TEMA 6: DISSENY I OPTIMITZACIÓ

DISSENY EXPERIMENTAL. FACTORS

Correlació: que no hi hagi correlació estadísticament significativa entre 2 variables no vol dir que no depenguin una de l'altre, o que no hi hagi una relació entre elles. Pot ser que depengui d'un model més complex: multivariant. Que depengui de més d'un factor: y – x1,x2,x3 ...

Problema químic: obtenir un valor donat, diana de la variable: pot ser mínim, màxim, aprox.

Factor: qualsevol element de les condicions experimentals que pot afectar el valor/resposta de la variable diana, objecte d'interès. Pot ser:

- **Controlat**: és conegut i els seu valor ve definit per l'experimentalista. Ex: pH, T, mètode
- **No controlat:** pot contribuir a la variabilitat: error aleatori o no aleatori, però no pot ser explícitament inclòs en el disseny experimental. Ex: condicions ambientals, fluctuacions instrumentals.
- **Quantitatius**: varien en un rang numèric. Ex: pH, T
- **Qualitatius**: relacionat amb una condició experimental no quantificable i no pot ser expressat numèricament. Ex: mètode, tipus de reactiu...

Nivell: cada valor que un factor pot prendre amb un domini experimental donat.

OBJECTIUS DEL DISSENY EXPERIMENTAL:

- Obtenir la màxima informació química d'un problema donat amb el **mínim nombre d'experiments representatius**. Marcar a priori el **domini experimental** de variació dels nivells dels factors.
- Els experiments es porten a terme variant factors que poden afectar la resposta de la variable objectiu.
- Variacions dels factors han de cobrir adequadament tot el domini/camp experimental.

OBJECTIUS DE L'OPTIMITZACIÓ: escenari avançat del disseny experimental

- Determinar les **condicions experimentals (nivell de significació dels factors)** on la resposta sigui òptima (ex: màxim, mínim o similar a un valor o rang predefinit). Poder arribar a obtenir el model que relaciona la variable amb els factors.
- Predir el **valor de resposta** dels factors de significació: coneixement de la correlació entre factors i resposta variable. Idealment en qualsevol moment.

PASSOS PER AL DISSENY I L'OPTIMITZACIÓ D'EXPERIMENTS

1. ***Screening* del factor: CRIBATGE DE FACTORS:** mínim nombre possible d'experiments
   Disseny dels experiments i avaluació de l'efecte del factor i les seves interaccions.
2. **Optimització:**
   Modelatge de la variació de la resposta d'acord amb els canvis significants de factors i interaccions.
   Recerca bibliogràfica de les condicions experimentals que porten a una resposta òptima.

DISSENY DELS EXPERIEMENTS:

1. Identificació dels factors que poden afectar els resultats. Recerca de factors que poden afectar.
2. Planejar els experiments per tal de **minimitzar** l'efecte de factors no controlats.
3. Definició del rang de valors (nivells) de factors controlats on els experiments es portaran a terme: domini experimental. Puc marcar nivell de factors controlats, dels no controlats no puc.
4. Dissenys d'experiments per screening factors: només es fan tests d'uns quants nivells!
5. Avaluació dels efectes dels factors controlats i de les seves interaccions.
6. També es fa cribratge de les interaccions dels factors.

OPTIMITZACIÓ = màxim, mínim o el valor que m'interessa.

- Disseny detallat dels experiments, incloent només factors significants i interaccions: es fan tests de més nivells (i el camp experimental es pot estendre) Dels 5 factors, em quedo amb 2: 2 nivells. Puc augmentar el nombre de nivells.
- Construcció i validació del model que descriu la resposta basat en factors significants i interaccions.
- Recerca dels valors (o rang de valors) de cada factor que porten a una resposta òptima.

DISSENY D'EXPERMIENTS AMB 1 FACTOR CONTROLAT sospito que un factor controlat afecta.

- **Disseny aleatori**:
- **Disseny en bloc**
  - **Disseny en bloc aleatori**
  - **Disseny quadrats-llatins:** una mica controlat (*analogia sistemàtic pur)*

→ Una seqüencia d'experiments ha de minimitzar l'error aleatori degut a factors no controlats.

→ Es duen a terme **replicats** per tal d'avaluar l'error aleatori. Ho necessito fer perquè sinó tindré problemes per estimar error aleatori.

