Toolbox Measurement & Regulation

Responsible : Philippe Breuil pbreuil@emse.fr

General description

The objective of this toolbox is for students to understand and evaluate or implement a measurement-control chain, from the sensor to the control loops and the measurement statistics.

- **UP1 : Measurement**
- **UP2 : Regulation**

Pedagogy

Short courses (maxi 1h) , Practical works (Excel, Matlab, Simulink), Tutorials

Keywords

Measurement, regulation, automatic, uncertainties, calibration

Evaluation: examination (2x1.5h, 05/29/2018) with PC, authorized documents, evaluated program fixed 1 week before.

Toolbox « Measurement & Regulation » 2018

13/3 & 27/3: Course + tutorials (Excel), Stats of measurement

3/4, 10/4 & 17/4: Course + tutorial (Excel, matlab, Simulink) „Rgulation“

13/4 (day): Measurement & Regulation, industry vision (N Caillet, Total)

15/5: Course + Tutorial „Advanced regulation“

18/5 (day): Course „Multivariate Analysis“+ tutorials (Matlab)

22/5: Course + industrial instrumentation problem solving « Sensors»

16/5: examination regulation et measurement (2 x 1h30, with PC)

Statistics of the Measurement

- ***Measurement*** *Mesure*
- ***Uncertainty*** *incertitude*
- ***Calibration*** *Etalonnage*

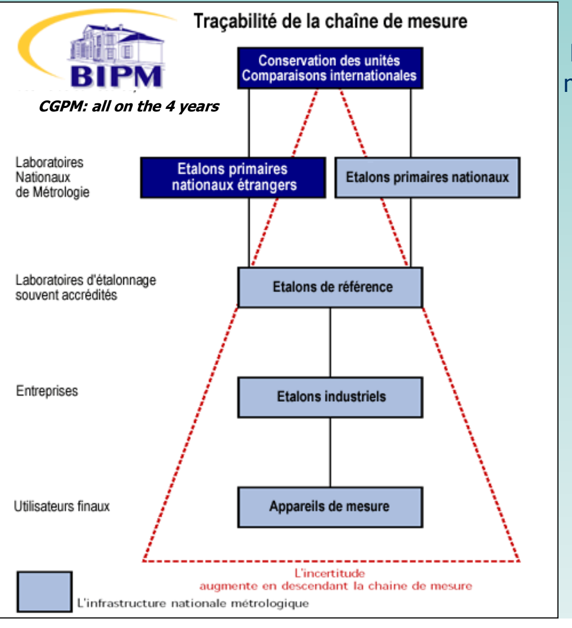
Philippe Breuil, March, 2018

Metrology = science of the measurement

- Physical and mathematical aspect
  - Statistics of the measurement:
    - -Calculation of the uncertainties
    - -Calibration
  - Ways of measure: sensors
- Legal aspect
  - Obligations during a commercial transaction
  - Obligations for the publication of a measurement
- Economic aspect
  - Traceability and reliability
  - Optimization of the quality

Metrology, measuring chain

NIST (US),
PTB (D),
NPL (UK) …



Standards

Physical quantities
m, kg, s, A, K, *cd, mol*

Laboratoire National d'Essais LNE

5

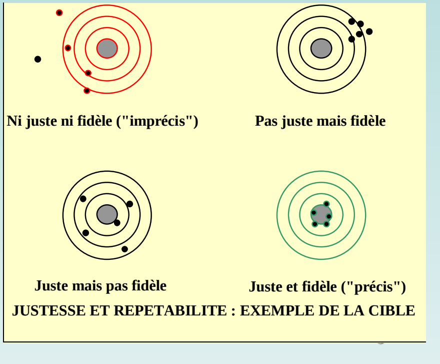
error and uncertainty …

Measurement x of a variable of real value $\mathbf{x_R}$: $$x_R = E(x)$$

(expectancy of x)
Espérance

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

- Systematic error: ε_B
  Erreur Systématique
- Accidental or random error: ε_A
  Erreur accidentelle ou aléatoire



Systematic or random error?

- By definition, the random error is unpredictable (and thus not corrigeable) and of null average value
- The systematic or random character of the error can depend on the context …

Systematic or random error?

- Ex: measure of a mass by means of a digital balance:

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

ε_a: random error explained no
ε_T : drift in temperature
ε_o : error "operator"
ε_u : error appropriate to the device
ε_s : systematic Error of the series (!)

Conditions of the experiences:	ε_a	ε_T	ε_o	ε_u	ε_s
1 operator the same day, 1 device					
Idem + spread over several days					
Idem + several operators					
Idem + tests on a set of devices					

Systematics

Random

Characterization of the measurement (*and of the random error*):

- Estimated mean of n measures of the same value x $_R$

Moyenne estimée

$$\bar{x} = \frac{1}{n}\sum_1^n x_i$$

$$\lim_{n\to\infty}(\bar{x}) = x_R = E(x)$$

x_R = *real value*

(Expectation of x)

Espérance

(Law of large numbers)

Loi des grands npmbres

(If random error only)

Standard deviation of the random error:

Ecart-type de l'erreur aléatoire

Variance *(of the random error)*: $V = E\left((x - E(x))^2\right) = \overline{(x - x_R)^2}$

Variance

Standard deviation of the random error $\sigma = \sqrt{V}$

$$\sigma = \sqrt{\left((x_i - E(x))^2\right)} = \sqrt{\frac{1}{n}\sum_1^n (x_i - E(x))^2}$$

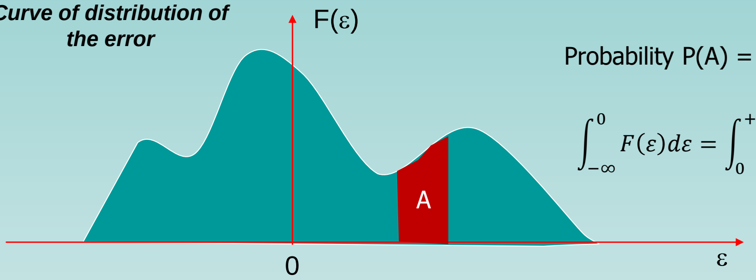
Real value, A priori unknown …

estimated **Standard deviation of the random error** $s = \sqrt{\frac{1}{n-1}\sum_1^n (x_i - \bar{x})^2}$ $E(s) = \sigma$

Relative standard deviation: $\sigma_r = \frac{\sigma}{|x|}$

Distributions of (random) errors

Curve of distribution of the error



Probability P(A) = $\int_A F(\varepsilon)d\varepsilon$

$$\int_{-\infty}^{0} F(\varepsilon)d\varepsilon = \int_{0}^{+\infty} F(\varepsilon)d\varepsilon = 0.5$$

