandlingarnas effekter och ett slumpmässigt fel. Detta fel antas oftast vara normalfördelat. Alla försökets faktorer jämförs mot en och samma varians, nämligen medelkvadratsumman (MSE) i variansanalystabellens enda residualrad (t.ex. tabell 2.3).

I fältförsök är det lätt att tänka sig att det finns flera varianskomponenter. I split-plot-försöket (avsnitt 6.1) delas fältet in i stora rutor och små, varvid det är rimligt att tänka sig att det finns en varians mellan de stora rutorna och en annan mellan de små rutorna. I ett rad-kolumn-försök skulle man kunna tänka sig att variansen är större mellan rader än mellan kolumner, särskilt om raderna lagts vinkelrätt mot en känd gradient i fältet. I en försöksserie omfattande ett antal försök utförda på olika platser i Skåne, är det rimligt att tänka sig att det finns en variation mellan platserna och en annan variation inom platserna, dvs. mellan replikaten och rutorna. Förmodligen varierar observationerna mer mellan platserna än inom platserna. Effekterna av platserna kan betraktas som slumpmässiga om platserna är eller liknar ett slumpmässigt urval av tänkbara platser i Skåne. I en försöksserie omfattande försök från flera år kan man betrakta effekterna av åren som slumpmässiga. Effekterna beror egentligen på variationer i väder och vind. I så fall har effekterna av åren en slumpmässig variation. Observationer från försök utförda under samma år varierar förmodligen mindre än observationer från försök utförda under olika år.

I alla dess fall och många flera är modeller med flera varianskomponenter användbara. Sådana modeller kallas blandade (på engelska: mixed models), eftersom de innehåller både fixa faktorer och slumpmässiga. Försökets olika behandlingar brukar vanligen modelleras som fixa. Med andra ord antas effekterna av behandlingarna vara fixa, vilket betyder att de inte varierar slumpmässigt. Det är framför allt dessa fixa effekter som vi är intresserade av att kunna skat-

ta och jämföra med varandra. I nya försök eller i praktiska tillämpningar ska vi kunna upprepa samma behandling och få samma effekt. I stora försök med många led, där syftet är att rangordna leden snarare än att göra parvisa jämförelser mellan dem, kan det dock vara bättre att modellera ledeffekterna som slumpmässiga än att modellera dem som fixa, för det kan ge bättre skattningar. I växtförädlingsförsök och större sortförsök är det därför inte ovanligt att man väljer att modellera behandlingarna som slumpmässiga (Smith, Cullis och Gilmour, 2001; Galwey, 2006).

Årseffekter kan gärna modelleras som slumpmässiga. I en försöksserie omfattande flera år har varje år en effekt på resultaten. Årsmånen varierar till synes slumpmässigt, för den är svår att förutsäga och går inte att upprepa.

Platser kan betraktas som fixa eller slumpmässiga. Om man är intresserad av just de platser försöksserien omfattar, och vill skatta effekterna av just de platserna och jämföra just de platserna med varandra, då ska de modelleras som fixa. Men om man snarare betraktar platserna som representativa för en större region av intresse bör de modelleras som slumpmässiga. Man kan säga att platserna i så fall modelleras som om de var slumpmässigt valda från regionen. Effekten av den första slumpmässigt valda platsen är slumpmässig, liksom effekten av den andra, och så vidare.

På samma sätt förhåller det sig med block, storrutor, rader och kolumner på fältet. Effekterna av dessa kan modelleras som fixa eller som slumpmässiga. I den traditionella variansanalysen betraktas de alla som fixa. Fördelen med att istället betrakta dem som slumpmässiga är framför allt att det möjliggör olika statistiska test för behandlingar som ligger på olika stora experimentenheter. I split-plot-försök (avsnitt 6.1) ligger den ena behandlingen på stora rutor och den andra på små rutor. Då bör faktorn som ligger på de stora rutorna jämföras mot variationen mellan de stora rutorna, och faktorn som ligger på de små rutorna bör jämföras mot variationen mellan de små rutorna. Den slumpmässiga variationen mellan de stora rutorna måste alltså skattas för att det ska bli möjligt att testa behandlingen som ligger på de stora rutorna. Med andra ord är det, för det statistiska testets skull, nödvändigt att modellera effekterna av de stora rutorna som slumpmässiga.

I rad-kolumn-försök med flera faktorer kan somliga faktorer ha randomiserats till rader och andra till kolumner eller rutor. Även i detta fall blir det för de statistiska testens skull nödvändigt att dela upp den totala variansen i varianskomponenter. Man måste skatta varianserna mellan rader, mellan kolumner och mellan rutor. Det betyder att rader och kolumner måste modelleras som slumpmässiga i det fallet, och inte fixa.

I ofullständiga blockförsök (kapitel 5) förekommer somliga par av behandlingar tillsammans i block fler gånger än andra par av behandlingar. Storleken på den slumpmässiga variationen mellan blocken är avgörande för hur väl skillnaden mellan två behandlingar som ligger i olika block skattas. Den slumpmässiga variationen mellan blocken måste skattas, dvs. blockeffekterna måste modelleras som slumpmässiga. I analys av gitterförsök (avsnitt 5.1), blir skattningarna av behandlingsskillnaderna ofta helt olika om blocken modelleras som slumpmässiga eller om de modelleras som fixa. I det randomiserade fullständiga

blockförsöket (kapitel 2), däremot, jämförs samtliga par av behandlingar inom blocken, och då spelar variationen mellan blocken ingen roll för jämförelserna. De statistiska testen för skillnaderna mellan blocken blir i det specifika fallet desamma oavsett om effekterna av blocken modelleras som slumpmässiga eller som fixa (förutsatt att det finns en variation mellan blocken). Men om någonting går fel i försöket, så att observationer från somliga block måste tas bort före analys, har valet av modell viss betydelse. I det fallet finns information om skillnaderna mellan behandlingarna inte bara inom blocken, utan också mellan blocken. Denna information tas till vara om blockeffekterna modelleras som slumpmässiga istället för fixa.

Varianskomponenter skattas vanligen med maximum likelihood-metoden, förkortat ML, eller ännu vanligare med den variant av maximum likelihood-metoden som brukar förkortas REML. Denna förkortning betyder restricted maximum likelihood eller residual maximum likelihood (båda benämningarna används). Enkelt uttryckt innebär maximum likelihood-metoden att man antar att observationerna följer en sannolikhetsfördelning med okända parametrar. I fältförsök är det ofta rimligt att anta att observationerna är normalfördelade med okända fixa medelvärden som beror på behandlingarna, och ett antal okända varianskomponenter som ger slumpmässiga effekter. Då är de fixa behandlingseffekterna samt varianskomponenterna fördelningens okända parametrar. Som skattning av parametrarna tar man de parametervärden för vilka observationerna är mest sannolika. Med varianten REML skattar man först varianskomponenterna, oberoende av de fixa effekterna, och därefter de fixa effekterna. Det matematiska optimeringsproblemet är mångdimensionellt, och det finns ingen explicit lösning, utan parametrarna måste skattas numeriskt med ett datorprogram utvecklat för uppgiften. Programmet söker sig fram till den bästa lösningen. Om inte modellen är alltför komplicerad eller antalet observationer alltför litet konvergerar algoritmen, och de parametervärden för vilka observationerna är mest sannolika, har hittats.

## 4.2 Ett enkelt sätt att specificera modeller

En linjär statistisk modell med två fixa faktorer, A och B, skrivs ofta

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij} , \quad (4.1)$$

där  $y_{ij}$  är den observation som hör till den  $i$ :te nivån av faktorn A och den  $j$ :te nivån av faktorn B, där vidare  $\alpha_i$  är den  $i$ :te effekten av faktorn A,  $\beta_j$  är den  $j$ :te effekten av faktorn B och  $e_{ij}$  är ett slumpmässigt fel, den så kallade residualen. Detta fel antas ofta vara normalfördelat med väntevärde noll och en okänd varians  $\sigma^2$ .

Det finns ett enklare sätt att specificera linjära modeller (se t.ex. Piepho, Büchse och Emrich, 2003) som vi kommer att använda det i det följande. Modellen (4.1) skrivs enligt detta sätt kortfattat

$$A + B . \quad (4.2)$$

I (4.2) är det underförstått att A och B är fixa faktorer och att det dessutom finns en slumpmässig term för residualen. För att vara tydligare kan man skriva ut faktorernas namn. Om A är en behandlingsfaktor och B en blockfaktor, kan (4.2) skrivas

$$\text{Led} + \text{Block} . \quad (4.3)$$

Modell (4.3), som alltså är samma som (4.2) och (4.1), är den traditionella modellen för analysen av det randomiserade fullständiga blockförsöket. Denna modell användes i avsnitt 2.2.

