

Introduction aux Plans Expérimentaux

Définitions et Bases Statistiques des Plans d'Expérience

Claude HOINARD

Faculté des Sciences Pharmaceutiques
« Philippe Maupas » -Tours

Les 2 grands types d'études

◆ les études non-expérimentales où la réalité est observée telle qu'elle se présente spontanément à l'observateur ; il s'agit d'études type enquête.

exemple : suivi de médecins britanniques fumeurs et non fumeurs pendant 20 ans pour étudier le lien entre tabagisme et cancer du poumon...

◆ Les études expérimentales diffèrent des précédentes par le fait que le chercheur choisit les variations du (ou des) facteur(s) étudié(s) pour observer l'effet sur une réponse

exemple : tester l'efficacité d'un médicament en comparant un groupe à qui on l'administre avec un autre groupe qui reçoit un placebo...

Les études expérimentales

La méthode expérimentale ne s'applique que lorsqu'on peut **fixer à volonté les conditions de l'observation et modifier une ou plusieurs d'entre elles de façon bien définie, réversible et avec possibilité de répétition.**

Pour désigner ces actions, on utilise les mots synonymes **expériences** ou **essais**.

C'est dans ce contexte d'études expérimentales que travaillent les **laboratoires industriels** (R et D , contrôle analytique , fabrication ...) , ainsi que les **laboratoires d'analyses** biologiques et médicales.

C'est aussi la situation habituelle pour les études sur l'**animal**.

Les études non expérimentales

Chez l'**homme**, en revanche, l'application de la méthode expérimentale est très limitée par des considérations évidentes d'éthique ...

Il existe, certes, des exceptions (essais cliniques ,essais thérapeutiques), mais dans la quasi totalité des cas le chercheur doit utiliser une autre approche : aucune intervention n'est autorisée, les résultats sont obtenus dans les conditions naturelles puis analysés en fonction de facteurs dits « d'exposition ».

Il existe de nombreux domaines où expérimenter est impossible ; par exemple :

⇒ **l'épidémiologiste** ne peut pas expérimenter : il doit travailler sur des faits tels qu'ils se présentent, en dehors de son contrôle.

⇒ **l'économiste** est également dans cette situation : lorsqu'il étudie la consommation d'un produit en fonction de son prix, il ne peut qu'observer périodiquement les consommations et les prix correspondants... il ne peut pas faire varier à son gré le prix pour voir ce que devient la consommation.

Toute étude, quelle qu'elle soit (non expérimentale ou expérimentale) doit être suivie d'une analyse par la méthode statistique. Selon son objectif, l'étude est soit à **visée descriptive**, soit à **visée explicative**.

Les études non expérimentales sont souvent à **visée descriptive** ; on peut ainsi décrire des variables par des paramètres résumés (moyenne, écart type, pourcentages ...), des intervalles de confiance, des graphiques pour visualiser la distribution ; on peut aussi mettre en évidence des **associations statistiques** entre, par exemple, une maladie et un facteur^(*), mais cela ne constitue qu'une première étape : la difficulté majeure est d'interpréter ensuite la signification de la dépendance obtenue.

(*) cancer du poumon et exposition à l'amiante

Il est important de noter que la méthode statistique en elle-même est incapable d'établir la preuve d'une relation causale entre 2 variables. C'est pourquoi **une analyse statistique doit toujours être suivie d'une interprétation.**

Pour discuter en termes de **causalité**, il est nécessaire que l'effet résultant soit **consistant** (reproductible), **spécifique** (qu'il s'agisse toujours du même effet) et que l'**explication** soit **cohérente** (biologiquement plausible).

De ce point de vue, on peut considérer qu'une étude expérimentale est plus appropriée qu'une étude non expérimentale.

I - Quelques considérations sur l'expérimentation

terminologie

- ◆ **Un facteur** (*quantitatif ou qualitatif*) est un état du système étudié dont la variation est susceptible d'en modifier le fonctionnement (*ex : température, concentration d'un réactif, nature d'un médicament...*)
- ◆ **La réponse** du système correspond au paramètre mesuré ou observé avec pour objectif de connaître l'effet du ou des facteur(s) étudié(s) sur cette réponse. Une réponse peut être **quantitative** (*ex : baisse de la pression artérielle, rendement d'une réaction ...*) ; elle peut aussi être **qualitative** (*ex : mort/survie d'un animal, type de lésion observée ...*) : elle s'exprime alors par un pourcentage.

Il existe 2 grands types de facteurs étudiés

Un facteur qualitatif se caractérise par des **états** (ce qu'on modifie), on dit aussi des **modalités** ou des **catégories** ou encore des **niveaux**.

exemples :

- ◆ Comparaison des effets de 2 médicaments : le facteur « médicament » est étudié sous 2 états M_1 et M_2
- ◆ étude du facteur « ordre d'administration » lorsque chaque sujet reçoit successivement 2 médicaments ; 2 niveaux (M_1 puis M_2 et M_2 puis M_1)
- ◆ comparaison de 3 médicaments à un placebo : le facteur est étudié sous 4 niveaux (placebo + médicaments)

Un facteur quantitatif correspond à une grandeur mesurable ; les valeurs x qu'on décide de choisir dans l'expérience sont les **niveaux**.

Exemples :

- ◆ lors de l'étude de l'influence de la température d'une réaction, 2 séries d'essais sont effectuées, à 20°C et à 37°C : le facteur « température » est étudié sous 2 niveaux
- ◆ lors de l'étalonnage d'une technique, 5 concentrations d'étalon sont utilisées : le facteur « concentration » est étudié sous 5 niveaux
- ◆ en cinétique, le degré d'avancement d'une réaction est mesuré à 6 instants : le facteur « temps » comprend 6 niveaux

La réponse (ou critère)

La réponse correspond aux résultats obtenus dans l'expérience. C'est toujours **une variable considérée comme aléatoire** qui doit être analysée au moyen de tests statistiques en fonction du ou des facteur(s) étudié(s).

Lorsqu'elle est **qualitative** (nominale ou ordinale), on utilise des **tests non paramétriques** (Khi-deux, Mann et Whitney, Wilcoxon ...)

Lorsqu'elle est **quantitative** (mesure), on utilise des **tests paramétriques**, mais à la condition que les conditions de validité de ces tests soient satisfaites. Dans la suite de ce document, nous nous limiterons à ce type de réponse quantitative.

Exemple : Résultats de 2 traitements A et B administrés à des malades.
Les 2 traitements ont-ils une efficacité significativement différente ?

	Trait A	Trait B	Total
Echecs	17	18	35
Succès	4	1	5
Total	21	19	40

Le résultat réponse qualitative



On utilise un test « non paramétrique », le test du χ^2

Exemple : Distribution du poids de nouveau-nés en fonction de la parité. Le poids dépend t-il de la parité ?

	Primipares	Multipares
Poids (kg)	2,100	2,500
	3,500	3,100
	5,100	3,900

Le poids réponse quantitative

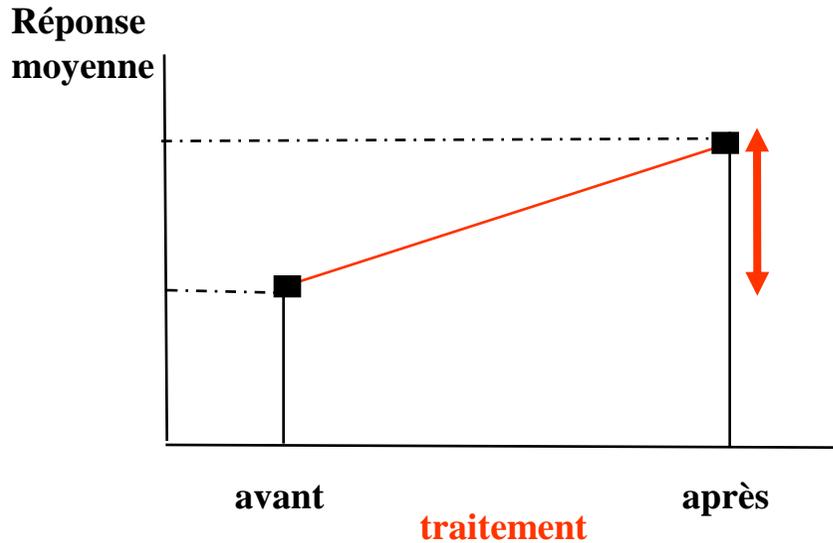


On utilise un test « paramétrique », le test t (comparaison de 2 moyennes)

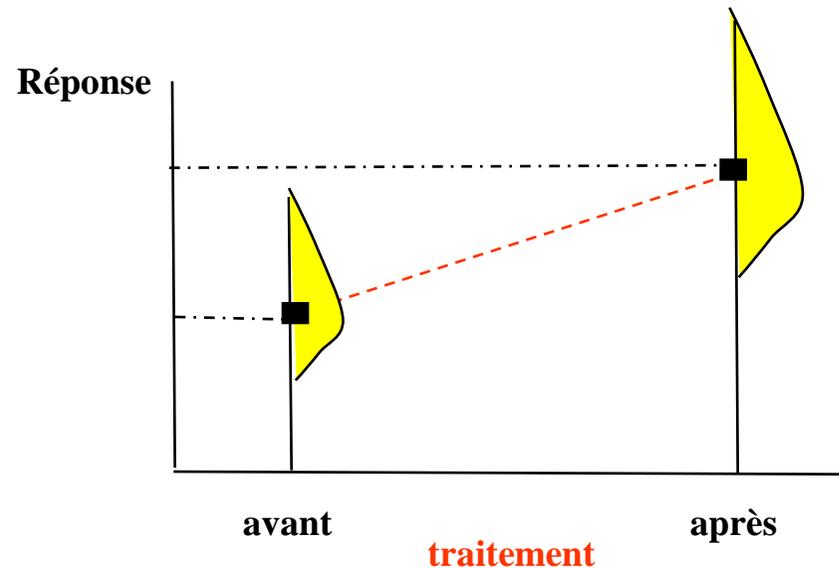
Il est important de souligner que **la variabilité des réponses individuelles** a pour effet de créer une « *sorte de bruit de fond de la réponse* » appelé **erreur expérimentale** ; celle-ci est primordiale à considérer car c'est sa grandeur (quantifiée par **l'écart type de l'erreur**) qui limite la possibilité de détecter l'effet du facteur étudié sur la réponse lors de l'utilisation des tests statistiques.

La plupart des tests statistiques fonctionnent en effet de manière analogue au rapport *signal / bruit* d'un appareil de mesure : lorsque le *signal* (**l'effet à tester**) est plus petit ou du même ordre de grandeur que le *bruit* (**l'écart type de l'erreur**), on ne peut pas affirmer l'existence de cet effet ...

Exemple : erreur expérimentale et effet significatif



Apparemment la réponse augmente nettement avec le traitement. Celui-ci est efficace



Mais si l'on fait apparaître les variations individuelles, la conclusion n'est plus forcément la même

Les origines de la variabilité des réponses individuelles sont multiples ... et sont propres à chaque type d'essais. Distinguons cependant :

- ◆ les **sciences appliquées**, qualifiées parfois « d'exactes » (chimie analytique et autres chimies, galénique industrielle, physique ...) dans lesquelles le soin apporté aux manipulations, l'imprécision des appareils de mesure, les conditions d'environnement (température, humidité, lumière ...) constituent les principales sources de fluctuations. L'**erreur expérimentale** est en général assez **faible**.
- ◆ Les **sciences biologiques**, dans lesquelles l'étude porte sur tout ou partie d'un matériel vivant qui possède son propre comportement, ce qui amène une variabilité importante de la réponse d'une unité à une autre ... et qui s'ajoute à celles mentionnées dans l'alinéa précédent. L'**erreur expérimentale** est souvent **très importante** et pose problème

Il en résulte des différences dans l'organisation des essais, selon que **l'erreur expérimentale** sur la réponse est faible ou élevée :

◆ **lorsqu'elle est faible**, on ne prévoit généralement **pas de répétitions en routine** pour chaque condition expérimentale (on dit aussi traitement) ; c'est, par exemple, le cas des dosages analytiques au laboratoire de contrôle ou dans un LABM, pour des raisons de coût d'analyse. Mais il faut bien sûr avoir préalablement vérifié que l'erreur est effectivement faible lors de la **validation de la méthode**.

◆ **Lorsqu'elle est élevée**, dans les essais biologiques par exemple, **la répétition de chaque traitement est indispensable** pour cerner un comportement moyen des unités pour chaque traitement et il est aussi intéressant de rechercher des dispositions pour essayer de diminuer cette erreur expérimentale.

Le plan expérimental

Il comprend toutes les mesures d'organisation prises pour effectuer les essais.

Il décrit l'objectif de l'étude, le ou les facteur(s) étudié(s) et leurs niveaux, la réponse et sa méthode de mesure, l'ordre d'exécution des essais et éventuellement les mesures prises pour diminuer l'erreur expérimentale.

Il doit donc préciser par écrit les méthodes, les modes opératoires, les appareils utilisés, les jours de manipulation, les opérateurs ...

Il est important de noter que le traitement statistique des résultats est très lié à la structure du plan expérimental : **c'est lors de l'élaboration de ce dernier qu'il faut penser aux méthodes statistiques qui permettront l'analyse des résultats ...**

Dans ce qui suit, nous étudierons ce qu'on peut espérer des **plans unifactoriels** (comportant l'étude d'un seul facteur), puis des **plans** comportant au moins **2 facteurs étudiés**.

De plus, nous distinguerons les expériences à **visée comparative** de celles dont l'objectif est la détermination d'une équation (**techniques de régression**).

Des dispositions pour diminuer l'erreur expérimentale

Lors de l'élaboration du plan expérimental, il peut arriver qu'il soit connu que la réponse dépende fortement de paramètres, autres que le facteur étudié, capables de créer de la dispersion dans la réponse s'ils varient pendant la durée des essais.

Par exemple, dans la comparaison des viscosités de plusieurs huiles, on sait que la température a un effet important sur la valeur de la viscosité ; dans certains dosages, le pH peut influencer fortement le résultat ...

La solution consiste à prévoir de maintenir constant le paramètre perturbateur dans tous les essais. Dans le cas de la température, on thermostatera le dispositif expérimental ; pour le pH, on tamponnera le milieu réactionnel.

De telles mesures ne sont malheureusement pas toujours réalisables ; les unités expérimentales dans les essais sur des êtres vivants (animaux, plantes, bactéries, virus ...) créent une dispersion importante de la réponse, inévitable. On tente, certes, de la réduire en choisissant par exemple des animaux d'une même souche ou d'un même sexe ou du même âge ou encore élevés dans une même cage ... etc, mais avec plus ou moins de succès dans la diminution de l'erreur expérimentale.

Une organisation particulière de l'expérimentation est celle où il est **possible d'appliquer à chaque unité expérimentale tous les traitements étudiés dans l'expérimentation**, on dit qu'il s'agit d'un **plan de mesures répétées**.