- Gaussian or normal *,
- Uniform,
- Poissonnian,
- Etc. …

- *Gaussian gold normal *,*
- *Uniform,*
- *Poissonnian,*

* The most prevalent, thanks to the limit central Theorem …

Central theorem limits

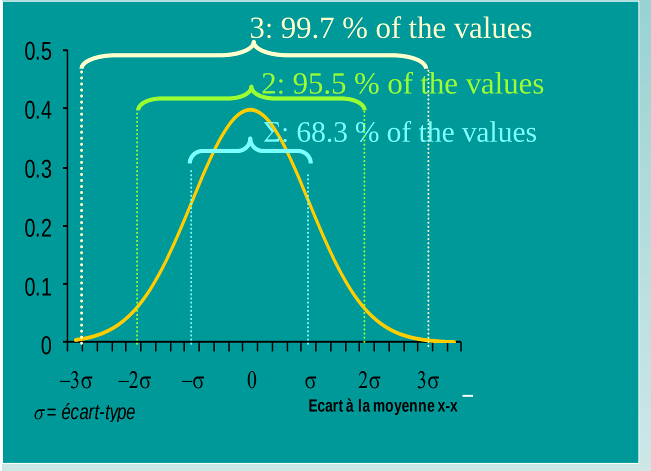
Limit Theorem exchange

- If a variable is the resultant of a large number of causes, small, with additive effect, this variable aims towards a ***normal law.***

- It is because of this interpretation that the normal law is very often used as model (unfortunately not always adequately).

Demo

The Gaussian distribution (or normal)



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

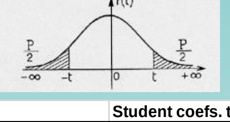
P	Width t
50 %	0,67·σ
68 %	1·σ
70 %	1,04·σ
87 %	1,5·σ
90 %	1,65·σ
95 %	1,96·σ
99 %	2,56·σ
99,7 %	3·σ
99,9 %	3,28·σ
99,999 999 8 %	6·σ

$$P = proba\left[\left|X - X_R\right| < t.\sigma\right]$$

Valid only if standard deviation calculated "comfortably" (n> 20)

Otherwise, Gauss → Student... 14

The coefficient t of Student



	Student coefs. t				
conf. Interv.	90.0%	95.0%	98.0%	99.0%	99.9%
p	0.1	0.05	0.02	0.01	0.001
N=nb deg lib					
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

Relation standard deviation / Gaussian: valid only if standard deviation perfectly known (n= ∞)
Practically, st. Dev. Determined from a statistical calculation:
-> Student's law
(tends to gaussian if n big, n > 20)

$$P = proba\left[\left|X - X_R\right| < t.\sigma\right]$$

P=TDIST(t;N;2)

t=LOI.STUDENT.INVERSE.BILATERALE(P,N)

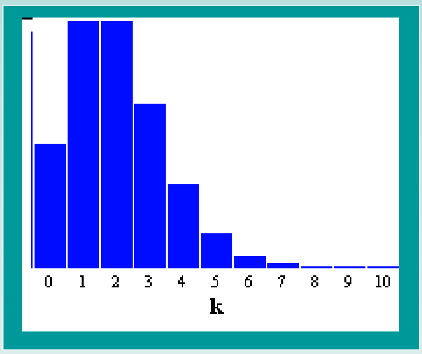
N = Nb samples used for the calculation of -1
= Nb of degrees of freedom

Reminder: valid only for gaussian law!

Http: // www.jybaudot.fr/Inferentielle/tstudent. HTML

Distribution of Poisson

- Ex: counting of not simultaneous events (radioactive decomposition, queue…)



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

Mean: μ
Variance: μ

Demo

The Significance tests in measurement

Tests d'hypothèse

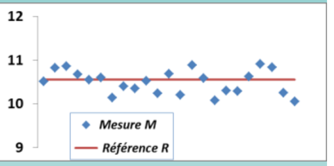
Tests of a hypothesis (" null hypothesis ", ex: 2 samples have the same mean) *) from a finite number* of measurements, with random error.
The result of the test is not absolute but is a probability which is a help to the validation or not of the initial hypothesis**, thus it never establishes a proof.**

Parametric tests *(Hypothesis on distribution + or - necessary)*	
1-sample T-test	**Comp. Sample with reference value, confidence interval**
2-samples T-test	**Comparison of 2 samples**
F-test	Comparison of the variance of 2 samples
ANOVA	Variance analysis: analysis of the variances of K samples, comparison of the means
Test Chi-public garden	Use in particular to verify a hypothesis of distribution
Grubbs ' test	Detection of the absurd values *("outliers"))*
Non parametric Tests *(No hypothesis on distribution)*	
Test Wilcoxon. Mr. W	Comparison of 2 samples, *method of row*

In red: described tests + far, otherwise to see biblio or google

"simplified" one-sample T-test

Comparison of the measurement M of a sample with a reference value R, it is assumed that the distribution is normal.

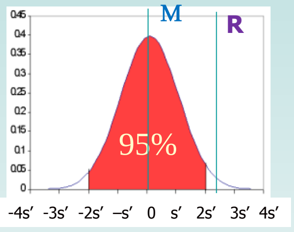


Hypothesis:

- *« is M est different from R »?* ***Yes, almost always, because of the random error …***
- *« Is M significantly different from R »?*
- *« The difference between **M** and **R** is "probably" not only due to random error. »*

This H1 difference hypothesis is used if its probability is greater than 95% (for example) (or if the probability of equality is less than 5% (H2)

1 measurement: **M**, standard deviation: s, Gaussian distr.



Hypothesis retained if:

$$|M - R| \geq t.s$$ *(t=Student's coef.)*

if s can be calculated « accuratly *»:

	H1: 95% (or H2: 5%)	H1: 99% (or H2: 1%)	H1: 99.9% (or H2: 0.1%)
t	1.96	2.56	3.28

****n>20, otherwise, Student's coefs table…***

« official » one-sample T-test

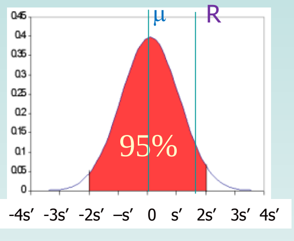
Comparison of the μ average of a sample with a reference value R, it is assumed that the distribution is normal.*
Hypothesis: « the difference between the mean of n measurements and a reference value is not due solely to random errors »

This hypothesis is retained if its probability is greater than 95% (for example)

n measurements: mean μ, estimated standard deviation: ***s***

The estimated standard deviation **of the mean** is shown (down…) to be:

$$s' = \frac{s}{\sqrt{n}}$$



Hypothesis retained if:

$$|\mu - R| \geq t\frac{s}{\sqrt{n}}$$ *(t=Student's coef.)*

if s can be calculated « accuratly »:

	H1: 95% (or H2: 5%)	H1: 99% (or H2: 1%)	H1: 99.9% (or H2: 0.1%)
t	1.96	2.56	3.28

one-sample T-test, example

*Hypothesis:"The difference between the mean of n weighing of a standard mass of 1 kg and the value of this mass **is not** due only to random errors".*

There is a "bias"...

