



HAL
open science

La Méthode statistique et la Foresterie - V - Ajustement

Léon Schaeffer

► **To cite this version:**

Léon Schaeffer. La Méthode statistique et la Foresterie - V - Ajustement. Revue forestière française, 1953, S, pp.53. <10.4267/2042/26921>. <hal-03384231>

HAL Id: hal-03384231

<https://hal.science/hal-03384231v1>

Submitted on 18 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

V. - Ajustement

Quand un observateur a pu relever un certain nombre de mesures, il s'empresse de les décrire et de les classer. Mais bien souvent il ne bornera pas là ses recherches et il aura pour ambition de traduire ses observations par une loi.

Il se demandera, par exemple, s'il n'y a pas proportionnalité entre deux séries de phénomènes. Dans ce cas, il pointera sur un graphique les différentes mesures effectuées et il tâchera de substituer à la distribution réelle une distribution théorique figurée par des points alignés. C'est ce qu'on exprime en disant qu'il ajustera une droite à une série de résultats numériques.

S'il a l'idée d'une autre liaison qui ne soit pas une loi de simple proportionnalité, mais qui puisse se représenter graphiquement par une courbe d'équation connue, ce ne sera plus une droite qu'il ajustera, mais une courbe plus ou moins compliquée : parabole, courbe de Gauss, etc...

Enfin, l'ajustement peut se faire après avoir fait subir aux données une transformation. Au lieu de porter sur le graphique des longueurs proportionnelles aux grandeurs mesurées, on peut porter des longueurs qui soient proportionnelles aux carrés de ces grandeurs, à leur logarithme ou à quelque autre fonction. On réalise ainsi une anamorphose et si, grâce à cet artifice, on est arrivé à remplacer une courbe par une droite, on a grandement simplifié le problème de l'ajustement.

Étudions successivement les trois cas qui viennent d'être cités.

RÉGRESSION LINÉAIRE

Portons sur un graphique toutes les mesures effectuées. Nous obtenons ainsi un « nuage de points ». La forme de ce nuage peut nous suggérer l'idée d'une loi de proportionnalité. Il nous faut donc tracer la droite qui s'ajuste le mieux à l'ensemble des points figuratifs (fig. 10). Nous serons en mesure de le faire si nous déterminons un point de cette droite et si nous connaissons sa pente.

Il vient tout d'abord à l'esprit que la somme des distances à la droite des points situés d'un côté soit égale à la somme des distances des points situés de l'autre côté.

$$\sum A A' = \sum B B' \text{ (fig. 11).}$$

Cela revient à dire que la droite doit passer par le centre de masses égales placées en chacun des points, ou « centre des moyennes distances ». Si M_x est la moyenne des abscisses et M_y la moyenne

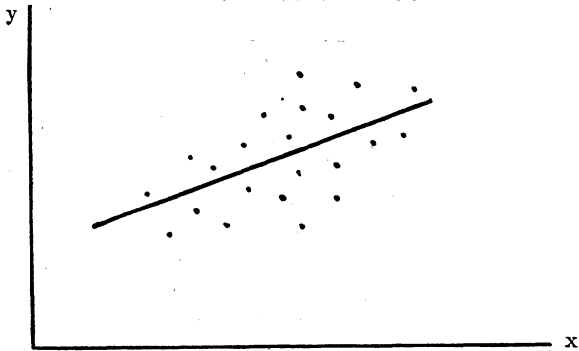


FIG. 10.

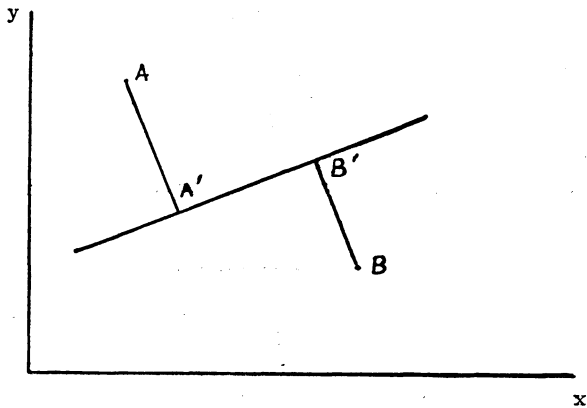


FIG. 11.

des ordonnées, la droite est assujettie à passer par le point de coordonnées (M_x, M_y) .

Il nous reste à préciser l'inclinaison de la droite. Par application de la méthode des moindres carrés, on a été amené à imposer que

la somme des carrés des distances y soit la plus petite possible. $\Sigma(PP')^2$ doit être minimum (fig. 12).

