

**MANAGEMENTUL ȘI MODELAREA
DATELOR EXPERIMENTALE**

**MASTERAT
BIOLOGIE APLICATĂ 1**

DESIGN EXPERIMENTAL

ÎN STUDIILE DE BIOLOGIE

Conf. univ. dr. Ioan Sîrbu

Design experimental =

= descrierea structurii logice a activității de testare a unor ipoteze.

= conceperea, proiectarea și implementarea judicioasă a unui set de observații și/sau experimente, care permit obținerea de date, prin care testăm ipotezele de lucru, descriptive și/sau funcționale, pentru a îndeplini obiectivele studiului.

= (simplu) modul în care concepem un studiu!

Proiectarea eronată a experimentelor produce:

- informații părtinitoare,
- date nerelevante,
- date redundante,
- imposibilitatea testării ipotezelor,
- cheltuieli inutile ale resurselor,
- eșecul întregului studiu,
- discreditarea cercetătorului sau a echipei.

Categorii majore de experimente:

1. Evaluative (Explorative)

Se execută **măsurători și evaluări** ale unor parametri sau variabile; biologul nu modifică starea sistemului sau a mediului.

2. Manipulative

Sunt aplicate **tratamente** (intervenții ale cercetătorului asupra unor unități de probă, care modifică unii parametri ai mediului sau ai sistemului).

EXPERIMENTELE MANIPULATIVE =

**= SINGURA CALE PRIN CARE PUTEM DEMONSTRA
RELAȚII DE TIP CAUZAL!**

În toate celelalte cazuri căutăm și analizăm relații, despre care în cel mai bun caz putem presupune că sunt de tip cauză - efect, dar NU vom putea demonstra acest lucru.

"Unitatea experimentală" = cea mai mică diviziune a materialului experimental, care este alocată unui singur tip de tratament, adică i se aplică un singur fel de intervenție și posedă o semnificație distinctă.

Identificarea acestor unități este primul pas în designul experimental.

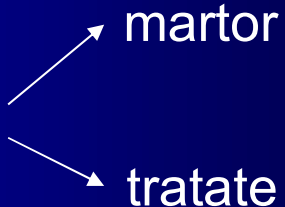
"Numărul de replicări" = câte unități experimentale au fost desemnate sau alocate pentru fiecare categorie de tratament.

Definirea unităților experimentale

Unitățile experimentale trebuie să fie sisteme de tipul celor asupra cărora cercetătorul intenționează să aplice concluziile experimentului.

Exemplu: efectul unor fertilizatori asupra bălților

Unitățile de probă = bălți întregi, dintre care unele



```
graph LR; A[Unitățile de probă = bălți întregi, dintre care unele] --> B[martor]; A --> C[tratate]
```

NU se va adăuga un fertilizator într-o parte a unei bălți și NICI nu se vor compara rezultatele din zona de deversare cu o altă parte a aceleiași bălți unde nu am adăugat substanța respectivă. Acesta ar fi un design eronat! Unitățile experimentale sunt bălți și nu părți ale unor bălți!

Experiențele manipulative implică unul sau mai multe tratamente, care modifică condițiile din **unitățile experimentale**, care sunt împărțite în cel puțin două categorii: **unități supuse tratamentului și unități martor (care nu sunt supuse tratamentului)**.

Regula 1: orice experiment manipulativ trebuie să conțină martori!

Efectele tratamentelor sunt evaluate prin măsurarea uneia, sau mai multor variabile de răspuns.

Designul unui experiment se referă la caracteristicile unităților experimentale, tipul de tratamente aplicate acestora, numărul de unități asupra cărora se aplică fiecare categorie de tratament și modul în care acestea sunt localizate în spațiu și timp.

- TRATAMENTE

Un tratament este o trăsătură a sistemului experimental, care este manipulată de către cercetător, pentru a determina maniera în care sistemul răspunde la alterarea sau modificarea acesteia.

În designul experimental trebuie manifestată grijă, ca tratamentul să nu prezinte componente multiple, dintre care unele nu ar fi evidente.

Asemenea "tratamente ascunse" pot influența rezultatele experimentale și pot conduce la interpretări false.

