**Faculté des Sciences Pharmaceutiques de Tours.**

*Laboratoire de Biophysique et Mathématiques*

# LES

# PLANS FACTORIELS

# COMPLETS

## Cours et énoncés d'exercices

Mai 2009 Claude HOINARD

# TABLE DES MATIERES

# GENERALITES CONCERNANT LES PLANS EXPERIMENTAUX

## *Introduction*

- **Qu'est ce qu'un plan expérimental ?**

- Une expérience est une intervention volontaire dans un système en fonctionnement pour mesurer les effets de cette intervention. L'expérimentation diffère de l'enquête par le fait que dans cette dernière on se contente d'observer passivement le fonctionnement du système. Seule l'expérience est capable d'apporter des renseignements sur les relations de causes à effets.
- Le plan d'expériences correspond à toutes les mesures d'organisation prises pour effectuer l'étude ; il comprend le schéma expérimental, définit de façon précise les méthodes, les appareils utilisés, les modes opératoires, les produits, les jours de manipulation, les opérateurs....

- **Un exemple d'étude de procédé et nécessité d'une stratégie adaptée à l'objectif : notion de plan efficace.**

Il n'y a pas si longtemps, le principe de base de l'expérimentation était « de ne faire varier qu'un facteur à la fois » ; Il est important de noter que les résultats de cette stratégie, certes sécurisante, sont médiocres. Il a été démontré qu'il est plus efficace de faire varier simultanément plusieurs facteurs à la condition de respecter certaines règles qui font précisément l'objet de cet enseignement.

### Le message de cette introduction

> Pour acquérir des connaissances précises sur un phénomène ou un procédé, la méthode des plans multifactoriels est celle qui apporte le plus d'efficacité.

## PETIT HISTORIQUE DES PLANS D'EXPERIENCES

- Les bases datent des années 1925 (R.A. FISHER).
- Dans le secteur Industriel, technique restée confidentielle jusqu'après la deuxième guerre mondiale ; développement au Japon entre les années 1950 et 1960 grâce notamment à TAGUCHI ; ses travaux sont diffusés aux Etats Unis et appliqués à grande échelle vers les années 1970.
- En Europe, les grandes Entreprises industrielles utilisent les plans expérimentaux à partir des années 1980. Actuellement cette méthode d'expérimentation est utilisée par l'ensemble des Industriels, petits et grands.

La méthode est maintenant considérée comme un **OUTIL DE LA QUALITE**, indispensable

- au stade de la conception des produits pour permettre de fixer les paramètres de développement du produit de façon optimale.
- au cours de la production, comme complément indispensable des méthodes de maîtrise des procédés.

| |
|---|
| « Une fois comprise, la méthode constitue une étape irréversible dans la carrière d'un technicien ou d'un ingénieur qui ne pourra plus envisager de réaliser des essais sans utiliser un plan d'expériences. »<br>RYAN (1990 - Conférence sur la Qualité) |

# LE VOCABULAIRE DE BASE

## 1- Facteur et Réponse.

- Un facteur est un paramètre ou un état du système étudié ( phénomène ou procédé ) dont la variation est susceptible de modifier le fonctionnement de ce système.
- La réponse (ou critère) du système correspond au paramètre mesuré ou observé pour connaître l'effet des facteurs étudiés sur le système. Nous n'envisagerons dans cet enseignement que le cas des réponses quantitatives mesurables.

## 2- Facteurs quantitatifs et qualitatifs ; niveaux (modalités).

- Un facteur quantitatif est un paramètre expérimental qui s'exprime par un nombre ; les niveaux du facteur correspondent aux différentes valeurs que l'expérimentateur décide de choisir dans son étude.
- Un facteur qualitatif se caractérise par un ensemble discontinu d'états ; les états choisis dans l'expérimentation sont appelés les modalités - on dit aussi les niveaux - du facteur étudié.

Dans un même plan expérimental, on peut étudier simultanément des facteurs quantitatifs et qualitatifs.

## 3- Plans multifactoriels dits factoriels.

Un plan est dit factoriel quand, <u>dans une même expérimentation</u>, il est étudié plusieurs facteurs dont les niveaux sont croisés.
Il est dit complet si tous les croisements possibles figurent dans l'expérimentation.

- *Nombre de traitements d'un plan factoriel complet*
  Un traitement est un <u>ensemble de conditions expérimentales définissant les conditions d'un essai</u> ; elles correspondent à un niveau pour chacun des facteurs.

  Pour obtenir le nombre total de traitements différents à expérimenter, il suffit de multiplier le nombre de niveaux de tous les facteurs étudiés.

- *Notation des plans factoriels complets.*
  Un plan $2*2$ ou $2^2$ signifie qu'on étudie 2 facteurs A et B chacun à 2 niveaux : il définit 4 traitements.
  Un plan $2^8$ correspond à l'étude de 8 facteurs chacun à 2 niveaux : représente 256 traitements.
  Un plan $3*2*4$ correspond à l'étude de 3 facteurs : A est à 3 niveaux, B en comporte 2 et C, 4 ; le plan comporte 24 traitements.

- *Nombre d'essais à effectuer ; cas où il existe des répétitions.*
  Quand chaque traitement est expérimenté une seule fois, alors :

  <u>nombre d'expériences = nombre de traitements</u>

  C'est un cas fréquent dans les essais industriels et au laboratoire de Contrôle.

  Pour certains types d'études, notamment biologiques, chaque traitement est expérimenté plusieurs fois ; on dit qu'il est fait des <u>répétitions</u>. Il est important de prévoir un même nombre **r** de répétitions pour tous les traitements du plan afin de respecter le bon équilibre du plan.
  Il vient alors :

  <u>nombre d'expériences = nombre de traitements * **r**</u>

  La réponse à un traitement est constituée dans ce cas par la moyenne des réponses des répétitions.

### 4- Facteurs étudiés (ou contrôlés) et facteurs non contrôlés dans le plan.

Les facteurs agissant sur la réponse du système sont plus ou moins nombreux selon la nature du système ; supposons les en nombre **k**, nombre évidemment inconnu de l'expérimentateur.

Dans un plan factoriel, on n'expérimente qu'un nombre limité, **n**, de facteurs dont l'influence sur la réponse sera quantifiée.

Cela signifie que les autres facteurs, en nombre **k - n**, inconnus de l'expérimentateur ou négligés par lui, sont susceptibles de changer d'état ou de niveau d'un essai à l'autre au cours de la manipulation et donc de faire fluctuer la valeur de la réponse : ces facteurs non contrôlés créent une variabilité de la réponse y.

Un des fondements des plans expérimentaux consiste à supposer que les fluctuations de la réponse **y** pour chaque traitement sont distribuées selon une loi Normale.

Pour essayer de diminuer la variabilité de la réponse, on a donc intérêt à « contrôler » le plus de facteurs possibles dans un même plan, soit qu'on les étudie, soit qu'on programme de les maintenir constants pendant l'expérimentation.

D'autre part, il faut aussi éviter que des facteurs non contrôlés aient la possibilité de jouer de manière systématique pendant le déroulement des expériences en créant des variations chronologiques systématiques (des dérives) de la réponse y ⇒ **randomisation** de l'ordre des essais (répartition au hasard).

# ASPECTS ALGEBRIQUES ET STATISTIQUES DES PLANS EXPERIMENTAUX

### 1- Aspects algébriques.

L'analyse d'un plan factoriel complet comportant au total n traitements nécessite la mesure de n réponses.

Ces n réponses permettent de définir n paramètres issus du modèle mathématique associé au plan :

- un de ces paramètres est la moyenne générale des n réponses mesurées.
- les n-1 autres paramètres sont appelés des **effets**.

En ce qui concerne ces effets, deux problèmes essentiels se posent : comment les calculer ?
Et, quelle signification concrète peuvent-ils avoir ?

Calcul.

Les valeurs des n paramètres sont les solutions d'un système de n équations simultanées du premier degré comportant n inconnues ;

⇒ des notions de calcul matriciel facilitent beaucoup la résolution de tels systèmes d'équations d'où une introduction aux bases de calcul matriciel dans cet enseignement.

Interprétation.

Donner une « signification concrète » à des paramètres effets définis de façon abstraite à partir du modèle n'est possible que pour quelques-uns ; d'abord ceux qui ne font intervenir qu'un facteur (on les appelle des **effets principaux**) et ils ont en général les valeurs numériques les plus importantes ; notons en passant qu'un facteur est dit sans action si son effet est nul (E=0) et que l'effet d'un facteur actif peut être positif (il fait augmenter la réponse quand le facteur passe du niveau inférieur au niveau supérieur) ou négatif, selon le sens de son action sur la réponse.

Ensuite, il est souvent considéré les effets qui font intervenir simultanément deux facteurs (ils sont appelés **interactions de 2$^{ème}$ ordre**) ; ces interactions complètent utilement les aspects pratiques des actions des facteurs.

En revanche, les autres effets, mettant en jeu simultanément trois facteurs et plus (**interactions** dites **d'ordres supérieurs** ), ne présentent la plupart du temps pas d'intérêt pratique lors de l'interprétation dans le domaine de l'expérimentation industrielle.

Remarquons que comme les n paramètres effets ne sont pas tous nécessaires lors de l'analyse et l'interprétation, il est superflu de disposer de n réponses expérimentales. C'est cette constatation qui a été à l'origine de la création des **plans factoriels** dits **fractionnaires** dans le but d'économiser la charge de travail expérimental.

Pour étudier, par exemple, 6 facteurs à 2 niveaux dans un même plan, on peut n'expérimenter que 32 traitements, voire 16 ou même 8 au lieu des 64 nécessités par le plan complet. La réduction du nombre d'expériences rend évidemment ces plans attrayants, mais il faut aussi préciser dès maintenant que l'interprétation peut être délicate et que la génération de ces plans est plus complexe : elle doit obéir à des règles de symétrie précises.

## 2- Aspects statistiques.

La répétition d'un même traitement ne conduit généralement pas exactement à la même valeur de la réponse mesurée. L'origine de cette fluctuation est double :

- imprécision de l'instrument de mesure qui a permis d'obtenir les réponses.
- influence des variations des niveaux des facteurs non contrôlés dans le plan entre les répétitions ; c'est souvent cette source de variation qui est prépondérante dans les essais industriels.

Les effets, calculés à partir des réponses mesurées, sont donc entachés de variabilité ; on considère que, comme pour les réponses, leurs fluctuations sont aléatoires et normalement distribuées.

Se pose alors la question fondamentale de savoir à partir de quelle valeur seuil, il faut considérer qu'un effet calculé a une valeur sûrement différente de 0. En d'autres termes, à partir de quelle limite peut-on considérer que le facteur (ou l'interaction) agit réellement ?

Ce sont les tests statistiques qui apportent ce type de renseignements : **test t** de comparaison de 2 moyennes dans les cas les plus simples, **analyse de variance à plusieurs facteurs croisés** (**ANOVA**) le plus souvent.

Ajoutons que la modélisation des résultats, nécessaire pour certaines applications, utilise comme outil privilégié la **régression linéaire multiple**.

> ⇒ Toutes ces considérations justifient l'approfondissement des connaissances statistiques dans cet enseignement.

# APERCU SUR LES APPLICATIONS DES PLANS FACTORIELS

Pour des objectifs définis, il existe des plans expérimentaux adaptés, efficaces et économiques. Citons par exemple :

## ⇒ La recherche (exhaustive) des facteurs agissant sur un procédé.

**L'idée force** : pour bien utiliser un procédé, il faut en avoir la « maîtrise », c'est à dire connaître tous les facteurs ayant une action sur la(les) réponse(s) (sens et intensité) du procédé ou du phénomène ainsi que les interactions de 2$^{ème}$ ordre.

Dans ce genre d'étude, il s'agit préalablement d'essayer de deviner tous les facteurs susceptibles d'avoir un effet puis de les tester dans une expérience multifactorielle pour identifier les facteurs provoquant des variations de réponse significatives…

Ce type d'expérimentation est dit de « **criblage des facteurs** » et s'effectue chronologiquement avant d'utiliser le procédé en routine. Il intervient aussi lors de la production pour « affiner » les réglages du procédé (il sera alors programmé de petites variations pour ne pas compromettre la qualité de la production).

Dans une telle expérimentation il y a presque toujours un grand nombre de facteurs (parfois plus d'une dizaine) et pour que le nombre d'essais ne soit pas prohibitif, il est intéressant et économique d'utiliser :

- deux niveaux seulement par facteur.
- des plans factoriels fractionnaires.

