

S1 METHODES STATISTIQUES

I.- INTRODUCTION

Lorsque l'on réalise une étude expérimentale, les résultats que l'on obtient sont en général entachés d'une incertitude. Deux questions essentielles surgissent alors; comment obtenir un résultat expérimental aussi proche que possible du résultat exacte et deuxièmement quelle est l'erreur d'une mesure unique? Pour obtenir ces informations, on a recours à des méthodes statistiques qui fournissent des outils simples qui permettent d'évaluer un résultat moyen, son écart par rapport à la valeur exacte, l'évolution de cet écart en fonction du nombre de mesures et finalement l'écart type pour une seule mesure.

II.- THEORIE

Supposons que lors d'une expérience, on a réalisé un ensemble de N mesures $S = \{x_1, x_2, x_3, \dots, x_N\}$. En général, on représente ces résultats sous la forme d'un histogramme où l'on reporte le nombre de fois que le résultat x est obtenu en fonction de x (voir exemple de la figure 1). Dans le cas où le résultat est un nombre réel, on divise l'axe des réels en intervalles et on réalise l'histogramme en portant le nombre de fois que le résultat apparaît sur un intervalle.

De cet ensemble S on peut calculer la valeur moyenne μ des mesures et l'écart quadratique moyen σ^2 ou l'écart standard σ :

$$(1) \quad \mu = \frac{\sum_{i=1}^N x_i}{N}$$

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

Remarque: on trouve parfois une définition moins rigoureuse de σ où le facteur N-1 est remplacé par N dans le calcul de σ . Si $N \gg 1$, les deux définitions de σ sont pratiquement identiques.

Statistique sur un ensemble de moyenne

De l'ensemble S, on peut extraire un sous-ensemble de n mesures que l'on appelle échantillonnage dont on peut calculer sa valeur moyenne \bar{x} .

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

On peut construire un nouvel ensemble $\bar{S} = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k\}$ contenant les valeurs moyennes de n mesures. On a alors une distribution des valeurs moyennes dont on peut calculer la valeur moyenne $\mu_{\bar{x}}$ et son écart standard $\sigma_{\bar{x}}$

$$(3) \quad \mu_{\bar{x}} = \frac{\sum_{i=1}^k \bar{x}_i}{k}$$

$$(4) \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad \text{où } \sigma \text{ est calculé à l'aide de (2)}$$

Ainsi l'écart standard des moyennes diminue avec la racine carrée du nombre de mesures de chaque moyenne. De plus quel que soit le type de distribution, on peut mathématiquement démontrer (théorème de la limite centrale) que la distribution des valeurs moyennes d'un grand nombre d'expériences tend vers une distribution unique appelée distribution normale ou de Gauss.

Illustrons ces concepts à l'aide d'un exemple.

Le tableau I donne les résultats d'une série de $N=100$ mesures du nombre n de désintégrations par seconde produites par un élément radioactif.

Tableau I

0	3	1	5	2	2	3	1	3	3	2	4	4	3	5	3	4	1	2	2
2	5	1	3	0	2	4	3	4	2	1	3	2	4	3	1	6	3	2	6
5	1	3	2	2	4	3	5	6	2	0	7	2	1	6	4	2	2	3	4
3	3	2	1	4	5	3	3	1	0	6	4	3	2	5	4	4	7	1	3
2	4	2	1	1	3	2	4	2	3	4	2	3	2	5	8	4	1	4	5

Le tableau II donne le nombre de fois N_n et la probabilité $P_n = N_n/N$ d'avoir n désintégrations.

Tableau II

n	0	1	2	3	4	5	6	7	8	9
N_n	4	14	24	23	18	9	5	2	1	0
P_n	0.04	0.14	0.24	0.23	0.18	0.09	0.05	0.02	0.01	0

La Figure 1 montre la distribution de N_n et de P_n en fonction de n . La forme de la distribution de P_n est identique à celle de N_n .

