

Inst för MARIN EKOLOGI, GU



STATISTISK ANALYS OCH UTFORMNING EXPERIMENT 3 VARIANSANALYS

Användningsområden	1
Enfaktors variansanalys (ANOVA).....	2
Antaganden för variansanalys	6
Multipla jämförelser ("Multiple-comparisons", "a posteriori-test").....	7
Hierarkisk ("nested") ANOVA.....	7
Exempel 1 - Bottenundersökning	7
Exempel 2 - Akvarieexperiment.....	9
ANOVA med ortogonal faktorer	12
Fler än 2 ortogonala faktorer.....	15
Generell metod att finna varianskomponenter	16
Utformning av stickprovsprogram och experiment.....	17
Analys av statistisk styrka.....	22

Användningsområden

Med t-test kan man testa om det är sannolikt om **två** stickprov kommer från samma eller två olika populationer. De flesta frågeställningar kräver dock betydligt mer komplicerade observationer eller experiment än bara en jämförelse mellan två stickprov. Exempel på mer komplexa situationer är:

1. Fler än två grupper som ska jämföras, d.v.s. en faktor har fler än två nivåer, t.ex. en jämförelse av förekomsten av en makroalg mellan tre lokaler (Fig. 1a).
2. När man vill undersöka på vilken rumskala variationen är störst genom att ta stickprov ordnade i en hierarkisk struktur med succesivt minskande skala. Figur 1b visar ett exempel med tre hierarkiska nivåer.

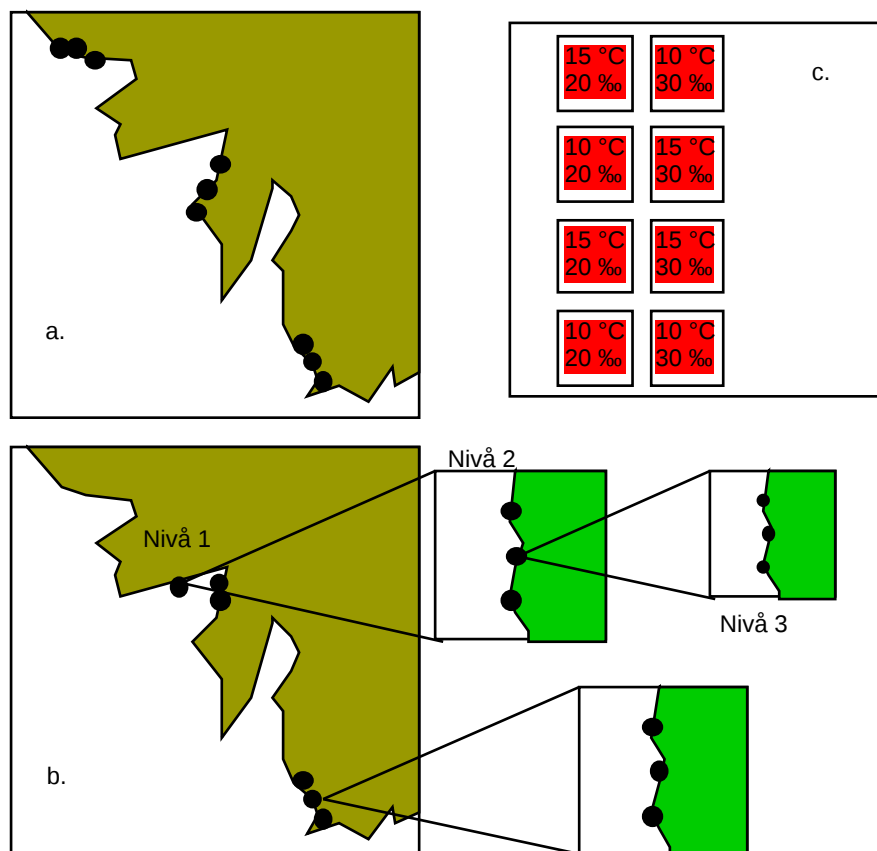


Fig. 1. Illustrationer av användningsområden för ANOVA inom ekologi.

3. När man vill undersöka hur flera faktorer påverkar en variabel tillsammans, t.ex. salinitet och temperatur. En sådan studie kan svara på om det finns synergistiska eller antagonistiska effekter mellan olika faktorer (Fig. 1c).

För att lösa ovanstående problem måste man använda andra analysmetoder än t-test. En kraftfull statistisk metod som ger oss dessa möjligheter är variansanalys (eng. analysis of variance = ANOVA).

Enfaktors variansanalys (ANOVA)

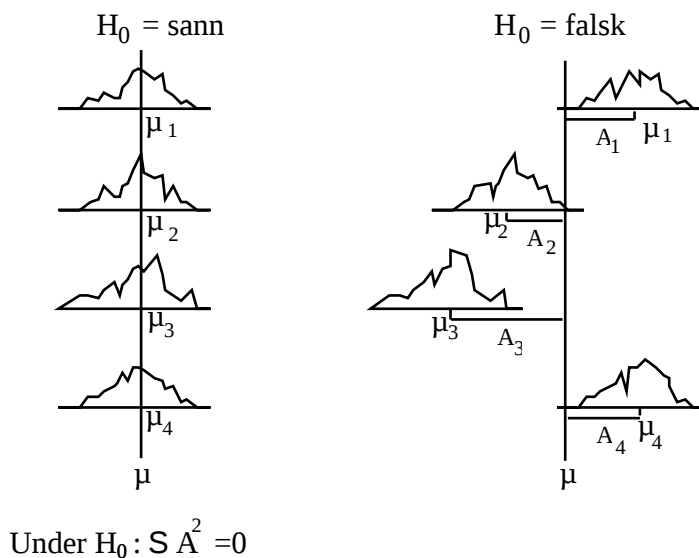
Om vi har flera nivåer i en faktor, t.ex. tre salthalter, skulle vi kunna göra tre parvisa t-tester mellan alla nivåer. Det finns två problem med detta:

1. Det blir snabbt många jämförelser när antal nivåer växer.
2. Vi gör oss skyldiga till ett allt större Typ I-fel med ökande antal jämförelser.

Antal nivåer	Antal jämförelser (k)	Typ 1-fel, $1-(1-\alpha)^k$
2	1	0.05
3	3	0.14
5	10	0.40
10	45	0.90
15	105	0.995

Det här problemet löste Ronald Fisher i början på 30-talet. Han började med ett antal stickprov (se Fig. 2) och ställde upp noll-hypotesen att $\mu_1 = \mu_2 = \mu_3 = \dots \mu_i$. Om istället noll-hypotesen är falsk är stickprovens medelvärden olika p.g.a. en behandlings-effekt med storleken $A_1, A_2, \dots A_i$ (Fig. 2). Under noll-hypotesen, H_0 , gäller att: $A_1^2 = A_2^2 = A_3^2 = \dots A_i^2 = 0$ eller, $\sum A_i^2 = 0$. För alternativhypotesen gäller att: $\sum A_i^2 > 0$

Logik bakom variansanalys



$$MS_A \text{ skattar: } s_e^2 + ns_A^2$$

$$MS_{\text{residual}} \text{ skattar: } s_e^2$$

$$\text{Test statistika: } F = \frac{MS_A}{MS_{\text{residual}}}$$

Fig. 2. Schematisk illustration av fördelningar när H_0 är sann respektive falsk.

Det fina här är att jämförelsen blir ensidig. Vi behöver egentligen bara beräkna faktorns effekt, d.v.s. SA^2 och se om den är större än vad som kan förväntas under H_0 .

Problemet är ju bara att vi inte kan skatta A utan vidare eftersom det till variationen p.g.a. faktorns effekt A också adderas den variation som finns mellan observationer inom en och samma nivå, d.v.s mellan replikaten. Vi kan liksom vid regressionsanalys uttrycka problemet som en linjär modell.

$$X_{ij} = \mu + A_i + e_{ij}$$

Frågan är om A förklarar tillräckligt av variationen i x i förhållande till residual-variationen, e_{ij} , inom grupperna (nivåerna). Hur kan vi då jämföra variationen p.g.a. A_i och e_{ij} ?

