

**UNIVERSITE MOHAMMED V-AGDAL  
FACULTE DES SCIENCES  
DEPARTEMENT DE CHIMIE  
RABAT**



**Cours de Métrologie & Assurance qualité**  
**Licence Professionnelle**  
**Génie Analytique (chimie)**

*Pr R. BCHITOU*

La métrologie existe depuis plus d'un siècle mais, à quelques exceptions près, cette discipline ne s'est appliquée à la chimie que durant les quinze dernières années.

Ce module permet à l'étudiant :

- ✚ de savoir identifier et affermir la place de la fonction "Métrologie" au sein d'un laboratoire de chimie, en relation avec le système d'assurance qualité en vigueur ou en projet dans cette structure,
- ✚ d'apprendre à avoir confiance et inspirer confiance dans des résultats de mesure, d'analyse ou d'essais,
- ✚ de maîtriser les outils associés,
- ✚ d'acquérir les fondements de base sur le contrôle de qualité et sa mise en application,
- ✚ de compléter sa formation dans les techniques d'expression et de communication.

## *E1 : Métrologie*

- I. Définitions
- II. Précision et exactitude
- III. Introduction à la pratique des Plans d'expériences
- IV. Plans de pesée
- V. Matrice d'Hadamard
- VI. Plans factoriels complets

## *E2 : Assurance qualité*

- I. Démarche qualité
- II. Carte de contrôle
- III. Justesse
- IV. Répétabilité
- V. Test de linéarité

*E1*

# *Métrologie*

# Historique de la métrologie

*"L'homme est la mesure de toutes choses"*

Protagoras (sophiste grec, 485- 411 av JC)

*"La métrologie est une des bases de la qualité"*

## ☞ Les unités de référence :

- ✓ Besoin d'unités de référence (unités de temps, poids, volume, température...).
- ✓ Une classification arbitraire des quantités fondamentales de longueur, masse et temps qui ont déjà une longue histoire.
- Exemple: l'origine de la mesure du Yard anglais était la longueur du bras du roi de l'époque (Henri 1<sup>er</sup>; 1100-1137). Il correspondait à la distance entre le bout des doigts et le nez du roi.
- Le premier « Yard étalon », constitué d'une barre de fer, a été établi entre (1272-1307). D'autres Yards étalons constitués (en argent ou en bronze) au cours des siècles suivants.
- Le Yard Impérial a été établi en 1878 par l'acte sur les Poids et Mesures avec des spécifications précises.

## ☞ Des bases de mesure :

- Le pied de roi ( **0,32483 m** ) se subdivise en 12 pouces,
- le pouce ( **2,706 cm** ) se subdivise en 12 lignes,
- la ligne ( **0,226 cm** ) se subdivise en 12 points
- le point ( **0,188 mm** ).

- Quelques membres de l'Académie Royale des Sciences ralliés à la révolution, proposent l'établissement d'une unité de longueur universelle sur des base géodésiques vue les différences de mesures d'une région à l'autre.

• Le 26 mars 1791 naît le mètre, dont la longueur est établie comme égale à la *dix millionième partie du quart du méridien terrestre*. (la circonférence de la terre est rigoureusement égale à 40 000 km).

# Métrologie Chimique: Validation des Méthodes d'Analyses Chimiques

*Métrologie en chimie est devenu une référence utilisée par les techniciens et chercheurs de laboratoires d'analyses, et d'instituts de recherche. Il est aussi particulièrement recommandé aux ingénieurs et étudiants de 3<sup>ème</sup> cycle (mastère, doctorat) en chimie analytique, environnement et qualité.*

## •Objectifs:

- Acquérir les outils nécessaires pour caractériser une méthode (tests statistiques, planification des expériences...).
- Savoir organiser les essais pour les optimiser d'un point de vue technique, mais également économique.
- Utiliser les données issues des travaux de caractérisation pour estimer l'incertitude du résultat de l'analyse et valider cette méthode.

## •Prérequis:

Connaissances de base en statistiques souhaitées.



# *Chapitre I*

# Définitions

# I/ Définition de la Métrologie

✓ La métrologie peut se définir comme étant "la science de la mesure associée à l'évaluation de son incertitude". Au sens large; c'est une science de la mesure.

**Mesure = résultat du mesurage d'un mesurande**

• **Résultat de la mesure**

**Action de mesurer**

**Grandeur mesurée**

✓ Ce terme est utilisé dans le domaine des mesures de paramètres physiques (par exemple température, pression, dimensions, temps, etc.), et plus fréquemment employé dans des calculs de constantes d'équilibre chimiques (identification/quantification d'éléments ou composés) ou biologiques (par exemple comptage d'organismes microbiens).

✓ C'est une composante de la qualité.

✓ Sa spécificité est dans la validation du résultat.

✓ La **chimie** analytique représente une des pierres de base de la **chimie** et englobe le concept récent de « **métrologie** chimique » dont les principes ont été trop souvent sous-estimés par les analystes.

**Objectif : Obtenir des mesures fiables ( niveau de confiance élevé)**

# La métrologie présente trois aspects

## Métrologie des équipements

- **Vérification du fonctionnement correct des appareils par mesure de grandeurs fondamentales**  
(ex : température, pression, masse, volume...)

### ✓ **Nécessité**

- locaux adaptés,
- personnel formé et compétent

## Contrôle qualité des mesures

- **Détermination des incertitudes**  
sur les mesures -> sur les analyses

## Validation des méthodes « alternatives »

La métrologie et ses exigences sont intégrées dans les normes qualité concernant les laboratoires : norme ISO 17025 et BPL.

# I.1/ Métrologie des équipements

- Vérification du bon fonctionnement des équipements avec :
  - Rendus de mesures *fidèles* et *justes*
  - Respect des spécifications du constructeur...

*-Mesures fidèles : même résultat obtenu ou mesures voisines en répétant le mesurage plusieurs fois.*

*-Mesures justes : mesure en accord avec la valeur attendue.*

- Vérification de chaque type d'appareil ( balance, thermomètre, spectrophotomètre...) selon un cahier des charges :
  - extrêmement précis et rigoureux,
  - faisant l'objet d'une norme constructeur.

## Un cahier de vie accompagne chaque instrument

Document qui recense toutes caractéristiques et les interventions faites sur un instrument : date d'achat, spécifications, calibration, réglage, réparations...

## **Exemple : Critères de vérification d'un spectrophotomètre**

- **Alignement optique**

Vérification du centrage correct du faisceau lumineux ainsi que de sa stabilité (forme et taille) dans le temps : maintenance en atelier

- **Présence de lumière parasite**

Vérification de l'unicité de la longueur d'onde du monochromateur arrivant sur la cellule photoélectrique : maintenance en atelier

- **Exactitude photométrique**

Vérification de la justesse de la longueur d'onde affichée

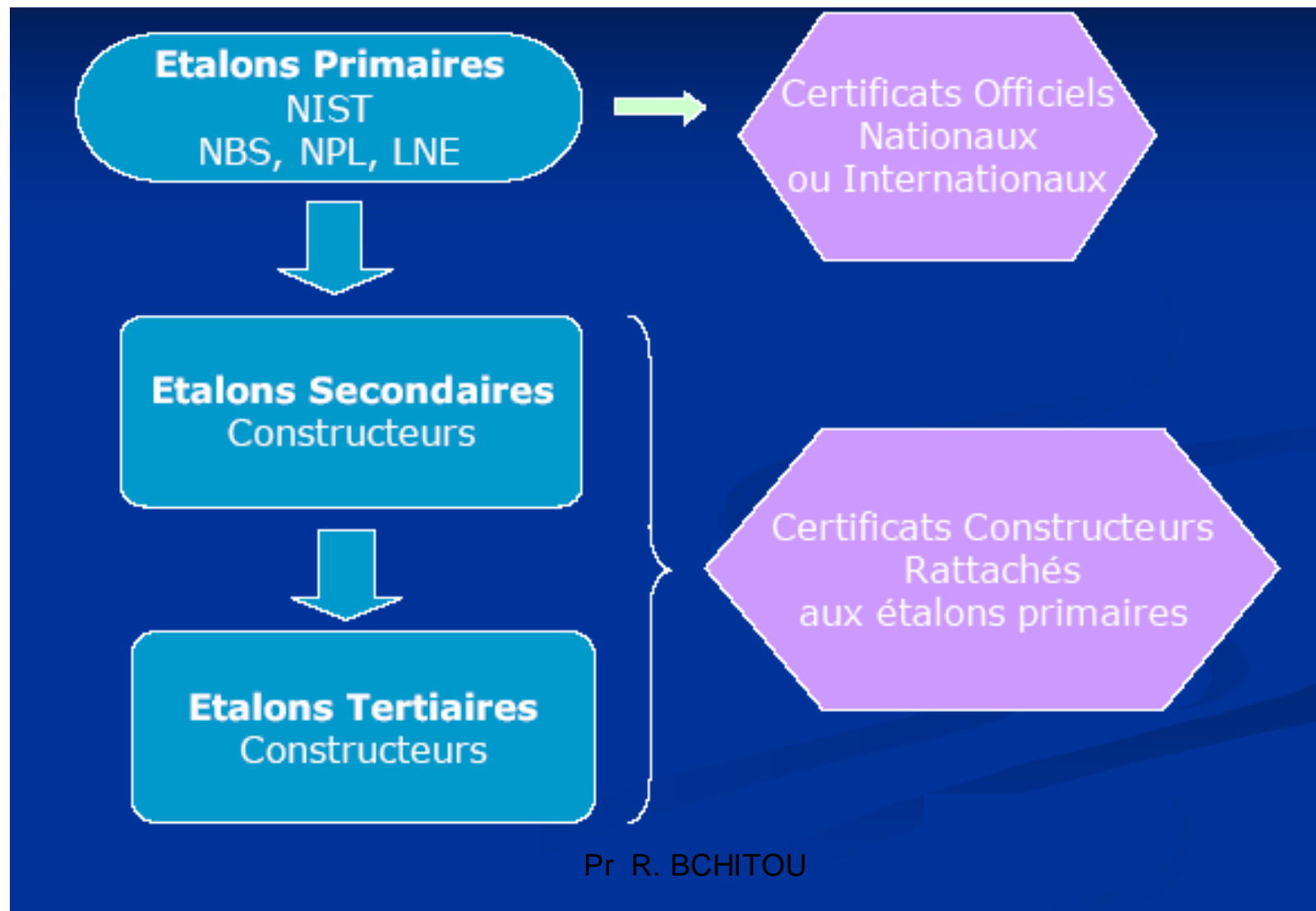
- **Exactitude de l'absorbance et de sa répétabilité**

Vérification de l'absorbance connue d'un échantillon et de sa stabilité dans le temps.

- **Résolution spectrale**

Contrôle de la capacité du spectrophotomètre à séparer deux longueurs d'onde très proches (valable pour des spectrophotomètres haut de gamme) correspondant aux spécifications de l'appareil.

# La métrologie d'un appareil est réalisée avec des **matériaux de référence** ou **des étalons certifiés**



## I.2/ Validation de méthodes alternatives

La validation d'une méthode alternative est basée sur :

✓ Méthodes de dosage dites de « **référence** » :

- reconnues par la communauté scientifiques et faisant l'objet de NORMES.  
Méthodes réputées **fidèles et justes et d'incertitude connue**
- mais souvent longues, chères, nécessitant des appareils sophistiqués...

✓ Utilisation par le laboratoire dans ce cas des méthodes « **alternatives** »:

- utilisée à la place d'une méthode de référence.
- pas forcément reconnue par la communauté scientifique ( pas de norme)
- pouvant être propre à un laboratoire.

Mais elle doit être **VALIDÉE**, c'est à dire contrôlée.

# Exemples de critères de validation de méthodes alternatives

## ➤ Fidélité

Une méthode est fidèle lorsqu'elle donne toujours le même résultat ou des résultats voisins si on la répète sur le même échantillon..

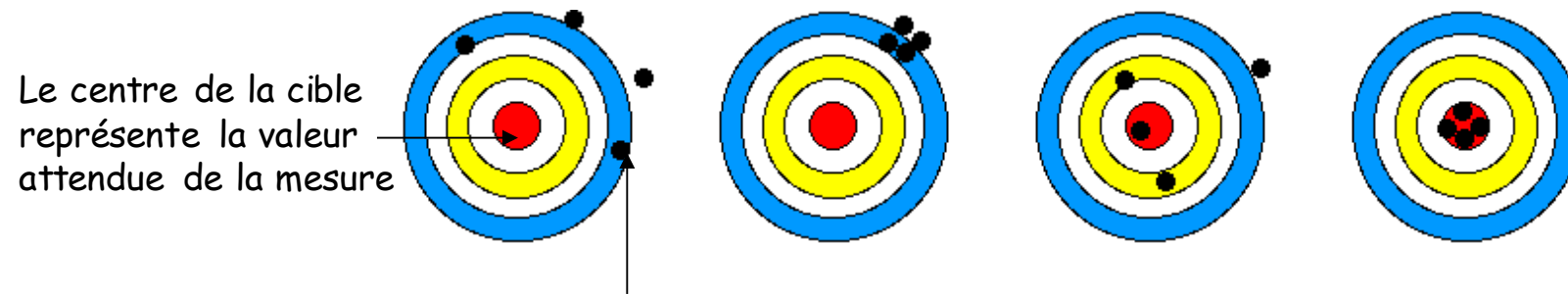
**-Répétabilité:** Fidélité sous des conditions de répétabilité (même méthode , même laboratoire, même opérateur, même équipement et pendant un court intervalle de temps)

**-Reproductibilité:** Fidélité sous des conditions de reproductibilité (même méthode dans différents laboratoires, avec différents opérateurs et utilisant des équipements différents).

## ➤ Justesse

Étroitesse d'accord entre le résultat d'une mesure et la valeur attendue (CIBLE ou valeur réputée vraie)





Les points représentent les résultats de différentes mesures  
(en conditions de répétabilité ou de reproductibilité)

**Non fidèle**  
**Non juste**

**Fidèle**  
**Non juste**

**Non fidèle**  
**Juste**

**Fidèle**  
**Juste**

**Représentation de la fidélité et de la justesse sous forme de cible**

## ➤ Sensibilité

Rapport entre la variation instrumentale et la variation de la concentration.

- Par exemple une méthode de dosage colorimétrique est sensible si elle est capable de donner des A franchement différentes pour des concentrations assez proches.

## ➤ Limite de détection

Plus petite quantité ou concentration distinguable d'un blanc.

## ➤ Limite de quantification

Plus petite quantité ou concentration pouvant être mesurée avec un risque d'erreur connu.

# Les concepts de base de la métrologie en chimie

la métrologie en chimie représente la voie à suivre pour améliorer la qualité des données analytiques. Elle englobe des concepts de base telles que : la traçabilité, l'incertitude, l'étalonnage et la validation des mesures, qui sont définies ci-dessous :

## *A/Traçabilité*

Afin d'améliorer la qualité de l'analyse chimique; la métrologie est construite sur la possibilité d'établir une relation de comparaison entre la grandeur à mesurer, le mesurande et une référence connue de même nature.

- Donc la traçabilité correspond à la possibilité de relier le résultat d'une mesure (ou de la valeur d'un étalon) à des références déterminées, habituellement des étalons internationaux ou nationaux afin d'obtenir un résultat quantifié.

**B/L'étalonnage** : est l'opération qui permet d'effectuer des mesures de grandeurs connues  $G_i$  avec l'instrument de mesure donnant les valeurs  $M_i$ . On établit alors une courbe donnant les écarts entre les valeurs données par l'appareil et les valeurs des grandeurs connues.

- On distingue deux types d'étalonnage:

➤ ***l'étalonnage de l'appareillage*** : qui correspond au test du paramètre physique mesuré par l'appareil.

➤ ***l'étalonnage de la méthode*** : qui permet d'établir la relation signal-quantité de substance ; ce dernier doit être réalisé avec des étalons et des matériaux de références adaptés.

## **C/ L'incertitude**

La valeur vraie d'une mesure ne peut jamais être connue, elle peut seulement être approchée.

➤ L'incertitude sur la mesure est un paramètre associé au résultat de cette mesure, qui caractérise la dispersion des valeurs qui pourraient raisonnablement être attribuées au mesurage.

**D/Validation d'une méthode d'analyse** : est l'opération par laquelle on s'assure que les résultats répondent au problème de manière satisfaisante pour l'utilisateur.

➤ Valider une méthode consiste à démontrer, avec un degré de confiance élevé et sous une forme documentée, que la méthode permet d'obtenir un résultat analytique qui atteint les spécifications définies à l'avance.

➤ La validation se poursuit par une évaluation des résultats obtenus par la méthode pour l'analyse d'un matériau de référence certifié, de manière interne au laboratoire.

# Systeme Internationale d'Unité (S.I)

• Les unités SI sont donc, dans un système métrologique idéal, *les références ou étalons primaires* auxquels toute mesure doit pouvoir être traçable.

• La métrologie est basée sur une référence internationale qui est le système international d'unités (SI) qui est composé des unités suivantes :

- le mètre (m), unité de longueur ;

- le kilogramme (kg), unité de masse ;

- la seconde (s), unité de temps ;

- l'ampère (A), unité d'intensité de courant ;

- le kelvin (K), unité de température thermodynamique;

- la mole (mol), unité de quantité de matière;

- la candela (cd), unité d'intensité lumineuse.

- En chimie, le mole et le kilogramme sont les références primaires.

- Les normes et les guides internationaux existants ont été établis selon des principes généraux et ne font pas de différences marquées entre les mesures physiques, chimiques ou biologiques.

# Liste des organisations concernées par la métrologie en chimie

**ISO** : International Organisation for Standardization. Genève

**BIPM** : Bureau International des Poids et Mesures.

**CIPM** : Comité International des Poids et Mesures.

**CCQM** : Comité Consultatif pour la Quantité de Matière.

**CGPM** : Conférence Générale des Poids et Mesures

**CITAC** : Coopération on International Traceability in Analytical Chemistry

**SMT** : Standard, Measurements and Testing Programme, Commission Européenne.  
Bruxelles

**REMCO** : ISO Committee for Reference Materials. Genève

**NIST** : National Institute of Standards and Technology. USA

**COMAR** : International Data Bank on Reference Materials, Laboratoire National  
d'Essais. Paris

**CEN** : Comité Européen de Normalisation. Genève

**AFNOR** : Association Française de Normalisation. Paris

**ILAC** : International Laboratory Accreditation Conference.

**EURACHEM** : European Focus for Analytical Chemistry, Teddington. U.K.

**ICH** : International Conference on Harmonization of Technical Requirements for the  
Registration of Pharmaceuticals for Human Use.

**IUPAC** : International Union of Pure and Applied Chemistry

**LNE** : Laboratoire National d'Essais, Paris

**FDA** : Food and Drug Administration, USA.



## Exemple :

Les normes se répandent dans l'industrie sont celles de la série des ISO 9000.

les normes qualité concernant les laboratoires : norme ISO 17025 et BPL

Des normes environnementales sont en cours de publication

# Normalisation des instruments de mesure au Maroc

Un comité de normalisation des instruments de mesure est institué afin de répondre aux besoins des industriels dans le domaine de la métrologie. Ce comité adopte des normes dans les domaines suivants:

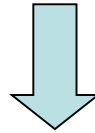
- généralités et définitions de la métrologie marocaine
- mesure dimensionnelle
- mesure de masse
- mesure physico-chimique
- mesure de température
- mesure de pression
- mesure de vitesse
- mesure de volume et de débit

# *Chapitre II*

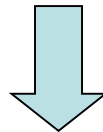
# *Précision et exactitude*

# Comment présenter les résultats de mesures ?

**Encadrement des résultats**  
(intervalles de **confiance**)



**Validité des résultats** (**tests d'hypothèses**)



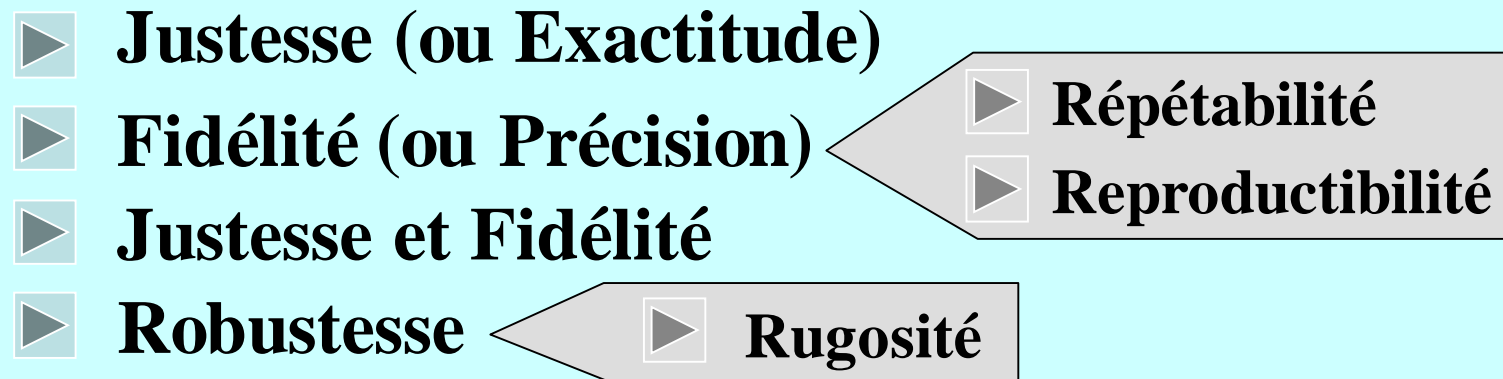
**Etablissement de prévisions** (utilisation de  
modèles prévisionnels)

## II.1/ Précision et exactitude

Lorsque l'on **réalise des mesures** (des analyses), on veut **rendre un résultat** en s'interrogeant sur la **validité** de ce que l'on présente.

*(Plus la précision est grande, plus les indications sont proches de la valeur vraie).*

Dans ce contexte, cela implique **l'évaluation des caractéristiques** suivantes :



## II.2/ Justesse, Fidélité et Erreur

**II.2.1/ Justesse ou *Exactitude*** : Une méthode est réputée juste quand la moyenne  $\bar{X}$  d'un grand nombre de mesures  $X_i$  est confondue avec la valeur  $X$  du mesurande , quelle que soit la dispersion .

✓ L'erreur de justesse  $J$  est définie par :

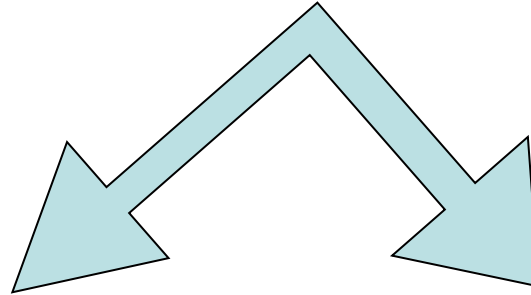
$$\boxed{J = \bar{X} - X} \quad \text{avec} \quad \bar{X} = \sum_{i=1}^n X_i/n \quad (\text{moyenne des } X_i)$$

**II.2.2/ Fidélité ou *Précision*** : caractérise la dispersion d'une série de mesures  $X_i$  d'une même grandeur (échantillon). Il est défini par l'écart type  $\sigma$  ou la variance  $\sigma^2$  :

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2$$

Pr R. BCHITOU

# Fidélité (ou Précision)

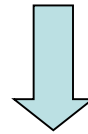


Pour un **PROCÉDE** :

à partir d'un **échantillonnage**  
**du(des) produit(s)** fabriqué(s)  
dans les conditions prescrites.

Pour une **METHODE** :

à partir de prélèvements multiples  
**d'un échantillon homogène** avec les  
conditions d'analyse prescrites.



La fidélité doit être étudiée en utilisant des **étalons** ou des **échantillons** authentiques **homogènes**.

**La fidélité peut être considérée à **trois** niveaux :  
Répétabilité, Reproductibilité et Précision intermédiaire .**

→ Répétabilité  
*(même série d'analyses)*

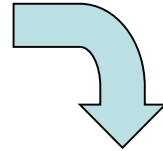
→ Reproductibilité  
*(opérateur et jour et appareillage différents)*

→ Précision intermédiaire  
*(opérateur, ou jour, ou appareillage différent)*

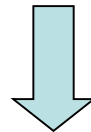


## II.2.2.1/ Répétabilité

### Définition



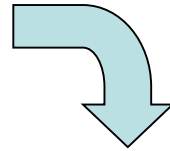
La répétabilité exprime la Fidélité pour les **mêmes conditions opératoires** pour une même série d'analyses dans un **court intervalle de temps**.



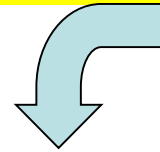
Elle est aussi appelée « **précision intra-essai** ».

## II.2.2.2/ Reproductibilité

### Définition



La Reproductibilité représente les **Variations INTER-Ateliers** ou **INTER-laboratoires**



C'est la mesure de la **dispersion obtenue par plusieurs opérateurs** qui analysent ou mesurent :

- dans des ateliers ou laboratoires différents,  
*et*
- dans des intervalles de temps différents,  
*et*
- éventuellement avec des types d'appareils différents.

## II.2.2.3/Précision intermédiaire

### Définition



la Précision intermédiaire, représente les **Variations INTER-Ateliers ou INTER-laboratoires**



C'est la mesure de la **dispersion obtenue par plusieurs opérateurs** qui analysent ou mesurent :

→ dans des ateliers ou laboratoires différents,

*ou*

→ dans des intervalles de temps différents,

*ou*

→ avec des types d'appareils différents.

## II.2.3/ Les Erreurs Expérimentales

### Deux types d'erreur

```
graph TD; A[Deux types d'erreur] --> B[Erreur Systématique]; A --> C[Erreur Aléatoire]; B --> D["elle varie toujours dans le même sens par rapport à la moyenne. i.e. une erreur qui reste constante pour des mesures effectuées dans des conditions identiques."]; C --> E["elle se répartit de part et d'autre de la valeur moyenne (variance, écart-type)."]; D --> F["Les erreurs systématiques affectent l'exactitude (Justesse)"]; E --> G["Les erreurs aléatoires sont relatives à la précision (Fidélité)"];
```

### Erreur Systématique

elle varie **toujours dans le même sens** par rapport à la moyenne. i.e. une erreur qui reste constante pour des mesures effectuées dans des conditions identiques.

