

Statistiques de la Mesure

- **Mesure** *Measurement*
- **Incertitude** *uncertainty*
- **Etalonnage** *Calibration*

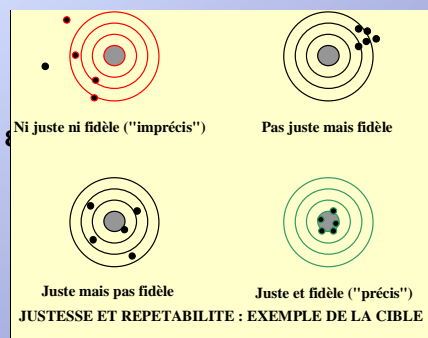
Philippe Breuil, novembre 2008

Quelques rappels sur l'erreur et l'incertitude...

Mesure x d'une variable de valeur **réelle** x_R :

Erreur: variable $\varepsilon = x_R - x$ $x = x_R + \varepsilon_B + \varepsilon_A$

- Erreur systématique: ε_B
Systematic error
- Erreur accidentelle ou aléatoire: ε_A
Random or accidental error



Erreur systématique ou aléatoire?

- Par définition, l'erreur aléatoire est imprévisible (et donc non corrigeable) et de valeur moyenne nulle
- Le caractère systématique ou aléatoire de l'erreur peut dépendre du contexte...

3

Erreur systématique ou aléatoire?

- Ex: mesure d'une masse à l'aide d'une balance numérique:

Erreur: $\varepsilon = \varepsilon_a + \varepsilon_T + \varepsilon_o + \varepsilon_u + \varepsilon_s$

ε_a : Erreur aléatoire non expliquée
 ε_T : Dérive en température
 ε_o : Erreur "opérateur"
 ε_u : Erreur propre à l'appareil
 ε_s : Erreur systématique de la série (!)

Conditions des expériences	ε_a	ε_T	ε_o	ε_u	ε_s
1 opérateur le même jour, 1 appareil					
Idem + étalé sur plusieurs jours					
Idem + plusieurs opérateurs					
Idem + tests sur un lot d'appareils					

4

Caractérisation de la mesure (et de l'erreur aléatoire):

- Moyenne estimée de n mesures d'une même valeur x_R

Estimated mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\lim_{n \rightarrow \infty} (\bar{x}) = x_R = E(x) \quad (\text{espérance de } x)$$

$n \rightarrow \infty$

$x_R = \text{valeur réelle}$

Expectation

(loi des grands nombres)

Law of large numbers

(Si erreur aléatoire)

5

Ecart-type de l'erreur aléatoire:

Standard deviation of random error

Ecart-type de l'erreur aléatoire

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

Valeur réelle, À priori inconnue...

Ecart-type estimé de l'erreur aléatoire

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$E(s) = \sigma$$

Variance (de l'erreur aléatoire): $V = \sigma^2$

Variance (of random error)

Ecart type relatif : $\sigma_r = \frac{\sigma}{|x|}$

6

Quelques calculs, justification du terme (n-1)

$$V(x) = E((x - E(x))^2) = E(x^2 - 2xE(x) + E(x)^2) = E(x^2) - E(x)^2$$

$$\rightarrow E(x^2) = E(x)^2 + V(x)$$

$$E(\bar{x}) = E(x)$$

Soit la grandeur: $v(x) = \overline{(x - \bar{x})^2} = \frac{1}{n} \sum_i (x_i - \bar{x})^2 = \overline{(x^2)} - \bar{x}^2$

$$E(v) = E(\overline{(x^2)} - \bar{x}^2) = E(x^2) - E(\bar{x}^2)$$

$$E(v) = E(x)^2 + V(x) - E(\bar{x})^2 - V(\bar{x})$$

Or, d'après le théorème de la moyenne: $V(\bar{x}) = \frac{V(x)}{n}$ (démonstration + loin, propagation d'erreur)

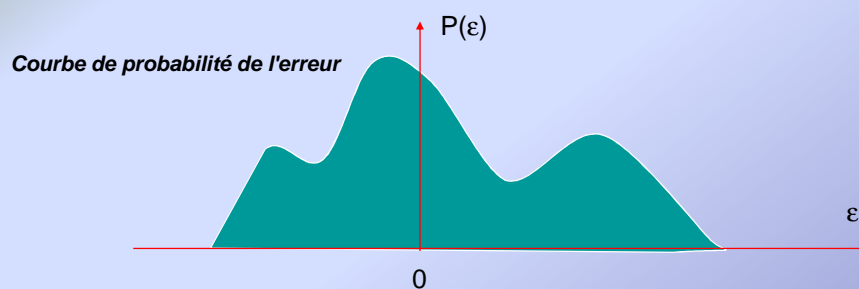
$$E(v) = E(x)^2 + V(x) - E(x)^2 - \frac{1}{n}V(x) = \frac{n-1}{n}V(x)$$

Donc le terme $\frac{n}{n-1}v(x) = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2$ a pour espérance $V(x)$...

7

Distributions d'erreurs

Distributions of errors



- Gaussienne ou normale*,
- Uniforme,
- Poissonnienne,
- Etc...