C'est ainsi qu'on pourra avoir une bonne connaissance des facteurs actifs et de leurs éventuelles interactions, ce qui pourra conduire, par exemple, à ajouter un thermostat au système pour minimiser l'action de la température si celle-ci a un effet important, à le faire fonctionner en atmosphère contrôlée si l'humidité ambiante joue un rôle néfaste….etc.

## ⇒ La recherche d'un optimum de réponse.

Il s'agit de **méthodes** dites **d'optimisation** dans lesquelles on recherche les valeurs des niveaux des facteurs qui conduisent à une réponse maximale ou minimale selon le cas.

- Cette optimisation est assez simple à effectuer dans le cas de **facteurs qualitatifs** : il suffit de rechercher parmi les modalités expérimentées celle qui conduit à la réponse la plus convenable (la plus grande ou la plus petite selon le cas). Un plan factoriel complet comportant jusqu'à 4 ou 5 facteurs convient parfois pour ce type d'étude si le nombre de modalités étudiées n'est pas trop grand ; dans le cas contraire, il est utilisé des plans factoriels fractionnaires.

- Avec des **facteurs tous quantitatifs**, il faut être capable de prédire numériquement la réponse en fonction des valeurs des niveaux des facteurs : cette opération s'appelle une modélisation. L'équation est nécessairement du 2$^{ème}$ degré pour être capable de présenter un maximum ou un minimum. Il faut donc que chaque facteur soit expérimenté sous 3 niveaux ; pour de telles études, les plans utilisés font intervenir un nombre limité de facteurs (pas plus de 4 ou 5) pour ne pas trop alourdir l'expérimentation et compliquer l'analyse ainsi que l'interprétation des résultats.

Ajoutons en ce qui concerne l'optimisation, qu'il existe, en concurrence avec les méthodes planifiées à l'avance et nécessitant une modélisation dont nous venons de donner un aperçu, des techniques d'optimisation itératives plus directes, plus simples à utiliser et interpréter. Elles ont pour nom : **méthode uniplex** quand on ne fait varier qu'un facteur dans l'optimisation et **méthode simplex** quand on fait varier plusieurs facteurs simultanément. Ces méthodes ont prouvé leur efficacité pour localiser un optimum.

### ⇒ La recherche de nominale.

Il s'agit de construire des plans dans lesquels on recherche les valeurs des niveaux des facteurs qui conduisent à une réponse égale à une **valeur cible** qui n'est ni un maximum ni un minimum.
Pour concrétiser ce cas, citons l'exemple de la recherche des réglages d'une machine destinée à fabriquer des bidons de plastique dont la contenance doit être exactement de 1 litre (la nominale).

Les plans utilisés sont souvent des plans factoriels complets comportant un nombre restreint de facteurs (au maximum 3 ou 4). Ces facteurs doivent être quantitatifs avec un nombre de niveaux suffisant pour « quadriller » convenablement le domaine expérimental. Ce qui permet d'estimer l'équation des variations de la réponse en fonction des facteurs et d'éventuelles interactions (modélisation) et ainsi de prédire les caractéristiques numériques des facteurs conduisant à la valeur cible.

## *En guise de conclusion…*

Ce qui précède ne constitue que quelques aperçus de plans permettant de résoudre des problèmes se rencontrant fréquemment dans le domaine industriel. Précisons que, plus généralement, pour chaque objectif défini il a été développé un plan « efficace » qui figure dans l'abondante littérature des plans expérimentaux. Prenons un seul exemple : les plans de mélange (dits de Scheffé) sont utiles à considérer pour un pharmacien industriel travaillant dans le domaine des pâteux.

# LES PLANS FACTORIELS COMPLETS

## *Présentation.*

Ce chapitre est principalement consacré aux plans factoriels $2^n$. En fin de chapitre, nous envisagerons le cas des plans factoriels complets dans lesquels des facteurs sont étudiés sous plus de deux niveaux.

Dans un plan factoriel complet $2^n$, n facteurs ( qualitatifs et / ou quantitatifs ) sont étudiés dans une même expérimentation, chacun à 2 niveaux ( souvent appelés niveau BAS et niveau HAUT ). Toutes les combinaisons des niveaux des facteurs doivent être étudiées : elles sont en nombre $2^n$ définissant autant de traitements différents (les conditions expérimentales).

Le nombre d'essais à réaliser est donc aussi égal à $2^n$ s'il n'est pas prévu de répéter chaque traitement ; dans le cas contraire, en appelant **r** le nombre de répétitions - qui doit être le même pour tous les traitements - le nombre total d'essais à réaliser dans le plan expérimental est :

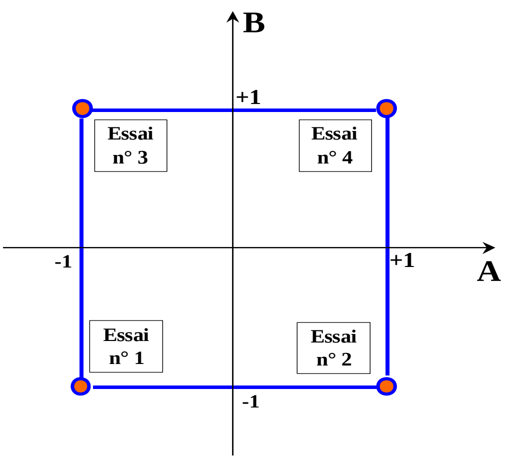
$$N = \mathbf{r} * 2^n$$

Un plan $2^n$ permet de mettre en évidence, parmi les n facteurs étudiés, ceux qui agissent sur la réponse mesurée **y** ; la méthode permet en plus d'évaluer diverses interactions susceptibles de se produire entre les actions des facteurs.

## **1-** DEFINITION DES EFFETS PRINCIPAUX ET DES INTERACTIONS – CALCULS.

### 1-1 <u>Le plan $2^2$</u> (facteurs étudiés : A et B)

- **domaine de variation des facteurs** ; modalités pour les facteurs qualitatifs ; valeurs (niveaux) pour les facteurs quantitatifs ; facteurs centrés réduits (notation de Yates).
  *cf exercice n°1 ( questions 1, 2 )*




Remarquons que les 4 points expérimentés sont aux extrémités du domaine expérimental qui est dans ce cas un carré.

### ❑ matrice d'expériences.

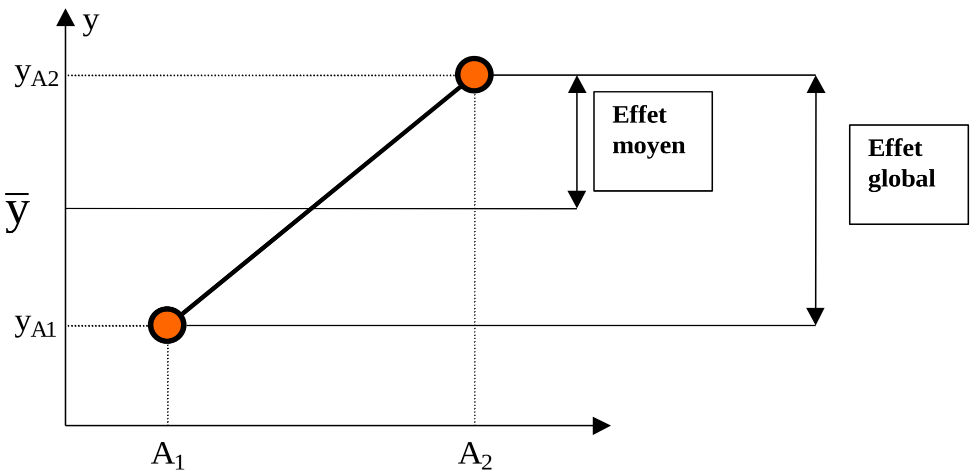
*cf exercice n°1 ( question 3 )*

Ce tableau récapitule les essais à effectuer : c'est le plan de manipulation. Il présente 4 lignes, une par essai, et 2 colonnes, une par facteur, indiquant les niveaux de chaque facteur dans l'essai (-1= niveau BAS ; +1= niveau HAUT).
Les essais sont identifiés par des numéros de 1 à 4.

### ❑ effet principal d'un facteur : expressions.

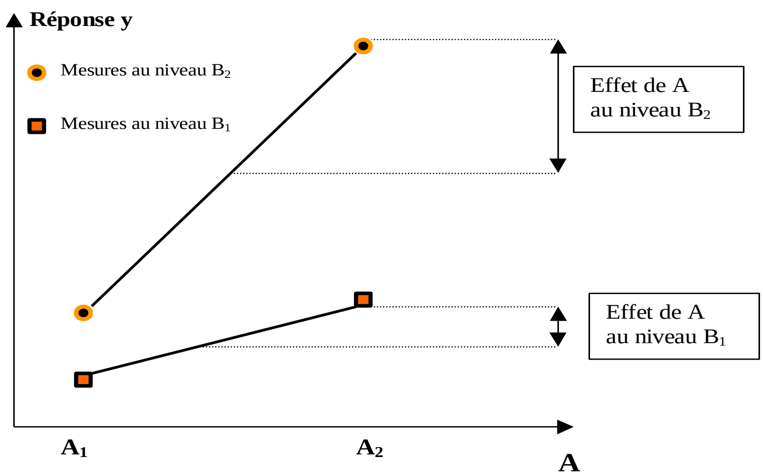
**EFFET GLOBAL** ⇒ il est évalué par la variation de réponse quand le facteur passe du niveau BAS au niveau HAUT. Sa signification est évidente ; un effet peut être positif (la réponse augmente) ou négatif (la réponse diminue).

**EFFET MOYEN** ⇒ différence entre la réponse y quand le facteur est au niveau HAUT et la moyenne générale $\overline{y}$ des réponses ; sa valeur est égale à la moitié de l'effet global comme on peut le voir sur le schéma suivant. C'est lui qui sera utilisé dans la suite de ce document.




### ❑ interaction entre 2 facteurs ; quantification.

Il existe une interaction lorsque l'effet d'un facteur dépend du niveau de l'autre facteur.



Le schéma précédent illustre la présence d'une interaction entre les facteurs A et B : l'effet de A au niveau $B_2$ n'est visiblement pas égal à l'effet de A au niveau $B_1$.

Une interaction est une action mutuelle : l'effet de B dépend réciproquement du niveau de A

Aspect quantitatif.

L'effet d'une interaction est évalué par la demi différence entre l'effet moyen d'un facteur au niveau HAUT de l'autre et l'effet moyen de ce même facteur au niveau BAS de l'autre.

### ❑ signification de l'effet d'un facteur quand il y a interaction.

Lorsqu'une interaction est présente, l'effet principal d'un facteur perd sa signification immédiate quand on le considère indépendamment de l'autre facteur puisqu'il dépend du niveau de ce dernier. Numériquement, sa valeur correspond à l'effet de ce facteur quand l'autre facteur est au niveau 0.

### ❑ pratique des calculs avec la notation de Yates : matrice des effets.

*cf exercice n°1 ( question 4 ) et exercice n°2 ( question 1 )*

La matrice des effets est conçue pour calculer les 4 paramètres que l'on peut extraire des résultats de l'expérimentation ; par rapport à la matrice d'expérience, elle comprend les colonnes supplémentaires suivantes :

- 1 colonne pour l'interaction AB, -1 ou +1 selon le produit des colonnes A et B.
- 1 colonne pour le calcul de la moyenne des réponses y, ne contenant que des +1.
- 1 colonne pour les valeurs individuelles des réponses y.

Les formules de calcul des paramètres sont :

$E_A = 1/4\ ( - y_1 + y_2 - y_3 + y_4 )$ effet principal de A

$E_B = 1/4\ ( - y_1 - y_2 + y_3 + y_4 )$ effet principal de B

$E_{AB} = 1/4\ ( + y_1 - y_2 - y_3 + y_4 )$ effet de l'interaction AB

$\bar{y} = 1/4\ ( + y_1 + y_2 + y_3 + y_4 )$ moyenne générale des réponses

Les séquences des facteurs centrés réduits (-1 ; +1) des colonnes de la matrice des effets pour les 4 essais reproduisent exactement celles des formules précédentes. Le calcul d'un paramètre effet se présente donc comme la somme des produits de la colonne réponse avec la colonne effet, divisée par 4.