Förloppet illustreras i flera steg:

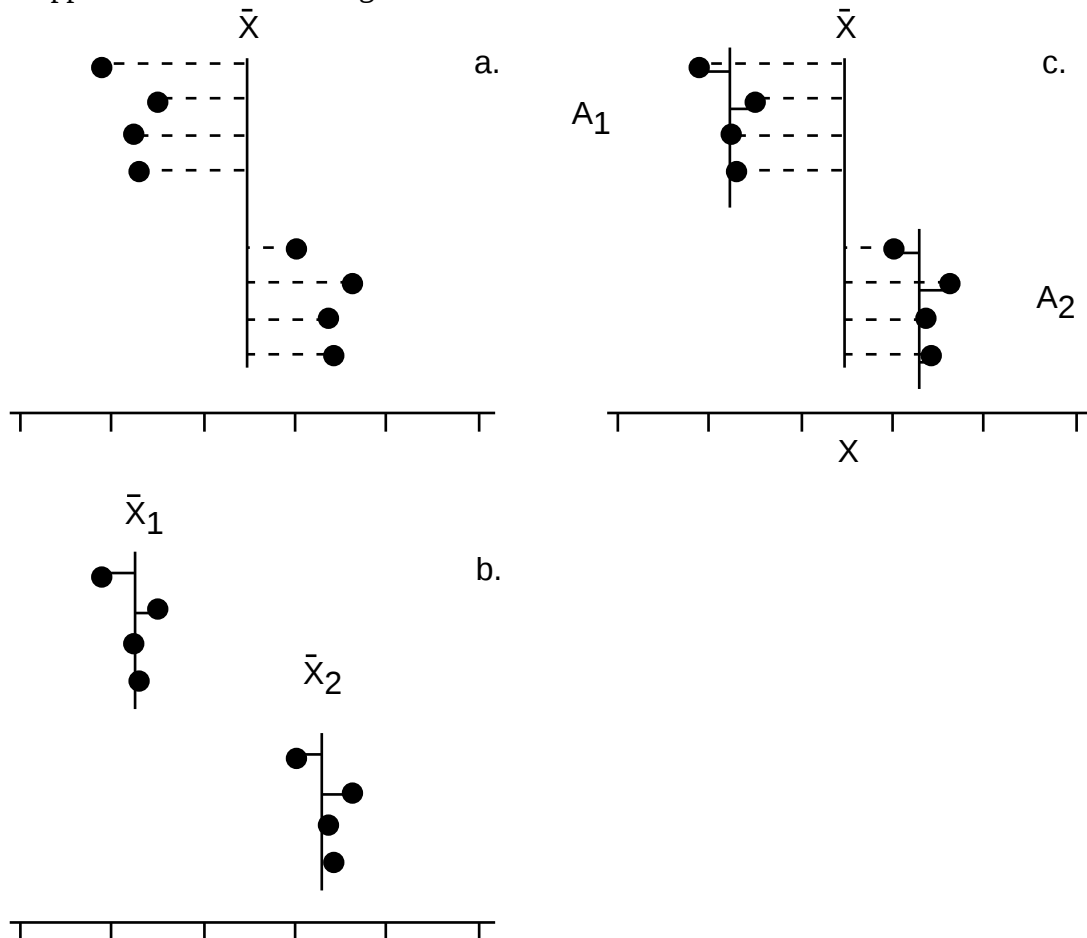


Fig. 3. Illustration av innebörd av SS för olika variationskällor. Se text för ytterligare detaljer.

1. Ett mått på den totala variationen mellan observationerna är deras avvikelse mot det gemensamma medelvärdet. Vi kallar denna den totala SS $S(x_{ij} - \bar{x})^2$ (Fig. 3a).
2. Den totala SS består av variationen p.g.a. av A_i och e_{ij} . Variationen p.g.a. e_{ij} fås genom att summera alla avvikelser mot respektive stickprovs medelvärde $S(x_{ij} - \bar{x}_i)^2$. Vi kallar den residual SS eller error SS (Fig. 3b).
3. Nu kommer vi till den variation som förklaras av faktor-effekterna A_i . Denna kan fås genom subtraktion av residual SS från total SS (Fig. 3c). Den kan också beräknas

som summan av de kvadrerade avvikelserna mellan gruppernas medelvärde och det totala medelvärdet $S(\bar{x}_i - \bar{x})^2$. Vi kallar denna summa för SS mellan grupperna. Denna summa ska dessutom multipliceras med stickprovets storlek (antal replikat). Varför? Jo det är för att SS mellan grupperna ska vara i samma "valuta" som de övriga så att de går att jämföra. Vi kan se detta om vi först antar att H_0 är sann och de olika grupperna har samma sanna medelvärde (vänster i Fig. 2). P.g.a. icke-representativitet kommer de skattade medelvärden att avvika något. När vi beräknar summan av avvikelserna av de olika $\bar{x}_i$ mot det totala medelvärdet, vad är det vi skattar då egentligen? Kom ihåg centralvärdesteoremet! Det är ju i själva verket $\frac{(n-1) s_e^2}{n}$ vilket skattas av $\frac{SS_{\text{residual}}}{n}$. Det innebär att om vi multiplicerar denna kvantitet med n så får vi SS residual; det är precis vad vi vill. Det innebär att om alla A_i är 0 kommer $n * SS$ mellan grupperna också att skatta SS residual.

4. Nu börjar vi bli klara för analysen av variationen i våra data enligt modellen ovan. Vi skriver upp alla variationsbidrag:

Bidrag	SS	df	MS	MS skattar
Mellan grupperna (Faktor A)	$S(\bar{x}_i - \bar{x})^2$	$a-1$	$\frac{SS_a}{a-1}$	$s_e^2 + \frac{nS A_i^2}{a-1}$
Residual (inom grupperna)	$S(x_{ij} - \bar{x}_i)^2$	$an-a$	$\frac{SS_{\text{res}}}{an-a}$	s_e^2
Total	$S(x_{ij} - \bar{x})^2$	$an-1$	$\frac{SS_{\text{tot}}}{an-1}$	

Nästa steg är att bilda ett variationsmått som är oberoende av stickprovsstorleken. Det görs genom att dela SS med antalet frihetsgrader, vilket ger det som kallas för MS (eng. mean square). Detta är analogt med en vanlig varians med den skillnaden att vissa MS skattar flera varianser. Vi skriver också ut till höger i tabellen vilka varianser MS skattar. Antalet frihetsgrader fås genom att ta det antal kvadrerade avvikelser som varje bidrag är baserat på och minska detta med antalet oberoende skattade parametrar som behövs för att beräkna avvikelserna.

Vi är nu redo att jämföra våra varianskomponenter och testa för en signifikant effekt av A. Vi skapar en **statistika**, **F**, genom att bilda en kvot mellan två MS. Den relevanta F-kvoten för test av eventuell effekt av faktor A är att dela MS_A med MS_{res} . Principen när man försöker hitta den relevanta F-kvoten är att nämnaren ska innehålla alla varianskomponenter som i täljaren, utom den man är intresserad av testa effekten för. Vi ser att under H_0 förväntas F-kvoten vara 1, medan en effekt p.g.a. A måste öka F. Frågan är hur stor F måste vara för att det ska vara osannolikt att täljare och nämnare skattar samma varians, d.v.s. att vi ska förkasta H_0 . I tabeller finns fördelningen av F

som är beroende av antalet frihetsgrader i både täljare och nämnare. Här kan man bedöma med vilken sannolikhet en funnen F-kvot bör leda till förkastande av H_0 eller ej.

Antaganden för variansanalys

1. Normalitet. Som i all parametrisk statistik krävs även här egentligen att stickproven ska komma från sanna populationer som är normalfördelade. I praktiken har det dock visat sig att variansanalys är ganska okänsligt för avvikelser från normalfördelning.
2. Homogena varianser (homoscedasticitet). Avvikelser från kravet att de sanna populationerna endast får skilja sig åt i sina medelvärden och inte i sina varianser är betydligt allvarligare. Testar man stickprov som kommer ur populationer med mycket olika varianser kommer man att göra ett typ I-fel av okänd storlek. I korthet kan sägas att F-tabellerna inte längre ger de korrekta kritiska F. Man bör därför alltid först testa H_0 att varianserna är lika med t.ex. Cochrans' test:

$$C = \frac{\max(s^2)}{S(s^2)}$$

Måste H_0 förkastas kan man inte använda en parametrisk variansanalys. En möjlighet är att s.k. transformera sina data på ett sådant sätt att varianserna stabiliseras (blir mer lika).