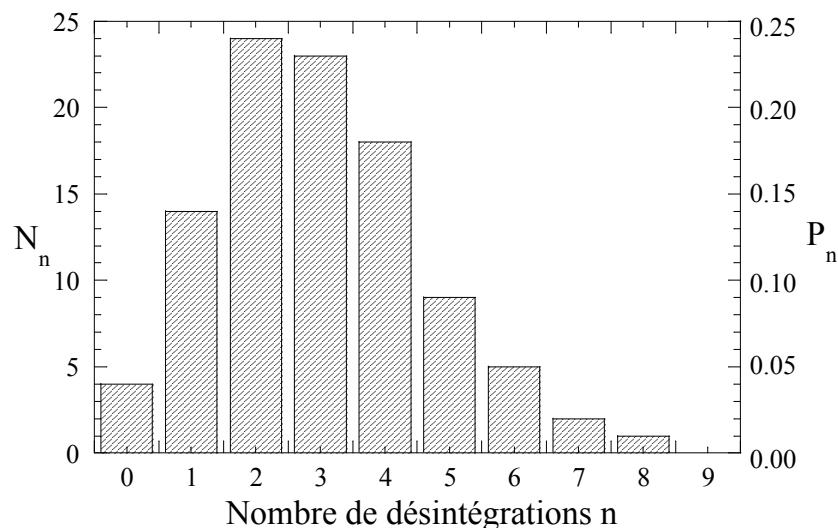


Figure 1.- Histogramme ou distribution des mesures de désintégrations

En utilisant les relations (1) et (2), on peut calculer la valeur moyenne du nombre de désintégrations ainsi que son écart quadratique moyen. Ces calculs sont illustrés dans le tableau III

Tableau III

n	N _n	n·N _n	N _n (n-μ) ²
0	4	0	36
1	14	14	56
2	24	48	24
3	23	69	0
4	18	72	18
5	9	45	36
6	5	30	45
7	2	14	32
8	1	8	25
9	0	0	0
$N = \sum N_n$		$\mu = \frac{\sum N_n \cdot n}{N}$	$\sigma^2 = \frac{\sum N_n (n - \mu)^2}{N - 1}$
100		3	2.75

On constate que le nombre moyen de désintégration est 3 avec un écart quadratique moyen de 2.75. Ainsi nous pouvons donner le résultat avec son incertitude sous la forme

$$n = 3 \pm \sqrt{2.75} \approx 3 \pm 2$$

La forme de la distribution dépend de la nature du phénomène étudié. Il existe beaucoup d'autres distributions mais nous restreindrons ce chapitre à l'étude de deux distributions: la distribution de Poisson et celle de Gauss (ou distribution normale)

Dans l'exemple du Tableau II, on constate que les mesures relatives à n=1,2,3,4 et 5 sont les mesures dont les intervalles d'incertitude (n±2) englobent la valeur moyenne μ = 3. Ces mesures représentent 87% des mesures totales. Ainsi l'écart standard donne une probabilité

de trouver un résultat expérimental dans un intervalle de largeur 2σ centré sur la valeur moyenne μ . La valeur de cette probabilité dépend de la forme de la distribution, mais pour une distribution normale ou de Gauss cette probabilité est de 68%.

Statistique sur les moyennes

Reprenons les mesures du tableau I pour déterminer un nouvel ensemble de 10 moyennes obtenus en calculant la moyenne des résultats pour 10 mesures.

										\bar{n}
0	3	1	5	2	2	3	1	3	3	2.3
2	4	4	3	5	3	4	1	2	2	3.0
2	5	1	3	0	2	4	3	4	2	2.6
1	3	2	4	3	1	6	3	2	6	3.1
5	1	3	2	2	4	3	5	6	2	3.3
0	7	2	1	6	4	2	2	3	4	3.1
3	3	2	1	4	5	3	3	1	0	2.5
6	4	3	2	5	4	4	7	1	3	3.9
2	4	2	1	1	3	2	4	2	3	2.4
4	2	3	2	5	8	4	1	4	5	3.8

On peut alors compter les résultats compris dans les intervalles $[j, j+1[$ ainsi que la moyenne des \bar{n} et son écart standard $\sigma_{\bar{n}}$.

$[0.5, 1.5 [$	$[1.5, 2.5 [$	$[2.5, 3.5 [$	$[3.5, 4.5 [$	$[4.5, 5.5 [$	$\mu_{\bar{n}}$	$\sigma_{\bar{n}}$
0	2	6	2	0	3	0.56

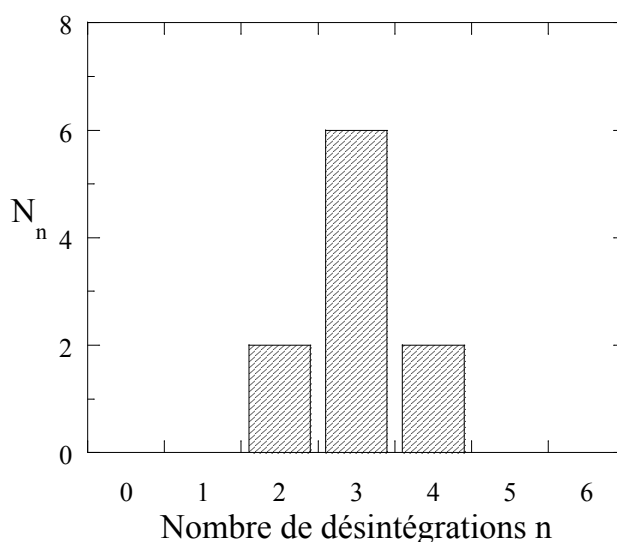


Figure 2.- Distribution des moyennes sur 10 mesures

La Figure 2 montre que la statistique des moyennes change de forme en devenant plus symétrique et que l'écart type de la statistique des moyennes sur 10 mesures diminue de 1.66 à 0.56.