Le plan de mesures répétées le plus simple est celui où il n'y a que 2 traitements (étude d'un facteur à 2 niveaux). Les résultats (les réponses) se présenteront sous forme de 2 séries dont les valeurs se correspondent 2 à 2 et l'analyse statistique s'effectue par un **test t de Student** particulier dit « **pour séries appariées** ». Notons que ce type de planification est généralisable au cas où l'expérimentation comporte plus de 2 traitements.

La diminution de l'erreur expérimentale lors de l'analyse statistique résulte du fait qu'il est possible de défalquer de celle-ci la variabilité due à l'hétérogénéité de comportement des diverses unités expérimentales.

Ce même principe est applicable dans le cas où on souhaite **tenir compte de facteurs d'hétérogénéité liés aux conditions de réalisation des essais** ; par exemple lorsque les essais sont réalisés dans 2 réacteurs différents et n'ayant peut être pas le même comportement. L'organisation des essais doit être spécifique et porte le nom de :

méthode des blocs (prise en compte d'un facteur d'hétérogénéité)

méthode des carrés latins (prise en compte de 2 facteurs d'hétérogénéité)

Plan de mesures répétées

Exemple : Numérisation globulaire mesurée 3 fois sur les mêmes sujets (valeurs divisées par 100 000)

Sujet	Janvier	Juin	Décembre
1	46	48	52
2	38	47	50
3	42	44	48
4	47	45	47
5	48	51	47
6	40	44	47
7	43	47	45
8	43	48	48
9	42	47	46
....

On teste l'effet du facteur « période » (3 niveaux)

Si l'on avait utilisé 3 groupes de sujets différents, la variabilité individuelle se serait ajoutée à celle (éventuelle) due au facteur.

Plan de mesures répétées

Exemple : On veut comparer 3 traitements A, B, C (facteur principal) en tenant compte de 2 facteurs secondaires à 3 modalités chacun, sujet et date d'administration.

	<i>date 1</i>	<i>date 2</i>	<i>date 3</i>
sujet 1	A	B	C
sujet 2	B	C	A
sujet 3	C	A	B

Disposition en « carré latin »

Traitement statistique : Analyse de variance à 2 facteurs croisés

L'opération de randomisation

Randomisation signifie répartition au hasard.

Dans les expériences biologiques, l'affectation des unités expérimentales aux divers traitements doit être effectuée par **tirage au sort**. Par ailleurs, l'ordre d'exécution des essais doit également être effectué par tirage au sort.

Cette opération a, certes, pour effet de compliquer le protocole de manipulation, mais elle est importante à réaliser car c'est elle qui assure la comparabilité des groupes de résultats en évitant des « biais de la réponse » lors du déroulement de l'expérimentation.

II - Expériences comportant l'étude d'un seul facteur

II - Expériences comportant l'étude d'un seul facteur



II - 1 : Plan à 1 seul facteur à 2 niveaux sans répétition

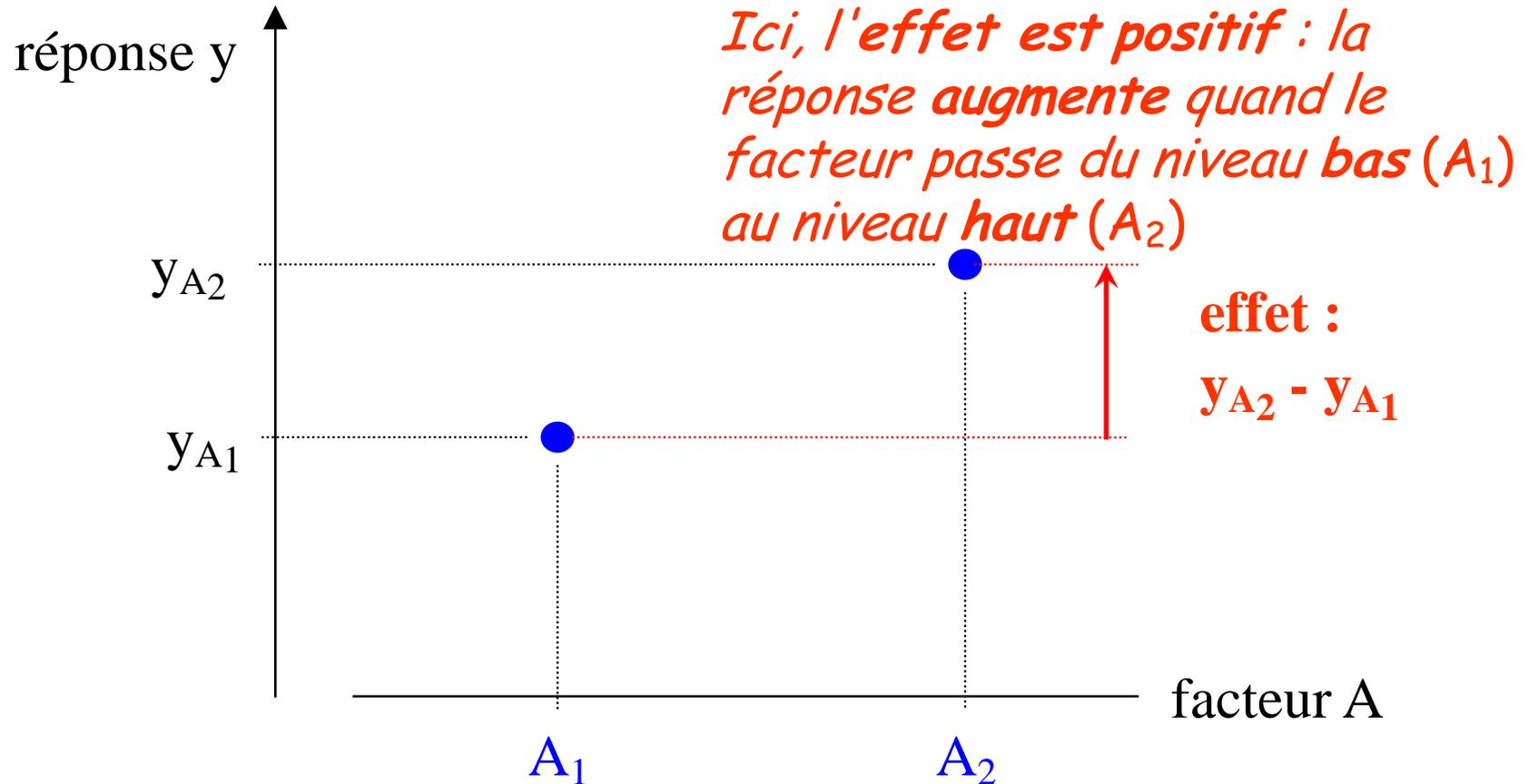
II - 2 : Plan à 1 seul facteur à 2 niveaux comportant des répétitions

II - 3 : Plan à 1 seul facteur comportant plus de 2 niveaux

II - 1 : Plan à 1 seul facteur à 2 niveaux sans répétition

- ◆ Dans ce type de plan expérimental, il n'y a que 2 essais à effectuer : l'un au niveau A_1 (réponse mesurée y_{A_1}) et l'autre au niveau A_2 (réponse y_{A_2}).
- ◆ Selon qu'il s'agit d'un **problème de comparaison** (facteur A qualitatif et parfois quantitatif) ou d'un **problème de régression** (facteur A quantitatif), la méthode d'analyse des résultats est différente.

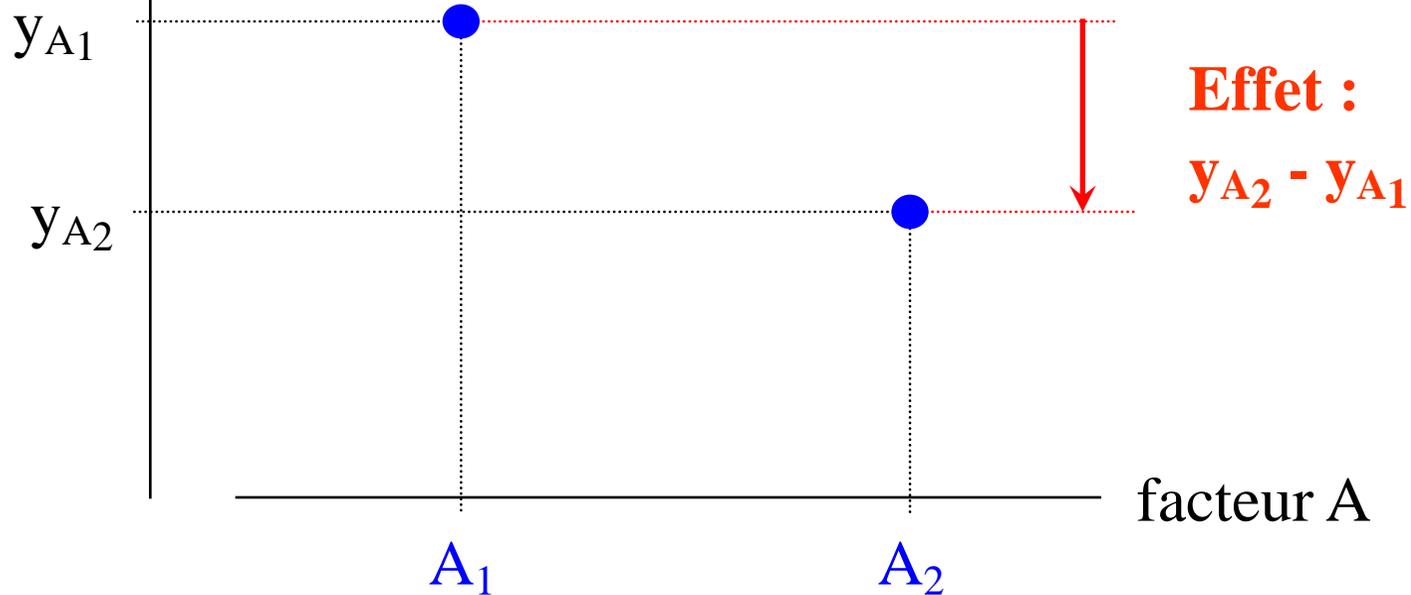
Dans un problème de comparaison, l'effet du changement de niveau du facteur est traduit par une éventuelle variation de la réponse



exemple : nombre de colonies pour 2 milieux de culture (facteur qualitatif)

réponse y

Il est des cas où l'effet est négatif : la réponse diminue quand le facteur passe du niveau bas (A_1) au niveau haut (A_2)

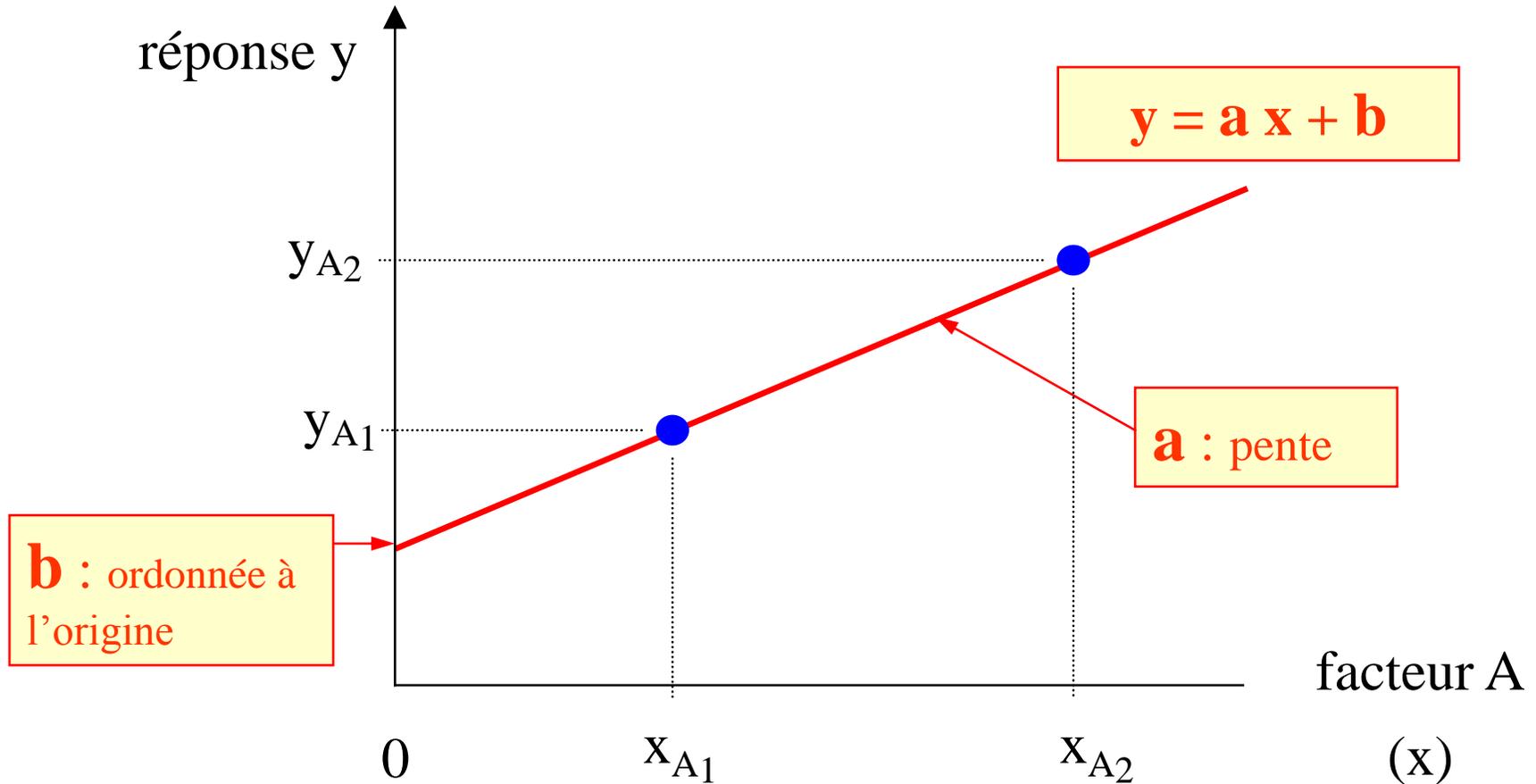


En pratique, estimer la valeur (positive ou négative) d'un effet ne constitue qu'une partie de l'analyse.

En effet, la détermination expérimentale des 2 réponses est entachée d'erreur aléatoire et si on ne connaît pas un ordre de grandeur de celle-ci, **il est impossible de savoir si l'effet obtenu diffère réellement de 0** ; en d'autres termes, si cet effet existe vraiment et n'est pas le reflet de l'imprécision ...

Or, avec seulement 1 détermination par niveau, il n'est pas possible d'estimer la dispersion. Il faut des renseignements complémentaires sur la variabilité des réponses pour pouvoir effectuer l'analyse statistique.