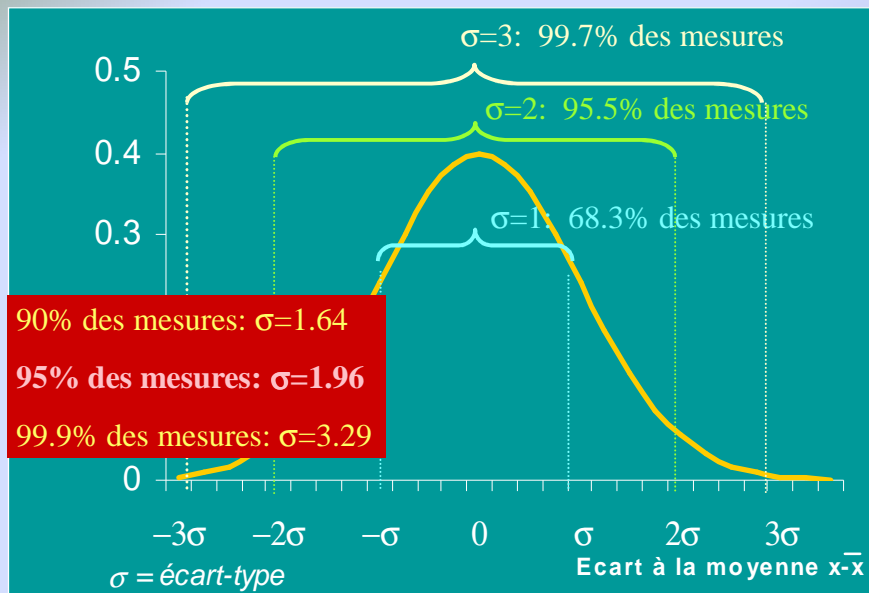
- Gaussian or normal*,
- Uniform,
- Poissonian,

* La plus répandue, grâce au Théorème central limite...

8

La distribution Gaussienne (ou normale)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma}\right)^2}$$



Théorème central limite

Central Limit Theorem

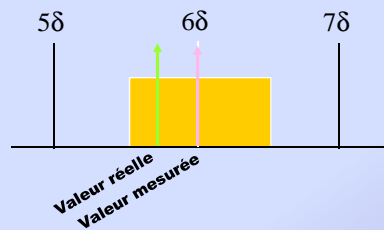
- Si une variable est la résultante d'un grand nombre de causes, petites, à effet additif, cette variable tend vers une **loi normale**.
- C'est à cause de cette interprétation que la loi normale est très souvent employée comme modèle (malheureusement pas toujours à raison).

Demo 10

Distribution d'erreur uniforme

Erreur équiprobable sur tout un intervalle

Peu courant, sauf erreur de discrétisation:

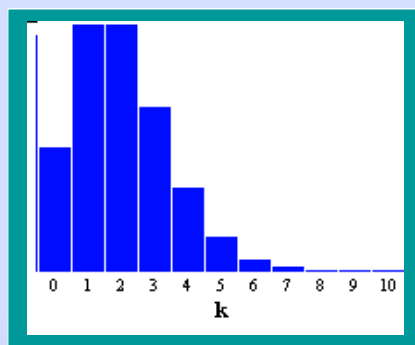


$$\sigma_d = \sqrt{\frac{1}{\delta} \int_{-\frac{\delta}{2}}^{\frac{\delta}{2}} x^2 dx} = \frac{\delta}{2\sqrt{3}} = 0.29\delta$$

11

Distribution de Poisson

- Ex: comptage d'évènements non simultanés (désintégration radioactive, queue...)



$$P(r) = \frac{\mu^r e^{-\mu}}{r!}$$

Moyenne: μ

Variance: μ

[demo](#)

12

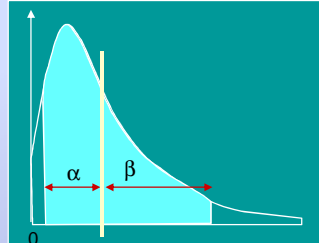
Intervalle de confiance *Confidence interval*

L'intervalle de confiance à p (95%) d'une mesure x est un intervalle de valeurs qui a une probabilité centrée p (95%) de contenir la vraie valeur x_R du paramètre estimé.

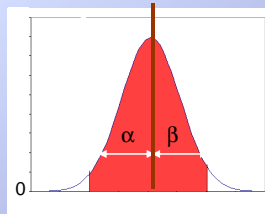
$$[x - \alpha, x + \beta]$$

$$P(x_R < \alpha) = P(x_R > \beta) = 1 - p / 2$$

α et β = incertitude à p (95%)



La calcul de α et β en fonction de la probabilité P (généralement 95%) dépend de la loi de distribution de l'erreur



- Distribution symétrique: $\alpha = \beta$
- Valeur moyenne de la mesure = la plus probable

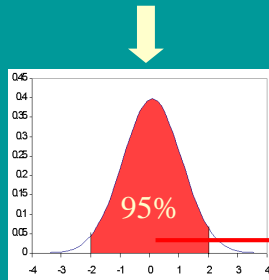
Écriture d'une mesure (distribution normale)

Le résultat d'une mesure doit comporter 4 éléments :

