JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

JEAN DUFRÉNOY

Résolutions graphiques et problèmes statistiques

Journal de la société statistique de Paris, tome 98 (1957), p. 126-139
<http://www.numdam.org/item?id=JSFS_1957__98__126_0>

VARIÉTÉS

Résolutions graphiques et problèmes statistiques

Introduction

« Pourquoi est-ce que je cherche
à tracer une courbe sans sinuosités? »

Cette question que se posait H. Poincaré, pourrait servir de titre à la première partie de cet exposé; lorsque nous nous efforçons de déterminer expérimentalement un alignement de points de coordonnées (x, y) c'est en général dans le but de représenter de façon aussi simple que possible la variation de la variable dépendante y en fonction de la variation de la variable indépendante x;

lorsque x représente le temps, t, « il est fréquent d'ajuster à une courbe observée une courbe présumée rendre compte de la tendance à long terme; pour éliminer les effets des variations à court terme (c'est-à-dire pour essayer d'obtenir une « courbe sans sinuosités ») il est classique de prendre les moyennes mobiles sur un certain nombre d'années ».

Le cas le plus simple est celui de l'ajustement sur une droite des moindres carrés; deux possibilités se présentent; 1° L'équation $y = a + b\,x$, comporte un terme b suffisamment grand pour imposer à la droite une pente qui caractérise la tendance, soit à la hausse soit à la baisse, ou 2° la droite est à peu près parallèle à l'axe des x (ou des t) et la dispersion des points indique seulement une fluctuation autour de la valeur moyenne $\bar{y}$.

Entre ces deux extrêmes se présentent les possibilités de faire passer, par les points expérimentaux, une courbe dont l'équation comporte, outre le terme en x, un terme en x^2.

Il est toujours possible de représenter, par un polynôme de degré p, une courbe expérimentale dont on connaît $p + 1$ points : on obtient un système linéaire dont les $p + 1$ inconnues sont les coefficients du polynôme.

Si les $p + 1$ points correspondent sur l'échelle des x (ou des t) à des espacements égaux (soit sur échelle arithmétique, soit sur échelle log) la méthode des polynômes orthogonaux permet très aisément de calculer successivement : 1° les coefficients d'un polynôme du 1er degré (définissant la tendance rectilinéaire); 2° d'un polynôme du second degré (définissant la fraction de la tendance correspondant à une courbe parabolique, ou, de façon générale, la « tendance à rendement moins que proportionnel »); et enfin, les coefficients correspondant à l'introduction successive de moments du 3e, 4e.... p ème degré, jusqu'à obtenir une courbe dont les sinuosités passent par les $p + 1$ points expérimentaux.

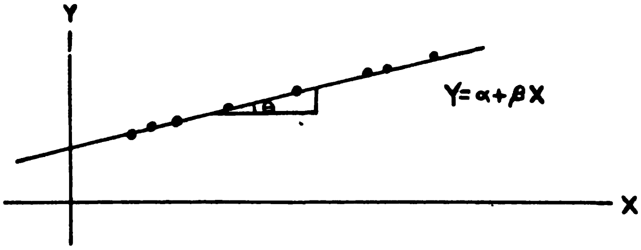
On peut également se contenter de calculer une courbe du second degré, et, par l'analyse de la variance, déterminer la part de la variance totale (correspondant à la dispersion des points expérimentaux autour de leur moyenne générale $\bar{y}$) qui relève de la régression rectilinéaire et de la régression parabolique, le résidu étant attribué aux fluctuations.

Ces fluctuations peuvent connaître une « périodicité » décelable par la technique du périodogramme, qui consiste à tester l'hypothèse de l'existence, dans la série des termes de moment supérieur à deux, d'un mouvement cyclique d'une période T déterminable (*Ann. des Mines*, IV, 209, av. 1957).

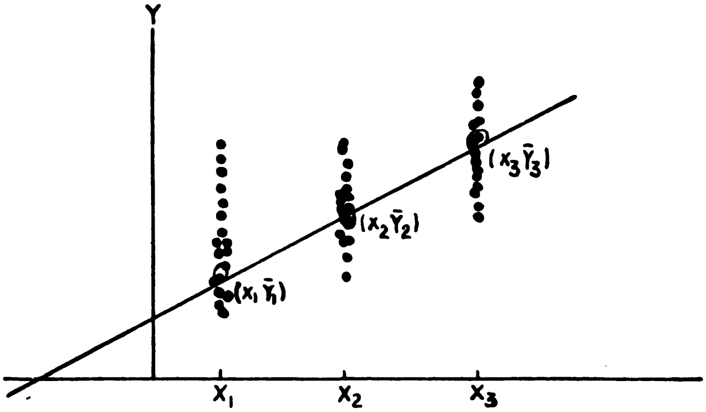
Au contraire, ce qu'il convient de superposer à la tendance (caractérisée par les termes en x et éventuellement en x^2) peut être une simple variation aléatoire; (le résidu aléatoire pouvant être additif ou multiplicatif).