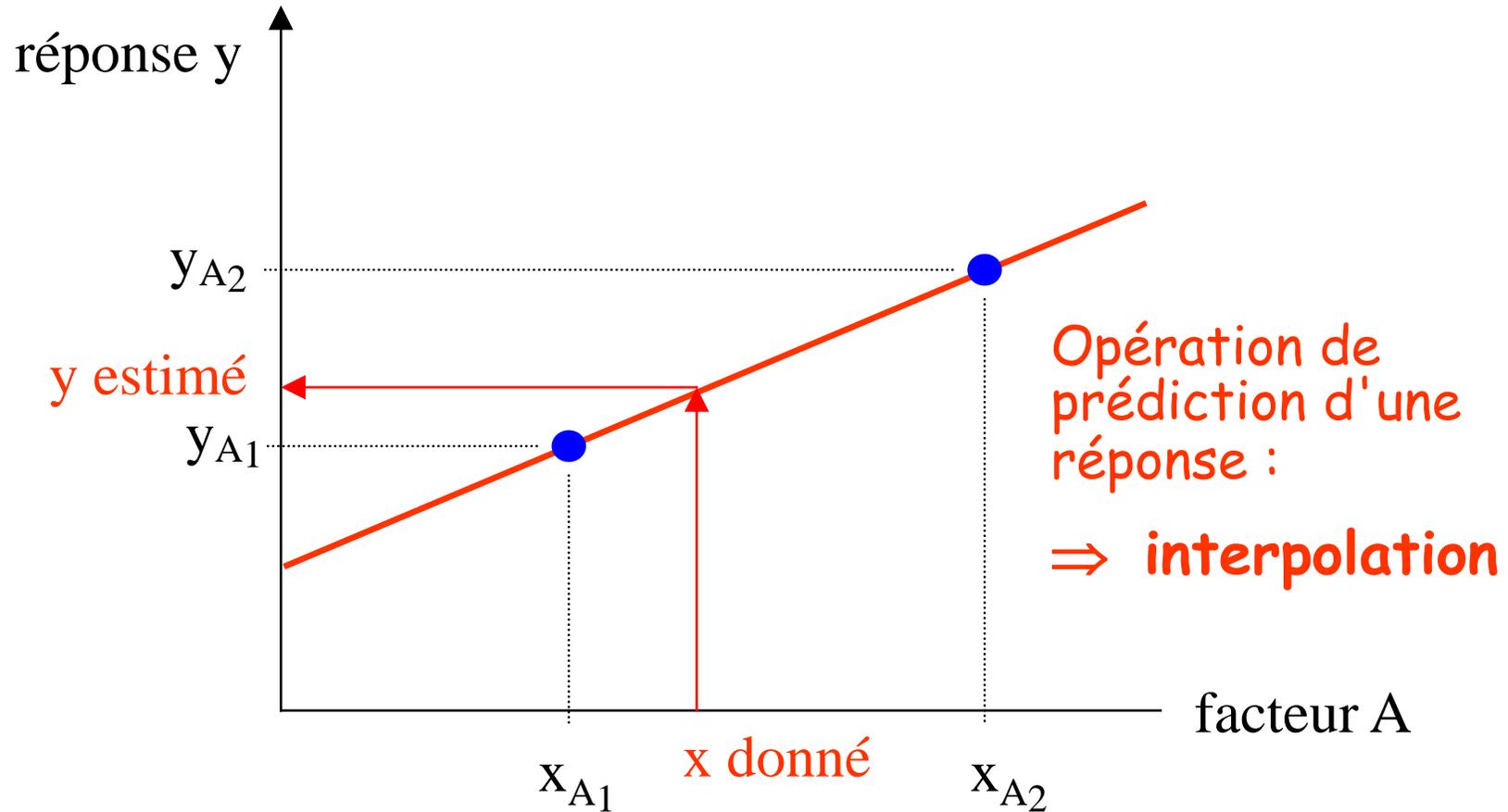
Dans un problème de régression, c'est l'équation décrivant la variation de y en fonction de x qui est intéressante : le plus simple est une variation linéaire. La droite passe exactement par les 2 points.



exemple : densité optique (réponse) et concentration (**facteur quantitatif**)

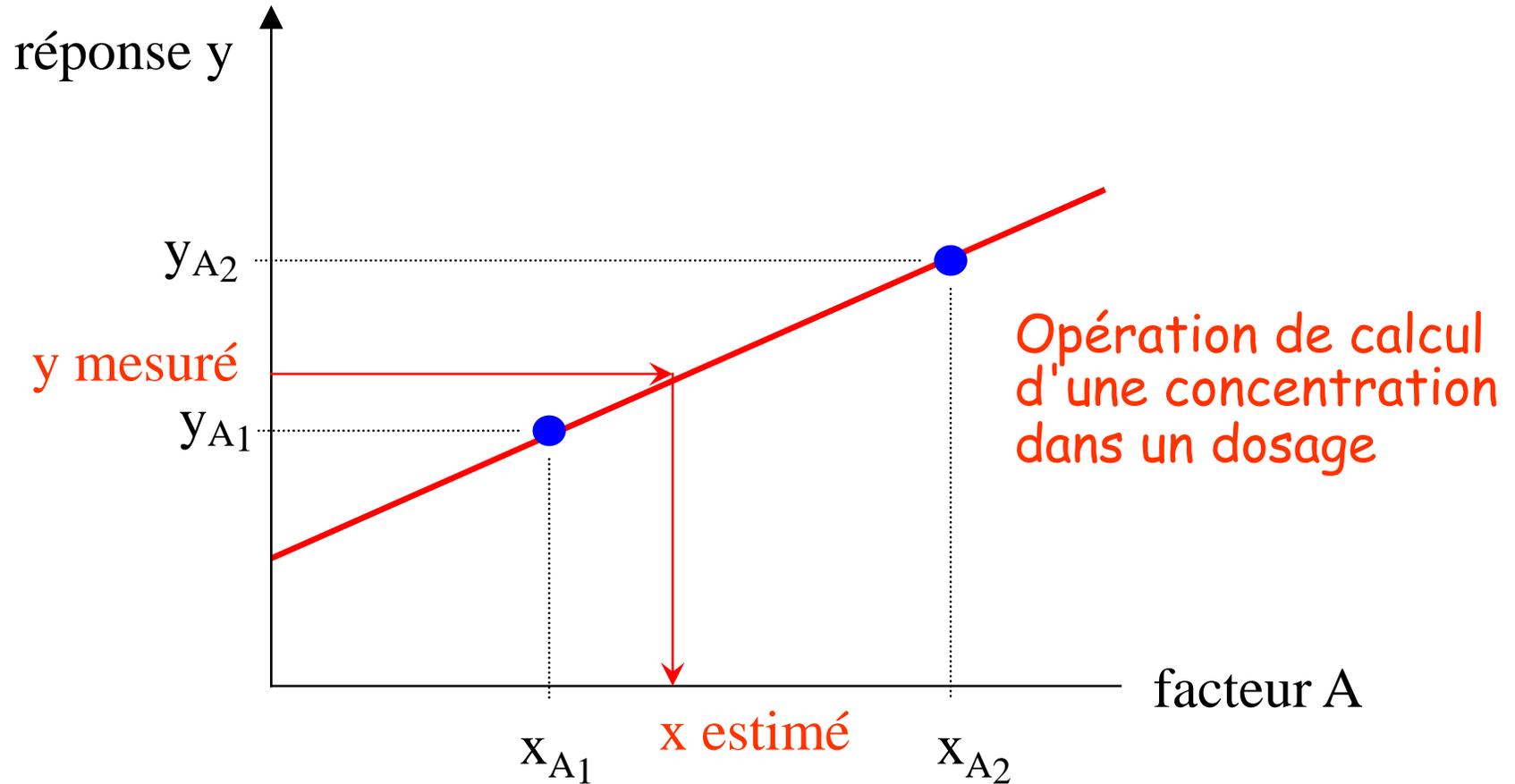
Les 2 utilisations possibles de la droite :

1) - estimer y pour une valeur x donnée



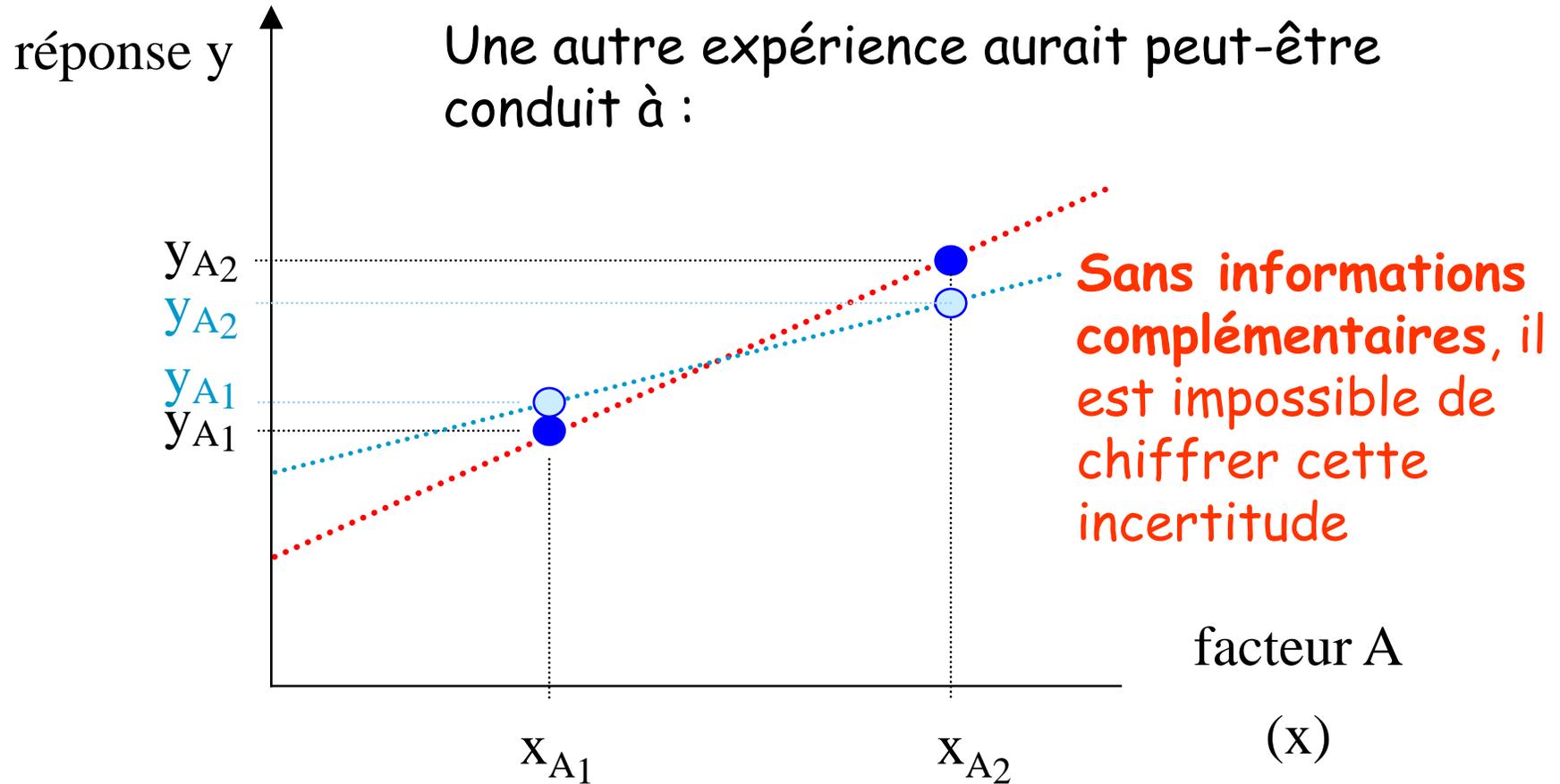
Les 2 utilisations de la droite :

- 1) - estimer y pour une valeur x donnée
- 2) - estimer x pour une valeur y mesurée



Attention !

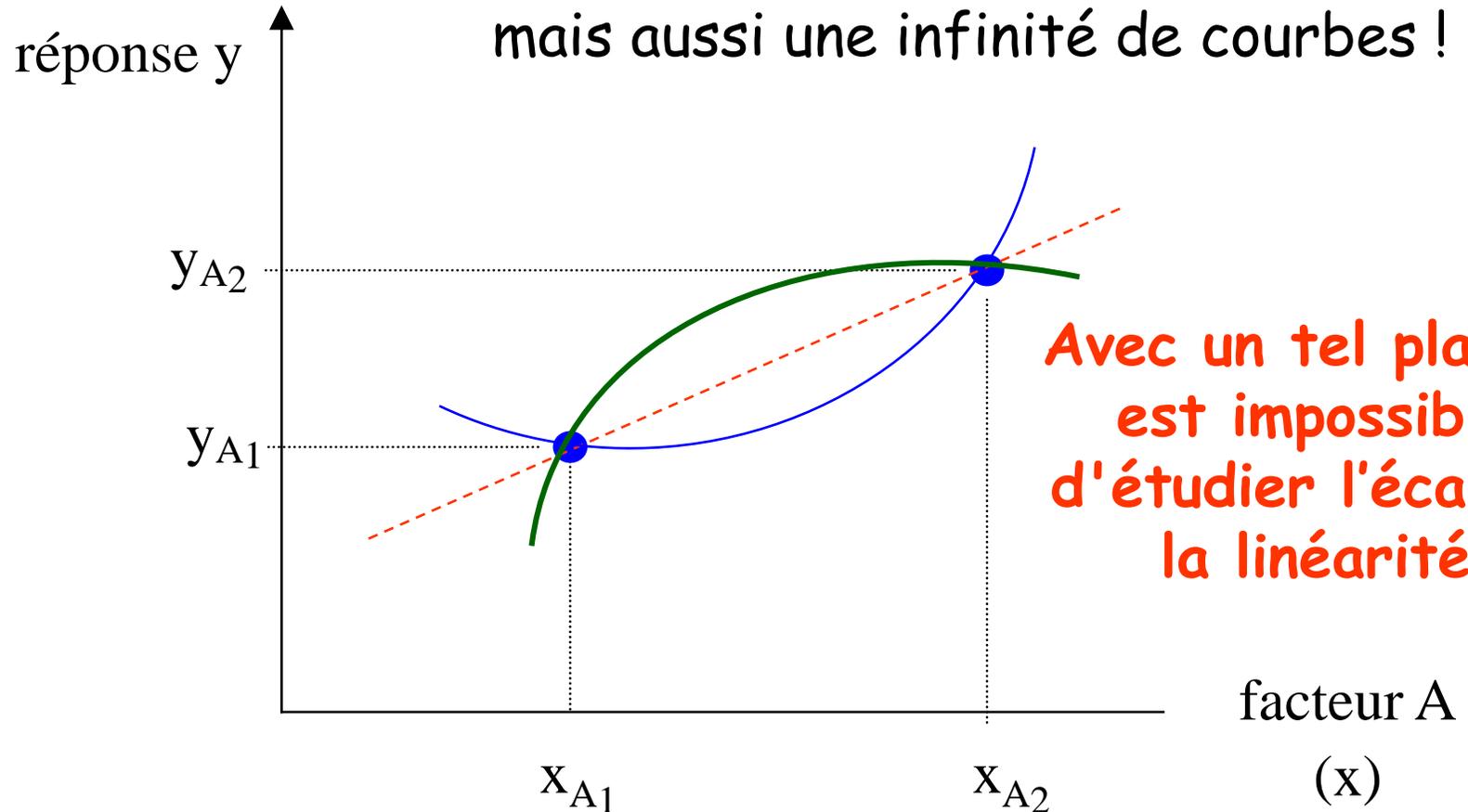
La position de la droite est incertaine à cause des erreurs aléatoires sur les déterminations de y ...