Ex : $C_{NO} = 125.3 \text{ ppb} \pm 1.7 \text{ ppb}$ (à 95% ou $k=2$)

- 1 : Valeur numérique avec un nombre correct de décimales
- 2 : Unité
- 3 : Incertitude élargie = $t \cdot \sigma$
- 4 : Le coefficient d'élargissement t utilisé

Probabilité en % pour que la mesure soit hors de l'intervalle $[x - ts, x + ts]$



Loi normale:

- 1 : Numerical value with a correct number of decimals
- 2 : Unit
- 3 : expanded uncertainty = $t \cdot s$
- 4 : Coverage factor t

		Coefs de Student t				
intervalle de conf.		90.0%	95.0%	98.0%	99.0%	99.9%
n	ag lib	0.1	0.05	0.02	0.01	0.001
	1	6.31	12.71	31.82	63.66	636.62
	2	2.92	4.30	6.96	9.92	31.60
	3	2.35	3.18	4.54	5.84	12.92
	4	2.13	2.78	3.75	4.60	8.61
	5	2.02	2.57	3.36	4.03	6.87
	6	1.94	2.45	3.14	3.71	5.96
	7	1.89	2.36	3.00	3.50	5.41
	8	1.86	2.31	2.90	3.36	5.04
	9	1.83	2.26	2.82	3.25	4.78
	10	1.81	2.23	2.76	3.17	4.59
12	1.78	2.18	2.68	3.05	4.32	
14	1.76	2.14	2.62	2.98	4.14	
17	1.74	2.11	2.57	2.90	3.97	
20	1.72	2.09	2.53	2.85	3.85	
30	1.70	2.04	2.46	2.75	3.65	
40	1.68	2.02	2.42	2.70	3.55	
50	1.68	2.01	2.40	2.68	3.50	
100	1.66	1.98	2.36	2.63	3.39	
	100000	1.64	1.96	2.33	2.58	3.29

Nombre de points - 1 pour le calcul de σ

Évaluation de l'incertitude :

Évaluation par analyse statistique de séries de mesures (« type A »)
(généralement mesure, mais aussi simulation)

Évaluation par calcul de l'effet sur l'incertitude finale des différentes sources d'incertitude (« type B »: « par tout autre moyen »!), elles même évaluées :

par une méthode de type A,

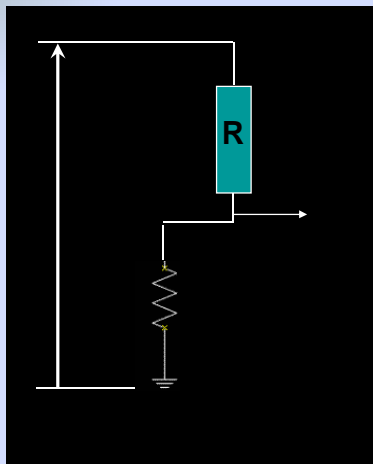
par des données constructeur, d'étalonnage etc...

Il est alors nécessaire de connaître les lois de propagation de l'erreur... →

15

Évaluation par analyse statistique de séries de mesures

Exemple: mesure d'une résistance:



$$V = E \frac{R_0}{R + R_0} \quad \longrightarrow \quad R = R_0 \frac{E - V}{V}$$

Mesure et calcul avec un grand nombre N de résistances « étalon » (R connu)

Calcul de l'écart-type σ

Incertitude à 95% = $t \cdot \sigma$, # 2σ si $N > 15$

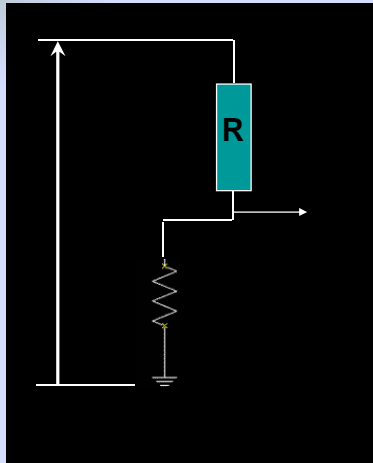
Ne tient pas compte:

- De l'incertitude du Vmètre (V)

- Des incertitudes sur E, R_0 , les résistances étalon...

16

Évaluation par calcul de l'effet sur l'incertitude finale des différentes sources d'incertitude



Exemple: mesure d'une résistance:

$$R = R_0 \frac{E - V}{V}$$

$\sigma(R_0), \sigma(E), \sigma(V)$ connus

Loi de propagation des écart-types



combined standard uncertainty

$$\sigma(R) = \sqrt{\left(\frac{E-U}{U} \sigma(R_0)\right)^2 + \left(\frac{R_0}{V} \sigma(E)\right)^2 + \left(\frac{R_0 E}{V^2} \sigma(V)\right)^2}$$

Explication →

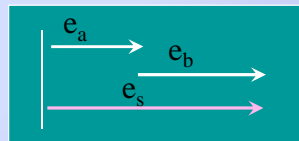
17

Lois de propagation des erreurs *aléatoires indépendantes* (1)

combined standard uncertainty for random & independant errors

Exemple de la somme: $s = a + b$

Rappel: les erreurs s'ajoutent algébriquement: $e_s = e_a + e_b$

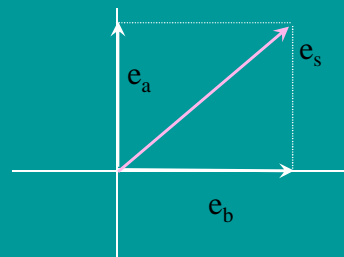


Et leur écart-type?