Soit m le coefficient angulaire cherché. Les coordonnées d'un point tel que P étant a et b , un calcul assez simple, mais qu'il est

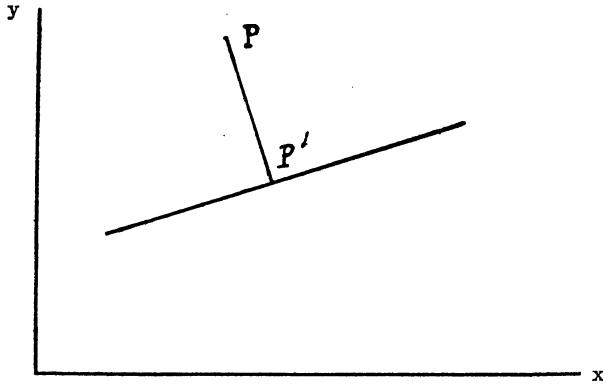


FIG. 12.

sans intérêt de reproduire ici, montre que la valeur de m est donnée par la formule

$$m = \frac{\Sigma (a - Mx) (b - My)}{\Sigma (a - Mx)^2}$$

En définitive, l'équation de la droite cherchée est :

$$Y - My = \frac{\Sigma (a - Mx) (b - My)}{\Sigma (a - Mx)^2} (X - Mx)$$

AJUSTEMENT D'UNE COURBE NORMALE A UNE DISTRIBUTION RÉELLE

Reprenons l'exemple déjà traité (p. 48) et cherchons à ajuster une courbe normale à cette distribution réelle. Dans ce but, nous allons superposer à l'histogramme (fig. 13) une courbe en cloche dont il s'agit de calculer les caractéristiques.

Nous savons déjà que la moyenne a pour valeur 7,214 et que l'écart-type est $\sigma = 2,23$. La position du sommet de la courbe est donnée par l'équation suivante qui indique la longueur de son ordonnée y_0 en fonction de σ et du nombre d'observations N :

$$y_0 = \frac{N}{\sigma \sqrt{2\pi}} = \frac{607}{2,23 \times 2,51} = 108,4$$

En utilisant l'équation de la courbe normale et en effectuant le calcul correspondant aux différentes classes de circonférence, on pourrait obtenir la longueur d'un certain nombre d'ordonnées en vue de la construction par points de la courbe.

Mais on gagnera du temps en faisant appel aux tableaux qui sont insérés dans tous les ouvrages de statistique. Sans nous attarder au détail de ce calcul, signalons simplement qu'on arriverait ainsi à tracer la courbe de la figure 13.

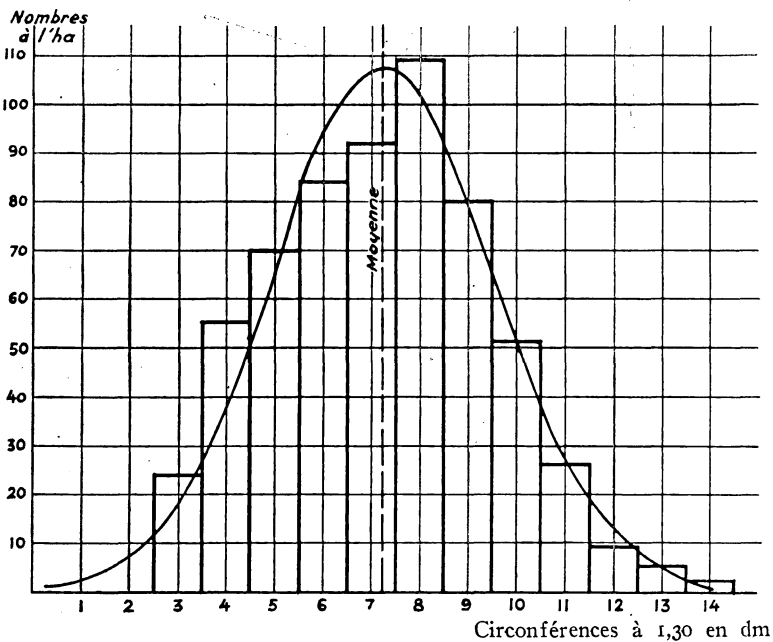


FIG. 13. — Ajustement d'une courbe normale à une distribution réelle.

DROITE DE HENRY ET PROBITS

Au lieu de considérer la fréquence dans chaque classe, il y a souvent intérêt à faire intervenir les fréquences cumulées.

Reportons-nous cette fois à l'exemple de la page 36, dans lequel nous avons la suite de nombres :

1 — 9 — 36 — 84 — 126 — 126 — 84 — 36 — 9 — 1

dont le total est 512.

Ajoutons à chaque nombre de cette suite tous ceux qui le précèdent, nous obtenons ainsi la suite :

1 — 10 — 46 — 130 — 256 — 382 — 466 — 502 — 511 — 512.