Un autre problème !

Le modèle du 1er degré en x est-il pertinent ?

Par 2 points, il passe une droite et une seule ...
mais aussi une infinité de courbes !



Avec un tel plan, il est impossible d'étudier l'écart à la linéarité

II - Expériences comportant l'étude d'un seul facteur

II - 1 : Plan à 1 seul facteur à 2 niveaux sans répétition



II - 2 : Plan à 1 seul facteur à 2 niveaux comportant des répétitions

II - 3 : Plan à 1 seul facteur comportant plus de 2 niveaux

II - 2 : Plan à 1 seul facteur à 2 niveaux comportant des répétitions

On parle de **répétitions** quand il est pratiqué plusieurs essais pour chacun des niveaux du facteur ; *attention de ne pas confondre répétition de la détermination avec répétition de la mesure seulement ... !*

Effectuer si possible le même nombre de répétitions pour A_1 et A_2

Le tableau de résultats permet de calculer les moyennes \bar{y}_{A1} et \bar{y}_{A2} qui sont les meilleures estimations de la réponse de chaque niveau.

	A_1	A_2
	y_{11}	y_{21}
Réponses Individuelles	y_{12}	y_{22}
(n déterminations)

	y_{1n}	y_{2n}
Moyennes :	\bar{y}_{A1}	\bar{y}_{A2}

L'utilité des répétitions est double :

- ◆ elles permettent d'obtenir une **réponse** à chaque niveau du facteur **plus précise**, puisqu'il s'agit d'une **moyenne**
- ◆ elles permettent de quantifier la dispersion des résultats dans les 2 groupes par l'intermédiaire des écarts types ; on peut alors appliquer le **test de comparaison de 2 moyennes (test t de Student)** qui permet de savoir si l'effet diffère statistiquement de 0.

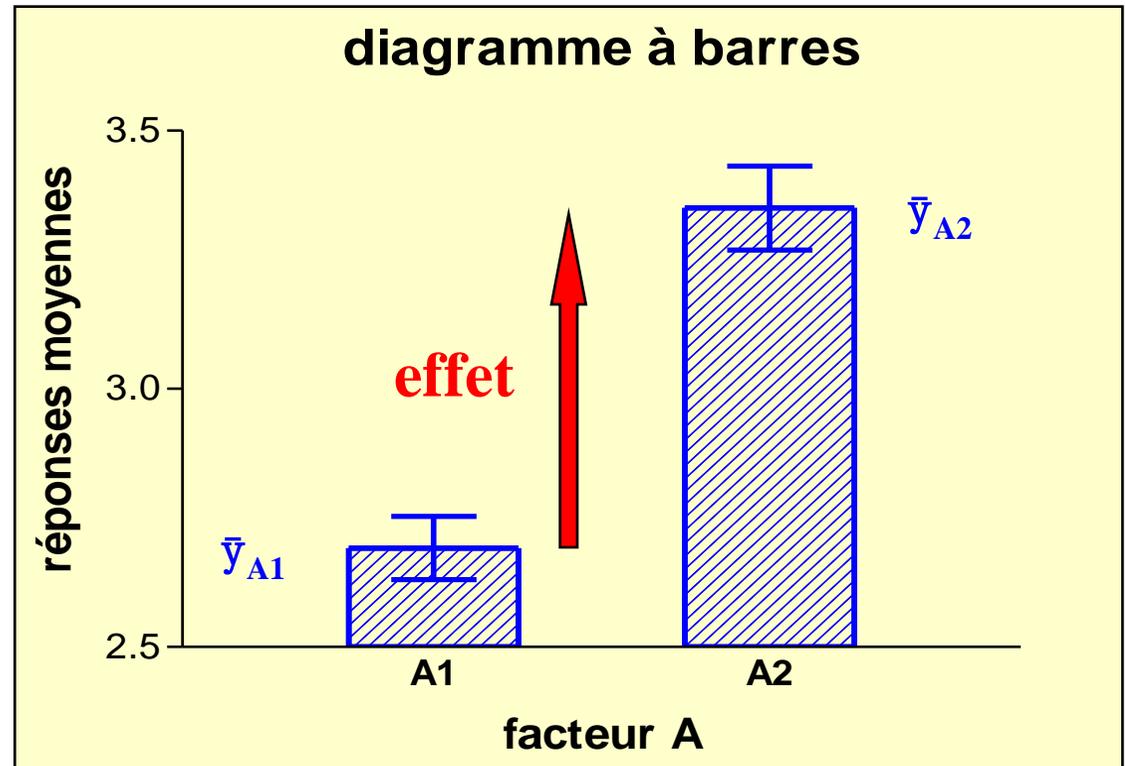
Ce test suppose que des conditions de validité soient remplies (normalité et homogénéité des variances) : il est donc utile d'examiner préalablement la dispersion des résultats obtenus.

Le diagramme à barres est un graphe synthétique montrant les différences entre les moyennes des groupes ; il est habituel d'y faire figurer les barres d'erreur, soit les écarts types sur les moyennes (SEM), soit les intervalles de confiance (à 95%)

C'est la représentation la plus utilisée pour les **facteurs qualitatifs**.

on voit immédiatement la différence de réponse entre les niveaux et on peut la comparer de visu aux barres d'erreur.

Généralement on indique le degré de signification du test de comparaison.



Dans un problème de régression, supposée du 1er degré en x (droite), $y = a x + b$

les paramètres a et b doivent être déterminés par la méthode des moindres carrés en utilisant l'ensemble des réponses individuelles des 2 groupes. La droite passe exactement par les 2 points correspondants aux moyennes des réponses.

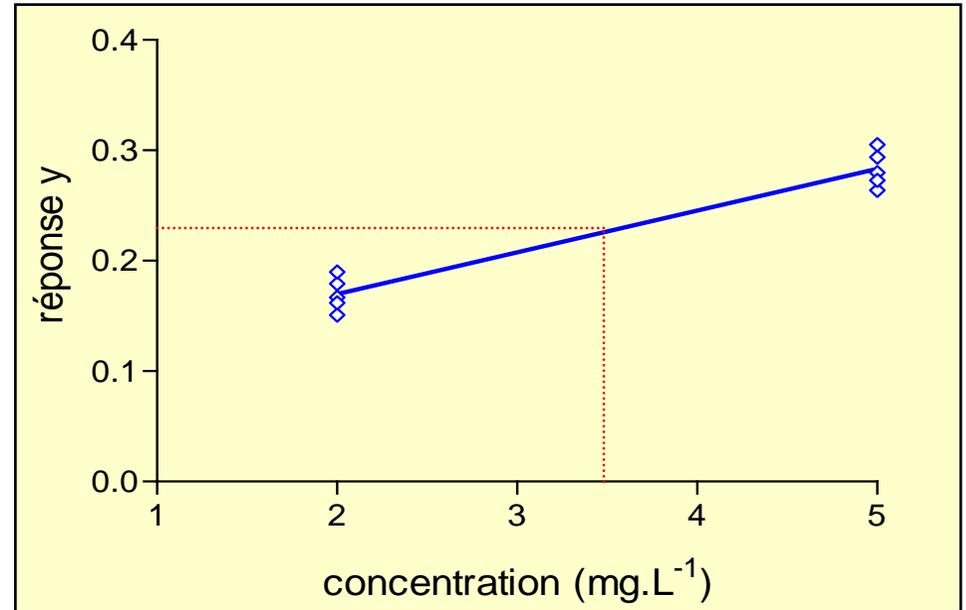
Le fait de disposer de répétitions permet d'estimer l'incertitude pesant sur la position de la droite ; la qualité de l'ajustement des points à la droite se traduit par l'écart type résiduel s_y et le coefficient de détermination R^2 .

On peut aussi calculer les écarts types s_A et s_B des paramètres a et b, ce qui permet de les comparer à des valeurs théoriques par des tests t classiques ; par exemple comparer la pente a à 0 permet de savoir si la réponse y varie effectivement entre x_{A1} et x_{A2} .

exemple

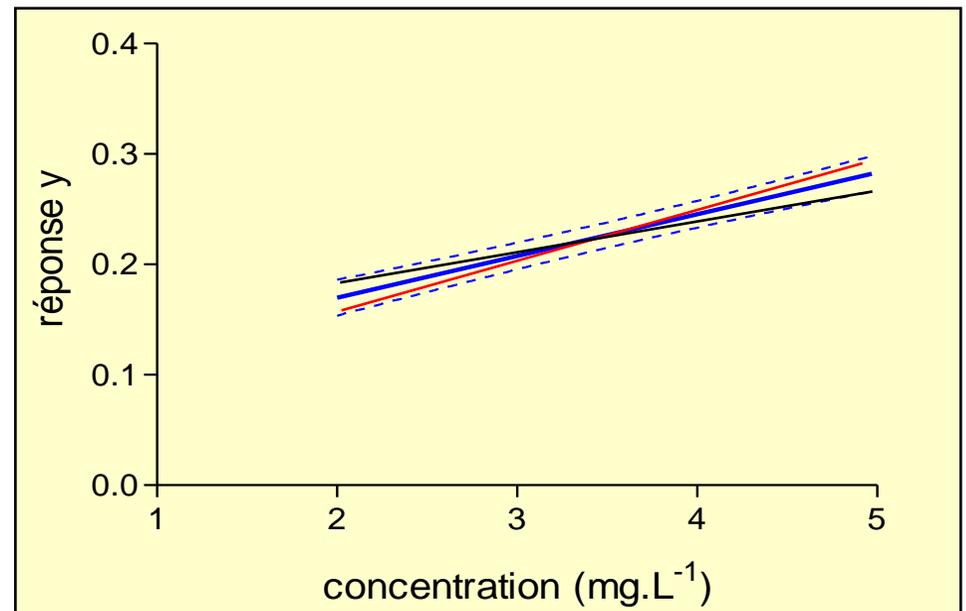
Pour $C = 2 \text{ mg.L}^{-1}$ et $C = 5 \text{ mg.L}^{-1}$, 5 répétitions dont les moyennes sont $\bar{y}(2) = 0,170$ et $\bar{y}(5) = 0,283$. La droite passe par ces 2 points moyens.

La pente a pour valeur $0,0378 \pm 0,003$ elle diffère significativement de 0. Le coefficient R^2 est égal à 0,9417



Cette droite n'est qu'une estimation. On peut visualiser l'incertitude sur sa position en traçant les enveloppes de part et d'autre de la ligne obtenue. Il s'agit d'hyperboles figurées ici en pointillé.

A titre d'illustration, 2 autres droites qui auraient pu être obtenues, car compatibles avec la dispersion obtenue.



La notion de facteur centré réduit

Lors de l'analyse des résultats, il est souvent intéressant de noter les niveaux A_1 et A_2 par les nombres -1 et $+1$

- ◆ Dans le cas d'un facteur qualitatif, ces 2 nombres ne sont que des codes ; toute autre valeur n'a pas de signification.
- ◆ Pour un facteur quantitatif, cela revient à changer d'origine et d'échelle ; entre les niveaux x du facteur et les valeurs X du facteur centré réduit, il existe la correspondance :

$$X = \frac{x - \bar{x}}{\Delta x} \quad \text{avec} \quad \Delta x = x_{A2} - \bar{x}$$

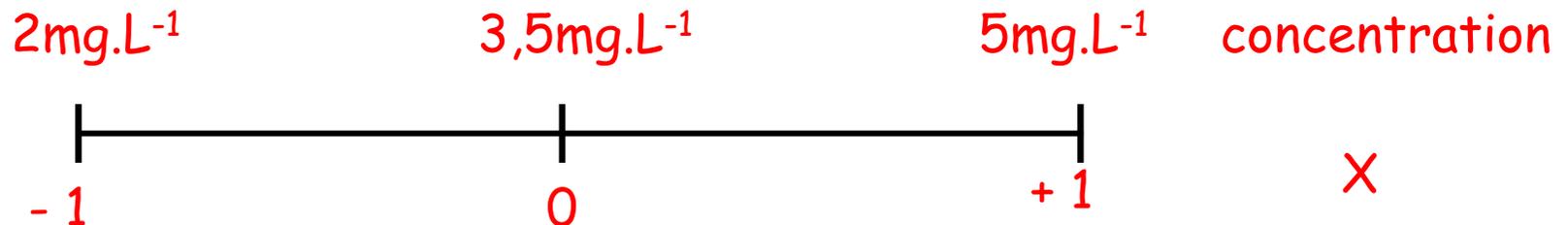
A titre d'illustration :

lors d'un étalonnage d'une méthode de dosage, deux concentrations ont été utilisées : $C_1 = 2\text{mg.L}^{-1}$ et $C_2 = 5\text{mg.L}^{-1}$

