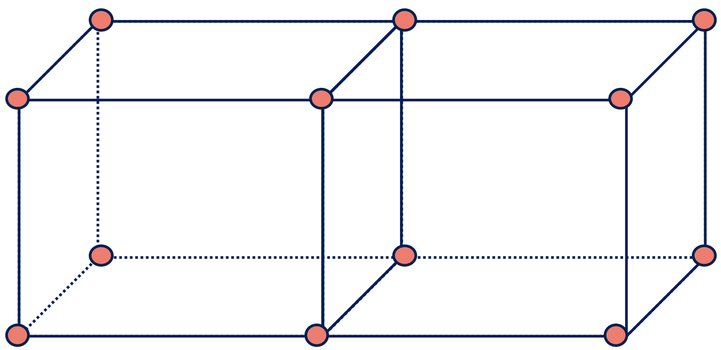
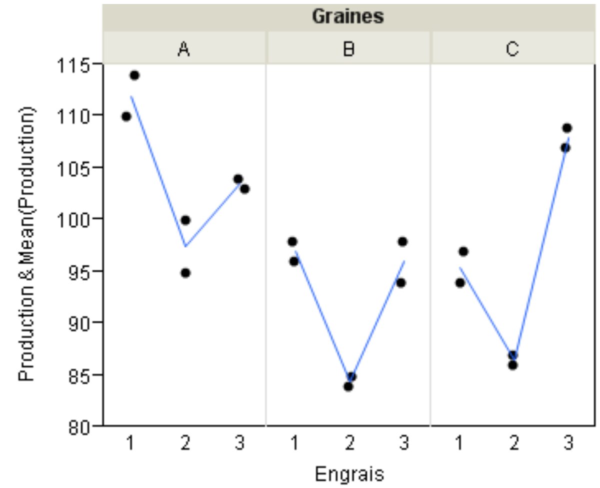
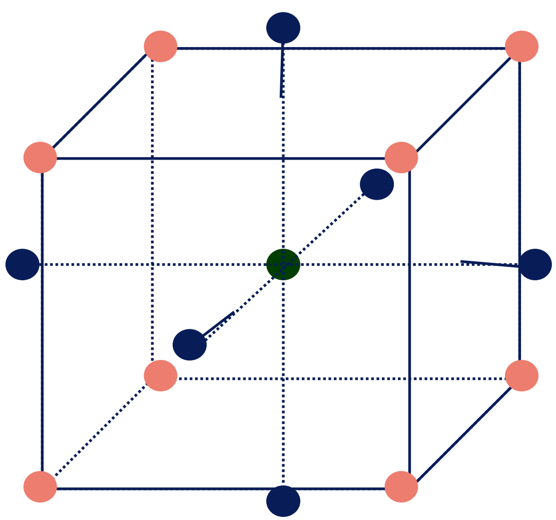
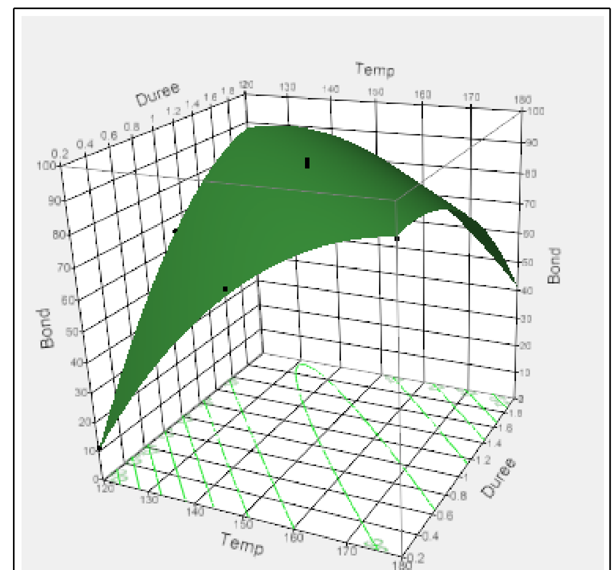
Plans d’expérience multi-facteurs









Objectifs/plan du chapitre

Objectifs du chapitre

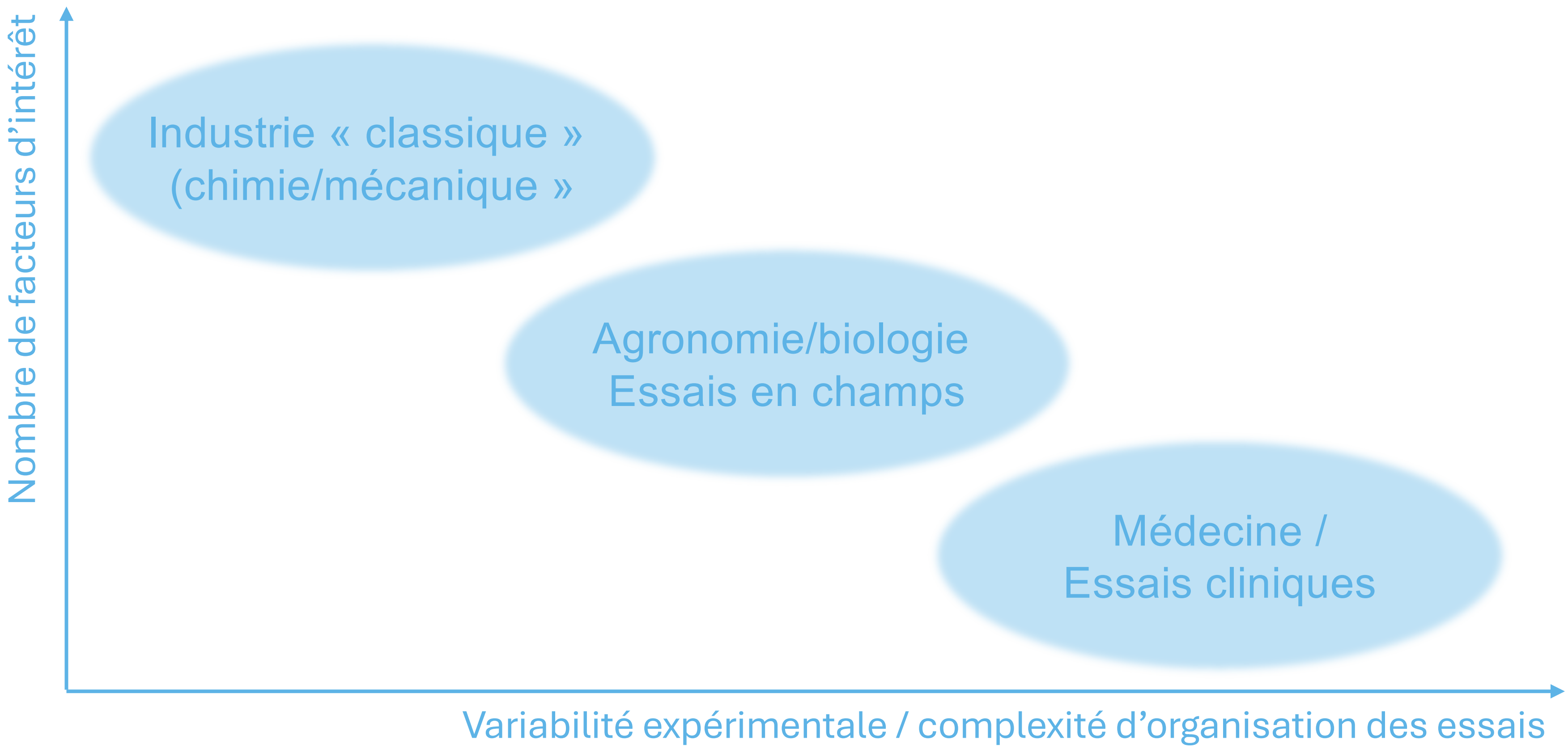
- Présenter des plans d'expérience pour le traitement de problèmes multi-facteurs avec applications industrielles/labo.
- Rester dans un cadre de facteurs fixes étudiés dans un milieu expérimental stable avec des unités expérimentales semblables et indépendantes.
- Ce cours ≃ résumé/condensé du cours LSTAT2320 (Plans d'expérience)

Plan

- Remise en contexte
- Exemples traités dans ce chapitre
- Rappel de vocabulaire et des composantes essentielles d'un design d'expérience
- Classes de plans multi-facteurs

Origine de la planification expérimentale

Les plans d'expériences utilisés en science expérimentale émanent de 3 domaines où les plans d'expériences se sont développés historiquement.



Etapes d'une recherche expérimentale

La question	Enoncé de la question à laquelle le chercheur désire répondre.
Formalisation	Etude détaillée et structurée du problème et des informations disponibles (biblio, matériel et méthode)
Planification/design	Définition des traitements testés et du dispositif expérimental
Expérimentation/observation	Exécution des essais du plan ou observation et collecte des données
Analyse des données	Analyse exploratoire et modélisation statistique des résultats.
Validation	Exécution éventuelle de quelques essais supplémentaires pour valider les résultats de l'analyse.
Conclusion	Conclusion de l'étude, écriture d'un rapport, initiation des études futures.

Plan du chapitre

Exemple 1 : Production de maïs (version simple)

Contexte (cours LBIRA2110)

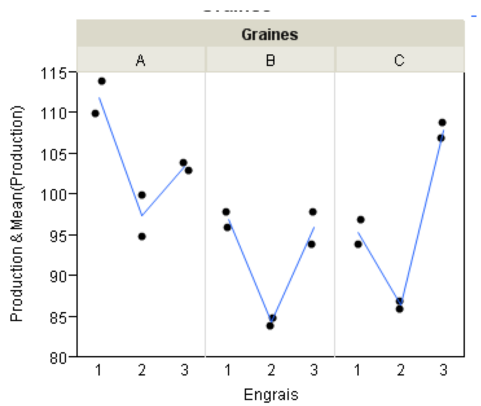
On s'intéresse à la production de parcelles de maïs pour 3 variétés de maïs et 3 types d'engrais.

18 parcelles ont été cultivées et 9 traitements appliquées aléatoirement à ces 18 parcelles (2 répétitions par traitement).

On ne dit rien sur ces parcelles et sur la période de l'étude → on suppose implicitement que les parcelles sont indépendantes, assez homogènes et représentent bien la population des parcelles à laquelle les résultats de l'étude visent d'être généralisés.

Plan et données : factoriel complet 3^2 (= 9) répété deux fois (= 18 expériences)

		Engrais	
Graines	1	2	3
A	110	95	103
	114	100	104
B	96	84	94
	98	85	98
C	94	86	107
	97	87	109



Exemple 1 : Résultats

Questions

- Y a-t-il des différences entre engrais ? Entre les variétés ?
- Quel est le meilleur engrais ? Quelle est la meilleure variété ?
- Le meilleur engrais diffère-t-il selon la variété utilisée ?

Plan et modèle lié

- Plan : factoriel complet 3^2 répété deux fois
- Modèle : Anova 2 croisée pour facteurs (catégoriels) fixes

Résultats statistiques principaux

Analysis of Variance Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Graines	2	432.3	216.2	50.53	1.278e-05
Engrais	2	628	314	73.4	2.676e-06
Graines:Engrais	4	243.7	60.92	14.24	0.0006255
Residuals	9	38.5	4.278	NA	NA

```
$contrasts
 contrast estimate   SE df t.ratio p.value
 A - B       11.83 1.19  9   9.910  <.0001
 A - C        7.67 1.19  9   6.420  0.0003
 B - C       -4.17 1.19  9  -3.489  0.0170

$contrasts
 contrast estimate   SE df t.ratio p.value
 1 - 2          12 1.19  9  10.049 <.0001
 1 - 3          -1 1.19  9  -0.837 0.6905
 2 - 3         -13 1.19  9 -10.887 <.0001
```

Exemple 2 : Production de maïs (version avancée)

Contexte et but de l'étude

- Etude de l'impact de la variété de maïs et de l'engrais utilisé sur le rendement d'une culture.
- On s'intéresse à l'évolution de la taille des plants en fonction du temps et au rendement final d'une parcelle (nb de kg de mais produit).
- 3 variétés de graines et 3 variété d'engrais sont visés.

En pratique

- On dispose de 2 champs (dans des lieux différents)
- On divise chaque champ en 18 parcelles
- En mesure tous les mois les hauteurs de 4 plants par parcelle.
- En fin de saison, le rendement de chaque parcelle sera mesuré
- Les expériences sont réalisées sur une seule année

Exemple 2 : facteurs/unité expérimentales/dispositif

Facteurs d'intérêt et traitements

- Variété (3 niveaux) et Engrais (3 niveaux)
- 9 traitements : plan croisé ou factoriel complet à 2 facteurs 3^2

Unité expérimentales , blocs, répétitions

- 2 champs, 18 parcelles par champs
- 2 x 18 = 36 parcelles → 2x2 = 4 répétitions possibles par traitement
- 4 plants mesurés par parcelle mesuré chacun 2x : 32 mesures par traitement.

Répartition des traitements entre les unités expérimentales

- Chaque traitement est testé deux fois par champ
- Les 2 x 9 traitements sont
  - soit alloués aléatoirement aux parcelles des champs
  - soit organisés dans chaque champ selon un dispositif spécial pour éviter une confusion entre de la position de la parcelle et les facteurs d'intérêt.

Exemple 2: répartition des traitements par champs

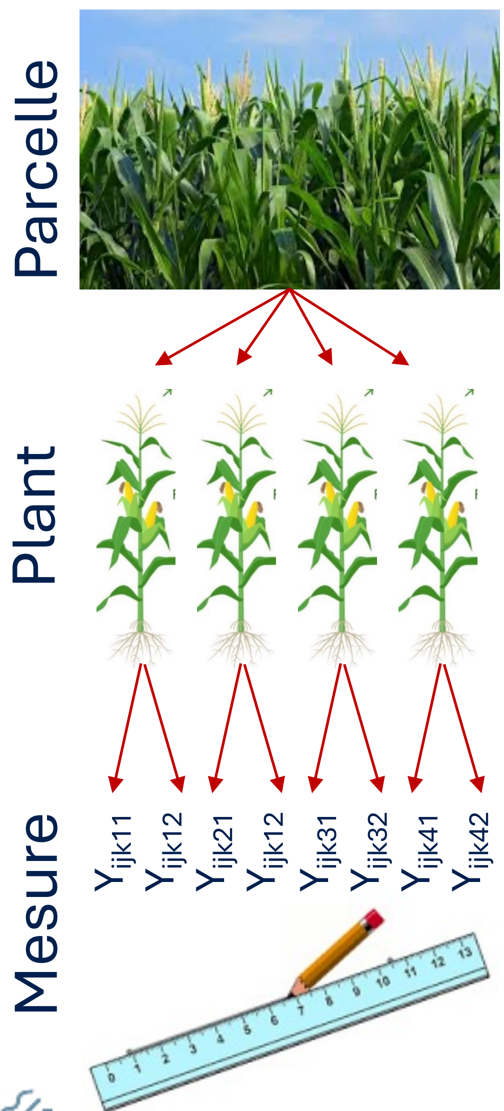
Randomisation des essais dans chaque champ

Champ 1				**Champ 2**		
V3-E2	V2-E1	V1-E2		V1-E1	V2-E3	V1-E3
V1-E3	V1-E1	V2-E3		V2-E3	V2-E2	V1-E1
V3-E1	V3-E3	V2-E2		V2-E1	V1-E3	V3-E3
V1-E3	V3-E3	V2-E1		V1-E2	V3-E2	V3-E1
V3-E2	V3-E1	V1-E2		V3-E2	V3-E3	V3-E1
V2-E2	V1-E1	V2-E3		V2-E2	V1-E2	V2-E1

Organisation de chaque champ selon une sorte de carré-latin
(champ est divisé en deux blocs et chaque variété/engrais apparait une fois par ligne et une fois par colonne dans chaque bloc)

Champ 1				**Champ 2**		
V1-E1	V2-E2	V3-E3		V2-E1	V3-E2	V1-E3
V2-E3	V3-E1	V1-E2		V3-E3	V1-E1	V2-E2
V3-E2	V1-E3	V2-E1		V1-E2	V2-E3	V3-E1
V2-E1	V3-E2	V1-E3		V3-E1	V1-E2	V2-E3
V3-E3	V1-E1	V2-E2		V1-E3	V2-E1	V3-E2
V1-E2	V2-E3	V3-E1		V2-E2	V3-E3	V1-E1

Exemple 2 : modèle (pour un seul champ)



Modèle complet mixte hiérarchisé

- Y_{ijklm} : hauteur du plant pour les graines *i*, l'engrais *j*, la parcelle *k*, le plant *l*, la mesure *m*
- Modèle :

$$y_{ijklm} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \delta_{ijk} + \nu_{ijkl} + \epsilon_{ijklm}$$