?

Probability > 95%!
hypothesis retained!

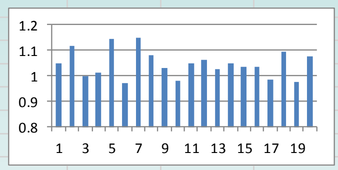
$$|\mu - R| \geq t.\frac{s}{\sqrt{n}}$$

True!

$$t = \frac{|\mu - R|}{\frac{s}{\sqrt{n}}}$$

1	1.04946993
2	1.11917309
3	0.99865798
4	1.01040713
5	1.14655592
6	0.97054733
7	1.14894136
8	1.08005111
9	1.03044121
10	0.98202478
11	1.05094267
12	1.06252219
13	1.02608401
14	1.04961426
15	1.03483134
16	1.03475702
17	0.98306671
18	1.09287381
19	0.97490694
20	1.07697103

Standard R (l	1.0000	
mean μ:	1.04614199	
Standard dev	0.05326172	
$\lvert\mu - R\rvert$	0.0461	
$t.\frac{s}{\sqrt{n}}$	0.02381937	(t=2)
		(n=20)



two-sample T-test

Comparison of 2 means of 2 samples, it is assumed that the distribution is normal.
*Hypothesis:"The **difference** between the two **means is not** just due to random errors."*

This assumption is used if its probability is greater than 95% (for example)

	Samp 1	Samp 2
Nb meas.	n_1	n_2
Est. Std. Dev.	s_1	s_2
Est. mean	μ_1	μ_2

Comparison of μ_1 and μ_2:
***Can be reduced to a 1-sample-test** by comparing $|\mu_1-\mu_2|$ to 0*

We use then as standard deviation of $|\mu_1-\mu_2|$:

$$s = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

(demonstrated in part "error propagation laws")

Hypothesis retained if: $|\mu_1 - \mu_2| \geq t.s$ $\quad |\mu_1 - \mu_2| \geq t\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

And the number of degrees of freedom is a first approximation: n_1+n_2-2*

(otherwise: $df = \frac{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$ *)!*

21

* Number of degrees of freedom (DOF)

*The number of degrees of freedom is the number of **independent** observations in a sample of data that are available to estimate a parameter of the population from which the sample is drawn.*

- Ex, n random variables: $X_1, \ldots, X_n$
  - Mean of the sample: $\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}$
- « Residuals » (« errors X_i - µ») : $X_i - \overline{X}_n$ $\quad \sum = 0$

If n-1 known residuals, we can know the last one…, so n-1 DOF! So for the calculation of the « residual » parameter, the number of DOF is n-1, same for standard deviation.

We must have DOF>0, so if n-i DOF, n=i+1 will be the minimum size for estimating the parameter.

Ex: calculation of a standard deviation: possible from n=2, no sense if n=1

https://en.wikipedia.org/wiki/Degrees_of_freedom_(statistics)

22

two-sample T-test: example

Hypothesis:"The difference between the averages of weighing of the same mass per 2 operators is not due only to random errors".

There is a "bias between the 2 operators"...

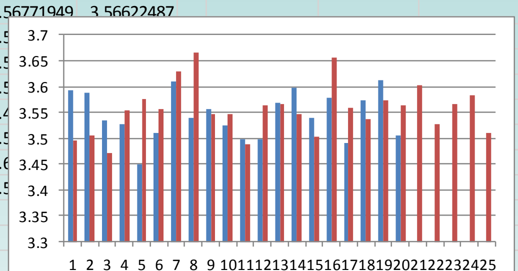
?

Probability < 95%! hypothesis rejected!

$$|\mu_1 - \mu_2| \geq t.\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

False!

op1	op2			op1	op2
3.59385514	3.49673556		mean µ	3.54463973	3.55551443
3.58810682	3.50481358		std. Dev. s	0.04510027	0.04809217
3.53371004	3.47201328				
3.52760094	3.553306			|µ1-µ2|	**0.0108747**
3.44867115	3.57512954				
3.51024893	3.55717682				
3.60880932	3.62969549				
3.53927667	3.66487781		$s = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$		0.0139361
3.55719746	3.54702026				
3.52552333	3.54598904		n:	43	t=2
3.49798625	3.48741103				
3.49706929	3.56409401		s.t:	**0.02787228**	
3.56771949	3.56622487				
3.5					
3.5					
3.5					
3.4					
3.5					
3.6					
3.5					
	3.58384532				
	3.50903522				



Confidence interval

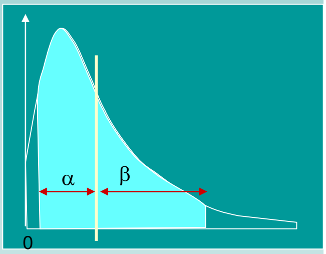
Intervalle de confiance

The confidence interval at *p* (95%) of an x-measurement is an interval of values that has a centered probability *p* (95%) of containing the true x_R value of the estimated parameter.

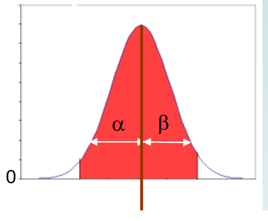
$$\left[x-\alpha, x+\beta\right]$$

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

α et β = uncertainty at *p* (95%)



The calculation of α and β as a function of probability P (usually 95%) depends on the distribution law of the error.



Symetric distribution:

- α=β
- Mean value of the measurement = the more probable

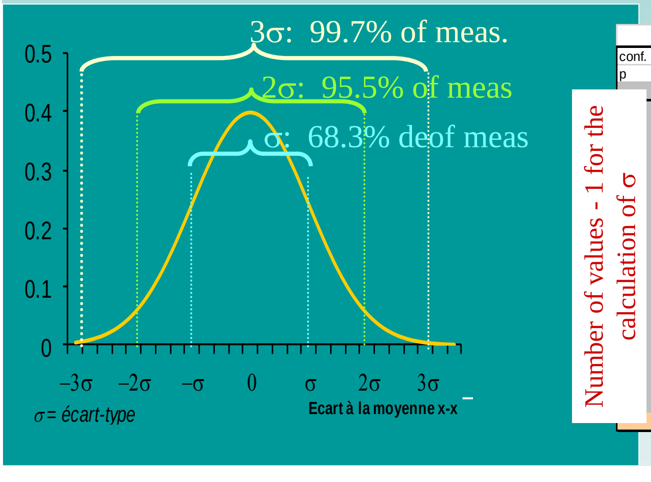
→ Normal distribution

24

Confidence interval, case of normal distribution:

The confidence interval at p (95%) of an x-measurement x is an interval of values that has a probability centered p (95%) of containing the true xR value of the estimated parameter.

$$\left[x - t.s, x + t.s\right]$$



Number of values - 1 for the calculation of σ

	Student's coefs t				
conf. interval	90.0%	95.0%	98.0%	99.0%	99.9%
p	0.1	0.05	0.02	0.01	0.001
deg freedom					
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

25

Writing a measurement (normal distribution)

The result of a measurement must have 4 elements:

Ex : C_{NO} = 125.3 ppb ± 1.7 ppb (à 95% *ou* k=2)

1 2 3 4

1 : **Numerical value with a correct number of decimals**
2 : **Unit**
3 : **expanded uncertainty** = t.s
4 : **Coverage factor t**

26

Evaluation of uncertainty* :

Evaluation by statistical analysis of series of measurements ("Type A") (usually measurement, but also simulation)

Calculation evaluation of the effect on the final uncertainty of the different sources of uncertainty *("Type B":"by any other means"!)*, **themselves evaluated:**

With a « type A » method,

by manufacturer data, calibration etc...