On peut alors définir la variance totale, représentant l'effet de la distribution (supposée normale), des y autour de la moyenne générale $\bar{y}$, puis la variance correspondant à la distribution (supposée normale) des y d'abscisses $x_1, x_2 \dots\dots\dots x_a$, x_{p+1} par rapport aux points correspondant de la courbe de régression (droite ou parabole...).

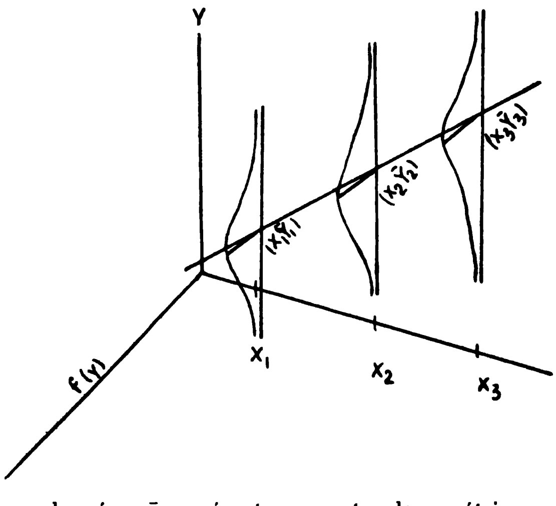
Graphiques I.



Régression linéaire simple : la variable y dépend uniquement de la variable x; tous les points de coordonnées (y, x) s'alignent sur la droite $y = \beta x + \alpha$ où $\beta = tg\ \theta = \frac{d}{dx} y$ représente le coefficient de régression.



Régression linéaire correspondant à n valeurs de y pour chaque valeur de x. Pour chaque valeur de x on détermine la moyenne $(\bar{y})$ de n valeurs correspondantes de y. C'est-à-dire un point de coordonnées x_1, $\bar{y}_1$; x_2, $\bar{y}_2$.... x_p, $\bar{y}_p$. La droite qui s'ajuste au mieux à ces p points, définit, pour chaque valeur de x, la valeur correspondante $\hat{y}$ correspondant à une régression rectilinéaire.



Chaque point de coordonnées $x\bar{y}$ représente un centre de symétrie, par rapport auquel les n valeurs de y se distribuent selon une distribution normale, qu'on peut représenter par une courbe en cloche, dans un espace à trois dimensions; les fréquences $f(y)$ sont représentées sur des axes perpendiculaires au plan défini par l'axe des x et l'axe des y.

I. — Distribution normale

La distribution normale; sa représentation graphique : la « courbe en cloche » la « sigmoïde », et son anamorphose en droite.

Sur la fig. I, chacune des trois « courbes en cloches » représente une « distribution normale » de y autour de leur moyenne d'ordonnée $\bar{y}_1$, $\bar{y}_2$, $\bar{y}_3$...

De façon générale, pour établir une courbe en cloche, on a rangé N individus par rang de taille, y_1, y_2.... y_n.

Il est classique de figurer la « courbe en cloche » comme une courbe théorique, ajustée à une distribution observée sous forme d'histogramme. Les N individus d'une population sont répartis entre un certain nombre de classes, comprenant chacune les individus compris entre une certaine valeur inférieure de y et une certaine valeur supérieure de y; chaque classe se représente alors graphiquement par une colonne dont la hauteur est proportionnelle à f (y) c'est-à-dire à la fréquence des y comptés dans cette classe; la courbe en cloche est la courbe enveloppant le polygone de fréquence constitué par l'ensemble des colonnes.

Il est d'ailleurs plus simple d'établir directement la courbe en cloche, en rangeant par exemple 100 individus par rang de taille y_1, y_2..... y_{100}, puis en représentant chacun, à son rang d'abscisses, 1, 2, ... 100, par une ordonnée y_1, y_2, ... y_{100}.

Il est généralement plus avantageux, ayant classé les 100 individus par rang de taille, d'établir la courbe intégrale (courbe sigmoïde) en représentant chaque individu de rang 1, 2, ... 100 par l'une des ordonnées cumulatives y_1, $y_1 + y_2$..., $y_{99} + y_{100}$. On obtient alors la courbe en S ou courbe sigmoïde.

Utilisations techniques des courbes sigmoïdes.

Une courbe sigmoïde est symétrique par rapport à la position 50 % : 1° Sur l'échelle verticale (où sont reportés sous forme de rapport : $\frac{(Y)}{(100 - Y)}$ les résultats des mesures ou des dénombrements); 2° Sur l'échelle horizontale, où sont reportés : *a*) les intervalles de temps (dans le cas d'une courbe de croissance); *b*) les intervalles de l'échelle des intensités (x) de la cause provoquant les effets mesurables (y). Cette position 50 % sur l'échelle verticale ou sur l'échelle horizontale correspond au K de l'équation :

$$X = K + f \log \frac{(Y)}{(100 - Y)}.$$

Chaque courbe sigmoïde est caractérisée par une valeur de K et par une valeur de f. La comparaison de la signification du f d'une courbe sigmoïde avec celle du f d'une autre courbe sigmoïde exige que, pour chacune de ces deux courbes, K soit ramené à une même valeur; on peut faire $K = 1$, c'est-à-dire adopter (K) comme unité commune, en divisant chaque f par le K correspondant pour obtenir une nouvelle caractéristique $F = f/K$.