  - graine et engrais sont des **facteurs fixes**.
  - parcelle (*k*), plant (*l*) et mesure (*m*) : **facteurs aléatoires**
  - **Hypothèses** sur les facteurs **aléatoires** : $\delta_{ijk} \sim iN(0, \sigma^2_{Par}), \nu_{ijkl} \sim iN(0, \sigma^2_{Pl}), \epsilon_{ijklm} \sim iN(0, \sigma^2_M)$

Modèle "pratique" et simple

- $\bar{Y}_{ijk}$= Moyenne des 8 mesures de la parcelle *k*
- Modèle

$$\bar{y}_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk} \text{ avec } \epsilon_{ijk} \sim iN(0, \sigma^2_T)$$

In fine les résultats seront comparables

Exemple 2 : comparaison de 3 modèles

Simulation de plan balancé avec 288 données : 3 x 3 x 4 x 4 x 2

1 champ, 3 types de graines, 3 engrais, 4 parcelles/GxE, 4 plants/PxGxE et 2 mesures

Anova 2 (faux) sur les Y_{ijklm}

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	8	467.06264	58.3828	33.1104
Error	279	491.95471	1.7633	Prob > F
C. Total	287	959.01735		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	14.000417	0.078246	178.93	<.0001*
graine[1]	-0.910625	0.110657	-8.23	<.0001*
graine[2]	0.0067708	0.110657	0.06	0.9513
engrais[1]	-1.083333	0.110657	-9.79	<.0001*
engrais[2]	-0.029479	0.110657	-0.27	0.7901
graine[1]*engrais[1]	0.4201042	0.156493	2.68	0.0077*
graine[1]*engrais[2]	0.2240625	0.156493	1.43	0.1533
graine[2]*engrais[1]	-0.767292	0.156493	-4.90	<.0001*
graine[2]*engrais[2]	-0.195521	0.156493	-1.25	0.2126

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
graine	2	2	158.03866	44.8139	<.0001*
engrais	2	2	231.63185	65.6822	<.0001*
graine*engrais	4	4	77.39213	10.9728	<.0001*

Modèle mixte sur les Y_{ijklm}

REML Variance Component Estimates

Random Effect	Var Ratio	Var Component	Std Error
parcelle	1045.4955	2.18465	0.5995736
plant[parcelle]	34.54049	0.0721752	0.0099647
Residual		0.0020896	0.0002463
Total		2.2589148	0.599615

Parameter Estimates

Term	Estimate	Std Error	DFDen	t Ratio	Prob>|t|
Intercept	14.000417	0.247373	27	56.60	<.0001*
graine[1]	-0.910625	0.349838	27	-2.60	0.0148*
graine[2]	0.0067708	0.349838	27	0.02	0.9847
engrais[1]	-1.083333	0.349838	27	-3.10	0.0045*
engrais[2]	-0.029479	0.349838	27	-0.08	0.9335
graine[1]*engrais[1]	0.4201042	0.494745	27	0.85	0.4033
graine[1]*engrais[2]	0.2240625	0.494745	27	0.45	0.6543
graine[2]*engrais[1]	-0.767292	0.494745	27	-1.55	0.1326
graine[2]*engrais[2]	-0.195521	0.494745	27	-0.40	0.6958

Fixed Effect Tests

Source	Nparm	DF	DFDen	F Ratio	Prob > F
graine	2	2	27	4.4837	0.0208*
engrais	2	2	27	6.5716	0.0047*
graine*engrais	4	4	27	1.0978	0.3777

Anova 2 sur les moyennes $\overline{Y}_{ijk}$

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	8	58.38283	7.29785	3.3128
Error	27	59.47979	2.20296	Prob > F
C. Total	35	117.86262		0.0091*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	14.000417	0.247373	56.60	<.0001*
graine[1]	-0.910625	0.349838	-2.60	0.0148*
graine[2]	0.0067708	0.349838	0.02	0.9847
engrais[1]	-1.083333	0.349838	-3.10	0.0045*
engrais[2]	-0.029479	0.349838	-0.08	0.9335
graine[1]*engrais[1]	0.4201042	0.494745	0.85	0.4033
graine[1]*engrais[2]	0.2240625	0.494745	0.45	0.6543
graine[2]*engrais[1]	-0.767292	0.494745	-1.55	0.1326
graine[2]*engrais[2]	-0.195521	0.494745	-0.40	0.6958

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
graine	2	2	19.754833	4.4837	0.0208*
engrais	2	2	28.953982	6.5716	0.0047*
graine*engrais	4	4	9.674016	1.0978	0.3777

Conclusions:

- Anova 2 simple : trop de degrés de liberté, effets trop ou faussement significatifs,
- Modèles mixte et Anova 2 sur moyennes équivalents
- mais le modèle mixte estime les composantes de variance + traite les cas non balancés

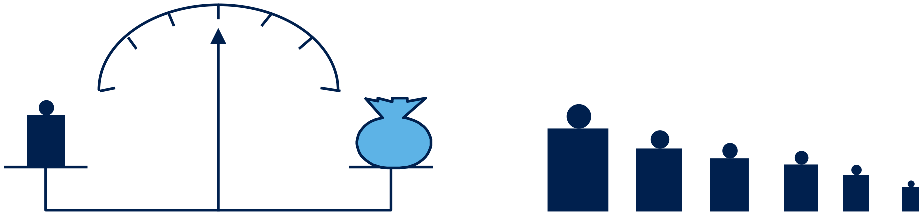
Exemple 3 : le problème des pesées

La question

Déterminer les poids de trois objets A, B et C en quatre pesées et avec un maximum de précision

Le matériel expérimental

Une balance à deux plateaux à équilibrer avec des poids



Budget : 4 essais

Exemple 4 : étude d'un procédé de fermentation

Le contexte

Les fermenteurs sont utilisés couramment dans la production de vaccins mais le bon choix de leurs paramètres de réglage est primordial pour assurer une production suffisante, de qualité et stable dans le temps.

But de l'étude

Rechercher quels facteurs de fermentation ont le plus d'effet sur la biomasse finale obtenue au terme d'une fermentation de 24h ainsi que le sens de ces effets.

Budget : Nombre maximum d'essais : 16 (=16 jours)

Facteurs d'intérêt : 10 facteurs quantitatifs fixes

Phase de croissance (0 à 18h) : X1 - Température de croissance, X2 – Surpression, X3 - Débit d'air , X4 - Vitesse d'agitation X5 – PH, X6 – Débit d'alimentation , X7 – Débit d'antimousse (>16h)

Phase d'induction (18-24h) : X8 – Moment d'induction, X9 – Température d'induction, X10 – Débit d'antimousse

Exemple 5 : Optimisation d’un processus industriel

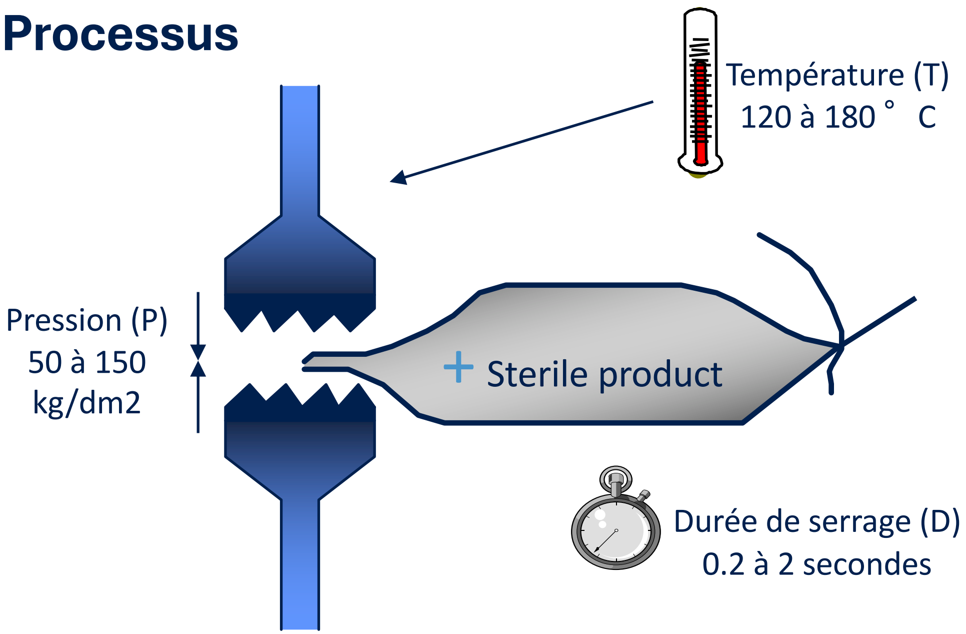
Contexte

Etude d’un processus industriel dont le but est de fermer, par soudure, des poches de produit (ex: produits stériles pharmaceutiques ou produits alimentaires)

But

Trouver les valeurs de la **pression**, de la **température** et de la **durée** de serrage qui optimisent la résistance de la soudure (qui doit être suffisamment résistante mais pas trop pour faciliter l’ouverture) et la qualité visuelle de la soudure

Processus



Exemple 5 : Les éléments

Les facteurs

- 3 facteurs quantitatifs fixes (temperature, pression, durée)
- Domaines : intervalles continus

Les unités expérimentales

- Les sachets.
- Pour une application donnée ils sont a priori considérés tous +/- identiques
- On peut considérer que les essais (et unités expérimentales) sont indépendants

Objectifs de l’étude

- Objectif premier : optimiser le procédé
- Objectif secondaire : comprendre l’effet des facteurs

Essais, budget, plan, modèle

- Budget : 20 essais. Un essai = un “traitement”
- Quels essais/plan choisir ?
- En quoi un modèle est-il utile ici ?

Plan du chapitre

- Remise en contexte
- Exemples traités dans ce chapitre
- Rappel de vocabulaire et des composantes essentielles d’un design d’expérience
- Description des problèmes traités dans ce chapitre
- Classes de plans multi-facteurs

Rappel de vocabulaire

Facteur – critère

Variable d'intérêt de l'étude dont on veut étudier l'effet sur une « réponse ».

Ex: Facteur catégoriel : Type d'engrais, facteur quantitatif : température etc…

Deux types de facteurs dans les modèles : fixe ou aléatoire

Niveaux/domaine

Valeurs que peut prendre un facteur

Ex: A, B ou C pour la variété ou [120,180] pour la température

Unité expérimentale

Objet qui « subit » l'expérience, le traitement: parcelle, animal, juge…

Traitement

Combinaison de niveaux des facteurs appliquée à une unité expérimentale

Ex: la variété A et l'engrais 2 sont appliqués à une parcelle

Réponse

Variable observée/mesurée au terme de l'expérience (peut être quantitative/catégorielle ou du comptage)

Ingrédients d'un design d'expérience

- **Les traitements** : quelles combinaisons des niveaux des facteurs d'intérêt seront testés ? = points supports du plan.

- **Répétitions** : Combien de fois chaque essai/traitement sera répété.

  Elles influencent la précision/puissance des résultats statistiques

- **Dispositif expérimental** :
  - Allocation des essais aux unités expérimentales
  - Choix de l'ordre des essais
  - Randomisation/organisation des essais en blocs

- **Modèle à estimer sur les résultats.** Adapté au design et destiné à répondre aux questions posées

Plan du chapitre

- Remise en contexte
- Exemples traités dans ce chapitre
- Rappel de vocabulaire et des composantes essentielles d'un design d'expérience
- Classes de plans multi-facteurs
    - Objectif général
    - Plans factoriels complets
    - Plans factoriels fractionnaires 2^{k-p} et pour le criblage de facteurs
    - Plans pour l'estimation de surfaces de réponse
        - factoriel complet, composite centré, Doehlert, Box Benken

Objectif général

Motivation

- Quand le nombre de facteurs dépasse 2 ou 3 et/ou que les facteurs sont quantitatifs, le choix de la liste des "traitements" (ou combinaisons de facteurs) à tester n'est pas automatique. Prendre toutes les combinaisons possibles peut être très coûteux.
- Dans le milieu industriel "Classique" (chimie, formulation de produits…) ou dans les essais de laboratoire les unités expérimentales sont plus homogènes, l'effet du milieu plus contrôlable et la répétabilité souvent bien meilleure MAIS le nombre de facteurs plus élevé.

Buts

Proposer des plans (et modèles liés) pour traiter des problèmes impliquant

- un nombre de facteurs + important (>2)
- des facteurs fixes quantitatifs et/ou catégoriels
- dans un but de modélisation mais aussi d'optimisation
- En considérant que la répétabilité est bonne, le milieu contrôlé et les unités semblables et indépendantes.

Matière de modélisation ⇒ modèles linéaires à effets fixes (cours LBIRA2110)

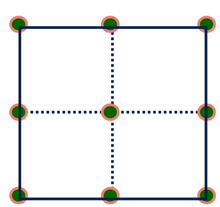
Plan du chapitre

- Remise en contexte
- Exemples traités dans ce chapitre
- Rappel de vocabulaire et des composantes essentielles d’un design d’expérience
- Classes de plans multi-facteurs
  - Objectif général
  - Plans factoriels complets
  - Plans factoriels fractionnaires 2^{k-p} et pour le criblage de facteurs
  - Plans pour l’estimation de surfaces de réponse
    - factoriel complet, composite centré, Doehlert, Box Benken

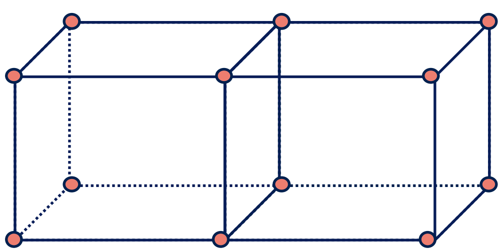
Plan factoriel complet : définition

Un plan factoriel complet $l_1 \times l_2 \times \ldots \times l_k$ à k facteurs est un plan formé de toutes les combinaisons des $l_1, l_2, \ldots l_k$ niveaux de k facteurs $X_1, X_2, \ldots X_k$.

- C'est le plan le plus utilisé
- Il s'utilise pour étudier l'effet de facteurs quantitatifs et/ou qualitatifs.
- Il sera répété *n* fois en fonction de la précision désirée pour l'étude.
- Il permet d'estimer un modèle avec les effets des facteurs et toutes les interactions possibles entre facteurs. Si le modèle contient des facteurs quantitatifs on peut ajouter des termes quadratiques X^2 ou cubiques X^3 en fonction du nombre de niveaux des facteurs.
- Ces plans sont toujours très bons mais coûteux quand le nombre de facteurs augmente.



Plan factoriel complet 3^2

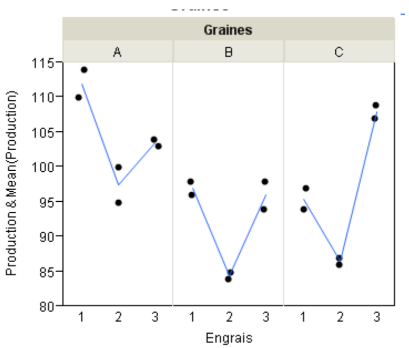


Plan factoriel complet 3 x 2 x 2

Exemple 1 : Etude de la production de maïs (simplifiée)

Plan factoriel complet à 2 facteurs balancé avec *n* = 2 répétitions par traitement.

