

CPH 316

Méthodes de la chimie physique

Analyse des données

Andrzej Lasia

Département de chimie
Université de Sherbrooke

2012

1. Introduction	3
1.1 Chiffres significatifs	3
2. Mesures d'erreurs	4
3. Types d'erreurs:	5
4. Distribution d'erreurs	5
4.1 Courbe de Gauss	5
4.2 L'intégration numérique	18
4.3 L'écart type de la population et d'une petite série des données	20
4.4 L'écart type de la valeur vraie ou la moyenne	21
4.5 Intervalle (limite) de confiance	22
4.5.1 Intervalles de confiance quand σ est connue.	22
4.5.2 Intervalle de confiance quand σ est inconnue	23
4.5.3 L'écart-type d'un ensemble des données (« pooling » ou observations couplées)	24
4.6 La moyenne pondérée	27
5. Propagation d'erreurs	29
5.1 L'erreur maximale	30
6. La méthode des moindres carrés pour une ligne droite	33
6.1 Propriétés de la méthode des moindres carrés	36
6.2 Écarts types et intervalles de confiance de \hat{y}_i calculée:	37
6.3 Écarts types et intervalles de confiance de y_i (expérimentale)	38
6.4 Coefficient de corrélation	39
6.4.1 Coefficient de détermination	39
6.5 La méthode de moindres carrés pour $y = ax$	42
6.6 Erreur de x_c de la ligne de régression	45
6.7 Méthode matricielle de moindres carrés	46
6.8 La méthode des moindres carrés pondérée	48
7. Tests statistiques pour une moyenne	48
7.1 Introduction	48
7.2 Test χ^2	49
7.3 Test Q	50
7.4 Test u	51
7.5 Test t	52
7.5.1 Comparaison avec le standard	52
7.5.2 Comparaison de deux moyennes	52
7.6 Test F	54
8. Tests statistiques pour une régression	56
8.1 Rejet d'un point dans la régression	56
8.2 Importance des paramètres de la régression	57
8.2.1 Test t d'importance des paramètres de la régression	57
8.2.2 Test F d'importance des paramètres de la régression	58
8.3 Analyse des variances	59
9. Lissage des courbes expérimentales	70
9.1 Filtres numériques simples	70
9.2 Méthode de Savitzky-Golay	72
10. Transformée de Fourier rapide (Fast Fourier Transform, FFT)	83
10.1 Théorie	83

10.2	Applications	90
10.3	« Aliasing »	96
10.4	« Leakage »	99
11.	Exercices avec Maple	109
12.	Fonctions statistiques d'Excel	120
	Fonction statistiques d'Excel	120
	Fonctions complexes	124
13.	Références	125
14.	Annexe	126
14.1	Table de la distribution intégrale normale (Gauss), $A_G(z)$	
	126	
14.2	Table des valeurs $z(\alpha)$	
	127	
14.3	Table de la distribution de Student, $t(\alpha, k)$ pour $\alpha = 0.1, 0.05$ et 0.01	
	127	
14.4	Table de la distribution F, $F(\alpha, D2, D1)$ pour $\alpha = 0.05$	
	128	
14.5	Table de la distribution $\chi^2(\alpha, k)$	
	129	
14.6	Valeurs de Q pour le rejet d'un point	
	129	

1. Introduction

1.1 Chiffres significatifs

Tous les résultats sont obtenus avec des erreurs. Le nombre des chiffres significatifs doit correspondre à la précision des résultats. Dans les mesures physico-chimiques ou analytiques on détermine un écart-type ou les limites de confiance (voir ci-dessous).

Une règle générale : **On fait tous les calculs avec une grande précision et on arrondit à la fin.**

Voici des exemples :

Arrondissements

37.56 → 37.6

37.54 → 37.5

37.65 → ?

Pour éviter une accumulation d'erreurs on arrondit en haut ou en bas. Si le dernier chiffre est 5 et avant dernier est paire on arrondit en bas et quand l'avant dernier chiffre est impair on arrondit en haut.

37.65 → 37.6

37.35 → 37.4

23.4 3 chiffres significatifs

12.40 4 chiffres significatifs

0.002 1 chiffre significatif

0.0027 2 chiffres significatifs

0.00270 3 chiffres significatifs

240 ??

2.4×10^2 2 chiffres significatifs

2.40×10^2 3 chiffres significatifs

Multiplication/division

Conserver le nombre des chiffres significatifs correspondants au chiffre le moins précis.

$7.643 \times 15.3 = 116.93794$ chiffres significatifs \times 3 chiffres significatifs = 3 chiffres significatifs
= 117

$7.8933 \times 15 = 118.3995$ 2 chiffres significatifs = 1.2×10^2

$68.233^2 = 4655.7423$ 5 chiffres significatifs = 4655.7

Addition/soustraction

Conserver les nombres de places décimales correspondant au terme le moins précis.

$386.0 + 67.241 = 453.241 \approx 453.2$

386	
67.241	
1.32	
+ 64.5	
<hr style="border: 1px solid black;"/>	
516.421 \approx 516	

2. Mesures d'erreurs

La statistique a pour objet de permettre de tirer des conclusions de données observées.

Les données expérimentales possèdent toujours des erreurs. Il y a deux mesures d'erreurs principales :

1) l'**exactitude**: différence entre la valeur mesurée et la valeur vraie:

une **erreur absolue** : $x_i - x_{\text{vrai}}$

une **erreur relative** : $(x_i - x_{\text{vrai}})/x_{\text{vrai}}$

2) la **précision**: caractérise la reproductibilité des résultats, si on répète la même mesure plusieurs fois de la même façon, (reliée a la distribution aléatoire de l'erreur)

l'**écart type** (la *déviatiion standard*), la *variance*

Mesures d'erreurs:

Précision reproductibilité, σ , σ^2

Exactitude une erreur absolue, $x_i - x_v$

une erreur relative. $\frac{x_i - x_v}{x_v}$

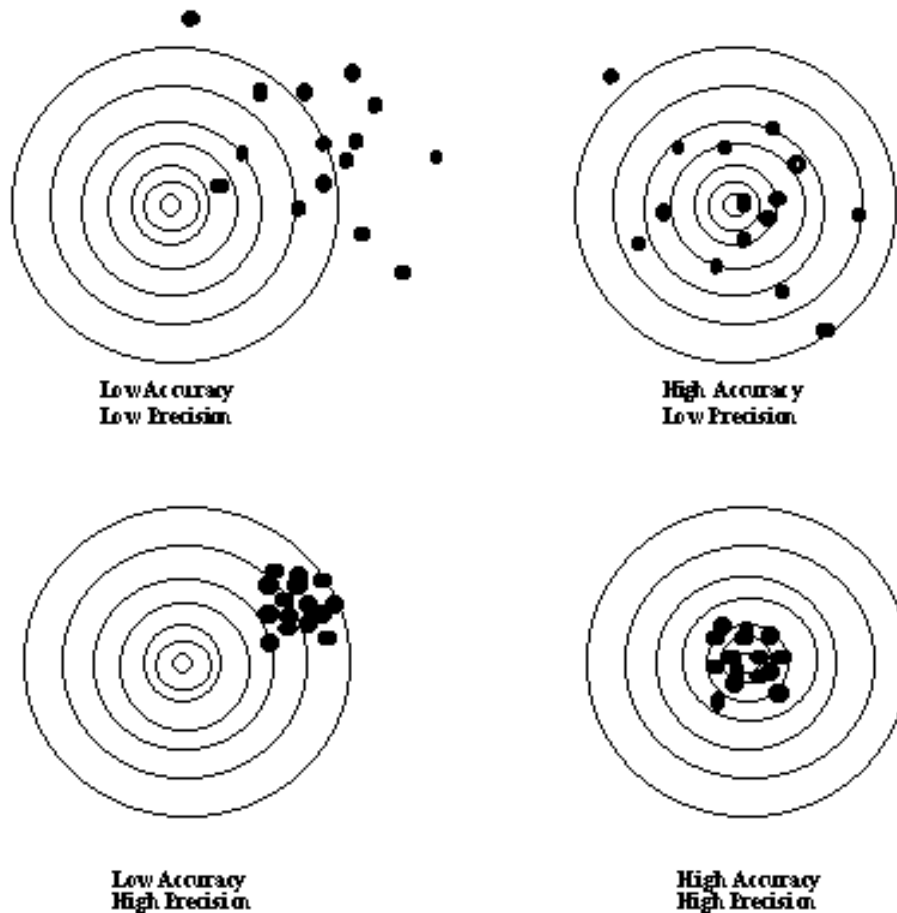


Figure 1. L'illustration de la précision et de l'exactitude.

3. Types d'erreurs:

I. systématiques, elles ont une cause déterminable et tendent à être reproductibles d'une mesure à une autre et elles peuvent être corrigées

I. 1. *erreurs d'instrumentation* : p. ex. une vieille pile, une grande résistance des contacts électriques, etc.

I. 2. *erreurs de la méthode* : causées par un comportement non idéal des réactifs, p. ex. une réaction chimique trop lente, une contamination, la décomposition des réactifs, des interférences chimiques

I.3 *erreurs d'opérateur* : p. ex. une erreur d'échelle, pas de réglage de zéro, etc.

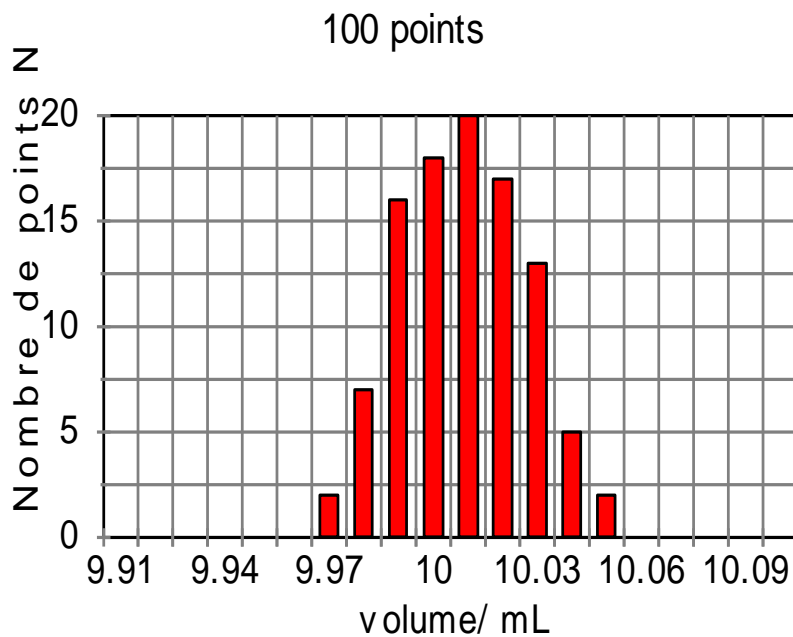
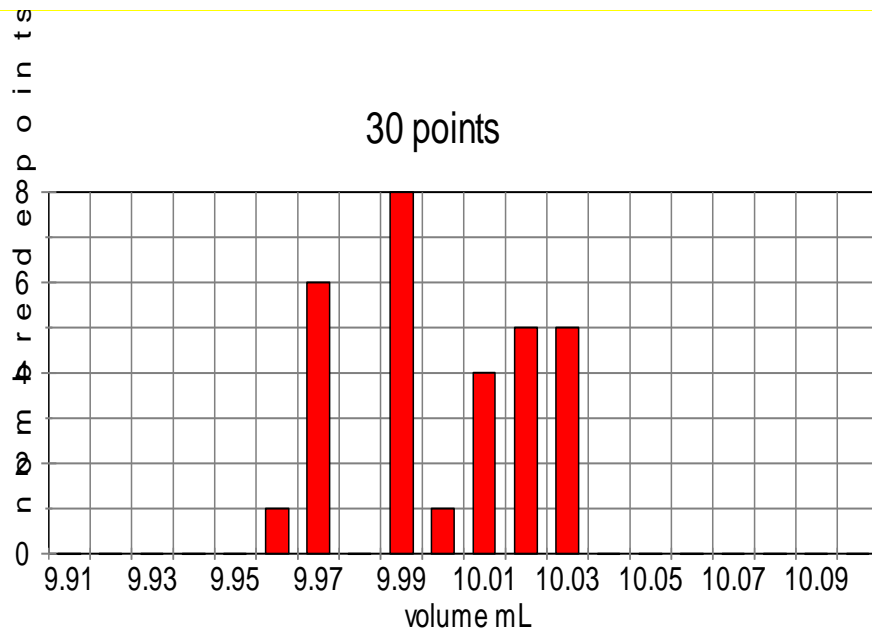
II. aléatoires, indéterminables, positives ou négatives causées par des fluctuations aléatoires, le bruit, etc., pas d'une cause.

4. Distribution d'erreurs

4.1 Courbe de Gauss

Population : Ensemble complet des données ou de toutes les valeurs possibles (mesures sur un groupe de personnes ou d'objets, en générale infinie); p. ex. la population des habitants du pays, ensemble des objets produits par une usine,

Échantillon : Série limitée de valeurs observées, tirées d'une façon aléatoire de la population. Quand on répète une mesure plusieurs fois on peut tracer la fréquence avec laquelle chaque valeur est obtenue en fonction de la différence entre la valeur mesurée et la valeur vraie. Les résultats obtenus sont illustrés dans la Figure 2.



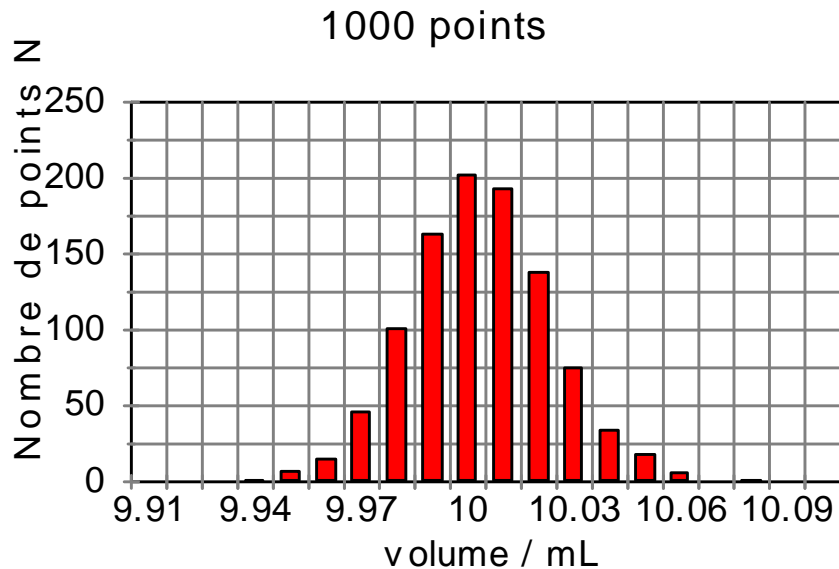


Figure 2. Distribution des valeurs de volume d'une pipette de 10 ml en fonction de nombre des mesures (30, 100 et 1000 mesures).

Quand le nombre de mesures $\rightarrow \infty$, cette courbe s'appelle la *courbe de Gauss*. Elle décrit la distribution normale et représente la densité de la probabilité en fonction de $x_i - x_{vraie}$. La densité de probabilité, P , est la probabilité qu'une valeur mesurée se trouve entre x_i et $x_i + dx$: $y = P(x_i, x_i + dx)$.

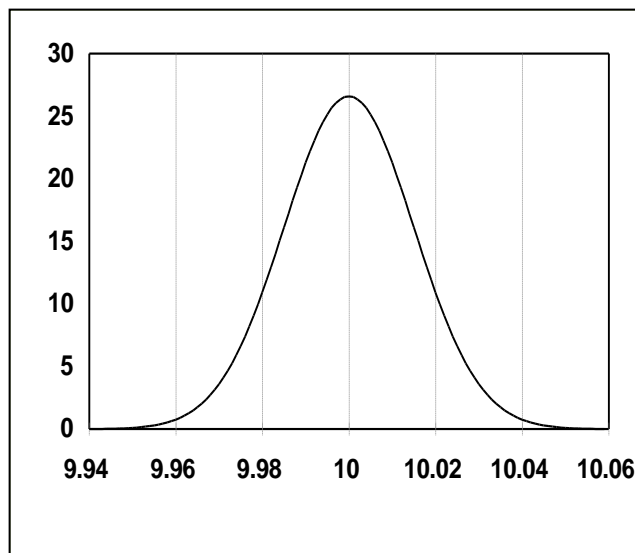


Figure 3. Distribution des valeurs de volume d'une pipette de 10 mL en fonction de nombre des mesures lorsque le nombre des mesures $N \rightarrow \infty$.

Les courbes gaussiennes non-normalisées et normalisées sont décrites comme :

$$P_G(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \quad P_G(z, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad (1)$$

où

$$z = \frac{(x_i - \mu)}{\sigma} \quad (2)$$

est une variable réduite, σ est l'**écart type de la population** et $\mu = x_{\text{vraie}}$, la **valeur vraie**, qui peut être estimée par la moyenne de la population. En utilisant variable z on peut transformer une distribution $P_G(x, \mu, \sigma)$ en $P_G(z, 0, 1)$, c'est à dire une distribution avec $\mu = 0$ et $\sigma = 1$.

La valeur vraie μ peut être estimée par une valeur moyenne des grands nombres des mesures N :

$$\mu = \lim_{N \rightarrow \infty} \left(\frac{x_1 + x_2 + \dots + x_N}{N} \right) = \frac{\sum_{i=1}^N x_i}{N} = \bar{x} \quad (3)$$

L'**écart type (la déviation standard) de la population** ($N \rightarrow \infty$, en pratique quand $N > 30$) est définie comme:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}} \quad (4)$$

Pour une distribution normale qui est caractérisée par la valeur vrai « μ » et l'écart type de la population « σ », 68.3% de la population se trouve dans les limites $\pm 1\sigma$ autour de la moyenne, 95.5% de la population se trouve dans les limites $\pm 2\sigma$ et 99.7% dans les limites $\pm 3\sigma$ autour de la moyenne.

L'aire de surface sous la courbe gaussienne:

$$\bar{x} \pm \sigma \quad 68.3\%$$

$$\bar{x} \pm 2\sigma \quad 95.5\%$$

$$\bar{x} \pm 3\sigma \quad 99.7\%$$

Cela signifie qu'il y a 95.5 chances sur 100 que la valeur vraie se trouve dans la zone $\bar{x} \pm 2\sigma$.

Le **coefficient de variation** (l'**écart type relative**), (RSD « relative standard déviation » ou CV « coefficient of variation » en anglais) :

$$CV = \frac{\sigma}{\mu} 100\% \approx \frac{s}{x} 100\% \quad (5)$$

La propriété de la moyenne est que la somme des déviations de chaque mesure de la moyenne est zéro :

$$\sum_{i=1}^N (x_i - \bar{x}) = 0 \quad \sigma^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N} \quad (6)$$

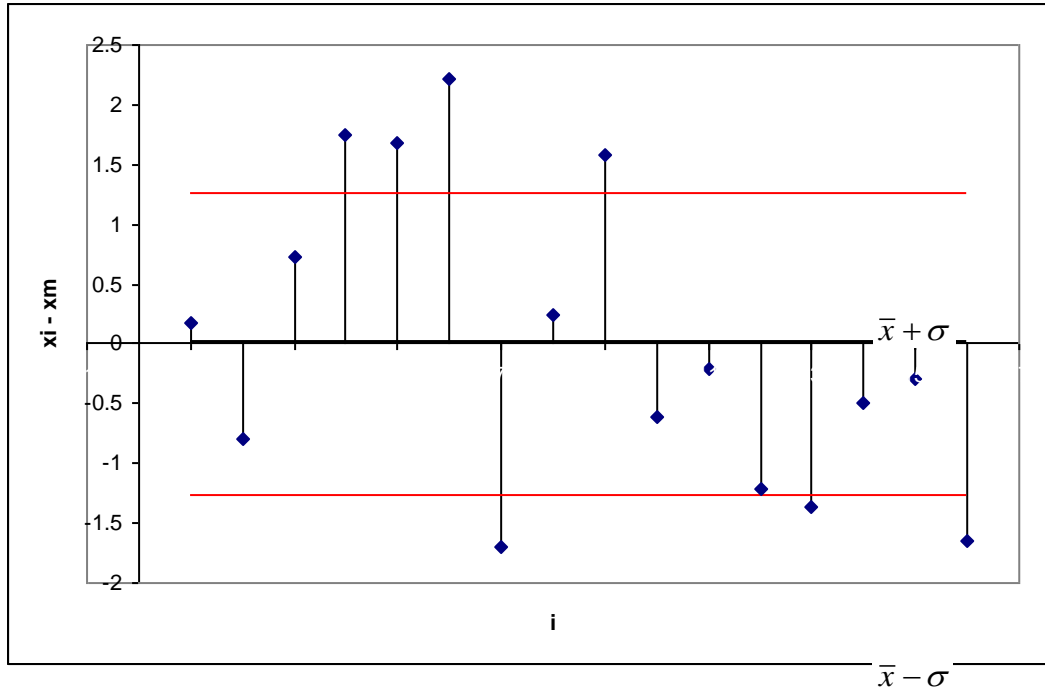


Figure 4. Propriétés de la moyenne.

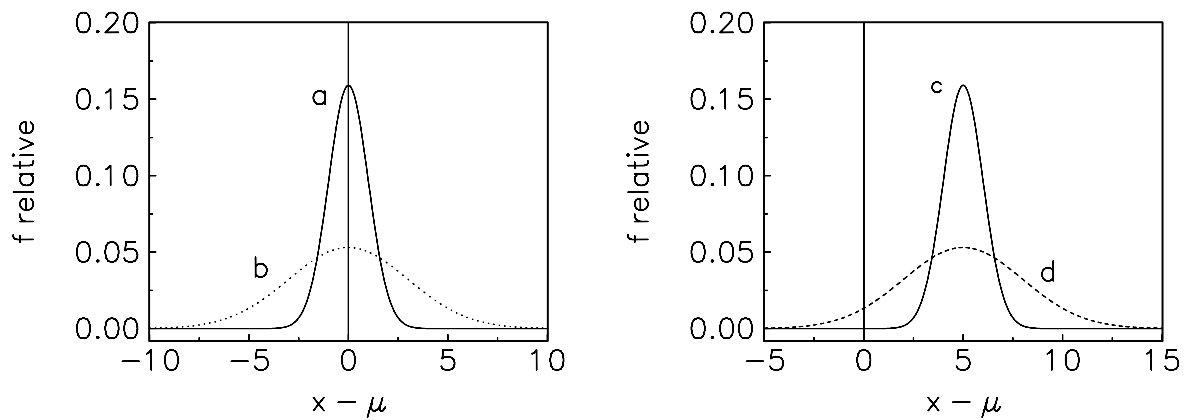


Figure 5. Courbes gaussiennes: a) bonne précision et exactitude, b) mauvaise précision et bonne exactitude, c) bonne précision et mauvaise exactitude, d) mauvaise précision et exactitude.

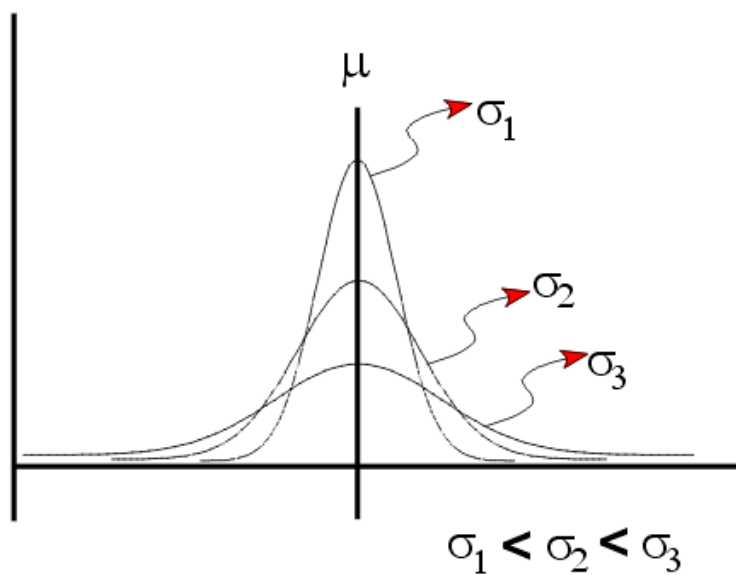


Figure 6.
des différentes valeurs de σ .

Courbes de Gauss pour

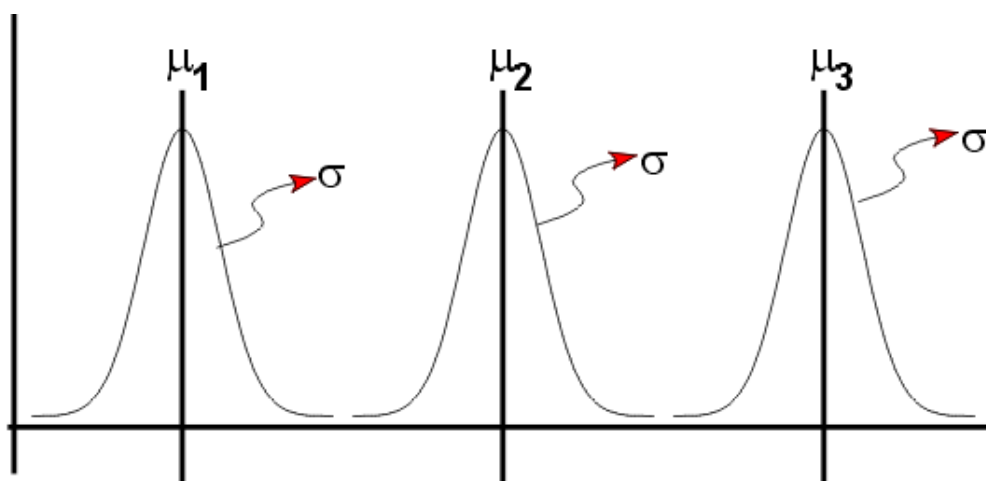


Figure 7. Courbes de Gauss pour des différentes valeurs de μ .

Les fonctions de Gauss peuvent être calculées en utilisant l'Excel :

$$P_G(x, \mu, \sigma) = \text{LOI.NORMALE}(x, \mu, \sigma, \text{FAUX}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

(7)

ou

$$P_G(x, \mu, \sigma) = \text{LOI.NORMALE}(z, 0, 1, \text{FAUX}) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right] \quad (8)$$

La fonction d'Excel LOI.NORMALE peut donner deux différentes variables (la fonction gaussienne et son intégrale) et la valeur du paramètre logique indique laquelle de deux est calculée. La valeur « FAUX » indique que la valeur de la fonction gaussienne est calculée.

Intégration de la fonction gaussienne

L'Excel permet de trouver l'intégrale de la fonction gaussienne :

$$P = \text{LOI.NORMALE}(x, \mu, \sigma, \text{VRAI}) = \int_{-\infty}^x P_G(x, \mu, \sigma) dx \quad (9)$$

où P est la probabilité de trouver une valeur entre $-\infty$ et x et la valeur logique « VRAI » indique qu'on calcule l'intégral de la fonction gaussienne. Les propriétés de la distribution gaussienne :

1. La distribution normale est continue;
2. Elle s'étend de $-\infty \leq x \leq +\infty$;
3. La fonction de Gauss est normalisée :

$$\int_{-\infty}^{\infty} P_G(x, \mu, \sigma) dx = 1 \quad (10)$$

Dans les tables statistiques la valeur de la fonction de répartition est habituellement présentée.

Probabilité intégrale (fonction de répartition)

Ce qui nous intéresse souvent est la probabilité qu'une mesure aléatoire sera située dans les limites $\mu \pm z \sigma$.

On peut donc intégrer cette fonction entre ces limites :

$$A_G = \int_{\mu-z\sigma}^{\mu+z\sigma} dP_G(x, \mu, \sigma) = \int_{\mu-z\sigma}^{\mu+z\sigma} P_G(x, \mu, \sigma) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu-z\sigma}^{\mu+z\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \quad (11)$$

Cette intégrale s'appelle la probabilité intégrale A_G . En utilisant la variable réduite z on peut écrire :

$$A_G(|z|, 0, 1) = \int_{-z}^z P_G(z, 0, 1) dz = \frac{1}{\sqrt{2\pi}} \int_{-z}^z e^{-\frac{z^2}{2}} dz \quad (12)$$

Les valeurs de A_G sont tabulées dans l'Annexe 12.1. On peut les calculer en utilisant l'Excel :

$$A_G(z) = \text{LOI.NORMALE}(z,0,1,\text{VRAI}) - \text{LOI.NORMALE}(-z,0,1,\text{VRAI}) =$$

$$= \int_{-\infty}^z P_G(z,0,1) dz - \int_{-\infty}^{-z} P_G(z,0,1) dz = \int_{-z}^z P_G(z,0,1) dz = \frac{1}{\sqrt{2\pi}} \int_{-z}^z e^{-\frac{z^2}{2}} dz$$

$$A_G(x) = \text{LOI.NORMALE}(x,\mu,\sigma,\text{VRAI}) - \text{LOI.NORMALE}(-x,\mu,\sigma,\text{VRAI}) = \int_{\mu-z\sigma}^{\mu+z\sigma} P_G(x,\mu,\sigma) dx$$

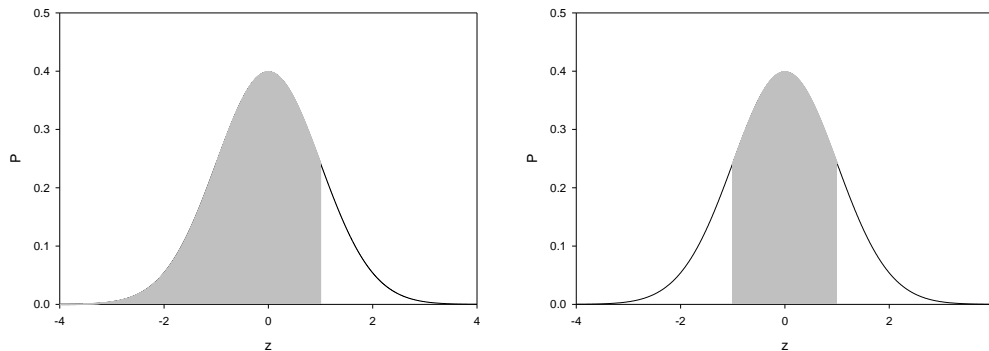


Figure 8. L'illustration de l'intégrale d'Excel et de la fonction de répartition.

Applications

Exemple 1

Quelle est la probabilité d'obtenir une mesure x déviant de la moyenne μ d'un écart type : $x = \mu \pm \sigma$?

$$z = \frac{x - \mu}{\sigma} = \frac{\mu \pm \sigma - \mu}{\sigma} = \pm 1.0$$

D'après les tables, pour $z = 1.0$, $A_G = 0.6827$. La probabilité est de 68.3%.

Exemple 2

On veut que $x = \mu \pm 2\sigma$:

$$z = \frac{x - \mu}{\sigma} = \frac{\mu \pm 2\sigma - \mu}{\sigma} = \pm 2.0$$

$A_G = 0.9545$, $P = 95.4\%$

Pour $x = \mu \pm 3\sigma$, $z = 3.0$, $P = 99.7\%$.

Exemple 3

Les données suivent la distribution normale avec $\mu = 25$, $\sigma = 5$. Trouvez :

a) $P(x \geq 20)$

On cherche une intégrale de 20 à ∞ :

$$P(x \geq 20) = \int_{20}^{\infty} P_G(x, 25, 5) dx$$

Au lieu de travailler avec une courbe non-normalisée $P_G(x, \mu, \sigma)$ on peut travailler avec une courbe normalisée $P_G(z, 0, 1)$ avec la moyenne $\mu = 0$ et l'écart type $\sigma = 1$ en utilisant la transformation : $z = (x - \mu) / \sigma$. Les deux courbes ont une aire de surface identique:

$$P(x \geq 20) = \int_{20}^{\infty} P_G(x, 25, 5) dx = P(z \geq -1) = \int_{-1}^{\infty} P_G(z, 0, 1) dz$$

$$z = \frac{20 - 25}{5} = -1.0, A_G(1, 0, 1) = 0.68$$

$$A_g(z, 0, 1) = \int_{-z}^z P_G(u, 0, 1) du$$

$$A_1 = \frac{1}{2} \int_{-z}^z P_G(u, 0, 1) du = \frac{0.68}{2} = 0.34$$

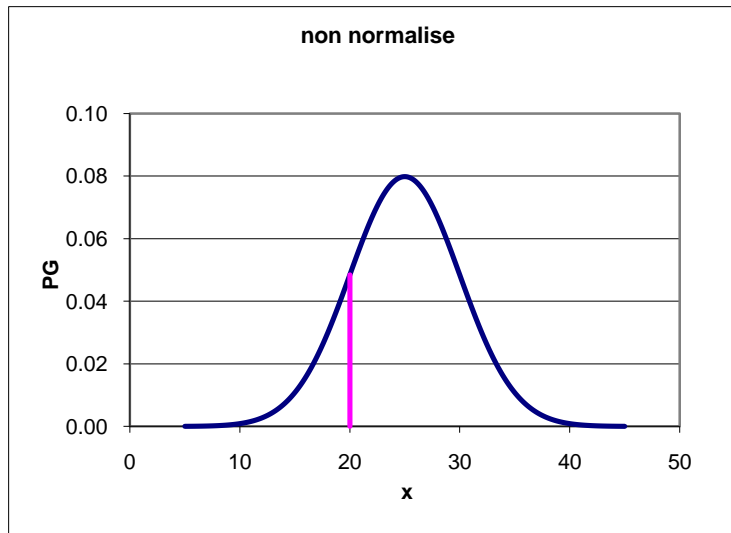
$$A_2 = 0.5$$

$$A = 0.34 + 0.5 = 0.84$$

Cependant, il est très facile de calculer en utilisant l'Excel :

$$A = 1 - \int_{-\infty}^{20} P_G(x, 25, 5) dx = 1 - 0.15866 = 0.84134$$

$$A = 1 - \int_{-\infty}^{-1} P_G(z, 0, 1) dz = 1 - 0.15866 = 0.84134$$

$P_G(x, 25, 5)$


$$\int_{-z}^z P_G(z, 0, 1) dz$$

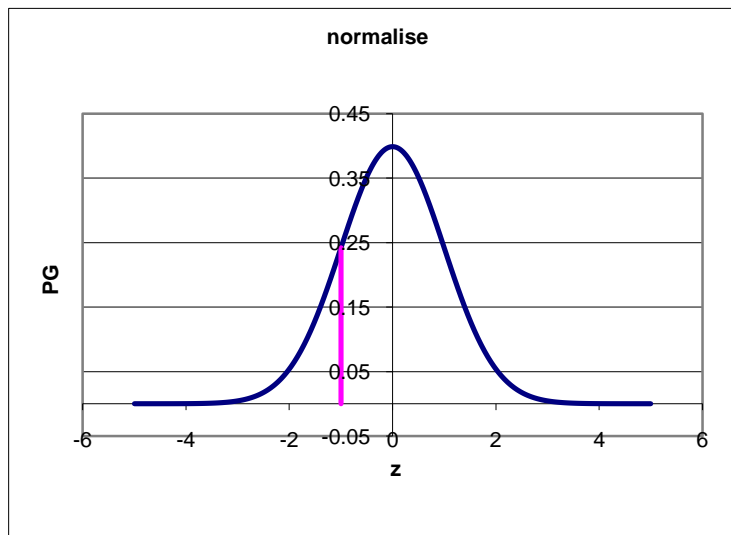


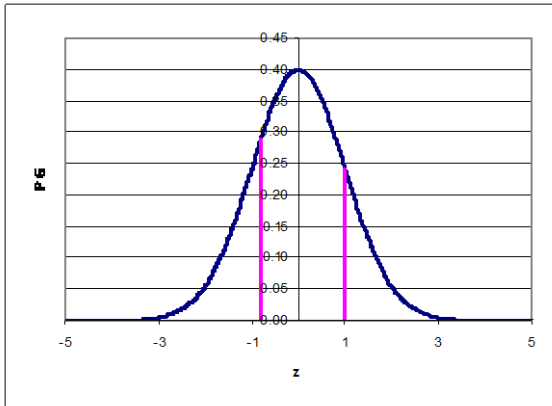
Figure 9. Distribution normale non normalisée et normalisée.

b) $P(x < 40)$

$$z = \frac{40 - 25}{5} = 3$$

$$A = \int_{-\infty}^3 P_G(u, 0, 1) du = \frac{A_G(3, 0, 1)}{2} + 0.5 = \frac{0.9973}{2} + 0.5 = 0.99865$$

$$A = \int_{-\infty}^{40} P_G(x, 25, 5) dx = \int_{-\infty}^3 P_G(z, 0, 1) dz = 0.99865$$

c) $P(21 \leq x \leq 30)$ 

$$z_1 = (21-25)/5 = -0.8; \quad z_2 = (30-25)/5 = 1$$

$$A = \int_{-0.8}^1 P_G(u, 0, 1) du = \frac{A_G(z=0.8)}{2} + \frac{A_G(z=1)}{2} = 0.28814 + 0.34134 = 0.62948$$

$$\begin{aligned} A &= \int_{-0.8}^1 P_G(u, 0, 1) du = \int_{-\infty}^1 P_G(u, 0, 1) du - \int_{-\infty}^{-0.8} P_G(u, 0, 1) du = \\ &= \int_{21}^{30} P_G(x, 25, 5) dx = \int_{-\infty}^{30} P_G(x, 25, 5) dx - \int_{-\infty}^{21} P_G(x, 25, 5) dx \end{aligned}$$

d) $P(18 \leq x \leq 23)$

$$z_1 = (18-25)/5 = -1.4; \quad z_2 = (23-25)/5 = -0.4$$

$$\begin{aligned} P &= \int_{-1.4}^{-0.4} P_G(u, 1, 0) du = \frac{A_G(z=1.4)}{2} - \frac{A_G(z=0.4)}{2} = \frac{0.83849}{2} - \frac{0.31084}{2} = 0.41924 - 0.15542 \\ &= 0.26382 \end{aligned}$$

$$A = \int_{-\infty}^{-0.4} P_G(z, 0, 1) dz - \int_{-\infty}^{-1.4} P_G(z, 0, 1) dz = \int_{-\infty}^{23} P_G(x, 25, 5) dx - \int_{-\infty}^{18} P_G(x, 25, 5) dx$$

Fonctions d'Excel :

$$\text{LOI.NORMALE}(z, 0, 1, \text{VRAI}) = \int_{-\infty}^z P_G(u, 0, 1) du$$

$$\text{LOI.NORMALE}(x, \mu, \sigma, \text{VRAI}) = \int_{-\infty}^x P_G(u, \mu, \sigma) du$$

$$A_G(z) = \text{LOI.NORMALE}(z, 0, 1, \text{VRAI}) - \text{LOI.NORMALE}(-z, 0, 1, \text{VRAI})$$

$$A_G(x) = \text{LOI.NORMALE}(x, \mu, \sigma, \text{VRAI}) - \text{LOI.NORMALE}(-x, \mu, \sigma, \text{VRAI})$$

Exemple 4

La moyenne d'un cours, mesurée pendant plusieurs années, est 65 et $\sigma = 15$.