3. Oberoende observationer. Som vi redan berört måste varje stickprov innehålla oberoende observationer. Detta är mycket viktigt. Ett beroende har i allmänhet helt oförutsägbara konsekvenser och varje analys blir tveksam. Detta är det absolut vanligaste felet i vetenskapliga studier. Några exempel på beroende:

Inom en nivå: Krabbor i ett akvarium stimulerar varandras födoaktivitet vilket minskar variationen i aktivitet mellan individerna.

Mellan nivåer: I ett fältexperiment manipuleras ytor genom att ta bort predatorer. Kontrolltytor lämnas omanipulerade. Under predationsförsöket sker hela tiden en återkolonisation av predatorer från angränsande områden. Tyvärr, ligger experiment- och kontrolltytor så tätt ihop att många av de predatorer som vandrar in i experimenttytorna kommer från kontrollområdet. Vad som sker i experiment- och kontrolltytor är inte längre oberoende av varandra. Effekten i detta fall är en underskattning av effekten av predatorer.
4. I de flesta fall bör man också se till att man har lika många observationer (replikat) i alla grupper, en s.k. balanserad analys. Obalanserade data kan ge upphov till olika allvarliga fel, främst vid analyser med flera faktorer (se nedan). Dessutom är alla balanserade analyser mer robusta mot heterogena varianser.

Multipla jämförelser ("Multiple-comparisons", "a posteriori-test")

I en ANOVA testades om det fanns en skillnad i antalet fångade byten mellan 5 predatortätheter. F-kvoten blev större än den kritiska och H_0 , att det inte fanns någon skillnad mellan behandlingarna, förkastades vid $\alpha = 0.05$. Vad säger då detta? Jo, endast att det finns en signifikant effekt av faktorn Predatortäthet, d.v.s. att antalet fångade byten beror på hur många predatorer det finns. Anovan säger inget om vilka av de 5 nivåerna (tätheterna) som skilde sig åt. Är man intresserad av mer detaljerad kunskap om vilka nivåer som skiljer sig åt kan man utföra s.k. "multi-comparison tests". Ett vanligt sådant test är Student-Newman-Keuls test (SNK) som innebär en stegvis, parvis jämförelse mellan alla nivåer. Multi-comparison test har alltid lägre statistisk styrka än Anovan. Det händer därför ibland att man får en signifikant effekt av en faktor, men i t.ex. ett SNK-test finner man ingen signifikant skillnad mellan nivåerna.

Hierarkisk ("nested") ANOVA**Exempel 1 - Bottenundersökning**

Vi introducerar nu variansanalys med fler än en faktor. En typ av experiment eller stickprovsprogram består av flera faktorer ordnade hierarkiskt. Detta är vanligt när man med stickprov vill undersöka om det finns en skillnad mellan olika platser. Vi börjar att ge ett exempel på en felaktig design för en sådan stickprovsanalys. En undersökning utföres för att ta reda på om antalet havskräftor skiljer sig åt mellan inre och yttre Skagerak. På 2 slumpmässigt utvalda lokaler i yttre och 2 i inre Skagerak stoppar vi båten och tar 5 hugg på varje ställe. Materialet sållas och räknas. När vi sedan testar H_0 att det inte är någon skillnad mellan yttre och inre Skagerak gör vi en variansanalys med 10 replikat i varje grupp:

	X		X		X		X
Inre lokal 1	10	Inre lokal 2	15	Yttre lokal 3	27	Yttre lokal 4	38
Inre lokal 1	23	Inre lokal 2	38	Yttre lokal 3	18	Yttre lokal 4	49
Inre lokal 1	45	Inre lokal 2	18	Yttre lokal 3	35	Yttre lokal 4	53
Inre lokal 1	12	Inre lokal 2	25	Yttre lokal 3	19	Yttre lokal 4	37
Inre lokal 1	21	Inre lokal 2	20	Yttre lokal 3	24	Yttre lokal 4	48

Följande medelvärden kan beräknas:

	Inre lokal 1	Inre lokal 2	Yttre lokal 1	Yttre lokal 2
$\bar{x}$ lokal	22.2	23.2	24.6	45
$\bar{x}$ inre/ytte	22.7		34.8	
$\bar{x}$ total	28.7			

F-kvoten mellan MS_{Skagerak} för inre/ytte Skagerak och MS_{residual} blir här 5.2 med 1 över 18 frihetsgrader vilket är större än det kritiska F på 4.4. Alltså förkastas H_0 .

Men, så som stickproven är tagna är det ologiskt att dra slutsatsen att det finns en skillnad mellan inre och yttre Skagerak. Varför? Problemet är att de fem huggen på varje plats är beroende. Alla 10 huggen i inre Skagerak var inte oberoende utplacerade utan hamnade i två oberoende grupper. Vi har alltså inte 1 över 18 frihetsgrader i testet utan bara 1 över 2. Nu har den statistiska styrkan sjunkit i vårt test och H_0 kan inte längre förkastas. I detta fall var alltså inte varje hugg ett äkta replikat utan ett pseudoreplikat. Vilken var då den rätta slutsatsen i den första testen med alla hugg som pseudoreplikat? Jo, förkastandet av H_0 berodde på att lokal 4 skiljde sig kraftigt från de övriga. Den råkade ligga i yttre Skagerak men det enda man kan säga är att det finns skillnader mellan lokaler.

Ofta tar man prover på det sätt som beskrivits ovan men man är då medveten om att det finns två rumsliga faktorer, en storskalig som beskriver skillnader mellan inre/ytte Skagerak, och en småskalig faktor som beskriver skillnader mellan lokaler. På detta sätt kan man få en hierarkisk serie av faktorer som beskriver variation på allt mindre rumskalor. Man säger att faktorn lokal ligger under faktorn Skagerak, eller att lokal är nestad under Skagerak, ofta betecknat med en parentes lokal(Skagerak). En linjär modell av exemplet ovan kan nu skrivas som:

$$X_{ijk} = \mu + \text{Skagerak}_i + \text{Lokal (Skagerak)}_{j(i)} + e_{ijk}$$

och analysen blir:

Bidrag	SS	df	MS	F	MS skattar
Skagerak (A)	732	1	732	1.4	$s_e^2 + nS^2_{B(A)} + \frac{nbS A^2_i}{a-1}$
Lokal (B(A))	1043	2	521	5.6	$s_e^2 + nS^2_{B(A)}$
Residual	1492	16	93		s_e^2
Total	3267	19			

Notera att rätt F-kvot för Skagerak fås genom att dela MS_{Skagerak} med MS_{lokal} . Felet vi gjorde i början var ekvivalent med att testa MS_{Skagerak} över MS_{residual} . Nu kan man också se fördelen med en nestad ANOVA eftersom den ger möjlighet att dela upp

variationen på olika rumskalor. Lokal förklarar här en signifikant del av variationen. En studie som syftar till att undersöka skillnader i olika geografiska områden kan vinna mycket på att i ett pilotskede inkludera flera hierarkiska faktorer. På så vis kan man undersöka vilka rumskalor som förklarar den största variationen. Med sådan information kan sedan ett effektivt stickprovsprogram utformas som ser till att stickproven fördelas på de skalor där man har mest variation.

Exempel 2 - Akvarieexperiment

Även i experiment finner man ofta hierarkiska faktorer. I ett försök undersöks om fiskar tillväxer olika bra beroende på vilken föda de får. Experimentet omfattar tre akvarier som får diet 1 och tre akvarier som får diet 2. I varje akvarium finns fem fiskar som vägs före och efter experimentet. Variationen i viktförändring efter experimentet analyseras enligt modellen:

$$x_{ijk} = \mu + \text{Diet}_i + \text{Akvarium}(\text{Diet})_{j(i)} + e_{ijk}$$

Med modellen har man visat att man förstått att akvarium är den oberoende experimentenheten och att varje fisk är ett pseudoreplikat eftersom de fem fiskarna i varje akvarium inte utgör oberoende stickprov. På samma sätt som i bottenundersökningen ovan kan man identifiera två faktorer som kan förklara variation i tillväxt, Diet och Akvarier, där Akvarier är nestad under Diet. Residualvariationen är den oförklarade variationen mellan fiskar i samma akvarium. Schematiskt kan man se hur Anovan fungerar i Figur 4.