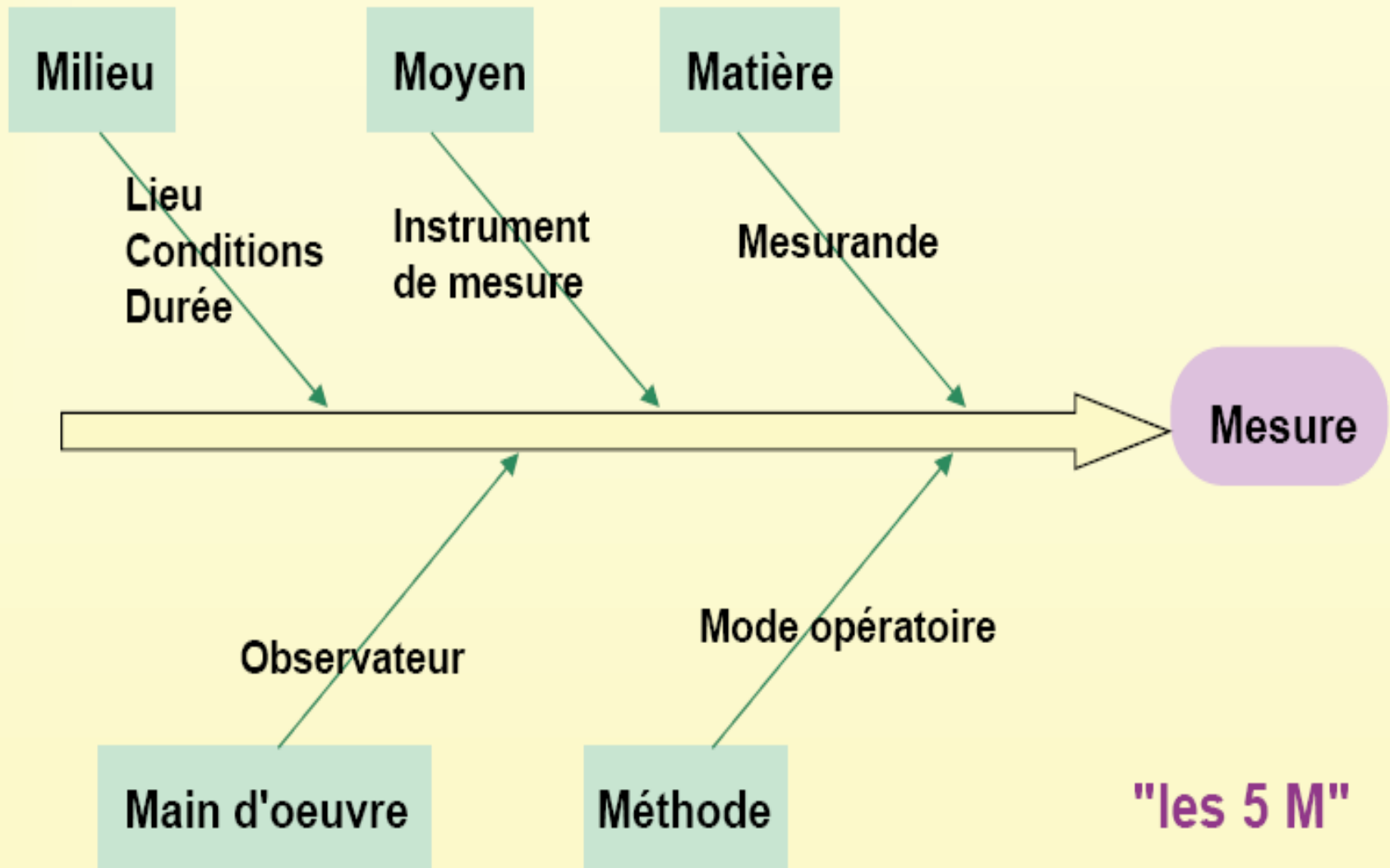
Les erreurs systématiques affectent l'exactitude  
(Justesse)

### Erreur Aléatoire

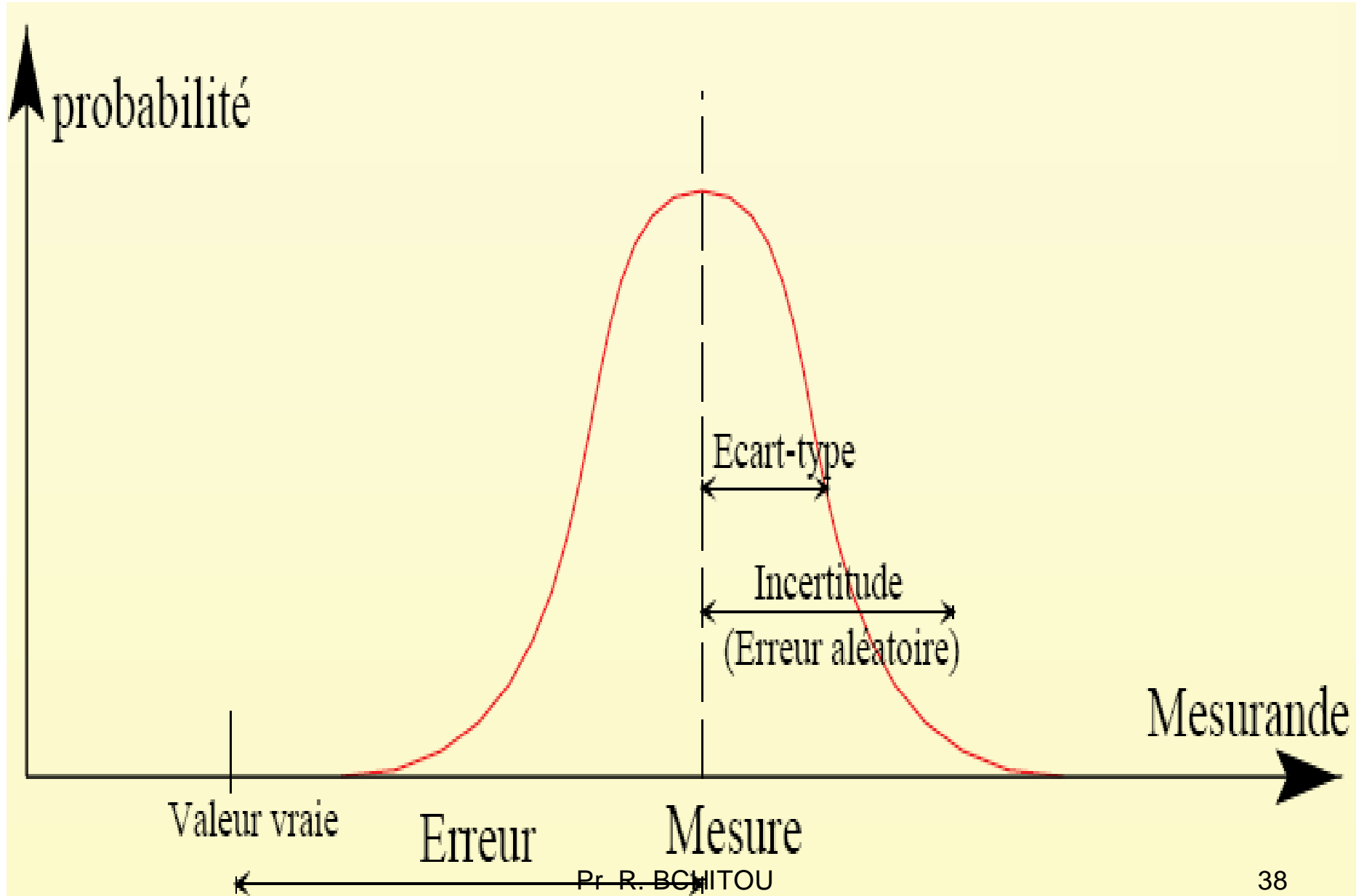
elle se répartit **de part et d'autre** de la valeur moyenne (*variance, écart-type*).

Les erreurs aléatoires sont relatives à la précision  
(Fidélité)

# Recherche des causes d'erreurs

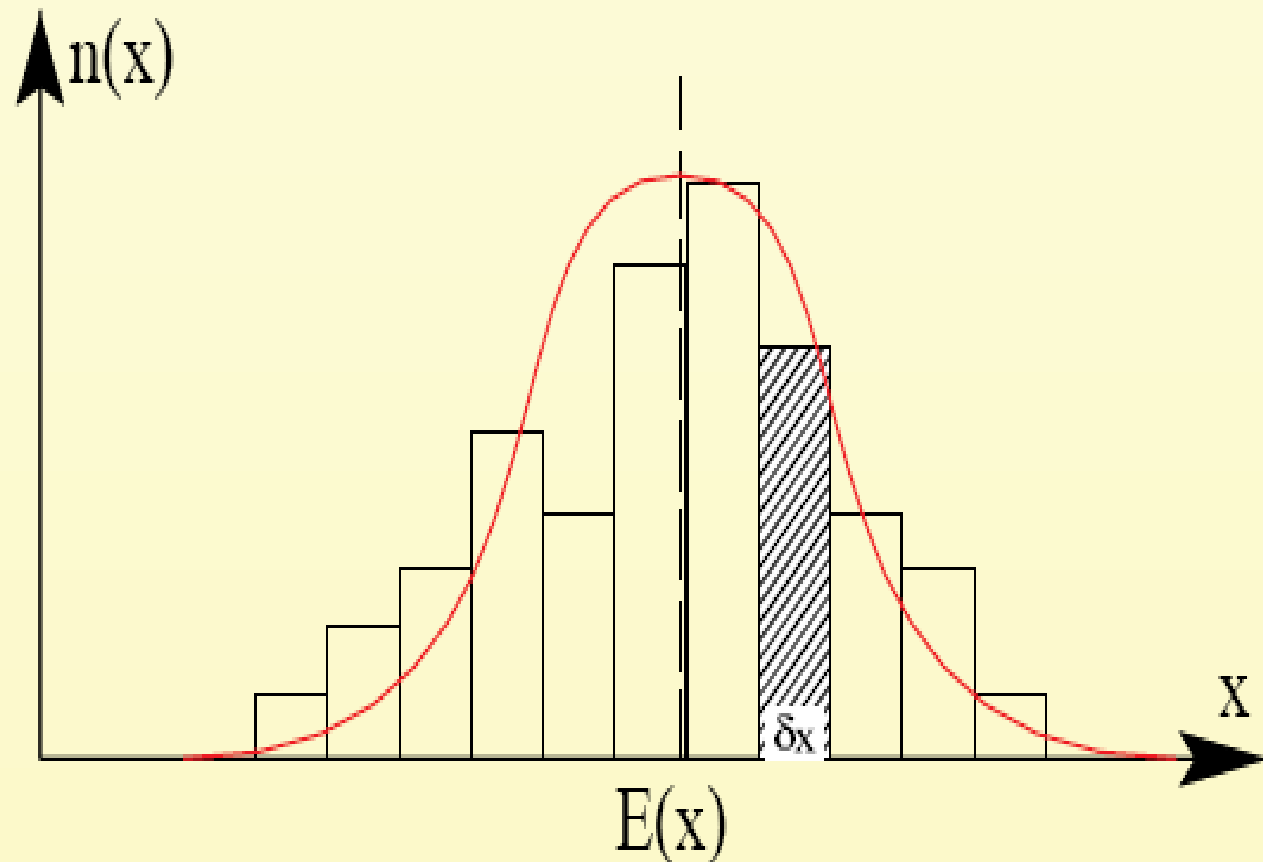


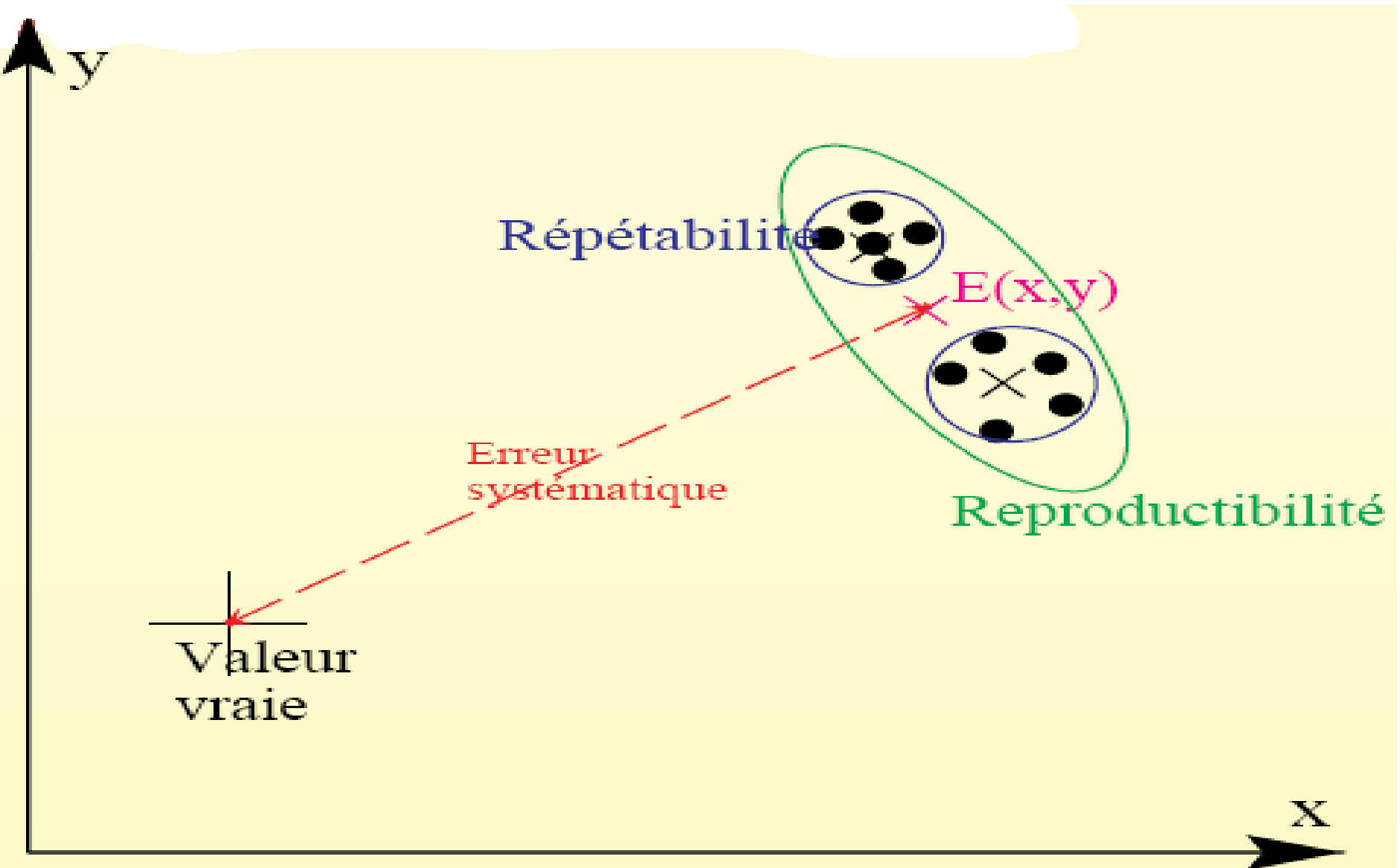
## Démarche de Mesurage



- Soit une série de  $n$  mesurages répétés.

–  $X_1, X_2, X_3, \dots, X_i, \dots, X_n$

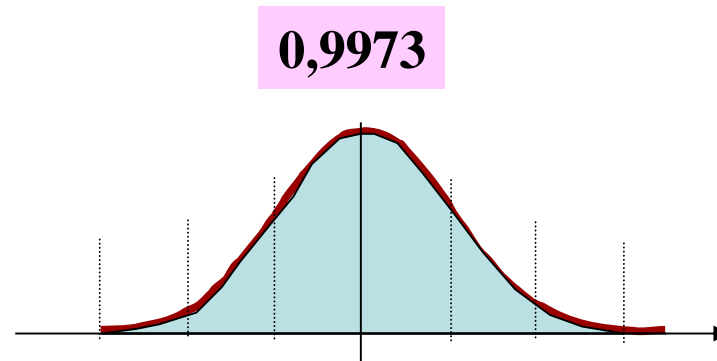




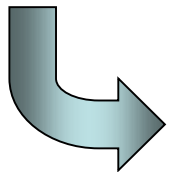


## II.3/ Intervalle de confiance

Pour encadrer un résultat on parlera d'**intervalle de confiance** :



Probabilité = 99,73% pour que  $x$  soit compris dans l'intervalle  $\mu \pm 3 \sigma$



il y a plus de 99% de chances d'obtenir un résultat dont la valeur est égale à la **valeur centrale  $\pm 3 \sigma$** .

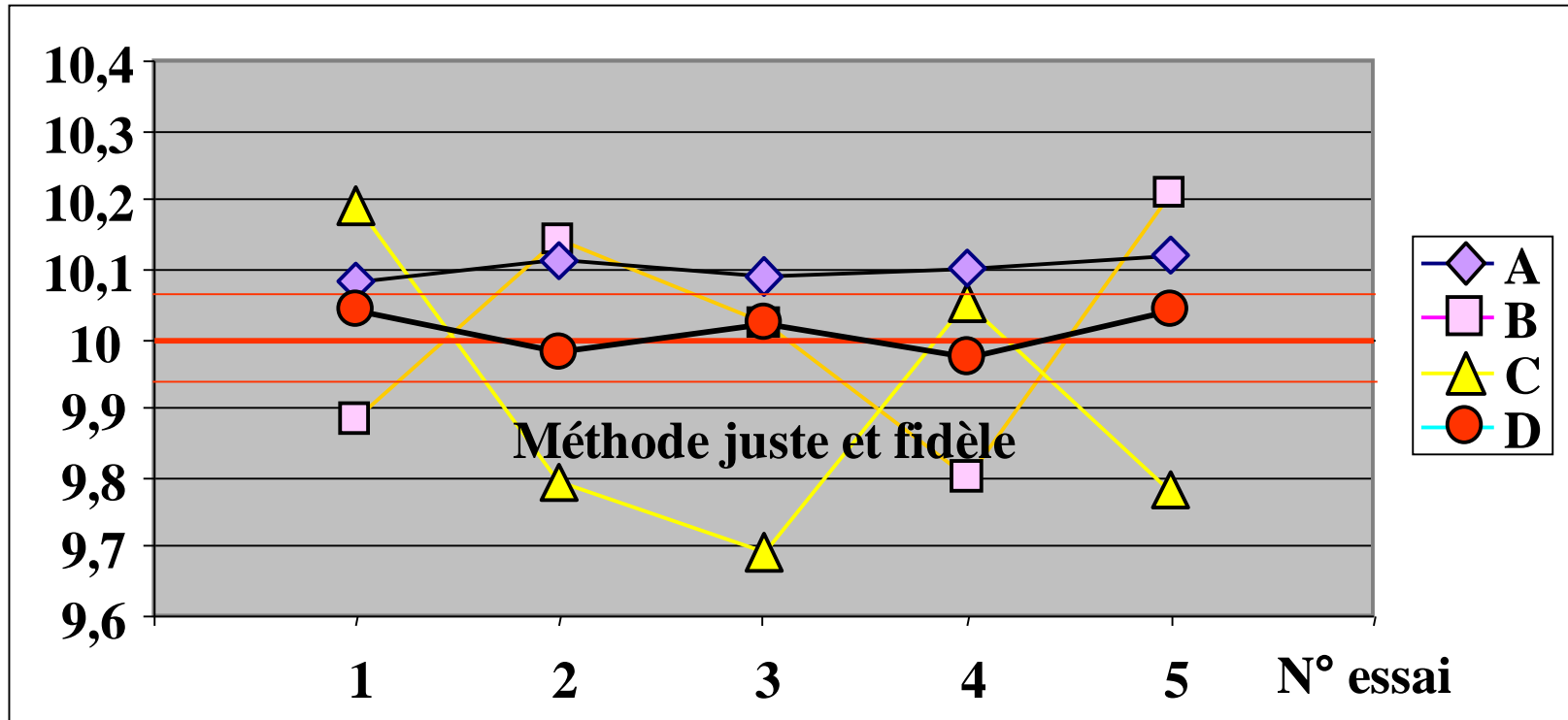
## Exemple :

Quatre opérateurs A, B, C et D dosent 10 ml de solution 0,1M de soude, mesurés exactement (précision instrumentale de 0,05 ml) avec une solution d'acide qui titre exactement 0,1 M . Quel est l'opérateur qui a travaillé avec précision et exactitude?

A	B	C	D
10,08	9,88	10,19	10,04
10,11	10,14	9,79	9,98
10,09	10,02	9,69	10,02
10,10	9,80	10,05	9,97
10,12	10,21	9,78	10,04

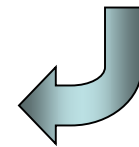
	Inexact	Exact	Inexact	Exact
moyenne	10,10	10,01	9,90	10,01
écart-type	0,016	0,172	0,21	0,033
	Précis	Imprécis	Imprécis	Précis

# Représentation graphique




Valeur **théorique** de **10 ml**, si on estime l'écart-type expérimental à **0,016 ml** (opérateur A) alors :

**99,7%** des valeurs expérimentales doivent être comprises dans l'intervalle **10 ml  $\pm$  0,048 ml**




## II.4/ Méthodologie

Pour la **Maîtrise** d'un Procédé ou d'une Méthode (d'analyse par exemple) **il faut** :



- Procurer une **connaissance totale et non biaisée des possibilités** du Procédé ou de la Méthode telles que : justesse, fidélité et robustesse.

- **Structurer le travail expérimental** de telle manière que les validations appropriées des caractéristiques du Procédé ou de la méthode puissent être **considérées simultanément**.



Procurer une **connaissance totale et non biaisée des possibilités** du Procédé ou de la Méthode telles que : **justesse, fidélité et robustesse.**

## Outils Statistiques

**Structurer le travail expérimental** de telle manière que les validations appropriées des caractéristiques du Procédé ou de la méthode puissent être **considérées simultanément.**

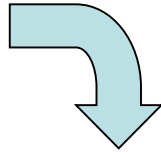
## Méthodologie des Plans d'Expériences

Voir E2 PIDD ET LIC PROF

## II.4/ Mise au point de méthodes

### II.4.1/ ROBUSTESSE

#### Définition



La robustesse d'un procédé ou d'une méthode est une mesure de son **aptitude à ne pas être affectée par de petites variations délibérées des paramètres de la méthode.**



Elle fournit une **indication de sa fiabilité** pour un usage normal.

## II.4.1.2/ Développement de Méthodes et Robustesse

"L'analyse", "Le procédé », qu'ils soient chimiques ou physico-chimiques, impliquent la **mise au point** et l'utilisation de **"méthodes"**.

**Il peut s'agir :**

✓ **d'adapter une méthode existante** au matériel dont on dispose ou à un nouveau type d'échantillons que l'on doit traiter (ajuster des volumes de réactifs, des temps et des températures de réaction et/ou des réglages d'appareils...) pour obtenir des performances satisfaisantes.

✓ **de mettre au point une méthode originale**



se terminer par une **optimisation**

Pr. R. BCHITOU





## II.4.1.3/ Notion additionnelle à la Robustesse

Pour la capacité d'un Procédé ou d'une Méthode à fournir des «produits» **conformes** on peut distinguer :

### La Robustesse

Faible sensibilité à une **légère variation** des facteurs expérimentaux **maîtrisables**.

Paramètres **du** Procédé :  
température, concentration,  
vitesse d'outils ...

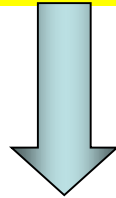
### La Rugosité

Faible sensibilité à une **légère variation** des facteurs expérimentaux **non maîtrisables**

Paramètres **hors** Procédé :  
temps, opérateur, espace,  
matériel, consommables...

## II. 5/ Loi normale (ou loi de Gauss)

Laplace et Gauss ont démontré que, pour la plupart des phénomènes physiques observables, **les mesures expérimentales suivent une même loi de probabilité :**



une même fonction de densité de probabilité appelée  
**Loi Normale.**

## II.5.1/ Forme analytique de la loi normale

Cette loi, qui décrit une **variable aléatoire**, est caractérisée par deux paramètres :

Un paramètre de position ou de centrage : la **moyenne  $\mu$**

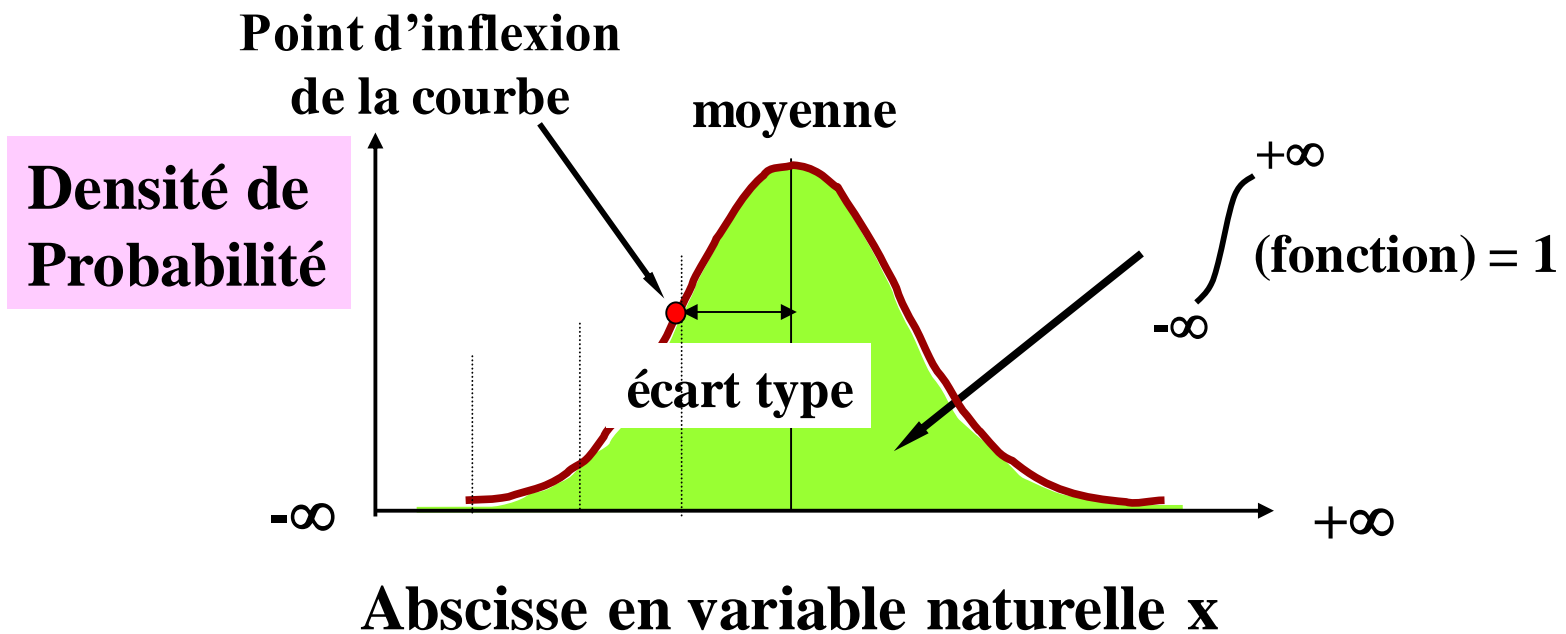
un paramètre de dispersion : **l'écart-type  $\sigma$** .

