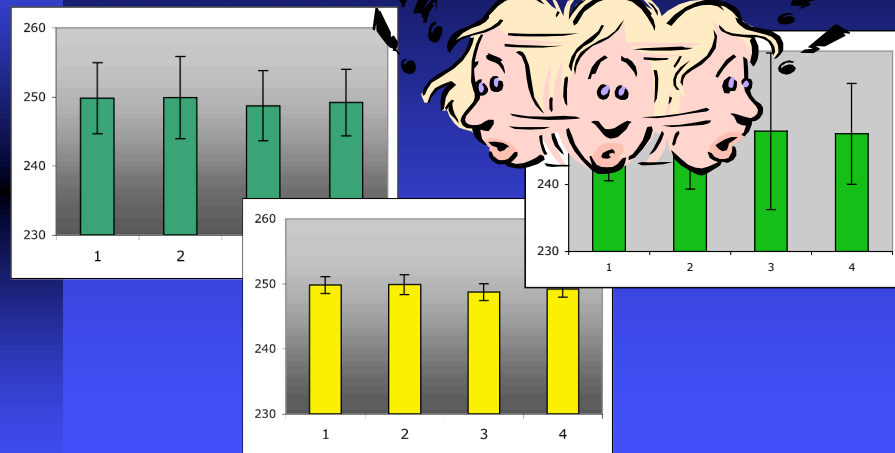


## Les barres d'erreur dans les histogrammes: pas si simple...



27/04/05

Life Science Seminar

1

Les barres d'erreurs sont incontournables dans les communications scientifiques. Cependant la rigueur avec laquelle elles sont élaborées, l'interprétation que l'auteur fait des résultats et les informations qu'il communique au lecteur sur la méthode employée pour les tracer laissent à désirer dans de nombreuses publications. Mais effectivement, ce n'est pas si simple...

## Deux approches

Intervalle de confiance



Test d'hypothèse



27/04/05

Life Science Seminar

2

L'inférence statistique est un ensemble de techniques qui permet de contrôler le risque d'erreur de la généralisation des observations réalisées dans un échantillon à la population dont il est issu. Deux grandes familles de techniques se complètent : l'intervalle de confiance et le test d'hypothèse. Si l'on prend l'exemple du client qui se rend dans un garage avec sa voiture en panne, l'intervalle de confiance répond à la question : combien la réparation va-t-elle coûter? (*Très probablement* entre 2.000 et 2.500 euros), ou : combien de kms vais-je pouvoir rouler après cette réparation (*Très probablement* entre 10.000 et 15.000 Kms). Il s'agit donc d'une fourchette qui fournit une base de réflexion. Le test d'hypothèse est une prise de décision : soit je répare, soit je ne répare pas et j'achète une nouvelle voiture.

Les barres d'erreurs sont essentiellement une représentation graphique de l'intervalle de confiance. L'utiliser comme « test d'hypothèse graphique » est nettement plus périlleux.



■ La variabilité d'une série d'observations représente l'écart entre des observations réalisées :

1. dans des conditions identiques
2. dans des conditions identiques sur des individus différents
3. dans des conditions différentes