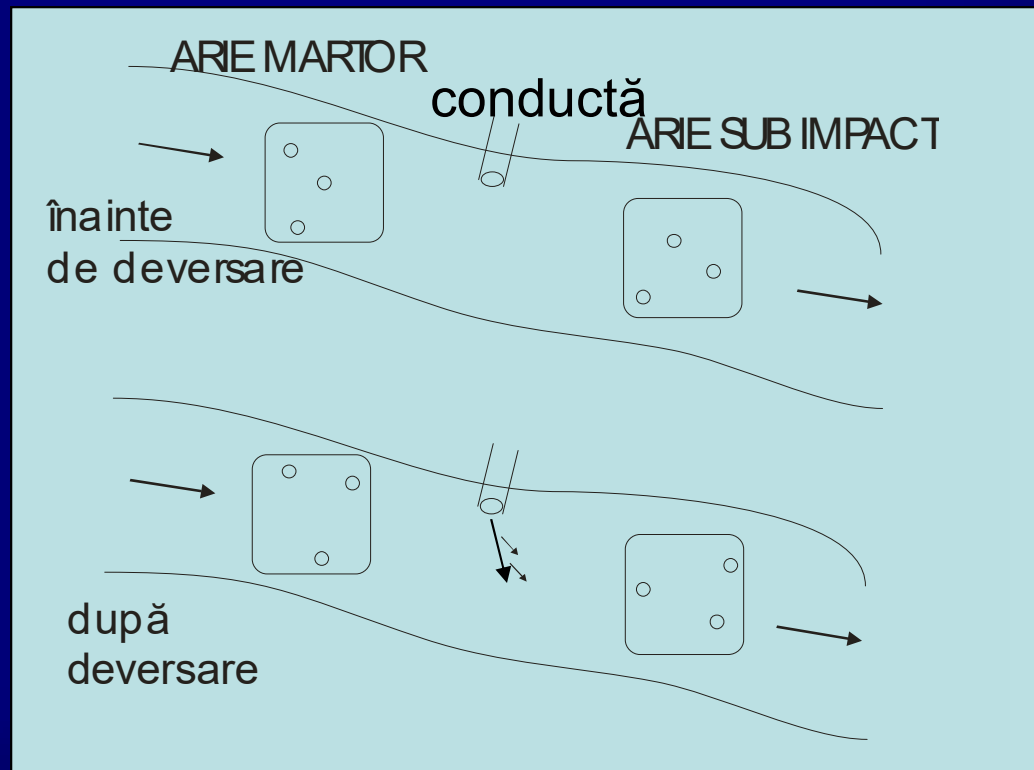
- MARTORII

sunt unități experimentale identice cu cele care sunt supuse tratamentului, dar asupra cărora nu se intervine în nici un fel, exceptând cazurile de factor critic de tratament.

Exemplu de factor critic de tratament și necesitatea acestuia: unități de probă sunt pulverizate cu o soluție apoasă a unei substanțe oarecare, al cărei efect este supus studiului. ATUNCI unitățile martor trebuie pulverizate cu aceeași cantitate de apă FĂRĂ SUBSTANȚA în cauză. Martorul trebuie să releve numai dacă apar diferențe exercitate de substanță și nu de adaosul de apă, eventual și efectele colaterale exercitate de către cercetători (modificarea habitatului, tasarea solului, ruperea vegetației etc.)

În studiile de teren există o imensă variație a tuturor condițiilor de la an la an, motiv pentru care se aplică regula 2:

Regula 2. Orice experiment de teren, dar frecvent (ideal) și în laborator, trebuie să prezinte unități martor contemporane.



Exemplu: design optim de impact - se eșantionează atât unitățile martor, cât și cele supuse impactului, atât înainte cât și după începerea deversărilor; unitățile martor sunt atât spațiale cât și temporale.

Reformulare: experiențele de teren trebuie să prezinte minim patru unități experimentale; DAR, această cerință fundamentală este încălcată de către cercetători în majoritatea studiilor.

Experimentele manipulative sunt supuse la diverse surse de variabilitate care pot genera confuzii și conduc la rezultate dificil sau imposibil de interpretat.