Ex: experiments a 4 valors de pH diferents (A,B,C,D). Es fan 4 mesures cada dia durant 4 dies consecutius.

day \ order	1	2	3	4
1	A	A	A	A
2	B	B	B	B
3	C	C	C	C
4	D	D	D	D

day \ order	1	2	3	4
1	A	B	C	D
2	A	B	C	D
3	A	B	C	D
4	A	B	C	D

CAS PITJOR: 3 Factors – pH (controlat)
- Dia
- Ordre

Com més desordenat sigui millor!! Millor forma de minimitzar un factor no controlat, és evitar una estructura repetida, ordenada pq sinó l'error aleatori es pot magnificar molt.

Proposta inicial: Disseny INCORRECTE: massa sensible a factors no controlats: els errors aleatoris no es minimitzen. No és el mateix comparar: s2 factor controlat ≈ 40 i s2aleatoria ≈ 30 es dispara

Alternatives millors:

Disseny ALEATORI PUR: ANOVA d'un factor del factor controlat (pH) i/o dels factors no controlats (dia/ordre). No és perfecte però està millor. No tots els dies faig de tot. Puc avaluar el que m'interessa l'efecte del factor controlat. Si sospito que un no controlat també afecta. ANOVA per separat!!

day \ order	1	2	3	4
1	A	B	B	D
2	C	A	B	D
3	D	C	C	A
4	A	B	D	C

day \ order	1	2	3	4
1	A	B	C	D
2	A	C	B	D
3	D	A	C	B
4	B	A	D	C

Disseny en bloc aleatori: m'asseguro poder mirar més d'un factor alhora. ANOVA d'un factor: el controlat pH, i/o ANOVA de dos factors (ex: pH i dia).
Blocs definits per dia. Factor controlat i del no controlat (només un) segons la definició prèvia que he fet. No es fa pH i ordre perquè em falten dades.
Si amb informació prèvia l'ordre no influeix millor disseny en bloc aleatori.

day \ order	1	2	3	4
1	A	B	C	D
2	B	C	D	A
3	C	D	A	B
4	D	A	B	C

Disseny Latin-quadrat: tots els nivells dels factors controlats apareixen en tots els nivells dels factors no controlats (distribució uniforme de les possibles interaccions) → ANOVA d'un factor del factor controlat i/o ANOVA de dos factors (ex. **pH i dia, pH i ordre). És un doble bloc: tant per ordre com per dia tinc tots els nivells del factor controlat per tots els nivells de factors no controlat.** MILLOR perquè tinc més info.

Disseny d'experiments amb F FACTORS CONTROLATS: screening de factors

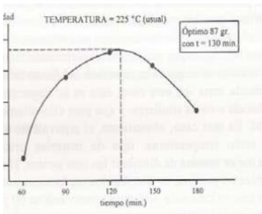
Screening de factors = CRIBATGE: té com a objectiu elucidar, amb un baix nombre d'experiments, quins factors afecten i interaccions.

- **Disseny seqüencial (Factor per factor**): canvis en la resposta són examinats quan variem els nivells d'un factor, mentre que la resta de factors es mantenen constants. Pocs experiments però molt afectat per la interacció entre factors.

  Requereix un nombre més baix d'experiments que el disseny factorial, però pot portar a resultats incorrectes si hi ha **interacció** entre factors no serveix per res (el nivell d'optimització d'un factor dependrà dels resultats dels experiments previs). Ex: quantitat de producte sintetitzat al màxim en el mínim temps i temperatura:

Perquè la info que obtinc no em serveix si interacciona amb l'altre. Si tinc informació prèvia, el tractament de les dades és més senzill i el nombre d'experiments pot ser inferior.

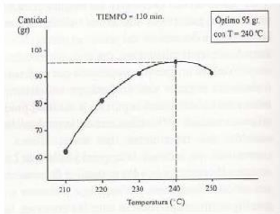
Fixo T, estudio 5 temps. Obtinc el màxim.
Fixo temps, estudio 5 T. Obtinc el màxim de 95g que és diferent..