En modell med två fixa faktorer, A och B, och ett samspel mellan de båda fixa faktorerna, kan skrivas

$$A + B + A \cdot B .$$

När det finns flera slumpmässiga faktorer, dvs. när modellen är blandad, kan man skriva de fixa faktorerna först och de slumpmässiga därefter med ett kolon emellan. Så om försöket, utöver de fixa faktorerna A och B och samspelet A · B innehåller effekter av rader och kolumner på fältet, kan modellen skrivas

$$A + B + A \cdot B : \text{Rad} + \text{Kolumn} . \quad (4.4)$$

Observera att modellen (4.4) innehåller tre varianskomponenter: Det finns en variation mellan raderna, en variation mellan kolumnerna och en variation mellan residualerna.

### 4.3 Slumpmässiga blockeffekter

Vid analys av randomiserade blockförsök brukar residualtermen ofta antas vara slumpmässig och normalfördelad. Residualvariansen är ett mått på den slumpmässiga variationen mellan rutorna inom blocken. Det är lätt att tänka sig att det finns slumpmässig variation inte bara mellan rutorna inom blocken utan också mellan blocken.

När effekterna av blocken i det randomiserade blockförsöket antas vara slumpmässiga används modellen

$$\text{Led} : \text{Block} . \quad (4.5)$$

*Exempel 4.1.* I exempel 2.1 användes modellen (4.3) för analys av ett randomiserat fullständigt blockförsök. Residualvariansen skattades till 56 930. Standardavvikelsen skattades således till 238,6 kg/ha, och variationskoefficienten till 2,8 %. Med modellen (4.5) skattas inte bara variansen mellan rutorna inom blocken, utan även variansen mellan blocken. Metoden REML ger variansen 56 930 mellan rutor inom block, precis som i exempel 2.1. Variansen mellan block skattas till 31 794. Standardavvikelsen är roten ur variansen, så standardavvikelsen mellan blocken skattas till 178,3 kg/ha, vilket i relation till

medelvärdet 8 460 kg/ha är 2,1 %. Variationskoefficienten, CV, är alltså 2,1 % mellan block och 2,8 % inom block.

REML-metoden har fördelen jämfört med ursprungliga ML-metoden att den ger samma skattning av residualvariansen som den traditionella analysen i fullständiga randomiserade blockförsök. Med ML-metoden skattas variansen inom block, dvs. residualvariansen, till 42 693, och variansen mellan block till 23 846. Därför brukar REML-metoden föredras framför ML-metoden.

De statistiska testen för ledeffekterna blir exakt desamma med modell (4.5) som med modell (4.3) när REML-metoden används. Hypotesen att det inte finns några skillnader mellan leden testas med ett F-test. Testvärdet är 2,48, vilket jämförs med en F-fördelning med 13 frihetsgrader i täljaren och 39 frihetsgrader i nämnaren, vilket ger sannolikhetsvärdet 0,014. Således förkastas hypotesen att det inte finns några skillnader mellan leden.

För slumpmässiga faktorer brukar man inte göra några signifikanstest.

## 4.4 Romerska kvadrater

Vi har redan sett exempel på hur man kan planera sitt försök så att det kompenserar för gradienter i två riktningar på fältet. Om antalet behandlingar  $v$  kan skrivas  $v = ks$ , kan block med  $k$  rader och  $s$  kolumner användas, så som figur 2.2 visade.

I en så kallad rad-kolumn-design är experimentenheterna ordnade i rader och kolumner. Den romerska kvadraten är en rad-kolumn-design. Den omfattar  $v^2$  rutor ordnade i  $v$  rader och  $v$  kolumner. På dessa rutor jämförs  $v$  behandlingar. De  $v$  behandlingarna randomiseras till rutorna på ett sådant sätt att varje behandling förekommer en gång i varje rad och en gång i varje kolumn. I figur 4.1 visas tre romerska kvadrater.

A	B	C	D
B	A	D	C
C	D	A	B
D	C	B	A

A	C	D	B
B	D	C	A
C	A	B	D
D	B	A	C

A	D	B	C
B	C	A	D
C	B	D	A
D	A	C	B

Figur 4.1: Tre parvist ortogonala romerska kvadrater med fyra behandlingar (A–D).

Före utläggning i fält ska den romerska kvadraten randomiseras. Först permuterar man slumpmässigt raderna, och därefter kolumnerna (eller tvärtom). I den statistiska analysen av den romerska kvadraten skattas såväl spridningen mellan raderna som spridningen mellan kolumnerna. Följande modell används:

$$\text{Led} : \text{Rad} + \text{Kolumn}$$

I försök med romerska kvadrater fördelas frihetsgraderna så som visas i tabell 4.1. Eftersom kvadraten består av  $v^2$  rutor finns totalt  $v^2 - 1$  frihetsgrader. Emellertid går  $v - 1$  frihetsgrader åt till skattningen av effekterna av raderna, och  $v - 1$  frihetsgrader till skattningen av effekterna av kolumnerna. Kvar finns  $(v^2 - 1) - 2(v - 1) = (v - 1)^2$  frihetsgrader. Men jämförelsen av behandlingarna kräver också  $v - 1$  frihetsgrader. Således finns bara  $(v - 1)^2 - (v - 1) = (v - 1)(v - 2)$  frihetsgrader kvar till skattningen av variansen mellan rutorna. Det är mot denna varians behandlingarna testas, och därför måste den vara väl skattad. Kravet att försöksfelets varians ska skattas med minst 10 frihetsgrader (se kapitel 3) skrivs  $(v - 1)(v - 2) > 10$ , vilket medför  $v > 4$ . Romerska kvadrater med färre än fem led bör därför undvikas, såvida det inte är möjligt att låta försöket omfatta flera romerska kvadrater. Det är vanligt att lägga ut flera kvadrater i försöket, eftersom det är ett enkelt sätt att öka antalet frihetsgrader i skattningen av residualvariansen.

Tabell 4.1: I en romersk kvadrat med  $v$  led finns totalt  $v^2 - 1$  frihetsgrader. Variansen mellan kolumnerna skattas med  $v - 1$  frihetsgrader, liksom variansen mellan raderna. Leden jämförs mot residualvariansen som skattas med  $v^2 - 3v + 2$  frihetsgrader.

Variationskomponent	Variationsorsak	Frihetsgrader
Kolumner		$v - 1$
Rader		$v - 1$
Rutor	Led Residual Totalt	$v - 1$ $(v - 1)(v - 2)$ $(v - 1)^2$
Totalt		$v^2 - 1$

När försöket omfattar flera romerska kvadrater för  $v$  behandlingar kan man randomisera varje romersk kvadrat för sig. I så fall permuteras först raderna inom varje kvadrat, och därefter kolumnerna inom varje kvadrat. I analysen skattar man spridningen mellan kvadraterna, mellan raderna inom kvadraterna och mellan kolumnerna inom kvadraterna, enligt följande modell:

$$\text{Led} : \text{Kvadrat} + \text{Kvadrat} \cdot \text{Rad} + \text{Kvadrat} \cdot \text{Kolumn}$$

Med  $r$  kvadrater erhålls  $(v - 1)(r(v - 1) - 1)$  frihetsgrader för skattningen av residualvariansen, enligt tabell 4.2.

Romerska kvadrater som är parvist ortogonala är av särskilt intresse. Om man lägger en romersk kvadrat på en annan får man par av behandlingar i



Tabell 4.2: I ett försök med  $r$  romerska kvadrater för  $v$  led finns totalt  $rv^2 - 1$  frihetsgrader. Om raderna och kolumnerna randomiserats separat inom varje block skattas variansen mellan kolumnerna och mellan raderna med  $r(v - 1)$  frihetsgrader vardera. Leden jämförs mot residualvariansen som skattas med  $(v - 1)(r(v - 1) - 1)$  frihetsgrader.

Varianskomponent	Variationsorsak	Frihetsgrader
Kvadrater		$r - 1$
Kolumner		$r(v - 1)$
Rader		$r(v - 1)$
Rutor	Led Residual Totalt	$v - 1$ $(v - 1)(r(v - 1) - 1)$ $r(v - 1)^2$
Totalt		$rv^2 - 1$

varje ruta. Om i så fall varje behandling från den ena kvadraten förekommer tillsammans med varje behandling från den andra kvadraten exakt en gång sägs kvadraterna vara parvist ortogonala. De tre romerska kvadraterna i figur 4.1 är parvist ortogonala. Det kan aldrig finnas fler än  $v - 1$  parvist ortogonala romerska kvadrater med  $v$  led. För somliga romerska kvadrater finns det inte någon romersk kvadrat sådan att de tillsammans är parvist ortogonala. Speciellt gäller detta alla romerska kvadrater med sex led. Parvist ortogonala romerska kvadrater utnyttjas vid design av gitterförsök, vilket beskrivs i kapitel 5.



