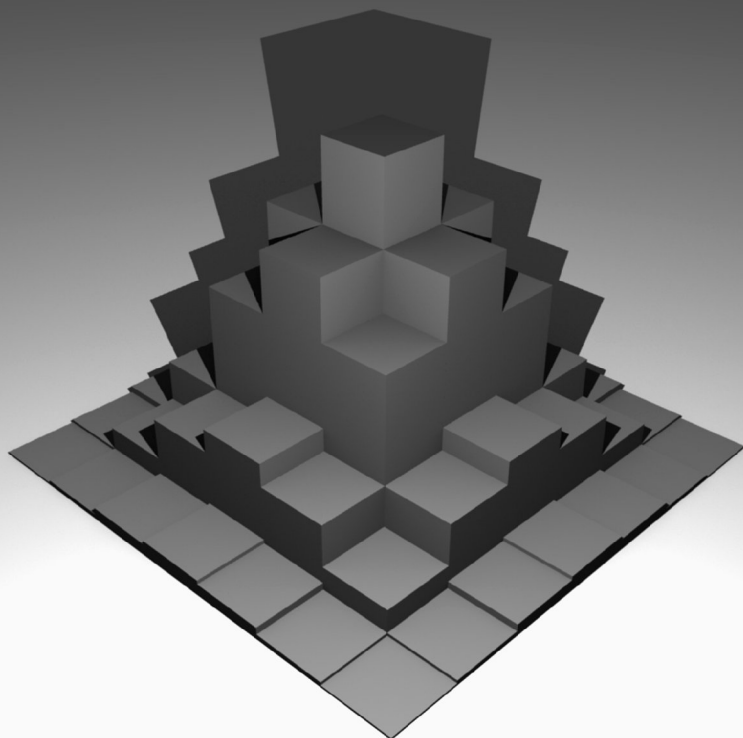


Probabilités, statistiques et analyses multicritères



Mathieu ROUAUD

Probabilités, statistiques et analyses multicritères

Application aux sciences expérimentales
(*exercices corrigés*)



Mathieu ROUAUD

Professeur Agrégé de Sciences Physiques
en classes préparatoires aux grandes écoles
d'ingénieurs.

Diplômé en Physique Théorique.

Vous pouvez reproduire et partager ce livre librement à condition de citer l'auteur et de ne pas en faire d'usage commercial. Mais pas de liberté sans solidarité. **Merci de soutenir mon travail de publication en achetant le livre ou en faisant un don.** Je pourrais ainsi écrire d'autres livres que j'ai en projet. Merci pour votre intérêt porté pour le livre et bonne lecture !

**Faire un
don**

**Acheter le
livre papier**

Pour un don envoyez un chèque à l'ordre de *Mathieu ROUAUD* à l'adresse suivante : Boudiguen 29310 QUERRIEN (FRANCE). Merci beaucoup pour votre aide ! En espérant publier de nouveaux livres. Site internet de l'auteur : www.science-et-decouvertes.fr

Pour un meilleur partage de la connaissance et l'accès au plus grand nombre, le livre est en licence libre, le livre numérique est gratuit et pour minimiser le coût de la version papier, il est imprimé en noir et blanc et sur papier économique.

livre numérique :
www.incertitudes.fr/proba-stat-acp/livre.pdf

livre papier
en auto-édition sur lulu.com

Ce livre est sous licence Creative Commons Attribution-Non Commercial 3.0.



Vous êtes libres :



de reproduire, distribuer et communiquer cette création au public ,



de modifier cette création .

Selon les conditions suivantes :



Attribution. Vous devez citer le nom de l'auteur original de la manière indiquée par l'auteur de l'œuvre ou le titulaire des droits qui vous confère cette autorisation (mais pas d'une manière qui suggérerait qu'ils vous soutiennent ou approuvent votre utilisation de l'œuvre).



Pas d'Utilisation Commerciale. Vous n'avez pas le droit d'utiliser cette création à des fins commerciales.

- A chaque réutilisation ou distribution de cette création, vous devez faire apparaître clairement au public les conditions contractuelles de sa mise à disposition. La meilleure manière de les indiquer est un lien vers cette page web : <http://creativecommons.org/licenses/by-nc/2.0/deed.fr>
- Chacune de ces conditions peut être levée si vous obtenez l'autorisation du titulaire des droits sur cette œuvre.
- Rien dans ce contrat ne diminue ou ne restreint le droit moral de l'auteur ou des auteurs.

Date de parution : juin 2012

Avant-propos

Cet ouvrage se veut accessible et pédagogique. Il est le fruit d'interrogations personnelles sur la nature probabiliste des mesures en sciences. Dans un cursus classique ces aspects ne sont pas, ou peu, abordés. Il est important que les fondements expérimentaux et pratiques des sciences soient complémentaires d'une science au tableau en cours magistraux. Il existe une beauté scientifique qui naît de l'interaction entre la théorie et l'expérience.

Tout en introduisant les principes fondamentaux de la statistique, les deux premiers chapitres expliquent comment déterminer les incertitudes dans différentes situations expérimentales. Beaucoup d'exemples sont issus de cours et travaux pratiques réalisés en math sup.

Le troisième chapitre aborde les analyses multicritères telles qu'elles ont été enseignées en deuxième année d'école d'ingénieur (École Hubert Curien / Maîtrise de l'efficacité énergétique).

Bonne lecture !

Remerciements :

Je remercie Grégoire BONNET (ingénieur charpentier) et Éric NOIZET (professeur Agrégé de Chimie en prépa) pour leurs multiples apports à la clarté pédagogique de l'ouvrage. Un grand merci à Reine pour sa relecture précise et consciencieuse. Pleins de mercis, à Aurélien SEMACH (étudiant) et, aux enseignants de sciences-physiques Françoise MARCADET (pour ses contributions en métrologie) et Julien BONVALET. Merci à la vie et à tous ceux qui m'ont précédé.

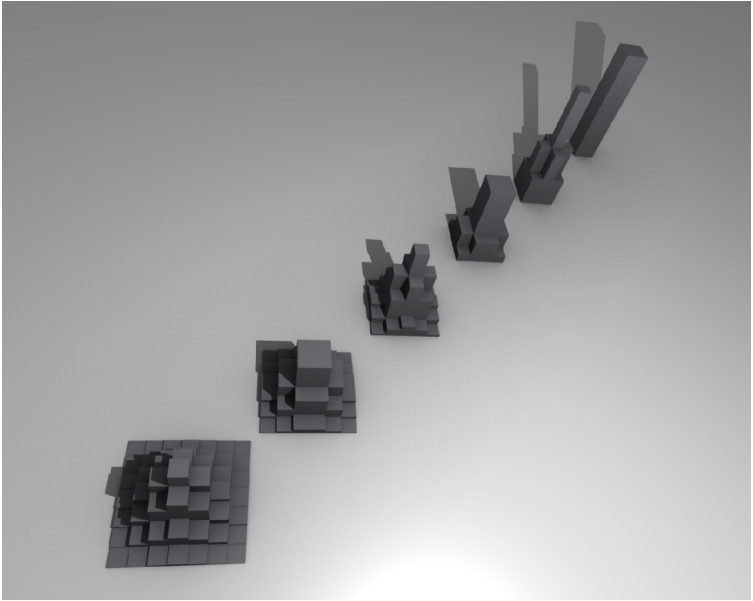


Table des matières

I. VARIABLE ALÉATOIRE.....	1
A. Grandeurs et mesures.....	1
B. Centre d'une distribution.....	1
C. Dispersion d'une distribution.....	3
D. Exemples de distributions.....	4
E. Théorème central limite.....	7
1) Population et échantillons.....	7
2) Le théorème central limite.....	10
3) Coefficient de Student et incertitude.....	12
4) Exemples.....	14
F. Distribution de Gauss.....	18
1) Définition d'une distribution continue.....	18
2) Courbe de Gauss.....	19
3) Loi normale standard.....	21
G. Test d'hypothèse.....	22
H. Test du Khi-deux.....	24
I. Sources des incertitudes.....	27
J. Exercices.....	31
II. CORRÉLATIONS ET INDÉPENDANCES.....	37
A. Coefficient de corrélation.....	37
B. Formule de propagation des incertitudes.....	42
1) Formule de propagation des écart-types.....	42
2) Calcul d'incertitude.....	43
C. Régression linéaire.....	48
1) Principe et formules.....	48

2) Détermination du zéro absolu.....	52
3) Régression avec barres d'erreurs.....	54
D. Exercices	57
III. APPROCHES MULTICRITÈRES.....	67
A. Méthode par pondération.....	68
B. Analyse en Composantes Principales.....	71
1) Principes.....	71
2) Une deuxième illustration.....	84
3) Résumé.....	87
C. Exercices.....	88
IV. COMPLÉMENTS.....	92
A. Mesure avec une règle.....	92
B. Métrologie.....	104
C. Thermodynamique.....	112
D. Indépendance des variables.....	118
V. DEVOIRS.....	120
A. Devoir Suricate.....	120
B. Devoir Narval.....	127
VI. CORRECTIONS.....	134
VII. OUTILS MATHÉMATIQUES.....	164
VIII. Bibliographie / Sources / Logiciels / Illustration de couverture.....	167
IX. TABLES / Index.....	172
A. Coefficients de Student.....	172
B. Valeurs critiques de Khi-deux.....	173



I. VARIABLE ALÉATOIRE

A. Grandeurs et mesures

Soit X une variable aléatoire et n réalisations $\{x_i\}$ de cette variable.

Nous pouvons simplement estimer une grandeur classique : par exemple, combien y-a-t-il de jours dans une semaine ? La réponse est sans ambiguïté. Par contre pour une grandeur statistique l'approche est plus subtile. Imaginons des étudiants qui font des expériences de calorimétrie pour mesurer la capacité thermique de l'eau¹. Les différents groupes mesurent les valeurs suivantes : {5100; 4230; 3750; 4560; 3980} J/K/kg. Que vaut alors la capacité ? Nous donnerons dans ce chapitre une réponse à cette question. Elle sera de nature probabiliste.

B. Centre d'une distribution

Nous cherchons une caractéristique du centre de la distribution des observations $\{x_i\}$. Il en existe plusieurs, le *mode*, par exemple, est facile à déterminer, il s'agit de la valeur la plus représentée (illustrations page 4). Nous

1 PHYSIQUE : Quantité d'énergie à fournir à un kilogramme d'eau pour que sa température s'élève de 1°C. L'eau emmagasine ainsi de l'énergie et peut la restituer par la suite en diminuant sa température. Tables : $c_{\text{eau}} = 4180$ Joules par degré Celsius et par kilogrammes.

avons aussi la *médiane* qui correspond à la valeur qui sépare la distribution en deux parties égales. Mais la plus utilisée est la *moyenne* qui représente au mieux le centre d'une distribution :

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_i + \dots + x_n}{n} \quad \text{soit} \quad \boxed{\bar{x} = \frac{\sum_{i=1}^n x_i}{n}} \quad 2$$

Pour la capacité thermique de l'eau nous obtenons :

$$\bar{c} = \frac{5100 + 4230 + 3750 + 4560 + 3980}{5} = 4324 \text{ J/K/kg}$$

Nous avons considéré la moyenne arithmétique. Nous aurions pu prendre la moyenne géométrique :

$$\bar{x} = \sqrt[n]{\prod x_i}$$

Par exemple, pour deux températures 20°C et 40°C, la moyenne géométrique est $\sqrt{20^\circ\text{C} \cdot 40^\circ\text{C}} \simeq 28,3^\circ\text{C}$ alors que la moyenne arithmétique est 30°C. Dans la pratique on constate que la moyenne arithmétique est mieux adaptée.

2 MATH : se dit «la moyenne de x est égale à la somme de l à n des x_i , le tout divisé par n ». Pour la moyenne géométrique nous considérons la racine $n^{\text{ième}}$ du produit des x_i . \bar{x} , "x moyen", se dit aussi "x barre".

C. Dispersion d'une distribution

Il s'agit d'estimer ce que nous pourrions aussi appeler la largeur d'une distribution. La grandeur la plus simple à déterminer est l'*étendue*, différence entre les valeurs maximale et minimale. Mais celle-ci est très sensible aux valeurs extrêmes qui ne sont pas toujours représentatives, et peuvent même parfois être absurdes.

Dans les faits, la grandeur la plus utilisée est l'*écart-type* :

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

Pour l'écart-type de la capacité thermique de l'eau nous obtenons :

$$s_c = \sqrt{\frac{(5100-4324)^2 + (4230-4324)^2 + (3750-4324)^2 + (4560-4324)^2 + (3980-4324)^2}{4}}$$

$$\text{soit } s_c \simeq 530 \text{ J/K/kg}$$

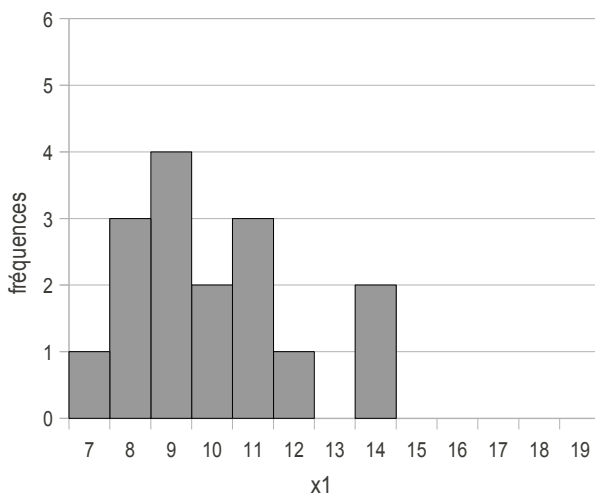
Nous pourrions aussi considérer l'*écart moyen* par rapport à la moyenne (voir l'exercice 1).

Pour l'écart-type si nous divisons par n au lieu de $n-1$, nous obtiendrions l'*écart quadratique moyen*. Le choix de l'écart-type sera justifié par la simplicité des formules qui en découleront. De plus nous travaillons souvent avec n grand et la différence entre les deux types d'écarts quadratiques est alors minime.

D. Exemples de distributions

Cas 1 :

	x_1
x_1^1	11
x_1^2	9
x_1^3	10
x_1^4	14
x_1^5	11
x_1^6	8
x_1^7	9
x_1^8	12
x_1^9	7
x_1^{10}	8
x_1^{11}	8
x_1^{12}	9
x_1^{13}	11
x_1^{14}	14
x_1^{15}	10
x_1^{16}	9



moyenne = 10

mode= 9

médiane= 9,5

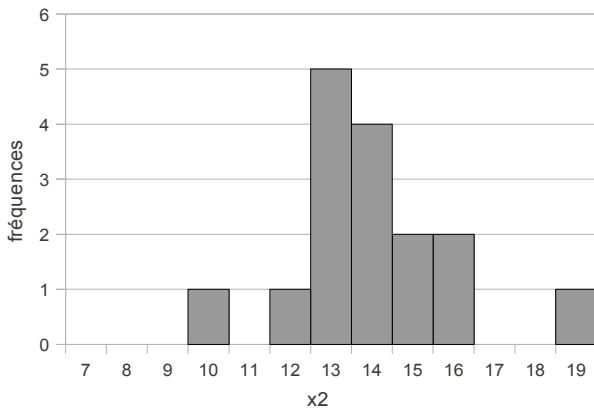
écart-type= 2,07

étendue= 7

écart quadratique moyen= 2,00

Cas 2 :

	X_2
X_2^1	15
X_2^2	13
X_2^3	12
X_2^4	13
X_2^5	14
X_2^6	13
X_2^7	16
X_2^8	19
X_2^9	13
X_2^{10}	14
X_2^{11}	10
X_2^{12}	16
X_2^{13}	14
X_2^{14}	15
X_2^{15}	13
X_2^{16}	14

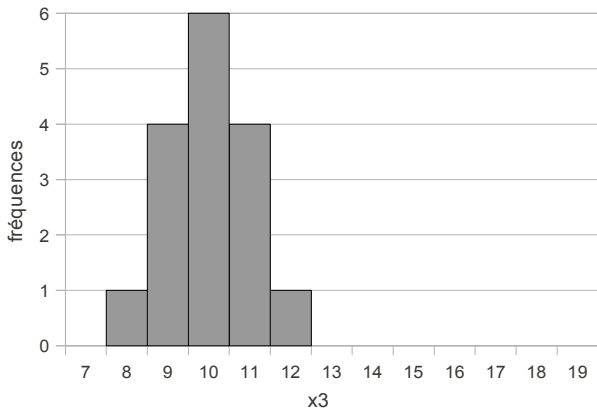


moyenne = 14
mode = 13
médiane = 14

écart-type = 2,00
étendue = 9
écart quadratique moyen = 1,94

Cas 3 :

	X_3
X_3^1	10
X_3^2	10
X_3^3	12
X_3^4	11
X_3^5	9
X_3^6	8
X_3^7	10
X_3^8	9
X_3^9	9
X_3^{10}	11
X_3^{11}	9
X_3^{12}	11
X_3^{13}	10
X_3^{14}	10
X_3^{15}	11
X_3^{16}	10



moyenne = 10
mode = 10
médiane = 10

écart-type = 1,03
étendue = 4
écart quadratique moyen = 1,00

La moyenne n'est pas toujours la valeur la plus représentée (*cas 1 et 2*) et elle peut même dans certains cas être absente. Dans le *cas 3* la courbe est symétrique ce qui implique l'égalité de la médiane et de la moyenne.

Sur les trois exemples certaines valeurs sont représentées plusieurs fois, on parle alors de la *fréquence* f_i d'une valeur x_i . Nous avons $n = \sum_{i=1}^c f_i$, où c correspond au nombre de valeurs de x_i différentes auxquelles nous attribuons une fréquence (dans la suite, c sera aussi le nombre de classes).

La moyenne et l'écart-type peuvent alors s'exprimer ainsi :

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

Parfois on rassemble les valeurs par *classe*, par exemple si nous nous intéressons à la taille des habitants d'une ville nous pouvons rassembler tous les habitants qui ont une taille comprise entre 160 cm et 170 cm dans le même ensemble appelé classe. Leur nombre dans cette classe est la fréquence (ou effectif de la classe) et la valeur est prise égale au milieu de la classe, ici 165 cm (démarche illustrée dans l'exercice 5).

Plus la courbe est ramassée sur son centre plus l'écart-type est faible (sur le *cas 3* l'écart est deux fois plus faible que sur les *cas 1 et 2*).

E. Théorème central limite

1) Population et échantillons

Considérons une ville d'un million d'habitants. Pour sonder la population nous pouvons interroger un échantillon de seulement mille personnes tirées au hasard. A partir de cet échantillon de $n=1000$ individus, nous pouvons, grâce à la statistique, avoir des informations sur la population toute entière. Plus la taille de l'échantillon est grand, plus les résultats seront précis. Nous appelons \bar{x} la moyenne de l'échantillon et s son écart-type. Pour la population nous notons μ (lettre grec mu) la moyenne et σ (sigma) l'écart-type. Plus l'échantillon est grand, plus les valeurs de \bar{x} et de s de cet échantillon sont amenées à se rapprocher de celles μ et σ de la population.

Comme dans le cas des sondages d'opinion avec des échantillons de l'ordre de mille personnes, si nous mesurons la taille de mille habitants qui sont choisis au hasard parmi la population d'une ville d'un million d'habitants, la moyenne de la taille sur cet échantillon a de forte chance d'être proche de celle sur l'ensemble de la population, mais n'a aucune raison de lui être égale.

Illustrons maintenant par le jeu du lancer de pièces. A chaque lancer nous obtenons une réalisation de la variable aléatoire *pile* ou *face*. Ici la population est infinie, nous pouvons répéter le lancer indéfiniment et avoir une infinité de mesures. De plus, les probabilités étant

connues, nous pouvons déterminer à l'avance les caractéristiques de la population.

Tout d'abord, quand la taille de l'échantillon devient très grande et donc infinie, l'échantillon s'identifie à la population : $\mu = \lim_{n \rightarrow \infty} \bar{x}$ ³.

Ensuite nous voyons apparaître la notion de probabilité :

$p_i = \lim_{n \rightarrow \infty} \frac{f_i}{n}$ où p_i est la probabilité de réalisation de l'évènement x_i .

D'après la formule page 6 nous avons ainsi l'expression de la moyenne pour la population : $\mu = \sum p_i \cdot x_i$.

Nous avons $\sum p_i = 1$, car nous considérons tous les événements possibles (1=100%).

Nous associons $x_0=0$ à l'observation *face*, et $x_1=1$ pour *pile*. Comme la pièce est équilibrée $p_0=p_1=1/2=0,5=50\%$ et $\mu=p_0 \cdot x_0 + p_1 \cdot x_1$. Nous avons jugé inexistant l'évènement *la pièce reste sur la tranche*.

De même nous avons : $\sigma = \lim_{n \rightarrow \infty} s$ et en prenant la limite de la formule pour s page 6 nous obtenons $\sigma = \sqrt{\sum p_i \cdot (x_i - \mu)^2}$ (avec pour n grand, $n-1$ pris égal à n).

Au final : $\mu=0,5$ et $\sigma=0,5$.

3 MATH : se lit « μ est égal à la limite de \bar{x} quand n tend vers l'infini».

Prélevons un échantillon en lançant neuf pièces :

{0; 1; 1; 0; 1; 1; 1; 0; 0}.

Nous avons alors $\bar{x} \simeq 0,56$ et $s \simeq 0,53$.

Supposons que nous tirions de nombreuses fois neuf pièces aléatoirement au sein de la même population. A chaque fois nous aurions un résultat différent pour \bar{x} .

Par exemple pour deux autres échantillons prélevés :

{1; 1; 0; 1; 1; 0; 1; 1; 0} avec $\bar{x} \simeq 0,67$ et $s \simeq 0,50$

et {0; 1; 0; 0; 0; 0; 1; 0; 0} avec $\bar{x} \simeq 0,22$ et $s \simeq 0,44$

Quelle serait alors la répartition de l'ensemble de ces résultats ? (appelée la distribution d'échantillonnage)

Les valeurs obtenues pour les échantillons sont en général différentes de celles de la population, mais plus l'échantillon est grand plus il est probable que les valeurs soient proches de celle de la population. Cas d'un échantillon de taille $n=50$ où $\bar{x} \simeq 0,520$ et $s \simeq 0,505$:

{00100011111110010111110001001101101000110000011101}

2) Le théorème central limite

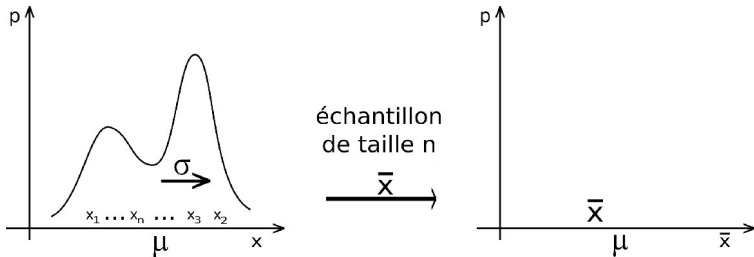
THÉORÈME CENTRAL LIMITE :

Nous prélevons au sein d'une population des échantillons aléatoires de taille n , la moyenne de l'échantillon \bar{x} varie autour de la moyenne de la population μ avec un écart-type égal à σ/\sqrt{n} , où σ est l'écart-type de la population.

Quand n croît la distribution d'échantillonnage de \bar{x} est de plus en plus concentrée autour de μ et devient de plus en plus proche d'une distribution de Gauss.

Nous décrivons prochainement ce qu'est une distribution de Gauss, dans l'immédiat nous considérerons simplement une courbe en cloche. C'est un théorème très important, **quelque soit** la forme de la distribution de la population, la distribution d'échantillonnage est **gaussienne**, et sa dispersion est donnée par le théorème central limite.

Illustrons sur des schémas :

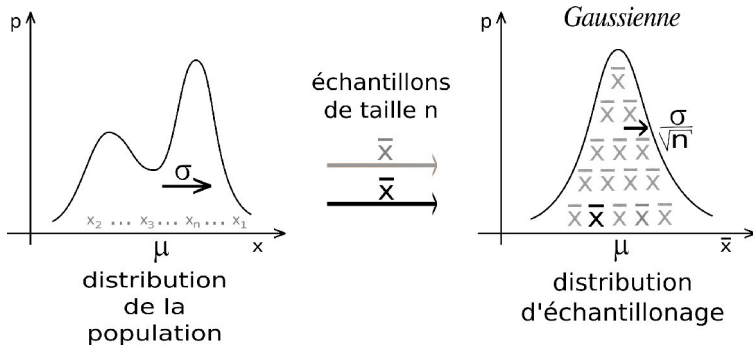


A gauche nous avons la probabilité p d'un événement x (distribution de la population). Par exemple, pour une po-

pulation d'un million d'habitants, pourrait être ici représenté la probabilité p qu'ils aient une taille donnée x . Si nous pouvions mesurer la taille de tous les habitants, nous pourrions déterminer exactement leur taille moyenne μ et son écart-type σ . C'est d'un point de vue pratique difficile, ou même impossible, alors nous décidons de mesurer la taille de mille habitants seulement. Pour que ce soit caractéristique nous tirons ces mille personnes au hasard. Nous obtenons mille mesures de tailles de x_1 à x_{1000} . De cet échantillon de taille $n=1000$ nous calculons une moyenne \bar{x} et un écart-type s . Nous pensons que \bar{x} est proche de μ , mais à la fois il n'y a aucune raison qu'il soit égale à μ . Nous plaçons cette valeur de \bar{x} sur la partie droite de la figure page 10.

Nous prenons un nouvel échantillon aléatoire de mille personnes et nous plaçons un nouveau \bar{x} .

Nous répétons ensuite cette opération un grand nombre de fois. Nous voyons à droite la distribution des échantillons obtenue :



3) Coefficient de Student et incertitude

Le théorème central limite s'applique dans la limite des grands nombres, mais on sait l'adapter pour n petit grâce aux *coefficients de Student*. En effet on ne peut pas appliquer directement ce théorème car ne connaissant pas σ nous l'estimons avec s . Du fait d'une statistique faible il y a alors un élargissement connu :

$$\mu = \bar{x} \pm t \cdot \frac{s}{\sqrt{n}}$$

Le coefficient de Student t dépend de n et inclut la *confiance* que nous voulons donner au résultat. Si la confiance est de 95%, nous avons 95 chances sur 100 que μ soit compris entre $\bar{x} - t \cdot s / \sqrt{n}$ et $\bar{x} + t \cdot s / \sqrt{n}$. Les valeurs de t sont lues dans la table page 172.

Nous reconnaissons ici la notion d'*incertitude* Δx ⁴:

$$x = \bar{x} \pm \Delta x \quad \text{avec} \quad \Delta x = t \cdot \frac{s}{\sqrt{n}}$$

Δx est aussi appelée l'*incertitude absolue* et $\Delta x / \bar{x}$ l'*incertitude relative*.

4 MATH : se lit "delta x".

Reprenons l'expérience de calorimétrie décrite page 1, supposons que nous voulions maintenant connaître la capacité thermique de l'eau avec une confiance de 95%. Nous trouvons dans la table pour quatre *degrés de liberté* ($ddl=n-1$) $t=2,78$.

D'où : $c = \bar{c} \pm t \cdot s_c / \sqrt{n} = 4320 \pm 660 \text{ J/K/kg}$ à 95%.

Ici suite à la dispersion des valeurs mesurées par les étudiants $\Delta c / \bar{c} \approx 15\%$. Les mesures en calorimétrie ne sont pas très précises. La valeur attendue, ici connue, est bien dans l'intervalle : $3660 < 4180 < 4980$.

Nous avons aussi l'expérience avec les pièces décrite sur les pages 7 à 9. Pour le premier échantillon de neuf pièces nous aurions avec une confiance de 90% : $\mu_1 = 0,56 \pm 1,86 \cdot 0,53 / \sqrt{9}$ d'où $\mu_1 = 0,56 \pm 0,33$. Pour le 2^{ème} : $\mu_2 = 0,67 \pm 0,31$. Le 3^{ème} : $\mu_3 = 0,222 \pm 0,273$. Nous devons avoir $\mu = 0,5$ dans l'intervalle, ce qui est facilement vérifié pour μ_1 et μ_2 . Pour μ_3 on y est presque, mais de toute façon un cas sur dix peut être en dehors (100% moins 90%). Avec une confiance de 95% μ_3 engloberait la valeur attendue.

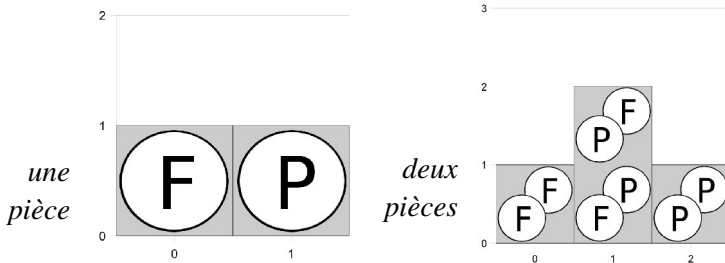
En sciences expérimentales nous nous efforçons de quantifier l'ensemble des phénomènes naturels. Mais, par la nature même de la démarche expérimentale, les différents paramètres qui permettent de décrire une situation ne sont pas parfaitement connus. Nous n'avons pas simplement une valeur numérique associée à chaque caracté-

ristique, mais un intervalle pour une confiance donnée. En toute rigueur, à toute grandeur expérimentale doit être associée son incertitude avec sa confiance.

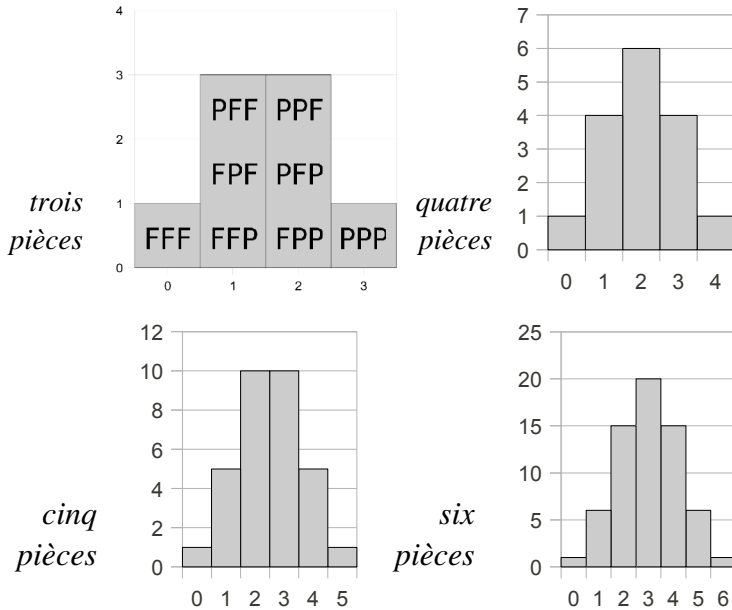
4) Exemples

Concrètement un grand nombre de facteurs aléatoires vont influencer sur la mesure d'une grandeur, facteurs indépendants, qui, quelque soient leurs natures vont au final générer une distribution gaussienne des valeurs. Prenons deux exemples, le lancer de pièces et celui de dés à six faces.

Pour les pièces nous comptons le nombre de piles à chaque lancer de plusieurs pièces. Nous dénombrons le nombre de possibilités pour un nombre de piles donné. Pour une pièce, une possibilité pour zéro pile (face, F) et une possibilité pour un pile (P). Pour deux pièces, une possibilité pour zéro pile (F F), deux possibilités pour un pile (F P ou P F) et une possibilité pour deux piles (P P). Plus le nombre de pièces lancées simultanément est grand, plus nous tendons vers une distribution gaussienne.



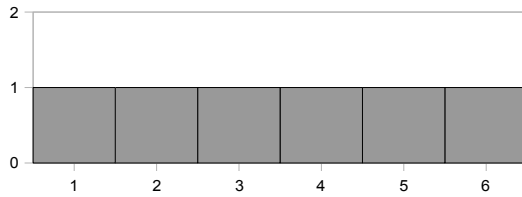
Nous pouvons obtenir en ordonnée la probabilité, en divisant par le nombre de possibilités 2^n , et en abscisse la moyenne pour chaque pièce, en divisant par le nombre de pièces n . Pour $n=1$ nous avons alors la distribution de la population et ensuite les distributions d'échantillonnage pour différentes valeurs de n .



De même pour les dés nous énumérons les possibilités pour leur somme et nous tendons aussi vers une gaussienne.

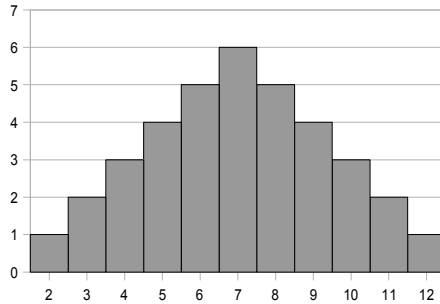
Pour un seul dé, la somme correspond tout simplement à la valeur du dé. Nous avons une possibilité pour chaque valeur :

un dé :

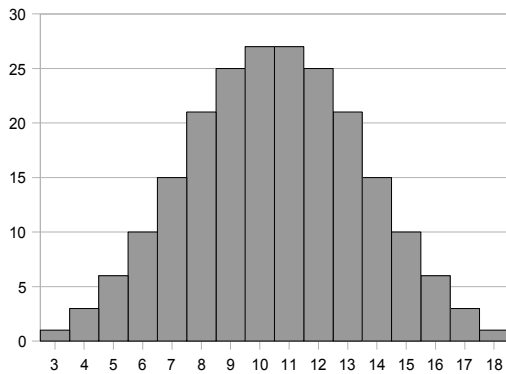


Pour deux dés, il y a une seule possibilité pour que la somme fasse deux : 1 pour le premier dé et 1 pour le deuxième dé. Pour que la somme fasse trois, il y a deux possibilités : 1 puis 2, ou, 2 puis 1. Le plus probable avec deux dés est d'obtenir 7 : (1,6) (6,1) (2,5) (5,2) (3,4) (4,3).

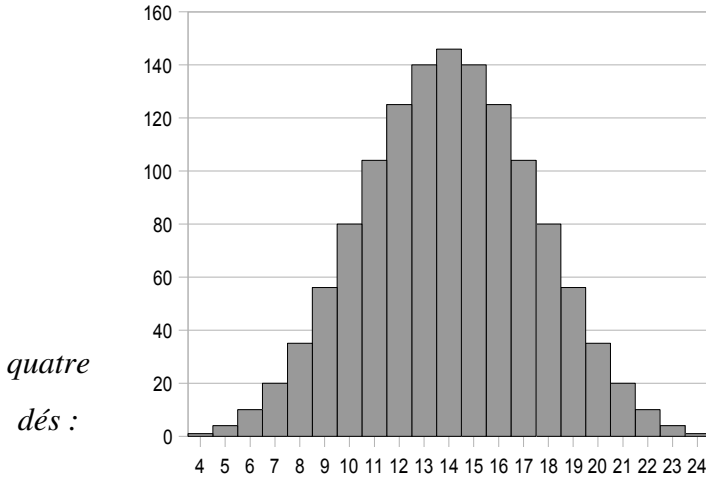
deux dés :



trois dés :



Pour quatre dés nous reconnaissons déjà bien la courbe en cloche et le profil est clairement de type gaussien :



Sur ce dernier exemple vérifions la validité du théorème central limite.

La moyenne pour la population est :

$$\mu_x = (1 + 2 + 3 + 4 + 5 + 6) / 6 = 3,5$$

La moyenne de la distribution d'échantillonnage est bien la même : $\mu_{\bar{x}} = 14 / 6 = 3,5$.

L'écart-type de la population :

$$\sigma_x = \sqrt{\sum p_i \cdot (x_i - \mu)^2} \text{ d'où}$$

$$\sigma_x = \sqrt{1/6 \cdot [(1-3,5)^2 + (2-3,5)^2 + (3-3,5)^2 + (4-3,5)^2 + (5-3,5)^2 + (6-3,5)^2]}$$

et $\sigma_x \approx 1,71$

d'où pour quatre dés $\sigma_x = \sigma_x / \sqrt{n} = \sigma_x / 2 \approx 0,85$. Or, sur la courbe du dessus, à 40% du maximum (explication page 20), nous avons un écart d'environ 3,5 (entre 3 et 4), soit en moyenne $3,5/4 \approx 0,88$. Il y a bien correspondance.

F. Distribution de Gauss

1) Définition d'une distribution continue

Il existe des grandeurs fondamentalement continues. Le temps en est une. Le temps s'écoule continuellement tel un fluide, il ne saute pas d'une valeur à l'autre. S'il semble parfois égrainer les secondes, comme les perles d'un chapelet, cela n'est dû qu'à l'appareil de mesure insuffisamment précis. Par contre quand nous lançons un dé six faces, impossible de dire "j'ai fait 2,35 !", c'est une valeur interdite, seules les valeurs entières de un à six sont permises.