Modèle d'ANOVA 2 croisé (facteurs principaux + interaction)



$$y_{ijk} = \mu_{ij} + \epsilon_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$$

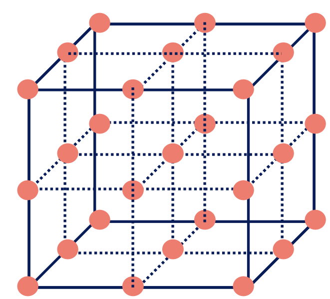
	GA	GB	GC
E1	n=2	n=2	n=2
E2	n=2	n=2	n=2
E3	n=2	n=2	n=2

Exemple 5 : Optimisation d'un processus de fermeture de sachets

Plan possible pour explorer le domaine : plan factoriel complet à 3 facteurs 3^3

(mais ce n'est pas le seul possible ⇒ affaire à suivre)

Modèle typique estimé sur ce type de plan



$$F = \beta_0 + \beta_1\ T + \beta_2\ P + \beta_3\ D + \beta_{12}\ T.P + \beta_{13}\ T.D + \beta_{23}\,PD + \beta_{11}\ T^2 + \beta_{22}P^2 + \beta_{33}\ D^2 + \varepsilon$$

Autres effets estimables : $X_iX_j^2, X_iX_jX_k \quad , X_i^2X_j^2, X_i^2X_j\ X_k , X_i^2\ X_j^2\ X_k \ , X_i^2\ X_j^2\ X_k^2$

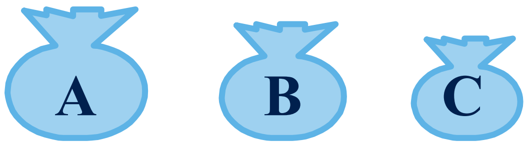
Plan du chapitre

- Remise en contexte
- Exemples traités dans ce chapitre
- Rappel de vocabulaire et des composantes essentielles d'un design d'expérience
- Classes de plans multi-facteurs
  - Objectif général
  - Plans factoriels complets
  - Plans factoriels fractionnaires 2^{k-p} et pour le criblage de facteurs
  - Plans pour l'estimation de surfaces de réponse
    - factoriel complet, composite centré, Doehlert, Box Benken

Exemple 3 : le problème des pesées

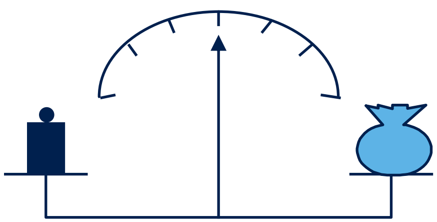
La question

Déterminer les poids de trois objets A, B et C en quatre pesées et avec un maximum de précision



Le matériel expérimental

Une balance à deux plateaux à équilibrer avec des poids



Budget : 4 essais

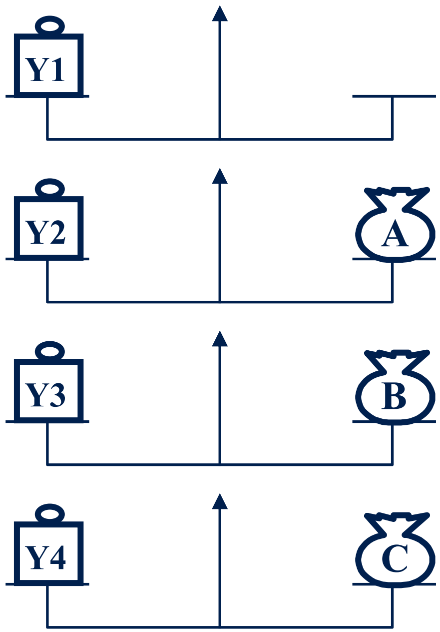
Hypothèses statistiques

- Les mesures sont entachées d'erreurs de variance constante, sont indépendantes et sont entachées d'un biais systématique
- Modèle

$$Y = \text{biais} + \mu + \varepsilon \qquad \text{Variance}(\varepsilon) = \sigma^2 \qquad \text{Covariance}(\varepsilon_i, \varepsilon_j) = 0$$

Stratégie 1

On pèse un objet à la fois



Matrice d'expérience

	A	B	C
E1	0	0	0
E2	1	0	0
E3	0	1	0
E4	0	0	1

0 : l'objet n'est pas sur la balance

1 : l'objet est sur le plateau de droite

-1 : l'objet est sur le plateau de gauche

À estimer :

μ_0 : Erreur systématique de la balance

μ_A, μ_B et μ_C : les poids des trois objets

Les estimateurs sont notés M_0, M_A, M_B, M_C

Estimation des poids

$M_0 = Y1$

$M_A = Y2 - Y1$

$M_B = Y3 - Y1$

$M_C = Y4 - Y1$

Quelle est la précision des mesures ?

Chaque pesée Y1, Y2, Y3, Y4 est entachée d'une erreur ε de variance constante $V(\varepsilon) = \sigma^2$.

Estimation des poids

$M_0 = Y1$

$M_A = Y2 - Y1$

$M_B = Y3 - Y1$

$M_C = Y4 - Y1$

Formules de combinaison de variances

$\mathbf{V}(X + Y) = \mathbf{V}(X) + \mathbf{V}(Y) + 2\,\mathbf{Cov}(X,Y)$

$\mathbf{V}(X - Y) = \mathbf{V}(X) + \mathbf{V}(Y) - 2\,\mathbf{Cov}(X,Y)$

$\mathbf{V}(aX + b) = a^2\,\mathbf{V}(X)$

Variances des estimateurs

$\mathbf{V}(M_0) = \sigma^2$

$\mathbf{V}(M_A) = \mathbf{V}(M_B) = \mathbf{V}(M_C) = 2\sigma^2$

Comment obtenir des mesures plus précises ?

Comment obtenir une meilleure précision ?

Solution possible

Répéter deux fois chaque pesée puis

Calculer la moyenne des résultats

$M_0 = (Y1 + Y1') / 2$

$M_A = (Y2 + Y2') / 2 - (Y1 + Y1') / 2$

$M_B = (Y3 + Y3') / 2 - (Y1 + Y1') / 2$

$M_C = (Y4 + Y4') / 2 - (Y1 + Y1') / 2$

Variance des estimateurs

$\mathbf{V}(M_0) = (\mathbf{V}(Y1) + \mathbf{V}(Y1')) / 4 = \sigma^2/2$

$\mathbf{V}(M_A) = \mathbf{V}(M_B) = \mathbf{V}(M_C) = (\mathbf{V}(Y2) + \mathbf{V}(Y2')) / 4 + (\mathbf{V}(Y1) + \mathbf{V}(Y1')) / 4 = \sigma^2$

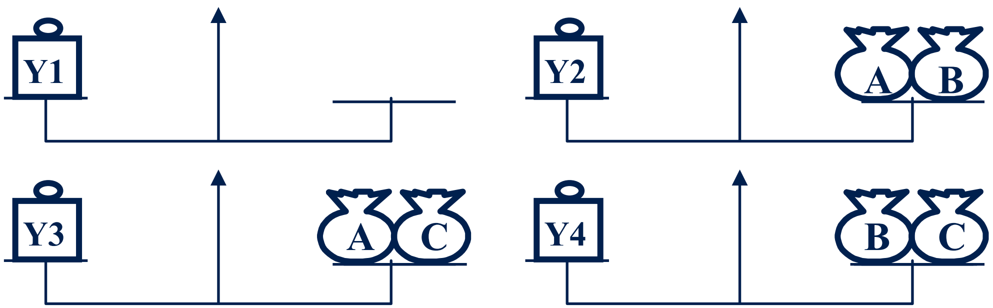
Et comment faire encore mieux ?

Nombre d'essais	4	8	16	32
V(MA)=V(MB)=V(MC)	$2\ \sigma^2$	σ^2	$\sigma^2 / 2$	$\sigma^2 / 4$

Mais peut-on améliorer la précision sans augmenter le nombre d'essais ?

Stratégie 2

Deux objets sont pesés à la fois



Matrice d'expérience

	A	B	C
E1	0	0	0
E2	1	1	0
E3	1	0	1
E4	0	1	1

Système d'équations à résoudre

$$M_0 = Y1$$
$$M_0 + M_A + M_B = Y2$$
$$M_0 + M_A + M_C = Y3$$
$$M_0 + M_B + M_C = Y4$$

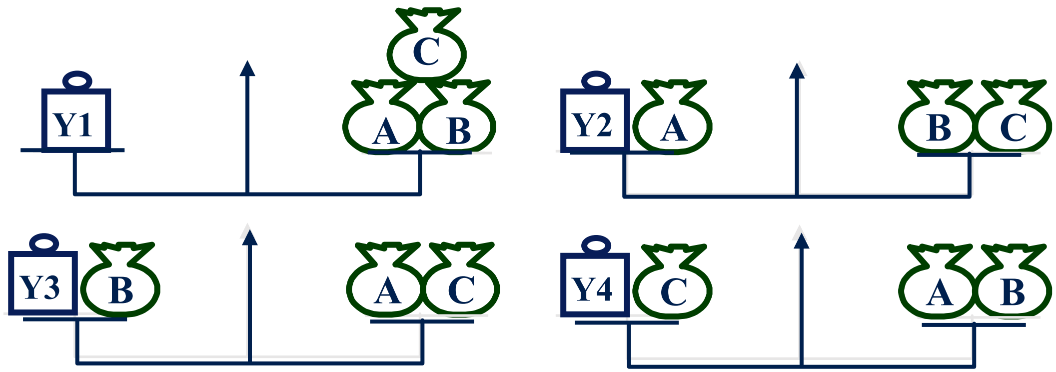
Estimation des poids

$$M_0 = Y1$$
$$M_A = (Y2 + Y3 - Y1 - Y4)/2$$
$$M_B = (Y2 + Y4 - Y1 - Y3)/2$$
$$M_C = (Y3 + Y4 - Y1 - Y2)/2$$

$V(M_A) = V(M_B) = V(M_C) = (\sigma^2 + \sigma^2 + \sigma^2 + \sigma^2) / 4 = \sigma^2$ **Coût = N = 4**

Stratégie 3

Trois objets sont pesés à la fois



Matrice d'expérience

	A	B	C
E1	1	1	1
E2	-1	1	1
E3	1	-1	1
E4	1	1	-1

Système d'équations à résoudre

$M_0 + M_A + M_B + M_C = Y1$

$M_0 - M_A + M_B + M_C = Y2$

$M_0 + M_A - M_B + M_C = Y3$

$M_0 + M_A + M_B - M_C = Y4$

Estimations des poids

$M_0 = (Y2 + Y3 + Y4 - Y1)/2$

$M_A = (Y1 - Y2)/2$

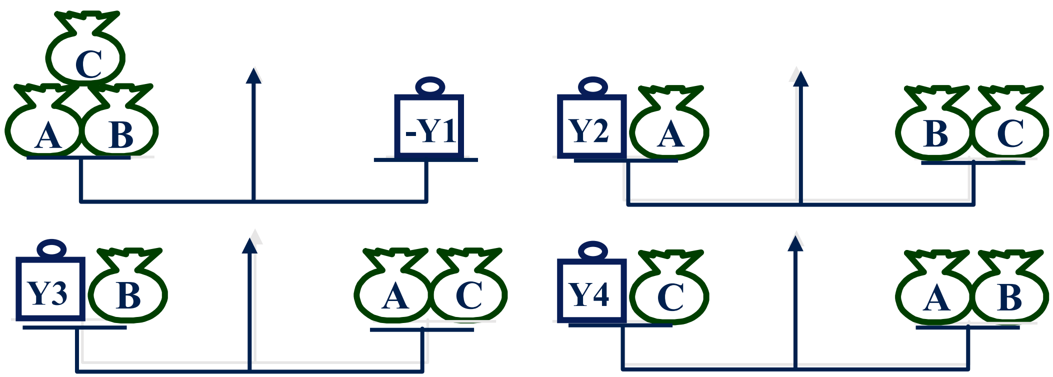
$M_B = (Y1 - Y3)/2$

$M_C = (Y1 - Y4)/2$

$$V(M_A) = V(M_B) = V(M_C) = (\sigma^2 + \sigma^2) / 4 = \sigma^2 / 2 \qquad N = 4$$

Stratégie 4

La première pesée est inversée



Matrice d'expérience

	A	B	C
E1	-1	-1	-1
E2	-1	1	1
E3	1	-1	1
E4	1	1	-1

Système d'équations à résoudre

$$M_0 - M_A - M_B - M_C = Y1$$
$$M_0 - M_A + M_B + M_C = Y2$$
$$M_0 + M_A - M_B + M_C = Y3$$
$$M_0 + M_A + M_B - M_C = Y4$$

Estimation des poids

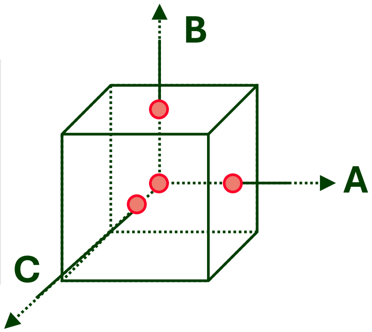
$$M_0 = (Y1 + Y2 + Y3 + Y4)/4$$
$$M_A = (-Y1 - Y2 + Y3 + Y4)/4$$
$$M_B = (-Y1 + Y2 - Y3 + Y4)/4$$
$$M_C = (-Y1 + Y2 + Y3 - Y4)/4$$

$$V(M_A) = V(M_B) = V(M_C) = (\sigma^2 + \sigma^2 + \sigma^2 + \sigma^2) / 16 = \sigma^2 / 4 \qquad N = 4$$

Pourquoi la stratégie 4 est-elle la meilleure ?

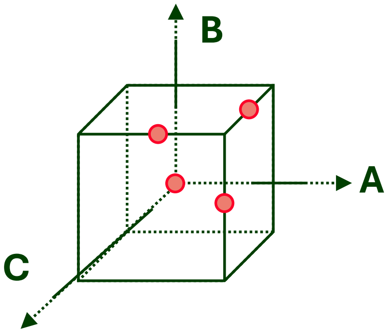
Stratégie 1

	A	B	C
Exp 1	0	0	0
Exp 2	1	0	0
Exp 3	0	1	0
Exp 4	0	0	1



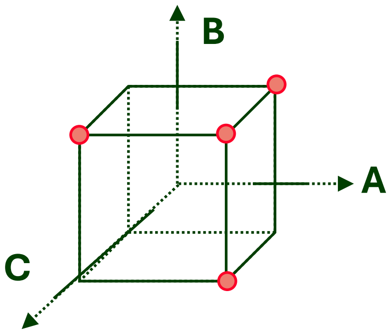
Stratégie 2

	A	B	C
Exp 1	0	0	0
Exp 2	1	1	0
Exp 3	1	0	1
Exp 4	0	1	1



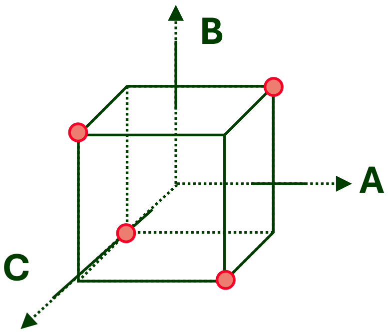
Stratégie 3

	A	B	C
Exp 1	1	1	1
Exp 2	-1	1	1
Exp 3	1	-1	1
Exp 4	1	1	-1



Stratégie 4

	A	B	C
Exp 1	-1	-1	-1
Exp 2	-1	1	1
Exp 3	1	-1	1
Exp 4	1	1	-1



Reformulation du problème des pesées

Problème

Etudier (ou PESER) l'effet de **3 facteurs**
(température [100, 200], vitesse [1, 3] et débit [20, 30])
sur le rendement d'une réaction chimique.

Place de l'objet sur la balance

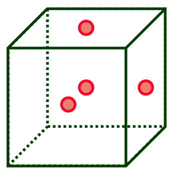
Objet sur le plateau mode **gauche** ⇔ Niveau **minimum** du facteur

Objet sur le plateau de **droite** ⇔ Niveau **maximum** du facteur

Objet **hors** de la balance ⇔ Niveau **moyen** du facteur

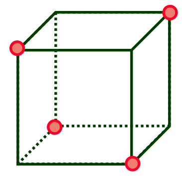
Stratégie 1 :

Un facteur varie à la fois



Stratégie 4 :

Plan factoriel fractionnaire



Ecriture du problème des pesées comme un problème de régression linéaire multiple

L'estimation des poids au terme de l'expérience des pesées peut s'écrire sous la forme d'une régression linéaire multiple.

$$Y = X\beta + \varepsilon \qquad Y = \mu_0 + \mu_1 x_1 + \mu_2 x_2 + \mu_3 x_3 + \varepsilon$$

Les poids/paramètres sont alors estimés par : $b = (X'X)^{-1} X'y$

Avec

$$Y = \begin{bmatrix} Y1 \\ Y2 \\ Y3 \\ Y4 \end{bmatrix} \quad X = \begin{bmatrix} 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix} \quad \beta = \begin{bmatrix} \mu 0 \\ \mu 1 \\ \mu 2 \\ \mu 3 \end{bmatrix} \quad b = \begin{bmatrix} M0 \\ M1 \\ M2 \\ M3 \end{bmatrix}$$

Matrice de variance covariance des paramètres : $V(b) = \sigma^2 (X'X)^{-1}$

	4	0	0	0			1/4	0	0	0
(X'X)=	0	4	0	0		(X'X)^-1=	0	1/4	0	0
	0	0	4	0			0	0	1/4	0
	0	0	0	4			0	0	0	1/4

Ce plan est optimal car il fournit des estimations de variance minimum et indépendantes (orthogonalité)

Exemple 4 : étude d'un procédé de fermentation

Le contexte

Les fermenteurs sont utilisés couramment dans la production de vaccins mais le bon choix de leurs paramètres de réglage est primordial pour assurer une production suffisante, de qualité et stable dans le temps.