## Kapitel 5

# Fältförsök med ofullständiga block

När antalet behandlingar är stort blir blocken stora i det randomiserade fullständiga blockförsöket. Detta är ett problem, för när blocken blir stora ökar sannolikt spridningen inom blocken. De täcker ju en större yta, och rutor som ligger långt från varandra är ofta mer olika varandra än rutor som ligger nära varandra på fältet. Av denna anledning bör inte blocken vara för stora. Det är som regel olämpligt att använda block med fler än 15 rutor.

Om man ändå vill jämföra många behandlingar i experimentet blir följden att blocken inte rymmer samtliga behandlingar. I denna situation är fältförsök med ofullständiga block användbara. Blocken sägs vara ofullständiga när de inte innehåller försökets samtliga behandlingar.

När man säger att ett försök är balanserat brukar man mena ungefär att det finns lika många observationer av alla behandlingar. Men för ofullständiga blockförsök har termen balans en särskild betydelse. Ofullständiga blockförsök sägs vara balanserade om och endast om samtliga par av behandlingar förekommer tillsammans i block lika många gånger. Figur 5.1 åskådliggör begreppet. Försöket i figur 5.1 är balanserat eftersom alla par av behandlingar förekommer tillsammans i tre block. Behandlingen A förekommer till exempel tillsammans med behandlingen C i tre block, nämligen i blocken II, IV och V. När det ofullständiga blockförsöket är balanserat jämförs samtliga par av behandlingar med samma precision, trots att blocken är ofullständiga. Detta är ofta önskvärt.

### 5.1 Gitterdesign

Gitterdesignen är den vanligaste ofullständiga blockdesignen. Den är mer känd under sitt engelska namn: lattice design. Figur 5.2 visar en gitterdesign för 16 behandlingar i 5 replikat. Block I är första kolumnen i första replikatet, block II är andra kolumnen i första replikatet, och så vidare. Block XX är således det fjärde blocket i det femte replikatet. I block I jämförs fyra behandlingar: A, B,

V	C	A	D	E
IV	D	A	C	B
III	D	C	B	E
II	B	E	C	A
I	B	E	A	D

Figur 5.1: Ett balanserat ofullständigt blockförsök med fem block (I–V) och fem behandlingar (A–E), men bara fyra rutor per block. Försöket är balanserat, för varje par av behandlingar förekommer tillsammans i tre block.

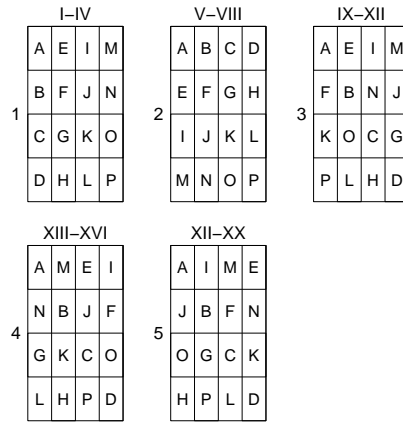
C och D, i block II jämförs behandlingarna E, F, G och H, och så vidare.

Ofta läggs replikatens rutor bredvid varandra på en lång rad på fältet, precis som i ett vanligt randomiserat blockförsök. Det är också vanligt att planen inte omfattar samtliga gitterdesignens replikat. Till exempel kan en försöksplan med två replikat, baserad på designen i figur 5.2, se ut som i figur 5.3. Lägga märke till att varje block i figur 5.3 innehåller precis samma led som motsvarande block i figur 5.2, men att leden har randomiserats till rutorna inom blocken. Blocken har också randomiserats inom replikaten, liksom replikaten till de båda platserna på fältet.

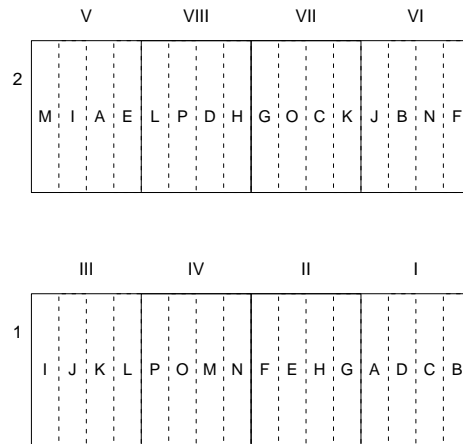
Det är inte nödvändigt att randomisera behandlingarna till ledbeteckningarna. Om försöket, som i figur 5.3, omfattar de två första replikaten från figur 5.2 kommer precisionen i jämförelsen av leden A och P att bli sämre än precisionen i jämförelsen av leden A och B. Detta beror på att leden A och P inte förekommer tillsammans i något block, vilket däremot leden A och B gör (nämligen i block I). Designen är därför lämplig när skillnaden mellan A och B är mer intressant att undersöka än skillnaden mellan A och P. När försöket planeras bör man se till att par av led som är särskilt intressanta att jämföra hamnar i samma block åtminstone någon gång.

Gitterdesignen har tre varianskomponenter:

1. Variation mellan replikaten
2. Variation mellan blocken inom replikaten



Figur 5.2: En gitterdesign med fem replikat (1-5), 20 block (I-XX) och 16 led (A-P).



Figur 5.3: En randomiserad enkel gitterdesign med åtta block (I-VIII) och 16 led (A-P).

## 3. Variation mellan rutorna inom blocken (dvs. residualvarians)

Dessa varianskomponenter skattas i den statistiska analysen. Modellen kan skrivas

$$\text{Led} : \text{Replikat} + \text{Replikat} \cdot \text{Block} \quad (5.1)$$

Lägg märke till att randomiseringen följer försökets varianskomponenter: Replikaten randomiseras, liksom blocken inom replikaten och rutorna inom blocken.

I det första replikatet i figurerna 5.2 och 5.3 jämförs leden A, B, C och D inom samma block. Replikat 1 ger därför god information om skillnaderna mellan leden A, B, C och D. Bara den slumpmässiga variationen mellan rutorna inom blocket skapar en osäkerhet i skattningarna av skillnaderna mellan A, B, C och D. Jämförelsen mellan leden A och M, till exempel, blir inte lika noggrann i replikat 1. Leden A och M ligger nämligen i olika block, och det finns en slumpmässig variation mellan blocken. Jämförelsen mellan leden A och M försvåras därmed inte bara av variationen mellan rutor, utan också av variationen mellan block.

Den relativt goda information replikat 1 ger om skillnaderna mellan leden A, B, C och D utnyttjas i den statistiska analysen för skattningen av variationen mellan blocken i replikat 2. I det replikatet ligger nämligen leden A, B, C och D i olika block. Därmed får vi god kunskap om blockskillnaderna i replikat 2. På samma sätt ger replikat 2 god information om skillnaderna mellan leden A, E, I och M eftersom dessa jämförs i samma block (V). Den informationen tas tillvara vid skattningen av variansen mellan blocken i replikat 1, där leden A, E, I och M ligger i olika block.

Blocken III, IV och V i figur 5.2 är konstruerade från de parvist ortogonala romerska kvadraterna i figur 4.1. Den första romerska kvadraten ger replikat III i figur 5.2, den andra ger replikat IV, och den tredje replikat V. I den första romerska kvadraten ligger led A i diagonalen. Därför utgör leden i diagonalen i replikat 1 första blocket (IX) i replikat 3. Vidare finns i block X de led från replikat 1 som har samma positioner som led B i den första romerska kvadraten, nämligen leden E, B, O och L. På samma sätt konstrueras alla övriga block. Det sista blocket (XX) består till exempel av de led i replikat 1 som finns på samma positioner som led D i den tredje romerska kvadraten. Eftersom det som mest kan finnas  $k - 1$  parvist ortogonala romerska kvadrater med sidan  $k$  finns som mest  $2 + (k - 1) = k + 1$  replikat i en gitterdesign för  $k^2$  led. Om det finns  $k + 1$  replikat är gitterdesignen balanserad. För  $k = 6$  finns inte två romerska kvadrater som är parvist ortogonala. Följaktligen har gitterdesignen för 36 led bara två replikat. En enkel gitterdesign är en gitterdesign bestående av de två första replikaten, så som i figur 5.3.

## 5.2 Rektangulär gitterdesign

Den vanliga kvadratiske gitterdesignen kräver att antalet behandlingar kan skrivas  $k^2$ , där  $k$  är ett heltal. När antalet behandlingar är  $k(k + 1)$  för något heltal

$k$  kan en rektangulära gitterdesign användas.