Ainsi, certaines distributions de probabilité sont discrètes et d'autres continues. Pour le lancé de dés la probabilité est discrète : $p_1 = \dots = p_6 = 1/6$ et $\sum_{i=1}^{n=6} p_i = 1$. Le résultat du lancer n'a que six valeurs possibles. Par contre si nous nous intéressons à la taille des habitants d'une ville, celles-ci sont toutes possibles, la distribution est continue. Nous parlons alors de densité de probabilité $p(x)$ avec $p(x)dx$ la probabilité d'un évènement d'être entre x et $x+dx$. dx est une petite variation, et $x+dx$ est infini-

ment voisin de x . La probabilité que l'évènement se réalise sur l'ensemble des valeurs possibles est de 100% :

$$\int_{x=-\infty}^{+\infty} p(x) dx = 1 \quad ^5$$

Calcul de la moyenne et de l'écart-type d'une distribution continue :

$$\mu = \int_{-\infty}^{+\infty} x \cdot p(x) dx \qquad \sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 p(x) dx$$

2) Courbe de Gauss

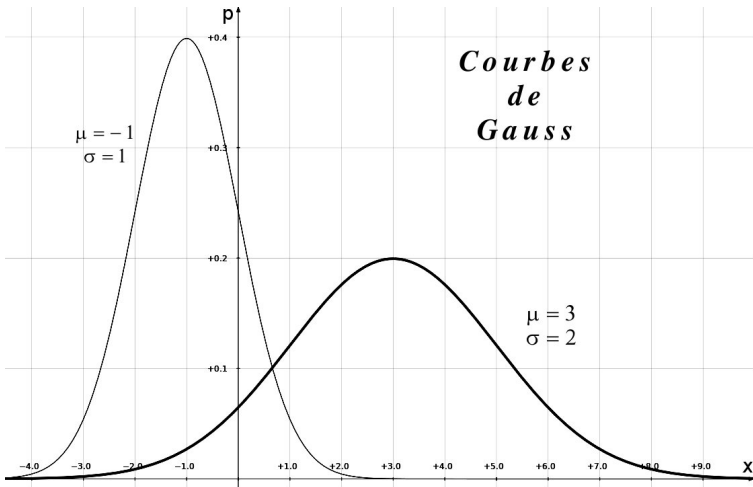
Le théorème central limite s'intéresse au cas où la taille de l'échantillon n est grande, et dans le cas limite où n tend vers l'infini nous considérons une distribution continue. Celle-ci est une gaussienne, l'expression mathématique est connue :

$$p(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2}$$

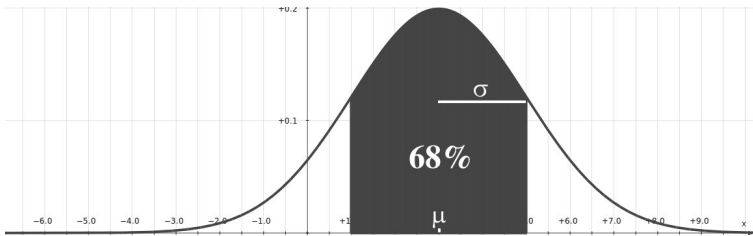
Dans le complément mathématique page 166 différents termes sont justifiés.

⁵ MATH : se lit «l'intégrale de p de x , x allant de moins l'infini à $+\infty$, est égale à 1 ».

Nous avons représenté, sur le graphe suivant, deux cas :



L'aire totale sous la courbe vaut toujours 1. Mais, si maintenant nous ne nous écartons que d'un écart-type par rapport à la moyenne, nous n'avons que 68% des possibilités :

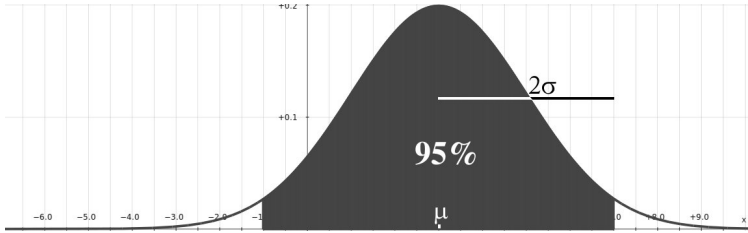


$$\int_{\mu-\sigma}^{\mu+\sigma} p(x) dx = 0,683 \dots \approx 68\%$$

L'écart-type peut s'évaluer ici, en effet à 60%. p_{max} :

$$p(\mu \pm \sigma) / p_{max} = 1/\sqrt{e} \approx 0,607$$

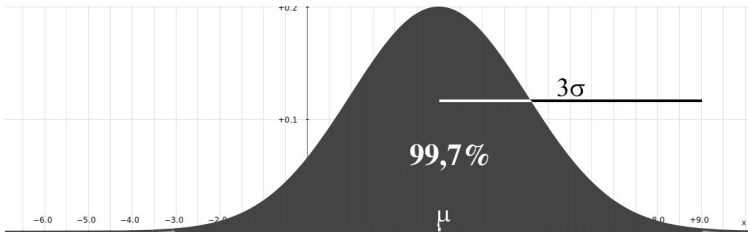
Élargissons à deux écart-types :



$$\int_{\mu-2\sigma}^{\mu+2\sigma} p(x) dx = 0,954 \dots \simeq 95 \%$$

On dit 95% de confiance, ou, une confiance à deux sigmas. On travaille souvent avec cette confiance.

Puis pour trois sigmas :



$$\int_{\mu-3\sigma}^{\mu+3\sigma} p(x) dx = 0,997 \dots > 99 \%$$

3) Loi normale standard

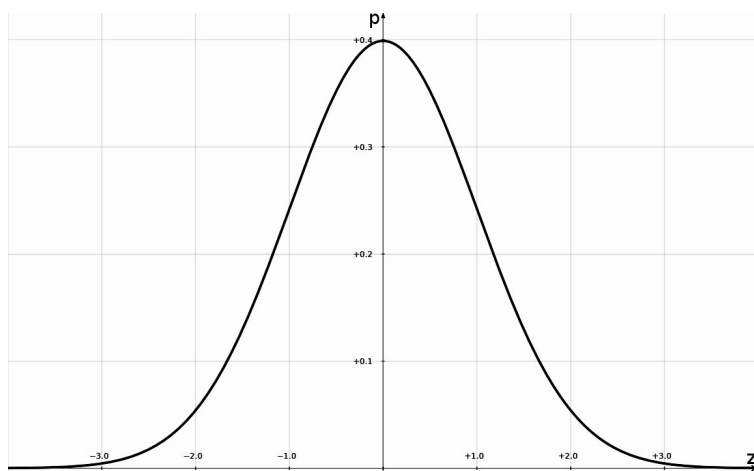
C'est la distribution normale de Gauss centrée et

réduite. Pour le recentrage nous soustrayons la moyenne :
 $x' = x - \mu$. Pour la réduction nous divisons par l'écart-type :

$$z = \frac{x - \mu}{\sigma}$$

d'où : $p(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

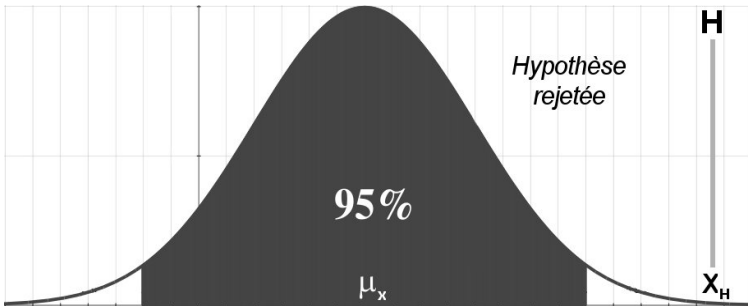
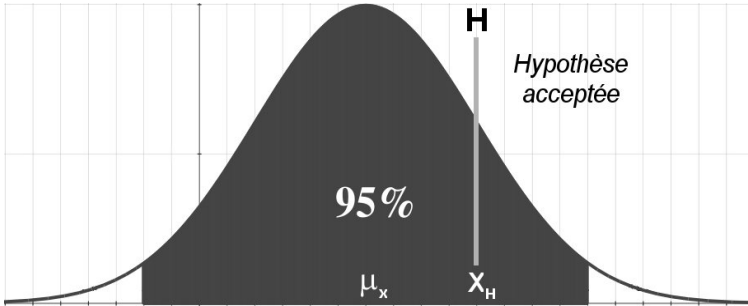
Nous avons alors une distribution normale de moyenne nulle et d'écart-type égale à un :



G. Test d'hypothèse

Nous voulons vérifier une hypothèse H . La valeur correspondant à la grandeur X est supposée égale à x_H . Si

après estimation de μ_x et σ_x , x_H est compris entre $\mu_x - 2\sigma_x$ et $\mu_x + 2\sigma_x$ alors l'hypothèse H est acceptée avec une confiance de 95%, sinon l'hypothèse est rejetée.



Si $x_H = \mu_x \pm 2\sigma_x$ nous avons 95% de chances d'avoir raison de soutenir l'hypothèse H et 5% de chance de nous tromper.

Pour travailler avec d'autres confiances il suffit de considérer le coefficient de Student t_∞ .

Dans l'exercice 4, page 32, nous testons un isolant thermique et nous vérifions les valeurs annoncées par le fabri-

cant.

L'hypothèse peut aussi être dissymétrique et nous ne considérons alors qu'une des deux parties de la queue de distribution : par exemple, la charge maximale d'un ascenseur est de 300 kg, la masse totale des occupants est de 280 ± 10 kg à σ . Quelle est la probabilité d'être en surcharge ? 25 chances sur mille.

H. Test du Khi-deux

Il s'agit d'un test d'hypothèse simple basé sur les différences entre les effectifs observés et ceux espérés :

Effectifs observés :

O_1	O_2	...	O_j	...	O_c
-------	-------	-----	-------	-----	-------

Effectifs espérés :

E_1	E_2	...	E_j	...	E_c
-------	-------	-----	-------	-----	-------

Nous calculons la sommes des écarts au carré par rapport à la valeur théorique attendue :

$$\chi^2 = \sum_{j=1}^c \frac{(O_j - E_j)^2}{E_j} \quad (\text{nommé "khi-deux"})$$

Nous avons ensuite une table (page 173) qui permet d'estimer la **probabilité que l'hypothèse soit juste**. Suivant la valeur de χ^2 et le nombre de degrés de liberté nous déterminons la justesse de l'hypothèse.

Les degrés de liberté se calculent ainsi :

$$ddl = c - 1 \quad (\text{nombre de valeurs moins une unité})$$

Illustrons avec les expériences réalisées par le botaniste Mendel. Il effectue des croisements entre des plants, ici des hybrides. Il croise entre eux des pois à fleurs roses. Sa théorie lui indique qu'il doit obtenir 25% de pois à fleurs rouges, 25% de pois à fleurs blanches et 50% de pois à fleurs roses. Ceci résulte de la rencontre aléatoire des gamètes. Imaginons qu'il observe sur mille fleurs les valeurs suivantes : 27% de blanches, 24% de rouges et 49% de roses. Doit-il continuer à croire en son hypothèse ?

Effectifs observés :

270	240	490
-----	-----	-----

Effectifs espérés :

250	250	500
-----	-----	-----

d'où

$$\chi^2 = \frac{(270-250)^2}{250} + \frac{(240-250)^2}{250} + \frac{(490-500)^2}{500} \simeq 2,2 \quad \text{et}$$
$$ddl = 3 - 1 = 2 \quad .$$

Donc, d'après la table, il y a plus de 30% de chance que l'hypothèse soit juste. Il n'y a alors aucune raison de la rejeter. En général on prend une probabilité critique de 5%, en dessous de laquelle on envisage de rejeter l'hypothèse.

Le test se généralise facilement pour un tableau.

Effectifs observés :

Effectifs espérés :

$$\begin{pmatrix} O_{11} & O_{12} & \dots & O_{1j} & \dots & O_{1c} \\ O_{21} & O_{22} & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ O_{i1} & \dots & \dots & O_{ij} & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ O_{l1} & \dots & \dots & \dots & \dots & O_{lc} \end{pmatrix} \quad \begin{pmatrix} E_{11} & E_{12} & \dots & E_{1j} & \dots & E_{1c} \\ E_{21} & E_{22} & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ E_{i1} & \dots & \dots & E_{ij} & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ E_{l1} & \dots & \dots & \dots & \dots & E_{lc} \end{pmatrix}$$

Les ddl tiennent compte du nombre de colonnes c et de lignes l : $ddl = (c - 1)(l - 1)$

Le χ^2 se calcul avec le même type de formule :

$$\chi^2 = \sum_{(i,j)} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

De plus nous utilisons la même table pour déterminer la validité de l'hypothèse.

I. Sources des incertitudes

Une variable aléatoire a une incertitude d'autant plus faible que la mesure est fidèle, juste et que le système d'acquisition a une bonne résolution.

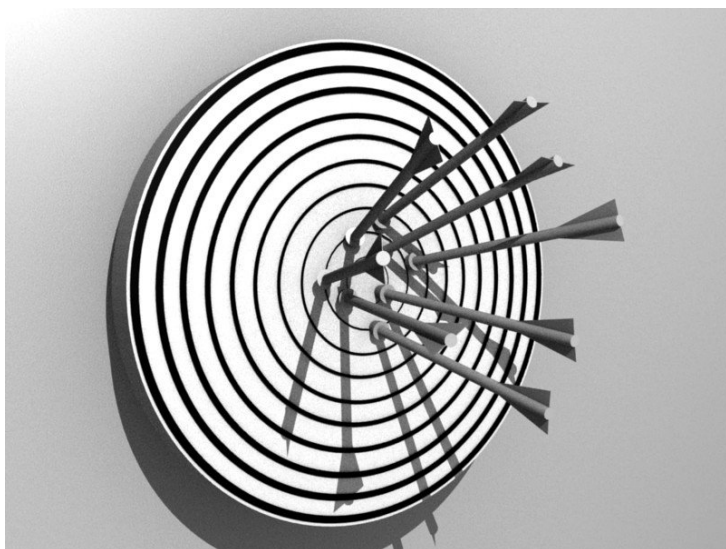
La justesse est assurée par l'absence d'erreurs systématiques. Il peut exister un biais qui rend la mesure inexacte (même si la dispersion est faible). Erreurs de lecture, absence de contrôle et de corrections de facteurs influents, incertitude due à la modélisation, etc. Tous les biais doivent être identifiés et estimés afin d'être ajoutés à la dispersion, le système devient alors juste.

La fidélité provient de la répétabilité et de la reproductibilité des mesures. Les valeurs d'un système fidèle sont peu dispersées. La dispersion peut provenir d'erreurs accidentelles ou d'un phénomène physique par essence aléatoire (comme par exemple la radioactivité). Les expérimentateurs par un travail propre, consciencieux et selon un protocole bien défini et réfléchi, pourront minimiser la dispersion. Les sources peuvent être innombrables, nous essaierons d'en identifier un maximum afin de les évaluer.

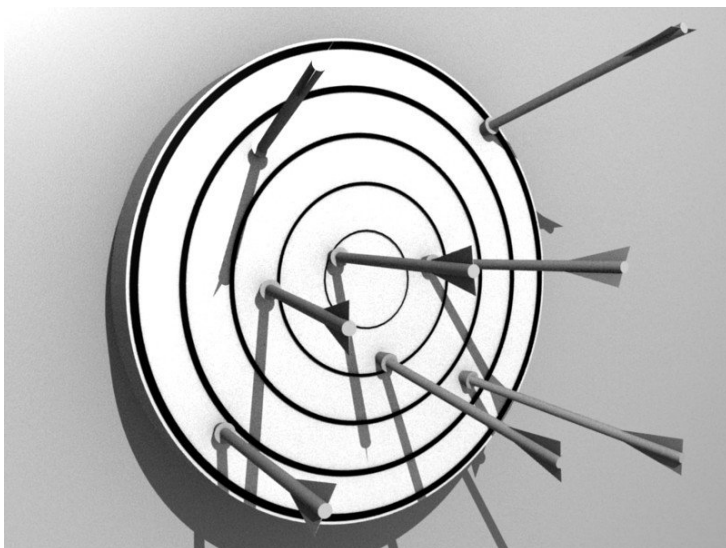
La résolution de l'instrument de mesure dépend de la taille des graduations, du type du vernier ou du nombre de digits de l'affichage. A l'incertitude due à la discrétisation des mesures peuvent s'ajouter d'autres facteurs. Il faudra se référer à la notice ou contacter le fabricant pour connaître au mieux la précision de votre appareil. On peut aussi effectuer un étalonnage avec un instrument de haute précision qui sert de référence.

L'influence de ces différentes sources d'incertitude peut être illustrée par une cible et des flèches. Le centre de la cible correspond à la grandeur à mesurer et les flèches représentent les différentes mesures. Si les flèches, dans leur ensemble, ne sont pas correctement centrées, la justesse n'est pas assurée. Le resserrement des flèches représente la fidélité. La distance entre les cercles sur la cible indique la résolution. La valeur notée est celle du cercle dont la flèche est le plus proche. L'expérimentateur voit les flèches et les cercles, par contre il ne sait pas où est la cible et son centre. Il tient l'arc et son désir d'être au plus proche du centre de la cible montre la qualité et la rigueur de son travail.

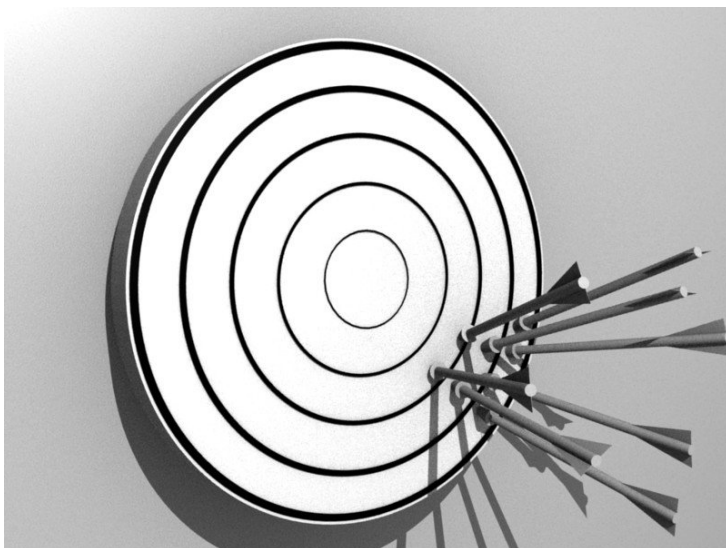
Mesure juste, fidèle et avec une bonne résolution :



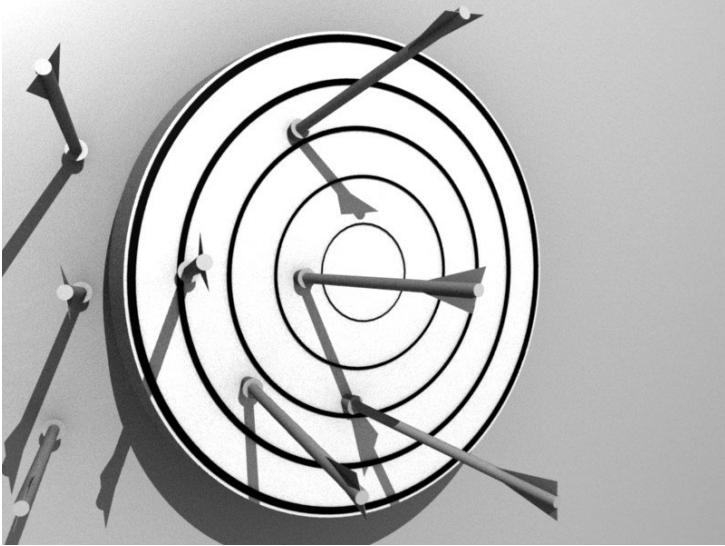
Mesure juste, mais peu fidèle et avec une faible résolution :



Mesure peu dispersée mais avec un biais, et mal résolue :



*Mesure avec un biais, fortement dispersée
et une faible résolution :*



L'écart-type complet sera déterminé à partir des écarts de chaque source en ajoutant les carrés (dus aux compensations explicitées au chapitre 2) :

$$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \dots}$$

J. Exercices

Exercice 1 : Âges

Soient les âges des étudiants d'une classe {18; 20; 18; 19; 18; 18; 18; 17; 18; 19; 17; 19; 17; 21; 18}. Déterminez le mode, la médiane, la moyenne arithmétique, la moyenne géométrique, l'étendue, l'écart-type, l'écart quadratique moyen et l'écart moyen⁶.

Exercice 2 : Cartes

Soit un jeu de 32 cartes. Nous tirons au hasard cinq cartes. Déterminez la probabilité d'avoir un carré d'as, puis celle d'avoir une couleur.

Exercice 3 : Champ de pesanteur

Des étudiants mesurent l'intensité g du champ de pesanteur terrestre au laboratoire. Les huit binômes mesurent les valeurs suivantes en m/s^2 : 6,20 ; 8,35 ; 13,00 ; 8,37 ; 8,54 ; 9,67 ; 9,75 ; 10,66.

- Quel commentaire général feriez-vous en voyant ces résultats.
- Calculez la moyenne et l'écart-type.
- La dispersion des valeurs donne quelle incertitude sur la moyenne (confiance de 95%) ? Le résultat est-il cohérent avec la valeur attendue ?
- Un neuvième binôme fait une nouvelle mesure dans les mêmes conditions expérimentales. Estimez la probabilité qu'ils obtiennent un résultat entre 8 et 12 m/s^2 .

⁶ $\text{Écart moyen} = \left(\sum |x_i - \bar{x}| \right) / n = \left(\sum \sqrt{(x_i - \bar{x})^2} \right) / n$

Exercice 4 : Test d'un isolant

Un fabricant indique pour un panneau isolant en cellulose une conductivité thermique de 0,039 W/m.K. La valeur est certifiée à 5% près. Vous voulez vérifier si c'est vrai, pour cela vous prenez dix panneaux au hasard et vous mesurez leurs conductivités respectives ($\text{mW}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$) :

39,1	38,8	39,5	39,2	38,9	39,1	39,2	41,1	38,6	39,3
------	------	------	------	------	------	------	------	------	------

Les valeurs sont-elles conformes à celles annoncées par le fabricant (il considère une confiance de 95% sur la marge donnée) ? Pourrait-il, selon vos résultats, annoncer une autre marge ?

Exercice 5 : Devoir

Pour un devoir les notes obtenues par des étudiants sont les suivantes (notation sur 20) :

10	5	13	7	6	9	5	5	10	15	5	3
15	12	11	1	3	13	11	10	2	7	2	8
2	15	4	11	11	5	8	12	10	18	6	

- Déterminez la moyenne et l'écart-type.
- Réalisez un diagramme avec les notes en abscisse et les fréquences en ordonnée.
- Faire un autre diagramme des fréquences en regroupant les données en classes : $[0, 1, 2]$, $[3, 4, 5]$... , $[18, 19, 20]$. On indiquera aussi les fréquences relatives. Comparez la lisibilité avec le premier diagramme.

Exercice 6 : Yams

Le Yams se joue avec cinq dés équilibrés à six faces.

1) On lance les cinq dès, quelle est la probabilité de faire un Yams directement (mêmes valeurs sur les cinq dés).

2) Pour un lancer quelle est la probabilité d'avoir un total inférieur à dix ?

3) Nous effectuons une série de lancers et nous obtenons les totaux suivants : 18, 15, 17, 22, 16, 12, 14, 22, 23, 14, 23, 14, 18, 21, 12, 15, 18, 13, 15, 18, 17, 15, 17, 21, 25, 16, 8, 15, 15, 13.

a) Calculez la moyenne et l'écart-type.

b) Quelle est l'estimation de la moyenne avec un intervalle de confiance de 95% ? Est-ce que ça correspond à la valeur théorique ?

c) Réalisez un graphe des valeurs avec leurs fréquences correspondantes.

d) Si nous effectuons un nouveau lancer, quelle est, selon vous, la probabilité d'avoir un résultat supérieur à 24 ?

Exercice 7 : Pièces

Nous effectuons un grand nombre de lancers de pièces, nous cherchons à déterminer si les probabilités d'obtenir un pile ou un face sont égales. Nous effectuons l'expérience avec trois pièces différentes, sont-elles équilibrées ? (réponses avec une confiance de 95%)

1) 42 piles et 58 faces.

2) 510 piles et 490 faces.

3) 420 piles et 580 faces.

Exercice 7 : Parité

La parité hommes-femmes est-elle respectée dans les deux chambres et le conseil constitutionnel ?

	Hommes	Femmes
Assemblée nationale	470	107
Sénat	272	76
Conseil constitutionnel	10	2

Exercice 9 : Naissances

Testons l'hypothèse suivante : les naissances en Suède se répartissent uniformément tout au long de l'année. Supposons que nous avons un échantillon aléatoire de 88 naissances. Les résultats sont regroupés selon des saisons de longueur variable : 27 naissances au printemps (avril-juin), 20 en été (juillet/août), 8 en automne (septembre/octobre) et 33 en hiver (novembre-mars).

Au seuil de 5% l'hypothèse peut-elle être rejetée ?

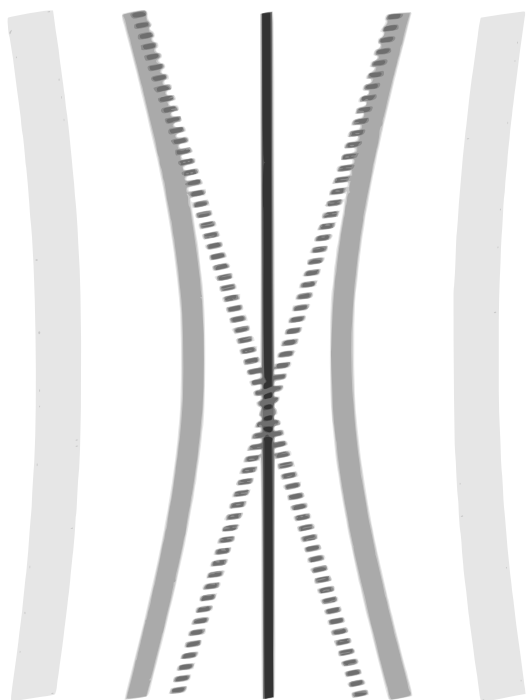
Sur le même exemple on collecte maintenant un très grand échantillon : 23385, 14978, 14106 et 35804.

Quelle est alors la conclusion ?

Pour vérifier vos calculs sur un tableur vous pouvez utiliser les fonctions suivantes :

sur OpenOffice :

- somme d'une zone sélectionnée * (ex.: B43:B53) : =SOMME(*)
- valeur fixée à la cellule B3 : \$B\$3
- =MOYENNE(*)
- carré : $*^2$
- racine carrée : $*^{(1/2)}$
- =ECARTYPE(*)
- coefficient de Student pour une confiance de 95% et n=20 : =LOI.STUDENT.INVERSE(0,05;19)
- probabilité que l'hypothèse soit vraie, tableau des valeurs observées * (ex.: B71:E72), celui des valeurs estimées ** : =TEST.KHIDEUX(*;**).



II. CORRÉLATIONS ET INDÉPENDANCES

Au chapitre précédent nous n'avions qu'une grandeur aléatoire X avec ses n réalisations $\{x_i\}$. Maintenant nous avons plusieurs grandeurs et un nouvel indice permet de les distinguer : X_j et ses mesures $\{x_{jk}\}$. X_j est la $j^{\text{ème}}$ grandeur et x_{jk} est la $k^{\text{ème}}$ observation de cette grandeur. Nous allons nous intéresser aux interactions entre différentes grandeurs.

Pour illustrer, considérons un échantillon de quatre individus qui possèdent trois caractéristiques, la taille X_1 , le poids X_2 et le mois de naissance X_3 . A priori, nous nous attendons à une corrélation entre la taille et le poids : plus on est grand plus on a, en général, une masse importante (corrélation positive). Par contre, nous pouvons penser que le mois de naissance n'a aucune incidence sur le poids et la taille (X_3 non corrélée avec X_1 et X_2).

A. Coefficient de corrélation

Le coefficient de corrélation r permet d'identifier s'il y a une relation linéaire entre deux variables X_i et X_j :

$$r_{ij} = \frac{\sum_k [(x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)]}{\sqrt{\sum_k [(x_{ik} - \bar{x}_i)^2]} \cdot \sqrt{\sum_k [(x_{jk} - \bar{x}_j)^2]}}$$

r varie entre -1 et $+1$. Si $|r|=1$ les variables sont parfaitement corrélées : $r=1$ même sens de variation, $r=-1$ sens opposé. Si $r=0$ il n'y a pas la moindre corrélation, les variables sont parfaitement indépendantes.

Calcul de r_{12} , r_{13} et r_{23} :

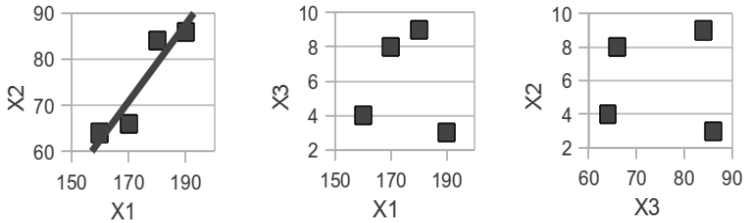
	X_1 (cm)	X_2 (kg)	X_3	$x_1 - \bar{x}_1$	$x_2 - \bar{x}_2$	$x_3 - \bar{x}_3$	$(x_1 - \bar{x}_1)^2$
1	160	64	4	-15	-11	-2	225
2	170	66	8	-5	-9	2	25
3	180	84	9	5	9	3	25
4	190	86	3	15	11	-3	225
\bar{x}	175	75	6			$\Sigma =$	500

et la suite des calculs :

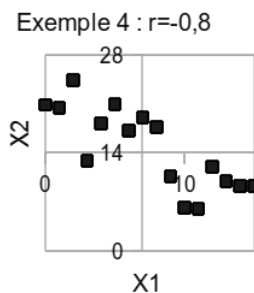
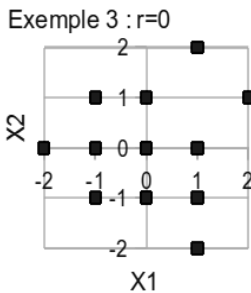
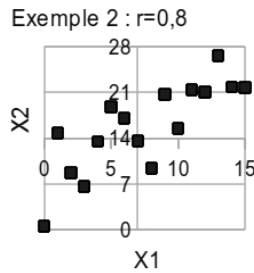
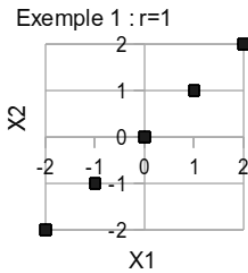
$(x_2 - \bar{x}_2)^2$	$(x_3 - \bar{x}_3)^2$	$(x_1 - \bar{x}_1)$ $\cdot (x_2 - \bar{x}_2)$	$(x_1 - \bar{x}_1)$ $\cdot (x_3 - \bar{x}_3)$	$(x_2 - \bar{x}_2)$ $\cdot (x_3 - \bar{x}_3)$
121	4	165	30	22
81	4	45	-10	-18
81	9	45	15	27
121	9	165	-45	-33
404	26	420	-10	-2

d'où : $r_{12} = \frac{420}{\sqrt{500}\sqrt{404}} \simeq 0,93$, $r_{13} \simeq -0,09$ et $r_{23} \simeq -0,02$.

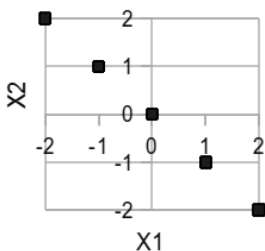
r_{12} est proche de +1, nous avons donc une corrélation positive importante. r_{13} et r_{23} sont proches de zéro : X_3 est indépendante de X_1 et X_2 :



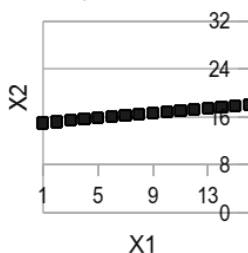
Exemples de nuages de points entre deux variables :



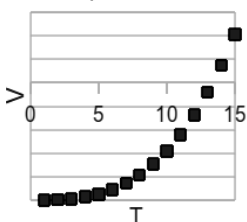
Exemple 5 : $r=-1$



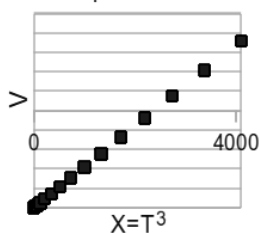
Exemple 6 : $r=1$



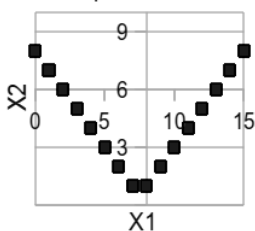
Exemple 7 : $r=0,92$



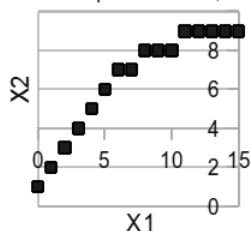
Exemple 8 : $r=1$



Exemple 9 : $r=0$



Exemple 10 : $r=0,94$



Nous voyons sur les exemples 7, 9 et 10 une forte corrélation entre les deux variables. Pourtant le coefficient de corrélation n'est pas aussi proche de -1 ou $+1$ que nous pourrions l'imaginer, il est même nul sur l'exemple 9. Ceci provient du fait que les corrélations ne sont pas linéaires.

Il peut y avoir des phénomènes de saturation (exemple 10) ou de seuil (un produit peut être bénéfique à faible dose et nocif pour des doses plus importantes - exemple 9). Il est préférable d'avoir identifié ces phénomènes en amont. Pour cela il est important de bien réfléchir à la pertinence des variables que l'on choisit avant de commencer une étude statistique.

Autre exemple : si nous étudions le volume V de différents objets en fonction de leur taille T , nous trouvons une corrélation positive. Mais celle-ci sera bien plus forte entre V et $X=T^3$ (graphes 7 et 8).

B. Formule de propagation des incertitudes

Soit un seau rempli d'un million de grains de sable. La masse d'un grain est de 10 mg à 1mg près. Qu'elle est la masse de sable contenue dans le seau ?

1) Formule de propagation des écart-types

Pour une approche générale du problème, soit une fonction f qui dépend de p variables indépendantes :

$$f(x_1, x_2, \dots, x_j, \dots, x_p)$$

A chacune de ces variables aléatoires est associée une valeur moyenne \bar{x}_j et un écart-type σ_j . Que valent f et σ_f ?

La statistique donne la réponse et démontre la formule de propagation des écart-types :

$$\sigma_f^2 = \sum_{j=1}^p \left[\left(\frac{\partial f}{\partial x_j} \right)^2 \sigma_j^2 \right]$$

2) Calcul d'incertitude

Pour les incertitudes (définies page 12) nous avons aussi une formule de propagation :

$$\Delta f^2 = \sum_{j=1}^p \left[\left(\frac{\partial f}{\partial x_j} \right)^2 \Delta x_j^2 \right]$$

La formule de propagation des incertitudes n'est pas exacte comme pour les écart-types, mais elle est très pratique et le plus souvent très proche du résultat exact.

Pour notre seau : $M(m_1, m_2, \dots, m_j, \dots, m_p)$

avec
$$M = \sum_{j=1}^p m_j$$

où nous appelons M la masse totale de sable dans le seau, m_j la masse de chaque grain et p le nombre de grains.

$$\Delta M^2 = \sum_{j=1}^p \left(\partial M / \partial m_j \right)^2 \Delta m_j^2$$

$$\partial M / \partial m_j = \partial m_1 / \partial m_j + \dots + \partial m_j / \partial m_j + \dots + \partial m_p / \partial m_j$$

$$\partial M / \partial m_j = 0 + \dots + 1 + \dots + 0 = 1$$

(les calculs de dérivées partielles sont expliqués page 165)

alors
$$\Delta M^2 = \left(\sum_{j=1}^p 1^2 \right) \Delta m^2$$

d'où $\Delta M^2 = p \cdot \Delta m^2$ avec $\Delta m = \Delta m_j$ quelque soit j .

Finalement : $\Delta M = \sqrt{p} \cdot \Delta m = \sqrt{1000000} \cdot 0,001 \text{ g}$.