$$\sigma_{\bar{n}} = 0.56 \cong \frac{\sigma}{\sqrt{k}} = \frac{1.66}{\sqrt{10}} = 0.52$$

Ce résultat montre que la relation (4) est bien vérifiée et donc que l'écart standard diminue comme la racine carrée du nombre de mesures. Nous pouvons alors utiliser cette observation pour donner la meilleure estimation de μ sur la base des 100 mesures:

$$(\text{meilleure estimation de } \mu) \quad \bar{n}_{100} \pm \frac{\sigma}{\sqrt{100}} = 3 \pm \frac{1.66}{10} \cong 3 \pm 0.2$$

DISTRIBUTION DE POISSON

Distribution binomiale

Considérons un événement qui a une probabilité p de se produire. Dans une population de M éléments on désire savoir la probabilité que l'événement se produise 0 fois, 1 fois, 2 fois etc... Si tous les éléments sont identiques, ces probabilités sont données par la distribution binomiale $p_M(n)$ où n est le nombre de fois que l'événement se produit:

$$p_M(n) = \frac{M!}{n!(M-n)!} p^n (1-p)^{(M-n)}$$

$$n! = n \cdot (n-1) \cdot \dots \cdot 3 \cdot 2 \cdot 1$$

Distribution de Poisson

La distribution de Poisson est une approximation de la distribution binomiale dans le cas où la probabilité p qu'un événement se réalise est très petite ($p \ll 1$) mais que la population M sur laquelle on va observer cet événement est très grande ($M \gg 1$). Dans ce cas, la valeur moyenne μ est donnée par:

$$\mu = p M$$

et la probabilité d'observer n événements est donnée par la loi de Poisson:

$$(5) \quad P(n) = \frac{\mu^n}{n!} \text{Exp}(-\mu) \quad n = 0, 1, 2, \dots, \infty \quad \text{avec} \quad \sum_{n=0}^{\infty} P(n) = 1$$

μ valeur moyenne

$\sigma = \sqrt{\mu}$ écart standard

Reprenons l'exemple précédent de la radioactivité où nous observons que le nombre de désintégrations dans un intervalle de temps fixé fluctue fortement. En répétant un grand nombre de fois la même mesure, on constate que chaque résultat apparaît avec une certaine probabilité qui suit une loi statistique bien précise : la loi de Poisson. La Figure 3 montre

que les prédictions d'une statistique de Poisson avec $\mu=3$ décrivent très bien les résultats expérimentaux

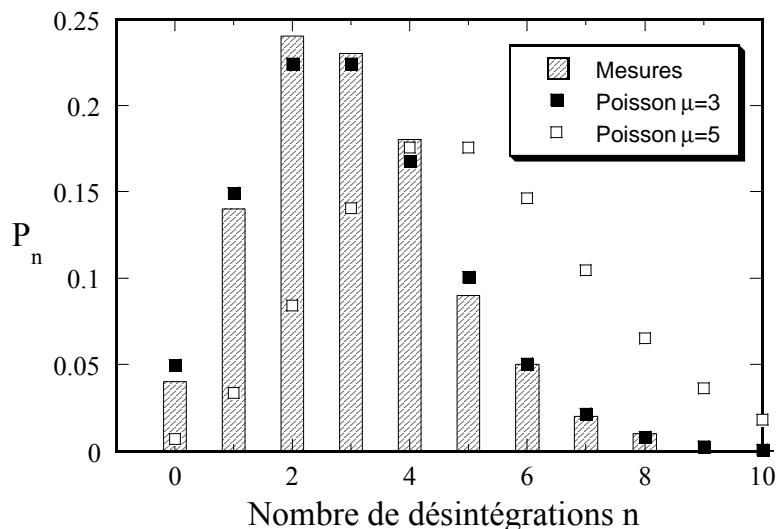


Figure 3.- Comparaison des résultats avec des distributions de Poisson

Lorsqu'une statistique suit la loi de Poisson, il n'est pas nécessaire de connaître expérimentalement la distribution pour en déterminer son écart quadratique moyen puisque $\sigma^2=\mu$. Pour illustrer ceci, calculons le nombre total d'événements $Z = \sum n = 300$ obtenus pendant 100 secondes. L'écart standard est $\sqrt{300} \cong 17.2$. Le nombre de désintégrations par seconde est alors donné par

$$\frac{Z \pm \sqrt{Z}}{100} = 3 \pm \frac{17.2}{100} \cong 3 \pm 0.2$$

C'est le même résultat que celui obtenu précédemment. Ceci montre que l'erreur relative diminue avec le nombre total d'événements mesurés.