Sa forme analytique est :

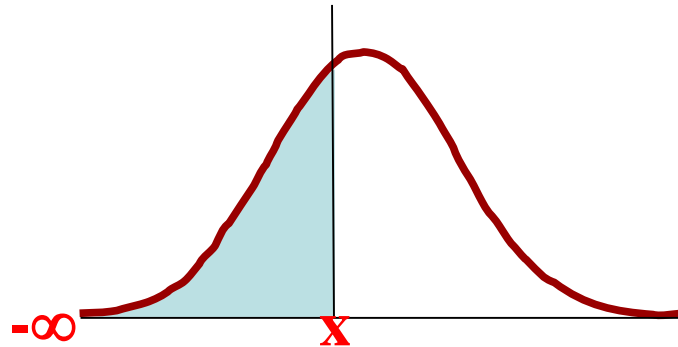
$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}$$

## II.5. 2/ Graphe de la Loi normale

### Distribution symétrique centrée sur la moyenne



## II.5.3/ Propriétés de la loi Normale

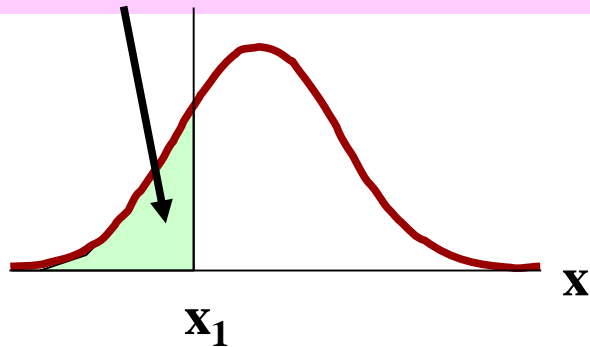


$\int_{-\infty}^x$  (fonction) = **probabilité** pour que la valeur de la variable **X** soit comprise entre  **$-\infty$**  et  **$x$**

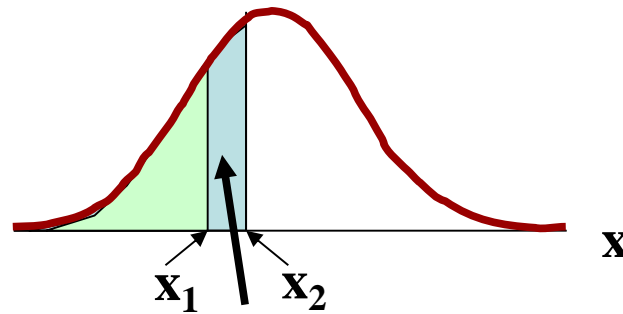
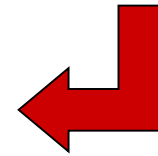
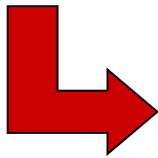
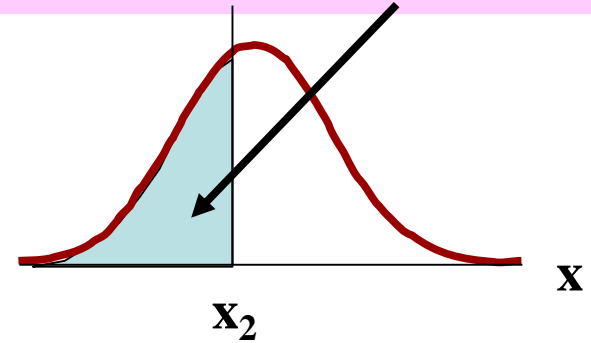


**Probabilité pour qu'une valeur d'abscisse soit comprise entre deux valeurs données ?**

Probabilité  $p_1$  pour qu'une valeur de  $x$  soit **inférieure** à  $x_1$

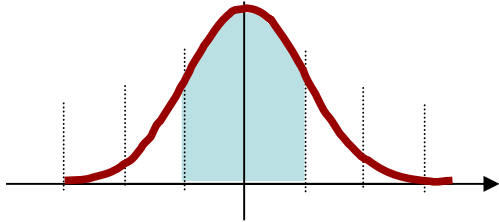


Probabilité  $p_2$  pour qu'une valeur de  $x$  soit **inférieure** à  $x_2$



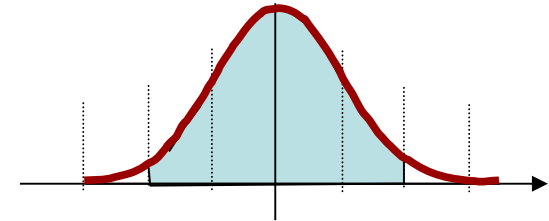
$p_2 - p_1 =$  Probabilité pour qu'une valeur de  $x$  soit **comprise entre  $x_1$  et  $x_2$**

**0,6827**



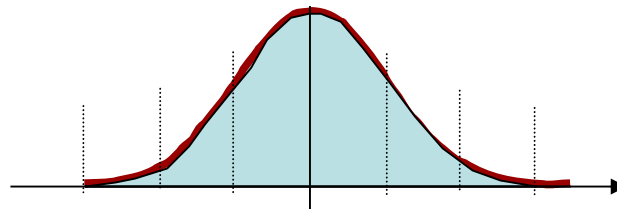
**Probabilité = 68,27% pour que x soit  
compris dans l'intervalle  $\mu \pm 1 \sigma$**

**0,9545**



**Probabilité = 95,45% pour que x soit  
compris dans l'intervalle  $\mu \pm 2 \sigma$**

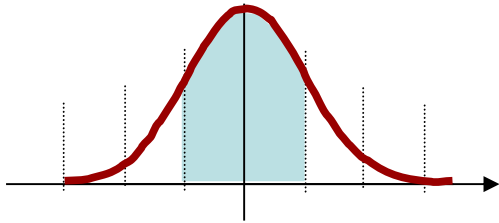
**0,9973**



**Probabilité = 99,73% pour que x  
soit compris dans l'intervalle  $\mu \pm 3 \sigma$**

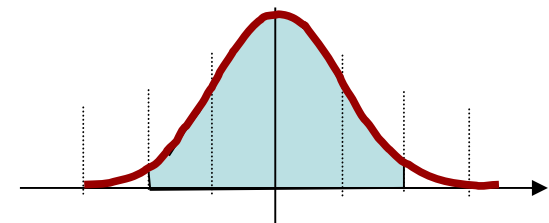


**0,6827**



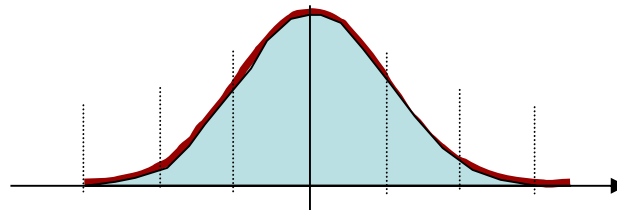
**Probabilité = 68,27% pour que x soit  
compris dans l'intervalle  $\mu \pm 1 \sigma$**

**0,9545**



**Probabilité = 95,45% pour que x soit  
compris dans l'intervalle  $\mu \pm 2 \sigma$**

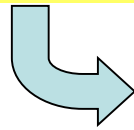
**0,9973**



**Probabilité = 99,73% pour que x  
soit compris dans l'intervalle  $\mu \pm 3 \sigma$**

## II.5.3/ Loi normale Standard

Telle que nous venons de la définir, la loi Normale est fonction de  $\mu$  et  $\sigma$  exprimés avec l'unité de la variable  $X$  :

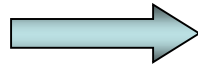


chaque cas est donc un cas particulier

On peut rendre la loi universelle à l'aide d'un changement de variable :



en prenant la moyenne  $\mu$  de la distribution pour origine de l'axe des  $x$ ,



avec l'écart type de la distribution comme unité de mesure.

**Cette nouvelle variable s'appelle variable centrée réduite z, elle est sans dimension :**

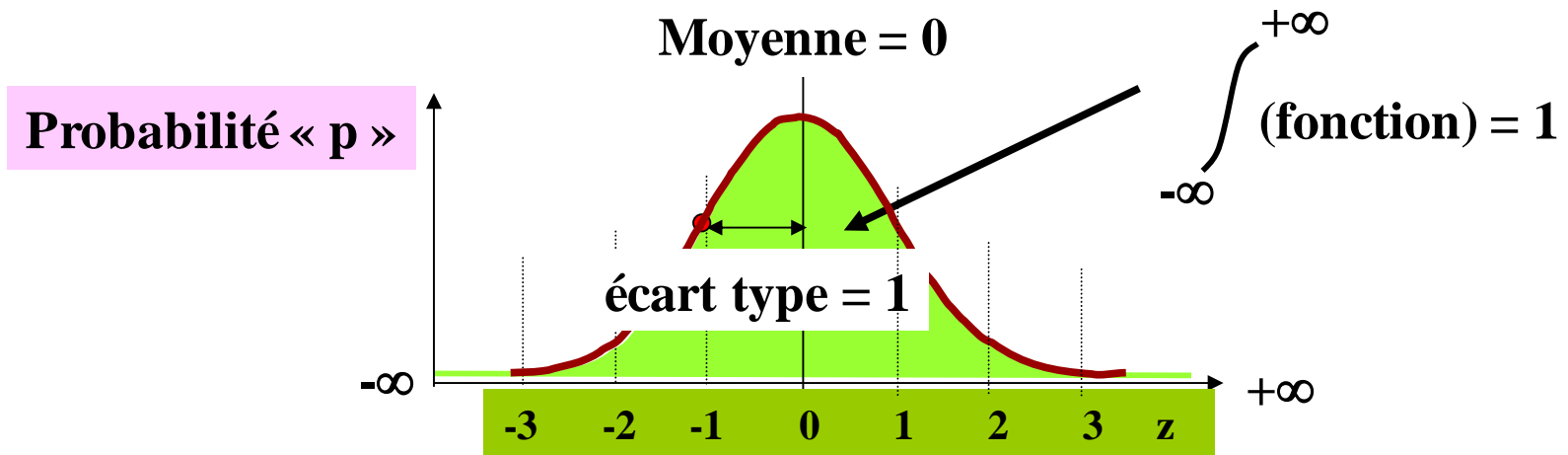
$$Z_i = \frac{X_i - \bar{X}}{\sigma}$$

**Les caractéristiques de Z sont : moyenne = 0 et écart type = 1**

**La forme analytique de la Loi Normale Standard est :**

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

# Loi normale standard



**Abscisse en variable centrée réduite z**

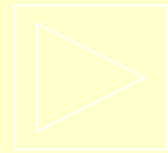
**C'est une loi universelle, indépendante des unités de la variable étudiée**

**Elle s'utilise de la même manière que la loi normale**

# Intervalles de confiance

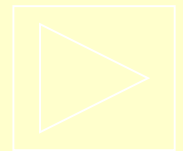
Valeur individuelle :

$$\bar{x} \pm z \sigma$$



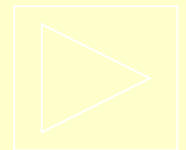
Grands Echantillons  
(> 30 répétitions) :

$$\bar{x} \pm z_c \frac{\sigma}{\sqrt{n}}$$



Petits Echantillons  
(< 30 répétitions) :

$$\bar{x} \pm t_c \frac{\sigma}{\sqrt{n}}$$



## II.5.3/ La distribution de Gauss ou « normale »

Si l'on réalise un **nombre infini de mesures sur le même échantillon et dans les mêmes conditions expérimentales**, on obtient une série de résultats qui se répartissent avec des fréquences qui suit une distribution normale.

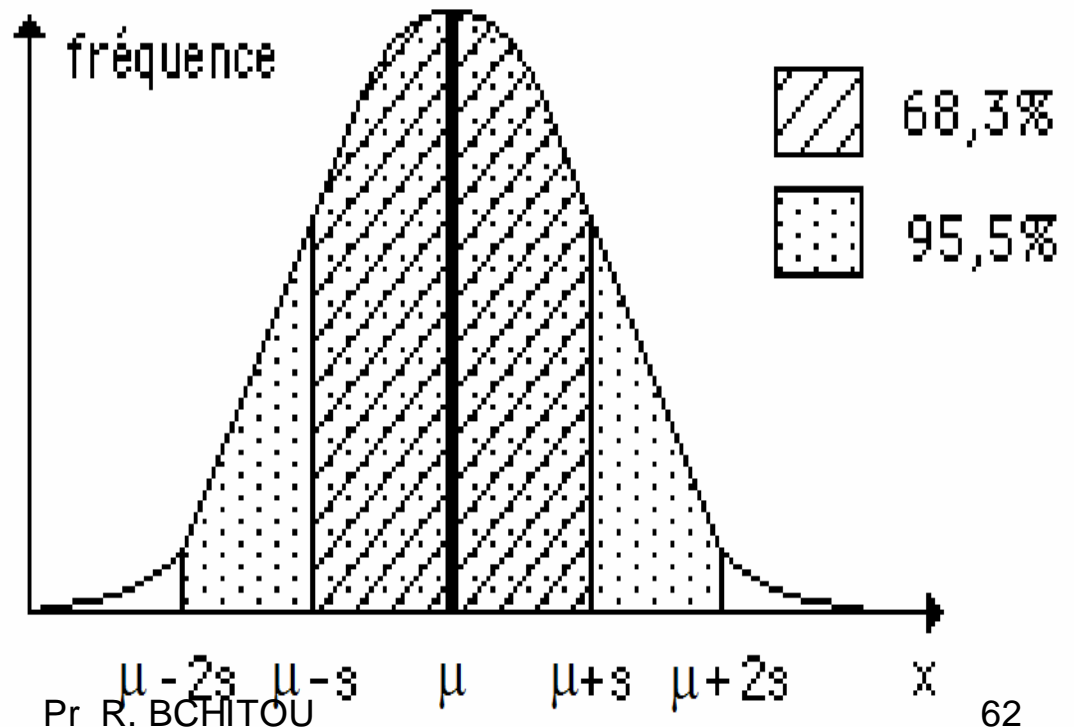
série de n mesures :  $x_1 x_2 x_3 \dots$

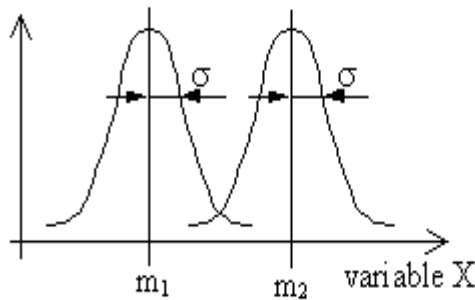
$$\text{moyenne } \mu = \frac{\sum x_i}{n} = \text{médiane}$$

$$\text{variance} = \sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

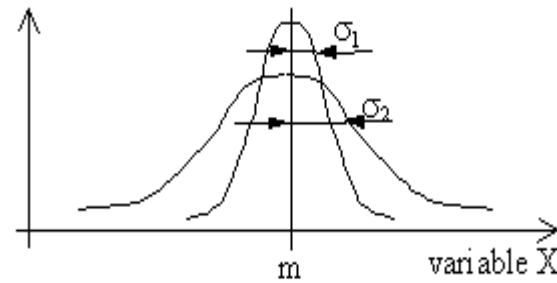
$$\text{écartype} = \sigma$$

Distribution "normale"





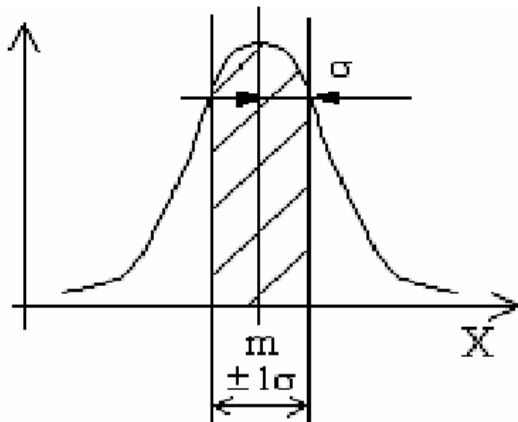
moyenne différentes  
dispersion identique



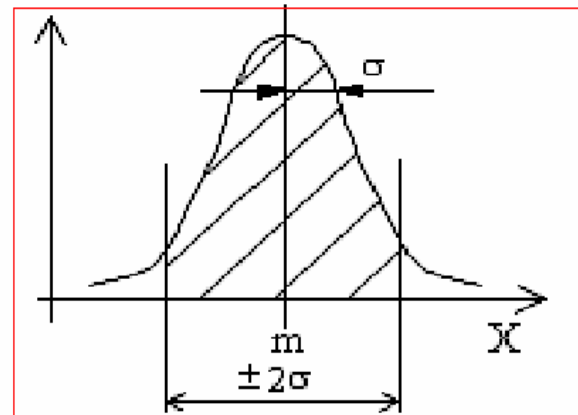
moyenne identique  
dispersions différentes

• Plus l'écart-type est **grand**, plus la dispersion est **grande**

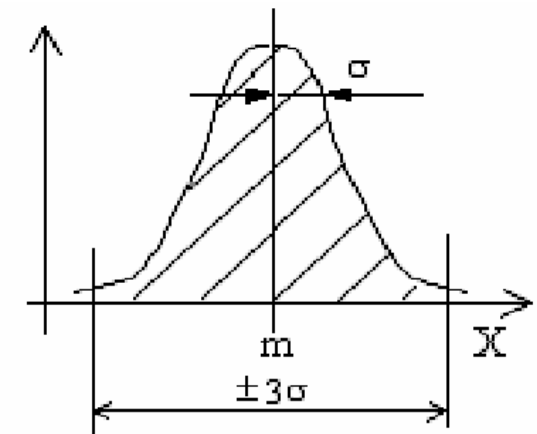
## Répartition des mesures, intervalle de « confiance »



environ 68% des mesures  
sont comprises dans l'intervalle  
 $m \pm 1\sigma$



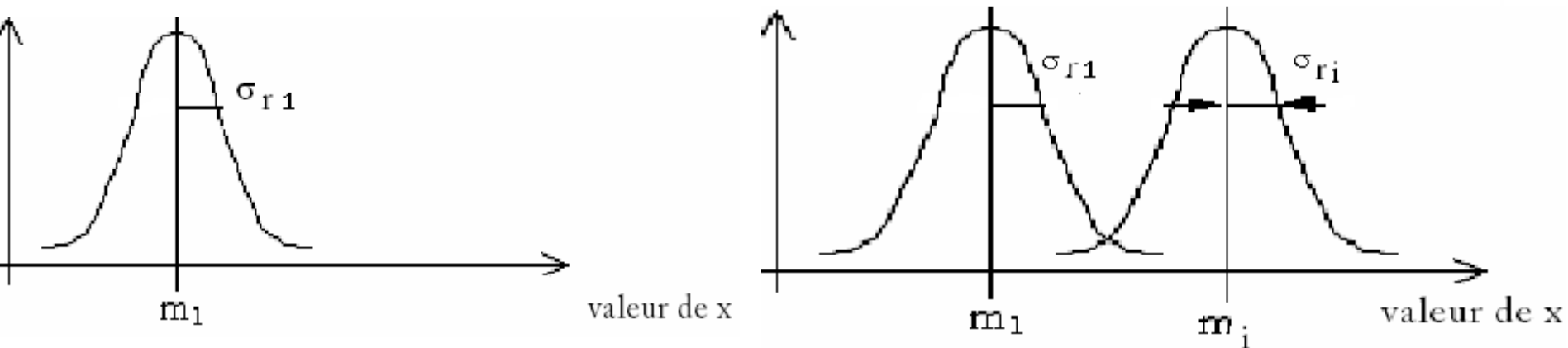
environ 95% des mesures  
sont comprises dans l'intervalle  
Pr. R. BCHITOU



environ 99,8% mesures  
sont comprises dans l'intervalle  
 $m \pm 3\sigma$  63

## II. 6/ Modèle statistique des essais inter-laboratoire

➤ Le laboratoire  $L_1$  effectue un dosage sur l'échantillon **E avec une méthode donnée en condition de répétabilité**, il obtient une série de mesure qui suit une distribution gaussienne: moyenne  $m_1$ , écart type de répétabilité  $\sigma_{r1}$

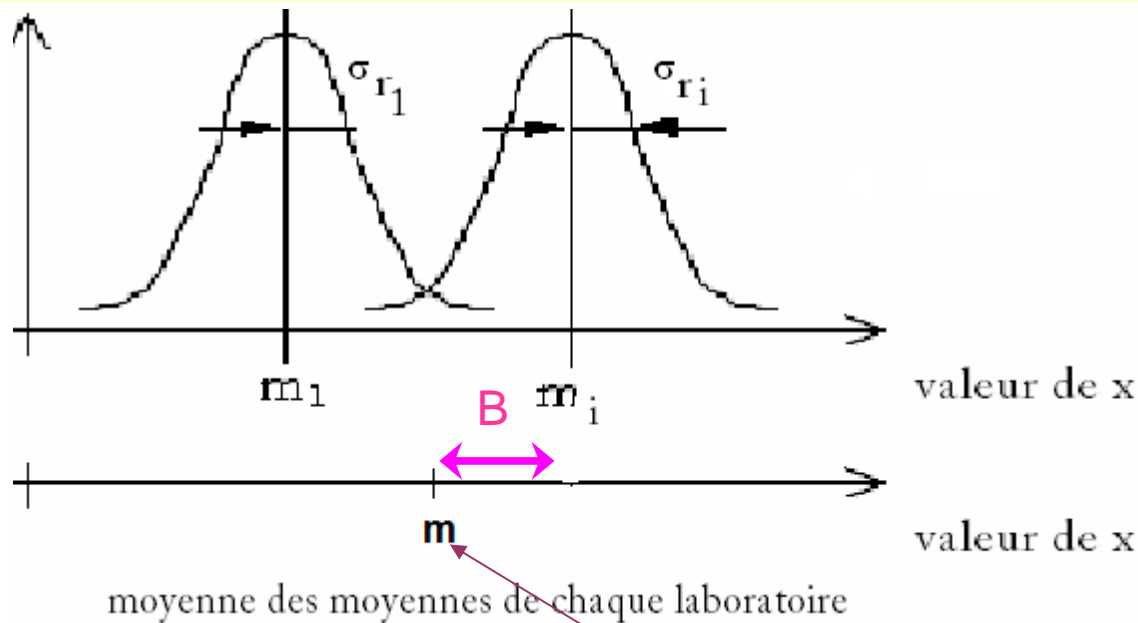


➤ Le laboratoire  $L_i$  effectue un dosage sur le **même échantillon E avec la même méthode et en condition de répétabilité**, il obtient une série de mesure qui suit une distribution gaussienne : moyenne  $m_i$ , écart type de répétabilité  $\sigma_{ri}$

Si la méthode est bien normalisée, **les écart types de répétabilité  $\sigma_{ri}$  sont identiques**. C'est cette valeur identique qui est nommée l'écart type de répétabilité de la méthode  $\sigma_r$



\* Mais les moyennes des laboratoires sont différentes

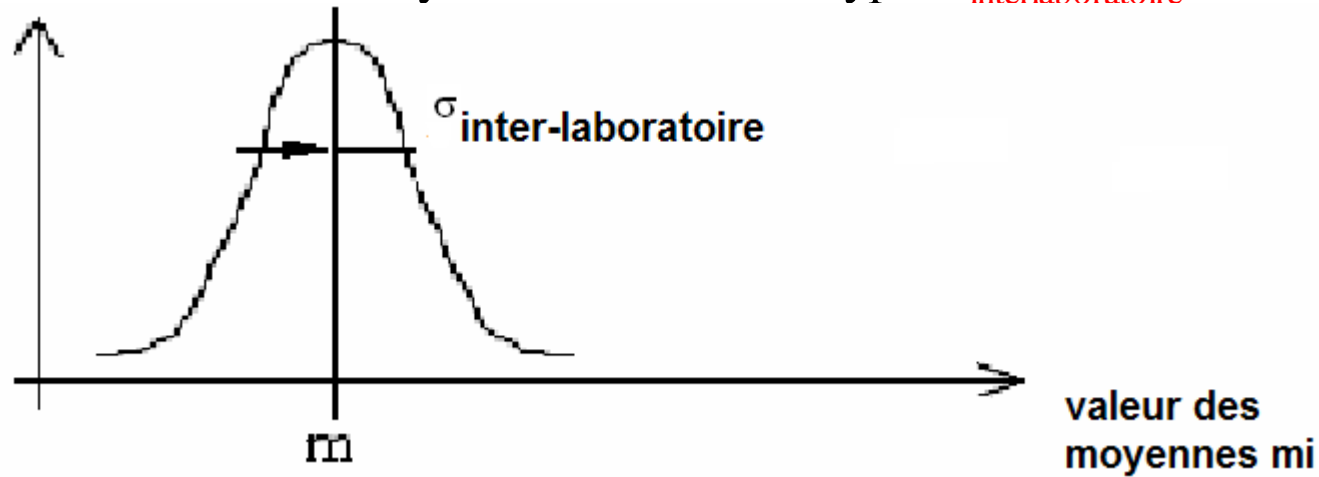


Si on fait la moyenne des moyennes :  $m = (m_1 + m_2 + \dots + m_n) / n$

**Le jour J, un laboratoire donné  $L_i$  donne un résultat qui s'écarte de la moyenne  $m$  avec une déviation  $B$  (biais du laboratoire le jour J)**

**$B$  prend une valeur aléatoire, varie tous les jours, sa moyenne est nulle**

- En reportant les moyennes de chaque laboratoire sur un histogramme, on obtient une distribution normale de moyenne  $m$  et d'écart type  $S_{\text{interlaboratoire}}$



• Soit  $X$  le résultat d'un dosage, effectué par un laboratoire  $L_i$ , un jour  $J$

$$X = m + B + e$$

Valeur supposée vraie →  $m$      
 Biais de  $L_i$ , le jour  $J$  →  $B$      
 Alea de répétabilité de la méthode →  $e$

Alea de reproductibilité

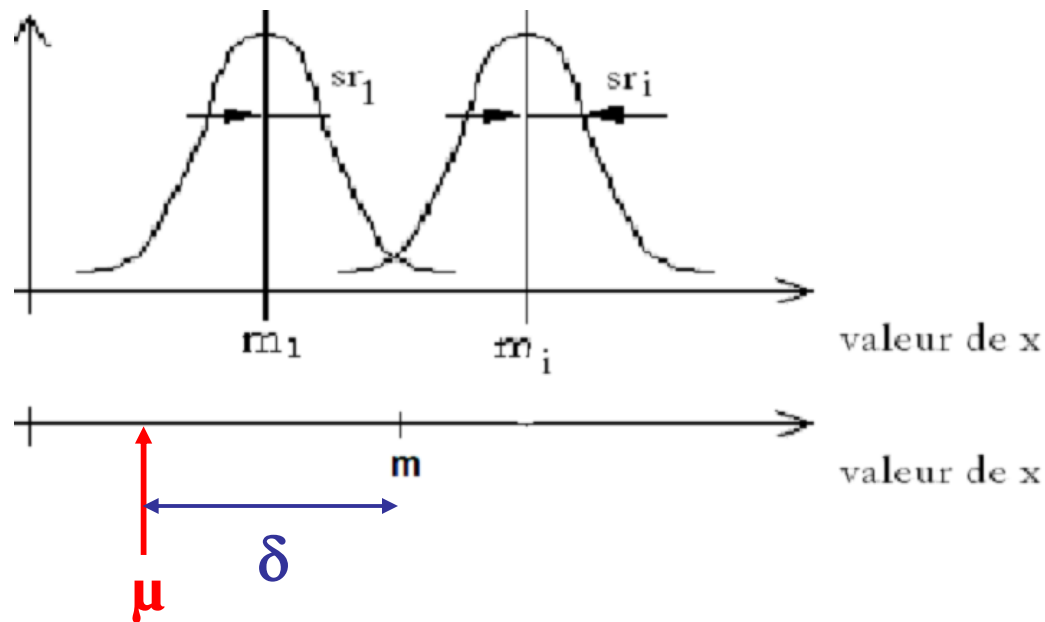
Mathématiquement :

$$\sigma_R = \sigma_{\text{reproductibilité}} = \sqrt{(\sigma_{\text{interlaboratoire}})^2 + (\sigma_{\text{répétabilité}})^2}$$

D'où :

$$\text{Variance } (B+e) = \text{Variance}(B) + \text{Variance}(e)$$

$$\text{ou } (\sigma_{\text{reproductibilité}})^2 = (\sigma_{\text{interlaboratoire}})^2 + (\sigma_{\text{répétabilité}})^2$$



**Mais quelle est la valeur « vraie » du dosage?**

- ✓ La valeur vraie n'est connue que si l'on dispose d'un matériau de référence ( MR). C'est à dire un échantillon dont la valeur de la mesure est parfaitement connue
- ✓ Les matériaux de référence peuvent être certifiés ( MRC). C'est à dire accompagné d'un certificat garantissant la valeur réelle:  $\mu$ . Ce sont des étalons garantis.
- ✓ La méthode de dosage peut alors donner un résultat différent de  $\mu$ , avec une déviation  $\delta$  ( biais de la méthode).

**Si on ne dispose de MRC, on admet que la moyenne des moyennes donnent la valeur vraie: m est confondue avec  $\mu$**

# Remarque importante

- Avant tout calcul statistique, il est nécessaire **d'éliminer les valeurs aberrantes**, sinon la moyenne, l'écart type n'ont plus de sens...
- Pour éliminer les aberrants, on réalise des **tests mathématiques** ( que nous n'aborderons pas dans ce cours).

**Test du  $\chi^2$**  : pour comparer les valeurs  $x_i$  intra-laboratoire et rejeter les aberrants avant de calculer moyenne et écart type.