Le seau pèse donc dix kilos à un gramme près. La précision sur la masse du seau est donc de 0,01%. Naïvement, nous aurions pu penser que l'incertitude globale sur la masse du seau était la somme des incertitudes de chaque grain, nous aurions alors une incertitude absolue d'un kilo et une relative de 10%, ce qui est très différent de la réalité et ignorerait les compensations.

Ici la formule de propagation est très précise car nous avons un très grand nombre de grains. Elle est même exacte, dès les petits nombres, si la distribution de la masse des grains est gaussienne⁷.

⁷ MATH : une combinaison linéaire de grandeurs gaussiennes est elle-même gaussienne (s'applique ici à une somme). Et dans la formule de propagation des incertitudes, si f et les x_i ont des lois de probabilités de même nature, la formule est tout aussi exacte que la formule de propagation des écart-types.

En pratique, certains cas se rencontrent souvent :

- Pour des sommes ou des différences les incertitudes absolues au carré s'ajoutent :

$$\Delta f^2 = \sum_{j=1}^p \Delta x_j^2$$

Par exemple si $d = x_2 - x_1$ avec $\Delta x_2 = \Delta x_1 = 1 \text{ cm}$ alors
 $\Delta d \simeq 1,4 \text{ cm}$.

- pour des produits ou des quotients les incertitudes relatives au carré s'ajoutent :

$$\left(\frac{\Delta f}{f} \right)^2 = \sum_{j=1}^p \left(\frac{\Delta x_j}{x_j} \right)^2$$

Par exemple si $R = U/I$ avec U et I à 1% alors R est connu à 1,4%.

Dans les cas plus complexes, il faut réaliser explicitement le calcul aux dérivées partielles.

A l'aide d'un ordinateur il est aussi possible de faire un calcul numérique. On peut générer des nombres aléatoires ou utiliser des paquets décorrélés globalement gaussiens. Cette dernière méthode est illustrée dans l'exercice 2 de ce

chapitre. Un tableur peut permettre de faire automatiquement les calculs (par exemple feuille 4 du tableur IncertitudesLibres sur www.incertitudes.fr).

On peut aussi donner des méthodes qui donnent des idées générales sur l'incertitude. Elles développent une certaine intuition mais il y a un risque de se tromper.

Par exemple comme les incertitudes s'ajoutent au carré, on peut considérer que l'incertitude la plus grande va l'emporter rapidement sur les autres. Dans l'exemple où $R=U/I$ si U est connu 1% et I à 0,1% alors R est connu à $1,005\% \approx 1\%$, on peut négliger l'incertitude sur I .

Pour des sommes ou des différences, on considère parfois que le terme qui a le dernier chiffre significatif le moins précis indique la précision du dernier chiffre significatif du résultat. Mais sur notre exemple de la masse du seau rempli de grains de sable, ça ne marche pas : la masse d'un grain est $m = 10$ mg or la masse du seau M est connue au gramme et non au milligramme !

Pour les produits ou des quotients, on considère parfois que le facteur qui a le nombre de chiffres significatifs le plus faible indique le nombre de chiffres significatifs du

résultat, mais là aussi il faut être très attentif.

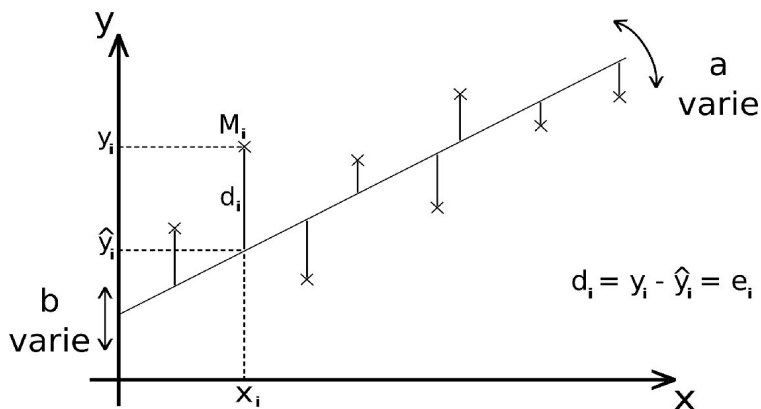
Une illustration très simple, si $H=2h$ avec $h=5,00m$ (h connu au cm), que vaut H ? D'après la règle du dessus H serait connu avec trois chiffres significatifs, donc $H=10,0m$. H ne serait connu qu'à $10cm$ près, il va de soit que l'on est plus vers le cm ...

N'utilisez pas ces recettes si vous n'avez pas conscience des pièges possibles.

C. Régression linéaire

Si maintenant nous avons deux variables corrélées nous pourrions vouloir déterminer la relation la plus adaptée entre elles. Les variables aléatoires seront nommées X et Y et nous chercherons la droite qui passe au mieux par le nuage de points $y(x)$. Par exemple, quelle est la relation la mieux adaptée entre la taille X et le poids Y dans notre exemple initial : $y = ax + b$? Quelles sont les incertitudes Δa et Δb ?

1) Principe et formules



La méthode choisie est celle des moindres carrés : la droite considérée la meilleure est celle qui minimise la somme des carrés des distances à la droite, distances prises selon y (écarts).

L'ensemble des points se note $M_i(x_i, y_i)$. Pour x_i don-

né, l'ordonnée estimée sur la droite s'écrit $\hat{y}_i = ax_i + b$.
 D'où la somme des distances au carré :

$$\sum d^2 = \sum_i (y_i - \hat{y}_i)^2$$

Les dérivées partielles de cette quantité selon a et b s'annulent pour la meilleure droite et nous obtenons les équations suivantes :

$$\sum_i (y_i - ax_i - b)x_i = 0 \quad \text{et} \quad \sum_i (y_i - ax_i - b) = 0 .$$

Nous obtenons ainsi les résultats désirés :

$$a = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2} \quad \text{et} \quad b = \bar{y} - a\bar{x} .$$

On nomme e_i le résidu tel que $y_i = \hat{y}_i + e_i$.

On trouve en statistique [vi] les différents écart-types suivants :

- pour les résidus $s_r = \sqrt{\frac{\sum_i e_i^2}{n-2}}$
- pour la pente $s_a = \frac{s_r}{\sqrt{\sum_i (x_i - \bar{x})^2}}$
- pour l'ordonnée à l'origine $s_b = s_r \sqrt{\frac{\sum_i x_i^2}{n \sum_i (x_i - \bar{x})^2}}$

Puis $\Delta a = t_{n-2} s_a$ et $\Delta b = t_{n-2} s_b$.

t_{n-2} : coefficients de Student pour $n-2$ degrés de liberté.

Vous êtes ainsi à même de mener tous les calculs.

Effectuons les calculs pour le poids en fonction de la taille:

$$\bar{x}\bar{y} = (160 \times 64 + 170 \times 66 + 180 \times 84 + 190 \times 86) / 4$$

$$\bar{x}^2 = (160^2 + 170^2 + 180^2 + 190^2) / 4$$

$$a = (13230 - 175 \times 75) / (30750 - 175^2) = 0,84 \text{ et}$$

$$b = 75 - 0,84 \times 175 = -72$$

$$s_r = \sqrt{[(64 - (0,84 \times 160 - 72))^2 + (-4,8)^2 + 4,8^2 + (-1,6)^2] / 2} \approx 5,06$$

$$s_a \approx 5,06 / \sqrt{(160 - 175)^2 + (-5)^2 + 5^2 + 15^2} \approx 0,226$$

$$\Delta a \approx 2,92 \times 0,226 \approx 0,66 \text{ avec une confiance de 90\%}$$

$$s_b \approx 5,06 \sqrt{(160^2 + 170^2 + 180^2 + 190^2) / [4(15^2 + 5^2 + 25 + 225)]} \approx 39,7$$

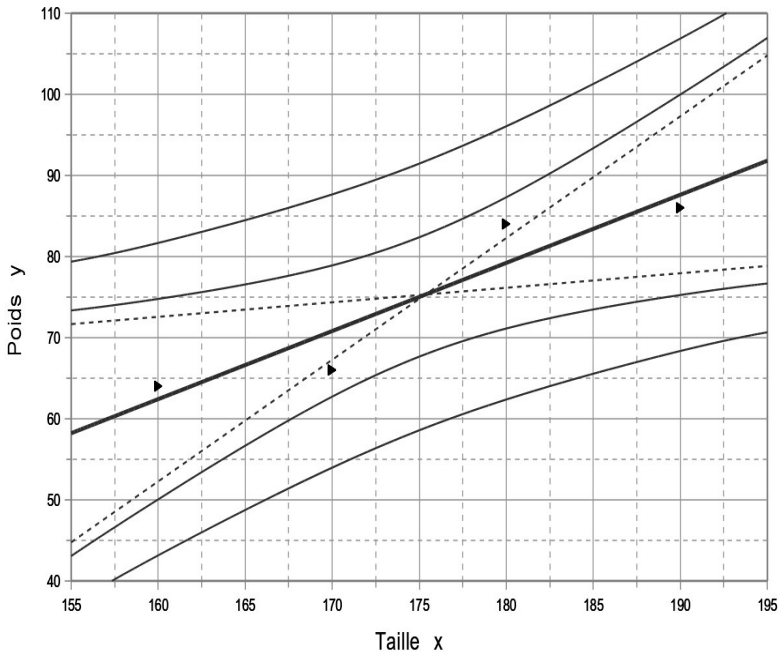
$$\Delta b \approx 2,92 \times 39,7 \approx 116 \text{ à 90\%}$$

d'où : $Poids = (0,84 \pm 0,66) Taille - (72 \pm 116)$ à 90%.

Formule ici très imprécise, ce qui n'est pas étonnant vu le peu de points et la dispersion des données. Mais la méthode de calcul est maintenant explicite et compréhensible.

Sur le graphique qui suit nous avons :

- au milieu la droite interpolée (le meilleur équilibre entre les points du dessus et ceux du dessous de cette droite).



- En pointillés sont représentés les deux droites extrêmes ($y = a_{min}x + b_{max}$ et $y = a_{max}x + b_{min}$).
- La première enveloppe correspond aux valeurs estimées de y . Intervalle de confiance de la moyenne de y_o pour une valeur x_o :

$$\Delta y_o = t_{n-2} s_r \sqrt{\frac{1}{n} + \frac{(x_o - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$$

Par exemple pour $x_o = 175$ cm nous avons $y_o = 75,0 \pm 7,4$ kg. On peut même avoir une estimation en dehors de l'intervalle par exemple pour $x_o = 195$ cm nous avons $y_o = 92 \pm 15$ kg.

- La deuxième enveloppe correspond à une prédiction si nous effectuons une nouvelle mesure. Intervalle de prédiction pour une observation y_o :

$$\Delta y_o = t_{n-2} s_r \sqrt{\frac{1}{n} + \frac{(x_o - \bar{x})^2}{\sum (x_i - \bar{x})^2} + 1}$$

Par exemple, il y a 90% de chance pour une personne de 175 cm que sa masse soit entre 58 et 92 kg (en moyenne 90% des points sont dans cette seconde enveloppe et 10% en dehors).

2) Détermination du zéro absolu

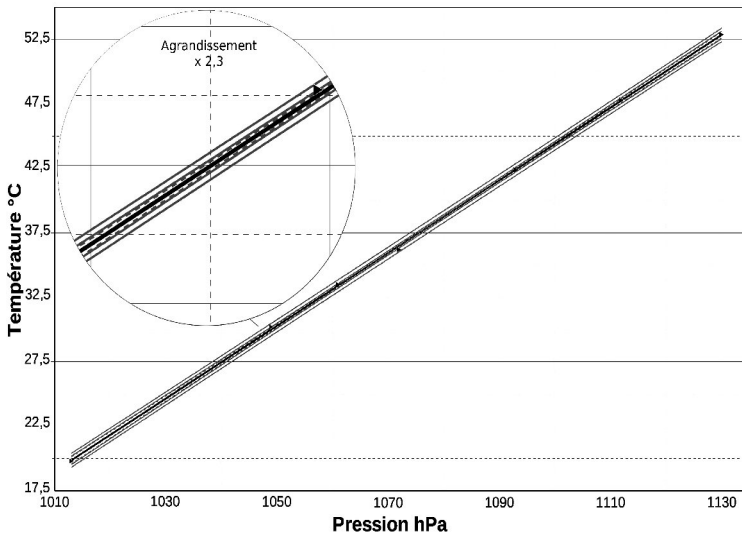
Nous étudions un gaz enfermé dans un récipient rigide de volume constant. Nous avons des sondes pour mesurer sa température et sa pression. Initialement le gaz est à température et pression ambiante, ensuite nous immergeons l'ensemble dans de l'eau chaude et nous laissons évoluer en prenant des mesures à la volée¹ :



¹ Cette expérience a réellement été réalisée le mardi 17 octobre 2006 à Bourges par M. ROUAUD et O. LEROY au Lycée Alain-Fournier.

heure	10h 15	10h 30	10h 40	10h 55	non notée	non notée	12h
température Θ (°C)	19,8	52,9	47,8	42,4	36,2	33,5	30,2
pression P (hPa)	1013	1130	1112	1093	1072	1061	1049

Nous supposons que le gaz obéit à l'équation d'état du gaz parfait $PV = nRT = nR(\Theta - \Theta_{0K})$. En traçant $\Theta(P)$ nous pouvons obtenir la température du zéro absolu : l'ordonnée à l'origine donne Θ_{0K} .



Nous voyons que la régression est bonne ($r=0,99991$) mais les points mesurés sont éloignés du zéro absolu. Par prolongement nous obtenons avec une confiance de 95% : $\Theta_{0K} = -266,0 \pm 4,8^\circ \text{C}$.

Or, nous savons que le zéro absolu est à $-273,15\text{ °C}$, ce qui n'est pas très cohérent. Nous pouvons donc supposer qu'il y a un biais et que nous n'avons pas considéré toutes les sources d'incertitudes.

3) Régression avec barres d'erreurs

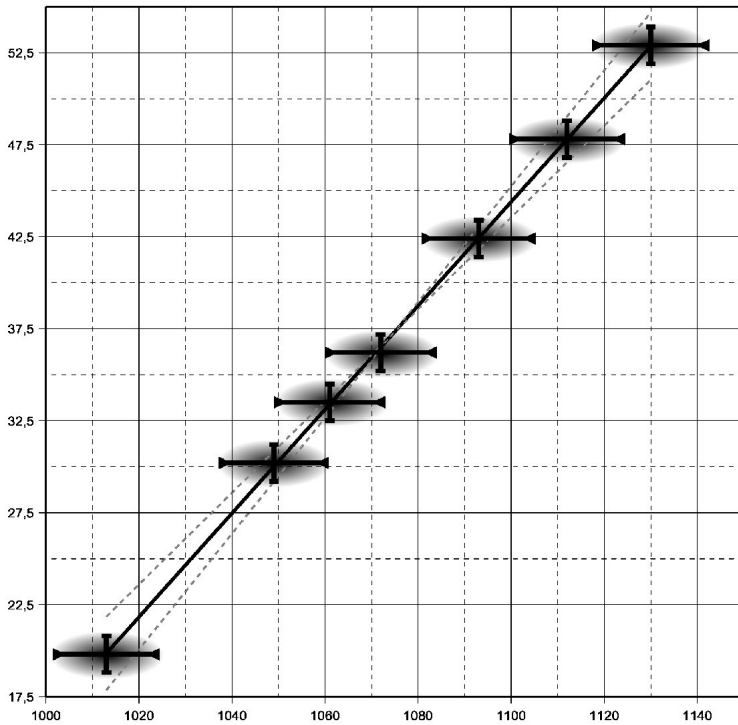
Les instruments de mesure ne sont pas parfaits et les notices indiquent leurs précisions. La précision du manomètre est à un pour cent près et celle du thermomètre au degré. Nous avons donc une incertitude sur x et y :

$$M_i(x_i \pm \Delta x_i, y_i \pm \Delta y_i)$$

Nous allons ajouter ce qu'on appelle des barres d'erreurs, nous n'avons plus des points mais des rectangles. Plus un point est précis, plus il joue un rôle important. La méthode des moindres carrés est modifiée, et la somme des carrés est pondérée par ce "poids" :

$$\sum_i w_i e_i^2 \quad \text{avec} \quad w_i = \frac{1}{(\Delta y_i)^2 + (a \Delta x_i)^2}$$

C'est une méthode itérative, nous mettons initialement une valeur de a estimée, puis la valeur de a obtenue la remplace jusqu'à l'égalité des valeurs.



Nous obtenons alors : $\Theta_{0K} = -266 \pm 35 \text{ } ^\circ\text{C}$ avec la même confiance que les incertitudes sur x_i et y_i . La valeur est, cette fois, satisfaisante. Les sources principales d'incertitudes semblent retenues.

Nous pourrions aussi considérer l'incertitude de modélisation induite par l'hypothèse de gaz parfait, mais, dans les conditions expérimentales de cette expérience, l'hypothèse est excellente. Cette source d'incertitude est négligeable devant les autres considérées ici. L'utilisation d'un modèle de gaz réel (Van der Waals par exemple) permettrait de le démontrer.

Formules [i] :

$$S^2 = \sum_i w_i [y_i - (ax_i + b)]^2$$

$$\frac{\partial S^2}{\partial b} = 0 \quad \text{et} \quad \frac{\partial S^2}{\partial a} = 0$$

conduit à :

$$b = \frac{\sum w_i y_i \sum w_i x_i^2 - \sum w_i x_i \sum w_i x_i y_i}{\Delta}$$

et

$$a = \frac{\sum w_i \sum w_i x_i y_i - \sum w_i x_i \sum w_i y_i}{\Delta}$$

avec

$$\Delta = \sum w_i \sum w_i x_i^2 - \left(\sum w_i x_i \right)^2$$

puis

$$\Delta b = \sqrt{\frac{\sum w_i x_i^2}{\Delta}}$$

et

$$\Delta a = \sqrt{\frac{\sum w_i}{\Delta}}$$

D. Exercices

Exercice 1 : Corrélations

1- On réalise neuf expériences, nous obtenons à chaque fois trois réalisations des grandeurs X_1 , X_2 et X_3 :

$i=$	1	2	3	4	5	6	7	8	9
X_1	$x_1^1 = -1$	-1	-1	0	0	0	1	1	1
X_2	$x_2^1 = 1$	0	-1	1	0	-1	1	0	-1
X_3	$x_3^1 = -1$	0	1	-1	0	1	-1	0	1

a) Sur l'ensemble des expériences déterminez les moyennes arithmétiques et les écart-types de ces trois grandeurs.

b) Tracez X_2 en fonction de X_1 . Puis $X_3(X_1)$ et $X_3(X_2)$.

c) Calculez les coefficients de corrélation r_{12} , r_{13} et r_{23} . Commentez.

2- Même chose avec les données suivantes :

X_1	0	1	-1	2	0	-1	0	-2	1
X_2	1	2	-1	2	0	-2	-1	-2	1

3- Même chose avec les données suivantes :

X_1	-1	2	-2	0	-2	2	1
X_2	-1	0	2	-2	0	2	-1

Exercice 2 : Volumes

À l'aide d'une pipette jaugée nous remplissons quatre béchers avec 100 mL d'eau chacun. Pour tester la pipette et connaître précisément la quantité d'eau, nous effectuons quatre pesées au décigramme et nous obtenons, ramené en mL, les résultats suivants pour les différents béchers :

$$V_1 = \{100,1 ; 100,0 ; 99,9 ; 100,0\}$$

1- Calculez la moyenne et l'écart-type de V_1 . Estimez la précision de la pipette avec une confiance de 95%.

Nous remplissons maintenant deux béchers et rassemblons le contenu des deux dans un seul :

$$V = V_1 + V_2.$$

Correspondant à V_1 , nous avons les mesures suivantes pour V_2 :

$$V_2 = \{100,0 ; 100,1 ; 100,0 ; 99,9\}$$

Par exemple pour la troisième mesure nous avons $V_1=99,9$ mL et $V_2=100,0$ mL.

2- Montrez que V_1 et V_2 sont des grandeurs indépendantes.

3- Calculez la moyenne de V , son écart-type et l'incertitude ΔV à 95%.

4- Pourriez-vous retrouver ce résultat avec la formule de propagation des incertitudes?

(Pour affiner le test il faudrait prendre plus de mesures, mais le principe reste le même, et les résultats restent valides car nous avons élargi avec le Student, considéré des données décorréelées et des paquets globalement gaussiens. Nous devrions aussi tenir compte des incertitudes sur les mesures -résolution- en plus de leur dispersion.)

Exercice 3 : Arbres

Nous voulons mesurer la distance d entre deux arbres. Pour cela nous disposons d'un bâton d'une longueur d'un mètre. D'un arbre à l'autre nous reportons le bâton cent fois. Nous estimons pour chaque report une incertitude de 1 cm.

Quelle est l'incertitude sur la valeur de d ?

Exercice 4 : Méthode de Bessel

C'est une méthode de focométrie qui permet de mesurer la distance focale f d'une lentille convergente. Pour cela on mesure la distance D entre un objet lumineux et son image sur un écran. Quand $D > 4f$ il existe deux positions où l'image est nette. La distance entre ces deux positions est notée d . Nous avons ensuite la distance focale de la lentille par la relation $f = (D^2 - d^2) / 4D$. Nous mesurons $D = 2000 \pm 10$ mm et $d = 536 \pm 20$ mm.

Quelle est alors l'incertitude sur f ?

Exercice 5 : Indice

Nous voulons déterminer expérimentalement l'indice n_2 d'un verre. Pour cela nous effectuons l'expérience de la réfraction d'un faisceau laser. Les lois de Descartes pour l'optique géométrique indique que $n_1 \cdot \sin(i_1) = n_2 \cdot \sin(i_2)$, avec n_i les indices lumineux et i_i les angles par rapport à la normale au dioptre. Nous avons $n_1 = n_{\text{air}} = 1$, $i_1 = 30 \pm 1^\circ$ et $i_2 = 20 \pm 2^\circ$.

Déterminez n_2 avec son incertitude.

Exercice 6 : Formule de Cauchy

Nous voulons mesurer la variation d'indice lumineux n en fonction de la longueur d'onde λ dans un milieu transparent (phénomène de dispersion). Pour cela nous utilisons un prisme, une lampe à sodium et un goniomètre. D'après la théorie la variation de $n(\lambda)$ doit suivre la formule de Cauchy dans le spectre visible :

$$n(\lambda) = A + \frac{B}{\lambda^2}$$

Le spectre de raie du sodium est connu. Pour chaque raie de longueur d'onde λ_i on calcule l'indice n_i correspondant en utilisant la formule du prisme suivante :

$$n_i = \frac{\sin\left(\frac{A + D_{m,i}}{2}\right)}{\sin\left(\frac{A}{2}\right)}$$

D_m est la déviation minimale. $A=60^\circ$ est l'angle au sommet du prisme. Ces deux angles sont mesurés à 2' près (1'=une minute d'arc et $1^\circ=60'$).

Nous obtenons ainsi le tableau de données suivant :

$\lambda(nm)$	615,7	589,2	568,5	515,2	498,1
Couleur	rouge	jaune	vert-jaune	vert	bleu-vert
D_m	$57^\circ 49,5'$	$58^\circ 9'$	$58^\circ 28'$	$59^\circ 26,5'$	$59^\circ 50'$
n	1,71276	1,71568	1,71852	1,72716	1,73060

1- Déterminez l'incertitude sur n (celle-ci est globalement constante d'une valeur à l'autre).

2- Acceptant l'expression en $1/\lambda^2$ de la formule de Cauchy, en déduire A et B avec leurs incertitudes. Que vaut le coefficient de régression r ?

3- Peut-être qu'en traçant n en fonction de $1/\lambda$ ou $1/\lambda^3$ nous aurions un meilleur alignement des points ? Nous voulons vérifier expérimentalement que la variation en $1/\lambda^2$ est bien la meilleure des relations polynomiales. Pour cela nous prenons la forme :

$$n(\lambda) = A + B \cdot \lambda^\alpha$$

Proposez une méthode pour déterminer A , B et α .

Nous pourrions vérifier le modèle car nous aurons α avec son incertitude.

Exercice 7 : Mur

Nous avons un mur d'une surface $S=72 \text{ m}^2$. La température extérieure est de 6°C et la température intérieure est maintenue à 18°C . Ce mur de 50 cm d'épaisseur est constitué de $e_p=40 \text{ cm}$ de paille compressée (conductivité thermique $\lambda_p=45 \text{ mW/K/m}$) et de $e_e=10 \text{ cm}$ d'enduit ($\lambda_e=200 \text{ mW/K/m}$). Les λ sont données à 10% près, les épaisseurs au cm et les température au demi degré.

1- Déterminez la résistance thermique, avec son incertitude, de la paille pour ce mur ($R \cdot \lambda \cdot S = e$)

2- Même chose pour l'enduit.

3- Sachant que les résistances thermiques s'associent comme les résistances électriques en série, déterminez la résistance thermique totale du mur avec son incertitude.

4- Quelle doit être la puissance minimale du chauffage de la maison rien que pour compenser les pertes par les murs ? ($\Delta T = R \cdot \Phi$)

Exercice 8 : Isolation et inertie

Nous avons une maison proche des préconisations pour une *maison énergie zéro*. Les résistances thermiques ramenées au m^2 , soit e/λ , sont pour le toit de $8 m^2.K/W$ pour les murs et le sol de $4 m^2.K/W$ et pour les huisseries de $1 m^2.K/W$ (les résistances sont connues avec une précision de 10%). La maison a une surface au sol de $36 m^2$, $54 m^2$ de toit, $82 m^2$ de mur et $8 m^2$ d'huisseries.

1- Les résistances équivalentes du toit, du mur, du sol et des huisseries étant en parallèle, déterminez la résistance thermique totale (en K/W) avec sa précision.

Les températures extérieures et intérieures sont constantes.

2- Quelle doit être la puissance minimale du chauffage de la maison pour maintenir la température intérieure constante en compensant les pertes ?

3- Nous coupons le chauffage et nous mesurons la température au cours du temps :

t en heures	0	1	2	4	5	6	8	9	10
T en °C	18	16	14	12	11	10	9	9	8

Expliquez pourquoi la décroissance ne peut être linéaire. Nous considérons la décroissance de type exponentielle : $T(t)=a.exp(-t/\tau)+b$. Déterminez b , a et τ , avec leurs incertitudes.

4- La maison est isolée par l'extérieur. Le flux thermique perdu correspond à une diminution de l'énergie emmagasinée dans la maison. Cette inertie est due à la capacité thermique C des matériaux (J/K).

a) En raisonnant sur un intervalle de temps infinitésimal dt trouvez l'équation différentielle

vérifiée par $T(t)$ et retrouvez l'expression du 3.

b) Quelle est la relation entre τ , R et C ? Déterminez C et son incertitude.

Au 3. nous pouvons aussi tenir compte des incertitudes de mesure : le temps peut-être considéré comme parfaitement connu et la température est mesuré avec un thermomètre à dilatation d'un liquide avec des graduations tous les degrés Celsius.

Pour simplifier nous avons considéré que la température extérieure reste constante (pour tenir compte des variations jour/nuit nous envisagerions des variations sinusoïdales et une approche harmonique).

Exercice 9 : Rendement

Des quantités d'engrais déterminées sont répandues sur des champs et nous obtenons les rendements suivants :

Engrais (kg/ha)	100	200	300	400	500	600	700
Rendement (Q/ha)	41	44	53	63	66	65	78

1- Déterminez la droite de régression qui passe par ce nuage de points. Pente, ordonnée à l'origine et incertitudes avec une confiance de 95%.

2- Pour 550 kg d'engrais par ha estimez le rendement.

3- Même chose en absence d'engrais.

4- Si un agriculteur répand 250 kg d'engrais par hectare, quelle est la probabilité qu'il obtienne 40 à 60 quintaux de céréales ?

Exercice 10 : Étude d'une pile

Nous cherchons à déterminer la tension à vide E et la résistance interne r .

Pour cela nous mesurons pour la pile différentes valeurs de U et I avec un voltmètre et un ampèremètre ($U = E - r.I$) :

calibre pour U :	unité : V précision $\pm 0,05\% \pm 0,003$									
U (V)	4,731	4,731	4,730	4,728	4,724	4,724	4,722	4,721	4,719	4,716
calibres pour I :	unité : μA précision $\pm 0,2\% \pm 0,03$			unité : mA précision $\pm 0,2\% \pm 0,0003$						
I	92,83	115,45	152,65	0,2352	0,4686	0,5200	0,5841	0,6661	0,7750	0,9264

1- Sans les barres d'erreurs : déterminez $E \pm \Delta E$ et $r \pm \Delta r$.

2- Même chose en incluant cette fois les incertitudes de chaque mesure indiquées dans la notice du fabricant du multimètre.

Exercice 11 : Focométrie

Nous désirons déterminer la distance focale f' d'une lentille convergente. La lentille est en O et fait l'image net en A' d'un point objet A . Nous mesurons OA , OA' et leurs incertitudes (sont incluses toute les sources d'incertitudes, géométriques, latitude de mise au point et modélisation). Nous considérons que la lentille vérifie la relation de conjugaison :

$$1/OA' + 1/OA = 1/f'$$

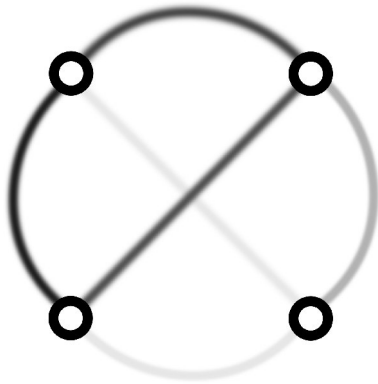
Déterminez f' par l'intermédiaire d'une régression linéaire. Tableau page suivante.

Données expérimentales (mm) :			
OA	ΔOA	OA'	$\Delta OA'$
635	5	150	15
530	5	160	17
496	5	164	15
440	5	172	18
350	5	191	20
280	5	214	25
210	5	292	28
150	5	730	102

Pour vérifier vos calculs de corrélation vous pouvez utiliser la fonction du tableur OOo : =COEFFICIENT.CORRELATION(*;**).

Courbes : Insertion>Diagramme...>XY(Dispersion)>etc.

Vous pouvez utiliser le fichier IncertitudesLibresOOo32.ods sur le site www.incertitudes.fr pour réaliser les régressions. Feuille 2, régression simple et feuille 3 avec barres d'erreurs.



III. APPROCHES MULTICRITÈRES

Les approches multicritères correspondent à un ensemble d'outils d'aide à l'analyse et à la décision en présence d'un nombre important de paramètres. Nous verrons deux méthodes. La méthode par pondération, rapide à comprendre et à mettre en œuvre, mais d'un pouvoir d'analyse sommaire. Puis nous étudierons l'analyse en composantes principales, plus difficile à maîtriser, mais qui permet une étude fine et approfondie.

Pour illustrer, nous nous appuierons dans les deux cas, sur l'exemple suivant du choix d'un isolant :

Données :

	Densité kg/m ³	Épais. cm	Prix €/m ²	Confort sur 20
1. Ouate de cellulose	55	25	14	16
2. Laine de roche en vrac	25	36	18	20
3. Perlite expansée	90	25	30	15
4. Polystyrène expansé	18	35	28	20
5. Laine de verre ou roche en rouleau	18	42	34	20
6. Laine de verre ou roche semi-lourde	70	21	36	18
7. Laine de chanvre en rouleau	25	32	41	20
8. Polystyrène extrudé	35	25	49	20
9. Laine de bois	150	18	50	13
10. Laine roche pour sol	130	19	54	15
11. Polyurethane	35	26	106	20

A. Méthode par pondération

Soit un ensemble d'individus qui ont un ensemble de caractéristiques. En statistique, nous comprenons par individus toutes entités semblables pourvues de propriétés multiples. Les caractéristiques ont des unités différentes. De plus les données pour chaque caractéristique sont différemment centrées et dispersées. Pour normaliser les différences, nous allons attribuer des notes :

- caractéristique positive : $note = \frac{V - V_{min}}{E} \times 20$
- caractéristique négative : $note = \frac{V_{max} - V}{E} \times 20$

V : valeur

V_{min} : valeur minimale

$E = V_{max} - V_{min}$: étendue

V_{max} : valeur maximale

Pour la note finale nous pondérons les différentes caractéristiques :

$$note_{finale} = \sum_j p_j n_j$$

avec $\sum_{j=1}^p p_j = 1$ et n_j : note de la $j^{\text{ème}}$ caractéristique.

Pour notre exemple, concentrons-nous sur trois caractéristiques des isolants : l'épaisseur, le prix et le confort.

Nous prenons deux cas, l'un, où le plus important c'est le prix (prix 50%, confort 30% et épaisseur 20%), et un second cas, où, avant tout le confort et l'encombrement du matériaux important sans trop regarder au prix (confort 50%, épaisseur 40% et prix 10%). Le choix final dépendra de ces priorités.

Priorités 1 :	Épaisseur			Prix			Confort			Note totale
	cm	Note	pondérée	€/m ²	Note	pondérée	sur 20	Note	pondérée	
Pondération :	0,2			0,5			0,3			
Ouate de cellulose	25	14	2,8	14	20	10,0	16	9	2,6	15
Laine de roche en vrac	36	5	1,0	18	19	9,6	20	20	6,0	17
Perlite expansée	25	14	2,8	30	17	8,3	15	6	1,7	13
Polystyrène expansé	35	6	1,2	28	17	8,5	20	20	6,0	16
Laine de verre ou roche en rouleau	42	0	0,0	34	16	7,8	20	20	6,0	14
Laine de verre ou roche semi-lourde	21	18	3,5	36	15	7,6	18	14	4,3	15
Laine de chanvre en rouleau	32	8	1,7	41	14	7,1	20	20	6,0	15
Polystyrène extrudé	25	14	2,8	49	12	6,2	20	20	6,0	15
Laine de bois	18	20	4,0	50	12	6,1	13	0	0,0	10
Laine roche pour sol	19	19	3,8	54	11	5,7	15	6	1,7	11
Polyurethane	26	13	2,7	106	0	0,0	20	20	6,0	9

moyennes :	28	12		42	14		18	14		14
étendues :	24	20		92	20		7	20		8
écart-types :	8	6		25	5		3	8		3

Priorités 2 :	Épaisseur			Prix			Confort			Note totale
	cm	Note	pondérée	€/m ²	Note	pondérée	sur 20	Note	pondérée	
Pondération :	0,4			0,1			0,5			
Ouate de cellulose	25	14	5,7	14	20	2,0	16	9	4,3	12
Laine de roche en vrac	36	5	2,0	18	19	1,9	20	20	10,0	14
Perlite expansée	25	14	5,7	30	17	1,7	15	6	2,9	10
Polystyrène expansé	35	6	2,3	28	17	1,7	20	20	10,0	14
Laine de verre ou roche en rouleau	42	0	0,0	34	16	1,6	20	20	10,0	12
Laine de verre ou roche semi-lourde	21	18	7,0	36	15	1,5	18	14	7,1	16
Laine de chanvre en rouleau	32	8	3,3	41	14	1,4	20	20	10,0	15
Polystyrène extrudé	25	14	5,7	49	12	1,2	20	20	10,0	17
Laine de bois	18	20	8,0	50	12	1,2	13	0	0,0	9
Laine roche pour sol	19	19	7,7	54	11	1,1	15	6	2,9	12
Polyurethane	26	13	5,3	106	0	0,0	20	20	10,0	15

Dans la première situation la laine de roche en vrac sera la mieux adaptée, dans la deuxième, ce sera le polystyrène extrudé.

Bien sûr, si nous ajoutons d'autres critères comme la densité ou le bilan écologique (traité dans un devoir page 130), la conclusion pourra évoluer.

B. Analyse en Composantes Principales

1) Principes

Soit X une caractéristique. Celle-ci prend des valeurs différentes suivant n individus. Ces réalisations, encore appelées mesures ou observations, sont notées :

$$\{x_1, x_2, \dots, x_i, \dots, x_n\}$$

ou plus simplement $\{x_i\}$ avec $1 \leq i \leq n$

Prenons maintenant p caractéristiques, nous avons donc pas seulement une, mais p variables aléatoires:

$$\{X_1, X_2, \dots, X_j, \dots, X_p\}$$

, et les réalisations de X_j se notent :

$$\{x_{1j}, x_{2j}, \dots, x_{ij}, \dots, x_{nj}\} \quad 8$$

La matrice des données regroupe l'ensemble des informations :

$$M = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{ip} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{np} \end{pmatrix}$$

8 Les indices sont ici inversés par rapport à la notation page 37.

Comme $n \neq p$ cette matrice n'est pas, en général, carrée.