Les sources de variabilité

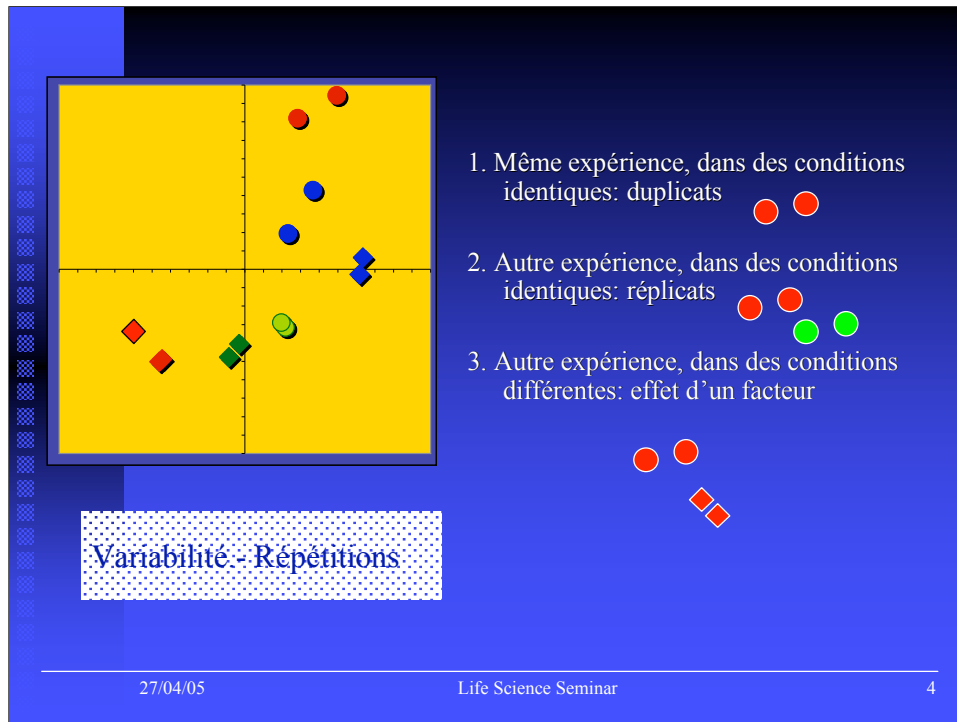
27/04/05 Life Science Seminar 3

Pour utiliser l'intervalle de confiance, il faut d'abord déterminer quelle est la variabilité que l'on veut mettre en évidence. Il y a trois type de variabilité représentant du « bruit » ou de « l'information »

1. Bruit : variabilité résiduelle, inexplicée
2. Bruit : variabilité individuelle, on dira aussi variabilité biologique, dans le domaine du vivant
3. Information : Variabilité factorielle, liée au facteur que l'on étudie

Variabilité totale = information + bruit, l'information étant liée à un modèle.

Sur le graphique, aucun modèle ne précise si les impacts sur la cible ont été tirés par le même tireur, avec la même carabine ...



un modèle permet d' identifier les différents niveaux de variabilité.

$$X(ij)k = \mu + a_i + B(ij) + E(ij)k$$

Sur ce graphique, les différents éléments du modèles sont représentés schématiquement. Dans la métaphore du tir sur une cible, les duplicats représentent des tirs réalisés par la même personne avec la même carabine, au même moment... La distance entre ces impacts représentent la variabilité résiduelle,  $E(ij)k$ . Les répliqués représentent des tirs réalisés par différentes personnes avec la même carabine. La distance entre ces impacts représentent la variabilité individuelle  $B(ij)$ . Le facteur représentent des tirs réalisés par différentes personnes avec une deux carabines, dont les viseurs est peuvent être réglés différemment. La distance entre ces impacts représentent la variabilité factorielle  $a_i$ .

$\mu$  représente le centre de la cible et  $x(ij)k$  la coordonnées du tir  $k$  réalisé par la personne  $j$  avec la carabine  $i$ .

Variabilité - Modélisation

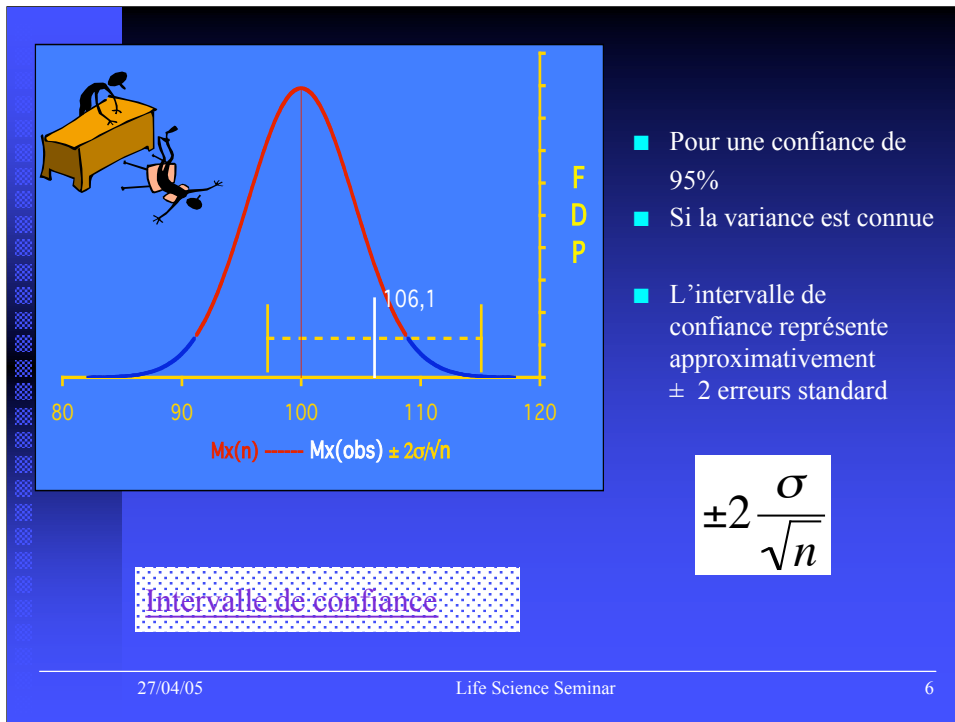
- La variabilité de la moyenne de plusieurs observations se modélise par la variable normale
- Plus le nombre d'observation est grand plus la dispersion est faible