Quel que soit le phénomène à représenter par une courbe sigmoïde, la position la plus facile à déterminer est celle qui correspond à K et qui se situe à mi-chemin (50 %) sur l'étendue utilisée de l'échelle verticale ou de l'échelle horizontale.

Le point K étant déterminé, il est facile de transformer l'S en droite :

1° en portant sur l'échelle verticale chaque valeur du rapport log (Y/100 — Y) à son rang (X — K) sur l'échelle horizontale (c'est-à-dire en prenant comme origine, sur l'échelle des abscisses, la position correspondant à K);

ou 2° en transformant toutes les valeurs mesurées Y en pourcentage du maximum, (Y) pris pour 100 %, et en portant ces pourcentages à leur rang, sur l'échelle de probabilité normale;

ou 3° en portant directement en ordonnées les valeurs mesurées, Y à leur rang d'abscisses sur échelle arc-tangente.

Nous ne considérerons ici que le deuxième procédé. Supposons que la courbe en S ait été tracée sur une feuille élastique (en caoutchouc, par exemple). Maintenons fixe l'horizontale correspondant à 50 % sur l'échelle des Y, et étirons la feuille vers le haut et vers le bas jusqu'à transformer en droite la courbe en S_1. L'échelle des Y aura dès lors été transformée en échelle de probabilité normale.

Détermination graphique de la moyenne $\overline{X}$ et de la déviation standard σ_x de n variables x distribuées normalement.

Rangées par ordre de taille croissante y_1, y_2, ... y_{100}, cent variables x se classent de la plus petite à la plus grande, sur l'échelle de probabilité normale, de manière à occuper chacune un centième de l'échelle; chaque variable se voit donc attribuer une position correspondant pour la 49e à 49,5 %, pour la 51e à 50,5 %, et ainsi de suite de proche en proche vers la gauche, jusqu'à la

N°	5	6	7	8	9	10	12	15	20	25
1	10	8,3	7,1	6,25	5,6	5	4,3	3,3	2,5	2
2	30	25	21,4	18,75	16,7	15	12,7	10	7,5	6
3	50	41,7	35,7	31,25	27,8	25	21	16,7	12,5	10
4	70	59,3	50	43,75	38,9	35	29,2	23,3	17,5	14
5	90	75	64,3	56,25	50	45	37,8	30,0	22,5	18
6		91,7	78,6	68,78	61,1	55	45,8	36,7	27,5	22
7			92,2	81,25	72,2	65	54,2	43,3	31,5	26
8				93,75	83,3	75	62,5	50	37,5	30
9					94,4	85	70,8	56,7	42,5	34
10						95	79,1	63,3	47,5	38
11							87,5	70	52,3	42
12							95,6	76,7		46
13								83,4		50
14								90		54
15								96,7		58
16										62
17										66
18										70
19										74
20										78
21										82
22										86
23										90
24										94
25										98

Référence : H. Rissik, *Probability Graph Paper : The Engineer;* 172 : 296-298, October 1941.

première qui occupe le niveau 0,5 %, et vers la droite jusqu'à la 100e qui occupe le niveau 99,5 %.

La seizième valeur se situe vers 16 % à une distance — S de la moyenne $\bar{x}$.

La 84e valeur se situe vers 84 % à une distance + S de la moyenne $\bar{x}$.

Pour 5, 6, ... 25 variables x, classées par ordre de taille croissante, les positions sur l'échelle de probabilité normale sont indiquées dans la page précédente.

Les positions du tableau précédent peuvent, pour 5, 6, 7... n variables, être repérées et matérialisées sur une règle en métal ou en plastique, par un trou dans lequel vient tomber un plongeur, ce qui permet de réaliser une sorte de règle à calcul pour la détermination automatique de la moyenne et de la déviation standard d'une distribution normale ou log normale.

Un bras vertical, portant l'échelle des ordonnées, est solidaire du plongeur.

Cinq variables, par exemple, ayant été classées par ordre de grandeur croissante, et la rainure de la règle correspondant à $n = 5$ étant disposée en regard du plongeur, le plongeur immobilise d'abord la règle au niveau 10 sur l'échelle de probabilité; on repère par un premier point l'ordonnée correspondant à la plus petite des cinq variables; on déplace le bras vertical jusqu'à ce que le plongeur l'immobilise au niveau 30 et l'on repère par un deuxième point l'ordonnée correspondante; de même pour les 3e, 4e, 5e points. Si la distribution est normale, parmi ces cinq points, on peut faire passer une droite qui intercepte l'ordonnée de niveau 50 à une valeur correspondant à la moyenne de la distribution; et les ordonnées de niveaux 16 et 84 à des distances qui, lues sur l'échelle des ordonnées, à partir de la moyenne, correspondent à $-s$ et $+s$ et mesurent la déviation standard affectant chacune des 5 variables.

Bien entendu la déviation standard de la moyenne est $s/\sqrt{5}$.