Le domaine expérimental s'étend entre 2 et 5mg.L^{-1}

Le centre du domaine est $C_0 = 3,5\text{mg.L}^{-1}$

Le pas de variation est $\Delta C = C_2 - C_0 = 1,5\text{mg.L}^{-1}$



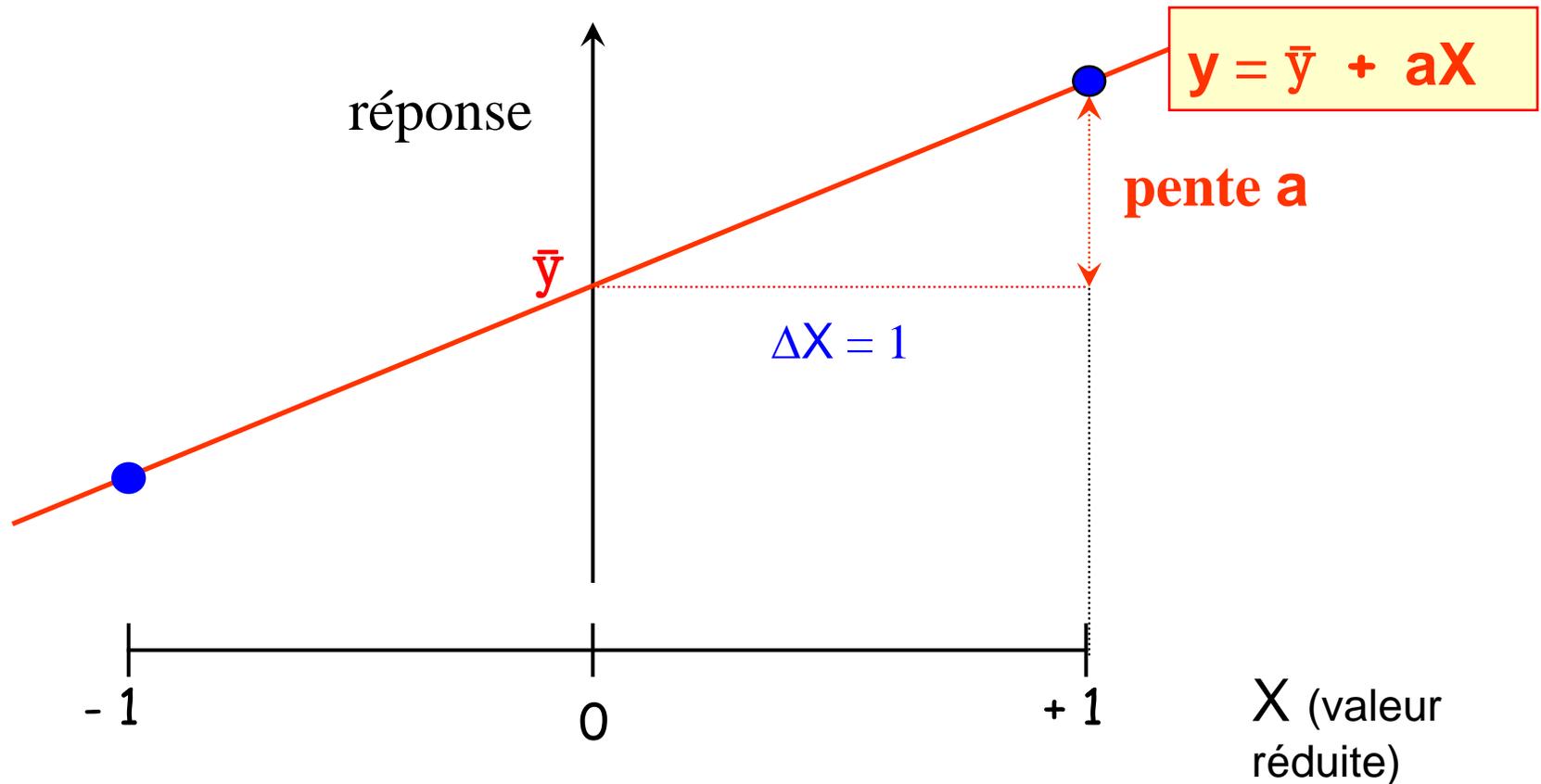
La correspondance entre X et C s'écrit :

$$X = (C - 3,5) / 1,5 \quad \text{ou} \quad C = 3,5 + 1,5 X$$

Pour $C = 4,0 \text{ mg.L}^{-1}$ on a $X = + 0,33$

pour $C = 2,5 \text{ mg.L}^{-1}$ on a $X = - 0,67$

Pour une **régression supposée linéaire**, avec cette nouvelle variable X , l'ordonnée à l'origine qui correspond à $X = 0$ devient \bar{y} moyenne générale des réponses pour les 2 niveaux X_{A_1} (-1) et X_{A_2} (+1). L'équation s'écrit :



L'effet d'un niveau

Soit \bar{y} la moyenne générale de toutes les réponses mesurées.

Une autre façon de caractériser l'effet d'un **facteur qualitatif** consiste à définir **un effet par niveau** en rapportant la réponse de ce niveau à la moyenne générale \bar{y} .

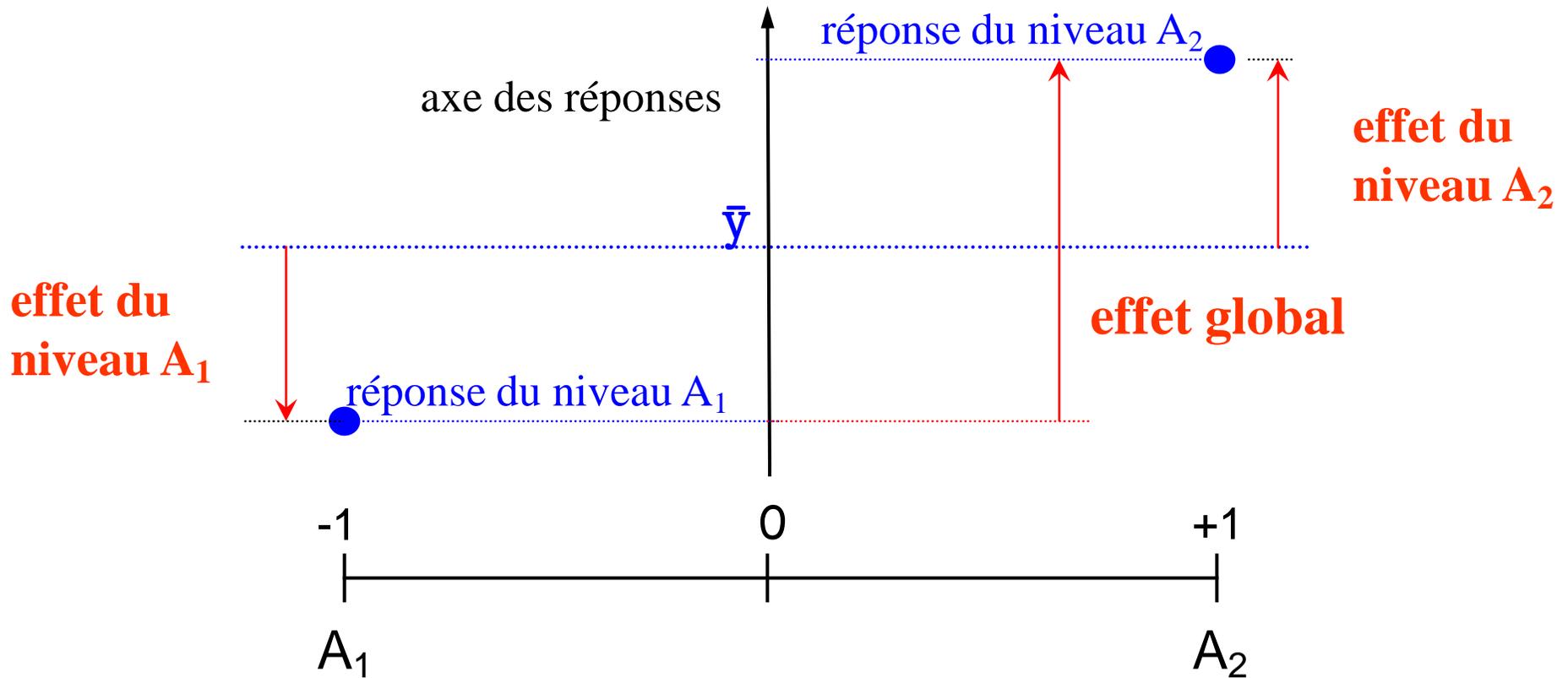
Ainsi pour un facteur à 2 niveaux A_1 et A_2 , il est défini 2 effets :

l'effet de $A_1 = \mathbf{E}_{A_1} =$ réponse moyenne de $A_1 - \bar{y}$

l'effet de $A_2 = \mathbf{E}_{A_2} =$ réponse moyenne de $A_2 - \bar{y}$

Le schéma suivant montre que ces 2 effets sont opposés et égaux en valeur absolue : $E_{A_2} + E_{A_1} = 0$

$E_{A_2} = - E_{A_1}$ est appelé **effet moyen** ; il vaut la moitié de l'effet global précédemment défini.



Quel intérêt de définir un effet par niveau ?

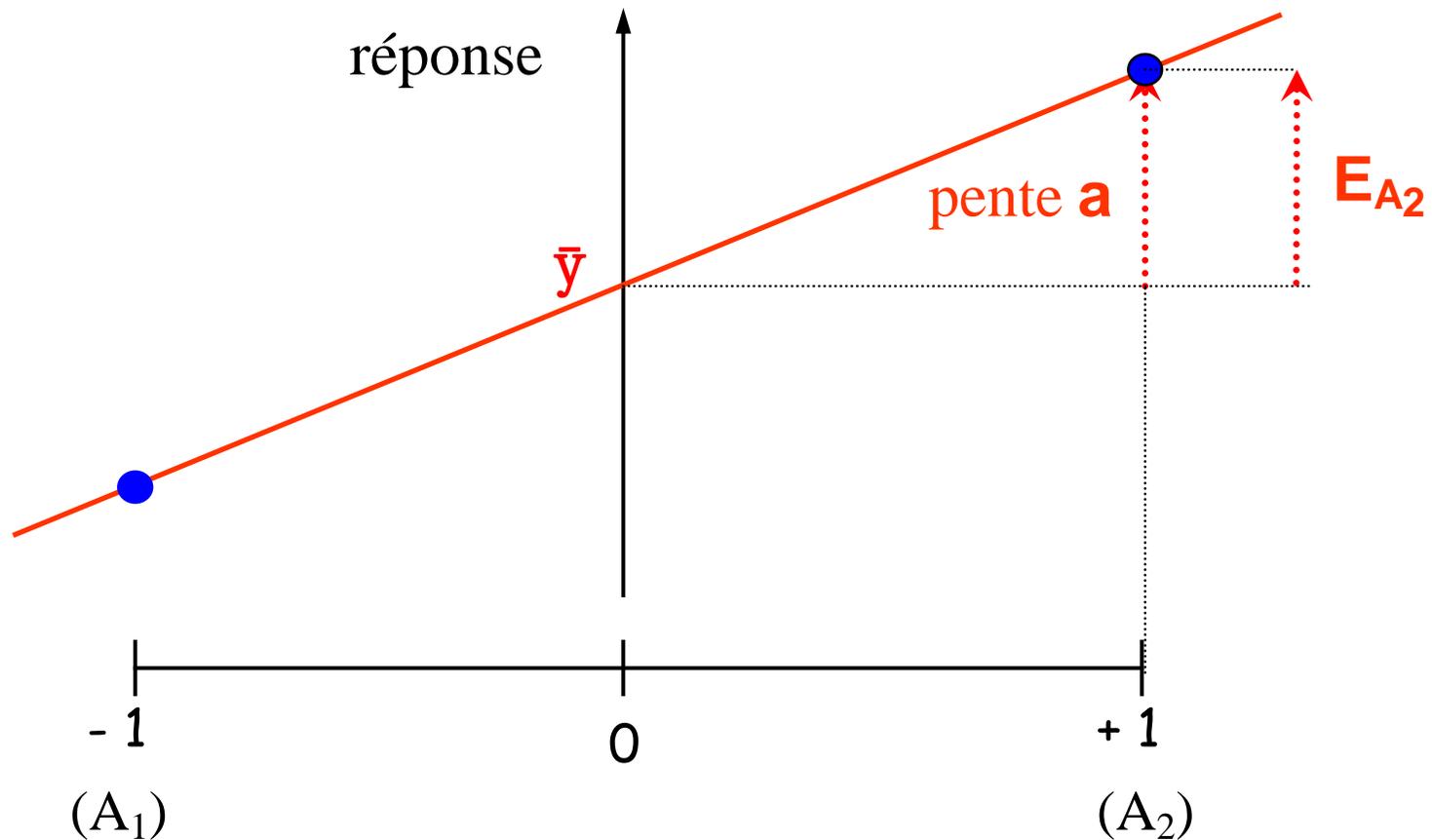
Nous verrons ultérieurement que ce concept « un effet par niveau » est surtout intéressant quand un facteur qualitatif possède 3 niveaux ou plus.

Dans le cas de 2 niveaux, on peut y voir 2 intérêts :

1) - pour un facteur qualitatif, la réponse y estimée à un niveau (A_1 ou A_2) est liée à l'effet par la relation :

$$y = \bar{y} + \text{effet } (E_{A_1} \text{ ou } E_{A_2})$$

2) - la valeur de la pente **a** de la droite de régression est identique à celle de l'effet moyen E_{A_2} : il n'y a pas nécessité de faire la distinction entre facteur qualitatif ou quantitatif



II - Expériences comportant l'étude d'un seul facteur

II - 1 : Plan à 1 seul facteur à 2 niveaux sans répétition

II - 2 : Plan à 1 seul facteur à 2 niveaux comportant des répétitions



II - 3 : Plan à 1 seul facteur comportant plus de 2 niveaux

II - 3 : Plan à 1 seul facteur comportant plus de 2 niveaux

Il est des cas où 2 niveaux du facteur sont insuffisants pour obtenir dans une même série d'essais tous les renseignements souhaités par l'expérimentateur.

Par exemple, dans une expérience de comparaison traitement/témoin, il arrive qu'on soit obligé de prévoir 2 témoins (un témoin négatif et un témoin positif ^(*)), ce dernier étant destiné à vérifier que le système expérimental fonctionne correctement : le facteur étudié comporte alors 3 niveaux (Tr , Te⁻ , Te⁺).

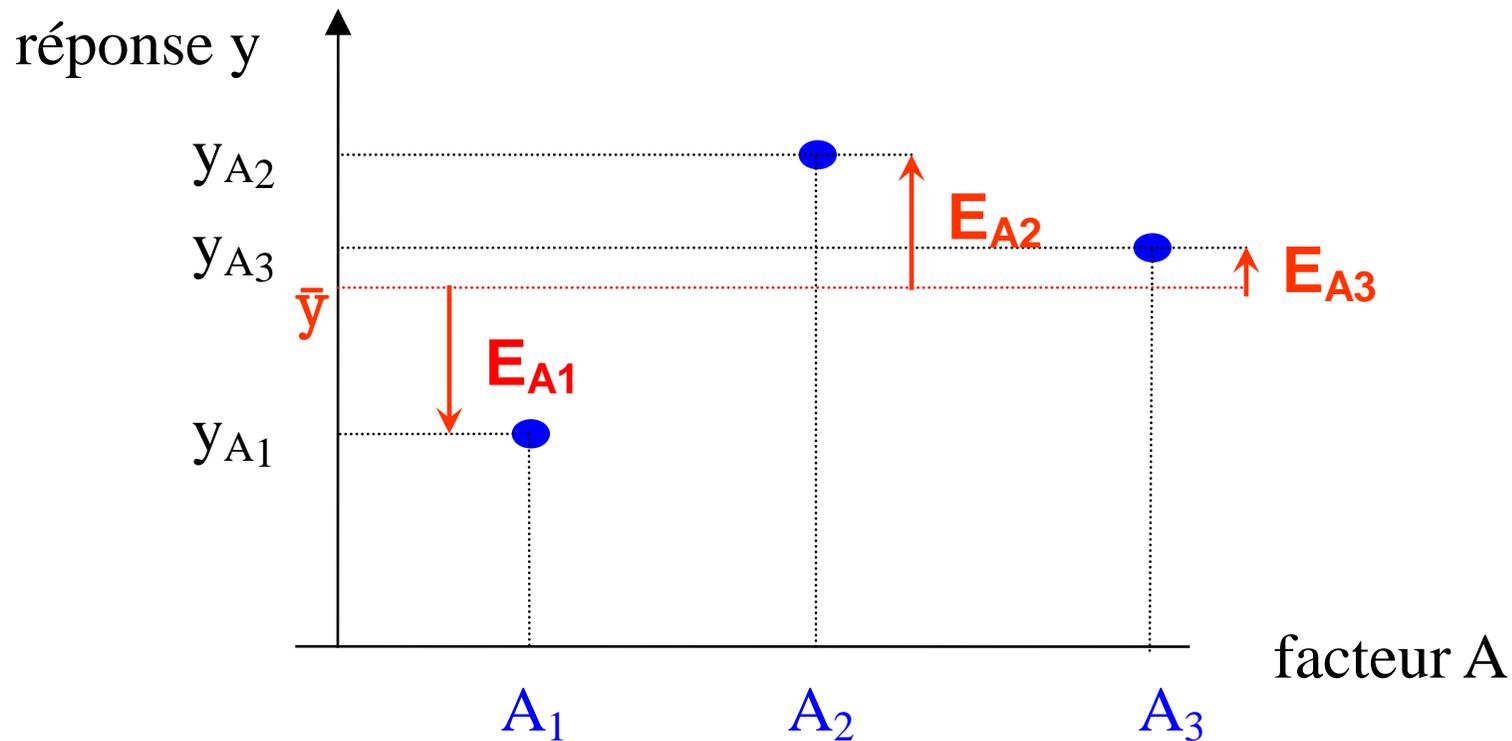
Nous nous limiterons ici à ce cas de 3 niveaux, en indiquant que la démarche est identique quand il y a 4 niveaux et plus.