**Test de Cochran**: pour comparer les écart types  $s_{r_i}$  et rejeter les aberrants avant de calculer  $s_r$

**Test de Grubbs**: pour comparer les moyennes  $m_i$  et rejeter les aberrants avant de calculer  $m$

**Observez les résultats avant de vous lancer dans les calculs**

# *Chapitre III*

## **Introduction à la pratique des Plans d'expériences**

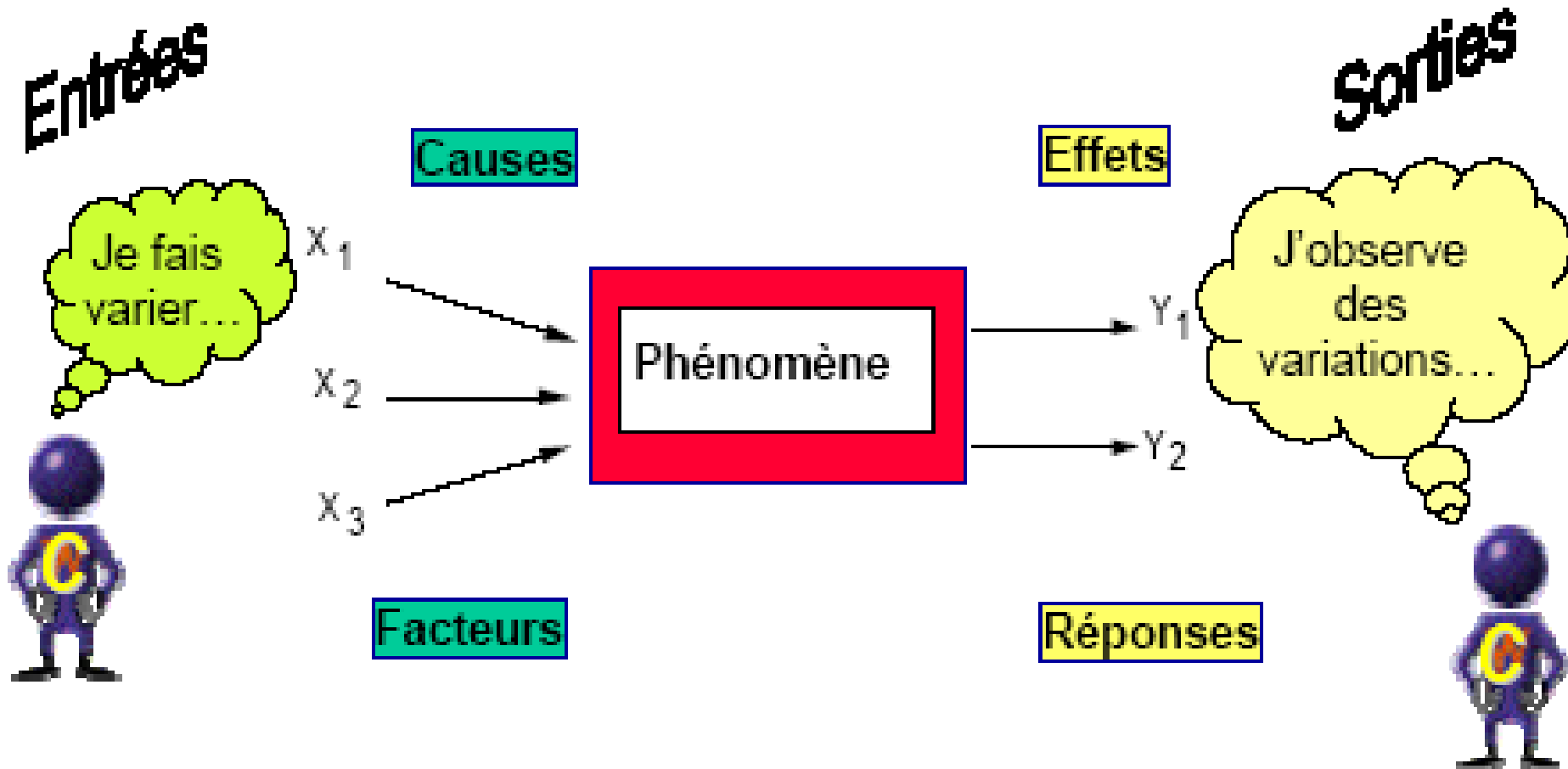
# LES PLANS D'EXPERIENCES

Outils **indispensables** pour le développement des Méthodes analytiques :

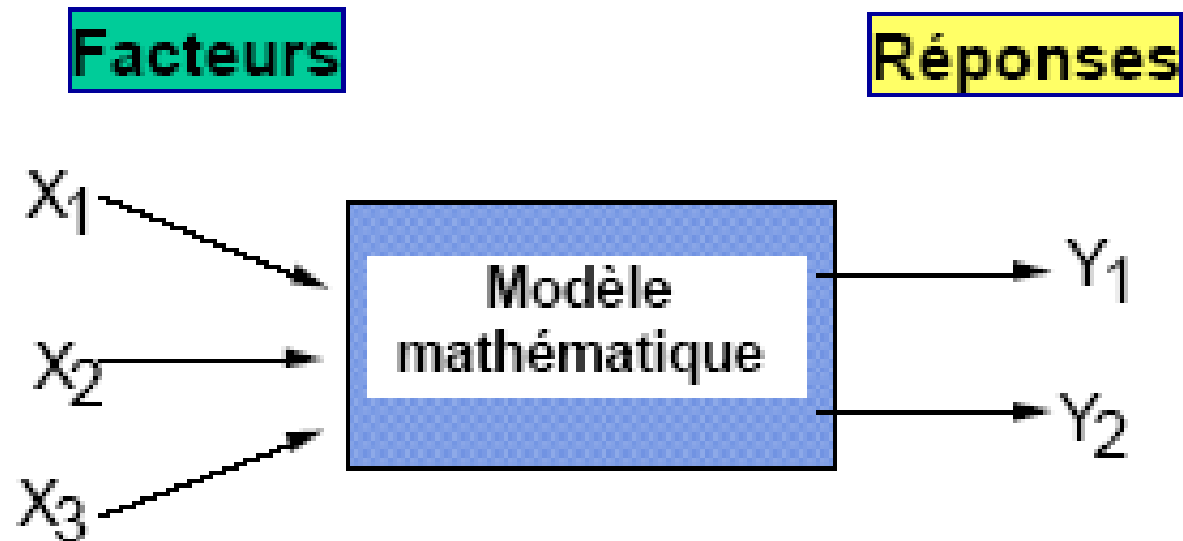
- au niveau du **procédé de préparation** de l'échantillon analytique
- au niveau de **l'optimisation** de la méthode

D'une manière générale : **à mettre en œuvre** au niveau de **toutes les étapes expérimentales de développement** et de **validation** de la méthode.

# Expériences



# Plans d 'expérience



Variation  
Raisonnée  
des Facteurs

Calcul des Effets  
sur les réponses



# *Rappel de quelques définitions concernant Les plans d'expériences*

## ● Matrice d'expérience :

Matrice d'expérience est un tableau de  $n$  lignes et  $k$  colonnes, regroupant les conditions expérimentales d'un plan d'expériences.  $n$  et  $k$  correspondent respectivement au nombre d'expériences et au nombre des variables codées.

## ● Variables explicatives et notion d'interaction :

Les variables explicatives d'une étude sont les paramètres susceptibles de modifier les réponses de cette étude. Si l'effet d'une variable explicative dépend du niveau d'une autre variable explicative, on dit qu'il y a interaction entre ces deux variables explicatives.

## ● **Niveaux d'une variable explicative :**

Les niveaux d'une variable explicative sont les différents états que peut prendre cette variable explicative

## ● **Notion d'effet significatif :**

L'effet d'une variable explicative sur la réponse  $y$  s'obtient en comparant les deux résultats de mesure  $y_1$  et  $y_2$  de réponse, mesurée lorsque la variable explicative passe d'un niveau (0) à un niveau (+). Si l'écart entre  $y_1$  et  $y_2$  est important on dit que le facteur est influent ou significatif.

## ● **Variables codées et variables naturelles :**

Les variables naturelles  $x_i$  sont les valeurs qui correspondent à chaque niveau d'une variable explicative. Pour comparer les effets des variables naturelles sur la réponse, il est nécessaire de les remplacer par les variables codées  $X_i$  qui sont sans unité.

## ● Réponse:

Le résultat mesuré d'une étude. A chaque point du domaine d'étude correspond une réponse. L'ensemble des réponses forme la surface de réponse.

## ● Courbes d'isoréponses :

Après la détermination du modèle et la vérification de sa validité, les courbes d'isoréponses peuvent être tracé à l'intérieur du domaine expérimental. Ces courbes représentent des plans pour surfaces de réponse c'est à dire la représentation graphique des résultats (modèle estimé) pour pouvoir en tirer des optimums.

# Variance-covariance

$$\hat{\sigma}_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Variance de **x**

$$\hat{\sigma}_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

Variance de **y**

$$\hat{\sigma}_{x,y}^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})$$

Covariance **xy**

# ***III.1/ MATRICES D'EXPERIENCES et METHODOLOGIE***

## ***Les objectifs :***

- **Le criblage des facteurs** : classement hiérarchisé des facteurs;
- **Les études quantitatives des facteurs** : quantification des influences principales et des synergies éventuelles;
- **Les études quantitatives des réponses** : modélisation prévisionnelle du phénomène étudié;
- **L'optimisation** : déterminer un ou plusieurs points de fonctionnement optimaux .

### III .1.1/ Les Modèles : polynômes

Soit X un facteur quantitatif. Il peut être représenté par un polynôme dans les modèles :

**Premier degré**

$$Y \text{ (réponse)} = b_0 + b_1 X_1 + b_2 X_2$$

Effet principal

**Synergique**

$$Y \text{ (réponse)} = b_0 + b_1 X_1 + b_2 X_2 + b_{12} X_1 X_2$$

Effet d'interaction

**Quadratique**


$$Y \text{ (réponse)} = b_0 + b_1 X_1 + b_2 X_2 + b_{12} X_1 X_2 + b_{11} X_1^2 + b_{22} X_2^2$$


Surface de réponse


« Courbure »

# III .1.2/ Les Matrices

## MATRICES

Premier degré :  $Y (\text{réponse}) = b_0 + b_1 X_1 + b_2 X_2$   Hadamard (Plackett-Burman)

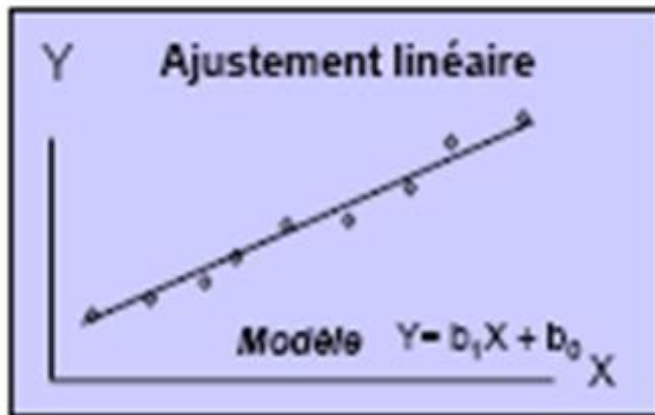
Synergique :  $Y (\text{réponse}) = b_0 + b_1 X_1 + b_2 X_2 + b_{12} X_1 X_2$   Factorielles

Quadratique :  $Y (\text{réponse}) = b_0 + b_1 X_1 + b_2 X_2 + b_{12} X_1 X_2 + b_{11} X_1^2 + b_{22} X_2^2$   Composite (Doehlert)

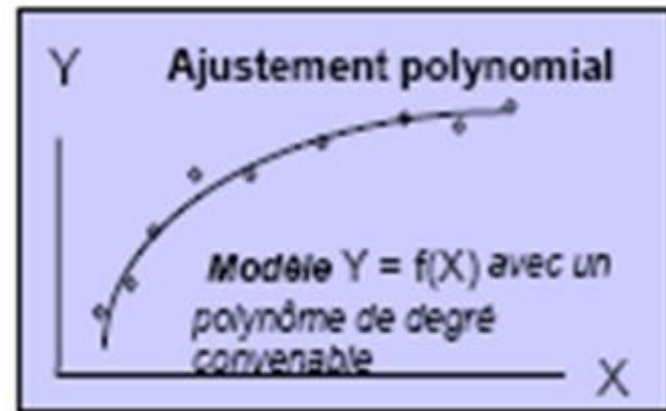
# III. 2/ MODELISATION

**Modéliser** : utiliser des données expérimentales pour **prévoir une information quantitative** inconnue Y à partir de mesures de X via une certaine « **fonction mathématique** » :

Le modèle mathématique postulé peut être :



Une droite si Y varie linéairement avec X.



Sinon un polynôme de degré convenable.



**Exemple** : Etude de la stabilité d'une suspension.

➤ Chercher un modèle liant les facteurs expérimentaux.

Facteurs expérimentaux	Variables	Domaine Expérimental	
		Minimum	Maximum
Tensio-actif : Mouillant 1	$M_1$	20 g/l	40 g/l
Tensio-actif : Mouillant 2	$M_2$	5 "	15 "
Epaississant : Structurant 1	$S_1$	5 "	20 "
Epaississant : Structurant 2	$S_2$	0 "	10 "

Réponse étudiée :

**Y = % de séparation de la suspension en deux phases**

# Tableau des résultats

Plan d'expérimentation

Réponse

Essais	M <sub>1</sub>	M <sub>2</sub>	S <sub>1</sub>	S <sub>2</sub>	Y
1	40	15	20	0	10
2	40	15	5	10	16
3	40	5	20	0	15
4	20	15	5	0	38,7
5	40	5	5	10	30,5
6	20	5	20	10	18
7	20	15	20	10	13
8	20	5	5	0	32

Modèle postulé

$$Y = a_0 + a_1 M_1 + a_2 M_2 + a_3 S_1 + a_4 S_2$$

## III. 2. 1/ Codage des variables :

- Variables naturelles :  $U$
- Variables codées :  $X$

### $U_i$

Toutes les variables naturelles ont leurs propres plage de variation qui dépendent des unités

Les coefficients du modèle **ne peuvent pas** se comparer directement

**Interprétation difficile**

### $X_i$

Toutes les variables codées ont la même plage de variation entre -1 et +1 indépendante des unités

Les coefficients du modèle **peuvent** se comparer directement

**Interprétation facile**

= **moyenne** des valeurs maximum et minimum que peut prendre la variable  $U_j$ .

$U_{ij}$  = valeur de la variable naturelle  $j$  à l'expérience  $i$ .

$$X_{ij} = \frac{U_{ij} - U_j^0}{\Delta U_j}$$

$U_j^0$  = valeur de la variable naturelle  $j$  au **centre du domaine**.

$\Delta U_j$  = **pas de variation** de la variable naturelle  $j$ .

$X_{ij}$  = valeur de la variable codée  $j$  pour l'expérience  $i$ .

= **demi étendue** : moitié de l'écart entre la valeur maximum et la valeur minimum que peut prendre la variable  $U_j$ .

Réciproquement :

$$U_{ij} = U_j^0 + X_{ij} \cdot \Delta U_j$$

# Valeurs codées

$U_1$  est le **mouillant 1** qui varie entre  
20 et 40 g/l

centre du domaine  $U_1$   
 $= (20+40)/2 = 30 \text{ g/l}$

Pas de variation  $\Delta U_1$   
 $= (40-20)/2 = 10 \text{ g/l}$

à 20 g/l :  $X_{i1} = (20 - 30)/10 = -1$   
à 40 g/l :  $X_{i1} = (40 - 30)/10 = 1$

$U_4$  est le **structurant 2** qui varie entre  
0 et 10 g/l

centre du domaine  $U_4$   
 $= (0+10)/2 = 5 \text{ g/l}$

Pas de variation  $\Delta U_4$   
 $= (10-0)/2 = 5 \text{ g/l}$

à 0 g/l :  $X_{i4} = (0-5)/5 = -1$   
à 10 g/l :  $X_{i4} = (10-5)/5 = 1$

# Régression avec des variables codées

Matrice d'expériences

Essais	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>
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

Modèle postulé

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4$$

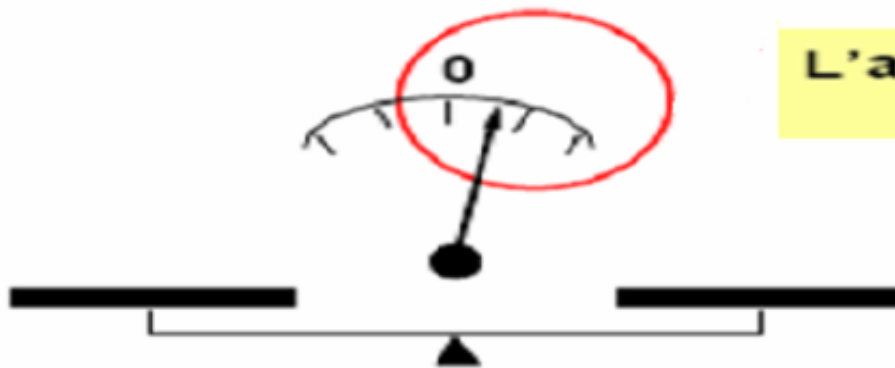
Comment déterminer  $b_0$ ,  $b_1$ ,  $b_2$ ,  $b_3$  et  $b_4$ ?

# *Chapitre IV*

# Plans de pesée

## Expérience

### Balance



L'aiguille **n'est pas** sur le zéro

Nous convenons d'appeler "expérience" **le fait de lire la position de l'aiguille** sur le cadran.

Comme pour toute expérience, **il existe une erreur expérimentale** :



erreur de parallaxe, épaisseurs relatives de la pointe de l'aiguille et des traits de graduation etc.



# Efficacité

Etre efficace c'est rendre minimum :

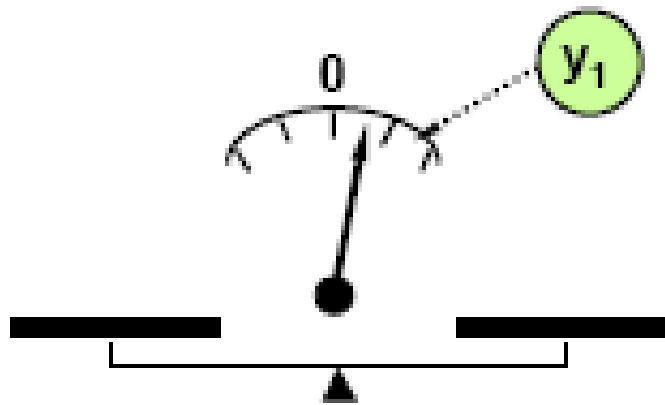
non pas l'erreur de lecture

MAIS

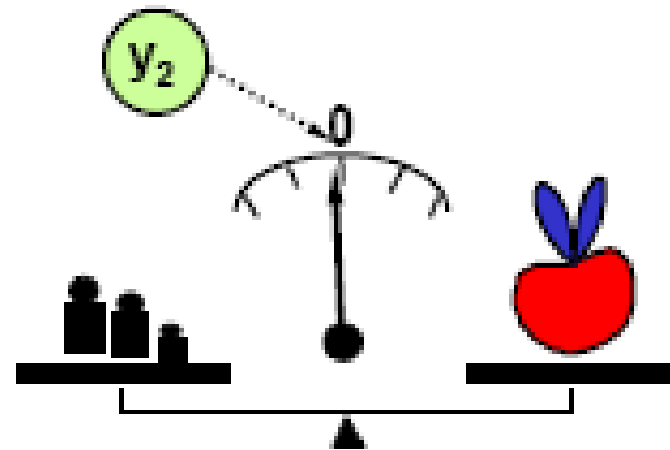
sa répercussion sur l'estimation des masses,

nous admettrons que l'expérimentateur prend toutes les précautions utiles pour la rendre minimale

(sur le calcul). En d'autres termes c'est cerner au plus près la masse vraie des objets à peser.



Lecture à vide



Lecture avec la pomme

$$\text{Masse de la pomme} = y_2 - y_1$$

Chaque pesée, matérialisée par la lecture  $Y_i$  (réponse), est :

indépendante des autres lectures

une variable aléatoire car  $Y_i$  est une mesure expérimentale.

$Y_i$  est en réalité la somme de deux valeurs :

$$Y_i = \eta_i + e_i$$

$\eta_i$  = valeur vraie

$e_i$  = erreur de mesure aléatoire

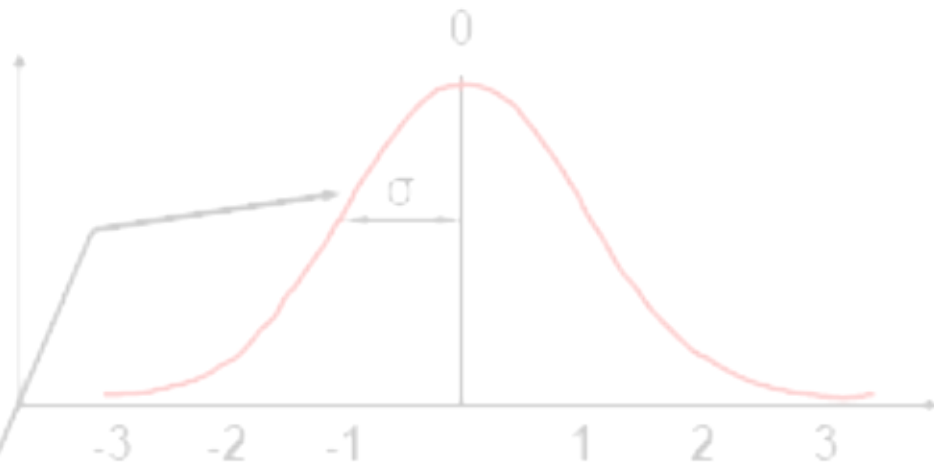
# Caractéristiques de l'erreur expérimentale $r_i$

Distribution de Gauss centrée sur zéro

En **moyenne**,  
l'erreur **est nulle**

Sa **dispersion** est  
mesurée par sa  
variance :  $\text{var}(r_i) = \sigma^2$

ou par l'écart-type  $\sigma$ .



# Qualité de l'estimation de la masse de la pomme

l'estimation de la Masse de la pomme  $m = y_2 - y_1$

$$\begin{aligned}\text{var}(m) &= \text{var}(Y_2 - Y_1) = \text{var}(Y_2) + \text{var}(Y_1) \\ &= \text{var}(\eta_2 + e_2) + \text{var}(\eta_1 + e_1) \\ &= \text{var}(\eta_2) + \text{var}(e_2) + \text{var}(\eta_1) + \text{var}(e_1)\end{aligned}$$

$= 0$

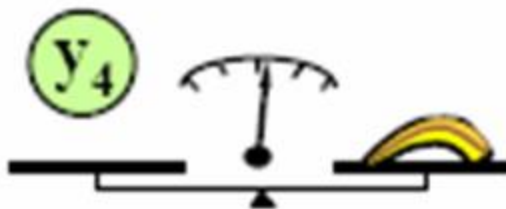
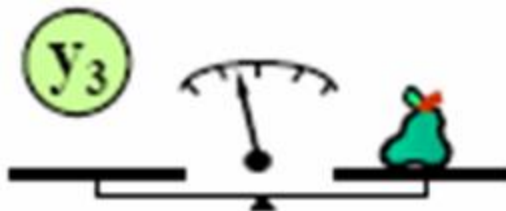
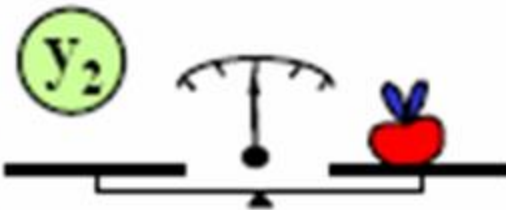
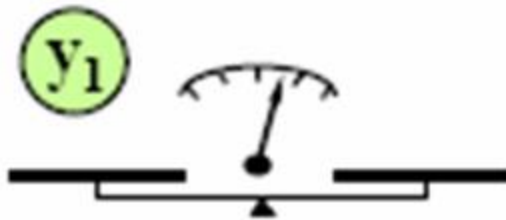
$= \sigma^2$   
(variance de  $e_1$ )

$$\text{var}(\hat{m}) = 2\sigma^2$$

# Pesée de plusieurs objets

- **Si nous avons trois objets, comment les peser ?**
- **Quel est le prix minimum de l'expérimentation?**

# Expérimentateur n°1



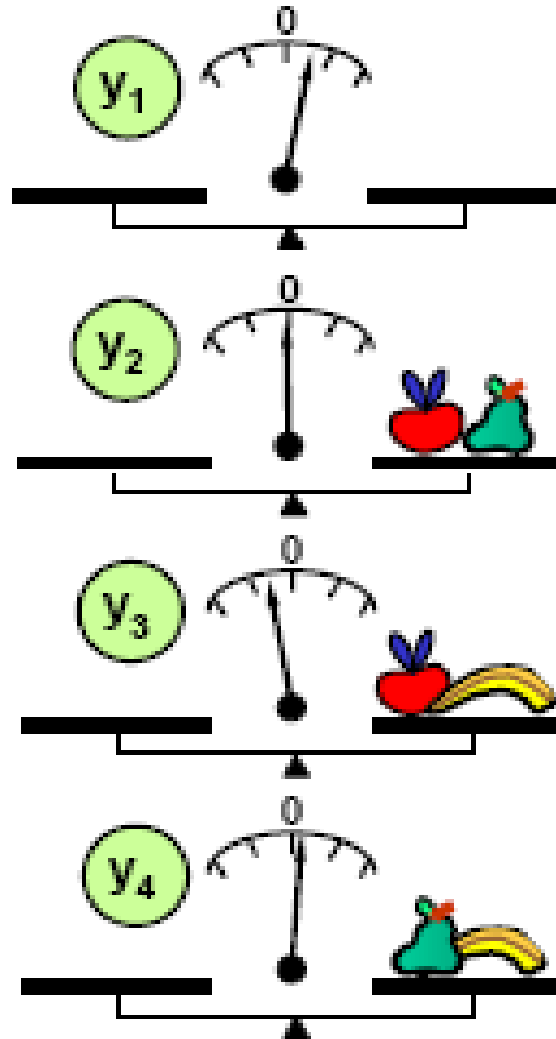
Lectures

Masse d'un objet :  $\hat{m}_i = y_{i+1} - y_i$

$$\text{var}(\hat{m}_i) = \text{var}(Y_{i+1}) + \text{var}(Y_i)$$

$$\text{var}(\hat{m}_i) = 2\sigma^2$$

# Expérimentateur n°2



Lectures

Masse d'un objet ?

Difficile à lire directement

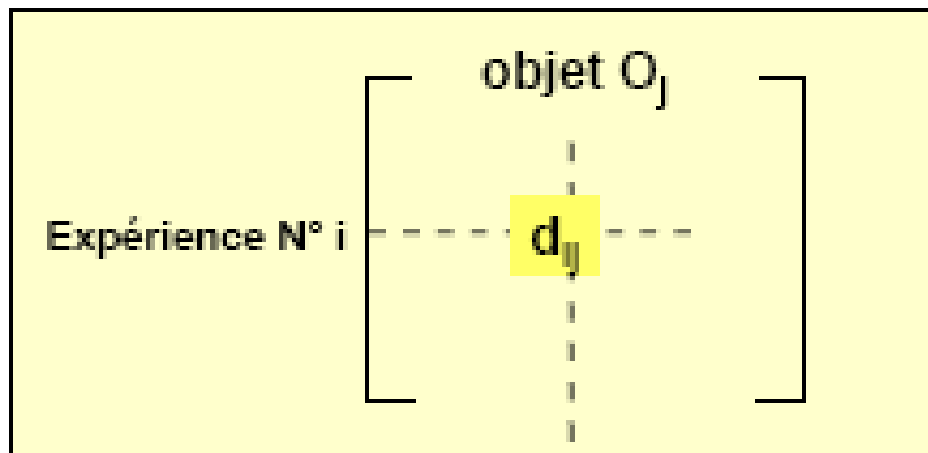


Introduction de la notion de matrice d'expériences



# Matrice d'expériences

Une matrice d'expériences est un **tableau** permettant de **décrire une expérimentation** en donnant, pour chaque expérience, les valeurs des facteurs expérimentaux.