II. — Distribution log normale

Tout ce qui précède, concernant la détermination graphique de la moyenne ou de la déviation standard d'une distribution normale s'applique à une distribution log normale, par simple substitution d'échelle log à l'échelle arithmétique en ordonnées.

1° *Détermination graphique des éléments intéressant une distribution log normale (Ajustement de la droite de Henry).*

Utilisation du papier logarithmique gaussien, par exemple « Bloc précis de la Cie Fr. des Diagrammes » : gradué, pour l'axe des abscisses en log x; pour l'axe des ordonnées en 10.000 G (z).

G (z) = F (x) définit la probabilité d'avoir $X > x$.

Par exemple, dans le cas étudié par M. Leveau (*Ann. des Mines*, avr. 1957, IV, 223-7), G (z) représente le pourcentage de minerai supérieur à une teneur $X > x$; le % de métal contenu dans cette portion est G ($z - s$) où s est la racine carrée de la variance, c'est-à-dire l'erreur standard ou écart type.

Tracé de la droite : Répartir les N observations en p classes de limites décroissantes x_p, x_{p-1} . . . . . x_0.

Chaque classe telle que (x_i, x_{i-1}) compte n_i observations. La fréquence cumulée des observations supérieures à x_i est $F(x_i) = \frac{1}{N}(n_{i+1} + n_{i+2} + \cdots + n\, p)$.

Porter les valeurs de chaque $F(x_i)$ en ordonnée, pour chaque valeur de x_i portée en abscisse. Tracer la « droite de Henry » passant par les points obtenus. L'abscisse de son intersection avec l'ordonnée 5.000 (correspondant à 50 %) donne la valeur de la médiane (M) : sur le papier « Bloc précis » les graduations de l'axe des abscisses correspondent à 1 cm pour 10 log x, celles de l'axe des ordonnées à $\eta = 5\,x$; dès lors il suffit de porter dans le sens des ordonnées 11,5 cm, pour déterminer (d) en cm.

Calculer d/2 et d/4; l'erreur standard s = d/10.

En effet, si la distribution de x est log normale, $z = \frac{1}{\mu} \mathrm{L} \frac{x}{\mathrm{M}}$, la variance s^2 de Lx se détermine graphiquement

$$\text{d'après l'égalité } \mu = 10\,\frac{\mathrm{L}\,10}{2} = 11.513 \text{ cm}$$
$$\text{d'où } \mathrm{M} - \mathrm{M}_o = \mathrm{d} = 10\ s$$

La moyenne $m = \mathrm{M}\, e^{s^2/2}$ donc $\mathrm{L}\frac{m}{\mathrm{M}} = \frac{s^2}{2}$ pour $x = m$, $\eta = \frac{5}{s}\mathrm{L}\frac{m}{\mathrm{M}}$ d'où $\eta = \frac{5}{s} \times \frac{s^2}{2} = \frac{5\,s}{2} = \frac{d}{4}$

Prévisions à moyen terme.

Soit une variable u_t telle que l'on puisse écrire $u_t = \mathrm{X}.\,f(t)$; on se propose de prévoir la valeur de u aux temps $t+1$, $t+2$, ... $t+n$, connaissant u au temps t. A cette fin on étudie la variable $\frac{u_{t+1}}{u_t}$, variable sans dimension, et qui, d'ailleurs, rend compte des hausses ou baisses en %, par rapport aux années précédentes.

Un « essai de prévision à moyen terme des cours des métaux non ferreux » (Cuivre) a conduit E. Ventura (*Ann. des Mines*, IV, 205-222, av. 1957) à considérer $(u_{t+1})/u_t$ comme une variable log normale, et à poser

$$z_1 = \log(u_{t+1}) - \log u_t.$$

Prévisions. On admet que les régularités statistiques, relevées dans le passé, se poursuivront pendant les prochaines années, c'est-à-dire que les valeurs des $\mathrm{U}_t/(\mathrm{U}_{t+1})$ seront indépendantes les unes des autres et qu'elles correspondront à une loi log normale, pour une distribution de même moyenne (m) et de même déviation standard (s) que pour le passé. Posons $\mathrm{Z}_1 = \log(\mathrm{U}_{57}/\mathrm{U}_{56})$.

Les cours en 1956 étant U_{56}, les cours en 1957, 1958.... seront donnés par

$$\begin{aligned}
&\log \mathrm{U}_{57} = \log \mathrm{U}_{56} + \mathrm{Z}_1\\
&\log \mathrm{U}_{58} = \log \mathrm{U}_{57} + \mathrm{Z}_2 = \log \mathrm{U}_{56} + \mathrm{Z}_1 + \mathrm{Z}_3\\
&\cdots\cdots\\
&\log \mathrm{U}_{56+n} = \log \mathrm{U}_{56} + \mathrm{V}_n
\end{aligned}$$

V_n étant une variable normale dont la moyenne est égale à n fois la moyenne de z, et dont la déviation standard est $s_n.\!/n$