But de l'étude: Problème de criblage

Rechercher quels facteurs de fermentation ont le plus d'effet sur la biomasse finale obtenue au terme d'une fermentation de 24h ainsi que le sens de ces effets.

Budget : Nombre maximum d'essais : 16 (=16 jours)

Facteurs d'intérêt : 10 facteurs quantitatifs fixes

Phase de croissance (0 à 18h) :

X1 - Température de croissance, **X2** – Surpression, **X3** - Débit d'air , **X4** - Vitesse d'agitation **X5** – pH, **X6** – Débit d'alimentation , **X7** – Débit d'antimousse C (>16h)

Phase d'induction (18-24h) :

X8 – Moment d'induction, **X9** – Température d'induction, **X10** – Débit d'antimousse de C

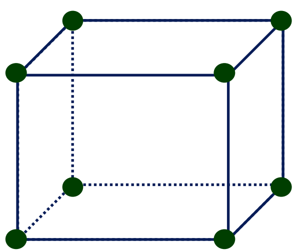
Quels sont les plans possibles ?

Modèle classiquement utilisé dans un problème de «criblage»

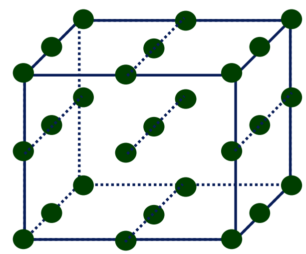
16 essais, modèle d'ordre 1. $p = 11 < N = 16$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_{10} X_{10} + \varepsilon$$

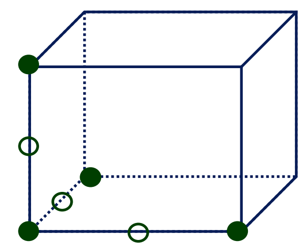
Plans possibles pour **3 facteurs**



Factoriel complet 2^k



Factoriel complet 3^k



1 facteur varie à la fois

Pour **10 facteurs**

Plan factoriel complet à **2 niveaux** : 1024 essais = 3 ans

Plan factoriel complet à **3 niveaux** : 59049 essais = 164 ans

Plan "1 facteur à la fois" : 11 essais → mais pas précis

Plans proposés :

Plan **factoriel fractionnaire 2^{k-r}** à **16 essais** ou
plan de Plackett et Bruman à 12 essais (non vu).

Plan factoriel fractionnaire 2^{k-r}

Définition

Un plan **factoriel fractionnaire 2^{k-r}** à k facteurs est un plan résultant de r fractionnements successifs d'un plan factoriel complet 2^k

Principe du fractionnement

Définir r « **générateurs** » qui permettent de fractionner successivement le plan en 2 puis 4 puis .... 2^r fractions de 2^{k-r} essais. Une des fractions est utilisée comme plan.

Générateur

Un générateur est une formule (par exemple $X_1 X_2 X_3=1$) qui indique qu'on sélectionne dans le plan de départ les essais tels que le produit $X_1X_2X_3$ des valeurs des facteurs (en valeurs standardisées) est égal à 1.
La partie gauche du générateur est toujours un produit de facteurs et la partie droite vaut 1 ou -1 selon que l'on garde l'une ou l'autre partie du plan.

Propriétés des plans factoriels fractionnaires 2^{k-r}

Orthogonaux et de variance minimale mais il faut bien choisir les générateurs.

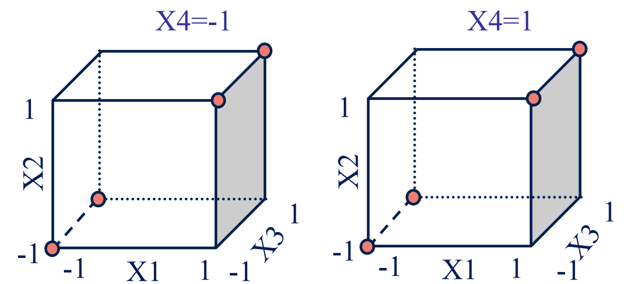
Exemples de plans factoriels fractionnaires 2^{4-1} à 8 essais

Plan complet

	Valeurs standardisées			
Essais	**X1**	**X2**	**X3**	**X4**
1	-1	-1	-1	-1
2	1	-1	-1	-1
3	-1	1	-1	-1
4	1	1	-1	-1
5	-1	-1	1	-1
6	1	-1	1	-1
7	-1	1	1	-1
8	1	1	1	-1
9	-1	-1	-1	1
10	1	-1	-1	1
11	-1	1	-1	1
12	1	1	-1	1
13	-1	-1	1	1
14	1	-1	1	1
15	-1	1	1	1
16	1	1	1	1

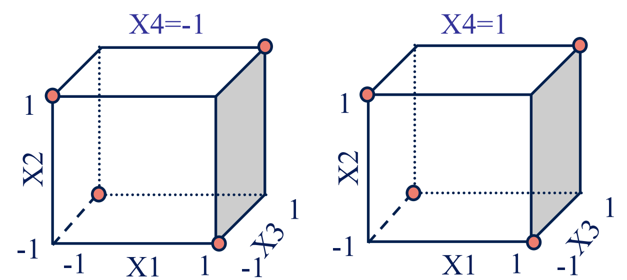
Générateur 1 : $X_1X_2=1$

	Valeurs standardisées			
Essais	**X1**	**X2**	**X3**	**X4**
1	-1	-1	-1	-1
4	1	1	-1	-1
5	-1	-1	1	-1
8	1	1	1	-1
9	-1	-1	-1	1
12	1	1	-1	1
13	-1	-1	1	1
16	1	1	1	1



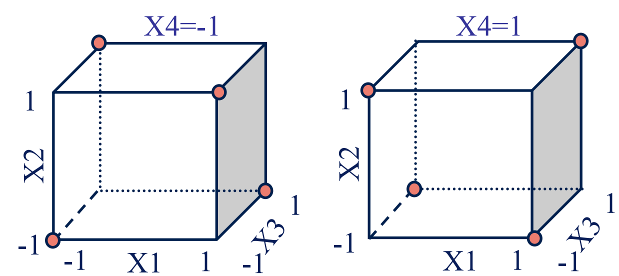
Générateur 2 : $X_1X_2X_3=1$

	Valeurs standardisées			
Essais	**X1**	**X2**	**X3**	**X4**
2	1	-1	-1	-1
3	-1	1	-1	-1
5	-1	-1	1	-1
8	1	1	1	-1
10	1	-1	-1	1
11	-1	1	-1	1
13	-1	-1	1	1
16	1	1	1	1



Générateur 3 : $X_1X_2X_3X_4=1$

	Valeurs standardisées			
Essais	**X1**	**X2**	**X3**	**X4**
1	-1	-1	-1	-1
4	1	1	-1	-1
6	1	-1	1	-1
7	-1	1	1	-1
10	1	-1	-1	1
11	-1	1	-1	1
13	-1	-1	1	1
16	1	1	1	1



Ces trois générateurs permettent-ils d'estimer un modèle d'ordre 1 ?

Quel est le générateur le plus adéquat ?

Tableau de générateurs pour le fractionnement de plans

Quand on doit choisir plus d'un générateur, des **tables** existent.

La résolution détermine quel modèle pourra être estimé

R=3 (III) → modèle d'ordre 1 (plus éventuellement qq interactions)

R=4 (IV) → modèle d'ordre 1 et quelques interactions X_iX_j

R=5 (V) → modèle d'ordre 1 et toutes les interactions

k	*r*	*N*	Résolution	Générateurs
3	1	4	III	123 = ±1
4	1	8	IV	1234 = ±1
5	2	8	III	124 = ±1, 135 = ±1
	1	16	V	12345 = ±1
6	3	8	III	124 = ±1, 135 = ±1, 236 = ±1
	2	16	IV	1235 = ±1, 2346 = ±1
7	4	8	III	124 = ±1, 135 = ±1, 236 = ±1, 1237 = ±1
	3	16	IV	1235 = ±1, 2346 = ±1, 1347= ±1
8	4	16	IV	2345 = ±1, 1346 = ±1, 1237 = ±1, 1248 = ±1
9	5	16	III	1235 = ±1, 2346 = ±1, 1347 = ±1, 1248 = ±1, 12349 = ±1
10	6	16	III	1235 = ±1, 2346 = ±1, 1347 = ±1, 1248 = ±1, 12349 = ±1, 12(10) = ±1

Voir : Box, Hunter et Hunter : Statistics for experimenters

Exemple : Plan factoriel fractionnaire : construction

Les 6 générateurs ci-dessous permettent d’extraire **16 essais** du plans factoriel complet à 1024 (= 2^{10}) essais

Générateurs : $G_1: X_1X_2X_3X_5 = 1$ $G_2: X_2X_3X_4X_6 = 1$ $G_3: X_1X_3X_4X_7 = 1$

$G_4: X_1X_2X_4X_8 = 1$ $G_5: X_1X_2X_3X_4X_9 = 1$ $G_6: X_1X_2X_{10} = 1$

Plan factoriel fractionnaire 2^{10-6}										
Facteurs	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
Générateurs					**123**	**234**	**134**	**124**	**1234**	**12**
1	-1	-1	-1	-1	-1	-1	-1	-1	1	1
2	1	-1	-1	-1	1	-1	1	1	-1	-1
3	-1	1	-1	-1	1	1	-1	1	-1	-1
4	1	1	-1	-1	-1	1	1	-1	1	1
5	-1	-1	1	-1	1	1	1	-1	-1	1
6	1	-1	1	-1	-1	1	-1	1	1	-1
7	-1	1	1	-1	-1	-1	1	1	1	-1
8	1	1	1	-1	1	-1	-1	-1	-1	1
9	-1	-1	-1	1	-1	1	1	1	-1	1
10	1	-1	-1	1	1	1	-1	-1	1	-1
11	-1	1	-1	1	1	-1	1	-1	1	-1
12	1	1	-1	1	-1	-1	-1	1	-1	1
13	-1	-1	1	1	1	-1	-1	1	1	1
14	1	-1	1	1	-1	-1	1	-1	-1	-1
15	-1	1	1	1	-1	1	-1	-1	-1	-1
16	1	1	1	1	1	1	1	1	1	1

Randomization

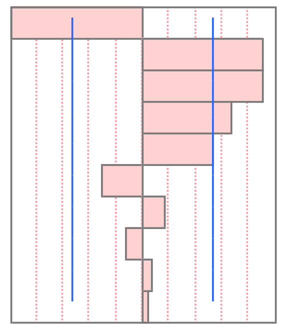
- Attention, selon les designs, on ne veut pas forcément faire les experiences dans l'ordre du design pour ne pas avoir de confusion entre certains facteurs et des facteurs aléatoires (mesurés ou non) qui varient avec le temps.

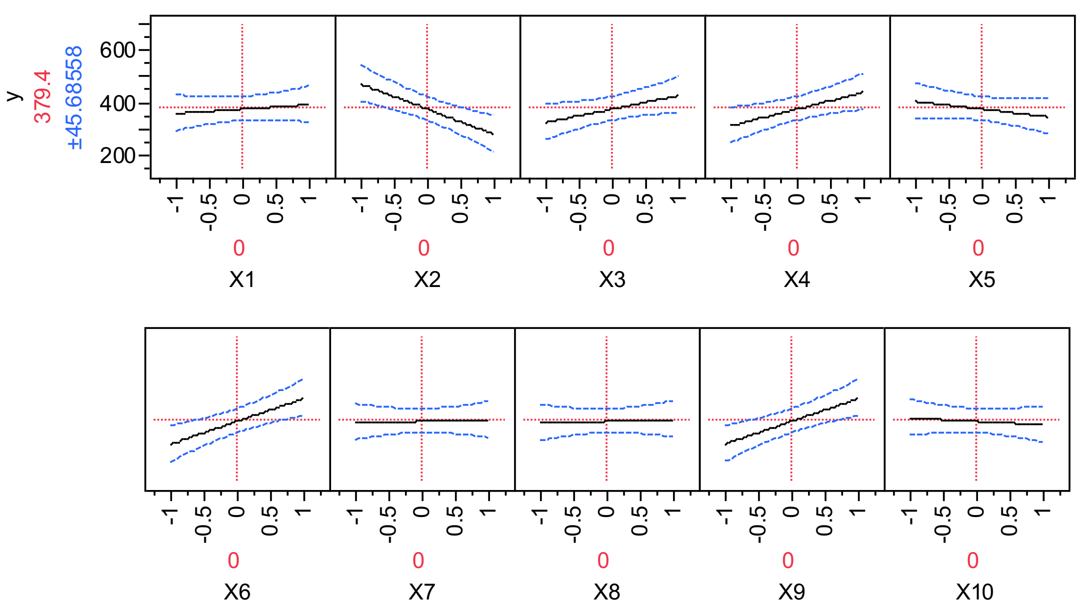
Résultat typique d'un criblage

Le plan de criblage permet uniquement d'étudier si l'effet est présent et son signe

Sorted Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
X2	-95.58333	18.39692	-5.20	0.0065*
X6	87.083333	18.39692	4.73	0.0091*
X9	86.416667	18.39692	4.70	0.0093*
X4	63.583333	18.39692	3.46	0.0259*
X3	51.416667	18.39692	2.79	0.0491*
X5	-28.75	18.39692	-1.56	0.1931
X1	16.583333	18.39692	0.90	0.4183
X10	-12.41667	18.39692	-0.67	0.5367
X8	6.75	18.39692	0.37	0.7323
X7	4.0833333	18.39692	0.22	0.8352

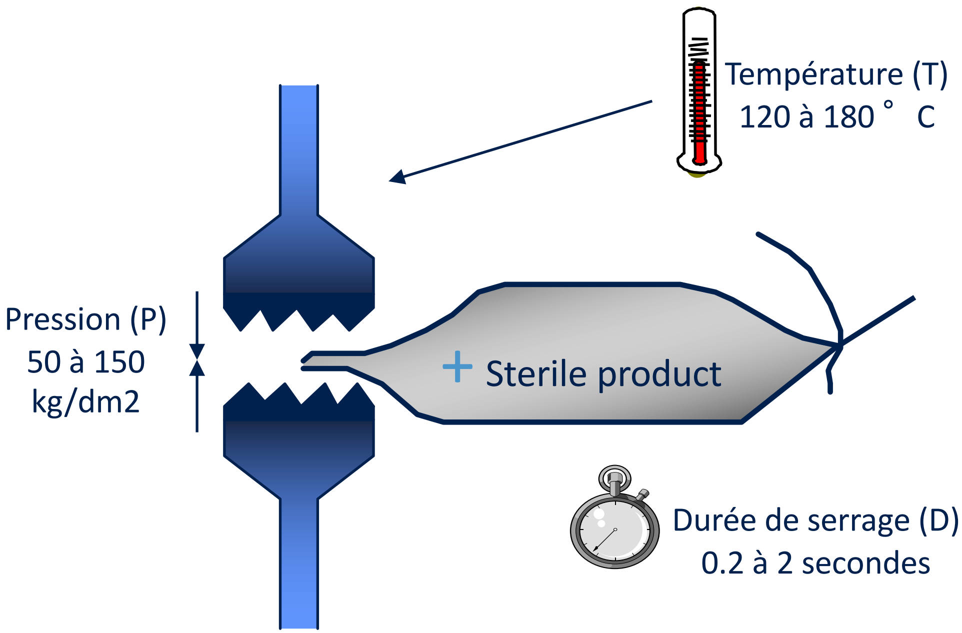




Plan du chapitre

- Remise en contexte
- Exemples traités dans ce chapitre
- Rappel de vocabulaire et des composantes essentielles d'un design d'expérience
- Classes de plans multi-facteurs
  - Objectif général
  - Plans factoriels complets
  - Plans factoriels fractionnaires 2^{k-p} et pour le criblage de facteurs
  - Plans pour l'estimation de surfaces de réponse
    - factoriel complet, composite centré, Doehlert, Box Benken

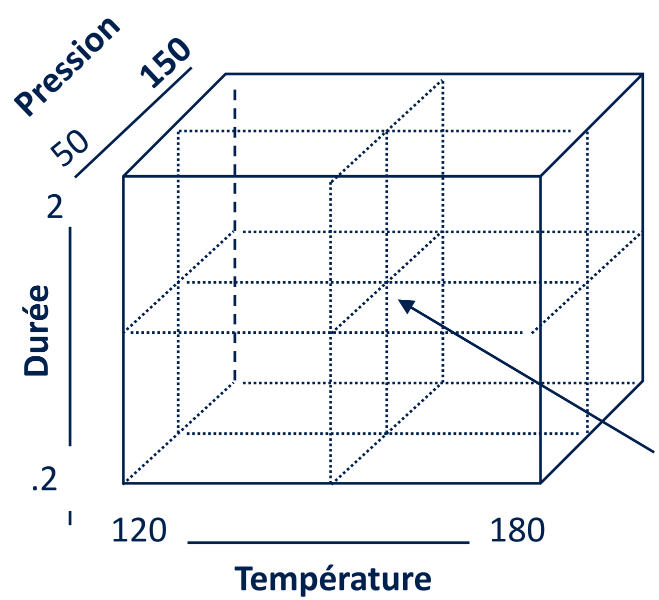
Exemple 3 : optimisation d’un processus de soudure



But :

Rechercher des conditions optimales de fabrication qui assurent

- une **résistance** à la soudure proche de 85 (kg/dm^2)
- une **qualité visuelle** de la soudure PRINT > 4



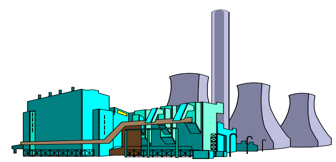
Question:

Comment explorer le domaine et utiliser la modélisation pour répondre à la question ?