Bidrag	SS	df	MS	MS skattar
Diet		a-1		$s_e^2 + n s_{B(A)}^2 + \frac{nb \sum A_i^2}{a-1}$
Akvarier(Diet)		ab-a		$s_e^2 + n s_{B(A)}^2$
Residual		abn-ab		s_e^2
Total		abn-1		

Genom att skriva ut vilka variationskomponenter som MS skattar ser man tydligt hur de relevanta F-kvoterna ska bildas. F-kvoten för Diet blir $MS_{\text{diet}}/MS_{\text{akvarier}}$ och F-kvoten för Akvarier blir $MS_{\text{akvarier}}/MS_{\text{residual}}$. En signifikant effekt för Akvarier betyder att olika akvarier med samma diet gav olika tillväxt, vilket kan indikera att det är något med försöksupställningen som inte är bra. Kanske vissa akvarier står på en utsatt plats i försöksrummet, kanske har man inte fördelat fiskarna slumpmässigt mellan akvarierna. Ett exempel på det senare är om man plockar upp fiskar från ett lagringsakvarium till sitt experiment och fyller på med fiskar akvarium för akvarium. Troligtvis hamnar då de slöaste fiskarna i det första och de piggaste i det sista. Hur bör man göra istället (fundera själv)?

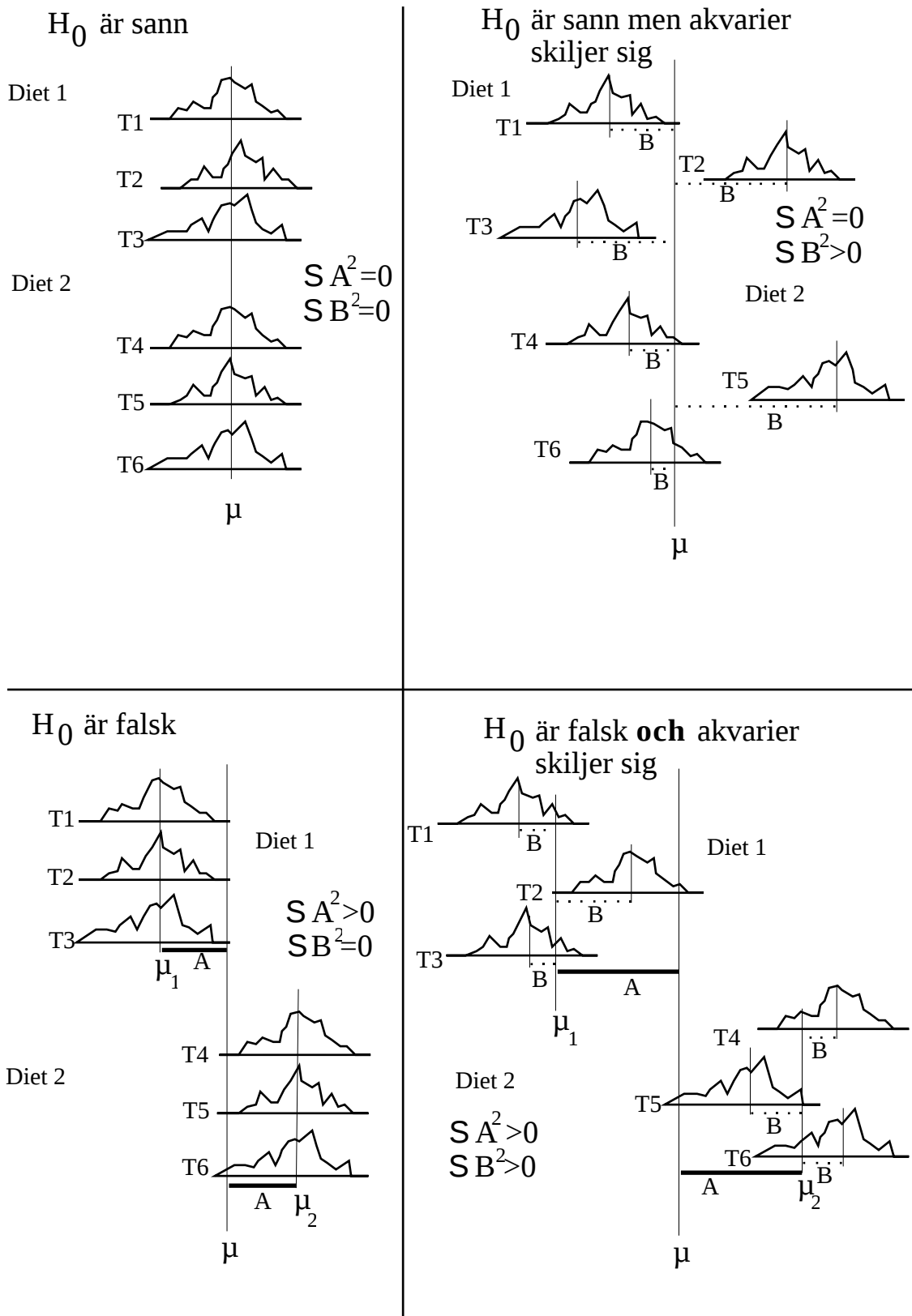


Fig. 4. Schematisk bild av olika utfall vid test av hypoteser med nestad ANOVA.

Om akvarier inte skiljer sig åt och förklarar en liten del av variationen jämfört med residual-variationen kan man eventuellt lägga ihop SS för Akvarier och Residual och bilda

en ny, s.k. "pooled", MS genom att dela med summan av frihetsgraderna för Akvarier och Residual. När nu effekten av Diet testas över den "poolade" MS ökas antalet frihetsgrader i testet vilket ger en ökad statistisk styrka, d.v.s. en ökad möjlighet att upptäcka en skillnad.

Varför nestade faktorer?

Det finns flera skäl till att uforma experimentet med en nestad faktor som i exemplen ovan.

1. Man kanske har ett begränsat antal akvarier men istället för att tillsätta en fisk i varje akvarium väljer man att tillsätta fler fiskar som ger möjlighet att undersöka om olika replikat varierar mycket. Stor variation mellan replikat indikerar att det finns andra icke identifierade faktorer som påverkar variationen i experimentet, och man kanske måste modifiera utformningen av experimentet.
2. I allmänhet har man begränsade resurser (tid, pengar) när man utför experiment. I exemplet ovan finns en kostnad för att analysera varje fisk och en kostnad för att hålla igång varje akvarium. När man kan dela upp variationen i flera nestade faktorer kan man göra en s.k. "cost-benefit" analys. Frågan är här om man ska ha många akvarier med en fisk i varje eller få akvarier med många fiskar. Svaret beror på den relativa kostnaden mellan fiskanalys och akvariehantering, samt den relativa variationen mellan fiskar och mellan akvarier. Man kan visa att för en begränsad kostnad, C , blir antal akvarier per behandling, b , och antal fiskar per akvarium, n :

$$b = \frac{C}{C_b + nC_n} \quad n = \sqrt{\frac{C_b S_e^2}{C_n S_{B(A)}^2}}$$

där C_b är kostnad per akvarium, C_n är kostnad per fisk, $S_{B(A)}^2$ är variationen förklarad av akvarier och S_e^2 är residual variationen (mellan fiskar i samma akvarium).

Använder vi ekvationerna ovan för att göra en "cost-benefit"-analys för exemplet med bottenundersökningen kan det se ut på följande sätt. Vi har 100000 kr till förfogande. Att analysera ett bottenhugg tar ca 1 dag eller 1000 kr i lönekostnad. Att besöka en ny lokal tar ca 3 h i båttid vilket kostar 3000 kr. Från pilotstudien ovan ser vi att S_e^2 (mellan hugg) är 93 och $S_{B(A)}^2$ (mellan lokaler) är 86. Enligt ekvationen ovan ska vi besöka ca 20 lokaler och ta 2 hugg på varje. Eftersom vi har två nivåer av den fixerade faktorn inre/ytte blir det totalt 10 lokaler i inre Skagerak och 10 i yttre.

ANOVA med ortogonal faktorer

Vi har sett hur man kan inkludera nestade faktorer under varandra. I allmänhet betyder inte de olika nivåerna i den nestade faktorn något särskilt. Det är bara en annan lokal eller ett annat akvarium. Syftet är inte att t.ex. se hur ett visst akvarium med en viss diet påverkar tillväxten. Istället vill man undersöka hur mycket tillväxten varierar beroende på att man mäter i olika akvarier. Den beskrivningen stämmer ju bra in på en slumpad faktor, och nestade faktorer är alltid slumpade. En nestad faktor utmärker sig på ytterligare ett sätt. I exemplet med akvarieförsöket så finns varje nivå av faktor Akvarium bara i en nivå av Diet; ett akvarium kan ju bara representera en nivå av diet. Nu ska vi börja inkludera flera faktorer där varje nivå i en faktor finns i varje nivå av en annan, eller med andra ord alla kombinationer av faktorerna är med.