Temperature: 225 °C

Time: 60, 90, 120, 150, 180 (min)

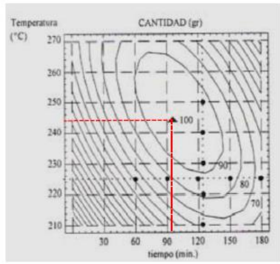
Optimized time: 130 min



Time: 130 min

Temperature: 210, 220, 230, 240, 250 (°C)

Optimized temperature: 240 °C



Contour plots: amount produced wrt time and temperature

There is interaction between the two factors

Quan faig un disseny combinat: diagrama de contorn, isoquantitats → puc produir més de cent grams, a 95 min a una T major.

DISSENY FACTORIAL (simultani): canvis en la resposta són examinats quan combinem diferents nivells (n) de tots els factors considerats (f).

→ **Disseny factorial complet (n^f experiments):**

- Els experiments reflecteixen totes les possibles combinacions de factors i nivells: només es fan tests d'uns quants factors i nivells.
- Disseny factoria a **2 o 3 nivells (2^f / 3^f) de f factors** són els que es fan servir habitualment.
- Dissenys de dos nivells **(- ; +)** són especialment útils en estudis preliminars, quan busquem tendències principals, basats en una interpretació simple dels resultats.

→ **Disseny factorial fraccionat (n^{f-r} experiments)**: quan vull augmentar el nombre de factors però no vull que es dispari nombre d'experiments.

- Un nou factor pot ser important.
- Un grup d'experiments representatius per a uns certs factors f>3 és seleccionat.
- Els experiments avaluen el efecte de factors principals.

Disseny factorial complet (n^f experiments): 2 factors, 2 nivells: 2^2 experiments

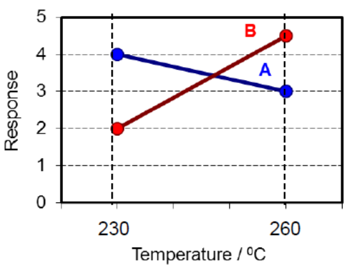
Ex: efecte de dos catalitzadors (A i B) a dues temperatures diferents (230 i 260ºC). 2 nivells + i -

Example: Effect of two catalysts (C; A and B) at two temperatures (T; 230 and 260 °C)

▪ **Coding of the levels of the factors (contrast coefficients): + / -**

	C_- A	C_+ B
T_- 230 °C	4	2
T_+ 260 °C	3	4.5

→ **The means of the (one-order) linear functions are calculated**



Effect of the catalyst factor $\dfrac{\Sigma C_+ - \Sigma C_-}{2} \quad \dfrac{(2 + 4.5) - (4 + 3)}{2} = -0.25$

Effect of the temperature factor $\dfrac{\Sigma T_+ - \Sigma T_-}{2} \quad \dfrac{(3 + 4.5) - (4 + 2)}{2} = 0.75$

Effect of the interaction $\dfrac{(T_+ C_+ + T_- C_-) - (T_+ C_- + T_- C_+)}{2} \quad \dfrac{(4.5 + 4) - (3 + 2)}{2} = 1.75$

$$F_{calc} = \frac{S^2_{effect}}{S^2_{random}}$$

d.o.f. overall = N - 1

d.o.f. effect = n_{level} - 1

d.o.f. random = (N - 1) - ∑ (levels - 1)

3 factors, 2 nivells: 2^3 experiments

Ex: optimització del rendiment Y d'una reacció variant 3 factors FTAB 1 CUA!!!

Example: Optimization of the yield Y of a reaction from the variation of three factors

Codifiquem les variables segons si són + o – i mirem la seva interacció.

Factor		level -	level +
T	Temperature (°C)	160	180
C	Concentration (%)	20	40
K	Catalyst	A	B

3 factors (T, C, K), 2 levels

2^3 = 8 experiments

In this case, two replicates per type of experiment (each result is the resulting mean value)

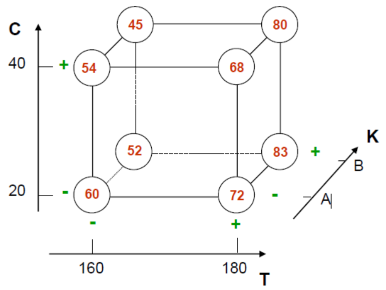
Original variables → **Coded variables: contrast coefficients**

	T	C	K	Yield (Y)
1	160	20	A	60
2	180	20	A	72
3	160	40	A	54
4	180	40	A	68
5	160	20	B	52
6	180	20	B	83
7	160	40	B	45
8	180	40	B	80

▪ Experiments should be carried out randomly

	T	C	K	Yield (Y)
1	-	-	-	60
2	+	-	-	72
3	-	+	-	54
4	+	+	-	68
5	-	-	+	52
6	+	-	+	83
7	-	+	+	45
8	+	+	+	80

Representació gràfica:



Coeficients de contrast per a factors i interaccions.