And Standard deviations?

Erreur aléatoire:

$$\sigma(s) \leq \sigma(a) + \sigma(b)$$



Écart type de l'erreur aléatoire de la somme de variables

Exemple de la somme: $s = a + b$ (suite)

Variance: $Var(a) = \frac{1}{n-1} \sum (a_i - \bar{a})^2$ $Var(b) = \frac{1}{n-1} \sum (b_i - \bar{b})^2$

$$Var(s) = \frac{1}{n-1} \sum (a_i - \bar{a} + b_i - \bar{b})^2 = \frac{1}{n-1} \left(\sum (a_i - \bar{a})^2 + \sum (b_i - \bar{b})^2 + 2 \sum ((a_i - \bar{a})(b_i - \bar{b})) \right)$$

$$Var(a + b) = Var(a) + Var(b) + 2Cov(a, b)$$

= 0 si erreurs indépendantes

Les variances s'ajoutent,

Et les écarts-types?

$$\sigma(a+b) = \sqrt{\sigma(a)^2 + \sigma(b)^2}$$

Somme **quadratique** des écart-types

19

Lois de propagation des erreurs *aléatoires indépendantes* (2)

Combinaison linéaire: $y = \sum_i a_i x_i$ $\sigma(y)^2 = \sum_i (a_i \sigma(x_i))^2$

Ex: **somme ou différence:**

$$y = a + b - c$$



$$\sigma(y) = \sqrt{\sigma(a)^2 + \sigma(b)^2 + \sigma(c)^2}$$

Application fondamentale en instrumentation: La moyenne

20

Lois de propagation des erreurs *aléatoires indépendantes* (3)

Application fondamentale en instrumentation: **La moyenne**

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \longrightarrow \quad \sigma(\bar{x}) = \frac{1}{n} \sqrt{\sum_{i=1}^n \sigma(x)^2} = \frac{\sigma(x)}{\sqrt{n}}$$

Théorème de la moyenne: *Theorem of the mean*
Si une série de mesures a une erreur aléatoire et indépendante d'écart-type σ , alors la moyenne de n de ces mesures a une erreur aléatoire dont l'écart-type est divisé par racine carrée de n .

Loi "des grands nombres"

Demo 21 Esc

Lois de propagation des erreurs aléatoires (3)

Produits et puissances :

Propriété intéressante des écart-types relatifs: $\frac{\sigma(x)}{x}$

$$y = A \prod_i x_i^{\alpha_i} \quad \longrightarrow \quad \left(\frac{\sigma(y)}{y} \right)^2 = \sum_i \left(\alpha_i \frac{\sigma(x_i)}{x_i} \right)^2$$

Somme quadratique des écart-types *relatifs*

22

Lois de propagation des erreurs aléatoires (4)

Fonction quelconque:

$$y = f(x_1, \dots, x_i, \dots)$$

$$\sigma^2(y) = \sum_1^n \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma^2(x_i)$$

Cas général (variables non indépendantes):

$$\sigma^2(y) = \sum_1^n \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma^2(x_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \text{Cov}(x_i, x_j) \quad !$$

(rarement utilisé...)

23

Calcul de l'incertitude par simulation de l'erreur aléatoire

$$\text{Ex: } y = f(x_1, x_2),$$

x_1 et x_2 ont des erreurs aléatoires indépendantes caractérisées par leur écart-type s_1 et s_2

- Simuler un certain nombre d'expériences avec les mêmes valeurs x_1 et x_2 + erreur aléatoire

$$y_i = f(x_1 + \varepsilon_1, x_2 + \varepsilon_2) \quad N \text{ fois}$$

- Les ε_i sont calculés à l'aide d'un générateur de nombre aléatoires à distribution adéquate (généralement normale)

- Calcul de l'écart type des y_i

Algorithme de Box-Müller: $\varepsilon = \sqrt{-2 \text{Ln}(\text{Rnd})} * \cos(2\pi \text{Rnd}')$

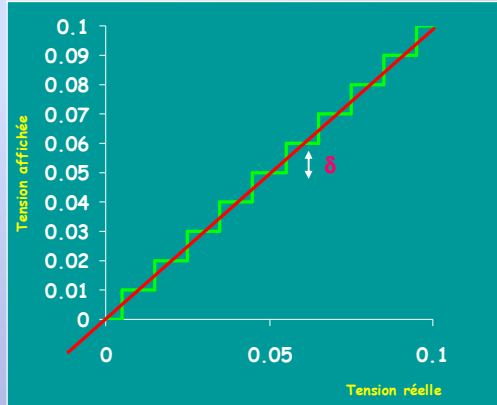
Rnd et Rnd' : distributions uniformes 0-1 $\bar{\varepsilon} = 0$ $\sigma\varepsilon = 1$