## 1-2 <u>Le plan $2^3$</u> (facteurs A, B et C)

### ❑ domaine expérimental ; matrice d'expériences.

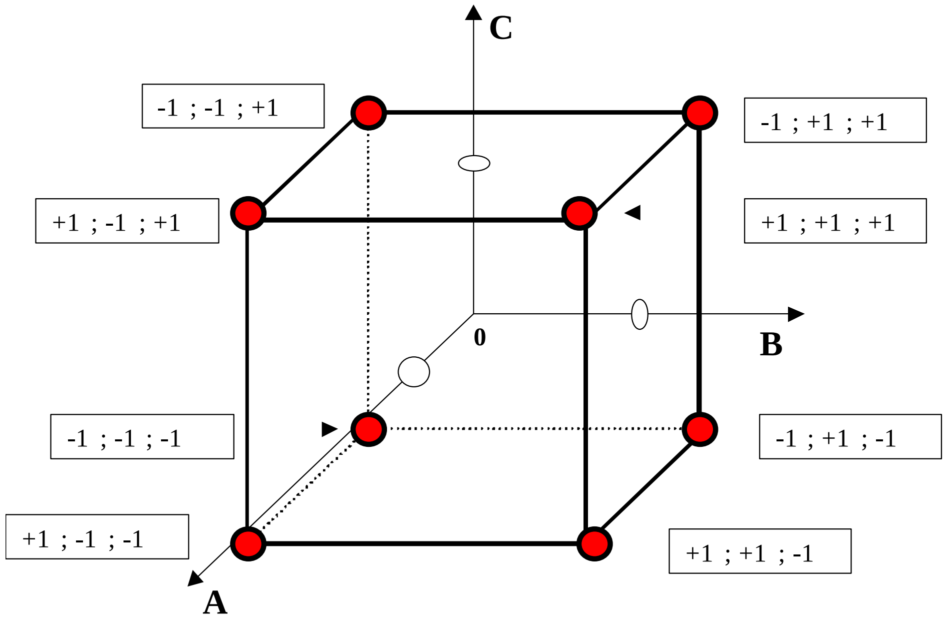
*cf exercice n°3 ( question 1 )*

Le tableau des 8 essais (un essai par ligne) contient maintenant 3 colonnes, une par facteur.

Le schéma de la page suivante fait apparaître que les 8 points expérimentaux se situent aux extrémités du domaine, dans ce cas les sommets d'un cube.

**Cette disposition « aux extrémités » est générale pour tous les plans factoriels dont les facteurs sont expérimentés à 2 niveaux.**

C'est elle qui procure les avantages essentiels des plans factoriels : **indépendance des effets calculés** les uns par rapport aux autres (propriété dite d'orthogonalité), **précision optimale** de ces estimations d'effets.




Il est possible de montrer mathématiquement que toute autre disposition des points expérimentaux est moins efficace pour obtenir les effets.

*Voir à ce propos l'exercice 2 ( question 3).*

## effets des actions ( facteurs et interactions ) ; signification.

3 effets principaux ( A, B, C ).
3 interactions de 2$^{ème}$ ordre ( AB, AC, BC ) mettant en jeu 2 facteurs.
1 interaction de 3$^{ème}$ ordre ( ABC ) mettant en jeu les 3 facteurs.

Les formules de calcul des paramètres à partir des 8 réponses s'écrivent :

$$E_A = 1/8( - y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8 )$$
$$E_B = 1/8( - y_1 - y_2 + y_3 + y_4 - y_5 - y_6 + y_7 + y_8 )$$
$$E_C = 1/8( - y_1 - y_2 - y_3 - y_4 + y_5 + y_6 + y_7 + y_8 )$$
$$E_{AB} = 1/8( + y_1 - y_2 - y_3 + y_4 + y_5 - y_6 - y_7 + y_8 )$$
$$E_{AC} = 1/8( + y_1 - y_2 + y_3 - y_4 - y_5 + y_6 - y_7 + y_8 )$$
$$E_{BC} = 1/8( + y_1 + y_2 - y_3 - y_4 - y_5 - y_6 + y_7 + y_8 )$$
$$E_{ABC} = 1/8( - y_1 + y_2 + y_3 - y_4 + y_5 - y_6 - y_7 + y_8 )$$
$$\overline{y} = 1/8( + y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7 + y_8 )$$

## pratique des calculs : matrice des effets.

*cf exercice n°3 ( questions 1 et 2 ).*

La matrice des effets contient la colonne des réponses y et 8 colonnes correspondant chacune à un paramètre.

La méthode de calcul d'un effet est la même que pour le plan $2^2$ : on effectue la somme des produits des colonnes réponse et effet, puis on divise par 8.

## 1-3 Les plans $2^n$.

- **conditions expérimentales ; nombre d'effets d'actions à calculer** (effets principaux et interactions de divers ordres ).

Généralisation des plans précédents, l'expérimentation comprend $2^n$ traitements différents et autant d'essais.

En négligeant la moyenne des réponses mesurées, paramètre particulier, ces plans permettent de définir et de calculer $2^n - 1$ effets d'actions :

- n effets principaux.
- $C_n^2$ interactions du 2ème ordre.
- le reste est constitué par des interactions dites d'ordres supérieurs, en nombre $C_n^3$ pour les interactions d'ordre 3, …etc.

Il est difficile de donner une signification concrète aux interactions d'ordres supérieurs qui ne correspondent mathématiquement qu'à des termes correctifs provenant de la résolution du système d'équations ; il est considéré généralement que ces interactions n'ont pas d'existence réelle. Il leur est donc attribué un effet théorique nul. Leurs effets calculés sont cependant différents de 0 par suite de la dispersion expérimentale et sont utiles à connaître pour des raisons statistiques ; nous y reviendrons.

En pratique, l'analyse et l'interprétation sont focalisées essentiellement sur les effets principaux et les interactions d'ordre 2.

Remarquons enfin que, du point de vue manipulation, ces plans présentent l'inconvénient d'être vite très « consommateurs » d'expériences : chaque fois que n augmente d'une unité, le nombre d'essais est multiplié par deux.

- **matrices d'expériences et des effets ; pratique des calculs d'effets.**

*exercice n°4 ( question 1 ).*

La matrice d'expériences comporte n colonnes, une par facteur. Elle ne comprend que des 1 accompagnés des signes – ou +.

- Pour le facteur A, on a une suite de signes – et + alternés, commençant par un signe -.
- La suite des signes du facteur B est constituée de deux signes -, suivis de deux signes +, puis deux signes -, à nouveau deux signes + …etc.
- Pour le facteur C, la suite des signes est constituée de quatre signes -, suivis de quatre signes +, …etc.
- Les facteurs suivants ont 8, 16, 32… signes -, suivis de 8, 16, 32 … signes +.

Pour chaque facteur, il y a autant de signes – que de signes +.

Pour construire la matrice des effets, il faut ajouter $2^n$–n colonnes, correspondant au nombre total d'interactions et à la moyenne. Le contenu de chaque colonne est obtenu en multipliant les valeurs –1 ou +1 des facteurs participant à l'interaction pour chaque essai. Il faut aussi créer la colonne des moyennes qui ne contient que des +1 et la colonne des réponses y.

La technique de calcul des effets est analogue à celle utilisée pour les plans $2^2$ et $2^3$.

- **calcul des effets quand le plan comporte des répétitions.**

*exercice n°4 ( question 2 ).*

Lorsque le plan comporte des répétitions, **la réponse représentative d'un traitement est constituée par la moyenne des réponses individuelle obtenues pour ce traitement**. Il faut donc préalablement calculer les $2^n$ moyennes. Puis on effectue le calcul des effets avec la matrice des effets en utilisant comme réponses les moyennes précédemment déterminées.

Il est important que chaque moyenne représente le même « poids » dans le calcul des effets, **il faut donc que le nombre de répétitions soit le même** pour toutes les conditions expérimentales.

# 2- ANALYSE DES RESULTATS : QUELS EFFETS SONT SIGNIFICATIFS ?

Par suite de la variabilité des réponses expérimentales, un effet sans action n'a pas une valeur calculée exactement égale à 0 ; on peut cependant espérer que son résultat ne s'écartera pas trop de 0. La question se pose donc de savoir **à partir de quelle valeur seuil**, on peut considérer qu'**un effet est significatif**, autrement dit que le facteur ou l'interaction a une action réelle.

## *Relation fondamentale* :

- Entre l'écart type $\sigma_y$ des n réponses individuelles y, supposé le même en tous les points du domaine expérimental et l'erreur type $\sigma_E$ commune à tous les effets, il existe la relation :

$$\sigma_E = \frac{\sigma_y}{\sqrt{n}}$$

- Dans la pratique, divers cas peuvent se produire mais le plus souvent $\sigma_y$ sera remplacé par son estimation $S_y$ . On écrit alors : $S_E = \frac{Sy}{\sqrt{n}}$

La solution du problème de signification des effets est statistique ; il y a autant de tests à effectuer qu'il y a d'effets estimés. Un avantage des plans factoriels à 2 niveaux est que chaque test de comparaison d'un effet calculé à la valeur 0 est mathématiquement indépendant des autres tests (les comparaisons sont dites orthogonales). On peut ainsi utiliser une série de tests z (loi normale) ou de tests t (loi de Student) selon les informations dont on dispose concernant la variabilité de la réponse.

## 2-1 Estimation des effets significatifs : $\sigma$y est connu par des expériences antérieures.

*cf exercice n°3 ( question 3 ).*

Les conditions des expériences antérieures doivent évidemment être comparables à celles de l'expérimentation actuelle et en nombre suffisant pour que $\sigma_y$ soit considéré comme connu avec précision. Dans ces conditions, on calcule $\sigma_E$ et chaque effet estimé est comparé à la valeur théorique 0 par un test fondé sur l'emploi de la loi normale. Un effet est déclaré significatif au risque 0,05 si la valeur absolue calculée de z est supérieure à 1,96.

## 2-2 Estimation des effets significatifs : $\sigma$y n'est pas connu, mais il est effectué des essais complémentaires.

- **quelques essais complémentaires en un même point à l'intérieur du domaine expérimental** (de préférence au centre du domaine ).
  *Cf exercice n°2 ( question 2 ).*

Lors de l'élaboration du plan, il a été prévu un nombre N de répétitions (essais correspondant à un même point à l'intérieur du domaine). Il est intéressant de choisir le centre du domaine (point 0 des facteurs centrés réduits) lorsque les facteurs sont quantitatifs mais ce n'est pas une obligation ($\sigma_y$ est supposé constant dans le domaine). Le nombre N doit être au minimum égal à 5 pour que l'estimation Sy ne soit pas trop imprécise. Il est ensuite calculé $S_E$.

Les tests des comparaisons des effets à 0 utilisent la loi de Student à N-1 degrés de liberté. Tout effet dont le test t conduit à un degré de signification inférieur à 0.05 est déclaré significatif.

- **répétition entière du plan factoriel.**
  *Cf exercice n°4 ( question 3 ).*

On dispose alors de groupes de résultats comprenant chacun plusieurs valeurs comparables (un groupe correspond à un traitement). Il faut estimer une variance $S^2y$ commune à l'ensemble des groupes (par addition des SCE et des ddl de chaque groupe), puis Sy ; après avoir calculé SE, il est procédé aux comparaisons des effets calculés à la valeur 0 au moyen de tests de Student ; le nombre de degrés de liberté de chaque test est égal à celui qui a permis d'estimer $S^2y$.

On peut aussi utiliser l'analyse de variance, méthode équivalente aux multiples tests t dans les plans factoriels puisque les effets sont tous indépendants les uns des autres ; les 2 méthodes donnent des degrés de signification identiques pour les effets.

## 2-3 Estimation des effets significatifs : $\sigma y$ n'est pas connu et il n'a pas été effectué d'essais complémentaires.

- **principe de l'estimation de $S_E$ à partir des interactions d'ordres élevés.**
  (à partir du 3ème ordre).
  *cf exercice n°4 ( question 4 ).*

Dans ce cas, on prend comme hypothèse que les interactions d'ordres élevés (3 et plus) n'ont pas d'effet ; leurs valeurs calculées ne sont supposées différer de 0 que par suite de la dispersion expérimentale et constituent donc des mesures d'écarts à 0 aléatoires. Ce qui permet d'estimer l'erreur type SE des effets et de tester la signification des autres effets du plan (principaux et du 2ème ordre) par une série de tests de Student.

Cette méthode n'est conseillée que pour des plans factoriels comportant au minimum 4 facteurs (il y a alors 5 effets d'ordre supérieurs) pour que l'imprécision de SE ne soit pas trop grande.

- **méthode graphique : le graphe de probabilité normale.**
  *cf exercice n°5 ( questions 1, 2, 3 et 4 ).*

Les facteurs et interactions inactifs ont des effets calculés supposés être distribués normalement autour de 0 avec l'écart type $\sigma_{E.}$. Pour les distinguer des actions ayant un effet réel (valeur différente de 0 et très supérieure à $\sigma_E$ en grandeur), on peut utiliser une méthode graphique, variante de la droite de Henry.