$$\frac{\Delta Z}{Z} = \frac{\sqrt{Z}}{Z} = \frac{1}{\sqrt{Z}}$$

On peut donc augmenter la précision d'une mesure d'intensité de rayonnement nucléaire en augmentant l'intervalle de temps, par conséquent aussi le nombre Z de particules comptées.

DISTRIBUTION DE GAUSS OU DISTRIBUTION NORMALE

Lorsque le résultat d'une mesure varie aléatoirement et de manière continue autour d'une certaine valeur moyenne μ , on observe souvent que cette distribution est correctement décrite par la distribution de Gauss.

$$(6) \quad f_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \text{Exp} \left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}$$

où μ est la valeur moyenne et σ l'écart standard.

La Figure 4 montre la prédiction théorique pour une expérience mesurant l'accélération terrestre avec une valeur moyenne de 9.8 m/s^2 et un écart standard $\sigma=0.1$ pour la courbe en pointillé et de 0.2 pour la courbe continue.

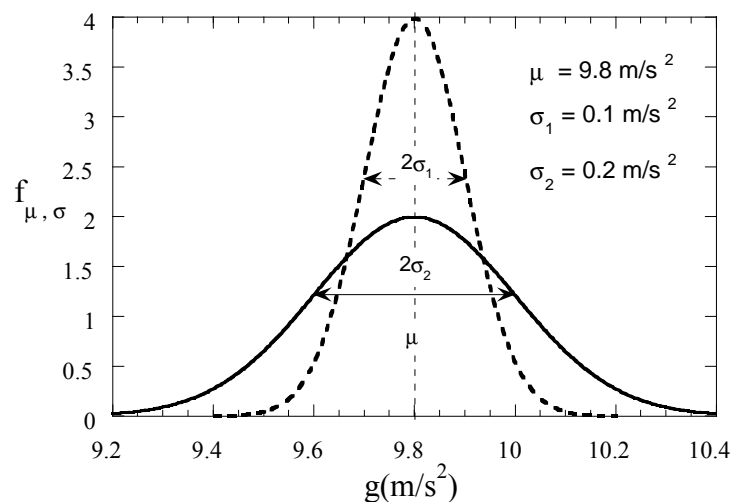


Figure 4.- Distribution normale ou de Gauss pour deux écarts types σ

$f_{\mu,\sigma}(x)$ ne représente pas la probabilité d'avoir le résultat x mais ce que l'on appelle une densité de probabilité. En effet demander quelle est la probabilité d'obtenir le résultat $g=9.8$ n'a pas de sens, sauf si on donne un intervalle autour de cette valeur, par exemple ± 0.1 qui correspond à l'écart standard. Ainsi la probabilité de trouver un résultat compris entre 9.7 et 9.9 est une question sensée. Cette probabilité est donnée par la surface sous la courbe $f_{\mu,\sigma}(x)$ dans l'intervalle $[9.7,9.9]$. Sur la figure 5 cette surface est hachurée. Mathématiquement cela se traduit par l'évaluation d'un intégrale de la forme:

$$P([a, b]) = \int_a^b f_{\mu,\sigma} dx$$

Ainsi, la probabilité de trouver un résultat entre $]-\infty, +\infty[$ doit être de 100%, ceci quels que soient la valeur moyenne et l'écart standard. Ce résultat se démontre mathématiquement en calculant l'intégrale ci-dessus. Pour terminer mentionnons que la probabilité de trouver un résultat dans l'intervalle 2σ entourant la valeur moyenne ($9.8 \pm 0.1 \text{ m/s}^2$ dans l'exemple de la Figure 5) est approximativement de 68%. Donner cet intervalle de confiance ne signifie pas qu'un résultat en dehors de cet intervalle est un mauvais résultat mais que l'on a environ une chance sur 3 d'obtenir un résultat qui ne se situera pas dans l'intervalle de confiance.

