Chapitre 1

Introduction sur la théorie d'erreurs

Buts

1. savoir traiter des erreurs de mesure
2. comprehension de certaines notions de statistique
3. traiter et interpréter des mesures(Excell)

1.1 Erreur de mesure

Une quantité physique peut être mesurée mais sur chaque mesure il y a un écart, peu importe la précision de votre instrument. Il y a un certain niveau de confiance pour chaque mesure. La valeur exacte de votre mesure se trouve dans un interval de confiance. Nous notons une mesure avec son interval de confiance B de manière suivante

$$x \pm B_x$$

Si on fait des calculs avec ces mesures le résultat aura un écart sur la valeur réelle. L'écart se propage dans les calcules.

On definie une

1. erreur absolue : l'écart maximal qui existe entre la valeur mesurée et la valeur réelle. Pour une mesure on prend l'échelle de l'instrument.
2. erreur relative : l'erreur absolue divisée par la valeur de la mesure

Exemple : Nous mesurons la largeur d'une feuille A1 à l'aide d'une régle classique. Nous mesurons 21.1 cm et la division de la régle est en mm. L'erreur est donc 1mm et nous notons

1. resultat de mesure : 21.1 ± 0.1 cm

2. erreur absolue : AF=1mm
3. erreur relative : RF=0.0473933 of 4.73933%

Il faut encore mentionner qu' il faut arrondir mais que l'arrondi est de nouveau une source d'erreurs qui aura une influence sur les résultats.On arrondi deux chiffres considérables dépendant de la précision de la mesure.

exemple :

Nous mesurons 15.329m avec une intervalle de confiance B=0.26m. On va noter le résultat comme 15.33 ± 0.26 m

Il y a deux types d'erreur

1. **erreur systématique** : Une erreur inhérent de l'experience comme par exemple une lecture fausse par parrallax ou une mauvaise calibration de l'instrument. Ce type d'erreur ne peut être trouvée que en comparant le résultats avec des résultats d'autres rechercheurs ou refaire l'experience avec d'autres instruments ou rechercheurs.
2. **erreur aléatoire** : Si cette erreur est irrégulière, on apelle ceci un bruit sur la mesure. On peut compenser par refaire plusieurs fois l'experience et calculer la moyenne.Les mesures auront une étendue autour de la moyenne.

Il faut encore definir deux notions qui donnent de temps en temps beaucoup de confusion

1. **precision**(eng. accuracy) : est une notion qui représente la différence entre la valeur réelle et la valeur mesurée. Cette notion est relaté á l'erreur systématique.
2. **fidélité**(eng.precision) :est une notion pour l'erreur aléatoire et est désignée par l'erreur-type.

1.2 Calcule de l'intervalle de confiance

1.2.1 Fonction connue

Nous avons une fonction qui designe la quantité que nous voulons mesurer. (cf. une volume) Nous allons calculer le mesure et l'interval de confiance á l'aide de cette fonction. Attention, dans ce cas nous supposons que les variabeles sont indépendant unes des autres, sinon il faut prendre la correlation des paramètres dans les calcules.

$$y = f(x_i)$$

l'intervalle de confiance devient

$$B_y = \sqrt{\sum_{i=1}^{n} (\frac{\partial f}{\partial x_i} B_{xi})^2}$$

De ceci suivent un nombre de régles de calculs

1. somme et différence :

$$y = ax_1 \pm bx_2$$

devient

$$B_y^2 = a^2.B_{x1}^2 + b^2.B_{x2}^2$$

2. produit en quotient :

$$RF_y^2 = RF_{x1}^2 + RF_{x2}^2$$

et l'intervalle de confiance devient

$$B_y = y.RF_y$$

3. exposant :

$$y = a.x^n$$

et

$$RF_y = |n|.RF_x$$

4. sinus :y=a. sinx et

$$B_y = |a.cosx|.B_x$$

(rem. :si y=a.cosx les functions changent)