Celle ci a pour but de vérifier la normalité d'une distribution de données en examinant le profil des points classés par valeurs croissantes, profil qui doit être sensiblement celui d'une droite. On peut ainsi apprécier visuellement un écart à la normalité de la distribution (les points seront disposés selon une courbe) et des valeurs « aberrantes », données individuelles très éloignées des autres données qui, elles, sont sensiblement alignées et peuvent être considérées comme distribuées normalement.

Dans cette représentation, les actions et interactions sans effet devront donc être sensiblement alignées et les effets significatifs éloignés de cette droite, ce qui permet de les identifier.
Pour plus détail sur le principe de cette représentation, consulter l'annexe de ce chapitre.

# 3 - LE MODELE LINEAIRE ET LES EXPERIENCES FACTORIELLES $2^n$.

La méthode précédente de calcul des effets utilise de façon sous jacente les propriétés d'un modèle appelé modèle linéaire. Celui-ci ne permet pas seulement de quantifier des effets et de détecter ceux qui sont significatifs, il a d'autres objectifs, notamment de prédiction, que nous nous contenterons ici d'esquisser.

L'équation du modèle est par exemple indispensable à considérer quand le but de l'étude est de prédire la réponse en n'importe quel point du domaine expérimental, dans des conditions où aucune mesure n'a été effectuée, ou encore pour choisir les valeurs des niveaux des facteurs en vue d'expériences futures destinées à se diriger le plus rapidement possible vers un optimum de réponse.

## 3-1 L'équation du modèle linéaire associé aux plans factoriels $2^n$ .

Dans le modèle linéaire associé aux plans factoriels $2^n$, la réponse prédite est la somme de plusieurs termes.

Chaque terme représente un effet significatif, facteur ou interaction.

- Pour un facteur, le terme est une fonction du 1er degré par rapport au facteur centré réduit X, le coefficient de proportionnalité étant égal à l'effet principal de ce facteur (effet moyen).
- Pour une interaction, le terme est constitué par les produits des facteurs centrés réduits intervenant dans l'interaction et l'effet de cette interaction.

### ❑ Cas d'un plan $2^2$ sans interaction.

L'équation du modèle s'écrit :

$$\hat{Y} = \bar{y} + E_A X_A + E_B X_B$$

$\bar{y}$ est la réponse prédite au centre du domaine expérimental (elle correspond à $X_A$=0 et $X_B$=0) ; sa valeur est égale à la moyenne des réponses mesurées des 4 essais du plan.

$E_A$, $E_B$ désignent les effets principaux calculés qui doivent être **significatifs** pour figurer dans l'équation de prédiction de la réponse. Les variables explicatives $X_A$, $X_B$ sont les facteurs centrés réduits des facteurs A et B.

$\hat{Y}$ symbolise la valeur de la réponse prédite par le modèle pour des valeurs données de $X_A$ et $X_B$.

### ❑ Cas d'un plan $2^2$ avec interaction.

L'équation du modèle contient un terme supplémentaire représentant l'interaction entre les facteurs A et B dont l'effet est $E_{AB}$ .

Elle s'écrit :

$$\hat{Y} = \bar{y} + E_A X_A + E_B X_B + E_{AB} X_A X_B$$

Cette équation complète ne présente d'intérêt que si les 3 effets $E_A$, $E_B$ et $E_{AB}$ sont significatifs. Dans ces conditions, les réponses prédites pour les 4 points expérimentés dans le plan sont exactement égales aux réponses mesurées s'il n'y a pas de répétitions.

En revanche, lorsqu'un des effets n'est pas significatif, $E_B$ par exemple, il est inutile d'écrire son terme dans l'équation du modèle qui se simplifie et devient :

$$\hat{Y} = \bar{y} + E_A X_A + E_{AB} X_A X_B$$

Les réponses prédites pour les points expérimentés sont alors en général différentes des réponses mesurées, qu'il y ait des répétitions ou non ; les différences entre réponses mesurées et prédites sont appelées des résidus. Nous y reviendrons.

### ❑ généralisation : cas d'un plan $2^n$

*cf exercice n°5 ( question 5 ).*

Pour des plans factoriels comprenant un nombre croissant de facteurs, le nombre total d'effets à calculer [principaux (nombre n) et interactions] augmente rapidement avec n. Cependant, en pratique, il a été constaté que le nombre d'effets significatifs d'une expérimentation se limite très souvent à quelques facteurs et interactions de 2ème ordre. L'équation du modèle reste donc relativement simple puisqu'il n'y figure que des sommes de termes d'effets significatifs.

**Cette équation constitue le résumé le mieux adapté à l'interprétation lorsque les facteurs significatifs sont tous quantitatifs continus**. Les variables explicatives X ont été expérimentées pour –1 et +1 mais toute autre valeur comprise entre ces limites définit un point du domaine : l'équation numérique permet l'estimation de la valeur de la réponse en ce point. L'ensemble des réponses prédites dans tout le domaine expérimental définit un ensemble de surfaces appelées surfaces de réponses. Nous venons de donner un aperçu d'un outil statistique de prévision appartenant aux méthodes dites de **modélisation par surfaces de réponses**.

Notons que ce modèle simple, posé 'a priori', a ses limitations. Il est en défaut lors de certaines situations expérimentales. Du 1er degré par rapport à chacun des facteurs, il est par exemple incapable de simuler un maximum ou un minimum de réponse dans le domaine exploré ; il faudrait pour cela utiliser une équation contenant au moins un terme du 2ème degré en X. Et le plan expérimental devrait donc avoir été conçu pour comporter un minimum de 3 niveaux par facteur.

Remarquons aussi que lorsque **les facteurs sont qualitatifs**, **les valeurs –1 et +1 des facteurs centrés réduits ne constituent que des codes de désignation des 2 modalités expérimentées**, utiles uniquement pour le calcul des effets. **Toute autre valeur de X n'a, en revanche, pas de signification concrète**.

L'équation du modèle, écrite sous cette forme mathématique, est trompeuse puisqu'elle n'a de sens que pour deux valeurs particulières de X. Nous étudierons dans un prochain paragraphe une technique de modélisation en représentation qualitative, mieux adaptée à ce cas.

## 3-2 Utilisation de la régression linéaire multiple

*cf exercice n°4 ( question 5 ).*

La régression linéaire multiple a pour objet d'expliquer une variable (ici la réponse y) par plusieurs variables explicatives (ici les facteurs et interactions) au moyen d'une relation linéaire. C'est un outil statistique de prévision.

Cette méthode convient particulièrement bien pour la modélisation des plans factoriels $2^n$ comportant des facteurs quantitatifs : les variables explicatives sont indépendantes entre elles (par construction du plan et il est facile de le vérifier en remarquant que les corrélations entre les colonnes de la matrice des effets, prises deux à deux, sont toutes nulles). Il en résulte que les coefficients de l'équation (les effets estimés) sont aussi indépendants les uns des autres et ne dépendent pas du nombre de variables explicatives introduites dans la régression.

En pratique, dans la matrice des effets, on choisit habituellement comme variables explicatives les colonnes correspondant aux facteurs et aux interactions d'ordre 2. En procédant ainsi, on admet donc que les interactions d'ordres supérieurs, négligées dans l'explication de la réponse y, n'ont pas d'effet. Un intérêt de cette supposition est statistique : il est ainsi possible d'estimer l'écart type des effets comme nous l'avons déjà vu en (2-3) et de tester les effets principaux et les interactions d'ordre 2.

Des programmes de calcul de la régression linéaire multiple existent dans la plupart des logiciels de statistique et dans quelques tableurs (notamment Excel). L'avantage de l'utilisation d'un tel programme est évident : il donne immédiatement les estimations des effets moyens, l'erreur type de ces effets, les effets significatifs accompagnés du degré de signification. Les résultats peuvent comprendre en plus l'étude des résidus, importante pour détecter d'éventuelles mesures aberrantes, incompatibles avec le modèle choisi.

Ajoutons que la régression linéaire multiple fournit pour tous les plans factoriels à deux niveaux des résultats identiques à ceux qui ont été obtenus plus laborieusement dans les paragraphes 2 et 3.

## 3-3 Validation du modèle.

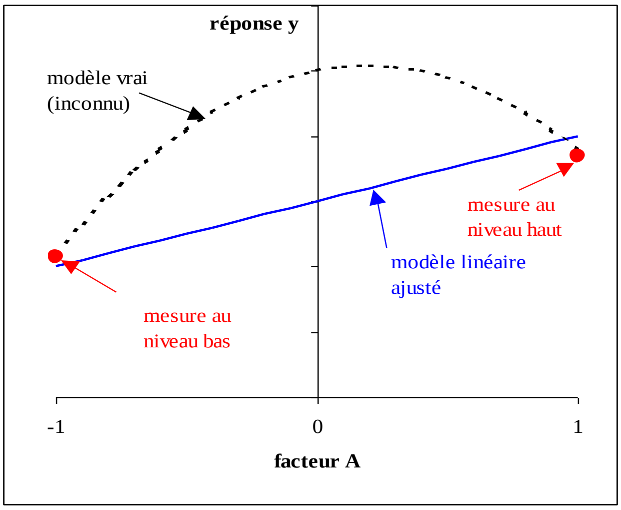
Les méthodes de calcul des effets (matrice des effets ou régression linéaire multiple) utilisent les propriétés du modèle linéaire du premier degré. Celui-ci repose sur un certain nombre de suppositions, à la fois mathématiques et statistiques, dont il est nécessaire de vérifier qu'elles sont plausibles.

- **Test du modèle.**

*cf exercice n°2 ( question 4 ).*

En choisissant les points expérimentaux aux extrémités du domaine et en ajustant l'équation numérique grâce à ces seuls points, l'utilisateur fait en sorte que les réponses prédites $\hat{Y}$ soient le plus proche possible des résultats expérimentaux y.

Le choix de ce modèle est cependant empirique : il s'agit de l'équation la plus simple (du premier degré, ne tenant compte que des effets significatifs) et il arrive parfois qu'elle constitue une approximation trop grossière pour pouvoir s'appliquer à l'ensemble des réponses y mesurées.
De plus, en pratique, rien ne prouve que le modèle est capable de s'ajuster correctement à l'intérieur du domaine expérimental comme illustré sur le schéma suivant :




Supposons que le modèle « vrai » du système expérimenté corresponde à une équation du 2ème degré. L'approximation par le modèle linéaire, matérialisé par la droite ajustée n'est convenable qu'aux 2 extrémités du domaine du facteur A ; il est visible qu'il existe des écarts importants entre les deux modèles à l'intérieur du domaine expérimental, dans lequel aucune mesure n'a été effectuée.

⇒ Le modèle linéaire n'est évidemment pas dans ce cas, un bon outil de prévision des réponses hors des points expérimentés.

Pour s'assurer de la validité de l'équation de régression (linéarité) dans l'ensemble du domaine expérimental, il est nécessaire **d'exécuter des expériences complémentaires - des répétitions –au centre du domaine expérimental**.
Il devient alors possible de comparer statistiquement la réponse moyenne des répétitions mesurées au centre du domaine, à la valeur prédite par le modèle du premier degré. Rappelons que la valeur de cette dernière est la moyenne $\bar{y}$ des essais du plan. Ces 2 moyennes doivent peu différer si le modèle est valide.

Lorsque l'expérimentateur obtient des valeurs très différentes, s'écartant de plusieurs fois l'écart type, il faut mettre en cause l'équation du modèle linéaire adopté. Un modèle empirique plus complexe, où les facteurs significatifs n'interviennent pas seulement au 1er degré mais aussi au 2ème degré, rend probablement mieux compte des résultats du plan factoriel.

Signalons enfin que la mise en œuvre de ce test de linéarité de l'équation, au centre du domaine, n'est réalisable que si tous les facteurs significatifs étudiés sont quantitatifs continus : le centre du domaine (X=0 pour tous les facteurs) doit évidemment correspondre à un traitement physiquement ou chimiquement possible. Or cela n'est pas le cas avec des facteurs qualitatifs dont les modalités varient par bond. Pour ces derniers, la prédiction des réponses à l'intérieur du domaine ne présente pas de signification concrète.

## ❑ Étude des résidus.

*cf exercice n°5 ( question 6 )*

On appelle résidu la différence entre la réponse mesurée et la réponse prédite par le modèle. Pour l'essai n° i, la valeur de son résidu $r_i$ est égale à :

$$r_i = y_i - \hat{y}_i$$

$y_i$ et $\hat{y}_i$ désignent respectivement les valeurs mesurées et prédites pour l'essai i.