Estimez :

- 1) % des étudiants avec la moyenne $x \geq 85$
- 2) % des étudiants avec la moyenne $x \leq 50$
- 3) % des étudiants avec la moyenne $x \geq 60$
- 4) 20% des étudiants auront la moyenne inférieure ou égale à quelle valeur?

Re. 1.

Méthode de la probabilité intégrale :

Transformation dans la distribution normalisée :

$$z = (85-65)/15 = 1.333$$

$$P = \int_{1.33}^{\infty} P_G(z, 0, 1) dz = \int_0^{\infty} P_G(z, 0, 1) dz - \int_0^{1.33} P_G(z, 0, 1) dz$$

$$0.5 - \frac{A_G(z=1.33)}{2} = 0.5 - \frac{0.8165}{2} = 0.092$$

la moyenne $x \geq 85$ 9.2% de population

Méthode d'intégrale (Excel), fonction normalisée :

$$P(x \geq 85) = 1 - \int_{-\infty}^{1.33} P_G(x, 0, 1) dx = 1 - \text{LOI.NORMALE}(1.33, 0, 1, \text{VRAI}) =$$

$$= 1 - 0.9082 = 0.092$$

fonction non normalisée

$$P(x \geq 85) = \int_{85}^{\infty} P(x, 65, 15) dx = 1 - \int_{-\infty}^{85} P(x, 65, 15) dx = 1 - \text{LOI.NORMALE}(85, 65, 15, \text{VRAI}) =$$

$$= 1 - 0.9082 = 0.092$$

Re. 2.

$$z = (50-65)/15 = -1$$

$$P = \int_{-\infty}^{-1} P_G(z, 0, 1) dz = 0.5 - \frac{A_G(z=1)}{2} = 0.5 - \frac{0.6827}{2} = 0.5 - 0.341 = 0.159$$

la moyenne $x \leq 50$ 15.9%

$$A = \int_{-\infty}^{50} P_G(x, 65, 15) dx = \text{LOI.NORMALE}(50, 65, 15, \text{VRAI}) = 0.159$$

$$A = \int_{-\infty}^{-1} P_G(z, 0, 1) dz = \text{LOI.NORMALE}(-1, 0, 1, \text{VRAI}) = 0.159$$

Re. 3.

$$z = (60 - 65) / 15 = -0.333$$

$$A = \int_{-0.333}^{\infty} P_G(z, 0, 1) dz = \int_{-0.333}^0 P_G(z, 0, 1) dz + \int_0^{\infty} P_G(z, 0, 1) dz =$$

$$\frac{A_G(z = 0.333)}{2} + 0.5 = \frac{0.2586}{2} + 0.5 = 0.129 + 0.5 = 0.629$$

la moyenne $x \geq 60$ 62.9%

$$P(x \geq 60) = 1 - \int_{-\infty}^{60} P_G(x, 65, 15) dx = 1 - \text{LOI.NORMALE}(60, 65, 15, \text{VRAI}) = 0.630$$

$$P(z \geq -0.333) = 1 - \int_{-\infty}^{-0.333} P_G(z, 0, 1) dz = 1 - \text{LOI.NORMALE}(-0.333, 0, 1, \text{VRAI}) = 0.630$$

Re. 4.

$$\int_{-\infty}^z P_G(z, 0, 1) dz = 0.20$$

$$z = -0.84, \quad z = \frac{x - \mu}{\sigma} \quad -0.84 = \frac{x - 65}{15} \quad x = 52.4$$

ou

$$0.5 - 0.2 = 0.3$$

$$A_G(z) = 0.60 \quad |z| = 0.84 \quad z = -0.84$$

20% des étudiants auront la moyenne inférieure ou égale à 52.4

Méthode d'Excel

$$\text{LOI.NORMALE.STANDARD.INVERSE}(\alpha) = z(\alpha)$$

$$\int_{-\infty}^z P_G(z, 0, 1) dz = \alpha$$

ou

$$\text{LOI.NORMALE.INVERSE}(\alpha, \mu, \sigma) = x(\alpha)$$

$$\int_{-\infty}^x P_G(x, \mu, \sigma) dx = \alpha$$

LOI.NORMALE.STANDARD.INVERSE(0.2) = -0.841

LOI.NORMALE.INVERSE(α, μ, σ) = LOI.NORMALE.INVERSE(0.2, 65, 15) = 52.4

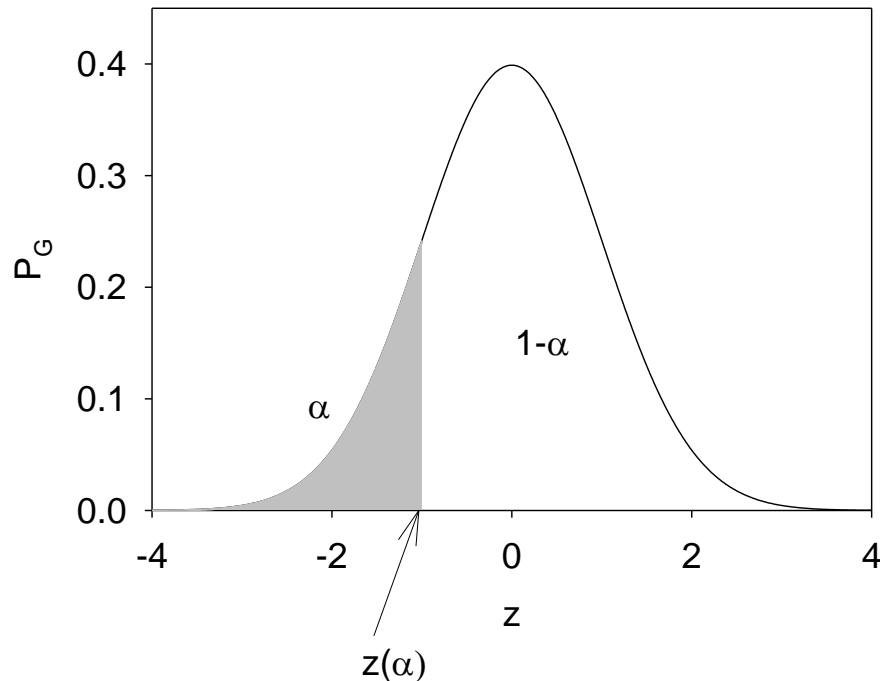


Figure 10. Illustrations de la fonction LOI.NORMALE.STANDARD.INVERSE(α) qui nous donne la valeur $z(\alpha)$.

4.2 L'intégration numérique

Le but de la méthode est d'intégrer une fonction donnée numériquement comme une série des points x_i, y_i . Ces valeurs sont obtenue d'habitude avec un appareil d'acquisition des données.

Il existent différentes méthodes d'intégration numérique des données expérimentales. Considérons une série de données qui sont enregistrées aux intervalles identiques, par exemple :

x	y
0.5	0.25
1.0	1.00
1.5	2.25
2.0	4.00
2.5	6.25
3.0	9.00

Méthode des rectangles

La méthode de rectangles considère construction des rectangles à chaque point en débutant par le premier point, comme dans le graphique ci-dessous. Dans ce cas on approxime l'aire sous la courbe par la somme des aires de rectangles. Bien sur, dans ce cas l'aire des rectangles est plus petite que l'aire sous la courbe, mais cette erreur diminue avec la diminution de Δx .

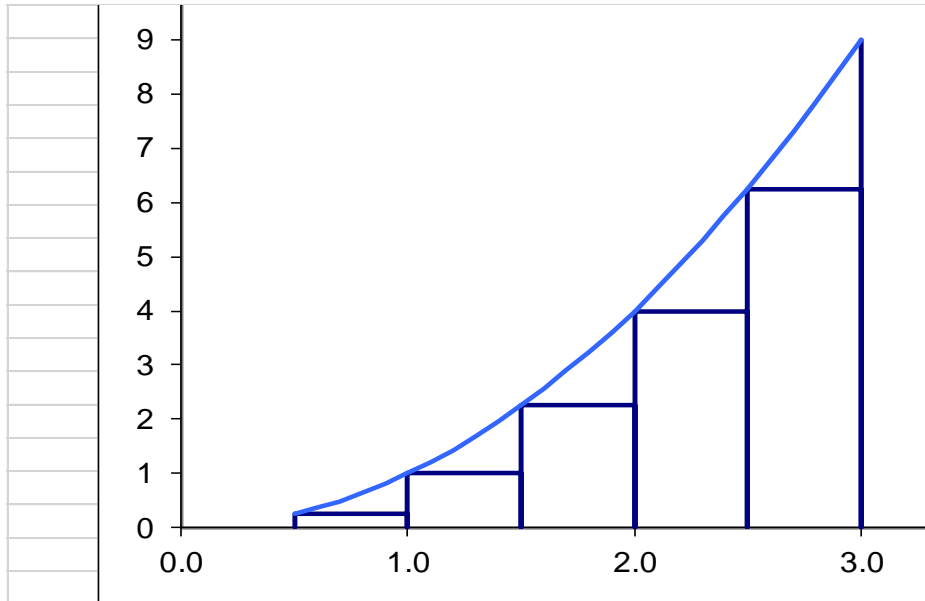


Figure 11. Intégration par la méthode des rectangles. On remplace l'aire sous la courbe par la somme des aires de rectangles.

L'aire totale des rectangles est :

$$\int_{x_1}^{x_N} f(x)dx = y_1\Delta x + y_2\Delta x + \dots + y_{N-1}\Delta x = \Delta x \sum_{i=1}^{N-1} y_i$$

ou dans le cas particulier :

$$\int_{0.5}^3 f(x)dx = (0.25 + 1 + 2.25 + 4 + 6.25)0.5 = 6.875$$

L'intégral de la courbe calculée analytiquement est 8.958333, donc l'erreur de la méthode des rectangles est dans ce cas particulier: -23.3%. Cependant; quand on diminue Δx à 0.1, l'erreur diminue à -4.8%.

La méthode des trapèzes

La méthode des rectangles n'est pas très précise. Une meilleur approximation de l'aire de surface peut être obtenue en utilisant la méthode des trapèzes. Dans ce cas on construit un trapèze entre chaque deux points. L'aire d'un trapèze est égale à $A = \Delta x(y_1 + y_2)/2$. Donc, l'aire sous la courbe est :

$$\int_{x_1}^{x_N} f(x)dx = \left[\frac{y_1 + y_2}{2} + \frac{y_2 + y_3}{2} + \dots + \frac{y_{N-1} + y_N}{2} \right] \Delta x = \left[\frac{y_1}{2} + \sum_{i=2}^{N-1} y_i + \frac{y_N}{2} \right] \Delta x$$

Dans notre cas particulier on peut écrire :

$$\int_{0.5}^3 f(x)dx = \left(\frac{0.25}{2} + 1 + 2.25 + 4 + 6.25 + \frac{9}{2} \right) 0.5 = 9.0625$$

Le graphique correspondant est montré ci-dessous.

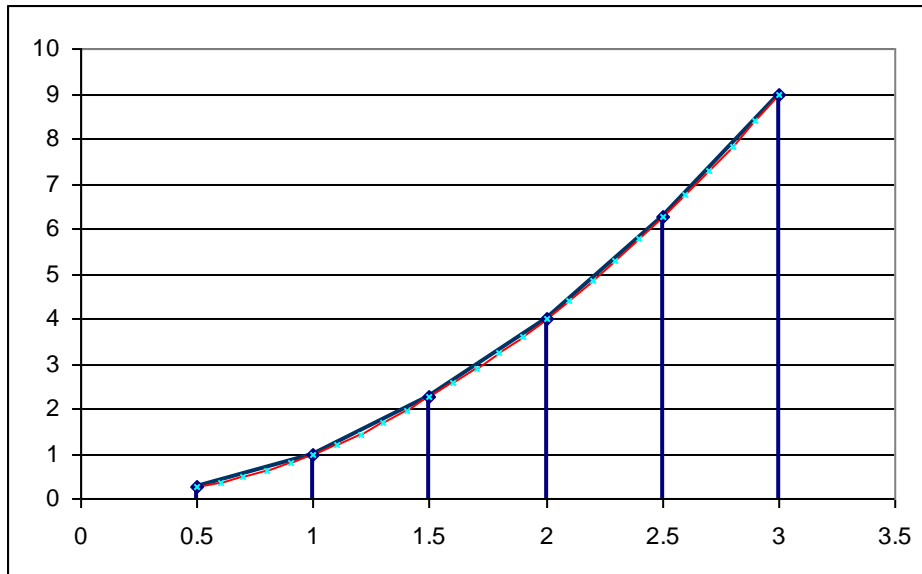


Figure 12. Intégration par la méthode des trapèzes. On remplace l'aire sous la courbe par la somme l'aire des trapèzes.

Méthode des trapèzes donne l'erreur de 1.2%. En diminuant Δx à 0.1 diminue cette erreur à 0.046%. On voit, que la méthode des trapèzes est beaucoup plus précise. Il existent d'autres méthodes qui utilisent approximation locale par une parabole, polynôme cubique, etc.

La méthode des paraboles

On peut approximer une fonction qui passe par trois points par une parabole. L'aire de surface sous une parabole est donnée par :

$$\int_{x_1}^{x_3} f(x)dx = \left(\frac{y_1 + 4y_2 + y_3}{3} \right) \Delta x$$

et on peut développer une formule pour la somme des aires des paraboles adjacentes :

$$\int_{x_1}^{x_N} f(x)dx = \left(\frac{y_1 + 4y_2 + 2y_3 + 4y_4 + 2y_5 + 4y_6 + \dots + y_N}{3} \right) \Delta x$$

pour $N = 3 + 2i$ points.

4.3 L'écart type de la population et d'une petite série des données

Le but de la méthode statistique est de déterminer la moyenne, l'écart type, l'écart type de la moyenne et les limites de confiance. Il y a deux méthodes de calculer l'écart type :

- 1) Quand on prend beaucoup des mesures (en pratique $N \geq 30$) on peut estimer la **valeur vrai**, μ , et l'**écart type de la population**, σ :

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}} \quad (13)$$

où la somme des déviations des mesures individuelles de la valeur vrai au carrée est divisée par le nombre d'observations N .

- 2) En général $\bar{x} \neq \mu$ mais $\bar{x} \rightarrow \mu$ quand $N \rightarrow \infty$. Quand $N < 30$ on ne peut pas déterminer de σ . On peut déterminer la **écart type d'une petite série des données, s** :

$$s = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}} = \sqrt{\frac{\sum x_i^2 - (\sum x_i)^2 / N}{N-1}} \quad (14)$$

La valeur s est plus grande que σ . Dans l'équation (14) on a le nombre de degrés de liberté $N-1$ au lieu de N parce qu'une équation a été utilisée pour la détermination de la moyenne (le nombre de degré de liberté = $N-1$).

L'écart type peut être calculée dans l'Excel comme fonction : ECARTYPE.

4.4 L'écart type de la valeur vraie ou la moyenne

L'écart type de la valeur vraie ou la moyenne (μ, \bar{x}) est plus petite que la valeur d'une mesure (σ, s) :

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}} \quad s_{\bar{x}} = \frac{s_x}{\sqrt{N}} \quad (15)$$

Les valeurs de la moyenne, l'écart type et l'écart type de la moyenne sont illustrées ci-dessus.

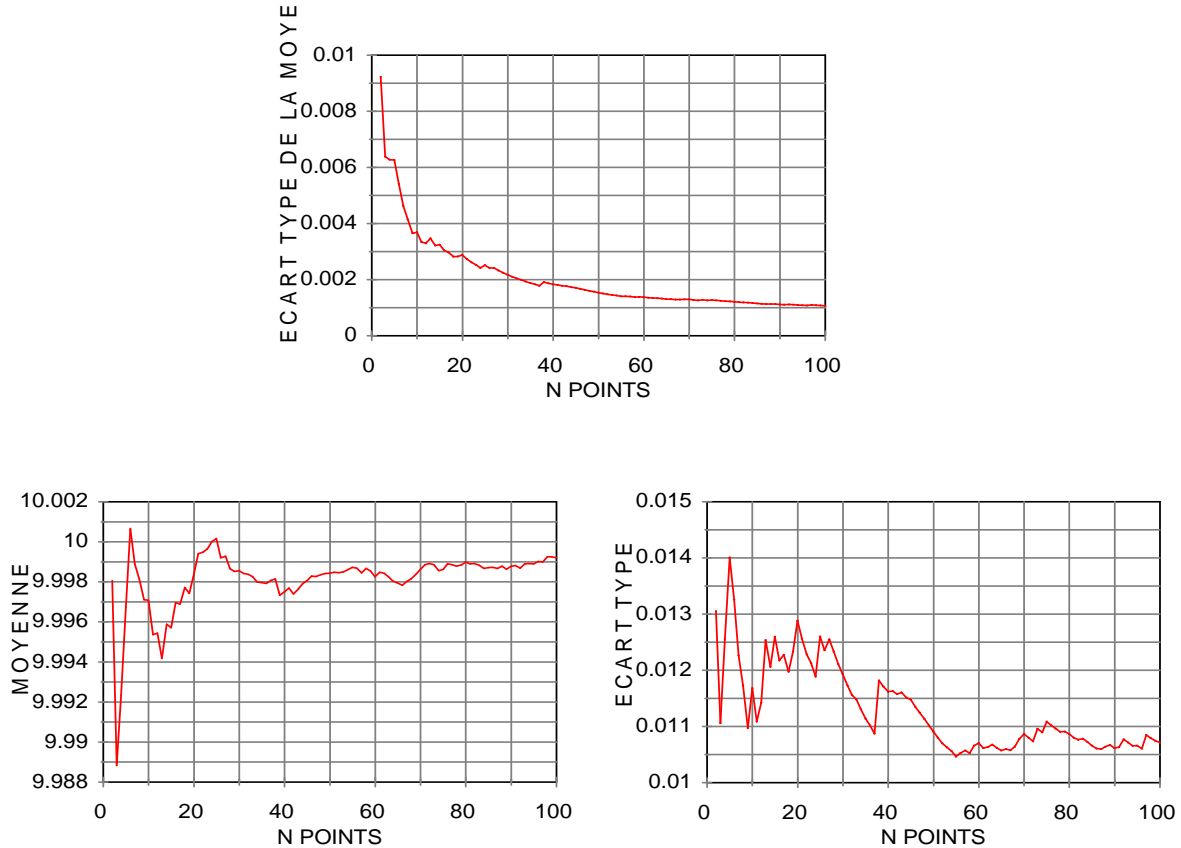


Figure 13. Dépendance de la moyenne, de l'écart type et de l'écart type de la moyenne en fonction de nombre des mesures.

4.5 Intervalle (limite) de confiance

Les valeurs de x_i sont dispersées autour d'une valeur moyenne, \bar{x} . En utilisant des méthodes statistiques on peut spécifier une intervalle autour de \bar{x} dans lequel on veut être confiant de situer μ . Cet intervalle s'appelle l'*intervalle* (ou *limite*) *de confiance*. Donc, on peut déterminer avec certaine probabilité que la valeur moyenne se trouve entre valeurs limites: $a \geq \bar{x} \geq b$.

4.5.1 Intervalles de confiance quand σ est connue.

Quand on fait plus que 30 mesures, la valeur s est une bonne estimation de σ . Dans ce cas on utilise la **distribution de Gauss** les intervalles de confiance sont définis comme:

$$IC(\mu) = x_i \pm z(\alpha)\sigma \quad (\text{pour une mesure})$$

$$\text{ou } IC(\mu) = \bar{x} \pm z(\alpha) \frac{\sigma}{\sqrt{N}} = \bar{x} \pm z(\alpha)\sigma_{\bar{x}} \quad (\text{pour } N \text{ mesures}) \quad (16)$$

où $z(\alpha)$ est définie par l'équation (12) et donné dans la Table en Annexe 14.2 et α est un *niveau* (ou *degré*) *de confiance*. Dans ce cas on utilise la distribution normale (de Gauss); . Entre $-z$ et z il y a $(1-\alpha)\%$ de l'aire totale sous la courbe de Gauss. Par exemple, la degré de confiance de 95% ($\alpha =$

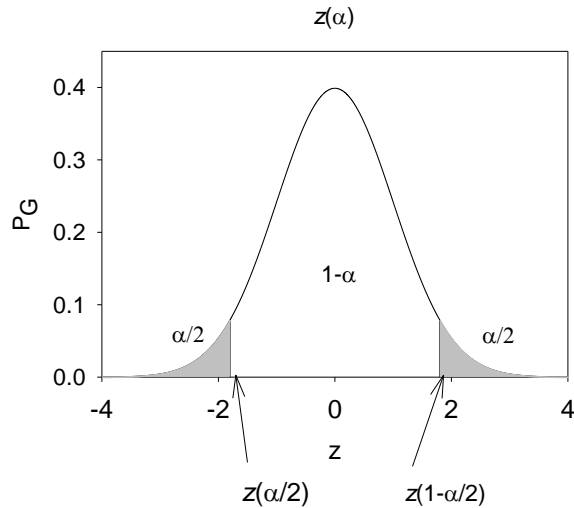
0.05) signifie qu'il y a une probabilité de 95% que la valeur vraie se trouve dans une intervalle : $x_i \pm z\sigma$.

Remarque : Les valeurs de la fonction $z(\alpha)$ peuvent être obtenues par Excel;

$z(\alpha) = \text{LOI.NORMALE.STANDARD.INVERSE}(1-\alpha/2)$

p. ex. $z(0.05) = \text{LOI.NORMALE.STANDARD.INVERSE}(0.975) = 1.95996$.

Dans la version anglaise : $z(\alpha) = \text{NORMSINV}(1-\alpha/2)$.



Exemple

Le cuivre a été déterminé par la spectroscopie atomique. On a trouvé la moyenne de 3 mesures $\bar{x} = 2.30$ ppm et la déviation standard de la méthode (déterminé auparavant) $\sigma = 0.20$ ppm. Calculer des intervalles de confiance de cette valeur pour le niveau de confiance $\alpha = 95\%$ et 99% .

À partir du Tableau 14.2 (Annexe) : $z(95\%) = 1.96$ et $z(99\%) = 2.58$.

Pour $\alpha = 95\%$, $IC(\mu) = \bar{x} \pm z(95\%) \sigma_{\bar{x}} = 2.30 \pm 1.96 \times 0.20 / \sqrt{3} = 2.30 \pm 0.23$

Pour $\alpha = 99\%$, $IC(\mu) = \bar{x} \pm z(99\%) \sigma_{\bar{x}} = 2.30 \pm 2.58 \times 0.20 / \sqrt{3} = 2.30 \pm 0.30$.

Il y a 5 chances sur 100 que la valeur vraie se trouve à l'extérieur des limites 2.30 ± 0.23 . Il y a aussi une chance sur 100 que la valeur vraie se trouve à l'extérieur des limites 2.30 ± 0.30 .

4.5.2 Intervalle de confiance quand σ est inconnue

Quand σ est inconnue on doit utiliser s et au lieu de la distribution de Gauss on doit utiliser la **distribution de Student**. Elle donne des intervalles de confiance plus grandes. Seulement quand $N \rightarrow \infty$, la distribution de Student devient la distribution gaussienne. La fonction de la distribution de Student est illustrée dans la Figure 6.

On calcule les intervalles de confiance en utilisant une formule suivante:

$$IC(\mu) = \bar{x} \pm t(\alpha, k) \frac{s}{\sqrt{N}} = \bar{x} \pm t(\alpha, k) s_{\bar{x}} \quad (17)$$

où k est le nombre de degrés de liberté, $k = N - 1$, et les valeurs de t sont présentées dans le Tableau en Annexe 14.3.

Remarque : Les valeurs de t peuvent être trouvées en utilisant une fonction d'Excel : $t(\alpha, k) = \text{LOI.STUDENT.INVERSE}(\alpha, k)$, p. ex. $\text{LOI.STUDENT.INVERSE}(0.05, 5) = 2.5706$. Dans la version anglaise : $t(\alpha, k) = \text{TINV}(\alpha, k)$.

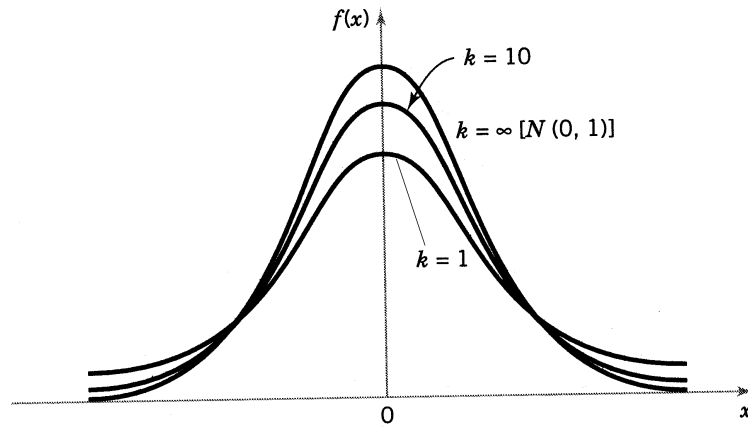


Figure 14. La distribution de Student; pour $k = \infty$ elle devient la distribution normale de Gauss.

La valeur moyenne, l'écart type, l'écart type de la valeur moyenne, des intervalles de confiance, etc. peuvent être calculées en utilisant : Statistiques descriptives dans Outils, Outils d'analyse.

Exemple

Le dosage de l'alcool dans le sang a donné les résultats suivants : 0.084%, 0.089% et 0.079%. Calculer des intervalles de confiance pour le degré de confiance $\alpha = 95\%$

a) en utilisant les trois données seulement

b) en supposant que σ de la méthode est connu, $\sigma = 0.006\%$

ad. a)

3 mesures, la distribution de Student

$$\bar{x} = (0.084 + 0.089 + 0.079)/3 = 0.084$$

$$s = \sqrt{\frac{\sum x_i^2 - (\sum x_i)^2 / N}{N-1}} = \sqrt{\frac{0.021218 - (0.252)^2 / 3}{3-1}} = 0.005$$

$$t(95\%, 2) = 4.30$$

$$\text{IC} = 0.084 \pm 4.30 * 0.005 / \sqrt{3} = 0.084 \pm 0.012$$

ad. b), σ connu

$$z(95\%) = 1.96$$

$$\text{IC} = 0.084 \pm 1.96 * 0.006 / \sqrt{3} = 0.084 \pm 0.007$$

4.5.3 L'écart-type d'un ensemble des données (« pooling » ou observations couplées)

On peut déterminer L'écart-type à partir d'une série des données.

Exemple

On a déterminé la concentration de fer dans 7 échantillons différents, chaque échantillon a été analysé de $N_i = 2$ à 6 fois, avec une moyenne rapportée dans la troisième colonne. Trouver L'écart-type de l'ensemble de ces données.

# d'échantillon	N_i	\bar{x}	$\sum (x_i - \bar{x})^2$
1	3	1.673	0.0259
2	4	1.015	0.0115
3	2	3.240	0.0242
4	6	2.018	0.0611
5	4	0.570	0.0114
6	5	2.482	0.0658
7	4	1.130	0.0170
	$N = 28$		$\sum = 0.2169$

$$s_{ens} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N - n}} = \sqrt{\frac{0.2169}{28 - 7}} = 0.10 \quad (18)$$

Si $N - n \geq 30$, on peut estimer l'écart type de la population σ .

Exemple

Une série des mesures de la concentration du fer en solution par l'absorption atomique donne les résultats suivants : 3.2, 2.9, 3.0, 3.3, 3.1 ppm. Calculez la moyenne, l'écart type, l'écart type de la moyenne et les intervalles de confiance pour le niveau de confiance de 95% et 99%.

En utilisant une fonction d'Excel : « Statistiques descriptives » dans l'Utilitaire d'analyse on trouve :

Pour le niveau de confiance de 95%

<i>Colonne 1</i>		
Moyenne	3.1	
Erreur-type	0.070710678	Écart-type de la moyenne 0.071
Médiane	3.1	
Mode	#N/A	
Écart-type	0.158113883	Écart-type 0.16
Variance de l'échantillon	0.025	
Kurtosis (Coefficient d'aplatissement)	-1.2	
Coefficient d'assymétrie	8.69675E-15	
Plage	0.4	
Minimum	2.9	
Maximum	3.3	
Somme	15.5	
Nombre d'échantillons	5	
Niveau de confiance(95.0%)	0.196324723	Intervalle de confiance de la moyenne

Résultats :

$$\bar{x} = 3.10, s_x = 0.16, s_{\bar{x}} = 0.07$$

$$IC : 3.10 - 0.20 = 2.90 \leq \bar{x} \leq 3.10 + 0.20 = 3.30$$

Il y a 5 chances sur 100 que la moyenne est à l'extérieur de ces intervalles.

Pour le niveau de confiance de 99%.

<i>Colonne 1</i>		
Moyenne	3.1	3.100
Erreur-type	0.070710678	0.071
Médiane	3.1	
Mode	#N/A	
Écart-type	0.158113883	0.16
Variance de l'échantillon	0.025	
Kurtosis (Coefficient d'aplatissement)	-1.2	
Coefficient d'assymétrie	8.69675E-15	
Plage	0.4	
Minimum	2.9	
Maximum	3.3	
Somme	15.5	
Nombre d'échantillons	5	
Niveau de confiance(99.0%)	0.325557651	0.33

$$IC : 3.10 - 0.33 = 2.77 \leq \bar{x} \leq 3.10 + 0.33 = 3.43$$

Il y a une chance sur 100 que notre moyenne est à l'extérieur de ces limites.

4.6 La moyenne pondérée

Lorsque les mesures possèdent une précision différente on est obligé d'utiliser la moyenne pondérée. Supposons, qu'on veut calculer la moyenne d'une série des mesures, et pour chaque mesure on connaît l'écart type correspondante :

$$\begin{array}{l} x_1 \quad s_{x_1} \\ x_2 \quad s_{x_2} \\ x_3 \quad s_{x_3} \\ \dots \\ x_N \quad s_{x_N} \end{array}$$

Le poids statistique de chaque mesure est définie comme suit : $w_i = 1/s_{x_i}^2$. La moyenne pondérée est calculée comme :

$$\bar{x} = \frac{\sum w_i x_i}{\sum w_i} \quad (19)$$

et la variance de la moyenne pondérée :

$$s_{\bar{x}}^2 = \frac{1}{\sum w_i} = \frac{1}{\sum \frac{1}{s_{x_i}^2}} \quad (20)$$

Exemple

Dans le cas de désintégration radioactive l'écart type est égale à la racine carrée de la valeur mesurée : $s_{x_i} = \sqrt{x_i}$. Les mesures suivantes ont été obtenues par un expérimentateur :

Temps d'acquisition t_i / min	Nombre des désintégrations (impulsions) mesuré x_i
5	10255
20	41200
2	4084
10	20650

Calculez l'activité radioactive par minute de cette échantillon.

On détermine l'activité par minute, $r_i = X_i/t$ et son écart-type : $s_{r_i} = \frac{s_{x_i}}{t} = \frac{\sqrt{x_i}}{t_i}$

Activité $r_i = x_i / t_i$	Écart type $\sqrt{x_i} / t_i$	w_i	$w_i x_i$	$\sum w_i$
2051	20	0.0025	5.1275	0.0186
2060	10	0.010	20.6	
2042	32	0.001	2.042	
2065	14	0.0051	10.5315	

En utilisant l'éqn. (19) et (20) on obtient :

$$\bar{x} = 2059 \text{ impulsions/min}, \quad s_{\bar{x}} = 7 \text{ imp./min.}$$

Dans certains cas, l'écart type de la moyenne semble être trop petit par rapport aux différences entre les valeurs mesurées. Cela peut arriver quand la distribution des valeurs mesurées n'est pas normale. Il est possible de vérifier si la distribution est normale en utilisant un test χ^2 (kchi-deux), voir chapitre 7.2.

Ce problème arrive souvent dans l'expérience : vitesse de son. Dans ce cas on peut augmenter la valeur de l'écart type de la moyenne en la multipliant par un facteur de correction :

$$s_{\bar{x}, \text{corrigée}} = s_{\bar{x}} \sqrt{\frac{\chi^2}{N-1}} \quad (21)$$

où

$$\chi^2 = \sum_{i=1}^N \frac{(x_i - \bar{x})^2}{s_{x_i}^2} = \sum_{i=1}^N w_i (x_i - \bar{x})^2 \quad (22)$$

Exemple

Cinq mesures expérimentales, avec ces écart-types sont présentées ci-dessous :

x_i	s_{x_i}
1.4	0.2
0.9	0.15
3.0	0.3
1.8	0.2
2.5	0.25

Calculez la moyenne et l'écart type.

Les calculs présentés dans le chapitre 7.2 indiquent que la distribution n'est pas normale. On doit augmenter l'écart type de la moyenne en utilisant l'éq. (21). Les résultats des calculs sont présentés dans la table :

x_i	s_{x_i}	w_i	$w_i x_i$	$w_i (x_i - \bar{x})^2$
1.4	0.2	25	35	0.907195
0.9	0.15	44.44444	40	21.19029
3	0.3	11.11111	33.33333	22.07454
1.8	0.2	25	45	1.097323
2.5	0.25	16	40	13.23523
somme		121.5556	193.3333	58.50457

$$\bar{x} = \mathbf{1.59} \quad \chi^2(0.05, 4) =$$

$$s_{\bar{x}} = \mathbf{0.09}$$

Correction

$$\sqrt{\frac{\chi^2}{4}} = 3.824414 \quad \bar{x} = 1.59$$

$$s_{\bar{x},corr} = 0.346879 \quad s_{\bar{x},corr} = 0.35$$

Les valeurs mesurées avec ces écarts-types sont présentées dans le graphique.

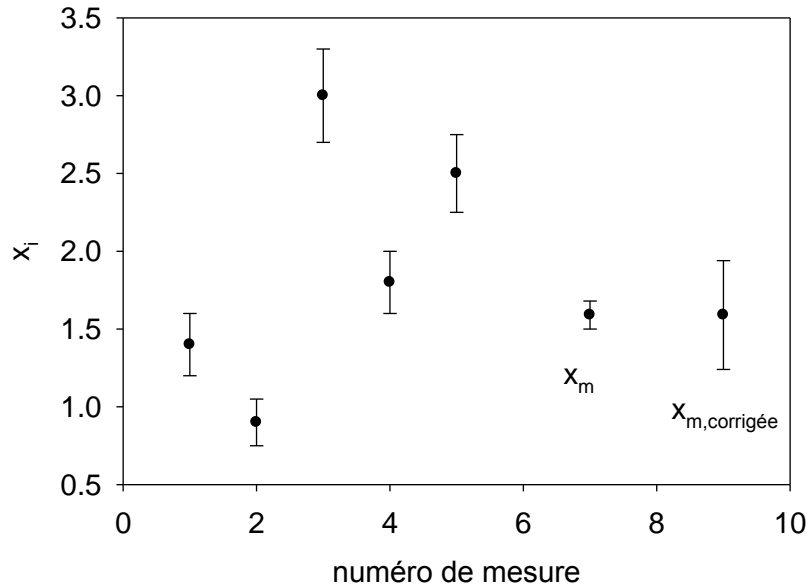


Figure 15. Calculs de la moyenne pondérée; l'écart type de la moyenne x_m est trop petit mais il devient correct après la correction, Éq. (21).

On voit que l'écart type calculée en utilisant l'équation (20) semble être trop petite, mais celle calculée de l'éq. (21) est plus grande et plus correcte.

5. Propagation d'erreurs

Souvent on calcule une fonction « $z(x_1, x_2, x_3, \dots, x_p)$ » des plusieurs paramètres (x_i) et chaque paramètre est déterminé avec certain écart-type.

La question est la suivante :

Quelle est l'écart type de la valeur z calculée indirectement en utilisant les paramètres x_i ? C'est un problème de la propagation des erreurs.