Vid konstruktionen av en rektangulär gitterdesign utgår man från en genererande matris av storleken  $(k + 1) \times (k + 1)$  med olika ledbeteckningar i alla rutor utom i diagonalen. Figur 5.4 är en genererande matris för en rektangulär gitterdesign med  $3(3 + 1) = 12$  behandlingar.

	A	B	C
D		E	F
G	H		I
J	K	L	

Figur 5.4: Genererande matris för en rektangulär gitterdesign med 12 behandlingar.

Matrisen i figur 5.4 genererar den rektangulära gitterdesign som visas i figur 5.5. Det första replikatet utgörs av den genererande matrisens rader, och det andra replikatet utgörs av den genererande matrisens kolumner. Tredje och fjärde replikatet konstrueras med hjälp av parvisa ortogonala romerska kvadrater med sidan  $k + 1$  som är sådana att de  $k + 1$  leden i huvuddiagonalen alla är olika. Om det finns  $k$  parvist ortogonala matriser är  $k - 1$  sådana att leden i diagonalen är olika. Tredje replikatet i figur 5.5 har konstruerats med hjälp av den andra romerska kvadraten i figur 4.1. Om man lägger den romerska kvadraten ovanpå den genererande matrisen i figur 5.4 hamnar rutorna med leden F, H och L i den genererande matrisen under rutorna med ledet A i den romerska kvadraten. Därför består första blocket (IX) i tredje replikatet av leden F, H och L. Under ledet B i den romerska kvadraten hittar man leden C, D och K. Slutligen, om man lägger den tredje romerska kvadraten från figur 4.1 på den genererande matrisen i figur 5.4 finner man leden A, F och J under ledet D i den romerska kvadraten. Därför utgör dessa led det sista blocket i replikat 4 i figur 5.5.

Om man så önskar kan man lägga ut det rektangulära gitterförsöket på fältet i vanliga avlänga replikat, på samma sätt som i figur 5.3. I planen (figur 5.5) ska

	I	II	III	IV		V	VI	VII	VIII
1	A	D	G	J	2	D	A	B	C
	B	E	H	K		G	H	E	F
	C	F	I	L		J	K	L	I
	IX	X	XI	XII		XIII	XIV	XV	XVI
3	F	C	A	B	4	E	B	C	A
	H	D	E	I		I	D	G	F
	L	K	G	J		K	H	L	J

Figur 5.5: En rektangulär gitterdesign med fyra replikat (1–4) och 16 block (I–XVI) för tolv behandlingar (A–L).

rutorna randomiseras inom blocken, och blocken randomiseras inom replikaten. Dessutom ska replikaten randomiseras till platserna på fältet. I den statistiska analysen skattas varianserna mellan replikat, mellan block inom replikat och mellan rutor inom block enligt modellen (5.1).

### 5.3 Alfadesign

Ofta är antalet behandlingar  $v$  så stort att ett randomiserat fullständigt blockförsök är olämpligt. Detta är särskilt vanligt i sortförsök. När  $v$  varken kan skrivas som  $k^2$  eller som  $k(k+1)$ , där  $k$  är ett heltal, finns ingen kvadratisk eller rektangulär gitterdesign. I de fallen bör man använda en så kallad alfadesign.

Alfadesignen bestäms av en genererande matris. Denna matris är designens så kallade alfamatrix. Ett exempel på en alfamatrix visas i figur 5.6. Alfamatrixens första kolumn bestämmer designens första replikat, alfadesignens andra kolumn bestämmer designens andra replikat, och så vidare. Alfamatrixen i figur 6.6 har tre kolumner och genererar således en design med tre replikat. Antalet rader i alfadesignen bestämmer antalet rutor per block. Alfamatrixen i figur 6.6 har fyra rader, och därför ger den en design med fyra rutor per block. Elementen i alfamatrixen är  $0, 1, 2, \dots, s-1$ , där  $s$  är antalet block per replikat i alfadesignen. Eftersom alfamatrixen i figur 5.6 innehåller talen 0, 1 och 2 består alfadesignen av tre block per replikat ( $s = 3$ ).



0	0	0
0	1	2
0	2	1
0	0	1

Figur 5.6: Genererande matris för alfadesignen i figur 6.8.

Alfadesignen skapas enklast från alfamatriken i två steg. Först bildar man mellanmatriserna, och från dessa bildar man den slutliga designen. Alfamatriken i figur 5.6 ger de mellanmatriser som visas i figur 5.7. Den första kolumnen i alfamatriken ger den första mellanmatrisen, den andra kolumnen ger den andra mellanmatrisen och så vidare. Mellanmatrisernas första kolumner är kolumnerna i alfamatriken. Mellanmatrisernas andra kolumner får man genom att addera 1 till den första kolumnen. Om emellertid summan blir  $s$ , skriver man 0 istället. Mellanmatrisernas element är därför  $0, 1, 2, \dots, s - 1$ , precis som i alfamatriken.

Man får slutligen alfadesignen genom att addera  $s$  till mellanmatrisernas andra rader uppifrån,  $2s$  till mellanmatrisernas tredje rader,  $3s$  till mellanmatrisernas tredje rader, och så vidare. Alfamatriken i figur 5.6 ger alltså mellanmatriserna i figur 5.7 som ger alfadesignen i figur 5.8. I detta exempel består alfadesignen av tre fullständiga replikat med tolv behandlingar. Designen omfattar  $s = 3$  block per replikat och fyra rutor per block.

Precis som i gitterdesignen ska replikaten randomiseras till försöksytorna på fältet. Blocken ska randomiseras inom replikaten och rutorna inom blocken. I den statistiska analysen skattas varianserna mellan replikat, mellan block inom replikat och mellan rutor inom block, enligt modell (5.1).

I designen i figur 5.8 förekommer somliga par av behandlingar tillsammans i samma block noll gånger (t.ex. leden 0 och 1), andra par förekommer tillsammans i block en gång (t.ex. leden 2 och 5 som är tillsammans endast i block III). En del par av behandlingar förekommer tillsammans i block två gånger (t.ex. leden 7 och 10 som är tillsammans i blocken II och VII). Inget par förekommer

0	1	2
0	1	2
0	1	2
0	1	2

0	1	2
1	2	0
2	0	1
0	1	2

0	1	2
2	0	1
1	2	0
1	2	0

Figur 5.7: Mellanmatriser för alfadesignen i figur 6.8.

1			2			3		
I	II	III	IV	V	VI	VII	VIII	IX
0	1	2	0	1	2	0	1	2
3	4	5	4	5	3	5	3	4
6	7	8	8	6	7	7	8	6
9	10	11	9	10	11	10	11	9

Figur 5.8: En alfadesign med tre replikat (1, 2 och 3) och nio block (I–IX) för tolv behandlingar (0–11).

tillsammans i block fler än två gånger. Designen i figur 6.8 sägs därför vara en  $\alpha(0,1,2)$ -design. Siffrorna inom parentes anger hur ofta par av behandlingar förekommer tillsammans i block. Det är som regel inte optimalt att somliga par (t.ex. 7 och 10) jämförs två gånger inom samma block medan andra par (t.ex. 0 och 1) aldrig jämförs inom samma block. Om man är lika intresserad av samtliga parvisa jämförelser bör så många par som möjligt jämföras inom block. Därför föredrar man oftast  $\alpha(0,1)$ -designer framför  $\alpha(0,1,2)$ -designer. Av särskilt intresse är alfadesigner som är så effektiva som möjligt i den meningen att de ger så liten genomsnittlig varians i de parvisa jämförelserna som möjligt.

## 5.4 Val av blockdesign

Fullständiga block är att föredra så länge blocken kan hållas så små att inte variationen inom blocken blir ett problem. Det går inte att generellt säga var gränsen går, men ofta är nog försök med ofullständiga block effektivare än försök med fullständiga block vid fler än ca 12 eller 15 behandlingar. Det är också möjligt att analysera ofullständiga blockförsök som fullständiga om det visar sig att variationen mellan blocken inom replikaten är försumbar. Man förlorar därför inget på att lägga upp försöket med ofullständiga block inom replikaten så länge replikaten är fullständiga, dvs. så länge varje replikat innehåller samtliga behandlingar.