La notion de résidu n'a pas de signification lorsqu'il est tenu compte dans l'équation du modèle de tous les effets calculés du plan factoriel : ils sont évidemment tous nuls pour un plan ne comportant pas de répétitions.

Lorsqu'on ne tient compte que des effets significatifs, les résidus ont des valeurs non nulles qui doivent être considérées comme des « termes d'erreur » ; ils correspondent par définition à la partie de la mesure non explicable par le modèle. Les causes de leur existence sont diverses : citons les variations des facteurs non contrôlés pendant l'expérimentation (cause souvent prépondérante), l'imprécision de la méthode de mesure et un éventuel mauvais choix du modèle.

Par construction, les résidus ont toujours comme moyenne la valeur 0 et comme écart type Sy, l'écart type des réponses individuelles (le paramètre à la base des tests des effets).

Les suppositions statistiques du modèle nécessitent que **les résidus soient distribués de façon aléatoire, selon une loi normale.**

Diverses méthodes simples, le plus souvent graphiques, sont utilisées pour essayer de vérifier la normalité et le caractère aléatoire ; Citons :

- Le graphe de probabilité normale, qui permet de détecter éventuellement un écart à la normalité ou un « résidu aberrant » (la valeur de la réponse mesurée est suspecte, est-ce une faute de manipulation ? …).
- L'étude des résidus réduits. Un résidu réduit $rr_i$ est défini comme le rapport $r_i$ / Sy . Les résidus réduits doivent en effet être compris entre –2 et +2 (il n'y a qu'une probabilité maximale de 0,05 pour qu'un résidu réduit dépasse par hasard la valeur 1,96, selon la loi normale).
- La représentation graphique des résidus en fonction de l'ordre chronologique des essais (rappelons que cet ordre doit être obtenu par randomisation) ne doit pas faire apparaître de tendance systématique à la baisse ou à la hausse des résidus. Sinon cela signifie qu'un facteur non contrôlé dans le plan a produit une dérive des réponses mesurées qui biaise les effets calculés.
- On peut aussi représenter ces mêmes résidus en fonction des réponses prédites par le modèle ou en fonction des niveaux des facteurs avec le même objectif que précédemment : dépister une structure non aléatoire des résidus.

L'étude des résidus aux points expérimentés s'effectue avec n'importe quelle catégorie de facteurs, qualitatifs et / ou quantitatifs. Il n'est pas inutile d'insister sur la nécessité de pratiquer ces études de validation 'a posteriori' : elles donnent de la crédibilité aux résultats obtenus.

## 3-4 Modélisation en représentation qualitative.

*cf exercice n°5 ( question 7 a) ).*

Pour les facteurs qualitatifs, on peut utiliser avec profit une présentation différente du modèle linéaire, mieux adaptée au caractère discontinu des modalités de cette catégorie de facteurs.

### ❑ Écriture du modèle.

- Les facteurs.

Jusqu'à présent, nous avons caractérisé l'action d'un facteur A (niveaux $A_1$ et $A_2$) par son effet moyen $E_A$. Rappelons que cette grandeur traduit la différence entre la réponse au niveau HAUT ($A_2$) et la moyenne générale des réponses ($\overline{y}$) des essais du plan. Pour un facteur qualitatif, cette définition est artificielle. Seule la variation de réponse résultant du passage de $A_1$ à $A_2$ présente de l'intérêt ; elle vaut $2E_A$ et s'appelle l'effet global.

Une approche intéressante consiste à définir **un effet moyen pour chaque niveau de chaque facteur** : il suffit de considérer, en plus, l'effet du niveau BAS ($A_1$) défini comme la différence entre la réponse au niveau BAS et la moyenne générale $\overline{y}$ ; sa valeur est donc égale à $-E_A$.

Le terme de ce facteur dans l'équation de modélisation s'écrit symboliquement : $[\,-E_A \;\; {+E_A}\,][A]$, le symbole [A] étant présent seulement pour indiquer que les valeurs entre crochets se réfèrent à ce facteur. A gauche est indiqué l'effet du niveau $A_1$ et à droite celui du niveau $A_2$. De telle sorte que lors du passage de $A_1$ à $A_2$, l'effet global se calcule immédiatement par l'opération : $+E_A - (-E_A) = 2E_A$.

Dans un modèle comprenant par exemple 2 facteurs A et B significatifs sans interaction, on écrira :

$$\hat{Y} = \overline{y} + [\,-E_A \;\; {+E_A}\,][A] + [\,-E_B \;\; {+E_B}\,][B]$$

Ce système de modélisation présente les avantages suivants : tous les effets des facteurs significatifs à chaque niveau sont représentés numériquement en clair et il est ainsi très facile de calculer la réponse prédite pour une configuration donnée de l'état des facteurs.

- Les interactions.

Nous limiterons ici l'étude du modèle aux interactions de $2^{\text{ème}}$ ordre. Les 2 facteurs intervenant dans l'interaction sont désignés par A et B. Selon les niveaux 1 ou 2 de ces facteurs, il y a 4 états possibles pour l'interaction ($A_1B_1$, $A_2B_1$, $A_1B_2$ et $A_2B_2$).

L'effet $E_{AB}$, défini et calculé précédemment, correspond aux configurations $A_2B_2$ et $A_1B_1$ ; on peut montrer que les deux autres configurations sont caractérisées par un effet égal à $-E_{AB}$ de telle sorte que le terme des effets d'interaction s'écrit de la façon matricielle suivante :

$$\begin{bmatrix} +E_{AB} & -E_{AB} \\ -E_{AB} & +E_{AB} \end{bmatrix} AB$$

La première ligne de la matrice correspond au niveau $A_1$ du facteur A et la deuxième au niveau $A_2$. La colonne de gauche correspond au niveau $B_1$ du facteur B et celle de droite au niveau $B_2$.

Lorsqu'on considère le modèle complet d'un plan $2^2$ avec interaction, on écrit :

$$\hat{Y} = \overline{y} + [\,-E_A \;\; {+E_A}\,][A] + [\,-E_B \;\; {+E_B}\,][B] + \begin{bmatrix} +E_{AB} & -E_{AB} \\ -E_{AB} & +E_{AB} \end{bmatrix} AB$$

Exemple d'utilisation.

Considérons l'équation numérique de prévision suivante :

$$\hat{Y} = 58{,}9 + [\ +25{,}0 \quad \mathbf{-25{,}0}][A] + [\ \mathbf{-13{,}6} \quad +13{,}6\ ][B] + \begin{bmatrix} +8.2 & -8.2 \\ -8.2 & +8.2 \end{bmatrix} AB$$

Quand A est au niveau $A_2$ et B au niveau $B_1$, la valeur de la réponse prédite est égale à :

$$\hat{Y} = 58{,}9 - 25{,}0 - 13.6 - 8{,}2 = \mathbf{12{,}1}.$$

Ce type matriciel d'équation de modélisation est généralisable à des plans factoriels contenant plus de 2 facteurs qualitatifs. On n'y fait généralement figurer que les effets principaux et interactions d'ordre 2 significatifs.
Notons qu'il peut aussi s'appliquer aux facteurs quantitatifs quand l'utilisateur n'est intéressé que par les réponses prédites aux points expérimentés.

- **Application : recherche des modalités des facteurs conduisant à une réponse optimale.**
  *cf exercice n°5 ( question 7b ).*

Il peut arriver d'avoir à choisir, parmi les traitements d'un plan $2^n$, celui qui conduit à une réponse optimale, un minimum ou un maximum selon le cas. Ce problème se produit fréquemment avec des facteurs qualitatifs : quelle est la meilleure machine ?   quelle est la meilleure méthode ? … etc.

Une façon très pragmatique consiste à examiner les réponses mesurées, ce qui permet de repérer immédiatement l'essai ayant donné « la meilleure » réponse. Cette manière de procéder pour trouver la configuration optimale des facteurs n'est que très approximative par suite de l'imprécision entachant les réponses individuelles mesurées.

Une façon de procéder plus rationnelle consiste à déterminer les effets significatifs du plan ; à établir une modélisation en représentation qualitative et à rechercher grâce à celle-ci les modalités des facteurs qui rendent la réponse prédite optimale.

# *Conclusion sur les plans $2^n$.*

Les plans factoriels complets $2^n$ présentent de nombreux avantages : ils sont efficaces, leur analyse et interprétation sont simples et sûres. Ces plans ont, de plus, des vertus pédagogiques dans la mesure où ils permettent de considérer une autre stratégie expérimentale que « un seul facteur variable étudié à la fois ».

Ils ne constituent pourtant pas la panacée des plans d'expériences :

- Ils conduisent vite à des expérimentations onéreuses et très consommatrices de temps quand le nombre de facteurs augmente.
- Ils sont en défaut lorsque certains traitements du plan ne sont pas physiquement (ou chimiquement) réalisables : des incompatibilités peuvent se manifester entre niveaux de facteurs. Un exemple : lors de l'étude d'une propriété d'un composé en solution, il est clair que si tel niveau de pH combiné avec telle température provoque la précipitation du composé, le traitement ne pourra pas être expérimenté et le plan factoriel ne pourra pas être complet.
- Deux niveaux par facteur (une mesure d'économie) peuvent se révéler insuffisants dans certaines études : on peut avoir à comparer 3 méthodes d'analyses, il faut plus de 2 concentrations pour bien estimer une courbe d'étalonnage …etc.

# ANNEXE : CRITERE GRAPHIQUE DE NORMALITE D'UNE DISTRIBUTION.APPLICATION A LA DETECTION DES EFFETS SIGNIFICATIFS D'UN PLAN FACTORIEL.

## Introduction

Les tests statistiques paramétriques supposent la normalité des distributions de données à comparer.

Une étude préalable de cette normalité, utile pour valider les conclusions des tests n'est pourtant pas systématiquement effectuée, d'une part parce que la plupart des tests statistiques sont « robustes » vis à vis des écarts à la normalité et, d'autre part parce que les méthodes classiques d'étude de la normalité nécessitent de nombreuses données, au moins 50 valeurs.

Pour des effectifs N supérieurs à 50, rappelons qu'il faut d'abord regrouper les données individuelles en classes de valeurs, construire ensuite l'histogramme des fréquences relatives qu'il est intéressant de superposer avec la loi de probabilité normale de même moyenne et écart type que ceux de la série de valeurs pour visualiser si l'ajustement est convenable ou pas. On peut aussi, sur ces données groupées en classes, effectuer des tests statistiques d'ajustement à la loi normale, le plus connu étant le test du $\chi^2$ (il en existe d'autres, par exemple celui de Kolmogorov).

Ces méthodes, bien que classiques, ne sont pas totalement satisfaisantes. L'histogramme et les tests de normalité sont en effet calculés à partir des effectifs des classes et ainsi dépendent du découpage de l'échelle de la variable en classes, défini empiriquement par l'utilisateur. Ce choix arbitraire a peu d'importance pour de grandes séries (effectif supérieur à 100) mais il joue un rôle d'autant plus grand que l'effectif total de la série est plus petit et il est d'ailleurs fortement à conseiller de ne pas faire de groupements en classes pour des séries de moins de 50 valeurs (la forme de l'histogramme peut varier notablement selon le découpage et les tests ne sont plus très fiables).

Il a été développé des méthodes d'appréciation de la normalité d'une distribution, fondées sur les valeurs individuelles, ne nécessitant pas de regroupements.

Parmi ces méthodes, il en est une, graphique, portant le nom de **graphe de probabilité normale** qui apparaît intéressante lors de l'analyse exploratoire des données pour détecter dans celles ci des écarts importants à la normalité ou une éventuelle présence de valeurs aberrantes.

Elle peut être utilisée pour déceler les effets significatifs d'un plan factoriel.

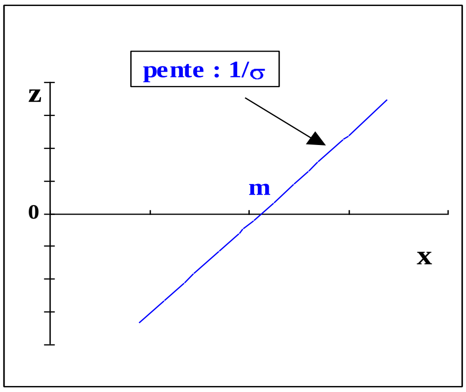
# 1- PRINCIPE DE LA REPRESENTATION.

Soit une série comportant N données numériques $x_1$, $x_2$, ...., $x_i$, ......, $x_N$.
Désignons par m et $\sigma$ la moyenne et l'écart type de ces N données.

A chaque donnée $x_i$, on peut faire correspondre sa valeur centrée réduite $z_i$ par :

$$z_i = \frac{x_i - m}{\sigma}$$

Z est une fonction linéaire de X, comme le montre le schéma de droite.
(ordonnée à l'origine = m et pente = $1/\sigma$).