Une colonne représente la variabilité d'une caractéristique selon les individus :

$$X_j = \begin{pmatrix} x_{1j} \\ \cdot \\ \cdot \\ x_{ij} \\ \cdot \\ \cdot \\ \cdot \\ x_{nj} \end{pmatrix}$$

Une ligne de la matrice représente les différentes caractéristiques d'un individu. Mais un individu peut aussi être représenté par un vecteur colonne :

$$\vec{I}_i = \begin{pmatrix} x_{i1} \\ \cdot \\ \cdot \\ x_{ij} \\ \cdot \\ \cdot \\ \cdot \\ x_{ip} \end{pmatrix}$$

Dans l'espace des individus, les individus sont un ensemble de points. Cet espace a pour axes les p caractéristiques et la base $\{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_j, \dots, \vec{e}_p\}$.

Nous allons commencer par calculer la matrice des données normalisée m (centrage et réduction) :

$$\hat{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}$$

Valeur

moyenne :

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

Écart-type :

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}}$$

Les caractéristiques fluctuent autour d'une valeur moyenne. Les valeurs moyennes et les dispersions dépendent de chaque caractéristique. Or l'analyse en composantes principales (ACP) est basée sur les coefficients de corrélations linéaires entre les différentes caractéristiques, et ces coefficients ne dépendent pas de la moyenne et de l'écart-type :

$$r_{jk} = \frac{\sum_{i=1}^n \hat{x}_{ij} \hat{x}_{ik}}{n-1}$$

De ce fait il est plus clair de travailler directement avec la matrice m normale. Si nous gardions M les résultats obtenus seraient les mêmes. Nous trouverons au final, dans un nouvel espace, une nouvelle matrice m' des données que nous pourrions aussi normaliser.

Il faut donc bien garder à l'esprit que pour l'ACP seules les corrélations linéaires comptent. Qu'une grandeur varie sur un millimètre, ou un mètre, autour d'un kilomètre, ou un micromètre, peu où prou, il importe seulement de savoir si cette grandeur varie, ou pas, dans le même sens que les autres. Il convient donc de réfléchir avant de commencer une étude afin de savoir si c'est bien ce type d'analyse qui convient à votre problématique.

Pour les isolants nous obtenons pour M et m :

	x_1	x_2	x_3	x_4
l_1	$x_1^1 = 55$	$x_2^1 = 25$	$x_3^1 = 14$	$x_4^1 = 16$
l_2	$x_1^2 = 25$	$x_2^2 = 36$	$x_3^2 = 18$	$x_4^2 = 20$
l_3	$x_1^3 = 90$	$x_2^3 = 25$	$x_3^3 = 30$	$x_4^3 = 15$
l_4	$x_1^4 = 18$	$x_2^4 = 35$	$x_3^4 = 28$	$x_4^4 = 20$
l_5	$x_1^5 = 18$	$x_2^5 = 42$	$x_3^5 = 34$	$x_4^5 = 20$
l_6	$x_1^6 = 70$	$x_2^6 = 21$	$x_3^6 = 36$	$x_4^6 = 18$
l_7	$x_1^7 = 25$	$x_2^7 = 32$	$x_3^7 = 41$	$x_4^7 = 20$
l_8	$x_1^8 = 35$	$x_2^8 = 25$	$x_3^8 = 49$	$x_4^8 = 20$
l_9	$x_1^9 = 150$	$x_2^9 = 18$	$x_3^9 = 50$	$x_4^9 = 13$
l_{10}	$x_1^{10} = 130$	$x_2^{10} = 19$	$x_3^{10} = 54$	$x_4^{10} = 15$
l_{11}	$x_1^{11} = 35$	$x_2^{11} = 26$	$x_3^{11} = 106$	$x_4^{11} = 20$

M =

Moyennes : 59,18 27,64 41,82 17,91
 Écart-types : 46,09 7,65 24,81 2,66

	x_1	x_2	x_3	x_4
l_1	$x_1^1 = -0,09$	$x_2^1 = -0,34$	$x_3^1 = -1,12$	$x_4^1 = -0,72$
l_2	$x_1^2 = -0,74$	$x_2^2 = 1,09$	$x_3^2 = -0,96$	$x_4^2 = 0,79$
l_3	$x_1^3 = 0,67$	$x_2^3 = -0,34$	$x_3^3 = -0,48$	$x_4^3 = -1,09$
l_4	$x_1^4 = -0,89$	$x_2^4 = 0,96$	$x_3^4 = -0,56$	$x_4^4 = 0,79$
l_5	$x_1^5 = -0,89$	$x_2^5 = 1,88$	$x_3^5 = -0,32$	$x_4^5 = 0,79$
l_6	$x_1^6 = 0,23$	$x_2^6 = -0,87$	$x_3^6 = -0,23$	$x_4^6 = 0,03$
l_7	$x_1^7 = -0,74$	$x_2^7 = 0,57$	$x_3^7 = -0,03$	$x_4^7 = 0,79$
l_8	$x_1^8 = -0,52$	$x_2^8 = -0,34$	$x_3^8 = 0,29$	$x_4^8 = 0,79$
l_9	$x_1^9 = 1,97$	$x_2^9 = -1,26$	$x_3^9 = 0,33$	$x_4^9 = -1,84$
l_{10}	$x_1^{10} = 1,54$	$x_2^{10} = -1,13$	$x_3^{10} = 0,49$	$x_4^{10} = -1,09$
l_{11}	$x_1^{11} = -0,52$	$x_2^{11} = -0,21$	$x_3^{11} = 2,59$	$x_4^{11} = 0,79$

m =

Moyennes : 0 0 0 0
 Écart-types : 1 1 1 1

L'idée fondamentale de cette étude est basée sur la matrice de corrélation. Nous cherchons les corrélations entre les différentes caractéristiques. Si un ensemble de caractéristiques sont très corrélées entre elles, cela signifie qu'une seule suffirait à décrire notre système sans perte importante d'information.

Comme nous sommes mieux habitué à visualiser les choses en deux dimensions, nous cherchons deux caractéristiques, combinaisons de toutes les autres, qui contiendraient la majorité de l'information. Ainsi nous pourrions nous ramener à l'essentiel d'un problème en peu de paramètres et simplifier la vision avec des caractéristiques mieux adaptées auxquelles nous n'aurions pas pensées à priori.

Coefficients de corrélations linéaires :

$$r_{jk} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{(n-1)\sigma_j\sigma_k}$$

avec $r_{jj} = 1$ et $r_{jk} = r_{kj}$

nous avons alors une matrice carrée appelée la matrice des corrélations :

$$r = \begin{pmatrix} 1 & r_{12} & r_{13} & \dots & r_{1p} \\ r_{12} & 1 & \dots & \dots & \dots \\ r_{13} & \dots & 1 & \dots & \dots \\ \dots & \dots & \dots & 1 & \dots \\ r_{1p} & \dots & \dots & \dots & 1 \end{pmatrix} \quad -1 \leq r_{ij} \leq 1$$

Nous diagonalisons ensuite la matrice r :

$$r' = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_j & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_p \end{pmatrix}$$

$$\text{avec } \sum_{j=1}^p \lambda_j = p = \text{tr}(r) = \text{tr}(r')$$

La somme des valeurs propres λ_j est p , le nombre de caractéristiques. La somme des valeurs sur la diagonale d'une matrice carrée est appelée trace de la matrice, celle-ci est invariante par changement de base. Nous les plaçons tel que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$.

Nous avons alors une nouvelle base, appelée base des vecteurs propres $\{e_{\vec{1}}, e_{\vec{2}}, \dots, e_{\vec{j}}, \dots, e_{\vec{p}}\}$, pour la projection de nos individus sur les nouvelles caractéris-

tiques $\{X'_1, X'_2, \dots, X'_j, \dots, X'_p\}$.

Dans ce nouvelle espace toutes les nouvelles caractéristiques sont décorréliées entre elles. C'est une grande avancée pour l'analyse, nous savons maintenant qu'elles sont les "vraies" caractéristiques. Nous avons le point de vue sur le problème qui donne la vision la plus simple et la plus claire.

De plus avec les valeurs propres nous avons un ordre d'importance des caractéristiques. Le pourcentage d'information contenu dans une caractéristique X'_j est :

$$\%o_j = \frac{\lambda_j}{p}$$

Il n'est pas rare que deux caractéristiques sur une dizaine contiennent à elles seules 90% de l'information totale d'un système de données. C'est un outil descriptif très puissant.

Les valeurs propres se calculent en résolvant l'équation $\det(r - \lambda I) = 0$. Nous calculons le déterminant de la matrice des corrélations à laquelle nous avons soustrait la matrice identité multipliée par la valeur propre recherchée. Il s'agit d'un problème classique de mathématiques abordé en début d'enseignement supérieur scientifique. Les calculs sont effectués sur des exemples dans le deuxième exercice. Nous avons ensuite l'habitude de classer les valeurs propres en ordre décroissant dans la matrice r' .

Les vecteurs propres sont déterminés en résolvant l'équation :

$$r \vec{e}'_j = \lambda_j \vec{e}'_j$$

Pour obtenir la nouvelle matrice réduite des données avec les individus en fonction des nouvelles caractéristiques nous avons besoin de la matrice de passage P . Celle-ci est une matrice carrée dont les colonnes sont les composantes des vecteurs propres. Ensuite $m P = m'$.

Ce travail effectué pour nos isolants donne :

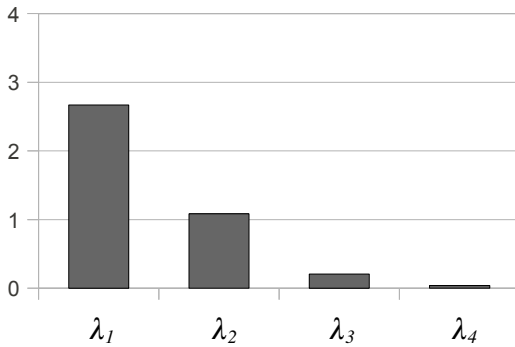
$$r = \begin{vmatrix} 1,000 & -0,810 & 0,114 & -0,936 \\ -0,810 & 1,000 & -0,307 & 0,725 \\ 0,114 & -0,307 & 1,000 & 0,097 \\ -0,936 & 0,725 & 0,097 & 1,000 \end{vmatrix}$$

$$r' = \begin{vmatrix} 2,671 & 0,000 & 0,000 & 0,000 \\ 0,000 & 1,083 & 0,000 & 0,000 \\ 0,000 & 0,000 & 0,206 & 0,000 \\ 0,000 & 0,000 & 0,000 & 0,040 \end{vmatrix}$$

Cumul des valeurs propres :

Valeurs propres		%	
		cumul	%
2,67	67	2,67	67
1,08	27	3,75	94
0,21	5	3,96	99
0,04	1	4	100

Éboulis des valeurs propres :



La matrice de passage permet d'avoir la nouvelle base en fonction de l'ancienne, par exemple :

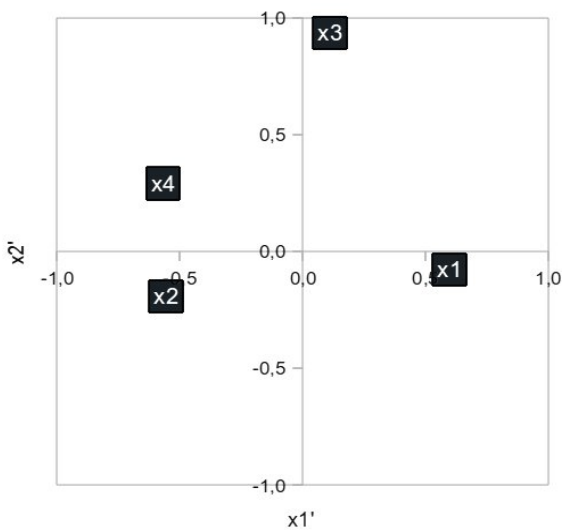
$$\vec{e}'_1 \simeq 0,60 \vec{e}_1 - 0,56 \vec{e}_2 + 0,11 \vec{e}_3 - 0,57 \vec{e}_4$$

$$P = \begin{vmatrix} 0,60 & -0,08 & -0,34 & -0,72 \\ -0,56 & -0,19 & -0,81 & -0,06 \\ 0,11 & 0,94 & -0,31 & 0,14 \\ -0,57 & 0,29 & 0,37 & -0,68 \end{vmatrix}$$

$$\text{norme :} \quad 1 \quad 1 \quad 1 \quad 1$$

Or X_1' et X_2' contiennent 94% de l'information du système.

Et tracer X_1 , X_2 , X_3 et X_4 en fonction de X_1' et X_2' donne une bonne vision globale de l'information.



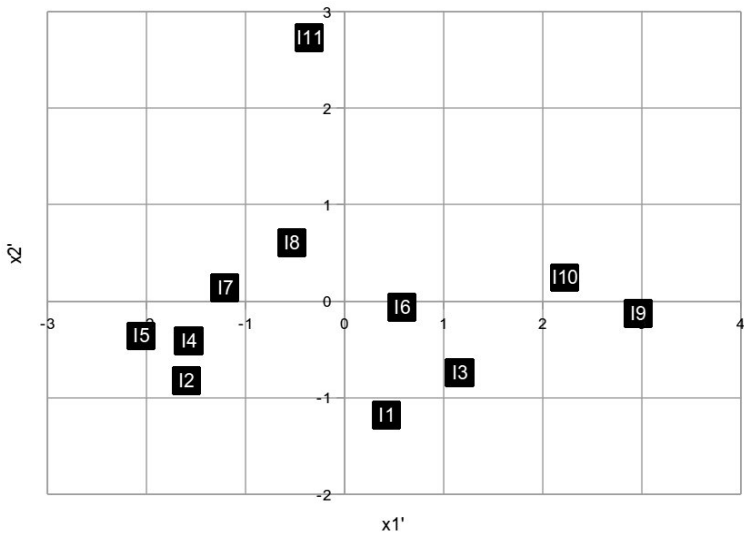
La perte d'information est de seulement 6%.

Nous obtenons ensuite la matrice des données en fonction des composantes principales :

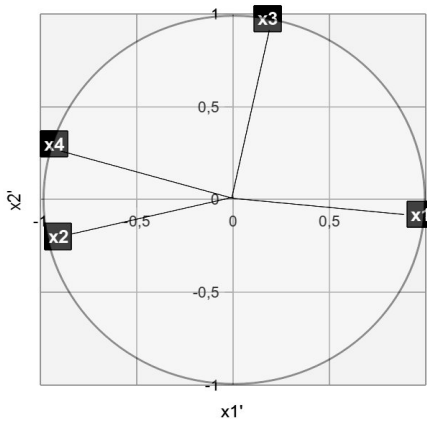
$m' =$

	x_1'	x_2'	x_3'	x_4'
l_1	0,42	-1,18	0,39	0,42
l_2	-1,60	-0,83	-0,04	-0,19
l_3	1,16	-0,74	-0,21	0,21
l_4	-1,58	-0,41	-0,01	-0,01
l_5	-2,06	-0,36	-0,82	-0,03
l_6	0,58	-0,06	0,70	-0,18
l_7	-1,21	0,14	0,09	-0,03
l_8	-0,54	0,60	0,66	-0,09
l_9	2,96	-0,13	-0,44	-0,06
l_{10}	2,22	0,24	-0,17	-0,24
l_{11}	-0,36	2,73	-0,15	0,21

Nous représentons enfin ces données en fonctions des composantes principales :



Pour trouver rapidement le sens de ces deux nouvelles composantes principales il est pratique de tracer un **cercle des corrélations** :



Pour cela nous avons calculé la matrice des corrélations entre anciennes et nouvelles variables (composantes principales) :

	x_1'	x_2'	x_3'	x_4'
x_1	$r_{11'} = 0,97$	$r_{12'} = -0,08$	$r_{13'} = -0,16$	$r_{14'} = -0,14$
x_2	$r_{21'} = -0,91$	$r_{22'} = -0,20$	$r_{23'} = -0,37$	$r_{24'} = -0,01$
x_3	$r_{31'} = 0,18$	$r_{32'} = 0,97$	$r_{33'} = -0,14$	$r_{34'} = 0,03$
x_4	$r_{41'} = -0,93$	$r_{42'} = 0,30$	$r_{43'} = 0,17$	$r_{44'} = -0,14$

L'interprétation peut maintenant commencer :

Tout d'abord les quatre caractéristiques sont proches des bords du cercle de rayon 1. Elles vont, donc, toutes jouer un rôle important. Si nous avons des caractéristiques proches du centre elles seraient peu ou pas prises en considération dans la description.

Nous voyons que X_1 , X_2 et X_4 sont très corrélés avec la première composante principale X_1' . X_1 positivement, X_2 et X_4 négativement. Tout d'abord, quand l'épaisseur augmente, le confort aussi (toujours en terme probabiliste, bien sûr). De plus, plus la densité d'un matériau est grande, moins il est confortable (on sait que les bons isolants sont souvent remplis d'air, donc peu denses). Aussi plus la densité d'un matériau est grande et moins il est épais (et volumineux, ce qui est logique aussi). Apparaît donc une caractéristique nouvelle qui combine ces trois paramètres et qui contient 67% de l'information du problème.

Sur le cercle des corrélations X_3 est selon X_2' (il est très corrélés positivement avec lui $r_{32'}$ vaut environ 0,97). La deuxième composante principale correspond donc simplement au prix (poids 27%).

Le prix des matériaux est sans lien avec leur confort. Des grandeurs décorréliées sur le cercle des corrélations sont à angle droit avec le centre. Plus elles sont corrélées dans le même sens, plus l'angle est faible ; et si elles sont corrélées en sens contraire, l'angle vaut 180° .

Regardons maintenant les individus et essayons de les regrouper en catégories :

Tout d'abord un isolé, l'individu 11, qui à un confort/épaisseur moyen et un prix élevé.

Le groupe le plus intéressant en bas à gauche (2, 5 et 4), qui correspond à des isolants à la fois pas cher et confortable (avec une bonne épaisseur).

Et puis ceux qui partent vers la droite, d'un prix moyen mais pas confortable, donc à éviter, à moins que le volume disponible pour l'isolation soit faible.

A vous de prendre votre décision ; cet outil permet de choisir son isolant en connaissance de cause.

Dans le deuxième devoir sont étudiées d'autres isolants avec l'éco-bilan, l'isolation et le confort d'été.

2) Une deuxième illustration

Nous considérons maintenant un exemple plus mathématique qui permet de visualiser et d'élargir notre compréhension de la méthode.

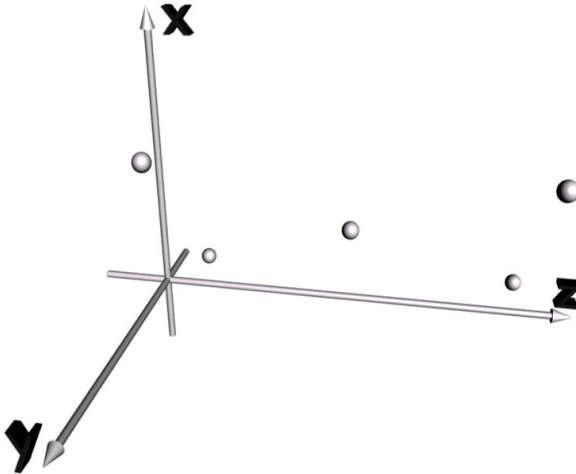
Nous prenons des points dans l'espace. Les caractéristiques prises sont les positions x , y et z . Nous avons cinq points : $M_1(1,1,1)$, $M_2(1,1,5)$, $M_3(2,2,3)$, $M_4(3,3,1)$ et $M_5(3,3,5)$.

Matrice de données :

$$M = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 5 \\ 2 & 2 & 3 \\ 3 & 3 & 1 \\ 3 & 3 & 5 \end{pmatrix}$$

Matrice réduite :

$$m = \begin{pmatrix} -1 & -1 & -1 \\ -1 & -1 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix}$$



Matrice des corrélations : $r = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

d'où : $\lambda_1=2$, $\lambda_2=1$ et $\lambda_3=0$.

Matrice de passage : $P = \begin{pmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \end{pmatrix}$

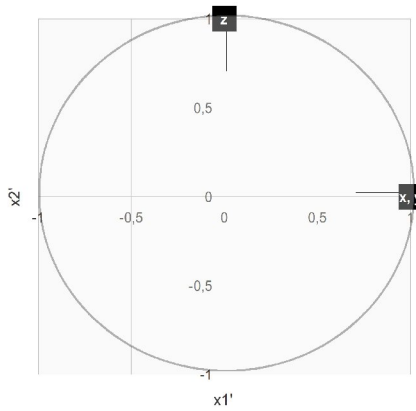
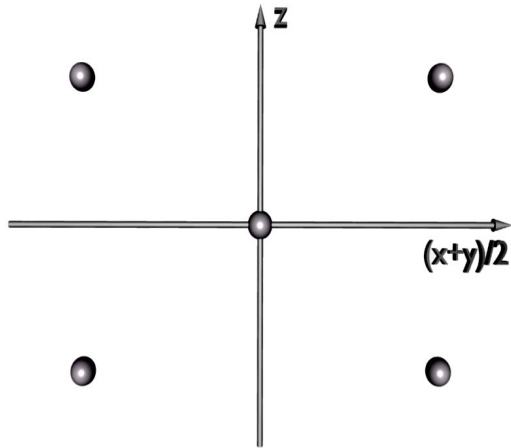
d'où : $\vec{e}'_1 = (\vec{e}_1 + \vec{e}_2)/\sqrt{2}$, poids $w_1=2/3$, $\vec{e}'_2 = \vec{e}_3$,
poids $w_2=1/3$ et $\vec{e}'_3 = (\vec{e}_2 - \vec{e}_1)/\sqrt{2}$, poids $w_3=0$.

Matrice des individus dans la nouvelle base :

$$m' = \begin{pmatrix} -\sqrt{2} & -1 & 0 \\ -\sqrt{2} & 1 & 0 \\ 0 & 0 & 0 \\ \sqrt{2} & -1 & 0 \\ \sqrt{2} & 1 & 0 \end{pmatrix}$$

Corrélations entre nouvelles et
anciennes caractéristiques : $\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$

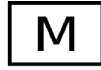
D'où le graphe avec les composantes principales $X'_1=(x+y)/2$ et $X'_2=z$ et le cercle des corrélations suivants :



Ressort ainsi la figure suivante : un carré avec un point au milieu. À partir d'un rectangle incliné dans l'espace, on a une figure simple dans un plan qui contient toutes les informations relatives.

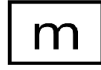
3) Résumé

Matrice des données
(individus selon les lignes et
grandeurs selon les colonnes)



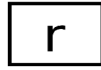
$$x_{ij}$$

Matrice des données
réduite



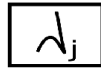
$$\hat{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}$$

Matrice des corrélations



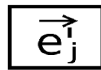
$$r_{jk} = \frac{\sum_{i=1}^n \hat{x}_{ij} \hat{x}_{ik}}{n-1}$$

Valeurs propres
(ordonnées dans l'ordre
décroissant)



$$\det(r - \lambda I) = 0$$

Vecteurs propres
correspondants



$$w_j = \frac{\lambda_j}{p}$$

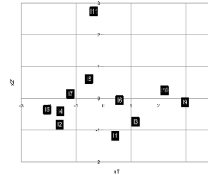
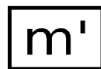
$$r \vec{e}_j = \lambda_j \vec{e}_j$$

Matrice de passage (les
colonnes sont les vecteurs
propres)



$$m' = m P$$

Matrice des données
dans la nouvelle base des
vecteurs propres
(possibilité de normaliser m')

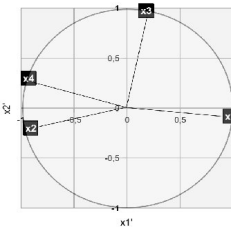


Représentation
des individus
dans
le nouvel espace

↓

Matrice
des corrélations
entre nouvelles
et anciennes
grandeurs :

**Cercle
des
corrélations**



	X_1'	X_2'	...	X_p'
X_1	$r_{11'}$	$r_{12'}$...	$r_{1p'}$
X_2	$r_{21'}$	$r_{22'}$...	$r_{2p'}$
...
X_p	$r_{p1'}$	$r_{p2'}$...	$r_{pp'}$

C. Exercices

Exercice 1 : Fournisseurs

Nous voulons choisir entre deux fournisseurs. Ceux-ci sont notés selon trois critères : les tarifs, la qualité et la solidité financière.

Pour le fournisseur 1 les notes sont, sur dix : 4, 8 et 5.
Pour le deuxième : 6, 3 et 6.

Quel fournisseur choisir si l'on pondère 50% pour les tarifs, 30% pour la qualité et le reste pour la solidité financière ?

Exercice 2 : Principe de l'ACP

Réalisez l'analyse mathématique pour les deux matrices de données suivantes (la première sans l'aide d'un ordinateur et pour la seconde vous pourrez utiliser le logiciel Scilab) :

	x_1	x_2	x_3	x_4
l_1	$x_1^1 = -2$	$x_2^1 = 4$	$x_3^1 = 0$	$x_4^1 = 8$
l_2	$x_1^2 = 0$	$x_2^2 = 3$	$x_3^2 = -3$	$x_4^2 = 3$
l_3	$x_1^3 = 0$	$x_2^3 = 3$	$x_3^3 = 3$	$x_4^3 = 13$
l_4	$x_1^4 = 2$	$x_2^4 = 2$	$x_3^4 = 0$	$x_4^4 = 8$

	x_1	x_2	x_3
l_1	$x_1^1 = 1$	$x_2^1 = 3$	$x_3^1 = 2$
l_2	$x_1^2 = 2$	$x_2^2 = 1$	$x_3^2 = 3$
l_3	$x_1^3 = 3$	$x_2^3 = 2$	$x_3^3 = 1$

Pour vous exercer davantage sur l'ACP vous avez deux cas concrets traités en devoir.

Pour faire vos calculs pour l'ACP vous avez Scilab un très bon logiciel de calcul numérique libre, gratuit et multiplateforme. Vous pouvez ensuite présenter vos résultats avec un tableur.

Par exemple pour le devoir Suricate : ouvrir le logiciel Scilab (prononcez : " saille lab", contraction de *Scientific Laboratory* en anglais) puis *Fichier > exécuter...* le fichier acpsuricate.sce :

```
function cov=covariance(A)
    moyenne = mean(A,"r");
    n = size(A,"r")
    Acent = A - ones(n,1)*moyenne;
    cov=(Acent'*Acent)/(n-1);
endfunction;

function c=correlation(A)
    cov=covariance(A);
    ectinv=(1)./sqrt(diag(cov));
    c=diag(ectinv)*cov*diag(ectinv);
endfunction;

function [c]=val_prop(A);
    [Diag,Base]=bdiag(A);
    c=gsort(diag(Diag));
endfunction;

function c=vect_prop(A);
    [Diag,Base]=bdiag(A);
    if Base(:,1)==-abs(Base(:,1)) then Base=-
1*Base; end;
    [Valpr,k]=gsort(diag(Diag));
    c = Base(:,k);
endfunction;
function c=acp_indiv(A);
    mat_cor=correlation(A);
    X=centrer_reduire(A);
```

```

    VectP=vect_prop(mat_cor);
    c = X*VectP;
endfunction;

```

```

function c=centrer_reduire(A)
    moyenne = mean(A,"r");
    cov=covariance(A);
    n = size(A,"r");
    Acent = A - ones(n,1)*moyenne;
    c=Acent*(diag((1)./sqrt(diag(cov))));
endfunction;

```

```

function c=acp_var(A);
    mat_cor=correlation(A);
    X=centrer_reduire(A);
    VectP=vect_prop(mat_cor);
    ValP=val_prop(mat_cor);
    c = VectP*diag(sqrt(ValP));
endfunction;

```

```

M=[78 357 24 4.5 10 1 5 7.2 ; 190 1300 85 44
38 10 165 6 ; 486 403 84 10 1187 3.2 9.1 7.2 ;
555 403 110 11 1479 4 14 7.2 ; 90 4368 11 322
174 132 1708 6.6 ; 202 402 43 8 336 2 4.7
7.5 ; 11.5 71 8 13.5 8.1 6.2 11.6 6.95 ; 103
2989 10 235 138 66 1172 6.8 ; 253 820 11 4 25
3 7 6]

```

```

m=centrer_reduire(M)
r=correlation(m)
val_prop(r)
vect_prop(r)
acp_indiv(m)
acp_var(m)

```


IV. COMPLÉMENTS

A. Mesure avec une règle

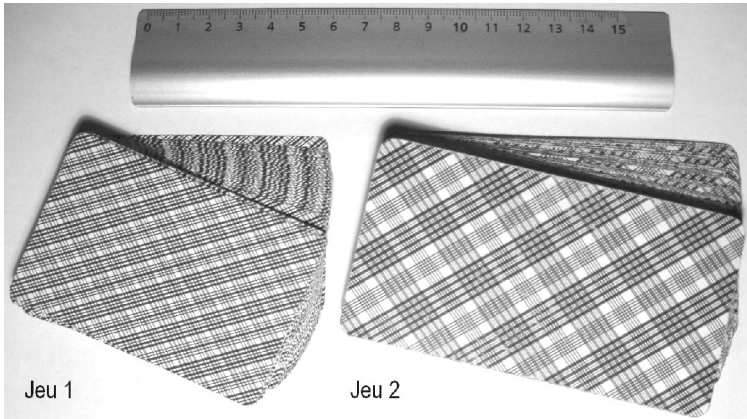
Article publié dans le BUP [ii].

RÉSUMÉ

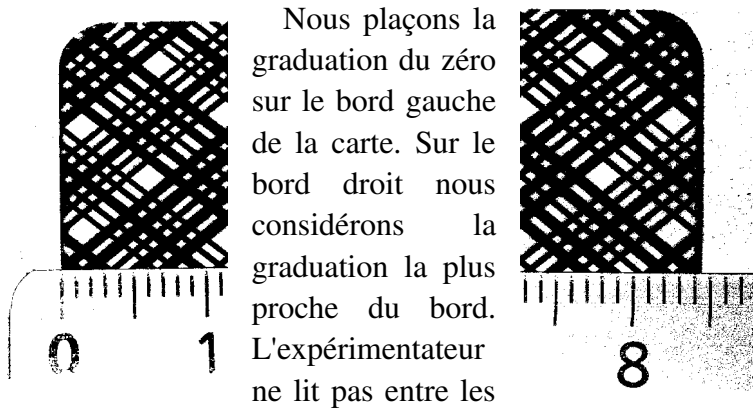
La mesure d'une grandeur par un système d'acquisition induit de part sa résolution une erreur de discrétisation. Nous nous attachons ici à la mesure d'une longueur avec une règle graduée. Ce type de mesure nous amène à considérer une loi de probabilité continue uniforme. Nous utilisons ensuite un produit de convolution pour déterminer l'incertitude avec sa confiance d'une somme de longueurs. Nous généralisons finalement au cas général du calcul d'incertitudes pour des variables aléatoires indépendantes en utilisant la formule de propagation des erreurs.

INTRODUCTION

Nous voulons mesurer des longueurs et évaluer les incertitudes le plus précisément possible. Incertitudes sur les valeurs mesurées et leurs sommes. Nous disposons d'une règle de 15cm graduée au millimètre et de deux jeux de cartes. La règle est supposée parfaite et les cartes de chaque jeu à l'identique.



1. MESURE DE LA LONGUEUR D'UNE CARTE

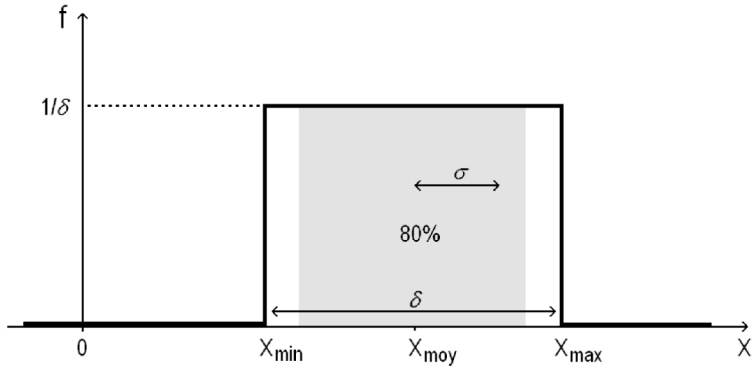


Nous plaçons la graduation du zéro sur le bord gauche de la carte. Sur le bord droit nous considérons la graduation la plus proche du bord. L'expérimentateur ne lit pas entre les graduations. L'épaisseur des traits qui délimitent une graduation est considérée comme négligeable devant la largeur de cette graduation. Nous obtenons ainsi pour le jeu 1:

$$x_1 = 8,4 \pm 0,05 \text{ cm.}$$

Pour le jeu 2:

$$x_2 = 11,2 \pm 0,05 \text{ cm}.$$



Nous acceptons une perte d'information due à la résolution $\delta = 1 \text{ mm}$ de la règle. Lorsque ultérieurement nous exploitons ces données toutes les valeurs entre $x_{\min} = x_{\text{moy}} - \delta/2$ et $x_{\max} = x_{\text{moy}} + \delta/2$ sont équiprobables. La loi de probabilité de la variable continue aléatoire X est uniforme. x est une réalisation de X . Cette distribution de probabilité a une étendue $E = x_{\max} - x_{\min}$ et nous vérifions pour la densité de probabilité $f(x)$:

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

La probabilité pour que la valeur de X soit comprise entre x et $x + dx$ est de $f(x) dx$. Le résultat sera compris avec certitude entre x_{\min} et x_{\max} : par exemple $x_1 = 8,4 \pm 0,05 \text{ cm}$ à 100% de confiance, mais $x_1 = 8,4 \pm 0,04 \text{ cm}$ avec une probabilité de 80%.

Pour caractériser l'étalement d'une distribution considérons l'étendue E et l'écart-type σ dont la définition pour une loi continue est:

$$V = \sigma^2 = \int (x - x_{moy})^2 f(x) dx ,$$

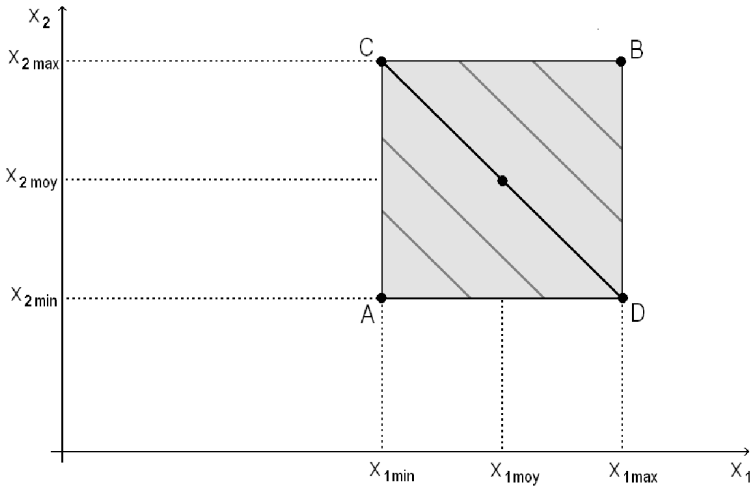
V est appelée la variance. Pour une loi uniforme:

$$\sigma = \delta / \sqrt{12} \approx 0,29 \delta ,$$

et nous avons $x = x_{moy} \pm \sigma$ avec une confiance de 58%. L'écart-type est une grandeur adéquate pour caractériser la largeur d'une distribution. L'étendue quant à elle est définie par les valeurs extrêmes qui peuvent être peu représentatives ou pire des valeurs aberrantes.

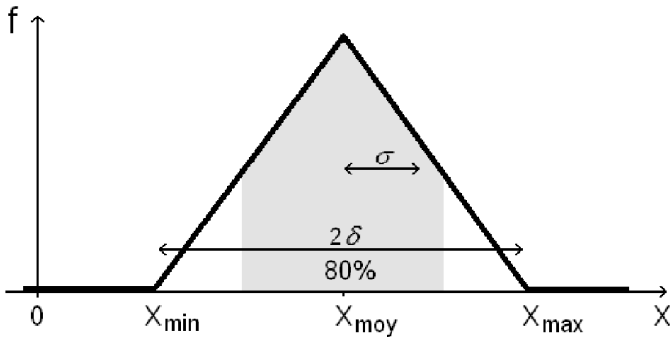
2. LONGUEUR DES DEUX CARTES MISES BOUT À BOUT

Nous souhaitons déterminer l'incertitude sur x avec $x = x_1 + x_2$. Si nous traçons x_2 en fonction de x_1 l'ensemble des points possibles forme un domaine carré. L'ensemble des points tel que x soit constant est une portion de droite de pente -1 et d'ordonnée à l'origine x : $x_2 = -x_1 + x$. Il n'y a qu'un cas qui réalise $x = x_{\min}$ soit $\left\{ x_1 = x_{1\min}; x_2 = x_{2\min} \right\}$ au point A sur la figure. Par contre sur l'ensemble du segment [CD] $x = x_{moy}$. Nous comprenons que toutes les valeurs de x ne sont pas équiprobables.