Gitterdesignen är den mest effektiva designen för försök med ofullständiga block, så om antalet led är en kvadrat (t.ex. 16, 25, 36, 49, 64 eller 81) bör den designen väljas. I andra hand väljer man en rektangulär gitterdesign. Om inte heller någon sådan finns för det antal behandlingar man vill undersöka bör man välja en alfadesign, och i så fall helst en  $\alpha(0,1)$ -design. Det finns  $\alpha(0,1)$ -designer för två, tre och fyra replikat, när antalet rutor per block är mindre än eller lika med antalet block per replikat, och antalet behandlingar är mindre än eller lika med 100. Undantaget är fallet med sex rutor per block, sex block per replikat och fyra replikat. För just den situationen finns ingen  $\alpha(0,1)$ -design, men däremot en  $\alpha(0,1,2)$ -design. Den praktiska slutsatsen är att antalet rutor per block inte bör vara större än antalet block per replikat. För ett givet val av antal behandlingar, antal replikat, antal block per replikat och antal rutor per block gäller det att hitta den effektivaste designen. I en effektiv design är den genomsnittliga variansen i de parvisa jämförelserna så liten som möjligt. Patterson, Williams och Hunter (1978) tillhandahåller en tabell med ett flertal alfamatriser för effektiva alfadesigner. Alternativt tar man hjälp av ett för syftet utvecklat datorprogram för att hitta en så effektiv alfadesign som möjligt.



## Kapitel 6

# Fältförsök med flera faktorer

### 6.1 Fördelar med flerfaktoriella försök

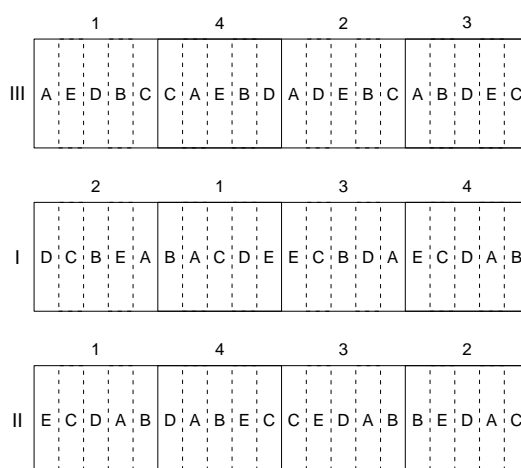
När man vill undersöka två eller flera faktorer är det generellt bättre att jämföra dem i samma försök än att göra ett försök per faktor.

*Example 6.1* Antag att vi vill undersöka sex sorters höstvetete och fyra olika kvävegivor. Först skulle vi kunna jämföra de sex vetesorterna i ett fältförsök med en av kvävegivorna. Om vi gör det i fyra block blir, enligt (1.7), variansen i de parvisa jämförelserna av sorterna  $2\sigma^2/4$ , där  $\sigma^2$  är variansen mellan rutorna i försöket. Detta försök kräver 24 rutor. Sedan skulle vi kunna jämföra de fyra kvävegivorna för en av sorterna. Om vi gör det i sex block blir, enligt samma formel, variansen i skillnaden mellan två kvävegivor  $2\sigma^2/6$ . Även detta försök kräver 24 rutor. Med samma försöksomfattning, 24 rutor + 24 rutor = 48 rutor, skulle vi kunna undersöka samtliga 24 kombinationer av sorter och kvävegivor i ett enda försök med två replikat. Variansen i skillnaden mellan två sorter, beräknat över samtliga kvävegivor, blir  $2\sigma^2/8$ , eftersom det finns åtta observationer av varje sort. På samma sätt blir variansen i skillnaden mellan två kvävegivor, beräknat över samtliga sorter,  $2\sigma^2/12$ , eftersom det finns 12 observationer av varje kvävegiva. Lägg märke till att båda varianserna är hälften så stora i det tvåfaktoriella försöket som i försöken med en faktor. Genom att undersöka båda faktorerna blir det också möjligt att se om skillnaderna mellan sorterna är olika för olika kvävegivor. När så är fallet säger man att det finns ett samspel mellan de båda faktorerna, dvs. mellan sorter och kvävegivor.

I ett flerfaktoriellt försök utgör kombinationerna av faktorernas nivåer försökets behandlingar. Dessa behandlingar kan randomiseras till försökets rutor fullständigt slumpmässigt. Alternativt delas fältet in i block, enligt avsnitt 3.1, varefter behandlingarna randomiseras inom blocken. Det är vanligen bra att använda block, för variationen är oftast mindre inom blocken än över hela försöksytan.

## 6.2 Split-plot-försök

Ofta kräver den ena faktorn större rutor än den andra. Man kan då dela in fältet i stora rutor som i sin tur delas in i små rutor. I ett split-plot-försök görs denna indelning, varefter den ena faktorn randomiseras till de stora rutorna och den andra faktorn till de små rutorna. Randomiseringarna görs inom block. Ett split-plot-försök kan se ut som i figur 6.1.



Figur 6.1: Ett randomiserat split-plot-försök med tre block (I, II, III), fyra stor-rutor per block och fem smårutor per storruta. Ena faktorn (F1) har fyra led (1–4), vilka randomiserats till de stora rutorna inom replikaten. Andra faktorn (F2) har fem led (A–E), vilka randomiserats till de små rutorna inom de stora rutorna.

Split-plot-försöket har tre varianskomponenter, för det kan finnas variation mellan blocken, mellan storrutorna inom blocken, och mellan de små rutorna inom de stora rutorna. Den statistiska modellen kan skrivas

$$F1 + F2 + F1 \cdot F2 : \text{Block} + \text{Block} \cdot \text{Storruta} , \quad (6.1)$$

där F1 betecknar faktorn som randomiserats till de stora rutorna, och F2 betecknar faktorn som randomiserats till de små rutorna. Ofta saknar datamängden en variabel som anger de stora rutorna. Istället för att skapa en ny sådan variabel kan man definiera storrutorna som samspelet mellan faktorerna Block och F1. Varje kombination av dessa båda variabler utgör ju en storruta. Modellen (6.1) är därför, i split-plot-försöket, samma som modellen

$$F1 + F2 + F1 \cdot F2 : \text{Block} + \text{Block} \cdot F1 . \quad (6.2)$$

Det totala antalet frihetsgrader delas upp på de tre varianskomponenterna. Frihetsgradsuppdelningen för försöket i figur 6.1 visas i tabell 6.1. Det totala antalet frihetsgrader är som vanligt lika med antalet observationer minus 1. Eftersom det finns tre block åtgår två frihetsgrader till skattningen av blockeffekterna. Vidare finns det fyra storrutor per block, så det går åt  $3(4 - 1) = 9$  frihetsgrader till storrutorna. Kvar finns  $59 - 2 - 9 = 48$  frihetsgrader till smårutorna. Faktorn F1 har fyra nivåer (1-4), så det krävs  $4 - 1 = 3$  frihetsgrader för att skatta dess effekter. Den har randomiserats till storrutor, så skillnaderna mellan dess nivåer ska jämföras mot storrutornas varians. Som beräkningen i tabell 6.1 visar skattas denna varians med  $9 - 3 = 6$  frihetsgrader. Faktorn F2 har fem nivåer (A-E), så den kräver  $5 - 1$  frihetsgrader. Antalet frihetsgrader för samspelet mellan F1 och F2 är lika med produkten av faktorernas frihetsgrader, dvs.  $(4 - 1)(5 - 1) = 12$ . Dessa faktorer jämförs mot variansen mellan de smårutorna. Denna varians skattas med 32 frihetsgrader, enligt beräkningen i tabell 6.1.

Tabell 6.1: Försöket i figur 6.1 omfattar tre varianskomponenter. Faktor F1 jämförs på storrutor. Variansen mellan storrutorna skattas med 6 frihetsgrader. Faktorn F2 och samspelet mellan F1 och F2 undersöks på smårutor. Variansen mellan smårutorna skattas med 32 frihetsgrader

Varianskomponent	Variationsorsak	Frihetsgrader
Block		2
Storrutor	F1	3
	Residual	6
	Totalt	9
Smårutor	F2	4
	F1 · F2	12
	Residual	32
	Totalt	48
Totalt		59

Split-plot-försökets största nackdel är att variationen mellan de stora rutorna ofta blir skattad med stor osäkerhet, helt enkelt på grund av att antalet stora rutor i försöket är litet. Försöket i figur 6.1 omfattar fyra storrutor per block, vilket enligt beräkningen i tabell 6.1, när det finns tre block, ger sex frihetsgrader till skattningen av variationen mellan de stora rutorna. Som vi såg i kapitel 3, brukar man vilja ha minst åtta eller tio frihetsgrader för skattningen av variansen. Om antalet frihetsgrader är litet blir det nämligen svårt att upptäcka skillnaderna mellan behandlingarna. Det är alltså vanligt att man i den statistiska analysen inte kan se några signifikanta skillnader mellan de behandlingar som randomiserats till de stora rutorna.