Cette relation, vérifiée quelle que soit la distribution de X, ne présente d'intérêt que s'il est possible d'évaluer Z autrement qu'à partir de la moyenne m et de l'écart type $\sigma$ de la série de données.

Remarquons alors que si les valeurs de X sont distribuées normalement, Z suit la loi normale centrée réduite (de moyenne 0 et d'écart type 1). La table de la fonction de répartition de cette loi établit la correspondance numérique entre une donnée connue $x_i$ et la probabilité pour qu'une valeur X (ou Z) soit inférieure à $x_i$ (ou $z_i$), puisqu'une propriété des lois normales s'écrit :

$$\text{Prob}\,(Z < z_i) = \text{Prob}\,(X < x_i)$$

L'estimation de la probabilité de la loi normale par son équivalent expérimental la fréquence relative cumulée jusqu'à la valeur $x_i$, permet ainsi d'obtenir une valeur $z_i$ , indépendante de la définition précédente.

Le report des valeurs de $z_i$ estimées pour les valeurs mesurées $x_i$ sur un graphe avec des échelles arithmétiques, ne doit conduire à des points alignés que si les valeurs de $x_i$ sont distribuées normalement.

C'est donc l'appréciation de l'alignement des points sur cette représentation qui permet de juger de l'écart à la normalité de la distribution observée des valeurs de X.

# 2 - CALCULS A EFFECTUER : FREQUENCES RELATIVES CUMULEES PUIS Z.

La fréquence relative cumulée correspondant à chaque valeur $x_i$ de la série se calcule de la façon suivante :

- Ranger les données $x_i$ par valeurs croissantes ; noter 1 pour la plus petite, 2 pour la deuxième, ...., et N pour la plus grande. Ces nombres sont appelés les rangs (1, 2, ..., $r_i$, ....N).
  Quand 2 données sont égales, ces 2 valeurs doivent avoir le même rang et il faut donc prendre la moyenne arithmétique des 2 rangs. Un rang commun est calculé de façon similaire (moyenne arithmétique des rangs) quand plus de 2 données sont égales.

- La fréquence relative cumulée jusqu'à la valeur $x_i$, à laquelle correspond le rang $r_i$, notée $FRC_i$, s'obtient par une formule que nous admettrons :

$$FRC_i = \frac{r_i - 3/8}{N + 1/4}$$

- La table de la loi normale permet enfin de déterminer la valeur $z_i$ correspondant à $FRC_i$..

Au total, les N valeurs x de la série conduisent à N couples de points ($x_i$ , $z_i$) qui devront être sensiblement alignés (se répartir de façon aléatoire autour d'une droite) si la distribution est normale.

En pratique, à la condition que l'étude porte **sur plus d'une dizaine de données**, il est possible d'avoir une assez bonne idée d'écarts importants à la normalité de la distribution et espérer détecter :

- une distribution très différente de la loi normale (courbure manifeste des points).
- des valeurs isolées appelées « aberrantes » dans le vocabulaire statistique ; elles correspondent à quelques données individuelles suffisamment éloignées des autres pour être incompatibles avec ces dernières. Elles résultent souvent d'une faute (manipulation ?, transcription ? ...) et doivent déclencher une enquête sur les conditions d'obtention de ces résultats ; elles ne doivent être écartées de l'étude que s'il est prouvé qu'il s'agit bien d'une erreur.

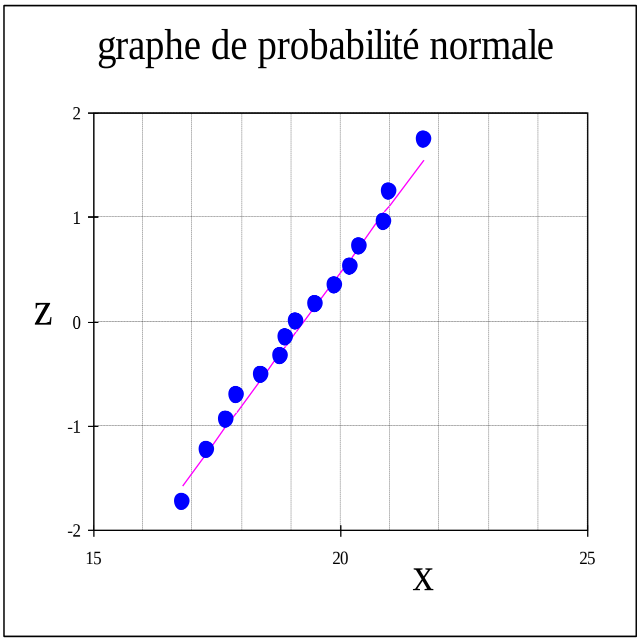
Remarquons enfin qu'une variante de cette méthode graphique d'étude de la normalité est fréquemment utilisée dans le cas de données nombreuses (effectif supérieur à 50, de préférence 100), groupées en classes. Elle porte le nom de **droite de Henry.** Cette représentation, très employée en contrôle de qualité, substitue à chaque valeur individuelle son centre de classe ; les effectifs des classes permettent d'obtenir les estimations des fréquences relatives cumulées, ce qui différencie les 2 méthodes. Pour le reste, les représentations sont analogues.

# 3- DES EXEMPLES DE GRAPHES DE PROBABILITE NORMALE.

## Exemple 1.

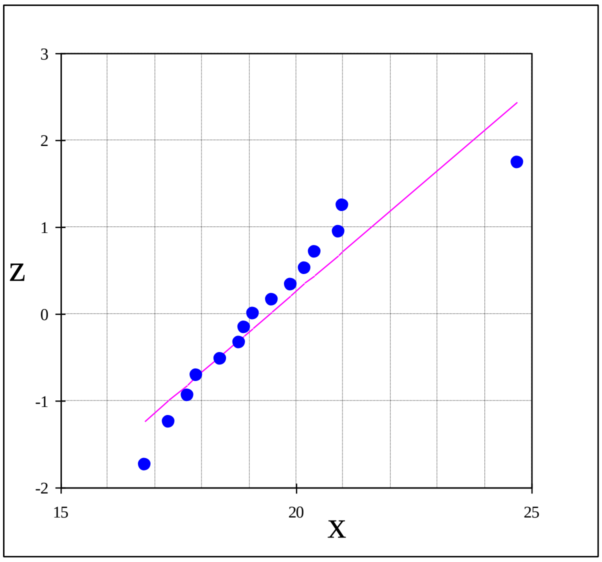
Soit la série des 15 valeurs de x suivantes déjà rangées en ordre croissant ; on souhaite savoir si l'écart à la normalité est important ou pas ....
Le tableau des calculs à effectuer est figuré ci dessous :

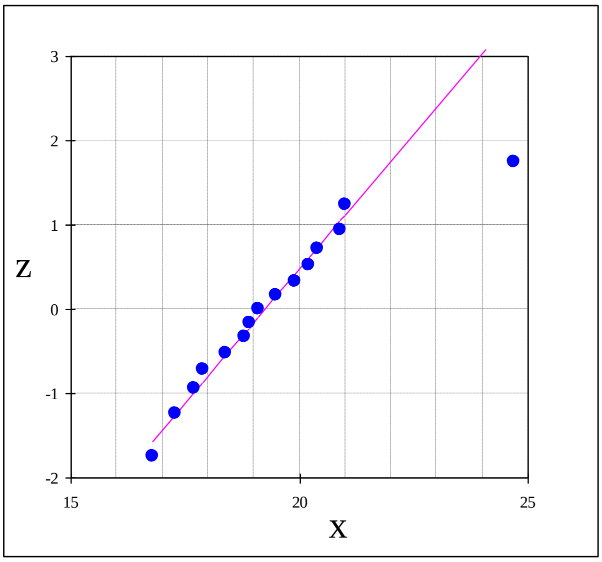
| x | rang r | FRC | z normale |
|---|---|---|---|
| **16.8** | 1 | 0.041 | -1.739 |
| **17.3** | 2 | 0.107 | -1.245 |
| **17.7** | 3 | 0.172 | -0.946 |
| **17.9** | 4 | 0.238 | -0.714 |
| **18.4** | 5 | 0.303 | -0.515 |
| **18.8** | 6 | 0.369 | -0.335 |
| **18.9** | 7 | 0.434 | -0.165 |
| **19.1** | 8 | 0.500 | 0.000 |
| **19.5** | 9 | 0.566 | 0.165 |
| **19.9** | 10 | 0.631 | 0.335 |
| **20.2** | 11 | 0.697 | 0.515 |
| **20.4** | 12 | 0.762 | 0.714 |
| **20.9** | 13 | 0.828 | 0.946 |
| **21.0** | 14 | 0.893 | 1.245 |
| **21.7** | 15 | 0.959 | 1.739 |



L'examen visuel du graphe montre que les 15 points ($z_i$ , $x_i$) sont répartis aléatoirement autour de la droite de régression sans s'en écarter notablement ; il est possible de considérer que ces 15 valeurs constituent un échantillon d'une population distribuée normalement.

Il n'en serait pas de même si la dernière donnée avait comme valeur 24,7 au lieu de 21,7 , les 14 autres données restant inchangées. Le graphe de gauche montre la discordance entre les 15 points et sa droite de régression calculée avec les 15 valeurs, alors qu'à droite le calcul de la régression n'a été effectué qu'avec les 14 premières valeurs.



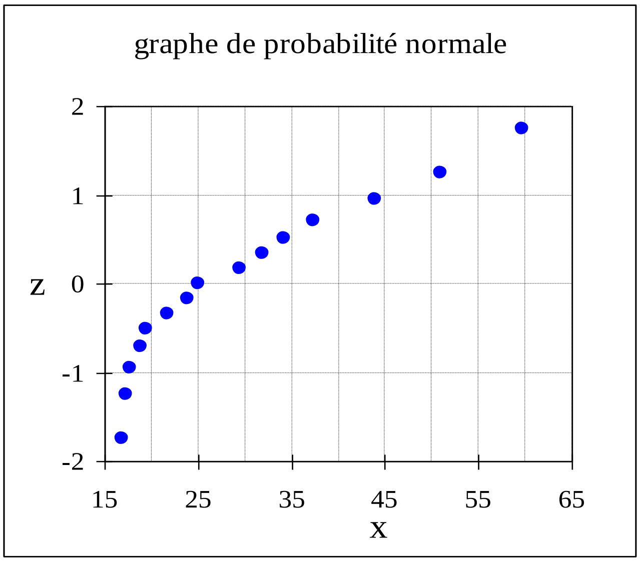

Ces 2 graphes ont pour but de montrer qu'un point « aberrant » est capable de modifier très nettement la représentation : à gauche, on voit qu'il « pèse » suffisamment dans l'équation de la droite de régression pour que les 14 points inchangés ne se répartissent plus de façon aléatoire autour de cette droite.

On voit mieux la singularité du 15ème point sur la représentation de droite, dont les 14 premiers points se répartissent de façon aléatoire autour de la droite de régression, calculée en ne tenant compte que d'eux seuls : le point aberrant s'éloigne de la droite, ce qui facilite sa mise en évidence.

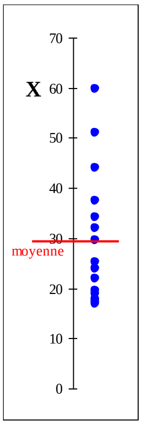
## <u>Exemple 2</u>.

Considérons maintenant la série des 15 données suivantes déjà ordonnées en ordre croissant. Les mêmes calculs et représentation que précédemment sont figurés ci dessous :

| x | rang r | FRC | z normale |
|---|---|---|---|
| **16.8** | 1 | 0.041 | -1.739 |
| **17.3** | 2 | 0.107 | -1.245 |
| **17.7** | 3 | 0.172 | -0.946 |
| **18.9** | 4 | 0.238 | -0.714 |
| **19.4** | 5 | 0.303 | -0.515 |
| **21.8** | 6 | 0.369 | -0.335 |
| **23.9** | 7 | 0.434 | -0.165 |
| **25.1** | 8 | 0.500 | 0.000 |
| **29.5** | 9 | 0.566 | 0.165 |
| **31.9** | 10 | 0.631 | 0.335 |
| **34.2** | 11 | 0.697 | 0.515 |
| **37.4** | 12 | 0.762 | 0.714 |
| **43.9** | 13 | 0.828 | 0.946 |
| **51.0** | 14 | 0.893 | 1.245 |
| **59.7** | 15 | 0.959 | 1.739 |



Un simple coup d'œil sur la disposition des points suffit pour constater qu'il n'y a pas linéarité : il est inutile d'essayer d'ajuster une droite des moindres carrés. La distribution n'est manifestement pas normale.