La loi de probabilité f de X se calcule à partir de celle f_1 de X_1 et f_2 de X_2 . Pour une somme de variables aléatoires indépendantes le résultat est donné par un produit de convolution [iii] :

$$f(x) = \int f_1(y) f_2(x-y) dy \Rightarrow \begin{cases} x < x_{\min} \Rightarrow f(x) = 0 \\ x_{\min} < x < x_{\text{moy}} \Rightarrow f(x) = \frac{(x - x_{\min})}{\delta^2} \\ x_{\text{moy}} < x < x_{\max} \Rightarrow f(x) = \frac{(x_{\max} - x)}{\delta^2} \\ x > x_{\max} \Rightarrow f(x) = 0 \end{cases}$$

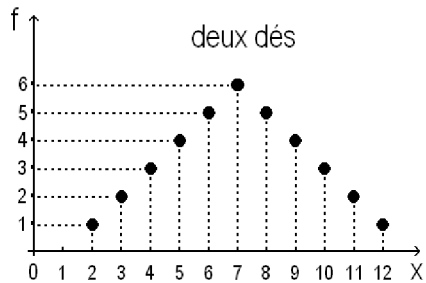
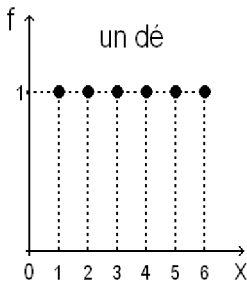


Nous avons alors une loi de probabilité triangulaire.

Nous obtenons $x = 19,6 \pm 0,1 \text{ cm}$ avec 100% de confiance, et $x = 19,6 \pm 0,055 \text{ cm}$ avec 80% de confiance.

3. ANALOGIE AVEC LE LANCER DE DEUX DÉS

Pour chaque dé les six valeurs sont équiprobables. Ici la loi de probabilité n'est plus continue mais discrète. Pour le lancer simultané de deux dés, la somme des valeurs obtenues est comprise entre deux et 12. Dans ce cas il n'y a plus d'équiprobabilité, une manière de faire deux avec un double un, deux manières de faire trois avec un et deux ou deux et un... Pour faire sept nous obtenons le maximum de possibilités. Nous retrouvons ainsi une loi triangulaire.



4. LONGUEUR DE DEUX CARTES D'UN MÊME JEU MISES BOUT À BOUT

Les cartes d'un jeu étant supposées identiques si la longueur de l'une d'elle est surestimée, il en sera de même pour la deuxième. Dans ce cas les erreurs s'ajoutent et ne peuvent pas se compenser. Pour deux cartes différentes, la première mesure pouvait être sous-estimée et la deuxième surestimée, une compensation pouvant alors se produire. Ici ce n'est plus le cas et pour $X = X_i + X_i'$ nous obtenons une loi de probabilité à nouveau uniforme de largeur 2δ . Nos variables aléatoires ne sont plus indépendantes.

Pour le jeu 1:

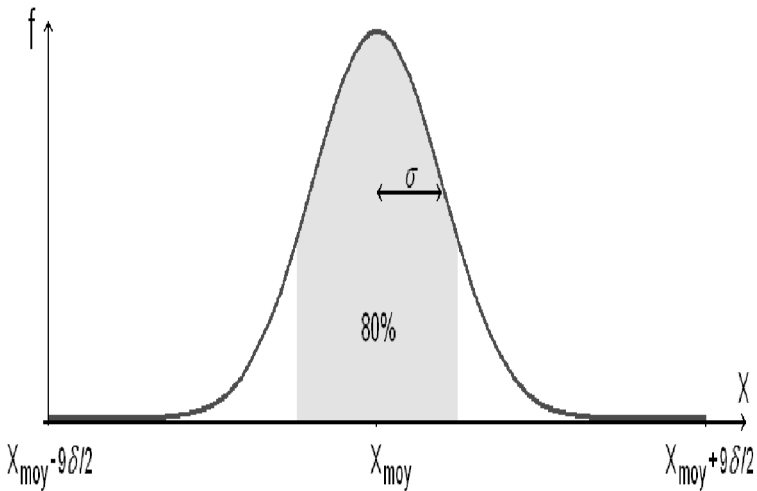
$$x_1 = 8,4 \pm 0,04 \text{ cm} \Rightarrow x = 2x_1 = 16,8 \pm 0,08 \text{ cm}$$

à 80% de confiance.

5. SOMME DE N LONGUEURS INDÉPENDANTES

Nous avons $X = \sum_{i=1}^N X_i$. Chaque longueur X_i

suit une loi uniforme de largeur δ . Pour la somme de neuf variables aléatoires indépendantes après itération du calcul nous obtenons la courbe suivante:



Nous avons dans ce cas $x = x_{moy} \pm 0,11 \text{ cm}$ à 80%. A 100% de confiance $x = x_{moy} \pm 0,45 \text{ cm}$ ce qui amène à considérer des domaines où la probabilité de présence de X est vraiment négligeable. Une incertitude de 0,45cm semble inutile alors que 99% des cas étaient déjà présents avec une incertitude de 0,22cm.

Raisonnement avec une confiance de 100% revient à considérer l'étendue, celle-ci est additive pour une somme de variables. L'étendue est proportionnelle à N.

	80%	95%	99%
N=1	0,40 δ	0,48 δ	0,50 δ
2	0,55 δ	0,78 δ	0,90 δ
3	0,66 δ	0,97 δ	1,19 δ
4	0,75 δ	1,12 δ	1,41 δ
5	0,84 δ	1,25 δ	1,60 δ
6	0,92 δ	1,38 δ	1,76 δ
7	0,99 δ	1,49 δ	1,91 δ
8	1,06 δ	1,59 δ	2,05 δ
9	1,12 δ	1,69 δ	2,18 δ
10	1,2 δ	1,8 δ	2,3 δ
20	1,7 δ	2,5 δ	3,3 δ
50	2,6 δ	4,0 δ	5,2 δ
100	3,7 δ	5,7 δ	7,4 δ

Mais cette approche ne tient pas compte d'une chose: la courbe se resserre autour de la moyenne quand N augmente. Il existe une autre grandeur additive: la variance. L'écart-type racine de la variance est proportionnel à \sqrt{N} et tient compte des compensations d'erreurs.

La courbe obtenue est ce qu'on appelle une courbe en cloche. Un théorème statistique, appelé théorème central limite, indique que pour N grand la courbe tend vers une gaussienne. L'étendue d'une gaussienne est infinie pour un écart-type fini.

Nous pouvons résumer l'évolution de l'incertitude sur la somme de N longueurs indépendantes mesurées avec une même résolution δ dans un tableau. En

italique, à partir de $N=10$, il s'agit de simulations numériques réalisées sur ordinateur par génération de nombres aléatoires.

Les résultats des mesures sont souvent donnés avec une confiance de 95%, ce qui correspond pour une gaussienne à une incertitude d'environ 2σ .

6. AUTRES APPLICATIONS

Un coureur souhaite mesurer son temps de parcours. Sa montre à affichage numérique indique qu'il part à $10\text{ h }52\text{ min}$ et qu'il arrive à $11\text{ h }11\text{ min}$. L'affichage est à la minute, il est donc parti entre $10\text{ h }52\text{ min }00\text{ s}$ et $10\text{ h }52\text{ min }59\text{ s}$. D'où la date de départ dans l'intervalle $t_1 = 10\text{ h }52\text{ min }30\text{ s} \pm 30\text{ s}$. La résolution est d'une minute. La durée du parcours est $\Delta t = t_2 - t_1$. Les résultats restent vrais pour une différence. Nous avons $N=2$ et $\Delta t = 19\text{ min} \pm 47\text{ s}$ avec 95% de confiance.

Même démarche si des étudiants mesurent une différence d'angles sur un goniomètre. Chaque mesure étant à la minute d'arc près l'incertitude du résultat est de 47 secondes d'arc à 95%.

Sept personnes veulent rentrer en même temps dans un ascenseur. Sa charge maximale est de 500kg. Leurs masses individuelles sont mesurées avec un pèse personne d'une résolution de un kilogramme. La masse totale est de 499kg. Quelle est la probabilité d'être en surcharge?

Pour $N=7$ l'incertitude atteint un kilogramme avec une confiance de 80%. Il y a donc une chance sur dix pour que l'ascenseur soit en surcharge.

Au laboratoire de nombreux appareils de mesure disposent d'affichages numériques. La résolution est au dernier digit près. Mais l'incertitude globale est bien supérieure. Il faut consulter la notice de chaque appareil.

CONCLUSION

La démarche générale consiste à combiner des lois de probabilités. L'outil mathématique utilisé est un changement de variables, puis une ou plusieurs intégrations. Pour la mesure avec une règle décrite dans cet article, il s'agissait d'une somme de deux variables aléatoires indépendantes et nous avons obtenu un produit de convolution.

Si l'on veut faire un calcul plus rapide une analyse de variance peut suffire. Nous avons une variable aléatoire X qui dépend de N variables aléatoires indépendantes X_i : $X = f(X_1, X_2, \dots, X_i, \dots, X_N)$. Nous appelons σ_i l'écart-type de X_i et σ celui de X . Pour des σ_i finis et de petites variations, nous avons la formule de propagation des écart-types [iv] :

$$\sigma^2 = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_i^2$$

Et, indépendamment des lois de probabilités, cette relation entre les variances reste vraie. On pourra ainsi donner son résultat avec une incertitude à 2σ ou 3σ .

Existe-t-il une formule analogue en terme de confiance? Oui, mais elle est approximative, c'est la formule de propagation des incertitudes:

$$\Delta f^2 = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)^2 \Delta x_i^2 ,$$

avec $x_i = x_{i\text{moy}} \pm \Delta x_i$, $f = f_{\text{moy}} \pm \Delta f$ et une confiance constante. Cette formule est très pratique et permet un calcul rapide et raisonnable des incertitudes combinées. Qui plus est, elle est exacte si la forme des distributions est la même pour X et les X_i . Par exemple si les X_i sont à distribution gaussienne toute combinaison linéaire l'est aussi.

Nous tenons ainsi compte des compensations et nous évitons d'utiliser la formule $\Delta f = \sum_{i=1}^n |\partial f / \partial x_i| \Delta x_i$ qui surestime les incertitudes, parfois même avec un tel excès que l'on en perd son sens physique. Cette dernière formule ne tient compte d'aucunes compensations, on a la pire des situations, statistiquement improbable. Ici, par exemple pour $N = 100$, on aurait une incertitude de 50δ , au lieu de $5,7 \delta$ dans la pratique (confiance de 95%).

Dans cet article nous nous sommes concentré sur la résolution d'un système d'acquisition qui donne une erreur de discrétisation. Mais on peut aussi être amené à considérer des erreurs systématiques et des erreurs aléatoires. Ici la règle était supposée parfaite, c'est à dire juste et fidèle.

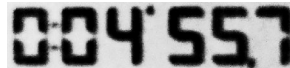
B. Métrologie

1. Exemples

Intéressons-nous à deux instruments de mesure particuliers : un chronomètre et un ampèremètre.

Mesure d'une durée avec un chronomètre :

Nous mesurons la durée d d'un phénomène. Nous supposons que l'expérimentateur appuie aux bons moments pour mesurer ce que l'on appelle les dates initiale t_1 et finale t_2 . Pour être un bon expérimentateur il faut avoir une connaissance minimale de ses instruments de mesure. Un chronomètre, du type utilisé ici, fonctionne avec un quartz très précis. La dérive est d'une quinzaine de secondes par an, soit environ 0,00004 secondes sur notre expérience. Ici, nous négligeons la dérive. Nous pouvons supposer le chronomètre parfait (juste et fidèle). Seule la résolution va intervenir ($\delta = 1/10$ éme de seconde).



$$3\text{min}32,75\text{s} < t_1 < 3\text{min}32,85\text{s}$$

$$4\text{min}55,65\text{s} < t_2 < 4\text{min}55,75\text{s}$$

$$\text{et } d = t_2 - t_1$$

Les lois de probabilité de t_1 et t_2 sont uniformes et celle de d triangulaire (voir article précédent).

D'où⁹ $\sigma_d = \delta/\sqrt{6} \approx 0,41\delta$, $\Delta d \approx 0,78 \delta$ (tableau page 100),

9 Calculons l'écart-type d'une distribution triangulaire (après recentrage) : $\sigma^2 = \int_{-\delta}^0 x^2 \frac{(x+\delta)}{\delta^2} dx + \int_0^{\delta} x^2 \frac{(\delta-x)}{\delta^2} dx = \frac{\delta^2}{6}$. Ou, si nous utilisons la formule de propagation des écart-types on trouve

$d = 82,900 \pm 0,078$ s et $\Delta d/d \approx 0,1\%$ avec 95% de confiance.

Maintenant, si le phénomène durait plusieurs mois, l'incertitude la plus grande serait la dérive, le chronomètre ne serait plus juste. Nous voyons que l'ensemble de la situation expérimentale est à prendre en compte et pas seulement l'instrument de mesure.

Nous pourrions incorporer cette incertitude systématique de dérive par la connaissance de la valeur du biais.

Si la dérive n'est pas connue, nous pouvons prendre 100 chronomètres identiques que nous lançons simultanément. Nous aurons au final une dispersion aléatoire des valeurs qui donne une incertitude liée à la dérive.

Mesure d'un courant avec un ampèremètre :

L'ampèremètre, placé en série dans un circuit, mesure le courant de la branche où il est situé.



Nous choisissons le calibre adapté, ici pour des courants de 20 à 200 mA, et nous mesurons : $I = 186,30$ mA.

Quelle est l'incertitude sur I ?

Nous pourrions penser, comme pour le chronomètre, qu'elle est sur le dernier digit, en fait la notice indique une plus grande

que pour une somme, ou différence, de n grandeurs de même écart-types nous avons $\sigma_n = \sqrt{n} \sigma_1$. Ici $\sigma_2 = \sqrt{2} \sigma_1 = \sqrt{2} \delta / \sqrt{12} = \delta / \sqrt{6}$.

incertitude :

MX579

COURANTS CONTINUS:

Cali- bre	Résolu- tion	Précision $\pm(n\%L + mUR^*)$	Coeff. de Ture.	Chute de tension	Protec- tion
200 μ A	10nA	0,2 % + 3 UR	2E-4/K	<300mV	fusible 2A
2mA	100nA	"	"	"	
20mA	1 μ A	"	"	"	
200mA	10 μ A	0,7 % + 3 UR	7E-4/K	"	10A
2A	100 μ A	"	"	<700mV	
10A	1mA	"	"	"	

D'où $\Delta I = 0,7\% \times 186,30\text{mA} + 3 \times 10\mu\text{A}$.

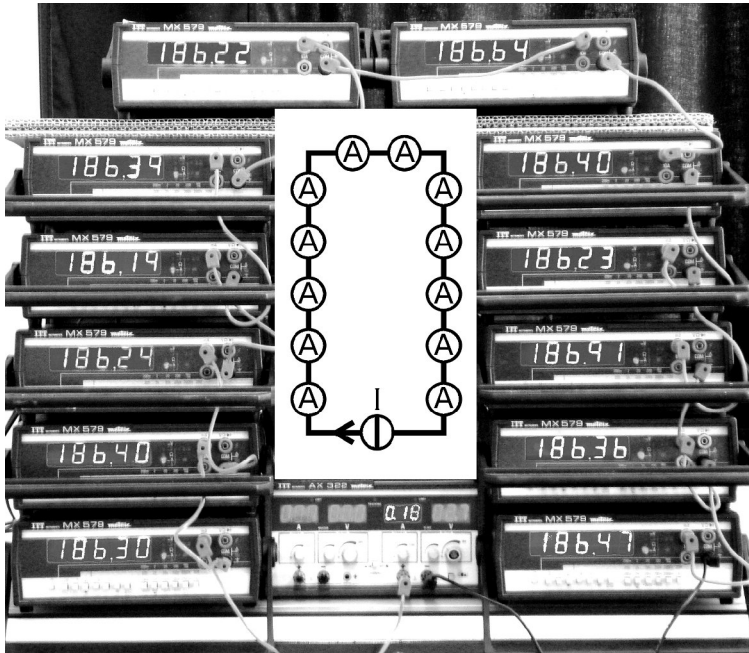
Pour simplifier nous pouvons considérer que les indications des fabricants sont données avec **une confiance de 95% et un profil gaussien.**

Alors $I = 186,30 \pm 1,34 \text{ mA}$

avec un niveau de confiance de 95%

(184,96 mA < I < 187,64 mA)

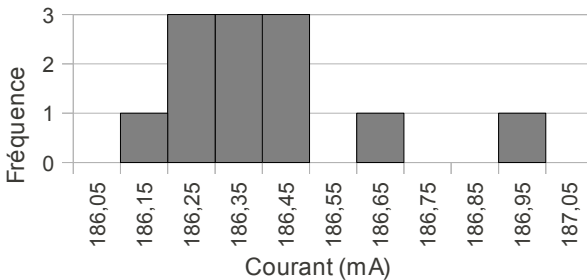
Essayons de retrouver l'erreur systématique donnée par le fabricant en regardant l'erreur aléatoire donnée par 12 ampèremètres de même modèle. Nous les plaçons en série et ainsi ils mesurent tous la même grandeur :



Le calcul de l'écart-type donne : $s_I = 0,21 \text{ mA}$, le coefficient de Student vaut $t = 2,2$, d'où $\Delta I = t \cdot s = 0,45 \text{ mA}$.

Soit $I = 186,40 \pm 0,45 \text{ mA}$ avec un niveau de confiance de 95% ($185,94 \text{ mA} < I < 186,85 \text{ mA}$).

L'incertitude est ici, à peu près, 3 fois plus faible, ce qui est normal car nous avons une meilleure connaissance de la moyenne ($\Delta \bar{I} = t \cdot s / \sqrt{n} = 0,13$



I (mA)
186,30
186,40
186,24
186,19
186,39
186,22
186,64
186,40
186,23
186,91
186,36
186,47

mA). Si nous considérons l'ampèremètre pour lequel $I=186,91\text{mA}$, seule l'incertitude systématique plus grande fournie par le constructeur garantie un bon encadrement.

Par ailleurs l'incertitude du constructeur concerne l'ensemble des ampèremètres de ce modèle qu'il produit et pas seulement ceux livrés au lycée. Aussi, nous devons considérer, pour la comparaison des résultats, que les 12 ampèremètres n'ont pas été recalibrés avec un étalon depuis leur achat d'il y a plus de 15 ans.

2. Normes

Pour déterminer rigoureusement l'incertitude, il faudrait que tous les fabricants donnent la confiance de leurs incertitudes et les profils des lois de probabilité.

Pour contribuer à une information complète sur l'expression de l'incertitude et fournir une base pour la comparaison internationale des résultats de mesure *Le comité international des poids et mesures* et *l'Organisation internationale de normalisation* ont développés un guide [vii].

Il s'agit de mettre en place des méthodes d'évaluation et d'expression des incertitudes harmonisées pour tous les pays et tous les domaines (de la santé, de la sécurité, industriel, commercial, etc.).

3. Vocabulaire

Que l'on parle avec le vocable de la statistique ou avec celui de la métrologie les notions sont les mêmes. Pour aider à faire la jonction, nous avons synthétisé certaines expressions courantes dans un tableau. Sur une même ligne vous trouverez les termes équivalents.

statistique [vi] [iv]		métrologie [v] [vii] [viii]		
grandeur à mesurer X		mesurande M		
mesure	action de mesurer		mesurage	
ensemble des mesures $\{x_i \pm \Delta x_i\}$		résultat du mesurage		
x_i	une mesure		m_i	
$X = \bar{x} \pm \Delta x$	résultat d'une mesure		$M = m \pm \Delta M$	
dispersion	fidélité (répétabilité et reproductibilité)	erreurs aléatoires (accidentelles)	erreurs de type A	
justesse	biais		erreurs systématiques	erreurs de type B
résolution (sensibilité)	(erreur de) discrétisation			
μ (espérance)	moyenne de la population (souvent inconnue)		la valeur vraie M_{vrai}	
\bar{x}	moyenne de l'échantillon		\bar{m} ou m	
σ : écart-type de la population (souvent inconnu)				
s	s_x	écart-type de l'échantillon	écart-type expérimental	s_{exp}
s / \sqrt{n}	$s_{\bar{x}}$	écart-type estimé de la distribution d'échantillonnage	incertitude-type	$s = s_{\text{exp}} / \sqrt{n}$
coefficient de Student t $\Delta x = t \cdot s / \sqrt{n}$		facteur d'élargissement k $\Delta M = k \cdot s$		
Δx	incertitude (incertitude absolue)		incertitude-type élargie	ΔM
$\Delta x / \bar{x}$	précision (incertitude relative)		$\Delta M / m$	

Remarques :

- Plus le biais est grand, moins la mesure est juste.
- Plus les mesures sont dispersées, moins la fidélité est bonne.
- L'écart-type de l'échantillon est un estimateur biaisé de l'écart-type de la population.
- Dans ce livre nous avons évité de parler d'*erreur*, nous avons préféré le terme *incertitude*. En métrologie, nous parlons de différentes erreurs (erreur de mesure, erreur aléatoire et erreur systématique).
- Nous utilisons le terme écart pour la différence entre la moyenne de la mesure et la valeur attendue (une valeur donnée par le fabricant ou dans les tables).

4. Méthode simplifiée d'évaluation de l'incertitude de type B

Si le fabricant donne des limites et que l'on ne dispose d'aucune autre information, il est raisonnable de supposer la probabilité égale sur l'intervalle et nulle en dehors. Nous utilisons alors les caractéristiques d'une loi uniforme ([vii] p23 et [v] p7) :

Cas :	Incertitude-type :	Incertitude élargie :
Appareil à cadran, lecture d'un réglelet, digit, etc. Résolution : $\delta = x_{max} - x_{min}$, $x = \bar{x} \pm \delta/2$	$s_{\bar{x}} = \frac{\delta}{\sqrt{12}}$	$\Delta x = k s_{\bar{x}}$
Le constructeur fournit la classe : $\pm a$ a correspond à $\delta/2$	$s_{\bar{x}} = \frac{a}{\sqrt{3}}$	($k=2$ à 95%)

Remarques :

- Pour le premier cas nous ne disposons pas d'information du constructeur, dans le second, il fournit ce que l'on appelle la classe a de l'appareil.
- Nous prenons $k=2$ avec une confiance de 95% comme si nous avions une gaussienne (approximation).
- Si la notice est conçue selon cette norme, nous aurons pour l'ampèremètre (étudié page 106), $s_I = 1,34 \text{ mA}$ et $I = 186,30 \pm 2,68 \text{ mA}$ à 95%.

- Sur le matériel de verrerie jaugé, la classe est toujours indiquée. **25ml**
 Par exemple, sur une pipette jaugée de 25 mL, nous avons les indications suivantes : **Ex 20°C**
 $\pm 0,06 \text{ mL}$
 La classe est 0,06 mL, d'où $s_V = 0,035 \text{ mL}$ et $V = 25 \pm 0,07 \text{ mL}$ avec une confiance de 95%. La précision sur le volume est de 0,28%.

- Pour une résistance R de 1000Ω et d'une précision de 5% (brun, noir, rouge, or), nous avons :
 $\Delta R = 2 \cdot 1000 \cdot (5/100) / \sqrt{3} \approx 58\Omega$ à 95% de confiance.
 $R = 1000 \pm 58\Omega$ et $\Delta R/R \approx 5,8 \%$.

- Pesée : une balance numérique, de résolution 0,1g, indique une masse m de 73,7g.
 $\Delta m = 2 \cdot 0,1 / \sqrt{12} \approx 0,06 \text{ g}$, $m = 73,70 \pm 0,06 \text{ g}$ à 95%. $\Delta m/m \approx 0,08 \%$

- Par la méthode d'autocollimation, nous mesurons la distance focale f' d'une lentille convergente mince. L'image est nette (latitude de mise au point) de $f'_{\min} = 195 \text{ mm}$ à $f'_{\max} = 203 \text{ mm}$. D'où $s_{\text{optique}} = 2,31 \text{ mm}$.
 $s_{\text{géométrique}} = 0,29 \text{ mm}$ (δ de 1mm du banc d'optique) et $s_{\text{modélisation}} = 0,58 \text{ mm}$ (épaisseur de 2mm de la lentille).

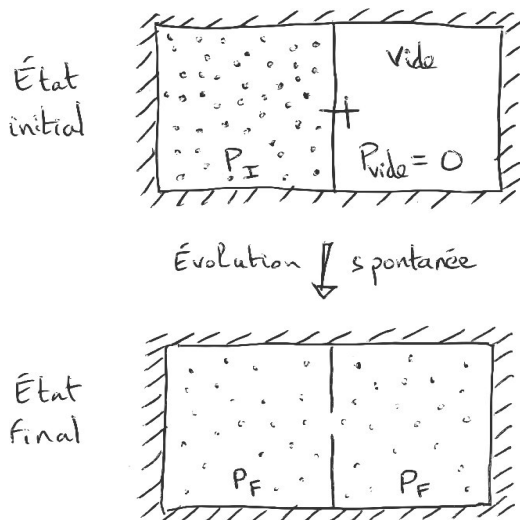
$$s = \sqrt{s_o^2 + s_g^2 + s_m^2} \quad f' = 199,0 \pm 4,8 \text{ mm} (95\%)$$

C. Thermodynamique

Nous allons étudier la détente de Joule et Gay-Lussac. Pour cette détente, nous allons utiliser deux méthodes pour déterminer les fluctuations de la pression dans l'état final : l'une classique en thermodynamique, l'autre en utilisant la formule de propagation. Dans ce cadre, nous montrerons que cette formule est en accord avec les lois de la physique.

1. Description de la détente de Joule Gay-Lussac

Nous avons deux enceintes, aux parois indéformables et calorifugées, pouvant communiquer au moyen d'un robinet. Initialement un fluide est enfermé dans l'un des compartiments, l'autre compartiment est vide. Nous ouvrons le robinet et une détente dans le vide se produit. La pression s'égalise.



Nous considérerons le cas où les deux enceintes ont le même volume.

De plus, le fluide étudié sera un gaz parfait : le premier principe de la thermodynamique permet alors de montrer que la température finale est égale à la température initiale : $T_F=T_I$; et pour les volumes $V_F=2V_I$.

2. Hypothèses du gaz parfait

Les molécules sont considérées ponctuelles, identiques et sans interactions entre elles. Elles évoluent donc, de manière tout à fait indépendante, elles s'ignorent mutuellement, et les seuls contacts des particules sont avec les parois des récipients. Leurs chocs sur les parois exercent une pression, celle-ci dépend de la densité des particules et de la température :

$$PV = nRT \Rightarrow P = n^* k_B T$$

$$\text{avec } N = n \cdot N_A \text{ et } n^* = N/V$$

n : nombre de moles de gaz dans la boîte

N : nombre de molécules de gaz dans la boîte

n^* : densité de particules

(nombre de particules par unité de volume)

$$k_B = R/N_A \text{ : constante de Boltzmann}$$

R : constante des gaz parfaits N_A : nombre d'Avogadro

Nous avons, ici, les mêmes températures initiale et finale, ainsi, $P_F=P_I/2$, et la pression est directement proportionnelle à la densité de particules :

$$P \propto n^*$$

3. Hypothèses de la physique statistique

Nous distinguons l'état macroscopique du gaz, appelé macroétat (décrit par les variables d'états de la thermodynamique, telles la pression et la température) de tous les états microscopiques correspondants, appelés microétats (décrit par les positions et les vitesses individuelles de toutes les molécules du gaz). Un expérimentateur (macroscopique) qui observe la détente ne peut mesurer que le nombre total de particules (et donc la pression) dans chaque compartiment. Il n'a pas accès à l'ensemble des informations microscopiques du système : par exemple, il ne sait pas si une particule donnée A est dans le compartiment de gauche ou de droite, il peut seulement en connaître la probabilité.

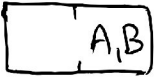
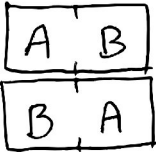
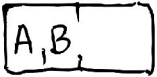
Nous noterons k le nombre total de molécules dans le compartiment de gauche.

Imaginons notre gaz constitué de cette seule particule, deux possibilités :

$$k=0 \left[\begin{array}{|c|} \hline A \\ \hline \end{array} \right] p=1/2 \quad \text{ou} \quad k=1 \left[\begin{array}{|c|} \hline A \\ \hline \end{array} \right] p=1/2$$


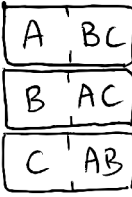


Les deux microétats possibles ont la même probabilité p , et correspondent chacun à un macroétat.

Notre gaz a maintenant deux particules, $N=2$:

macroétats	$k=0$	$k=1$	$k=2$
microétats			
probabilités	$p=1/4$	$p=1/2$	$p=1/4$

Le macroétat $k=1$ est représenté par deux microétats.

Regardons aussi le cas $N=3$:

macroétats	$k=0$	$k=1$	$k=2$	$k=3$
microétats				
probas	$p=1/8$	$p=3/8$	$p=3/8$	$p=1/8$

Allons encore plus loin pour $N=4$:

macroétats	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
microétats	(,ABCD)	(A,BCD) (B,ACD) (C,ABD) (D,ABC)	(AB,CD) (AC,BD) (AD,BC) (BC,AD) (BD,AC) (CD,AB)	(BCD,A) (ACD,B) (ABD,C) (ABC,D)	(ABCD,)
probas	$p=1/16$	$p=1/4$	$p=3/8$	$p=1/4$	$p=1/16$

Nous voyons que, plus le nombre de particules est grand, plus les macroétats autour de $k \simeq N/2$ sont probables.

Nous avons deux principes physiques :

- les microétats sont équiprobables.
- hypothèse ergodique : au cours du temps tous les microétats sont indifféremment balayés et le système reste un temps équivalent dans chacun deux.

4. État le plus probable et fluctuations

Nous comprenons maintenant que le système sera, la plupart du temps, dans le macroétat $k=N/2$ (prenant N pair, ce qui ne restreint pas notre étude car N est très grand et nous ne sommes donc pas à une particule près).

Nombre de microétats Ω et probabilité p

pour un macroétat k donné :

$$\Omega = C_N^k = \frac{N!}{k!(N-k)!} \quad \text{et} \quad p(k) = \frac{C_N^k}{2^N}$$

Finalement, cela va se passer autour de $N/2$, alors symétrisons, et posons $k=N/2+i$, d'où : $\Omega = N!/(N/2+i)!(N/2-i)!$

Nous cherchons l'expression de $p(i)$ pour N grand et i petit, utilisons la formule de Stirling $\ln(n!) \simeq n \ln(n) - n$ avec n grand et le développement limité $\ln(1+\varepsilon) \simeq \varepsilon$ avec ε petit ; nous avons alors : $\ln[p(i)] \simeq -4i^2/N$ et $p(i) \propto e^{-4i^2/N}$.

La probabilité obtenue obéit à une loi gaussienne.

Déterminons l'écart-type : $\frac{p(\pm\sigma)}{p_{max}} = \frac{1}{\sqrt{e}}$ d'où $\sigma \propto \frac{\sqrt{N}}{2}$.

Nous avons l'incertitude relative à un sigma pour le nombre de particules et donc aussi pour la pression :

$$\frac{\sigma_{N/2}}{N/2} = \frac{1}{\sqrt{N}} \quad \text{et} \quad \frac{\sigma_P}{P} = \frac{1}{\sqrt{N}}$$

N est grand, ainsi la pression fluctue très peu et est parfaitement définie.

5. Utilisation de la formule de propagation

Nous avons N particules de pression p_j qui contribuent, ou pas, à la pression du compartiment de droite : $P = \sum_{j=1}^N p_j$.

Si la particule est dans le compartiment de gauche $p_j=0$, si elle est à droite $p_j=P_F/(N/2)$. Donc $\sigma_p = P_F/N$ (même principe que les pièces page 8). En utilisant la formule de propagation des écart-types nous obtenons : $\sigma_P = \sqrt{N} \sigma_p$

$$\text{et } \frac{\sigma_P}{P} = \frac{1}{\sqrt{N}}$$

6. Conclusion

La formule de propagation permet de retrouver rapidement le résultat de la physique statistique.

Si nous faisons l'erreur de considérer que les incertitudes s'ajoutent nous aurions $P_F = P_F \pm P_F$. Ainsi la pression et toutes les variables d'état de la thermodynamique ne seraient pas définies. De plus, cette détente, démontrée irréversible en vertu du second principe, deviendrait réversible.

Nous voyons à nouveau, sur cet exemple, que la formule de propagation n'est pas seulement un outil rigoureux pour le calcul des incertitudes mais participe, aussi, aux fondements des lois de la nature.

D. Indépendance des variables

Considérons un système. Nous mesurons deux grandeurs associées à celui-ci. Normalement les deux processus de mesure s'effectuent indépendamment. L'action de mesurer la première grandeur fournit un ensemble d'observations relatives à cette grandeur. Ces observations ont un centre et une dispersion. Dans un second temps, nous mesurons le deuxième grandeur et il n'y a aucune raison que cette deuxième collecte d'informations soit perturbée par la première. Les fluctuations de la première ne sont pas corrélées avec celles de la deuxième.

En effet, la plupart du temps, le mesurage ne modifie pas le système et les différents processus de mesure n'interagissent pas entre eux.

Exemple : nous avons un système électrique en régime stationnaire constitué d'un générateur de tension continue U qui alimente une résistance R parcourue par un courant I . Nous avons la relation théorique $U=RI$ et nous souhaitons estimer R .

Nous effectuons 100 observations de la tension U (nous pouvons prendre 100 voltmètres différents) et nous en déduisons \bar{U} et ΔU . Dans un second temps, nous procédons de même pour I , d'où \bar{I} et ΔI . Nécessairement, le coefficient de corrélation r_{UI} entre $\{U_1, \dots, U_{100}\}$ et $\{I_1, \dots, I_{100}\}$ est nul. Les grandeurs U et I sont indépendantes. D'où $\bar{R}=\bar{U}/\bar{I}$ et $\Delta R/R=\sqrt{(\Delta U/U^2+\Delta I/I^2)}$.

Deuxième exemple : nous lançons quatre dés équilibrés six faces de couleurs différentes. Nous calculons la somme :

$S=d_1+d_2+d_3+d_4$. Si les lancers des trois premiers dés donnent à chaque fois 6 cela ne préjuge en rien la probabilité d'obtenir 6 avec le quatrième dé, c'est toujours une chance sur six. Nos quatre grandeurs sont indépendantes. Conclusion : quelque soit $i,j=1,2,3,4$ avec $i \neq j$ $r_{ij}=0$ et $\sigma_S^2=\sigma_1^2+\sigma_2^2+\sigma_3^2+\sigma_4^2$.

Il existe une généralisation de la formule de propagation des écart-types dans le cas de grandeurs corrélées mais, dans la plupart des cas, nous n'avons pas à l'utiliser :

$$\sigma_f^2 = \sum_j \left(\frac{\partial f}{\partial x_j} \right)^2 \sigma_j^2 + 2 \sum_{i < j} \left(\frac{\partial f}{\partial x_i} \right) \left(\frac{\partial f}{\partial x_j} \right) r_{ij} \sigma_i \sigma_j$$

Remarques :

- $\sigma_f \leq \sum |\partial f / \partial x_i| \sigma_i$.
- si $f=x_1+x_2$, $\sigma_1=\sigma_2$ et $r_{12}=-1$ alors $\sigma_f=0$.

V. DEVOIRS

Durée de composition de deux heures pour chaque devoir. Calculatrice autorisée. Documents, ordinateurs et téléphones interdits. Toutes les réponses doivent être justifiées et clairement explicitées. Les tables de Student et de khi deux sont fournies, ainsi qu'un formulaire avec la formule de l'écart-type et celle du coefficient de corrélation.

A. Devoir Suricate

Exercice 1 (3 pts) :

Nous lançons quatre pièces équilibrées.

Quelle est la probabilité d'en avoir quatre qui tombent du côté pile?

Même question pour avoir trois faces et un pile suite au lancer des quatre pièces.