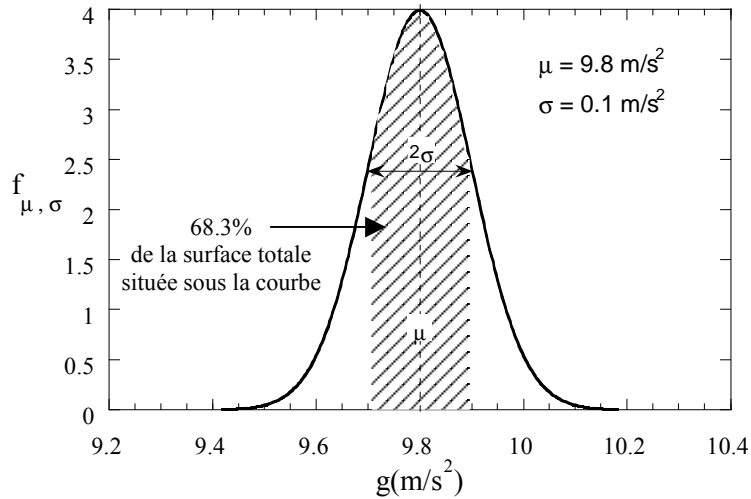


Figure 5.- Illustration de la notion de probabilité pour une distribution normale

Pour les physiciens :

Nous venons de voir que l'incertitude sur un ensemble de mesures est décrite par la déviation standard. Dans le chapitre d'introduction sur les erreurs de mesures, nous avons donné deux méthodes pour calculer l'erreur δF d'une fonction de plusieurs variables $F(x, \dots, z)$. Si les variables sont indépendantes l'erreur δF est donnée par

$$(7) \quad (\delta F)^2 = \left(\frac{\partial F}{\partial x} \delta x\right)^2 + \dots + \left(\frac{\partial F}{\partial z} \delta z\right)^2$$

mais dans tous les cas l'erreur δF est inférieure à la simple somme

$$(8) \quad \delta F = \left| \frac{\partial F}{\partial x} \delta x \right| + \dots + \left| \frac{\partial F}{\partial z} \delta z \right|$$

Exemple : Considérons la mesure de la surface d'un rectangle à l'aide d'un mètre roulant. Comme il est difficile de prévoir l'erreur réel de la mesure (erreur du mètre, du géomètre), on réalise une série de 10 mesures, pour chacune des dimensions du rectangle, pour déterminer les valeurs moyennes, l'erreur standard et l'erreur standard de la moyenne en appliquant les relations (1), (2) et (4)

Valeurs mesurées en [m]										μ	σ	σ_μ	
x	24.25	24.26	24.22	24.28	24.24	24.25	24.22	24.26	24.23	24.23	24.245	0.019	0.006
y	50.36	50.35	50.41	50.37	50.36	50.32	50.39	50.38	50.36	50.38	50.368	0.024	0.008

Si les variables x et y sont indépendantes, l'erreur obtenue est déduite de la relation (7) où $F(x,y)=xy$ alors $\frac{\partial F}{\partial x} = y$ et $\frac{\partial F}{\partial y} = x \Rightarrow (\delta F)^2 = (y\delta x)^2 + (x\delta y)^2 \equiv (\bar{y}\sigma_x)^2 + (\bar{x}\sigma_y)^2$

et la surface du rectangle sera de $(1221.2 \pm 0.4)\text{m}^2$

Le calcul basé sur la relation (8) donnerait une erreur légèrement plus grande de $\pm 0.5\text{m}^2$.

REGRESSION LINEAIRE

La régression linéaire d'un ensemble de points x_i, y_i est une droite $Y = Ax + B$ qui décrit les résultats x_i, y_i avec le minimum d'erreur. Pour un x_i fixé, l'écart entre le résultat y_i et la prédiction de la régression linéaire $Y_i = Ax_i + B$ est $\Delta Y_i = y_i - Y_i$. L'erreur moyenne ΔY et l'erreur quadratique moyenne $(\Delta Y)^2$ sont données par:

$$\Delta Y = \frac{1}{N} \sum_{i=1}^N \Delta Y_i$$

$$(\Delta Y)^2 = \frac{1}{N} \sum_{i=1}^N (\Delta Y_i)^2$$

Les coefficients A et B, tels que $\Delta Y = 0$ et $(\Delta Y)^2$ soit minimum, sont donnés par les deux relations suivantes

$$(9) \quad B = \frac{(\sum y_i)(\sum x_i^2) - (\sum x_i)(\sum x_i y_i)}{N \sum (x_i^2) - (\sum x_i)^2}$$