Posons z comme une fonction de p paramètres :

$$z = f(x_1, x_2, \dots, x_p)$$

La dérivée totale de cette fonction est :

$$dz = \left(\frac{\partial f}{\partial x_1} \right)_{x_2, \dots, x_p} dx_1 + \left(\frac{\partial f}{\partial x_2} \right)_{x_1, x_3, \dots, x_p} dx_2 + \left(\frac{\partial f}{\partial x_p} \right)_{x_1, \dots, x_{p-1}} dx_p \quad (23)$$

Nous avons supposé que les déviations dx_i sont beaucoup plus petites que les valeurs x_i , où dz est définie comme :

$$dz = z(x_1 + \Delta x_1, x_2 + \Delta x_2, \dots) - f(x_1, x_2, \dots, x_p)$$

$$dx_i = x_i - \mu$$

Le carrée de dz est

$$\begin{aligned} dz^2 &= \left(\frac{\partial f}{\partial x_1} \right)^2 dx_1^2 + \left(\frac{\partial f}{\partial x_2} \right)^2 dx_2^2 + \dots \\ &+ 2 \left(\frac{\partial f}{\partial x_1} \right) \left(\frac{\partial f}{\partial x_2} \right) dx_1 dx_2 + 2 \left(\frac{\partial f}{\partial x_1} \right) \left(\frac{\partial f}{\partial x_3} \right) dx_1 dx_3 + \dots \end{aligned}$$

Les valeurs dx_i et les dérivées partielles peuvent être positives ou négatives et dans la sommations elles peuvent s'annuler. Seulement les termes au carrée s'additionnent, donc :

$$\sum_{i=1}^N (dz_i)^2 \approx \left(\frac{\partial f}{\partial x_1} \right)^2 \sum_{i=1}^N (dx_{1,i})^2 + \left(\frac{\partial f}{\partial x_2} \right)^2 \sum_{i=1}^N (dx_{2,i})^2 + \dots \quad | : N$$

mais:

$$\frac{\sum dz^2}{N} = \frac{\sum (x_i - \mu)^2}{N} = \sigma_z^2$$

$$\sigma_z^2 = \left(\frac{\partial f}{\partial x_1} \right)^2 \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2} \right)^2 \sigma_{x_2}^2 + \dots \quad (24)$$

$$s_z^2 = \left(\frac{\partial f}{\partial x_1} \right)^2 s_{x_1}^2 + \left(\frac{\partial f}{\partial x_2} \right)^2 s_{x_2}^2 + \dots$$

5.1 L'erreur maximale

Quand on ne connaît pas des déviations standards de chaque paramètre on ne peut pas utiliser l'éqn. (24). Cependant, si on peut estimer une erreur maximale de chaque paramètre, on peut déterminer une erreur maximale de la fonction $z = f(x_1, x_2, \dots, x_p)$ avec $|\Delta x_i| \ll |x_i|$

$$dz = \left(\frac{\partial f}{\partial x_1} \right)_{x_2, \dots, x_p} dx_1 + \left(\frac{\partial f}{\partial x_2} \right)_{x_1, x_3, \dots, x_p} dx_2 + \left(\frac{\partial f}{\partial x_p} \right)_{x_1, \dots, x_{p-1}} dx_p$$

$$|dz| = \left| \left(\frac{\partial f}{\partial x_1} \right)_{x_2, \dots, x_p} \Delta x_1 \right| + \left| \left(\frac{\partial f}{\partial x_2} \right)_{x_1, x_3, \dots, x_p} \Delta x_2 \right| + \dots + \left| \left(\frac{\partial f}{\partial x_p} \right)_{x_1, \dots, x_{p-1}} \Delta x_p \right|$$

$$\Delta z = \sum_{i=1}^p \left| f'_{x_i}(x_1, x_2, \dots, x_p) \Delta x_i \right| \quad (25)$$

Δz représente une erreur maximale obtenue sans compensation d'erreurs aléatoires, c'est le pire cas possible. Cette méthode peut être utilisée pour une estimation rapide des erreurs possibles.

Exemple

Calculez l'écart type du volume d'une boîte de dimensions a , b et c .

$$V = a \times b \times c$$

$$a = 5.5 \quad b = 3.6 \quad c = 1.9 \quad s_a = s_b = s_c = 0.1$$

$$V = 37.62$$

$$s_v^2 = (f'_a \cdot s_a)^2 + (f'_b \cdot s_b)^2 + (f'_c \cdot s_c)^2 = (b \cdot c \cdot s_a)^2 + (a \cdot c \cdot s_b)^2 + (a \cdot b \cdot s_c)^2 = 6.49$$

$$s_v = 2.55$$

Réponse:

$$V = 37.6 \quad s_v = 2.6$$

Calculez l'erreur maximale de volume si des erreurs maximales des dimensions sont :

$$\Delta a = \Delta b = \Delta c = 0.1$$

$$\Delta V = ?$$

$$\Delta V = |f'_a \Delta a| + |f'_b \Delta b| + |f'_c \Delta c|$$

$$\Delta V = 3.78$$

$$V = 38 \quad \Delta V = 4 \quad \text{ou} \quad V = 37.6 \quad \Delta V = 3.8$$

Exemple

$$Y = a + b$$

$$s_y^2 = \left(\frac{\partial y}{\partial a} \right)^2 s_a^2 + \left(\frac{\partial y}{\partial b} \right)^2 s_b^2 = s_a^2 + s_b^2$$

Dans la somme on a l'addition des variances (variance = écart-type au carré).

$$y = ab$$

$$s_y^2 = b^2 s_a^2 + a^2 s_b^2 \quad ; \quad y^2 = a^2 \cdot b^2$$

$$\frac{s_y^2}{y^2} = \frac{s_a^2}{a^2} + \frac{s_b^2}{b^2} \quad s_y = y \sqrt{\left(\frac{s_a}{a} \right)^2 + \left(\frac{s_b}{b} \right)^2}$$

Dans le cas de multiplication ou division les variances relatives s'additionnent.

$$y = e^a$$

$$s_y^2 = (e^a)^2 s_a^2 \quad \frac{s_y}{y} = s_a$$

$$y = a^n$$

$$s_y^2 = (na^{n-1}s_a)^2 \quad \frac{s_y}{y} = \frac{na^{n-1}}{a^n} s_a = \frac{n}{a} s_a$$

Erreur de la moyenne

$$\mu = \frac{\sum_{i=1}^N x_i}{N} = \frac{1}{N} (x_1 + x_2 + \dots + x_N)$$

$$\begin{aligned} s_\mu^2 &= \sum_{i=1}^N \left(\frac{\partial \mu}{\partial x_i} \right)^2 s_x^2 = \left(\frac{1}{N} \right)^2 s_x^2 + \left(\frac{1}{N} \right)^2 s_x^2 + \dots + \left(\frac{1}{N} \right)^2 s_x^2 \\ &= \frac{1}{N^2} s_x^2 + \frac{1}{N^2} s_x^2 + \dots + \frac{1}{N^2} s_x^2 = \frac{N}{N^2} s_x^2 = \frac{s_x^2}{N} \end{aligned}$$

$$s_\mu = \frac{s_x}{\sqrt{N}}$$

6. La méthode des moindres carrés pour une ligne droite

Voilà des exemples de corrélations entre les données :

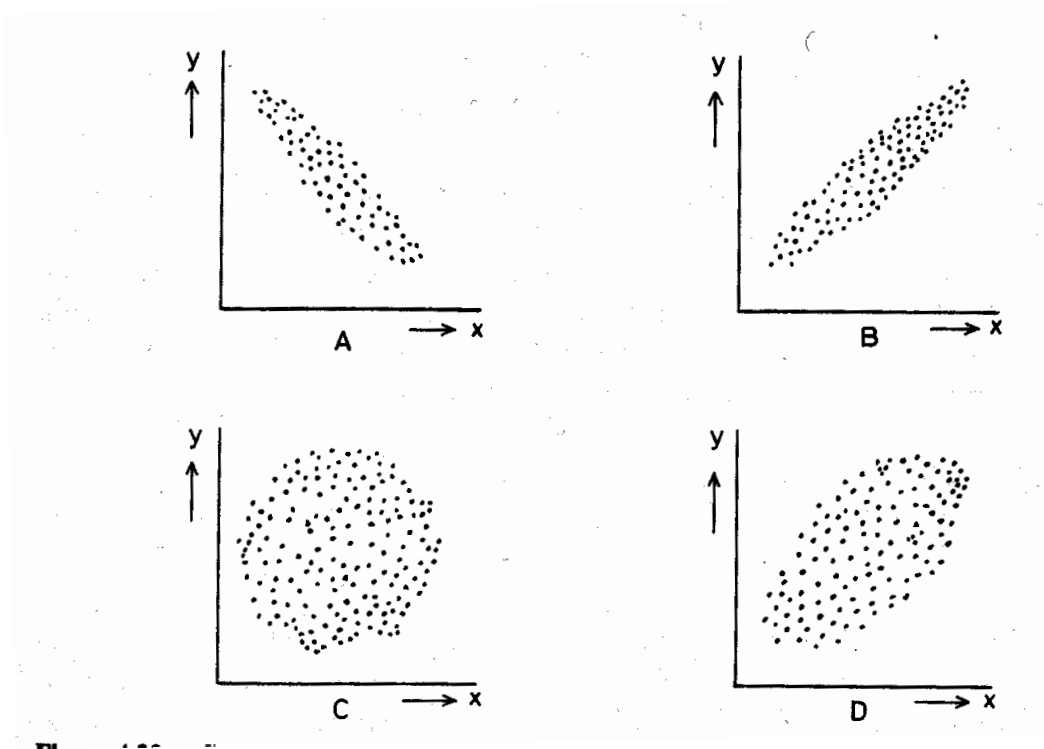


Fig. Corrélations entre x et y : a) et b) corrélations négatives ou positives avec des dispersions faibles; c) pas de corrélation, d) corrélation avec grande dispersion (pas très significative).

En chimie ou physique on observe souvent des relations linéaires ou celles qui peuvent être linéarisées; p. ex. la relation entre l'absorbance et la concentration, $\ln k$ et temps (pour la cinétique de premier ordre), etc.

Supposons que les valeurs y_i sont mesurées en fonction de paramètre x_i :

x_1 y_1
 x_2 y_2
 x_3 y_3
 \bullet \bullet
 \bullet \bullet
 x_N y_N

Seulement les valeurs mesurées y_i sont déterminées avec certaine erreur, les **valeurs x_i sont connues avec une grande précision (sans erreur)**. On suppose une relation linéaire entre y et x .

$$y = b_0 + b_1x \quad (26)$$

mais, en réalité, pour chaque point on doit écrire une équation :

$$y_i = b_0 + b_1x_i + \varepsilon_i \quad (27)$$

où ε_i est un écart entre la valeur expérimentale y_i et la valeur calculée en utilisant une équation linéaire. Il faut trouver un minimum de la fonction S^2 :

$$S^2 = \sum_i^N \varepsilon_i^2 = \sum_{i=1}^N (y_i - b_0 - b_1 x_i)^2 = \sum_{i=1}^N (y_i - \hat{y}_i)^2 = \min \quad (28)$$

c'est à dire des paramètres b_0 et b_1 qui minimise cette fonction.

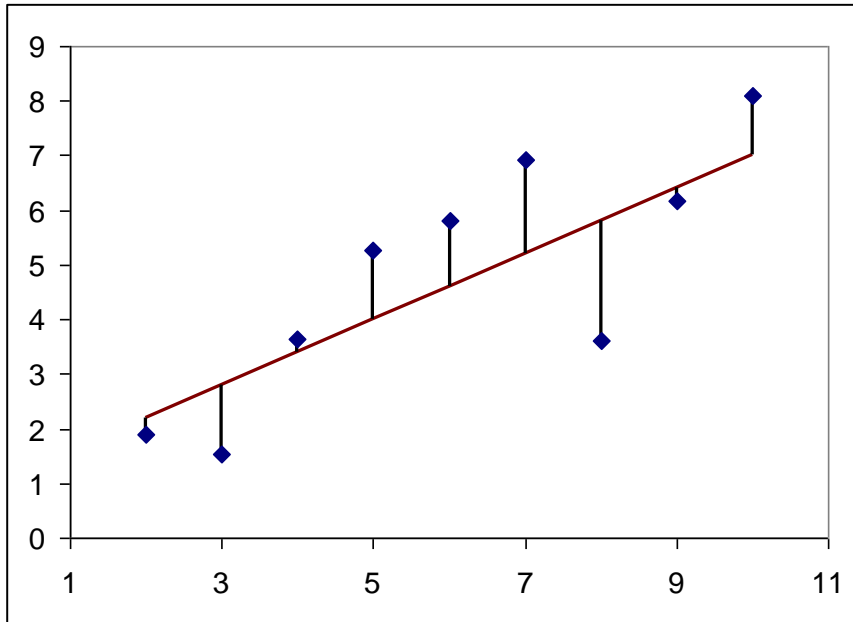


Fig. Les points expérimentaux et la ligne droite trouvée par minimisation de $\sum \varepsilon_i^2$.

Le minimum d'une fonction existe quand la dérivée est égale à zéro :

$$\left(\frac{\partial S^2}{\partial b_0} \right)_{b_1} = 0 \quad \left(\frac{\partial S^2}{\partial b_1} \right)_{b_0} = 0$$

Cela donnent :

$$\begin{aligned} \sum 2(y_i - b_0 - b_1 x_i) &= 0 \\ \sum 2(y_i - b_0 - b_1 x_i) x_i &= 0 \end{aligned}$$

$$\begin{aligned} \sum y_i - N b_0 - b_1 \sum x_i &= 0 \\ \sum_i x_i y_i - b_0 \sum_i x_i - b_1 \sum_i x_i^2 &= 0 \end{aligned}$$

On obtient deux équations avec deux inconnues :

$$\begin{aligned} N b_0 + \left(\sum x_i \right) b_1 &= \sum y_i \\ \left(\sum x_i \right) b_0 + \left(\sum x_i^2 \right) b_1 &= \sum x_i y_i \end{aligned}$$

$$\begin{bmatrix} N & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix} \quad (29)$$

La solution :

$$b_0 = \frac{\sum x^2 \sum y - \sum x \sum xy}{d} \quad (30)$$

$$b_1 = \frac{N \sum xy - \sum x \sum y}{d} \quad (31)$$

et

$$d = N \sum x^2 - (\sum x)^2 = NS_{xx} \quad (32)$$

On peut donner une autre forme de ces éqns. en utilisant des paramètres :

$$S_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{N} = \frac{d}{N} \quad (33)$$

$$S_{yy} = \sum (y_i - \bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{N}$$

$$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{\sum x_i \sum y_i}{N}$$

Formule pour S_{xx} :

$$\begin{aligned} S_{xx} &= \sum (x_i - \bar{x})^2 = \sum \left(x_i^2 - \frac{2x_i \sum x_i}{N} + \frac{(\sum x_i)^2}{N^2} \right) = \\ &= \left(\sum x_i^2 - \frac{2(\sum x_i)^2}{N} + \frac{N(\sum x_i)^2}{N^2} \right) = \\ &= \sum x_i^2 + \frac{-2(\sum x_i)^2 + (\sum x_i)^2}{N} = \sum x_i^2 - \frac{(\sum x_i)^2}{N} \end{aligned}$$

$$b_1 = \frac{S_{xy}}{S_{xx}} \quad (34)$$

L'écart type d'une valeur de y_i , $s_y = s_r$ (résiduelle) est calculée comme la somme des déviations entre y_i et la valeur $\hat{y} = b_0 + b_1 x_i$ calculée à partir de l'équation de la régression, pour $N - 2$ degré de liberté (**deux** équations ont été utilisées pour calculer les valeurs de b_0 et b_1) :

$$s_r = s_y = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{N - 2}} = \sqrt{\frac{\sum (y_i - b_0 - b_1 x_i)^2}{N - 2}} = \sqrt{\frac{S_{yy} - b_1^2 S_{xx}}{N - 2}} \quad (35)$$

L'écart type de b_1 est calculée à partir de la relation (34) en utilisant la loi de propagation d'erreurs :

$$\begin{aligned} s_{b_1}^2 &= \sum_i \left(\frac{\partial b_1}{\partial y_i} \right)^2 s_y^2 = s_y^2 \sum_i \left(\frac{\partial b_1}{\partial y_i} \right)^2 \\ \frac{\partial b_1}{\partial y_i} &= \frac{N x_i - \sum x_i}{d} \\ \left(\frac{\partial b_1}{\partial y_i} \right)^2 &= \frac{N^2 x_i^2 - 2N x_i \sum x_k + (\sum x_k)^2}{d^2} \\ s_{b_1}^2 &= \frac{N^2 \sum x_i^2 - 2N (\sum x_i)^2 + N (\sum x_i)^2}{d^2} s_y^2 = N \frac{N \sum x_i^2 - (\sum x_i)^2}{\left[N \sum x_i^2 - (\sum x_i)^2 \right]^2} s_y^2 = \frac{N}{d} s_y^2 \\ s_{b_1} &= \sqrt{\frac{N}{d}} s_y = \frac{s_y}{\sqrt{S_{xx}}} \end{aligned} \quad (36)$$

L'écart type de b_0 est calculée à partir de l'équation : $\bar{y} = b_0 + b_1 \bar{x}$:

$$\begin{aligned} b_0 &= \bar{y} - b_1 \bar{x} \\ s_{b_0}^2 &= \left(\frac{\partial b_0}{\partial \bar{y}} \right)^2 s_{\bar{y}}^2 + \left(\frac{\partial b_0}{\partial b_1} \right)^2 s_{b_1}^2 = 1 \cdot \frac{s_y^2}{N} + (-\bar{x})^2 \frac{N s_y^2}{d} = \frac{\sum x_i^2}{d} s_y^2 \\ s_{b_0} &= \sqrt{\frac{\sum x_i^2}{d}} s_y = \sqrt{\frac{\sum x_i^2}{N S_{xx}}} s_y \end{aligned} \quad (37)$$

6.1 Propriétés de la méthode des moindres carrés

$$\begin{aligned}
\sum_{i=1}^N (y_i - \hat{y}_i) &= \sum (y_i - b_0 - b_1 x_i) \\
&= \sum \left[\frac{y_i - \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{d}}{\frac{(N \sum x_i y_i - \sum x_i \sum y_i) x_i}{d}} \right] = \\
&= \frac{\sum y_i * d - N \sum x_i^2 \sum y_i + N \sum x_i \sum x_i y_i}{d} \\
&+ \frac{-N \sum x_i \sum x_i y_i + \sum y_i (\sum x_i)^2}{N} \\
&= \frac{N \sum x_i^2 \sum y_i - \sum y_i (\sum x_i)^2}{d} \\
&+ \frac{-N \sum x_i^2 \sum y_i + \sum y_i (\sum x_i)^2}{d} = 0
\end{aligned}$$

$$\sum_{i=1}^N (y_i - \hat{y}_i) = 0 \quad (38)$$

$$\sum y_i = \sum \hat{y}_i \quad (39)$$

$$\bar{y}_i = \bar{\hat{y}}_i \quad (40)$$

6.2 Écart types et intervalles de confiance de \hat{y}_i calculée:

$$\hat{y}_i = b_0 + b_1 x_i$$

$$\hat{y}_i = \bar{y} + b_1 (x_i - \bar{x})$$

$$\begin{aligned}
s_{\hat{y}_i}^2 &= s_y^2 + s_{b_1}^2 (x_i - \bar{x})^2 \\
s_{\hat{y}}^2 &= \frac{s_y^2}{N} \quad s_y = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{N-2}} \quad s_{b_1}^2 = \frac{s_y^2}{S_{xx}} \quad S_{xx} = \sum (x_i - \bar{x})^2 \\
s_{\hat{y}_i} &= \sqrt{\frac{1}{N} + \frac{(x_i - \bar{x})^2}{S_{xx}}} s_y \quad (41)
\end{aligned}$$

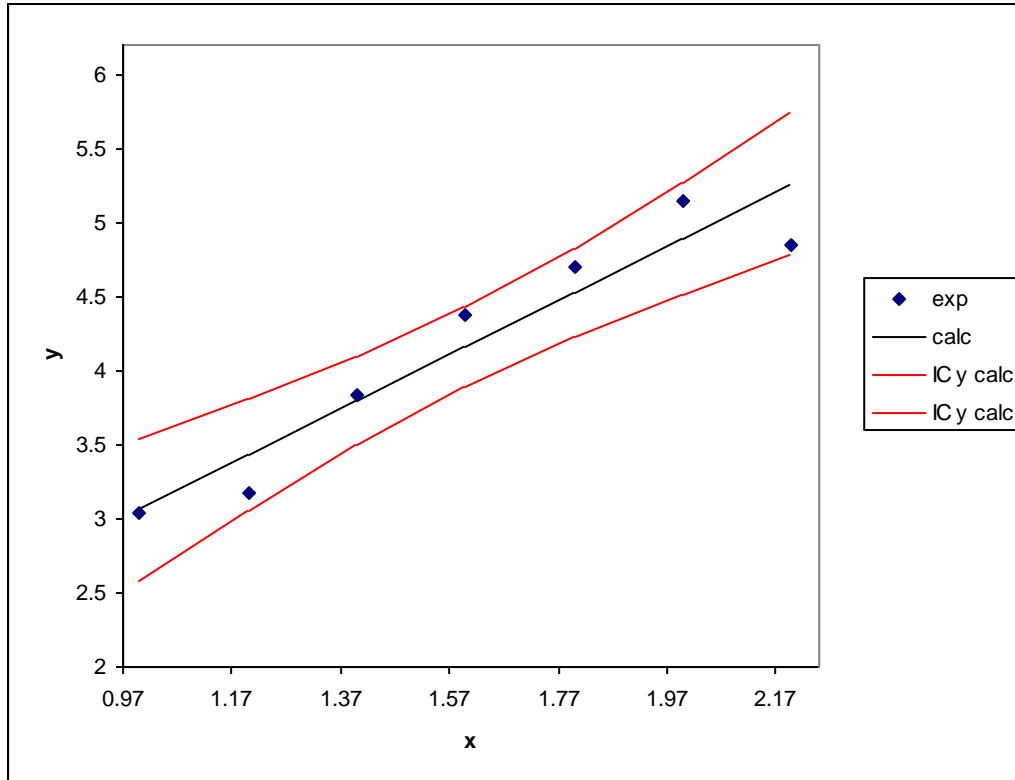


Fig. Distribution des valeurs des intervalles de confiance de \hat{y}_i , $\hat{y}_i \pm t_{0.05, N-2} s_{\hat{y}_i}$.

Erreurs:

$$b_0 \pm t_{0.05, N-2} s_{b_0}$$

$$b_1 \pm t_{0.05, N-2} s_{b_1}$$

$$\hat{y}_i \pm t_{0.05, N-2} s_{\hat{y}_i}$$

6.3 Écart types et intervalles de confiance de y_i (expérimentale)

De la même manière on peut déterminer les intervalles de confiance des points expérimentaux. L'écart d'un point expérimental, y_i , de la valeur prédite par la régression, \hat{y}_i , est Δ :

$\Delta = \varepsilon_i = y_i - \hat{y}_i$. L'écart type de Δ est :

$$s_{\Delta}^2 = s_y^2 + s_{\hat{y}_i}^2 = s_y^2 \left(1 + \frac{1}{N} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right)$$

$$s_{\Delta} = s_y \sqrt{1 + \frac{1}{N} + \frac{(x_i - \bar{x})^2}{S_{xx}}}$$

et les intervalles de confiance sont $y_i \pm s_{\Delta} t(\alpha, k)$. Ils sont illustrés dans la figure suivante.

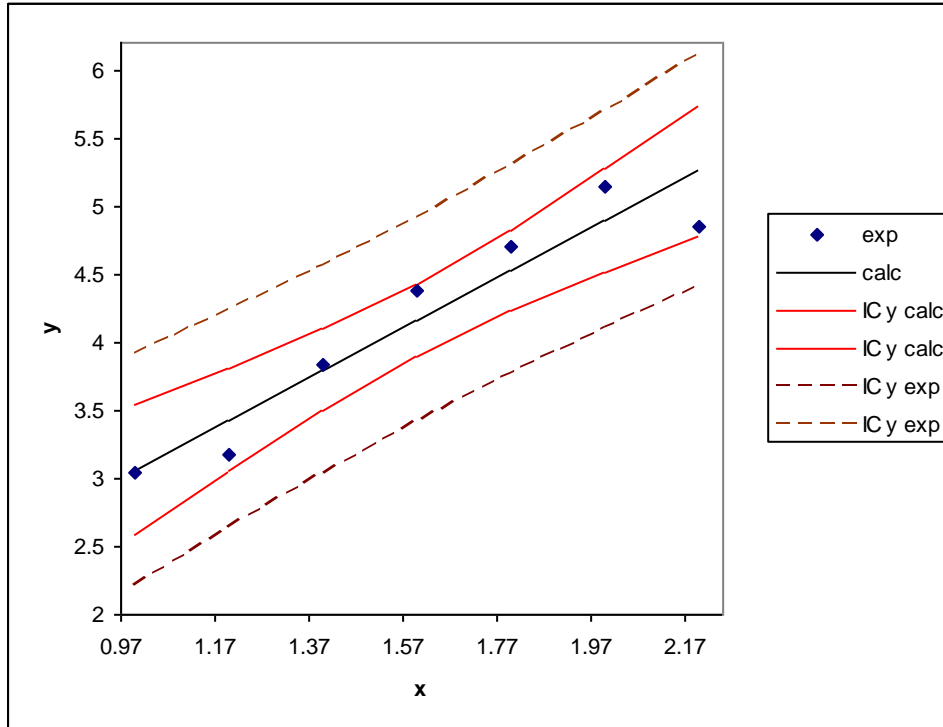


Figure 16. Comparaison de y expérimental, y_i , calculé par la régression \hat{y}_i , des intervalles de confiance de \hat{y}_i et y_i pour $\alpha = 0.05$.

6.4 Coefficient de corrélation

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} \quad (42)$$

$$-1 \leq r \leq 1$$

$$|r| > 0.95 \text{ bonne}$$

$$|r| > 0.99 \text{ très bonne}$$

6.4.1 Coefficient de détermination

100% r^2 - combien % de la variation totale de y peut être expliquée par l'équation $y = b_0 + b_1x$.

$$r^2 = \frac{\sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2} = \frac{\sum \text{des carrés dans la régression}}{\sum \text{des carrés hors de la moyenne (totale)}}$$

Exemple

Les mesures de y en fonction de x on données les résultats suivants :

x	y
-5	-7.4
-3	-4.3
-1	-0.4
1	3.3
3	6.7
5	10.2
7	12.4
9	16.4

Calculez l'équation de la régression linéaire.

En utilisant la Régression linéaire dans l'Excel on obtient les résultats présentés sur la page suivante.

$$y = b_0 + b_1x$$

$$b_0 = 1.22, s_{b_0} = 0.18, \text{ limite de confiance pour } \alpha = 95\% : 0.79 \leq b_0 \leq 1.65$$

$$b_1 = 1.698, s_{b_1} = 0.035, \text{ limite de confiance pour } \alpha = 95\% : 1.612 \leq b_1 \leq 1.784$$

$$R^2 = 0.9974303$$

Résultats de la régression linéaire :

Statistiques de la régression

Coefficient de
détermination
multiple 0.9987143

**Coefficient de
détermination**

R² 0.9974303

R²

Coefficient de
détermination

R² 0.9970020

Erreur-type 0.4561093

Observations 8

ANALYSE DE VARIANCE

	Degré de liberté	Somme des carrés	Moyenne des carrés	F	Valeur critique de F
Régression	1	484.500536	484.50053	2328.9	5.3076E-09
Résidus	6	1.24821429	0.20803571		
Total	7	485.74875			

	Coefficients	Erreur-type	Statistique t	Probabilité	Limite inférieure pour seuil de confiance = 95%	Limite supérieure pour seuil de confiance = 95%	Limite inférieure pour seuil de confiance = 95.0%	Limite supérieure pour seuil de confiance = 95.0%
Constante	1.216070	0.17594799	6.91153915	0.0004537	0.78554189	1.64660097	0.78554189	1.64660097
Variable X 1	1.69821	0.0351896	48.2589848	5.3076E-09	1.61210838	1.78432019	1.61210838	1.78432019

ANALYSE DES RÉSIDUS

Observation	Prévisions pour Y	Résidus
1	-7.275	-0.125
2	-3.87857	-0.4214285
3	-0.482142	0.0821428
4	2.914285	0.3857142
5	6.310714	0.3892857
6	9.707142	0.4928571
7	13.10357	-0.703571
8	16.5	-0.1

6.5 La méthode de moindres carrés pour $y = ax$

Dans certains cas on peut utiliser une régression simplifiée, $y = b_1x$, donc le terme libre $b_0 = 0$. Cette régression peut être utile pour modéliser une courbe de calibrage. On peut décrire le modèle comme suit :

$$y_i = b_1x_i + \varepsilon_i$$

Pour trouver la valeur optimale de b_1 il faut minimiser :

$$S^2 = \sum_{i=1}^N (y_i - b_1x_i)^2 = \sum_{i=1}^N (y_i^2 - 2b_1x_iy_i - b_1^2x_i^2) = \sum y_i^2 - 2b_1 \sum x_iy_i + b_1^2 \sum x_i^2$$

$$\frac{dS^2}{db_1} = -2 \sum x_iy_i + 2b_1 \sum x_i^2 = 0$$

$$b_1 = \frac{\sum x_iy_i}{\sum x_i^2} \quad (43)$$

L'écart type de la pente :

$$s_{b_1}^2 = \sum \left(\frac{db_1}{dy_i} \right)^2 s_y^2$$

$$\frac{db_1}{dy_i} = \frac{x_i}{\sum x_i^2} \quad \left(\frac{db_1}{dy_i} \right)^2 = \frac{x_i^2}{\left(\sum x_i^2 \right)^2}$$

$$s_{b_1}^2 = \frac{\sum x_i^2}{\left(\sum x_i^2\right)^2} s_y^2 = \frac{s_y^2}{\sum x_i^2} \quad (44)$$

$$s_{b_1} = \frac{s_y}{\sqrt{\sum x_i^2}}$$

$$s_y^2 = \frac{\sum (y_i - \hat{y}_i)^2}{N-1} = \frac{\sum (y_i - b_1 x_i)^2}{N-1}$$

L'écart-type de \hat{y}_i calculé, $\hat{y}_i = ax_i$:

$$\hat{y}_i - \bar{y} = b_1 (x_i - \bar{x})$$

$$\hat{y}_i = \bar{y} + b_1 (x_i - \bar{x})$$

$$s_{\hat{y}_i}^2 = s_{\bar{y}}^2 + s_{b_1}^2 (x_i - \bar{x})^2$$

$$s_{\hat{y}_i}^2 = \frac{s_y^2}{N} + \frac{(x_i - \bar{x})^2}{\sum x_i^2} s_y^2$$

$$s_{\hat{y}_i} = \sqrt{\frac{1}{N} + \frac{(x_i - \bar{x})^2}{\sum x_i^2}} s_y \quad (45)$$

Comme auparavant on peut aussi trouver les limites de confiance de $\Delta = y_i - \hat{y}_i$:

$$s_{y_i} = \sqrt{1 + \frac{1}{N} + \frac{(x_i - \bar{x})^2}{\sum x_i^2}} s_y$$

Exemple

On mesure le sodium par l'émission atomique. Pour les concentrations $c = x$ en ppm nous obtenons les signaux donnés par y (voir le tableau ci-dessus) :

x	y
0.2	0.62
0.4	1.17
0.6	1.79
0.8	2.43
1	3.1
1.2	3.58

Calculez la ligne de régression qui passe par l'origine.

Réponse :

$$y = b_1 x$$

$b_1 = 3.024$, $s_{b_1} = 0.024$, intervalles de confiance (95%) : $2.961 \leq b_1 \leq 3.086$

$$R^2 = 0.998346$$

SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.999173
R Square	0.998346
Adjusted R Square	0.798346
Standard Error	0.046191
Observations	6

R²

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	6.438682	6.43868	3017.71	6.57E-07
Residual	5	0.010668	0.00213	4	
Total	6	6.44935			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A
X Variable 1	3.024	0.024	124.887	7.624E-10	2.961391	3.085862

6.6 Erreur de x_c de la ligne de régression

Calculer une erreur de la valeur x_c déterminée à partir de la régression:

$$y_c = \bar{y} + b_1(x_c - \bar{x})$$

1) quand les données expérimentales sont décrites par l'équation $\hat{y} = b_0 + b_1x$

$$x_c = \bar{x} + \frac{y_c - \bar{y}}{b_1}$$

$$s_{x_c}^2 = \left(\frac{\partial x}{\partial y_c}\right)^2 s_y^2 + \left(\frac{\partial x}{\partial \bar{y}}\right)^2 s_{\bar{y}}^2 + \left(\frac{\partial x}{\partial b_1}\right)^2 s_{b_1}^2$$

$$s_{\bar{y}}^2 = \frac{s_y^2}{N} \quad s_{b_1}^2 = \frac{s_y^2}{S_{xx}}$$

$$\frac{\partial x}{\partial y_c} = \frac{1}{b_1}$$

$$\frac{\partial x}{\partial \bar{y}} = -\frac{1}{b_1}$$

$$\frac{\partial x}{\partial b_1} = -\frac{(y_c - \bar{y})}{b_1^2}$$

$$s_{x_c} = \frac{s_y}{b_1} \sqrt{1 + \frac{1}{N} + \frac{(y_c - \bar{y})^2}{b_1^2 S_{xx}}} \quad (46)$$

Quand la valeur y_c est mesurée m fois on obtient:

$$s_{x_c} = \frac{s_y}{b_1} \sqrt{\frac{1}{m} + \frac{1}{N} + \frac{(y_c - \bar{y})^2}{b_1^2 S_{xx}}} \quad (47)$$

2) quand les données expérimentales sont décrites par l'équation $\hat{y} = b_1x$, $s_{b_1} = s_y / \sqrt{\sum x_i^2}$ et on obtient :

$$s_{x_c} = \frac{s_y}{b_1} \sqrt{\frac{1}{m} + \frac{1}{N} + \frac{(y_c - \bar{y})^2}{b_1^2 \sum x_i^2}}$$

6.7 Méthode matricielle de moindres carrées

La description matricielle simplifie la description des problèmes des moindres carrés. Considérons une matrice \mathbf{X} , \mathbf{Y} , \mathbf{b} et matrice d'erreurs $\boldsymbol{\varepsilon}$ définies comme :

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \dots & \\ 1 & x_N \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_N \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \dots \\ \varepsilon_N \end{bmatrix} \quad (48)$$

Notre modèle linéaire peut être écrit comme : $\mathbf{Y} = \mathbf{X} \mathbf{b} + \boldsymbol{\varepsilon}$. En utilisant la définition des matrices transposées :

$$\mathbf{X}' = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_1 & x_2 & x_3 & \dots & x_N \end{bmatrix}$$

$$\mathbf{Y}' = [y_1 \quad y_2 \quad y_3 \quad \dots \quad y_N]$$

etc. on peut écrire :

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} N & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \quad \mathbf{X}'\mathbf{Y} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix} \quad (49)$$

et

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y} \quad (50)$$

qui est équivalent à l'éqn. (29). Pour résoudre cette équation il faut la multiplier par la matrice inverse :

$$(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})$$

La solution de cette équation est :

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \quad (51)$$

On peut aussi développer cette équation à partir d'une condition de moindres carrés. La somme des carrés est donnée par :

$$\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{Y} - \mathbf{X}\mathbf{b})'(\mathbf{Y} - \mathbf{X}\mathbf{b})$$

et

$$\frac{\partial(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})}{\partial \mathbf{b}} = -\mathbf{X}'(\mathbf{Y} - \mathbf{X}\mathbf{b}) - (\mathbf{Y} - \mathbf{X}\mathbf{b})'\mathbf{X} = 0$$

La réponse est un scalaire, donc $\mathbf{u}' = \mathbf{u}$

$$[(\mathbf{Y} - \mathbf{X}\mathbf{b})'\mathbf{X}]' = \mathbf{X}'(\mathbf{Y} - \mathbf{X}\mathbf{b})$$

d'où on obtient eqn. (50) :

$$\frac{\partial(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon})}{\partial \mathbf{b}} = -2\mathbf{X}'(\mathbf{Y} - \mathbf{X}\mathbf{b}) = 0$$

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}$$

Les valeurs calculées en utilisant la régression sont :

$$\hat{\mathbf{Y}} = \mathbf{X} \mathbf{b} \quad (52)$$

Cette méthode peut être utilisée pour déterminer la déviation standard des paramètres de la régression. La matrice des covariances est :

$$\mathbf{C}_b = (\mathbf{X}'\mathbf{X})^{-1} s_y^2 = s_y^2 \begin{bmatrix} \frac{\sum x_i^2}{NS_{xx}} & -\frac{\sum x_i}{NS_{xx}} \\ -\frac{\sum x_i}{NS_{xx}} & \frac{1}{S_{xx}} \end{bmatrix} \quad (53)$$

avec $s_y^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} / (N-2)$.

Les éléments diagonaux de cette matrice sont les variances (déviations standards au carrée) des paramètres \mathbf{b} et les éléments hors diagonale sont les covariances.

La valeur de \hat{y}_k calculée pour x_k peut être obtenue en introduisant un vecteur $\mathbf{X}'_k = [1 \quad x_k]$:

$$\hat{y}_i = \mathbf{X}'_k \mathbf{b} = \mathbf{b}' \mathbf{X}_k$$

et sa variance :

$$C_k = \mathbf{X}'_k \mathbf{C}_b \mathbf{X}_k = \mathbf{X}'_k (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_k s_y^2$$

En utilisant la méthode matricielle on peut définir une régression pour d'autres fonctions. Par exemple, pour une régression $y_i = b_0 + b_1 x_i + b_2 x_i^2$ la matrice \mathbf{X} est :

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ \dots & & \\ 1 & x_N & x_N^2 \end{bmatrix} \quad (54)$$

En général, la matrice \mathbf{X} est une matrice des dérivées de y_i par rapport aux paramètres de l'équation :

$$x_{ij} = \frac{\partial y_i}{\partial b_j} \quad \mathbf{X} = \begin{bmatrix} \frac{\partial y_1}{\partial b_0} & \frac{\partial y_1}{\partial b_1} & \frac{\partial y_1}{\partial b_2} \\ \frac{\partial y_2}{\partial b_0} & \frac{\partial y_2}{\partial b_1} & \frac{\partial y_2}{\partial b_2} \\ \frac{\partial y_3}{\partial b_0} & \frac{\partial y_3}{\partial b_1} & \frac{\partial y_3}{\partial b_2} \\ \dots & \dots & \dots \\ \frac{\partial y_N}{\partial b_0} & \frac{\partial y_N}{\partial b_1} & \frac{\partial y_N}{\partial b_2} \end{bmatrix} \quad (55)$$

6.8 La méthode des moindres carrés pondérée

Il est possible, que les données utilisées pour la régression sont déterminées avec une précision différente, c'est à dire chaque valeur de y_i possède un écart type différent. Dans ce cas on définit la matrice diagonale des poids statistiques des valeurs de y_i , égales à $1/s_{y_i}^2$:

$$\mathbf{G}_y = \begin{bmatrix} 1/s_{y_1}^2 & 0 & 0 & 0 & 0 \\ 0 & 1/s_{y_2}^2 & 0 & 0 & 0 \\ 0 & 0 & 1/s_{y_3}^2 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 1/s_{y_N}^2 \end{bmatrix} \quad (56)$$

Le problème est semblable à l'éqn. suivante :

$$\mathbf{X}'\mathbf{G}_y\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{G}_y\mathbf{Y} \quad (57)$$

avec la solution :

$$\mathbf{b} = (\mathbf{X}'\mathbf{G}_y\mathbf{X})^{-1} \mathbf{X}'\mathbf{G}_y\mathbf{Y} \quad (58)$$

et la matrice des covariances est égale à :

$$\mathbf{C}_b = (\mathbf{X}'\mathbf{G}_y\mathbf{X})^{-1} s_y^2 \quad (59)$$

avec $s_y^2 = \boldsymbol{\varepsilon}'\mathbf{G}_y\boldsymbol{\varepsilon}/(N-2)$.