Plans pour « surface de réponse »

Objectifs de la méthodologie des « surfaces de réponse »

- Étudier le sens, l'amplitude et de la forme des effets de facteurs QUANTITATIFS sur des réponses via un modèle statistique.
- Utiliser le modèle pour réaliser des prédictions et optimisation des réponses.

Facteurs →  **Processus** → **Réponses**

Modèle : $\hat{y} = f(x_1, x_2 \dots x_k; \theta)$

Modèle visé (le plus souvent) : Modèle polynomial quadratique complet (c'est pour ce modèle que les plans RSD classiques sont prévus)

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \beta_{11} X_1^2 + \dots + \beta_{kk} X_k^2$$
$$+ \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \dots + \beta_{(k-1)k} X_{k-1} X_k + \varepsilon$$

pour $k = 3$, $p = 10$

Besoin :

Choisir un "bon" plan qui donne des estimations et prédictions précises

Choix d'une classe de plan possible

Etape 1

Quels plans sont capables de répondre à la question ?

Conditions

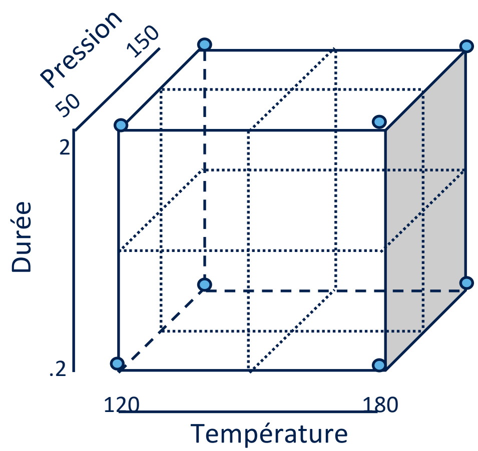
→ Permettre d'estimer le modèle

→ Ne pas dépasser le budget disponible (20 essais).

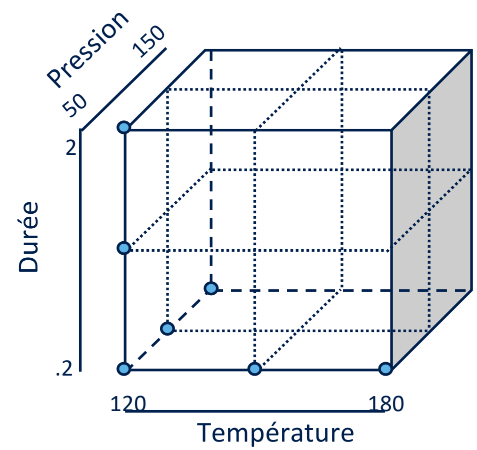
Etape 2

Dans les plans possibles, quel plan choisir ?

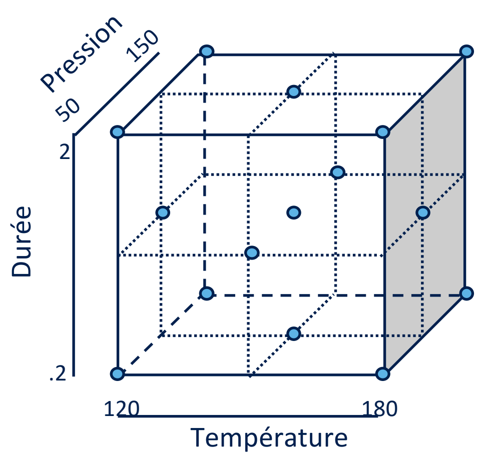
Il y a-t-il des alternatives ?



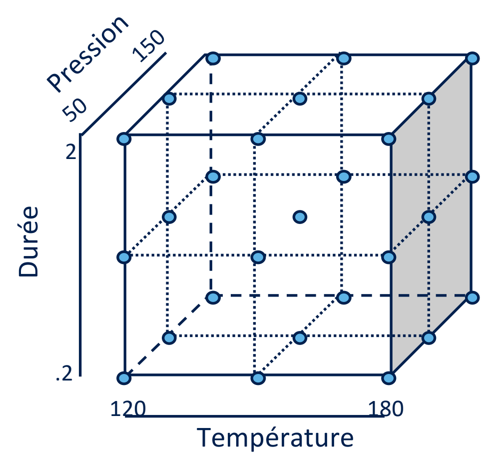
Plan 1



Plan 2



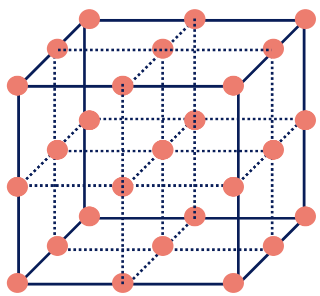
Plan 3



Plan 4

Plans à 3 facteurs pour estimer un modèle d'ordre 2

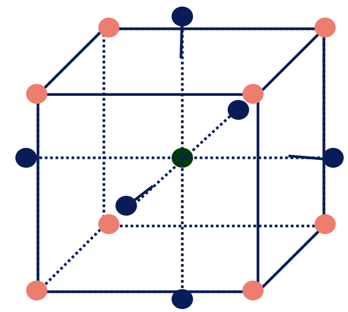
Plan factoriel complet 3^k



Un traitement = un point support

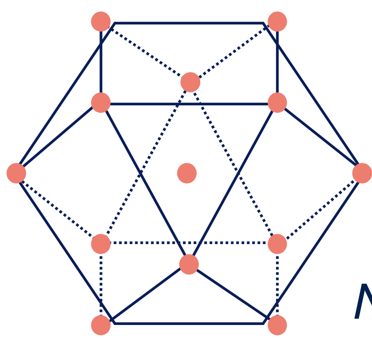
$N = 27$

Plan composite centré



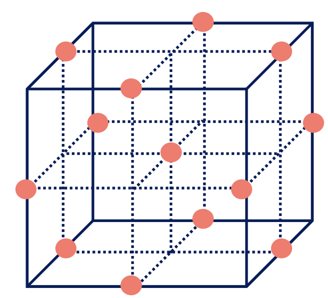
$N = 15$

Plan de Doehlert



$N = 13$

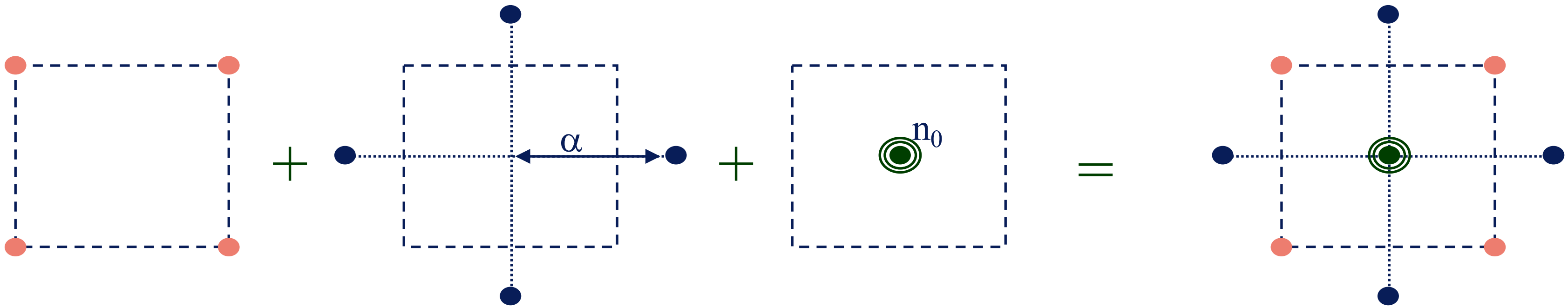
Plan de Box et Behnken



$N = 13$

Comment choisir un plan ou l'autre ?

Plan composite centré : définition



C'est le plan le plus utilisé pour l'estimation de modèles quadratiques.

Définition : Un plan composite centré à k facteurs est composé de

- ***N_f*** essais d'un plan factoriel complet ou fractionnaire 2^{k-r} pour estimer les effets principaux et les interactions.
- ***2k*** essais en étoile sur les axes à une distance α du centre du domaine pour estimer les effets quadratiques.
- **n_0** essais au centre du domaine (pour estimer la variabilité)

Coût : $N = 2^{k-r} + 2k + n_0$

k	2	3	4	5
n	9	15	25	27
p	6	10	15	21

avec
$r = 0$,
$n_0=1$

Comparaison de la qualité des plans

On compare le plus souvent les plans sur base de leur coût, de la qualité des estimateurs des paramètres (puissances) et des prédictions.

Comparaison des plans à 3 facteurs : Composite, Box-Behnken et factoriel avec $N_0 = 3$ et $\alpha = 1$ pour le plan composite.

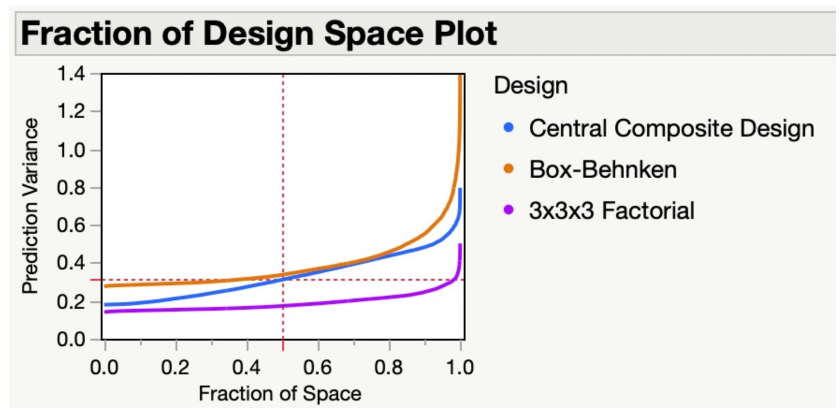
Power Analysis

Significance Level 0.05

Anticipated RMSE 1

N=17 N=15 N=29

Term	Anticipated Coefficient	Central Composite Design Power	Box-Behnken Power	3x3x3 Factorial Power
Intercept	1	0.522	0.291	0.632
X1	1	0.774	0.623	0.980
X2	1	0.774	0.623	0.980
X3	1	0.774	0.623	0.980
X1*X2	1	0.681	0.368	0.907
X1*X3	1	0.681	0.368	0.907
X2*X3	1	0.681	0.368	0.907
X1*X1	1	0.294	0.345	0.687
X2*X2	1	0.294	0.345	0.687
X3*X3	1	0.294	0.345	0.687

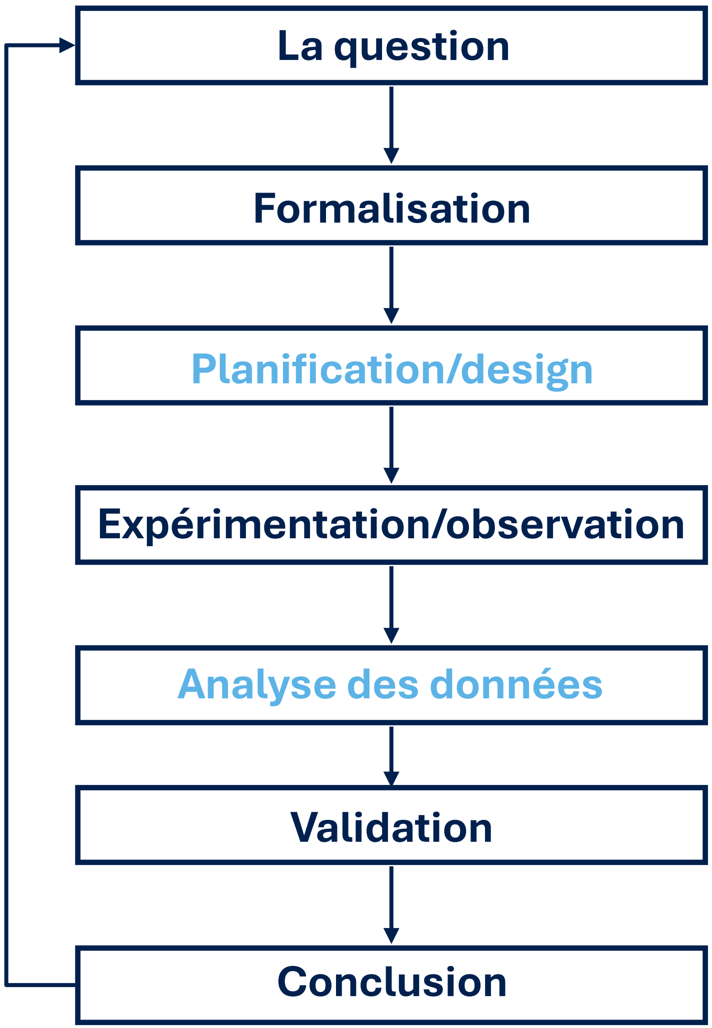


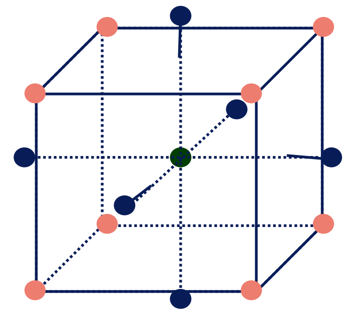
Design Diagnostics

	Efficiency of Central Composite Design Relative to Box-Behnken	Efficiency of Central Composite Design Relative to 3x3x3 Factorial
D-efficiency	1.277	0.564
G-efficiency	1.756	0.632
A-efficiency	1.148	0.525
I-efficiency	1.190	0.579
Additional Run Size	2	-12

Good 1.50 1.25 0.80 0.67 Bad

Retour à l’exemple 5

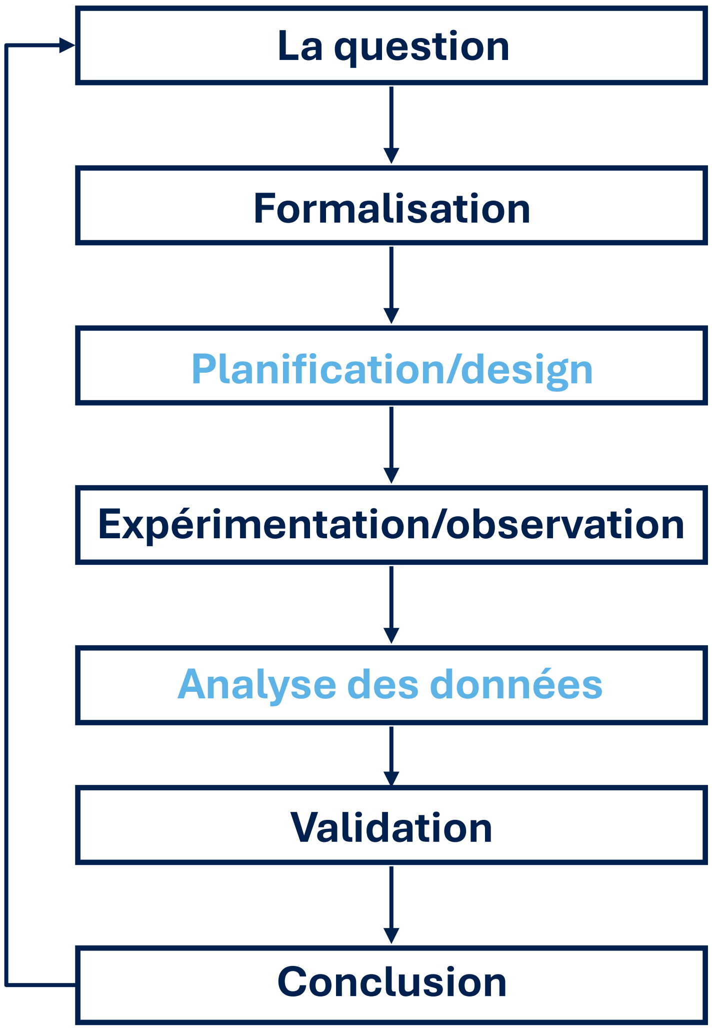


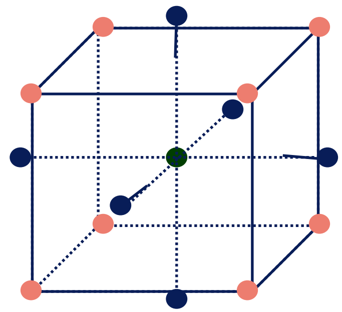


N = 15 + 4 rép au centre

	Temp	Press	Duree	Ts	PS	DS	Bond	Print
1	120	50	0.2	-1	-1	-1	13.2	3.35
2	120	50	2	-1	-1	1	89.85	2.04
3	120	100	1.1	-1	0	0	65.32	3.87
4	120	150	0.2	-1	1	-1	11.25	4.48
5	120	150	2	-1	1	1	89.97	3.01
6	150	50	1.1	0	-1	0	91.45	3.14
7	150	100	0.2	0	0	-1	70.53	3.54
8	150	100	1.1	0	0	0	86.89	3.81
9	150	100	1.1	0	0	0	91.03	3.63
10	150	100	1.1	0	0	0	93.11	3.46
11	150	100	1.1	0	0	0	89.41	3.74
12	150	100	1.1	0	0	0	88.71	3.62
13	150	100	2	0	0	1	80.92	2.46
14	150	150	1.1	0	1	0	93.29	4.36
15	180	50	0.2	1	-1	-1	92.94	2.04
16	180	50	2	1	-1	1	44.53	1.11
17	180	100	1.1	1	0	0	81.55	2.32
18	180	150	0.2	1	1	-1	91.53	2.77
19	180	150	2	1	1	1	41.83	2.07