It is then necessary to know the laws of propagation of uncertainty...

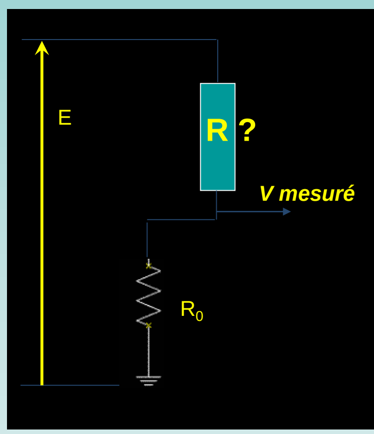
***GUM : Guide to the expression of Uncertainly in Measurement (1993)**

Norme Française X 07-020 AFNOR Août 1999

27

Evaluation by statistical analysis of series of measurements

Example: Evaluation of resistance measurement:



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

Measurement and calculation with a large number of N "standard" resistors (R known)

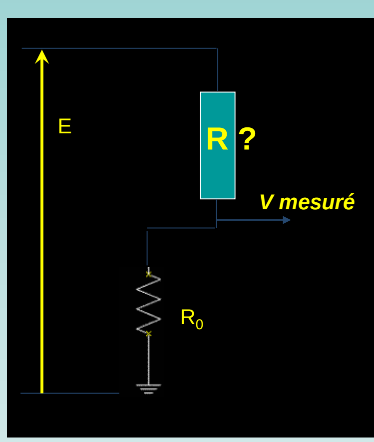
Calculation of standard deviation σ

95% uncertainty = t.σ, # 2σ if N>15

Ignoring:

- ***The uncertainty of the Vmeter,***
- ***Uncertainties about E, R_0, standard resistors...***

Calculation-based assessment of the effect on the final uncertainty of the different sources of uncertainty



Example: measurement of a resistor:

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

σ(R_0), σ(E), σ(V) known

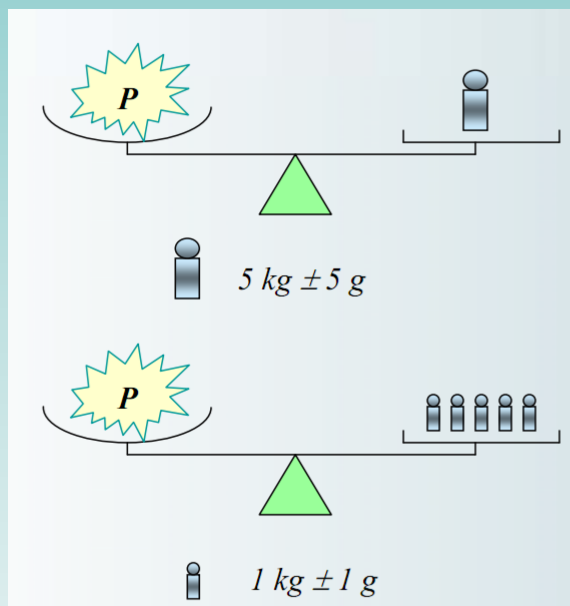
combined standard uncertainty

Loi de propagation des écart-types

$$\sigma(R) = \sqrt{\left(\frac{E-U}{U}\sigma(R_0)\right)^2 + \left(\frac{R_0}{V}\sigma(E)\right)^2 + \left(\frac{R_0E}{V^2}\sigma(V)\right)^2}$$

explanation

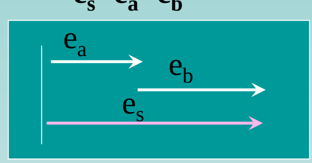
What is the most accurate weighing?



combined standard uncertainty for random & independant errors (1)

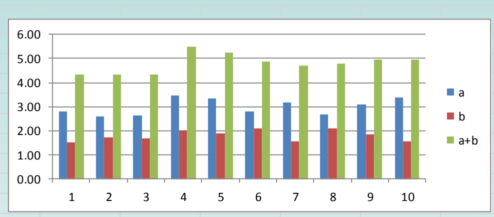
Example of the sum: **s=a+b**

Reminder: errors are added algebraically: $e_s = e_a + e_b$



And Standard deviations?

	a	b	a+b	
	2.82	1.51	4.33	
	2.61	1.71	4.32	
	2.64	1.68	4.32	
	3.45	2.02	5.48	
	3.32	1.90	5.23	
	2.79	2.08	4.88	
	3.17	1.55	4.72	
	2.70	2.09	4.79	
	3.10	1.86	4.96	
	3.40	1.56	4.95	
mean	**3.00**	**1.80**	**4.80**	**4.80 sum of means**
variance	0.11	0.05	0.16	**0.16** sum of variances
Std. Dev.	**0.33**	**0.23**	**0.39**	**0.55 sum of standard deviations**



combined standard uncertainty for random & independant errors (2)

Example of the sum: s=a+b

***Random error*:**

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

Standard deviations are not added!!

$$Var(a+b) = Var(a) + Var(b)$$ ***Variances do!!****

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

Standard deviations are added quadratically*

****If the errors are independent and random***

Addition of variances, demonstration:

Example of the sum: s=a+b

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

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

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

=0 if independent errors

Variances can be added,

And for standard deviations?

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

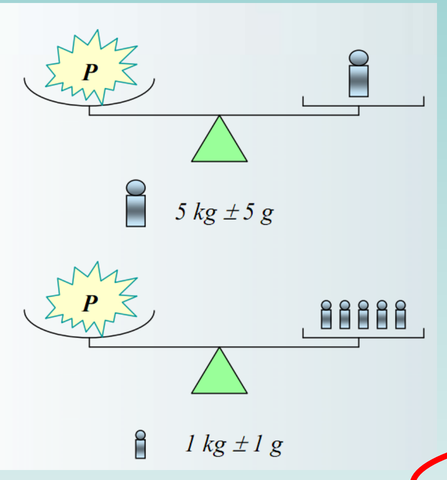
Quadratic **sum of standard deviations**

Uncertainties: si I=t.σ, $I(a+b) = \sqrt{I(a)^2 + I(b)^2}$

(if same distributions...)