La matrice des covariances des valeurs de \hat{y}_k calculées est déterminée comme :

$$\mathbf{C}_y = \mathbf{X}'_k \mathbf{C}_b \mathbf{X}_k \quad (60)$$

7. Tests statistiques pour une moyenne

7.1 Introduction

Les deux principaux champs d'étude de l'inférence statistique sont l'estimation de paramètres et les tests statistiques sur les paramètres. Le problème de l'estimation statistique consiste essentiellement

dans le choix d'une statistique, par exemple la moyenne. Les tests statistiques concernent les hypothèses qui sont formulées sur les paramètres qui nous intéressent. A partir d'une méthode d'échantillonnage appropriée, nous évaluons ensuite la (ou les) statistique(s) correspondante(s) pour déterminer si la différence entre la valeur de la statistique et celle du paramètre postulé est significative ou pas. Si oui, l'hypothèse est rejetée, si non, on considère qu'il n'y a pas suffisamment d'évidence statistique pour rejeter l'hypothèse postulée.

Une **hypothèse statistique** est un énoncé (une affirmation) quantitatif concernant les caractéristiques (valeurs des paramètres, forme de distribution des observations) d'une population.

Supposons que nous nous intéressons à la moyenne μ de la caractéristique d'un procédé, et que la moyenne hypothétique de ce procédé est μ_0 . On aimerait éprouver, en tirant un échantillon aléatoire de taille n du procédé, si le procédé est encore centré à μ_0 ou s'il y a eu déplacement de la tendance centrale (en plus ou en moins) et ceci à l'aide de la moyenne de l'échantillon.

Distinguons d'abord entre les hypothèses H_0 et H_1 qui est une notation conventionnelle en statistique pour identifier la formulation de l'hypothèse nulle H_0 et l'hypothèse alternative H_1

Un **test d'hypothèse** (ou test statistique) est une démarche qui a pour but de fournir une règle de décision permettant, sur la base de résultats d'échantillon, de faire un choix entre deux hypothèses statistiques. On compare deux hypothèses statistiques : l'hypothèse nulle (H_0) et l'hypothèse alternative (H_1).

L'hypothèse selon laquelle on fixe à priori un paramètre de la population à une valeur particulière (p. ex. zéro) s'appelle l'**hypothèse nulle** H_0 . N'importe quelle autre hypothèse qui diffère de l'hypothèse H_0 s'appelle l'**hypothèse alternative** H_1 . Les deux hypothèses sont exclusives : une de deux est toujours vraie, mais nous ne savons jamais laquelle avec certitude. La décision de choix a un caractère probabiliste : il y a un certain risque qu'elle soit erronée. Ce risque est donné par le seuil de signification du test. Le risque de rejeter à tort l'hypothèse nulle H_0 alors qu'elle est vraie s'appelle le **seuil de signification** du test α (souvent 1% ou 5%).

Ainsi, les résultats d'une décision peuvent se présenter de la façon suivante.

		Situation	
		H_0 vraie	H_1 vraie
Décision	Ne pas rejeter H_0	Bonne décision Probabilité $1 - \alpha$	Erreur de deuxième espèce, type II Risque = ???
	Rejeter H_0	Erreur de première espèce, type I Risque d'erreur = seuil = α	Bonne décision

7.2 Test χ^2

Le test χ^2 (khi-deux) est utilisé pour vérifier si les erreurs sont distribués d'une façon normale (toutes les calculs statistiques présentés ici sont valident pour la distribution normale). Si la distribution d'erreurs n'est pas normale, l'écart type calculé n'est pas correct.

Pour vérifier cela il faut tester deux hypothèses :

Les hypothèses à tester sont : **H_0 : la distribution des erreurs est normale**

H₁ : la distribution des erreurs n'est pas normale

$$\chi^2 = \sum_{i=1}^N \frac{(x_i - \bar{x})^2}{s_{x_i}^2}$$

Il faut comparer la valeur de χ^2 calculée des données expérimentales avec celle des tables $\chi^2(\alpha, N-1)$. Si, $\chi_{\text{exp}}^2 < \chi^2(\alpha, N-1)$ on garde l'hypothèse H₀ et la distribution est normale.

Cette fonction est dans l'Excel : KHIDEUX.INVERSE(α, k); p. ex. KHIDEUX.INVERSE(0.05,10) = 18.307, version anglaise : CHIINV(α, k).

Annexe 14.5 – table de la distribution $\chi^2(\alpha, N-1)$.

Exemple

On va tester les mesures présentés dans le chapitre 4.6.

x_i	s_{x_i}
1.4	0.2
0.9	0.15
3.0	0.3
1.8	0.2
2.5	0.25

La valeur de χ^2 (expérimentale), calculée dans le chapitre 4.6 est égale à 58.50 et $\chi^2(\alpha, N-1) = \chi^2(0.05, 4) = 9.49$. **Donc $\chi_{\text{exp}}^2 > \chi^2(\alpha, N-1)$, l'hypothèse H₀ doit être rejetée et la distribution des erreurs n'est pas normale. Dans ce cas, pour calculer l'écart type on a utilisé l'éq. (21).**

7.3 Test Q

Souvent on veut vérifier rapidement si un point n'est pas correct (erreur trop élevée). Pour cela on utilise le test Q. Il répond à la question : est-ce qu'il faut rejeter ce point?

L'hypothèse à tester H₀ : on garde ce point
contre H₁ : on rejette ce point

On détermine la valeur Q :

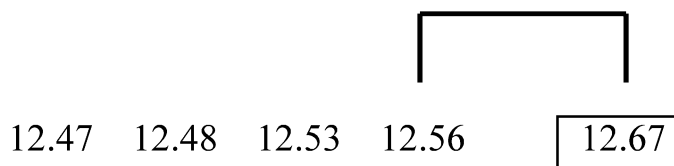
$$Q_{\text{calculé}} = \frac{\text{écart de ce point des données}}{\text{écart des données}}$$

Quand $Q_{\text{calculé}} > Q_{\text{Table}}$ on rejette ce point.

Exemple

Peut-on rejeter la plus grande valeur?

écart d'un point = 0.11



$$Q_{\text{calculé}} = \frac{0.11}{0.20} = 0.55$$



écart des données = 0.20

Dans le Table 14.6 pour $N = 5$, $Q_{\text{Table}}(95\%) = 0.71$ et $Q(99\%) = 0.76$, donc : $Q_{\text{calculé}} < Q_{\text{Table}}$ et on ne rejette pas de ce point.

Exemple

Peut-on rejeter les valeurs extrêmes de la série?

5.00 5.10 5.10 5.15 5.20 5.30 6.20

Test pour la valeur la plus petite et la plus grande :

$$Q_1 = \frac{|5.00 - 5.10|}{|6.20 - 5.00|} = 0.083 \qquad Q_7 = \frac{6.20 - 5.30}{6.20 - 5.00} = 0.75$$

$N = 7$, $Q(95\%) = 0.57$ et $Q(99\%) = 0.68$, donc $Q_1 < Q(\text{Table})$ et $Q_7 > Q(\text{Table})$, donc il faut rejeter le point 7, c.à.d. 6.20.

7.4 Test u

Ce test est utilisé pour comparer la valeur expérimentale avec une valeur vraie quand l'écart type de la population est connue.

L'hypothèse à tester $H_0 : \mu_x = \mu_0$
 contre $H_1 : \mu_x \neq \mu_0$

Il faut comparer la valeur :

$$u = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma_x} \sqrt{N} \qquad (61)$$

avec la valeur $z(\alpha)$. Quand $u < z(\alpha)$ on n'a pas des raisons de rejeter l'hypothèse H_0 .

Exemple

Selon le manufacturier la fiole jaugée a le volume $V_0 = 100.0$ ml. Pour vérifier cela, on a mesuré le volume $N = 5$ fois et on a obtenu les résultats suivants : 99,89 ml, 100,42 ml, 100,11 ml, 99,96 ml et 100,33 ml. Le déviation standard de calibrage (mesuré séparément) $\sigma_x = 0.2$ ml. Est-ce qu'on peut dire que le volume de la fiole est de 100.0 ml au niveau de confiance de 95% ?

L'hypothèse initiale **$H_0 : V = 100.0$ ml**
et l'hypothèse alternative **$H_1 : V \neq 100.0$ ml**

Le volume moyen :

$$\bar{V} = \frac{99,98 + 100,42 + 100,11 + 99,96 + 100,33}{5} = 100,142 \text{ ml}$$

et la valeur de variable u :

$$u = \frac{100,142 - 100,00}{0,2} \sqrt{5} = 1,59$$

Cette valeur doit être comparée avec la valeur de $z(0,95) = 1,96$.

Ceci nous indique que :

$u = 1,59 < z(0,95) = 1,96$. Donc, il n'y a pas de raisons de rejeter l'hypothèse H_0 , et le volume de la fiole égale 100.0 mL.

7.5 Test t

7.5.1 Comparaison avec le standard

Ce test est utilisé pour comparer la valeur expérimentale avec une valeur vraie quand l'écart type de la population n'est pas connue. C'est un test de la tendance ("bias"), pour détecter des erreurs systématiques.

L'hypothèse à tester $H_0: \bar{x} = \mu_0$

contre $H_1: \bar{x} \neq \mu_0$

Il faut comparer la valeur de t calculée comme suit :

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{\bar{x} - \mu}{s_x} \sqrt{N} \quad (62)$$

avec celle des tables : $t(\alpha, k)$. **Quand $t_{\text{exp}} < t(\alpha, k)$ on n'a pas de raisons de rejeter H_0 .**

Exemple

La valeur vraie $\mu = 0.123$. Quatre mesures ont donné les valeurs suivantes: 0.112, 0.118, 0.115 et 0.119. Est ce qu'on peut dire que la moyenne est égale à la valeur vraie ?

donc $\bar{x} = 0.116$, $s = 0.0032$, et $\bar{x} - \mu = -0.007$, $t = 0.007 * \sqrt{4} / 0.0032 = 4.38$

Pour le niveau de confiance $\alpha = 95\%$, $t(\alpha, N-1) = t(0.05, 3) = 3.18$ et pour $\alpha = 0.99$, $t(0.01, 3) = 5.84$.

Donc, pour le niveau de confiance de 95% $\alpha = 0.05$, $t(\text{calculé}) = 4.38 > t(\alpha, k)$ (tables) = 3.18 et l'hypothèse H_0 doit être rejetée, c'est à dire il y a une erreur systématique.

Pour $\alpha = 0.99$, $t(\text{calculé}) = 4.38 < t(\text{table}) = 5.84$ et l'hypothèse H_0 ne peut pas être rejetée et il n'y a pas d'erreur systématique.

7.5.2 Comparaison de deux moyennes

Souvent on pose la question suivante : **est-ce que les deux méthodes expérimentales différentes nous donnent la même réponse?**

Pour vérifier cela on peut utiliser le test t . Il y a deux tests : a) quand les variances de deux méthodes sont statistiquement identiques et b) quand les variances sont différentes.

Test d'égalité des espérances: deux observations de variances identiques

Dans ce cas : $H_0 : \mu_1 = \mu_2$ et $H_1 : \mu_1 \neq \mu_2$. On calcule :

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \quad \text{avec} \quad s^2 = \frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2} \quad (63)$$

et nombre de degrés de liberté $N_1 + N_2 - 2$. Il faut comparer la valeur de t calculée avec celle des tables $t(\alpha, N_1 + N_2 - 2)$. Si $t_{\text{exp}} < t_{\text{table}}$ on garde H_0 .

Test d'égalité des espérances: deux observations de variances différentes

Dans ce cas on calcule la valeur suivante :

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \quad (64)$$

ou D = nombre de liberté calculé selon la formule donnée ci-dessous (formule utilisée dans l'Excel) :

$$D = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\left(\frac{s_1^2}{N_1}\right)^2 \frac{1}{N_1 - 1} + \left(\frac{s_2^2}{N_2}\right)^2 \frac{1}{N_2 - 1}} \quad (65)$$

On teste une l'hypothèse $H_0 : \mu_1 = \mu_2$
 contre $H_1 : \mu_1 \neq \mu_2$.
 en comparant la valeur de t calculée avec celle des tables $t(\alpha, D)$.

Note : Il existe une autre formule pour estimer le nombre de degrés de liberté, développé par d'autres chercheurs:

$$D = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\left(\frac{s_1^2}{N_1}\right)^2 \frac{1}{N_1 + 1} + \left(\frac{s_2^2}{N_2}\right)^2 \frac{1}{N_2 + 1}} - 2$$

Exemple

On a déterminé la teneur en eau d'un produit en utilisant deux méthodes différentes, possédant les variances différentes. Est-ce qu'il y a de différences systématiques entre les deux méthodes ? Utilisez les résultats suivants.

i	x_{1i}	x_{2i}
1	6.4	6.5
2	6.2	6.7
3	6.2	6.5
4	6.5	6.1

5	6.3	6.0
6	6.4	6.8
7	6.4	6.2
\bar{x}	6.34	6.40
s_x	0.11	0.31

Les calculs en utilisant les équations (64) et (65) donnent :

$t = 0.46$, $D=8.16 \cong 8$; de Tables $t(0.05, 8) = 2,31$, donc $t(\text{expérimentale}) < t(0.05, 8)$ l'hypothèse H_0 est accepté : il n'y a pas de différence entre les résultats de deux méthodes.

On peut aussi utiliser l'Excel :

Test d'égalité des espérances: deux observations de variances différentes

	Variable 1	Variable 2
Moyenne	6.342857	6.4
Variance	0.012857	0.093333
Observations	7	7
Différence hypothétique des moyennes	0	
Degré de liberté	8	
Statistique t	-0.46395	
P(T<=t) unilatéral	0.327524	
Valeur critique de t (unilatéral)	1.859548	
P(T<=t) bilatéral	0.655049	
Valeur critique de t (bilatéral)	2.306006	

7.6 Test F

Ce test est utilisé pour la comparaison des variances. On pose la question suivante : Est-ce qu'une méthode a une déviation standard plus petite que l'autre ?

$$H_0: s_1^2 = s_2^2$$

$$H_1: s_1^2 \neq s_2^2$$

On doit déterminer la valeur de la fonction F de Fischer-Snydecor définie comme :

$$F = \frac{s_2^2}{s_1^2} \geq 1 \quad (66)$$

où s_1 et s_2 sont les déviations standards de deux méthodes. Les valeurs de la fonction $F(\alpha, D_2, D_1)$, où D_1 et D_2 sont les nombres de degrés de liberté des deux échantillons (égale à N_1-1 et N_2-1 pour les simples moyennes), sont montrées dans la Annexe 14.4. Si $F_{\text{exp}} > F_{\text{tables}}$ H_0 est rejetée, les deux variances sont différentes.

Remarque : Dans la Table 14.4 il y des valeurs $F(\alpha, D_2, D_1) = s_2^2 / s_1^2$. Dans l'Excel on peut trouver les valeurs de F comme : $F(\alpha, D_2, D_1) = \text{INVERSE.LOIF}(\alpha, D_2, D_1)$, p. ex. $\text{INVERSE.LOIF}(0.05, 5, 10) = 3.3258$; version anglaise : $\text{FINV}(\alpha, D_2, D_1)$.

Exemple du chapitre 7.5.2.

Test d'égalité des variances (F-Test)

	<i>Variable 1</i>	<i>Variable 2</i>
Moyenne	6.4	6.342857
Variance	0.093333	0.012857
Observations	7	7
Degré de liberté	6	6
F	7.259259	
P(F<=f) unilatéral	0.014682	
Valeur critique pour F (unilatéral)	4.283862	

La valeur de test F expérimentale = 7.2593 > $F(0.05,6,6) = 4.284$. Les deux échantillons ont les variances différentes.

Exemple

Le titan a été déterminé en utilisant deux méthodes différentes. Les résultats sont présentés dans le tableau suivant. Vérifiez si les deux méthodes ont la même précision.

i	$x_{1,i}$	$x_{2,i}$
1	1.00	1.12
2	1.22	1.05
3	1.29	1.19
4	1.11	1.06
5	1.10	1.10
6	1.24	1.24
\bar{x}	1.16	1.127
s_x	0.11	0.075
s_x^2	0.0121	0.00562

L'hypothèse à tester est $H_0 : s_1^2 = s_2^2$
 contre $H_1 : s_1^2 \neq s_2^2$

$F(\text{expérimentale}) = 0,0121/0,00562 = 2,15$; Table : $F(0.05,5,5)=5,05$; donc :

$F(\text{expérimentale}) < F(0.05,5,5)$ et il n'y a pas de raisons de rejeter l'hypothèse H_0 et statistiquement les deux méthodes ont la même précision.

Test d'égalité des variances (F-Test)

	<i>Variable 1</i>	<i>Variable 2</i>
Moyenne	1.16	1.126667
Variance	0.01172	0.005587
Observations	6	6
Degré de liberté	5	5

F	2.097852
P(F<=f) unilatéral	0.217712
Valeur critique pour F (unilatéral)	5.050339

$F_{\text{exp}} = 2.098 < F(0.05, 5, 5) = 5.050$, les variances sont identiques.

Test d'égalité des espérances: deux observations de variances égales

	Variable 1	Variable 2
Moyenne	1.16	1.126667
Variance	0.01172	0.005587
Observations	6	6
Variance pondérée	0.008653	
Différence hypothétique des moyennes	0	
Degré de liberté	10	
Statistique t	0.620651	
P(T<=t) unilatéral	0.274356	
Valeur critique de t (unilatéral)	1.812462	
P(T<=t) bilatéral	0.548712	
Valeur critique de t (bilatéral)	2.228139	

$t_{\text{exp}} = 0.62065 < t(0.05, 10) = 2.2281$, les moyennes sont identiques.

8. Tests statistiques pour une régression

8.1 Rejet d'un point dans la régression

La différence entre la valeur d'un point, y_i , et sa valeur calculée, \hat{y}_i est Δ :

$$\Delta = y_i - \hat{y}_i$$

L'hypothèse à tester H_0 : on garde ce point

contre H_1 : on rejette ce point

$$s_{\Delta}^2 = s_y^2 + s_{\hat{y}_i}^2 = s_y^2 \left(1 + \frac{1}{N} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right)$$

$$s_{\Delta} = s_y \sqrt{1 + \frac{1}{N} + \frac{(x_i - \bar{x})^2}{S_{xx}}} \quad (67)$$

$$s_{\Delta} = s_y \sqrt{\frac{1}{m} + \frac{1}{N} + \frac{(x_i - \bar{x})^2}{S_{xx}}} \quad (68)$$

Il faut comparer la valeur t calculée :

$$t = \frac{|y_i - \hat{y}_i|}{s_{\Delta}} = \frac{|\Delta|}{s_{\Delta}} \quad (69)$$

avec celle des tables : $t(k, \alpha)$ et $k = N - 2$. Si $t(\text{expérimentale}) > t(k, \alpha)$ il faut rejeter ce point et accepter H_1 . Cela veut dire que si notre valeur expérimentale est à l'extérieur de l'intervalle $\hat{y}_i \pm s_{\Delta} t(\alpha, k)$ (voir le graphique dans le chapitre 6.3) on peut rejeter ce point.

8.2 Importance des paramètres de la régression

Dans le cas de la régression linéaire nous supposons que les données expérimentales peuvent être décrites par une équation :

$$y = b_0 + b_1 x.$$

Cependant, on peut aussi proposer deux autres modèles :

$$1) y = b_0 = \bar{y} \text{ et}$$

$$2) y = b_1 x.$$

Dans le premier cas nous supposons que nos données peuvent être modélisées par une moyenne (la pente n'est pas importante) et dans d'autre cas nous supposons que l'ordonnée à l'origine n'est pas importante.

Dans la modélisation des données expérimentales il faut toujours garder le nombre des paramètres ajustables au minimum!

Cela veut dire, qu'il faut trouver si un terme additionnel est vraiment nécessaire dans l'équation de la régression. Pour faire un choix, on peut utiliser le test t ou le test F .

8.2.1 Test t d'importance des paramètres de la régression

Pour tester l'importance du paramètre b_1 on fait deux hypothèses :

$$H_0: \quad b_1 = 0 \quad y = b_0$$

$$\text{et } H_1: \quad b_1 \neq 0. \quad y = b_0 + b_1 x$$

On peut définir le test t de Student comme :

$$t = \frac{b_1}{s_{b_1}} \quad (70)$$

et comparer la valeur de $t_{\text{calculé}}$ avec $t(\alpha, N-2)$ des tables. **Si notre valeur t calculée est plus grande que t_{tab} (sa valeur est beaucoup plus grande que son écart-type), b_1 est statistiquement important et on doit rejeter l'hypothèse H_0 , et accepter le paramètre b_1 .**

De la même façon on teste l'importance de terme b_0 (la ligne de régression passe par l'origine) :

	$H_0:$	$b_0 = 0$	$y = b_1x$
contre	$H_1:$	$b_0 \neq 0.$	$y = b_0 + b_1x$

$$t = \frac{b_0}{s_{b_0}} \quad (71)$$

et on compare la valeur de $t_{\text{calculé}}$ avec $t(\alpha, N-2)$ des tables. **Si t (expérimentale) < $t(\alpha, N-2)$ on garde l'hypothèse H_0 et rejette H_1 : le paramètre b_0 n'est pas statistiquement significatif.**

8.2.2 Test F d'importance des paramètres de la régression

Ici on peut utiliser le test F de comparaison des variances : l'addition d'un terme additionnel doit améliorer la somme des carrés de manière importante. Supposons, que:

$H_0:$	$b_1 = 0$	$y = b_0 = \bar{y}$
$H_1:$	$b_1 \neq 0$	$y = b_0 + b_1x$

On détermine la somme des carrés,

S_1^2 (somme des carrés hors la moyenne) pour une approximation : $y = \hat{y} = b_0$, $S_1^2 = \sum (y_i - \bar{y})^2$, nombre de degrés de liberté $N - 1$, et

S_2^2 (somme des carrés résiduelle) pour : $y = b_0 + b_1x$, $S_2^2 = \sum (y_i - \hat{y}_i)^2$, nombre de degrés de libertés $N - 2$.

Dans ce cas $S_1^2 = \sum (y_i - b_0)^2 = \sum (y_i - \bar{y})^2$ (la moyenne décrit l'ensemble des donnés, $b_0 \equiv \bar{y}$).

Ensuite on calcule la valeur de F , c.à.d. le rapport des deux variances :

$$F = \frac{s_1^2}{s_2^2} = \frac{\frac{S_1^2 - S_2^2}{1}}{\frac{S_2^2}{N - 2}} = \frac{S_1^2 - S_2^2}{s_y^2} \quad (72)$$

où :

$$s_1^2 = \frac{S_1^2 - S_2^2}{(N - 1) - (N - 2)} = \frac{S_1^2 - S_2^2}{1} = S_1^2 - S_2^2 \quad (73)$$

$$s_2^2 = \frac{S_2^2}{N - 2} \quad (74)$$

Quand on augmente le nombre des paramètres dans la régression (ici d'un (b_0) à deux (b_0 et b_1)) la somme des carrés diminue. Cette diminution, c.à.d. la différence $S_1^2 - S_2^2$ doit être statistiquement importante en comparaison avec l'écart type résiduelle (de la régression) s_y^2 . Cette différence possède $(N-1)-(N-2) = 1$ degré de liberté. On compare F_{calc} avec $F(\alpha, 1, N-2)$. Quand le valeur $F_{\text{calc}} > F_{\text{table}}$ on accepte l'hypothèse H_1 , c.à.d. b_1 est important. Ce test est formellement identique avec le test t , parce que : $F(\alpha, 1, D_1) = t^2(\alpha, D_1)$. Cette valeur peut être trouvé dans la table « Analyse de variance » ou ANOVA.

De la même façon on procède pour tester l'importance de b_0 .

$$\begin{array}{lll} H_0: & b_0 = 0 & y = b_1x \\ \text{contre} & H_1: & b_0 \neq 0 & y = b_0 + b_1x \end{array}$$

Le test F est plus générale et permet de tester l'importance de plus que deux paramètres dans la courbe de régression (nécessaire dans certains cas).

8.3 Analyse des variances

On peut exprimer la somme des carrés de la différence entre les valeurs expérimentales et leur moyenne, sous forme d'une somme de deux termes comme suit :

$$\sum_{i=1}^N (y_i - \bar{y})^2 = \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \sum_{i=1}^N (\hat{y}_i - \bar{y})^2 \quad (75)$$

Cela signifie :

$$\begin{aligned} (\sum \text{ des carrés hors de la moyenne}) &= (\sum \text{ des carrés hors de la régression}) \\ &+ (\sum \text{ des carrés dans la régression}) \end{aligned}$$

$$\begin{aligned} \text{Variance totale} &= \text{Variance inexpliqué par la régression} + \text{Variance expliquée par la régression} \\ &(\text{résidu}) \qquad \qquad \qquad (\text{régression}) \end{aligned}$$

Nombre de degrés de liberté de ces termes est:

$$(N-1) = (N-2) + 1$$

On construit le tableau des variances :

Source	Degré de liberté	Somme des carrés	Moyenne des carrés	Test F
Régression	1	$\sum (\hat{y}_i - \bar{y})^2$	$MS_R = \frac{\sum (\hat{y}_i - \bar{y})^2}{1}$	$F = \frac{MS_R}{s_y^2}$
Résidus	$N - 2$	$SS = \sum (y_i - \hat{y}_i)^2$	$s_y^2 = \frac{SS}{N - 2}$	
Total par rapport de la moyenne	$N - 1$	$\sum (y_i - \bar{y})^2$		

Test F présenté ci-dessus est un test d'importance de paramètre b_1 , c.à.d. compare des hypothèses :

$$H_0: b_1 = 0, y = b_0 = \bar{y}$$

$$H_1: b_1 \neq 0, y = b_0 + b_1x$$

Cette valeur doit être comparée avec $F(\alpha, 1, N-2)$. Si la valeur expérimentale est plus grande que la valeur calculée, l'ajout de la pente améliore de la manière importante l'approximation.

Il est donné comme : $F = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \hat{y}_i)^2}$. On peut remarquer, que

$\sum_{i=1}^N (y_i - \bar{y})^2 = \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \sum_{i=1}^N (\hat{y}_i - \bar{y})^2$, donc : $\sum_{i=1}^N (y_i - \bar{y})^2 - \sum_{i=1}^N (y_i - \hat{y}_i)^2 = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2$ et cette formule est équivalente à l'Éq. (72).

Exemple d'Excel :

ANALYSE DE VARIANCE

	Degré de liberté	Somme des carrés	Moyenne des carrés	F
Régression	1	1.0041827	1.00418273	58.6725
Résidus	9	0.1540355	0.01711505	
Total	10	1.1582182		

SS "sum of squares", somme des carrés

MS_R "mean square", moyenne des carrée

$$\hat{y}_i - \bar{y} = b_1(x_i - \bar{x})$$

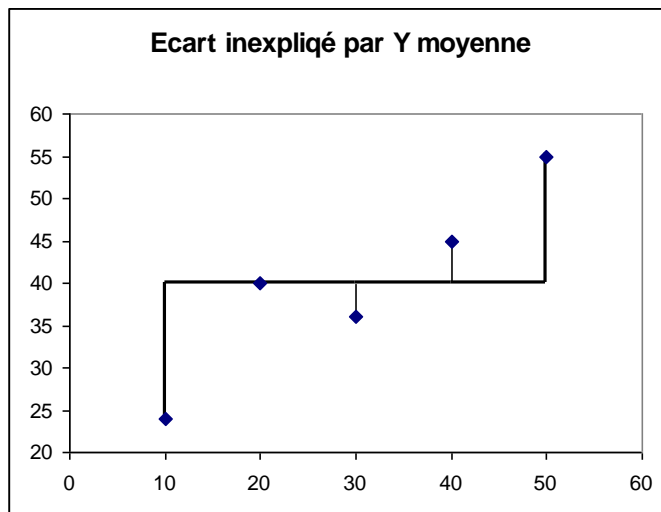
$$(\hat{y}_i - \bar{y})^2 = b_1^2(x_i - \bar{x})^2$$

$$b_1 = \frac{S_{xy}}{S_{xx}}$$

$$\sum_{i=1}^N (\hat{y}_i - \bar{y})^2 = b_1^2 \sum_{i=1}^N (x_i - \bar{x})^2 = b_1^2 S_{xx} = b_1 S_{xy}$$

$$\sum_{i=1}^N (y_i - \bar{y})^2 = S_{yy} = \sum_{i=1}^N y_i^2 - \frac{\left(\sum_{i=1}^N y_i\right)^2}{N}$$

L'analyse des variances est expliquée dans les graphiques suivants.



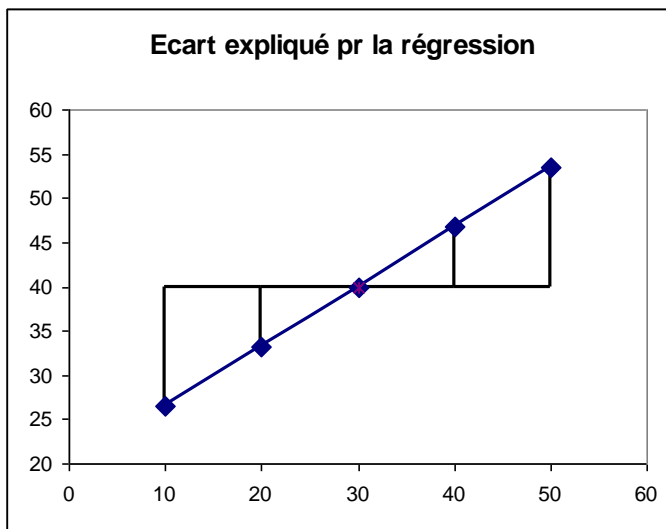
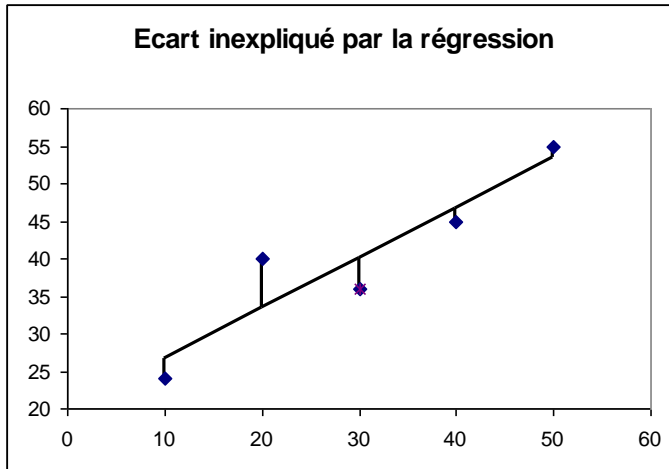


Figure 17. Illustration de l'Éq. (75).

Exemple

Vérifiez, si on peut décrire les données par la régression linéaire et déterminez les paramètres de cette régression. Déterminez l'importance de b_1 .

x	y
-2	0.27
-1.8	0.19
-1.6	0.36
-1.4	0.49
-1.2	0.5
-1	0.57
-0.8	0.2
-0.6	0.42
-0.4	0.57
-0.2	0.37
0	0.43

Test t de l'importance de b_1 :

$$H_0: b_1 = 0$$

$$H_1: b_1 \neq 0$$

On utilise le test t :

$t_{\text{exp}} = b_1 / s_{b_1} = 1.25$ (voit table ci-dessus), $t(0.05, 9) = 2.62$; $t_{\text{exp}} < t(0.05, 9)$, donc il n'y a pas raisons de rejeter l'hypothèse H_0 et $b_1 = 0$. Donc, les données peuvent être décrites par une moyenne : $\bar{y} = 0.397 \pm 0.090$ pour $\alpha = 0.05$.

Test F

Pour H_0 $S_1^2 = \sum_{i=1}^{11} (y_i - \bar{y})^2 = 0.1806$; pour H_1 S_2^2 est déterminé de la régression linéaire (Analyse

des variances), $S_2^2 = 0.1540$. Le test F est :

$$F_{\text{exp}} = \frac{S_1^2 - S_2^2}{s_y^2} = (0.1806 - 0.154) / 0.017115 = 1.553$$

$F(0.05, 1, 9) = 5.12$; $F_{\text{exp}} < F(0.05, 1, 9)$ donc une addition du terme b_1 n'améliore pas de l'approximation et ce terme n'est pas statistiquement important.

x	y	(yi-ymoyenne)^2
-2	0.27	0.016198347
-1.8	0.19	0.042961983
-1.6	0.36	0.001389256
-1.4	0.49	0.008598347
-1.2	0.5	0.010552893
-1	0.57	0.029834711
-0.8	0.2	0.038916529
-0.6	0.42	0.000516529
-0.4	0.57	0.029834711
-0.2	0.37	0.000743802
0	0.43	0.001071074
y moyenne=	0.397273	0.180618182
		=somme S^2

<u>Colonne1</u>	
Moyenne	0.397272727
Erreur-type	0.040521395
Médiane	0.42
Mode	0.57
Écart-type	0.134394264
Variance de l'échantillon	0.018061818
Kurstosis (Coefficient d'aplatissement)	-1.024076383
Coefficient d'assymétrie	-0.326324322
Plage	0.38
Minimum	0.19
Maximum	0.57
Somme	4.37
Nombre d'échantillons	11
Niveau de confiance(95.0%)	0.090287311

RAPPORT DÉTAILLÉ

$$F \text{ exp} = \frac{= (0.1806 - 0.154)}{0.017115} = 1.55317843$$

$$F(0.05, 1, 9) = 5.117357205$$

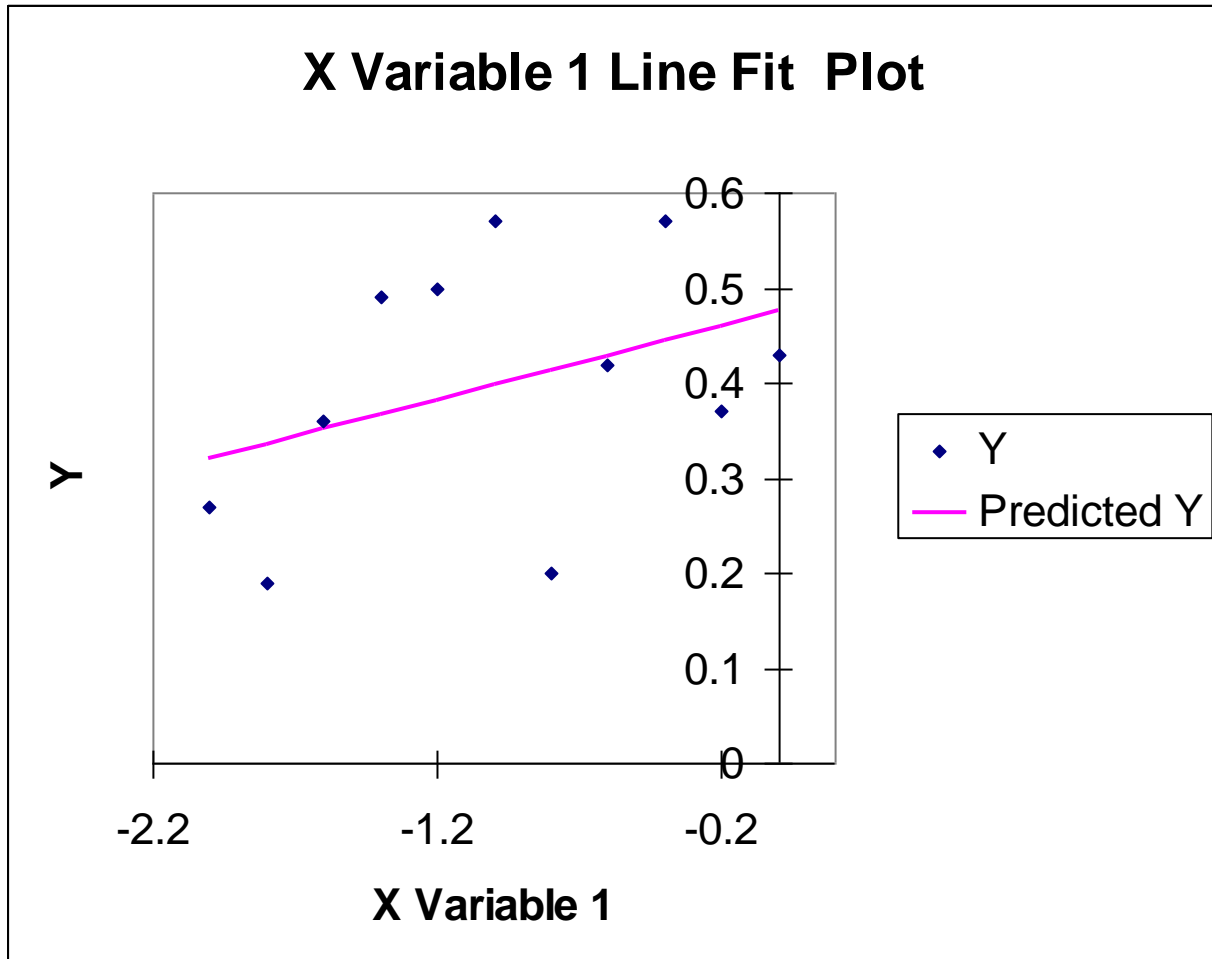
Statistiques de la régression

Coefficient de détermination multiple	0.383636	
Coefficient de détermination R^2	0.147176	=R^2
Coefficient de détermination R^2	0.052418	
Erreur-type	0.130825	
Observations	11	

ANALYSE DE VARIANCE

	Degré de liberté	Somme des carrés	Moyenne des carrés	F
Régression	1	0.026582727	0.026582727	1.55317843
Résidus	9	0.154035455	0.017115051	
Total	10	0.180618182		

	Coefficients	Erreur-type	Statistique t	Probabilité
Constante	0.475	0.073794972	6.436752902	0.000119987
Variable X 1	0.077727	0.062368135	1.246265794	0.244128754
t=b0/s0=	6.436753		t(0.05,9)=	2.262158887
t=b1/s1=	1.246266			



Exemple

Vérifiez, si on peut décrire les données par la régression linéaire et déterminez les paramètres de cette régression.