Ofta vill man testa hur effekten av en faktor påverkas av effekten av en annan faktor. Enda sättet är att inkludera alla kombinationer av de båda faktorerna i en s.k. ortogonal design. En statistisk metod som klarar experiment med flera ortogonala faktorer är ANOVA. Det finns två viktiga saker att vinna med att analysera flera faktorer samtidigt:

1. Större statistisk styrka p.g.a. fler frihetsgrader i analysen.
2. Man får information om interaktioner mellan faktorer.

Vi startar med ett konkret exempel. Man önskar undersöka hur överlevnaden hos en mussla påverkas av temperatur och salinitet. Baserat på kunskap om de naturliga förhållandena den lever under bestämmer man sig för att studera två nivåer av temperatur, 10° och 20°, och tre nivåer av salinitet, 20‰, 25‰ och 30 ‰. Vi börjar med att ställa upp en linjär modell för hur variationen i överlevnad kan tänkas förklaras.

$$x_{ijk} = \mu + T_i + S_j + T*S_{ij} + e_{ijk}$$

Variansen har här delats upp på variation som förklaras av en ändring i temperatur, T_i , en ändring i salinitet, S_j , en eventuell interaktion (synergi/antagonism) betecknad med $T*S_{ij}$, och slutligen en oförklarad variation mellan individer e_{ijk} . Vi anser att både temperatur och salinitet är fixerade faktorer eftersom vi skulle ha valt samma nivåer om vi gjort om experimentet. Resultatet från experimentet kanske ser ut som i Figur 5.

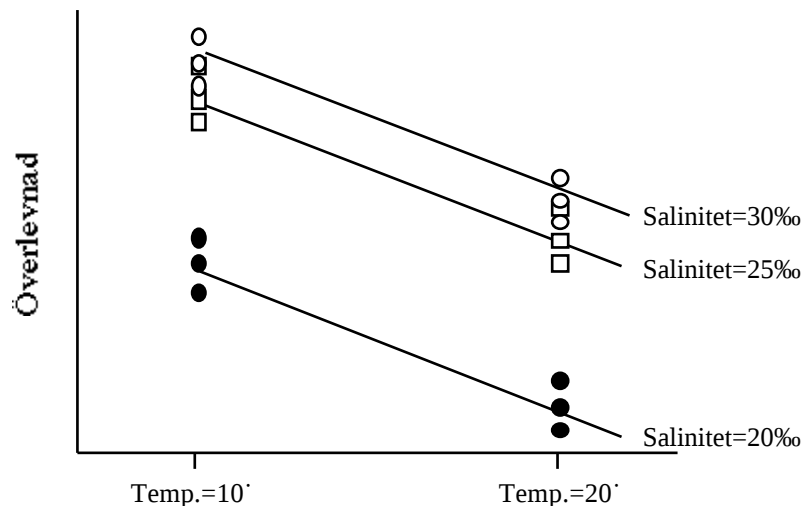


Fig. 5 Resultat från experiment med två faktorer.

En variansanalys ställs upp enligt:

Bidrag	SS	df	MS	MS skattar
Temperatur (A)		a-1		$s_e^2 + \frac{nbS A^2_i}{a-1}$
Salinitet (B)		b-1		$s_e^2 + \frac{naS B^2_j}{b-1}$
A*B		(a-1)(b-1)		$s_e^2 + \frac{nS AB^2_{ij}}{(a-1)(b-1)}$
Residual		ab(n-1)		s_e^2
Total		abn-1		

När man lyckats härleda vilka varianskomponenter som MS skattar är det lätt att bilda F-kvoter för test av effekten av temperatur, salinitet och interaktionen dem emellan.

Två intressanta aspekter belyses tydligt med denna ortogonal 2-faktors Anova, ökning av den statistiska styrkan och test av interaktionen. Vi ser att testet för temperatur sker genom att bilda en F-kvot mellan $MS_T/MS_{\text{residual}}$ med a-1 över ab(n-1) frihetsgrader. Det faktum att temperatur analyseras samtidigt som salinitet gör alltså att testet för temperatur får b gånger fler frihetsgrader i nämnaren. Den statistiska styrkan för test av faktorn temperatur ökar här betydligt. Anledningen är givetvis att för testet av temperatur ovan så finns det b*n replikat för varje nivå av temperatur istället för bara n replikat om salinitet inte hade inkluderats i analysen. Som nämnts ovan är detta faktum ett starkt skäl varför man bör analysera flera faktorer samtidigt.

Möjligheten att analysera **interaktioner**, dvs om effekten av en viss faktor påverkas av en annan faktor utgör kanske den största styrkan med ANOVA. Särskilt i biologiska

system med komplicerad dynamik är interaktioner mycket vanliga. Ingen annan statistisk analys ger samma möjlighet till analys av interaktioner. Vad innebär då en interaktion? I exemplet ovan är båda faktorerna temperatur och salinitet signifikanta men det finns ingen signifikant interaktion (Fig. 5). I ord kan det uttryckas som att effekten av temperatur, d.v.s. att minska överlevnaden, är densamma oavsett vilken salinitet musslorna är utsatta för. Omvänt är effekten av salinitet, d.v.s. att öka överlevnaden, densamma oavsett temperatur. Figur 6 är ett exempel på ett annat resultat med en signifikant interaktion.

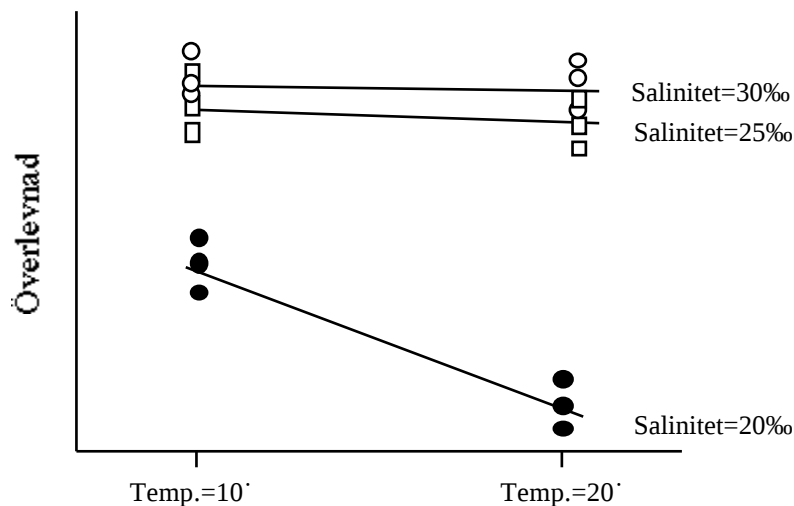


Fig. 6. Utfall av experiment med två faktorer där det finns en interaktionseffekt.

Här ser vi att effekten av temperatur är starkt beroende av vilken nivå av salinitet musslorna befinner sig i. Det är mycket viktigt att notera att när man har en signifikant interaktion kan man inte på ett entydigt sätt tolka huvudeffekterna (eng. main effects) temperatur och salinitet. För resultatet i Fig. 5 är det innehållslöst att hävda att temperatur påverkar överlevnaden hos de studerade musslorna. Interaktionen visar att verkligheten är mer komplicerad, d.v.s. att effekten av temperatur beror på saliniteten. När man utför ANOVA med två eller flera faktorer börjar man alltid med att testa interaktionen. Är interaktionen signifikant innebär det att man bör avstå eller vara försiktig när man testar huvudeffekterna. Ibland är bara interaktionen signifikant men ingen av huvudeffekterna (Fig. 7), och det kan i vissa fall endast vara interaktionen som är av intresse att testa.