Bättre blir jämförelserna för den andra faktorn, som ju randomiserats till de små rutorna. Skillnaderna mellan den faktorns nivåer jämförs mot variansen

mellan de små rutorna, vilken vanligen blir väl skattad. Styrkan i testet av smårutefaktorn blir därmed bra, liksom styrkan i testet av samspelet mellan de båda faktorerna. Dessa test har bättre styrka i split-plot-försök än i randomiserade fullständiga blockförsök av samma storlek. Styrkan i testet av den faktor som undersöks på de stora rutorna är däremot sämre än i motsvarande randomiserade fullständiga blockförsök.

Det finns fem olika slags parvisa jämförelser i ett split-plot-försök. Man kan vilja jämföra

1. Ett led som ligger på storrutor jämfört med ett annat led som ligger på storrutor. Denna jämförelse kan göras
  - a beräknat genomsnittligt över samtliga led som ligger på smårutor, eller
  - b för något specifikt led som ligger på smårutor.
2. Ett led som ligger på smårutor jämfört med ett annat led som ligger på smårutor. Denna jämförelse kan göras
  - a beräknat genomsnittligt över samtliga led som ligger på storrutor, eller
  - b för något specifikt led som ligger på storrutor.
3. Två led som är helt olika kombinationer av de två faktorerna.

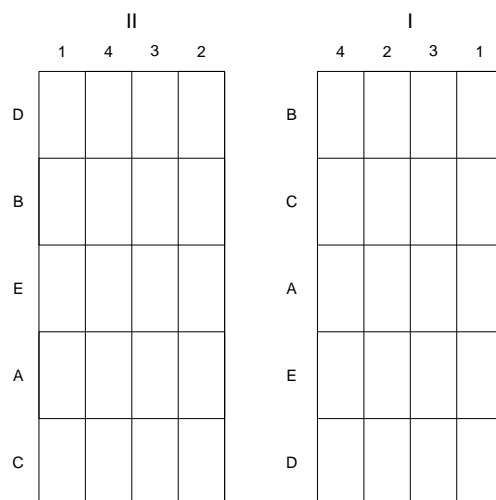
Jämförelserna 2a och 2b får bättre precision i ett split-plot-försök än i ett randomiserat fullständigt blockförsök, men övriga jämförelser får sämre precision. Exakta statistiska test saknas för jämförelserna 1b och 3, så man måste avstå från att göra dessa jämförelser eller använda approximativa test. Randomiserade fullständiga försök och randomiserade fullständiga blockförsök är enklare och har inte detta problem.

Det är naturligtvis möjligt att dela in smårutorna i ännu mindre rutor och randomisera en tredje faktor till dessa minsta rutor. Då får man ett så kallat split-split-plot-försök. Ett sådant försök delar split-plot-försökets styrkor och svagheter.

### 6.3 Split-block-försök

I split-block-försök delas varje block in i rader och kolumner. I varje block randomiseras en faktors nivåer till raderna och en annan faktors nivåer till kolumnerna. Split-block-försök kallas även criss-cross-försök eller strip-plot-försök. Benämningen criss-cross-försök är särskilt vanlig när försöket inte är indelat i block. I så fall randomiseras upprepningarna av de båda faktorerna fullständigt på de tillgängliga raderna respektive kolumnerna. I Figur 6.2 visar hur ett split-block-försök kan se ut. Den första rutan i block II har till exempel behandlats med kombinationen D1, den andra rutan i samma rad med kombinationen D4, och så vidare.





Figur 6.2: Ett randomiserat split-block-försök med två block (I och II). Ena faktorn (F1) har fyra nivåer (1–4), randomiserade till kolumner inom block. Andra faktorn (F2) har fem nivåer (A–E), randomiserade till rader inom block.

Split-block-försöket analyseras med modellen

$$F1 + F2 + F1 \cdot F2 : \text{Block} + \text{Block} \cdot \text{Rad} + \text{Block} \cdot \text{Kolumn} .$$

Tabell 6.2 innehåller frihetsgradsanalysen för split-block-försöket i figur 7.2. Försöket omfattar sammanlagt 40 rutor, så totala antalet frihetsgrader är  $40 - 1 = 39$ . Det finns två block, så det åtgår  $2 - 1 = 1$  frihetsgrad till skattningen av skillnaden mellan blocken. Skillnaderna mellan kolumnerna skattas med  $4 - 1 = 3$  frihetsgrader i varje block, eftersom det finns fyra kolumner per block. Det blir  $2(4 - 1) = 6$  frihetsgrader totalt. Vidare finns fem rader per block, så skillnaderna mellan raderna skattas med  $2(5 - 1) = 8$  frihetsgrader, eftersom försöket omfattar 2 block. Ena faktorn (F1) jämförs med  $4 - 1 = 3$  frihetsgrader på kolumnerna, eftersom den faktorn har fyra nivåer och är randomiserad till kolumner. Andra faktorn (F2) jämförs med  $5 - 1 = 4$  frihetsgrader på raderna, eftersom den faktorn har fem nivåer och är randomiserad till rader. Samspelet mellan de båda faktorerna, dvs. effekterna av kombinationerna av de båda faktorerna, undersöks på rutorna. Antalet frihetsgrader är lika med produkten av de båda faktorernas frihetsgrader. Denna produkt är  $(4 - 1)(5 - 1) = 12$ . Variansen mellan block skattas med en frihetsgrad. Frihetsgraderna för övriga varianser beräknas per subtraktion. Variansen mellan kolumner skattas med tre frihetsgrader, variansen mellan rader med fyra frihetsgrader och variansen mellan rutor med tolv frihetsgrader.

Enligt beräkningen i tabell 6.2 skattas varianserna mellan kolumner och rad-

Tabell 6.2: Försöket i figur 6.2 omfattar fyra varianskomponenter. Faktorn F1 jämförs på kolumner. Variansen mellan kolumnerna skattas med 3 frihetsgrader. Faktorn F2 jämförs på rader. Variansen mellan raderna skattas med 4 frihetsgrader. Samspelet mellan F1 och F2 jämförs på rutor. Variansen mellan rutor skattas med 12 frihetsgrader

Varianskomponent	Variationsorsak	Frihetsgrader
Block		1
Kolumner	F1	3
	Residual	3
	Totalt	6
Rader	F2	4
	Residual	4
	Totalt	8
Rutor	F1 · F2	12
	Residual	12
	Totalt	24
Totalt		39

er i försöket i figur 6.2 med få frihetsgrader, närmare bestämt med tre respektive fyra stycken. Med så få frihetsgrader blir osäkerheten i skattningarna av varianserna stor. Följaktligen riskerar skattningarna av behandlingsskillnaderna att bli dåliga, och det blir svårt att få signifikanta resultat i testen av behandlingarna. Samspelet testas dock mot tolv frihetsgrader, vilket är tillfredsställande.

Följande fem slag av parvisa jämförelser kan vara aktuella i ett split-block-försök:

1. Ett led som ligger på rader jämfört med ett annat led som ligger på rader. Denna jämförelse kan göras
  - a beräknat genomsnittligt över samtliga led som ligger på kolumner, eller
  - b för något specifikt led som ligger på kolumner.
2. Ett led som ligger på kolumner jämfört med ett annat led som ligger på kolumner. Denna jämförelse kan göras
  - a beräknat genomsnittligt över samtliga led som ligger på rader, eller
  - b för något specifikt led som ligger på rader.
3. Ett led som ligger på rader i kombination med ett led som ligger på kolumner jämfört med ett annat led som ligger på rader i kombination med ett annat led som ligger på kolumner.

Exakta statistiska test saknas för jämförelserna 1b, 2b och 3. Antingen avstår man från att testa dessa jämförelser statistiskt, eller använder man ett approximativt test.

## Kapitel 7

# Försöksserier

Med en försöksserie menar vi en mängd försök som ska analyseras tillsammans. Det är som regel mycket lämpligt att lägga upp försöken som ska analyseras tillsammans på liknande sätt. Helst bör alla försök innehålla samma behandlingar och utföras med samma design. Diskussionen i det här kapitlet kräver dock inte detta. I detta kapitel föreslås modeller för statistisk analys av olika slags försöksserier bestående av ettåriga försök med en enda behandlingsfaktor. Serierna behöver inte vara strikt balanserade för att modellerna ska fungera, men de ingående leden måste i stor utsträckning förekomma i många försök.

Försöken kan ha utförts på en eller flera lokaler. Om försöken bara kommer från en enda lokal kan de förstas bara ge information om just den lokalen. Om de däremot har utförts på olika lokaler kan försöksserien ge information om den region lokalerna tillhör, nämligen om lokalerna kan betraktas som representativa för regionen i den meningen att de skulle kunna ha varit slumpmässigt valda från en större population av tänkbara lokaler i regionen. Utgångspunkten är därför att effekterna av lokalerna kan betraktas som slumpmässiga (kapitel 4).