Ex 24
Minitab

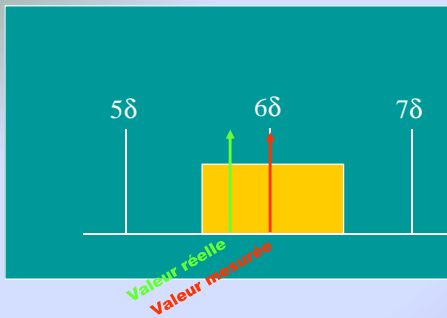
l'incertitude due à la discrétisation :

Discretization uncertainty

- La discrétisation intervient lorsque les valeurs finales constituent un ensemble discret.
- *Résolution* δ du système d'acquisition, à ne pas confondre avec l'*incertitude*, ou la *sensibilité*.



l'incertitude due à la discrétisation : (2)



Erreur maxi: $\delta/2$

$$\sigma_d = \sqrt{\frac{1}{\delta} \int_{-\frac{\delta}{2}}^{\frac{\delta}{2}} x^2 dx} = \frac{\delta}{2\sqrt{3}} = 0.29\delta$$

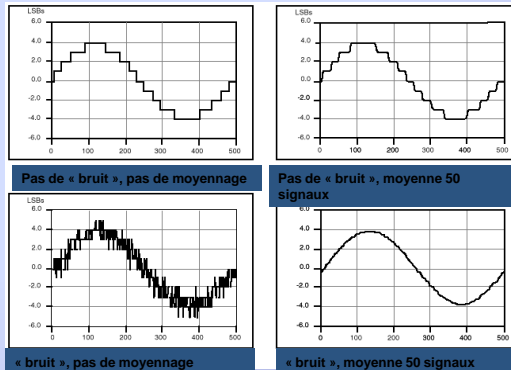
Ecart type de l'erreur due à la discrétisation: $\sigma_d = \frac{\delta}{2\sqrt{3}}$

Ecart type de l'erreur totale:

$$\sigma(X) = \sqrt{\sigma^2(x) + \sigma_d^2}$$

Le « Dithering »

La méthode du moyennage **peut** réduire aussi l'erreur de discrétisation (= quantification):



On peut gagner 1 bit de résolution chaque fois que l'on multiplie par 4 la fréquence d'échantillonnage...

... à condition que le signal avant numérisation contienne un bruit (aléatoire et indépendant) de valeur efficace supérieure à la résolution initiale...

La présence d'erreur aléatoire permet ici d'améliorer la « précision » de la mesure!!!

27

l'incertitude due à la discrétisation : (3) les décimales significatives

Écriture d'un nombre avec un nombre de chiffres *fini* → Erreur de discrétisation

Pour une mesure, l'écart type de l'erreur introduite doit être petit devant celui de l'erreur de mesure initiale

Mais les systèmes numériques peuvent donner beaucoup de chiffres significatifs...

Ex: mesure = 438.2659872 Écart-type mesure = 0.55 (soit 0.125 % en relatif)



438.2659872 ?

Doit-on écrire:

438.265 ?

438 ?

28

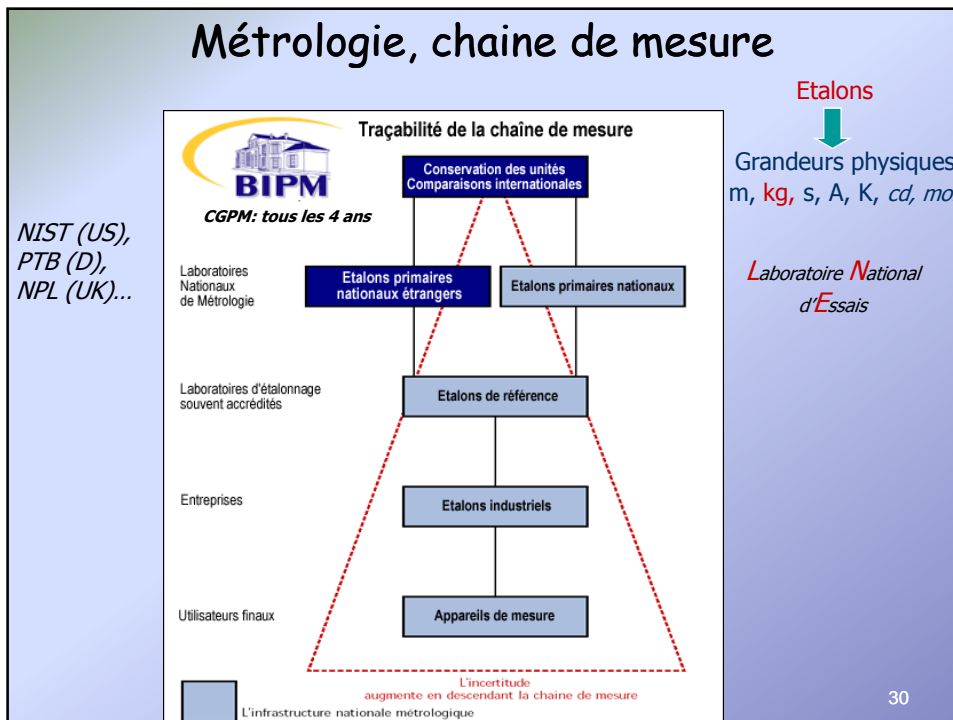
l'incertitude due à la discrétisation : (3) les décimales significatives:
exemple