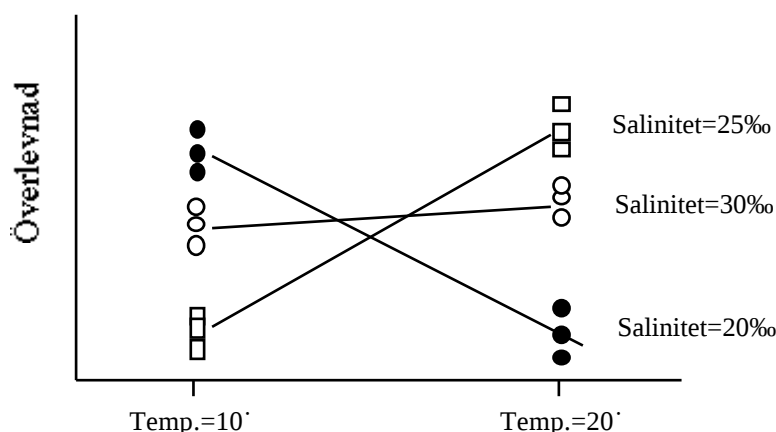


Fig. 7. Utfall av experiment med två faktorer där det finns en kraftig interaktionseffekt men ingen effekt av de enskilda faktorerna var för sig.

Om man får en signifikant interaktion och man önskar testa vilka kombinationer av t.ex. temperatur och salinitet som skiljer sig åt kan man utnyttja "multi-comparison"-test, t.ex. det ovan nämnda SNK.

Fler än 2 ortogonala faktorer

Det fascinerande med ANOVA är att det går att analysera även mycket komplexa experiment och stickprovsundersökningar genom att bygga in alla faktorer, såväl fixerade som slumpade, med alla sina interaktioner. Som ett exempel kan vi bygga ut experimentet med fiskars tillväxt med, förutom två olika dieter, t.ex. tre slumpade temperaturer och honor och hanar. Utformar vi experimentet på ett ortogonalt sätt d.v.s. att alla kombinationer av diet, temperatur och kön finns och att varje akvarium representerar en sådan kombination fås följande variansanalys (utan koefficienter):

Bidrag	df	MS skattar
Diet (fixed)	a-1	$s_e^2 + s^2_{A(D,K,T)} + s^2_{D*T} + k^2_D$
Kön (fixed)	b-1	$s_e^2 + s^2_{A(D,K,T)} + s^2_{K*T} + k^2_K$
Temp. (random)	c-1	$s_e^2 + s^2_{A(D,K,T)} + s^2_T$
Diet*Kön	(a-1)(b-1)	$s_e^2 + s^2_{A(D,K,T)} + s^2_{D*K*T} + k^2_{D*K}$
Diet*Temp.	(a-1)(c-1)	$s_e^2 + s^2_{A(D,K,T)} + s^2_{D*T}$
Kön*Temp	(b-1)(c-1)	$s_e^2 + s^2_{A(D,K,T)} + s^2_{K*T}$
Diet*Kön*Temp.	(a-1)(b-1)(c-1)	$s_e^2 + s^2_{A(D,K,T)} + s^2_{D*K*T}$
Akvarium (Diet, Kön, Temp)	abc(d-1)	$s_e^2 + s^2_{A(D,K,T)}$
Residual	abcd(n-1)	s_e^2
Total	abcdn-1	

Detta ser kanske något skrämmande ut men är inte så besvärligt om man tar det stegvis. Det som framför allt har blivit besvärligare från tidigare är att det är svårt att förstå vilka varianskomponenter MS skattar. Det blir särskilt svårt när vi som i exemplet ovan har två fixerade faktorer, en slumpad (temp.) och en nestad (akvarium). Eftersom vi måste veta vilka varianskomponenter som MS skattar för att bilda rätt F-kvoter är den här informationen avgörande för att analysen ska bli riktig. Som en varning kan nämnas att mycket få statistiska datorprogram klarar av detta på ett riktigt sätt. För att kunna skriva ut de varianskomponenter som MS skattar för en godtycklig utformning av experimentet finns några enkla minnesregler.

Generell metod att finna varianskomponenter

1. Gör en tabell som inkluderar alla de ortogonala och nestade faktorer och interaktioner som ingår i analysen. Ett exempel visas i tabellen nedan.

Identifierad (s2)	Residual	ABC	BC	AC	AB	C	B	A
Koefficient	1	n	na	nb	nc	nab	nac	nbc
Bidrag								
A f	1	-		1	-			1
B f	1	-	1		-		1	
C s	1	-	-	-		1		
AB	1	1			1			
AC	1	-		1				
BC	1	-	1					
ABC	1	1						
Residual	1							

2. Avgör först vilka faktorer som är fixerade och vilka som är slumpade. Indikera detta med ett litet f eller s. I exemplet ovan är A och B fixerade och C slumpad.
3. Avgör vilka nestade faktorer som finns och under vilka andra faktorer de är nestade. Indikera detta med parenteser t.ex. om C är nestad under A fås C(A). C(A) skrivs ut även i interaktioner t.ex. B*C(A).
4. Alla variationsbidrag innehåller residualen. Ange detta med en etta i kolumnen för residual.

5. Alla variationsbidrag innehåller sin egen varianskomponent. Ange även detta med en etta i respektive cell.
6. Nu ska man avgöra om eventuella övriga komponenter ska med. För varje bidrag t.ex. A markeras vilka andra komponenter som innehåller A. I fallet ovan blir det AB, AC och ABC. För AB blir det bara ABC som är potentiellt intressant. Alla dessa potentiella komponenter betecknas så länge med ett streck.
7. Om de potentiella komponenterna ska vara med avgörs genom att se på de faktorer som inte omfattas av det aktuella bidraget. För A i AB är det B, För A i ABC är det B och C etc. Är någon av de övriga faktorerna en fixerad faktor ska komponenten inte vara med. Är alla de övriga faktorerna slumpade ska komponenten vara med vilket indikeras med en etta. Om den aktuella faktorn är nestad är det bara faktorer utanför parenteserna som ska bedömas. T.ex. om C är nestad under A, d.v.s. C(A), och vi har identifierat B*C(A) som en potentiell komponent så ska den med om B är slumpad men inte med om B är fixerad.

Nu kan man skriva ut vilka varianskomponenter respektive MS skattar, d.v.s. alla med ettor i sina celler.

Utformning av stickprovsprogram och experiment

Variationsanalys är ett mycket värdefullt verktyg i många vetenskapliga studier, både i samband med observationer av mönster och experiment. För att det ska finnas en logisk koppling mellan hypotes, experiment och test är det helt avgörande:

1. att det finns en tydlig frågeställning med en eller flera hypoteser man önskar testa. I allmänhet innebär det att man identifierar tänkbara faktorer som kan förklara variation.
2. att man specificerar önskad generalitet. Ska resultatet vara giltigt för ett visst temperatur-område, ett visst geografiskt område etc?
3. att man identifierar den oberoende replikationsenheten. Är det t.ex. akvariet eller fisken i akvariet som är replikationsenheten? Se upp med pseudoreplikat!
4. att man undviker sammanblandning ("confounding") p.g.a. beroende mellan faktorer eller p.g.a. att vissa kontroller saknas. Ligger t.ex. provtytor för två olika behandlingar så nära att en effekt i den ena kan påverka den andra, eller har man i ett transplantationsförsök glömt att inkludera kontrollen för en hanteringseffekt?
5. att man kan översätta sin faktiska experimentutformning till en varians-modell (linjär modell) på ett riktigt sätt. Är faktorerna fixerade eller slumpade? Är någon faktor nestad under någon annan?
6. att man klarar av att finna vilka varianskomponenter MS skattar så att man kan bilda korrekta F-kvoter för de test man är intresserade av.
7. att man om möjligt beräknar hur många oberoende replikat man behöver för att med tillräcklig statistisk styrka upptäcka en specificerad skillnad. Detta steg kräver att man

har någon, åtminstone ungefärlig, information om MS i nämnaren i den F-kvot man är intresserad av.

Det är viktigt att man går igenom alla dessa punkter redan innan man går igång med sin studie. Ofta avslöjas här problem som man inte kommer att tänka på innan det är försent. Som exempel kan nämnas att i många komplexa experiment, där man har fler än två ortogonala faktorer och minst en är slumpad, finns det inte F-kvoter för alla faktorer eller interaktioner.

Formulering av en varians-modell och uppställning av variansanalysen på planeringsstadiet gör det också tydligare hur t.ex. skillnader i hur man lägger ut sina provytor i ett fältexperiment förändrar den logiska tolkningen av resultaten. Exempelvis, om lokal är en slumpad faktor är generaliteten större än om man anser den fixerad men ofta till priset av minskad statistisk styrka i testet (se nedan).