**d<sub>ij</sub>** caractérise l'état de l'**objet j**  
au cours de l'**expérience i**

# Expérimentateurs n°1 & n°2

**Codage** des objets : objet **absent** : état **0**  
objet **présent** : état **1**

Exp.	O <sub>1</sub>	O <sub>2</sub>	O <sub>3</sub>
1	0	0	0
2	1	0	0
3	0	1	0
4	0	0	1

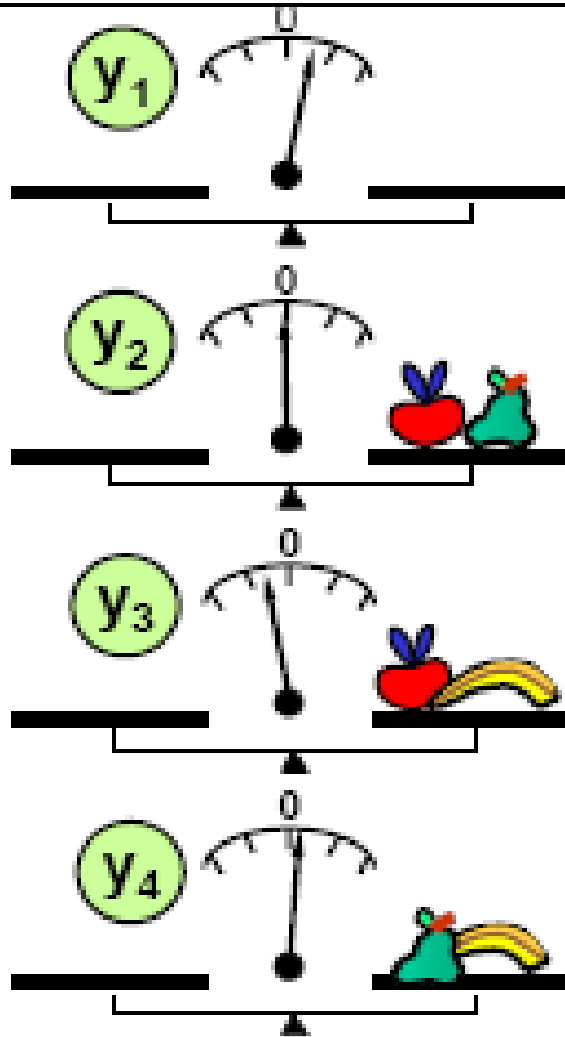
Plan N°1 (D<sub>1</sub>)

Exp.	O <sub>1</sub>	O <sub>2</sub>	O <sub>3</sub>
1	0	0	0
2	1	1	0
3	1	0	1
4	0	1	1

Plan N°2 (D<sub>2</sub>)

# Expérimentateur n° 2

Lectures



## Système d'équations

$$\begin{aligned}y_1 &= \eta_0 + e_1 = \widehat{m}_0 \\y_2 &= \eta_1 + \eta_2 + e_2 = \widehat{m}_0 + \widehat{m}_1 + \widehat{m}_2 \\y_3 &= \eta_1 + \eta_3 + e_3 = \widehat{m}_0 + \widehat{m}_1 + \widehat{m}_3 \\y_4 &= \eta_2 + \eta_3 + e_4 = \widehat{m}_0 + \widehat{m}_2 + \widehat{m}_3\end{aligned}$$

# Matrice d'expériences et système d'équations

Exp.	O <sub>1</sub>	O <sub>2</sub>	O <sub>3</sub>
1	0	0	0
2	1	1	0
3	1	0	1
4	0	1	1

$$y_1 = \widehat{m}_0 + 0 \widehat{m}_1 + 0 \widehat{m}_2 + 0 \widehat{m}_3$$

$$y_2 = \widehat{m}_0 + 1 \widehat{m}_1 + 1 \widehat{m}_2 + 0 \widehat{m}_3$$

$$y_3 = \widehat{m}_0 + 1 \widehat{m}_1 + 0 \widehat{m}_2 + 1 \widehat{m}_3$$

$$y_4 = \widehat{m}_0 + 0 \widehat{m}_1 + 1 \widehat{m}_2 + 1 \widehat{m}_3$$

Le **codage** utilisé pour écrire la matrice d'expériences permet de **faire correspondre**



les éléments de la matrice aux **coefficients des inconnues** du système d'équations



Inconnues = **masses à estimer**

# Matrice du modèle et système d'équations

$$Y = m_0 + m_1 X_1 + m_2 X_2 + m_3 X_3$$

Avec la **notation matricielle**, on peut écrire de manière compacte le **système** de 4 équations linéaires qui lient les réponses observées aux états des objets :

$$Y = Xm$$

Vecteur réponse

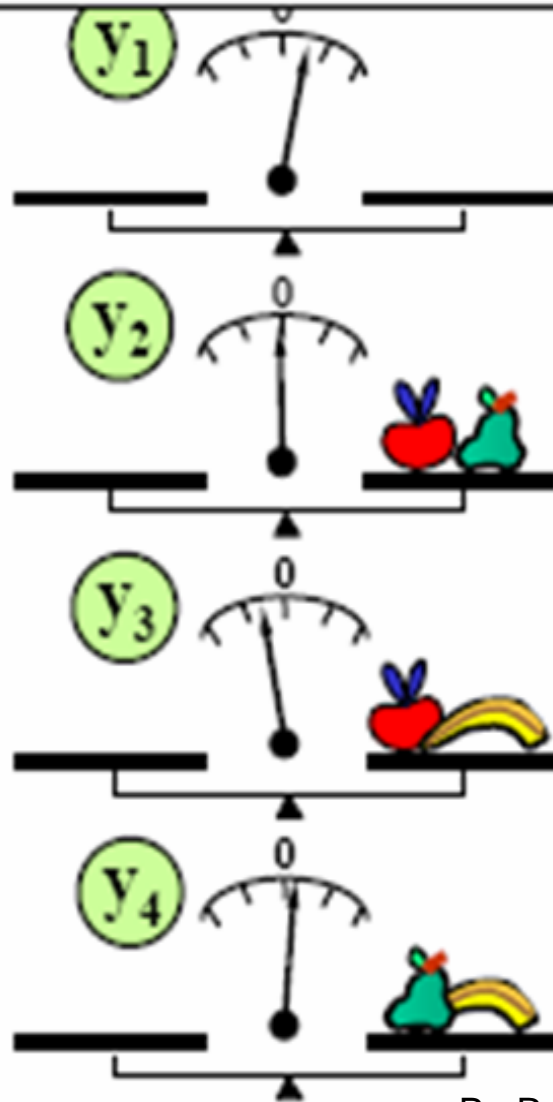
$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix} * \begin{bmatrix} m_0 \\ m_1 \\ m_2 \\ m_3 \end{bmatrix}$$

Matrice du modèle

Vecteur des coefficients

# Expérimentateur n°2

Lectures



Résolution

$$\hat{m}_0 = y_1$$

$$\hat{m}_1 = (-y_1 + y_2 + y_3 - y_4) / 2$$

$$\hat{m}_2 = (-y_1 + y_2 - y_3 + y_4) / 2$$

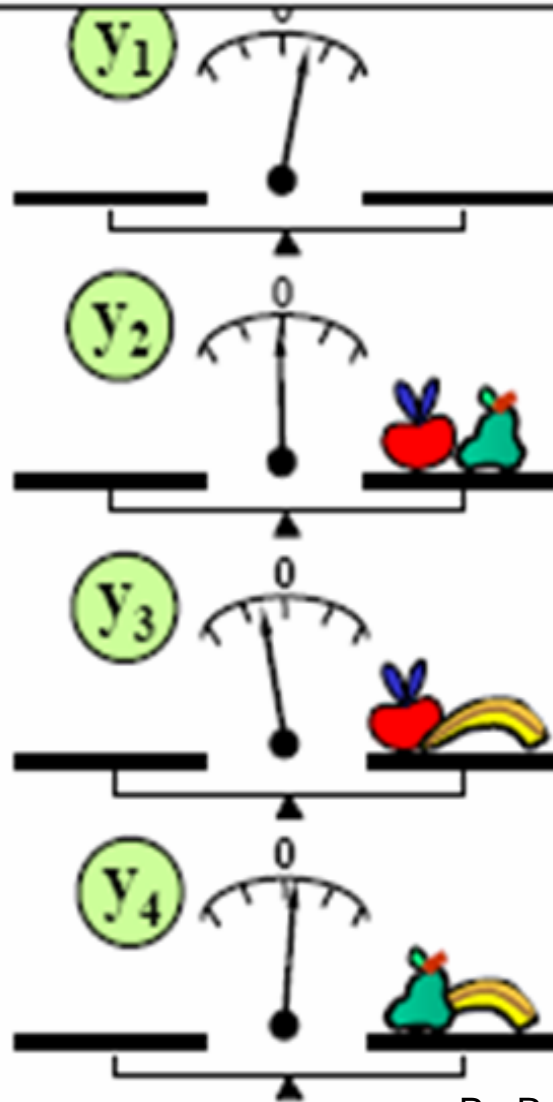
$$\hat{m}_3 = (-y_1 - y_2 + y_3 + y_4) / 2$$

$$\begin{aligned} \text{var}(\hat{m}_i) &= \text{var}[1/2(\Sigma_{\text{algèbr. de 4 } Y_i})] \\ &= 1/4 [\text{var}(\Sigma_{\text{algèbr. de 4 } Y_i})] \end{aligned}$$

$$\text{var}(\hat{m}_i) = \sigma^2$$

# Expérimentateur n°2

Lectures



Résolution

$$\hat{m}_0 = y_1$$

$$\hat{m}_1 = (-y_1 + y_2 + y_3 - y_4) / 2$$

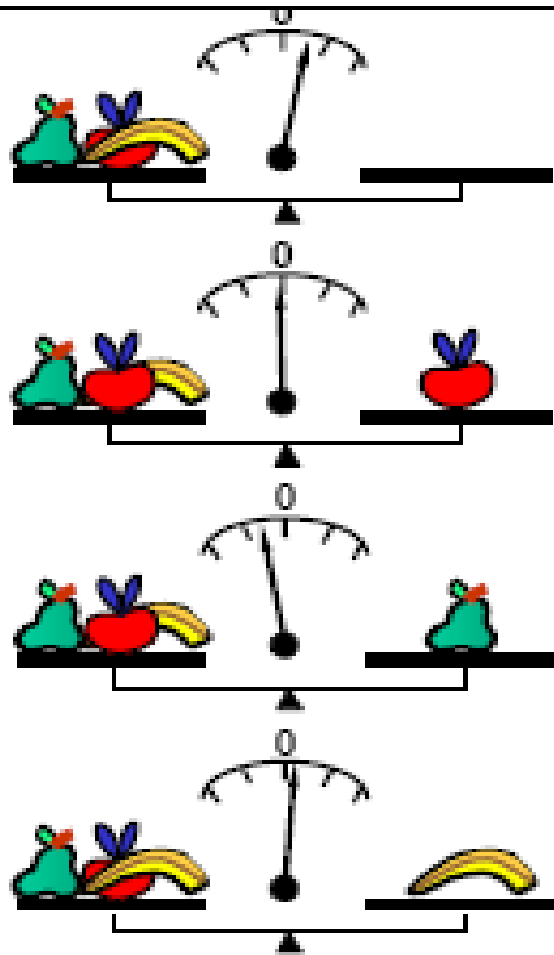
$$\hat{m}_2 = (-y_1 + y_2 - y_3 + y_4) / 2$$

$$\hat{m}_3 = (-y_1 - y_2 + y_3 + y_4) / 2$$

$$\begin{aligned} \text{var}(\hat{m}_i) &= \text{var}[1/2(\Sigma_{\text{algèbr. de 4 } Y_i})] \\ &= 1/4 [\text{var}(\Sigma_{\text{algèbr. de 4 } Y_i})] \end{aligned}$$

$$\text{var}(\hat{m}_i) = \sigma^2$$

# Expérimentateur n°3



**Codage** des objets :

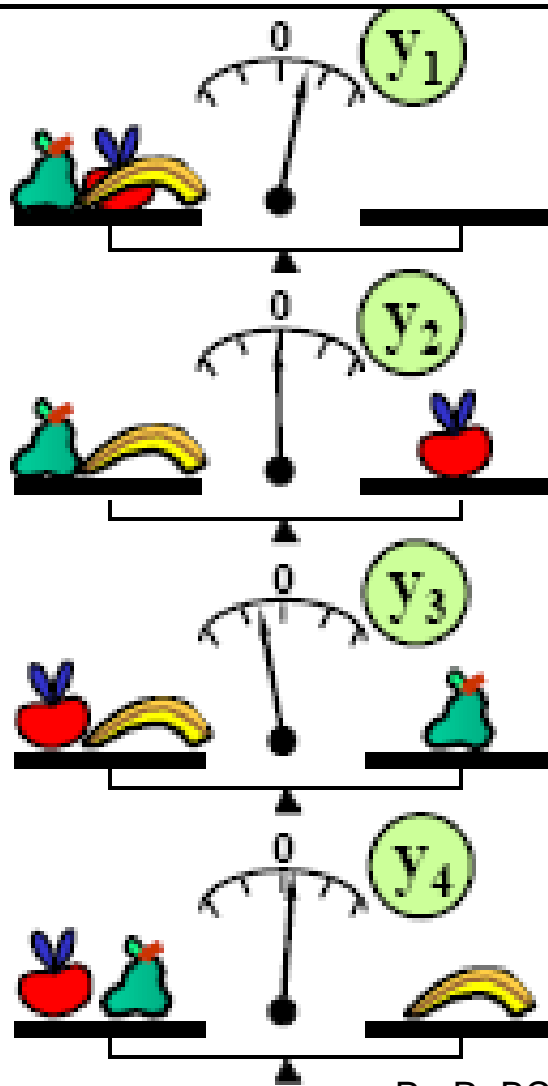
objet **absent** : état **0**  
objet présent **à droite** : état **1**  
objet présent **à gauche** : état **-1**

Exp.	O <sub>1</sub>	O <sub>2</sub>	O <sub>3</sub>
1	-1	-1	-1
2	1	-1	-1
3	-1	1	-1
4	-1	-1	1

Plan N°3 (D<sub>3</sub>)



# Expérimentateur n°3



## Systeme d'equations

$$y_1 = \hat{m}_0 - \hat{m}_1 - \hat{m}_2 - \hat{m}_3$$

$$y_2 = \hat{m}_0 + \hat{m}_1 - \hat{m}_2 - \hat{m}_3$$

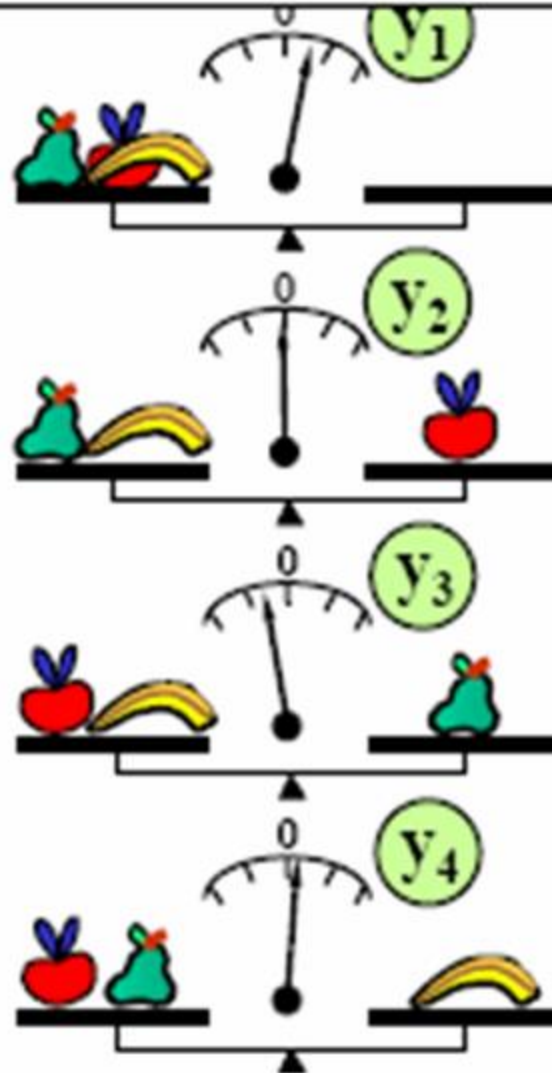
$$y_3 = \hat{m}_0 - \hat{m}_1 + \hat{m}_2 - \hat{m}_3$$

$$y_4 = \hat{m}_0 - \hat{m}_1 - \hat{m}_2 + \hat{m}_3$$

Lectures

# Expérimentateur n°3

Lectures



Résolution

$$\hat{m}_0 = (-y_1 + y_2 + y_3 + y_4) / 2$$

$$\hat{m}_1 = (y_2 - y_1) / 2$$

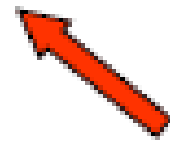
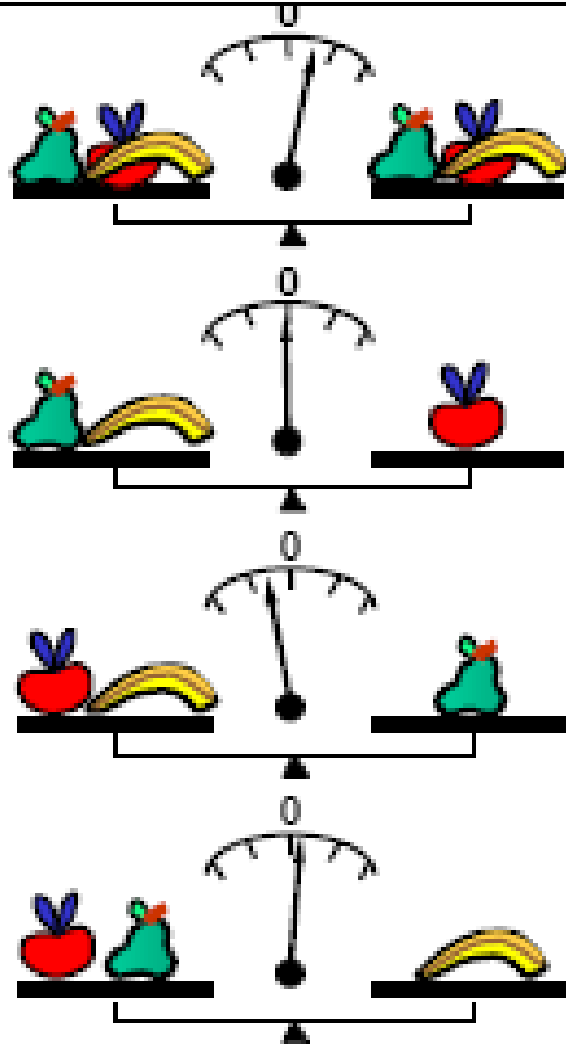
$$\hat{m}_2 = (y_3 - y_1) / 2$$

$$\hat{m}_3 = (y_4 - y_1) / 2$$

$$\begin{aligned} \text{var}(\hat{m}_i) &= \text{var}[1/2(\Sigma_{\text{algèbr. de 2 } Y_i})] \\ &= 1/4 [\text{var}(\Sigma_{\text{algèbr. de 2 } Y_i})] \end{aligned}$$

$$\text{var}(\hat{m}_i) = \sigma^2/2$$

# Expérimentateur n°4



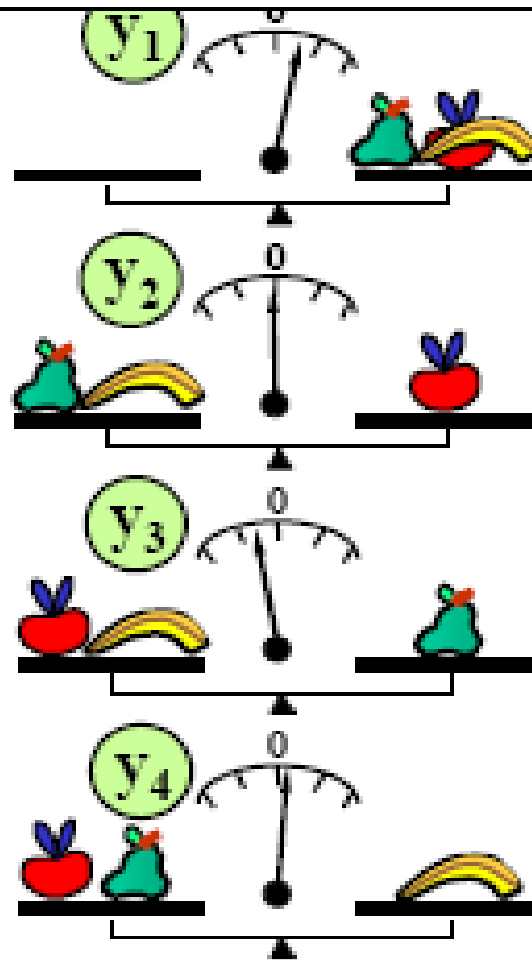
Seule l'expérience 1 change

Exp.	O <sub>1</sub>	O <sub>2</sub>	O <sub>3</sub>
1	1	1	1
2	1	-1	-1
3	-1	1	-1
4	-1	-1	1

Plan N°4 (D<sub>4</sub>)

# Expérimentateur n°4

Lectures



## Système d'équations

$$y_1 = \widehat{m}_0 + \widehat{m}_1 + \widehat{m}_2 + \widehat{m}_3$$

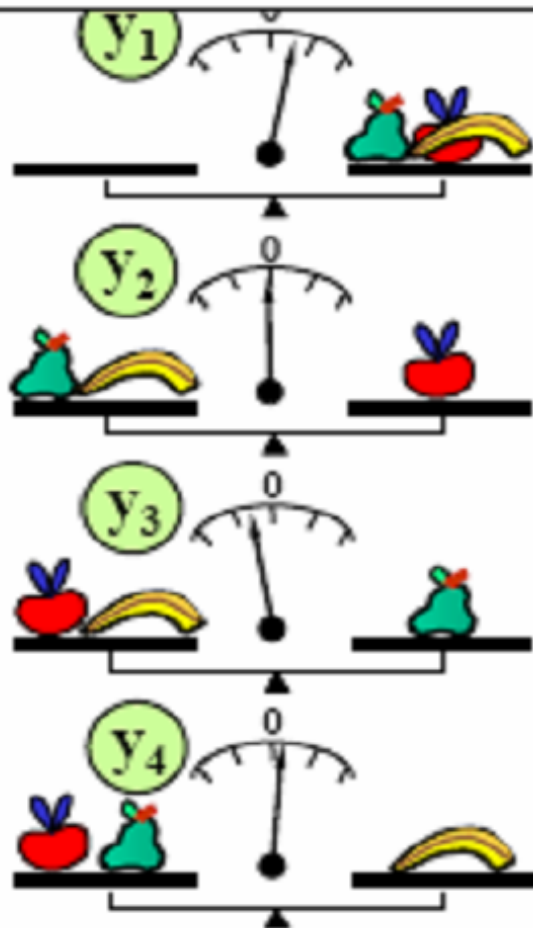
$$y_2 = \widehat{m}_0 + \widehat{m}_1 - \widehat{m}_2 - \widehat{m}_3$$

$$y_3 = \widehat{m}_0 - \widehat{m}_1 + \widehat{m}_2 - \widehat{m}_3$$

$$y_4 = \widehat{m}_0 - \widehat{m}_1 - \widehat{m}_2 + \widehat{m}_3$$

# Expérimentateur n°4

Lectures



Résolution

$$\hat{m}_0 = (+y_1 + y_2 + y_3 + y_4) / 4$$

$$\hat{m}_1 = (+y_1 + y_2 - y_3 - y_4) / 4$$

$$\hat{m}_2 = (+y_1 - y_2 + y_3 - y_4) / 4$$

$$\hat{m}_3 = (+y_1 - y_2 - y_3 + y_4) / 4$$

$$\begin{aligned} \text{var}(\hat{m}_i) &= \text{var}[1/4(\Sigma_{\text{algèbr. de 4 } Y_j})] \\ &= 1/16 [\text{var}(\Sigma_{\text{algèbr. de 4 } Y_j})] \end{aligned}$$

$$\text{var}(\hat{m}_i) = \sigma^2/4$$

# Conclusion

Nous n'avons pas eu besoin des valeurs des  $y_i$  pour prévoir la qualité des estimations.

La qualité de l'information expérimentale ne dépend que du choix des essais (de la matrice d'expériences).

Cette réflexion préalable peut être généralisée à toute expérimentation : c'est avant d'expérimenter qu'il faut s'interroger sur la qualité de l'expérimentation projetée.

## THÉORÈME

Si on fait **N pesées**, la **variance** sur les coefficients est  $\geq \sigma^2/N$

Comment obtenir une matrice optimale pour minimiser la variance sur les masses estimées?

# Modélisation

La relation entre la mesure  $Y$  (la lecture du cadran) et les différentes masses dont on cherche à déterminer la valeur est une relation linéaire qui s'écrit :

$$y_i = m_0 + m_1 X_{1i} + m_2 X_{2i} + m_3 X_{3i} + e_i$$

masses des objets

$m_0, m_1, m_2$  et  $m_3$

état des masses dans  
l'expérience  $j$  (la pesée)

$X_1, X_2$  et  $X_3$  présence ou absence  
sur les plateaux (-1, 1 ou 0).

erreur



# Matrices d 'Expériences

Exp.	O <sub>1</sub>	O <sub>2</sub>	O <sub>3</sub>
1	0	0	0
2	1	0	0
3	0	1	0
4	0	0	1

Plan N°1 (D<sub>1</sub>)

Exp.	O <sub>1</sub>	O <sub>2</sub>	O <sub>3</sub>
1	0	0	0
2	1	1	0
3	1	0	1
4	0	1	1

Plan N°2 (D<sub>2</sub>)

Exp.	O <sub>1</sub>	O <sub>2</sub>	O <sub>3</sub>
1	-1	-1	-1
2	1	-1	-1
3	-1	1	-1
4	-1	-1	1

Plan N°3 (D<sub>3</sub>)

Exp.	O <sub>1</sub>	O <sub>2</sub>	O <sub>3</sub>
1	1	1	1
2	1	-1	-1
3	-1	1	-1
4	-1	-1	1

Plan N°4 (D<sub>4</sub>)

# Expression matricielle de la régression

$$Y = m_0 + m_1 X_1 + m_2 X_2 + m_3 X_3 + e$$

Avec la **notation matricielle**, on peut écrire de manière compacte le **système** de 4 équations linéaires qui lient les réponses observées aux états des objets :

$$Y = X m + e$$

Vecteur réponse

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix} * \begin{bmatrix} m_0 \\ m_1 \\ m_2 \\ m_3 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix}$$

Vecteur  
des erreurs

Matrice du modèle

Vecteur des coefficients

# Expression matricielle de la régression

La matrice  $X$  représente la **matrice des coefficients des inconnues  $m_j$**  du système d'équations.

Calcul du vecteur des coefficients

$X'$  étant la **matrice transposée** de  $X$ , si  $X'X$  n'est pas singulière (c'est à dire s'il est possible de calculer la matrice inverse correspondante) :

on aura l'**estimation** du vecteur des coefficients du modèle au sens des moindres carrés par :

$$m = (X'X)^{-1}X'Y$$

# Qualité des Matrices d'Expériences

Exp.	O <sub>1</sub>	O <sub>2</sub>	O <sub>3</sub>
1	0	0	0
2	1	0	0
3	0	1	0
4	0	0	1

Matrice D

Exp.	O <sub>0</sub>	O <sub>1</sub>	O <sub>2</sub>	O <sub>3</sub>
1	1	0	0	0
2	1	1	0	0
3	1	0	1	0
4	1	0	0	1

Matrice X du modèle :

$$Y = m_0 + \sum m_i X_i$$

masse de l'objet i

état de l'objet i  
dans la pesée considérée

Matrice de **variance-covariance** :  $(X'X)^{-1}$

# Analyse quantitative et Étalonnage

## Les données sont toujours en nombre limité

➤ Elles ne représentent donc qu'un échantillon de la population de toutes les mesures de la teneur en analyse de l'étalon que l'on pourrait effectuer.