Sursele de confuzie se pot reduce prin proceduri statistice care alcătuiesc "controlul designului":

- randomizare;
- replicare;
- utilizarea de unități experimentale cât mai omogene;
- optimizarea și eficientizarea designului. ;

- RANDOMIZAREA

- Cele mai multe teste și aplicații statistice presupun că observațiile (măsurătorile repetate) sunt independente.
- Această presupunere contravine frecvent realității.
- Principala tehnică de respectare a regulii este **randomizarea**, adică extragerea aleatoare a unei probe din populația statistică, colectarea la întâmplare a probelor, sau alocarea tratamentelor și unităților martor în mod aleator.
- Randomizarea reduce părtinirea și crește acuratețea estimărilor asupra efectelor tratamentelor.
- Randomizarea completă este adesea imposibilă; există alternative (exemplu: colectare sistematică).

Regula 3. Randomizați oricând și oricât este posibil; studiați alternativele!

- REPLICAREA = repetarea și multiplicarea unităților experimentale

"Numărul de replicări" = câte unități experimentale au fost desemnate sau alocate pentru fiecare categorie de tratament.

Ideal, unitățile experimentale sunt cât mai similare, sau chiar identice, sub aspectul dimensiunii, al trăsăturilor structurale și al condițiilor.

Această condiție este adesea violată sau imposibil de îndeplinit! Erorile apar din cauza heterogenității mediului, erori întâmplătoare sau cele datorate cercetătorului.

Aceasta din urmă este numită eroare experimentală și poate fi evaluată, respectiv eliminată, dacă unitățile experimentale sunt **replicate**, adică dacă sunt desemnate mai multe unități pentru fiecare tratament sau condiție, inclusiv pentru martor.

Prin replicare adecvată:

- se estimează și se controlează eroarea experimentală;
- se realizează evaluarea nivelului de semnificație;
- se determină limitele de confidență;
- crește precizia estimării parametrului cantitativ;
- se reduce riscul de intruziune al "zgomotelor" (perturbări nesistematice ale vieții sau mediului, care afectează designul experimental și rezultatele)

Surse de variabilitate în experiențe și proceduri pentru reducerea confuziilor cauzate de acestea

Surse de variabilitate	Reducerea efectelor prin design experimental
1. Variabilitate între unitățile experimentale	Replicare, interspersie, măsurători simultane
2. Eroare randomizată în măsurarea variabilelor de răspuns	Replicare
3. Schimbarea condițiilor în timp	Unități martor și control
4. Efecte colaterale ale tratamentelor	Unități martor și control
5. Părtinirea cercetătorului	Alocarea randomizată a unităților martor și a celor tratate (a tratamentelor)
6. Influențe întâmplătoare asupra experiențelor în derulare	Replicare și interspersie

REPLICAREA (condiție esențială)

nu este sinonimă cu

PSEUDOREPLICAREA

(care este o eroare frecventă

sau o sursă imensă de erori)

Exemplu:

O baltă este evaluată sub aspectul unor parametri fizico-chimici și biologici,

apoi este tratată cu fertilizatori,

după care este evaluată în diferite locuri și perioade de timp sub aspectul efectelor,

constituie un exemplu de PSEUDOREPLICARE.

Este o eroare majoră, nu permite verificarea ipotezelor și nici aplicarea analizei statistice. Rezultatul este un fals!
Sau, rezultatul nu se aplică asupra bălților în general, ci doar asupra bălții particulare studiate.

Pseudoreplicarea apare în diferite variante, voite sau nu.

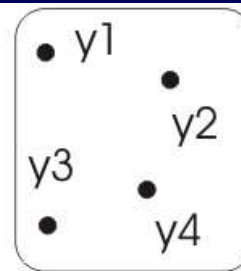
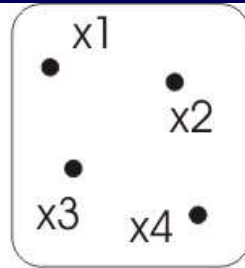
Toate sunt erori!!!!!!

Pseudoreplicarea simplă implică măsurarea multiplă a variabilei de răspuns, într-o singură unitate de experiență, și analiza datelor ca și cum ar proveni din diferite unități.