Nous allons maintenant nous intéresser à la composition de différentes eaux minérales :

	Calcium mg/L	Bicarbonates mg/L	Magnésium mg/L	Chlorures mg/L	Sulfates mg/L	Potassium mg/L	Sodium mg/L	pH
Evian	78	357	24	4,5	10	1	5	7,2
Badoit	190	1300	85	44	38	10	165	6
Contrex	486	403	84	10	1187	3,2	9,1	7,2
Hépar	555	403	110	11	1479	4	14	7,2
St-Yorre	90	4368	11	322	174	132	1708	6,6
Vittel	202	402	43	8	336	2	4,7	7,5
Volvic	11,5	71	8	13,5	8,1	6,2	11,6	6,95
Vichy célestin	103	2989	10	235	138	66	1172	6,8
Salvetat	253	820	11	4	25	3	7	6

Exercice 2 (7 pts) :

Nous allons tout d'abord considérer la quantité B de bicarbonates en mg/L :

- Déterminez la valeur moyenne \bar{B} de B.
- Calculez l'écart-type σ_B sur B.
- Avec une confiance de 95%,
quelle est l'incertitude ΔB ?

Nous considérons maintenant la concentration S en sulfates.

- Les grandeurs B et S sont-elles corrélées ? (si la valeur absolue du coefficient de corrélation r est inférieure à la valeur critique r_c ici égale à 0,52 les grandeurs peuvent être considérées comme indépendantes).
- Déterminez la teneur moyenne totale $\{B \oplus S\}$ en mg/L des bicarbonates et sulfates avec son incertitude.

Exercice 3 (3 pts) :

Nous considérons une population d'un pays constituée de pauvres et de non-pauvres. Les pays est découpé en quatre régions. Dans chaque région nous tirons 1000 personnes au hasard et nous obtenons la répartition suivante :

Zone géographique :	Nord - Est	Centre - Nord	Sud	Ouest
Pauvres :	112	105	154	113
Non pauvres :	888	895	846	887

Nous faisons l'hypothèse que la pauvreté est uniformément répartie sur tout le pays.

L'hypothèse est-elle soutenable avec une probabilité critique de 5% ?

Exercice 4 (3 pts) :

Nous traçons maintenant pour les différentes eaux minérales la quantité de sodium en fonction de celle de chlorure. Une régression nous donne la pente et l'ordonnée à l'origine avec leurs incertitudes à 95% de confiance. Pour l'ordonnée à l'origine nous trouvons : $b = -43 \pm 25$ mg/L. Nous vous proposons des possibilités pour la pente dont une est la bonne, expliquez votre choix et pourquoi les autres ne peuvent convenir.

Pour ce faire vous devez vous aider d'une matrice de l'exercice 5. D'autres données qui peuvent vous aider : $M_{\text{Na}}=23,0\text{g/mol}$, $M_{\text{Cl}}=35,5\text{g/mol}$. Vous pourrez conclure sur l'information que l'on en tire sur les eaux minérales en général.

Cas :	1)	2)	3)	4)	5)	6)	7)	8)
pente a :	-5,34	5,34	-0,65	0,65	5,34	-5,34	0,65	-0,65
Δa :	1,19	0,19	0,19	1,19	1,19	0,19	0,19	1,19

Exercice 5 (8 pts dont 4 en bonus) :

Nous voulons réaliser une analyse en composantes principales avec les eaux minérales mentionnées sur la première page du devoir. Montrez sur l'exemple de x_7^2 comment on passe de la matrice de données \mathbf{M} à la matrice de données réduite \mathbf{m} suivante :

-0,753	-0,597	-0,475	-0,569	-0,660	-0,539	-0,532	0,693
-0,154	0,044	1,059	-0,238	-0,609	-0,339	-0,281	-1,542
1,430	-0,566	1,034	-0,523	1,454	-0,490	-0,525	0,693
1,799	-0,566	1,687	-0,514	1,979	-0,472	-0,518	0,693
-0,689	2,132	-0,802	2,089	-0,365	2,371	2,139	-0,424
-0,089	-0,567	0,003	-0,540	-0,074	-0,517	-0,532	1,252
-1,109	-0,792	-0,877	-0,494	-0,663	-0,424	-0,521	0,228
-0,619	1,194	-0,827	1,361	-0,430	0,905	1,298	-0,052
0,183	-0,282	-0,802	-0,573	-0,633	-0,495	-0,529	-1,542

Quel est l'intérêt d'une telle transformation ?

Nous construisons ensuite la matrice des corrélations \mathbf{r} .
Complétez les valeurs manquantes :

?	-0,339	0,825	-0,375	0,922	-0,358	-0,373	0,204
?	?	-0,373	0,982	?	0,969	0,984	-0,336
?	?	?	-0,413	0,781	-0,411	-0,421	0,168
?	?	?	?	-0,234	0,980	0,999	-0,201
?	?	?	?	?	-0,213	-0,236	0,471
?	?	?	?	?	?	0,986	-0,194
?	?	?	?	?	?	?	-0,200
?	?	?	?	?	?	?	?

Nous diagonalisons ensuite la matrice \mathbf{r} et nous donnons les quatre premières valeurs propres :

$$\lambda_1 = 4,76; \lambda_2 = 2,05; \lambda_3 = 0,95 \text{ et } \lambda_4 = 0,20.$$

Que pensez-vous des autres valeurs propres en terme d'information ?

Expliquez le principe qui permet d'obtenir les valeurs propres.

Nous obtenons ensuite la matrice de passage \mathbf{P} (en dessous), les colonnes sont les vecteurs propres en fonction des vecteurs de la base initiale.

-0,305	0,472	-0,223	0,481	-0,457	-0,283	-0,278	0,192
0,422	0,258	-0,108	-0,063	-0,442	-0,173	0,502	-0,512
-0,312	0,402	-0,286	-0,806	0,006	-0,050	-0,051	0,068
0,422	0,265	0,050	-0,059	-0,104	0,493	-0,657	-0,244
-0,267	0,553	0,089	0,289	0,492	0,366	0,316	-0,236
0,417	0,272	0,055	0,015	0,531	-0,652	-0,203	0,030
0,423	0,264	0,053	-0,022	-0,085	0,256	0,307	0,762
-0,169	0,171	0,917	-0,166	-0,231	-0,143	-0,025	-0,002

Expliquez le principe qui permet d'obtenir la matrice de passage.

Comment à partir de cette matrice déterminons-nous la nouvelle matrice \mathbf{m}' des individus avec les nouvelles caractéristiques ?

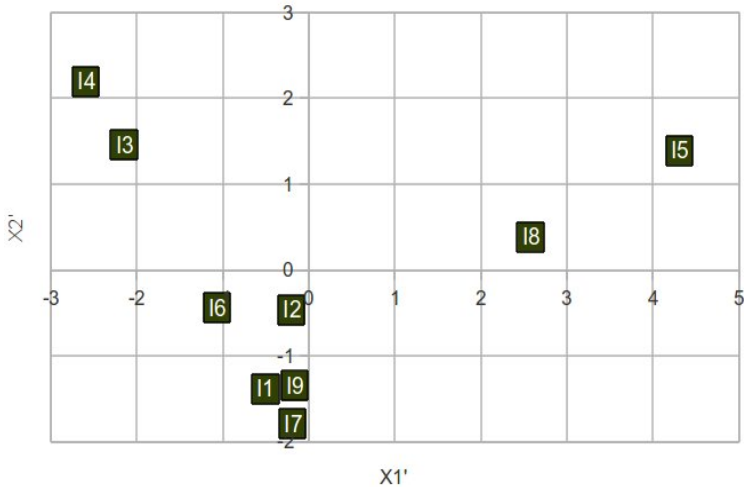
Illustrez en calculant l'élément $x'_2{}^1$.

Nous avons restitué les trois premières colonnes de \mathbf{m}' :

-0,50	-1,38	0,86	...
-0,20	-0,47	-1,79	...
-2,15	1,46	0,13	...
-2,59	2,19	-0,09	...
4,31	1,39	0,08	...
-1,07	-0,44	1,14	...
-0,19	-1,79	0,66	...
2,58	0,39	0,35	...
-0,17	-1,35	-1,33	...

Finalement nous obtenons la distribution des individus :

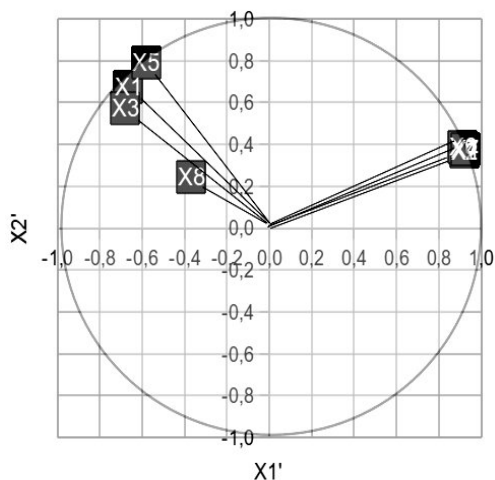
Eaux selon les deux composantes principales



Quel sens donneriez-vous à X_1' et X_2' ?

Vous pouvez aussi vous aider des corrélations entre nouvelles et anciennes variables, et du cercle des corrélations.

	X_1'	X_2'	X_3'
X_1	-0,666	0,675	-0,218 ...
X_2	0,920	0,369	-0,105 ...
X_3	-0,680	0,575	-0,279 ...
X_4	0,921	0,379	0,049 ...
X_5	-0,583	0,791	0,087 ...
X_6	0,910	0,389	0,054 ...
X_7	0,923	0,378	0,051 ...
X_8	-0,368	0,245	0,893 ...



Quels enseignements tirez-vous concernant les différents types d'eaux ?

Apparaît-il une classification particulière ? Par exemple selon la minéralisation, eau gazeuse, etc.

Que pensez-vous de X_3' ?

B. Devoir Narval

Exercice 1 (3 pts) :

Nous lançons deux dés équilibrés à six faces.

Quelle est la probabilité que la somme des deux dés fasse 12 ?

Même question pour une somme égale à 7.

Exercice 2 (7 pts) :

Pour huit maisons du même modèle nous mesurons les résistances thermiques. Pour les murs nous mesurons, de la première à la huitième maison, les valeurs R_1 suivantes en mK/W :

39	41	41	43	41	37	39	39
----	----	----	----	----	----	----	----

- Déterminez la valeur moyenne \bar{R}_1 de R_1 .
- Calculez l'écart-type σ_1 sur R_1 .
- Avec une confiance de 95%, quelle est l'incertitude ΔR_1 ?

Nous mesurons maintenant les résistances thermiques R_2 des huisseries. Pour les mêmes maisons, dans le même ordre, les valeurs mesurées en mK/W :

104	102	98	102	98	102	96	98
-----	-----	----	-----	----	-----	----	----

- d) Les grandeurs R_1 et R_2 sont-elles corrélées ?
- e) Déterminez la résistance thermique R de l'ensemble {murs \oplus huisseries} avec son incertitude.

Exercice 3 (3 pts) :

Nous faisons l'hypothèse d'égalité des salaires hommes-femmes au sein de l'entreprise suivante :

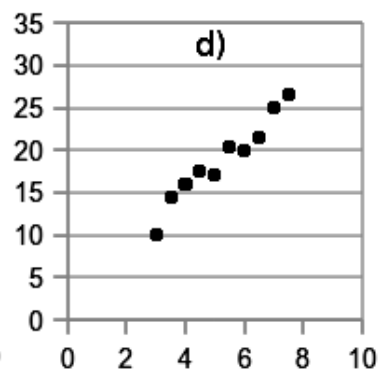
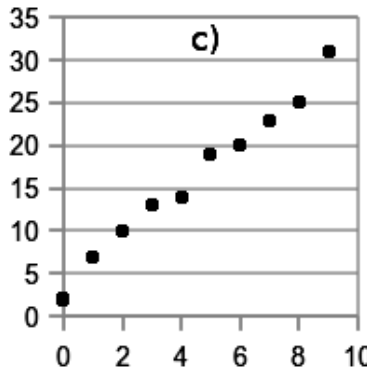
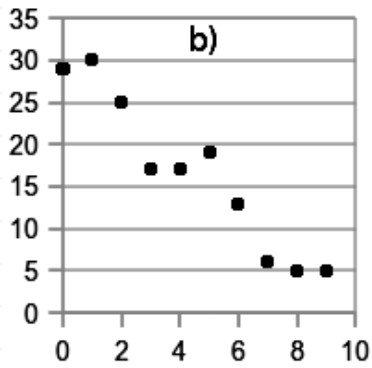
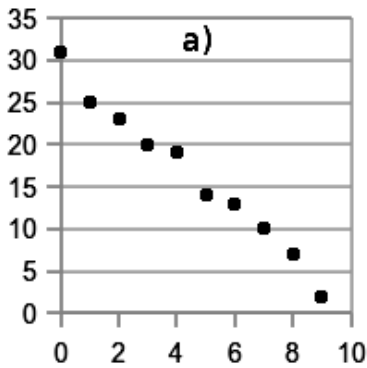
Salaires en € :	1000 à 1999	2000 à 2999	3000 à 3999	4000 à 4999
Hommes (effectifs):	139	124	68	29
Femmes (effectifs) :	261	66	22	11

L'hypothèse est-elle soutenable avec une probabilité critique de 5% ?

Exercice 4 (3 pts) :

Pour les courbes suivantes de température T (°C) dans une habitation en fonction du temps t (h), indiquez si des cas du tableau correspondent à des graphes. Dans les cas positifs donnez la correspondance, sinon, expliquez pourquoi cela n'est pas possible. N'oubliez pas que vos réponses doivent être scientifiquement justifiées.

Cas :	1)	2)	3)	4)
pente a :	3,16	-5,80	2,90	-3,01
incertitude sur a :	0,63	0,31	0,31	0,69
ordonnée à l'origine b :	2,27	20,44	3,36	30,13
incertitude sur b :	3,40	1,70	1,70	3,70
coeff. de corrélation :	0,9712	-0,9914	0,9914	-0,9627



Exercice 5 (8 pts dont 4 en bonus) : Isolants

Nous voulons réaliser une analyse en composantes principales avec les huit isolants du tableau :

Données :

	éco-bilan		iso- la- tion	con- fort d'été
	constr- uction	démo- lition		
1. Vermiculite	-	.	-	.
2. Polyuréthane	--	--	++	.
3. Liège expansé	.	+	+	++
4. Laine de roche	-	-	+	.
5. Laine de mouton	.	+	+	+
6. Isolant mince	--	--	--	--
7. Béton cellulaire	.	.	+	++
8. Botte de paille	+	+	+	++

Pour chiffrer, nous transformons les appréciations (--, -, ., ., +, ++) en notes (-2, -1, 0, 1, 2).

Montrez sur l'exemple de x_1^1 comment on passe de la matrice de données \mathbf{M} à la matrice de données réduite \mathbf{m} suivante :

	X_1	X_2	X_3	X_4
l_1	$x_1^1 = ?$	$x_2^1 = 0,195$	$x_3^1 = -1,146$	$x_4^1 = -0,444$
l_2	$x_1^2 = -1,296$	$x_2^2 = -1,365$	$x_3^2 = 1,146$	$x_4^2 = -0,444$
l_3	$x_1^3 = 0,589$	$x_2^3 = 0,975$	$x_3^3 = 0,382$	$x_4^3 = 0,977$
l_4	$x_1^4 = -0,354$	$x_2^4 = -0,585$	$x_3^4 = 0,382$	$x_4^4 = -0,444$
l_5	$x_1^5 = 0,589$	$x_2^5 = 0,975$	$x_3^5 = 0,382$	$x_4^5 = 0,266$
l_6	$x_1^6 = -1,296$	$x_2^6 = -1,365$	$x_3^6 = -1,909$	$x_4^6 = -1,864$
l_7	$x_1^7 = 0,589$	$x_2^7 = 0,195$	$x_3^7 = 0,382$	$x_4^7 = 0,977$
l_8	$x_1^8 = 1,532$	$x_2^8 = 0,975$	$x_3^8 = 0,382$	$x_4^8 = 0,977$

Quel est l'intérêt d'une telle transformation ?

Nous construisons ensuite la matrice des corrélations \mathbf{r} .

Complétez les valeurs manquantes :

?	?	0,360	0,873
?	?	0,255	0,811
?	?	?	0,659
?	?	?	?

Nous diagonalisons ensuite la matrice \mathbf{r} et nous obtenons les valeurs propres :

$\lambda_2 = 0,8552$; $\lambda_3 = 0,0779$ et $\lambda_4 = 0,0588$.

Que vaut λ_1 ?

Expliquez le principe qui permet d'obtenir les valeurs propres. Nous obtenons ensuite la matrice de passage \mathbf{P} , les colonnes sont les vecteurs propres en fonction des vecteurs de la base initiale :

0,544	0,285	0,488	0,620
0,520	0,411	-0,747	-0,056
0,347	-0,856	-0,251	0,287
0,560	-0,127	0,375	-0,728

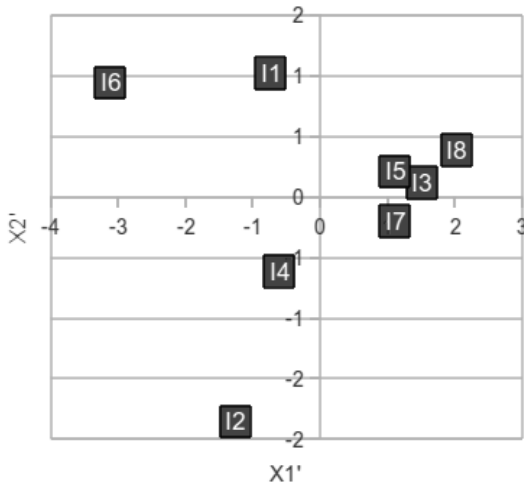
Expliquez le principe qui permet d'obtenir la matrice de passage.

Comment à partir de cette matrice nous déterminons la nouvelle matrice \mathbf{m}' des individus avec les nouvelles caractéristiques ?

Illustrez en calculant l'élément $x'_2{}^2$.

	X_1'	X_2'	X_3'	X_4'
I_1	-0,738	1,017	-0,197	-0,236
I_2	-1,265	-1,856	-0,068	-0,075
I_3	1,507	0,118	-0,171	-0,291
I_4	-0,612	-0,612	0,002	0,247
I_5	1,109	0,208	-0,437	0,226
I_6	-3,122	0,941	0,168	0,081
I_7	1,101	-0,203	0,412	-0,247
I_8	2,019	0,387	0,290	0,294

Finalement nous obtenons la distribution suivante des individus:

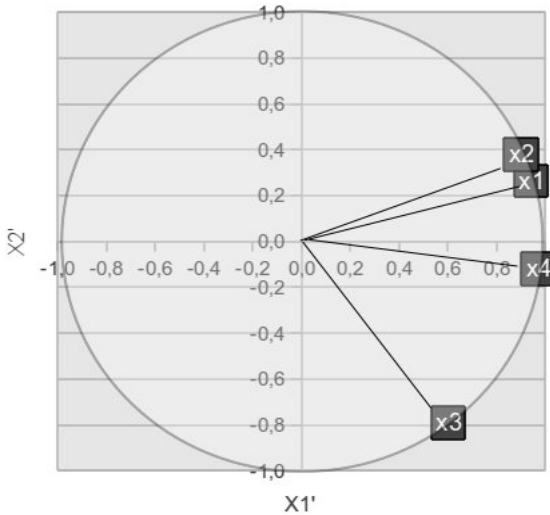


Quel sens donneriez-vous à X_1' et X_2' ?

En omettant X_3' et X_4' avons-nous perdu de l'information ?

Vous pouvez aussi vous aider des corrélations entre nouvelles et anciennes variables, et du cercle des corrélations :

	X_1'	X_2'	X_3'	X_4'
X_1	$r_{11}' = 0,943$	$r_{12}' = 0,264$	$r_{13}' = 0,136$	$r_{14}' = 0,150$
X_2	$r_{21}' = 0,901$	$r_{22}' = 0,380$	$r_{23}' = -0,208$	$r_{24}' = -0,014$
X_3	$r_{31}' = 0,602$	$r_{32}' = -0,792$	$r_{33}' = -0,070$	$r_{34}' = 0,070$
X_4	$r_{41}' = 0,972$	$r_{42}' = -0,117$	$r_{43}' = 0,105$	$r_{44}' = -0,177$



Quels enseignements tirez-vous concernant les différents types d'isolants ?

Apparaît-il une classification particulière selon que nous ayons une isolation répartie, un isolant synthétique, une laine minérale ou un isolant sain (biosourcé) ?

VI. CORRECTIONS

Corrections complètes ou sommaire suivant les cas.

Chapitre I : Variable Aléatoire

E1 : Âges

$n=15$; mode=18; médiane=18; moyenne $\approx 18,333$; moy. géométrique $\approx 18,303$; étendue=4; écart-type $\approx 1,113$; écart quadratique moyen $\approx 1,075$; écart moyen $\approx 0,844$.

E2 : Cartes

Nb de tirages possibles : $32 \times 31 \times 30 \times 29 \times 28$ divisé par toutes les permutations possibles des 5 cartes $5 \times 4 \times 3 \times 2 \times 1 = 5!$, soit **201 376** jeux.

Nb de tirages correspondant aux carrés d'As : une seule manière d'avoir les 4 As, et 28 possibilités pour la carte qui les accompagne, donc **28** jeux possibles (un jeu = 120 tirages possibles).

D'où la probabilité $p = 28/201376 = 139$ chances sur un million = 1 chance sur 7192 = **0,014%**.

Pour une couleur : chaque couleur a 8 cartes. Nb de manières de choisir 5 cartes parmi 8 = $8! / (5! (8-5)!) = C_8^5$ (combinaison) = 56.

Comme il y a 4 couleurs : $4 \times 56 = 224$.

D'où la probabilité $p = 224 / 201\ 376 = 1$ chance sur 899 = **0,11%**.

E3 : Champ de pesanteur

a) Très dispersé. On pourrait indiquer moins de chiffres significatifs.

b) $\bar{g} = 9,32 \text{ m/s}^2$. $\sigma_g = 1,994 \text{ m/s}^2 \approx 2 \text{ m/s}^2$

c) $\Delta g = 2,36.1,994/\sqrt{8} = 1,66 \text{ m/s}^2$. D'où $g=9,32\pm 1,66\text{m/s}^2$ et la valeur attendue appartient à l'intervalle :

$$7,66 \text{ m/s}^2 < g = 9,81\text{m/s}^2 < 10,98 \text{ m/s}^2$$

d) On est environ à σ de $g \approx 10\text{m/s}^2$ d'où 68% de chance.

E4 : Test d'un isolant

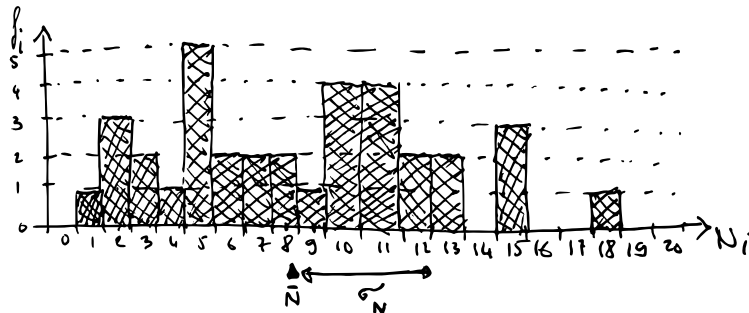
$x=39,28\pm 0,49 \text{ mW/m/K}$ à 95%. Résultats conformes. Là la marge est d'environ 1,3% au lieu de 5%. Il s'est assuré une marge de sécurité.

E5 : Devoir

a) $n=35$, $\bar{N} \approx 8,29$ et $\sigma_N \approx 4,40$.

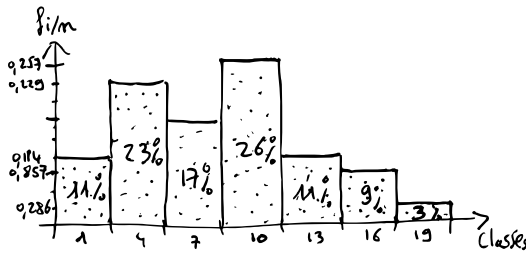
b) Notes avec leur fréquences :

Notes	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Fréq.	0	1	3	2	1	5	2	2	2	1	4	4	2	2	0	3	0	0	1	0	0



c) Classes, fréquences et fréquences relatives :

Classes	[0,1,2]	[3,4,5]	[6,7,8]	[9,10,11]	[12,13,14]	[15,16,17]	[18,19,20]
Fréq.	4	8	6	9	4	3	1
F. rel.	11,4%	22,9%	17,1%	25,7%	11,4%	8,57%	2,86%



En regroupant les données par classes il y a une perte d'information. La perte est ici faible. Mais par ailleurs nous avons un gain de visibilité. Nous pouvons plus facilement faire ressortir un comportement global et des catégories d'étudiants. Les fréquences relatives indiquent directement le pourcentage d'étudiant dans chaque classe, par exemple 34% des étudiants semblent en grande difficulté, ce qui est plus difficile d'interprétation sur le premier diagramme.

E6 : Yams

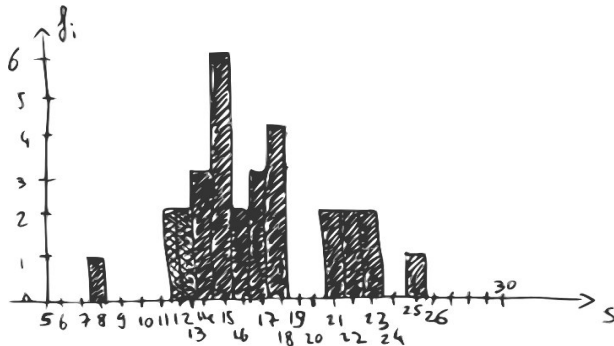
1) $1/1296$

2) $(1+5+15+35+70)/6^5 = 1,65\%$

3) a) $n=30$, $\bar{S} \approx 16,73$ et $\sigma_s \approx 3,88$.

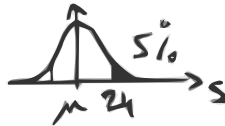
b) $\Delta \bar{S} \approx 1,45$. $\bar{S} = 16,73 \pm 1,45$ avec une confiance de 95%. Ici la population est infinie et connue. On trouve $\mu_{\text{pop}} = 17,5$. C'est cohérent. Il n'y a pas de raison de penser que les dés sont pipés.

c)



d) La somme est le résultat de cinq lancers indépendants, nous pouvons penser d'après le théorème central limite que nous nous appro-

chons d'une gaussienne : $(24-\mu)/\sigma_S \approx 1,68$. On est proche de $t_{\infty,90\%}$, donc la moitié 5% sur la queue de gaussienne de droite. Donc **1 chance sur vingt** (sur nos lancers ici : 1/30, cohérent).



E7 : Pièces

1) $\chi^2 \approx 2,56 < 3,84$. Il n'y a pas lieu de remettre en cause l'équilibre des pièces. La différence par rapport à l'équiprobabilité s'explique par les fluctuations statistiques.

Ou par un test classique :

$$|(\bar{x} = 42/100) - (\mu = 0,5)| \leq t_{\infty,95\%}(\sigma = 0,5) / \sqrt{n} = 100$$

2) $\chi^2 \approx 0,4 < 3,84$. Idem. p.c. (probabilité critique) $\approx 20,8\%$, il y a une chance sur 5 que l'hypothèse soit juste, on ne peut pas la rejeter.

3) $\chi^2 \approx 25,6 \gg 3,84$! La pièce n'est sûrement pas équilibrée. Une chance sur deux millions que l'hypothèse soit juste, on la rejette !

E8 : Parité

Valeurs espérées		$\chi^2 = \frac{(470-288,5)^2}{288,5} + \frac{(107-288,5)^2}{288,5} + \dots$	et
288,5	288,5		
174	174	$\chi^2 \approx 344$, avec ddl=2x1=2 la probabilité que
6	6		l'hypothèse soit vraie est très inférieure à une
			chance sur 1000 (13,8 \ll 344 pour 0,1%). <u>La parité n'est pas respectée.</u>

Comprenant ces résultats il faudrait être d'une mauvaise foi extrême pour prétendre qu'il n'y a aucune raison de penser que la parité n'est pas respectée en justifiant les écarts par les fluctuations statistiques.

E9 : Naissances

$\chi^2 \approx 6,16 < 7,8$. L'hypothèse ne peut être abandonnée, on a besoin

de plus d'effectif pour trancher.

$\chi^2 \approx 128 \gg 7,8$. Hypothèse totalement rejetée.

Chapitre II : Corrélations et indépendances

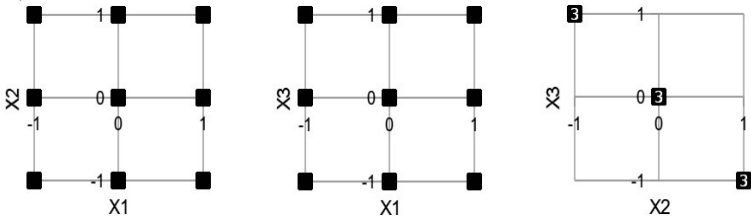
E1 : Corrélations

1- a) $\bar{x}_1 = (-1-1-1+0+0+0+1+1+1)/9=0$ de même $\bar{x}_2=0$ et $\bar{x}_3=0$.

$\sigma_1 = \sqrt{\sum_{i=1}^9 (x_{i1} - \bar{x}_1)^2 / (9-1)} = \sqrt{6 \times 1 / 8}$ d'où

$\sigma_1 = \sigma_2 = \sigma_3 = \sqrt{3}/2 \approx 0,87$.

b)



c) $r_{12}=0$. $r_{13}=0$. $r_{23}=-1$. X_1 et X_2 ne sont pas corrélées. De même pour X_1 et X_3 , X_2 et X_3 sont dépendantes et totalement corrélées.

2- a) $\bar{x}_1=0$ et $\bar{x}_2=0$. $\sigma_1 \approx 1,22$ et $\sigma_2 \approx 1,58$.

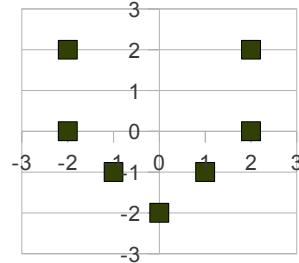
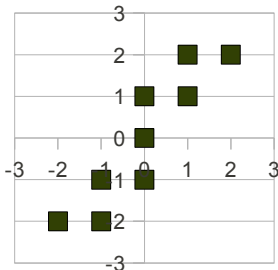
c) $r_{12}=0,904$. Grandeurs globalement corrélées positivement.

3- a) $\bar{x}_1=0$ et $\bar{x}_2=0$. $\sigma_1 \approx 1,73$ et $\sigma_2 \approx 1,53$.

c) $r_{12}=0$. Grandeurs totalement décorélées, ne pas oublier que les corrélations recherchées sont ici **linéaires**. Il y a une corrélation en V.

2- b)

3- c)



E2 : Volumes

1-

$$\bar{V}_1 = (100,1 + 100,0 + 99,9 + 100,0) / 4 \text{ d'où } \bar{V}_1 = \mathbf{100,0 \text{ mL}}$$

$$\sigma_1 = \sqrt{\frac{0,1^2 + 0^2 + (-0,1)^2 + 0^2}{4-1}} = \sqrt{\frac{2}{3}} \cdot 0,1 \text{ mL} \text{ d'où } \sigma_1 \approx 0,082 \text{ mL}$$

D'après le théorème central limite : $\Delta V = t \cdot \sigma / \sqrt{n} = 3,18 \times 0,0816 / 2$

$t_{ddl=3; 95\%} = 3,18$ soit $\Delta \bar{V}_1 \approx \mathbf{0,13 \text{ mL}}$ et $\Delta \bar{V}_1 / \bar{V}_1 \approx 0,13 / 100$

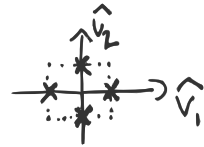
La pipette, avec une confiance de 95%, est à 0,13 mL près, soit pour 100 mL à 0,13% de précision.

2-

$$\hat{V}^i = V^i - \bar{V} \text{ et}$$

$$\sum_i [(V_1^i - \bar{V}_1)(V_2^i - \bar{V}_2)] = \sum_i [\hat{V}_1^i \hat{V}_2^i] = 0,1 \cdot 0 + 0,0,1 + (-0,1) \cdot 0 + 0,0,1 = 0$$

$$\text{or } r_{12} = \frac{\sum_i [(V_1^i - \bar{V}_1)(V_2^i - \bar{V}_2)]}{\sqrt{\sum_i (V_1^i - \bar{V}_1)^2 \sum_i (V_2^i - \bar{V}_2)^2}}$$



soit $r_{12} = \mathbf{0}$, les grandeurs sont totalement décorrélées et donc indépendantes.

3-

$$V = \{200,1 ; 200,1 ; 199,9 ; 199,9\} \text{ mL d'où } \bar{V} = \mathbf{200 \text{ mL}}$$

$$\sigma_V = \sqrt{\frac{0,1^2 + 0,1^2 + (-0,1)^2 + (-0,1)^2}{4-1}} = \frac{2}{\sqrt{3}} \cdot 0,1 \text{ mL}$$

d'où $\sigma_V \approx 0,115 \text{ mL}$ et $\Delta \bar{V} \approx \mathbf{0,183 \text{ mL}}$ et $\Delta \bar{V} / \bar{V} \approx 0,09\%$

4-

$V(V_1, V_2)$ d'où :

$$\Delta V^2 = \left(\frac{\partial V}{\partial V_1}\right)^2 \Delta V_1^2 + \left(\frac{\partial V}{\partial V_2}\right)^2 \Delta V_2^2 = \Delta V_1^2 + \Delta V_2^2$$

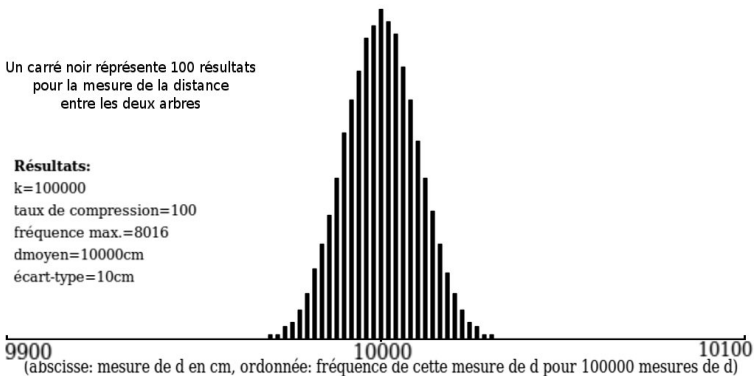
et $\Delta V = \sqrt{2} \cdot \Delta V_1 \approx 0,18$ même résultat qu'au 3-.

E3 : Arbres

Nous avons $d = \sum x_i$ (soit $d = x_1 + x_2 + \dots + x_i + \dots + x_n$), où les x_i sont les mesures de longueur pour chaque report de bâton. Δx_i est l'incertitude sur chaque mesure x_i . Ici $\Delta x_i = 1\text{cm}$ et $x_i = 100 \pm 1\text{cm}$.

Quelle est alors l'incertitude Δd ?

Spontanément nous pourrions penser que $\Delta d = \sum \Delta x_i$, d'où ici $\Delta d = 100\text{cm}$ soit 1m pour 100m mesurés d'un arbre à l'autre. Mais problème, ce n'est absolument pas ce qu'indique la simulation ci-dessous !



Elle indique plutôt $\Delta d = 0,1\text{m}$. En effet un simple calcul de probabilités montre qu'il est extrêmement peu probable d'obtenir pour d , 99m ou 101m , alors que l'on a une chance sur deux, pour x_i , d'avoir $0,99\text{m}$ ou $1,01\text{m}$. Or nous désirons avoir Δd à probabilité égale de Δx_i .

Ici $\partial d / \partial x_i = 1$ et $\Delta x_i = \Delta x$, quelquesoit $i = 1 \dots n$, ainsi $\Delta d = \sqrt{n} \cdot \Delta x$ soit $\Delta d = 0,1\text{m}$, ce qui correspond bien au résultat attendu !

E4 : Méthode de Bessel

$(\partial f/\partial D)_d = (D^2 + d^2)/4d^2$ et $(\partial f/\partial d)_D = -d/2D$, d'où

$$\Delta f = \sqrt{[(D^2 + d^2)/4d^2]^2 (\Delta D)^2 + (d/2D)^2 (\Delta d)^2}$$

puis $f = 464,1 \pm 3,8$ mm

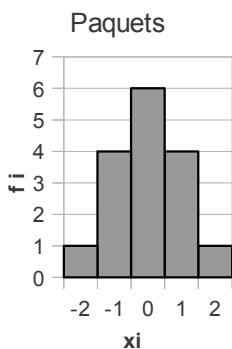
et $\Delta f/f = 0,8\%$.