27/04/05
Life Science Seminar
5

Le théorème de la limite centrale nous dit que la moyenne d'une variable quelconque dans un échantillon de taille  $n$  suivra une distribution normale (si  $n$  est suffisamment grand), de variance  $\sigma^2/n$ .

Le graphique montre la superposition de la fonction de Gauss-Laplace, en rouge, avec la distribution de fréquence de la moyenne de 16 lancers de dés.

La valeur  $\sigma$  s'appelle l'écart-type et la valeur  $\sigma/n^{0,5}$  s'appelle l'erreur standard



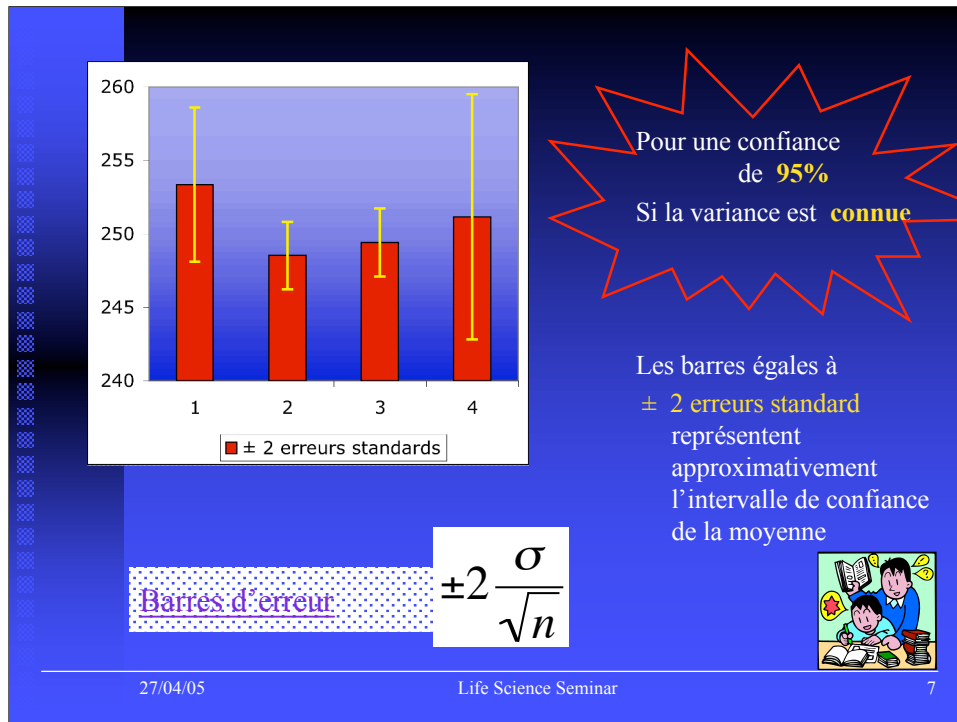
Le modèle Gaussien permet de prédire que dans une population de moyenne  $\mu$  et de variance  $\sigma$  connue:

95% des observations individuelles se trouvent dans un intervalle approximativement égal à :

$$\mu \pm 2 \sigma$$

95% des moyennes observées sur  $n$  observations se trouvent dans un intervalle approximativement égal à :

$$\mu \pm 2 \sigma/n^{0,5}$$



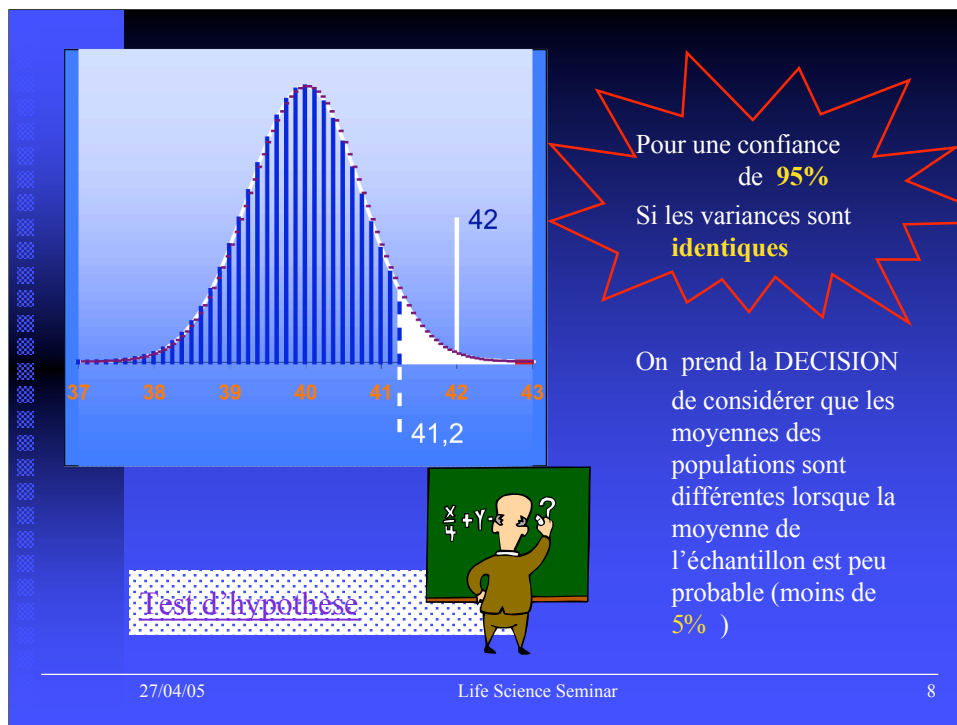
Les barres d'erreurs peuvent classiquement représenter différentes longueurs:

$$s, \sigma, s/n^{0,5}, \sigma/n^{0,5}, 2s, 2\sigma, 2s/n^{0,5}, 2\sigma/n^{0,5}$$

*Il n'est donc pas exagéré de demander à l'auteur de le préciser ;-)*

Elles représentent un écart-type ( $\sigma$ ), résiduel ou individuel suivant qu'il est calculé sur des duplicats ou des réplicats), ou une erreur-type ( $\sigma/n^{0,5}$ ), la variabilité étant connue ( $\sigma$ ) ou estimée ( $s$ )