Likaså kan försöken ha utförts under ett och samma år eller under olika år. Årseffekterna kan modelleras som slumpmässiga (kapitel 4). Om försöken kommer från ett och samma år kan de naturligtvis inte ge någon information om några andra år än just det året. De statistiska slutsatserna måste begränsa sig till det enda året. Om däremot försöken har utförts under olika år blir syftet med den statistiska analysen annorlunda. I så fall syftar analysen till att undersöka långsiktiga skillnader mellan behandlingarna.

Det är viktigt att undersöka försöksseriens samspel. När serien omfattar flera lokaler kan det hända att behandlingen samspelar med lokalerna, så att skillnaderna mellan behandlingarna är olika på olika lokaler. Likaså, när serien omfattar flera år kan det hända att behandlingsskillnaderna varierar slumpmässigt mellan åren. Något år gynnas någon särskild behandling, ett annat år gynnas en annan. Detta yttrar sig som ett slumpmässigt samspel mellan år och behandlingar. Om samspelet mellan år och behandlingar är stort krävs det många år för att kunna urskilja de systematiska, dvs. långsiktiga, skillnaderna mellan behandlingarna. När det finns en systematisk skillnad mellan två behandlingar

ger den ena behandlingen genomsnittligt, beräknat över en längre period, högre värden än den andra. Enskilda år kan dock den andra behandlingen ge högre värden än den första, särskilt om samspelet är stort.

## 7.1 Statistiska modeller för försöksserier

Det är ofta praktiskt att analysera försöksserier i två steg, särskilt om försöken är av olika slag, dvs. om de har olika design. Först analyseras de enskilda försöken. I analysen justeras medelvärdena för eventuell obalans som orsakats till exempel av att blocken varit ofullständiga. De justerade medelvärdena kallas minsta kvadratmedelvärden. Den statistiska analysen av försöksserien görs sedan på dessa minsta kvadratmedelvärden.

I det följande definieras fem olika statistiska modeller användbara för 13 olika typer av försöksserier. Led modelleras genomgående som en fix faktor. Som vi redan nämnt i kapitel 4 kan det i stora växtförädlingsförsök vara lämpligare att anta att behandlingseffekterna är normalfördelade, dvs. att modellera även led som en slumpmässig faktor.

Den enklaste modellen för statistisk analys av försöksserier kan skrivas

$$\text{Led} : \text{Försök} \quad (7.1)$$

Modell (7.1) är användbar i följande situationer:

- i. Det finns försök från ett enda år och en enda lokal.
- ii. Det finns försök från flera år men bara från en enda lokal, och det finns bara ett försök per år.
- iii. Det finns försök från flera lokaler men bara från ett enda år, och det finns bara ett försök per lokal
- iv. Det finns försök från många år och lokaler, men bara ett försök per år och ett försök per lokal.

När försöksserien omfattar många år är följande modell ofta lämplig:

$$\text{Led} : \text{År} + \text{Led} \cdot \text{År} + \text{Försök} \quad (7.2)$$

Närmare bestämt är modell (7.2) användbar i följande situationer:

- v. Det finns försök från flera år och flera försök per år. Försöken är från en enda lokal.
- vi. Det finns försök från flera år och från flera lokaler. Det finns bara ett försök per lokal. Det finns flera försök per år.

- vii. Det finns försök från flera lokaler och flera år. Det finns flera försök per lokal och år. Alla försök från en specifik lokal är från samma år. Alla försök från ett specifikt år är från samma lokal.

Termen  $\text{Led} \cdot \text{År}$  i modell (7.2) anger att skillnaderna mellan leden varierar slumpmässigt från år till år. Detta är ofta ett mycket rimligt antagande. Effekterna av leden testas mot variansen i samspelet  $\text{Led} \cdot \text{År}$ . Om samspelet är stort är det svårt att se de långsiktiga systematiska skillnaderna mellan leden, särskilt om dessa är små eller om serien bara omfattar ett fåtal år.

Modell (7.3) är samma som modell (7.2), men med  $\text{År}$  utbytt mot  $\text{Lokal}$ .

$$\text{Led} : \text{Lokal} + \text{Led} \cdot \text{Lokal} + \text{Försök} \quad (7.3)$$

Modell (7.3) är användbar i följande situationer:

- viii. Det finns försök från flera lokaler och flera försök per lokal. Försöken är från ett enda år.
- ix. Det finns försök från flera lokaler och från flera år. Det finns bara ett försök per år. Det finns flera försök per lokal.
- vii. Det finns försök från flera lokaler och flera år. Det finns flera försök per lokal och år. Alla försök från en specifik lokal är från samma år. Alla försök från ett specifikt år är från samma lokal.

I situation vii är det omöjligt att skilja på effekter av år och effekter av lokaler. Då ger modellerna (7.2) och (7.3) samma resultat.

När serien omfattar såväl flera år som flera lokaler är modellerna (7.4) och (7.5) lämpliga.

$$\text{Led} : \text{År} + \text{Lokal} + \text{Led} \cdot \text{År} + \text{Led} \cdot \text{Lokal} + \text{Försök} \quad (7.4)$$

Modell (7.4) är användbar i följande situationer:

- x. Det finns försök från flera år och flera lokaler. Det finns flera lokaler per år och flera försök per lokal. Alla försök från en specifik lokal är från samma år.
- xi. Det finns försök från flera år och flera lokaler. Det finns flera år per lokal och flera försök per år. Alla försök från ett specifikt år är från samma lokal.
- xii. Det finns försök från flera år och flera lokaler. Det finns flera försök per lokal och per år, men bara ett försök per år och lokal.

Modell (7.5) innehåller dessutom samspelet mellan år och lokaler:

$$\text{Led} : \text{År} + \text{Lokal} + \text{Led} \cdot \text{År} + \text{Led} \cdot \text{Lokal} + \text{År} \cdot \text{Lokal} + \text{Försök} \quad (7.5)$$

Modell (7.5) är används när det finns flera försök per år och lokal, dvs. när samspelet mellan år och lokaler kan skattas:

- xiii. Det finns försök från flera år och flera lokaler. Det finns flera försök per lokal och per år. Det finns flera försök per år och lokal.

## 7.2 Dimensionering av enkla försöksserier

I enkla försöksserier kan modell (7.1) användas för statistisk analys, närmare bestämt i situationerna i–iv ovan. När man planerar en sådan försöksserie kan man, givet kunskap om hur stor den slumpmässiga variationen är, välja antalet försök så att minsta signifikanta skillnaden (LSD) i försöket sannolikt blir tillräckligt liten. I praktiken kan man förstås inte veta hur stor spridningen kommer att bli i serien, men man kan ofta göra en uppskattning baserad på tidigare erfarenheter. Givet en uppskattning av hur stor variationskoefficienten (CV) kommer att bli i analysen med modell (7.1) kan ett förväntat LSD beräknas så här:

$$\text{LSD}(\%) = t \sqrt{\frac{2 \text{CV}(\%)^2}{n}}. \quad (7.6)$$

I ekvation (7.6) står  $n$  för antalet försök och  $t$  för t-fördelningens 97,5:te percentil vid  $(n-1)(v-1)$  frihetsgrader, där  $v$  är antalet led i försöket. Om antalet försök eller antalet led är stort är  $t \approx 2$ . Lägg märke till att ekvation (7.6) är samma som ekvation (3.2). I analysen av den enkla försöksserien fungerar försöken som block. Observationerna är medelvärden från de enskilda försöken.

*Exempel 7.1.* En försöksserie planeras. Ett stort antal behandlingar ska jämföras i  $n$  försök. Frågan är hur många försök som måste utföras, dvs. hur stort  $n$  måste vara. Från analyser av tidigare försöksserier av liknande slag vet vi att variationskoefficienten (dvs. roten ur residualvariansen dividerad med medelvärdet), brukar vara lägre än 5 %. Låt säga att vi önskar att skillnader mellan två led ska vara signifikanta på 5 % -nivån om de skattas till 5 % eller mer, givet att variationskoefficienten är 5 %. Eftersom antalet led är stort kan vi räkna med  $t = 2$  i ekvation (7.6), och eftersom

$$2 \sqrt{\frac{2 \cdot (5\%)^2}{8}} = 5\%$$

måste serien, enligt ekvation (7.6), omfatta minst  $n = 8$  försök.