(*) test cutané pour dire si un produit est allergène ou non - témoin positif : allergène connu, témoin négatif : solution sans allergène

Le plan ne comporte pas de répétition

- ◆ Il n'y a que 3 essais à effectuer : l'un au niveau A_1 (réponse mesurée y_{A_1}), un autre au niveau A_2 (réponse y_{A_2}) et un troisième au niveau A_3 (réponse y_{A_3})
- ◆ La méthode d'analyse diffère selon qu'il s'agit d'un **problème de comparaison** (facteur A qualitatif et parfois quantitatif) ou d'un **problème de régression** (facteur A quantitatif).

Dans un problème de comparaison, il est possible de calculer 3 effets E_{A1} , E_{A2} et E_{A3} par rapport à la moyenne \bar{y} des 3 réponses ; ils ne sont pas tous indépendants, puisque $E_{A1} + E_{A2} + E_{A3} = 0$ par construction.



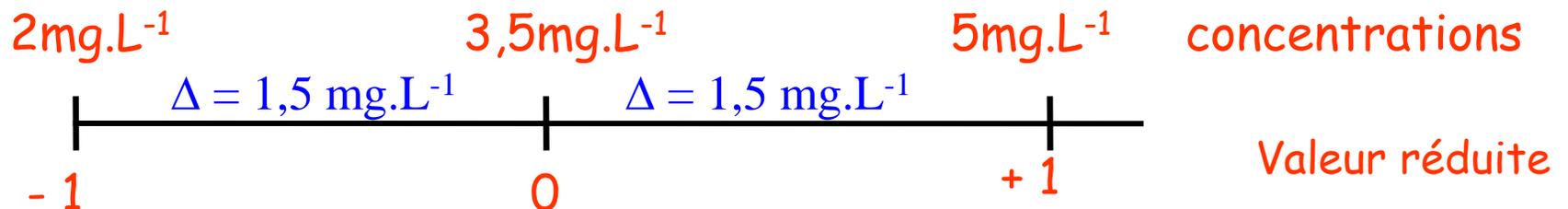
Mais ... **il est impossible de savoir si ces effets diffèrent significativement de 0** (autrement dit si les 3 réponses y_{A_1} , y_{A_2} et y_{A_3} diffèrent réellement entre elles), car ces réponses sont entachées d'erreurs aléatoires dues à la manipulation et à la mesure ...

Il est impératif de prévoir des répétitions pour chaque niveau du facteur étudié pour avoir une idée de la grandeur de la dispersion de la réponse (**erreur expérimentale**) ce qui permettra de la confronter aux valeurs des effets.

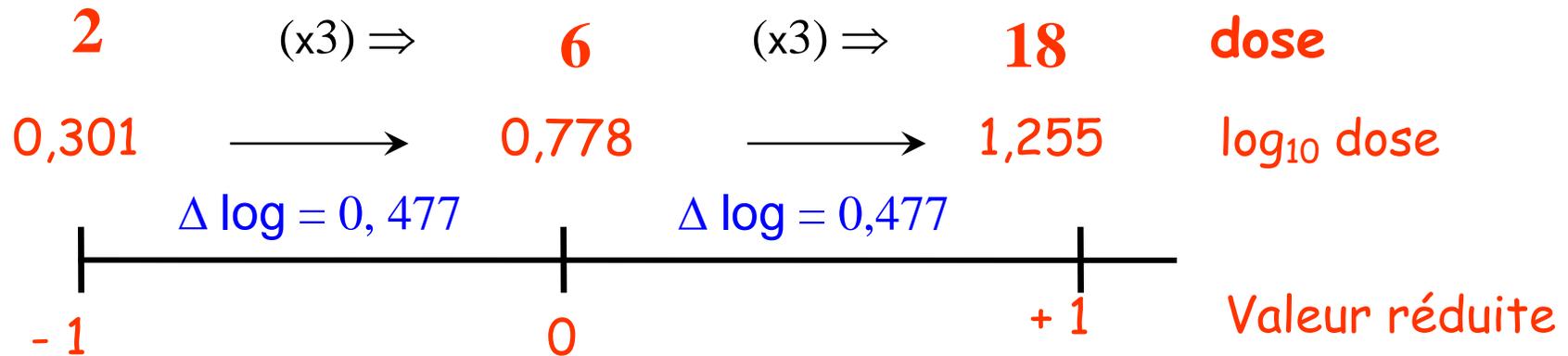
Dans un problème de régression, pour lequel le facteur étudié est quantitatif (variable x), il est judicieux de prévoir des niveaux régulièrement espacés pour obtenir une relation linéaire avec la réponse.

Par exemple,

◆ dans un étalonnage où s'applique la loi de Beer-Lambert (linéarité entre l'absorbance et la concentration), on prendra 3 concentrations en progression arithmétique, afin que la concentration médiane soit au milieu de l'intervalle :



◆ Dans les dosages biologiques où il est fréquent que l'effet soit proportionnel au logarithme de la dose, on prendra 3 doses espacées en progression géométrique :



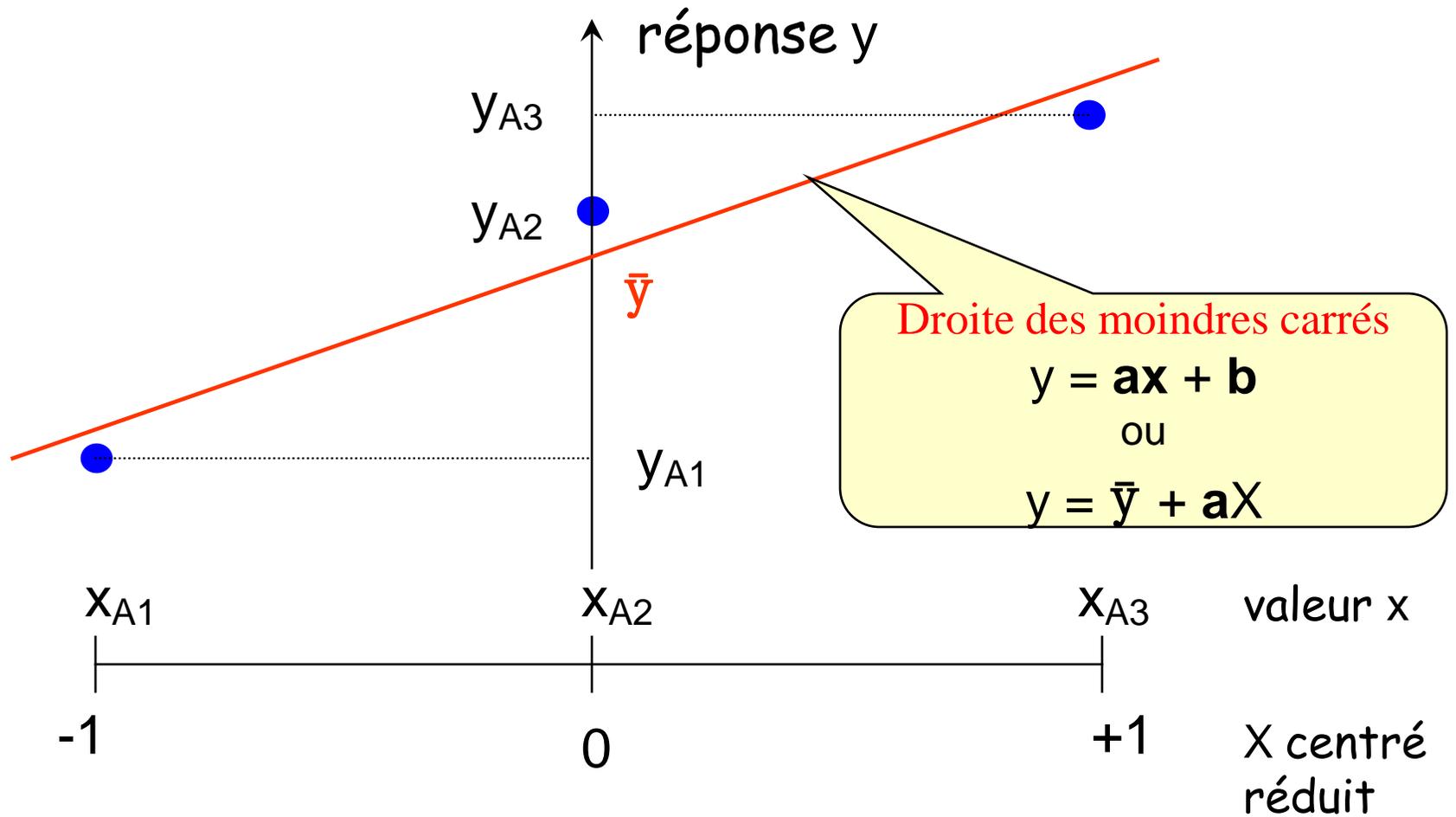
C'est avec des doses régulièrement espacées que la régression linéaire fournit les informations statistiques les plus efficaces.

$$\log 6 = \log (2 \times 3) = \log 2 + \log 3 = 0,301 + 0,477$$

Quels renseignements peut-on obtenir par l'étude de la régression ?

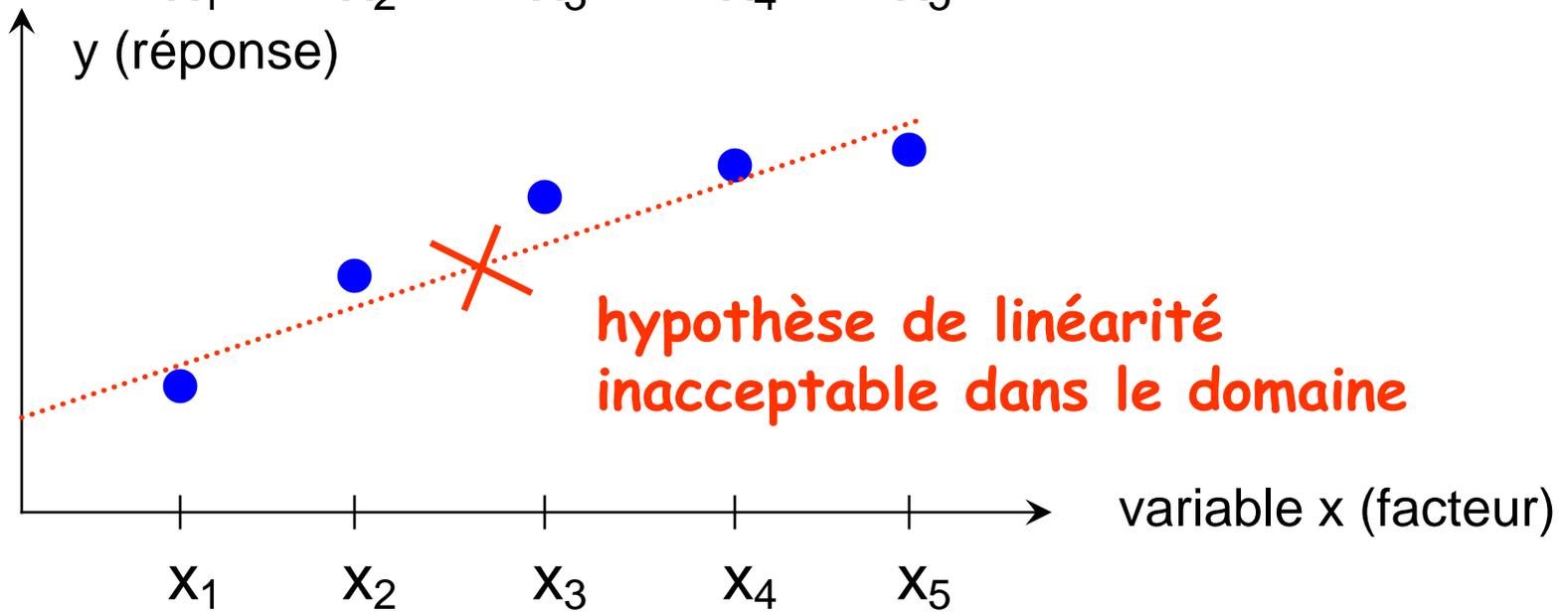
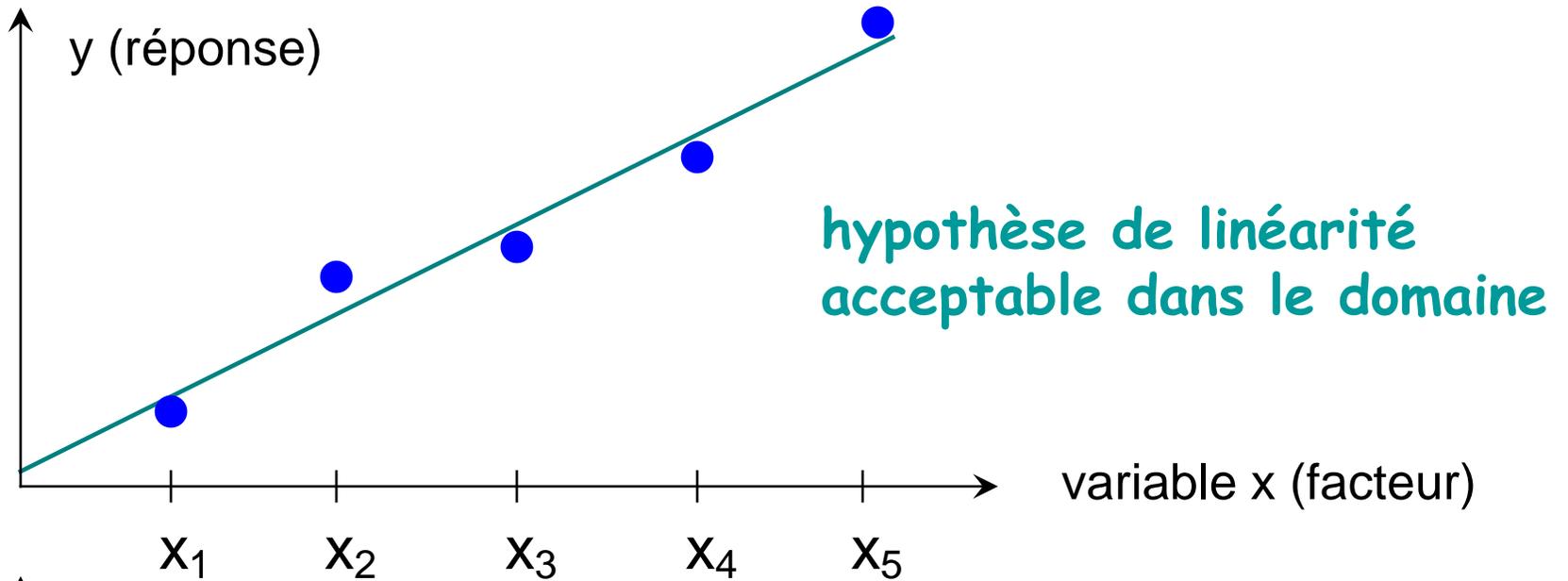
Même si l'hypothèse de linéarité de la relation entre la réponse et la variable x contrôlée (le facteur) est réelle, les 3 points ne sont généralement pas exactement alignés à cause des erreurs aléatoires entachant les déterminations des réponses.