$$A = \frac{(\sum y_i) - NB}{(\sum x_i)}$$

On définit alors l'écart standard σ_y qui représente la valeur moyenne de l'écart quadratique moyen par rapport à Y. Remarquez le facteur N-2 dans la définition de σ_y . Cela provient du fait qu'avec 2 points, on obtient une droite passant exactement par ces 2 points et que la notion d'erreur ne peut débiter qu'avec le troisième point. Comme la régression linéaire est basée sur la distribution de Gauss des erreurs, l'écart standard nous indique que 68% des résultats expérimentaux y_i sont dans un intervalle 2σ centré sur la droite. Ainsi la droite $Y = Ax + B$ nous permet de prédire que $Y_i \pm \sigma$ à 68% de chance d'englober le résultat expérimental y_i .

$$(10) \quad \sigma_y = \sqrt{\frac{\sum (y_i - Y_i)^2}{N - 2}}$$

Une mesure du *bon sens* de la régression est fournie par le coefficient de corrélation r

$$(11) \quad r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad \text{avec } \bar{x} = \frac{\sum x_i}{N} \quad \text{et } \bar{y} = \frac{\sum y_i}{N}$$

$r = \pm 1$ signifie que la corrélation est parfaite.

$r = 0$ nous informe que la régression n'a aucun sens.

Un r positif signifie que y croît avec x, alors qu'un r négatif indique que y décroît avec x

Finalement, dans de nombreuses situations on désire donner les coefficients A et B avec une incertitude σ_A et σ_B . En utilisant les règles de propagation des erreurs, on obtient les expressions suivantes:

$$(12) \quad \sigma_A^2 = \frac{N\sigma_y^2}{N\sum x_i^2 - (\sum x_i)^2}$$

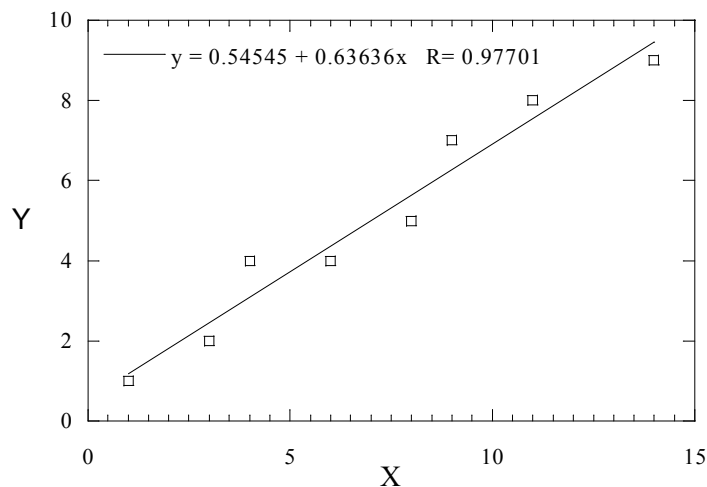
$$\sigma_B^2 = \sigma_y^2 \frac{\sum x_i^2}{N\sum x_i^2 - (\sum x_i)^2}$$

Exemple : Considérons les 8 points x,y donnés dans le tableau ci-dessous

x	1	3	4	6	8	9	11	14	$\sum x = 56$
y	1	2	4	4	5	7	8	9	$\sum y = 40$
x^2	1	9	16	36	64	81	121	196	$\sum x^2 = 524$
xy	1	6	16	24	40	63	88	126	$\sum xy = 364$
y^2	1	4	16	16	25	49	64	81	$\sum y^2 = 256$

$$B = \frac{40 \cdot 524 - 56 \cdot 364}{8 \cdot 524 - 56^2} = \frac{6}{11} = 0.545$$

$$A = \frac{40 - 8 \cdot \frac{6}{11}}{56} = \frac{7}{11} = 0.636$$



Le calcul de l'écart type se fait avec la relation (10) et l'équation $Y = Ax + B$

x	1	3	4	6	8	9	11	14	$\bar{x} = 7$
y	1	2	4	4	5	7	8	9	$\bar{y} = 5$
Y	1.1818	2.4545	3.0909	4.3636	5.6364	6.2727	7.5455	9.4545	
y-Y	-0.1818	-0.4545	0.9091	-0.3636	-0.6364	0.7273	0.4455	-0.4545	$\sum (x - Y) = 0$
$(y-Y)^2$	0.0331	0.2066	0.8264	0.1322	0.4050	0.5289	0.2066	0.2066	$\sum (y - Y)^2 = 1.4793$

$$\sigma_y = \sqrt{\frac{\sum (y - \bar{Y})^2}{N - 2}} = \sqrt{\frac{1.4793}{8 - 2}} = 0.651$$