A gauche nous avons reporté sur un axe gradué les 15 valeurs individuelles, montrant que la distribution est très nettement dissymétrique par rapport à la moyenne des 15 valeurs. Comme une des propriétés essentielles d'une loi normale est la symétrie autour de la moyenne, les données ne peuvent évidemment pas être distribuées normalement, ce que révèle aussi le graphe de probabilité normale.

# 4 - APPLICATION A LA DETECTION DES EFFETS SIGNIFICATIFS D'UN PLAN FACTORIEL.

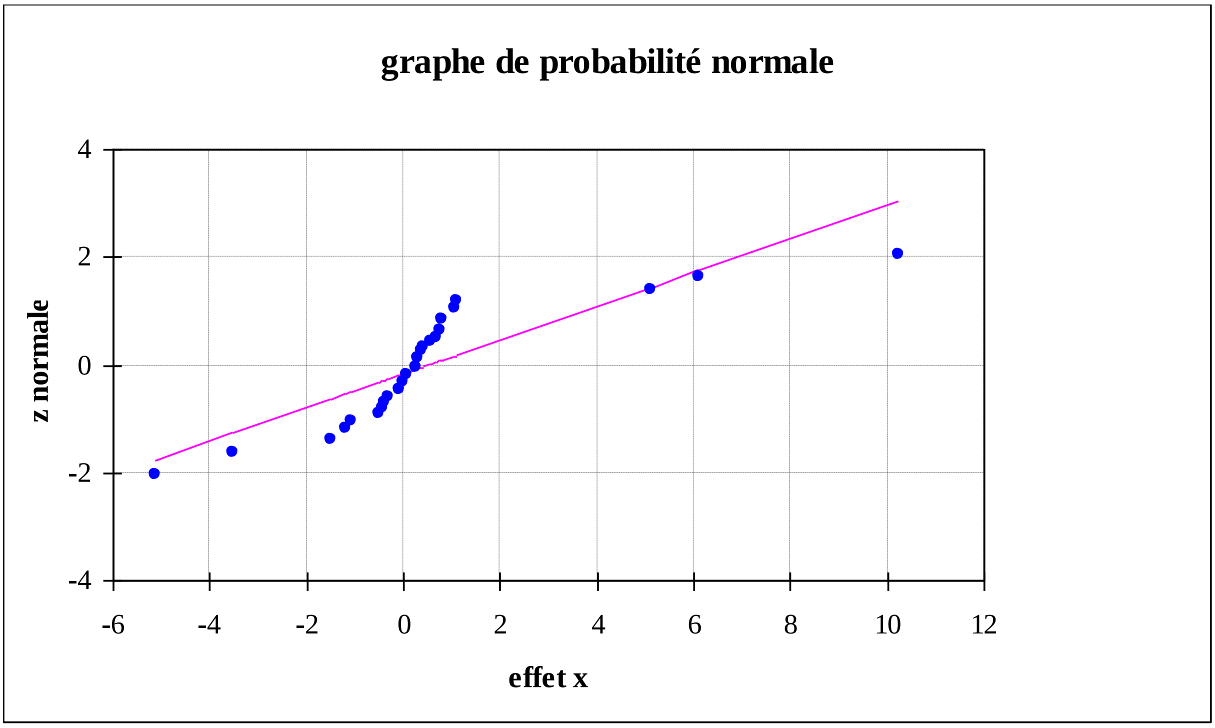
Les effets calculés à partir d'un plan factoriel constituent une série de données ; nous désirons, au moyen du graphe de probabilité normale, différencier ceux qui sont significatifs de ceux qui ne le sont pas.

- les effets inactifs dans le plan factoriel sont supposés être distribués normalement avec comme moyenne 0 et comme écart type $\sigma_E$ . Ils doivent donc conduire à des points sensiblement alignés dans le graphe de probabilité normale ; on pourra grâce à eux tracer une droite de régression.
- les effets significatifs ne seront pas situés sur cette droite ou même aux alentours de cette droite puisqu'ils ont un effet non nul ; ils se comporteront comme des valeurs aberrantes dans la série des effets, ce qui doit permettre de les identifier.

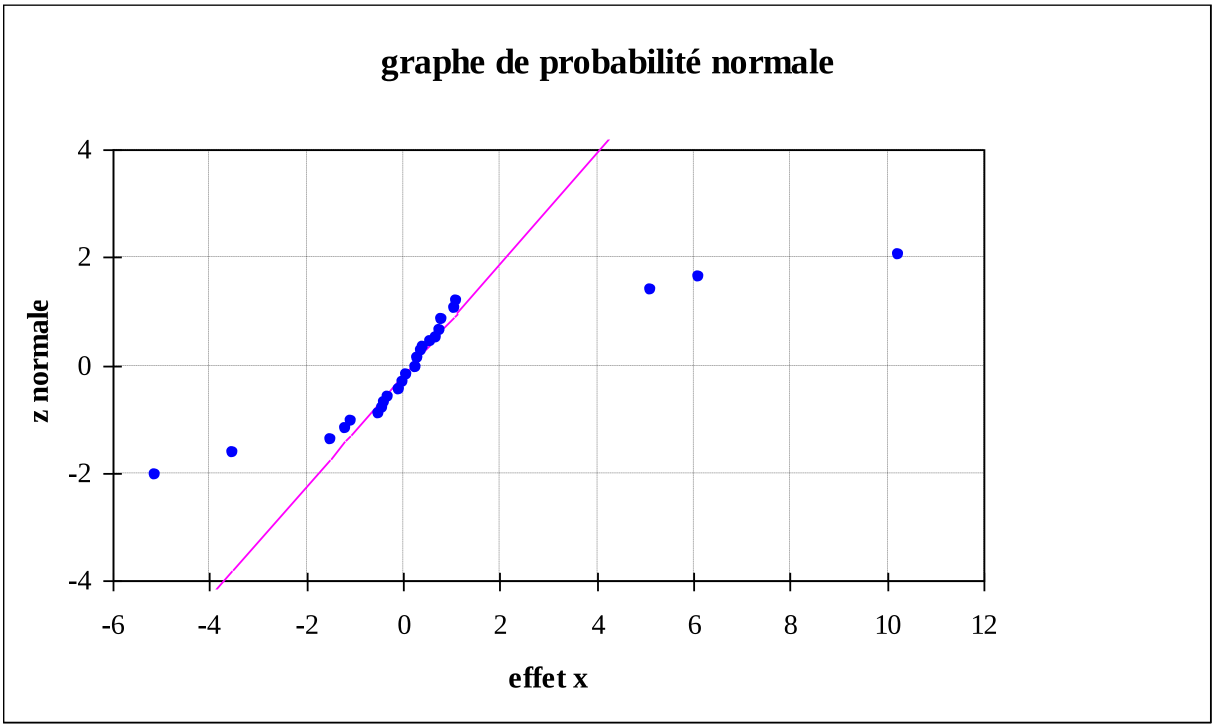
A titre d'exemple, considérons les 31 effets du plan factoriel $2^5$ de l'exercice 5 dont les valeurs triées par ordre croissant sont données ci dessous avec les calculs nécessaires pour effectuer le graphe de probabilité normale.

| désignation | effet | rang | FRC | Z normale | droite |
|---|---|---|---|---|---|
| DF | -5.125 | 1 | 0.020 | -2.054 | -1.762 |
| F | -3.500 | 2 | 0.052 | -1.626 | -1.254 |
| BCF | -1.500 | 3 | 0.084 | -1.379 | -0.630 |
| ABF | -1.188 | 4 | 0.116 | -1.195 | -0.532 |
| BD | -1.063 | 5 | 0.148 | -1.045 | -0.493 |
| BCD | -0.500 | 6 | 0.180 | -0.915 | -0.317 |
| BF | -0.438 | 7 | 0.212 | -0.800 | -0.298 |
| ABCD | -0.375 | 8 | 0.244 | -0.693 | -0.278 |
| B | -0.313 | 9 | 0.276 | -0.595 | -0.259 |
| C | -0.063 | 10.5 | 0.324 | -0.457 | -0.181 |
| ACF | -0.063 | 10.5 | 0.324 | -0.457 | -0.181 |
| ABCDF | 0.000 | 12 | 0.372 | -0.327 | -0.161 |
| CF | 0.063 | 13.5 | 0.420 | -0.202 | -0.142 |
| ACDF | 0.063 | 13.5 | 0.420 | -0.202 | -0.142 |
| ADF | 0.250 | 15.5 | 0.484 | -0.040 | -0.083 |
| ABCF | 0.250 | 15.5 | 0.484 | -0.040 | -0.083 |
| ABD | 0.313 | 17.5 | 0.548 | 0.121 | -0.064 |
| ACD | 0.313 | 17.5 | 0.548 | 0.121 | -0.064 |
| AF | 0.375 | 19 | 0.596 | 0.243 | -0.044 |
| CDF | 0.438 | 20 | 0.628 | 0.327 | -0.025 |
| CD | 0.563 | 21 | 0.660 | 0.412 | 0.014 |
| AC | 0.688 | 22 | 0.692 | 0.502 | 0.053 |
| BC | 0.750 | 23.5 | 0.740 | 0.643 | 0.073 |
| BCDF | 0.750 | 23.5 | 0.740 | 0.643 | 0.073 |
| BDF | 0.813 | 25.5 | 0.804 | 0.856 | 0.092 |
| ABDF | 0.813 | 25.5 | 0.804 | 0.856 | 0.092 |
| AB | 1.063 | 27 | 0.852 | 1.045 | 0.171 |
| ABC | 1.125 | 28 | 0.884 | 1.195 | 0.190 |
| D | 5.125 | 29 | 0.916 | 1.379 | 1.439 |
| AD | 6.125 | 30 | 0.948 | 1.626 | 1.751 |
| A | 10.250 | 31 | 0.980 | 2.054 | 3.040 |

Le graphe de probabilité normale avec les 31 points et la droite de régression calculée à partir de tous les effets (ordonnée à l'origine = -0,161 et pente =  0,312) présente l'allure suivante :




Il apparaît nettement que 5 points, 3 à droite  et 2 à gauche, influencent fortement la position de la droite de régression. En recalculant celle ci avec les 26 points restants (ordonnée à l'origine égale à -0,177 et pente égale à 1,030), on obtient alors le graphe suivant :




L'ensemble des 26 points se répartit convenablement autour de la droite et on peut conclure que 5 effets sont significatifs : les 3 facteurs A, D et F et les 2 interactions de 2$^{ème}$ ordre AD et DF.

# EXERCICES SUR LES PLANS FACTORIELS COMPLETS

## Exercice 1

On étudie l'influence de la température θ et de la concentration C d'un réactif sur le rendement y (en %) d'une réaction chimique.
Il a été décidé d'expérimenter la température entre 60°C et 80°C et la concentration entre $10 g.L^{-1}$ et $15 g.L^{-1}$ en se limitant à 2 niveaux par facteur.

1ère question
Dans une expérience factorielle $2^2$, quelles sont les conditions expérimentales à réaliser ?
Quel nombre d'essais faut-il effectuer sachant qu'on n'a pas prévu de répétitions ?

2ème question
La température est appelée facteur A et la concentration facteur B.
En utilisant le concept de facteurs centrés réduits
* où se situe le centre du domaine expérimental ?
* où se situe le point ($x_A = + 0,5$ ; $x_B = - 0,6$)

3ème question
Construire (avec EXCEL) la matrice d'expériences. Tableau de résultats :

| essai n° | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Température (°C) | 60 | 80 | 60 | 80 |
| Concentration ($g.L^{-1}$) | 10 | 10 | 15 | 15 |
| rendement (%) | 60 | 70 | 80 | 90 |

4ème question : matrice des effets.
Construire (avec EXCEL) la matrice des effets et calculer tous les effets d'un tel plan ( $\bar{y}_C$ , $E_\theta$, $E_{conc}$, $E_{\theta\,.\,conc}$).

## Exercice 2

On effectue la même étude que celle de l'exercice 1 réalisée cette fois en présence d'un catalyseur ; les autres conditions expérimentales sont inchangées.
Tableau de résultats :

| essai n° | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| rendement (%) | 60 | 70 | 80 | 95 |

1ère question : calcul des effets.
Construire (avec EXCEL) la matrice des effets et calculer tous les effets.

2$^{ème}$ question : signification des effets.