33

What is the most accurate weighing?



⇒ **Result of measurement :**

$\boldsymbol{P = 5\ kg \pm 5\ g}$

⇒ **Result of measurement :**

$$P = P_1 + P_2 + P_3 + P_4 + P_5$$

$$c_{i=1...5} = 1$$

$$s(P) = \sqrt{\sum_{i=1}^{5} s(P_i)^2} = \sqrt{5 \cdot (0.5)^2} = 1.12\,g$$

$\boldsymbol{P = 5\ kg \pm 2.24\ g}$

combined standard uncertainty for random & independant errors (3)

Fundamental pplication in instrumentation: **The mean**

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

Theorem of the mean

If a series of measurements has a random and independent standard deviation s error, then the mean of n of these measurements has a random error whose standard deviation is divided by square root of n.

Demo 35 Exc

Reminder: law of propagation of a "small" error

$$y = f(x) \qquad f(x+\Delta x) = f(x) + \frac{df}{dx}.\Delta x$$

$$\Delta y = \frac{df}{dx}.\Delta x$$

Any function: $y = f(x_1,..x_i,..)$

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

Special case: product-power, logarithmic derivative...

$$y = A\prod_i x_i^{\alpha_i} \quad \Rightarrow \quad \frac{\Delta(y)}{y} = \sum_i \left(\alpha_i \frac{\Delta(x_i)}{x_i}\right)$$

Relative error

NOT VALID FOR STANDARD DEVIATION!!

Law of propagation of variance and standard deviation

Any function: $y = f(x_1,..x_i,..)$

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

The variance is additive, if the variables x_i are independent,"we show that":

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

Std. Dev. **: V(x)=(σ(x))²** $$\sigma^2(y) = \sum_1^n \left(\frac{\partial f}{\partial x_i}\right)^2 \sigma^2(x_i)$$

General case (non-independent variables):

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

Uncertainty Propagation Laws (1)

Any function: $y = f(x_1, .. x_i, ..)$

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

Uncertainty for normal distr. : I(y)=t.σ(y), generally, t~2

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

If:

- ***same coefs t***
- ***Indépendant variables x_i***

Uncertainty Propagation Laws (1)

Products & powers :

Interesting property of relative standard deviations: $\frac{\sigma(x)}{x}$

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

*Quadratic sum of **relative** standard deviations*

Example: $y = \frac{a.b.c^3}{d^{1/2}}$

$$\Rightarrow \frac{\sigma(y)}{y} = \sqrt{\left(\frac{\sigma(a)}{a}\right)^2 + \left(\frac{\sigma(b)}{b}\right)^2 + \left(3\frac{\sigma(c)}{c}\right)^2 + \left(1/2\frac{\sigma(d)}{d}\right)^2}$$

Calculation of uncertainty by simulation of random error or Monte Carlo method

Ex: $y=f(x_1,x_2)$,

x_1 and x_2 have random errors (not necessarily independent) characterized by their standard deviations s_1 and s_2

- Simulate a certain number of experiments with the same values x_1 and x_2 + random error ($\neq$ drawings)

$y_i=f(x_1+\varepsilon_1,x_2+\varepsilon_2)$ → ***N times***

- The ε_i are calculated using a random number generator with adequate distribution (usually normal*)

- **Calculation of the standard deviation of the y_i**

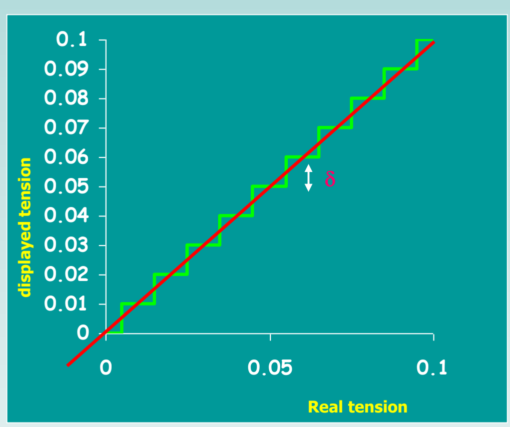
**Excel: =NORMSINV (RAND();0;std dev)*

TP4 **Excel: =LOI.NORMALE.INVERSE(ALEA();0;écart-type)*

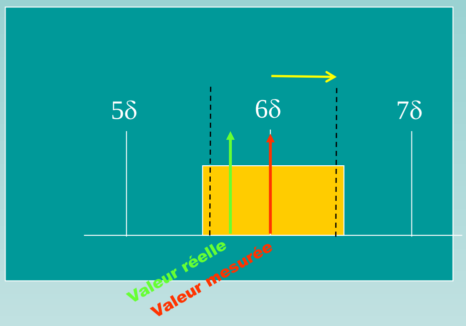
Discretization uncertainty

L'incertitude due à la discrétisation

- **Discretization occurs when the final values constitute a discrete set.**
- ***Resolution* δ of the acquisition system, not to be confused with *uncertainty* or *sensitivity*.**



Discretization uncertainty (2)



Maxi error: δ/2

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

Standard deviation of the error due to discretization:

$$\sigma_d = \frac{\delta}{2\sqrt{3}}$$

95% uncertainty due to discretization: $2\sigma_d$?

No!

$$I_d^{95\%} = 0.975\,\delta/2$$

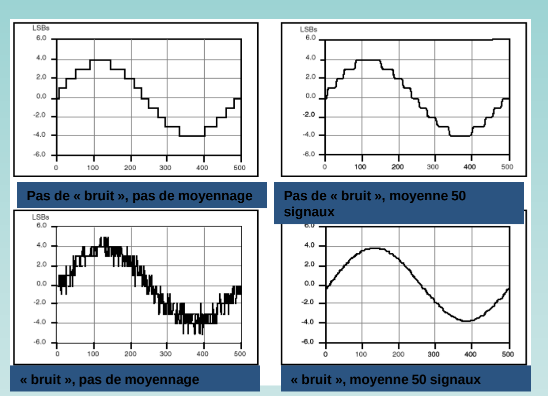
0.58δ

0.485δ

$$I_d^{95\%} \approx \delta/2$$

The « Dithering »

The averaging method can also reduce the discretization error (= quantification):



You can gain 1 resolution bit every time you multiply the sampling rate by 4…

… provided that the signal before digitizing contains a noise (random and independent) of RMS value higher than the initial resolution…

*The presence of random error here allows to improve the "accuracy" **of the measurement!!***

demo

Discretization uncertainty : (3) significant decimal places

Writing a number with a finite number of digits → discretisation error

For a measurement, the standard deviation of the introduced error must be small compared to the standard deviation of the initial measurement error.