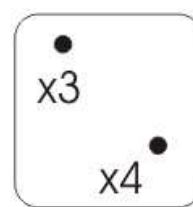
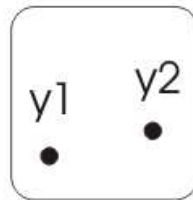
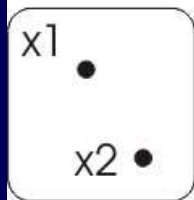
Pseudoreplicarea temporală semnifică realizarea de măsurători în aceeași unitate de experiență, la diferite intervale de timp, și tratarea lor ca și cum ar proveni din diferite unități experimentale.

Pseudoreplicare sacrificială apare atunci când există posibilitatea de a distinge variabilitatea în și dintre unitățile experimentale, dar aceasta este sacrificată prin combinarea tuturor datelor într-o singură grupă de analiză.

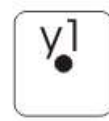
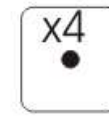
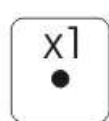
Cele mai comune trei tipuri de pseudoreplicare



Pseudoreplicare simplă



Pseudoreplicare sacrificială



Timp



Pseudoreplicare temporală

Unitățile experimentale supuse tratamentelor și martorii trebuie să fie interspersate în spațiu și timp, pentru a reflecta efectele mediului, a dinamicii modificărilor și ale condițiilor heterogene.

Ambele categorii (în timp-spațiu) ar trebui să fie distribuite la întâmplare.

Când numărul de unități experimentale este mare, o procedură de selectare aleatoare servește scopului urmărit, însă nu și când numărul acestora este mic. Frecvent există limitări și dificultăți în desemnarea unui număr mare de unități.

Soluția este interspersia sistematică sau semisistematică (adică în blocuri randomizate).

Schema generală a celor mai uzuale tipuri de design experimental (după Ch. Krebs, 1989), cu modalități bune sau corecte (A) de interspersie a replicatelor celor două tratamente (în alb și negru), precum și modalități negative sau incorecte (B), care contravin principiului interspersiei.

Tip de design experimental	Schema
A 1 - Randomizare completă	
A 2 - Blocuri randomizate	
A 3 - Sistematică	
B 1 - Segregare simplă	
B 2 - Segregare agregată	
B 3 - Segregare izolativă	
B 4 - Randomizat, dar cu replicare interdependente	
B 5 - Fără replicare	

Design randomizat în blocuri

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

Design în pătrat latin

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

*Reprezentarea schematică a configurării spațiale și metodologice a unităților experimentale într-un **design randomizat de blocuri** și în **pătrat latin** (după Cox, 2002).*

Literale A, B, C și D indică patru tratamente diferite (factori) și martorul.

- Design complet randomizat

Cel mai simplu, recomandat de multe teste. Când unitățile sunt puține, apar probleme atunci când există condiții cu variație clinală.

- Design în blocuri randomizate

Unitățile experimentale sunt grupate în blocuri care sunt relativ uniforme în interior, pe când diferențele dintre blocuri pot fi mai mari sau mai mici. Este un design excelent pentru experimente de teren, deoarece produce o interspersie optimă a tratamentelor, reducând efectele potențiale ale evenimentelor eractice asupra rezultatelor. Un alt avantaj este că blocuri întregi se pot pierde, fără compromiterea experiențelor.

- Design sistematic

Atinge interspersia maximă a tratamentelor, cu riscul apariției de de erori în mediu, sau la intervale de timp, în care factorii oscilează periodic. Va fi evitat când variabilele vor prezenta periodicitate.

Regula 4: se va aloca un număr egal de blocuri pentru fiecare tratament (inclusiv pentru martori)!

EXPERIMENTE FĂRĂ REPLICĂRI (SAU PERTURBĂRI NEREPLICABILE)

- activități antropice planificate sau nu, cu impact potențial major (exemple: deversări accidentale de poluanți, baraje, fabrici, exploatări etc.)
- acestea sunt "tratamente singulare"; excepții de la regulile cercetării și ale designului experimental, prezentate anterior.
- soluții:
 - selectare arii de eșantionare și studii realizate înainte și după construcție sau poluare;
 - selectare arii din vecinătate asupra cărora construcția/perturbarea nu are efecte = unități martor;
 - intervale lungi între măsurători (monitoring îndelungat) pentru ca observațiile să fie independente.