5. logaritme :

$$y = lnx$$

et

$$B_y = \frac{B_x}{x}$$

6. exponentiel :

$$y = e^x$$

et

$$B_y = e^x.B_x$$

1.2.2 Exercise

Nous mettons deux résistances en parallèle. Calculez la résistance remplaçante et son intervalle de confiance. $R_1 = (4.1 \pm 0.2)k\Omega$ *et* $R_2 = (1.01 \pm 0.01)k\Omega$

1.3 Notions statistiques

1.3.1 Distributions

Chaque mesure est une prise d'un ensemble de valeurs possibles. On apelle cet ensemble une distribution de probabilité.

Ces distributions peuvent être discrétes, il y a donc une certaine 'constatabilité' des éléments.

Ces distributions peuvent aussi être continu donc il n'a plus de 'constatabilité'.

D'aprés le comportement des éléments,donc mesures, on obtient une autre distribution.

Une propriété importante est que les distributions de probabilité sont normalisées. C'est á dire que la somme de toutes les probabilités est 1.

– distribution discrete :

$$\sum_{k=1}^{n} f(k) = 1$$

– distribution continue :

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1$$

Comme mentioné il y a des différentes distributions et chacun a son propre domaine d'application.

1. **Distribution binomiale**

   Si une mesure ne peut que prendre deux valeurs possibles (vrai/faut, 0/1,..) la probabilité que de n mesures k ont la valeur 1, est décrite par

$$f(k) = \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k}$$

2. **Distribution Poisson**

   Cette distribution est utilisée quand on compte des nombres comme le nombre de photons qui passe par un trou pendant une certaine periode ou le nombre de colonies de bactéries dans un plat Petri. Si μ est la moyenne expecté pour le recensement la probabilité que k événements démontrent

$$f(k) = \frac{\mu^k}{k!} e^{-\mu}$$

   Cette distribution est un cas limite de la distribution binomiale et pour des larges nombres de k la distribution approximera la distribution normale.

   Une propriété importante de cette distribution est que l'étendue est de valeur $\sqrt{\mu}$. Elle est beaucoup utilisé pour des processus dépendant du temps.(cf. usure) En Excell on utilise la fonction EXPONDIST .

3. **Distribution normale**
   On retrouve cette distribution régulierement dans la nature et par la thèse de la limite centrale, cette distribution peut être utilisé pour des cas ou on a un grand nombre de mesures.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp(-\frac{(x-\mu)^2}{2\sigma^2})$$

4. **Distribution Chi-quadrat**
   Cette distribution est appliquée pour calculer des intervalles de confiance pour des écarts entre les valeurs mésurées et les valeurs prévues. L'écart-type des mesures doit être connue.
5. **Distribution Student t**
   Cette distribution est appliquée pour calculer des intervalles de confiance pour la moyenne des valeurs mésurées. L'écart-type des mesures ne doit pas être connue. En Excell c'est la fonction TDIST.
6. **Distribution Fisher**
   Cette distribution est appliquée pour comparer deux théories si on a les données.

1.3.2 Fonction génératrice des moments

C'est une méthode générale pour calculer la moyenne et la variance (et encore d'autres proprietées) d'une distribution.

Par definition le moment d' ordre n d'une distribution de probabilité est

$$\mu_n = E[x^n]$$

et le moment centré de ordre n

$$\mu_n^c = E[(x-\mu)^n]$$

avec

$$E[g(x)] = \sum_{k=1}^{n} g(k)f(k)$$

pour une distribution discrete et

$$E[g(x)] = \int_{-\infty}^{+\infty} g(x)f(x)\,dx$$

pour une distribution continue.

Ainsi le moment centré d'ordre trois est lié à l'asymmétrie (skewness) de la distribution et le moment centré d'ordre quatre est lié au kurtosis.

1.3.3 Moyenne

La moyenne $\bar{x}$ est le premier moment centré d'une distribution de probabilité.

Pour une suite de mesures on peut calculer cette moyenne rapidement en Excell avec la fonction 'MEAN'.