Déterminez l'importance des paramètres de la régression.

x	y
-2	-0.93
-1.8	-0.93
-1.6	-0.68
-1.4	-0.47
-1.2	-0.38
-1	-0.23
-0.8	-0.52
-0.6	-0.22
-0.4	0.01
-0.2	-0.11
0	0.03

Les résultats de l'analyse sont données dans le Tableau à la page suivante.

$$y = b_0 + b_1 x$$

$$b_0 = 0.075, s_{b_0} = 0.074$$

$$b_1 = 0.478, s_{b_1} = 0.062$$

L'importance de b_1 :

$$H_0: b_1 = 0$$

$$H_1: b_1 \neq 0$$

On utilise le test t :

$t_{\text{exp}} = b_1 / s_{b_1} = 7.66$ (voir table ci-dessus), $t(0.05, 9) = 2.62$; $t_{\text{exp}} > t(0.05, 9)$, donc l'hypothèse H_0 est rejetée et la pente est statistiquement importante.

Test de l'importance de b_0 :

$$H_0: b_0 = 0$$

$$H_1: b_0 \neq 0$$

$t_{\text{exp}} = b_0 / s_{b_0} = 1.016$, $t_{\text{exp}} < t(0.05, 9)$, donc il n'y a pas raisons de rejeter l'hypothèse H_0 et l'ordonnée à l'origine est nul. Dans ce cas il faut recalculer la régression linéaire avec l'ordonnée qui passe par l'origine (0,0) :

$$y = b_1 x$$

$$b_1 = 0.442, s_{b_1} = 0.033, \text{l'intervalles de confiance (95\%)} : 0.350 < b_1 < 0.499.$$

On peut aussi utiliser le test F pour vérifier l'importance des paramètres de la régression.

L'importance de b_1 :

$$H_0: b_1 = 0$$

$$H_1: b_1 \neq 0$$

$F_{\text{exp}} = 58.672495$ (de l'Analyse des variances); $F(0.05, 1, 9) = 5.117$, $F_{\text{exp}} > F(0.05, 1, 9)$, donc l'hypothèse H_0 est rejetée et la pente est importante.

Test F pour l'importance de b_0 : $F_{\text{exp}} = (0.1717 - 0.154) / 0.0171 = 1.033$; $F_{\text{exp}} < F(0.05, 1, 9)$, donc b_0 n'est pas important.

x	y
-2	-0.93
-1.8	-0.93
-1.6	-0.68
-1.4	-0.47
-1.2	-0.38
-1	-0.23
-0.8	-0.52
-0.6	-0.22
-0.4	0.01
-0.2	-0.11
0	0.03

RAPPORT DÉTAILLÉ

<i>Statistiques de la régression</i>	
Coefficient de détermination multiple	0.931131854
Coefficient de détermination R ²	0.86700653 =1.0042/1.1582
Coefficient de détermination R ²	0.852229478
Erreur-type	0.130824503
Observations	11

ANALYSE DE VARIANCE

	<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>
Régression	1	1.004182727	1.004182727	58.67249571
Résidus	9	0.154035455	0.017115051	
Total	10	1.158218182		

	<i>Coefficients</i>	<i>Erreur-type</i>	<i>Statistique t</i>	<i>Probabilité</i>
Constante	0.075	0.073794972	1.016329406	0.336014287
Variable X 1	0.477727273	0.062368135	7.659797367	3.12689E-05
		t(0.05,9)=	2.262158887	

RAPPORT DÉTAILLÉ

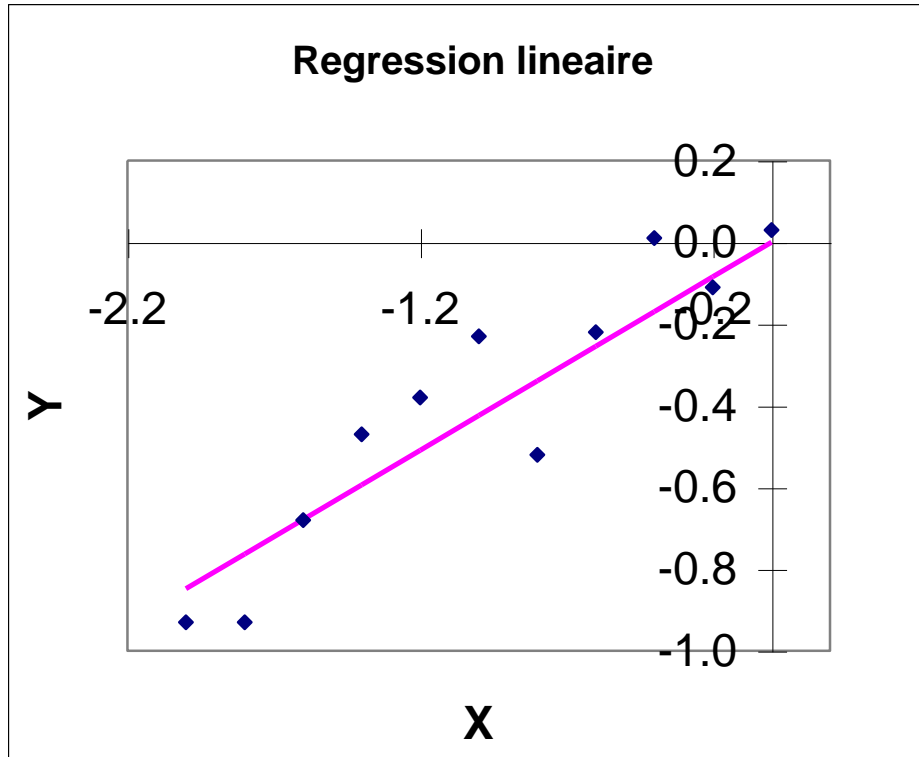
<i>Statistiques de la régression</i>	
Coefficient de détermination multiple	0.922899202
Coefficient de détermination R ²	0.851742937 =0.9865/1.1582
Coefficient de détermination R ²	0.751742937
Erreur-type	0.131039699
Observations	11

ANALYSE DE VARIANCE

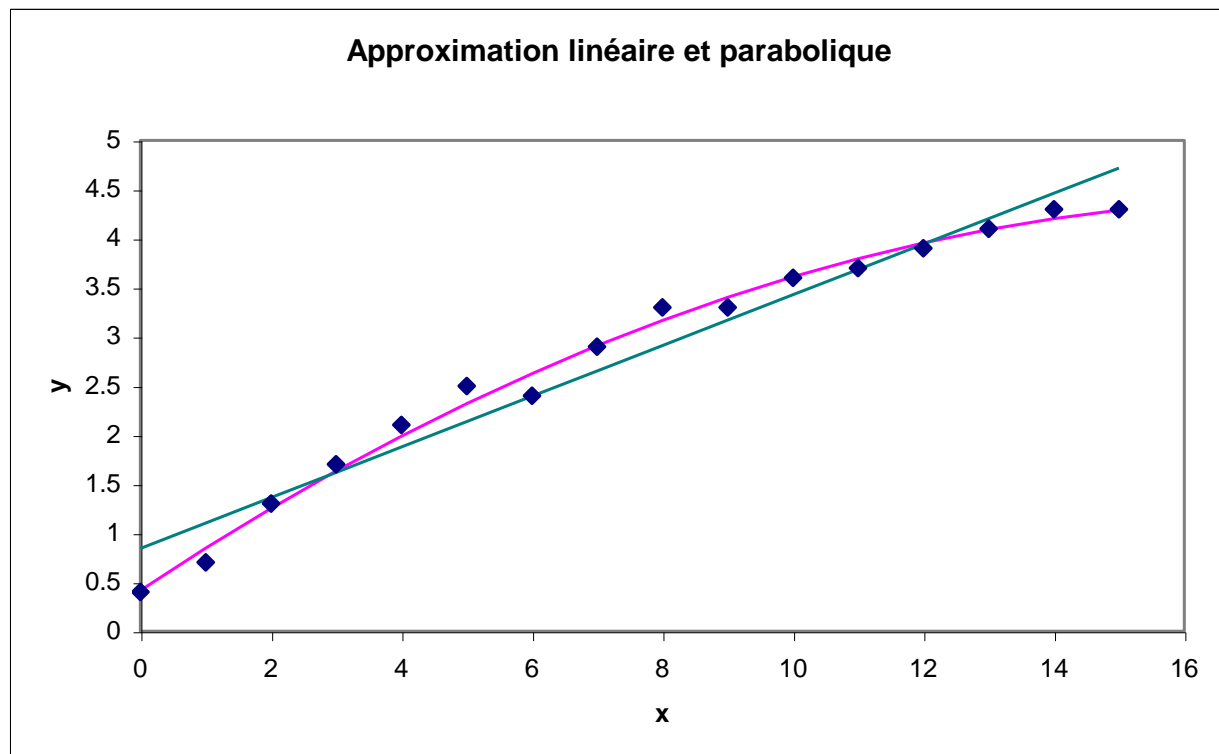
	<i>Degré de liberté</i>	<i>Somme des carrés</i>	<i>Moyenne des carrés</i>	<i>F</i>
Régression	1	0.986504156	0.986504156	57.45041212
Résidus	10	0.171714026	0.017171403	
Total	11	1.158218182		

	<i>Coefficients</i>	<i>Erreur-type</i>	<i>Statistique t</i>	<i>Probabilité</i>
Constante	0	#N/A	#N/A	#N/A
Variable X 1	0.424155844	0.033392007	12.70231652	1.70818E-07

Test F $= (0.1717 - 0.154) / 0.0171$
 $1.032925461 = t^2$
F(0.05, 1, 9) 5.117357205



Exemple : L'approximation parabolique



SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.99630542
	1
R Square	0.99262449
Adjusted R Square	1
	R 0.99148979
	7
Standard Error	0.11601702
	2
Observations	16

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	2	23.54939566	11.7747	874.7951	1.38E-14
Residual	13	0.174979342	0.01346		
Total	15	23.724375			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	0.416	0.077	5.394798	0.000122	0.24959	0.58300
						8
X Variable 1	0.441	0.024	18.49245	1.02E-10	0.38985	0.49298
						6
X Variable 2	-0.0122	0.0015	-7.94908	2.4E-06	-0.01552	-0.00889

$$t(0.05,13)= \quad \mathbf{2.16036824}$$

La valeur de test F dans le tableau : Analyse de variance correspond au test :

$$H_0: \quad b_1 = b_2 = 0, \quad y = \bar{y} = b_0$$

$$H_1: \quad b_1 \neq 0, \quad b_2 \neq 0, \quad y = b_0 + b_1x + b_2x^2$$

9. Lissage des courbes expérimentales

Les mesures expérimentales possèdent toujours certain bruit aléatoire et les courbes ne sont pas lissées. Il y a certaines techniques de lissage des courbes expérimentales. Pour effectuer le lissage il faut acquérir numériquement une courbe expérimentale aux intervalles constants.

9.1 Filtres numériques simples

Par exemple, pour lissage numérique on peut utiliser une équation suivante :

$$\hat{y}_j = \sum_{i=-m}^{i=m} \frac{C_i y_i}{N} \quad j = 1, 2, \dots, NP \quad (76)$$

où NP est le nombre des points expérimentaux, C_i est le paramètre de lissage et N est le facteur de normalisation. Quelques exemples des fonctions de lissage sont montrés dans la figure :

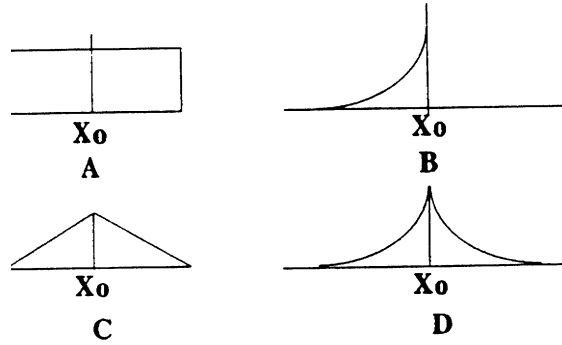


Figure 18. Exemples des filtres numériques.

Un filtre le plus simple est le filtre carré A, par exemple pour 5 points expérimentaux on a :

$$\hat{y}_j = \sum_{i=-2}^{i=2} \frac{y_i}{5} = \frac{1}{5}(y_{-2} + y_{-1} + y_0 + y_1 + y_2) \quad (77)$$

Malheureusement, ce filtre n'est pas très approprié pour la majorité des données. Un filtre exponentiel B est utilisé dans le lissage des signaux expérimentaux, car il correspond à un filtre R-C analogue. Un exemple de ce filtre est donné ci-dessous.

$$C_i = \exp(-i/3) \quad \text{pour } i = -4 \dots 0$$

$$\hat{y}_0 = \frac{1}{\sum_{i=-4}^{i=0} C_i} (C_{-4}y_{-4} + C_{-3}y_{-3} + C_{-2}y_{-2} + C_{-1}y_{-1} + C_0y_0)$$

$$C_{-4} = \exp(-4/3) = 0.26$$

$$C_{-3} = \exp(-3/3) = 0.37$$

$$C_{-2} = \exp(-2/3) = 0.51$$

$$C_{-1} = \exp(-1/3) = 0.72$$

$$C_0 = \exp(-0/3) = 1$$

(78)

Ce filtre introduit une distorsion des résultats. Un exemple pour une courbe Gaussienne est montré ci-dessous.

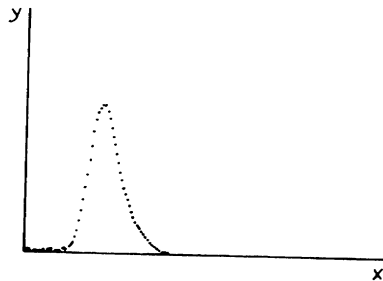


Figure 19. Déformation de la courbe gaussienne avec un filtre exponentiel.

Les filtres symétriques (C, D) sont meilleures. Ils sont construits de la même façon. Les filtres présentés ci-dessus modifient la valeur du pic et doivent être utilisés avec prudence. Il faut toujours comparer la courbe lisse avec la courbe originale.

9.2 Méthode de Savitzky-Golay

On peut utiliser la méthode des moindres carrées pour faire les lissages des courbes. On peut approximer des fragments de la courbe expérimentale par une parabole. Une parabole passe exactement par 3 points. Pour obtenir un effet de lissage il faut calculer une parabole, qui passe par 5, 7, 9, ou autre nombre des points. Deux exemples de calculs pour une parabole, qui passe par 5 et par 7 points sont présentés ici.

a) Méthode à 5 points : trouvez les paramètres b_i pour une parabole : $\hat{y} = b_2x^2 + b_1x + b_0$, qui passe les plus près de tous les points par la méthode des moindres carrées.

```
> restart;
```

```
> with(linalg);
```

Warning, the protected names norm and trace have been redefined and unprotected

[BlockDiagonal, GramSchmidt, JordanBlock, LUdecomp, QRdecomp, Wronskian, addcol, addrow, adj, adjoint, angle, augment, backsub, band, basis, bezout, blockmatrix, charmat, charpoly, cholesky, col, coldim, colspace, colspan, companion, concat, cond, copyinto, crossprod, curl, definite, delcols, delrows, det, diag, diverge, dotprod, eigenvals, eigenvalues, eigenvectors, eigenvecs, entermatrix, equal, exponential, extend, ffgausselim, fibonacci, forwardsub, frobenius, gausselim, gaussjord, geneqns, genmatrix, grad, hadamard, hermite, hessian, hilbert, htranspose, ihermite, indexfunc, innerprod, intbasis, inverse, ismith, issimilar, iszero, jacobian, jordan, kernel, laplacian, leastsqrs, linsolve, matadd, matrix, minor, minpoly, mulcol, mulrow, multiply, norm, normalize, nullspace, orthog, permanent, pivot, potential, randmatrix, randvector, rank, ratform, row, rowdim, rowspace, rowspan, rref, scalarmul, singularvals, smith, stackmatrix, submatrix, subvector, subbasis, swapcol, swaprow, sylvester, toeplitz, trace, transpose, vandermonde, vecpotent, vectdim, vector, wronskian]

Créez une matrice \mathbf{X} : $\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_1^2 \\ 1 & x_3 & x_3^2 \\ \dots & & \\ 1 & x_N & x_N^2 \end{bmatrix}$ pour les points de -2 à 2 . Le point central est $x = 0$.

```
> X:=array([[1,-2,4],[1,-1,1],[1,0,0],[1,1,1],[1,2,4]]);
```

$$X := \begin{bmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}$$

$X^T = X'$, une matrice transposée :

```
> XT:=transpose(X);
```

$$X^T := \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \\ 4 & 1 & 0 & 1 & 4 \end{bmatrix}$$

$$X^T X = X'X$$

> **XTX:=multiply(XT,X);**

$$XTX := \begin{bmatrix} 5 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 34 \end{bmatrix}$$

Matrice des valeurs pour lissage (5 points) :

Y:=array([[y_2],[y_1],[y0],[y1],[y2]]);

$$Y := \begin{bmatrix} y_2 \\ y_1 \\ y_0 \\ y_1 \\ y_2 \end{bmatrix}$$

$$X^T Y = X'Y$$

> **XTY:=multiply(XT,Y);**

$$XTY := \begin{bmatrix} y_2 + y_1 + y_0 + y_1 + y_2 \\ -2 y_2 - y_1 + y_1 + 2 y_2 \\ 4 y_2 + y_1 + y_1 + 4 y_2 \end{bmatrix}$$

Inversion de la matrice $X'X$:

> **XTX_1:=inverse(XTX);**

$$XTX_1 := \begin{bmatrix} \frac{17}{35} & 0 & \frac{-1}{7} \\ 0 & \frac{1}{10} & 0 \\ \frac{-1}{7} & 0 & \frac{1}{14} \end{bmatrix}$$

$$\mathbf{b} = (X'X)^{-1}X'Y, \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} :$$

> **b:=multiply(XTX_1,XTY);**

$$b := \begin{bmatrix} -\frac{3}{35}y_2 + \frac{12}{35}y_1 + \frac{17}{35}y_0 + \frac{12}{35}y_1 - \frac{3}{35}y_2 \\ -\frac{1}{5}y_2 - \frac{1}{10}y_1 + \frac{1}{10}y_1 + \frac{1}{5}y_2 \\ \frac{1}{7}y_2 - \frac{1}{14}y_1 - \frac{1}{7}y_0 - \frac{1}{14}y_1 + \frac{1}{7}y_2 \end{bmatrix}$$

Valeur lisse à $x = 0$:

$$\hat{y}(x=0) = b_0 = \frac{1}{35}(-3y_{-2} + 12y_{-1} + 17y_0 + 12y_1 - 3y_2)$$

$$\hat{y}'(x=0) = b_1 = \frac{1}{10}(-2y_{-2} - y_{-1} + y_1 + 2y_2)$$

$$y''(x=0) = 2b_2 = \frac{1}{7}(2y_{-2} - y_{-1} - 2y_0 - y_1 + 2y_2)$$

Méthode de Golay-Savitzky à 7 points : trouvez les paramètres b_i pour une parabole : $\hat{y} = b_2x^2 + b_1x + b_0$, qui passe les plus près de tous les points par la méthode des moindres carrés.

> **restart;**

> **with(linalg);**

Warning, the protected names norm and trace have been redefined and unprotected

[*BlockDiagonal, GramSchmidt, JordanBlock, LUdecomp, QRdecomp, Wronskian, addcol, addrow, adj, adjoint, angle, augment, backsub, band, basis, bezout, blockmatrix, charmat, charpoly, cholesky, col, coldim, colspace, colspan, companion, concat, cond, copyinto, crossprod, curl, definite, delcols, delrows, det, diag, diverge, dotprod, eigenvals, eigenvalues, eigenvectors, eigenvects, entermatrix, equal, exponential, extend, ffgausselim, fibonacci, forwardsub, frobenius, gausselim, gaussjord, geneqns, genmatrix, grad, hadamard, hermite, hessian, hilbert, htranspose, ihermite, indexfunc, innerprod, intbasis, inverse, ismith, issimilar, iszero, jacobian, jordan, kernel, laplacian, leastsqrs, linsolve, matadd, matrix, minor, minpoly, mulcol, mulrow, multiply, norm, normalize, nullspace, orthog, permanent, pivot, potential, randmatrix, randvector, rank, ratform, row, rowdim, rowspan, rref, scalarmul, singularvals, smith, stackmatrix, submatrix, subvector, subbasis, swapcol, swaprow, sylvester, toeplitz, trace, transpose, vandermonde, vecpotent, vectdim, vector, wronskian*]

Créez une matrice \mathbf{X} : $\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ \dots & & \\ 1 & x_N & x_N^2 \end{bmatrix}$ pour les points de -3 à 3 :

>**X:=array([[1,-3,9],[1,-2,4],[1,-1,1],[1,0,0],[1,1,1],[1,2,4],[1,3,9]]);**

$$X := \begin{bmatrix} 1 & -3 & 9 \\ 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}$$

> **XT:=transpose(X) ;**

$$XT := \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \\ 9 & 4 & 1 & 0 & 1 & 4 & 9 \end{bmatrix}$$

> **XTX:=multiply(XT,X) ;**

$$XTX := \begin{bmatrix} 7 & 0 & 28 \\ 0 & 28 & 0 \\ 28 & 0 & 196 \end{bmatrix}$$

Y:=array([[y_3],[y_2],[y_1],[y0],[y1],[y2],[y3]]);

$$Y := \begin{bmatrix} y_3 \\ y_2 \\ y_1 \\ y0 \\ y1 \\ y2 \\ y3 \end{bmatrix}$$

> **XTY:=multiply(XT,Y) ;**

$$XTY := \begin{bmatrix} y_3 + y_2 + y_1 + y0 + y1 + y2 + y3 \\ -3 y_3 - 2 y_2 - y_1 + y1 + 2 y2 + 3 y3 \\ 9 y_3 + 4 y_2 + y_1 + y1 + 4 y2 + 9 y3 \end{bmatrix}$$

> **XTXI:=inverse(XTX) ;**

$$XTXI := \begin{bmatrix} \frac{1}{3} & 0 & \frac{-1}{21} \\ 0 & \frac{1}{28} & 0 \\ \frac{-1}{21} & 0 & \frac{1}{84} \end{bmatrix}$$

➤ **B:=multiply(XTXI,XTY) ;**

$$B := \begin{bmatrix} -\frac{2}{21}y_{-3} + \frac{1}{7}y_{-2} + \frac{2}{7}y_{-1} + \frac{1}{3}y_0 + \frac{2}{7}y_1 + \frac{1}{7}y_2 - \frac{2}{21}y_3 \\ -\frac{3}{28}y_{-3} - \frac{1}{14}y_{-2} - \frac{1}{28}y_{-1} + \frac{1}{28}y_1 + \frac{1}{14}y_2 + \frac{3}{28}y_3 \\ \frac{5}{84}y_{-3} - \frac{1}{28}y_{-1} - \frac{1}{21}y_0 - \frac{1}{28}y_1 + \frac{5}{84}y_3 \end{bmatrix}$$

$$\hat{y}(x=0) = b_0 = \frac{1}{21}(-2y_{-3} + 3y_{-2} + 6y_{-1} + 7y_0 + 6y_1 + 3y_2 - 2y_3)$$

$$\hat{y}'(x=0) = b_1 = \frac{1}{28}(-3y_{-3} - 2y_{-2} - y_{-1} + y_1 + 2y_2 + 3y_3)$$

$$\hat{y}''(x=0) = 2b_2 = \frac{1}{42}(5y_{-3} - 3y_{-1} - 4y_0 - 3y_1 + 5y_3)$$

Les valeurs des paramètres pour d'autres nombres des points sont présentés dans les tableaux.

SAVITZKY GOLAY
Lissage des courbes

POINTS	25	23	21	19	17	15	13	11	9	7	5
-12	-253										
-11	-138	-42									
-10	-33	-21	-171								
-9	62	-2	-76	-136							
-8	147	15	9	-51	-21						
-7	222	30	84	24	-6	-78					
-6	287	43	149	89	7	-13	-11				
-5	342	54	204	144	18	42	0	-36			
-4	387	63	249	189	27	87	9	9	-21		
-3	422	70	284	224	34	122	16	44	14	-2	
-2	447	75	309	249	39	147	21	69	39	3	-3
-1	462	78	324	264	42	162	24	84	54	6	12
0	467	79	329	269	43	167	25	89	59	7	17
1	462	78	324	264	42	162	24	84	54	6	12
2	447	75	309	249	39	147	21	69	39	3	-3
3	422	70	284	224	34	122	16	44	14	-2	
4	387	63	249	189	27	87	9	9	-21		
5	342	54	204	144	18	42	0	-36			
6	287	43	149	89	7	-13	-11				
7	222	30	84	24	-6	-78					
8	147	15	9	-51	-21						
9	62	-2	-76	-136							
10	-33	-21	-171								
11	-138	-42									
12	-253										
NORME	5175	805	3059	2261	323	1105	143	429	231	21	35

Coefficients pour calculer la première dérivée.

DERIVEE liere

SAVITZKY GOLAY

POINTS	25	23	21	19	17	15	13	11	9	7	5
-12	-12										
-11	-11	-11									
-10	-10	-10	-10								
-9	-9	-9	-9	-9							
-8	-8	-8	-8	-8	-8						
-7	-7	-7	-7	-7	-7	-7					
-6	-6	-6	-6	-6	-6	-6	-6				
-5	-5	-5	-5	-5	-5	-5	-5	-5			
-4	-4	-4	-4	-4	-4	-4	-4	-4	-4		
-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	
-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3	
4	4	4	4	4	4	4	4	4	4		
5	5	5	5	5	5	5	5	5			
6	6	6	6	6	6	6	6				
7	7	7	7	7	7	7					
8	8	8	8	8	8						
9	9	9	9	9							
10	10	10	10								
11	11	11									
12	12										
NORME	1300	1012	770	570	408	280	182	110	60	28	10

Coefficients pour calculer la deuxième dérivée.
DERIVEE 2ieme

SAVITZKY GOLAY

POINTS	25	23	21	19	17	15	13	11	9	7	5
-12	92										
-11	69	77									
-10	48	56	190								
-9	29	37	133	51							
-8	12	20	82	34	40						
-7	-3	5	37	19	25	91					
-6	-16	-8	-2	6	12	52	22				
-5	-27	-19	-35	-5	1	19	11	15			
-4	-36	-28	-62	-14	-8	-8	2	6	28		
-3	-43	-35	-83	-21	-15	-29	-5	-1	7	5	
-2	-48	-40	-98	-26	-20	-44	-10	-6	-8	0	2
-1	-51	-43	-107	-29	-23	-53	-13	-9	-17	-3	-1
0	-52	-44	-110	-30	-24	-56	-14	-10	-20	-4	-2
1	-51	-43	-107	-29	-23	-53	-13	-9	-17	-3	-1
2	-48	-40	-98	-26	-20	-44	-10	-6	-8	0	2
3	-43	-35	-83	-21	-15	-29	-5	-1	7	5	
4	-36	-28	-62	-14	-8	-8	2	6	28		
5	-27	-19	-35	-5	1	19	11	15			
6	-16	-8	-2	6	12	52	22				
7	-3	5	37	19	25	91					
8	12	20	82	34	40						
9	29	37	133	51							
10	48	56	190								
11	69	77									
12	92										
NORME	26910	17710	33649	6783	3876	6188	1001	429	462	42	7

Utilisation d'une grande fenêtre des points expérimentaux peut causer une distorsion et diminution du pic. Dans la figure suivante on trouve lissage d'une courbe Gaussienne avec la méthode de Savitzky-Golay pour 5, 9, 17 et 25 points et un filtre carré (filtre A ci-dessus). On observe, que la méthode de Savitzky-Golay cause plus petite distorsion du pic que la méthode de moyennes (filtre carré).

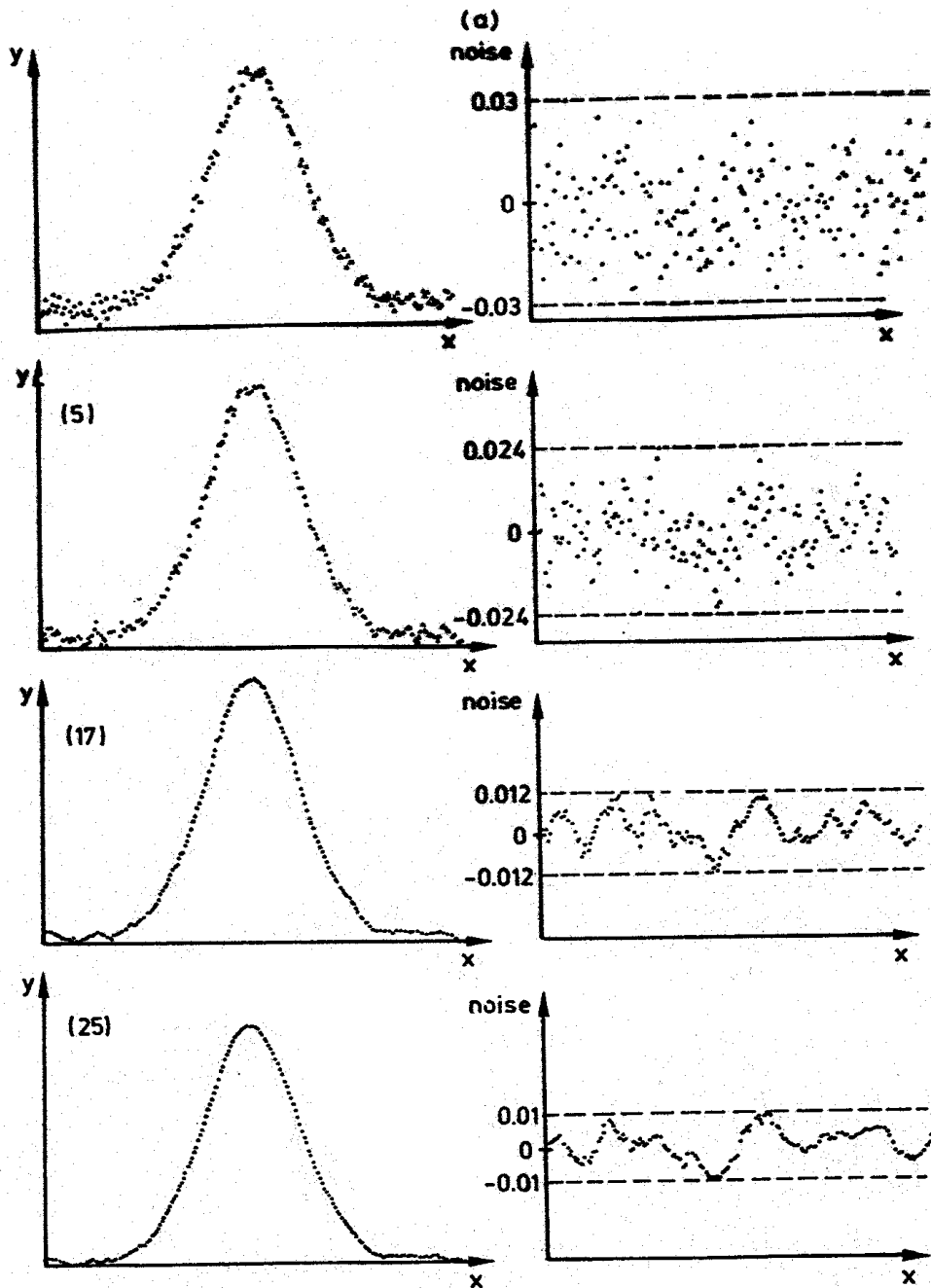


Fig. 19. (a) Five-point, 17-point, and 25-point quadratic smoothing of a Gaussian peak, noise: $N(0,3\%)$.
 (b) S/N after smoothing as a function of the window size.

Figure 20. (a) Lissage d'une courbe de Gauss contenant 0.3% de bruit, en utilisant des fenêtres de 5, 17 et 25 points, (b) le rapport signal/bruit après lissage.

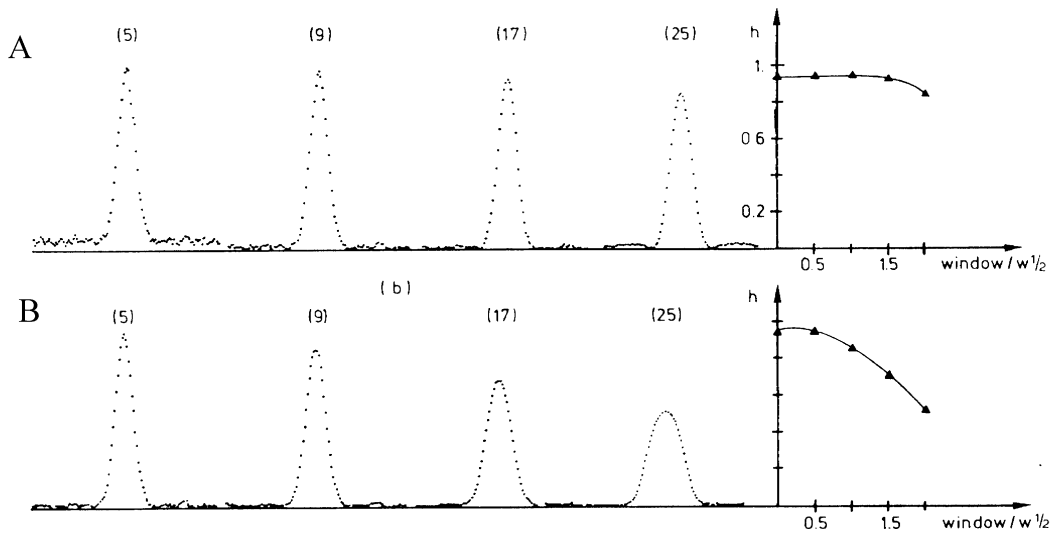


Figure 21. Filtrés : A – Savitzky Golay, B – carré pour (n) points; h est une mesure de la distorsion.

On peut aussi approximer les courbes expérimentales par les polynômes des plus grand ordre. Avec augmentation du ordre la distorsion diminue mais le bruit augmente.

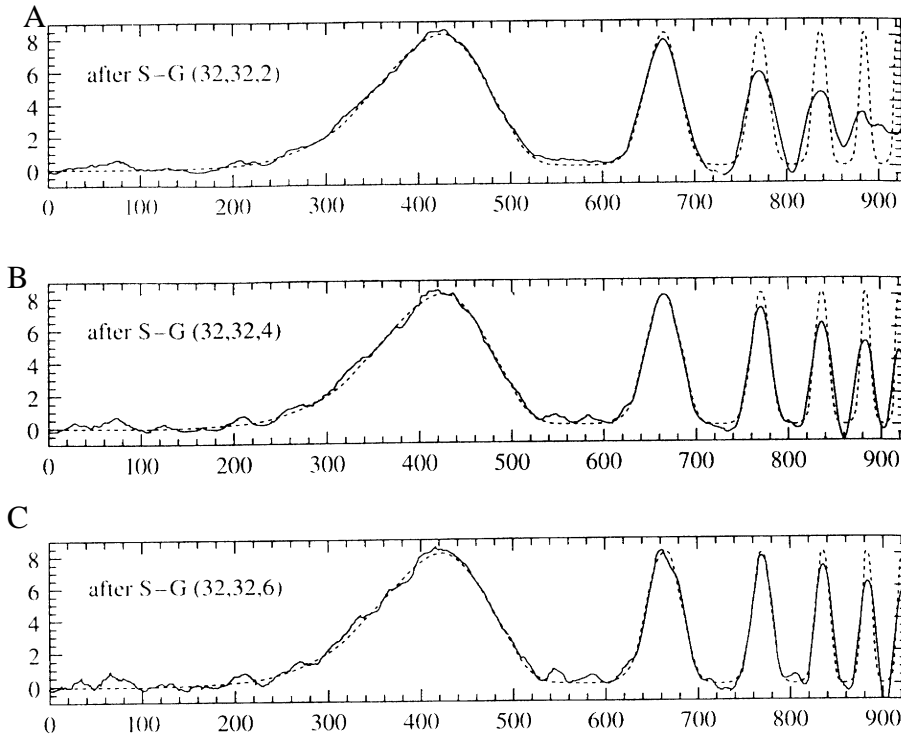


Figure 22. Lissage d'une courbe avec des filtres à 65 points. A : Savitzky-Golay, deuxième ordre, B : Savitzky-Golay, quatrième ordre, C : Savitzky-Golay sixième ordre.