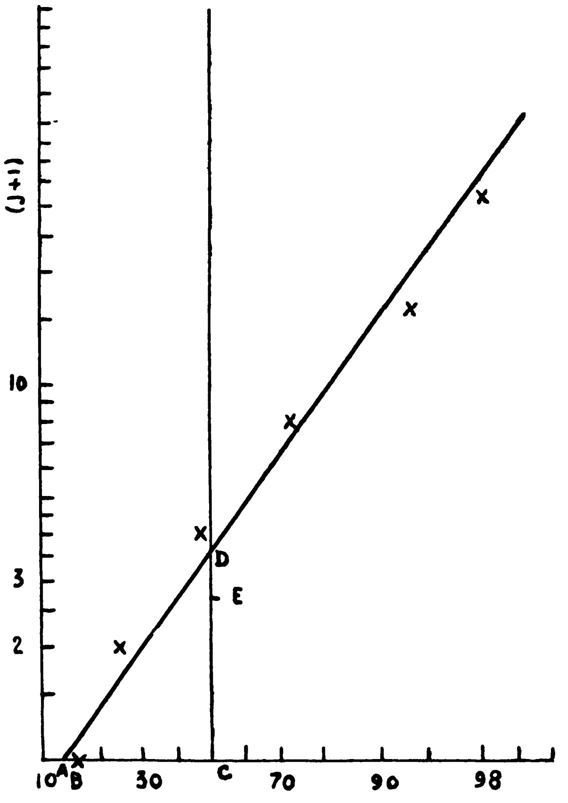
Prix moyen d'une période de 5 années à venir :

Soit S la somme des cours de 5 années successives; log (S/ 5 u_t) peut être considérée comme une variable normale.

On peut dans une intervalle à venir aussi grand que l'on veut, par exemple 1956 à 1964, déterminer k valeurs de (S/5). Sur ces k valeurs 68 % seront au plus inférieures de $-s$ ou au plus supérieures de $+s$, à la moyenne; 95 % seront comprises entre la moyenne et ± 2 s.

2° *Séries de Poisson :*

Ajustement graphique de distribution log-normale (*Taylor and Williams*, *N. Z. j. Sc. Technol.*, nov. 1956, p. 249). On dénombre les fréquences de « quadrats » comptant les nombres $j = 0, 1$, (2 ou 3) (4 à 7)... (2^{i-1} à $2^i - 1$) de



En ordonnées (échelle log, 2 cycles) « nombres de manifestations + 1 », soit (j + 1).
En abscisses (échelle de probabilité normale) pourcentages de « quadrats » montrant moins de (j + 1) manifestations (Taylor and Williams, *N. Z. J. Sc. Techn.*, p. 248, nov. 1956).

« manifestations ». Sur papier portant en ordonnées une échelle logarithmique et en abscisses l'échelle de probabilité normale (c'est-à-dire l'échelle correspondant à la distribution des erreurs cumulées) nous portons : en ordonnées les nombres ($j + 1$) et en abscisses les % de quadrats comptant moins de ($j + 1$) manifestations. Exemple :

($j + 1$)	1	2	4	8	16	32	64
%	15.7	24.1	46.3	71.3	93.5	98.1	100

Le premier point du graphique se situe sur la ligne de base, (correspondant à l'ordonnée 1 sur échelle log et au niveau de probabilité 15.7. Le deuxième point se situe à l'ordonnée 2, et au niveau de probabilité 24.1.... Le sixième point à l'ordonnée 64 sur l'échelle log, et au niveau de probabilité 98.1.

La droite passant par les 6 points intercepte l'axe vertical élevé du niveau 50 % en D et intercepte l'axe horizontal (correspondant à l'ordonnée 1 sur l'échelle log.) en A. Dans le cas présent A se situe vers 13 %; sur l'échelle des abscisses la distance entre B (situé au niveau 15.9 %), et C mesure la valeur de la déviation standard *sigma;* sur l'échelle log des ordonnées nous pouvons prendre comme unité la longueur CE = e = 2.718; les deux paramètres peuvent alors être déterminés géométriquement; moyenne : $\bar{m} = \frac{\mathrm{DC}}{\mathrm{CE}}$ $a = \frac{\mathrm{AC}}{\mathrm{CB}}$.

3° *Distribution log normale d'un indice moyen; applications.*

1° La « condition » : ce terme désigne, pour le biologiste, l'aptitude d'un animal à survivre et à se reproduire et, pour le technologiste, l'aptitude à fournir telle ou telle quantité de chair consommable, ou à recevoir telle ou telle utilisation industrielle : par exemple quatre harengs de longueur inférieure à 18 cm empliront une boîte demi-ovale, qui ne recevrait que trois harengs de longueur