Soit par un calcul numérique avec un tableur et des macros (pas besoin alors d'effectuer des calculs de dérivées partielles) :

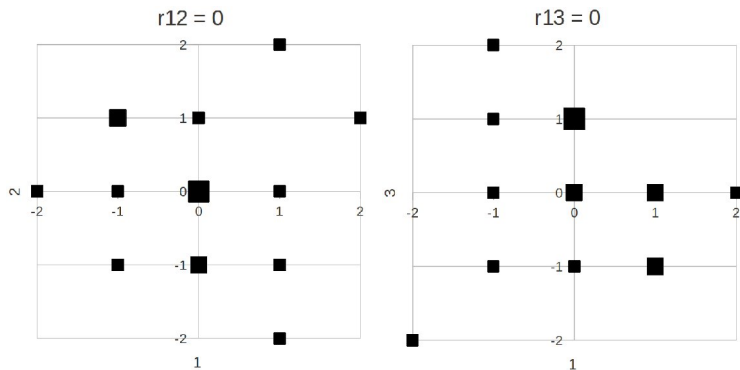
The screenshot shows a spreadsheet window titled "IncertitudesLibres00o32.ods - OpenOffice.o". The menu bar includes "Fichier", "Édition", "Affichage", "Insertion", "Format", "Outils", and "Donné". The formula bar shows the function $f(x) = \sum = (x^2 - y^2)/(4 * x)$. Below the formula bar, a grey box titled "Calcul d'incertitudes" contains the following data:

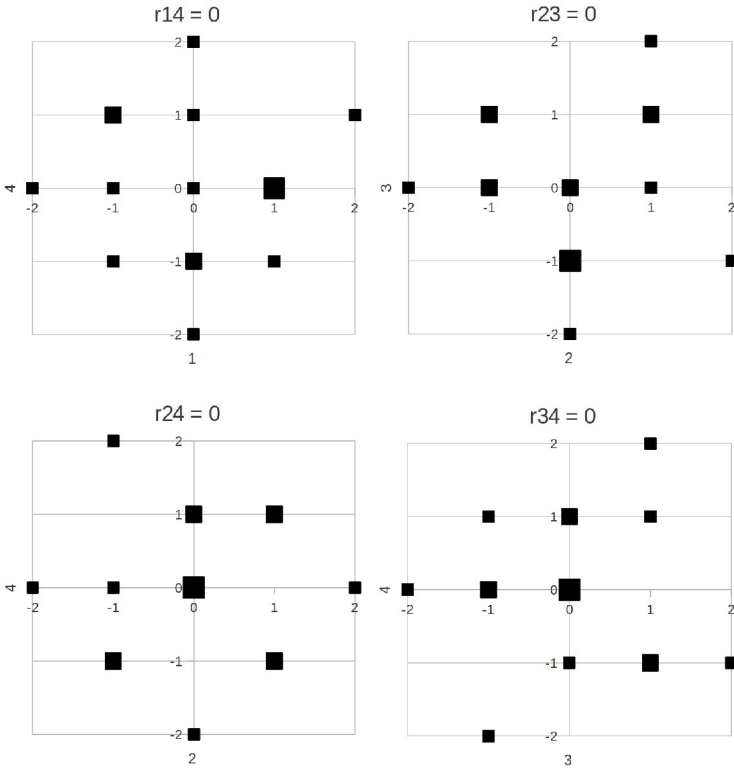
x=	2000	+/-	10
y=	536	+/-	20
z=	0	+/-	0
w=	0	+/-	0
f=	464,09	+/-	3,77
		soit à	0,81 %

Le tableur
 utilise ici 2 des 4 paquets
 globalement gaussiens
 parfaitement décorrélés
 disponibles :



1	2	3	4	x	y	f
-2	0	-2	0	1973,71	536	457,04
-1	-1	0	0	1986,86	509,71	464,02
-1	0	-1	1	1986,86	536	460,56
-1	1	1	1	1986,86	562,29	456,93
-1	1	2	-1	1986,86	562,29	456,93
0	0	0	0	2000	536	464,09
0	0	-1	-2	2000	536	464,09
0	-1	1	-1	2000	509,71	467,52
0	1	1	-1	2000	562,29	460,48
0	-1	1	2	2000	509,71	467,52
0	0	0	1	2000	536	464,09
1	-2	0	0	2013,14	483,43	474,26
1	0	-1	0	2013,14	536	467,61
1	-1	0	-1	2013,14	509,71	471,02
1	2	-1	0	2013,14	588,57	460,27
2	1	0	1	2026,29	562,29	467,56





E5 : Indice

D'après la loi de Descartes pour la réfraction : $n_2 = n_1 \frac{\sin(i_1)}{\sin(i_2)}$

d'où $n_2 \approx 1,462$.

$n_2(i_1, i_2)$ alors

$$\Delta n_2^2 = \left(\frac{\partial n_2}{\partial i_1} \right)_{i_2}^2 \Delta i_1^2 + \left(\frac{\partial n_2}{\partial i_2} \right)_{i_1}^2 \Delta i_2^2$$

(la variation en fonction de l'indice n_1 n'est pas incluse car il est supposé connu avec une grande précision)

$$\text{On a } \left(\frac{\partial n_2}{\partial i_1} \right)_{i_2} = \frac{n_1}{\sin(i_2)} \cos(i_1) \quad \text{et} \quad \left(\frac{\partial n_2}{\partial i_2} \right)_{i_1} = n_1 \sin(i_1) \frac{-\cos(i_2)}{\sin^2(i_2)} .$$

$$\text{soit } \Delta n_2 = \left[\left(\frac{\cos(30^\circ)}{\sin(20^\circ)} \frac{\pi}{180} \right)^2 + \left(\frac{\sin(30^\circ) \cos(20^\circ)}{\sin^2(20^\circ)} \frac{2\pi}{180} \right)^2 \right]^{1/2}$$

(les angles sont mis dans leur unité naturelle, les radians, sans unité :
 $\pi \text{ rad} = 180^\circ$)

Finalement $\Delta n_2 = 0,1470$ avec une confiance de 95%,
 $n_2 = 1,46 \pm 0,15$ et $\Delta n_2/n_2 = 10\%$.

Avec le tableur : $n_2 = 1,46 \pm 0,16$ et $\Delta n_2/n_2 = 11\%$. La méthode numérique est approximative mais nous vérifions ainsi que nous n'avons pas fait d'erreur de calcul.

Par ailleurs avec un goniomètre nous pouvons faire des mesures d'angles bien plus précises à quelques minutes d'arc près (un minute d'arc = $1' = 1^\circ/60$).

E6 : Formule de Cauchy

1 -

$$\Delta n^2 = \left(\frac{\partial n}{\partial D_m} \right)_A^2 \Delta D_m^2 + \left(\frac{\partial n}{\partial A} \right)_{D_m}^2 \Delta A^2$$

$$\text{avec } \left(\frac{\partial n}{\partial D_m} \right)_A = \frac{\cos[(A+D_m)/2]}{2 \sin(A/2)} \quad \text{et}$$

$$\left(\frac{\partial n}{\partial A} \right)_{D_m} = \frac{\cos[(A+D_m)/2] \sin(A/2) - \sin[(A+D_m)/2] \cos(A/2)}{2 \sin^2(A/2)}$$

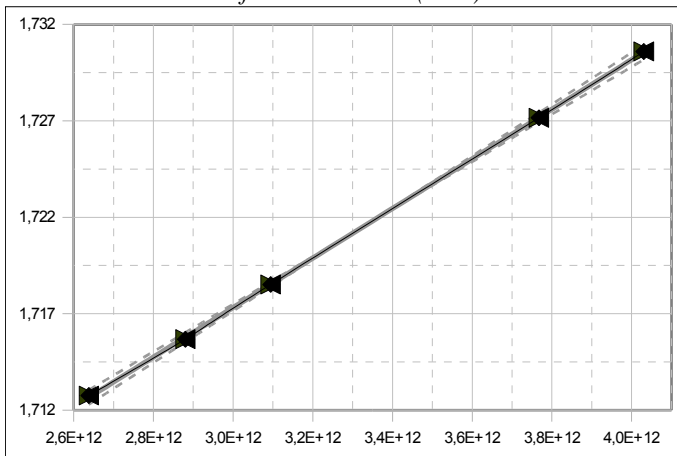
avec $A=60^\circ \simeq D_m$: $\left(\frac{\partial n}{\partial D_m}\right)_A \simeq \frac{1}{2}$ et $\left(\frac{\partial n}{\partial A}\right)_{D_m} \simeq -1$

d'où $\Delta n \simeq 0,00065$ et $\Delta n/n \simeq 0,04\%$
 (même chose en calcul numérique)

2-

Avec les barres d'erreurs, $\Delta n=0,00065$ et $\Delta(1/\lambda^2)=2\Delta\lambda/\lambda^3$ avec $\Delta\lambda=0,05\text{nm}$, on trouve **$A=1,67877 \pm 0,00182$** ; $\Delta A/A=0,109\%$;
 $B=(1,285 \pm 0,055) \cdot 10^{-14} \text{m}^2$ et $\Delta B/B=4,26\%$

n en fonction de $1/\lambda^2 (1/\text{m}^2)$:



La régression simple donne **$r=0,99995...$**

(Si on oublie les barres d'erreurs $A=1,67877 \pm 0,00075$;
 $\Delta A/A=0,044\%$; $B=(1,285 \pm 0,022) \cdot 10^{-14} \text{m}^2$ et $\Delta B/B=1,75\%$)

3- En régression simple :

- $r=0,99947...$ pour $n(1/\lambda)$
- $r=0,99995...$ pour $n(1/\lambda^2)$
- $r=0,99982...$ pour $n(1/\lambda^3)$

La loi en $1/\lambda^2$ est donc meilleure,
 mais qu'elle valeur de α permet d'optimiser la régression ?

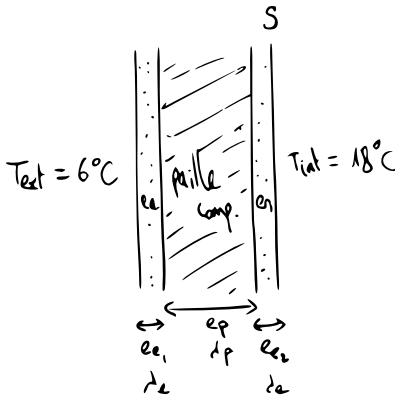
Pour cela nous allons tracer $y = \ln(n-A) = \alpha x + \ln(B)$ avec
 $x = \ln(\lambda)$. α et $\ln(B)$ seront obtenus avec leur incertitudes et A
 sera choisi de manière à avoir r_{max} .

Les résultats sont les suivants $r_{max} = -0,99997024...$ pour
 $A = 1,6843744$ et nous trouvons : $\alpha = -2,306 \pm 0,033$. Puis avec les
 barres d'erreurs : $\alpha = -2,3 \pm 0,1$. Conclusion $\alpha = -2$ n'est pas dans
 l'intervalle le modèle en $1/\lambda^2$ n'est pas validé.

Ceci dit il fonctionne correctement. La différence peut
 s'expliquer simplement car le modèle de l'électron élastiquement lié,
 utilisé ici, comporte d'autres termes (il s'agit d'un développement
 limité d'une fonction plus complexe) : $n = A + \frac{B}{\lambda^2} + \frac{C}{\lambda^4}$. Une étude
 plus avancée pourrait le vérifier en incluant le nouveau terme en $1/\lambda^4$.

Il faudrait aussi vérifier nos incertitudes expérimentales, il
 est possible que nous ayons négligé, ou sous-estimé, des sources
 d'incertitudes (par exemple en triplant les incertitudes sur les angles à
 $6'$: $\alpha = -2,3 \pm 0,3 ...$).

E7 : Mur



1-

$$R_p = \frac{e_p}{\lambda_p S} = \frac{0,4 \text{ m K m}}{0,045 \text{ W} \cdot 72 \text{ m}^2}$$

d'où $R_p = 0,123 \text{ K/W}$

$$\left(\frac{\Delta R_p}{R_p}\right)^2 = \left(\frac{\Delta e_p}{e_p}\right)^2 + \left(\frac{\Delta \lambda_p}{\lambda_p}\right)^2$$

et $\Delta R_p / R_p = 0,103$

$$\underline{R_p = 0,123 \pm 0,013 \text{ K/W}}$$

2- $R_e = 0,0069 \pm 0,0010 \text{ K/W}$

3- Les résistances sont en série : $R = R_p + R_e$

et $\Delta R = \sqrt{(\Delta R_p)^2 + (\Delta R_e)^2}$ d'où $R = 0,130 \pm 0,015 \text{ K/W}$

4- $\Phi = \frac{\Delta T}{R} = \frac{12}{0,130} = 92,3 \text{ W}$

$$(\Delta \Phi / \Phi)^2 = (\Delta(\Delta T) / \Delta T)^2 + (\Delta R / R)^2 \quad \Delta T = T_{\text{intérieur}} - T_{\text{extérieur}}$$

d'où $\Delta(\Delta T) = \sqrt{2} \cdot 0,5^\circ \text{C}$, $\Delta \Phi / \Phi = 12,9\%$ et

$80 \text{ W} \leq \Phi \leq 104,3 \text{ W}$ d'où une puissance minimale de chauffage :

$$P_{\text{min}} = 105 \text{ W}$$

E8 : Isolation et inertie

1- En parallèle : $\frac{1}{R_{\text{éq}}} = \sum_i \left(\frac{1}{R_i} \right) = \sum_i \left(\frac{\lambda_i S_i}{e_i} \right)$

ou en terme de conductances thermiques $G = 1/R$: $G_{\text{éq}} = \sum_i G_i$

donc dans notre cas :

$$\frac{1}{R} = \frac{S_s}{(e_s / \lambda_s)} + \frac{S_t}{(e_t / \lambda_t)} + \frac{S_m}{(e_m / \lambda_m)} + \frac{S_h}{(e_h / \lambda_h)} = \frac{36}{4} + \frac{54}{8} + \frac{82}{4} + \frac{8}{1}$$

pour les incertitudes :

$$\Delta G = \sqrt{\sum_i \Delta G_i^2} \quad , \quad \Delta G_i = S_i \frac{\Delta(e_i / \lambda_i)}{(e_i / \lambda_i)^2} \quad , \quad \Delta G \simeq 2,471 \text{ W/K}$$

et $\Delta R = R^2 \Delta G \simeq 1,26 \text{ mK/W}$

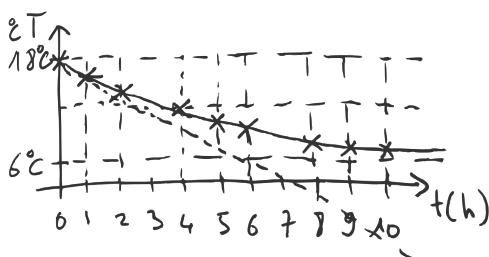
finalement $R = 22,6 \pm 1,3 \text{ mK/W}$ et $\Delta R / R = 5,6 \%$

$$2- \Phi = \frac{\Delta T}{R} = \frac{12}{22,6 \cdot 10^{-3}} \approx 531 \text{ W} \pm 8,1\% \text{ (comme à l'exercice Mur)}$$

d'où une puissance minimale de chauffage :

$$P_{\min} = 575 \text{ W}$$

3-

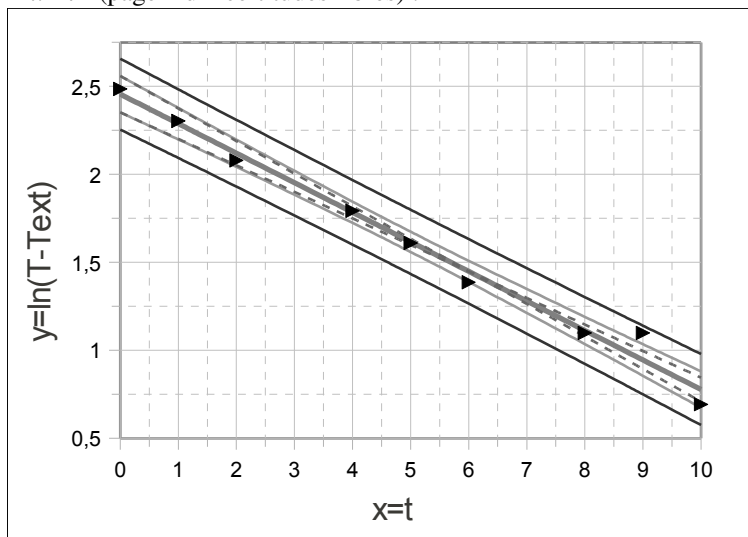


Nous avons nécessairement une asymptote à T_{ext} : la température ne peut descendre spontanément en dessous de T_{ext} .

Si linéaire la température tendrait vers moins l'infini avec le temps.

$$T - b = a e^{-t/\tau} \text{ d'où } \ln(T - b) = \ln a - \frac{1}{\tau} t ,$$

on trouve $b = T_{\text{ext}} = 5,7^\circ\text{C}$ (coefficient de corrélation r maximal),
 puis on fait une régression sur un tableur avec $y = \ln(T - b)$ et
 $x = t$ (page 2 d'IncertitudesLibres) :



coefficient directeur = $-1/\tau = -0,156 \pm 0,0157 \text{ h}^{-1}$
 ordonnée à l'origine = $\ln a = \ln (T_{\text{int}} - T_{\text{ext}}) = 2,474 \pm 0,0949$
 $r_{\text{max}} = -0,9937170474$ (meilleur alignement des points)

Nous avons : $\frac{\Delta \tau}{\tau} = \frac{\Delta (1/\tau)}{(1/\tau)}$ et $\Delta x = x \cdot \Delta (\ln x)$

D'où : $\tau = 6 \text{ h } 25 \text{ min à } 10,1\%$ et $T_{\text{int}} = 17,5 \pm 1,2 \text{ }^\circ\text{C}$.

4-

a)

flux sortant d'énergie pendant dt

= énergie calorifique perdue pendant ce même intervalle de temps

soit $\Phi dt = \frac{\Delta T}{R} dt = \frac{T(t) - T_{\text{ext}}}{R} dt = C dT$

et $\frac{dT}{dt} + \frac{1}{\tau} T = \frac{1}{\tau} T_{\text{ext}}$ avec $\tau = RC$

où l'expression du 3- est la solution générale.

b) $C = \tau / R$ et $(\Delta C / C)^2 = (\Delta \tau / \tau)^2 + (\Delta R / R)^2$

d'où $C = 1034 \pm 119 \text{ kJ/K}$

E9 : Rendement

1-

Sur un tableur (droite de régression sur le graphique) :

n = 7

Conf.: 95 %

t = 2,57

x	y	(xi-xm)	(xi-xm) ²	yim	(yi-yim) ²	xi ²	(xi-xm) ²	(yi-yim) ²	(xi-xm) *(yi-yim)
100	41	5271,43	90000	40,79	0,05	10000	90000	308,76	5271,4
200	44	2914,29	40000	46,71	7,37	40000	40000	212,33	2914,3
300	53	557,14	10000	52,64	0,13	90000	10000	31,04	557,14
400	63	0	0	58,57	19,61	160000	0	19,61	0
500	66	742,86	10000	64,5	2,25	250000	10000	55,18	742,86
600	65	1285,71	40000	70,43	29,47	360000	40000	41,33	1285,7
700	78	5828,57	90000	76,36	2,70	490000	90000	377,47	5828,6

$xm = ym = Cov(x,y) = Var(x) =$
 400 58,57 16600 280000

$sy/x = sx^2 = sx = sy = s2xy =$
 3,51 1400000 216,02 13,20 2766,7

da = 0,017

da/|a| = 29%

db = 8

db/|b| = 22%

a = 0,059 + ou - 0,017

b = 35 + ou - 8

0,042 < a < 0,076

27 < b < 42

dy =

r = 0,9701

9,021

r: coefficient de corrélation
linéaire algébrique.

sa = 0,00663

sb = 2,96579

Coefficient de corrélation

linéaire critique:

rc = 0,669

Test de corrélation (linéaire):

si $|r| < rc$ alors grandeur non corrélés

si $|r| > rc$ alors grandeur corrélés

lci les grandeurs sont corrélées.

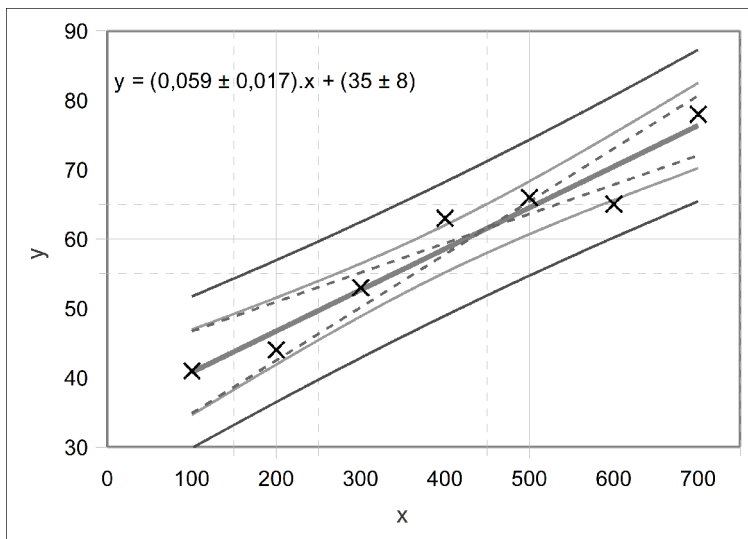
lci les grandeurs sont corrélées.

R = 0,941

R: coefficient de détermination ($R=r^2$).

94,1% de la variation de y est
expliquée par x

Rendement en fonction de la quantité d'engrais :



2- Le rendement serait alors de 67 Q/ha à 4 Q/ha près avec une confiance de 95 pour cent.

Pour $x_0 = 550$
Estimation de la moyenne de y_0 :
 $63 < y_0 < 72$
 $y_{om} = 67$
 $dy_{om} = 4$
 $dy_{om}/|y_{om}| = 6\%$

3- Le rendement serait alors de 35 Q/ha à 8 Q/ha près avec une confiance de 95 pour cent (ordonnée à l'origine).

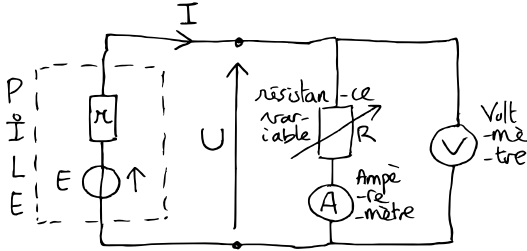
Pour $x_0 = 0$
Estimation de la moyenne de y_0 :
 $27 < y_0 < 42$
 $y_{om} = 35$
 $dy_{om} = 8$
 $dy_{om}/|y_{om}| = 22\%$

4- La probabilité est de 95% (intervalle de prédiction à 2 sigmas).

Pour $x_0 = 250$
Prédiction pour y_0 :
 $40 < y_0 < 60$
 $y_{om} = 50$
 $dy_{om} = 10$
 $dy_{om}/|y_{om}| = 20\%$

E10 : Étude d'une pile

Nous pouvons comme pour le tracé de la caractéristique de tout dipôle utiliser le montage suivant :



En faisant varier R le montage parcourt différents points de fonctionnement du dipôle. Nous avons choisi le montage longue dérivation ($R_V \rightarrow \infty$ et $R_A \rightarrow 0$).

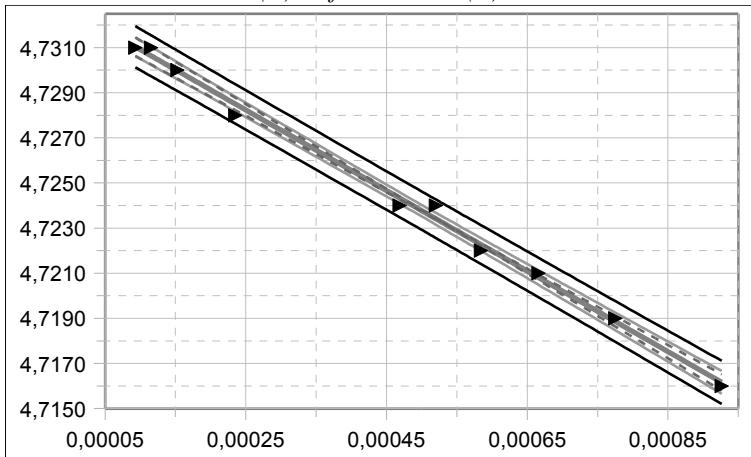
1- Le tableau donne :

$a = -17,85$ + ou - $0,92$
 $b = 4,7327$ + ou - $0,00049$
 $r = -0,9980$ Confiance: 95%

d'où :

$E = 4,7327 \pm 0,0005 \text{ V}$
 précision : $0,010\%$
 $r = 17,85 \pm 0,92 \Omega$
 précision : $5,16\%$

$U(V)$ en fonction de $I(A)$:



2-

n = 10

a estimé= -17,851

I(A)	dI(A)	U(V)	dU(V)	pois wi	wi/S
0,000093	2,16E-07	4,731	0,00537	347,4E+2	10%
0,000115	2,61E-07	4,731	0,00537	347,4E+2	10%
0,000153	3,35E-07	4,730	0,00537	347,4E+2	10%
0,000235	7,70E-07	4,728	0,00536	347,6E+2	10%
0,000469	1,24E-06	4,724	0,00536	347,8E+2	10%
0,000520	1,34E-06	4,724	0,00536	347,8E+2	10%
0,000584	1,47E-06	4,722	0,00536	347,9E+2	10%
0,000666	1,63E-06	4,721	0,00536	348,0E+2	10%
0,000775	1,85E-06	4,719	0,00536	348,1E+2	10%
0,000926	2,15E-06	4,716	0,00536	348,3E+2	10%

S=

347,8E+3

da = 6,09

da/|a| = 34,1%

sa = 6,09

db = 0,00324

db/|b| = 0,0685%

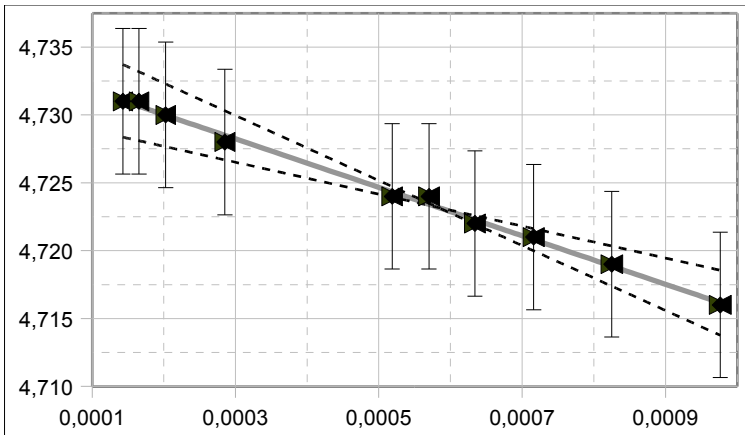
sb = 0,00324

a =	-17,85	+ ou -	6,1
b =	4,7327	+ ou -	0,0032

d'où : $E = 4,7327 \pm 0,0032 \text{ V}$, précision : 0,07 %

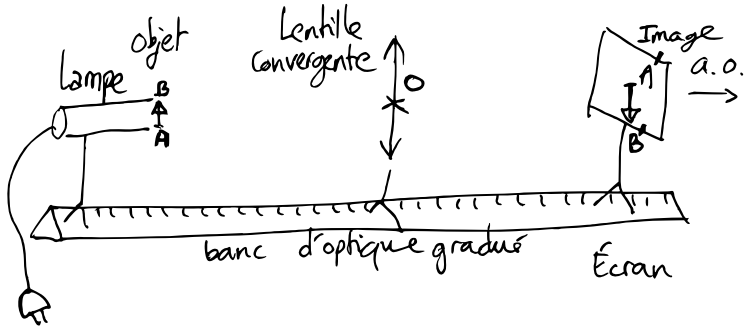
et $r = 17,85 \pm 6,1 \Omega$, précision : 34 %

U(V) en fonction de I(A) :



Ce sont bien sûr les résultats du 2- qui font foi, ceux du 1- n'ont pas tenu compte de sources d'incertitudes importantes.

E11 : Focométrie



Dans les conditions de Gauss, la relation de conjugaison au centre de la lentille mince sphérique s'écrit :

$$\frac{1}{OA'} - \frac{1}{OA} = \frac{1}{f'} \quad , f' : \text{distance focale image.}$$

Nous avons ici un objet et une image réels et suivant le choix d'axe optique $\overline{OA} < 0$ et $\overline{OA'} > 0$. Ainsi pour travailler avec des grandeurs positives nous posons : $\overline{OA} = -OA$ et $\overline{OA'} = OA'$.

La méthode la plus rapide pour mesurer la focale d'une lentille convergente est la méthode par auto-collimation. Ensuite nous avons la méthode de Bessel, plus précise mais plus longue à mettre en œuvre. L'avantage avec l'autre méthode de cet exercice, certes longue, est que nous ne travaillons pas avec une seule mesure mais avec n mesures. Plus le nombre mesure est grand plus le résultat est précis.

ΔOA contient les incertitudes géométriques (mesures de longueurs) et de modélisation (épaisseur de la lentille, conditions de Gauss). $\Delta OA'$ contient en plus les incertitudes optiques (latitudes de mise au point) :

$$\Delta OA' = \sqrt{(\Delta OA'_{\text{géo}})^2 + (\Delta OA'_{\text{mod}})^2 + (\Delta OA'_{\text{opt}})^2}$$

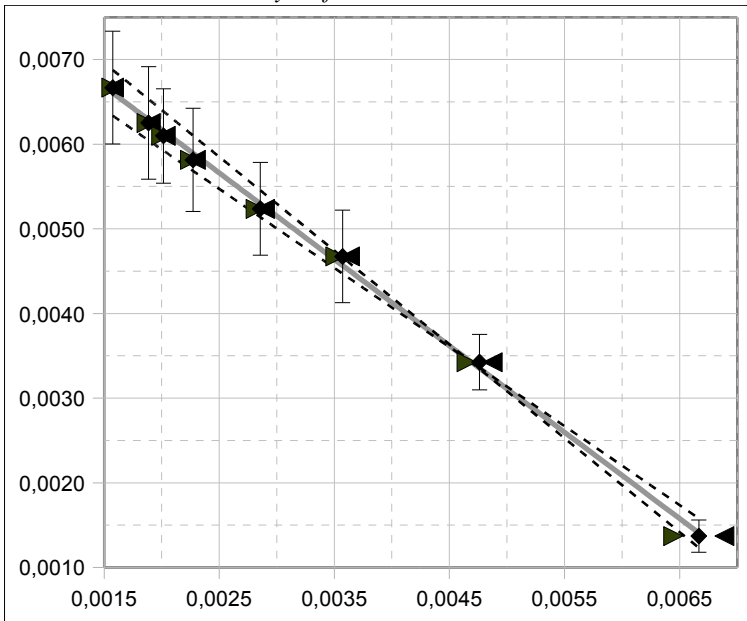
La relation entre OA et OA' est hyperbolique, nous considérons donc $x=1/OA$ et $y=1/OA'$. Nous avons alors une relations linéaire $y=ax+b$ où le coefficient directeur a doit être égal à -1 et l'ordonnée à l'origine b à $1/f$. $\Delta x = \Delta(1/OA) = \Delta OA/OA^2$, de même pour Δy .

En utilisant un tableur (ici www.incertitudes.fr/proba-stat-acp/TPlentille.ods) nous obtenons :

$a = -1,02 \pm 0,09$ et $b = 0,0082 \pm 0,0004$ soit $f' = 122 \pm 6 \text{ mm}$ (cohérent avec l'indication du fabricant $f' = 125 \text{ mm}$).

$a = -1$ est bien dans l'intervalle, nous vérifions aussi la validité de cette relation linéaire entre $1/OA$ et $1/OA'$. Pour préciser davantage la validité de la relation de conjugaison il faudrait aussi montrer que cette relation linéaire est celle qui donne la meilleure corrélation.

y en fonction de x :



n = 8		a estimé= -1,0210		a= -1,0210	
x	dx	y	dy	poids wi	wi/S
0,00157	0,0000124	0,00667	0,000667	2249189	6%
0,00189	0,0000178	0,00625	0,000664	2265985	6%
0,00202	0,0000203	0,00610	0,000558	3210643	9%
0,00227	0,0000258	0,00581	0,000608	2696211	7%
0,00286	0,0000408	0,00524	0,000548	3308045	9%
0,00357	0,0000638	0,00467	0,000546	3308566	9%
0,00476	0,0001134	0,00342	0,000328	8248034	23%
0,00667	0,0002222	0,00137	0,000191	11349080	31%

S=
36635752

da = 0,0871 da/|a| = 8,5% sa = 0,0871
db = 0,000408 db/|b| = 5,0% sb = 0,0004

a = -1,021 + ou - 0,087
b = 0,00821 + ou - 0,00041
 -1,11 < a < -0,93

f=1/b=	122
df=d(1/b)=	6
df/f=d(1/b)/ 1/b =	5%

Chapitre III : Approches multicritères

E1 : Fournisseurs

Nous avons déjà des notes, il n'est pas ici utile de normaliser, surtout qu'il n'y a que deux individus.

Fournisseur 1 : $4/10 \times 0,5 + 8/10 \times 0,3 + 5/10 \times 0,2 = 5,4/10$.

Fournisseur 2 : $6/10 \times 0,5 + 3/10 \times 0,3 + 6/10 \times 0,2 = 5,1/10$.

Choix du premier fournisseur.

E2 : Principe de l'ACP

$$\text{Cas 1 : } M = \begin{pmatrix} -2 & 4 & 0 & 8 \\ 0 & 3 & -3 & 3 \\ 0 & 3 & 3 & 13 \\ 2 & 2 & 0 & 8 \end{pmatrix} \quad m = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 \end{pmatrix}$$

$$r = \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

$$\det(r - \lambda I) = \det \begin{pmatrix} 1-\lambda & -1 & 0 & 0 \\ -1 & 1-\lambda & 0 & 0 \\ 0 & 0 & 1-\lambda & 1 \\ 0 & 0 & 1 & 1-\lambda \end{pmatrix}$$

$$\text{soit } \det(r - \lambda I) = \det \begin{pmatrix} 1-\lambda & -1 \\ -1 & 1-\lambda \end{pmatrix} \det \begin{pmatrix} 1-\lambda & 1 \\ 1 & 1-\lambda \end{pmatrix}$$

d'où $((1-\lambda)^2 - 1)(1-\lambda)((1-\lambda)^2 - 1)(1-\lambda) = 0$ et $\lambda_1 = \lambda_2 = 2$

$$\lambda_3 = \lambda_4 = 0 \quad \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x_1 - x_2 \\ -x_1 + x_2 \\ x_3 + x_4 \\ x_3 + x_4 \end{pmatrix} = \lambda_j \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \quad \text{d'où}$$

$$\vec{e}'_1 = (\vec{e}_1 - \vec{e}_2) / \sqrt{2}, \quad \vec{e}'_2 = (\vec{e}_3 + \vec{e}_4) / \sqrt{2}, \quad \vec{e}'_3 = (\vec{e}_1 + \vec{e}_2) / \sqrt{2} \quad \text{et}$$

$$\vec{e}'_4 = (\vec{e}_3 - \vec{e}_4) / \sqrt{2}. \quad \text{Puis avec } P: \quad m' = \begin{pmatrix} -\sqrt{2} & 0 & 0 & 0 \\ 0 & -\sqrt{2} & 0 & 0 \\ 0 & \sqrt{2} & 0 & 0 \\ \sqrt{2} & 0 & 0 & 0 \end{pmatrix}$$

$$\vec{I}'_1 = \begin{pmatrix} -\sqrt{2} \\ 0 \end{pmatrix}, \quad \vec{I}'_2 = \begin{pmatrix} 0 \\ -\sqrt{2} \end{pmatrix}, \quad \vec{I}'_3 = \begin{pmatrix} 0 \\ \sqrt{2} \end{pmatrix} \quad \text{et} \quad \vec{I}'_4 = \begin{pmatrix} \sqrt{2} \\ 0 \end{pmatrix}.$$

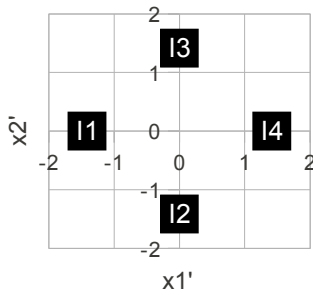
Pour le cercle des corrélations, nous calculons les corrélations entre

les nouvelles et les anciennes valeurs

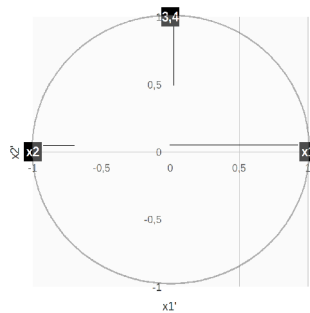
$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

(nous l'obtenons en calculant les corrélations entre les colonnes de m et m').