10. Transformée de Fourier rapide (Fast Fourier Transform, FFT)

10.1 Théorie

Toute fonction périodique du temps $f(t)$ peut être décomposée en une somme infinie des fonctions sin et cos des fréquences différentes (en commençant à zéro) :

$$f(t) = a_0 + \sum_{k=1}^{\infty} a_k \cos(2\pi k v_1 t) + b_k \sin(2\pi k v_1 t) \quad (79)$$

où v_1 est la fréquence fondamentale et la période du signal égale $T = 1/v_1$. Le signal contient un terme de fréquence 0 (a_0), une fréquence fondamentale v_1 ($k = 1$) et des fréquences harmoniques : $k v_1$ ($k = 2, 3, \dots$). Les coefficients de la série sont déterminés à partir de :

$$a_k = \frac{2}{T} \int_0^T f(t) \cos(2\pi k v_1 t) dt \quad (80)$$

$$b_k = \frac{2}{T} \int_0^T f(t) \sin(2\pi k v_1 t) dt \quad (81)$$

La transformée de Fourier est définie comme une intégrale :

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \quad (82)$$

et elle représente une transformation de la fonction de temps, $f(t)$, en fonction de fréquence, $F(\omega)$ (son image dans l'espace de fréquences). En pratique on utilise une intégration entre $t = 0$ et T :

$$F(\omega) = \int_0^T f(t) e^{-j\omega t} dt = \int_0^T f(t) \cos(\omega t) dt + j \int_0^T f(t) \sin(\omega t) dt \quad (83)$$

L'analyse des signaux expérimentaux est effectuée à l'aide d'un ordinateur. Le signal est enregistré dans la forme numérique, en utilisant un convertisseur analogue-numérique (A/D analog to digital convertor).

Une fonction continue $f(t)$ est changée en série de N points à l'intervalle constant de Δt :

$$f(0), f(\Delta t), f(2\Delta t), \dots, f((N-1)\Delta t)$$

ou

$$f(0), f(1), f(2), \dots, f(N-1)$$

La Transformée de Fourier discrète est une transformée de la fonction de temps à une fonction de la fréquence selon la formule :

$$F(u) = \frac{1}{N} \sum_{i=0}^{N-1} f(i) \exp(-j\omega_u t_i) = \frac{1}{N} \sum_{i=0}^{N-1} f(i) \exp\left(-\frac{j2\pi u i}{N}\right) \quad (84)$$

où $\omega_u t_i = 2\pi v_u t_i = \frac{2\pi u}{T} i \Delta t = \frac{2\pi u i \Delta t}{N \Delta t} = \frac{2\pi u i}{N}$, $j = \sqrt{-1}$ et $u = 0, 1, 2, \dots, N-1$. Le terme exponentiel

est un nombre complexe et peut être représenté comme :

$$\exp(-j2\pi u i / N) = \cos(2\pi u i / N) - j \sin(2\pi u i / N) \quad (85)$$

La fréquence fondamentale est un inverse de temps de mesures : $v_1 = 1/T = 1/(N\Delta t)$ et les fréquences harmoniques sont $2v_1, 3v_1, 4v_1, \dots$. Donc, chaque fréquence est $v_u = u/(N\Delta t)$. La transformée de Fourier contient les informations pour les fréquences jusqu'à $u = N/2$, c'est à dire $f_{N/2} = N/(2N\Delta t) =$

$1/(2\Delta t)$. La plus grande fréquence s'appelle une **fréquence de Nyquist**. La transformée de Fourier est symétrique autour de point $N/2$.

Exemple

$$\nu_u = \frac{u}{N\Delta t} = \frac{u}{T} \quad u = 0, \dots, \frac{N}{2}$$

$$T = N\Delta t = 8 * 0.02 \text{ s} = 0.16 \text{ s} \quad \nu_1 = \frac{1}{T} = \frac{1}{0.16} \text{ s}^{-1} = 6.25 \text{ s}^{-1}$$

$$f = \cos(2\pi t / 0.08 + \pi / 3) \quad \nu = 1 / 0.08 = 12.5 \text{ s}^{-1}$$

i	t/s	f(t)	u	freq/Hz	F'	F''
0	0	0.5	0	0	0	0
1	0.02	-0.866	1	6.25	0	0
2	0.04	-0.5	2	12.5	0.25	0.433
3	0.06	0.866	3	18.75	0	0
4	0.08	0.5	4	25	0	0
5	0.1	-0.866	5	31.25	0	0
6	0.12	-0.5	6	37.5	0	0
7	0.14	0.866	7	43.75	0.25	-0.43

Propriétés des nombres complexes :

$$j^2 = -1$$

$$j^3 = -j$$

$$j^4 = 1$$

$$\exp(j\theta) = \cos(\theta) + j \sin(\theta)$$

Chaque nombre complexe peut être représenté comme :

$$z = x + jy = r \exp(j\theta) = r \cos(\theta) + jr \sin(\theta)$$

$$z^2 = (x + jy)(x - jy) = x^2 + y^2$$

et les fonctions trigonométriques :

$$\cos(\theta) = \frac{1}{2} (e^{j\theta} + e^{-j\theta})$$

$$\sin(\theta) = \frac{1}{2} (e^{j\theta} - e^{-j\theta})$$

Exemple

Calculez la transformée de Fourier de la fonction $\cos(2\pi t f_0)$, avec $\Delta t = 1 \text{ s}$ et la fréquence fondamentale $f_0 = 1/(N\Delta t) = 1/(4*1 \text{ s}) = 0.25 \text{ s}^{-1}$.

Au début il faut générer la fonction, avec $t = i\Delta t$, $\cos(2\pi t f_0) = \cos(2\pi i * 0.25)$:

t/s	i	$\cos(\pi i/2)$
0	0	1
1	1	0
2	2	-1
3	3	0

Calculs de la transformée de Fourier :

$$\begin{aligned}
 F(0) &= \frac{1}{4} \sum_{i=0}^3 f(i) \exp(-j2\pi 0i/4) \\
 &= \frac{1}{4} \sum_{i=0}^3 f(i) \exp(0) = \frac{1}{4} \sum_{i=0}^3 f(i) \\
 &= \frac{1}{4} (1 + 0 - 1 + 0) = 0
 \end{aligned}$$

$$\begin{aligned}
 F(1) &= \frac{1}{4} \sum_{i=0}^3 f(i) \exp(-j2\pi 1i/4) \\
 &= \frac{1}{4} \sum_{i=0}^3 f(i) \exp(-j\pi i/2) \\
 &= \frac{1}{4} (1 \exp(0) + 0 \exp(-j\pi/2) - 1 \exp(-j\pi 2/2) + 0 \exp(-j\pi 3/2)) \\
 &= \frac{1}{4} (1 + 0 - 1(-1) + 0) = \frac{1}{2}
 \end{aligned}$$

$$\begin{aligned}
 F(2) &= \frac{1}{4} \sum_{i=0}^3 f(i) \exp(-j2\pi 2i/4) \\
 &= \frac{1}{4} \sum_{i=0}^3 f(i) \exp(-j\pi i) \\
 &= \frac{1}{4} (1 \exp(0) + 0 \exp(-j\pi) - 1 \exp(-j\pi 2) + 0 \exp(-j\pi 3)) \\
 &= \frac{1}{4} (1 + 0 - 1(+1) + 0) = 0
 \end{aligned}$$

$$\begin{aligned}
 F(3) &= \frac{1}{4} \sum_{i=0}^3 f(i) \exp(-j2\pi 3i/4) \\
 &= \frac{1}{4} \sum_{i=0}^3 f(i) \exp(-j\pi 3i/2) \\
 &= \frac{1}{4} (1 \exp(0) + 0 \exp(-j\pi 3/2) - 1 \exp(-j\pi 6/2) + 0 \exp(-j\pi 9/2)) \\
 &= \frac{1}{4} (1 + 0 - 1(-1) + 0) = \frac{1}{2}
 \end{aligned}$$

La transformée de Fourier est :

$F(0) = 0$, $F(1) = 0.5$, $F(2) = 0$, $F(3) = 0.5$, pour les fréquences : 0, 0.25 et 0.5 s^{-1} et 0.25 s^{-1} . La fréquence maximale est obtenue pour le troisième point. Ici, la transformée de Fourier contient les valeurs réelles seulement.

Exemple

Calculez la transformée de Fourier de la fonction $\sin(2\pi t f_0)$, avec $\Delta t = 1$ s et $f_0 = 0.25$ s^{-1} (la fréquence fondamentale) et $T_0 = 4$ s.

Au début il faut générer la fonction, avec $t = i\Delta t$, $\sin(2\pi t f_0) = \sin(2\pi i * 0.25)$:

t / s	i	$\sin(\pi i / 2)$
0	0	0
1	1	1
2	2	0
3	3	-1

Calculs de la transformée de Fourier :

$$\begin{aligned} F(0) &= \frac{1}{4} \sum_{i=0}^3 f(i) \exp(-j2\pi 0i / 4) \\ &= \frac{1}{4} \sum_{i=0}^3 f(i) \exp(0) = \frac{1}{4} \sum_{i=0}^3 f(i) \\ &= \frac{1}{4} (0 + 1 + 0 - 1) = 0 \end{aligned}$$

$$\begin{aligned} F(1) &= \frac{1}{4} \sum_{i=0}^3 f(i) \exp(-j2\pi 1i / 4) \\ &= \frac{1}{4} \sum_{i=0}^3 f(i) \exp(-j\pi i / 2) \\ &= \frac{1}{4} (0 \exp(0) + 1 \exp(-j\pi / 2) + 0 \exp(-j\pi 2 / 2) - 1 \exp(-j\pi 3 / 2)) \\ &= \frac{1}{4} (0 - j + 0 - j) = -\frac{1}{2} j \end{aligned}$$

$$\begin{aligned} F(2) &= \frac{1}{4} \sum_{i=0}^3 f(i) \exp(-j2\pi 2i / 4) \\ &= \frac{1}{4} \sum_{i=0}^3 f(i) \exp(-j\pi i) \\ &= \frac{1}{4} (0 \exp(0) + 1 \exp(-j\pi) + 0 \exp(-j\pi 2) - 1 \exp(-j\pi 3)) \\ &= \frac{1}{4} (0 - 1 + 0 - 1 * (-1)) = 0 \end{aligned}$$

$$\begin{aligned}
 F(3) &= \frac{1}{4} \sum_{i=0}^3 f(i) \exp(-j2\pi 3i/4) \\
 &= \frac{1}{4} \sum_{i=0}^3 f(i) \exp(-j\pi 3i/2) \\
 &= \frac{1}{4} (0 \exp(0) + 1 \exp(-j\pi 3/2) + 0 \exp(-j\pi 6/2) - 1 \exp(-j\pi 9/2)) \\
 &= \frac{1}{4} (1 + j + 0 + j) = \frac{1}{2} j
 \end{aligned}$$

La transformée de Fourier est :

$F(0) = 0$, $F(1) = -j/2$, $F(2) = 0$, $F(3) = j/2$, pour les trois fréquences seulement : 0, 0.25 s⁻¹, 0.5 s⁻¹ et 0.25 s⁻¹. Ici, la transformée de Fourier contient les valeurs imaginaires seulement.

Résultats :

t/s	i	$\cos(\pi i/2)$	f/s^{-1}	FFT	
				réelle	imaginaire
0	0	1	0	0	0
1	1	0	0.25	0.5	0
2	2	-1	0.5	0	0
3	3	0	(0.25)	0.5	0

t/s	i	$\sin(\pi i/2)$	f/s^{-1}	FFT	
				réelle	imaginaire
	0	0	0	0	0
1	1	1	0.25	0	-0.5
2	2	0	0.5	0	0
3	3	-1	(0.25)	0	0.5

On définit aussi la transformée de Fourier inverse, qui permet d'obtenir la fonction initiale à partir de sa transformée de Fourier :

$$f(i) = \sum_{u=0}^{N-1} F(u) \exp\left(\frac{j2\pi ui}{N}\right) \quad (86)$$

On peut aussi représenter cette équation dans une forme :

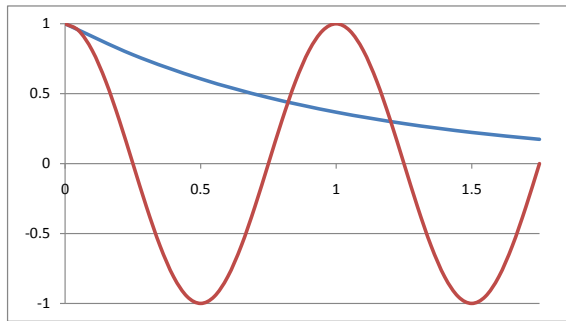
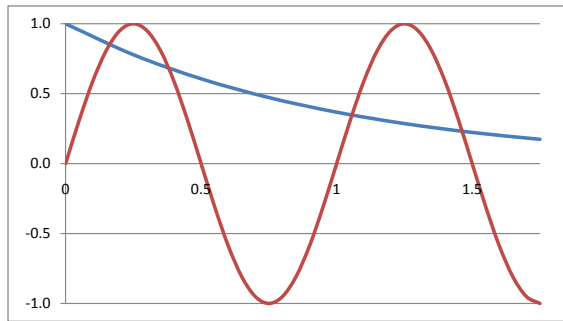
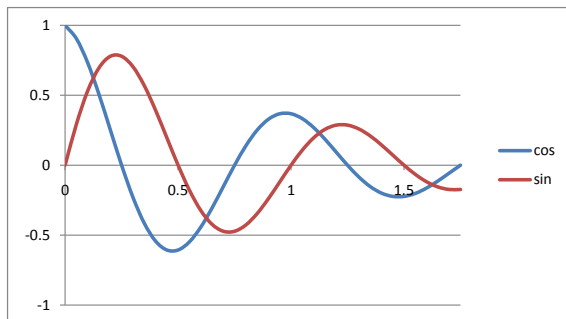
$$f(t_k) = \sum_{i=1}^{N-1} |F_i| \cos(\omega_i t_k + \varphi_i) \quad (87)$$

où $\omega = 2\pi\nu = 2\pi u/(N\Delta t)$ et $\varphi = \text{atan}\left(\frac{F''}{F'}\right)$.

Exemple

Série des 8 mesures :

$$f = \exp(-t)$$

$f(t), \cos(\omega t)$  $f(t), \sin(\omega t)$  $f(t)*\cos(\omega t); f(t)*\sin(\omega t)$ Intégration pour différent valeurs de ω

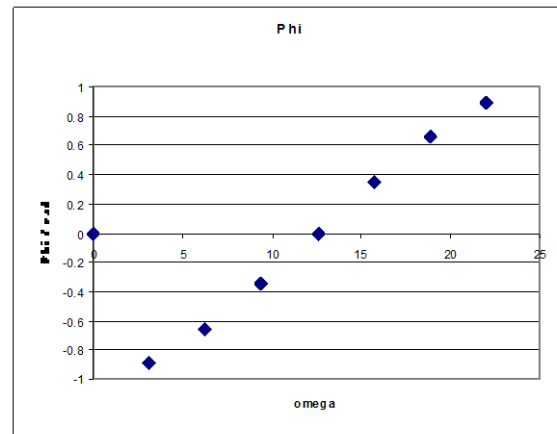
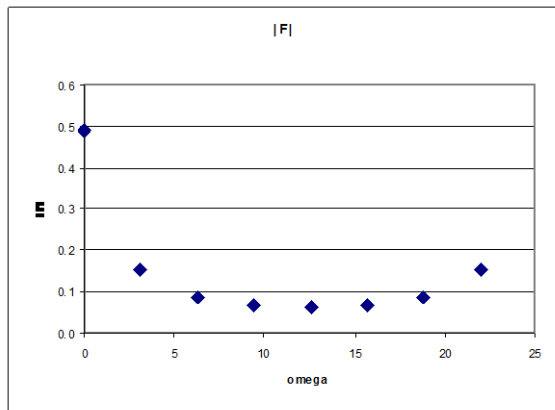
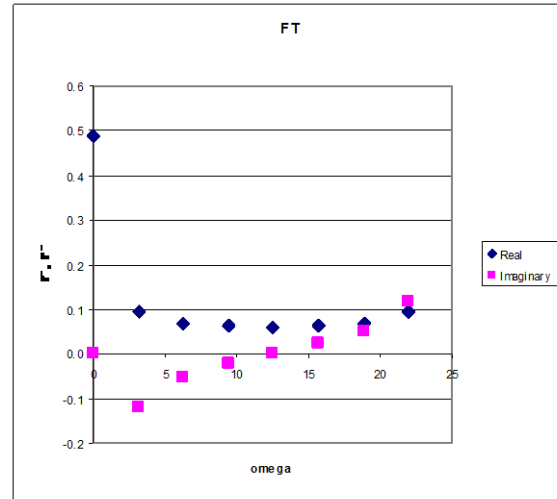
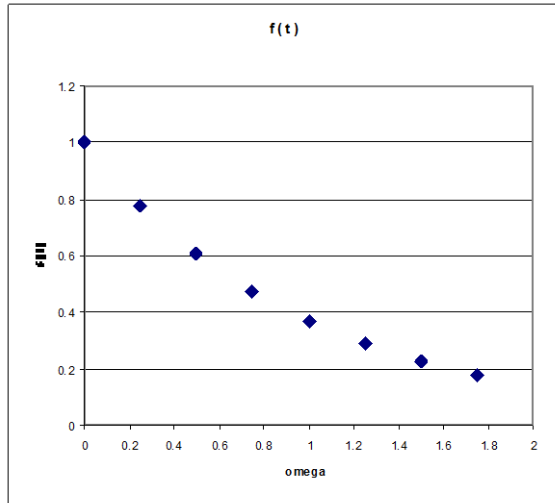
i	t/s	f	ω	Re	Im	$ F $	Phi / rad	f
0	0	1	0	0.48862	0.00000	0.48862	0	1.00000
1	0.25	0.778801	3.141593	0.09614	-0.11783	0.15207	-0.88644	0.778801
2	0.5	0.606531	6.283185	0.06728	-0.05240	0.08527	-0.66168	0.606531
3	0.75	0.472367	9.424778	0.06189	-0.02198	0.06568	-0.34124	0.472367
4	1	0.367879	12.56637	0.06076	0.00000	0.06076	0.00000	0.367879
5	1.25	0.286505	15.70796	0.06189	0.02198	0.06568	0.34124	0.286505
6	1.5	0.22313	18.84956	0.06728	0.05240	0.08527	0.66168	0.223130
7	1.75	0.173774	21.99115	0.09614	0.11783	0.15207	0.88644	0.173774

f

ici $i = 0..7$

$$f(0) = 0.47762 \cos(0+0) + 0.15207 \cos(3.14159*0-0.88644) + 0.085273 \cos(6.2818*0-0.66168) + \dots = 1.00000$$

$$f(1) = 0.47762 \cos(0+0) + 0.15207 \cos(3.14159*0.25-0.88644) + 0.085273 \cos(6.28318*0.25-0.66168) + \dots = 0.778801$$



DFT et FFT

- Discrete Fourier Transform – la transformé de Fourier discrète, N points
- Fast Fourier Transform – la transforme de Fourier rapide, l'algorithme de calculs rapide, $N = 2^k$, k nombre entier, $k \geq 2$.

10.2 Applications

- Higher sample rates allow the waveform to be more accurately represented.

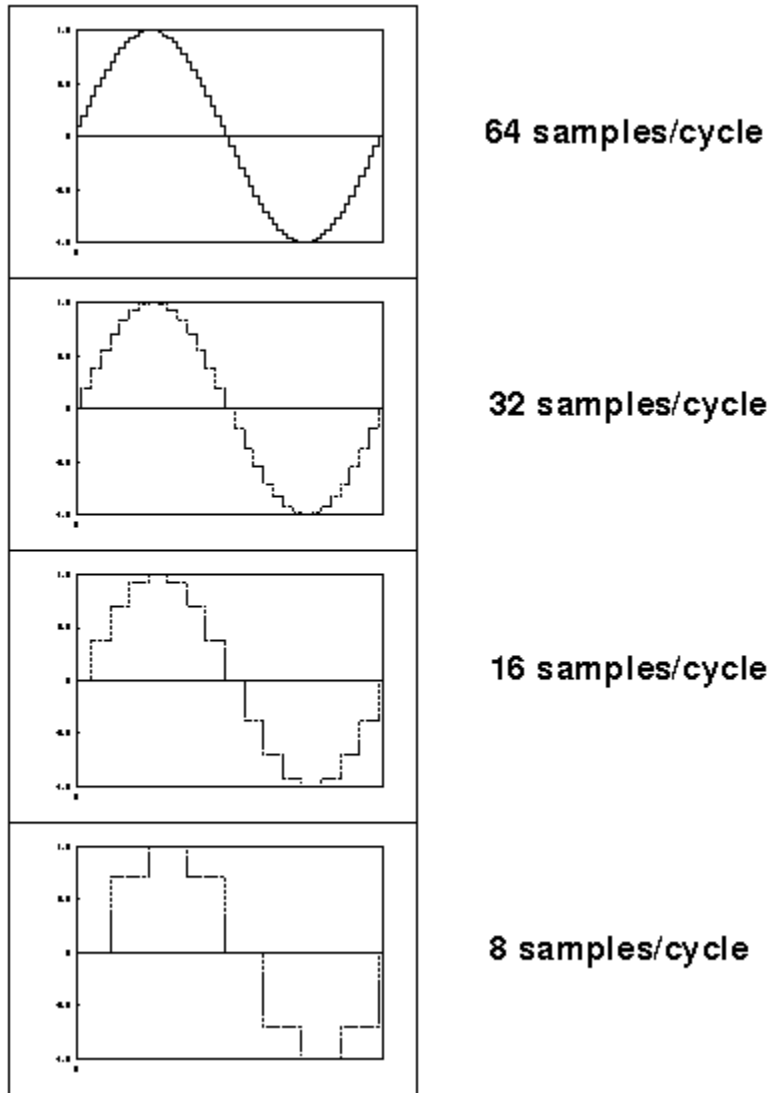


Figure 23. Digitalisation d'une courbe expérimentale avec différentes vitesses d'échantillonnage (nombre des points).

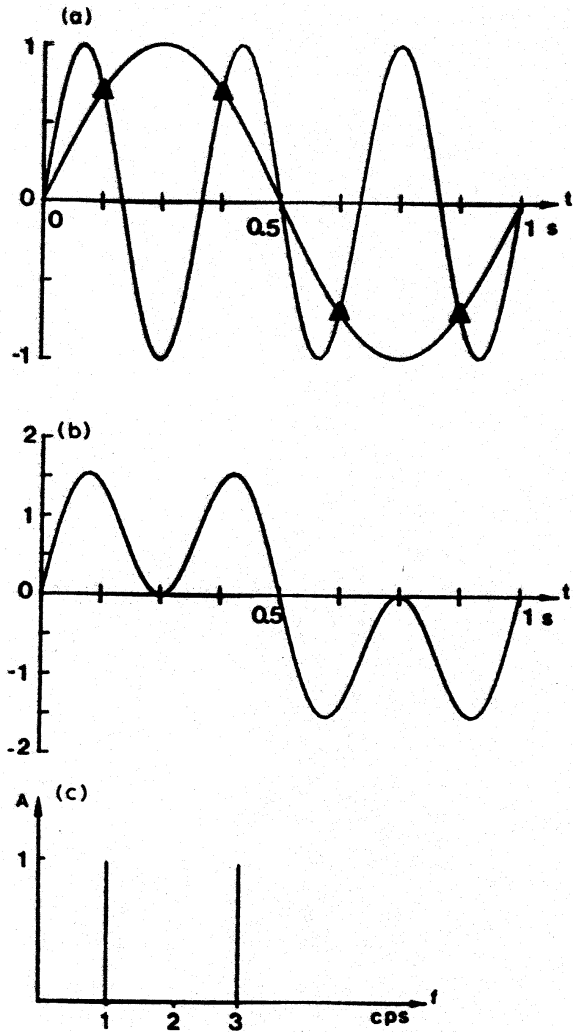


Fig. 1. A signal consisting of a sum of two periodic functions with $T = 1$ s and $T = 1/3$ s (the time axis is considered to be infinite). (a) The two sine functions in the time domain. (b) The sum of the two sine functions. (c) Representation of (b) in the frequency domain. A is the amplitude of the sine functions.

Figure 24. FFT d'une somme de deux fonctions sin.

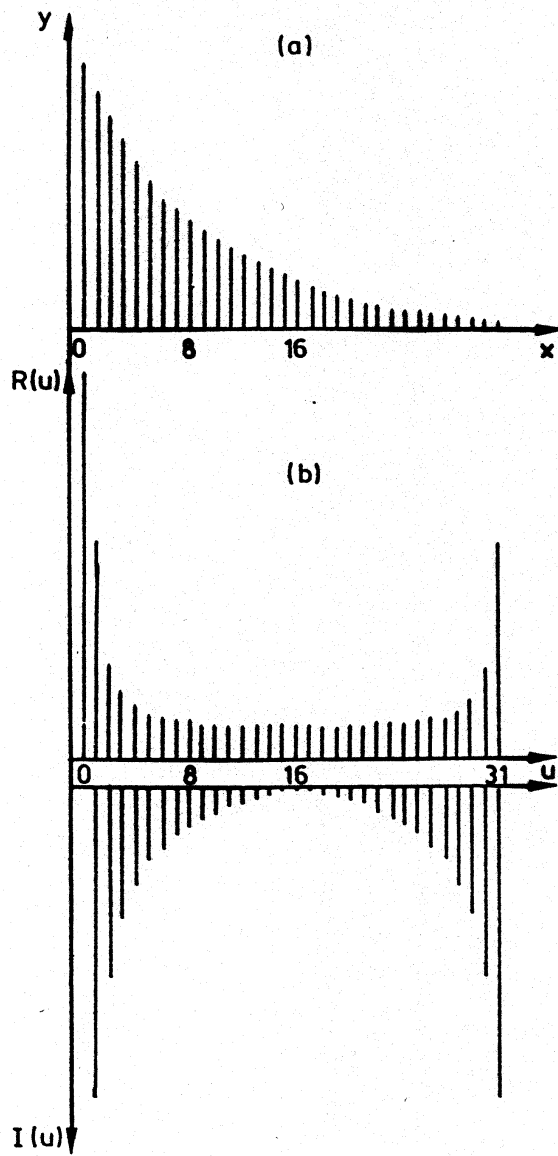


Fig. 4. (a) Exponential signal with $n = 32$ and (b) its FT. $R(u)$ and $I(u)$ are the real and imaginary parts, respectively.

Figure 25. Fonction (x) (a) et sa transformée de Fourier (b).

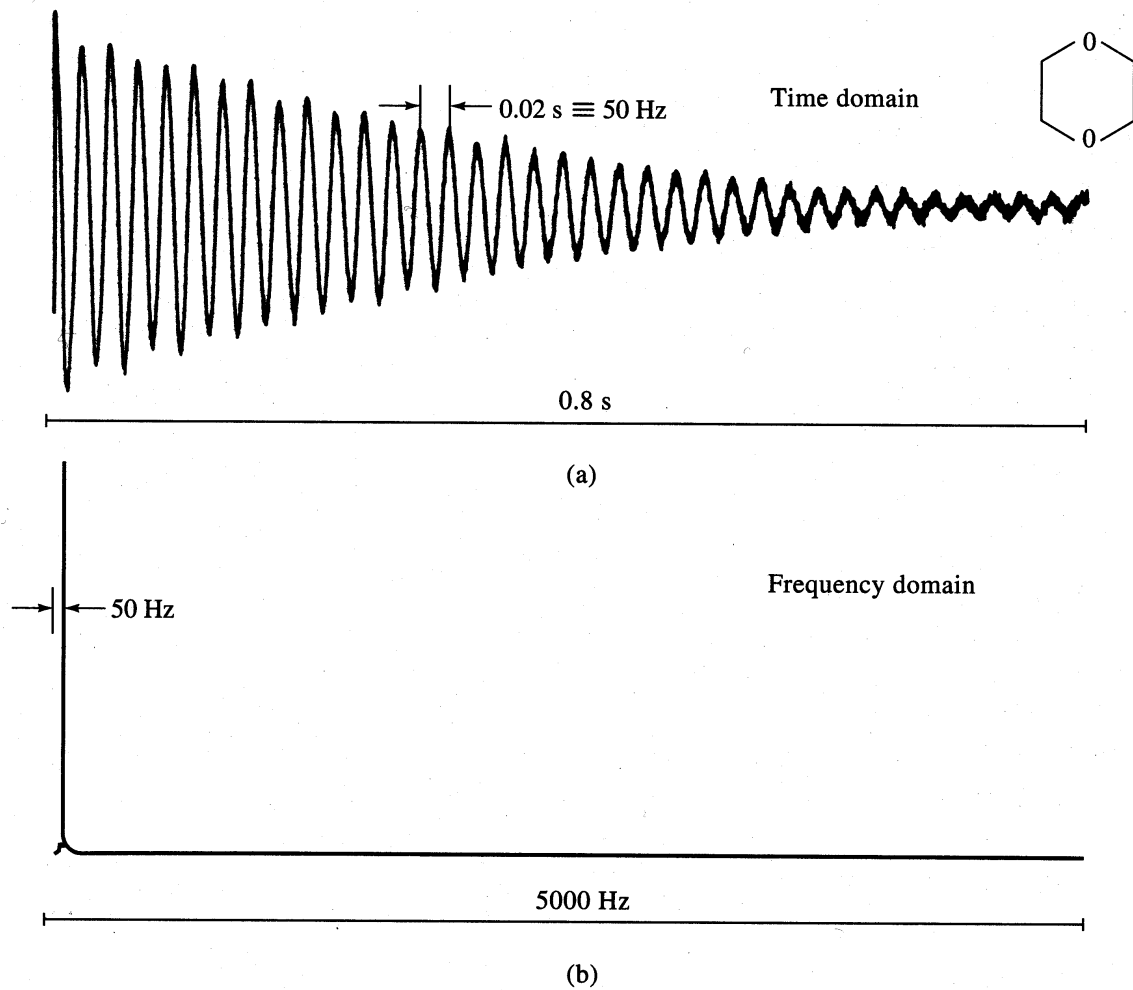


Figure 19-9 (a) FID signal for ^{13}C in dioxane when pulse frequency differs from Larmor frequency by 50 Hz; (b) Fourier transform of (a). (From R. J. Abraham, J. Fisher, and P. Loftus, *Introduction to NMR Spectroscopy*, p. 90. New York: Wiley, 1988. Reprinted by permission of John Wiley & Sons, Inc.)

Figure 26. L'application de la transformée de Fourier à la RMN : Une fréquence dans le système (un type d'atomes de ^{13}C).

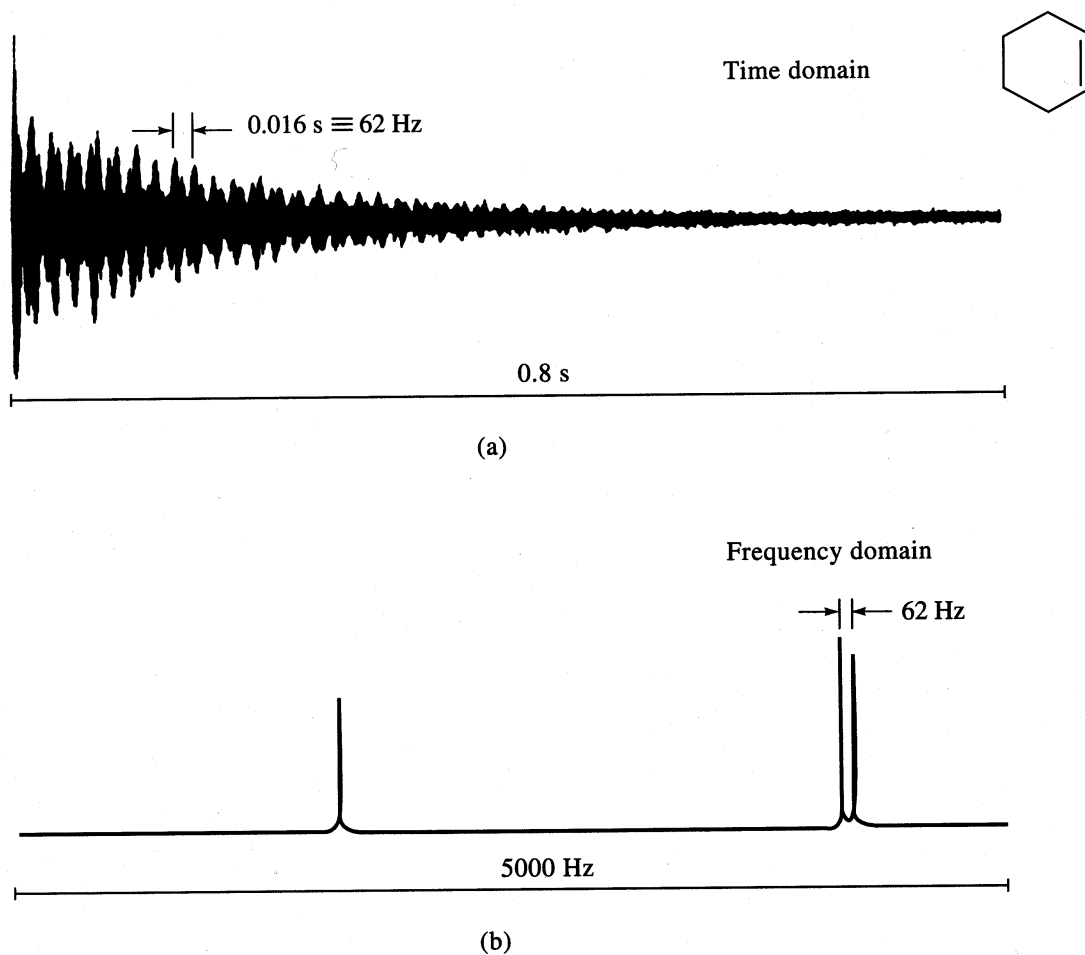


Figure 19-10 (a) FID signal for ^{13}C in cyclohexane; (b) Fourier transform of (a). (From R. J. Abraham, J. Fisher and P. Loftus, *Introduction to NMR Spectroscopy*, p. 91. New York: Wiley, 1988. Reprinted by permission of John Wiley & Sons, Inc.)

Figure 27. L'application de la transformée de Fourier à la RMN : Trois fréquences dans le système (trois types d'atomes de ^{13}C).

Exemple

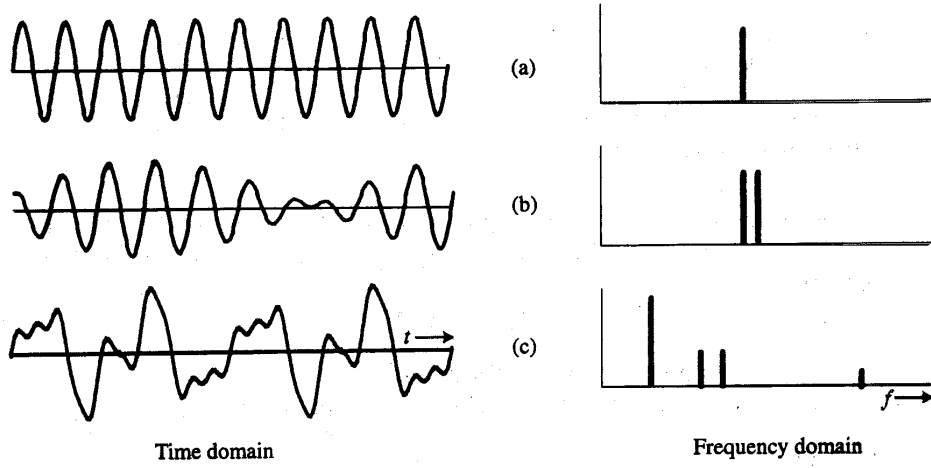
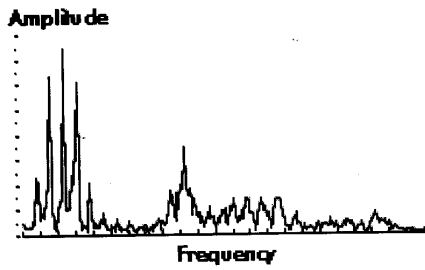


Figure 13.4 Fourier transformations of: (a) a simple sine wave; (b) the product of two sine waves; (c) a complex signal.



• a recording of speech...



• can be analysed to show the spectrum

10.3 « Aliasing »

Un de problèmes rencontrés dans la FFT et « aliasing ». Comme on a vu, la transformée de Fourier contient l'information sur des fréquences jusqu'à :

$$f_{\max} = f_{N/2} = N/(2N\Delta t) = 1/(2\Delta t)$$

Dans notre exemple ci-dessus, $f_{\max} = 1/(2 \cdot 1 \text{ s}) = 0.5 \text{ s}^{-1}$. On n'a pas d'information sur les fréquences plus grandes. Cette fréquence maximale s'appelle la fréquence de Nyquist. Cela veut dire, que ce théorème exige au moins 2 points par période pour déterminer une fréquence.