	T	C	K	TC	TK	CK	TCK	Yield (Y)
1	-	-	-	+	+	+	-	60
2	+	-	-	-	-	+	+	72
3	-	+	-	-	+	-	+	54
4	+	+	-	+	-	-	-	68
5	-	-	+	+	-	-	+	52
6	+	-	+	-	+	-	-	83
7	-	+	+	-	-	+	-	45
8	+	+	+	+	+	+	+	80

Efecte dels factors i de les seves interaccions:

Effects of factors and their interactions

Effect of T factor $= \frac{\sum T_+ - \sum T_-}{4} = \frac{(68 + 80 + 72 + 83) - (54 + 45 + 60 + 52)}{4} = 23$

Effect of C factor $= \frac{\sum C_+ - \sum C_-}{4} = \frac{(54 + 45 + 68 + 80) - (60 + 52 + 72 + 83)}{4} = -5$

Effect of K factor $= \frac{\sum K_+ - \sum K_-}{4} = \frac{(45 + 80 + 52 + 83) - (54 + 68 + 60 + 72)}{4} = 1.5$

Avaluació qualitativa a partir de la comparació dels valors dels efectes.

$$TC = \frac{60-72-54+68+52-83-45+80}{4} = 1.5$$

$$TK = \frac{60-72+54-68-52+83-45+80}{4} = 10$$

$$CK = \frac{60+72-54-68-52-83+45+80}{4} = 0$$

$$TCK = \frac{-60+72+54-68+52-83-45+80}{4} = 0.5$$

FER TAULA D'EFECTES

Avaluació de la significació dels factors i de les interaccions:

$$t = \frac{efecte\ del\ factor/interacció}{S_{aleatori}}$$

H_0: effect = 0 $t_{calc} < t_{crit}$

H_1: effect ≠ 0 $t_{calc} > t_{crit}$

$S_{aleatori}$: dels replicats de tots els tipus d'experiments o del punt central. ***Es pot estimar fent la variància de replicats i la mitjana de les variàncies.***

Dof: Nexp (nreplicats-1). Ex: 3 replicats del punt central: 1 (3-1) =2

Ex: 2 replicats de tots els experiments d'un disseny 2^3: 8 (2-1) =8

Ús de l'ANOVA: hi ha dificultats a l'hora d'aplicar ANOVA per a més de dos factors → usar algoritme de Yates.

$$SS_{effect} = \frac{N \cdot (effect)^2}{n_{design} / n_{levels}}$$

N = overall number of experiments (including replicates)

n_{design} = number of type of experiments

n_{level} = number of levels

S^2 =ssefecte/dof

algoritme de Yates (ANOVA) és més restrictiu que el test t

Disseny factorial fraccionat ($n^{f\text{-}r}$ experiments):

$n^{f\text{-}r}$: uns quants factors addicionals poden ser parcialment introduïts si s'han d'examinar més factors (Amb un nombre menor d'experiments que un disseny factorial complet).

	T	C	K	TC	TK	CK	D TCK
1	-	-	-	+	+	+	-
2	+	-	-	-	-	+	+
3	-	+	-	-	+	-	+
4	+	+	-	+	-	-	-
5	-	-	+	+	-	-	+
6	+	-	+	-	+	-	-
7	-	+	+	-	-	+	-
8	+	+	+	+	+	+	+

Si unes quantes **interaccions de major ordre s'assumeixen com a negligibles**, podem introduir nous factors en el experiment.
Ex: **2^{4-1}** experiments (encara que fem test de 4 factors!)
Els codis de la interacció de 3-factors s'assignen a un nou (quart) factor: D factor en lloc d'interacció TCK . es realitzen 8 experiments en comptes de 16. És una aproximació útil per a avaluar un major nombre de factor, encara que està subjecte a l'aparició d'efectes confusos (contribució del **factor i de la interacció de major ordre**).

*Perdre informació d'interaccions de grau superior. → **$N^{f\text{-}r}$:** s'introdueixen factors addicionals (r)*
Ex: 2^{4-1} factors T, C (concentració), K (catalitzador) i TKD → D : interacció de grau superior. He de jugar amb els signes (multiplicar) per deduir-ho. D: és el 4t factor que addiciono. Coeficient de contrast, multiplico i veig si dóna + o -.