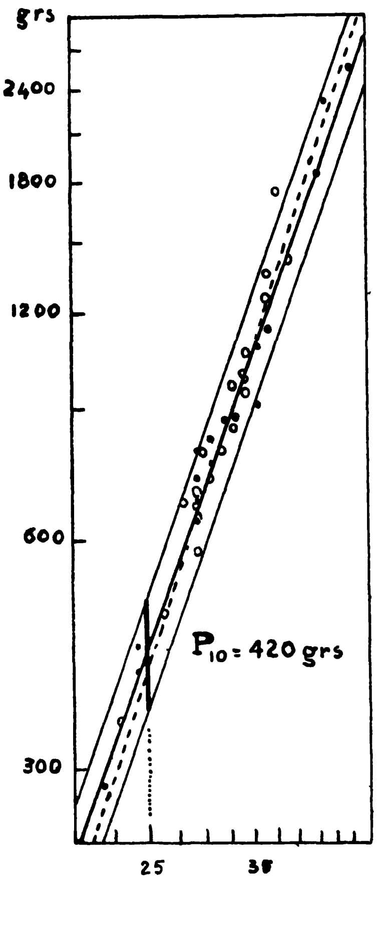


Fig. 3.
En ordonnée (échelle log) poids des poissons (en gr) en abscisses (échelle log) longueur des poissons (en cm).

comprise entre 18 et 21 cm, les harengs dépassant 21 cm ne convenant pas pour la conserve et devant être dirigés vers la salaison.

La condition peut s'évaluer en fonction d'un indice, qui met en cause une fonction de la longueur (mesure linéaire ou du « 1er degré ») et du poids, lequel, comme le volume, est une mesure « cubique »; P étant le poids, L la longueur d'un animal donné, le « facteur de condition » peut s'écrire $C = P/L^3$ R. M. CASSIE (*N. Z. J. Sc. Technol.*, jan. 1957, pp. 375-88) a étudié statistiquement et graphiquement la signification de ce « facteur » dans le cas d'un poisson (*Chrysophrys auratus*) : une espèce de poisson faisant l'objet d'une pêche industrielle représente une population soumise à échantillonnage selon un système défini, dans chaque région de pêche, et pour chaque saison, par la technique employée pour cette pêche.

Sous sa forme la plus générale le facteur de condition peut être défini par l'équation $K = P/L^b$ où l'exposant b est généralement inférieur à 3; la valeur la plus probable de b est celle du coefficient de régression dans l'équation $\log K = \log P - b \log L$.

Le facteur K permet de comparer différents échantillons entre eux, mais ne donne pas d'information sur les poids, en grammes de tel ou tel poisson considéré. Il est donc opportun d'introduire un « indice pondéral » tel que $P_{10} = 10^b K = 10^b P/L^b$ ou, sous forme logarithmique, $\log P_{10} = \log P + b\,(1 - \log L)$.

L'« indice moyen » d'une population de N individus sera donc défini par $\log P_{10} = (S \log P)/N + b\ (1 - (S \log L)/N)$. Représentation graphique : Les longueurs mesurées sont portées en abscisses sur échelle log; les poids en ordonnées sur échelle log; Cassie obtient 40 points par lesquels il fait passer une droite qui définit le coefficient de régression, de l'ordre de 2.69, (fig. 3).

Analyse statistique :

a) le coefficient de régression b, ne doit pas varier sensiblement d'un groupe à l'autre.

b) P_{10} doit connaître une distribution normale, c'est-à-dire que les valeurs de P_{10} étant portées en ordonnées sur échelle log les fréquences cumulatives des valeurs successives de P_{10} étant portées en abscisses sur « échelle des probabilités » on obtient des points s'alignant sur une droite, représentant une « distribution log normale de même moyenne et même variance que celles de la population étudiée » (fig. 4).

L'indice moyen de poissons d'une certaine espèce pourrait trouver une application dans la technique de triage automatique, en plusieurs catégories, en fonction des dimensions : selon le Brevet nº 1.134.468, du 11 avril 1957, la Soc. anon. Mather et Platt décrit un procédé qui consiste : « à déplacer, dans une direction donnée, les poissons disposés, les uns à la suite des autres, perpendiculairement à cette direction et avec une même orientation; à soumettre successivement chacun de ces poissons à l'action d'un jet d'eau... transversal à la dite direction de manière à déplacer ledit poisson dans le sens de sa longueur d'une quantité fonction de ses dimensions.... » par exemple, dans le cas du hareng destiné à la conserve il suffit de séparer quant à la longueur les poissons P_1 de moins de 18 cm, P_2 de 18 à 21, et P_3 de plus de 21. Ces longueurs correspondent à des épaisseurs maximum du hareng, qui sont, pour P_1 inférieures à

26 mm, pour P_2, entre 26 et 30 mm et pour P_3, supérieures à 30 mm; on peut donc trier les poissons par un jet réglé soit en fonction de la longueur, soit en fonction de l'épaisseur du poisson : dans le premier cas une surface de butée est

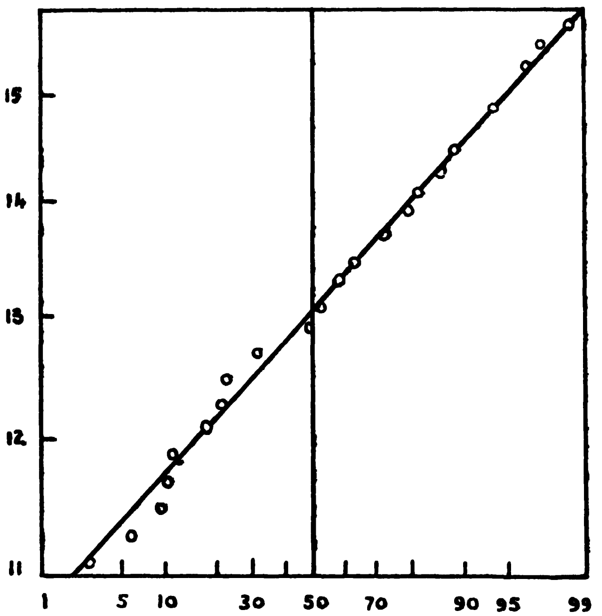


Fig. 4.