Designul experimental care se poate defini în aceste cazuri este de tipul metodelor ” Martor-Impact, Anterior-Ulterior” (sau perechi MIAU)

(traducere pe sensuri după expresia:

Before-After Control-Impact - BACI,

sau *Before-After Control-Impact Pairs Method* - BACIP).

CONTROLUL DESIGNULUI

Unul dintre obiectivele designului experimental este reducerea erorii experimentale, sau - altfel spus - creșterea preciziei concluziilor.

Există cinci tehnici generale pentru atingerea acestui deziderat:

1. utilizarea de unități experimentale cât mai omogene;
2. utilizarea informației oferite de variabile relaționate care se măsoară în fiecare unitate;
3. utilizarea cât mai multor replicări;
4. utilizarea unui design experimental cât mai eficient,
5. selectarea testelor statistice adecvate designului.

TIPURI DE DESIGN EXPERIMENTAL

Clasificările pot fi **fixe sau randomizate**. În analiza de varianță (ANOVA) dacă un **factor este fix sau aleator**, este crucială pentru testarea tuturor ipotezelor.

Factor fix

- (1) Toate nivelurile de clasificare sunt în experiment, sau
- (2) sunt numai nivelurile de interes pentru cercetător, sau
- (3) nivelurile din experiment au fost alese în mod deliberat și sunt nerandomizate.

Factor aleator

Toate nivelurile din experiment sunt o probă randomizată din toate nivelurile posibile.

Exemplu de factor fix: sexul (când se studiază ambele sexe)

Exemplu de factor aleator: temperatura (se aleg la întâmplare valori)

1. Modele aditive liniare

Toate proiectările complexe utilizate în analiza de varianță pot fi simplu descrise prin asemenea modele.

Se presupune că măsurătoarea obținută, când un tratament particular este aplicat la una dintre unitățile experimentale, este:

$$\left\{ \begin{array}{l} \text{o cantitate dependentă} \\ \text{numai de o unitate} \\ \text{experimentală particulară} \end{array} \right\} + \left\{ \begin{array}{l} \text{o cantitate dependentă} \\ \text{de tratamentul aplicat} \end{array} \right\}$$

Trăsături:

- efectele tratamentului se adaugă la termenii unității;
- efectele sunt constante pentru toate unitățile experimentale;
- unitățile experimentale operează independent.

Exemplu: se studiază densitatea plantulelor de stejar în 6 unități de teren care au fost supuse anterior unui incendiu și 6 martor.

Posibil modelul aditiv:

Densitatea de
plantule de stejar =
pe unitate probă

Densitatea medie
a plantulelor de stejar
în unitățile de probă +
atât incendiate cât și
cele martor

Efectul
existenței +
sau nu a
incendiului

eroare
experimentală

Interesul se concentrază de obicei pe efectele tratamentului, care se estimează din mediile aritmetice observate:

Efectele incendiului
asupra densității =

Densitatea medie
a plantulelor de stejar -
în unitățile incendiate

Media densității
în toate probele

În mod similar se evaluează efectul absenței incendierii asupra densității

Același lucru definit în termenii unei ecuații matematice

$$Y_{ij} = \mu + T_i + e_{ij}$$

unde:

Y = variabila măsurată

μ = valoarea medie a variabilei Y

T = efectul tratamentului

e = eroarea experimentală (variabilitatea dintre replicare)

i = numărul tratamentului (1 = incendiat, 2 = neincendiat)

j = numărul replicării (1, 2, ..., 6)

Modelele aditive se scriu adesea ca deviații: $Y_{ij} - \mu = T_i + e_{ij}$

Exemplu:

Nr. unitate neincendiată / densitate plantule						Nr. unitate incendiată / densitate plantule					
1	2	3	4	5	6	1	2	3	4	5	6
6	9	5	8	11	9	1	2	2	1	4	2
medie = 8						medie = 2					

medie generală = 5

Efectul incendierii asupra densității (T_1) = 2 - 5 = -3

Efectul neincendierii asupra densității (T_2) = 8 - 5 = 3

Se observă că efectele tratamentelor sunt întotdeauna relative, iar cercetătorii se concentrează adesea pe efectele unui tratament în comparație cu celălalt. De exemplu:

Diferența dintre cele două tratamente = $T_1 - T_2 = -3 - 3 = -6$ arbori/m²

2. Design factorial

În multe studii se urmăresc efectele mai multor factori, care acționează în același timp asupra variabilei de răspuns.