Ex: mesure = 438.2659872 Écart-type mesure = $\sigma(x) = 0.55$ (soit 0.125 % en relatif)

Mesure affichée <i>résolution</i>	E.T. erreur discrétisation $\sigma_d = \frac{\delta}{2\sqrt{3}}$	E.T. erreur finale $\sigma(X) = \sqrt{\sigma^2(x) + \sigma_d^2}$	% d'erreur due à la discrétisation
438.265987 0.000001	2.9 10 ⁻⁷	0.55	< 10 ⁻³ %
438.27 0.01	2.9 10 ⁻³	0.550008	10 ⁻³ %
438.3 0.1	0.029	0.5506	0.14 %
438 1	0.29	0.62	13 %
440 10	2.89	2.94	434 %

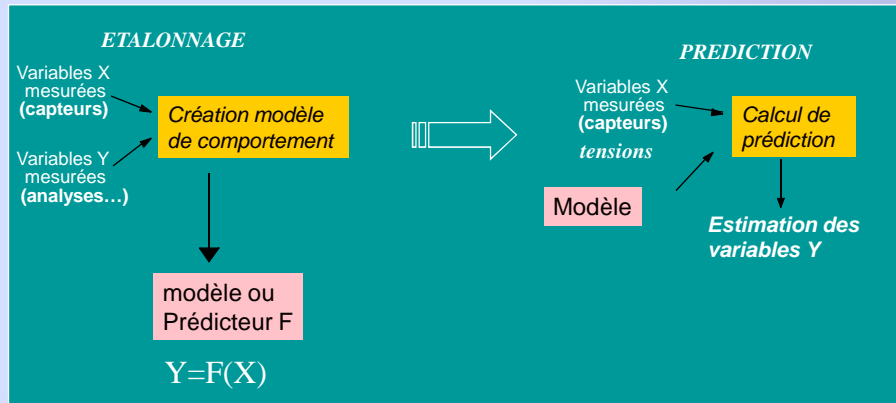
- On peut choisir la résolution immédiatement plus petite que l'écart-type de la mesure.
- L'écart-type (ou l'incertitude) affiché ne doit pas avoir plus de 2 chiffres significatifs.

Métrieologie, chaîne de mesure



Etalonnage: Création d'un modèle de comportement

Calibration: Creation of a *behavior model*



Les moindres carrés

The least Squares

Méthode des moindres carrés :

On veut modéliser: $Y = F(X)$, en fait, à partir des expériences, on a:
 $Y = F(X) + E$ où E = « résidu » ou erreur

$$E = Y - F(X) = (y_j^i - f_j(x_1^i, \dots, x_k^i)) \quad \text{minimal}$$

On utilise généralement la distance Euclidienne:

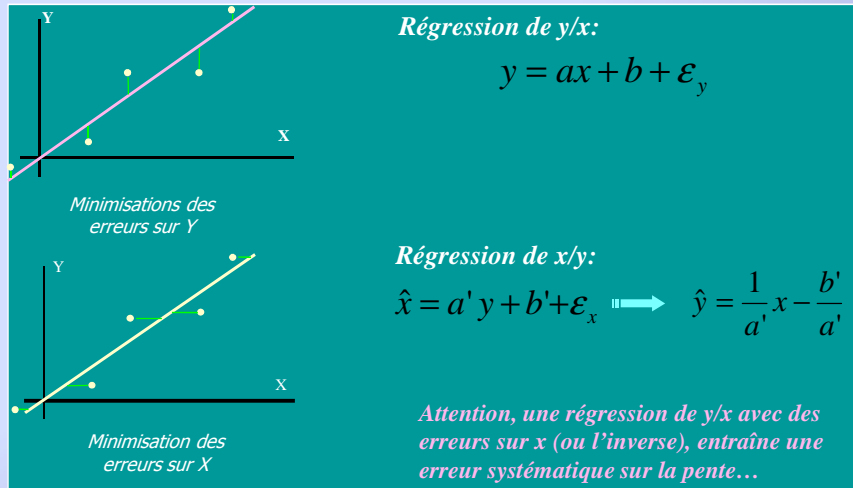
$$\sum_{i,j} (y_j^i - f_j(x_1^i, \dots, x_k^i))^2 \quad \text{« minimale »}$$

32

La régression linéaire

Linear regression

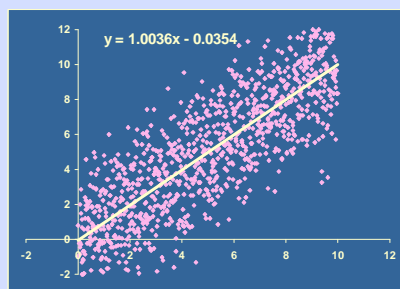
- Monovariante ($y=f(x)$) + hypothèse linéaire ($y=ax+b$)
- 2 cas:



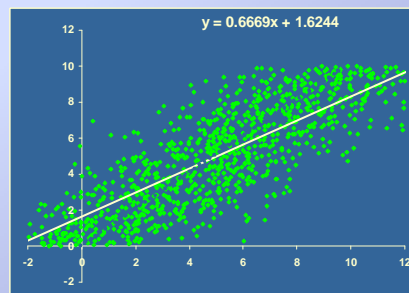
La régression linéaire: y/x ou x/y ?