- Si  $X$  représente une teneur connue en analyse

-  $Y$  représente le résultat observé

La relation linéaire postulée devient :  $Y = b_0 + b_1 X$

Avec uniquement une "estimation" des coefficients  $b_0$  et  $b_1$  du modèle postulé.

$$\hat{Y} = \beta_0 + \beta_1 X$$

# Expression matricielle

$$\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \\ \dots & \dots \\ 1 & x_n \end{pmatrix} * \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} + \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ \dots \\ r_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \dots \\ y_n \end{pmatrix}$$

$\hat{y}_1$   
 $\hat{y}_2$   
 $\hat{y}_3$   
 $\hat{y}_4$   
 $\dots$   
 $\hat{y}_n$

Vecteur des résidus :  $r$

Vecteur de la réponse expérimentale

$XB + r = Y$

# Estimation des coefficients :

Les  $\beta_i$  sont les inconnus que nous devons estimer :  
( $b_i$  est l'estimation calculée de  $\beta_i$  )

1. Au sens des moindres carrés (résolution algébrique) :

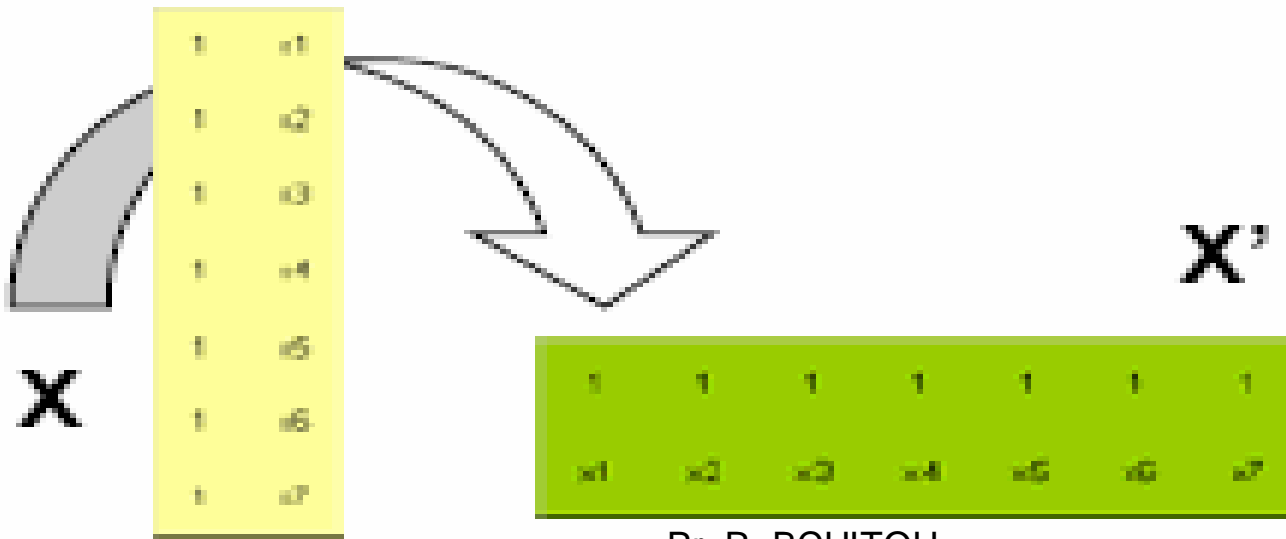
$$\hat{\beta}_1 = b_1 = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \quad \text{et} \quad \hat{\beta}_0 = b_0 = \bar{y} - b_1 \bar{x}$$

## 2. Au sens des moindres carrés (résolution matricielle) :

➤ En notant  $X'$  la matrice transposée de  $X$ , on aura l'estimation du vecteur des coefficients du modèle au sens des moindres carrés par :

$$B = (X'X)^{-1}X'Y$$

Vecteur des coefficients





# *Chapitre V*

## *Matrice d'Hadamard*

## **But :**

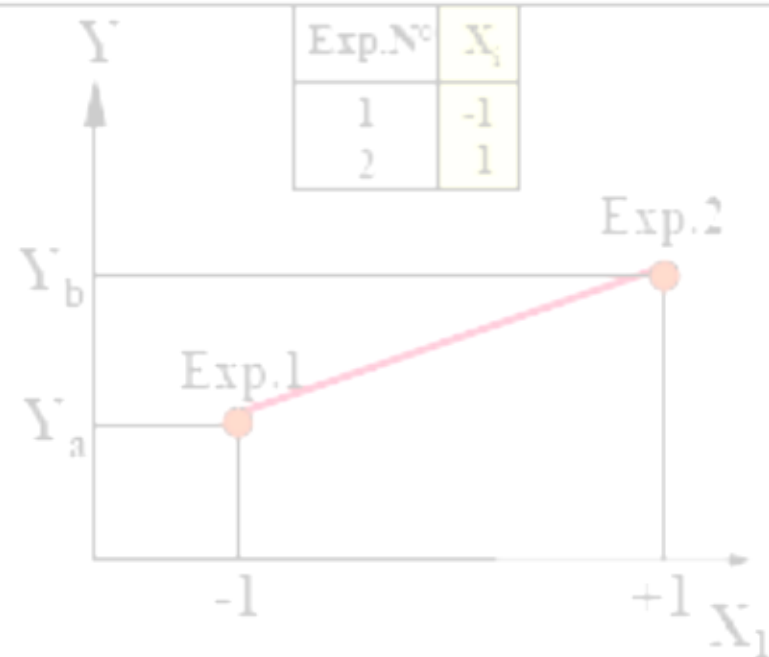
- Le criblage des facteurs : classement hiérarchisé des facteurs**

# Effets des Facteurs

Effet  
d'un facteur

Effet de  $X_1$  :  $E_1$

$$E_1 = Y_b - Y_a$$



**Effet de X :**

(valeur de la réponse  $Y$  au niveau +1)

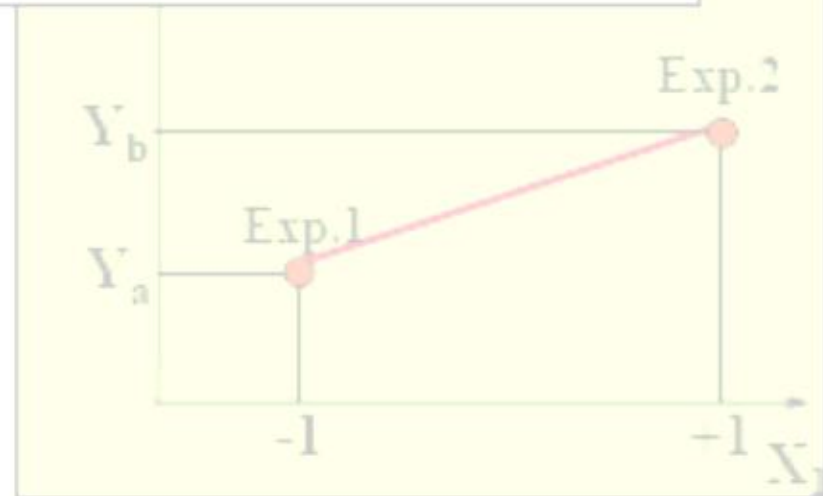
-

(Valeur de la réponse  $Y$  au niveau -1)

# Poids des Facteurs



Exp.N°	$X_1$
1	-1
2	1



Modèle :  $Y = b_0 + b_1 X_1$

$$b_1 = \frac{Y_b - Y_a}{2}$$

« poids du facteur  $X_1$  »  $b_1 = \frac{E_1}{2}$

$$b_0 = \frac{Y_b + Y_a}{2}$$

« valeur moyenne » "réponse théorique"  
au centre du domaine

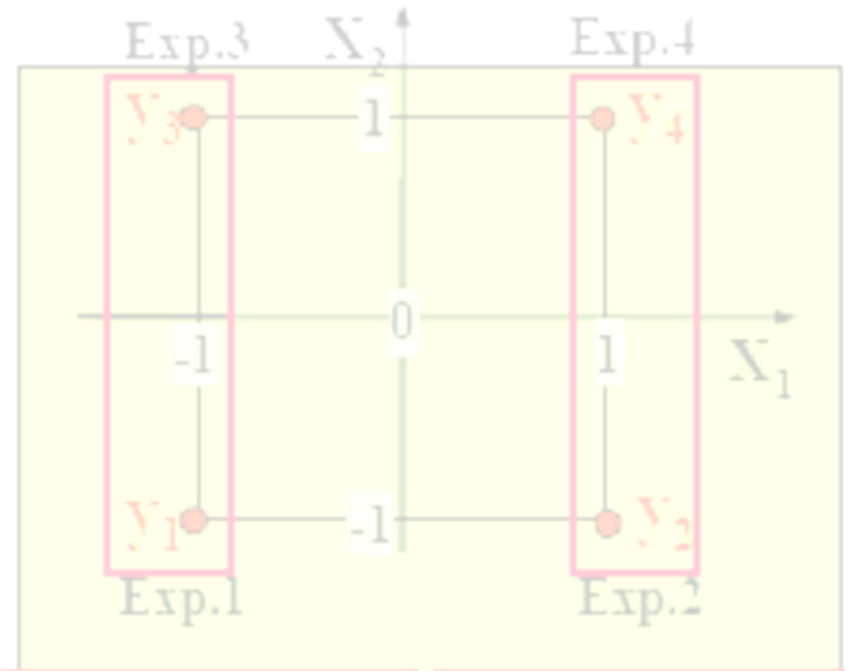
# Effets des Facteurs

## Cas de deux facteurs

Exp.N°	$X_1$	$X_2$
1	-1	-1
2	1	-1
3	-1	1
4	1	1

Pour  $X_1$

$E_1/2 =$  Poids du facteur  $X_1 :$   
 $[(y_2+y_4)/2] - [(y_1+y_3)/2]$



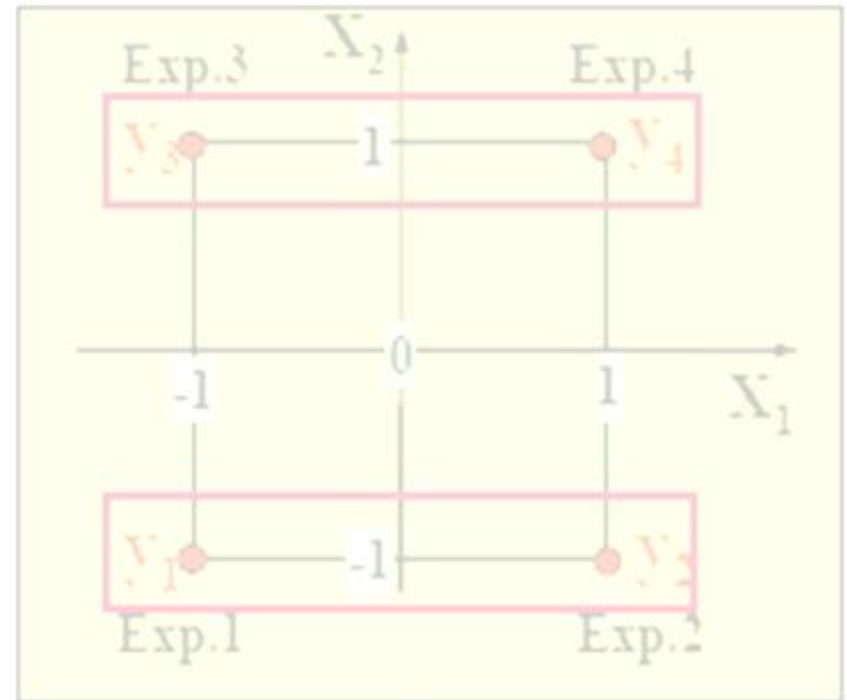
Réponse moyenne pour  $X_1 = -1$   
 $(y_1+y_3)/2$

Réponse moyenne pour  $X_1 = 1$   
 $(y_2+y_4)/2$



poids de  $X_1 :$   $E_1/2 = \frac{-y_1 + y_2 - y_3 + y_4}{2}$

Pour le cas de  $X_2$



$E_2 / 2 =$  Poids du facteur  $X_2$ :

$$[(y_3 + y_4) / 2] - [(y_1 + y_2) / 2]$$

Réponse moyenne  
pour  $X_2 = -1$

$$(y_1 + y_2) / 2$$

Réponse moyenne  
pour  $X_2 = 1$

$$(y_3 + y_4) / 2$$

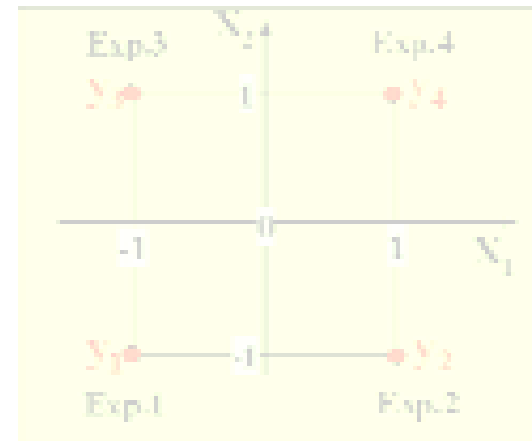
poids de  $X_2$ :

$$E_2 / 2 = \frac{-y_1 - y_2 + y_3 + y_4}{2}$$

# Poids des Facteurs

**Poids pour  
aux facteurs**

Exp.N°	$X_1$	$X_2$
1	-1	-1
2	1	-1
3	-1	1
4	1	1



Modèle :  $Y = b_0 + b_1 X_1 + b_2 X_2$

**Poids de  $X_1$**   $b_1 = \frac{E_1}{2} = \frac{-y_1 + y_2 - y_3 + y_4}{4}$

**Poids de  $X_2$**   $b_2 = \frac{E_2}{2} = \frac{-y_1 - y_2 + y_3 + y_4}{4}$

Valeur moyenne  $b_0 = \frac{+y_1 + y_2 + y_3 + y_4}{4}$

**Poids = Pente = Coefficient**

**Modèle pour un facteur :  $Y = b_0 + b_1X$**

**Modèle pour deux facteurs :  $Y = b_0 + b_1X_1 + b_2X_2$**

**Modèle pour K facteurs :  $Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 \dots + b_kX_k$**

Pour avoir la **variance des estimations minimale**  $\sigma^2/N$ ,  
la **condition nécessaire et suffisante** est que l'on ait :

$$X'X = N I_N$$



# Plans de Pesées et Matrices d 'Hadamard

Cette propriété est obtenue avec les **matrices d'Hadamard**, qui ne contiennent que des valeurs -1 ou +1 .

Elles n'existent que pour **N multiple de 4** :  
4, 8, 12, 16, 20 ...



Jacques Salomon  
HADAMARD  
(1865 - 1963)

## Construction de Plackett et Burman

$k \leq 3$  (N=4) : + + -

$k \leq 7$  (N=8) : + + + - + - -

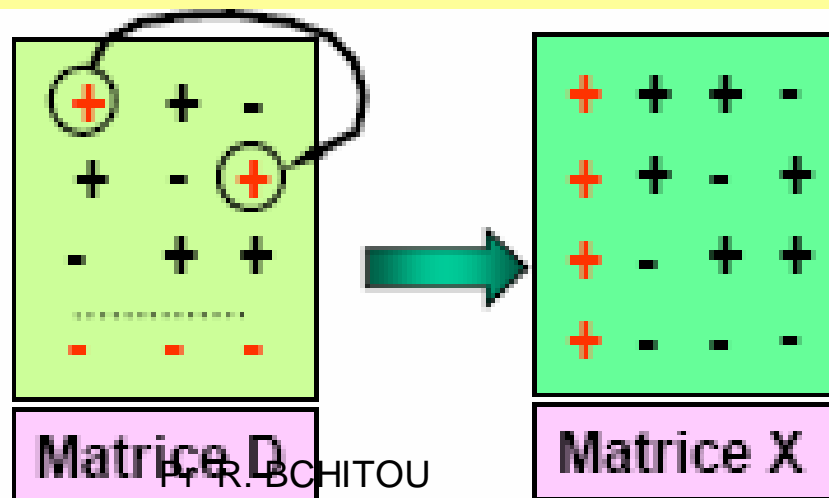
$k \leq 11$  (N=12) : + + - + + + - - - + -

$k \leq 15$  (N=16) : + + + + - + - + + - - + - - -

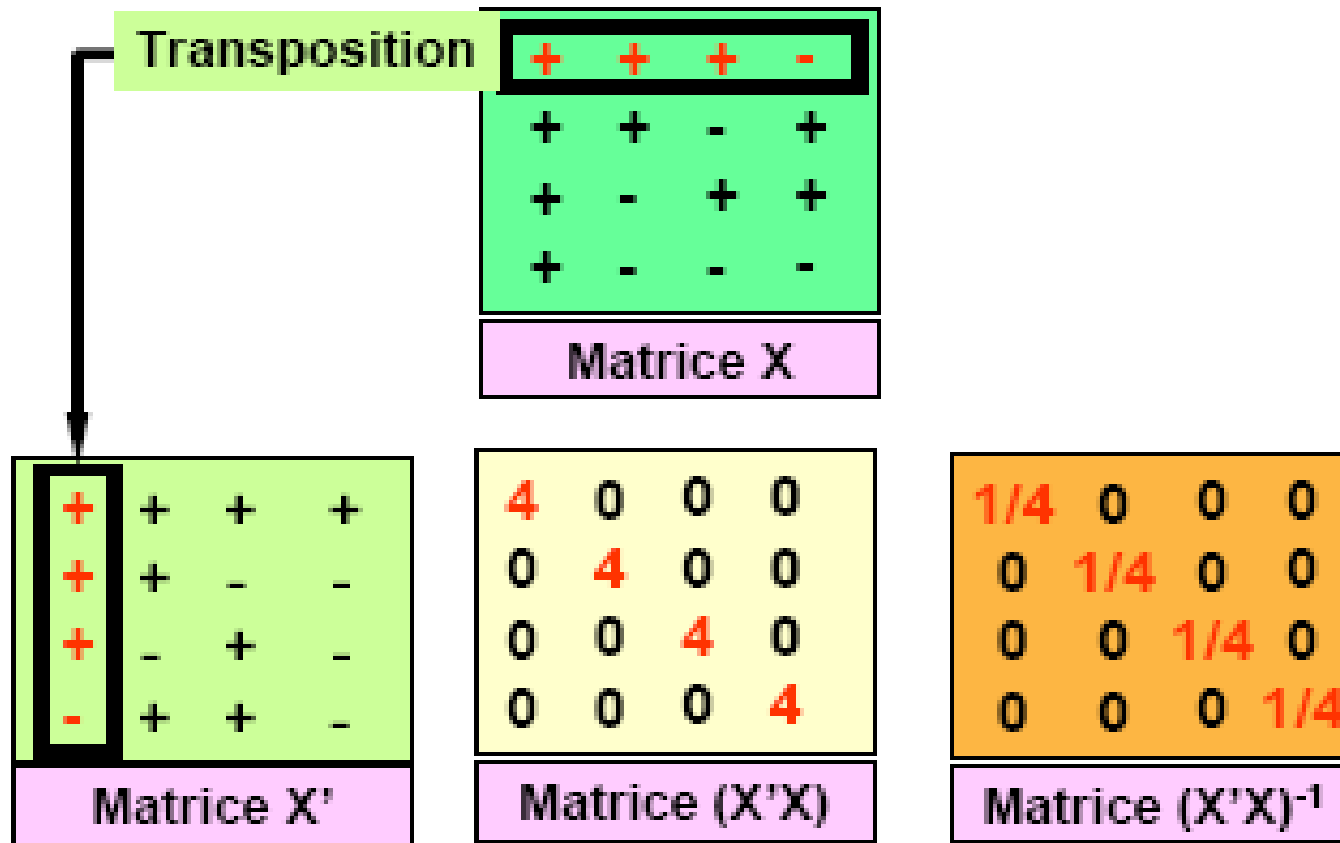
$k \leq 19$  (N=20) : + + - - + + + + - + - + - - - - + + -

$k \leq 23$  (N=24) : + + + + + - + - + + - - + + - - + - + - - - -

Exemple N = 4  
après N-2 permutations



# Plans de Pesées et Matrices d'Hadamard



# Matrice d'expérience

$k \leq 7$  (N=8)

: + + + - + - -

Décalage droite

N° Exp	X1	X2	X3	X4	X5	X6	X7
1	1	1	1	-1	1	-1	-1
2		1	1	1	-1	1	-1
3							
4							
5							
6							
7							
8							

-1

# Matrice d'expérience

$k \leq 7$  (N=8)

: + + + - + - -

**Décalage droite**

N° Exp	X1	X2	X3	X4	X5	X6	X7
1	1	1	1	-1	1	-1	-1
2	-1	1	1	1	-1	1	-1
3		-1	1	1	1	-1	1
4							
5							
6							
7							
8							

-1

# Matrice d'expérience

$k \leq 7$  (N=8)

: + + + - + - -

N° Exp	X1	X2	X3	X4	X5	X6	X7
1	1	1	1	-1	1	-1	-1
2	-1	1	1	1	-1	1	-1
3	-1	-1	1	1	1	-1	1
4	1	-1	-1	1	1	1	-1
5	-1	1	-1	-1	1	1	1
6	1	-1	1	-1	-1	1	1
7	1	1	-1	1	-1	-1	1
8	-1	-1	-1	-1	-1	-1	-1

Dernière ligne : - - - - -

Matrice complète

# *Chapitre VI*

## **Plans factoriels complets FFD**

Les plans factoriels complets sont caractérisés par :

## Matrices

synergique

Effet d'interaction

$$Y (\text{réponse}) = b_0 + b_1 X_1 + b_2 X_2 + b_{12} X_1 X_2$$

Factorielles

Pour  $k = 2$ , on obtient :

- *Une matrice d'expériences  $2^2$*  : soit  $N = 4$  lignes et 2 colonnes.
- *La matrice du modèle  $X$*  :  $N = 4$  lignes et 4 colonnes.



# Modèle synergique : deux facteurs

Plan factoriel complet pour 2 facteurs avec interaction

Exp.N°	X <sub>1</sub>	X <sub>2</sub>
1	-1	-1
2	1	-1
3	-1	1
4	1	1

Matrice d'expériences

Plan factoriel complet 2<sup>2</sup>

Toutes les combinaisons de 2 facteurs pouvant prendre 2 niveaux

$$\text{Modèle : } Y = b_0 + b_1 X_1 + b_2 X_2 + b_{12} X_1 X_2$$

Exp.N°	X <sub>0</sub>	X <sub>1</sub>	X <sub>2</sub>	X <sub>1</sub> X <sub>2</sub>
1	1	-1	-1	1
2	1	1	-1	-1
3	1	-1	1	-1
4	1	1	1	1

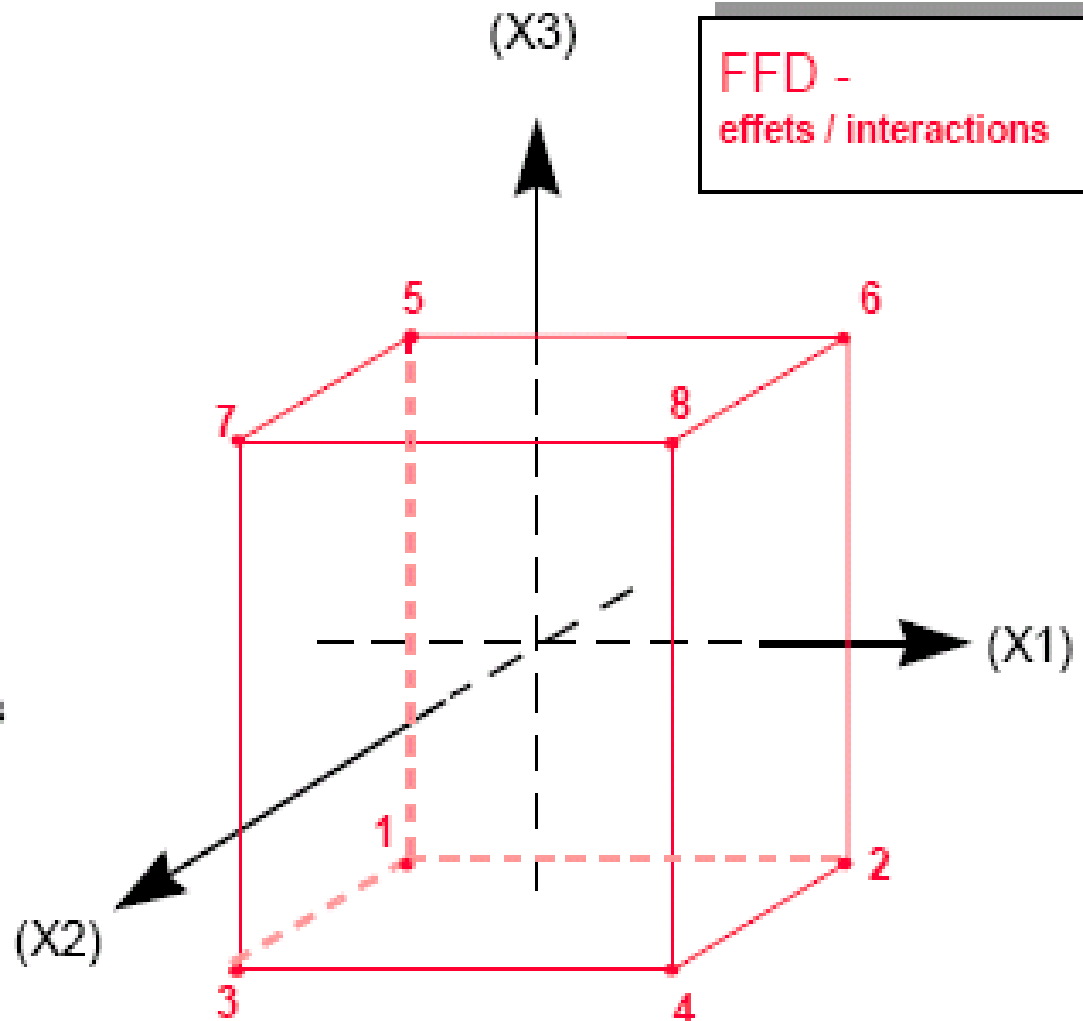
Matrice du modèle

# Matrice d'expériences: 3 facteurs

Essai	X1	X2	X3
1	-1	-1	-1
2	1	-1	-1
3	-1	1	-1
4	1	1	-1
5	-1	-1	1
6	1	-1	1
7	-1	1	1
8	1	1	1

**Matrice d'expériences**  
**Plan factoriel complet  $2^3$**   
Toutes les combinaisons  
de 3 facteurs pouvant  
prendre 2 niveaux

**Algorithme de Yates**



# Modèle synergique : 3 facteurs

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_{12} X_1 X_2 + b_{13} X_1 X_3 + b_{23} X_2 X_3 + b_{123} X_1 X_2 X_3$$

1 terme constant

3 effets principaux  $b_i$

1 effet d'interaction du second ordre  $b_{ijk}$

3 effets d'interaction du premier ordre  $b_{ij}$

N°exp.	$X_0$	$X_1$	$X_2$	$X_3$	$X_1 X_2$	$X_1 X_3$	$X_2 X_3$	$X_1 X_2 X_3$
1	1	-1	-1	-1	1	1	1	-1
2	1	1	-1	-1	-1	-1	1	1
3	1	-1	1	-1	-1	1	-1	1
4	1	1	1	-1	1	-1	-1	-1
5	1	-1	-1	1	1	-1	-1	1
6	1	1	-1	1	-1	1	-1	-1
7	1	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1	1

Matrice du modèle

En conclusion, pour  $k=3$ , on obtient :

Matrices d'expériences : matrice de taille  $2^3$   
soit  $N=8$  lignes et 3 colonnes

Matrice du modèle  $X$  : 8  
lignes et 8 colonnes

N°exp.	$X_0$	$X_1$	$X_2$	$X_3$	$X_1X_2$	$X_1X_3$	$X_2X_3$	$X_1X_2X_3$
1	1	-1	-1	-1	1	1	1	-1
2	1	1	-1	-1	-1	-1	1	1
3	1	-1	1	-1	-1	1	-1	1
4	1	1	1	-1	1	-1	-1	-1
5	1	-1	-1	1	1	-1	-1	1
6	1	1	-1	1	-1	1	-1	-1
7	1	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1	1

## Exemple d'utilisation d'une matrice factorielle complète : étude quantitative des facteurs

La réaction enzymatique est réalisée par la culture de la souche microbienne sélectionnée, dans un milieu nutritif contenant le substrat à déshydrogéner :  
c'est le développement de la souche qui libère le système enzymatique dans le milieu nutritif et qui "métabolise" le substrat.