De exemplu:

- probe de fitoplancton din diferite lacuri în diferite sezoane;
- rata de eclozare în funcție de salinitate și temperatură;
- biodiversitatea relaționată cu altitudinea și factorii edafici.

Conceptul de factorial este acela conform căruia toate tratamentele unui factor ar trebui studiate sub aspectul interacțiunii cu toate tratamentele celorlalți factori.

Exemplu: dacă se studiază rata de depunere a ponteii de către o specie de pești, la trei valori ale salinității și două temperaturi, se experimentează efecte ale interacțiunii tuturor combinațiilor.

Testul pentru un aranjament factorial se face sub forma unui tabel:

		salinitate		
		mică	medie	mare
temperatură	mică			
	mare			

- fiecare celulă trebuie să conțină date pentru un design factorial
- ideal fiecare celulă va avea același număr de probe
(= *design echilibrat*)
- factorii pot opera independent, dar de obicei acest lucru nu se întâmplă; spunem că ei "interacționează" (interferează etc.)
- interacțiunile produc dureri de cap în ceea ce privește statistica, dar care sunt compensate printr-o știință și cunoaștere interesantă.

Se studiază dinamica densității medii a unei specii din două arii (A1, A2) în fiecare din cele patru sezoane ale unui an (B1, B2, B3, B4).

O analiză bifactorială ar putea fi reprezentată prin:

$$Y_{ijk} - \mu = A_i + B_j + AB_{ij} + e_{ijk}$$

unde:

$Y_{ijk} - \mu$ = deviația valorii observate de la media generală μ

A_i = efectul principal al nivelului i al factorului A

B_j = efectul principal al nivelului j al factorului B

AB_{ij} = termenul interacțiunii dintre A_i și B_j

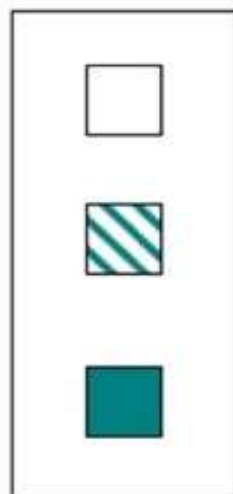
e_{ijk} = eroarea experimentală

- Fiecare dintre acești termeni se estimează și sunt testați în ceea ce privește semnificația, prin analiza de varianță (asemănător cu maniera de lucru la modelarea statistică neliniară).
- Deși designul factorial nu are limite teoretice, există limite practice în studii, care restrâng studiile de laborator sau de teren la design în 2-3 factori, fiecare cu maxim 4-5 niveluri.

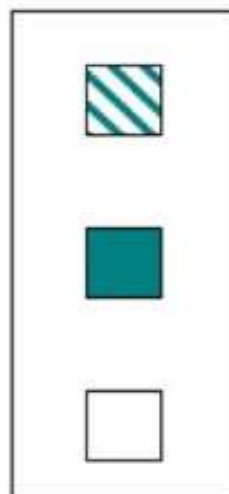
3. Design în blocuri randomizate

- Identificarea blocurilor, ca grupe relativ omogene de unități experimentale.
- Se pot defini sau construi în cadrul sau în jurul oricărei surse de variație.
- Exemple de blocuri posibile:
 - o pajiște pe un versant sudic
 - o arie terestră cu grad ridicat de omogenitate
 - un acvariu într-un laborator
 - o zi din săptămână
 - un grup de animale de aceeași masă și vârstă
 - datele obținute cu un anumit instrument de măsură etc.