En ordonnées (échelle log) « Indices moyens » (P_{18}) en abscisses (échelle de probabilité normale) fréquences cumulées pour cent (d'après CASSIE, *N. Z. J. Techn.*, p. 381, janvier 1957).

disposée latéralement au transporteur, et chaque poisson attaqué par le jet vient buter sur cette surface par une partie déterminée de son corps; dans le deuxième, chaque poisson, projeté dans un espace cunéiforme, vient buter dans une position transversale qui dépend de son épaiseur.

III. Transformation Mosteller-Turkey

Application :

Problème : Soit une population de N Levures, dont « z » manifestent une certaine caractéristique : par exemple « levure bourgeonnant ». La probabilité de manifestation de cette caractéristique est $P = \frac{z_0}{N}$ 100 %. Si sur 2.000 Levures examinées on en trouve 80 « bourgeonnant » la probabilité P = 4 %; si nous divisons cette population de 2.000 Levures en lots de 25 Levures chacun nous pouvons espérer observer une Levure en voie de bourgeonnement dans chaque lot.

Exemple I) distribution de fréquences (x) de levures bourgeonnant par lot de 60 Levures; (y) fréquences de lots montrant (x) Levures bourgeonnant;

x	0	1	2	3	4	5	6	7	8	9
y		13	14	8	11	6	3	3	0	1

$P = \frac{187}{8393}$ 223. %.

Exemple II) Soit n le nombre de levures par champ microscopique examiné,

x;	50-69	70-89	90-109	110-129	130-149	150-169	170-189	190-209	210-229	230
k	2	2	10	14	11	10	3	3	4	1

Solution graphique; soit z le nombre de Levures bourgeonnant dans un lot de n Levures; $p = z/n$

Coordonnées : poser

$$x = \sqrt{n - z} \qquad y = \sqrt{z} \qquad \text{et } r = \sqrt{x^2 + y^2} = \sqrt{n}$$

$$x/y = \operatorname{tg} \varphi = \sqrt{z/_{(n-z)}} = \sqrt{p/_{(1-p)}}$$

$$p = \sin^2 \varphi \qquad q = 1 - p = \cos^2 \varphi.$$

Chaque point de coordonnées (x, y) est situé sur la circonférence de rayon $\sqrt{n}$ (pour chaque lot de n Levures) et sur la droite partant de l'origine des coordonnées ($\sqrt{z}$ en ordonnées et $\sqrt{n - z}$ en abscisses) et de pente définie par $\operatorname{tg} \varphi = \sqrt{p/_{(1-p)}}$.

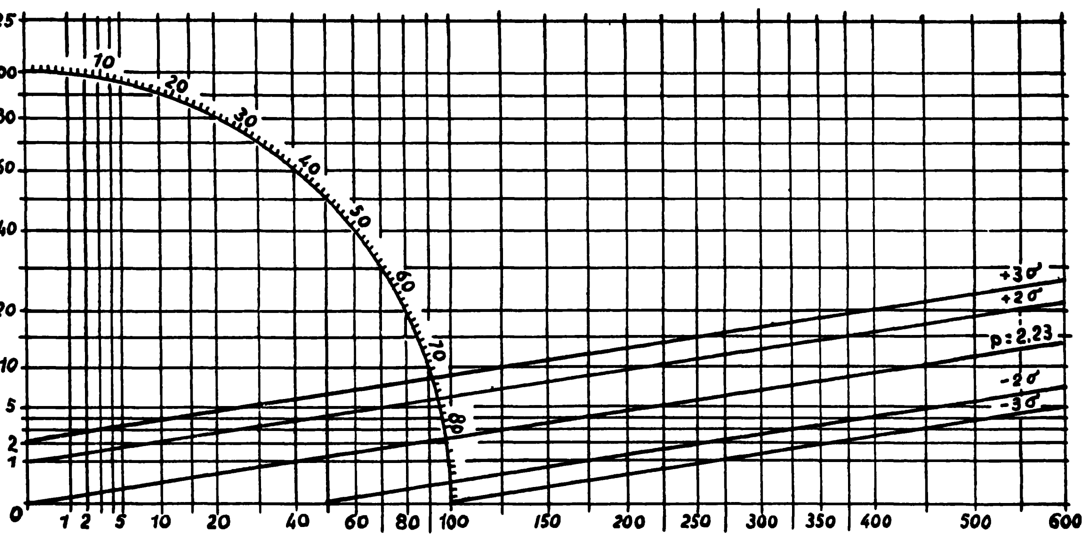


Fig. 5.