Facteurs	Bornes	
	(-1)	(1)
$U_1$ : Quantité de liqueur de maïs en g/l	10	20
$U_2$ : Durée de réaction en heures	24	48
$U_3$ : Quantité de glucose en g/l	5	10

RÉPONSE Y étudiée :

Quantité de substrat (mg) déshydrogénée enzymatiquement. 141

# Etude d'une déshydrogénation enzymatique

Matrice d'expériences

Exp.N°	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	Y
1	-1	-1	-1	230
2	1	-1	-1	205
3	-1	1	-1	110
4	1	1	-1	70
5	-1	-1	1	270
6	1	-1	1	220
7	-1	1	1	110
8	1	1	1	70

Plan d'expérimentation

U <sub>1</sub> Quantité liq.maïs	U <sub>2</sub> Durée réaction	U <sub>3</sub> Qté glucose	mg subst transf.
10	24	5	230
20	24	5	205
10	48	5	110
20	48	5	70
10	24	10	270
20	24	10	220
10	48	10	110
20	48	10	70

## Matrice du modèle

Exp.N°	$X_0$	$X_1$	$X_2$	$X_3$	$X_1X_2$	$X_1X_3$	$X_2X_3$	$X_1X_2X_3$	Y
1	1	-1	-1	-1	1	1	1	-1	230
2	1	1	-1	-1	-1	-1	1	1	205
3	1	-1	1	-1	-1	1	-1	1	110
4	1	1	1	-1	1	-1	-1	-1	70
5	1	-1	-1	1	1	-1	-1	1	270
6	1	1	-1	1	-1	1	-1	-1	220
7	1	-1	1	1	-1	-1	1	-1	110
8	1	1	1	1	1	1	1	1	70

$I_1$	$I_2$	$I_3$	$I_4$	$I_5$	$I_6$	$I_7$	$I_8$
-------	-------	-------	-------	-------	-------	-------	-------

# Exploitation des résultats :

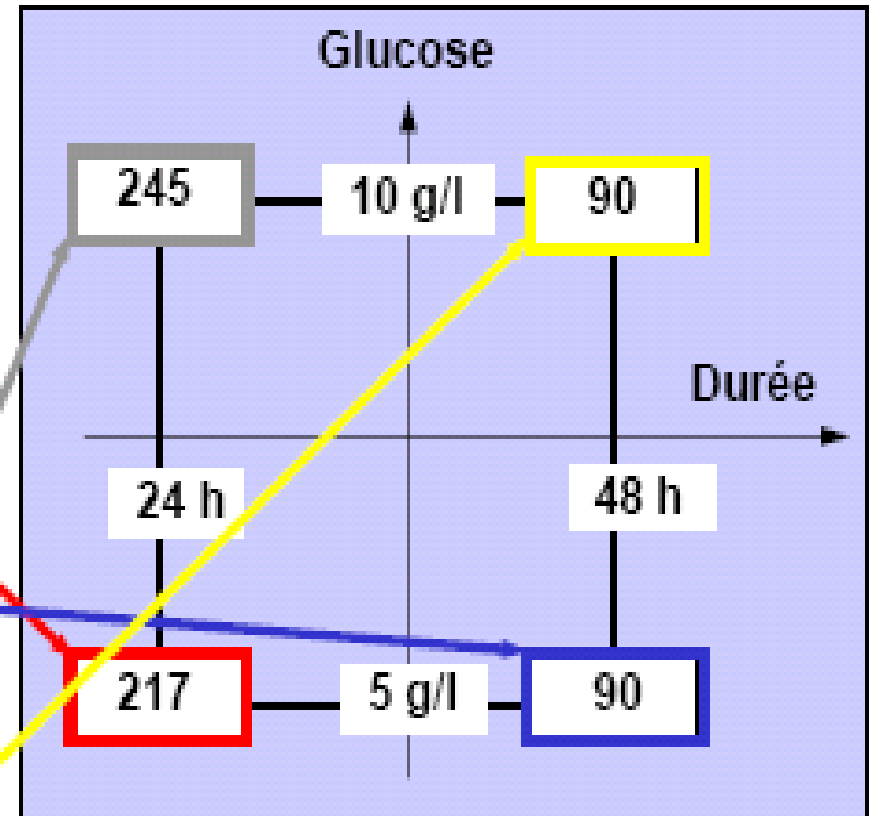
- a/ Calcul des coefficients (*Déterminer une estimation ponctuelle des effets principaux sans et avec interaction*).
- b/ Analyse Statistique des résultats
- c/ Significativité des coefficients



# Représentation graphique d'une interaction

## Matrice d'expériences

$X_1$	$X_2$	$X_3$	Y
-1	-1	-1	230
1	-1	-1	205
-1	1	-1	110
1	1	-1	70
-1	-1	1	270
1	-1	1	220
-1	1	1	110
1	1	1	70



## a/ Calcul de coefficients

Nom	Coefficients
$b_0$	160,6
$b_1$	-19,4
$b_2$	-70,6
$b_3$	6,9
$b_{12}$	-0,6
$b_{13}$	-3,1
$b_{23}$	-6,9
$b_{123}$	3,1

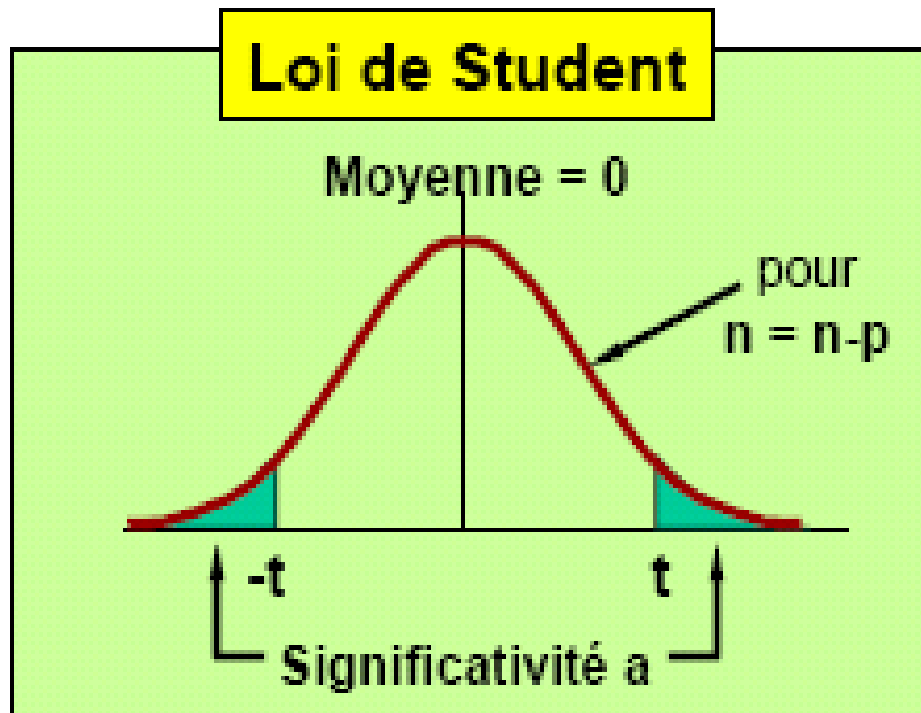
La durée de la réaction (facteur  $X_2$ ) a le plus d'influence : en moyenne le rendement diminue de 140 g/l quand on passe de 24 à 48 h.

Cela signifie que le produit de déshydrogénation est à son tour métabolisé, il ne faut pas laisser la culture trop se développer.

La quantité de liqueur de maïs (facteur  $X_1$ ) doit également être choisi à son niveau minimum (10 g/l),

La quantité de glucose (facteur  $X_3$ ) n'a proportionnellement que peu d'effet.

# c/ Significativité des coefficients



$$t = \frac{b_i - b_i^0}{\text{é.t.}(b_i)}$$

$$t = \frac{b_i}{\text{é.t.}(b_i)}$$

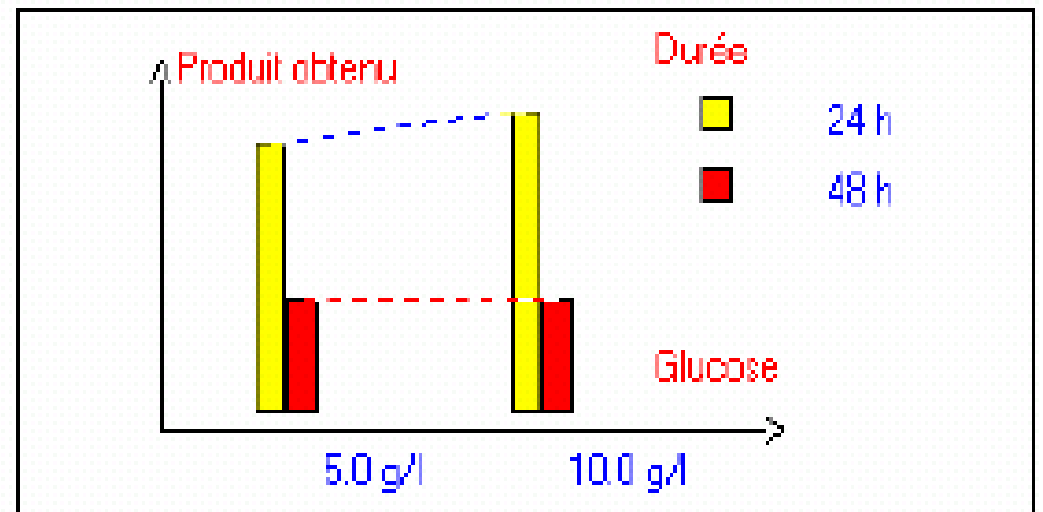
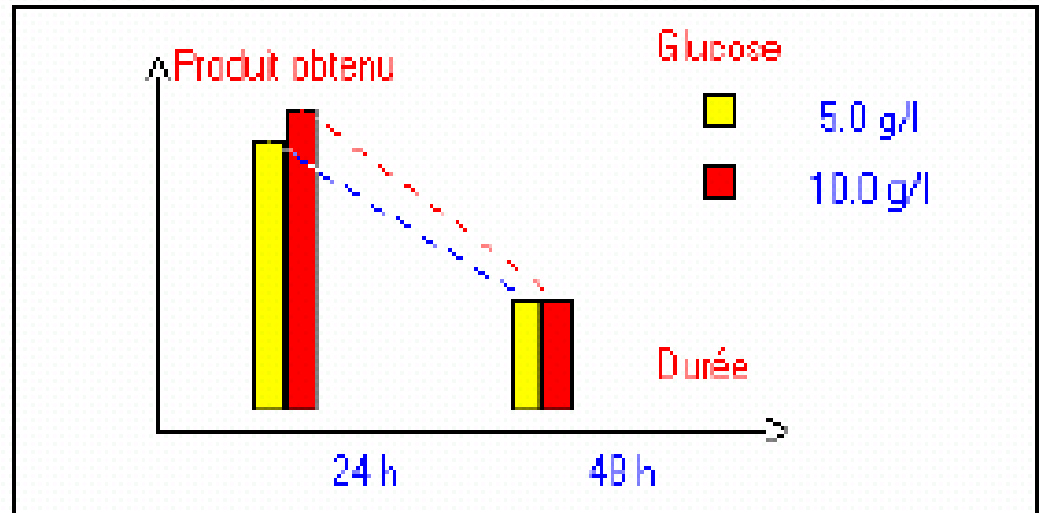
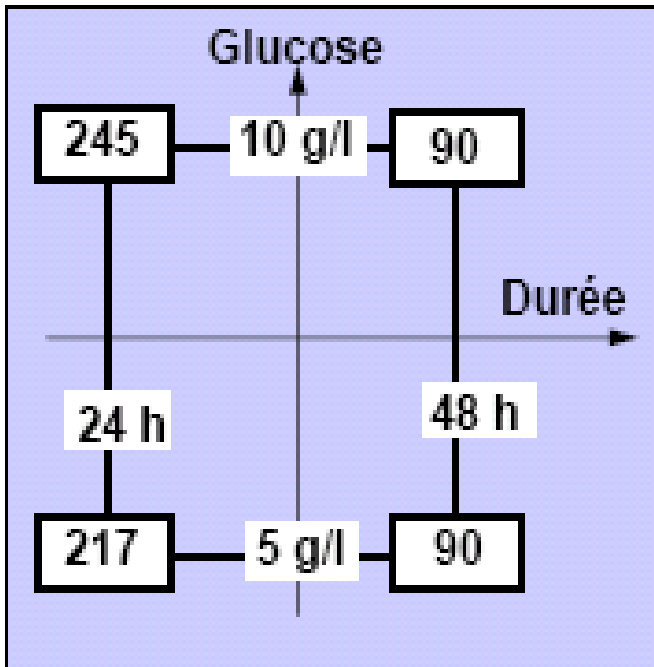
## Exploitation des résultats

Réexamen «statistique» en négligeant  $b_{123}$  comme coefficient :

➔ sa valeur est alors considérée comme **une estimation** de  $\frac{\sigma}{\sqrt{N}}$

Nom	Coefficient	Ecart-type	t.exp.	Signif. %
$b_0$	160,6	3.1	51.40	*
$b_1$	-19,4	3.1	-6.20	10.6
$b_2$	-70,6	3.1	-22.60	*
$b_3$	6,9	3.1	2.20	27.9
$b_{12}$	-0,6	3.1	-0.20	86.8
$b_{13}$	-3,1	3.1	-1.00	50.0
$b_{23}$	-6,9	3.1	-2.20	27.9

# Représentation graphique d'une interaction



# Conclusion

## Grandes étapes d'une étude

- Recherche des variables importantes

- Parmi toutes les variables susceptibles d'influencer le phénomène étudié :
  - lesquelles ont une influence significative ou non?
  - que vaut cette influence? (positif ou négatif )
  - y a-t-il des interactions entre les variables expérimentales ?

- Modélisation

- Chercher l'allure de cette influence : linéaire ou courbe; quelle équation permettant de décrire les variations du phénomène étudié avec les variables importantes

- Optimisation

- chercher les conditions expérimentales qui donnent le meilleur résultat pour un contrainte.

# *Chapitre VII*

## **Analyse en composante principale      ACP**

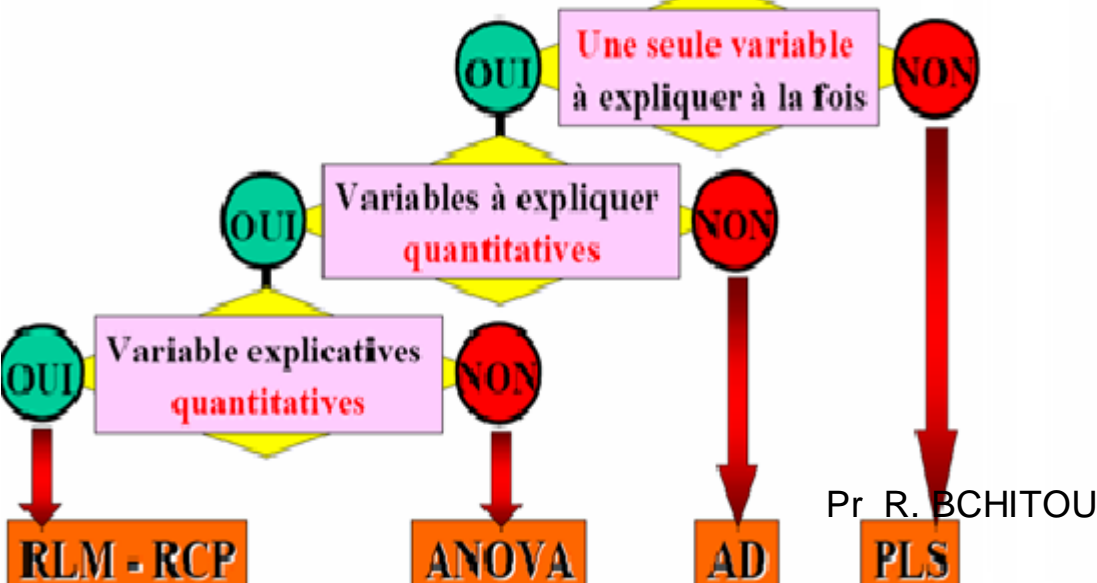
Variables explicatives  
et à expliquer **séparées** :

OUI

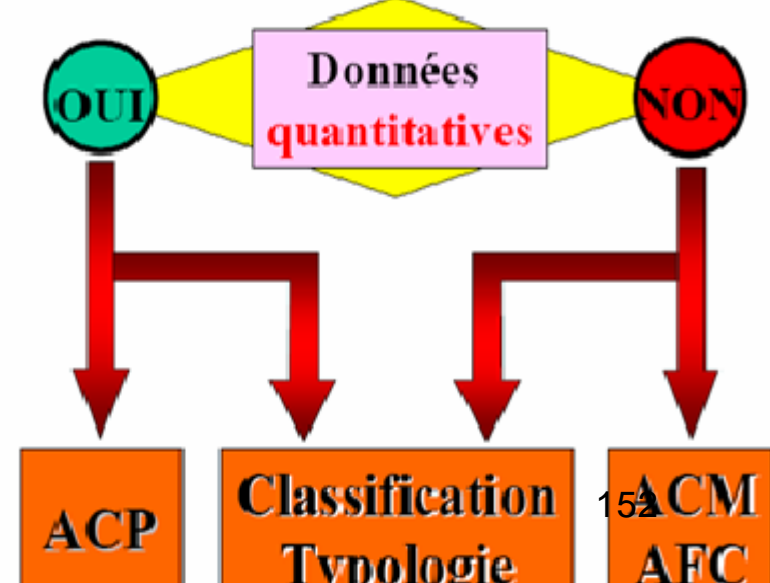
NON

Méthodes **explicatives**

Méthodes **descriptives**



Pr. R. BCHITOU





# Problème de la Compression des Données

Comment  
prendre en compte :

Un **grand nombre** de variables

Un **nombre limité** d'individus

l'information pertinente étant noyée dans une  
**masse d'informations** partiellement **redondantes**  
et polluées par **toutes sortes d'erreurs** ?

Peut-on **réduire la masse** des données  
**sans perdre trop d'information** ?  
et **améliorer la lisibilité** du contenu des données ?

# L'Analyse en Composantes Principales (ou Principal Component Analysis)

Elle consiste à **transformer** les **m** variables quantitatives initiales toutes plus ou moins **corrélées** entre elles

en **m** nouvelles variables quantitatives **non corrélées** appelées **composantes principales**.

Puis, si possible, **ne conserver qu'un nombre k réduit** de composantes.

## Principe de l'ACP

A partir de la matrice  $X$  des données de départ on construit la **matrice M de corrélation** :

$$M = 1/n (X'X)$$

Avec  $X'$  matrice transposée de  $X$  ;

*N.B. :  $X$  peut être préalablement transformée en variables centrées, en variables centrées réduites.*

La **matrice M de corrélation** est **diagonalisée** pour fournir :

une matrice diagonale  $D$  des **valeurs propres**  $\lambda_1, \lambda_2, \dots, \lambda_m$

une matrice  $P$  des **vecteurs propres**.

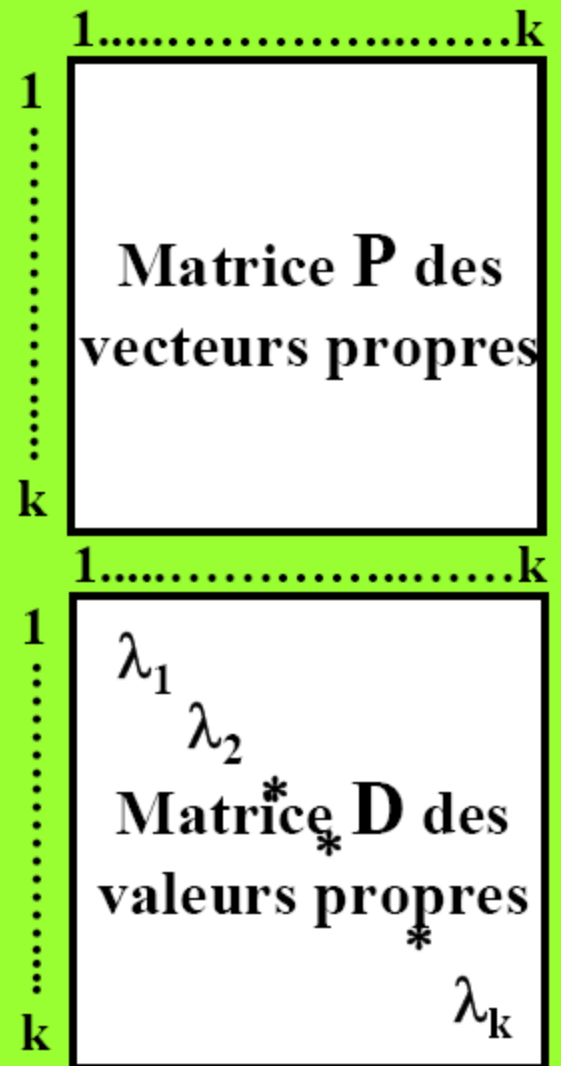
Les **vecteurs colonnes** de la matrice **P** représentent les **combinaisons linéaires** des variables de départ conduisant aux **nouveaux axes factoriels** :



les **Composantes Principales**.

Les **valeurs propres**  $\lambda_1, \lambda_2, \dots, \lambda_m$  représentent les **variances des individus** sur les composantes principales correspondantes.

$$D = P^{-1} M P$$

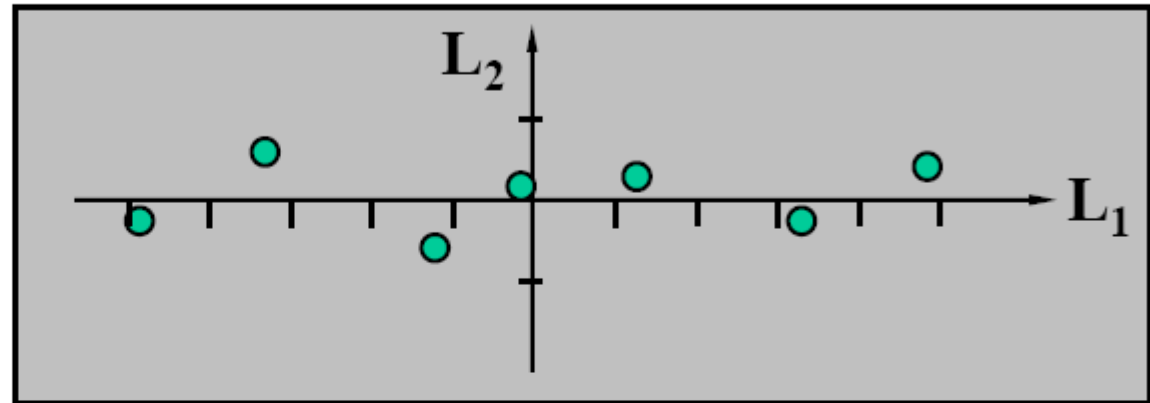


**Le rapport** de chaque valeur propre à la somme de toutes les valeurs propres donne **la part de toute l'information initiale** visible sur chaque axe :

$$\% \text{ information} = \frac{\lambda_i}{\sum \lambda_i}$$

On peut alors faire une "projection" d'un espace à k dimensions initiales à un espace plus réduit (2, 3) **en sachant quel pourcentage de l'information est conservé** (90 % par exemple).

$L_1$	$L_2$
-2.124	0.101
-1.420	-0.274
-0.569	0.249
-0.024	-0.057
0.570	-0.084
1.417	0.174
2.149	-0.109



Les descripteurs dans ce nouveau système d'axes sont **totallement décorrélés** :

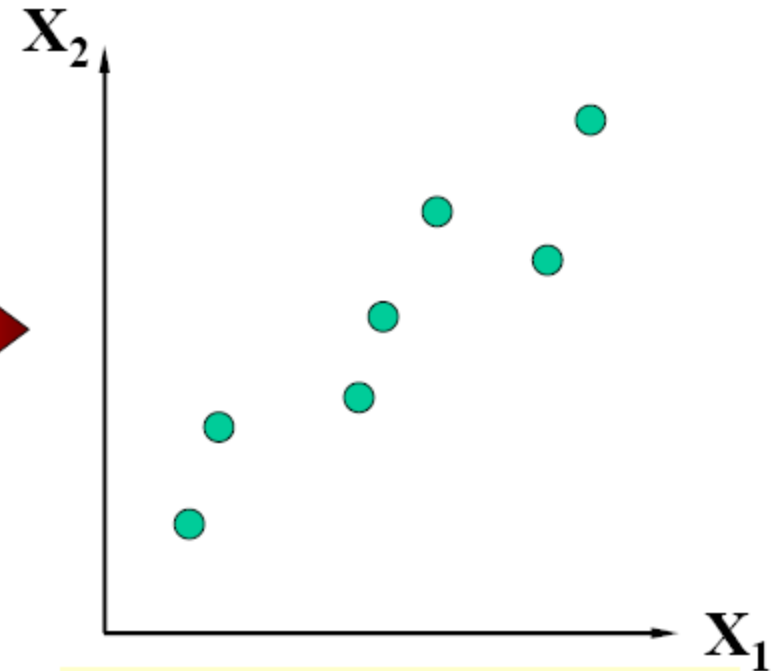


Matrice de  
**corrélacion**

	$L_1$	$L_2$
$L_1$	1	
$L_2$	<b>0.00</b>	1

# Composantes Principales dans le cas de 2 variables

$X_1$	$X_2$
2.8	13.3
3.9	16.6
8.5	17.6
9.3	20.2
11.2	22.1
14.9	23.9
16.4	27.0



Soit 7 individus décrits par deux variables  $X_1$  et  $X_2$  fortement corrélées



Matrice de **corrélation**

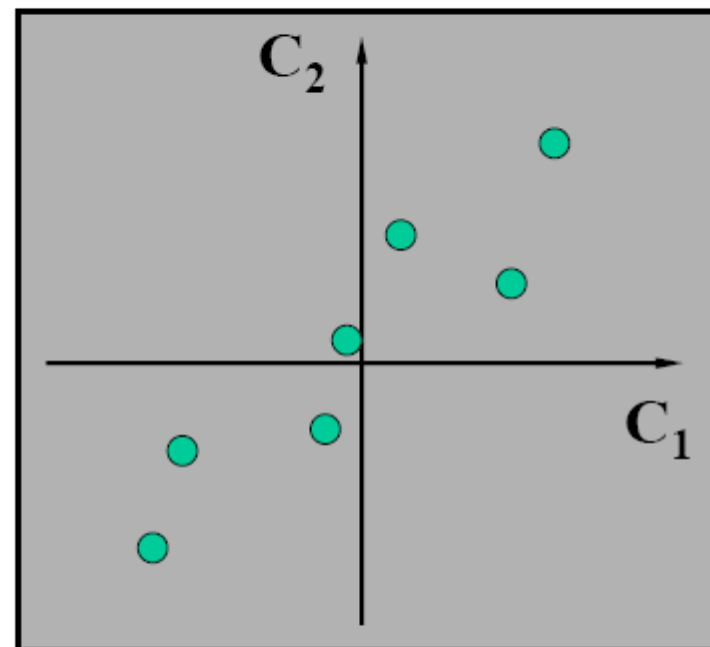
	$X_1$	$X_2$
$X_1$	1	
$X_2$	0.97	1

## Composantes Principales dans le cas de 2 variables

$X_1$	$X_2$
2.8	13.3
3.9	16.6
8.5	17.6
9.3	20.2
11.2	22.1
14.9	23.9
16.4	27.0



$C_1$	$C_2$
-6.77	-6.80
-5.67	-3.50
-1.07	-2.50
-0.27	0.10
1.63	2.00
5.33	3.80
6.83	6.90



Les **données** sont **centrées**  
autour de la moyenne

$$c_i = (x_i - \bar{x})$$



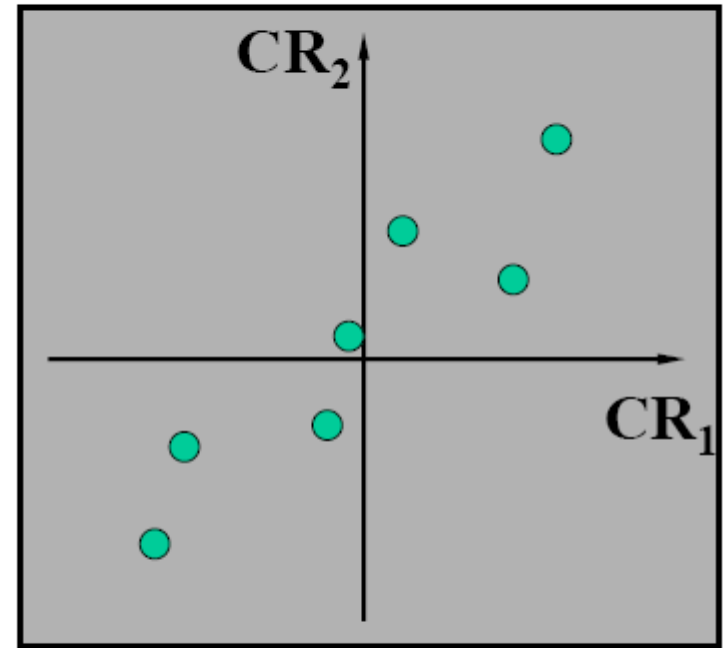


## Composantes Principales dans le cas de 2 variables

$X_1$	$X_2$
2.8	13.3
3.9	16.6
8.5	17.6
9.3	20.2
11.2	22.1
14.9	23.9
16.4	27.0



$CR_1$	$CR_2$
-1.32	-1.46
-1.11	-0.75
-0.21	-0.54
-0.05	0.02
0.32	0.43
1.04	0.81
1.34	1.48

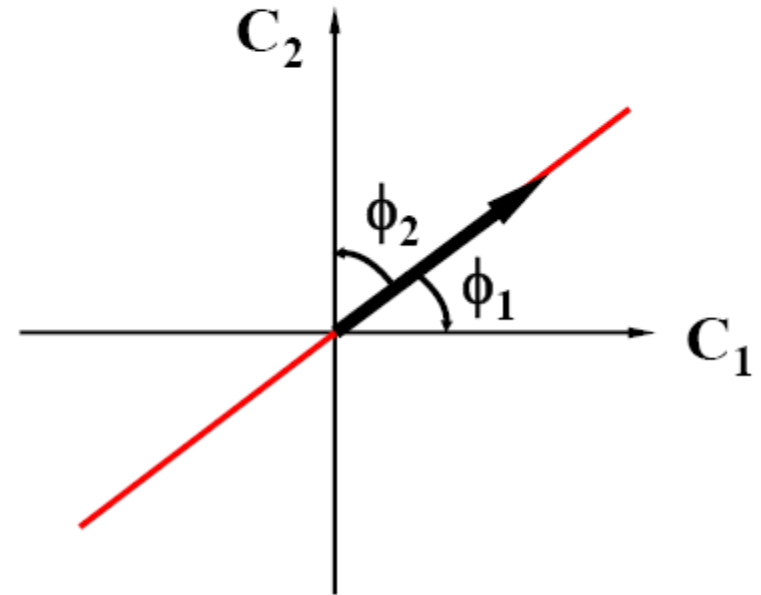
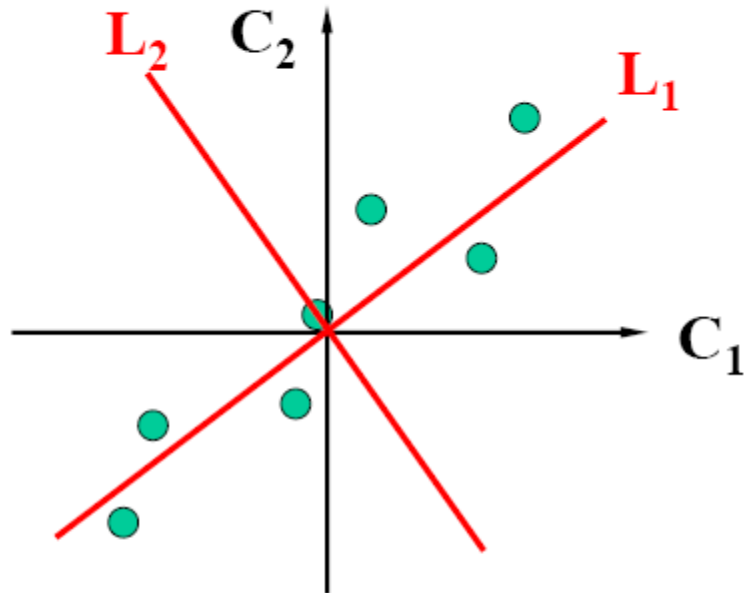


Ou bien, les **données** sont  
**centrées réduites**

$$cr_i = (x_i - \bar{x})/\sigma$$



## Retenons le passage aux variables **centrées réduites**



La **première composante principale** est la **droite des moindres carrés**.