La courbe représentative de cette nouvelle suite n'est plus une courbe en cloche, mais une courbe en « S » (fig. 14).

Cette courbe a, elle aussi, été étudiée par les mathématiciens. Elle n'est pas en elle-même d'un tracé plus aisé que la courbe en cloche, mais ce qui fait ici son intérêt pour nous, c'est que, par un procédé d'anamorphose, il est facile de lui substituer une ligne droite.

Nous ne nous étendrons pas sur la « méthode des probits » décrite par LAMOTTE et pour laquelle il faut faire appel aux « statistical tables » de FISHER et YATES, mais nous mentionnerons seulement le procédé, rigoureusement identique dans le fond, imaginé par

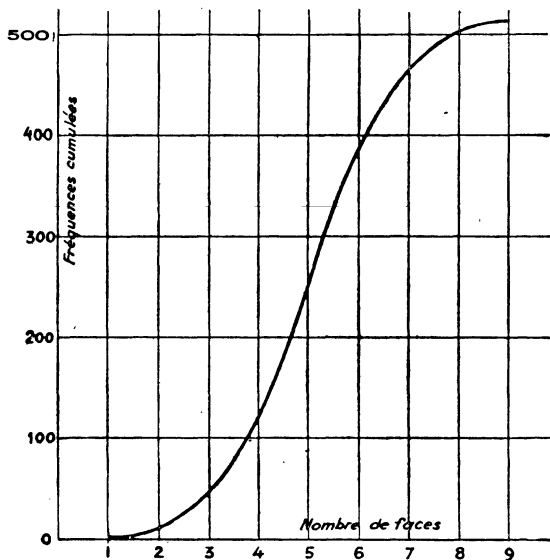


FIG. 14. — Courbe des fréquences cumulées.

HENRY, quand il était professeur de tir à l'Ecole d'application de l'artillerie à Fontainebleau.

On peut faire soi-même, mais on trouve aussi dans le commerce du papier dit « de probabilité ». En abscisses, les classes sont portées sur une échelle métrique ou arithmétique du type ordinaire. Il en est par contre tout autrement pour les ordonnées. Là, les graduations sont symétriques par rapport au trait 50 %, mais inégalement serrées : une même longueur qui, au milieu du graphique représente 10 % ne représentera plus que 1 % un peu plus loin et beaucoup moins encore quand on s'écarte davantage du milieu.

Si une série d'observations suit la loi normale, les points figura-

tifs se trouvent alignés, et la pente de la droite obtenue est inversement proportionnelle à l'écart-type.

Il en a parfois été fait de curieuses applications aux notes attribuées aux candidats lors des concours. Portons sur ce papier les notes en abscisses et les fréquences cumulées en ordonnées. Qu'il s'agisse d'une épreuve écrite ou d'un examen oral, la répartition

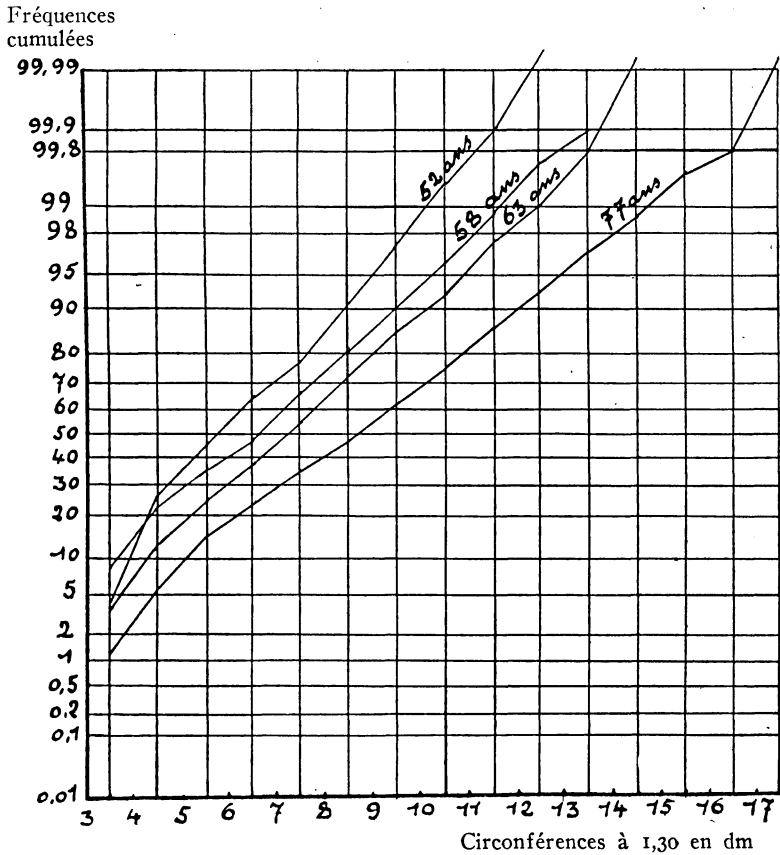


FIG. 15. — Utilisation du papier de probabilité.

des notes suit à peu près la loi normale : les points sont alignés. Mais on observe qu'à l'oral, la droite coupe le trait 50 % plus à droite et se trouve avoir une pente plus forte qu'à l'écrit. On en conclut à une sélection efficace effectuée par les examinateurs de l'écrit et se traduisant à l'oral par une amélioration générale et une moindre dispersion des notes.