Quand, dans un spectre existe une ou des fréquences plus grandes que la fréquence de Nyquist, on peut trouver des fréquences fantômes. Regardons le graphique suivant :

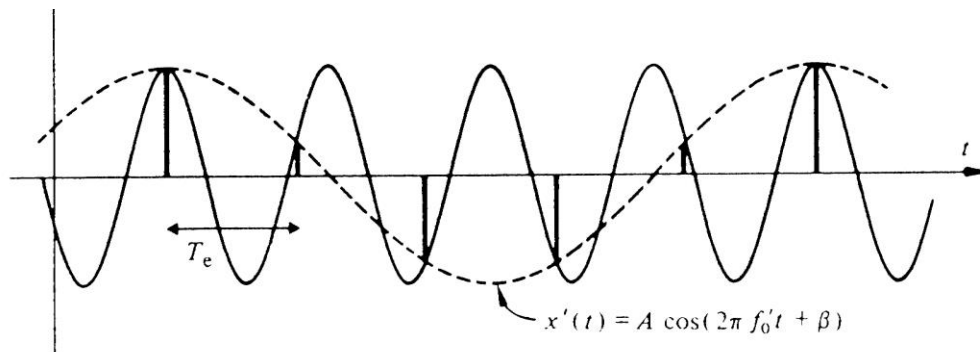
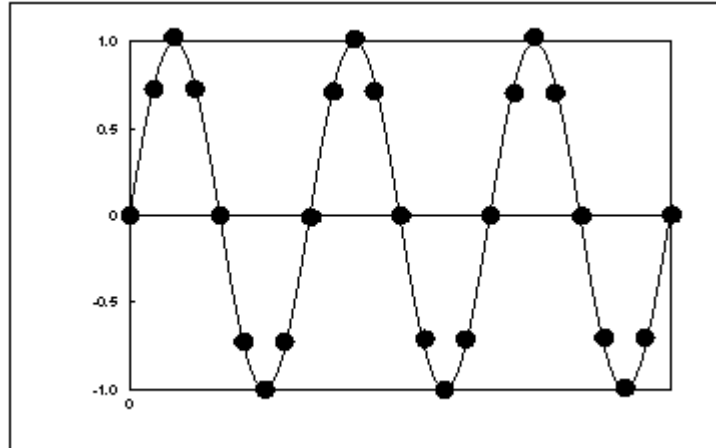


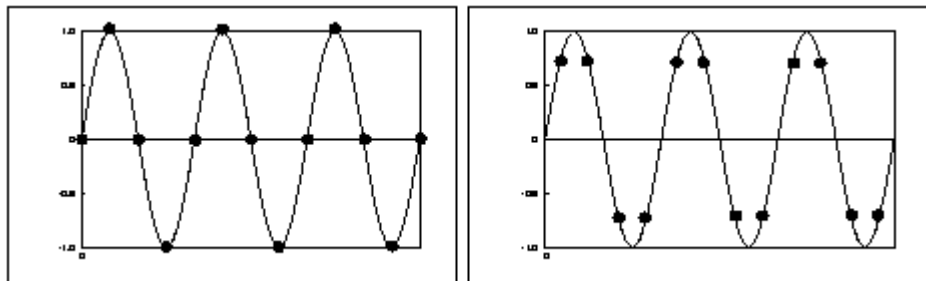
Figure 28. « Aliasing » avec une période d'échantillonnage T_e plus faible que $2/f_0$ on n'est pas capable de reproduire de la courbe originale.

Dans ce cas on a une fréquence élevée, mais on l'échantillonne avec la période T_e . La ligne pointillée, qui rejoint nos points forme une sinusoïde de fréquence inférieure, qui n'existe pas en réalité dans notre spectre. C'est une fréquence fantôme et fausse. Pour éviter ce problème il faut augmenter la vitesse d'échantillonnage ou utiliser un filtre électronique, qui coupera toutes les fréquences plus grandes que la fréquence de Nyquist.

- Graphical Example:
 - SR = 20,000 Hz, Nyquist Frequency = 10,000 Hz
 - $f = 2,500$ Hz (no aliasing)

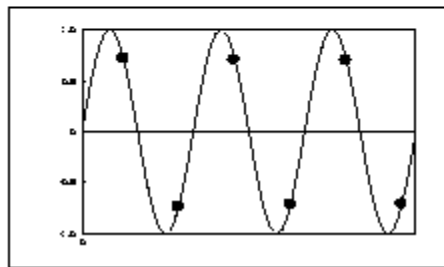
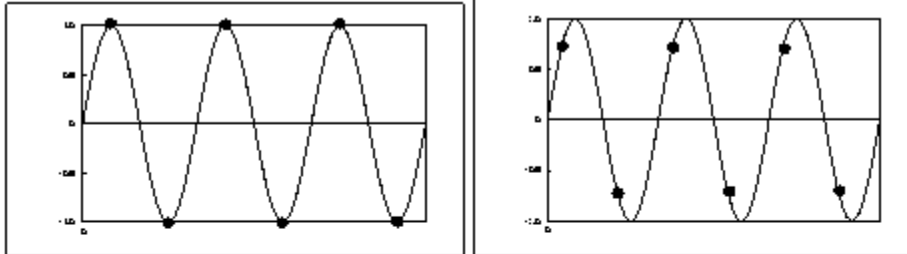


- $f = 5,000$ Hz (no aliasing)

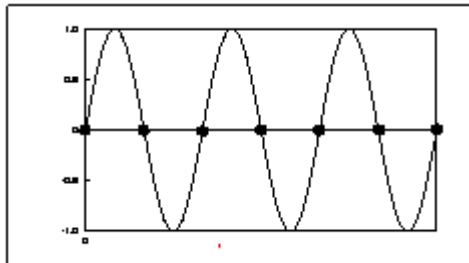


(left and right figures above have same frequency, but have different sampling points)

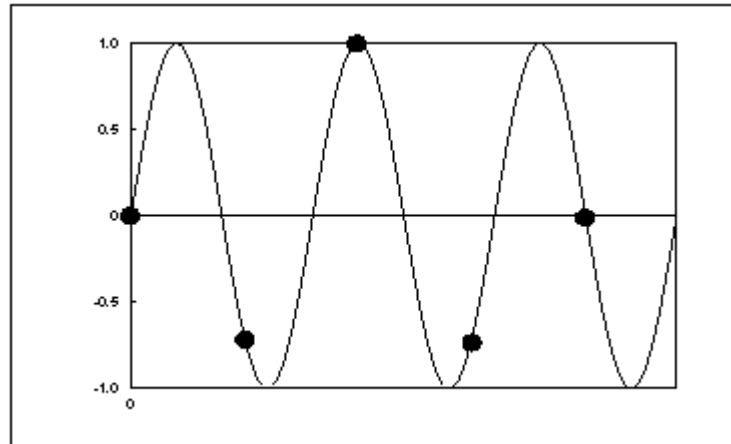
- Graphical Example 2:
- SR = 20,000 Hz, Nyquist Frequency = 10,000 Hz
- $f = 10,000$ Hz (no aliasing)



BUT, if sample points fall on zero-crossings the sound is completely cancelled out.



- **Graphical Example 3:**
 - **SR = 20,000 Hz, Nyquist Frequency = 10,000 Hz**
 - **$f = 12,500$ Hz, $f' = 7,500$**



Fitting the simplest sine wave to the sampled points gives an aliased waveform (dotted line below):

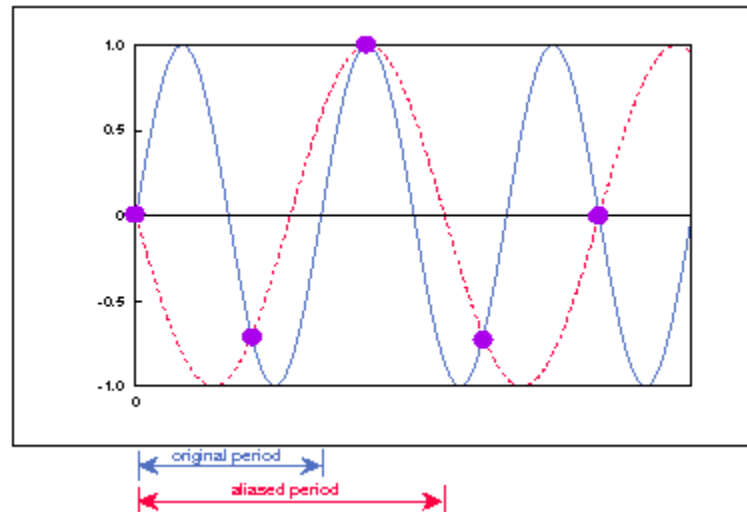


Figure 29. Illustration de l'influence de la vitesse d'échantillonnage sur les courbes obtenus.

10.4 « Leakage »

Un autre problème est rencontré quand le temps total d'acquisition des données n'est pas une multiplication entière de la période de chaque fonction. La transformée de Fourier est définie comme intégrale de $-\infty$ à ∞ , Eq. (82), mais en pratique on intègre de 0 à T . Mais on toujours suppose que la fonction est périodique.

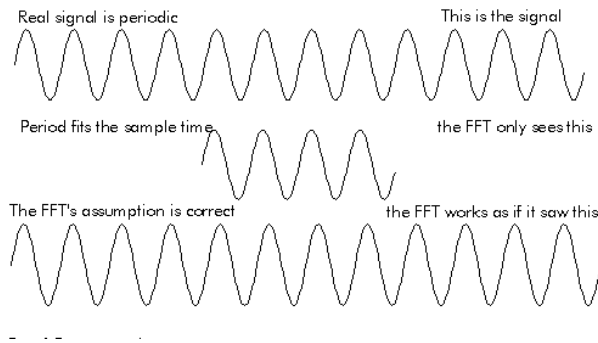


Figure 30. Acquisition des données finies et reconstruction de la fonction périodique : le nombre entier des cycles pendant une acquisition.

Dans le cas quand la portion mesurée ne contient pas de nombre entier de périodes on obtient des discontinuités.

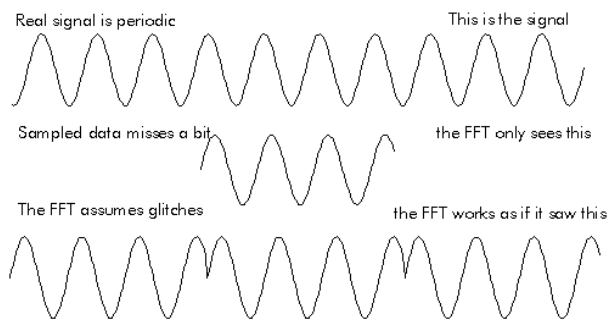


Figure 31. Acquisition des données finies et reconstruction de la fonction périodique : il n'y a pas de nombre entier des cycles pendant une acquisition.

Dans ce cas on obtient une distribution des fréquences, au lieu d'une fréquence. Par exemple la somme de 3 fonction cosinus :

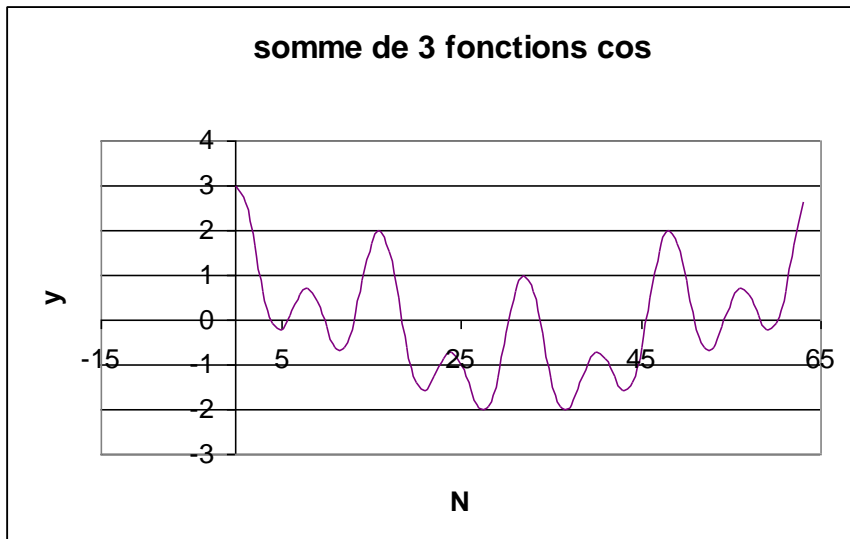
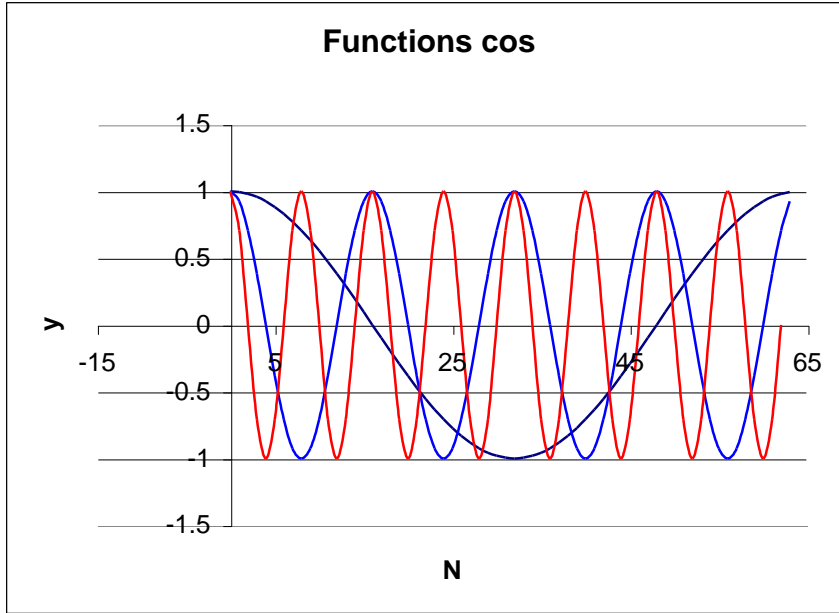


Figure 32. La somme de trois fonctions cosinus.

La transformée FFT est :

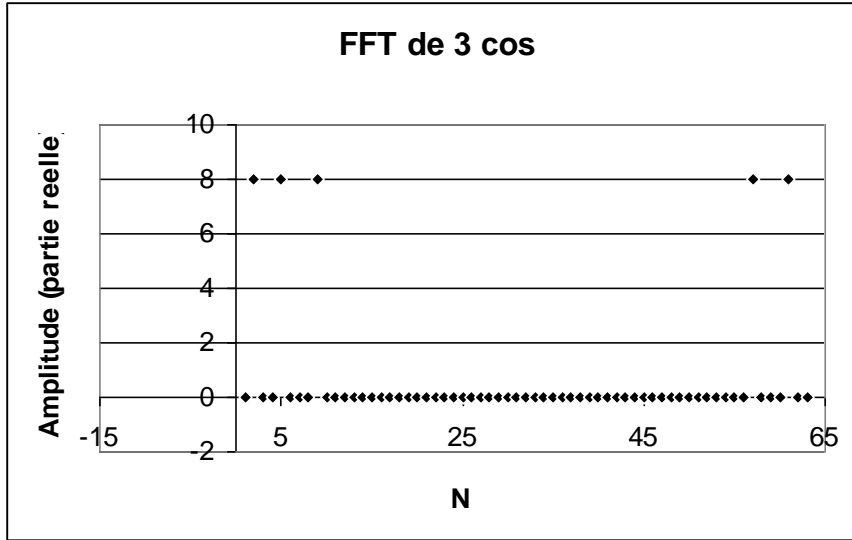


Figure 33. La transformée de Fourier de la courbe dans la Figure 32.

Cependant, pour une fonction dans la Figure 34 la transformée FFT est présentée dans la Figure 35.

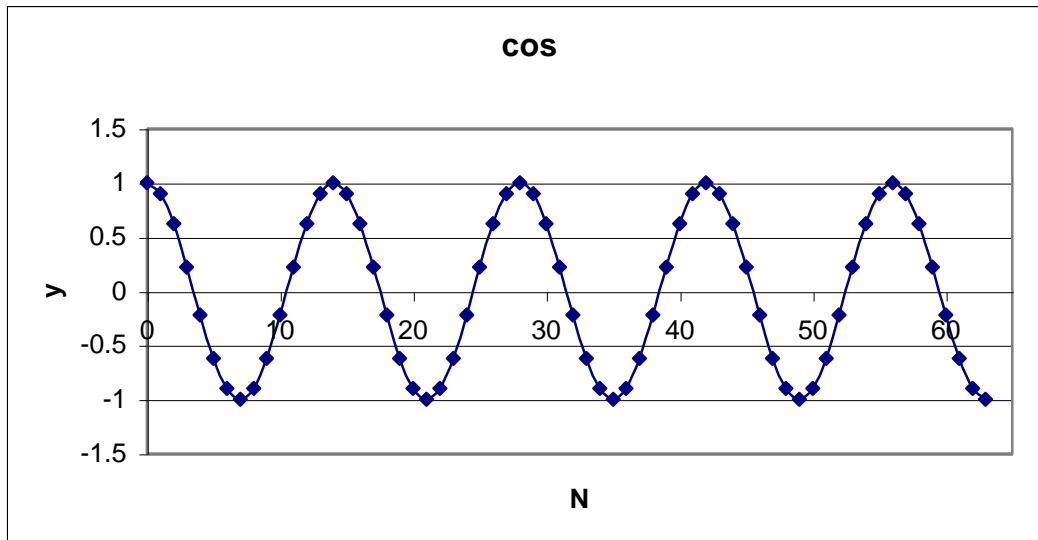


Figure 34. Fonction périodique cosinus qui n'a pas un nombre entier de périodes durant le temps total mesuré.

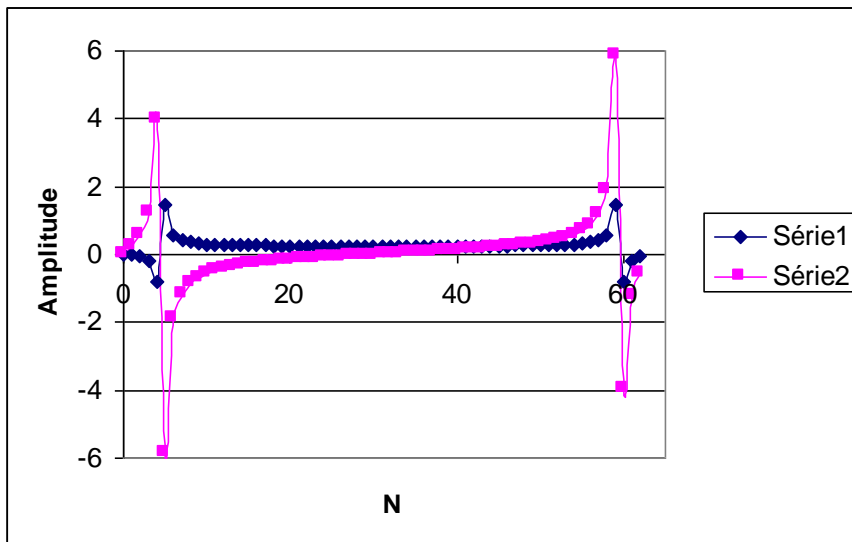
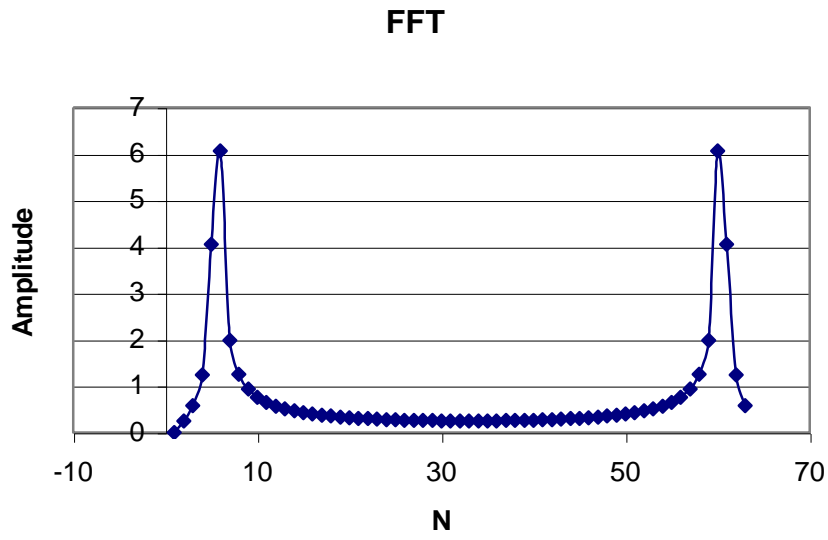
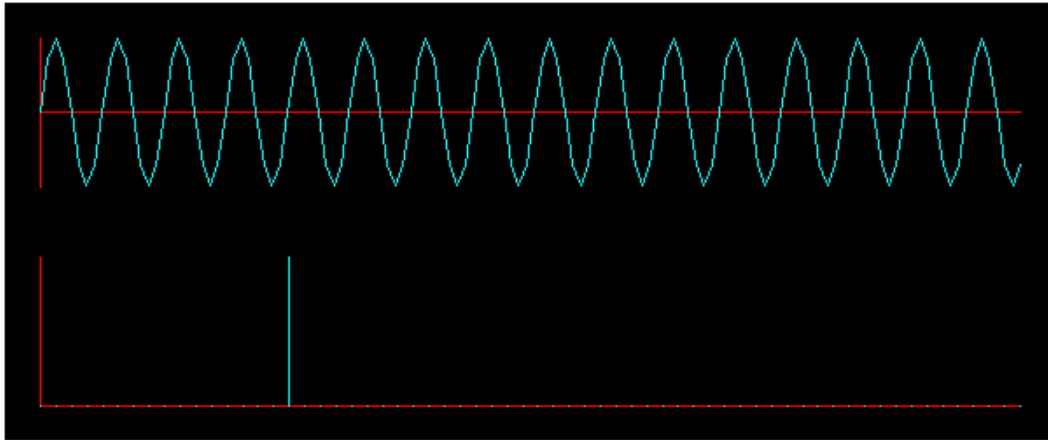


Figure 35. Transformée de Fourier de la fonction dans la Figure 34.

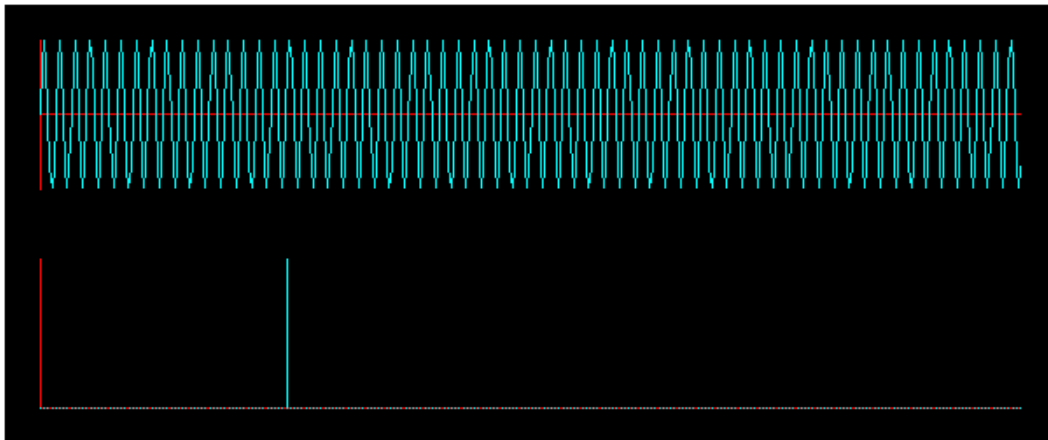
Ce problème est aussi illustré dans la série des exemples ci dessous.



Number of samples:

Sampling rate: samples / s

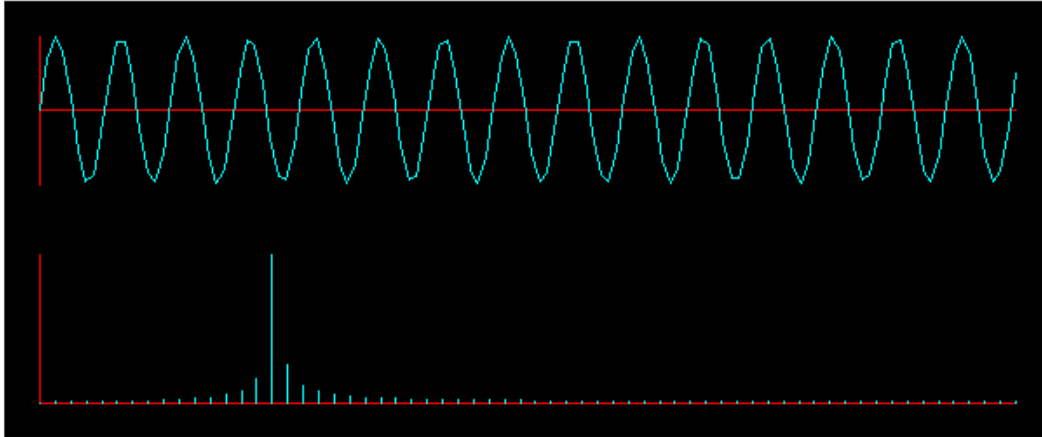
Signal waveform expression:



Number of samples:

Sampling rate: samples / s

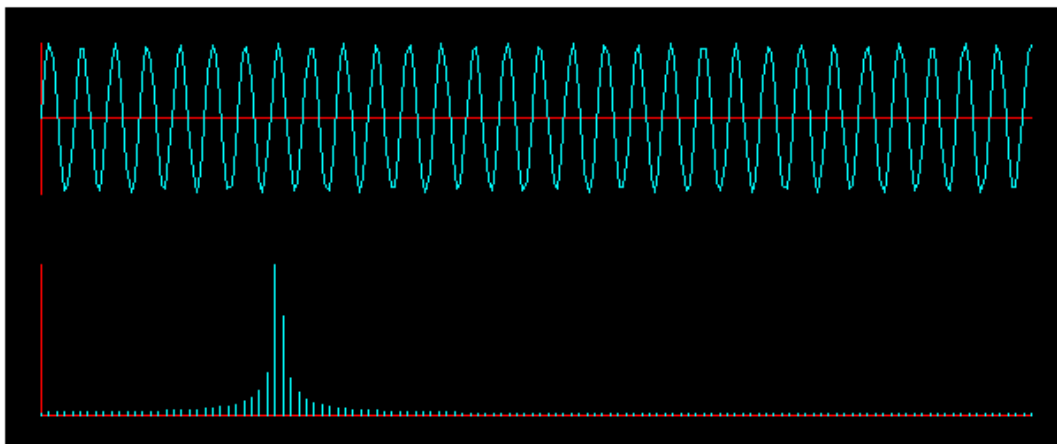
Signal waveform expression:



Number of samples:

Sampling rate: samples / s

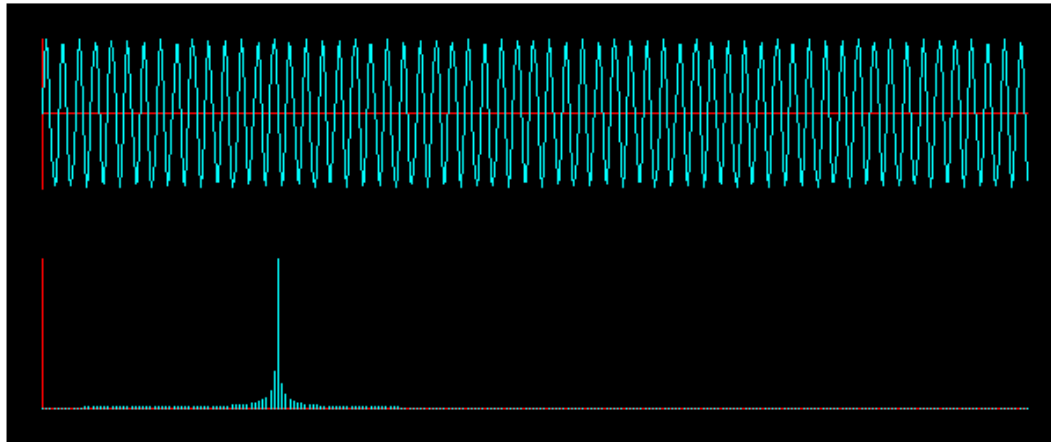
Signal waveform expression:



Number of samples:

Sampling rate: samples / s

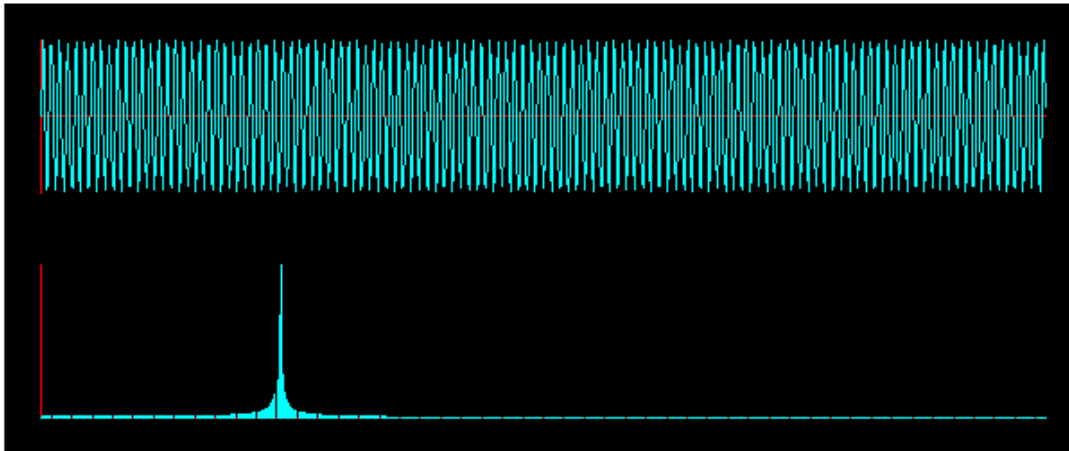
Signal waveform expression:



Number of samples:

Sampling rate: samples / s

Signal waveform expression:



Number of samples:

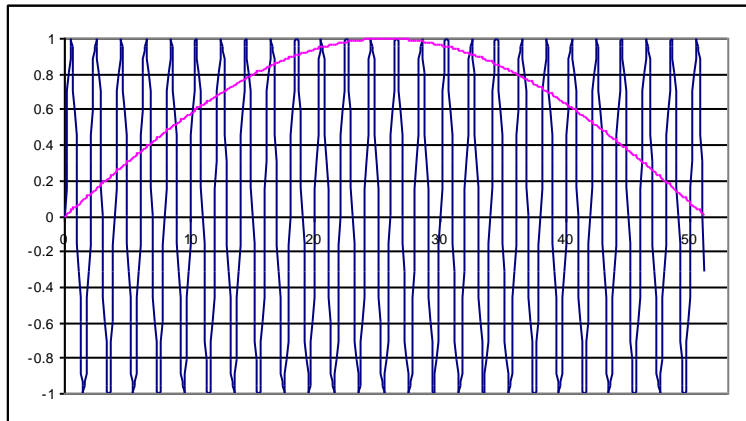
Sampling rate: samples / s

Signal waveform expression:

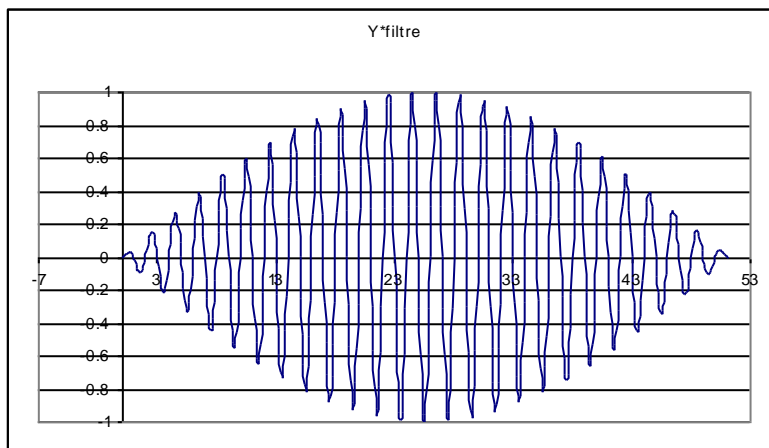
Figure 36. Illustration de problème de leakage ». En augmentant le nombre des échantillons on est capable de mieux déterminer la fréquence originale du signal. ¹

Pour éviter le problème de « leakage » il faut synchroniser la vitesse d'échantillonnage avec la période de la fonction étudiée, augmenter le nombre de points expérimentaux ou utiliser un filtre, qui diminue l'importance des valeurs au début et à la fin de la période d'échantillonnage.

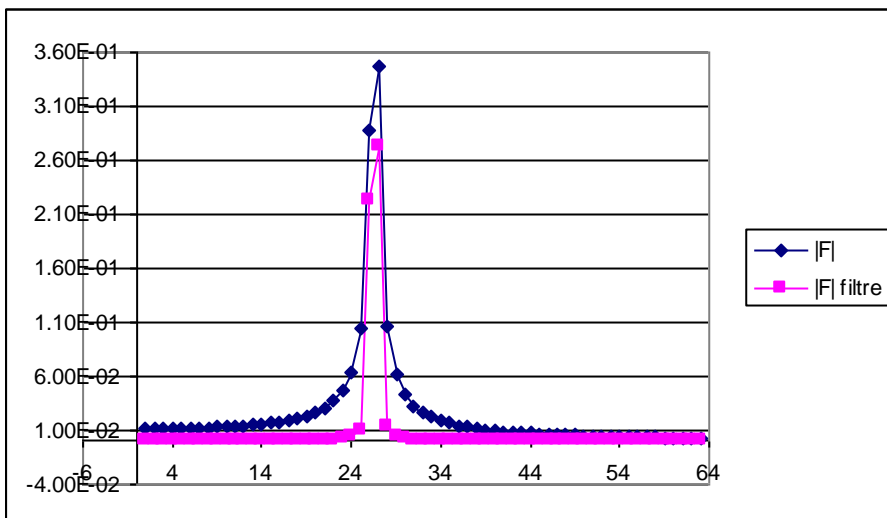
Utilisation d'un filtre pour éviter les problèmes de « leakage » :
Le filtre débute et finie à zéro, on multiplie le signal par le filtre



Le signal obtenu :

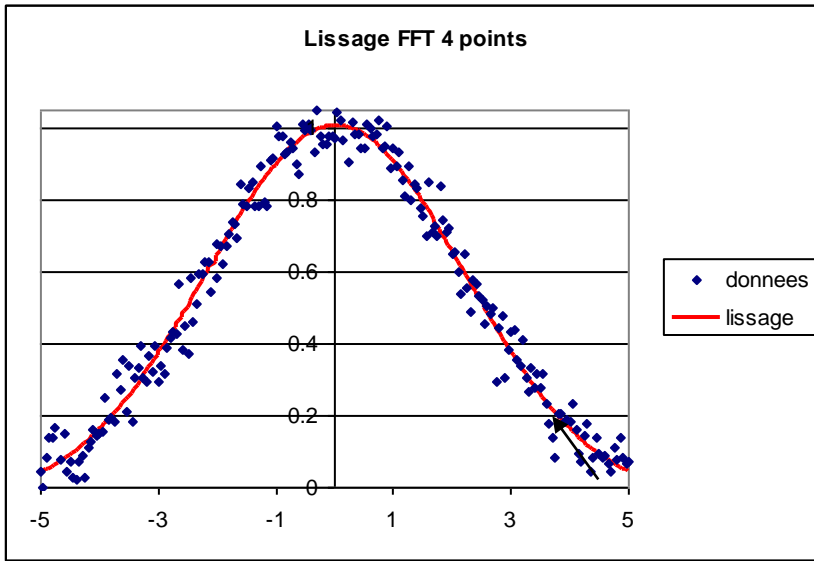


Comparaison de la FFT de signal brut et filtré :

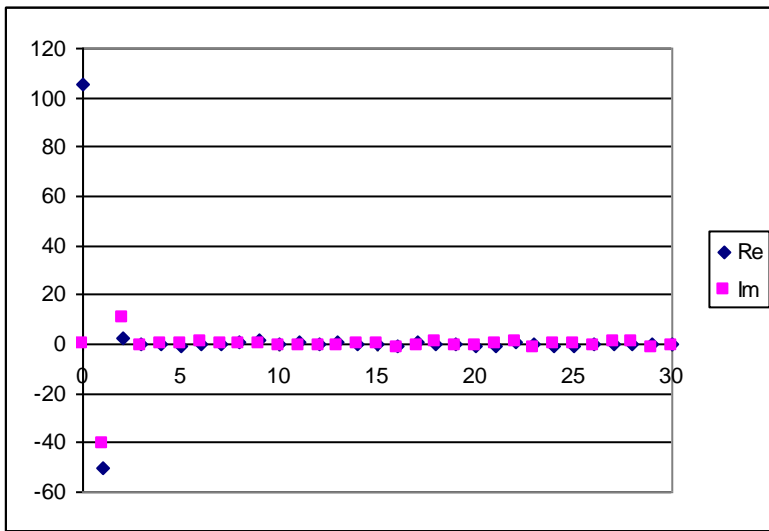


Lissage par FFT

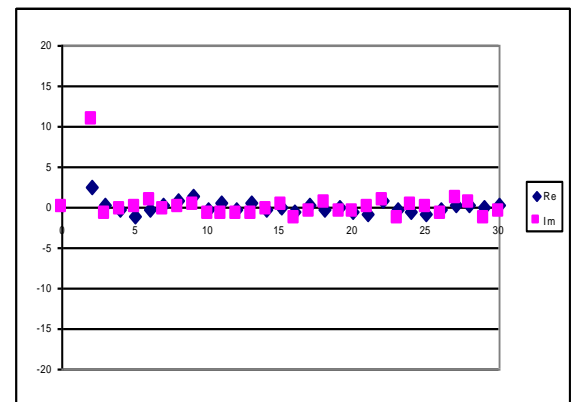
Données brut :



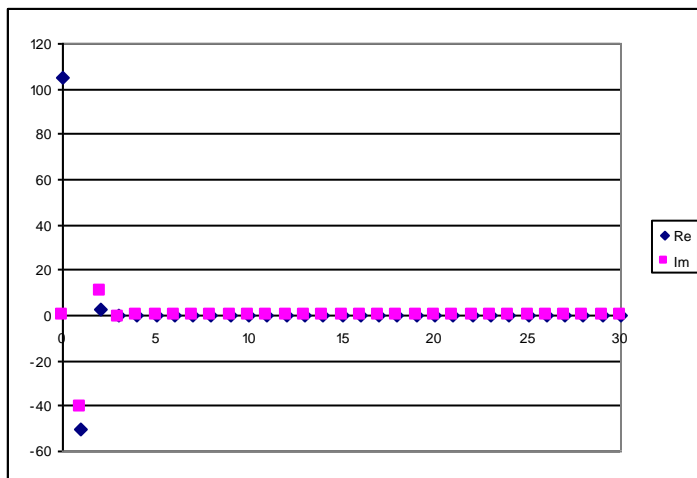
La FFT



détail



Filtrage FFT, la FFT inverse produit une courbe lisse.



11. Exercices avec Maple

Determination de l'écart type.