But digital systems can give a lot of significant numbers...

Ex: measure = 438.2659872 Std. Dev. mesure = 0.55 (0.125 % in relatif)

Must we write:

438.2659872 ?

438.265 ?

438.3 ?

438?

Discretization uncertainty : (4) significant decimal places, example

Ex: measure = 438.2659872 Std. Dev. measure = σ(x)= 0.55 (0.125 % in relatif)

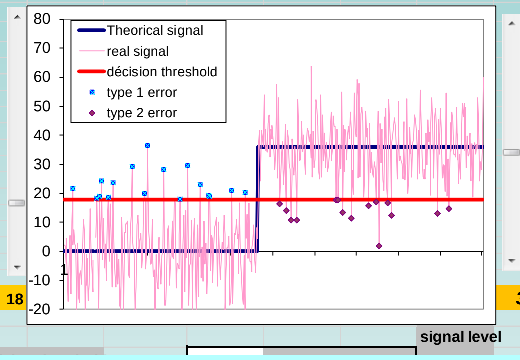
Displayed meas. ***resolution***	S.D. discretisation error $I_d^{95\%} = 0.95\,\delta/2$	S.D. final error $\sigma(X) = \sqrt{\sigma^2(x) + \sigma_d^2}$	% of error due to discretisation
438.265987 ***0.000001***	2.4 10-7	0.55	< 10^{-3} %
438.27 ***0.01***	2.4 10-3	0.550008	10^{-3} %
438.3 ***0.1***	0.024	0.5506	0.14 %
438 ***1***	0.24	0.62	*13 %*
440 ***10***	2.4	2.94	*434 %*

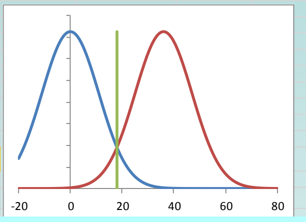
- ***You can choose the resolution immediately smaller than the standard deviation of the measurement.***
- ***The standard deviation (or uncertainty) displayed must not exceed 2 significant digits.***

Limit of detection (LOD)

Detect = determine if measurement is zero or not

LOD = Smallest non-zero value detectable with a certain probability of error, equalizing the risk of false >0 and false <0





If normal distr. and probability dofe 95%:

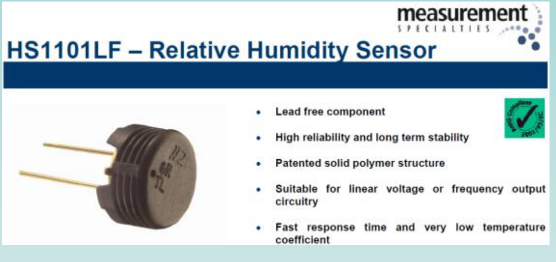
$LOD=3.3\sigma$

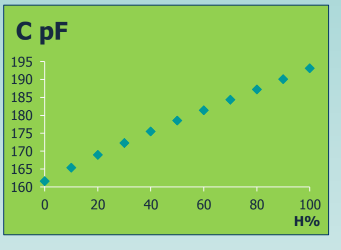
We often take: $LOD=3\sigma$

σ = standard deviation of the "close to zero" measurement

Calibration of a measuring system

Calibration:"Operation consisting in establishing a relationship between the values of the quantity indicated by a measuring instrument and the corresponding values of the quantity produced by **standards" (VIM)**





POLYNOMIAL RESPONSE OF HS1101LF

$$C\ (pF)=C@55\ \%*(3.903\ 10^{-8}*RH^{3}-8.294\ 10^{-6}*RH^{2}+2.188\ 10^{-3}*RH+0.898)$$

REVERSE POLYNOMIAL RESPONSE OF HS1101LF

$$RH\ (\%) = -3.4656\ 10^{+3}*X^{3}+1.0732\ 10^{+4}*X^{2}-1.0457\ 10^{+4}*X+3.2459\ 10^{+3}$$

With X=C(read) / C@55%RH

Sensors & Instruments

« variables transformer »

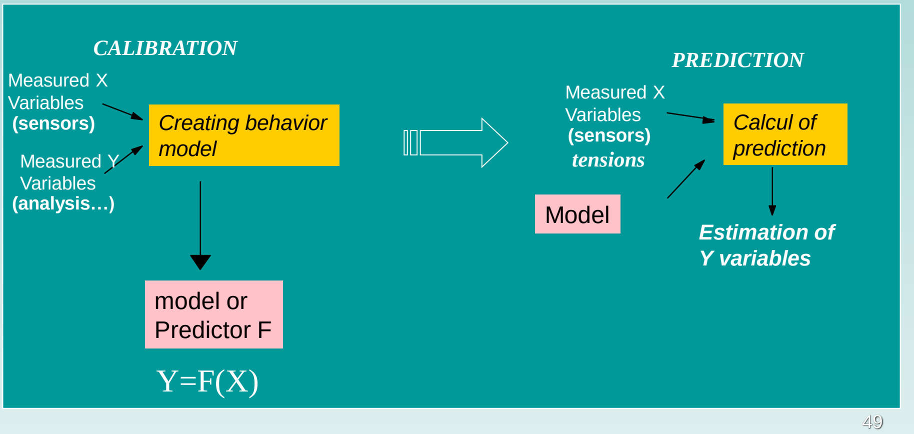
Input ("Yvariables"):
- **System status variable**
- **Difficult to measure**
- **(ex: pressure…)**

Output ("Xvariables"):
- **Variable output from sensor**
- **easy to measure**
- **(e. g. electrical signal…)**

48

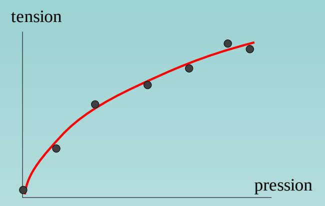
Calibration: Creation of a behavior model

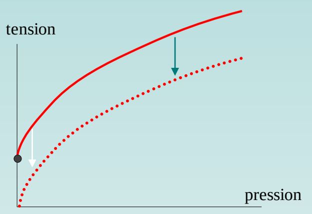
Etalonnage: Création d'un modèle de comportement



49

Calibration - Gauging (zero adjustement)





- **Calibration (*Etalonnage!*)** The complete calculation of the model requires a "large" number of experiments.

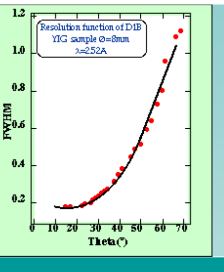
- **Gauging or zero adjustement (*calibrage*) = The "realignment" of the model requires 1 experiment ("zero"), or even 2 ("zero + gain").**
  - •Correction of manufacturing variations
  - •Drift correction
  - •Correction of experimental conditions (temperature, balance calibration...)
  - •Etc...