Retour à l'exemple 5 (alternative + analyses avec R)





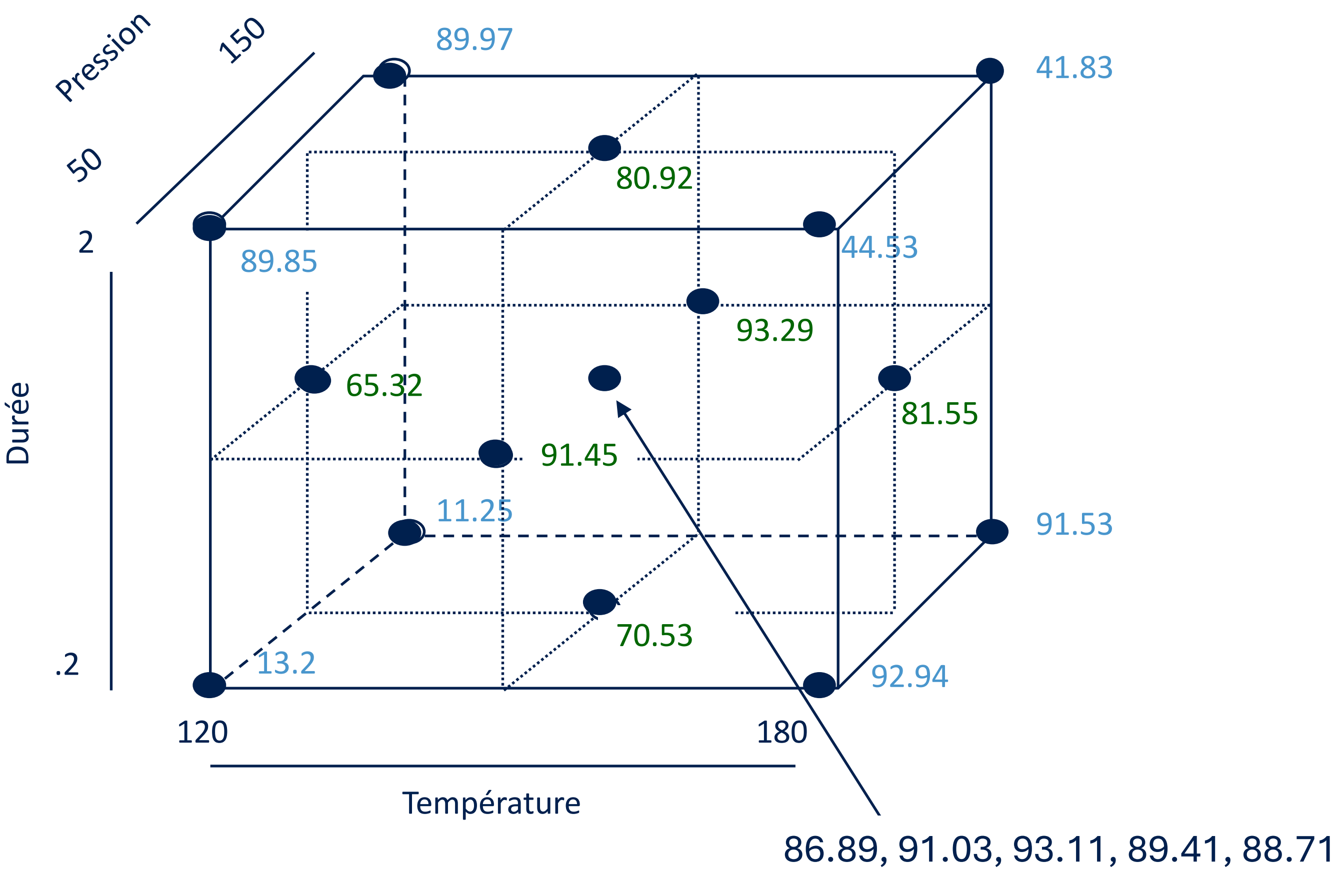
On garde 3 essais (N = 3) pour "valider" nos résultats.

N = 15 + 2 rép au centre

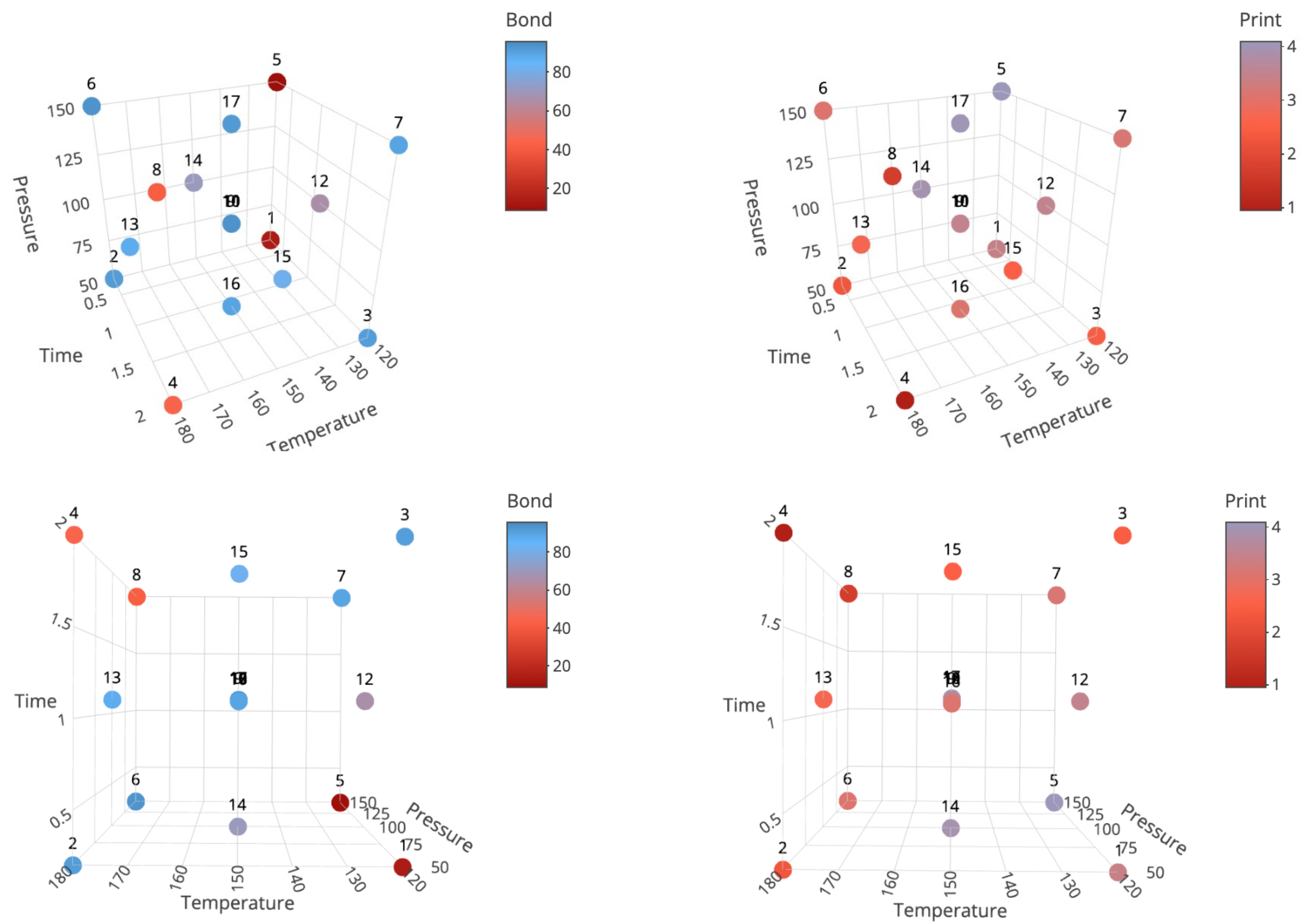
Trial_number	Temperature	Time	Pressure	Bond	Print
1	120	0.2	50	13.65	3.41
2	180	0.2	50	91.46	2.27
3	120	2.0	50	91.17	2.45
4	180	2.0	50	44.73	0.95
5	120	0.2	150	8.66	4.09
6	180	0.2	150	93.49	3.07
7	120	2.0	150	89.58	3.16
8	180	2.0	150	41.14	1.67
9	150	1.1	100	95.74	3.79
10	150	1.1	100	89.75	3.59
11	150	1.1	100	93.72	3.51
12	120	1.1	100	65.82	3.52
13	180	1.1	100	87.65	2.70
14	150	0.2	100	70.68	3.93
15	150	2.0	100	81.65	2.51
16	150	1.1	50	89.71	3.10
17	150	1.1	150	91.93	4.04

Analyse visuelle des résultats (bond)

Que peut-on dire sur l'effet des facteurs ?



Analyse visuelle des résultats



Analyse statistique des résultats

Résultats expérimentaux

Essais	ξ_1	ξ_2	Y
1	1	3	86.89
2	4	5	93.29
3	4	1	91.03
4	3	3	91.45
5	5	3	65.32
6	2	5	89.85
7	2	1	13.20
8	3	3	89.97

Modèle choisi dans l'étape de formalisation

$$Y=f(X_1, X_2 \ldots X_k; \beta_1, \beta_2, \ldots \beta_p)+\varepsilon$$

Analyse visuelle des résultats

Ajustement du modèle aux données
par régression linéaire multiple

Vérification de la **qualité** du modèle

Interprétation des effets des facteurs sur les réponses

Utilisation du modèle estimé pour faire des **prédictions**

Recherche de conditions de fabrication **optimales**

Réponse Bond : Modèle de régression

Summary of Fit

RSquare	$\boxed{0.997131}$
RSquare Adj	0.994262
Root Mean Square Error	$\boxed{2.024217}$
Mean of Response	74.06895
Observations (or Sum Wgts)	19

99.7% de la variance du Bond expliquée par le modèle. ☺

Estimation de l'écart-type des erreurs

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	9	12817.009	1424.11	347.5601
Error	9	36.877	4.10	**Prob > F**
C. Total	18	12853.886		<.0001*

Le modèle explique quelque chose: au moins un terme est significatif

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	41.396	3.612	11.46	<.0001*
Temp	0.276	0.021	12.93	<.0001*
Press	-0.008	0.013	-0.64	0.5378
Duree	7.517	0.711	10.57	<.0001*
(Temp-150)*(Temp-150)	-0.019	0.001	-14.14	<.0001*
(Temp-150)*(Press-100)	-0.000	0.000	-0.40	0.6997
(Press-100)*(Press-100)	0.001	0.000	1.32	0.2188
(Temp-150)*(Duree-1.1)	-1.174	0.027	-44.27	<.0001*
(Press-100)*(Duree-1.1)	0.002	0.016	0.14	0.8946
(Duree-1.1)*(Duree-1.1)	-18.55	1.512	-12.27	<.0001*

Equation du modèle

Termes liés à la pression non-significatifs *p*-valeur >> 0.05

Réponse Bond : Modèle de régression

```
Bond ~ Temperature + Time + Pressure + Temperature:Time + Temperature:Pressure
+ Time:Pressure + I(Temperature^2) + I(Time^2) + I(Pressure^2)
```

Equation du modèle

```
Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          -5.734e+02  4.066e+01 -14.100 2.14e-06 ***
Temperature           6.796e+00  5.634e-01  12.062 6.14e-06 ***
Time                  2.314e+02  7.423e+00  31.173 9.03e-09 ***
Pressure              5.998e-02  1.677e-01   0.358    0.731
I(Temperature^2)     -1.742e-02  1.858e-03  -9.376 3.27e-05 ***
I(Time^2)            -2.006e+01  2.065e+00  -9.717 2.58e-05 ***
I(Pressure^2)        -6.389e-04  6.690e-04  -0.955    0.371
Temperature:Time     -1.192e+00  3.585e-02 -33.256 5.76e-09 ***
Temperature:Pressure  4.183e-04  6.453e-04   0.648    0.537
Time:Pressure        -6.167e-03  2.151e-02  -0.287    0.783
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.738 on 7 degrees of freedom
Multiple R-squared:  0.996,     Adjusted R-squared:  0.9908
F-statistic: 192.7 on 9 and 7 DF,  p-value: 1.517e-07
```

Termes liés à la pression non-significatifs *p*-valeur >> 0.05

Estimation de l'écart-type des erreurs

99.6% de la variance du Bond expliquée par le modèle. ☺

Le modèle explique quelque chose: au moins un terme est significatif

Bond: Vérification de la qualité du modèle de régression

Le modèle n'est qu'une « imitation » du phénomène réel.

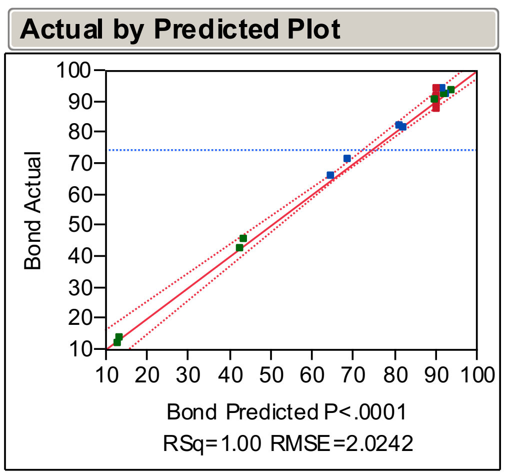
Il est donc indispensable de le valider et de vérifier les hypothèses statistiques sous-jacentes !

« All models are wrong, but some are useful » (Box)

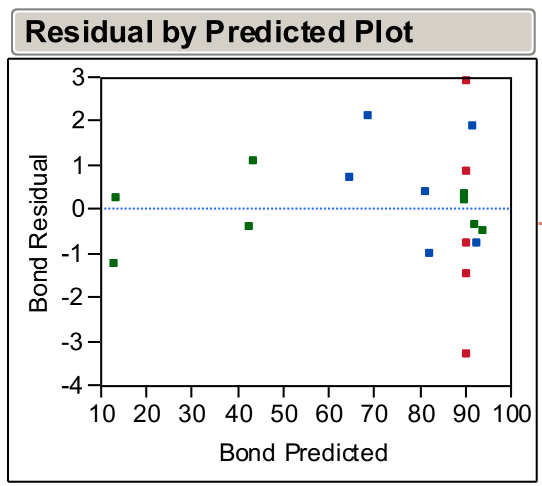
Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	5	14.604303	2.92086	0.5246
Pure Error	4	22.272800	5.56820	Prob > F 0.7523
Total Error	9	36.877103		Max RSq 0.9983

L'équation du modèle semble valide car la variance des résidus est proche de la variance expérimentale



Les valeurs prédites par le modèle sont proche des points observés. ☺



Les résidus doivent, si possible, être répartis de manière homogène autour de 0.