I planeringsarbetet har man också möjlighet att göra en "power" analys och "cost-benefit" analyser för att se om man verkligen klarar av att nå en viss målsättning. Ofta inser man redan vid planeringen att tid och pengar inte tillåter alla önskvärda faktorer, eller att ambitionsnivån måste sänkas när det gäller vilken faktoreffekt som kan upptäckas.

Det tar tid att lära sig utforma experiment och analyser på ett riktigt sätt. Bäst är att träna på konkreta problem. För att belysa hur den specifika frågeställningen är knuten till modell och analys ska vi se på tre snarlika men olika utformningar av ett fältexperiment. Figuren 7, 8 och 9 visar tre olika fältexperiment där man är intresserad av att undersöka effekter på strandsnäckan A:s överlevnad. Faktorerna är också likartade, konkurrens (från art B), predation och lokal. För att utföra experimentet märks rutor ut som får olika kombinationer av dessa behandlingar. På vilket sätt detta sker skiljer sig starkt åt mellan de tre experimenten.

I Fig. 7 är lokal en fixerad faktor, liksom predation och konkurrens. Man är verkligen intresserad av en eventuell skillnad just mellan de 3 lokalerna. Båda faktorerna predation och konkurrens är fullständigt ortogonala mot lokal, d.v.s. varje nivå av en viss faktor finns i alla nivåer av en annan faktor. Alla kombinationer finns, och varje ruta är slumpmässigt placerad i varje lokal. Två oberoende replikat finns. En modell för experimentet kan skrivas:

$$X_{ijkl} = \mu + P_i + K_j + L_k + P*K_{ij} + P*L_{ik} + K*L_{jk} + P*K*L_{ijk} + e_{ijkl}$$

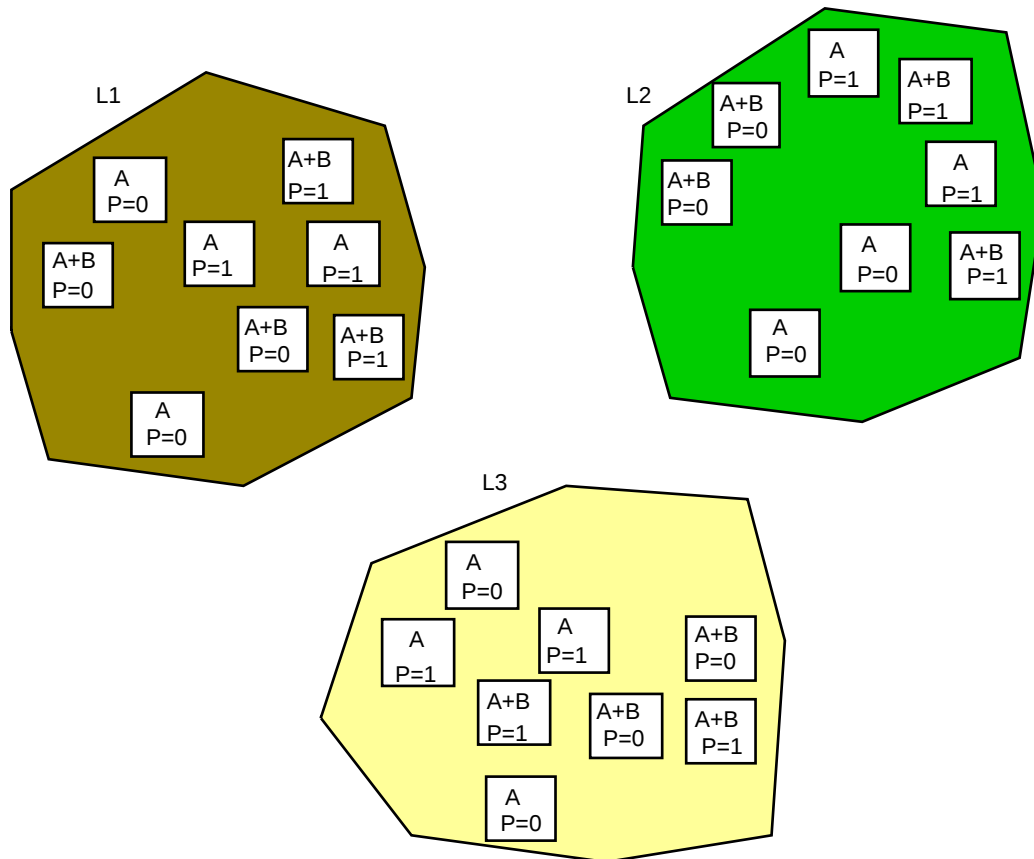


Fig. 7. Schematisk uppställning av experiment om effekter av predation och konkurrens på tre lokaler.

Genom att använda reglerna för att finna vilka varians-komponenter som MS skattar finner vi följande variansanalys (pröva själv):

Bidrag	df	MS skattar
Predation (fixed)	1	$S^2_e + k^2_p$
Konkurrens (fixed)	1	$S^2_e + k^2_K$
Lokal (fixed)	2	$S^2_e + k^2_L$
P*K	1	$S^2_e + k^2_{P*K}$
P*L	2	$S^2_e + k^2_{P*L}$
K*L	2	$S^2_e + k^2_{K*L}$
P*K*L	2	$S^2_e + k^2_{P*K*L}$
Residual	12	S^2_e
Total	23	

I figur 8 står lokal för något annat. Här är lokal en slumpad faktor inom målområdet och representerar bara ett stickprov av olika platser. I varje lokal finns två experimenttytor som har en nivå av konkurrens och en nivå av predation. Varje "nivå" av lokal innehåller bara en nivå av konkurrens och predation och är alltså nestad under båda dessa (jämför akvarier och fiskar). Den oberoende experimentenheten är alltså lokal och de två ytorna

inom lokal är pseudoreplikat som kan användas för att testa om lokaler skiljer sig åt.

Figur 8 är en klassisk design när man vill öka generaliteten till ett större målområde av resultaten från experiment med predation och konkurrens. Modellen kan skrivas:

$$X_{ijkl} = \mu + P_i + K_j + L(P, K)_{k(ij)} + P*K_{ij} + e_{ijkl}$$

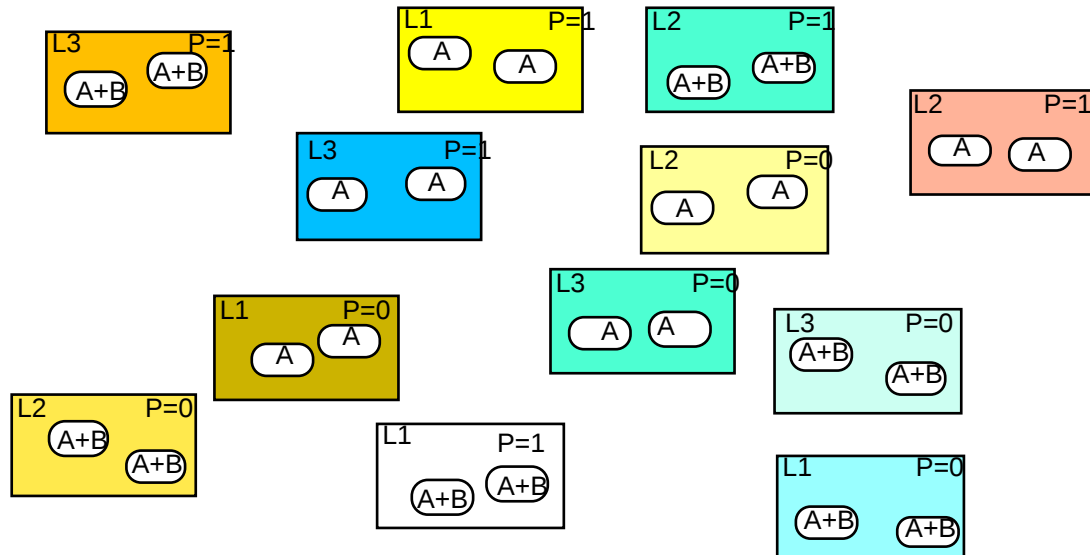


Fig. 8. Schematisk uppställning av experiment om effekter av predation och konkurrens. Endast en kombination av predation och konkurrens finns på varje lokal.