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} = \frac{84}{\sqrt{132 \cdot 56}} = 0.977$$

$$\sigma_A = \sigma_y \sqrt{\frac{N}{N \sum x^2 - (\sum x)^2}} = 0.651 \sqrt{\frac{8}{8 \cdot 524 - 56^2}} = 0.0567$$

$$\sigma_B = \sigma_y \sqrt{\frac{\sum x^2}{N \sum x^2 - (\sum x)^2}} = 0.651 \sqrt{\frac{524}{8 \cdot 524 - 56^2}} = 0.459$$

III.- EXERCICES ET EXPERIENCES

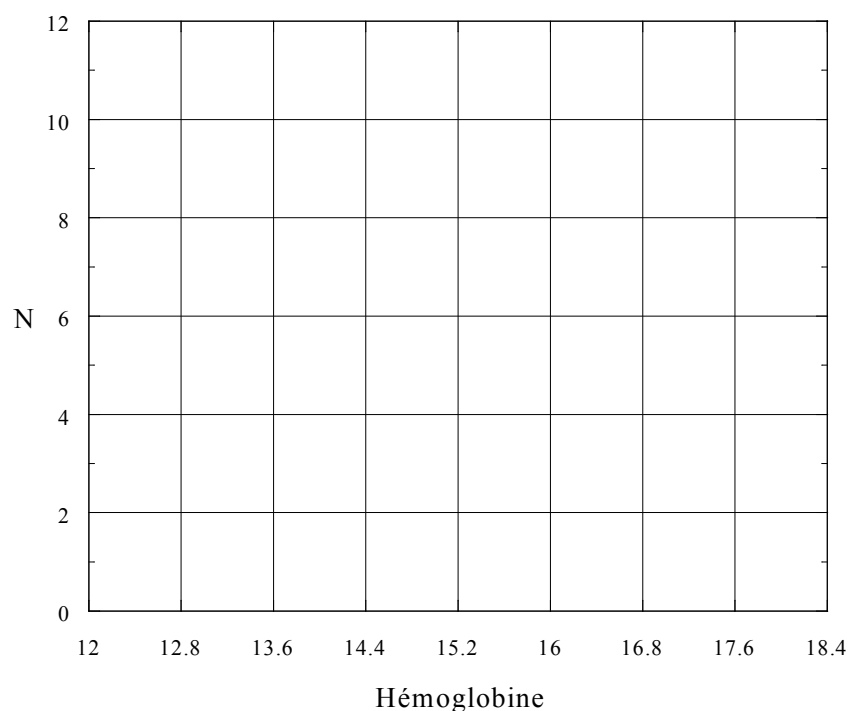
Mesure du taux d'hémoglobine sur 30 patients.

15.6	14.1	12.1	15.0	16.9	17.5
15.3	14.5	13.6	12.8	16.2	17.5
15.1	14.2	14.0	12.6	14.8	15.1
16.8	16.4	14.4	13.0	16.2	15.0
13.7	13.4	15.3	15.0	16.1	15.2

a) Calculer la valeur moyenne et l'écart standard de cet ensemble statistique.

b) Compléter le tableau ci-dessous et donner les résultats sous la forme d'un histogramme.

Intervalle	[12.0, 12.8[[12.8, 13.6[[13.6, 14.4[[14.4, 15.2[[15.2, 16.0[[16.0-16.8[[16.8, 17.6[
N							



LOI DE POISSON

1- Supposons que 10% d'un lot de pilule sont inefficaces. Trouver la probabilité que deux pilules soient inefficaces dans un lot de 10 en utilisant:

- a) la distribution binomiale
- b) la distribution de Poisson

2- Supposons que la probabilité qu'un individu présente des complications suite à un vaccin soit de 0.001. Sur une population de 2000 individus, déterminer la probabilité que:

- a) exactement 3 personnes présentent des complications,
- b) aucune personne ne présente des complications,
- c) au moins une personne présente des complications

Expériences

Pour cette expérience, nous avons mesuré la radioactivité d'une source pendant un intervalle de temps court afin que le nombre d'événements soit petit. Ces données sont stockées dans l'ordinateur.

- Suivre l'évolution de la distribution des mesures et remarquez à partir de combien d'intervalles, l'allure de cette distribution perd son caractère aléatoire.
- Une distribution selon la loi de Poisson est ensuite dessinée avec pour paramètre la moyenne calculée. Faites varier la moyenne arbitrairement et constatez l'effet sur la distribution calculée.