FACTORIAL FRACCIONAT PASSOS:

1. *Decidir quin NO faré complet (D)*
2. *T,C,K veure tots contra tots, són complets, totes les combinacions possibles.*
3. *El que fracciono (D) és T*C*K: **combinació de contrastos que faig fraccionat;** no totes veuen tot contra tot.*
4. *+ o – és el codi resultat de la interacció de grau superior dels factors de grau complert.*

*Si fes un disseny factorial complet: 2 nivells^{4factors} = 16 experiments *2/3replicats (32,48)*
*No val la pena perquè vull fer un **cribratge, screening** : millor fer 1 factor parcial. Tal que: 2^{4-1} = 8 experiments *2/3 (16,24). Sense rèpliques i amb rèpliques redueixo a la meitat el nº d'experiments.*
D és parcial. Trio els nivells del signe del grau d'interacció.
M'interessa fer aquest disseny quan tinc molts experiments i vull reduir feina experimental.

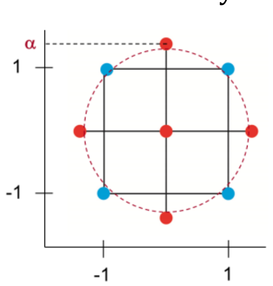
- → *Si el resultat de D és completament inequívoc, tot OK. Si les conclusions, resultats pel D no són clares, no em fio massa.*
- → *Si T,C.K afecten molt i D afecta poc, és possible que vegi un fals positiu perquè veig l'efecte de la interacció dels que sí que afecten.*

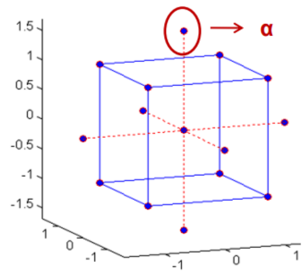
DISSENYS EXPERIMENTALS DIRIGITS A L'OPTIMITZACIÓ

Ara que ja tenim 1 o 2 factors que afecten hem d'**optimitzar**: trobar el model matemàtic (lineal, si és possible) que ajusti el model.

- Només tenen en compte els factors més rellevants i significatius (de la informació del cribratge previ) amb un major nombre de nivells (> 2).
- Els experiments permeten modelar la resposta en les bases de correlacions entre la resposta i factors amb equacions lineals.

Disseny central compost circumscrit: es basa en mesclar el disseny factorial complert + disseny d'estrella.
Disseny d'estrella: inclou punts més enllà del domini experimental que jo havia considerat amb el disseny factorial complert. A més inclourà el punt central.
Incremento el disseny amb **3 nivells nous**:





1. Punt central **N**

Punts α: punts més enllà del disseny experimental del complet:

2. Punts per sobre
3. Punts per sota

Disseny factorial complet: 2^2 = 4 nivells 4 punts.

Disseny d'estrella: **α** = $(2^f)^{1/4} = 2^{1/2} = 1{,}4$ **α depèn del factor**

Circumscrita: dins Central: distància **α-N** és igual a=b

Si faig 3 factors tinc un cub que va de -1,1 i – **α,+ α**

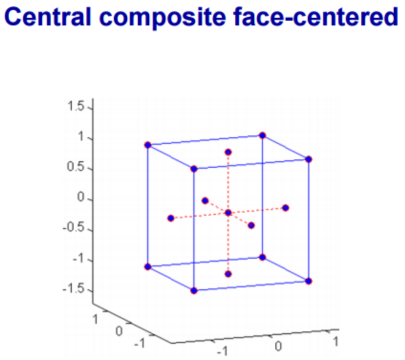
Faré experiments a 5 nivells: **- α,-1,0,+1,+ α** **Nº d'experiments: 2^f + 2f +N**

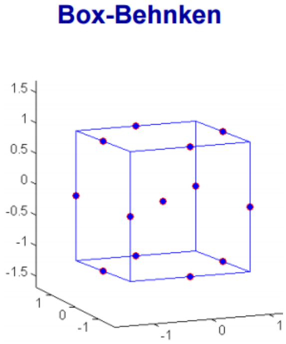
2f: fa referència a les 2 α

N: punt central: millor fer rèpliques del punt central

Ex: 3 factors / 5 nivells → 15 experiments. Si fes un disseny factorial complet tindria 5^3 =125 experiments.