Il faut remarquer qu'on peut toujours calculer une moyenne pondérée s'il y a des mesures dans la suite qui sont plus importantes que des autres. On peut attribuer des pondérées à chaque mesure et on calcule la moyenne par

$$\langle g \rangle = \frac{1}{w} \sum_{i=1}^{n} w_i x_i$$

et

$$w = \sum_{i=1}^{n} w_i$$

1.3.4 Médiane

La médiane d'un ensemble est la valeur où il y a autant de valeurs qui se trouvent au dessus qu'au dessous. Pour une distribution F c'est le point m ou F(m)=0.5. Si la distribution consiste d'une nombre impaire de mesures c'est la mesure au milieu i la distribution consiste d'un nombre de mesures paire on prend les deux nombres au milieu ou la moyenne de ces deux nombres.

1.3.5 L'étendue

L'étendue est la racine carré de la variance de la distribution de probabilité. La variance est le deuxième moment centré de sette distribution.

Pour une suite on applique en excell la fonction 'VAR',ensuite on prend la racine de cette valeur.

1.3.6 Observations aberrantes

Observations aberrantes (ang. outliers) sont en faite des brouilleurs en digérant les mesures parce qu'ils vont déformer les résultats malgré que se sont des valeurs veritablement mesurées. Il y a beaucoup de causes pour ces observations mais les resultats des calcules seront déformés et donc des conclusions fauses. On a fait déjà beaucoup de recherche mais nous allons garder les choses simple. Il faut être prudent si on rejete des mesures à base d'un écart trop grand. On accepte comme règle d'or qu'on peut rejeter une valeur à la base de l'écart-type. Le critère doit être choisie qu' une telle mesure aura lieu accidentellement. la limite dépend donc de la nombre de mesures.

Comme régle d'or on prend :

nombre de mesures	**rejeter á**
moins que 10	2.5σ
10 á 50	3σ
plus que 50	3.5σ

1.4 Regression

Si on cherche une liaison entre des résultats, on veut donc une fonction mathématique pour décrire cette liaison. Sur la base des mesures on arrive presque jamais á une fonction claire et net unique mais on voit un nuage de points dans laquelle on calcule cette fonction.

On a trouvé une méthode statistique pour calculer cette fonction de cette nuage de points et cette méthode est apellée regression.

La méthode que nous utilisons est appellée 'la méthode des moindres carrés'. Si on applique cette technique on dessine une fonction arbitraire connue parmi les points et on va minimaliser la somme des distances de chaque point vers la fonction. On va donc minimaliser l'erreur totale sur l'approximation. L'écart peut être calculée par le coefficient de correlation.

Un cas spécial est la regression lineaire. On suppose donc que la fonction qu'on cherche comme liaison entre les paramétres est lineaire.

L'équation de la meilleure droite est dans le cas générale

$$y = a.x + b$$

Il faut donc calculer a et b et pour atteindre ceci on applique les formules suivantes

$$A = \sum x_i.y_i - n.\bar{x}.\bar{y}$$

$$D = \sum x_i^2 - n.(\bar{x})^2$$

$$E = \sum y_i^2 - n.(\bar{y})^2$$

de lesquelles suivent les valeurs pour a et b

$$a = \frac{A}{D}$$

en

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

La correlation r est calculée par

$$r = \frac{A}{\sqrt{D.E}}$$

. Cette chiffre vous dit si l'approximation est assez bien.

En Excell on utilise 'scatterplot'. Puis on choisit 'layout' puis 'trendline' et 'more trendline options'.On indique en bas du fenêtre 'display equation on chart' et 'display R-squared value on chart'.

Si on digére des données on calculera les moyennes de chaque point de surveillance et puis on utilise ces moyennes pour calculer la grafique de regression optimale. On va calculer et utiliser aussi l'écart type á chaque point de surveillance et l'écart type sera mis dans la grafique avec ce qu' on apelle un bar d'erreur. Ce bar d'erreur est mis sur le graphique en excell par 'trendline' puis 'error bar' on descend á 'more options' et on prend 'custom' et on choisi les valeurs des écarts type en 'positive values' et 'negative values'.