If the model is not linear, the zero adjustment can only correct small variations...

50

The least Squares

Les moindres carrés



Least squares method (monovariate):

We want to model: $\hat{y}_i = f(x_i)$, in fact, based on experiments, we have: $y_i = f(x_i) + e_i$ *where e_i= "residual" or error.*

$$e_i = y_i - f(x_i) = y_i - f(x_i, \theta_1..\theta_j)$$

Minimal, for the set of calibration measurements, by finding the optimal funtion f with parameters $\theta_1..\theta_j$

θ_j represents the parameter(s) of the function to be optimized, f (ex: coefs polynôm)

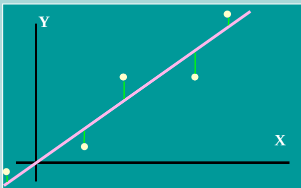
We usually use the Euclidean distance:

$$\sum_i \left(y_i - f(x_i, \theta_1..\theta_j)\right)^2 \quad \text{« minimale »}$$

Linear regression

La régression linéaire

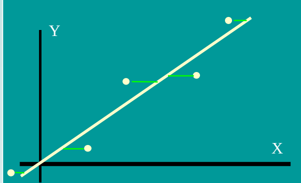
- Monovariate (y=f(x)) + linear hypothesis (y=ax+b)
- 2 cases:



Minimisations of errors on Y

Regression of y/x:

$$y = ax + b + \varepsilon_y$$



Minimisation of errors on X

Regression of x/y:

$$\hat{x} = a'y + b' + \varepsilon_x \Longrightarrow \hat{y} = \frac{1}{a'}x - \frac{b'}{a'}$$

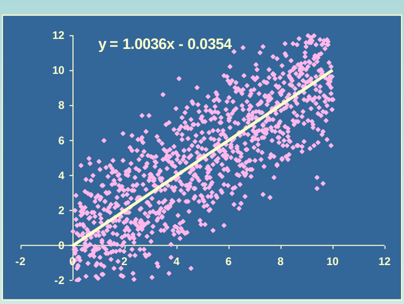
Attention, a regression of y/x with errors on x (or vice versa) leads to a systematic error on the slope...

52

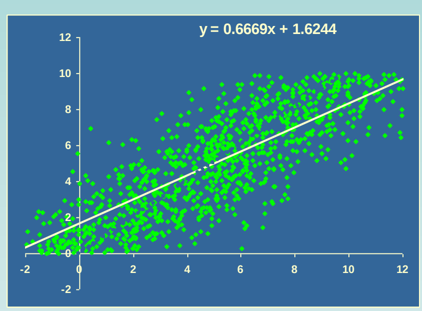
Linear regression: y/x ou x/y?

Attention, a regression of y/x with errors on x (or vice versa) leads to a systematic error on the slope...

Regression Y/X



Regression X/Y



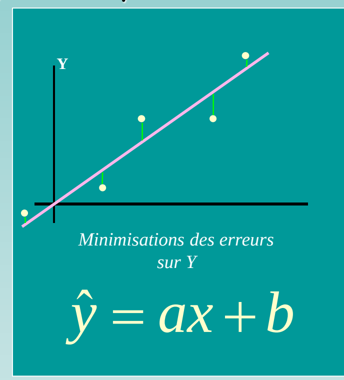
Negligible difference in most cases...

53

Characteristics of regression y/x:

- Estimations of a and b:

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

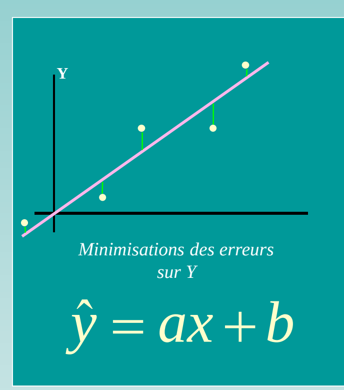


$$b = \bar{y} - a\bar{x}$$

Excel calculation using:
- Trend curve
- LINEST (French, droitereg)

Measurement-related error:

Error that we would have even if the model were perfect, because of the random error on x...



Standard deviation of residuals:

$$s_{y/x} = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{n-2}}$$ (function linest)

S.D. of the measurement error
= S.D. of prediction that we would have (same conditions) if the model was error-free!

But the model is never error-free!

Modelisation error:

The model $(\hat{y} = ax + b)$ *can't be perfect since it is made from experiments…*

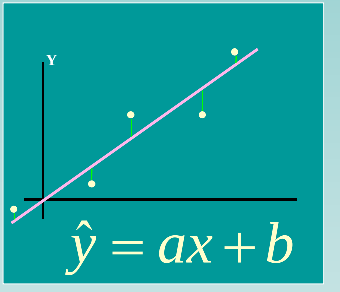
- Estimation uncertainties of a and b:

Standard deviation of the slope a:

$$s_a = \frac{s_{y/x}}{\sqrt{\sum_i (x_i - \bar{x})^2}}$$

Standard deviation of the intercept b:

$$s_b = s_{y/x}\sqrt{\frac{\sum_i x_i^2}{n\sum_i (x_i - \bar{x})^2}}$$



Excel calculation using the function « linest »(french « droitereg »)

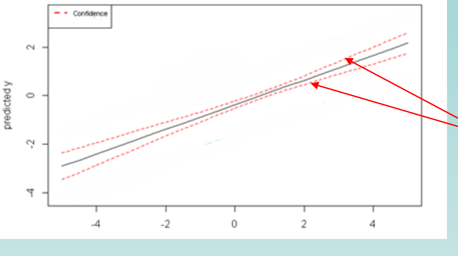
Final error = "Measurement-related error" + "Model error"

$$\textbf{Final uncert.} = \sqrt{(\textbf{Final error} \;=\; \text{"}\textbf{Meas.}-\textbf{related uncert.}\text{"})^2 + (\textbf{Model error uncert})^2}$$

56

Uncertainty of prediction of a linear model

I(pred)²=I(meas)² + I(modelisation)²



Meas. *Model*

$$I(\hat{Y}) = t.s_{y/x}\sqrt{1 + \frac{1}{m} + \frac{(X-\bar{x})^2}{\sum (x_i - \bar{x})^2}}$$

- m= nb calibration samples
- x_i = calibration values, which mean is $\bar{x}$
- X = X Value of prediction sample
- Sy/x: = Standard deviation of residuals

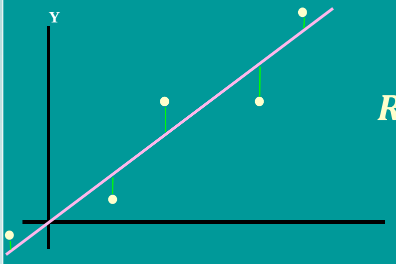
- Depends (a little…) on X
- If « perfect » model (or m great), $I(\hat{Y}) = t.s_{y/x}$
- Uncertainty increases if X goes away from $\bar{x}$ (problem of extrapolation)