LOI DE GAUSS

Expérience

Pour mettre cette distribution en évidence, nous allons regarder l'évolution de la valeur moyenne lorsque l'on réalise une moyenne sur 1,2,5,10,20,50 mesures. On verra que l'écart standard $\sigma(n)$ diminue avec le nombre de mesures n comme:

$$\sigma(n) = \frac{\sigma(1)}{\sqrt{n}} \quad \text{où } \sigma(1) \text{ est l'écart standard d'une mesure unique.}$$

Déterminer l'écart standard $\sigma(n)$ en fonction du nombre n de mesures sur lesquelles la moyenne est effectuée pour les distributions suivantes.

1. Distribution de Poisson pour un comptage de désintégration.

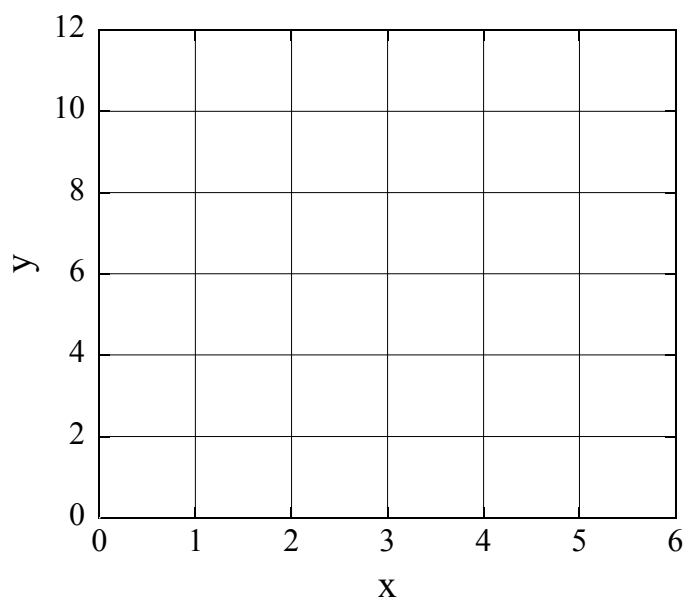
2. Distribution d'un tirage aléatoire d'un nombre compris entre 0 et 3.
3. Distribution de Gauss des résultats d'une mesure de l'accélération terrestre.

n	Poisson		Aléatoire		Gauss	
	$\sigma(n)$	$\sigma(1)/\sqrt{n}$	$\sigma(n)$	$\sigma(1)/\sqrt{n}$	$\sigma(n)$	$\sigma(1)/\sqrt{n}$
1						
2						
5						
10						
20						
50						

REGRESSION LINEAIRE:

	x	y	x^2	xy	Y	(y-Y)	(y-Y) ²
	0	2					
	1	0					
	2	4					
	3	6					
	4	6					
	5	12					
Somme							
Moyenne							

- Reporter sur le graphique les points (x,y) du tableau ci-dessus et estimer la droite la plus probable qui décrit le mieux l'ensemble des points



- Calculer les coefficients A et B de la régression linéaire pour les points (x,y) avec x comme variable indépendante.
- Comparer ce résultat avec celui obtenu en traçant visuellement la droite la plus probable.
- Comment interpréter l'écart quadratique moyen .

EXPERIENCES: ANALYSES DE DONNEES BIOSTATISTIQUES

Il s'agit de caractériser, par la méthode de la régression linéaire simple, la relation pouvant exister entre d'une part, la chaleur dissipée totale ("métabolisme au repos", "métabolisme de maintenance") et, d'autre part, certaines des variables disponibles restantes. Le travail se fait sur une base de données d'un effectif de 337 sujets (hommes et femmes). Dans cette approche, la chaleur dissipée totale Q sera toujours désignée comme la variable expliquée y, tandis que les autres variables disponibles dans le fichier (poids, âge, taille, surface cutanée, fréquence cardiaque) serviront de variables explicatives x. Pour chaque choix de variable x, le programme affiche un graphique de Q(x) pour l'ensemble des 337 sujets. Une régression linéaire est calculée et les résultats sont affichés.

- Afin de se convaincre que la droite de la régression est la plus vraisemblable, après chaque régression, il est possible de modifier les paramètres (a,b) et de faire tracer une autre droite. Constaté la différence dans les sommes de carrés résiduelles.
- Relever pour chaque variable explicite, les coefficients de corrélation trouvés par la régression et déduire de quelles variables la chaleur dégagée est dépendante.
- En augmentant le nombre de données, pensez-vous que vos conclusions sur les relations entre la chaleur dégagée et les paramètres physiologiques seront différentes.