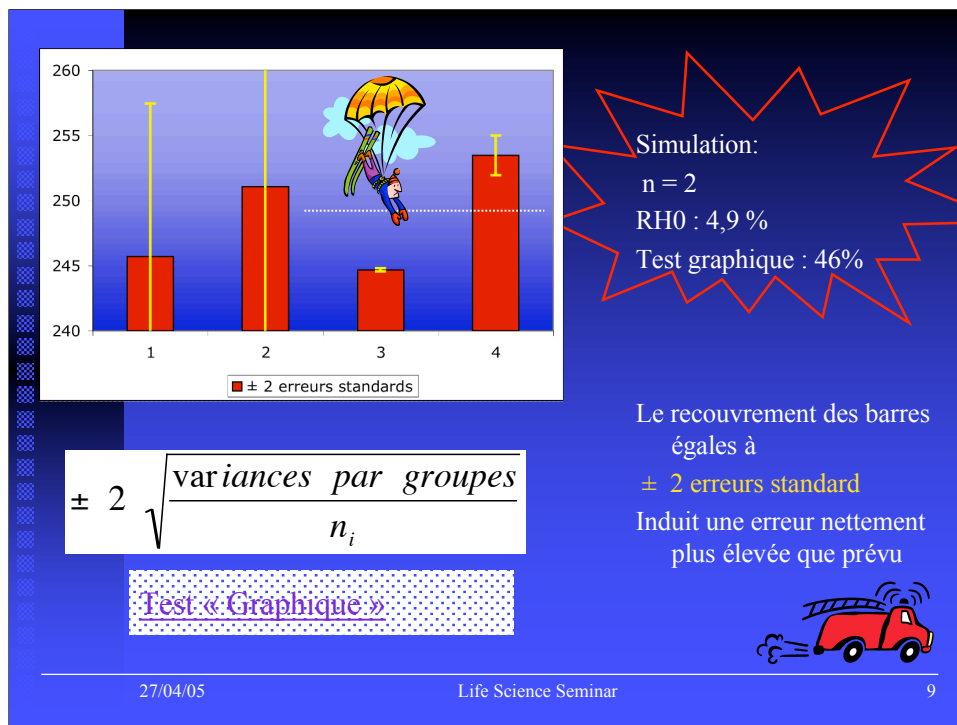
Elles représentent approximativement un intervalle de confiance, lorsque l'écart-type ou l'erreur type est multiplié par deux. Le seul intervalle de la moyenne correct est celui de deux erreurs types, lorsque la variance est connue. (En arrondissant à 2 la valeur de  $Z_{0,975} = 1,96$ )



Le test d'hypothèse sur les moyennes est une prise de décision. Le seuil alpha est garanti seulement si les variances sont homogènes (supposées égales dans les différentes populations testées).

Toute comparaison de moyennes doit être précédée d'un test de comparaison des variances.

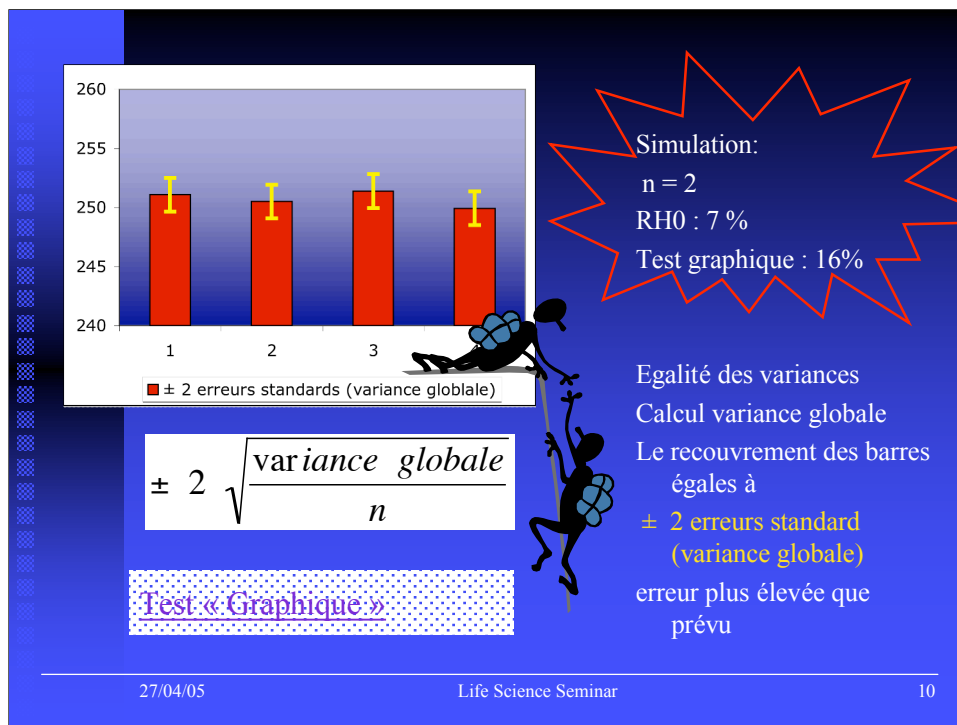




Un « test graphique » est réalisé en considérant que les moyennes sont significativement différentes lorsque les barres d'erreur ne se recouvrent pas.