1.5 Test d' hypothèse

1.5.1 Introduction général

On calcule d'une suite de mesures la moyenne et la variance mais il faut se rendre compte que ce n'est qu'une approximation de la valeur réelle. Il faut donc calculer la fiabilité des mesures et calcules et on va calculer un intervalle de confiance pour cette approximation. On appelle cette technique tester une hypthèse. On suppose une hypothèse que l'on appelle H_0 et on calcule si on peut accepter cette hypothèse pour un certain pourcentage, nommé la valeur seuil, (99% ou 95%). Puis on compare avec une hypthèse alternative nommé H_1. Ce pourcentage est nommé le niveau de confiance.(eng. confidence level)

*Chaque test peut être calculé au dessous OU au dessus un niveau et on appelle ça un test **unilatéral**. Si on cherche un intervalle de deux valeurs, donc au dessus ET au dessous ou DIFFERENT d'une valeur arbitraire on calcule un test **bilatéral**.*

Le seuil de signification est appellé $(1-\alpha)$. La probabilité avec laquelle on rejete injustement H_0 est donc α.

De ceci suit le tableau suivant

	H_0 **est vrai**	H_1 **est vrai**
H_0 **est accepté**	décision correcte probabilité= $1-\alpha$	décision incorrecte probabilité= β
H_0 **est rejeté**	décision incorrecte probabilité= α	décision correcte probabilité= $1-\beta$

Pour tester de hypothèses on applique une procédure standard

1. Formulez une nulhypothèse
2. Cherchez la variable de test et une distribution adaptée
3. déterminez le niveau de confiance (donc α)
4. Déterminez la zone d'acceptation
5. Déterminez la valeur actuelle de la variable de test et comparez avec la zone d'acceptation
6. Conclusion si H_0 est rejetée

1.5.2 Certains test qui sont beaucoup appliqués

1. **test u**

   Ce test est appliqué pour chercher si la différence entre la moyenne d'une échantillonage $\bar{x}$ et la valeur de référence μ_0 de cette moyenne est significante. Ici la variance de cette moyenne σ et la moyenne de l'échantillonage doit être connue. Le nombre d'elements dans l'échantillonage est n.

   En Excell on applique la function NORMINV.

   La variable statistique est dans ce cas ci

$$u = (\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}})$$

et elle est distribué normale de moyenne. La zone d'acceptation est calculée en Excell par $NORMINV(\alpha;0;1)$

2. **test t**

   (a) **pour une moyenne d'échantillonage**

   Ce test est appliqué pour chercher si la différence entre une moyenne d'échantillonage $\bar{x}$ et une valeur de référence de cette moyenne μ est significative. Dans ce cas ci la moyenne et la variance s^2 de l'échantillonage doivent être connues.

   C'est la situation normale pour les campagnes de mesures !

   En Excell on applique la fonction TINV.

   La variable statistique est dans ce cas

   $$t = (\frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}})$$

   et elle est distribuée student avec une degrée de liberté $\nu = n - 1$. La zone d' acceptation est trouvé avec Excell avec la fonction $TINV(\alpha;\nu)$

   (b) **pour deux moyenne d'échantillonage**

   Dans ce cas il ya deux possibilités , ou les écart-types sont le même (homoscédastiques) ou elles sont différentes (hétéroscédastiques).

   Aussi ici on applique en Excell la function TINV

3. **test khi deux**

   Ce test est appliqué pour chercher si la différence entre une variance déchantillonage et sa valeur de référence est significante. La variance d l'échantillonage est connue.