La méthode des moindres carrés permet d'obtenir une équation numérique du 1er degré en x qui correspond à l'estimation de la position de la droite la plus proche des 3 points.



Avec seulement 3 points, il est possible d'effectuer une analyse statistique de la régression. Les valeurs numériques des estimations des 2 coefficients, ordonnée à l'origine (**b**) et pente (**a**), peuvent être enrichis par un intervalle de confiance ; les écarts verticaux entre les points expérimentaux et la droite se traduisent par le coefficient de détermination **R²**, nombre compris entre les limites 0 et 1 et d'autant plus proche de 1 que les points sont plus proches de la droite (les chimistes utilisent plus volontiers comme paramètre sa racine carrée **R**).

Cependant, il est évident qu'avec un plan comportant 4 , 5 (voire plus) valeurs de la variable contrôlée x - **régulièrement espacées** - l'estimation de la position de la droite (**a** et **b**) est meilleure : il y a plus d'information apportée. Il est surtout plus facile de juger sur le graphique la linéarité (absence de courbure systématique dans la disposition des points).



Car l'hypothèse de linéarité de la relation $y = f(x)$ ne peut pas être testée statistiquement si le plan expérimental ne comporte pas de répétitions : les écarts verticaux entre les points expérimentaux et la droite de régression peuvent en effet provenir de 2 causes : l'imprécision des déterminations de la réponse y (l'erreur expérimentale) d'une part, un choix erroné du type de relation (linéaire) entre y et x d'autre part et il est impossible de séparer la part qui revient à chacune de ces 2 causes ...

Globalement l'ensemble des écarts des points à la droite de régression, appelé **erreur résiduelle**, est quantifié par **l'écart type résiduel** lors de l'analyse statistique.

En pratique, un tel plan est utilisé couramment lors de l'**étalonnage des méthodes analytiques en routine** (à la condition qu'une étude préalable lors de la **validation de la méthode** ait montré que l'étalonnage est bien linéaire).

L'utilisation de la droite pour calculer des concentrations de solutions inconnues nécessite que les points soient très proches de la droite, donc que le R^2 soit très proche de 1 (les analystes choisissent en général $R^2 \geq 0,999$) ; la méthode doit être suffisamment précise, ce qui est en général réalisé en chimie analytique mais pas pour les dosages biologiques ...

Quant au nombre de concentrations d'étalon à choisir (les niveaux du facteur) il résulte d'un compromis entre coût financier, durée de manipulation et confiance dans les résultats : prendre 5 concentrations semble raisonnable, mais il est des cas où on n'en prendra que 3, voire 2 . Ex : CPG , HPLC ...

Le plan comporte des répétitions

- ◆ Le plan expérimental comprend **3 séries d'essais à effectuer** : l'une au niveau A_1 , une autre au niveau A_2 et la troisième au niveau A_3 du facteur. Le tableau de résultats permet de calculer les moyennes \bar{y}_{A1} , \bar{y}_{A2} et \bar{y}_{A3} qui représentent la réponse de chaque niveau.
- ◆ Il est très souhaitable que les 3 séries comportent le même nombre de déterminations, pour des raisons évidentes d'équilibre de l'information apportée par les 3 moyennes et aussi à cause des conditions d'application de certains tests statistiques utilisés dans l'analyse (égalité des effectifs).

Dans une étude à visée comparative, il est possible grâce aux moyennes des réponses des 3 séries de calculer 3 effets E_{A1} , E_{A2} et E_{A3} par rapport à la moyenne générale \bar{y} de l'ensemble des réponses individuelles ; ces effets ne sont pas indépendants, puisqu'on a déjà vu que $E_{A1} + E_{A2} + E_{A3} = 0$.

L'information supplémentaire apportée par les expériences avec répétitions est que l'on peut estimer la dispersion des réponses à l'intérieur de chaque série : en combinant ces 3 estimations (sous réserve que les variances des 3 séries soient égales, donc qu'il y ait homogénéité des variances), on obtient **l'erreur expérimentale** qui traduit globalement l'imprécision des réponses dans l'expérimentation. Il est alors possible de comparer les valeurs des effets à cette erreur expérimentale grâce aux tests statistiques.

Du point de vue tests statistiques :

Ce qu'il est incorrect de faire !

Comparer les 3 moyennes 2 par 2 à l'aide de tests t de Student ...

Ce qu'il faut faire :

1) - effectuer une analyse de variance à un facteur (une ANOVA dans le langage des spécialistes) qui permet de savoir si **globalement, au moins 1 effet diffère significativement de 0** (autrement dit si parmi les 3 moyennes, au moins une diffère réellement des autres). Lorsque ce n'est pas le cas, l'analyse est terminée : les niveaux du facteur n'agissent pas différemment sur la réponse.

Dans le cas contraire, l'ANOVA se révèle significative, l'expérimentateur souhaite en général comparer les moyennes des séries en fonction des objectifs particuliers de son étude ... Cela est réalisé dans une 2ème étape de l'analyse au moyen de tests a posteriori

2) - tests a posteriori

Les tests à utiliser dépendent de la façon dont a été organisée l'expérience. 2 cas sont fréquents en pratique :

- ◆ une des 3 séries est un « groupe témoin » et on souhaite comparer les 2 autres groupes à ce groupe de référence. Il faut utiliser **le test de Dunnett**.
- ◆ Toutes les comparaisons peuvent se révéler utiles ; on utilisera alors **le test de Newman-Keuls**

Dans une étude de régression, un intérêt de répéter les déterminations pour chaque valeur de x est qu'on peut utiliser la dispersion dans chaque groupe, exactement comme dans une ANOVA, pour estimer **l'erreur expérimentale** (à la condition que les séries de réponses aient même variance) ; en la comparant à **l'erreur résiduelle** de la régression linéaire, on peut ainsi **tester statistiquement l'écart à la linéarité** (si une relation du 1er degré entre y et x convient ou non).

Une telle étude est toujours **effectuée lors de la validation d'une méthode analytique** : elle nécessite de choisir un minimum de 3 concentrations d'étalon (régulièrement espacées). Le nombre de répétitions par concentration dépend de l'imprécision de la méthode (souvent voisin de 5) ; Le but est de s'assurer qu'on pourra utiliser une relation linéaire lors des étalonnages en routine.

La répétition des concentrations (ou des doses) est aussi nécessaire quand l'imprécision des réponses individuelles est grande : on utilise alors comme réponse pour chaque valeur de x , la moyenne \bar{y} des réponses individuelles du groupe.

En procédant ainsi, la droite est estimée avec d'avantage de précision (le coefficient R^2 est plus proche de 1 que si on utilisait l'ensemble des réponses individuelles dans la régression).

C'est la méthode utilisée, même en routine lors de l'étalonnage dans les dosages biologiques : la variabilité et la diversité de comportement du matériel vivant sont telles qu'il est indispensable de procéder ainsi.

III - Expériences comportant l'étude de 2 facteurs et plus

Il peut arriver que l'objectif de l'expérimentation nécessite d'étudier **les effets de plusieurs facteurs**, notés **A , B , C ...** sur **une même réponse y**.

Ces facteurs étudiés peuvent être qualitatifs ou quantitatifs. Les problèmes posés peuvent être comparatifs ou de régression.

Comme précédemment, les niveaux choisis par l'expérimentateur, notés $[A_1 A_2]$ pour le facteur **A** et $[B_1 B_2]$ pour le facteur **B**, peuvent être codés $[-1 , +1]$ lorsqu'il n'y a que 2 niveaux et $[-1 , 0 , +1]$ s'il y en a 3.

Notions d'interaction entre 2 facteurs

L'étude conjointe de 2 facteurs est plus riche en information que celle apportée par 2 études séparées avec chacun des facteurs ...

Considérons le cas où les 2 facteurs A et B sont étudiés chacun à 2 niveaux : A_1 , A_2 et B_1 , B_2 .

L'effet de A sur la réponse y peut être estimé de 2 façons : E_A au niveau B_1 du facteur B et E_A au niveau B_2 ; si ces 2 effets sont égaux (ou voisins par suite de l'inévitable dispersion expérimentale de la réponse), c'est que le facteur B n'a pas d'influence sur l'effet de A.

Réciproquement l'effet du facteur **B** n'est pas non plus influencé par le niveau du facteur **A** dans ce cas.

On dit alors que **les 2 facteurs A et B agissent indépendamment sur la réponse y** ; les effets de **A** et **B** sont simplement additifs (il n'y a pas d'interaction).

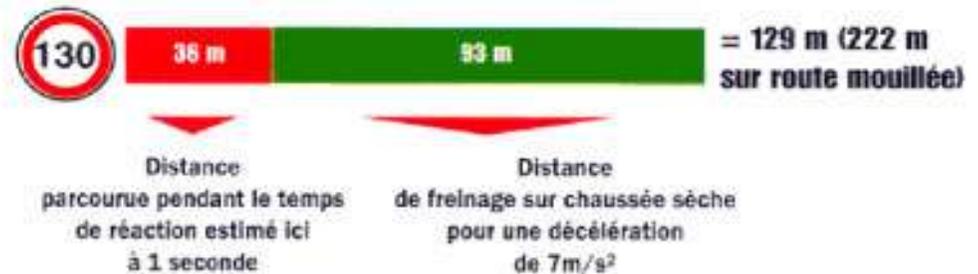
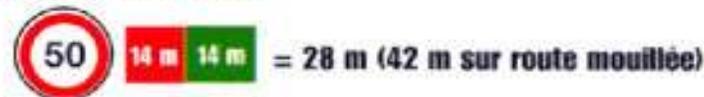
En revanche, lorsque les 2 effets de **A** aux niveaux **B₁** et **B₂** sont significativement différents, **c'est qu'il existe une interaction entre les 2 facteurs**, notée **AB**, dont l'analyse permet de chiffrer la contribution E_{AB} sur la détermination de la réponse **y** ; **l'effet d'une interaction peut être positive** (la réponse est alors plus grande que la simple addition des effets des facteurs) **ou négative** (la réponse est plus petite).

2 illustrations de la notion d'interaction

1) - **exemple emprunté à la sécurité routière** : étude de la distance d'arrêt d'une automobile lors d'un freinage (c'est la réponse y exprimée en mètres) en fonction de l'état de la chaussée (facteur **A** à 2 niveaux : route sèche et route mouillée) et de la vitesse du véhicule (facteur **B** à 2 niveaux : 50 km/h et 130 km/h).

Un extrait d'une documentation de la sécurité routière → → →

- Les **distances d'arrêt augmentent** considérablement avec la vitesse.



Le tableau suivant résume les résultats :

vitesse	Distance d'arrêt (m)	
	Route sèche	Route mouillée
50 km/h	28	42
130 km/h	129	222

L'effet de l'état de la route sur la distance d'arrêt du véhicule est $E_A = 14 \text{ m}$ à 50 km/h et devient $E_A = 93 \text{ m}$ à 130 km/h . Bien que la sécurité routière ne précise pas l'imprécision sur les distances d'arrêt, **il est clair qu'il y a interaction entre la vitesse et l'état de la route**

On peut aussi raisonner sur l'autre facteur : l'effet global de la vitesse (entre 50 et 130 km/h) est $E_B = 101 \text{ m}$ sur route sèche et qu'il devient $E_B = 180 \text{ m}$ sur route mouillée ...

Distance d'arrêt (m)		
vitesse	Route sèche	Route mouillée
50 km/h	28	42
130 km/h	129	222
	+ 101	+ 180

On retrouve la même différence de **79 m** entre les 2 effets du même facteur, valeur qui caractérise l'interaction.

- Effet de l'état de la route dépend de la vitesse : **93** versus **14**
- Effet de la vitesse dépend de l'état de la route : **180** versus **101**

Il y aurait absence d'interaction si les données avaient été les suivantes →→

	Distance d'arrêt (m)	
vitesse	Route sèche	Route mouillée
50 km/h	28	42
130 km/h	129	143

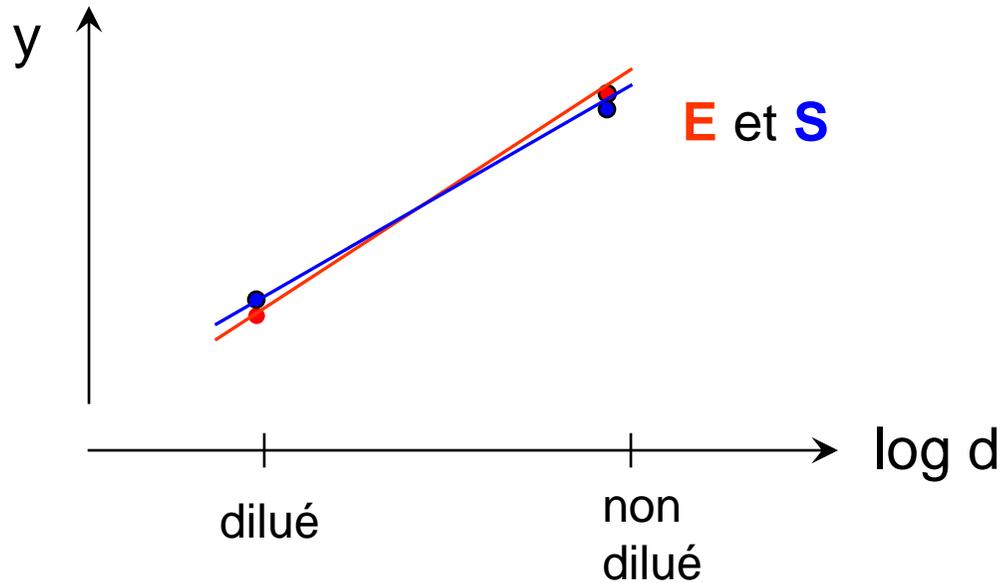
2) - autre exemple : cas d'un dosage biologique

Soit par exemple une préparation d'antibiotique S dont l'activité est évaluée en boîte de Pétri, sur une gélose ensemencée par un germe et comportant des disques de papier imbibé d'un volume déterminé d'antibiotique. La réponse est le diamètre mesuré y (en 1/10 de mm) de la zone d'inhibition autour de chaque disque.