Bien que cette pratique ne soit pas recommandée, elle est tellement courante, et publiée, qu'il est indispensable d'en fixer les limites de validité.

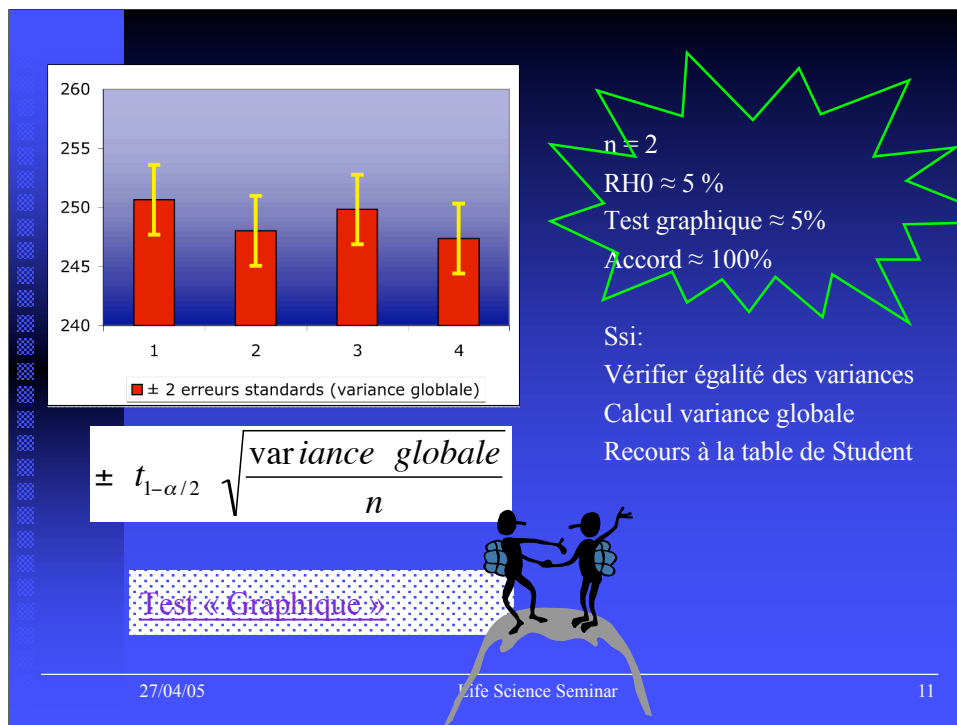
Tout d'abord, seul l'intervalle de deux erreurs standards est comparable à un test d'hypothèses bi-directionnel au seuil 5%. Si n est petit et que les variances sont estimées, le test graphique est totalement invalide. Dans la simulation reproduite ci-dessus, le taux d'erreur réel est de 46% au lieu des 5% annoncés.



Si l'on calcule une variance globale pour estimer l'erreur standard le test est amélioré.

Un « test d'hypothèse graphique » ne peut pas être réalisé lorsque l'on constate que les barres d'erreurs n'ont pas la même longueur!

Toutefois lorsque n est petit, la longueur de l'intervalle ne peut pas être estimée par 2 erreurs standards car 2 approxime un  $Z_{0,975}$  mais pas un  $t_{0,975}$  (le recours au t de Student est dû à l'estimation de la variance).



Enfin, si l'homogénéité des variances est testée, si une variance globale (CMR, d.l. = Nombre d'observations moins nombre de groupes) est estimée et si la longueur de l'intervalle est

$t_{d.l. \ 1-\alpha/2}$  alors l'accord entre le test d'hypothèse graphique et le test d'égalité des moyennes est quasi parfait.

Notons cependant qu'en cas de comparaison multiples 1 vs 2, 1 vs 3, 2 vs 3 etc... le seuil alpha n'est plus garanti et que la seule voie possible est celle des contrastes (par exemple contrastes de Scheffé).

En conclusion, la plus grande prudence est recommandée pour interpréter les barres d'erreur comme un test d'hypothèse!