I försöksserier som sträcker sig över flera år är modell (7.2) ofta användbar. I sådana serier har inte bara residualvariansen betydelse för säkerheten i skattningarna mellan leden, utan också variansen för samspelet mellan år och

led. För att kunna bestämma hur många år och hur många försök per år serien måste omfatta för att tillgodose något krav på precisionen i resultaten är det nödvändigt att ha en uppfattning om hur stora dessa båda varianser är. Förhoppningsvis har försök av liknande slag utförts någon gång tidigare så att det är möjligt att skatta varianserna. Om variansen för samspelet mellan år och led skattas till  $s_A^2$  i försöksserien, och variansen för residualspridningen skattas till  $s_E^2$ , blir medelfelet för skillnaden mellan två behandlingar

$$\sqrt{2\left(\frac{s_A^2}{a} + \frac{s_E^2}{an}\right)}, \quad (7.7)$$

där  $4a$  är antalet år och  $n$  antalet försök per år. Ofta är det lättare att göra ungefärliga beräkningar i procentenheter. Som följd av (7.7) är minsta signifikanta skillnaden, uttryckt i procent av totala medelvärdet,

$$\text{LSD}(\%) = t\sqrt{2\left(\frac{\text{CV}_A(\%)^2}{a} + \frac{\text{CV}_E(\%)^2}{an}\right)}. \quad (7.8)$$

där  $\text{CV}_A$  är variationskoefficienten för samspelet mellan år och led, och  $\text{CV}_E$  är variationskoefficienten för residualerna. Dessa erhålls från standardavvikelserna  $s_A$  respektive  $s_E$  genom division med totala medelvärdet. I ekvation (7.8) betecknar  $t$  den 97,5:te percentilen i en t-fördelning med  $(a-1)(v-1)$  frihetsgrader, där  $v$  är antalet led. Om antalet frihetsgrader är stort är  $t \approx 2$ .

*Exempel 7.2.* En försöksserie omfattande flera år ska planeras. Man måste räkna med att skillnaderna mellan leden varierar mellan åren. Serien syftar till att ta reda på om det finns systematiska, långsiktiga, skillnader mellan leden. Med andra ord vill man kunna urskilja de långsiktiga skillnaderna mellan leden från de årliga slumpmässiga skillnaderna mellan leden. Man planerar göra  $n = 3$  försök per år och önskar att de skattade skillnaderna mellan leden, beräknade genomsnittligt över åren, ska vara signifikanta om de är större än 5 %. Frågan är hur många år som behövs för att klara detta krav. För att kunna besvara den frågan krävs kunskap om hur stor variationen är mellan och inom år. Närmare bestämt måste man veta hur mycket skillnaderna mellan leden varierar slumpmässigt mellan åren och hur stor residualspridningen är. I en tidigare studie skattades variansen för samspelet mellan år och led till 36 000, och residualvariansen till 62 500. Medelvärdet var 6 400. Räknat i procent av medelvärdet är standardavvikelserna 3,0 % respektive 3,9 %. Eftersom försöket innehåller många led kan vi räkna med  $t = 2$ . Med dessa siffror kan ekvation (7.8) skrivas

$$\text{LSD}(\%) = 2\sqrt{2\left(\frac{(3,9\%)^2}{a} + \frac{(3,0\%)^2}{a \cdot 3}\right)}$$

Kravet  $\text{LSD} < 5\%$  är uppfyllt om  $a \geq 5$ , men inte om  $a \leq 4$ , så serien bör omfatta minst fem år.

I verkligheten finns det ingen garanti att variansskattningarna blir desamma i en ny serie som i en tidigare. Därför är det vid planeringen av försöksserier klokt att räkna med lite större varianser än man tidigare upplevt. Det är också möjligt att dimensionera serier utifrån krav på styrka i signifikanstesten. Kapitel 3 visade hur detta går till för randomiserade blockförsök, och samma metod kan användas i enkla serier.

Ekvation (7.7) kan användas för att avgöra huruvida det är bäst att göra flera försök med ett fåtal replikat eller få försök med ett flertal replikat. Formeln avser visserligen medelfelet för en skillnad mellan två led skattade i modell (7.2). Men om man i modell (7.2) byter ut  $\bar{A}$  mot Försök och Försök mot Replikat avser  $a$  i formel (7.7) antalet försök och  $n$  antalet replikat.

*Exempel 7.3.* Vilket av följande två alternativ ger den säkraste skattningen av effekten av fungicidbehandlingen, i de enskilda sortförsöken respektive i seriesammanställningarna:

1. Tio försök med fyra storrutor varav två behandlas mot svamp, eller
2. Fem försök med åtta storrutor varav fyra behandlas mot svamp?

*Svar:* I de enskilda försöken är det alltid bäst med så många replikat som möjligt. Behandlingseffekten i ett försök skattas bättre med fyra replikat än med två, dvs. bättre med alternativ 2 än med alternativ 1. I en försöksserie kan behandlingseffekten variera mellan försöken, så att  $s_A^2 > 0$  i ekvation (7.7). I så fall är det bäst att göra så många försök som möjligt, förutsatt att totala antalet replikat,  $a \cdot n$ , är konstant. Alternativ 1 är alltså bäst om behandlingseffekten varierar mellan försöken, vilket den väl troligen gör. Men om det mot förmodan inte finns någon varians i behandlingseffekten mellan försöken, dvs. om  $s_A^2 = 0$  i ekvation (7.7), spelar det inte någon roll om man gör många försök med ett fåtal replikat eller få försök med ett flertal replikat. I så fall är det bara det totala antalet replikat,  $a \cdot n$ , som har betydelse, och de två alternativen är lika bra.



# Referenser

- Engstrand, U., Olsson, U. (2003). *Variansanalys och försöksplanering* Lund: Studentlitteratur.
- Galwey, N. W. (2006). *Introduction to mixed modelling: Beyond regression and analysis of variance*. Chichester: Wiley.
- Kupper, L. L., Hafner, K. B. (1989). How appropriate are popular sample size formulas? *The American Statistician* 43, 101–105.
- Patterson, H. D., Williams, E. R., Hunter, E. A. (1975). Block designs for variety trials. *Journal of Agricultural Science, Cambridge* 90, 395–400.
- Piepho, H.-P., Büchse, A., Emrich, K. (2003). A hitchhiker's guide to mixed models for randomized experiments. *Journal of Agronomy & Crop Science* 189, 310–322.
- Seeger, P. *Handbok för jordbruksförsök: statistisk planering och analys*. Speciala skrifter 18, Uppsala: Sveriges lantbruksuniversitet.
- Smith, A., Cullis, B., Gilmour, A. (2001). The analysis of crop variety evaluation data in Australia. *Australian & New Zealand Journal of Statistics* 43, 129–145.

# Sakregister

- alfadesign, 48
- balans, 43
- bestämmande försök, 7
- blandade modeller, 35, 38
- block, 7
- criss-cross-försök, 56
- CV, 12, 24
- dimensionering, 27, 62
- effektivitet, 51
- ekvivalensstudier, 15
- experimentenheter, 7
- F-fördelning, 24
- F-test, 24
- faktorer, 10
- fix faktor, 35
- flerfaktoriella försök, 53
- frihetsgrader, 10
- fullständigt randomiserad design, 29
- förklarande variabler, 10
- försöksserie, 35
- försöksserier, 59
- gitterdesign, 43, 46
- gitterförsök, 36, 43
- hypotesprövning, 14
- jämförande försök, 7, 28
- konfidensintervall, 25
- kvadratsumma, 24
- lattice design, 43
- lokaler, 59
- LSD, 25, 28
- maximum likelihood-metoden, 37
- medelfel, 13
- medelvärde, 11
- mixed models, 35
- ML, 37
- MSE, 24, 35
- nivåer, 10
- normalfördelning, 14
- observationsenheter, 9
- ofullständiga blockförsök, 36, 43
- parameter, 10
- primär variabel, 15
- pseudoreplikat, 10
- rad-kolumn-försök, 35, 36, 39, 56
- randomiserat fullständigt blockförsök, 19
- randomisering, 11
- rektangulär gitterdesign, 46
- REML, 37, 39
- replikat, 7
- residual, 23
- residualspridning, 24
- responsvariabler, 10
- romerska kvadrater, 39
- samspel, 25, 38, 53, 59
- sannolikhetsvärde, 16
- signifikansnivå, 16
- slumpmässig faktor, 35
- slumpmässiga block, 38
- split-block-försök, 56

split-plot-försök, 35, 36, 54  
split-split-plot-försök, 56  
spridningsmått, 11  
standardavvikelse, 12  
stickprovsstorlek, 27  
strata, 10  
strip-plot-försök, 56  
styrka, 31

t-fördelning, 16  
t-test, 15, 17  
testvärde, 15

upprepningar, 7

varians, 11  
variansanalys, 23  
varianskomponenter, 35  
variationskoefficient, 12, 24

årseffekter, 59