Individus selon X_1' et X_2' :



Cercle des corrélations :



Cas 2 : $\bar{x}_1=2$, $\bar{x}_2=2$ et $\bar{x}_3=2$.

$$M = \begin{pmatrix} 1 & 3 & 2 \\ 2 & 1 & 3 \\ 3 & 2 & 1 \end{pmatrix} \quad m = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{pmatrix} .$$

Matrice des corrélations : $r = \begin{pmatrix} 1 & -0,5 & -0,5 \\ -0,5 & 1 & -0,5 \\ -0,5 & -0,5 & 1 \end{pmatrix} .$

Valeurs propres : $\lambda_1 = \lambda_2 = 1,5$ et $\lambda_3 = 0$.

100% de l'information sera représentée par X_1' et X_2' .

Vecteurs propres : $\vec{e}'_1 \simeq -0,118\vec{e}_1 - 0,648\vec{e}_2 + 0,766\vec{e}_3$,
 $\vec{e}'_2 \simeq 0,816\vec{e}_1 - 0,408\vec{e}_2 - 0,408\vec{e}_3$ et $\vec{e}'_3 \simeq -0,577(\vec{e}_1 + \vec{e}_2 + \vec{e}_3)$.

Matrice de passage : $P \simeq \begin{pmatrix} -0,118 & 0,816 & -0,577 \\ -0,648 & -0,408 & -0,577 \\ 0,766 & -0,408 & -0,577 \end{pmatrix}$ et

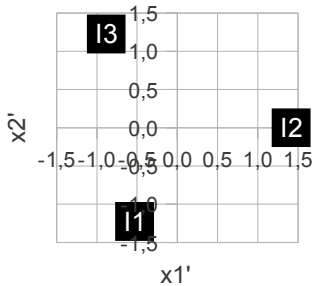
$$m' \simeq \begin{pmatrix} -0,530 & -1,225 & 0 \\ 1,414 & 0 & 0 \\ -0,884 & 1,225 & 0 \end{pmatrix} .$$

$$\vec{I}'_1 \simeq \begin{pmatrix} -0,530 \\ -1,225 \end{pmatrix} , \quad \vec{I}'_2 \simeq \begin{pmatrix} 1,414 \\ 0 \end{pmatrix} \quad \text{et} \quad \vec{I}'_3 \simeq \begin{pmatrix} -0,884 \\ 1,225 \end{pmatrix} .$$

Corrélations entre nouvelles et anciennes caractéristiques :

$$\begin{pmatrix} -0,144 & 1 & 0 \\ -0,794 & -0,5 & 0 \\ 0,938 & -0,5 & 0 \end{pmatrix}$$

Individus selon X_1' et X_2' :



Cercle des corrélations :



Devoir Suricate :

E1 : Pièces

Nb de lancers possibles : $2 \times 2 \times 2 \times 2 = 16$.

Une seule possibilité correspond à 4 piles d'où une probabilité de **une chance sur 16** soit 6,25%.

Quatre possibilités pour un pile : **PFFF, FPF, FFP, FFFP**. D'où

$p=4/16$ soit **une chance sur quatre** ($p=0,25=25\%$).

E2 : Eaux minérales (bicarbonates et sulfates)

a) $\bar{B}=1234,8$ mg/L.

b) $\sigma_B=1469$ mg/L.

c) $\Delta B=t \cdot \sigma_B / \sqrt{n}$. $t(\text{ddl}=8;95\%)=2,31$. $\Delta B=1131$ mg/L. Grande dispersion, la teneur en bicarbonate est très variable d'une eau à l'autre.

d) $r_{BS} = -0,265 < 0,52$. Les grandeurs peuvent être considérées comme **non corrélées**.

e) D'après la formule de propagation des incertitudes : si $T=B+S$ on a $\Delta T=\sqrt{(\Delta B^2+\Delta S^2)}$.

D'où $\Delta T=1209,5$ mg/L, $\bar{T}=\bar{B}+\bar{S}$, $T=1612 \pm 1209$ mg/L à 95%, très dispersé.

E3 : Population (pauvres / non-pauvres)

Valeurs observées :

$O_{11}=112$	$O_{12}=105$	$O_{13}=154$	$O_{14}=113$
$O_{21}=888$	895	...	
1000	1000	1000	1000

Pauvres sur le pays : $112+105+154+113=484$ sur 4000 personnes au total, d'où 12,1% de pauvres. Nous en déduisons le tableau espéré pour l'hypothèse de répartition uniforme:

$E_{11}=121$	$E_{12}=121$	$E_{13}=121$	$E_{14}=121$
$E_{21}=879$	$E_{22}=879$	$E_{23}=879$	$E_{24}=879$
1000	1000	1000	1000

$$\chi^2 = \frac{(112-121)^2}{121} + \frac{(105-121)^2}{121} + \frac{(154-121)^2}{121} + \frac{(113-121)^2}{121} \simeq 14,0$$

Le nb de degrés de liberté : ddl=(4-1)(2-1) d'où $\chi^2 = 7,815$ au maximum pour que la probabilité que l'hypothèse soit vraie soit supérieure à 5%. Ici on est bien au dessus, **l'hypothèse est donc rejetée** (pour être plus précis, la proba que l'hyp. soit juste n'est que de 0,29%) .

E4 : Eaux minérales (sodium et chlorures) 2)

E5 : Eaux minérales (analyse en composantes principales)

$$\tilde{\bar{x}}_7^2 (x_7^2 - \bar{x}_2) / \sigma_2 \simeq (71 - 1234,8) / 1469 \simeq -0,792.$$

Nous avons réalisé une renormalisation qui permet de comparer à égalité les différentes caractéristiques (grandeurs sans dimension, même centre et même dispersion).

$r_{25} = -0,266$ (corrélation entre sulfates et bicarbonates).

Ensuite la matrice est symétrique : $r_{ij} = r_{ji}$. Et $r_{ii} = 1$.

$\sum_{j=1}^8 \lambda_j = 8 = \text{tr}(r)$. Le pourcentage d'information contenu dans les caractéristiques X'_5 , X'_6 , X'_7 et X'_8 est $\%_{5,6,7,8} = \frac{8 - \lambda_1 - \lambda_2 - \lambda_3 - \lambda_4}{8} = \frac{8 - 7,96}{8} = 0,5\%$ d'information perdue si on les ignore.

Pour obtenir les valeurs propres on résout $\det(r - \lambda I) = 0$.

Les colonnes de la matrice de passage sont les vecteurs propres déterminés en résolvant :

$$r \vec{e}'_j = \lambda_j \vec{e}'_j$$

$$m' = m P \cdot x_2' = -0,154x(-0,305) + 0,044x0,422 + \dots = -0,202.$$

X_1' et X_2' sont les deux premières composantes principales. Elles

contiennent $\%_{1,2} = (4,76+2,05)/8 = 85,1\%$ de l'information du système.

Depuis le cercle des corrélations nous constatons deux groupes d'ions très corrélés :

X_1' représente principalement X_2, X_4, X_6 et X_7 (Bicarbonates / Cl et K / Na -ions alcalins).

X_2' représente principalement X_1, X_3 et X_5 (Ca / Mg -dureté de l'eau -ions alcalino-terreux- et Sulfates).

X_8 se rapproche du centre du cercle des corrélations, ainsi il est peu corrélé à X_1' et X_2' , et joue un rôle faible.

Aussi X_3' est très décorrélé des autres caractéristiques sauf de X_8 (le pH). Comportement intéressant, le pH évolue à part et pourrait être sorti de l'étude. Une interprétation précise peut être faite par un spécialiste en eaux minérales.

Deux remarques :

- $\text{pH} = -\log [\text{H}^+]$, donc le pH n'a pas une relation linéaire avec $[\text{H}^+]$. Il serait donc mieux adapté de prendre $[\text{H}^+]$ et non le pH pour X_8 . Toutes les autres grandeurs étant aussi des ions. Que donnerait l'ACP avec $[\text{H}^+]$ et aurions-nous le même résultat (X_8 décorrélée des autres grandeurs) ? L'étude faite, nous constatons cette déconnexion du pH, ou de H^+ , confirmée.
- la minéralisation (résidu sec) manque dans l'étude. En effet en renormalisant les caractéristiques la quantité totale en ions a disparue. La minéralisation de l'eau correspond globalement à la somme de toutes les concentrations. Après étude, en remplaçant le pH par la minéralisation, nous constatons que cette nouvelle grandeur joue un rôle important et que l'analyse est à reprendre dans son ensemble (par exemple, dans ce cas, les deux composantes principales contiennent à elles seules 96% de l'information et la minéralité est très corrélée aux bicarbonates, chlorures, potassium et sodium).

Devoir Narval :

E1 : Dès

36 possibilités. 12 : $p=1/36=2,8\%$. 7 : $p=6/36=1/6=16,7\%$.

E2 : Maisons

a) b) c) d) $\bar{R}_1=40$ mK/W ; $\sigma_1=1,85$ mK/W ; $\Delta R_1=1,55$ mK/W ; $r_{12}=0$

e) $R=28,6 \pm 0,8$ mK/W

E3 : Salaires

$\chi^2=86,5 \gg 7,8$ hypothèse non soutenable.

E4 : Courbes

1) d); 3) c) ; 4) b); 2) aucunes correspondances.

E5 : Isolants

$\bar{x}_1=-0,625$; $\sigma_1=1,06$; $x_{11}=-0,354$; $r_{12}=0,92$; $\lambda_1=3,0081$

sans X'_3 et X'_4 , 3,4% d'information perdue.

VII. OUTILS MATHÉMATIQUES

Les outils mathématiques utilisés au chapitre 1 sont du niveau lycée. Les dérivées partielles utilisées au chapitre 2 sont de la première année d'enseignement supérieur mais on comprend très vite le lien avec les dérivées du lycée. C'est au chapitre 3, avec l'utilisation des matrices, que nous plongeons dans l'enseignement universitaire. Mon but n'est pas ici de revoir ou d'introduire toutes ces notions, seulement un petit formulaire qui pourra être utile pour résoudre les exercices proposés.

A - Dérivées

1- Définition : $f'(x) = \lim_{\epsilon \rightarrow 0} \left(\frac{f(x+\epsilon) - f(x)}{\epsilon} \right)$

Par exemple si $f(x) = x^2$:

$$(x+\epsilon)^2 - x^2 = x^2 + 2x\epsilon + \epsilon^2 - x^2 \simeq 2x\epsilon \quad \text{et} \quad f'(x) = 2x.$$

Graphiquement la dérivée correspond à la pente de la courbe en un point.

2- Principales dérivées utilisées :

fonction f	dérivée f'	Δf
ax	a (constante)	$\Delta(ax) = a \Delta x$
x^α	$\alpha x^{\alpha-1}$	$\Delta(x^\alpha) = \alpha x^{\alpha-1} \Delta x$
$\sin(x)$	$\cos(x)$	$\Delta(\sin(x)) = \cos(x) \Delta x$

$\cos(x)$	$-\sin(x)$	$\Delta(\cos(x)) = \sin(x) \Delta x$
e^x	e^x	$\Delta(e^x) = e^x \Delta x$
$\ln(x)$	$1/x$	$\Delta(\ln(x)) = \Delta x / x$
$u + v$	$u' + v'$	(u et v fonctions de x)
$u v$	$u' v + v' u$	
$\frac{u}{v}$	$\frac{u' v - v' u}{v^2}$	
$f(g(x))$	$g'(x) f'(g(x))$	

- $\frac{1}{x} = x^{-1}$ donc $(\frac{1}{x})' = (-1)x^{-1-1} = -\frac{1}{x^2}$.
- $\sqrt{x} = x^{\frac{1}{2}}$ donc $(\sqrt{x})' = \frac{1}{2} x^{\frac{1}{2}-1} = \frac{1}{2\sqrt{x}}$.
- $(\sin(x^2))' = (x^2)' \cos(x^2) = 2x \cos(x^2)$

B - Dérivées partielles

Celles-ci concernent les fonctions à plusieurs variables. Par exemple considérons la fonction à trois variables :

$$f(x, y, z) = x^2 - 2z + xy$$

Nous pouvons regarder les variations de cette fonction par rapport à une variable tout en considérant les autres constantes. On procède alors comme pour une dérivée. On a donc :

$$\left(\frac{\partial f}{\partial x}\right)_{y,z} = 2x + y \quad , \quad \left(\frac{\partial f}{\partial y}\right)_{x,z} = x \quad \text{et} \quad \left(\frac{\partial f}{\partial z}\right)_{x,y} = -2$$

On parle de dérivée partielle et par exemple pour la première on dit : "dé rond" f sur "dé rond" x , à (avec) x et y constants.

C - Intégrales

On montre en math que :
$$\int_{x=-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi} .$$

Pour la loi normale standard, nous pouvons vérifier que la moyenne est nulle et que l'écart-type est égal à 1 :

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \text{et} \quad \mu = \int_{-\infty}^{+\infty} x \cdot p(x) dx = 0 \quad \text{car l'intégrale sur}$$

un intervalle symétrique d'une fonction impaire est nulle.

$$\sigma^2 = \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (-x)(-x e^{-\frac{x^2}{2}}) dx \quad \text{d'où}$$

$$\sigma^2 \sqrt{2\pi} = [-x e^{-\frac{x^2}{2}}]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} (-1) e^{-\frac{x^2}{2}} dx = 0 + \sqrt{2} \int_{-\infty}^{+\infty} e^{-x^2} dx$$

d'où l'écart-type égal à l'unité.

Nous avons utilisé une intégration par parties :

$$\boxed{\int_a^b u(x) v'(x) dx = [u(x) v(x)]_a^b - \int_a^b u'(x) v(x) dx}$$

puis un changement de variable $x' = x/\sqrt{2}$.

Nous pouvons aller plus loin, en calculant, le *coefficient de*

dissymétrie $\mu_3 = \int_{-\infty}^{+\infty} x^3 \cdot p(x) dx$, le *coefficient d'aplatissement*

$$\mu_4 = \int_{-\infty}^{+\infty} x^4 \cdot p(x) dx \quad , \quad \text{et, de manière générale, les moments}$$

d'ordre n $\mu_n = \int_{-\infty}^{+\infty} x^n \cdot p(x) dx$. L'ensemble de ces moments

permettent de caractériser une distribution.

Pour une gaussienne $\mu_3 = 0$ (symétrique). Si ce coefficient est négatif la courbe s'étale vers la gauche, s'il est positif la courbe s'étale vers la droite. Pour une gaussienne $\mu_4 = 3$. Si ce coefficient est inférieur à 3 la courbe est plus aplatie qu'une gaussienne. Pour une loi binomiale : $\mu_4 = 3 - 8/n$.

VIII. Bibliographie / Sources / Logiciels / Illustration de couverture

Livres

[iv] PROTASSOV Konstantin. *Probabilités et incertitudes dans l'analyse des données expérimentales*. Presses Universitaires de Grenoble, 1999. 128 p.

[vi] WONNACOTT. *Statistique Économie - Gestion - Sciences - Médecine*. Economica, 1972. 919 p.

[iii] SAPORTA Gilbert. *Probabilités, analyse des données et statistique*. Technip, 2006. 622 p.

Sites

Placez *http://* devant le nom des sites.

La plupart des fichiers sont en copie dans le dossier

<www.incertitudes.fr/proba-stat-acp/>.

ROUAUD Mathieu. *Calculs d'incertitudes*. <www.incertitudes.fr/Incertitudes.html>

[viii] MARCADET Françoise. *Estimer une incertitude et Mesures et incertitudes (accompagnement personnalisé en 1^{ère}S)*. Professeur de Sciences Physiques appartenant au groupe académique. Lycée Alain Fournier de Bourges. Sur le site <physique.ac-orleans-tours.fr/accompagnement_personnalise/>

[v] *Nombres, mesures et incertitudes*. mai 2010.
<http://www.educnet.education.fr/rnchimie/recom/mesures_incertitudes.pdf>

[vii] *Évaluation des données de mesure — Guide pour l'expression de l'incertitude de mesure.* <www.bipm.org/utills/common/documents/jcgm/JCGM_100_2008_F.pdf>, 2008. 132p.

BREUIL P, DI BENEDETTO D. *Incertitude et étalonnage.*2000. 16p.

[i] <tice.univ-nc.nc/~garrigos/tp001.pdf>, page 7 ,issu d'un article en anglais , K.S.KRANE et L.SCHECTER American J. of Physics 50,82 (1982).

Articles

DUBY C., ROBIN .S. *Analyse en Composantes Principales.* <www.agroparistech.fr/IMG/pdf/ACP2006.pdf>

[ii] ROUAUD Mathieu. *Mesure avec une règle.* Bulletin de l'Union des Physiciens n°913, avril 2009.

Logiciels

Tous les logiciels utilisés sont libres et gratuits.

- Traitement de texte et tableur *OpenOffice*
- Graphisme : *Gimp* (points), *Inkscape* (vectoriel) et *Blender* (3D).
- Calcul : *Scilab* (numérique), *XCAS* (symbolique) et programmes en *PHP* (sur serveur).
- Tracé de courbes : *KmPlot*.
- Système d'exploitation : *Ubuntu*.

Illustration de couverture

Représentations 3D de la couverture de devant et celle située avant la table des matières.

Un marcheur aléatoire se déplace dans un plan. Il lance deux pièces et regarde les résultats : deux piles, un pile et un face, un face et un pile ou deux faces. La première pièce lui indique s'il doit faire un premier pas vers l'Est ou l'Ouest, la deuxième s'il doit faire un second pas vers le Sud ou le Nord. Ainsi à chaque intervalle de temps Δt il se déplace dans le plan de deux pas. Pendant Δt la distance parcourue par le marcheur est $\Delta d = 2p$ (p longueur d'un pas).

À qu'elle distance du point de départ se trouve le marcheur à l'instant t ? (après n intervalles de temps : $t = n \Delta t$)

Pour $n=1$, nous traçons les tableaux suivants :

0,5	1	1
-0,5	1	1
	-0,5	0,5

25%	25%
25%	25%

Le centre du tableau est son point de départ. En abscisse x (direction Est-Ouest) le déplacement est de plus ou moins un pas ($x=\pm p$, $p=\Delta d/2$ et on a pris $\Delta d=1$). De même en ordonnée : $y=\pm p$. Dans chaque est indiqué le nombre de possibilités de se retrouver à cet endroit. Le second tableau indique les probabilités ($1/4$, $4=2 \times 2$).

Dans les quatre possibilités il est à une distance $\sqrt{2}/2$ du point d'origine. Et, en terme d'écart-type, la distance caractéristique est

$$s = \sqrt{(4 \cdot 1/\sqrt{2})^2 / (4-1)} \approx 0,816 \quad .$$

Pour $n=2$, nous traçons les tableaux suivants :

1	1	2	1
0	2	4	2
-1	1	2	1
	-1	0	1

6%	13%	6%
13%	25%	13%
6%	13%	6%

Par exemple pour être en $(x=0 ; y=-1)$, il y a deux chemins possibles : (PF,PP) et (FP,PP). Écart-type $s_2 \approx 1,033$.

Le marcheur a une chance sur quatre d'être revenu au point de départ.

Pour $n=3$, nous traçons les tableaux suivants :

1,5	1	3	3	1
0,5	3	9	9	3
-0,5	3	9	9	3
-1,5	1	3	3	1
	-1,5	-0,5	0,5	1,5

2%	5%	5%	2%
5%	14%	14%	5%
5%	14%	14%	5%
2%	5%	5%	2%

$s_3 \approx 1,234$

Pour $n=4$, nous traçons les tableaux suivants :

2	1	4	6	4	1
1	4	16	24	16	4
0	6	24	36	24	6
-1	4	16	24	16	4
-2	1	4	6	4	1
	-2	-1	0	1	2

0,4%	1,6%	2,3%	1,6%	0,4%
1,6%	6,3%	9,4%	6,3%	1,6%
2,3%	9,4%	14%	9,4%	2,3%
1,6%	6,3%	9,4%	6,3%	1,6%
0,4%	1,6%	2,3%	1,6%	0,4%

Pour $n=5$, nous traçons les tableaux suivants :

2,5	1	5	10	10	5	1
1,5	5	25	50	50	25	5
0,5	10	50	100	100	50	10
-0,5	10	50	100	100	50	10
-1,5	5	25	50	50	25	5
-2,5	1	5	10	10	5	1
	-2,5	-1,5	-0,5	0,5	1,5	2,5

0,1%	0,5%	1,0%	1,0%	0,5%	0,1%
0,5%	2,4%	4,9%	4,9%	2,4%	0,5%
1,0%	4,9%	9,8%	9,8%	4,9%	1,0%
1,0%	4,9%	9,8%	9,8%	4,9%	1,0%
0,5%	2,4%	4,9%	4,9%	2,4%	0,5%
0,1%	0,5%	1,0%	1,0%	0,5%	0,1%

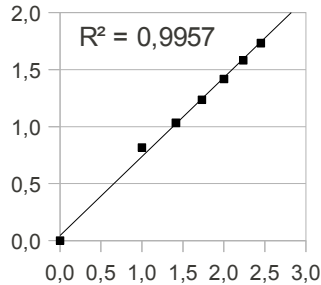
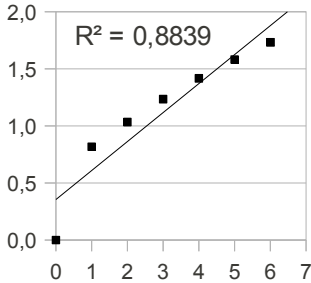
Pour $n=6$, nous traçons les tableaux suivants :

3	1	6	15	20	15	6	1	0,0%	0,1%	0,4%	0,5%	0,4%	0,1%	0,0%
2	6	36	90	120	90	36	6	0,1%	0,9%	2,2%	2,9%	2,2%	0,9%	0,1%
1	15	90	225	300	225	90	15	0,4%	2,2%	5,5%	7,3%	5,5%	2,2%	0,4%
0	20	120	300	400	300	120	20	0,5%	2,9%	7,3%	9,8%	7,3%	2,9%	0,5%
-1	15	90	225	300	225	90	15	0,4%	2,2%	5,5%	7,3%	5,5%	2,2%	0,4%
-2	6	36	90	120	90	36	6	0,1%	0,9%	2,2%	2,9%	2,2%	0,9%	0,1%
-3	1	6	15	20	15	6	1	0,0%	0,1%	0,4%	0,5%	0,4%	0,1%	0,0%
	-3	-2	-1	0	1	2	3							

D'où l'évolution de la distance quadratique moyenne du point d'origine en fonction du temps :

t (Δt)	0	1	2	3	4	5	6
\sqrt{t}	0,00	1,00	1,41	1,73	2,00	2,24	2,45
s (2p)	0,000	0,816	1,033	1,234	1,417	1,582	1,732

Nous traçons la courbe les points pour s en fonction de t , puis celle pour s en fonction de \sqrt{t} :



Nous constatons une bien meilleure corrélation en \sqrt{t} . En effet nous voyions directement que la distance à l'origine n'évoluait pas proportionnellement au temps, pour $n=2$ nous sommes à environ une unité de l'origine, nous devrions donc être vers 3 pour $n=6$.

Cette variation en \sqrt{t} est caractéristique des phénomènes de diffusion et trouve ici son analogie avec les compensations d'erreurs en \sqrt{n} .

IX. TABLES / Index

A. Coefficients de Student

Coefficient de Student t		Confiance (%)								
		50	80	90	95	98	99	99,5	99,8	99,9
Degrés de liberté (taille de l'échantillon moins une unité)	1	1,00	3,08	6,31	12,7	31,8	63,7	127	318	637
	2	0,82	1,89	2,92	4,30	6,96	9,92	14,1	22,3	31,6
	3	0,76	1,64	2,35	3,18	4,54	5,84	7,45	10,2	12,9
	4	0,74	1,53	2,13	2,78	3,75	4,60	5,60	7,17	8,61
	5	0,73	1,48	2,02	2,57	3,36	4,03	4,77	5,89	6,87
	6	0,72	1,44	1,94	2,45	3,14	3,71	4,32	5,21	5,96
	7	0,71	1,41	1,89	2,36	3,00	3,50	4,03	4,79	5,41
	8	0,71	1,40	1,86	2,31	2,90	3,36	3,83	4,50	5,04
	9	0,70	1,38	1,83	2,26	2,82	3,25	3,69	4,30	4,78
	10	0,70	1,37	1,81	2,23	2,76	3,17	3,58	4,14	4,59
	11	0,70	1,36	1,80	2,20	2,72	3,11	3,50	4,02	4,44
	12	0,70	1,36	1,78	2,18	2,68	3,05	3,43	3,93	4,32
	13	0,69	1,35	1,77	2,16	2,65	3,01	3,37	3,85	4,22
	14	0,69	1,35	1,76	2,14	2,62	2,98	3,33	3,79	4,14
	15	0,69	1,34	1,75	2,13	2,60	2,95	3,29	3,73	4,07
	16	0,69	1,34	1,75	2,12	2,58	2,92	3,25	3,69	4,01
	17	0,69	1,33	1,74	2,11	2,57	2,90	3,22	3,65	3,97
	18	0,69	1,33	1,73	2,10	2,55	2,88	3,20	3,61	3,92
	19	0,69	1,33	1,73	2,09	2,54	2,86	3,17	3,58	3,88
	20	0,69	1,33	1,72	2,09	2,53	2,85	3,15	3,55	3,85
	22	0,69	1,32	1,72	2,07	2,51	2,82	3,12	3,50	3,79
	24	0,68	1,32	1,71	2,06	2,49	2,80	3,09	3,47	3,75
	26	0,68	1,31	1,71	2,06	2,48	2,78	3,07	3,43	3,71
	28	0,68	1,31	1,70	2,05	2,47	2,76	3,05	3,41	3,67
	30	0,68	1,31	1,70	2,04	2,46	2,75	3,03	3,39	3,65
	40	0,68	1,30	1,68	2,02	2,42	2,70	2,97	3,31	3,55
	50	0,68	1,30	1,68	2,01	2,40	2,68	2,94	3,26	3,50
	60	0,68	1,30	1,67	2,00	2,39	2,66	2,91	3,23	3,46
	70	0,68	1,29	1,67	1,99	2,38	2,65	2,90	3,21	3,44
	80	0,68	1,29	1,66	1,99	2,37	2,64	2,89	3,20	3,42
90	0,68	1,29	1,66	1,99	2,37	2,63	2,88	3,18	3,40	
100	0,68	1,29	1,66	1,98	2,36	2,63	2,87	3,17	3,39	
200	0,68	1,29	1,65	1,97	2,35	2,60	2,84	3,13	3,34	
300	0,68	1,28	1,65	1,97	2,34	2,59	2,83	3,12	3,32	
500	0,67	1,28	1,65	1,96	2,33	2,59	2,82	3,11	3,31	
1000	0,67	1,28	1,65	1,96	2,33	2,58	2,81	3,10	3,30	
∞	0,67	1,28	1,64	1,96	2,33	2,58	2,81	3,09	3,29	

B. Valeurs critiques de Khi-deux

χ^2	Probabilité pour que l'hypothèse soit vraie (%)										
	99	90	75	50	30	20	10	5	2	1	0,1
1	0,000157	0,0158	0,102	0,455	1,07	1,64	2,71	3,84	5,41	6,63	10,8
2	0,0201	0,211	0,575	1,39	2,41	3,22	4,61	5,99	7,82	9,21	13,8
3	0,115	0,584	1,21	2,37	3,66	4,64	6,25	7,81	9,84	11,3	16,3
4	0,297	1,06	1,92	3,36	4,88	5,99	7,78	9,49	11,7	13,3	18,5
5	0,554	1,61	2,67	4,35	6,06	7,29	9,24	11,1	13,4	15,1	20,5
6	0,872	2,20	3,45	5,35	7,23	8,56	10,6	12,6	15,0	16,8	22,5
7	1,24	2,83	4,25	6,35	8,38	9,80	12,0	14,1	16,6	18,5	24,3
8	1,65	3,49	5,07	7,34	9,52	11,0	13,4	15,5	18,2	20,1	26,1
9	2,09	4,17	5,90	8,34	10,7	12,2	14,7	16,9	19,7	21,7	27,9
10	2,56	4,87	6,74	9,34	11,8	13,4	16,0	18,3	21,2	23,2	29,6
11	3,05	5,58	7,58	10,3	12,9	14,6	17,3	19,7	22,6	24,7	31,3
12	3,57	6,30	8,44	11,3	14,0	15,8	18,5	21,0	24,1	26,2	32,9
13	4,11	7,04	9,30	12,3	15,1	17,0	19,8	22,4	25,5	27,7	34,5
14	4,66	7,79	10,2	13,3	16,2	18,2	21,1	23,7	26,9	29,1	36,1
15	5,23	8,55	11,0	14,3	17,3	19,3	22,3	25,0	28,3	30,6	37,7
16	5,81	9,31	11,9	15,3	18,4	20,5	23,5	26,3	29,6	32,0	39,3
17	6,41	10,1	12,8	16,3	19,5	21,6	24,8	27,6	31,0	33,4	40,8
18	7,01	10,9	13,7	17,3	20,6	22,8	26,0	28,9	32,3	34,8	42,3
19	7,63	11,7	14,6	18,3	21,7	23,9	27,2	30,1	33,7	36,2	43,8
20	8,26	12,4	15,5	19,3	22,8	25,0	28,4	31,4	35,0	37,6	45,3
21	8,90	13,2	16,3	20,3	23,9	26,2	29,6	32,7	36,3	38,9	46,8
22	9,54	14,0	17,2	21,3	24,9	27,3	30,8	33,9	37,7	40,3	48,3
23	10,2	14,8	18,1	22,3	26,0	28,4	32,0	35,2	39,0	41,6	49,7
24	10,9	15,7	19,0	23,3	27,1	29,6	33,2	36,4	40,3	43,0	51,2
25	11,5	16,5	19,9	24,3	28,2	30,7	34,4	37,7	41,6	44,3	52,6
26	12,2	17,3	20,8	25,3	29,2	31,8	35,6	38,9	42,9	45,6	54,1
27	12,9	18,1	21,7	26,3	30,3	32,9	36,7	40,1	44,1	47,0	55,5
28	13,6	18,9	22,7	27,3	31,4	34,0	37,9	41,3	45,4	48,3	56,9
29	14,3	19,8	23,6	28,3	32,5	35,1	39,1	42,6	46,7	49,6	58,3
30	15,0	20,6	24,5	29,3	33,5	36,3	40,3	43,8	48,0	50,9	59,7
31	15,7	21,4	25,4	30,3	34,6	37,4	41,4	45,0	49,2	52,2	61,1
32	16,4	22,3	26,3	31,3	35,7	38,5	42,6	46,2	50,5	53,5	62,5
33	17,1	23,1	27,2	32,3	36,7	39,6	43,7	47,4	51,7	54,8	63,9
34	17,8	24,0	28,1	33,3	37,8	40,7	44,9	48,6	53,0	56,1	65,2
35	18,5	24,8	29,1	34,3	38,9	41,8	46,1	49,8	54,2	57,3	66,6
36	19,2	25,6	30,0	35,3	39,9	42,9	47,2	51,0	55,5	58,6	68,0
37	20,0	26,5	30,9	36,3	41,0	44,0	48,4	52,2	56,7	59,9	69,3
38	20,7	27,3	31,8	37,3	42,0	45,1	49,5	53,4	58,0	61,2	70,7
39	21,4	28,2	32,7	38,3	43,1	46,2	50,7	54,6	59,2	62,4	72,1
40	22,2	29,1	33,7	39,3	44,2	47,3	51,8	55,8	60,4	63,7	73,4

Index

- Analyse en composantes
principales.....67
- Approches multicritères.....67
- Calcul d'incertitude.....43
- Capacité thermique.....1
- Cercle des corrélations.....81
- Classe.....**6**, 32
- Coefficient d'aplatissement 166
- Coefficient de corrélation.....37
- Coefficient de dissymétrie..166
- Coefficient de Student...**12**, 50,
172
- Compensations.....44, 100
- Conductivité thermique.....61
- Confiance.....12
- Densité de probabilité.....**18**, 94
- Dérivée partielle.....49, 165
- DÉS.....15, 33, 97
- Diffusion.....171
- Dispersion.....109
- Distribution de Gauss.....10, **18**
- Droites extrêmes.....51
- Eaux minérales.....120
- Écart.....110
- Écart moyen.....3, 31
- Écart quadratique moyen.....4
- Écart-type.....3, 109
- Équation différentielle.....62
- Erreur accidentelle.....27, 109
- Erreur aléatoire...103, 109, 110
- Erreur de discrétisation..27, 92,
109
- Erreur de mesure.....110
- Erreur de type A.....109
- Erreur de type B.....109
- Erreur systématique.....27, 103,
109, 110
- Espace des individus.....72
- Estimation.....51
- Étendue.....3, 94
- Facteur d'élargissement.....109
- Fidélité.....27
- Focométrie.....59, 64, 111
- Formule de Cauchy.....60
- Formule de propagation des
écart-types.....42, 102
- Formule de propagation des
incertitudes.....42, 103
- Fréquence.....6
- Gaz parfait.....55, 113
- Hypothèse ergodique.....115
- Incertaince.....12
- Incertaince absolue.....**12**, 45
- Incertaince relative.....**12**, 45
- Incertaince-type.....109

Indice.....	59	Nuages de points.....	39
Intégrale.....	19, 166	Ordonnée à l'origine.....	49
Intervalle de confiance.....	51	Parfaite.....	103
Intervalle de prédiction.....	52	Pente.....	49
Isolant.....	32, 67, 130	Physique statistique.....	114
Justesse.....	27, 109	Pile électrique.....	64
Loi de probabilité triangulaire	97, 104	Prédiction.....	52
Loi de probabilité uniforme..	94	Prisme.....	60
Macroétat.....	114	Produit de convolution..	92, 96
Matrice de corrélation.....	75	Régression avec barres d'erreurs.....	54
Matrice de passage.....	78	Régression linéaire.....	48
Matrice des données.....	71	Répétabilité.....	27, 109
Matrice des données normalisée.....	72	Reproductibilité.....	27, 109
Matrice réduite.....	84	Résidu.....	49
Médiane.....	2	Résistance thermique...61, 127	
Mesurage.....	109	Résolution.....	27, 92, 109
Mesurande.....	109	Scilab.....	89
Méthode par pondération.....	67	Spectre de raie.....	60
Métrologie.....	104	Test d'hypothèse.....	22
Microétat.....	114	Test du Khi-deux.....	24
Mode.....	1	Théorème central limite.....	10, 100
Moindres carrés.....	48, 54	Valeur vraie.....	109
Moment.....	166	Valeurs propres.....	76
Montage longue dérivation.	152	Variance.....	95, 100
Moyenne.....	2	Vecteurs propres.....	76
Moyenne arithmétique.....	2	Zéro absolu.....	52
Moyenne géométrique.....	2		

Ce livre s'adresse à tout public intéressé par les sciences expérimentales et les mathématiques. Le champ d'application déborde largement sur des thèmes très présents dans notre société : les sondages, les modélisations économiques et l'aide à la décision. Par exemple l'analyse en composantes principales est un outil très utilisé.

Le fil directeur suit l'enchaînement d'un cours donné en école d'ingénieur, ainsi ce livre pourra être très utile à l'étudiant tout comme à l'enseignant. Aussi, je souhaite que l'expérimentateur trouve tous les outils nécessaires à des calculs d'incertitudes simples et rigoureux.

En effet, la science s'efforce de lier les phénomènes naturels à une logique mathématique. Cette quête d'une vérité et d'une cohérence globale nécessite d'exercer un esprit critique qui s'appuie sur des mesures accompagnées de leurs incertitudes.

3 1415926535897932384626433832795028841971693
 99375 105820974944592307816406286208998628034
 8253421 1706798214808651328230664709384460955
 058223172535940812848 1117450284102701938521

$$\chi^2 = \sum_{(i,j)} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$\det(r - \lambda I) = 0$$

$$\Delta a = t_{n-2} S a$$

$$\Delta f^2 = \sum_{j=1}^p \left[\left(\frac{\partial f}{\partial x_j} \right)^2 \Delta x_j^2 \right]$$

$$r \vec{e}_j = \lambda_j \vec{e}_j$$

$$w_i = \frac{1}{(\Delta y_i)^2 + (a \Delta x_i)^2}$$

$$p(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\mu = \bar{x} \pm t \frac{s}{\sqrt{n}}$$

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