Attention, une régression de y/x avec des erreurs sur x (ou l'inverse), entraîne une erreur systématique sur la pente...

Régression Y/X



Régression X/Y



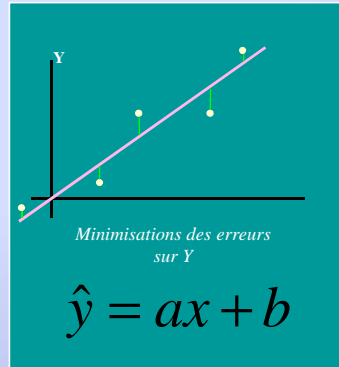
Différence négligeable dans une majorité de cas...

Caractéristiques de la régression y/x:

- Estimations de a et b:

$$a = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \frac{Cov(x, y)}{Var(x)}$$

$$b = \bar{y} - a\bar{x}$$



Ecart type de la pente: $s_a = \frac{s_{y/x}}{\sqrt{\sum_i (x_i - \bar{x})^2}}$ avec $s_{y/x} = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{n-2}}$

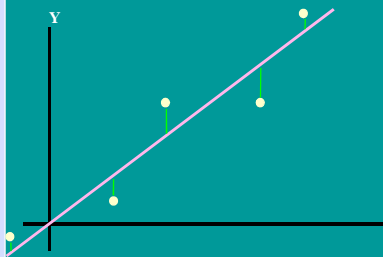
Ecart type de l'ordonnée à l'origine: $s_b = s_{y/x} \sqrt{\frac{\sum_i x_i^2}{n \sum_i (x_i - \bar{x})^2}}$

35

Mesure de la qualité d'ajustement d'une régression linéaire:

$$\text{Var}(Y) = \text{Var}(aX+b) + \text{Var}(\epsilon)$$

information totale = information modélisée + information résiduelle



Coef. De détermination:

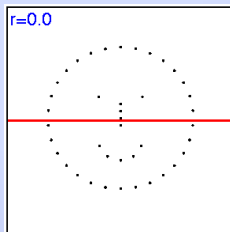
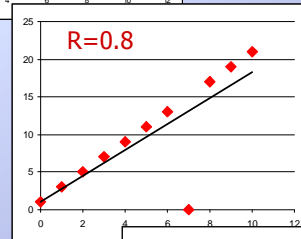
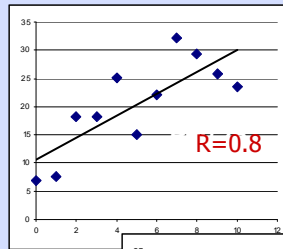
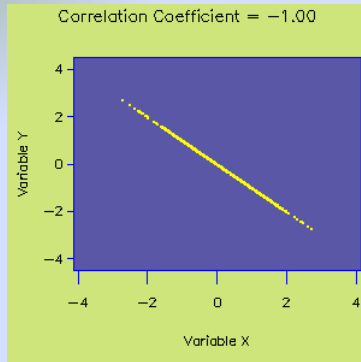
$$R_d = \frac{\text{Var}(\text{info modélisée})}{\text{Var}(\text{info totale})}$$

Coef. de corrélation:

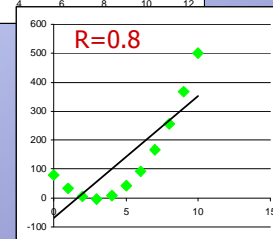
$$R_d = R_c^2$$

$$R_c = \frac{Cov(X, Y)}{\sigma(X)\sigma(Y)}$$

Le coef. de corrélation, utilisation:

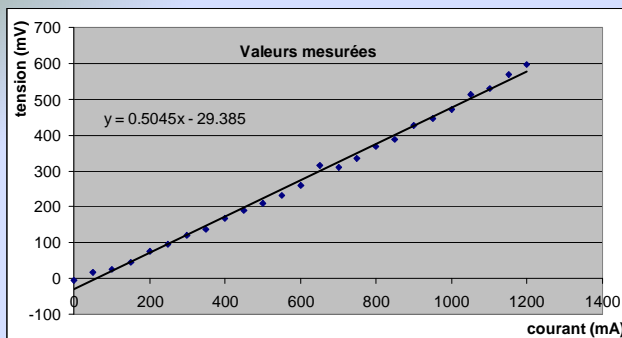


Le coefficient de corrélation ne sert à quantifier que les relations **linéaires** entre X et Y variables



Régression linéaire: examen des résidus

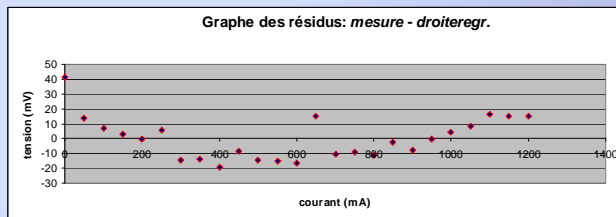
residuals



Résidu = information non modélisée, idéalement erreur aléatoire

$$R_i = y_i - (a \cdot x_i + b)$$

Détection points aberrants (*outliers*)
A éliminer après vérification et avec précautions...