> **restart;**

> **y := (a-b) / (a+b);**

$$y := \frac{a-b}{a+b}$$

derivé partielle dy/da

> **diff(y, a);**

$$\frac{1}{a+b} - \frac{a-b}{(a+b)^2}$$

> **dyda := simplify(%);**

$$dyda := 2 \frac{b}{(a+b)^2}$$

> **dydb := simplify(diff(y, b));**

$$dydb := -2 \frac{a}{(a+b)^2}$$

> **err := sqrt((sa*dyda)^2 + (sb*dydb)^2);**

$$err := 2 \sqrt{\frac{sa^2 b^2}{(a+b)^4} + \frac{sb^2 a^2}{(a+b)^4}}$$

> **simplify(%);**

$$2 \sqrt{\frac{sa^2 b^2 + sb^2 a^2}{(a+b)^4}}$$

> **sa := 0.1;**

$$sa := .1$$

> **sb := 0.2;**

$$sb := .2$$

> **a := 4;**

$$a := 4$$

> **b := 2;**

$$b := 2$$

> **eval(err);**

$$.04581228472$$

> **eval(y);**

$$\frac{1}{3}$$

> **evalf(y);**

.3333333333

Une autre fonction.> **restart;**> **y:=(a+b)/(c-d);**

$$y := \frac{a+b}{c-d}$$

> **dyda:=diff(y,a);**

$$dyda := \frac{1}{c-d}$$

> **dydb:=diff(y,b);**

$$dydb := \frac{1}{c-d}$$

> **dydc:=diff(y,c);**

$$dydc := -\frac{a+b}{(c-d)^2}$$

> **dydd:=diff(y,d);**

$$dydd := \frac{a+b}{(c-d)^2}$$

> **std:=sqrt((sa*dyda)^2+(sb*dydb)^2+(sc*dydc)^2+(sd*dydd)^2);**

$$std := \sqrt{\frac{sa^2}{(c-d)^2} + \frac{sb^2}{(c-d)^2} + \frac{sc^2(a+b)^2}{(c-d)^4} + \frac{sd^2(a+b)^2}{(c-d)^4}}$$

> **sa:=0.1;**

$$sa := .1$$

> **sb:=0.2;**

$$sb := .2$$

> **sc:=0.05;**

$$sc := .05$$

> **sd:=0.15;**

$$sd := .15$$

> **a:=3;**

$$a := 3$$

> **b:=2;**

$$b := 2$$

> **c:=4;**

$$c := 4$$

> **d:=1;**

```

d := 1
> eval(std);
.1152024520
> evalf(y);
1.666666667

```

Une autre fonction:

```

> restart;
> y:=a*b^c;
y := a b^c
> dyda:=diff(y,a);
dyda := b^c
> dydb:=diff(y,b);
dydb :=  $\frac{a b^c c}{b}$ 
> dydb:=simplify(%);
dydb := a b^(c-1) c
> dydc:=diff(y,c);
dydc := a b^c ln(b)
> std:=sqrt((sa*dyda)^2+(sb*dydb)^2+(sc*dydc)^2);
std :=  $\sqrt{sa^2 (b^c)^2 + sb^2 a^2 (b^{(c-1)})^2 c^2 + sc^2 dydc^2}$ 
> a:=2.1;
a := 2.1
> b:=3.1;
b := 3.1
> c:=1.5;
c := 1.5
> sa:=0.1;
sa := .1
> sb:=0.15;
sb := .15
> sc:=0.2;
sc := .2
> std;
>

```


2.777939088

> **evalf(y)** ;

11.46203778

> *Exemple*> **restart** ;> **y := (a+b^2) * log[10] (a+b^2) ;**

$$y := \frac{(a + b^2) \ln(a + b^2)}{\ln(10)}$$

> **dyda := diff(y, a) ;**

$$dyda := \frac{\ln(a + b^2)}{\ln(10)} + \frac{1}{\ln(10)}$$

> **dydb := diff(y, b) ;**

$$dydb := 2 \frac{b \ln(a + b^2)}{\ln(10)} + \frac{2b}{\ln(10)}$$

> **err := sqrt((dyda*sa)^2 + (dydb*sb)^2) ;**

$$err := \sqrt{\left(\frac{\ln(a + b^2)}{\ln(10)} + \frac{1}{\ln(10)}\right)^2 sa^2 + \left(2 \frac{b \ln(a + b^2)}{\ln(10)} + \frac{2b}{\ln(10)}\right)^2 sb^2}$$

> **simplify(%) ;**

1.900000000

> **a := 2.7 ;***a := 2.7*> **b := 1.0 ;***b := 1.0*> **sa := 0.2 ;***sa := .2*> **sb := 0.25 ;***sb := .25*

>

>

> **evalf(err) ;**

.5398607286

> **evalf(y) ;**

2.102346380

Intégration

```
> restart;
```

```
> y:=exp(-x^2);
```

$$y := e^{(-x^2)}$$

Integration de y de 0 à 1

```
> iy:=int(y,x=0..1);
```

$$iy := \frac{1}{2} \operatorname{erf}(1) \sqrt{\pi}$$

```
> evalf(iy);
```

.7468241330

```
> iy:=int(y,x=0..infinity);
```

$$iy := \frac{1}{2} \sqrt{\pi}$$

```
> evalf(iy);
```

.8862269255

```
> Pg:=1/sqrt(2*Pi)*exp(-z^2/2);
```

$$Pg := \frac{1}{2} \frac{\sqrt{2} e^{(-1/2 z^2)}}{\sqrt{\pi}}$$

```
> Ag:=int(Pg,z=-x..x);
```

$$Ag := \operatorname{erf}\left(\frac{1}{2} \sqrt{2} x\right)$$

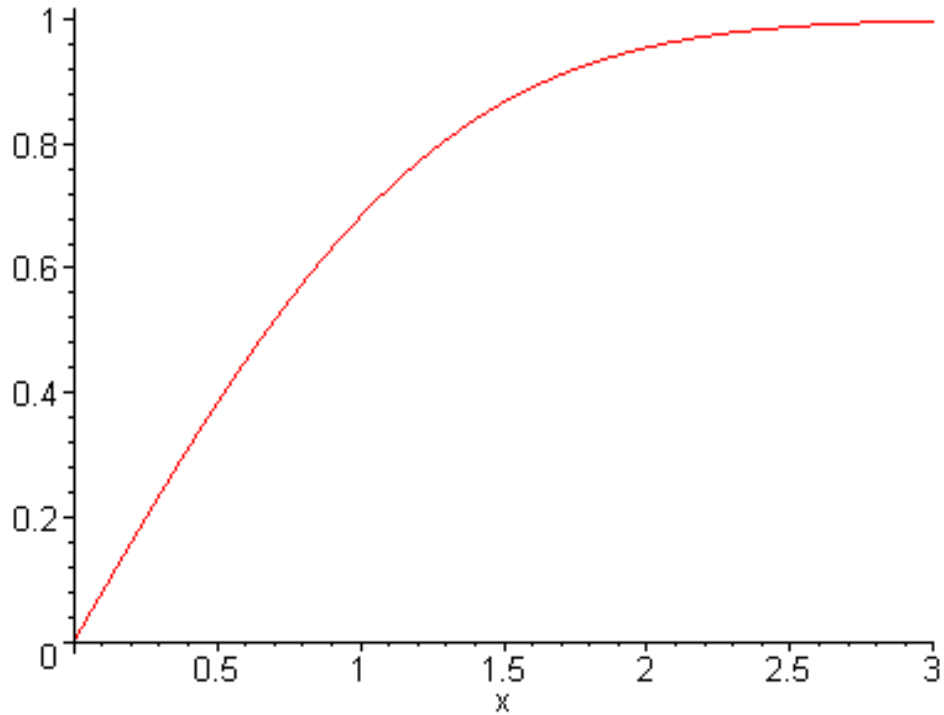
```
> Ag1:=int(Pg,z=-1..1);
```

$$Ag1 := \operatorname{erf}\left(\frac{1}{2} \sqrt{2}\right)$$

```
> evalf(Ag1);
```

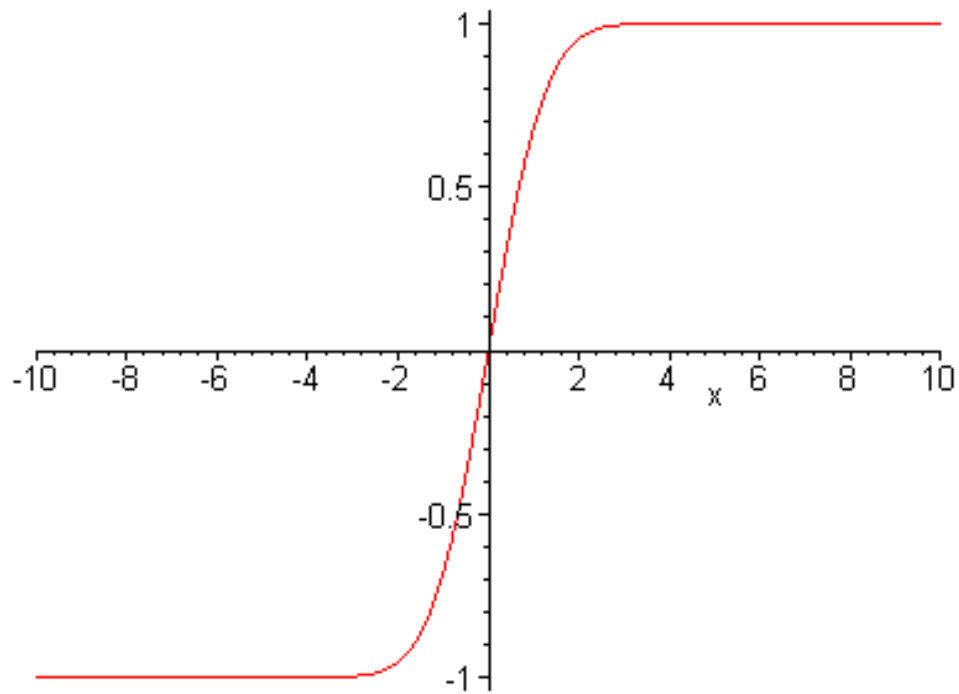
.6826894920

```
> plot(Ag,x=0..3);
```



Cliquez l'équation Ag de bouton droite et exécutez le graphique automatique.

> **smartplot(Ag) ;**



> **evalf(erf(0.5)) ;**

.5204998778

> **evalf(erf(1.)) ;**

.8427007929

```

> evalf(erf(2.)) ;
.9953222650

> f:=x->Ag;
f:=x → Ag

> Ag;
 $\operatorname{erf}\left(\frac{1}{2}\sqrt{2} x\right)$ 

> f(x) ;
 $\operatorname{erf}\left(\frac{1}{2}\sqrt{2} x\right)$ 

> eval(f(x), x=2) ;
 $\operatorname{erf}(\sqrt{2})$ 

> evalf(%);
.9544997360

> restart;
> y:=x/(1+x)^2;
 $y := \frac{x}{(1+x)^2}$ 

> iy:=int(y, x) ;
 $iy := \frac{1}{1+x} + \ln(1+x)$ 

> x:=1;
x:=1

> eval(iy) ;
 $\frac{1}{2} + \ln(2)$ 

> iy;
 $\frac{1}{2} + \ln(2)$ 

> evalf(%);
1.193147181

>

```

Équations différentielles

```

> restart;
> y:=diff(C(t), t)+k*C(t)=0;

```

$$y := \left(\frac{\partial}{\partial t} C(t) \right) + k C(t) = 0$$

> **dsolve** ({y, C(0)=Co}, C(t));

$$C(t) = Co e^{(-kt)}$$

> **restart**;

> **y:=diff**(C(t), t)+k*C(t)^2=0;

$$y := \left(\frac{\partial}{\partial t} C(t) \right) + k C(t)^2 = 0$$

> **dsolve** ({y, C(0)=Co}, C(t));

$$C(t) = \frac{1}{kt + \frac{1}{Co}}$$

Solution des équations non-linéaires, pH

Calculez le pH du phénol $5 \cdot 10^{-5}$ M, $K_a = 10^{-10}$ (acide très faible). Comparez avec une solution approximative (en négligeant la dissociation de l'eau). Quelle est le taux de dissociation du phénol?

Pour une réaction de dissociation dans le milieu aqueux :



on peut écrire les équilibres suivants :

$$K_a = \frac{[H_3O^+][A^-]}{[HA]} \quad (89)$$

$$[H^+][OH^-] = k_w \quad (90)$$

et les balances : de charges

$$[H^+] = [OH^-] + [A^-] \quad (91)$$

et de masses

$$[HA] + [A^-] = C_a \quad (92)$$

> **restart**;

> **Ka:=1e-10**;

$$Ka := .1 \cdot 10^{-9}$$

> **kw:=1e-14**;

$$kw := .1 \cdot 10^{-13}$$

> **C:=5e-5**;

$$C := .00005$$

> **eq1:=Ka=CH*CA/CHA**;

```

eq1 := .1 10-9 =  $\frac{CH \cdot CA}{CHA}$ 
> eq2 := CH*COH=kw;
eq2 := CH COH = .1 10-13
> eq3 := CH=COH+CA;
eq3 := CH = COH + CA
> eq4 := C=CHA+CA;
eq4 := .00005 = CHA + CA
>
sol := solve ({eq1, eq2, eq3, eq4, CH>0, CA>0, COH>0, CHA>0}, {CH, CA, COH, CHA
});
sol := { CA = .407970667610-7, COH = .816607639110-7, CHA = .00004995920293
CH = .122457830710-6 }
> CH := .1224578307e-6;
CH := .122457830710-6
> pH := -log10 (CH) ;
pH := 6.912013438

```

pH = 6.91

Solution alternative sans restrictions donne 3 séries des racines:

```

> sol1 := solve ({eq1, eq2, eq3, eq4}, {CH, CA, COH, CHA}) ;
sol1 := { COH = -.816385416910-7, CHA = .00005004085262 CA = -.408526223310-7,
CH = -.122491164010-6 }, { CHA = .00004995920293 CA = .407970667610-7,
COH = .816607639110-7, CH = .122457830710-6 }, { COH = -.0001500000222
CHA = -.00009999995556 CA = .00014999995556 CH = -.666666567910-10 }
> CH := .1224578307e-6;
CH := .122457830710-6
> pH := -log10 (CH) ;
pH := 6.912013438

```

On peut aussi réarranger les équations; l'éqn (5) nous donne:

$$[HA] = C_a - [A^-] \quad (93)$$

En combinant avec les eqns. (3) et (4) nous obtenons :

$$[H^+] = \frac{k_w}{[H^+]} + [A^-] \quad (94)$$


```
> kw:=1e-14;
                                kw:=.1 10-13
> Ca:=5e-5;
                                Ca :=.00005
> sol:=solve({CH^2+CH*Ka-Ka*Ca=0},{CH});
                                sol := { CH = .706606958010-7 }, { CH = -.707606958010-7 }

sol:=solve({CH^2+CH*Ka-Ka*Ca=0,CH>0},{CH});
                                sol := { CH = .706606958010-7 }
> CH:=.7066069580e-7;
                                CH := .706606958010-7
> pH:=-log10(CH);
                                pH := 7.150822090
```

pH=7.15!

12. Fonctions statistiques d'Excel

Pour installer l'Utilitaire d'analyse

Ouvrir:

Outils

Macro complémentaire

Utilitaire d'analyse

Fonction statistiques d'Excel

$$P_G(x, \mu, \sigma) = \text{LOI.NORMALE}(x, \mu, \sigma, \text{FAUX}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

$$P_G(z) = \text{LOI.NORMALE}(z, 0, 1, \text{FAUX}) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

English : NORMDIST($x, \mu, \sigma, \text{FALSE}$)

$$\text{LOI.NORMALE}(z, 0, 1, \text{VRAI}) = \int_{-\infty}^z P_G(u, 0, 1) du$$

$$\text{LOI.NORMALE}(x, \mu, \sigma, \text{VRAI}) = \int_{-\infty}^x P_G(u, \mu, \sigma) du$$

English : NORMDIST($x, \mu, \sigma, \text{TRUE}$)

$$\begin{aligned} A_G(z) &= \text{LOI.NORMALE}(z, 0, 1, \text{VRAI}) - \text{LOI.NORMALE}(-z, 0, 1, \text{VRAI}) \\ &= \int_{-z}^z P_G(z, 0, 1) dz = \frac{1}{\sqrt{2\pi}} \int_{-z}^z e^{-\frac{z^2}{2}} dz \end{aligned}$$

$$A_G(x) = \text{LOI.NORMALE}(x, \mu, \sigma, \text{VRAI}) - \text{LOI.NORMALE}(-x, \mu, \sigma, \text{VRAI})$$

$$\text{LOI.NORMALE.STANDARD.INVERSE}(\alpha) = z(\alpha) \qquad \int_{-\infty}^z P_G(z, 0, 1) dz = \alpha$$

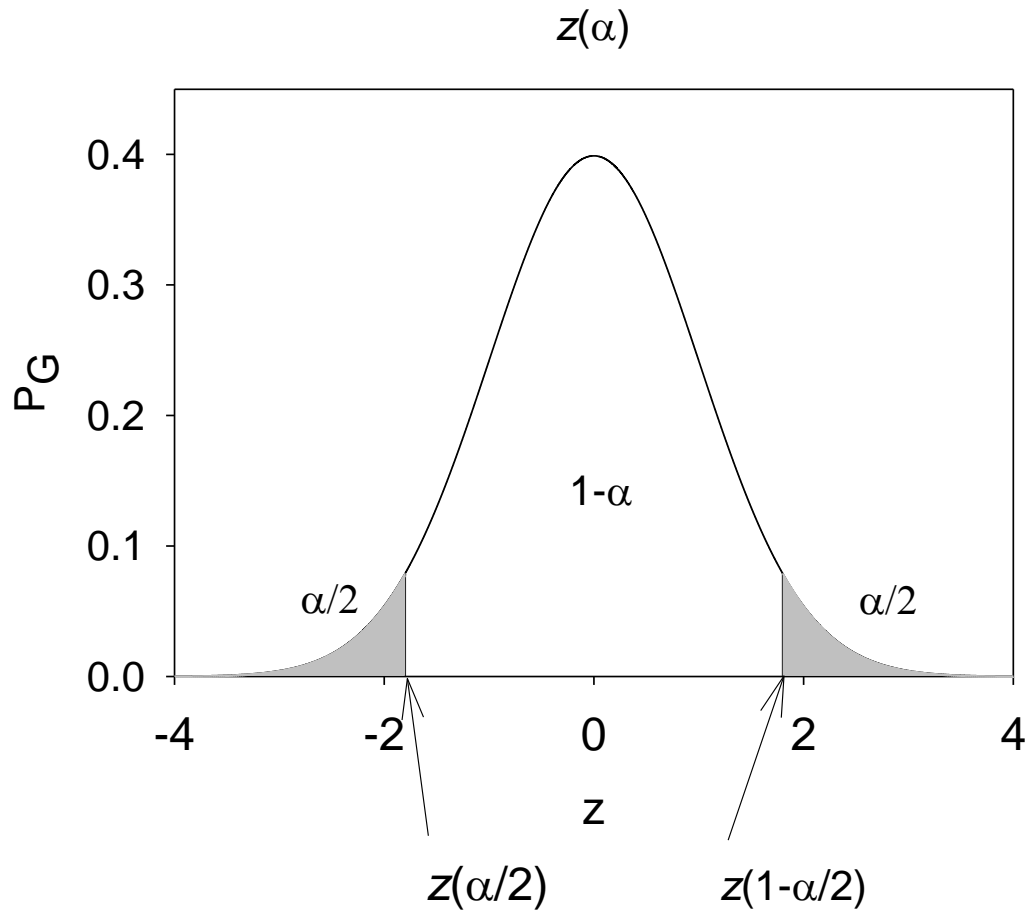
English : NORMSINV(α)

LOI.NORMALE.INVERSE(α, μ, σ) = $x(\alpha)$

$$\int_{-\infty}^x P_G(x, \mu, \sigma) dz = \alpha$$

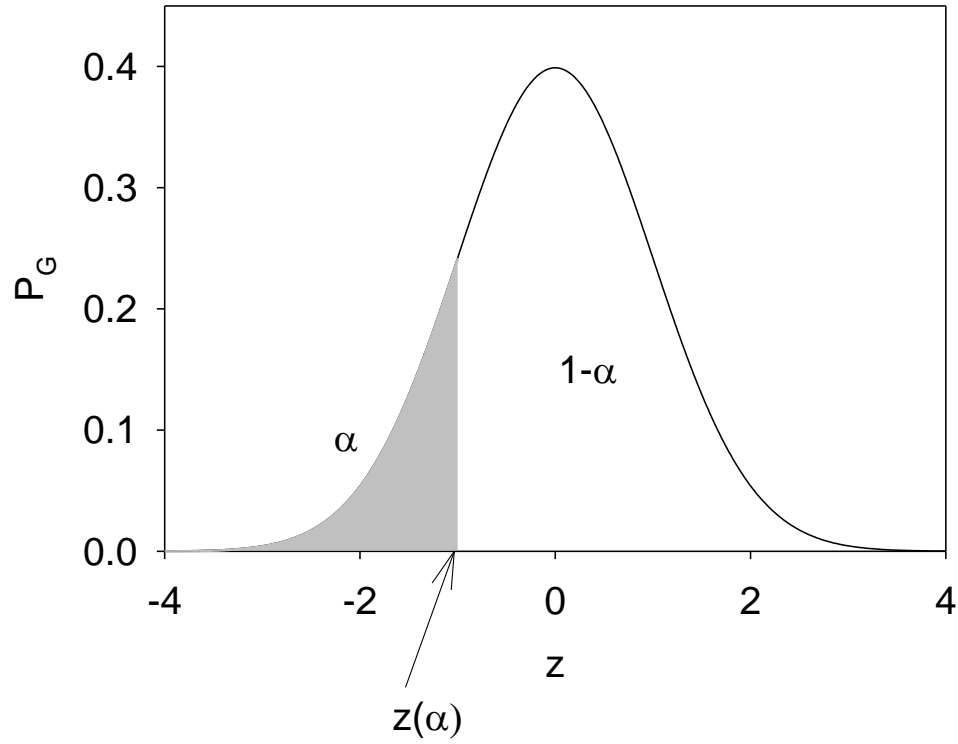
English : NORMINV(α, μ, σ)

$z(\alpha) = \text{LOI.NORMALE.STANDARD.INVERSE}(1-\alpha/2)$ voir la figure



$$\text{LOI.NORMALE.STANDARD.INVERSE}(\alpha) = z(\alpha) \int_{-\infty}^z P_G(z, 0, 1) dz = \alpha$$

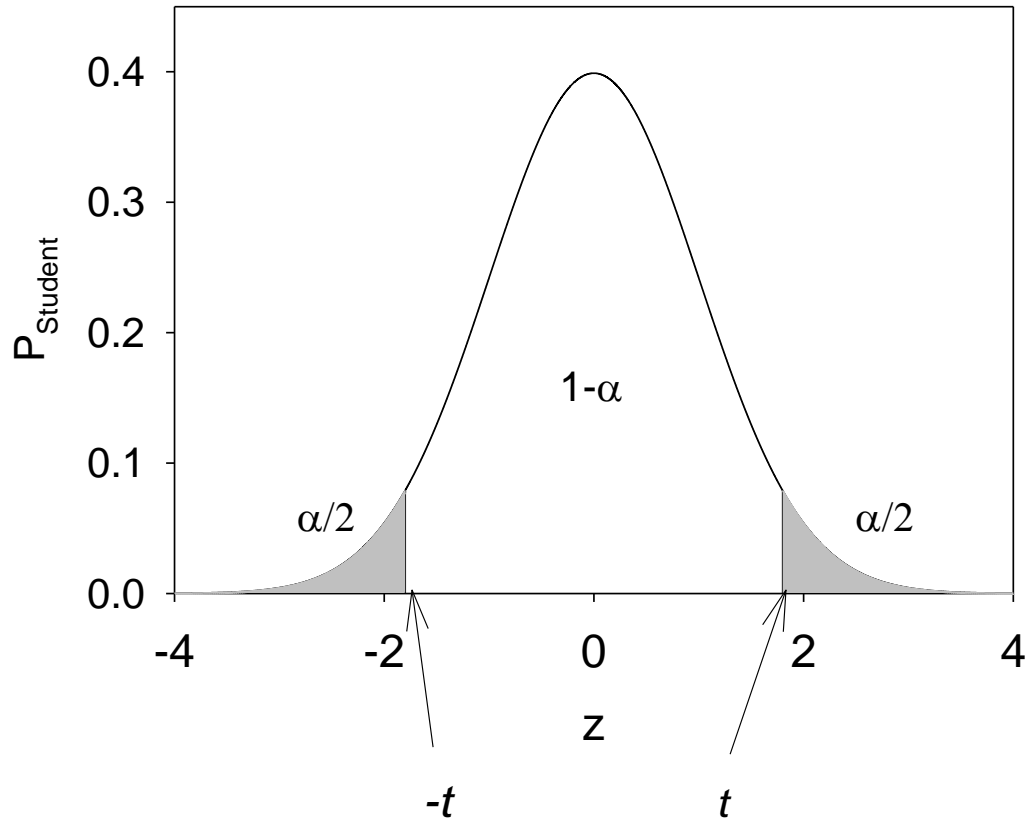
English : NORMSINV(α)



$t(\alpha, k) = \text{LOI.STUDENT.INVERSE}(\alpha, k)$

English : $\text{TINV}(\alpha, k)$

$t(\alpha, k)$



Français

English

$F(\alpha, D_1, D_2) = \text{INVERSE.LOI.F}(\alpha, D_1, D_2)$

$\text{FINV}(\alpha, D_1, D_2)$

$\chi^2 = \text{KHIDEUX.INVERSE}(\alpha, k)$

$\text{CHIINV}(\alpha, k)$

ECARTYPE; p. ex. ECARTYPE (A1:A10)

STDEV

SOMME

SUM

MOYENNE

AVERAGE

Fonctions complexes

Insérer fonction f_x , choisir Scientifiques
(les cellules A1, A2... sont montrées à titre d'exemple)

COMPLEXE.REEL(A1) extraire la partie réelle d'un nombre complexe (texte)

COMPLEXE.IMAGINAIRE(A1) extraire la partie imaginaire d'un nombre complexe (texte)

Exemple :

$$A1 = 3 - 4i$$

$$\text{COMPLEXE.REEL}(A1) = 3$$

$$\text{COMPLEXE.IMAGINAIRE}(A1) = -4$$

COMPLEXE(A1,A2) changer en nombre complexe deux nombres réels, A1 partie réelle, A2 partie imaginaire

Exemple :

$$B1 = 3 \quad B2 = -4$$

$$\text{COMPLEXE}(B1,B2) = 3 - 4i$$

COMPLEXE.MODULE(A1) calculer un module d'un nombre complexe (texte)

$$A1 = 3 - 4i$$

$$\text{COMPLEXE.MODULE}(A1) = 5$$

COMPLEXE.ARGUMENT(A1) calculer un angle en radians d'un nombre complexe (texte)

$$A1 = 3 - 4i$$

$$\text{COMPLEXE.ARGUMENT}(A1) = -0.9273 \text{ rad} = -53.13^\circ$$

13. Références

G. Baillargeon, Probabilité, statistique et techniques de régression, Les Éditions SMG, 1989.

D.B. Hibberg, J.J. Gooding, Data Analysis for Chemistry: An Introductory Guide for Students and Laboratory Scientists, OXFORD UNIVERSITY PRESS, disponible sur l'Internet de l'Université (pdf).

E. Oran Brigham, The Fast Fourier Transform, Prentice-Hall, Englewood Cliffs, New Jersey, 1974.

J.N. Miller, J.C. Miller, Statistics and Chemometrics for Analytical Chemistry, Prentice Hall, 2000.

J. Maurice, Jugement statistique sur échantillons en chimie, Polytechnica, 1993.

N. Gilbert, J.G. Savard, Statistiques, Édition Études Vivantes, 1992.

Autres

- R.G. Brereton, Chemometrics. Data Analysis for the Laboratory and Chemical Plant, Wiley, 2003.
- Data Reduction and Error Analysis for the Physical Sciences Second Edition. P. R. Bevington, D. Keith Robinson, McGraw-Hill.
- P.C. Meier, R.E. Zünd, Statistical Methods in Analytical Chemistry, Wiley, 2000.
- Mathematics for Physical Chemistry. Robert G. Mortimer, MacMillan.
- Essentials of Statistics for Scientists and Technologists. C. Mack, Édition Plenum/Rosetta.
- The Analysis of Physical Measurements, Pugh-Windslow, Addison-Wesley.

14.2 Table des valeurs $z(\alpha)$

<u>%</u>	α	z
50	0.5	0.67449
80	0.2	1.28155
90	0.1	1.64485
95	0.05	1.95996
99	0.01	2.57583
99.9	0.001	3.29048

14.3 Table de la distribution de Student, $t(\alpha, k)$ pour $\alpha = 0.1, 0.05$ et 0.01

%		90%	95%	99%
k		0.1	0.05	0.01
1	6.313749	12.70615	63.6559	
2	2.919987	4.302656	9.924988	
3	2.353363	3.182449	5.840848	
4	2.131846	2.776451	4.60408	
5	2.015049	2.570578	4.032117	
6	1.943181	2.446914	3.707428	
7	1.894578	2.364623	3.499481	
8	1.859548	2.306006	3.355381	
9	1.833114	2.262159	3.249843	
10	1.812462	2.228139	3.169262	
11	1.795884	2.200986	3.105815	
12	1.782287	2.178813	3.054538	
13	1.770932	2.160368	3.012283	
14	1.761309	2.144789	2.976849	
15	1.753051	2.131451	2.946726	
16	1.745884	2.119905	2.920788	
17	1.739606	2.109819	2.898232	
18	1.734063	2.100924	2.878442	
19	1.729131	2.093025	2.860943	
20	1.724718	2.085962	2.845336	
21	1.720744	2.079614	2.831366	
22	1.717144	2.073875	2.818761	
23	1.71387	2.068655	2.807337	
24	1.710882	2.063898	2.796951	
25	1.70814	2.059537	2.787438	
26	1.705616	2.055531	2.778725	
27	1.703288	2.051829	2.770685	
28	1.70113	2.048409	2.763263	
29	1.699127	2.045231	2.756387	
30	1.69726	2.04227	2.749985	

14.4 Table de la distribution F, $F(\alpha, D_2, D_1)$ pour $\alpha = 0.05$

		D ₂ (numérateur)																		
D ₁ (denominateur)																				
	1	2	3	4	5	6	7	8	9	10	12	14	16	20	24	30	40	60	100	
1	161.4	199.5	215.7	224.6	230.2	234	236.8	238.9	240.5	241.9	243.9	245.4	246.5	248	249.1	250.1	251.1	252.2	253	
2	18.51	19	19.16	19.25	19.3	19.33	19.35	19.37	19.38	19.4	19.41	19.42	19.43	19.45	19.45	19.46	19.47	19.48	19.49	
3	10.13	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.785	8.745	8.715	8.692	8.66	8.638	8.617	8.594	8.572	8.554	
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.912	5.873	5.844	5.803	5.774	5.746	5.717	5.688	5.664	
5	6.608	5.786	5.409	5.192	5.05	4.95	4.876	4.818	4.772	4.735	4.678	4.636	4.604	4.558	4.527	4.496	4.464	4.431	4.405	
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.06	4	3.956	3.922	3.874	3.841	3.808	3.774	3.74	3.712	
7	5.591	4.737	4.347	4.12	3.972	3.866	3.787	3.726	3.677	3.637	3.575	3.529	3.494	3.445	3.41	3.376	3.34	3.304	3.275	
8	5.318	4.459	4.066	3.838	3.688	3.581	3.5	3.438	3.388	3.347	3.284	3.237	3.202	3.15	3.115	3.079	3.043	3.005	2.975	
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.23	3.179	3.137	3.073	3.025	2.989	2.936	2.9	2.864	2.826	2.787	2.756	
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.02	2.978	2.913	2.865	2.828	2.774	2.737	2.7	2.661	2.621	2.588	
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.788	2.739	2.701	2.646	2.609	2.57	2.531	2.49	2.457	
12	4.747	3.885	3.49	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.687	2.637	2.599	2.544	2.505	2.466	2.426	2.384	2.35	
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.604	2.554	2.515	2.459	2.42	2.38	2.339	2.297	2.261	
14	4.6	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.534	2.484	2.445	2.388	2.349	2.308	2.266	2.223	2.187	
15	4.543	3.682	3.287	3.056	2.901	2.79	2.707	2.641	2.588	2.544	2.475	2.424	2.385	2.328	2.288	2.247	2.204	2.16	2.123	
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.425	2.373	2.333	2.276	2.235	2.194	2.151	2.106	2.068	
17	4.451	3.592	3.197	2.965	2.81	2.699	2.614	2.548	2.494	2.45	2.381	2.329	2.289	2.23	2.19	2.148	2.104	2.058	2.02	
18	4.414	3.555	3.16	2.928	2.773	2.661	2.577	2.51	2.456	2.412	2.342	2.29	2.25	2.191	2.15	2.107	2.063	2.017	1.978	
19	4.381	3.522	3.127	2.895	2.74	2.628	2.544	2.477	2.423	2.378	2.308	2.256	2.215	2.155	2.114	2.071	2.026	1.98	1.94	
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.278	2.225	2.184	2.124	2.082	2.039	1.994	1.946	1.907	
21	4.325	3.467	3.072	2.84	2.685	2.573	2.488	2.42	2.366	2.321	2.25	2.197	2.156	2.096	2.054	2.01	1.965	1.916	1.876	
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297	2.226	2.173	2.131	2.071	2.028	1.984	1.938	1.889	1.849	
23	4.279	3.422	3.028	2.796	2.64	2.528	2.442	2.375	2.32	2.275	2.204	2.15	2.109	2.048	2.005	1.961	1.914	1.865	1.823	
24	4.26	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.3	2.255	2.183	2.13	2.088	2.027	1.984	1.939	1.892	1.842	1.8	
25	4.242	3.385	2.991	2.759	2.603	2.49	2.405	2.337	2.282	2.236	2.165	2.111	2.069	2.007	1.964	1.919	1.872	1.822	1.779	
26	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321	2.265	2.22	2.148	2.094	2.052	1.99	1.946	1.901	1.853	1.803	1.76	
27	4.21	3.354	2.96	2.728	2.572	2.459	2.373	2.305	2.25	2.204	2.132	2.078	2.036	1.974	1.93	1.884	1.836	1.785	1.742	
28	4.196	3.34	2.947	2.714	2.558	2.445	2.359	2.291	2.236	2.19	2.118	2.064	2.021	1.959	1.915	1.869	1.82	1.769	1.725	
29	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278	2.223	2.177	2.104	2.05	2.007	1.945	1.901	1.854	1.806	1.754	1.71	
30	4.171	3.316	2.922	2.69	2.534	2.421	2.334	2.266	2.211	2.165	2.092	2.037	1.995	1.932	1.887	1.841	1.792	1.74	1.695	
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.18	2.124	2.077	2.003	1.948	1.904	1.839	1.793	1.744	1.693	1.637	1.589	
50	4.034	3.183	2.79	2.557	2.4	2.286	2.199	2.13	2.073	2.026	1.952	1.895	1.85	1.784	1.737	1.687	1.634	1.576	1.525	
60	4.001	3.15	2.758	2.525	2.368	2.254	2.167	2.097	2.04	1.993	1.917	1.86	1.815	1.748	1.7	1.649	1.594	1.534	1.481	
80	3.96	3.111	2.719	2.486	2.329	2.214	2.126	2.056	1.999	1.951	1.875	1.817	1.772	1.703	1.654	1.602	1.545	1.482	1.426	
100	3.936	3.087	2.696	2.463	2.305	2.191	2.103	2.032	1.975	1.927	1.85	1.792	1.746	1.676	1.627	1.573	1.515	1.45	1.392	
120	3.92	3.072	2.68	2.447	2.29	2.175	2.087	2.016	1.959	1.91	1.834	1.775	1.728	1.659	1.608	1.554	1.495	1.429	1.369	
∞	3.843	2.998	2.607	2.374	2.216	2.1	2.011	1.94	1.882	1.833	1.754	1.694	1.646	1.573	1.519	1.461	1.396	1.321	1.247	

14.5 Table de la distribution $\chi^2(\alpha, k)$

k	α		
	0.1	0.05	0.01
1	2.705541	3.841455	6.634891
2	4.605176	5.991476	9.210351
3	6.251394	7.814725	11.34488
4	7.779434	9.487728	13.2767
5	9.236349	11.07048	15.08632
6	10.64464	12.59158	16.81187
7	12.01703	14.06713	18.47532
8	13.36156	15.50731	20.09016
9	14.68366	16.91896	21.66605
10	15.98717	18.30703	23.20929
11	17.27501	19.67515	24.72502
12	18.54934	21.02606	26.21696
13	19.81193	22.36203	27.68818
14	21.06414	23.68478	29.14116
15	22.30712	24.9958	30.57795
16	23.54182	26.29622	31.99986
17	24.76903	27.5871	33.40872
18	25.98942	28.86932	34.80524
19	27.20356	30.14351	36.19077
20	28.41197	31.41042	37.56627
21	29.61509	32.67056	38.93223
22	30.81329	33.92446	40.28945
23	32.00689	35.17246	41.63833
24	33.19624	36.41503	42.97978
25	34.38158	37.65249	44.31401
26	35.56316	38.88513	45.64164
27	36.74123	40.11327	46.96284
28	37.91591	41.33715	48.27817
29	39.08748	42.55695	49.58783
30	40.25602	43.77295	50.89218

14.6 Valeurs de Q pour le rejet d'un point

Nombre d'observations	Q (90% de confiance)	Q (95% de confiance)	Q (99% de confiance)
3	0.941	0.970	0.994
4	0.765	0.829	0.926
5	0.642	0.710	0.821
6	0.560	0.625	0.740
7	0.507	0.568	0.680
8	0.468	0.526	0.634
9	0.437	0.493	0.598
10	0.412	0.466	0.568