Bond: Vérification de la qualité du modèle de régression

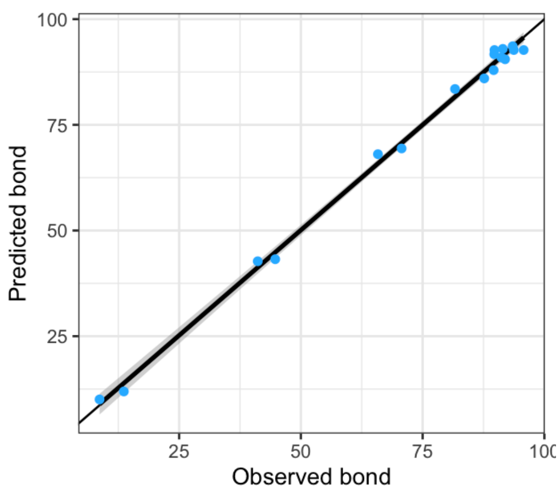
Le modèle n'est qu'une « imitation » du phénomène réel.

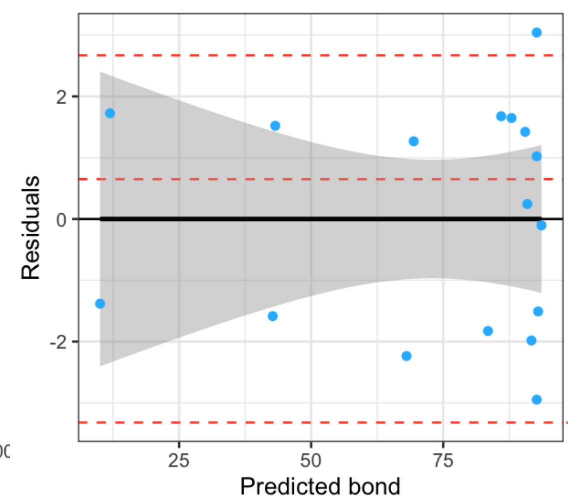
Il est donc indispensable de le valider et de vérifier les hypothèses statistiques sous-jacentes !

source	df	SSE	MSE	F_stat	p_value
pure error	2	18.57	9.287	NA	NA
lack of fit	5	33.89	6.779	0.7299	0.5271
model residuals	7	52.47	7.495	NA	NA

L'équation du modèle semble valide car la variance des résidus est proche de la variance expérimentale

« All models are wrong, but some are useful » *(Box)*





Les résidus doivent être répartis de manière homogène autour de 0.

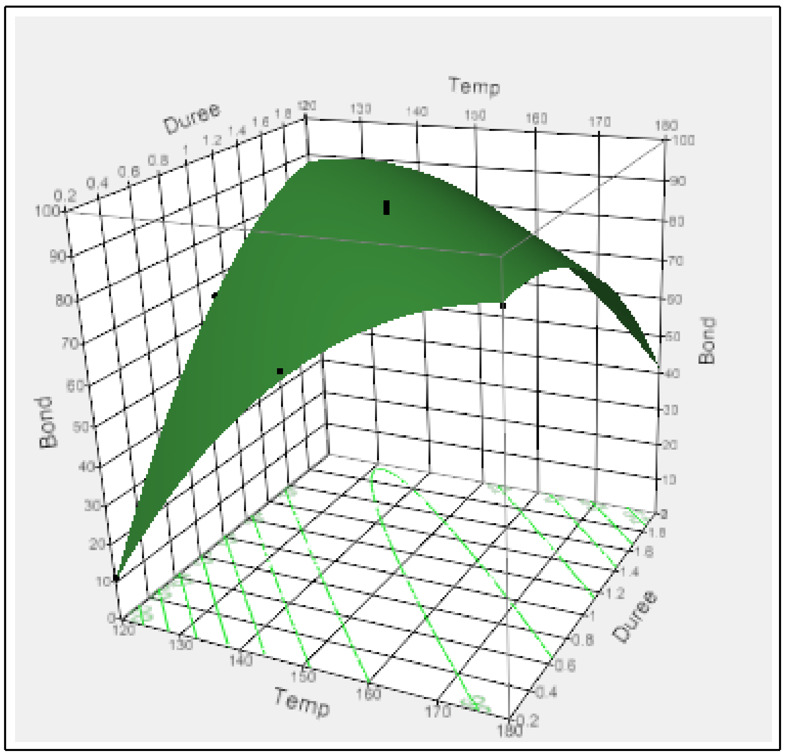
Les 3 lignes rouges représentent la différence entre les valeurs de bond des répétitions au centre et leur moyenne (= variabilité expérimentale)

Les valeurs prédites par le modèle sont proche des points observés. ☺

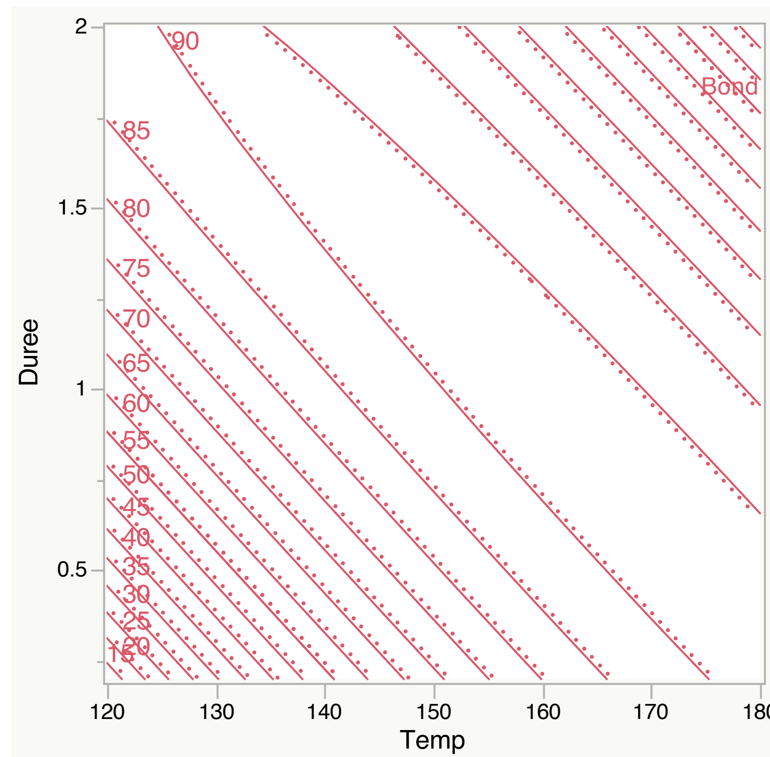
Visualisation du modèle réponse Bond

Puisque la pression ne semble pas utile pour prédire le bond, on peut représenter le bond en fonction de la température et de la durée.

Surface de réponse



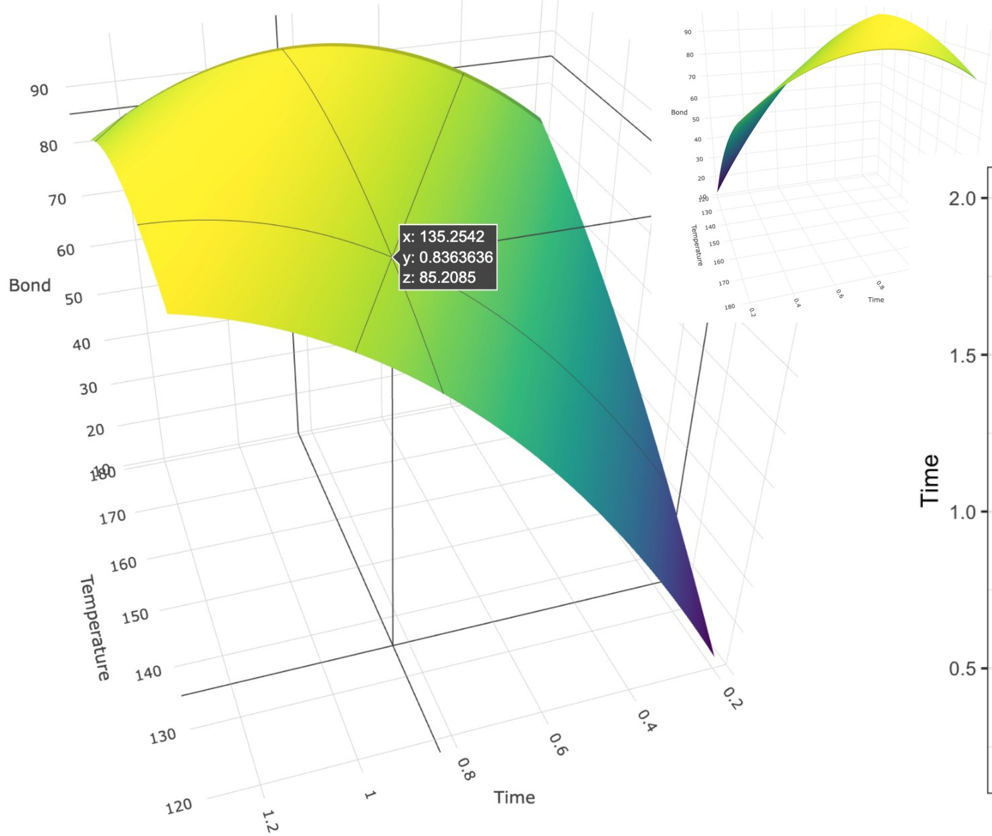
Courbes de niveau



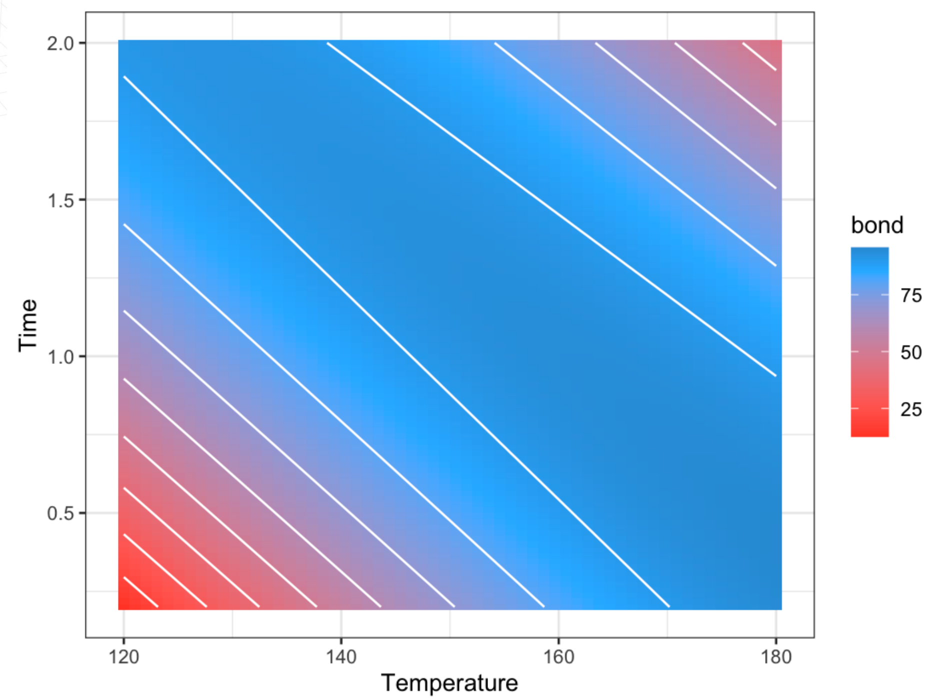
Visualisation du modèle réponse Bond

On observe qu’il y a de nombreuses combinaison the temperature & durée qui nous permettent d’avoir un bond attendu de ~85.

Surface de réponse



Courbes de niveau



Modélisation de la réponse « Print »

Summary of Fit

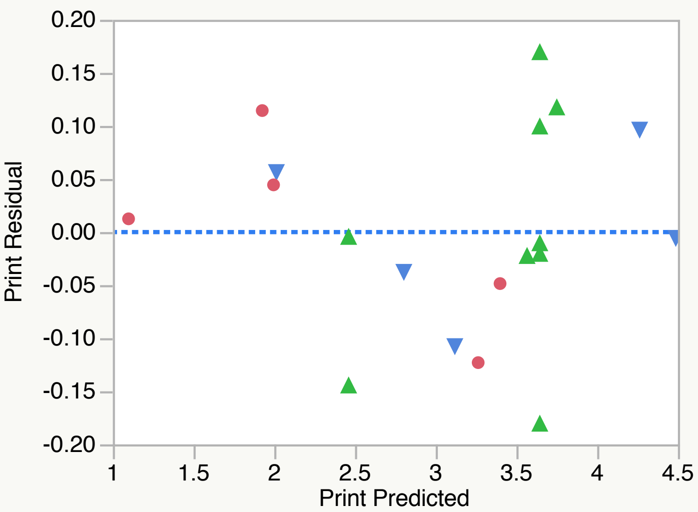
RSquare	0.988508
RSquare Adj	0.977016
Root Mean Square Error	0.135316
Mean of Response	3.095789
Observations (or Sum Wgts)	19

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	5	0.09371281	0.018743	1.0547
Pure Error	4	0.07108000	0.017770	**Prob > F**
Total Error	9	0.16479281		0.4932
				Max RSq
				0.9950

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	6.5298247	0.241431	27.05	<.0001*
Temp	-0.021467	0.001426	-15.05	<.0001*
Press	0.01002	0.000856	11.71	<.0001*
Duree	-0.61	0.047545	-12.83	<.0001*
(Temp-150)*(Temp-150)	-0.000591	0.000091	-6.50	0.0001*
(Temp-150)*(Press-100)	-3.417e-5	3.189e-5	-1.07	0.3119
(Press-100)*(Press-100)	4.9258e-5	3.274e-5	1.50	0.1668
(Temp-150)*(Duree-1.1)	0.0053241	0.001772	3.00	0.0148*
(Press-100)*(Duree-1.1)	0.0001944	0.001063	0.18	0.8589
(Duree-1.1)*(Duree-1.1)	-0.773896	0.101063	-7.66	<.0001*

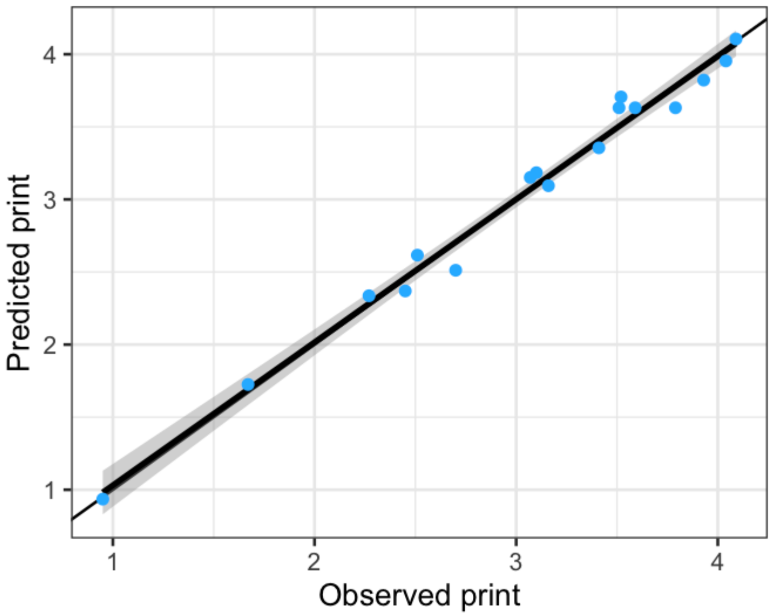


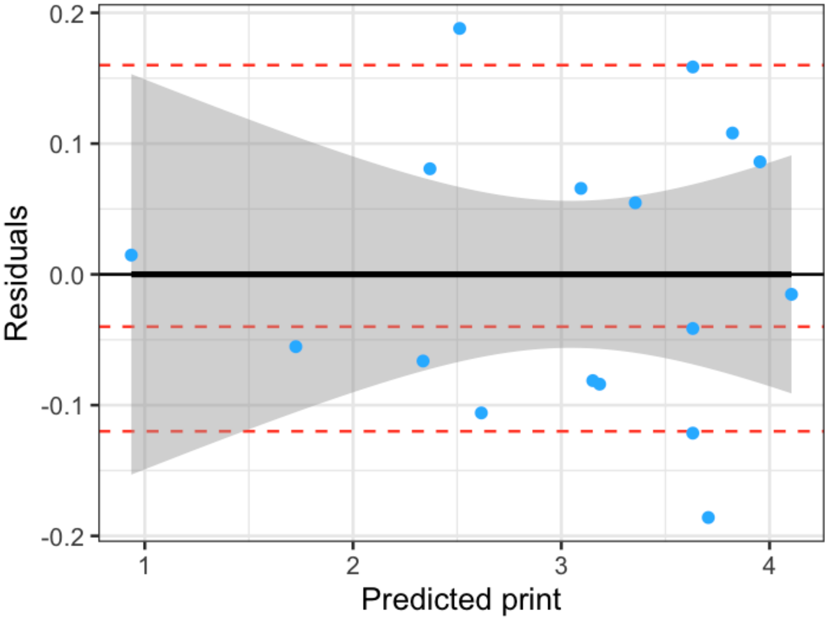
Modélisation de la réponse « Print »

```
Print ~ Temperature + Time + Pressure + Temperature:Time + Temperature:Pressure
+ Time:Pressure + I(Temperature^2) + I(Time^2) + I(Pressure^2)


Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)          -7.831e+00  2.363e+00  -3.314  0.01288 *
Temperature           1.574e-01  3.274e-02   4.807  0.00195 **
Time                  1.041e+00  4.314e-01   2.412  0.04665 *
Pressure              1.122e-02  9.748e-03   1.151  0.28735
I(Temperature^2)     -5.805e-04  1.080e-04  -5.375  0.00104 **
I(Time^2)            -5.092e-01  1.200e-01  -4.243  0.00383 **
I(Pressure^2)        -2.499e-05  3.888e-05  -0.643  0.54097
Temperature:Time     -3.843e-03  2.084e-03  -1.844  0.10768
Temperature:Pressure  1.083e-05  3.750e-05   0.289  0.78106
Time:Pressure        -1.389e-04  1.250e-03  -0.111  0.91466
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1591 on 7 degrees of freedom
Multiple R-squared:  0.9852,      Adjusted R-squared:  0.9661
F-statistic: 51.66 on 9 and 7 DF,  p-value: 1.422e-05
```

On voit que pour la qualité visuelle de la soudure, la pression ne semble pas importante non plus. Et comme pour le « bond » il semble qu’on ait un bon modèle. Il faut quand même le vérifier.