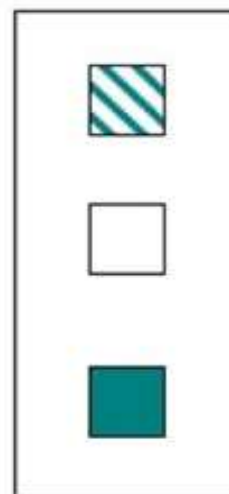
Gradient al mediului



Bloc 1



Bloc 2



Bloc 3



Bloc 4

- Variația dintre blocuri este de mai mică importanță pentru că se știe că aceasta există; din acest motiv ea este înlăturată din termenul de eroare experimentală într-o ANOVA, crescând precizia experimentului.
- Există multe tipuri de design în blocuri;
- Cel mai comun este **designul în blocuri complet randomizate**, în care fiecare tratament apare o singură dată în fiecare bloc, motiv pentru care un bloc conține același număr, maxim, de tratamente aplicate.
- Se cunosc și variante de blocuri incomplete, în care nu sunt prezente toate tratamentele.

4. Design ierarhic sau imbricat (*nested design*)

Toate analizele de varianță se ocupă de unul sau mai mulți factori, respectiv de variabile de clasificare. Factorii pot fi de două feluri:

1. *Efecte principale*: fiecare nivel se poate identifica independent de oricare alt nivel (design factorial)
2. *Efecte ierarhice*: dacă fiecare nivel al unui efect principal poate fi divizat în subgrupe selectate aleator, clasificarea grupelor se numește imbricată sau ierarhică.

Atenție! Designul ierarhic se poate ușor confunda cu cel factorial, de aceea trebuie să ne asigurăm din faza de planificare a studiului despre categoria acestuia. Cele două categorii folosesc o statistică diferită.

Exemplu:

Dieta1		Dieta2		Dieta3	
Nr. pui	Nr. pui	Nr. pui	Nr. pui	Nr. pui	Nr. pui
A	B	C	D	E	F

Nr.	1.	---	---	etc.
crt.	2.	---	---	
individ	3.	---	---	
reprodu-	4.	---	---	
cător				

În acest exemplu numărul de urmași sunt grupați în serii pe diferite diete, și ar fi o eroare să considerăm numărul de pui ca un efect principal al unui design factorial.

Alte exemple de design serial: *lacuri* în cadrul unor *provincii geografice*, sau *adâncimi* în cadrul *lacurilor* de diferite configurații, *probe de polen* la fiecare *adâncime* pentru fiecare *carotă palinologică* etc.

Cel mai simplu model pentru design ierarhic se poate scrie prin ecuația:

$$Y_{ijk} - \mu = A_i + B_{j(i)} + e_{ijk}$$

unde:

Y_{ijk} = valoarea pentru replicata k a nivelului j al factorului serial în nivelul i al factorului principal

μ = media generală a tuturor datelor

A_i = efectul principal al nivelului i al factorului A

$B_{j(i)}$ = Efectul nivelului j al factorului serial în nivelul i al factorului A

e_{ijk} = termenul de eroare

Multe designuri ierarhice sunt rezultatul realizării de subprobe, iar o aplicație importantă este căutarea surselor de variație în măsurători, astfel încât să se poată realiza analize de tipul costuri -beneficii pentru optimizarea efortului de eșantionare.

Metoda statistică = ANOVA ierarhică sau imbricată (***nested ANOVA***)

5. Design în pătrat latin

A	B	C
B	C	A
C	A	B

Design în blocuri randomizate se aplică în general atunci când se cunoaște o sursă de variație înainte de realizarea experienței.

În unele cazuri se cunosc două surse de variație care acționează (de exemplu altitudinea și productivitatea covorului vegetal) și care trebuie considerate în corectarea designului.

O variantă de corecție este utilizarea pătratului latin, extensie a designului în bloc randomizat complet. Fiecare tratament apare o singură dată pe un rând și o singură dată pe o coloană.

Principala constrângere a pătratului latin este că numărul de niveluri al fiecărui factor trebuie să fie același. Dacă această simetrie nu există, se aplică analiza factorială.

Cel mai mic pătrat latin este 3x3 și rar se aplică mai mult de 5x5.