Altres dissenys: van més enllà de 2 nivells. No és tothom contra tothom, es veuen les que tenen més pes estadístic: **punts clau**. El rang experimental inicial es manté. 3 nivells/f





Disseny central compost centrat a la cara: tenim punts extres que no són **α** sinó que són la d a les cares del cub.

Box Behnken: disseny central compost als vèrtexs.

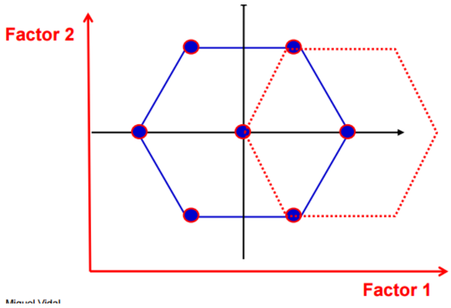
- Tres nivells per factor
- El rang inicial de l'experiment es manté.

<u>Disseny Doehlert:</u>

Es fa de 2 factors normalment, també es podria de 3 però és més difícil.

Case of 2 factors

Factor 1	Factor 2
0	0
1	0
0,5	0,866
-1	0
-0,5	-0,866
0,5	-0,866
-0,5	0,866
5 levels	**3 levels**



Si tinc 2 factors: tinc una superfície que es pot moure per trobar el disseny òptim. Es pot moure per l'espai fins trobar uns valors de tot el domini experimental que donin els valors òptims.

Factor 1: -1,-0,5,0,+0,5,+1 → 5 nivells

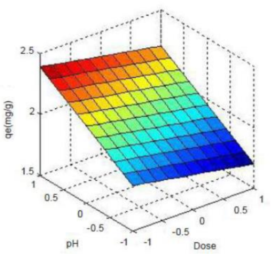
Factor 2: -0,866,0,0,866 → 3 nivells

Explorar zones veïnes del domini experimental que s'han de **SOLAPAR**. Agafar 4 punts que ja conec (-feina) i 3 punts nous. Mirar dreta, esquerra, a dalt i a baix.

- Nº d'experiments: **f^2 + f + N**
- Tots els experiments són equidistants al punt central: N.

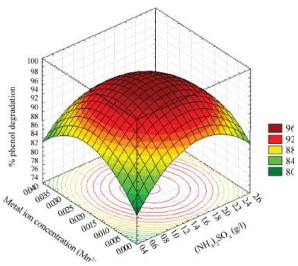
El nº d'experiments nous que faig és limitat perquè alguns ja els he fet.

Modelling the response surface



- First-order linear models require a minimum of 2 levels of each factor
  - Without interactions:
    $R = a_0 + a_1f_1 + a_2f_2 + ...$
  - With interactions:
    $R = a_0 + a_1f_1 + a_2f_2 + a_3f_1f_2 ...$

Two-order linear models require more than 2 levels of each factor to fit the curvature of the surface
- Without interactions:
  $R = a_0 + a_1f_1^2 + a_2f_2^2 + a_3f_1 + a_4f_2 ...$
- With interactions:
  $R = a_0 + a_1f_1^2 + a_2f_2^2 + a_3f_1 + a_4f_2 + a_5f_1f_2 ...$



<u>SUPERFÍCIE DE RESPOSTA:</u>

Una vegada tinc els factors que afecten la resposta confirmats, hem de buscar models matemàtics.

- **Predir els valors de resposta** de cada nivell de factor
- Visualitzar els **resultats en gràfics de superfície de resposta**, al llarg de tot el **<u>domini</u>** experimental examinat.
- Seleccionar els nivells de factors que porten a una **<u>resposta òptim</u>**a, basant-nos en el criteri preestablert.

Sé els factors que afecten després de l'*screening*, examino la correlació entre variables: modelar i predir la variable en funció dels factors que afecten.

El model ha de predir els valors de la resposta amb un ús químic, econòmic, ... valors de factors que fan òptima la resposta.

Correlació bona és de primer ordre: molts punts i molts factors, tinc superfície plana (2 nivells).

Correlació multivariant entre variables: de grau 2/3 comportament no pla: calen més nivells per tal de predir la curvatura. buscar resposta òptima, faig derivades parcials i trobo màxim, mínim, punts diana..