La **seconde composante** est **orthogonale** à la première.

**Projection orthogonale du vecteur unitaire :**

$$p_1 = \cos \phi_1$$

$$p_2 = \cos \phi_2$$

Retenons le passage aux variables **centrées réduites** qui conduit à des nombres sans dimensions :

Ce changement de variable confère **le même poids** à chacun des descripteurs.



Le passage aux variables centrées réduites étant une **transformation linéaire**, il n'y a **pas de changement** de la matrice de corrélation.



Matrice de **corrélation**

	$X_1$	$X_2$
$X_1$	1	
$X_2$	0.97	1

Valeur propre	1.971	0.029
% variance	98.571	1.429
% cumulé	98.571	100

### Vecteurs propres

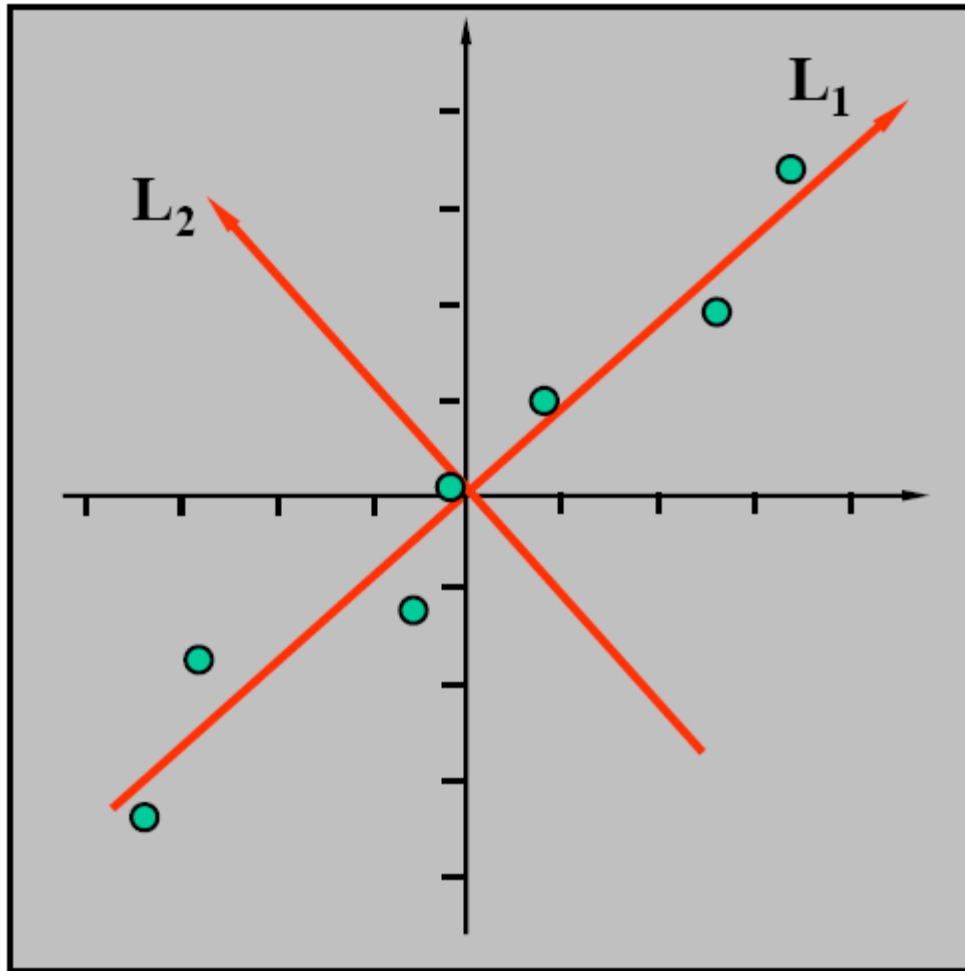
V1	V2
0.707	0,707
0.707	-0,707

V1	V2
a11	a12
a21	a22

### Calcul des coordonnées en composantes principales

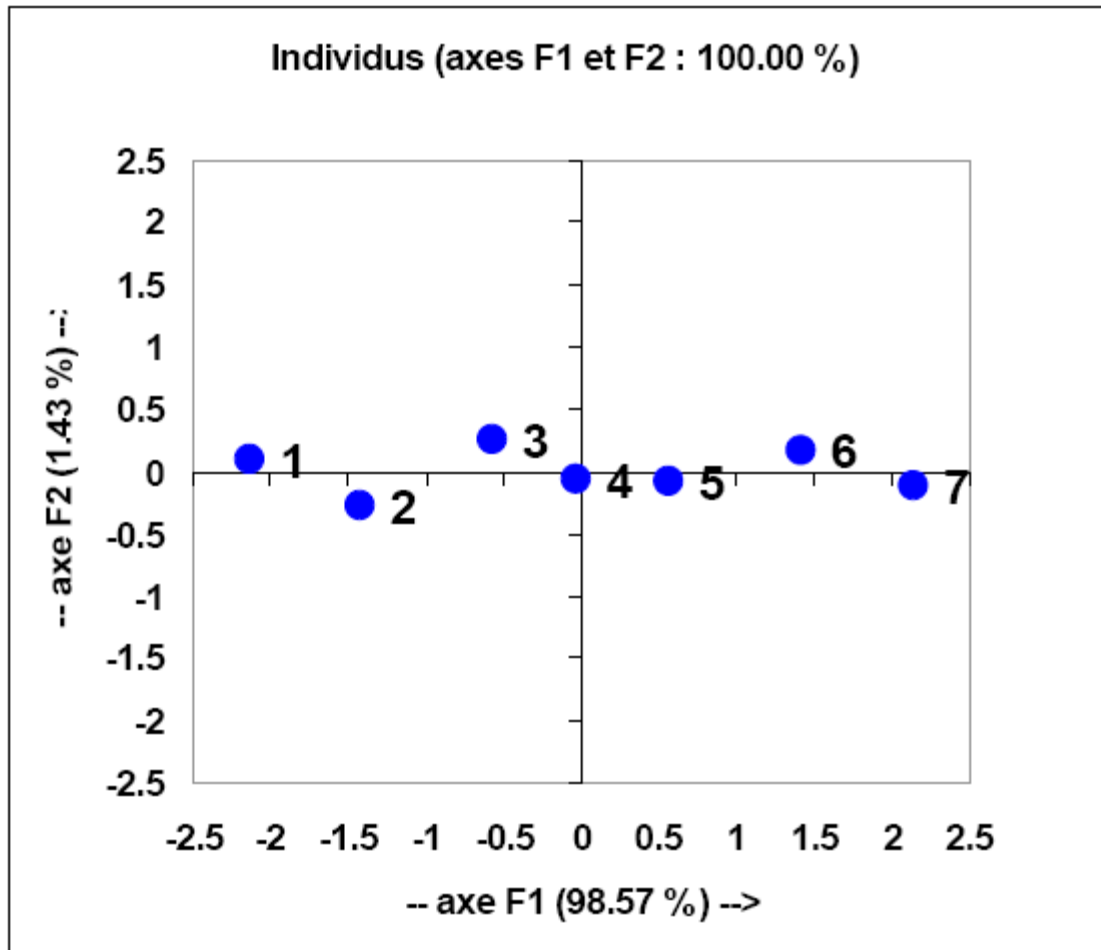
$$L1 = CR1 * a11 + CR2 * a21$$

$$L2 = CR1 * a12 + CR2 * a22$$



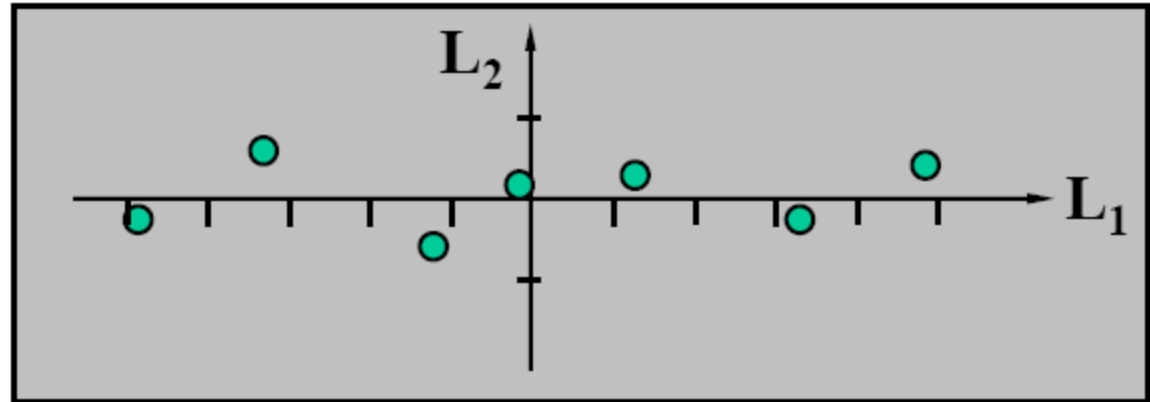
Les **coordonnées** des individus dans **l'espace des composantes** sont appelées les « **SCORES** ».

$L_1$	$L_2$
-2.124	0.101
-1.420	-0.274
-0.569	0.249
-0.024	-0.057
0.570	-0.084
1.417	0.174
2.149	-0.109



$F_1$	$F_2$
-2.124	0.101
-1.420	-0.274
-0.569	0.249
-0.024	-0.057
0.570	-0.084
1.417	0.174
2.149	-0.109

$L_1$	$L_2$
-2.124	0.101
-1.420	-0.274
-0.569	0.249
-0.024	-0.057
0.570	-0.084
1.417	0.174
2.149	-0.109

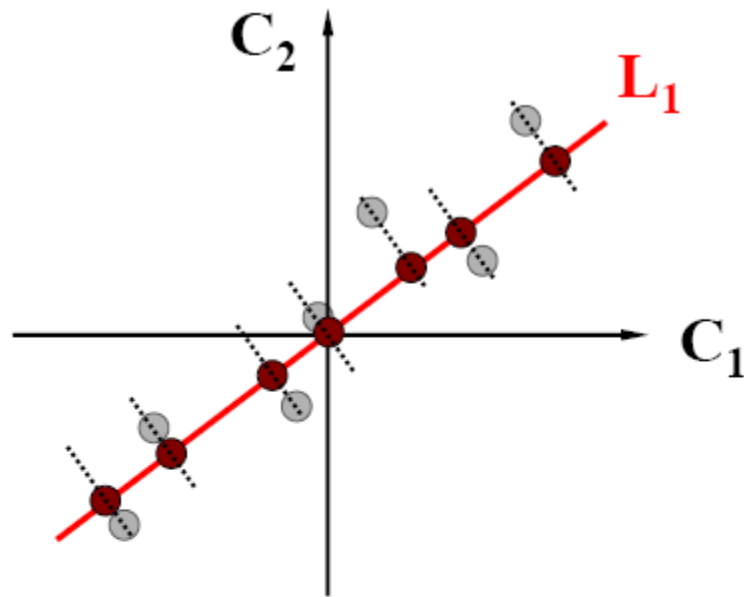


Les descripteurs dans ce nouveau système d'axes sont **totallement d corr l s** :

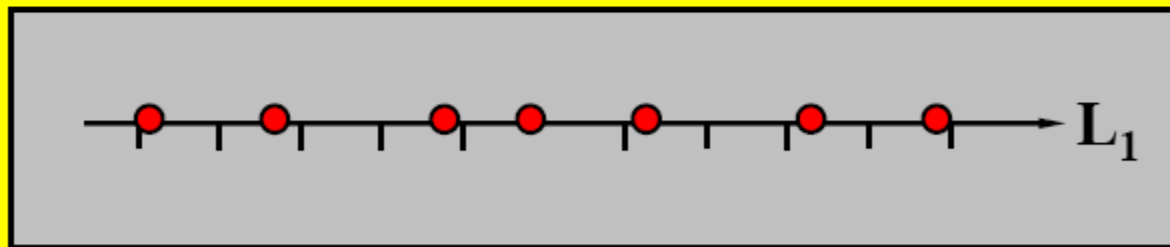


Matrice de  
**corr lation**

	$L_1$	$L_2$
$L_1$	1	
$L_2$	<b>0.00</b>	1

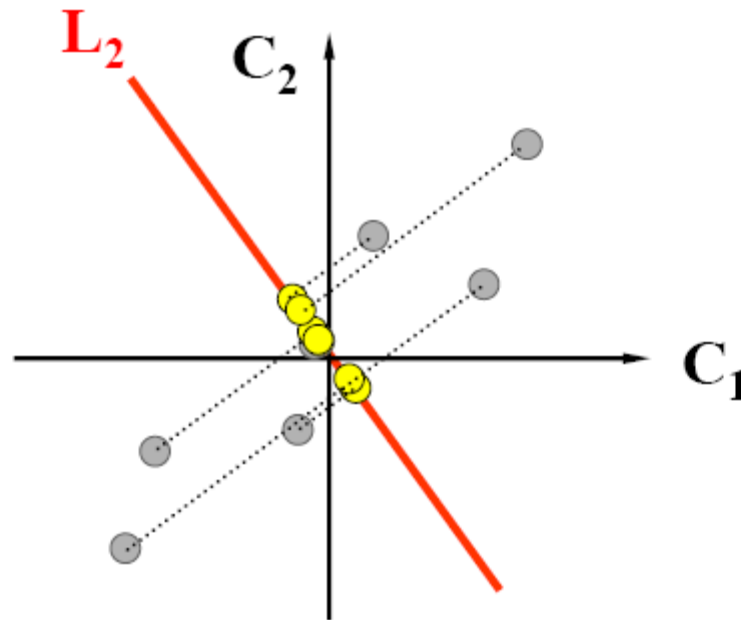


**A elles seules, les projections des individus sur la première composante :**

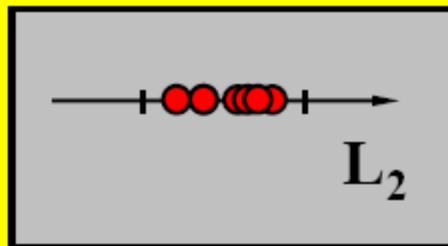


« visualisent ou **expliquent** » la quasi totalité de la « **dispersion** » des individus.



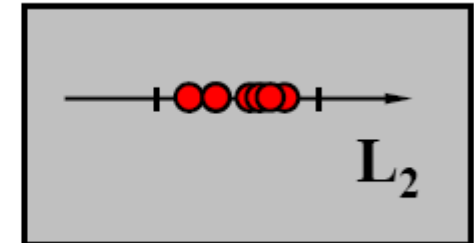
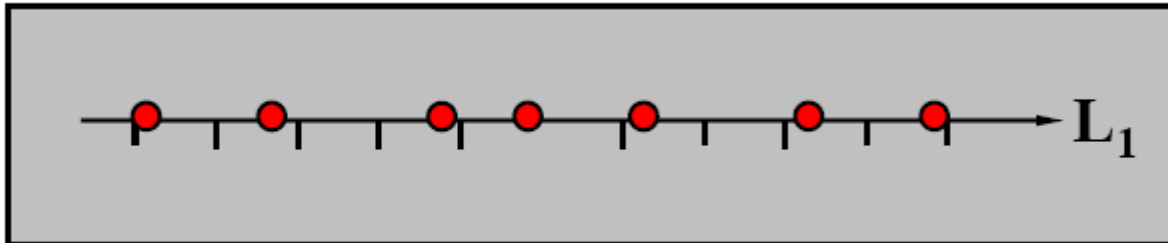


Le fait de négliger la part de « dispersion » expliquée par les projections des individus sur la seconde composante :



constitue une **perte d'information négligeable.**

**L'information** apportée par une projection, dépend fortement de la direction, donc ici de la composante principale considérée ...



# Application

Le levain remplace la levure boulangère utilisée dans les pains industriels et traditionnels. Les levures et les bactéries lactiques du levain font monter la pâte et donnent du goût au pain.

il s'agit d'un mélange de farine , d'eau et de sel qu'on laisse pétrir pendant une vingtaine de minutes dans un pétrin mécanique pour obtenir une pâte élastique qui sera abandonnée une quinzaine de minutes dans le pétrin.

Puis on a mesuré l'**acidité** , **la viscosité**, **la durée et la température** .

**Acidité Viscosité Durée Température**

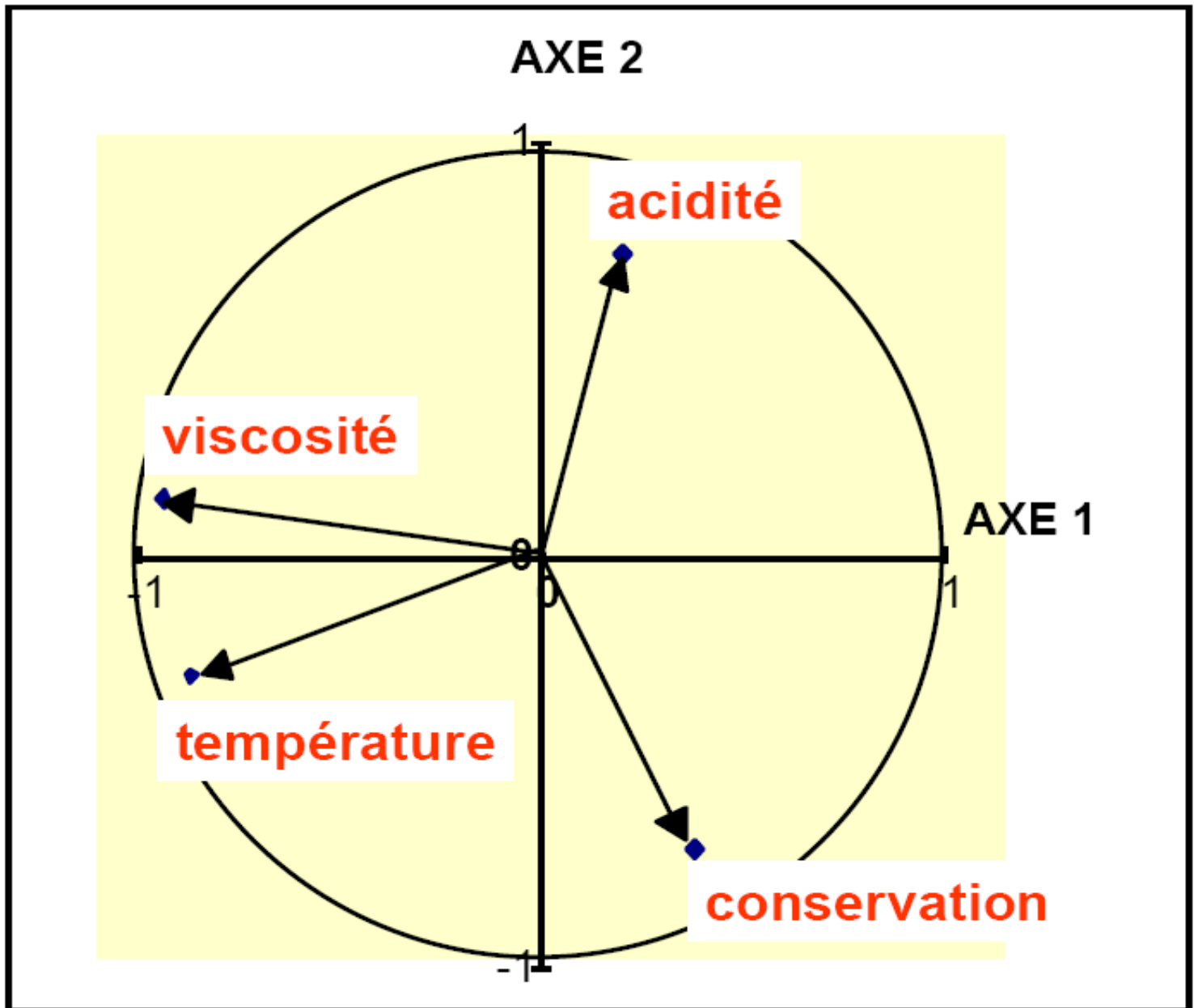
<b>individus</b>	<b>1</b>	<b>121,5</b>	<b>8100</b>	<b>1</b>	<b>8</b>	<b>L1</b>	<b>Type de levains</b>
	<b>2</b>	<b>126</b>	<b>9360</b>	<b>1</b>	<b>8</b>	<b>L2</b>	
	<b>3</b>	<b>124</b>	<b>13000</b>	<b>1</b>	<b>22</b>	<b>L1</b>	
	<b>4</b>	<b>131,5</b>	<b>12260</b>	<b>1</b>	<b>22</b>	<b>L2</b>	
	<b>5</b>	<b>147,5</b>	<b>6260</b>	<b>7</b>	<b>8</b>	<b>L1</b>	
	<b>6</b>	<b>149,5</b>	<b>8750</b>	<b>7</b>	<b>8</b>	<b>L2</b>	
	<b>7</b>	<b>153,5</b>	<b>11250</b>	<b>7</b>	<b>22</b>	<b>L1</b>	
	<b>8</b>	<b>154,25</b>	<b>12766</b>	<b>7</b>	<b>22</b>	<b>L2</b>	
	<b>9</b>	<b>149,5</b>	<b>7633</b>	<b>15</b>	<b>8</b>	<b>L1</b>	
	<b>10</b>	<b>150,5</b>	<b>9366</b>	<b>15</b>	<b>8</b>	<b>L2</b>	
	<b>11</b>	<b>155</b>	<b>8800</b>	<b>15</b>	<b>22</b>	<b>L1</b>	
	<b>12</b>	<b>157,75</b>	<b>11100</b>	<b>15</b>	<b>22</b>	<b>L2</b>	
	<b>13</b>	<b>152</b>	<b>6800</b>	<b>21</b>	<b>8</b>	<b>L1</b>	
	<b>14</b>	<b>155</b>	<b>9466</b>	<b>21</b>	<b>8</b>	<b>L2</b>	
	<b>15</b>	<b>157</b>	<b>9400</b>	<b>21</b>	<b>22</b>	<b>L1</b>	
	<b>16</b>	<b>159</b>	<b>9700</b>	<b>21</b>	<b>22</b>	<b>L2</b>	

## CORRELATIONS DES VARIABLES INITIALES

	ACIDITE	VISCOSITE	CONSERVATION	TEMPERATURE
ACIDITE	1			
VISCOSITE	-0,02	1		
CONSERVATION	-0,072	-0,341	1	
TEMPERATURE	-0,194	0,726	0	1

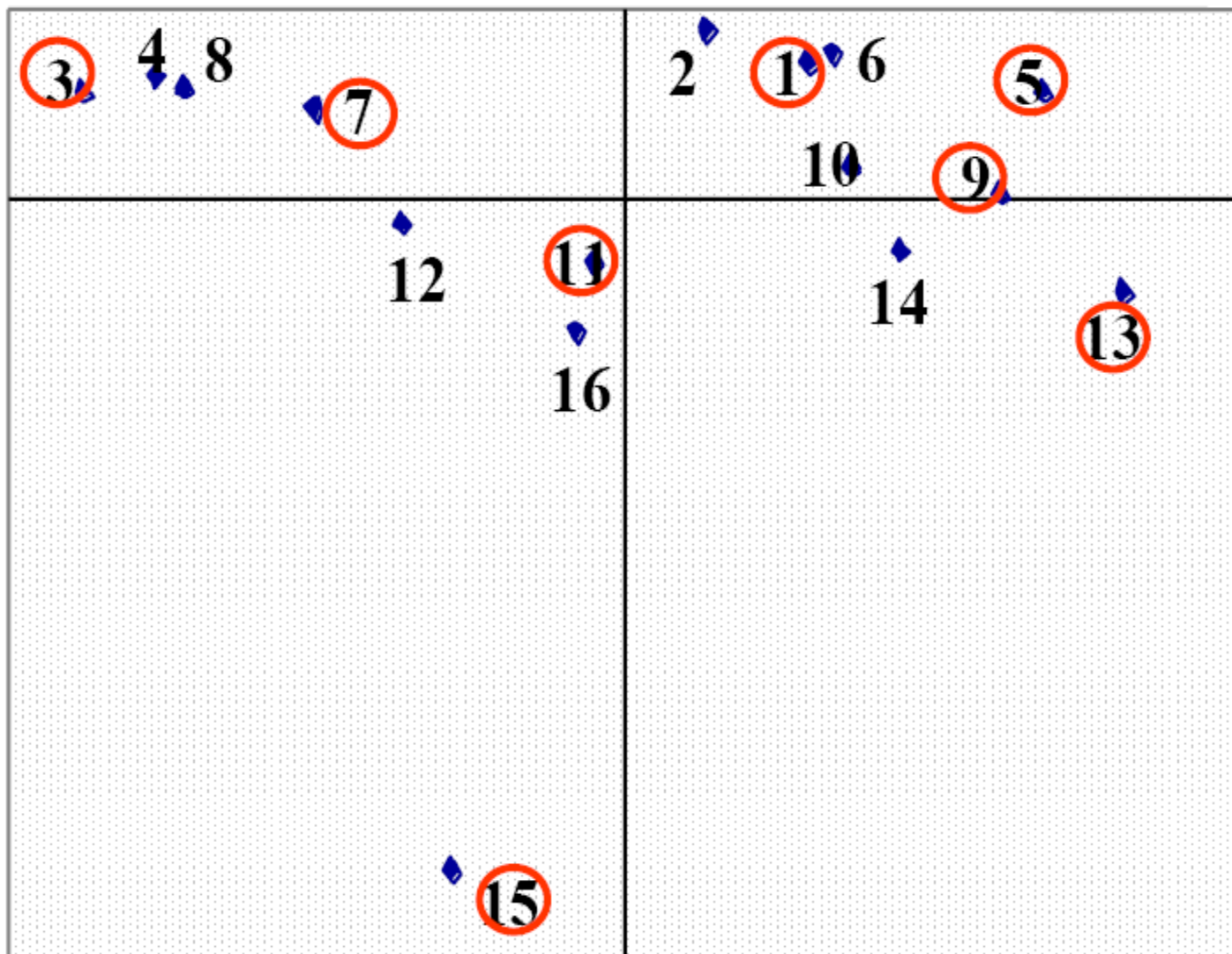
## ANALYSE DES VARIABLES SYNTHETIQUES

	AXE 1	AXE 2	AXE 3	AXE 4
ACIDITE	0,2	0,73	-0,65	0,05
VISCOSITE	-0,94	0,14	-0,12	-0,3
CONSERVATION	0,38	-0,71	-0,58	-0,12
TEMPERATURE	-0,87	-0,29	-0,27	0,28



<b>Indiv.</b>	<b>Axe 1</b>	<b>Axe 2</b>	<b>Axe 3</b>	<b>Axe 4</b>
1	0,74	0,73	1,47	0,22
2	0,32	0,89	1,31	-0,2
3	-2,2	0,57	0,54	-0,22
4	-1,91	0,66	0,45	-0,06
5	1,69	0,57	0,62	0,72
6	0,84	0,77	0,43	-0,14
7	-1,27	0,48	-0,38	0,26
8	-1,79	0,59	-0,49	-0,27
9	1,51	0,03	-0,14	-0,04
10	0,91	0,16	-0,27	-0,64
11	-0,13	-0,33	-0,89	0,82
12	-0,91	-0,13	-1,09	0,03
13	2,02	-0,49	-0,62	0,04
14	1,11	-0,26	-0,84	-0,88
15	-0,71	-3,55	1,42	-0,06
16	-0,21	-0,7	-1,51	0,3

Axe 2



Axe 1