En ordonnées, z, nombre de « manifestations » (ici, levures en voie de bourgeonnement); en abscisses, n-z, nombre de non-manifestations (ici nombre de levures non-bourgeonnantes) sur échelles de « racines carrées ».

Application de la méthode de transformation Mosteller-Turkey (*J. Am. Statist. Assoc.* 44, pp.174-212, 1949).

La droite de régression de pente p = 2.8 a été tracée au travers du nuage des points expérimentaux (non figurés). En extrapolant on trouverait que cette droite intercepte par exemple l'horizontale d'ordonnée z = 28 à la verticale d'abscisse n — z = 972.

Cette droite est encadrée, au-dessus et en dessous, par les parallèles correspondant à l'écarteur + 2 sigma, (comprenant entre elles 95% des points expérimentaux), puis par les parallèles correspondant à l'écart ± 3 sigma (comprenant entre elles 99,7 % des points expérimentaux).

On obtient les différents points expérimentaux en élevant de l'abscisse (n — z) une verticale s'élevant jusqu'à l'ordonnée z sur papier de « Probabilité binomiale » (M. Hengst et G. et Ch. Koblinsky, *Die Brauerei*, 688, n° 101-102).

V. Courbes de « rendements moins que proportionnels »

Courbe d'efficacité d'un produit (utilisé pour la Production Agricole). Courbe $y = a - b\, r^x$ où r est compris entre O et 1, les valeurs de (x) formant une progression arithmétique, $x_1 - x_0 = \ldots\ x_n - x_{n-1}$.

A chaque « dose » (x) correspond un « effet » mesurable (y).

Dans le cas d'un témoin (x_0) et de 3 doses $x_1\ x_2\ x_3$

$$r = \frac{4\, y_3 + y_2 - 5\, y_1}{4\, y_2 + y_1 - 5\, y_0}.$$

En général $-\, b = \dfrac{\mathrm{S}\,(y - \bar{y})\ (r^x - \bar{r}^x)}{\mathrm{S}\,(r^x - \bar{r}^x)^2}$ où $\bar{y}$ = val. moy. de y, $\bar{r}^x$ = val. moy. de r^x.

Exemple numérique (F. Bertossi, *Not. s. Mal. delle piante*, n° 40-41, p. 178)

doses (anticryptogamiques)	x	0	1	2	3	
effets (kgs. fruits sains)	y	13	20	25	28	$\bar{y} = 21.5$

$r = \dfrac{4 \times 28 + 25 - 5 \times 20}{4 \times 25 + 20 - 5 \times 13} = 0.672$

r^x	0.672	0.672^1	0.672^2	0.672^3
	$= 1$	$= 0.672^0$	$= 0.451$	$= 0.303$

$$\bar{r}^x = \frac{1 + 0.672 + 0.451 + 0.303}{4} = 0.606$$

$$-\, b = \frac{(13 - 21{,}5)\ (1 - 0.606) + (20 - 21.5)\ (0.676 - 0.606) + \ldots\ (28 - 21.5)\ (0.303 - 0.606)}{(1 - 0.606)^2 + \ldots\ldots + (0.303 - 0.606)^2}$$

$= -\, 12.161$

L'équation générale $a = \bar{y} + b\ \bar{r}x$ devient $a = 21.5 + (12.621 \times 0.606)$ L'équation $y = a - br^x$ devient $y = 28.869 - 12.161 \times 0.672^x$ qui définit la courbe exponentielle tendant asymptotétiquement vers un plafond (mais pouvant, pour des valeurs excessives, phytotoxiques, de (x), décroître).

Remarque : Un meilleur ajustement serait obtenu avec une courbe du second degré $y = ax^2 + bx + c$, facile à calculer par la méthode des polynômes orthogonaux :

x	y	z_1	yz_1	z_2	yz_2	Ycalc.
0	13	-3	-39	$+1$	13	$21.5 - (3 \times 2.5) - 1 = 13$
1	20	-1	-20	-1	-20	$21.5 - 2.5 + 1 = 20$
2	25	$+1$	25	-1	-25	$21.5 + 2.5 + 1 = 25$
3	28	-3	-84	$+1$	28	$21.5 + (3 \times 2.5) - 1 = 28$

$y = 21.5$ $\quad$ $\mathrm{B} = 50/20 = 2.5$ $\quad$ $\mathrm{C} = -\, 4/4 = -\, 1$

$\mathrm{Y} = 21.5 + 2.5\, z_1 - z_2.$

CONCLUSIONS

Nous emprunterons notre conclusion à Paul Valéry (Eupalinos ou l'Architecte).

« Ce qu'il faut, c'est que, par une seule proposition, le mouvement soit défini de façon si précise qu'il ne reste au corps mobile d'autre liberté que de le tracer, et lui seul. Et il faut que cette proposition soit obéie de toutes les parties de ce mouvement... »

Jean Dufrénoy.

***