57 TP6

Measuring the fit quality of a linear regression:

Var(Y) = Var(aX+b) + Var(ε)

Total information = modelised information + residual information



***Coef. of determination*:**

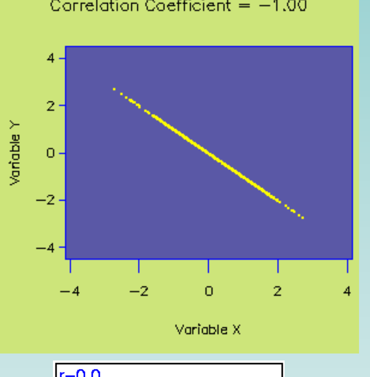
$$R_d = \frac{Var(modelised\ info)}{Var(total\ info)}$$

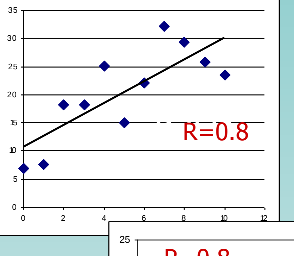
Correlation coef.

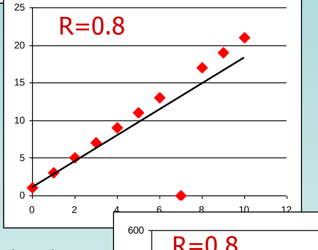
$$R_c = \frac{Cov(X,Y)}{\sigma(X)\sigma(Y)}$$

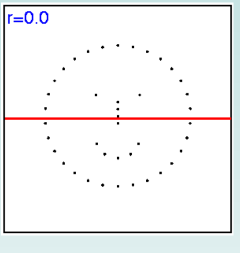
$R_d = R_c^{\,2}$

The correlation coef. Usage:

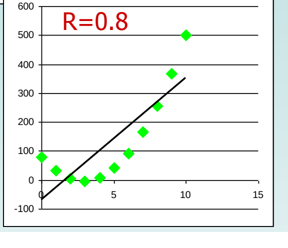




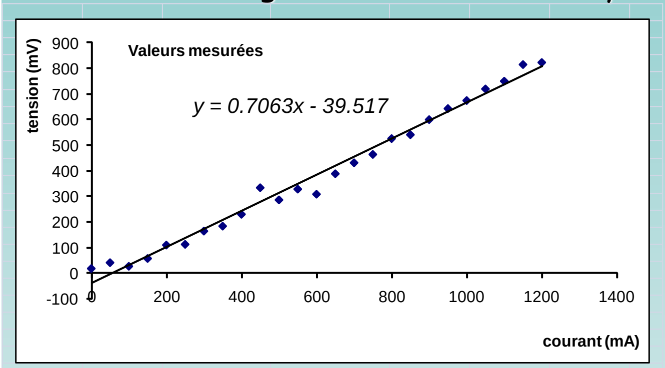


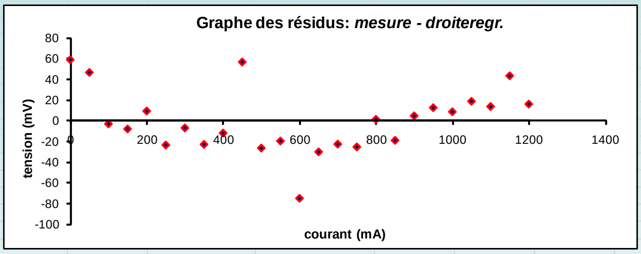


The correlation coefficient is only used to quantify the linear relationships between X and Yvariables with random errors.



Linear regression: residuals analysis





Residuals = information not modelised, ideally random error

$$R_i = y_i - (a.x_i + b)$$

Detection of outliers

To be removed after verification and with care...

Verification of linearity

Possibly to be corrected with non-linear regression

60

Other behavioural modelling methods:

- Multivariates:
  - Linear multivariate
  - Multivariate regression, principal components analysis
  - Partial Least Squares (PLS)
  - Neural Network
- Non-linear monovariate
- Polynomial, power distribution, any function

61

Linear multivariate regression, simplified approach:

- n Xvariables, p experiments (or samples)

Écriture matricielle pour les p expériences: **Y= A.X+ε_y** To be minimised

$$\begin{cases} y_1 = a_1 x_{1,1} + \ldots + a_n x_{n,1} + \varepsilon_1 \\ y_p = a_1 x_{1,p} + \ldots + a_n x_{n,p} + \varepsilon_p \end{cases}$$

- **p<n: no solution…**
- **p=n: maybe a solution:** **A= Y.X^{-1}** Residual ε_y nul…

(Ex: 2 points to determine 1 straight line…)

- **p>n: X not square…**

But maybe even a least-squares solution!

A=Y.X^T.(X.X^T)$^{-1}$ ***$X^T.(X.X^T)^{-1}$ is called« pseudo-inverse » of X***

•Excel « LINEST » function

Points developed at the end of TB…

62

Non-linear regression (simplified…)

model Y=F(X), F not linear

1: Linearization of the problem by variable change

Beer-Lambert: I(λ)=I$_0$(λ)exp(β(λ).C), non linéaire...

devient linéaire si on prend comme Xvariables les absorbances A(λ)=-log$_{10}$(I(λ)/I$_0$(λ))

2: polynomial regression

"Any continuous function can be approximated using a polynomial"

Principle: add quadratic variables to obtain a linear combination, and apply linear methods

ex: Y=f(x_1, x_2) not linear can be « approximated » by:

$\mathbf{Y=a_{1,1}x_1+a_{1,2}x_2+a_{2,1}x_1^2+a_{2,2}x_2^2+b_{1,2}x_1x_2}$ etc…

*x_1, x_2, x_1^2, x_2^2 et x_1x_2 are **linearly** independant…*

Multilinear regression…

Excel trend curve

63

Biblio & links

Books:

"Statistics for analytical Chemistry" 3 rd ed. J.C. Miller and J.N. Miller John Wiley & Sons, 1998, la Bible…

"Multivariate Statistical Methods, A Primer" B.F.J. Monley, Chapman & Hall 1986.

"Modélisation et estimation des erreurs de mesure", M Neuilly, CETAMA, Lavoisier, Paris 1993.

Websites:

http://physics.nist.gov/cuu/Uncertainty/index.html: An excellent, simple and succinct site, but a reference on the calculation of uncertainties, this document was very much inspired by it. A more complete statistics course is available at : http://www.itl.nist.gov/div898/handbook/

http://www.deltamu.fr/Publications: Deltamu, SMEs in metrology offering many resources

http://statpages.org/javasta2.html: Collection of links and especially of Freewares on statshttp://www.jybaudot.fr/General/indexstats.html: Website of Jean-Yves Baudot with precise faith, popularizer and fun, a challenge in statistics!