TABLEAU 3

Circonférence en dm	52 ans			58 ans			63 ans			77 ans		
	Nombre de tiges	%	Fréquences cumulées (en %)	Nombre de tiges	%	Fréquences cumulées (en %)	Nombre de tiges	%	Fréquences cumulées (en %)	Nombre de tiges	%	Fréquences cumulées (en %)
3	50	4,7	4,7	56	7,0	7,0	24	4,0	4,0	6	1,1	1,1
4	210	19,9	24,6	122	15,2	22,2	55	9,1	13,1	28	5,0	6,1
5	213	20,2	44,8	117	14,5	36,7	70	11,5	24,6	54	9,6	15,7
6	202	19,2	64,0	104	12,9	49,6	84	13,8	38,4	53	9,4	25,1
7	144	13,7	77,7	135	16,8	66,4	92	15,2	53,6	65	11,6	36,7
8	138	13,1	90,8	113	14,0	80,4	109	18,0	71,6	66	11,7	48,4
9	67	6,4	97,2	90	11,2	91,6	80	13,2	84,8	79	14,1	62,5
10	22	2,1	99,3	35	4,4	96,0	51	8,4	93,2	70	12,5	75,0
11	6	0,6	99,9	22	2,7	98,7	26	4,3	97,5	62	11,0	86,0
12	1	0,1	100,0	8	1,0	99,7	9	1,5	99,0	48	8,5	94,5
13				2	0,2	99,9	5	0,8	99,8	16	2,8	97,3
14							2	0,3	100,1	7	1,2	98,5
15										6	1,1	99,6
16										1	0,2	99,8
17										1	0,2	100,0
	1 053			804			607			562		

N. B. — Les calculs n'ont pas été poussés avec un nombre suffisant de décimales pour qu'on aboutisse toujours en fin de colonne, comme le voudrait un calcul rigoureux, exactement à 100,0. En réalité, peu importe : les écarts constatés sont sans signification.

Utilisons un de ces papiers de probabilité pour y figurer l'évolution du peuplement de la place d'expériences dont il a déjà été question à plusieurs reprises. Il nous faudra reporter les fréquences observées aux différents inventaires, mais après avoir remplacé les nombres absolus de tiges par les pourcentages présents dans les différentes classes.

Reproduites simultanément, les courbes ne donneraient pas une image claire de l'évolution de cette placette. Les courbes s'enchevêtreraient et on n'aurait qu'une idée assez confuse du déplacement du sommet des courbes et de leur aplatissement progressif. Il en sera tout autrement si nous calculons les fréquences cumulées (tableau 3).

Reportant les fréquences sur le papier de probabilité, nous obtenons les 4 lignes reproduites sur la figure 15, construite en tenant compte du fait que les pourcentages cumulés correspondent aux limites supérieures de classes. Nous pouvons y lire d'une façon approchée quelles sont les valeurs médianes de la circonférence aux 4 âges considérés. La valeur de l'écart-type peut elle-même se discerner avec une certaine précision.

Considérons en particulier la répartition, déjà analysée en détail, (p. 48 et sq.) du peuplement âgé de 63 ans. Le tracé correspond à cet âge et nous montre que la médiane (50 %) est un peu en dessous de la limite supérieure de la classe 7, soit donc voisine de 7,2 ou 7,3, mettons 7,25.

D'autre part, on sait que, quand on passe de 50 % à 84 % de l'effectif, on franchit un écart-type (p. 40). Sur le graphique, on voit que pendant ce temps-là, on passe de la circonférence 7,25 à une valeur un peu inférieure à la limite supérieure de la classe 9, soit donc 9,4. Par différence, on a : $9,4 - 7,25 = 2,15$, valeur peu éloignée de la valeur trouvée par le calcul (2,23) page 51.

Notons enfin que le papier de probabilité se prête fort bien à une interpolation graphique. On reconstituerait avec une grande vraisemblance ce qu'était la répartition des tiges entre les classes de grosseurs à un âge intermédiaire tel que 60 ou 70 ans.

L'utilisation du papier de probabilité peut donc, à juste titre, être considéré comme d'un grand secours pour l'établissement des tables de production.

L. SCHAEFFER.