Notera att en faktor som är nestad under en annan faktor aldrig kan användas för att skatta interaktionen dem emellan. En interaktion mellan lokal och konkurrens i Fig. 8 skulle innebära att man jämför hur effekten av konkurrens varierar mellan lokaler. Men varje lokal innehåller ju bara en nivå av konkurrens, t.ex. bara A, och interaktionen kan inte skattas. Därför faller alla interaktionstermer bort i modellen som innehåller lokal. I detta fall skattar MS följande varians-komponenter (pröva själv):

Bidrag	SS	df	MS	MS skattar
Predation (fixed)		1		$S^2_e + S^2_{L(P,K)} + k^2_P$
Konkurrens (fixed)		1		$S^2_e + S^2_{L(P,K)} + k^2_K$
P*K		1		$S^2_e + S^2_{L(P,K)} + k^2_{P*K}$
Lokal (P, K)		8		$S^2_e + S^2_{L(P,K)}$
Residual		12		S^2_e
Total		23		

I Fig. 9 är det ännu lite klurigare. Även här är lokal en slumpad faktor och betyder inget annat än olika platser i rummet. Som i Fig. 8 finns bara en nivå av konkurrens i varje "nivå" av lokal, men däremot finns nu alla nivåer av predation. Lokal är alltså nestad under konkurrens men ortogonal mot predation. Denna design kanske hade valts om det var mycket arbete förknippat med att manipulera ytor med bara A, bara B och A + B.

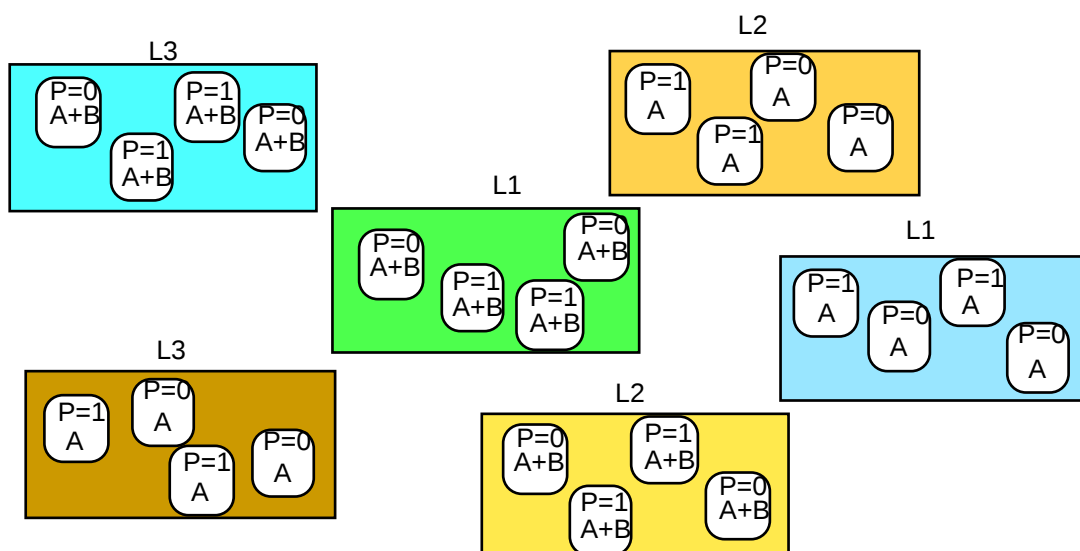


Fig. 9. Schematisk uppställning av experiment om effekter av predation och konkurrens. På varje lokal finns båda nivåer av predation men endast en nivå av konkurrens.

I Fig. 9 klarar man sig med ett fåtal sådana ytor. Men notera också att det finns ett högt pris eftersom de oberoende replikationsenheterna nu har sjunkit till bara 6 st. Det syns också i ANOVA tabellen nedan där t.ex. testet för predation bara kommer att ha 1 över 4 frihetsgrader, ett verkligt svagt test. Jämför detta med Fig. 7 där testet för predation har 1 över 12 frihetsgrader. Modellen för Fig. 9 kan skrivas:

$$x_{ijkl} = \mu + P_i + K_j + L(K)_{k(j)} + P*K_{ij} + P*L(K)_{ik(j)} + e_{ijkl}$$

Nu finns möjligheten att skatta interaktionen mellan lokal och predation.

På nästa sida ses vilka varianskomponenter som MS skattar (pröva själv).

Bidrag	df	MS skattar
Predation (fixed)	1	$S^2_e + k^2_P + S^2_{P*L(K)}$
Konkurrens (fixed)	1	$S^2_e + k^2_K + S^2_{L(K)}$
Lokal (K) (random)	4	$S^2_e + S^2_{L(K)}$
P*K	1	$S^2_e + k^2_{P*K} + S^2_{P*L(K)}$
P*L(K)	4	$S^2_e + S^2_{P*L(K)}$
Residual	12	S^2_e
Total	23	

Dessa exempel visar kopplingen mellan experimentets utformning, analys och slutsatser.

Analys av statistisk styrka

Vi har redan nämnt de olika statistiska fel vi kan göra när vi testar våra resultat mot en noll-hypotes. En mycket viktig del i det vetenskapliga arbetet är att beräkna den statistiska styrkan, eller "power". Power uttrycker sannolikheten att man förkastar H_0 när H_0 är falsk. Det kan också uttryckas som sannolikheten att man ska upptäcka en viss sann skillnad i sitt experiment. Vet man att det är biologiskt relevant att kunna upptäcka en predationseffekt i en bytespopulation på 20% är det givetvis avgörande att det experiment man utformar också kan upptäcka en sådan effekt. Varje gång som ett test av en faktor eller interaktion inte förkastar H_0 bör man ange vilken statistisk styrka som fanns i testet för att upptäcka engiven effektstorlek (minsta intressanta skillnad). Ett exempel på en "power"-analys kan se ut på följande sätt.

I en fältundersökning är man intresserad av att testa om biomassan av fintrådiga alger skiljer sig mellan ett kontrollområde, ett reningsverk med fosforrening och ett reningsverk med både fosfor- och kväverening. Målet är att kunna upptäcka en eventuell förändring med 10 g torrsvikt m^{-2} jämfört med kontrollområdet. Följande pilotstudie gjordes:

Biomassa av fintrådiga alger (g torrsvikt m^{-2})

	Kontrollområde	Reningsverk med P	Reningsverk med P + N
	10	35	19
	20	18	8
	15	28	25
$\bar{x}$	15	27	17.3
s^2	25	73	74.3

Först kalkyleras residualvariationen som ju betecknar det "brus" av oförklarad variation som en effekt måste upptäckas mot.

$$\text{Total residual-varians, } s_e^2 = \frac{\sum (x_{ij} - \bar{x})^2}{3(3-1)} = 57.4$$

"Power" är en funktion av statistikan F : $F = \sqrt{\frac{n D^2}{2a s_e^2}}$ med $3(n-1)$ frihetsgrader

där n är antal replikat, D är den specificerade skillnaden man vill upptäcka, a är antal nivåer i faktorn. För en specificerad skillnad på 10 g torrsvikt m^{-2} och $\alpha = 0.05$ blir power för ökande antal replikat följande:

Antal replikat	Φ	Frihetsgrader	Power (tabellvärde)
2	0.76	3	-
3	0.93	6	-
4	1.08	9	-
5	1.20	12	0.3
6	1.32	15	0.4
7	1.42	18	0.45
8	1.52	21	0.52
9	1.62	24	0.6
10	1.71	27	0.65
12	1.87	33	0.8
14	2.02	39	0.9

Om vi istället nöjer oss med att kunna upptäcka en skillnad på 15 g torrsvikt m^{-2} fås följande:

Antal replikat	Φ	Frihetsgrader	Power (tabellvärde)
2	1.14	3	-
3	1.40	6	0.3
4	1.62	9	0.5
5	1.81	12	0.6
6	1.98	15	0.8
8	2.29	21	0.9

Vi ser alltså att om vi vill kunna upptäcka en effekt på 10 g torrsvikt m^{-2} behöver vi 14 replikat för att få en statistisk styrka på 0.90. Nöjer vi oss med att upptäcka en skillnad på 15 g torrsvikt m^{-2} bör ca. 8 replikat räcka. Statistisk styrka kan beräknas för de flesta utformningar av experiment om man kan specificera en önskad effektstorlek och man har någon information om MS för F-kvotens nämnare (residualvariansen).