Choisissons de prendre 2 doses de la préparation de S à titrer (non dilué, dilué au 1/2 par exemple); on supposera que la représentation de y en fonction du log de la dilution (variable x) est linéaire. La pente de la droite caractérise la variation d'activité de S quand la dose varie (la sensibilité du germe).

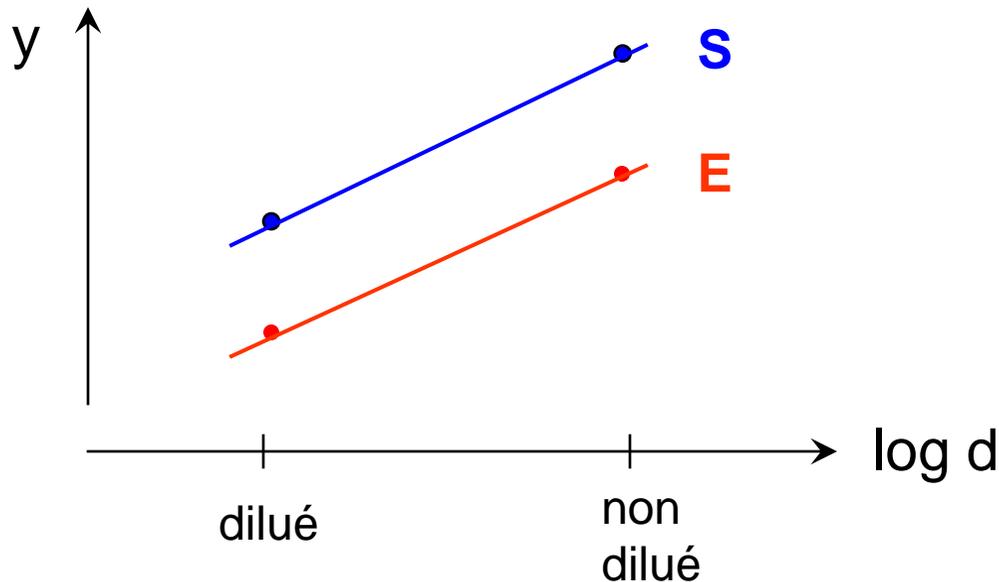
Pour essayer de s'affranchir le plus possible des sources d'hétérogénéité, nombreuses dans ce type d'essai, **on compare ces 2 réponses de S à celles obtenues avec l'étalon E d'antibiotique (de titre connu) dilué de la même façon que S, dans des conditions expérimentales identiques.** On prévoit donc 4 disques par boîte de Pétri, 2 pour S et 2 pour E.

Et, à cause de l'imprécision des déterminations dans la manipulation, **on répète ces déterminations** en utilisant plusieurs boîtes de Pétri (en général entre 5 et 10) ; on obtient ainsi **2 réponses moyennes pour S et 2 pour E**, générant 2 droites que l'on peut comparer et qui permettent d'estimer le titre de S.



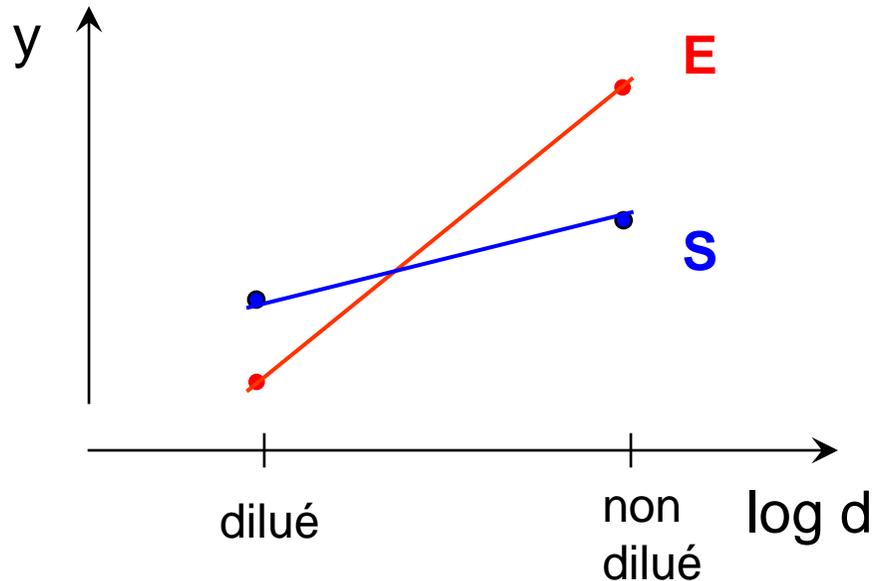
1er cas : les 2 droites sont confondues :

Le titre de S est le même que celui de E



2ème cas : les 2 droites sont parallèles :

L'activité de S est ici supérieure à celle de E et de même nature (le titrage est **valide**) ; il est licite de déduire le titre de S en se référant à E.



3ème cas : les 2 droites ne sont pas parallèles

L'activité de S n'est pas de même nature que celle de E : le titrage n'est pas valide.

Dans ce type de plan, les effets de 2 facteurs chacun à 2 niveaux sont étudiés : l'un, qualitatif, est la nature de l'antibiotique (S et E) et l'autre, quantitatif, est la dose.

L'interaction entre les 2 facteurs est due au fait que la sensibilité du germe aux antibiotiques S et E diffère qualitativement ; cette différence se traduit ici par le fait que les 2 droites de dilution n'ont pas la même pente.

Lorsqu'en revanche, les préparations S et E ont le même type d'action sur la souche incorporée dans la gélose, mais pas la même teneur (**différence seulement quantitative**), les 2 droites de dilution seront parallèles : **le parallélisme est un critère de validité important à vérifier.**

Du point de vue analyse statistique, un tel dosage, noté 2×2 , présente des avantages : les répétitions de chaque dose pour chaque préparation permettent de tester l'hypothèse d'**homogénéité des variances** nécessaire en régression et de tester l'**écart au parallélisme** (l'interaction), car on conçoit qu'en pratique les droites ne sont jamais exactement parallèles à cause de la dispersion expérimentale. La seule chose qu'on ne peut pas tester ici est la linéarité des 2 droites de dilution (il n'y a que 2 doses !)

C'est pourquoi les dosages biologiques les plus utilisés sont les essais 3×3 où **3 doses espacées en progression géométrique** (dilutions) sont utilisées ; on teste en plus l'**écart à la linéarité**.

Planifier une expérience factorielle

Un plan comportant **l'étude simultanée de plusieurs facteurs croisés** dans une même expérimentation est appelé **plan factoriel**.

Son élaboration se fait en combinant tous les niveaux des facteurs : **toutes les combinaisons doivent être présentes** et déterminent autant d'essais ou de groupes d'essais (lorsqu'il est prévu des répétitions) appelés **traitements**.

Les 2 plans précédents (sécurité routière et dosage biologique) sont des plans factoriels croisant 2 facteurs à 2 niveaux et il y a **4 traitements** (A_1B_1 , A_2B_1 , A_1B_2 , A_2B_2).

Quand on décide d'utiliser 3 doses dans l'essai biologique de comparaison de S et E, le plan est noté 2×3 - le facteur A comporte deux niveaux et B trois - il faut prévoir **6 traitements** (A_1B_1 , A_2B_1 , A_1B_2 , A_2B_2 , A_1B_3 , A_2B_3).

Comme en biologie on effectue toujours des répétitions, le nombre d'expériences à prévoir est égal au produit du nombre de traitements par le nombre r de répétitions par traitement. **Il est important de prévoir le même nombre r de répétitions pour chaque traitement.**

Quelques mots sur l'analyse statistique de ces plans

L'exploitation statistique des résultats des expériences factorielles est réalisée au moyen **d'une analyse de variance à 2 facteurs** pour une expérience comparative : elle permet de savoir si chacun des facteurs a un effet significatif et s'il existe ou non une interaction entre ceux-ci.

Dans un dosage biologique où l'intérêt est centré sur la régression, la technique d'exploitation adéquate est **l'analyse de variance de la régression**. Voisine de la précédente, elle est plus complexe à mettre en œuvre.

Attention aux traitements choisis !

Un plan factoriel correspond à des conditions expérimentales bien définies ...

Par exemple, dans une étude des influences du pH et de la concentration d'un réactif dans une réaction de coloration (la réponse est l'absorbance), il a été décidé d'expérimenter le pH entre 7,0 et 9,0 et la concentration entre 1 mg.L^{-1} et 3 mg.L^{-1} .

2 expérimentateurs ont planifié pour cette étude 4 expériences différentes ... et qui n'apportent pas la même information.

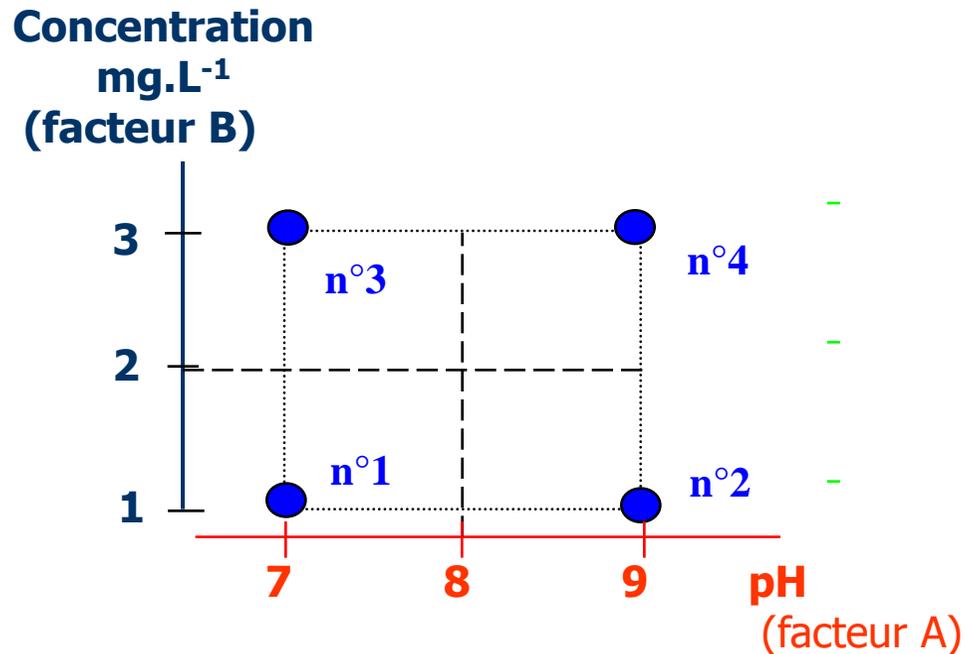
Expérimentateur 1

- ◆ essai n°1 : pH = 7,0 et $C = 1 \text{ mg.L}^{-1}$
- ◆ essai n°2 : pH = 9,0 et $C = 1 \text{ mg.L}^{-1}$
- ◆ essai n°3 : pH = 7,0 et $C = 3 \text{ mg.L}^{-1}$
- ◆ essai n°4 : pH = 9,0 et $C = 3 \text{ mg.L}^{-1}$

Expérimentateur 2

- ◆ essai n°1 : pH = 7,0 et $C = 2 \text{ mg.L}^{-1}$
- ◆ essai n°2 : pH = 9,0 et $C = 2 \text{ mg.L}^{-1}$
- ◆ essai n°3 : pH = 8,0 et $C = 1 \text{ mg.L}^{-1}$
- ◆ essai n°4 : pH = 8,0 et $C = 3 \text{ mg.L}^{-1}$

Plan de l'expérimentateur 1

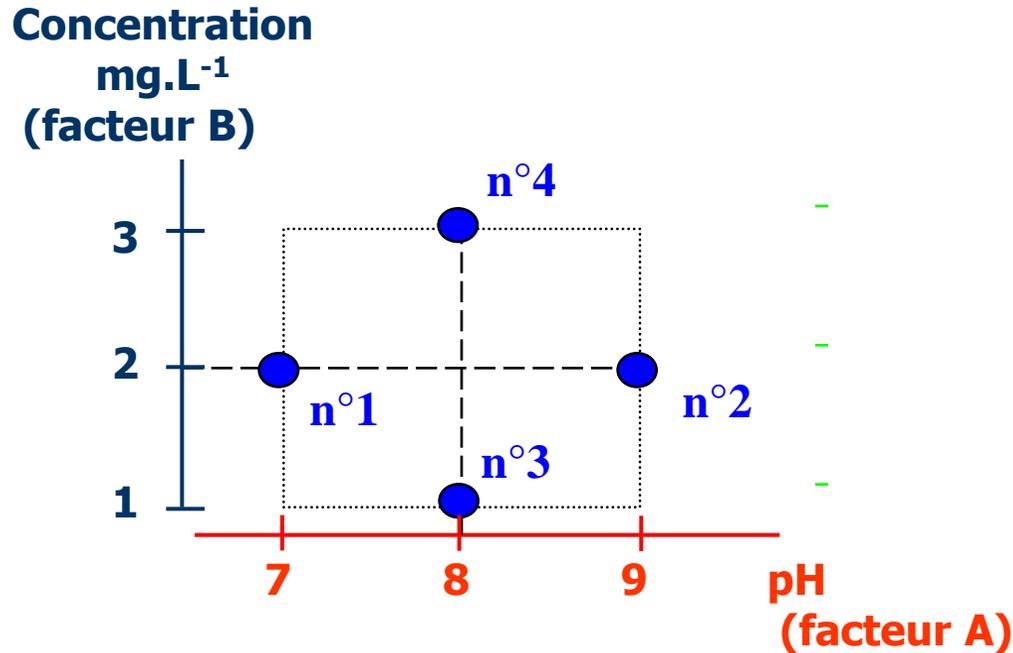


Ces 4 essais correspondent aux combinaisons des 2 niveaux des facteurs : **il s'agit d'un plan factoriel.**

Il est identifiable par le fait que les **4 traitements** sont **aux extrémités du domaine expérimental.**

Ce plan permet d'estimer l'effet du pH et de la concentration sur l'absorbance, ainsi que l'importance de l'interaction [conc, pH].

Plan de l'expérimentateur 2



Les 4 essais ne sont pas aux extrémités du domaine expérimental : **il ne s'agit pas d'un plan factoriel.**

Cette expérimentation a été conçue pour faire varier 1 seul facteur à la fois : les essais 1 et 2 permettent d'obtenir l'effet du pH (à 2 mg.L⁻¹), les essais 3 et 4 l'effet de la concentration (à pH = 8). **Mais on ne peut pas avoir de renseignement sur une éventuelle interaction.**

En conséquence ...

Intuitivement, pour évaluer les effets que peuvent avoir différents facteurs sur une réponse, on a tendance à imaginer de **faire varier ces facteurs un à un en maintenant le niveau des autres constant** ; c'est une méthode prudente mais qui n'est pas efficace, car avec un tel plan il est impossible de mettre en évidence les **interactions** qui peuvent exister entre les facteurs. Or dans de nombreuses études, ces interactions sont loin d'être négligeables ...

Il est très préférable pour étudier plusieurs facteurs de choisir un plan factoriel.