Vérification linéarité
A corriger éventuellement avec régression non linéaire

38

Autres méthodes de modélisation:

- Multivariables:
 - Multivariable linéaire
 - Regression multivariable, analyse composantes principales
 - Moindres carrés partiels (PLS)
- Monovariaire non linéaire
- Polynomial, loi puissance, fonction quelconque

39

Régression multivariables linéaire:

- n Xvariables, p expériences

Écriture matricielle pour les p expériences:

$$Y = A.X + \epsilon_y \quad \text{À minimiser}$$

$$\begin{cases} y_1 = a_1x_{1,1} + \dots + a_nx_{n,1} + \epsilon_1 \\ y_p = a_1x_{1,p} + \dots + a_nx_{n,p} + \epsilon_p \end{cases}$$

➤ $p < n$: pas de solution...

➤ $p = n$: peut-être une solution: $A = Y.X^{-1}$ Résidu ϵ_y nul...

(Ex: 2 points pour déterminer 1 droite...)

➤ $p > n$: X non carrée...

Mais peut-être quand même une solution au sens des moindres carrés!

$$A = Y.X^T.(X.X^T)^{-1} \quad X^T.(X.X^T)^{-1} \text{ est appelée « pseudo-inverse » de X}$$

➔ fonction "DROITEREG" d'Excel

Utilisable pour une quantité faible de Xvariables faiblement corrélées entre elles, sinon, méthodes d'analyse de facteurs (PCR, PLS...)

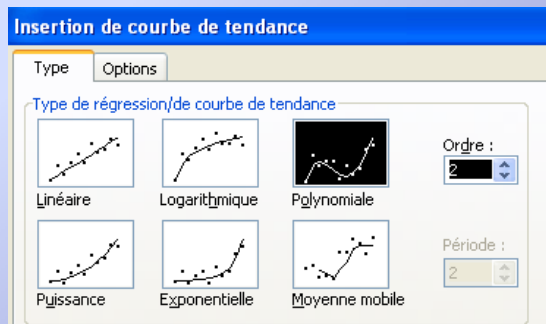
40

Régression non linéaire: 1 cas "simple"

- La fonction f se décompose en une combinaison linéaire de fonctions monovariées, ex polynôme:
 - $y=(a_0)+a_1.x+a_2.x^2+...+a_n.x^n$
- Alors les grandeurs x, x^2, \dots, x^n peuvent être considérées comme *linéairement* indépendantes et le problème est équivalent au cas précédent (multivariable linéaire)

Ex: courbe de tendance d'Excel:

[excel](#)



41

Régression non linéaire: cas général

Il n'y a pas de solution analytique au problème des moindres carrés dans le cas général.

On doit alors utiliser des algorithmes d'optimisation.

Ex: avec le [solveur d'Excel](#)

Ou des réseaux de neurones...

42

Biblio & liens utiles

Bouquins:

- "Statistics for analytical Chemistry" 3rd ed. J.C. Miller and J.N. Miller John Wiley & Sons, 1998.
"Multivariate Statistical Methods, A Primer" B.F.J. Monley, Chapman & Hall 1986.
"Modélisation et estimation des erreurs de mesure", M Neully, CETAMA, Lavoisier, Paris 1993.

Sites WWW: (la plupart en anglais...)

- <http://www.emse.fr/%7Epbreuil/capmes/index.htm> Ce document + exercices sous Excel
<http://physics.nist.gov/cuu/Uncertainty/index.html>: excellent site simple et succinct mais de référence sur la calcul des incertitudes, ce document s'en est beaucoup inspiré. Un cours plus complet de statistiques est disponible à : <http://www.itl.nist.gov/div898/handbook/>
<http://rfv.insa-lyon.fr/~jolon/STAT/poly.html> : cours complet et touffu de probas & stats.
<http://math.uc.edu/~brycw/classes/147/blue/tools.htm> : collection de liens (dont beaucoup morts...), books online et applets.
<http://www.helsinki.fi/~jpuranen/links.html#psy>: très grande collection de liens « stats » classés.
<http://my.execpc.com/~helberg/statistics.htm>: même chose...
http://www.math-info.univ-paris5.fr/smel/simulations/cadre_simulations.html: cours, articles exemples et surtout "applets" sur les stats

43

La norme AFNOR la plus utilisée:

normalisation française

UNM 00-000
Avril 1996

Indice de classement UNM 00-000

Grandeurs et unités

Système d'unités pifométriques.

E : Quantities and units - Nosemetrical system of units

D : Grossen und Einheiten - Nasimetrischeinheitensystem.



Norme interne, totalement expérimentale et non éditable par l'afnor.

Les observations relatives au présent document peuvent être adressées à l'UNM à tout moment

correspondance Il est étonnant que le présent document soit unique, mais aucune étude internationale ne s'est encore pointée à l'horizon.

PDF

http://www.qualiteonline.com/norme_afnor_pifometrique.pdf

44/62