A une température de 70°C et une concentration de 12,5 g.L$^{-1}$ (centre du domaine), il a été décidé d'effectuer 6 essais complémentaires. Les 6 rendements obtenus sont les suivants (en %) :

77,3 - 79,1 - 77,8 - 77,0 - 77,7 - 79,1

* - quelle estimation faites-vous de l'écart type des réponses individuelles ?
* - quelle est l'estimation de l'erreur type d'un effet ?
* - quels effets sont significatifs au risque de 5 % (utiliser un test t).

3$^{ème}$ question

Un autre expérimentateur, à qui on avait demandé la même étude, décide d'effectuer les 4 essais suivants

| essai n° | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Température (°C) | 60 | 80 | 70 | 70 |
| Concentration (g.L$^{-1}$) | 12,5 | 12,5 | 10 | 15 |
| rendement mesuré (%) | 71 | 83,5 | 66 | 88,5 |

* - s'agit-il des conditions d'un plan factoriel ?
* - calculer les effets principaux des facteurs et comparer avec les résultats de la 1$^{ère}$ question
* - que peut-on dire concernant l'interaction ?

4$^{ème}$ question : test du modèle.

Compte tenu des résultats de la 2$^{ème}$ question, écrire l'équation numérique du modèle linéaire associé au plan factoriel.

La moyenne des réponses mesurées des 6 répétitions diffère t'elle significativement (risque 0,05) de la valeur prédite au centre du domaine, égale à la moyenne des réponses des 4 essais du plan.

Que peut-on en conclure ?

*NB : utiliser le test de Student de comparaison des moyennes de 2 séries indépendantes, d'effectifs 6 et 4.*

## Exercice 3

Dans une solution habituellement fabriquée à 30°C sous agitation (200 tours/min) un léger trouble apparaît.

L'expérimentateur désire en connaître la (les) cause(s) et pense que 3 facteurs peuvent jouer :

- la température
- la vitesse d'agitation
- la concentration d'un additif habituellement présent à 0,30 % (p/v)

Le trouble se mesure par un indice d'opacité traduisant l'impression visuelle que donne l'intensité du « louche », indice d'autant plus grand que la solution est trouble.

Il est décidé d'organiser une expérience factorielle $2^3$ :

| facteurs | niveaux | |
|---|---|---|
| A = température | 20°C | 40°C |
| B = vitesse agitation | 100 t/min | 300 t/min |
| C = concentration additif | 0,1 % | 0,5 % |

1$^{ère}$ question

Combien y-a-t-il de conditions expérimentales et d'essais à effectuer sachant qu'on ne fera pas de répétitions ?

2$^{ème}$ question : calcul des effets.

Construire (avec EXCEL) la matrice des effets et calculer tous les effets d'un tel plan sachant que les mesures d'opacité ont donné les résultats suivants :

| essai n° | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| indice d'opacité | 0 | 4,7 | 0 | 11,5 | 9 | 14,5 | 5,1 | 18,7 |

<u>3$^{ème}$ question</u> : signification des effets.

L'expérimentateur avait constaté depuis longtemps la présence de ce trouble lors des préparations successives dans les conditions habituelles (30°C, agitation 200 t/min., additif 0,3 %).

A chaque préparation, il mesurait l'opacité pour s'assurer qu'elle ne dépassait pas une limite fixée. En étudiant la distribution de 50 mesures d'opacité dans ces conditions, il trouve que les valeurs individuelles semblent se répartir suivant une loi normale

- de moyenne 7,85
- d'écart-type 2,45

* - calculer l'erreur type $\sigma_E$ sur un effet de l'expérience factorielle.

* - quels effets peuvent être considérés comme significatifs au risque de 5 % (utiliser la loi normale) ?

# Exercice 4

Lors de l'étude de l'optimisation des conditions expérimentales d'une réaction de précipitation, on a retenu comme facteurs à étudier susceptibles d'agir sur la masse de précipité pesée dans des conditions identiques :

| facteurs | niveaux | |
|---|---|---|
| | **0** | **1** |
| A = température de la réaction | 60°C | 70°C |
| B = concentration d'un réactif | 1 g.L$^{-1}$ | 2 g.L$^{-1}$ |
| C = temps de contact | 30 min. | 45 min. |
| D = débit de lavage du précipité | 1 L.min$^{-1}$ | 0,5 L.min$^{-1}$ |

Il est décidé d'effectuer une expérience factorielle complète sans répétition. Les résultats sont les suivants :

| conditions expérimentales | Masse précipité | conditions expérimentales | Masse précipité |
|---|---|---|---|
| $A_0B_0C_0D_0$ | 60,6 | $A_0B_0C_0D_1$ | 59,6 |
| $A_1B_0C_0D_0$ | 61,0 | $A_1B_0C_0D_1$ | 61,1 |
| $A_0B_1C_0D_0$ | 60,3 | $A_0B_1C_0D_1$ | 60,7 |
| $A_1B_1C_0D_0$ | 61,7 | $A_1B_1C_0D_1$ | 61,3 |
| $A_0B_0C_1D_0$ | 62,0 | $A_0B_0C_1D_1$ | 61,6 |
| $A_1B_0C_1D_0$ | 61,5 | $A_1B_0C_1D_1$ | 61,9 |
| $A_0B_1C_1D_0$ | 61,7 | $A_0B_1C_1D_1$ | 62,3 |
| $A_1B_1C_1D_0$ | 62,4 | $A_1B_1C_1D_1$ | 62,8 |

<u>1$^{ère}$ question</u> : calcul des effets.

Construire (avec EXCEL) la matrice des effets et calculer la valeur de ces effets. Faire un graphique (histogramme) représentant la grandeur des effets pour les diverses conditions.

<u>2$^{ème}$ question</u> : plan comportant des répétitions.

Ayant constaté que le débit de lavage du précipité n'a visiblement pas d'effet sur le poids de précipité, on est ainsi ramené à un plan $2^3$ (A, B, C) comportant 2 répétitions.

* - calculer la moyenne de réponse des "duplicates" pour chaque condition expérimentale.

* - calculer les effets de ce plan $2^3$ dupliqué (vérifier que les valeurs obtenues sont identiques à celles de la 1$^{ère}$ question).

3ème question : signification des effets.

* - en se servant du tableau des réponses individuelles de la 2ème question, calculer les sommes des carrés des écarts à la moyenne (SCE) pour chacun des 8 groupes puis la variance $s^2_y$ d'une réponse du plan factoriel et enfin l'écart type $s_y$.

* - en déduire l'erreur type $s_E$ d'un effet du plan factoriel.

* - quels sont les effets significatifs au risque de 5 % (on utilisera la loi de Student à 8 d.d.l.).

4ème question : signification des effets, autre méthode.

Pour estimer l'erreur type $s_E$ des effets, il a été décidé d'utiliser les 5 interactions d'ordre 3 et 4 de l'expérience factorielle $2^4$ (donc sans utiliser l'information que l'effet de D est nul).

* - calculer (avec EXCEL) l'estimation $s_E$ à partir de ces 5 effets.

* - en déduire les effets significatifs (on utilisera la loi de Student à 5 d.d.l.).

5ème question :

Excel contient un programme de régression linéaire dans le menu **outils / utilitaire d'analyse**.

En utilisant les colonnes de la matrice des effets, effectuer la régression linéaire multiple en prenant comme variable y la colonne des réponses et comme variables explicatives (xn) :

- les 15 colonnes d'effets.
- les 10 colonnes correspondant aux 4 facteurs et aux 6 interactions d'ordre 2 (comme à la 4ème question).

Vérifier que la régression linéaire multiple permet de retrouver l'ensemble des résultats déjà obtenus.

*NB : lors de la sélection des colonnes, on inclura l'intitulé de ces colonnes (A, B, …, AB, …) et on cochera la case « intitulé présent » pour que le rapport édité par Excel soit plus lisible. Dans les options de sortie, on cochera la case « insérer une nouvelle feuille ». Ignorer les les options « analyse des résidus » et « probabilité normale ».*

# Exercice 5

Pour déterminer les meilleures conditions expérimentales de mesure de l'activité de la phosphatase alcaline du sérum au moyen d'un auto-analyseur , on a d'abord recherché les facteurs influençant la réponse mesurée y, celle ci étant transformée en activité de la phosphatase par un étalonnage.

Il a été décidé lors de l'établissement du plan expérimental d'étudier les 5 facteurs suivants, chacun à 2 niveaux :

| **facteur** | **nom du facteur** | **unité** | **niveau bas (-)** | **niveau haut (+)** |
|---|---|---|---|---|
| A | sulfate de Zinc | $\mu mol.L^{-1}$ | 40 | 80 |
| B | sulfate de Magnésium | $\mu mol.L^{-1}$ | 1,50 | 2,50 |
| C | pH | - | 10,0 | 10,7 |
| D | p nitrophényl phosphate disodique | $mmol.L^{-1}$ | 10 | 20 |
| F | 2 amino 2 méthyl 1 propanol | $mol.L^{-1}$ | 0,20 | 0,60 |

Les 32 essais définissant le plan factoriel complet ont été effectués sur l'auto-analyseur dans un ordre tiré au sort et ont conduit au tableau de résultats suivant :

| n° essai | A | B | C | D | F | Réponse y | n° essai | A | B | C | D | F | Réponse y |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | - | - | - | - | - | 109 | 17 | - | - | - | - | + | 106 |
| 2 | + | - | - | - | - | 113 | 18 | + | - | - | - | + | 120 |
| 3 | - | + | - | - | - | 103 | 19 | - | + | - | - | + | 113 |
| 4 | + | + | - | - | - | 113 | 20 | + | + | - | - | + | 115 |
| 5 | - | - | + | - | - | 103 | 21 | - | - | + | - | + | 109 |
| 6 | + | - | + | - | - | 104 | 22 | + | - | + | - | + | 117 |
| 7 | - | + | + | - | - | 106 | 23 | - | + | + | - | + | 105 |
| 8 | + | + | + | - | - | 123 | 24 | + | + | + | - | + | 115 |
| 9 | - | - | - | + | - | 119 | 25 | - | - | - | + | + | 96 |
| 10 | + | - | - | + | - | 146 | 26 | + | - | - | + | + | 128 |
| 11 | - | + | - | + | - | 111 | 27 | - | + | - | + | + | 95 |
| 12 | + | + | - | + | - | 143 | 28 | + | + | - | + | + | 127 |
| 13 | - | - | + | + | - | 116 | 29 | - | - | + | + | + | 99 |
| 14 | + | - | + | + | - | 145 | 30 | + | - | + | + | + | 131 |
| 15 | - | + | + | + | - | 110 | 31 | - | + | + | + | + | 92 |
| 16 | + | + | + | + | - | 148 | 32 | + | + | + | + | + | 132 |

## 1ère question

Pourquoi a t'on tiré au sort l'ordre d'exécution des essais et comment s'appelle cette opération ?

## 2ème question

Combien d'effets peut on calculer avec les résultats de cette expérimentation ?

- combien d'effets principaux ?
- combien d'interactions d'ordre 2 ?
- combien d'interactions d'ordres supérieurs ? (préciser le nombre d'interactions de 3ème ordre, de 4ème ordre et de 5ème ordre).

## 3ème question

Utiliser la matrice des effets fournie pour calculer les valeurs de tous les effets ainsi que la moyenne générale des réponses.

## 4ème question

a)- Sachant qu'on ne dispose d'aucun renseignement sur la précision de la mesure des réponses, quels effets sont, à votre avis, significatifs ?

b)- Utiliser une représentation graphique pour estimer les effets significatifs (variante de la méthode de Henry dite graphique de probabilité normale)

## 5ème question

Ecrire l'équation numérique du modèle linéaire de prédiction des réponses.

- quelle réponse peut on prédire au centre du domaine expérimental ?
- quelle réponse peut on prédire pour la condition expérimentale suivante :

| nom du facteur | niveau |
|---|---|
| sulfate de Zinc | 70 µmol.L$^{-1}$ |
| sulfate de Magnésium | 1,75 µmol.L$^{-1}$ |
| pH | 10,0 |
| p nitrophényl phosphate disodique | 20 mmol.L$^{-1}$ |
| 2 amino 2 méthyl 1 propanol | 0,40 mol.L$^{-1}$ |

6ème question

Calculer les 32 réponses prédites par le modèle et les 32 résidus (par différence avec les réponses mesurées).

Construire le graphe de probabilité normale des résidus ; détecte t'on une non normalité ou une réponse mesurée suspecte invalidant les calculs d'effets ?

7ème question

a)- Ecrire l'équation de modélisation en représentation qualitative.

b)- Quel est, parmi les traitements expérimentés, celui qui permet de prédire la réponse la plus élevée (préciser le niveau de chacun des facteurs) ?

Quelle est la valeur estimée de cette réponse maximale ?