Modélisation de la réponse « Print »



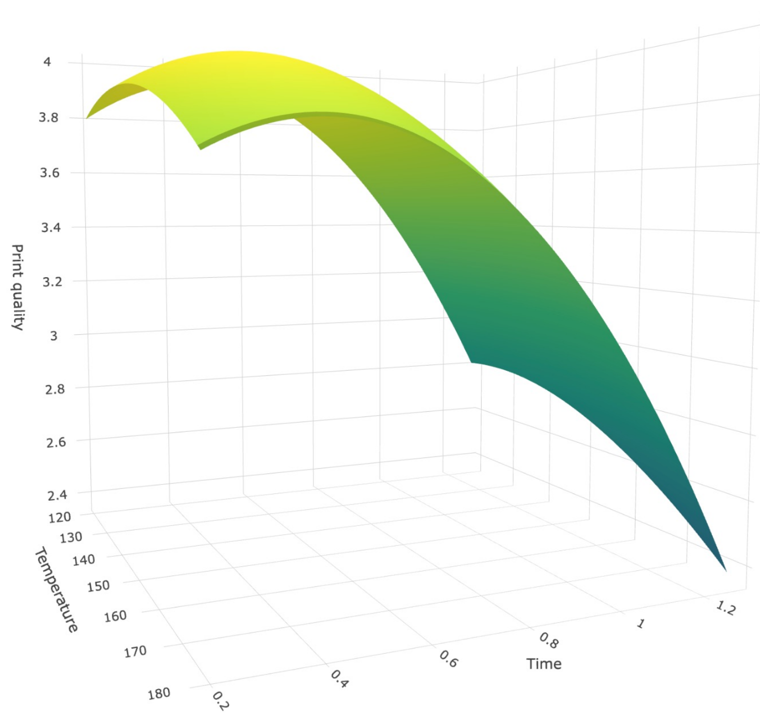


source	df	SSE	MSE	F_stat	p_value
pure error	2	0.0416	0.0208	NA	NA
lack of fit	5	0.1356	0.02713	1.304	0.3501
model residuals	7	0.1772	0.02532	NA	NA

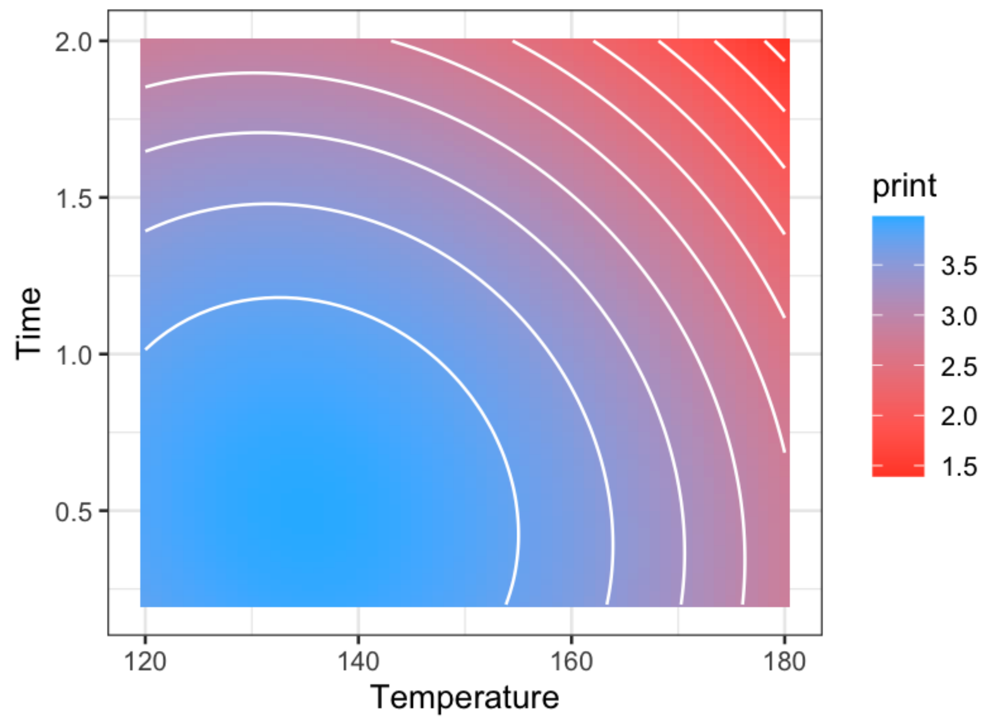
Le modèle semble bien fitter les données.

Visualisation du modèle « Print »

Surface de réponse



Courbes de niveau



Pour le print, puisqu'on cherche à le maximiser, il n'y a qu'un seul point considéré comme optimum.

Optimisation par recherche de zone de compromis

En général les optima individuels des réponses sont contradictoires et il faudra rechercher une ou plusieurs solutions de compromis.

Principe simple d'optimisation:

Donner pour chaque réponse un objectif et une zone d'acceptation.
Rechercher graphiquement les **zones qui répondent aux critères.**

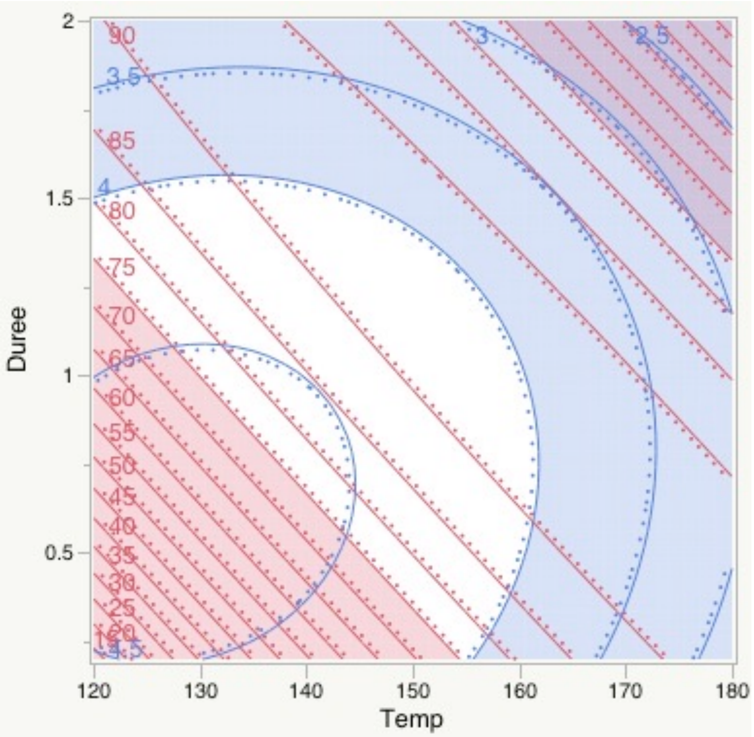
Exemple :

Bond : cible = 85. Accepté à 85 ± δ,
Print : à maximiser avec un min de μ

Optimum proposé

Temp = 140, Durée = 1, Pression=150
Bond prédit = 84.8   Print prédit = 4.5
Intervalles de prédiction à 95% :
[79.2,90.3]   et   [4.12, 4.87]

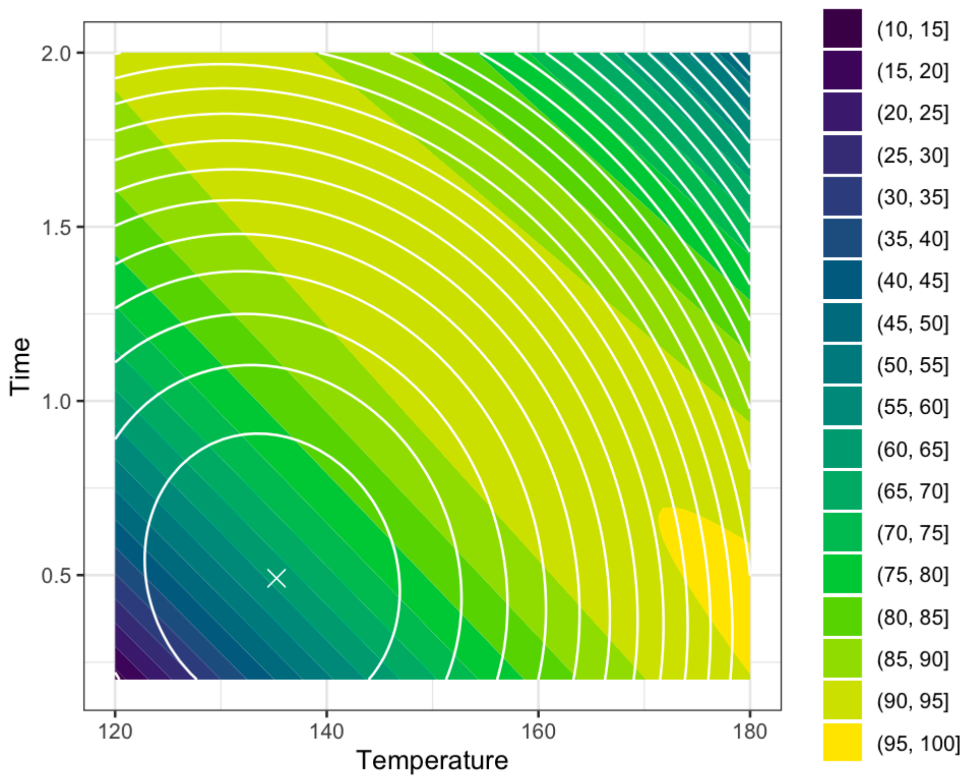
Bond et Print avec Pression = 150

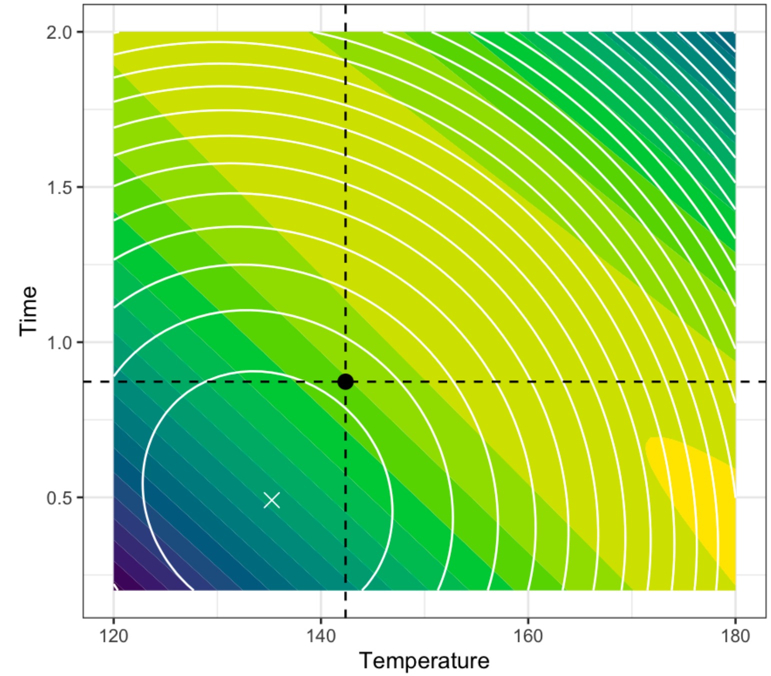


Optimisation par recherche de zone de compromis

En général les optima individuels des réponses sont contradictoires et il faudra rechercher une ou plusieurs solutions de compromis.

Couleurs = bond; contours blanc = print (max sur la croix) pour pression = 100





Optimum proposé (pour ces données)

Temp = 142, Durée = 0.9, Pression = 100

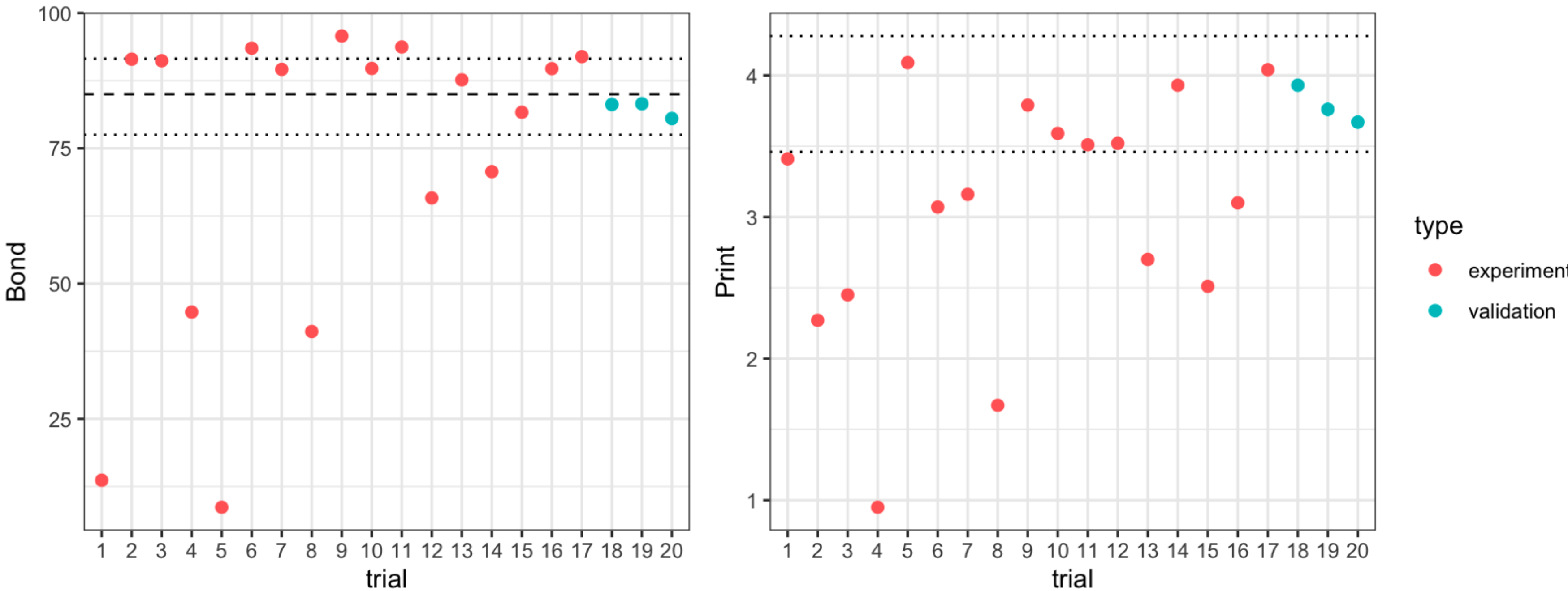
Bond prédit = 84.5   Print prédit = 3.9

Intervalles de **prédiction** à 95% :

[77.49, 91.56]   et   [3.46, 4.28]

Validation

Dans le cas où on a gardé 3 essais pour « valider » nos résultats, on peut répéter 3 essais à l’optimum trouvé et examiner nos mesures.



Pointillés: intervalles de prédiction