   En Excell on applique la function CHINV

   La variable statistique est dans ce cas

   $$\chi^2 = \nu \frac{s^2}{\sigma^2}$$

   avec $\nu = n-1$. En Excell on calcule la zone d'acceptation avec $CHINV(\alpha, \nu)$

4. **test F**

   Ce test est utilisé pour chercher si la différence entre deux variances d'échantillonage est significative. Les variances de l'échantillonage sont connues.Les populations *ne doivent pas* avoir la même moyenne.

   En Excell on applique la function FINV.

Α α	alpha	Ν ν	nu
Β β	beta	Ξ ξ	ksi
Γ γ	gamma	Ο ο	omicron
Δ δ	delta	Π π	pi
Ε ε	epsilon	Ρ ρ	rho
Ζ ζ	zeta	Σ σς	sigma
Η η	eta	Τ τ	tau
Θ θ	theta	Υ υ	upsilon
Ι ι	iota	Φ φ	phi
Κ κ	kappa	Χ χ	chi
Λ λ	lambda	Ψ ψ	psi
Μ μ	mu	Ω ω	omega

USEFUL CONVERSION FACTORS AND RELATIONSHIPS

Length

SI unit: meter (m)

1 km = 0.62137 mi
1 mi = 5280 ft
= 1.6093 km
1 m = 1.0936 yd
1 in. = 2.54 cm (exactly)
1 cm = 0.39370 in.
1 Å = 10^{-10} m

Mass

SI unit: kilogram (kg)

1 kg = 2.2046 lb
1 lb = 453.59 g
= 16 oz
1 amu = $1.66053873 \times 10^{-24}$ g

Temperature

SI unit: Kelvin (K)

0 K = −273.15°C
= −459.67°F
K = °C + 273.15
$°C = \frac{5}{9}(°F - 32°)$
$°F = \frac{9}{5}°C + 32°$

Energy (derived)

SI unit: Joule (J)

1 J = 1 kg-m^2/s^2
1 J = 0.2390 cal
= 1 C × 1 V
1 cal = 4.184 J
1 eV = 1.602×10^{-19} J

Pressure (derived)

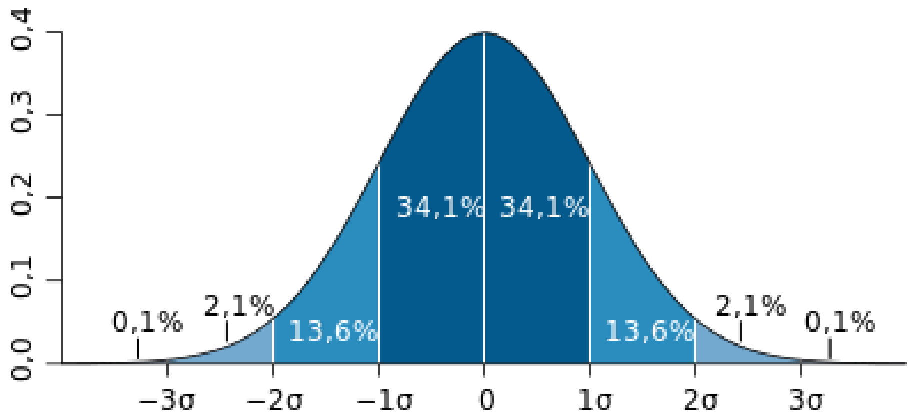
SI unit: Pascal (Pa)

1 Pa = 1 N/m^2
= 1 kg/m-s^2
1 atm = 101,325 Pa
= 760 torr
= 14.70 lb/in^2
1 bar = 10^5 Pa
1torr = 1 mm Hg

Volume (derived)

SI unit: cubic meter (m^3)

1 L = 10^{-3} m^3
= 1 dm^3
= 10^3 cm^3
= 1.0567 qt
1 gal = 4 qt
= 3.7854 L
1 cm^3 = 1 mL
1 in^3 = 16.4 cm^3


0,4
0,3
0,2
0,1
0,0
34,1%
34,1%
13,6%
13,6%
2,1%
2,1%
0,1%
0,1%
−3σ
−2σ
−1σ
0
1σ
2σ
3σ

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.
(Revised March 2003)